

**Departamento de Ciencias Políticas y Sociales de la
UNIVERSIDAD POMPEU FABRA**

Facultad de Ciencias Sociales y de la Comunicación

Programa de doctorado: Teoría Política y Social 1ª edición

Desarrollado en el Bienio: 1993-1995

Tesis Doctoral

**EL ESTUDIO DEL COMPORTAMIENTO ELECTORAL EN
ESPAÑA: SU UBICACIÓN DENTRO DE LA CIENCIA, SU
RELACIÓN CON LA ESTADÍSTICA Y LAS NUEVAS
POSIBILIDADES DE ANÁLISIS QUE SE OFRECEN AL
POLITÓLOGO**

Presentada por la doctorando:

Alicia CODURAS MARTÍNEZ

Para optar al título de:

DOCTOR POR LA UNIVERSIDAD POMPEU FABRA

La directora de la tesis ha sido la:

Catedrática: **Dra. M^a Rosa VIRÓS I GALTIER**

Barcelona 1998

CAPITULO V

ANÁLISIS SIMULTÁNEO DE DATOS AGREGADOS E INDIVIDUALES

1 INTRODUCCIÓN

En los capítulos anteriores hemos relacionado los datos agregados y desagregados con una serie de técnicas estadísticas, algunas más novedosas que otras en cuanto a su aparición en el tiempo, pero muy poco empleadas en general en el ámbito del Estudio del Comportamiento Electoral y muy necesarias y aprovechables dentro del mismo.

Para redondear el propósito de esta tesis no podía faltar un capítulo dedicado a lo que realmente es una de las grandes aportaciones de los últimos años a la Estadística en general y a las Ciencias Sociales por su relación con esta materia. Se trata, como indica el título de este capítulo de la posibilidad de trabajar simultáneamente con datos agregados y desagregados, un avance largamente esperado, intuido por muchos y posibilitado por el avance de la informática.

La modelización estadística ha tenido una gran participación en la explicación del Comportamiento Electoral. Un ligero repaso a los artículos publicados en los últimos cinco años en las revistas más importantes pone inmediatamente de manifiesto que el modelo de regresión y todas las variantes relacionadas con él, es el más empleado en los Estados Unidos en investigaciones relacionadas con el comportamiento político y electoral.

Así, por ejemplo, simplemente la APSR está llena de artículos relacionados con lo que se viene afirmando. Baste mencionar algunos casos en que se emplea la modelización inferencial para demostrarlo:

Beyond SES: A Resource Model of Political Participation

APSR, Nº 2, Vol 89, Junio 1995, pg. 271

En él, Brady, Verba y Leheman emplean modelos de regresión ordinarios y bietápicos para explicar la participación política.

The Responsive Voter: campaign information and the dynamics of candidate evaluation.

APSR, Nº 2, Vol 89, Junio 1995

Lodge, Steebergen y Bran aplican un modelo OL (On Line) para analizar los efectos de la campaña sobre el votante. Usan coeficientes de correlación para analizar la relación entre evaluación de candidatos y mensajes de cada tipo empleados por los partidos (Republicano y Demócrata en este caso). También emplean modelos de regresión y derivados en el mismo contexto.

The Electoral Cycle and Institutional Sources of Divided Presidential Government.

APSR, Nº 2, Vol 89, Junio 1995

M. Soberg Shupart emplea modelos de regresión y series temporales para analizar elecciones presidenciales en distintos ámbitos geográficos.

Shaping Mexico's Electoral Arena: the construction of partisan cleavages in the 1988 and 1991 national elections.

APSR, Nº 2, Vol 89, Mayo 1995

Dominguez y McCann explican el comportamiento electoral mejicano en esas dos elecciones. Se basan en encuestas de opinión públicas nacionales y utilizan un modelo logit para investigar los factores que han influido en que los electores votasen a los partidos de la oposición.

Party cleavages and welfare effort in the American States

APSR, Vol 89, Nº 1, Mayo 1995

Robert D. Brown utiliza modelos de regresión cross-seccionales, heteroscedásticos y autorregresivos propuestos por Kmenta en 1986 que sirven para eliminar la heteroscedasticidad y la autocorrelación.

The Concept of a Critical Realignment, Electoral Behavior and Political Change

APSR, Vol 89, Nº 1, Mayo 1995

Peter F. Nardulli usa el análisis de series temporales y mapas comparativos para ver la evolución del voto republicano contrastado con el demócrata.

El número de ejemplos sería interminable, pero el denominador común de todos ellos es el empleo de técnicas de regresión clásicas y reconocidas.

Hay estudios con datos agregados y con datos de encuesta individual, pero, evidentemente, lo que tanto tiempo se ha deseado: el poder estudiar simultáneamente ambos tipos de datos es actualmente una realidad gracias a los modelos llamados en general Multilevel y que comenzaron a experimentarse a principios de los noventa, existiendo en estos momentos aplicaciones estadísticas que están conduciendo a un nuevo desarrollo de la Estadística.

El tema que se va a tratar en esta parte de la Tesis es el de las posibilidades que representa la modelización multilevel en la parcela de estudio de Comportamiento y Resultados electorales, la forma en que afecta su aparición al planteamiento de una investigación de esta esfera y el potencial que representa para el futuro inmediato dentro de la disciplina.

Para situarnos en el contexto de los modelos multilevel se presenta en primer lugar un apartado en que se explican brevemente los antecedentes de estas técnicas para, posteriormente, comentar los principales rasgos del modelo más simple y, seguidamente, tratar acerca de las posibles relaciones que se pueden establecer entre estos modelos y el estudio del Comportamiento y los Resultados Electorales.

1.1 ANTECEDENTES

A finales de los años setenta y, a lo largo de los ochenta, autores como Crombach, Webb y Burstein, apuntaron la importancia de considerar el estudio de las relaciones entre datos recogidos a distintos niveles de una estructura jerárquica.

Hay muchas formas de establecer una estructura jerárquica, pero en el caso de un investigador de comportamiento electoral, no hay duda de que es posible plantear algunas como las siguientes:

Nivel 1 (inferior) = votante o elector
Nivel 2 (superior) = municipio en el que vota

Nivel 1 (inferior) = municipio donde se producen los resultados
Nivel 2 (superior) = comarca a que pertenece el municipio

Nivel 1 (inferior) = elector
Nivel 2 (medio) = sección censal
Nivel 3 (superior) = distrito

Y muchas otras de las que se tratará más adelante.

Sin embargo, las primeras experiencias que se llevaron a cabo en relación con los modelos multilevel, pertenecen al terreno del sistema educativo, debido a que éste presentaba una estructura jerárquica natural fácilmente aprovechable.

En un principio, la pregunta clave fue: ¿de qué forma se relacionan las características individuales del estudiante con la clase a la que pertenece y con los factores que definen el colegio al que asiste?

Evidentemente, desde la óptica del investigador político, la pregunta puede traducirse en, por ejemplo: ¿de qué forma se relacionan las características del elector con la sección en que vota y con el distrito que la contiene?

Durante mucho tiempo, este tipo de preguntas se ha ido resolviendo mediante la aplicación de regresiones múltiples, trabajando por separado con datos agregados y desagregados. Se han establecido perfiles mediante otras técnicas multivariantes: cluster, componentes principales, discriminante, etc. Sin embargo, un reciente desarrollo teórico unido al avance informático ha permitido una rápida experimentación de nuevos procedimientos que, poco a poco, sustituirán a los tradicionales.

Aitkin y su equipo llevaron a cabo un estudio acerca de las formas de enseñanza, que supuso un primer paso en la consecución de modelos más adecuados para el análisis de este tipo de problemas. Sus logros dieron lugar a la aparición de extensiones y ramificaciones que permitían analizar estructuras jerárquicas de distintos tipos, como por ejemplo: las longitudinales y las multivariantes. Esto último, desde el punto de vista de los temas de estudio aquí tratado, significa poder trabajar con estructuras como las siguientes:

Modelo longitudinal

Nivel 1 = voto en distintas elecciones
Nivel 2 = elecciones
Nivel 3 = elector

Modelo multivariante

Nivel 1 = elecciones

Nivel 2 = electores

Variables dependientes diversas

Los modelos multilevel comenzaron a emplearse con cierta regularidad a partir de 1987, asociados sobre todo, a la técnica de la regresión.

Esta familia de modelos puede utilizarse desde muchos puntos de vista y tiene un futuro tan prometedor que ha sido calificada por algunos como "The next great unifying step investigación Statistics" porque aglutina a todas las modelizaciones.

Sin embargo, por el momento y, pese a los rápidos avances que se están llevando a cabo en este terreno, su difusión a nivel práctico en España es difícil porque lo que se publica proveniente del extranjero (especialmente Gran Bretaña y los Estados Unidos) casi siempre tiene un carácter muy técnico. Para que los investigadores sociales conozcan el desarrollo y posibilidades de esta nueva parcela, tiene que existir un intermediario que efectúe una introducción a la misma y, por el momento, las aplicaciones estadísticas existentes son particulares y no están contenidas en los grandes paquetes estadísticos.

Una de las características destacables de esta modelización es que el concepto de varianza y también el de covarianza ocupan el primer plano de la atención. En referencia a este hecho, se puede ver una evolución de la metodología estadística que primero puso el acento en el empleo de la media, posteriormente en el de la mediana y ahora establece el poder de la varianza como principal elemento configurador del análisis.

2 LOS MODELOS MULTILEVEL Y EL ANÁLISIS DEL COMPORTAMIENTO ELECTORAL

2.1 INTRODUCCIÓN

Aunque la literatura acerca del análisis multilevel se ha centrado especialmente en temas relacionados con la educación y el rendimiento académico y, recientemente se ha infiltrado en los temas relacionados con la sanidad, también ha aparecido alguna vez algún artículo que explora el empleo de la modelización multilevel en el análisis de datos electorales.

Concretamente, Kelvin Jones, R.B. Johnston y C.J. Pattie²⁰², presentaron el artículo "People, places and regions: exploring the use of multilevel modeling investigación the analysis of electoral data en B.J. Pol, S,22, 1992.

El marco de trabajo en que se basaron era el siguiente: a fines de los ochenta hubo un considerable debate, acerca de la importancia del contexto local como una influencia sobre las actitudes políticas y el voto en Gran Bretaña. Las resoluciones de este debate fueron difíciles de exponer porque los métodos analíticos no eran válidos para evaluar la importancia relativa, tanto de las características individuales del votante como de las características de su ambiente (entorno) como correlacionadores independientes de actitudes y comportamiento. La técnica de la

²⁰² Departamentos de Geografía del Politécnico de Portsmouth, de la Universidad de Sheffield y de la Universidad de Nottingham respectivamente.

modelización multilevel ha sido desarrollada por investigadores en el terreno educacional justamente para lograr esto.

En el artículo mencionado, se introduce este análisis a partir de datos de las elecciones generales británicas de 1987 y los resultados preliminares sugieren que el lugar de residencia, claramente actúa como un componente del proceso que influye en las elecciones de los votantes.

Los autores sabían que el análisis del Comportamiento Electoral en Gran Bretaña había estado buscando, como en todas partes, explicaciones para las variaciones del voto en una serie de niveles que abarcaban desde el mero individuo hasta una región.

Ahora, casi todo el mundo reconoce que la comprensión de los patrones de voto requiere análisis sofisticados que tengan conjuntamente presentes estos niveles y que además, usen simultáneamente datos agregados y desagregados. Pocos han reconocido las dificultades metodológicas relacionadas con los análisis y procedimientos más comúnmente utilizados como, por ejemplo, el problema de la autocorrelación espacial, que les lleva a imponer en sus análisis una estructura que no es necesariamente consistente con la modelización intuitiva del proceso que se desea describir.

Los recientes desarrollos en la modelización estadística del fenómeno multilevel proporcionan una oportunidad de evaluar algunos de los modelos intuitivos de forma más exacta que antes.

Por tanto, para introducirse en el mundo de la explicación del voto utilizando la modelización multilevel, vamos a tratar de presentar los conceptos más sencillos de forma que se entienda la idea principal que luego puede ser desarrollada hasta el infinito.

Conceptos básicos que relacionan el modelo multilevel con el estudio del Comportamiento Electoral

En principio, no hay duda de que una de las unidades fundamentales en cualquier estudio electoral es el elector individual. Los electores individuales no han sido políticamente socializados a partir del vacío. Aprenden sus actitudes políticas en una variedad de contextos.

Según Rose y McAllister, el más importante de estos contextos es el entorno familiar: características de la familia en que crecen, actitudes políticas de los padres y sus propias predisposiciones al voto. Como en los estudios de comportamiento electoral se observan individuos en lugar de familias pero, a estos individuos se les puede preguntar sobre el entorno familiar del que proceden, estas características pueden incorporarse sin problemas a los modelos operacionales como aspectos del individuo.

Fuera de la familia, la socialización política está influenciada por otras características tales como su posición económica y social, el sector en el que trabajan (especialmente público o privado)²⁰³ si están o no afiliados a Sindicatos y el tipo de propiedad de su vivienda²⁰⁴.

²⁰³ P. Dunleavy, *The Political Implications of Sectorial Cleavages and The Growth of State Employment*, *Political Studies*, 28, 1980.

²⁰⁴ Para más información sobre esto, se recomienda la consulta de R.J. Johnston, *A Note on Housing Tenure and Voting* investigación Britain, 1983, *Housing studies*, 2, 1987.

La afiliación a Sindicatos, la posición que ocupa el individuo en la estructura de clase ocupacional, el tipo de vivienda, la clase social y el sector en que está empleado, son factores considerados por los británicos como medios de socialización, aunque reconocen que algunos tienen más poder que otros.

Sin embargo, de nuevo, debido a que la mayoría de los estudios sobre voto se basan en datos recogidos a partir de individuos seleccionados aleatoriamente, todas estas características son tratadas como propias del individuo. Así, por ejemplo, la probabilidad de votar Comunista se considera que no sólo varía por la posición laboral del individuo sino por si viene de una familia con disposición a votar a esa fuerza política o no, vive en una casa de alquiler, trabaja en las capas bajas del sector público y pertenece a un Sindicato. Juntos, estos factores comprenden las principales influencias de nivel individual en la formación de la actitud política y en el comportamiento de voto identificado (de momento dejamos aparte el tipo de elección, municipal, autonómica o general en que además de los factores mencionados hay otros que entran en juego).

Gran parte de la socialización política se da en niveles particulares y, recientes escritos acerca del comportamiento electoral británico, efectuado desde el punto de vista geográfico, han enfatizado su aparente importancia. Intentos de establecer patrones de voto usando solamente las características individuales han fallado durante mucho tiempo, ya que la gente con características similares es más propensa a votar al Partido Laborista en algunas zonas que en otras donde vive el mismo tipo de gente.

Por tanto, hay que poner más atención a los contextos espaciales dentro de los cuales tiene lugar la socialización y los cuales forman la arena de la modelización política.

Para comenzar está bien centrarse en dos niveles particulares: el elector y el distrito o la sección censal en el que reside, por ejemplo. Gran parte de la primera socialización de los votantes tiene lugar, no sólo en su hogar, sino en su vecindario, en el cual, tanto instituciones formales (escuelas, edificios públicos, etc.) como elementos informales (amigos, vecinos, etc.) actúan como agentes socializantes.

A través de todo ello, la gente observa un amplio abanico de actitudes, las cuales predisponen, aunque no determinen necesariamente, tanto su interpretación de la estructura de la sociedad, como su posición dentro de ella. Este aprendizaje influye no sólo sus futuras carreras sino las imágenes de la sociedad en que habitan. Una gran consecuencia de estas influencias, directa e indirectamente, es su impacto donde la gente vive, tanto si han sido ellos los que escogido vivir allí como si se han visto obligados a ello por las circunstancias.

En una sociedad que está fuertemente segmentada, especialmente por la clase ocupacional, la mayoría de las personas viven en vecindarios ocupados por otras similares a ellos. En este sentido, el lugar de residencia es un continuo contexto de refuerzo personal para la socialización política.

En el proceso de socialización hay dos lugares especialmente importantes: el lugar en que un individuo ha crecido, especialmente durante su adolescencia que es cuando las actitudes políticas tienen un primer desarrollo y, el lugar donde vive habitualmente el individuo. También hay otros que pueden ser importantes, como el lugar donde se llevan a cabo los estudios medios y superiores (los que los hacen) y, posiblemente hayan llevado a cabo su primera acción de votar en unas elecciones generales, pero estos lugares son más difíciles de incorporar en modelos del proceso de socialización.

Desgraciadamente, pocos conjuntos de datos contienen suficiente información acerca del primero de estos lugares²⁰⁵. Sin embargo, cada vez es más posible mezclar datos de individuos con información sobre el barrio en que viven habitualmente, obtenidos a través del Padrón o fuentes censales por el estilo.

A pesar de que algunos estudios del llamado efecto vecindario usan meramente variables nominales u ordinales (clase social, clase profesional, etc.), otros consideran la relación entre comportamiento del votante y el lugar de residencia como un continuo. Así, por ejemplo, puede hacerse la hipótesis de que la probabilidad de que un individuo de la clase trabajadora vote comunista aumenta tanto como la proporción de personas de clase trabajadora que viven en ese vecindario.

¿Pero cuál es la medición del vecindario? ¿Qué variables o escalas se pueden utilizar? ¿Qué define el lugar en el cual vive y se socializa una persona? La respuesta es que, en primer lugar, hay que tener presente que para un individuo dado no existe claramente un solo lugar, sino que hay varios. Los modelos completamente especificados pueden llegar a incorporarlos a todos, pero ello podría introducir problemas analíticos sustanciales a causa de la multicolinealidad. Un modelo completamente especificado es casi imposible de construir porque tiene muchas restricciones respecto a los datos. Los datos censales son válidos en muchos tipos de escalas de medición y los desarrollos en el uso de Sistemas de Información Geográfica, cada vez permiten un mayor acceso a los mismos.

Hasta la actualidad, la mayor parte de los análisis se han efectuado utilizando datos de un solo nivel, el individuo seleccionado aleatoriamente en un distrito electoral. Este hecho tiene consecuencias en los análisis estadísticos. Además, ya se ha demostrado que el ligamen entre clases y partidos ya no es tan claro como hace años y que en las décadas recientes el electorado es cada vez más versátil.

El distrito electoral es el lugar o contexto local dentro del cual los individuos viven su vida diaria. Este lugar está a su vez relacionado con otros que también tienen un impacto en los individuos pero que pertenecen a un nivel superior: municipio, comarca, región o país. Pero aún considerando todas estas relaciones, lo cierto es que existe una tendencia a estrategias de voto denominadas egocéntricas y sociotrópicas.

Las estrategias egocéntricas se relacionan con la evaluación de la actuación del gobierno en el pasado y en el futuro potencial. Desde una óptica muy personal, muy del elector desde sí mismo. El elector se hace a sí mismo un informe que relaciona, especialmente con sus condiciones de trabajo y su situación personal y a partir de él toma una decisión.

Las estrategias sociotrópicas están relacionadas con las decisiones de los votantes basadas en evaluaciones de las actuaciones pasadas y del futuro próximo, del gobierno pero como reflejo de la opinión de la sociedad en conjunto.

Por tanto, si en general se considera que se dan estas dos estrategias de voto, ello supone dificultades para la modelización estadística, porque hay que poder determinar aspectos como los siguientes: ¿qué definiciones territoriales hay que usar? ¿cuáles son relevantes en el sentido de que los electores sigan una estrategia sociotrópica?

²⁰⁵ Una excepción se halla en la publicación de G.C. Wright Jr. Contextual Models of Electoral Behavior: The southern Wallace vote, APSR, 71, 1977, pg 497-508.

