Universidad de Valladolid

Facultad de Ciencias Económicas y Empresariales

Grado en Marketing e Investigación de Mercados

Análisis factorial con variables categóricas

Presentado por:

Irene Barajas Marcos

Tutelado por:

Mercedes Prieto Alaiz

Valladolid, 1 de Julio de 2015

Índice

1. INTRODUCCIÓN

El análisis factorial es una técnica multivariante de reducción de datos que resume la información de un conjunto de variables en un número menor de factores, los cuales se pueden utilizar en otro tipo de análisis, aunque también pueden tener interés propio debido a su interpretación y sentido.

Esta técnica se desarrolla habitualmente a partir de la matriz de correlaciones de Pearson, y su uso está recomendado para variables cuya distribución se aproxime a una normal.

En aquellos casos en los que se trabaje con variables dicotómicas, lo más adecuado desde el punto de vista metodológico es, según diversos autores (Ferrando, 1996; Elosua y Zumbo, 2008), la utilización de correlaciones tetracóricas. En este tipo de correlaciones se asume que existen unas variables latentes continuas que subyacen a las variables dicotómicas observadas. Su uso proporciona un mejor ajuste y una mayor robustez de los modelos, además de mejores estimaciones en los análisis de fiabilidad y mejores niveles de significación en las relaciones existentes entre las variables (Choi et. al., 2010; Holgado Tello et. al., 2010), obteniéndose por tanto una estimación mucho más precisa de las dimensiones subyacentes al conjunto de variables.

No obstante, también existen aquellos que se posicionan en contra de esta técnica, argumentando para ello que las correlaciones tetracóricas no equivalen algebraicamente a las habituales correlaciones de Pearson (Morales, 2006) entre otros inconvenientes que presentan este tipo de correlaciones a los que se hará referencia a lo largo de la investigación.

El objetivo de este trabajo es, por tanto, el desarrollo de la metodología de las correlaciones tetracóricas aplicadas al análisis factorial cuando se trabaja con variables dicotómicas. Además, se expone la aplicación a un caso práctico en el que se resumirá la información de 9 variables basadas en la privación de material en España en un número más reducido de factores usando este tipo de correlaciones.

El desarrollo del trabajo comienza por la explicación de la técnica del análisis factorial, haciéndose referencia a sus antecedentes y a las hipótesis y propiedades del modelo, así como a los distintos métodos disponibles para estimar las cargas factoriales y para rotar los factores.

A continuación, se aborda el tema de las correlaciones tetracóricas, definiéndolas y estableciendo tanto las ventajas como los inconvenientes de su utilización. Seguidamente, se presenta la aplicación a un caso práctico que ejemplificará el uso de la metodología anteriormente expuesta, para finalizar con un resumen de las principales conclusiones extraídas de la investigación.

2. ANÁLISIS FACTORIAL

2.1 INTRODUCCIÓN

El análisis factorial es una técnica de interdependencia cuyos orígenes se remontan a principios del siglo XX, momento en el que fue aplicado en estudios psicométricos que trataban de medir la inteligencia.

Aunque varios autores han contribuido al desarrollo de dicha técnica (Karl Pearson, Francis Galton o Louis Leon Thurstone), hoy en día se considera a Charles Spearman como el padre del análisis factorial. Actualmente, es una técnica muy extendida y utilizada en múltiples campos como la Economía o la Biología, además de la Psicología y Psicometría.

El análisis factorial es una técnica de reducción de datos que sirve para explicar las correlaciones de un conjunto de variables observadas mediante un número reducido de factores no observables. Por lo general, estos factores deben ser pocos e interpretables, ya que resulta interesante encontrar el sentido y dar una explicación a dichos factores.

El análisis factorial puede ser exploratorio o confirmatorio, en función de si se tiene o no conocimiento a prioi de los factores. En el análisis factorial exploratorio no se conoce de antemano el número de factores a utilizar, sino que se determina en el desarrollo del análisis. Por el contrario, en el análisis factorial confirmatorio los factores están fijados previamente a la aplicación de la técnica, y son corroborados mediante distintos test de hipótesis.

2.2 FORMULACIÓN E HIPÓTESIS DEL MODELO

Supongamos que se dispone de *n* observaciones de *p* variables aleatorias: X_1, $X_2,\ldots X_p$. Sin pérdida de generalidad, se planteará el modelo en función de las variables tipificadas Z_1, $Z_2,\ldots Z_p$.

$$Z_1 = \alpha_{11}F_1 + \ldots + \alpha_{1m}F_m + u_1$$

$$\vdots$$

$$Z_p = \alpha_{p1}F_1 + \ldots + \alpha_{pm}F_m + u_p$$

donde:

1) F_j, j=1…m son los *m<p* factores comunes o variables latentes no observables. Se supone que tienen media cero, varianza unitaria, están mutuamente incorrelacionados y se distribuyen conjuntamente como una normal. Por lo tanto, el vector F de orden (mx1) formado por los *m* factores se distribuye como una N (0, I).
2) u_i, i=1…p son las perturbaciones no observadas o factores únicos. Se supone que tienen media 0, están incorrelacionados entre sí e incorrelacionados con los factores comunes, y conjuntamente se distribuyen como una normal. En base a esto, el vector U de orden (px1) formado por los factores únicos sigue una distribución N_p (0, φ), donde φ es una matriz diagonal.
3) a_{ij} son los pesos factoriales o constantes desconocidas que describen cómo afectan los factores a las variables observadas. Por lo tanto, la matriz A de orden (pxm) recogerá dichas cargas factoriales.
4) Z_1, $Z_2,\ldots Z_p$, i=1…p son variables con media 0 y varianza la unidad. El vector Z de orden (pxn) formado por las variables observables se distribuye de la siguiente manera: Z $\rightarrow$N (0, R_{pob}), donde R_{pob} es la matriz de correlaciones de las variables originales X_1, $X_2,\ldots X_p$, o la matriz de varianzas y covarianzas de las variables tipificadas.

Expresando el modelo de forma matricial se obtiene:

$$\begin{pmatrix} Z_1 \\ \vdots \\ Z_i \\ \vdots \\ Z_p \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{1j} & \alpha_{1m} \\ \vdots & \vdots & \vdots \\ \alpha_{i1} & \alpha_{ij} & \alpha_{im} \\ \vdots & \vdots & \vdots \\ \alpha_{p1} & \alpha_{pj} & \alpha_{pm} \end{pmatrix} \begin{pmatrix} F_1 \\ \vdots \\ F_j \\ \vdots \\ F_m \end{pmatrix} + \begin{pmatrix} u_1 \\ \vdots \\ u_i \\ \vdots \\ u_p \end{pmatrix}$$

Por tanto:

$$\boxed{Z = AF + u}$$

2.3 PROPIEDADES DEL MODELO

El modelo factorial tiene dos propiedades fundamentales:

- La varianza de cada variable se divide en dos partes: la primera es aquella parte que Z_i tiene en común con el resto de las variables a través de los factores comunes, es decir, su comunalidad, y la segunda es la especificidad de Z_i.

