

**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

4 ¿QUÉ PUEDEN APORTAR ACTUALMENTE LOS MÉTODOS Y TÉCNICAS ESTADÍSTICAS APLICADAS AL ESTUDIO DE DATOS AGREGADOS?

4.1 LAS NUEVAS POSIBILIDADES

En primer lugar, es inevitable comentar que, el fondo de la metodología estadística es prácticamente el mismo, pero que el avance de la informática y la ampliación de las aplicaciones estadísticas está poniendo en manos de los investigadores herramientas que les permiten desarrollar puntos como los ya establecidos, pero de forma mucho más rápida y eficaz.

Todo ello, proporciona un beneficio: el investigador puede profundizar más en sus estudios porque puede responder y plantear muchas más preguntas en menos tiempo que antes y además, ahorra muchas energías en cuanto a la presentación de resultados y a su interpretación. Pero, también tiene el inconveniente de que ante tanta diversidad puede perderse fácilmente o recurrir a lo que ya conoce.

Este es uno de los principales motivos por los que se ha abordado esta Tesis. Si bien no puede organizar la Ciencia Política, ni dar una idea exacta de la ubicación de los estudios de Comportamiento Electoral, sí que trata de orientar al máximo tanto a investigadores de esta parcela como a estadísticos que tengan que colaborar con ellos en cuanto a las nuevas posibilidades técnicas y sus campos de aplicación.

4.2 EL ANÁLISIS EXPLORATORIO

4.2.1 INTRODUCCIÓN

Dentro del panorama actual del tratamiento de datos agregados, nunca se le ha dado en España, la importancia que tienen a las herramientas relacionadas directamente con el llamado "Análisis Exploratorio de datos" cuyo principal desarrollo se debe a John Tukey y que se remonta en sus inicios a la época de la segunda guerra mundial, momento en que éste gran científico comienza a introducirse en el terreno de la Estadística.

Esta parcela, que algunos consideran que incluso está fuera de la Estadística, resulta de gran utilidad cuando se trata de llevar a cabo la primera fase de un estudio con datos agregados.

La exploración es una filosofía de observación de los datos, una actitud hacia los problemas metodológicos.

Tukey defiende con energía en muchos de sus escritos la importancia de la simple observación frente a los análisis de tipo confirmatorio que pasan por encima, en muchas ocasiones, de la detección de estructuras simples y de datos muy sencillos pero informativos subyacentes en la información de base. Un ejemplo de su forma de expresarse acerca de estos temas puede ilustrar su persecución de la sencillez y la simplificación: "Érase una vez que los estadísticos únicamente exploraban. Entonces aprendieron a confirmar exactamente, a confirmar unas pocas cosas exactamente, cada una bajo unas muy específicas circunstancias"¹⁸⁸.

¹⁸⁸ John Tukey, The collected works, 1965-1985, Edited by William S. Cleveland.

Se puede afirmar que, en parte, la filosofía del análisis exploratorio surge cuando el estadístico "olvida" sus orígenes y se deja llevar por la facilidad que supone en la actualidad realizar cualquier tipo de análisis, por sofisticado que sea, con el concurso de un ordenador y una aplicación estadística.

Tukey nunca se opuso al progreso y al avance de la Estadística, pero insistió mucho en que el estadístico debe actuar con cautela. Antes de aplicar técnicas, digamos de alto nivel, hay que conocer los datos, hay que explorarlos y hay que extraer de ellos toda la información que se pueda.

El resultado de la exploración permite, entre otras cosas, saber si es adecuado continuar en dirección ascendente o si, por el contrario, nos enfrentamos a unos datos que poco pueden proporcionar en cuanto a resultados consistentes.

La exploración se basa en estadísticos robustos, gráficos dinámicos y todo tipo de herramientas sencillas que permiten: destacar estructuras, conocer sus características, ver las necesidades metodológicas de cada caso etc.

El explorador actúa como un "detective" y no se enfrenta a una contrastación de hipótesis al no haber formulaciones de este tipo en esta parcela de la Estadística.

La Estadística clásica, en cambio, trata de fenómenos bien conocidos, de hipótesis y de su confirmación, lo cual hace actuar al estadístico como un "juez". Tukey pensaba que a medida que los estadísticos ponían más énfasis en la confirmación exacta, sus técnicas se hicieron cada vez menos flexibles.

La conexión con la mayoría de las técnicas más utilizadas en el pasado se fue rompiendo. Sin embargo, hay que tener en cuenta que precisamente en el pasado, el coste computacional era enorme y conducía por naturaleza al resumen de los datos, cosa que actualmente se ha hecho innecesaria desde el punto de vista mecánico.

A pesar de todo, desde el punto de vista racional, el resumen sigue siendo necesario. Cualquier técnica que conduzca a la confirmación sin más, no es tan útil como puede parecer a primera vista.

En la realidad, cuando investigamos, nos encontramos en una situación intermedia, es decir, no podemos ser sólo "detectives" o "jueces", sino parte de ambos. La exploración, en este sentido, tiene que suponer la antesala de la confirmación. No es bueno olvidar este aspecto de la Estadística como tampoco sería aconsejable estancarse en él.

Según Tukey, la palabra "estadística" se puede usar de varias formas: como ciencia, como resultado, como análisis de datos, etc. El cómo se "vende" la estadística, cómo la entienden los estadísticos y cuestiones por el estilo son temas delicados a los que hay que prestar atención: la estadística se emplea actualmente a través de los ordenadores y no se suele reflexionar demasiado acerca de las técnicas.

El análisis exploratorio conduce a la descripción y, por tanto, a los estadísticos descriptivos, al análisis multivariable no confirmatorio y a las clasificaciones o análisis cluster. Estos métodos no requieren hipótesis ni condiciones de partida y son insensibles a las violaciones de hipótesis que condicionan una modelización, porque tampoco modelizan y pueden ayudar, en cambio, a determinar las técnicas de tratamiento de alto nivel más adecuadas.

Como ya es sabido en general, los métodos confirmatorios requieren una batería de hipótesis sólidas, son muy sensibles a sus violaciones y su principal objetivo es poder proporcionar un modelo estadístico de comportamiento. La confirmación se basa en la Estadística real y clásica y no tiene en cuenta los datos.

El profesor Eugène Horber del Departamento de Ciencias Políticas de la Universidad de Ginebra creó un programa específico llamado EDA especializado en el análisis exploratorio de datos agregados abarcando aspectos descriptivos y de análisis multivariante, todo él basado en las ideas y planteamientos de Tukey, del que fue alumno. Su utilización resulta muy adecuada para presentar algunas de las nuevas posibilidades de que estamos hablando.

Sin embargo, la mayoría de las aplicaciones estadísticas actuales (SPSS, Minitab, SAS, etc.) proporcionan partes de análisis exploratorio. Las más conocidas: diagramas de tallo y hojas, diagramas de caja y poca cosa más. El EDA, en cambio es una aplicación muy especializada que vale la pena conocer a fondo, porque a pesar de que las ideas de Tukey provienen de los años cuarenta en adelante, nadie las había presentado de forma tan completa y desarrollada como lo ha hecho Horber en esta aplicación.

El marco de trabajo del EDA es el siguiente:

| | |
|---|--|
| CLASIFICACIÓN | |
| Situaciones de investigación Necesidades de detección Ver que herramientas necesitamos | |
| EXPLORACIÓN | CONFIRMACIÓN |
| Detección de estructuras Conocimiento de sus características No hay Hipótesis "Trabajo de detective" | Confirmación de hipótesis El fenómeno es bien conocido Hipótesis consistentes "Trabajo de juez" |

4.2.2 LOS DATOS PARA LLEVAR A CABO ALGUNAS APLICACIONES PRÁCTICAS

Para trabajar la parte práctica de este capítulo, se dispone de los datos electorales de un municipio catalán con 7 distritos electorales y 130 secciones censales.

A partir de estos datos vamos a establecer diferentes tipos de preguntas. Las variables electorales son en este caso: el porcentaje sobre el número total de electores para cada fuerza política relevante (CIU, IC, ERC, PSC, PP) en diferentes elecciones, más la abstención correspondiente a las mismas, más el total de votos emitidos y el total de electores posibles de cada convocatoria electoral. La variable identificadora de los casos es, para las primeras experiencias, la sección censal de cada distrito.

También se dispone de algunas variables de tipo social. En este caso, como se trata de un municipio catalán, es importante tener en cuenta el idioma, de forma que se tiene: el nivel de estudios, los grupos de edad y el nivel de conocimiento del catalán.

Para facilitar la presentación de estas herramientas se va a trabajar no con las 130 secciones generales, sino con las correspondientes a cada distrito. De esta forma, resulta que cada distrito tiene el siguiente número de secciones censales:

| | |
|------------|-----------------------|
| Distrito 1 | 29 secciones censales |
| Distrito 2 | 16 secciones censales |
| Distrito 3 | 23 secciones censales |
| Distrito 4 | 20 secciones censales |
| Distrito 5 | 13 secciones censales |
| Distrito 6 | 23 secciones censales |
| Distrito 7 | 6 secciones censales |

4.2.3 LAS HERRAMIENTAS DEL ANÁLISIS EXPLORATORIO

Las herramientas básicas de la exploración nos ayudan a:

- Poder describir conjuntos de números
- Ampliar nuestra capacidad de hacernos preguntas estadísticas
- Poder descubrir patrones de comportamiento
- Poder descubrir estructuras en los datos

Entre las más conocidas podemos destacar: la codificación, los diagramas de tallo y hojas (stem & leaf), los diagramas de caja (box plots), todo tipo de estadísticos descriptivos, clásicos y "nuevos", rectas de regresión basadas en la mediana y muchas otras más.

A continuación se presentan aplicaciones de todo ello a los datos agregados electorales reales con el objetivo de dar a conocer una metodología ya consagrada en la teoría pero escasamente puesta en práctica, especialmente en España.

4.2.3.1 EL LISTADO Y LOS DIAGRAMAS DE TALLO Y HOJAS (STEM AND LEAF)

A través de la exploración se trata de captar ideas fundamentales acerca de los datos agregados, establecer conceptos, pensar acerca de las herramientas, de su utilidad y construir una perspectiva.

Las preguntas básicas que podemos hacer cuando tenemos ante nosotros una lista de secciones censales y los resultados porcentuales de votos de una fuerza política cualquiera son:

- Valor máximo y mínimo
- Valor típico: moda, mediana, media
- Tendencia general de los números
- Variación alrededor del valor típico
- Excepciones
- Características especiales

Para responder a estas preguntas, lo que podemos hacer en primer lugar es, si los datos originales tienen este aspecto:

(Estamos viendo de izquierda a derecha, las columnas de número de caso, sección censal general de un distrito de un municipio y porcentaje sobre participación de IC

en las municipales de 1995). El distrito es el número 1 y tiene 29 secciones censales.

```
29 cases
variable listing
cases      SECCG PEICM
           EN    95
  1  1      1.00  5.94
  2  2      2.00 17.42
  3  3      3.00 21.05
  4  4      4.00 25.91
  5  5      5.00 17.31
  6  6      6.00 13.75
  7  7      7.00 22.21
  8  8      8.00 20.06
  9  9      9.00 19.04
 10 10     10.00 16.30
 11 11     11.00 21.32
 12 12     12.00 23.38
 13 13     13.00 26.39
 14 14     14.00 24.21
 15 15     15.00 20.68
 16 16     16.00 21.27
 17 17     17.00 24.15
 18 18     18.00 30.16
 19 19     19.00 20.70
 20 20     20.00 32.29
 21 21     21.00 24.80
 22 22     22.00 28.37
 23 23     23.00 31.18
 24 24     24.00 28.22
 25 25     25.00 34.17
 26 26     26.00 28.31
 27 27     27.00 32.30
 28 28     28.00 27.39
 29 29     29.00 29.89
      >>>LIST 1 2
```

Simplificarlos ordenándolos en sentido descendente o ascendente tomando como variable de referencia el porcentaje de votos y redondear quitando los decimales para tener una idea global más simple.

```

29 cases
variable listing
cases SE PE
CC ICM95
25 25 25 34
27 27 27 32
20 20 20 32
23 23 23 31
18 18 18 30
29 29 29 30
22 22 22 28
26 26 26 28
24 24 24 28
28 28 28 27
13 13 13 26
4 4 4 26
21 21 21 25
14 14 14 24
17 17 17 24
12 12 12 23
7 7 7 22
11 11 11 21
16 16 16 21
3 3 3 21
19 19 19 21
15 15 15 21
8 8 8 20
9 9 9 19
2 2 2 17
5 5 5 17
10 10 10 16
6 6 6 14
1 1 1 6

```

```
>>>LIST 1 2 SORT KEY=2 DESCEN DECIMALS=0
```

De esta forma podemos ver rápidamente, el máximo y el mínimo, hacemos una idea del valor típico, la tendencia general de los números, la dispersión, los casos excepcionales y las características especiales si las hay y en qué secciones del distrito sucede todo ello.

En este caso, el máximo alcanzado por IC es un 34% y el mínimo un 6%. El valor típico está alrededor del 25%, hay cierta dispersión, pero ésta se ve afectada sobre todo por el caso más excepcional, el de la sección 1 con un 6% que se diferencia con mucho de los restantes. Y, en cuanto a características especiales, destacar que sólo 6 secciones quedan por debajo del 20%.

La misma información puede obtenerse en orden ascendente si se desea:

```
29 cases
variable listing
cases      SE PE
           CC ICM95
   1  1      1  6
   6  6      6 14
  10 10     10 16
   5  5      5 17
   2  2      2 17
   9  9      9 19
   8  8      8 20
  15 15     15 21
  19 19     19 21
   3  3      3 21
  16 16     16 21
  11 11     11 21
   7  7      7 22
  12 12     12 23
  17 17     17 24
  14 14     14 24
  21 21     21 25
   4  4      4 26
  13 13     13 26
  28 28     28 27
  24 24     24 28
  26 26     26 28
  22 22     22 28
  29 29     29 30
  18 18     18 30
  23 23     23 31
  20 20     20 32
  27 27     27 32
  25 25     25 34
>>>LIST 1 2 SORT KEY=2 DECIM=0
```

En segundo lugar, se recomendaría confeccionar un diagrama de tallo y hojas, cuya principal característica es la de resumir los datos, proporcionar también el máximo y el mínimo, la mediana y los casos especiales, añadiendo el detalle de que nos da una idea de la forma de la distribución.

