UNIVERSITAT DE LLEIDA

ESCOLA TÈCNICA SUPERIOR D'ENGINYERIA AGRÀRIA

DEPARTAMENT D'ADMINISTRACIÓ D'EMPRESES I GESTIÓ ECONÒMICA DELS RECURSOS NATURALS

Tesis Doctoral

SISTEMAS DE AYUDA A LA MODELIZACIÓN DE LA PRODUCCIÓN EN LA EMPRESA AGRARIA

DIRECTOR DE TESIS:
Dr. Francisco Juárez Rubio

DOCTORANDA:
Maria Mercè Clop i Gallart

Mayo de 2000

UNIVERSITAT DE LLEIDA

ESCOLA TÈCNICA SUPERIOR D'ENGINYERIA AGRÀRIA

DEPARTAMENT D'ADMINISTRACIÓ D'EMPRESES I GESTIÓ ECONÒMICA DELS RECURSOS NATURALS

Tesis Doctoral

SISTEMAS DE AYUDA A LA MODELIZACIÓN DE LA PRODUCCIÓN EN LA EMPRESA AGRARIA

Tesis que presenta para optar al grado de Doctor Ingeniero Agrónomo Dª. Maria Mercè Clop i Gallart, bajo la dirección del Dr. D. Francisco Juárez Rubio, Catedrático de Universidad del Departament d'Administració d'Empreses i Gestió Econòmica dels Recursos Naturals de la Universitat de Lleida.

Firma de la Doctoranda

Firma del Director de Tesis

Mayo de 2000

A D. Enrique Ballestero, en agradecimiento
por los caminos que abrió y que ahora transito.

Pero el mayor honor recae en aquellos que prevén,
como muchos en efecto prevén,
que Efialtes el Traidor aparecerá finalmente,
y entonces los persas podrán
atravesar el estrecho desfiladero.

K. Kavafis ("Termopilas", 1901)

AGRADECIMIENTOS

En primer lugar, quiero agradecer al Director de la Tesis, el Dr. Francisco (Paco) Juárez, su ayuda y orientación a lo largo de la investigación, así como la gran ayuda prestada en la elaboración de los Capítulos 2 y 3.

En ciertos aspectos relacionados con la programación, que se estudiaron en un principio pero que no se decidieron incluir en el documento final, debo agradecer los consejos y recomendaciones del Dr. Janos Mayer, en cuanto a programación estocástica, y del Dr. Vicenç Sanjosé, en cuanto a conjuntos borrosos y algoritmos de cálculo con probabilidades.

A Belén de Pablo le agradezco el enlace del generador de variables aleatorias con el programa Lindo, incluido en el Anejo 9.

Al Dr. Lloyd D. Teigen debo agradecerle la bibliografía facilitada sobre estudios de funciones de densidad de rendimientos de cultivos en Estados Unidos.

La parte experimental no hubiera sido posible sin la participación de todas las personas encuestadas, a las cuales expreso mi gratitud: en primer lugar, a los agricultores "víctimas" de las entrevistas en profundidad, por su paciencia y colaboración desinteresada. En segundo lugar, a los agricultores que colaboraron en las encuestas generales, y en tercer lugar, a todos los expertos que amablemente respondieron a nuestras encuestas. Finalmente, al ingeniero agrónomo Toni Teixidó, por la realización de las entrevistas en profundidad del segundo grupo de agricultores, y a los estudiantes de Economía de la Empresa de los cursos 1998/99 y 1999/00 que colaboraron en las encuestas generales a agricultores.

Agradecer al Dr. Julio Berbel y la Dra. M. Àngels Colomer sus valiosas sugerencias, y al Dr. Joaquín Domingo y al Dr. Miquel Subirachs su ayuda puntual.

Mi más sincera gratitud hacia las personas que me han ayudado en la entrada de datos y en el aspecto final del documento, principalmente Carmina Badia, y también Belén de Pablo y Pilar Juárez.

Por último, agradecer a las personas que tengo más cerca, especialmente mi familia, su gran paciencia y apoyo incondicional.

ÍNDICE

RESUMEN – RESUM – SUMMARY

INTRODUCCIÓN

CAPÍTULO 1. PROGRAMACIÓN EN CONDICIONES DE RIESGO EN AGRICULTURA: DE LA PROGRAMACIÓN ESTOCÁSTICA A LOS MODELOS DE RIESGO

#1.1.	PROGRAMACIÓN ESTOCÁSTICA	12
#1.2.	MODELOS DE RIESGO	14
#1.3.	SEGURIDAD PRIMERO	24
#1.4.	PROGRAMACIÓN MULTIOBJETIVO	29
#1.5.	EFICIENCIA ESTOCÁSTICA	32
#1.6.	OTROS MODELOS	38

CAPÍTULO 2. PROBABILIDADES SUBJETIVAS Y RENDIMIENTOS DE CULTIVOS

#2.1.	FORMACIÓN DE LAS DISTINTAS INTERPRETACIONES DE LA PROBABILIDAD	44
#2.2.	TEORIA DE LA DECISIÓN Y UTILIDAD	46
#2.3.	MARCO CONCEPTUAL DE LAS PROBABILIDADES SUBJETIVAS	50
#2.4.	REVISIÓN DE LAS PROBABILIDADES SUBJETIVAS Y TEOREMA DE BAYES	52
#2.5.	LEYES A PRIORI, A POSTERIORI Y FUNCIONES DE VEROSIMILITUD	53
#2.6.	DETERMINACIÓN DE LAS LEYES A PRIORI	55
#2.7.	DETERMINACIONES EXPERIMENTALES DE FUNCIONES DE DENSIDAD A PRIORI	59
#2.8.	ESTIMACIÓN MEDIANTE APROXIMACIONES A FUNCIONES	64
#2.9.	CALIBRACIÓN	68
#2.10.	FUNCIONES DE DENSIDAD DEL RENDIMIENTO DE LOS CULTIVOS	73
#2.11.	FUNCIONES DE DENSIDAD SUBJETIVAS DE	75

RENDIMIENTOS

CAPÍTULO 3. HEURÍSTICAS Y SESGOS EN LA ASIGNACIÓN DE PROBABILIDADES

#3.1. APORTACIONES DE LA SICOLOGÍA 78
#3.2. LA APROXIMACIÓN COGNITIVA 80
#3.3. RESOLUCIÓN DE PROBLEMAS 82
#3.4. ASIGNACIÓN DE PROBABILIDADES 84
#3.5. EL CONCEPTO DE HEURÍSTICA 86
#3.6. HEURÍSTICA DE REPRESENTATIVIDAD 89
#3.7. ACCESIBILIDAD O DISPONIBILIDAD 101
#3.8. ANCLAJE 110

CAPÍTULO 4. PLANTEAMIENTO DEL TRABAJO EXPERIMENTAL Y PRINCIPALES RESULTADOS

#4.1. PLANTEAMIENTO DEL TRABAJO EXPERIMENTAL 114
#4.2. PROBLEMAS COMUNES EN LAS RESPUESTAS E INTERPRETACIONES 118
#4.3. CONVENCIONES COMUNES 128
#4.4. PRINCIPALES RESULTADOS DE LA INVESTIGACIÓN EXPERIMENTAL 130
#4.5. VARIABLES Y TÉRMINOS 143

CAPÍTULO 5. ENCUESTA A LOS EXPERTOS

#5.1. ENCUESTA A LOS EXPERTOS 146
#5.2. ESTABILIDAD DE LAS “RECETAS” (preguntas 1 a 5) 148
#5.3. FUNCIONES DE DENSIDAD DE RENDIMIENTOS CON INTERVALOS ORDINALES 152
#5.4. FUNCIONES DE DENSIDAD DE RENDIMIENTOS CON INTERVALOS NUMÉRICOS 155
#5.5. ESTIMACIÓN DIRECTA DE LAS MODAS Y DE LA FORMA DE LA F.D.P. 158
#5.6. APROXIMACIÓN A FUNCIONES TRIANGULARES Y BETAS 161
#5.7. PROBLEMAS EN LA INDAGACIÓN SOBRE INTERVALOS 171

NUMÉRICOS
#5.8. COHERENCIA DE LOS EXPERTOS 180
#5.9. PRINCIPALES CONCLUSIONES 183

CAPÍTULO 6. ENTREVISTAS CON AGRICULTORES

#6.1. ENTREVISTAS CON GRUPOS DE AGRICULTORES 185
#6.2. ESTIMACIÓN DE FUNCIONES DE DENSIDAD 187
#6.3. COMPARACIÓN DE VALORES MEDIOS 191
#6.4. COMPARACIÓN DE VALORES EXTREMOS 195

CAPÍTULO 7. ENCUESTAS GENERALES A LOS AGRICULTORES

#7.1. ENCUESTAS A LOS AGRICULTORES 198
#7.2. ENCUESTA DEL PRIMER AÑO (1998 – 1999) 198
#7.3. APROXIMACIONES A FUNCIONES TRIANGULARES Y BETAS 204
#7.4. FUNCIONES DE DENSIDAD DE RENDIMIENTOS CON INTERVALOS ORDINALES 215
#7.5. FUNCIONES DE DENSIDAD DE RENDIMIENTOS CON INTERVALOS NUMÉRICOS 221
#7.6. PROBLEMAS EN LA DETERMINACIÓN DE INTERVALOS 229
#7.7. ANÁLISIS DE LOS RENDIMIENTOS NORMALES Y MÁS FRECUENTES 248
#7.8. COHERENCIA DE LAS ESTIMACIONES 250

CAPÍTULO 8. ENCUESTAS GENERALES A LOS AGRICULTORES. ENCUESTA DEL SEGUNDO AÑO (1999 – 2000)

#8.1. PLANTEAMIENTO 264
#8.2. RESULTADOS 265
#8.3. APROXIMACIONES A FUNCIONES TRIANGULARES Y BETAS 269
#8.4. FUNCIONES DE DENSIDAD DE RENDIMIENTOS CON INTERVALOS ORDINALES 280
#8.5. FUNCIONES DE DENSIDAD DE RENDIMIENTOS CON 285

INTERVALOS NUMÉRICOS
#8.6. PROBLEMAS EN LA DETERMINACIÓN DE INTERVALOS 289
#8.7. ANÁLISIS DEL CENTRO DEL RECORRIDO 308
#8.8. COHERENCIA DE LAS ESTIMACIONES 309

ANEJOS

ANEJO 1: ENCUESTA A LOS EXPERTOS 324
ANEJO 2: ENCUESTA EN PROFUNDIDAD A LOS AGRICULTORES 331
ANEJO 3: ENCUESTA EN PROFUNDIDAD A LOS AGRICULTORES. COMPARACIÓN GRÁFICA DE LOS RESULTADOS DE LOS DIAS 1 Y 2 333
ANEJO 4: ENCUESTA EN PROFUNDIDAD A LOS AGRICULTORES. COMPARACIÓN DE TODOS LOS AGRICULTORES POR DÍA 343
ANEJO 5: ENCUESTA EN PROFUNDIDAD A LOS AGRICULTORES. COMPARACIÓN DE LOS RENDIMIENTOS DEL 2º DÍA Y PARTICIÓN DE INTERVALOS 347
ANEJO 6: ENCUESTA EN PROFUNDIDAD A LOS AGRICULTORES. GRÁFICAS DE LOS RESULTADOS DE SIMULACIÓN 358
ANEJO 7: ENCUESTA A LOS AGRICULTORES REALIZADA POR ESTUDIANTES EN EL CURSO 1998-1999 366
ANEJO 8: ENCUESTA A LOS AGRICULTORES REALIZADA POR ESTUDIANTES EN EL CURSO 1999-2000 376
ANEJO 9: PROGRAMA FUENTE DE SIMULACIÓN – PROGRAMACIÓN LINEAL 386

BIBLIOGRAFÍA 395

RESUMEN

En los próximos años se emprenderá el desarrollo de sistemas de ayuda en la toma de decisiones (SATD) que aconsejen a los agricultores en la programación en condiciones de riesgo y de incertidumbre. Esta es una cuestión ampliamente examinada en la Economía Agraria, donde se han realizado importantes avances. Por ello, en el Capítulo 1, se han revisado las principales corrientes teóricas y los modelos propuestos, a fin de hacer balance de las opciones disponibles, de los supuestos implícitos, de los principales problemas y de las exigencias de programación. Posiblemente, para los objetivos a los que pretende contribuir esta Tesis Doctoral, la principal conclusión de la revisión sea que las mayores necesidades de desarrollo son conceptuales, ya que una integración de la programación lineal y de los simuladores aleatorios (como la que hemos desarrollado en el Anejo 9), permite disponer de un núcleo de cálculo suficiente.

En muchos casos la función de densidad de probabilidad (f.d.p.) de las variables aleatorias de los modelos no sólo será desconocida, sino que será imposible determinarla o no tendrá interés práctico plantear esa tarea, por lo que la programación se tendrá que realizar con probabilidades subjetivas, en un contexto estadístico bayesiano. En la literatura económica agraria los estudios sobre determinación de probabilidades subjetivas no han sido muy numerosos, por lo que existe una gran necesidad de experimentación en este área de actividad científica. Debido a ello, en el Capítulo 2, se han revisado las principales líneas de desarrollo de la Teoría Personalista de la Probabilidad, y en relación más directa con los objetivos inmediatos de la investigación, las principales conclusiones en torno a la forma de las f.d.p. de rendimientos de cultivos y la estimación subjetiva de esas funciones.

Un programa que obtenga la información mediante preguntas a uno o varios seres humanos, debe tener en cuenta la tendencia de los sujetos a cometer errores o sesgos sistemáticos, permanentes y difícilmente corregibles (si es que es posible la corrección, existiendo amplísima evidencia empírica contraria). Este tema ha sido escasamente tratado en la literatura económica, por lo que en el Capítulo 3 se han revisado las principales formulaciones, desarrolladas en la Sicología.

A partir del Capítulo 4 se ha discutido el trabajo experimental, que ha consistido en la realización de diversas encuestas: a expertos en ámbitos próximos a la Fitotecnia; a un grupo de diez agricultores, entrevistados en profundidad; y a un amplio grupo de agricultores. Las encuestas trataban de identificar problemas en los métodos de estimación de funciones de densidad de probabilidades de rendimientos de cultivos a partir de la información subjetiva de los sujetos.

En el capítulo 4 se han discutido los problemas encontrados en la investigación y, sobre todo, los principales resultados de la misma, a fin de subrayar el objetivo principal de la investigación: es una exploración muy básica sobre problemas de interpretación en los diferentes métodos de estimación de las f.d.p. mediante interrogación a sujetos, que pretende identificar problemas y conjeturas antes que obtener conclusiones.

Los métodos estudiados de estimación de la f.d.p. son: (1) estimaciones puntuales de los rendimientos mínimos, más frecuentes y máximos; (2) atribución de frecuencias a intervalos ordinales, (lenguaje natural); y (3) traducción directa a una escala numérica. Los resultados más interesantes tienen que ver con: (a) estimación coherente de valores medios con la aproximación beta – PERT, superior a la triangular, y posible contraste de los recorridos y de la forma de la f.d.p.; (b) coherencia y buena estimación de la forma de las f.d.p. en escalas ordinales; (c) dificultades de interpretación de las escalas ordinales (que se entienden en sus significados materiales antes que en su función en la descripción de intervalos), y dificultad en la traducción a escalas numéricas, especialmente de los rendimientos extremos; (d) en general, una descripción suficiente coherente de las f.d.p., en relación a la escasa sofisticación de los métodos empleados; y (e) posibles sesgos en direcciones contrarias de los métodos que utilizan escalas ordinales y numéricas. Es importante subrayar la constatación de problemas relacionados con los significados lingüísticos.

Los buenos resultados experimentales obtenidos con técnicas que pueden considerarse una aproximación preliminar al problema central discutido, alientan a perfeccionar esas técnicas, a fin de obtener una metodología de obtención de información de los sujetos que pueda utilizarse de forma eficaz en los SATD.

RESUM

En els propers anys s'emprendrà el desenvolupament de sistemes d'ajut a la presa de decisions (SATD) que aconsellin els agricultors en la programació en condicions de risc i d'incertesa. Aquesta és una qüestió amplament examinada dins l'Economia Agrària, on s'han realitzat importants avenços. És per això que, en el Capítol 1, s'han revisat els principals corrents teòrics i els models proposats, per tal de fer balanç de les opcions disponibles, dels supòsits implícits, dels principals problemes i de les exigències de programació. Possiblement, per als objectius als que pretén contribuir aquesta Tesi Doctoral, la principal conclusió de la revisió sigui que les majors necessitats de desenvolupament són conceptuals, donat què una integració de la programació lineal i dels simuladors aleatoris (com la que hem desenvolupat a l'Annex 9), permet disposar d'un nucli de càlcul suficient.

En molts casos la funció de densitat de probabilitat (f.d.p.) de les variables aleatòries dels models no només serà desconeguda, sinó que serà impossible determinar-la o no tindrà interès pràctic plantejar aquesta tasca, per la qual cosa la programació s'haurà de realitzar amb probabilitats subjectives, en un context estadístic baiesià. En la literatura econòmica agrària els estudis sobre determinació de probabilitats subjectives no han estat massa nombrosos, per la qual cosa hi ha una necessitat d'experimentació en aquesta àrea d'activitat científica. És per això que, en el Capítol 2, s'hi han revisat les principals línies de desenvolupament de la Teoria Personalista de la Probabilitat, i en relació més directa amb els objectius immediats de la investigació, les principals conclusions al voltant de la forma de les f.d.p. de rendiments de cultius i l'estimació subjectiva d'aquelles funcions.

Un programa que obtingui la informació mitjançant preguntes a un o diversos éssers humans, ha de tenir en compte la tendència dels subjectes a cometre errors o biaixos sistemàtics, permanents i difícilment corregibles (si és que és possible la correcció, existint amplíssima evidència empírica contrària). Aquest tema ha estat escassament tractat a la literatura econòmica, per la qual cosa en el Capítol 3 s'hi han revisat les principals formulacions, desenvolupades a la Psicologia.

A partir del Capítol 4 s'ha discutit el treball experimental, que ha consistit a la realització de diverses enquestes: a experts en àmbits propers a la Fitotècnia; a un grup de deu agricultors, entrevistats en profunditat; i a un grup ampli d'agricultors. Les enquestes tractaven d'identificar problemes en els mètodes d'estimació de funcions de densitat de probabilitats de rendiments de cultius a partir de la informació subjectiva dels subjectes.

En el Capítol 4 s'hi han discutit els problemes trobats en la investigació i, sobretot, els principals resultats de la mateixa, per tal de subratllar l'objectiu principal de la investigació: és una exploració molt bàsica sobre problemes d'interpretació en els diferents mètodes d'estimació de les f.d.p. mitjançant interrogació a subjectes, que pretén identificar problemes i conjectures abans que obtenir conclusions.

Els mètodes estudiats d'estimació de la f.d.p. són: (1) estimacions puntuals dels rendiments mínims, més freqüents i màxims; (2) atribució de freqüències a intervals ordinals (llenguatge natural); i (3) traducció directa a una escala numèrica. Els resultats més interessants estan relacionats amb: (a) estimació coherent de valors mitjans amb l'aproximació beta – PERT, superior a la triangular, i possiblement dels recorreguts i de la forma de la f.d.p.; (b) coherència i bona estimació de la forma de les f.d.p. en escales ordinals; (c) dificultats d'interpretació de les escales ordinals (que s'entenen en els seus significats materials abans que en la seva funció en la descripció d'intervals), i dificultat en la traducció a escales numèriques, especialment dels rendiments extrems; (d) en general, una descripció suficient coherent de les f.d.p. en relació a l'escassa sofisticació dels mètodes emprats; i (e) possibles biaixos en direccions contràries dels mètodes que utilitzen escales ordinals i numèriques. És important subratllar la constatació de problemes relacionats amb els significats lingüístics.

Els bons resultats experimentals obtinguts amb tècniques que poden considerar-se una aproximació preliminar al problema central discutit, encoratgen a perfeccionar aquelles tècniques, per tal d'obtenir una metodologia d'obtenció d'informació dels subjectes que pugui utilitzar-se de forma eficaç en els SATD.

SUMMARY

The development of decision making support systems (SATD) will be undertaken in the forthcoming years, in order to advise farmers in programming for conditions of risk and uncertainty. This issue has been widely assessed in Agrarian Economy, and has had an important progress. This is why Chapter 1 contains a revision of the main theoretical trends and proposed models in order to make a balance of the available options, the implicit assumptions, the main problems and programming demands. According to the objectives this PhD Thesis envisages to contribute to, the main conclusion of the revision is that development needs may be essentially conceptual, since an integration of linear programming and random simulators (like the one developed in Annex 9) allows to have enough calculation nucleus.

In many cases, the density function of probability (f.d.p.) of the models random variables will not only be unknown, but also impossible to determine, or will be of no practical interest, then programming will have to be done with subjective probabilities in a Bayesian statistical context. In agrarian economic literature, studies on subjective probability determination have not been numerous, so there is a need for experimentation in this area of scientific activity. This is the reason why in Chapter 2, the main lines of development of the Personalistic Theory of Probability have been revised, and according to the most direct relation with the immediate research objectives, the main conclusions related to the crop yields f.d.p. shape and the subjective estimation of those functions.

A program getting information out of questioning one or more individuals needs to take into account the individuals tendency to commit permanent, systematic and difficult to correct mistakes or bias (if correction is possible when there is wide empirical evidence of the contrary). This issue has hardly been studied in economic literature, therefore, Chapter 3 contains a revision of its main formulations developed in Psychology.

From Chapter 4 onwards, experimental work is discussed. It consisted of different surveys: to the experts in fields related to Phytotechnics; a deeper survey to a group of ten farmers; and to a numerous group of farmers. The surveys sought to identify problems in the estimation methods of yield probability density functions out of the subjective information of the individuals.

In Chapter 4 problems found in the research have been discussed, and, especially, the main results obtained, in order to emphasize the main objective of the research: that is a basic exploration of the interpretation problems in the different estimation methods of the f.d.p. by means of individuals questioning, that attempts to identify problems and conjectures than obtaining conclusions.

The assessed methods of f.d.p. estimation are: (1) punctual estimations through minimum, most frequent, and maximum yields; (2) frequencies attribution to ordinary intervals (natural language); and (3) direct translation into a numerical scale. The most interesting results are related to: (a) a coherent estimation of average values with a beta - PERT approximation, superior to the triangular, and contrast possible with f.d.p. ranges and shape; (b) coherence and good estimation of the f.d.p. shape in ordinal scales; (c) difficulties in interpreting ordinal scales (understood before in their material significance than in their function in describing intervals) , and difficulties in translation into numerical scales, specially of extreme yields; (d) in general, a coherent sufficient description of f.d.p. in report to the little sophistication of the methods used; and (e) possible bias in opposite directions of the methods that use ordinal and numerical scales. The problems related to linguistic meanings are remarkable.

The good experimental results obtained with techniques that can be considered as a preliminary approximation to the core problem discussed, are encouraging to perfect those techniques, so that a methodology for obtaining information out of the individuals that can be efficiently used in SATD can be achieved.

INTRODUCCIÓN

En los últimos años se han realizado importantes progresos en el desarrollo y difusión de sistemas de tratamiento y transmisión de la información, tanto en lo que se refiere a los aspectos físicos (hardware) como lógicos (software). Las amplias opciones disponibles, y su abaratamiento, permiten imaginar un futuro próximo en el que la formación y la toma de decisiones, en extensos dominios de la acción humana, se apoyen en esas tecnologías.

En cada ámbito específico habrá que evaluar los requisitos necesarios para desarrollar los sistemas de ayuda en la toma de decisiones (SATD) que se anuncian. La cuestión no será sencilla, porque los sistemas mostrarán de una forma más notable que actualmente, con la experiencia, los límites de los conocimientos disponibles y las dificultades de su integración en un sistema coherente capaz de interpretar la información adquirida y deducir una acción.

En el ámbito de la formación, los resultados de las tecnologías actuales son francamente positivos, aunque su utilización no se haya generalizado todavía. La mera comunicación de información, semejante en su pasividad a los textos actuales, es una tecnología prácticamente gratuita y cada día son más los títulos disponibles. Los sistemas más sofisticados, que integran interactividad en los textos, gráficos, videos y audio, se encuentran disponibles a precios inimaginables hace una década. Todos estos sistemas se adaptan perfectamente a las nuevas tecnologías de distribución de la información en red. Cualquier actividad que exija un entrenamiento previo es susceptible de ser expuesta y simulada digitalmente, frecuentemente de forma más eficaz que mediante los instrumentos, la maquinaria, las instalaciones o las instituciones reales.

En el ámbito de los sistemas de cálculo y de representación se dispone también de un amplio repertorio de programas (hojas de cálculo, bases de datos, programación, simulación, estimación estadística, representación gráfica, GIS, CAD, etc.). Posiblemente se disponga de instrumentos lógicos suficientes para la aplicación automática de los modelos de decisión desarrollados en las últimas cinco décadas.

Desde el punto de vista de los instrumentos lógicos (software), su rápido desarrollo y mejora es, paradójicamente, una de las principales fuentes de incertidumbre a la hora de imaginar un proyecto ambicioso de desarrollo de sistemas de tomas de decisiones, debido a la rapidez con que los diferentes programas se ven afectados por la obsolescencia. Sin embargo es preciso reconocer que el principal obstáculo para el desarrollo de sistemas eficaces de ayuda en la toma de decisiones no se derivará en los próximos años de la capacidad de cálculo disponible (hardware y software), ni de su coste, sino de la dificultad de diseñar algo parecido a un sujeto virtual que sea competente en el ámbito de decisión en el que se intente desarrollar el correspondiente sistema.

Cuando los problemas a resolver tienen una estructura muy definida, y los protocolos de comunicación con el decisor se han normalizado (es decir, la tarea a resolver es muy mecánica y existe una formación formal muy codificada), se han desarrollado sin mayores problemas programas para la aplicación de los modelos. Así, por ejemplo, se dispone de programas capaces de calcular de forma muy competitiva con los seres humanos la estructura de edificios, el camino crítico de un proyecto o el diseño de un sistema industrial de síntesis química, incluso en condiciones de aleatoriedad, etc. Con la utilización de técnicas de inteligencia artificial o de sistemas expertos, los programas pueden incluso ser eficaces en el diagnóstico de enfermedades, en el diseño de políticas fiscales en las empresas o en la simulación de procesos de producción por lotes, etc.

Pese a la innegable capacidad de esos programas, es evidente que los límites en los que se puede obtener ayuda eficaz en la toma de decisiones se encuentran con relativa facilidad, y coinciden frecuentemente con problemas lingüísticos o de interpretación o con limitaciones de las teorías o de los modelos de decisión desarrollados a partir de las mismas. Por ello es importante revisar los diferentes campos de conocimiento desde la perspectiva del desarrollo de sistemas de ayuda, para tomar conciencia de dónde se encuentran los principales problemas, las restricciones esenciales en el desarrollo de los SATD.

Las ventajas prácticas de los sistemas de ayuda a la toma de decisiones en la producción agraria son evidentes, y los inconvenientes son difíciles de imaginar. Un inconveniente importante que se puede conjeturar, en una primera aproximación, es que la utilización de esos sistemas podría ofrecer las mismas soluciones a un conjunto muy amplio de usuarios, a partir de la información disponible en el momento de tomar la decisión, y como consecuencia desestabilizar todo el sistema de precios, contribuyendo a la creación de ciclos como los descritos en el teorema de la telaraña. Sin embargo es más verosímil que los programas contribuyan a conseguir una distribución más adecuada de las actividades en función de las ventajas comparativas de que se disponga de los diferentes insumos, incluidos los aspectos espaciales y medioambientales. Por ello se ha considerado que vale la pena avanzar en la definición y la resolución de problemas en esa dirección.

LA PROGRAMACIÓN DE CULTIVOS

En esta Tesis Doctoral se ha partido de un problema aparentemente elemental, como es el de la programación de cultivos, un subconjunto del problema más general de programación de la producción en las explotaciones agrarias. Es un problema que no se puede considerar resuelto satisfactoriamente, aunque se hayan producido desarrollos conceptuales valiosos y adaptado instrumentos de cálculo sofisticados. La prueba más evidente es que la modelización nacida en los centros de investigación no ha sustituido a los sistemas, aparentemente simples o arbitrarios, empleados tradicionalmente por los agricultores para adoptar sus decisiones de producción.

Es posible imaginar un sistema de información, de bajo coste, que combine bases de datos, programas de cálculo y técnicas como GPS y GIS, en el que se definan con precisión los límites y la topografía de las parcelas, las características edáficas, el registro de operaciones realizadas sobre los cultivos, las condiciones climáticas que se presentaron, los resultados obtenidos en cada zona (p.e. dotando de GPS a las maquinas recolectoras), etc. Incluso, en un paso más, podemos imaginar la adquisición de información sobre condiciones de mercado, legales, fiscales o de otro tipo a través de las redes de comunicación. Las relaciones entre insumos y productos, la gestión de almacenes, las políticas de renovación de inmovilizados, el control de calidad u otros resultados, como comparaciones entre la calidad de los insumos, los precios de diferentes proveedores, etc. podrían ser estadísticamente tratadas de forma casi automática. El registro contable de las diferentes operaciones, la gestión de tesorería, el pago de nóminas y a proveedores, la emisión de facturas, la liquidación de tributos, etc., completarían las prestaciones disponibles. Algunos de los programas comerciales disponibles han avanzado en varios de estos aspectos, y su integración no parece que deba encontrar problemas especialmente interesantes.

El siguiente paso en el desarrollo de sistemas de ayuda en la toma de decisiones parece que debería ser incorporar capacidades para planificar la producción, partiendo de la información interna así generada, y de la externa, derivada de los mercados, de la investigación y de la experimentación. Se conseguiría de esta forma una vieja aspiración de los círculos académicos que trabajaron en la formulación de los modelos: transformarlos en herramientas auxiliares para la toma de decisiones en las explotaciones agrarias.

Son muchos los problemas que se deberán resolver para avanzar en esa dirección, y en este trabajo se han analizado problemas relacionados con una cuestión clásica: **las necesidades derivadas del núcleo de la programación matemática.**

Un primer problema tiene que ver con la clase de programación matemática a incorporar al núcleo del programa.

Desde las primeras formulaciones del modelo de programación de la producción agraria se identificó el contexto de riesgo como el relevante en la formulación y la resolución de los problemas, inspirando un gran número de trabajos teóricos y experimentales bien conocidos. Aunque las primeras modelizaciones, basadas en la propuesta de Markowitz, se plantearon como programas cuadráticos, avances sucesivos, a partir de Hazell, recuperaron la programación lineal. Es razonable, por tanto, conjeturar que un hipotético sistema de ayuda a la toma de decisiones en la programación de cultivos no exigirá un algoritmo de cálculo diferente al de la programación lineal, del que existen versiones comerciales suficientes a precios razonables. Aunque, en general, no se podrá suponer normalidad en las f.d.p. de rendimientos, ni probablemente en otras derivadas de ellas, como las de ingresos, esta circunstancia no contradice la idea de que los modelos lineales son una opción suficiente, siempre que se disponga de simuladores aleatorios integrados con la programación. Una combinación de programación lineal y de sistemas de generación de valores de variables aleatorias mediante simulación dotará al núcleo de suficiente capacidad de cálculo para abordar un amplio grupo de modelos de programación.

A esta cuestión se han referido varios trabajos en la literatura, en muchos casos con el desarrollo de programas, limitados al caso concreto discutido. Para comprobar la viabilidad de esta idea con los actuales programas comerciales se ha desarrollado una integración que se ha descrito en el Anejo 9, y que no se discute con mayor profundidad, porque consiste en la constatación de la factibilidad de esas soluciones.

Mayor problema será dilucidar el enfoque más conveniente, el modelo en torno al que se organizará el núcleo de la programación. Las diferentes aproximaciones han sido discutidas con detalle en la revisión de la literatura realizada en el Capítulo 1.

En este ámbito los desarrollos existentes en la literatura económica agraria son francamente abundantes, por lo que nuestro interés se ha centrado en determinar los problemas de cálculo y de necesidad de información que precisa cada modelo, sus supuestos implícitos, sus limitaciones conceptuales, etc. Podemos aceptar, al menos provisionalmente, que no serán las necesidades de programación matemática en el núcleo del SATD el principal obstáculo en el desarrollo futuro del modelo más adecuado, y que la herramienta básica de programación puede ser un programa lineal más un simulador aleatorio. No podemos conjeturar por ahora la forma que adoptará el modelo empleado en el núcleo.

FUNCIONES DE DENSIDAD DE LOS RENDIMIENTOS

Otro problema general consiste en determinar cómo alimentar el núcleo lineal con la información relevante disponible en cada caso.

El modelo lineal básico se puede formular como:

$$max \sum_{i=1}^{n} CX$$

$$s.a.$$

$$A\,X \leq B$$

$$X \geq 0$$

donde las $c_j \in C$ se pueden interpretar como el resultado obtenido con cada actividad j (por ejemplo, el margen bruto del cultivo j), las $x_j \in X$ como la amplitud de cada actividad (p.e. la superficie dedicada al cultivo j), $a_{ij} \in A$ los coeficientes técnicos que expresan la cantidad del recurso i necesaria por cada unidad de actividad j, y $b_i \in B$ la disponibilidad de recursos (en este conjunto se incluyen también restricciones derivadas de aspectos legales, institucionales, de mercado, reglas de sucesión, etc.).

En un problema en condiciones de riesgo, cualquiera de los coeficientes de C, A o B puede ser una variable aleatoria, y para obtener las correspondientes distribuciones de las soluciones X es necesario conocer la distribución de probabilidad de esos coeficientes aleatorios.

Una primera dificultad de alimentar con la información relevante el núcleo de programación será que en la mayoría de los casos de interés, no se conocerán las **funciones de densidad de probabilidad** de las variables aleatorias, no en el sentido de que no se conocen porque no han sido todavía determinadas por la experimentación, sino en el de que **no es posible determinarlas o que no tiene interés práctico su determinación.**

Para ilustrar esta tesis supongamos que sean aleatorias las variables c_j, por ejemplo, que los márgenes brutos sean aleatorios. En el caso más sencillo, en el que los precios sean relativamente estables, gran parte de la aleatoriedad de c_j estará determinada por la variabilidad de los rendimientos físicos de cada actividad o cultivo. Aparentemente la programación podría basarse en la estimación de las f.d.p. de los diferentes cultivos, un programa de investigación al que, en los ámbitos fitotécnicos, se ha prestado una importante atención.

Si se revisa la literatura, por ejemplo J.W. Pease (1992) o B. K. Goodwin y A. P. Ker (1998), se observará que ésta no es una tarea elemental. De hecho es significativo que en la literatura económica agraria no se disponga de funciones de densidad para los cultivos más frecuentes y generalizados, tal como han puesto de manifiesto Just y Weninger (1999). Sin embargo, los modelos de decisión en contexto de riesgo, revisados en el Capítulo 1, suponen implícitamente que se conoce la función de densidad de las variables aleatorias o que alguna variable relevante se distribuye uniformemente. Éstas son hipótesis que seguramente no se verifican en la mayoría de los casos de programación de cultivos.

Posiblemente la variable – raíz de los modelos de programación de la producción sea el rendimiento de los distintos cultivos en función de las características físico – químicas o edáficas de las parcelas, de la climatología, de los insumos y las técnicas empleados, etc. Pero esta variable tan sencilla y básica es una gran desconocida en su comportamiento en el contexto de riesgo.

Existe una amplia literatura que ha tratado de estimar:

- simplemente la forma de la función de densidad de probabilidad de los rendimientos;

- las variables que determinan la variabilidad de los rendimientos de los cultivos (climáticas, técnicas, económicas, etc.);

- las relaciones funcionales entre la aleatoriedad de las variables causales y la de los rendimientos.

Sería ocioso decir que en las dos últimas líneas de investigación no se disponen de resultados útiles para ser incorporados a un modelo de toma de decisiones. Incluso en la simple determinación de la función de densidad de probabilidad de los rendimientos de los cultivos, se dispone de escasos resultados que puedan ser aprovechados en la planificación (Just y Weninger, 1999).

Existen dos argumentos importantes para cuestionar la utilidad, en la planificación de explotaciones concretas, de los resultados que pudieran obtenerse de la investigación experimental sobre funciones de densidad de rendimientos de cultivos, incluso sin especificar su relación con las variables determinantes de las mismas.

El primer argumento tiene que ver con las metodologías utilizadas para la obtención de la información.

Existen tres fuentes principales clásicas de datos que pueden ser utilizados para determinar las distribuciones de rendimientos (Pease, 1992):

- series provinciales;

- experimentos agronómicos;

- series históricas de la propia explotación.

Las series provinciales son un pobre reflejo de las distribuciones en cada explotación (y no digamos en cada parcela), debido a los problemas de agregación, algunos de los cuales, por ejemplo, se discuten en Eisgruber y Schuman (1963). Es importante señalar que la metodología seguida en la mayoría de los países, entre ellos España, de hecho equivale en gran medida a una estimación subjetiva.

Los experimentos agronómicos no se consideran, en general, representativos de las prácticas culturales realizadas en cada explotación. La propia artificialidad del diseño de los experimentos, las condiciones en que se realiza cada operación, etc., hacen que los resultados normalmente no reflejen las prácticas de los agricultores, ni las que sería más conveniente desarrollar en función de la estructura de las explotaciones particulares.

Las series históricas de rendimientos de las propias explotaciones generalmente son cortas, incompletas o dudosas.

El segundo argumento en nuestra opinión es más importante y tiene que ver con la propia naturaleza de la variable a medir (los rendimientos de cultivos herbáceos).

Existe una hipótesis implícita en la interpretación de los resultados experimentales como pertenecientes a una función de densidad probabilística, y es que las distribuciones de las variables que la determinan, y las formas funcionales a través de las que se combinan, permanecen prácticamente estables en el tiempo (la variable a lo largo de la cual se realizan los registros) o se mueven siguiendo sendas paralelas (como dando saltos a lo largo del tiempo entre niveles), o que la senda que recorren en el tiempo es predecible.

Pero los períodos de tiempo necesarios para la determinación de las funciones de densidad en muchos cultivos, especialmente en los que presentan una cosecha anual, puede abarcar más de veinte años, y si las condiciones tecnológicas o de mercado no son estables en ese tiempo, los resultados serán extraordinariamente inadecuados.

Es cierto que existe evidencia histórica de que han existido épocas de gran estabilidad en las condiciones que determinan las producciones agrarias, hasta el punto de permitir el desarrollo de sistemas legales codificados de reparto del producto en función de los insumos aportados por diferentes agentes a la producción, como en el caso de los contratos de aparcería. Así, por ejemplo, en ciertos tipos de contratos en los que participan el propietario de la tierra, el propietario del agua, el propietario de las plantaciones, el propietario de los insumos utilizados anualmente para la producción y el encargado de aportar la mano de obra y la gestión de la producción, existían reglas legales que atribuían diferentes porcentajes del producto a cada agente. Esto supone no solamente un conocimiento empírico muy avanzado de la función de producción del cultivo y de la variabilidad relativa de los rendimientos (función de densidad de probabilidad de los mismos), sino también una gran estabilidad de las distintas variables físicas y sociales en las que se desarrollaban los cultivos, hasta el punto de permitir inferir empíricamente los coeficientes de reparto y considerar esa estimación tan segura que se incorporara a los códigos legales.

Esa estabilidad, sin embargo, no es algo que caracterice las condiciones de producción actuales. La idea general es que si en una finca particular se inicia un programa de experimentación para utilizar los resultados obtenidos en la planificación de la misma, cuando se disponga de suficientes datos para realizar una inferencia, posiblemente las condiciones técnicas o económicas hayan cambiado de forma tan decisiva que esa información sea obsoleta. En efecto, el cambio de variedades, de las técnicas de cultivo, de las sustituciones entre insumos por efectos técnicos o económicos, etc. seguramente provocarán de forma continuada cambios en la función de densidad subyacente.

El anterior argumento no niega importancia a los resultados obtenidos experimentalmente, que también deberán tenerse en cuenta en la planificación. La tesis que se sostiene es que no son suficientes, y que no lo serán en el futuro, por lo que habrá que completar esa información. Es por ello que diversos autores han propuesto que en la planificación de las actividades agrarias en condiciones de riesgo sea el propio agricultor o un experto próximo el que estime las funciones de densidad de los rendimientos mediante probabilidades subjetivas. Anderson et al. (1977) llegan incluso a afirmar que esas probabilidades son la única medida válida del riesgo en los estudios positivos o empíricos.

Los trabajos existentes sobre la utilización de probabilidades subjetivas en la programación de la producción agraria no son especialmente numerosos, y en algunos casos han puesto de manifiesto dificultades serias para la aplicación de esas técnicas. Sin embargo, consideramos que la probabilidad subjetiva es una clara frontera en la Economía Agraria y que todo trabajo que se realice en la misma contribuirá al avance en el objetivo principal de esta Tesis.

La importancia para la investigación de los enfoques basados en probabilidades subjetivas ha sido recientemente expuesta por Huirne, Hardaker, Dijkhuizen (1997) quienes han señalado que, pese a existir algunos artículos de gran calidad sobre la incorporación de las probabilidades subjetivas en la literatura agraria, este aspecto de cuantificación del riesgo ha sido tratado superficialmente. Concluyen que es necesario, por ello, trabajar para mejorar la calidad de los juicios probabilísticos. Entre otras cuestiones plantean:

- si los agricultores, los agentes de extensión, los investigadores o los políticos podrían obtener un mayor provecho de aquellas situaciones en las que existe poca información;

- si esos agentes están en condiciones de evitar los sesgos asociados con las estimaciones de probabilidades subjetivas.

En este trabajo se insiste en la importancia de abordar esos problemas en la investigación futura y en la necesidad de desarrollar un ‘código de buenas prácticas’ para realizar juicios acertados sobre probabilidades en una amplia variedad de situaciones, especialmente cuando la información es escasa y sesgada, o cuando existen, en el tiempo, oportunidades de aprendizaje sobre la incertidumbre. Una reivindicación de la utilidad del trabajo con probabilidades subjetivas se puede leer también en Horst, Huirne, Dijkhuizen y Steenkamp (1997).

Anderson (1997), por su parte, señaló que “I shall, however, be extremely surprised if the methods that will be advocated venture far from the subjective view of probability as the only practical language of analysed uncertainty (Anderson, Dillon and Hardaker, 1977; Morgan and Henrion, 1990), and thus eschew the temptations of fuzzy (Zadeh, 1965; Kosko, 1994) and other approaches”. Según este autor, el principal logro en el campo de los conocimientos analíticos aplicado sobre el riesgo ha sido el modelo de Utilidad Esperada Subjetiva (SEU) (tal como creía Shoemaker, 1982), aunque su adopción entre los autores que estudian los problemas asociados al riesgo en la agricultura esté resultando algo lenta.

En el Capítulo 2 de la Tesis se ha procedido a revisar el marco conceptual derivado de las probabilidades subjetivas o enfoque personalista de la probabilidad, ya que, en la línea de la argumentación anterior, se considera que jugará un papel esencial en la estimación de las principales variables de los modelos de decisión. Se completa esa revisión con la relativa a las funciones de densidad de probabilidad de los rendimientos y sus estimaciones subjetivas.

LOS SUJETOS COMO ESTIMADORES DE VARIABLES ALEATORIAS

Si no es posible disponer de la información sobre la función de densidad probabilística objetiva de los rendimientos o de otras variables relevantes, será necesario apoyarse en estimaciones subjetivas, y en este caso el sujeto decisor se sitúa en el centro de la escena, no ya en la elección final de una solución eficiente, como se había considerado tradicionalmente en los modelos de decisión, según su función de utilidad, sino también como fuente de información. Pero si el decisor o, las personas que colaboren con él, van a constituir una fuente primaria, y esencial, de información para la programación que prepare la toma de decisiones, es fundamental analizar con detalle la problemática que se puede presentar asociada a esta circunstancia.

La Sicología experimental ha puesto de manifiesto que los sujetos, a la hora de establecer juicios, comenten errores o sesgos sistemáticos y permanentes, provocados por los mecanismos cognitivos humanos. Como caso particular, esos sesgos se manifiestan también en la asignación de probabilidades subjetivas.

Esta es una cuestión escasamente explorada en la literatura económica agraria, aunque existen algunas referencias, como Barry (1984), que sin embargo consideramos que no han puesto de manifiesto el alcance y la naturaleza del problema. Es por ello que se ha realizado una revisión de los principales resultados en el Capítulo 3, donde se siguen especialmente las contribuciones de D. Kahneman; P. Slovic; A. Tversky (1984) y Evans (1989).

La literatura sobre resultados experimentales en la asignación de probabilidades subjetivas ha puesto de manifiesto la existencia de sesgos debidos a los mecanismos cognitivos humanos, pero también ha aportado suficiente evidencia de que la forma de plantear las cuestiones, las preconcepciones, los marcos interpretativos, los fenómenos de fijeza mental, etc. tienen una gran influencia sobre los juicios que expresan los sujetos.

Este papel central del decisor es necesario que sea tenido en cuenta también en la interfaz de comunicación entre el programa de ayuda a la toma de decisiones y el sujeto. Al menos se deberán tener en cuenta problemas relacionados con:

- la presentación, comodidad, simplicidad, etc., o, en general, la ergonomía de los sistemas de interrogación;

- la eficacia de los métodos empleados para obtener la información subjetiva para construir las inferencias;

- los resultados de la Sicología cognitiva sobre sesgos en el razonamiento, y específicamente en la asignación de probabilidades, a fin de evitarlos o corregirlos en la medida de lo posible, y, eventualmente, tenerlos en cuenta en toda la programación lógica;

- los supuestos implícitos en las microculturas de los potenciales usuarios, de los desarrolladores, de los expertos informantes, etc.;

La segunda cuestión cobra una gran importancia en el contexto que discutimos. Previamente a la posible corrección de los sesgos implicados por los mecanismos cognitivos, es preciso evitar los inducidos por los supuestos implícitos en la comunicación hombre – máquina y entre técnicos de diferentes formaciones y culturas. Estas cuestiones, muy específicas de cada ámbito particular de problemas, no han sido tratadas de forma sistemática en la literatura especializada (aunque se encuentren referencias en la experimentación didáctica), posiblemente por las circunstancias en que normalmente se aborda la resolución de problemas, mediante lenguajes relativamente precisos y muy formalizados, en los que han sido (de) formados los usuarios.

En el amplio campo delimitado por ese problema es el que esta Tesis ha planteado su trabajo experimental.

TRABAJO EXPERIMENTAL

El problema planteado en la Tesis es la comprobación de algunas metodologías, muy básicas, que podían utilizarse en la estimación de las funciones de densidad de probabilidad de los rendimientos de cultivos. Se planteó la contrastación de tres métodos:

(1) Aproximación a la f.d.p. mediante la estimación de tres valores (rendimientos mínimos, más frecuentes y máximos);

(2) Obtención de la f.d.p. mediante la definición de los rendimientos en una escala ordinal, expresada en lenguaje natural, y asignación de frecuencias a los distintos intervalos así definidos;

(3) Traducción directa de las escalas ordinales a numéricas, y obtención de la f.d.p. definida en (2) en una escala numérica;

Como es conocido, esos métodos constituyen una primera aproximación al problema de la obtención de las funciones de densidad de probabilidad, pero el objetivo no era la determinación de las funciones, sino la evaluación de las dificultades, sobre todo ligado a problemas previos lingüísticos y de estimación. Por ello los métodos se han aplicado a diferentes poblaciones y en diferentes momentos. La idea que subyace en nuestro trabajo es que hay que hacer una exploración sistemática, desde las técnicas más sencillas a las más sofisticadas, en las que se vayan discutiendo los problemas que aparezcan de determinación, y que no dependan solamente de las técnicas empleadas, sino de las cuestiones relacionadas con la calidad del conocimiento específico disponible, la forma de preguntar para obtener esa información y las interpretaciones que introduzcan los diferentes sujetos y contextos.

El objetivo final sería obtener un protocolo de interrogación que permitiera realizar las estimaciones de las f.d.p. de una forma relativamente sencilla, mecánica y eficaz. Esta Tesis Doctoral es un paso preliminar en esa dirección.

CAPÍTULO 1

PROGRAMACIÓN EN CONDICIONES DE RIESGO EN AGRICULTURA: DE LA PROGRAMACIÓN ESTOCÁSTICA A LOS MODELOS DE RIESGO

#1.1.- PROGRAMACIÓN ESTOCÁSTICA

Un modelo de programación se puede formular de forma genérica como:

$Max\ f_i(X)$

s. a.

$G_j(X) \leq B$

$X \geq 0$

Esa forma general se puede particularizar de distintas maneras, siendo la más simple el programa lineal (con un solo objetivo):

$Max\ CX$

s.a.

$AX \leq B$

$X \geq 0$

Donde los valores $c_j \in C$, $a_{ij} \in A$ y $b_i \in B$ pueden interpretarse de diferentes formas, pero en el contexto de la programación de la producción, x_i es el valor que toma la actividad *i*, c_j es el logro alcanzado por cada unidad de actividad *j*, a_{ij} es el coeficiente técnico que indica el consumo del recurso *i* por la actividad *j*, y b_i es la cantidad disponible del recurso *i*.

Obviamente, cualquiera de los coeficientes de *C*, *B* o de *A* puede ser una variable aleatoria, y el modelo sería estocástico.

Inicialmente, la modelización de la programación estocástica consideró aleatorios sólo los coeficientes de la única función objetivo. Se trata de modelos que abordan el riesgo considerando un único criterio optimizador (minimizar alguna medida adecuada de la dispersión de los resultados, como la Varianza, o maximizar alguna medida del logro del objetivo, como por ejemplo la Esperanza Matemática). El problema planteado consistía en elegir un vector *X* antes de conocer los valores que tomarán los coeficientes aleatorios c_j.

Este problema fue abordado de diferentes formas:

#1.1.1.- MODELO E: CRITERIO DEL VALOR ESPERADO

Siendo $\overline{c}_j$ la esperanza matemática de c_j, que se supone conocida, y admitiendo la hipótesis de que la distribución de los elementos c_j es independiente de las variables a optimizar, la decisión X vendrá dada por la solución del problema:

$$Maximizar \quad Z = \sum_{j=1}^{n} \overline{c}_j x_j$$

s. a:* $$\sum_{j=1}^{n} a_{ij} x_j \le b_i \quad i = 1,2,...,m$$

$$x_j \ge 0 \quad j = 1,2,..,n$$

De ahora en adelante, se sustituirá el conjunto de restricciones indicado (*) por la notación $X \in F_d$.

#1.1.2.- MODELO V: CRITERIO DE VARIANZA MÍNIMA

$$Minimizar \quad V$$

s. a: $$\sum_{j=1}^{n} \overline{c}_j x_j \ge Z_0$$

$$X \in F_d$$

siendo $Z_0=C_0X_0$.

#1.1.3.-. MODELO P

Introducido por *Charnes y Cooper* (1963), consiste en maximizar la probabilidad de que la función objetivo sea menor que un valor constante k prefijado anteriormente. En este caso los valores c_j son costes (variables aleatorias de función de probabilidad conocida). *Bereanu* (1964) llamó a este modelo de mínimo riesgo a nivel k.

El problema original:

$$Minimizar \quad \sum_{j=1}^{n} c_j x_j$$

s. a: $X \in F_d$

se convierte en:

$$Maximizar \quad P\left\{ \sum_{j=1}^{n} c_j x_j \le k \right\}$$

s. a: $X \in F_d$

#1.1.4.- MODELO DE KATAOKA

Puede considerarse una variante del modelo anterior en la cual se fija la probabilidad α y se desea minimizar el valor *k*. Este modelo ha sido resuelto para el caso en que las variables aleatorias c_j estén normalmente distribuidas, siendo su distribución conjunta la normal multivariante $N(\{\bar{c}_j\},V)$

Puede escribirse: $\alpha = P\left\{\sum c_j x_j \le k\right\} = P\left\{\frac{\sum c_j x_j - M}{s} \le \frac{k-M}{s}\right\}$

donde $\frac{\sum c_j x_j - M}{s}$ es una variable aleatoria normalmente distribuida, de media 0 y varianza 1. Por tanto, fijado α se puede determinar el valor de la abscisa correspondiente en las tablas de la distribución normal. Si la abscisa se representa como $\lambda\alpha$, se tiene:

$$\lambda\alpha = \frac{k-M}{s};\ k = \lambda\alpha s + M$$

y por tanto:

$$Minimizar\ k = \lambda\alpha\left(X'VX\right)^{1/2} + \sum_{j=1}^{n} \bar{c}_j x_j$$

s. a: $X \in F_d$

#1.2. MODELOS DE RIESGO

Los modelos de riesgo consideran básicamente la dualidad renta-riesgo en el planteamiento de los problemas de programación, ya sea de forma explícita a través de la programación multicriterio o bien a través de otros modelos, con la finalidad de calcular el conjunto de soluciones eficientes.

En primer lugar, los modelos E-V y MOTAD plantean la minimización de una medida de dispersión (Varianza y MAD respectivamente) de la variable aleatoria renta con una restricción paramétrica para la esperanza matemática de la misma, llegando en el primer caso a una programación cuadrática y en el segundo a una programación lineal. Una variante del modelo MOTAD que introduce la posibilidad de considerar el 'downside risk' la constituye el modelo Target-MOTAD, que mide las desviaciones respecto a un nivel dado de ingresos mediante una programación también lineal. El modelo Media – DAP, complementario del modelo Target-MOTAD, pretende encontrar el conjunto eficiente de mayor amplitud posible, dado un nivel de ingresos fijado en el valor maximin de los mismos, además de calcular la probabilidad de fracaso. La programación compromiso acota el conjunto eficiente según las distancias de los diferentes puntos al punto ideal.

Se revisan también los modelos de eficiencia estocástica, que abordan la selección de carteras utilizando las funciones de distribución completas de las variables aleatorias, y otros modelos como el del índice simple o la simulación.

#1.2.1.- EL MODELO CUADRÁTICO (E-V)

Harry M. Markowitz (1952), en su trabajo sobre la selección de carteras de valores, desarrolló un modelo que permite abordar el riesgo mediante el siguiente programa:

$$min\, V = \sum_j \sum_k X_j X_k \sigma_{jk}$$

sujeto a:

$$\sum_j \overline{c}_j X_j = \lambda$$

$$X \in F_d$$

El sumatorio $\sum_j \overline{c}_j X_j$ es la esperanza del margen bruto total esperado, *E*, que se iguala a un parámetro λ. Dando diferentes valores a λ, se obtiene la frontera eficiente, o puntos donde la varianza es mínima para cada esperanza matemática.

Como medida de riesgo, *Markowitz* utilizó la varianza de la renta, *V*. El agricultor manifestaría sus preferencias entre los diversos planes de producción alternativos en función de la renta esperada de cada plan productivo, *E*, y de la varianza asociada, *V*, mediante una función de utilidad *U* = *E* - *V*.

La elección de la solución óptima, de entre las distintas soluciones eficientes, se realiza mediante la introducción de la función de utilidad del agricultor (que se supone cuadrática o exponencial). Dicha función toma la forma común *U* = *E* - *kV*, siendo *k* el coeficiente de aversión al riesgo.

Entre las ventajas de este modelo, se encuentran:

1) Puede utilizarse para describir la frontera eficiente E-V (*Scott y Baker*, 1972).

2) Incorpora las varianzas y covarianzas de los ingresos de las posibles combinaciones empresariales (*Scott y Baker*, 1972; *Romero y Rehman*, 1989).

3) Si la distribución de los ingresos es simétrica y el inversor es adverso al riesgo, el modelo E-V es razonable (Schrage, 1997).

Aunque se trata de un modelo ampliamente utilizado y comentado en la literatura, tiene algunos inconvenientes (*Scott y Baker*, 1972; *Porter*, 1973; *Barry y Robison*, 1975; *Romero*, 1976; *Alonso*, 1977; *Anderson et al.*, 1977; *Fishburn*, 1977; *Levy y Markowitz*, 1979; *Paris*, 1979; *Robison y Brake*, 1979; *Dios et al.*, 1980; *Sumpsi*, 1980; *Holthausen*, 1981; *Musser y Stamoulis*, 1981; *Tauer*, 1983; *El-Nazer*, 1984; *Selley*, 1984; *Berbel*, 1987; *Hassan y Hallam*, 1990; *Hanson y Ladd*, 1991, entre otros):

1) Supone que las variables aleatorias están distribuidas normalmente.

2) Es un modelo cuadrático, y necesita programación no lineal para su resolución.

3) La programación cuadrática supone que las curvas de indiferencia son convexas (aversión al riesgo). Una función de utilidad cuadrática se caracteriza por una aversión al riesgo creciente, y por un valor máximo de ingresos más allá del cual la utilidad marginal se vuelve negativa.

4) En la especificación de la función de utilidad (cuadrática para los ingresos: $U = E - kV$), k es un coeficiente de naturaleza sicológica, y desconocido a priori.

5) Independientemente del valor de k, afectar de un signo negativo al término que comprende la varianza prejuzga el comportamiento del empresario.

6) La actitud del empresario ante el riesgo depende también de factores económicos, en especial de las disponibilidades financieras, no siendo constante para un mismo individuo.

7) Utilizar la varianza en la función de utilidad supone una simetría de comportamiento del agricultor, rechazando tanto un resultado excepcionalmente bueno como uno excepcionalmente malo, lo cual es cuestionable.

8) En el modelo sólo se tiene en cuenta la variabilidad en la función objetivo, no considerándose la variación de las disponibilidades de recursos B ni de los coeficientes a_{ij} de la matriz A de las restricciones.

9) El indicador de utilidad $U = E - kV$ no obedece al concepto de orden parcial, por lo que no puede servir más que como una aproximación.

10) En bastantes casos se produce un incremento de la varianza de la muestra al aumentar el tamaño de la misma (muchas series económicas y estadísticas de precios están construidas a partir de muestras de varianza infinita). Si esto se produce, la desviación estándard de una muestra, aún siendo necesariamente finita, no es demasiado fiable y es una medida pobre del riesgo.

11) En general, se prefiere no utilizar series históricas de datos para estimar los valores de las distribuciones de E y V del modelo. La utilización de datos históricos es arriesgada, pues se debería tener en cuenta la información específica y los cambios ocurridos. Una forma de abordar este problema (*Baumol*, 1963) sería no utilizar datos anuales, como hace *Markowitz*, sino otros obtenidos en períodos de tiempo inferiores. Los modelos más simples de predicción, que utilizan la media ponderada de los ingresos netos para los últimos años, como los de *Mapp et al.* (1979) y *Adams et al.*(1980), permiten predecir mejor los ingresos esperados de los siguientes períodos que si se utilizara la media de los ingresos de, por ejemplo, los últimos veinte años.

12) También puede suceder que los datos de que disponga el agricultor sean demasiado antiguos o inexistentes.

13) El decisor debe elegir un valor concreto de k. (Sea como sea la distribución, la desigualdad de Tchebyshev permite interpretar cualquier valor de k en términos de que la probabilidad de las observaciones estén comprendidas en el intervalo $E \pm k\sigma$).

14) Algunos autores como *Porter* (1974), *Berbel* (1987) o Schrage (1997) creen que el criterio media-varianza supone una violación del comportamiento racional, que lleva a decisiones no justificadas empíricamente, puesto que lo que el decisor desea realmente es medir el riesgo de ruina que corre al realizar su actividad empresarial.

15) El modelo E-V, o el modelo E-σ, penaliza las grandes desviaciones por debajo de la media (incluso las positivas), a diferencia de los modelos de seguridad primero.

16) Para dos variables aleatorias X e Y tales que $E(X) = E(Y)$, la variable aleatoria X es más arriesgada que la variable aleatoria Y si y sólo si la utilidad esperada de X es menor que la utilidad esperada de Y (*Antle*, 1983, considerando la definición de riesgo de *Rothschild y Stiglitz*, 1970). Según esta

definición, la varianza no es generalmente una medida válida del riesgo, dado que la utilidad esperada de *X* puede ser mayor que la utilidad esperada de *Y* incluso cuando la varianza de *X* sea mayor que la varianza de *Y*. *Antle*, en consecuencia, indica que sería apropiado ordenar variables aleatorias en términos de media y varianza sólo cuando la utilidad esperada del beneficio dependa sólo de la media y la varianza del mismo.

Freund (1956) aplicó por primera vez el modelo de *Markowitz* de selección de carteras eficientes a la planificación de actividades agrarias bajo riesgo. Propuso elegir la distribución de probabilidad que ofrezca la máxima (esperanza matemática) utilidad esperada. Para ello, utilizó una función de utilidad exponencial con unos ingresos que se distribuyen normalmente. Dicha función toma la forma:

$E(U[M]) = a - be^{-kM}$, siendo *b* y λ mayores de cero.

Si los ingresos *M* están normalmente distribuidos, se verifica que:

$$E(U[M]) = a - be^{-kE(M)+(k^2/2)V(M)}$$

De esta forma, según *Freund*, maximizar esa función equivale a maximizar $E(U[M]) = E(M) - (k/2)V(M)$, y el modelo se plantearía como:

Programa sin riesgo:	*Programa con riesgo:*
$max \sum_j \bar{c}_j X_j$ *s. a:* $X \in F_d$	$max \sum_j \bar{c}_j X_j - \frac{k}{2} \sum_j \sum_k X_j X_k \sigma_{jk}$ *s. a:* $X \in F_d$

siendo $\sum_j \bar{c}_j X_j$ los ingresos totales esperados.

Este modelo presenta las ventajas y los inconvenientes comentados en el modelo general de *Markowitz*.

#1.2.2.- EL MODELO CUADRÁTICO EN LA LITERATURA ECONÓMICA AGRARIA

El modelo cuadrático se ha utilizado ampliamente en la literatura, pudiéndose citar los siguientes trabajos:

Frankfurter et al. (1971), *Dickenson* (1974), *Klien y Bawa* (1976) y *Jobson y Korkie* (1980) examinaron la imprecisión de las soluciones eficientes E-V ocasionada por el error de muestreo de los estimadores *E* y *V*.

Scott y Baker (1972) combinaron la programación cuadrática y la teoría de *Baumol* (1963).

Romero (1976) aplicó el modelo de Markowitz a la selección de planes óptimos de variedades de manzanos en la provincia de Lleida.

Alonso (1977) estudió el modelo de Markowitz como modelo de riesgo y lo comparó con modelos de incertidumbre de la Teoría de Juegos.

Fishburn (1977), *Holthausen* (1981) y *Tauer* (1983) examinaron las limitaciones del modelo E-V.

Paris (1979) utilizó el modelo E-V teniendo en cuenta que los precios y la disponibilidad de recursos son aleatorios. Utilizó la función de utilidad de *Freund* (1956) con modelos primal y dual, e introdujo las covarianzas entre los precios y la disponibilidad de recursos en otro supuesto.

Dios, Cañas y Rodríguez (1980) aplicaron a los seguros de cosechas los conceptos de planes eficientes, utilizando programación cuadrática.

Sumpsi (1980) realizó una revisión de los métodos de programación cuadrática bajo riesgo.

La programación cuadrática con preferencias de riesgo parametrizadas fue utilizada por *Musser y Stamoulis* (1981) para evaluar programas agrícolas (Agricultural Commodity Programs), en un estudio empírico de incentivos para una empresa agrícola representativa en el centro-sur de Georgia para el año 1977.

Lazarus y Swanson (1983) utilizaron el modelo de *Freund* (1956) que maximiza la utilidad esperada.

El-Nazer (1984) aproximó la función de utilidad del empresario mediante los dos primeros momentos (media y varianza) de la distribución de los márgenes brutos para todas las actividades. Esto es así dado que cualquier momento superior al segundo no será significativo según el teorema central del límite (*Anderson et al.*, 1977).

Juárez (1985) propuso una metodología de búsqueda del plan óptimo, como segunda etapa en los modelos E-V o MOTAD.

Según *Romero y Rehman* (1989) entre otros autores, una de las ventajas de los modelos E-V y E-σ es que tienen en cuenta las relaciones de covarianza entre los márgenes brutos de las actividades.

Hassan y Hallam (1990) aplicaron el modelo E-V incorporando tecnología aleatoria y midiendo los efectos del riesgo sobre los factores de producción en explotaciones de regadío en Sudán.

Hanson y Ladd (1991) estudiaron la aplicación del modelo media-varianza en el caso de que la distribución de probabilidad de los ingresos esté truncada.

Alaejos y Cañas (1992), determinaron planes de cultivo eficientes, en el sentido de Markowitz, en la provincia de Córdoba.

Arias (1994) comparó los modelos E-V y MOTAD, en una aplicación a la comarca de la Campiña (Guadalajara), obteniendo resultados muy semejantes entre ambos métodos.

Berg (1997) utilizó el modelo del valor esperado-varianza para investigar la influencia de la reforma de la PAC sobre el riesgo.

Chaherli (1997) realizó una selección de cultivos bajo riesgo mediante el modelo E-V, teniendo en cuenta la elección bajo distintos programas de política agraria.

Pennings y Meulenberg (1997) utilizaron el modelo valor esperado-varianza en un mercado de futuros de leche.

Rister et al. (1997) utilizaron el modelo E-V en el estudio de estrategias de marketing.

Robison y Hanson (1997) discutieron las ventajas e inconvenientes de los modelos de utilidad esperada y de esperanza-varianza.

Stevens (1998) caracterizó la inversa de la matriz de covarianzas del análisis de la cartera de valores.

#1.2.3.- ANÁLISIS MEDIA - SEMIVARIANZA

El *análisis media-semivarianza* es una regla de decisión que utiliza la zona inferior de la distribución. La semivarianza del beneficio referida a un objetivo *d*, SV_d, es el valor esperado de las desviaciones al cuadrado del beneficio por debajo de *d*, $E(M-d \mid M<d)^2$. Es decir:

$$SV_d(M) = \sum_i min(M_i - d, 0)^2 P(M_i)$$

El análisis media-semivarianza puede ser calculado por referencia a la media o a otro punto.

Según *Selley* (1984), cuando la distribución de beneficios es asimétrica, es preferible este análisis al de media-varianza.

Como crítica al modelo, hay que señalar que el análisis media-semivarianza referido a la media no es consistente con la maximización de la utilidad esperada, igual que ocurre con el análisis media-desviación absoluta referido a la media, mientras que la media-semivarianza referida a un punto concreto puede serlo. Además, según *Hogan y Warren* (1972), para distribuciones simétricas el modelo media-semivarianza requiere información de λ (grado de aversión al riesgo bajo utilidad exponencial) y *d* para representar la aversión al riesgo, mientras que el modelo media-varianza tiene la ventaja de requerir sólo un parámetro para representar el nivel de aversión al riesgo.

#1.2.4.- APROXIMACIONES LINEALES AL MODELO CUADRÁTICO: MODELO MOTAD

Dentro de las aproximaciones lineales al modelo cuadrático de Markowitz, el modelo MOTAD, desarrollado por *Hazell* (1971), propone el uso de estimadores de varianza basados en la Desviación Absoluta Media de la muestra (MAD).

El estimador MAD es:

$$\tilde{V} = F\left\{(1/T)\sum_t \left|\sum_j c_{jt} X_j - \sum_j \bar{c}_j X_j\right|\right\}^2 = F\left\{(1/T)\sum_t \left|Y_t - \bar{Y}\right|\right\}^2$$

donde la expresión entre llaves es el MAD muestral, y F es una constante que relaciona el MAD muestral con la varianza poblacional. En concreto, $F = T\pi / 2(T-1)$, donde π es la constante matemática.

El atractivo del estimador MAD es que si sustituimos su expresión por la de $min\, V = \sum_j \sum_k X_j X_k \sigma_{jk}$ en el modelo de programación cuadrática, entonces se obtiene un modelo de programación lineal. Si la desviación del beneficio de la explotación respecto a su media para el año t se representa por Z_t^+ si es positiva, y por Z_t^- si es negativa, entonces:

$$Z_t^+ - Z_t^- = \sum_j c_{jt} X_j - \sum_j \bar{c}_j X_j, \quad \forall\, t$$

Hay que tener en cuenta que Z_t^+ y Z_t^- son ambas no negativas en esta formulación, y por tanto miden la dimensión absoluta de la desviación del beneficio respecto a su media. Además, sólo una de ellas puede ser superior a cero cada año, puesto que la desviación no puede ser positiva y negativa a la vez. Por lo tanto, $\sum_t (Z_t^+ + Z_t^-)$ mide la suma de los valores absolutos de las desviaciones del beneficio, y el estimador MAD de la varianza se puede expresar como:

$$\tilde{V} = F\left\{(1/T)\sum_t (Z_t^+ + Z_t^-)\right\}^2$$

Dado que F / T^2 es una constante en cada problema concreto, se puede dividir $\tilde{V}$ entre F / T^2 y obtener:

$$W = (T^2 / F)\tilde{V} = \left\{\sum_t (Z_t^+ + Z_t^-)\right\}^2$$

También se puede extraer la raíz cuadrada de W, con lo cual se formula el siguiente programa lineal alternativo al cuadrático:

$$min W^{1/2} = \sum_t (Z_t^+ + Z_t^-)$$

sujeto a:

$$\sum_j (c_{jt} - \bar{c}_j) X_j - Z_t^+ + Z_t^- = 0, \quad \forall\, i$$

$$\sum_j \bar{c}_j X_j = \lambda$$

$$X \in F_d$$

$$Z_t^+, Z_t^- \geq 0, \quad \forall\, j,\, t$$

Dado que la función objetivo del modelo es la Minimización de las Desviaciones Absolutas Totales, Hazell lo llamó modelo MOTAD. Este modelo se puede resolver mediante programación lineal paramétrica para obtener el conjunto de planes eficientes E - V.

Se puede obtener una versión más compacta del modelo MOTAD, dado que la suma de las desviaciones negativas del beneficio por debajo de la media, $\sum_t Z_t^-$, debe ser siempre igual a la suma de las desviaciones positivas por encima de la media, $\sum_t Z_t^+$.

Por otro lado, es suficiente minimizar cualquiera de esas dos sumas y multiplicar el resultado por 2 para obtener $W^{½}$. Puede comprobarse este enunciado en el siguiente modelo MOTAD (donde se ha seleccionado la minimización de la suma de las desviaciones negativas):

$$min \quad 0.5\, W^{1/2} = \sum_t Z_t^-$$

sujeto a:

$$\sum_t (c_{jt} - \overline{c}_j) X_j + Z_t^- \geq 0, \quad \forall\ t$$

$$\sum_j \overline{c}_j X_j = \lambda$$

$$X \in F_d$$

$$Z_t^+, Z_t^- \geq 0, \quad \forall\ j,\ t$$

Entre las ventajas del modelo MOTAD se encuentran:

1) Se trata de un modelo lineal, en el que se obtienen resultados similares a los del programa cuadrático del que se deriva (*Fisher* 1920, *Hazell* 1971).

2) Cuando los beneficios se distribuyen normalmente, MAD es un estimador muestral tan bueno como la varianza muestral, especialmente con tamaños muestrales reducidos (*Thomson y Hazell*, 1972).

3) Dado que el decisor está más interesado en la falta de logro que en el exceso, en términos de ingreso esperado, MOTAD presenta ventajas en las distribuciones de ingresos asimétricas (*Kennedy y Francisco*, 1974).

4) Para distribuciones simétricas, el análisis media-desviación absoluta esperada ofrece en teoría el mismo ránking ordinal de alternativas que el análisis media-varianza (*Johnson y Boehlje,* 1981).

5) El análisis media-desviación absoluta no penaliza las grandes desviaciones, como hace el análisis media-varianza (*Barry,* 1984).

No obstante sus ventajas, se han señalado algunos inconvenientes y críticas:

1) El estimador MAD de la muestra es un estimador de la varianza de la población menos eficiente que la varianza muestral (*Fisher* 1920, *Hazell* 1971) en el caso de que MAD se use como estimador y no como fin en sí mismo.

2) Cuando las distribuciones de los beneficios son asimétricas, MAD a veces no es muy fiable (*Thomson y Hazell*, 1972).

3) El análisis media-desviación absoluta no es consistente con la maximización de la utilidad esperada (*Barry*, 1984), ya que genera funciones de utilidad no derivables, entre otras razones.

4) Con los avances en técnicas de optimización no lineal, aproximaciones como MOTAD no parecen necesarias para problemas con menos de 1500 variables (*McCarl,* 1986).

5) Igual que en el análisis E-V, MOTAD supone que el decisor es adverso al riesgo (*Mjelde et al.,* 1990).

#1.2.5.- MOTAD - RINOCO

Como una ampliación al modelo MOTAD, *Wicks y Guise* (1978) desarrollaron el modelo MOTAD con RINOCO (Risky Input-Output Coefficients), añadiendo aleatoriedad a los coeficientes a_{ij} de las restricciones:

$$maxE = \sum_j \bar{c}_j X_j$$

sujeto a:

$$\hat{\sigma}_i = (2F^{1/2} / T)\sum_t Z_{it}^+, \quad \forall \ i$$

$$\sum_j (a_{ijt} - \bar{a}_{ij}) X_j - Z_{it}^+ \leq 0, \quad \forall \ t$$

$$\sum_j \bar{a}_{ij} X_j + K_\alpha \hat{\sigma}_i \leq b_i, \quad \forall \ i$$

$$X_j, Z_{it}^+, \hat{\sigma}_i \geq 0, \quad \forall \ i, \ j, t$$

donde $\bar{a}_{ij}$ denota el valor medio de la muestra para el coeficiente a_{ij} y F es la constante de Fisher, $\pi T/2(T\text{-}1)$. Las variables Z_{it}^+ miden la cantidad en que las necesidades del plan de cultivos para el recurso *i* en el año *t* exceden a las necesidades de la media para aquel recurso y σ_i es el estimador MAD de la desviación estándar de las necesidades de recursos de la planificación.

En este caso el MAD no constituye una aproximación a la programación cuadrática tan buena como lo es la forma paramétrica del modelo MOTAD.

#1.2.6.- EL MODELO MOTAD EN LA LITERATURA ECONÓMICA AGRARIA

Brink y McCarl (1978) utilizaron la formulación de Hazell incorporando a la función objetivo la suma de las desviaciones totales negativas anuales, multiplicadas por un coeficiente de aversión al riesgo y por una constante $K = \frac{2}{s}\sqrt{\frac{s\Pi}{2(s-1)}}$ (para transformar el término en un estimador de las desviaciones estándard; más información en *Simmons y Pomareda*, 1975), y restadas de los márgenes brutos. Todo ello sujeto a las restricciones del modelo MOTAD inicial. Parametrizando el coeficiente de aversión al riesgo, obtuvieron el conjunto eficiente de planes, concluyendo que la aversión al riesgo no era, en general, un factor importante en la elección de los cultivos a implantar en el grupo estudiado por ellos.

Hanf y Mueller (1979) utilizaron tres modelos derivados de MOTAD (MOTAD inicial o de *Hazell*, 1971; MOTAD con coste de infactibilidad, debido a *Schiefer*, 1978; y MOTAD con RINOCO, debido a *Wicks y Guise* , 1978) para estudiar tres explotaciones representativas de Alemania.

Brandao et al. (1984) utilizaron como medida del riesgo el modelo de *Hazell* (1971) con las modificaciones posteriores de *Brink y McCarl* (1978) en su planificación de cultivos en Brasil.

El-Nazer y McCarl (1986) utilizaron una variación del modelo MOTAD de Hazell, minimizando la desviación negativa de *Brink y McCarl* (1978).

McCarl y Önal (1989) compararon la programación no lineal con aproximaciones lineales como MOTAD o la programación separable, concluyendo que la solución no lineal es más eficiente que las aproximaciones lineales cuando la no linealidad está en la función objetivo. En cambio, cuando ésta se encuentra en las restricciones, es mejor realizar la aproximación a modelos lineales.

Arias (1994) comparó los modelos E-V y MOTAD, en una aplicación a la comarca de la Campiña (Guadalajara), obteniendo resultados muy semejantes, señalando las ventajas del segundo modelo.

Gómez-Limón y Berbel (1995) utilizaron el modelo MOTAD como medida de riesgo en un programa multicriterio aplicado al regadío cordobés.

Carvalho et al. (1997) utilizaron un modelo matemático basado en programación estocástica discreta asociada al modelo MOTAD, para incluir los efectos de la variabilidad en los rendimientos sobre la producción animal, considerando 16 estados de la naturaleza.

Oglethorpe (1997), a partir de una encuesta realizada a agricultores del norte de Inglaterra, obtuvo información sobre las estrategias de manejo que luego utiliza para estimar las fronteras ingreso-varianza generadas con el modelo MOTAD.

#1.2.7. ANÁLISIS MEDIA-DESVIACIÓN ESTÁNDARD

El criterio media-desviación estándard (E-σ) está muy relacionado con el modelo E-V. Dado que σ es la raíz cuadrada de *V*, el conjunto de planes eficientes E-σ es idéntico al del conjunto E-V.

Como criterio de decisión, el modelo E-σ está sujeto a las mismas críticas que la aproximación E-V.

Tsiang (1972) argumentó que el criterio E-σ es una buena aproximación a los mejores criterios de decisión si el riesgo asumido es pequeño en relación a la economía del agricultor. Pero mientras que esa condición puede ser conseguida fácilmente en numerosas explotaciones comerciales, probablemente no se cumpla para explotaciones pequeñas, de subsistencia, en países en desarrollo.

Baumol (1963) propuso un criterio de eficiencia dentro del análisis media - desviación estándard como alternativa a la selección de la cartera de Markowitz. El modelo se basa en el criterio E-L, siendo $L=E-k\sigma$ el límite de confianza inferior para los ingresos del inversor, de forma que cuanto mayor sea *k*, más se aproximarán los conjuntos eficientes E-L y E-σ y más conservador será el inversor. Por otro lado, ninguna cartera puede ser E-L eficiente sin ser E-σ eficiente, pues el primero es un subconjunto del segundo.

Un inconveniente del análisis E-L es que impone al usuario del análisis un juicio explícito que toma una forma a la cual está poco habituado.

Según *Hazell y Norton* (1986), existen dos problemas principales con esta aproximación para estimar *k*. El primero es que existe una gran probabilidad de que *k* esté sesgado, debido a los defectos del modelo y a los errores de los datos; y el segundo que si los agricultores tienen acceso a instituciones que compartan el riesgo, tales como seguros de cultivos o mercados de futuros, entonces sus decisiones en la planificación de la explotación no reflejará sus preferencias reales frente al riesgo, con lo que se subestimará probablemente el valor de *k*.

Scott y Baker (1972) integraron el criterio de Baumol y la programación cuadrática para calcular las alternativas significativas, como ya se ha comentado en la revisión de la programación cuadrática.

#1.3. SEGURIDAD PRIMERO

Los modelos de seguridad primero (*safety-first*) proporcionan al decisor un criterio que tiene en cuenta la probabilidad de quedarse por debajo de un nivel mínimo de ingresos. Según *Berbel* (1988), todos los modelos de seguridad primero requieren de estimadores de probabilidad para el extremo inferior de la distribución correspondiente. Es decir, se desea determinar, en general, la probabilidad de que los ingresos no alcancen un nivel de “desastre” o de “ruina” predeterminado, que puede ser estimado utilizando la desigualdad de Tchebyshev:

$$Pr[/x - E(x)/ > k\,s] \leq 1 / k^2$$

donde $E(x)$ es el valor esperado de los ingresos, *s* es la desviación estándard, y *k* es un factor que determina el límite de probabilidad.

#1.3.1.- MODELO DE ROY

Roy (1952) no rechazó la varianza como indicador de la dispersión, pero avanzó la idea de que el objetivo más importante de un empresario prudente es minimizar la probabilidad de ruina antes que maximizar sus ganancias. Por ello, su modelo propone minimizar la probabilidad de que los ingresos *z* se encuentren por debajo de un nivel dado $\bar{z}$, y utiliza un límite superior basado en la desigualdad de Tchebyshev, en lugar de un límite de probabilidad preciso, cuando la distribución de probabilidad de los beneficios es desconocida:

$$min\ Pr(z \leq \bar{z})$$

Este modelo comporta las dificultades de estimar $\bar{z}$ y la ley de probabilidad de z (*Boussard*, 1969).

#1.3.2.- MODELO DE TELSER

Telser (1955) propuso el siguiente modelo, que utiliza la desigualdad de Tchebyshev:

$Max\ \mu$

sujeto a:

$Pr(z \leq \bar{z}) \leq \alpha$

donde μ es el valor esperado de los ingresos z y α la probabilidad de desastre.

#1.3.3.- MODELO DE KATAOKA

Kataoka (1963) propuso el siguiente modelo:

$Max\ \bar{z}$

sujeto a:

$Pr(z \leq \bar{z}) \leq \alpha$

planteando encontrar el nivel máximo de ingresos $\bar{z}$ para el cual la probabilidad de que éstos estén por debajo de $\bar{z}$ sea inferior a un nivel preespecificado α.

Los Criterios de Telser y Kataoka implican reforzar las restricciones probabilísticas $Pr(Z < g) \leq 1/L^*$, pero esto puede ser problemático en un modelo de optimización si (a) la distribución de los ingresos no es normal o (b) los valores obtenidos de la distribución de los ingresos se ven afectados por decisiones endógenas al modelo. Estas dificultades son discutidas por *Sengupta* (1969).

#1.3.4.-MODELO DE BOUSSARD Y PETIT

Boussard y Petit (1967) se basaron en el concepto de *"focus loss"* desarrollado por *Shackle* (1949, 1961), para adaptar las exigencias de los modelos de seguridad primero a un modelo lineal.

El *focus loss* de una actividad arriesgada se define como el nivel de pérdida que un decisor estaría "muy sorprendido" de asumir. En la práctica, *Boussard y Petit* aproximaronn los valores de "focus loss" para diferentes actividades agrarias utilizando "catrástofes decenales", es decir, el peor margen bruto que se podría producir en una década. Dado este peor margen bruto (llamado c_j^*) para cada actividad, y el margen bruto esperado, $\bar{c}_j$, la pérdida focal se define como $f_j = \bar{c}_j - c_j^*$, $\forall j$.

Para cualquier plan de cultivos, la máxima pérdida posible (llamada LOSS) se define como la diferencia entre el margen bruto total esperado, $\sum_j \bar{c}_j X_j$ y el mínimo ingreso (MINI) requerido para cubrir los costes fijos de la explotación, los costes esenciales para mantener a la familia y el pago de las deudas. Es decir,

$$LOSS = \sum_j \bar{c}_j X_j - MINI$$

Boussard y Petit impusieron la condición de que ninguna actividad simple debe tener una pérdida focal $f_j X_j$ superior a *1/k* de la máxima pérdida permitida en el plan de cultivos. Estas restricciones se expresan como $f_j X_j \leq 1 / k\ (LOSS),\ \forall\ j$.

Como ventajas del modelo se pueden destacar:

1) Ofrece una base diferente para describir la percepción del riesgo en el contexto de un modelo de asignación de recursos (*Scott y Baker*, 1972).

2) Se puede resolver mediante programación lineal (*Hazell y Norton*, 1986).

3) Necesita relativamente poca información sobre los posibles márgenes brutos obtenidos (*Hazell y Norton,* 1986).

Entre los inconvenientes del modelo se encuentran:

1) Debilidad conceptual, debido a que ignora las relaciones de covarianza entre los márgenes brutos de las actividades (*Scott y Baker,* 1972; *Hazell y Norton,* 1986).

2) Dificultades de medida de los coeficientes de pérdida focal f_j y del valor de *k (Hazell y Norton,* 1986).

#1.3.5.-MODELO DE LOW

Low (1974) propuso una modificación del criterio de la Teoría de Juegos *maximin,* que selecciona el plan de cultivos con unos ingresos iguales o superiores a Y_0 para cada estado de la naturaleza, y que maximiza el ingreso esperado *E*:

$$maxE = \sum_j \bar{c}_j X_j$$

sujeto a:

$$\sum_j c_{jt} X_j \geq Y_0, \qquad \forall\ t$$

$$X \in F_d$$

El modelo tiene la misma solución que el modelo maximin si Y_0 es igual al máximo valor alcanzable de *M* (maximin). Una dificultad que tiene este modelo (*Hazell y Norton,* 1986) es que puede que no exista una solución factible si Y_0 es grande comparado con el máximo *E* alcanzable y/o si el agricultor opera en un ambiente de alto riesgo. Una aproximación alternativa es tratar Y_0 como un objetivo (como en la programación por metas) y buscar el plan de cultivos que se desvíe lo mínimo posible de éste.

#1.3.6.-MODELO TARGET-MOTAD

Tauer (1983) desarrolló el modelo Target MOTAD como adaptación del modelo MOTAD al contexto de la seguridad primero. Se formula como:

$$maxE = \sum_j \bar{c}_j X_j$$

sujeto a:

$$Y_0 - \sum_j c_{jt} X_j - Z_t^- \leq 0, \quad \forall\ t$$

$$\sum_t p_t Z_t^- = \lambda$$

$$X \in F_d$$

$$Z_t^- \geq 0, \quad \forall\ j, t$$

Las variables Z_t^- miden el valor de cualquier desviación en el ingreso por debajo del objetivo (Y_0).

El modelo pretende maximizar *E* sujeto a alcanzar un nivel satisfactorio (determinado por λ) de cumplimiento del objetivo, de forma que parametrizando λ se obtiene un conjunto de planes de cultivo eficientes.

Las ventajas de este modelo en la literatura son:

1) Comparando el modelo Target MOTAD con el modelo MOTAD:

Tauer (1983) señaló que sólo las soluciones Target MOTAD son eficientes en segundo grado de dominancia estocástica, demostrando la superioridad del modelo Target MOTAD sobre el MOTAD. También *Pope y Ziemer* (1984) demostraron que el conjunto eficiente generado con Target MOTAD es un subconjunto del segundo criterio de dominancia estocástica. *Berbel* (1993) indicó que a diferencia del modelo Target MOTAD, los modelos MOTAD y E-V no cumplen necesariamente el criterio de segundo grado de dominancia estocástica.

McCamley y Kliebenstein (1986) aplicaron en la comparación dos tests a los ejemplos propuestos por *Tauer* (1983) y *Watts et al.* (1984), identificando más soluciones MOTAD eficientes que dichos autores. El primer test es condición suficiente para que una solución MOTAD lo sea también Target MOTAD, utilizando precios sombra. El segundo test se basa en una condición necesaria para la eficiencia en primer grado de dominancia estocástica de las soluciones MOTAD.

2) Puede incluirse en un programa lineal.

3) La función de utilidad relacionada con Target MOTAD es superior a la relacionada con MOTAD (*Watts et al.,* 1984).

Entre los inconvenientes cabe destacar que los criterios de seguridad primero son en general incompatibles con la teoría de la utilidad estándard (*Bigman*, 1996), aunque según *Berbel* (1987), para modelizar las decisiones bajo condiciones de riesgo e incertidumbre, deben ser rechazados los métodos de optimización de la utilidad esperada, al ser muy costosos en tiempo e información. Este autor reivindica en un enfoque más empírico y realista del comportamiento.

#1.3.7.- LOS MODELOS DE SEGURIDAD PRIMERO EN LA LITERATURA ECONÓMICA AGRARIA

Varios autores han debatido el uso de desigualdades estocásticas, como *Roy* (1952), *Telser* (1955), *Kennedy y Francisco* (1974), *Anderson et al.* (1977), *Gabriel y Baker* (1980), *Sengupta* (1969), y *Berck y Hihn* (1982). Dichos estudios (con la excepción de Berck y Hihn) utilizaron la desigualdad de Tchebyshev para imponer restricciones probabilísticas a las variables aleatorias.

Boussard (1969) repitió el experimento de *Freund* (1956), ligeramente modificado, aplicando criterios alternativos a los de *Markowitz*, como el de *Roy* (1952) --que como puede resultar demasiado conservador, lo modifica y maximiza el ingreso medio, bajo la restricción de que la probabilidad de ruina no exceda un nivel dado--, de juegos contra la naturaleza (*McInerney*, 1967), o del pago de una estrategia (*Shackle*, 1961).

Pyle y Turnovsky (1970) transformaron los modelos de seguridad primero de *Roy* (1952), *Telser* (1955) y *Kataoka* (1963) en modelos que no utilizan la desigualdad de Tchebyshev (tal y como hacen los dos primeros), introduciendo parámetros de esperanza matemática y desviación estándard para caracterizar las distribuciones de las variables.

Kennedy y Francisco (1974) siguieron la aproximación de *Boussard y Petit* (1967), sin el supuesto de que la varianza del margen bruto total asociado con el plan de cultivos sea independiente de los niveles de las actividades. Compararon en un ejemplo tres formulaciones: dos MOTAD y una *Focus Loss Constrained Programming* (FLCP), que minimizaban el margen bruto total esperado sujeto a restricciones de *focus loss* y niveles de ingresos necesarios para cubrir gastos, que son parametrizados. Concluyen que aunque los coeficientes *loss* en las restricciones *focus loss* reflejan mejor las expectativas de los agricultores de lo que permiten los resultados basados en MOTAD, si se dispone de datos de series temporales en márgenes brutos, entonces MOTAD es más sencillo de utilizar y tiene la ventaja de la flexibilidad en la interpretación.

Barry y Robison (1975) adaptaron a la línea de *Scott y Baker* (1972) los criterios de seguridad primero, combinando la programación cuadrática y el criterio de Baumol.

Fishburn (1977) sugirió medidas de seguridad primero del riesgo basadas en el parámetro del momento parcial inferior, definido como:

$$\rho(\varepsilon, z) = \int_{-\infty}^{z} (z-x)^{\varepsilon} f(x)dx$$

donde $\varepsilon \geq 0$. Cuando $z = \mu$ y $\varepsilon = 2$, este parámetro es la media semivarianza; para un z dado y $\varepsilon = 2$, este parámetro es idéntico a la semivarianza de *Porter* (1974). Fishburn demuestra que si $\varepsilon \geq 1$, entonces una evaluación del intercambio entre μ y el parámetro $\rho(\varepsilon, z)$ generaría soluciones pertenecientes al subconjunto del conjunto de dominancia estocástica de segundo grado.

Berck y Hihn (1982) y *Atwood* (1985) sugirieron métodos para estimar las probabilidades de desastre (incluso sin conocimiento detallado de los estimadores empíricos de la función de distribución acumulada de las variables), partiendo de la desigualdad de Tchebyshev. Cuando la distribución de los pagos es asimétrica, *Berck y Hihn* sugieren un estimador distinto de los límites de probabilidad que se basa en la media y la semivarianza (retomando la idea avanzada por *Markowitz* de que la media-semivarianza, donde sólo las desviaciones por debajo de la media se elevan al cuadrado, es teóricamente más atractiva como medida del riesgo que la varianza). El problema principal con una medida de probabilidad del riesgo es que no refleje la profundidad del desastre dado que proporciona la misma medida del riesgo independientemente de si la cantidad por debajo de *z* es muy grande o muy pequeña.

Según *Barry* (1984), los modelos de seguridad primero no penalizan las grandes desviaciones por debajo de la media como hacen los modelos E-V y E-σ.

Watts (1984), aplicó el modelo Target MOTAD maximizando el ingreso y dejando la suma de las desviaciones como una restricción paramétrica, para poder así generar el conjunto eficiente.

Los resultados de un estudio realizado por *Patrick et al.* (1985) indican que los modelos de seguridad primero pueden ser importantes en las decisiones empresariales.

Berbel (1987, 1990) comenta que el agricultor ve el riesgo “hacia abajo” y lo que le interesa principalmente es la probabilidad de no alcanzar unos ingresos dados, lo que es más comprensible que algunos parámetros estadísticos como la varianza.

Atwood et al. (1988) creen que los modelos de seguridad primero constituyen una alternativa en la cual el decisor se encuentra con una probabilidad de fracaso de lograr sus metas. Presentan un método en que las restricciones probabilísticas pueden ser reforzadas fácilmente en un modelo Target MOTAD ligeramente modificado, en el que se utiliza la desigualdad estocástica del momento parcial inferior lineal (LPM) presentada por *Atwood* (1985). El procedimiento es flexible y sólo requiere que la variable aleatoria sea finitamente discreta (o se pueda aproximar como tal).

Zwingli et al. (1989) aplicaron el modelo Target MOTAD a rotaciones de vegetales. Descartan modelos alternativos como MOTAD o la programación cuadrática y eligen este modelo por su fácil implementación y porque la utilización de ingresos ‘objetivo’ permite simular la forma en que muchos agricultores toman sus decisiones de producción. Además, el modelo Target MOTAD también refleja el interés de los productores por las desviaciones de los ingresos por debajo de un nivel esperado, y no en las desviaciones absolutas respecto al nivel de ingresos medios, tal como se realiza en el modelo MOTAD.

Novak et al. (1990) utilizaron el modelo Target MOTAD y diez años de datos experimentales para determinar un esquema de rotación que minimice el riesgo para un nivel aceptable de beneficio.

Flichman, Garrido y Varela-Ortega (1994) utilizaron la aproximación de *Freund* y la combinaron con el modelo Target MOTAD para analizar las consecuencias de la reforma de la política agraria comunitaria, examinando los efectos económicos y medioambientales, así como las sustituciones entre ellos.

Bigman (1996) propuso definir una medida cardinal en línea con los criterios de seguridad primero, y que represente la probabilidad de no alcanzar el objetivo, la amplitud del desastre, y la pérdida esperada de utilidad por no alcanzar el objetivo mínimo. Esa medida tiene las propiedades de la función de utilidad de *Von Neumann-Morgenstern*.

#1.4.- PROGRAMACIÓN MULTIOBJETIVO

Berbel (1987) cree que cualquier análisis del riesgo es por su misma naturaleza de carácter multicriterio.

#1.4.1.- MODELO MEDIA-DAP

Berbel (1988) presentó el modelo Media-DAP, como una combinación del modelo Target MOTAD (*Tauer*, 1983) y el propuesto por *Atwood* (1985) en un contexto de programación multiobjetivo. Este último autor propuso utilizar momentos parciales de orden inferior para mejorar los resultados de la desigualdad de la semivarianza generalizada de *Berck y Hihn* (1982), introducida para evitar el efecto conservador de la desigualdad de Tchebyshev.

Realizando algunos cambios a la misma, se llega a la expresión:

$Pr(x < g) \leq Q(1, t) : (t-g)$

siendo g el nivel umbral de seguridad, t un parámetro auxiliar para estimar la probabilidad de fracaso (debería cumplirse $t > g$), $Pr(x < g)$ la probabilidad de fracaso y $Q(1, t)$ la Desviación Absoluta Parcial o DAP, base del modelo propuesto por *Berbel*.

La minimización de la probabilidad de fracaso equivale a la minimización del DAP. La probabilidad de fracaso se calcula a partir de la expresión anterior sustituyendo $Q(1, t)$ por (DAP/m), siendo m igual al número de años o períodos considerados:

$Pr(x < g) \leq DAP / m(t -g)$

Berbel utilizó en el modelo Target MOTAD t = *maximin*, argumentando que bajo este nivel el conjunto eficiente generado será (salvo raras excepciones) el de mayor amplitud posible. De esta forma, el conjunto eficiente riesgo-ingreso está acotado por dos extremos: el de mayor ganancia (y también mayor riesgo) y el de menor riesgo (maximin). En este segundo caso, las desviaciones son nulas (DAP=0). Una vez acotado el problema, para calcular el conjunto eficiente puede utilizarse cualquiera de los métodos habituales (ver *Cohon et al.* 1978), aunque Berbel utilizó el método NISE (NonInferior Set Estimation, propuesto por dichos autores).

En definitiva, la estructura del modelo es:

$Eff\ (MB, DAP)$

$MB = MX$

$DAP = \Sigma Pr(i)N(i)$

Sujeto a:

$X \in F_d$

$S(i)X + N(i) \geq t \quad \forall i = 1, 2, ..., m$

donde *Eff* significa conjunto eficiente, *MB* el margen bruto, *M* el vector margen bruto esperado por unidad de nivel de actividad, *X* el vector de niveles de actividad, *DAP* la suma de desviaciones negativas desde t para los m años (estados de la naturaleza), ponderadas mediante probabilidades, $S(i)$ el vector de márgenes brutos para los m años (estados de la naturaleza), $N(i)$ el vector de desviaciones negativas y t un parámetro (escalar).

#1.4.2.- EL MODELO MEDIA-DAP EN LA LITERATURA ECONÓMICA AGRARIA

Berbel (1987, 1988, 1990a, 1990b, 1990c, 1993) introdujo el modelo Media-DAP y realizó diversas aplicaciones del mismo. *Millán, J. S.* (1992) combinó este modelo con la programación con restricciones probabilísticas y lo aplicó a los regadíos del Bembézar. *Alaejos* (1993) lo utilizó para seleccionar planes de cultivo en contexto de riesgo en el Bembézar (Córdoba), y *Millán y Berbel* (1994), junto con la programación con restricciones aleatorias, lo utilizaron para describir mediante programación multicriterio la agricultura en el Valle del Guadalquivir. Finalmente *Millán y Ruiz* (1995) aplicaron el modelo como medida de riesgo dentro de una programación multiobjetivo, en la que se consideran objetivos públicos y privados en Córdoba.

#1.4.3.- PROGRAMACIÓN COMPROMISO

Según *Ballestero y Romero* (1991), la programación compromiso es una aproximación de la decisión multicriterio desarrollada por *Yu* (1973) y *Zeleny* (1973). La idea básica es la identificación de la solución ideal como punto donde cada atributo considerado alcanza su valor óptimo. Cuando existe un conflicto entre los atributos considerados (frecuente en la realidad), el punto ideal, aún siendo infactible o inalcanzable, se utiliza como referencia, pues la programación compromiso define la solución óptima como la solución eficiente más cercana al punto ideal.

Según *Romero et al.* (1988), la utilización de la programación riesgo-compromiso permite evitar el problema, prácticamente irresoluble, de la determinación de la función de utilidad del decisor. Según *Zeleny* (1973), es posible identificar los límites del conjunto eficiente donde se produce la tangencia entre la curva de iso-utilidad y la de eficiencia, sin necesidad de supuestos sobre la forma de las curvas de utilidad.

El concepto de distancia se utiliza en este caso como medida práctica para las preferencias humanas, más que en su sentido geométrico. Se utiliza una familia de métricas *Lp* para calcular distancias entre soluciones pertenecientes al conjunto eficiente y el punto ideal. *Yu* (1973) demostró que los puntos definidos con todas las métricas pertenecen al conjunto de puntos definidos por las métricas L_1 y L_∞.

Según *Romero* (1993) si se representan por W_j las preferencias que el centro decisor asocia a la discrepancia existente entre la realización del objetivo *j*-ésimo y su ideal, la programación compromiso se convierte en el siguiente problema de optimización:

$$Min\, L_p = \left[\sum_{j=1}^{n} W_j^p \left| \frac{f_j^* - f_j(x)}{f_j^* - f_{*j}} \right|^p \right]^{1/p}$$

sujeto a: $x \in F_d$

siendo f_j el objetivo *j*-ésimo, f_j^* el ideal de dicho objetivo *j*-ésimo y f_{*j} el anti-ideal de dicho objetivo.

Para la *métrica p=1*, la mejor solución compromiso o punto más próximo al ideal se puede obtener resolviendo el siguiente programa lineal:

$$Min\, L_1 = \sum_{j=1}^{n} W_j \ \frac{f_j^* - f_j(x)}{f_j^* - f_{*j}}$$

sujeto a: $x \in F_d$

Para la *métrica* $p=\infty$, se minimiza la máxima desviación de entre todas las desviaciones individuales, pues para dicha métrica sólo la mayor desviación influye en el proceso de minimización. En este caso, la mejor solución compromiso o punto más próximo al ideal se puede obtener resolviendo el siguiente programa lineal:

$$\begin{aligned}
&Min\ L_\infty = d \\
&sujeto\ a: \\
&x \in F_d \\
&W_1 \frac{f_1^* - f_1(x)}{f_1^* - f_{*1}} \leq d \\
&\vdots \\
&W_i \frac{f_i^* - f_i(x)}{f_i^* - f_{*i}} \leq d \\
&\vdots \\
&W_q \frac{f_q^* - f_q(x)}{f_q^* - f_{*q}} \leq d
\end{aligned}$$

donde *d* representa la mayor desviación.

#1.4.4.- LA PROGRAMACIÓN COMPROMISO EN LA LITERATURA ECONÓMICA AGRARIA

Romero et al. (1988) utilizaron la programación riesgo-compromiso como una nueva aproximación para tratar el riesgo en decisiones de asignación de recursos en agricultura, sintetizando los modelos de *Markowitz* o MOTAD en un modelo de programación compromiso, para generar las mejores soluciones compromiso. *Ballestero y Romero* (1991) establecieron las condiciones que debe satisfacer la función de utilidad para poder garantizar que su máximo pertenece al conjunto compromiso. Más tarde, *Ballestero y Romero* (1993) probaron la superioridad de un sistema de pesos comúnmente utilizado en programación compromiso, concretamente los pesos inversamente proporcionales al ideal, dado que pueden ser identificados con los precios sombra de los escenarios económicos.

Millán y Ruiz (1995) utilizaron la programación compromiso para la determinación de soluciones eficientes en planificación de cultivos con objetivos públicos y privados en Córdoba.

#1.5.- EFICIENCIA ESTOCÁSTICA

Las técnicas de eficiencia estocástica permiten evaluar el impacto del riesgo cuando es imposible obtener las funciones de utilidad de los agricultores (*Anderson et al.*, 1977). Como ventaja de la dominancia estocástica (SD) hay que decir que constituye una aproximación a la selección de carteras que utiliza la distribución de probabilidad entera, en lugar de utilizar un número finito de momentos (*Booth et al*., 1985).

Como crítica, comentar que la gran cantidad de información necesaria para operativizar dichos métodos (hace falta conocer la distribución de probabilidades), junto con el carácter normativo de los mismos, los hacen poco interesantes para un modelo de decisión que trate de ser realista (*Berbel*, 1987).

#1.5.1.- EFICIENCIA ESTOCÁSTICA EN PRIMER GRADO

El concepto de *eficiencia estocástica en primer grado (FSD)*, elaborado por *Quirk y Saposnik* (1962), y formalizado por *Fishburn* (1964), se basa en la idea de que si x es una medida sin escala como puede ser el beneficio, los decisores siempre preferirán un x mayor. Más formalmente, esto equivale a suponer una función de utilidad monótona creciente en que la derivada primera sea estrictamente positiva, $U_1(x)>0$, permitiendo por otro lado preferencia, indiferencia, o aversión al riesgo.

La eficiencia estocástica en primer grado considera un par de funciones continuas F_1 y G_1, definidas en el rango $[a, b]$ y asociadas respectivamente a dos actuaciones o situaciones arriesgadas F y G. Se dice que F domina a G en el sentido de primer grado de dominancia estocástica si $F_1(R) \leq G_1(R)$ para todos los posibles R en el rango $[a, b]$ con una desigualdad estricta por lo menos para un valor de R.

Según *Mjelde, Lacewell, Talpaz y Taylor* (1990), el criterio de primer orden de dominancia estocástica (FSD) no es demasiado restrictivo y se aplica tanto a decisores adversos como a amantes al riesgo.

#1.5.2.- EFICIENCIA ESTOCÁSTICA EN SEGUNDO GRADO

Otro concepto dentro de este ámbito es el de *eficiencia estocástica en segundo grado (SSD)*, que consigue reducir el conjunto eficiente. Las reglas para identificar estos casos fueron determinadas por *Fishburn* (1964), *Hammond* (1968), *Hadar y Russell* (1969) y *Hanoch y Levy* (1969) y dependen sólo del supuesto de que el decisor se comporta como adverso al riesgo. La función de utilidad en el rango $[a, b]$ de los posibles resultados no es solamente monótona creciente sino también estrictamente cóncava. Esto es equivalente a $U_1(x)>0$ y $U_2(x)<0$, siendo SSD un criterio más discriminatorio que FSD.

Las distribuciones dominadas son ineficientes en el sentido de que nunca serán preferidas por un decisor adverso al riesgo que busque maximizar la utilidad.

Se define el segundo grado de dominancia estocástica acumulado para una distribución F_1 como:

$$F_2(R) = \int_0^R F_1(x)dx$$

La distribución F domina a G en el sentido de segundo grado de dominancia estocástica si $F_2(R) \leq G_2(R)$ para todos los posibles valores R , con una desigualdad estricta por lo menos.

En el caso de las distribuciones normales, es más interesante la dominancia estocástica de segundo grado que la de primer grado. Específicamente, una F normal domina a una G normal en el sentido de segundo grado de dominancia estocástica si $E_F(x) \geq E_G(x)$ y $V_F(x) \leq V_G(x)$ con una desigualdad estricta como mínimo, donde E es el operador esperanza matemática y V es el operador varianza. En variables normales es equivalente el conjunto eficiente E-V al conjunto eficiente estocástico de segundo grado (*Hannoch y Levy*, 1969; *McCarl et al*., 1987).

Según *Barry* (1984), la eficiencia del criterio SSD, así como la de los criterios E-V y MOTAD, puede ser irreal si no se cumple el supuesto de aversión al riesgo. *Pope y Ziemer* (1984) utilizaron el segundo grado de dominancia estocástica porque este grado ha recibido más atención en el análisis del riesgo en agricultura.

#1.5.3.- EFICIENCIA ESTOCÁSTICA EN TERCER GRADO

El concepto de *tercer grado de dominancia estocástica (TSD)* supone que la tercera derivada de la función de utilidad es positiva ($U_3>0$). Se debe a *Whitmore* (1970) y *Hammond* (1974), y requiere la definición de otro tipo de función acumulada:

$$F_3(R) = \int_0^R F_2(x)dx$$

La distribución F domina a G en el sentido de dominancia estocástica de tercer grado si $F_3(R) \leq G_3(R)$ para todos los posibles R con una desigualdad estricta por lo menos y si $F_2(b) \leq G_2(b)$, entonces b es el rango superior, o equivalentemente, $E_F(x) \geq E_G(x)$.

TSD incorpora los supuestos de FSD y SSD, añadiendo que la tercera derivada de la función de utilidad respecto a los ingresos (x), sea no negativa en todo el dominio ($U_3 \geq 0$).

Varios autores desaconsejan el tercer grado de dominancia estocástica frente al segundo, dado que el coste adicional no cubre el beneficio marginal de identificar un conjunto eficiente ligeramente inferior.

#1.5.4.- EFICIENCIA MEDIA-VARIANZA

Según *Barry* (1984), la eficiencia media-varianza (E-V), debida a *Markowitz* (1959), es el criterio de eficiencia más familiar y ampliamente utilizado. Al igual que el criterio SSD, la eficiencia E-V requiere que el decisor sea adverso al riesgo. Además, las distribuciones deben ser normales o la función de utilidad del decisor cuadrática. Cuando se cumplen estas condiciones, el conjunto eficiente E-V es equivalente al conjunto eficiente SSD.

Según este criterio, la distribución F, de media E_F y varianza V_F, domina a la distribución G, de media E_G y varianza V_G si $E_F \geq E_G$ y $V_F \leq V_G$ y si una de estas desigualdades es estricta.

Tsiang (1972) y *Levy y Markowitz* (1979) demostraron que el criterio E-V sirve a menudo como aproximación razonable a SSD. Según *Barry* (1984), la eficiencia de este criterio, además de los criterios SSD y MOTAD, puede ser irreal si no se cumple la hipótesis de aversión al riesgo.

#1.5.5.- EFICIENCIA MEDIA-DESVIACIÓN ABSOLUTA

El criterio de la desviación absoluta media (MOTAD) es una aproximación a la eficiencia E-V que puede ser modelizada mediante programación lineal (*Hazell*, 1971). Aunque no se pueden establecer vínculos directos analíticamente entre el criterio MOTAD y la forma de la función de utilidad del decisor, se supone que la eficiencia MOTAD se verifica para decisores adversos al riesgo, y cuando las distribuciones ordenadas son aproximadamente normales, el conjunto eficiente MOTAD es muy parecido al conjunto eficiente E-V.

Bajo este criterio, se utilizan la media *E* y la desviación absoluta media *A* de las distribuciones de los rendimientos para ordenar las alternativas. Se define la desviación absoluta media, *A*, como:

$$A = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \mu_i\right|$$

siendo n el número de niveles de rendimientos observados, y_i el nivel de rendimiento i observado, y μ_i el valor esperado de la variable rendimiento. Una distribución F, de media E_F y desviación absoluta media A_F, domina a otra distribución G, de media E_G y desviación absoluta media A_G, si $E_F \geq E_G$ y $A_F \leq A_G$ y si una de estas desigualdades es estricta.

Según *Barry* (1984), la eficiencia del criterio MOTAD, además de la de los criterios SSD y E-V, puede ser irreal si no se verifica la hipótesis de aversión al riesgo.

#1.5.6.- DOMINANCIA ESTOCÁSTICA RESPECTO A UNA FUNCIÓN

La *dominancia estocástica respecto a una función* (SDRF), también conocida como *dominancia estocástica generalizada*, es un criterio que ordena alternativas arriesgadas en intervalos de preferencia al riesgo seleccionados y definidos por el coeficiente de aversión al riesgo absoluto de Pratt-Arrow. La función de aversión al riesgo absoluta se define como:

$R_a(y) = -U''(y)/U'(y)$

siendo $U''(y)$ la segunda derivada de la función de utilidad y $U'(y)$ la primera.

Más formalmente, SDRF establece las condiciones necesarias y suficientes bajo las cuales la función de distribución acumulada $F(y)$ es preferida a la función $G(y)$ para aquellos individuos cuyas funciones de aversión al riesgo absolutas se encuentran entre los límites inferior y superior $r_1(y)$ y $r_2(y)$.

Tal y como *Meyer* (1977a) lo desarrolló, el procedimiento de solución requiere la identificación de la función de utilidad $u_0(y)$, que minimiza:

$$\int_{-\infty}^{\infty}\left[G(y) - F(y)\right]u'(y)dy$$

sujeto a la restricción

$r_1(y) \leq -u''(y) / u'(y) \leq r_2(y) \;\; \forall y.$

La primera expresión se refiere a la diferencia entre las utilidades esperadas de las distribuciones *F(y)* y *G(y)*. Si el mínimo de esta diferencia es positivo, entonces *F(y)* es unánimemente preferida a *G(y)* (esta condición implica que la utilidad esperada de *F(y)* es siempre mayor que la de *G(y)*). Si el mínimo es cero, el decisor estará indiferente entre las dos alternativas, y si el mínimo es negativo, *F(y)* no será unánimemente preferida a *G(y)*.

El criterio SDRF es más discriminatorio (*Barry*, 1984) que los criterios FSD, SSD, E-V y MOTAD, pues ninguno reduce de manera fiable el gran número de opciones a un conjunto eficiente que pueda ser ordenado directamente por el decisor. Además, permite mayor flexibilidad representando las preferencias individuales (*Meyer*, 1977b), aunque también requiere información más detallada sobre ellas.

Según *Maynard et al.* (1997), una característica atractiva de SDRF para el investigador es que no requiere un conocimiento específico de la función de utilidad de los individuos. Otra ventaja es su capacidad para evaluar el rango completo de preferencias de riesgo, desde el amante hasta el adverso.

#1.5.7.- LA DOMINANCIA ESTOCÁSTICA EN LA LITERATURA ECONÓMICA AGRARIA

Según *Porter* (1973) una consecuencia importante de las críticas a la aproximación E-V ha sido el desarrollo de los criterios de dominancia estocástica para la selección de carteras. *Porter* (1974) demostró que puede utilizarse un punto de referencia fijo de la semivarianza para generar soluciones que constituyan un subconjunto del segundo grado estocásticamente no dominado. *Anderson* (1974) discutió el concepto de dominancia estocástica.

Bawa (1975), *Bey* (1979), y *Nantell, et al.* (1982) analizaron detalladamente la relación entre momentos parciales inferiores y dominancia estocástica.

Fishburn (1977) presentó una forma general del momento parcial inferior, en la cual, variando el valor de la constante α se pueden generar soluciones pertenecientes al segundo grado de dominancia estocástica si $\alpha \geq 1$, y al tercer grado si $\alpha \geq 2$.

Kramer y Pope (1981) definieron la dominancia estocástica como una herramienta popular para el análisis de la eficiencia en la elección discreta. Cuando existe, por ejemplo, una mezcla continua de beneficios correspondientes a distintos cultivos, los procedimientos de dominancia estocástica deben comparar un número infinito de planes de cultivo, no constituyendo una técnica útil por no tener un número finito y reducido de opciones de cultivo.

Según *Barry* (1984), los criterios de eficiencia son útiles en situaciones relacionadas con un único decisor de preferencias desconocidas, en situaciones que impliquen a varios decisores cuyas preferencias difieran de acuerdo con un conjunto específico de restricciones, y en el análisis de políticas alternativas o de recomendaciones de extensión que afecten a varios individuos. También son útiles en la derivación de resultados teóricos de amplia aplicación.

Pope y Ziemer (1984) plantean en su artículo una comparación en términos de eficiencia estocástica del criterio E-V y el criterio SSD, comprobando que para variables aleatorias que no se distribuyan normalmente, ambos análisis dan lugar a diferentes conjuntos eficientes. Para elecciones que comporten un número finito de comparaciones, es preferible la aproximación SSD dado que no requiere la condición de normalidad.

Zacharias y Grube (1984) utilizaron las técnicas de dominancia estocástica para estudiar las rotaciones de cultivos como herramienta de control de malas hierbas.

Booth et al. (1985) expusieron que los criterios de eficiencia estocástica E-V son apropiados sólo si la función de utilidad del decisor es cuadrática con derivadas segundas no positivas o si la distribución de los ingresos puede ser caracterizada completamente por los dos primeros momentos (por ejemplo, distribuciones normal, uniforme o binomial). Cualquiera es suficiente para que la aproximación E-V sea consistente con la maximización de la utilidad de *Von Neuman-Morgenstern*.

Musser et al. (1985) aplicaron una programación lineal generalizada a varias rotaciones potenciales, debido a la facilidad para examinar carteras de rotaciones que tienen las aproximaciones de programación matemática sobre la dominancia estocástica.

Brown (1987) encontró una estrecha correspondencia entre el comportamiento del productor y los resultados de dominancia estocástica en un estudio de la agricultura en Saskatchewan.

McCarl et al. (1987) comentaron que, bajo normalidad, el análisis E-V es preferible dado que se tiene en cuenta explícitamente la diversificación. Desarrollaron unas reglas indicando cuándo la dominancia de una alternativa sobre otra implica la dominancia sobre todas las combinaciones lineales de las dos, llegando a un teorema. Según estos autores, el criterio SSD es demasiado conservador, y por ello desarrollan una regla alternativa basada en la teoría de la utilidad esperada, en forma de teorema. Señalaron también que los criterios de dominancia estocástica deben aplicarse con precaución cuando se comparan alternativas no mutuamente excluyentes o perfectamente correlacionadas. *Samuelson* (1967) y *Hadar y Russell* (1971, 1974) habían demostrado que la diversificación es óptima cuando existe independencia.

Tolley y Pope (1988) desarrollaron un test no paramétrico para la dominancia estocástica de segundo grado.

Williams (1988) defendió la dominancia estocástica argumentando que no requiere que la distribución sea normal y es más flexible que el análisis E-V. Además, es consistente con la hipótesis de la utilidad esperada como modelo general para la toma de decisiones bajo incertidumbre.

Lowenberg-DeBoer y Cherney (1989) utilizan criterios de dominancia estocástica para determinar las estrategias eficientes obtenidas mediante simulación.

Mjelde et al. (1990) utilizaron criterios de dominancia estocástica.

Berbel (1993) comentó que los criterios de dominancia estocástica constituyen un enfoque diferente de la programación bajo riesgo, y consisten en que esencialmente no hay restricciones en las preferencias del decisor.

Anderson (1997) comentó que en los últimos años han existido estructuras analíticas basadas en la utilidad (como el análisis de eficiencia estocástica), que quizás no han alcanzado lo que deberían.

Hardaker et al. (1997) comentaron que una vez los métodos de análisis de eficiencia estocástica se hayan mejorado y extendido, cobrarán menor importancia los problemas sobre las limitaciones de las funciones de utilidad y su determinación.

Maynard et al. (1997) realizaron un análisis de dominancia estocástica de cinco rotaciones de cultivos, utilizando datos de rendimientos correspondientes a veintiún años. Señalaron que la elevada incidencia relativa del comportamiento amante del riesgo en los estudios de *Wilson y Eidman* (1983) y *Tauer* (1986) es un factor que influye en la decisión de utilizar la dominancia estocástica generalizada en su trabajo.

