Universidad Nacional Agraria La Molina
Dpto. de Estadística e Informática
Curso: Técnicas Multivariadas I

ANALISIS FACTORIAL[1]

I. El Modelo de Análisis Factorial

Formulación del Modelo

Considerando que $X_1, X_2, \ldots, X_p$ son variables estandarizadas; es decir, son variables con media 0 y variancia 1.

El modelo de análisis factorial se define de la siguiente forma:

$$
\begin{aligned}
X_1 &= l_{11}F_1 + l_{12}F_2 + \ldots + l_{1m}F_m + e_1 \\
X_2 &= l_{21}F_1 + l_{22}F_2 + \ldots + l_{2m}F_m + e_2 \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
X_p &= l_{p1}F_1 + l_{p2}F_2 + \ldots + l_{pm}F_m + e_p
\end{aligned}
$$

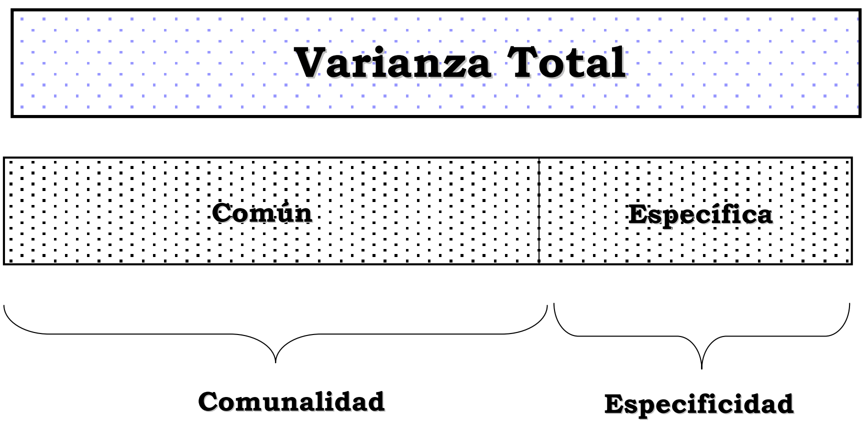
Donde $F_1, F_2, \ldots, F_m$ son factores comunes; $e_1, e_2, \ldots, e_p$ son factores únicos o específicos; l_{jh} es el peso del factor h en la variable j. A los coeficientes de este tipo se les denomina cargas factoriales.

Cada una de las variables observables es una combinación lineal de m factores comunes (m < p) y de una factor único.

Matricialmente, se tiene:

$$
\begin{bmatrix} X_1 \\ X_2 \\ . \\ . \\ X_p \end{bmatrix} = \begin{bmatrix} l_{11} & l_{12} & \ldots & l_{1m} \\ l_{21} & l_{22} & \ldots & l_{2m} \\ . & . & . & . \\ . & . & . & . \\ l_{p1} & l_{p2} & \ldots & l_{pm} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ . \\ . \\ F_m \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ e_p \end{bmatrix}
$$

En forma resumida, se tiene que: $X = LF + e$



1. Adaptado del libro de "Análisis Multivariante Aplicado. Aplicaciones al marketing, investigación de mercados, economía, dirección de empresas y turismo" de Uriel, Ezequiel & Aldas, Joaquín. Editorial Thomson. 2005. España.

Hipótesis del Modelo

Las hipótesis sobre los factores comunes son las siguientes:

La esperanza de cada uno de los factores comunes es nula, es decir, $E(f) = 0$

La matriz de covariancias de los factores comunes es $E(ff')=I$

La matriz de covariancias de los factores comunes es la matriz identidad, lo que implica que la variancia de cada uno de los factores es 1 y que los factores están no correlacionados entre sí, ya que todos los elementos de fuera de la diagonal principal son nulos. Así pues, los factores comunes son variables estandarizadas de media 0 y variancia 1, y que además no están correlacionadas entre sí.

Las hipótesis sobre los factores únicos son las siguientes:

La esperanza de cada uno de los factores únicos es nula, es decir, $E(e) = 0$

La matriz de covariancias de los factores comunes es $E(ee')=\Omega$

Donde Ω es una matriz diagonal.

La matriz de covariancias de los factores únicos es una matriz diagonal, lo que implica que las variancias de los factores únicos pueden ser distintas y también que los factores no están correlacionados entre sí.

La hipótesis sobre la relación entre los factores comunes y únicos es la siguiente: $E(fe')=0$

Para poder realiza inferencias que permitan distinguir, para cada variable, entre los factores comunes y el factor único, es necesario postular que los primeros no estén correlacionados con este último.

Propiedades del Modelo

Como las variables X son variables estandarizadas, su matriz de covariancias es igual a la matriz de correlación poblacional R_p; es decir,

$$E(XX') = R_p = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{21} & 1 & \dots & \rho_{2p} \\ . & . & . & . \\ \rho_{p1} & \rho_{p2} & \dots & 1 \end{bmatrix}$$

Como se trata de variables estandarizadas, la variancia de cada una de ellas es igual a 1. La matriz de correlación poblacional se puede descomponer de la siguiente forma: $R_p = LL' + \Omega$, la cual puede descomponerse de la siguiente forma:

$$\begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{21} & 1 & \dots & \rho_{2p} \\ . & . & . & . \\ \rho_{p1} & \rho_{p2} & \dots & 1 \end{bmatrix} = \begin{bmatrix} l_{11} & l_{12} & \dots & l_{1m} \\ l_{21} & l_{22} & \dots & l_{2m} \\ . & . & . & . \\ l_{p1} & l_{p2} & \dots & l_{pm} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & \dots & l_{p1} \\ l_{12} & l_{22} & \dots & l_{p2} \\ . & . & . & . \\ l_{1m} & l_{2m} & \dots & l_{pm} \end{bmatrix} + \begin{bmatrix} \omega_1^2 & 0 & \dots & 0 \\ 0 & \omega_2^2 & \dots & 0 \\ . & . & . & . \\ 0 & 0 & \dots & \omega_p^2 \end{bmatrix}$$

El primer elemento de la diagonal principal del primer miembro, que es la variancia de variable estandarizada X_1, puede descomponerse de la siguiente forma:

$$1 = l_{11}^2 + l_{12}^2 + \dots + l_{1m}^2 + \omega_1^2$$

De forma genérica la variancia de la variable estandarizada X_j se puede descomponer de la siguiente forma:

$$1 = l_{j1}^2 + l_{j2}^2 + \dots + l_{jm}^2 + \omega_j^2$$

La suma de los “m” primeros términos va a ser designada por h_j^2, es decir,

$$h_j^2 = l_{j1}^2 + l_{j2}^2 + \dots + l_{jm}^2$$

De tal forma, que: $1 = h_j^2 + \omega_j^2$

Donde h_j^2 es la **comunalidad**, que se define como la parte de la variancia que es debida a los factores comunes, mientras que w_j^2 es la **especificidad**, que se define como la parte de la variancia que es debida a los factores únicos.
Se puede obtener el coeficiente de correlación entre par de variable originales como función de los factores comunes,

$$\rho_{hj} = l_{h1} l_{j1} + l_{h2} l_{j2} + \dots + l_{hm} l_{jm}$$

El problema que se plantea en el análisis factorial es la estimación de los coeficientes l_{jh}. A los coeficientes estimados se les denomina ***cargas factoriales***. Para estimar dichas cargas se usan varios métodos:

- Método de componentes principales
- Método de componentes principales iteradas o ejes principales
- Mínimos cuadrados no ponderados
- Mínimos cuadrados generalizados
- Máxima verosimilitud

Estimación usando el método de componentes principales

Usando ACP las "p" componentes se pueden expresar como:

$$Y_1 = u_{11}X_1 + u_{12}X_2 + \ldots + u_{1p}X_p$$
$$Y_2 = u_{21}X_1 + u_{22}X_2 + \ldots + u_{2p}X_p$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$Y_p = u_{p1}X_1 + u_{p2}X_2 + \ldots + u_{pp}X_p$$

En conjunto de ecuaciones es reversible, pudiéndose demostrar que es posible expresar las variables X en función de las componentes Z.

$$X_1 = u_{11}Y_1 + u_{21}Y_2 + \ldots + u_{p1}Y_p$$
$$X_2 = u_{12}Y_1 + u_{22}Y_2 + \ldots + u_{p2}Y_p$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$X_p = u_{1p}Y_1 + u_{2p}Y_2 + \ldots + u_{pp}Y_p$$

En ACP las componentes Z no están estandarizadas, mientras que los factores comunes si lo están. Para superar este problema se pueden utilizar las componentes estandarizadas

$$F_h = \frac{Y_h}{\sqrt{\lambda_h}}, h = 1, 2, \ldots, p$$

$$Y_h = F_h\sqrt{\lambda_h}, h = 1, 2, \ldots, p$$

Con lo que la ecuación j-ésima puede expresarse así:

$$X_j = u_{1j}\sqrt{\lambda_1}F_1 + u_{2j}\sqrt{\lambda_2}F_2 + \ldots + u_{pj}\sqrt{\lambda_p}F_p$$

Teniendo en cuenta que $u_{hj}\sqrt{\lambda_h}$ es precisamente el coeficiente de correlación entre la variable j-ésima y la componente h-ésima, se puede expresar de la siguiente forma:

$$X_j = r_{1j}F_1 + r_{2j}F_2 + \ldots + r_{pj}F_p$$
$$X_j = r_{1j}F_1 + r_{2j}F_2 + \ldots + r_{mj}F_m + (r_{m+1,j}F_{m+1} + \ldots + r_{pj}F_p)$$

Se tiene también que:

$$X_j = l_{j1}F_1 + l_{j2}F_2 + \ldots + l_{jm}F_m + e_j$$

Los "m" factores F_h se estiman mediante las "m" primeras componentes principales estandarizadas Y_h y la estimación de los coeficientes l_{jh} viene dada por:

$$\hat{l}_{jh} = r_{1j}$$
$$\hat{l}_{j2} = r_{2j}$$
$$.$$
$$.$$
$$\hat{l}_{jm} = r_{mj}$$

Una vez estimados los coeficientes anteriores, se puede estimar la comunalidad de la variable X_j de la siguiente forma:

$$\hat{h}_j^2 = \hat{l}_{j1}^2 + \hat{l}_{j2}^2 + \ldots + \hat{l}_{jm}^2 ..$$

La estimación del factor único e_1 viene dada por:

$$\hat{e}_1 = r_{m+1,j}F_{m+1} + \ldots + r_{pj}F_p$$

La especificidad se puede estimar directamente mediante la siguiente expresión:

$$\hat{\omega}_j^2 = 1 - \hat{h}_j^2 .$$

Puntuaciones de los factores

Denominados scores factoriales. Una vez que se tienen los factores puede interesar conocer que puntuación obtendrían los sujetos en estos factores. Para ello hay que calcular lo que se conoce como puntuaciones factoriales de cada individuo.