  En principio, tanto la parte común como la parte específica están formadas por parámetros desconocidos que se estimarán a partir de las cargas factoriales.

$$\text{Var}(Z_i) = 1 = \underbrace{\alpha_{i1}^2 + \alpha_{i2}^2 + \ldots + \alpha_{ij}^2 + \alpha_{im}^2}_{\textit{Comunalidad}} + \underbrace{\text{Var}(u_i)}_{\textit{Unicidad}} \qquad \forall\ i = 1 \ldots p$$

- La correlación entre las variables originales X_h y X_i depende únicamente de la presencia de factores comunes:

$$\text{Corr}(X_h X_i) = \text{Corr}(Z_h Z_i) = \alpha_{h1}\alpha_{i1} + \alpha_{h2}\alpha_{i2} + \ldots + \alpha_{hj}\alpha_{ij} + \ldots + \alpha_{hm}\alpha_{im} = \sum_{j=1}^{m} \alpha_{hj}\alpha_{ij}$$

  En función de esto, se ha de tener en cuenta que dos variables estarán altamente correlacionadas si existen cargas factoriales elevadas en los mismos factores comunes.

De estas dos propiedades se deriva que la matriz de correlaciones de las variables originales (o la matriz de varianzas y covarianzas de las variables tipificadas) se puede expresar de la siguiente forma:

$$\boxed{R_{pob} = AA' + \varphi}$$

Como se puede observar, dicha matriz admite la siguiente descomposición:

- ❖ La matriz AA', determinada por las cargas factoriales, depende solo de los factores comunes.
- ❖ La matriz φ, determinada por los factores específicos.

2.4 MÉTODOS PARA LA ESTIMACIÓN DE LAS CARGAS FACTORIALES

El objetivo del análisis factorial, es conseguir estimaciones de A y φ, ya que son parámetros a priori desconocidos.

$$R_{pob} = AA' + \varphi$$

$$R_{pob} = \begin{pmatrix} 1 & \rho_{x_1x_2} & \rho_{x_1x_p} \\ \vdots & & \\ \rho_{x_1x_2} & 1 & \rho_{x_2x_p} \\ \vdots & \vdots & \\ \rho_{x_1x_p} & \rho_{x_2x_p} & 1 \end{pmatrix}$$

Los elementos de la matriz de correlaciones poblacionales de las variables se pueden estimar partiendo de las correlaciones muestrales (en general, se utilizan las correlaciones de Pearson, pero en este trabajo se propone la utilización en este punto de las correlaciones tetracóricas):

$$\widehat{R_{pob}} = \begin{pmatrix} 1 & r_{x_1x_2} & r_{x_1x_p} \\ \vdots & & \\ r_{x_1x_2} & 1 & r_{x_2x_p} \\ \vdots & \vdots & \\ r_{x_1x_p} & r_{x_2x_p} & 1 \end{pmatrix}$$

Para obtener las estimaciones de A y φ, se ha de tener en cuenta la posible aparición de dos problemas:

1) Grados de libertad: es necesario que el número de ecuaciones sea mayor o igual que el número de parámetros a estimar.

$$\frac{p(p+1)}{2} \geq p\ (m+1)$$

2) No existe una solución única, es decir, las cargas no quedan determinadas de una sola forma. Por lo tanto, el modelo factorial está indeterminado ante rotaciones ortogonales[1].

Existen diferentes métodos para obtener las cargas factoriales:

2.4.1 Análisis de componentes principales:

Para determinar las cargas factoriales con el método de componentes principales se parte de *p* variables Z_1, Z_2,…, Z_p de las cuales se tiene *n* observaciones tipificadas, y se obtienen *p* nuevas variables Cp_1, Cp_2,…, Cp_p que son las componentes principales resultado de la combinación lineal de las variables originales. Dichas componentes recogerán la máxima información posible de las variables iniciales.

Los factores extraídos a través de este método están incorrelacionados entre sí y también están incorrelacionados con los factores específicos, sin embargo, los factores específicos no están incorrelacionados entre ellos, lo cual implica que uno de los supuestos del análisis factorial no se sostiene. Se ha de tener en cuenta que esto no tiene mucha importancia cuando las comunalidades son altas.

2.4.2 Ejes principales:

El método de ejes principales es una variante del de componentes principales, y se trata de un proceso iterativo que evita tener que resolver las ecuaciones de máxima verosimilitud (que resultan más complejas).

Este método trata de que los factores expliquen la máxima varianza posible y estén a su vez incorrelacionados.

[1] Como se verá más adelante, esto se puede convertir en una ventaja, ya que permite elegir entre todas las posibles soluciones aquella que facilite más la interpretación de los factores.

2.4.3 Mínimos cuadrados no ponderados:

Consiste en estimar las cargas factoriales minimizando la suma de cuadrados de las diferencias entre los elementos de la matriz de correlación muestral y la matriz de correlación reproducida, sin incluir los elementos de la diagonal.

2.4.4 Mínimos cuadrados ponderados:

En este caso se estiman las cargas factoriales de igual forma que en el método anterior, minimizando la suma de cuadrados de las diferencias entre los elementos de la matriz de correlación muestral y la matriz de correlación reproducida, sin incluir los elementos de la diagonal, ponderando además las correlaciones con la inversa de las especifidades de las variables.

Con esto se conseguirá que aquellas variables con una alta especifidad tengan un menor peso en la solución que aquellas que tengan una baja especifidad.

2.4.5 Máxima verosimilitud:

Para la aplicación de este método, se estiman las cargas factoriales y las varianzas de los factores específicos que maximicen la función de verosimilitud de una muestra de tamaño *n* de $Z = (Z_1\ Z_2\ldots..Z_p)'$, siendo Z una variable normal p-dimensional, cuya media es la matriz columna de ceros y cuya matriz de varianzas y covarianzas es $R_{pob}= AA'+ \varphi$.

2.5 ROTACIÓN DE LOS FACTORES

Como ya se ha explicado anteriormente, el análisis factorial trata tanto de determinar los factores comunes que subyacen a las variables observadas como de dar sentido a los mismos. Esto se lleva a cabo teniendo en cuenta la relevancia que cada una de las variables tiene en cada factor (medido a través del coeficiente de correlación). Frecuentemente, la solución inicial es compleja de interpretar, ya que puede ocurrir que las variables tengan presencia en prácticamente todos los factores.

Para interpretar los factores se ha tener en cuenta por tanto que la solución al análisis factorial no es única: hay infinitas soluciones para cada modelo. La rotación consiste en transformar los factores iniciales en otros nuevos que sean más fáciles de interpretar mediante la multiplicación de la matriz de cargas factoriales por otra matriz ortogonal.

A continuación, se exponen distintos criterios de rotación dentro de las dos formas genéricas de rotar los factores existentes: la rotación ortogonal o rígida y la rotación oblicua.

2.5.1 Rotación ortogonal

En este tipo de rotación, los ejes resultantes en el nuevo espacio factorial son ortogonales, por lo tanto, se mantiene la incorrelación de los factores. Dentro de la rotación ortogonal se pueden encontrar distintos criterios de rotación: Varimax, Equamax, Quartimax,…etc.

De ellas, el método más utilizado es el Varimax, en el cual los ejes se hallan mediante la maximización de la suma de las varianzas de las cargas factoriales al cuadrado de cada factor.

Lo que se consigue con este método es que cada variable esté altamente representada en el menor número de factores posible (lo ideal sería que estuviese únicamente en 1 de ellos). El inconveniente de este método es que pueden dejar de detectarse algún factor subyacente a todas las variables.

2.5.2 Rotación oblicua

En la rotación oblicua se permite que los nuevos ejes sean “moderadamente” no ortogonales, es decir, los factores presentarán cierta correlación con la finalidad de obtener una interpretación más clara.

En este caso, el método más utilizado es el Oblimin.