¿Qué preguntas podemos responder mediante su utilización?. Las mismas que antes, pero añadiendo más información:

- Forma aproximada de la distribución
- Diferentes disposiciones de los datos
- Posibilidad de manipular el gráfico para aclarar dudas
- Relaciones con otras variables

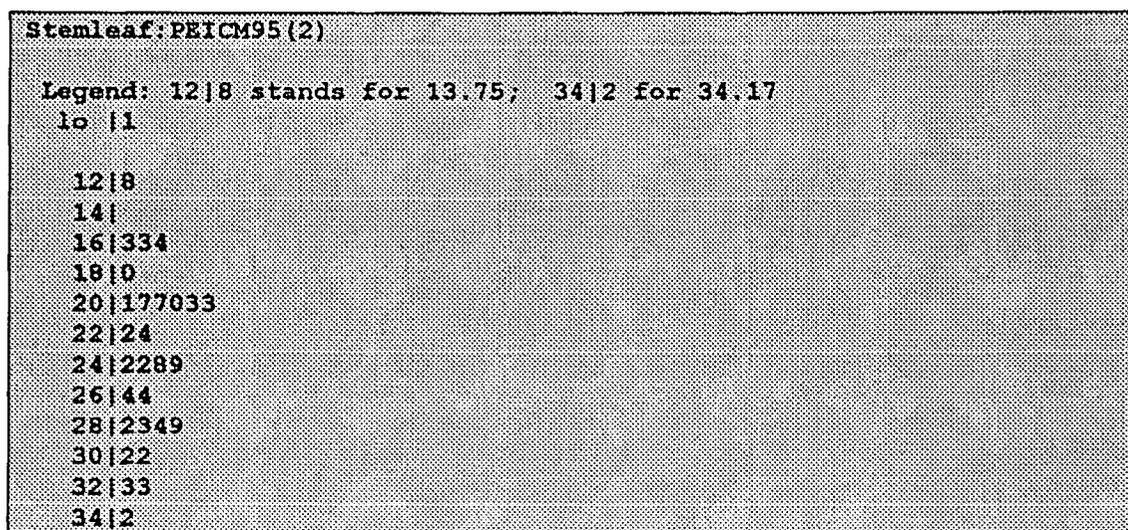
Con este conjunto de datos electorales reales, vamos a tratar de responder a estas preguntas y completaremos la descripción que veníamos haciendo a través del listado. Nos concentraremos en los resultados del distrito 1 para las municipales del 95 del municipio catalán con 7 distritos y 130 secciones censales.

Los siguientes son los resultados más simples que pueden obtenerse a través del diagrama de tallo y hojas. Hay uno para cada fuerza política y la abstención en el distrito 1 del municipio que, como recordaremos tiene 29 secciones censales. El diagrama muestra al principio el nombre y número de la variable con que va a trabajar. En este caso, porcentaje sobre electores de IC en las municipales del 95.

Seguidamente, nos indica que los valores 12|18 y 34|2 corresponden exactamente a 13.75% y 34.17% siendo los valores mínimo y máximo de la distribución general, casos extremos aparte.

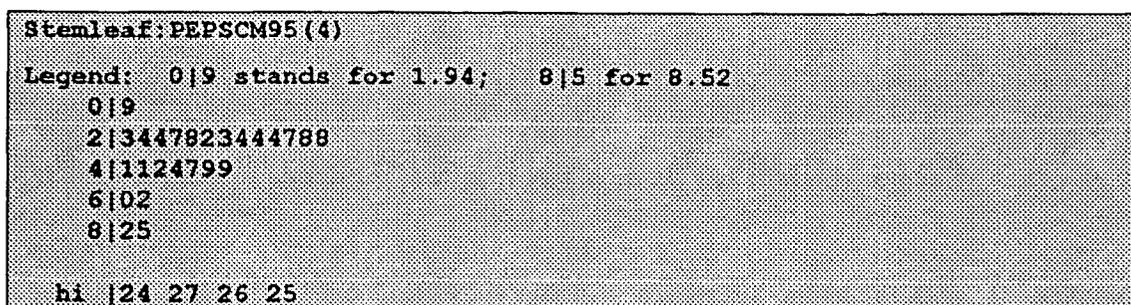
Después aparece la expresión: lo |1. Esto indica que existe un caso extremo por debajo de los mostrados a bastante distancia (si miramos los datos originales al final, veremos que se trata de la sección 1 de este distrito en que IC obtuvo aproximadamente un 5% de votos sobre el total del electorado).

Por tanto, dejando aparte este caso extremo, el gráfico nos muestra la forma de la distribución de los % obtenidos por IC. Como vemos se trata de una distribución algo irregular con un pico que destaca en el 20. De forma que 6 secciones censales rondaron el 20% de votos sobre electores, mientras que sólo 5 se aproximaron a la mediana de IC en el municipio que era 30.1156%.



La distribución general del PSC es mucho más concentrada y se mueve entre 1.94% y 8.53%. De forma que en este distrito no supuso un rival para IC, El comportamiento más habitual de las secciones está en torno al 2 y al 3%.

Sin embargo, existen cuatro secciones que destacan como valores extremos por encima del máximo general. Se trata de las secciones 24, 27, 26 y 25, que es donde el PSC obtuvo mejores resultados.



La distribución de CIU oscila entre 4.56% y 35.09% es un poco más regular que la de IC pero también más concentrada. Los porcentajes más habituales están alrededor del 25-30%, aunque también destacan secciones entre el 10 y el 17%.

```
Stemleaf: PECIUM95 (6)
Legend: 0|5 stands for 4.56; 3|5 for 35.09
0|
0|59
1|002234
1|556677
2|4444
2|5556789
3|023
3|5
```

La distribución del PP oscila entre el 2.67% y el 10.28% con un caso extremo que corresponde a la sección 6 que quedó por encima de este máximo. los porcentajes más habituales están entre el 4 y el 8%.

```
Stemleaf: PEPPM95 (8)
Legend: 2|7 stands for 2.67; 10|3 for 10.28
2|77
3|48
4|255788
5|467
6|004
7|2234678
8|3
9|001
10|3
hi |6
```

La distribución de ERC oscila entre 1.21 y 7.23%, no habiendo casos extremos y siendo lo más habitual, secciones entre el 3 y el 6% de votos sobre electores.

```
Stemleaf: PEERC95 (10)
Legend: 1|2 stands for 1.21; 7|2 for 7.23
1|2
2|1234
3|133779
4|0157
5|013456777
6|023
7|12
```

Finalmente, la abstención del distrito fue alta en casi todas las secciones y partió de un 30.63% llegando al 43.48% y con un caso extremo que sobrepasó este máximo y que fue la sección 1 del distrito.

Stemleaf: PABSM95 (73)

Legend: 30|6 stands for 30.63; 42|5 for 43.48

30|6
32|06928
34|45768
36|11415
38|11119
40|0123
42|7145

hi |1

Seguidamente, se facilitan los resultados originales para poder efectuar consultas y comprobaciones.

variable listing

| cases | PEICM 95 | PEPSC M95 | PECIU M95 | PEPPM 95 | PEER CM95 | PABSM 95 |
|-------|-------------|--------------|--------------|-------------|--------------|-------------|
| 1 1 | 5.94 | 2.33 | 16.64 | 2.67 | 2.43 | 69.54 |
| 2 2 | 17.42 | 2.41 | 30.37 | 7.18 | 6.00 | 34.73 |
| 3 3 | 21.05 | 4.07 | 33.37 | 2.67 | 5.47 | 31.98 |
| 4 4 | 25.91 | 3.17 | 24.13 | 7.39 | 3.33 | 34.52 |
| 5 5 | 17.31 | 2.68 | 31.71 | 7.32 | 5.66 | 33.75 |
| 6 6 | 13.75 | 2.41 | 35.09 | 10.71 | 5.45 | 30.63 |
| 7 7 | 22.21 | 3.76 | 24.31 | 9.06 | 5.30 | 32.60 |
| 8 8 | 20.06 | 4.94 | 25.30 | 10.28 | 4.55 | 32.91 |
| 9 9 | 19.04 | 1.94 | 25.33 | 8.31 | 6.22 | 36.13 |
| 10 10 | 16.30 | 2.80 | 28.57 | 7.76 | 5.75 | 37.11 |
| 11 11 | 21.32 | 3.41 | 25.05 | 4.84 | 5.71 | 37.47 |
| 12 12 | 23.38 | 3.73 | 27.36 | 4.47 | 6.30 | 33.23 |
| 13 13 | 26.39 | 4.93 | 16.27 | 7.21 | 5.10 | 39.14 |
| 14 14 | 24.21 | 3.44 | 24.14 | 4.17 | 7.10 | 35.55 |
| 15 15 | 20.68 | 3.30 | 25.55 | 4.72 | 7.23 | 36.40 |
| 16 16 | 21.27 | 3.38 | 27.55 | 7.69 | 4.00 | 34.38 |
| 17 17 | 24.15 | 4.19 | 17.15 | 6.44 | 5.56 | 40.98 |
| 18 18 | 30.16 | 4.35 | 12.12 | 6.01 | 3.32 | 43.42 |
| 19 19 | 20.70 | 4.71 | 24.44 | 8.97 | 3.89 | 36.10 |
| 20 20 | 32.29 | 3.85 | 12.88 | 5.64 | 4.65 | 39.89 |
| 21 21 | 24.80 | 8.23 | 15.87 | 9.04 | 4.98 | 35.81 |
| 22 22 | 28.37 | 7.17 | 13.96 | 3.78 | 3.08 | 43.10 |
| 23 23 | 31.18 | 8.52 | 8.14 | 6.01 | 2.13 | 42.74 |
| 24 24 | 28.22 | 8.66 | 10.29 | 5.75 | 2.32 | 43.48 |
| 25 25 | 34.17 | 14.12 | 4.56 | 3.42 | 1.21 | 41.15 |
| 26 26 | 28.31 | 9.80 | 12.13 | 7.60 | 2.21 | 39.09 |
| 27 27 | 32.30 | 8.91 | 9.86 | 4.75 | 3.68 | 39.07 |
| 28 28 | 27.39 | 5.96 | 15.02 | 5.45 | 3.68 | 41.09 |
| 29 29 | 29.89 | 4.06 | 14.67 | 4.48 | 4.13 | 41.32 |

>>>list 2 4 6 8 10 73

Evidentemente, el diagrama de tallo y hojas presentado hasta el momento es ya muy conocido y utilizado y no aporta nada nuevo, incluso, el SPSS, el SAS o el minitab, no sólo lo contienen, sino que su presentación es incluso más agradable. Veamos por ejemplo la sección 1 del distrito 1 confeccionada por el SPSS y comparada con el EDA:

```

% SE IC M95
% SE IC M95 Stem-and-Leaf Plot

Frequency      Stem & Leaf

  1,00 Extremes      (<=6)
  1,00          1 .  3
  4,00          1 .  6779
 11,00          2 .  00011123444
  7,00          2 .  5678889
  5,00          3 .  01224

Stem width:      10,00
Each leaf:       1 case(s)

```

```

Stemleaf: PEICM95 (2)
Legend: 12|8 stands for 13.75;  34|2 for 34.17
lo |1

 12|8
 14|
 16|334
 18|0
 20|177033
 22|24
 24|2289
 26|44
 28|2349
 30|22
 32|33
 34|2

```

El SPSS proporciona algunos detalles pero tiende a redondear mucho más los datos que el EDA, de forma que resulta más exacto el segundo diagrama comparado con el primero. Lo que sucede es que el EDA no proporciona frecuencias y fragmenta más los datos.

Sin embargo, la cuestión no es que aplicación lo hace mejor o peor, sino las posibilidades. El SPSS no llega más lejos y lo conocido del diagrama de tallo y hojas es lo que se ha presentado. Ahora se trata de ver novedades, de saber que Tukey extrajo muchas más posibilidades de este gráfico de las que acabamos de ver.

En la siguiente representación, que es la más sencilla que se puede pedir, se puede ver en primer lugar, la variable analizada, que es el porcentaje sobre electores que votaron a IC en las municipales del 95 en el distrito 1 que tiene 29 secciones censales.

Seguidamente, la leyenda nos informa de que 12|8 representa 13.75% y que 34|2 representa a 34.17%.

Debajo, la expresión lo | 1 indica que hay un caso extremo por abajo (ya sabemos que es la sección 1 que no llega al 6%).

Finalmente se ve la distribución, con los tallos (cuerpo del porcentaje) a la izquierda y los decimales a la derecha de los separadores.

```
Stemleaf: PEICM95 (2)
Legend: 12|8 stands for 13.75; 34|2 for 34.17
lo |1

12|8
14|
16|334
18|0
20|177033
22|24
24|2289
26|44
28|2349
30|22
32|33
34|2

>>>stemle 2
```

Así, vemos que el mínimo está por debajo de 12.8% y es un caso especial u "outlier" (representado por la sección 1), que el máximo es aproximadamente 34.2%. El valor típico está alrededor del 24% y la mayor parte de las secciones se mueven en cifras del 20 al 30%. Es una distribución algo asimétrica a la izquierda, de forma que en este distrito, la mayor parte de las secciones proporcionan valores elevados dentro del rango.