El cálculo de las puntuaciones factoriales se realiza a partir de la matriz factorial rotada y se basa en el modelo de la regresión múltiple.

Las puntuaciones factoriales exactas sólo pueden calcularse estrictamente cuando el método de extracción ha sido el de Análisis de Componentes Principales. Con los otros métodos sólo podrán hacerse estimaciones por medio de algún método correlacionado. Estas estimaciones se pueden realizar por distintos métodos. Los procedimientos más conocidos son el de Regresión, Anderson-Rubin y Bartlett.

Ejemplo de Aplicación Nº 1

1	2	3	4	5	6	7	8	9	10	11	12
X1	X2	Z1	Z2	Puntuación ACP 1	Puntuación ACP 2	Carga AF 1	Carga AF 2	Coeficiente AF 1	Coeficiente AF 2	Puntuación AF 1	Puntuación AF 2
775104	23795	1.25752	0.55653	1.28273	0.49567	0.87921	0.47643	0.56869	1.04947	1.03163	0.73567
775218	58778	1.25838	1.92752	2.25277	-0.47315	0.87921	-0.47643	0.56869	-1.04947	1.81179	-0.70224
700963	1531	0.69751	-0.31600	0.26977	0.71666					0.21696	1.06365
674063	-12756	0.49432	-0.87591	-0.26982	0.96890					-0.21700	1.43802
631003	14729	0.16908	0.20123	0.26185	-0.02274					0.21059	-0.03375
537744	9059	-0.53534	-0.02098	-0.39337	-0.36371					-0.31637	-0.53981
489155	12541	-0.90235	0.11548	-0.55640	-0.71972					-0.44748	-1.06819
448465	13495	-1.20969	0.15287	-0.74729	-0.96348					-0.60100	-1.42998
445853	-34824	-1.22942	-1.74076	-2.10023	0.36157					-1.68911	0.53663

Correlación: X1, X2

```
Correlación de Pearson de X1 y X2 = 0.546
```

A) Análisis de componente principal: X1, X2

```
Análisis de los valores y vectores propios de la matriz de correlación

Valor propio  1.5460  0.4540
Proporción     0.773   0.227
Acumulada      0.773   1.000


Variable    PC1     PC2
X1        0.707   0.707
X2        0.707  -0.707
```

B) Estadísticas descriptivas: Puntuacion ACP 1, Puntuacion ACP 2

```
Variable          N  Media  Desv.Est.  Varianza
Puntuacion ACP 1  9  0.000      1.243     1.546
Puntuacion ACP 2  9  0.000      0.674     0.454
```

C) Análisis factorial: X1, X2

```
Análisis factorial del componente principal de la matriz de correlación

Cargas de factores no rotados y comunalidades

Variable  Factor1  Factor2  Comunalidad
X1          0.879    0.476        1.000
X2          0.879   -0.476        1.000

Varianza   1.5460   0.4540       2.0000
% Var       0.773    0.227        1.000


Coeficientes de puntuación de factores

Variable  Factor1  Factor2
X1          0.569    1.049
X2          0.569   -1.049
```

D) Estadísticas descriptivas: Puntuacion AF 1, Puntuacion AF 2

```
Variable         N  Media  Desv.Est.  Varianza
Puntuacion AF 1  9  0.000      1.000     1.000
Puntuacion AF 2  9  0.000      1.000     1.000
```

Ejemplo de Aplicación Nº 2[2]

Id	X1	X2	X3	X4	X5	X6	Z1	Z2	Z3	Z4	Z5	Z6
1	4,1	0,6	6,9	4,7	2,4	5,2	0,4429	-1,4753	-0,7169	-0,4843	-0,3453	-1,1172
2	1,8	3,0	6,3	6,6	4,0	8,4	-1,2985	0,5319	-1,1497	1,1950	1,7315	0,9014
3	3,4	5,2	5,7	6,0	2.7	8,2	-0,0871	2,3719	-1,5824	0,6647	0,0441	0,7753
4	2,7	1,0	7,1	5,9	2,3	7,8	-0,6171	-1,1408	-0,5727	0,5763	-0,4751	0,5229
5	6,0	0,9	9,6	7,8	4,6	4,5	1,8815	-1,2244	1,2304	2,2556	2,5102	-1,5588
.												
.												
.												
100	2,5	1,8	9,0	5,0	3,0	6,0	-0,7685	-0,4717	0,7977	-0,2192	0,4335	-0,6125

A) ANÁLISIS DE FACTORES CON 6 VARIABLES SIN ROTACIÓN USANDO MINITAB

Análisis factorial: X1, X2, X3, X4, X5, X6

Análisis factorial del componente principal de la matriz de **correlación**

Cargas de factores no rotados y comunalidades

```
Variable  Factor1  Factor2  Comunalidad
X1         -0.627    0.514        0.658
X2          0.758   -0.068        0.580
X3         -0.730    0.335        0.646
X4          0.494    0.799        0.882
X5          0.424    0.832        0.872
X6          0.767   -0.167        0.616

Varianza   2.5127   1.7397       4.2524
% Var       0.419    0.290        0.709
```

Coeficientes de puntuación de factores

```
Variable  Factor1  Factor2
X1         -0.250    0.295
X2          0.302   -0.039
X3         -0.291    0.192
X4          0.196    0.459
X5          0.169    0.478
X6          0.305   -0.096
```

Puntuaciones de los individuos

```
Puntuacion 1  Puntuacion 2
-0.84169     -0.22961
 1.62044      0.66475
 1.57214     -0.17131
 0.16879     -0.26087
-0.80632      3.22710
     .           .
```

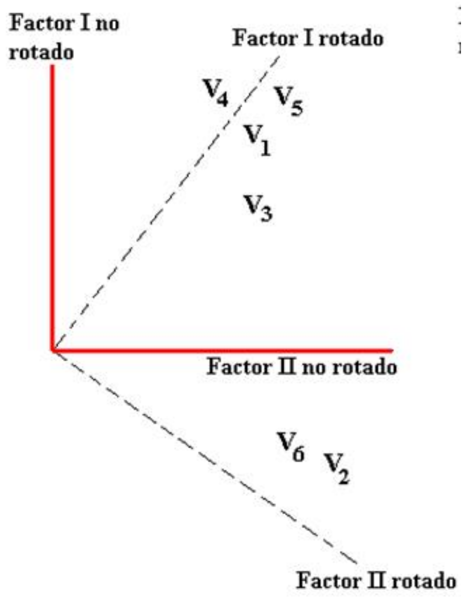
[2] Basado en Hair, Joseph; Anderson, Rolph; Thatam, Ronald & Black, William. "Análisis Multivariante". Editorial Prentice Hall. 1999. España.

II. Rotación de Factores

La interpretación se efectúa considerando las correlaciones del factor con las variables observadas iniciales, analizando cuáles de estas variables contribuyen en mayor medida a la formación del mismo y a facilitar la identificación de un nombre para el factor.

Se presentan casos en los que el coeficiente de correlación de las variables iniciales con el factor no permite una clara interpretación por no destacarse algunas variables especialmente correlacionadas.

Se puede proceder a efectuar rotaciones para aumentar el valor de los coeficientes de correlación de algunas de las variables con los factores y facilitar la interpretación.



Entre los métodos de rotación más comunes se tiene:

- **Varimax**
  Maximiza la suma de variancias de las cargas factoriales dentro de cada factor.
- **Quartimax**
  Trata de conseguir que una variable tenga una carga alta con un factor y baja con los demás.
- **Equimax**
  Solución intermedia entre los anteriores.

Para interpretar la matriz de rotación se debe tener en cuenta:

- El objetivo de la matriz factorial rotada consiste en identificar cada una de las dimensiones extraídas.
- Se efectúa escogiendo para cada factor, las variables originales (X's) cuyas correlaciones con el factor sean las más elevadas (cercanas a +1 o a -1).
- Hay que tener en cuenta que el total de la variabilidad retenida permanece constante, pero varía la variabilidad retenida por cada uno de los factores.

B) ANÁLISIS DE FACTORES CON 6 VARIABLES Y CON ROTACIÓN VARIMAX USANDO MINITAB

Análisis factorial: X1, X2, X3, X4, X5, X6

```
Análisis factorial del componente principal de la matriz de correlación


Cargas de factores rotados y comunalidades
Rotación Varimax

Variable  Factor1  Factor2  Comunalidad
X1         -0.787    0.195        0.658
X2          0.714    0.264        0.580
X3         -0.803   -0.011        0.646
X4          0.103    0.933        0.882
X5          0.025    0.934        0.872
X6          0.764    0.178        0.616

Varianza   2.3701   1.8823       4.2524
% Var       0.395    0.314        0.709


Coeficientes de puntuación de factores

Variable  Factor1  Factor2
X1         -0.352    0.159
X2          0.289    0.094
X3         -0.345    0.049
X4         -0.020    0.499
X5         -0.053    0.504
X6          0.317    0.044
```

III. Contrastes en el Modelo Factorial

Los siguientes contrastes tratan de analizar la pertinencia de aplicación del análisis factorial a un conjunto de variables observables: matriz de correlaciones, contraste de esfericidad de Bartlett y la medida de adecuación muestral de Kaiser, Meyer y Olkin.

Matriz de Correlaciones

Comprobar si a simple vista el número de correlaciones superiores a 0,5 es considerable.

