

6. Aportación y propuesta para al estudio de noticias en prensa con técnicas estadísticas multivariantes: Análisis cluster

En este estudio sobre discapacidad queremos realizar una propuesta de aplicación del procedimiento estadístico multivariante cluster, también conocido como análisis de conglomerados, cuyo objetivo es realizar clasificaciones jerárquicas que nos muestren las semejanzas entre los datos, con el objetivo de determinar grupos con características semejantes.

La propuesta que realizamos tiene como objetivo clasificar los periódicos en grupos según el tratamiento que dan a la discapacidad, y todo ello a partir de una base de artículos.

1. Se parte de la base de datos de los 1.539 artículos analizados para este estudio. Cada artículo es considerado un caso en esta base, en la cual se toman como variables las de la ficha de estudio (ver anexo número 1).
2. Se calcula de las variables más relevantes, según la práctica periodística, sus medias (si son variables en escala) o porcentajes (si son variables nominales). En nuestro caso se quería clasificar en función del tratamiento –mejor o peor– que los medios realizan sobre la discapacidad. Consideramos que las siguientes variables son indicadores del tratamiento:
 - Número de artículos (suponemos que un mayor número es un indicador positivo).
 - Tamaño medio de los artículos (se calcula la media, y no los casos, ya que los diferentes periódicos han publicado diferente número de noticias).

- En qué porcentaje las personas con discapacidad aparecen como activas vs. pasivas (se calcula el porcentaje, ya que los diferentes periódicos han publicado diferente número de artículos).
- Porcentaje de los artículos en páginas impares.
- Número medio de fuentes.
- Porcentaje de artículos publicados en fin de semana.
- Número de portadas y contraportadas.

Para la variable número de artículos se toma como indicador el número de casos. También seguimos este procedimiento en la variable portada, ya que sólo había 21 contraportadas y portadas en total y poner éstas en porcentaje, según el número de informaciones publicadas en el periódico, ofrecería cifras muy bajas.

Proponemos que se tome el porcentaje como valor ya que no existe un número similar de artículos por periódico (este cálculo se puede realizar con la prueba Chi-cuadrado). De todas formas, aunque sí existiera un número similar de artículos, el poner el número de casos en vez del porcentaje no mejoraría el estudio estadístico.

Para las variables medidas en escala (cuantitativas de intervalo o de razón) se pueden tomar como indicadores la media, la mediana y la desviación típica. En nuestro caso se optó por la media, ya que era el indicador que más diferencias mostraba entre los periódicos. La desviación típica también podría ser un buen índice en otros estudios ya que, por ejemplo, si tomamos la variable número de fuentes, la media nos indicaría el periódico que por término medio acude a más fuentes, pero es importante tener en cuenta que una alta variabilidad (es decir alta desviación típica) sería indicador de una gran desigualdad en el tratamiento de los artículos, ya que en unos casos habría muchas fuentes y en otros pocas. Así, las medidas de variabilidad podrían ayudar a diferenciar a periódicos con medidas de tendencia central similares.

3. Se realiza una base de datos con los periódicos como casos, y como variables, los estadísticos obtenidos anteriormente. En nuestro caso tenemos una base con 25 periódicos y las 9 variables anteriormente citadas:
4. Cálculo de análisis de conglomerados. Para ello nos servimos de la opción Conglomerados Jerárquicos, que está en el programa SPSS. Se encuentra en la opción clasificar. En nuestro caso realizamos un primer análisis sólo con la variable número de artículos. Las opciones fueron:

Gráfico 44

conglomeradosperiodicos - Editor de datos SPSS

Archivo Edición Ver Datos Transformar Análisis Gráficos Utilidades Ventana ?

25 - dicapoc

	perio	finde	noticia	var0	tamaño	discapac	impar	fuentes	portadas
1	D. ÁVILA	329	81	2	247	12,300	368	136	0
2	D. LEÓN	184	90	2	194	29,100	101	160	0
3	CRÓNICA-MUNDO-LEÓN	239	93	2	178	27,200	54	151	0
4	OPINIÓN-CORREO-ZAMOR	333	58	2	203	43,600	155	127	0
5	ABC	.	1	1	300	.	.	500	0
6	NORTE-ZAMORA	283	64	2	214	36,400	136	120	0
7	PRENSA-ZAMORA	167	6	1	250	20,000	.	100	0
8	GACETA-SALAMANCA	208	161	2	193	31,100	366	127	2
9	ADELANTO-SALAMANCA	261	140	2	188	25,900	429	136	2
10	TRIBUNA-SALAMANCA	291	114	2	205	33,900	445	132	4