2.6 PUNTUACIÓN DE LOS FACTORES

Además de los objetivos ya señalados del análisis factorial, esta técnica no solo tiene la finalidad de extraer y posteriormente interpretar los factores comunes, sino que es utilizada también para exportar las puntuaciones factoriales, utilizándolas como variables en otros procedimientos multivariantes, sabiendo que estas variables van a permitir reducir la dimensionalidad de la información inicial y asegurando de esta manera el trabajo con variables incorrelacionadas.

En caso de utilizar el método de componentes principales, se obtienen de manera inmediata las puntuaciones factoriales (ya que estas se corresponden con las puntuaciones de las *m* primeras componentes tipificadas).

Si por el contrario, se utilizaran otros métodos de extracción, habría que estimar las puntuaciones factoriales a través de distintos métodos, pero teniendo en cuenta que en este caso se puede incumplir alguno de los supuestos teóricos del modelo.

3. CORRELACIONES ENTRE VARIABLES DICOTÓMICAS

El método más habitual para estimar la matriz de correlaciones necesaria para aplicar el análisis factorial es el que utiliza la matriz formada por los coeficientes de correlación de Pearson (este es el método implantado por defecto en la mayoría de programas estadísticos que trabajan esta técnica).

Este método está tradicionalmente destinado a ser utilizado con variables que se distribuyen aproximadamente como una normal, ya que su uso cuando se dispone de variables ordinales y dicotómicas genera a menudo resultados distorsionados y menos precisos que los deseables, tal y como señala Pearson (1900).

Es por esto que se han propuesto diferentes formas de estudiar el grado de correlación entre las variables ordinales y/o dicotómicas. La tabla 3.1.1 muestra los diferentes tipos de correlación recomendados en función de la naturaleza de las variables.

TABLA 3.1 CORRELACIONES EN FUNCIÓN DE LA NATURALEZA DE LAS VARIABLES

		Y_2		
		CONTINUA	**ORDINAL**	**DICOTÓMICA**
	CONTINUA	Correlación de Pearson	Correlación policórica	Punto biserial
Y_1	**ORDINAL**	-	Correlación policórica	Correlación policórica
	DICOTÓMICA	-	-	Correlación tetracórica

Fuente: SECQ

Esta investigación se va a centrar en la forma de estudiar la correlación entre variables dicotómicas. Consecuentemente, se abordará la utilización de la correlación tetracórica.

La implantación y el desarrollo de las correlaciones recogidas en la tabla anterior (policórica, punto biserial y tetracórica) supone un grado mayor de dificultad que el tradicional (Pearson), sin embargo, su utilización arroja resultados más fiables, precisos y metodológicamente más correctos. A través de estudios comparativos entre la metodología de Pearson y la de las correlaciones tetracóricas y policóricas, se han encontrado resultados significativamente más favorables hacia estas últimas, haciéndose referencia a un mejor ajuste y una mayor robustez de los modelos, así como a la obtención de mejores niveles de significación en las relaciones entre variables y mejores estimaciones en análisis de fiabilidad (Choi et. al., 2010; Holgado Tello et. al., 2010).

Supongamos que se dispone de dos variables dicotómicas observables Y_1 e Y_2, cuya distribución conjunta se puede ver en la siguiente tabla:

TABLA 3.2 FRECUENCIAS

		Y_2: 0	Y_2: 1	
Y_1	0	a	b	a+b
	1	c	d	c+d
		a+c	b+d	N

donde:

1) a, b, c y d representan las frecuencias de cada par de combinaciones de las variables Y_1 e Y_2.
2) N es el tamaño de la población.

El supuesto básico de las correlaciones tetracóricas es que existen dos variables latentes, Z_1 y Z_2, no observables[2], originalmente continuas y distribuidas de acuerdo a una normal bivariante que definen los valores de Y_1 y de Y_2 en función de unos umbrales (h y k).

$$Y_1 = \begin{cases} 0 \text{ si } Z_1 \text{ es} \leq h \\ 1 \text{ si } Z_1 \text{ es} > h \end{cases}$$

$$Y_2 = \begin{cases} 0 \text{ si } Z_2 \text{ es} \leq k \\ 1 \text{ si } Z_2 \text{ es} > k \end{cases}$$

La correlación tetracórica entre Y_1 e Y_2 es por tanto la correlación existente entre las variables latentes (véase Pearson 1900).

[2] Sin pérdida de generalidad, se supondrá que Z_1 y Z_2 están tipificadas.

Teniendo en cuenta la distribución conjunta de Z_1 y Z_2, se puede observar que:

$$p\ (Y_1=1)=p\ (Z_1>h)=\frac{c+d}{N}=\frac{1}{\sqrt{2\pi}}\int_k^{\infty} e^{-\frac{1}{2}z_1^2}\,dz_1$$

$$p\ (Y_2=1)=p\ (Z_2>k)=\frac{b+d}{N}=\frac{1}{\sqrt{2\pi}}\int_h^{\infty} e^{-\frac{1}{2}z_2^2}\,dz_2$$

Por lo tanto, $h=\Phi^{-1}\left(\frac{c+d}{N}\right)$ y $k=\Phi^{-1}\left(\frac{b+d}{N}\right)$, donde $\Phi^{-1}(u)$ es la inversa de una normal estándar.

Consecuentemente, el coeficiente de correlación tetracórico se podría hallar resolviendo la siguiente ecuación:

$$\frac{d}{N}=p\ (Y_1=1,\ Y_2=1)=p\ (Z_1>h,\ Z_2>k)=$$

$$\frac{1}{2\pi\sqrt{(1-\varphi^2)}}\cdot\int_h^{\infty}\int_k^{\infty} e^{-\frac{1}{2}\frac{1}{1-\varphi^2}(z_1^2+z_2^2-2z_1z_2)}\,dz_1\,dz_2$$

Sin embargo, resolver dicha ecuación no es algo trivial, por lo que diversos autores han propuesto diferentes aproximaciones para el cálculo de este coeficiente.

El propio Pearson propuso algunas de ellas, una de las cuales es también conocida como el coeficiente de asociación de Yule:

$$Q_1=\frac{ad-bc}{ad+bc}$$

Otras de las recomendaciones aportadas por Pearson son:

$$Q_3=\sin\frac{\pi}{2}\frac{\sqrt{ad}-\sqrt{bc}}{\sqrt{ad}+\sqrt{bc}}$$

$$Q_4=\sin\frac{\pi}{2}\left(1+\frac{2bcN}{(ad-bc)(b+c)}\right)^{-1}$$

$$Q_5=\sin\frac{\pi}{2}\left(1+k^2\right)^{-1/2}\text{, donde:}$$

$$k^2 = 4\text{abcd}N^2 / [(\text{ad - bc})^2(\text{a} + \text{d})\ (\text{b} + \text{c})]$$

Las aproximaciones propuestas para el cálculo del coeficiente de correlación tetracórico aportadas por Yule y Digby (véase Edwards, 1957) son las siguientes:

$$\text{Y de Yule} = \frac{(\alpha^{\frac{1}{2}} - 1)}{(\alpha^{\frac{1}{2}} + 1)} \quad \text{donde } \alpha = \frac{ad}{bc}$$

$$\text{H} = \frac{(\alpha^{\frac{3}{4}} - 1)}{(\alpha^{\frac{3}{4}} + 1)}$$

$$\text{J} = \frac{(\alpha^{\frac{\pi}{4}} - 1)}{(\propto^{\frac{\pi}{4}} + 1)}$$

En este trabajo, se utilizará la estimación que se obtiene a partir del programa STATA, que emplea la estimación máximo verosímil obtenida a partir del modelo probit bivariante sin variables explicativas. De acuerdo con Ferrando (1996), la correlación tetracórica obtenida sobre las dicotomías resultantes es el estimador máximo verosímil de la correlación de Pearson que había entre las variables continuas originales.