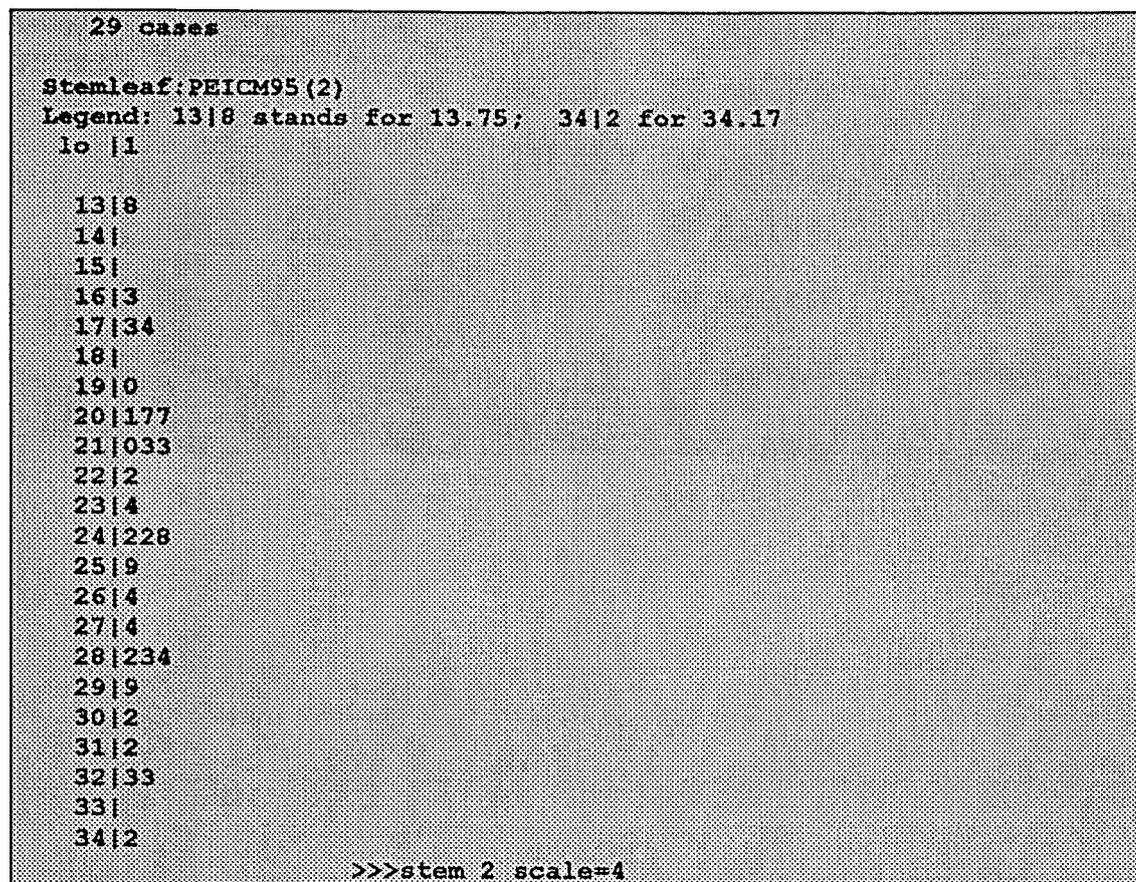
A diferencia de otras aplicaciones estadísticas, el EDA permite resumir o detallar más los mismos datos a través del control de la escala de tallos. Así, por ejemplo, en el siguiente diagrama hemos simplificado más la información:
29 cases

```
Stemleaf: PEICM95 (2)
Legend: 1|4 stands for 13.75; 3|4 for 34.17
lo |1

1|4
1|6779
2|0111112344
2|5667888
3|001224

>>>stem 2 scale=0.5
```

Y, por el contrario, en este diagrama la hemos detallado más:



Y, en el siguiente, todavía más:

```
29 cases
Stemleaf: PEICM95 (2)
Legend: 13|8 stands for 13.75; 34|2 for 34.17
10 |1

13|8
14|
14|
15|
15|
16|3
16|
17|34
17|
18|
18|
19|0
19|
20|1
20|77
21|033
21|
22|2
22|
23|4
23|
24|22
24|8
25|
25|9
26|4
26|
27|4
27|
28|234
28|
29|
29|9
30|2
30|
31|2
31|
32|33
32|
33|
33|
34|2

>>>stem 2 scale=8
```

Por otro lado, el EDA permite ver los datos en orden ascendente (es el defecto) o descendente, como se muestra a continuación:

```

>>>set sort desc

 29 cases
Stemleaf: PEICM95 (2)
Legend: 12|8 stands for 13.75; 34|2 for 34.17

 34|2
 32|33
 30|22
 28|9432
 26|44
 24|9822
 22|42
 20|330771
 18|0
 16|433
 14|
 12|8

lo |1

>>>stem 2

```

Sin embargo, una de las utilidades más atractivas que ofrece es la posibilidad de hacer diagramas comparativos. Por ejemplo, el siguiente compara los resultados porcentuales de IC y del PSC en las municipales del 95, en las secciones de este distrito:

```

Stemleaf: PEICM95 (2)
with      : PEPSCM95 (4)

Legend: 0|2 stands for 1.94; 3|4 for 34.17

      PEICM95                PEPSCM95
                                1|lo |
                                | 0|22223333333344444444
                                | 0|5556789
                                4| 1|
                                9776| 1|
                                4432111110| 2|
                                8887665| 2|
                                422100| 3|

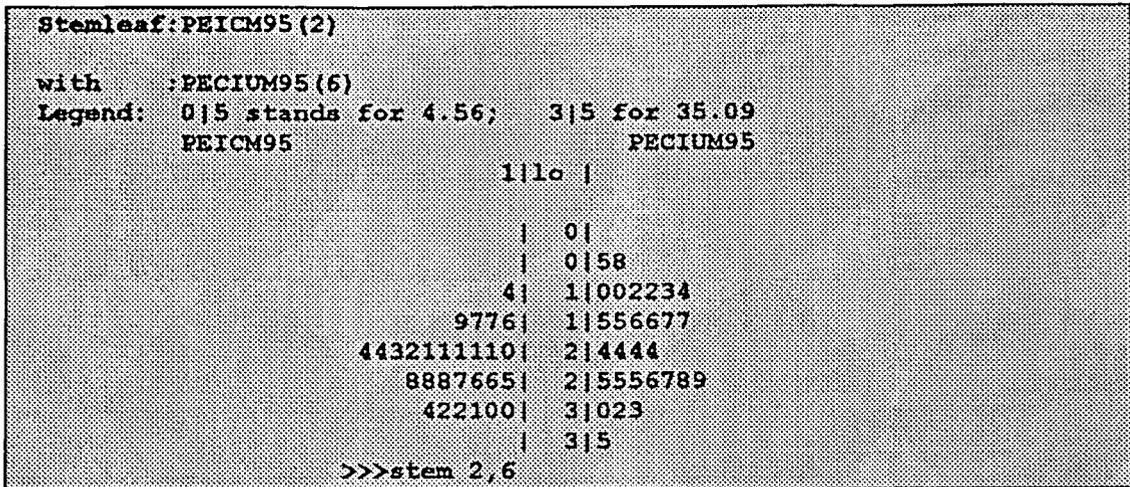
                                |hi |24 27 26 25
                                >>>stem 2,4

 29 cases

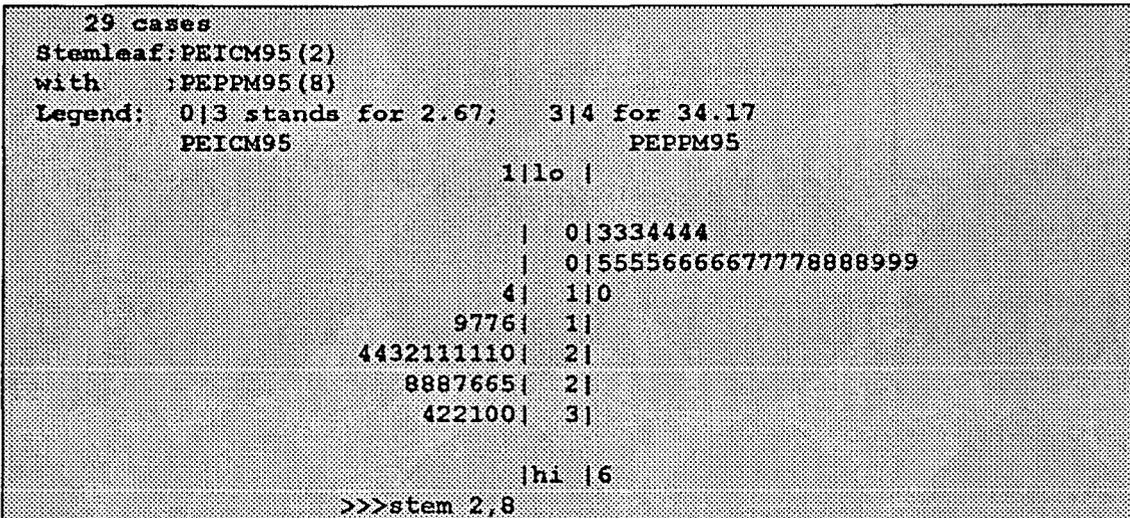
```

Tal y como se puede apreciar observando las explicaciones de la leyenda, los porcentajes del PSC se mueven entre el 2 y el 9% aproximadamente para las 29 secciones, mientras que los de IC parten del 14 hasta el 34% aproximadamente. En IC hay casos extremos: uno por debajo (sección 1), alejado del 14% y en el PSC cuatro por encima del 9% y corresponden a las secciones 24, 27, 26 y 25 (por orden de menor a mayor).

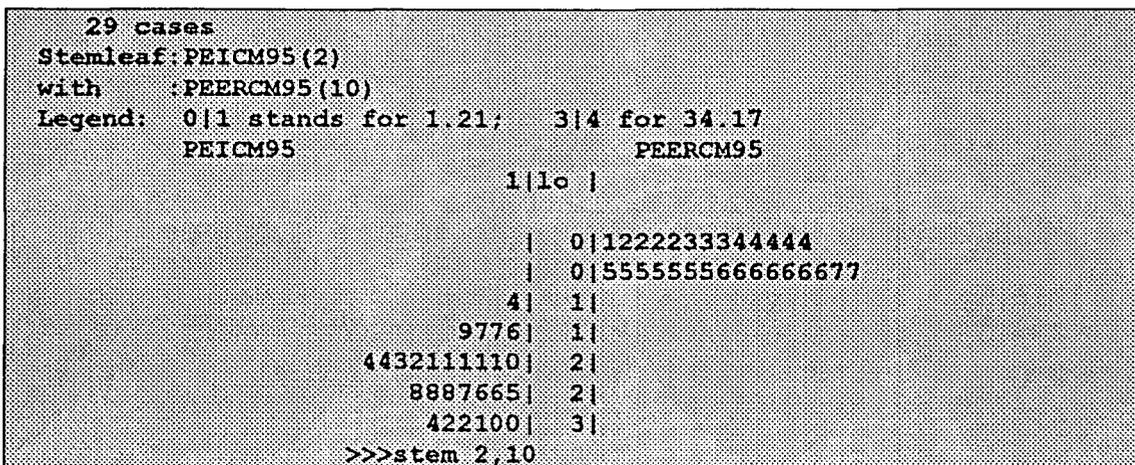
Si se efectúa la misma operación de comparación con el resto de fuerzas políticas, podemos ver los diagramas comparativos de IC con CIU, PP y ERC:



Los porcentajes de CIU se mueven entre el 4.56% y el 35.09%. IC tiene como hemos dicho anteriormente, una sección "outlier" que es la 1 y, en el resto de secciones, sus porcentajes de votos oscilan entre el 14 y el 34%.



El diagrama que compara IC con el PP muestra que esta fuerza política se mueve entre el 2.67% y el 10%, con un valor extremo por arriba corresponde a la sección 6 (10.71%), mientras que IC, tiene una sección por debajo, que es la 1 y el resto se mueven entre el 14 y el 34%



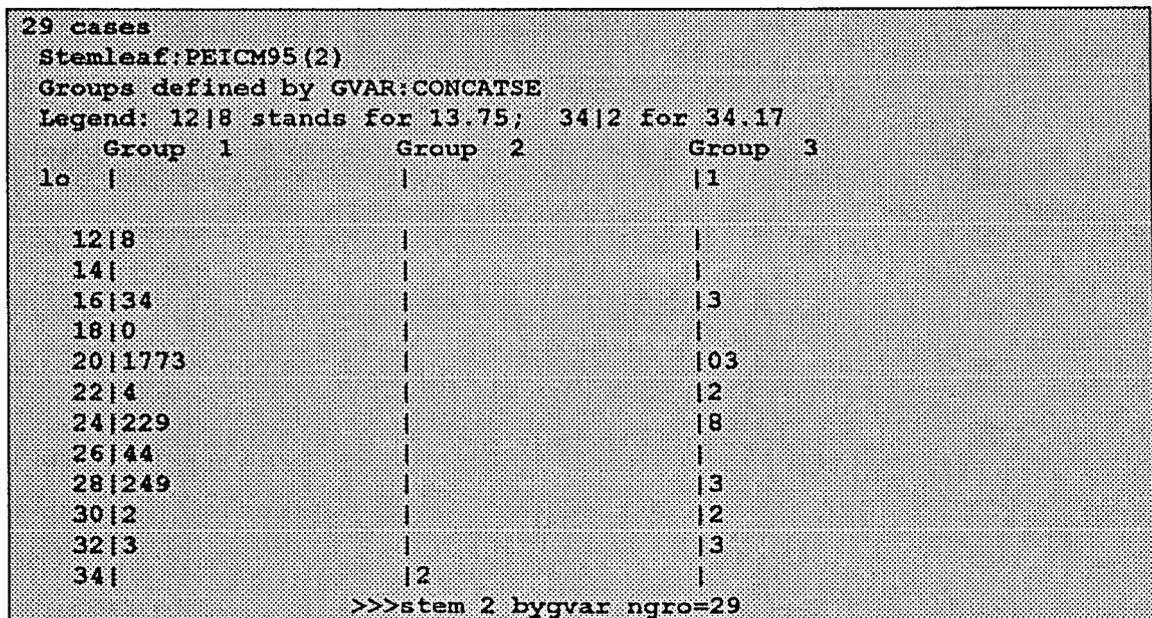
En la comparación con ERC, IC muestra los valores a que ya se ha hecho referencia y ERC oscila entre 1.21% y 7%, aproximadamente.

También es posible efectuar de forma rápida agrupaciones en función de variables que sean de nuestro interés para relacionarlas con los datos electorales. Por ejemplo, se ha confeccionado una variable llamada "CONCATSE" que recoge 4 niveles de conocimiento del catalán en las secciones de este municipio.

Las que llevan un 1 son las de mayor dominio del idioma. Las secciones que llevan un código 2 son las de menor dominio de la lengua. En las secciones que llevan un 3 convive gente que no sabe hablar, leer o escribir con nivel pero que se defiende entendiendo y más o menos hablando y las que llevan el código 4 son secciones mixtas donde vive gente que habla, escribe y lee bien el idioma con otras personas que no lo saben o no lo entienden

En el distrito 1 sólo han salido secciones del tipo 1, 2 y 3. En los siguientes diagramas podemos ver su relación con el porcentaje de votos obtenido por cada fuerza política: Cada "group" representa un valor de la variable sobre dominio del idioma.

Así, la sección donde IC obtuvo menos votos, fue del tipo 3 (mixta) y donde obtuvo más votos fue en una sección del tipo 2 (menor dominio del idioma). El resto se repartieron entre secciones tipo 1 donde se acumulan porcentajes entre el 12.8 y el 26.4% (14, con gran dominio de la lengua) ó 3 (5 secciones mixtas, gente que entiende pero con nivel bajo de dominio) entre el 16.3% y el 24.8%.