McCamley y Rudel (1997) sugirieron una aproximación gráfica a la dominancia estocástica generalizada, con la idea de sustituir o complementar los análisis de sensibilidad.

Saatkamp et al. (1997) realizaron un análisis de dominancia estocástica respecto a una función, utilizando el programa desarrollado por *Goh et al.* (1989), y aplicado al estudio del control de enfermedades contagiosas.

#1.6.- OTROS MODELOS

#1.6.1.- ÍNDICE SIMPLE

Una aproximación destacable la constituye el modelo de *Sharpe* (1963), diagonal o de índice simple, que permite reducir el trabajo de cálculo en las múltiples comparaciones que deben realizarse en los modelos estándard de cartera de valores. La principal característica de este modelo es el supuesto de que los beneficios de las distintas acciones están vinculados sólo a través de relaciones comunes con algunos factores básicos. El beneficio de cualquier acción viene dado por:

$$R_i = A_i + B_i I + C_i$$

donde A_i y B_i son parámetros, C_i es una variable aleatoria de valor esperado nulo y varianza Q_i, e I es el nivel de un índice dado. Dicho índice I puede ser el Producto Nacional Bruto, un índice de precios o cualquier otro factor que pueda influir de forma decisiva en ese contexto.

El nivel futuro de I viene determinado en parte por factores aleatorios:

$$I = A_{n+1} + C_{n+1}$$

donde A_{n+1} es un parámetro y C_{n+1} es una variable aleatoria con valor esperado nulo y varianza Q_{n+1}. Se supone que la covarianza entre C_i y C_j es nula para todos los valores de i y j ($i \neq j$).

En el modelo diagonal el analista debe hacer las siguientes predicciones: por un lado los valores de A_i, B_i y Q_i para cada una de las N acciones, y por otro los valores de A_{n+1} y Q_{n+1} para el índice I, reduciéndose significativamente en este modelo el número de estimadores, comparado con el de *Markowitz* (1952).

Una vez especificados los parámetros del modelo diagonal, pueden derivarse todas las variables necesarias en el análisis de la cartera de valores estándard. Dichas relaciones son:

$$E_i = A_i + B_i (A_{n+1})$$

$$V_i = (B_i)^2 (Q_{n+1}) + Q_i$$

$$C = (B_i)(B_j)(Q_{n+1})$$

#1.6.2.- EL MODELO DE ÍNDICE SIMPLE EN LA LITERATURA ECONÓMICA AGRARIA

Alonso y Rodríguez (1983) aplicaron este modelo al estudio de la relación entre los rendimientos y un índice pluviométrico en zonas de secano del Duero. *Alonso et al.* (1987) lo utilizaron para la determinación de los períodos críticos, cuando se consideran los índices pluviométricos y termométricos en los cultivos de secano más importantes de la Comunidad Autónoma de Madrid. También estimaron el riesgo económico de los cultivos y los índices de "performance" de los mismos. *Serrano* (1987) adaptó este modelo al sector agrario para los cultivos de secano.

Millán y Millán (1995) analizaron la aplicabilidad del modelo de Sharpe en planificación agraria. No discutieron la validez teórica del modelo, sino algunos problemas generados a la hora de aplicarlo, como la elección de los índices, aunque subrayaron su robustez. *Millán y Millán* (1996) lo aplicaron a los cultivos de regadío de Córdoba, y lo compararon con el de *Markowitz*.

#1.6.3.-SIMULACIÓN

Lutgen y Helmers (1979) utilizaron la simulación de una explotación agraria de Nebraska para generar relaciones de ingresos arriesgadas en combinaciones alternativas de producción y marketing. La simulación fue también utilizada por *Mapp et al.* (1979) para evaluar planes derivados del modelo MOTAD. La simulación completa el modelo MOTAD evaluando los efectos acumulativos de la planificación de cultivos en el tiempo.

Dean et al. (1980) analizaron el efecto de los programas gubernamentales de gestión del riesgo sobre la supervivencia y el crecimiento de las explotaciones. *Richardson y Nixon* (1981), en cooperación con el USDA, simularon los de ingresos una explotación en función de diversas políticas (FLIPSIM).

Calatrava y Domingo (1982) emplearon la simulación en la resolución de programas estocásticos. Concretamente estudiaron el caso de los programas lineales estocásticos con función objetivo aleatoria, y los aplicaron a la planificación de cultivos en una explotación hortícola de la costa mediterránea de Andalucía. *Calatrava* (1986) realizó algunas reflexiones sobre la simulación aplicada a explotaciones hortícolas, adjuntando una revisión bibliográfica.

Consiglio y Zenios (1999) han integrado modelos de simulación Monte Carlo y modelos de optimización en el diseño de carteras de productos financieros.

Por nuestra parte, en los trabajos llevados a cabo en el desarrollo de la Tesis, se han desarrollado rutinas para enlazar generadores de funciones de densidad de cualquier coeficiente con el programa LINDO de resolución de programas lineales (Anejo 9). De esta forma:

(1) Se genera un valor para cada coeficiente aleatorio, siguiendo las leyes correspondientes.

(2) Esos valores se introducen en el programa lineal, que se resuelve mediante LINDO, obteniéndose una solución óptima del programa, que se almacena.

(3) Se vuelve al punto (1) las veces que sea necesario para generar la función de densidad de las soluciones (2).

#1.6.4.- ALEATORIEDAD EN LAS RESTRICCIONES

Las técnicas más habituales en este tipo de problemas son la programación estocástica discreta, la programación con restricciones aleatorias y la programación en dos etapas bajo incertidumbre.

#1.6.4.1.- PROGRAMACIÓN ESTOCÁSTICA DISCRETA

Según *Hazell y Norton* (1986), todos los riesgos en el conjunto de restricciones deberían ser transferidos a la función objetivo del modelo y así aplicar sólo un criterio de decisión. Pero esto sólo puede hacerse si los recursos son intercambiables libremente, y entonces cualquier discrepancia aleatoria entre las necesidades en recursos y la oferta de los mismos puede ser trasladada a la función objetivo a través de la compra y la venta de actividades.

En este modelo, se supone que las decisiones son tomadas después de observarse el estado de la naturaleza. El principal inconveniente es que necesita muchos datos, y además pronto toma grandes dimensiones dada la necesidad de tener filas y columnas separadas para recursos y actividades en cada estado de la naturaleza.

Cuando los recursos no son intercambiables, el riesgo ya no puede ser transferido a la función objetivo mediante compra y venta de actividades, llegando a planes de cultivo conservadores, que pueden ser incompatibles con la aversión al riesgo de los agricultores.

Cocks (1968) presentó un método para resolver problemas de programación lineal con distribuciones discretas de probabilidad en los coeficientes *c*, *A* o *b*, formulando la función objetivo en términos de varianza y/o esperanza matemática. *Rae* (1971a, 1971b), demostró que la programación estocástica discreta puede ser utilizada para modelizar secuencias de decisiones en agricultura.

#1.6.4.2.-. PROGRAMACIÓN CON RESTRICCIONES ALEATORIAS

La programación con restricciones aleatorias, desarrollada por *Charnes y Cooper* (1959), expresa las necesidades de factibilidad en términos probabilísticos. Una restricción aleatoria para el recurso *i* toma la forma

$$\Pr\left\{\sum_j a_{ij} X_j \le b_i\right\} \ge 1-\alpha_i$$

donde α es una probabilidad preestablecida (normalmente 10% o menos). La restricción indica que las necesidades totales para el recurso *i* no deben exceder su disponibilidad en más de un α por ciento del tiempo. Para poder incorporar esta ecuación a un modelo de programación lineal, es necesario convertirla en una restricción de desigualdad lineal, cosa que suele hacerse suponiendo que todos los coeficientes aleatorios están normalmente distribuidos. Pueden distinguirse dos casos:

1) Que los **coeficientes a_{ij}** sean **estocásticos**, con medias $\bar{a}_{ij}$ y covarianzas *cov* (a_{ij}, a_{ik}), siendo los coeficientes b_i determinísticos. En este caso, la suma

$$\sum_j a_{ij} X_j$$

(la cual denotamos por Z_i) constituye una variable aleatoria normalmente distribuida de media

$$E\left[Z_i\right] = \sum_j \bar{a}_{ij} X_j$$

y desviación estándar

$$\sigma_{zi} = \left[\sum_j \sum_k X_j X_k \operatorname{cov}(a_{ij}, a_{ik}) \right]^{1/2}$$

La ecuación de la restricción probabilística se convierte en:

$$\Pr\{Z_i \le b_i\} \ge 1 - \alpha_i$$

o de forma equivalente

$$\Pr\left\{ \left(Z_i - E\left[Z_i\right]\right) / \sigma_{zi} \le \left(b_i - E\left[Z_i\right]\right) / \sigma_{zi} \right\} \ge 1 - \alpha_i$$

Ahora $(Z_i - E[Z_i])/\sigma_{zi}$ es una variable aleatoria normal estándard con media cero y desviación estándard uno, por lo que las tablas para la distribución normal acumulada pueden utilizarse para buscar, para cualquier α_i, una constante K_α tal que

$$\Pr\{(Z_i - E[Z_i])/\sigma_{zi} \le K_\alpha\} = 1 - \alpha_i$$

Comparando las expresiones última y antepenúltima, la restricción aleatoria se satisfará si $K_\alpha \le \left(b_i - E\left[Z_i\right]\right) / \sigma_{zi}$ o de forma equivalente, si $E[Z_i] + K_\alpha \sigma_{zi} \le b_i$. Sustituyendo las expresiones para $E[Z_i]$ y σ_{zi}, la restricción aleatoria se convierte en

$$\sum_j \bar{a}_{ij} X_j + K_\alpha \left[\sum_j \sum_k X_j X_k \operatorname{cov}(a_{ij}, a_{ik}) \right]^{1/2} \le b_i$$

Añadiendo esta ecuación al modelo queda de programación cuadrática, aunque puede linealizarse mediante la transformación de *Wicks y Guise* (1978), ya estudiada en los modelos de riesgo en la función objetivo (ver #1.2.5.).

2) Que la **aleatoriedad** se encuentre **en la disponibilidad de recursos $\mathbf{b_i}$**. Se opera de la misma forma llegando en este caso a un programa lineal y no cuadrático como venimos de obtener.

En este caso, siendo $E(b_i)$ y σ_{bi} la esperanza matemática y la desviación estándard respectivamente de la disponibilidad de recursos b_i, la restricción correspondiente queda:

$$\Pr\left\{ \left(\sum a_{ij} X_j - E(b_i)\right) / \sigma_{bi} \le \left(b_i - E(b_i)\right) / \sigma_{bi} \right\} \ge 1 - \alpha_i$$

donde $(b_i - E(b_i)) / \sigma_{bi}$ es una variable aleatoria normal estándard con media cero y desviación estándard uno. Por lo tanto, las tablas para la distribución normal acumulada pueden utilizarse para buscar, para cualquier α_i, una constante K_α tal que

$$\Pr\{K\alpha_i \leq (b_i - E(b_i)/\sigma_{bi}\} = 1 - \alpha_i$$

Comparando las expresiones última y antepenúltima se deduce que

$$K_{\alpha i} \geq (\sum a_{ij} X_j - E(b_i))/\sigma_{bi}$$

con lo que la restricción probabilística queda

$$\sum a_{ij} X_j \leq E(b_i) + K_{\alpha i}\sigma_{bi}$$

que es una ecuación lineal.

Charnes et al. (1979) utilizaron este modelo para evaluar la respuesta a los desastres debidos a la contaminación marina, utilizando programación por metas.

Las principales ventajas de la programación con restricciones aleatorias son:

1) Necesita mucha menos información que la programación estocástica discreta.

2) Es muy fácil de integrar en cualquier formulación de programación lineal.

3) Pueden considerarse distintos comportamientos del agricultor frente al riesgo teniendo en cuenta distintos valores de α_i.

A pesar de las ventajas de las restricciones aleatorias, existen también algunos inconvenientes (*Hazell y Norton*, 1986):

1) Se supone que las variables aleatorias siguen una ley de distribución normal, aunque existe algún método basado en técnicas no paramétricas (*Millán*, 1995), que permite aplicar la programación con restricciones aleatorias a casos en los que no se puede determinar su distribución.

2) Si el modelo incorpora más de una restricción aleatoria, entonces es necesario suponer que los coeficientes estocásticos de cada restricción son estadísticamente independientes de dichos coeficientes en todas las demás restricciones.

3) Cuando dos o más restricciones aleatorias sean estadísticamente independientes, habrá problemas debido a que la probabilidad de que todas las restricciones sean factibles al mismo tiempo será el producto de sus probabilidades individuales.

4) Dado que debe ser seleccionado un valor de α para cada restricción aleatoria, dichos valores pueden ser inconsistentes con el tipo y grado de aversión al riesgo asumido en la función objetivo del modelo.

5) La programación con restricciones aleatorias no proporciona ayuda respecto a cómo debería actuar un agricultor en el α por ciento de los años en los cuales la planificación sea infactible.

6) En aquellos estados de la naturaleza favorables en relación con la actitud del centro decisor se producirá una infrautilización de los recursos que puede cuestionar la conducta maximizadora de dicho centro decisor.

#1.6.4.3.- PROGRAMACIÓN EN DOS ETAPAS BAJO INCERTIDUMBRE

Dantzig (1955) desarrolló el método de programación bajo incertidumbre con dos o más etapas. En el caso de un modelo de transporte se divide el coste en dos partes, una referente al coste de transporte y otra referente al coste en que se incurre cuando no se satisfacen las demandas desconocidas. La estructura de este tipo de problemas es, para la primera etapa:

$$\sum_{j=1}^{n} x_{ij} = a_i$$

$$\sum_{i=1}^{m} b_{ij} x_{ij} = u_j$$

donde x_{ij} representa la cantidad del recurso *i* asignado al destino *j* y b_{ij} representa el número de unidades de demanda en el destino *j* que pueden ser satisfechas por unidad de recurso *i*. Para la segunda etapa

$d_j=u_j+v_j-s_j \ (j=1, 2,\ldots, n)$

donde v_j es el déficit de oferta y s_j es el exceso de oferta.

La función de costes totales tiene la forma:

$$C = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} + \sum_{j=1}^{n} \alpha_j v_j$$

y depende de la elección x_{ij} y de los déficits v_j (que dependen de las asignaciones u_j y de las demandas d_j). α es el coeficiente de proporcionalidad.

Este modelo puede ser generalizado a varias etapas (ver por ejemplo *Kall y Wallace*, 1994).

Beale (1955) desarrolló la programación lineal con coeficientes aleatorios en las restricciones, llegando a la introducción de unos ponderadores en la función objetivo tal y como hacen *Kall y Wallace* (1994). *Madansky* (1962) propuso, para reducir los efectos de la incertidumbre, replantear el problema como si tuviera dos etapas (compensando en la segunda etapa las inexactitudes de las actividades correspondientes a la primera etapa, obteniendo la llamada solución ‘slack’), o bien sustituir los elementos aleatorios por sus valores esperados (solución del valor esperado), o bien sustituir los elementos aleatorios por las estimaciones pesimistas de sus valores (solución ‘fat’).

CAPÍTULO 2

PROBABILIDADES SUBJETIVAS Y RENDIMIENTOS DE CULTIVOS

#2.1.- FORMACIÓN DE LAS DISTINTAS INTERPRETACIONES DE LA PROBABILIDAD

La probabilidad es un concepto teórico, algo que no tiene un referente real, o como señaló de Finetti, que no existe. Por ello los modelos probabilísticos se expresan mediante distintas interpretaciones de la probabilidad.

Las interpretaciones de la probabilidad han dado origen a dos grandes corrientes en la Estadística (inferencias y decisiones), conocidas como clásica y bayesiana. La primera se basa en una interpretación que pretende ser objetiva, y estar determinada por el universo de los sucesos, independiente del sujeto que realiza un juicio sobre la misma. La estadística bayesiana se basa en una concepción subjetivista o personalista de la probabilidad, en las creencias del sujeto, y recibe esta denominación por el papel central que juega en la misma el teorema de Bayes, publicado hacia 1763.

Históricamente, y hasta bien avanzado el siglo XIX, coexistieron los enfoques subjetivistas y objetivistas de la probabilidad, aunque, como es conocido, el segundo predominó posteriormente sobre el primero, porque se consideró que la probabilidad objetiva tenía una mejor fundamentación (objetiva y experimental) que la subjetiva.

La evolución de la Estadística, como la de otras disciplinas, se puede explicar en función de los problemas formulados internamente y de los que se le forzó a considerar desde el exterior. Inicialmente las mayores demandas de soluciones se derivaron de otras disciplinas, y así en el siglo XVII fueron los problemas actuariales los más importantes, en el XVIII las cuestiones relacionadas con mediciones inciertas en astronomía, y desde los últimos años del siglo XIX los análisis de datos biológicos. Estas condiciones predominantes en el nacimiento del concepto de probabilidad determinaron el espacio de teorías posibles (Hacking, 1993).

El nacimiento del concepto de probabilidad se suele situar en torno a mediados del siglo XVII, asociada a los nombres de Huygens, Pascal y su famoso argumento sobre la conveniencia de creer en la existencia de Dios, Leibniz y su trabajo sobre combinaciones, J. Hudde y J. de Witt y los problemas actuariales relacionados con las finanzas públicas de las ciudades holandesas, y J. Graunt y las inferencias estadísticas a partir de los registros de mortandad. En estos trabajos iniciales la probabilidad se discutió en el contexto de las leyes que gobiernan los procesos aleatorios, y los aspectos epistemológicos relacionados con el grado de creencia en una proposición.

La interpretación clásica frecuentista de la probabilidad fue formulada por Laplace (1795, 1951, 1985), y es la más antigua y simple. Se basa en una interpretación objetiva y lógica, aplicada a juegos de azar y a otras situaciones claramente especificadas, que se pueden descomponer en un número de casos igualmente probables. Por ejemplo, en los juegos de cartas se supone que cada carta tiene la misma probabilidad de ser extraída; en los juegos de dados, se supone que cada cara tiene la misma probabilidad de salir, etc. Se supone que cada proposición tiene asociada un conjunto finito de posibilidades. Dada una proposición p y las posibilidades asociadas con ella, aquellas que la verifican se llaman p-casos. La probabilidad de p se define como la razón entre el número de p-casos y el número total de casos o posibilidades. La probabilidad de un suceso es la razón entre el número de casos favorables del mismo y todos los casos posibles.

Como las probabilidades son en última instancia razones entre posibilidades abstractas, la interpretación clásica permite proposiciones cuyas implicaciones no se refieren a sucesos reales y por tanto no se pueden contrastar. Para poder definir la probabilidad en términos de sucesos reales (experimentales) Venn desarrolló la aproximación frecuentista, que pretende ser objetivista y empírica.

En esta interpretación, la probabilidad se define como el límite de la frecuencia relativa de un suceso conforme aumenta el número de casos producidos (siendo el suceso un resultado de varios posibles).

En situaciones con un número reducido de manifestaciones, o en casos singulares o no observados, se ha utilizado el concepto de "propensión" (p.e. se puede hablar de que una moneda tiene una propensión a salir cara de ½). Este enfoque, sin embargo es limitado. Por ejemplo, los médicos saben que un fumador tiene mayor propensión que un no fumador a sufrir enfermedades coronarias. Pero para un sujeto concreto, fumador, es difícil establecer si tiene una mayor propensión que otro sujeto concreto no fumador.

Estas dos interpretaciones se suelen denominar como lógica o empírica respectivamente, según se defina la probabilidad en términos de estructuras lógicas o matemáticas, o en términos de propiedades empíricamente definidas.

La interpretación lógica no es adecuada en las situaciones en que no se dispone de recursos para analizar los supuestos. La interpretación frecuentista es inadecuada en situaciones con pocos casos, casos singulares o casos inobservados, cuando no se pueden asignar propensiones.

La estadística clásica, basada en la probabilidad objetiva, se ha desarrollado en una amplia gama de intereses, entre ellos los problemas de inferencia. En este ámbito, que es especialmente interesante para la fundamentación de las decisiones, coexistieron dentro de la estadística clásica concepciones diferentes, que se pueden describir abreviadamente en dos líneas, la primera ligada a los trabajos de R. A. Fisher y la segunda a los de J. Neyman y E. Pearson.

Con anterioridad a los trabajos de Neyman, la inferencia estadística clásica consistía en un conjunto de métodos para justificar la aceptación de proposiciones a la luz de la evidencia aportada por los datos. Los enunciados típicos estaban relacionados con la estimación puntual (la mejor estimación de un parámetro desconocido μ es m), la estimación por intervalos (μ se encuentra entre m_1 y m_2), y la aceptación de hipótesis ($\mu \neq 0$), etc. Neyman reclamó la importancia de la acción, punto de vista reclamado con fuerza, desde una perspectiva decisional, por Wald (1950) desde los años 40, y rechazada por Fisher (1956). La discusión se centró, precisamente, en torno a si los problemas económicos son un buen marco para examinar los problemas de inferencia estadística.

La orientación bayesiana de la Estadística consiste en un conjunto de técnicas para expresar una opinión, comprobar su consistencia interna, y revisarla a la luz de la información adicional obtenida. En principio este enfoque no se apartaría de las formulaciones derivadas de la Teoría de la Decisión y de la moderna Teoría de la Inferencia, en la estadística clásica, excepto en la consideración de la noción de probabilidad subjetiva o personal.

La Estadística Bayesiana se ha fundamentado recientemente sobre dos grandes corrientes: la concepción subjetivista de la probabilidad y la Teoría de la Decisión. Estas dos líneas no pueden considerarse como paralelas, sino que a lo largo de su desarrollo se han influido mutuamente.

La interpretación subjetiva de la probabilidad se vio favorecida por los cambios experimentados a finales del siglo XIX en la concepción metodológica de las ciencias formales (que liberó a los sistemas axiomáticos de una correspondencia "evidente" con la "realidad"). En efecto, la interpretación objetiva de la probabilidad considera a ésta como la medida de algo independiente de los juicios de los sujetos, y contrastable interpersonalmente, por lo que la primera ruptura con esa concepción sólo se podía dar en el seno de la evolución de la concepción de los sistemas formales. Así, el desarrollo de la interpretación de la probabilidad subjetiva se debe a Borel (1924, 1964), B. de Finetti (1937, 1964), Ramsey (1926,1980), B. O. Koopman (1940), Good (1950, 1960), Grayson (1960), Anscombe (1961), Lindley (1961), Pratt (1961), Raiffa y Schlaifer (1961), Savage et al. (1962), etc.

Existe una relación, muy conocida en la Sicología, entre creencia, deseo y acción (una persona que cree que el agua ha sido envenenada, se resistirá a beberla aunque esté sediento; si una persona cruza una calle, podemos inferir razonablemente que desea ir a la otra acera y que considera seguro hacerlo). F. Ramsey construyó una teoría subjetiva de la probabilidad, donde los grados de creencia (o confianza) en las proposiciones están relacionados con la certeza de las acciones (lo mejor que se puede hacer). Por ejemplo, una persona que cree que un caballo ganará una carrera, estará dispuesto a aceptar apuestas con premios desequilibrados (40 u.m. si gana, - 60 u.m. si pierde): cuanto más seguro esté de que el caballo ganará, aceptará apuestas más desequilibradas. Este autor desarrolló una concepción en la que la probabilidad subjetiva y la utilidad se expresaban conjuntamente.

En la fundamentación de la Teoría de la Decisión, construida en torno al núcleo de la Teoría de Juegos, se suelen citar a Borel (1921), von Neumann (1928), y von Neumann y Morgenstern (1944) y las construcciones en Estadística, inicialmente realizadas por Neyman, Wald (1950), etc. Estos enfoques y el trabajo de Ramsey (1926), sirvieron de base al trabajo de Savage (1954) sobre fundamentación de la Estadística. Una recopilación bibliográfica de las primeras aportaciones se puede encontrar en Savage (1954) y Savage et al. (1962).

B. de Finetti (1964) señaló que no es necesario suponer que la probabilidad de un suceso tome valor determinado, sino que la probabilidad expresa las opiniones de cada sujeto, y por tanto tiene sentido para el sujeto que expresa la probabilidad. Esta interpretación es la más reciente fundamentación de la probabilidad subjetiva o personalista.

La interpretación subjetivista, aunque necesaria en un gran número de situaciones decisionales reales, no fue admitida de forma significativa en los círculos estadísticos hasta bien avanzada la década de los años 80, cuando culminaron y se divulgaron los extraordinarios avances conseguidos en la fundamentación de la probabilidad subjetiva, pero también conforme se comprendieron las enormes dificultades de la estadística clásica, basadas en diferentes interpretaciones de la probabilidad objetiva, especialmente en el ámbito de la inferencia. Las dificultades de la estadística clásica y los avances de la bayesiana habrían contribuido a oscurecer la frontera entre ambas aproximaciones, regresándose a una situación que había predominado en la concepción de la probabilidad hasta principios del siglo XX.

En los textos de Berger (1985), Robert (1992, 1997) o Bernardo y Smith (1997), por ejemplo, se pueden consultar los resultados obtenidos por la estadística bayesiana, su evolución histórica y un resumen de las discusiones sobre su fundamentación.

El desarrollo de las aplicaciones del Teorema de Bayes (aunque sin aceptar la concepción personalista de la probabilidad) fue realizado por Jeffreys (1931, 1939).

En el ámbito de la aplicación, el primer texto que se suele citar es el de R. Schlaifer (1959), **Probability and Statistics for Business Decisions**. Era un manual que presentaba una aproximación práctica de las ideas bayesianas, sobre la base que la probabilidad es una opinión fundada que se puede revisar a partir de la evidencia adicional disponible utilizando el teorema de Bayes. Un segundo texto, H. Raiffa y R. Schlaifer (1961), amplió estas ideas en el ámbito de la decisión empresarial, y el texto de Lindley (1965) consolidó esta orientación. Manuales como el de B. W. Morgan (1971), publicado inicialmente en lengua inglesa en 1964, contribuyeron a mantener vivo el interés en estas técnicas en distintas aplicaciones. El heredero directo de estos esfuerzos es el magnifico texto de Pratt, J.W., H. Raiffa, R. Schlaifer (1965, 1995).

#2.2.- TEORIA DE LA DECISIÓN Y UTILIDAD

La Teoría de la Decisión es un conjunto de teorías filosóficas, lógicas y matemáticas que tratan de explicar los mecanismos utilizados por los sujetos racionales a la hora de tomar sus decisiones. Trata tanto aspectos de toma de decisiones individuales como en grupos, en juegos contra la naturaleza o en competición. Entre otras, incluye las teorías sobre la utilidad, la Teoría de Juegos y la Teoría de la Elección Social.

La Teoría de la Decisión ha desarrollado hipótesis sobre la racionalidad de los decisores y de las decisiones, y sobre el comportamiento humano, en lugar de estudiar los aspectos de formulación o de verificación de hipótesis, que han sido más propios de las teorías estadísticas. En la Estadística, la Teoría de la Decisión se emplea tanto para elegir entre hipótesis alternativas como para determinar la acción óptima en función de los resultados de un experimento.

Una importante contribución en el seno de la teoría la constituyen los trabajos matemáticos y lógicos que investigan las consecuencias de las diferentes reglas de decisión, y las características matemáticas de las diferentes descripciones del comportamiento de los agentes idealmente racionales, capaces de adquirir, almacenar y procesar cantidades ilimitadas de información, de aplicar correctamente las reglas lógicas y matemáticas adecuadas y conocer las consecuencias lógicas de sus creencias.

Un segundo grupo de contribuciones tiene que ver con las investigaciones acerca del comportamiento real de los sujetos a la hora de tomar las decisiones.

La elección humana ha sido atribuida en distintas teorías a aspectos como el instinto, el inconsciente, el altruismo, el reforzamiento, la pasión ciega, el deber, la maldad o los imperativos morales. La Teoría de la Decisión postula que la elección se realiza por mediación de la razón, entendida como algo propio de la acción humana, que constituye una explicación general del comportamiento.

La racionalidad, o lo que es razonable, está expuesta a interpretaciones distintas, en función del contexto disciplinar (una persona racional no se deja arrastrar por su temperamento o por la emoción, o comprende sus motivaciones, o adecua su razonamiento a las reglas formales de inferencia), pero en un contexto económico, el hombre racional es aquel cuya decisión (racional u óptima) es la más favorable, de varias posibles, en el sentido de producirle los mayores beneficios. La decisión optima depende de la información disponible y del criterio adoptado para juzgarla, y puede tomar valores distintos para diferentes sujetos.

El problema típico de la teoría de la decisión especifica un conjunto de actos, a_i, de estados de la naturaleza θ_j y de resultados r_{ij}, estos últimos obtenidos como consecuencia de adoptar la decisión a_i cuando se manifiesta el estado de la naturaleza θ_j. Tanto los actos como los estados de la naturaleza se supone que son mutuamente excluyentes y que agotan todas las posibilidades. Dado el conjunto de resultados r_{ij}, y una regla o criterio que permita ordenarlos, aquel que resulte preferido sobre todos los demás determina la acción a_i* que se considera óptima, de acuerdo con esa regla o criterio.

Esta forma de identificar una acción optima se conoce como “analytical approach” (Raiffa, 1968), y se especifica en un conjunto de pasos: (1) se identifican todas las posibles decisiones o acciones y sucesos o estados de la naturaleza; (2) se calculan los resultados de cada combinación de decisiones y sucesos; (3) se asigna un valor, medido en utilidades, de logro a los resultados; (4) se establece la probabilidad de que ocurra cada resultado; y (5) se identifican los resultados no dominados óptimos que maximizan la esperanza matemática de las utilidades.

En general, se suelen distinguir tres contextos de decisión en la teoría. En el primero, llamado contexto de certidumbre, se supone que se conoce que cada acción (decisión) a_i produce un único resultado r_i. En el segundo, llamado contexto de riesgo, se supone que cada acción a_i puede producir n resultados distintos r_{ij}, cada uno de ellos asociado con un estado de la naturaleza θ_j, y una probabilidad $P(\theta_j)$, con j = 1, 2,.., n. En el tercero, llamado contexto de incertidumbre, la situación es similar a la de riesgo, excepto que se desconoce la probabilidad $P(\theta_j)$ de que se manifieste cada estado θ_j. Mayores niveles de incertidumbre estarían asociados con fallos en el conocimiento de todas las acciones posibles o de los estados de la naturaleza que se pueden presentar.

Los resultados r_{ij} son una medida del logro obtenido como consecuencia de adoptar la decisión a_i si posteriormente se presenta el estado de la naturaleza θ_j. Medir significa asignar una cantidad a un objeto o suceso, y la medida expresa el grado en que un atributo se encuentra en lo medido. La medición se puede realizar con diferentes escalas, cada una relacionada con el grado de arbitrariedad en los números o cantidades asignados al realizarla. El tipo de escala depende del tipo de operaciones realizadas para efectuar la medida, y así algo como la estatura de un grupo de personas, que se mide utilizando una regla, se expresa en una escala razón, pero si se ordenan se necesita solamente una escala ordinal para ello. En general no existe una relación biunívoca entre un atributo y el tipo de escala utilizada para su medida.

Los métodos de medida se clasifican de acuerdo con el grado de arbitrariedad en las medidas. El grado de arbitrariedad lo determinan las transformaciones matemáticas que permite obtener un conjunto de medidas. Se suelen distinguir cuatro escalas básicas: absoluta (ningún grado de arbitrariedad), razón o ratio (arbitrariedad reducida a multiplicar por un número positivo), intervalo (transformación lineal positiva) y ordinal (cualquier transformación monótona positiva).

Es importante entender que unas escalas son casos particulares de otras. Por ejemplo, la escala intervalo es un caso especial de la escala ordinal: una clase particular de transformación monótonamente creciente, la lineal. De igual forma, la escala razón es un caso particular de la escala intervalo. Una escala ordinal se puede transformar de muchas maneras en una escala intervalo, y esta última se puede transformar de muchas formas en una escala razón.

Las reglas o criterios de ordenación de los resultados pueden exigir una u otra escala de medida de los resultados. Así, la regla min – max o criterio de Wald necesita que los resultados r_{ij} estén expresados, como mínimo, en una escala ordinal, de tal forma que basta con que el decisor sepa que prefiere un resultado a otro para poder ordenarlos.

Sin embargo una regla como la de Savage, que sigue utilizando el principio min – max de ordenación, pero sobre una matriz derivada de la inicial, llamada de arrepentimiento, exige una medida que actúe sobre valores obtenidos mediante una transformación lineal positiva de los originales, es decir, una escala intervalo. El decisor necesita mucha más información sobre sus preferencias para aplicar esta regla de decisión.

El criterio de Hurwicz actúa ordenando combinaciones lineales del peor y el mejor resultado que se puede obtener para en cada acción a_i. De esta forma cada acción a_i está asociada con una media de resultado $r_i = \alpha \min (r_{ij}) + (1-\alpha) \max (r_{ij})\ \forall\ j,\ \alpha \in [0,1]$. Una transformación de este tipo supone también una escala intervalo de utilidad.

El criterio de Laplace se corresponde con el principio de razón insuficiente de la Estadística Bayesiana clásica, y puede ordenar los resultados mediante una escala intervalo de utilidad.

Si imaginamos una escala que mida la información disponible, el criterio de Laplace o principio de razón insuficiente, establecería una frontera entre la Teoría de Juegos, a su izquierda, donde la falta de información no permite asignar probabilidades a cada estado de la naturaleza θ_j, y la Teoría de la Decisión Bayesiana, a su derecha, donde el decisor, que no conoce la probabilidad objetiva de cada estado de la naturaleza $P(\theta_j)$, sí es capaz de realizar un juicio y aplicar probabilidades subjetivas.

Conforme se adquiera información adicional, el decisor conocerá la probabilidad de cada estado de la naturaleza, $P(\theta_j)$, y diremos que dispone de información completa (o que su información es 1, en una escala de 0 a 1) en el contexto de riesgo. Pero como consecuencia de cada acción pueden resultar n posibles resultados, cada uno asociado con la realización de cada estado de la naturaleza, y diremos que dispone de información completa pero que le falta influencia en este contexto (de riesgo).

El contexto de certidumbre se caracterizaría porque el decisor dispone de información completa (como en el de riesgo) y de influencia completa (cada acción está asociada a un único resultado).

En esta clasificación coexisten las dos interpretaciones de la probabilidad, objetiva y subjetiva. Más aún, la probabilidad objetiva puede interpretarse como un caso límite de la subjetiva, aquel en el que toda la información está disponible.

Cualquiera que sea la concepción de la probabilidad, la elección en contexto de riesgo y de incertidumbre (excepto con el criterio de Wald) implicará la necesidad de escalas intervalo de utilidad (pero no escalas superiores). El decisor no solamente debe ordenar los resultados, sino que es necesario mantener las distancias relativas o intervalos de preferencia entre ellos. Al utilizar escalas intervalo para representar las preferencias, no se puede suponer automáticamente que las operaciones aritméticas que se realizan habitualmente en las escalas razón (para velocidades, pesos o distancias) tengan sentido para las utilidades.

En general no existirá una biyección entre unidades monetarias y utilidades. Sin embargo a veces la escala monetaria se comporta como una escala intervalo, al menos en un determinado rango o intervalo: cuando el decisor se muestra indiferente entre incrementos (o reducciones) iguales de precios dentro de un determinado rango. En ese rango tiene sentido tomar decisiones en contexto de riesgo con escalas monetarias y el criterio de valores esperados monetarios (EMV). Sin embargo, en la mayoría de los casos esa indiferencia no se verificará, y la decisión debe tomarse con el criterio de esperanza matemática de la utilidad (EMU).

Von Neumann y Morgenstern (1944) desarrollaron una aproximación a la utilidad que evitaba las objeciones planteadas al utilizar de forma general el criterio EMV, y que además presentaba la ventaja sobre la aportación anterior de Ramsey de separar la probabilidad y la utilidad. Esa aproximación se conoce como N-M, y se basa en la idea de medir la fuerza de la preferencia de una persona por una cosa por el riesgo que está dispuesta a asumir para conseguirla.

Es importante entender que una escala de utilidad es únicamente una representación numérica de las preferencias de un agente. Por ello se debe tener cuidado a la hora de proyectar sobre las propiedades de un agente las propiedades numéricas de la escala de utilidad. Es importante entender que los agentes no tienen preferencias como consecuencia de las propiedades de su escala de utilidad, sino que sus escalas de utilidad tienen determinadas propiedades porque reflejan las preferencias del agente.

El teorema de la utilidad esperada en el sistema N-M es un “teorema de representación”, es decir, que muestra que una estructura no numérica puede ser representada numéricamente. Concretamente dice que si las preferencias de un agente tienen una estructura suficientemente rica, la estructura puede representarse numéricamente por medio de funciones de utilidad intervalo que poseen la propiedad de utilidad esperada. Para ello se asignan números a cada premio básico y a las diferentes loterías y se comprueba que la escala numérica resultante posee las propiedades deseadas.

Sin embargo, si las preferencias del decisor no verifican algunas de las condiciones del teorema, entonces no se puede asegurar que la construcción tenga las propiedades deseadas (p.e. la propiedad de utilidad esperada). El teorema parte de las preferencias del agente y reformula esa información en términos numéricos. Para representar la regla de maximización de la utilidad esperada se necesita una escala intervalo de utilidad. Las escalas monetarias son insatisfactorias porque frecuentemente no son compatibles con las verdaderas preferencias y porque fallan en situaciones de una sola decisión.

Las escalas de utilidad asignan “verdaderos valores” en el sentido de que las utilidades se mantienen en una relación uno a uno con las preferencias del decisor. Cuando un decisor aplica la regla de maximización de la utilidad esperada simplemente elige el acto cuya utilidad es más alta. Si un acto es preferido a otro, la utilidad del primero es superior a la del segundo. Cuando el decisor elige un acto, elige su opción preferida. Esto es verdad incluso en el caso de una decisión única o no repetida.

Lo anterior justifica la utilización de utilidades esperadas en las decisiones en condiciones de riesgo, sobre todo en caso de una sola decisión. Eligiendo un acto que tiene la mayor utilidad esperada un decisor solamente dice que es eso lo que desea elegir. En realidad, un decisor que tuviera una estructura muy definida de preferencias, podría no necesitar una teoría de la utilidad.

#2.3.- MARCO CONCEPTUAL DE LAS PROBABILIDADES SUBJETIVAS

En la Teoría de la Decisión (de Finetti, 1937, 1967, 1974; Savage, 1954) se considera a las probabilidades subjetivas como una opinión cuantificada realizada por una persona. Más formalmente, una probabilidad subjetiva es un grado de creencia en la verdad de una proposición. Estos autores desarrollaron una teoría axiomática donde se demuestra que las probabilidades subjetivas se corresponden con una medida de probabilidad.

La probabilidad que una persona otorga a un suceso o hipótesis particular se dice que mide su "grado de creencia" en la misma. Si esos "grados de creencia" se asignan cuantitativamente de una forma coherente, el sujeto realiza una asignación de probabilidades subjetivas. Es importante entender que en esta interpretación de la probabilidad no se trata de establecer las opiniones que las personas deberían tener en una circunstancia concreta, sino de cómo se deben expresar las opiniones y cómo modificarlas en función de la (nueva) información adicional disponible. Como señaló de Finetti: "the true... subjective probability... problem consists in the investigations concerning the ways in which probabilities are assesed by more or less educated people, and the way in which such abilities may be improved." (citado en Hogarth, 1975).

Generalmente, se considera que un conjunto coherente o consistente de probabilidades subjetivas es aquel que no permite formular una serie de apuestas contra el decisor de tal forma que éste pierda cualquiera que sea el resultado. Un conjunto de probabilidades subjetivas no coherente se conoce como "Dutch Book".

Las condiciones necesarias y suficientes para asegurar la coherencia o consistencia son las siguientes:

$$0 \leq P(A) \leq P(S) = 1$$

$$P(A \cup B) = P(A) + P(B)$$

donde S es el conjunto universal, y A y B son dos sucesos incompatibles y disjuntos (excluyentes).

En la vida real las personas a veces realizan elecciones que violan esas reglas, especialmente la segunda. Por ello se insiste en la concepción personalista de la probabilidad en que la probabilidad es una opinión consistente, y no solamente una opinión.

Winkler y Murphy (1968) señalaron dos tipos de "bondad" en la evaluación de las probabilidades subjetivas realizadas por un sujeto:

- la bondad normativa, que expresa el grado en el que la estimación de un sujeto expresa las verdaderas creencias del mismo y se ajusta a los axiomas de la teoría de la probabilidad, y

- la bondad sustantiva, que refleja la cantidad de conocimiento sobre el área de conocimiento implicada en la estimación.

Winkler (1967 b) y A. Tversky y D. Kahneman (1984,1) señalaron que el criterio de coherencia o de consistencia interna no es suficiente para evaluar la bondad de un juicio sobre un conjunto de probabilidades subjetivas, ya que puede ser incompatible con otras creencias sostenidas por el individuo. Un decisor racional debería añadir a los criterios de consistencia interna otros como la compatibilidad con su conocimiento sobre la materia tratada, las leyes de la probabilidad y sus propios heurísticos y sesgos. Así, por ejemplo, una persona que incurra en la falacia del jugador (cree que después de aparecer cuatro caras seguidas en el lanzamiento de una moneda es más probable que, en la siguiente tirada, aparezca una cruz) podría establecer un conjunto de probabilidades internamente consistente con esa creencia, pero su asignación de probabilidades sería manifiestamente errónea.

Si se identifican las probabilidades subjetivas de un sujeto con las apuestas que está dispuesto a aceptar, se pueden medir sus probabilidades subjetivas mediante una serie de apuestas sobre la verdad de un conjunto inicial de proposiciones, sus negaciones, sus conjunciones, sus disfunciones, y sus apuestas condicionales. Todas esas proposiciones forman lo que se conoce como el cierre de de Finetti ("de Finetti closure") del conjunto inicial de proposiciones. El cierre de de Finetti del conjunto inicial A, DC(A) puede formalmente definirse como sigue:

1. Cada proposición en A lo es también en DC(A)
2. Si las proposiciones S y W están en DC(A), también lo están las proposiciones no S, ($S \cup W$) y ($S \cap W$).
3. Si S y W están en DC(A), también está la frase S dado W.
4. No existen otros elementos en DC(A) que no sean los descritos en 1, 2 y 3.

El agente debe asignar apuestas, de la forma "a contra b" (odds), a cada uno de los elementos de DC(A). Una vez realizadas esas asignaciones, un agente organizador de la apuesta ("booki") determina el resto de las características del juego, incluyendo las cantidades monetarias asignadas a cada apuesta a favor y en contra.

Así, por ejemplo, sea una única proposición, p. Se aceptan apuestas a favor y en contra de p, de tal forma que todo lo que gana (pierde) la persona que apuesta a favor lo pierde (gana) la persona que apuesta en contra. Si S es la cantidad puesta en juego, y a es un número, denominado "cociente de la apuesta", la razón a / (1-a) es la apuesta, expresada de la forma a contra (1-a) u "odds", que se hace a favor de la verdad de la proposición p.

La proposición p solo puede tomar dos valores, verdadero o falso. Si se apuesta por p y p es falso, el pago obtenido es -aS (el organizador "booki", que apuesta contra p, gana aS). Si p es verdadera, el jugador gana (1-a) S y el organizador ("booki") pierde - (1-a) S. De esta forma se tiene la matriz de pagos:

p	**Pago por p**	**Pago contra p**
V	(1-a)S	- (1-a)S
F	-aS	aS

Si se apuesta a contra (1-a) a favor de la proposición p, es obvio que se está dispuesto a perder una fracción a de la cantidad S puesta en juego si la proposición p es falsa. Obviamente, cuanto mayor es la confianza en la verdad de p, mayor será el valor a que se estará dispuesto a apostar contra (1-a).

En la tabla anterior el sujeto que apuesta sólo puede determinar el valor de "a". El organizador de la apuesta ("booki") fija el valor puesto en juego S, y quién apuesta a favor de p y quién en contra. Sin embargo, no puede apostar simultáneamente por p y contra p (tampoco los jugadores).

Para aclarar el concepto de apuestas "a" contra "1-a" (odds) veamos el siguiente ejemplo. Suponga que el Círculo Ecuestre admite apuestas de 99 a 1 a que el caballo "Solsona" ganará la carrera. Esto significa que nos pagará 100 $ por cada $ 1 que hayamos comprado dando a "Solsona" como ganador. La cantidad puesta en juego por cada ticket de 1$ es 100 $. Al comprarlo, arriesgamos 1/100 del valor del ticket, y el organizador de la carrera arriesga 99/100.

En una situación más general, si alguien acepta apuestas de a contra b, acepta un riesgo de perder dado por a / (a+b) de la cantidad puesta en juego en caso de no acertar. En general, las apuestas son a / (a+b) contra b / (a+b) y el cociente de apuesta es a / (a+b).

Aunque la probabilidad subjetiva se determina a partir de preferencias entre apuestas, normalmente no se forma así sicológicamente. En realidad, la probabilidad subjetiva determina las preferencias entre las apuestas y no se derivan de ellas, como establece la teoría axiomática (Savage, 1954). Una persona apuesta por el equipo A en lugar de hacerlo por el equipo B porque cree que es más probable que gane el equipo A. De esta forma no deriva sus creencias de sus preferencias entre apuestas.

#2.4.- REVISIÓN DE LAS PROBABILIDADES SUBJETIVAS Y TEOREMA DE BAYES

La revisión de las probabilidades (subjetivas) a priori mediante la consideración de la información (muestral) adicional se realiza normalmente utilizando el teorema de Bayes. La aplicación del teorema permite obtener una probabilidad revisada (posterior) a partir de la probabilidad subjetiva (anterior) y la evidencia disponible. De esta forma, aunque la teoría personalista de la probabilidad acepta que dos sujetos puedan tener inicialmente opiniones distintas sobre un suceso (y por tanto asignarán probabilidades a priori distintas al mismo), mantiene que la corrección de ambas opiniones con la información adicional hará que las probabilidades posteriores lleguen a ser indistinguibles.

La probabilidad condicional $P(D/H)$ de un suceso D dado otro H se puede interpretar como la cantidad de dinero que se estaría dispuesto a pagar si se recibiera un premio de 1 u.m. cuando D sea verdad, con la condición de que todas las apuestas se cancelarán a menos que se verifique que H es verdad. Es fácil comprobar que $P(D \cap H) = P(D/H)\ P(H)$, donde $D \cap H$ es el suceso "D y H" y son ambas verdaderas. De esta forma se tiene:

$$P(D/H) = P(D \cap H) : P(H) \quad \text{excepto que } P(H) = 0$$

y la probabilidad condicional expresaría lo que se ha aprendido mediante la experiencia. La probabilidad de D, una vez que se sabe (se ha aprendido) que H es verdadera, es precisamente $P(D/H)$. De esta forma, una vez que se ha aprendido que H es verdadera, el nuevo sistema de probabilidades $P(D/H)$, para unas condiciones específicas de H, juega el mismo papel que $P(D)$ en un sistema incondicional.

El teorema de Bayes establece la relación que existe entre la probabilidad condicional $P(H_i/D)$ y las probabilidades $P(H_i)$ y $P(D/H_i)$, donde H_i puede interpretarse como las hipótesis cuyas probabilidades se desean determinar, una vez observados un conjunto de datos D. Suponiendo que se desea establecer la probabilidad de un suceso o hipótesis H_i, siendo las hipótesis posibles exhaustivas y mutuamente excluyentes, $H = H_1 \cup H_2 \cup \ldots \cup H_i \cup \ldots \cup H_n$, con $H_i \cap H_j = \varnothing\ \forall i \neq j$, el teorema de Bayes establece que:

$$[1] \qquad P(H_i/D) = \frac{P(H_i)\ P(D/H_i)}{\sum_{i=1}^{n} P(H_i)\ P(D/H_i)}$$

$P(H_i/D)$ es la probabilidad posterior de la hipótesis H_i después de observar los datos D. El valor $P(H_i)$ es la probabilidad a priori o anterior, es decir, la probabilidad que se atribuye a la hipótesis H_i antes de observar los datos D. La expresión $P(D/H_i)$ es la probabilidad de observar los datos D cuando la hipótesis H_i es verdadera. El teorema de Bayes, por tanto, permite calcular la probabilidad posterior de un suceso o hipótesis H_i, $P(H_i/D)$, partiendo de información previa o probabilidad anterior, $P(H_i)$, corregida con nuevos datos o información adicional de muestra, D.

El denominador de la expresión [1] es la probabilidad incondicional de los datos, $P(D)$, y puede ser considerada una constante para normalizar, para hacer que las probabilidades posteriores sumen uno cuando se consideran varias hipótesis.

$$[2] \qquad P(D) = \sum_{i}^{n} P(H_i)\ P(D/H_i)$$

En principio, la probabilidad anterior o "a priori", $P(H_i)$ puede ser una probabilidad objetiva, pero esto no es necesario. También se puede expresar como una impresión personal o subjetiva, en forma de probabilidad subjetiva. Algo semejante ocurre con $P(D/H_i)$, que se puede deducir de probabilidades objetivas o mediante la expresión de asociaciones intuitivas derivadas de procesos de aprendizaje o de la experiencia.

La elección de las particiones H_i tiene una gran importancia práctica, pero sin embargo es algo arbitrario. Por ejemplo, podemos decir que el tiempo que hará mañana será “bueno” o “malo”, pero como es obvio, estas hipótesis pueden refinarse todavía mucho más. El teorema de Bayes es verdadero para todas las particiones, pero es más útil en unas particiones que en otras. Conforme el conocimiento avanza, algunas particiones antes insospechadas resultan de una importancia decisiva.

En muchas aplicaciones existe un amplio consenso sobre el valor de $P(D/H_i)$, que puede considerarse público, ya que se deriva de algún modelo analítico que los científicos aceptan en su ámbito de actuación, de donde provienen los datos.

La probabilidad P(D) en principio tiene un interés limitado, y frecuentemente se elimina de los cálculos. Una forma de hacerlo es expresar el teorema de Bayes en forma de apuestas (“odds”). Suponiendo dos hipótesis mutuamente excluyentes H_A y H_B, sometidas a una evidencia común, se puede escribir:

[3] $$\frac{P(H_A / D)}{P(H_B / D)} = \frac{P(D / H_A)\,P(H_A)}{P(D / H_B)\,P(H_B)}$$

o $\Omega_1 = L\ \Omega_0$

donde Ω_1 son las apuestas (“odds”) posteriores a favor de H_A sobre H_B, Ω_0 son las apuestas (“odds”) a priori, y L es la razón de verosimilitud.

$P(H_i/D)$, el producto del teorema de Bayes, es la información que se desea determinar para guiar la acción futura, a la vista de la información obtenida en el proceso de observación experimental. Indica la probabilidad de que la hipótesis H_i sea verdadera, dada la información obtenida mediante observación, D.

Dado que $P(D/H_i)$ es una probabilidad razonablemente pública, y que $P(H_i/D)$ es lo que el científico desea conocer, la estadística clásica no utiliza el teorema de Bayes debido a que considera que no se conoce la probabilidad $P(H_i)$. Sin embargo, aunque la probabilidad a priori $P(H_i)$ sea frecuentemente vaga y variable, el impacto de este desconocimiento y de su variabilidad difiere mucho de unos problemas a otros. Si las observaciones son precisas, las propiedades de la distribución a priori tienen una influencia poco importante en la distribución posterior. Igualmente, la subjetividad de una opinión sobre un parámetro disminuye conforme se dispone de más información (datos).

#2.5.- LEYES A PRIORI, A POSTERIORI Y FUNCIONES DE VEROSIMILITUD

Si se dispone de una distribución a priori sobre θ, $\pi(\theta)$ y de una ley sobre las observaciones x, $f(x / \theta)$, a partir de esas distribuciones, se pueden construir:

- la ley de probabilidades conjuntas, $\varphi(\theta, x) = f(x / \theta)\, \pi(\theta)$

- la ley marginal de x, $m(x) = \int \varphi(\theta, x)\, d\theta = \int f(x / \theta)\, \pi(\theta)\, d\theta$

- la ley a posteriori de θ, obtenida mediante la fórmula de Bayes,

$$\pi\,(\theta / x) = \frac{f(x/\theta)\,\pi(\theta)}{\int f(x/\theta)\,\pi(\theta)\; d\theta} = \frac{f(x/\theta)\,\pi(\theta)}{m(x)}$$

La ley posterior representa la actualización de la información a priori, π (θ), a la vista de la información contenida en las observaciones, L(θ / x).

Así, en un ejemplo ofrecido por Robert (1992), si se desea calcular la distribución posterior, sabiendo que:

$$\pi(p) = \frac{1}{B(\alpha,\beta)}\, p^{\alpha-1}\,(1-p)^{\beta-1} \text{ con } 0 \le p \le 1$$

$$f(x/p) = C_n^x\, p^x\,(1-p)^{n-x}, \text{ con } x = 0, 1, ..., n$$

A partir de las funciones anteriores se pueden calcular:

- la distribución conjunta $\varphi(x,p) = \pi(p)\, f(x/p) = \frac{C_n^x}{B(\alpha,\beta)}\, p^{\alpha+x-1}\,(1-p)^{n-x+\beta-1}$

- la distribución marginal de x, $m = \frac{C_n^x}{B(\alpha,\beta)}\, B(\alpha+x, n-x+\beta) = C_n^x\, \frac{\Gamma(\alpha+\beta)\,\Gamma(n-x+\beta)}{\Gamma(\alpha)\Gamma(\beta)\,\Gamma(\alpha+\beta+n)}$

- la distribución posterior de p:

$$\pi(p/x) = \frac{p^{\alpha+x-1}\,(1-p)^{\beta+n-x-1}}{B(\alpha+x, \beta+n-x)} \text{ que es una Be}(\alpha+x, \beta+n-x)$$

En ciertos casos, puede que las observaciones no modifiquen algunos parámetros de la ley a priori, por ejemplo cuando la distribución de x no depende de esos parámetros o cuando el número de parámetros es muy elevado en comparación con el número de observaciones.

En los primeros años de desarrollo de la Estadística Bayesina, se realizó un gran esfuerzo en construir tablas en las cuales se daban familias de funciones de densidad a priori, distribuciones de la muestra y la correspondiente función de densidad a posteriori resultante de la conjunción de las anteriores familias de distribuciones, todo ello con objeto de simplificar los cálculos de los potenciales usuarios. Robert (1992) ha señalado que la búsqueda de leyes a posteriori analíticamente agradables es algo artificial y superado, ya que los instrumentos informáticos y numéricos de los que se dispone permiten utilizar leyes a priori más elaboradas y por tanto más próximas a la información a priori.

Las leyes naturales conjugadas de las familias exponenciales clásicas se resumen en la tabla 1.

Tabla 1.

Leyes posteriores obtenidas a partir de las leyes a priori y de las observaciones

f(x\\θ)	**$\pi(\theta)$**	**$\pi(\theta$\x)**
Normal N(θ, σ^2)	Normal N(μ, τ^2)	Normal N($\rho(\sigma^2\,\mu + \tau^2\,x), \rho\,\sigma^2\,\tau^2$) con $(1/\rho) = \sigma^2 + \tau^2$
Poisson P(θ)	Gamma G(α, β)	Gamma G(α+x, β+1)

Gamma $G(\nu, \theta)$	Gamma $G(\alpha, \beta)$	Gamma $G(\alpha+\nu,\beta+x)$
Binomial $B(n, \theta)$	Beta $Be(\alpha, \beta)$	Beta $Be(\alpha+x, \beta+n-x)$
Negativa Binomial $NB(m, \theta)$	Beta $Be(\alpha, \beta)$	Beta $Be(\alpha+m, \beta+ x)$
Multinomial $M_k(\theta_1, ... , \theta_k)$	Dirichlet $D(\alpha_1, \alpha_2, ... \alpha_k)$	Dirichlet $D(\alpha_1+x_1, \alpha_2+x_2, ... \alpha_k+x_k)$
Beta $Be(\alpha, \theta)$	Exponencial Exp (λ)	Exponencial Exp $(\lambda\text{-}\log(1\text{-}x))$
Normal $N(\mu, 1/\theta)$	Gamma $G(\alpha, \beta)$	Gamma $G(\alpha+0.5, \beta+(\mu\text{-}x)^2 /2$

Fuente: Robert (1992).

#2.6.- DETERMINACIÓN DE LAS LEYES A PRIORI

Un aspecto importante del análisis bayesiano es el de la elección de una ley o distribución a priori de los parámetros. Raramente se dispone de suficiente información como para que se pueda determinar con precisión esa ley, por lo que es necesario utilizar alguna aproximación para estimarla.

En la estimación subjetiva de las probabilidades a priori es fundamental contrastar la coherencia de los valores obtenidos y conocer que, en general, las personas tienen una gran dificultad para establecer los valores bajos de probabilidad, normalmente asociados a colas de las distribuciones.

Se han propuesto distintos métodos o alternativas para determinar una función a priori de densidad π para θ. Si Θ es discreto, el problema consiste en determinar la probabilidad subjetiva de cada elemento de Θ. Si Θ es continuo, entonces el problema de determinar $\pi(\theta)$ es más complicado.

A la hora de establecer probabilidades, en general será también más fácil hacerlo en aquellas situaciones donde θ pueda considerarse aleatoria en la interpretación frecuentista, antes que en los casos donde θ se considere aleatoria solamente en el contexto subjetivista, especialmente cuando se trata de determinar la probabilidad de casos que no han ocurrido o son únicos (p. e. cuál es la probabilidad de que estalle una guerra nuclear).

La forma más sencilla de determinar las probabilidades subjetivas es comparar los distintos sucesos y establecer la verosimilitud relativa de cada uno. Así, por ejemplo, si se desea determinar la probabilidad del suceso E, se puede comparar con E^c, y si el primero parece dos veces más probable que ocurra que el segundo, deducir que $P(E) = 2/3$.

Cuando Θ es un intervalo de R, una de las formas más intuitivas de asignación de probabilidades es construir el histograma de frecuencias. Para ello se divide Θ en intervalos y se determina la probabilidad subjetiva de cada intervalo. A partir de este histograma a veces es posible esbozar una función de densidad $\pi(\theta)$. En estos casos se pueden truncar las colas de forma inadvertida, y si éstas juegan algún papel en la predicción se perderá esa información, y ello puede tener importantes consecuencias en la determinación de la función de densidad posterior.

En otras ocasiones es posible construir una escala de verosimilitudes relativas de θ, de forma semejante a como se construye la función de utilidad. Si la clasificación es coherente (p.e. transitiva) se puede deducir una distribución a priori de θ. La existencia de distribuciones subjetivas derivadas de la ordenación de las verosimilitudes relativas es importante para evitar las justificaciones frecuentistas, muchas veces inaplicables.

Cuando Θ es un subconjunto de R se puede comparar la verosimilitud relativa de varios puntos de Θ para esbozar la función $\pi(\theta)$. Para ello se puede seguir una secuencia de varios pasos, ilustrada con el siguiente ejemplo de Robert (1992):

Sea $\Theta = [0,1]$.

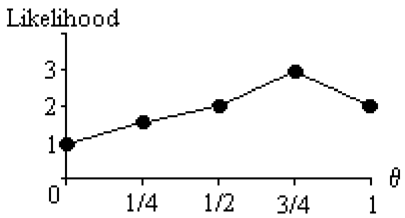
Paso 1.- Determinar los puntos para los que se presenta la probabilidad más alta y la más baja. Supongamos que en el ejemplo son $\theta = 3/4$ y $\theta = 0$, respectivamente.

Paso 2.- Se determina cuántas veces es más probable un punto que otro. En este caso, por ejemplo, se considera que la probabilidad de $\theta = 3/4$ es tres veces mayor que $\theta = 0$.

Paso 3.- En general, es suficiente determinar la verosimilitud relativa de otros tres puntos para obtener un esbozo suficiente de la función de densidad. Supongamos que esos puntos son $\theta = 1/4$, $\theta = 1/2$ y $\theta = 1$.

Paso 4.- La probabilidad de cada uno de esos puntos se compara con la de $\theta = 0$. Supongamos que consideramos que $\theta = 1/2$ y $\theta = 1$ son dos veces más probables que $\theta = 0$, mientras que $\theta = 1/4$, es 1.5 veces más probable que $\theta = 0$.

Paso 5.- Dando el valor una densidad anterior de 1 a $\theta = 0$, se obtiene el siguiente esquema de la función de densidad:



Paso 6.- Puede comprobarse la coherencia de la información representada, p.e. comprobar si la verosimilitud de $\theta = 1/2$ y $\theta = 1$ es la misma.

Paso 7.- Si se desea mayor precisión, pueden determinarse nuevos puntos y calcular la verosimilitud de cada uno de ellos.

De esta forma se puede obtener una función de densidad no normalizada, que puede normalizase dividiendo por la correspondiente constante.

Robert (1992) ha señalado que si el experimentador es capaz de establecer la verosimilitud de un suceso, se puede suponer que dispone, aunque sea imperfectamente, de un modelo probabilístico subyacente. El decisor puede disponer de un orden de preferencias, y por tanto de una ley a priori, aunque tan aproximada que sea imposible utilizarla con fines prácticos. Sin embargo es importante disponer de una axiomática que justifique la existencia de una ley a priori como representación del orden de verosimilitudes, sin tener que recurrir a un modelo frecuentista de repetición de las observaciones, que es poco realista.

La elección de una ley a priori a la vista de un histograma de verosimilitudes no es algo que se pueda hacer mecánicamente, porque varias formas de funciones de densidad pueden adecuarse al histograma observado y conducir cada una a inferencias muy diferentes.

Si Θ no tiene limites finitos ("unbounded"), se realiza la determinación de la verosimilitud en una región finita acotada previamente. Como primer paso se establece un juicio sobre la proporción de probabilidad de cada una de las regiones (dentro de los límites y fuera), y a partir de ahí se procede de forma similar a como se ha expuesto anteriormente.

También se pueden determinar las distribuciones a priori o de probabilidades subjetivas mediante la elección entre apuestas o reglas de puntuación ("scoring rules"). Si se desea calcular la probabilidad de E se puede imaginar un juego en el que se pierden z u.m. si E ocurre y se ganan (1-z) unidades si no ocurre (con $0 \leq z \leq 1$). Se trata de elegir un valor de z tal que el sujeto considere que el juego es "justo" (que tiene una utilidad nula), lo que implica, para valores z pequeños, donde se puede suponer que la función de utilidad es localmente lineal:

$$U(-z)P(E) + U(1-z)(1-P(E)) = 0 \Rightarrow P(E) = U(1-z) / [U(1-z) - U(-z)] \Rightarrow P(E) \approx 1-z$$

Otras alternativas se basan en la estimación de probabilidades a partir de equivalencias entre juicios sobre muestras y probabilidad de sucesos.

Hasta los trabajos de Jeffreys (1961) no se dispuso de una metodología de generación de leyes a priori que no dependiera de la distribución de las observaciones (aunque esta aproximación ha sido discutida por Lindley, 1971).

Frecuentemente se conjetura un conjunto de familias (llamadas leyes conjugadas) para las probabilidades anteriores (como la normal, la gamma, la beta, etc.) debido a que se disponen de resultados de composición de estas leyes con las de los datos observados para obtener la distribución posterior. En esta estrategia se sacrifica a un tratamiento matemáticamente más sencillo la elección de una ley que mejor describa la información a priori disponible.

Algunas situaciones favorecen la determinación de leyes a priori, como por ejemplo cuando se conoce el mecanismo físico, económico, biológico, etc. de generación del parámetro θ, en cuyo caso se puede conjeturar la ley a priori con bastante aproximación utilizando esa información.

En otros casos, en los que θ no tiene un referente, sino que se corresponde con una parametrización que describe el fenómeno aleatorio observado, la ley a priori π es un medio de resumir las informaciones disponibles sobre esa ley y la incertidumbre. En estos casos se suele proceder por aproximaciones.

Cuando el espacio de los parámetros Θ es finito, en muchos casos es posible realizar evaluaciones subjetivas de las probabilidades de los elementos θ_i. Así, por ejemplo, en algunos casos las probabilidades se pueden atribuir según la frecuencia relativa determinada en experiencias pasadas similares. Esta aproximación no es posible en caso de sucesos únicos o no observados (como determinar la probabilidad de que estalle una guerra nuclear).

Una alternativa frecuentemente propuesta es la de limitarse arbitrariamente a una forma parametrizada de densidad y evaluar los parámetros a partir de los momentos o de los fractiles. Por ejemplo, a partir de una evaluación subjetiva de la mediana y del cuartil 75% se pueden estimar dos parámetros de una ley normal. El método de los momentos, sin embargo, lleva frecuentemente a dificultades prácticas, principalmente la obtención de valores "imposibles" de los parámetros, p.e varianzas estimadas negativas.

El ajuste a una forma funcional se realiza con mucha frecuencia, aunque en general es un método inadecuado. Se supone que $\pi(\theta)$ viene determinada por una forma funcional concreta (p. e. se puede suponer que $\pi(\theta) \sim Be(\alpha, \beta)$, y se determinan los valores de α y β). La determinación de los parámetros (en el ejemplo, α y β) puede realizarse de varias formas:

- Estimar los parámetros de la función a priori a partir de la estimación a priori de algunos momentos. Por ejemplo, se pueden determinar α y β en $Be(\alpha, \beta)$ si se estima el valor de la media μ y la varianza σ^2, ya que $\mu = \alpha / (\alpha+\beta)$ y $\sigma^2 = \alpha \beta / [(\alpha+\beta)^2 (\alpha+\beta+1)]$.

Sin embargo la estimación de los momentos a priori es una tarea extremadamente delicada e incierta. Ello es debido a que las colas de las funciones de densidad tienen una gran influencia en los valores de los parámetros. Por ello este método es más aconsejable cuando el espacio es limitado, como $\Theta = [0,1]$.

- Estimar los parámetros de la función a priori a partir de la estimación subjetiva de varios fractiles de la función de densidad a priori. A partir de ahí se pueden elegir los parámetros de la f.d.p. a priori que mejor se ajusten a la estimación de los fractiles.

Dada una forma funcional para la f.d.p., normalmente se necesitan muy pocos fractiles para determinar los parámetros de la f.d.p. a priori. Si, dado un conjunto de fractiles, se determina un f.d.p. que verifica un subconjunto de los mismos, y el resto de los fractiles no encajan en la forma determinada, la conclusión es que se debe ajustar otra forma, tras el establecimiento de un modelo adecuado.

Un problema es que varias funciones de densidad a priori pueden ajustarse a las estimaciones existentes sobre los fractiles. Si las diferentes formas son equivalentes, y esto significa que la determinación de la función posterior es prácticamente la misma, esto no supone un problema mayor. En otro caso es necesario reconstruir un modelo adecuado para basar en el mismo la estimación de la f.d.p. que se considere más adecuada.

Si no se dispone de información que permita proponer aproximaciones a los momentos o los fractiles, se puede también utilizar la ley marginal de x

$$m(x) = \int_{\Theta} f(x \backslash \theta)\pi(\theta)d\theta$$

para estimar π. Ello es debido a que a veces el modelo de la variable observada aparece como un caso particular de un metamodelo, para el cual se dispone de informaciones más precisas. Esta aproximación se puede hacer siguiendo diferentes aproximaciones (Berger 1985).

Además del método de los momentos se puede utilizar el método de máximo de entropía o ML_II (Good, 1983), cuando se conocen algunas características de la ley de θ (momentos, fractiles). La ley a priori π que maximiza la entropía se corresponde con la minimización de la información a priori aportada por π. En distribuciones discretas, si no existen restricciones, la ley del máximo de entropía es una ley uniforme sobre Θ. En casos de leyes continuas la correspondencia no es tan mecánica.

Algunas aproximaciones que intentan resolver el problema de la determinación de funciones a priori son los métodos bayesianos jerárquicos y los empíricos. Los primeros modelizan la falta de información sobre los parámetros de la ley a priori por una nueva distribución de esos parámetros (hiperdistribución a priori). Los segundos utilizan las observaciones y la ley marginal para estimar los parámetros de la ley a priori. Es un método muy utilizado por los frecuentistas, porque lo consideran un método no bayesiano.

Cuando no se dispone de ninguna información acerca del modelo es imposible construir una ley a priori a partir de consideraciones subjetivas. Los métodos que se utilizan en estos casos se llaman no informativos, y fueron introducidos muy pronto, por ejemplo cuando Laplace reflejaba su ignorancia de la ley a priori utilizando como tal una ley uniforme. La equiprobabilidad de todos los sucesos no contiene información sobre el parámetro θ. Se han realizado dos criticas a esta aproximación:

- que las leyes uniformes no existen en los espacios no acotados, y que por tanto es necesario utilizar leyes generalizadas. Pero esta crítica no es importante, ya que no es necesario reducir la modelización del espacio de los parámetros a las distribuciones de probabilidad.

- una segunda critica, basada en la invarianza por parametrización, es más importante, habiéndose propuesto como alternativa la utilización de leyes a priori no informativas de Jeffreys.

#2.7.- DETERMINACIONES EXPERIMENTALES DE FUNCIONES DE DENSIDAD A PRIORI

La utilización de probabilidades subjetivas en la toma de decisiones implica la necesidad de obtener funciones de densidad de los decisores para las distintas variables relevantes, especialmente las distribuciones a priori en caso de emplear el teorema de Bayes. Sin embargo, las personas no están familiarizadas con la estadística o el calculo de probabilidades, y les resulta difícil expresar sus probabilidades a priori en la forma exigida por la teoría, aunque tengan una noción intuitiva suficiente. El problema, por tanto, se plantea en el desarrollo de métodos que permitan que el sujeto pueda expresar sus estimaciones de forma comprensible para un asesor.

Autores como Good (1965), Pratt - Raiffa - Schlaifer (1965), etc. sugirieron un conjunto de métodos para asignar probabilidades a priori y determinar las distribuciones correspondientes, que en parte se corresponden con algunos de los expuestos anteriormente. Estos métodos, como señaló Winkler (1967 a) suponen que los sujetos son estadísticos competentes, y que el decisor y el asesor son la misma persona. Otros trabajos, como los de Edwards (1953, 1954, 1964), Green et al. (1965) y Marschak (1964), investigaron la forma en que las personas asignan las probabilidades a priori, pero con objeto de discutir el comportamiento de los sujetos a la hora de realizar juicios que exigían la aplicación del teorema de Bayes.

Las primeras medidas experimentales subjetivas del riesgo fueron publicadas por Winkler (1967a), Bessler (1980) y Lindner y Gibbs (1990), y se basan en metodologías para la determinación de probabilidades subjetivas a partir de la información facilitada por los sujetos. En principio, existen dos tipos de problemas de estimación: la probabilidad (subjetiva) que se asigna a la ocurrencia de un suceso singular (estimación puntual), y la determinación de los valores posibles de una variable aleatoria (determinación de una distribución).

Para la determinación de la probabilidad subjetiva de un suceso particular se pueden utilizar dos métodos generales: la interrogación directa o la deducción de la probabilidad a partir de las preferencias del sujeto entre posibles decisiones. Los primeros se conocen como métodos directos y los segundos como métodos indirectos o conductistas (W. Lee, 1971; J.M. Hampton et al. 1973).

Los métodos directos se basan en preguntar a un sujeto más o menos directamente sobre la fuerza de sus creencias en diferentes sucesos o proposiciones, por ejemplo en una escala entre 0 y 1. La respuesta, sin embargo, no necesita obtenerse necesariamente como una probabilidad, sino que se puede expresar mediante sistemas equivalentes (por ejemplo, visualmente, haciendo que el sujeto divida un segmento en dos partes, señalando la probabilidad que asigna a dos sucesos disjuntos; o utilizando cartas, dados, etc. para a través de su conocimiento de los juegos expresar la probabilidad).

Otras veces se pide a los sujetos que expresen apuestas (“odds”) sobre distintas posibilidades. Por ejemplo se puede apostar 3 contra 1 a que llueva mañana, y suponiendo dos sucesos disjuntos (E_1 = llueve mañana, E_2 = no llueve mañana) se tendría que el sujeto asigna una probabilidad subjetiva de $P(E_1) = 3/4$.

En algunas situaciones puede que no se necesite una asignación cardinal de probabilidades, ya que en determinados problemas bastará con una ordenación de las probabilidades, llamadas por W. Lee (1971) “probabilidades cualitativas”. La valoración de la confianza en un enunciado puede bastar también en muchos casos, y se puede establecer mediante respuestas del tipo “es cierto con absoluta seguridad”, “creo que es cierto”, “posiblemente sea cierto”, etc.

Los métodos indirectos, o “conductistas”, en un sentido más reducido empleado por Savage (1954), en gran parte se basan en la interpretación de los resultados de las apuestas aceptadas por los sujetos y en situaciones de indiferencia entre varias alternativas. Preston y Baratta (1948) utilizaron técnicas llamadas de subasta. Lichtenstein (1965) utilizó el concepto de “oferta honesta” individual. Toda (1963) discutió sistemas de arrepentimiento.

La aproximación indirecta o conductista supone someter a un sujeto un conjunto de apuestas (loterías), similares a las descritas para el cierre de de Finetti, que debe aceptar o rechazar. Este procedimiento necesita mucho tiempo y trabajo para concluir sobre la probabilidad de un suceso, por lo que es más útil cuando se trata de determinar un conjunto reducido de estimaciones. Sin embargo, gran parte del trabajo experimental se realizó inicialmente con estos métodos.

El concepto de “lotería de referencia” trata de contrastar las probabilidades asignadas mediante la indiferencia entre apuestas, sustituyendo alguna de ellas por una opción que implique un mecanismo capaz de generar las probabilidades aceptadas en la indiferencia. Si las apuestas no se mantienen es necesario explorar otra situación de indiferencia.

Una aproximación sugerida por Raiffa (1968) se basa en el concepto de “urna equivalente”, donde el sujeto indica el número de bolas etiquetadas con cada posible suceso (p.e. “ventas altas”, “medias” y “bajas”, o E ocurre y E no ocurre) en una urna que es equivalente a su percepción de la probabilidad de ocurrencia de cada suceso. Una idea semejante subyace en los métodos basados en el concepto de “muestra equivalente”.

La construcción de distribuciones se basa en la idea de que se puedan derivar de un conjunto reducido de estimaciones puntuales como las discutidas anteriormente.

La aplicación de las técnicas comentadas para asignar probabilidades subjetivas es problemática, dado que los individuos no tienen práctica en realizar ese tipo de juicios, y por tanto encuentran dificultades para expresar sus opiniones en términos probabilísticos (Hogarth, 1975). Debido a ello se han realizado diferentes trabajos para evaluar los méritos relativos de cada técnica, como Stael von Holstein (1970), Ludke et al. (1977), etc.

Dos trabajos pioneros, experimentales, de determinación de distribuciones subjetivas de probabilidad por personas no expertas estadísticamente, fueron realizados a finales de los años 60 por R. L. Winkler (1967a) y M. Alpert y H. Raiffa (este último realizado en 1969 y publicado en 1984), influyendo de forma decisiva en las aproximaciones posteriores.

Posiblemente el método más utilizado por los experimentadores haya sido el de los fractiles. Sea x^* el valor verdadero u objetivo de alguna cantidad, desconocido para el asesor. La estimación del fractil $k \in [0, 1]$ de x^* es el número x_k tal que la probabilidad que se asigna al suceso $(x^* \leq x_k)$ es k. Es decir, $P(x^* \leq x_k) = k$. Los valores x_{50}, x_{25} y x_{75} son la mediana, y los cuartiles inferior y superior, respectivamente.

Para encontrar el valor x_{50}, el sujeto debe pensar un valor tal que estime que es tan probable que sea superior a que x^* como que sea inferior a x^*. De esta forma x_{50} divide el recorrido en dos intervalos, juzgados como igualmente probables. Para determinar el cuartil inferior, x_{25}, se divide el intervalo $(-\infty, x_{50})$ en dos intervalos que se juzgan equiprobables. Para determinar el cuartil superior, x_{75}, se divide el intervalo (x_{50}, ∞) en dos intervalos que se consideran equiprobables.

La consistencia o coherencia de la división requiere que el sujeto crea: (a) que cada uno de los cuatro intervalos $(-\infty, x_{25})$, (x_{25}, x_{50}), (x_{50}, x_{75}), (x_{75}, ∞) tienen la misma probabilidad de incluir el verdadero valor de x^*; y (b) que es tan probable que el valor de x^* se encuentre en el intervalo (x_{25}, x_{75}) como fuera de él. Normalmente el intervalo (x_{25}, x_{75}) se denomina “rango intercuartil”.

De igual forma se pueden estimar otros valores (x_k, k), y a partir de los mismos derivar una curva de probabilidades acumuladas.

Morrison (1967) propuso un conjunto de cuestiones para obtener del sujeto interrogado los fractiles en las funciones de distribución (acumulada) de probabilidades (CDF). Con esta técnica se sugiere que finalmente se realice la representación gráfica de los valores y que se compruebe su coherencia mediante apuestas.

Schlaifer (1969) propuso dos métodos:

a) Ajuste directo de una curva que exprese la función de densidad, cuando (1) el sujeto estima que esa curva discurre suavemente a derecha e izquierda de la moda, (2) cree que gran parte de la probabilidad se encuentra concentrada en torno a la moda, y (3) las colas están relativamente equilibradas.

b) Ajuste de datos históricos o escasos.

El ajuste de datos históricos permite realizar una estimación de las frecuencias a largo plazo mediante histogramas, fractiles o técnicas basadas en los diferentes momentos. Es necesario, sin embargo, realizar una interpretación subjetiva de los datos históricos, generalmente mediante una función unimodal.

Cuando los datos son escasos se puede inferir una distribución mediante técnicas que no supongan un tipo de función subyacente. Así, cuando se dispone de N observaciones de una variable aleatoria continua, si se ordenan de forma creciente, la observación k – ésima es un buen estimador del fractil k / (N+1), tal como se discute en Graybill (1963) y Barnett (1975). A partir de estas estimaciones de fractiles se puede construir también la función de distribución.

Anderson et al. (1974, 1977) señaló que había realizado experimentos con esa técnica a partir de distribuciones conocidas, y que con 3 – 5 observaciones obtuvo un 50% de probabilidades de estimar una función de distribución muy parecida a la verdadera. Con 12 o más datos la probabilidad crecía de forma importante.

Tanto Schlaifer (1969) como Raiffa (1968) pensaban que sería más intuitiva la estimación de funciones de distribución (CDF o acumuladas) que de funciones de densidad (PDF o f.d.p.). Semejante idea se encuentra en Anderson (1977).

Similar a estas metodologías es la ordenación sicométrica propuesta por Smith (1967) que se basa en la idea de establecer ordenaciones o rangos para los distintos intervalos o clases considerados, y aplicar la cuantificación de Kendall (1962) para traducir los rangos a probabilidades. Un método semejante había sido utilizado por Kruskal (1964) y Separd (1966) para escalas no métricas multidimensionales (J.M. Hampton et al. 1973). Morrison (1967) no lo consideraba práctico.

Tanto Winkler (1967 a) como M. Alpert y H. Raiffa (1984, 21) trabajaron con 5 percentiles (0,01, 0,25, 0,5, 0,75, 0,99), que determinan 6 categorías. En Winkler (1967 a) se dibujaron las curvas ajustando los cinco puntos. Las funciones se representaron para valores de p $\in$ [0.01, 0.99], sistema conocido como CDF truncada.

El trabajo de Winkler (1967 a) utilizó un grupo de estudiantes y profesores de la Universidad de Chicago, con conocimientos muy dispares de calculo de probabilidades, y se les sometió cuestiones relacionadas con atributos de los estudiantes de esa universidad (qué porcentaje existe de católicos, cuántos tienen menos de 21 años, cuántos son becarios o ciudadanos americanos, etc.). Se entrenó a los sujetos en las cuatro técnicas de estimación de distribuciones de probabilidad que iban a utilizar:

- CDF : función de distribución acumulada, mediante la asignación de fractiles directamente a las diferentes subdivisiones, dibujando la curva de la CDF.

- PDF: funciones de densidad (probability density function), establecidas mediante cuestiones relacionadas con densidades de intervalos o áreas relativas, y dibujando la función correspondiente.

- HFS: (hypothetical future samples) o determinación de la probabilidad asignada por los sujetos a los resultados futuros hipotéticos obtenidos mediante muestreo. Equivale al “device of imaginary results” de Good (1965).

- EPS: (equivalent prior sample information) o juicios expresados mediante muestras a priori equivalentes.

CDF y PDF son técnicas directas. Las técnicas HFS y EPS son técnicas indirectas, porque es difícil para el sujeto ver la relación de sus estimaciones con las distribuciones resultantes.

En la estimación de la función de distribución (acumulada) CDF se trata de construir juicios en torno a cuestiones del tipo $P(x \leq x^*)$ o $P(x \geq x^*)$, donde x^* un valor particular de la variable aleatoria x.

En la técnica del impacto visual, el intervalo continuo de la variable x se divide en intervalos mutuamente excluyentes. Para cada valor seleccionado de x^* se estima $P(x \leq x^*)$. Las probabilidades conjuntas $P(x^{**} < x \leq x^*)$ se determinan con relativa facilidad en la función de distribución CDF, y a partir de ellas es posible calcular las probabilidades condicionales mediante la expresión:

$$P(x > x^{**} / x \leq x^*) = P(x^{**} < x \leq x^*) / P(x \leq x^*)$$

La función de densidad PDF se puede, en teoría, derivar de la función de distribución, simplemente estimando el valor de sus tangentes, aunque es más práctico construir un histograma de frecuencia calculando las áreas correspondientes a partir de la información contenida en la CDF.

Winkler encontró que la función de densidad PDF parece más intuitiva que la función de distribución CDF, pero curiosamente mientras que las funciones de densidad eran unimodales en las respuestas obtenidas, las CDF parecían consistentes con la existencia de dos modas. Winkler atribuye el efecto a la escasa formación recibida por los sujetos encuestados sobre las relaciones entre las funciones de densidad y las de distribución.

La estimación de probabilidades subjetivas mediante el método HFS (hypothetical future samples o mecanismo de resultados imaginarios de Good), se basa en la idea de que una vez que el sujeto ha realizado una estimación de una proporción, mediante este método se examina el efecto sobre la distribución de la evidencia adicional disponible, y para ello el sujeto debe imaginar el efecto de la información aportada por una muestra aleatoria. La metodología procede mediante preguntas del tipo:

Supongamos que usted estima que la probabilidad de que un estudiante elegido al azar sea hombre es I_1. Seguramente deseará disponer de información adicional relevante para establecer su distribución. Suponga que esa información proviene de un conjunto de muestras independientes no acumulables extraídas al azar entre los estudiantes de la universidad. Si los resultados de alguna muestra no le parecen razonables puede ignorarlos. Pero para las muestras que considere razonables, indique cómo se vería afectada su estimación inicial por la nueva información aportada por la muestra, e indíquelo mediante I_{10}.

Las muestras se presentan de la siguiente forma general: Supongamos que hemos extraído una muestra de n estudiantes y que en ella hay m hombres. ¿Cuál es la probabilidad de que el siguiente estudiante elegido al azar en el conjunto de la población sea hombre?. Indíquelo en I_h. Las preguntas se construyen con pares (n, m), concretamente (10,10), (50, 25), (5,0), (100, 85), (1,0), (100, 60).

El método EPS (equivalent prior sample information) o juicios expresados mediante muestras a priori equivalentes, es similar al HFS. Las preguntas se formulan como:

Determine, en función de su conocimiento de los estudiantes de la Universidad, dos números, r y n, tal que usted considere equivalente observar r hombres en una muestra aleatoria de n estudiantes. (Observe que la razón r/n debe estar próxima a su estimación de p, y que cuanto mayor sea n será mayor será su confianza en la estimación).

Una cuestión interesante introducida en el estudio fue la claridad con que era percibida cada técnica por los sujetos (su nivel de comprensión y la facilidad de responder). En una primera sesión el orden de claridad fue HPS > EPS > PDF > CDF. En la segunda sesión el orden se mantuvo, pero en general se consideraron todas las técnicas más claras.

En el análisis de datos, el principal punto de interés se centró en la medida de la dispersión relativa introducida por las diferentes técnicas para cada sujeto y cuestión, utilizándose dos intervalos de confianza como medidas de dispersión ($p_{0,75} - p_{0,25}$ o intervalo del 50% y $p_{0,99} - p_{0,01}$ o intervalo del 98%).

En bastantes casos los sujetos daban respuestas compatibles con una distribución normal, posiblemente como consecuencia de la influencia de la formación recibida.

Una conclusión general del estudio fue que los sujetos tendían a sobrestimar los cuartiles inferiores y a subestimar los superiores. Es decir, tienden a cerrar la distribución en torno a la mediana.

En general, las técnicas directas (CDF y PDF) producen mayores intervalos que las indirectas.

Los resultados obtenidos se correspondían con la hipótesis de Edwards de que los sujetos son procesadores conservadores de información. Por ello dan a las muestras hipotéticas HPS menor peso del que debieran, obteniendo distribuciones mas estrechas. Cuando utilizan las técnicas directas, no son capaces de utilizar todo su conocimiento disponible y obtienen distribuciones más dispersas. Las diferencias indicarían la incapacidad de procesar intuitivamente toda la información disponible.

A un grupo reducido de sujetos se les mostró sus resultados, indicándoles las relaciones entre las funciones de distribución acumulada y las funciones de densidad. Cuando se estuvo seguro de que comprendían las relaciones, fueron sometidos a dos nuevas cuestiones, y las nuevas respuestas redujeron ligeramente la dispersión, haciendo menores las diferencias entre técnicas.

Tanto Winkler (1967 a) como M. Alpert y H. Raiffa (1984, 21) encontraron que cuando se discutía con los sujetos los sesgos de sus respuestas se mejoraban las estimaciones posteriores.

En un segundo paso, Winkler (1967 a) planteó un conjunto de apuestas para comprobar si las elecciones de los sujetos eran consistentes con sus distribuciones. De un total de 74 apuestas, solamente 9 fueron inconsistentes con las probabilidades asignadas, indicando que los sujetos generalmente eligen las decisiones que son consistentes con su asignación de probabilidades.

Se midió también la máxima discrepancia vertical entre pares de funciones de distribución acumuladas, utilizando una media similar al test de Kolmogorov - Smirnov. En general las diferencias eran menores cuanto mayores conocimientos estadísticos tenían los sujetos, indicando que las personas con una buena formación estadística tienen mayor facilidad para traducir sus juicios en probabilidades.

#2.8.- ESTIMACIÓN MEDIANTE APROXIMACIONES A FUNCIONES

En algunos casos se ha propuesto la estimación de distintos parámetros mediante aproximaciones a ciertas funciones, ajustadas empíricamente.

Una regla muy conocida para la determinación aproximada de los valores medios y la varianza de las duraciones de las actividades de un proyecto es la utilizada en el PERT aleatorio. La duración media de las actividades y su varianza se estima a partir de tres valores (a, m, b) que debe determinar subjetivamente el decisor o su asesor. En este caso se supone que la duración de la actividad se distribuye siguiendo una ley beta, cuyos parámetros son tales que el valor medio viene dado por [4] y su desviación típica por [5].

[4] $\bar{x} = \dfrac{a+4m+b}{6}$

[5] $\sigma = (1/6)\,(b - a)$

siendo (a, m, b) las estimaciones de las duraciones menor, moda y mayor, respectivamente.

La función de densidad de la ley beta es $f(x) = K\,(x-a)^{\alpha}\,(b-x)^{\beta}$, $x \in [a, b]$, y la especificación de los extremos a, b, la moda m y que la desviación típica es 1/6 del recorrido permiten determinar los dos parámetros α y β. Para simplificar los cálculos se suele aproximar el valor de la media al dado en [4].

Frecuentemente se da la función de densidad suponiendo un intervalo [0,1], y en este caso se tiene:

$$f(X) = \frac{1}{B(\alpha,\beta)} X^{\alpha-1}\,(1-X)^{\beta-1} \quad \text{con } 0< X <1$$

y entonces:

$$E(x) = \frac{\alpha}{\alpha+\beta}$$

$$V^2(x) = \frac{\alpha\beta}{(\alpha+\beta)^2\,(\alpha+\beta+1)}$$

$$\text{Moda }(x) = \frac{\alpha-1}{\alpha+\beta-2}$$

Existe, en general, una relación empírica entre la media, la moda y la mediana que también se puede utilizar para determinar la mediana de forma aproximada en este caso: mediana $\cong$ (2 media + moda) / 3.

Si se tiene un intervalo (a, m, b) y se transforma en un intervalo (0, 1), el valor de m queda en el nuevo intervalo como $m=\dfrac{m-a}{b-a}$. Como la moda en el intervalo (0, 1) es $\dfrac{\alpha-1}{\alpha+\beta-2}$, se tendrá:

$$\frac{m-a}{b-a} = \frac{\alpha-1}{\alpha+\beta-2} = \frac{\alpha-1}{(\alpha-1)+(\beta-1)}$$ e igualando se obtiene $\begin{cases}\alpha-1=m-a\\ \beta-1=b-m\end{cases}$ y por tanto

$$\begin{cases}\alpha = m-a+1\\ \beta = b-m+1\end{cases}$$

McCrimmon y Ryavec (1964) publicaron una de las primeras discusiones detalladas de los supuestos aleatorios del PERT, y señalaron las ventajas de la ley beta: es continua, unimodal, y se extiende entre dos extremos no negativos que cortan al eje de abscisas. A diferencia de la normal, que queda determinada por el valor de la media y de la desviación típica, la función beta necesita los parámetros adicionales α y β para su determinación. Para la descripción de la variabilidad de ciertas medidas (como la duración de las actividades, o los rendimientos de los cultivos), la función beta es muy flexible, ya que permite asimetrías a la derecha, a la izquierda o nulas. Presenta además la ventaja de que existen todos los momentos de la distribución (Nelson y Preckel, 1989).

Con anterioridad al trabajo de McCrimmon y Ryavec (1964), los supuestos aleatorios habían sido introducidos en los apéndices del informe inicial del PERT. Se puede consultar la bibliografía de este trabajo pionero para comprobar los trabajos previos que se habían realizado. Estos autores, además de exponerlos de forma sistemática, comprobaron la amplitud de los posibles errores introducidos por los supuestos.

Si la “verdadera” función de densidad es, como la beta, unimodal, continua y corta a la abscisa en dos valores positivos, y a = 0 y b = 1 (con $m \leq 1/2$), entonces el error máximo fue estimado en (1/3) (1-2m) para la media. Si m = a = 0 o m = b = 1 (la moda cerca de alguno de los extremos), el error puede alcanzar el 33%. Si la moda está más centrada, con $|1/2 - m| \leq 1/6$, entonces el error en la estimación de la media puede ser como máximo del 11%.

El error máximo en la desviación típica es del orden de 1/6, es decir, del 17%, y no depende del valor de la moda.

Estos errores, de producirse, se manifestarán en su valor absoluto, y por tanto pueden tener signo positivo o negativo. En el conjunto de las actividades de un proyecto es de esperar que unos errores se compensen con otros de signo contrario. La amplitud de las compensaciones dependerá del número de actividades que se realizan en un camino, de la amplitud de la variación de las duraciones y de la asimetría.

Las aproximaciones a la media y a la desviación típica, dadas por [4] y [5], pueden estar expuestas a errores máximos del 33% en la primera y del 17% en la segunda. Con valores $1 \leq \alpha \leq 6$ y modas del orden de $|1/2 - m| \leq 1/6$, los errores respectivos se reducen al 4% y al 7%.

Los anteriores errores se han calculado suponiendo que las estimaciones de a, m y b son relativamente exactas. Es posible que estos valores sean estimados, a su vez, con un cierto grado de error que influya en una estimación errónea de la media y de la desviación típica. Las estimaciones de McCrimmon y Ryavec (1964) dan una idea del impacto (es obvio que grandes errores en los valores de a, m y b pueden tener impactos enormes).

Suponiendo que los verdaderos valores sean (a, m, b) y que los estimados sean (αa, ηm, βb), la diferencia entre la media dada por el primer conjunto de valores y la obtenida como consecuencia de la estimación será:

$$[6] \qquad (1/6)\,[(\alpha-1)a,\ 4\,(\eta-1)m,\ (\beta-1)\,b]$$

Como α es el cociente entre el valor estimado a* y el verdadero valor a, cuando se estime el valor de a con exactitud, $\alpha = 1$, y el término correspondiente en [6] se anulará. Cuando a* = 0,9a (un error del 10%, subestimando el coeficiente), será $\alpha = 0{,}9$. Cuando a*=1,1a (un error del 10%, sobrestimando el coeficiente), será $\alpha = 1{,}1$. El correspondiente valor (α -1) en [6] será $\pm$ 0,1. Es decir, si todos los errores tuvieran el mismo signo (que sería la situación más desfavorable), el error en la estimación de la media vendría dado por [6], donde los coeficientes α, η, β expresan los errores respectivos de medida (expresada en tanto por 1) de cada valor (a, m, b).

En [6] se observa que un mismo error porcentual es más importante en la estimación de b que en la de a. El valor más alto posible, b, por tanto, puede ser una fuente de error importante y habría que prestar especial atención a su estimación. El impacto de un determinado nivel de error sobre b en comparación con m depende del valor de b. Si b > 4m, los errores cometidos en b tienen más peso en el error total que los cometidos en m (esto supone m muy a la izquierda).

El error absoluto para la desviación típica sería del orden de:

[7] $(1/6)\ [(\beta\text{-}1)\ b - (\alpha\text{-}1)a]$

De nuevo b aparece como una fuente potencial de grandes errores en la desviación típica, aunque su propia estimación no se realice con un gran error.

J.J. Moder y E. G. Rodgers (1968) señalaron que la definición de Malcolm et al. (1959) de los valores a, b en el PERT se referían a los extremos de la función de densidad beta supuesta, y por tanto dejaban el 100% de la probabilidad entre esos extremos. El problema que se plantearon resolver era cuál sería el valor de la media y la desviación típica si los valores de a y b se referían a algunos percentiles de la función de densidad.

En el resultado se supone que influirá el percentil considerado, la situación de la moda y la forma de la función de distribución. Para estudiar la variabilidad de esta última cuestión señalan que se pueden estudiar cinco distribuciones que cubren todo el rango de formas entre la normal y la exponencial (son la normal, la beta, la triangular, la uniforme y la potencial).

En su estudio comprobaron que $(b_{100} - a_0) / \sigma$ es muy sensible a la forma de la distribución (y moderadamente sensible a la localización de la moda en las funciones triangulares). La variación de la razón $(b_k - a_k) / \sigma$ es menor para otros percentiles. Por ello señala que la utilización de los percentiles del 5% y del 95% para los valores de a y b, o de 10% y 90%, permite una estimación de la desviación típica más robusta en relación con la localización de la moda y la forma de la función.

Si los valores que se estiman de la función de densidad de probabilidades son los percentiles $b_{100\text{-}\alpha}$ y a_α, entonces la media y la desviación típica vendrán dadas por:

[8] $x_\alpha = (a_\alpha + 4\ m + b_{100\text{-}\alpha}) / 6$

[9] $\sigma_\alpha = (b_{100\text{-}\alpha} - a_\alpha)/ K_\alpha$

con $K_{0,05} = 3,2$ y $K_{0,1}=2,7$.

J.J. Moder y E. G. Rodgers (1968) realizaron un experimento muy interesante. Seleccionaron cuatro grupos de 25 sujetos cada uno. El primer grupo estaba formado por sujetos con experiencia en la utilización de las estimaciones aleatorias en el PERT; el segundo por personas con una buena formación estadística, aunque sin experiencia específica en las estimaciones del PERT; el tercero formado por personas sin formación en ninguno de esos ámbitos, pero con experiencia profesional; y el cuarto similar al tercero, pero sin experiencia profesional.

A los sujetos se les suministró una lista de duraciones de tres actividades (1,3), (2,3) y (3,4), ocho observaciones para la primera y la última y siete para la (2,3), en tres bloques o repeticiones. Las actividades (1,3) y (2,3) se podían realizar en paralelo, y ambas precedían a la actividad (3,4). Con este esquema se pedía a los sujetos que estimaran la duración media del proyecto (del vértice 1 al vértice 4), el valor de m y los valores a_α, $b_{100\text{-}\alpha}$, con $\alpha \in [0;\ 0,05;\ 0,1]$. De los distintos resultados concluyeron la recomendación de utilizar estimaciones con $\alpha = 0,05$, a fin de mejorar las estimaciones de la varianza (que, por otra parte, como hemos visto, serían bastante independientes de la forma de la función de densidad).

C. Perry y I.D. Greig (1975) señalaron que las mejores estimaciones con los percentiles 5% y 95% posiblemente se deba a que pueden formar parte de la experiencia de los sujetos, mientras que los extremos 0% y 100% puede que nunca hayan sido observados (si existen). Por otra parte, consideran que las personas tienden a redondear más la forma de la función que los picos de la misma, por lo que proponen $K_{0,05} = 3,25$ en [9].

El problema de obtener una buena aproximación a la media, que no dependa excesivamente de la forma funcional que se suponga, les lleva a proponer:

[10] $x= (p_{0,05} + 0,95\ m + p_{0,95}) / 2,95$

que da errores inferiores al 5% con la mayoría de las funciones beta, gamma, y lognormal con desviaciones típicas no muy extremas y no excesivamente asimétricas. Pero [10] es prácticamente la media de una función triangular.

Entre la media y la mediana existe una muy buena aproximación, muy independiente de la forma de la función, dada por:

[11] $x = p_{50} + 0,185\ (p_{0,95}+p_{0,05}-2p_{0,50})$

Esta aproximación da errores inferiores al 0,5% para funciones que no presenten una leptocurtosis extrema.

Estos autores concluyen que las fórmulas [10] y [11], con K = 3,25, se pueden aplicar en problemas prácticos cuando se puede ignorar la forma de la función de densidad.

Otra aproximación muy popular en la literatura ha sido la ley triangular. McCrimmon y Ryavec (1964), al determinar los niveles de error al utilizar funciones beta en el PERT, señalaron que los niveles de error que se podrían haber producido si se hubieran supuesto leyes triangulares en su lugar hubieran sido del mismo orden. Destacan su completa caracterización mediante a, m y b, teniendo fórmulas precisas de la media y de la desviación estándard. Algo semejante se puede leer en Anderson et al. (1977), quienes señalan la ventaja de disponer de una función de distribución CDF cuadrática. No obstante, en las aplicaciones agrarias, Pease (1992) se mostró escéptico sobre su utilidad, excepto en casos especiales.

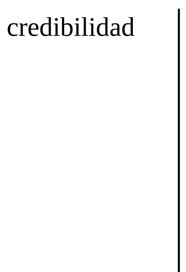
La media y la desviación típica de la función triangular viene dada por:

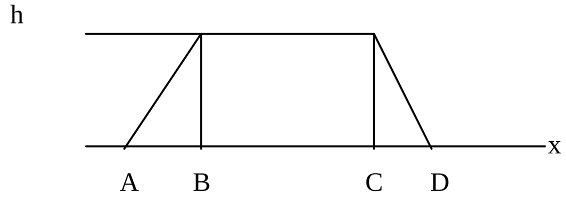
[12] $x = (a+m+b)/3$

[13] $\sigma = \sqrt{\frac{1}{18}\left[(b-a)^2 + (m-a)(m-b)\right]}$

Otra aproximación muy popular ha sido la de Allen (1968), quien propuso, para el caso de situaciones no repetibles o únicas, una aproximación funcional basada en los conceptos de credibilidad y sorpresa potencial de Shackle (1961). Este último autor había sostenido que en la selección de proyectos era imposible hacer una enumeración de todas las alternativas, paso previo para asignar la probabilidad de ocurrencia a cada una de ellas. Por ello, la selección de proyectos se debería basar en las credibilidades, que no necesitan sumar uno y que se pueden determinar con una experiencia incompleta.

Para determinar un diagrama de credibilidad de una variable se realiza el siguiente razonamiento: “Con la información disponible actualmente considero cualquier valor x entre B y C como completamente creíble (o alternativamente, no me sorprendería que el valor de x tomara valores entre B y C), y considero totalmente increíble que x tome valores menores que A o mayores que D…”.





Se puede suponer que la figura representa una función de densidad, y por tanto h = 2 / (AD+BC).

El concepto de sorpresa ha sido criticado por su ambigüedad, y debido a su naturaleza cualitativa. El concepto de credibilidad puede dar origen a problemas de coherencia. Se puede enlazar la idea, sin embargo, con una función de densidad a priori difusa.

#2.9.- CALIBRACIÓN

Cuando se establece una probabilidad subjetiva de forma coherente y teniendo en cuenta la información relevante disponible, no es razonable calificar alguna proposición así establecida como "equivocada". Sin embargo, en muchas situaciones, es posible verificar la verdad o falsedad de las proposiciones en las cuales interviene una probabilidad, en función de los datos posteriormente observados. De esta forma se puede determinar qué proposiciones han sido más o menos precisas y "calibrar" a algunos sujetos en sus predicciones. Por ejemplo, si un sujeto predice que "mañana lloverá", es posible determinar la exactitud con la que realiza la asignación de probabilidades, y esta operación se conoce como "calibración".

M. Alpert y H. Raiffa (1984, 21) señalaron que desde el punto de vista de la teoría, si mediante algún método obtenemos de un sujeto una distribución de probabilidades de una cantidad incierta para él, pero que nosotros conocemos, y se comprueba que las distribuciones no coinciden, carece de sentido decir que la distribución es incorrecta. La distribución de probabilidad es una expresión formal de lo que el sujeto conoce o no conoce sobre una cantidad. No se puede decir en el seno de la teoría que esa distribución sea "demasiado cerrada", "demasiado abierta", "inclinada hacia la derecha", etc.

Sin embargo, si el sujeto genera cientos de distribuciones de diferentes cantidades y si cada uno de los valores reales se encuentra entre los fractiles 0,01 y 0,99, podemos decir que el sujeto no está externamente calibrado, que sus distribuciones tienden a ser demasiado cerradas.

Tal como lo expresan Lichtenstein et al. (1984, 22), si una persona estima que la probabilidad de que una proposición sea verdadera es 0,7 y después encuentra que la proposición es falsa, esto no invalida su estimación. Sin embargo, si asigna una probabilidad de 0,7 a 10.000 proposiciones independientes y resulta después que solamente 25 son verdaderas, existe algo equivocado en sus estimaciones. En este caso se dice que al sujeto le falta calibración o realismo (Bronw y Shuford, 1973) o validación externa o realismo o confianza (Adams y Adams, 1961, Oskamp, 1962), o validación secundaria (Murphy y Winkler, 1971) o confianza (Murphy, 1973).

Formalmente, un juicio es calibrado si, a largo plazo, para todas las proposiciones a las que se les ha asignado una probabilidad, la proporción de verdaderas es igual a la probabilidad asignada.

La precisión de las proposiciones se ha medido utilizando, principalmente, tres métodos: (1) Experimentos de laboratorio, donde los sujetos expresan sus probabilidades a priori, que son comparadas con las probabilidades "objetivas" conocidas por el experimentador. (2) Evaluación mediante funciones de penalización, que dependen de la probabilidad asignada a una proposición y de los resultados obtenidos. No es un método adecuado para decisiones que no se realizan una sola vez o en un número escaso de ocasiones. (3) Comparando la evaluación subjetiva de la probabilidad de varias series de ensayos con las frecuencias relativas empíricamente obtenidas. El problema es que no se dispone de una definición del largo plazo, a la hora de calcular las frecuencias relativas.

En general, los trabajos sobre el establecimiento de inferencias se han visto afectados por: (1) considerar situaciones muy simples, cuyos resultados no son generalizables; (2) por utilizar sujetos que no se pueden considerar expertos (sustantivos o normativos) en las áreas objeto de examen; y (3) por emplear una información de base totalmente especificada, en lugar de la que interesa a los sujetos encuestados.

Una cuestión central en la discusión sobre estimaciones de funciones de densidad es la de si existen incentivos o no a que el asesor exprese sus estimaciones verdaderas (es decir, que sea honesto y coherente con el resto de sus creencias). Los asesores en muchas situaciones pueden tener incentivos para ofrecer respuestas que no se corresponden con sus verdaderos juicios, ya que pueden percibir (correcta o incorrectamente) un esquema de recompensas y de penalizaciones que le lleven a expresar intencionadamente juicios distintos a los que cree.

Así, por ejemplo, si en una situación particular el asesor percibe que se considera un error más importante la sobrestimación que la subestimación, tenderá a ofrecer los menores valores de sus estimaciones. Esta percepción puede derivarse de su experiencia pasada.

El problema de estimar las funciones de densidad (u otros parámetros) no se reduce a obtener la información que se desea, p. e. mediante los métodos de interrogación directa, sino que deben también existir mecanismos para incentivar que se revelen los juicios subjetivamente verdaderos (que se exprese lo que realmente se cree).

Winkler, R.L. (1967b) puso de manifiesto que es imposible determinar en qué medida un juicio sobre probabilidad es coherente con el resto de juicios sostenidos por el sujeto. Aunque no se puede verificar esa coherencia, sugirió métodos para ayudar a establecer las estimaciones de acuerdo con el resto de sus juicios.

En un primer paso se puede interrogar directamente sobre las probabilidades que se estiman (estimación de gráficos de funciones, apuestas, loterías, muestras hipotéticas, etc.). Esta aproximación tiene el inconveniente de que los asesores no tienen incentivos para realizar asignaciones de probabilidades coherentes con el resto de sus juicios. Así, si pregunta a un sujeto cuál es la probabilidad de que llueva mañana, puede que conteste con cualquier cifra, porque no existe ninguna recompensa o multa por lo acertado de su predicción.

Debido a ello sugirió el empleo de métodos adicionales que implican apuestas y funciones de penalización para incentivar unas respuestas más cuidadosas.

En la literatura se han discutido diversas medidas, especialmente en Winkler (1967 b), Raiffa (1969), Murphy (1973), Hampton, Moore y Thomas (1973), Murphy y Winkler (1974), Lichtenstein et al. (1984, 22), etc.

Winkler (1967 b) postuló que un buen asesor:

(1) No viola los postulados de coherencia: nunca acepta una apuesta en la que pierde cualquiera que sea el resultado, tal como postulaba Savage et al. (1962).

(2) Comprende los métodos utilizados para asignar las probabilidades, y estos métodos favorecen las estimaciones cuidadosas o, lo que es lo mismo, comprende las diferentes alternativas y las implicaciones de cada uno.

(3) Tiene una función de utilidad lineal respecto a las cantidades monetarias en el rango relevante de decisión. Sus elecciones se realizan de forma que maximiza la utilidad esperada.

Si el asesor no es neutral al riesgo (en el rango relevante de decisión), esta circunstancia puede afectar a la dirección de corrección de las funciones de penalización, de igual forma que afecta a las situaciones en las que se emplean apuestas. Una persona amante del riesgo tenderá a expresar estimaciones muy concentradas en torno a un punto, mientras que una persona adversa al riesgo intentará aproximarse al valor 0,5 en una dicotomía o asignar una función de distribución uniforme.

Si f(x) es la función de densidad de una variable aleatoria X, y se considera una partición X_1, X_2 …, X_n sobre la recta real, entonces f(x) puede resumirse mediante una serie p_1, p_2, …, p_n con $p_i = P(x \in X_i)$.

Supongamos que se estima subjetivamente que la distribución de probabilidades es r(x), con $r_i = P(x \in X_i)$, cumpliendo los requisitos de coherencia. El problema que plantea Winkler (1967 b) es qué métodos llevarían al asesor a hacer r(x) idéntica a f(x), o lo que es lo mismo, hacer $r_i = p_i\ \forall i$. Este autor revisó las siguientes alternativas:

1) Métodos basados en apuestas. Se trata de ofrecer al asesor un conjunto de apuestas reales, en el sentido de que las cantidades monetarias implicadas realmente cambien de manos, tal como planteó Goods (1965). En cada caso presentado, el asesor elige entre pares de apuestas, sin que se pueda abstener de apostar.

Cada par de apuestas permite obtener desigualdades, y para determinar el valor de la probabilidad con un determinado grado de precisión puede que sea necesario formular un gran número de apuestas. Para evitarlo, Winkler (1967 b), en la misma dirección que las propuestas de Savage y de Finetti, proponía realizar una primera aproximación mediante asignaciones directas y después utilizar las apuestas para verificar los resultados obtenidos.

Si se emplean k diferentes métodos de interrogación directa, que producen resultados distintos, podemos tomar las respuestas r_{jn} de cada método concreto j (j= 1,2,…, k) y mediante un conjunto adecuado de apuestas determinar cuáles son consistentes con todas ellas. Por ejemplo, si se tienen dos sucesos X_1 y X_2, y se emplean dos métodos de interrogación directa, se tendrán las respuestas: r_{11}, $r_{21}=1-r_{11}$, and r_{12}, $r_{22}=1-r_{12}$. Supongamos que se utiliza el juego 1 con $q = (r_{11}+r_{12})/2$. La elección del asesor determinará en qué medida la respuesta r_{11} o r_{12} es mejor aproximación a p_1.

2) Métodos basados en funciones de penalización o métodos de puntuación ("scoring rules"). Esta metodología fue propuesta por de Finetti (1965), Shuford et al. (1966), Toda (1963), y van Naerssen (1962). Las funciones de penalización o de puntuación suponen el cálculo de una puntuación de acuerdo con una regla diseñada para obligar al asesor a revelar sus creencias verdaderas, ya que en otro caso obtendrá una puntuación menor.

La recompensa o la penalización es proporcional a la puntuación obtenida (pudiendo ser la propia puntuación la recompensa). De esta forma el asesor intentará maximizar su puntuación esperada: una vez comprendida la función de penalización, realizará la mejor asignación automáticamente.

Las funciones de penalización son muy interesantes y valiosas porque no son tan costosas en consumo de tiempo como los métodos basados en apuestas.

Las reglas de puntuación son esquemas de pagos que dependen de la distribución estimada y la que realmente resulta posteriormente. En un conjunto de n estados mutuamente excluyentes y exhaustivos, si las creencias del asesor son p={p_1, p_2, ..., p_n}, pero las expresa como r ={r_1, r_2, ..., r_n}, se dice que este juicio es honesto si r = p. Una regla de puntuación asigna una puntuación o recompensa $S_i(r)$ si se presenta el estado i, y de esta forma la puntuación esperada es $S(r,p) = \Sigma\, p_i\, S_i(r)$. Las reglas de puntuación admisibles serán aquellas para las cuales S(r, p) alcanza su máximo cuando r = p (Anderson, 1977).

Estas reglas se emplean en situaciones en las que es posible un aprendizaje adaptativo en función de la retroalimentación, por ejemplo, en la investigación clínica, en la predicción meteorológica, etc.

De Finetti (1965) propuso una regla general de puntuación, donde, si el verdadero valor de X es X_h:

$S_1 = 2r_h - \sum_{j=1}^{n} r_j^2$ cuyo recorrido es [-1,1]

A partir de esta formulación general sugirió dos variantes que permiten cambiar de escala, dado un valor r_i que cumpla el postulado de coherencia:

$S'_1=(S_1+1)/2$ cuyo recorrido es [0,1]

y $S''_1=(S_1-1)/2$ cuyo recorrido es [-1,0]

En general una combinación aS_1+b, con a>0, tendrá un recorrido dado por [-a+b, a+b].

Una regla de puntuación de la forma aS_1+b, con a>0, incentiva a los asesores a expresar sus juicios verdaderos, ya que la maximización de la esperanza matemática de esa regla implica $r_i = p_i$ (i=1,2,...,n).

En el caso n = 2, se tiene una dicotomía, y $S_2 = 1- (1-r_h)^2$ es la regla de puntuación de Brier (1950), que se utiliza en la calibración de las predicciones meteorológicas.

En el caso n > 2, la expresión $S_1+(n-1)$ equivale a la suma de una serie de puntuaciones S_2. Una partición en n intervalos puede ser interpretada como un conjunto de dicotomías, o conjunto de pares de particiones, donde cada una de ellas está formada por un elemento y el resto de los n-1 elementos. De esta forma se tiene:

$S_{2i} = 1- (1-r_i)^2$ si $X \in X_i$

$S_{2i} = (1-r_i)^2$ si $X \notin X_i$ i=1,2,...,n

En el primer caso, si $X \in X_h$ se verificará que $\Sigma_{i=h}\, S_{2i} = 1 - (1-r_h)^2 + \Sigma_{i\neq h}\, (1-r_h)^2 = 2r_h - \Sigma_i\, r_i^2 + n - 1 = S_1 + (n-1)$.

Toda (1963) propuso la puntuación $S''_2=S_2-1$ en dicotomías (n=2) cuando la función de pérdida es cuadrática. Igualmente propuso las siguientes expresiones:

Para funciones de ganancia esférica:

$$S_3 = \frac{r_h}{(\sum_{j=1}^{n} r_j^2)^{1/2}}$$

Para funciones logarítmicas de pérdida:

$S_4 = \log r_h$

o, en general, $S_3(r) = a \log_e r_i + b$, $a>0$. Obviamente, $r_i = 0 \Rightarrow S_i(r) = -\infty$ y $S_i(r)$ depende solamente de r_i.

Se ha desarrollado la correspondiente prueba de que tanto S_3 como S_4, y sus transformaciones lineales positivas, obligan al asesor a expresar sus verdaderos juicios.

Algunas reglas de puntuación parecen razonables, pero no cumplen los requisitos necesarios. Así, parece razonable la regla $S=r_h$ (puntuación igual a la probabilidad asignada al intervalo que contiene el valor verdadero). Este esquema fue, por ejemplo, utilizado por Grisley y Kellog (1983) en un contexto agrario. Sin embargo, en este caso el asesor puede maximizar su puntuación esperada dando un valor 1 al r_i correspondiente al caso que considere más probable y un valor 0 al resto de los valores r_i . Así, por ejemplo, en una dicotomía en la que $p_1 = 0{,}6$ y $p_2 = 0{,}4$, si el asesor predice $r_i = p_i$, sus asignaciones verdaderas, entonces la puntuación esperada es $0{,}6^2 + 0{,}4^2 = 0{,}52$, pero si hace la estimación oportunista $r_1 = 1$ y $r_2 = 0$, entonces la puntuación esperada es $1(0{,}6) + 0(0{,}4) = 0{,}6$. Por ello el trabajo de Grisley y Kellog (1983) fue criticado por Knight, Johnson y Finley (1985).

Este ejemplo muestra que no es suficiente que una regla parezca intuitivamente razonable, sino que además debe forzar al asesor a expresar su estimación verdadera.

McCarthy (1956) demostró que una regla S es admisible si y sólo si adopta la forma $S = \delta s(r_1,r_2, ..., r_n)/ \delta r_h$, donde s es una función convexa respecto a r_i, y homogénea de grado 1. Como consecuencia, cualquier combinación lineal positiva de reglas admisibles es también una regla admisible.

Como las funciones de penalización dependen de r_i ($i = 1, 2, \ldots, n$), también dependerán de la partición en n intervalos que se realice. En algunas situaciones la naturaleza de X sugiere una partición obvia, pero en otras situaciones debe buscarse un compromiso entre el deseo de obtener la máxima información del asesor (haciendo una partición muy fina) y la necesidad de simplificar los cálculos.

En el caso de que no se desee obtener la función de distribución del asesor, las funciones de penalización se pueden utilizar para estimar otros parámetros de la distribución. Por ejemplo, si el asesor tiene que estimar el verdadero valor de X y se utiliza una función de penalización proporcional al cuadrado de su error, entonces su mejor estrategia es revelar la media de su distribución. Si la penalización es proporcional al error absoluto, entonces su mejor estrategia es revelar la mediana de su distribución, etc.

A veces lo que se trata de determinar son intervalos del verdadero valor de la variable, y la puntuación dependerá de que el valor real caiga o no en un intervalo particular. Para estas situaciones Toda (1965) desarrolló dos métodos, conocidos como juego del intervalo y juego de particiones.

Lichtenstein et al (1984, 22), para el caso de proposiciones discretas, indican la forma de obtener una curva o gráfica de calibración. Para ello parte de una descripción del número de alternativas que se pueden ofrecer como respuesta en las proposiciones: ninguna y se describe una única respuesta para ser evaluada como verdadera (1) o falsa (0); dos alternativas, y en este caso la verdadera es marcada con 1 y la falsa con 0; y más de dos alternativas, y en este caso: (a) el asesor puede seleccionar la respuesta más probable de k posibles y asignarle una probabilidad mayor que 1/k; o (b) puede dar a cada alternativa un valor entre 0 y 1. El método consiste en:

1.- Recopilar varias estimaciones de probabilidad para cuestiones cuya respuesta correcta es conocida o que puede ser fácilmente conocida por el experimentador.