Además de lo anteriormente expuesto, se ha de tener en cuenta la posible aparición de ciertos inconvenientes o dificultades a la hora de trabajar con las correlaciones tetracóricas, los cuales unidos a la complejidad del procedimiento han llevado posiblemente a su escasa utilización en la mayoría de estudios que aplican la técnica del análisis factorial.

En primer lugar, se ha de considerar que la matriz de correlaciones tetracóricas resultante puede no ser definida positiva, circunstancia bajo la cual no se podría aplicar el análisis factorial en aquellos métodos que requieren de la inversión de dicha matriz (como puede ser, por ejemplo, el de máxima verosimilitud). Este problema sin embargo puede reducirse fácilmente ya sea mediante la utilización de grandes muestras, o aplicando técnicas que suavizan la matriz de correlaciones para hacerla definida positiva y no alteran prácticamente en nada la solución factorial.

Además de esto, puede darse el caso de que aparezcan en el análisis soluciones impropias (estimación de parámetros con valores fuera del rango 0 - 1). Diversos autores (Ferrando, P. y Lorenzo, U., 1994; Flora, D. y Curran, P., 2004) han estudiado las causas que llevan a la aparición de este tipo de soluciones, concluyendo que esto viene determinado en gran parte por el método de estimación escogido (este problema predomina cuando se utiliza mínimos cuadrados ponderados, mínimos cuadrados generalizados o máxima verosimilitud).

Por otro lado, también es posible que aparezcan problemas cuando se dispone de frecuencias muy bajas en alguna de las celdas de las tablas de contingencia, supuesto en el que se obtendrían estimaciones distorsionadas de las correlaciones tetracóricas (si alguna de las celdas tuviera un valor de 0, la correlación no se podría estimar). Como solución a este problema, se suele dar un valor bajo a aquellas celdas que se encuentren en esta situación para poder así realizar el análisis.

Otro de los inconvenientes que puede generar el uso de este tipo de matrices es la variabilidad de los resultados en función del método de estimación que se utilice. Aun teniendo en cuenta esta sensibilidad, se considera que la estimación de la matriz de correlaciones tetracóricas es más adecuada que la aplicación de la tradicional matriz de correlaciones de Pearson.

4. CASO PRÁCTICO

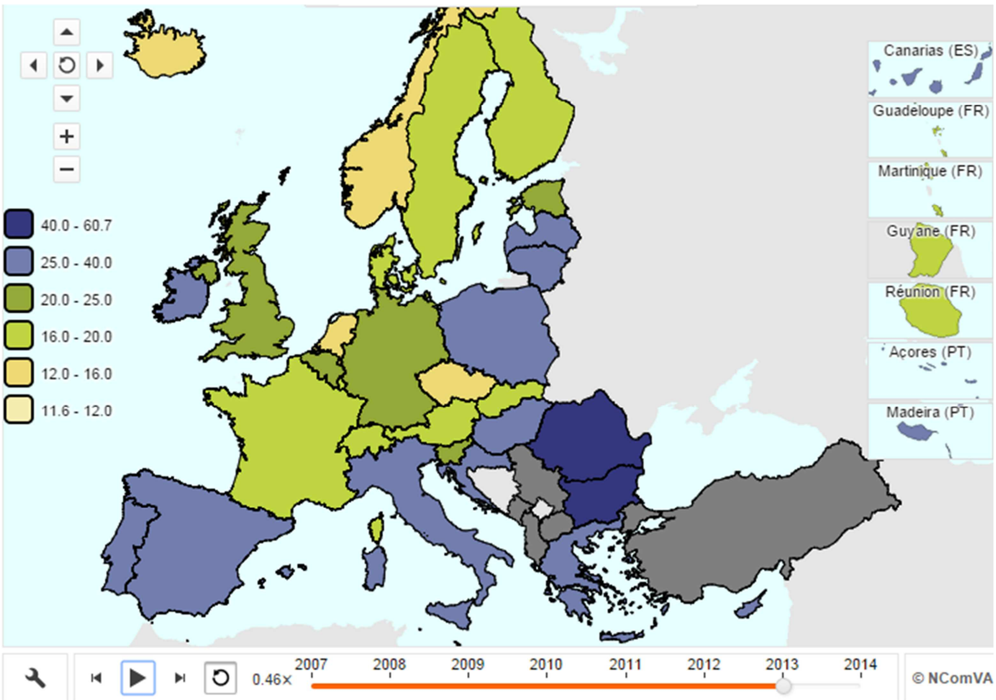
Con el fin de ilustrar el uso de la metodología anteriormente expuesta, se muestra a continuación el análisis factorial destinado a resumir la información extraída de la encuesta de condiciones de vida para España durante el año 2013.

En primer lugar, se hará un breve repaso por la definición y los últimos datos disponibles sobre privación de material. Con la finalidad de corroborar las ventajas anteriormente señaladas que hacen de las matrices tetracóricas la técnica más válida desde el punto de vista metodológico, se procede en segundo lugar a resumir la información de las variables utilizadas en el estudio a través del análisis factorial utilizando dos métodos; en primer lugar aplicando la técnica con la matriz de correlaciones tetracóricas, y en segundo lugar a través de la matriz de correlaciones habitual de Pearson, con la finalidad de comparar dichos resultados. Para concluir con el análisis del caso práctico, se hará referencia a la variabilidad de la solución en función del método utilizado para estimar las cargas factoriales.

4.1 PRIVACIÓN DE MATERIAL

En el año 2008, 80 millones de personas residentes en la Unión Europea vivían por debajo del umbral de la pobreza, siendo predominante la incidencia en mujeres y niños. El recrudecimiento de esta cifra a raíz de la crisis sufrida en el continente ha derivado en la necesidad de la toma de decisiones por parte de las autoridades competentes.

IMAGEN 4.1.1. PORCENTAJE DE POBLACIÓN EN RIESGO DE POBREZA O EXCLUSIÓN SOCIAL



Fuente: Eurostat.

En este contexto, en el año 2010 se aprueba la Plataforma Europea contra la Pobreza y la Exclusión Social, incluida dentro de la estrategia Europa 2020 de crecimiento. Dicha estrategia se desarrolla durante 10 años en Europa con objetivos en 5 ámbitos: empleo, educación, innovación, clima o energía e integración social. Uno de los pilares más relevantes de los que se recogen en esta propuesta es la erradicación de la elevada pobreza en el continente.

En este ámbito aparece el concepto de privación de material, que se basa en la imposibilidad de pagar determinados bienes y servicios que son considerados por los individuos como deseables o necesarios. Con la finalidad de desarrollar los indicadores oficiales de privación de material, se hace referencia en la encuesta a los siguientes elementos, que serán las variables utilizadas en el análisis:

TABLA 4.1.1 VARIABLES UTILIZADAS EN EL ANÁLISIS

NOMBRE VARIABLE	DESCRIPCIÓN
1. Préstamos	¿En los últimos doce meses, ha tenido su hogar atrasos en el pago de plazos, cuotas u otros pagos de algún préstamo por dificultades financieras?
2. Vacaciones	¿Puede su hogar permitirse el lujo de irse de vacaciones una semana, lejos de casa, incluyendo estancias en segunda vivienda o con amigos y familiares?
3. Comida	¿Puede su hogar permitirse una comida con carne, pollo, pescado (o equivalente vegetariano) cada dos días?
4. Gastos inesperados	¿Puede su hogar permitirse un gasto inesperado y pagarlo con sus propios recursos?
5. Caliente	¿Puede su hogar permitirse el lujo de mantener su casa adecuadamente caliente?
6. Teléfono	¿Tiene su hogar teléfono (fijo o móvil)?
7. Televisión	¿Tiene su hogar televisión a color?
8. Lavadora	¿Tiene su hogar lavadora?
9. Coche	¿Tiene su hogar coche o furgoneta para uso privado?