Los datos originales se ponen a continuación para efectuar comprobaciones:

```

29 cases variable listing
cases CONCATSE PEICM95
  1 1      3  5.94
  2 2      1 17.42
  3 3      3 21.05
  4 4      1 25.91
  5 5      1 17.31
  6 6      1 13.75
  7 7      3 22.21
  8 8      1 20.06
  9 9      1 19.04
 10 10     3 16.30
 11 11     3 21.32
 12 12     1 23.38
 13 13     1 26.39
 14 14     1 24.21
 15 15     1 20.68
 16 16     1 21.27
 17 17     1 24.15
 18 18     3 30.16
 19 19     1 20.70
 20 20     1 32.29
 21 21     3 24.80
 22 22     1 28.37
 23 23     1 31.18
 24 24     1 28.22
 25 25     2 34.17
 26 26     3 28.31
 27 27     3 32.30
 28 28     1 27.39
 29 29     1 29.89
  
```

Y, seguidamente se ofrece el mismo tipo de gráfico para el resto de fuerzas políticas.

El PSC presenta un "outlier" que es la sección 25. El resto de votos se reparten entre secciones de tipo 1 y tipo 3, pero en niveles bajos (1.94%-8.52%). La sección 25 que nos ha llamado la atención, es donde el PSC triunfa con un 14.12% de los votos, cifra muy superior a las que muestra el gráfico y es una sección de tipo 2 (bajo dominio del catalán). Las secciones 26 y 27 superaron el 8.52%, pero no son tan extremas como la 25.

```

29 cases
Stemleaf: PEPSCM95(4)
Groups defined by GVAR: CONCATSE
Outliers not shown: 25
Legend: 0|9 stands for 1.94; 8|5 for 8.52
      Group 1                      Group 3
                9| 0|
      874432744| 2|3848
                99721| 4|14
                20| 6|
                5| 8|2
                24|hi |27 26
>>>stem 4 bygvar ngro=3
  
```

En el caso de CIU, la mayor concentración de votos se da en secciones tipo 1 y algo menos en las de tipo 3 (alto dominio del catalán y cierto conocimiento). De secciones en que no haya dominio del catalán y CIU tenga votos se da un caso con 4.56% de votos, una cifra muy baja.

```

29 cases
Stemleaf: PECIUM95 (6)
Groups defined by GVAR: CONCATSE
Legend: 0|5 stands for 4.56; 3|5 for 35.09

```

| Group 1 | Group 2 | Group 3 |
|---------|---------|---------|
| 0 | | |
| 0 8 | 5 | |
| 1 034 | | 022 |
| 1 5567 | | 67 |
| 2 444 | | 4 |
| 2 55678 | | 59 |
| 3 02 | | 3 |
| 3 5 | | |

>>>stem 6 bygvar ngro=3

El caso del PP se parece al de CIU, sólo que los porcentajes sobre participación son mucho más bajos.

```

29 cases
Stemleaf: PEPPM95 (8)
Groups defined by GVAR: CONCATSE
Legend: 2|7 stands for 2.67; 10|3 for 10.28

```

| Group 1 | Group 2 | Group 3 |
|---------|---------|---------|
| 2 | | 77 |
| 3 8 | 4 | |
| 4 2557 | | 88 |
| 5 467 | | |
| 6 04 | | 0 |
| 7 22347 | | 68 |
| 8 3 | | |
| 9 0 | | 01 |
| 10 3 | | |
| hi 6 | | |

>>>stem 8 bygvar ngro=3

El caso de ERC, está bastante claro, dado que, aunque con valores mucho más bajos que el resto de fuerzas políticas, el voto tiende a concentrarse en secciones de tipo 1 y en pocas del tipo 3. No deja de ser curioso, sin embargo, que haya una sección en que el dominio es muy bajo pero ERC obtenga, un 1.21% de votos. Sería un caso a investigar y se trata de la sección 25, que ya se ha destacado por ser donde mayor porcentaje de votos han sacado, tanto IC como el PSC.

```

29 cases
Stemleaf: PEERCM95 (10)
Groups defined by GVAR: CONCATSE
Legend: 1|2 stands for 1.21; 7|2 for 7.23

```

| Group 1 | Group 2 | Group 3 |
|---------|---------|---------|
| 1 | 2 | |
| 2 13 | | 24 |
| 3 1379 | | 37 |
| 4 0157 | | |
| 5 1467 | | 03577 |
| 6 023 | | |
| 7 12 | | |

>>>stem 10 bygvar ngro=3

Viendo estos gráficos, se puede concluir que en estas municipales, el conocimiento del idioma es un factor con escasa influencia sobre el voto en este distrito. Por eso, el análisis exploratorio nos sugiere que si hemos descubierto "pistas" interesantes, sigamos indagando en el fenómeno para confirmarlo o rechazarlo.

Mediante una técnica inferencial podemos intentar ver lo que da de sí una hipótesis formal, como, por ejemplo: el porcentaje de votos obtenido por determinadas fuerzas políticas depende del nivel de dominio del catalán en la sección. Esto es lo que nos conduce a ver que la exploración es un buen camino a seguir para probar posteriormente otras herramientas más sofisticadas.

La exploración nos da a entender que el idioma, en este caso no es muy relevante, de forma que si, por ejemplo, si efectuamos un contraste de diferencia de medias, mediante la aplicación SPSS:

COMPARACIÓN DE MEDIAS

| CONOCIMIENTO DEL CATALÁN EN LA SECCI | | % SE CIU M95 | % SE ERC M95 | % SE IC M95 | % SE PP M95 | % SE PSC M95 |
|---|------------|-----------------|-----------------|----------------|----------------|-----------------|
| SECCION DE ALTO NIV DE CONQC. Y USO DEL CATALÁN | Media | 21,5439 | 4,7568 | 23,7691 | 6,6293 | 4,3917 |
| | N | 19 | 19 | 19 | 19 | 19 |
| | Desv. ttp. | 7,7343 | 1,4973 | 5,1358 | 1,9980 | 1,9498 |
| SECCION DE MAS BAJ CONOCIMIENTO Y USO DEL CATALÁN | Media | 4,5558 | 1,2149 | 34,1686 | 3,4169 | 14,1230 |
| | N | 1 | 1 | 1 | 1 | 1 |
| | Desv. ttp. | , | , | , | , | , |
| ENTIENDEN EL CATALA PERO NO LO DOMINAN | Media | 19,7710 | 4,3158 | 22,4876 | 6,0451 | 5,2941 |
| | N | 9 | 9 | 9 | 9 | 9 |
| | Desv. ttp. | 8,2993 | 1,4214 | 7,9831 | 2,4820 | 2,8581 |
| Total | Media | 20,4079 | 4,4978 | 23,7300 | 6,3372 | 5,0073 |
| | N | 29 | 29 | 29 | 29 | 29 |
| | Desv. ttp. | 8,2532 | 1,5683 | 6,2892 | 2,1717 | 2,8336 |

Resulta que en IC la lengua es influyente pero poco, ya que sólo destaca que obtiene una media superior a las otras en la sección de menor dominio. El candidato satisfacía a todas las secciones y la fuerza política era especialmente apoyada en la zona de menor conocimiento del catalán.

CIU tiene, por el contrario, la media más baja en las secciones de tipo 2, es decir donde menos arraigo tiene el catalán, mientras que se acerca a los resultados de IC en las demás. Lo mismo sucede con ERC y el PP. En cambio, el PSC tiene su mayor apoyo en las secciones donde menos se domina el catalán, teniendo resultados bajos en las restantes.

Por tanto, existen ciertos indicios de mayor o menor apoyo según el dominio de la lengua, pero no son suficientemente significativos como para aceptar la hipótesis de dependencia. En todo caso, a nivel descriptivo ya se ha dicho hasta dónde tiene importancia el fenómeno.

Finalmente, los diagramas de tallo y hojas comparativos también pueden resultar de gran utilidad para comparar los resultados de una misma fuerza política en distintas elecciones, resultados de una misma fuerza política en distintos tipos de elecciones (autonómicas, municipales y generales), y todo tipo de combinaciones que se nos ocurran al efecto.

Está claro que el electorado tiende a votar de forma diferente en las elecciones de distinto tipo. Para comenzar la descripción de un fenómeno como éste se puede

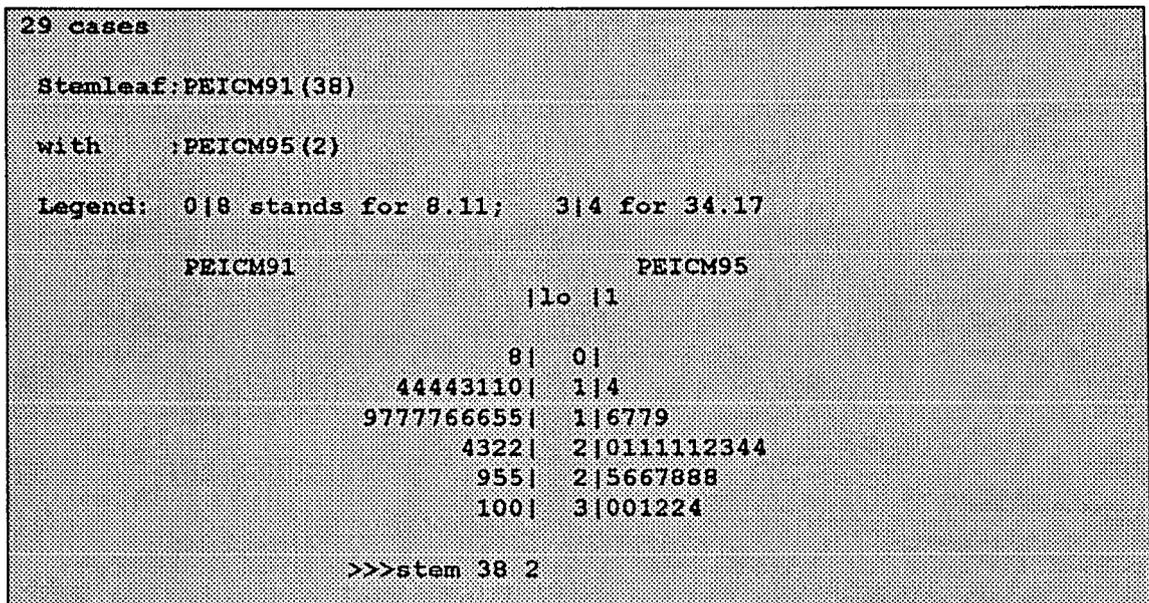
realizar la observación de unos cuantos gráficos comparativos y ver si realmente hay algo interesante en lo que valga la pena profundizar:

Por ejemplo, en el siguiente diagrama se puede ver la comparación de IC entre dos convocatorias electorales municipales sucesivas: las del 91 y las del 95. Inspeccionando el gráfico puede apreciarse que, en este distrito, en 1991 el valor más bajo fue 8.11%, mientras que en 1995 fue 5.9%

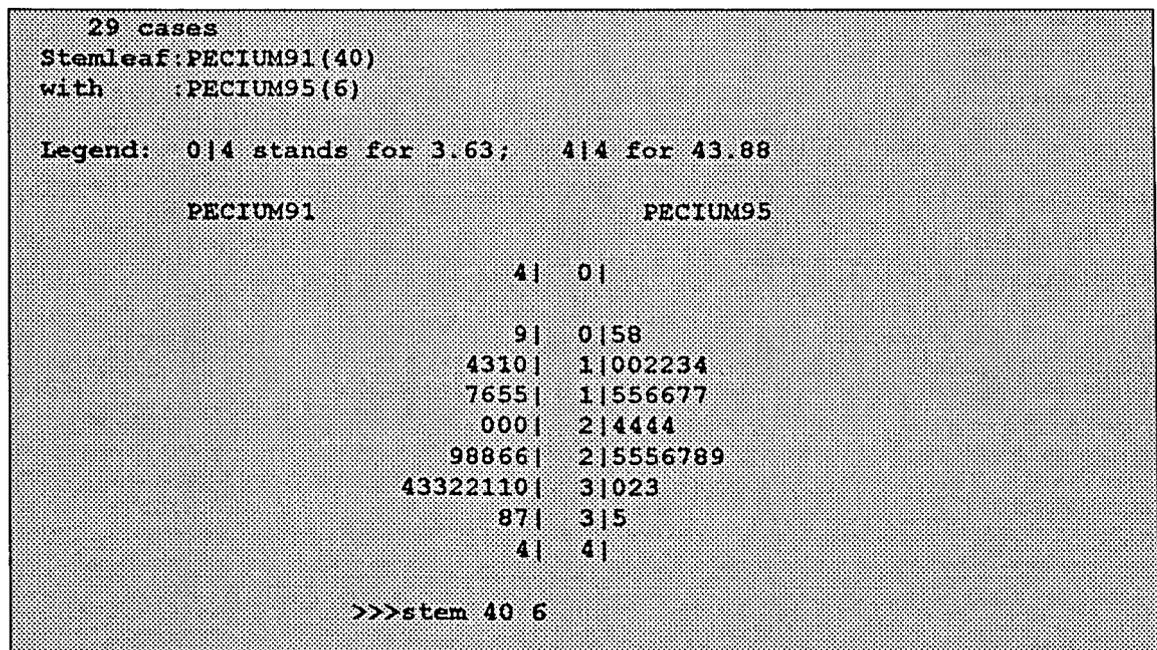
En cambio, globalmente, la fuerza política ha mejorado sus resultados sensiblemente de una elección a otra.

Si esto es o que se observa, el investigador puede seguir adelante y preguntarse ¿por qué?, ¿qué factores pudieron incidir en esta mejora?