En relación con este tema, los investigadores británicos (Johnston, Pattie, Jones, etc.) proponen la consideración de dos tipos de regiones: regiones geográficas y regiones funcionales. Las primeras son las divisiones territoriales normales administrativas de un país o zona que pueden o no suponer un cierto grado de identificación personal de sus habitantes. Las segundas son áreas donde se dan unas características predefinidas, independientemente de su localización geográfica.

Ambos tipos de regiones pueden ser relevantes en relación con las decisiones de voto y hay que tratar de incorporar las dos posibles visiones a los análisis. La forma de hacerlo es a través de variables nominales que indiquen, por ejemplo, que la probabilidad de votar a la fuerza en el poder es superior (1) en ciertos distritos que en otros (0). También es posible tomar algún indicador económico (inversión en obras públicas en cada distrito, por ej.) e introducirlo como variable continua.

La gente de un municipio, por ejemplo, puede captar de la mayoría de los residentes quejas respecto al gobierno, o elogios o cualquier otro tipo de información que, en principio podría influir en su voto pero, por otro lado, es posible que su situación personal no concuerde con lo que dice la gente en general y que decida lo contrario en defensa de sus intereses. Lo mismo puede suceder a nivel de distrito: en unos se han hecho unas mejoras considerables, el gobierno municipal está bien considerado y en otros sigue habiendo muchos cambios y reformas por efectuar, de manera que no han percibido la bondad de ese gobierno, votando en su contra.

En definitiva, en el estudio del voto en España hay diversos niveles de interés: el individuo, el lugar donde reside, la ciudad, la Comunidad Autónoma y el País, e incluso hay quien añadiría la Comarca.

Sin embargo, para introducirse en el entorno de este tipo de modelización, ya hemos explicado que lo mejor es comenzar por considerar únicamente dos niveles y que el único requisito para su aplicación es determinar que la población objeto de estudio responde a algún tipo de estructura jerárquica por sencilla que sea

Desde el punto de vista del entorno Multilevel, lo que importa es cómo afecta esta estructura a las mediciones que nos interesa evaluar. Así, si estamos midiendo resultados educacionales, ya sabemos que la nota media varía de una escuela a otra. Esto significa que los niños dentro de una escuela pueden parecerse más entre ellos en promedio, que los estudiantes de diferentes escuelas. Asimismo, si estamos midiendo resultados electorales, sabemos que la media de votos a una fuerza política puede variar de un distrito a otro, de forma que los electores dentro de un distrito pueden parecerse más entre ellos en promedio que los electores de diferentes distritos.

2.2 LA APLICACIÓN INFORMÁTICA ML3

El programa ML3 fue la primera aplicación que surgió a nivel comercial tras otras versiones experimentales dentro del ámbito de la modelización multilevel. Con este programa se podían plantear modelos de hasta tres niveles y, actualmente está disponible el programa MLn que permite analizar estructuras de "n" niveles y que incluso ya se puede hallar en una versión preparada para trabajar con el sistema operativo Windows.

Sin embargo, estas aplicaciones informáticas todavía no son muy conocidas en nuestro país y, aunque existen muchos investigadores aplicando técnicas multilevel,

su difusión y conocimiento deben alcanzar cotas más altas para lograr su popularización.

El programa ML3 es de procedencia británica, pero otros grupos y autores de los Estados Unidos han trabajado paralelamente a los británicos ofreciendo sus propias herramientas informáticas²⁰⁸.

En esta tesis se ha escogido esta aplicación por ser la introductoria en el tema y por ser la que la doctorando tuvo oportunidad de conocer y trabajar con el Dr. Kelvin Jones en la Universidad de Essex.

Se trata de una versión que todavía funciona con el sistema operativo MSDOS, lo cual hace que sus resultados sean, a nivel de presentación, un tanto rudimentarios. Sin embargo, es perfectamente válido para introducirse en el tema y posteriormente seguir con otras versiones más actualizadas.

2.3 LA ESTRUCTURA DE UNA BASE DE DATOS QUE SIRVA PARA EFECTUAR UN ANÁLISIS MULTILEVEL

Para poder aplicar un análisis de este tipo, se necesita construir una base de datos diferente de las que estamos acostumbrados a tratar. De momento, como no se ha efectuado ninguna encuesta real pensada para este tipo de modelización en España, usaremos, en primer lugar, dos bases de datos que convertiremos en una sola que sirva para nuestros propósitos y también otra base de datos que permita otras experiencias cuando la primera proporcione resultados poco satisfactorios.

Por tanto, comentando la creación de la primera base, se parte de dos: una de datos de encuesta y otra de datos agregados construyéndose una sola en que la clave de unión es el código de distrito y el de la sección censal.

Entre las variables de la base de datos individuales destacan las siguientes, para nuestros propósitos:

Grado de satisfacción de vivir en el municipio
¿Le agradaría que la gestión del actual alcalde siguiese cuatro años más?
Valoración de la gestión del equipo de gobierno municipal actual
¿A qué fuerza política votó en las pasadas elecciones municipales?
¿A qué fuerza política votaría si se celebrasen ahora?
¿A qué fuerza política piensa votar en las próximas autonómicas?
Género del entrevistado
Edad del entrevistado
Profesión del entrevistado
Idioma empleado en la entrevista
Distrito
Sección censal

Entre las variables de la base de datos agregados destacan las siguientes como interesantes para la prueba que vamos a efectuar:

Número de hombres de la sección
Número de mujeres de la sección
Número total de electores de la sección
Número de personas de 15 a 19 años de edad

²⁰⁸ Por ejemplo, existe el programa VARCL debido al profesor Nicholas Longford y su equipo.

Número de personas de 20 a 24 años de edad
 Número de personas de 25 a 29 años de edad
 Número de personas de 30 a 34 años de edad
 Número de personas de 35 a 39 años de edad
 Número de personas de 40 a 44 años de edad
 Número de personas de 45 a 49 años de edad
 Número de personas de 50 a 54 años de edad
 Número de personas de 55 a 59 años de edad
 Número de personas de 60 a 64 años de edad
 Número de personas nacidas en Barcelona
 Número de personas nacidas en Girona
 Número de personas nacidas en Lleida
 Número de personas nacidas en Tarragona
 Número de personas nacidas en cada una de las Comunidades Autónomas por separado
 Número de personas nacidas en el resto del estado
 Número de extranjeros
 Número de personas que no saben leer
 Número de personas sin estudios
 Número de personas que han cursado EGB 1A
 Número de personas que han cursado EGB 2A
 Número de personas que han cursado FP 1
 Número de personas que han cursado FP 2
 Número de personas que han cursado BUP/COU
 Número de personas con título medio
 Número de personas con título superior
 Número de personas que han realizado el servicio militar
 Número de ocupados
 Número de parados que buscan el primer trabajo
 Número de jubilados
 Número de incapacitados
 Número de estudiantes
 Número de amas de casa
 Número de personas que entienden el catalán
 Número de personas que saben hablar el catalán
 Número de personas que saben leer el catalán
 Número de personas que saben escribir el catalán
 Número de personas que no entienden el catalán

Concretamente, la base de datos para trabajar con modelos multilevel debe cumplir con una serie de requisitos. Para mostrar un pequeño ejemplo de lo que representan estos requisitos, hemos tomado algunos casos de la muestra de 300 y unas pocas variables (agregadas y desgregadas) del total de que disponemos y así, podremos ver cuál es la disposición de los datos en una matriz que conjuntamente contiene información individual y agregada.

La siguiente tabla muestra que hemos introducido diez de las variables que deseamos emplear en las primeras demostraciones prácticas más el código identificador de cada elector en su sección censal y la constante que necesita el programa para efectuar las regresiones y que es una columna de unos llamada CONS. De izquierda a derecha tenemos el número y nombre de cada variable. La columna "n" indica el número de casos y la columna missing el de casos de no respuesta. Seguidamente, tenemos dos columnas en que se muestran los valores mínimo y máximo de cada variable. En el caso de la variable ELECTOR, el máximo que aparece es 66 en lugar de 300, porque ese es el número máximo de electores que hay en una sección. En una base multilevel, cada nivel debe tener una lista de

los individuos que vuelva a comenzar por 1 sean o no iguales los tamaños de los grupos que forman el nivel.

NAMES					
	Name	n	missing	min	max
1	ELECTOR	300	0	1.0000	66.000
2	CONS	300	0	1.0000	1.0000
3	DISTRIC	300	0	1.0000	2.0000
4	SECC	300	0	1.0000	7.0000
5	SATISF	300	0	2.0000	9.0000
6	VALORA	300	0	1.0000	9.0000
7	VOTOS95	300	0	753.00	1201.0
8	CIUA95	300	0	371.00	656.00
9	PSCA95	300	0	142.00	244.00
10	ICA95	300	0	34.000	77.000
11	ERCA95	300	0	64.000	184.00
12	CONTINU	300	0	1.0000	3.0000

Seguidamente, para tener una idea de cómo es la base de datos, hemos seleccionado los primeros casos para algunas variables y una parte de los que comienzan la segunda sección censal para que se vea cómo vuelve a comenzar por uno la numeración:

PRINT C1-C7					
N =	ELECTOR	CONS	DISTRIC	SECC	SATISF
1	1.0000	1.0000	1.0000	1.0000	7.0000
2	2.0000	1.0000	1.0000	1.0000	7.0000
3	3.0000	1.0000	1.0000	1.0000	5.0000
4	4.0000	1.0000	1.0000	1.0000	9.0000
5	5.0000	1.0000	1.0000	1.0000	9.0000
6	6.0000	1.0000	1.0000	1.0000	7.0000
7	7.0000	1.0000	1.0000	1.0000	7.0000
8	8.0000	1.0000	1.0000	1.0000	9.0000
9	9.0000	1.0000	1.0000	1.0000	7.0000
10	10.000	1.0000	1.0000	1.0000	5.0000
11	11.000	1.0000	1.0000	1.0000	9.0000
etc etc.					
67	1.0000	1.0000	1.0000	2.0000	8.0000
68	2.0000	1.0000	1.0000	2.0000	9.0000
69	3.0000	1.0000	1.0000	2.0000	6.0000
70	4.0000	1.0000	1.0000	2.0000	9.0000
71	5.0000	1.0000	1.0000	2.0000	7.0000
72	6.0000	1.0000	1.0000	2.0000	7.0000
73	7.0000	1.0000	1.0000	2.0000	9.0000
74	8.0000	1.0000	1.0000	2.0000	7.0000
75	9.0000	1.0000	1.0000	2.0000	8.0000
76	10.000	1.0000	1.0000	2.0000	9.0000
77	11.000	1.0000	1.0000	2.0000	9.0000
78	12.000	1.0000	1.0000	2.0000	9.0000
79	13.000	1.0000	1.0000	2.0000	8.0000
80	14.000	1.0000	1.0000	2.0000	7.0000

Tal y como puede apreciarse en la tabla anterior, el primer requisito que debe cumplir la base es estar ordenada. En este caso, la ordenación se lleva a cabo mediante las variables distrito y sección censal. En el cuadro se puede ver, de izquierda a derecha, el número de caso 1 que pertenece al distrito 1 y a la sección

censal 1 de dicho distrito, que ha puntuado con un 7 la satisfacción de vivir en el municipio (variable desagregada). Al lado de esta variable aparecerían otras como la variable agregada VOTOS95, que contiene el número total de votos emitidos en las autonómicas del 95 en ese distrito y sección censal. A medida que la base de datos avanzase, esta cifra cambiaría con la de sección censal.

Por tanto, en una base de datos de estas características, los datos correspondientes a las variables agregadas se repiten tantas veces como sea necesario para todos los entrevistados de un distrito y sección censal. Cuando cambia la sección censal, varían los datos y así sucesivamente. Ahora bien, repetimos que el orden de los individuos acerca de esas dos variables ha de ser perfecto.

Una vez presentada la idea del formato que debe tener una matriz a la que se pueda aplicar un análisis multilevel, se puede comenzar a presentar el desarrollo de este tipo de modelización por medio de aplicaciones concretas, sistema que lo hará más agradable y comprensible.

2.4 EL DESARROLLO DE UN MODELO MULTILEVEL A TRAVÉS DE UN CASO SIMPLE

Como en cualquier investigación, deberíamos de haber definido el objetivo de la misma, los datos a recopilar, la metodología de dicha recopilación y realizado todos los pasos previos de que se ha hablado en los primeros capítulos. Sin embargo, tendremos que dar por supuesto todo ello y partir de una base de datos real que hemos tenido que confeccionar a partir de otras dos que en su día ya sirvieron para efectuar determinados estudios electorales.

El centro de atención de un análisis multilevel como el que vamos a llevar a cabo no sería el conjunto de secciones censales de nuestra muestra, sino estimar el patrón de variación que subyace en la población de secciones censales de un territorio determinado, pero para llevar a cabo el experimento de forma realista necesitaríamos muchos más datos. Una vez realizada la experiencia, entraríamos en un estadio científico, si todos los pasos se hubiesen dado correctamente y, podríamos intentar explicar el patrón de comportamiento en términos de las características generales de la población objetivo de secciones censales, incorporando para ello, más variables en el modelo. Este podría ser un planteamiento correcto de aplicación si ésta se planificase desde el principio.

Pero, a nuestro nivel, podemos seguir presentando la técnica y sus utilidades, en la creencia de que pronto se podrán desarrollar experiencias bien diseñadas en nuestro país, para un amplio y bien construido conjunto de datos.

Las consecuencias de ignorar una estructura jerárquica en diversos niveles, cuando ésta existe, pueden conducir a la infravaloración de determinadas variables como factores explicativos de un modelo.

En nuestro ejemplo, se espera analizar la relación entre la valoración de un equipo de gobierno municipal y la satisfacción de residir en un municipio que tienen sus electores. Para ello, contamos con 300 datos de electores en 7 secciones censales. Las técnicas que se han venido aplicando habitualmente, nos habrían obligado a tomar una decisión: llevar a cabo el análisis a nivel de sección censal (agregado) o a nivel de elector (nivel desagregado). Con el análisis multilevel se pretende demostrar que ambas opciones son insatisfactorias y que, aunque finalmente llegásemos a la conclusión de que la estructura jerárquica (sección censal, elector) no existe en realidad, al menos lo habríamos probado.

A nivel agregado, tendríamos que haber calculado la media de satisfacción y de valoración para cada sección y la regresión ordinaria se habría usado para estimar la relación entre las medias de la satisfacción y de la valoración. El principal problema que presenta este enfoque es que resulta muy difícil interpretar cualquier relación que se halle ya que, cualquier interpretación causal debería incorporar a los electores individuales y su información ha quedado eliminada al agregar los datos.

En la práctica es posible hallar una gran variedad de modelos que ajusten bien los datos, pero pueden proporcionar estimaciones muy diferentes entre sí y la interpretación de los resultados de tales análisis resulta muy difícil porque depende de una elección esencialmente arbitraria del modelo.

Efectuar el análisis en el nivel de los electores implicaría obtener una estimación de la media de la relación de las dos variables: satisfacción y valoración para los 300 casos de que se dispone. La variación entre secciones censales podría haberse modelizado incorporando términos separados para cada una de ellas. Este procedimiento es ineficiente e inadecuado cuando el propósito último es llegar a una generalización científica. Es ineficaz porque en el proceso se estiman muchas veces más coeficientes que usando la técnica multilevel y porque no reconoce que las secciones censales serían una muestra aleatoria que no proporciona información acerca de la variación subyacente en la población total de secciones censales.

El análisis multilevel centra su atención en los niveles jerárquicos de la población, permitiendo al investigador comprender qué efectos se están produciendo y en qué forma lo hacen. Además, proporciona mejores estimaciones en respuesta a las preguntas que pueden hacerse tanto a nivel simple como para estructuras de más de un nivel.

Finalmente, comentar que, llevar a cabo un análisis sin tener presente una estructura jerárquica existente en la población objetivo puede dar lugar a serios problemas. Por ejemplo, ante una estructura jerárquica no considerada, podemos provocar que los errores estándar de los coeficientes de la regresión queden infravalorados (su estimación tenga un valor inferior al que debería).

Por ejemplo, si consideramos que los electores están agrupados en secciones censales y éstas en distritos electorales, si los errores estándar son inferiores a lo que deberían ser, podría inferirse que existía una preferencia real por un candidato, o por una acción en lugar de otra, cuando de hecho, tal preferencia, estimada a nivel muestral puede deberse al azar. Los errores estándar correctos sólo pueden estimarse si la variación a nivel de sección y de distrito fuese considerada dentro del análisis. Por eso, la modelización multilevel proporciona una forma eficiente de lograr este propósito permitiendo al modelo indagar en la influencia de la sección censal y del distrito sobre el comportamiento electoral, así como trabajar características individuales como la clase social.

Por tanto, teniendo presentes nuestras limitaciones para el caso que vamos a desarrollar, ahora, a la vista de las variables de que disponemos, debemos diseñar un experimento multilevel que tenga cierta lógica y que sea coherente con el material de que disponemos. Por supuesto, si la investigación se iniciase en este momento, la forma de proceder debería ser diferente.

2.5 MODELO "VARIANCE COMPONENTS MODEL": PRIMERA EXPERIENCIA

Para introducir la parte práctica vamos a considerar un modelo de dos niveles:

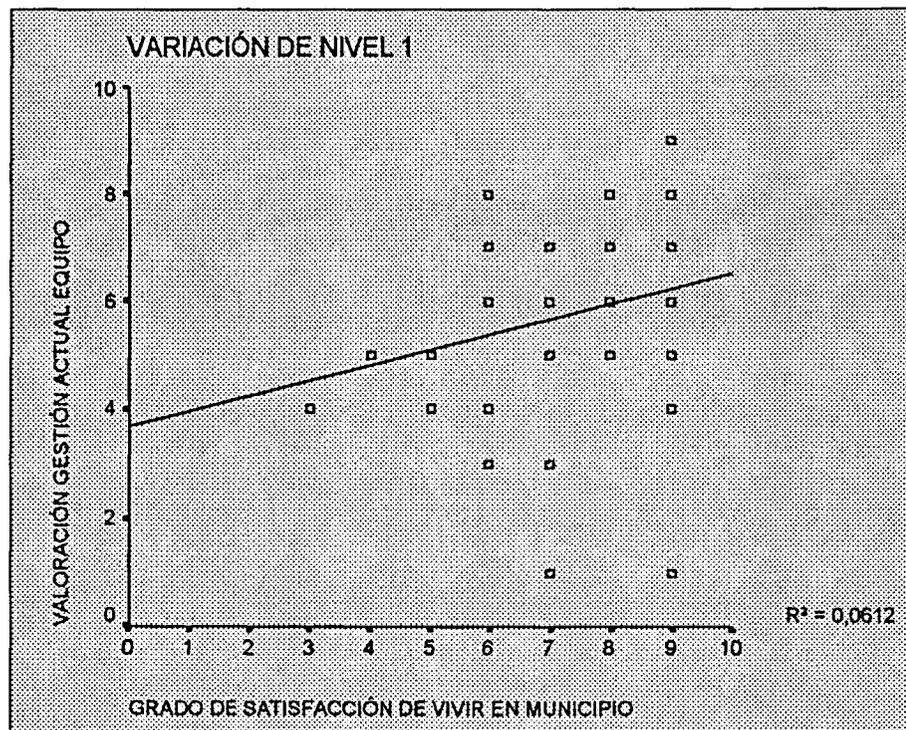
Nivel 1: electores

Nivel 2: secciones censales

Vamos a suponer que el grado de satisfacción de vivir en el municipio puede tener alguna capacidad para explicar la valoración que efectúan los electores del actual equipo de gobierno al frente del Ayuntamiento. La técnica más elemental para llevar a cabo una primera aproximación es la de regresión simple calculada mediante mínimos cuadrados ordinarios.