A aquellos individuos que no puedan permitirse el pago de cuatro o más de estos elementos, se les considera en una situación de privación material grave.

Según datos ofrecidos por Eurostat, y tal y como puede verse en la Tabla 4.1.2, en el año 2013 había en España un 6,2% de la población en situación de privación de material (de ellos, un 6,1% son mujeres y un 6,3% son hombres). Este dato se sitúa por debajo de la media europea, siendo ésta del 9,6% en 2013. Atendiendo a la evolución a lo largo de los años desde 2010, se puede ver las cifras siguen una tendencia por lo general creciente tanto en España como en Europa, por lo tanto, queda patente la necesidad de tomar medidas en este ámbito.

TABLA 4.1.2 CARENCIA MATERIAL POR SEXOS. ESPAÑA Y UE-28 (% DE LA POBLACIÓN TOTAL)

	2013		**2012**		**2011**		**2010**	
	España	**UE-28**	**España**	**UE-28**	**España**	**UE-28**	**España**	**UE-28**
Mujeres	6,1	9,8	5,5	10,2	4,6	9,2	5,1	8,6
Hombres	6,3	9,4	6,2	9,6	4,5	8,6	4,7	8,2
Total	**6,2**	**9,6**	**5,8**	**9,9**	**4,5**	**8,9**	**4,9**	**8,4**

Fuente: Eurostat

4.2 IMPLEMENTACIÓN EN SPSS

Para alcanzar los objetivos de la investigación, se llevará a cabo un análisis factorial exploratorio con la finalidad de encontrar las dimensiones que subyacen a las 9 variables dicotómicas que se utilizan para medir la privación de material en España para una muestra de 11.891 individuos.

El paquete estadístico SPSS, que es el que se utilizará en este trabajo, no dispone de un procedimiento ya implantado para realizar el análisis factorial a través de matrices tetracóricas, por lo tanto, se procede al cálculo de la misma a través del programa STATA.

Una vez que se dispone de los datos necesarios (la sintaxis utilizada para ello se recoge en el ANEXO 1), éstos se utilizan como input en la sintaxis del SPSS para llevar a cabo las técnicas propuestas.

4.3 ANÁLISIS FACTORIAL CON CORRELACIONES TETRACÓRICAS

El método utilizado para la extracción de las cargas factoriales es el de componentes principales, que aunque no se considere un método para estimar dichas cargas como tal, sino que más bien es un procedimiento destinado a la reducción de la información, resulta ser el método más adecuado cuando lo que se pretende con el análisis es encontrar el mayor porcentaje de explicación de la varianza (Elosua y Zumbo, 2008).

En este caso, la solución más interpretable era aquella en la que se aplicaba la rotación Varimax.

Se puede ver en la Tabla 4.3.1 de correlaciones tetracóricas, que efectivamente las variables introducidas en el análisis están correlacionadas entre sí con valores lo suficientemente altos como para dar pie a la creación de factores.

TABLA 4.3.1 MATRIZ DE CORRELACIONES TETRACÓRICAS

	Préstamos	Vacaciones	Comida	Gastos inesperados	Caliente	Teléfono	TV	Lavadora	Coche
Préstamos	1,000	0,548	0,350	0,652	0,417	0,332	0,163	0,420	0,426
Vacaciones	0,548	1,000	0,551	0,773	0,592	0,405	0,374	0,192	0,480
Comida	0,350	0,551	1,000	0,518	0,746	0,494	0,403	0,508	0,446
Gastos inesperados	0,652	0,773	0,518	1,000	0,560	0,565	0,298	0,489	0,535
Caliente	0,417	0,592	0,746	0,560	1,000	0,540	0,391	0,593	0,434
Teléfono	0,332	0,405	0,494	0,565	0,540	1,000	0,742	0,642	0,385
TV	0,163	0,374	0,403	0,298	0,391	0,742	1,000	0,676	0,371
Lavadora	0,420	0,192	0,508	0,489	0,593	0,642	0,676	1,000	0,616
Coche	0,426	0,480	0,446	0,535	0,434	0,385	0,371	0,616	1,000

A partir del análisis factorial exploratorio realizado, se observa en la tabla 4.3.2 que la solución óptima es aquella que resume la información contenida en las 9 variables dicotómicas originales en dos únicos factores.

Cada uno de ellos explica respectivamente un 36,3% y un 32,9% de la varianza total, llegándose a alcanzar prácticamente el 70% de manera conjunta. El resultado obtenido es por tanto muy bueno, ya que se consigue reducir el conjunto de la información disponible en 2 dimensiones que recogen por si mismas una gran parte de la varianza total a explicar, superando ambas el 30% de varianza explicada.

TABLA 4.3.2 VARIANZA TOTAL EXPLICADA

	Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
Componente	**Total**	**% de la varianza**	**% acumulado**	**Total**	**% de la varianza**	**% acumulado**
1	4,944	54,932	54,932	3,270	36,328	36,328
2	1,295	14,391	69,324	2,970	32,995	69,324

A continuación, puede verse en la matriz de comunalidades (Tabla 4.3.3) cómo queda recogida cada una de las variables originales en los nuevos factores creados. Se puede observar que aquellas variables que mejor quedan recogidas en el modelo son “gastos inesperados”, “televisión”, “lavadora”, “vacaciones” y “teléfono”. Por el contrario, puede decirse que sobre todo las variables “coche” y “comida” no tienen tanta representación en los nuevos factores a los que pertenecen.

TABLA 4.3.3 COMUNALIDADES

	Inicial	Extracción
Gastos inesperados	1,000	0,805
TV	1,000	0,800
Lavadora	1,000	0,782
Vacaciones	1,000	0,765
Teléfono	1,000	0,747
Caliente	1,000	0,644
Préstamos	1,000	0,624
Comida	1,000	0,579
Coche	1,000	0,493

En la Tabla 4.3.4 se puede ver cuáles son las variables que forman cada uno de los dos factores obtenidos. Se observa que hay dos dimensiones latentes que subyacen a las variables originales a través de las cuales se mide la privación de material. Con la finalidad de simplificar el análisis, se tendrá en cuenta a la hora de decidir la pertenencia a cada factor aquellos valores superiores a 0,5.

En primer lugar, el factor 1 hace referencia a una dimensión de carácter más financiero, estando representada sobre todo por los ítems “préstamos”, “vacaciones”, y “gastos inesperados”, y en menor medida por los ítems “comida”, “caliente” y “coche” (éste último, aunque tiene un valor ligeramente superior en el factor 1, tiene valores prácticamente iguales en ambos factores, por lo tanto, no aporta demasiado a la solución). Como se puede ver, estas variables con mayor peso hacen alusión a la solvencia financiera que las familias tienen (o no) para hacer frente por un lado a gastos corrientes, como pueden ser pagos de un préstamo o unas vacaciones, y por otro lado, a gastos inesperados.

