

XXVII Congreso de la Asociación Latinoamericana de Sociología. VIII Jornadas de Sociología de la Universidad de Buenos Aires. Asociación Latinoamericana de Sociología, Buenos Aires, 2009.

Inconvenientes metodológicos del análisis de correspondencias para el tratamiento de variables mixtas. Soluciones aplicadas al estudio de la estratificación social en Chile.

Adrián Leguina.

Cita:

Adrián Leguina (2009). *Inconvenientes metodológicos del análisis de correspondencias para el tratamiento de variables mixtas. Soluciones aplicadas al estudio de la estratificación social en Chile.* XXVII Congreso de la Asociación Latinoamericana de Sociología. VIII Jornadas de Sociología de la Universidad de Buenos Aires. Asociación Latinoamericana de Sociología, Buenos Aires.

Dirección estable: <https://www.aacademica.org/000-062/1157>

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.

Inconvenientes metodológicos del análisis de correspondencias para el tratamiento de variables mixtas

Soluciones aplicadas al estudio de la estratificación social en Chile

Adrián Leguina

*Departamento de Estadística, Facultad de Matemática,
Pontificia Universidad Católica de Chile
anleguin@uc.cl*

1. Introducción

En estudios de caracterización y/o construcción de perfiles, es usual utilizar variables cualitativas y cuantitativas. A este tipo de variables multidimensionales, se les conoce como variables mixtas. Para el tratamiento conjunto de variables cualitativas y cuantitativas, encontramos diferentes tipos de discretización para variables numéricas, las que generalmente consisten en transformar las variables continuas en variables cualitativas multiestado, asumiendo que este proceso implica perder información. Dentro de estas posibilidades encontramos la codificación de variables continuas de Escofier (Escofier 1979).

Para el análisis estadístico de variables mixtas, se propone aplicar la codificación de Escofier a variables cuantitativas, transformar cada categoría de variables cualitativas multiestado en variables dicotómicas, y finalmente aplicar análisis de correspondencias (AC) a toda la matriz de datos. Comparando los resultados del AC sólo para variables cuantitativas codificadas, con la aplicación de análisis de componentes principales a las variables originales, es posible verificar el grado de información perdida debido a la codificación de variables (Leguina 2009).

Esta propuesta se aplicó al estudio de la problemática de estratificación social en Chile (Sepúlveda y Leguina 2008). Los resultados muestran que al aplicar AC a variables continuas codificadas utilizando la propuesta de Escofier, se obtienen resultados considerablemente mejores que los encontrados a través de realizar la aplicación a datos codificados arbitrariamente. De esta forma, se pone de manifiesto la importancia de esta etapa previa al análisis de datos.

2. VARIABLES MIXTAS Y PROBLEMAS EN SU UTILIZACIÓN

Usualmente, se trabaja en forma simultáneamente con grandes matrices de datos multivariantes, que contienen variables de distinta naturaleza. A este tipo de variables multidimensionales se le conoce como “variables mixtas”. Los investigadores sociales, constantemente analizan e interpretan este tipo de variables, a pesar de no ser siempre conscientes de ello. Por ejemplo, en modelos lineales (de cualquier tipo y/o complejidad) se pueden utilizar variables dicotómicas (dummies) para estudiar la capacidad predictiva de una variable independiente, no necesariamente continua, sobre una variable dependiente.

Según Benzécri (2002) “se subentiende la distinción entre variable cualitativa y cuantitativa: no deberíamos decir que un valor continuo corresponde a datos cuantitativos, y un valor que toma una cantidad finita de modalidades equivale a datos aproximadamente cualitativos. Eso no siempre será cierto, ya que dependiendo de la unidad estadística de análisis, no siempre somos capaces de medir variables numéricas con la precisión esperada. La distinción entre variables debe provenir de su significado”. Lo anterior, hace que plantear el tipo de variable que se necesita y utiliza en una investigación, sea una tarea más difícil de lo que parece, produciendo que los propios investigadores no sean capaces de reconocer los tipos de variables que operacionalizan. Por lo tanto, el sustento teórico de la investigación debe guiar la conceptualización de las variables utilizadas, incluyendo su posible (re)codificación.

El procedimiento clásico para el manejo de datos en AC, a través de la construcción de la tabla disyuntiva completa, considera que todas las variables son de tipo cualitativas. Su aplicación directa a datos cuantitativos continuos o discretos¹ genera discontinuidades en la nueva variable, que implican pérdida de información. El paso progresivo de una categoría a otra implica que valores cercanos a la frontera de dos categorías no son muy diferentes en sus valores originales.