2.- Agrupar las estimaciones similares en rangos (p.e. las estimaciones entre 0.6 y 0.69 en una misma categoría).

3.- En cada categoría, calcular la proporción de estimaciones correctas (p.e. la proporción de cuestiones para las que la proposición es correcta o la alternativa es correcta).

4.- Para cada categoría, dibujar la respuesta media (en abscisas) contra la proporción de correctas (en la ordenada).

La calibración perfecta debería dar un conjunto de puntos unidos por la recta de pendiente unidad.

En el caso (a) las estimaciones mal calibradas pueden ser por sobreconfianza, cuando la proporción correcta es menor que la estimada, y en este caso la curva de calibración está por debajo de la recta de pendiente unidad, o subconfianza, cuando la proporción correcta es mayor que la estimada y la curva de calibración esta por encima de la recta de pendiente unidad.

En el caso (b) (full-range) la sobreconfianza tiene dos posible significados:

- el asesor tiene un exceso de confianza en lo correcto de su estimación (la curva de calibración por debajo de la recta de pendiente unidad)

- el asesor tiene un exceso de confianza en su capacidad de discriminar entre proposiciones verdaderas y falsas (en este caso la curva de calibración está por debajo de la recta de pendiente unidad en la región por encima de 0,5 y sobre la recta unidad en la región por debajo de 0,5).

#2.10.- FUNCIONES DE DENSIDAD DEL RENDIMIENTO DE LOS CULTIVOS

La referencia clásica sobre la forma de la función de densidad de probabilidad de los rendimientos de los cultivos es Day (1965), quien señaló que no se distribuyen según una función de densidad normal, como se suponía generalmente con anterioridad. Los datos de este trabajo provenían de una estación experimental en Mississippi y se referían a los cultivos de maíz, algodón y cebada, e indicaban claramente distribuciones asimétricas (asimetría negativa). Con esos resultados, sugirió el empleo de funciones tipo I (como la beta), asimétricas y con forma de campana, como hipótesis a priori razonables para investigaciones posteriores.

Los resultados anteriores le permitieron señalar que las funciones de utilidad basadas en esperanzas matemáticas y varianzas son claramente inadecuadas en la programación de cultivos, por lo que es necesario considerar también la asimetría de las funciones de densidad. Un resultado significativo de este trabajo es la reivindicación de estimadores como la moda o la mediana de los rendimientos, como alternativa a la media.

Teigen (1985) dedujo funciones de rendimientos de maíz y de soja en función de diferentes variables, en un enfoque relacionado con la obtención de funciones de producción. Thompson (1986) calculó mediante regresión los rendimientos del maíz en función del tiempo. Campbell et al. (1988) publicaron los resultados de una regresión rendimiento-humedad el para trigo en Saskatchewan, mientras que Thompson (1988) los publicó para los rendimientos de maíz y soja en U.S. en función del clima y Kaylen y Koroma (1991) para los de maíz. Love y Buccola (1991) estudiaron el rendimiento del maíz en función de los fertilizantes aplicados.

Teigen (1991) publicó una relación del rendimiento de la soja con el clima (precipitación y temperatura), pero poniendo de manifiesto que las medidas estadísticas del clima normalmente no reflejan todos los sucesos agronómicamente significativos para el cultivo (por ejemplo, calor o lluvia durante la floración y formación de la vaina). Teigen (1991b) relacionó el rendimiento del maíz con el clima. Unos años después, Teigen y Thomas Jr. (1995) señalaron que el clima es el factor más importante que afecta a la producción de los cultivos, de forma que la relación entre clima y rendimiento explica más del 90% de la variación del rendimiento en la mayoría de los casos estudiados. La distribución de los rendimientos depende de la distribución de probabilidad del clima y de una función de respuesta no lineal. Las temperaturas siguen una distribución de probabilidad simétrica (normal), y las precipitaciones una distribución gamma (no simétrica), por lo que los rendimientos de los cultivos tienen componentes de funciones normales, log-normales, chi-cuadrado y otras variables aleatorias.

Gallagher (1987) confirmó que los rendimientos de los cultivos no son simétricos y que muestran una asimetría negativa, por lo que la distribución normal es inapropiada para describirlos. Por ello, utilizó una distribución gamma para modelizar los rendimientos de la soja.

Nelson y Preckel (1989) utilizaron una distribución beta condicional como modelo paramétrico de la distribución de probabilidad de los productos, y estudiaron la respuesta del rendimiento del maíz a las aplicaciones de fertilizante. Las distribuciones de maíz obtenidas mostraron una asimetría negativa. Esta modelización de los productos como una ley beta permitía obtener mejores resultados que los de Antle et al. (1984), que utilizaron un modelo basado en momentos.

Babcock (1990) utilizó una distribución beta con tres parámetros para describir los rendimientos asimétricos del cacahuete. Nelson (1990) admitió que las funciones de densidad de los rendimientos de los cultivos tienen asimetría negativa, y utilizó una distribución beta para modelizar los rendimientos del maíz. Taylor (1990) propuso como función de distribución acumulada de los rendimientos una función tangente hiperbólica. Swinton y King (1991) admitieron la asimetría negativa como típica de las funciones de densidad de probabilidad en los rendimientos de trigo. Moss y Shonkwiler (1993) señalaron que está suficientemente reconocido que los rendimientos no se distribuyen normalmente. Por ello, utilizan la transformación seno hiperbólico inverso para incorporar la asimetría negativa en un modelo de rendimientos de maíz. Babcock y Hennessy (1996), utilizaron una distribución beta, que permite asimetría a ambos lados.

Kaufmann y Snell (1997) integraron variables climáticas, económicas y técnicas para describir los rendimientos de maíz. Ramírez (1997) encontró que los rendimientos del maíz y de la soja son asimétricos, presentan curtosis, y muestran varianzas diferentes con el tiempo, mientras que los rendimientos de trigo parecen normales, pero heteroscedásticos. Parametrizó una transformación seno hiperbólico inverso para desarrollar y definir formalmente una función de densidad multivariada, no normal, que tenga en cuenta la asimetría, la curtosis y la heteroscedasticidad detectadas en la correlación.

Goodwin y Ker (1998) señalaron que la mayoría de las funciones de densidad obtenidas mediante técnicas no paramétricas muestran una asimetría negativa. Como es sabido, las técnicas no paramétricas de estimación de funciones de densidad no suponen previamente ninguna forma particular para las distribuciones de rendimientos, sino que permiten seleccionar la representación más apropiada en función de los datos.

R. E. Just y Q. Weninger (1999), finalmente, han subrayado la ausencia de consenso sobre la forma de la función de densidad de los rendimientos o de los ingresos, y las importantes implicaciones que supone el abandono de la normalidad (entre otras, se invalidan las aproximaciones media - varianza E-V en la maximización de la utilidad esperada).

La tesis de estos autores es que con la evidencia actual no se puede concluir sobre la no normalidad de los rendimientos de los cultivos, y se apoyan en tres posibles fuentes de problemas: la mala especificación de los componentes no aleatorios de los rendimientos; escasa evidencia sobre la significación estadística; y la utilización de series de tiempo agregadas.

En la discusión sobre el aislamiento de los componentes realmente aleatorios de la distribución de rendimientos, estos autores ponen de manifiesto uno de los principales problemas que sustenta la presente investigación: la evolución tecnológica puede dificultar la detección de la forma de la f.d.p.

Tabla 2.

Densidad de probabilidad de los rendimientos conjeturadas.

Autor	**Cultivos**	**Forma función de densidad**
Day (1965)	cebada, maíz, algodón	asimetría positiva – beta
Gallaguer (1987)	soja	asimetría negativa - gamma
Nelson y Preckel (1989)	maíz	asimetría negativa – beta
Babcock (1990)	cacahuete	beta tres parámetros
Nelson (1990)	maíz	asimetría negativa - beta
Taylor (1990)		derivada de una función tangente hiperbólica
Swinton y King (1991)	trigo	asimetría negativa
Moss y Shonkwiler (1993)	maíz	T. seno hiperbólico inverso
Teigen y Thomas Jr. (1995)		normal + log – normal + Chi cuadrado + otras
Babcock y Hennessy (1996)		beta
Kaufmann y Snell (1997)	maíz	
Ramirez (1997)	maíz, soja	T. seno hiperbólico inverso – asimetría negativa
Goodwin y Ker (1998)		asimetría negativa
Just y Wninger (1999)		no es posible rechazar normalidad

#2.11.- FUNCIONES DE DENSIDAD SUBJETIVAS DE RENDIMIENTOS

Las probabilidades subjetivas no han sido un tema que haya sido tratado con frecuencia en la literatura económica agraria, aunque, como en otros campos, se hayan realizado intentos de incorporar esta metodología, con frecuencia en torno a problemas de estimación subjetiva de funciones de densidad.

En la literatura se suelen citar como antecedentes los trabajos de Carlson (1970), con un trabajo sobre control de enfermedades en los cultivos, Francisco y Anderson (1972) y Bessler (1980), quien realizó una comparación entre probabilidades subjetivas de rendimientos y valores experimentales, pero para la que no obtuvo resultados positivos (aunque con datos agregados los resultados fueron mejores). Bessler había realizado en 1977 su Tesis Doctoral en el ámbito de la previsión y del razonamiento inductivo en cultivos.

Unos años después, Grisley y Kellogg (1983), construyeron distribuciones de probabilidad subjetivas de precios, rendimientos e ingresos netos a partir de la información facilitada por los agricultores de una región rural de Thailandia. El artículo procedía de la Tesis Doctoral de Grisley, y utilizaba una metodología en la que se introducía una compensación monetaria para los agricultores participantes. La técnica se derivaba de la apuntada por Francisco y Anderson (1972), y expuesta en el libro de Anderson, Dillon y Hardaker (1977), y denominada de "contaje visual". A cada sujeto se le pedía que estableciera su estimación de, por ejemplo, los rendimientos máximos y mínimos, y este rango se dividía en 5 intervalos gráficamente sobre una hoja de papel. A continuación se le daban al sujeto 25 monedas (equivalentes en conjunto a un jornal medio en la zona), que tenía que distribuir en los cinco intervalos de acuerdo con la probabilidad que estimaba que tenían de presentarse en la próxima cosecha.

La novedad del experimento de Grisley y Kellogg (1983) consistía en que al agricultor se le recompensaba con la cantidad de monedas que hubiera depositado sobre el intervalo que finalmente se manifestara. Para facilitar la decisión del agricultor, éste no debía comprometer una respuesta hasta una semana después de conocer el contenido del experimento. Para contrastar la información, debía declarar la cuenta que tenía prevista en su plan de cultivos.

En general, se encontró que los agricultores sobrestimaban los precios y subestimaban los rendimientos, lo cual podía tener sentido desde la perspectiva de la obtención de los ingresos y las rentas.

El método de recompensas utilizado por estos autores fue discutido por Knight, Johnson y Finley (1985), quienes señalaron que no se ajustaba a las familias de reglas de puntuación ("scoring rules") eficientes determinadas por la teoría. La respuesta de Grisley y Kellog (1985) señaló que lo principal era que los sujetos desearan participar y cooperar, siendo en este caso secundario si se utilizaba o no una regla de puntuación acorde con la teoría.

En los primeros años de la década de los años 80 se publicaron también los trabajos de Wright (1983) y de Pingali y Carlson (1985), estos últimos para la determinación de probabilidades subjetivas para la previsión de daños en horticultura. Estos autores determinaron distribuciones de probabilidad subjetiva para la previsión de daños en las cosechas de un grupo de horticultores de North California, utilizando la distribución triangular, por considerarla de fácil manejo. Compararon las distribuciones subjetivas con la opinión de expertos y con observaciones históricas, mediante regresión del error absoluto del daño predicho en el fruto respecto a un conjunto de variables demográficas y de manejo.

Norris y Kramer (1986) publicaron una revisión de la estimación de probabilidades subjetivas en la agricultura.

Nelson y Bessler (1989), discutieron la utilidad de las 'scoring rules' para inducir a las personas encuestadas a revelar las probabilidades que creen realmente. Concluyeron que es evidente que las personas pueden aprender a mejorar sus estimaciones de probabilidad siguiendo unas reglas de puntuación adecuadas, pero que también es verdad que los sujetos no entrenados en la utilización de esas reglas pueden contestar de forma adecuada cualquiera que sea el esquema de puntuación introducido, especialmente si no deben hacer estimaciones continuadas en el tiempo (como hacen los meteorólogos). Por tanto, en trabajos de campo en el ámbito agrario, es razonable utilizar sistemas de recompensa lineales, ya que éstos son más fácilmente comprendidos por los agricultores que las reglas normativas de puntuación.

Lindner y Gibbs (1990) compararon estimaciones subjetivas y objetivas en un contexto relacionado con el aprendizaje y adopción de variedades de trigo, pero suponiendo que debía existir una percepción del riesgo próximo a las probabilidades objetivas.

Pease (1992) señaló algunos de los inconvenientes de estimar las funciones de rendimientos a partir de estimaciones provinciales, experimentos agronómicos o datos históricos de las explotaciones, y propuso una estimación subjetiva. En su trabajo, entrevistó una muestra de 98 empresarios agrícolas de Kentucky, utilizando una técnica basada en estimaciones subjetivas de rendimientos. Encontró respuestas compatibles con una asimetría negativa para el 84% de las distribuciones históricas de maíz y de soja. Mediante un un test de normalidad (estadístico Shapiro - Wilk W), rechazó que la mayor parte de las distribuciones de maíz y de soja estén distribuidas normalmente. Discutió también que la distribución triangular represente mejor la incertidumbre subjetiva, tal como sostuvieron Pingali y Carlson (1985), puesto que sólo se estiman tres puntos de la distribución, aunque comenta que los buenos resultados conseguidos por Pingali y Carlson con dicha distribución, pueden ser debidos a que el daño ocasionado por insectos sea menos variable y más predecible que el rendimiento de los cultivos. Sin duda es uno de los trabajos que plantea con mayor claridad la necesidad de estimar los rendimientos de los cultivos a partir de estimaciones subjetivas.

Smith y Mandac (1995) realizaron un trabajo de estimación de probabilidades subjetivas y comparación con objetivas en Filipinas, con incentivos monetarios. Estos autores insistieron en la importancia de la incorporación de las "scoring rules" para favorecer una respuesta sincera.

Estos autores presentaron un resumen de los estudios empíricos sobre la influencia de la aversión al riesgo de los agricultores sobre las prácticas productivas utilizadas. Un primer grupo de estudios, realizados sobre la producción tropical de arroz, mostraron que la aversión al riesgo no impide una correcta asignación de los recursos (Roumasset, 1976; Rosegrant & Roumasset 1985; Smith & Umali

1985; Smith et al. 1989). Contrariamente, trabajos como los de Just (1974), Moscardi y de Janvry (1977), Ryan y Perrin (1973), Foster y Rausser (1991) señalaron claramente que la aversión al riesgo tenía importantes efectos sobre las decisiones de los agricultores. Esta contradicción es el punto de partida de su estudio, preguntándose hasta qué punto no es consecuencia de diferencias en la metodología utilizada para la medida del riesgo.

CAPÍTULO 3

HEURÍSTICAS Y SESGOS EN LA ASIGNACIÓN DE PROBABILIDADES

#3.1.- APORTACIONES DE LA SICOLOGÍA

El razonamiento ha sido uno de los temas de interés de la Sicología, y en esa línea, el estudio de los límites de los sujetos a la hora de realizar inferencias. Desde antiguo se sabía que las personas incurrían en falacias, a veces sutiles y con frecuencia burdas, en sus razonamientos. Una gran cantidad de evidencia indicaba que los sujetos cometen sistemáticamente errores en los procesos de establecimiento de juicios, en su razonamiento, en la selección de estrategias, etc., y que esos errores se presentan en todos los contextos: certidumbre, riesgo e incertidumbre, entre personas educadas y legos, en los ámbitos de especialización y en los generales.

Los temas de interés y los enfoques metodológicos en la Sicología han ido evolucionando a lo largo del tiempo, determinando la importancia concedida al estudio de los sesgos en el razonamiento, su naturaleza y las explicaciones.

Hasta mediados de la década de 1950 los paradigmas predominantes en la Sicología fueron el conductismo y el neoconductismo, que propugnaban análisis asociacionistas de la conducta y negaban el valor funcional de los procesos mentales. Los trabajos de J. S. Bruner, G. A. Miller, D. E. Broadbent y otros contribuyeron a que se aceptase que los procesos mentales son un objeto legítimo de estudio. Estos trabajos constituyeron el punto de partida de una nueva orientación conocida como Sicología cognitiva, de la que se suele decir que Miller, Galanter y Pribram presentaron el manifiesto fundacional en 1960.

Hasta ese momento en la investigación sicológica sobre el razonamiento había predominado el estudio de los errores en el razonamiento deductivo, y ello porque se disponía de un modelo formal de comparación: las técnicas desarrolladas en ese ámbito, como el silogismo. Los trabajos, en gran parte, examinaron distintas formas de juicios y los errores sistemáticos o sesgos producidos al violar los sujetos las reglas formales lógicas. Una breve descripción de estos trabajos se puede consultar en J. St. B.T. Evans (1989).

La aproximación general de esos trabajos es conocida como logicista o racionalista, que significa que los errores eran discutidos en un marco más general de análisis relacionado con el debate sobre la racionalidad de los seres humanos. El modelo de razonamiento subyacente en los seres humanos se suponía que era el descrito por las reglas formales que gobernaban los razonamientos óptimos, y el trabajo del sicólogo consistía en estudiar en qué medida el pensamiento real se alejaba de ese modelo, tanto en sujetos ingenuos como experimentados (el hombre de la calle y el científico), para establecer la racionalidad de los sujetos.

Al igual que los juicios en contexto de certidumbre, la Sicología abordó el estudio de aquellos realizados en contexto de riesgo y de incertidumbre. Durante los años 50 y 60 se publicaron diferentes descripciones de sesgos en los juicios probabilísticos o en condiciones de incertidumbre, que violaban algunas reglas de los modelos formales (adición, conjunción, etc. de probabilidades). Todos estos trabajos seguían también un enfoque logicista o racionalista.

Los autores racionalistas sostuvieron frecuentemente tesis que implicaban la infalsabilidad de sus teorías, entre ellas que la irracionalidad del ser humano no se puede demostrar experimentalmente. Para sostener esta tesis desarrollaron argumentos que son útiles para comprender ciertos problemas en este ámbito de estudio.

Un primer argumento tiene que ver con la distinción competencia - calidad (“competence - performance”), empleada por Chomsky en su teoría lingüística, y que fue adoptada en el estudio del razonamiento. Los errores en el razonamiento pueden ser atribuidos a factores de calidad (“performance”) antes que a aspectos de competencia lógica subyacente. Si se entiende que el interés de la discusión no se centra en la cuestión de si los seres humanos son racionales o no, sino en la calidad de los juicios, esta distinción carece de utilidad práctica, y marca perfectamente los límites de nuestro interés.

Un segundo argumento se apoya en la dicotomía representaciones mentales - procesos. Una persona podría seguir las reglas lógicas en sus razonamientos y obtener consecuencias erróneas porque éstos estarían provocados por representaciones propias del problema (la persona no razona con el problema formulado, sino con representaciones personalizadas del mismo, añadiendo o eliminando premisas, etc.)

Un tercer argumento tiene que ver con las actividades experimentales utilizadas por los sicólogos, que han sido calificadas de artificiales, arbitrarias, confusas, y no representativas del razonamiento en la vida real. Sin embargo, no se ha encontrado que en los problemas importantes la calidad de los juicios se haya visto muy afectada por la presentación de la actividad o de las instrucciones.

Un cuarto argumento esgrimido tiene que ver con los sesgos en las publicaciones, en el sentido de que se publican los casos de fallos en el razonamiento y no las inferencias correctas obtenidas experimentalmente.

Un último argumento tiene que ver con el hecho de que la orientación racionalista era tributaria de las teorías normativas que se empleaban para discriminar entre inferencias correctas e incorrectas, distinción que se apoyaba en algún criterio que frecuentemente implica medidas, y que era discutible. Como la descripción de un error depende del sistema normativo, y éste es seleccionado por el experimentador, podría suceder que esa decisión ignorase que los sujetos razonan de forma racional utilizado algún otro sistema lógico alternativo, no identificado por el sicólogo.

A finales de los años 60 se produjeron importantes cambios en la orientación de los estudios sobre el razonamiento humano, de tal forma (a) que las aproximaciones logicistas o racionalistas perdieron actualidad, aunque se mantuvieron algunas de sus teorías; y (b) se produjo una integración de los estudios sobre el razonamiento con otras investigaciones cognitivas sobre la memoria, la atención, etc.

En nuestro ámbito interés, los sesgos en el razonamiento en condiciones de riesgo e incertidumbre, seguramente no es exagerado decir que los mayores avances se han derivado de la Sicología cognitiva. Aunque se suelen citar también como antecedentes de las aproximaciones modernas los trabajos de W. Edwards y P. Meehl, es indudable que la principal fundamentación teórica y los programas de investigación se han derivado de la Sicología cognitiva.

La Sicología cognitiva investigó los procesos mentales o internos (como la percepción, la memoria, la atención, el razonamiento, etc.), el procesamiento de la información en el cerebro, las limitaciones mentales existentes en el procesamiento de la información, y la forma en las que los procesos de razonamiento se ven afectados por esas limitaciones. La Sicología cognitiva está interesada en la organización funcional de la mente y no en los aspectos orgánicos.

Una importante aportación de la Sicología cognitiva en los años 70 fue el abandono de las explicaciones logicistas (las estructuras mentales son análogas a las de un sistema formal normativo). El logicismo de las teorías racionalistas del razonamiento tenía dificultades para explicar la evidencia disponible sin abandonar una aproximación apriorista, que llevaba necesariamente a la infalsabilidad de las explicaciones que se construían. En este contexto, H. Simon y J. S. Bruner propusieron que las estrategias simplificadoras utilizadas para reducir la complejidad de los procesos implicados en la formulación de juicios, a fin de que sean tratables por la mente, explican las desviaciones observadas con respecto a los juicios óptimos, deducidos de las reglas lógicas formales. Abrieron así el camino para explicar los sesgos en función de las heurísticas que utilizan los sujetos en la formulación de los juicios y sus efectos, lo que ha permitido conseguir una integración conceptual que explica los juicios correctos y los erróneos. Esta aproximación describe el razonamiento humano como un proceso dirigido por heurísticas y por esquemas de contenido.

Inicialmente se postuló en la Sicología cognitiva que el sistema cognitivo opera mediante una serie de estadios discretos, que la información se procesa secuencial o linealmente, como en los ordenadores. Este esquema se utilizó inicialmente para desarrollar modelos de atención, memoria o específicos de tarea (verificación de frases, lectura, etc.). La acumulación de anomalías llevó en los años 70 a abandonar también la estricta analogía mente-ordenador y la concepción lineal que sugirió el procesamiento de la información. La reformulación teórica se basó en concepciones constructivistas, en las que se admite que los procesos mentales operan a veces en paralelo, o interactivamente, y que los diferentes niveles combinan funcionalmente la información de los datos y del propio sistema.

Otra importante contribución de la Sicología cognitiva tiene que ver con la relación entre la formación de juicios y la acción. Aunque no existe una concepción unitaria de los procesos mentales en la Sicología cognitiva, todos ellos tienen en común la consideración de los fenómenos mentales como agentes causales del comportamiento. Este último aspecto diferencia la Sicología cognitiva de la Sicología evolutiva de Piaget, al percibir de forma diferente el procesamiento de la información. La segunda supone que el sujeto construye su conocimiento del mundo a partir de la acción, y la primera supone la existencia de una relación inversa entre conocimiento y acción: las representaciones mentales del individuo determinan su comportamiento. El sujeto de la Sicología genética tiene una orientación epistemológica (pretende comprender la realidad), mientras que el sujeto de la Sicología cognitiva tiene una orientación pragmática (pretende controlar la realidad alcanzando metas).

H. Kelley (1967) formalizó una teoría cognitiva, como síntesis de los resultados obtenidos en la investigación sobre los procesos internos, conocida como teoría de la atribución. Esa teoría, que culminó los trabajos iniciados por F. Heider (1944) sobre la formación de juicios y la atribución de causalidad, construye modelos de sujetos considerados como sicólogos ingenuos (o científicos o estadísticos ingenuos), que infieren causas de los efectos observados, y esas causas que atribuye a los fenómenos determinan su visión del mundo, y ésta su comportamiento.

P. Meehl (1954), llevó a cabo un estudio fundamental, señalando las discrepancias entre predicciones clínicas y estadísticas. Este autor puso de manifiesto que el porcentaje de aciertos en los diagnósticos clínicos, medidos estadísticamente, era bastante inferior a la percepción que de sus aciertos tenían los expertos. La explicación construida por Meehl de los errores y de las desviaciones tiende a ser cognitiva: los modelos se construyen sobre ilusiones y no sobre desilusiones.

#3.2.-LA APROXIMACIÓN COGNITIVA

La Sicología cognitiva ha estudiado los procesos internos de percepción, memorización, formación de conceptos, pensamiento, aprendizaje, formulación de juicios, toma de decisiones, resolución de problemas, etc. Estos temas están relacionados con el procesamiento humano de la información, y las principales conclusiones en este campo han sido:

(1) que las personas tienen una capacidad limitada de procesamiento de la información; y que por ello

(2) seleccionan parte de la misma, anticipándose a lo que percibirán, lo que determina de forma importante su percepción, y

(3) la formación de juicios se apoya en algoritmos heurísticos y otras estrategias cognitivas simplificadoras, y no en reglas lógicas o en procesos de optimización.

La investigación cognitiva puso de manifiesto que procesos como la percepción o la memoria suponen tanto actividades de selección como de interpretación, de construcción y de reconstrucción.

La percepción no está determinada exclusivamente por la capacidad sensorial o los procesos sensoriales, sino que también interviene el conocimiento que posee el sujeto y las expectativas generadas por ese conocimiento. Lo que se percibe está influido por lo que se sabe, como puso de manifiesto, entre otros, Bruner y Minturn (1955) con experimentos de figuras ambiguas (p.e. una figura que puede ser interpretada como una B o como un 13).

La información se procesa siguiendo dos grandes esquemas: (1) de abajo a arriba o guiado por el estímulo (por ejemplo, la conducción cuando las condiciones de visibilidad son buenas) y (2) de arriba abajo o guiado conceptualmente (cuando nos apoyamos más en la información interna que en la sensorial, por ejemplo, al conducir en condiciones de escasa visibilidad).

La percepción no supone solamente la habilidad de discriminar, sino también la formación de hipótesis, la aplicación de principios organizadores y la toma de decisiones, es decir, una combinación de procesos sensoriales y de decisión, muchos de ellos no sometidos a la consciencia. Las decisiones implicadas en la percepción varían según el contexto, por ejemplo en función de la prudencia que exija una circunstancia concreta (sesgo de respuesta). De esta forma, la realidad, tal como la conocemos, es en parte una construcción individual.

En el caso de la memoria, las personas pueden recordar mejor el significado de los hechos que los detalles de los mismos, influyendo el significado sobre los detalles recordados. Incluso cuando se recuerdan bien los detalles, este recuerdo no es permanente, sino que puede alterarse (así cuando un sujeto presencia un accidente de automóvil y posteriormente es interrogado sobre si el vehículo se detuvo antes o después del árbol, inserta un árbol en el recuerdo, incluso si no lo había en la realidad).

Para explicar el funcionamiento de la memoria se ha propuesto un modelo de tres almacenes. El almacén sensorial recibe la información de los sentidos y la retiene un segundo mientras decide a qué se le va a prestar atención. Aquello a lo que se presta atención se transfiere al almacén a corto plazo y el resto de la información se pierde rápidamente. El almacén a corto plazo tiene una capacidad aproximada de siete elementos, y la información vieja es sustituida rápidamente por la nueva. A lo que se continúa prestando atención, pasa al almacén a largo plazo.

Los recuerdos similares, sin embargo, acaban confundiéndose o interfiriendo entre sí cuando se intenta recuperarlos. Por ello, los cumpleaños acaban confundiéndose entre sí, y lo que se recuerda tiene relación con la significación de los cumpleaños, más que con lo que ocurrió exactamente cuando se cumplieron 5, 10 o 15 años. Los significados generales son más importantes que los detalles, excepto que algo provoque que esos detalles sean especialmente relevantes (cuando se cumplieron los cuarenta años, la fiesta sorpresa, etc.).

Uno de los componentes principales del pensamiento son los conceptos, las abstracciones que resumen lo que conocemos. Se forman tanto a partir de la experiencia con objetos y situaciones como del contacto con símbolos o signos de esos objetos o situaciones. Podemos aprender de un cultivo tanto mediante su práctica como a partir de lecturas o narraciones sobre el mismo.

La teoría del prototipo propone que los conceptos se organizan en torno a prototipos. Un prototipo es un conjunto de rasgos característicos. Cuanto más difiere un objeto de su prototipo, más difícil resulta aprenderlo, recordarlo y reconocerlo. Esta teoría, sin embargo, tiene dificultades para explicar el uso de conceptos abstractos como “talento”, para lo que no existe un prototipo evidente.

Los conceptos no necesitan ser construcciones precisas y fijas, sino eficaces para pensar, razonar y comunicar. Se supone que las reglas de la lógica y de la gramática permiten realizar esas actividades de forma más eficaz, pero el éxito de estas actividades depende de otros muchos determinantes.

No siempre es útil pensar. Actividades que inicialmente exigían pensar cuidadosamente (como aprender a conducir), posteriormente, con una práctica prolongada, se convierten en automáticas. Esas actividades se relegan al subconsciente y se pueden realizar sin pensar, aunque sea a cambio de cometer algún error. Estos errores o fallos cognitivos aumentan con la fatiga o la confusión, y se pueden reducir pensando de nuevo sobre esas actividades.

Mediante este mecanismo muchas de las decisiones que se toman son inconscientes, y las experiencias de “ceguera mental” indican que es muy útil tomar consciencia del pensamiento cuando se ha de realizar una elección difícil, cuando se presentan acontecimientos que no se pueden tratar de forma automática o sentimientos inesperados.

Un componente importante del pensamiento es el razonamiento, que supone utilizar la información disponible para resolver problemas. Hay dos clases de razonamiento, el inductivo y el deductivo. El razonamiento inductivo significa obtener una conclusión general de un conjunto de datos; el deductivo consiste en obtener una conclusión de un conjunto dado de afirmaciones iniciales o premisas. Ambos tipos de razonamiento están expuestos a errores o falacias bien conocidos.

El hecho de que el pensamiento esté sometido a sesgos ilógicos no quiere decir que éstos no sean útiles. Por ejemplo, si un amigo nos anuncia que después del partido de fútbol irá al bar a celebrar el triunfo de su equipo, y conocemos posteriormente por la radio que ese equipo ha perdido el partido, de esa información no se deduciría que nuestro amigo haya ido a su casa después del encuentro. Pero nuestro sesgo ilógico nos llevará a buscarlo precisamente en su casa después de la derrota de su equipo.

Razonar es una tarea compleja y difícil que supone una gran carga para la memoria, por lo que en la práctica se utilizan heurísticos o reglas toscas, pero eficaces, para guiar el pensamiento. De esta forma los heurísticos ayudan a resolver problemas complejos.

#3. 3.- RESOLUCIÓN DE PROBLEMAS

Por “resolución de problemas” se entienden aquellas actividades que exigen procesos de razonamiento complejos, y no una actividad asociativa y rutinaria. Simon (1978) definió que una persona se enfrenta a un problema cuando acepta una tarea, pero no sabe de antemano cómo realizarla.

En el estudio de la resolución de problemas existen tres tradiciones importantes: la Gestalt, la tradición conductista y la cognitiva.

La Gestalt insistió en los aspectos holísticos en la resolución de problemas, como la reestructuración del problema y las nuevas formas de combinación de elementos. Contrariamente a la opinión del asociacionismo, estimaba que la resolución de los problemas no es un proceso mecánico basado en la experiencia pasada, sino que hace surgir algo nuevo (pensamiento productivo), una configuración perceptiva alcanzada bruscamente, mediante reorganización súbita (“insight”).

La tradición conductista estudió la resolución de problemas desde la perspectiva de dividirlos en procesos más sencillos de respuestas de aprendizaje a estímulos y conseguir incrementalmente una solución.

La tradición cognitiva o del procesamiento de la información ha dominado la investigación reciente, especialmente desde que Newell y Simon publicaron **Human Problem Solving** (1972).La teoría se aplicó en problemas bien estructurados (problemas de ordenación o de transformación). Otros problemas más “creativos” y peor definidos se excluyeron inicialmente (Simon, 1978).

Uno de los temas centrales en el ámbito de la resolución de problemas es cómo influye la experiencia pasada, la información almacenada en la memoria, ya que los problemas se resuelven más fácilmente cuando se tiene mayor experiencia sobre los mismos (“efecto de transferencia positiva”).

Expertos y novatos parten de un período de incubación en el que no se piensa en el problema. Cuando se identifica una estrategia para resolver un problema, puede que necesite alguna habilidad para aplicarla y cierta habilidad del razonamiento para evaluar el progreso que se obtiene. Los expertos resultan mejores en el reconocimiento de patrones, en el recuerdo de las reglas adecuadas, y en la eliminación de reglas que no llevan a ninguna solución.

Los expertos atacan la resolución de los nuevos problemas provistos de un "conjunto mental" que les ahorre redescubrir las reglas ya conocidas para la resolución de problemas. Pero cuando encuentran un problema que no se resuelve mediante aquellas reglas, el experto también falla en su resolución ("fijeza mental").

La "fijeza mental" es un tipo de "conjunto mental" que impide la eficaz resolución de problemas. En este caso se piensa en los objetos en términos de sus funciones. La creatividad es la vía de encontrar nuevas funciones (por ejemplo un sobre para la correspondencia y para el envasado de azúcar). Se suele hablar de un pensamiento divergente o creativo, como aquel en el que se exploran las ideas con libertad, generando muchas soluciones. El pensamiento convergente parece seguir unas pautas que llevan a una única solución correcta del problema.

En el estudio de la resolución de problemas se ha prestado una gran atención al llamado "sesgo confirmatorio", uno de los más universales y persistentes de los identificados. Todos los sujetos tienden a subrayar la evidencia favorable a sus explicaciones o hipótesis, y a descuidar la evidencia desfavorable.

Este sesgo tiene una gran influencia sobre el aprendizaje. Los experimentos muestran que las personas aprenden de la experiencia de una forma selectiva y fortuita, y que una vez que se han formado una idea están poco dispuestos a abandonarla, magnificando cualquier prueba que la confirme, sin importar lo artificial que pueda ser. En general las pruebas disconfirmadoras no tienen efecto y las confirmadoras tienen un efecto fortalecedor.

Gran parte de los experimentos consisten en plantear problemas - tipo relacionados con el razonamiento inductivo, cuya resolución supone falsar una hipótesis (mediante la generación de evidencia incompatible con la misma). Sin embargo, los sujetos intentan verificar la hipótesis de forma sistemática, sin percibir que se puede realizar un gran número de verificaciones de una hipótesis incorrecta. (Klayman y Ha, 1987). Este es un sesgo habitual entre los científicos, quienes sistemáticamente intentan verificar sus hipótesis.

Una consecuencia es que la presentación de los dos lados de una cuestión no reduce las diferencias entre quienes sostienen opiniones opuestas, antes al contrario, las amplía. Las personas tienden a seleccionar y a elegir pruebas, creyendo sólo lo que se ajusta a sus opiniones (reforzando su posición) y rechazando lo que no se ajusta (Lord, Lepper, Ross, 1979).

Einhorn (1984, 19) ha subrayado la importancia de estudiar el aprendizaje para comprender las heurísticas y el comportamiento a la hora de realizar una elección. Piensa que se ha concedido demasiada importancia a factores ambientales (lo que Ross denominó error fundamental de atribución), estudiando especialmente la actividad para comprender el comportamiento, y descuidando la propia estructura de la actividad. Sin conocimiento de la estructura de la tarea, la retroalimentación del resultado puede ser irrelevante o incluso perjudicial en la corrección de las heurísticas pobres. Más aún, una retroalimentación positiva sin conocimiento de las actividades puede ocultarnos la pobreza de nuestras reglas, ya que no existe una motivación para investigar cómo se obtuvo el resultado.

Si se conoce bajo qué condiciones se aprendió una regla podemos predecir cuando se utilizará (por ejemplo, si se aprende una regla en situaciones en las que se dispone de poco tiempo para tomar la decisión, podremos predecir que esa regla será utilizada en esos casos). De forma semejante, simulando las condiciones en que se produjo el aprendizaje se puede ayudar a las personas a modificar las reglas de una forma eficaz (Fischhoff, Slovic, Lichtenstein, 1978; 1980).

#3.4.- ASIGNACIÓN DE PROBABILIDADES

La predicción de una cantidad desconocida o la evaluación de la probabilidad de un suceso incierto es un proceso complejo, que comprende una interpretación del problema, una búsqueda de información relevante y la selección de una respuesta apropiada. Se ha desarrollado una extensa literatura sociológica experimental sobre cómo se perciben, procesan y evalúan las probabilidades de sucesos inciertos en diversas circunstancias (aprendizaje de la probabilidad, utilización intuitiva de los conocimientos estadísticos, toma de decisiones en contexto de riesgo e incertidumbre, etc.), y sobre la formación de creencias y opiniones.

Las principales líneas de interés han sido:

- comparación de los juicios probabilísticos con los resultados derivados de los modelos formales;

- comparación de las probabilidades asignadas subjetivamente con las objetivas correspondientes;

- factores que afectan a la asignación de las probabilidades;

- consistencia de las asignaciones de probabilidad.

Algunos trabajos de referencia obligada son Edwards (1968, 1984, 25), Kahneman y Tversky (1984, 30), Slovic (1972), Slovic, Fischhoff y Lichtenstein (1977), Tversky y Kahneman (1984, 1).

Inicialmente gran parte de la investigación sobre sesgos en los juicios probabilísticos se centró en la comparación con los resultados correspondientes de los modelos formales. Los distintos sesgos experimentales que se identificaron en el razonamiento probabilístico culminaron en los años 60 en una línea especialmente productiva, relacionada con los juicios que implicaban probabilidades a priori, y que podían corregirse mediante información adicional. Esta experimentación sociológica en el marco bayesiano, aunque fue practicada por un gran número de autores, está asociada al nombre de W. Edwards, quien introdujo la estadística bayesiana en el ámbito de la Sicología. El teorema de Bayes le permitía comparar las inferencias realizadas en condiciones de incertidumbre con el correspondiente modelo normativo, a fin de investigar la lógica seguida en la construcción sicológica de las inferencias.

Los distintos autores, entre ellos W. Edwards, realizaron un gran número de trabajos experimentales en los que comparaban el comportamiento humano con los resultados esperados aplicando el teorema de Bayes, que se consideraba la regla formal óptima para revisar las opiniones (o probabilidades subjetivas) en función de la información adicional disponible. La conclusión más general señalaba que los sujetos son muy conservadores a la hora de estimar las probabilidades posteriores, que concluyen en sus inferencias un nivel de certeza inferior al que le garantizan los datos, o lo que es lo mismo, que la información adicional tiene menor impacto del que debiera (Phillips y Edwards, 1966; Pitz, Downing y Reinhold, 1967; Kozielecki, 1970). Un problema típico puede tener la siguiente forma:

Se dispone de dos urnas con fichas, una contiene 300 rojas y 700 azules (urna A), y la otra 700 rojas y 300 azules (urna R). Se elige una de las urnas lanzando una moneda al aire, y de ella se extraen 12 fichas, de las cuales 8 son rojas y 4 azules. ¿Cuál es la probabilidad de que se haya elegido la urna R?

Sin hacer operaciones, la mayoría de la gente asigna una probabilidad entre 0,7 y 0,8. Pero la probabilidad calculada es 0,97. Ni siquiera personas familiarizadas con el teorema de Bayes suelen hacer una estimación próxima a la probabilidad calculada.

Edwards (1968, 1984, 25) avanzó la hipótesis de que los hombres siguen reglas bayesianas a la hora de establecer las probabilidades subjetivas, pero que fallan a la hora de apreciar el impacto de la evidencia, y como consecuencia se manifiestan conservadores en sus conclusiones (este sesgo es explicado actualmente mediante la heurística de anclaje). Las personas perciben adecuadamente cada dato, y son capaces de entender su significado individual en un diagnóstico, pero no tienen la capacidad de combinar varios datos para revisar de forma adecuada sus opiniones.

Paradójicamente, si en la inferencia bayesiana unietápica los sujetos se comportan conservadoramente, en las inferencias multietápicas (donde un sujeto debe realizar una serie de inferencias, una por etapa, y el resultado de cada etapa es utilizado como punto de partida de la inferencia siguiente), se comportan de forma absolutamente contraria. Aunque sería lógico esperar la acentuación de la actitud conservadora, al comportarse de forma conservadora en cada etapa, los resultados señalan un comportamiento contrario: los sujetos muestran mayor confianza en sus inferencias de las que autoriza el modelo formal MBT (Modified Bayes Theorem; Gettys y Willke, 1969).

También es sorprendente que cuando se realizan juicios técnicos que suponen un gran número de alternativas, aparece un exceso de influencia de la información adicional (Kozielicki, 1970). Las explicaciones iniciales se basaban en que las personas tienen dificultades para conocer el valor de diagnóstico de la información adicional (Beach, 1966; Lathrop, 1970), o en que conociéndola, tienen dificultad para combinar de forma adecuada las probabilidades (Phillips, 1965; Edwards, 1968; Peterson y Phillips, 1966; DuCharme y Peterson, 1968).

Violaciones semejantes de reglas formales de la estadística se habían descrito anteriormente, como la adición de probabilidades o la conjunción, y se tendía a explicarlas como limitaciones de la racionalidad de los sujetos. Así, por ejemplo, se señalaba que los sujetos perciben perfectamente que la probabilidad es un número comprendido entre cero y uno, y que expresa la fuerza con que una persona cree que una proposición es verdad. Sin embargo, cuando juzga varias probabilidades aditivas, la aditividad tiene consecuencias tan exigentes que pocas personas serían capaces de aceptarlas.

Existe un gran número de resultados de la experimentación sicológica que muestran que las predicciones de los sujetos se apartan de cualquier regla formal estadística. Las personas tienden a ignorar la información previa o la probabilidad a priori, la influencia del tamaño de la muestra, los fenómenos de regresión estadística observados en el mundo real, etc .

Conocer cuál es la exactitud de las predicciones subjetivas aporta poco a la hora de determinar las características esenciales de los procesos mentales que siguen los sujetos para establecer sus juicios. Si los sujetos no obtienen las consecuencias que se derivan del teorema de Bayes, y al mismo tiempo son insensibles al tamaño de las muestras en sus inferencias, o sistemáticamente violan un gran número de reglas formales estadísticas, no parece razonable concluir solamente que el hombre es conservador, sino más bien que no es bayesiano. La tendencia a elaborar teorías de competencia para salvaguardar la racionalidad del ser humano, en línea con la concepción del mismo como un “estadístico intuitivo” (Piaget, Ihhelder, Edwards...), originaba que se formularan explicaciones en las que la estadística bayesiana era una descripción adecuada de los procesos predictivos de ese estadístico intuitivo o ingenuo, pero sin posibilidades de falsar esa hipótesis, y por tanto llevando a la teoría a un callejón sin salida, a una mera descripción de la colección de errores.

La ruptura de esa inercia logicista fue la gran aportación de autores como H. Simon, J. S. Bruner, y sobre todo, de D. Kahneman, P. Slovic, y A. Tversky. Estos últimos autores cambiaron la concepción anterior de las personas como “estadísticos intuitivos”, al sostener que en lugar de cómputos estadísticos, la predicción ingenua se basa en heurísticas o reglas que se aplican espontáneamente como parte de un proceso de evaluación natural desarrollado rutinariamente en la percepción y comprensión de la información. De esta forma, las probabilidades subjetivas, que juegan un papel central en la vida de las personas a la hora de tomar decisiones, construir explicaciones o concluir sobre relaciones causales, actividades todas que dependen de juicios sobre la verosimilitud de algunos sucesos inciertos implicados, se asignan y revisan de forma intuitiva o subjetiva antes que siguiendo modelos formales de cálculo de probabilidades.

En la comparación de las probabilidades subjetivas asignadas por los sujetos con las correspondientes objetivas, un resultado clásico, formulado por Preston y Barratta (1948) y comprobado por Griffiths (1949), Vail (1952), Attneave (1953) y Sprowls (1953), señalaba que los sujetos tienden a sobrestimar las probabilidades bajas o pequeñas y a subestimar las altas. La técnica empleada fue la de subasta de apuestas.

La interpretación de este resultado, sin embargo, no esta exenta de discusión. A. Tversky (1967), en estudios empíricos dirigidos a medir conjuntamente la probabilidad subjetiva y la utilidad, planteó juegos en los que se podía ganar x u.m. en caso de que ocurriera el suceso E y nada si no ocurría, donde el sujeto debía determinar la cantidad mínima cierta por la que estaría dispuesto a ceder su derecho a jugarlos. Los resultados parecían indicar que los sujetos guiaban sus decisiones por el principio de utilidad esperada subjetiva (SEU), donde las cantidades monetarias x_j se sustituyen por sus utilidades $u(x_j)$ en el calculo de la esperanza matemática (J.M. Hampton, P.G.Moore, H. Thomas, 1973; Coombes, 1964; Davidson y Suppes , 1956).

En problemas de este tipo existen demasiados parámetros implicados para que se pueda extraer una conclusión indiscutible. En cualquier caso, la determinación de las relaciones entre probabilidades objetivas y subjetivas carece de interés, ya que la probabilidad subjetiva refleja exclusivamente las creencias de un sujeto establecidas con la información disponible en el momento de plantear la decisión.

Más interesantes son las investigaciones sobre la forma en que las personas establecen las probabilidades subjetivas y cómo influyen sus creencias u otros factores instrumentales (la forma de estar planteado el problema, el nivel de dificultad, la interacción posible, el tiempo disponible para resolverlo, limitaciones para efectuar los cálculos necesarios, disponibilidad u olvido de alguna información esencial, etc.), lo que se consideraba que podría ser útil para la formulación de programas de entrenamiento (Vail, 1952; Howell, 1967; Martin, 1969; Strub, 1969). La comparación entre la forma de asignar las probabilidades los sujetos expertos y los no expertos en un determinado campo no siempre fue favorable a los primeros, lo que llevó a conjeturar inicialmente que la dificultad se encontraba en la traducción del conocimiento disponible en probabilidades (Cohen et al 1956, 1957). Sin embargo, esta vía de escape volvía a contradecir la noción de probabilidad subjetiva como expresión de una opinión.

Es importante entender que no toda respuesta que contradiga un hecho o una regla aceptada es un error de juicio, ya que la diferencia puede estar causada por malentendidos por parte del sujeto sobre la cuestión o un error de juicio del experto sobre la respuesta. La descripción de una respuesta como errónea supone un conjunto de supuestos sobre la comunicación experimentador - sujeto, y se deberían evitar tanto las interpretaciones estrictas, que clasifiquen las respuestas razonables como errores, como las interpretaciones caritativas que intenten racionalizar cualquier respuesta (D. Kahneman y A. Tversky, 1984, 34).

Igualmente, la calidad del pensamiento de un sujeto no se puede juzgar simplemente por su resultado, ya que existe una tendencia humana a la “sabiduría ex-post”, a ver clara una opción una vez que se ha manifestado. Las malas consecuencias no se derivan solamente de pensamientos deficientes, ni las buenas decisiones de un razonamiento correcto. Los sesgos inducidos por una propiedad lógicamente irrelevante de una actividad pueden producir, en ocasiones, un resultado coincidente con el determinado por las reglas lógicas formales. Como consecuencia, no se puede simplemente decidir en qué situaciones las personas son competentes a la hora de realizar una inferencia mediante su habilidad para responder a una clase de problemas que incorporen el principio estudiado. Es necesario estudiar un conjunto variado de problemas a fin de identificar la naturaleza de los procesos mentales implicados y la amplitud y naturaleza de los sesgos que se manifiestan (J. St. B.T. Evans, 1989).

#3.5.- EL CONCEPTO DE HEURÍSTICA

A comienzos de la década de los años 70, Kahneman, Slovic y Tversky (1984) señalaron que para realizar juicios en condiciones de incertidumbre (establecer la probabilidad o la verosimilitud de un suceso aleatorio o predecir un valor), las personas no siguen los principios del cálculo de probabilidades o de la teoría estadística, aunque dispongan de evidencia estadística, sino que se limitan a utilizar un conjunto de reglas de inferencia muy generales (heurísticas) o principios heurísticos, descritos a veces como “atajos” (“shortcuts”). De esta forma una heurística es una regla simple, en comparación con las reglas formales de la teoría de la decisión, que resultarían demasiado complejas y pesadas de aplicar de forma intuitiva (Tversky y Kahneman, 1975; Anderson et al, 1977). Este marco conceptual, conocido como “heurísticas y sesgos", ha predominado en el estudio de la inferencia estadística desde los años 70.

Los sujetos asignan probabilidades de forma semejante a como se estiman otras cantidades, como las distancias o los tamaños, es decir, utilizando indicios o claves identificados en la percepción. Los juicios se forman con datos de validez limitada, que son procesados con arreglo a reglas derivadas de la heurística (p.e. la proximidad se estima por la claridad con que se percibe un objeto, y de esta forma la distancia normalmente se sobrestima cuando la visibilidad es mala y se subestima cuando la visibilidad es buena). Esquemáticamente el proceso de formación de juicios sería:

Datos de validez limitada + Principios heurísticos → Juicios

Las personas utilizan esas reglas u heurísticas para reducir las tareas mentales difíciles a otras más simples. Las heurísticas responden a estrategias adaptativas en los ámbitos ecológicos y sociales en que aparecen, y por tanto son válidas y eficientes en muchas circunstancias, pero también conducen a numerosos errores sistemáticos o sesgos cuando se emplean para juzgar fenómenos inusuales y relaciones complejas.

Un sesgo es una tendencia sistemática a tener en cuenta factores irrelevantes o a ignorar otros relevantes. Los sesgos cognitivos inducidos por la utilización de heurísticas para el establecimiento de los juicios, son un fenómeno bien establecido, sistemático y de difícil corrección. Los errores o sesgos no se pueden atribuir a cuestiones relacionadas con la motivación de los sujetos (espejismos o distorsiones producidas en el juicio por ilusiones, recompensas o penalizaciones), ya que se repiten incluso cuando se anima a los sujetos a obtener soluciones correctas y se les recompensa por ello. Igualmente, afectan a las personas expertas y no expertas. Aunque un experto estadístico no incurrirá en errores elementales, como la falacia del jugador, sus juicios intuitivos se ven afectados por falacias similares, aunque de una forma intrincada y en problemas poco transparentes (por ejemplo, es bastante frecuente que olvide tener en cuenta las probabilidades a priori a la hora de establecer un juicio).

Esta explicación supone que existen mecanismos mentales o procesos síquicos que imponen restricciones a la forma de razonar y de resolver los problemas, que se manifiestan en una racionalidad imperfecta a través de los sesgos que comparten los sujetos. El pensamiento científico, como el vulgar, parte de representaciones mentales “naturales” (que no lógicas), tales como las imágenes mentales, y resuelve problemas por analogía, inducción, atribución de causalidad, etc. Pero por encima de estas manifestaciones existe una competencia lógica general, que permite evaluar el rendimiento del razonamiento humano comparándolo con los modelos formales correspondientes. De esta forma no se necesita el supuesto de competencia lógica, ya que los principios lógicos son irrelevantes para explicar los procesos mentales implicados en el razonamiento cotidiano.

Se han aducido un gran numero de razones para explicar por qué, en los contextos de riesgo e incertidumbre, las personas han desarrollado heurísticas o atajos en la resolución de sus problemas antes que generalizaciones a partir de las leyes estadísticas. Un argumento interesante es que muchas leyes de la probabilidad no son ni intuitivas ni fáciles de aplicar. Los principios estadísticos no se aprenden con la experiencia diaria, debido a que las instancias relevantes no están codificadas de forma adecuada. Por ejemplo, cuando una persona lee, no presta atención a la longitud individual de cada palabra, por lo que no puede apreciar que en el texto dos líneas difieren más en la longitud media de las palabras que en el caso de dos páginas sucesivas. De esta forma las personas no aprenden la relación entre el tamaño de la muestra y la variabilidad, aunque los datos sean abundantes.

Las personas podrían aprender a calibrar sus juicios registrando la proporción de sucesos que realmente ocurren de entre aquellos a los que asignó probabilidades idénticas. Sin embargo, no es habitual agrupar los sucesos por la probabilidad que se les atribuye. En ausencia de este agrupamiento, es imposible para un individuo descubrir, por ejemplo, que solamente el 50% de las predicciones a las que asignó una probabilidad del 90% o mayor realmente ocurren.

Muchos de los sujetos no formados estadísticamente tienden a aceptar una regla cuando se le explica de forma abstracta, pero cuando entra en conflicto con sus impresiones o representaciones, la niegan. La formación estadística no protege en caso de formulaciones poco claras. El error de conjunción es un error de aplicación en que incurren sistemáticamente individuos formados en técnicas estadísticas.

La intuición estadística, como han puesto de manifiesto varios trabajos, está altamente correlacionada con el nivel de entrenamiento estadístico. Las intuiciones estadísticas varían con la inteligencia, la experiencia y la educación. Como otro tipo de conocimiento, lo que es intuitivo para los expertos no lo es para el resto de la gente. Sin embargo, algunos resultados estadísticos (coincidencia de cumpleaños en un grupo de 23 personas o el cambio de racha en el lanzamiento de monedas) son contraintuitivos incluso para personas formadas en la teoría de la probabilidad.

El pensamiento científico manifiesta ventajas de tipo educativo (método científico) que incrementan la capacidad de tratar la información. El método evita y corrige algunos tipos de errores, y amplifica y optimiza los procesos mentales. Pero aunque se ha encontrado que los modelos estadísticos son superiores a las reglas heurísticas en muchos problemas, los primeros apenas se emplean en un gran número de situaciones prácticas. Así, por ejemplo, en la practica clínica, los modelos estadísticos no se consideran útiles para la predicción. Ello es debido a que los especialistas en un campo son más hábiles seleccionando y codificando información que integrándola. Los modelos estadísticos pueden integrar información de una manera óptima, pero siempre es una persona (juez, clínico, etc.) la que elige las variables. Más aún, es el juez humano el que conoce la relación direccional entre las variables predictivas y los criterios de interés o el que puede codificar las variables de tal forma que haya una relación clara.

Dada la limitada capacidad del ser humano para procesar la información, necesita estructurar claramente su entorno (mediante componentes de simetría, cierre, proximidad, continuidad, destino común, etc.), y en ese proceso frecuentemente intenta soslayar la presencia de lo aleatorio.

Aunque las personas no se sienten cómodas con la incertidumbre, han aprendido a convivir con ella y a elegir en su presencia. La adaptación se ha producido mediante el desarrollo de estrategias heurísticas, pero también de tecnologías e instituciones para acomodarse a los efectos de la incertidumbre. Los pronósticos meteorológicos, los paraguas de bolsillo, las normas de construcción o los seguros son algunos de los productos de esa adaptación. El problema es que no siempre las heurísticas y las estrategias desarrolladas son tan eficaces como se supone.

Cuando en un proceso de razonamiento, en el que intervienen variables en contexto de riesgo o de incertidumbre, se comete un error o sesgo, frecuentemente se es incapaz de detectarlo, y muchas ocasiones es imposible aprender incluso de las situaciones más desastrosas. También es normal que en muchos tipos de problemas, especialmente los relacionados con decisiones sociales y políticas, se proceda como si se viviera en un mundo perfectamente predecible, siendo así que la incertidumbre asociada con los análisis en estos ámbitos es mucho mayor que los descritos en otros campos del conocimiento, como en las ciencias naturales. Posiblemente en ello tenga que ver que existen también factores culturales que tienden a enmascarar la aleatoriedad. Así, por ejemplo, Savage señaló la confusión en el lenguaje occidental entre certidumbre y creencia, que impulsa a los sujetos a razonar y a actuar como si estuvieran en un contexto de certidumbre.

De esta forma la explicación general de los sesgos se basa en procesos selectivos de tratamiento de la información. Las personas consideran la información de un problema de forma selectiva, determinada por los procesos de formación de las representaciones mentales de la información disponible en el problema, y por ello las personas a menudo consideran cuestiones lógicamente irrelevantes del problema tratado e ignoran otras relevantes.

Un pensamiento selectivo es, paradójicamente, un pensamiento inteligente. La reducción de la masiva información disponible en base a reglas heurísticas apropiadas es una característica del comportamiento inteligente (Newell y Simon, 1972). Los procesos cognitivos son necesariamente selectivos, y la selección es fundamental para la inteligencia, como consecuencia de la limitación de la capacidad cognitiva.

J. St. B.T. Evans (1989) propuso un esquema bietápico del proceso de razonamiento, con una primera etapa denominada como “proceso heurístico” y una segunda como “proceso analítico”. Se partiría de la información asociada a un problema, y en la etapa heurística se obtendría una representación de las características consideradas relevantes del problema. En la segunda etapa, mediante el proceso analítico, se obtendrían finalmente las inferencias o los juicios. En este contexto, el termino heurística se debe entender como heurística representacional, para no confundirla con la descrita por Kahneman y Tversky.

El esquema heurístico - analítico del razonamiento, según J. St. B.T. Evans (1989), permitiría explicar de forma más adecuada por qué los sujetos que entienden las leyes estadísticas y las aceptan, en bastantes situaciones incurren en sesgos que las violan. La primera etapa heurística, de selección de la información relevante, estaría expuesta a poderosas influencias no lógicas. Así, por ejemplo, los procesos de comprensión lingüística pueden dar origen a poderosas influencias no lógicas en el razonamiento verbal, estando esta circunstancia en el origen de ciertas clases de sesgos.

En lo que sigue se expondrán las tres clases generales de heurísticas (representatividad, disponibilidad y anclaje) identificadas por A. Tversky y D. Kahneman (1975), principalmente siguiendo el conjunto de trabajos recopilados en Kahneman, Slovic y Tversky (1984).

#3.6.- HEURÍSTICA DE REPRESENTATIVIDAD

La asignación de probabilidades subjetivas, la realización de predicciones intuitivas, y otros juicios en condiciones de incertidumbre, se realizan frecuentemente a partir de expectativas e impresiones fundadas en la representatividad.

El heurístico de la representatividad es uno de los procedimientos que se pueden utilizar para recuperar, interpretar y evaluar la información de la memoria a largo plazo, y es utilizado con frecuencia porque el conocimiento conceptual muchas veces está organizado y se procesa en forma de prototipos (los ejemplares más perfectos), estereotipos (concepciones muy simplificadas) o ejemplos ilustrativos.

Al realizar una predicción, se seleccionan y ordenan los resultados (o estados de la naturaleza) que se creen compatibles con los rasgos esenciales de la evidencia actual (con los datos disponibles). Esa ordenación se realiza mediante la representatividad: lo que importa es el grado en que un resultado es representativo de la evidencia. Por ejemplo, dada la descripción de los rasgos de personalidad de un sujeto (evidencia), las diferentes profesiones que se le pueden atribuir (resultados) se ordenan por el grado en que cada profesión se considera representativa de esa descripción de la personalidad (no se imagina a una persona tímida como relaciones públicas, por ejemplo).

Mediante la representatividad se evalúa el grado de correspondencia o similitud entre la evidencia X y un modelo M (p.e. una muestra y una población), para determinar la pertenencia a una categoría o realizar la predicción de resultados. El modelo M es un conjunto, una clase o una categoría y X es un elemento de la misma, que en caso de pertenencia categorial puede ser único y monoatributo, o único multiatributo, o varios elementos multiatributo, y en el caso de la predicción de resultados, un efecto. Cada una de estas posibilidades da origen a diferentes casos.

Un primer caso es cuando X es un valor individual monoatributo de una variable definida en M (p.e. los ingresos de los profesores de universidad). El valor más representativo será la media, la moda o la mediana de la distribución, y la relación de representatividad estará determinada por lo que se conoce sobre esa distribución de frecuencias.

Un segundo caso se presenta cuando X es un ejemplar o elemento multiatributo de M. La representatividad está determinada por la similitud del ejemplar X a otros ejemplares prototípicos o estereotípicos del conjunto M. Así, la mayor parte de las personas estarían de acuerdo en que Fleming es más representativo de los investigadores que Keynes, porque este juicio no se basa en la frecuencia, sino que refleja el grado en el que los distintos atributos de un investigador (campos, temas, ideas…) son centrales en la concepción de la investigación científica. Un ejemplar es representativo de una categoría si posee características esenciales de los miembros de la misma, pero también si carece de características esenciales que no son compartidas por los otros miembros. Por ser un científico social, Keynes no es percibido como un investigador tan puro como Fleming (como el científico natural).

Entre un elemento típico o modal y un elemento prototípico puede haber una gran diferencia. El prototipo de mujer francesa, por ejemplo, es parisina, joven, estilizada, elegante, etc.; sin embargo, la mujer francesa típica es seguramente de mediana edad, de provincias y rechoncha.

Un tercer caso es aquel en el que X son varios elementos multiatributo de M (p. e. un subconjunto o muestra de M). La representatividad está determinada por la similitud de la muestra a la población (los estudiantes de Astronomía son menos representativos que los de Economía en el conjunto de los universitarios).

Los criterios de representación no son los mismos para un subconjunto que para un ejemplar individual, porque un ejemplar sólo puede representar la tendencia central de los atributos, mientras que un subconjunto puede también representar la variabilidad de un intervalo. Un hombre cuyo peso, altura y edad coincidan con los valores medios de la población es, claramente, representativo de toda la población, pero un grupo de 100 hombres con las mismas características podría no representar la variabilidad de los atributos.

Cuando la clase M está constituida por diferentes grupos (“clusters”), cada uno de ellos con una variabilidad relativamente baja en relación a la existente entre los distintos grupos, cada grupo puede ser considerado un ejemplar más que un subconjunto, y entonces el segundo caso discutido puede considerarse un caso particular del caso examinado en tercer lugar.

En el cuarto caso, M es un sistema (causal) o causa y X es una (posible) consecuencia o efecto. Este caso difiere de los anteriores en que M no es un conjunto de objetos o ejemplares, sino un sistema que produce varios efectos. Por ejemplo, M puede ser la economía de un país y X la tasa de inflación, o M puede ser una persona y X un acto de la misma (divorcio, suicidio, elección de profesión, etc.). En este caso la representatividad está determinada por la asociación frecuente de X con M (p.e. la fiebre alta en caso de neumonía) o por la creencia (válida o no) de que existe una relación causal entre ambos (la pena de muerte reduce los asesinatos).

La representatividad es una relación direccional, en el sentido de que se dice que una muestra es más o menos representativa de una población o que un acto es más representativo de una persona, y no al revés. En algunos problemas, sin embargo, es posible invertir los papeles del modelo M y del resultado X (p.e. cuando se evalúa si una persona es representativa de un estereotipo de profesión o si una profesión es representativa de una persona).

La representatividad, como la similitud, puede ser establecida empíricamente, p.e. preguntando a la gente cuál de dos sucesos X_1, X_2 es más representativo de algún modelo M, o si un suceso X es más representativo del modelo M_1 o del modelo M_2.

Los estudios experimentales han puesto de manifiesto las relaciones existentes entre los juicios realizados por los sujetos y la información de la que disponían en forma de estereotipos (p.e. el perfil de un ingeniero) antes del experimento, y las expectativas sobre la relación entre las variables.

Inicialmente se avanzó la hipótesis extrema de que algunos juicios probabilísticos están basados exclusivamente en la representatividad (p.e. que las distribuciones subjetivas de muestras fueran sistemáticamente independientes del tamaño de las mismas, sugería que se evaluaba la probabilidad de una muestra por la similitud que presentara su estadístico con el correspondiente parámetro de la población). Las investigaciones posteriores han puesto de manifiesto una hipótesis más matizada, en el sentido de que los juicios probabilísticos y las predicciones no están completamente dominados por la representatividad, pero que son muy sensibles a la misma.

En los años 70 Tversky y Kahneman avanzaron dos hipótesis que incorporaban el concepto de representatividad, pero que son conceptualmente distintas.

La primera hipótesis afirmaba que las personas esperan que las muestras sean muy parecidas a las poblaciones y que representen la aleatoriedad del proceso de muestreo, lo que explica la creencia en que los procesos aleatorios se autocorrigen, la confianza exagerada en la estabilidad de los resultados observados en pequeñas muestras, la falacia del jugador, la creencia en la ley de los pequeños números (pequeñas muestras son representativas de sus poblaciones), etc. Todo lo anterior indicaría que la concepción ingenua del azar incorpora la creencia de que incluso las muestras pequeñas son representativas de la población de la que provienen. Una hipótesis similar explicaría la exagerada confianza en el valor predictivo de los rasgos de personalidad, y la correlación sobrestimada entre variables similares y comportamiento.

La segunda hipótesis señalaba que los sujetos utilizan la representatividad como heurístico para realizar juicios y predicciones sobre la probabilidad de un suceso incierto o de una muestra (por lo similar que es a la población o por reflejar las características esenciales del proceso que la ha generado). Cuando predomina la representatividad en la asignación de probabilidades, los sujetos esperan que un suceso o una muestra X sea similar a las propiedades esenciales de la población M a la que pertenece, o que refleje las propiedades más llamativas del proceso que lo genera.

Los sujetos utilizan la heurística de representatividad en la asignación de probabilidades porque:

(1) La representatividad de un ejemplar X de una clase M es más accesible y fácil de evaluar que su probabilidad. Los elementos más representativos (típicos o prototipos o estereotipos) de una categoría son mejor aprendidos, recordados y reconocidos como elementos de la misma. Si el elemento más representativo no coincide con el más frecuente, la representatividad puede inducir sesgos en los juicios sobre frecuencias.

(2) Los sucesos más probables son frecuentemente más representativos que los menos probables (p.e. una muestra que se parece a la población es generalmente más probable que una muestra altamente atípica del mismo tamaño). El suceso más representativo puede ser en muchas ocasiones el más probable, pero no siempre será así, ya que sobre la probabilidad también influyen otros factores de la evidencia (como la probabilidad a priori, la precisión que razonablemente se puede esperar en la inferencia, etc.). Por ello si se asignan probabilidades a un suceso en función de la representatividad, se pueden producir errores sistemáticos al dar mayor a importancia a las variables que afectan a la representatividad del mismo, pero no a su probabilidad, e ignorar las variables que afectan a su probabilidad, pero no a su representatividad.

(3) La creencia de que las muestras son generalmente representativas de sus poblaciones induce a los sujetos a sobrestimar la correlación entre la frecuencia y la representatividad, o entre la asociación estadística y su similitud connotativa.

En definitiva, la representatividad sustituye a la probabilidad porque: es más accesible; representatividad y probabilidad frecuentemente están correlacionadas; y los sujetos sobrestiman esa correlación.

La lógica de la representatividad, sin embargo, no coincide totalmente con la de la probabilidad, por lo que una ordenación realizada por representatividad no reflejará siempre la verosimilitud de cada resultado o la probabilidad de su ocurrencia. Reducir la probabilidad a la representatividad supone un salto conceptual injustificado. Los errores pueden ser especialmente importantes cuando la evidencia es de validez limitada o el suceso altamente específico.

Un resultado que es altamente representativo puede ser improbable en el caso de que el modelo se base en evidencia de validez limitada, como cuando se predice que un candidato realizará con éxito una tarea, en base a la buena impresión que ha causado durante una entrevista, ya que es conocido que las impresiones basadas en entrevistas son notoriamente falibles y a que el éxito o el fracaso en el trabajo está determinado por numerosos factores que no son predecibles en la entrevista (el éxito no estará correlacionado en general con la impresión causada por el candidato).

Igualmente, un resultado altamente específico o detallado puede ser muy representativo, pero poco probable, precisamente por la especificidad. El incremento de especificidad de un suceso no disminuye el grado de representatividad, pero sí lo hace más improbable. Así, un peso por debajo de 70 quilos es atípico para un hombre de edad mediana; sin embargo, un peso de 80 quilos es típico, pero altamente especifico. El segundo es muy representativo, pero poco probable, mientras que el primero es poco representativo, pero más probable que el segundo.

Aunque la representatividad induce a errores, seguramente es útil en numerosos pronósticos realizados por los seres humanos, lo que justificaría su amplia utilización, ya que la información particular que cada sujeto posee debe tener algún valor de diagnóstico respecto a ciertas categorías (p.e. la personalidad de una persona y su profesión). Una ventaja evidente es que la representatividad produce inferencias válidas a partir de pocos datos, lo que es aceptable siempre que se trate con poblaciones homogéneas.

En general, cuando se realizan inferencias estadísticas intuitivamente a partir de los datos, los sujetos han mostrado cierta capacidad para estimar la tendencia central, pero una capacidad menor para determinar la varianza. Esto sugiere que las personas están más familiarizas con valores medios que con la varianza. Cuando las personas piensan en variabilidad no lo hacen en el sentido estadístico de la varianza. Así, Lathrop (1967) concluyó que las estimaciones intuitivas de la varianza están sujetas a efectos contextuales, ya que al estimar la variabilidad relativa las personas tienden a estar influidas por la media del estímulo, y por tanto estiman mejor el coeficiente de variación que la varianza. Beach y Scopp (1968) realizaron experimentos sobre la estimación intuitiva de la varianza con muestras obtenidas de poblaciones normalmente distribuidas y encontraron que las estimaciones no se correspondían con la desviación típica, sino con la varianza elevada a un exponente menor (0,39). Esto se supone que es debido a que la mayoría de los datos se encuentran próximos a la media. Por el contrario, en distribuciones más “uniformes” se estimaba una potencia superior a 2.

Nisbett et al. (1983) mostraron que los sujetos también son capaces de tener en cuenta la variabilidad de la población a la hora de realizar pronósticos por representatividad, lo que haría disminuir la tendencia comentada a ofrecer resultados erróneos.

En un experimento de esos autores, se decía a los sujetos que al llegar a una isla desconocida se observaban ciertos objetos, personas, acontecimientos, etc., y se pedía que establecieran la frecuencia en la isla de los mismos. Por ejemplo, se señalaba que se había observado la presencia de un nativo obeso, y se pedía que se señalara cuál era el porcentaje de obesos en la población de la isla. Igualmente, se había observado que una sustancia llamada floridium ardía con llama verde. Se solicitaba qué porcentaje de veces este elemento ardía con ese color de llama. La respuesta a la primera cuestión fue el 35% y a la segunda el 90%.

Un importante fenómeno puesto de manifiesto en los estudios sobre el heurístico de representatividad es el llamado **efecto diluido** (Nisbett et al., 1981). Si se ofrece a los sujetos una cantidad reducida de información (aunque no sea informativa), el sujeto considera que tiene valor de diagnóstico respecto a un modelo y predecirá un resultado sin realizar ningún cálculo estadístico. Por ejemplo, si se dice que Juan “es amigo de los juegos lógicos”, se considera una información de diagnóstico que permite considerar más probable que sea ingeniero, antes que escultor. Sin embargo, cuando se dispone de mucha información, tanto de diagnóstico como irrelevante para el pronóstico, se produce un efecto diluido, de forma que los pronósticos se hacen menos extremos, atenuando el sesgo de la representatividad. Por ejemplo, si se dice que Juan es amigo de los juegos lógicos, que es extremeño, que le gusta jugar al fútbol, etc., el sujeto considera menos probable que sea un ingeniero que cuando solamente se le señalaba que era amigo de los juegos lógicos.

Se conjetura que la causa del efecto diluido es que la introducción de información que no es de diagnóstico disminuye la similitud global entre la descripción y el prototipo o modelo. El efecto diluido, sin embargo, no se puede considerar una corrección de la representatividad, y en la atribución de probabilidades puede originar errores. El exceso de información puede oscurecer el valor predictivo de la información de diagnóstico disponible.

El hecho de que para establecer una conclusión se considere un reducido número de criterios, que buscan la coincidencia con el estereotipo, y se ignore la mayor parte de la información o evidencia disponible, tiende a producir errores importantes. Los tipos de sesgos provocadas se pueden agrupar en: insensibilidad a las probabilidades a priori de los resultados, insensibilidad al tamaño de la muestra, falsa concepción del azar, insensibilidad a la exactitud predictiva, la ilusión de validez, y una concepción errónea de la regresión.

#3.6.1. INSENSIBILIDAD A LAS PROBABILIDADES A PRIORI DE LOS RESULTADOS

La predicción estadística depende de: (1) la información a priori o de contexto; (2) la evidencia específica relacionada con el caso examinado; y (3) de la precisión esperada en la predicción. Normativamente se sabe que la precisión esperada es la que determina el peso relativo asignado a la evidencia específica y a la información a priori. Conforme decrece la precisión esperada, la predicción es más regresiva, es decir, la predicción se debe parecer más a la probabilidad a priori.

En la teoría estadística se puede ignorar la información a priori o de contexto solamente cuando se espera una predicción infalible a partir de la evidencia. En otros casos, se debe buscar un compromiso entre la ordenación sugerida por la descripción o evidencia específica y la derivada de las probabilidades a priori.

La frecuencia base de los resultados o a priori, por tanto, tiene un gran efecto en la asignación de las probabilidades, pero es un factor que no influye sobre la representatividad. Por ello, cuando en la evaluación de la probabilidad de un suceso se utiliza el heurístico de representatividad, no se tienen en cuenta las probabilidades a priori, aún cuando sean conocidas, y esto provoca sesgos en las predicciones.

Por ejemplo, un agricultor puede concluir que el actual período de sequía es el comienzo de un período prolongado de sequía, porque las circunstancias actuales se parecen mucho a las observadas al comienzo de otros períodos de sequía prolongada. Este juicio será sesgado si no tiene en cuenta los períodos parecidos al actual que no fueron seguidos por sequías prolongadas. Si los períodos de sequía prolongados en la región son raros, y son mas frecuentes los períodos cortos secos seguidos de lluvias, esta última información debería tener un peso importante en la inferencia (Anderson et al., 1977).

Es importante señalar que la tendencia a ignorar las probabilidades a priori se produce también por otras causas distintas del heurístico de representatividad, ya que se han descrito numerosos casos de formulación de juicios que no pueden ser interpretados en términos de representatividad.

P. E. Meehl y de A. Rosen a mediados de los años 50 evaluaron la influencia de la probabilidad a priori sobre los juicios intuitivos en diagnósticos clínicos. En este tipo de problemas, el experto debe asignar probabilidades a un suceso – meta (p.e. el diagnóstico de un paciente o las ventas de un producto) basándose en la frecuencia de base del resultado en la población y en la evidencia específica sobre el caso analizado (p.e. las respuestas del paciente al test de diagnóstico, o el resultado de estudios de mercado), y para ello puede utilizar el teorema de Bayes. Estos autores encontraron que los expertos sobrestimaban la probabilidad de sucesos raros (p. e. un suicidio), ya que apenas tenían en cuenta la probabilidad a priori.

Esta tendencia a basar el juicio exclusivamente sobre la (falible) evidencia específica del caso analizado, permite la construcción de problemas cuyos resultados son aparentemente paradójicos, algunos de ellos popularizados por M. von Savant (1998). Un ejemplo típico señala que se dispone de un test farmacéutico que ofrece un 95% de fiabilidad en la detección del consumo de ciertas drogas, consumidas por el 5% de la población. Si se supone que no se producen errores en el laboratorio (que podrían corregirse con la repetición de la prueba), es posible que una persona que no haya consumido drogas dé un falso positivo en el test. El problema consiste en determinar cuál es la probabilidad de que haya consumido la droga una persona elegida al azar que ha sido sometida al test y ha dado positivo. La mayoría de las personas contestan que el 95%, porque no tienen en cuenta las probabilidades a priori. En realidad la probabilidad es del 50%:

$$\text{P(D/P)} = \frac{P(D)\ P(P/D)}{P(D)P(P/D) + P(D')P(P\ /\ D')} = \frac{(0{,}05)\ (0{,}95)}{(0{,}05)(0{,}95) + (0{,}95)(0{,}05)} = 0{,}5$$

Resultados parecidos habían sido obtenidos por W. Casscells, A. Schoenberger y T. Grayboys (1978), Eddy (1984,18), etc.

Los experimentos típicos para poner de manifiesto la influencia de la representatividad en la ignorancia de la probabilidad a priori o frecuencia de base, son variantes del problema de la determinación de la profesión de una persona a partir de la descripción del perfil de la misma. Las personas juzgan la verosimilitud de que una persona tenga una profesión determinada (p.e. abogado o ingeniero) por lo representativa que sea la descripción de rasgos de la persona de los estereotipos de la profesión, sin tener en cuenta las probabilidades a priori de que una persona pertenezca o no a cada una de esas profesiones. Debido a ello, si se hace una descripción de una persona, aunque esta descripción no sea informativa, buscará la similitud con un estereotipo y asignará la probabilidad, sin tener en cuenta la probabilidad a priori. Por el contrario, cuando no existe una evidencia específica, las personas tienden a asignar mejor las probabilidades en función de las probabilidades a priori.

Los estudios que han explorado cuestiones sobre procedimientos, han puesto de manifiesto, entre otras cuestiones, que dado que los sujetos, cuando aprenden sobre probabilidades, tienden a determinar la distribución de resultados mediante la repetición de pruebas, en estas situaciones tienden a tener en cuenta las frecuencias base, lo que no ocurre en general en el caso de la formulación de juicios sobre sucesos que se manifiestan de forma no repetitiva.

En los estudios sobre la información disponible o evidencia, se ha encontrado que la frecuencia base puede ser percibida por los sujetos como causal o como incidental. Por ejemplo, tal como planteó Ajzen (1977), cuando en un experimento se señala que en un Centro el 75% de los aspirantes al ingreso suspenden el examen, esta información se suele tener en cuenta en preguntas posteriores, porque se interpreta como causal. Sin embargo, cuando se dice que de los estudiantes que se han presentado al examen de ingreso en un Centro, un experimentador ha seleccionado una muestra en la que ha incluido un 75% de estudiantes que han suspendido, los sujetos tienden a considerar esa evidencia como incidental, y no suelen tenerla en cuenta a la hora de realizar inferencias.

La frecuencia base se considera causal si sugiere la existencia de un factor causal que explique por qué es más probable que resulte un suceso antes que otro, y será tenida en cuenta en las inferencias en las que intervenga. Una frecuencia base se considerará incidental si no sugiere la existencia de ese factor causal, y tenderá a ser ignorada en las inferencias.

D. Kahneman y A. Tversky (1972) plantearon un problema que ha permitido diversas versiones para contrastar hipótesis sobre las características percibidas de la evidencia, conocido como “caso de los taxis”.

#3.6.2.- SIMILITUD DE LAS MUESTRAS A LAS POBLACIONES

Al evaluar la probabilidad de obtener un resultado particular en una muestra, las personas utilizan el heurístico de representatividad. Creen que una muestra obtenida aleatoriamente de una población debe ser, en primer lugar, altamente representativa de la misma, es decir, similar a la población en todas sus características esenciales (similitud o proximidad de los estadísticos de la muestra a los parámetros de la población). En consecuencia, por ejemplo, esperan que dos muestras obtenidas de una población se parezcan a ella y entre sí, en contra de lo que predice la teoría estadística, al menos para muestras pequeñas. En segundo lugar esperan que una muestra represente la aleatoriedad del proceso de selección, y por tanto la representatividad de una muestra depende también de los procedimientos empleados para obtenerla.

Es importante distinguir entre la representatividad de la muestra y la representatividad de los resultados de la muestra. Inicialmente se pensó que la representatividad era un juicio que se hacía sobre los resultados de la muestra (cómo de representativa es una muestra de una población, conocidas las características esenciales de ambas, o proximidad entre estadísticos y parámetros). Pero hay otro sentido de la representatividad: las muestras son más representativas si se juzgan más probables de ser representativas en el primer sentido.

Las características que hacen que una muestra sea similar o representativa de la población de la que se extrae son que se preserve la relación mayoría – minoría existente en la población, y que parezca aleatoria, es decir, que mantenga las propiedades del proceso aleatorio que la genera. Intuitivamente se espera que las muestras posean dos propiedades generales: irregularidad y representatividad local.

#3.6.2.1 INSENSIBILIDAD AL TAMAÑO DE LAS MUESTRAS

Las personas no son sensibles a los efectos inducidos por el tamaño de la muestra.
Así, si a una persona se le indica que construya muestras hipotéticas de una población, tenderá a construir muestras cuyas medias coincidan con la media de la población, independientemente de su tamaño.

Cuando se describe una muestra en términos de un solo estadístico (p.e. la media o una proporción), el grado de representatividad de la muestra de la población se determina por la similitud de los valores del estadístico en ambas. Como el tamaño de la muestra no modifica ninguna propiedad de la población, tampoco modifica la representatividad, y por tanto se espera que la similitud del parámetro de la muestra al de la población no varíe con el tamaño de la muestra.

La influencia del tamaño de la muestra también se ignora en la evaluación de sucesos en los que se llama la atención sobre el particular, como en los experimentos del tipo "evaluar qué suceso es más probable, encontrar más de 600 varones en una muestra de 1000 bebés o más de 60 varones en una muestra de 100 bebés".

#3.6.2.2.- ESTIMACIONES CONSERVADORAS

Otra manifestación de la insensibilidad al tamaño de la muestra se produce al juzgar las probabilidades posteriores (es decir, en la asignación de la probabilidad de que una muestra haya sido extraída de una población, dadas dos poblaciones distintas). Este efecto está normalmente asociado con la tendencia a subestimar el impacto de la evidencia sobre las probabilidades a posteriori, sesgo conocido como estimación conservadora, y estudiado por W. Edwards (1968) y P. Slovic y S. Lichtenstein (1971).

D. Kahneman y A. Tversky (1972) realizaron encuestas entre estudiantes israelíes, en los que describían poblaciones abstractas binomiales (p.e. una distribución por sexos o binomial $p = 0,5$) o distribuciones más próximas a la experiencia de los sujetos, como la de alturas de las personas, y se pedía la probabilidad para diferentes intervalos de valores, dados varios valores de N (tamaño de la muestra). Al representar la mediana estimada para cada intervalo, se observó que el tamaño de la muestra N no afectaba a la distribución propuesta (las distribuciones en cada caso para cada N eran muy parecidas). En las distribuciones binomiales, la media generalmente coincide con la moda, y aunque la moda de las distribuciones subjetivas de las muestras estaba en general situada correctamente en el valor más representativo, p, la media estaba desplazada hacia la cola más larga. J. Cohen y C.E.M. Hansel (1956), y C.R. Peterson, W.M. DuCharme y W. Edwards (1968) habían descrito un efecto semejante.

Otros experimentos de los mismos autores, pusieron de manifiesto que la idea de que la varianza de la muestra decrece al incrementarse el tamaño de la misma no es algo que pueda considerarse intuitivo. Las personas tienden a considerar más verdadero un resultado dado en porcentaje, obtenido en una muestra pequeña, que otros resultados provenientes de muestras más amplias (p. e. se critica el I.P.C. porque no representa la media de toda la población).

#3.6.2.3. FALSA CONCEPCIÓN DEL AZAR

De igual forma que las personas creen que una muestra es representativa de las características esenciales de la población (proximidad de los estadísticos a los parámetros), también espera que una secuencia de sucesos aleatorios sea similar al proceso que la genera, que refleje las características esenciales del proceso, y no sólo globalmente en la muestra o en la secuencia completa, sino también localmente, en cada una de sus partes, incluso en secuencias cortas (ley de los pequeños números).

De esta forma, si se lanza una moneda varias veces y se anotan las caras (C) y las cruces (X) obtenidas, los sujetos creen que la secuencia CXCXCX es más probable que la secuencia CCCXXX, que no parece aleatoria, y también más probable que la secuencia CCCCXC, que no parece que provenga de una moneda equilibrada (que produciría un 50% de caras a largo plazo).

Cohen y Hansel (1955) indicaron que “Instead of a tendency towards a distribution, we find a tendency away from it. This might be a accounted for by the maked preference by balance and the avoidance of asymmetrical arrangements”.

Existe evidencia de que las personas adultas tienen dificultades con el concepto de independencia y particularmente en lo que se refiere a la capacidad de las personas para detectar secuencias aleatorias de las no aleatorias. Wagenaar (1970, 1972) ha concluido que “Subjects are unable to produce a randomized sequence of two, three or more alternatives even if they are explicity instructed and motivated to do so. Generally S´s show a tendency towards too few reptitions and too many alterations.”

Como consecuencia de una incorrecta concepción del azar los sujetos incurren en varias falacias: la del jugador (donde se considera el azar un proceso autocorregido), y la ley de los pequeños números (cuando se considera que muestras pequeñas son altamente representativas de las poblaciones de las que se extraen).

#3.6.2.3.1. PRESERVACIÓN DE LA RELACIÓN MAYORÍA – MINORÍA

Las muestras que mantienen la relación mayoría – minoría de la población son más representativas de la misma, y por tanto se consideran más probables que otras que no presenten esa relación y son, objetivamente, más probables.

Los ejemplos típicos de experimentos para comprobar la representatividad de la relación mayoría – minoría suelen tener la siguiente estructura: En un Centro se cursan dos carreras, A y B, con el mismo número de estudiantes, donde los chicos son mayoría en la primera (65%) y minoría en la segunda (45%). Si usted entra en una clase al azar y observa que el 55% de los estudiantes son chicos, ¿con qué probabilidad diría que la clase es de la carrera A?. La mayoría de los sujetos señalan que es más probable que se trate de una clase de la carrera A que de la carrera B, pero objetivamente es ligeramente más probable la segunda opción, debido a que la varianza para p = 0,45 es mayor que para p = 0,65.

La mayoría de los sujetos muestran asombro cuando se les explica que en un grupo de 23 personas, considerado pequeño, la probabilidad de que al menos dos de ellos cumplan los años el mismo día supera el 50%. El razonamiento que seguramente hacen los sujetos es que con 23 personas el número esperado de cumpleaños por día es menor que 1/15, y por tanto que en un mismo día coincidan dos cumpleaños, habiendo 343 días “vacíos”, no es muy representativo y, por tanto, resulta improbable.

En general, los sujetos no ignoran el orden de la información, pero están muy atentos a los valores que se consideran representativos de la población. Así, si se plantea que en una ciudad se han examinado todas las familias que tienen 6 hijos, y se pregunta cuál de las dos secuencias exactas de nacimientos HVHVVH y VHVVVV (con H= hembra y V = varón) será más frecuente, aunque objetivamente las dos secuencias son prácticamente igual de probables, la primera es más representativa, ya que la fracción 3 hembras y 3 varones es más próxima a la de hembras y varones que se espera en la población. La segunda secuencia falla a la hora de reflejar esa proporción. Esquemas semejantes fueron descritos por J. Cohen y C.E.M. Hansel (1956) y F. Alberoni (1962).

#3.6.2.3.2. IRREGULARIDAD DE LAS SERIES

Una propiedad de la apariencia de aleatoriedad es la ausencia de patrones sistemáticos. Una secuencia regular no se considera representativa de un proceso aleatorio. Las personas esperan alguna irregularidad en la aparición de los sucesos, no sólo en el orden de los resultados, sino también en su distribución. Por ejemplo, secuencias de caras y cruces como CXCXCXCX o CCXXCCXXCCXX son muy regulares, y por tanto se suponen poco representativas y poco probables. La presencia de alguna perturbación contribuye a aumentar la representatividad y la verosimilitud de los sucesos.

Si en lugar de preguntar al sujeto por secuencias que fallen en representar el parámetro fundamental de la población, ese parámetro toma el mismo valor en las dos secuencias, entonces los sujetos consideran más representativa a la que presente alguna irregularidad. Así, si se les pide que asignen probabilidades a las secuencias VVVHHH y VHHVHV, los sujetos considerarán más probable la segunda, porque la primera parece menos aleatoria.

Un experimento clásico consiste en decir a los sujetos encuestados que se han distribuido aleatoriamente n objetos entre s personas, y mostrarles dos resultados alternativos de esa distribución (por ejemplo, una tabla en la que se señala el número de objetos recibido por cada sujeto en cada alternativa examinada). Una de las alternativas es una distribución uniforme, donde cada persona recibió n / s objetos, y la otra muestra una asignación irregular (unas personas reciben dos objetos más que otras). Se pregunta a los sujetos que indiquen cual de las dos alternativas es más probable.

Aunque la distribución uniforme sea más probable que la distribución irregular con la que se compara, las personas asignan mayor probabilidad a la segunda porque es más representativa de una distribución aleatoria. Las personas perciben la aleatoriedad como impredecible, pero esencialmente justa o equilibrada. Por ello esperan que en una asignación aleatoria cada sujeto reciba aproximadamente (pero no exactamente igual) el mismo número de objetos.

#3.6.2.3.3. REPRESENTATIVIDAD LOCAL

W.K. Estes (1964), G.S. Tune (1964) y W.A. Wagenaar (1970, 1972) mostraron que cuando las personas simulan un proceso aleatorio (p.e. lanzamientos con una moneda), producen secuencias que son localmente representativas, y que consideran poco probables, o las rechazan como no aleatorias, secuencias en la cuales existe una distribución correcta en series largas, seguramente porque las series largas no son localmente representativas. Sin embargo, una secuencia localmente representativa se desviará sistemáticamente de la secuencia esperada (o intuición de aleatoriedad), ya que, en el caso de pocas manifestaciones, existen muchas alteraciones y muy pocos grupos homogéneos.

Cuando a un sujeto se le pide que genere una secuencia aleatoria de una hipotética serie de lanzamientos de una moneda, produce secuencias donde la proporción entre caras y cruces está más cerca de 0,5 de la que se produciría en un proceso aleatorio. Cada segmento de la secuencia de respuestas en las muestras parece reflejar la verdadera proporción en la población, lo que es altamente representativo del proceso aleatorio, de la "justicia" de la moneda, de que la moneda está "equilibrada".

Una muestra que contiene todos los sucesos posibles es más representativa que una a la que falta algún suceso. Por ejemplo, en un proceso binomial con $p = 4/5$ la mayor parte de los sujetos cree que una muestra de 10 aciertos y 0 fallos es menos probable que una muestra de 6 aciertos y 4 fallos, aunque la primera muestra es objetivamente más probable.

#3.6.2.3.4.- FALACIA DEL JUGADOR

Una consecuencia de esperar secuencias localmente representativas es la conocida falacia del jugador. Los sujetos actúan como si cada segmento de una secuencia aleatoria debiera reflejar la frecuencia existente en la población; si la secuencia obtenida se aparta de la proporción de la población, introduce un sesgo para corregir la desviación, incurriendo en la falacia. Así, después de observar en una serie de lanzamientos de una moneda una secuencia CCCCCC se piensa que es más probable que en la próxima tirada aparezca una X en lugar de una C.

Algunas leyes de la naturaleza actúan corrigiendo desviaciones de un equilibrio estable, mediante la actuación de fuerzas que restauran el equilibrio, pero este no es el caso de las leyes del azar o de los procedimientos de muestreo. El azar no es un proceso autocorrectivo, donde una desviación en una dirección induce otra de sentido opuesto, cancelándose para restaurar el equilibrio. Las desviaciones no se corrigen en un proceso aleatorio, sino que se diluyen.

Igualmente, las personas tienden a creer, erróneamente, en la imparcialidad de las leyes del azar (se espera que la imparcialidad de la moneda corrija una desviación con otra, de tal forma que se cancelen pronto: pero la moneda no tiene memoria para ser imparcial).

#3.6.2.3.5.- FALACIA DE LA LEY DE LOS PEQUEÑOS NÚMEROS

La ley de los pequeños números señala que las pequeñas muestras son altamente representativas de las poblaciones de las que son extraídas. La confianza en la ley de los pequeños números lleva a sobrestimar la confianza en los resultados obtenidos en muestras pequeñas.

La ley de los grandes números garantiza que las muestras muy grandes sean altamente representativas de las poblaciones de las que han sido extraídas. La intuición de las personas tiende a creer que también se verifica lo anterior para pequeñas muestras, porque creen que las muestras son muy similares entre ellas y con la población, y que existen procesos autocorrectivos. Como consecuencia, la variabilidad esperada en las muestras pequeñas es menor que la verdadera, y las personas creen en la ley de los pequeños números.

Así, por ejemplo, bastantes científicos esperan que una vez establecida una hipótesis, las características de la población descrita por la hipótesis se manifiesten en todas las muestras, cualquiera que sea su tamaño, sobrestimando la verosimilitud y replicabilidad de sus resultados. En la investigación este sesgo lleva a la selección de muestras de tamaños reducidos y a sobreinterpretar los hallazgos.

La creencia en la ley de los pequeño números lleva a una práctica científica como la siguiente:

1.- Se avanzan hipótesis en base a los resultados de muestras pequeñas, sin darse cuenta que la probabilidad de error es muy alta. De esta forma se sobrestima el poder predictivo.

2.- Se muestra una confianza excesiva en las primeras tendencias detectadas (p.e. los datos de los primeros sujetos) y en la estabilidad de los patrones observados. Se sobrestima la significación (significance).

3.-Se tiene una gran confianza en la replicabilidad de los resultados. Se subestima la amplitud de los intervalos de confianza.

4.-Raramente se atribuyen las desviaciones de los resultados a las expectativas de variabilidad de la muestra, porque se encuentra una "explicación" causal para cualquier discrepancia. De esta forma se pierde la oportunidad de identificar la influencia de la variación de la muestra, con lo que se refuerza la fe en la ley de los pequeños números.

#3.6.3. INSENSIBILIDAD A LA PRECISIÓN PREDICTIVA

Las personas frecuentemente deben hacer predicciones numéricas sobre un valor futuro (el resultado de una acción, la demanda de un bien, el resultado de un partido de fútbol, la producción de leche de una vaca, etc.). En la teoría estadística, la bondad de una predicción y el rango al que se puede extender se deriva de consideraciones relacionadas con el poder predictivo. Cuando no es posible realizar ninguna predicción, se puede asignar el mismo valor a cualquier caso (por ejemplo, atribuir a un conjunto de empresas estudiadas los beneficios medios). Si la predictibilidad es perfecta, el valor predicho será el valor real, y el rango de las predicciones será igual al rango de los valores predichos.

Normalmente las personas ignoran, al hacer sus predicciones, el poder predictivo del contexto en el que juzgan, pero son sensibles a la representación. Al predecir, por ejemplo, los resultados de una empresa (beneficios futuros), las personas tienden a dejarse influir por cuestiones como la descripción de la compañía. Una descripción muy favorable hará que se considere más representativo de la misma el que obtenga un beneficio alto, mientras que una pobre descripción se asociará con resultados bajos. Sin embargo, una descripción favorable no tiene por que ser fidedigna. Otro ejemplo (Anderson et al., 1997) señala que a veces una persona puede pagar un precio alto por una vaca al sobrestimar la influencia en la producción de leche de factores tales como el tamaño de la ubre o el número de veces que se ha quedado preñada.

En un experimento típico se presentaba a los sujetos encuestados una descripción de una clase práctica impartida por un profesor ayudante. A un grupo de sujetos se les pidió que **evaluaran** la calidad de la lección descrita (puntuación mediante percentiles), mientras que a los sujetos del otro grupo se les pidió que **predijeran** la situación del profesor dentro de 5 años. El juicio emitido en ambos casos fue idéntico. La predicción de un resultado remoto (éxito del profesor después de 5 años) fue idéntica a la evaluación de la información en que se basaba la predicción (la calidad de la lección práctica).

#3.6.3.1. ILUSIÓN DE VALIDEZ

Las personas frecuentemente realizan una predicción seleccionando el resultado (p.e. la profesión de una persona) que es más representativo de los datos (p.e. la descripción de una persona). La confianza que se tiene en una predicción depende principalmente del grado de representatividad (cómo coinciden el resultado y los datos), sin prestar atención a los factores que pueden limitar la exactitud de la predicción.

Se llama **ilusión de validez** a la confianza infundada producida por una buena correlación entre el resultado predicho y la información utilizada, a la confianza depositada por las personas en juicios altamente falibles. La ilusión persiste incluso cuando el juicio sobrepasa ampliamente los factores que limitan la bondad de la predicción. Esta ilusión ha sido descrita, por ejemplo, en el caso de las entrevistas realizadas por los sicólogos en la selección de personal para puestos de trabajo, donde se ha podido demostrar la escasa influencia en los resultados de los juicios formados mediante esa metodología.

Las pautas o patrones seguidos por las variables de partida se consideran altamente consistentes cuando son muy redundantes o están muy correlacionados. Por ello, las personas tienden a tener una gran confianza en las predicciones basadas en variables de partida redundantes. Sin embargo, se sabe por la teoría que se obtienen mejores resultados cuando las variables son independientes entre sí que cuando son redundantes o existe correlación entre ellas.

Un determinante de la representatividad en el contexto de predicciones numéricas con variables independientes multiatributo, es la consistencia o coherencia y la extremalidad de esas variables. Como la confianza se incrementa con la consistencia, la confianza será generalmente mayor cuando las variables estén altamente correlacionadas. Se produce así una contradicción: la alta correlación entre las variables dependientes incrementa la confianza y disminuye la validez. Así, por ejemplo, cuando se entrevista a un candidato para un puesto, la mayoría de nosotros tiene una gran confianza en su predicción sobre el buen rendimiento futuro de su candidato favorito, pese a que se sabe que este tipo de entrevistas son altamente falibles.

#3.6.3.2. CONCEPCIÓN ERRÓNEA DE LA REGRESIÓN

Existe un fenómeno general conocido como regresión hacia la media. Si se tienen dos variables aleatorias X e Y que tienen la misma distribución, si se seleccionan un conjunto de individuos cuya media se desvía de la media de X en k unidades, entonces la media de sus puntuaciones en Y normalmente se desviará de la media en menos de k unidades. Así, por ejemplo, si un grupo numeroso de niños ha sido examinado con dos versiones equivalentes de un test de actitud, y se seleccionan 10 niños de entre los que han obtenido puntuaciones por encima de la media en el primer test, y se comprueba su posición en el segundo test, nos encontraremos con que han obtenido una menor puntuación.

Este fenómeno fue observado por Galton y todos estamos expuestos en nuestra experiencia a multitud de manifestaciones del mismo. Casos de regresión hacia la media se han observado en la estatura de los padres y de los hijos, en la inteligencia de maridos y esposas, en las notas de los individuos en exámenes consecutivos, etc. Padres de inteligencia excepcional tienen hijos decepcionantes; esposas brillantes tienen maridos tontos; puntuaciones altas en una prueba van seguidas de otras sorprendentemente bajas, etc.

Pese a la exposición diaria a estos fenómenos, las personas no han desarrollado una intuición adecuada de la regresión (no entienden correctamente esa idea o no detectan la regresión implícita en los datos), posiblemente debido a que: (1) en muchas situaciones las personas no esperan que se presente el fenómeno de regresión hacia la media; (2) la noción de regresión es muy difícil de adquirir; y (3) cuando se intuye o reconoce que existe la regresión hacia la media, se atribuye a explicaciones causales espurias, en lugar de entender un hecho muy simple: normalmente las correlaciones entre variables son inferiores a 1. Por ejemplo, el hecho de que algunos estudiantes excepcionales no sean tan buenos cuando se incorporan a una profesión suele explicarse acudiendo a argumentos como “tiene una gran competencia, pero le falta garra en el mundo social”, “después del prestigio ganado como estudiante, se durmió en los laureles”, “tuvo poco apoyo de sus colegas y bastante mala suerte”, etc. Otras explicaciones del mismo tipo se buscan para el caso opuesto, es decir, aquellos profesionales brillantes que habían sido malos estudiantes.

La razón de los sesgos descritos es que los efectos de la regresión violan la intuición de que el resultado predicho es muy representativo de la información de partida disponible (por representatividad se supondría que los padres inteligentes deben tener hijos también inteligentes; los buenos estudiantes deben ser buenos profesionales; etc.). Los resultados se piensa que deberían ser tan extremos como el valor tomado por la variable independiente.

En general, parece ignorarse que dadas dos variables de igual varianza, los enunciados siguientes son equivalentes: (a) Y es regresiva con respecto a X, (b) la correlación entre Y y X es menor que la unidad. Explicar la regresión equivale a explicar por qué la correlación es menor que la unidad.

Este sesgo puede tener consecuencias muy importantes. Por ejemplo, en el entrenamiento de pilotos se suele observar que una alabanza después de un buen aterrizaje suele ir seguido de otro aterrizaje peor, y que una recriminación después de un mal aterrizaje suele ir seguida de un aterrizaje mejor. A partir de ello los instructores concluyen que las recompensas van en detrimento del aprendizaje, mientras que el castigo es beneficioso, contrariamente a lo aceptado por la Sicología. Precisamente porque el instructor ha alabado un buen aterrizaje (que irá seguido de otro peor) y ha recriminado uno malo (que irá seguido de otro mejor), cree que la recriminación está asociada con los buenos aterrizajes futuros (en lugar de ser el actual mal aterrizaje la explicación).

En un experimento típico se plantea un problema de estimación, en el que se supone que un individuo elegido al azar ha obtenido una puntuación de 140 en un test de I.Q. (C.I.) estándard (donde 100 es el valor medio de la población). La puntuación es la suma de su verdadera puntuación y de un error aleatorio de medida que está distribuido normalmente. Se pide al sujeto que indique su mejor predicción de las fronteras del intervalo que con un 95% de probabilidad contenga el valor de la verdadera puntuación. Es decir, que dé una estimación superior de tal forma que esté seguro a un nivel del 95% de que la puntuación verdadera es menor que la estimada, y una estimación inferior de tal forma que este seguro a un nivel del 95% de que la puntuación verdadera es mayor.

Como el valor obtenido es mayor que la media de la población, es más probable que el error tenga una componente positiva y que el sujeto obtenga un valor menor en otras mediciones. La mayoría de los sujetos, sin embargo, establecen intervalos de confianza simétricos alrededor de 140, mostrando unas falsas expectativas sobre la regresión. Existe así una tendencia a realizar predicciones como si la información de las variables independientes estuviera libre de errores.

#3.6.3.2.1.- FALACIA DE LA CONJUNCIÓN

Una ley básica de la probabilidad señala que la especificación solamente puede reducir la probabilidad. Así, la probabilidad de que una persona sea republicana y artista es menor que la probabilidad que sea artista. Es decir, $P(A \cap B) \leq P(B)$.

Sin embargo, $P(A \cap B) \leq P(B)$ no se verifica en la representatividad, donde la especificidad incrementa la representatividad, provocando la "falacia de la conjunción".

Cuando se incurre en la falacia de la conjunción se puede contestar que la descripción de una persona es más probable que se refiera a un profesor de gimnasia que un profesor, considerando así más probable una conjunción que uno de sus componentes, porque la conjunción puede ser considerada más representativa de un resultado que los elementos aislados.

#3.7.- ACCESIBILIDAD O DISPONIBILIDAD

En ocasiones las personas determinan la frecuencia de una clase o la probabilidad de un suceso por la facilidad con que los ejemplos o asociaciones de los mismos les vienen a la mente o son evocados.

Por ejemplo, se puede juzgar la probabilidad de sufrir un ataque de corazón por lo que se recuerda de los sufridos por nuestros conocidos, o la viabilidad de un negocio por las dificultades que somos capaces de imaginar que encontremos si lo emprendemos, o estimar más probable que nos roben después de hablar con un amigo que ha sido víctima de un robo. Este tipo de heurística se denomina accesibilidad o disponibilidad.

Las personas que utilizan la heurística de la disponibilidad juzgan intuitivamente un suceso como probable o frecuente si sus instancias son fáciles de imaginar o de recordar. La fuerza de una asociación (imaginabilidad, facilidad de recuperación, etc.) permite al sujeto inferir la frecuencia de un suceso, categoría o relación. Precisamente, los vínculos asociativos se refuerzan por la repetición, y puede que esta sea la ley más antigua que el hombre conoce sobre la memoria. La heurística de la disponibilidad explota la inversa de esa ley: utiliza la fuerza de la asociación como base para juzgar la frecuencia. Los sujetos explotan de modo intuitivo su conocimiento de las leyes asociativas.

La heurística de la disponibilidad es útil en la asignación de frecuencias porque las instancias de las clases numerosas se encuentran más fácil y rápidamente (son más accesibles) que las instancias de las clases menos frecuentes, y porque, en general, los sucesos más frecuentes son más fáciles de recuperar, recordar o imaginar que los sucesos raros.

Sin embargo, la disponibilidad puede verse afectada por otros factores que no están relacionados con la ocurrencia o la frecuencia real, y por tanto distorsionar la frecuencia percibida y producir sesgos sistemáticos. La probabilidad objetiva de los sucesos evocados no cambia por los recuerdos o conversaciones recientes sobre ellos, pero sí que incrementan la saliencia cognitiva o accesibilidad de esos sucesos, afectando las expectativas. Por ejemplo, un desastre reciente o una película que nos impresiona podría distorsionar fácilmente nuestro juicio.

Existen tres clases principales de sesgos en la disponibilidad que pueden afectar a los juicios:

(1) Los datos sobresalientes son más disponibles y tienen una influencia desproporcionada sobre nuestros juicios.

(2) Los sesgos del proceso de recuperación pueden hacernos basar nuestro juicio en datos no representativos.

(3) Las estructuras cognitivas resistentes, tales como creencias, valores, preconcepciones, pueden hacernos poco sensibles a la evidencia disponible.

A la hora de establecer las estrategias a través de las que se aplica la disponibilidad, se pueden distinguir dos tipos de operaciones mentales para "extraer cosas de la mente": la recuperación de asociaciones e instancias, y la construcción de ejemplos o de escenarios. Recuperar y construir son estrategias diferentes, que siguen reglas distintas y que se utilizan para responder a cuestiones dispares.

La estrategia de recuperación de instancias consiste en utilizar el número de instancias recuperadas en un período corto de tiempo para estimar el número de instancias que podrían recuperarse en un período más largo. En este tipo de problemas, las personas pueden asignar índices de disponibilidad de forma rápida y exacta. Estos índices se pueden contrastar experimentalmente, porque cada problema tiene una respuesta correcta, y se puede comparar con ella la estimación efectuada por los sujetos en los estudios experimentales.

En caso de recuperación de instancias, la disponibilidad puede utilizar diferentes aproximaciones:

- Facilidad de recuperación. Las clases cuyas instancias son recordadas con mayor facilidad se consideran más numerosas que aquellas del mismo tamaño que no son tan accesibles.

- Conteo de ocurrencias. Al juzgar la probabilidad de que una pareja particular se divorcie, se pueden repasar los casos que se recuerdan de parejas similares y anotar cuántas de ellas se divorciaron. Se concluirá que un divorcio es probable si entre las parejas recordadas prevalecen los casos de divorcio.

- Reglas de construcción: a partir de una regla de construcción de instancias los sujetos estiman la frecuencia relativa de las mismas. En estos casos, los sujetos no pueden construir y enumerar la totalidad de las instancias de cada clase, y por ello generan un subconjunto y estiman la frecuencia mediante su disponibilidad. Por ello las clases que son más fáciles de imaginar o percibir se consideran más frecuentes que las clases cuyas instancias son menos accesibles.

Un ejemplo típico es cuando se pregunta dónde es más frecuente encontrar la letra R, al principio, en la tercera letra, al final de una palabra, etc. Igualmente se utiliza esta estrategia en problemas relacionados con soluciones combinatorias, u otras donde existe un mecanismo que genera un número muy elevado de instancias.

La construcción de escenarios se utilizará en otros casos, especialmente cuando los sucesos sean únicos para los sujetos.

#3.7.1.- FACILIDAD EN LA RECUPERACIÓN DE INSTANCIAS

Una clase cuyos miembros sean recordados más fácilmente será considerada más numerosa que otra clase cuyas instancias sean más difíciles de recuperar. Los factores que afectan a la facilidad de recuperación son la familiaridad, la prominencia, la actualidad, etc.

Cuando se somete a los individuos una lista de hombres famosos y de mujeres menos famosas, y se les pregunta la frecuencia de cada sexo, responden que hay más hombres, cuando la frecuencia era la misma. La relativa facilidad con que se evocan los nombres más famosos induce a sobrestimar su frecuencia. Igualmente, los sucesos actuales son más fácilmente accesibles que los antiguos. Las probabilidades asignadas a que una casa se incendie se juzgan más altas cuando se acaba de contemplar un incendio o cuando se leen noticias sobre incendios en los periódicos.

Sesgo egocéntrico. La asignación de responsabilidades en una empresa común (p.e. el orden de firmantes de un artículo), entre participantes bien intencionados, se realiza recordando la contribución de cada uno al producto final. Sin embargo, algunos aspectos de la interacción pueden ser recordados más fácilmente, por ser más disponibles, que otros. Lo que puede ser recordado más fácilmente no es sino un subconjunto aleatorio del total. En un proyecto, independientemente del éxito con que haya sido culminado, una persona recordará en mayor proporción su propia contribución al mismo, subestimando la del resto de participantes.

En la experiencia diaria las tendencias egocéntricas pueden distorsionar la evaluación de las empresas comunes en la asignación de responsabilidad. Es un foco de disensiones, y los individuos raramente son capaces de dirimir sus diferencias mediante una evaluación honesta. Para evitar los conflictos derivados de este sesgo, estas situaciones, a veces, se resuelven por convenciones, como la de firmar los artículos por orden alfabético.

Facilidad de búsqueda. Si se pregunta a un sujeto qué palabras de tres letras o más son más frecuentes, las que comienzan por r o las que la tercera letra es una r, como es mucho más fácil recordar palabras que comienzan por r que encontrar palabras cuya tercera letra sea una r, los sujetos dirán que las primeras son más numerosas. Esto ocurre con letras como la r o la k que tienen en inglés mayor probabilidad de aparecer en la tercera posición que en la primera.

La gente, igualmente, cree que son más numerosas en un diccionario las palabras abstractas (amor, pensamiento) que las concretas (puerta, ventana), porque tiene mayor facilidad de formar palabras de la primera clase en los contextos que imagina. Sin embargo ambas tienen la misma probabilidad.

Imaginabilidad. A veces se trata de establecer la frecuencia de clases cuyas instancias no están almacenadas en la memoria, pero que se pueden generar de acuerdo con una regla dada. En tales circunstancias, se generan varias instancias y se evalúa su frecuencia por la facilidad con que se puedan construir las instancias. Sin embargo, la facilidad de construir instancias no es un indicador de la frecuencia real, y este modo de evaluación lleva a la aparición de sesgos.

Si se desean formar comités de k miembros ($2 \leq k \leq 8$) con 10 personas, el número de los mismos que se pueden formar viene dado por el coeficiente binomial $\binom{10}{k}$. Con k = 5 miembros se pueden formar 252 comités. El número de comités de (10 - k) miembros es igual al número de comités de (k - 10) miembros, porque al formar un comité de (10 - k) miembros se tiene otro comité de (k - 10) "no miembros".

Los comités de 2 miembros son más fáciles de imaginar que los de 8 miembros. Los pequeños comités, en general, parecerán más frecuentes que los formados con muchos miembros. Por ejemplo, la estimación media del número de comités de 2 miembros fue 70, mientras que de 8 miembros fue 20 (siendo así que en ambos casos se pueden formar 45).

El riesgo de una expedición o de una aventura se evalúa imaginando las contingencias que no están cubiertas. Se pueden imaginar un gran número de desastres que, sin embargo, tienen una pequeña probabilidad de suceder. Por el contrario, el riesgo asociado con una empresa puede subestimarse si algunas dificultades son difíciles de concebir o simplemente no nos vienen a la mente.

#3.7.2.- ILUSIÓN DE CORRELACIÓN

Cuando dos sucesos están asociados conceptualmente, los sujetos tienden a sobrestimar la frecuencia con que co-ocurren en la realidad. Esta correlación ilusoria fue descrita por Chapman y Chapman (1969) en el campo de la Sicología clínica.

La frecuencia con que co-ocurrren dos sucesos se estima por la fuerza de la asociación entre ellos. Cuando la asociación es fuerte, el sujeto concluye que los sucesos aparecen frecuentemente asociados. La conexión asociativa entre sucesos es reforzada por la frecuencia de la co-ocurrencia. De esta forma, disponemos de un procedimiento (heurística) para estimar lo numerosa que es una clase, la verosimilitud de un suceso, o la frecuencia de las co-ocurrencias, mediante la facilidad con que se puedan realizar las operaciones mentales de recordar, construir o asociar.

Las operaciones de codificar, almacenar y recuperar información, sin embargo, no están exentas de problemas específicos, que añaden factores adicionales y dificultades al establecer covariaciones. Las expectativas o las hipótesis constituyen una fuente potencial de sesgos en cada etapa del procesamiento de la información.

Tablas de contingencia. Las tablas de contingencia simple 2 x 2 constituyen un dispositivo experimental normativo para juzgar la eficacia de los sujetos en la detección de covariaciones o en el reconocimiento de relaciones funcionales. Estas tablas resumen el número de instancias presentes y ausentes de las variables X (p.e. una enfermedad) e Y (p.e. un síntoma), siendo X e Y dos variables presumiblemente asociadas.

Los resultados experimentales obtenidos muestran que las personas juzgan pobremente las relaciones funcionales. Las dificultades para reconocerlas se derivan de utilizar estrategias que ignoran una o más de las 4 celdas de las tablas de contingencia. Un fallo bastante común consiste en comparar solamente el tamaño de la celda presente - presente con la población total. De esta forma mucha gente dice que el síntoma Y está asociado con la enfermedad X simplemente porque la tabla de contingencia presenta un gran número de casos en los que se presentan la enfermedad y el síntoma.

Cuando se trata de decidir si las personas con pelo rojo tienen un carácter fuerte, en primer lugar hay que decidir qué datos deben considerarse, o recuperarse: (a) una muestra de personas con el pelo rojo y carácter fuerte; o (b) solamente la gente de pelo rojo y anotar cuáles tienen carácter fuerte; o (c) personas de carácter fuerte para comprobar cuántos tienen el pelo rojo. Seguramente se adoptará una de esas estrategias o una combinación de ellas, sin que se caiga en la cuenta de que las personas morenas y pacíficas puedan ser también relevantes para la solución del problema. Esta forma de proceder contrasta con las tablas de contingencia 2x2 (normativas) donde todos los datos relevantes están disponibles en igual medida.

Pero aunque los sujetos entiendan que necesitan los datos de las cuatro celdas: ¿Como diseñar el tamaño adecuado de muestra para generar las frecuencias relevantes de las celdas? (¿el primer grupo de personas que le vengan a la mente; toda la gente de una determinada clase, como sus familiares, sus amigos o vecinos; un método de generar aleatoriamente casos?). Los sesgos potenciales de los sistemas informales de muestreo son evidentes. Las muestras de las personas que primero recordemos estarán expuestas seriamente a la disponibilidad, pero las muestras de amigos o vecinos mostrarán los mismos sesgos, además de violar los requerimientos de independencia (ya que los amigos o vecinos pueden tender a compartir tendencias de comportamiento, características físicas o relaciones entre ellos).

#3.7.3.- IMPACTO DE LAS PRECONCEPCIONES

L. J. Chapman y J.P. Chapman (1967, 1969) estudiaron el impacto de las preconcepciones en las percepciones, en un trabajo donde señalaron que los sicólogos tienen una confianza excesiva en los tests proyectivos como el de Roscharch o el Draw-a-Person (DAP) para el diagnóstico de enfermedades mentales, pese a la abrumadora evidencia empírica en contra de su eficacia en el diagnóstico. Sin embargo, multitud de sicólogos experimentados los continúan utilizando como instrumentos de diagnóstico, suponiendo una correlación entre, por ejemplo, los rasgos de un dibujo y ciertos síntomas síquicos del sujeto que lo ha realizado. Así, se supone que los pacientes con suspicacia paranoica subrayan los ojos en sus dibujos, y las personas dependientes exageran la boca.

Las conclusiones de esos estudios han acuñado principios como "los resultados de los tests que aplica a sus pacientes son lo que usted piensa". La correlación ilusoria es una consecuencia del heurístico de accesibilidad, ya que, por ejemplo, la suspicacia de un paranoico evoca más fácilmente los ojos que cualquier otra parte del cuerpo. Pero los resultados de estas asociaciones semánticas son covariaciones ficticias.

Los autores anteriores concluyeron, al estudiar la capacidad de los sujetos para detectar covariaciones entre signos clínicos y síntomas, que las covariaciones detectadas parecen reflejar las verdaderas covariaciones en menor medida que las preconcepciones basadas en modelos teóricos o construcciones semánticas, es decir, aquellas que teóricamente "deberían" existir.