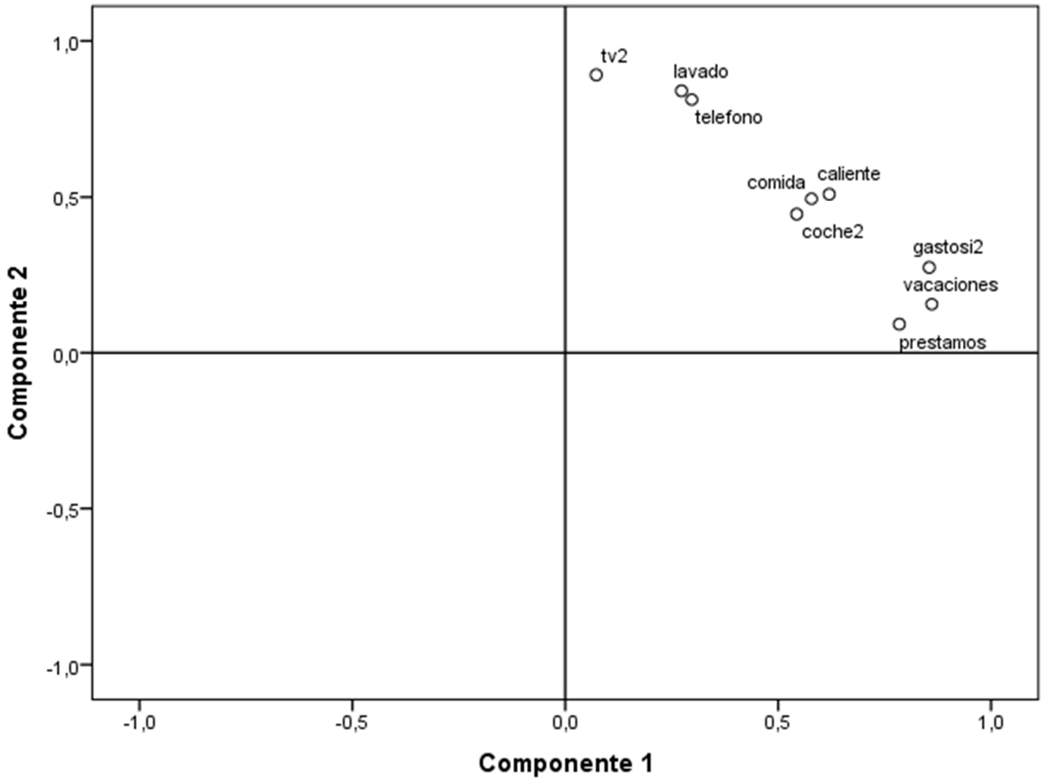
La segunda dimensión se encuentra formada por los ítems “teléfono”, “televisión” y “lavadora”, estando en este caso las 3 variables altamente representadas en el factor con valores significativamente elevados y muy próximos a 1. El segundo factor hace referencia por tanto a una dimensión más material, es decir, a la capacidad que tienen o no los hogares españoles de permitirse tener estos objetos, hoy en día prácticamente imprescindibles en el día a día de los individuos.

TABLA 4.3.4 MATRIZ DE COMPONENTES ROTADOS

	Componente	
	1	**2**
Vacaciones	0,861	
Gastos inesperados	0,855	
Préstamos	0,785	
Caliente	0,620	
Comida	0,578	
Coche	0,543	
TV		0,892
Lavadora		0,841
Teléfono		0,812

Por último, puede comprobarse en el Gráfico 4.3.1 que, como ya se ha visto, los ítems más relacionados con la primera componente son “préstamos”, “vacaciones” y “gastos inesperados”, y los más relacionados con la segunda componente son, las variables “televisión”, “lavadora” y “teléfono”. Tal y como se vió en la matriz de comunalidades y en la matriz de componentes rotados, las variables “comida”, “caliente” y “coche” no son muy relevantes a la hora de crear los factores, y aunque se clasifican en cada uno de los dos factores creados, no se puede decir que encajen de manera tan clara en la solución.

GRÁFICO 4.3.1 GRÁFICO DE COMPONENTES EN ESPACIO ROTADO



4.4 ANÁLISIS COMPARADO

Para poder corroborar las ventajas que presentan las matrices tetracóricas en la aplicación del análisis factorial con variables categóricas con respecto al método habitualmente utilizado con la matriz de correlaciones de Pearson, se expone a continuación un análisis comparado entre los dos métodos.

Del mismo modo que el caso anterior, el método utilizado para la extracción de los factores es el de componentes principales, y la rotación utilizada es la Varimax.

Como puede verse en la Tabla 4.4.1, las correlaciones tetracóricas obtenidas son, en todos los casos, superiores a las obtenidas en la matriz producto-momento de Pearson. Esto es un primer indicio de la mejora que se produce al utilizar éstas correlaciones, ya que la validez del análisis factorial parte de la existencia de correlaciones significativas y suficientemente elevadas entre las variables originales, es decir, esta técnica únicamente tiene sentido cuando las variables están muy correlacionadas entre sí.

TABLA 4.4.1 MATRIZ DE CORRELACIONES DE PEARSON

	Préstamos	Vacaciones	Comida	Gastos inesperados	Caliente	Teléfono	Televisión	Lavadora	Coche
Préstamos	1,000	0,250	0,135	0,316	0,176	0,051	0,028	0,040	0,189
Vacaciones	0,250	1,000	0,156	0,559	0,246	0,042	0,030	0,047	0,173
Comida	0,135	0,156	1,000	0,162	0,396	0,089	0,071	0,127	0,145
Gastos inesperados	0,316	0,559	0,162	1,000	0,247	0,054	0,041	0,052	0,208
Caliente	0,176	0,246	0,396	0,247	1,000	0,085	0,079	0,102	0,175
Teléfono	0,051	0,042	0,089	0,054	0,085	1,000	0,163	0,131	0,086
Televisión	0,028	0,030	0,071	0,041	0,079	0,163	1,000	0,152	0,084
Lavadora	0,040	0,047	0,127	0,052	0,102	0,131	0,152	1,000	0,109
Coche	0,189	0,173	0,145	0,208	0,175	0,086	0,084	0,109	1,000

A continuación, se puede ver en la Tabla 4.4.2 que en este caso, el análisis factorial exploratorio sugiere la extracción de 3 factores, que en su conjunto explican un 51,4% de la varianza total. El primero de ellos es el que más contribuye a la explicación de dicha varianza, teniendo los dos restantes valores más bajos y similares entre ellos.

Al utilizar el método de Pearson, el porcentaje de varianza total explicada por los factores es significativamente inferior al obtenido con las matrices tetracóricas, que llegaba a ser prácticamente del 70%, es decir, se ha reducido en un 20% la explicación de la varianza habiéndose introducido un factor más en la solución. Por lo tanto, se puede ver que la aplicación de este método con las correlaciones de Pearson lleva a la obtención de una solución menos precisa.

TABLA 4.4.2 VARIANZA TOTAL EXPLICADA

	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
Componente	**Total**	**% de la varianza**	**% acumulado**	**Total**	**% de la varianza**	**% acumulado**	**Total**	**% de la varianza**	**% acumulado**
1	2,290	25,446	25,446	2,290	25,446	25,446	1,897	21,081	21,081
2	1,301	14,452	39,898	1,301	14,452	39,898	1,399	15,548	36,629
3	1,037	11,528	51,426	1,037	11,528	51,426	1,332	14,797	51,426
4	,881	9,788	61,214						
5	,859	9,546	70,760						
6	,829	9,216	79,976						
7	,774	8,603	88,579						
8	,592	6,581	95,160						
9	,436	4,840	100,000						

Como ya se ha dicho anteriormente, y según puede verse en la Tabla 4.4.3, en este caso se extraen 3 factores, en los cuales pueden apreciarse de la misma manera que en el caso anterior, las dimensiones financiera y material acompañadas por una nueva que recoge los ítems “comida” y “caliente”, que podría hacer referencia a algunas de las “comodidades” que, aunque parezcan muy básicas y habituales (poder hacer una comida de carne, pollo, pescado o el equivalente vegetariano cada 2 días o ser capaz de mantener la casa caliente), muchos hogares actualmente no se las pueden permitir.