¹ En estricto rigor, no es posible construir una tabla disyuntiva completa a variables continuas con recorrido en los números reales, debido a que tendrían infinitas modalidades.

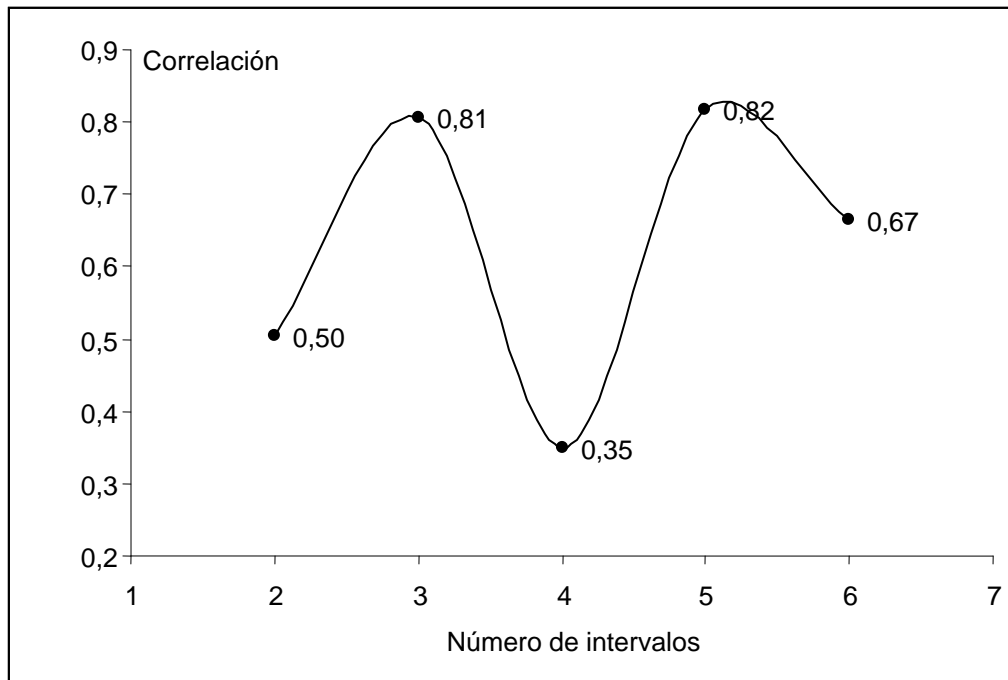
Tabla 1. Ejemplo de codificaciones arbitrarias

Calidad de TV	2 categorías	3 categorías	4 categorías	5 categorías	6 categorías
8	2	3	4	4	6
4,2	1	2	2	3	3
7	2	3	3	4	6
10	2	3	4	6	6
5,4	2	2	3	3	4

Fuente: Elaboración propia

Lo anterior es posible demostrarlo empíricamente a través de un pequeño ejemplo. La tabla 1 contiene información de la variable cuantitativa “percepción de la calidad del canal de televisión que ve con mayor frecuencia” para las primeras 5 observaciones. Imaginemos que deseamos codificar esta variable en categorías discretas. El primer problema que se enfrenta es definir el número óptimo de categorías para discretizar la variable. En este caso se codificará utilizando 2, 3, 4, 5 y 6 categorías. Ahora, ¿Qué ancho tendrán cada uno de los intervalos? Esto corresponde a otra posible interrogante dentro de la discretización arbitraria, que se podría resolver (sin claros beneficios) definiendo intervalos de igual ancho. Notemos que aplicando este procedimiento se pierde la posibilidad de interpretar el sentido de orden presente en la variable cuantitativa original², además de existir una clara pérdida de información.

Gráfico 1. Correlaciones entre distancias



Fuente: Elaboración propia

² Este problema se ve acentuado al aplicar AC sobre una variable discretizada de esta manera.

Para enfatizar aún más en este problema, se calculará la distancia³ entre individuos, utilizando la variable original y la variable discretizada en distinta cantidad de intervalos. Si la codificación arbitraria fuera buena, debería capturar la mayor cantidad de información acerca de las diferencias entre individuos, es decir, estar altamente correlacionada con las distancias entre individuos respecto a la variable original. El gráfico 1 muestra la evolución de la correlación de Spearman entre la distancia euclídeana estandarizada y el coeficiente de disimilaridad⁴ de emparejamiento simple extendido⁵ para distinta cantidad de intervalos. En el gráfico podemos observar que las correlaciones varían considerablemente, dependiendo del número de intervalos utilizados. De esta forma, se demuestra que la aplicación de una codificación arbitraria para discretizar una variable continua, no es un procedimiento adecuado, ni entregará los resultados esperados. Adicionalmente notemos que a medida que se aumenta el número de intervalos, la correlación no aumenta, tendiendo a fluctuar y disminuir sinusoidalmente.