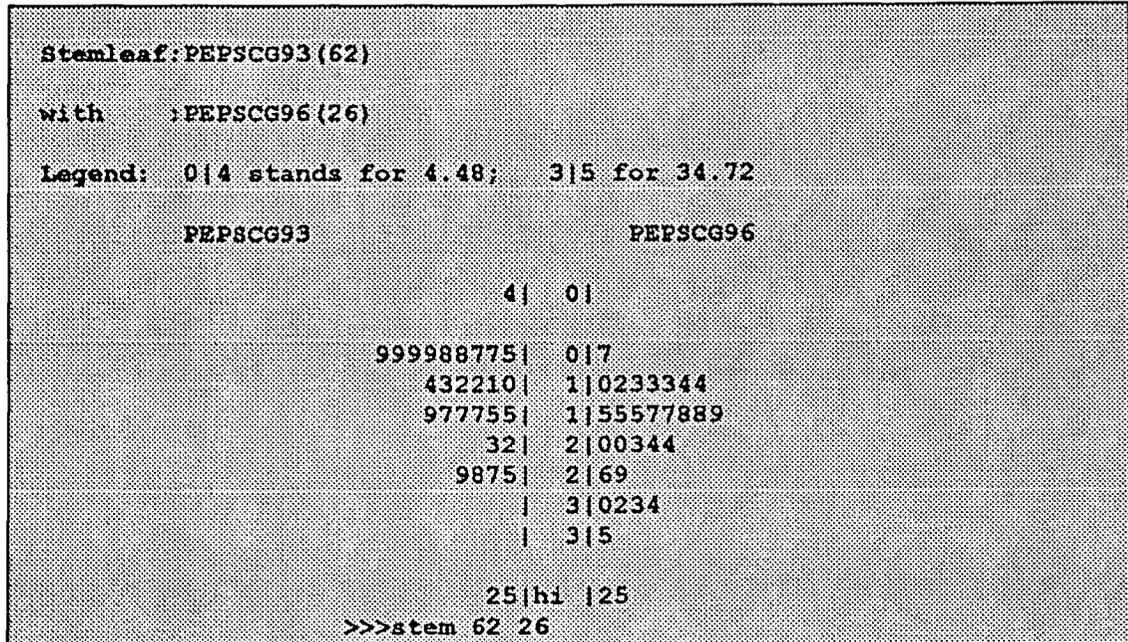
Esta parte ya se saldría de los límites del análisis exploratorio, pero no hay duda de que éste nos ha llevado a formularnos más preguntas y por tanto a avanzar.



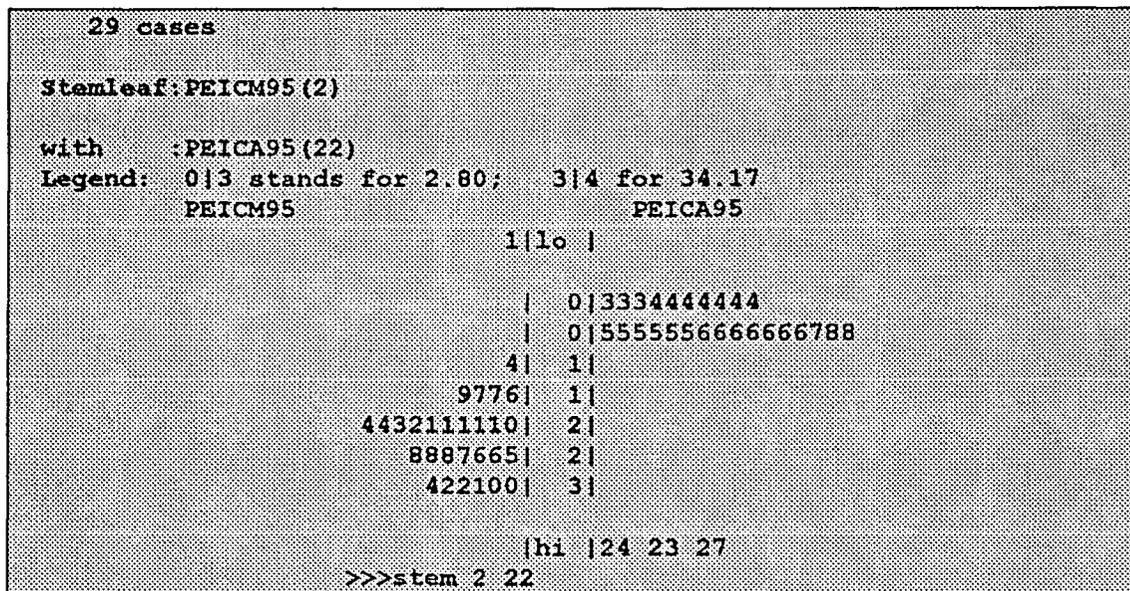
El siguiente caso compara los resultados de CIU en las municipales del 91 con las del 95. Según puede apreciarse, CIU tuvo un retroceso en el 95 y mejores resultados globales en el 91 para este distrito.



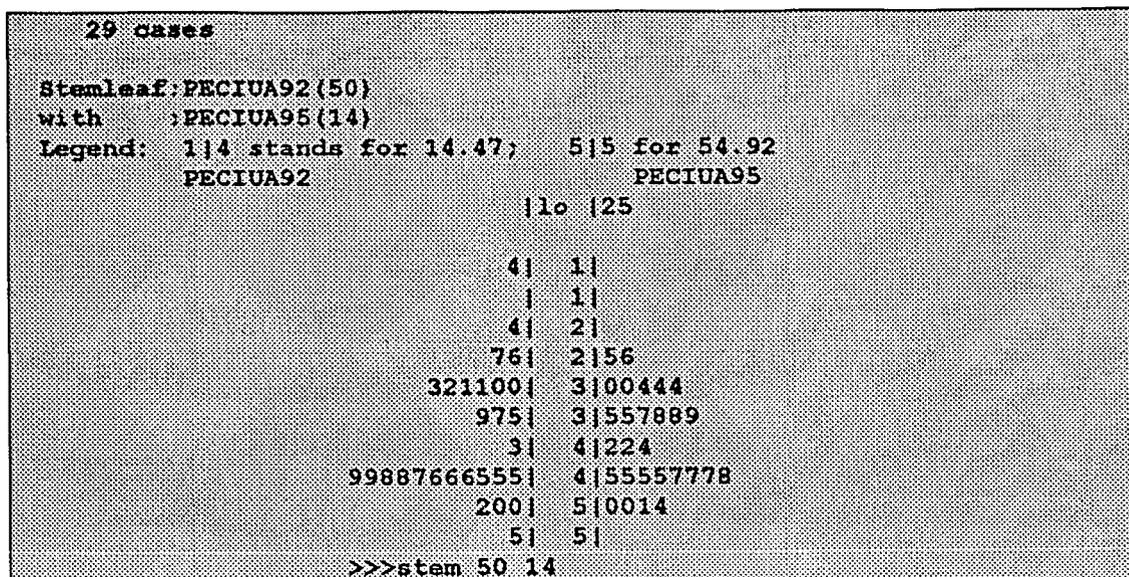
En el siguiente gráfico se comparan los resultados del PSC en dos elecciones generales consecutivas: 1993 y 1996. También se aprecian diferencias, habiendo mejorado los resultados en el 96.



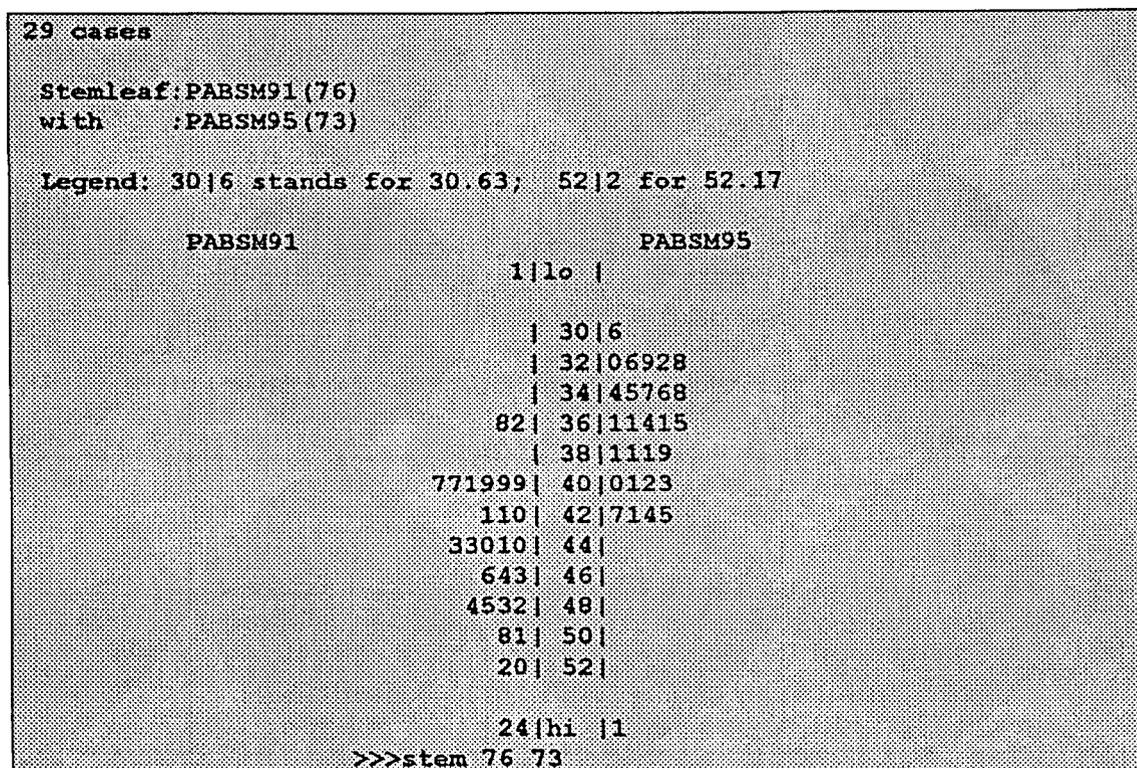
El siguiente caso muestra los indicios de comportamiento electoral diferente según el tipo de elección en un mismo año electoral. Así, IC, obtiene, sin duda, muchos mejores resultados en las municipales que en las autonómicas, de forma que a partir de esta constatación se puede comenzar a indagar en la fidelidad de voto y en los factores que la influyen.



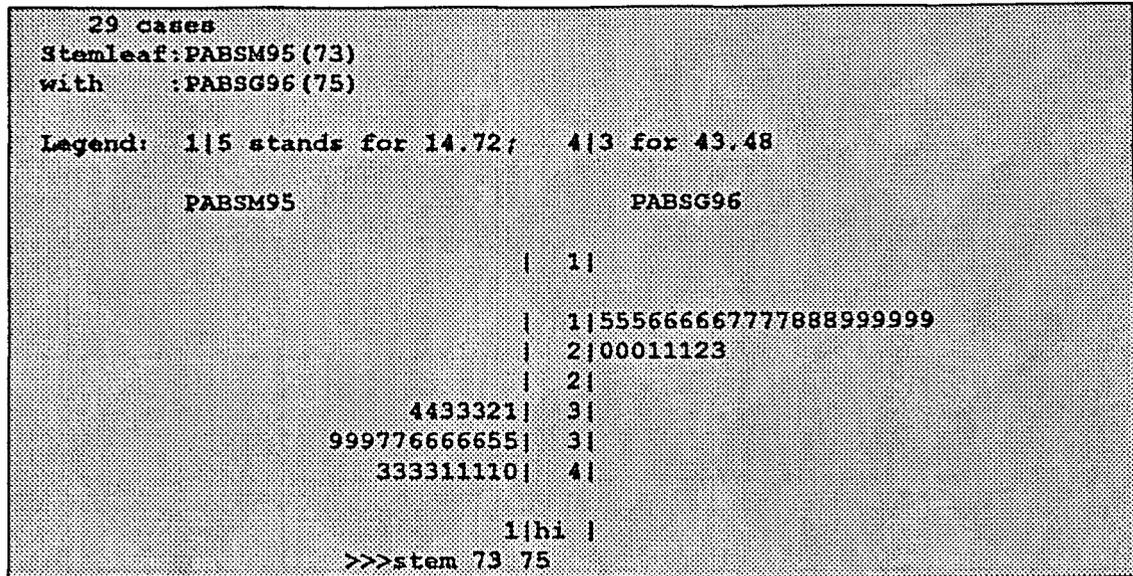
El siguiente gráfico compara los resultados de CIU entre dos autonómicas, las de 1992 y las de 1995. Como puede apreciarse, existen menos diferencias de comportamiento entre el electorado en ambas convocatorias.



En el siguiente diagrama se compara la abstención entre las municipales del 91 y las del 95. Como puede apreciarse, en la última convocatoria mejoró mucho el estado de la abstención aún siendo elevada, ya que los porcentajes tienden a tener cifras más bajas en la segunda convocatoria.



Finalmente, se compara la abstención entre unas municipales (las del 95) y unas generales (las del 96), viéndose que ésta aumenta mucho en las municipales con relación a las generales.



El diagrama de tallo y hojas es, por tanto, una fuente inagotable de exploración en busca de indicios de todo tipo de preguntas que, de ser hallados, nos permiten ir más allá de la fase 4 del método científico.

4.2.3.2 LA CODIFICACIÓN

En muchas ocasiones resulta preferible reemplazar los números por símbolos que nos proporcionen visualmente la información que nos interesa, de forma rápida y eficaz. La codificación es una herramienta clave del análisis exploratorio y muy aplicable a datos electorales agregados como se mostrará a continuación.

Existen diversos tipos de codificación dentro del análisis exploratorio:

Codificación distribucional: los símbolos que se muestran reflejan la posición de cada observación dentro de la distribución de una variable. Por ejemplo, los símbolos se usan para saber si un caso es muy extremo, extremo o cercano a un valor determinado. Normalmente, se usan diferentes símbolos para los valores que quedan por encima o por debajo de la mediana.

Codificación basada en divisiones de la distribución (percentiles, cuartiles, etc.): los casos se agrupan según la división establecida de la variable de acuerdo con algún criterio y se usa un símbolo diferente para cada parte. Cada división contiene el mismo número de observaciones (codificación fráctil), cada división corresponde a un intervalo de igual longitud (codificación interválica), o también es posible establecer divisiones definidas por el usuario que es el que marca sus límites.

Codificación referencial: cada símbolo refleja la posición de un caso respecto a un criterio determinado. El criterio utilizado más a menudo es el de la mediana. Si se usa ese criterio, entonces, son marcados de forma diferente los casos que quedan por encima o por debajo de la mediana. También es posible, en lugar de marcar los casos por encima o por debajo, indicar la distancia de un caso a un valor de referencia, ya sea usando diferentes símbolos o utilizando más de un símbolo, por ejemplo, un + para un caso cercano a la mediana, y 2,3, etc. signos + para casos que se alejen más.

Codificación de marcas: todos los casos que responden a un mismo criterio son marcados con símbolos especiales y el resto de casos no se marcan.

La codificación se usa para resaltar aspectos importantes de la información que estamos tratando de obtener, reemplazando la forma numérica (no siempre rápida o agradable de leer y tratar) por un conjunto de símbolos.

Se recomienda su presentación horizontal cuando el número de casos es relativamente pequeño (unos 100 como máximo) y la vertical cuando el número de casos es superior a esa cifra.

En la codificación distribucional los símbolos que aparecen muestran la posición de cada observación dentro de la distribución de una variable.

