

Una Extensión Del Análisis De Tablas Múltiples De Variables Cualitativas

Juan Ignacio Modroño
Universidad del País Vasco

Karmele Fernández-Aguirre
Universidad del País Vasco

M. Isabel Landaluce
Universidad de Burgos

E-mail: imh@alcib.bs.ehu.es , kf@alcib.bs.ehu.es , iland@ubu.es

Palabras Clave: Análisis Factorial Múltiple, Análisis de Correspondencias Múltiples, Variables Cualitativas

Resumen

El núcleo central de este trabajo lo constituye el diseño experimental de una técnica que, basada en la filosofía de los recientes métodos de análisis de tablas múltiples desarrollados en el seno de la escuela francesa de análisis de datos, permita realizar estudios comparativos de matrices de naturaleza cualitativa. Estas matrices deben tener el mismo número de filas, no existiendo restricción alguna sobre el número de columnas.

Para ello, se ha realizado un ensayo exploratorio de las encuestas que sobre desigualdad social ha realizado el ISSP (International Social Survey Programme), durante el año 1992, en España e Italia.

1. Introducción

Este documento está inmerso en el contexto de tablas múltiples. En este campo, destacan dos metodologías, tanto por sus posibilidades de aplicación como por su riqueza interpretativa: El Análisis Factorial Múltiple (Escofier y Pagès [1988]) y la

metodología STATIS (Escoufier, Dazy *et al.* [1996] y Lavit [1988]). La primera, simbolizada como AFM, permite un análisis comparativo de un grupo de tablas referidas a los mismos individuos (filas) y a las mismas ó diferentes variables (columnas), siendo las variables cuantitativas ó categóricas (con la condición de que dentro de cada tabla todas las variables deben de ser del mismo tipo). Por otro lado, la metodología STATIS se presenta en tres versiones diferentes: Analisis Triadique de Jaffrenou, adecuado para el estudio de conjuntos de tablas relacionados a los mismos individuos y a las mismas variables, STATIS, diseñado para el análisis de tablas de datos cuantitativos, conteniendo información de los mismos individuos sobre las mismas ó diferentes variables, y STATIS DUAL, el cual permite la comparación de variables cuantitativas referidas a las mismas variables y a diferentes grupos de individuos.

Como puede fácilmente verse, estas metodologías permiten el análisis de diferentes tipos de tablas múltiples. Sin embargo, existen limitaciones cuando se trata de conjuntos de variables de naturaleza cualitativa: ninguno de los procedimientos antes citados está diseñado para el análisis de grupos de tablas de datos conteniendo información sobre distintos grupos de individuos, sobre quienes las mismas variables nominales han sido medidas. Este tipo de variables constituyen una parte fundamental en encuestas realizadas en todos los campos del saber humano, especialmente en las ciencias sociales.

2. Objetivo y Metodologia

Nuestro objetivo en este trabajo es el diseño experimental de un método que, basado en las metodologías antes expuestas, nos permita cubrir el vacío existente en el análisis de tablas de variables cualitativas. Para ilustrar este método, hemos elegido la Encuesta de Desigualdad Social, proveniente de la base de datos del International Social Survey Programme (ISSP), recopilada por el Zentralarchiv für Empirische Sozialforschung, Universidad de Colonia, Alemania. Dicha encuesta está realizada en 1992, y las dos tablas de datos escogidas como ejemplo de aplicación, son las correspondientes a las encuestas realizadas en España e Italia, respectivamente.

El procedimiento que proponemos tiene su objetivo final en el análisis de tablas múltiples referidas a los mismos individuos (filas), no siendo necesario la igualdad de dimensiones en columnas. Se realiza en dos pasos:

Paso 1: Se realiza un Análisis de Correspondencias Múltiples sobre cada una de las tablas objeto del análisis, en este caso, dos países. Cada una de estas tablas tiene como filas los individuos encuestados en cada país, y por columnas las variables objeto de la encuesta que han sido elegidas para el análisis. No es necesario que el número de individuos que aparecen en cada tabla sea el mismo, pero sí que las encuestas sean homogéneas en el sentido de que en el análisis han de incurirse exactamente el mismo número de cuestiones que además deben contener el mismo número de modalidades, para que el análisis posterior tenga sentido. De cada uno de los ACM correspondientes, se extrae un número reducido de factores interpretables con inercia destacable. No es formalmente necesario extraer el mismo número de factores en cada tabla, aunque creemos que puede ser una buena práctica de cara a la interpretación futura de los elementos del análisis.

Paso 2: Se yuxtaponen las tablas de los factores obtenidos en el paso anterior, formando una gran tabla que, por filas, tiene las modalidades de las variables elegidas para el análisis, que, necesariamente han de ser las mismas. Por columnas, tiene los factores elegidos en el paso anterior para todos los países, formando teóricos grupos geográficos, no necesariamente de la misma dimensión. Entonces, se realiza un AFM de esta segunda tabla, de naturaleza ya cuantitativa, donde cada grupo representa un país, para obtener una vista completa y simplificada de diferencias y similitudes entre las tablas estudiadas.