3. CODIFICACIÓN DE ESCOFIER

La codificación de Escofier (Escofier 1979), utilizada como paso previo al AC, tiene como fin poder analizar en forma simultánea variables cuantitativas y cualitativas. El procedimiento consiste en transformar cada variable cuantitativa en dos nuevas variables:

$$\begin{cases} (1-x_i)/2 \\ (1+x_i)/2 \end{cases}$$

Donde x_i es el valor estandarizado de la variable para el individuo i . Notamos que al igual que para una variable cualitativa binaria, la variable recodificada queda definida por dos columnas, que para una misma variable siempre la suma será igual a 1. Las variables cuantitativas codificadas de esta manera se pueden analizar simultáneamente con variables cualitativas con 2 ó más categorías presentadas en forma de tabla disyuntiva completa.

³ En estadística, la distancia cuantifica la proximidad o lejanía entre individuos o poblaciones (Cuadras 2007).

⁴ Se le llama disimilaridad al cálculo de la distancia entre individuos cuando se utiliza variables cualitativas.

⁵ Este coeficiente se calcula como $1 - (\# \text{ de similitudes} / \# \text{ variables totales})$ (Cuadras 2007)

4. ANÁLISIS DE CORRESPONDENCIAS

El análisis de correspondencias⁶ puede ser definido como una técnica descriptiva para casi cualquier tipo de matriz de datos con entradas no negativas, especialmente útil para representar tablas de contingencia que resuman información cualitativa, de una forma exploratoria y reduciendo sus características multivariantes a la menor cantidad de dimensiones posibles. El método resulta ser una forma sencilla de graficar los resultados, permitiendo su rápida comprensión e interpretación (Greenacre 2007). A pesar de que AC es muy flexible para trabajar con distintos tipos de variables, es la codificación la que aporta aun más a esta capacidad (Murtagh 2005).

Dentro de las principales ventajas de la técnica, se encuentra que su principio geométrico permite representar simultáneamente individuos y variables en un mismo espacio, construyendo una cartografía del espacio social de los encuestados, y caracterizando mediante las variables principales (o factoriales) resultantes la existencia de distintos perfiles. Para enriquecer los resultados e interpretaciones, AC permite incluir variables y/o individuos “suplementarios”, que no participan en la construcción del espacio geométrico de las variables. Además, sus resultados gráficos facilitan la comprensión e interpretación de los datos.

5. ANÁLISIS DE DATOS GEOMÉTRICOS, SOCIOLOGÍA Y ESTRATIFICACIÓN SOCIAL

Desde sus principios, el análisis de datos geométricos (ADG) ha sido utilizado en las más diversas disciplinas de las ciencias. El análisis de correspondencias, es la técnica más importante de ADG, donde es posible encontrar que una parte importante de su historia, está muy relacionada con la sociología. Para explicar la razón, en primer lugar se debe tener en cuenta una de las diferencias más importantes entre la metodología propuesta por el ADG y la estadística “tradicional” aplicadas a la sociología (Roux y Rouanet 2005):

- En la estadística tradicional, los indicadores numéricos como coeficientes de regresión y valores p, son el centro de todo análisis. Desde esta perspectiva se busca la *sociología de las variables*.

⁶ En este artículo no se presentará la formulación matemática del análisis de correspondencias. Para una completa definición se recomienda consultar Murtagh (2005), Roux y Rouanet (2005) y Greenacre (2007).

- En ADG, las nubes de puntos son el centro de atención. Desde esta perspectiva se busca construir el *espacio social de las variables*.

Las características especiales del ADG (y específicamente de AC) resultaron ser bastante deseables por toda una escuela de sociología. El marco conceptual del libro “La Distinción” del sociólogo francés Pierre Bourdieu (1930–2002), considerado uno de los más importantes sociólogos contemporáneos, nace del uso exhaustivo de teoría y evidencia empírica de tipo cualitativa (entrevistas) y cuantitativa (encuestas). A grandes rasgos, Bourdieu (1979) especifica tres tipos de capitales (económico, cultural y social⁷), como las dimensiones básicas del espacio social de los estilos de vida. De esta forma, una serie de interrelaciones entre los capitales se asocian al “lugar” o “posición” de los individuos en la estratificación social, definiendo la forma en que se relacionan y diferencian entre sí. En este contexto, el AC, junto con el uso de metodologías cualitativas, fueron capaces de caracterizar mediante gráficos de sus ejes factoriales, con notable precisión y detalle, a la sociedad francesa de la década de los 60s. No solo en esta área Bourdieu aplicó el análisis de correspondencias, sino que en gran parte de sus investigaciones empíricas.