En una investigación bien planificada, el interés debe centrarse en usar la variación entre las secciones censales de la muestra para efectuar inferencias acerca de la variación de la población objetivo subyacente.

En nuestro caso, podemos comenzar por ver qué sucede al respecto en una de las secciones censales, por ejemplo, la primera:



Cada uno de los puntos de este gráfico representa un par de valores de un elector individual. El grado de dispersión de estos puntos en torno a la línea ajustada representa la variación aleatoria al nivel del elector individual acerca de la posible relación entre su satisfacción de vivir en el municipio y su valoración de equipo de gobierno municipal. En el ámbito de estos modelos, a esta variación se la conoce como nivel 1 de variación (level 1 variation) porque se refiere a la variación que hay entre las unidades del nivel 1, es decir, los electores.

La técnica de la regresión simple aquí empleada proporciona elementos sobradamente conocidos:

$$y_i = a + b x_i + e_i \quad (1.1)$$

$$y_i^* = a + b x_i \quad (1.2)$$

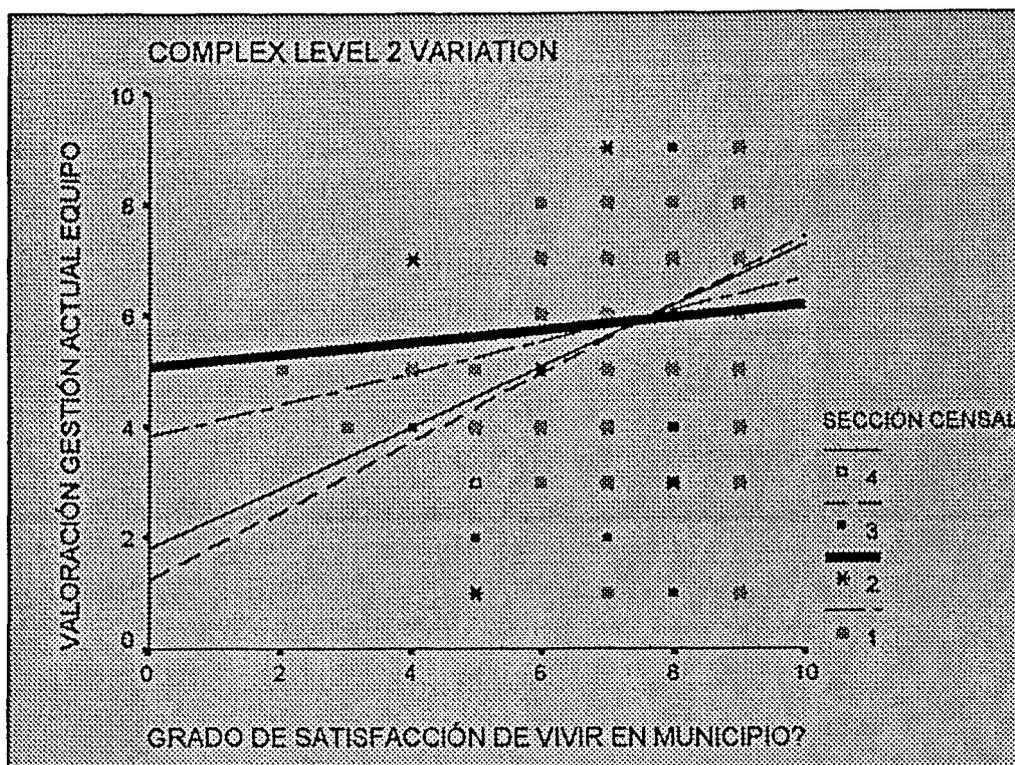
La ecuación (1.1) representa la relación de las variables valoración y satisfacción para el elector i , siendo e_i el residuo o parte de la valoración no explicada por la recta de regresión. En el nivel 1, cuando consideramos los electores de una sola sección censal, la variación es precisamente la varianza de estos residuos.

La ecuación (1.2) representa la recta de regresión ajustada a los datos, siendo a el punto de corte y b la pendiente.

Esta recta ajustada se puede interpretar como el resumen a nivel de una sección censal de las relaciones entre las dos variables implicadas (satisfacción y valoración) de los electores de la sección. Este resumen puede variar de sección a sección y, ello se podría dar de dos formas diferentes:

Que se diese una variación simple en el nivel 2 (secciones censales), de forma que fuese constante entre ellas y no dependiese de las puntuaciones de satisfacción de los electores, es decir, que para una valoración del equipo de gobierno siempre existiese una única correspondencia del valor de la satisfacción de vivir en el municipio. Este tipo de variación, recibe el nombre de Simple level 2 variation en el ámbito de los modelos multilevel, pero no es muy realista esperar que se de en este caso. Su representación gráfica daría lugar a un conjunto de rectas de regresión, una para cada sección censal con distintos puntos de corte e igual pendiente. Este caso es el que corresponde al llamado modelo Variance Components con pendientes constantes y puntos de corte diferentes para todas las secciones y es el primero que vamos a probar.

Que se diese una variación compleja en el nivel 2, de forma que no fuese constante entre ellas y que dependiese de las puntuaciones de satisfacción y de valoración del equipo por parte de los entrevistados. En el siguiente gráfico, se muestran algunas regresiones en que tanto los puntos de corte como las pendientes varían: es un ejemplo de esta variación, llamada "Complex level-2 variation", o variación compleja del nivel 2. En este caso, las diferencias entre secciones dependen de los grados de valoración que se den en cada una de las secciones. Así, la mejor sección, en términos de valoración esperada para electores que muestren una satisfacción de 6 puntos, no es la mejor para aquellos electores que muestren una satisfacción de 8 puntos. Este es el segundo modelo que probaremos y que, como se verá parece mucho más realista para el conjunto de datos de que disponemos si en él, se da realmente una estructura de dos niveles.



En el caso multilevel, cuando las secciones censales son consideradas como una muestra de la población de secciones censales de un territorio, podemos expresar las relaciones que describen las ecuaciones (1.1) y (1.2) de la siguiente forma:

$$y_{ij} = a_j + b_{xij} + e_{ij} \quad (1.3)$$

$$y_{ij}^* = a_j + b_{xij} \quad (1.4)$$

donde la j toma valores de 1 a 7, uno para cada sección censal, de manera que x_{ij} es la satisfacción del elector i de la sección censal j .

En general, los subíndices i serán los referentes a los electores y los subíndices j los referentes a las secciones censales. Si un ítem lleva solamente un subíndice j , quiere decir que varía a lo largo de las diferentes secciones censales pero que tiene el mismo valor para todos los electores dentro de cada sección censal. Si un ítem tiene subíndices ij , entonces varía de elector en elector dentro de cada sección censal. Finalmente, si un ítem no tiene subíndices, entonces se le considera constante tanto para los electores como para las secciones censales.

Por tanto, x_{ij} es la satisfacción del elector i en la sección j ; y_{ij} es la valoración del elector i en la sección j ; a_j es el punto de corte de la regresión correspondiente a la sección j y b es la pendiente, que en este primer modelo se asume como constante a través de todas las secciones. Los e_{ij} representan los residuos, en este caso, del elector i en la sección j y las y_{ij}^* las predicciones de valoración de los electores en cada sección.

De ello se desprende que es posible escribir:

$$e_{ij} = y_{ij} - y_{ij}^*$$

En un análisis multilevel, los elementos del nivel 2, las secciones censales en este caso, son tratados como una muestra aleatoria. Por tanto, podemos re-expresar la ecuación (1.4) de la siguiente forma:

$$y_{ij}^* = a + bx_{ij} + u_j \quad (1.5)$$

donde $u_j = a_j - a$ es la diferencia del punto de corte de la sección j respecto del punto de corte a que proviene de la ecuación en que se consideraban todas las secciones, como si el nivel 2 no existiese. Por tanto, a y u_j lo podemos identificar como el representante de los residuos del nivel 2 (hay que tener presente que a no tiene subíndice y que por tanto es constante para todos los electores y secciones y que u_j tiene subíndice j que indica que varía en las diferentes secciones, pero que es igual para todos los electores de la sección j).

Así, el modelo general, puede escribirse como:

$$y_{ij} = a + bx_{ij} + u_j + e_{ij} \quad (1.6)$$

ecuación en la que, tanto u_j como e_{ij} son términos aleatorios y, por tanto, sus medias, son igual a cero, tal y como sabemos que se impone en las hipótesis de la regresión clásica.

En este contexto, se asume que habiendo diferentes niveles, estas variables están incorrelacionadas y siguen una distribución aproximadamente Normal. Por tanto, es suficiente con estimar sus correspondientes varianzas para determinar el modelo, ya que los términos a y b son comunes a todos los miembros de la población o, dicho de otra forma: son fijos, no aleatorios y, por tanto, no necesitan ser estimados.

Es precisamente la existencia de las dos variables aleatorias u_j y e_{ij} en la ecuación (1.6) la que nos indica que estamos en presencia de un modelo multilevel. Las varianzas de estos términos, que simbolizaremos como σ_u^2 y σ_e^2 son los llamados parámetros aleatorios del modelo y en cambio, a y b se conocen como parámetros fijos del modelo.

Un modelo multilevel de tipo simple, como éste, en que los únicos parámetros aleatorios son las varianzas de los residuos de cada nivel, se llama Variance Components Model.

Seguidamente, vamos a estimar este modelo para los datos de 7 secciones censales y 300 electores de que disponemos.

Los parámetros a estimar en nuestro primer modelo son:

Los parámetros fijos a y b

Los parámetros aleatorios σ_u^2 y σ_e^2

El programa ML3 necesita que cada parámetro esté asociado con una variable explicativa. Así, el parámetro fijo b , está asociado con el grado de satisfacción de vivir en el municipio. Para el resto, es necesario construir una variable ficticia, que sea una constante con tantos 1 como casos haya en el nivel 1. A esta variable se la llama CONS y, a partir de todo esto, el modelo que vamos a estimar se puede escribir así:

$$(Valoración)_{ij} = a(Cons) + b(Satisfacción)_{ij} + u_j(Cons) + e_{ij}(Cons) \quad (1.7)$$

Es usual referirse a los términos con parámetros fijos como “la parte fija del modelo” y como “parte aleatoria del modelo” a los términos con parámetros aleatorios. Ello está relacionado con las órdenes que se da al programa para especificar un modelo multilevel en ML3. Por ejemplo, vamos a ver el aspecto que tiene la parte de especificaciones de un modelo en la aplicación ML3 antes de llenarla con nuestras instrucciones:

```

ML3 - Software for three-level analysis.

sett
EXPLanatory variables in
FPARameters
FMEANs
RMEANs
RESPonse variable in
IDENTifying codes for level 1:          level 2:          level 3:
RESEtting covariances level 1: ON      level 2: ON        level 3: ON
MAXIteratIons 5      TOLERance 2      METHod is IGLS    BATCH is OFF

LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX unspecified
LEVEL 1 RANDOM PARAMETER MATRIX unspecified

```

En la línea que pone “IDENTifying codes for level 1: level 2: level3: “ está esperando que coloquemos las variables que servirán para identificar a los sujetos del nivel 1 (los electores) y a los del nivel 2 (las secciones). El nivel 3 de momento no vamos a utilizarlo.

Por tanto, vamos a recordar la estructura completa de la base de datos y a colocar los nombres de las variables que precisa la plantilla anterior.

NAMES					
	Name	n	missing	min	max
1	ELECTOR	300	0	1.0000	66.000
2	CONS	300	0	1.0000	1.0000
3	DISTRIC	300	0	1.0000	2.0000
4	SECC	300	0	1.0000	7.0000
5	SATISF	300	0	2.0000	9.0000
6	VALORA	300	0	1.0000	9.0000
7	VOTOS95	300	0	753.00	1201.0
8	CIUA95	300	0	371.00	656.00
9	PSCA95	300	0	142.00	244.00
10	ICA95	300	0	34.000	77.000
11	ERCA95	300	0	64.000	184.00
12	CONTINU	300	0	1.0000	3.0000

En el comando de la primera línea de la tabla siguiente indicamos al programa que identifique como códigos de los individuos del nivel 1 a los contenidos en la variable “ELECTOR” y como códigos de los individuos del nivel 2 a los contenidos en la variable “SECC”. Ante esta petición seguida del comando SUMM, obtenemos un resumen del número de electores que hay en cada una de las 7 secciones censales de este municipio y el porcentaje que representan en el total de la muestra.

IDEN CODES FOR LEVEL 1 IN "ELECTOR" AND LEVEL 2 IN "SECC"							
SUMM							
7 level 2 units		300 level 1 units					
ID2:	1	2	3	4	5	6	7
N :	66	37	30	43	48	36	40
% :	22	12	10	14	16	12	13

Seguidamente indicamos que las variables explicativas son "CONS" (la constante cuya columna sólo contiene unos) y "SATISF". Después de esta operación, el comando SETT nos proporciona una plantilla en la que ya figura esta información:

```

EXPL "CONS" "SATISF"
SETT
EXPLANATORY variables in      CONS      SATISF
FPARAMETERS                   CONS      SATISF
PMEANS
RMEANS
RESPONSE variable in
IDENTIFYING codes for level 1: ELECTOR level 2: SECC level 3:
RESETTING covariances level 1: ON level 2: ON level 3: ON
MAXITERATIONS 5 TOLERANCE 2 METHOD is IGLS BATCH is OFF
LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX unspecified
LEVEL 1 RANDOM PARAMETER MATRIX unspecified

```

Sin embargo, falta por indicar las variables a las que vamos a asociar las variables aleatorias del modelo. Por tanto, el proceso continúa con las órdenes pertinentes (en este caso SETV que sirve para asociar la variación del nivel 2 y luego del 1 a CONS) y cuando ya se ha indicado cuál es la variable dependiente (VALORA en este caso y mediante el comando RESPONSE), un nuevo comando SETT nos proporciona la plantilla completa lista para poder efectuar el análisis.

```

SETV 2 "CONS"
SETV 1 "CONS"
RESPONSE "VALORA"
SETT
EXPLANATORY variables in      CONS      SATISF
FPARAMETERS                   CONS      SATISF
PMEANS
RMEANS
RESPONSE variable in          VALORA
IDENTIFYING codes for level 1: ELECTOR level 2: SECC level 3:
RESETTING covariances level 1: ON level 2: ON level 3: ON
MAXITERATIONS 5 TOLERANCE 2 METHOD is IGLS BATCH is OFF

LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
CONS
CONS 1
LEVEL 1 RANDOM PARAMETER MATRIX
CONS
CONS 1

```

La orden Start pone en marcha el proceso iterativo de estimación de los cuatro parámetros que hemos especificado en esta plantilla: a, b, σ_u^2 y σ_e^2 . En el siguiente cuadro se puede ver el tipo de información que proporciona el programa mientras realiza las iteraciones. Primero se para y cuando ponemos el modo BATCH ON, comienza de nuevo a partir de unos valores iniciales que antes no tenía. Al finalizar la segunda iteración, ha alcanzado la convergencia.

```

START
Iteration number 1 in progress
Iteration number 1 completed
Convergence not achieved

BATCH
BATCH mode is ON

NEXT
Iteration number 2 in progress
Iteration number 2 completed
Convergence achieved

```

Los resultados de las estimaciones se piden por separado: primero los correspondientes a la parte fija del modelo y después los de la parte aleatoria:

FIXED PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE
CONS	3.246	0.5323	3.246
SATISF	0.3495	0.06735	0.3495

RANDOM				
LEVEL 3				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
LEVEL 2				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
CONS /CONS	0	0	0	1
LEVEL 1				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
CONS /CONS	2.581	0.2107	2.581	1

Los resultados de todas las operaciones anteriores son los que comentaremos a continuación. En primer lugar se ven las estimaciones de la parte fija del modelo, es decir a (3.246) y b (0.3495) con sus errores estándar (0.5323 y 0.06735 respectivamente) y las estimaciones previas en la anterior iteración.

La regresión simple, de la parte fija del modelo tiene por lo tanto el siguiente aspecto:

$$\text{"VALORA*"} = 3.246 + 0.3495 * \text{"SATISF"}$$

De forma que, cuando la satisfacción de vivir en el municipio es cero, el ciudadano valora por término medio en 3.246 puntos (en una escala de 1 a 9), la actuación del actual gobierno. A medida que la satisfacción aumenta en un punto, la valoración del actual gobierno aumenta en 0.3495 puntos.

Los errores estándar de estos dos parámetros fijos no son elevados, teniendo presente que trabajamos con escalas de valoración del 1 al 9. Especialmente, se indica que la precisión de la estimación de b es bastante mejor, proporcionalmente que la de a , aunque el punto de corte existe y no es cero (es decir, no sucede poblacionalmente que a satisfacción cero, valoración del gobierno cero).

Seguidamente, se puede ver la parte aleatoria del modelo, que es la que lo diferencia y lo convierte en multilevel. La zona reservada a los resultados del nivel 3 está vacía puesto que sólo hemos trabajado con dos niveles. Seguidamente aparece la estimación de la varianza de los residuos del nivel 2 (secciones), que es cero y que tiene un error estándar cero. Esto quiere decir que bajo el supuesto de que todas las rectas de regresión de las secciones tienen la misma pendiente, no hay variación en los puntos de corte y por tanto, el segundo nivel no existe y no influye en la variación de la variable VALORA.

Finalmente, se puede ver la estimación de la varianza de los residuos del nivel 1 que es 2.581 con un error estándar de 0.2107 y la estimación en la iteración previa. Esta variación es, en este caso, la representante de la dispersión de los electores alrededor de la recta global de regresión, puesto que bajo estos supuestos de trabajo no se ha podido confirmar la existencia del nivel 2. Para poder interpretar con propiedad la varianza de los residuos del nivel 1, es mejor extraer la raíz cuadrada (1.6) y así, poder decir que, por término medio, los electores se alejan de la recta ajustada en 1.6 puntos (positivos y negativos), lo cual, en una escala de 1 a 9 representa una dispersión algo notable e indica que el ajuste no es muy bueno.

Por tanto, de la estimación de la ecuación

$$(\text{Valoración})_{ij}^* = a(\text{Cons}) + b(\text{Satisfacción})_{ij} + u_j(\text{Cons}) + e_{ij}(\text{Cons}) \quad (1.7)$$

hemos extraído lo siguiente:

$$(\text{Valoración})_{ij}^* = 3.246(\text{Cons}) + 0.3495(\text{Satisfacción})_{ij} + u_j(\text{Cons}) + e_{ij}(\text{Cons}) \quad (1.7)$$

$$\text{con } \sigma_u^2 = 0 \text{ y } \sigma_e^2 = 2.581$$

Al ser cero la varianza de los residuos del segundo nivel, los resultados indican que bajo el supuesto de estar en un modelo de Variance Components, en realidad no existe una estructura jerárquica y que los electores no se comportan de forma diferente en secciones en que se supone que las varianzas son constantes entre ellas.

Como comentario adicional, conviene saber que el criterio de convergencia que tiene implementado el programa por defecto, es el siguiente:

$$(\text{nueva estimación} - \text{estimación previa}) / (\text{estimación previa}) < 10^{-2}$$

que deriva de un nivel de tolerancia igual a 2. Se puede cambiar dicho nivel si el investigador lo considera conveniente.

En la columna NCONV aparece el número de iteraciones anterior al que se ha logrado la convergencia.

A continuación, vamos a aplicar la segunda posibilidad de este modelo, es decir aquella en que se supone que las pendientes varían de sección en sección. En el contexto del análisis multilevel, a esta posibilidad se la conoce como "random slopes regression".