La ilusión de correlación no se reduce solamente en campo del diagnóstico clínico, sino que muchas supersticiones se basan en supuestas asociaciones entre acciones particulares o sucesos y los consiguientes resultados positivos o negativos. Los estereotipos raciales, regionales, religiosos, u ocupacionales son creencias fuertemente asentadas sobre covariaciones, que son muy resistentes al impacto de la evidencia contraria.

Los estudios pusieron de manifiesto que las preconcepciones hacen más disponibles los datos más próximos a las mismas, pero no pudieron explicar cómo ocurre ese fenómeno. No se sabe si el sujeto "ve" realmente la relación de la que informa, o si simplemente informa de la relación que espera que se presente en los datos, o si se trata de un "compromiso" entre sus percepciones subjetivas y las esperadas. Parece bien establecido, sin embargo, que los datos no eliminan las preconcepciones subjetivas. Incluso después de una prolongada exposición a relaciones negativas, los sujetos siguen informando sobre correlaciones positivas.

Esto plantea dos cuestiones: (1) ¿Cuál es el mecanismo mediante el cual las creencias erróneas o exageradas sobre relaciones funcionales pueden sobrevivir frente a una lógica convincente o a los desafíos de la evidencia? (2) ¿Cómo pueden reconciliarse los anteriores resultados con la capacidad obvia de los organismos de aprender y con la exactitud y adecuación de muchas de nuestras actuales creencias y estrategias sociales?

Un mecanismo que podría explicar lo anterior es que las personas actúan según sus propias creencias, y ello hace discordante que las actualicen a la luz de la nueva evidencia. Las creencias o expectativas pueden constituirse, de esta forma, en una profecía autorrealizada, es decir, que una creencia provoque que un sujeto se comporte de un modo que produzca los resultados o datos que finalmente validan esa creencia. Otro mecanismo tendría que ver con la propensión a asimilar nueva información en función de su consistencia con las creencias anteriores, y de esta forma la asimilación de datos desarrolla una explicación causal (considerada correcta) del estado del mundo conjeturado.

Sin embargo es innegable que los sujetos son capaces de percibir ciertas covariaciones de una forma muy precisa. Por ejemplo, la madre que ofrece el pecho al oír llorar a su hijo, etc., prueba que los organismos reconocen covariaciones entre estímulos, por lo que es lógico preguntarse si los resultados de Chapman y otros autores se refieren a covariaciones raras, precisamente aquellas para las que los sujetos no son eficientes. Nisbett y Ross (1980) han contestado que no: las covariaciones eficientes detectadas son la excepción de una regla más general que indica que las personas perciben de forma poco eficiente las covariaciones.

Preconcepciones sociales y error fundamental de atribución.

Frecuentemente se construyen hipótesis sobre comportamientos sociales, por ejemplo intentando adivinar cómo se comportará un amigo que acaba de saber que su mujer le ha abandonado. Al juzgar por disponibilidad, intentaremos dar respuesta a ese interrogante recordando sucesos pasados similares o recuperando instancias de como ese mismo amigo trató en el pasado las crisis, a fin de inferir cómo la tratará ahora. También podemos construir una realidad social con la que comparar el suceso actual.

Los sesgos provocados por la presencia de un atributo sobresaliente (lo coloreado, lo dinámico u otro estímulo distintivo que atrae desproporcionadamente la atención) afectan desproporcionadamente a los juicios. Es el caso del **error fundamental de atribución** (Ross, 1977). En una situación social en la que cualquier persona o cualquier variable situacional es candidata a ser una causa plausible, existe un sesgo que hace ver a las personas como agentes causales. Algunas personas son más llamativas que otras, y eso puede crear un sesgo en el juicio social. Vestidos, colores, etc. llaman desproporcionadamente la atención.

Los sesgos inducidos por las estructuras cognitivas (preconcepciones) se derivan del hecho de que las personas tienen estructuras resistentes para procesar la información, y que estas estructuras son utilizadas frecuentemente como hábitos cognitivos. Así, por ejemplo, los académicos discriminan entre las personas por la inteligencia, mientras que los aficionados al deporte consideran más las habilidades deportivas que la inteligencia. Las personas emplean sus construcciones y esquemas personales cuando se enfrentan a nuevas situaciones o cuando deben hacer predicciones acerca del futuro. La utilización de reglas, esquemas y construcciones personales pueden llevar a un sujeto a hacer inferencias que no haría otro individuo que no comparta las mismas estructuras cognitivas.

El fenómeno más representativo de estas circunstancias es el del estereotipo. Los estereotipos son una clase particular de expectativas que pueden funcionar como guía y dar forma a la realidad. Los estereotipos, una vez formados, pueden distorsionar la recogida y almacenamiento de información, y las impresiones consecuentes, produciendo sesgos. Se pueden realizar inferencias injustificadas sobre grupos sociales o individuos y perpetuarlas en ausencia de base empírica o ignorándola.

#3.7.4.- LA ILUSIÓN DE CONTROL

Las actividades que necesitan habilidad exigen una ligazón causal entre el comportamiento y el resultado, mientras que los resultados positivos aleatorios ocurren fortuitamente, no siendo controlable el éxito en esas condiciones. Esta distinción, sin embargo, no es reconocida generalmente, y en muchas situaciones de azar las personas se comportan como si fuera controlable. Las personas a menudo fallan a la hora de responder de forma diferenciada a los sucesos controlables e incontrolables. Ese fenómeno se conoce como ilusión de control, porque se supone que es un comportamiento derivado de la necesidad de control que tienen los individuos.

Goffman (1967) y Henslin (1967) realizaron observaciones sobre personas que tratan las cuestiones de azar como controlables. Goffman estudió los croupiers de la Vegas y Henslin los jugadores de dados. Estos sujetos creen que tienen algún control sobre los resultados de los juegos, y que el esfuerzo y la concentración dan buenos resultados. Se comportan como si un juego de azar fuera un juego de habilidad.

#3.7.5.- ESCENARIOS

Cuando se trata de estimar la probabilidad de sucesos únicos para el sujeto (la ocurrencia de una recesión económica o de un divorcio, el éxito de una operación quirúrgica, etc.), ésta no puede ser evaluada por un simple recuento de instancias. Se puede, sin embargo, aplicar también la heurística de la disponibilidad para evaluar la verosimilitud de los sucesos únicos, mediante la construcción de historias o escenarios. La construcción de escenarios es una actividad cotidiana dirigida al pronóstico de acontecimientos futuros, la explicación de sucesos pasados de origen incierto, o la elaboración de planes.

Los escenarios son historias que llevan de la situación actual al suceso meta. La plausibilidad, la facilidad o la dificultad con que pueden ser traídos a la mente esos escenarios, puede proveer una base para juzgar la verosimilitud del suceso meta. Si no se puede imaginar ningún escenario, entonces el suceso se considera altamente improbable. Pero si en una situación se imaginan muchos escenarios, o si uno es especialmente convincente, entonces el suceso se juzgará altamente probable. La probabilidad de un resultado se juzga por la facilidad con que se pueden construir uno o varios escenarios de acontecimientos concatenados que conduzcan a ese resultado.

Por ejemplo, sea un médico que teme que un paciente cometa suicidio. La probabilidad subjetiva que atribuirá a ese suceso posiblemente dependa de la similitud con los casos anteriormente experimentados que le vengan a la mente al examinar el caso actual. A veces solamente le vendrá a la mente una instancia relevante (memorable), con la que comparará el caso actual. En otras ocasiones, por el contrario, se recuerdan varias instancias, y habrá que sopesarlas en función de su similitud con el caso actual. Un primer problema que tendrá que dilucidar es cuál de los casos recordados debe tener en cuenta para considerar el caso actual: (1) los de pacientes que se parecen al que se está tratando, o (2) los pacientes que se suicidaron, o (3) los pacientes que se parecen al que se está tratando y que se suicidaron.

Desde un punto de vista actuarial, desde luego, la clase relevante es (1), y el estadístico relevante es la frecuencia de intentos de suicidio en esa clase.

Pero la búsqueda en la memoria puede seguir otras reglas. Como el intento de suicidio es un suceso dramático y sobresaliente, se recordará con más facilidad a los pacientes que cometieron suicido antes que a los pacientes depresivos que no intentaron suicidarse. Consecuentemente, el médico puede que compare el caso actual con los pacientes suicidas y juzgue la verosimilitud de que el paciente actual intente un suicidio por el grado de similitud que tenga con los casos recordados, produciéndose el correspondiente sesgo. Dado que los pacientes que se suicidaron sufrían una severa depresión, puede que incluso concluya que los pacientes que sufren una fuerte depresión se suicidarán. Contrariamente, puede que concluya que un paciente no se suicidará porque no se parece a los casos de suicidio que recuerda. Igualmente, puede pensar solamente en los pacientes que eran simultáneamente depresivos y suicidas.

En general, diversos estudios han mostrado que la contingencia entre dos variables (p.e. un síntoma y una enfermedad) se juzga por la frecuencia en la que coocurren, prestando poca atención a los casos en los que el síntoma o la enfermedad no están presentes.

Es importante el impacto de incidentes dramáticos sobre los escenarios (después de contemplar un accidente, se considera más probable que ocurra) o la preocupación continuada por un resultado (que puede incrementar su disponibilidad y ser percibido como más verosímil). Las personas están preocupadas por resultados deseados (ganar la lotería) o altamente no deseados (accidentes de avión), y la disponibilidad podría explicar por qué la ocurrencia de sucesos asociados a utilidades extremas (o desutilidades) pueden parecer más probables de lo que realmente son.

La producción de escenarios convincentes es en muchos casos, paradójicamente, una restricción al pensamiento o a la percepción futura. Existe una amplia evidencia de que una vez que se ha percibido o interpretado la incertidumbre de una determinada manera, es bastante difícil verla de otra. La generación de un escenario específico dificulta la visión posterior de otros escenarios, especialmente aquellos que dan resultados diferentes.

Falacia de la iniciativa.

La tendencia a considerar solamente los escenarios relativamente más simples puede tener efectos particularmente sobresalientes en caso de conflicto. Nuestro humor y planes son para nosotros más disponibles que los de nuestro oponente. El jugador tiende a considerar la estrategia de su oponente como relativamente constante e independiente de sus propios movimientos, incurriendo en la **falacia de la iniciativa** (una tendencia a atribuir menor iniciativa e imaginación al oponente que a sí mismo).

#3.7.6.- HEURISTICA DE SIMULACIÓN

En bastantes situaciones la probabilidad de los sucesos se establece mediante una operación similar a un modelo de simulación. Una simulación no produce una sola historia, que comienza y acaba con un resultado definitivo, sino que se construye el resultado de la simulación como un enunciado que refleja la facilidad con que un modelo puede producir diferentes resultados, dadas las condiciones iniciales y los parámetros de operación. La simulación mental produce una medida de la propensión de nuestro modelo de la situación a generar varios resultados, de igual forma que las tendencias de un modelo estadístico pueden ser establecidas por las técnicas de simulación Monte Carlo. La facilidad con la que se produce la simulación de un sistema en un estado particular se utiliza eventualmente para juzgar la propensión del sistema (real) a producir ese estado.

En el contexto de predicción y planificación bajo incertidumbre, la manipulación deliberada de los modelos mentales parece ser suficientemente importante para que se haya propuesto una heurística de la simulación.

Las alteraciones que las personas introducen en las historias se pueden clasificar en "cuesta abajo", "cuesta arriba" o en "cambios horizontales". Un cambio "cuesta abajo" (downhill) es aquel que elimina aspectos inesperados de la historia, incrementando su coherencia interna. Un cambio "cuesta arriba" (uphill) es el que introduce sucesos improbables. Un cambio horizontal es aquel en el que se toma un valor arbitrario de una variable y se reemplaza por otro valor arbitrario, que no es ni más ni menos probable que el anterior.

Un buen escenario es aquel que cubre la diferencia entre el estado inicial y el suceso meta mediante una serie de sucesos intermedios, con una tendencia general *downhill* y escasa utilización de la *uphill*. La evidencia sugiere que la plausibilidad de un escenario depende mucho más de la plausibilidad de sus ligazones fuertes que del número de ligazones. Un escenario es especialmente satisfactorio cuando el camino que lleva del punto inicial al final no es inmediatamente evidente, de forma que la introducción de pasos intermedios realmente incrementa las probabilidades subjetivas del suceso meta.

Cualquier escenario es esquemático e incompleto y las uniones son sucesos escasamente redundantes, pero muy significativos causalmente. Un suceso no redundante representa un mínimo local en la predictibilidad de la secuencia, un punto en el que se pueden alcanzar alternativas significativas. Un suceso causalmente significativo es el que al ocurrir altera los valores que son considerados normales por otros sucesos en la cadena que lleva al objetivo del escenario.

Tendemos a considerar un resultado improbable si puede ser alcanzado solamente invocando hipótesis de tipo *uphill,* sucesos raros y coincidencias extrañas.

D. Kahneman y A. Tversky (1984, 14) avanzaron las siguientes hipótesis:

(1) La búsqueda de uniones no redundantes y causalmente significativas en la construcción de escenarios puede conducir a sesgos en los escenarios, donde sucesos dramáticos marcan transiciones causales. Existirá en general una tendencia a subestimar la verosimilitud de los sucesos que son producidos por cambios incrementales bajos.

(2) La utilización de escenarios para asignar probabilidades puede ocasionar sesgos a favor de los sucesos para los que se puede imaginar un escenario plausible. En estos escenarios, la heurística de la simulación es más propensa a considerar los cambios *downhill*. Estos sesgos son especialmente perniciosos en el contexto de planificación, porque pueden incrementar el optimismo sobre la probabilidad de realización del plan. Por su propia naturaleza, un plan es una cadena de ligazones plausibles. Cualquier punto de la cadena es sensible a la expectativa de que los sucesos se realizarán como han sido planificados. Sin embargo, la probabilidad acumulada de al menos un fallo fatal podría tener mucha importancia, aunque la probabilidad de cada causa individual de fallo sea despreciable. Los planes fallan por las sorpresas, por las ocasiones en que se presentan cambios *uphill*.

#3.7.7.- EXPERIENCIA Y EXCESO DE CONFIANZA

La heurística de la disponibilidad puede hacer que se perciba un riesgo mayor en sucesos objetivamente de baja probabilidad, por efecto de los recuerdos o de la imaginación. Este es el origen, por ejemplo, de las falsas concepciones y decisiones que la gente adopta en relación con los riesgos naturales.

Lichtenstein et al. (1978), en un estudio clásico, estudiaron la percepción por las personas de la frecuencia de varios sucesos letales. Se indicó a los sujetos (estudiantes universitarios y mujeres de la L.W.V.) que 50.000 personas morían al año en USA a causa de los accidentes de tráfico, y se les preguntaba que estimasen la frecuencia de otras 40 causas de muerte. En otro estudio se preguntaba cuáles de dos causas de muerte eran más frecuentes. En ambos estudios los juicios fueron relativamente buenos, indicando que las personas conocen cuáles son los sucesos más y menos frecuentes. Sin embargo, los sujetos sufrieron desviaciones que reflejaban la influencia de la disponibilidad: en general, los casos raros fueron sobrestimados y los más comunes subestimados.

La heurística de la disponibilidad exagera la importancia de la experiencia vital como determinante del riesgo percibido, ya que se infiere en muchas ocasiones como si se considerara que la experiencia no está expuesta a sesgos, y que por tanto la percepción basada en ella es muy probable que sea correcta. Claramente este supuesto es una fuente de sesgos muy importante.

Mucha de la información a la que la gente está expuesta le ofrece una imagen distorsionada de los peligros del mundo. Combs y Slovic (1979) examinaron las causas de muerte reflejadas en dos periódicos de las costas Este y Oeste de USA durante 1 año, encontrando que los periódicos, en su información sobre sucesos, contenían sesgos similares. Las causas más frecuentes de muerte apenas fueron citadas en los periódicos, prestando atención solamente a las causas violentas o catastróficas.

También ocurre que aunque las personas juzguen relativamente bien, sin ser especialistas, la frecuencia de las distintas causas de muerte, tienden a considerarse a sí mismos **inmunes al azar**. La mayoría de individuos cree que él mismo es mejor conductor que la media de los conductores, más parecido a los sujetos que han vivido más de 80 años, y menos expuesto al daño producido por los productos que utiliza. De esta forma los riesgos se reducen en la perspectiva de la experiencia de cada sujeto. Esto explica que, en ocasiones, tiendan a rechazar tomar medidas protectoras (p.e. ponerse los cinturones de seguridad).

En algunas situaciones la incapacidad para apreciar los límites de los datos “disponibles” lleva a la gente a un exceso de confianza o complacencia. En un árbol de decisión, las ramas que no se representan tienden a ser ignoradas (no visto, no pensado).

Un aspecto especialmente pernicioso de la heurística es que la gente normalmente tiene gran confianza en los juicios basados en su propia experiencia. En experimentos con apuestas, cuanto más injustificado estaba el exceso de confianza, los sujetos incrementaban las apuestas. Incluso sometidos a fuertes pérdidas, seguían manifestando exceso de confianza. Aunque las bases sicológicas para la certidumbre injustificada son complejas, un elemento clave de explicación es que la gente no comprende que su conocimiento se basa en supuestos débilmente fundamentados.

Los expertos están tan expuestos al exceso de confianza como las personas corrientes. Un experimento en el que se juzgaban las posibles causas de fallos en el encendido de los automóviles fue sometido a mecánicos y universitarios, obteniéndose resultados parecidos en ambos colectivos. Hynes y Vanmarcke (1976) plantearon un problema de su especialidad a 7 geólogos y no obtuvieron una respuesta mejor que la que podría obtenerse de otros sujetos.

#3.8.- ANCLAJE

El anclaje es un efecto particular de la accesibilidad de una información irrelevante presente en la situación de prueba o generada por el propio sujeto con motivo de un cómputo incompleto. Este anclaje a una información poco fiable e irrelevante es causa de errores y sesgos, y es especialmente frecuente cuando se deben realizar juicios difíciles mediante introspección: cuando se nos ocurre algún valor particular (o es sugerido por otro), se tiende a anclar ese valor.

#3.8.1.- AJUSTE INSUFICIENTE

En un experimento donde se preguntaba a los participantes sobre el valor de una determinada cantidad (p.e. el porcentaje de países africanos en la ONU), se obtenía en su presencia aleatoriamente un número entre 0 y 100 en una ruleta. El encuestado debía decir si su estimación estaba por encima o por debajo de ese número, y después se ajustaba su estimación horquillando. Por término medio, los sujetos que vieron el número 10 en la ruleta estimaron un porcentaje del 25%, mientras que los que vieron el 65 estimaron un 45%, y en todos los casos el número inicial extraído al azar tuvo un efecto importante en el porcentaje dado por el grupo. La introducción de pagos por la obtención de buenas estimaciones no redujo el sesgo de anclaje.

El efecto de anclaje no ocurre solamente cuando se suministra el número inicial al sujeto encuestado, sino también cuando se obtiene mediante un cálculo insuficiente. Así, por ejemplo, distintos grupos de estudiantes evaluaron la expresión 8 x 7 x 6 x 5 x 4 x 3 x 2 x 1 y 1 x 2 x 3 x 4 x 5 x 6 x 7 x 8 en pocos segundos. Los sujetos realizaban unos pocos cálculos iniciales y, por falta de tiempo, estimaban la respuesta mediante extrapolación o ajuste. En general se subestimó el resultado, pero como el resultado del reducido número de multiplicaciones que daba tiempo a realizar era mayor en la secuencia decreciente que en la creciente, se daban valores mayores a la primera (2250) que a la segunda (512). La respuesta correcta es 40 320.

En cuestiones del tipo: ¿Cuántos coches extranjeros se importaron en USA en 1968?

(a) Establezca una cifra alta de tal forma que sólo exista una probabilidad del 1% de que la cifra real sea superior a la dada por usted.
(b) Realice una estimación baja, de tal forma que exista una probabilidad de solamente un 1% de que la verdadera respuesta esté por debajo de su estimación.

En teoría estas preguntas determinan intervalos de confianza del 98%. Como resultado se encontró que del 40 al 50 % de las respuestas no incluían en el verdadero intervalo de la respuesta. En bastantes casos se debe a que esto se pregunta a personas que no son ni expertos sustantivos ni normativos.

Johnson (1972) y Woodworth (1938), entre otros, han mostrado que en una amplia variedad de juicios, incluyendo la mera traslación de los valores de una escala a otra, los sujetos tienden a evitar las respuestas extremas y a reducir la variabilidad de sus juicios (Stevens y Greenbaum, 1966). Debido a este sesgo, los juicios serán regresivos, al comparar la evidencia y las predicciones (“inputs” y “outcomes”).

Igualmente, los expertos tienden a describir en los ensayos de laboratorio distribuciones simétricas, siguiendo el modelo de la normal. Parece como si, sicológicamente, la simetría fuera una forma poderosa de organización conceptual. Esto es sobre todo verdad con experimentos de laboratorio donde los sujetos no están familiarizados con los mecanismos que subyacen en la generación de las probabilidades que deben determinar.

#3.8.2. - EVALUACIÓN DE SUCESOS CONJUNTOS Y DISJUNTOS

Bar-Hillel (1973) planteó un experimento típico consistente en invitar a un sujeto a realizar la estimación de la probabilidad de (a) sucesos simples (obtener una bola roja de una urna que contiene 50 bolas rojas y 50 blancas); (b) sucesos conjuntos (p.e. obtener una bola roja 7 veces consecutivas, en una extracción con reemplazamiento); y (c) sucesos disjuntos (p.e. probabilidad de obtener al menos una bola roja en 7 extracciones consecutivas con reemplazamiento). En general los sujetos tienden a sobrestimar las probabilidades de los sucesos conjuntos y a subestimar la probabilidad de los sucesos disjuntos.

Como el ajuste desde el punto de partida es normalmente insuficiente, la probabilidad final es estimada como próxima a la probabilidad de los sucesos elementales. La probabilidad de un suceso conjunto tiende a ser menor que la de cada suceso elemental, por lo que en este caso se sobrestima la probabilidad correcta. Lo contrario ocurre con sucesos disjuntos.

Este fenómeno tiene una gran influencia en las estimaciones realizadas en proyectos (p.e. el desarrollo de un nuevo producto), en los que existen muchas tareas realizadas secuencialmente. En general, las personas tienden a realizar estimaciones muy optimistas, debido a que la probabilidad de las conjunciones tiende a ser sobrestimada. Lo contrario, la subestimación del riesgo, suele ocurrir en problemas con una estructura de actividades paralelas (p. e. en la evaluación de riesgos en una central nuclear, donde existen muchos sistemas, cada uno de ellos con una probabilidad baja de fallo, la unión de las diferentes posibilidades puede dar origen a una probabilidad significativa de fallo general).

#3.8.3.-ANCLAJES EN LA ESTIMACIÓN DE DISTRIBUCIONES DE PROBABILIDADES

Cuando se pregunta a un experto sobre la distribución de un determinado valor que se desea estimar, frecuentemente se determinan percentiles para los diferentes valores, que expresan la probabilidad subjetiva que se asigna a los diferentes valores. Por ejemplo, se puede preguntar por un valor X_{90}, tal que su probabilidad subjetiva de que este número sea mayor que el valor de interés (el precio del trigo el año que viene, por ejemplo) sea 0,9.

En base a las distribuciones establecidas por un experto es posible realizar una calibración del mismo. Un experto está adecuadamente calibrado si un p % de los verdaderos valores caen en el rango establecido para Xp. Por ejemplo, los verdaderos valores deberían caer por debajo de $X_{0.01}$ un 1% de los casos, y por encima de $X_{0.99}$ un 1% de los casos. De esta forma, el 98% de los casos deberían caer entre $X_{0.01}$ y $X_{0.99}$.

En muchos estudios empíricos (Alpert y Raiffa, 1969, 1984; Staël von Holstein, 1971; Winkler, 1967a) se han encontrado desviaciones grandes y sistemáticas de los valores compatibles con una adecuada calibración. Los expertos en general cierran las distribuciones, reducen la varianza, mostrando una mayor confianza en sus estimaciones de las que está justificada por su conocimiento de las cantidades predichas. Este sesgo es muy general, y no suele corregirse con la introducción de incentivos para mejorar la calibración.

Para estimar el valor $X_{0.90}$ del precio del trigo el año que viene, es natural comenzar estimando el valor que se considera la mejor estimación, y después ajustar ese valor hacia arriba para establecer el valor $X_{0.90}$. Como este ajuste es insuficiente, $X_{0.90}$ no será suficientemente extremo. Algo semejante ocurre en la estimación de $X_{0.10}$. La consecuencia es que el intervalo entre $X_{0.1}$ y $X_{0.9}$ será mas corto que el real, y la función de densidad será más cerrada.

Las distribuciones de probabilidades subjetivas para una determinada cantidad (p.e. el precio del trigo el próximo año) se pueden obtener de dos manera distintas:

- (1) Indicando al sujeto que seleccione los valores del precio que se corresponden con los percentiles especificados de su distribución de probabilidad.

- (2) Indicando al sujeto que asigne la probabilidad de que un determinado valor del precio sea superior a un valor dado.

Los dos procedimientos son formalmente equivalentes y deberían dar las mismas distribuciones. Sin embargo cada procedimiento sugiere diferentes puntos de anclaje. En (1) el ancla es la mejor estimación de la cantidad cuya distribución se trata de determinar; y en (2) el ancla puede ser el valor establecido en la pregunta. Alternativamente podría utilizar la estimación 50 - 50 (la mediana) como ancla, que es un punto natural para la estimación de la verosimilitud de un valor. En cualquier caso el procedimiento (2) debería sugerir apuestas (odds) menos extremas que el procedimiento (1).

#3.8.4.-ESTIMACIÓN CONSERVADORA

En general, las personas perciben cada dato correctamente y son capaces de ver el diagnóstico que se deriva de él, pero tienen dificultades para combinar ese diagnóstico con el ofrecido por otros datos a fin de revisar sus opiniones. Esta revisión insuficiente de las opiniones a la luz de la nueva evidencia se conoce como estimación conservadora.

Sean dos urnas, una conteniendo 700 fichas rojas y 300 azules (urna R), y la otra 300 rojas y 700 azules (urna A). Se elige una de las urnas lanzando una moneda al aire. Se extraen de la bolsa elegida 12 fichas y resulta que 8 son rojas y 4 azules. ¿Cuál es la probabilidad de que se haya elegido la urna R? Sin hacer operaciones, la mayoría de la gente asigna una probabilidad entre 0.7 y 0.8. Pero la probabilidad calculada es 0.97. Ni siquiera personas familiarizadas con el teorema de Bayes suelen hacer una estimación próxima a ella.

CAPÍTULO 4

PLANTEAMIENTO DEL TRABAJO EXPERIMENTAL Y PRINCIPALES RESULTADOS

#4.1.- PLANTEAMIENTO DEL TRABAJO EXPERIMENTAL

Un esquema de ayuda a la toma de decisiones debe tener en cuenta una cadena potencial de competencias que va desde los expertos a los agricultores. Se supone que los expertos disponen de conocimientos científico – técnicos, series de datos históricos, fuentes de información, etc. que podrían ser relevantes para completar la información empírica acumulada por los agricultores en sus explotaciones concretas y la adquirida en su contacto con otros agricultores, con los medios de comunicación, con los agentes de extensión o comerciales, etc.

Todo el conjunto de percepciones o interpretaciones de la realidad discutida, el significado de los conceptos, el planteamiento de los problemas, la fijeza mental en la búsqueda de las soluciones, las convenciones, los sesgos, los prejuicios, y demás elementos sicológicos y culturales de cada uno de los grupos y de los sujetos, tendrá su reflejo en la interacción que se estudia.

En el esquema del proceso de razonamiento en dos etapas, como se ha discutido en el capítulo anterior, J. St. B. T. Evans (1989) sugirió una primera etapa heurística, muy sensible, entre otros, a los procesos de comprensión lingüística, y una segunda etapa analítica que se supone más dirigida por concepciones metodológicas. Si bien en ambas etapas se pueden producir errores y sesgos, se comprende que en la primera etapa pueden ser importantes aquellas cuestiones derivadas de problemas en los supuestos relacionados con la comunicación, con la formulación de las preguntas y de las respuestas, con la interpretación de ambas, con los malos entendidos, etc. Es por tanto importante determinar el nivel de comprensión de los términos y esquemas empleados en la interrogación, y cómo son comprendidos e interpretados en cada uno de los eslabones de la cadena humana descrita.

Por otra parte, de los resultados discutidos en el capítulo 2, sobre el establecimiento de funciones de densidad, se sabe:

(a) que los resultados dependen en gran parte de la comprensión por parte de los sujetos de la formulación de los problemas y de la metodología que se trata de aplicar;

(b) que los conceptos estadísticos implicados no se puede suponer que formen una base común compartida por todas las personas, incluso:

(b1) después de una formación técnica en ámbitos distintos a la Estadística,

(b2) en los ámbitos estadísticos (ya que los sujetos entrenados en estos ámbitos comprenden bien los métodos, pero entienden con dificultad la formulación de los problemas en otros campos o desconocen valores comúnmente compartidos por los expertos de otras áreas o ignoran la correspondencia problema – método); y

(c) que el entrenamiento de los sujetos mejora su capacidad para el establecimiento de las mismas.

En función de lo anterior, es razonable suponer que, incluso antes de que se puedan manifestar los sesgos discutidos en el capítulo 3, es necesario investigar los problemas de comunicación que se pueden establecer en las primeras etapas de obtención de la información y los supuestos que, pensándose generales, no son compartidos.

La determinación de las funciones de densidad de probabilidad de los rendimientos, por ejemplo, supondrá un esquema de comunicación ordenador – agricultor, semejante, por un lado, al estudiante de últimos años – agricultor para las tareas más rutinarias, y al experto agrónomo – agricultor en cuestiones de incorporación de información nueva adicional.

Es pertinente, por tanto, experimentar:

- qué supuestos implícitos se manifestarán en la cadena de comunicación en los primeros pasos para la determinación subjetiva de las funciones de densidad de probabilidad de un hipotético sistema de ayuda a la toma de decisiones,

- y cuál es la calidad de los resultados obtenidos en función de las diferentes técnicas empleadas.

Los resultados se consideran de gran importancia para el diseño de la interfaz ordenador – agricultor, es decir, de la forma de interrogar al agricultor y a los expertos, las palabras y los conceptos implicados en esa actividad.

El trabajo experimental se ha centrado en esos problemas, y específicamente en las tareas de obtención de funciones de densidad de probabilidad de los rendimientos de cultivos anuales y de la alfalfa.

La razón de incluir la alfalfa es la de explorar las respuestas en un cultivo que puede servir de transición al estudio de los mismos problemas en los cultivos plurianuales herbáceos, que será la continuación lógica de los trabajos emprendidos en esta Tesis. En lo sucesivo, sin embargo, nos referiremos al conjunto de los cultivos estudiados como “anuales”.

En los cuestionarios también se han incluido preguntas relacionadas con el comportamiento de los agricultores en el uso de los insumos, problema sobre el que se desea disponer de alguna información previa. La razón es que, posiblemente, a la hora de desarrollar el sistema de ayuda a la toma de decisiones, haya que tener en cuenta aspectos relacionados con lo que se puede aprender de esa investigación, aunque sólo sea para facilitar su adaptación a las prácticas culturales de los potenciales usuarios.

El problema de la variabilidad de las “recetas” de insumos se plantea en torno a la siguiente cuestión: Si se resuelve un problema de programación en un instante t_0, y como consecuencia se obtiene una formulación F_t para el conjunto de variables decisión que el agricultor aplica sistemáticamente (una “receta” relativamente fija), este sistema se puede considerar estático. Pero si, a causa de que el conjunto de operaciones de F_t se aplica a lo largo de un período dilatado de tiempo, y las circunstancias en un instante t_1 permiten mejorar la previsión, parece razonable ajustar F_t a la nueva información o a las nuevas expectativas, y entonces el esquema tendrá una fuerte componente dinámica.

Aunque éste no es el problema que se analiza en esta Tesis, se han incluido algunas preguntas en la dirección comentada para obtener una impresión preliminar.

#4.1.1.- HIPÓTESIS

El trabajo experimental de esta Tesis está principalmente relacionado con la determinación de funciones de densidad de probabilidades (f.d.p.) de rendimientos de cultivos anuales a partir de la información subjetiva disponible en expertos y agricultores, y más específicamente con la evaluación de algunas técnicas desde el punto de vista de su comprensión y facilidad de utilización. El punto de vista predominante es que los resultados deben servir para el diseño de interfaces de comunicación de un programa de ordenador en un sistema de ayuda en la toma de decisiones de planificación de la producción.

Para contribuir al objetivo comentado, se ha planteado la obtención de f.d.p. de rendimientos de cultivos anuales mediante encuestas a agricultores y a expertos, y concretamente la determinación de su forma. Recuérdese que el interés no es tanto la descripción de la forma como la identificación de los problemas de interpretación en la metodología empleada.

La hipótesis de partida es que existe un conjunto de términos del lenguaje ordinal que facilitan la asignación de frecuencias, por tener los sujetos codificada la información en torno a esos términos (accesibilidad). En el caso de los rendimientos, se supone que se organizan en torno a conceptos como “rendimiento muy bueno”, “rendimiento malo”, etc.

En el caso particular de los rendimientos esa escala es ordinal, y por tanto la asignación de frecuencias a cada clase o intervalo nos daría una imagen de la f.d.p. referida a la misma. Aunque se puedan cuestionar las comparaciones entre las formas y parámetros de las diferentes estimaciones, se ha discutido la forma y parámetros de las f.d.p. en las escalas ordinales. Implícitamente se acepta la hipótesis de que en un mismo medio ecológico, las escalas ordinales serán traducidas en escalas numéricas indistinguibles prácticamente. En esta Tesis no se ha planteado la contrastación de esta hipótesis, que exige la obtención de una gran cantidad de encuestas en un mismo medio ecológico para cultivos concretos, una estrategia distinta a la seguida.

La segunda hipótesis tiene que ver con la idea de que los sujetos son capaces de traducir su escala ordinal a una escala numérica. No se ha explorado sistemáticamente la validez de esta hipótesis en esta Tesis, pero se piensa que puede aceptarse sin mayor discusión, ya que no es un problema más complicado que el de la determinación de funciones de utilidad, en el que existe suficiente evidencia sobre la capacidad de los sujetos para realizar esas traducciones. Por otra parte, dado que los rendimientos son observados en una escala numérica, si los sujetos no fueran capaces de hacer esas traducciones sin dificultad no podrían codificar sus resultados en escalas ordinales, como es evidente que sí hacen.

Como consecuencia de la traducción de las escalas ordinales a escalas numéricas, se dispone de una estimación de la f.d.p. Como el número de clases o intervalos ordinales no será en general muy extenso (y en esta Tesis se ha trabajado con cinco intervalos), se conjetura que posiblemente se hayan producido truncamientos de las colas. Aunque algunos autores señalan la importancia de las colas, como se ha puesto de manifiesto en el capítulo 2, se estima que en el contexto de programación de cultivos posiblemente el truncamiento no sea tan importante en sus consecuencias prácticas. Esta es una cuestión que se debe investigar, pero existe evidencia empírica suficiente para aceptar la tesis de que no son importantes las colas en las funciones de densidad de rendimientos (y, por otra parte, existen mecanismos institucionales para paliar sus manifestaciones, como la declaración de zonas catastróficas, etc.).

Debido a lo anterior, la caracterización de las funciones mediante las encuestas se ha realizado describiendo su forma, especialmente a través de la determinación del número y posición de las modas y de la posición de la mediana.

Una segunda técnica de determinación de f.d.p. se basa en una asignación gráfica directa de frecuencias. Sobre un papel se señalan un número determinado de intervalos, con sus límites definidos. Se indica al sujeto que deposite sobre cada intervalo un número de monedas (o de garbanzos u otro indicador) proporcional a la frecuencia con que se presenta cada intervalo de rendimientos. Este método había sido empleado por W. Grisley y E. D. Kellogg (1983), tal como se ha discutido en ζ 2.11.

En lo que se refiere al sistema de recompensas (reglas de puntuación o “scoring rules”) creemos que, en un sistema de ayuda a la toma de decisiones, no tiene sentido afirmar que el sujeto no tiene incentivos para declarar sus verdaderas convicciones. Por tanto, nos interesan los incentivos morales, que se han intentado crear mediante la existencia de algún tipo de relación entre el encuestado y el encuestador. Puede que la respuesta no refleje exactamente las convicciones del sujeto encuestado, pero nuestro objetivo no es determinar los valores, sino detectar los problemas de comprensión o de interpretación o de adecuación de los métodos de obtención de la información.

Esta técnica se empleó, con garbanzos en lugar de monedas, en las entrevistas en profundidad a los dos grupos de agricultores entrevistados por un postgraduado y por la autora de la Tesis.

Es evidente que los agricultores, en ausencia de otros incentivos, pueden tener interés en expresar opiniones distintas de las que creen. Por ejemplo, un agricultor puede entender que mediante la entrevista se juzga su calidad como empresario, y declarar rendimientos sistemáticamente mayores a los realmente obtenidos, o frecuencias mayores para los rendimientos superiores. Igualmente, si desea transmitir el mensaje de que la agricultura necesita ayudas, puede que declare rendimientos inferiores a los reales o que exagere la influencia de los rendimientos inferiores. Pero también es posible que entienda que le entrevista un técnico que necesita información precisa para la realización de estudios que son importantes para él, y que por razones de simpatía u otras el agricultor decida cooperar.

Como la hipótesis de la fuerza de los incentivos morales necesita bastante trabajo experimental para ser aceptada y determinar sus límites de validez, se avanzó en esa dirección mediante la realización de las entrevistas en varios días. Implícitamente se admite que lo declarado un día será recordado de forma poco precisa una o dos semanas después. Si se detectan diferencias sustanciales entre los valores dados a una misma variable en dos momentos distintos, indagando sobre las diferencias se pueden detectar las causas de la divergencia (no se expresan sinceramente las verdaderas convicciones, no se conoce con suficiente detalle aquello que se contesta, etc.)

En el contexto que nos interesa de utilización de los resultados, es evidente que no serán de utilidad las técnicas basadas en ofrecer premios monetarios por la expresión sincera de las convicciones, ya que este problema no se planteará, porque en general coincidirá el sujeto que informa al sistema con el interesado en los resultados. En la obtención e intercambio de información, sin embargo, puede que los problemas de los incentivos recobren importancia. En cualquier caso, es interesante "per se" la exploración de los sistemas de incentivos morales.

Una tercera técnica tiene que ver con los trabajos experimentales desarrollados en otros ámbitos, como el PERT, para la estimación de valores medios y de varianzas, a partir de la estimación de los valores mínimo y máximo del recorrido de la variable aleatoria y del más frecuente o moda. Como es conocido, dados esos valores, a partir de un conjunto de simplificaciones, se pueden calcular los valores comentados por aproximación a una función triangular o a una beta. Si se pudiera tener confianza en que los agricultores y los expertos realizan buenas estimaciones de estos valores, y que al menos son tan buenas como sus estimaciones de los valores medios, se abriría un campo importante, útil al menos para contrastar valores de las funciones de densidad establecidas por otros métodos.

#4.1.2.- ENCUESTAS

En el año 1997 se realizaron dos tipos de encuestas dirigidas a detectar las dificultades a la hora de construir f.d.p. de rendimientos de cultivos anuales (y alfalfa) a partir de información ofrecida por los agricultores y los expertos.

La primera encuesta (Anejo 1) se dirigió, por correo y mediante entrevistas personales, a expertos en fisiología y producción vegetal, mediante un cuestionario cerrado. Se enviaron 185 cuestionarios y se recibieron 38 respuestas. El detalle de las encuestas enviadas puede consultarse también en el mismo Anejo 1.

La segunda encuesta se dirigió a una población de agricultores que se consideraban representativos de los de la zona de Les Garrigues, en la provincia de Lleida. El cuestionario era relativamente abierto (Anejo 2), al objeto de determinar hasta qué punto la información de que disponían los agricultores era relevante para la determinación de las funciones de densidad y ayudar en la confección de futuros cuestionarios. Por ello las principales preocupaciones eran determinar:

- la estabilidad y coherencia de las respuestas;
- la posible influencia del entrevistador.

Estas preocupaciones sugirieron la realización de la encuesta en varios días y por dos encuestadores diferentes: la autora de la tesis y un postgraduado específicamente contratado para ello.

En función de los resultados, se realizó un conjunto de encuestas a los agricultores a partir de 1998 – 99. Como encuestadores actuaron estudiantes de segundo ciclo de las carreras de Ingeniero Agrónomo e Ingeniero de Montes. Se supone que es una población que puede representar bastante bien a los sistemas lógicos - formales de obtención de información mediante la interrogación a sujetos con experiencia práctica.

Los estudiantes entrevistaron a los agricultores en los cursos 1998 – 99 (Anejo 7) y 1999 – 00 (Anejo 8), por lo que se dispone de dos conjuntos de encuestas, correspondientes a esos dos años, el primero compuesto por 52 encuestas y el segundo por 44 encuestas.

Cada encuesta dirigida a los agricultores, constaba de dos cuestionarios (encuesta del primer día y encuesta del segundo día), que aunque recogían información similar, para comprobar la estabilidad de las respuestas, se diseñaron de tal forma que no le pareciese al agricultor que el segundo era una clara repetición del primero.

Un agricultor era sometido al primer cuestionario por un estudiante, y, al menos una semana después, volvía a ser encuestado con el segundo cuestionario por otro estudiante distinto.

Entre el agricultor y alguno de los dos estudiantes que lo entrevistaban sucesivamente existía algún tipo de relación (familiar, vecinal, etc.), a fin de favorecer la veracidad de las respuestas, fomentando las motivaciones morales en el encuestado. El incentivo de los estudiantes era material: la correcta realización de la encuesta suponía una parte de la calificación de una asignatura troncal.

De esta forma la realización de estas encuestas (denominadas “encuesta del primer día” y “encuesta del segundo día”) durante los dos cursos académicos supuso la realización de 52 + 52 = 104 entrevistas a 52 agricultores distintos durante 1998 - 99, y 44 + 44 = 88 entrevistas a 44 agricultores distintos durante 1999 – 00.

Para justificar la hipótesis de que los estudiantes, cuando se ven forzados a dividir un recorrido en diferentes intervalos, tienden sistemáticamente a darles la misma amplitud a todos los intervalos, se realizaron dos encuestas, con formulaciones distintas, durante el curso 1999 – 2000.

#4.2.- PROBLEMAS COMUNES EN LAS RESPUESTAS E INTERPRETACIONES

En las diferentes encuestas, algunas preguntas han dado origen a respuestas que han sido problemáticas de analizar, en unos casos sorprendiendo a los autores de la encuesta y en otros como respuesta a un experimento planteado.

Debido a ello se han debido realizar clasificaciones o se han tenido que definir hipótesis que se aplican a los distintos cuestionarios, pero cuya discusión sería tediosa de realizar en cada ocasión. Por ello se considera de interés discutir estos problemas inicialmente, para dejar sentado el planteamiento del problema, las interpretaciones sugeridas y las hipótesis introducidas en el análisis. De esta forma, se puede analizar el contenido de las distintas encuestas haciendo referencia al marco discutido en esta sección.

#4.2.1.- PROBLEMAS EN LA DETERMINACIÓN DE LOS RECORRIDOS

Supongamos que tratamos de determinar la función de densidad de probabilidad (f.d.p.) de los rendimientos de un cultivo.

En algunas preguntas se ha definido el recorrido de la f.d.p. mediante una escala ordinal, de tal manera que se identificaban cinco intervalos ordinales: rendimientos "muy malos", "malos", "normales", "buenos" y "muy buenos". A esta forma de definir los intervalos con frecuencia la hemos denominado "definición mediante lenguaje natural" o "escala ordinal".

Una hipótesis implícita en el anterior planteamiento es que las personas disponen de una atribución de frecuencias a cada una de esas clases, aunque sea de forma aproximada. De hecho, es un dato de experiencia que muchas personas expresan la naturaleza de los resultados de un año utilizando términos semejantes, en sentencias del tipo "este año ha sido muy bueno", "o las lluvias de este año han arruinado la cosecha, y tendremos un mal año", etc.

Posteriormente, en otra pregunta, se ha interrogado a los sujetos sobre qué valores de rendimientos asignarían a cada clase definida en la anterior escala ordinal. Se trata de buscar una regla de correspondencia o traducción de la escala ordinal a otra cardinal, mediante la técnica de atribuir un intervalo de variación de los rendimientos a cada intervalo ordinal. Por ejemplo, el agricultor podría concretar que considera "malos" los rendimientos de cebada de secano comprendidos entre 600 y 800 kg/ha.

Francamente, a la hora de plantear este tipo de experimento se pensaba que las principales dificultades se derivarían de cuestiones relacionadas con la vaguedad del lenguaje natural y la propia incertidumbre asociada a ese tipo de transformaciones.

Normalmente se preparaba una serie de preguntas que implicaban:

- determinar el recorrido de los rendimientos del cultivo considerado (R_{min}, R_{max});
- atribuir frecuencias a los cinco intervalos ordinales: rendimientos "muy malos", "malos", "normales", "buenos", "muy buenos";
- establecer una correspondencia entre esos cinco intervalos ordinales y sus correspondientes estimaciones cardinales.

Por tanto confiábamos en que los sujetos establecerían correspondencias como las siguientes:

Tabla 3.

Correspondencia entre las escalas ordinal y numérica.

Rendimientos de un cultivo	
Escala ordinal	**Escala numérica**
Muy malo	$(R_{min} \pm \varepsilon , r_1 \pm \varepsilon)$
Malo	$(r_1 \pm \varepsilon, r_2 \pm \varepsilon)$
Normal	$(r_2 \pm \varepsilon, r_3 \pm \varepsilon)$
Bueno	$(r_3 \pm \varepsilon, r_4 \pm \varepsilon)$
Muy bueno	$(r_4 \pm \varepsilon, R_{max} \pm \varepsilon)$

donde $\pm\ \varepsilon$ indica el nivel de incertidumbre o de confusión en el establecimiento de valores cuantitativos en la proximidad de las fronteras ordinales.

Sin embargo aparecieron otros problemas relacionados con la forma en que los sujetos encuestados (expertos y agricultores) conciben e interpretan los intervalos, que debe ser considerada seriamente en el diseño de los futuros interfaces de comunicación.

Normalmente, en las encuestas realizadas se obtuvieron cuatro tipos básicos de respuestas al traducir la escala ordinal a otra numérica:

- **Respuestas de la Clase I**. En este caso, los sujetos dieron un intervalo numérico para cada clase definida mediante lenguaje natural, definiendo con precisión la frontera inferior y la superior de cada intervalo.

- **Respuestas de la Clase II**. Las respuestas se dieron en forma de intervalos numéricos, pero los intervalos inferior y superior se dieron abiertos. Por ejemplo, diciendo que rendimientos "*muy malos*" son aquellos que son inferiores o iguales a 1500 kg/ha y "*muy buenos*" mayores o iguales a 5000 kg/ha.

- **Respuestas de la Clase IIa**. Solamente el intervalo inferior se dio abierto. Así, se dice que los rendimientos *"muy malos"* son aquellos iguales o inferiores a una cota, pero para los *"muy buenos"* se da el intervalo completo.

- **Respuestas de la Clase IIb**. Solamente el intervalo superior se dio abierto. Así, se dice que los rendimientos *"muy buenos"* son aquellos iguales o superiores a una cota, pero para los *"muy malos"* se da el intervalo completo.

- **Respuestas de la Clase III**. En este caso el sujeto ofrece un único punto como respuesta para cada clase.

- **Respuestas de la Clase IV**. Las respuestas son puntuales, pero las clases inferior y superior se dan abiertas (como en la clase II).

- **Respuestas de la Clase IVa**. Las respuestas son puntuales, pero la clase inferior se da abierta (como en la clase IIa).

- **Respuestas de la Clase IVb**. Las respuestas son puntuales, pero la clase superior se da abierta (como en la clase IIb)

Las respuestas de la Clase II significan de hecho, la pérdida de los intervalos extremos, el inferior y el superior (o uno solo de ellos, en los casos IIa y IIb), en los análisis en los que esté implicado el recorrido de la f.d.p. En el caso de las clases III y IV la pérdida de información es aún mayor, existiendo una fuerte incertidumbre entre las fronteras de los distintos intervalos, lo que puede invalidar la totalidad de la información obtenida de esa forma.

Se puede pensar que la pérdida de esa información es fácilmente recuperable, al menos en la Clase II, si en otra respuesta el informante ha establecido los límites de recorrido de la f.d.p. Pero, en general, no se admitirá esta hipótesis sin discutirla.

#4.2.1.1.- RECUPERACIÓN DE RECORRIDOS MEDIANTE LA RELACION ENTRE R_C Y R_L

Como consecuencia de la búsqueda de posibles métodos de recuperación de la información perdida en las contestaciones de la Clase II, una de las aproximaciones consiste en suponer que los sujetos mantienen una cierta relación entre la amplitud del conjunto de las tres clases centrales y la amplitud total o recorrido que asignan a la f.d.p. Para comprobar esa hipótesis la metodología consiste en comparar la amplitud de los dos intervalos en aquellos casos en que ambos son conocidos con seguridad (en la Clase I de respuestas), mediante la introducción de las:

Definiciones:

Recorrido. El recorrido definido por el entrevistado en las respuestas de la clase I se calcula como la diferencia entre el extremo superior de la clase rendimientos *"muy buenos"* y el extremo inferior de la clase rendimientos *"muy malos"*. Esta medida, como es natural, no se puede aplicar al resto de las clases de respuestas. En las respuestas de la clase II, sin embargo, se puede determinar el recorrido entre el extremo inferior de la clase de rendimientos *"muy buenos"* y la frontera superior de la clase de rendimientos *"muy malos"*. Para distinguir entre ambas medidas, se ha introducido la notación *"recorrido largo"* (que se corresponde con el concepto estadístico de recorrido) y *"recorrido corto"*.

Recorrido largo (R_L). La diferencia entre el valor superior de la clase rendimientos *"muy buenos"* y el valor inferior de la clase rendimientos *"muy malos"*.

Recorrido corto (R_C). La diferencia entre el valor inferior de la clase rendimientos *"muy buenos"* (o el superior de la clase rendimientos *"buenos"*) y el valor superior de la clase rendimientos *"muy malos"* (o el inferior de la clase rendimientos *"malos"*).

Δ

En las respuestas de la Clase II, si se pudiera mantener la hipótesis de que el recorrido corto R_C está estrechamente relacionado con el recorrido total o recorrido largo R_L, en el sentido de que el recorrido de las tres clases centrales determina de forma suficientemente precisa el recorrido total declarado por los sujetos, entonces se podría aprovechar ese conocimiento para estimar el recorrido (largo) en esas encuestas. De esta forma se dispondría de información adicional para la discusión de cuestiones en las que es importante conocer o tener estimado el recorrido (por ejemplo, para estudiar la forma de la función de densidad de probabilidades, la posición de la moda, de la mediana, etc.)

La estimación de R_L a partir de R_C se hace en las encuestas de las que se dispone de toda la información (encuestas con respuestas de la Clase I). Si los resultados son positivos, entonces a partir de los valores R_C de las encuestas con respuestas de la Clase II se obtienen sus correspondientes valores R_L estimados.

En las clases IIa y IIb, la diferencia $R_L – R_C$ indica la amplitud estimada del único intervalo no definido con precisión por el sujeto. En el caso de las respuestas de la Clase II, sin embargo, la diferencia indica la suma de la amplitud de los dos intervalos extremos no definidos. Se ha adoptado la convención de atribuir la mitad de la diferencia $R_L – R_C$ a cada uno de esos dos intervalos extremos.

#4.2.1.2.- RECUPERACIÓN MEDIANTE INFORMACIÓN DEL RECORRIDO

En los cuestionarios normalmente se incluye una pregunta para que el sujeto estime el valor mínimo de los rendimientos, el más frecuente y el máximo (R_{min}, R_{mf}, R_{max}). Se dispone, por tanto, de una estimación del recorrido del cultivo, aunque se trata de una estimación realizada en otro contexto al de la traducción de intervalos ordinales en numéricos.

En teoría, ambas estimaciones deberían coincidir o ser muy próximas, pero existen razones que pueden provocar que esto no sea así. Es por ello que se ha investigado:

(1) la correspondencia entre el recorrido $R_{max} - R_{min}$ con el recorrido largo R_L, a fin de estimar el recorrido largo mediante esa relación;

(2) la correspondencia entre R_{max} y el extremo superior de la clase "muy buenos" y R_{min} con el extremo inferior de la clase "muy malos", por el propio interés para explorar la coherencia de las estimaciones de rendimientos, pero también para poder aprovechar esa información en la recuperación de la información de las clases de respuestas II, III y IV.

En el caso de la Clase II es inmediato utilizar los valores de los extremos, a partir de su estimación para recomponer el recorrido (largo). La operación no es tan inmediata con las clases de respuestas III y IV.

#4.2.1.3.- HIPÓTESIS DE DISTANCIA MEDIA EN LAS FRONTERAS DE LOS INTERVALOS

En las respuestas de la Clase III, el sujeto encuestado da una estimación puntual para cada uno de los cinco intervalos. Si se divide el recorrido en cinco intervalos, en lugar de dar sus correspondientes fronteras, responde con cinco valores v_i, uno para cada intervalo i (i = 1, 2, 3, 4, 5). Para estimar una frontera entre los distintos intervalos, supondremos que la frontera entre v_i y v_{i+1} se encuentra en el punto medio de esos puntos.

Esta convención divide el recorrido en intervalos que pueden ser desiguales (lo que es normal en las contestaciones de los agricultores y de algunos expertos), y tiene el inconveniente de que el valor representativo del intervalo, dado por el sujeto encuestado, v_i, no será en general el punto medio de su intervalo.

La ventaja de adoptar esta convención es su sencillez. La adopción de un supuesto de intervalos simétricos en torno a los correspondientes v_i provoca problemas de estimación que puede que no justifiquen su utilización.

#4.2.1.4.- RESUMEN DE CORRECCIONES DE LOS TIPOS DE INTERVALO

Resumiendo la discusión anterior, cuando los resultados de las respuestas sean un conjunto de clases como las descritas en ζ 4.2.1, las operaciones que se proponen para estimar los recorridos y la amplitud de los diferentes intervalos son las que se describen a continuación.

En respuestas de la Clase I tanto el recorrido como las amplitudes de los intervalos han sido facilitados por el encuestado, y por tanto no son datos estimados posteriormente por las operaciones que se describen.

En las respuestas de la Clase II, el sujeto encuestado no ha definido con precisión la frontera inferior (tanto del intervalo extremo inferior como del recorrido), o no ha definido con precisión la frontera superior (del intervalo extremo superior como del recorrido) o no ha definido ninguna de las dos fronteras.

La corrección en este caso se basará:

(1) En los resultados de las regresiones R_L f (R_C) con la información de las respuestas de la Clase I, de tal forma que a partir del valor R_C calculado con los valores de la Clase II se puede estimar su correspondiente valor R_L.

(2) A partir de la información dada por los encuestados acerca de estimaciones directas del recorrido, mediante (R_{min}, R_{mf}, R_{max}). Se estudia la correlación de R_{min} y R_{max} con los extremos inferiores y superiores de las respuestas de la Clase I. Con los resultados se estiman los extremos de las respuestas de la Clase II, a partir de sus correspondientes R_{min} y R_{max}.

(3) Se correlaciona el recorrido $R_{max} – R_{min}$ con R_L.

En el caso de los cuestionarios con respuestas de la Clase IIa, solamente es necesario estimar el extremo inferior. En el caso de respuestas de la Clase IIb sólo es necesario estimar la frontera superior. En la Clase II, sin embargo, existe un problema de asignación de la cantidad estimada $R_L – R_C$ a los dos intervalos extremos, por lo que por convención se realizará asignando a cada uno de ellos una amplitud (R_L-R_C)/2. Esta convención se justifica porque se desea utilizar la información en la estimación de medidas que en general estarán sobre intervalos más interiores, y aunque la simplificación puede afectar a esas medidas, no se espera que el impacto sea muy importante (no se discuten en esta Tesis problemas relacionados con las colas de la distribución).

En los cuestionarios con respuestas de la Clase III se utilizará la hipótesis de distancias medias en las fronteras. Para las fronteras inferior y superior del intervalo, se supondrá (1) que están estimadas por los valores R_{min} y R_{max}, cuando sea posible; y (2) que son simétricas, en otros casos.

#4.2.2.- DISCONTINUIDAD DE INTERVALOS

Normalmente se espera que los intervalos recubran totalmente y de forma continua la amplitud del recorrido, es decir, que entre dos intervalos contiguos no existan “solapamientos” ni “huecos”. Esta propiedad la hemos denominado *continuidad de los intervalos*.

Cuando los sujetos establecen intervalos, con frecuencia definen sus fronteras de tal forma que dos intervalos contiguos no tienen fronteras comunes. Así, por ejemplo, puede que para el conjunto de los rendimientos *“normales”* un experto dé el intervalo (1500, 2000) y para el intervalo contiguo de rendimientos *“buenos”* dé el intervalo (3000, 4000), dejando un “hueco” entre ambas clases contiguas dado por (2000, 3000).

Este comportamiento no mantiene la propiedad de continuidad de los intervalos, y por tanto define intervalos discontinuos, cuyos límites precisos son difíciles de establecer, circunstancia que puede impedir la utilización de la información definida de esta forma, por ejemplo, en los análisis relacionados con los recorridos de las f.d.p.

Igualmente, los sujetos frecuentemente definen sus fronteras de tal forma que se superponen o solapan puntos de dos intervalos contiguos (no existe un único punto frontera entre los dos intervalos). De este modo, puede que para el conjunto de los rendimientos *“normales”* un experto dé el intervalo (1500, 2500) y para el intervalo contiguo de rendimientos *“buenos”* dé el intervalo (2000, 3000), introduciendo un “solapamiento” de 500 kg/ha entre ambas clases.

Posiblemente este comportamiento esté asociado a que el sujeto busca mentalmente un anclaje para la marca de clase del intervalo, y la describe mediante una “bola” centrada en torno al anclaje, o como un subintervalo significativo del intervalo total, pero sin prestar atención a la definición de las fronteras de cada intervalo.

Alternativamente se puede conjeturar que presta atención a un conjunto de valores representativos del intervalo, sin prestar atención o siendo incapaz de precisar las fronteras del intervalo completo.

#4.2.2.1.- ÍNDICE DE CONTINUIDAD DE INTERVALOS

Para identificar en qué cuestionarios se producen los fenómenos de discontinuidad de intervalos, se ha elaborado el Índice de continuidad de los intervalos (ICI).

Si en las respuestas se definen intervalos, es deseable que éstos sean continuos, es decir que no existan "huecos" entre intervalos, ni tampoco "solapamientos". Una clase y su inmediata inferior (superior) deben tener en común un punto de frontera o dos puntos contiguos. El índice, por tanto, se calcula midiendo la diferencia entre los extremos superior e inferior de clases contiguas. Cuando el índice toma el valor 0 quiere decir que existe continuidad entre las diferentes clases definidas. En otro caso, indica la existencia de discontinuidades.

Obviamente, el índice se puede aplicar solamente a las respuestas de las clases I y II, que son aquellas en las que se han definido intervalos indicando sus fronteras.

#4.2.2.2.- INTERPRETACIÓN Y CORRECCIÓN DE DISCONTINUIDADES

Provisionalmente, se ha interpretado la definición de intervalos no continuos como una insuficiencia de clases y los solapamientos como un exceso de clases. La idea, en el caso de discontinuidad, es que al forzar al sujeto a recubrir el recorrido de los rendimientos con cinco clases (*rendimientos muy malos, malos, normales, buenos y muy buenos*), implícitamente se le obliga a aceptar un recorrido relativamente amplio para cada clase, que no coincide con la amplitud que él juzga intuitivamente como definitoria de la clase. Lo contrario ocurre en los solapamientos.

En el caso de discontinuidad (por **"huecos"**), si un sujeto dice que los rendimientos *"normales"* abarcan el intervalo (1500, 2000) y para el intervalo contiguo de rendimientos *"buenos"* da el intervalo (3000, 4000), dejando un "hueco" entre ambas clases, ese "hueco" se interpreta como una clase que no se ha definido previamente. Es decir, el sujeto no juzga con las cinco clases definidas en la pregunta, sino que aumenta el número de clases en sus respuestas.

Para corregir, interpretaremos que las fronteras inferiores de los intervalos son seguras. Así, por ejemplo, si un resultado *"normal"* se considera como (1500, 2000) y uno *"bueno"* como (3000, 4000), si no se enlaza la frontera inferior del bueno con la superior del normal, es porque no se desea extender el resultado bueno más allá del valor 3000. A partir de ese valor, se introduce la clase (2000, 3000) "no definida" en la pregunta.

Consecuentemente, para enlazar con la situación definida por la pregunta, se adoptará la siguiente convención: se respeta el límite inferior de cada clase y se extiende el superior hasta enlazar con la clase contigua, corrigiendo de esta forma las discontinuidades observadas. Así, por ejemplo, en el caso que estamos discutiendo, se consideraría el rendimiento *"normal"* como (1500, 3000), en lugar de (1500, 2000), y de esta forma se enlazaría con la clase de rendimientos *"buenos"* (3000, 4000).

Cuando un sujeto introduce "**solapamientos"** entre rendimientos, planteando, por ejemplo, que los rendimientos normales abarcan el intervalo (1500, 2500) y los buenos abarcan el intervalo (2000, 3000), dejando una parte común, dicho "solapamiento" puede interpretarse como un exceso de clases, es decir, que el sujeto no juzga con las cinco clases definidas en la pregunta, sino que reduce el número de clases en sus respuestas. Consecuentemente, entre las dos clases en que se produce el "solapamiento" se ha respetado el límite superior de la clase *j* y se ha sustituido este valor en el límite inferior de la clase *j+1*. Así, por ejemplo, en el caso que estamos discutiendo como ejemplo, se consideraría el rendimiento *"normal"* como (1500, 2500), tal como estaba, y el *"bueno"* como (2500, 3000), en lugar de (2000, 3000).

#4.2.3.- REDUCCIÓN DE INTERVALOS

Otro comportamiento detectado en las distintas encuestas, tanto entre los expertos como entre los agricultores, es el denominado "reducción de intervalos", fenómeno para el que se da la siguiente

Definición:

Reducción del intervalo. Se dice que en una respuesta se reducen los intervalos dados por la pregunta, si a alguno de los intervalos, particularmente a los extremos, se le atribuye una frecuencia nula.

Así, por ejemplo, supongamos que se pide a un sujeto que asigne una frecuencia a cada uno de los intervalos de rendimientos definidos en lenguaje natural como rendimientos "muy malos", "malos", "normales", "buenos" y "muy buenos", si como respuesta asigna una frecuencia nula a una o varias de esas clases o intervalos, diremos que ha realizado una reducción de intervalo.

La reducción de intervalos posiblemente indique varias circunstancias distintas.

Una primera interpretación puede basarse en la asociación establecida entre el calificativo de cada año y la influencia de factores que determinen el resultado. Por ejemplo, puede que un sujeto crea que las condiciones climáticas son tan estables que no se puede hablar de años "muy malos", y por tanto expresará esa idea asignando un valor nulo a la frecuencia del intervalo.

La asociación del calificativo "muy malo", "malo", etc. del año puede ser incompatible con la visión de un determinado cultivo muy ligado a ciertas condiciones (meteorológicas, tecnológicas, etc.) para las que no se conciben esas calificaciones.

Por tanto, la analogía que se desea establecer entre los rendimientos, cuya variabilidad se cubre con cinco intervalos, y unos calificativos en lenguaje natural, que intentan etiquetar de forma relativa a cada uno de los intervalos, no es aceptada por el sujeto, para quien los calificativos tienen significados que transcienden y predominan sobre su utilidad para indicar diferencias relativas de rendimientos.

En ocasiones, puede que el intervalo de variación de los rendimientos sea tan reducido, que no se conciba una división en cinco intervalos, ya que el umbral cuantitativo que hace que un tipo de rendimiento se distinga de otro sea relativamente tan amplio que no se pueda hacer una división significativa en cinco intervalos. Así, por ejemplo, supongamos que los rendimientos de un cultivo varían entre 6000 y 7000 kg / ha, y que una variación significativa de rendimientos tiene que ser superior a 400 kg / ha (este es el umbral para que el sujeto distinga entre clases de rendimientos), las clases (6000, 6200), (6200, 6400), (6400, 6600), (6600, 6800), (6800, 7000), cuyas amplitudes son inferiores al umbral sicológico comentado, puede que sean percibidas como excesivas. Esta representación mental de la aplicación de la escala ordinal a la cardinal no será cómoda, y puede que el sujeto opte por definir solamente tres clases (6000, 6300), (6300, 6600), (6600, 7000), identificándolas con las tres clases ordinales centrales ("malo", "normal", "bueno"), y por tanto amputando los dos intervalos extremos ("muy malos" y "muy buenos").

Posiblemente haya implícitas otras razones e interpretaciones para truncar el número de intervalos ordinales a la hora de asignar frecuencias, que habrá que establecer en investigaciones futuras.

Relacionado con este fenómeno se tiene la siguiente:

Conjetura: Las definiciones de los intervalos en un lenguaje natural pueden ser problemáticas si los sujetos disponen de referentes distintos para los mismos términos, y pueden interpretar de forma distinta cada significado.
Δ

#4.2.4.- HIPÓTESIS SOBRE AMPLITUD IGUAL DE LOS INTERVALOS

Hipótesis sobre amplitud igual de los intervalos. Dado que la responsabilidad de definir los intervalos es principalmente de los encuestadores, y que éstos eran estudiantes universitarios, admitiremos la hipótesis de que éstos tienden a definir amplitudes iguales para todos los intervalos.

En general, en la literatura predominan los histogramas de frecuencias en los que las amplitudes de los distintos intervalos es la misma. Casi todos los problemas académicos que implican la división de un recorrido en intervalos se suelen desarrollar dando igual amplitud a todos los intervalos.

Solamente en problemas relacionados con el mejor ajuste de un histograma a una f.d.p. continua, en la que cada frecuencia está determinada por la amplitud de su base (d_i) y por su altura (h_i), se suele discutir si el mejor ajuste se podría obtener con valores desiguales de las bases d_i o amplitudes de los intervalos. Pero finalmente muchos de estos problemas se acaban resolviendo mediante histogramas de frecuencia con intervalos de igual amplitud.

No es exagerado decir que el mundo académico está poblado exclusivamente por histogramas construidos sobre intervalos de igual amplitud.

Cuando se realizaron las encuestas y se pidió una traducción de las escalas ordinales del tipo rendimientos "muy malos", "malos", "normales", "buenos" y "muy buenos" a intervalos numéricos, la mayor parte de los agricultores, y algunos expertos, utilizaron intervalos de amplitud distinta para cada clase o intervalo ordinal. Las respuestas sugerían que las personas externas al mundo académico tienden a emplear clases de longitud distinta, que se ajusten a su percepción, antes que ajustar su percepción a las exigencias de los intervalos de igual amplitud.

En el problema que discutimos, la hipótesis de igual amplitud simplifica bastante las potenciales interpretaciones de los datos ofrecidos por los estudiantes – encuestadores. Es por ello que se afirma de forma general la validez de esa conjetura (que los estudiantes tienden a emplear intervalos iguales). Esta tendencia, en algunas encuestas, se tiende a potenciar mediante el sistema de instrucciones.

#4.2.5.- DIFERENCIAS ENTRE MEDIAS Y VARIANZAS DE LAS APROXIMACIONES FUNCIONALES

Los valores medios de las funciones triangular y beta – PERT vienen dados, respectivamente, por las expresiones [14] y [15], mientras que las varianzas son [16] y [17].

[14] $RMT = (1/3)\,[R_{min} + R_{mf} + R_{max}]$

[15] $RMB = (1/6)\ [R_{min} + 4\ R_{mf} + R_{max}]$

[16] $VT = (1/18)\ [(R_{max} – R_{min})^2 + (R_{mf} – R_{min})\ (R_{mf} – R_{max})]$

[17] $VB = (1/36)\ (R_{max} – R_{min})^2$

Para las diferentes estimaciones (R_{min}, R_{mf}, R_{max}) se han calculado los valores RMT y RMB, y en los casos que ha sido posible se han comparado con estimaciones directas de los rendimientos medios R_m (puntuales) realizadas por los sujetos encuestados.

La comparación entre RMT y RMB no tiene un interés especial, ya que dados los valores (R_{min}, R_{mf}, R_{max}) la diferencia entre ambas estimaciones queda analíticamente determinada. Así, si d_x es la diferencia entre RMT y RMB se tiene que:

[19] $d_x = RMT – RMB = (1/3)\ [R_{min} + R_{mf} + R_{max}] - (1/6)\ [R_{min} + 4\ R_{mf} + R_{max}] = (1/6)\ (R_{max} + R_{min} - 2\ R_{mf})$

que alcanza su valor mínimo para $R_{mf} = (1/2)\ (R_{min} + R_{max})$

El anterior resultado indica que cuando el valor más frecuente R_{mf} coincide con el punto medio del recorrido (R_{max}, R_{min}), los valores RMT y RMB coinciden. La diferencia será tanto mayor cuanto mayor sea la distancia de R_{mf} al punto medio del intervalo.

La cuestión de qué aproximación funcional, triangular o beta – PERT, ofrece la mejor estimación de un valor R_m a partir de la estimación de tres valores puntuales (R_{min}, R_{mf}, R_{max}) se resuelve finalmente en la respuesta: “depende de la posición del valor más frecuente R_{mf} con respecto a la media del intervalo (R_{max}, R_{min})”, ya que $(RMT – R_m) – (RMB – R_m) = RMT – RMB$.

Por tanto, en las comparaciones entre RMT y RMB en realidad lo que se está midiendo es la posición de R_{mf} en relación al recorrido de la función.

Para el caso de la varianza ocurre algo semejante. Así, si la diferencia entre las varianzas es d_s:

[20] $d_s = VT – VB = (1/18)\ [(R_{max} – R_{min})^2 + (R_{mf} – R_{min})\ (R_{mf} – R_{max})] - (1/36)\ (R_{max} – R_{min})^2 =$

$(1/36)\ [(R_{max} – R_{min})^2 - 2\ (R_{mf} – R_{min})\ (R_{max} - R_{mf})] =$

$(1/36)\ [(R_{max} – R_{mf} + R_{mf} – R_{min})^2 – 2\ (R_{mf} – R_{min})\ (R_{max} - R_{mf})] =$

$(1/36)\ [(R_{max} – R_{mf})^2 + (R_{mf} – R_{min})^2\ ]$

expresión que es siempre positiva (y por tanto VT > VB) y que alcanza un valor mínimo para $R_{mf} = (1/2)\ (R_{max} + R_{min})$. La diferencia de las varianzas será tanto mayor cuanto más se aleje la moda o valor más frecuente del punto medio del intervalo.

Como consecuencia, las diferencias entre las estimaciones de las medias y de las varianzas de la estimación triangular o la beta - PERT constituyen en realidad un test sobre la situación del valor más frecuente en el recorrido de la f.d.p.

#4.3.- CONVENCIONES COMUNES

En el análisis de los diferentes cuestionarios se han empleado sistemática y repetidamente técnicas de clasificación que resultarán tediosas de leer si se explican en los correspondientes apartados. Es por eso que se considera conveniente establecer en este apartado las claves interpretativas, a fin de ahorrar su exposición futura.

#4.3.1.- FORMA DE LA FUNCIÓN DE DENSIDAD CON INTERVALOS ORDINALES

A fin de dar una imagen gráfica de la forma de la función de densidad de probabilidad se utilizan en las descripciones unas tablas, cuyos encabezamientos son un conjunto de convenciones que se describen en lo que sigue.

En general se ha trabajado con recorridos divididos en 5 intervalos. Por convención, los intervalos propuestos se numeran comenzando por el intervalo inferior (número 1) y finalizando por el intervalo superior (número 5), con lo que el intervalo central tiene asignado el número 3.

En caso de que el recorrido sea dividido en un número mayor (menor) de intervalos, la numeración seguirá la misma convención (número 1 el intervalo inferior, y número n el intervalo superior, en caso de división en n intervalos).

Reducción del intervalo. Se dice que en una respuesta se reducen los intervalos dados por la pregunta, si a alguno de los extremos se les atribuye una frecuencia nula (tal como se ha discutido en ζ 4.2.4). La leyenda *"A"* indica los casos en que se asignan frecuencias nulas a los dos intervalos extremos (inferior y superior) simultáneamente; bajo la etiqueta *"I"* se indican los casos en los que se asigna una frecuencia nula solamente el extremo inferior o izquierdo; y bajo la etiqueta *"D"* se denotan los casos en los que se asigna una frecuencia nula solamente al extremo superior o derecho.

Bajo la etiqueta de *"uniforme"*, en la misma tabla, se indica cuántas contestaciones dan el mismo porcentaje a los intervalos considerados. Se ha establecido la convención de que cuando se presenta reducción de intervalo y distribución uniforme de los intervalos con frecuencias positivas, entonces el caso se considera de distribución uniforme.

Bajo la leyenda *"extremos < 15 %"* se indica en cuántas respuestas la frecuencia asignada a los dos intervalos extremos, inferior y superior, no llega a sumar el 15% (y por tanto los n-2 intervalos centrales tienen asignada una frecuencia superior al 85%).

Con el título *"simetría"* se indican cuántas respuestas han asignado frecuencias simétricas en torno al intervalo central. Es decir, por ejemplo, se ha asignado la misma probabilidad a las dos clases extremas, de rendimientos *"muy malos"* y *"muy buenos"*; y también se ha asignado la misma frecuencia a los intervalos de rendimientos *"malos"* y *"buenos"*.

Bajo el título *"inter 1 = inter 5"* se indica el número de casos en los que la frecuencia asignada al intervalo extremo inferior es igual a la frecuencia atribuida al intervalo extremo superior. Por tanto, a ambos intervalos extremos se les ha asignado la misma frecuencia.

Con el título *"inter 2 = inter 4"* se indica el número de casos en los que la frecuencia asignada al intervalo rendimientos *"malos"* coincide con la frecuencia atribuida al intervalo rendimientos *"buenos"*.

En general, con "inter p = inter q" se indica el número de casos en los que la frecuencia asignada al intervalo p coincide con la frecuencia atribuida al intervalo q.

#4.3.2.- MEDIANAS EN FUNCIONES DE DENSIDAD CON INTERVALOS ORDINALES

Definición:

Una mediana referida a medidas ordinales de intervalos se define como el punto en el que se tiene la mitad de las frecuencias acumuladas a su izquierda. Este punto se expresa por un entero que indica la posición relativa de la mediana con respecto a los intervalos y una parte fraccionaria que indica un tanto por uno de la amplitud del intervalo. Así, en caso de cinco intervalos numerados de 1 a 5, una mediana de valor 2,5 indica que se encuentra en la mitad del intervalo 3, es decir, en el centro de la distribución de 5 intervalos, dejando a su izquierda dos intervalos y medio. Una mediana de valor 2 indicaría que se encuentra en la frontera derecha del segundo intervalo. Una mediana de valor 3,3 indicaría que se encuentra en el primer tercio del cuarto intervalo (dejando a su izquierda 3 intervalos y extendiéndose hasta 0,3 veces la amplitud del cuarto intervalo).

#4.3.3.- NÚMERO Y POSICIÓN DE LAS MODAS EN INTERVALOS NUMÉRICOS

Al analizar la forma de la f.d.p., una vez que se refieren las frecuencias a intervalos numéricos, se ha seguido una metodología de análisis común.

Se han construido tablas en las que se ha indicado en cuántas ocasiones la moda se encuentra a la izquierda del punto medio del recorrido (leyenda *I*), centrada o coincidiendo con el punto medio del recorrido (leyenda *C*), o a su derecha (leyenda *D*). También se reflejan los casos en los que la distribución es uniforme (leyenda *Un*), los casos en que se presentan dos modas no contiguas (leyenda *Bi*) y el total de casos examinados.

La **posición relativa de la moda** se ha indicado mediante un índice (IM) construido de la siguiente manera: Si el sujeto ha definido la moda en el intervalo i, de un conjunto de n intervalos, se ha calculado la amplitud de los intervalos que quedan a la izquierda del mismo y se le ha restado la amplitud de los intervalos que quedan a la derecha. Esta diferencia se ha dividido por la amplitud del intervalo para el que se ha definido la moda, para normalizar.

Así, el recorrido con referencia a ese intervalo i de amplitud A(i) viene dado por R = A(1, i -1) + A(i) + A(i+1, n), donde A(i, j) indica la suma de las amplitudes de los intervalos contiguos i, i+1,..., j, con $j \geq i$. El índice por tanto vendrá dado por: IM = [A(1, i-1) – A(i+1, j)] : A(i).

Por tanto, un valor p del índice significa: p A(i) = A(1, i-1) – A(i+1, j).

De esta forma un valor negativo indica que la moda se encuentra a la izquierda del punto medio del recorrido (el “brazo derecho” es más grande que el “brazo izquierdo”), un valor nulo indica una moda centrada y un valor positivo indica una moda situada a la derecha del punto medio del recorrido.

Un valor negativo –1 indica que la moda se encuentra a la izquierda del punto medio del recorrido, y a una distancia una vez la amplitud del intervalo para el cual se ha señalado que se encuentra la moda.

#4.4.- PRINCIPALES RESULTADOS DE LA INVESTIGACIÓN EXPERIMENTAL

Se discuten en este apartado los principales resultados obtenidos en el trabajo experimental, al objeto de facilitar una visión de conjunto del mismo, que permita un mejor seguimiento de la discusión más detallada que se realiza en cada capítulo para cada encuesta. Se prefiere una discusión en un apartado que no sea estrictamente de conclusiones, debido a que los resultados de nuestra investigación son más conjeturas que hipótesis, dada su naturaleza de exploración.

Como se ha discutido anteriormente, el planteamiento del trabajo experimental analizado en esta investigación ha sido el siguiente:

(1.1) Determinación por el sujeto de estimaciones puntuales de los rendimientos de algunos cultivos, concretamente los rendimientos mínimos, más frecuentes y máximos, (R_{min}, R_{mf}, R_{max}), para realizar estimaciones de los valores medios y de las varianzas a partir de las aproximaciones funcionales triangular y beta – PERT (de esta forma denominaremos a la aproximación a una función beta con las hipótesis normalmente admitidas en el PERT).

(1.2) En la mayoría de los casos, estimación puntual directa de los rendimientos medios por parte del sujeto, a fin de contrastar las estimaciones (1.1) con esta estimación directa.

(1.3) No se ha solicitado de los sujetos una estimación de la varianza, por ser un valor del que existe suficiente evidencia empírica de que no está sicológicamente disponible.

(1.4) Sin embargo, cuando se han obtenido distribuciones de frecuencias en intervalos numéricos, se han comparado las varianzas calculadas a partir de las distribuciones y de las aproximaciones funcionales triangular y beta – PERT.

El siguiente experimento consistía en:

(2.1) Determinación de la frecuencia (en tanto por ciento) de cada uno de los cinco intervalos ordinales expresados en lenguaje natural como rendimientos “muy malos”, “malos”, “normales”, “buenos” y “muy buenos”. A partir de estas estimaciones se han discutido las formas funcionales (en intervalos ordinales) resultantes, especialmente su caracterización mediante la situación de la moda y de la mediana (en este tipo de intervalos). Se conjetura que esta descripción posiblemente esté asociada a la visión de los rendimientos del cultivo en la memoria del decisor, y que eventualmente podría ser la base de algún tipo de cálculo o comparación mentales a la hora de realizar sus decisiones de producción. (Las importantes reducciones de intervalos que han realizado algunos sujetos sugieren la plausibilidad de esa conjetura. Sin embargo, en esta investigación no se ha explorado específicamente).

(2.2) Traducción de los intervalos en escalas ordinales a intervalos numéricos. Mediante esta traducción, el sujeto debe indicar cuál es el recorrido que considera para cada clase definida ordinalmente (p.e. los rendimientos “normales” del trigo en secano se corresponden con el intervalo (1000, 1300 kg/ha)).

(2.3) Teóricamente se podría haber solicitado que los sujetos señalaran las frecuencias que estimaban para cada intervalo numéricamente definido, al objeto de comparar los resultados de las frecuencias sobre intervalos ordinales + traducción de los intervalos ordinales en numéricos con las correspondientes frecuencias sobre intervalos numéricos. Esta comparación triangular, sin embargo, no se ha planteado porque en esta primera aproximación se temía que el sujeto identificara que se le estaba solicitando la misma información de diferentes formas, y ello le motivara a extremar su coherencia, y como consecuencia decidiera anclarse en algunos valores. Como consecuencia altos niveles de coherencia podrían haber estado asociados a fuertes anclajes, y habría dificultado la obtención de conclusiones válidas o, lo que sería peor, la identificación de problemas en las estimaciones.

(2.4) A partir de las frecuencias (2.1) trasladadas a los intervalos numéricos obtenidos mediante la traducción (2.2), se ha examinado la forma de la función de densidad, principalmente mediante la determinación de la posición de las modas y de las medianas, en intervalos normalizados [0,1].

Una primera medida de coherencia se ha derivado de la comparación entre los valores de los rendimientos medios obtenidos de las aproximaciones funcionales discutidas en (1.x) y de las estimaciones de frecuencias sobre intervalos numéricos explicadas en (2.x).

La cuestión de si los agricultores emplean una “receta productiva” fija o si la varían en función de las circunstancias del año no forma parte del análisis principal que se ha realizado en esta investigación, pero se ha incluido su discusión en el capítulo 5, dedicado al análisis de la encuesta a los expertos, porque pone claramente de manifiesto que las preguntas son interpretadas en función de contextos, y subraya la importancia de conocer más sobre el particular en los ámbitos de nuestro interés, para el desarrollo de sistemas eficaces de comunicación. Esta conclusión no se podía analizar de forma tan gráfica en el conjunto de cuestiones discutidas en torno a la definición de las funciones de densidad, que es el principal eje de la presente investigación, aunque existe abundante evidencia en esa dirección.

Igualmente, en el capítulo 5 se ha realizado un esfuerzo para explicitar las conjeturas y conclusiones que sugieren los análisis, a fin de llamar la atención de los lectores sobre los principales puntos. Dado que las principales líneas señaladas en ese capítulo se dibujan claramente en el resto, se ha evitado la repetición.

#4.4.1.- RESULTADOS DE LAS APROXIMACIONES FUNCIONALES

En la Tabla Resumen 4 se sintetizan los principales resultados obtenidos de las estimaciones puntuales, (R_{min}, R_{mf}, R_{max}) y rendimiento medio R_m realizadas por los sujetos encuestados.

La primera cuestión de interés tiene que ver con la situación de la moda (el valor más frecuente R_{mf}) en los intervalos normalizados [0,1]. Los expertos la sitúan mayoritariamente en la zona centro – izquierda, tanto en secano como en regadío. Solamente en un 20% de los casos de secano y un 27% de los casos de regadío los expertos dan modas centradas.

Las diferencias entre los valores medios calculados mediante la aproximación triangular (RMT) y la beta – PERT (RMB) con el rendimiento medio estimado puntualmente, R_m, son relativamente reducidas: en el 75% de los casos de secano y el 80% de los casos de regadío, la aproximación triangular ofrece diferencias entre ambos valores inferiores al 15%. Con la aproximación beta – PERT los resultados son ligeramente mejores.

Con la aproximación triangular el 50% de las estimaciones de los expertos presentan diferencias inferiores al 7% en secano y al 6% en regadío, mientras que con la aproximación beta esa diferencia relativa se reduce al 6% y 5%.

La regresión lineal de R_m con las diferentes estimaciones funcionales (forzando al término b, término independiente de la recta, a ser nulo) ofrece valores del coeficiente a (pendiente de la recta) del orden o superiores a 0,94.

Los mismos resultados para los agricultores son igualmente favorables para las diferencias entre el valor puntual estimado de los rendimientos medios y las estimaciones funcionales (excepto para algunos casos de alfalfa, de la que se disponía de pocas observaciones y que es un cultivo plurianual). En general, la aproximación beta – PERT está más próxima a R_m que la triangular. La regresión de las diferentes estimaciones ofrece en la inmensa mayoría de los casos resultados mejores que los discutidos para los expertos.

Esta buena aproximación se mantiene en el tiempo, tanto para los mismos agricultores entrevistados dos veces, por diferentes encuestadores, en un plazo del orden o superior a una semana, como en el conjunto de las poblaciones encuestadas en dos años sucesivos.

Se puede concluir, por tanto, que las estimaciones puntuales ofrecidas por los expertos y los agricultores son bastante coherentes en general, y sugieren que existe un conocimiento sustantivo que posiblemente se pueda obtener sin grandes desviaciones mediante la mejora de los sistemas de obtención de la información (de los cuestionarios, de la comprensión de las preguntas, de la metodología implícita, etc.). En efecto, posiblemente términos como “rendimiento más bajo posible” o “el mayor rendimiento observado” o “el rendimiento normal o más frecuente”, etc. sean demasiado ambiguos para personas no acostumbradas a realizar estimaciones con esta metodología. Parece existir evidencia sobre que algunas estimaciones de los rendimientos mínimos y máximos no son adecuadas, por no comprenderse con claridad el concepto implicado. Como es sabido, no se trata de describir como mínimo un potencial valor muy bajo de los rendimientos, que nunca ha sido observado ni probablemente se observará, simplemente porque el significado del término empleado no señale como manifiestamente errónea la estimación. Lo mismo ocurre con el término “rendimiento máximo”.

Tabla 4.

Resultados de las estimaciones puntuales, (R_{min}, R_{mf}, R_{max}) y rendimiento medio R_m realizadas por los sujetos encuestados.

			Posición de la Moda				**RMT-R_m**		**RMB-R_m**		**R_m = a RMX**	
		Total casos	**0.2-0.4**	**0.4-0.6**	**0.6-0.8**	**Posición**	**<15%**	**Mediana**	**<15%**	**Mediana**	**RMT a=**	**RMB a=**
Expertos												
	Secano	48	20	24:10	3	CI	36	0,071	41	0,062	0,97	0,94
	Regadío	44	17	23:12	3	CI	35	0,063	36	0,049	0,94	0,96
Agricultores general 1998-99 (segundo día y R_{m2})												
Cebada	Secano	45	8	27:7	10	C	34	0,071	38	0,048	0,99	0,99
Trigo	Secano	24	7	14:3	3	CI	20	0,05	21	0,035	0,98	0,99
Trigo	Regadío	13	1	9:3	3	CD	11	0,058	11	0,036	1,02	1,01
Maíz	Regadío	20	2	13:4	4	CD	16	0,038	18	0,033	0,99	0,99
Alfalfa	Regadío	12	3	5:2	4	C	12	0,094	12	0,063	1	1,01
Agricultores general 1998-99 (primer día y R_{m2})												
Cebada	Secano	46	7	24:8	14	CD	35	0,077	37	0,062	1	0,99
Trigo	Secano	27	9	9:3	7	C	19	0,071	20	0,046	0,99	0,99
Trigo	Regadío	13	2	9:3	2	C	12	0,046	12	0,036	1,01	1,01
Maíz	Regadío	20	1	14:2	4	CD	17	0,06	18	0,038	1,01	1,01
Alfalfa	Regadío	14		3:2	6	D	10	0,112	10	0,062	1,03	1,01
Agricultores general 1999-00 (primer día y R_{m1})												
Cebada	Secano	22	1	17:4	3	CD	19	0,046	18	0,039	0,99	0,99
Trigo	Secano	12	3	7:1	2	C	10	0,043	8	0,05	0,96	0,96
Trigo	Regadío	5	0	4:2	1	C	5	0,031	5	0,031	1	0,99
Maíz	Regadío	9	1	5:3	3	CD	9	0,05	10	0,036	0,98	0,98
Alfalfa	Regadío	7	0	5:3	2	CD	6	0,035	7	0,044	1,01	0,99
Agricultores general 1999-00 (segundo día y R_{m2})												
Cebada	Secano	22	6	12:2	4	CI	18	0,065	19	0,069	0,98	0,98
Trigo	Secano	12	3	6:1	3	C	12	0,033	12	0,033	1,02	1,01
Trigo	Regadío	5	0	3:0	2	CD	5	0,025	5	0,031	0,99	0,98
Maíz	Regadío	10	1	6:2	2	C	10	0,05	10	0,042	1	0,98
Alfalfa	Regadío	7	0	4:0	2	CD	7	0,058	7	0,058	1,06	1,07

(< 15% significa número de casos en los que la diferencia es igual o inferior al 15%; Mediana: mide la mediana de la distribución de diferencias; la lectura para la columna ‘posición de la moda’ correspondiente a los valores 0.4-0.6, es número de casos : número de casos con moda = 0.5).

La imagen de la función de densidad de los agricultores, marcada por la posición de la moda, sin embargo, parece diferente a la descrita por los expertos, ya que predominan los casos de modas en los intervalos centrales o centro – derecha (con 2 casos de 20 en el centro – izquierda para cultivos de secano).

Para subrayar la permanencia o estabilidad de las estimaciones realizadas por los agricultores entrevistados en dos días diferentes por encuestadores distintos, se han calculado las diferencias entre los rendimientos medios estimados mediante aproximaciones funcionales con los valores (R_{min}, R_{mf}, R_{max}) dados en cada ocasión (tabla resumen 5). Como puede observarse, todos los indicadores utilizados toman valores muy favorables a la conjetura de que las estimaciones de los valores de rendimientos mínimo, más frecuente y máximo, efectuadas en días distintos y en diferentes poblaciones de agricultores, son muy próximas cuando se comparan mediante el valor medio calculado con las aproximaciones funcionales triangular y beta - PERT. (Existe una ligera ventaja de la estimación beta sobre la triangular).

Tabla 5.

Resultados de los valores medios calculados a partir de las estimaciones puntuales de rendimientos (R_{min}, R_{mf}, R_{max}) realizadas por los agricultores en días diferentes.

			|RMT1 -RMT2|: RMT2		**|RMB1 -RMB2|: RMB2**		**RMT1= a RMT2**	**RMB1 = a RMB2**
		Total	**<15%**	**Mediana**	**<15%**	**Mediana**	**a=**	**a=**
AGRICULTORES GENERAL 1998-99								
Cebada	Secano	46	42	0,04	42	0,036	0,99	0,99
Trigo	Secano	24	21	0,038	21	0,038	0,98	0,99
Trigo	Regadío	13	11	0,036	11	0,036	1,01	1
Maíz	Regadío	20	18	0,038	18	0,036	0,98	0,98
Alfalfa	Regadío	14	10	0,044	9	0,039	0,99	0,99
AGRICULTORES GENERAL 1999-00								
Cebada	Secano	22	15	0,075	16	0,067	1	0,99
Trigo	Secano	12	12	0,027	12	0,03	0,99	0,98
Trigo	Regadío	5	5	0,031	5	0,031	0,97	0,96
Maíz	Regadío	10	8	0,05	9	0,083	0,96	0,98
Alfalfa	Regadío	7	6	0,058	6	0,044	1,03	1,05

(< 15% significa número de casos en los que la diferencia es igual o inferior al 15%; Mediana: mide la mediana de la distribución de diferencias).

Los resultados obtenidos al comparar las varianzas calculadas a partir de las estimaciones puntuales (R_{min}, R_{mf}, R_{max}) sugieren una mayor diferencia y variabilidad (tabla resumen 6). No obstante en 1998 – 99 los resultados de las regresiones de los valores obtenidos el primer día con los del segundo día sugieren poblaciones muy relacionadas. Se conjetura que la varianza es muy sensible a los valores extremos, que normalmente son los peor estimados.

Tabla 6.

Resultados de las varianzas calculados a partir de las estimaciones puntuales de rendimientos (R_{min}, R_{mf}, R_{max}) realizadas por los agricultores en días diferentes.

			|VT1 - VT2| : VT2		**|VB1 - VB2| : VB2**		**VT1=a VT2**	**VB1 = a VB2**
		Total	**<20%**	**Mediana**	**<20%**	**Mediana**	**a=**	**a=**
AGRICULTORES GENERAL 1998-99								
Cebada	Secano	46	23	0.2	22	0,22	0,97	0,98
Trigo	Secano	24	9	0.275	9	0,3	0,97	0,96
Trigo	Regadío	13	9	0.093	9	0,093	0,97	0,96
Maíz	Regadío	20	15	0.11	15	0,083	1	1
Alfalfa	Regadío	14	7	0.2	7	0,2	1	1
AGRICULTORES GENERAL 1999-00								
Cebada	Secano	22	13	0,267	14	0,25	0,69	0,68
Trigo	Secano	12	9	0,1	9	0,086	0,87	0,87
Trigo	Regadío	5	0	0,718	0	0,75	0,57	0,62
Maíz	Regadío	10	3	0,7	3	0,7	0,83	0,82
Alfalfa	Regadío	7	2	0,35	2	0,375	0,38	0,38

(< 15% significa número de casos en los que la diferencia es igual o inferior al 15%; Mediana: mide la mediana de la distribución de diferencias).

#4.4.2.- RESULTADOS EN FUNCIONES SOBRE INTERVALOS ORDINALES

Dados cinco intervalos ordinales definidos en lenguaje natural (rendimientos “muy malos”, “malos”, “normales”, “buenos” y muy buenos”) los resultados de la asignación de frecuencias a cada uno de ellos por los sujetos permiten dibujar una forma funcional referida a los mismos. Los resultados de las distintas estimaciones se resumen en la tabla 7.

Los expertos solamente en 1/3 de los casos definieron distribuciones de frecuencias simétricas respecto al intervalo ordinal central. Sin embargo en este caso (como se discute en el capítulo 5) se les solicitó una estimación directa de la forma de la función de densidad de probabilidades, y en esa estimación directa aumentaron el porcentaje de respuestas con distribuciones simétricas (desde 1/3 al 50 – 60 %).

En el mayor número de casos los expertos también definieron funciones unimodales (85% de los casos de secano y 95% de los de regadío), con la moda situada sobre el intervalo central (94 - 97% de los casos) de rendimientos “normales”.

Entre los agricultores, en la encuesta del primer año normalmente se definieron mayoritariamente distribuciones no simétricas respecto al intervalo central (normalmente en menos de 1/3 de los casos se definían como simétricas). Esta tendencia general se alteró el segundo año, cuando se incrementó de forma notable el número de casos que establecían distribuciones simétricas. En cualquier caso siguieron predominando las descripciones que definían frecuencias en los distintos intervalos que no eran simétricas respecto al intervalo central.

Las descripciones unimodales predominaron claramente entre los agricultores los dos años, tal como se ha descrito también para los expertos. En general, más del 80% de los casos presentan distribuciones unimodales.

Tabla 7.

Formas de las funciones de densidad de rendimientos definidas sobre intervalos ordinales.

Ordinal					**Modas**				**Medianas**				
		Resp.	**R.I.**	**Sim.**	**1 Moda**	**M**	**N**	**B**	**1.5-2**	**2-2.5**	**2.5-3**	**3-3.5**	**Sitúa**
Expertos (estimación de frecuencias)													
	Secano	33	7	11	28	2	31	3	2	21	10		C
	Regadío	34	23	11	32	0	33	3		13	20	1	CD
Expertos (estimación directa)													
	Secano	27		13	21								
	Regadío	28		16	26								
Agricultores general, primer día (1998-99)													
Cebada	Secano	46	3	12	43	7	33	8	3	24	13	3	CD
Trigo	Secano	25	2	4	22	2	16	7		8	12	3	CD
Trigo	Regadío	13	3	3	12	2	8	4	1	4	6	2	CD
Maíz	Regadío	20	10	7	16	2	17	5	1	8	7	3	CD
Alfalfa	Regadío	17	6	3	12	0	13	7		3	10	3	CD
Agricultores general, segundo día (1998-99)													
Cebada	Secano	46	3	12	41	6	36	5	4	22	16	3	CI
Trigo	Secano	23	1	7	21	2	17	4		12	7	2	CI
Trigo	Regadío	13	5	4	12	1	10	3		5	6	1	C
Maíz	Regadío	20	8	6	18	2	16	4	1	9	7	2	CD
Alfalfa	Regadío	16	6	5	13	0	11	6		5	6	4	CD
Agricultores general, primer día (1999-00)													
Cebada	Secano	22	3	8	19	2	18	2	5	12	5	0	CI
Trigo	Secano	12	2	3	11	1	12	1	1	8	3	0	CI
Trigo	Regadío	5	1	2	4	0	5	1	0	2	3	0	C
Maíz	Regadío	10	4	4	10	0	6	3	1	4	2	2	CI
Alfalfa	Regadío	7	2	4	7	0	5	2	0	5	0	2	CI
Agricultores general, segundo día (1999-00)													
Cebada	Secano	22	2	5	16	4	17	2	5	13	4	0	CI
Trigo	Secano	12	2	3	11	1	9	2	1	8	3	0	CI
Trigo	Regadío	5	1	2	5	0	4	1	0	2	2	1	C
Maíz	Regadío	10	4	4	10	0	7	3	0	4	3	2	C
Alfalfa	Regadío	7	2	2	6	0	5	3	0	2	3	2	CD

(Resp. = número de casos o respuestas; R. I. = reducción de intervalos, ζ 4.2.3; Sim. = casos que presentan simetría en torno al intervalo central en las frecuencias, ζ 4.3.1; 1 Moda = casos que presentan una moda y su situación N = intervalo de rendimientos normales, M = intervalo de rendimientos malos, B = intervalo de rendimientos buenos; Medianas = situación de la mediana expresada por el número de intervalo, ζ 4.3.2; Sitúa = situación relativa de la mediana: Centrada (C), a la izquierda del intervalo central (I), a la derecha del intervalo central (D)).

Mayoritariamente la moda se sitúa sobre el intervalo central o de rendimientos “normales”, seguido en importancia por el de rendimientos “buenos”. De los casos en los que los sujetos han situado la moda en alguno de los tres intervalos centrales (existen algunos casos poco numerosos en que sitúan la moda sobre alguno de los dos intervalos extremos), el 75% de ellos hacen coincidir moda y rendimientos “normales”, el 17% moda y rendimientos “buenos”, y el 8% moda y rendimientos “malos”. Se observa una clara interferencia del significado material de los términos “bueno”, “malo”, “normal”, etc. con su significado relativo en una distribución de frecuencias. Un 25% de los sujetos no identifican los rendimientos normales como los más frecuentes a la hora de asignar frecuencias.

Esta interferencia del significado del lenguaje natural se puede apreciar también en la importancia de las reducciones de intervalos (ζ 4.2.3) introducidas por los sujetos. En el caso de los expertos, este fenómeno afecta a un 21% de las respuestas de secano y a un 68% de las de regadío. Esta situación se repite cualitativamente en las respuestas de los agricultores: número moderado de casos de reducción de intervalos en los cultivos de secano (4% - 17% de los casos) e importante en los de regadío (20 – 50 % de los casos). Es como si los sujetos no estuvieran de acuerdo en aceptar rendimientos “muy malos” o “muy buenos” como descriptores del recorrido de los cultivos, especialmente en el caso de regadío. Por tanto, no toman su significado en relación a su función (meras marcas para dividir un recorrido de rendimientos), sino en relación a otros significados sustantivos.

La imagen de la función de densidad que transmite la situación de las modas se ha intentado precisar mediante la definición de una mediana referida a intervalos ordinales.

En el caso de los expertos en los cultivos de secano la mediana tiende a situarse en una posición central, estando equilibrados en número de casos a la izquierda o derecha del intervalo central. En los cultivos de regadío por el contrario las medianas tienden a situarse en el centro o en la derecha del punto medio del intervalo central. En ambos casos no se alejan mucho de ese intervalo.

En el caso de los agricultores, la posición central de la mediana se vuelve a confirmar, pero predominando el primer día del primer año de encuesta la posición centro – derecha para todos los cultivos, y el segundo día la posición centro – izquierda para el secano y la centro o centro – derecha para el regadío. En el segundo año, la posición centro – izquierda predomina en los diferentes cultivos en la encuesta del primer día, y la posición centro – izquierda para los de secano y centro para los de regadío. No existe una tendencia clara conjunta, no siquiera por grupos de cultivos. Lo que sí es claro, que la mayoría de las respuestas sitúan la mediana a la izquierda o a la derecha del intervalo central (como ocurría con las modas), y por tanto no se puede señalar que una función de densidad de probabilidades para una parcela concreta tienda a ser simétrica, sino todo lo contrario. Pero no se puede señalar que en unos cultivos predominen claramente las posiciones a la izquierda sobre las posiciones a la derecha.

#4.4.3.- RESULTADOS EN FUNCIONES SOBRE INTERVALOS NUMÉRICOS

Como consecuencia de la definición de los intervalos ordinales en términos de intervalos numéricos realizada por los sujetos (“traducción” de los intervalos), se han construido las distintas f.d.p. sobre estos últimos. El resumen de los principales resultados obtenidos se puede consultar en la tabla resumen 8. La mayoría de las modas y de las medianas se encuentran relativamente centradas, aunque sin definir funciones simétricas.

La forma de las funciones que se derivan de la traducción de los intervalos no es muy distinta a la que se discutió en el caso de los intervalos ordinales. Tampoco existen diferencias profundas con la descripción que se hizo a partir de las estimaciones puntuales (R_{min}, R_{mf}, R_{max}), caso en el que quizás la mayor diferencia se presenta en los expertos, que estimaban los cultivos de regadío con modas predominantemente centro – izquierda y que describen frecuencias más compatibles con modas centro – derecha.

Existe un predominio claro de las posiciones centrales, con una cierta tendencia a predominar las posiciones centro – derecha de la moda, lo que parece compatible con las descripciones predominantes en la literatura de funciones de densidad de probabilidad de rendimientos de cultivos con asimetría negativa.

Se debe matizar que esa predominancia no significa que no exista un número elevado de casos con las modas hacia la izquierda.

El estudio de la posición de la mediana indica que, en un recorrido normalizado [0,1], los expertos sitúan la mediana en la región central (0,4 – 0,6), que equivaldría a la amplitud del intervalo central en una partición de cinco intervalos iguales, en un 75 – 80% de los casos de secano y en un 70% de los casos de regadío. Sin embargo los agricultores de los estudios en profundidad solamente situaron la mediana en esa región un 40 – 45% de los casos.

Esa tendencia se repitió en las encuestas generales a los agricultores. El primer año, en el conjunto de encuestas recuperadas, la mediana se situó en la región central (0,4 – 0,6) entre un 44% - 63% de casos en las respuestas del primer día, mientras que en las del segundo día ese porcentaje osciló entre el 56% - 75% (excluidas alfalfa y maíz para los que había pocos datos). En el segundo año, el primer día se observaron valores muy bajos en los cultivos de maíz y alfalfa de regadío (del orden del 30%), cultivos para los que había pocos datos, y relativamente altos en los cultivos de secano y en el trigo de regadío (del orden del 50 – 67%). En el segundo día los cultivos con más información presentaron valores del orden del 60 – 75%.

Tabla 8.

Resumen de formas de las funciones de densidad de rendimientos definidas sobre intervalos numéricos.

Intervalo Numérico			**Modas**								**Medianas**			
Clases	**Cultivo**	**Total***	**I**	**C**	**D**	**-2,-1**	**-1- 0**	**0 -1**	**1-2**	**Sitúa**	**0,3-0,4**	**0,4-0,5**	**0,5-0,6**	**0,6-0,7**
Expertos (estimación de frecuencias)														
Clase I	Secano	9	4	1	4	1	4	4		C		4	4	
	Regadío	4	1	1	2		2	2		CD		1	2	1
Todas	Secano	32	18	3	11	7	14	10	1	CI	3	13	10	2
	Regadío	27	9	4	14	1	12	12		CD	2	7	12	4
Expertos (estimación directa)														
	Secano	27	7	13	7									
	Regadío	28	2	16	10									
Profundidad														
Cebada	Secano	20	10	0	10					CD	4	8	1	1
Trigo	Regadío	20	1	1	18					CD			8	8
Maíz	Regadío	16	3	1	12					CD	1	2	5	6
Agricultores general, primer día (1998-99). Clase I														
Cebada	Secano	18	6	2	8	3	5	4	1	C	2	5	4	4
Trigo	Secano	13	3	1	7		4	3		CD	1	4	2	4
Trigo	Regadío	4	1	0	3			1	1	CD	1		2	
Maíz	Regadío	5	2	1	2	1	2	1		C	2	2	1	1
Alfalfa	Regadío	5	1	0	4	1		4		CD	1		1	2
Agricultores general, primer día (1998-99). Todas														
Cebada	Secano	45	19	4	20	9	12	10	3	CI	7	13	10	8
Trigo	Secano	25	8	5	10		8	6	2	C	1	7	4	9
Trigo	Regadío	12	3	1	8		3	2	4	CD	2	2	5	1
Maíz	Regadío	19	9	2	8	2	7	5	1	CI	5	5	7	3
Alfalfa	Regadío	16	1	0	15	1	0	12	2	CD	1	0	7	6
Agricultores general, segundo día (1998-99). Clase I														
Cebada	Secano	12	4	1	5	1	4	3	1	C	0	5	3	2

Intervalo Numérico			**Modas**									**Medianas**			
Clases	**Cultivo**	**Total***	**I**	**C**	**D**	**-2,-1**	**-1- 0**	**0 -1**	**1-2**	**Sitúa**	**0,3-0,4**	**0,4-0,5**	**0,5-0,6**	**0,6-0,7**
Trigo	Secano	7	1	1	3	0	2	1	1	C	1	1	2	2
Trigo	Regadío	4	2	0	2	1	1	0	1	C	1	1	1	0
Maíz	Regadío	4	1	0	3	0	1	2	1	CD	0	1	3	0
Alfalfa	Regadío	1	0	0	1	0	0	1	0	C	0	0	0	0
Agricultores general, segundo día (1998-99). Todas														
Cebada	Secano	45	20	4	19	5	17	14	3	CI	4	16	14	7
Trigo	Secano	23	8	2	11	1	8	3	3	CI	1	6	7	6
Trigo	Regadío	13	5	1	7	2	4	3	2	C	2	3	6	1
Maíz	Regadío	20	9	1	10	4	5	4	5	C	1	8	7	3
Alfalfa[+]	Regadío													
Agricultores general, primer día (1999-00). Todas														
Cebada	Secano	21	6	2	11	1	5	9	2	CD	2	5	9	3
Trigo	Secano	12	5	1	6	2	1	6	0	CI	4	2	4	1
Trigo	Regadío	5	1	1	3	0	2	2	1	C	0	2	1	2
Maíz	Regadío	10	2	2	6	1	2	1	1	C	2	1	2	1
Alfalfa	Regadío	7	1	1	5	1	1	1	1	CD	1	2	0	3
Agricultores general, segundo día (1999-00). Todas														
Cebada	Secano	21	7	3	9	2	5	7	1	C	3	6	7	4
Trigo	Secano	12	4	3	5	1	4	4	1	C	1	4	5	1
Trigo	Regadío	5	3	0	2	1	2	0	2	C	0	3	0	2
Maíz	Regadío	10	2	3	5	1	4	0	1	CI	1	2	2	2
Alfalfa	Regadío	7	1	0	6	1	0	1	2	CD	0	1	0	4

* Total = casos excluidas las distribuciones uniformes; Modas = número de casos que presentan la moda a la izquierda (I) del intervalo central, en el intervalo central (C) o a la derecha del intervalo central (D); situación en una escala relativa ζ 4.3.3.; Sitúa= situación relativa de la moda con respecto al intervalo central; Medianas= situación relativa de la mediana en el intervalo normalizado [0,1].

[+] 2 observaciones.

Se puede concluir que las f.d.p. mayoritariamente (del 50% al 75%) presentan sus medianas en el intervalo central (0,4 – 0,6), acentuando la percepción de funciones centradas, que por los sesgos educativos evidentes han sido tradicionalmente descritas como próximas a la normalidad.

#4.4.4.- RESULTADOS SOBRE RECORRIDOS

En la tabla 9 se han sintetizado algunas de las relaciones exploradas para comparar los diferentes recorridos.

La relación $R_L = a\ R_C$, trataba de buscar una correspondencia entre la amplitud de los tres intervalos centrales (R_C) y el recorrido total de los cinco intervalos. Aunque inicialmente esta exploración se realizó para recuperar la información perdida sobre el recorrido en las encuestas con respuestas de la clase II (ζ 4. 2. 1), después se comprobó que tiene cierto interés para estudiar las estrategias de traducción de intervalos ordinales a numéricos. En efecto, como se discutirá en los capítulos correspondientes, no parece que exista una tendencia sistemática a dar amplitudes distintas a cada uno de los intervalos ordinales cuando se realiza su traducción a numéricos, excepto en el caso del intervalo central (rendimientos “normales”), que parece que tiende a ser definido como más amplio que los demás, posiblemente constituyendo un punto de anclaje al asignar las amplitudes.

Si la traducción de la escala ordinal a la numérica se hiciera dando la misma amplitud a los cinco intervalos, se tendría una relación R_L = a R_C dada por a= 5/3 = 1,66. Valores inferiores a 1,66 indicarían intervalos extremos más cortos que los centrales, y lo contrario significarían los valores superiores a 1,66.

En el caso de los expertos, se han podido utilizar un número reducido de encuestas para esta exploración (respuestas de la Clase I), pero los resultados son compatibles con la uniformidad de amplitudes en el secano, y en el regadío con su tendencia a reducir intervalos (ζ 4.2.3).

En las encuestas a los agricultores se observa en el primer año una cierta estabilidad para el valor de la cebada de secano (1,49 el primer día y 1,48 el segundo), valores semejantes a los del segundo año (1,49 – 1,52, el primer y segundo día). En la traducción se estaría reduciendo la amplitud media de los intervalos extremos (rendimientos “muy malos” y “muy buenos”) a aproximadamente una amplitud ¾ de la del intervalo central.

En el caso del trigo de secano los valores correspondientes (1,66 – 1,55; 1,57 – 1,53) y del trigo de regadío (1,67 – 1,61; 1,57 – 1,67) indican una tendencia a expresar los extremos con amplitudes semejantes a las de los intervalos centrales, y en conjunto, una mayor uniformidad en las amplitudes de los cinco intervalos.

En el caso del maíz de regadío (para el que se ha podido recuperar el segundo año un número de casos razonable), los valores 1,52 – 1,46 subrayarían la tendencia comentada a la reducción de intervalos en los cultivos de regadío, tanto entre los expertos como entre los agricultores (términos como rendimientos “muy buenos” o “muy malos” serían juzgados como carentes de sentido en condiciones de regadío).

La siguiente relación explorada tiene que ver con las amplitudes definidas en la traducción de los intervalos ordinales a numéricos (R_L) y la definida mediante estimaciones puntuales (R_{max}, R_{min}). Creemos que, por razones lógicas, la relación más significativa es R_L = a (R_{max} – R_{min}), que será la discutida.

En el caso de los expertos, la amplitud definida mediante atribución de frecuencias es significativamente superior (1,41) a la deducida de la estimación puntual en el caso de secano, mientras que ambos recorridos son similares para el caso de regadío (1,02). Esto sugiere una mayor confianza de los expertos a la hora de realizar las estimaciones del recorrido en los cultivos de regadío, que es coherente con su tendencia a reducir intervalos. Las condiciones de secano, donde su propia experiencia forma parte de una sometida a una fuerte varianza, les llevaría a anclar los valores R_{max} y R_{min} sobre el intervalo central, y a un reconocimiento de mayores amplitudes al examinar intervalos distintos ordinales a la hora de asignar la frecuencia.

A partir de lo anterior se conjetura la importancia de los intervalos ordinales en la obtención de información. Al forzar a los sujetos a considerar los años “muy malos” o “muy buenos” posiblemente se les impulse a considerar extremos que en otros casos no hubieran considerado, anclando sus respuestas en torno a los valores centrales.

Los mismos valores, en el caso de los agricultores (excepto en el caso del trigo de regadío, donde podemos dudar del valor por los pocos datos disponibles) señalan el primer año valores relativamente próximos tanto en secano como en regadío, posiblemente porque reflejan más estrechamente su propia experiencia. En el segundo año, sin embargo, los valores se disparan al alza, indicando grandes diferencias de recorrido, curiosamente mayores en los cultivos de regadío.

Esos resultados subrayan la importancia de estimaciones cruzadas de información, y sugieren anclajes en torno a los valores centrales de las estimaciones puntuales (R_{max}, R_{min}).

Tabla 9.

Resumen de las relaciones entre recorridos.

		Casos	R_L=a R_C	R_L = a (R_{max} – R_{min}) + b	R_L=a (R_{max} – R_{min})	R_{min} = a $R_{<MM}$	R_{max} = a $R_{>MB}$	$R_{mf} \subset R_{NN}$ I y II	R_{NN} = R_{mf} III, IV
Expertos									
	Secano	10	a= 1,65	a= 1,29	a= 1,41	a= 1	a= 0,92	95%	27%
	Regadío	4	a= 1,44	a= 0,96	a= 1,02	a= 1	a= 0,99	92%	75%
Agricultores general , primer día 1998-99									
Cebada	Secano	19	a= 1,49	a= 0,882	a= 1,10	a= 1,05	a= 0,97	87%	62%
Trigo	Secano	13	a= 1,66	a= 0,907	a= 1,10	a= 0,79	a= 0,98	94%	57%
Trigo	Regadío	5	a= 1,67	a= 0,942	a= 1,40	a= 1,06	a = 0,91	88%	20%
Maíz	Regadío	6	a= 1,91	a= 0,945	a= 1,02	a= 1,01	a= 0,99	83%	38%
Alfalfa	Regadío	5	a= 1,37	a= 1,04	a= 1,02	a= 1,00	a= 0,99	78%	25%
Agricultores general , segundo día 1998-99									
Cebada	Secano	12	a= 1,48	a= 0,16	a= 1,16	a= 1,1	a= 0,94	78%	68%
Trigo	Secano	7	a= 1,55	a= 0,03	a= 1,18	a= 1,06	a= 1,01	73%	78%
Trigo	Regadío	4	a= 1,61	a= 0,47	a= 1,10	a= 0,98	a= 0,93	86%	33%
Maíz	Regadío	4	a= 1,64	a= 0,54	a= 1,08	a= 1,20	a= 0,96	67%	27%
Alfalfa	Regadío	2	*	*	*	*	*	50%	62%
Agricultores general , primer día 1999-00									
Cebada	Secano	22	a= 1,49	a= 0,41	a= 1,36	a= 1,34	a= 0,88	86%	
Trigo	Secano	12	a= 1,57	a= 0,66	a= 1,30	a= 1,20	a= 0,89	83%	
Trigo	Regadío	5	a= 1,57	a= 0,34	a= 1,53	a= 1,67	a= 0,97	60%	
Maíz	Regadío	10	a= 1,52	a= 0,25	a= 1,49	a= 1,35	a= 0,83	80%	
Alfalfa	Regadío	7	a= 1,64	a= 1,36	a= 1,59	a= 1,45	a= 0,89	43%	
Agricultores general , segundo día 1999-00									
Cebada	Secano	22	a= 1,52	a= 0,33	a= 1,25	a= 1,27	a= 0,91	77%	
Trigo	Secano	12	a= 1,53	a= 0,50	a= 1,20	a= 1,17	a= 0,95	67%	
Trigo	Regadío	5	a= 1,67	a= 0,18	a= 1,25	a= 1,37	a= 0,99	40%	
Maíz	Regadío	10	a= 1,46	a= 0,76	a= 1,41	a= 1,08	a= 0,86	60%	
Alfalfa	Regadío	7	a= 1,60	a= 1,07	a= 1,18	a= 1,20	a= 0,98	71%	

*2 observaciones

La siguiente cuestión tiene que ver con la proximidad de las estimaciones puntuales (R_{max}, R_{min}) y los extremos de los recorridos definidos numéricamente al traducir los intervalos ordinales.

En el caso de R_{min} y el extremo inferior del recorrido en el que se realiza la asignación de frecuencias, $R_{<MM}$, o frontera izquierda de la clase de rendimientos “muy malos”:

En los expertos ambos valores tienden a coincidir, lo cual es propio de los criterios de coherencia más simples de controlar a la hora de contestar una encuesta (aunque en los extremos superiores no se haya mantenido esta tendencia).

En los agricultores, en general, se advierte una tendencia clara $R_{min} > R_{<MM}$, coincidente con la hipótesis de reducción de los recorridos comentada anteriormente, e incrementada por la tendencia a dar valores cero al extremo inferior de los rendimientos “muy malos”, lo cual es coherente con el significado del término, pero no con la descripción del recorrido.