La obra de Bourdieu y sus seguidores, ha tenido un gran impacto sobre el AC. Sin lugar a dudas, entre los más importantes encontramos el dar a conocer la técnica a nivel mundial, generar discusión e interés a nivel teórico, tanto sociológico como estadístico, lo cual produjo muchos avances teóricos paralelos en la obra de Benzécri.

Desde el punto de vista estadístico, se puede notar que en estudios de estratificación social es difícil diferenciar las variables entre “dependientes” e “independiente”⁸, de la forma en la que usualmente se hace en modelos lineales u otras técnicas. Definir relaciones causales, o incluso explicativas en el estudio de la estratificación social, ha sido cuestionado incluso antes de la aparición de La Distinción (Bourdieu 1979; Savage et al. 2005), donde aun no se tiene del todo clara la existencia de relaciones explicativas entre los capitales, y mucho menos la forma de esta relación.

En Chile también se puede ver un gran interés por el tema, donde en los últimos años se han generado varios proyectos relacionados. Por último, y no menos importante, esta investigación es clara la influencia de la obra de Benzécri y Bourdieu en el apartado estadístico y aplicado respectivamente.

⁷ Para sus definiciones, se recomienda ver Sepúlveda y Leguina (2009).

⁸ Esto es válido específicamente para la línea originada por la obra de P. Bourdieu. Otros investigadores en estratificación social, como John Goldthorpe y Erik Olin Wright, discrepan al respecto.

6. APLICACIÓN A DATOS REALES DE ESTRATIFICACIÓN SOCIAL EN CHILE⁹

La investigación propuesta por Sepúlveda y Leguina (2008) tiene como objetivo aportar a los estudios de estratificación social, mediante un concepto de desigualdad social, que define una distinción global y otra local. Dichos conceptos, provenientes de investigaciones anteriores, consideran que además de las variables tradicionalmente utilizadas en estratificación social (ingreso, educación, posesión de bienes, etc.), existen otros factores para distinguir a las personas en grupos sociales. Bajo este contexto se profundizó la instrumentalización de los conceptos de “nivel de información” que tiene una persona, el cuál está sujeto a la calidad, la rapidez, el tipo de información y qué valor le genera en la interacción con su entorno (Sepúlveda 2008) y “acceso a la información” que tiene relación a la disponibilidad de los distintos medios informacionales (Sepúlveda 2008). Desde esta mirada, se pretendió demostrar estadísticamente que los sujetos se pueden distinguir a través de nuevos factores no analizados bajo el enfoque clásico de los estudios de estratificación social.

Con el fin de caracterizar tanto el nivel como el acceso a la información se lleva a cabo un análisis de correspondencias múltiples, seleccionando las variables de la encuesta¹⁰ que definen de la mejor manera posible su nivel y acceso a la información. Para su implementación se utilizó el programa estadístico R (R Development Core Team 2008). El detalle de las dimensiones, sus indicadores y definiciones técnicas pueden verse en Sepúlveda y Leguina (2008).

El problema detectado en Leguina y Sepúlveda (2008), es que junto a variables cualitativas nominales, se analiza variables de tipo cuantitativas ordinales (entre 5 y 6 categorías, además de continuas) recodificadas en 3 ó 4 categorías consideradas como nominales. Esta recodificación se llevó a cabo arbitrariamente. Leguina (2009) presenta un estudio detallado acerca de los inconvenientes de realizar este procedimiento arbitrariamente, sufriendo severos problemas de pérdida de información, que llevan a una incorrecta construcción de la nube de puntos, falta de ortogonalidad de la solución factorial, y finalmente, a interpretaciones incorrectas.

⁹ El desarrollo completo de la investigación aplicada, incluyendo la conceptualización sociológica, se encuentra en Sepúlveda (2008), Sepúlveda y Leguina (2008), y en la presentación realizada por Denisse Sepúlveda y Adrián Leguina en el GT-3 de este Congreso.