En la prueba anterior habíamos considerado que las pendientes eran iguales para todas las secciones censales de la población, es decir, que las diferencias entre secciones con electores que otorgan buenas valoraciones eran iguales que las diferencias con electores que otorgan valoraciones bajas. Ahora vamos a ver qué pasa cuando suponemos que las pendientes varían de sección en sección.

Al igual que antes, nuestro interés no se centra en la estimación de las pendientes de cada una de las secciones sino en la estimación de la varianza de dichas pendientes. Si existe tal varianza, las secciones existen como segundo nivel e influyen en la valoración de los electores y, si no existe tal varianza, las secciones no existen como segundo nivel y el municipio es uniforme en cuanto a comportamiento de valoración del equipo de gobierno: en todas las secciones se valora por término medio de la misma forma, ya sea buena o mala dicha valoración.

Para comprender a nivel matemático la idea que vamos a desarrollar, partamos del modelo anterior:

$$y_{ij} = a_j + b_{xij} + e_{ij} \quad (1.3)$$

$$y_{ij}^* = a_j + b_{xij} \quad (1.4)$$

En este caso, la b es constante, por tanto, para permitir que varíe entre las diferentes secciones censales podemos escribir:

$$y_{ij} = a_j + b_{jxij} + e_{ij} \quad (1.9)$$

donde b_j será la pendiente de la sección j . Ahora, podemos re-expresar b_j tal y como hicimos con $a_j = a + u_j$, de la siguiente forma:

$$b_j = b + v_j$$

Así, las constantes a y b siguen siendo cantidades fijas y representan los valores medios poblacionales del punto de corte y la pendiente (valoración media del gobierno de toda la población cuando la satisfacción es cero e incremento de la valoración cada vez que la satisfacción aumenta en un punto). Por otro lado, u_j y v_j representan el punto hasta el cual la sección j se aleja de la media.

Tanto u_j como v_j son variables aleatorias y ambas varían en el nivel 2 (las secciones). De nuevo se asume que siguen una distribución Normal y que tienen media cero y unas varianzas que son objeto de nuestra estimación.

Al variar al mismo nivel, la consecuencia es que u_j y v_j pueden estar correlacionadas, de forma que resulta necesario estimar la covarianza existente entre ambas. Por tanto, el modelo puede escribirse de la siguiente forma:

$$y_{ij} = (a+u_j) + (b+v_j)x_{ij} + e_{ij} \quad (1.10)$$

y para que quede más claro cuál es la parte fija y cual la aleatoria, se puede organizar de esta manera:

$$y_{ij} = (a + b_{xij}) + (u_j + v_{jxij} + e_{ij}) \quad (1.11)$$

Ahora, el número de parámetros a estimar es superior al del primer modelo:

Hay que estimar los parámetros fijos a y b.

Hay que estimar la varianza de la variable v_j que simbolizaremos como σv^2 .

Hay que estimar la covarianza entre las variables u_j y v_j que simbolizaremos mediante σ_{uv} .

Hay que estimar la varianza de la variable e_{ij} que simbolizaremos como σe^2 .

Veamos la forma de poner en práctica todo esto en el ML3, que además, proporciona, en este caso, un contraste de la significación de los términos aleatorios extra que se han añadido al modelo.

En primer lugar, fuera del programa ML3, comprobamos el nivel de correlación entre las variables VALORA y SATISF. Como puede apreciarse en la siguiente tabla, dicha correlación existe y, aunque no es muy intensa, es positiva y significativa. Por tanto, una cifra tan baja puede dar como resultado un modelo en que de nuevo se ponga de manifiesto que las secciones no influyen en los resultados y que nos hallamos frente a una estructura no jerárquica de estos datos.

Correlaciones

		SATISFACCIÓN	VALORACIÓN
Correlación de Pearson	SATISFACCIÓN	1,000	,300**
	VALORACIÓN	,300**	1,000
Sig. (bilateral)	SATISFACCIÓN	,	,000
	VALORACIÓN	,000	,
N	SATISFACCIÓN	300	297
	VALORACIÓN	297	297

** . La correlación es significativa al nivel 0,01 (bilateral).

Es evidente que al investigador le agradaría hallar un modelo que mostrase justamente lo contrario, pero recordemos que ya hemos dicho en otras secciones de la tesis que un resultado negativo es tan importante como uno positivo y que llegar a la conclusión de que no hay varios niveles es tan importante como llegar a la contraria.

Por tanto, aplicamos el análisis de variación compleja al nivel 2, sospechando, de entrada que rechazaremos la hipótesis de existencia de diferencias entre los electores por secciones respecto a la variable VALORA explicada por la variable SATISF.

```

SETV 2 "SATISF"
SETT
EXPLANATORY variables in      CONS      SATISF
FParameters                   CONS      SATISF
FMEANS
RMEANS
RESPONSE variable in          VALORA
IDENTIFYING codes for level 1: ELECTOR  level 2: SECC  level 3:
RESETTING covariances level 1: ON      level 2: ON    level 3: ON
MAXITERATIONS 5  TOLERANCE 2  METHOD is IGLS  BATCH is ON

LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
      CONS      SATISF
CONS  1
SATISF 1          1
LEVEL 1 RANDOM PARAMETER MATRIX
      CONS
CONS  1
    
```

Al colocar el comando SETV 2 "SATISF" hemos permitido la variación en el nivel dos de las pendientes. Ahora vamos a estimar la varianza que nos dirá si efectivamente dicha dispersión existe y, por tanto, se puede hablar de comportamiento distinto en las secciones censales.

Al poner en marcha la estimación, aparecen mensajes que indican que no se logra la convergencia y de que aunque aumentemos el número de iteraciones, ésta no se logrará porque los términos aleatorios del modelo son linealmente dependientes.

```
START
Iteration number 1 in progress
fixed parameter ssp matrix -ve definite
reconstruction used
Iteration number 1 completed
Convergence not achieved
MAXI 20
NEXT

caution - random parameters may be linearly dependent

fixed parameter ssp matrix -ve definite
reconstruction used
Iteration number 2 completed
Convergence not achieved
NEXT

Etc. etc.

Iteration number 23 in progress
random parameter ssp matrix -ve definite
reconstruction used

caution - random parameters may be linearly dependent

Iteration number 23 completed
Convergence not achieved
```

Por tanto, si detenemos el proceso y pedimos los resultados de estimación de la parte fija y aleatoria del modelo, podremos ver que la parte fija proporciona una a de 3.383 y una b de 0.3321, muy parecidas a las del primer modelo, pero con errores estándar más elevados y que en la parte aleatoria se dan los valores iniciales antes de comenzar las iteraciones, con una covarianza negativa entre las dos variables aleatorias u_j v_j , difícil de ignorar y que indica que el modelo no va a ajustar bien por problemas de multicolinealidad. Las estimaciones de las varianzas σ_u^2 y σ_v^2 indican que hay variación en el nivel 1 y que no hay variación significativa en el nivel 2, teniendo que concluir que no existe una estructura jerárquica en nuestros datos y que la poca o mucha dependencia de la valoración del gobierno municipal de la satisfacción de vivir en el municipio es homogénea en todas las secciones censales.

FIXED			
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE
CONS	3.383	0.9341	3.39
SATISF	0.3321	0.1094	0.3427

RANDOM				
LEVEL 3				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
LEVEL 2				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
CONS /CONS	3.896	4.248	5.479	0
SATISF /CONS	-0.4371	0.4323	-0.5073	0
SATISF /SATISF	0.04894	0.04346	0.04238	0
LEVEL 1				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
CONS /CONS	2.609	0.2216	2.656	0

2.6 MODELO "VARIANCE COMPONENTS MODEL": SEGUNDA EXPERIENCIA

La primera aplicación de una modelización multilevel nos ha conducido a la conclusión de que si no existe una estructura jerárquica en los datos, el análisis no superará lo que puede proporcionarnos una regresión clásica. Esto se puede comprobar fácilmente realizando la regresión simple de las variables VALORA y SATISF con el SPSS, cuyos resultados se ofrecen a continuación y donde se puede ver que a vale 3.143 (3.246 con ML3), su error estándar 0.535 (0.5323 con ML3) y que b vale 0.365 (0.3495 con ML3) con un error estándar de 0.068 (0.06735 en ML3). El error estándar de la regresión es de 1.61 puntos, que coincide con la raíz cuadrada que hemos extraído de la estimación de la varianza de los residuos del nivel 1 en ML3 ($\sqrt{2.609} = 1.61$).

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error tip. de la estimación
1	,300 ^a	,090	,087	1,61

a. Variables predictoras: (Constante), SATISFACCIÓN DE VIVIR EN EL MUNICIPIO

Coefficientes^a

Modelo		Coefficients no estandarizados		Coefficients estandarizados	t	Sig.
		B	Error tip.	Beta		
1	(Constante)	3,143	,535		5,877	,000
	SATISFACCIÓN DE VIVIR EN EL MUNICIPIO	,365	,068	,300	5,396	,000

a. Variable dependiente: VALORACIÓN DE LA GESTIÓN DEL ACTUAL EQUIPO DE GOBIERNO

Una vez comprobado este aspecto, pasamos a presentar una segunda experiencia con otra base de datos.

Se dispone de una encuesta electoral efectuada a 887 electores de una comarca catalana, seleccionados aleatoriamente en 48 secciones censales de diversos municipios de la misma. En este caso, se puede aplicar un análisis multilevel porque se sabe a qué sección electoral pertenecía cada entrevistado.

De todas las variables que componen esta encuesta, se han seleccionado las siguientes para efectuar una primera aproximación a la modelización multilevel:

```

ML3 - Software for three-level analysis.  Sun Jan 18 15:03:02 1998
NAMES
  Name          n  missing  min      max
1 SECC          887      0      1.0000  48.000
2 ELECTOR       887      0      1.0000  62.000
3 ORIGEN       887      0      0.0000   1.0000
4 VALORA91     887      0      4.0000  36.000
5 VALORA95     887      0      5.0000  40.000
6 GENERO       887      0      0.0000   1.0000
7 CONS         887      0      1.0000   1.0000
    
```

En primer lugar se tiene la sección general que toma valores de 1 a 48. En segundo lugar los electores, que para la primera sección tienen códigos del 1 al 62. En tercer lugar, el origen, siendo 1 catalán y 0 no catalán. Seguidamente vienen las respuestas a una misma pregunta efectuada en las campañas de las elecciones del 91 y del 95: ¿en qué porcentaje piensa usted que el gobierno elegido cumplirá sus promesas electorales?. Como puede verse, en el 91 las respuestas oscilan del 4 al 36% y en el 95 del 5 al 40%. Finalmente está la variable género (siendo 1 los hombres y 0 las mujeres) y una constante que ya hemos explicado que necesita el modelo de regresión y que es una columna formada por unos.

La parte de los 20 primeros casos de la base de datos tiene el siguiente aspecto:

```
PRINT C1-C7
```

N =	SECC 887	ELECTOR 887	ORIGEN 887	VALORA91 887	VALORA95 887
1	1.0000	1.0000	0.00000	23.000	23.000
2	1.0000	2.0000	1.0000	22.000	39.000
3	1.0000	3.0000	0.00000	14.000	32.000
4	1.0000	4.0000	0.00000	16.000	11.000
5	1.0000	5.0000	0.00000	17.000	26.000
6	1.0000	6.0000	0.00000	21.000	28.000
7	1.0000	7.0000	0.00000	32.000	32.000
8	1.0000	8.0000	1.0000	29.000	36.000
9	1.0000	9.0000	0.00000	34.000	33.000
10	1.0000	10.000	0.00000	27.000	33.000
11	1.0000	11.000	0.00000	21.000	30.000
12	1.0000	12.000	0.00000	17.000	25.000
13	1.0000	13.000	0.00000	22.000	17.000
14	1.0000	14.000	0.00000	15.000	18.000
15	1.0000	15.000	0.00000	25.000	33.000
16	1.0000	16.000	0.00000	21.000	20.000
17	1.0000	17.000	1.0000	20.000	13.000
18	1.0000	18.000	0.00000	16.000	25.000
19	1.0000	19.000	0.00000	34.000	27.000
20	1.0000	20.000	1.0000	20.000	21.000

Por tanto, la primera fila corresponde a un elector entrevistado en la sección 1, de origen no catalán, que pensó que el gobierno cumpliría sus promesas en un 23% y que piensa lo mismo de cara a las elecciones del 95.

Todos los electores que se ven en la tabla anterior son de la sección 1. A medida que se avanza en la visión de los datos, esta columna va cambiando hasta pasar, ordenadamente por todos los números hasta el 48. Es imprescindible que la base esté ordenada para que el algoritmo del programa pueda trabajar correctamente.

En nuestro ejemplo, se espera analizar la relación entre la valoración anterior y futura (aunque ambas fechas ya sean muy anteriores) de un equipo de gobierno autonómico por parte del electorado. Para ello, contamos con 887 datos de electores en 48 secciones censales.

Dando por explicados los principales conceptos relacionados con el modelo Variance Components, pasamos a aplicarlo directamente.

En primer lugar se presenta la tabla con el número de electores entrevistados en cada sección censal y el porcentaje que representan sobre el total.

IDEN CODES FOR LEVEL 1 IN "ELECTOR" AND LEVEL 2 IN "SECC"										
SUMM										
49 level 2 units					887 level 1 units					
ID2:	1	2	3	4	5	6	7	8	9	10
N :	25	10	8	24	25	13	10	26	21	11
% :	2.8	1.1	0.9	2.7	2.8	1.5	1.1	2.9	2.4	1.2
ID2:	11	12	13	14	15	16	17	18	19	20
N :	22	19	13	7	15	6	16	12	12	28
% :	2.5	2.1	1.5	0.79	1.7	0.68	1.8	1.4	1.4	3.2
ID2:	21	22	23	24	25	26	27	28	29	30
N :	14	18	21	11	20	19	14	12	20	35
% :	1.6	2	2.4	1.2	2.3	2.1	1.6	1.4	2.3	3.9
ID2:	31	32	33	34	35	36	37	38	39	40
N :	22	39	25	16	27	16	11	13	8	10
% :	2.5	4.4	2.8	1.8	3	1.8	1.2	1.5	0.9	1.1
ID2:	41	42	43	44	45	46	47	48		
N :	38	5	11	13	30	62	22	12		
% :	4.3	0.56	1.2	1.5	3.4	7	2.5	1.4		

Seguidamente se especifica un modelo de Variance components:

```

iden codes for level 1 in "ELECTOR" AND LEVEL 2 IN "SECC"
EXPLA C7 C4
SETT
EXPLANATORY variables in      CONS      VALORA91
FPARAMETERS                   CONS      VALORA91
FMEANS
RMEANS
RESPONSE variable in
IDENTIFYING codes for level 1: ELECTOR level 2: SECC level 3:
RESETTING covariances level 1: ON level 2: ON level 3: ON
MAXITERATIONS 5 TOLERANCE 2 METHOD is IGLS BATCH is OFF

LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX unspecified
LEVEL 1 RANDOM PARAMETER MATRIX unspecified

```

A continuación se asocian los términos de los modelos a las variaciones de los niveles, suponiendo que en el nivel 2 las pendientes de las rectas de regresión son constantes para todas las secciones, es decir, que el incremento de valoración del porcentaje que cumplirá el gobierno tras las elecciones del 95 es el mismo en todas las secciones cuando la valoración anterior aumenta en un 1%.

```

SETV 2 C7
SRTV 1 C7
SETT
EXplanatory variables in      CONS      VALORA91
FPARameters                  CONS      VALORA91
FMEAns
RMEAns
RESPonse variable in
IDENTifying codes for level 1: ELECTOR level 2: SECC level 3:
RESetting covariances level 1: ON level 2: ON level 3: ON
MAXIterations 5 TOLERance 2 METHOD is IGLS BATCH is OFF

LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
CONS
CONS 1
LEVEL 1 RANDOM PARAMETER MATRIX
CONS
CONS 1
RESPONSE C5

```

Seguidamente se pone en marcha el proceso iterativo y se logra la convergencia en la tercera iteración.

```

START

Iteration number 1 in progress
Iteration number 1 completed
Convergence not achieved

BATCH
BATCH mode is ON
NEXT

Iteration number 2 in progress
Iteration number 2 completed

Iteration number 3 in progress
Iteration number 3 completed
Convergence achieved

```

Seguidamente se pueden observar los resultados de la parte fija y aleatoria del modelo en que se han estimado cuatro parámetros: a, b (fijos) y σ^2 , $\sigma\epsilon^2$ (aleatorios).

FIXED	ESTIMATE	S. ERROR	PREV. ESTIMATE
PARAMETER			
CONS	15.14	0.8994	15.14
VALORA91	0.6088	0.03264	0.6088

En la parte fija se nos informa de que a vale 15.14, es decir, que por término medio, el elector piensa que el gobierno cumplirá lo prometido en un 15.14% a partir del 95 cuando la valoración del 91 es que no cumplirá nada. El error estándar es bajo (0.8994) y, en general se puede considerar que la estimación es buena. Por otro lado, el incremento que se produce en la valoración del 95 es de 0.6088% a medida que aumenta en un 1% la valoración del 91. El error estándar de esta estimación también es muy bajo (0.03264), lo cual indica que es una buena estimación. Esta

parte se refiere al nivel 1, es decir sólo al de los electores sin considerarlos incorporados a una sección censal en particular.

RANDOM				
LEVEL 3				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
LEVEL 2				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
CONS /CONS	4.027	1.181	4.019	1
LEVEL 1				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
CONS /CONS	28.13	1.373	28.13	1

En la parte aleatoria, la estimación de la varianza de los residuos del primer nivel es es 28.13, es decir, que los electores se alejan de cada una de las rectas de cada sección censal correspondiente en un 5.3% (raíz cuadrada de 28.13) por encima y por debajo, lo cual indica una dispersión algo notable a esta escala de medición entre las opiniones dentro de cada sección. En el nivel dos, la varianza es mucho menor: 4.027, lo cual indica una desviación estándar del 2% (raíz cuadrada de 5.157) entre las secciones. Los errores estándar de ambas estimaciones no llegan a un 2%, lo cual significa que son estimaciones bastante buenas.

Por tanto, el modelo:

$$(\text{Valora95})_{ij}^* = a(\text{Cons}) + b(\text{Valora91})_{ij} + u_j(\text{Cons}) + e_{ij}(\text{Cons}) \quad (1.7)$$

proporciona la siguiente estimación:

$$(\text{Valora95})_{ij}^* = 15.14(\text{Cons}) + 0.6088(\text{Valora91})_{ij} + u_j(\text{Cons}) + e_{ij}(\text{Cons}) \quad (1.7)$$

$$\text{con } \sigma_u^2 = 4.027 \quad \sigma_e^2 = 28.13$$

Por tanto, los parámetros fijos constituyen un resumen o línea promedio para todos los electores de todas las secciones. La estimación del punto de corte, asociada con la variable CONS es de 15.14%. La estimación de la pendiente, asociada con la variable VALORA91 es 0.6088, de forma que dicha línea promedio es:

$$(\text{VALORA95})^* = 15.14 + 0.6088(\text{VALORA91})$$

Sin embargo, nosotros deseamos probar que la población de secciones censales varía alrededor de esta línea y que son necesarios los restantes términos que convierten al modelo en un multilevel. Nuestro modelo ha estimado esta variación. Se trata de $\sigma_u^2 = 4.027$ que, como ya hemos interpretado, una vez extraída la raíz cuadrada y leída en forma de desviación estándar, indica que las secciones están, en promedio, por encima y por debajo de la línea global en un 2%. Ahora hay que calibrar si esta variación es suficiente o no para considerar que las secciones censales y, por tanto, sus características, influyen en las valoraciones de las promesas del gobierno que hacen los electores.