Este procedimiento de comparación de tablas cualitativas creemos que tiene la ventaja de dejar que sean los propios datos los que nos sugieran cuales son los factores más relevantes para cada país, cuál es su importancia relativa de cara a un análisis global y cómo de próximos están unos de otros (en definitiva, unas tablas de otras).

3. Datos y Resultados

Como se ha comentado antes, la encuesta objeto del estudio es la Encuesta de Desigualdad Social del ISSP y, dentro de ella, datos para dos países, España e Italia. Hemos elegido un conjunto de 18 variables cualitativas todas con el mismo número de modalidades (3 ó 5, según el caso) para los dos países hasta un total de 92. Las variables originales son del tipo “¿Cuánto cree usted que debería ganar un ...?” y “¿Cuánto cree usted que realmente gana un ...?” donde los puntos suspensivos son reemplazados por diversas profesiones, incluyendo las correspondientes a varios estratos sociales como gerentes, jueces, etc, por un lado, y agricultores, dependientes de almacén, obreros, etc., por otro. Dichas variables en la encuesta original eran continuas, pero fueron recodificadas a cualitativas con un criterio de tratar de preservar en lo posible la igualdad de clases ó tramos en que quedaba dividida la muestra. Así, un “1” indica un valor bajo y un “5” (ó un “3”, si es una continua recodificada en 3 clases) los valores más altos. En un principio, realizamos el ACM inicial sobre estas variables y sobre el resto disponible en la encuesta, pero decidimos prescindir del resto puesto que su importancia relativa era, a posteriori, muy pequeña.

En los ACM iniciales de los dos países, encontramos 4 factores interpretables que son los que finalmente pasamos a la segunda fase del análisis junto con el quinto factor para cada país, pensando, al incluir este último que es quizás mejor incluir alguno más de los necesarios de cara a facilitar la comparación de los países en esa segunda etapa. La extracción de 5 factores, recoge el 30.24% de la inercia para España y el 33.9% para Italia, siendo porcentajes bastante similares. Notamos que la interpretación de los dos primeros para España e Italia son similares (el primero refleja la no respuesta y el segundo un factor talla de las respuestas), aunque esta similitud no parece trasladarse al tercero y cuarto. En el caso de España, el tercero parece oponer categorías extremas a las medias, y el cuarto parece asociar grupos de individuos que piensan que determinadas profesiones “de cuello blanco” ganan ó deberían de ganar mucho a la vez que otras profesiones, principalmente dependientes de tiendas ó almacenes ganan ó deberían ganar poco y viceversa. Para Italia, estos dos factores parecen estar invertidos.

En la segunda fase, realizamos el AFM sobre una tabla en la que aparecen, por columnas, los cinco factores que hemos extraído para cada uno de los dos países. En la

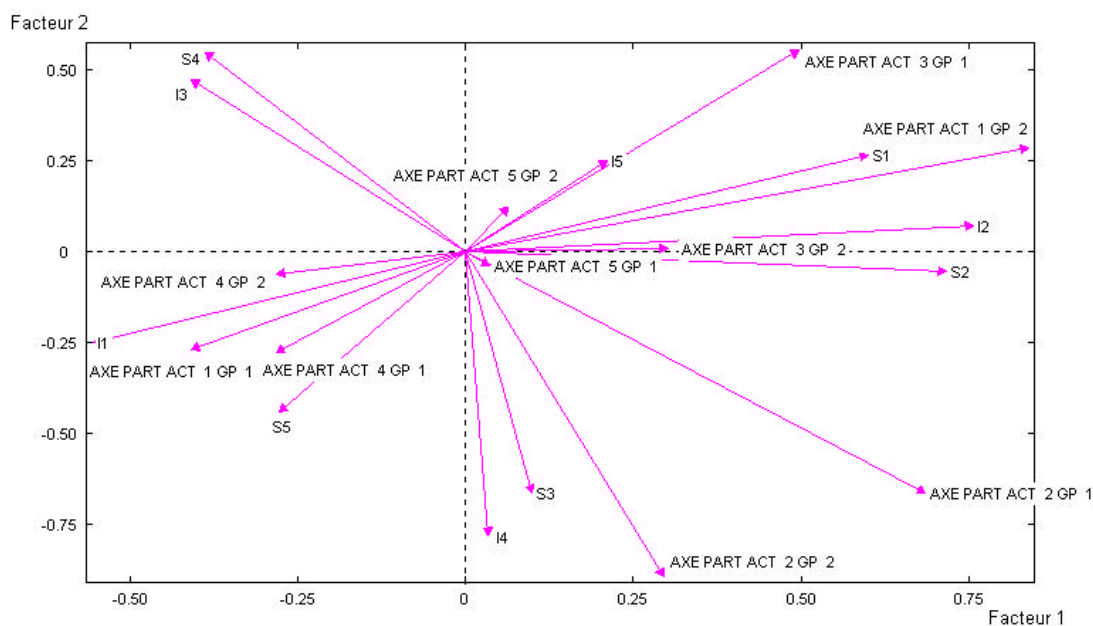
matriz de correlaciones entre factores se aprecia una muy elevada correlación entre los dos primeros factores de los dos países, aunque muy baja entre el tercero y el cuarto. Pero sí es cierto que el tercero de España está muy correlacionado con el cuarto de Italia y viceversa, que era lo que se apuntaba en el ACM separado anterior. En el caso del quinto factor, la correlación ya va decreciendo en importancia, alrededor de -0.5 , que es ya un valor no muy significativo. Este AFM no hace más que resaltar la similitud entre ambos países al quedar, en la proyección sobre los dos primeros ejes globales, siempre los factores similares originales par los dos países muy próximos.