¹⁰ Los datos provienen de la 3° Encuesta Nacional de Opinión Pública de la Universidad Diego Portales (2007), cuyo objetivo es contribuir a la comprensión de los cambios en las percepciones de los chilenos(as) en diversas áreas donde el país evidencia transformaciones cruciales

7. RESULTADOS PRINCIPALES

Para el análisis codificando arbitrariamente las variables cuantitativas, la solución con tres variables principales es capaz de explicar un 61,6 % de la variabilidad total de los datos. Los resultados se resumen en la tabla 2. A continuación, para cada variable principal, se descompone su variabilidad según las siete componentes de nivel y acceso a la información, de manera de determinar cual(es) de ella(s) mejor la explican.

Tabla 2. Contribuciones relativas de los componentes de nivel y acceso a la información por las variables principales aplicando codificaciones arbitrarias y de Escofier.

	C. arbitraria			C. Escofier			
	Vp1	Vp2	Vp3	Vp1	Vp2	Vp3	Vp4
<i>% de varianza expresado por la variable principal</i>	36,4 %	14,5 %	10,7 %	17,9 %	13,6 %	9,0 %	6,8 %
Rapidez/Frecuencia de la información	23,5 %	2,7 %	4,0 %	26,4 %	2,8 %	7,6 %	11,0 %
Tipo de información	14,2 %	0,9 %	3,0 %	13,0 %	1,6 %	32,2 %	36,6 %
Influencia de los medios	9,3 %	39,3 %	46,5 %	6,8 %	39,8 %	4,4 %	0,4 %
Influencia de la elite	7,1 %	32,1 %	33,3 %	4,8 %	32,5 %	5,0 %	0,2 %
Influencia propia	2,7 %	2,7 %	11,1 %	2,8 %	2,8 %	0,2 %	0,0 %
Calidad de la información	14,2 %	1,8 %	2,0 %	12,9 %	1,5 %	32,2 %	43,9 %
Acceso a la información	29,0 %	20,5 %	0 %	33,2 %	19,1 %	18,3 %	7,8 %
TOTAL	100 %	100 %	100 %	100 %	100 %	100 %	100 %

Fuente: Leguina y Sepúlveda (2008)

La codificación de datos mediante la técnica de Escofier se llevó a cabo tal como se presentó en la sección 4. La tabla 2 presenta el resumen de los resultados de la aplicación del AC. Se opta por una solución con cuatro variables principales, capaces de explicar un 47,3 % de la variabilidad total. La componente de influencia propia no figura en ninguna de las primeras cuatro variables principales, considerando que no tiene importancia suficiente como para ser interpretada dentro del análisis.

Gracias a la comparación de los resultados anteriores, podemos visualizar las diferencias entre los análisis. Por ejemplo, codificando arbitrariamente pareciera que se explica un mayor porcentaje de varianza total, pero eso se debe a la violación del supuesto de ortogonalidad¹¹ de las variables principales. Aunque se explique un menor porcentaje de la varianza total, utilizando codificación de Escofier se obtiene una solución más realista, donde cada variable principal está asociada a distintas modalidades componentes del nivel y acceso a la información.

¹¹ Las variables principales construidas mediante AC son independientes entre sí. Si dos variables principales se conforman con las mismas modalidades, estas no serán independientes, ya que estarán asociadas a través de la variación en la respuesta de estas modalidades comunes.

Desde el punto de vista interpretativo, la solución mediante codificación de Escofier nos lleva a un correcto análisis del nivel de importancia que adquiere cada una de las modalidades sobre las variables principales, anulando el efecto de “influencia propia”, pero incluyendo la percepción de “calidad de la información”, e incluso diferenciando entre medios, incluyendo a diarios (VP3) y televisión (VP4) como factores jerarquizados y diferenciadores sociales. Esto es de gran importancia para comprender adecuadamente los fenómenos sociales detrás de la posición que adquieren los individuos en la estratificación social, mediante el nivel y acceso de la información. Leguina (2009) analiza en detalle los resultados numéricos, profundizando en las ventajas de aplicar la codificación de Escofier.

Para decidir si efectivamente es la estrategia que presenta mejores resultados, es necesario preguntarse si existe pérdida de información en la codificación. Para demostrarlo empíricamente, se propone verificar qué ocurre con los porcentajes de varianza explicados y calidad de representación, comparándolos con la aplicación de un análisis de componentes principales (ACP) sólo para las variables cuantitativas. No se detallará la teoría de ACP, pero esta técnica es posible entenderla como realizar un AC sobre variables cuantitativas, donde se busca una proyección geométrica según la cual los datos sean mejor representados en términos de mínimos cuadrados.