Algo similar ocurre con la regresión de R_{max} = a $R_{>MB}$ (este último el extremo superior del intervalo de rendimientos “muy buenos”), donde las estimaciones directas tienden a estimar recorridos más reducidos.

Por último se investiga el porcentaje de veces en que la estimación puntual de rendimientos más frecuentes R_{mf} estaría situada en el intervalo central de los rendimientos "normales", observándose cierta disparidad en los valores.

El análisis anterior pone de manifiesto el interés de seguir explorando las relaciones entre diversos sistemas de definición de las amplitudes de los intervalos y del recorrido de los rendimientos, ya que es posible que existan anclajes poderosos en todos los casos.

#4.4.5.- RESULTADOS SOBRE MEDIAS

Es posible comparar el rendimiento medio calculado a partir de la distribución de frecuencias, R_{xf}, con el estimado directamente, R_m, o con los calculados a partir de las aproximaciones funcionales RMT (triangular) y RMB (beta – PERT).

La encuesta en profundidad a los agricultores se realizó de forma suficientemente cuidadosa para que los valores estimados por los agricultores sean bastante representativos de sus opiniones. En la tabla 10 se han comparado los valores calculados R_{xf} con las estimaciones puntuales de rendimiento medio. El valor R_{xf1} se refiere a la distribución de frecuencias realizada por el agricultor con un recorrido establecido por él mismo y expresión gráfica de las frecuencias con un número limitado de diez monedas. El valor R_{xf2} se obtuvo con un recorrido muy amplio común para todos los agricultores y sin limitación de monedas.

Tabla 10.

Diferencias entre valores de los rendimientos medios. Encuesta a los agricultores (en profundidad), 1997-98.

			$(R_{xf1}$-$R_m)$:R_m			**$(R_{xf2}$-$R_m)$:R_m**			**$(R_{xf1}$-$R_{xf2})$:min (R_{xfi})**		
Cultivo		**Total**	**<15%**	**15-25%**	**25-50%**	**<15%**	**15-25%**	**25-50%**	**<15%**	**15-25%**	**25-50%**
Agricultores profundidad											
Cebada	Secano	10	5	3	2	5	1	4	6	3	1
Trigo	Regadío	10	8	2	0	8	2	0	10	0	0
Maíz	Regadío	8	7	1	0	7	1	0	7	1	0

(<15% diferencias inferiores al 15%; 15-25%, diferencias entre el 15 y el 25%; 25-50% diferencias entre el 25 y el 50%).

Obsérvese que en general existe una tendencia a obtener valores de los rendimientos medios relativamente próximos, frecuentemente por debajo de una diferencia del 15%. Estos resultados señalan una gran coherencia entre las estimaciones puntuales de rendimientos medios (R_m), los rendimientos calculados a partir de las distribuciones de frecuencia (R_{xf}), y entre estos valores calculados con datos de diferentes días y condiciones de asignación.

En la tabla 11 se presentan los valores de las medianas de las distribuciones de las diferencias entre los valores calculados de los rendimientos medios a partir de las estimaciones de frecuencias (R_{xf}) y los señalados puntualmente por los sujetos, R_m. Así, por ejemplo, en el 50% de los casos, las estimaciones de frecuencias de los expertos y su traducción de intervalos ordinales a numéricos determinan valores de los rendimientos medios R_{xf} que se diferencian de los directamente estimados R_m en menos de un 18,3% para los cultivos de secano, y un 11% para los cultivos de regadío. Aunque esta estimación en el caso de los cultivos de regadío no se juzga especialmente buena, en el conjunto de los resultados señala una gran coherencia en las estimaciones de las medias.

Tabla 11.

Valores de las medianas de las distribuciones de diferencias porcentuales $|R_{xf} – R_m|: R_{xf}$.

	Medianas de la distribuciones $|R_{xf} – R_m|: R_{xf}$				
	Secano		**Regadío**		
Expertos	0,183		0,110		
	Cebada	**Trigo**	**Trigo**	**Maíz**	**Alfalfa**
Agricultores 98-99 (primer día)	*	*	*	*	*
Agricultores 98-99 (segundo día)	0,068	0,081	0,038	0,081	**
Agricultores 99-00 (primer día)	0,092	0,075	0,125	0,067	0,125
Agricultores 99-00 (segundo día)	0,112	0,075	0,138	0,062	0,075

*No se dispone de la estimación de R_m.
**2 observaciones.

En la tabla 12, se presenta la comparación entre los rendimientos medios calculados a partir de las distribuciones de frecuencia (R_{xf}) y los calculados a partir de las estimaciones puntuales (R_{min}, R_{mf}, R_{max}), mediante las aproximaciones funcionales triangular y beta – PERT (RMT y RMB respectivamente, y RMX genéricamente). De nuevo, excepto las estimaciones de los expertos para los cultivos de secano y en cultivos con pocas observaciones, se obtiene una coherencia razonable en las estimaciones de los valores medios mediante las diferentes aproximaciones empleadas en la investigación.

Tabla 12.

Valores de las medianas de las distribuciones de diferencias porcentuales $|R_{xf} – RMT|: R_{xf}$ y $|R_{xf} – RMB|: R_{xf}$.

	Medianas distribuciones $|R_{xf} – RMX|: R_{xf}$				
	TRIANGULAR				
	Secano		**Regadío**		
Expertos	0,160		0,071		
	Cebada	**Trigo**	**Trigo**	**Maíz**	**Alfalfa**
Agricultores 98-99 (primer día)	0.09	0.096	0.092	0.06	0.088
Agricultores 98-99 (segundo día)	0.10	0.058	0.068	0.112	*
Agricultores 99-00 (primer día)	0.10	0.10	0.125	0.10	0.175
Agricultores 99-00 (segundo día)	0.12	0.083	0.138	0.042	0.044
	BETA				
	Secano		**Regadío**		
Expertos	0.138		0.050		
	Cebada	**Trigo**	**Trigo**	**Maíz**	**Alfalfa**
Agricultores 98-99 (primer día)	0.100	0.105	0.075	0.075	0.100
Agricultores 98-99 (segundo día)	0.120	0.081	0.041	0.100	*
Agricultores 99-00 (primer día)	0.100	0.100	0.125	0.075	0.175
Agricultores 99-00 (segundo día)	0.112	0.100	0.175	0.050	0.075

* No se consideró por tener sólo dos observaciones.

#4.4.6.- RESULTADOS SOBRE VARIANZAS

Si bien los resultados anteriores sugieren una buena estimación de los valores centrales, y concretamente de los rendimientos medios, que posiblemente podrían mejorarse mediante el desarrollo de protocolos adecuados de interrogación y de interpretación, las comparaciones entre las varianzas calculadas a partir de las distribuciones de frecuencias, V_{xf}, y calculadas a partir de las estimaciones funcionales triangular y beta – PERT son francamente diferentes (tabla 13). Ello nos lleva a conjeturar, por un lado, que existen problemas de anclajes en la estimación de intervalos, y por otro que posiblemente parte de la diferencia porcentual detectada sea un artificio derivado de las propias ecuaciones.

Tabla 13

Valores de las medianas de las distribuciones de diferencias porcentuales de las varianzas calculadas $|V_{xf} – VT|: V_{xf}$ y $|V_{xf} – VB|: V_{xf}$.

	Medianas distribuciones $\|V_{xf} – VX\|: V_{xf}$				
	TRIANGULAR				
	Secano		**Regadío**		
Expertos	0.371		0.300		
	Cebada	**Trigo**	**Trigo**	**Maíz**	**Alfalfa**
Agricultores 98-99 (primer día)	0.606	0.531	0.631	0.533	0.500
Agricultores 98-99 (segundo día)	0.621	0.538	0.638	0.650	*
Agricultores 99-00 (primer día)	0.700	0.633	0.675	0.900	0.850
Agricultores 99-00 (segundo día)	0.720	0.650	0.750	0.700	0.650
	BETA				
	Secano		**Regadío**		
Expertos	0.633		0.480		
	Cebada	**Trigo**	**Trigo**	**Maíz**	**Alfalfa**
Agricultores 98-99 (primer día)	0.428	0.550	0.508	0.333	0.400
Agricultores 98-99 (segundo día)	0.507	0.438	0.438	0.533	*
Agricultores 99-00 (primer día)	0.500	0.500	0.525	0.850	0.750
Agricultores 99-00 (segundo día)	0.533	0.450	0.725	0.500	0.450

* No se consideró por tener sólo dos observaciones.

#4.5.- VARIABLES Y TÉRMINOS

Beta PERT = función beta, con las hipótesis habitualmente admitidas en el PERT

Bi = función de densidad de probabilidad bimodal (ver #4.3.3)

C (posición de la moda) = moda en el punto medio del recorrido o centrada (ver #4.3.3)

Clase (de respuesta) I. (ver #4.2.1)

Clase (de respuesta) II, IIa, IIb. (ver #4.2.1)

Clase (de respuesta) III. (ver #4.2.1)

Clase (de respuesta) IV, IVa, IVb. (ver #4.2.1)

D (posición de la moda) = moda a la derecha del punto medio del recorrido (ver #4.3.3)

Discontinuidad de intervalos (ver #4.2.2)

f.d.p. = Función de densidad de probabilidad

Función de densidad = Función de densidad de probabilidad (f.d.p.)

Hipótesis sobre amplitud igual de los intervalos (ver #4.2.4.1)

I (posición de la moda) = moda a la izquierda del punto medio del recorrido (ver #4.3.3)

IM = índice de posición relativa de la moda (ver #4.3.3)

Indice de continuidad de intervalos (ver #4.2.2.1)

Indice de posición relativa de la moda (ver #4.3.3)

Inter p = Inter q: indica el número de casos en los que la frecuencia asignada al intervalo p coincide con la frecuencia atribuida al intervalo q

$\mathbf{R_C}$ = Recorrido corto (ver #4.2.1)

$\mathbf{R_L}$ = Recorrido largo (ver #4.2.1)

$\mathbf{R_m}$ = Rendimientos medios estimados puntualmente

$\mathbf{R_{max}}$ = $\mathbf{r_{max}}$ = Rendimientos máximos

$\mathbf{R_{mf}}$ = Rendimientos más frecuentes o normales

$\mathbf{R_{min}}$ = $\mathbf{r_{min}}$ = Rendimientos mínimos

$\mathbf{R_{>MB}}$ = extremo superior de la clase de rendimientos “muy buenos” (numérico)

$\mathbf{R_{<MM}}$ = extremo inferior de la clase de rendimientos “muy malos” (numérico)

$\mathbf{R_{NN}}$ = intervalo de rendimientos “normales” (numérico)

$\mathbf{R_{xf}}$ = valor medio del rendimiento calculado a partir de una distribución de frecuencias (Σ marca de clase x frecuencia de la clase)

RMB = (1/6) (R_{min} + 4 R_{mf} + R_{max}) valor del rendimiento medio suponiendo una aproximación mediante una función beta con las hipótesis aceptadas en el PERT (beta – PERT) a (R_{min}, R_{mf}, R_{max})

RMT= (1/3) (R_{min} + R_{mf} + R_{max}) valor del rendimiento medio suponiendo una aproximación triangular a (R_{min}, R_{mf}, R_{max})

Rr = rendimientos en cultivos de regadío

Rs = rendimiento en cultivos de secano

Reducción de intervalo (ver #4.2.3)

Reducción de intervalo “A”, “I”, “D” (ver #4.3.1)

Solapamiento de intervalos (ver #4.2.2)

Superposición de intervalos (ver #4.2.2)

Un = función de densidad de probabilidad uniforme (ver #4.3.3)

VB = $(1/36)\ (R_{max} - R_{min})^2$, valor de la varianza del rendimiento suponiendo una aproximación beta – PERT a $(R_{min}, R_{mf}, R_{max})$

$\mathbf{V_{xf}}$ = varianza calculada a partir de la distribución de frecuencia (Σ (media – marca clase)^2 x frecuencia de clase)

VT = $(1/18)\ [(R_{max} - R_{min})^2 + (R_{mf} - R_{min})\ (R_{mf} - R_{max})]$, valor de la varianza del rendimiento suponiendo una aproximación triangular a $(R_{min}, R_{mf}, R_{max})$

CAPÍTULO 5

ENCUESTA A LOS EXPERTOS

#5.1.- ENCUESTA A LOS EXPERTOS

En el Anejo 1 se han reproducido las instrucciones, el listado de destinos y el cuestionario utilizados para interrogar a los expertos sobre cuestiones relacionadas con la relación insumos – producto y la forma de la función de densidad de probabilidad de los rendimientos.

Se enviaron 185 encuestas a expertos de áreas afines a la Fitotecnia de las diversas Escuelas Técnicas y Superiores de Ingeniería Agraria, de los cuales han respondido 38, y algunos disculparon la falta de respuesta. Las encuestas se enviaron el 12 de junio y el 4 de noviembre de 1997, con el mismo cuestionario.

El cuestionario constaba de una introducción con instrucciones y de 12 cuestiones.

Inicialmente se pedía a los expertos que se clasificaran, según su especialidad, por tipos de cultivos. El resultado de esta clasificación se refleja en la tabla 14.

Tabla 14.

Perfil de los expertos encuestados (1997-98).

PERFIL	**NÚMERO (AUTOCLASIFICACIÓN)**
Cultivos anuales herbáceos extensivos	27
Cultivos anuales herbáceos intensivos	20
Cultivos herbáceos muy intensivos (invernadero)	7
Horticultura	9
Fruticultura	12
Cultivos herbáceos plurianuales extensivos (pastos, etc.)	7
Cultivos herbáceos plurianuales intensivos (alfalfa, etc.)	13
Cultivos plurianuales leñosos en secano (viña, olivar, etc.)	12
Cultivos anuales leñosos en regadío	13

Las zonas a las que se refieren los datos facilitados en las encuestas, pueden consultarse en la tabla 15.

Tabla 15.

Respuestas por referencias geográficas. Encuesta a expertos (1997-98).

Región	**Número de encuestas**
Andalucía	8
Aragón	2
C. La Mancha	2
C. León	9
Cataluña	8
Centro - Madrid	3
Navarra - Rioja	2
Valencia- Murcia	4
TOTAL	**38**

Los cultivos a los que se refieren los datos aportados por los encuestados se reflejan en las tablas 16 y 17.

Tabla 16.

Cultivos de secano referidos en las contestaciones. Encuesta a expertos (1997-98).

CULTIVOS DE SECANO	NÚMERO DE ENCUESTAS
Almendro	1
Avena	1
Cebada	13
Centeno	1
Garbanzo	2
Girasol	6
Lenteja	1
Olivo	3
Trigo	13
Veza	2
Vid	3
TOTAL	**46**

Tabla 17.

Cultivos de regadío referidos en las contestaciones Encuesta a expertos (1997-98).

CULTIVOS DE REGADÍO	NÚMERO DE ENCUESTAS
Albaricoquero	1
Alcachofa	2
Alfalfa	2
Algodón	1
Bróculi	1
Cebada	3
Ciruelo	1
Coliflor	1
Espárrago	1
Fresa	1
Girasol	2
Limonero	2
Maíz	8
Manzano	1
Melocotonero	2
Olivo	2
Patata	2
Peral	1
Pimiento	1
Remolacha	7
Tomate	1
Trigo	3
TOTAL	**46**

Las cinco primeras preguntas del cuestionario tenían por objeto establecer la opinión de los expertos sobre la utilización de los insumos por parte de los agricultores: sin variación en la composición relativa de una formulación ("receta") o adaptándola a las distintas circunstancias tecnológicas y de mercado. La cuestión se planteaba por su propio interés, y porque en los manuales de estos especialistas normalmente las recomendaciones sobre utilización de insumos para un determinado cultivo se suelen dar en forma de "recetas de cocina", en enunciados similares al siguiente: "El cultivo en secano del trigo en tal región exige la aplicación de tantos kg. / ha de abono de fondo en tales épocas, y de una dosis de tantos kg. / ha de abono de cobertera….". Esta percepción puede derivarse de que, dadas unas condiciones tecnológicas y de precios relativos de los insumos y del producto, los agricultores estarán utilizando una determinada combinación de insumos, y que esta combinación no varía o lo hace muy lentamente.

Si este fuera el caso, la producción se caracterizaría por "apostar" a una determinada combinación de insumos, dejando que sean las condiciones climáticas y económicas del año las que determinaran el resultado. Esta imagen se correspondería con un planteamiento estático "ex – ante" de los problemas de optimización.

Sin embargo, si algunos insumos se emplean adaptándose a ciertas condiciones particulares de cada año, por ejemplo, las condiciones climáticas (en un año seco no conviene añadir a los cereales tanto abono nitrogenado de cobertera como en uno húmedo), los agricultores irían tomando sus decisiones de dosis de insumos en función de las manifestaciones que de forma secuencial se fueran presentando a lo largo del año, y nos enfrentaríamos a una situación de decisión relativamente dinámica.

Las seis preguntas últimas (desde la pregunta número 6 a la número 12) tratan de determinar los recorridos y la forma de las funciones de densidad de probabilidad de los rendimientos de los cultivos, especialmente el número de modas y su situación.

Dado que se supone que los expertos tienen una formación estadística suficiente para entender el concepto de función de densidad, de moda, de recorrido, etc., se interroga de forma directa, con ciertas ayudas gráficas en las preguntas 6, 7 y 8. En la pregunta 9 se interroga apelando a sus recuerdos de conceptos estadísticos (funciones binomiales, normales, gamma, beta, etc.). Si un experto ha reflexionado alguna vez sobre la forma funcional de la distribución de los rendimientos, probablemente haya considerado alguna de estas funciones. La pregunta 10 trata de obtener información sobre los factores que se consideran más importantes para explicar la variabilidad de los rendimientos (es una pregunta para reunir información para experimentos futuros). La pregunta 11 intenta obtener valores de los rendimientos en la forma de (mínimo, más frecuente, máximo) utilizada, por ejemplo en la estimación de tiempos en el PERT, y una estimación directa del rendimiento medio, para explorar cuestiones relacionadas con las aproximaciones funcionales (beta, triangular, etc.) y sus valores medios. La cuestión 12 se introdujo para tratar de traducir a una escala cardinal los valores de los intervalos dados en lenguaje natural (escala ordinal) en la pregunta 6.

#5.2.- ESTABILIDAD DE LAS "RECETAS" (preguntas 1 a 5)

Las tres primeras preguntas pretendían identificar si, en opinión de los expertos, los agricultores tienden a utilizar una receta fija de insumos, o si la varían en función de las circunstancias de cada año. La primera y la segunda preguntas eran la misma, expresadas de distinta manera. La tercera pretendía aportar evidencia adicional en caso de incoherencia en las respuestas anteriores.

Las preguntas cuatro y cinco pretendían aportar información adicional y aclaraciones sobre los significados aprehendidos por los expertos en las preguntas 1 – 3.

Una contestación coherente a las cinco preguntas supondría un esquema similar al siguiente (siendo V enunciado verdadero, F enunciado falso, y O enunciado en unos casos verdadero y en otros falso).

Tabla 18.

Esquema de contestaciones coherentes a las cinco primeras preguntas. Encuesta a los expertos (1997-98).

Pregunta	**Respuesta**		
1	V	O	F
2	F	O	V
3	Un valor para cada insumo	Un valor para algunos insumos y un rango para otros insumos	Un rango para todos los insumos
4	No tiene sentido	Alguna respuesta V, O	Alguna respuesta V, O
5	No tiene sentido	Alguna respuesta	Alguna respuesta

En la **pregunta 1** se pedía al experto que señalara si los agricultores utilizan una "receta" de insumos fija, independientemente de las circunstancias particulares del año.

De las 37 respuestas obtenidas, 12 expertos consideran que el enunciado es falso (F), 12 que en algunos casos es verdadero y en otros falso (O), y 13 que es un enunciado verdadero (V). Las respuestas se pueden interpretar como que no existe una opinión general muy precisa. Dos expertos han comentado su respuesta: uno considera el enunciado más verdadero en cultivos de secano; otro señala que el enunciado es verdadero en cultivos de regadío (que están muy "controlados" técnicamente).

La **pregunta 2** (con 38 respuestas) es una formulación alternativa (positiva) de la primera cuestión. Se pregunta si el experto cree que los agricultores utilizan variaciones de una receta, que adaptan en función de las circunstancias particulares del año. Curiosamente en este caso 9 expertos consideran el enunciado falso, 6 (16%) verdadero en algunos casos y falso en otros, y 23 lo consideran verdadero (60%). En esta pregunta se dispone de una respuesta más que en la primera, y esta respuesta es V, por lo que respecto a la primera contestación, en conjunto 4 expertos de 13 han cambiado su opinión de receta fija a variable y 6 expertos de unas veces fija y otras variable a variable.

La coherencia de los expertos entre lo que contestan a la primera cuestión y a la segunda se puede apreciar en la siguiente tabla – resumen (tabla 19) (el número que precede a las claves de respuesta es el correspondiente a las preguntas 1 y 2), donde las combinaciones equivalentes son FV, VF, OO (ya que el sentido de ambas preguntas era opuesto, y por tanto una contestación F en la primera pregunta equivale a una contestación V en la segunda).

Tabla 19.

Respuestas a la primera y segunda preguntas de la encuesta a expertos (1997-98).

	Respuesta a la primera cuestión		
Respuesta a la segunda cuestión	1 V	1 O	1 F
2 V	3	8	11
2 O	3	3	0
2 F	7	1	1

Las combinaciones FV, OO y VF se corresponden con el número de expertos que han mantenido su opinión invariante en las dos formulaciones de la pregunta. De las 37 respuestas a las 2 cuestiones, 21 expertos mantienen su opinión inicial (57%), y de ellos 7 creen que se utilizan recetas fijas, 11 que las recetas son variables y 3 que en unos casos son fijas y en otros variables.

Los 12 expertos (32%) que en la primera pregunta contestaron que la receta podía ser fija en unos casos y variable en otros (O), 3 mantienen su opinión en la segunda, 1 la cambia a receta fija, y 8 a receta variable.

Las respuestas FO y VO se corresponden con personas que cambian su opinión, pero posiblemente matizándola. De estos casos solamente se ha dado el segundo, el VO, es decir, 3 expertos que inicialmente señalaron que la receta era fija y después han matizado que unas veces es fija y otras variable.

Las respuestas FF y VV se corresponden con personas que han contestado de forma contradictoria. Solamente existe un caso FF y 3 VV.

En resumen, de los 33 expertos que han mantenido una opinión compatible en las preguntas 1 y 2, el 33% mantiene una opinión invariante sobre que la receta varía, y el 24% cambia de O a considerarla variable. Los resultados anteriores parecen avalar la siguiente:

Conjetura:

Existen dificultades de comprensión del significado de los distintos conceptos puestos en juegos en las preguntas. Aunque parecen conceptos muy gráficos los de “receta de insumos fija”, “variaciones de una receta de insumos”, “independientemente de las circunstancias particulares del año” o “en función de las circunstancias particulares del año”, el cambio de opinión de los expertos entre las dos formulaciones alternativas de la pregunta no se puede explicar por su exclusivo conocimiento o desconocimiento de las prácticas seguidas por los agricultores, sino más probablemente por los significados atribuidos a las formulaciones, que en bastantes casos han sido percibidas como distintas.

Posiblemente haya un umbral de variación de los insumos que unos expertos consideren como variación y que otros no lo consideren de igual manera. Igualmente, puede que en la formulación de los insumos haya algunos que se mantengan relativamente fijos y otros que varíen en función de las circunstancias (aunque esta situación se podía contestar sin dificultad mediante O). También puede que influyan percepciones de los expertos que no tengan relación con las practicas seguidas por los agricultores.

Δ

Con la **pregunta 3** se intentaba (por otra vía) obtener confirmación sobre las opiniones expresadas en las preguntas 1 y 2. En efecto, si un experto señala en la pregunta 1 que los agricultores utilizan una receta de insumos fija (1V), y en la segunda señala como falso que los agricultores utilicen variaciones de una receta de insumos para adecuarse a las circunstancias del año (2F), sería coherente que después no señalara rangos de variación de insumos en la pregunta 3 (respondiera con valores fijos). (Obsérvese que no podría concluirse incoherencia si diera un rango de variación para los insumos, ya que podría estar refiriéndose a las practicas de los agricultores de una zona, donde cada uno utilizaría una combinación fija, pero que no coincidiese con otra fija de un vecino).

De los 38 encuestados, 12 no respondieron a esta pregunta número 3. Se considera coherente que se haya contestado F o O a la primera cuestión y V o O a la segunda y se haya reflejado en la pregunta 3 una variación de insumos. En la tabla 20 se refleja la casuística existente (donde VO podría también ser compatible con respuestas coherentes con variaciones en la pregunta 3).

Tabla 20.

Combinación de respuestas a las preguntas 1, 2 y 3 (encuesta a expertos, 1997-98).

Pregunta 3	**Coherentes variables**				**Coherentes fijos**		**Incoherentes**			
“Receta”	FV	OV	FO	OO	OF	VF	VO	VV	FF	**TOTAL**
Varía (v)	9	4	0	2	1	1	1	2	0	20
Fija (f)	0	0	0	0	0	4	1	1	0	6
No responde	2	4	0	1	0	2	1	0	1	11
TOTAL	11	8	0	3	1	7	3	3	1	**37**

De las personas que han respondido en las preguntas 1 y 2 coherentemente respuestas compatibles con la existencia de combinaciones o recetas variables de insumos, 7 no han respondido la pregunta 3, y las 16 personas que la han contestado han mantenido consecuentemente su opinión.

De las personas que sostienen en las dos primeras preguntas combinaciones o recetas fijas, dos no contestan, otras dos señalan variaciones de las recetas en la pregunta 3 (que como hemos dicho no puede concluirse que sean respuestas incoherentes, aunque posiblemente lo sean) y 4 responden la pregunta 3 consecuentemente con sus opiniones anteriores.

En la **pregunta 4** se pedía la opinión de los expertos sobre cuáles eran las causas del cambio de la "receta productiva" de insumos utilizada por el agricultor. Las respuestas obtenidas se resumen en la tabla 21. Obsérvese que la mayoría de los encuestados opinan que los cambios están provocados por causas técnicas y económicas (31 respuestas de 38 posibles, 20 de ellas sin matizaciones), y que solamente en 1 caso se mantiene la opinión de que no hay cambio, y que dos personas creen que los cambios están provocados por causas técnicas y no por causas económicas. Ningún experto señala solamente causas económicas.

Tabla 21.

Opinión de los expertos sobre causas técnicas o económicas en los cambios de "receta productiva" (encuesta a expertos, 1997-98).

Cambios: causas económicas	**Los cambios están provocados por causas técnicas**				
	V	O	F	No contesta	**TOTAL**
V	20	4	0	3	27
O	5	2	0	0	7
F	2	0	1	0	3
No contesta	0	0	0	1	1
TOTAL	27	6	1	4	**38**

La **pregunta 5** indagaba en qué tipos de cultivos pensaba el experto que se producían principalmente los cambios discutidos en la pregunta 4. Así, para un conjunto de 6 posibles clases de cultivos (anuales de secano, anuales de regadío, herbáceos plurianuales, leñosos plurianuales, otros cultivos…), el experto debía decir si existe variación o no de la "receta productiva". Eran posibles tres tipos básicos de respuestas: existe siempre variación en los cultivos de los que se dispone de información (V en todos esos cultivos), existe variación clara y condicionada (conjuntos de V y O), no existe nunca variación (todas las respuestas F), etc.

Desde el punto de vista de nuestro interés, las respuestas se evalúan en función de si el experto admite variación o no en las recetas productivas. Para ello se han resumido las distintas respuestas dadas para cada tipo de cultivos (V, O, F) mediante las operaciones $X \cap X = X$ y $X \cap Y = O$, con $X \neq Y$ y X, Y = {V,O,F}. En particular, $V \cap O = O$, $F \cap O = O$, $V \cap F = O$. Al realizar el análisis se obtuvieron 29 casos de V, 8 casos de O y ninguno de F en 37 respuestas obtenidas.

Así las respuestas obtenidas en esta pregunta (número 5) son compatibles únicamente con la opinión de que la "receta" productiva varía, al menos en algunas ocasiones.

Como en un proceso de destilación, se ha comenzado con una pregunta que obtenía respuestas V, O, F muy dispares en la misma proporción (siguiendo una distribución uniforme, que indica la máxima dispersión de opiniones). En la segunda pregunta aumentó el número de respuestas favorables a la variación de las recetas, pero aún con un importante porcentaje de negaciones. En las preguntas 3 y 4 los expertos matizaban sus respuestas. Finalmente en la pregunta 5 todos los expertos han expresado que las recetas productivas siempre varían o que no varían para unos cultivos, pero sí para otros. Parece, por tanto, razonable concluir:

Conclusión:

Las preguntas son interpretadas por los expertos, que posiblemente se ven enfrentados a una opinión que si no tiene los matices que normalmente introducirían les induce a negarla (o a afirmarla si confirma sus matices). La formulación de la pregunta puede inducir un tipo u otro de respuesta, sobre todo cuando los términos con que se formulan no son precisos en el ámbito de especialidad del experto.

Los expertos mantienen una opinión favorable a la utilización de "recetas variables de insumos", pero esta expresión es extraordinariamente ambigua y posiblemente poco útil para la obtención de información y su discusión.

Una cuestión, por tanto, no debe darse por bien interpretada por el hecho de estar bien formulada. Para obtener una respuesta precisa a una cuestión es conveniente introducir un conjunto de preguntas cruzadas sobre el particular. El problema es qué álgebra puede introducirse para obtener una conclusión en función de las respuestas a cada pregunta de un grupo diseñado para la determinación de una respuesta precisa a una cuestión concreta.

Posiblemente fuera de interés el desarrollo de una sistemática de formulación de cuestiones (afirmaciones, negaciones, intersecciones, etc.) de forma semejante a como se diseñan los conjuntos de apuestas en la obtención de las probabilidades subjetivas.

Δ

#5.3.- FUNCIONES DE DENSIDAD DE RENDIMIENTOS CON INTERVALOS ORDINALES

En la **pregunta 6** se ensayó una formulación basada en una escala ordinal para obtener una función de densidad de los rendimientos de los cultivos. En el cuestionario se definieron cinco intervalos en lenguaje natural (porcentaje de años *muy malos*, *malos*, *normales*, *buenos* y *muy buenos*) y a los que el experto debía asignar una frecuencia relativa. Se obtuvieron 33 respuestas de 38 posibles (y en 5 de ellas los porcentajes atribuidos a las diferentes clases de años no sumaban 100 %).

Los cultivos especificados por los expertos en sus respuestas fueron los indicados en las Tablas 16 y 17, donde se indican los de secano y de regadío. Dada la gran variabilidad en los cultivos utilizados en las respuestas por los expertos, resulta poco útil considerarlos de forma individual, por lo que se estudiarán de forma conjunta, diferenciando entre secano y regadío, y excluyendo frutales y leñosos, por tratarse de cultivos no anuales, único objeto del interés en este trabajo.

Dadas las asignaciones de frecuencia de rendimientos para cada cultivo en los cinco intervalos definidos en lenguaje natural, queda dibujada la forma de una función de densidad de probabilidad, aunque referida a un recorrido dividido en clases ordinales (intervalos definidos cualitativamente). En la tabla 22 se resumen algunos de los rasgos que se pueden deducir sobre la forma de la f.d.p. a partir de las contestaciones de los expertos.

Tabla 22.

Forma de las funciones de densidad (intervalos ordinales). Encuesta a los expertos (1997-98).*

Cultivos	**Reducción de intervalo**			**Uniforme**	**Extremos < 15%**	**Simetría**	**Inter 1 = Inter 5**	**Inter 2 = Inter 4**	**Total**
	I	A	D						
Secano	1	3	3	1	10	11	19	12	33
Regadío	15	8	0	0	22	11	17	19	34

* Significados de los encabezamientos: consultar ζ 4.3.1

Una primera cuestión que llama la atención es que, en los cultivos de regadío, en 23 casos de 34 (67,7%) los expertos han efectuado reducciones de intervalos (ζ 4.3.1), amputando algunas clases de la distribución asignándoles una frecuencia nula. En 8 casos se ha asignado una frecuencia nula a los dos intervalos extremos y en 15 al intervalo izquierdo. Es como si los expertos no estuvieran de acuerdo con la existencia de años "muy malos" en regadío (23 casos de 34) o de años "muy buenos" (8 casos de 34).

El mismo fenómeno se presenta en 7 casos de 33 para los cultivos de secano. Estos 7 expertos han asignado una frecuencia nula a los intervalos de rendimiento "años muy malos" y/o "años muy buenos", amputándolos de la distribución.

Para explicar este comportamiento, avanzamos la siguiente

Conjetura:

Los expertos no consideran que se esté preguntando sobre cinco intervalos de la función de densidad de probabilidades, sino sobre un significado material concreto y definido en un marco distinto al de la mera división en intervalos de un recorrido. Así, en los cultivos de regadío, 15 expertos creen que los rendimientos están tecnológicamente tan "controlados" que no se presentan rendimientos que puedan ser considerados correspondientes a años "muy malos", y 8 incluso no admiten los intervalos extremos "años muy malos" y "años muy buenos". En los cultivos de secano este efecto no es tan dramático, pero en 7 casos de 33 se producen también reducciones de intervalos.

Es decir, los expertos amputan uno o dos intervalos, posiblemente para (a) indicar que la influencia climatológica no es tan esencial (no existen "años" muy buenos o muy malos en ciertas condiciones) o, (b) que la estabilidad de los rendimientos es tan grande que no es necesario distinguir cinco intervalos.

Por ser utilizado como etiqueta, un término no pierde otros significados que tenga, además de los relativos a su función en el contexto de la definición de intervalos.

Δ

Pese a la reducción de intervalos, casi ningún experto ha expresado una f.d.p. uniforme para los rendimientos (solamente un caso en secano). Los expertos admiten la variabilidad de los rendimientos, y frecuencias distintas en los intervalos, y algunos de ellos han truncado los intervalos porque han dado un significado a las etiquetas con el que materialmente no estaban de acuerdo.

En los cultivos de regadío se considera que los intervalos extremos izquierdo y derecho acumulan menos del 15% de la frecuencia en 22 de 34 casos (64,7%), mientras que esta circunstancia se declara en 10 casos de 33 en cultivos de secano (30,3%). Por tanto, se supone que las funciones de densidad de los cultivos en regadío tienden a tener, en recorridos ordinales, un coeficiente de curtosis mayor que el de las correspondientes funciones para los cultivos de secano. La mayor rapidez de caída de la función en los cultivos de regadío sería visualmente "estadísticamente compatible" con el truncamiento de los intervalos extremos.

Igualmente, se puede observar en la tabla 22 que en los cultivos de secano se atribuye simetría a la f.d.p en 11 de 33 casos (33%), y por tanto mayoritariamente se señalan distribuciones no simétricas. Valores semejantes se observan en regadío.

La simetría de los 3 intervalos centrales se declara en 12 de 33 casos de secano (36,4 % de los casos) y en 19 de 34 casos de regadío (55,9 %). De lo anterior se puede concluir que:

Conclusión:

Aunque las f.d.p. de los cultivos en regadío tienden a ser vistas por los expertos como más simétricas que las correspondientes a los cultivos de secano, es necesario concluir que, al menos en relación con los intervalos definidos ordinalmente en la pregunta 6, la descripción de las f.d.p. tiende mayoritariamente a no admitir la simetría de las mismas.

Δ

A partir de las frecuencias asignadas a los distintos intervalos, en la pregunta 6 se ha deducido el número y posición de la moda (tabla 23).

Tabla 23.

Número y posición de las modas (intervalos ordinales). Encuesta a los expertos (1997-98).

Cultivos	**1 Moda**	**2 Modas**	**Moda en 2 intervalos**	**Moda en intervalo**				
				MM	M	N	B	MB
Secano	28	0	4	0	2	31	3	0
Regadío	32	0	2	0	0	33	3	0

En la mayoría de los cultivos de secano y de regadío se declararon funciones unimodales, con las modas mayoritariamente situadas en el intervalo central (“N” o de rendimientos “normales”).

Conjetura:

En el lenguaje natural el rendimiento “normal” se corresponde con el más frecuente o moda.

Δ

A fin de determinar la simetría en la distribución de las masas de frecuencia, se ha determinado la mediana referida a los 5 intervalos ordinales (Tabla 24; definición en ζ 4.3.2). La numeración de los intervalos va de 1 para el intervalo “muy malos” o “MM”, hasta 5 para el intervalo “muy buenos” o “MB”.

Tabla 24.

Valor estimado de la mediana (relativa a intervalos ordinales). Encuesta a los expertos (1997-98).

Cultivos	**Valor de la mediana (expresada por el número del intervalo)**									
	0-0.5	**0.5-1**	**1-1.5**	**1.5-2**	**2-2.5**	**2.5-3**	**3-3.5**	**3.5-4**	**4-4.5**	**4.5-5**
Secano				2	21	10				
Regadío					13	20	1			

Los cultivos de secano presentan 31 casos de 33 (94%) con la mediana en el intervalo central. En ese intervalo central, 11 están sobre el punto medio (valor 2,5), 10 a su izquierda y 10 a su derecha De esta forma, la mediana se encontraría a la izquierda del punto medio del recorrido en 12 casos de los 33, coincidiendo con ese punto en 11 ocasiones, y a su derecha en 10 casos. El valor de la mediana indica una tendencia mayoritaria a asignar masas de probabilidad centradas en los cultivos de secano.

En el caso de los cultivos de regadío, 33 casos de 34 presentan su mediana en el intervalo central. En ese intervalo, a la izquierda de su punto medio se encuentran 2 observaciones, otras 11 en el valor 2,5 y 20 a la derecha del punto medio. De esta forma, la mediana se encontraría a la izquierda del punto medio del recorrido en 2 casos de los 34, coincidiendo con ese punto en 11 ocasiones, y a la derecha del mismo en 21 casos. El valor de la mediana indica una tendencia mayoritaria a asignar masas de probabilidad relativamente centradas, aunque la parte derecha del intervalo central tiene un mayor peso en los cultivos de regadío.

#5.4.- FUNCIONES DE DENSIDAD DE RENDIMIENTOS CON INTERVALOS NUMÉRICOS

En la **pregunta 12** se pidió a los expertos que establecieran una correspondencia entre cada uno de los cinco intervalos ordinales (rendimientos "muy malos", "malos", etc.) y una medida numérica que expresara los rendimientos físicos (kg. / ha). Es decir, se trataba de decir qué intervalo de rendimientos físicos se considera compatible con el calificativo de cada clase definida ordinalmente.

Como se discutirá más adelante, en ζ 5.7, un gran número de respuestas de los expertos se expresaron de tal forma que no son de la Clase I (donde están perfectamente definidos cinco intervalos numéricos), en la denominación introducida en ζ 4.2.1. La principal consecuencia es que el número de respuestas que se puede utilizar directamente (tabla 59), sin grandes problemas de interpretación, se reduce a 10 casos de secano (de 43 disponibles) y 4 de regadío (de 35 disponibles)

En la tabla 25 se ha resumido la posición de la moda en las respuestas de Clase I dadas por los expertos, relacionadas con la **pregunta 6**, una vez hecha la traducción a escalas numéricas con ayuda de las respuestas de la **pregunta 12**. La posición de la mediana, para las mismas respuestas, se puede consultar en la tabla 26.

Tabla 25.

Posición de la moda (relativa a intervalos numéricos). Respuestas Clase I. Encuesta a los expertos (1997-98).*

Cultivos	**I**	**C**	**D**	**Un**	**Bi**	**Total casos**	**IM = Índice posición relativa de la moda**								
							≤ -2	**- 2 - -1**	**-1 - 0**	**0 - 1**	**1 - 2**	**2 - 3**	**3 - 4**	**4 - 5**	**>5**
Secano	4	1	4	1	0	10		1	4	4					
Regadío	1	1	2	0	0	4			2	2					

* Significado de las leyendas en ζ 4.3.3

En los cultivos de secano se observa que de los 10 casos disponibles, 9 de ellos presentan una única moda, en cuatro casos a la izquierda del punto medio del intervalo, en 1 caso centrada y en 4 casos a la derecha del punto medio del intervalo. Esas modas se encuentran relativamente centradas y, en conjunto, las modas se distribuyen de forma bastante simétrica con respecto al punto medio del recorrido. Las modas se encuentran muy cercanas al centro del recorrido, expresada esa distancia en términos de la amplitud del intervalo para el que se ha señalado la presencia de la moda.

En los cultivos de regadío se observa que, de los 4 casos disponibles, todos ellos presentan una única moda. Esa moda se encuentra en la mitad de los casos a la izquierda o centrada y en la otra mitad a la derecha. Dichas modas se encuentran cercanas al centro del recorrido, expresada esta distancia en términos de la amplitud del intervalo para el que se ha señalado la presencia de la moda.

En lo que respecta a la **mediana**, se ha calculado su posición relativa para cada caso, habiendo normalizado los recorridos de tal forma que todos tengan una longitud unidad. El resultado se ha resumido en la tabla 26, donde se indica el número de casos en que se presenta un valor de la mediana comprendido en cada intervalo.

Tabla 26.

Posición relativa de la mediana. Respuestas de la Clase I. Encuesta a los expertos (1997-98).

	Situación relativa en [0 – 1]									
Cultivos	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Secano			1		4	4				
Regadío					1	2	1			

En los cultivos de secano, de las 10 observaciones que contenían respuestas de la clase I, una se corresponde con una distribución uniforme, y 8 presentan sus modas en puntos próximos al centro del recorrido, pero ninguna respuesta sitúa la mediana en ese punto. De esta forma, en 5 casos la mediana de la función de densidad definida por el experto se encuentra a la izquierda del punto medio del recorrido (que en esta escala es el punto 0.5), y en todos los casos con valores mayores de 0.25. Los 4 casos restantes se encuentran comprendidos en el intervalo (0.5, 0.75]. En el intervalo [0.4, 0.6], por tanto, se encuentran 8 de los 9 casos, y de ellos 5 a la izquierda del punto medio y 4 a su derecha.

En los cultivos de regadío, en un caso la mediana se encuentra a la izquierda del punto medio del recorrido (punto 0.5), con un valor superior a 0.25. En ningún caso la mediana coincide con el punto medio del recorrido, y los 3 casos restantes se encuentran incluidos en el intervalo (0.5, 0.75]. En el intervalo [0.4, 0.6] se encuentran 3 de los 4 casos, y de ellos 1 a la izquierda del punto medio y dos a su derecha.

El anterior análisis es muy limitado, ya que se ha realizado con solamente 10 respuestas en secano (de 43 disponibles) y 4 de regadío (de 35 disponibles). Por ello en ζ 5.7, y en especial en ζ 5.7.4, mas adelante, se han estudiado las condiciones en las que se podría estimar la información que falta en las clases de respuestas II, III y IV para aumentar el número de casos de los que poder obtener conclusiones.

Consideramos la clase II como una clase de respuestas relativamente fiable, especialmente los casos en que se han podido establecer sus recorridos. Sin embargo, se acepta que la calidad de su información no es la misma que la de las respuestas de la Clase I, por lo que ambos casos han sido estudiados separadamente (tablas 25 y 27), antes de estudiarlos conjuntamente (tabla 28).

Tabla 27.

Posición relativa de la moda. Clase II. Encuesta a los expertos (1997-98).

Cultivos	**I**	**C**	**D**	**Un**	**Bi**	**Total**	**Posición relativa de la moda**								
							≤ -2	**-2- -1**	**-1- 0**	**0- 1**	**1- 2**	**2- 3**	**3- 4**	**4- 5**	**>5**
Secano	7	0	3	0	0	10		1	6	2	1				
Regadío	3	0	5	0	0	8			3	4				1	

Tabla 28.

Posición relativa de la moda. Clases I y II. Encuesta a los expertos (1997-98).

Cultivos	**I**	**C**	**D**	**Un**	**Bi**	**Total**	**Posición relativa de la moda**								
							≤ -2	**-2- -1**	**-1- 0**	**0- 1**	**1- 2**	**2- 3**	**3- 4**	**4- 5**	**>5**
Secano	11	1	7	1	0	20		2	10	6	1				
Regadío	4	1	7	0	0	12			5	6				1	

Las clases III y IV de respuestas exigen mayor número de supuestos para reconstruir su recorrido, por lo que su nivel de fiabilidad es ciertamente menor. No obstante, se ha realizado un análisis conjunto de todas las clases (tabla 29), una vez recuperados los recorridos de las clases distintas a la Clase I.

Tabla 29.

Posición relativa de la moda. Clases I, II, III y IV. Encuesta a los expertos (1997-98).

Cultivos	**I**	**C**	**D**	**Un**	**Bi**	**Total**	**Posición relativa de la moda**								
							≤ -2	**-2- -1**	**-1- 0**	**0- 1**	**1- 2**	**2- 3**	**3- 4**	**4- 5**	**>5**
Secano	18	3	11	1	0	33		7	14	10	1				
Regadío	9	4	14	0	0	27		1	12	12			1	1	

Los mismos análisis se han realizado para la estimación de la mediana, y los resultados se han resumido en las tablas 30, 31 y 32.

Las respuestas de la clase II que se han podido recuperar, en comparación con la clase I, en el caso de los cultivos de secano, tienden a tener las modas en mayor proporción a la izquierda del punto central, mientras que en el caso de regadío (excepto por un punto muy a la derecha) la situación de las modas es prácticamente idéntica a la de la Clase I.

En el conjunto de las clases I y II (recuperada), en los cultivos de secano, la moda tiende a situarse ligeramente a la izquierda en un 60% de los casos o ligeramente a la derecha (37%) de los casos, En un solo caso la moda está centrada. Para el caso de cultivos de regadío la situación se invierte, predominando las modas ligeramente a la derecha sobre las situadas ligeramente a la izquierda, en proporciones prácticamente contrarias a las descritas para el secano. En cualquier caso, las distribuciones tienden a estar centradas en torno a 1,5 veces la amplitud del intervalo central.

En el conjunto de todas las encuestas que se han podido recuperar, de todas las clases, la situación es semejante a la descrita anteriormente. En los cultivos de secano, en un 56% de los casos la moda se presenta a la izquierda, y en un 34% a la derecha (con sólo un 10% de casos centrada). En los cultivos de regadío los porcentajes prácticamente se intercambian: las modas a la izquierda son el 33% de los casos, y a la derecha el 52% de los casos (y solamente un 15% de modas centradas). Pero, mientras que los cultivos de regadío tienden a concentrar sus modas en torno a 1,5 veces la amplitud del intervalo central, las de los cultivos de secano tienden a desbordar esta cota hacia la izquierda.

Tabla 30.

Posición relativa de la mediana. Clase II. Encuesta a los expertos (1997-98).

Cultivos	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Secano			1	1	5	3				
Regadío					1	4	2	1		

Tabla 31.

Posición relativa de la mediana. Clases I y II. Encuesta a los expertos (1997-98).

Cultivos	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Secano			2	1	9	7				
Regadío					2	6	3	1		

Tabla 32.

Posición relativa de la mediana. Clases I, II, III y IV. Encuesta a los expertos (1997-98).

Cultivos	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Secano		1	3	3	13	10	2			
Regadío				2	7	12	4	2		

En el caso de la mediana, con las respuestas de la Clase I, en el caso de los cultivos de secano, la mayor parte de los casos (89%) la sitúan en el intervalo relativo [0,4 – 0,6], es decir en el 1/5 del recorrido central. Este mismo patrón se repite conforme se amplía el número de encuestas con las recuperadas de las clases II, III y IV, pero apareciendo algunos casos a la izquierda de ese intervalo central (22% de los casos) y a la derecha (6%). Predominan por tanto las medianas centradas, en un radio 1/10 del recorrido (72%).

En los cultivos de regadío está más definida la tendencia de la mediana a localizarse en el centro. En el conjunto de los datos, en un 73% de los casos la mediana se encuentra centrada, en un radio 1/10 del recorrido. Contrariamente al caso de cultivos de secano, existe un mayor número de casos a la derecha del mismo (22% a la derecha y 7% a la izquierda).

Conclusión:

No existen grandes diferencias entre los resultados obtenidos con las diferentes clases, después de un proceso de recuperación de cuestionarios en las clases II, III y IV, mediante estimación de sus recorridos. Este resultado indica que las técnicas descritas de recuperación de respuestas truncadas en sus recorridos posiblemente no introduzcan alteraciones excesivas en relación a la información del resto de las respuestas.

Una segunda conclusión tiene que ver con la situación de la moda en los cultivos, en opinión de los expertos. Existe una gran coincidencia de predominio de las modas a la izquierda (56%) para los cultivos de secano, aunque sin descartar las modas a la derecha (34%). Proporciones prácticamente inversas se producen en el caso de los cultivos de regadío, donde predominan las modas a la derecha (52%), pero sin descartar las modas a la izquierda (33%). Si se suman los casos de secano y de regadío, prácticamente se obtiene una distribución simétrica.

En el caso de la mediana, los resultados coinciden en señalar una tendencia a estar situada en una región próxima al centro del recorrido, en un radio de 1/10 la amplitud del mismo. En el caso de secano, las modas a la izquierda de esa bola predominan sobre las situadas a la derecha, y esa tendencia se invierte en el caso del regadío. La suma de todos los casos tiende a dar una distribución simétrica.

Δ

#5.5.- ESTIMACIÓN DIRECTA DE LAS MODAS Y DE LA FORMA DE LA F.D.P.

En la **pregunta 7** (donde se obtuvieron 22 respuestas de 38 posibles) se interrogó directamente al experto sobre el número de modas y su situación. Se pedía también la opinión de los expertos respecto a la forma de la función de densidad, indicando cuántas modas presenta, si es simétrica o no, y en caso de no serlo, indicar si la moda (o modas) se encuentra (n) a la izquierda o a la derecha del punto medio del recorrido. Pueden consultarse en la Tabla 33 los resultados de esta pregunta.

Tabla 33.

Simetría y forma de la función de densidad. Encuesta a los expertos (1997-98).

Cultivos	**Distribución**				**Número de modas**			
	Simétrica	**Moda a la izquierda**	**Moda a la derecha**	**TOTAL**	**1**	**2**	**3**	**TOTAL**
Secano	13 (48%)	7 (26%)	7 (26%)	27	21 (75%)	5 (18%)	2 (7%)	28
Regadío	16 (57%)	2 (7%)	10 (36%)	28	26 (84%)	3 (10%)	2 (6%)	31

Puede observarse un predominio de la distribución simétrica sobre cualquier otra, tanto en cultivos de secano como en regadío, así como de la distribución unimodal. Con relación a la asimetría, no está muy clara en secano, puesto que se atribuye el mismo número de respuestas a la moda a la izquierda que a la moda a la derecha. En cambio, en regadío, predomina la asimetría hacia la derecha sobre la asimetría hacia la izquierda.

En base a lo anterior, podemos concluir:

Conjetura:

Los expertos tienden a tener una visión conjunta de los rendimientos distribuidos en torno a una moda central, y cuando se admite una desviación de ese valor, en los cultivos de secano se consideran igualmente probables las modas a la derecha y a la izquierda del punto medio del recorrido; mientras que en los cultivos de regadío mayoritariamente se perciben las modas a la derecha del punto medio.

Δ

En la **pregunta 8** (35 respuestas de 38 posibles) se intentó establecer para un mismo cultivo si: (a) los rendimientos mínimos son inferiores en secano que en regadío; (b) los rendimientos máximos en secano son inferiores a los máximos en regadío; y (c) el recorrido de rendimientos en secano es inferior al caso de regadío.

En general, los expertos consideran ciertas las tres afirmaciones (tabla 34).

Tabla 34.

Número de respuestas. Encuesta a los expertos (1997-98).

Respuesta	**min Rs < min Rr**	**max Rs < max Rr**	**(max Rs – min Rs) < (max Rr – min Rr)**
V	33	32	20
F	3	4	11
Depende / No Responde	3	4	11

NOTA: (Rs = rendimiento en secano; Rr = rendimiento en regadío).

Que el recorrido de los cultivos de regadío es mayor que el recorrido de los cultivos de secano contradice una de las hipótesis que hemos considerado y que los expertos asumían cuando efectuaban reducciones de intervalos.

Conjetura:

No se puede suponer que los recorridos de los cultivos en secano y regadío sigan, en su relación, una regla evidente. (El recorrido de los rendimientos de los cultivos en secano se percibe mayoritariamente como menor que el correspondiente a los cultivos de regadío, pero en bastantes casos, 1 de cada 3, se discute esa percepción).

En cualquier caso, la medida significativa no sería en términos absolutos (tantos kg. / ha de recorrido), sino posiblemente una variación porcentual.

Δ

En la **pregunta 9**, (a) se interrogaba directamente a los expertos sobre la forma de la función de densidad (sugiriéndose binomial, normal, gamma, beta, etc.) en los cultivos de secano y en los de regadío; y (b) se les solicitaba información bibliográfica sobre el particular. De los 38 expertos, 19 no contestaron esta pregunta, 5 señalaron que no sabían la contestación o que no disponían de datos, y 14 respondieron con alguna respuesta en cultivos de secano o de regadío (las dos opciones de la pregunta). Las respuestas obtenidas se pueden consultar en la tabla 35.

Tabla 35.

Forma funcional de la distribución de probabilidades de los rendimientos de los cultivos. Encuesta a los expertos (1997-98).

Función de densidad de probabilidad	**Cultivos de Secano**	**Cultivos de Regadío**
Normal	12	11
Binomial	1	1
No responde	25	26

La practica totalidad de los expertos que señalan una determinada función de densidad, lo hacen refiriéndose a la normal. A este grupo de expertos que declaran que la función de densidad de rendimientos es normal pertenecen 3 que no contestaron la pregunta número 6 (la frecuencia de cada uno de los cinco intervalos definidos), con lo que comparativamente con la discusión anterior, habría que considerar los resultados de la tabla 36.

Tabla 36.

Características de los cultivos señalados por los expertos que atribuyeron distribuciones normales. Encuesta a los expertos (1997-98).

Cultivos	**Número**	**Simétricas**	**I+A+D**	**$I_1 = I_5$**	**$I_2 = I_4$**
Secano	9	7	4	5	1
Regadío	8	4	5	1	4

De los 33 expertos que contestaron la pregunta 6, 11 señalaron posteriormente funciones normales para 17 cultivos (9 de secano y 8 de regadío). En esos cultivos, asignaron frecuencias simétricas en 7 cultivos de secano y 4 de regadío. No existe, por tanto, una coherencia muy elevada entre los expertos que señalan distribuciones normales y las distribuciones de frecuencia que asignan.

En base a los comentarios anteriores, parece razonable suponer que:

Conjetura:

Los expertos más habituados a trabajar en contextos estadísticos son los que tienden a definir f.d.p. Aunque no existe un consenso sobre la simetría de la función de densidad, bastantes elementos tienden a reforzar esa imagen (moda centrada, medianas no muy alejadas de los puntos medios de los intervalos), y posiblemente exista un sesgo en la formación estadística de los expertos que les hace tender a identificar los rendimientos de los cultivos como normalmente distribuidos.

Δ

En la **pregunta 10** (donde se obtuvieron 34 respuestas de 38 posibles) se pedía a los expertos su opinión sobre los factores que determinan que un año sea bueno o malo, desde el punto de vista de los rendimientos físicos de los cultivos, así como los factores que determinan la variabilidad de los rendimientos en secano y en regadío. En general, las respuestas a las tres preguntas tenían un denominador común, que era la climatología y las enfermedades.

#5.6.- APROXIMACIÓN A FUNCIONES TRIANGULARES Y BETAS

En la **pregunta 11** (donde se obtuvieron 30 respuestas de las 38 posibles) se pedía una estimación del tipo rendimiento (mínimo, más frecuente, máximo) o (R_{min}, R_{mf}, R_{max}), y del rendimiento medio, para estudiar su relación a través de las aproximaciones funcionales simplificadas normalmente propuestas.

Una primera cuestión que se examina es la posición de la moda en el intervalo definido. A partir de las respuestas dadas por los expertos, se ha calculado la posición de la moda normalizando todos los recorridos, de tal forma que se expresen en [0, 1]. La posición de la moda se ha determinado mediante la fórmula $(R_{mf} – R_{min}) : (R_{max} – R_{min})$. Cuando el valor de la expresión anterior es mayor que 0,5, eso indica que el segmento $(R_{mf} - R_{min})$ es mayor que el segmento $(R_{max} - R_{mf})$, y por tanto la moda se sitúa a la derecha del punto medio del intervalo. En caso de que el cociente sea inferior a 0,5, entonces la moda se sitúa a la izquierda del punto medio del intervalo. Los resultados obtenidos se han resumido en la Tabla 37.

Tabla 37.

Posición de la moda en la estimación directa de recorrido y moda. Número de casos en cada intervalo. Encuesta a los expertos (1997-98).

	Intervalo del valor $(R_{mf} – R_{min}) / (R_{max} - R_{min})$				
Cultivos	**0-0.2**	**0.2-0.4**	**0.4-0.6**	**0.6-0.8**	**0.8-1**
Secano	1 (2%)	20 (42%)	24 (50%)	3 (6%)	0
Regadío	1 (2%)	17 (39%)	23 (52%)	3 (7%)	0

Se puede observar que, en general, las distribuciones esquematizadas mediante tres puntos para los diferentes cultivos de secano y de regadío son bastantes semejantes, tendiendo la moda a estar situada a la izquierda del punto medio del intervalo.

En la tabla 38 se indican las respuestas con moda igual a 0,5 y el porcentaje que representan sobre el total.

Tabla 38.

Respuestas con moda igual a 0,5 y porcentaje respecto al total. Encuesta a los expertos (1997-98).

Cultivos	**Respuestas con moda = 0,5**	**Total respuestas**	**Porcentaje (%)**
Secano	10	48	21
Regadío	12	44	27

No se ha observado que exista una correlación entre el valor del recorrido y la posición de la moda, ni en secano ni en regadío. La distribución de la nube de puntos suele parecerse a la reproducida en la figura 1.

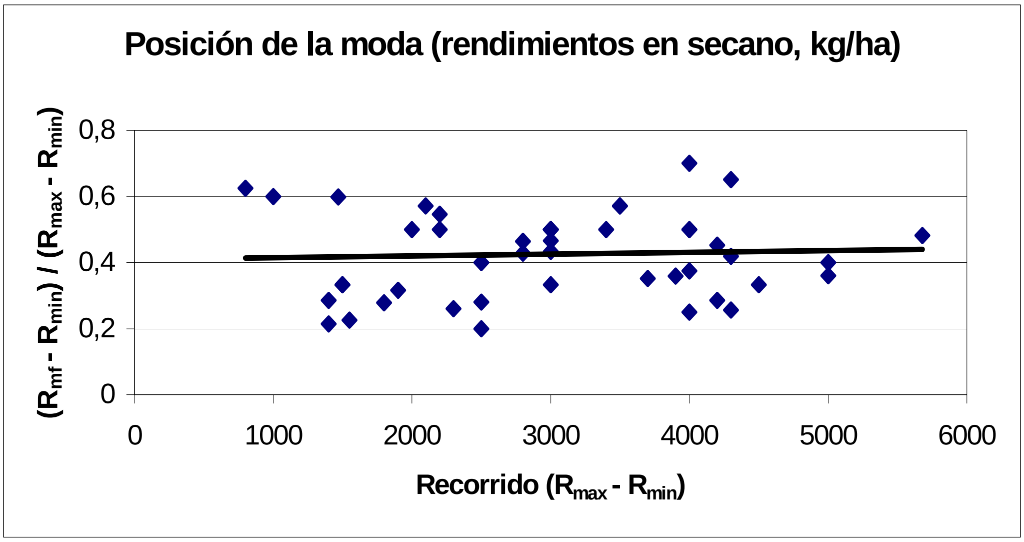


Figura 1.

Posición de la moda en función del recorrido. Encuesta a los expertos (1997 - 98).

Una segunda cuestión importante es la relación existente entre las diferentes estimaciones, en función de la forma funcional implícitamente supuesta en las estimaciones. Para ello inicialmente se han comparado los valores medios estimados por los expertos y los calculados a partir de la estimación directa de los rendimientos mínimo (R_{min}), máximo (R_{max}) y más frecuente (R_{mf}), según la fórmula del valor medio para las distribuciones de densidad triangular RMT= (1/3) (R_{min} + R_{mf} + R_{max}) y beta, simplificada según las hipótesis aceptadas en el método PERT, RMB = (1/6) (R_{min} + 4 R_{mf} + R_{max}).

Las diferencias entre esos valores y el valor medio directamente estimado por los expertos, $R_{m,}$ normalizada dividiendo por el valor medio estimado, se han representado en la tabla 39. Este valor expresa el nivel de error entre esas dos estimaciones, expresado en tanto por uno.

Tabla 39.

Error relativo de las estimaciones puntuales, (RMT – R_m): R_m o (RMB – R_m): R_m. Encuesta a los expertos (1997-98).

TRIANGULAR	**Número de casos**											
Cultivos	**0 - 0.05**	**0.05 - 0.10**	**0.10 - 0.15**	**0.15 - 0.20**	**0.20 - 0.25**	**0.25 - 0.30**	**0.30 - 0.35**	**0.35 - 0.40**	**0.40 - 0.45**	**0.45 - 0.50**	**> 0.50**	**Mediana**
Secano	18	12	6	6	1	2	1					0.071
Regadío	18	13	4	4	1	1					2	0.063
BETA	**Número de casos**											
Cultivos	**0 - 0.05**	**0.05 - 0.10**	**0.10 - 0.15**	**0.15 - 0.20**	**0.20 - 0.25**	**0.25 - 0.30**	**0.30 - 0.35**	**0.35 - 0.40**	**0.40 - 0.45**	**0.45 - 0.50**	**> 0.50**	**Mediana**
Secano	20	12	9	2	2		1					0.062
Regadío	22	8	6	3	1	1			1		1	0.049

Los valores de la tabla 39 indican que los expertos que han estimado tres valores “clásicos” de las funciones de densidad (el valor máximo, el mínimo y el más frecuente o moda), si hubieran utilizado las estimaciones analíticas de los valores medios a partir de esos tres valores, el error que hubieran encontrado con sus estimaciones directas de la media hubiera sido inferior al 15% en 41 casos de 46 para el caso de cultivos de secano y estimación beta – PERT.

Los mismos datos indican, para todos los cultivos y las distribuciones triangular y beta - PERT, unas diferencias relativas de los valores medios calculados y estimados cercanas a cero, ligeramente mejores para la comparación con la distribución beta - PERT. En general, existen diferencias menores del 15% en más del 80% de los casos.

La buena correspondencia entre los valores medios calculados a partir de las estimaciones de los valores mínimos, más frecuentes y máximos, y los estimados como medias directamente por los expertos, hacen muy atractiva la consideración de la aproximación beta – PERT en la estimación de medias. Dado que esta recomendación puede estar en el origen de importantes problemas de estimación, se realizó a una comparación entre las poblaciones de valores medios estimados y calculados a partir de las distribuciones triangular y beta, obteniéndose los resultados de la tabla 40.

Tabla 40.

Relaciones entre valores calculados RMT y RMB y estimados R_m de las medias. RMT con R_m y RMB con R_m. Encuesta a los expertos (1997-98).

Comparación con la distribución triangular			
Cultivos	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Secano	0.964	No	Diferencias
Regadío	0.973	No	No diferencias
Comparación con la distribución beta - PERT			
Cultivos	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Secano	0.971	No	No diferencias
Regadío	0.969	No	No diferencias

También se ha realizado una regresión entre los valores de rendimientos medios (R_m) estimados y los calculados a partir de distribuciones triangular (*RMT*) y beta – PERT (*RMB*), obteniéndose los resultados esquematizados en las tablas 41 a 44, que indican unos ajustes aceptables.

Tabla 41.

Regresión $R_m = a\ RMT + b$. Encuesta a los expertos (1997-98).

	CULTIVOS DE SECANO	**CULTIVOS DE REGADÍO**
a	0.96	0.92
b	44.2	608.9
R^2	0.99	0.98
Fc	16027.4	1738.6
Ft	4.06	4.07

Tabla 42.

Regresión $R_m = a\ RMB + b$. Encuesta a los expertos (1997-98).

	CULTIVOS DE SECANO	CULTIVOS DE REGADÍO
a	0.93	0.95
b	209.6	530.0
R^2	0.99	0.98
Fc	18784.2	2001.2
Ft	4.06	4.07

Tabla 43.

Regresión $R_m = a\ RMT$ (b = 0). Encuesta a los expertos (1997-98).

	CULTIVOS DE SECANO	CULTIVOS DE REGADÍO
a	0.97	0.94
b	0	0
R^2	0.99	0.99
Fc	21662.31	3731.42
Ft	4.06	4.07

Tabla 44.

Regresión $R_m = a\ RMB$ (b = 0). Encuesta a los expertos (1997-98).

	CULTIVOS DE SECANO	CULTIVOS DE REGADÍO
a	0.94	0.96
b	0	0
R^2	0.99	0.99
Fc	18814.82	4296.79
Ft	4.06	4.07

Los resultados obtenidos en las regresiones anteriores parecen indicar que se pueden obtener valores muy similares tanto con la aproximación triangular como con la beta-PERT, mejores en aquellas aproximaciones lineales con término b = 0. Con esta restricción, no existen diferencias significativas entre los errores cometidos con cultivos de secano y de regadío.

Para identificar los pesos que los ajustes podrían introducir, se realizó a continuación una regresión lineal múltiple entre los rendimientos medios (R_m) y los rendimientos mínimos (R_{min}), máximos (R_{max}) y más frecuentes (R_{mf}). La regresión tenía por ecuación: $R_m = a_1\ R_{min} + a_2\ R_{max} + a_3\ R_{mf} + b$, para los casos en que b fuera igual o diferente de cero. Los resultados se especifican en las Tablas 45 y 46.

Tabla 45.

Regresión $R_m = a_1 R_{min} + a_2 R_{max} + a_3 R_{mf} + b$. Encuesta a los expertos (1997-98).

	Cultivos secano	Cultivos regadío
b	217.16	660.64
a_1	0.12	0.36
a_2	0.16	0.07
a_3	0.63	0.63
R^2	0.99	0.98
Fc	5991.84*	677.99*
Ft	4.07	4.09
t_1	1.22	1.88*
t_2	2.31*	0.84
t_3	5.51*	3.27*
t_t	1.68	1.68

donde a_j (j = 1,..., 3) son los coeficientes que acompañan a las variables independientes R_{min}, R_{max} y R_{mf}, t_j (j = 1,..., 3) los valores de t de Student para cada variable independiente, F_t y t_t los valores críticos obtenidos en las respectivas tablas de F y t, y R^2 el coeficiente de determinación. El asterisco indica que la variable es significativa a un nivel $\alpha = 0.05$.

Se obtiene en secano una regresión significativa en conjunto, aunque el rendimiento mínimo no es significativo individualmente. En regadío, también se obtiene una regresión significativa en conjunto, aunque con el rendimiento máximo no es significativo individualmente.

Tabla 46.

Regresión $R_m = a_1 R_{min} + a_2 R_{max} + a_3 R_{mf}$ ($b = 0$). Encuesta a los expertos (1997-98).

	Cultivos secano	Cultivos regadío
b	0	0
a_1	0.27	0.34
a_2	0.28	0.07
a_3	0.41	0.65
R^2	0.99	0.98
Fc	*	*
Ft	4.06	4.08
t_1	3.03*	1.79*
t_2	5.24*	0.89
t_3	5.05*	3.42*
t_t	1.68	1.68

En este caso, se obtuvo una regresión significativa en conjunto para ambos tipos de cultivos, con los tres coeficientes significativos para los cultivos de secano y el rendimiento máximo no significativo para los cultivos de regadío. La relación entre los coeficientes a_i parece estar más próxima a una función beta – PERT que a una triangular.

Lo anterior parece sostener la siguiente:

Conjetura:

Existe una relación muy estrecha entre los valores de la media estimada y la calculada a partir de las tres estimaciones (máximo, normal, mínimo) realizados por los expertos para los rendimientos de los cultivos, siendo algo mejor la aproximación beta – PERT que la triangular.

Δ

#5.6.1.- COHERENCIA DE LAS ESTIMACIONES

A partir de las estimaciones de la frecuencia de cada intervalo numérico y de las estimaciones puntuales de los valores (R_{max}, R_{mf}, R_{min}) es posible calcular los valores de las medias y de las varianzas respectivas, compararlas entre sí y con los valores de los rendimientos medios R_m estimados directamente por los expertos (puntualmente). Llamando x_i al valor de la marca de clase del intervalo i en la estimación de frecuencias, y f_i la frecuencia correspondiente, en lo que sigue se comparan los siguientes valores medios:

$R_{xf} = (\Sigma\ x_i\ f_i)$ con los correspondientes valores medios obtenidos a partir de las estimaciones de (R_{max}, R_{mf}, R_{min});

$R_{xf} = (\Sigma\ x_i\ f_i)$ con la estimación directa puntual del valor medio, R_m.

Una vez estudiadas las diferencias entre las estimaciones directas e indirectas de los valores medios, se realizará el correspondiente estudio para el caso de la varianza, mediante la comparación de

$V_{xf} =(x_i – R_{xf})^2 f_i$ con los valores de las varianzas obtenidos de las estimaciones (R_{max}, R_{mf}, R_{min}).

#5.6.1.1.- COHERENCIA DE LOS VALORES MEDIOS

En la tabla 47 se comparan las diferencias de la media calculada en base a las frecuencias atribuidas a cada intervalo $R_{xf} = (\Sigma\ x_i\ f_i)$, y las medias estimadas a partir de (R_{max}, R_{mf}, R_{min}) mediante las aproximaciones triangular y beta – PERT, obteniéndose un mejor resultado en regadío para ambas comparaciones, y para la aproximación beta – PERT.

Tabla 47.

Diferencias relativas de medias calculadas $|(R_{xf}$ - RMT)$:R_{xf}|$ y $|(R_{xf}$ -RMB)$:R_{xf}|$. Encuesta a los expertos (1997-98).

APROXIMAC.												
TRIANGULAR	**Número de casos**											
Cultivos	**0-**	**0.05-**	**0.10-**	**0.15-**	**0.20-**	**0.25-**	**0.30-**	**0.35-**	**0.40-**	**0.45-**	**>0.5**	**Mediana**
Secano	7	2	3	5	1		1	1		1	5	0.160
Regadío	9	7	4	1	1			2				0.071
BETA	**Número de casos**											
Cultivos	**0-**	**0.05-**	**0.10-**	**0.15-**	**0.20-**	**0.25-**	**0.30-**	**0.35-**	**0.40-**	**0.45-**	**>0.5**	**Mediana**
Secano	5	5	4	3	2	2					5	0.138
Regadío	12	5	3	1	2			1				0.050

De las 26 estimaciones realizadas para los cultivos de secano, 12 presentan diferencias en sus valores medios R_{xf} y RMT (estimación triangular) inferiores al 15%, y el 50% de los casos presentan diferencias inferiores o iguales al 16%. No obstante 8 casos presentan diferencias muy elevadas. En el caso de la estimación beta – PERT, 14 casos presentan diferencias inferiores al 15%, con el 50% de los casos con diferencias inferiores al 5%. Cinco estimaciones son francamente malas.

De las 24 estimaciones realizadas para los cultivos de regadío, en el caso de la aproximación triangular, 20 casos presentan diferencias inferiores al 15%, y el 50% de los casos presentan diferencias inferiores al 7,1%. Con la aproximación beta-PERT estos resultados son aún mejores: 20 casos con diferencias inferiores al 15%, solamente dos casos con diferencias superiores al 20%, y el 50% de los casos con diferencias inferiores al 5%.

Se puede decir que para los cultivos de regadío las respuestas sobre atribución de frecuencias y traducción de intervalos son bastante consecuentes con las estimaciones puntuales de (R_{max}, R_{mf}, R_{min}). En el caso del secano, posiblemente se puedan mejorar los resultados mediante mejoras en los sistemas de obtención de la información. El examen de las poblaciones de valores medios estimados por ambos métodos, resumido en la tabla 48, confirma que los resultados en caso de regadío se pueden considerar bastante razonables, mientras que en el caso del secano se tienen dificultades.

Tabla 48.

Comparación entre las poblaciones de rendimientos medios (RMT con R_{xf} y RMB con R_{xf}). Encuesta a los expertos (1997-98).

Comparación con la distribución triangular			
Cultivos	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Secano	0.716	No	Diferencias
Regadío	0.987	No	Diferencias
Comparación con la distribución beta			
Cultivos	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Secano	0.727	No	Diferencias
Regadío	0.987	No	No diferencias

En la tabla 49 se resume el resultado de las regresiones realizadas entre las diferentes estimaciones. En el caso de los cultivos de regadío los resultados son extraordinariamente buenos, sugiriendo un buen nivel de coherencia. En el caso del secano los resultados no pueden considerarse desfavorables, sugiriendo la posibilidad de desarrollar métodos para mejorar la coherencia (o de explicaciones de si estos resultados son compatibles con algún concepto riguroso de coherencia). El mismo análisis para el caso b = 0, resumido en la tabla 50, parece confirmar claramente esta conjetura. La restricción b = 0 mejora sensiblemente el coeficiente "a" de la regresión, señalando en el secano diferencias entre el 7 – 11%, y prácticamente nulas para el regadío.

Tabla 49.

Regresión $RMT = a\, R_{xf} + b$ y $RMB = a\, R_{xf} + b$. Encuesta a los expertos (1997-98).

DISTRIBUCIÓN TRIANGULAR		
	Cultivos Secano	**Cultivos Regadío**
a	0.70	1.01
b	1158.09	731.99
R^2	0.46	0.98
Fc	20.0	1411.3
Ft	4.26	4.30
DISTRIBUCIÓN BETA		
	Cultivos Secano	**Cultivos Regadío**
a	0.68	0.99
b	1126.59	682.87
R^2	0.42	0.99
Fc	17.2	1686.9
Ft	4.26	4.30

Tabla 50.

Regresión *RMT = a RMF* (*b* = 0) y *RMB = a RMF* (*b* = 0). Encuesta a los expertos (1997-98).

DISTRIBUCIÓN TRIANGULAR		
	Cultivos Secano	**Cultivos Regadío**
a	1.11	1.03
b	0	0
R^2	0.90	0.99
Fc	213.8	3345.7
Ft	4.24	4.28
DISTRIBUCIÓN BETA		
	Cultivos Secano	**Cultivos Regadío**
a	1.07	1.01
b	0	0
R^2	0.88	0.99
Fc	190.2	3986.2
Ft	4.24	4.28

Las comparaciones que siguen (tabla 51) se realizan entre el valor de los rendimientos medios estimados directamente por los expertos (puntualmente), R_m, y los calculados a partir de sus estimaciones de frecuencias y valores numéricos de los correspondientes intervalos.

Tabla 51.

COCIENTE $|(R_{xf}-R_m):R_{xf}|$. Encuesta a los expertos (1997-98).

	Número de casos											
Cultivos	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**>0.5**	**Mediana**
Secano	3	2	6	3	3		2	3			4	0.183
Regadío	7	4	5		3	2	2	1				0.110

Como en la discusión anterior, los resultados son mejores para los cultivos de regadío. De los 26 casos de secano examinados, 11 señalan diferencias inferiores al 15%, y 14 casos presentan diferencias inferiores al 20%. Existe, sin embargo, un grupo de 9 encuestas con diferencias superiores al 30%. En el caso del regadío, de los 24 casos, 16 presentan diferencias inferiores al 15%, y el resto de los casos se encuentra en errores en torno al 30%.

La tabla 52 señala la semejanza entre las dos poblaciones de valores medios, mientras que en las tablas 53 y 54 se reflejan los resultados de realizar una regresión entre ambas estimaciones. De nuevo, para el caso b = 0, los resultados obtenidos en ambas regresiones son buenos, aunque, como antes, las estimaciones en regadío se muestran mejores que en secano.

Tabla 52.

Comparación entre las poblaciones de rendimientos medios (R_{xf} con R_m). Encuesta a los expertos (1997 –98).

Comparación			
Cultivos	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Secano	0.697	No	No diferencias
Regadío	0.963	No	No diferencias

Tabla 53.

Regresión $R_m = a\, R_{xf} + b$. Encuesta a los expertos (1997-98).

$R_m = a\, R_{xf} + b$		
	Cultivos Secano	**Cultivos Regadío**
a	0.60	0.92
b	1366.74	1414.67
R^2	0.29	0.97
Fc	9.76	618.0
Ft	4.26	4.30

Tabla 54.

Regresión $R_m = a\, R_{xf}$ ($b = 0$). Encuesta a los expertos (1997-98).

$R_m = a\, R_{xf}$		
	Cultivos Secano	**Cultivos Regadío**
a	1.08	0.95
b	0	0
R^2	0.85	0.99
Fc	137.4	1478.9
Ft	4.24	4.28

#5.6.1.2.- COHERENCIA DE LAS VARIANZAS

En la tabla 55 se realiza la comparación de las varianzas calculadas a partir de la distribución de frecuencias $V_{xf} = (\Sigma\, (x_i - R_{xf})^2\, f_i)$, y las derivadas de las aproximaciones a funciones triangulares y beta-PERT, obteniéndose un mejor resultado en regadío para ambas comparaciones, aunque en este caso mejor para la aproximación triangular que para la beta - PERT. En general, los errores medidos mediante la mediana de las diferencias son superiores en el caso de las varianzas que en el anteriormente discutido de las medias.

Tabla 55.

Diferencias relativas de las varianzas calculadas $|(V_{xf}\text{-VT}):V_{xf}|$ y $|(V_{xf}\text{-VB}):V_{xf}|$. Encuesta a los expertos (1997 –98).

APROXIMAC.												
TRIANGULAR	**Número de casos**											
Cultivos	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1.0**	**>1.0**	**Mediana**
Secano	3	3	2	7	1	5	1			2	2	0.371
Regadío	6	4	2	3	2	2		1			4	0.300
BETA	**Número de casos**											
Cultivos	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1.0**	**>1.0**	**Mediana**
Secano		2	2	4	3	6	5	3			1	0.633
Regadío	1		4	3	5	2	2	4	1		2	0.480

En el caso de los cultivos de secano, de los 26 casos examinados, en 6 casos la diferencia es inferior al 20% entre V_{xf} y VT. El 50% de los casos presenta diferencias inferiores al 37,1%. Con la aproximación beta – PERT la situación es peor: el 50% de los casos presentan diferencias superiores 63,3%. No se puede decir que exista una aproximación razonable a las varianzas, y esta cuestión plantea la necesidad de buscar una explicación sobre los factores que producen este fenómeno.

En la tabla 56 se resume el estudio de las poblaciones de varianzas obtenidas por los diferentes métodos, obteniéndose mejores resultados para la aproximación triangular en los cultivos de regadío. Esto sugiere que, pese a las importantes diferencias absolutas de la tabla 55, posiblemente las poblaciones no sean tan dispares.

Tabla 56.

Comparación entre las poblaciones de varianzas (VT con V_{xf} y VB con V_{xf}). Encuesta a los expertos (1997-98).

Comparación con la distribución triangular			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Secano	0.905	No	Diferencias
Regadío	0.906	No	No diferencias
Comparación con la distribución beta			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Secano	0.905	No	Diferencias
Regadío	0.903	No	Diferencias

Las tablas 57 y 58 muestran los resultados de realizar la regresión de las diferentes estimaciones. En general se obtienen mejores resultados en regadío, pero existen diferencias demasiado elevadas para poder conjeturar directamente la coherencia de las estimaciones. Sin embargo, tampoco, a la vista de los datos disponibles, se puede concluir sobre la incoherencia de las estimaciones.

Tabla 57.

Regresión $VT = a\ V_{xf} + b$ y $VB = a\ V_{xf} + b$. Encuesta a los expertos (1997-98).

DISTRIBUCIÓN TRIANGULAR		
	Cultivos Secano	**Cultivos Regadío**
a	0.46	0.58
b	147157.69	12875721.17
R^2	0.73	0.56
Fc	64.2	28.0
Ft	4.26	4.30
DISTRIBUCIÓN BETA		
	Cultivos Secano	**Cultivos Regadío**
a	0.29	0.38
b	94103.10	8215917.24
R^2	0.72	0.57
Fc	60.8	29.6
Ft	4.26	4.30

Tabla 58.

Regresión $VT = a\ V_{xf}$ (b = 0) y $VB = a\ V_{xf}$ (b = 0). Encuesta a los expertos (1997 – 98).

DISTRIBUCIÓN TRIANGULAR		
	Cultivos Secano	**Cultivos Regadío**
a	0.57	0.68
b	0	0
R^2	0.88	0.71
Fc	189.8	57.4
Ft	4.24	4.28
DISTRIBUCIÓN BETA		
	Cultivos Secano	**Cultivos Regadío**
a	0.37	0.45
b	0	0
R^2	0.88	0.72
Fc	181.9	60.1
Ft	4.24	4.28

Conjetura:

Posiblemente las formas asociadas al histograma de frecuencia y a las aproximaciones funcionales generen una diferencia entre las varianzas calculadas para cada una de ellas, que podríamos denominar "de base", y que expliquen un porcentaje importante de las diferencias observadas. Este valor introducido por la mera diferencia funcional disfrazaría las verdaderas diferencias de varianzas estimadas por cada método. Así, podría ocurrir que $VT - V_{xf} = s^2_{base} + s^2_0$, donde s^2_{base} fuera un artefacto introducido por la diferencia funcional y s^2_0 la diferencia entre las varianzas implícitamente determinadas por cada estimación.

Δ

#5.7.- PROBLEMAS EN LA INDAGACIÓN SOBRE INTERVALOS NUMÉRICOS

En la **pregunta 12** (donde se obtuvieron 28 respuestas de 38 posibles) se pidió a los expertos que establecieran una correspondencia entre el lenguaje natural referido a rendimientos (rendimiento muy malo, malo, normal, bueno y muy bueno) y una escala numérica que expresara los valores que entendían que se correspondían con cada uno de los intervalos ordinales.

Se pensaba que esta traducción sería mecánica y precisa (intervalos iguales, contiguos, etc.). Sin embargo, a la hora de examinar las respuestas se observó una casuística muy heterogénea, que obligó a realizar una clasificación de las respuestas en clases. La metodología asociada con este tipo de problemas se expuso en ζ 4.2.1. El número de respuestas obtenidas de cada clase se indica en la tabla 59 y en la 60 cuántas respuestas presentaban discontinuidades (respuestas con discontinuidad / respuestas totales).

Tabla 59.

Clases de respuestas obtenidas. Encuesta a los expertos (1997-98).

Cultivos	**Clase I**	**Clase II**			**Clase III**	**Clase IV**			**Total**
		II	**IIa**	**IIb**		**IV**	**IVa**	**IVb**	
Secano	10	7	0	6	11	3	4	2	43
Regadío	4	5	0	3	15	4	4	0	35

Tabla 60.

Número de respuestas que presentan discontinuidades / número de respuestas totales. Encuesta a los expertos (1997-98).

Cultivos	**Índice de Continuidad de Intervalos**		
	Clase I	**Clase II**	**Total**
Secano	2/10	3/13	5/23
Regadío	0/4	1/8	1/12

Inicialmente, con los casos de respuestas de la Clase I, se han realizado diversas regresiones lineales, para estudiar la relación entre el recorrido largo R_L y el recorrido corto, R_C. (tablas 61 y 62), que indican en general un buen ajuste, sobre todo cuando b = 0. Desgraciadamente el número de datos de la clase I es tan reducido (tabla 59, sobre todo para el caso de los cultivos de regadío), que no se puede obtener una aproximación segura entre ambos términos.

Tabla 61.

Regresión $R_L = a\ R_C + b$. Clase de respuestas I. Encuesta a los expertos (1997-98).

	b ≠ 0		**b = 0**	
	C. de secano	**C. de regadío**	**C. de secano**	**C. de regadío**
a	1.46	1.34	1.65	1.44
b	488.7	340.0	0	0
Fc	46.33	1104.9	1048.97	965.22
Ft	5.32	18.51	5.12	10.13
R^2	0.85	0.99	0.99	0.99

Tabla 62.

Regresión $(R_L – R_C) = a\ R_C + b$. Clase de respuesta I. Encuesta a los expertos (1997-98).

	b ≠ 0		**b = 0**	
	C. de secano	**C. de regadío**	**C. de secano**	**C. de regadío**
a	0.46	0.34	0.65	0.44
b	488.7	340.0	0	0
Fc	4.60	70.44	162.33	88.67
Ft	5.32	18.51	5.12	10.13
R^2	0.36	0.97	0.95	0.97

Los cocientes relativos y su distribución se pueden consultar en las tablas 63 y 64.

Tabla 63.

Cociente R_C / R_L. Clase de respuesta I. Encuesta a los expertos (1997-98).

	Número de casos									
Cultivos	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Secano				1	1	4	3	1		
Regadío						2	1	1		

Tabla 64.

Cociente $(R_L - R_C) / R_L$. Clase de respuesta I. Encuesta a los expertos (1997-98).

	Número de casos									
Cultivos	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Secano			1	4	3	2				
Regadío			1	2	1					

Con los resultados anteriores, se puede establecer la

Conjetura:

En el caso de los expertos, las respuestas obtenidas de la clase I (completa definición y contigüidad de los límites de los intervalos) sugieren que, siempre que se puedan definir "recorridos cortos" en una distribución de intervalos, estos recorridos cortos se pueden utilizar como estimadores del recorrido de la distribución.

Δ

#5.7.1.- COHERENCIA DE LOS RECORRIDOS

En la pregunta **número 11** los expertos definieron el recorrido de los rendimientos de los diferentes cultivos, mediante la identificación de dos extremos, el mínimo (R_{min}) y el máximo (R_{max}). En la pregunta **número 12**, los expertos asignaron intervalos numéricos a las cinco clases expresadas en lenguaje natural. En el caso de las respuestas de la Clase I existe una definición precisa de los valores extremos, por lo que es posible comparar el recorrido (largo) de la pregunta 12 con el definido en la pregunta 11. Se han comparado esas dos medidas del recorrido $R_L = a\ (R_{max} - R_{min}) + b$. En general, se obtiene un ajuste razonable para los cultivos de secano y mejor para los de regadío.

Todo parece indicar que la amplitud del recorrido es de un mismo orden cuando se define directamente o a través de intervalos expresados en lenguaje natural, especialmente en los cultivos de regadío, aunque esta afirmación se debe matizar por las pocas observaciones de la Clase I disponibles para los cultivos de regadío.

Tabla 65.

Regresión $R_L = a\ (R_{max} - R_{min}) + b$. Clase de respuesta I. Encuesta a los expertos (1997-98).

	b ≠ 0	
	C. de secano	**C. de regadío**
a	1.29	0.96
b	317.3	302.0
Fc	14.95	321.0
Ft	5.32	18.51
R^2	0.65	0.99

También se ha calculado la recta de regresión del recorrido corto de la pregunta 12 para la clase de respuestas I y los resultados de la pregunta número 11, mediante la ecuación $R_C = a\ (R_{max} - R_{min}) + b$.

Tabla 66.

Regresión $R_C = a\ (R_{max} - R_{min}) + b$. Clase de respuesta I. Encuesta a los expertos (1997-98).

	CULTIVOS DE SECANO	CULTIVOS DE REGADÍO
a	0.81	0.72
b	110.4	-31.0
Fc	16.42	940.31
Ft	5.32	18.51
R^2	0.67	0.99

Obteniéndose también un ajuste razonable para los cultivos de secano y aún mejor para los de regadío.

#5.7.2.- AMPLITUD DE LOS INTERVALOS

En general, los expertos definen de la misma amplitud los intervalos numéricos de los rendimientos, y no se ha identificado un patrón (por ejemplo, no siempre los intervalos extremos son más amplios, ni tampoco lo es el intervalo de los rendimientos normales). Para las respuestas de la clase I, se han normalizado los datos sobre la amplitud de los intervalos de cada encuesta, dando el valor 1 al intervalo más amplio. El resto de los intervalos se han expresado como una fracción del más amplio, y, por ejemplo, cuando un intervalo toma el valor 0,8 significa que la amplitud de ese intervalo es 0,8 veces la del más amplio en la respuesta a la pregunta 12 en esa encuesta.

Tabla 67.

Número de casos en que el intervalo ( ] toma el valor. Encuesta a los expertos (1997-98).

	Mayor	**0-0.2**	**0.2-0.4**	**0.4-0.6**	**0.6-0.8**	**0.8-1**	**Mediana**	**Media**
Cultivos de secano								
Muy Malo	2/10	1	1	2	4	2	0,73	0,66
Malo	4/10	1	1	3	1	4	0,6	0,67
Normal	4/10	0	0	0	4	6	0,85	0,85
Bueno	3/10	0	0	3	3	4	0,67	0,74
Muy Bueno	6/10	0	0	1	2	7	1	0,87
Cultivos de regadío								
Muy Malo	2/4	0	0	2	0	2	0,75	0,75
Malo	2/4	0	0	2	0	2	0,8	0,78
Normal	2/4	0	0	1	1	2	0.9	0,82
Bueno	2/4	0	0	2	0	2	0,8	0,78
Muy Bueno	1/4	0	0	3	0	1	0,55	0,65

Tanto para los cultivos de secano como para los de regadío, se observa bastante uniformidad respecto a las amplitudes de los intervalos, sin que pueda destacarse ningún intervalo como más amplio sistemáticamente que los demás.

Los intervalos máximo y mínimo y se expresaron como porcentaje del recorrido (largo) (Tabla 68) y como porcentaje del punto medio del intervalo de rendimientos *"normales"* (Tabla 69), para las respuestas de la clase I.

Tabla 68.

Amplitud del intervalo máximo y del mínimo en las diferentes encuestas, expresada como porcentaje del recorrido (largo). Clase de respuesta I. Encuesta a los expertos (1997-98).

		Número de casos									
Cultivos	**Amplitud**	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Secano	Máxima		1	5	3	1					
	Mínima	1	8	1							
Regadío	Máxima		1	2	1						
	Mínima		4								

Tabla 69.

Amplitud del intervalo máximo y del mínimo en las diferentes encuestas, expresada como porcentaje del punto medio del intervalo de rendimientos *"normales"*. Clase de respuesta I. Encuesta a los expertos (1997-98).

		Número de casos									
Cultivos	**Amplitud**	**0 - 0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Secano	Máxima				2	4	1	2	1		
	Mínima	1	2	1	5	1					
Regadío	Máxima		2		1	1					
	Mínima	1	2	1							

Aunque no parece observarse una pauta regular, posiblemente se pueda avanzar la

Conjetura:

Posiblemente la amplitud del intervalo central sirva de anclaje para las amplitudes de los demás intervalos, cuando estos no tienen la misma amplitud. Igualmente, la amplitud de los intervalos puede ser un indicador de la precisión con que se desean indicar los rendimientos.

Δ

#5.7.3.- ANÁLISIS DE LOS EXTREMOS

Es posible comparar los extremos del recorrido en las respuestas obtenidas en las preguntas 11 y 12, en este último caso para respuestas de la Clase I. Se ha comparado el rendimiento mínimo R_{min} de la pregunta 11 con el rendimiento extremo inferior del recorrido, $R_{<MM}$ (extremo inferior del intervalo de rendimientos "muy malos") , de la pregunta 12, y el rendimiento máximo R_{max} con el rendimiento extremo superior del recorrido, $R_{>MB}$ (extremo superior del intervalo de rendimientos "muy buenos").

Inicialmente, se han calculado el coeficiente de correlación de Spearman, el test de independencia (significación $\alpha = 0.05$) y el test de Wilcoxon (significación $\alpha = 0.05$), a fin de determinar si cada par de poblaciones están relacionadas y si se puede aceptar la hipótesis de que sus valores medios pertenecen a la misma distribución (Tabla 70).

Tabla 70.

Comparación entre poblaciones de rendimientos máximos y mínimos. Encuesta a los expertos (1997-98).

Cultivos	**Comparación**	**Spearman (r_S)**	**Muestras independientes**	**Wilcoxon (α=0.05)**
Secano	R_{min} con $R_{<MM}$	0.133	Sí	Existen diferencias entre las 2 poblaciones
	R_{max} con $R_{>MB}$	0.945	No	No existen diferencias
Regadío	R_{min} con $R_{<MM}$	1	*	*
	R_{max} con $R_{>MB}$	1	*	*

* No se realizaron dichas comparaciones por tener sólo 4 observaciones

Se observa que los cultivos de secano tienen una semejanza mayor entre las poblaciones de rendimientos máximos que entre las de rendimientos mínimos, para los tres casos.

Se efectuó también la regresión entre el *rendimiento mínimo* y el *rendimiento muy malo* y el *rendimiento máximo* y el *rendimiento muy bueno*, para los casos en que la constante fuera igual o diferente de cero (tabla 71).

Tabla 71.

Regresión entre rendimientos máximos y mínimos. Clase de respuesta I. Encuesta a los expertos (1997-98).

Cultivos	**Comparación**	**y = a x + b, b ≠ 0**	**y = ax, b = 0**
Secano	R_{min} con $R_{<MM}$	$R_{min} = -0.21\ R_{<MM} + 946.3$ $R^2 = 0.03$ Fc = 0.24, Ft = 5.32	$R_{min} = 1.00\ R_{<MM}$ $R^2 = 0.17$ Fc = 1.88, Ft = 5.12
	R_{max} con $R_{>MB}$	$R_{max} = 0.89\ R_{>MB} + 96.2$ $R^2 = 0.97$ Fc = 263.5, Ft = 5.32	$R_{max} = 0.92\ R_{>MB}$ $R^2 = 0.99$ Fc = 2527.65, Ft = 5.12
Regadío	R_{min} con $R_{<MM}$	$R_{min} = 0.99\ R_{<MM} + 70.8$ $R^2 = 0.99$ Fc = 5935.5, Ft = 18.51	$R_{min} = 1.00\ R_{<MM}$ $R^2 = 1.00$ Fc = 9373.34, Ft = 10.13
	R_{max} con $R_{>MB}$	$R_{max} = 1.01\ R_{>MB} – 174.6$ $R^2 = 0.99$ Fc = 1240.7, Ft = 18.51	$R_{max} = 0.99\ R_{>MB}$ $R^2 = 0.99$ Fc = 3402.79, Ft = 10.13

La regresión es significativa en todos los casos, excepto en los extremos inferiores de los cultivos de secano.

Diferencias en los extremos inferiores del recorrido. Se han examinado también las diferencias entre los valores de R_{min} (pregunta 11) y $R_{<MM}$, expresadas como porcentaje del valor R_{min}, para normalizar. La distribución del número de casos, según la cuantía de la diferencia, se reproduce en la tabla 72.

Tabla 72.

Cociente $|(R_{<MM} - R_{min})/R_{min}|$. Clase de respuesta I. Encuesta a los expertos (1997-98).

	Número de casos									
Cultivos	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Secano	3									7
Regadío	3					1				

En los cultivos de secano, en 3 casos se obtuvieron diferencias nulas ($R_{min} = R_{<MM}$), y en los 7 restantes se observaron diferencias del 100%, pudiéndose considerar "outliers", ya que se corresponden con respuestas $R_{<MM} = 0$, que obviamente es una forma de hablar más que de describir un recorrido. Es importante subrayar que en todos los casos $R_{min} \geq R_{<MM}$.

En los cultivos de regadío, en tres casos se obtuvieron diferencias nulas($R_{min} = R_{<MM}$), y en un caso del 60%. Es importante destacar que en todos los casos $R_{min} \geq R_{<MM}$.

Conclusión:

Al traducir del lenguaje natural a un conjunto de intervalos numéricos los sujetos pueden expresar como límite inferior del recorrido valores "razonables", como 0, pero nunca observados o sin significación estadística. Estos datos constituyen verdaderos "outliers" y se debe establecer algún sistema para evitarlos, tal como las precauciones que se adoptan en las estimaciones PERT.

Es decir, si se supone que el sujeto considera resultados muy malos aquellos que sean inferiores a 1000 kg. / ha, la clase de rendimientos "muy malos" debería definirse como el mínimo valor observado y esa cota, es decir, como $[r_{min}, 1000)$. Pero es obvio que si r_{min} es un resultado "muy malo" también lo es cualquier valor inferior, en el significado lingüístico del término. Esto puede ampliar hacia la izquierda los intervalos de la f.d.p. estimada.

Δ

Diferencias en los extremos superiores del recorrido. Se ha estudiado el número de casos según la cuantía de las diferencias entre los valores de R_{max} (pregunta 11) y $R_{>MB}$ (pregunta 12), expresadas como porcentaje del valor R_{max}. Los resultados se reproducen en la tabla 73.

Tabla 73.

Cociente $|(R_{>MB} - R_{max}) / R_{max}|$. Clase de respuesta I. Encuesta a los expertos (1997-98).

	Número de casos									
Cultivos	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Secano	5	4	1							
Regadío	3	1								

En todos los cultivos $R_{>MB} \geq R_{max}$. En los cultivos de secano se obtuvieron cuatro (de 10 respuestas) coincidencias ($R_{>MB} = R_{max}$) y en regadío tres (de cuatro).

Conclusión:

También en la clase de rendimientos "muy buenos" es posible que los sujetos encuestados den límites superiores no observados, en función del significado lingüístico del término.

Δ

Como resumen de la discusión anterior, se resumen en la tabla 74 los casos obtenidos.

Tabla 74.

Valores de los extremos inferiores R_{min}, R_{mm} y superiores R_{max}, R_{MB}. Clase de respuesta I. Encuesta a los expertos (1997-98).

Cultivos	$\mathbf{R_{mm} = 0}$	$\mathbf{R_{min} = R_{<MM}}$ $\mathbf{R_{max} = R_{>MB}}$	$\mathbf{R_{min} = R_{<MM}}$ $\mathbf{R_{max} > R_{>MB}}$	$\mathbf{R_{min} = R_{<MM}}$ $\mathbf{R_{max} < R_{>MB}}$	$\mathbf{R_{max} = R_{>MB}}$ $\mathbf{R_{min} > R_{<MM}}$	$\mathbf{R_{max} = R_{>MB}}$ $\mathbf{R_{min} < R_{<MM}}$	$\mathbf{R_{max} < R_{>MB}}$ $\mathbf{R_{min} > R_{<MM}}$	$\mathbf{R_{max} > R_{>MB}}$ $\mathbf{R_{min} < R_{<MM}}$	**Total**
Secano	7	3	0	0	0	0	0	0	10
Regadío	0	2	0	1	1	0	0	0	4

Conclusión:

Se observa para los dos tipos de cultivos una clara tendencia a incrementar el recorrido cuando está expresado en términos de lenguaje natural. Es decir, la amplitud definida al traducir las escalas ordinales es superior a la definida por las estimaciones puntuales R_{max} y R_{min}.

Δ

#5.7.4.- RECUPERACIÓN DE CASOS

Las respuestas de la clase II (13 de secano y 8 de regadío, en la tabla 59) en principio no deberían de presentar grandes problemas de recuperación, dado que solamente les falta uno o ambos de los extremos inferior o superior del recorrido. Las técnicas de recuperación se han discutido en ζ 4.2.1.4.

En esta clase, por tanto, se puede estimar el recorrido (a) mediante el recorrido corto R_C; (b) a partir de los valores definidos en la pregunta 11 (R_{min}, R_{mf}, R_{max}). De las diferentes estimaciones, se adopta la media de aquellas que resulten coherentes con el resto de los valores disponibles, o la mejor con arreglo a ese criterio. Una vez estimado el recorrido total o largo, se puede conocer la suma de las amplitudes de los dos intervalos extremos, y adoptar la convención de atribuir la misma longitud a ambos en las clases II. Naturalmente, esta estimación de los recorridos presenta un mayor nivel de incertidumbre que el de las Clases I, por lo que inicialmente se han separado los análisis.

Se estudió de forma individual los casos en que eran "recuperables" las respuestas de la Clase II, obteniéndose los resultados resumidos en la tabla 75, donde el "número de casos" se da de la forma **x(y)**, donde **x** es el número de cultivos e **(y)** es el número de encuestas en que se describen esos casos.

Tabla 75.

Estimación de recorridos de las respuestas de la Clase II. Encuesta a los expertos (1997-98).

SECANO					
		CRITERIO			
Número casos	**CLASE**	$\mathbf{R_L = R_C}$	$\mathbf{R_{min} = R_{<MM}}$	$\mathbf{R_{max} = R_{>MB}}$	$\mathbf{R_L = (R_{max} - R_{min})}$
8 (5)	II	BIEN	MAL	BIEN	MAL
2 (1)	IIb	MAL	MAL	BIEN	MAL
3 (1)	IIb	BIEN	MAL	BIEN	MAL

REGADÍO					
2 (1)	II	BIEN	BIEN	BIEN	BIEN
2 (2)	II	BIEN	MAL	BIEN	BIEN
1 (1)	II	BIEN	MAL	BIEN	MAL
3 (1)	IIb	BIEN	MAL	BIEN	MAL

Con las convenciones sobre estimación de recorridos, se han podido recuperar 7 respuestas en secano (de 13 posibles) y 8 en regadío (la totalidad).

En las clases de respuestas III y IV se ha procedido de forma similar.

Obsérvese que la mala "traducción" del extremo inferior izquierdo ("rendimientos muy malos"), donde conjeturamos que se han expresado "valores muy malos" en lugar de "valores muy malos observados", ha provocado que no se haya podido utilizar de forma eficaz la información contenida en las estimaciones de R_{min}.

#5.7.5.- ANÁLISIS DE LOS RENDIMIENTOS NORMALES Y MÁS FRECUENTES

Diferencias en el centro del recorrido. Se examinó en cuántas ocasiones los rendimientos más frecuentes (R_{mf}) de la pregunta 11, se encontraban en el intervalo de rendimientos "normales" de la pregunta 12, para las clases de respuestas I, II, IIa y IIb (tabla 76).

Tabla 76.

Respuestas R_{mf} perteneciente al intervalo normal R_{NN}. Clases de respuesta I, II, IIa y IIb. Encuesta a los expertos (1997-98).

Cultivos	**Respuestas pertenecientes al intervalo normal**	**Total respuestas**	**Porcentaje (%)**
Secano	21	22	95
Regadío	11	12	92

Para las respuestas de las clases III, IV, IVa y IVb, se midió la diferencia entre los rendimientos normal (R_{NN}) y más frecuente (R_{mf}), expresada como porcentaje del rendimiento más frecuente, obteniendo los resultados de la Tabla 77.

Tabla 77.

Cociente $|(R_{NN} - R_{mf}) / R_{mf}|$. Clases de respuesta III, IV, IVa y IVb. Encuesta a los expertos (1997-98).

	Número de casos									
Cultivos	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Secano	7	8		1		1	3			
Regadío	19	3								

Observándose unas diferencias relativamente pequeñas entre los rendimientos considerados, aunque algo mayores en los cultivos de secano.

En la Tabla 78 se expresan las coincidencias entre los rendimientos normal y más frecuente para cada cultivo y para las clases de respuesta III, IV, IVa y IVb.

Tabla 78.

Número de casos en que $R_n = R_{mf}$. Clases de respuesta III, IV, IVa y IVb.

Cultivos	**$R_{NN} = R_{mf}$**	**Casos totales**	**Porcentaje (%)**
Secano	6	22	27
Regadío	15	20	75

Observándose unas mayores coincidencias en los cultivos de regadío.

Conclusión:

En general se observa una relación estrecha entre la estimación de los rendimientos más frecuentes y el intervalo de los rendimientos "normales".
Δ

#5.8.- COHERENCIA DE LOS EXPERTOS

Se han comparado las respuestas de los expertos a las diferentes preguntas, comprobando si se producen diferencias.

Preguntas 1 y 3.- *Comparación entre receta fija de utilización de insumos y rango de variación de los mismos.*

De las 38 encuestas, han respondido a ambas preguntas 26 expertos (no han respondido los expertos 2, 7, 9, 12, 14, 17, 19, 20, 26, 29, 34 y 35), de los cuales 6 no son coherentes entre ambas respuestas (se trata de los expertos números 4, 8, 16, 26, 32 y 36).

Preguntas 6 y 7.- *Comparación entre porcentaje de años muy malos, malos, normales, buenos y muy buenos y forma de la función de densidad de los rendimientos (número de modas y simetría)*

En este caso, 5 de los 38 expertos no han respondido a la pregunta 6; 16 de ellos no han respondido a la pregunta 7; 4 de ellos no ha respondido ni a la pregunta 6 ni a la 7; 12 de ellos ha respondido a la pregunta 6 pero no a la 7; 1 de ellos ha respondido a la pregunta 7 y no a la 6.

En cuanto a la casuística de respuestas, 5 de ellos han respondido con valores de rendimientos en lugar de porcentajes a la pregunta 6; 1 de ellos ha respondido con fracciones en lugar de porcentajes a la pregunta 6; 3 de ellos han respondido sólo parcialmente a la pregunta 7, es decir, sólo indicando el número de modas que deben tener las distribuciones pero sin indicar la forma de las distribuciones de rendimientos; 5 de ellos han incluido los porcentajes de la pregunta 6 sin que el total sume el 100%, y finalmente 1 de ellos ha utilizado datos propios para responder a la pregunta 6.

Es decir, que basándonos en la falta de respuesta o en la incoherencia de la misma, 23 expertos han tenido dificultades para expresar las funciones de rendimientos de los cultivos. Sólo 6 de los 15 expertos han dado respuestas 'válidas', en el sentido de ajustarse a los supuestos implícitos en las normas del encuestador.

De los expertos que han respondido a ambas preguntas, se producen diferencias en 13 encuestas, consistiendo dichas diferencias en cambios en la forma de las funciones de densidad de los rendimientos, presentando asimetrías en una pregunta y simetrías en la otra, o bien presentando más de una moda en una pregunta y una sola moda en la otra, etc.

Preguntas 6 y 9.- *Comparación entre el porcentaje de años muy malos, malos, normales, buenos y muy buenos y el ajuste de la función de densidad a una variable estadística conocida*

De las 38 encuestas realizadas, han respondido a ambas preguntas 12 expertos (5, 18, 19, 20, 22, 23, 24, 30, 31, 32, 37 y 38), de los cuales 5 (5, 18, 24, 30 y 38) son coherentes en cultivos de secano y 3 (18, 24 y 37) en cultivos de regadío. El resto (19, 20, 22, 23, 31 y 32) son incoherentes en ambos tipos de cultivos.

Preguntas 7 y 9.- *Comparación entre la forma de la función de densidad de los rendimientos y ajuste a una variable estadística conocida*

Comparando las respuestas a dichas preguntas, de los 38 expertos encuestados 13 responden a la pregunta 9, 12 de los cuales asignan distribuciones normales a las funciones de densidad, 10 (encuestas 1, 3, 4, 11, 15, 16, 21, 25, 27 y 34) responden a la pregunta 7 y no a la 9. Por otro lado, sólo 9 de ellos (encuestas 5, 13, 19, 20, 23, 24, 31, 32 y 37) responden a ambas preguntas, de los cuales sólo 3 (20, 24 y 32) son coherentes.

Hay cinco expertos que responden a la pregunta 9 pero no a la 7 (se trata de las encuestas 18, 22, 26, 30 y 38). Todos ellos coinciden en sus respuestas, ajustando las funciones de densidad a variables normales. Entre ellos, 4 responden a la pregunta 6 excepto uno (encuesta 26).

Los 14 expertos restantes no responden a ninguna de dichas preguntas.

Preguntas 8 y 11.- *Coherencia en la estimación del recorrido de los rendimientos*

Dividimos dicho apartado en tres partes:

1.Coherencia en rendimiento mínimo en secano inferior a rendimiento mínimo en regadío (Mín Rs < Mín Rr)

Un total de 33 encuestas responden afirmativamente a esta pregunta, por lo menos para algunos cultivos. De ellas, sólo algunos expertos (9, de números 1, 5, 14, 15, 16, 19, 28, 36 y 37) expresan luego rendimientos para un mismo cultivo en secano y en regadío.

De nueve expertos, se produce una pequeña incoherencia en las encuestas 14 y 16 (22%).

2.Coherencia en rendimiento máximo en secano inferior a rendimiento máximo en regadío (Máx Rs < Máx Rr)

Un total de 32 encuestas responden afirmativamente a esta pregunta, por lo menos para algunos cultivos. De ellas, sólo 9 expertos expresan luego rendimientos para un mismo cultivo en secano y en regadío.

Para las 9 encuestas, se produce incoherencia en la encuesta 16 (11%).

3.Coherencia en recorrido en secano inferior a recorrido en regadío (Máx Rs - Mín Rs) < (Máx Rr - Mín Rr)

Un total de 20 encuestas responden afirmativamente a esta pregunta, por lo menos para algunos cultivos. De ellas, sólo 8 expertos expresan luego rendimientos para un mismo cultivo en secano y en regadío.

Dentro de las 8 encuestas, tres de ellas (15, 16 y 37) son incoherentes (38%).

Preguntas 11 y 12.- *Comparación entre rendimientos mínimo, máximo, más frecuente y medio e intervalo de rendimientos muy malo, malo, normal, bueno y muy bueno*

Tabla 79.

Incoherencias en las preguntas 11 y 12. Encuestas a expertos (1997 – 98).

	MOTIVO DE LA INCOHERENCIA			
ENCUESTA NÚMERO	**RENDIMIENTO MÍNIMO P11**	**INTERVALO MUY MALO P12**	**RENDIMIENTO MÁXIMO P11**	**INTERVALO MUY BUENO P12**
3	30000	20000	-	-
4	20000	10000-20000	-	-
4	5000	2000-5000	-	-
4	10000	2000-5000	-	-
6	800	0-1200	-	-
6	600	0-1000	-	-
10	800	0-1200	-	-
10	400	0-600	-	-
10	55000	<50000	-	-
12	2200	500	-	-
12	2000	500	-	-
12	1200	250	-	-
15	9000	7500	-	-
16	3000	2000	-	-
16	2800	1500	-	-
16	8000	6000	12000	14000
16	2000	1000	-	-
18	2000	1500	-	-
18	12000	<10000	-	-
24	12000	8000	22000	25000
24	10000	6000	18000	20000
24	8000	5000	18000	20000
25	2500	1500	-	-
25	2500	2200	-	-
25	450	<400	-	-
27	20000	<15000	-	-
28	20000	0	-	-
28	7000	0	-	-
31	820	<800	-	-
32	1500	800-1500	-	-
32	1000	500-1000	-	-
32	500	200-300	-	-
32	40000	<35000	-	-
33	1700	0-1000	-	-
33	1400	0-800	4200	3800-4800
33	1400	0-800	3500	3300-4000
33	600	0-400	-	-
33	500	0-400	2300	2000-2800
34	5000-6000	4000-5000	-	-
34	5000-6000	4000-5000	-	-
36	2000	1500	-	-
36	1800	1500	-	-
36	6000	5000	-	-
37	50000	<40000	-	-
37	-	-	90000	100000
38	250	100-350	-	-

#5.9.- PRINCIPALES CONCLUSIONES

De las encuestas realizadas a los expertos, se puede concluir de forma general que:

♦ Moda única y situada en el intervalo central mayoritariamente (en secano 31 de 33 casos o 94% y en regadío en 33 de 34 casos o 97%). Nunca se encuentran modas en los intervalos de rendimientos MM (muy malo) y MB (muy bueno), ni en secano ni en regadío, y en regadío, tampoco existe ningún caso con moda situada en el intervalo M (malo) de rendimientos.

♦ Reducción del intervalo: mayoritariamente reducción en regadío del intervalo inferior y en secano del superior.

♦ Mediana en secano centrada, con poca dispersión fuera del intervalo central. En regadío, mayor tendencia hacia la derecha del intervalo central.

♦ En comparación con la pregunta referente a los intervalos de rendimientos definidos mediante lenguaje natural, moda relativamente simétrica y poca dispersión respecto al centro del recorrido. En regadío, la mitad de las encuestas a la izquierda o en el centro y la otra mitad a la derecha.

♦ La mediana, está relativamente centrada en secano, pero ningún caso exactamente en el centro. En regadío, más bien se encuentra hacia la derecha del centro del recorrido.

♦ En secano, se observan principalmente funciones de densidad centradas, mientras que en regadío se observa asimetría ligeramente hacia la derecha.

♦ Regresión $R_L = a\ R_C + b$, significativa con b igual o distinta de cero.

♦ Regresión $(R_L - R_C) = a\ R_C + b$, no significativa en secano para b distinta de cero y significativa en regadío, con b igual o distinta de cero.

♦ Regresión $R_L = a\ (R_{max} - R_{min}) + b$, significativa en todos los casos, pero mejores ajustes en regadío.

♦ Regresión $R_C = a\ (R_{max} - R_{min}) + b$, significativa en todos los casos, pero mejores ajustes en regadío.

♦ No destaca ningún intervalo como más amplio que los demás.

♦ Poblaciones de extremos superiores (R_{max} y $R_{>MB}$) e inferiores (R_{min} y $R_{<\mathrm{MM}}$): en secano, mayor semejanza entre las poblaciones de rendimientos superiores. En regadío, sólo se calcula el coeficiente de Spearman, obteniéndose con valor 1, y no pudiendo comparar los demás indicadores por contar sólo con 4 observaciones para la clase I.

♦ Regresión $R_{min} = R_{<\mathrm{MM}} + b$, significativa en regadío y no significativa en secano, para b igual o distinta de cero.

♦ Regresión $R_{max} = R_{>MB} + b$, significativa en todos los casos, para b igual o distinta de cero.

♦ $R_{min} \geq R_{<\mathrm{MM}}$ en secano y regadío. Se producen diferencias nulas en un 30% en los cultivos de secano y en un 75% en los de regadío.

♦ $R_{max} \leq R_{>MB}$ en secano y regadío. Se producen diferencias nulas en un 40% en los cultivos de secano y en un 75% en los de regadío.

- Se dan mayores diferencias relativas en extremos inferiores $\left|\dfrac{R_{<MM} - R_{min}}{R_{min}}\right|$ que en superiores $\left|\dfrac{R_{>MB} - R_{max}}{R_{max}}\right|$.

- En la distribución de porcentajes en lenguaje natural, se observa una simetría del orden del 32-33% en ambos tipos de cultivos.

- Se observa un incremento de los recorridos cuando éstos vienen expresados mediante lenguaje natural.

- Altos porcentajes de pertenencia del rendimiento más frecuente de la estimación triangular de rendimientos al intervalo normal de rendimientos expresados mediante lenguaje natural (superiores al 92%).

- Diferencias relativas $\left|\dfrac{R_{NN} - R_{mf}}{R_{mf}}\right|$ pequeñas, pero algo mayores en secano.

- Coincidencia $R_{NN} = R_{mf}$ alta en regadío (75%) y más baja en secano (27%).

- Comparación rendimientos medios estimados por los expertos y rendimientos medios calculados a partir de las distribuciones triangular y beta - PERT: diferencias bajas y poblaciones mayoritariamente dependientes, con altos coeficientes de Spearman.

- Buenos ajustes en las regresiones $R_m = a\ RMT + b$ y $R_m = a\ RMB + b$, con b igual o distinta de cero (siendo R_m el rendimiento medio estimado, y RMT y RMB los rendimientos medios calculados a partir de las distribuciones triangular y beta, respectivamente.

- No se observa correlación entre el recorrido y la posición de la moda.

- Regresión $R_m = a_1\ R_{min} + a_2\ R_{max} + a_3\ R_{mf} + b$ significativa en conjunto para los cultivos de secano, aunque el rendimiento mínimo no es significativo individualmente. En regadío, también se obtiene una regresión significativa en conjunto, aunque con el rendimiento máximo no significativo individualmente. Para el caso de b nula, se obtuvo una regresión significativa en conjunto para ambos tipos de cultivos, con los tres coeficientes significativos para los cultivos de secano y el rendimiento máximo no significativo para los de regadío.

CAPÍTULO 6

ENTREVISTAS CON AGRICULTORES

#6.1.- ENTREVISTAS CON GRUPOS DE AGRICULTORES

Se intentó derivar las funciones de densidad de los cultivos anuales más conocidos por los agricultores mediante una entrevista personal, realizada a dos grupos de cinco agricultores cada uno, siguiendo un cuestionario relativamente abierto. Las características del cuestionario se pueden consultar en el Anejo 2.

Se entrevistó a diez agricultores de la comarca de Les Garrigues (Lleida), donde los cultivos anuales más frecuentes son la cebada en secano, y el trigo y el maíz en regadío.

El primer grupo fue entrevistado por la autora de la tesis durante el otoño - invierno de 1997, y el segundo grupo por un postgraduado contratado y entrenado para la realización de las correspondientes entrevistas, durante el invierno – primavera de 1998. Las entrevistas eran individuales, pero cada miembro del grupo sabía que se estaban realizando entrevistas a otros cuatro agricultores conocidos de él.