Para ello, una forma posible es considerar que al haber supuesto que los puntos de corte de las líneas de regresión de las secciones censales se distribuyen normalmente, el 97% de los puntos de corte de estas líneas debe hallarse entre ± 2 desviaciones estándar del corte de la línea global de todos los electores. La relación

matemática que describe esto es la ecuación (1.5) en que u_j es la diferencia entre la pendiente de la recta global a y la de la recta de la sección censal j .

$$y_{ij}^* = a + bx_{ij} + u_j \quad (1.5)$$

con $u_j = a_j - a$

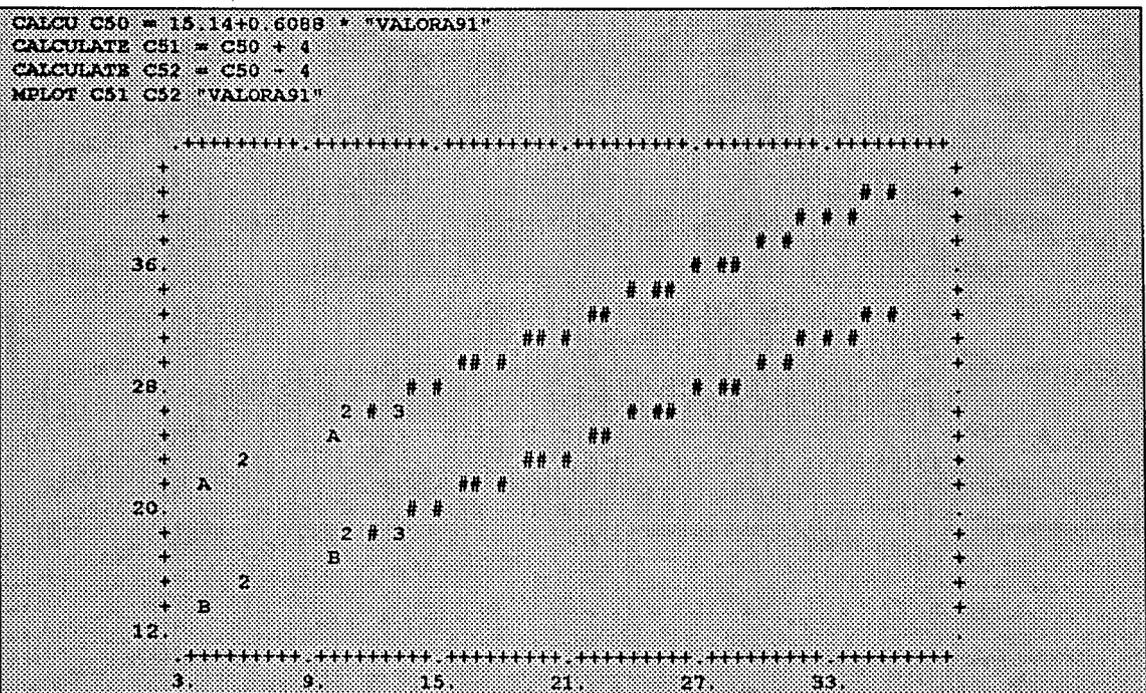
Ahora, si calculamos el error estándar de la regresión global, es decir de la recta:

$$(\text{VALORA95})^* = 15.14 + 0.6088(\text{VALORA91})$$

podremos ver la amplitud del intervalo de los puntos de corte. Para ello, proponemos al ML3 que nos calcule las predicciones de la variable VALORA95 a partir la recta anterior mediante las siguientes órdenes:

```
CALCUL C50 = 15.14 + 0.6088 * "VALORA91"
NAME C50 "PREDVAL95"
```

Y, posteriormente, como la altura de la línea media está en C50 podemos fácilmente crear un gráfico de este intervalo calculando las rectas que lo limitan. Así, la variable C51 se obtiene restando a cada uno de los valores de C50 dos veces la desviación estándar de u_j (aproximadamente $2 \times 2 = 4$) y C52 se obtiene sumando a los valores de C50 la misma cantidad. La representación gráfica de las rectas que señalan el intervalo de confianza de los puntos de corte que contiene el 97% de las rectas de las secciones generales es, por tanto:



De manera que los puntos de corte se mueven entre 12 y 20, aproximadamente, lo cual indica que aunque existe diferencia entre las secciones, ésta no es muy notable o significativa.

Por otro lado, los electores también se dispersan alrededor cada una de las rectas de cada sección censal y su variación viene estimada por medio de $\sigma_e^2 = 28.13$, que comparada con la varianza de las secciones, $\sigma_u^2 = 4.027$, es 7 veces más

grande. Ello nos conduce a observar que los electores se dispersan más alrededor cada una de las rectas de su sección que las secciones censales alrededor de la recta global.

Resumiendo, se puede afirmar que la parte interpretativa de los resultados de un análisis de este tipo no es sencilla con las herramientas de que se dispone y que hay que pensar detenidamente en las escalas en que nos movemos y en las dimensiones de las cifras que aparecen en los resultados. Si tuviésemos que extraer una conclusión, acerca de este caso, en nuestra opinión lo resumiríamos así:

Se trata de una población de electores agrupados en 48 secciones censales que opinan en forma de porcentaje el nivel al que ha cumplido sus promesas un gobierno autonómico elegido en 1991 y la predicción que hacen de lo que cumplirá el que salga elegido tras las autonómicas del 95.

Nuestra hipótesis es que la opinión del 95 depende de la del 91 y que, además, influye en la opinión el hecho de pertenecer a una sección censal u otra.

Los resultados indican que, efectivamente, la opinión del 95 depende de la del 91 y que los electores tienden a ser más homogéneos en sus opiniones en las secciones a que pertenecen que considerados como población sin dividir. Por tanto, se puede hablar de estructura jerárquica a dos niveles en este caso y, de momento, se ha llegado a probar todo ello bajo el supuesto de que el incremento del porcentaje de promesas cumplidas es constante entre las secciones.

Ahora, se trataría de ir más allá y ver qué sucede si se deja variar a las pendientes de las rectas de las secciones. Para ello, ponemos las especificaciones del modelo de varianza compleja al nivel 2 y observamos los resultados:

```

SETV 2 C4
SETT
EXPLANATORY variables in      CONS      VALORA91
FPARAMETERS                   CONS      VALORA91
FMEANS
RMEANS
RESPONSE variable in          VALORA95
IDENTIFYING codes for level 1: ELECTOR level 2: SECC level 3:
RESETTING covariances level 1: ON level 2: ON level 3: ON
MAXITERATIONS 5 TOLERANCE 2 METHOD is IGLS BATCH is ON

LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
      CONS      VALORA91
CONS      1
VALORA91 1      1
LEVEL 1 RANDOM PARAMETER MATRIX
      CONS
CONS      1

```

En estas especificaciones, podemos ver que en el nivel 2 la matriz de parámetros aleatorios a estimar, contiene tres en lugar de los dos del modelo anterior. Recordemos que vamos a estimar el modelo:

$$y_{ij} = (a + bx_{ij}) + (u_j + vx_{ij} + e_{ij}) \quad (1.11)$$

En el modelo que se ha estimado antes, se estaba suponiendo que las diferencias entre secciones con porcentajes altos de cumplimiento de promesas eran iguales que las diferencias de las secciones con electores que proporcionaban porcentajes de bajo cumplimiento. Ahora vamos a explorar un modelo en que estas diferencias (representadas por las pendientes de las rectas de las secciones), pueden variar de sección a sección.

El proceso de puesta en marcha de los cálculos es el siguiente:

```

START
Iteration number 1 in progress
Iteration number 1 completed

Iteration number 2 in progress
Iteration number 2 completed

Iteration number 3 in progress
Iteration number 3 completed

Iteration number 4 in progress
Iteration number 4 completed

Iteration number 5 in progress
Iteration number 5 completed
Convergence achieved
  
```

La convergencia se alcanza tras cinco iteraciones y los resultados de la parte fija del modelo son los siguientes:

FIXED			
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE
CONS	15.04	1.318	15.04
VALORA91	0.6124	0.04285	0.6124

Como se puede observar, el hecho de añadir dos términos aleatorios en el nivel 2 no cambia apenas la regresión simple o promedio para toda la población de electores. La a vale 15.04 (y valía 15.14) y la b vale 0.6124 (y valía 0.6088). Los errores estándar son un poco más altos, de forma que indican que estas estimaciones son un poco peores que las del modelo anterior.

En la parte aleatoria se obtienen las siguientes estimaciones:

RANDOM				
LEVEL 3				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
LEVEL 2				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
CONS /CONS	45.02	16.36	45.05	2
VALORA91 /CONS	-1.231	0.5209	-1.233	2
VALORA91 /VALORA91	0.03429	0.017	0.03433	1
LEVEL 1				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
CONS /CONS	26.96	1.343	26.97	3

Por tanto, en el nivel 2, o de las secciones, se obtiene una estimación de la varianza de las u_j de valor 45.02, una varianza de las v_j de 0.03429 y una covarianza entre u_j y v_j de -1.231. Las diferencias con el modelo anterior son notables:

σ^2 ha pasado de 4.027 a 45.02. Esto ya es un síntoma de que los resultados son inconsistentes ya que en este modelo se apreciaría 10 veces más variación entre secciones que en el otro.

Por tanto, ¿qué ha sucedido para que se den estas diferencias tan enormes?. La explicación está en que el punto de corte de la recta de cada sección es la altura de dicha línea en VALORA95 cuando VALORA91 vale cero. Al no haber intervenido antes la variable VALORA91 como explicativa en el nivel dos, no se ha tenido en cuenta la dispersión que presentan sus valores y, en cambio, ahora aparece en toda su dimensión. Por ejemplo, el valor mínimo de VALORA91 es 3 y antes estábamos estimando la varianza de puntos de corte en valores de VALORA91 como éste que no estaban interviniendo, por estar el nivel 2 asociado únicamente a la variable CONS.

Por tanto, para que los resultados de este análisis sean comparables a los del anterior, la solución está en efectuar el centrado de la variable explicativa VALORA91 respecto de su media, para que esté en las mismas condiciones que la variable CONS. Haciendo esta operación y volviendo a estimar el modelo, podremos interpretar la varianza de los puntos de corte como aquella que indica la variación en las alturas de las líneas de las secciones a partir de la media de la variable VALORA91, variación que será mucho más reducida que si la consideramos alrededor de VALORA91 = 0. Por tanto, cabe esperar que σ^2 se reduzca bastante con esta forma de proceder.

Así, si antes hemos limitado a las líneas de las diferentes secciones a ser paralelas entre sí, a diferentes alturas, ahora vamos a permitir que tengan diferentes pendientes. El modelo anterior no se habría visto afectado por el centrado de VALORA91 porque no intervenía en los cálculos del nivel 2.

El conjunto de operaciones que hay que efectuar para seguir adelante, se detalla a continuación:

En primer lugar, se calcula la media de VALORA91 para poder proceder al centrado de la variable.

AVERAGE "VALORA91"	
Count	= 887
Average	= 25.361
S.D.	= 5.7836
S.E.M.	= 0.19419

En segundo lugar se realiza el centrado y se repite la estimación del modelo.

```
CALCULATE "VALORA91" = "VALORA91" - 25.361
IDEN CODES FOR LEVEL 1 IN "ELECTOR" AND LEVEL 2 IN "SECC"
EXPLA "CONS" "VALORA91"
SETV 2 "CONS"
SETV 1 "CONS"
RESPONSE "VALORA95"
START
```

Las consecuencias para la parte fija del modelo afectan al valor de a que ha pasado a ser el doble: 30.58 y no afectan a b , ni a σ_u^2 ni a σ_e^2 de la parte aleatoria porque el nivel 2 estaba asociado solamente con la constante y no con VALORA91. Por tanto, el punto de corte de la recta general ha sufrido una traslación y, en lugar de medir su altura desde cero lo hace desde de la media de VALORA91.

FIXED			
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE
CONS	30.58	0.3489	30.58
VALORA91	0.6088	0.03264	0.6088

RANDOM				
LEVEL 3				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
LEVEL 2				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
CONS /CONS	4.027	1.181	4.019	1
LEVEL 1				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
CONS /CONS	28.13	1.373	28.13	1

Seguidamente, se prosigue con la estimación del modelo de variación compleja en el nivel 2, asociando la variable VALORA91 a dicho nivel:

```

SETV 2 "VALORA91"
START
Iteration number 1 in progress
Iteration number 1 completed

Iteration number 2 in progress
Iteration number 2 completed

Iteration number 3 in progress
Iteration number 3 completed

Iteration number 4 in progress
Iteration number 4 completed

Iteration number 5 in progress
Iteration number 5 completed
Convergence achieved

```

La convergencia se logra tras cinco iteraciones y los resultados son los siguientes:

FIXED			
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE
CONS	30.57	0.3665	30.57
VALORA91	0.6124	0.04285	0.6124

Tal y como era de esperar, la parte fija apenas sufre cambios, respecto del modelo Variance Components. La a vale 30.57 (y valía 30.58) y la b vale 0.6124 (y valía 0.6088).

En cambio, en la parte aleatoria, ahora, las cifras tienen más sentido. La σ^2 vale 4.608 (y valía 4.027), de manera que el centrado de la variable VALORA91 no cambia la varianza de las pendientes ($\sigma^2=0.03429$, valor muy bajo) o la varianza de los electores alrededor de la línea de su sección censal (que ahora vale $\sigma^2=26.96$ y que antes valía 28.13).

RANDOM				
LEVEL 3				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
LEVEL 2				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
CONS /CONS	4.608	1.302	4.601	2
VALORA91 /CONS	-0.3619	0.1193	-0.362	2
VALORA91 /VALORA91	0.03429	0.01701	0.03434	1
LEVEL 1				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
CONS /CONS	26.96	1.343	26.97	3

La existencia de dos o más variables aleatorias en el mismo nivel hace que el modelo se denomine de variación compleja (Complex variation Model).

Para el modelo que acabamos de estimar, el resumen de las conclusiones es el siguiente:

Hemos permitido que las pendientes varíen entre las secciones censales o nivel 2.

Las secciones con valores altos de punto de corte de sus rectas tienden a tener pendientes de valores más bajos que aquellas que tienen valores bajos del punto de corte. Esto se puede apreciar observando el valor negativo de la covarianza entre u_j y v_j (-0.3619) y calculando la correlación entre los puntos de corte y las pendientes mediante:

$$\sigma_{uv}/\sigma_u\sigma_v = -03619/(2.146)(0.185) = -0.911 .$$

La dispersión entre las líneas o, dicho en otras palabras, la variación total del nivel 2, disminuye a medida que los valores de VALORA91 aumentan. Esto se justifica mediante lo siguiente:

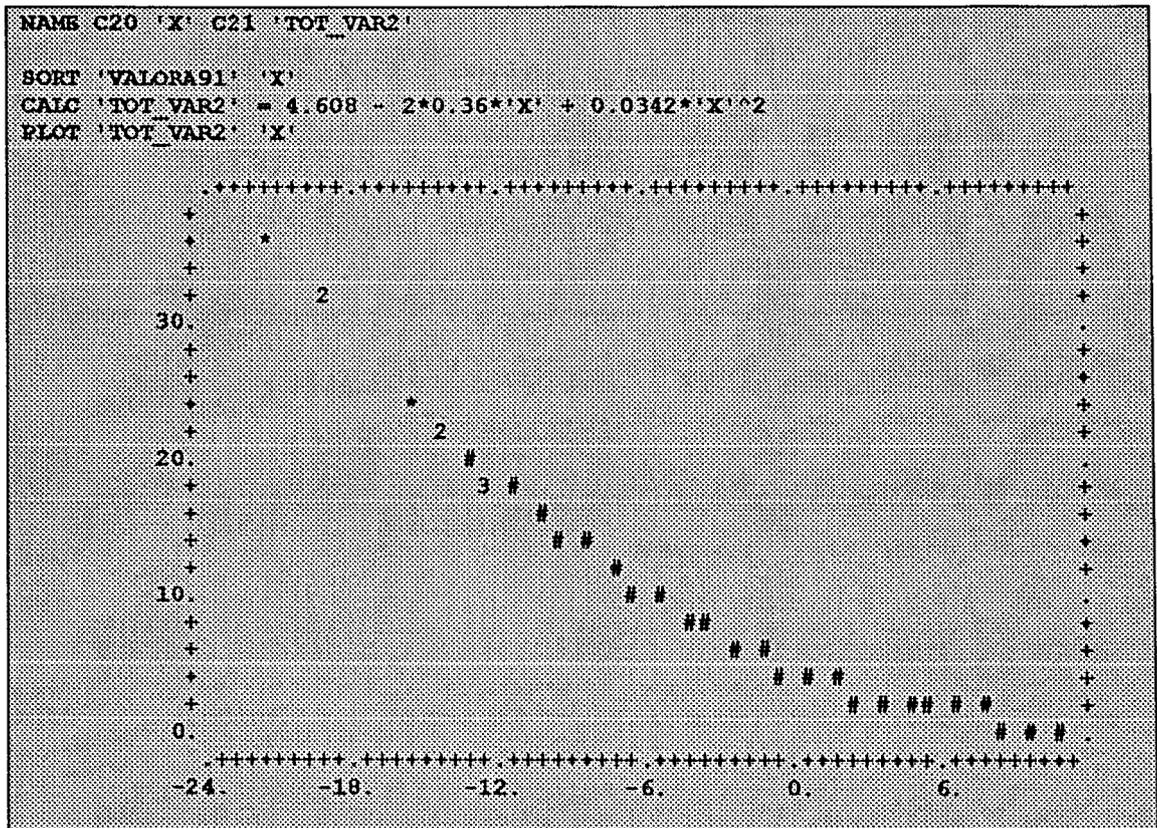
La varianza estimada para las pendientes de este modelo es 0.03429.

La varianza total en el nivel 2 es la varianza de la suma de dos variables aleatorias y se puede expresar como:

$$\text{Var}(u_j + v_j x_{ij}) = \sigma_u^2 + 2(\sigma_{uv} \cdot x_{ij}) + \sigma_v^2 x_{ij}^2 = 4.608 - 2 \times 0.3619(x_{ij}) + 0.03429(x_{ij})^2$$

Siendo x_{ij} la variable VALORA91 en este caso. Por tanto, la varianza total del nivel 2 es una función cuadrática de VALORA91 cuyos coeficientes son los tres parámetros que hemos estimado en el nivel 2. Si se calcula y representa esta función, se puede ver claramente que la varianza entre secciones disminuye a medida que los valores de VALORA91 aumentan.

À continuación se desarrollan dichos cálculos y se representa la función en que en el eje vertical están los valores de la varianza total del nivel 2 y en el eje horizontal los valores de la variable VALORA91.



Como puede apreciarse, no hubiese bastado con calcular la correlación lineal entre las variables Varianza total del nivel 2 y VALORA91 porque la relación que presentan no es de tipo lineal, de manera que, en este caso, el gráfico resulta mucho más descriptivo.

A nivel de interpretación práctica, lo que nos indica esta función es que la dispersión entre las secciones disminuye drásticamente a medida que aumenta el porcentaje de promesas que los electores pensaron que cumpliría el gobierno. Por tanto, son menos homogéneos los electores que pertenecen a secciones donde se han dado valores bajos de VALORA91 que donde se han dado valores altos.