Contraste de esfericidad de Bartlett

Consiste en comprobar que la matriz de correlaciones es significativamente distinta de la matriz identidad. En caso que fuera una matriz identidad no habría correlación entre variables y no tendría sentido seguir con el análisis factorial

$$\begin{cases} H_0 : |R_p| = 1 \\ H_1 : |R_p| \neq 1 \end{cases}$$

$$\chi^2_{0.5(p^2-p)} = -\left[n - 1 - \frac{1}{6}(2p+5)\right] Ln|R|$$

Medida de Adecuación Muestral

Los estadísticos Kaiser, Meyer y Olkin propusieron una medida de adecuación de la muestra al análisis factorial (KMO).

Un coeficiente de correlación parcial mide la correlación entre dos variables, una vez que se han descontado los efectos lineales de otras variables. En un modelo factorial se pueden interpretar esos efectos de otras variables como los correspondientes a los factores comunes. Por lo tanto, el coeficiente de correlación parcial entre dos variables sería equivalente al coeficiente de correlación entre los factores únicos de dos variables. Según el modelo de análisis factorial, los coeficientes de correlación teóricos calculados entre cada par de factores únicos son nulos por hipótesis. Si los coeficientes de correlación parcial constituyen una aproximación a dichos coeficientes teóricos, deben estar próximos a 0.

La medida de KMO se define de la siguiente forma:

$$KMO = \frac{\sum\sum_{i \neq j} r_{ij}^2}{\sum\sum_{i \neq j} r_{ij}^2 + \sum\sum_{i \neq j} a_{ij}^2}$$

Los r_{ij} son coeficientes de correlación observados entre variables originales,, mientras que los a_{ij} son coeficientes de correlación parcial entre variables originales

El valor de KMO variará entre 0 y 1

KMO	Intepretación
0.8 ó 0.9	Bueno
0.7	Intermedio
0.5	Aceptable
Menos de 0.5	Inaceptable

Basada en el KMO, se puede calcular también una medida de adecuación muestral individual para cada una de las variables. Esta medida, denominada MSA (Measure of Sampling Adequacy) se define de la siguiente forma:

$$MSA_i = \frac{\sum_{j \neq i} r_{ij}^2}{\sum_{j \neq i} r_{ij}^2 + \sum_{j \neq i} a_{ij}^2}$$

Un valor próximo a 1 para una variable indicará que dicha variables es adecuada para su tratamiento en el análisis factorial con el resto de variables.

C) ANÁLISIS DE FACTORES CON 7 VARIABLES USANDO SPSS

Matriz de correlaciones[a]

		Rapidez de servicio	nivel de precios	flexibilidad de precios	Imagen del fabricante	Imagen de los vendedores	Calidad del producto	Servicio
Correlación	Rapidez de servicio	1.000	-.349	.509	.050	.077	-.483	.612
	nivel de precios	-.349	1.000	-.487	.272	.185	.470	.513
	flexibilidad de precios	.509	-.487	1.000	-.116	-.035	-.448	.067
	Imagen del fabricante	.050	.272	-.116	1.000	.788	.200	.299
	Imagen de los vendedores	.077	.185	-.035	.788	1.000	.177	.240
	Calidad del producto	-.483	.470	-.448	.200	.177	1.000	-.055
	Servicio	.612	.513	.067	.299	.240	-.055	1.000
Sig. (Unilateral)	Rapidez de servicio		.000	.000	.309	.222	.000	.000
	nivel de precios	.000		.000	.003	.032	.000	.000
	flexibilidad de precios	.000	.000		.125	.366	.000	.255
	Imagen del fabricante	.309	.003	.125		.000	.023	.001
	Imagen de los vendedores	.222	.032	.366	.000		.039	.008
	Calidad del producto	.000	.000	.000	.023	.039		.293
	Servicio	.000	.000	.255	.001	.008	.293	

a. Determinante = .003

```
# Lectura de datos con 7 variables
datos=read.delim("hatco-factorial.txt")
str(datos)

# No considerar la primera columna Id
datos1=datos[,-1]
str(datos1)

# Analisis descriptivo y Analisis de Correlacion
library(psych)
describe(datos1)
```

```
   vars   n mean   sd median trimmed  mad min  max range  skew kurtosis   se
x1    1 100 3.52 1.32   3.40    3.53 1.48 0.0  6.1   6.1 -0.08    -0.59 0.13
x2    2 100 2.36 1.20   2.15    2.30 1.19 0.2  5.4   5.2  0.46    -0.59 0.12
x3    3 100 7.89 1.39   8.05    7.95 1.70 5.0 10.0   5.0 -0.28    -1.12 0.14
x4    4 100 5.25 1.13   5.00    5.23 1.04 2.5  8.2   5.7  0.21    -0.04 0.11
x5    5 100 2.67 0.77   2.60    2.63 0.59 1.1  4.6   3.5  0.48    -0.02 0.08
x6    6 100 6.97 1.59   7.15    7.01 1.85 3.7 10.0   6.3 -0.22    -0.91 0.16
x7    7 100 2.92 0.75   3.00    2.94 0.74 0.7  4.6   3.9 -0.36     0.01 0.08
```

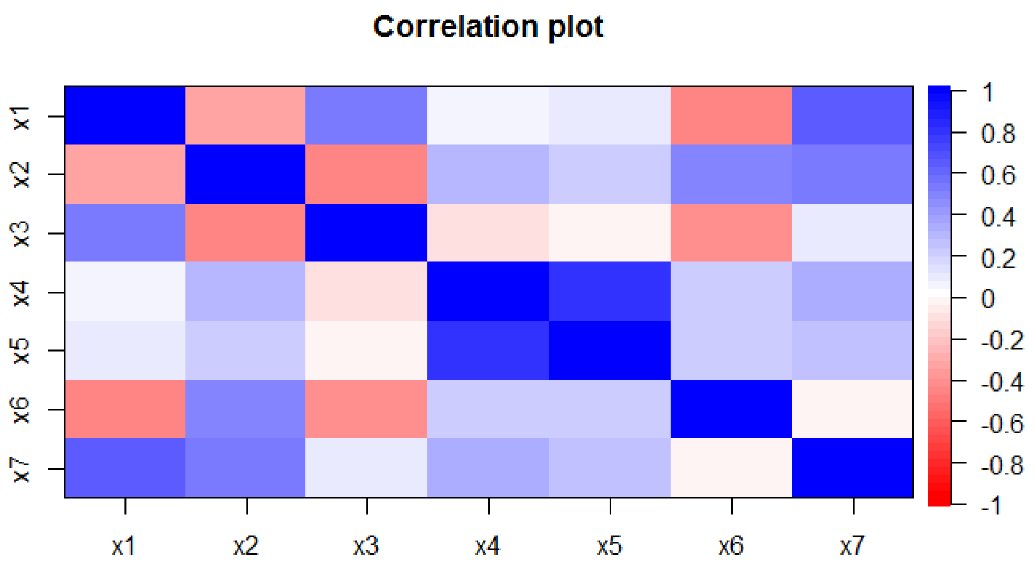
```
cor(datos1)
```

```
            x1         x2          x3         x4          x5         x6          x7
x1  1.00000000 -0.3492251  0.50929519  0.0504142  0.07742782 -0.4826309  0.61190069
x2 -0.34922515  1.0000000 -0.48721259  0.2721868  0.18533024  0.4697458  0.51298082
x3  0.50929519 -0.4872126  1.00000000 -0.1161041 -0.03479587 -0.4481120  0.06661728
x4  0.05041420  0.2721868 -0.11610408  1.0000000  0.78813516  0.1999811  0.29867737
x5  0.07742782  0.1853302 -0.03479587  0.7881352  1.00000000  0.1766130  0.24042771
x6 -0.48263094  0.4697458 -0.44811201  0.1999811  0.17661305  1.0000000 -0.05516130
x7  0.61190069  0.5129808  0.06661728  0.2986774  0.24042771 -0.0551613  1.00000000
```

corr.test(datos1)

```
Probability values (Entries above the diagonal are adjusted for multiple tests.)
     x1   x2   x3   x4   x5   x6   x7
x1 0.00 0.00 0.00 1.00 1.00 0.00 0.00
x2 0.00 0.00 0.00 0.07 0.52 0.00 0.00
x3 0.00 0.00 0.00 1.00 1.00 0.00 1.00
x4 0.62 0.01 0.25 0.00 0.00 0.41 0.03
x5 0.44 0.06 0.73 0.00 0.00 0.55 0.16
x6 0.00 0.00 0.00 0.05 0.08 0.00 1.00
x7 0.00 0.00 0.51 0.00 0.02 0.59 0.00
```

cor.plot(cor(datos1))



KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		.446
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	567.467
	gl	21
	Sig.	.000

\# Prueba de Esfericidad de Bartlett

library(rela)

cortest.bartlett(cor(datos1),n=dim(datos1))

```
$chisq
[1] 567.4674  22.6987

$p.value
[1] 1.094129e-106  3.602512e-01

$df
[1] 21
```

Matrices anti-imagen

		Rapidez de servicio	nivel de precios	flexibilidad de precios	Imagen del fabricante	Imagen de los vendedores	Calidad del producto	Servicio
Covarianza anti-imagen	Rapidez de servicio	.028	.028	.002	.015	-.006	-.002	-.025
	nivel de precios	.028	.032	.021	.014	-.005	-.020	-.026
	flexibilidad de precios	.002	.021	.608	.043	-.039	.086	-.011
	Imagen del fabricante	.015	.014	.043	.347	-.275	-.018	-.015
	Imagen de los vendedores	-.006	-.005	-.039	-.275	.371	-.044	.005
	Calidad del producto	-.002	-.020	.086	-.018	-.044	.623	.010
	Servicio	-.025	-.026	-.011	-.015	.005	.010	.023
Correlación anti-imagen	Rapidez de servicio	.345[a]	.957	.018	.148	-.059	-.016	-.978
	nivel de precios	.957	.330[a]	.155	.133	-.043	-.141	-.975
	flexibilidad de precios	.018	.155	.914[a]	.094	-.083	.139	-.091
	Imagen del fabricante	.148	.133	.094	.558[a]	-.766	-.040	-.172
	Imagen de los vendedores	-.059	-.043	-.083	-.766	.552[a]	-.091	.051
	Calidad del producto	-.016	-.141	.139	-.040	-.091	.927[a]	.088
	Servicio	-.978	-.975	-.091	-.172	.051	.088	.288[a]

a. Medida de adecuación muestral

Indicador Kaiser-Meyer-Olkinn KMO y MSA

KMO(datos1)

```
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = datos1)
Overall MSA =  0.45
MSA for each item = 
  x1   x2   x3   x4   x5   x6   x7 
0.34 0.33 0.91 0.56 0.55 0.93 0.29 
```

Matriz de Correlación Reproducida

En el análisis factorial se parte del supuesto de que las variables originales están correlacionadas entre sí. La matriz de correlación muestral refleja la correlación directa existente entre cada par de variables. El motivo de que las variables estén correlacionadas entre sí se debe a que comparten unos mismos factores comunes.