En análisis exploratorio es muy frecuente utilizar el criterio de comparación con la mediana como valor de referencia. Así, en una codificación distribucional podemos distinguir:

@ este símbolo que representa los valores que están más allá de 1.5 veces la dispersión en torno a la mediana y son los valores superiores más alejados del centro (high far out value)

este símbolo que representa a los valores que están entre 1 y 1.5 desviaciones de la mediana por la parte superior de la distribución de una variable (high out value)

+ este símbolo que representa a los valores que son adyacentes a la mediana pero que quedan algo por encima de ella (high adjacent)

* este símbolo que representa a los valores que prácticamente coinciden con la mediana (in values)

- este símbolo que representa a los valores que son adyacentes a la mediana pero que quedan algo por debajo de ella (low adjacent)

= este símbolo que representa a los valores que están entre 1 y 1.5 desviaciones de la mediana por la parte inferior de la distribución (low out value)

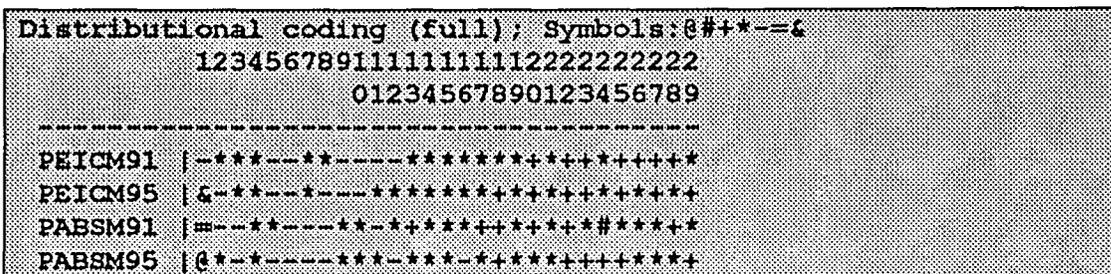
& este símbolo que representa a los valores que están más allá de 1.5 desviaciones de la mediana por la parte inferior de la distribución (far low out values)

El tipo de preguntas a que puede dar respuesta la codificación distribucional es muy amplio y puede tratarse, además, a diferentes niveles de complejidad.

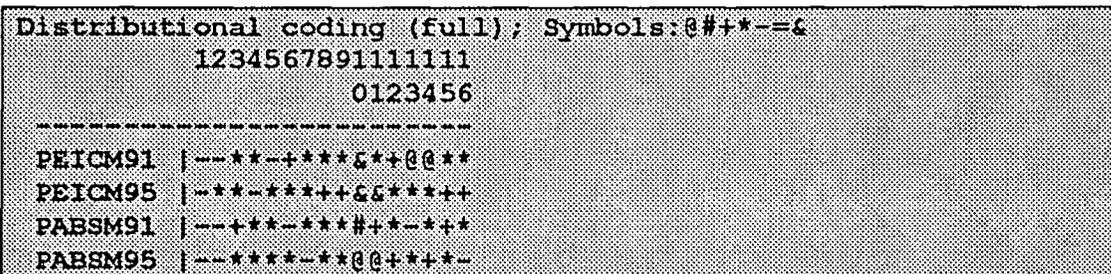
Siguiendo con los datos que se han utilizado en la sección anterior, se sabe que en este municipio, la alcaldía está actualmente en poder de IC. Supongamos que deseamos ver el comportamiento electoral en los distintos distritos y además, comparar los resultados entre dos convocatorias de elecciones municipales: 1991 y 1995. La codificación distribucional lo permite fácilmente:

En primer lugar se puede ver la codificación distribucional del primer distrito que tiene 29 secciones censales. El título del gráfico indica que se está efectuando una codificación distribucional, con los símbolos que antes hemos comentado y las secciones del distrito están señalizadas mediante números verticales que van del 1 al 29. Debajo de cada sección están los símbolos correspondientes según el comportamiento de la fuerza política. Como puede apreciarse rápidamente por estos símbolos, hubo cambios en diferentes secciones y, la situación de la

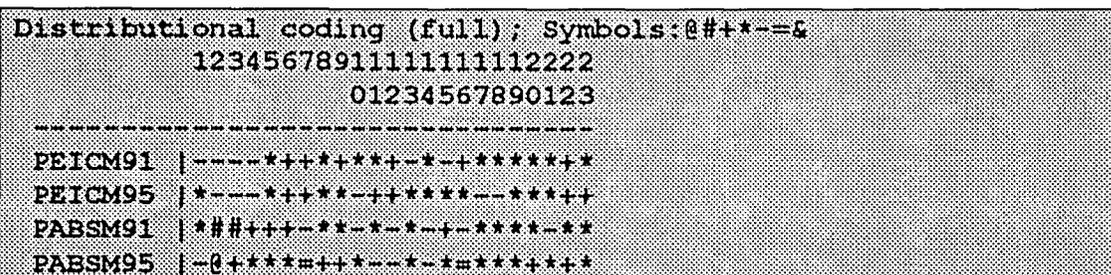
abstención entre ambas elecciones es algo diferente. En la sección 1 IC empeoró más sus resultados y la abstención subió brutalmente, de estar bastante por debajo de su mediana al máximo extremo alejado de ella por encima. En la sección 2 IC también pierde votos, pasa de estar en su mediana a estar algo por debajo de ella. En las secciones 3 y 4 se mantiene. Y, de esta forma, se puede leer todo el gráfico y extraer alguna conclusión sobre el comportamiento en este distrito de los votantes de IC en dos convocatorias municipales sucesivas. En total empeora en 5 secciones y mejora en 4. Ese es el balance para este distrito.



En el distrito 2 el balance es el siguiente: mejora en 7 secciones y empeora en 5, destacando las pérdidas en las secciones 11,13 y 14, en que baja mucho. En cuanto a la abstención, la situación tiende a empeorar en todo el distrito, aumentado mucho en las secciones 10 y 11.



En el distrito 3: mejora en 5 secciones y empeora en otras 5, pero las pérdidas y ganancias no son muy sustanciales. La abstención es muy variable entre las dos elecciones en este distrito, en algunas zonas mejora pero en general es abundante.



En el distrito 4, IC también ha sufrido ciertas variaciones: mejora en 4 secciones y empeora en 6. Hay dos descensos fuertes en las secciones 4 y 12. La abstención es un poco más estable en cuanto a diferencias, pero sigue siendo amplia en general.

```
Distributional coding (full); Symbols: @#+-=&
      1234567891111111112
      01234567890
-----
PEICM91 | +**@*****-+*#@***-&=-
PEICM95 | **--+**+**+**+**#***&-*
PABSM91 | *-**+**+**+**+**+&***-+*
PABSM95 | +**=*-*-**+*-+**+**
```

En el distrito 5, los cambios han sido: mejorar en tres secciones, especialmente en la 10 y empeorar en 2, aunque no de forma excesiva. La abstención baja mucho en las secciones 3 y 6 y en el resto permanece bastante estable.

```
Distributional coding (full); Symbols: @#+-=&
      1234567891111
      0123
-----
PEICM91 | **-*+--+*+***
PEICM95 | *--**--+**#*+*
PABSM91 | +***-*-+*+*
PABSM95 | **&***+*-***+*
```

En el distrito 6, mejora en 7 secciones, especialmente en la 11 y empeora, pero no excesivamente, en 4. La abstención, tendió a bajar en este distrito.

```
Distributional coding (full); Symbols: @#+-=&
      123456789111111112222
      01234567890123
-----
PEICM91 | *-**-**+**+**+**+**+**+**+
PEICM95 | *-**-**+**+**+**+**+**+**+
PABSM91 | **+***---*+*+**+**+&+**
PABSM95 | -+***+*-+**+**+**+**+**+*
```

Finalmente, en el distrito 7, IC mejoró en la sección 4 y empeoró en la 3, permaneciendo estables las demás. La abstención permaneció sin cambios entre las dos convocatorias, siendo la sección 3 la más abstencionista.

```
Distributional coding (full); Symbols: @#+-=&
      123456
-----
PEICM91 | **+***
PEICM95 | **+***
PABSM91 | *-+***
PABSM95 | *-+***
```

Una vez visto el resultado de la codificación, el investigador, estadístico en este caso, se pregunta acerca del método: ¿es científico o no? ¿es cualitativo o cuantitativo? ¿se basta a sí mismo para explicar los resultados o simplemente ha

llegado en su explicación hasta donde ha podido? ¿qué más debería plantearse para seguir adelante con la investigación? ¿qué está buscando realmente?

Todas estas preguntas nos devuelven al principio:

El investigador, en este caso, sabe la técnica, ha descrito una situación, pero no ha llegado mucho más lejos. Por tanto, debe plantearse las cosas de otra forma mucho más estructurada, ya que, en caso contrario, el método que ha aplicado resulta prácticamente estéril.

Por tanto, repasemos las etapas del método científico y veamos hasta dónde se puede llegar con estas herramientas:

El primer paso, como ya se apuntó en su momento, consiste en observar lo que puede ser observado y aceptar o recusar las observaciones como suficientes. Pero, para observar, necesitamos al menos, determinar qué deseamos observar y para qué, es decir, establecer al menos una pregunta. La experiencia anterior se ha presentado como un simple ejercicio de descripción: ver el comportamiento de los votantes de IC y de la abstención en los 7 distritos. Realmente, ese objetivo se ha cumplido, pero no llega más lejos y, por tanto, nada tiene de científico.

Por consiguiente, vamos a planteamos una cuestión sencilla y a aplicar el método científico lo más cuidadosamente posible. Sea la pregunta: ¿cuáles son las secciones censales dentro de cada distrito electoral en que cada fuerza política superó la mediana de IC en las elecciones municipales de 1995 en este municipio?

Observación: elegimos como método de observación el de la codificación con marcas de todos los partidos en los 7 distritos. Así, una sección será marcada cuando supere la mediana del porcentaje de votos a IC sobre el total de electores en todas las secciones del municipio

Resultados:

En primer lugar, calculamos la mediana del porcentaje sobre electores de IC para todo el municipio:

Estadísticos

| | N | | Mediana |
|-------------|---------|----------|---------|
| | Válidos | Perdidos | |
| % SE IC M95 | 130 | 0 | 30,1156 |

Ahora, imponemos el criterio a la codificación, de que marque la sección censal si supera este valor.

```

Mark values greater than 30.12; Symbols: @
1234567891111111112222222222
01234567890123456789
-----
PECIDM95 | @@ @@
PEPSCM95 |
PEPPM95 |
PEICM95 | @ @ @ @ @
PEERC95 |
PABSM95 | @@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
    
```

```

Mark values greater than 30.12; Symbols:@
      12345678911111111
          0123456
-----
PECIUM95|
PEPSCM95|
PEPPM95 |
PEICM95 |           @           @
PEERC95 |
PABSM95 | @@@@@@@@@@@@@@@@@@

```

```

Mark values greater than 30.12; Symbols:@
      1234567891111111112222
          01234567890123
-----
PECIUM95|
PEPSCM95|
PEPPM95 |
PEICM95 | @           @@@@@@@@@@@@@@@ @@@@@@@
PEERC95 |
PABSM95 | @@@@@@@@@@@@@@@@@@

```

```

Mark values greater than 30.12; Symbols:@
      12345678911111111112
          01234567890
-----
PECIUM95|
PEPSCM95|
PEPPM95 |
PEICM95 | @@@@@@@@@@@@@@@@@@@ @
PEERC95 |
PABSM95 | @@@@@@@@@@@@@@@@@@

```

```

Mark values greater than 30.12; Symbols:@
      1234567891111
          0123
-----
PECIUM95|
PEPSCM95|
PEPPM95 |
PEICM95 |           @ @ @ @
PEERC95 |
PABSM95 | @@@@@@@@@@@@@@@

```

```

Mark values greater than 30.12; Symbols:@
12345678911111111112222
01234567890123
-----
PECIUM95 |
PEPSCM95 |
PEPPM95 |
PEICM95 |e      @@@@@@      @@@
PEERC95 |
PABSM95 |@@@@@@@@@@@@@@@@@@@@@@@@@@@@

```

```

Mark values greater than 30.12; Symbols:@
123456
-----
PECIUM95 |
PEPSCM95 |
PEPPM95 |
PEICM95 |@@@@@@@
PEERC95 |
PABSM95 |e @@@@

```

Ahora, para seguir adelante, debemos aceptar o recusar las observaciones como suficientes. Para ello, el investigador puede comprobar estos resultados mediante otras fuentes o técnicas, más tradicionales y laboriosas, pero que le ayudarán a validar estos primeros resultados.

Mediante la aplicación estadística SPSS, podemos construir unas tablas que tienen la misma filosofía que la codificación del EDA, pero que son más engorrosas de obtener y que marcan las secciones con un 100% cuando cumplen el criterio impuesto, es decir, superar la mediana 30.1156%. Para el primer distrito, la tabla nos indica que CIU superó esa mediana en las secciones 2, 3, 5 y 6. Que IC lo hizo en las secciones 18, 20, 23, 25 y 27 y que, la abstención, superó esa mediana en todas las secciones del distrito, resultados que coinciden plenamente con los de la codificación del primer distrito realizada por medio del programa EDA.

| SECCIÓN | % SE CIU M95 > | % SE PSC M95 > | % SE PP M95 > | % SE IC M95 > | % SE ERC M95 > | % SE ABS M95 > |
|---------|----------------------|----------------------|---------------------|---------------------|----------------------|----------------------|
| | 30,1156 | 30,1156 | 30,1156 | 30,1156 | 30,1156 | 30,1156 |
| 1 | ,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |
| 2 | 100,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |
| 3 | 100,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |
| 4 | ,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |
| 5 | 100,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |
| 6 | 100,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |
| 7 | ,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |
| 8 | ,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |
| 9 | ,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |
| 10 | ,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |
| 11 | ,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |
| 12 | ,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |
| 13 | ,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |
| 14 | ,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |
| 15 | ,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |
| 16 | ,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |
| 17 | ,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |
| 18 | ,0% | ,0% | ,0% | 100,0% | ,0% | 100,0% |
| 19 | ,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |
| 20 | ,0% | ,0% | ,0% | 100,0% | ,0% | 100,0% |
| 21 | ,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |
| 22 | ,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |
| 23 | ,0% | ,0% | ,0% | 100,0% | ,0% | 100,0% |
| 24 | ,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |
| 25 | ,0% | ,0% | ,0% | 100,0% | ,0% | 100,0% |
| 26 | ,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |
| 27 | ,0% | ,0% | ,0% | 100,0% | ,0% | 100,0% |
| 28 | ,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |
| 29 | ,0% | ,0% | ,0% | ,0% | ,0% | 100,0% |

Por tanto, la metodología empleada anteriormente resulta más eficaz en cuanto a presentación, ocupa mucho menos espacio y proporciona la misma información, usando un solo comando para obtenerla, mientras que en el SPSS se ha tenido que trabajar más para llegar al mismo resultado, ocupando mucho más espacio y siendo la presentación mucho más engorrosa.