Sin embargo y, a pesar de que estamos trabajando a nivel inferencial, todavía no hemos visto en ningún momento un contraste de hipótesis. De momento todo son conjeturas y pistas que conducen al investigador a ir extrayendo conclusiones con

las que no se muestra tan seguro como si tuviese un estadístico de prueba que mostrar y que fuese algo más contundente que lo anterior.

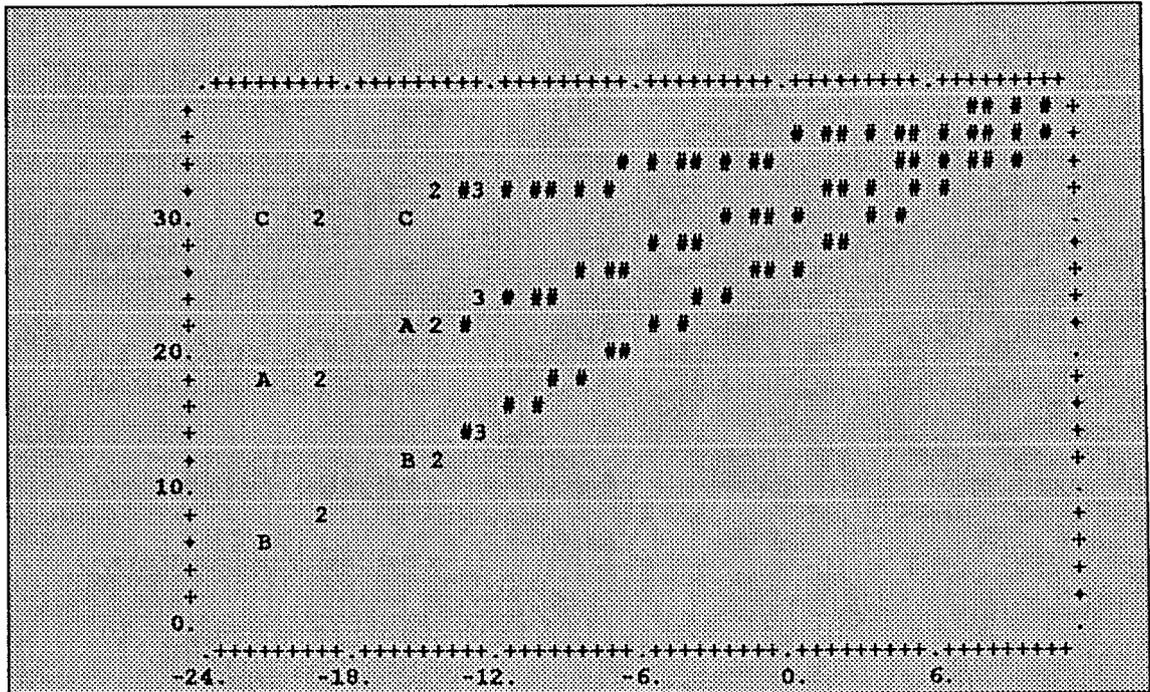
El ML3 y, en general, en el ámbito de la modelización multilevel existen estadísticos de prueba y contrastes de hipótesis, como en cualquier terreno inferencial y, es a partir de ahora cuando entramos en ese terreno.

Para comenzar, hemos calculado un modelo con pendientes supuestamente constantes y seguidamente hemos añadido nuevos parámetros al modelo que nos han permitido variar las pendientes. Parece que este modelo es mejor que el primero, pero podemos comprobarlo mediante el siguiente proceso:

En primer lugar vamos a calcular y a representar, como hemos hecho en el primer modelo, las líneas que delimitan el intervalo de confianza al 97%. El programa es el que sigue y, en él, podemos apreciar de nuevo el cálculo de la variación total del nivel 2, el cálculo de las predicciones de la nueva recta obtenida en la parte fija de este modelo (que irán a parar a la variable C22), el cálculo de la distancia de dos desviaciones estándar que irá a parar a C23, el etiquetado de C22 como LINEAMED o línea media para toda la población y de C23 como dos veces la desviación estándar, el cálculo de las líneas límite del intervalo de confianza restando y sumando dos desviaciones estándar a la línea media y el etiquetado de estas dos líneas que quedan, respectivamente en C24 y C25 como BAJA y ALTA, de forma que se entienda que son los límites inferior y superior del intervalo al 97%. Finalmente se pide la representación gráfica de las tres líneas y, en ella se puede ver, en el centro la línea media y encima y debajo, la región dentro de la cual, en este modelo, podemos asegurar con un 97% de probabilidad que estará la línea correspondiente a una sección censal de esta población.

```
NAME C20 'X' C21 'TOT_VAR2'  
SORT 'VALORA91' 'X'  
CALC 'TOT_VAR2' = 4.608 - 2*0.36*'X' + 0.0342*'X'^2  
CALC C22 = 30.59 + 0.612*'X'  
CALC C23 = 2*'TOT_VAR2'^0.5  
NAME C22 'LINEAMED' C23 '2SD'  
CALCU C24 = 'LINEAMED' - '2SD'  
CALCU C25 = 'LINEAMED' + '2SD'  
NAME C24 'BAJA'  
NAME C25 'ALTA'  
MPLOT 'LINEAMED' 'BAJA' 'ALTA' AGAINST 'X'
```

En el gráfico se aprecia claramente que la región varía y se estrecha a medida que crecen los valores de la variable VALORA91, que es la que representa el eje horizontal. Los valores negativos no deben sorprendernos dado que la variable VALORA91 ha sido centrada con respecto a su media.



Por lo que respecta al contraste de hipótesis, la herramienta de que dispone el ML3 se basa en que a medida que añadimos parámetros para estimar en los modelos, sería factible calcular la probabilidad de que el nuevo parámetro añadido fuese cero y no un valor causado por el azar y el estar trabajando con muestras. Por tanto, el planteamiento es idéntico al de cualquier otro contraste de hipótesis.

Para lograr este propósito, el investigador podría seguir dos caminos:

El primero es más rudimentario y consiste en comparar el parámetro estimado con su error estándar (que es lo que hemos venido practicando) y que funciona bastante bien cuando trabajamos con la parte fija del modelo. A pesar de ser rudimentario, tiene un buen fundamento y es que, con una muestra suficientemente amplia, el cociente entre un parámetro fijo y su error estándar se distribuye de forma aproximadamente Normal (0,1).

El segundo es más aconsejable, sobre todo para los parámetros aleatorios, ya que, en su caso, el cociente anterior puede alejarse mucho de seguir una distribución Normal estándar. En ese caso, el estadístico de prueba deriva del llamado "Likelihood ratio statistic" (estadístico calculado a partir del concepto de máxima verosimilitud) que es igual a $-2 \cdot \log(lh)$. Así, el programa ML3 calcula este valor para cada modelo y se le puede pedir mediante la instrucción:

```
LIKE B40
-2*log(lh) is      5535.8
```

El valor que aparece en el cuadro anterior corresponde al modelo simple, al primero que hemos calculado y el que aparece en el siguiente cuadro corresponde a este último modelo que hemos estimado.

```
LIKE B40  
-2*log(lh) is      5514.16
```

Respecto a este modelo, nuestra hipótesis sería que los valores σ_v^2 y σ_{uv} son cero contra la hipótesis alternativa de que su valor es distinto de cero y no se debe al azar.

El estadístico de prueba de este contraste es la diferencia entre el valor del ratio de máxima verosimilitud del primer modelo y el del segundo. En este caso:

$$5535.8 - 5514.16 = 21.64$$

Este estadístico sigue una distribución Chi cuadrado con un número de grados de libertad igual al número de nuevos parámetros añadidos en el modelo. En este caso, hemos añadido 2 parámetros. Ahora, podemos pedir al programa que calcule la probabilidad de que obtengamos un valor de 21.64 de la Chi cuadrado debido al azar. La probabilidad es:

```
CPRO OF 21.6 WITH 2 DF  
0.000020399
```

de forma que podemos rechazar tranquilamente la hipótesis formulada y aconsejar que los parámetros sean incluidos en el modelo²⁰⁷.

Con esta herramienta, queda bastante claro que el camino a seguir en una investigación multilevel es comenzar por los modelos más sencillos e irlos complicando hasta que el contraste de la hipótesis que acabamos de proporcionar indique que no existe fundamento para añadir más parámetros a estimar y que por tanto, no intervienen más efectos en el fenómeno estudiado que se quedará con la estructura jerárquica a que se nos haya permitido llegar.

Para terminar con la visión completa de esta experiencia, debemos llevar a cabo, al igual que lo haríamos con cualquier análisis de la regresión, una exploración de los residuos y de su comportamiento.

Por tanto, debemos asegurarnos de que se cumple que los residuos del nivel 1 tienen varianza constante²⁰⁸. El programa ML3 tiene las siguientes herramientas para estimar y examinar los residuos:

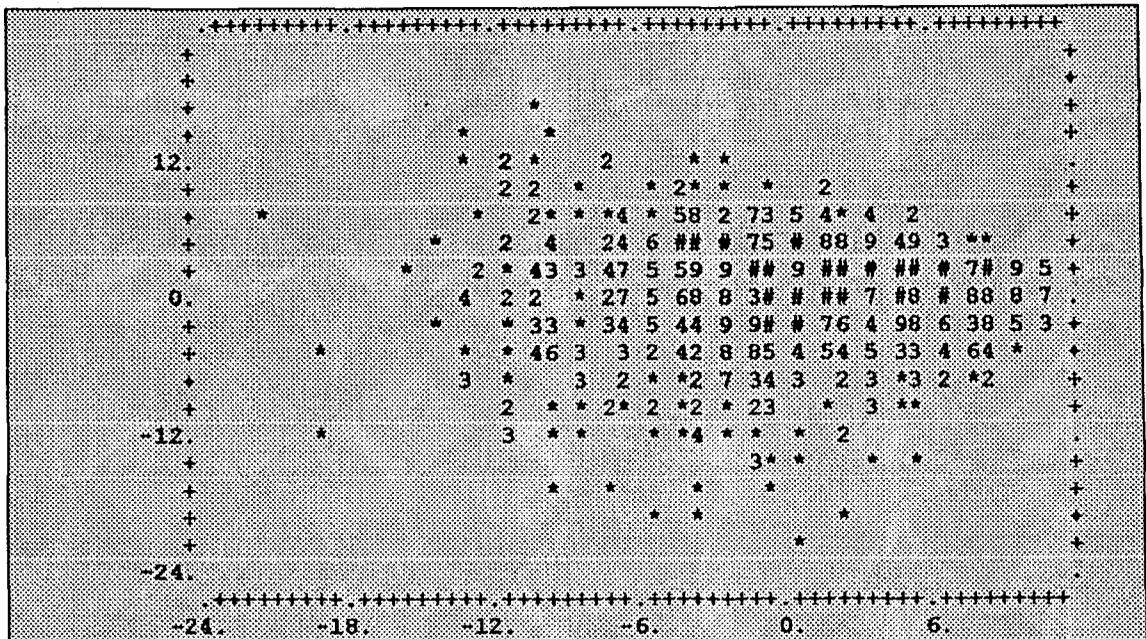
```
RESIDUALS FOR 1 LEVEL 1 VARIABLE ASSOCIATED WITH 'CONS' RESULTS TO  
C30  
PLOT C30 'VALORA91'
```

En primer lugar, calcular los residuos del nivel 1 y colocarlos en una variable, que en este caso llamamos C30.

²⁰⁷ Como es familiar entre los estadísticos, cuando la probabilidad es inferior a 0.05, nos hallamos en zona de rechazo de la hipótesis nula cuando menos al 95% de confianza.

²⁰⁸ Hipótesis básica de cualquier modelo de regresión.

En segundo lugar, representar gráficamente los residuos contra la variable explicativa VALORA91 y observar si el comportamiento sigue un patrón o es aleatorio.



La inspección del gráfico, que tiene en el eje vertical los valores de los residuos de todos los electores y en horizontal sus porcentajes (centrados) de la variable VALORA91, indica que la varianza de los residuos del nivel 1 no es constante y disminuye a medida que aumentan los valores de la variable independiente.

Por tanto, existe una dependencia aparente entre los residuos y la variable explicativa que hay que tratar adecuadamente.

En el contexto del análisis multilevel, a esta fase se la conoce como la modelización de la varianza del nivel 1.

En muchos modelos sólo se ajusta una varianza constante en el nivel 1. Si, como en nuestro caso, el examen de los residuos de dicho nivel sugiere que esta varianza no es constante, entonces debe ser modelizada como una función de los valores de la variable explicativa que le afecta.

La forma de proceder es muy parecida a la empleada para construir un modelo de variación compleja del nivel 2, pero su interpretación es más complicada y limitada.

Por tanto, al igual que antes, se trata de añadir otro término aleatorio al modelo, al que simbolizaremos mediante f_{ij} . La ecuación será ahora:

$$y_{ij} = (a + bx_{ij}) + (u_j + v_j x_{ij} + e_{ij} + f_{ij} x_{ij}) \quad (1.12)$$

Esta formulación muestra claramente que además de los parámetros fijos a y b , ahora hay cuatro variables aleatorias en el modelo, dos en el nivel 2 y dos en el nivel 1. Ahora estimaremos las varianzas de las cuatro, junto con las covarianzas de las parejas de cada nivel. Por tanto, vamos a estimar tres parámetros aleatorios en el nivel 2 y tres en el nivel 1.

Una formulación alternativa del modelo que lo clarifica respecto de los nombres de las variables originales es la siguiente:

$$(VALORA95)_{ij} = (a+u_j+e_{ij})(CONS) + (b+v_j+f_{ij})(VALORA91)_{ij} \quad (1.13)$$

Los parámetros aleatorios a estimar en el nivel 1 son:

σ_e^2 , la varianza de la variable e_{ij} que es una variable aleatoria del nivel 1 asociada a la variable CONS

σ_f^2 , la varianza de la variable f_{ij} que es una variable aleatoria del nivel 1 asociada con la variable VALORA91

σ_{ef} , la covarianza entre e_{ij} y f_{ij}

Por tanto, la expresión de la varianza total del nivel 1 es:

$$\text{Var} (e_{ij} + f_{ij}x_{ij}) = \sigma_e^2 + 2 \sigma_{ef}x_{ij} + \sigma_f^2x_{ij}^2 = \sigma_e^2 + 2 \sigma_{ef}(VALORA91)_{ij} + \sigma_f^2(VALORA91)_{ij}^2$$

Ecuación (1.14).

De esta forma, la varianza total del nivel 1 es una función cuadrática de la variable VALORA91. El único propósito por el cual se ha introducido el término extra $f_{ij}x_{ij}$ en la ecuación (1.12) es para poder plantear esta expresión de la varianza total del nivel 1, no existiendo otra interpretación a nivel práctico de f_{ij} como sucedía con v_j que podía interpretarse como la diferencia de la pendiente de la sección j de la pendiente de la recta media.

Ahora, se procede a programar el ML3 para que incorpore este término y estime el modelo. Las instrucciones y los resultados se ofrecen a continuación.

```
SETV 1 'VALORA91'
START
```

```
Iteration number 1 in progress
Iteration number 1 completed
```

```
Iteration number 2 in progress
Iteration number 2 completed
```

```
Iteration number 3 in progress
Iteration number 3 completed
```

```
Iteration number 4 in progress
Iteration number 4 completed
```

```
Iteration number 5 in progress
Iteration number 5 completed
```

```
Convergence not achieved
```

```
FIXED
```

PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE
CONS	30.56	0.3524	30.56
VALORA91	0.6169	0.0328	0.6166

RANDOM				
LEVEL 3				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
LEVEL 2				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
CONS /CONS	4.143	1.204	4.142	2
VALORA91 /CONS	-0.2348	0.0953	-0.2337	1
VALORA91 /VALORA91	0.01499	0.009819	0.0151	1
LEVEL 1				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
CONS /CONS	27.66	1.467	27.61	2
VALORA91 /CONS	-1.208	0.1384	-1.208	3
VALORA91 /VALORA91	0.001131	0.02747	0.001953	0

La parte fija del modelo no cambia y, en la parte aleatoria, si comparamos el valor de la varianza de σ^2 (0.001131) con su error estándar (0.02747), vemos que se trata de un parámetro no significativo y que su valor debería ser cero, es decir, deberíamos aceptar la hipótesis de que es cero. Para mayor seguridad, si pedimos el ratio de máxima verosimilitud de este modelo, tendremos el valor:

LIKE B40	
-2*log(lh) is	5375.42

que indica que la diferencia con el valor del modelo de varianza constante en el nivel 1 ($5514.16 - 5375.42 = 138.74$) es demasiado grande como para retener la covarianza σ_{ef} en el modelo. Es decir, al añadir nuevos parámetros, el contraste individual indica que uno de ellos es cero, mientras que el otro es significativo y debería incluirse en el modelo.

Por tanto, una vez efectuada esta comprobación, lo que podemos hacer, finalmente, es estimar un modelo en que permanezca σ_{ef} pero que no incluya σ^2 en el nivel 1. En ese caso, la variación total del nivel 1 será una función de este tipo:

$$\text{Var}(e_{ij} + f_{ij}) = \sigma_e^2 + 2 \sigma_{ef} x_{ij} = \sigma_e^2 + 2 \sigma_{ef} (\text{VALORA91})_{ij} \quad (1.15)$$

Ahora, para estimar el modelo, sacamos el término asociado a VALORA1 del nivel 1 mediante la orden CLRE y estimamos de nuevo, obteniendo los siguientes resultados:

CLRE 1 'VALORA91' 'VALORA91'			
NEXT			
Iteration number 6 in progress			
Iteration number 6 completed			
Convergence achieved			
FIXED			
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE
CONS	30.56	0.3522	30.56
VALORA91	0.6171	0.03269	0.6169

RANDOM				
LEVEL 3				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
LEVEL 2				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
CONS /CONS	4.137	1.203	4.143	3
VALORA91 /CONS	-0.236	0.09463	-0.2348	2
VALORA91 /VALORA91	0.01493	0.009551	0.01499	2
LEVEL 1				
PARAMETER	ESTIMATE	S. ERROR	PREV. ESTIMATE	NCONV
CONS /CONS	27.7	1.423	27.66	3
VALORA91 /CONS	-1.205	0.0784	-1.208	4

```

LIKE B40
-2*log(lh) is      5375.42

```

Valor que mantiene la diferencia que hace apropiada la inclusión de la covarianza entre e y f en el nivel 1 y que confirma que la varianza σ^2 no debe estar en el mismo.

La estimación de σ_{ef} (-1.205) es varias veces su error estándar (0.0784) y es significativa en la notable reducción del ratio de máxima verosimilitud que se ha logrado desde el primer modelo planteado hasta llegar a éste último.