Existe otro modo de definir la correlación entre dos variables originales, derivada precisamente de esos factores comunes que comparten, y que consiste en utilizar las correlaciones entre los factores y las variables.

A nivel teórico, la correlación entre la variable X_h y X_j viene dada por la siguiente expresión:

$$\rho_{hj} = l_{h1}l_{j1} + l_{h2}l_{j2} + \ldots + l_{hm}l_{jm}$$

A esa correlación teórica le corresponde una correlación muestral, en la que los parámetros entre los coeficientes "l" son sustituidos por sus correspondientes estimaciones. A la matriz formada por dichos elementos se le denomina matriz de correlación reproducida.

Si el modelo factorial es adecuado a los datos, entonces la diferencia para cada par de variables entre el coeficiente de correlación muestral directo y el coeficiente de correlación reproducido será muy pequeña.

D) ANÁLISIS DE FACTORES CON 6 VARIABLES Y SIN ROTACIÓN USANDO SPSS Y R

Matriz de correlaciones[a]

		Rapidez de servicio	nivel de precios	flexibilidad de precios	Imagen del fabricante	Imagen de los vendedores	Calidad del producto
Correlación	Rapidez de servicio	1.000	-.349	.509	.050	.077	-.483
	nivel de precios	-.349	1.000	-.487	.272	.185	.470
	flexibilidad de precios	.509	-.487	1.000	-.116	-.035	-.448
	Imagen del fabricante	.050	.272	-.116	1.000	.788	.200
	Imagen de los vendedores	.077	.185	-.035	.788	1.000	.177
	Calidad del producto	-.483	.470	-.448	.200	.177	1.000
Sig. (Unilateral)	Rapidez de servicio		.000	.000	.309	.222	.000
	nivel de precios	.000		.000	.003	.032	.000
	flexibilidad de precios	.000	.000		.125	.366	.000
	Imagen del fabricante	.309	.003	.125		.000	.023
	Imagen de los vendedores	.222	.032	.366	.000		.039
	Calidad del producto	.000	.000	.000	.023	.039	

a. Determinante = .118

KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		.665
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	205.902
	gl	15
	Sig.	.000

Matrices anti-imagen

		Rapidez de servicio	nivel de precios	flexibilidad de precios	Imagen del fabricante	Imagen de los vendedores	Calidad del producto
Covarianza anti-imagen	Rapidez de servicio	.629	.047	-.210	-.046	-.022	.208
	nivel de precios	.047	.650	.190	-.078	.013	-.162
	flexibilidad de precios	-.210	.190	.613	.037	-.038	.092
	Imagen del fabricante	-.046	-.078	.037	.358	-.281	-.012
	Imagen de los vendedores	-.022	.013	-.038	-.281	.372	-.046
	Calidad del producto	.208	-.162	.092	-.012	-.046	.628
Correlación anti-imagen	Rapidez de servicio	.721[a]	.074	-.338	-.098	-.046	.331
	nivel de precios	.074	.787[a]	.301	-.161	.027	-.253
	flexibilidad de precios	-.338	.301	.749[a]	.079	-.079	.149
	Imagen del fabricante	-.098	-.161	.079	.542[a]	-.769	-.025
	Imagen de los vendedores	-.046	.027	-.079	-.769	.532[a]	-.096
	Calidad del producto	.331	-.253	.149	-.025	-.096	.779[a]

a. Medida de adecuación muestral

```
#################################
#                               #
# Análisis Exploratorio con 6 variables #
#                               #
#################################

# Lectura de datos con 6 variables

datos=read.delim("hatco-factorial.txt")
str(datos)

# No considerar la primera columna Id ni la última variable X7
datos2=datos[,c(-1,-8)]
str(datos2)

# Analisis descriptivo y Analisis de Correlacion
library(psych)
describe(datos2)
```

```
   vars   n mean   sd median trimmed  mad min  max range  skew kurtosis   se
x1    1 100 3.52 1.32   3.40    3.53 1.48 0.0  6.1   6.1 -0.08    -0.59 0.13
x2    2 100 2.36 1.20   2.15    2.30 1.19 0.2  5.4   5.2  0.46    -0.59 0.12
x3    3 100 7.89 1.39   8.05    7.95 1.70 5.0 10.0   5.0 -0.28    -1.12 0.14
x4    4 100 5.25 1.13   5.00    5.23 1.04 2.5  8.2   5.7  0.21    -0.04 0.11
x5    5 100 2.67 0.77   2.60    2.63 0.59 1.1  4.6   3.5  0.48    -0.02 0.08
x6    6 100 6.97 1.59   7.15    7.01 1.85 3.7 10.0   6.3 -0.22    -0.91 0.16
```

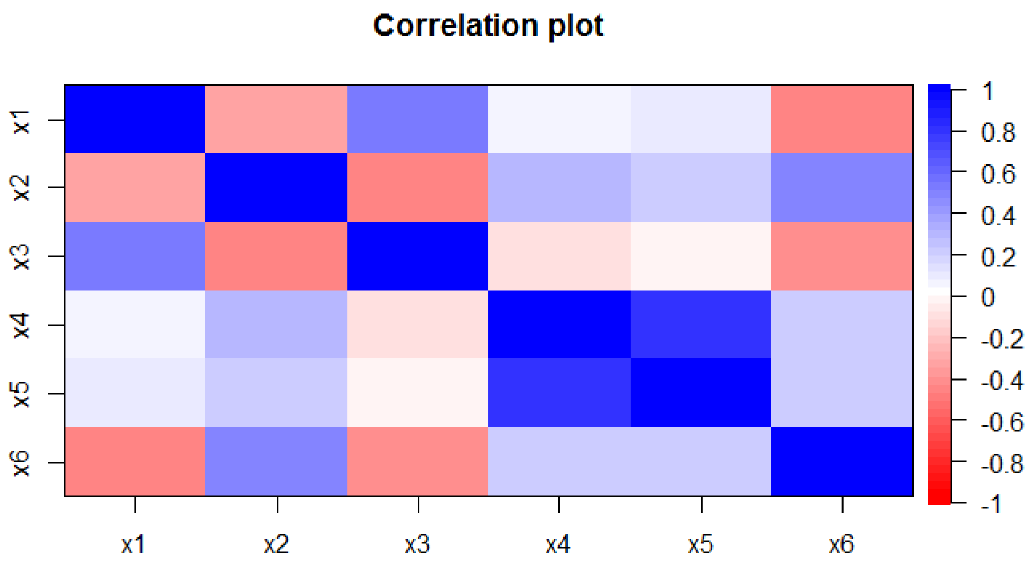
```
cor(datos2)
```

```
            x1         x2          x3         x4          x5         x6
x1  1.00000000 -0.3492251  0.50929519  0.0504142  0.07742782 -0.4826309
x2 -0.34922515  1.0000000 -0.48721259  0.2721868  0.18533024  0.4697458
x3  0.50929519 -0.4872126  1.00000000 -0.1161041 -0.03479587 -0.4481120
x4  0.05041420  0.2721868 -0.11610408  1.0000000  0.78813516  0.1999811
x5  0.07742782  0.1853302 -0.03479587  0.7881352  1.00000000  0.1766130
x6 -0.48263094  0.4697458 -0.44811201  0.1999811  0.17661305  1.0000000
```

corr.test(datos2)

```
      x1   x2   x3   x4   x5   x6
x1 0.00 0.00 0.00 1.00 1.00 0.00
x2 0.00 0.00 0.00 0.05 0.39 0.00
x3 0.00 0.00 0.00 1.00 1.00 0.00
x4 0.62 0.01 0.25 0.00 0.00 0.32
x5 0.44 0.06 0.73 0.00 0.00 0.39
x6 0.00 0.00 0.00 0.05 0.08 0.00
```

cor.plot(cor(datos2))



# Prueba de Esfericidad de Bartlett
library(rela)
cortest.bartlett(cor(datos2),n=dim(datos2))

```
$chisq
[1] 205.902077   4.639042

$p.value
[1] 1.339449e-35 9.947665e-01

$df
[1] 15
```

# Indicador Kaiser-Meyer-Olkinn KMO y MSA
KMO(datos2)

```
Call: KMO(r = datos2)
Overall MSA =  0.66
MSA for each item = 
  x1   x2   x3   x4   x5   x6 
0.72 0.79 0.75 0.54 0.53 0.78 
```

Comunalidades

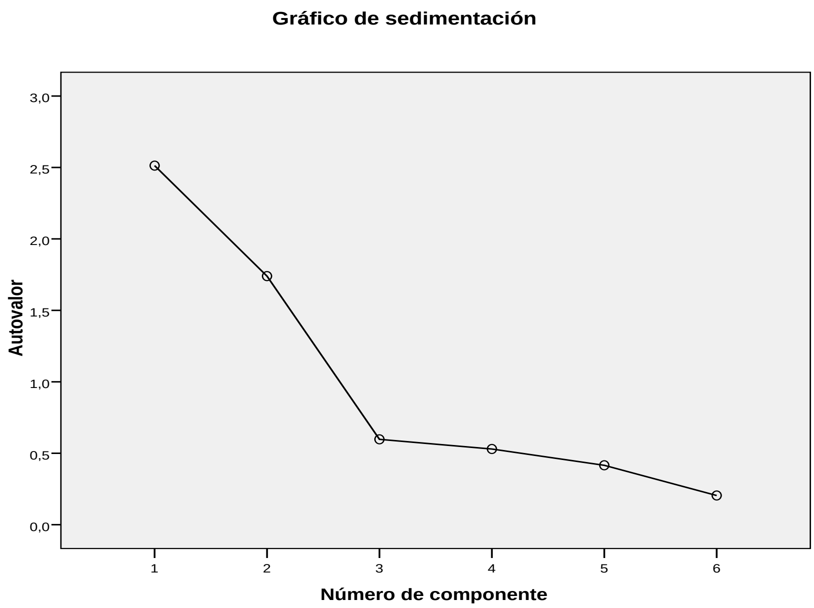
	Inicial	Extracción
Rapidez de servicio	1.000	.658
nivel de precios	1.000	.580
flexibilidad de precios	1.000	.646
Imagen del fabricante	1.000	.882
Imagen de los vendedores	1.000	.872
Calidad del producto	1.000	.616

Método de extracción: Análisis de Componentes principales.

Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	2.513	41.885	41.885	2.513	41.885	41.885
2	1.740	28.994	70.879	1.740	28.994	70.879
3	.598	9.959	80.838			
4	.530	8.830	89.667			
5	.416	6.928	96.595			
6	.204	3.405	100.000			

Método de extracción: Análisis de Componentes principales.



Matriz de componentes[a]

	Componente	
	1	2
Rapidez de servicio	-.627	.514
nivel de precios	.759	-.068
flexibilidad de precios	-.730	.336
Imagen del fabricante	.494	.799
Imagen de los vendedores	.424	.832
Calidad del producto	.767	-.167

Método de extracción: Análisis de componentes principales.

a. 2 componentes extraídos

Correlaciones reproducidas

		Rapidez de servicio	nivel de precios	flexibilidad de precios	Imagen del fabricante	Imagen de los vendedores	Calidad del producto
Correlación reproducida	Rapidez de servicio	.658[b]	-.511	.630	.101	.162	-.567
	nivel de precios	-.511	.580[b]	-.577	.320	.265	.593
	flexibilidad de precios	.630	-.577	.646[b]	-.092	-.030	-.616
	Imagen del fabricante	.101	.320	-.092	.882[b]	.874	.245
	Imagen de los vendedores	.162	.265	-.030	.874	.872[b]	.186
	Calidad del producto	-.567	.593	-.616	.245	.186	.616[b]
Residual[a]	Rapidez de servicio		.161	-.121	-.050	-.084	.084
	nivel de precios	.161		.089	-.048	-.080	-.123
	flexibilidad de precios	-.121	.089		-.024	-.004	.168
	Imagen del fabricante	-.050	-.048	-.024		-.086	-.045
	Imagen de los vendedores	-.084	-.080	-.004	-.086		-.009
	Calidad del producto	.084	-.123	.168	-.045	-.009	

Método de extracción: Análisis de Componentes principales.

a. Los residuos se calculan entre las correlaciones observadas y reproducidas. Hay 10 (66.0%) residuales no redundantes con valores absolutos mayores que 0,05.

b. Comunalidades reproducidas

Matriz de coeficientes para el cálculo de las puntuaciones en las componentes

	Componente	
	1	2
Rapidez de servicio	-.250	.296
nivel de precios	.302	-.039
flexibilidad de precios	-.290	.193
Imagen del fabricante	.197	.459
Imagen de los vendedores	.169	.478
Calidad del producto	.305	-.096

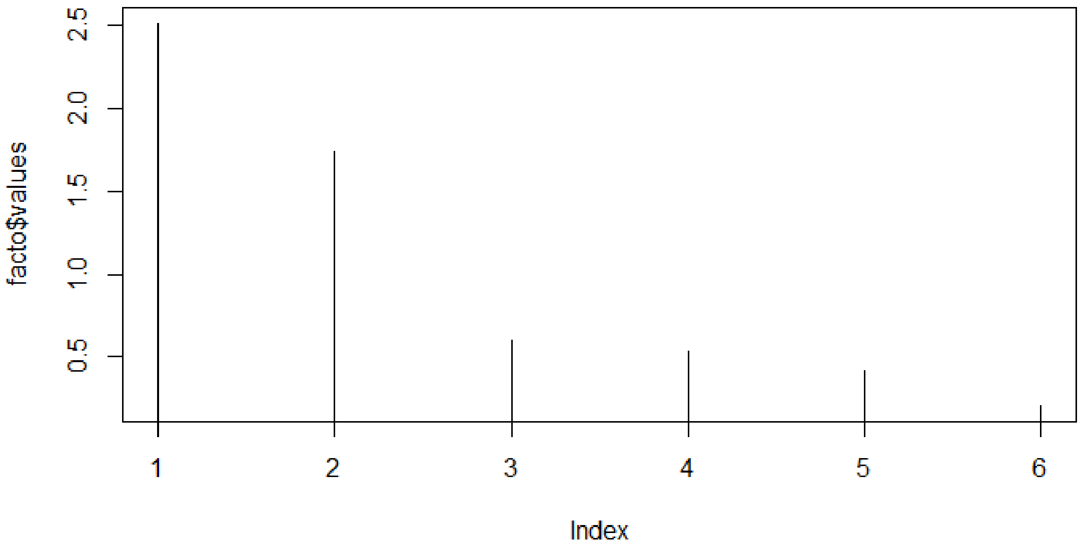
Método de extracción: Análisis de componentes principales.

Análisis Factorial sin rotación con función principal

```
library(psych)
facto=principal(r=datos2,nfactors=2,rotate="none")
str(facto)
facto$values
```

```
[1] 2.5131167 1.7396167 0.5975188 0.5297714 0.4156729 0.2043035
```

```
plot(facto$values,type="h")  # Grafica de Valores propios
```




```
facto$communality           # Comunalidades
```

```
       x1        x2        x3        x4        x5        x6 
0.6575783 0.5800456 0.6456241 0.8816649 0.8722222 0.6155984 
```

```
facto$loadings              # Cargas Factoriales, Correlaciones Factor,Componente
```

```
Loadings:
   PC1    PC2   
x1 -0.627  0.514
x2  0.759       
x3 -0.730  0.336
x4  0.494  0.799
x5  0.424  0.832
x6  0.767 -0.167

                 PC1   PC2
SS loadings    2.513 1.740
Proportion Var 0.419 0.290
Cumulative Var 0.419 0.709
```

E) ANÁLISIS DE FACTORES CON 6 VARIABLES Y CON ROTACIÓN VARIMAX USANDO SPSS Y R

Comunalidades

	Inicial	Extracción
Rapidez de servicio	1.000	.658
nivel de precios	1.000	.580
flexibilidad de precios	1.000	.646
Imagen del fabricante	1.000	.882
Imagen de los vendedores	1.000	.872
Calidad del producto	1.000	.616

Método de extracción: Análisis de Componentes principales.

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	2.513	41.885	41.885	2.513	41.885	41.885	2.370	39.499	39.499
2	1.740	28.994	70.879	1.740	28.994	70.879	1.883	31.380	70.879
3	.598	9.959	80.838						
4	.530	8.830	89.667						
5	.416	6.928	96.595						
6	.204	3.405	100.000						

Método de extracción: Análisis de Componentes principales.

Matriz de componentes[a]

	Componente	
	1	2
Rapidez de servicio	-.627	.514
nivel de precios	.759	-.068
flexibilidad de precios	-.730	.336
Imagen del fabricante	.494	.799
Imagen de los vendedores	.424	.832
Calidad del producto	.767	-.167

Método de extracción: Análisis de componentes principales.

a. 2 componentes extraídos

Matriz de componentes rotados[a]

	Componente	
	1	2
Rapidez de servicio	-.787	.194
nivel de precios	.714	.265
flexibilidad de precios	-.803	-.011
Imagen del fabricante	.102	.933
Imagen de los vendedores	.025	.934
Calidad del producto	.764	.179

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización Varimax con Kaiser.

a. La rotación ha convergido en 3 iteraciones.

Correlaciones reproducidas

		Rapidez de servicio	nivel de precios	flexibilidad de precios	Imagen del fabricante	Imagen de los vendedores	Calidad del producto
Correlación reproducida	Rapidez de servicio	.658[b]	-.511	.630	.101	.162	-.567
	nivel de precios	-.511	.580[b]	-.577	.320	.265	.593
	flexibilidad de precios	.630	-.577	.646[b]	-.092	-.030	-.616
	Imagen del fabricante	.101	.320	-.092	.882[b]	.874	.245
	Imagen de los vendedores	.162	.265	-.030	.874	.872[b]	.186
	Calidad del producto	-.567	.593	-.616	.245	.186	.616[b]
Residual[a]	Rapidez de servicio		.161	-.121	-.050	-.084	.084
	nivel de precios	.161		.089	-.048	-.080	-.123
	flexibilidad de precios	-.121	.089		-.024	-.004	.168
	Imagen del fabricante	-.050	-.048	-.024		-.086	-.045
	Imagen de los vendedores	-.084	-.080	-.004	-.086		-.009
	Calidad del producto	.084	-.123	.168	-.045	-.009	

Método de extracción: Análisis de Componentes principales.

a. Los residuos se calculan entre las correlaciones observadas y reproducidas. Hay 10 (66.0%) residuales no redundantes con valores absolutos mayores que 0,05.