Tabla 3. Comparación de resultados para variables cuantitativas

		1° VP	2° VP	3° VP	4° VP
ACP	Valor propio	2,777	1,640	1,090	1,015
	% relativo	27,800	16,400	10,900	10,200
	% relativo acumulado	27,800	44,200	55,100	65,200
C.arbitraria	Valor propio	0,215	0,157	0,103	0,082
	% relativo	11,041	9,365	7,281	5,241
	% relativo acumulado	11,041	20,406	27,688	32,962
C. Escofier	Valor propio	0,277	0,164	0,109	0,101
	% relativo	27,800	16,400	10,900	10,200
	% relativo acumulado	27,800	44,200	55,100	65,200

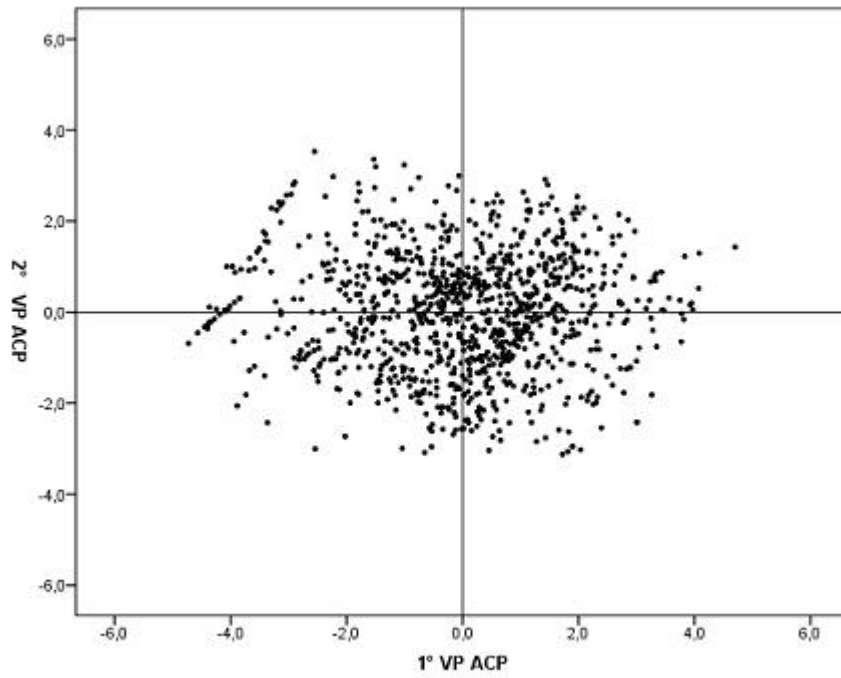
Fuente: Leguina (2009)

La investigación de Bramardi et al. (2006) demuestra matemáticamente este resultado para la codificación de Escofier. Este resultado, lo podemos corroborar en la tabla 3, que corresponde al resumen de los resultados obtenidos en la comparación, donde los valores propios, aunque en otra escala, y porcentajes de varianza son idénticos a los obtenidos con ACP. Codificar arbitrariamente, explica un 32% menos de variabilidad que al utilizar ACP.

Los gráficos 2, 3 y 4 muestran la posición de los(as) encuestados(as) en los dos primeros ejes factoriales para los distintos análisis realizados. La nube que se utiliza como referencia, es la obtenida mediante ACP, ya que corresponde a la técnica que se aplicaría originalmente. Visualmente, es posible verificar la precisión de las soluciones bajo las distintas técnicas de codificación, comparando la forma de las nubes de puntos representadas. Una buena estrategia de codificación será la que entregue resultados similares en calidad y porcentajes de varianza explicada, a la obtenida con ACP. La comparación de las representaciones gráficas de los resultados, se convierte en una importante herramienta de análisis en la evaluación de la calidad de codificación de datos.