El investigador decide dar por válida la información obtenida y considera que lo observado mediante el primer método es correcto y suficiente.

La segunda etapa consiste en la descripción de lo observado y aceptación o recusación provisional de la descripción como correcta y adecuada.

La descripción que se puede efectuar de la observación de los resultados se resume a continuación:

La abstención superó el 30.1156% sobre el total de electores en todas las secciones censales de todos los distritos excepto la 2 del distrito 7.

CIU superó la mediana de IC en el distrito 1 secciones 2, 3, 5 y 6. Ninguna otra fuerza política superó ese porcentaje en ninguna sección censal de ningún distrito a excepción de IC.

Los resultados para IC fueron de superación de su mediana en:

Distrito 1, secciones: 18, 20, 23, 25, 27 (17.24% de las secciones)

Distrito 2, secciones: 9, 15 (12.5% de las secciones)

Distrito 3, secciones: 1,5,6,7,8,9,10,11,12,13,14,15,17,18,19,20,21,22,23 (82.6% de las secciones)

Distrito 4, secciones: 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,20 (90% de las secciones)

Distrito 5, secciones: 4,5,7,8,10,12 (46.15% de las secciones)

Distrito 6, secciones: 1,7,8,9,10,11,17,18,19 (39.13% de las secciones)

Distrito 7, secciones: 1,2,3,4,5,6 (100% de las secciones)

Por tanto, puede decirse que siendo la abstención la que más iguala los resultados de IC, los distritos en que triunfa son, especialmente el 7, el 4 y el 3.

Desde el punto de vista metodológico esta información ya es más precisa y concreta y además, es mapificable en cualquier momento, de forma que puede interaccionar con la geografía electoral sin problemas.

También existen otras muchas posibilidades de obtener información acerca del éxito de cada fuerza política en los distritos. Por ejemplo, la siguiente presentación indica las 11 secciones en que las diversas opciones políticas obtuvieron mejores resultados para el distrito 1:

| Listing top | 11 cases | | | | | | | | | | |
|-------------|----------|----|----|----|----|----|----|----|----|----|----|
| PEICM95 : | 25 | 27 | 20 | 23 | 18 | 29 | 22 | 26 | 24 | 28 | 13 |
| PEFSCM95 : | 25 | 26 | 27 | 24 | 23 | 21 | 22 | 28 | 8 | 13 | 19 |
| PECTUM95 : | 6 | 3 | 5 | 2 | 10 | 16 | 12 | 15 | 9 | 8 | 11 |
| PEPPM95 : | 6 | 8 | 7 | 21 | 19 | 9 | 10 | 16 | 26 | 4 | 5 |
| PERRCM95 : | 15 | 14 | 12 | 9 | 2 | 10 | 11 | 5 | 17 | 3 | 6 |
| PABSM95 : | 1 | 24 | 18 | 22 | 23 | 29 | 25 | 28 | 17 | 20 | 13 |

Los resultados que va facilitando esta metodología son evidentes, de forma que aceptaremos la descripción como válida.

Ahora, el método científico sigue con la etapa de medición de lo que pueda ser medido; cierto que esto no es más que una forma especial de la observación y la descripción, pero es lo suficientemente característica e importante como para considerarla por separado.

Recordemos la pregunta objeto de investigación: ¿cuáles son las secciones censales dentro de cada distrito electoral en que cada fuerza política superó la mediana de IC en las elecciones municipales de 1995 en este municipio?

Está claro que la pregunta ya ha sido respondida, por tanto, ¿qué puede plantearse el investigador en esta etapa? ¿qué es lo que puede o debe medir?

Inmediatamente se nos pueden ocurrir muchas cosas: ¿a qué distancia quedaron realmente las demás fuerzas políticas de la mediana de IC?. Esta podría ser una cuestión que complementaría la información obtenida y susceptible de medición. Con la abstención ya lo hemos hecho en cierto modo y hemos podido constatar que, excepto en una sección censal, en todas las demás alcanzó y superó la mediana de IC. También es cierto que no sabemos si en mucho o en poco, de forma que esa parte se podría mejorar. Por tanto, vamos a intentar medir esas distancias mediante codificación.

Una forma muy sencilla y descriptiva que ofrece el EDA, consiste en confeccionar una lista de las secciones censales, viendo, por orden, en cada una qué opción ganó a las restantes.

Así, según esta lista, en la sección censal 1 del distrito 1, el porcentaje más elevado fue para la abstención, seguida de CIU, seguida de IC, PP, ERC y PSC. En la sección censal 2 también ganó la abstención, seguida de CIU, de IC, del PP, de ERC y del PSC. El resto se lee fácilmente y el investigador puede tener una idea rápida del orden en que quedaron las fuerzas políticas en cada sección y distrito.

La lectura de la tabla en vertical también permite ver las fuerzas políticas y la abstención en la posición que han quedado, de la primera a la sexta, con mayor frecuencia. Así, la abstención está en primer lugar en todas las secciones, con excepción de la 3 y la 6 en que se ve superada por IC, etc. De todo ello, el investigador puede extraer nuevas preguntas y ampliar su indagación hacia puntos concretos que le llamen la atención.

| Listing top | 6 variables | | | | | | |
|-------------|-------------|----------|----------|----------|----------|---------|---------|
| 1 | : | PABSM95 | PECIUM95 | PEICM95 | PEPPM95 | PEERC95 | PEPSC95 |
| 2 | : | PABSM95 | PECIUM95 | PEICM95 | PEPPM95 | PEERC95 | PEPSC95 |
| 3 | : | PECIUM95 | PABSM95 | PEICM95 | PEERC95 | PEPSC95 | PEPPM95 |
| 4 | : | PABSM95 | PEICM95 | PECIUM95 | PEPPM95 | PEERC95 | PEPSC95 |
| 5 | : | PABSM95 | PECIUM95 | PEICM95 | PEPPM95 | PEERC95 | PEPSC95 |
| 6 | : | PECIUM95 | PABSM95 | PEICM95 | PEPPM95 | PEERC95 | PEPSC95 |
| 7 | : | PABSM95 | PECIUM95 | PEICM95 | PEPPM95 | PEERC95 | PEPSC95 |
| 8 | : | PABSM95 | PECIUM95 | PEICM95 | PEPPM95 | PEPSC95 | PEERC95 |
| 9 | : | PABSM95 | PECIUM95 | PEICM95 | PEPPM95 | PEERC95 | PEPSC95 |
| 10 | : | PABSM95 | PECIUM95 | PEICM95 | PEPPM95 | PEERC95 | PEPSC95 |
| 11 | : | PABSM95 | PECIUM95 | PEICM95 | PEERC95 | PEPPM95 | PEPSC95 |
| 12 | : | PABSM95 | PECTUM95 | PEICM95 | PEERC95 | PEPPM95 | PEPSC95 |
| 13 | : | PABSM95 | PEICM95 | PECIUM95 | PEPPM95 | PEERC95 | PEPSC95 |
| 14 | : | PABSM95 | PEICM95 | PECIUM95 | PEERC95 | PEPPM95 | PEPSC95 |
| 15 | : | PABSM95 | PECIUM95 | PEICM95 | PEERC95 | PEPPM95 | PEPSC95 |
| 16 | : | PABSM95 | PECIUM95 | PEICM95 | PEPPM95 | PEERC95 | PEPSC95 |
| 17 | : | PABSM95 | PEICM95 | PECIUM95 | PEPPM95 | PEERC95 | PEPSC95 |
| 18 | : | PABSM95 | PEICM95 | PECIUM95 | PEPPM95 | PEPSC95 | PEERC95 |
| 19 | : | PABSM95 | PECIUM95 | PEICM95 | PEPPM95 | PEPSC95 | PEERC95 |
| 20 | : | PABSM95 | PEICM95 | PECIUM95 | PEPPM95 | PEERC95 | PEPSC95 |
| 21 | : | PABSM95 | PEICM95 | PECTUM95 | PEPPM95 | PEPSC95 | PEERC95 |
| 22 | : | PABSM95 | PEICM95 | PECTUM95 | PEPSC95 | PEPPM95 | PEERC95 |
| 23 | : | PABSM95 | PEICM95 | PEPSC95 | PECIUM95 | PEPPM95 | PEERC95 |
| 24 | : | PABSM95 | PEICM95 | PECTUM95 | PEPSC95 | PEPPM95 | PEERC95 |
| 25 | : | PABSM95 | PEICM95 | PEPSC95 | PECIUM95 | PEPPM95 | PEERC95 |
| 26 | : | PABSM95 | PEICM95 | PECIUM95 | PEPSC95 | PEPPM95 | PEERC95 |
| 27 | : | PABSM95 | PEICM95 | PECIUM95 | PEPSC95 | PEPPM95 | PEERC95 |
| 28 | : | PABSM95 | PEICM95 | PECIUM95 | PEPSC95 | PEPPM95 | PEERC95 |
| 29 | : | PABSM95 | PEICM95 | PECIUM95 | PEPPM95 | PEERC95 | PEPSC95 |

Como comprobación, se presenta la lista de las mismas variables pero con los porcentajes reales:

| variable listing | | PEICM | PECIU | PEPSC | PEPPM | PEER | PABSM |
|------------------|----|-------|-------|-------|-------|------|-------|
| cases | | 95 | M95 | M95 | 95 | CM95 | 95 |
| 1 | 1 | 5.94 | 16.64 | 2.33 | 2.67 | 2.43 | 69.54 |
| 2 | 2 | 17.42 | 30.37 | 2.41 | 7.18 | 6.00 | 34.73 |
| 3 | 3 | 21.05 | 33.37 | 4.07 | 2.67 | 5.47 | 31.98 |
| 4 | 4 | 25.91 | 24.13 | 3.17 | 7.39 | 3.33 | 34.52 |
| 5 | 5 | 17.31 | 31.71 | 2.68 | 7.32 | 5.66 | 33.75 |
| 6 | 6 | 13.75 | 35.09 | 2.41 | 10.71 | 5.45 | 30.63 |
| 7 | 7 | 22.21 | 24.31 | 3.76 | 9.06 | 5.30 | 32.60 |
| 8 | 8 | 20.06 | 25.30 | 4.94 | 10.28 | 4.55 | 32.91 |
| 9 | 9 | 19.04 | 25.33 | 1.94 | 8.31 | 6.22 | 36.13 |
| 10 | 10 | 16.30 | 28.57 | 2.80 | 7.76 | 5.75 | 37.11 |
| 11 | 11 | 21.32 | 25.05 | 3.41 | 4.84 | 5.71 | 37.47 |
| 12 | 12 | 23.38 | 27.36 | 3.73 | 4.47 | 6.30 | 33.23 |
| 13 | 13 | 26.39 | 16.27 | 4.93 | 7.21 | 5.10 | 39.14 |
| 14 | 14 | 24.21 | 24.14 | 3.44 | 4.17 | 7.10 | 35.55 |
| 15 | 15 | 20.68 | 25.55 | 3.30 | 4.72 | 7.23 | 36.40 |
| 16 | 16 | 21.27 | 27.55 | 3.38 | 7.69 | 4.00 | 34.38 |
| 17 | 17 | 24.15 | 17.15 | 4.19 | 6.44 | 5.56 | 40.98 |
| 18 | 18 | 30.16 | 12.12 | 4.35 | 6.01 | 3.32 | 43.42 |
| 19 | 19 | 20.70 | 24.44 | 4.71 | 8.97 | 3.89 | 36.10 |
| 20 | 20 | 32.29 | 12.88 | 3.85 | 5.64 | 4.65 | 39.89 |
| 21 | 21 | 24.80 | 15.87 | 8.23 | 9.04 | 4.98 | 35.81 |
| 22 | 22 | 28.37 | 13.96 | 7.17 | 3.78 | 3.08 | 43.10 |
| 23 | 23 | 31.18 | 8.14 | 8.52 | 6.01 | 2.13 | 42.74 |
| 24 | 24 | 28.22 | 10.29 | 8.66 | 5.75 | 2.32 | 43.48 |
| 25 | 25 | 34.17 | 4.56 | 14.12 | 3.42 | 1.21 | 41.15 |
| 26 | 26 | 28.31 | 12.13 | 9.80 | 7.60 | 2.21 | 39.09 |
| 27 | 27 | 32.30 | 9.86 | 8.91 | 4.75 | 3.68 | 39.07 |
| 28 | 28 | 27.39 | 15.02 | 5.96 | 5.45 | 3.68 | 41.09 |
| 29 | 29 | 29.89 | 14.67 | 4.06 | 4.48 | 4.13 | 41.3 |

Analizando las cifras se puede ver la correspondencia con la codificación, pero ésta resulta mucho más cómoda y descriptiva.

Sin embargo, esta medición no nos da una idea de las distancias que quedaron entre las diferentes fuerzas políticas y sólo llega a la ordenación. Sin embargo, hay otras variedades de codificación que sí que lo permiten:

En el tipo de codificación que se presenta a continuación, tenemos a la izquierda las secciones censales del distrito 1, a continuación el porcentaje de votos sobre el total de electores de IC en las municipales del 95 y a partir de esa columna que es la que está ordenada de menor a mayor, la posición del resto de las fuerzas políticas en esa sección, respecto a su propia mediana.

Para no proporcionar la gran cantidad de información estadística que representan los cálculos para los 7 distritos y dado que el objeto de la Tesis es presentar posibilidades y no analizar comportamientos electorales en sí, se ha trabajado esta parte únicamente con el distrito 7, que es el que menor número de secciones censales tiene.