Gráfico 45

Análisis de conglomerados jerárquico

Variables:

PERIODICO [periodic]
 PERIODICO [var0000]
 Porcentaje en fineser
 número de noticias: dic
 tamaño de la noticia (r
 LAS PERSONAS CON
 IMPAR (%) [impar]
 numero medio fuentes
 N° de portadas + contr

número de noticias [notic]

Etiquetar los casos mediante:

PERIODICO [perio]

Conglomerar

Casos Variables

Mostrar

Estadísticos Gráficos

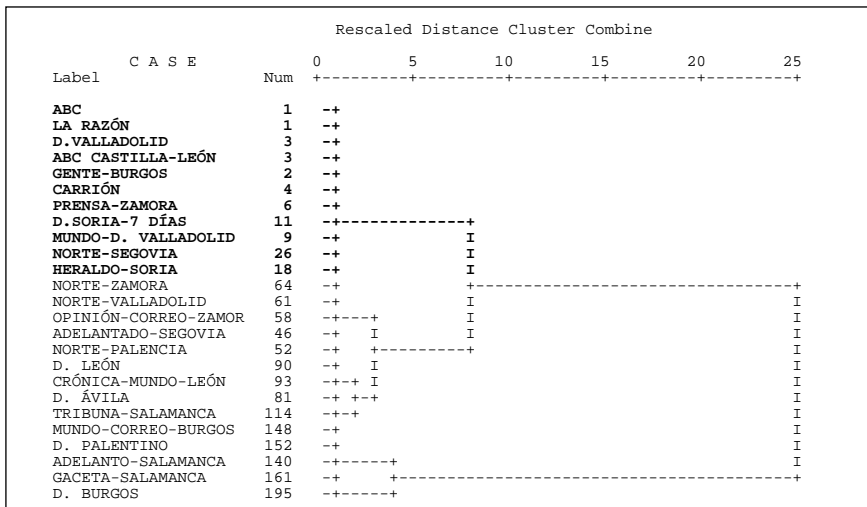
Aceptar
 Pegar
 Restablecer
 Cancelar
 Ayuda

Estadísticos... Gráficos... Método... Guardar...

Gráficos: dendograma;
Método: vinculación Inter-grupos,
Medida: Intervalo, distancia euclídea al cuadrado

Obteniéndose el siguiente dendograma:
 Se ha querido presentar este dendograma por un doble motivo:

Gráfico 46



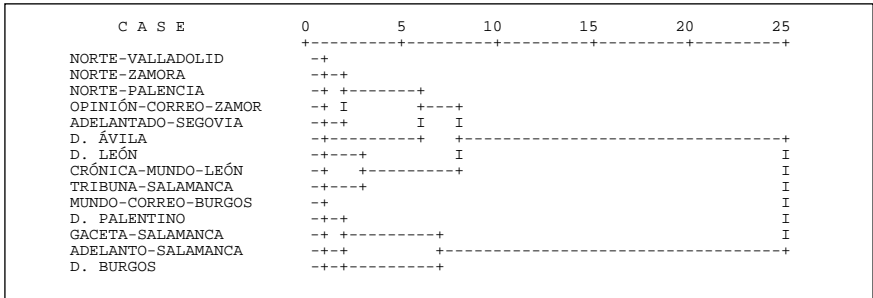
1. Permite comprender intuitivamente el objetivo del procedimiento estadístico de conglomerados jerárquicos, ya que se observa cómo agrupa los periódicos en grupos jerárquicos según el número de artículos publicados sobre discapacidad. Se observa cómo se podrían establecer tres grupos claramente:
 - Un grupo de periódicos que casi no publican sobre discapacidad, que son 11 de los 25, y que se muestran en negro (entre uno y 26 artículos).
 - Un grupo que publica regularmente sobre discapacidad, que a su vez podría subdividirse en dos: un primer grupo que publica entre 46 y 64 artículos; y otro de cuatro que publica entre 80 y 114 artículos.

- Existe un tercer grupo, que es el que más publica, con la indicación que dentro de este grupo existe un periódico que destaca: Diario de Burgos con 194 artículos. Los otros cuatro oscilan entre 140 y 160.

Nos permite excluir aquellos periódicos que por su baja cantidad de artículos publicados distorsionarían posteriores análisis donde se trabaja con distintas variables, ya que las medias y porcentajes en éstos podrían situarlos dentro del grupo de los periódicos con mejor tratamiento a la discapacidad. Así pues, en posteriores análisis el primer grupo será excluido de los análisis jerárquicos.