Cada encuestador realizó 6 entrevistas a cada agricultor de su grupo, en días distintos cada entrevista, con intervalos del orden de una semana entre cada par de entrevistas. Entre el primer grupo de entrevistas y el segundo grupo transcurrieron 3 meses.

El encuestador seguía un cuestionario, al que era sometido cada agricultor individualmente. La entrevista tenía por objeto obtener la información recogida en el cuestionario, pero también se prestaba especial atención a las informaciones que estaba en disposición de facilitar el sujeto, cuáles eran las circunstancias particulares que podían explicar su plan de producción y su sistema de toma de decisiones, qué cuestiones o giros entendía mejor, cuáles eran incomprensibles, qué convicciones mantenía, etc. Estas apreciaciones cualitativas debían servir a la hora de analizar los resultados y de planificar experimentos futuros.

La entrevista tendía a procurar un tipo de conversación que no resultara excesivamente fatigosa para el entrevistado. Debido a todas estas características, a esta encuesta se le ha denominado coloquialmente “encuesta en profundidad”.

Como puede apreciarse por las explicaciones anteriores, se pretendía con la realización de estas entrevistas conseguir una cierta familiaridad con las pautas de comprensión de su trabajo por parte de los agricultores y con la forma de expresar sus conocimientos.

La razón para realizar las entrevistas con dos encuestadores es que se sospechaba que, en una entrevista tan extensa en el tiempo, buscando la complicidad del agricultor, evitando su fatiga, etc., era posible que se produjeran influencias importantes de aquello que el encuestador creyera sobre las cuestiones discutidas, y que por tanto las respuestas en realidad estuvieran sugeridas en su contenido, marco conceptual, forma, etc. por el propio entrevistador. Como es conocido, se han descrito sesgos de complacencia que se han tratado de detectar mediante cambios de encuestador. (Se avanza la conjetura de que es posible que se hayan manifestado sesgos de esa naturaleza, ya que en general las respuestas obtenidas por el postgraduado mostraron un nivel de incoherencia mayor. No obstante no se puede deducir de esa circunstancia la conjetura comentada, ya que otras alternativas son igualmente compatibles con los resultados).

Se siguió el siguiente esquema de cuestiones:

PRIMER DÍA

1.1.- Operaciones realizadas en general para los tres cultivos (cebada de secano y trigo y maíz de regadío) durante el año.

1.2. - Rendimiento mínimo, máximo y medio para los tres cultivos estudiados.

1.3. - A partir del recorrido expresado en el punto 1.2, dividirlo en cinco partes iguales y pedir al sujeto entrevistado que distribuya diez garbanzos entre los intervalos de rendimientos así definidos. El número de garbanzos depositados sobre un intervalo representa la frecuencia relativa de los distintos rendimientos, según la experiencia del sujeto.

La cuestión 1.1 pretendía definir las peculiaridades de las técnicas de cultivo seguidas por el agricultor, en gran parte determinadas por cuestiones estructurales particulares de cada explotación, y de esta forma situar al encuestador en la racionalidad seguida por el sujeto.

Las cuestiones 1.2 y 1.3 tenían por objeto intentar la construcción de una primera aproximación a la función de densidad de probabilidad de rendimientos para los distintos cultivos, según la experiencia del sujeto, expresando las frecuencias de cada uno de los cinco intervalos mediante la asignación a cada uno de ellos de un número de garbanzos.

Este método había sido sugerido por el trabajo de W. Grisley y E. D. Kellogg (1983). Estos autores señalaban gráficamente cinco intervalos en un papel y le entregaban al agricultor veinticinco monedas. El sistema de incentivos o "scoring rules" en nuestro caso se pretendía construir mediante un clima de cooperación entre el entrevistado y el entrevistador, en cierta manera avalado por el conocimiento de que se estaban entrevistando sobre las mismas cuestiones a otras personas conocidas.

Se decidió utilizar garbanzos para indicar la frecuencia relativa de cada intervalo. Aunque W. Grisley y E. D. Kellogg (1983) entregaban a los agricultores 25 monedas, se consideró que en una primera aproximación interesaba que se marcasen claramente los pesos relativos de cada intervalo, por lo que se limitó a 10 el número de garbanzos en esta primera entrevista. Este número de piezas es suficiente para expresar una forma de la función de densidad del tipo 1:2:4:2:1 (o sus correspondientes variaciones, como 1:4:2:2:1, etc.).

SEGUNDO DÍA

En las respuestas del primer día, cada agricultor declaró para cada cultivo un recorrido determinado, que en general no tenía por qué coincidir con los declarados por el resto de los sujetos entrevistados. A partir de la información obtenida el primer día en cuanto a valores mínimos y máximos de rendimientos, se preparó un recorrido común que se facilitó el segundo día a los cinco agricultores del grupo. Este recorrido común para cada cultivo se determinó eligiendo como valor mínimo del mismo el menor valor de los rendimientos declarados por todos los agricultores el primer día, y como valor máximo, el mayor. Entonces:

2.1.- Se dividió el recorrido común de nuevo en cinco partes iguales y se pidió a cada agricultor que distribuyera los garbanzos representando la frecuencia de los distintos rendimientos, según su experiencia, pero esta vez sin límites en el número de garbanzos que necesitara para expresar el perfil de la función de densidad de probabilidad de los rendimientos.

2.2.- Se dividió este nuevo recorrido en dos mitades exactas, y a su vez cada mitad en cinco intervalos. Es decir, el recorrido total que se había plasmado gráficamente en una hoja, ahora se reproducía en dos hojas diferentes, cada una con una mitad del recorrido total. El agricultor debía asignar frecuencias a cada uno de los cinco intervalos de cada una de las dos mitades.

2.3.- Se pidió a cada agricultor que se pusiera en el lugar de otro agricultor del grupo, y que estimara los rendimientos de los cultivos del mismo, repitiendo los puntos 2.1 y 2.2.

La cuestión 2.1 intentaba crear un marco de referencia que (a) fuera común para los agricultores, y (b) planteara un contexto diferente al que el sujeto pudiera recordar del día anterior, de manera que tuviera que repetir los mismos valores en un nuevo marco que no le facilitara el anclaje en los antiguos valores, pretendiendo mantener una opinión “coherente”, pero posiblemente no sincera.

La cuestión 2.2. se planteó como un test de coherencia. Si los sujetos hubieran sido capaces de definir la f.d.p. mediante dos construcciones independientes y complementarias, de tal forma que hubiesen podido unir por sus fronteras para reconstruir la f.d.p. completa, hubiera sido un indicador de definición muy precisa de los valores declarados. Supongamos que se ha partido de un recorrido $R_{max} – R_{min}$, que inicialmente fué dividido en cinco intervalos, a los que el agricultor ha asignado las correspondientes frecuencias en (2.1). El punto medio de ese intervalo es $R_{1/2} = (R_{max} + R_{min}) / 2$.

El encuestador confecciona ahora dos hojas, la primera con un recorrido $R_{max} – R_{1/2}$, y la segunda con un recorrido $R_{1/2} – R_{min}$. Cada uno de estos semi - recorridos (los mismos para todos los agricultores) se dividía en cinco intervalos iguales.

Se presentaba al agricultor la primera hoja, con el recorrido $R_{1/2} – R_{min}$ y se le pedía que asignara la frecuencia de cada uno de los cinco intervalos señalados en la misma. Una vez finalizada esa asignación, se retiraba esta primera hoja, y se le sometía una segunda, en la que se tenía el resto del recorrido, $R_{max} – R_{1/2}$, divido en cinco intervalos. De nuevo se solicitaba al agricultor que indicara la frecuencia de cada intervalo.

La cuestión 2.3 pretendía explorar la opinión del agricultor sobre las formas de las funciones de densidad, en un contexto en que lo que ocurre no es de su responsabilidad (ya que estima los rendimientos de otro sujeto, concreto, miembro del grupo). Igualmente, puede sugerir, por comparación con sus propios valores, si existe alguna particularidad que le afecte a él especialmente.

Las cuestiones discutidas durante el tercer, cuarto, quinto y sexto día tenían que ver con la utilización de una “receta” de insumos fija (programación estática) o si variaba en función de las circunstancias (cuáles, programación dinámica). Esta línea de investigación se dejó abierta y no se examina en la presente Tesis.

#6.2.- ESTIMACIÓN DE FUNCIONES DE DENSIDAD

Inicialmente se discuten las formas de las f.d.p. que sugieren las distribuciones de frecuencias asignadas por los agricultores, tanto en la entrevista del primer día (cinco intervalos, con diez garbanzos a distribuir entre ellos), como el segundo día (cinco intervalos sin límite de garbanzos). Estos resultados se encuentran en las tablas 80 a 85 en las columnas encabezadas, respectivamente, como “Función día 1º” y “Función día 2º”.

Una representación gráfica de los resultados obtenidos para las funciones de cada agricultor el día 1 y el día 2 se puede consultar en el Anejo 3. En estas gráficas, las líneas continuas se refieren a los resultados de la entrevista del primer día y las líneas discontinuas a los resultados del segundo día. En general, la convención adoptada para identificar las curvas es del tipo x.y, que significa agricultor x, día y.

En el Anejo 4 se comparan, para cada día, las respuestas de los distintos agricultores.

Como se ha comentado en el apartado anterior, una segunda prueba de asignación de frecuencias era realizada sobre los dos semi - recorridos (izquierdo y derecho) en que se dividía posteriormente el recorrido total. Cuando se asignaban frecuencias a un semi – recorrido no se visualizaba el otro. Se considera que en el punto de unión de los semi - recorridos izquierdo y derecho, una buena asignación de frecuencias debería presentar la curva de frecuencias sin discontinuidad. Cuanto mayor es la discontinuidad ("salto"), se considera que existe peor estimación. Los resultados se han reflejado cualitativamente en la columna encabezada como "Salto entre los dos recorridos" en las tablas 80 a 85.

En general, los sujetos mostraron una gran dificultad para asignar probabilidades de esta forma (en dos semi - recorridos contiguos), como puede observarse de las gráficas reflejadas en el Anejo 5, donde se compara el perfil de la curva obtenida para el recorrido completo (línea continua), y las dos curvas obtenidas para cada uno de los semi - intervalos (líneas discontinuas).

Por último, en las tablas 80 a 85, la columna encabezada por "Coherencia" recoge la impresión de la doctoranda sobre la coincidencia de los valores dados por el agricultor en las distintas entrevistas de los días 1 y 2. Es un juicio subjetivo sobre coherencia de los agricultores con respecto a los valores de los intervalos y de las frecuencias dados en dos días diferentes.

En las Tablas 80 y 81 se indica, para la cebada de secano, el tipo de función de densidad de probabilidades estimada cualitativamente en los días primero y segundo, comprobando si se produce algún salto en el punto de unión de los dos intervalos (derecho e izquierdo) del recorrido, y si se produce coherencia entre las estimaciones.

Tabla 80.

Caracterización de las funciones de densidad estimadas por el primer grupo de agricultores para la cebada de secano. Encuesta en profundidad (1997-98).

AGRICULTOR	FUNCIÓN DÍA 1º	FUNCIÓN DÍA 2º	SALTO ENTRE LOS DOS RECORRIDOS	COHERENCIA
1	BETA CON MODA LIGERAMENTE A LA IZQUIERDA	BETA CON MODA A LA IZQUIERDA	NULO	COHERENTE
2	BETA CON MODA A LA IZQUIERDA	BETA CON MODA A LA IZQUIERDA	IMPORTANTE	COHERENTE
3	BETA CON MODA A LA IZQUIERDA	BETA CON MODA A LA IZQUIERDA	NULO	COHERENTE
4	BIMODAL HACIA LA IZQUIERDA	BETA CON MODA A LA IZQUIERDA	IMPORTANTE	COHERENTE
5	BETA CON MODA A LA IZQUIERDA	BETA CON MODA A LA IZQUIERDA	PEQUEÑO	COHERENTE

Los perfiles señalados por los agricultores en ambos días sugieren una forma similar a una función beta con la moda a la izquierda.

Tabla 81.

Caracterización de las funciones de densidad estimadas por el segundo grupo de agricultores para la cebada de secano. Encuesta en profundidad (1997-98).

AGRICULTOR	FUNCIÓN DÍA 1º	FUNCIÓN DÍA 2º	SALTO ENTRE LOS DOS INTERVALOS	COHERENCIA
6	BETA CON MODA A LA DERECHA	BETA CON MODA A LA DERECHA	PEQUEÑO	MUY COHERENTE
7	BIMODAL SUAVE HACIA LA DERECHA	BETA CON MODA A LA DERECHA	IMPORTANTE	COHERENTE
8	BETA CON MODA LIGERAMENTE A LA DERECHA	BETA CON MODA LIGERAMENTE A LA IZQUIERDA	NULO	COHERENTE
9	BETA CON MODA A LA DERECHA	BETA CON MODA A LA DERECHA	PEQUEÑO	COHERENTE
10	BETA CON MODA A LA DERECHA	BIMODAL SUAVE HACIA LA DERECHA	IMPORTANTE	COHERENTE

En este caso los agricultores parecen sugerir funciones de tipo beta, pero con la moda situada a la derecha. También se indica una función bimodal (hacia la derecha).

El cultivo del trigo se da en la zona sobre todo en las parcelas de regadío. En las Tablas 82 y 83 se indica la función de densidad estimada para este cultivo.

Tabla 82.

Caracterización de las funciones de densidad estimadas por el primer grupo de agricultores para el trigo de regadío. Encuesta en profundidad (1997-98).

AGRICULTOR	FUNCIÓN DÍA 1º	FUNCIÓN DÍA 2º	SALTO ENTRE LOS DOS INTERVALOS	COHERENCIA
1	BETA CON MODA A LA DERECHA	BETA CON MODA A LA DERECHA	PEQUEÑO	COHERENTE
2	BETA CON MODA A LA DERECHA	BETA CON MODA A LA DERECHA	IMPORTANTE	COHERENTE
3	BETA CON MODA A LA DERECHA	BETA CON MODA A LA DERECHA	IMPORTANTE	COHERENTE
4	BETA CON MODA A LA DERECHA	BETA BASTANTE CENTRADA	PEQUEÑO	COHERENTE
5	BETA CON MODA A LA DERECHA	BETA CON MODA A LA DERECHA	PEQUEÑO	MUY COHERENTE

En las respuestas predominan formas funcionales beta con modas desplazadas hacia la derecha.

Tabla 83.

Caracterización de las funciones de densidad estimadas por el segundo grupo de agricultores para el trigo de regadío. Encuesta en profundidad (1997-98).

AGRICULTOR	FUNCIÓN DÍA 1º	FUNCIÓN DÍA 2º	SALTO ENTRE LOS DOS INTERVALOS	COHERENCIA
6	BETA CON MODA A LA DERECHA	BETA CON MODA A LA DERECHA	PEQUEÑO	MUY COHERENTE
7	BIMODAL SUAVE HACIA LA DERECHA	BETA CON MODA A LA DERECHA	IMPORTANTE	COHERENTE
8	BETA CON MODA LIGERAMENTE A LA DERECHA	BETA CON MODA LIGERAMENTE A LA IZQUIERDA	NULO	COHERENTE
9	BETA CON MODA A LA DERECHA	BETA CON MODA A LA DERECHA	PEQUEÑO	COHERENTE
10	BETA CON MODA A LA DERECHA	BIMODAL SUAVE HACIA LA DERECHA	IMPORTANTE	COHERENTE

A diferencia de la cebada de secano, en el trigo de regadío existe mayor uniformidad para este grupo de agricultores en la forma asignada a la distribución de los rendimientos, observándose que podría tratarse de una distribución beta con la moda desplazada hacia la derecha. También se observa mayor coherencia en las respuestas que en la cebada de secano.

El caso del maíz de regadío se describe en las Tablas 84 y 85. (Los agricultores 2 y 9 no practicaban este cultivo).

Tabla 84.

Caracterización de las funciones de densidad estimadas por el primer grupo de agricultores para el maíz de regadío. Encuesta en profundidad (1997-98).

AGRICULTOR	FUNCIÓN DÍA 1º	FUNCIÓN DÍA 2º	SALTO ENTRE LOS DOS INTERVALOS	COHERENCIA
1	BETA CON MODA A LA DERECHA	TRIMODAL HACIA LA DERECHA	PEQUEÑO	COHERENTE
2				
3	BETA CON MODA A LA DERECHA	BETA CON MODA A LA DERECHA	PEQUEÑO	COHERENTE
4	BETA CENTRADA	BIMODAL LIGERAMENTE A LA IZQUIERDA	IMPORTANTE	COHERENTE
5	BETA CON MODA A LA IZQUIERDA	BETA CON MODA A LA DERECHA	IMPORTANTE	CIERTA INCOHERENCIA

Aunque no se observa un patrón de formas funcionales tan regular como en el caso del trigo en regadío, se puede apreciar cierta predominancia de las funciones beta, con la moda desplazada hacia la derecha.

Tabla 85.

Caracterización de las funciones de densidad estimadas por el segundo grupo de agricultores para el maíz de regadío. Encuesta en profundidad (1997-98).

AGRICULTOR	FUNCIÓN DÍA 1º	FUNCIÓN DÍA 2º	SALTO ENTRE LOS DOS INTERVALOS	COHERENCIA
6	BETA CON MODA A LA DERECHA	BETA CON MODA A LA DERECHA	IMPORTANTE	COHERENTE
7	BETA CON MODA LIGERAMENTE A LA DERECHA	BETA CON MODA A LA IZQUIERDA	PEQUEÑO	CIERTA INCOHERENCIA
8	BETA CON MODA A LA DERECHA	BETA CON MODA LIGERAMENTE A LA DERECHA	NULO	COHERENTE
9				
10	BETA CON MODA A LA DERECHA	BETA CON MODA A LA DERECHA	IMPORTANTE	COHERENTE

En este caso parecen predominar las funciones beta con moda desplazada hacia la derecha.

Cuando se pidió a los agricultores, durante el segundo día, que se pusieran en el lugar de otro agricultor del grupo e intentara estimar la distribución de los rendimientos obtenidos por el mismo, se observó que las funciones de densidad descritas eran similares a las expresadas para sus propios rendimientos. (El agricultor 4 no participó en esta actividad de simulación). Los resultados se han representado en el Anejo 6, donde **x (y)** indica la estimación del agricultor **x** poniéndose en el lugar del agricultor **(y).**

En lo que respecta a la mediana, el análisis se resume en la tabla 86.

Tabla 86.

Posición relativa de la mediana. Encuesta en profundidad (1997-98).

DIA 1										
Cultivos	0-0.1	0.1-0.2	0.2-0.3	0.3-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1
Cebada Secano			1	3	5	1				
Trigo Regadío						4	4	2		
Maíz Regadío (*)				1	1	2	2	2		
DIA 2										
Cebada Secano		1	4	1	3		1			
Trigo Regadío						4	4	2		
Maíz Regadío (*)					1	3	4			

(*) En este caso sólo se disponía de la opinión de 8 agricultores.

En el caso de la mediana de la cebada de secano, existe un gran número de modas a la izquierda del intervalo [0,4 – 0,6], el intervalo central de amplitud 2 x 1/10 del recorrido. Lo contrario ocurre en los cultivos de regadío.

#6.3.- COMPARACIÓN DE VALORES MEDIOS

Se calculó el valor medio de los rendimientos a partir de la distribución de frecuencias en los distintos intervalos ($R_{xf} = \Sigma f_i x_i$, con f_i = frecuencias y x_i = marca de clase del intervalo). Los resultados de estos cálculos, para los valores medios de cada día, se reproducen en la tabla 88. El valor medio del primer día se ha representado como R_{xf1} y el del segundo día como R_{xf2}.

Además de estos valores calculados se disponía de una estimación puntual directa de los rendimientos medios (R_m) de cada cultivo que se solicitó el primer día al sujeto entrevistado (además de los límites del recorrido, R_{max} y R_{min}).

Al comparar los valores medios calculados R_{xf} debe tenerse en cuenta que el primer día se establecieron las frecuencias de cada intervalo con una restricción en el número de garbanzos (limitados a 10), mientras que el segundo día esa limitación no existía.

Aunque con las funciones de densidad estimadas cada día es posible también calcular las correspondientes varianzas, se considera que la del segundo día es más creíble que la del primero, por las razones de limitación en el número de garbanzos discutida anteriormente. En la tabla 87 se reproduce el cálculo del valor medio R_{xf2} y de la varianza para las distribuciones obtenidas el segundo día.

Tabla 87.

Valores medios R_{xf2} (kg. / ha) y desviaciones típicas para las distribuciones de rendimientos del segundo día. Encuesta en profundidad (1997-98).

	CEBADA SECANO			TRIGO REGADÍO			MAÍZ REGADÍO		
	1	**1(2)**	**2**	**1**	**1(2)**	**2**	**1**	**1(2)**	**2**
MEDIA	3128	3473	2868	5704	6057	6090	11797	12091	10764
D. TÍPICA	915	1026	877	878	770	680	1625	1443	1476
	2	**2(1)**	**1**	**2**	**2(1)**	**1**	**2**	**2(1)**	**1**
MEDIA	2868	3910	3128	6090	6716	5704	10764	10810	11797
D. TÍPICA	877	1150	915	680	624	878	1476	1455	1625
	3	**3(2)**	**2**	**3**	**3(2)**	**2**	**3**	**3(2)**	**2**
MEDIA	2927	3321	2868	6524	6164	6090	11864	12535	10764
D. TÍPICA	855	1035	877	729	828	680	1222	1298	1476
	5	**5(2)**	**2**	**5**	**5(2)**	**2**	**5**	**5(2)**	**2**
MEDIA	3067	4621	2868	5126	6440	6090	11193	13018	10764
D. TÍPICA	887	988	877	831	582	680	2400	1878	1476
	CEBADA SECANO			**TRIGO REGADÍO**			**MAÍZ REGADÍO**		
	6	**6(10)**	**10**	**6**	**6(10)**	**10**	**6**	**6(10)**	**10**
MEDIA	3588	3588	2986	5405	5175	5586	11756	11500	11500
D. TÍPICA	855	855	1188	1035	1178	1367	1695	1626	2387
	7	**7(9)**	**9**	**7**	**7(9)**	**9**	**7**	**7(9)**	**9**
MEDIA	4761	3667	3367	5827	6670	5662	10427	11082	-
D. TÍPICA	1631	1140	1189	1755	1467	1054	3127	2153	-
	8	**8(10)**	**10**	**8**	**8(10)**	**10**	**8**	**8(10)**	**10**
MEDIA	3657	3450	2986	4830	5175	5586	10494	10580	11500
D. TÍPICA	825	946	1188	1274	1211	1367	1141	1127	2387
	9	**9(10)**	**10**	**9**	**9(10)**	**10**	**9**	**9(10)**	**10**
MEDIA	3367	3488	2986	5662	5485	5586	-	-	-
D. TÍPICA	1189	1196	1188	1054	1506	1367	-	-	-
	10	**10(9)**	**9**	**10**	**10(9)**	**9**	**10**	**10(9)**	**9**
MEDIA	2986	3273	3367	5586	5257	5662	11500	10969	-
D. TÍPICA	1188	1230	1189	1367	1303	1054	2387	1839	-

NOTA: **x(y)**: estimación del agricultor número x poniéndose en lugar del agricultor número y.

Los rendimientos calculados para cada distribución, del primer y del segundo día, R_{xf1} y R_{xf2} (tabla 88), se pueden comparar con la estimación puntual realizada el primer día, R_m. En las columnas encabezadas como "Comparaciones (%)", se han reflejado los siguientes cálculos:

- COMP1: la comparación del rendimiento medio calculado a partir de la distribución de frecuencias del primer día (R_{xf1}) respecto al rendimiento medio estimado directamente por el agricultor (R_m), dividido por el rendimiento medio estimado, o sea, $(R_{xf1}-R_m) / R_m$;

- COMP2: la comparación del rendimiento medio calculado a partir de la distribución de frecuencias del segundo día (R_{xf2}) respecto al rendimiento medio estimado, dividido por el rendimiento medio estimado, o sea, $(R_{xf2}-R_m) / R_m$;

- COMP3: la comparación del rendimiento medio calculado a partir de las distribuciones de los dos días respecto al menor de ellos, es decir, $(R_{xf1}-R_{xf2}) / MIN (R_{xf1}, R_{xf2})$

Tabla 88.

Comparaciones entre rendimientos medios calculados y estimados para la cebada de secano. Encuesta en profundidad (1997-98).

						COMPARACIONES (%)		
AGRICULTOR	**R_{max}**	**R_{min}**	**R_m**	**R_{xf1}**	**R_{xf2}**	**COMP1**	**COMP2**	**COMP3**
1	5750	920	2300	3232	3128	40.5	36	3.3
2	7590	1380	4370	3740	3128	14.4	28.4	19.5
3	4600	1610	2070	2926	2927	41.3	41.4	0.0
4	3450	1150	2530	2116	2852	16.4	12.7	34.8
5	4600	920	2530	3054	3067	20.7	21.2	0.4
6	5405	1725	4140	3643	3588	12	13.3	1.5
7	6325	2070	3565	3809	4761	6.8	33.5	25
8	5980	2070	3680	3400	3657	7.6	0.6	7.5
9	4600	1610	2990	2801	3367	6.3	12.6	20.2
10	4600	1840	3330	3330	2986	20.6	8.2	11.5

COMP1 y COMP2 indican porcentualmente las diferencias entre las estimaciones directas de rendimientos medios (realizadas el primer día), y los rendimientos medios que se deducen de la asignación de frecuencias a cada uno de los cinco intervalos (realizadas el primer y segundo día). Para la estimación de frecuencias realizada el primer día, en todos los casos esas diferencias son superiores al 5% del valor de la estimación directa puntual del rendimiento medio, R_m. En cinco casos son inferiores al 15%. En dos casos son del orden del 20% y en otros dos casos alcanzan una cifra francamente alta, del orden del 40%.

En las estimaciones de frecuencia del segundo día, existe una muy buena aproximación (agricultor 8) y en cinco casos las diferencias son inferiores al 15%. El resto de las estimaciones toman valores por encima del 20% y en dos casos valores del orden del 36 – 40%.

La conjetura de que la diferencia se debe a "visiones" aisladas de la distribución de frecuencias y la estimación puntual del rendimiento medio se ve rechazada por los datos ofrecidos en la tabla con la razón COMP3. Para las distribuciones de frecuencias ofrecidas cada día, las diferencias entre los valores medios que se deducirían de las mismas son aceptables en 5 casos (COMP3 inferior al 15%), pero existen 4 estimaciones con diferencias superiores al 20%.

Excepto el agricultor 8, no se puede decir que los agricultores mantengan una calidad similar en sus estimaciones de un día a otro, ni que sean coherentes entre sus estimaciones numéricas puntuales y por intervalos. Conjeturamos, sin embargo, que de esta discusión no se puede concluir que los agricultores no sean coherentes en sus estimaciones, ya que puede que las diferencias se deriven de la metodología utilizada para la obtención de la información. Es necesario seguir profundizando en estos aspectos en trabajos futuros.

Comparaciones semejantes se han realizado en las tablas 89 y 90, para los cultivos de regadío (trigo y maíz), donde los resultados son bastante mejores (por no decir francamente buenos), sin que tampoco a partir de ellos se pueda realizar una conclusión general sin más experimentación.

Tabla 89.

Comparaciones entre rendimientos medios calculados y estimados para el trigo de regadío. Encuesta en profundidad (1997-98).

	OPINIÓN			RESULTADOS		COMPARACIONES (%)		
AGRICULTOR	**R_{max}**	**R_{min}**	**R_m**	**R_{m1}**	**R_{m2}**	**COMP1**	**COMP2**	**COMP3**
1	6900	3910	5750	5704	5584	0.8	2.9	2.1
2	7360	4600	5750	6532	5888	13.6	2.4	10.9
3	6900	4600	5290	6164	6523	16.5	23.3	5.8
4	7360	2760	5060	5428	5444	7.3	7.6	0.3
5	6670	3450	5750	5126	5124	10.8	10.9	0.0
6	7360	3450	6210	5538	5405	10.8	12.9	2.5
7	7475	4025	6325	6394	5827	1.1	7.9	9.7
8	9200	2300	6325	5336	4830	15.6	23.6	10.5
9	6900	4600	5750	5750	5662	0.0	1.5	1.6
10	6900	4600	5750	5750	5586	0.0	2.8	2.9

Tabla 90.

Comparaciones entre rendimientos medios calculados y estimados para el maíz de regadío. Encuesta en profundidad (1997-98).

	OPINIÓN			RESULTADOS		COMPARACIONES (%)		
AGRICULTOR	**R_{max}**	**R_{min}**	**R_m**	**R_{m1}**	**R_{m2}**	**COMP1**	**COMP2**	**COMP3**
1	13800	8050	10350	11385	11650	10	12.5	2.3
3	13800	6900	10350	12144	11863	17.3	14.6	2.4
4	13800	8050	10925	10925	10321	0.0	5.5	5.8
5	14720	6900	11500	10184	11194	11.4	2.7	9.9
6	13225	9200	11500	11684	11756	1.6	2.2	0.6
7	10350	5175	8625	8740	10427	1.3	20.6	19.3
8	12650	10005	11500	10796	10494	6.1	8.7	2.9
10	16100	9200	11500	12098	11500	5.2	0.0	5.2

En la tabla 91 se ha calculado el valor medio y la desviación típica de las razones COMP1, COMP2 y COMP3 a fin de ordenarlas en función de la calidad de sus resultados.

Tabla 91.

Comparaciones entre estimaciones de rendimientos medios (en tanto por uno). Encuesta en profundidad (1997-98).

	MEDIA	**DESVIACIÓN ESTÁNDARD**	**ORDEN RESULTADOS (DE MEJOR A PEOR)**
Cebada de secano			
COMP1	0.2200	0.1269	2
COMP2	0.2350	0.1285	3
COMP3	0.1500	0.1025	1
Trigo de regadío			
COMP1	0.1150	0.0594	2
COMP2	0.1200	0.0748	3
COMP3	0.0800	0.0400	1
Maíz de regadío			
COMP1	0.1000	0.0500	2
COMP2	0.1125	0.0649	3
COMP3	0.0875	0.0484	1

En todos los casos, las diferencias entre los valores calculados a partir de las estimaciones de frecuencias son menores que las diferencias de cada estimación de frecuencias con la correspondiente estimación puntual del valor medio. Existe una menor diferencia entre el rendimiento medio derivado de la estimación de frecuencias del primer día y la estimación puntual del mismo día, como si hubiera existido algún valor de anclaje el primer día que hubiera variado el segundo día.

Las mejores estimaciones por cultivos, se tiene el siguiente orden para cada razón:

COMP1: CEBADA < TRIGO < MAÍZ
COMP2: CEBADA < TRIGO < MAÍZ
COMP3: CEBADA < MAÍZ < TRIGO

La cebada (de secano) es el cultivo peor estimado en todas las comparaciones realizadas.

#6.4.- COMPARACIÓN DE VALORES EXTREMOS

Se ha estudiado también la relación existente entre las estimaciones de los extremos de los recorridos realizada por los agricultores el primer y el segundo día. En la estimación del primer día, los agricultores definían el recorrido (el rendimiento máximo y el mínimo) sin ninguna influencia exterior, tal como lo percibían. En el segundo día, por el contrario, se les presentaba un recorrido de partida (construido a partir de los valores máximo y mínimo dados por el conjunto de los agricultores el primer día).

En las tablas 92, 93 y 94 se reproducen las estimaciones realizadas por el agricultor los diferentes días. En las columnas encabezadas como “Resultado (día 2)” se indica si durante la entrevista del segundo día el agricultor señaló algún valor, fuera del intervalo definido el primer día, con un porcentaje significativo. Así, por ejemplo, en la tabla 92, el agricultor 4 que el primer día definió un recorrido [1150 – 3450], sorprende el segundo día indicando que la clase del rendimiento 4370 tiene un 5% de probabilidad y la de 3680 un 16%, atribuyendo de esta forma un 21% a una región que el primer día había sido situada fuera del recorrido de los rendimientos.

En las columnas encabezadas por “Comparación” se han expresado, en tanto por uno, las diferencias en los extremos superior e inferior declarados el primer día y desbordados el segundo día, normalizadas mediante división de esa diferencia por el valor del rendimiento medio estimado puntualmente el primer día.

Tabla 92.

Diferencias relativas de rendimientos extremos para la cebada de secano. Encuesta en profundidad (1997-98).

AGRICULTOR	RESULTADO (día 1)			RESULTADO (día 2)		COMPARACIÓN	
	R_{max}	R_{min}	R_m	R_{max}	R_{min}	DIFMAX / R_m	DIFMIN / R_m
1	5750	920	2300				
2	7590	1380	4370				
3	4600	1610	2070				
4	3450	1150	2530	4370 (5%)*		0.3636	
5	4600	920	2530				
6	5405	1725	4140				
7	6325	2070	3565	6900 (19%)	1380 (6%)	0.1613	0.1935
8	5980	2070	3680				
9	4600	1610	2990	4692 (30%)	1380 (20%)	0.0308	0.0769
10	4600	1840	3330	4692 (18%)	1380 (27%)	0.0276	0.1381

*Asigna un 5% a 4370 kg/ha y un 16% a 3680 kg/ha.

Donde los números entre paréntesis significan el peso que da el agricultor a ese rendimiento.

El valor medio de las discrepancias (incoherencias) para el caso de la cebada de secano es de 0,1458 para el máximo y de 0,1362 para el mínimo, con unas desviaciones típicas de 0,1368 y de 0,0480 respectivamente. El porcentaje de incoherencias es del 40% en el caso del máximo (4 agricultores de 10) y del 30% en el caso del mínimo (3 agricultores de 10).

Tabla 93.

Diferencias relativas entre rendimientos extremos para el trigo de regadío. Encuesta en profundidad (1997-98).

AGRICULTOR				RESULTADO		COMPARACIÓN	
	R_{max}	R_{min}	R_m	$R_{max'}$	$R_{min'}$	DIFMAX / R_m	DIFMIN / R_m
1	6900	3910	5750				
2	7360	4600	5750				
3	6900	4600	5290	7360(36%)		0.0870	
4	7360	2760	5060				
5	6670	3450	5750				
6	7360	3450	6210				
7	7475	4025	6325	8050(20%)	2300(7%)*	0.0909	0.2727
8	9200	2300	6325				
9	6900	4600	5750		3450(8%)		0.2000
10	6900	4600	5750	8050(7%)	3450(14%)	0.2000	0.2000

*Asigna un 7% a 2300 kg/ha y un 13% a 3450 kg/ha.

El valor medio de las discrepancias (incoherencias) para el caso del trigo de regadío es de 0,1260 para el máximo y de 0,2242 para el mínimo, con unas desviaciones típicas de 0,0524 y de 0,0343 respectivamente. El porcentaje de incoherencias es del 30% (3 agricultores de 10) en el máximo y en el mínimo.

Tabla 94.

Diferencias relativas entre rendimientos extremos para el maíz de regadío. Encuesta en profundidad (1997-98).

AGRICULTOR				RESULTADO		COMPARACIÓN	
	R_{max}	R_{min}	R_m	$R_{max'}$	$R_{min'}$	DIFMAX/R_m	DIFMIN/R_m
1	13800	8050	10350				
3	13800	6900	10350				
4	13800	8050	10925		6900(13%)		0.1053
5	14720	6900	11500				
6	13225	9200	11500	13800(33.3%)		0.0500	
7	10350	5175	8625	16100(6.7%)*	4600(6.7%)	0.6667	0.0667
8	12650	10005	11500				
10	16100	9200	11500		6900(7.7%)		0.2000

*Asigna un 6.7% a 16100 kg/ha y un 20% a 13800 kg/ha.

El valor medio de las discrepancias (incoherencias) para el caso del maíz de regadío es de 0,3583 para el máximo y de 0,1240 para el mínimo, con unas desviaciones típicas de 0,3083 y de 0,0560 respectivamente. El porcentaje de incoherencias es del 25% (2 agricultores de 8) en el máximo y del 37,5% en el mínimo.

En general se han realizado peores estimaciones de los valores extremos en los cultivos de regadío.

CAPÍTULO 7

ENCUESTAS GENERALES A LOS AGRICULTORES

#7.1.- ENCUESTAS A LOS AGRICULTORES

Durante dos cursos académicos, el 1998 – 1999 y el 1999 – 2000, se realizaron un conjunto de encuestas a los agricultores, siendo los encuestadores estudiantes de los cursos terminales de segundo ciclo de las carreras de Ingeniero Agrónomo e Ingeniero de Montes. Los encuestadores seleccionaron y entrevistaron a agricultores con los que mantenían algún tipo de relación (familiar, de amistad, de vecindad, etc.), generalmente entre Noviembre y Enero de cada curso escolar.

La encuesta tenía como objetivo evaluar la coherencia de las respuestas sobre rendimientos de diferentes cultivos, obtenidas mediante encuestas convencionales, realizadas por personal sin una preparación especializada en los temas objeto de estudio. La versión del primer año se reproduce en el Anejo 7, y estaba estructurada en dos cuestionarios. Se sometía al agricultor a cada cuestionario en días distintos.

El primer cuestionario o “Encuesta del primer día” tenía por objeto determinar la función de densidad de los rendimientos de algunos cultivos. El segundo cuestionario o “Encuesta del segundo día” tenía por objeto obtener información adicional para juzgar la persistencia y coherencia de las respuestas obtenidas en ambos días. Cada agricultor era entrevistado con el primer cuestionario por un estudiante. Al menos una semana después, otro estudiante distinto realizaba con el segundo cuestionario otra entrevista al mismo agricultor. Durante el curso 1998 – 99 se realizaron 52 encuestas.

La revisión de las respuestas obtenidas el primer año puso de manifiesto distintos problemas, que son discutidos en el texto más adelante, y originó que se modificaran los cuestionarios para las encuestas realizadas durante el curso 1999 – 2000. Esta encuesta modificada se reproduce en el Anejo 8, y con ella se realizaron 44 encuestas, siguiendo una metodología similar a la expuesta anteriormente.

#7.2.- ENCUESTA DEL PRIMER AÑO (1998 – 1999)

Las características de las 52 explotaciones de los agricultores entrevistados durante el curso 1998-99 se han esquematizado en las Tablas 95 a 97.

Tabla 95.

Características de las explotaciones de los agricultores entrevistados (1998 – 99).

			Superficie secano (ha)			Superficie regadío (ha)			Superficie		Otras
Nº	**Población**	**Provincia**	**Herbáceos**	**Leñosos**	**Total**	**Herbáceos**	**Leñosos**	**Total**	**Total (ha)**	**Ganado**	**actividades**
13	Barrax	Albacete	70	0	70	0	0	0	70		
22	Las Peñas	Albacete	50	2,5	52,5	39,7	0	39,7	92,2	Ov	
45	María	Almería	10	0	10	0	0	0	10		
14	Villanueva del Aceral	Ávila	29	0	29	0	0	0	29	Bo,Por,Av	
6	Sa Roca	Balears	162	20	182	10	0	10	192	Por	ES
16	Caldes de Montbui	Barcelona	20	3,2	23,2	8	3	11	34,2	Bo,Por,Ov	
23	Avinyó	Barcelona	190	0	190	0	0	0	190	Bo	
26	Rajadell	Barcelona	80	0	80	0	0	0	80		ES

Nº	Población	Provincia	Superficie secano (ha)			Superficie regadío (ha)			Superficie	Ganado	Otras
			Herbáceos	Leñosos	Total	Herbáceos	Leñosos	Total	Total (ha)		actividades
33	Pont Vilomara i Rocafort	Barcelona	0,5	3,9	4,4	1,5	1,5	3	7,4		
39	Sta. Maria de Merles	Barcelona	14	0	14	0	0	0	14		
50	Canyamars	Barcelona	25	0	25	22	0	22	47	Bo	FOR
19	Pau	Girona	38	0	38	35	0	35	73		
48	Osona	Girona	71	0	71	0	0	0	71	Bo	
49	Vic	Girona	44	0	44	0	0	0	44		
52	Ordís	Girona	108	0	108	29	0	29	137	Bo	
4	Gurrea de Gállego	Huesca	160	0	160	66	0	66	226		
8	Peralta Sal	Huesca	120	15	135	3	0	3	138	Ov	
11	Betesa	Huesca	40	0	40	0	0	0	40	Ov	
12	Lascasas	Huesca	37	0	37	333	0	333	370		
21	Altorricó	Huesca	12	0	12	18	0	18	30	Ov	
43	Tamarite de Litera	Huesca	30	0	30	17	0	17	47	Bo	
1	Menàrguens	Lleida	20,5	0	20,5	8,74	3,74	12,48	32,98		
37	Bell-lloc	Lleida	0	0	0	3	0	3	3	Av,Ot	
2	Castellnou de Seana	Lleida	0	0	0	9,6	4,8	14,4	14,4		
10	Alfarràs	Lleida	41	0	41	16	2,4	18,4	59,4	Bo,Por	
3	Almenar	Lleida	4	0	4	22	0	22	26	Bo	
18	Almenar	Lleida	5	0	5	15	0	15	20		
32	Vilanova de la Barca	Lleida	0	0	0	26	0	26	26		
44	Alcanós	Lleida	80	0	80	0	0	0	80		
27	Balaguer	Lleida	28	0	28	23	2,4	25,4	53,4		
42	Artesa de Segre	Lleida	10	0	10	1	0	1	11		
40	Ivars de Noguera	Lleida	22	4,5	26,5	0	3,5	3,5	30		
31	Ponts	Lleida	8,5	0	8,5	0	0	0	8,5		
34	Ponts	Lleida	0	0	0	0	0	0	0		
30	Guimerà	Lleida	30	9	39	0	0	0	39		
38	Castellserà	Lleida	0	0	0	4	1	5	5		
25	Sant Guim de Freixenet	Lleida	150	0	150	0	0	0	150	Por,Av	
46	Les Oluges	Lleida	33	0	33	0	0	0	33	Por	
5	Salàs de Pallars	Lleida	69,9	15	84,9	71,3	0	71,3	156,2	Ov	
7	Isona	Lleida	50	0	50	0	0	0	50	Bov	
9	Beire	Navarra	33	0	33	47,6	0	47,6	80,6		
20	Tafalla	Navarra	95	10	105	0	0	0	105		
47	Oteiza de la Solana	Navarra	4,1	3,9	8	1	0	1	9		
51	Irujo	Navarra	25	0	25	0	0	0	25		
15	Iratxeta	Navarra	21	0	21	0	0,2	0,2	21,2	Por	
17	Berriosuso	Navarra	45	0	45	0	0	0	45	Por	
28	Naveros de Pisuerga	Palencia	260	0	260	54	0	54	314		
35	Cenicero	Rioja	3	0	3	0	0	0	3		
24	Pira	Tarragona	45	67	112	0	0	0	112		BO

Nº	Población	Provincia	Superficie secano (ha)			Superficie regadío (ha)			Superficie		Otras
			Herbáceos	Leñosos	Total	Herbáceos	Leñosos	Total	Total (ha)	Ganado	actividades
36	Banyeres del Penedès	Tarragona	0	0	0	75	0	75	75	Bo	
29	Calanda	Teruel	45	0	45	0	0	0	45		
41	Layana	Zaragoza	204	0	204	58,7	0	58,7	262,7	Ov	

Donde **Ov** indica ovino, **Por** porcino, **Bo**/**BO** bovino, **Av** avícola, **FOR** forestal y **ES** empresa de servicios.

Tabla 96.

Superficies (ha) de los principales tipos de cultivos de secano en las explotaciones encuestadas (1998 –99).

	SECANO						
NÚMERO	Cereales	Forrajes	Oleaginosas	Otros	Herbáceos	Leñosos	TOTAL
13	70	0	0	0	70	0	70
22	50	0	0	0	50	2,5	52,5
45	10	0	0	0	10	0	10
14	22	0	7	0	29	0	29
6	162	0	0	0	162	20	182
16	14	0	6	0	20	3,2	23,2
23	190	0	0	0	190	0	190
26	70	10	0	0	80	0	80
33	0,5	0	0	0	0,5	3,9	4,4
39	7	0	0	7	14	0	14
50	15	10	0	0	25	0	25
19	18	0	20	0	38	0	38
48	36	35	0	0	71	0	71
49	22	22	0	0	44	0	44
52	88	20	0	0	108	0	108
4	120	20	0	20	160	0	160
8	120	0	0	0	120	15	135
11	20	20	0	0	40	0	40
12	37	0	0	0	37	0	37
21	0	12	0	0	12	0	12
43	30	0	0	0	30	0	30
1	20,5	0	0	0	20,5	0	20,5
37	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
10	1	0	40	0	41	0	41
3	4	0	0	0	4	0	4
18	5	0	0	0	5	0	5
32	0	0	0	0	0	0	0
44	80	0	0	0	80	0	80
27	26	0	2	0	28	0	28
42	10	0	0	0	10	0	10
40	22	0	0	0	22	4,5	26,5
31	8,5	0	0	0	8,5	0	8,5

	SECANO						
NÚMERO	**Cereales**	**Forrajes**	**Oleaginosas**	**Otros**	**Herbáceos**	**Leñosos**	**TOTAL**
34	0	0	0	0	0	0	0
30	30	0	0	0	30	9	39
38	0	0	0	0	0	0	0
25	150	0	0	0	150	0	150
46	33	0	0	0	33	0	33
5	44,8	17,5	7,6	0	69,9	15	84,9
7	40	5	5	0	50	0	50
9	33	0	0	0	33	0	33
20	95	0	0	0	95	10	105
47	4,1	0	0	0	4,1	3,9	8
51	25	0	0	0	25	0	25
15	21	0	0	0	21	0	21
17	45	0	0	0	45	0	45
28	260	0	0	0	260	0	260
35	3	0	0	0	3	0	3
24	45	0	0	0	45	67	112
36	0	0	0	0	0	0	0
29	45	0	0	0	45	0	45
41	204	0	0	0	204	0	204

Tabla 97.

Superficies (ha) de los principales tipos de cultivos en las explotaciones de regadío encuestadas (1998 – 99).

	REGADÍO						
NÚMERO	**Cereales**	**Forrajes y Pastos**	**Horticultura e Industria**	**Otros**	**Herbáceos**	**Leñosos**	**TOTAL**
13	0	0	0	0	0	0	0
22	0	39,7	0	0	39,7	0	39,7
45	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0
6	0	10	0	0	10	0	10
16	3	5	0	0	8	3	11
23	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0
33	0,5	0	1	0	1,5	1,5	3
39	0	0	0	0	0	0	0
50	7	15	0	0	22	0	22
19	10	25	0	0	35	0	35
48	0	0	0	0	0	0	0
49	0	0	0	0	0	0	0
52	0	20	0	9	29	0	29
4	44	20	0	2	66	0	66
8	0	3	0	0	3	0	3
11	0	0	0	0	0	0	0
12	293	0	0	40	333	0	333

NÚMERO	REGADÍO						
	Cereales	**Forrajes y Pastos**	**Horticultura e Industria**	**Otros**	**Herbáceos**	**Leñosos**	**TOTAL**
21	8	10	0	0	18	0	18
43	3	14	0	0	17	0	17
1	8,74	0	0	0	8,74	3,74	12,48
37	3	0	0	0	3	0	3
2	9,6	0	0	0	9,6	4,8	14,4
10	0	16	0	0	16	2,4	18,4
3	22	0	0	0	22	0	22
18	15	0	0	0	15	0	15
32	13	13	0	0	26	0	26
44	0	0	0	0	0	0	0
27	16	7	0	0	23	2,4	25,4
42	1	0	0	0	1	0	1
40	0	0	0	0	0	3,5	3,5
31	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0
38	4	0	0	0	4	1	5
25	0	0	0	0	0	0	0
46	0	0	0	0	0	0	0
5	40,8	13,2	0	17,3	71,3	0	71,3
7	0	0	0	0	0	0	0
9	32,4	15,2	0	0	47,6	0	47,6
20	0	0	0	0	0	0	0
47	0	0	1	0	1	0	1
51	0	0	0	0	0	0	0
15	0	0	0	0	0	0,2	0,2
17	0	0	0	0	0	0	0
28	18	20	16	0	54	0	54
35	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0
36	50	0	0	25	75	0	75
29	0	0	0	0	0	0	0
41	29,6	16,3	0	12,8	58,7	0	58,7

En la tabla 98 se muestra el número de respuestas obtenidas, en las 52 encuestas, sobre cada uno de los cultivos.

Tabla 98.

Número de respuestas para los diferentes cultivos. Encuesta a agricultures (1998 – 99).

	NÚMERO DE ENCUESTAS	
CULTIVO	**SECANO**	**REGADÍO**
Alfalfa	5	17
Almendro	1	
Avena	3	
"Branya"(Trigo+Ray-Grass+Veza)	1	
Cáñamo	1	
Cebada	46	6
Colza	1	
Esparceta	2	
Espárrago	1	
Forrajes	1	
Girasol	5	3
Lino No Textil	1	1
Maíz	2	20
Manzana		1
Melocotonero		1
Mijo	1	
Olivo	4	
Patata		1
Peral		2
Ray-Grass	1	
Remolacha		1
Sorgo	3	
Trigo	20	13
Trigo+Ray-Grass		1
Veza	1	
Viña	6	
Total	**107**	**67**

Para la discusión se han seleccionado los cultivos de los que se obtuvieron mayor número de respuestas, concretamente la cebada de secano (46 respuestas), el maíz de regadío (20 respuestas), el trigo de secano (20) y de regadío (13) y la alfalfa de regadío (17).

Como se ha comentado anteriormente, la encuesta del primer día (Anejo 7) tenía como objetivo determinar la función de densidad de rendimientos para los cultivos más conocidos por el agricultor. La primera pregunta, cuyas respustas se han resumido en esta sección, se refería a los datos identificativos de la explotación y de los encuestadores.

#7.3.- APROXIMACIONES A FUNCIONES TRIANGULARES Y BETAS

En la **Pregunta 2** del cuestionario del primer día se pedía al agricultor, para cada cultivo, una estimación directa de los rendimientos mínimo (R_{min}), máximo (R_{max}) y más frecuente (R_{mf}) para cada cultivo, con objeto de determinar el intervalo que el agricultor consideraba razonable para los rendimientos de los cultivos en su zona, y una primera aproximación a la distribución de los mismos. En el cuestionario del primer día no se solicitó una estimación directa puntual del rendimiento medio (R_m), estimación que sí se solicitó en la entrevista del segundo día.

A partir de las respuestas dadas por los agricultores, se ha calculado la posición de la moda normalizando los recorridos, de tal forma que se atribuye a todos el valor 1. La posición de la moda se ha determinado mediante la fórmula ($R_{mf} - R_{min}$) : ($R_{max} - R_{min}$). Cuando el valor de la expresión anterior es mayor que 0,5 indica que la moda se sitúa a la derecha del punto medio del recorrido. Los resultados obtenidos se han resumido en la tabla 99.

Tabla 99.

Posición de la moda en la estimación directa de recorrido y moda. Número de casos en cada intervalo. Encuesta a los agricultores, primer día (1998-99).

	Intervalo del valor ($R_{mf} - R_{min}$) / (R_{max} - R_{min})				
Cultivo	**0-0.2**	**0.2-0.4**	**0.4-0.6**	**0.6-0.8**	**0.8-1**
Cebada secano	1 (2%)	7 (15%)	24 (53%)	14 (30%)	0
Trigo secano	0	9 (36%)	9 (36%)	7 (28%)	0
Trigo regadío	0	2 (15%)	9 (70%)	2 (15%)	0
Maíz regadío	1 (5%)	1 (5%)	14 (70%)	4 (20%)	0
Alfalfa regadío (*)	1 (6%)	0	3 (20%)	6 (37%)	6 (37%)

(*) Habiendo eliminado una de las observaciones por ser incompleta.

De las modas relativamente centradas (para valores 0,4 – 0,6 en la tabla 99), conviene identificar aquellas que se sitúan exactamente en la mitad del recorrido. Por ello, en la tabla 100 se indica el número y porcentaje de encuestas cuya moda toma el valor 0,5 en la escala numérica discutida.

Tabla 100.

Posición central de la moda en la estimación directa de recorrido y moda. Número de casos en cada intervalo. Encuesta a los agricultores, primer día (1998-99).

Cultivos	**Moda=0.5**	**Total encuestas**	**Porcentaje (%)**
Cebada secano	8	46	17
Trigo secano	3	25	12
Trigo regadío	3	13	23
Maíz regadío	3	20	15
Alfalfa regadío (*)	2	16	12

(*) Habiendo eliminado una de las observaciones por ser incompleta.

No se ha observado que exista una correlación entre el valor del recorrido y la posición de la moda. La distribución de la nube de puntos se reproduce en la Figura 2.

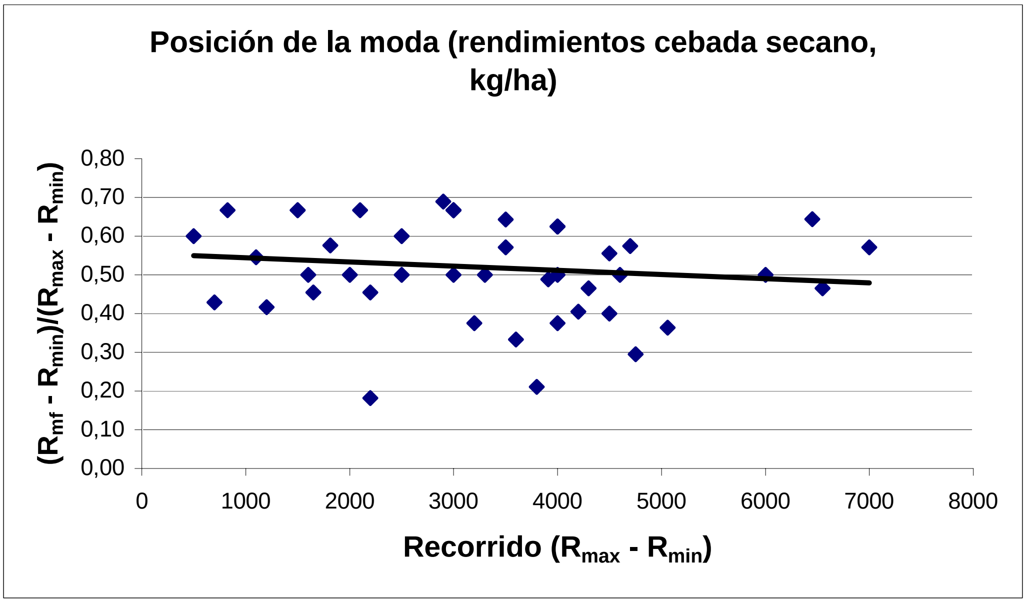


Figura 2.

Posición de la moda en función del recorrido. Encuesta a los agricultores, primer día (1998-99).

#7.3.1.- RESULTADOS EN EL SEGUNDO DÍA

En la encuesta del segundo día (realizada al menos una semana después) en la **pregunta 2** se pedía una estimación directa puntual de los rendimientos medios, R_m. Más tarde, en la **pregunta 9**, se volvía a solicitar una estimación directa de los rendimientos mínimos (R_{min}), más frecuentes (R_{mf}) y máximos (R_{max}).

La disposición de las preguntas intentaba minimizar la influencia de un posible anclaje común de todos esos valores.

Se ha calculado la posición relativa de la moda (R_{mf}) en intervalos normalizados, calculada mediante $(R_{mf} - R_{min}) : (R_{max} - R_{min})$, y los resultados se han resumido en las tablas 101 y 102.

Tabla 101.

Posición de la moda en la estimación directa de recorrido y moda. Número de casos en cada intervalo. Encuesta a los agricultores, segundo día (1998-99).

	Intervalo del valor $(R_{mf} - R_{min}) / (R_{max} - R_{min})$				
Cultivo	**0-0,2**	**0,2-0,4**	**0,4-0,6**	**0,6-0,8**	**0,8-1**
Cebada secano (*)	0	8(18%)	27(60%)	10(22%)	0
Trigo secano	0	7 (29%)	14 (58%)	3 (13%)	0

Trigo regadío	0	1 (8%)	9 (69%)	3 (23%)	0
Maíz regadío	0	3 (15%)	13 (65%)	4 (20%)	0
Alfalfa regadío (**)	0	3 (25%)	5 (42%)	4 (33%)	0

(*) No se considera una de las observaciones, por dar el rendimiento más frecuente inferior al rendimiento medio.
(**) No se consideran 3 de las observaciones por ser incompletas o dar valores del rendimiento más frecuente inferiores al rendimiento mínimo.

En la tabla 102 se indican las encuestas con moda igual a 0.5 y el porcentaje que representan sobre el total.

Tabla 102.

Posición central de la moda en la estimación directa de recorrido y moda. Número de casos en cada intervalo. Encuesta a los agricultores, segundo día (1998-99).

Cultivos	**Moda=0.5**	**Total encuestas**	**Porcentaje (%)**
Cebada secano (*)	7	45	16
Trigo secano	3	24	12
Trigo regadío	3	13	23
Maíz regadío	4	20	20
Alfalfa regadío (*)	2	12	17

(*) Habiendo eliminado una de las observaciones por ser incompleta.

La comparación entre el valor medio del rendimiento estimado directamente por el agricultor, y los valores medios calculados a partir de la aproximación funcional a una función triangular (RMT) y a una beta – PERT (RMB) se ha realizado en la tabla 103. Todos estos valores corresponden al segundo día.

Tabla 103.

Distribución del número de casos según las diferencias relativas entre valores medios (estimación directa y calculado). Encuesta a los agricultores, segundo día (1998-99).

TRIANGULAR	**Número de casos**										
Cultivo	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**Mediana**
Cebada secano	17	14	3	5	4	1	1	1			0.071
Trigo secano	12	7	1	2		1		1			0.050
Trigo regadío	6	3	2	2							0.058
Maíz regadío	13	3		3			1				0.038
Alfalfa regadío	7		5			1	1	1			0.094
BETA	**Número de casos**										
Cultivo	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**Mediana**
Cebada secano	24	13	1	3	2	1	1	1			0.048
Trigo secano	17	3	1	2	1						0.035
Trigo regadío	9	2		1	1						0.036
Maíz regadío	15	2	1	1				1			0.033
Alfalfa regadío	7	5		1		1				1	0.063

Se observa para todos los cultivos una menor diferencia relativa para la distribución beta – PERT, en comparación con la triangular. En la mitad de los casos (mediana) de la aproximación triangular los errores son inferiores al 10%, y en el caso de la distribución beta – PERT , inferiores al 7%.

Se realizó a continuación una comparación entre las poblaciones de rendimientos medios estimados R_m y calculados a partir de las distribuciones triangular (RMT) y beta - PERT (RMB), obteniendo los resultados resumidos en la Tabla 104, que indican una gran coincidencia entre esos valores. De nuevo la aproximación beta – PERT ofrece mejores resultados que la triangular. Todos los datos son del segundo día.

Tabla 104.

Comparación entre las poblaciones de rendimientos medios (estimación directa y calculados). Encuesta a los agricultores, segundo día (1998-99).

Comparación con la distribución triangular			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.893	No	No diferencias
Trigo secano	0.924	No	No diferencias
Trigo regadío	0.902	No	No diferencias
Maíz regadío	0.855	No	No diferencias
Alfalfa regadío	0.712	No	No diferencias
Comparación con la distribución beta			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.912	No	No diferencias
Trigo secano	0.964	No	No diferencias
Trigo regadío	0.887	No	No diferencias
Maíz regadío	0.885	No	No diferencias
Alfalfa regadío	0.705	No	No diferencias

Una tercera comprobación se realizó mediante regresión (tabla 105), con todos los datos del segundo día.

Tabla 105.

Regresión $R_m = a\ RMT + b$ y $R_m = a\ RBT + b$. Encuesta a los agricultores, segundo día (1998-99).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	1.03	1.02	0.99	1.00	0.78
b	-115.9	-140.5	204.5	-173.6	3838.5
R^2	0.87	0.93	0.78	0.99	0.81
Fc	286.7	305.4	38.7	1661.3	56.1
Ft	4.06	4.3	4.84	4.41	4.54
DISTRIBUCIÓN BETA					
a	1.02	1.02	1.03	1.00	0.80
b	-106.25	-127.0	-77.6	-308.1	3541.1
R^2	0.90	0.97	0.82	0.99	0.74
Fc	384.24	663.8	50.02	2235.5	37.0
Ft	4.06	4.3	4.84	4.41	4.54

Las mismas comprobaciones se han realizado para el caso en que se fuerza a que b = 0, resumiéndose los resultados en la tabla 106, para datos del segundo día. Se puede comprobar la fuerte correspondencia entre los valores medios directamente estimados en la pregunta 2 y los calculados a partir de las tres estimaciones (R_{min}, R_{mf}, R_{max}) de la **pregunta 9**, para todos los cultivos, en esta última regresión.

Tabla 106.

Regresión $R_m = a\ RMT$ (b = 0) y $R_m = a\ RMB$ (b = 0). Encuesta a los agricultores, segundo día (1998-99).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.99	0.98	1.02	0.99	1.00
b	0	0	0	0	0
R^2	0.99	0.99	0.99	0.99	0.98
Fc	3161.31	2839.87	1472.95	4834.81	573.68
Ft	4.06	4.28	4.75	4.38	4.49
DISTRIBUCIÓN BETA					
a	0.99	0.99	1.01	0.99	1.01
b	0	0	0	0	0
R^2	0.99	0.99	0.99	0.99	0.97
Fc	4102.05	5898.09	1819.07	6260.46	474.72
Ft	4.06	4.28	4.75	4.38	4.49

#7.3.2.- COMPARACIONES ENTRE R_m DEL SEGUNDO DÍA Y VALORES MEDIOS CALCULADOS DEL PRIMER DÍA

Las anteriores estimaciones se han realizado para los valores obtenidos el segundo día para la estimación directa puntual del rendimiento medio (R_m), en la **pregunta 2**, y de los rendimientos (R_{min}, R_{mf}, R_{max}), en la pregunta 9. Aunque la separación entre las preguntas pretendía evitar que se utilizara en la respuesta un anclaje común, es obvio que las estimaciones seguramente no serán independientes.

A continuación se realizará la comparación entre el valor del rendimiento medio (R_m) estimado directamente en el cuestionario del segundo día (**pregunta 2**), con los valores obtenidos, el primer día (también **pregunta 2**), de la distribución (R_{min}, R_{mf}, R_{max}). Entre la obtención de estos valores, había transcurrido al menos una semana de tiempo, y los datos se habían obtenido con dos encuestadores diferentes.

En la tabla 107 se señalan las diferencias relativas encontradas entre R_m (estimación del segundo día) y RMT y RMB (con datos del primer día), expresadas mediante los valores absolutos de (RMT – R_m): R_m y de (RMB – R_m): R_m.

Tabla 107.

Distribución del número de casos según las diferencias relativas entre valores medios (estimación directa R_m del segundo día y calculados a partir de estimaciones R_{min}, R_{mf}, R_{max} del primer día). Encuesta a los agricultores, primer y segundo día (1998-99).

APROXIMACIÓN											
TRIANGULAR	**Número de casos**										
Cultivo	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**>0.5**	**Mediana**
Cebada secano	17	11	7	4	3	1	1	2				0.077
Trigo secano	9	7	3	3	3		1			1		0.071
Trigo regadío	7	3	2	1								0.046
Maíz regadío	9	5	3	2		1						0.060
Alfalfa regadío	6		4							1	3	0.112
BETA	**Número de casos**											
Cultivo	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**>0.5**	**Mediana**
Cebada secano	20	12	5	5	1	2		1				0.062
Trigo secano	13	5	2	2	1			1				0.046
Trigo regadío	9	2	1	1								0.036
Maíz regadío	13	3	2	1			1					0.038
Alfalfa regadío	6	4				1					3	0.062

De nuevo se observa para todos los cultivos una menor diferencia relativa para la distribución beta, concentrando mayor cantidad de casos alrededor de cero. En el caso de la alfalfa, se obtiene una mayor dispersión en ambos casos, por lo que se opta por eliminar, a partir de ahora, las 4 observaciones que ocasionan las mayores diferencias, al considerarlas "outliers".

Con los datos provenientes de cuestionarios de los dos días diferentes (R_{min}, R_{mf}, R_{max} del primer día y R_m del segundo día) se realizó la comparación entre las poblaciones de valores medios (tabla 108):

Tabla 108.

Comparación entre poblaciones de valores medios (estimación directa R_m del segundo día y calculados a partir de estimaciones R_{min}, R_{mf}, R_{max} del primer día). Encuesta a los agricultores, primer y segundo día (1998-99).

Comparación con la distribución triangular			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.886	No	No diferencias
Trigo secano	0.915	No	No diferencias
Trigo regadío	0.879	No	No diferencias
Maíz regadío	0.923	No	No diferencias

Alfalfa regadío (*)	0.900	No	No diferencias
Comparación con la distribución beta			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.928	No	No diferencias
Trigo secano	0.941	No	No diferencias
Trigo regadío	0.876	No	No diferencias
Maíz regadío	0.952	No	No diferencias
Alfalfa regadío (*)	0.945	No	No diferencias

(*) Habiendo eliminado cuatro observaciones por ser incoherentes.

Obteniéndose en general correlaciones altas (mejores con la distribución beta) y dependencia en todas las poblaciones comparadas.

Los resultados de la **regresión** entre los valores de rendimientos medios (R_m) estimados en el día 2 y los calculados a partir de distribuciones triangular (*RMT*) y beta (*RMB*) con los datos del día 1, se reflejan en la tabla 109.

Tabla 109.

Regresión $R_m = a\ RMT + b$. (Estimación directa R_m del segundo día y calculados a partir de estimaciones R_{min}, R_{mf}, R_{max} del primer día). Encuesta a los agricultores, primer y segundo día (1998-99).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO (*)**
a	1.00	0.99	1.03	0.99	1.02
b	-1.98	-0.72	-75.03	331.7	229.57
R^2	0.86	0.89	0.82	0.99	0.83
Fc	266.1	170.7	50.5	2038.0	39.7
Ft	4.06	4.3	4.84	4.41	5.32
DISTRIBUCIÓN BETA					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO (*)**
a	1.00	0.98	1.03	0.99	1.02
b	-23.76	29.86	-128.4	241.27	-258.03
R^2	0.90	0.93	0.85	0.99	0.94
Fc	406.5	277.2	63.7	3497.9	132.0
Ft	4.06	4.3	4.84	4.41	5.32

(*) Habiendo eliminado cuatro observaciones por ser incoherentes.

Obteniéndose buenos ajustes, ligeramente mejores con la distribución beta.

Las mismas regresiones, forzando a que b = 0, dan los resultados resumidos en la tabla 110.

Tabla 110.

Regresión $R_m = a\ RMT + b$, *con* $b = 0$. (Estimación directa R_m del segundo día y calculados a partir de estimaciones R_{min}, R_{mf}, R_{max} del primer día). Encuesta a los agricultores, primer y segundo día (1998-99).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO (*)**
a	1.00	0.99	1.01	1.01	1.03
b	0	0	0	0	0
R^2	0.98	0.99	0.99	0.99	0.99
Fc	2983.1	1693.9	1834.6	5676.6	1621.2

Ft	4.06	4.28	4.75	4.38	5.12
DISTRIBUCIÓN BETA					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO (*)**
a	0.99	0.99	1.01	1.01	1.01
b	0	0	0	0	0
R^2	0.99	0.99	0.99	0.99	0.99
Fc	4352.4	2639.8	2224.6	9773.2	4757.3
Ft	4.06	4.28	4.75	4.38	5.12

(*) Habiendo eliminado cuatro observaciones por ser incoherentes.

#7.3.3.- VALORES DE APROXIMACIONES FUNCIONALES (DÍAS 1 Y 2)

Para los valores (R_{min}, R_{mf}, R_{max}) obtenidos cada día (pregunta 2 del primer día y pregunta 9 del segundo día), se ha procedido a calcular las diferencias relativas obtenidas de las estimaciones de valores medios mediante la utilización de las aproximaciones funcionales (tabla 111). Concretamente se han calculado los valores absolutos de las estimaciones triangular | RMT1 – RMT2): RMT2| y beta | RMB1 – RMB2): RMB2, siendo RMTi la aproximación triangular con datos del dia i y RMBi la correspondiente con la beta (i = 1, 2).

Tabla 111.

Valores de |(RMT1 – RMT2):RMT2| y | (RMB1 – RMB2):RMB2|. Encuesta a los agricultores, primer y segundo día (1998-99).

TRIANGULAR	**Número de casos**											
Cultivo	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**>0.5**	**Mediana**
Cebada secano	29	7	6	2	1	1						0.040
Trigo secano	16	3	2	3								0.038
Trigo regadío	9	2		1	1							0.036
Maíz regadío	13	4	1	1			1					0.038
Alfalfa regadío	8	1	1								4	0.044
BETA	**Número de casos**											
Cultivo	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**>0.5**	**Mediana**
Cebada secano	32	6	4	1	2	1						0.036
Trigo secano	16	4	1	2	1							0.038
Trigo regadío	9	2		2								0.036
Maíz regadío	14	2	2		1	1						0.036
Alfalfa regadío	9			1		1					3	0.039

Las diferencias relativas obtenidas mediante la aproximación a una función beta-PERT son menores o iguales a las obtenidas en el caso de la aproximación triangular, en todos los cultivos. En caso de la alfalfa de regadío, se observa una cierta dispersión, ocasionada por 4 observaciones que de ahora en adelante serán eliminadas por considerarse “outliers”.

La comparación entre las poblaciones de valores medios para las aproximaciones a las distribuciones triangular y beta en los dos días se resume en la tabla 112.

Tabla 112.

Comparación entre poblaciones de valores medios (RMT1 con RMT2, y RMB1 con RMB2). Encuesta a los agricultores, primer y segundo día (1998-99).

Comparación con la distribución triangular			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.992	No	No diferencias
Trigo secano	0.969	No	No diferencias
Trigo regadío	0.922	No	No diferencias
Maíz regadío	0.839	No	No diferencias
Alfalfa regadío (*)	0.948	No	No diferencias
Comparación con la distribución beta			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.959	No	No diferencias
Trigo secano	0.968	No	No diferencias
Trigo regadío	0.901	No	No diferencias
Maíz regadío	0.867	No	No diferencias
Alfalfa regadío (*)	0.912	No	No diferencias

(*) Habiendo eliminado cuatro observaciones por ser incoherentes.

Se obtienen coeficientes de correlación altos y dependencia en todas las poblaciones comparadas.

La **regresión** entre los valores de rendimientos medios calculados los días 1 y 2 mediante las fórmulas de las distribuciones triangular (*RMT1* y *RMT2*, respectivamente) y beta (*RMB1* y *RMB2*, respectivamente), se han reproducido en la tabla 113.

Tabla 113.

Regresión *RMT1 = a RMT2 + b*. Encuesta a los agricultores, primer y segundo día (1998-99).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO (*)**
a	0.99	0.97	0.89	0.99	1.06
b	13.15	39.11	605.11	-49.66	-1145.00
R^2	0.93	0.95	0.82	0.99	0.97
Fc	621.5	446.4	51.5	1596.9	264.7
Ft	4.06	4.3	4.84	4.41	5.32
DISTRIBUCIÓN BETA					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO (*)**
a	0.99	0.99	0.93	1.00	1.12
b	1.66	6.81	394.96	-500.81	-1906.9
R^2	0.94	0.95	0.84	0.99	0.94
Fc	743.0	449.9	56.3	1655.2	116.7
Ft	4.06	4.3	4.84	4.41	5.32

(*) Habiendo eliminado cuatro observaciones por ser incoherentes.

Obteniéndose unos buenos ajustes en todos los casos, ligeramente superiores en general para la distribución beta.

En caso de que se fuerce *b* = 0, se obtienen los resultados de la tabla 114.

Tabla 114.

Regresión *RMT1 = a RMT2* (*b* = 0) *RMB1 = a RMB2* (*b* = 0). Encuesta a los agricultores, primer y segundo día (1998-99).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO (*)**
a	0.99	0.98	1.01	0.98	0.99
b	0	0	0	0	0
R^2	0.99	0.99	0.99	0.99	0.99
Fc	7422.5	4615.9	2151.2	4242.3	9184.4
Ft	4.06	4.28	4.75	4.38	5.12
DISTRIBUCIÓN BETA					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO (*)**
a	0.99	0.99	1.00	0.98	0.99
b	0	0	0	0	0
R^2	0.99	0.99	0.99	0.99	0.99
Fc	8499.5	4407.4	2379.2	4345.7	3919.7
Ft	4.06	4.28	4.75	4.38	5.12

(*) Habiendo eliminado cuatro observaciones por ser incoherentes.

#7.3.4.- VARIANZAS

Al disponer de estimaciones de la forma (R_{min}, R_{mf}, R_{max}) para los dos días (preguntas 2 y 9 del primer y del segundo día, respectivamente), es posible comprobar la relación entre las varianzas estimadas mediante la simplificación triangular y la beta – PERT.

Como es conocido, la varianza de la función triangular viene dada por:

$$\left(VT = \frac{1}{18}\left[\left(R_{max} - R_{min}\right)^2 + \left(R_{mf} - R_{min}\right)\left(R_{mf} - R_{max}\right) \right] \right)$$

y la aproximación PERT a la beta por VB = (1/36) $(R_{max} - R_{min})^2$.

La varianza triangular del primer día se denotará VT1, y la del segundo día como VT2. Las varianzas beta se denominarán respectivamente VB1 y VB2.

En la tabla 115 se han resumido los resultados de las comparaciones |(VT1 – VT2) : VT2| y |(VB1 – VB2) : VB2|.

Tabla 115.

Diferencias relativas entre varianzas calculadas |(VT1 – VT2) : VT2| y |(VB1 – VB2) : VB2|. Encuesta a los agricultores, primer y segundo día (1998-99).

TRIANGULAR	**Número de casos**											
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**	**>1**	**Mediana**
Cebada secano	20	3	4	1	4	3	2	2	1	2	4	0.200
Trigo secano	8	1	4	4	2			3	1		1	0.275
Trigo regadío	7	2	1	1	2							0.093
Maíz regadío	11	4	1		1	1	1				1	0.110

Alfalfa regadío	6	1		2	1		1			2	1	0.200
BETA	**Número de casos**											
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**	**>1**	**Mediana**
Cebada secano	19	3	5	1	3	4	4			2	5	0.220
Trigo secano	8	1	3	3	4			3	1		1	0.300
Trigo regadío	7	2	1	2	1							0.093
Maíz regadío	12	3		1	2		1				1	0.083
Alfalfa regadío	6	1		2			1			3	1	0.200

En general, las estimaciones de la varianza realizadas por la aproximación triangular y la beta – PERT son muy semejantes. En bastantes casos el 50% de las medidas indican diferencias en las estimaciones del orden del 20 – 30%.

En la tabla 116 se han comparado las poblaciones de varianzas.

Tabla 116.

Comparación de las poblaciones de varianzas calculadas |(VT1 – VT2) : VT2| y |(VB1 – VB2) : VB2|. Encuesta a los agricultores, primer y segundo día (1998-99).

Comparación con la distribución triangular			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.875	No	No diferencias
Trigo secano	0.891	No	No diferencias
Trigo regadío	0.988	No	Diferencias
Maíz regadío	0.830	No	No diferencias
Alfalfa regadío (*)	0.988	No	No diferencias
Comparación con la distribución beta			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.876	No	No diferencias
Trigo secano	0.878	No	No diferencias
Trigo regadío	0.993	No	No diferencias
Maíz regadío	0.829	No	No diferencias
Alfalfa regadío (*)	0.997	No	Diferencias (**)

(*) Habiendo eliminado cuatro observaciones por ser incoherentes.
(**) En este caso, sin eliminar las cuatro observaciones incoherentes no se obtienen diferencias.

En general para todos los cultivos muestran altos coeficientes de correlación, existiendo dependencia en las poblaciones estudiadas.

El análisis anterior se completa con el estudio de la regresión entre los valores estudiados, los resultados del cual se resumen en la tabla 117.

Tabla 117.

Regresiones *VT1 = a VT2 + b y VB1 = a VB2 +b*. Encuesta a los agricultores, primer y segundo día (1998-99).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO (*)**
a	0.90	0.93	1.01	0.99	0.96
b	73749.58	44208.48	-72429.51	134463.11	372519.55

R^2	0.81	0.82	0.96	0.99	0.99
Fc	184.9	98.7	284.4	4563.4	1052.1
Ft	4.06	4.3	4.84	4.41	5.32
DISTRIBUCIÓN BETA					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO (*)**
a	0.90	0.93	1.00	0.99	0.96
b	49330.24	26462.44	-46447.48	81282.92	236784.40
R^2	0.81	0.81	0.96	0.99	0.99
Fc	191.1	93.6	272.2	4947.1	1146.0
Ft	4.06	4.3	4.84	4.41	5.32

(*) Habiendo eliminado cuatro observaciones por ser incoherentes.

Cuando se fuerza que b = 0, los resultados de las regresiones se resumen en la tabla 118.

Tabla 118.

Regresiones *VT1 = a VT2 y VB1 = a VB2.* Encuesta a los agricultores, primer y segundo día (1998-99).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO (*)**
a	0.97	0.97	0.97	1.00	1.00
b	0	0	0	0	0
R^2	0.92	0.92	0.98	0.99	0.99
Fc	482.4	284.4	614.3	5592.4	2405.8
Ft	4.06	4.28	4.75	4.38	5.12
DISTRIBUCIÓN BETA					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO (*)**
a	0.98	0.96	0.96	1.00	1.00
b	0	0	0	0	0
R^2	0.92	0.92	0.98	0.99	0.99
Fc	489.5	267.8	588.2	6055.8	2598.6
Ft	4.06	4.28	4.75	4.38	5.12

(*) Habiendo eliminado cuatro observaciones por ser incoherentes.

#7.4.- FUNCIONES DE DENSIDAD DE RENDIMIENTOS CON INTERVALOS ORDINALES

En la **Pregunta 4** se pedía al agricultor que definiera el porcentaje en que, en su opinión, se presentan cada uno de los casos de rendimientos definidos en lenguaje natural (*muy malos*, *malos*, *normales*, *buenos* y *muy buenos*), a fin de obtener una primera aproximación a la función de densidad, con las abscisas definidas en lenguaje natural. En las respuestas, como en el caso de la encuesta de los expertos, se dieron casos de reducción de intervalos (fenómeno discutido en ζ 4. 2. 3), cuya importancia se resume en la tabla 119.

Tabla 119.

Forma de las funciones de densidad (intervalos ordinarios), primer día. Encuesta a los agricultores (1998 – 99).

Cultivo	**Reducción de intervalo**			**Uniforme**	**Extremos < 15%**	**Simetría**	**Inter 1 = Inter 5**	**Inter 2 = Inter 4**	**Total**
	I	**A**	**D**						

Cebada secano	1	1	1	1	9	12	18	12	46
Trigo secano	1	1	0	0	5	4	11	5	25
Trigo regadío	1	1	1	1	5	3	5	3	13
Maíz regadío	5	3	2	2	11	7	6	6	20
Alfalfa regadío	3	3	0	0	7	3	6	4	17

En el caso de la cebada de secano, en 3 casos de 46 se redujo la amplitud del intervalo, y en un caso se dan frecuencias uniformes a los cinco intervalos. Dieciocho respuestas de 46 igualan las frecuencias de los intervalos extremos, y de ellas 12 se corresponden con distribuciones simétricas. Mayoritariamente (34 de 46 casos, el 74% de los mismos) los agricultores no atribuyen simetría a las distribuciones, y en la mayor parte de los casos incluso discriminan la frecuencia individual de cada intervalo.

En el caso del trigo de secano, en 2 casos de 25 se redujo la amplitud del intervalo, y en ningún caso se dan frecuencias uniformes a los cinco intervalos. Once respuestas de 25 igualan las frecuencias de los intervalos extremos, y de ellas 4 se corresponden con distribuciones simétricas. Igual que para la cebada de secano, mayoritariamente (21 de 25 casos o 84% de los mismos) los agricultores no atribuyen simetría a las distribuciones, y en la mayor parte de los casos incluso discriminan la frecuencia individual de cada intervalo.

En trigo y maíz de regadío, una misma encuesta asignó las frecuencias reduciendo el intervalo (ambos extremos) y atribuyendo valores uniformes en el resto de intervalos.

En el caso del trigo de regadío, en 3 casos de 13 se redujo la amplitud del intervalo, y en un caso se presenta simultáneamente reducción de intervalo y distribución uniforme. Cinco respuestas de 13 igualaron las frecuencias de los intervalos extremos, y de ellas 3 se corresponden con distribuciones simétricas. Igual que para los cultivos anteriores, mayoritariamente (10 de 13 casos o 77% de los mismos) los agricultores no atribuyen simetría a las distribuciones, y en la mayor parte de los casos incluso discriminan la frecuencia individual de cada intervalo.

En el caso del maíz de regadío, en 10 casos de 20 se redujo la amplitud del intervalo, y en dos se presenta simultáneamente reducción de intervalo y distribución uniforme. Seis respuestas de 20 igualan las frecuencias de los intervalos extremos, y 7 son distribuciones simétricas. Aunque en menor grado que en los cultivos anteriores, mayoritariamente (13 de 20 casos o 65% de los mismos) los agricultores no atribuyen simetría a las distribuciones.

En el caso de la alfalfa de regadío, en 6 de los 17 casos se reduce la amplitud del intervalo, y en ningún caso se dan frecuencias uniformes. Seis respuestas de 17 igualan las frecuencias de los intervalos extremos, y 3 son distribuciones simétricas. De la misma forma que para los cultivos anteriores, mayoritariamente (14 de 17 casos o 82% de los mismos) los agricultores no atribuyen simetría a las distribuciones.

Se puede observar que, porcentualmente, la reducción de intervalos es más importante en los cultivos de regadío que en los de secano, señalando una mayor dificultad para aceptar términos como “rendimientos muy malos” y / o “rendimientos muy buenos” como descriptores de los rendimientos físicos.

Determinación de la moda. Los intervalos que consideran los agricultores más frecuentes, se reproducen en la tabla 120, donde el intervalo MM se refiere al de rendimientos muy malos, el M a rendimientos malos, el N a normales, el B a buenos y el MB a muy buenos.

Tabla 120.

Número y posición de las modas (intervalos ordinales), primer día. Encuesta a los agricultores (1998 – 99).

Cultivo	**1 Moda**	**2 Modas**	**Moda en 2 intervalos**	**Moda en intervalo**				
				MM	**M**	**N**	**B**	**MB**
Cebada secano	43	2	7	1	7	33	8	1
Trigo secano	22	2	3	1	2	16	7	1
Trigo regadío	12	0	0	0	2	8	4	1
Maíz regadío	16	0	2	0	2	17	5	2
Alfalfa regadío	12	0	5	0	0	13	7	2

De los 46 casos examinados en la cebada de secano, en uno de ellos la distribución es uniforme, y no se ha asignado la moda a ningún intervalo. En los 7 casos en los que se presentan modas en 2 intervalos, contiguos o no, se ha asignado la moda a cada uno de los intervalos, por lo que la suma de casos supera el número de 45 respuestas. Mayoritariamente se considera el intervalo central como más frecuente, y cuando no lo es éste, lo son los dos intervalos contiguos al mismo.

De los 25 casos examinados en el trigo de secano, en 3 se presentan modas en dos intervalos (contiguos o no), de los cuales dos casos corresponden a distribuciones bimodales. Mayoritariamente se considera el intervalo central como más frecuente, y cuando no lo es éste, lo es el intervalo derecho contiguo al mismo.

De los 13 casos examinados en trigo de regadío, en uno de ellos la distribución es uniforme, no asignándose la moda a ningún intervalo. No se presentan modas en más de un intervalo, y mayoritariamente se considera el intervalo central como más frecuente, y cuando no lo es éste, lo es principalmente el contiguo derecho y en segundo lugar el contiguo izquierdo.

En maíz de regadío, en dos de los 20 casos examinados la distribución es uniforme, no asignándose la moda a ningún intervalo. En dos casos se presentan modas en dos intervalos contiguos, y en ninguno distribuciones bimodales. Mayoritariamente se considera el intervalo central como más frecuente, y cuando no lo es éste, lo es el contiguo derecho, y en segundo lugar el contiguo izquierdo.

En alfalfa de regadío, en 5 de los 17 casos examinados se presentan modas en dos intervalos, y no se observa ninguna distribución bimodal. Mayoritariamente se considera el intervalo central como más frecuente, y cuando no lo es éste lo es su contiguo derecho.

Determinación de la mediana. A fin de determinar la simetría en la distribución de las masas de frecuencia, se ha determinado la mediana en base a los 5 intervalos (tabla 121). Los intervalos se han numerado de 1 a 5, siendo el intervalo 1 el extremo inferior, y el 5 el superior. Así, una mediana de valor 2,5 indica que se encuentra en la mitad del intervalo 3, es decir, en el centro de la distribución en 5 intervalos, dejando a su izquierda dos intervalos y medio.

Tabla 121.

Valor de la mediana (expresada en función del número de intervalo), primer día. Encuesta a los agricultores (1998 – 99).

Cultivo	**Valor de la mediana (expresada en número del intervalo)**									
	0-0.5	0.5-1	1-1.5	1.5-2	2-2.5	2.5-3	3-3.5	3.5-4	4-4.5	4.5-5
Cebada de secano			2	3	24	13	3	1		
Trigo de secano			1		8	12	3	1		
Trigo de regadío				1	4	6	2			
Maíz de regadío				1	8	7	3	1		

Alfalfa regadío					3	10	3	1		

En la cebada de secano, el mayor número de casos (33 de 46) presenta su moda en el intervalo central (el tercer intervalo). En ese intervalo central, a la izquierda de su punto medio se encuentran 12 observaciones, otras 12 en el valor 2,5 y 13 observaciones a la derecha del punto medio. De esta forma, la mediana se encontraría a la izquierda del punto medio del recorrido en 17 casos de 46, coincidiendo con ese punto en 12 ocasiones, y a la derecha del mismo en 17 casos. El valor de la mediana indica una tendencia mayoritaria a asignar masas de probabilidad relativamente centradas, pero sin que el centro del recorrido concentre más de 1/3 de los casos.

Para el trigo de secano, el mayor número de casos (16 de 25) presenta su moda en el intervalo central. En ese intervalo, a la izquierda de su punto medio se encuentran 4 observaciones, otras 4 en el valor 2,5 y 12 a la derecha del punto medio. De esta forma, la mediana se encontraría a la izquierda del punto medio del recorrido en 5 casos de 25, coincidiendo con ese punto en 4 ocasiones, y a la derecha del mismo en 16 casos. El valor de la mediana indica una tendencia mayoritaria a asignar masas de probabilidad relativamente centradas, aunque la parte derecha del intervalo central tiene un mayor peso y en conjunto la distribución se decanta hacia la derecha.

Para el trigo de regadío, el mayor número de casos (8 de 13) presenta su moda en el intervalo central. En ese intervalo, a la izquierda de su punto medio se encuentra 1 observación, otras 3 en el valor 2.5 y 6 a la derecha del punto medio. De esta forma, la mediana se encontraría a la izquierda del punto medio del recorrido en 2 casos de 13, coincidiendo con ese punto en 3 ocasiones, y a la derecha del mismo en 8 casos. El valor de la mediana indica una tendencia mayoritaria a asignar masas de probabilidad relativamente centradas, aunque la parte derecha del intervalo central tiene un mayor peso y en conjunto la distribución se decanta hacia la derecha.

Para el maíz de regadío, el mayor número de casos (17 de 20 casos) presenta su moda en el intervalo central. En ese intervalo, a la izquierda de su punto medio se encuentra una observación, 7 en el valor 2.5 y 7 a la derecha del punto medio. De esta forma, la mediana se encontraría a la izquierda del punto medio del recorrido en 2 casos de 20, coincidiendo con ese punto en 7 ocasiones, y a la derecha del mismo en 11 casos. El valor de la mediana indica una tendencia mayoritaria a asignar masas de probabilidad relativamente centradas, aunque se observa una tendencia clara hacia los valores situados a la derecha del punto medio del recorrido de los rendimientos.

Para la alfalfa de regadío, el mayor número de casos (13 de 17) presenta su moda en el intervalo central. En ese intervalo, a la izquierda de su punto medio no se presenta ninguna observación, presentándose 3 para el valor 2.5 y 10 a su derecha. De esta forma, la mediana no se encontraría a la izquierda del recorrido en ningún caso, coincidiría con el punto medio en 3 casos y estaría a su derecha en 14 casos.

#7.4.1- FRECUENCIAS EN EL SEGUNDO DÍA (ESCALA ORDINAL)

En el cuestionario del segundo día se introdujo la misma pregunta (también con el **número 4**). El número de reducciones de intervalos y el estudio de simetría de las respuestas obtenidas se resume en la tabla 122.

Tabla 122.

Forma de las funciones de densidad (intervalos ordinales), segundo día. Encuesta a los agricultores (1998 – 99).

	Reducción de intervalo			**Uniforme**	**Extremos < 15%**	**Simetría**	**Inter 1 = Inter 5**	**Inter 2 = Inter 4**	**Total**
	I	**A**	**D**						
Cebada secano	1	1	1	0	15	12	20	12	46
Trigo secano	0	1	0	0	5	7	9	5	23
Trigo regadío	2	1	2	0	6	4	4	3	13
Maíz regadío	5	2	1	0	11	6	6	7	20
Alfalfa regadío (*)	4	2	0	0	8	5	7	5	16

(*) Habiendo eliminado una observación por sumar más de 100 %.

En el caso de la cebada de secano, en 3 casos de 46 se reduce la amplitud del intervalo, y en ningún caso se dan frecuencias uniformes a los cinco intervalos. 20 respuestas de 46 igualan las frecuencias de los intervalos extremos, y de ellas 12 se corresponden con distribuciones simétricas. Mayoritariamente (34 de 46 casos, el 74% de los mismos) los agricultores no atribuyen simetría a las distribuciones, y en la mayor parte de los casos incluso discriminan la frecuencia individual de cada intervalo.

En el caso del trigo de secano, en 1 caso de 23 se reduce la amplitud del intervalo, y en ningún caso se dan frecuencias uniformes a los cinco intervalos. Nueve respuestas de 23 igualan las frecuencias de los intervalos extremos, y de ellas 7 se corresponden con distribuciones simétricas. Igual que para la cebada de secano, mayoritariamente (16 de 23 casos o 70%) los agricultores no atribuyen simetría a las distribuciones, y en la mayor parte de los casos incluso discriminan la frecuencia individual de cada intervalo.

En el caso del trigo de regadío, en 5 casos de 13 se reduce la amplitud del intervalo. Cuatro respuestas de 13 igualan las frecuencias de los intervalos extremos, correspondiéndose con distribuciones simétricas. Igual que para los cultivos anteriores, mayoritariamente (9 de 13 casos o 69%) los agricultores no atribuyen simetría a las distribuciones, y en la mayor parte de los casos incluso discriminan la frecuencia individual de cada intervalo.

En el caso del maíz de regadío, en 8 casos de 20 se reduce la amplitud del intervalo. Seis respuestas de 20 igualan las frecuencias de los intervalos extremos, y 6 son distribuciones simétricas. Igual que en los cultivos anteriores, mayoritariamente (14 de 20 casos o 70%) los agricultores no atribuyen simetría a las distribuciones.

En el caso de la alfalfa de regadío, en 6 de los 16 casos se reduce la amplitud del intervalo, y en ningún caso se dan frecuencias uniformes. Cinco respuestas de 16 igualan las frecuencias de los intervalos extremos, y 8 son distribuciones simétricas. De la misma forma que para los cultivos anteriores, (12 de 17 casos o 70%) los agricultores no atribuyen simetría a las distribuciones.

Determinación de la moda. Los intervalos que consideran los agricultores más frecuentes, se reproducen en la tabla 123, donde se indica: el número de casos en que señalan un solo intervalo como más frecuente (1 moda), el número de casos en que identifican dos modas (dos intervalos no contiguos con las mayores frecuencias), y el número de casos en los que dos intervalos contiguos tienen la misma frecuencia, y ésta es la mayor de todas ellas. Por último se da el número de casos en los que la moda se asigna a cada uno de los intervalos (siendo el intervalo *MM* el de los rendimientos *"muy malos"*, el intervalo *M* el de los rendimientos *"malos"*, el *N* el de los rendimientos *"normales"*, el *B* el de los rendimientos *"buenos"*, y finalmente, el *MB* el de los rendimientos *"muy buenos"*).

Tabla 123.

Número y posición de las modas (intervalos ordinales), segundo día. Encuesta a los agricultores (1998 – 99).

Cultivo	1 Moda	2 Modas	Moda en 2 o más intervalos	Moda en intervalo				
				MM	M	N	B	MB
Cebada secano	41	2	3	3	6	36	5	1
Trigo secano	21	2	0	1	2	17	4	1
Trigo regadío	12	0	1	0	1	10	3	1
Maíz regadío	18	0	2	0	2	16	4	1
Alfalfa regadío	13	0	3	0	0	11	6	2

En cebada de secano, en los 3 casos en los que se presentan modas en 2 intervalos contiguos o no, se ha asignado la moda a cada uno de los intervalos, por lo que la suma de casos supera el número de 46 respuestas. Mayoritariamente se considera el intervalo central como más frecuente, y cuando no lo es éste, lo son los dos intervalos contiguos al mismo.

De los 23 casos examinados en trigo de secano, en 2 se presentan modas en dos intervalos (contiguos o no). Mayoritariamente se considera el intervalo central como más frecuente, y cuando no lo es éste, lo son los dos intervalos contiguos al mismo.

De los 13 casos examinados en trigo de regadío, en 1 caso se presentan modas en tres intervalos, y mayoritariamente se considera el intervalo central como más frecuente, y cuando no lo es éste, lo es principalmente el contiguo derecho.

En maíz de regadío en un caso se presentan modas en dos intervalos contiguos y en otro en tres intervalos (precisamente el mismo agricultor que distribuyó también la moda en tres intervalos en el trigo de regadío), y en ninguno distribuciones bimodales. Mayoritariamente se considera el intervalo central como más frecuente, y cuando no lo es éste, lo son los dos intervalos contiguos al mismo.

En alfalfa de regadío, en 3 de los 16 casos examinados se presentan modas en dos intervalos, y no se observa ninguna distribución bimodal. Mayoritariamente se considera el intervalo central como más frecuente, y cuando no lo es éste lo es su contiguo derecho.

Determinación de la mediana. A fin de determinar la simetría en la distribución de las masas de frecuencia, se ha determinado la mediana basándose en 5 intervalos (Tabla 124).

Tabla 124.

Valor de la mediana (relativa a intervalos ordinales), segundo día. Encuesta a los agricultores (1998 – 99).

Cultivo	Valor de la mediana (expresada en número del intervalo)									
	0-0.5	0.5-1	1-1.5	1.5-2	2-2.5	2.5-3	3-3.5	3.5-4	4-4.5	4.5-5
Cebada de secano			1	4	22	16	3			
Trigo de secano			1		12	7	2	1		
Trigo de regadío					5	6	1	1		
Maíz de regadío				1	9	7	2	1		
Alfalfa regadío					5	6	4	1		

Para la cebada de secano, el mayor número de casos (36 de 46) presenta su moda en el intervalo central (el tercer intervalo). En ese intervalo central, a la izquierda de su punto medio se encuentran 10 observaciones, otras 12 en el valor 2.5 y 16 observaciones a su derecha. De esta forma, la mediana se encontraría a la izquierda del punto medio del recorrido en 15 casos de 46, coincidiendo con ese punto en 12 ocasiones, y a su derecha en 19 casos. El valor de la mediana indica una tendencia mayoritaria a asignar masas de probabilidad relativamente centradas, pero sin que el punto medio del recorrido recoja más de un tercio de los casos.

Para el trigo de secano, el mayor número de casos (17 de 23) presenta su moda en el intervalo central. En ese intervalo, a la izquierda de su punto medio se encuentran 5 observaciones, otras 7 en el valor 2.5 y 7 a su derecha. De esta forma, la mediana se encontraría a la izquierda del punto medio del recorrido en 6 casos de 23, coincidiendo con ese punto en 7 ocasiones, y a su derecha en 10 casos. El valor de la mediana indica una tendencia mayoritaria a asignar masas de probabilidad relativamente centradas, aunque la parte derecha del intervalo central tiene un mayor peso y en conjunto la distribución se decanta hacia la derecha.

Para el trigo de regadío, el mayor número de casos (10 de 13) presenta su moda en el intervalo central. En ese intervalo, a la izquierda de su punto medio se encuentra 1 observación, otras 4 en el valor 2.5 y 6 a su derecha. De esta forma, la mediana se encontraría a la izquierda del punto medio del recorrido en 1 caso de 13, coincidiendo con ese punto en 4 ocasiones, y a su derecha en 8 casos. El valor de la mediana indica una tendencia mayoritaria a asignar masas de probabilidad relativamente centradas, aunque la parte derecha del intervalo central tiene un mayor peso y en conjunto la distribución se decanta hacia la derecha.

Para el maíz de regadío, el mayor número de casos (16 de 20 casos) presenta su moda en el intervalo central. En ese intervalo, a la izquierda de su punto medio se encuentran 3 observaciones, 6 en el valor 2.5 y 8 a su derecha. De esta forma, la mediana se encontraría a la izquierda del punto medio del recorrido en 4 casos de 20, coincidiendo con ese punto en 6 ocasiones, y a su derecha en 10 casos. El valor de la mediana indica una tendencia mayoritaria a asignar masas de probabilidad relativamente centradas, aunque se observa una tendencia clara hacia los valores altos de rendimientos, cercanos al centro del recorrido.

Para la alfalfa de regadío, el mayor número de casos (11 de 16) presenta su moda en el intervalo central. En ese intervalo, a la izquierda de su punto medio no se presenta ninguna observación, presentándose 5 para el valor 2.5 y 6 a su derecha. De esta forma, la mediana no se encontraría a la izquierda del recorrido en ningún caso, coincidiría con el punto medio en 5 casos y estaría a su derecha en 11 casos. El valor de la mediana indica una tendencia mayoritaria a asignar masas de probabilidad relativamente centradas, aunque se observa una tendencia clara hacia los valores altos de rendimientos, cercanos al centro del recorrido.

#7.5.- FUNCIONES DE DENSIDAD DE RENDIMIENTOS CON INTERVALOS NUMÉRICOS

El anterior análisis se ha realizado considerando qué frecuencia se ha asignado a cada intervalo definido en lenguaje natural en la **pregunta 4**, tanto del primer como del segundo día. Se dispone de una correspondencia entre los intervalos ordinales y los intervalos numéricos, a través de la **pregunta 3**, en ambos días. En lo que sigue se analiza la asignación de frecuencias, hecha sobre los intervalos ordinales, y traducidos a los numéricos.

Como se recordará del caso estudiado de la encuesta a los expertos (ζ 5), donde se utilizó la misma técnica, las respuestas de los sujetos (en este caso los agricultores) ha seguido una casuística, discutida en ζ 4.2.1, que obliga a distribuirlas por clases, en función de la definición del recorrido y de los diferentes intervalos. En la tabla 125 se resume, por clases, el número de respuestas obtenidas el primer día. Para el segundo día se refleja la misma información en la tabla 134.

Tabla 125.

Clases de respuestas obtenidas (encuesta primer día). Encuesta a los agricultores (1998 – 99)

Cultivo	**Clase I**	**Clase II**			**Clase III**	**Clase IV**			**Total**
		II	**IIa**	**IIb**		**IV**	**IVa**	**IVb**	

Cebada secano	19	5	0	6	14	1	1	0	46
Trigo secano	13	3	0	2	5	2	0	0	25
Trigo regadío	5	0	2	1	5	0	0	0	13
Maíz regadío	6	3	1	2	8	0	0	0	20
Alfalfa regadío	5	1	1	2	8	0	0	0	17

Las respuestas de la Clase I son las más seguras, en el sentido de que la amplitud del recorrido y de cada intervalo han sido definidas por los agricultores. En el resto de respuestas se estiman los recorridos, y para las clases III y IV los límites de los intervalos, a partir de los supuestos discutidos en ζ 4.2.1. Las estimaciones de la Clase II son bastante razonables, y más discutibles las del resto de las clases. Debido a ello se analizan por separado las Clases I y II, y después conjuntamente esas dos clases y, finalmente, el conjunto de las clases (en todos los casos, con las respuestas para las que se han podido estimar los recorridos). La discusión del proceso de estimación de los recorridos para las clases analizadas en este apartado se realiza en ζ 7.6.

En la tabla 126 se ha indicado en cuántos casos la moda se encuentra a la izquierda del punto medio del recorrido (leyenda *I*), centrada o coincidiendo con el punto medio del recorrido (leyenda *C*), o a su derecha (leyenda *D*). También se reflejan los casos en los que la distribución es uniforme (leyenda *Un*), los casos en que se presentan dos modas no contiguas (leyenda *Bi*) y el total de casos examinados (de la clase de respuestas I a la pregunta número 3), y la posición relativa de la moda (explicada en ζ 4.3.3).

Tabla 126.

Posición de la moda (relativa a intervalos numéricos). Respuestas de la Clase I. Encuesta a los agricultores (1998-99, primer día).

Cultivo	**I**	**C**	**D**	**Un**	**Bi**	**Total**	**Posición relativa de la moda**							
							≤ -2	**-2- -1**	**-1- 0**	**0- 1**	**1- 2**	**2- 3**	**3- 4**	**4- 5**
Cebada secano	6	2	8	1	2	19		3	5	4	1	1	2	
Trigo secano	3	1	7	0	2	13			4	3		2		2
Trigo regadío	1	0	3	1	0	5	1			1	1			1
Maíz regadío	2	1	2	1	0	6		1	2	1		1		
Alfalfa regadío	1	0	4	0	0	5		1		4				

En la cebada de secano se observa que de los 19 casos disponibles, 16 de ellos presentan una única moda. Esa moda se encuentra a la izquierda o centrada en la mitad de los casos, y a la derecha en la otra mitad. Individualmente, por tanto, son minoría los agricultores que presentan funciones con la moda centrada, pero en conjunto las modas se distribuyen de forma relativamente simétrica con respecto al punto medio del recorrido.

Cuando las modas se han señalado a la izquierda, en general no se encuentran muy alejadas del centro del recorrido, expresada esa distancia en términos de la amplitud del intervalo para el que se ha señalado la presencia de la moda. Las modas a la derecha se encuentran más dispersas, y tienden a alejarse del punto medio del recorrido, en algunos casos, a distancias relativamente altas.

El trigo de secano presenta una mayor importancia relativa de las modas situadas a la derecha del punto medio del recorrido. El trigo de regadío, con pocos datos, presenta un perfil muy abierto, pero equilibrado, anunciando la transición observada en los cultivos de regadío, donde predominan de nuevo las modas centradas, con cierta tendencia a ser más numerosas a la izquierda del punto medio del intervalo.

La **mediana** se expresa en la tabla 127 por su posición relativa normalizada en recorridos de amplitud unidad.

Tabla 127.

Posición relativa de la mediana. Clase de respuesta I. Encuesta a los agricultores (1998 – 99, primer día).

Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano	1		1	2	5	4	4	2		
Trigo secano				1	4	2	4	2		
Trigo regadío			1	1		2			1	
Maíz regadío				2	2	1	1			
Alfalfa regadío				1		1	2	1		

En el caso de la cebada de secano, se encuentran 9 casos de 19 en el intervalo [0.4, 0.6], y de ellos 4 a la izquierda del punto medio, uno coincidiendo con el punto medio, y otros 4 a la derecha de ese punto. La totalidad de los casos, excepto uno, se concentran en valores del intervalo [0.25, 0.75].

Para el trigo de secano, en 4 casos la mediana se encuentra a la izquierda del punto medio del recorrido (punto 0.5), y en los 4 con valores mayores de 0.25. En un caso la mediana coincide con el punto medio del recorrido, y los 7 casos restantes se encuentran comprendidos en el intervalo (0.5, 0.75]. De esta forma, excepto en un caso, la mediana se sitúa en valores del intervalo [0.25, 0.75]. En el intervalo [0.4, 0.6] se encuentran 6 de 13 casos, y de ellos 3 a la izquierda del punto medio, uno coincidiendo con él, y dos a su derecha.

En caso del trigo de regadío, en 2 casos la mediana se encuentra a la izquierda del punto medio del recorrido (punto 0.5), y en los dos con valores mayores de 0.25. En ningún caso la mediana coincide con el punto medio del recorrido, y de los 3 casos restantes, dos se encuentran comprendidos en el intervalo (0.5, 0.75]. De esta forma, excepto en un caso, la mediana se sitúa en valores del intervalo [0.25, 0.75]. En el intervalo [0.4, 0.6] se encuentran 2 de 5 casos, y ambos a la derecha del punto medio.

En maíz de regadío, en tres casos la mediana se encuentra a la izquierda del punto medio del recorrido (punto 0.5), y en los tres con valores mayores de 0.25. En un caso la mediana coincide con el punto medio del recorrido, y en los dos casos restantes se encuentra comprendida en el intervalo (0.5, 0.75]. De esta forma, en todos los casos la mediana se sitúa en valores del intervalo [0.25, 0.75]. En el intervalo [0.4, 0.6] se encuentran 3 casos, uno a la izquierda del punto medio, otro en el punto medio y otro a la derecha del punto medio.

En alfalfa de regadío, en un caso la mediana se encuentra a la izquierda del punto medio del recorrido (punto 0.5), y por encima de 0.25. En ningún caso la mediana coincide con el punto medio del recorrido, y en cuatro valores se encuentra dentro del intervalo (0.5, 0.75]. De esta forma, en todos los casos la mediana se sitúa en valores del intervalo [0.25, 0.75]. En el intervalo [0.4, 0.6] se encuentran 2 casos, uno a la izquierda del punto medio y otro a su derecha.

El análisis de la posición de la moda para las clases de respuestas de la Clase II que se han podido reconstruir (para las que se ha podido estimar un recorrido total coherente con el resto de los datos) se resume en la tabla 128. La tabla 129 resume el análisis de los casos de las clases I y II, y la tabla 130 el análisis conjunto de todas las clases.

Tabla 128.

Posición de la moda (relativa a intervalos numéricos). Respuestas de la Clase II. Encuesta a los agricultores (1998-99, primer día).

Cultivo	I	C	D	Un	Bi	Total	Posición relativa de la moda							
							≤ -2	-2- -1	-1- 0	0- 1	1- 2	2- 3	3- 4	4- 5
Cebada secano	7	0	4	0	0	11		3	4	2	1		1	
Trigo secano	2	0	3	0	0	5			2	3		2		2
Trigo regadío	2	0	1	0	0	3			2		1			
Maíz regadío	4	1	1	1	0	6	1	1	3	1				
Alfalfa regadío	0	0	4	0	0	4		0		4				

Tabla 129.

Posición de la moda (relativa a intervalos numéricos). Respuestas de las Clases I y II. Encuesta a los agricultores (1998-99, primer día).

Cultivo	I	C	D	Un	Bi	Total	Posición relativa de la moda							
							≤ -2	-2- -1	-1- 0	0- 1	1- 2	2- 3	3- 4	4- 5
Cebada secano	13	2	12	1	2	30		6	9	6	2	1	3	
Trigo secano	5	1	10	0	2	18			6	6		2		2
Trigo regadío	3	0	4	1	0	8	1		2	1	2			1
Maíz regadío	6	2	3	1	0	12	1	2	5	2		1		
Alfalfa regadío	1	0	4	0	0	9		1		8				

Tabla 130.

Posición de la moda (relativa a intervalos numéricos). Respuestas de las Clases I, II, III y IV. Encuesta a los agricultores (1998-99, primer día).

Cultivo	I	C	D	Un	Bi	Total	Posición relativa de la moda							
							≤ -2	-2- -1	-1- 0	0- 1	1- 2	2- 3	3- 4	4- 5
Cebada secano	19	4	20	1	2	46	2	9	12	10	3	4	3	
Trigo secano	8	5	10	0	2	25	1		8	6	2	3		3
Trigo regadío	3	1	8	1	0	13	1		3	2	4	1		1
Maíz regadío	9	2	8	1	0	20	2	2	7	5	1	2		
Alfalfa regadío	1	0	15	0	0	16		1		12	2			1

En lo que respecta a la **mediana**, se ha calculado su posición relativa para cada caso, habiendo normalizado los recorridos de tal forma que todos tengan una longitud unidad. El resultado se ha resumido en las tablas 131 a 133, donde se indica el número de casos en que se presenta un valor de la mediana comprendido en cada intervalo.

Tabla 131.

Posición relativa de la mediana. Clase de respuestas II. Encuesta a los agricultores (1998 – 99, primer día).

Cultivo	0-0.1	0.1-0.2	0.2-0.3	0.3-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1
Cebada secano				4	3	1	3			
Trigo secano					2	2	1			
Trigo regadío					2			1		
Maíz regadío				2	2	1	1			
Alfalfa regadío						3	1			

Tabla 132.

Posición relativa de la mediana. Clase de respuestas I y II. Encuesta a los agricultores (1998 – 99, primer día).

Cultivo	0-0.1	0.1-0.2	0.2-0.3	0.3-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1
Cebada secano	1		1	6	8	5	7	2		
Trigo secano				1	6	4	5	2		
Trigo regadío			1	1	2	2		1	1	
Maíz regadío				4	4	2	2			
Alfalfa regadío				1		4	3	1		

Tabla 133.

Posición relativa de la mediana. Clases de respuestas I, II, III y IV. Encuesta a los agricultores (1998 – 99, primer día).

Cultivo	0-0.1	0.1-0.2	0.2-0.3	0.3-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1
Cebada secano	1	1	3	7	13	10	8	3		
Trigo secano			1	1	7	4	9	3		
Trigo regadío			1	2	2	5	1	1	1	
Maíz regadío				5	5	7	3			
Alfalfa regadío				1		7	6	2		

#7.5.1.- RENDIMIENTOS CON INTERVALOS NUMÉRICOS (SEGUNDO DÍA)

En la tabla 134 se recoge el número de respuestas de las diferentes clases obtenidas con el cuestionario del segundo día.

Tabla 134.

Clases de respuestas obtenidas. Encuesta a los agricultores (1998 – 1999, segundo día).

Cultivo	Clase I	Clase II			Clase III	Clase IV			Total
		II	IIa	IIb		IV	IVa	IVb	

Cebada secano	12	7	0	4	22	1	0	0	46
Trigo secano	7	5	0	3	9	0	0	0	24
Trigo regadío	4	1	1	1	5	1	0	0	13
Maíz regadío	4	3	0	2	11	0	0	0	20
Alfalfa regadío	2	3	0	1	9	0	0	0	15

La posición de la moda en las respuestas de la Clase I se ha resumido en la tabla 135. El significado de los encabezamientos se puede consultar en el análisis anterior realizado para las respuestas del primer día.

Tabla 135.

Posición de la moda (relativa a intervalos numéricos). Respuestas de la Clase I. Encuesta a los agricultores (1998-99, segundo día).

Cultivo	**I**	**C**	**D**	**Un**	**Bi**	**Total**	**Posición relativa de la moda**								
							≤ -2	**-2 -1**	**-1 0**	**0 1**	**1 2**	**2 3**	**3 4**	**4 5**	**>5**
Cebada de secano	4	1	5	0	2	12		1	4	3	1				1
Trigo de secano	1	1	3	0	2	7			2	1	1				1
Trigo de regadío	2	0	2	0	0	4		1	1		1				1
Maíz de regadío	1	0	3	0	0	4			1	2	1				
Alfalfa de regadío	0	0	1	0	0	1				1					

En la cebada de secano se observa que de los 12 casos disponibles, 10 de ellos presentan una única moda. Los casos unimodales están distribuidos simétricamente. Individualmente, por tanto, son minoría los agricultores que presentan funciones con la moda centrada, pero en conjunto las modas se distribuyen de forma relativamente simétrica con respecto al punto medio del recorrido.

Cuando las modas se han señalado a la izquierda del centro del recorrido, en general no se encuentran muy alejadas de ese punto, expresada esa distancia en términos de la amplitud del intervalo para el que se ha señalado la presencia de la moda. Las modas señaladas a la derecha del centro del recorrido se encuentran más dispersas, y tienden a alejarse del punto medio, en algunos casos, a distancias relativamente altas.

En lo que respecta a la **mediana**, se ha resumido su posición en la tabla 136.

Tabla 136.

Posición relativa de la mediana. Clase de respuesta I. Encuesta a los agricultores (1998 – 99, segundo día).

Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano			1		5	3	2		1	

Trigo secano				1	1	2	2		1	
Trigo regadío				1	1	1		1		
Maíz regadío					1	3				
Alfalfa regadío								1		

En la cebada de secano, de las 12 observaciones que contienen respuestas de la clase I, 8 presentan sus medianas en puntos próximos al centro del recorrido, pero solamente una respuesta sitúa la mediana en ese punto. En el intervalo [0.4, 0.6] se encuentran 8 casos (de 12), y de ellos 4 a la izquierda del punto medio, uno coincidiendo con él, y otros tres a su derecha. Excepto en dos casos, la mediana se sitúa en valores del intervalo [0.25, 0.75].

Para el trigo de secano, en 2 casos la mediana se encuentra a la izquierda del punto medio del recorrido (punto 0.5), y en los dos con valores mayores de 0.25. En el intervalo [0.4, 0.6] se encuentran 3 de 7 casos, y de ellos 1 a la izquierda del punto medio y dos a su derecha. En ningún caso la mediana coincide con el punto medio del recorrido, y los 5 casos restantes se encuentran incluidos en el intervalo (0.5, 0.85]. De esta forma, excepto en un caso, la mediana se sitúa en valores del intervalo [0.25, 0.75].

En caso del trigo de regadío, en 2 casos la mediana se encuentra a la izquierda del punto medio del recorrido (punto 0.5), y en los dos con valores mayores de 0.25. En el intervalo [0.4, 0.6] se encuentran 2 de los 4 casos. En ningún caso la mediana coincide con el punto medio del recorrido, y los 2 casos restantes se encuentran incluidos en el intervalo (0.5, 0.75]. De esta forma, la mediana se sitúa en valores del intervalo [0.25, 0.75].

En maíz de regadío, en tres casos la mediana se encuentra a la derecha del punto medio del recorrido (punto 0.5), y por debajo de 0.75. En el intervalo [0.4, 0.6] se encuentran todos los casos, uno a la izquierda del punto medio, y otros tres a su derecha. En un caso se encuentra a su izquierda con un valor superior a 0.25, y en ningún caso la mediana coincide con el punto medio del recorrido. De esta forma, en todos los casos la mediana se sitúa en valores del intervalo [0.25, 0.75].

En la alfalfa de regadío, en el único caso perteneciente a la clase I, la mediana toma el valor 0.75.

En lo que sigue se analiza la posición de la moda y de la mediana para el resto de las respuestas, que como se ha comentado repetidamente, supone la estimación previa de sus recorridos y amplitud de intervalos.

Tabla 137.

Posición de la moda (relativa a intervalos numéricos). Respuestas de la Clase II. Encuesta a los agricultores (1998-99, segundo día).

Cultivo	**I**	**C**	**D**	**Un**	**Bi**	**Total**	**Posición relativa de la moda**								
							≤ -2	**-2 -1**	**-1 0**	**0 1**	**1 2**	**2 3**	**3 4**	**4 5**	**> 5**
Cebada de secano	5	0	6	0	0	11		2	2	4	2	0	1		
Trigo de secano	3	0	4	0	0	7		1	2	1	1	2			
Trigo de regadío	1	0	2	0	0	3			1	2					
Maíz de regadío	3	0	2	0	0	5	1		2	1	1				

*En este caso no se pudo recuperar la alfalfa de regadío.

Tabla 138.

Posición de la moda (relativa a intervalos numéricos). Respuestas de las Clases I y II. Encuesta a los agricultores (1998-99, segundo día).

Cultivo	**I**	**C**	**D**	**Un**	**Bi**	**Total**	**Posición relativa de la moda**								
							≤ -2	**-2 -1**	**-1 0**	**0 1**	**1 2**	**2 3**	**3 4**	**4 5**	**> 5**
Cebada de secano	9	1	11	0	2	23		3	6	7	3		1		1
Trigo de secano	4	1	7	0	2	14		1	4	2	2	2			1
Trigo de regadío	3	0	4	0	0	7		1	2	2	1				1
Maíz de regadío	4	0	5	0	0	9	1		3	3	2				

*En este caso no se pudo recuperar la alfalfa de regadío.

Tabla 139.