El distrito 7:

Para analizar la codificación en este distrito, se presentan, en primer lugar las medianas y su dispersión de las variables electorales:

| | | | | |
|--------------|----------|-------|-----------|-------|
| PEICM95 (2) | : Median | 39.12 | Midspread | 15.79 |
| PECIUM95 (6) | : Median | 1.45 | Midspread | 1.15 |
| PEPSCM95 (4) | : Median | 19.04 | Midspread | 6.35 |
| PEPPM95 (8) | : Median | 2.32 | Midspread | 0.55 |
| PEERM95 (10) | : Median | 0.21 | Midspread | 0.12 |
| PABSM95 (73) | : Median | 34.85 | Midspread | 7.98 |

>>>display 2 6 4 8 10 73

Seguidamente, se presentan dos tipos de codificación respecto a la mediana de cada variable:

Esta primera codificación tiene de izquierda a derecha las secciones censales, los resultados de IC(numéricos) y la posición en que quedan las demás fuerzas políticas y la abstención respecto de sus propias medianas. Así, en la primera fila vemos que CIU quedó algo por encima de su mediana en la sección 3 (la mediana de CIU en el distrito es 1.45 y el resultado real que obtuvo 3.09). El PSC se mantiene muy cerca de su mediana (19.04 con un resultado real de 18.06). El PP también queda cerca de su mediana (2.32 con un resultado real de 2.81). ERC también se mantiene cerca de su mediana (0.21 con un resultado real de 0.22). Y, finalmente, la abstención queda algo por encima de su mediana (34.85 con un resultado real de 43.53).

```
Showing : PEICM95 (2)

Legend for coded values: (HI far)@ # + - = & (LO far)

cases    PEICM95  PEI  PEPS  PEPP  PEER  PABS
          UM95  CM95  M95  CM95  M95
3 3      31.73  +
1 1      34.28  - +
4 4      38.87  # @
5 5      39.36
6 6      50.08  - - =
2 2      50.81  -

>>>show 2 6 4 8 10 73 coded
```

2) Esta segunda codificación es muy parecida a la anterior pero con otros símbolos.

```
Showing : PEICM95 (2)

Legend for coded values: (HI far)a l t . s y m (LO far)

cases    g#  PEICM95  PEI  PEPS  PEPP  PEER  PABS
          g#  UM95  CM95  M95  CM95  M95
3 3      31  31.73  t  .  .  .  t
1 1      34  34.28  s  t  .  .  .
4 4      38  38.87  .  .  l  a  .
5 5      39  39.36  .  .  .  .  .
6 6      50  50.08  .  s  s  y  .
2 2      50  50.81  .  .  .  .  s

>>>show 2 6 4 8 10 73 coded "alt.symbols"
```

A continuación se presenta otro tipo de codificación más preciso en que los símbolos indican a cuantas 0.5 desviaciones por encima o por debajo están las fuerzas políticas de su mediana en el distrito:

```
variable listing
Units of 1/2.00 midspread
cases  q#  PEICM95 PECTUM95 PEPSCM95  PEPPM95 PEERC95  PABSM95
1 1    34      -      -      +
2 2    50      +      -
3 3    31      ++
4 4    38      +++ $++++
5 5    39      +
6 6    50      +      +      -      -      -
```

>>>list 2 6 4 8 10 73 coded

Así, en la primera fila se puede ver que en la sección 1, IC quedó muy cerca de su mediana (39.12, no sobrepasando una mitad de su desviación $15.79/2=7.85$, pues la diferencia entre la mediana y el resultado real, no alcanza esta cifra $39.12-34.28=4.84$). CIU quedó a una distancia por debajo de su mediana cercana a la mitad de su desviación ($1.15/2=0.575$, siendo la diferencia entre la mediana y el resultado real $1.145-0.79=0.66$, que sobrepasa un poco el 0.575). El PSC quedó por encima de su mediana a una distancia de aproximadamente media desviación ($6.35/2=3.175$, siendo la diferencia entre la mediana y el resultado real $19.04-22.79 = -3.75$). Pasando a otros casos más llamativos, tenemos que ERC en la sección 4 quedó a unas 15 veces la mitad de su desviación, por encima de su mediana. El signo \$ representa que hay que multiplicar por dos las + que aparecen y que cada una de ellas representa $\frac{1}{2}$ de la desviación de la variable respecto a su mediana. Haciendo cálculos, resulta que la desviación de ERC es 0.12, que dividida por 2 da $0.12/2=0.06$. La diferencia entre la mediana y el resultado real es $0.21-0.92=-0.71$. Si tomamos esta cifra en valor absoluto y la dividimos por 0.06, resulta que ERC quedó a una distancia por encima de unas 12 veces la mitad de su desviación.

Finalmente se pueden ver los resultados reales para efectuar comprobaciones:

```
variable listing
cases  PEICM PEICI PEPSC PEPP PEER PABSM
      95   UM95 M95   M95  CM95 95
1 1    34.28 0.79 22.79 2.26 0.20 38.31
2 2    50.81 1.37 14.65 2.25 0.16 29.79
3 3    31.73 3.09 18.06 2.81 0.22 43.53
4 4    38.87 1.18 20.03 3.53 0.92 34.69
5 5    39.36 1.54 21.01 2.38 0.28 35.01
6 6    50.08 2.33 14.46 1.87 0.00 30.33
      >>>list 2 6 4 8 10 73
```

Por tanto, hemos podido efectuar mediciones a nivel descriptivo y el investigador puede extraer conclusiones de las mismas.

Si se desean cifras exactas, se puede pedir la diferencia entre el porcentaje de votos sobre electores de IC en las municipales del 95 y el resto de las fuerzas políticas en el mismo tipo de unidades para las 6 secciones del distrito 7. Los resultados son:

| Showing : PEICM95 (2) | | | | | | | |
|-----------------------|-------|-------|-------|-------|-------|--------|--|
| cases | PEICM | CIUIC | PSCIC | DPPIC | ERCIC | ABSICM | |
| | 95 | M95 | M95 | M95 | M95 | 95 | |
| 3 3 | 31.73 | 28.63 | 13.67 | 28.92 | 31.51 | -11.80 | |
| 1 1 | 34.28 | 33.50 | 11.49 | 32.02 | 34.09 | -4.03 | |
| 4 4 | 38.87 | 37.70 | 18.85 | 35.34 | 37.96 | 4.19 | |
| 5 5 | 39.36 | 37.82 | 18.35 | 36.97 | 39.08 | 4.34 | |
| 6 6 | 50.08 | 47.74 | 35.61 | 48.21 | 50.08 | 19.75 | |
| 2 2 | 50.81 | 49.44 | 36.15 | 48.55 | 50.64 | 21.01 | |

Las secciones censales han quedado ordenadas de menor a mayor porcentaje sobre electores para IC y las columnas que siguen por la derecha, son las diferencias de los porcentajes de cada fuerza política con la primera. Por tanto, en la sección 3 IC tuvo una diferencia positiva (por encima) de todas las fuerzas políticas bastante elevada, si bien, la abstención superó notablemente al porcentaje sobre electores. Las mejores secciones censales en cuanto a resultados para IC, en el distrito 7, fueron la 2 y la 6, sobrepasando el porcentaje de CIU en un 49.44%, el del PSC en un 36.15%, el del PP en un 48.55%, el de ERC en un 50.64% y la abstención en un 21.01%.

El siguiente paso que sigue en el esquema del método científico es el de la aceptación o recusación (provisional) de los resultados de la observación, descripción y medición como hechos o realidad. En este caso, el investigador acepta los resultados y sigue adelante.

El quinto paso es el de la generalización inductiva (provisional) de los hechos aceptados en la cuarta fase como "hipótesis fáctica". De nuevo el investigador, recuerda la pregunta objeto de investigación: ¿cuáles son las secciones censales dentro de cada distrito electoral en que cada fuerza política superó la mediana de IC en las elecciones municipales de 1995 en este municipio?

En este momento, se cuestiona si esta pregunta es o no una "Hipótesis Fáctica" y, por supuesto, llega a la conclusión de que no lo es. No está demostrando ninguna teoría ni tratando de explicar unos hechos, de manera que por más que se empeñe en seguir un método científico, no está realmente haciendo algo científico. Se está quedando en una mera exposición doctrinal de hechos, o dicho de otra forma, en una descripción lo más precisa posible, pero nada más.

Por tanto, queda demostrado que a partir de este punto ya no se puede seguir con el resto de etapas del método: puesto que no hay hipótesis fáctica no puede haber intento de explicación de la misma, ni deducción lógica, ni comprobación, ni, en definitiva, aportación científica alguna en el sentido estricto de la palabra.

Entonces, ¿qué ha estado haciendo el investigador?. Según nuestro punto de vista, el investigador ha formulado una pregunta que cae dentro del dominio de la geografía

electoral y ha descrito lo más claramente posible y con las diversas herramientas de que ha hecho uso, una situación del reparto de votos en un territorio. Para ello se ha valido del método científico, pero no puede desarrollarlo hasta el final, de manera que se puede afirmar que el análisis exploratorio sirve para cubrir el inicio de una investigación, prácticamente hasta la cuarta fase, y que, si en ese punto nos informa de que los datos pueden dar más de sí, es recomendable: o bien pararse en la descripción, o bien plantear una hipótesis y seguir hasta el final.

Tukey ya lo explicó en su libro "Exploratory Data Analysis": se trata de efectuar una primera descripción y de simplificar al máximo la información que contienen los datos. Este análisis no tiene nada que ver con el confirmatorio y, por tanto, sólo puede cubrir la primera parte de cualquier aplicación del método científico o dejamos en el estadio de la simple descripción cualitativa.

Muchas veces se parte de un planteamiento que luego es imposible demostrar cuantitativamente. Tukey pone de manifiesto que es importante, a la vista de la simple observación, cambiar de punto de vista si ello es necesario. En lugar de mantenernos firmes en la creencia de que muchas de las cuestiones clave que nos planteamos, en nuestra parcela de estudio, tarde o temprano exigirán una respuesta de "¿en qué cantidad?", también podemos cambiar de actitud y preguntamos simplemente: "¿en qué dirección?".

Las siguientes herramientas que se presentan van un poco más allá de la codificación, pero siguen siendo no inferenciales y por tanto, tampoco llegan más allá de la cuarta etapa del método.

Estadísticos descriptivos

Cualquier estadístico o científico social conoce los principales estadísticos descriptivos: moda, mediana, media, cuartiles, etc. Pero, de nuevo, se trata no de definirlos y presentarlos, sino de ver su empleo en el ámbito de la exploración, que puede resultar diferente del que efectuamos habitualmente.

Tukey y los seguidores de las técnicas exploratorias se basan más en la mediana que en cualquier otro estadístico de tendencia central. La razón es que ésta no se ve afectada por los valores extremos, mientras que la media sí. La media no es un estadístico robusto y la mediana sí que lo es.

Cuando analizamos un conjunto de datos y deseamos obtener un valor típico que represente a toda la distribución, si la distribución no es simétrica, la media no es interesante porque carece de representatividad. En cambio, la mediana es resistente, tiene lo que Tukey llamaba un "high break down point", algo así como el número de valores arbitrariamente grandes que pueden añadirse a una serie ordenada hasta que su resumen descriptivo se rompe y no sirve para nada.

La aplicación EDA ofrece estadísticos exploratorios diferentes de los que podemos hallar en cualquier otra aplicación estadística no especializada. El vocabulario que Tukey emplea en éste ámbito es el siguiente:

LETTER VALUES: son letras que se identifican con determinados puntos de corte de la distribución.

La O representa los extremos, la M la mediana, la H los cuartiles, la E los octavos, etc. Mediante el análisis de estos puntos se puede observar con mucho detalle una distribución.

```

O-----M-----O
---H-----H---
25% 25% 25% 25%

```

H-Spread = recorrido intercuartilar

O-Spread = Rango

MID-H = Mid Hinge = punto medio del recorrido intercuartilar Este punto sirve para saber si la distribución es simétrica o no, ya que en caso de serlo, coincide con la mediana.

MID-O = punto medio del rango

TRIMEAN = (HL + 2M + HU)/4

Mediante estos puntos podemos observar las distribuciones y usar estas herramientas junto a determinados gráficos.

Para comenzar, vamos a conocer alguna utilidad de los resúmenes numéricos más sencillos que proporciona el EDA. Se trata de una pequeñas tablas donde se puede ver en la parte superior externa la mediana de la variable. En la primer fila de la tabla (H) el primer y tercer cuartil (Hinges) y en la fila inferior (O) el mínimo y el máximo.

Seguimos trabajando con los mismos datos electorales y comenzamos por efectuar una comparación de estas tablas para todas las fuerzas políticas en las tres últimas elecciones en las secciones censales del distrito 1:

| | |
|------------------------|-------------|
| 29 cases | |
| Summary: PECIUM95 (6) | |
| 24.13 | |
| +-----+ | |
| H | 13.96 25.55 |
| O | 4.56 35.09 |
| 29 cases | |
| Summary: PECIUA95 (14) | |
| 42.40 | |
| +-----+ | |
| H | 33.91 46.88 |
| O | 18.69 54.36 |
| 29 cases | |
| Summary: PECIUG96 (28) | |
| 39.86 | |
| +-----+ | |
| H | 31.96 45.38 |
| O | 13.08 52.77 |

Con la simple observación de las cifras ya se ve que en este distrito, CIU tiene su máxima fuerza en las autonómicas, seguida de las generales y de las municipales. Los cuartiles son muy diferentes en las municipales de las autonómicas y generales, alcanzando CIU cifras más potentes en las dos últimas elecciones en el 25%, 50% y 75% de la distribución. El rango es amplio en todas las elecciones, pero se parte de un mínimo distanciado del de las municipales y se alcanzan máximos superiores a los de éstas.

| | |
|------------------------|-------------|
| 29 cases | |
| Summary: PEPSCM95 (4) | |
| 4.06 | |
| +-----+ | |
| H | 3.30 5.96 |
| O | 1.94 14.12 |
| 29 cases | |
| Summary: PEPSCA95 (16) | |
| 7.64 | |
| +-----+ | |
| H | 5.07 11.02 |
| O | 2.80 24.57 |
| 29 cases | |
| Summary: PEPSCG96 (26) | |
| 18.32 | |
| +-----+ | |
| H | 14.34 25.64 |
| O | 7.12 45.93 |

El PSC obtiene sus mejores resultados en este distrito en las generales, si bien, los porcentajes son discretos para el primer y tercer cuartil. Las autonómicas le van peor que las generales y las municipales presentan resultados muy bajos.

| | |
|-----------------------|------------|
| 29 cases | |
| Summary: PEPPM95 (8) | |
| 6.01 | |
| +-----+ | |
| H | 4.72 7.69 |
| O | 2.67 10.71 |
| 29 cases | |
| Summary: PEPPA95 (18) | |
| 6.92 | |
| +-----+ | |
| H | 5.20 8.80 |
| O | 3.01 11.18 |
| 29 cases | |
| Summary: PEPPG96 (30) | |
| 12.01 | |
| +-----+ | |
| H | 9.74 13.63 |
| O | 6.56 19.16 |