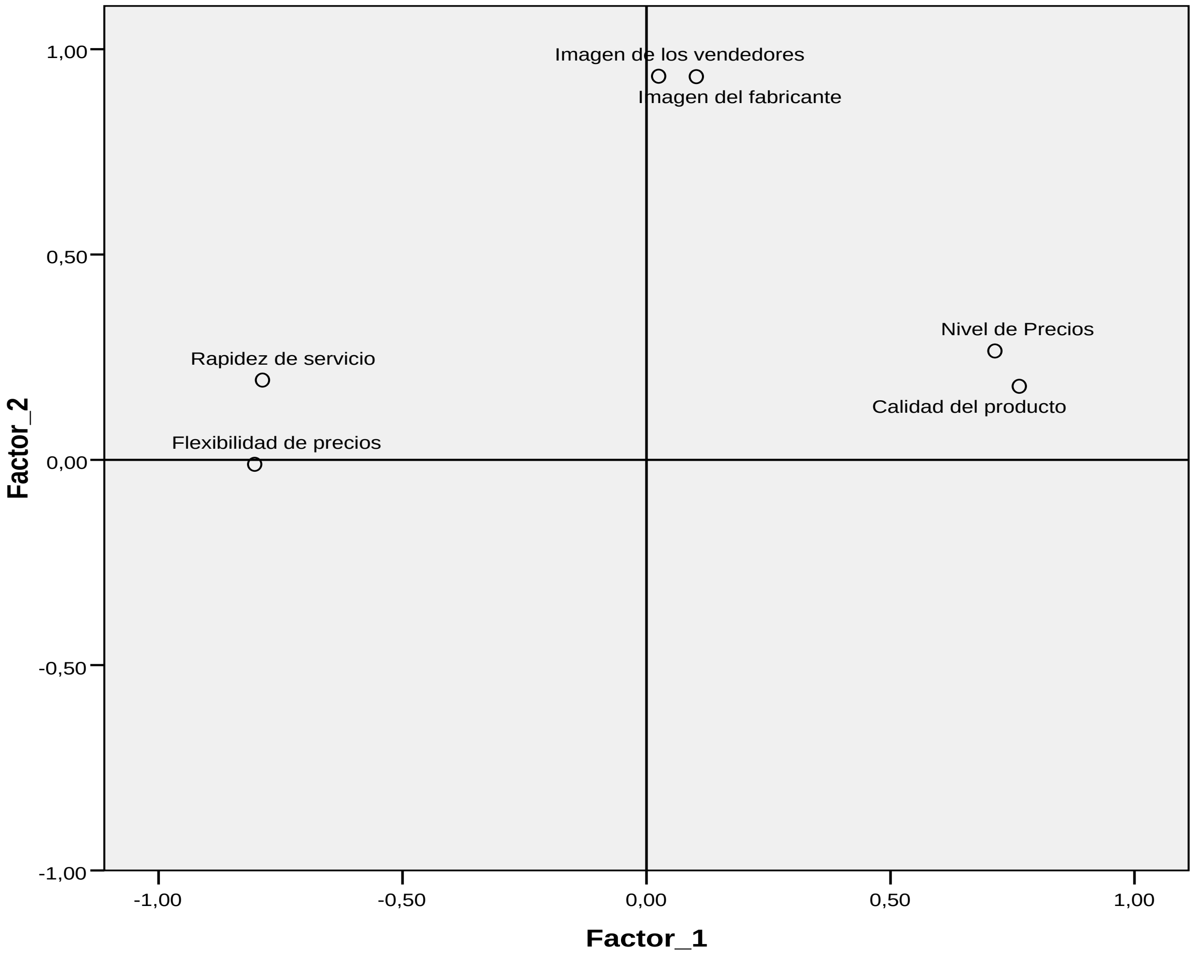
b. Comunalidades reproducidas

Matriz de coeficientes para el cálculo de las puntuaciones en las componentes

	Componente	
	1	2
Rapidez de servicio	-.352	.159
nivel de precios	.289	.095
flexibilidad de precios	-.345	.049
Imagen del fabricante	-.020	.499
Imagen de los vendedores	-.053	.504
Calidad del producto	.317	.044

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización Varimax con Kaiser.



# Análisis Factorial con rotacion con función principal
library(psych)
facto=principal(r=datos2,nfactors=2,rotate="varimax")
str(facto)
facto$values

```
[1] 2.5131167 1.7396167 0.5975188 0.5297714 0.4156729 0.2043035
```

facto$communality

```
       x1        x2        x3        x4        x5        x6 
0.6575783 0.5800456 0.6456241 0.8816649 0.8722222 0.6155984
```

facto$loadings

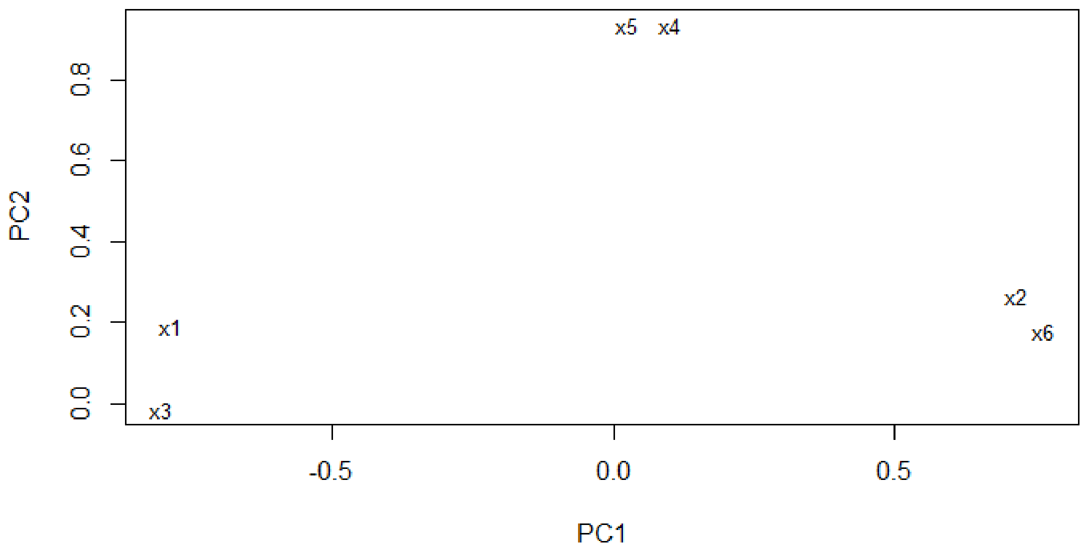
```
Loadings:
   PC1    PC2  
x1 -0.788  0.193
x2  0.714  0.266
x3 -0.803       
x4  0.101  0.933
x5         0.934
x6  0.764  0.179

                 PC1   PC2
SS loadings    2.369 1.883
Proportion Var 0.395 0.314
```

```
Cumulative Var 0.395 0.709
```

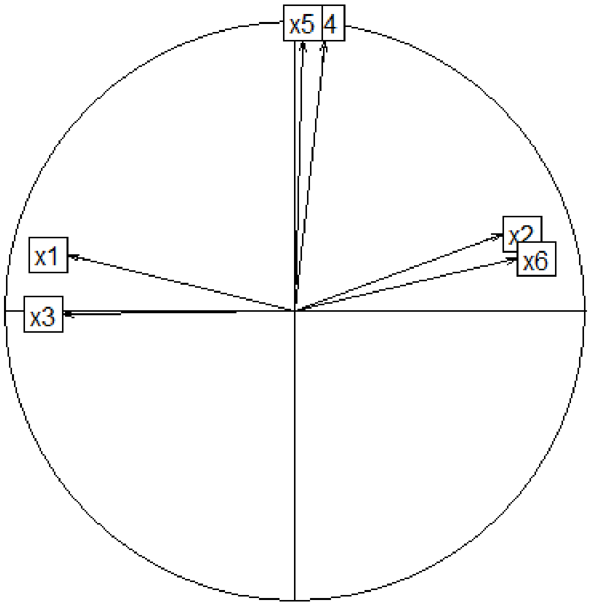
facto$scores

```
                PC1          PC2
  [1,] -0.657020717 -0.602644207
  [2,]  1.176158938  1.297526812
  [3,]  1.492440611  0.522742786
            .              .
            .              .
            .              .
[99,]  0.609275093  0.773018106
[100,] -0.353753147 -0.045820886
```

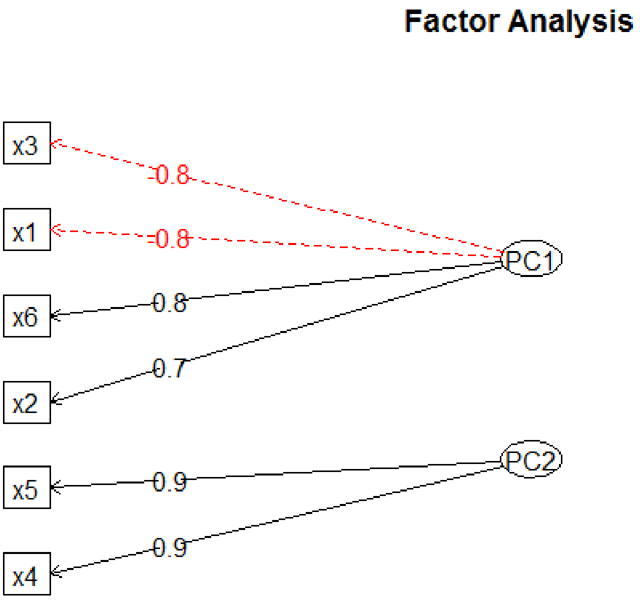
```
# Grafica de individuos sobre el primer plano de componentes
load <- facto$loadings[,1:2]
plot(load,type="n") # set up plot
text(load,labels=names(datos2),cex=0.8) # add variable names
```




```
# Grafica de circulo de correlaciones
library(ade4)
s.corcircle(load,grid=FALSE)
```



fa.diagram(facto)




```
# Almacena los datos y los resultados de los scores en un archivo CSV
datosf=cbind(datos2,facto$scores)
datosf
str(datosf)
write.csv(datosf,"hatco-datos-factorial-con-acp.csv")
```

Ejemplo de Aplicación Nº 3[3]

Una empresa especializada en el diseño de automóviles de turismo desea calcular cuáles son los deseos del público que compra automóviles.

Diseña una encuesta con 10 preguntas donde se le pide a cada uno de los 20 encuestados que valores de 1 a 5 si una característica es o no muy importante. Los encuestados deberán contestar con 5 si la característica es muy importante, un 4 si es importante, un 3 si tiene regular importancia, un 2 si es poco importante y un 1 si no es nada importante.