Posición de la moda (relativa a intervalos numéricos). Respuestas de las Clases I, II, II y IV. Encuesta a los agricultores (1998-99, segundo día).

Cultivo	**I**	**C**	**D**	**Un**	**Bi**	**Total**	**Posición relativa de la moda**								
							≤ -2	**-2 -1**	**-1 0**	**0 1**	**1 2**	**2 3**	**3 4**	**4 5**	**>5**
Cebada de secano	20	4	19	1	2	46	1	5	17	14	3	1	1		1
Trigo de secano	8	2	11	0	2	23	1	1	8	3	3	4			1
Trigo de regadío	5	1	7	0	0	13		2	4	3	2	1			1
Maíz de regadío	9	1	10	0	0	20	1	4	5	4	5	1			

En lo que se refiere a las medianas, los resultados se resumen en las tablas 140 a 142.

Tabla 140.

Posición relativa de la mediana. Clase de respuesta II. Encuesta a los agricultores (1998 – 99, segundo día).

Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano				1	2	5	3			
Trigo secano					1	4	1	1		
Trigo regadío				1		2		1		
Maíz regadío				1	2	2	1			

*En este caso no se pudo recuperar la alfalfa de regadío.

Tabla 141.

Posición relativa de la mediana. Clases de respuestas I y II. Encuesta a los agricultores (1998 – 99, segundo día).

Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano			1	1	7	8	5		1	
Trigo secano				1	2	6	3	1	1	
Trigo regadío				2	1	3		1		

Maíz regadío				1	3	5				

*En este caso no se pudo recuperar la alfalfa de regadío.

Tabla 142.

Posición relativa de la mediana. Clases de respuestas I, II, III, IV. Encuesta a los agricultores (1998 – 99, segundo día).

Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano		2	2	4	16	14	7		1	
Trigo secano		1		1	6	7	6	1	1	
Trigo regadío				2	3	6	1	1		
Maíz regadío			1	1	8	7	3			

*Dado que la alfalfa de regadío no se pudo recuperar en la clase II, no se consideró tampoco en este caso.

#7.6.- PROBLEMAS EN LA DETERMINACIÓN DE INTERVALOS

De igual forma que ocurrió en la traducción por los expertos de las escalas ordinales a numéricas, donde se definieron los recorridos y los intervalos de diversas formas, algunas de ellas difíciles de utilizar en la discusión de problemas donde sea importante el conocimiento preciso del recorrido y de las fronteras entre intervalos, este mismo fenómeno se repitió en el caso de los agricultores encuestados el primer día y segundo día, en la encuesta llevada a cabo el primer año (1998 – 99).

En las tablas 125 y 134 se resumieron los tipos de respuesta obtenidos, respectivamente, durante el primer y segundo día.

A fin de encontrar relaciones que permitan inferir la amplitud del recorrido a partir de otras informaciones, se analiza en lo que sigue la información contenida en las respuestas de la Clase I. Los resultados se utilizan para la recuperación de la información perdida en las Clases II, III y IV.

#7.6.1.- RESULTADOS CON LA INFORMACIÓN DEL PRIMER DIA

En la tabla 143 se resume el análisis de discontinuidades encontradas en las clases I y II (que deberían ser continuas), y en las tablas 144 a 147 se reflejan los resultados de las regresiones realizadas para poder estimar el recorrido (largo) R_L a partir de la información disponible del recorrido corto R_C (ver ζ 4.2.1).

Tabla 143.

Número de respuestas que presentan discontinuidades / número de respuestas totales. Encuesta a los agricultores (1998 – 99, primer día).

Cultivo	**Índice de Continuidad de Intervalos**		
	Clase I	**Clase II**	**Total**
Cebada secano	6/19	0/11	6/30
Trigo secano	4/13	1/5	5/18
Trigo regadío	1/5	0/3	1/8
Maíz regadío	0/6	0/6	0/12
Alfalfa regadío	2/5	0/4	2/9

Tabla 144.

Regresión $R_L = a\ R_C + b$. Clase de respuestas I. Encuesta a los agricultores (1998 – 99, primer día).

CULTIVO					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	1.23	1.49	1.79	2.08	1.35
b	826.42	483.93	-488.04	-2068.4	250.74
Fc	64.54	16.48	57.54	540.89	41.89
Ft	4.45	4.84	10.13	7.71	10.13
R^2	0.79	0.60	0.95	0.99	0.93

La misma regresión, forzando que el término independiente *b* sea nulo, ofrece los resultados de la tabla 145, obteniéndose un resultado aceptable.

Tabla 145.

Regresión $R_L = a\ R_C$ (con $b = 0$). Clase de respuesta I. Encuesta a los agricultores (1998 – 99, primer día).

CULTIVO					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	1.49	1.66	1.67	1.91	1.37
b	0	0	0	0	0
Fc	497.05	195.86	509.2	511.9	376.3
Ft	4.41	4.75	7.71	6.61	7.71
R^2	0.96	0.94	0.99	0.99	0.99

También se ha realizado la regresión $(R_L - R_C) = a\ R_C + b$. Los resultados se reflejan en la tabla 146, donde se puede apreciar que se obtienen solamente resultados significativos ($\alpha = 0{,}05$) para el trigo y el maíz de regadío, y en la tabla 147 (forzando b = 0), donde se obtienen ajustes significativos para todos los cultivos.

Tabla 146.

Regresión $(R_L - R_C) = a\ R_C + b$. Clase de respuesta I. Encuesta a los agricultores (1998 – 99, primer día).

CULTIVO					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.229	0.495	0.789	1.078	0.350
b	826.42	483.93	-488.04	-2068.39	250.74
Fc	2.24	1.81	11.19	145.55	2.82
Ft	4.45	4.84	10.13	7.71	10.13
R^2	0.12	0.14	0.79	0.97	0.48

Tabla 147.

Regresión *$(R_L – R_C) = a R_C$ (b = 0)*. Clase de respuesta I. Encuesta a los agricultores (1998 – 99, primer día).

	CULTIVO	
	CEBADA SECANO	**TRIGO SECANO**
a	0.492	0.657
b	0	0
Fc	54.11	30.81
Ft	4.41	4.75
R^2	0.75	0.72

Por último, se ha estudiado la distribución de los valores $(R_L - R_C) / R_L$.

Tabla 148.

Cociente R_C / R_L. Clase de respuesta I. Encuesta a los agricultores (1998 – 99, primer día).

	Número de casos									
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano					4	2	6	6		1
Trigo secano				2		3	4	3	1	
Trigo regadío						3	2			
Maíz regadío					1	2	3			
Alfalfa regadío							1	4		

Tabla 149.

Cociente *$(R_L - R_C) / R_L$*. Clase de respuesta I. Encuesta a los agricultores (1998 – 99, primer día).

	Número de casos									
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano	1		6	6	4	2				
Trigo secano		1	3	5	2		2			
Trigo regadío				2	3					
Maíz regadío			1	2	3					
Alfalfa regadío			4	1						

#7.6.1.1.- COHERENCIA DE LOS RECORRIDOS

En la pregunta **número 2** del cuestionario del primer día, el agricultor definió el recorrido de los rendimientos de los diferentes cultivos, mediante la identificación de dos extremos, el mínimo (R_{min}) y el máximo (R_{max}). En la pregunta **número 3** definió los intervalos a partir de las cinco clases expresadas en lenguaje natural. En el caso de las respuestas de la clase I existe una definición precisa de los valores extremos, por lo que es posible comparar el recorrido (largo) de la pregunta 3 con el definido en la pregunta 2.

Las regresiones entre los diferentes extremos y los recorridos (tablas 150, 151 y 152) dan unos resultados razonables, excepto en el caso del trigo. Todo parece indicar que la amplitud del recorrido es del mismo orden cuando se define directamente o a través de intervalos expresados en lenguaje natural, aunque no coinciden exactamente.

Tabla 150.

Regresión $R_L = a\ (R_{max} - R_{min}) + b$. Clase de respuesta I. Encuesta a los agricultores (1998 – 99, primer día).

	CULTIVO				
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.882	0.907	0.942	0.943	1.04
b	944.6	923.95	2263.36	2010.11	-321.72
Fc	54.14	32.68	6.03	352.61	778.12
Ft	4.45	4.84	10.13	7.71	10.13
R^2	0.76	0.75	0.67	0.99	0.99

Tabla 151.

Regresión $R_L = a\ (R_{max} - R_{min})$ (b = 0). Clase de respuesta I. Encuesta a los agricultores (1998 – 99, primer día).

	CULTIVO				
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	1.10	1.10	1.40	1.02	1.02
b	0	0	0	0	0
Fc	418.2	275.9	48.1	366.7	5840.0
Ft	4.41	4.75	7.71	6.61	7.71
R^2	0.96	0.96	0.92	0.99	0.99

A fin estudiar la influencia de sustituir el recorrido largo por el corto, se ha calculado la recta de regresión del recorrido corto de la **pregunta 3** para la clase de respuestas I y los resultados de la pregunta **número 2**, mediante la ecuación $R_C = a\ (R_{max} - R_{min}) + b$, obteniéndose resultados aceptables.

Tabla 152.

Regresión $R_C = a\ (R_{max} - R_{min}) + b$. Clase de respuesta I. Encuesta a los agricultores (1998 – 99, encuesta primer día).

	CULTIVO				
	CEBADA SEC	**TRIGO SEC**	**TRIGO REG**	**MAÍZ REG**	**ALFALFA REG**
a	0.586	0.468	0.464	0.452	0.709
b	559.61	833.72	1774.04	1983.82	344.97
Fc	30.20	31.79	3.62	320.95	27.64
Ft	4.45	4.84	10.13	7.71	10.13
R^2	0.64	0.74	0.55	0.99	0.90

#7.6.1.2.- AMPLITUD DE LOS INTERVALOS

Como ocurría también en el caso de la encuesta a los expertos (epígrafe ζ 5.7), los agricultores al traducir los distintos intervalos ordinales a numéricos, normalmente no dan la misma amplitud a cada intervalo.

Para las respuestas de la clase I, se han normalizado los datos sobre la amplitud de los intervalos de cada encuesta, dando el valor 1 al intervalo más amplio. El resto de los intervalos se han expresado como una fracción del más amplio. Los distintos casos se han resumido en la tabla 153.

Tabla 153.

Número de veces en que el intervalo ( ] toma el valor. Clase de respuestas I. Encuesta a los agricultores (1998 – 99, encuesta primer día).

	Mayor	**0-0.2**	**0.2-0.4**	**0.4-0.6**	**0.6-0.8**	**0.8-1**	**Mediana**	**Media**
Cebada secano								
Muy Malo	4/19	3	5	5	2	4	0,5	0,51
Malo	4/19	2	2	6	4	5	0,54	0,61
Normal	7/19	0	4	5	2	8	0,67	0,7
Bueno	4/19	0	6	5	3	5	0,5	0,61
Muy Bueno	2/19	2	3	7	3	4	0,5	0,56
Trigo Secano								
Muy Malo	5/13	2	2	2	2	5	0,64	0,64
Malo	6/13	0	2	4	1	6	0,8	0,74
Normal	7/13	1	1	3	1	7	1	0,75
Bueno	2/13	1	2	4	3	3	0,5	0,59
Muy Bueno	3/13	1	2	4	3	3	0,6	0,6
Trigo regadío								
Muy Malo	3/5	0	1	1	0	3	1	0,78
Malo	3/5	0	0	2	0	3	1	0,8
Normal	3/5	0	0	1	1	3	1	0,85
Bueno	3/5	0	0	2	0	3	1	0,8
Muy Bueno	4/5	1	0	0	0	4	1	0,8
Maíz regadío								
Muy Malo	2/6	0	1	3	0	2	0,55	0,66
Malo	2/6	0	0	4	0	2	0,55	0,68
Normal	2/6	0	0	3	1	2	0,63	0,71
Bueno	3/6	0	0	1	2	3	0,9	0,83
Muy Bueno	3/6	0	0	2	1	3	0,83	0,78
Alfalfa regadío								
Muy Malo	1/5	1	2	1	0	1	0,4	0,5
Malo	2/5	0	2	0	1	2	0,7	0,7
Normal	2/5	0	1	0	1	3	0,8	0,8
Bueno	2/5	0	1	1	1	2	0,8	0,7
Muy Bueno	0/5	2	1	2	0	0	0,3	0,4

En el caso de la cebada de secano, de las 19 respuestas de la clase I obtenidas, 7 asignan el intervalo más amplio a la clase de rendimientos *"normal"*, y esta clase presenta la mitad de las observaciones por encima de un valor relativamente elevado, 0,67. Aunque la clase *"normal"* es la que parece recibir las mayores amplitudes de intervalo, la distribución de frecuencias de la clase a la cual se asigna el mayor valor del intervalo es relativamente uniforme, sin que se pueda concluir la existencia de una tendencia general a presentar como más amplia alguna de las cinco clases definidas en la pregunta **número 3**.

Algo semejante se observa para el caso del trigo de secano, donde la clase *"normal"* tiende a ser la más amplia. En general, en los cultivos de secano, las clases *"malo"* y *"muy malo"* presentan mayores amplitudes que las de *"bueno"* y *"muy bueno"*.

En el caso de los cultivos de regadío se dispone de pocas observaciones, aunque en el trigo existe una cierta tendencia a dar la misma amplitud a los intervalos, y tanto en el trigo como en el maíz tienden a presentar mayores valores las clases *"bueno"* y *"muy bueno"*, contrariamente a lo que ocurre en secano, mientras que la alfalfa presenta una distribución bastante uniforme, ligeramente desplazada en el mismo sentido que los cultivos de secano.

A la vista de los resultados anteriores, no se puede afirmar que algunas clases tiendan a concentrar las mayores amplitudes, aunque la clase de rendimientos *"normal"* posiblemente actúe como un anclaje para el establecimiento de las amplitudes.

Probablemente la amplitud del intervalo esté asociada a la precisión con la que se desea puntualizar algunos rendimientos.

A continuación se midió la amplitud de los intervalos máximo y mínimo y se expresó como porcentaje del recorrido (largo) (Tabla 154) y como porcentaje del punto medio del intervalo de rendimientos *normales* (Tabla 155), para las respuestas de la clase I.

Tabla 154.

Amplitud del intervalo máximo y del mínimo en las diferentes encuestas, expresada como porcentaje del recorrido (largo). Clase de respuestas I. Encuesta a los agricultores (1998 – 99, primer día).

		Número de casos									
Cultivo	**Amplitud**	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano	Máxima			6	10	3					
	Mínima	6	13								
Trigo secano	Máxima		1	6	3	2	1				
	Mínima	5	8								
Trigo regadío	Máxima			4	1						
	Mínima	2	3								
Maíz regadío	Máxima			5	1						
	Mínima	1	5								
Alfalfa regadío	Máxima			1	4						
	Mínima	3	2								

Tabla 155.

Amplitud del intervalo máximo y del mínimo en las diferentes encuestas, expresada como porcentaje del punto medio del intervalo de rendimientos *normales*. Clase de respuestas I. Encuesta a los agricultores (1998 – 99, primer día).

		Número de casos									
Cultivo	**Amplitud**	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano	Máxima*	1		2	6	3	3	1	2		
	Mínima	7	7	5							
Trigo secano	Máxima	1	1	2	2	4	1	1	1		
	Mínima	3	7	3							
Trigo regadío	Máxima		1	2	1	1					
	Mínima	1	4								

Maíz regadío	Máxima		3	2					1		
	Mínima	3	2		1						
Alfalfa regadío	Máxima			3	1	1					
	Mínima	3	2								

*Habiendo eliminado una de las observaciones (número 10).

#7.6.1.3.- ANÁLISIS DE LOS EXTREMOS

En la **pregunta 2** de la encuesta del primer día se pedía al agricultor una estimación del recorrido de los rendimientos de cada cultivo: mínimo *(R_{min})*, más frecuente *(R_{mf})* y máximo *(R_{max})*. En la **pregunta 3**, en las respuestas de la clase I se dispone del recorrido entre el extremo inferior de la clase *"rendimientos muy malos"* ($R_{<MM}$) y el extremo superior de la clase *"rendimientos muy buenos"* ($R_{>MB}$). Es posible comparar ambos valores, y para ello, inicialmente, se han calculado el coeficiente de correlación de Spearman, el test de independencia (para un nivel de significación de α = 0.05) y el test de Wilcoxon (a un nivel de significación de α = 0.05), a fin de determinar si las dos poblaciones están relacionadas y si se puede afirmar que se acepta la hipótesis de que sus valores medios pertenecen a la misma distribución (tabla 156). Se observa en general una semejanza entre la población de rendimientos máximos que en la de rendimientos mínimos.

Tabla 156.

Comparación entre poblaciones de rendimientos máximos y mínimos. Respuestas de la Clase I. Encuesta a los agricultores (1998 –99, primer día).

Cultivo	**Comparación**	**Spearman (r_S)**	**Muestras independientes**	**Wilcoxon (α=0.05)**
Cebada secano	R_{min} con $R_{<MM}$	0.634	No	No existen diferencias entre las 2 poblaciones
	R_{max} con $R_{>MB}$	0.98	No	id
Trigo secano	R_{min} con $R_{<MM}$	0.508	No	id
	R_{max} con $R_{>MB}$	0.959	No	id
Trigo regadío	R_{min} con $R_{<MM}$	0.475	No	id
	R_{max} con $R_{>MB}$	0.975	No	id
Maíz regadío	R_{min} con $R_{<MM}$	0.028	Sí	id
	R_{max} con $R_{>MB}$	0.886	No	id
Alfalfa regadío	R_{min} con $R_{<MM}$	1	No	id
	R_{max} con $R_{>MB}$	1	No	id

Para profundizar en la misma idea, se ha realizado una regresión, para los cultivos cebada y trigo de secano, entre el *rendimiento mínimo* y el *rendimiento muy malo* y el *rendimiento máximo* y el *rendimiento muy bueno*, y los resultados se han resumido en la tabla 157, pudiéndose apreciar que se obtienen mejores resultados para los rendimientos máximos que para los mínimos.

Tabla 157.

Regresión entre rendimientos máximos y mínimos. Clase de respuestas I. Encuesta a los agricultores (1998 – 99, primer día).

Cultivo	**Comparación**	**y = a x + b, b ≠ 0**	**y = ax, b = 0**

Cultivo	**Comparación**	**y = a x + b, b ≠ 0**	**y = ax, b = 0**
Cebada secano	R_{min} con $R_{<MM}$	$R_{min} = 0.791\ R_{<MM} + 625.40$ $R^2 = 0.74$ Fc = 47.41, Ft = 4.45	$R_{min} = 1.05\ R_{>MM}$ $R^2 = 0.86$ Fc = 106.26, Ft = 4.41
	R_{max} con $R_{>MB}$	$R_{max} = 0.95\ R_{>MB} + 127.82$ $R^2 = 0.88$ Fc = 121.62, Ft = 4.45	$R_{max} = 0.97\ R_{>MB}$ $R^2 = 0.99$ Fc = 1663.46, Ft = 4.41
Trigo secano	R_{min} con $R_{<MM}$	$R_{min} = 0.689\ R_{<MM} + 856.93$ $R^2 = 0.62$ Fc = 18.18, Ft = 4.84	$R_{min} = 0.79\ R_{<MM}$ $R^2 = 0.95$ Fc = 229.82, Ft = 4.75
	R_{max} con $R_{>MB}$	$R_{max} = 1.13\ R_{>MB} – 949.87$ $R^2 = 0.94$ Fc = 165.36, Ft = 4.84	$R_{max} = 0.98\ R_{>MB}$ $R^2 = 0.99$ Fc = 1774.44, Ft = 4.75
Trigo regadío	R_{min} con $R_{<MM}$	$R_{min} = 0.26\ R_{<MM} + 2638.75$ $R^2 = 0.20$ Fc = 0.73, Ft = 10.13	$R_{min} = 1.06\ R_{<MM}$ $R^2 = 0.62$ Fc = 6.4, Ft = 7.71
	R_{max} con $R_{>MB}$	$R_{max} = 1.02\ R_{>MB} – 841.8$ $R^2 = 0.83$ Fc = 15.1, Ft = 10.13	$R_{max} = 0.91\ R_{>MB}$ $R^2 = 0.99$ Fc = 319.8, Ft = 7.71
Maíz regadío	R_{min} con $R_{<MM}$	$R_{min} = 0.98\ R_{<MM} + 625.74$ $R^2 = 0.99$ Fc = 1309.5, Ft = 7.71	$R_{min} = 1.01\ R_{<MM}$ $R^2 = 0.99$ Fc = 1756.5, Ft = 6.61
	R_{max} con $R_{>MB}$	$R_{max} = 1.02\ R_{>MB} – 1473.9$ $R^2 = 0.99$ Fc = 2785.9, Ft = 7.71	$R_{max} = 0.99\ R_{>MB}$ $R^2 = 0.99$ Fc = 2439.2, Ft = 6.61
Alfalfa regadío	R_{min} con $R_{<MM}$	$R_{min} = R_{<MM}$ $R^2 = 1$ Fc = -, Ft = 10.13	$R_{min} = 1.00\ R_{<MM}$ $R^2 = 1.00$ Fc = - , Ft = 7.71
	R_{max} con $R_{>MB}$	$R_{max} = 0.94\ R_{>MB} – 974.6$ $R^2 = 0.99$ Fc = 789.2, Ft = 10.13	$R_{max} = 0.99\ R_{>MB}$ $R^2 = 1.00$ Fc = 11582.0, Ft = 7.71

Diferencias en los extremos inferiores del recorrido. Se ha examinado también la cuantía de las diferencias entre los valores de R_{min} (pregunta 2) y $R_{<MM}$, expresadas como porcentaje del valor R_{min}, para normalizar. La distribución del número de casos según la cuantía de la diferencia se reproduce en la tabla 158.

Tabla 158.

Cociente $|(R_{<MM}-R_{min})/R_{min}|$. Clase de respuesta I. Encuesta a los agricultores (1998 – 99, primer día).

	Número de casos

Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano	10	1	1	2	1		1			3
Trigo secano	10				1					2
Trigo regadío	2			1						2
Maíz regadío	4	1								1
Alfalfa regadío	5									

En la cebada de secano, en 9 casos se obtuvieron diferencias apreciables, especialmente en 7 casos en los que la diferencia es superior al 30% del valor de R_{min}. De ellas, los tres casos con diferencias del 100% se pueden considerar "outliers", ya que se corresponden con respuestas $R_{<MM} = 0$, que obviamente es una forma de hablar más que de describir un recorrido. El resto de los valores se corresponden con rendimientos modestos, por lo que cualquier diferencia apreciable resulta porcentualmente importante. Es importante subrayar que en todos los casos $R_{min} \geq R_{<MM}$.

En el caso del trigo de secano, en 2 observaciones la diferencia es inferior o igual al 10%, y en un caso $R_{<MM} < R_{min}$ y en el otro ocurre lo contrario. En el caso en que la diferencia es del orden del 44%, se tiene que $R_{<MM} < R_{min}$.

Para el trigo de regadío, en 1 caso la diferencia es del 32%, con $R_{<MM} < R_{min}$. En el caso del maíz de regadío, en la única observación en que la diferencia no es del 0 ni del 100% entre los rendimientos mínimo y muy malo, se produce que $R_{<MM} < R_{min}$, con una diferencia del orden del 20%. En el caso de la alfalfa de regadío, se dispone de 5 respuestas de la clase I, y en todas ellas $R_{min} = R_{<MM}$.

En la Tabla 159 pueden verse los casos en que se observó una diferencia nula o máxima (100%) entre los dos valores de rendimientos, para las respuestas de la clase I.

Tabla 159.

Diferencias nula y máxima entre $R_{<MM}$ y R_{min}. Clase de respuestas I. Encuesta a los agricultores (1998 – 99, primer día).

Cultivo	**Encuestas clase I**	**Diferencia $\lvert R_{<MM} - R_{min} \rvert$**	
		0%	**100%**
Cebada secano	19	8	3
Trigo secano	13	8	2
Trigo regadío	5	2	2
Maíz regadío	6	4	1
Alfalfa regadío	5	5	0

En el 42% de los casos se obtiene una diferencia nula para la cebada de secano, en un 62% para el trigo de secano, en un 40% para el trigo de regadío, en un 67% para el maíz de regadío y en un 100% para la alfalfa de regadío.

Diferencias en los extremos superiores del recorrido. Se ha estudiado el número de casos según la cuantía de las diferencias entre los valores de R_{max} (pregunta 2) y $R_{>MB}$ (pregunta 3), expresadas como porcentaje del valor R_{max}. Los resultados se reproducen en la tabla 160.

Tabla 160.

Cociente $\lvert(R_{>MB}-R_{max})/R_{max}\rvert$. Clase de respuestas I. Encuesta a los agricultores (1998 – 99, primer día).

	Número de casos									
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano	12	3	4							
Trigo secano	11	1				1				
Trigo regadío	3	1			1					
Maíz regadío	3	1	1		1					
Alfalfa regadío	5									

En cebada de secano, de los 6 casos en que la diferencia es del orden del 20 – 30%, en 5 casos $R_{>MB} > R_{max}$, y solamente en 1 ocurre lo contrario. En los 2 casos en los que las diferencias son del orden del 5 – 10%, $R_{max} > R_{>MB}$.

En el caso del trigo de secano, de los 2 casos en que la diferencia es inferior al 10%, en uno $R_{max} > R_{>MB}$, y en el otro ocurre lo contrario. En los casos en que la diferencia es superior al 20%, las respuestas son $R_{>MB} > R_{max}$. En el caso del trigo de regadío, de las 5 respuestas disponibles de la clase I, en tres se produce $R_{>MB} > R_{max}$.

En el maíz de regadío, de las 6 respuestas disponibles de la clase I, en 3 $R_{>MB} > R_{max}$. En la alfalfa de regadío, en una observación se produce $R_{>MB} > R_{max}$ con una diferencia relativa del 4%.

En la Tabla 161 pueden verse los casos en que se observó $R_{MB} = R_{max}$, para las respuestas de la clase I.

Tabla 161.

Casos en que $R_{MB} = R_{max}$. Clase de respuestas I. Encuesta a los agricultores (1998 – 99, primer día).

Cultivo	**Encuestas clase I**	**$R_{>MB} = R_{max}$**
Cebada secano	19	11
Trigo secano	13	9
Trigo regadío	5	2
Maíz regadío	6	3
Alfalfa regadío	5	4

En un 58% de los casos la diferencia es nula para la cebada de secano, en un 69% para el trigo de secano, en un 40% para el trigo de regadío, en un 50% para el maíz de regadío y en un 80% para la alfalfa de regadío.

En definitiva, se observaron en general mayores diferencias relativas para los extremos inferiores que para los superiores en todos los cultivos.

Diferencias en los dos extremos del recorrido. Se han analizado simultáneamente las diferencias de los extremos inferiores y superiores, resumidas en la tabla 162.

Tabla 162.

Valores de los extremos inferiores R_{min}, $R_{<MM}$ y superiores R_{max}, $R_{>MB}$. Clase de respuesta I. Encuesta a los agricultores (1998 – 99, primer día).

Cultivo	**$R_{<MM}$=0**	**$R_{min}=R_{<MM}$ $R_{max}=R_{>MB}$**	**$R_{min}=R_{<MM}$ $R_{max}>R_{>MB}$**	**$R_{min}=R_{<MM}$ $R_{max}<R_{>MB}$**	**$R_{max}=R_{>MB}$ $R_{min}>R_{<MM}$**	**$R_{max}=R_{>MB}$ $R_{min}<R_{<MM}$**	**$R_{max}<R_{>MB}$ $R_{min}>R_{<MM}$**	**$R_{max}>R_{>MB}$ $R_{min}<R_{<MM}$**	**Total**
Cebada secano	3	6	1	1	3	0	5	0	19
Trigo secano	2	6	0	2	3	0	0	0	13
Trigo regadío	2	1	0	1	1	0	0	0	5
Maíz regadío	1	3	0	1	0	0	1	0	6
Alfalfa regadío	0	4	0	1	0	0	0	0	5

De las 19 respuestas analizadas en el caso de la cebada de secano, en 6 casos los agricultores daban los mismos valores de $R_{min} = R_{<MM}$ y $R_{max} = R_{>MB}$. De los 5 casos restantes en que $R_{max} = R_{>MB}$, en 2 los agricultores han respondido con $R_{<MM} = 0$, en otros 2 las diferencias entre R_{min} y $R_{<MM}$ son del orden del 10% y en 1 la diferencia es del 50%.

En los otros 2 casos en que $R_{min} = R_{<MM}$, en uno $R_{max} > R_{>MB}$ (diferencia del 5%), y en otro $R_{max} < R_{>MB}$ (diferencia del 30%). De los 6 casos en los que $R_{min} > R_{<MM}$, en 5 resulta $R_{>MB} > R_{max}$ (diferencias del orden del 20 – 25%), y el otro caso es una respuesta con $R_{<MM} = 0$.

Se observa una clara tendencia a incrementar el recorrido cuando se define en términos de lenguaje natural, ya que de los 10 casos en los que no son iguales los extremos, existen 9 casos en los que $R_{min} \geq R_{<MM}$ y $R_{max} \leq R_{>MB}$.

En los demás cultivos, también se observa una clara tendencia a ampliar los recorridos al expresarlos en intervalos definidos mediante lenguaje natural.

#7.6.2.- RESULTADOS CON LA INFORMACIÓN DEL SEGUNDO DIA

En la tabla 163 se indican los resultados de aplicar el índice de continuidad de intervalos a los datos obtenidos con el cuestionario del segundo día (1998 –99).

Tabla 163.

Número de respuestas que presentan discontinuidades / número de respuestas totales. Encuesta a los agricultores (1998 – 99, segundo día).

	Índice de Continuidad de Intervalos		
Cultivo	**Clase I**	**Clase II**	**Total**
Cebada secano	4/12	1/11	5/23
Trigo secano	3/7	1/8	4/15
Trigo regadío	1/4	0/3	1/7
Maíz regadío	1/4	1/5	2/9
Alfalfa regadío	0/2	0/4	0/6

Los resultados de las comparaciones entre el recorrido (largo) R_L y el corto R_C, se han resumido en las tablas 164 a 169.

Tabla 164.

Regresión $R_L = a\ R_C + b$. Clase de respuestas I. Encuesta a los agricultores (1998 – 99, segundo día).

CULTIVO					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO (*)**
a	1.09	0.98	1.08	1.41	-
b	1218.7	1799.2	1316.7	1161.1	-
Fc	42.1	8.37	26.7	14.9	-
Ft	4.96	6.61	18.51	18.51	-
R^2	0.81	0.63	0.93	0.88	-

(*) No se tiene en cuenta este cálculo para dos observaciones solamente.

Se obtuvo una regresión significativa para cebada y trigo de secano, y trigo de regadío. La misma regresión, forzando que el término independiente *b* fuera nulo, ofrece los resultados de la tabla 165.

Tabla 165.

Regresión $R_L = a\ R_C$ (con $b = 0$). Clase de respuesta I. Encuesta a los agricultores (1998 – 99, segundo día).

CULTIVO					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO (*)**
a	1.48	1.55	1.61	1.64	-
b	0	0	0	0	-
Fc	336.56	110.94	250.9	193.4	-
Ft	4.84	5.99	10.13	10.13	-
R^2	0.97	0.95	0.99	0.99	-

(*) No se tiene en cuenta este cálculo para dos observaciones solamente

Obteniéndose una regresión significativa para todos los cultivos. A continuación, se realizó la regresión $(R_L - R_C) = a\ R_C + b$.

Tabla 166.

Regresión $(R_L - R_C) = a\ R_C + b$. Clase de respuesta I. Encuesta a los agricultores (1998 – 99, segundo día).

CULTIVO					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO (*)**
a	0.09	-0.02	0.08	0.41	-
b	1218.7	1799.2	1316.7	1161.1	-
Fc	0.27	0	0.15	1.27	-
Ft	4.96	6.61	18.51	18.51	-
R^2	0.03	0	0.07	0.39	-

(*) No se considera el resultado con dos observaciones solamente.

Se obtuvo un mal resultado, con ningún cultivo significativo para esta regresión y se repitió la regresión para el caso de $b = 0$.

Tabla 167.

Regresión $(R_L – R_C) = a\ R_C\ (b = 0)$. Clase de respuesta I. Encuesta a los agricultores (1998 – 99, segundo día).

	CULTIVO	
	CEBADA SECANO	**TRIGO SECANO**
a	0.49	0.55
b	0	0
Fc	35.95	13.98
Ft	4.84	5.99
R^2	0.77	0.70

Obteniéndose un ajuste significativo en ambos casos.

Tabla 168.

Cociente R_C / R_L. Clase de respuesta I. Encuesta a los agricultores (1998 – 99, segundo día).

	Número de casos									
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0 .5**	**0.5- 0.6**	**0.6- 0.7**	**0.7- 0.8**	**0.8- 0.9**	**0.9-1**
Cebada secano					2	4	2	4		
Trigo secano				1	2		2	2		
Trigo regadío						3		1		
Maíz regadío					1	1	2			
Alfalfa regadío							1	1		

Tabla 169.

Cociente $(R_L - R_C) / R_L$. Clase de respuesta I. Encuesta a los agricultores (1998 – 99, segundo día).

	Número de casos									
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano			4	3	4	1				
Trigo secano			2	2	2		1			
Trigo regadío			1		3					
Maíz regadío				2	2					
Alfalfa regadío			1	1						

#7.6.2.1.- COHERENCIA DE LOS RECORRIDOS

Se comparó, mediante regresión, el recorrido largo de la distribución, según los resultados de la **pregunta 3** para las clases de respuestas I, y el recorrido calculado a partir de las respuestas a la **pregunta número 2**, mediante la ecuación $R_L = a\ (R_{max} - R_{min}) + b$. A diferencia del primer día, se obtuvo en este caso y para todos los cultivos (excepto alfalfa de regadío, en que no se realizó por contar sólo con dos observaciones) un mal ajuste en esta regresión. Para el caso de b = 0, los ajustes mejoraron, obteniéndose coeficientes "a" cercanos a la unidad.

Tabla 170.

Regresión $R_L = a\ (R_{max} - R_{min}) + b$. Clase de respuesta I. Encuesta a los agricultores (1998 – 99, segundo día).

	CULTIVO				
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO(*)**
a	0.16	0.03	0.47	0.54	-
b	3536.4	4364.4	2291.6	4256.8	-
Fc	0.24	0	3	1.83	-
Ft	4.96	6.61	18.51	18.51	-
R^2	0.02	0	0.60	0.48	-

(*) No se tiene en cuenta este cultivo por tener sólo dos observaciones.

Tabla 171.

Regresión $R_L = a\ (R_{max} - R_{min})$ (b = 0). Clase de respuesta I. Encuesta a los agricultores (1998 – 99, segundo día).

	CULTIVO				
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO(*)**
a	1.16	1.18	1.10	1.08	-
b	0	0	0	0	-
Fc	43.7	22.9	46.0	21.7	-
Ft	4.84	5.99	10.13	10.13	-
R^2	0.80	0.79	0.94	0.88	-

(*) No se tiene en cuenta este cultivo por tener sólo dos observaciones.

En la regresión $R_C = a\ (R_{max} - R_{min}) + b$ se obtuvo un mal resultado.

#7.6.2.2.- AMPLITUD DE LOS INTERVALOS

En las tablas 172 y 174 se resume el análisis realizado sobre la amplitud de los intervalos.

Tabla 172.

Número de veces en que el intervalo ( ] toma el valor. Clase de respuestas I. Encuesta a los agricultores (1998 – 99, encuesta segundo día).

	Mayor	**0-0.2**	**0.2-0.4**	**0.4-0.6**	**0.6-0.8**	**0.8-1**	**Mediana**	**Media**
Cebada secano								
Muy Malo	3/12	2	1	3	2	4	0.60	0,63
Malo	4/12	1	0	4	3	4	0.65	0.70
Normal	5/12	1	1	2	3	5	0.69	0.71

	Mayor	0-0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1	Mediana	Media
Bueno	4/12	1	4	0	2	5	0.69	0.65
Muy Bueno	3/12	1	2	3	2	4	0.63	0.66
Trigo Secano								
Muy Malo	3/7	1	1	1	1	3	0.64	0.67
Malo	2/7	0	2	2	1	2	0.60	0.66
Normal	2/7	1	2	1	1	2	0.50	0.59
Bueno	1/7	1	2	2	1	1	0.50	0.54
Muy Bueno	1/7	2	1	2	1	1	0.50	0.49
Trigo regadío								
Muy Malo	1/4	1	0	1	0	2	0.70	0.65
Malo	1/4	0	0	1	1	2	0.90	0.83
Normal	1/4	1	0	1	1	1	0.63	0.61
Bueno	3/4	1	0	0	0	3	1	0.80
Muy Bueno	3/4	0	1	0	0	3	1	0.85
Maíz regadío								
Muy Malo	2/4	0	0	2	0	2	0.80	0.78
Malo	3/4	0	0	1	0	3	1	0.88
Normal	2/4	0	0	2	0	2	0.76	0.76
Bueno	1/4	0	0	3	0	1	0.50	0.62
Muy Bueno	2/4	0	0	2	0	2	0.75	0.75
Alfalfa regadío								
Muy Malo	1/2	0	1	0	0	1	0.67	0.67
Malo	1/2	0	1	0	0	1	0.67	0.67
Normal	1/2	0	0	0	1	1	0.83	0.83
Bueno	1/2	0	0	1	0	1	0.75	0.75
Muy Bueno	0/2	0	1	1	0	0	0.42	0.42

Igual que en el primer día, no se observa una tendencia a asignar mayores amplitudes a un intervalo concreto.

A continuación se midió la amplitud de los intervalos máximo y mínimo y se expresó como porcentaje del recorrido (largo) (Tabla 173) y como porcentaje del punto medio del intervalo de rendimientos *normales* (Tabla 174), para las respuestas de la clase I.

Tabla 173.

Amplitud del intervalo máximo y del mínimo en las diferentes encuestas, expresada como porcentaje del recorrido (largo). Clase de respuestas I. Encuesta a los agricultores (1998 – 99, segundo día).

		Número de casos									
Cultivo	**Amplitud**	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano	Máxima		1	6	3	1	1				
	Mínima	3	9								
Trigo secano	Máxima			3	1	2	1				
	Mínima	2	5								
Trigo regadío	Máxima			3	1						
	Mínima	2	2								
Maíz regadío	Máxima			3	1						
	Mínima		4								
Alfalfa regadío	Máxima			1	1						
	Mínima		2								

Tabla 174.

Amplitud del intervalo máximo y del mínimo en las diferentes encuestas, expresada como porcentaje del punto medio del intervalo de rendimientos *normales*. Clase de respuestas I. Encuesta a los agricultores (1998 – 99, segundo día).

		Número de casos									
Cultivo	**Amplitud**	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano	Máxima	1	1	1	2	3	3		1		
	Mínima	3	5	4							
Trigo secano	Máxima	1			1	2		3			
	Mínima	1	5	1							
Trigo regadío	Máxima		1	3							
	Mínima	2	2								
Maíz regadío	Máxima		2	2							
	Mínima	2	2								
Alfalfa regadío	Máxima			1		1					
	Mínima		2								

#7.6.2.3.- ANÁLISIS DE LOS EXTREMOS

Se ha comparado el valor R_{min} de la pregunta 2 con el valor $R_{<MM}$ de la **pregunta 3**, y el valor R_{max} con $R_{>MB}$, y los resultados se han reproducido en las tablas 175 a 181.

Tabla 175.

Comparación entre poblaciones de rendimientos máximos y mínimos. Respuestas de la Clase I. Encuesta a los agricultores (1998 –99, segundo día).

Cultivo	**Comparación**	**Spearman (r_S)**	**Muestras independientes**	**Wilcoxon (α=0.05)**
Cebada secano	R_{min} con $R_{<MM}$	0.524	No	Existen diferencias entre las 2 poblaciones
	R_{max} con $R_{>MB}$	0.886	No	No existen diferencias
Trigo secano	R_{min} con $R_{<MM}$	-0.143	Sí	No existen diferencias
	R_{max} con $R_{>MB}$	0.902	No	No existen diferencias
Trigo de regadío	R_{min} con $R_{<MM}$	0.4	(*)	(*)
	R_{max} con $R_{>MB}$	0.8	(*)	(*)
Maíz de regadío	R_{min} con $R_{<MM}$	1	(*)	(*)
	R_{max} con $R_{>MB}$	1	(*)	(*)
Alfalfa de regadío	R_{min} con $R_{<MM}$	1	(*)	(*)
	R_{max} con $R_{>MB}$	1	(*)	(*)

(*) Debido al tamaño de muestra (entre 2 y 4 observaciones), no se realizó este test.

Se efectuó la regresión entre el rendimiento *mínimo* y el rendimiento *muy malo* y el rendimiento *máximo* y el rendimiento *muy bueno*, para los casos en que la constante era igual o diferente de cero (tabla 176).

Tabla 176.

Regresión entre rendimientos máximos y mínimos. Clase de respuestas I. Encuesta a los agricultores (1998 – 99, segundo día). (*)

Cultivo	Comparación	y = a x +b, b ≠ 0	y = ax, b = 0
Cebada secano	R_{min} con $R_{<MM}$	$R_{min} = 0.57\ R_{<MM} + 1243.12$ $R^2 = 0.26$ Fc = 3.43, Ft = 4.96	$R_{min} = 1.1\ R_{<MM}$ $R^2 = 0.63$ Fc = 18.90, Ft = 4.84
	R_{max} con $R_{>MB}$	$R_{max} = 0.93\ R_{>MB} + 78.04$ $R^2 = 0.65$ Fc = 18.31, Ft = 4.96	$R_{max} = 0.94\ R_{>MB}$ $R^2 = 0.98$ Fc = 656.21, Ft = 4.84
Trigo secano	R_{min} con $R_{<MM}$	$R_{min} = 0.3\ R_{<MM} + 1939.5$ $R^2 = 0.06$ Fc = 0.31, Ft = 6.61	$R_{min} = 1.06\ R_{<MM}$ $R^2 = 0.41$ Fc = 4.10, Ft = 5.99
	R_{max} con $R_{>MB}$	$R_{max} = 1.52\ R_{>MB} – 2855.4$ $R^2 = 0.90$ Fc = 46.2, Ft = 6.61	$R_{max} = 1.01\ R_{>MB}$ $R^2 = 0.99$ Fc = 666.36, Ft = 5.99
Trigo regadío	R_{min} con $R_{<MM}$	$R_{min} = 0.95\ R_{<MM} + 203.4$ $R^2 = 0.99$ Fc = 157.4, Ft = 18.51	$R_{min} = 0.98\ R_{<MM}$ $R^2 = 0.99$ Fc = 2512.7, Ft = 10.13
	R_{max} con $R_{>MB}$	$R_{max} = 1.25\ R_{>MB} – 2214.64$ $R^2 = 0.82$ Fc = 8.99, Ft = 18.51	$R_{max} = 0.93\ R_{>MB}$ $R^2 = 0.98$ Fc = 155.1, Ft = 10.13
Maíz regadío	R_{min} con $R_{<MM}$	$R_{min} = 1.68\ R_{<MM} – 3325.88$ $R^2 = 0.89$ Fc = 15.8, Ft = 18.51	$R_{min} = 1.20\ R_{<MM}$ $R^2 = 0.95$ Fc = 59.8, Ft = 10.13
	R_{max} con $R_{>MB}$	$R_{max} = 1.06\ R_{>MB} – 1530.14$ $R^2 = 0.99$ Fc = 140.9, Ft = 18.51	$R_{max} = 0.96\ R_{>MB}$ $R^2 = 0.99$ Fc = 979.6, Ft = 10.13

(*) No se incluyó la alfalfa por tener sólo 2 observaciones.

Obteniéndose en general mejores ajustes para los extremos superiores que para los inferiores, en ambos casos de *b* igual o distinta de cero.

Diferencias en los extremos inferiores del recorrido. Se examinó también la cuantía de las diferencias entre los valores de R_{min} (pregunta 2) y $R_{<MM}$, expresadas como porcentaje del valor R_{min}, para normalizar. La distribución del número de casos según la cuantía de la diferencia se reproduce en la tabla 177.

Tabla 177.

Cociente $|(R_{<MM}-R_{min})/R_{min}|$. Clase de respuesta I. Encuesta a los agricultores (1998 – 99, segundo día).

	Número de casos

Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano	4		1	2	1		2	1		1
Trigo secano	3	1							2	1
Trigo regadío	4									
Maíz regadío	2	1		1						
Alfalfa regadío	2									

De las 12 respuestas de la clase I obtenidas para la cebada de secano, en 4 casos los agricultores dieron el mismo valor a R_{min} y $R_{<MM}$. Sin embargo, en 7 casos se obtuvieron diferencias apreciables, con una diferencia superior al 30% del valor de R_{min}. De ellas, el caso con diferencias del 100% se puede considerar "outlier", ya que se corresponde con respuestas $R_{<MM} = 0$, que es una forma de hablar más que de describir un recorrido. El resto de los valores se corresponden con rendimientos modestos, por lo que cualquier diferencia apreciable resulta porcentualmente importante. Es importante subrayar que excepto en un caso $R_{min} \geq R_{<MM}$.

En el caso del trigo de secano, de las 7 observaciones con respuestas de la clase I, en 3 casos se han dado los mismos valores a los extremos inferiores, es decir, $R_{min} = R_{<MM}$, y en un caso $R_{<MM} < R_{min}$. El caso con diferencias del 100% es consecuencia de dar el valor $R_{<MM} = 0$.

Para el trigo de regadío se dispone de 4 observaciones con respuestas de la clase I, y de ellas 3 con $R_{min} = R_{<MM}$.. En todos los casos, $R_{min} \geq R_{<MM}$.

Para el maíz de regadío se dispone de 4 observaciones, de las cuales en 2 se verifica que $R_{min} = R_{<MM}$ y en una $R_{<MM} > R_{min}$.

Por último, en el caso de la alfalfa de regadío se dispone de 2 respuestas de la clase I, y en todas ellas $R_{min} = R_{<MM}$.

En la Tabla 178 pueden consultarse los casos en que se observó una diferencia nula o máxima (100%) entre los dos valores de rendimientos, para las respuestas de la clase I.

Tabla 178.

Diferencias nula y máxima entre $|R_{<MM} - R_{min}|$. Clase de respuestas I. Encuesta a los agricultores (1998 – 99, segundo día).

Cultivo	**Encuestas clase I**	**Diferencia $\|R_{<MM} - R_{min}\|$**	
		0%	**100%**
Cebada secano	12	4	1
Trigo secano	7	3	1
Trigo regadío	4	3	0
Maíz regadío	4	2	0
Alfalfa regadío	2	2	0

Diferencias en los extremos superiores del recorrido. Se estudió el número de casos según la cuantía de las diferencias entre los valores de R_{max} (**pregunta 2**) y $R_{>MB}$, expresadas como porcentaje del valor R_{max}. Los resultados se reproducen en la tabla 179.

Tabla 179.

Cociente $|(R_{>MB}-R_{max})/R_{max}|$. Clase de respuestas I. Encuesta a los agricultores (1998 – 99, segundo día).

	Número de casos									
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano	6	4	2							
Trigo secano	5	1		1						
Trigo regadío	3				1					
Maíz regadío	3		1							
Alfalfa regadío	2									

De las 12 respuestas de la clase I obtenidas para la cebada de secano, en 6 casos los agricultores dieron el mismo valor para R_{max} y R_{MB}. Excepto en un caso (diferencia del 16%), $R_{>MB} \geq R_{max}$. En el caso del trigo de secano, de las 7 respuestas obtenidas, en 5 $R_{max} = R_{>MB}$. Excepto en un caso (diferencia del 33.3%), $R_{>MB} \leq R_{max}$.

En el caso del trigo de regadío, de las 4 respuestas disponibles de la clase I, en tres de ellas $R_{max} = R_{>MB}$, y en las dos restantes $R_{>MB} > R_{max}$. En el maíz de regadío, de las 4 respuestas disponibles de la clase I, en 2 se verificó $R_{max} = R_{>MB}$, y en las 2 restantes $R_{>MB} > R_{max}$. En la alfalfa de regadío, en las 2 observaciones disponibles se verificó $R_{max} = R_{>MB}$.

En la Tabla 180 pueden verse los casos en que se observó $R_{>MB} = R_{max}$, para las respuestas de la clase I.

Tabla 180.

Casos en que $R_{>MB} = R_{max}$. Clase de respuestas I. Encuesta a los agricultores (1998 – 99, segundo día).

Cultivo	**Encuestas clase I**	**$R_{>MB} = R_{max}$**
Cebada secano	12	6
Trigo secano	7	5
Trigo regadío	4	2
Maíz regadío	4	2
Alfalfa regadío	2	2

De nuevo se han observado en general mayores diferencias relativas para los extremos inferiores que para los superiores.

Diferencias en los dos extremos del recorrido. Se analizaron simultáneamente las diferencias de los extremos inferiores y superiores, resumidas en la tabla 181.

Tabla 181.

Valores de los extremos inferiores R_{min}, $R_{<MM}$ y superiores R_{max}, $R_{>MB}$..Clase de respuesta I. Encuesta a los agricultores (1998 – 1999, segundo día).

Cultivo	**$R_{mm} = 0$**	**$R_{min} = R_{mm}$ $R_{max} = R_{MB}$**	**$R_{min} = R_{mm}$ $R_{max} > R_{MB}$**	**$R_{min} = R_{mm}$ $R_{max} < R_{MB}$**	**$R_{max} = R_{MB}$ $R_{min} > R_{mm}$**	**$R_{max} = R_{MB}$ $R_{min} < R_{mm}$**	**$R_{max} < R_{MB}$ $R_{min} > R_{mm}$**	**$R_{max} > R_{MB}$ $R_{min} < R_{mm}$**	**Total**

Cebada secano	1	4	0	0	1	0	5	1	12
Trigo secano	1	2	0	1	2	0	0	1	7
Trigo regadío	0	1	0	2	1	0	0	0	4
Maíz regadío	0	1	0	1	0	1	1	0	4
Alfalfa regadío	0	2	0	0	0	0	0	0	2

De las 12 respuestas analizadas en el caso de la cebada de secano, en 4 casos los agricultores daban los mismos valores de $R_{min} = R_{mm}$ y $R_{max} = R_{MB}$.

Se observó una clara tendencia a incrementar el recorrido cuando se definía en términos de lenguaje natural, ya que de los 7 casos en los que no eran iguales los extremos, existían 6 en los que $R_{min} \geq R_{mm}$ y $R_{max} \leq R_{MB}$.

En el caso del trigo de secano, se observó una clara tendencia a ampliar los recorridos al expresarlos en intervalos definidos mediante lenguaje natural (con $R_{min} \neq R_{mm}$ y $R_{max} \neq R_{MB}$), hecho observado en todos los agricultores excepto en uno.

Para trigo, maíz y alfalfa de regadío, se observó también la tendencia de ampliar los recorridos al expresarlos en intervalos definidos mediante lenguaje natural (con $R_{min} \neq R_{mm}$ y $R_{max} \neq R_{MB}$), para todos los agricultores.

#7.7. ANÁLISIS DE LOS RENDIMIENTOS NORMALES Y MÁS FRECUENTES

Mediante este análisis se pretende estimar la coherencia de las respuestas cuando se refieren a los rendimientos más frecuentes, en las estimaciones puntuales del tipo (R_{min}, R_{mf}, R_{max}), y los rendimientos declarados numéricamente como equivalentes a los definidos de forma ordinal como “normales”, y que hemos denominado R_{NN}.

El análisis se realiza para las respuestas de cada día.

#7.7.1.- RESPUESTAS DEL PRIMER DÍA

Se comparan los rendimientos normales (R_{NN}) y los rendimientos más frecuentes (R_{mf}), para las clases de respuestas I, II, IIa y IIb. Para ello, se comprobó si el rendimiento más frecuente se encontraba dentro del intervalo de rendimientos normales. Los resultados pueden consultarse en la tabla 182, donde se puede comprobar que gran parte de las estimaciones de rendimiento más frecuentes pertenecen al intervalo de rendimientos normales.

Tabla 182.

Respuestas R_{mf} pertenecientes al intervalo “rendimientos normales”. Clases de respuesta I, II, IIa y IIb. Encuesta a los agricultores (1998 – 99, primer día).

Cultivo	**Respuestas R_{mf} pertenecientes al intervalo**	**Total respuestas**	**Porcentaje (%)**

	normal		
Cebada secano	26	30	87
Trigo secano	17	18	94
Trigo regadío	7	8	88
Maíz regadío	10	12	83
Alfalfa regadío	7	9	78

Para completar el análisis anterior (aunque con datos de menor calidad), con las respuestas de las clases III, IV, IVa y IVb, se midió la diferencia entre los rendimientos normal (R_{NN}) y más frecuente (R_{mf}), expresada como porcentaje del rendimiento más frecuente, obteniendo los resultados de la Tabla 183, encontrándose diferencias pequeñas.

Tabla 183.

Cociente $|(R_{NN} - R_{mf}) / R_{mf}|$. Clases de respuesta III, IV, IV a y IV b. Encuesta a los agricultores (1998 – 99, primer día).

	Número de casos									
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano	12	2	1	1						
Trigo secano	5	1		1						
Trigo regadío	1	2	2							
Maíz regadío	6	1	1							
Alfalfa regadío	4	3	1							

En la Tabla 184 se expresan las coincidencias entre los rendimientos normal y más frecuente para cada cultivo y para las clases de respuesta III, IV, IVa y IVb. Las coincidencias son más numerosas en los cultivos de secano.

Tabla 184.

Número de casos en que $R_{NN} = R_{mf}$. Clases de respuesta III, IV, IVa y IVb. Encuesta a los agricultores (1998 – 99, primer día).

Cultivo	**$R_{NN} = R_{mf}$**	**Casos totales**	**Porcentaje (%)**
Cebada secano	10	16	62
Trigo secano	4	7	57
Trigo regadío	1	5	20
Maíz regadío	3	8	38
Alfalfa regadío	2	8	25

#7.7.2.- RESPUESTAS DEL SEGUNDO DÍA

La comparación de las respuestas R_{mf} (rendimientos más frecuentes) y los valores del intervalo de rendimientos “normales” (el intervalo central de las distribuciones estimadas), con los datos del segundo día, se ha sintetizado en la tabla 185.

Tabla 185.

Respuestas R_{mf} pertenecientes al intervalo normal. Clases de respuesta I, II, IIa y IIb. Encuesta a los agricultores (1998 – 99, segundo día).

Cultivo	**Respuestas pertenecientes al intervalo normal**	**Total respuestas**	**Porcentaje (%)**
Cebada secano	18	23	78
Trigo secano	11	15	73
Trigo regadío	6	7	86
Maíz regadío	6	9	67
Alfalfa regadío	3	6	50

Para las respuestas de las clases III, IV, IVa y IVb, se midió la diferencia entre los rendimientos normal (R_{NN}) y más frecuente (R_{mf}), expresada como porcentaje del rendimiento más frecuente, obteniendo los resultados de la Tabla 186.

Tabla 186.

Cociente $|(R_{NN}-R_{mf}):R_{mf}|$. Clases de respuesta III, IV, IVa y IVb. Encuesta a los agricultores (1998 – 99, segundo día).

	Número de casos									
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano (*)	17	3	1	1						
Trigo secano	8			1						
Trigo regadío	2	3	1							
Maíz regadío	7	2	2							
Alfalfa regadío (*)	6	2								

(*) Habiendo eliminado una observación por ser incoherente.

Observándose unas diferencias relativamente pequeñas entre los rendimientos considerados.

En la Tabla 187 se expresan las coincidencias entre los rendimientos normal y más frecuente para cada cultivo y para las clases de respuesta III, IV, IVa y IVb.

Tabla 187.

Número de casos en que $R_{NN} = R_{mf}$. Clases de respuesta III, IV, IVa y IVb. Encuesta a los agricultores (1998 – 99, segundo día).

Cultivo	**$R_{NN} = R_{mf}$**	**Casos totales**	**Porcentaje (%)**
Cebada secano (*)	15	22	68
Trigo secano	7	9	78
Trigo regadío	2	6	33
Maíz regadío	3	11	27
Alfalfa regadío (*)	5	8	62

(*) Habiendo eliminado una observación por ser incoherente.

Observándose unas coincidencias más elevadas en los cultivos de secano, igual que ocurría en el primer día, pero contrariamente a lo que ocurría en los expertos.

#7.8.- COHERENCIA DE LAS ESTIMACIONES

A partir de las estimaciones de la frecuencia de cada intervalo numérico y de las estimaciones puntuales de los valores (R_{max}, R_{mf}, R_{min}) es posible calcular los valores de las medias y de las varianzas respectivas, compararlas entre sí y con los valores de los rendimientos medios R_m estimados directamente por los agricultores (puntualmente). Llamando x_i al valor de la marca de clase del intervalo i en la estimación de frecuencias, y f_i la frecuencia correspondiente, en lo que sigue se comparan los siguientes valores medios:

$R_{xf} = (\Sigma\ x_i\ f_i)$ con los correspondientes valores medios obtenidos a partir de las estimaciones de (R_{max}, R_{mf}, R_{min}), para los dos días de la encuesta;

$R_{xf} = (\Sigma\ x_i\ f_i)$ con la estimación directa puntual del valor medio, R_m, para el segundo día de la encuesta;

$R_{xf} = (\Sigma\ x_i\ f_i)$ entre los días primero y segundo de la encuesta.

Una vez estudiadas las diferencias entre las estimaciones directas e indirectas de los valores medios, se realizará el correspondiente estudio para el caso de la varianza, mediante la comparación de

$V_{xf} = (x_i - R_{xf})^2\ f_i$ con las correspondientes varianzas obtenidas a partir de las estimaciones de (R_{max}, R_{mf}, R_{min});

$V_{xf} = (x_i - R_{xf})^2\ f_i$ entre los dos días de la encuesta.

#7.8.1- COHERENCIA DE LOS VALORES MEDIOS

En la tabla 188 se comparan las diferencias de la media calculada en base a las frecuencias atribuidas a cada intervalo $R_{xf} = (\Sigma\ x_i\ f_i)$, y las medias estimadas a partir de (R_{max}, R_{mf}, R_{min}) mediante las aproximaciones triangular y beta – PERT, obteniéndose diferencias del 10% como máximo en la mitad de los casos. Comparado con los expertos, se observan medianas inferiores en los agricultores para los cultivos de secano y la tendencia contraria para los de regadío. Se mantiene en general la tendencia de medianas inferiores para los cultivos de regadío que para los de secano.

Tabla 188.

Diferencias relativas de medias calculadas $|(R_{xf} - RMT):R_{xf}|$ y $|(R_{xf} - RMB):R_{xf}|$. Encuesta a los agricultores, primer día (1998-99).

APROXIMACIÓN												
TRIANGULAR	**Número de casos**											
Cultivo	**0-**	**0.05-**	**0.10-**	**0.15-**	**0.20-**	**0.25-**	**0.30-**	**0.35-**	**0.40-**	**0.45-**	**>0.5**	**Mediana**
Cebada secano	11	15	5	5	3	2			2	2	1	0.090
Trigo secano	7	6	4	2	3	1		2				0.096
Trigo regadío	4	3	3	1	1			1				0.092
Maíz regadío	9	5	2	2	1	1						0.060
Alfalfa regadío	5	4	3	2		1					1	0.088

BETA	**Número de casos**											
Cultivo	**0-**	**0.05-**	**0.10-**	**0.15-**	**0.20-**	**0.25-**	**0.30-**	**0.35-**	**0.40-**	**0.45-**	**>0.5**	**Mediana**
Cebada secano	12	11	9	8		2		1			3	0.100
Trigo secano	9	3	5	3	1	4						0.105
Trigo regadío	5	3	1	2	1				1			0.075
Maíz regadío	7	6	3	1	2		1					0.075
Alfalfa regadío	5	3	3	1	3						1	0.100

El examen de las poblaciones de valores medios estimados por ambos métodos viene resumido en la tabla 189, confirmando los buenos resultados anteriores, obteniendo sólo diferencias en el caso de la cebada de secano, en ambas comparaciones, triangular y beta - PERT.

Tabla 189.

Comparación entre las poblaciones de rendimientos medios (RMT con R_{xf} y RMB con R_{xf}). Encuesta a los agricultores, primer día (1998-99).

Comparación con la distribución triangular			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.916	No	Diferencias
Trigo secano	0.884	No	No diferencias
Trigo regadío	0.780	No	No diferencias
Maíz regadío	0.870	No	No diferencias
Alfalfa regadío	0.765	No	No diferencias
Comparación con la distribución beta			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.914	No	Diferencias
Trigo secano	0.922	No	No diferencias
Trigo regadío	0.736	No	No diferencias
Maíz regadío	0.863	No	No diferencias
Alfalfa regadío	0.749	No	No diferencias

En la tabla 190 se resume el resultado de las regresiones realizadas entre las diferentes estimaciones, obteniéndose resultados semejantes en ambas aproximaciones, con ajustes algo peores en trigo y alfalfa de regadío. Igual que para los expertos en los cultivos de secano, para estos cultivos con peores ajustes quizás pueda encontrarse algún método que mejore la coherencia. El mismo análisis para el caso b = 0, resumido en la tabla 191, parece confirmar claramente esta conjetura, obteniendo resultados altamente significativos para todos los cultivos y con unos coeficientes “a” mejores que en el caso anterior.

Tabla 190.

Regresión $RMT = a\ R_{xf} + b$ y $RMB = a\ R_{xf} + b$. Encuesta a los agricultores, primer día (1998-99).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.81	0.87	0.72	0.86	0.94
b	773.19	542.50	1486.88	1187.1	-6.71

R^2	0.87	0.86	0.60	0.99	0.66
Fc	296.2	136.1	16.3	1721.7	27.7
Ft	4.06	4.28	4.84	4.41	4.60
DISTRIBUCIÓN BETA					
a	0.83	0.91	0.71	0.86	0.98
b	716.42	390.95	1574.76	1277.1	-408.19
R^2	0.88	0.89	0.56	0.99	0.66
Fc	327.2	182.1	14.1	1261.8	27.4
Ft	4.06	4.28	4.84	4.41	4.60

Tabla 191.

Regresión $RMT = a\ R_{xf}$ ($b = 0$) y $RMB = a\ R_{xf}$ ($b = 0$). Encuesta a los agricultores, primer día (1998-99).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA**
a	1.03	1.00	1.00	0.92	0.94
b	0	0	0	0	0
R^2	0.98	0.98	0.99	0.99	0.95
Fc	2336.3	1354.7	816.2	2921.2	308.6
Ft	4.06	4.26	4.75	4.38	4.54
DISTRIBUCIÓN BETA					
a	1.04	1.01	1.00	0.92	0.96
b	0	0	0	0	0
R^2	0.98	0.99	0.98	0.99	0.95
Fc	2570.8	1573.7	734.4	2300.6	289.0
Ft	4.06	4.26	4.75	4.38	4.54

#7.8.2- COHERENCIA DE LAS VARIANZAS

En la tabla 192 se realiza la comparación de las varianzas calculadas a partir de la distribución de frecuencias $V_{xf} = (\Sigma\ (x_i - R_{xf})^2\ f_i)$, y las derivadas de las aproximaciones a funciones triangulares y beta-PERT, En general, los errores medidos mediante la mediana de las diferencias son superiores en el caso de las varianzas que en el anteriormente discutido de las medias, obteniéndose diferencias relativas inferiores en la aproximación beta – PERT, excepto en trigo de secano.

Tabla 192.

Diferencias relativas de las varianzas calculadas $|(V_{xf}\text{-VT}):V_{xf}|$ y $|(V_{xf}\text{-VB}):V_{xf}|$. Encuesta a los agricultores, primer día (1998-99).

APROXIMACIÓN											
TRIANGULAR	**Número de casos**										
Cultivo	**0-**	**0.1-**	**0.2-**	**0.3-**	**0.4-**	**0.5-**	**0.6-**	**0.7-**	**0.8-**	**0.9-**	**>1.0**	**Mediana**
Cebada secano	1	4	3	2	3	9	9	8	6	1		0.606
Trigo secano	2	1	2	2	3	4		6	4	1		0.531
Trigo regadío			1		1	2	4	3	1	1		0.631
Maíz regadío		1	2	2	1	6	3	2	2		1	0.533

Alfalfa regadío	1	2	1	1	3	2	1	2		3		0.500
BETA	**Número de casos**											
Cultivo	**0-**	**0.1-**	**0.2-**	**0.3-**	**0.4-**	**0.5-**	**0.6-**	**0.7-**	**0.8-**	**0.9-**	**>1.0**	**Mediana**
Cebada secano	3	3	8	5	7	5	7	2	6			0.428
Trigo secano	2	3	4	2		3	5	2	2	1	1	0.550
Trigo regadío		1	1	2	2	3	2	1	1			0.508
Maíz regadío	3	2	3	3	1	3	2	2			1	0.333
Alfalfa regadío	1	5		2	1	2	1	1	2	1		0.400

En la tabla 193 se resume el estudio de las poblaciones de varianzas obtenidas por los diferentes métodos, obteniéndose malos resultados para ambas aproximaciones en todos los cultivos.

Tabla 193.

Comparación entre las poblaciones de varianzas (VT con V_{xf} y VB con V_{xf}). Encuesta a los agricultores, primer día (1998-99).

Comparación con la distribución triangular			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.783	No	Diferencias
Trigo secano	0.678	No	Diferencias
Trigo regadío	0.908	No	Diferencias
Maíz regadío	0.872	No	Diferencias
Alfalfa regadío	0.864	No	Diferencias
Comparación con la distribución beta			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.774	No	Diferencias
Trigo secano	0.685	No	Diferencias
Trigo regadío	0.908	No	Diferencias
Maíz regadío	0.849	No	Diferencias
Alfalfa regadío	0.865	No	Diferencias

Las tablas 194 y 195 muestran los resultados de realizar la regresión de las diferentes estimaciones. Se obtienen mejores ajustes en los cultivos de regadío, aunque las diferencias obtenidas son grandes excepto en maíz de regadío, tanto para b igual o distinta de cero.

Tabla 194.

Regresión $VT = a\ V_{xf} + b$ y $VB = a\ V_{xf} + b$. Encuesta a los agricultores, primer día (1998-99).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.32	0.34	0.47	1.08	0.37
b	78524.8	118435.04	-129394.83	-2506638.31	656211.31
R^2	0.44	0.36	0.90	0.94	0.72
Fc	35.1	13.0	104.9	263.4	36.9
Ft	4.06	4.28	4.84	4.41	4.60

DISTRIBUCIÓN BETA					
a	0.49	0.52	0.71	1.63	0.57
b	123398.42	180480.66	-193254.53	-3711090.70	1047690.19
R^2	0.44	0.37	0.91	0.94	0.71
Fc	35.2	13.67	112.2	274.9	34.4
Ft	4.06	4.28	4.84	4.41	4.60

Tabla 195.

Regresión $VT = a\ V_{xf}$ (b = 0) y $VB = a\ V_{xf}$ (b = 0). Encuesta a los agricultores, primer día (1998-99).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	TRIGO SECANO	TRIGO REGADÍO	MAÍZ REGADÍO	ALFALFA REGADÍO
a	0.38	0.43	0.42	0.94	0.42
b	0	0	0	0	0
R^2	0.76	0.72	0.95	0.91	0.87
Fc	138.3	62.3	210.2	183.2	95.1
Ft	4.06	4.26	4.75	4.38	4.54
DISTRIBUCIÓN BETA					
a	0.58	0.66	0.63	1.42	0.65
b	0	0	0	0	0
R^2	0.76	0.73	0.95	0.91	0.86
Fc	139.6	64.9	224.2	190.9	89.3
Ft	4.06	4.26	4.75	4.38	4.54

#7.8.3.- COHERENCIA DE LOS VALORES MEDIOS EN EL SEGUNDO DÍA

En la tabla 196 se comparan las diferencias de la media calculada en base a las frecuencias atribuidas a cada intervalo $R_{xf} = (\Sigma\ x_i\ f_i)$, y las medias estimadas a partir de (R_{max}, R_{mf}, R_{min}) mediante las aproximaciones triangular y beta – PERT, obteniéndose resultados semejantes para ambos métodos, con diferencias relativas inferiores en trigo de secano y de regadío.

Tabla 196.

Diferencias relativas de medias calculadas $|(R_{xf}$ - RMT):$R_{xf}|$ y $|(R_{xf}$ -RMB):$R_{xf}|$. Encuesta a los agricultores, segundo día (1998-99).

APROXIMACIÓN												
TRIANGULAR	**Número de casos**											
Cultivo	**0-**	**0.05-**	**0.10-**	**0.15-**	**0.20-**	**0.25-**	**0.30-**	**0.35-**	**0.40-**	**0.45-**	**>0.5**	**Mediana**
Cebada secano	14	9	8	3	4	3	3				2	0.100
Trigo secano	11	3	2	4	2	1						0.058
Trigo regadío	5	4	2					1			1	0.068
Maíz regadío	5	4	4	3		1	1			1	1	0.112
Alfalfa regadío*												

BETA	**Número de casos**											
Cultivo	**0-**	**0.05-**	**0.10-**	**0.15-**	**0.20-**	**0.25-**	**0.30-**	**0.35-**	**0.40-**	**0.45-**	**>0.5**	**Mediana**
Cebada secano	11	8	10	9	3	1	2			1	1	0.120
Trigo secano	9	4	4	4		2						0.081
Trigo regadío	8	1	2					1			1	0.041
Maíz regadío	5	5	3	3			1			2	1	0.100
Alfalfa regadío*												

(*) No se pudo incluir la alfalfa debido a la imposibilidad de reconstruir los intervalos, al tener sólo dos observaciones de la clase I.

El examen de las poblaciones de valores medios estimados por ambos métodos, resumido en la tabla 197, parece confirmar la tendencia observada anteriormente de diferencias menores en trigo de secano y de regadío.

Tabla 197.

Comparación entre las poblaciones de rendimientos medios (RMT con R_{xf} y RMB con R_{xf}). Encuesta a los agricultores, segundo día (1998-99).

Comparación con la distribución triangular			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.888	No	Diferencias
Trigo secano	0.923	No	No diferencias
Trigo regadío	0.488	No	No diferencias
Maíz regadío	0.789	No	Diferencias
Alfalfa regadío*			
Comparación con la distribución beta			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.887	No	Diferencias
Trigo secano	0.941	No	No diferencias
Trigo regadío	0.504	No	No diferencias
Maíz regadío	0.771	No	Diferencias
Alfalfa regadío*			

(*) No se pudo incluir la alfalfa debido a la imposibilidad de reconstruir los intervalos, al tener sólo dos observaciones de la clase I.

En la tabla 198 se resume el resultado de las regresiones realizadas entre las diferentes estimaciones. Contrariamente a lo observado anteriormente respecto al trigo de regadío, en este caso el ajuste no es demasiado bueno y el coeficiente “a” tampoco. En los demás cultivos el ajuste es mejor. El mismo análisis para el caso b = 0 viene resumido en la tabla 199, observándose buenos ajustes para todos los cultivos, con diferencias pequeñas.

Tabla 198.

Regresión $RMT = a\ R_{xf} + b$ y $RMB = a\ R_{xf} + b$. Encuesta a los agricultores, segundo día (1998-99).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO***
a	0.85	0.95	0.60	0.84	
b	614.9	250.70	1751.11	692.11	

R^2	0.85	0.88	0.23	0.97	
Fc	252.9	152.3	3.22	613.4	
Ft	4.06	4.28	4.84	4.41	
DISTRIBUCIÓN BETA					
a	0.85	0.98	0.57	0.84	
b	625.27	130.84	1886.8	541.74	
R^2	0.86	0.89	0.22	0.96	
Fc	266.4	173.1	3.15	395.2	
Ft	4.06	4.28	4.84	4.41	

(*) No se pudo incluir la alfalfa debido a la imposibilidad de reconstruir los intervalos, al tener sólo dos observaciones de la clase I.

Tabla 199.

Regresión $RMT = a\ R_{xf}$ (b = 0) y $RMB = a\ R_{xf}$ (b = 0). Encuesta a los agricultores, segundo día (1998-99).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO***
a	1.03	1.01	0.93	0.87	
b	0	0	0	0	
R^2	0.98	0.99	0.96	0.99	
Fc	2458.3	1523.3	261.9	1505.3	
Ft	4.06	4.26	4.75	4.38	
DISTRIBUCIÓN BETA					
a	1.03	1.01	0.93	0.87	
b	0	0	0	0	
R^2	0.98	0.99	0.96	0.98	
Fc	2553.17	1667.5	273.9	1000.1	
Ft	4.06	4.26	4.75	4.38	

(*) No se pudo incluir la alfalfa debido a la imposibilidad de reconstruir los intervalos, al tener sólo dos observaciones de la clase I.

Las comparaciones que siguen se realizan entre el valor de los rendimientos medios estimados directamente por los agricultores (puntualmente), R_m, y los calculados a partir de sus estimaciones de frecuencias y valores numéricos de los correspondientes intervalos. Únicamente se pudo realizar dicha comparación en el caso del segundo día del curso 98-99, debido a que el primer día no se pidió la estimación del rendimiento medio a los agricultores.

Tabla 200.

COCIENTE $|(R_{xf}-R_m):R_{xf}|$. Encuesta a los agricultores, segundo día (1998-99).

	Número de casos											
Cultivo	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**>0.5**	**Mediana**
Cebada secano	19	11	4	5	3	2	1				1	0.068
Trigo secano	6	9	3	2	2	1						0.081
Trigo regadío	8	2	2									0.038
Maíz regadío	5	8	2	4	1							0.081
Alfalfa regadío*												

(*) No se pudo incluir la alfalfa debido a la imposibilidad de reconstruir los intervalos, al tener sólo dos observaciones de la clase I.

Obteniéndose en general diferencias relativas inferiores que en el caso de las comparaciones entre los rendimientos R_{xf} y RMF o RMB.

La tabla 201 señala la semejanza entre las dos poblaciones de valores medios, mientras que en las tablas 202 y 203 se reflejan los resultados de realizar una regresión entre ambas estimaciones. Excepto para la cebada de secano, se observan buenos resultados entre poblaciones. A pesar de ello, en las regresiones se obtienen unos ajustes buenos en todos los cultivos cuando b es nula.

Tabla 201.

Comparación entre las poblaciones de rendimientos medios (R_{xf} con R_m). Encuesta a los agricultores, segundo día (1998-99).

Comparación			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.899	No	Diferencias
Trigo secano	0.933	No	No diferencias
Trigo regadío	0.607	No	No diferencias
Maíz regadío	0.846	No	No diferencias
Alfalfa regadío (*)			

(*) No se pudo incluir la alfalfa debido a la imposibilidad de reconstruir los intervalos, al tener sólo dos observaciones de la clase I.

Tabla 202.

Regresión $R_m = a\, R_{xf} + b$. Encuesta a los agricultores, segundo día (1998-99).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO***
a	0.92	1.00	0.73	0.82	
b	393.45	48.62	1512.47	1689.85	
R^2	0.89	0.91	0.41	0.98	
Fc	362.8	217.8	7.7	1017.7	
Ft	4.06	4.28	4.84	4.41	

(*) No se pudo incluir la alfalfa debido a la imposibilidad de reconstruir los intervalos, al tener sólo dos observaciones de la clase I.

Tabla 203.

Regresión $R_m = a\, R_{xf}$ (b = 0). Encuesta a los agricultores, segundo día (1998-99).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO***
a	1.03	1.01	1.02	0.89	
b	0	0	0	0	
R^2	0.99	0.99	0.98	0.99	
Fc	3425.4	2028.9	493.6	1631.8	
Ft	4.06	4.26	4.75	4.38	

(*) No se pudo incluir la alfalfa debido a la imposibilidad de reconstruir los intervalos, al tener sólo dos observaciones de la clase I.

#7.8.4- COHERENCIA DE LAS VARIANZAS EN EL SEGUNDO DÍA

En la tabla 204 se realiza la comparación de las varianzas calculadas a partir de la distribución de frecuencias $V_{xf} = (\Sigma\ (x_i - R_{xf})^2\ f_i)$, y las derivadas de las aproximaciones a funciones triangulares y beta-PERT. Los errores medidos mediante la mediana de las diferencias son superiores en el caso de las varianzas que en el anteriormente discutido de las medias, y mayores en el caso de la comparación con la aproximación triangular, igual que en el primer día.

Tabla 204.

Diferencias relativas de las varianzas calculadas $|(V_{xf}\text{-VT}):V_{xf}|$ y $|(V_{xf}\text{-VB}):V_{xf}|$. Encuesta a los agricultores, segundo día (1998-99).

APROXIMACIÓN TRIANGULAR	**Número de casos**											
Cultivo	**0-**	**0.1-**	**0.2-**	**0.3-**	**0.4-**	**0.5-**	**0.6-**	**0.7-**	**0.8-**	**0.9-**	**>1.0**	**Mediana**
Cebada secano	2	5		3	3	7	7	9	7	2	1	0.621
Trigo secano	2		1	2	2	6	3	4	1	1	1	0.538
Trigo regadío		1	2			2	2	3	1	2		0.638
Maíz regadío			1	3	2	3	2	4	3	1	1	0.650
Alfalfa regadío*												
BETA	**Número de casos**											
Cultivo	**0-**	**0.1-**	**0.2-**	**0.3-**	**0.4-**	**0.5-**	**0.6-**	**0.7-**	**0.8-**	**0.9-**	**>1.0**	**Mediana**
Cebada secano	3	1	6	6	6	7	6	5	4	1	1	0.507
Trigo secano	3		1	6	2	5	2	1		2	1	0.438
Trigo regadío		2	1	2	2	1	2	1		1	1	0.438
Maíz regadío	3	2	2	2		3	2	3		1	2	0.533
Alfalfa regadío*												

(*) No se pudo incluir la alfalfa debido a la imposibilidad de reconstruir los intervalos, al tener sólo dos observaciones de la clase I.

En la tabla 205 se resume el estudio de las poblaciones de varianzas obtenidas por los diferentes métodos, viéndose malos resultados en todos los cultivos y ambas aproximaciones.

Tabla 205.

Comparación entre las poblaciones de varianzas (VT con V_{xf} y VB con V_{xf}). Encuesta a los agricultores, segundo día (1998-99).

Comparación con la distribución triangular			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.634	No	Diferencias
Trigo secano	0.687	No	Diferencias
Trigo regadío	0.791	No	Diferencias
Maíz regadío	0.795	No	Diferencias
Alfalfa regadío*			

Comparación con la distribución beta			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.639	No	Diferencias
Trigo secano	0.709	No	Diferencias
Trigo regadío	0.795	No	Diferencias
Maíz regadío	0.792	No	Diferencias
Alfalfa regadío*			

(*) No se pudo incluir la alfalfa debido a la imposibilidad de reconstruir los intervalos, al tener sólo dos observaciones de la clase I.

Las tablas 206 y 207 muestran los resultados de realizar la regresión de las diferentes estimaciones, obteniéndose malos ajustes excepto en maíz de regadío, aunque con diferencias considerables medidas mediante el coeficiente "a".

Tabla 206.

Regresión $VT = a\ V_{xf} + b$ y $VB = a\ V_{xf} + b$. Encuesta a los agricultores, segundo día (1998-99).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO***
a	0.30	0.35	0.11	0.63	
b	72876.79	96767.08	360508.39	-867435.64	
R^2	0.46	0.56	0.14	0.99	
Fc	37.8	26.9	1.78	1714.8	
Ft	4.06	4.28	4.84	4.41	
DISTRIBUCIÓN BETA					
a	0.46	0.54	0.17	0.95	
B	114657.86	146127.37	550542.67	-910410.68	
R^2	0.46	0.57	0.14	0.98	
Fc	37.6	27.76	1.78	953.9	
Ft	4.06	4.28	4.84	4.41	

(*) No se pudo incluir la alfalfa debido a la imposibilidad de reconstruir los intervalos, al tener sólo dos observaciones de la clase I.

Tabla 207.

Regresión $VT = a\ V_{xf}$ ($b = 0$) y $VB = a\ V_{xf}$ ($b = 0$). Encuesta a los agricultores, segundo día (1998-99).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO***
a	0.36	0.42	0.22	0.61	
b	0	0	0	0	
R^2	0.76	0.81	0.47	0.98	
Fc	141.7	91.1	10.5	1268.2	
Ft	4.06	4.26	4.75	4.38	
DISTRIBUCIÓN BETA					
a	0.54	0.63	0.33	0.92	
b	0	0	0	0	
R^2	0.76	0.81	0.47	0.98	
Fc	142.4	93.62	10.6	1009.7	
Ft	4.06	4.26	4.75	4.38	

(*) No se pudo incluir la alfalfa debido a la imposibilidad de reconstruir los intervalos, al tener sólo dos observaciones de la clase I.

#7.8.5. COHERENCIA ENTRE LOS DÍAS PRIMERO Y SEGUNDO. VALORES MEDIOS

Las comparaciones que siguen se realizan entre el valor de los rendimientos medios calculados a partir de sus estimaciones de frecuencias y valores numéricos de los correspondientes intervalos, para los dos días de la encuesta.

En primer lugar, se realiza en la tabla 208 la comparación relativa entre ambos rendimientos medios, obteniendo unas diferencias máximas del 4.6% para la mitad de la población, indicando un alto grado de coherencia en todos los cultivos.

Tabla 208.

COCIENTE $|(R_{xf1} – R_{xf2}):R_{xf1}|$. Encuesta a los agricultores, días primero y segundo (1998-99).

	Número de casos											
Cultivo	**0-**	**0.05-**	**0.10-**	**0.15-**	**0.20-**	**0.25-**	**0.30-**	**0.35-**	**0.40-**	**0.45-**	**>0.5**	**Mediana**
Cebada secano	25	11	5	1	2			1			1	0.046
Trigo secano	16	5	1	1								0.036
Trigo regadío	7	4	1		1							0.046
Maíz regadío	14	4	2									0.036
Alfalfa regadío*												

(*) No se pudo incluir la alfalfa debido a la imposibilidad de reconstruir los intervalos, al tener sólo dos observaciones de la clase I.

R_{xf1} y R_{xf2} corresponden a los valores calculados para los días 1 y 2, respectivamente.

En la tabla 209 pueden observarse las comparaciones entre poblaciones de rendimientos medios, observándose unos buenos resultados para todos los cultivos, confirmando la coherencia anterior.

Tabla 209.

Comparación entre las poblaciones de rendimientos medios (R_{xf1} con R_{xf2}). Encuesta a los agricultores, días primero y segundo (1998-99).

Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.944	No	No diferencias
Trigo secano	0.972	No	No diferencias
Trigo regadío	0.918	No	No diferencias
Maíz regadío	0.965	No	No diferencias
Alfalfa			

(*) No se pudo incluir la alfalfa debido a la imposibilidad de reconstruir los intervalos, al tener sólo dos observaciones de la clase I.

En las tablas 210 y 211 se realizan las regresiones entre los rendimientos medios de ambos días, confirmándose de nuevo los buenos resultados de coherencia para todos los cultivos.

Tabla 210.

Regresión $R_{xf1} = a R_{xf2} + b$. Encuesta a los agricultores, días primero y segundo (1998-99).

	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO***
a	0.99	1.03	0.97	0.96	
b	-29.61	-70.30	141.47	224.63	
R^2	0.91	0.96	0.82	0.99	
Fc	441.2	481.1	50.3	10044.9	
Ft	4.06	4.28	4.84	4.41	

(*) No se pudo incluir la alfalfa debido a la imposibilidad de reconstruir los intervalos, al tener sólo dos observaciones de la clase I.

Tabla 211.

Regresión $R_{xf1} = a R_{xf2}$ (b = 0). Encuesta a los agricultores, días primero y segundo (1998-99).

	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA**
a	0.99	1.01	1.00	0.97	
b	0	0	0	0	
R^2	0.99	0.99	0.99	0.99	
Fc	3703.9	4196.7	1941.1	22848.5	
Ft	4.06	4.26	4.75	4.38	

(*) No se pudo incluir la alfalfa debido a la imposibilidad de reconstruir los intervalos, al tener sólo dos observaciones de la clase I.

#7.8.6. COHERENCIA ENTRE LOS DÍAS PRIMERO Y SEGUNDO. VARIANZAS

Las comparaciones que siguen se realizan entre el valor de las varianzas de los rendimientos calculados a partir de sus estimaciones de frecuencias y valores numéricos de los correspondientes intervalos, para los dos días de la encuesta.

En la tabla 212 pueden observarse las diferencias relativas para ambas varianzas, referidas a la varianza del primer día. Se observan mayores diferencias en cultivos de secano.

Tabla 212.

COCIENTE $|(V_{f1}-V_{f2}):V_{f1}|$. Encuesta a los agricultores, días primero y segundo (1998-99).

	Número de casos											
Cultivo	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**>0.5**	**Mediana**
Cebada secano	17	2	6	1	7	3	2	1	2	2	3	0.133
Trigo secano	8	2	5	1		1	1		1		4	0.115
Trigo regadío	6	3			1	1			1	1		0.058
Maíz regadío	11	2	3		1			2	1			0.045
Alfalfa regadío*												

(*) No se pudo incluir la alfalfa debido a la imposibilidad de reconstruir los intervalos, al tener sólo dos observaciones de la clase I.

V_{f1} y V_{f2} corresponden a los valores calculados para los días 1 y 2, respectivamente.

En la tabla 213 se comparan las poblaciones de varianzas entre los dos días, obteniendo muy buenos resultados en todos los cultivos, indicando coherencia.

Tabla 213.

Comparación entre las poblaciones de varianzas (V_{f1} con V_{f2}). Encuesta a los agricultores, días primero y segundo (1998-99).

Cultivo	Spearman	Independencia	Wilcoxon muestras relacionadas (α=0.5)
Cebada secano	0.740	No	No diferencias
Trigo secano	0.769	No	No diferencias
Trigo regadío	0.835	No	No diferencias
Maíz regadío	0.878	No	No diferencias
Alfalfa regadío*			

(*) No se pudo incluir la alfalfa debido a la imposibilidad de reconstruir los intervalos, al tener sólo dos observaciones de la clase I.

En las tablas 214 y 215 se realizan las regresiones entre las varianzas del primer día y del segundo, para b igual o distinta de cero, obteniendo en general buenos ajustes, sobre todo en éste último caso.

Tabla 214.

Regresión $V_{f1} = a\ V_{f2} + b$. Encuesta a los agricultores, días primero y segundo (1998-99).

	CEBADA SECANO	TRIGO SECANO	TRIGO REGADÍO	MAÍZ REGADÍO	ALFALFA REGADÍO*
a	0.73	0.69	0.59	0.55	
b	214098.70	229874.68	671582.21	1941506.44	
R^2	0.60	0.67	0.66	0.95	
Fc	66.4	43.3	21.8	332.2	
Ft	4.06	4.28	4.84	4.41	

(*) No se pudo incluir la alfalfa debido a la imposibilidad de reconstruir los intervalos, al tener sólo dos observaciones de la clase I.

Tabla 215.

Regresión $V_{f1} = a\ V_{f2}$ (b = 0). Encuesta a los agricultores, días primero y segundo (1998-99).

	CEBADA SECANO	TRIGO SECANO	TRIGO REGADÍO	MAÍZ REGADÍO	ALFALFA REGADÍO*
a	0.88	0.84	0.78	0.60	
b	0	0	0	0	
R^2	0.85	0.87	0.82	0.94	
Fc	251.2	143.9	53.64	284.9	
Ft	4.06	4.26	4.75	4.38	

(*) No se pudo incluir la alfalfa debido a la imposibilidad de reconstruir los intervalos, al tener sólo dos observaciones de la clase I.

CAPÍTULO 8

ENCUESTAS GENERALES A LOS AGRICULTORES. ENCUESTA DEL SEGUNDO AÑO (1999 – 2000)

#8.1.- PLANTEAMIENTO

De igual forma que en el curso académico 1998 – 1999, en el 1999 – 2000, se repitieron las encuestas a los agricultores, siendo los encuestadores estudiantes de los cursos terminales de segundo ciclo de las carreras de Ingeniero Agrónomo e Ingeniero de Montes. Los encuestadores seleccionaron y entrevistaron a agricultores con los que mantenían algún tipo de relación (familiar, de amistad, de vecindad, etc.), generalmente entre Noviembre y Enero de cada curso escolar.

La encuesta tenía como objetivo evaluar la coherencia de las respuestas sobre rendimientos de diferentes cultivos, obtenidas mediante encuestas convencionales, realizadas por personal sin una preparación especializada en los temas objeto de estudio. La versión del segundo año se reproduce en el Anejo 8, y estaba estructurada en dos cuestionarios, sometiéndose al agricultor a cada cuestionario en días distintos.

El primer cuestionario o "Encuesta del primer día", de la misma forma que la del curso académico 1998 – 1999, tenía por objeto determinar la función de densidad de los rendimientos de algunos cultivos. El segundo cuestionario o "Encuesta del segundo día" tenía por objeto obtener información adicional para juzgar la persistencia y coherencia de las respuestas obtenidas en ambos días. Cada agricultor era entrevistado con el primer cuestionario por un estudiante, y al menos una semana después, otro estudiante distinto realizaba con el segundo cuestionario otra entrevista al mismo agricultor. Durante el curso 1999 – 2000 se realizaron 44 encuestas en total.

La revisión de las respuestas obtenidas el primer año puso de manifiesto distintos problemas, y originó que se modificaran los cuestionarios para las encuestas realizadas durante el curso 1999 – 2000. Esta encuesta modificada se reproduce en el Anejo 8.

La modificación de las encuestas consistió en:

Pregunta 2.- En la encuesta del primer año no se incluyó el rendimiento medio en la estimación de los rendimientos máximo, mínimo y más frecuente. Se decidió incluir esa estimación puntual en la encuesta del segundo año.

Preguntas 5 y 6.- Dado que en el primer año no se consiguió que se entendieran estas preguntas, se decidió incrementar las instrucciones para los estudiantes.

Otras cuestiones.- Se trataba básicamente de cuestiones no analizadas en esta encuesta, relacionadas con la "receta" productiva y su estabilidad, que se han pospuesto para estudios posteriores.

#8.2.- RESULTADOS

Las características de las 44 explotaciones de los agricultores entrevistados durante el curso 1999 - 00 se han esquematizado en las Tablas 216 a 218.

Tabla 216.

Características de las explotaciones de los agricultores entrevistados (1999 – 00).

			Superficie secano (ha)			Superficie regadío (ha)			Superficie		Otras
Nº	**Población**	**Provincia**	**Herbáceos**	**Leñosos**	**Total**	**Herbáceos**	**Leñosos**	**Total**	**Total (ha)**	**Ganado**	**actividades**
36	Balones	Alacant	0	4	4	0	0	0	4		
5	Los Anguijes	Albacete	25	0	25	14	0	14	39		AP
3	Mataró	Barcelona	0	0	0	4,8	0	4,8	4,8		
9	Sallent	Barcelona	57,9	0	57,9	0	0	0	57,9	Por	FOR
17	Gallecs	Barcelona	60	0	60	20	0	20	80		ES
26	Tiana	Barcelona	0	0	0	1,5	0	1,5	1,5		
10	La Guàrdia de Sagàs	Barcelona	67	0	67	29	0	29	96	Bo,Por, Av	AGROT
16	Miranda de Ebro	Burgos	100	0	100	13	0	13	113		
27	Medina de Pomar	Burgos	80	0	80	3	0	3	83		
6	Castelló de la Plana	Castelló	0	0	0	0	6,8	6,8	6,8		
21	La Vall d'Uixó	Castelló	0	0	0	0	4	4	4		
34	Canalejas del Arroyo	Cuenca	32	6	38	0	0	0	38		
22	La Pinya	Girona	7	0	7	5	0	5	12	Bo	
25	Ribes de Freser	Girona	26	0	26	0	0	0	26	Bo	
1	Fraga	Huesca	18	0	18	0	0	0	18		
4	Quicena	Huesca	43	0	43	8,8	0	8,8	51,8	Ov	
12	Las Casas	Huesca	11	0	11	12	0	12	23		
23	Betesa	Huesca	28	0	28	0	0	0	28	Ov	GA
24	Huesca	Huesca	100	0	100	67	0	67	167	Ov	
33	San Esteban de Litera	Huesca	6	0	6	13	0	13	19		
40	Novales	Huesca	72	0	72	12	0	12	84		
43	Huesca	Huesca	25	0	25	6	0	6	31		
15	Logroño	La Rioja	0	5	5	1,3	3,4	4,7	9,7	Por,Av	
41	Los Molinos de Ocón	La Rioja	0	10,8	10,8	0,4	0	0,4	11,2		GA AUT
11	Albesa	Lleida	25,3	0	25,3	11,5	9,6	21,1	46,4	Por	
31	Alcoletge	Lleida	0	0	0	5	0	5	5		
32	Torrebesses	Lleida	0	3	3	0	0	0	3		
37	Sunyer	Lleida	33	0	33	0	0,4	0,4	33,4		
39	Tornabous	Lleida	0	0	0	0	0	0	0		
13	Tàrrega	Lleida	36,3	0	36,3	13,8	0	13,8	50,1		
18	Vila-Sana	Lleida	0	0	0	0	6,1	6,1	6,1		
2	Ossó de Sió	Lleida	43,7	1,7	45,4	0	0	0	45,4		
7	Corçà	Lleida	307	7,5	314,5	0	5	5	319,5		
38	Santiago de la Puebla	Salamanca	76,4	0	76,4	0	0	0	76,4		
28	Tejado	Soria	20	0	20	20	0	20	40		
35	Tasahuerce	Soria	150	0	150	0	0	0	150	Ov	
8	Miravet	Tarragona	0	1,7	1,7	0	0	0	1,7		VIN

Nº	Población	Provincia	Superficie secano (ha)			Superficie regadío (ha)			Superficie	Ganado	Otras
			Herbáceos	Leñosos	Total	Herbáceos	Leñosos	Total	Total (ha)		actividades
	d'Ebre										
30	Amposta	Tarragona	0	0	0	200	0	200	200		
14	Cullera	Valencia	0	0	0	7	0	7	7		
42	Calaceit	Valencia	0	34	34	0	0	0	34		
44	Ròtova	València	0	0	0	0	1,3	1,3	1,3		
19	Lechón	Zaragoza	195	0	195	0	0	0	195		
20	San Mateo de Gállego	Zaragoza	66	0	66	14,5	0	14,5	80,5	Bo	
29	Castejón de Alarba	Zaragoza	11	1	12	0	6,5	6,5	18,5		

Donde **Ov** indica ovino, **Por** porcino, **Bo** bovino, **Av** avícola, **FOR** forestal, **ES** empresa de servicios, **AP** arriendo de pastos, **AGROT** agroturismo, **GA** ganado, **GA AUT** ganado para autoconsumo y **VIN** empresa vinícola.

Tabla 217.

Superficies (ha) de los principales tipos de cultivos de secano en las explotaciones encuestadas (1999 –00).

ENCUESTA	SECANO						
NÚMERO	Cereales	Forrajes	Oleaginosas	Otros	Herbáceos	Leñosos	TOTAL
5	25	0	0	0	25	0	25
36	0	0	0	0	0	4	4
3	0	0	0	0	0	0	0
9	48,9	0	0	9	57,9	0	57,9
17	30	0	30	0	60	0	60
26	0	0	0	0	0	0	0
10	27	0	0	40	67	0	67
16	100	0	0	0	100	0	100
27	80	0	0	0	80	0	80
6	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0
34	20	0	12	0	32	6	38
22	7	0	0	0	7	0	7
25	0	26	0	0	26	0	26
1	18	0	0	0	18	0	18
4	33	6	4	0	43	0	43
12	0	7	0	4	11	0	11
23	0	28	0	0	28	0	28
24	100	0	0	0	100	0	100
33	6	0	0	0	6	0	6
40	72	0	0	0	72	0	72
43	25	0	0	0	25	0	25
15	0	0	0	0	0	5	5
41	0	0	0	0	0	10,8	10,8
2	43,7	0	0	0	43,7	1,7	45,4
7	25	282	0	0	307	7,5	314,5
11	25,3	0	0	0	25,3	0	25,3
13	36,3	0	0	0	36,3	0	36,3

ENCUESTA NÚMERO	SECANO						
	Cereales	Forrajes	Oleaginosas	Otros	Herbáceos	Leñosos	TOTAL
18	0	0	0	0	0	0	0
31	0	0	0	0	0	0	0
32	0	0	0	0	0	3	3
37	33	0	0	0	33	0	33
39	0	0	0	0	0	0	0
38	76,4	0	0	0	76,4	0	76,4
28	20	0	0	0	20	0	20
35	150	0	0	0	150	0	150
8	0	0	0	0	0	1,7	1,7
30	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0
42	0	0	0	0	0	34	34
44	0	0	0	0	0	0	0
19	95	0	100	0	195	0	195
20	66	0	0	0	66	0	66
29	11	0	0	0	11	1	12

Tabla 218.

Superficies (ha) de los principales tipos de cultivos en las explotaciones de regadío encuestadas (1999 – 00).

ENCUESTA NÚMERO	REGADÍO						
	Cereales	Forrajes y pastos	Horticultura e industria	Otros	Herbáceos	Leñosos	TOTAL
5	14	0	0	0	14,0	0,0	14,0
36	0	0	0	0	0,0	0,0	0,0
3	0	0	4,8	0	4,8	0,0	4,8
9	0	0	0	0	0,0	0,0	0,0
17	10	0	10	0	20,0	0,0	20,0
26	0	0	0	1,5	1,5	0,0	1,5
10	0	0	0	29	29,0	0,0	29,0
16	0	0	1	12	13,0	0,0	13,0
27	0	0	3	0	3,0	0,0	3,0
6	0	0	0	0	0,0	6,8	6,8
21	0	0	0	0	0,0	4,0	4,0
34	0	0	0	0	0,0	0,0	0,0
22	5	0	0	0	5,0	0,0	5,0
25	0	0	0	0	0,0	0,0	0,0
1	0	0	0	0	0,0	0,0	0,0
4	3,8	3	0	2	8,8	0,0	8,8
12	0	12	0	0	12,0	0,0	12,0
23	0	0	0	0	0,0	0,0	0,0
24	27	40	0	0	67,0	0,0	67,0
33	13	0	0	0	13,0	0,0	13,0
40	12	0	0	0	12,0	0,0	12,0
43	4	2	0	0	6,0	0,0	6,0
15	0	0	0,3	1	1,3	3,4	4,7
41	0	0	0	0,4	0,4	0,0	0,4
2	0	0	0	0	0,0	0,0	0,0

ENCUESTA NÚMERO	REGADÍO						
	Cereales	Forrajes y pastos	Horticultura e industria	Otros	Herbáceos	Leñosos	TOTAL
7	0	0	0	0	0,0	5,0	5,0
11	11,5	0	0	0	11,5	9,6	21,1
13	9,1	4,8	0	0	13,8	0,0	13,8
18	0	0	0	0	0,0	6,1	6,1
31	5	0	0	0	5,0	0,0	5,0
32	0	0	0	0	0,0	0,0	0,0
37	0	0	0	0	0,0	0,4	0,4
39	0	0	0	0	0,0	0,0	0,0
38	0	0	0	0	0,0	0,0	0,0
28	20	0	0	0	20,0	0,0	20,0
35	0	0	0	0	0,0	0,0	0,0
8	0	0	0	0	0,0	0,0	0,0
30	200	0	0	0	200,0	0,0	200,0
14	7	0	0	0	7,0	0,0	7,0
42	0	0	0	0	0,0	0,0	0,0
44	0	0	0	0	0,0	1,3	1,3
19	0	0	0	0	0,0	0,0	0,0
20	3,5	11	0	0	14,5	0,0	14,5
29	0	0	0	0	0,0	6,5	6,5

En la tabla 219 se muestra el número de respuestas obtenidas, en las 44 encuestas, sobre cada uno de los cultivos.

Tabla 219.

Número de respuestas para los diferentes cultivos. Encuesta a agricultores (1999 – 00).

CULTIVO	NÚMERO DE ENCUESTAS	
	SECANO	REGADÍO
Alcachofa		1
Alfalfa	2	7
Alfalfa+Esparceta+Veza	1	
Almendro	4	1
Apio exterior		1
Apio invernadero		1
Arroz		1
Avena	1	
"Branya"(trigo+ray-grass+veza)	1	
Cáñamo	2	
Cebada	22	3
Cerezo		1
Cítricos		2
Coliflor		1
Colza	1	
Flor cortada		1
Girasol	3	
Guisante		1
Lino no textil	1	1

CULTIVO	NÚMERO DE ENCUESTAS SECANO	REGADÍO
Maíz	1	10
Manzana	1	1
Melocotonero		2
Olivo	6	
Patata		2
Pepino		1
Peral	3	
Rábano		1
Ray-grass	1	
Remolacha		1
Tomate exterior		1
Tomate invernadero		1
Trigo	12	5
Veza	2	
Viña	3	1
Total	**67**	**48**

Para la discusión se han seleccionado los mismos cultivos de la encuesta del año 1998 - 1999, que por otro lado también han sido, en general, los que han obtenido mayor número de respuestas, concretamente la cebada de secano (22 respuestas), el maíz de regadío (10 respuestas), el trigo de secano (12) y de regadío (5) y la alfalfa de regadío (7).

Como se ha comentado anteriormente, la encuesta del primer día (Anejo 8) tenía como objetivo determinar la función de densidad de rendimientos para los cultivos más conocidos por el agricultor. La primera pregunta, cuyas respuestas se han resumido en esta sección, se refería a los datos identificativos de la explotación y de los encuestadores.

#8.3.- APROXIMACIONES A FUNCIONES TRIANGULARES Y BETAS

En la **Pregunta 2** del cuestionario del primer día se pedía al agricultor, para cada cultivo, una estimación directa de los rendimientos mínimo (R_{min}), máximo (R_{max}), más frecuente (R_{mf}) y medio (R_m) para cada cultivo, con objeto de determinar el intervalo que el agricultor consideraba razonable para los rendimientos de los cultivos en su zona, y una primera aproximación a la distribución de los mismos.

A partir de las respuestas dadas por los agricultores, se ha calculado la posición de la moda normalizando todos los recorridos (valor 1). La posición de la moda se ha determinado mediante la fórmula $(R_{mf} - R_{min}) : (R_{max} - R_{min})$. Los resultados obtenidos se han resumido en la tabla 220.

Tabla 220.

Posición de la moda en la estimación directa de recorrido y moda. Número de casos en cada intervalo. Encuesta a los agricultores, primer día (1999-00).

	Intervalo del valor $(R_{mf} - R_{min}) / (R_{max} - R_{min})$				
Cultivos	**0-0.2**	**0.2-0.4**	**0.4-0.6**	**0.6-0.8**	**0.8-1**
Cebada secano	1 (4.5%)	1 (4.5%)	17 (77%)	3 (14%)	0
Trigo secano	0	3 (25%)	7 (58%)	2 (17%)	0
Trigo regadío	0	0	4 (80%)	1 (20%)	0
Maíz regadío (*)	0	1 (11%)	5 (56%)	3 (33%)	0
Alfalfa regadío	0	0	5 (71%)	2 (29%)	0

(*) Habiendo eliminado una de las observaciones por ser incompleta.

De las modas relativamente centradas (para valores 0.4 – 0.6 en la tabla 220), conviene identificar aquellas que se sitúan exactamente en la mitad del recorrido. Por ello, en la tabla 221 se indica el número y porcentaje de encuestas cuya moda toma el valor 0,5 en la escala numérica discutida.

Tabla 221.

Posición central de la moda en la estimación directa de recorrido y moda. Número de casos en cada intervalo. Encuesta a los agricultores, primer día (1999-00).

Cultivos	**Moda=0.5**	**Total encuestas**	**Porcentaje (%)**
Cebada secano	4	22	18
Trigo secano	1	12	8
Trigo regadío	2	5	40
Maíz regadío (*)	3	9	33
Alfalfa regadío	3	7	43

(*) Habiendo eliminado una de las observaciones por ser incompleta.

No se ha observado que exista una correlación entre el valor del recorrido y la posición de la moda. La distribución de la nube de puntos se reproduce en la Figura 3.

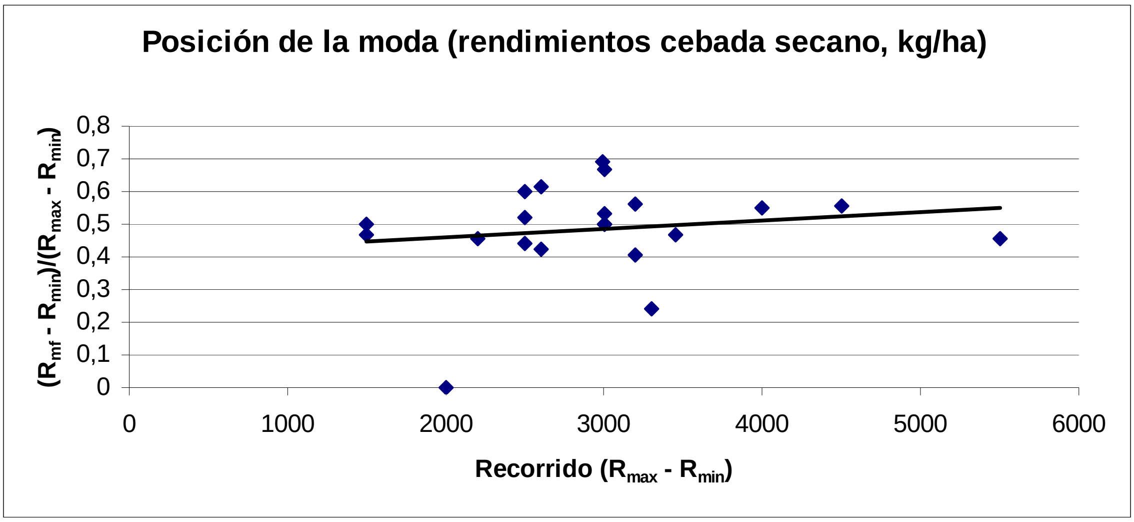


Figura 3.

Posición de la moda en función del recorrido. Encuesta a los agricultores, primer día (1999-00).

Una comparación entre el valor medio del rendimiento estimado directamente por el agricultor, y los valores medios calculados a partir de la aproximación funcional a una función triangular (RMT) y a una beta – PERT (RMB) se ha realizado en la tabla 222.

Tabla 222.

Distribución del número de casos según las diferencias relativas entre valores medios (estimación directa y calculado). Encuesta a los agricultores, primer día (1999-00).

TRIANGULAR	**Número de casos**										
Cultivo	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**Mediana**
Cebada secano	12	7		3							0.046
Trigo secano	7	3				1	1				0.043
Trigo regadío	4	1									0.031
Maíz regadío	5	3	1	1							0.050
Alfalfa regadío	5		1		1						0.035
BETA	**Número de casos**										
Cultivo	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**Mediana**
Cebada secano	14	2	2	4							0.039
Trigo secano	6	1	1	2	2						0.050
Trigo regadío	4	1									0.031
Maíz regadío	7	1	2								0.036
Alfalfa regadío	4	1	2								0.044

Se observa para todos los cultivos y las dos distribuciones estudiadas unas diferencias relativas para la mitad de los casos (mediana) inferiores al 5%.

Se realizó a continuación una comparación entre las poblaciones de rendimientos medios estimados R_m y calculados a partir de las distribuciones triangular (RMT) y beta (RMB), obteniendo los resultados resumidos en la Tabla 223, que indican una gran coincidencia entre esos valores.

Tabla 223.

Comparación entre las poblaciones de rendimientos medios (estimación directa y calculados). Encuesta a los agricultores, primer día (1999-00).

Comparación con la distribución triangular			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.926	No	No diferencias
Trigo secano	0.930	No	No diferencias
Trigo regadío	0.900	Sí	No diferencias
Maíz regadío	0.915	No	No diferencias
Alfalfa regadío	0.964	No	No diferencias
Comparación con la distribución beta			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.904	No	No diferencias
Trigo secano	0.935	No	No diferencias
Trigo regadío	0.900	Sí	No diferencias
Maíz regadío	0.988	No	No diferencias
Alfalfa regadío	0.964	No	No diferencias

Una tercera comprobación se realizó mediante regresión (Tabla 224).

Tabla 224.

Regresión $R_m = a\ RMT + b$ y $R_m = a\ RBT + b$. Encuesta a los agricultores, primer día (1999-00).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.95	1.00	0.81	0.94	0.99
b	179.3	-120.4	909.8	327.0	449.8
R^2	0.89	0.91	0.93	0.87	0.99
Fc	166.6	130.1	37.9	54.2	855.8
Ft	4.35	4.96	10.13	5.32	6.61
DISTRIBUCIÓN BETA					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.90	0.92	0.85	1.05	1.00
b	356.8	117.7	662.4	739.2	-125.7
R^2	0.84	0.88	0.95	0.94	0.99
Fc	106.0	76.1	63.6	123.5	1741.7
Ft	4.35	4.96	10.13	5.32	6.61

Las mismas comprobaciones se han realizado para el caso en que se fuerza a que b = 0, resumiéndose los resultados en la tabla 225. Se puede comprobar la fuerte correspondencia entre los valores medios directamente estimados y los calculados a partir de las tres estimaciones (R_{min}, R_{mf}, R_{max}), para todos los cultivos, en esta última regresión.

Tabla 225.

Regresión $R_m = a\ RMT$ ($b = 0$) y $R_m = a\ RMB$ ($b = 0$). Encuesta a los agricultores, primer día (1999-00).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.99	0.96	1.00	0.98	1.01
b	0	0	0	0	0
R^2	0.99	0.99	0.99	0.99	0.99
Fc	4593.99	1485.97	2133.72	1436.58	2376.63
Ft	4.32	4.84	7.71	5.12	5.99
DISTRIBUCIÓN BETA					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.99	0.96	0.99	0.98	0.99
b	0	0	0	0	0
R^2	0.99	0.99	0.99	0.99	0.99
Fc	2958.39	1131.52	3801.67	2877.89	5127.17
Ft	4.32	4.84	7.71	5.12	5.99

#8.3.1.- RESULTADOS EN EL SEGUNDO DÍA

En la encuesta del segundo día (realizada al menos una semana después) en la **pregunta 2** se pedía una estimación directa puntual de los rendimientos medios, R_m. Más tarde, en la **pregunta 9**, se volvía a solicitar una estimación directa de los rendimientos mínimos (R_{min}), más frecuentes (R_{mf}) y máximos (R_{max}).

La disposición de las preguntas intentaba minimizar la influencia de un posible anclaje común de todos esos valores.

Se ha calculado la posición relativa de la moda (R_{mf}) en intervalos normalizados, calculada mediante ($R_{mf} - R_{min}$): ($R_{max} - R_{min}$), y los resultados se han resumido en las tablas 226 y 227.

Tabla 226.

Posición de la moda en la estimación directa de recorrido y moda. Número de casos en cada intervalo. Encuesta a los agricultores, segundo día (1999-00).

	Intervalo del valor (R_{mf} - R_{min}) : (R_{max} - R_{min})				
Cultivo	**0-0,2**	**0,2-0,4**	**0,4-0,6**	**0,6-0,8**	**0,8-1**
Cebada secano	0	6 (27%)	12 (55%)	4 (18%)	0
Trigo secano	0	3 (25%)	6 (50%)	3 (25%)	0
Trigo regadío	0	0	3 (60%)	2 (40%)	0
Maíz regadío	0	1 (10%)	6 (60%)	2 (20%)	1 (10%)
Alfalfa regadío	1 (14%)	0	4 (57%)	2 (29%)	0

En la tabla 227 se indican las encuestas con moda igual a 0.5 y el porcentaje que representan sobre el total.

Tabla 227.

Posición central de la moda en la estimación directa de recorrido y moda. Número de casos en cada intervalo. Encuesta a los agricultores, segundo día (1999-00).

Cultivos	**Moda=0.5**	**Total encuestas**	**Porcentaje (%)**
Cebada secano	2	22	9
Trigo secano	1	12	8
Trigo regadío	0	5	0
Maíz regadío	2	10	20
Alfalfa regadío	0	7	0

La comparación entre el valor medio del rendimiento estimado directamente por el agricultor, y los valores medios calculados a partir de la aproximación funcional a una función triangular (RMT) y a una beta – PERT (RMB) se ha realizado en la tabla 228. Todos estos valores corresponden al segundo día.

Tabla 228.

Distribución del número de casos según las diferencias relativas entre valores medios (estimación directa y calculado). Encuesta a los agricultores, segundo día (1999-00).

TRIANGULAR	**Número de casos**										
Cultivo	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**Mediana**
Cebada secano	10	7	1	2		1	1				0.065
Trigo secano	9	2	1								0.033
Trigo regadío	5										0.025
Maíz regadío	5	5									0.050
Alfalfa regadío	3	3	1								0.058
BETA	**Número de casos**										
Cultivo	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**Mediana**
Cebada secano	9	7	3	1	1		1				0.069
Trigo secano	9	3									0.033
Trigo regadío	4		1								0.031
Maíz regadío	6	2	2								0.042
Alfalfa regadío	3	3	1								0.058

Se observa para todos los cultivos y las dos distribuciones estudiadas unas diferencias relativas para la mitad de los casos (mediana) inferiores al 7%.

Se realizó a continuación una comparación entre las poblaciones de rendimientos medios estimados R_m y calculados a partir de las distribuciones triangular (RMT) y beta (RMB), obteniendo los resultados resumidos en la Tabla 229, que indican una gran coincidencia entre esos valores. Todos los datos son del segundo día.

Tabla 229.

Comparación entre las poblaciones de rendimientos medios (estimación directa y calculados). Encuesta a los agricultores, segundo día (1999-00).

Comparación con la distribución triangular			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.822	No	No diferencias
Trigo secano	0.981	No	No diferencias
Trigo regadío	1	No	No diferencias
Maíz regadío	0.961	No	No diferencias
Alfalfa regadío	0.964	No	No diferencias
Comparación con la distribución beta			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.798	No	No diferencias
Trigo secano	0.981	No	No diferencias
Trigo regadío	1	No	No diferencias
Maíz regadío	0.985	No	No diferencias
Alfalfa regadío	0.928	No	No diferencias

Una tercera comprobación se realizó mediante regresión (Tabla 230), con datos del segundo día.

Tabla 230.

Regresión $R_m = a\ RMT + b$ y $R_m = a\ RBT + b$. Encuesta a los agricultores, segundo día (1999-00).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	1.06	1.18	1.05	1.10	1.09
b	-275.7	-537.7	-260.8	-1036.2	-672.3
R^2	0.82	0.96	0.95	0.96	0.99
Fc	89.4	219.9	59.1	213.9	1632.0
Ft	4.35	4.96	10.13	5.32	6.61
DISTRIBUCIÓN BETA					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.97	1.09	0.96	1.04	1.12
b	5.9	-258.2	59.7	-669.1	-1404.1
R^2	0.83	0.98	0.87	0.96	0.99
Fc	95.2	506.3	20.04	204.2	3381.9
Ft	4.35	4.96	10.13	5.32	6.61

Las mismas comprobaciones se han realizado para el caso en que se fuerza a que b = 0, resumiéndose los resultados en la tabla 231, para datos del segundo día. Se puede comprobar la fuerte correspondencia entre los valores medios directamente estimados en la pregunta 2 y los calculados a partir de las tres estimaciones (R_{min}, R_{mf}, R_{max}) de la **pregunta 9**, para todos los cultivos, en esta última regresión.

Tabla 231.

Regresión $R_m = a\ RMT$ ($b = 0$) y $R_m = a\ RMB$ ($b = 0$). Encuesta a los agricultores, segundo día (1999-00).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.98	1.02	0.99	1.00	1.06
b	0	0	0	0	0
R^2	0.99	0.99	0.99	0.99	0.99
Fc	2248.82	1879.92	4296.50	2173.57	4267.20
Ft	4.32	4.84	7.71	5.12	5.99
DISTRIBUCIÓN BETA					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.98	1.01	0.98	0.98	1.07
b	0	0	0	0	0
R^2	0.99	0.99	0.99	0.99	0.99
Fc	2432.26	4807.12	1678.24	2318.95	3313.52
Ft	4.32	4.84	7.71	5.12	5.99

#8.3.2.- VALORES DE APROXIMACIONES FUNCIONALES (DÍAS 1 Y 2)

Para los valores (R_{min}, R_{mf}, R_{max}) obtenidos cada día (pregunta 2 del primer día y pregunta 9 del segundo día), se ha procedido a calcular las diferencias relativas obtenidas de las estimaciones de valores medios mediante la utilización de las aproximaciones funcionales (tabla 232). Concretamente se han calculado los valores absolutos de las estimaciones triangular |(RMT1 – RMT2):RMT2| y beta |(RMB1 – RMB2):RMB2|, siendo RMTi la aproximación triangular con datos del dia i y RMBi la correspondiente con la beta (i = 1, 2).