TABLA 4.4.3 MATRIZ DE COMPONENTES ROTADOS

	Componente		
	1	**2**	**3**
Gastos inesperados	0,828		
Vacaciones	0,791		
Préstamos	0,607		
Coche	0,391		
Comida		0,847	
Caliente		0,772	
TV			0,694
Teléfono			0,650
Lavadora			0,584

En este punto se ha de tener en cuenta que la aplicación del análisis factorial con la matriz de correlaciones de Pearson cuando se trabaja con variables de naturaleza categórica puede dar lugar a la aparición de lo que algunos autores han denominado como “factores de dificultad”, es decir, factores que aparecen agrupados por sus características estadísticas, lo cual puede derivar en la presencia de alguna dimensión que no contenga información sustantiva y que por tanto, no sea fácilmente interpretable (Gaviria, 1990; Nunnally y Bernstein, 1994).

Para finalizar el análisis, se presenta en la Tabla 4.4.6 una comparativa entre las comunalidades obtenidas utilizando cada uno de los dos métodos. Se puede ver que todas las variables quedan mejor recogidas en el modelo cuando se utilizan las correlaciones tetracóricas, llegando a alcanzarse en algún caso más del doble de representación que en el caso de las matrices de correlación de Pearson.

TABLA 4.4.6 COMPARATIVA DE COMUNALIDADES

	Comunalidades TTC	Comunalidades PEARSON
	Extracción	**Extracción**
Préstamos	0,624	0,375
Vacaciones	0,765	0,636
Comida	0,579	0,729
Gastos inesperados	0,805	0,694
Caliente	0,644	0,656
Teléfono	0,747	0,427
Televisión	0,800	0,483
Lavadora	0,782	0,376
Coche	0,493	0,254

4.5 SENSIBILIDAD AL MÉTODO DE ESTIMACIÓN

Como ya se ha dicho anteriormente, uno de los inconvenientes que presentan las correlaciones tetracóricas es la posible variabilidad de los resultados en función del método de estimación de las cargas factoriales utilizado, por lo tanto, se presenta a continuación una comparativa entre distintos métodos utilizados por el programa SPSS para estimar las cargas factoriales con la finalidad de determinar la magnitud de dicha sensibilidad.

En primer lugar, se puede ver en la Tabla 4.5.1 que los valores de la varianza total explicada efectivamente difieren en función del método de estimación utilizado, siendo el de componentes principales el que obtiene un mayor valor (como ya se dijo, éste es el método más adecuado cuando lo que se quiere conseguir es un alto valor de varianza explicada). El resto de métodos tienen valores más parecidos entre ellos, y aunque no coinciden en todos los casos, la variabilidad no es muy significativa (todos se sitúan en el intervalo 60%-70%).

TABLA 4.5.1 COMPARATIVA VARIANZA TOTAL EXPLICADA

MÉTODO DE ESTIMACIÓN	VARIANZA TOTAL EXPLICADA
Componentes pricipales	69,32%
Factorización Alfa	61,64%
Mínimos cuadrados generalizados	62,26%
Factorización imagen	65,40%
Factorización de ejes principales	61,64%
Mínimos cuadrados no ponderados	61,64%

En segundo lugar, se analiza en qué medida las comunalidades obtenidas con el análisis factorial difieren según el método de extracción utilizado. Se observa en la tabla 4.5.2 mediante la comparación horizontal de sus filas, que las comunalidades varían prácticamente en todos los casos en mayor o menor medida, por lo tanto, habrá algunos métodos que recojan más información de las variables originales en los factores que se creen que otros.

Por ejemplo, para la variable préstamos se obtiene un valor de 0,624 utilizando el análisis de componentes principales, mientras que si el método escogido es el de factorización imagen se consigue únicamente un 0,387, valor significativamente inferior llegando a ser prácticamente la mitad que el anterior.

TABLA 4.5.2. COMPARATIVA COMUNALIDADES

	Componentes pricipales	Factorización Alfa	Factorización imagen	Ejes principales	MCNP
Préstamos	0,624	0,444	0,387	0,441	0,441
Vacaciones	0,765	0,700	0,981	0,735	0,735
Comida	0,579	0,509	0,460	0,507	0,507
Gastos inesperados	0,805	0,808	0,778	0,797	0,796
Caliente	0,644	0,591	0,565	0,583	0,583
Teléfono	0,747	0,647	0,628	0,658	0,658
Televisión	0,800	0,686	0,691	0,653	0,653
Lavadora	0,782	0,752	0,947	0,761	0,762
Coche	0,493	0,412	0,449	0,413	0,413

Por último, es importante señalar que la agrupación de las variables originales en los distintos factores, así como su posterior interpretación varía del mismo modo en función del método escogido, llegando en algunos casos a soluciones muy difícilmente interpretables.

Se concluye en función de lo anterior, que efectivamente existen notables diferencias en los resultados dependiendo del método utilizado para la estimación de las cargas factoriales usando para el análisis la matriz de correlaciones tetracóricas, siendo éste uno de los inconvenientes que se argumentan en contra de la utilización de dichas correlaciones.

5) CONCLUSIONES

El objetivo fundamental de esta investigación es la presentación y aplicación de las correlaciones tetracóricas al análisis factorial en presencia de variables de naturaleza categórica.

En este tipo de correlaciones se asume que existen unas variables latentes continuas que subyacen a las variables dicotómicas observadas. La correlación tetracórica mide por tanto la relación existente entre dichas variables latentes.

A través de la aplicación de esta metodología, se ha podido observar que los resultados obtenidos con las correlaciones tetracóricas son significativamente mejores que los obtenidos a partir de las correlaciones habitualmente utilizadas de Pearson.

Utilizando el método de componentes principales para la extracción, y a través de la solución rotada, se han podido identificar adecuadamente las dimensiones latentes que subyacen a las variables observadas en el caso analizado.

Por un lado, aparece una dimensión financiera relacionada con la capacidad de los hogares para hacer frente a distintos pagos, y por otro lado existe una segunda dimensión que hace referencia a algunos objetos materiales que los hogares pueden (o no) poseer. Cuando se utiliza la matriz de correlaciones de Pearson, aparece una tercera dimensión que hace referencia a ciertas comodidades que pueden existir en los hogares, como son el hecho de poder mantener la casa caliente o el poderse permitir una comida de carne, pollo, o pescado (o el equivalente vegetariano) cada dos días. En este caso, se ha observado que, además de incluirse una nueva dimensión en la solución, la varianza explicada se ha visto considerablemente reducida.

Si bien es cierto que este método presenta algunos inconvenientes que no se han de obviar, como la sensibilidad que existe al método de estimación de las cargas factoriales, se pone de manifiesto que su uso es, metodológicamente hablando, y de cara a obtener unos mejores resultados, el más adecuado cuando se trabaja con este tipo de variables.