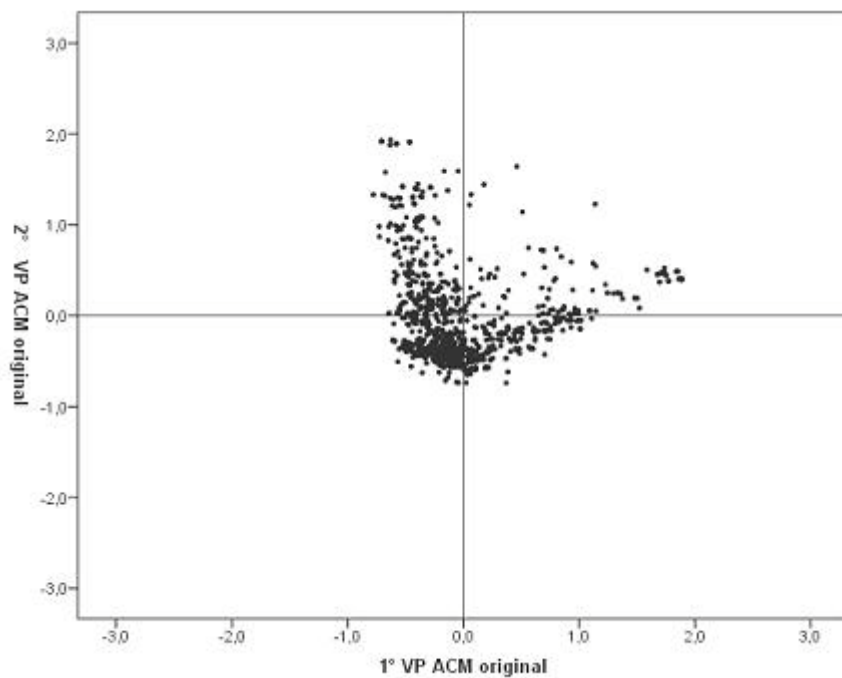
Comparando los resultados de la propuesta de codificación (gráfico 4) con la nube de referencia (gráfico 2) y el análisis original (gráfico 3), es posible sustentar la superioridad de la codificación de Escofier. Su representación gráfica es idéntica a la obtenida con ACP, excepto por un cambio de escala, lo cual había sido puesto en manifiesto por Bramardi et al. (2006). Resulta interesante observar que el análisis con variables codificadas arbitrariamente (gráfico 3) resulta tan diferente a lo que debió obtenerse. Esto pone en manifiesto lo crítico de la selección de la estrategia de codificación. Una inadecuada codificación de datos, conducirán a cometer errores en la solución, y finalmente, en las interpretaciones y conclusiones desprendidas de esta.

Gráfico 2. Resultados de ACP para encuestados. Se utilizaron solo las variables cuantitativas.



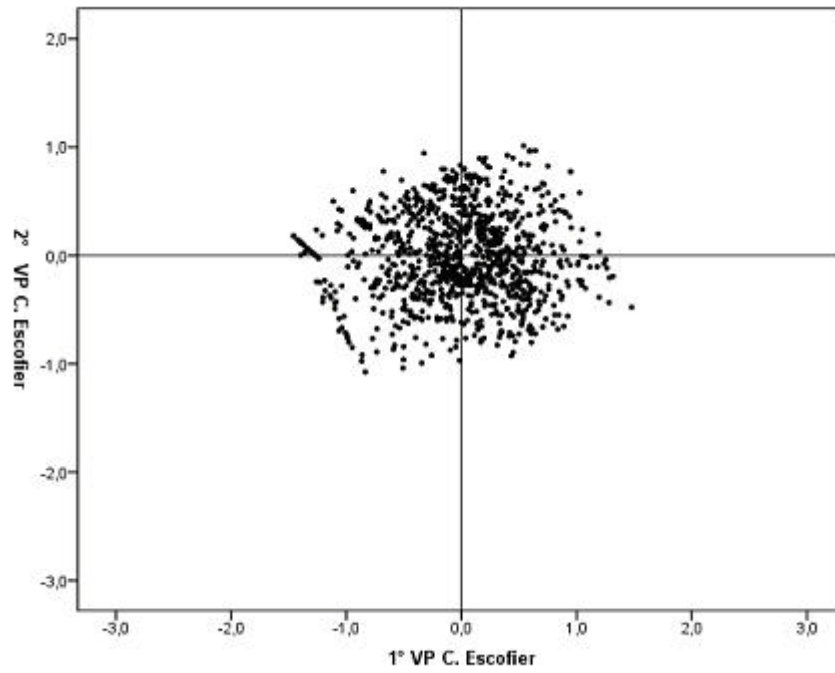
Fuente: Leguina (2009)

Gráfico 3. Resultados de ACM con codificación arbitraria para encuestados. Se utilizaron solo las variables cuantitativas.



Fuente: Leguina (2009)

Gráfico 4. Resultados de AC para encuestados. Se utilizaron solo las variables cuantitativas aplicando codificación de Escofier.