Tabla 232.

Valores de |(RMT1 – RMT2):RMT2| y |(RMB1 – RMB2):RMB2|. Encuesta a los agricultores, primer y segundo día (1999-00).

TRIANGULAR	**Número de casos**											
Cultivo	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**>0.5**	**Mediana**
Cebada secano	8	6	1	4		2	1					0.075
Trigo secano	11	1										0.027
Trigo regadío	4	1										0.031
Maíz regadío	5	1	2	1							1	0.050
Alfalfa regadío	3	3									1	0.058
BETA	**Número de casos**											
Cultivo	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**>0.5**	**Mediana**
Cebada secano	10	3	3	3	2			1				0.067
Trigo secano	10	1	1									0.030
Trigo regadío	4		1									0.031
Maíz regadío	3	3	3								1	0.083
Alfalfa regadío	4	2									1	0.044

Las diferencias relativas obtenidas son de un orden semejante en ambas aproximaciones (beta – PERT y triangular), y en el 50% de los casos se aprecian diferencias inferiores al 8%.

La comparación entre las poblaciones de valores medios para las aproximaciones a las distribuciones triangular y beta en los dos días se resume en la tabla 233.

Tabla 233.

Comparación entre poblaciones de valores medios. Encuesta a los agricultores, primer y segundo día (1999-00).

Comparación con la distribución triangular			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.756	No	No diferencias
Trigo secano	0.956	No	No diferencias
Trigo regadío	1	No	No diferencias
Maíz regadío	0.738	No	No diferencias
Alfalfa regadío	0.893	No	No diferencias
Comparación con la distribución beta			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.805	No	No diferencias
Trigo secano	0.951	No	No diferencias
Trigo regadío	0.900	No	No diferencias
Maíz regadío	0.861	No	No diferencias

Alfalfa regadío	0.875	No	No diferencias

Se obtienen coeficientes de correlación altos y dependencia en todas las poblaciones comparadas.

La **regresión** entre los valores de rendimientos medios calculados los días 1 y 2 mediante las fórmulas de las distribuciones triangular (*RMT1* y *RMT2*, respectivamente) y beta (*RMB1* y *RMB2*, respectivamente), se han reproducido en la tabla 234.

Tabla 234.

Regresión *RMT1 = a RMT2 + b* y *RMB1 = a RMB2 + b*. Encuesta a los agricultores, primer y segundo día (1999-00).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.82	1.01	1.12	0.85	1.13
b	615.46	-75.84	-792.36	1174.9	-2583.77
R^2	0.55	0.99	0.99	0.45	0.97
Fc	24.9	932.0	926.8	6.52	178.4
Ft	4.35	4.96	10.13	5.32	6.61
DISTRIBUCIÓN BETA					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.81	0.98	1.01	1.02	1.15
b	647.53	25.07	-264.99	-425.64	-2723.25
R^2	0.61	0.98	0.93	0.51	0.97
Fc	31.5	412.8	39.66	8.34	160.1
Ft	4.35	4.96	10.13	5.32	6.61

Obteniéndose ajustes significativos en todos los casos, aunque en ciertos cultivos los coeficientes de determinación son algo bajos.

En caso de que se fuerce $b = 0$, se obtienen los resultados de la tabla 235.

Tabla 235.

Regresión *RMT1 = a RMT2* (*b* = 0) *RMB1 = a RMB2* (*b* = 0). Encuesta a los agricultores, primer y segundo día (1999-00).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	1.00	0.99	0.97	0.96	1.03
b	0	0	0	0	0
R^2	0.98	0.99	1.00	0.96	0.98
Fc	1055.2	14120.3	8125.9	204.0	366.9
Ft	4.32	4.84	7.71	5.12	5.99
DISTRIBUCIÓN BETA					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.99	0.98	0.96	0.98	1.05
b	0	0	0	0	0
R^2	0.98	0.99	0.99	0.96	0.98
Fc	1119.4	5874.6	2818.7	219.3	346.8
Ft	4.32	4.84	7.71	5.12	5.99

#8.3.3.- VARIANZAS

Al disponer de estimaciones de la forma (R_{min}, R_{mf}, R_{max}) para los dos días (preguntas 2 y 9 del primer y del segundo día, respectivamente), es posible comprobar la relación entre las varianzas estimadas mediante la simplificación triangular y la beta – PERT.

Como es conocido, la varianza de la función triangular viene dada por:

$$\left(VT = \frac{1}{18}\left[\left(R_{max} - R_{min}\right)^2 + \left(R_{mf} - R_{min}\right)\left(R_{mf} - R_{max}\right)\right]\right)$$

y la aproximación PERT a la beta por VB = (1/36) (R_{max} –R_{min} $)^2$.

La varianza triangular del primer día se denotará VT1, y la del segundo día como VT2. Las varianzas beta se denominarán respectivamente VB1 y VB2.

En la tabla 236 se han resumido los resultados de las comparaciones |(VT1 – VT2):VT2| y |(VB1 – VB2):VB2|.

Tabla 236.

Diferencias relativas entre varianzas calculadas |(VT1 – VT2):VT2| y |(VB1 – VB2):VB2|. Encuesta a los agricultores, primer y segundo día (1999-00).

TRIANGULAR	**Número de casos**										
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**	**>1**	**Mediana**
Cebada secano	6	1	6	2					1	1	5	0.267
Trigo secano	6	1	2	1	1						1	0.100
Trigo regadío					1		1	3				0.718
Maíz regadío	2		1	1			1	3			2	0.700
Alfalfa regadío	1		1	1	1		1	1	1			0.350
BETA	**Número de casos**											
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**	**>1**	**Mediana**
Cebada secano	6	2	6	1					1	1	5	0.250
Trigo secano	7		2	1	1						1	0.086
Trigo regadío					1		1	2	1			0.750
Maíz regadío	2		1	1			1	2		1	2	0.700
Alfalfa regadío	1	1		2			1	1	1			0.375

En general, las estimaciones de la varianza realizadas por la aproximación triangular y la beta – PERT son muy semejantes.

En la tabla 237 se han comparado las poblaciones de varianzas.

Tabla 237.

Comparación de las poblaciones de varianzas calculadas |(VT1 – VT2):VT2| y |(VB1 – VB2):VB2|. Encuesta a los agricultores, primer y segundo día (1999-00).

Comparación con la distribución triangular			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.032	Sí	No diferencias
Trigo secano	0.671	No	No diferencias
Trigo regadío	0.400	Sí	Diferencias
Maíz regadío	0.529	Sí	No diferencias
Alfalfa regadío	0.884	No	No diferencias
Comparación con la distribución beta			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.002	Sí	No diferencias
Trigo secano	0.638	No	No diferencias
Trigo regadío	0.575	Sí	No diferencias
Maíz regadío	0.482	Sí	No diferencias
Alfalfa regadío	0.893	No	No diferencias

Se observa mayoritariamente dependencia entre las poblaciones estudiadas.

El análisis anterior se completa con el estudio de la regresión entre los valores estudiados, los resultados de la cual se resumen en la tabla 238.

Tabla 238.

Regresiones *VT1 = a VT2 + b y VB1 = a VB2 +b*. Encuesta a los agricultores, primer y segundo día (1999-00).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.32	0.60	0.01	0.74	0.31
b	268498.19	134873.63	494569.34	294721.28	2381064.22
R^2	0.22	0.41	0	0.45	0.68
Fc	5.51	7.06	0.0	6.52	10.5
Ft	4.35	4.96	10.13	5.32	6.61
DISTRIBUCIÓN BETA					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.33	0.61	0.03	0.76	0.32
b	174405.48	86033.37	320666.29	148143.56	1411263.91
R^2	0.22	0.41	0	0.46	0.73
Fc	5.64	6.9	0	6.8	13.5
Ft	4.35	4.96	10.13	5.32	6.61

Cuando se fuerza que b = 0, los resultados de las regresiones se resumen en la tabla 239.

Tabla 239.

Regresiones *VT1 = a VT2 y VB1 = a VB2* . Encuesta a los agricultores, primer y segundo día (1990-00).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.69	0.87	0.57	0.83	0.38
b	0	0	0	0	0
R^2	0.65	0.82	0.50	0.70	0.78
Fc	38.3	50.1	4.1	20.8	20.7
Ft	4.32	4.84	7.71	5.12	5.99
DISTRIBUCIÓN BETA					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.68	0.87	0.62	0.82	0.38
b	0	0	0	0	0
R^2	0.64	0.82	0.53	0.70	0.81
Fc	37.5	50.5	4.5	20.7	25.4
Ft	4.32	4.84	7.71	5.12	5.99

#8.4.- FUNCIONES DE DENSIDAD DE RENDIMIENTOS CON INTERVALOS ORDINALES

En la **Pregunta 4** se pedía al agricultor que definiera el porcentaje en que, en su opinión, se presentan cada uno de los casos de rendimientos definidos en lenguaje natural (*muy malos*, *malos*, *normales*, *buenos* y *muy buenos*), a fin de obtener una primera aproximación a la función de densidad, con las abscisas definidas en lenguaje natural. En las respuestas, como en el caso de la encuesta de los expertos, se dieron casos de reducción de intervalos (fenómeno discutido en ζ 4. 2. 3), cuya importancia se resume en la tabla 240.

Tabla 240.

Forma de las funciones de densidad (intervalos ordinales), primer día. Encuesta a los agricultores (1999 – 00).

Cultivo	**Reducción de intervalo**			**Uniforme**	**Extremos < 15%**	**Simetría**	**Inter 1 = Inter 5**	**Inter 2 = Inter 4**	**Total**
	I	**A**	**D**						
Cebada secano	0	1	2	1	4	8	11	8	22
Trigo secano	0	1	1	0	4	3	6	5	12
Trigo regadío	1	0	0	0	1	2	3	3	5
Maíz regadío	3	1	0	0	3	4	4	5	10
Alfalfa regadío	0	1	1	0	3	4	5	5	7

En el caso de la cebada de secano, en 3 casos de 22 se reduce la amplitud del intervalo, y en 1 caso se dan frecuencias uniformes a los cinco intervalos. Once respuestas de 22 igualan las frecuencias de los intervalos extremos, y de ellas 8 se corresponden con distribuciones simétricas. Mayoritariamente (14 de 22 casos, el 64% de los mismos) los agricultores no atribuyen simetría a las distribuciones.

En el caso del trigo de secano, en 2 casos de 12 se reduce la amplitud del intervalo, y en ningún caso se dan frecuencias uniformes a los cinco intervalos. Seis respuestas de 12 igualan las frecuencias de los intervalos extremos, y de ellas 3 se corresponden con distribuciones simétricas. Igual que para la cebada de secano, mayoritariamente (9 de 12 casos o 75% de los mismos) los agricultores no atribuyen simetría a las distribuciones.

En el caso del trigo de regadío, en 1 caso de 5 se reduce la amplitud del intervalo. Tres de las cinco respuestas igualan las frecuencias de los intervalos extremos, y de ellas 2 se corresponden con distribuciones simétricas. Igual que para los cultivos anteriores, mayoritariamente (3 de 5 casos o 60% de los casos) los agricultores no atribuyen simetría a las distribuciones.

En el caso del maíz de regadío, en 4 casos de 10 se reduce la amplitud del intervalo. Cuatro respuestas de 10 igualan las frecuencias de los intervalos extremos, y 4 son distribuciones simétricas. Igual que en los cultivos anteriores, mayoritariamente (6 de 10 casos o 60%) los agricultores no atribuyen simetría a las distribuciones.

En este cultivo, una de las encuestas presenta toda la masa concentrada en el intervalo normal de rendimientos.

En el caso de la alfalfa de regadío, en 2 de los 7 casos se reduce la amplitud del intervalo, y en ningún caso se dan frecuencias uniformes. Cinco respuestas de 7 igualan las frecuencias de los intervalos extremos, y 4 son distribuciones simétricas. En tres de los 7 casos o 43% los agricultores no atribuyen simetría a las distribuciones. En este cultivo, una de las encuestas presenta toda la masa concentrada en el intervalo bueno de rendimientos (correspondiendo a un agricultor distinto al del caso del maíz de regadío).

Determinación de la moda. Los intervalos que consideran los agricultores más frecuentes, se reproducen en la tabla 241, donde el intervalo MM se refiere al de rendimientos muy malos, el M a rendimientos malos, el N a normales, el B a buenos y el MB a muy buenos.

Tabla 241.

Número y posición de las modas (intervalos ordinales), primer día. Encuesta a los agricultores (1999 – 00).

Cultivo	**1 Moda**	**2 Modas**	**Moda en 2 intervalos**	**Moda en intervalo**				
				MM	**M**	**N**	**B**	**MB**
Cebada secano	19	2	1	2	2	18	2	0
Trigo secano	11	0	1	0	1	12	1	0
Trigo regadío	4	0	1	0	0	5	1	0
Maíz regadío	10	0	0	1	0	6	3	0
Alfalfa regadío	7	0	0	0	0	5	2	0

En el caso en que se presentan modas en 2 intervalos contiguos o no, se ha asignado la moda a cada uno de los intervalos, con lo que el total de modas supera al total de encuestas.

En cebada de secano, mayoritariamente se considera el intervalo central como más frecuente.

De los 12 casos examinados en trigo de secano, en 1 se presentan modas en dos intervalos contiguos, y del mismo modo que en cebada de secano, se considera el intervalo central como más frecuente.

De los 5 casos examinados en trigo de regadío, en 1 caso se presentan modas en más de un intervalo, y mayoritariamente se considera el intervalo central como más frecuente.

En maíz de regadío, en ningún caso se presentan modas en dos intervalos contiguos ni distribuciones bimodales. Mayoritariamente, se considera el intervalo central como más frecuente, y cuando no lo es éste, lo es principalmente su contiguo derecho.

En alfalfa de regadío, en ninguno de los 7 casos examinados se presentan modas en más de un intervalo. Mayoritariamente se considera el intervalo central como más frecuente, y cuando no lo es éste lo es su contiguo derecho.

Determinación de la mediana. A fin de determinar la simetría en la distribución de las masas de frecuencia, se ha determinado la mediana sobre la base de 5 intervalos (tabla 242).

Tabla 242.

Valor de la mediana (expresada en función del número de intervalo), primer día. Encuesta a los agricultores (1999 – 00).

Cultivo	**Valor de la mediana (expresada en número del intervalo)**									
	0-0.5	**0.5-1**	**1-1.5**	**1.5-2**	**2-2.5**	**2.5-3**	**3-3.5**	**3.5-4**	**4-4.5**	**4.5-5**
Cebada de secano				5	12	5				
Trigo de secano				1	8	3				
Trigo de regadío					2	3				
Maíz de regadío				1	4	2	2	1		
Alfalfa regadío					5		2			

Para la cebada de secano, el mayor número de casos (18 de 22) presenta su moda en el intervalo central (el tercer intervalo). En ese intervalo central, a la izquierda de su punto medio se encuentran 4 observaciones, otras 8 en el valor 2.5 y 5 observaciones a su derecha. De esta forma, la mediana se encontraría a la izquierda del punto medio del recorrido en 9 casos de 22, coincidiendo con ese punto en 8 ocasiones, y a su derecha en 5 casos. Puede observarse una tendencia mayoritaria a asignar masas de probabilidad relativamente centradas.

Para el trigo de secano, todos los casos tienen su moda en el intervalo central. En ese intervalo, a la izquierda de su punto medio se encuentran 5 observaciones, otras 3 en el valor 2.5 y 3 a su derecha. De esta forma, la mediana se encontraría a la izquierda del punto medio del recorrido en 6 casos de 12, coincidiendo con ese punto en 3 ocasiones, y a la derecha del mismo en 3 casos. El valor de la mediana indica una tendencia mayoritaria a asignar masas de probabilidad relativamente centradas, aunque con mayor peso hacia la izquierda del intervalo central.

Para el trigo de regadío, todos los casos (5 de 5) presentan su moda en el intervalo central. En ese intervalo, a la izquierda de su punto medio no se encuentra ninguna observación, dos en el valor 2.5 y 3 a su derecha. De esta forma, la mediana no se encontraría a la izquierda del punto medio del recorrido en ningún caso, coincidiendo con ese punto en 2 ocasiones, y a la derecha del mismo en 3 casos. El valor de la mediana indica una tendencia mayoritaria a asignar masas de probabilidad relativamente centradas, aunque la parte derecha del intervalo central tiene un mayor peso.

Para el maíz de regadío, el mayor número de casos (6 de 10 casos) presenta su moda en el intervalo central. En ese intervalo, a la izquierda de su punto medio no se encuentra ninguna observación, 4 en el valor 2.5 y 2 a la derecha del mismo. De esta forma, la mediana se encontraría a la izquierda del punto medio del recorrido en 1 caso, coincidiendo con ese punto en 4 ocasiones, y a la derecha del mismo en 5 casos. El valor de la mediana indica una tendencia mayoritaria a asignar masas de probabilidad relativamente centradas, aunque se observa una preferencia clara hacia los valores altos de rendimientos, dentro del intervalo central.

Para la alfalfa de regadío, el mayor número de casos (5 de 7) presenta su moda en el intervalo central. En ese intervalo, a la izquierda de su punto medio se presenta 1 observación, 4 para el valor 2.5 y ninguna a su derecha. De esta forma, la mediana se encontraría a la izquierda del recorrido en 1 caso, coincidiría con el punto medio en 4 casos y estaría a su derecha en 2 casos. El valor de la mediana indica una tendencia mayoritaria a asignar masas de probabilidad relativamente centradas.

#8.4.1- FRECUENCIAS EN EL SEGUNDO DÍA (ESCALA ORDINAL)

En el cuestionario del segundo día se introdujo la misma pregunta (también con el **número 4**). El número de reducciones de intervalos y el estudio de simetría de las respuestas obtenidas se resume en la tabla 243.

Tabla 243.

Forma de las funciones de densidad (intervalos ordinales), segundo día. Encuesta a los agricultores (1999 – 00).

	Reducción de intervalo			**Uniforme**	**Extremos < 15%**	**Simetría**	**Inter 1 = Inter 5**	**Inter 2 = Inter 4**	**Total**
	I	**A**	**D**						
Cebada secano	0	1	1	1	4	5	10	8	22
Trigo secano	0	1	1	0	3	3	5	3	12
Trigo regadío	1	0	0	0	1	2	2	1	5
Maíz regadío	3	1	0	0	5	4	4	4	10
Alfalfa regadío	1	1	0	0	3	2	3	3	7

En el caso de la cebada de secano, en 2 casos de 22 se reduce la amplitud del intervalo, y en un caso se dan frecuencias uniformes a los cinco intervalos. Diez de las 22 respuestas igualan las frecuencias de los intervalos extremos, y de ellas 5 se corresponden con distribuciones simétricas. Mayoritariamente (17 de 22 casos, el 77% de los mismos) los agricultores no atribuyen simetría a las distribuciones.

En el caso del trigo de secano, en 2 de los 12 casos se reduce la amplitud del intervalo, y en ningún caso se dan frecuencias uniformes a los cinco intervalos. Cinco respuestas de 12 igualan las frecuencias de los intervalos extremos, y de ellas 3 se corresponden con distribuciones simétricas. Igual que para la cebada de secano, mayoritariamente (9 de 12 casos o 75% de los mismos) los agricultores no atribuyen simetría a las distribuciones.

En el caso del trigo de regadío, en 1 de los 5 casos se reduce la amplitud del intervalo. Dos respuestas de 5 igualan las frecuencias de los intervalos extremos, y en ellas 2 se corresponden con distribuciones simétricas. Igual que para los cultivos anteriores, mayoritariamente (3 de 5 casos o 60% de los mismos) los agricultores no atribuyen simetría a las distribuciones.

En el caso del maíz de regadío, en 4 de los 10 casos se reduce la amplitud del intervalo. Cuatro respuestas de 10 igualan las frecuencias de los intervalos extremos, y 4 son distribuciones simétricas. Igual que en los cultivos anteriores, mayoritariamente (6 de 10 casos o 60%) los agricultores no atribuyen simetría a las distribuciones.

En el caso de la alfalfa de regadío, en 2 de los 7 casos se reduce la amplitud del intervalo, y en ningún caso se dan frecuencias uniformes. Tres respuestas de 7 igualan las frecuencias de los intervalos extremos, y 2 son distribuciones simétricas. De la misma forma que para los cultivos anteriores, mayoritariamente (5 de 7 casos o 71% de los mismos) los agricultores no atribuyen simetría a las distribuciones.

Determinación de la moda. Los intervalos que consideran los agricultores más frecuentes, se reproducen en la tabla 244, donde se indica: el número de casos en que señalan un solo intervalo como más frecuente (1 moda), el número de casos en que identifican dos modas (dos intervalos no contiguos con las mayores frecuencias), y el número de casos en los que dos intervalos contiguos tienen la misma frecuencia, y ésta es la mayor de todas ellas. Por último se da el número de casos en los que la moda se asigna a cada uno de los intervalos (siendo el intervalo *MM* el de los rendimientos *"muy malos"*, el intervalo *M* el de los rendimientos *"malos"*, el *N* el de los rendimientos *"normales"*, el *B* el de los rendimientos *"buenos"*, y finalmente, el *MB* el de los rendimientos *"muy buenos"*).

Tabla 244.

Número y posición de las modas (intervalos ordinales), segundo día. Encuesta a los agricultores (1999 – 00).

Cultivo	**1 Moda**	**2 Modas**	**Moda en 2 intervalos**	**Moda en intervalo**				
				MM	**M**	**N**	**B**	**MB**
Cebada secano	16	2	3	3	4	17	2	0
Trigo secano	11	0	1	1	1	9	2	0
Trigo regadío	5	0	0	0	0	4	1	0
Maíz regadío	10	0	0	0	0	7	3	0
Alfalfa regadío	6	0	1	0	0	5	3	0

En cebada de secano, en los 5 casos en los que se presentan modas en 2 intervalos contiguos o no, se ha asignado la moda a cada uno de los intervalos, por lo que la suma de casos supera el número de 22 respuestas. Mayoritariamente se considera el intervalo central como más frecuente, y cuando no lo es éste, lo es principalmente el intervalo situado a su izquierda.

De los 12 casos examinados en trigo de secano, en 1 se presentan modas en dos intervalos. Mayoritariamente se considera el intervalo central como más frecuente.

De los 5 casos examinados en trigo de regadío, sólo se presenta la moda en un intervalo, y mayoritariamente se considera el intervalo central como más frecuente.

En maíz de regadío, en los 10 casos estudiados se obtiene una sola moda. Mayoritariamente se considera el intervalo central como más frecuente, y cuando no lo es éste, lo es su contiguo derecho.

En alfalfa de regadío, en 1 de los 7 casos examinados se presentan modas en dos intervalos, y no se observa ninguna distribución bimodal. Mayoritariamente se considera el intervalo central como más frecuente, y cuando no lo es éste lo es su contiguo derecho.

Determinación de la mediana. A fin de determinar la simetría en la distribución de las masas de frecuencia, se ha determinado la mediana basándose en los 5 intervalos (Tabla 245), siendo el intervalo 1 el extremo inferior y el 5 el superior, del mismo modo que para la encuesta del primer día.

Tabla 245.

Valor de la mediana (relativa a intervalos ordinales), segundo día. Encuesta a los agricultores (1999 – 00).

Cultivo	**Valor de la mediana (expresada en número del intervalo)**									
	0-0.5	**0.5-1**	**1-1.5**	**1.5-2**	**2-2.5**	**2.5-3**	**3-3.5**	**3.5-4**	**4-4.5**	**4.5-5**
Cebada secano				5	13	4				
Trigo secano				1	8	3				
Trigo regadío					2	2	1			
Maíz regadío					4	3	2	1		
Alfalfa regadío					2	3	2			

Para la cebada de secano, el mayor número de casos (17 de 22) presenta su moda en el intervalo central (el tercer intervalo). En ese intervalo central, a la izquierda de su punto medio se encuentran 8 observaciones, otras 5 en el valor 2.5 y 4 a su derecha. De esta forma, la mediana se encontraría a la izquierda del punto medio del recorrido en 13 de los 22 casos, coincidiendo con ese punto en 5 ocasiones, y a su derecha en 4 casos. El valor de la mediana indica una tendencia mayoritaria a asignar masas de probabilidad relativamente centradas, sin que el punto medio concentre más de un tercio de los casos.

Para el trigo de secano, el mayor número de casos (9 de 12) presenta su moda en el intervalo central. En ese intervalo, a la izquierda de su punto medio se encuentran 5 observaciones, otras 3 en el valor 2.5 y 3 a su derecha. De esta forma, la mediana se encontraría a la izquierda del punto medio del recorrido en 6 casos de 12, coincidiendo con ese punto en 3 ocasiones, y a su derecha en 3 casos. El valor de la mediana indica una tendencia mayoritaria a asignar masas de probabilidad relativamente centradas, con mayor peso para los valores situados a la izquierda del punto medio.

Para el trigo de regadío, el mayor número de casos (4 de 5) presenta su moda en el intervalo central. En ese intervalo, no se encuentra ninguna observación a la izquierda de su punto medio, 2 en el valor 2.5 y 2 a su derecha. De esta forma, la mediana no se encontraría en ningún caso a la izquierda del punto medio del recorrido, coincidiendo con ese punto en 2 ocasiones, y a su derecha en 3 casos. El valor de la mediana indica una tendencia mayoritaria a asignar masas de probabilidad relativamente centradas, aunque la parte derecha de la zona central tiene un mayor peso.

Para el maíz de regadío, el mayor número de casos (7 de 10 casos) presenta su moda en el intervalo central. En ese intervalo, no se encuentra ninguna observación a la izquierda de su punto medio, 4 en el valor 2.5 y 3 a su derecha. De esta forma, la mediana no se encontraría a la izquierda del punto medio del recorrido en ninguno de los 10 casos, coincidiendo con ese punto en 4 ocasiones, y a su derecha en 6 casos. El valor de la mediana indica una tendencia mayoritaria a asignar masas de probabilidad relativamente centradas, aunque se observa una tendencia clara hacia los valores superiores y cercanos al punto medio del recorrido de los rendimientos.

Para la alfalfa de regadío, el mayor número de casos (5 de 7) presenta su moda en el intervalo central. En ese intervalo, a la izquierda de su punto medio no se presenta ninguna observación, 2 en el valor 2.5 y 3 a su derecha. De esta forma, la mediana no se encontraría a la izquierda del recorrido en ningún caso, coincidiría con este punto en 2 casos y estaría a su derecha en 5 casos. El valor de la mediana indica una tendencia mayoritaria a asignar masas de probabilidad relativamente centradas, aunque se observa una tendencia clara hacia los valores superiores y cercanos al punto medio del recorrido.

#8.5.- FUNCIONES DE DENSIDAD DE RENDIMIENTOS CON INTERVALOS NUMÉRICOS

El anterior análisis se ha realizado considerando qué frecuencia se ha asignado a cada intervalo definido en lenguaje natural en la **pregunta 4**, tanto del primer como del segundo día. Se dispone de una correspondencia entre los intervalos ordinales y los intervalos numéricos, a través de la **pregunta 3**, en ambos días. En lo que sigue se analiza la asignación de frecuencias, hecha sobre los intervalos ordinales, y traducidos a los numéricos.

Dada la casuística obtenida en la encuesta a los expertos (ζ 5), y la de los agricultores del curso 1998 – 1999 (ζ 7), sobre las respuestas de los sujetos (en este caso los agricultores), discutida en ζ 4.2.1, que obligaba a distribuirlas por clases, en función de la definición del recorrido y de los diferentes intervalos, se dio instrucciones a los estudiantes (los encuestadores) para que los agricultores dieran los dos extremos de cada intervalo del recorrido de rendimientos. Por ello, todas las respuestas obtenidas en este caso pertenecían a la Clase que hemos denotado como I.

En la tabla 246 se ha indicado en cuántos casos la moda se encuentra a la izquierda del punto medio del recorrido (leyenda *I*), centrada o coincidiendo con el punto medio del recorrido (leyenda *C*), o a su derecha (leyenda *D*). También se reflejan los casos en los que la distribución es uniforme (leyenda *Un*), los casos en que se presentan dos modas no contiguas (leyenda *Bi*) y el total de casos examinados, además de la posición relativa de la moda (explicada en ζ 4.3.3).

Tabla 246.

Posición de la moda (relativa a intervalos numéricos). Encuesta a los agricultores (1999 - 00, primer día).

Cultivo	**I**	**C**	**D**	**Un**	**Bi**	**Total**	**Posición relativa de la moda**								
							≤ -2	**-2- -1**	**-1- 0**	**0- 1**	**1- 2**	**2- 3**	**3- 4**	**4- 5**	**>5**
Cebada secano	6	2	11	1	2	22	2	1	5	9	2				
Trigo secano	5	1	6	0	0	12	3	2	1	6					
Trigo regadío	1	1	3	0	0	5			2	2	1				
Maíz regadío	2	2	6	0	0	10	1	1	2	1	1	1	1	1	1
Alfalfa regadío	1	1	5	0	0	7		1	1	1	1	2		1	

En cebada de secano se observa que de las 22 encuestas, 18 de ellas presentan una única moda. Esa moda se encuentra en conjunto algo decantada hacia la derecha del recorrido. Individualmente es minoría los agricultores que presentan funciones con la moda centrada. Cuando las modas se han señalado a la izquierda del centro del recorrido, se encuentran algo alejadas de ese punto, expresada esa distancia en términos de la amplitud del intervalo para el que se ha señalado la presencia de la moda. Las modas señaladas a la derecha del centro del recorrido se encuentran muy poco dispersas respecto al punto medio.

En trigo de secano sólo una encuesta sitúa la moda centrada, y el resto la sitúa de forma relativamente simétrica respecto al punto medio del recorrido. En cuanto a las modas señaladas a izquierda y derecha del centro del recorrido, se observa la misma tendencia que en cebada de secano.

En trigo de regadío, la moda se encuentra algo decantada hacia la derecha del recorrido. Individualmente, se observa que es minoría los agricultores que presentan funciones con la moda centrada. En cuanto a las modas, se presentan poco alejadas del centro del recorrido.

En maíz de regadío, se observa la moda algo decantada hacia la derecha del centro del recorrido, observándose cierta dispersión a derecha e izquierda, respecto al punto medio del recorrido, expresada esa distancia en términos de la amplitud del intervalo para el que se ha señalado la presencia de la moda.

En alfalfa de regadío, pueden generalizarse los comentarios relativos al maíz de regadío.

La **mediana** se expresa en la tabla 247 por su posición relativa normalizada en recorridos de amplitud unidad.

Tabla 247.

Posición relativa de la mediana. Encuesta a los agricultores (1999 – 00, primer día).

Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano		1	2	2	5	9	3			
Trigo secano		1		4	2	4	1			
Trigo regadío					2	1	2			
Maíz regadío				2	1	2	1	2	2	
Alfalfa regadío				1	2		3		1	

En cebada de secano, de las 22 observaciones disponibles, 14 presentan sus modas en puntos próximos al centro del recorrido, pero solamente una respuesta sitúa la mediana en ese punto. Así, en 9 casos la mediana de la función de densidad definida por el agricultor se encuentra a la izquierda del punto medio del recorrido (que en esta escala es el punto 0.5), y en 8 casos con valores mayores de 0.25. En un caso la mediana coincide con el punto medio del recorrido, y los 12 casos restantes se encuentran incluidos en el intervalo (0.5, 0.75]. De esta forma, excepto en un caso, la mediana se sitúa en valores del intervalo [0.25, 0.75]. En el intervalo [0.4, 0.6] se encuentran 14 casos (de 22), y de ellos 4 a la izquierda del punto medio, uno coincidiendo con él, y otros nueve a su derecha.

Para el trigo de secano, en 7 casos la mediana se encuentra a la izquierda del punto medio del recorrido (punto 0.5), y en 6 casos con valores mayores de 0.25. En ningún caso la mediana coincide con el punto medio del recorrido, y los 5 casos restantes se encuentran comprendidos en el intervalo (0.5, 0.75]. De esta forma, excepto en un caso, la mediana se sitúa en valores del intervalo [0.25, 0.75]. En el intervalo [0.4, 0.6] se encuentran 6 de los 12 casos, dos de ellos a la izquierda del punto medio y cuatro a su derecha.

En caso del trigo de regadío, en 2 casos la mediana se encuentra a la izquierda del punto medio del recorrido (punto 0.5), y en los dos con valores mayores de 0.25. En ningún caso la mediana coincide con el punto medio del recorrido, y los 3 casos restantes se encuentran comprendidos en el intervalo (0.5, 0.75]. De esta forma, la mediana se sitúa siempre en valores del intervalo [0.25, 0.75]. En el intervalo [0.4, 0.6] se encuentran tres de los cinco casos.

En maíz de regadío, en dos casos la mediana se encuentra a la izquierda del punto medio del recorrido (punto 0.5), y en los dos casos con un valor superior a 0.25. También en dos casos la mediana coincide con el punto medio del recorrido, y en 3 casos se encuentra comprendida en el intervalo (0.5, 0.75]. De esta forma, en 7 casos la mediana se sitúa en valores del intervalo [0.25, 0.75]. En el intervalo [0.4, 0.6] se encuentran 3 casos, dos coincidiendo con el punto medio y uno a su derecha.

En alfalfa de regadío, en un caso la mediana se encuentra a la izquierda del punto medio del recorrido (punto 0.5), y por encima de 0.25. En dos casos la mediana coincide con el punto medio del recorrido, y en 3 casos se encuentra dentro del intervalo (0.5, 0.75]. De esta forma, en 6 casos la mediana se sitúa en valores del intervalo [0.25, 0.75]. En el intervalo [0.4, 0.6] se encuentran los dos casos que coinciden con el punto medio.

#8.5.1.- RENDIMIENTOS CON INTERVALOS NUMÉRICOS (SEGUNDO DÍA)

La posición de la moda se ha resumido en la tabla 248. El significado de los encabezamientos se puede consultar en el análisis anterior realizado para las respuestas del primer día.

Tabla 248.

Posición de la moda (relativa a intervalos numéricos). Encuesta a los agricultores (1999-00, segundo día).

Cultivo	**I**	**C**	**D**	**Un**	**Bi**	**Total**	**Posición relativa de la moda**								
							≤ -2	**-2 -1**	**-1 0**	**0 1**	**1 2**	**2 3**	**3 4**	**4 5**	**> 5**
Cebada secano	7	3	9	1	2	22	3	2	5	7	1				1
Trigo secano	4	3	5	0	0	12	2	1	4	4	1				
Trigo regadío	3	0	2	0	0	5		1	2		2				
Maíz regadío	2	3	5	0	0	10		1	4		1	1	1	1	1
Alfalfa regadío	1	0	6	0	0	7		1		1	2	1		1	1

En cebada de secano se observa que de los 22 casos, 19 de ellos presentan una única moda. Esa moda se encuentra a la izquierda o centrada en la mitad de los casos, y a la derecha en la otra mitad. Individualmente, por tanto, es minoría los agricultores que presentan funciones con la moda centrada, pero en conjunto las modas se distribuyen de forma relativamente simétrica con respecto al punto medio del recorrido. Cuando las modas se han señalado a la izquierda del centro del recorrido, en general se encuentran algo alejadas de ese punto, expresada esa distancia en términos de la amplitud del intervalo para el que se ha señalado la presencia de la moda. Las modas señaladas a la derecha del centro del recorrido se encuentran más cercanas a él.

En trigo de secano, todas las distribuciones tienen una única moda, y respecto a la dispersión de las modas respecto al centro del recorrido, pueden generalizarse los comentarios realizados para la cebada de secano.

En trigo de regadío, todas las distribuciones tienen una única moda, y la distribución de modas alrededor del punto medio del recorrido es aproximadamente simétrica.

En maíz de regadío, también todas las distribuciones tienen una única moda, y la distribución de modas respecto al punto medio se realiza de forma simétrica, aunque la dispersión de los valores situados a la izquierda es inferior a la de los valores situados a la derecha.

En alfalfa de regadío, también todas las distribuciones tienen una única moda, y la distribución de las modas está claramente decantada hacia la derecha, aunque con una cierta dispersión.

En lo que respecta a la **mediana**, se ha resumido su posición en la tabla 249.

Tabla 249.

Posición relativa de la mediana. Encuesta a los agricultores (1999 – 00, segundo día).

Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano		1	1	3	6	7	4			
Trigo secano		1		1	4	5	1			
Trigo regadío					3		2			
Maíz regadío				1	2	2	2	1	2	
Alfalfa regadío					1		4	1	1	

En cebada de secano, 13 de las 22 observaciones presentan sus medianas en puntos próximos al centro del recorrido, pero solamente una respuesta sitúa la mediana en ese punto. Así, en 10 casos la mediana de la función de densidad definida por el agricultor se encuentra a la izquierda del punto medio del recorrido (que en esta escala es el punto 0.5), y en 8 casos con valores mayores de 0.25. En un caso la mediana coincide con el punto medio del recorrido, y los 11 casos restantes se encuentran incluidos en el intervalo (0.5, 0.75]. De esta forma, excepto en dos casos, la mediana se sitúa en valores del intervalo [0.25, 0.75]. En el intervalo [0.4, 0.6] se encuentran 13 casos (de 22), y de ellos 5 a la izquierda del punto medio, uno coincidiendo con él, y otros siete a su derecha.

Para el trigo de secano, en 6 casos de los 12 la mediana se encuentra a la izquierda del punto medio del recorrido (punto 0.5), y en cinco con valores mayores de 0.25. En ningún caso la mediana coincide con el punto medio del recorrido, y los 6 casos restantes se encuentran incluidos en el intervalo (0.5, 0.75]. De esta forma, excepto en un caso, la mediana se sitúa en valores del intervalo [0.25, 0.75]. En el intervalo [0.4, 0.6] se encuentran 8 de los 12 casos, y de ellos tres a la izquierda del punto medio y cinco a su derecha.

En caso del trigo de regadío, en 3 de los 5 casos la mediana se encuentra a la izquierda del punto medio del recorrido (punto 0.5), y en los tres con valores mayores de 0.25. En ningún caso la mediana coincide con el punto medio del recorrido, y los 2 casos restantes se encuentran incluidos en el intervalo (0.5, 0.75]. De esta forma, la mediana se sitúa en valores del intervalo [0.25, 0.75]. En el intervalo [0.4, 0.6] se encuentran 3 de los 5 casos.

En maíz de regadío, en dos de los diez casos la mediana se encuentra a la izquierda del punto medio del recorrido (punto 0.5), y con un valor superior a 0.25. En un caso coincide con él, y en 7 casos se encuentra a su derecha. En 4 de estos 7 casos se encuentra incluida en el intervalo (0.5, 0.75]. De esta forma, en 7 casos la mediana se sitúa en valores del intervalo [0.25, 0.75]. Concretamente, en el intervalo [0.4, 0.6] se encuentran cuatro casos, uno a la izquierda del punto medio, uno coincidiendo con él y dos a su derecha.

En alfalfa de regadío, en un caso de los 7 disponibles, la mediana se encuentra a la izquierda del punto medio del recorrido (punto 0.5), y por encima de 0.25. En ningún caso la mediana coincide con el punto medio del recorrido, y en cuatro casos se encuentra dentro del intervalo (0.5, 0.75]. De esta forma, excepto en dos casos la mediana se sitúa en el intervalo [0.25, 0.75]. En el intervalo [0.4, 0.6] se encuentra un caso, a la izquierda del punto medio.

#8.6.- PROBLEMAS EN LA DETERMINACIÓN DE INTERVALOS

En la traducción de las escalas ordinales a numéricas, en el caso de los expertos y de las encuestas a agricultores del primer año (1998 – 99), se definieron los recorridos y los intervalos de diversas formas, algunas de ellas difíciles de utilizar en la discusión de problemas donde sea importante el conocimiento preciso del recorrido y de las fronteras entre intervalos.

En la encuesta del segundo año (1999 – 00), aunque no se producen estos problemas, por las instrucciones dadas, se repiten en lo que sigue las comprobaciones para determinar si existe continuidad en los resultados.

#8.6.1.- RESULTADOS CON LA INFORMACIÓN DEL PRIMER DÍA

En la tabla 250 se resume el análisis de discontinuidades encontradas (que deberían ser continuas), y en las tablas 251 a 254 se reflejan los resultados de las regresiones realizadas para poder estimar el recorrido (largo) R_L a partir de la información disponible del recorrido corto R_C (ver ζ 4.2.1).

Tabla 250.

Número de respuestas que presentan discontinuidades / número de respuestas totales. Encuesta a los agricultores (1999 – 00, primer día).

Cultivo	**Índice de Continuidad de Intervalos**
Cebada secano	8/22
Trigo secano	6/12
Trigo regadío	1/5
Maíz regadío	5/10
Alfalfa regadío	4/7

Tabla 251.

Regresión $R_L = a\ R_C + b$. Encuesta a los agricultores (1999 – 00, primer día).

CULTIVO					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	1.09	0.93	0.85	1.02	1.59
b	1219.54	1571.91	2575.11	3938.97	710.14
Fc	38.43	2.51	3.80	14.94	21.66
Ft	4.35	4.96	10.13	5.32	6.61
R^2	0.66	0.20	0.56	0.65	0.81

La misma regresión, forzando que el término independiente *b* sea nulo, ofrece los resultados de la tabla 252, obteniéndose un resultado aceptable.

Tabla 252.

Regresión $R_L = a\ R_C$ (con $b = 0$). Encuesta a los agricultores (1999 – 00, primer día).

CULTIVO					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	1.49	1.57	1.57	1.52	1.64
b	0	0	0	0	0
Fc	690.14	140.48	94.4	157.5	101.5
Ft	4.32	4.84	7.71	5.12	5.99
R^2	0.97	0.93	0.96	0.95	0.94

También se ha realizado la regresión $(R_L – R_C) = a\ R_C + b$. Los resultados se reflejan en la tabla 253, donde se puede apreciar que se obtienen unos malos resultados, y en la tabla 254 (forzando b = 0), donde se obtienen ajustes significativos para los cultivos estudiados.

Tabla 253.

Regresión $(R_L – R_C) = a\ R_C + b$. Encuesta a los agricultores (1999 – 00, primer día).

CULTIVO					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.1	-0.07	-0.14	0.02	0.59
b	1219.54	1571.91	2575.11	3938.97	710.14
Fc	0.31	0.01	0.11	0.00	2.99
Ft	4.35	4.96	10.13	5.32	6.61
R^2	0.02	0.00	0.04	0.00	0.37

Tabla 254.

Regresión $(R_L - R_C) = a\ R_C\ (b = 0)$. Encuesta a los agricultores (1999 – 00, primer día).

	CULTIVO	
	CEBADA SECANO	**TRIGO SECANO**
a	0.50	0.57
b	0	0
Fc	76.34	18.36
Ft	4.32	4.84
R^2	0.78	0.62

Por último, se ha estudiado la distribución de los valores R_C / R_L y (R_L - R_C) / R_L (Tablas 255 y 256, respectivamente).

Tabla 255.

Cociente $R_C\ /\ R_L$. Encuesta a los agricultores (1999 – 00, primer día).

	Número de casos									
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano					3	2	9	7	1	
Trigo secano				1	2	1	2	3	3	
Trigo regadío					1	2	1	1		
Maíz regadío			1			5	1	2	1	
Alfalfa regadío			1		1	1	2	1	1	

Tabla 256.

Cociente $(R_L - R_C) / R_L$. Encuesta a los agricultores (1999 – 00, primer día).

	Número de casos									
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano		1	7	9	3	2				
Trigo secano		3	3	2	1	2	1			
Trigo regadío			2		2	1				
Maíz regadío		1	2	2	4		1			
Alfalfa regadío		1	1	2	2		1			

#8.6.1.1.- COHERENCIA DE LOS RECORRIDOS

En la pregunta **número 2** del cuestionario del primer día, el agricultor definió el recorrido de los rendimientos de los diferentes cultivos, mediante la identificación de dos extremos, el mínimo (R_{min}) y el máximo (R_{max}). En la pregunta **número 3** definió los intervalos a partir de las cinco clases expresadas en lenguaje natural. Por tanto, es posible comparar el recorrido (largo) de la pregunta 3 con el definido en la pregunta 2.

Las regresiones entre los diferentes extremos y los recorridos (tablas 257 y 258) dan unos resultados no demasiado buenos, a diferencia de los obtenidos en las mismas regresiones para la encuesta del año 1998 – 99, que parecían indicar que la amplitud del recorrido era del mismo orden cuando éste se definía directamente o a través de intervalos expresados en lenguaje natural, aunque no coincidían exactamente.

Tabla 257.

Regresión $R_L = a\ (R_{max} - R_{min}) + b$. Encuesta a los agricultores (1999 – 00, primer día).

CULTIVO					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.41	0.66	0.34	0.25	1.36
b	3067.78	1964.48	4227.68	9287.53	3306.46
Fc	2.18	5.61	0.16	0.43	13.59
Ft	4.35	4.96	10.13	5.32	6.61
R^2	0.1	0.36	0.05	0.05	0.73

Tabla 258.

Regresión $R_L = a\ (R_{max} - R_{min})$ (b = 0). Encuesta a los agricultores (1999 – 00, primer día).

CULTIVO					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	1.36	1.30	1.53	1.49	1.59
b	0	0	0	0	0
Fc	189.4	121.9	49.4	24.2	61.4
Ft	4.32	4.84	7.71	5.12	5.99
R^2	0.90	0.92	0.93	0.73	0.91

A fin estudiar la influencia de sustituir el recorrido largo por el corto, se ha calculado la recta de regresión del recorrido corto de la **pregunta 3** y los resultados de la pregunta **número 2**, mediante la ecuación $R_C = a\ (R_{max} - R_{min}) + b$, obteniéndose resultados no demasiado buenos (tabla 259).

Tabla 259.

Regresión $R_C = a\ (R_{max} - R_{min}) + b$. Encuesta a los agricultores (1999 – 00, encuesta primer día).

CULTIVO					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.31	0.08	0.36	0.46	0.80
b	1881.63	2120.72	2040.75	4038.81	2084.87
Fc	2.24	0.24	0.25	3.21	20.33
Ft	4.35	4.96	10.13	5.32	6.61
R^2	0.10	0.02	0.08	0.29	0.80

#8.6.1.2.- AMPLITUD DE LOS INTERVALOS

Como ocurría también en el caso de la encuesta a los expertos (epígrafe ζ 5.7), los agricultores al traducir los distintos intervalos ordinales a numéricos, normalmente no dan la misma amplitud a cada intervalo.

Se han normalizado los datos sobre la amplitud de los intervalos de cada encuesta, dando el valor 1 al intervalo más amplio. El resto de los intervalos se han expresado como una fracción del más amplio. Los distintos casos se han resumido en la tabla 260.

Tabla 260.

Número de veces en que el intervalo ( ] toma el valor. Encuesta a los agricultores (1999 – 00, encuesta primer día).

	Mayor	**0-0.2**	**0.2-0.4**	**0.4-0.6**	**0.6-0.8**	**0.8-1**	**Mediana**	**Media**
Cebada secano								
Muy Malo	2/22	1	5	7	5	4	0,56	0,57
Malo	12/22	0	2	5	2	13	1	0,79
Normal	5/22	1	4	4	4	9	0,68	0,68
Bueno	8/22	0	2	7	4	9	0,72	0,72
Muy Bueno	5/22	2	5	6	4	5	0,50	0,59
Trigo secano								
Muy Malo	0/12	2	4	4	1	1	0,44	0,43
Malo	3/12	2	1	3	2	4	0,58	0,59
Normal	2/12	0	3	5	2	2	0.48	0,55
Bueno	4/12	0	3	2	3	4	0,67	0,67
Muy Bueno	4/12	3	3	0	2	4	0,50	0,57
Trigo regadío								
Muy Malo	1/5	0	2	0	2	1	0.67	0,57
Malo	2/5	1	0	1	1	2	0.67	0,68
Normal	0/5	0	0	2	2	1	0.67	0,66
Bueno	0/5	0	2	1	2	0	0.56	0.51
Muy Bueno	2/5	1	0	2	0	2	0.50	0,63
Maíz regadío								
Muy Malo	3/10	1	3	0	2	4	0,69	0,63
Malo	4/10	1	1	1	2	5	0,80	0,72
Normal	4/10	2	0	2	1	5	0,81	0,71
Bueno	2/10	1	4	1	2	2	0,45	0,54
Muy Bueno	3/10	1	4	1	1	3	0,45	0,57
Alfalfa regadío								
Muy Malo	2/7	0	3	0	0	4	0,89	0,70
Malo	4/7	2	0	0	1	4	1	0,74
Normal	1/7	1	1	0	4	1	0,67	0,62
Bueno	2/7	1	1	2	1	2	0,60	0,63
Muy Bueno	1/7	1	3	1	0	2	0,40	0,51

En el caso de la cebada de secano, de las 22 respuestas obtenidas, 12 asignan el intervalo más amplio a la clase de rendimientos *"malo"*, y esta clase presenta la mitad de las observaciones por encima de uno. La clase *"malo"* es, por lo tanto, la que parece recibir las mayores amplitudes de intervalo, con un mayor peso en la distribución de frecuencias para dicha clase.

Para el trigo de secano, la clase *"bueno"* tiende a ser la más amplia, aunque la distribución de frecuencias tiende a ser más bien uniforme.

En el caso de los cultivos de regadío se dispone de pocas observaciones, aunque existe una cierta tendencia a dar una amplitud semejante a los intervalos, a diferencia de los cultivos de secano, que difieren más en las frecuencias. Se observa también una tendencia a presentar mayores valores en las clases *"muy malo"*, *"malo"* y *"normal"* para todos los cultivos de regadío.

A la vista de los resultados anteriores, no se puede afirmar que algunas clases tiendan a concentrar las mayores amplitudes, aunque la clase de rendimientos *"normal"* posiblemente actúe como un anclaje para el establecimiento de las amplitudes.

Probablemente la amplitud del intervalo esté asociada a la precisión con la que se desea puntualizar algunos rendimientos.

A continuación se midió la amplitud de los intervalos máximo y mínimo y se expresó como porcentaje del recorrido (largo) (Tabla 261) y como porcentaje del punto medio del intervalo de rendimientos *normales* (Tabla 262).

Tabla 261.

Amplitud del intervalo máximo y del mínimo en las diferentes encuestas, expresada como porcentaje del recorrido (largo). Encuesta a los agricultores (1999 – 00, primer día).

		Número de casos									
Cultivo	**Amplitud**	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano	Máxima			10	11	1					
	Mínima	9	13								
Trigo secano	Máxima			2	8	1	1				
	Mínima	8	4								
Trigo regadío	Máxima			2	2	1					
	Mínima	3	2								
Maíz regadío	Máxima		1	3	3	3					
	Mínima	5	5								
Alfalfa regadío	Máxima			3	3	1					
	Mínima	4	3								

Tabla 262.

Amplitud del intervalo máximo y del mínimo en las diferentes encuestas, expresada como porcentaje del punto medio del intervalo de rendimientos *normales*. Encuesta a los agricultores (1999 – 00, primer día).

		Número de casos										
Cultivo	**Amplitud**	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**	**>1**
Cebada secano	Máxima		2	3	7	2	6	1				1
	Mínima	3	16	3								
Trigo secano	Máxima			3	2	1	3	1	1			1
	Mínima	6	5	1								
Trigo regadío	Máxima			1	1	1	1		1			
	Mínima		4	1								

		Número de casos										
Cultivo	**Amplitud**	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**	**>1**
Maíz regadío	Máxima		2	2	2	1	2		1			
	Mínima	5	5									
Alfalfa regadío	Máxima		2	2	1	1			1			
	Mínima	4	3									

#8.6.1.3.- ANÁLISIS DE LOS EXTREMOS

En la **pregunta 2** de la encuesta del primer día se pedía al agricultor una estimación del recorrido de los rendimientos de cada cultivo: mínimo *(R_{min})*, más frecuente *(R_{mf})* y máximo *(R_{max})*. En la **pregunta 3**, se dispone del recorrido entre el extremo inferior de la clase *"rendimientos muy malos"* ($R_{<MM}$) y el extremo superior de la clase *"rendimientos muy buenos"* ($R_{>MB}$). Es posible comparar ambos valores, y para ello, inicialmente, se han calculado el coeficiente de correlación de Spearman, el test de independencia (para un nivel de significación de α = 0.05) y el test de Wilcoxon (a un nivel de significación de α = 0.05), a fin de determinar si las dos poblaciones están relacionadas y si se puede afirmar que se acepta la hipótesis de que sus valores medios pertenecen a la misma distribución (tabla 263).

Tabla 263.

Comparación entre poblaciones de rendimientos máximos y mínimos. Encuesta a los agricultores (1999 – 00, primer día).

Cultivo	**Comparación**	**Spearman (r_S)**	**Muestras independientes**	**Wilcoxon (α=0.05)**
Cebada secano	R_{min} con $R_{<MM}$	0.542	No	Existen diferencias
	R_{max} con $R_{>MB}$	0.582	No	Existen diferencias
Trigo secano	R_{min} con $R_{<MM}$	0.633	No	Existen diferencias
	R_{max} con $R_{>MB}$	0.787	No	Existen diferencias
Trigo regadío	R_{min} con $R_{<MM}$	0.575	Sí	No diferencias
	R_{max} con $R_{>MB}$	0.525	Sí	No diferencias
Maíz regadío	R_{min} con $R_{<MM}$	0.591	No	No diferencias
	R_{max} con $R_{>MB}$	0.718	No	Existen diferencias
Alfalfa regadío	R_{min} con $R_{<MM}$	0.616	Sí	No diferencias
	R_{max} con $R_{>MB}$	0.911	No	Existen diferencias

Para profundizar en la misma idea, se ha realizado una regresión entre el *rendimiento mínimo* y el *rendimiento muy malo* y el *rendimiento máximo* y el *rendimiento muy bueno*, y los resultados se han resumido en la tabla 264, obteniéndose mejores resultados cuando b es nula. Se observa claramente la tendencia de ampliación del recorrido expresado mediante lenguaje natural.

Tabla 264.

Regresión entre rendimientos máximos y mínimos. Encuesta a los agricultores (1999 – 00, primer día).

Cultivo	Comparación	y = a x + b, b ≠ 0	y = ax, b = 0
Cebada secano	R_{min} con $R_{<MM}$	$R_{min} = 0.71 R_{<MM} + 1064.64$ $R^2 = 0.48$ Fc = 18.75, Ft = 4.35	$R_{min} = 1.34 R_{<MM}$ $R^2 = 0.82$ Fc = 93.87, Ft = 4.32
	R_{max} con $R_{>MB}$	$R_{max} = 0.46 R_{>MB} + 2318.64$ $R^2 = 0.26$ Fc = 6.96, Ft = 4.35	$R_{max} = 0.88 R_{>MB}$ $R^2 = 0.97$ Fc = 793.30, Ft = 4.32
Trigo secano	R_{min} con $R_{<MM}$	$R_{min} = 0.79 R_{<MM} + 757.40$ $R^2 = 0.72$ Fc = 25.10, Ft = 4.96	$R_{min} = 1.20 R_{<MM}$ $R^2 = 0.89$ Fc = 87.33, Ft = 4.84
	R_{max} con R_{MB}	$R_{max} = 0.92 R_{MB} – 146.75$ $R^2 = 0.56$ Fc = 12.74, Ft = 4.96	$R_{max} = 0.89 R_{MB}$ $R^2 = 0.98$ Fc = 447.11, Ft = 4.84
Trigo regadío	R_{min} con $R_{<MM}$	$R_{min} = 0.44 R_{<MM} + 2492.58$ $R^2 = 0.25$ Fc = 1.02, Ft = 10.13	$R_{min} = 1.67 R_{<MM}$ $R^2 = 0.70$ Fc = 9.5, Ft = 7.71
		$R_{max} = 0.47 R_{MB} + 3294.12$ $R^2 = 0.12$ Fc = 0.4, Ft = 10.13	$R_{max} = 0.97 R_{MB}$ $R^2 = 0.98$ Fc = 252.5, Ft = 7.71
Maíz regadío	R_{min} con $R_{<MM}$	$R_{min} = 0.66 R_{<MM} + 4405.25$ $R^2 = 0.27$ Fc = 2.9, Ft = 5.32	$R_{min} = 1.35 R_{<MM}$ $R^2 = 0.79$ Fc = 34.7, Ft = 5.12
		$R_{max} = 0.46 R_{MB} + 5949.86$ $R^2 = 0.65$ Fc = 14.62, Ft = 5.32	$R_{max} = 0.83 R_{MB}$ $R^2 = 0.98$ Fc = 440.7, Ft = 5.12
Alfalfa regadío	R_{min} con $R_{<MM}$	$R_{min} = 1.49 R_{<MM} – 574.05$ $R^2 = 0.84$ Fc = 25.8, Ft = 6.61	$R_{min} = 1.45 R_{<MM}$ $R^2 = 0.91$ Fc = 63.6, Ft = 5.99
		$R_{max} = 0.96 R_{MB} – 2268.76$ $R^2 = 0.98$ Fc = 216.1, Ft = 6.61	$R_{max} = 0.89 R_{MB}$ $R^2 = 0.99$ Fc =535.3, Ft = 5.99

Diferencias en los extremos inferiores del recorrido. Se ha examinado también la cuantía de las diferencias entre los valores de R_{min} (pregunta 2) y $R_{<MM}$, expresadas como porcentaje del valor R_{min}, para normalizar. La distribución del número de casos según la cuantía de la diferencia se reproduce en la tabla 265.

Tabla 265.

Cociente $|(R_{<MM} - R_{min}):R_{min}|$. Encuesta a los agricultores (1999 – 00, primer día).

	Número de casos										
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**	**>1**
Cebada secano	6	1	2	5	1	1	1	2	1	2	
Trigo secano	3	1	4	1		1		1		1	
Trigo regadío			1	1	1				1	1	
Maíz regadío	2	1	1		4		1			1	
Alfalfa regadío	1		1	1			1			1	2

En cualquiera de los cultivos, los casos obtenidos con diferencias del 100% (ver Tabla 266) se pueden considerar “outliers”, ya que se corresponden con respuestas $R_{<MM} = 0$, que obviamente es una forma de hablar más que de describir un recorrido.

En cebada de secano, es importante subrayar que, excepto en un caso en que $R_{min} < R_{<MM}$, en el resto de los casos $R_{min} \geq R_{<MM}$.

En trigo de secano, en 3 observaciones la diferencia es inferior o igual al 10% y en todos los casos se observa que $R_{<MM} \leq R_{min}$.

Para el trigo de regadío, en 1 caso se obtiene un rendimiento mínimo nulo, y en todos los casos la diferencia es negativa, con $R_{<MM} < R_{min}$.

Para el maíz de regadío, en 1 observación el rendimiento mínimo es nulo, y excepto en un caso, se observa que $R_{<MM} < R_{min}$.

En el caso de la alfalfa de regadío, en un caso el rendimiento mínimo es nulo, y excepto en dos casos, se observa $R_{min} > R_{<MM}$. También en dos casos, la diferencia es superior a la unidad.

En la Tabla 266 pueden verse los casos en que se observó una diferencia nula o máxima (100%) entre los dos valores de rendimientos.

Tabla 266.

Diferencias nula y máxima entre $R_{<MM}$ y R_{min}. Encuesta a los agricultores (1999 – 00, primer día).

Cultivo	**Encuestas**	**Diferencia $\|R_{<MM} - R_{min}\|$**	
		0%	**100%**
Cebada secano	22	2	2
Trigo secano	12	2	1
Trigo regadío	5	0	1
Maíz regadío	10	0	1
Alfalfa regadío	7	0	1

Obteniéndose una baja cantidad de respuestas nulas en todos los cultivos.

Diferencias en los extremos superiores del recorrido. Se ha estudiado el número de casos según la cuantía de las diferencias entre los valores de R_{max} (pregunta 2) y $R_{>MB}$ (pregunta 3), expresadas como porcentaje del valor R_{max}. Los resultados se reproducen en la tabla 267.

Tabla 267.

Cociente $|(R_{>MB}-R_{max}):R_{max}|$. Encuesta a los agricultores (1999 – 00, primer día).

	Número de casos										
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**	**>1**
Cebada secano	13	5	1	1	1			1			
Trigo secano	6	4		1		1					1
Trigo regadío	3	1	1								
Maíz regadío	5	2	1	1		1					
Alfalfa regadío	3	3									1

En cebada de secano, y maíz y alfalfa de regadío, se obtiene en todas las encuestas $R_{>MB} \geq R_{max}$.

En el caso del trigo de secano y de regadío, excepto en un caso, se obtiene $R_{>MB} \geq R_{max}$.

En la Tabla 268 pueden verse los casos en que se observó $R_{>MB} = R_{max}$.

Tabla 268.

Casos en que $R_{>MB} = R_{max}$. Encuesta a los agricultores (1999 – 00, primer día).

Cultivo	**Encuestas**	**$R_{>MB} = R_{max}$**
Cebada secano	22	6
Trigo secano	12	4
Trigo regadío	5	2
Maíz regadío	10	1
Alfalfa regadío	7	1

Obteniéndose unos porcentajes bajos de coincidencias de rendimientos.

En definitiva, se observaron mayores diferencias relativas para los extremos inferiores que para los superiores en todos los cultivos.

Diferencias en los dos extremos del recorrido. Se han analizado simultáneamente las diferencias de los extremos inferiores y superiores, resumidas en la tabla 269.

Tabla 269.

Valores de los extremos inferiores R_{min}, $R_{<MM}$ y superiores R_{max}, $R_{>MB}$. Encuesta a los agricultores (1999 – 00, primer día).

Cultivo	$R_{<MM}=0$	$R_{min}=R_{<MM}$ $R_{max}=R_{>MB}$	$R_{min}=R_{<MM}$ $R_{max}>R_{>MB}$	$R_{min}=R_{<MM}$ $R_{max}<R_{>MB}$	$R_{max}=R_{>MB}$ $R_{min}>R_{<MM}$	$R_{max}=R_{>MB}$ $R_{min}<R_{<MM}$	$R_{max}<R_{MB}$ $R_{min}>R_{<MM}$	$R_{max}>R_{>MB}$ $R_{min}<R_{<MM}$	Total
Cebada secano	2	3	0	2	2	0	13	0	22
Trigo secano	1	2	0	0	2	0	7	0	12
Trigo regadío	1	0	0	0	2	0	2	0	5
Maíz regadío	1	0	0	0	1	0	8	0	10
Alfalfa regadío	1	0	0	0	0	0	4	0	7

En el caso de la alfalfa de regadío, en dos casos se da un rendimiento *mínimo* inferior al rendimiento *muy malo*, junto con un rendimiento *máximo* inferior al rendimiento *muy bueno*, situación no reflejada en la tabla.

Se observa una clara tendencia a incrementar el recorrido cuando se define en términos de lenguaje natural en todos los cultivos.

#8.6.2.- RESULTADOS CON LA INFORMACIÓN DEL SEGUNDO DÍA

En la tabla 270 se indican los resultados de aplicar el índice de continuidad de intervalos a los datos obtenidos con el cuestionario del segundo día (1999 – 00).

Tabla 270.

Número de respuestas que presentan discontinuidades / número de respuestas totales. Encuesta a los agricultores (1999 – 00, segundo día).

Cultivo	Índice de Continuidad de Intervalos
Cebada secano	9/22
Trigo secano	4/12
Trigo regadío	2/5
Maíz regadío	5/10
Alfalfa regadío	4/7

Los resultados de las comparaciones entre el recorrido (largo) R_L y el corto R_C, se han resumido en las tablas 271 a 274.

Tabla 271.

Regresión $R_L = a\ R_C + b$. Encuesta a los agricultores (1999 – 00, segundo día).

	CULTIVO				
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	1.03	1.07	1.22	0.82	1.60
b	1391.24	1129.0	1283.67	4681.85	2.05
Fc	39.88	13.62	2.4	14.55	20.52
Ft	4.35	4.96	10.13	5.32	6.61
R^2	0.67	0.58	0.44	0.64	0.80

La misma regresión, forzando que el término independiente *b* fuera nulo, ofrece los resultados de la tabla 272.

Tabla 272.

Regresión $R_L = a\ R_C$ (con $b = 0$). Encuesta a los agricultores (1999 – 00, segundo día).

	CULTIVO				
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	1.52	1.53	1.67	1.46	1.60
b	0	0	0	0	0
Fc	607.32	295.86	116.8	145.4	97.5
Ft	4.32	4.84	7.71	5.12	5.99
R^2	0.97	0.96	0.97	0.65	0.59

Obteniéndose un buen ajuste para todos los cultivos, un poco peor en maíz y alfalfa de regadío. A continuación, se realizó la regresión $(R_L - R_C) = a\ R_C + b$ (Tabla 273).

Tabla 273.

Regresión $(R_L - R_C) = a\ R_C + b$. Encuesta a los agricultores (1999 – 00, segundo día).

	CULTIVO				
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.03	0.07	0.22	-0.18	0.60
b	1391.24	1129.01	1283.67	4681.85	2.05
Fc	0.04	0.06	0.08	0.74	2.89
Ft	4.35	4.96	10.13	5.32	6.61
R^2	0.00	0.01	0.02	0.08	0.37

Se obtuvo un mal resultado, con ningún cultivo significativo para esta regresión y se repitió la regresión para el caso de $b = 0$ en cebada y trigo de secano (Tabla 274).

Tabla 274.

Regresión $(R_L – R_C) = a\ R_C\ (b = 0)$. Encuesta a los agricultores (1999 – 00, segundo día).

CULTIVO		
	CEBADA SECANO	**TRIGO SECANO**
a	0.52	0.53
b	0	0
Fc	70.29	35.65
Ft	4.32	4.84
R^2	0.77	0.76

Obteniendo un ajuste significativo en ambos cultivos.

También se calcularon los cocientes R_C / R_L y $(R_L - R_C) / R_L$ (Tablas 275 y 276, respectivamente).

Tabla 275.

Cociente R_C / R_L. Encuesta a los agricultores (1999 – 00, segundo día).

	Número de casos									
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0 .5**	**0.5- 0.6**	**0.6- 0.7**	**0.7- 0.8**	**0.8- 0.9**	**0.9-1**
Cebada secano				1	2	4	11	3	1	
Trigo secano					1	4	4	2	1	
Trigo regadío					1	2	1	1		
Maíz regadío			1		1	2	3	2	1	
Alfalfa regadío			1		1		2	2	1	

Tabla 276.

Cociente $(R_L - R_C) / R_L$. Encuesta a los agricultores (1999 – 00, segundo día).

	Número de casos									
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano		1	3	14	2	2				
Trigo secano		1	2	5	3	1				
Trigo regadío			2		2	1				
Maíz regadío		1	2	5		1	1			
Alfalfa regadío		1	2	2	1		1			

#8.6.2.1.- COHERENCIA DE LOS RECORRIDOS

Se comparó, mediante regresión, el recorrido largo de la distribución, según los resultados de la **pregunta 3** y el recorrido calculado a partir de las respuestas a la pregunta **número 2**, mediante la ecuación $R_L = a\ (R_{max} - R_{min}) + b$, con *b* igual o distinta de cero (tablas 277 y 278). Se obtuvo, en general, un mal ajuste en esta regresión para *b* distinta de cero, aunque mejor para *b* igual a cero, en la línea de confirmar el mayor recorrido para las estimaciones mediante lenguaje natural.

Tabla 277.

Regresión $R_L = a\ (R_{max} - R_{min}) + b$. Encuesta a los agricultores (1999 – 00, segundo día).

	CULTIVO				
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.33	0.50	0.18	0.76	1.07
b	3114.58	2190.43	4027.67	4887.24	2281.26
Fc	2.61	4.19	0.13	10.48	317.33
Ft	4.35	4.96	10.13	5.32	6.61
R^2	0.12	0.30	0.04	0.57	0.98

Tabla 278.

Regresión $R_L = a\ (R_{max} - R_{min})$ ($b = 0$). Encuesta a los agricultores (1999 – 00, segundo día).

	CULTIVO				
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	1.25	1.20	1.25	1.41	1.18
b	0	0	0	0	0
Fc	138.9	109.7	22.9	128.6	625.9
Ft					
R^2	0.87	0.91	0.85	0.94	0.99

También en la regresión $R_C = a\ (R_{max} - R_{min}) + b$ se obtuvo en general un mal resultado (Tabla 279).

Tabla 279.

Regresión $R_C = a\ (R_{max} - R_{min}) + b$. Encuesta a los agricultores (1999 – 00, segundo día).

	CULTIVO				
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.30	0.22	0.34	0.75	0.55
b	1726.72	1654.82	1608.73	1398.80	2998.61
Fc	3.59	1.32	3.47	10.68	23.63
Ft	4.35	4.96	10.13	5.32	6.61
R^2	0.15	0.12	0.54	0.57	0.82

#8.6.2.2.- AMPLITUD DE LOS INTERVALOS

En la tabla 280 se resume el análisis realizado sobre la amplitud de los intervalos.

Tabla 280.

Número de veces en que el intervalo ( ] toma el valor. Encuesta a los agricultores (1999 – 00, encuesta segundo día).

	Mayor	**0-0.2**	**0.2-0.4**	**0.4-0.6**	**0.6-0.8**	**0.8-1**	**Mediana**	**Media**
Cebada secano								
Muy Malo	8/22	1	5	3	4	9	0.72	0,67
Malo	13/22	0	4	2	1	15	1	0.81
Normal	10/22	0	5	3	3	11	0.80	0.73
Bueno	9/22	1	1	4	5	11	0.82	0.76
Muy Bueno	6/22	2	4	5	4	7	0.61	0.63
Trigo Secano								
Muy Malo	5/12	0	4	1	1	6	0.78	0.70
Malo	6/12	1	1	1	2	7	0.91	0.78
Normal	6/12	0	3	2	0	7	0.94	0.74
Bueno	3/12	0	1	1	6	4	0.70	0.75
Muy Bueno	3/12	1	3	0	4	4	0.67	0.64
Trigo regadío								
Muy Malo	1/5	2	1	0	1	1	0.25	0.44
Malo	0/5	1	1	1	1	1	0.56	0.50
Normal	2/5	0	1	1	0	3	0.89	0.74
Bueno	0/5	0	3	2	0	0	0.33	0.41
Muy Bueno	2/5	1	0	1	1	2	0.67	0.67
Maíz regadío								
Muy Malo	2/10	1	4	2	1	2	0.45	0.51
Malo	5/10	1	1	1	1	6	0.93	0.74
Normal	4/10	2	0	1	2	5	0.81	0.72
Bueno	3/10	1	6	0	0	3	0.30	0.50
Muy Bueno	4/10	1	4	1	0	4	0.45	0.61
Alfalfa regadío								
Muy Malo	1/7	1	2	0	1	3	0.70	0.63
Malo	4/7	2	0	0	1	4	1	0.74
Normal	1/7	1	0	2	3	1	0.64	0.64
Bueno	1/7	2	0	3	1	1	0.60	0.54
Muy Bueno	0/7	1	5	0	0	1	0.40	0.40

Igual que en el primer día, no se observa una tendencia a asignar mayores amplitudes a un intervalo concreto.

A continuación se midió la amplitud de los intervalos máximo y mínimo y se expresó como porcentaje del recorrido (largo) (Tabla 281) y como porcentaje del punto medio del intervalo de rendimientos *normales* (Tabla 282).

Tabla 281.

Amplitud del intervalo máximo y del mínimo en las diferentes encuestas, expresada como porcentaje del recorrido (largo). Encuesta a los agricultores (1999 – 00, segundo día).

		Número de casos									
Cultivo	**Amplitud**	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Cebada secano	Máxima		3	10	8	1					
	Mínima	11	11								
Trigo secano	Máxima		1	6	4		1				
	Mínima	3	9								
Trigo regadío	Máxima			1	3	1					
	Mínima	4	1								
Maíz regadío	Máxima		1	3	2	4					
	Mínima	7	3								
Alfalfa regadío	Máxima			2	3	2					
	Mínima	5	2								

Tabla 282.

Amplitud del intervalo máximo y del mínimo en las diferentes encuestas, expresada como porcentaje del punto medio del intervalo de rendimientos *normales*. Encuesta a los agricultores (1999 – 00, segundo día).

		Número de casos										
Cultivo	**Amplitud**	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**	**>1**
Cebada secano	Máxima		5	1	6	2	4	2		1		1
	Mínima	6	13	3								
Trigo secano	Máxima		4	1	1	2	3					1
	Mínima	2	8	2								
Trigo regadío	Máxima			1	2	1			1			
	Mínima	2	3									
Maíz regadío	Máxima		2	3	1	2	1		1			
	Mínima	6	4									
Alfalfa regadío	Máxima		2	1	2	1			1			
	Mínima	5	2									

#8.6.2.3. ANÁLISIS DE LOS EXTREMOS

Comparación de los extremos. Se comparó el valor R_{min} de la pregunta 2 con el valor $R_{<MM}$ de la pregunta 3, y el valor R_{max} con $R_{>MB}$ (tabla 283).

Tabla 283.

Comparación entre poblaciones de rendimientos máximos y mínimos. Encuesta a los agricultores (1999 – 00, segundo día).

Cultivo	**Comparación**	**Spearman (r_S)**	**Muestras independientes**	**Wilcoxon (α=0.05)**
Cebada secano	R_{min} con R_{mm}	0.520	No	Existen diferencias
	R_{max} con R_{MB}	0.562	No	Existen diferencias
Trigo secano	R_{min} con R_{mm}	0.614	No	No diferencias
	R_{max} con R_{MB}	0.846	No	No diferencias
Trigo de regadío	R_{min} con R_{mm}	0.475	Sí	No diferencias
	R_{max} con R_{MB}	0.9	Sí	No diferencias
Maíz de regadío	R_{min} con R_{mm}	0.639	No	No diferencias
	R_{max} con R_{MB}	0.594	No	Existen diferencias
Alfalfa de regadío	R_{min} con R_{mm}	0.928	No	Existen diferencias
	R_{max} con R_{MB}	0.938	No	No diferencias

No se observa una pauta clara en la semejanza entre poblaciones.

Se efectuó también la regresión entre el rendimiento *mínimo* y el rendimiento *muy malo* y el rendimiento *máximo* y el rendimiento *muy bueno*, para los casos en que la constante fuera igual o diferente de cero (Tabla 284):

Tabla 284.

Regresión entre rendimientos máximos y mínimos. Encuesta a los agricultores (1999 – 00, segundo día).

Cultivo	**Comparación**	**y = a x +b, b ≠ 0**	**y = ax, b = 0**
Cebada secano	R_{min} con $R_{<MM}$	$R_{min} = 0.57\ R_{<MM} + 1198.26$ $R^2 = 0.36$ Fc = 11.10, Ft = 4.35	$R_{min} = 1.27\ R_{<MM}$ $R^2 = 0.78$ Fc = 73.12, Ft = 4.32
	R_{max} con $R_{>MB}$	$R_{max} = 0.51\ R_{>MB} + 2103.19$ $R^2 = 0.24$ Fc = 6.30, Ft = 4.35	$R_{max} = 0.91\ R_{>MB}$ $R^2 = 0.97$ Fc = 717.65, Ft = 4.32
Trigo secano	R_{min} con $R_{<MM}$	$R_{min} = 0.02\ R_{<MM} + 1682.04$ $R^2 = 0.00$ Fc = 0.01, Ft = 4.96	$R_{min} = 1.17\ R_{<MM}$ $R^2 = 0.82$ Fc = 50.00, Ft = 4.84
	R_{max} con $R_{>MB}$	$R_{max} = 1.24\ R_{>MB} – 1426.15$ $R^2 = 0.64$ Fc = 18.10, Ft = 4.96	$R_{max} = 0.95\ R_{>MB}$ $R^2 = 0.98$ Fc = 509.29, Ft = 4.84
Trigo regadío	R_{min} con $R_{<MM}$	$R_{min} = 0.50\ R_{<MM} + 2262.52$ $R^2 = 0.22$ Fc = 0.83, Ft = 10.13	$R_{min} = 1.37\ R_{<MM}$ $R^2 = 0.75$ Fc = 12.0, Ft = 7.71
	R_{max} con $R_{>MB}$	$R_{max} = 1.35\ R_{>MB} – 2315.0$ $R^2 = 0.82$ Fc = 13.4, Ft = 10.13	$R_{max} = 0.99\ R_{>MB}$ $R^2 = 0.99$ Fc = 1570.5, Ft = 7.71
Maíz regadío	R_{min} con $R_{<MM}$	$R_{min} = 0.54\ R_{<MM} + 3740.79$ $R^2 = 0.34$ Fc = 4.17, Ft = 5.32	$R_{min} = 1.08\ R_{<MM}$ $R^2 = 0.83$ Fc = 43.4, Ft = 5.12

Cultivo	**Comparación**	**y = a x +b, b ≠ 0**	**y = ax, b = 0**
	R_{max} con $R_{>MB}$	$R_{max} = 0.58\ R_{>MB} + 4341.87$ $R^2 = 0.29$ Fc = 3.26, Ft = 5.32	$R_{max} = 0.86\ R_{>MB}$ $R^2 = 0.96$ Fc = 190.5, Ft = 5.12
Alfalfa regadío	R_{min} con $R_{<MM}$	$R_{min} = 1.06\ R_{<MM} + 2249.94$ $R^2 = 0.96$ Fc = 122.2, Ft = 6.61	$R_{min} = 1.20\ R_{<MM}$ $R^2 = 0.97$ Fc = 196.5, Ft = 5.99
	R_{max} con $R_{>MB}$	$R_{max} = 0.99\ R_{>MB} – 494.38$ $R^2 = 0.99$ Fc = 1126.4, Ft = 6.61	$R_{max} = 0.98\ R_{>MB}$ $R^2 = 0.99$ Fc = 3628.8, Ft = 5.99

Obteniéndose mejores ajustes en el caso de los extremos superiores, confirmándose la tendencia ya observada de ampliación de recorridos al ser expresados mediante lenguaje natural.

Diferencias en los extremos inferiores del recorrido. Se examinó también la cuantía de las diferencias entre los valores de R_{min} (pregunta 2) y $R_{<MM}$, expresadas como porcentaje del valor R_{min}, para normalizar. La distribución del número de casos según la cuantía de la diferencia se reproduce en la tabla 285.

Tabla 285.

COCIENTE $|(R_{<MM}-R_{min}):R_{min}|$

	Número de casos										
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**	**>1**
Cebada secano	6	2	1		5	2	2	1	1	2	
Trigo secano	1		2	1			2	2	1	2	1
Trigo regadío		1		2		1				1	
Maíz regadío	1					2	1	1	1	1	3
Alfalfa regadío	2	1	1	1					1	1	

En cebada de secano, en 13 casos se obtuvieron diferencias apreciables, con una diferencia superior al 40% del valor de R_{min}. De ellos, el caso con diferencias del 100% se puede considerar "outlier", ya que se corresponde con respuestas $R_{<MM} = 0$, que obviamente es una forma de hablar más que de describir un recorrido. Es importante subrayar que excepto en un caso $R_{min} \geq R_{<MM}$.

En el caso del trigo de secano, excepto en 3 casos, $R_{<MM} \leq R_{min}$. En un caso la diferencia es superior a la unidad.

Para el trigo de regadío, excepto en un caso, $R_{min} \geq R_{<MM}$.

Para el maíz de regadío, excepto en 4 casos $R_{min} \geq R_{<MM}$ y en tres se observa una diferencia entre rendimientos superior a la unidad.

Por último, en el caso de la alfalfa de regadío, en todos los casos se observa $R_{min} \geq R_{<MM}$.

En la Tabla 286 pueden verse los casos en que se observó una diferencia nula o del 100% entre los dos valores de rendimientos.

Tabla 286.

Diferencias nula y máxima entre $R_{<MM}$ y R_{min}.

Cultivo	**Encuestas**	**Diferencia $\|R_{<MM} - R_{min}\|$**	
		0%	**100%**
Cebada secano	22	5	2
Trigo secano	12	1	2
Trigo regadío	5	0	1
Maíz regadío	10	0	1
Alfalfa regadío	7	0	1

Diferencias en los extremos superiores del recorrido. Se estudió el número de casos según la cuantía de las diferencias entre los valores de R_{max} (pregunta 2) y $R_{>MB}$, expresadas como porcentaje del valor R_{max}. Los resultados se reproducen en la tabla 287.

Tabla 287.

COCIENTE $|(R_{>MB} - R_{max}):R_{max}|$.

	Número de casos										
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**	**>1**
Cebada secano	14	3	2	1	1			1			
Trigo secano	8	3									1
Trigo regadío	5										
Maíz regadío	6	2	1								1
Alfalfa regadío	6	1									

En cebada y trigo de secano, excepto en tres casos (dos de ellos correspondientes al mismo agricultor, y el tercero distinto por no tener simultáneamente los dos cultivos), se observó $R_{>MB} \geq R_{max}$.

En trigo y alfalfa de regadío, excepto en un caso se observó $R_{>MB} \geq R_{max}$.

En maíz de regadío, en todos los casos se observó $R_{>MB} \geq R_{max}$.

En la Tabla 288 pueden verse los casos en que se observó $R_{>MB} = R_{max}$.

Tabla 288.

Casos en que $R_{>MB} = R_{max}$.

Cultivo	**Encuestas**	**$R_{>MB} = R_{max}$**
Cebada secano	22	8
Trigo secano	12	4
Trigo regadío	5	2
Maíz regadío	10	3
Alfalfa regadío	7	2

En definitiva, se observaron mayores diferencias relativas para los extremos inferiores que para los superiores, para todos los cultivos.

Diferencias en los dos extremos del recorrido. Se analizaron simultáneamente las diferencias de los extremos inferiores y superiores, resumidas en la tabla 289.

Tabla 289.

Valores de los extremos inferiores R_{min}, $R_{<MM}$ y superiores R_{max}, $R_{>MB}$.

Cultivo	$R_{<MM} = 0$	$R_{min}=R_{<MM}$ $R_{max}=R_{>MB}$	$R_{min}=R_{<MM}$ $R_{max}>R_{>MB}$	$R_{min}=R_{<MM}$ $R_{max}<R_{>MB}$	$R_{max}=R_{>MB}$ $R_{min}>R_{<MM}$	$R_{max}=R_{>MB}$ $R_{min}<R_{<MM}$	$R_{max}<R_{>MB}$ $R_{min}>R_{<MM}$	$R_{max}>R_{>MB}$ $R_{min}<R_{<MM}$	Total
Cebada secano	2	4	1	0	4	0	9	1	22
Trigo secano	1	1	0	0	2	1	4	2	12
Trigo regadío	1	0	0	0	1	1	2	0	5
Maíz regadío	1	0	0	0	2	0	3	0	10
Alfalfa regadío	1	0	0	0	1	0	4	0	7

En cebada y trigo de secano, y alfalfa de regadío, se observa en un caso un rendimiento *mínimo* superior al *muy malo* y a la vez un rendimiento *máximo* superior al *muy bueno*, situación no reflejada en la Tabla 289.

En maíz de regadío, se observa en 4 ocasiones que el rendimiento *mínimo* es inferior al rendimiento *muy malo* y a la vez que el rendimiento *máximo* es inferior al *muy bueno*, situación no reflejada en la Tabla 289.

Se observó en todos los cultivos una clara tendencia a incrementar los recorridos al expresar los intervalos mediante lenguaje natural.

#8.7.- ANÁLISIS DEL CENTRO DEL RECORRIDO

#8.7.1.- ANÁLISIS DE LOS RENDIMIENTOS NORMALES Y MÁS FRECUENTES (PRIMER DÍA)

Diferencias en el centro del recorrido. Se compararon los rendimientos normales (R_{NN}) y los rendimientos más frecuentes (R_{mf}). Para ello, se comprobó si el rendimiento más frecuente se encontraba dentro del intervalo de rendimientos normales. Los resultados pueden consultarse en la tabla 290, donde se puede apreciar un porcentaje elevado para todos los cultivos.

Tabla 290.

Respuestas pertenecientes al intervalo normal. (Encuesta primer día, 1999 – 00).

Cultivo	Respuestas pertenecientes al intervalo normal	Total respuestas	Porcentaje (%)
Cebada secano	19	22	86
Trigo secano	10	12	83
Trigo regadío	3	5	60
Maíz regadío	8	10	80
Alfalfa regadío	3	7	43

#8.7.2.- RESPUESTAS DEL SEGUNDO DÍA

La comparación de las respuestas R_{mf} (rendimientos más frecuentes) y los valores del intervalo de rendimientos "normales" (el intervalo central de las distribuciones estimadas), con los datos del segundo día, se ha sintetizado en la tabla 291.

Tabla 291.

Respuestas R_{mf} pertenecientes al intervalo normal. Encuesta a los agricultores (1999 – 00, segundo día).

Cultivo	**Respuestas pertenecientes al intervalo normal**	**Total respuestas**	**Porcentaje (%)**
Cebada secano	17	22	77
Trigo secano	8	12	67
Trigo regadío	2	5	40
Maíz regadío	6	10	60
Alfalfa regadío	5	7	71

#8.8.- COHERENCIA DE LAS ESTIMACIONES

A partir de las estimaciones de la frecuencia de cada intervalo numérico y de las estimaciones puntuales de los valores (R_{max}, R_{mf}, R_{min}) es posible calcular los valores de las medias y de las varianzas respectivas, compararlas entre sí y con los valores de los rendimientos medios R_m estimados directamente por los agricultores (puntualmente). Llamando x_i al valor de la marca de clase del intervalo i en la estimación de frecuencias, y f_i la frecuencia correspondiente, en lo que sigue se comparan los siguientes valores medios, para los dos días de la encuesta:

$R_{xf} = (\Sigma\, x_i\, f_i)$ con los correspondientes valores medios obtenidos a partir de las estimaciones de (R_{max}, R_{mf}, R_{min});

$R_{xf} = (\Sigma\, x_i\, f_i)$ con la estimación directa puntual del valor medio, R_m;

$R_{xf} = (\Sigma\, x_i\, f_i)$ entre los dos días de la encuesta.

Una vez estudiadas las diferencias entre las estimaciones directas e indirectas de los valores medios, se realizará el correspondiente estudio para el caso de la varianza, mediante la comparación de:

$V_{xf} = (x_i - R_{xf})^2\, f_i$ con las correspondientes varianzas obtenidas a partir de las estimaciones de (R_{max}, R_{mf}, R_{min});

$V_{xf} = (x_i - R_{xf})^2\, f_i$ entre los dos días de la encuesta.

#8.8.1- COHERENCIA DE LOS VALORES MEDIOS

En la tabla 292 se compararon las diferencias de la media calculada en base a las frecuencias atribuidas a cada intervalo $R_{xf} = (\Sigma\, x_i\, f_i)$, y las medias estimadas a partir de (R_{max}, R_{mf}, R_{min}) mediante las aproximaciones triangular y beta – PERT, obteniéndose prácticamente el mismo resultado en ambas aproximaciones, ligeramente mejor en los cultivos de secano.

Tabla 292.

Diferencias relativas de medias calculadas $|(R_{xf} - RMT):R_{xf}|$ y $|(R_{xf} - RMB):R_{xf}|$. Encuesta a los agricultores, primer día (1999-00).

APROXIMACIÓN												
TRIANGULAR	**Número de casos**											
Cultivo	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**>0.5**	**Mediana**
Cebada secano	3	8	3	4	1	1			1	1		0.100
Trigo secano	4	2	2	1	1	1	1					0.100
Trigo regadío		2	1	1		1						0.125
Maíz regadío	3	2	1	1	2		1					0.100
Alfalfa regadío	1	1	1	1	1	1					1	0.175
BETA	**Número de casos**											
Cultivo	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**>0.5**	**Mediana**
Cebada secano	4	7	5	1	2		2			1		0.100
Trigo secano	4	2	3	1	1			1				0.100
Trigo regadío		2	1	1		1						0.125
Maíz regadío	4	2	1		2	1						0.075
Alfalfa regadío	2	1		1	2						1	0.175

El examen de las poblaciones de valores medios estimados por ambos métodos, resumido en la tabla 293, indica mejores resultados para los cultivos de regadío que para los de secano, tal como ocurría con los expertos.

Tabla 293.

Comparación entre las poblaciones de rendimientos medios (RMT con R_{xf} y RMB con R_{xf}). Encuesta a los agricultores, primer día (1999-00).

Comparación con la distribución triangular			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.830	No	Diferencias
Trigo secano	0.967	No	Diferencias
Trigo regadío	0.100	Sí	No diferencias
Maíz regadío	0.512	Sí	No diferencias
Alfalfa regadío	0.893	No	No diferencias
Comparación con la distribución beta			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.839	No	Diferencias
Trigo secano	0.965	No	Diferencias
Trigo regadío	0.100	Sí	No diferencias
Maíz regadío	0.648	No	No diferencias
Alfalfa regadío	0.893	No	No diferencias

En la tabla 294 se resume el resultado de las regresiones realizadas entre las diferentes estimaciones. Aunque en los cultivos de secano se observaban diferencias entre poblaciones, los ajustes obtenidos no son del todo malos, y la realización del mismo análisis para el caso b = 0, resumido en la tabla 295, mejora aún los resultados para todos los cultivos.

Tabla 294.

Regresión *RMT = a* R_{xf} *+ b y RMB = a* R_{xf} *+ b*. Encuesta a los agricultores, primer día (1999-00).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.82	0.97	3.76	0.57	1.18
b	845.92	367.40	-11444.32	4188.14	-3846.01
R^2	0.70	0.90	0.71	0.50	0.91
Fc	46.52	88.83	7.25	7.84	52.24
Ft	4.35	4.96	10.13	5.32	6.61
DISTRIBUCIÓN BETA					
a	0.85	1.02	3.56	0.60	1.18
b	762.69	217.53	-10513.82	3824.84	-3407.78
R^2	0.70	0.88	0.68	0.64	0.93
Fc	47.8	72.1	6.32	14.08	61.76
Ft	4.35	4.96	10.13	5.32	6.61

Tabla 295.

Regresión *RMT = a* R_{xf} (*b* = 0) y *RMB = a* R_{xf} (*b* = 0). Encuesta a los agricultores, primer día (1999-00).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	1.09	1.09	1.11	0.93	1.03
b	0	0	0	0	0
R^2	0.98	0.99	0.99	0.97	0.95
Fc	1326.3	1339.5	287.1	268.42	125.71
Ft	4.32	4.84	7.71	5.12	5.99
DISTRIBUCIÓN BETA					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	1.09	1.09	1.13	0.93	1.04
b	0	0	0	0	0
R^2	0.98	0.99	0.99	0.98	0.96
Fc	1338.8	1090.9	317.3	392.6	156.0
Ft	4.32	4.84	7.71	5.12	5.99

#8.8.2.- COHERENCIA DE LAS VARIANZAS

En la tabla 296 se realiza la comparación de las varianzas calculadas a partir de la distribución de frecuencias $V_{xf} = (\Sigma (x_i - R_{xf})^2 f_i)$, y las derivadas de las aproximaciones a funciones triangulares y beta-PERT. Se observan, de la misma forma que en la encuesta a expertos y la de los agricultores del año 1998 – 99, unas diferencias relativas superiores en el caso de las varianzas que en el caso de los valores medios calculados anteriormente. También se observan unos errores superiores en la aproximación triangular, de la misma forma que ocurría en los agricultores del año 1998 – 99, y a diferencia de lo que ocurría en los expertos.

Tabla 296.

Diferencias relativas de las varianzas calculadas $|(V_{xf}\text{-}VT):V_{xf}|$ y $|(V_{xf}\text{-}VB):V_{xf}|$. Encuesta a los agricultores, primer día (1999-00).

APROXIMACIÓN												
TRIANGULAR	**Número de casos**											
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1.0**	**>1.0**	**Mediana**
Cebada secano			1		4	3	3	5	5	1		0.700
Trigo secano		1		1	1	2	3	2	1	1		0.633
Trigo regadío					1		2	1	1			0.675
Maíz regadío				1			3		1	3	2	0.900
Alfalfa regadío				1			2		1	2	1	0.850
BETA	**Número de casos**											
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1.0**	**>1.0**	**Mediana**
Cebada secano	1	3	4		3	4	1	3	2	1		0.500
Trigo secano	1	1	2		2	2	1	1	1	1		0.500
Trigo regadío		1			1	2			1			0.525
Maíz regadío	1				3				2	2	2	0.850
Alfalfa regadío		1			2			1	2		1	0.750

En la tabla 297 se resume el estudio de las poblaciones de varianzas obtenidas por los diferentes métodos. Se obtienen buenos resultados sólo en el caso de la alfalfa de regadío en la aproximación triangular, y en el caso del maíz y la alfalfa de regadío en la aproximación beta – PERT.

Tabla 297.

Comparación entre las poblaciones de varianzas (VT con V_{xf} y VB con V_{xf}). Encuesta a los agricultores, primer día (1999-00).

Comparación con la distribución triangular			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.403	No	Diferencias
Trigo secano	0.456	Sí	Diferencias
Trigo regadío	0.475	Sí	Diferencias
Maíz regadío	-0.164	Sí	Diferencias
Alfalfa regadío	0.348	Sí	No diferencias
Comparación con la distribución beta			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.449	No	Diferencias
Trigo secano	0.448	Sí	Diferencias
Trigo regadío	0.300	Sí	Diferencias
Maíz regadío	-0.076	Sí	No diferencias
Alfalfa regadío	0.348	Sí	No diferencias

Las tablas 298 y 299 muestran los resultados de realizar la regresión de las diferentes estimaciones, que son más bien malos.

Tabla 298.

Regresión $VT = a\ V_{xf} + b$ y $VB = a\ V_{xf} + b$. Encuesta a los agricultores, primer día (1999-00).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.07	0.11	0.02	-0.02	0.12
x	193571.53	158154.55	300762.75	1248276.42	2117883.58
R^2	0.09	0.14	0.01	0.01	0.51
Fc	1.87	1.64	0.03	0.06	5.21
Ft	4.35	4.96	10.13	5.32	6.61
DISTRIBUCIÓN BETA					
a	0.11	0.17	0.03	-0.03	0.18
b	298815.56	249475.48	459747.45	912698.19	3533257.85
R^2	0.08	0.13	0.01	0.01	0.45
Fc	1.84	1.46	0.03	0.05	4.08
Ft	4.35	4.96	10.13	5.32	6.61

Tabla 299.

Regresión $VT = a\ V_{xf}$ (b = 0) y $VB = a\ V_{xf}$ (b = 0). Encuesta a los agricultores, primer día (1999-00).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.21	0.28	0.22	0.09	0.16
b	0	0	0	0	0
R^2	0.61	0.66	0.68	0.15	0.62
Fc	33.28	21.8	8.48	1.6	10.0
Ft	4.32	4.84	7.71	5.12	5.99
DISTRIBUCIÓN BETA					
a	0.32	0.43	0.33	0.15	0.25
b	0	0	0	0	0
R^2	0.62	0.66	0.68	0.16	0.58
Fc	33.8	20.9	8.42	1.68	8.30
Ft	4.32	4.84	7.71	5.12	5.99

#8.8.3- COHERENCIA DE LOS VALORES MEDIOS EN EL SEGUNDO DÍA

En la tabla 300 se comparan las diferencias de la media calculada en base a las frecuencias atribuidas a cada intervalo $R_{xf} = (\Sigma\ x_i\ f_i)$, y las medias estimadas a partir de (R_{max}, R_{mf}, R_{min}) mediante las aproximaciones triangular y beta – PERT, obteniéndose mejores resultados en maíz y alfalfa de regadío para ambas aproximaciones.

Tabla 300.

Diferencias relativas de medias calculadas $|(R_{xf}$ - RMT)$:R_{xf}|$ y $|(R_{xf}$ - RMB)$:R_{xf}|$. Encuesta a los agricultores, segundo día (1999-00).

APROXIMACIÓN												
TRIANGULAR	**Número de casos**											
Cultivo	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**>0.5**	**Mediana**
Cebada secano	8	1	5	4	1	1	1				1	0.120
Trigo secano	4	3	2	1		1					1	0.083
Trigo regadío		1	2		1		1					0.138
Maíz regadío	6		1	2			1					0.042
Alfalfa regadío	4	1		1			1					0.044

BETA	Número de casos											
Cultivo	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**>0.5**	**Mediana**
Cebada secano	6	4	4	4	2			1			1	0.112
Trigo secano	4	2	3		2						1	0.100
Trigo regadío	1	1		1	1		1					0.175
Maíz regadío	5	2		1	1	1						0.050
Alfalfa regadío	2	3			1	1						0.075

El examen de las poblaciones de valores medios estimados por ambos métodos viene resumido en la tabla 301, donde, al igual que en el primer día, se obtienen buenos resultados en los cultivos de regadío, para ambas aproximaciones.

Tabla 301.

Comparación entre las poblaciones de rendimientos medios (RMT con R_{xf} y RMB con R_{xf}). Encuesta a los agricultores, segundo día (1999-00).

Comparación con la distribución triangular			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.774	No	Diferencias
Trigo secano	0.825	No	Diferencias
Trigo regadío	0.300	Sí	No diferencias
Maíz regadío	0.673	No	No diferencias
Alfalfa regadío	1.000	No	No diferencias
Comparación con la distribución beta			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.688	No	Diferencias
Trigo secano	0.839	No	Diferencias
Trigo regadío	0.500	Sí	No diferencias
Maíz regadío	0.661	No	No diferencias
Alfalfa regadío	0.964	No	No diferencias

En la tabla 302 se resume el resultado de las regresiones realizadas entre las diferentes estimaciones. El mismo análisis para el caso b = 0 viene resumido en la tabla 303. Se obtienen buenos resultados para todos los cultivos en el caso de b igual a cero, con coeficientes "a" muy cercanos a la unidad.

Tabla 302.

Regresión $RMT = a\ R_{xf} + b$ y $RMB = a\ R_{xf} + b$. Encuesta a los agricultores, segundo día (1999-00).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.85	1.15	0.23	0.64	1.03
b	798.36	-432.13	3930.98	3405.13	-1017.14
R^2	0.70	0.58	0.03	0.65	0.93
Fc	47.5	14.01	0.11	14.7	67.0
Ft	4.35	4.96	10.13	5.32	6.61
DISTRIBUCIÓN BETA					
a	0.86	1.31	0.53	0.70	1.00
b	761.10	-818.79	2679.7	2973.88	-415.32
R^2	0.61	0.63	0.17	0.69	0.94
Fc	30.8	17.1	0.62	17.43	75.5
Ft	4.35	4.96	10.13	5.32	6.61

Tabla 303.

Regresión *RMT* = *a* R_{xf} (*b* = 0) y *RMB* = *a* R_{xf} (*b* = 0). Encuesta a los agricultores, segundo día (1999-00).

	CEBADA SECANO	TRIGO SECANO	TRIGO REGADÍO	MAÍZ REGADÍO	ALFALFA REGADÍO
DISTRIBUCIÓN TRIANGULAR					
a	1.11	1.01	1.09	0.93	0.99
b	0	0	0	0	0
R^2	0.99	0.93	0.98	0.98	0.98
Fc	1638.4	154.9	169.3	419.6	239.9
Ft	4.32	4.84	7.71	5.12	5.99
DISTRIBUCIÓN BETA					
a	1.11	1.04	1.12	0.95	0.99
b	0	0	0	0	0
R^2	0.98	0.93	0.98	0.98	0.98
Fc	1122.4	147.6	231.1	469.5	289.7
Ft	4.32	4.84	7.71	5.12	5.99

#8.8.4- COHERENCIA DE LAS VARIANZAS EN EL SEGUNDO DÍA

En la tabla 304 se realiza la comparación de las varianzas calculadas a partir de la distribución de frecuencias $V_{xf} = (\Sigma (x_i - R_{xf})^2 f_i)$, y las derivadas de las aproximaciones a funciones triangulares y beta-PERT. Los errores medidos mediante la mediana de las diferencias son superiores en el caso de las varianzas que en el anteriormente discutido de las medias, siendo también superiores en el caso de la aproximación triangular, de la misma forma que sucedía en el resto de comparaciones de agricultores (primer y segundo día del año 1998 – 99 y primer día del año 1999 – 00), y contrariamente a lo que ocurría con los expertos.

Tabla 304.

Diferencias relativas de las varianzas calculadas $|(V_{xf}\text{-}VT):V_{xf}|$ y $|(V_{xf}\text{-}VB):V_{xf}|$. Encuesta a los agricultores, segundo día (1999-00).

APROXIMACIÓN TRIANGULAR	**Número de casos**											
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1.0**	**>1.0**	**Mediana**
Cebada secano			3	1	2	2	2	5	5	2		0.720
Trigo secano		1	1	2		1	2	2	1	2		0.650
Trigo regadío				1		1		1	2			0.750
Maíz regadío				2	1	1	1	2			3	0.700
Alfalfa regadío					2		3	1			1	0.650
BETA	**Número de casos**											
Cultivo	**0-0.1**	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1.0**	**>1.0**	**Mediana**
Cebada secano	2	3	2	1	2	3	2	3	3	1		0.533
Trigo secano	1	1	1	2	2	1	1		2	1		0.450
Trigo regadío				1			1	2			1	0.725
Maíz regadío	2	1		1	1		2				3	0.500
Alfalfa regadío		2			3		1				1	0.450

En la tabla 305 se resume el estudio de las poblaciones de varianzas obtenidas por los diferentes métodos. En este caso, coinciden los buenos resultados obtenidos en alfalfa de regadío para la aproximación triangular, y mejoran los resultados para la distribución beta respecto al primer día de este año 1999 – 00, obteniendo buenos resultados para todos los cultivos de regadío, y no sólo para el maíz y la alfalfa como ocurría en el primer día.

Tabla 305.

Comparación entre las poblaciones de varianzas (VT con V_{xf} y VB con V_{xf}). Encuesta a los agricultores, segundo día (1999-00).

Comparación con la distribución triangular			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.363	No	Diferencias
Trigo secano	0.556	No	Diferencias
Trigo regadío	0.300	Sí	Diferencias
Maíz regadío	0.762	No	Diferencias
Alfalfa regadío	0.429	Sí	No diferencias
Comparación con la distribución beta			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.396	No	Diferencias
Trigo secano	0.556	No	Diferencias
Trigo regadío	0.300	Sí	No diferencias
Maíz regadío	0.752	No	No diferencias
Alfalfa regadío	0.464	Sí	No diferencias

Las tablas 306 y 307 muestran los resultados de realizar la regresión de las diferentes estimaciones. Se obtienen resultados no demasiado buenos, aunque algo mejores en maíz y alfalfa de regadío.

Tabla 306.

Regresión $VT = a\ V_{xf} + b$ y $VB = a\ V_{xf} + b$. Encuesta a los agricultores, segundo día (1999-00).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.20	0.08	-0.15	0.34	0.50
b	87950.30	160949.45	462951.25	475381.84	445013.97
R^2	0.27	0.09	0.02	0.57	0.88
Fc	7.6	1.0	0.08	10.5	38.4
Ft	4.35	4.96	10.13	5.32	6.61
DISTRIBUCIÓN BETA					
a	0.30	0.13	-0.27	0.51	0.76
b	137029.36	259669.33	747602.8	756762.19	764724.42
R^2	0.28	0.07	0.03	0.55	0.88
Fc	7.7	0.8	0.1	9.7	37.8
Ft	4.35	4.96	10.13	5.32	6.61

Tabla 307.

Regresión $VT = a\ V_{xf}$ (b = 0) y $VB = a\ V_{xf}$ (b = 0). Encuesta a los agricultores, segundo día (1999-00).

DISTRIBUCIÓN TRIANGULAR					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.26	0.22	0.35	0.44	0.51
b	0	0	0	0	0
R^2	0.67	0.52	0.51	0.81	0.92
Fc	42.1	12.1	4.1	37.1	68.9
Ft	4.32	4.84	7.71	5.12	5.99
DISTRIBUCIÓN BETA					
a	0.40	0.35	0.55	0.66	0.78
b	0	0	0	0	0
R^2	0.67	0.49	0.49	0.80	0.92
Fc	43.2	10.6	3.8	35.0	69.7
Ft	4.32	4.84	7.71	5.12	5.99

#8.8.5.- COHERENCIA DE LOS VALORES MEDIOS CALCULADOS Y LAS ESTIMACIONES

Las comparaciones que siguen se realizan entre el valor de los rendimientos medios estimados directamente por los agricultores (puntualmente), R_m, y los calculados a partir de sus estimaciones de frecuencias y valores numéricos de los correspondientes intervalos, para los dos días de la encuesta.

Tabla 308.

COCIENTE $|(R_{xf}-R_m):R_{xf}|$. Encuesta a los agricultores, días primero y segundo (1999-00).

DÍA 1	**Número de casos**											
Cultivo	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**>0.5**	**Mediana**
Cebada secano	6	6	3	1	2		2	1			1	0.092
Trigo secano	5	2	2	2			1					0.075
Trigo regadío		2	1	1		1						0.125
Maíz regadío	4	3	1	1			1					0.067
Alfalfa regadío	1	2	1	1		1					1	0.125
DÍA 2	**Número de casos**											
Cultivo	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**>0.5**	**Mediana**
Cebada secano	6	3	8		1	1	1	1			1	0.112
Trigo secano	5	2	1	2	2							0.075
Trigo regadío	1			2	1		1					0.138
Maíz regadío	4	4	1					1				0.062
Alfalfa regadío	3	1		2		1						0.075

Obteniéndose diferencias relativas semejantes para ambas aproximaciones, con la mitad de las observaciones con una diferencia relativa máxima del 13.8%.

La tabla 309 señala la semejanza entre las dos poblaciones de valores medios, mientras que en las tablas 310 y 311 se reflejan los resultados de realizar una regresión entre ambas estimaciones. No se observan diferencias en las poblaciones comparadas de cultivos de regadío, excepto en el trigo para el día 1. En las regresiones, se obtienen buenos ajustes en todos los cultivos cuando b es nula, con coeficientes "a" cercanos a la unidad.

Tabla 309.

Comparación entre las poblaciones de rendimientos medios (R_{xf} con R_m). Encuesta a los agricultores, días primero y segundo (1999-00).

Comparación día 1			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.818	No	Diferencias
Trigo secano	0.918	No	No diferencias
Trigo regadío	0.075	Sí	Diferencias
Maíz regadío	0.709	No	No diferencias
Alfalfa regadío	0.857	No	No diferencias
Comparación día 2			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.666	No	Diferencias
Trigo secano	0.956	No	Diferencias
Trigo regadío	0.300	Sí	No diferencias
Maíz regadío	0.718	No	No diferencias
Alfalfa regadío	0.964	No	No diferencias

Tabla 310.

Regresión $R_m = a\ R_{xf} + b$. Encuesta a los agricultores, días primero y segundo (1999-00).

Día 1					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.78	1.00	3.08	0.66	1.18
b	967.72	155.56	-8495.88	2906.86	-3473.11
R^2	0.63	0.88	0.66	0.67	0.92
Fc	34.2	70.9	5.94	16.0	55.4
Ft	4.35	4.96	10.13	5.32	6.61
Día 2					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.95	1.17	0.10	0.77	1.13
b	429.24	-233.36	4490.75	1899.54	-1885.03
R^2	0.64	0.92	0.01	0.67	0.94
Fc	35.2	120.1	0.02	16.14	75.5
Ft	4.35	4.96	10.13	5.32	6.61

Tabla 311.

Regresión $R_m = a\ R_{xf}$ (b = 0). Encuesta a los agricultores, días primero y segundo (1999-00).

Día 1					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	1.09	1.05	1.12	0.92	1.04
b	0	0	0	0	0
R^2	0.98	0.99	0.99	0.98	0.96
Fc	1061.2	1054.1	429.5	422.8	140.8
Ft	4.32	4.84	7.71	5.12	5.99
Día 2					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	1.08	1.09	1.09	0.93	1.05
b	0	0	0	0	0
R^2	0.98	0.99	0.97	0.98	0.98
Fc	1109.7	1459.9	136.3	407.6	239.8
Ft	4.32	4.84	7.71	5.12	5.99

Obteniéndose buenos resultados en todos los casos.

#8.8.6. COHERENCIA ENTRE LOS DÍAS PRIMERO Y SEGUNDO. VALORES MEDIOS

Las comparaciones que siguen se realizan entre el valor de los rendimientos medios calculados a partir de sus estimaciones de frecuencias y valores numéricos de los correspondientes intervalos, para los dos días de la encuesta (R_{xf1} y R_{xf2}).

En la tabla 312 pueden observarse las diferencias relativas entre dichos rendimientos medios, que son mayores en los cultivos de secano, y del 5% como máximo para la mitad de las observaciones. Son diferencias del orden de las obtenidas para la misma comparación en el año 1998 – 99.

Tabla 312.

COCIENTE $|(R_{xf1} - R_{xf2}):R_{xf1}|$. Encuesta a los agricultores, días primero y segundo (1999-00).

Cultivo	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**>0.5**	**Mediana**
Cebada secano	11	5	4				1			1		0.050
Trigo secano	6	5	1									0.050
Trigo regadío	3	1				1						0.042
Maíz regadío	7	2	1									0.036
Alfalfa regadío	6	1										0.029

En la tabla 313 puede observarse la comparación de las poblaciones de rendimientos medios, dando en general buenos resultados en todos los cultivos, de la misma forma que en el año 1998 – 99.

Tabla 313.

Comparación entre las poblaciones de rendimientos medios (R_{xf1} con R_{xf2}). Encuesta a los agricultores, días primero y segundo (1999-00).

Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.853	No	No diferencias
Trigo secano	0.972	No	No diferencias
Trigo regadío	0.600	Sí	No diferencias
Maíz regadío	0.976	No	No diferencias
Alfalfa regadío	1.00	No	No diferencias

En las tablas 314 y 315 se realizan las regresiones entre rendimientos cuando b es diferente o igual a cero, obteniéndose muy buenos resultados en ambos casos, excepto en trigo de regadío cuando b es distinta de cero. Cuando b es nula, los ajustes son muy buenos en todos los casos, y los coeficientes “a” están cercanos a la unidad.

Tabla 314.

Regresión $R_{xf1} = a\, R_{xf2} + b$. Encuesta a los agricultores, días primero y segundo (1999-00).

	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.89	0.97	0.06	0.98	0.99
b	335.34	-12.12	4063.91	-87.64	37.92
R^2	0.69	0.94	0.03	0.97	0.99
Fc	43.5	154.2	0.1	294.6	3711.9
Ft	4.35	4.96	10.13	5.32	6.61

Tabla 315.

Regresión $R_{xf1} = a\, R_{xf2}$ (b = 0). Encuesta a los agricultores, días primero y segundo (1999-00).

	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	1.00	0.97	0.95	0.98	0.99
b	0	0	0	0	0
R^2	0.98	0.99	0.99	0.99	1.00
Fc	1320.9	2240.2	313.6	5481.5	14825.5
Ft	4.32	4.84	7.71	5.12	5.99

#8.8.7. COHERENCIA ENTRE LOS DÍAS PRIMERO Y SEGUNDO. VARIANZAS

Las comparaciones que siguen se realizan entre el valor de las varianzas de los rendimientos calculados a partir de sus estimaciones de frecuencias y valores numéricos de los correspondientes intervalos, para los dos días de la encuesta (V_{f1} y V_{f2}).

En la tabla 316 se observan las diferencias relativas entre varianzas, con valores inferiores en el caso de los cultivos de regadío, tal como ocurría también en las encuestas del año 1998 – 99.

Tabla 316.

COCIENTE $|(V_{f1}-V_{f2}):V_{f1}|$. Encuesta a los agricultores, días primero y segundo (1999-00).

Cultivo	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**>0.5**	**Mediana**
Cebada secano	6	7	3	4	1		1					0.086
Trigo secano	5	1	4	2								0.100
Trigo regadío	3		1					1				0.042
Maíz regadío	5	1						1	1	1	1	0.050
Alfalfa regadío	3	1						1			2	0.075

En la comparación de poblaciones de varianzas realizada en la tabla 317, se observan buenos resultados excepto en maíz de regadío.

Tabla 317.

Comparación entre las poblaciones de varianzas (V_{f1} con V_{f2}). Encuesta a los agricultores, días primero y segundo (1999-00).

Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.925	No	No diferencias
Trigo secano	0.916	No	No diferencias
Trigo regadío	0.400	Sí	No diferencias
Maíz regadío	0.709	No	Diferencias
Alfalfa regadío	1.000	No	No diferencias

En las tablas 318 y 319 se realizan las regresiones de varianzas cuando b es igual o distinta de cero, obteniendo ajustes bastante buenos para b nula, excepto en maíz de regadío, como ya indicaban las diferencias entre poblaciones.

Tabla 318.

Regresión $V_{f1} = a\ V_{f2} + b$. Encuesta a los agricultores, días primero y segundo (1999-00).

	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.98	0.74	0.29	1.23	1.11
b	40237.48	144416.87	955436.99	1185407.31	-2225192.60
R^2	0.87	0.94	0.02	0.31	0.96
Fc	134.7	145.5	0.05	3.66	118.87
Ft	4.35	4.96	10.13	5.32	6.61

Tabla 319.

Regresión $V_{f1} = a\ V_{f2}$ (b = 0). Encuesta a los agricultores, días primero y segundo (1999-00).

	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	1.01	0.87	1.34	1.47	1.06
b	0	0	0	0	0
R^2	0.96	0.96	0.73	0.59	0.96
Fc	475.9	293.4	11.0	12.9	163.7
Ft	4.32	4.84	7.71	5.12	5.99

#8.8.8. COHERENCIA EN LAS ESTIMACIONES DE RENDIMIENTOS MEDIOS

A partir de las estimaciones de rendimientos medios realizadas por los agricultores en los dos días de la encuesta del año 1999 – 00 (R_{m1} y R_{m2}, respectivamente), se realizaron las siguientes comparaciones:

En la tabla 320 se miden las diferencias relativas entre valores medios estimados en ambos días, obteniendo unas diferencias máximas del 17.5% para el 50% de los casos.

Tabla 320.

COCIENTE $|(R_{m1} - R_{m2}):R_{m1}|$. Encuesta a los agricultores, días primero y segundo (1999-00).

	Número de casos											
Cultivo	**0-0.05**	**0.05-0.10**	**0.10-0.15**	**0.15-0.20**	**0.20-0.25**	**0.25-0.30**	**0.30-0.35**	**0.35-0.40**	**0.40-0.45**	**0.45-0.50**	**>0.5**	**Mediana**
Cebada secano	6	5	3	4	2	1		1				0.100
Trigo secano	3	2	1		1				4		1	0.150
Trigo regadío	3		2									0.042
Maíz regadío	2	3	1		1			1		1	1	0.100
Alfalfa regadío*	2	1		1						1	2	0.175

En la tabla 321 se comparan las poblaciones de rendimientos medios estimados, obteniéndose muy buenos resultados.

Tabla 321.

Comparación entre las poblaciones de rendimientos medios (R_{m1} con R_{m2}). Encuesta a los agricultores, días primero y segundo (1999-00).

Comparación			
Cultivo	**Spearman**	**Independencia**	**Wilcoxon muestras relacionadas (α=0.5)**
Cebada secano	0.847	No	No diferencias
Trigo secano	0.958	No	No diferencias
Trigo regadío	0.900	No	No diferencias
Maíz regadío	0.918	No	No diferencias
Alfalfa regadío	0.929	No	No diferencias

En la tabla 322 se realizan las regresiones entre rendimientos medios estimados en los dos días, para b igual o distinta de cero, obteniendo en general buenos resultados, sobre todo cuando b es nula, indicando coherencia de forma clara.

Tabla 322.

Regresión $R_{m1} = a\ R_{m2} + b$ y $R_{m1} = a\ R_{m2}$ (b = 0). Encuesta a los agricultores, días primero y segundo (1999-00).

R_{m1} = a R_{m2} + b					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	0.74	0.82	0.85	0.75	1.03
b	902.55	385.75	603.61	2228.9	-1426.1
R^2	0.70	0.86	0.92	0.77	0.97
Fc	47.1	60.8	35.0	26.4	165.8
Ft	4.35	4.96	10.13	5.32	6.61
R_{m1} = a R_{m2} (b = 0)					
	CEBADA SECANO	**TRIGO SECANO**	**TRIGO REGADÍO**	**MAÍZ REGADÍO**	**ALFALFA REGADÍO**
a	1.00	0.93	0.97	0.95	0.98
b	0	0	0	0	0
R^2	0.99	0.99	0.99	0.99	0.99
Fc	1405.9	832.5	2687.9	638.7	444.4
Ft	4.32	4.84	7.71	5.12	5.99