Se puede concluir por tanto, que las diferencias existentes en la aplicación de los dos métodos propuestos indican a grandes rasgos, que el uso de correlaciones tetracóricas en aquellos estudios en los que se utilizan variables de naturaleza categórica deriva en una mejora de las estimaciones obtenidas. Con este método, se ha obtenido un mayor porcentaje de explicación de la varianza total de las variables, llegando ésta a alcanzar prácticamente el 70% con dos únicos factores formados a partir de las 9 variables dicotómicas originales, frente al 51% conseguido con la estimación habitual de Pearson. Además, se ha de tener en cuenta que, en cuanto a las comunalidades, prácticamente la totalidad de las variables quedan recogidas en mayor medida cuando se utilizan las correlaciones tetracóricas para resumir la información.

6) REFERENCIAS BIBLIOGRÁFICAS

Anglim, J. (2009): "Tetrachoric Correlations | Overview and Resources". Disponible en: http://jeromyanglim.blogspot.com.es/2009/09/tetrachoric-correlations-overview-and.html [Consulta: 20/05/2015]

Burga, A. (2006): "La unidimensionalidad de un instrumento de medición: perspectiva factorial" *Revista de Psicología,* Vol. 14, Nº1, pp. 55-80.

Choi, J. et. al (2010): "Correlational analysis of ordinal data: from Pearson´s r to Bayesian polychoric correlation". *Asia Pacific, Vol.11*, pp. 459-466.

Dekkers, G. (2008): "Are you unhappy? Then you are poor! Multi-dimensional poverty in Belgium" *International Journal of Sociology and Social Policy,* Vol. 28, pp. 502-515.

Digby, P. (1983): "Approximating the Tetrachoric Correlation Coefficient", *Biometrics,* Vol. 39, Nº3, pp. 753-757.

Divgi, D. (1979): "Calculation of the tetrachoric correlation coefficient", *Psychometrika,* Vol. 44, Nº 2, pp.169-172.

Domínguez, S. (2014): "Matrices policóricas/tetracóricas o matrices Pearson? Un estudio metodológico" *Revista Argentina de Ciencias del Comportamiento,* Vol. 6, Nº 1, pp. 39-48.

Edwards, A. (1963): "The measure of association in a 2 x 2 table". *Journal of the Royal Statistical Society*, Vol. 126, pp. 109-114.

Edwards, J. (1957): "A note on the practical interpretation of 2 x 2 tables". *British Journal of preventive and Social Medicine,* Vol. 11, pp. 73-78.

Elosua, P. y Zumbo, B. (2008): "Coeficientes de fiabilidad para escalas de respuesta categórica ordenada". *Psicothema,* Nº 20 (4), pp. 896-901.

Eurostat (2009): "EU-SILC module on Material Deprivation". Disponible en: http://ec.europa.eu/eurostat/documents/1012329/6071326/2009_Module_Material_deprivation.pdf/8a3e135d-f926-41bd-9f0c-3f4f87fd5356 [Consulta: 27/04/2015]

Eurostat (2015): "Income and living conditions". Disponible en: http://ec.europa.eu/eurostat/web/income-and-living conditions/data/database?p_p_id=NavTreeportletprod_WAR_NavTreeportletprod_INSTANCE_CEM7npyJJgVL&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_col_id=column-2&p_p_col_count=1 [Consulta: 27/04/2015]

Ferrando, P. (1996): "Evaluación de la unidimensionalidad de los ítems mediante análisis factorial" *Psicothema,* Vol. 8, Nº 2, pp. 397-410.

Ferrando, P. y Anguiano-Carrasco C. (2010): "El análisis factorial como técnica de investigación en psicología". Disponible en: http://www.papelesdelpsicologo.es/pdf/1793.pdf [Consulta: 10/06/2015].

Ferrando, P. y Lorenzo, U. (1993): "Algunas relaciones entre el modelo de un factor común y el modelo logístico de dos parámetros" *Psicothema,* Vol. 5, Nº 2, pp. 403-412.

Ferrando, P. y Lorenzo, U. (1994): "Recuperación de la solución factorial a partir de variables dicotomizadas". *Psicothema,* Vol. *6* (3), pp. 483-491.

Flora, D. y Curran, P. (2004): "An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data". *Psychological Methods,* Vol. *9* (4), pp. 466-491.

Freiberg hoffmann, A. et al (2013): "Correlaciones policóricas y tetracóricas en estudios factoriales exploratorios y confirmatorios", *Ciencias Psicológicas*, Vol. 7 (2), pp. 151-164.

Gavira, J. L. (1990): "Factores de dificultad en el análisis de ítems. Qué son, por qué aparecen y posibles soluciones" *Revista Complutense de Educación,* Vol. 1 (1), pp. 95-108.

Hofacker, C. y Muthén, B. (1988): "Testing the assumptions underlying tetrachoric Correlations". *Psychometrika*, Vol. 53 (4), pp. 563-578.

Holgado, T., **et al (2010):** "Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables". *Quality & Quantity,* Vol. *44* (1), pp. 153-166.

Juras, J. y Pasarié, Z. (2006): "Application of tetrachoric and polychoric correlation coefficients to forecast verification", *Geofizika,* Vol. 23, Nº1, pp. 60-82.

Lorenzo, U. y Ferrando, P. (2012): "A comprehensive SPSS program for estimating the tetrachoric correlation", *Behavior Research Methods,* Vol. 44, pp. 1191-1196.

Manly, B. (2000): "Multivariate Statistical Methods", Chapman & Hall/CRC.

Morales, P. (2006): "Medición de actitudes en psicología y educación: Construcción de escalas y problemas metodológicos" *Universidad Pontifica de Comillas.*

Nunnally, J., y Bernstein, I. (1994): "*Psychometric theory",* McGraw-Hill.

Pearson, K. (1900): "Philosophical transations" University College, London.

Peña, D. (2002): *Análisis de datos multivariantes,* McGraw-Hill.

SECQ: "Correlación". Disponible en: http://www.seqc.es/dl.asp?175.145.205.255.15.30.27.21.118.133.24.113.255.173.42.13.166.147.73.155.249.7.59.165.241.44.201.87.76.120.115.248.0.229.154.28.69.158.114.106.167.74.239.206 [Consulta: 04/03/2015].

Westberg, M.: "Tetrachoric Correlation Coefficient" Disponible en: http://ec.europa.eu/eurostat/data/database?node_code=ilc_mddd11 [Consulta: 10/02/2014]

7) ANEXO 1

```
MATRIX DATA VARIABLES= prestamos  vacaciones  comida  gastosi2  caliente  telefono  tv2
lavado coche2

/contents=corr

/N=11891.

BEGIN DATA.

1

0,5481  1

0,3503  0,5505  1

0,6523  0,7731  0,5180  1

0,4166  0,5922  0,7458  0,5597  1

0,3317  0,4050  0,4941  0,5645  0,5397  1

0,1630  0,3741  0,4026  0,2976  0,3910  0,7418  1

0,4202  0,1918  0,5077  0,4887  0,5930  0,6418  0,6762  1

0,4259  0,4796  0,4458  0,5353  0,4344  0,3849  0,3713  0,6161  1

END DATA.

EXECUTE.

  FACTOR MATRIX=IN(COR=*)

  /PRINT= EXTRACTION ROTATION CORRELATION REPR

 /PRINT INITIAL KMO EXTRACTION

      /PLOT EIGEN ROTATION

   /CRITERIA MINEIGEN(1) ITERATE(25)

  /EXTRACTION=PC

   /ROTATION=VARIMAX

   /METHOD=CORRELATION.
```