MATRICE DES CORRELATIONS ENTRE COLONNES ACTIVES					
	S1	S2	S3	S4	S5
I1	-0.95	0.03	-0.13	-0.04	-0.04
I2	0.22	0.86	-0.03	-0.13	-0.22
I3	-0.04	-0.11	-0.04	0.77	0.01
I4	-0.08	0.09	0.75	-0.13	0.25
I5	0.05	0.07	-0.03	-0.01	-0.47

Nota: Las variables S1,...S5 representan los 5 factores extraídos inicialmente para España, mientras que las variables I1,...,I5 los 5 correspondientes a Italia.

En los gráficos siguientes, mostramos las coordenadas y los factores del ACM proyectados sobre los dos primeros ejes del AFM.

COORDONNEES ET AIDES A L'INTERPRETATION DES VARIABLES ACTIVES																
AXES 1 A 5																
VARIABLES		COORDONNEES					CONTRIBUTIONS					COSINUS CARRES				
IDENT./GROUPE	DISTO	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
GROUPE 1																
S1	1.00	0.6	0.3	0.7	-0.1	0.1	15.2	3.4	25.4	0.5	0.3	0.36	0.07	0.53	0.01	0.00
S2	1.00	0.7	-0.1	-0.5	0.4	0.2	21.6	0.1	9.7	10.3	1.9	0.51	0.00	0.20	0.17	0.03
S3	1.00	0.1	-0.7	0.4	0.4	-0.3	0.4	20.7	8.5	9.1	4.6	0.01	0.44	0.18	0.15	0.07
S4	1.00	-0.4	0.5	0.3	0.6	0.0	6.3	14.1	5.0	19.3	0.0	0.15	0.30	0.10	0.32	0.00
S5	1.00	-0.3	-0.4	0.4	-0.1	0.5	3.2	9.1	9.2	0.2	17.9	0.08	0.19	0.19	0.00	0.26
ENSEMBLE							46.8	47.4	57.8	39.4	24.7	0.22	0.20	0.24	0.13	0.07
GROUPE 2																
I1	1.00	-0.6	-0.2	-0.8	0.2	0.0	15.0	3.4	31.2	2.8	0.1	0.31	0.06	0.57	0.04	0.00
I2	1.00	0.8	0.1	-0.3	0.4	0.3	27.9	0.3	5.8	11.9	7.9	0.57	0.01	0.10	0.17	0.10
I3	1.00	-0.4	0.5	0.2	0.7	0.0	8.1	12.2	2.1	33.3	0.2	0.17	0.22	0.04	0.48	0.00
I4	1.00	0.0	-0.8	0.2	0.4	-0.3	0.1	33.3	2.4	12.6	6.1	0.00	0.61	0.04	0.18	0.08
I5	1.00	0.2	0.3	-0.1	0.0	-0.9	2.2	3.4	0.6	0.0	61.1	0.04	0.06	0.01	0.00	0.75
ENSEMBLE							53.2	52.6	42.2	60.6	75.3	0.22	0.19	0.15	0.17	0.19



4. Conclusiones

Somos conscientes que en esta nueva aproximación, los individuos originales pierden completamente su identidad. Sin embargo, es bien sabido que, en este tipo de encuestas, los individuos no tienen interés en sí mismos, siendo verdaderamente importantes las clases definidas a través de ellos en su diferente caracterización y/o opinión. De hecho, las categorías, que constituyen los nuevos individuos en la nueva matriz analizada, definen, de alguna forma, estas clases.

Los resultados, aplicados a la encuesta de desigualdad social, son esperanzadores. El hecho de poder comparar tablas cualitativas, aún con la pérdida de información que supone utilizar sólo unos pocos factores principales, consideramos que es de interés, ya que antes sólo se había realizado para tablas cuantitativas. Aunque no se aumente la información disponible en las dos encuestas, el poder analizar las dos tablas conjuntamente favorece la interpretación de los ejes parciales. El requisito inicial de contar con el mismo número inicial de variables y modalidades no nos parece excesivo siempre que se trate de comparar la misma encuesta sobre varias poblaciones diferentes, aunque, en cualquier caso, creemos que requiere una mayor profundización.

5. Bibliografía

- ESCOFIER B., PAGES J. [1988]. *Analyses factorielles simples et multiples: objectifs, méthodes et interpretation*. Dunod. Paris.
- DAZY F., LE BARZIC J.F. [1996]. *L'analyse des Données Evolutives. Méthodes et Applications*. Editions Technip. Paris.
- LAVIT C. [1988]. *Analyse conjointe de tableaux quantitatifs*. Masson. Paris.