Variable	Descripción
V1	Precio
V2	Financiación
V3	Consumo
V4	Combustible
V5	Seguridad
V6	Confort
V7	Capacidad
V8	Prestaciones
V9	Modernidad
V10	Aerodinámica

Id	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
1	4	1	4	3	3	2	4	4	4	4
2	5	5	4	4	3	3	4	1	1	3
3	2	1	3	1	4	2	1	5	4	5
4	1	1	1	1	4	4	2	5	5	4
5	1	1	2	1	5	5	4	3	3	2
6	5	5	5	5	3	3	4	2	2	1
7	4	5	4	4	2	2	5	1	1	1
8	3	2	3	1	4	4	2	5	5	5
9	4	4	4	3	4	4	3	1	1	1
10	5	5	5	5	2	2	3	2	2	2
11	2	2	2	1	5	4	4	3	4	3
12	4	4	5	5	4	5	5	2	1	2
13	3	2	2	1	4	5	4	4	3	3
14	5	5	4	4	5	4	4	1	2	2
15	4	3	3	1	4	4	5	3	4	4
16	5	5	4	4	4	5	4	2	1	1
17	4	4	5	2	4	5	5	4	4	2
18	5	5	4	4	2	2	1	2	2	3
19	3	3	2	2	4	4	5	4	5	4
20	5	5	4	4	4	5	4	3	2	1

[3] Basado en Pérez, César. "Técnicas de Análisis Multivariante de Datos. Aplicaciones con SPSS". Pearson Prentice Hall. 2004. España.

Estadísticos descriptivos

	Media	Desviación típica	N del análisis
Precio	3.70	1.342	20
Financiacion	3.40	1.635	20
Consumo	3.50	1.192	20
Combustible	2.80	1.576	20
Seguridad	3.70	.923	20
Confort	3.70	1.174	20
Capacidad	3.65	1.268	20
Prestaciones	2.85	1.387	20
Modernidad	2.80	1.473	20
Aerodinamica	2.65	1.348	20

Matriz de correlaciones

		Precio	Financiacion	Consumo	Combustible	Seguridad	Confort	Capacidad	Prestaciones	Modernidad	Aerodinamica
Correlación	Precio	1.000	.873	.823	.816	-.501	-.194	.213	-.648	-.645	-.497
	Financiacion	.873	1.000	.729	.829	-.439	-.071	.249	-.784	-.752	-.697
	Consumo	.823	.729	1.000	.812	-.478	-.226	.192	-.557	-.630	-.540
	Combustible	.816	.829	.812	1.000	-.550	-.262	.174	-.737	-.789	-.654
	Seguridad	-.501	-.439	-.478	-.550	1.000	.738	.175	.292	.341	.123
	Confort	-.194	-.071	-.226	-.262	.738	1.000	.421	.132	.055	-.236
	Capacidad	.213	.249	.192	.174	.175	.421	1.000	-.301	-.180	-.414
	Prestaciones	-.648	-.784	-.557	-.737	.292	.132	-.301	1.000	.886	.730
	Modernidad	-.645	-.752	-.630	-.789	.341	.055	-.180	.886	1.000	.785
	Aerodinamica	-.497	-.697	-.540	-.654	.123	-.236	-.414	.730	.785	1.000
Sig. (Unilateral)	Precio		.000	.000	.000	.012	.207	.183	.001	.001	.013
	Financiacion	.000		.000	.000	.026	.383	.145	.000	.000	.000
	Consumo	.000	.000		.000	.016	.169	.209	.005	.001	.007
	Combustible	.000	.000	.000		.006	.133	.232	.000	.000	.001
	Seguridad	.012	.026	.016	.006		.000	.230	.106	.071	.303
	Confort	.207	.383	.169	.133	.000		.032	.289	.409	.158
	Capacidad	.183	.145	.209	.232	.230	.032		.099	.223	.035
	Prestaciones	.001	.000	.005	.000	.106	.289	.099		.000	.000
	Modernidad	.001	.000	.001	.000	.071	.409	.223	.000		.000
	Aerodinamica	.013	.000	.007	.001	.303	.158	.035	.000	.000	

KMO y prueba de Bartlett

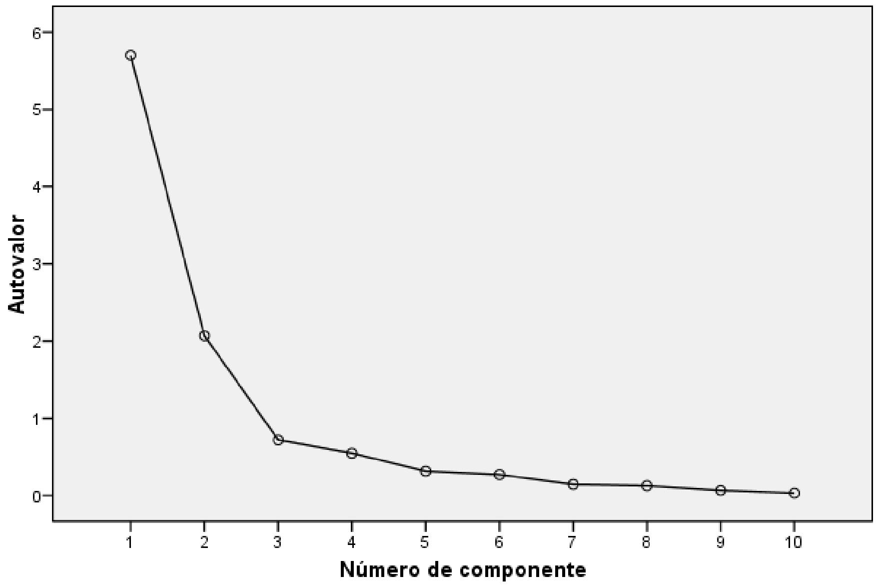
Medida de adecuación muestral de Kaiser-Meyer-Olkin.		.700
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	163.466
	gl	45
	Sig.	.000

Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	5.701	57.011	57.011	5.701	57.011	57.011
2	2.069	20.692	77.703	2.069	20.692	77.703
3	.720	7.205	84.908			
4	.548	5.478	90.386			
5	.316	3.158	93.544			
6	.271	2.707	96.251			
7	.146	1.464	97.715			
8	.128	1.280	98.995			
9	.068	.684	99.679			
10	.032	.321	100.000			

Método de extracción: Análisis de Componentes principales.

Gráfico de sedimentación



Comunalidades

	Inicial	Extracción
Precio	1.000	.781
Financiacion	1.000	.857
Consumo	1.000	.720
Combustible	1.000	.883
Seguridad	1.000	.803
Confort	1.000	.849
Capacidad	1.000	.528
Prestaciones	1.000	.767
Modernidad	1.000	.792
Aerodinamica	1.000	.791

Método de extracción: Análisis de Componentes principales.

Matriz de componentes[a]

	Componente	
	1	2
Precio	.878	-.099
Financiacion	.923	.064
Consumo	.841	-.116
Combustible	.933	-.109
Seguridad	-.532	.721
Confort	-.197	.900
Capacidad	.276	.672
Prestaciones	-.862	-.156
Modernidad	-.879	-.137
Aerodinamica	-.765	-.454

Método de extracción: Análisis de componentes principales.

a. 2 componentes extraídos

Usando rotación

Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	5.701	57.011	57.011	5.701	57.011	57.011	5.669	56.685	56.685
2	2.069	20.692	77.703	2.069	20.692	77.703	2.102	21.018	77.703
3	.720	7.205	84.908						
4	.548	5.478	90.386						
5	.316	3.158	93.544						
6	.271	2.707	96.251						
7	.146	1.464	97.715						
8	.128	1.280	98.995						
9	.068	.684	99.679						
10	.032	.321	100.000						

Método de extracción: Análisis de Componentes principales.

Matriz de componentes rotados[a]

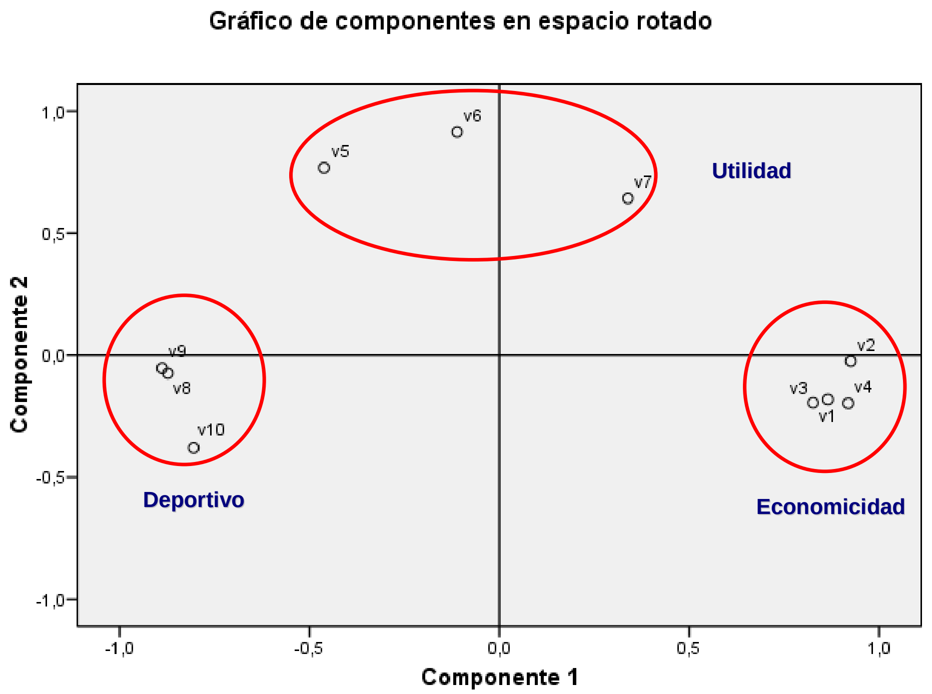
	Componente	
	1	2
Precio	.865	-.181
Financiacion	.925	-.024
Consumo	.826	-.195
Combustible	.919	-.197
Seguridad	-.462	.768
Confort	-.111	.915
Capacidad	.339	.643
Prestaciones	-.873	-.074
Modernidad	-.888	-.053
Aerodinamica	-.805	-.379

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización Varimax con Kaiser.

a. La rotación ha convergido en 3 iteraciones.