2. Un primer análisis de este primer dendograma nos hace pensar que los periódicos de un mismo ámbito geográfico tienden a parecerse y no influyen otras variables como podrían ser la orientación ideológica (progresista vs. conservador) o la pertenencia a grupos editoriales (Promecal, Grupo Zeta, etc.). Así, parece ser que los periódicos en Castilla-León lejos de intentar diferenciarse en contenidos de su vecino de capital, lo toman como referencia, es decir, tratan con extensión parecida los temas.
3. Realizamos un segundo análisis jerárquico con variables de agrupación que nos indican cantidad: número y tamaño medio de los artículos. Se obtiene el siguiente dendograma, que matiza los resultados anteriores:

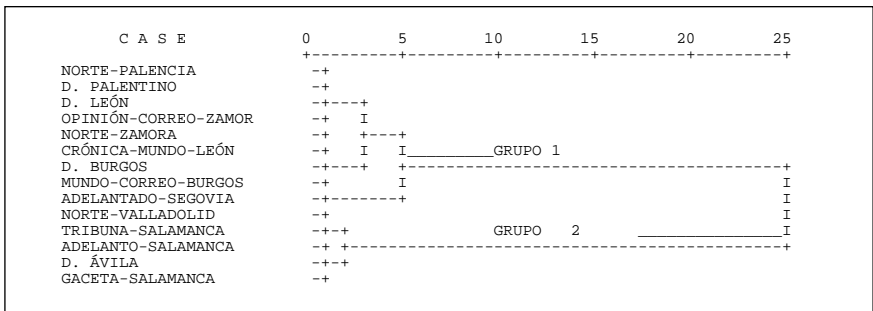
Gráfico 47



En este caso, Diario de Ávila baja al grupo de los que menos cantidad dedican a la discapacidad, y El Adelanto de Salamanca sube un poco, al situarse como el segundo periódico que más cantidad de información ofrece.

Presentamos otro dendograma realizado sólo con variables que son indicadores no de cantidad, sino de calidad de la información: número medio de fuentes, número de portadas y contraportadas, porcentaje de artículos en páginas impares y porcentaje de artículos donde las personas con discapacidad aparecen como activas.

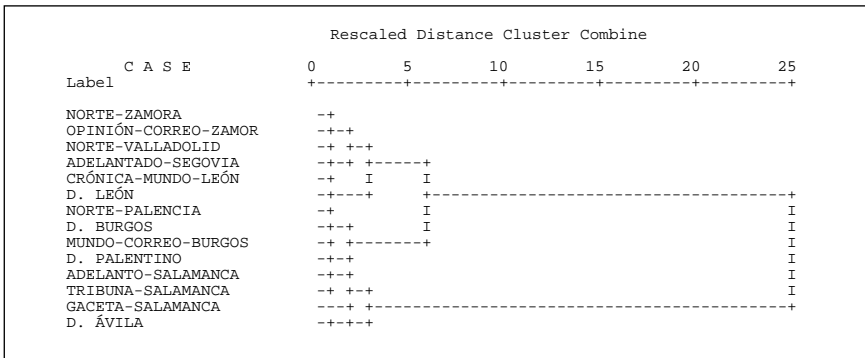
Gráfico 48



Este dendograma es bastante diferente al anterior. Muestra dos grupos claramente diferenciados. Para interpretar esta salida hay que acudir a la base de datos y observar qué grupos obtenían las puntuaciones más elevadas en dichas variables. Teniendo en cuenta estas variables, y tal y como ya se comentó en el capítulo anterior al presentar los resultados variable por variable, el grupo de periódicos de la provincia de Salamanca serían, conjuntamente con Diario de Ávila, los que informan con más calidad.

Por último, realizamos un análisis jerárquico con los dos grupos de variables: las de cantidad y calidad. Hay que tener en cuenta que al ser más numerosas las variables que introducimos relacionadas con la calidad (cuatro) que con la cantidad (dos), pesará más la cantidad que la calidad, tal y como se ha medido aquí de forma exploratoria.

Gráfico 49



En este tercer dendograma se observa un grupo en el que se encuentra Diario de Burgos, que en el primero aparecía como más destacado. Así, este nuevo grupo no realiza los artículos con tanta calidad como el grupo de Salamanca-Ávila, pero se diferencia de los otros en que sí le dedica mucho espacio al tema de la discapacidad, por lo que se demuestra que están también muy concienciados.