Fuente: Leguina (2009)

8. CONCLUSIONES

Un aspecto fundamental de cualquier investigación que requiera análisis estadístico de datos, es el tipo de variables a utilizar. Deben ser ellas las que guíen la selección de la técnica a usar. Dentro de este contexto, la codificación de datos es un aspecto usualmente dejado de lado, pero como ha sido demostrado en este artículo, ayuda de sobremanera a “enfocar la fotografía” de los datos. A través de la comparación de resultados, se concluye que la aplicación de alguna técnica adecuada de codificación para trabajar con variables cuantitativas en AC, aporta una considerablemente mejor calidad en la representación del espacio geométrico de las variables.

De forma empírica, y con datos de una investigación real, se demuestra en qué grado existe pérdida de información para las distintas codificaciones, o en el caso de Escofier, que no existe pérdida de información alguna, comparándolo con la aplicación de análisis de componentes principales. Además fue posible verificar las variaciones en la nube de puntos generadas para el análisis con variables codificadas arbitrariamente, demostrando que siempre es mejor implementar Escofier, en lugar de recodificar subjetivamente o dejar en su estado original las variables cuantitativas.

Para evaluar la calidad de una propuesta de codificación, se propone la utilización del análisis de componentes principales, y la comparación numérica y gráfica de sus resultados, con los obtenidos para las codificaciones alternativas. De esta forma, es posible cuantificar (al menos visualmente) la pérdida de información y calidad de la alternativa de codificación. Esta herramienta de análisis, es posible implementarla previa a la aplicación de cualquier análisis de datos que requiera codificar variables cuantitativas.

Bibliografía

- Benzécri, J.-P. (2002). In memoriam: P. Bourdieu.
- Bourdieu, P. (1979). *La distinction: Critique sociale du jugement*. (traducción al español: *La distinción: Criterio y bases sociales del gusto* (2006). Buenos Aires: Taurus). Paris: Editions de Minuit.
- Bramardi, S., Reeb, P., De Bernardin, F., Tassile, V., & Ferrer, M. (2006). Codificación de Escofier: una "discretización" sin pérdida de información. In *Actas del VII Congreso Latinoamericano de Sociedades de Estadística*. Rosario, Argentina.
- Cuadras, C. M. (2007). *Nuevos Métodos de Análisis Multivariante* Barcelona: CMC Editions.
- Escofier, B. (1979). Traitement simultané de variables qualitatives et quantitatives en analyse factorielle (qualitatives et quantitatives). *Les Cahiers de l'Analyse des Données*, 4, 137–146.
- Greenacre, M. (2007). *Correspondence analysis in practice* (Segunda ed.). Boca Raton: Chapman & Hall/CRC.
- Leguina, A. (2009). *Análisis de Correspondencias Múltiples para el Tratamiento de Variables Mixtas: Un Estudio Aplicado a la Problemática de Estratificación Social en Chile*. Tesis para optar al grado de Magíster en Estadística. Facultad de Matemáticas, Pontificia Universidad Católica de Chile.
- Leguina, A., & Sepúlveda-Sánchez, D. (2008). El nivel de información como una nueva forma de distinción social. Una aplicación de estadística multivariante. En *Actas de las VIII Jornadas Latinoamericanas de Sociedades de Estadística*. Montevideo, Uruguay.
- Murtagh, F. (2005). *Correspondence analysis and data coding with Java and R*. Boca Raton: Chapman & Hall/CRC.
- R Development Core Team. (2007). *R: A language and environment for statistical computing* [Computer software manual]. Vienna, Austria. Available from <http://www.R-project.org>.
- Roux, B., & Rouanet, H. (2005). *Geometric data analysis*. New York: Kluwer Academic Publishers.
- Savage, M., Gayo-Cal, M., Warde, A., & Tampubolon, G. (2005). Cultural capital in the uk: A preliminary report using correspondence analysis. *CREST Working Paper Series*, 4.
- Sepúlveda, D. (2008). Una nueva perspectiva de clases sociales en Chile: Global y local. En *IX Jornadas de estudiantes de postgrado en Humanidades, Artes, Ciencias Sociales y Educación: "América Latina en el Nuevo Milenio, Procesos, Crisis y Perspectivas"*. Santiago, Chile.
- Sepúlveda, D., Leguina A. (2008). El nivel de la información como una nueva forma de distinción social". *Actas del Encuentro preALAS Chile 2008 "50 años de Sociología en Chile"*, Santiago, Chile.
- Universidad Diego Portales. (2007). *Tercera Encuesta Nacional de Opinión Pública UDP 2007*. Santiago, Chile: ICSO-UDP.