Interpretación de los Factores



44

Factor 1: Economicidad-Deportivo

Se trata de un factor que une la economicidad (correlaciones altas positivas con precio, financiación, consumo, combustible) con el escaso interés en que el auto tenga aire deportivo (correlaciones altas negativas con prestaciones, modernidad, aerodinámica)

Este factor explica el 57% de la variancia.

Factor 2: Utilidad

Se trata de un factor que mide la utilidad del auto (correlaciones altas positivas con seguridad, confort, capacidad)

Este factor explica el 20.7% de la variancia.

Ejemplo de Aplicación Nº 4[4]

El estudio de mercado busca establecer la imagen que proyecta en los consumidores un supermercado determinado.

Para esto se les presentaron las siguientes variables:

- Lejanía / Cercanía al hogar (Hogar)
- Altos / Bajos de precios(Precios)
- Revistas / No Revistas (Revistas): Se refiere a si frecuentemente publica catálogos o revistas de sus productos.
- Alta / Baja calidad (Calidad)
- Casero / No Casero (Casero): Se refiere al grado de cercanía y amabilidad por parte del personal del supermercado.
- Cupones / No cupones (Cupones): Se refiere a la entrega de cupones para descuentos y/o sorteos.
- Publicidad / No publicidad (Publicidad).

Estas fueron evaluadas en una escala de Diferencial Semántico de 1 a 7.

Estadísticos descriptivos

	Media	Desviación típica	N del análisis
hogar	4.3200	1.86458	25
precios	4.0400	1.85921	25
revistas	3.9600	1.79072	25
calidad	3.9200	1.80093	25
casero	3.8400	1.92959	25
cupones	4.1600	1.84120	25
publicid	4.2000	1.89297	25

Matriz de correlaciones

		hogar	precios	revistas	calidad	casero	cupones	publicid
Correlación	hogar	1.000	-.004	.628	.082	.675	-.100	-.338
	precios	-.004	1.000	.151	-.248	.048	.582	-.251
	revistas	.628	.151	1.000	-.182	.480	.090	-.588
	calidad	.082	-.248	-.182	1.000	.272	.017	.469
	casero	.675	.048	.480	.272	1.000	-.110	-.082
	cupones	-.100	.582	.090	.017	-.110	1.000	.014
	publicid	-.338	-.251	-.588	.469	-.082	.014	1.000
Sig. (Unilateral)	hogar		.493	.000	.348	.000	.316	.049
	precios	.493		.236	.116	.409	.001	.113
	revistas	.000	.236		.192	.008	.334	.001
	calidad	.348	.116	.192		.094	.469	.009
	casero	.000	.409	.008	.094		.301	.348
	cupones	.316	.001	.334	.469	.301		.473
	publicid	.049	.113	.001	.009	.348	.473	

---

4 Basado en Galaz, I. & Otros. "Análisis multivariado enfocado al marketing. Guía práctica para posicionamiento, imagen y segmentación utilizando herramientas de Excel". Chile. 2003.

KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		.550
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	57.994
	gl	21
	Sig.	.000

Comunalidades

	Inicial	Extracción
hogar	1.000	.818
precios	1.000	.796
revistas	1.000	.790
calidad	1.000	.800
casero	1.000	.805
cupones	1.000	.841
publicid	1.000	.796

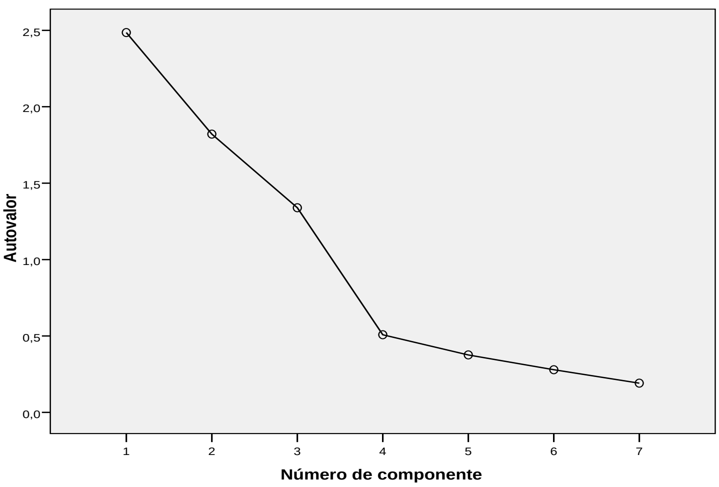
Método de extracción: Análisis de Componentes principales.

Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	2.485	35.505	35.505	2.485	35.505	35.505	2.315	33.076	33.076
2	1.821	26.013	61.518	1.821	26.013	61.518	1.731	24.729	57.805
3	1.339	19.131	80.649	1.339	19.131	80.649	1.599	22.844	80.649
4	.508	7.258	87.907						
5	.376	5.373	93.280						
6	.279	3.990	97.270						
7	.191	2.730	100.000						

Método de extracción: Análisis de Componentes principales.

Gráfico de sedimentación



Matriz de componentes(a)

	Componente		
	1	2	3
hogar	.817	.378	.087
precios	.279	-.714	.457
revistas	.887	-.027	-.043
calidad	-.204	.634	.597
casero	.664	.505	.329
cupones	.050	-.604	.689
publicid	-.684	.383	.426

Método de extracción: Análisis de componentes principales.
a 3 componentes extraídos

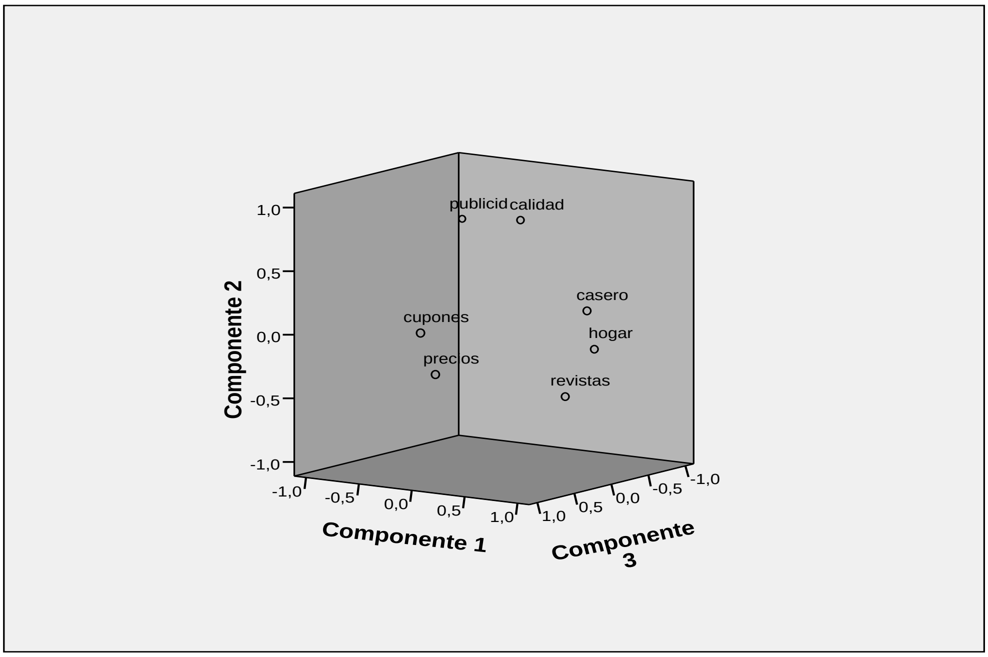
Interpretación de los Factores

Matriz de componentes rotados(a)

	Componente		
	1	2	3
hogar	.897	-.082	-.076
precios	.049	-.232	.860
revistas	.762	-.440	.125
calidad	.214	.867	-.052
casero	.868	.224	-.017
cupones	-.057	.091	.911
publicid	-.351	.817	-.073

Método de extracción: Análisis de componentes principales.
Método de rotación: Normalización Varimax con Kaiser.
a La rotación ha convergido en 4 iteraciones.

Gráfico de componentes en espacio rotado



Factor 1: Atención al cliente. Se compone de las variables hogar, revistas y casero.

Factor 2: Competitividad. Se compone de las variables calidad y publicidad.

Factor 3: Economía. Se compone de las variables precios y cupones.

IV. Limitaciones y Supuestos del Análisis Factorial[5]

El análisis factorial es una técnica de reducción de datos que depende especialmente de las correlaciones (o de las covariancias) y sus limitaciones y supuestos están, por tanto, ligados a ellas.

Tamaño de la muestra y datos perdidos

Los coeficientes de correlación son muy sensibles al tamaño de las muestras. Son de hecho, poco fiables cuando las muestras son pequeñas, lo que hace recomendable que el número de sujetos sea lo suficientemente grande para que el análisis produzca resultados estables.

La pérdida de sujetos también puede afectar a los coeficientes de correlación, particularmente cuando esta se produce de manera no aleatoria.

Normalidad

Es un método fundamental cuando el método de extracción es de máxima verosimilitud. En los demás casos su no cumplimiento puede ser importante sólo en lo relativo a la determinación del número de factores, particularmente si se emplean criterios estadísticos.

Linealidad

Los coeficientes de correlación miden la relación lineal entre variables y el supuesto de normalidad multivariada lo exige. Si las correlaciones son no lineales, el análisis no producirá resultados adecuados.

Ausencia de valores extremos

Los coeficientes de correlación son medidas no robustas de relación lineal, lo que implica que son afectados por la existencia de puntuaciones extremas. Su existencia compromete la solución factorial, por lo que es importante comprobar si existen, y eliminarlos o realizar transformaciones que limiten su influencia.

Factorizabilidad de la matriz de correlaciones

El análisis factorial produce resultados adecuados cuando la matriz de correlaciones contienen correlaciones altas entre pares de variables. En general, si ninguna correlación excede de 0.30 es bastante probable que no pueda llegarse a una solución aceptable.

5 Adaptado del libro “Análisis Multivariado. Un manual para investigadores” de Andrés Catena & Otros. Editorial Biblioteca Nueva. Madrid. 2003.