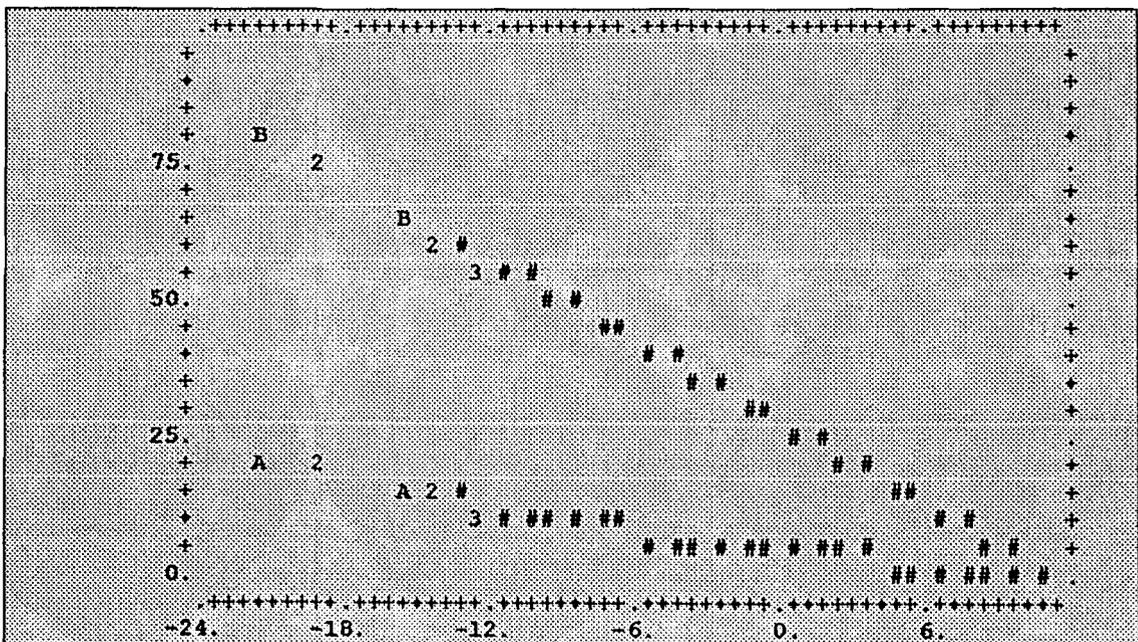
Seguidamente, podemos calcular y representar las funciones de la varianza total del nivel de las secciones y del nivel de los electores mediante los siguientes comandos:

```

CALC 'TOT_VAR2' = 4.137 - 2*0.236*'X' + 0.01493*'X'^2
CALC C26 = 27.7 - 2*1.205*'X'
NAME C26 'TOT_VAR1'
MPLOT 'TOT_VAR2' 'TOT_VAR1' 'X'

```

La varianza del nivel 2 es una función cuadrática (TOT_VAR2) y la del nivel 1 (C26 = TOT_VAR1) una línea recta.



Por tanto, la primera línea que aparece en el gráfico muestra cómo la varianza del nivel 1 (las secciones censales) varía (diminuyendo) a medida que cambian los valores de VALORA91 representada en el eje horizontal: los individuos de la población son más homogéneos a medida que los valores de los porcentajes de VALORA91 aumentan.

La segunda función, de tipo cuadrático, como ya hemos dicho, muestra que las secciones son más homogéneas cuando tienen electores que tienden a puntuar alto el porcentaje de promesas cumplidas.

2.7 CONCLUSIONES ACERCA DE LAS EXPERIENCIAS

Cuando se utilizan las técnicas convencionales de la regresión, el esfuerzo del investigador suele concentrarse en la parte fija del modelo que hemos presentado y sólo se permite que una variable aleatoria describa la variación residual, asumiéndose, además, que tiene varianza constante. Las técnicas multilevel, nos permiten modelizar la estructura de la variación residual, es decir, la que no es capaz de recoger la parte fija del modelo y eso representa una gran novedad respecto al análisis clásico y la posibilidad de obtener una explicación mayor del fenómeno que estamos estudiando.

Sin embargo, en este capítulo no se ha hecho más que introducir el tema de la modelización multilevel, ya que sus posibilidades son tan enormes que proporcionan material suficiente no para una, sino para muchas tesis.

Por ejemplo, a partir del punto en que hemos dejado la última experiencia, podríamos complicar más el modelo con otras posibilidades de estructurar la varianza de los residuos del nivel 1. Bastaría con considerar que dicha varianza fuese una función del género de los electores. Modelizar la varianza en función de una variable categórica en lugar de continua introduce nuevos retos que el análisis multilevel ya ha resuelto. Es decir, que podríamos llevar a cabo la experiencia puesto que disponemos de los datos de la variable género en nuestra base. El único cambio estaría en que la variable categórica impondría un límite sobre el número de parámetros aleatorios que podrían asociarse a dicha variable en el nivel en que fuese definida, pero la estimación de un modelo así ya es posible.

La dilatada exposición efectuada hasta el momento en los cinco capítulos y la gran diversidad de temas tratados imponen el poner un punto final a la reflexión. En caso contrario, éste capítulo debería convertirse en un libro de texto acerca de las técnicas multilevel y desarrollar todas y cada una de sus principales aplicaciones. Sin embargo, ya se ha dicho al principio que no es esa la pretensión de este trabajo y, no obstante, este último capítulo ya comenzaba a adolecer de un estilo de manual porque no hay referencias ni artículos elaborados en España con estas técnicas aplicadas a datos de Comportamiento Electoral.

En ese sentido, a la doctorando, no le ha quedado más remedio que comenzar la explicación por el principio y dejarla en el punto en que piensa que el lector ya ha podido iniciarse en el tema y ver su utilidad y afinidad con su campo de investigación.

El desarrollar más estas técnicas a nivel de esta tesis implicaría entrar en un terreno matemático más complicado (aunque no inasequible para los iniciados en

las técnicas de regresión de todo tipo) y un cambio de estilo y de objetivos que en ningún momento se ha deseado alcanzar.

Por tanto, lo único que se va a añadir a lo expuesto es que los modelos entran en el terreno inferencial y que permiten llevar a cabo investigaciones científicas cuantitativas en el más amplio sentido de estos términos, siempre que se planifiquen bien todas las etapas de los estudios y se siga el método científico. El alcanzar resultados favorables o desfavorables con respecto a nuestras hipótesis no es lo que más debe preocuparnos.

Como último comentario técnico, exponer que se ha repetido varias veces que estas técnicas son inferenciales y que, por tanto, su objetivo último debe ser inferir resultados de muestras a poblaciones. Sin embargo, en un tema como el que hemos escogido, puede suceder que exista alguna o algunas secciones censales de la muestra que tengan un interés particular. Por eso es interesante saber ilustrar gráficamente la variación existente entre las secciones y observar si en el gráfico destaca alguna en especial.

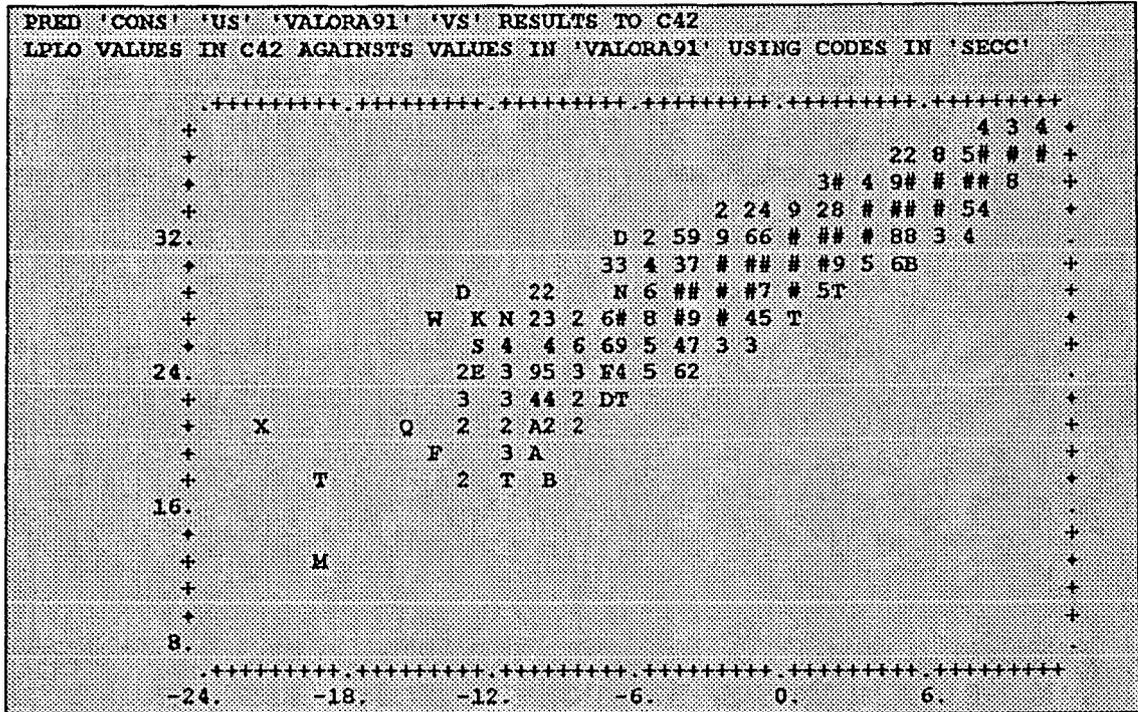
Para lograr esta representación, se puede desarrollar el siguiente programa:

```
REST FOR 2 LEVEL 2 VARI ASSOC WITH 'CONS' AND 'VALORA91' RESULTS TO  
C40 C41  
NAME C40 'US' C41 'VS'  
PRINT C40 C41
```

En primer lugar se envían los residuos del nivel 2 a unas variables que llamamos C40 y C41 y a las cuales ponemos el nombre de US y VS. Seguidamente, listamos estas variables, que por ser del nivel 2 tienen 48 casos:

N =	US	VS
1	-2.8187	0.13549
2	-0.27624	0.0075319
3	1.3742	-0.073651
4	-2.0015	0.12408
5	-0.097915	0.0057500
6	0.72316	-0.039570
7	1.7642	-0.095676
8	0.50033	-0.025702
9	-2.3269	0.15168
10	0.77469	-0.040519
11	-0.33270	0.028469
12	-0.47165	0.016166
13	-0.77040	0.044075
14	0.24001	-0.025137
15	-0.65014	0.040816
16	0.81034	-0.038289
17	-0.45487	0.029585
18	0.92224	-0.062448
19	-0.58937	0.032106
20	-3.5753	0.18816
21	-1.5657	0.094242
22	-0.34985	0.031317
23	2.6071	-0.16200
24	1.6756	-0.10352
25	0.90557	-0.052626
26	0.61362	-0.031062
27	-1.8649	0.11070
28	-4.5444	0.24454
29	0.85664	-0.051422
30	3.4963	-0.18935
31	0.31553	-0.019180
32	-1.5296	0.070295
33	2.7599	-0.15610
34	-0.45855	0.038608
35	3.2523	-0.18010
36	0.45457	-0.027927
37	2.7570	-0.16223
38	-1.5133	0.080878
39	-2.5122	0.14088
40	1.9085	-0.11394
41	-0.39479	0.020578
42	-0.25014	0.015297
43	-2.2161	0.14228
44	0.46854	-0.026000
45	1.5909	-0.089310
46	-0.59695	0.043967
47	0.44860	-0.0080859
48	0.94242	-0.063653

Por filas tenemos el punto de corte y la pendiente de los residuos de la primera sección censal, en la segunda fila los de la segunda sección censal y así sucesivamente hasta la sección 48. Por tanto, la columna US contiene puntos de corte y la VS pendientes. A partir de estos datos y de los residuos almacenados, podemos representar gráficamente las rectas de regresión de las 48 secciones censales. Desafortunadamente, esta versión del programa todavía proporciona gráficos de baja resolución, aunque ahora, con la nueva versión para Windows este problema ya está superado.



Es difícil distinguir aquí 48 líneas de regresión, pero están y muchos de los valores se solapan. En el eje vertical tenemos los valores de las predicciones de cada elector (VALORA95*) almacenadas en la variable C42 y en el eje horizontal tenemos la variable VALORA91 centrada. De esta forma, el gráfico muestra la relación entre las valoraciones del 95 y del 91 para una muestra de 48 secciones censales tal y como ha sido estimada por la aplicación ML3.

Este gráfico no se parecería en nada a uno en que hubiésemos hecho una regresión para los datos de cada sección por separado y luego hubiésemos representado las líneas en un solo gráfico. Emplear regresiones separadas es ineficiente ya que en términos técnicos, las estimaciones multilevel son predicciones de los residuos de las secciones censales y, como para efectuar estas estimaciones se usan datos de todos los electores y secciones al mismo tiempo, estas estimaciones serán siempre más precisas que las obtenidas mediante regresiones separadas. Además, esto resulta una importante ventaja cuando hay secciones en que se dispone de poca muestra de electores.

CONCLUSIONES

Al dar por finalizado el trabajo de exposición de los diferentes temas tratados a través de los cinco capítulos de que consta esta tesis, es necesario efectuar un balance que sea capaz de reflejar las metas conseguidas de aquellos objetivos que se fijaron en un principio y las conclusiones que es posible extraer de todo ello.

En cuanto a la forma de la tesis, una vez leída, se tiene la impresión de que es un trabajo dividido en dos grandes partes: una de preparación, documentación y discusión y otra de presentación y apertura de nuevas posibilidades de aplicaciones estadísticas prácticas dentro de la parcela fundamental acerca de la cual se ha centrado el estudio: el Comportamiento Electoral en España.

De la primera parte, se puede concluir que la Ciencia Política existe como tal y que es una especialidad dentro de las Ciencias Sociales en general.

Dentro de esta disciplina existe una parcela a la que se puede llamar Estudio de Resultados Electorales y de Comportamiento Electoral, pero hay que matizar que tanto la Ciencia Política como esta parcela, son materias que comparten espacios con otras ramas de las Ciencias Sociales siendo a veces invadidas y otras invasoras. La disciplina de la cual se hallan más cerca es la Sociología, Ciencia que aparece posteriormente a la Ciencia Política pero que es pionera en el establecimiento de la llamada Sociología Electoral, de la que posteriormente se desarrollará más detalladamente el Estudio de los Resultados y el Comportamiento Electoral desde el punto de vista de la Ciencia Política.

Ha quedado demostrado que los solapamientos y las confusiones son continuos desde la aparición de las Ciencias Sociales en general y que, la Ciencia Política, por ser una de las de más larga tradición, se ha visto más afectada que otras a lo largo del proceso.

Por si fuera poco, la posibilidad de recopilar datos sociales y de analizarlos cuantitativamente vino, por su novedad y rápido desarrollo, a confundir aún más la metodología de estas Ciencias avanzando más rápidamente los estudios aplicados que las reflexiones acerca del planteamiento y la cientificidad de los mismos.

Por tanto, otra conclusión importante a extraer es que la práctica se ha desarrollado más rápidamente que la teoría, produciendo un cierto caos en las disciplinas afectadas por ello.

La relación de la Estadística con las Ciencias Sociales se hace más y más evidente desde la Segunda Guerra Mundial en adelante, pero son pocas las precauciones que se toman ante la avalancha de técnicas y posibilidades, situación que todavía empeora más cuando se produce el gran desarrollo de la informática.

Así, otra conclusión importante a la que se llega tras leer la documentación es que ante tal desorganización ha sido muy natural que se produjesen debates y enfrentamientos entre los que dominaban y los que no dominaban las técnicas cuantitativas, queriendo tener todos la razón y teniéndola todos en parte.

Esa fase, aunque no lo parezca está a punto de finalizar, porque la formación de los nuevos licenciados y diplomados de las universidades a nivel mundial, les proporcionará una preparación que les permitirá ubicarse libremente en la parcela

que deseen y trabajar con uno o diversos enfoques según les parezca más adecuado. Al tener todas las herramientas tanto cuantitativas como cualitativas bajo su conocimiento, la pugna por estar en un bando o en otro por falta de conocimientos carecerá de sentido. De todo ello se espera que la planificación de las investigaciones sea en adelante y poco a poco mucho más coherente de lo que viene siendo desde los años cincuenta hasta los noventa. Probablemente estamos en el inicio de una nueva etapa en este sentido que finalizará con ciertos debates y, seguramente aportará otros nuevos.

Acerca de algunas de las preguntas que se han propuesto al inicio de la tesis, ahora se puede exponer una opinión, como conclusión de lo leído y aprendido.

En primer lugar hemos hecho referencia al tema: ¿Quién debe formular las hipótesis de trabajo en investigaciones sobre comportamiento electoral, científicos ligados a la concepción filosófica, a la científica o a ambas?

A la vista de lo expuesto y de los razonamientos de los muchos autores leídos, no hay duda de que las hipótesis de trabajo sobre Comportamiento Electoral las puede formular cualquier científico social (no sólo politólogo) que esté suficientemente formado en el tema ya siga una concepción cualitativa o cuantitativa de la investigación. Como ya se ha demostrado, una descripción cualitativa organizada según el método científico es tan válida como una inferencia cuantitativa a la cual se llegue mediante el mismo método. Por tanto, no hay restricciones acerca de qué tipo de científico puede formular hipótesis en este terreno.

Sobre la pregunta: ¿Sólo pueden efectuar estudios prácticos los científicos ligados a la visión científica? y, en ese caso, ¿están capacitados para interpretar los resultados o necesitan los conocimientos de los investigadores adscritos a la visión filosófica?

El material recopilado a lo largo del segundo capítulo responde que no hay una visión científica entendida al estilo en que insinúa la pregunta, sino que es científico todo trabajo que se desarrolle siguiendo las etapas del método científico hasta donde sea posible, aportando algo al cuerpo disciplinar, ya sea en sentido positivo o negativo. La interpretación de los resultados puede llevarla a cabo cualquier científico social preparado y formado para ello y, si en la investigación intervienen técnicos estadísticos o de otras disciplinas, su opinión puede ser un buen complemento siempre que tengan una formación de tipo generalista que les permita intervenir con propiedad.

En cuanto a la pregunta: ¿Se solapan los elementos que utilizan ambos tipos de científicos para analizar el comportamiento electoral?, la respuesta es que evidentemente así es. Los solapamientos son naturales y constantes, todos emplean elementos que son comunes de una u otra forma, lo que cambia a veces es el enfoque o el tratamiento de los mismos, pero, tal y como se ha apuntado antes, la homogeneización de la formación tenderá a disminuir la distancia entre lo que hasta ahora han sido dos concepciones antagónicas.

Por otro lado, al preguntamos: Si ambos tipos de científicos tienden a investigar por separado, ¿cómo va a ser posible establecer una línea metodológica coherente en análisis de Comportamiento Electoral?, la respuesta vuelve a estar en el mismo argumento, como la preparación de los futuros científicos será más homogénea y todos tendrán a su alcance las diversas técnicas, no habrá grandes problemas en cuanto a las líneas metodológicas a seguir. La base de todo estará en que no se olvide que la línea conductora es el método científico.

Asimismo, acerca de la pregunta: ¿Es suficiente el concurso de un científico político sea partidario de la concepción filosófica o de la científica para llevar a cabo un análisis electoral o hay que admitir la participación de científicos relacionados con otras disciplinas?, la tendencia que se está defendiendo es la multidisciplinar. Un buen equipo de trabajo acerca de estos temas puede estar formado por personas procedentes de la Ciencia Política, la Sociología, la Estadística, la Informática y también la Economía, el Derecho, la Antropología y la Psicología. Cuanto más participativa sea la persona que lleva a cabo una investigación sobre Comportamiento Electoral mejor y, si decide llevar a cabo la investigación en solitario, debe estar abierto a poder consultar a especialistas en cualquiera de las disciplinas mencionadas cuando ello sea necesario.

Acerca de si: ¿Es posible llegar a establecer un esquema de grupo de trabajo que unifique las dos tendencias para las investigaciones relacionadas con los temas electorales, en que claramente se definan los actores y las tareas y técnicas a emplear por cada uno de ellos?

La conclusión es que se podrían formar equipos de esta índole, pero no es estrictamente necesario y lo cierto es que o bien los investigadores individuales hallan su propia línea, o los equipos se van formando poco a poco de forma natural a medida que se van intercambiando experiencias y conocimientos para desarrollar los trabajos. Hasta ahora ha sido bastante típico que los investigadores llamados "cualitativos o filosóficos" trabajasen en solitario o con otros como ellos y que los llamados "científicos o cuantitativistas" trabajasen formando equipo con algún estadístico, pero esto en la actualidad no tiene porqué ser así.

Por lo que respecta a la relación de la Estadística con el Estudio de los Resultados y el Comportamiento Electoral, se extraen muchas conclusiones que, en el fondo no son más que el reflejo de lo que ya viene sucediendo desde hace muchos años.

En primer lugar, la falta de formación de la gente a la que van destinadas las informaciones generales sobre resultados y encuestas electorales propician el uso de las técnicas más sencillas, quedando ocultas para el gran público las investigaciones que se pueden realizar a nivel muy sofisticado con los datos que actualmente se pueden recoger.

De ello se ha derivado un cierto estancamiento en los investigadores no estadísticos y en ese sentido, no están aprovechando, ni enseñando a sus alumnos, el potencial de que se dispone en estos momentos.

A la vista de los estudios recientes, las novedades se plasman en cientos y cientos de artículos que pasan desapercibidos en las revistas especializadas que sólo se leen en círculos muy cerrados. Por tanto, no es que no se usen las nuevas técnicas, es que no se populariza y generaliza su uso entre los investigadores que no tienen un especial interés por llegar más lejos en su dominio de la estadística como herramienta de trabajo.

Por eso, en esta tesis se han escogido tres parcelas particulares en las que la doctorando piensa que se han quedado estancados los politólogos y a partir de las cuales pueden ampliar sus aplicaciones.

Como los datos de Resultados y Comportamiento Electoral se dividen en agregados y desagregados, para los primeros se ha recomendado encarecidamente el empleo de técnicas exploratorias ya que, aunque se sigan mapificando resultados y elaborando tablas de todo tipo, en esa parcela hallarán formas nuevas y originales de resumir los

datos y, lo que es más importante, una nueva valoración de las posibilidades cualitativas como paso previo a lanzarse en un terreno inferencial sin referencias.

Para los datos desagregados se ha recomendado el uso apropiado del Análisis de Componentes principales, tan utilizado y confundido con el Factorial, insistiendo en que ahora está al alcance de todos el Análisis de Componentes Principales no lineal y todas sus derivaciones que permiten trabajar con variables categóricas. En los ejemplos ya se ha demostrado que aunque el análisis clásico proporciona resultados aparentemente satisfactorios con todo tipo de variables, en realidad, cuando se conocen y aplican las nuevas técnicas, se gana en calidad de resultados e interpretación.

Finalmente, se ha efectuado una incursión en el terreno más novedoso: el análisis conjunto de datos agregados y desagregados mediante los análisis Multilevel. La enorme extensión de esta materia y su complejidad matemática sólo permiten una presentación y la realización de unos pocos ejemplos, pero lo visto es sólo la punta del iceberg.

Las técnicas multilevel permiten analizar variables cualitativas y cuantitativas conjuntamente, datos temporales, aplicar desde una regresión simple hasta un modelo logit o cualquier otra derivación de la regresión que se nos ocurra, pero para ello, si que hay que planificar muy bien la investigación y sobre todo la recopilación de los datos.

Si en un estudio van a intervenir diversos niveles, el cuestionario debe contener las casillas pertinentes para rellenar la información de esos niveles y no sólo las del individuo entrevistado. Cuestionarios así, todavía no se han probado en la práctica en España y, trabajar con bases de datos que más o menos permiten una aplicación multilevel es bastante inadecuado. En esta tesis se ha podido comprobar y para poder avanzar más en este tema, realmente apasionante, lo primero sería entrar en el terreno del diseño de un cuestionario apropiado que luego permitiese efectuar una buena experimentación.

Por tanto, se dejan muchas puertas abiertas y muchas sugerencias, especialmente en éste último terreno, que podría proporcionar grandes novedades y nuevas ideas acerca de los factores que pueden explicar el Comportamiento Electoral de diversos tipos de poblaciones en nuestro país.

La tesis expuesta no es una tesis que añada resultados concretos al estudio del Comportamiento Electoral, sino que es una reflexión y su aportación estriba precisamente en la intención de ayudar a la clarificación metodológica de la parcela. En este sentido, por tanto, no se explica en ningún momento cómo se comporta electoralmente ningún colectivo, sino las formas en que se puede enfocar la investigación desde diversos puntos de vista.

En especial, se ha puesto mucho énfasis en considerar la validez de los métodos cualitativos infravalorados durante mucho tiempo frente a los cuantitativos y se ha aportado, desde el punto de vista estadístico, que en esta Ciencia también hay que distinguir entre técnicas descriptivas no inferenciales y técnicas inferenciales. Por tanto, si uno de los pilares en que se fundamenta la investigación del Comportamiento y los Resultados Electorales ya se halla en sí mismo dividido en parcelas cualitativas y cuantitativas, resultaría absurdo pensar que la disciplina en la que se aplican no lo esté también.

Finalmente, la doctorando insiste en que la aportación de esta tesis al cuerpo disciplinario tiene, ante todo un carácter de ordenación, reflexión e investigación de

muchas de las causas que han producido el desorden en el que se mueve la investigación actual. Por tanto, aunque parezca que no es ninguna novedad, la gran conclusión a extraer de todo lo expuesto es que será científico todo aquello que se planifique y se desarrolle de acuerdo con el esquema del método científico y que no importa el estadio que se alcance mientras se siga la lógica unificadora de la inferencia a la que ya se ha hecho alusión. Toda contribución es válida tanto si el resultado es el esperado o el contrario, lo importante no es sólo el contenido, sino la forma en que se llega a su establecimiento.

REFERENCIAS BIBLIOGRÁFICAS

BIBLIOGRAFIA

AUTOR	DESCRIPCIÓN
Alker, Hayward	El uso de la matemática en el análisis político Ammorortu editores, 1969
Almond Gabriel ¹	A Discipline Divided: Schools and Sects in Political Science 1989, Sage Publications.
Anderson, C., Statecraft	An Introduction to Political Choice and Judgement, 1977, New York, John Willey
Barranco, F.J.	Técnicas de márketing político Pirámide
Brecht, A.	Teoría Política
Bryk, A.S., Raudenbush, S.W.	Hierarchical Linear Models Advanced Quantitative Techniques in the Social Sciences Series, 1 Sage Publications, 1992
Buttolph Johnson Janet y Richard A. Joslyn	Political Science Research Methods C.Q. Press, 1991, Washington
Caplow, T.	La investigación sociológica Editorial Laia, Barcelona, 1977
Cleveland, W.S.	Visualizing Data Murray Hill, AT&T Bell Laboratories/Hobart Press, 1993
Corbetta, P., Parisi, A.M., Schadee, H.M.A.	Elezioni in Italia Struttura e Tipologia delle Consultazioni Politiche Il Mulino, 1988
Erikson, B.H., Nosanchuk, T.A.	Understanding data: an introduction to exploratory and confirmatory data analysis for students in the social sciences. Milton Keynes, Open University Press, 1977, segunda edición 1992
Estudis Electorals	A cura de l'Equip de Sociologia Electoral (U.A.B.) Volums de l'1 al 10. Publicacions de la Fundació Jaume Bofill.

AUTOR	DESCRIPCIÓN
Fisichella, Domenico	Método Científico e Investigación Política NIS: La Nuova Italia Scientifica, 1985, Roma
Gablentz, Otto Heinrich von der	Introducción a la Ciencia Política Biblioteca Herder 1970, Sección de Ciencias Sociales
Gamble, A.,	Theories of British Politics, 1990a, Political Studies, XXXVIII, 3.
García Ferrando, Ibañez y Alvira	El análisis de la realidad social: métodos y técnicas de investigación Añianza Universidad Textos
Gettell, R.G	Introduction to political science, 1910.
Gettell, R.G.	Nature and Scope of present political theory, 1914
Gifi, Albert	Princals Department of Data Theory Leiden, 1985
Gifi, Albert (University Leiden, The Netherlands)	Nonlinear Multivariate Analysis John Wiley & Sons, Chichester, England, 1990
Gifi, Albert, Groenen Patrick	Anacor Department of Data Theory University of Leiden, 1991
Goldstein Harvey	Multilevel Models in Educational and Social Research Charles Griffin & Company Limited, London, Oxford University Press, New York, 1987
Gunnell, John G.	Political Theory: The evolution of a sub-field Political Science Edited by Ada W. Finifter 1983, APSA
Hagenaars, J.A.	Categorical longitudinal data Sage publications, New York, 1990
Himmelweit, H.T. et al.	How voters decide Open University Press, 1985

AUTOR	DESCRIPCIÓN
Hoaglin, Mosteller, Tukey	Understanding Robust and Exploratory Data Analysis New York, 1983, Wiley
Hoaglin, Mosteller, Tukey	Exploring data tables, trends and shapes Wiley, 1985
Judge,D., H. Wolman y G. Stoker,	Theories of Urban Politics, 1995, London: Sage.
Kim Jae-On, Nie Norman, Verba Sidney	A note on factor analyzing dichotomus variables: the case of political participation Political Methodology 1977
Kinder, D., Iyengar, D.R.	News that Matters The University of Chicago Press Chicago, 1987
King, G.	Unifying Political Methodology The Likelihood Theory of Statistical Inference Cambridge University Press, 1989
King, G., Keohane, R.O., Verba, S.	Designing Social Inquiry: Scientific Inference investigación Qualitative Research Princeton University Press, 1994
Kish, L.	Statistical design for research John Wiley New York, 1987
Krupp	La estructura de la Ciencia Económica Biblioteca de Ciencias Sociales 1973, Ed. Aguilar
Leftwich, C.	What is Politics?, 1984a, Oxford, ed. Basil Blackwell.
Manheim, Jarol B., Richard C. Rich	Análisis político empírico Métodos de investigación en Ciencia Política Alianza Universidad Textos, 1988
Marsh, Catherine	Exploring Data. An introduction to data analysis for Social Scientists. Cambridge, Polity Press, 1988

AUTOR	DESCRIPCIÓN
Marsh, D., Stoker Gerry	Theory and methods in political science Macmillan Press, 1995, London
Mínguez González, Santiago	La abstención electoral en Galicia Universidad de la Coruña
Morgenthaler, S., Tukey J.W.	Configural Polysampling A route to practical robustness Edited by Morgentahler & Tukey, 1991
Mosteller, Frederick, John W. Tukey	Data Analysis and Regression Addison –Wesley Publishing Company, 1977
Niemi, R.G., Weisberg, H.F.	Controversies in voting behavior. Second edition CQ Press, 1984
Nieuwbeerta, Paul, Ultee, W.	Explaining Differences investigación teh level of class voting investigación 20 western industrial nations 1945-1990 University of Nijmegen, The Netherlands, 1992
Ostrom, Elinor.	Beyond Positivism. In Ostrom (Ed.). Strategies of political inquiry. Beverly Hills: Sage, 1982.
Parry, G., Moyser, G, Day, N.	Political participation and Democracy investigación Britain Cambridege University Press Cambridge, 1992
Peña Daniel, Juan Romo	Introducción a la Estadística para las Ciencias Sociales McGrawHill, 1997
Primo Yúfera, Eduardo	Introducción a la Investigación Científica y tecnológica Alianza Universidad
Pulido San Román, A.	Estadística y técnicas de investigación social Pirámide
Raudenbush S. W., J.d. Willms	Schools, Classrooms, and Pupils International Studies of Schooling from a Multilevel Perspective Academic Press INC, Boston 1991

AUTOR	DESCRIPCIÓN
Rosenstone, S.J., Hansen, M.	Mobilization, apticipation and Democracy investigación America Macmillan, New York, 1993
Sartori Giovanni y Morlino Leonardo	La comparación en las ciencias sociales Alianza Universidad, 1994
Sartori, Giovanni	La política: lógica y método en las ciencias sociales, Fondo de cultura económica, México, 1987
Skinner, Quentin	Meaning and Understanding in the History of Ideas, Meaning and context, Ed. James Tully. Oxford: Basil Blackwell, 1989
Spence, I., Lewandowsky	Graphical Perception Modern Methods of Data Analysis Fox/Log (eds.) Newbury Park, Sage, 1990
Stoker, Gerry	Theory and Methods in Political Science, 1995, London, Macmillan Press LTD.
Tanur, Mosteller, Kruskal et al.	La estadística: una guía de lo desconocido Alianza Editorial
The Collected Works of John. W. Tukey	Volúmenes 1-5 Edited by William S. Cleveland 1965-1985
The collected works of John W. Tukey	Pacific Grove & Brooks/Cole, 1949-1983 Tukey, John W. Exploratory Data Analysis Addison-Wesley Publishing Company USA, 1977
Vallespín, Fernando	Historia de la Teoría Política, Vol. 6, pg.10, Madrid 1995, Alianza Editorial.
Van der Burg, Eeke	Canals Department of Data Theory University of Leiden, 1985
Van de Geer, John P.	Multivariate Analysis of Categorical Data: Theory Sage publications, Newbury Park, California 1993

AUTOR	DESCRIPCIÓN
Van de Geer, John P.	Multivariate Analysis of Categorical Data: Applications Sage publications, Newbury Park, California 1993
Van de Geer, John P.	Primals Department of Data Theory University of Leiden, 1985
Van de Geer, John P.	Homals Department of Data Theory University of Leiden, 1985
Velleman, P.F. & Hoaglin, D.C.	ABC of EDA: Applications, Basics and Computing of Exploratory Data Analysis. Boston, Mars, 1981 (Duxbury Press)
Verba, S. Nie, N.H.	Participation in America Political Democracy and Social Equality The University of Chicago Press, 1987
Verdegaal, Renée	Overals Department of Data Theory University of Leiden, 1986
Weisberg, Herbert F.	The Science of Politics and Political Change, Political Science, The Science of Politics, Edited by Herbert F. Weisberg, Agathon Press, Inc. New York, 1983.
Zuckerman,	Doing Political Science, 1991, Boulder, Col: Westview.

REVISTAS, ARTÍCULOS, CONFERENCIAS

REVISTA Y/O AUTOR	REFERENCIA
European Political data	Publicada conjuntamente por el European Consortium for Political Research y el Norwegian Social Science Data Services
European Journal of Political Science	
European Journal of Political Research	
American Political Science Review	
REIS (CIS)	
WORKING PAPERS (ICPS)	
Sondeig d'opinió Catalunya 1995	ICPS
Font, Virós, Rainer, Schultze, Corbetta, Parisi, Denters, Denver y Justel	Electoral Abstention in Europe ICPS, 1995
Quaderni dell' Osservatorio Elettorale	A cura della Giunta Regionale, dell' IRPET e del gruppo di studio sul comportamento elettorale investigación Toscana Firenze
Achen, C.H.	Toward Theories of Data: The State of Political Methodology Political Science, The State of the Discipline Ed. Ada W. Finifter, APSA, 1983
Carroll, John B.	The Nature of tha Data, or How to Choose a Correlation Coefficient Psychometrika 26: 347-72, 1961
Christofferson, A.	Factor Analysis of Dichotomized Variables Psychometrika 40: 5-32, 1975
Deutsch, Karl.	On political theory and political action. APSR, 1971, 65, pgs. 11-27.
Deutsch y Rieselbach.	Recent trends in political theory and political philosophy. Annals of the American Academy of Political and Social Science, 1965, 360, pgs. 139-62.
Echevarría Zabalza, Javier	Movilidad Social y Comportamiento Electoral Comunicación presentada en el V Congreso Español de Sociología, Granada 1995

REVISTA Y/O AUTOR	REFERENCIA
ESE	Notes per a una geografia electoral de Catalunya Perspectiva Social, 1978
Horber, Eugène	Exploring Aggregate Data European Political Data Newsletter, September 1987
Laitin, David D.	Disciplining Political Science APSR, vol 89, Nº 2, junio 1995
Hosmer, D.W., Lemeshow, Stanley Jr.	Applied logistic regression John Wiley & Sons, 1989
Jones, K., Johnston, R.J., Pattie, C.J.	People, Places and regions: Exploring the use of Multilevel Modelling in the analysis of electoral data, 1991
Joop, J., Hox, Ita, Kreft, G.G.	Multilevel Analysis Methods Sociological Methods & Research Sage Publications, 1994
Kim, J.O., Nie, N., Verba, S.	A note on factor analyzing dichotomus variables: the case os political participation Political Methodology, 1977
Laitin, David D.	The Qualitative-Quantitative Disputation: Gary King, Robert O. Kehoane, and Sidney Verba's designing social inquiry: scientific inference in qualitative research. APSR, Vol 89 Nº 2, 1995
Martínez de Luna Pérez de Arriba, Iñaki	La participación política en el País Vasco Servicio Central de Publicaciones del Gobierno Vasco, 1993
Merriam, Charles	The present state of the study of politics American Political Science Review, 1921, 15, pgs. 173-185.
Merriam, Charles	Report of the national conference on the science of politics, APSA, 1924, 18, 119-166.
Miller, Warren E.	The role of research in the unification of a discipline. APSR, 1981, 75, pgs. 9-16.
Montero, J.R., Pallarés, Francesc	Los estudios electorales en España: un balance bibliográfico ICP, WP Nº 49, Barcelona 1992

REVISTA Y/O AUTOR	REFERENCIA
Lynn, Naomi B	Self Portrait: Profile of Political Scientists, Political Science, ed. Finifter, 1983, APSA.
Lasswell, H.	A pre-view of policy sciences. New York: American Elsevier, 1971. pg. 5.
Mellenbergh, G.J.	European dissertation research in the methodology of social and behavioral sciences IOPS, May 1991
Padró Solanet, A. Colomer, J.M.	Espacio político – ideológico y Temas de campaña Revista de Estudios Políticos (Nueva Época) Nº 78, Octubre, Diciembre de 1992
Riba i Romeva, Clara	Dimensiones subyacentes en el debate político. Los electores ideológicamente indiferentes. Comunicación presentada en el grupo de trabajo sobre Comportamiento Político y Sistemas Electorales del II Congreso de la Asociación Española de Ciencia Política y de la Administración. Santiago de Compostela, 1996.
Ruiz Olabuénaga, J.I.	Alternativas metodológicas en la investigación política Catedrático de la Universidad de Deusto-Bilbao Conferencia en el área de Ciencia Política y de la Administración de la UPF en marzo de 1994.
Shalala Donna E.	"Politics and the uses of social science research", Political Science, The State of the discipline, editado por Ada W. Finifter, 1983, APSA.
Sociological Methodology	San Francisco Quantitative Applications investigación the Social Sciences Beverly Hills
Galston, William,	Political Theory in the 1980s, Political Science: The state of the discipline II, ed. Ada W. Finifter APSA, 1993.
Salvador Eugènia	Modernas tendencias de la sociología electoral
Stokes, Donald E	Basic Inquiry and applied use in the social sciences. APSA. 1983.

REVISTA Y/O AUTOR	REFERENCIA
-------------------	------------

Virós i Galtier M^a Rosa

Algunes reflexions sobre el comportament electoral a Catalunya
 Conferència donada el 10-VI-1983 a l'Associació Catalana de Sociologia

Whyte, W.F.

A Challenge to political scientists. APSR, 1943, 37, pgs. 692-97.

MANUALES DE APLICACIONES INFORMÁTICAS

APLICACIÓN INFORMÁTICA	REFERENCIA
EDA User's Manual	Horber, Eugène Département de science politique Université de Genève, 1991
SPSS Categories	SPSS Inc. USA, 1990
ML3	Software for Three-level Analysis User's guide for v.2 R. Prosser, J. Rabsbash, H. Goldstein Institute of education, University of London
A guide to ML3 for New Users	G. Woodhouse, J. Rasbash, H. Goldstein, M. Yang Institute of Education, University of London, 1992
Data Analysis with ML3	R. Prosser, J. Rabash, H. Goldstein Institutue of Education, University of London, 1991