ANALISIS DE COMPONENTES PRINCIPALES

I. Introducción[1]

Se presenta a continuación la matriz de correlación para 9 elementos (variables) de imagen de un establecimiento.

	V1	V2	V3	V4	V5	V6	V7	V8	V9
V1. Nivel de precios	1.000								
V2. Personal de establecimiento	0.427	1.000							
V3. Política de retorno	0.302	0.771	1.000						
V4. Disponibilidad del producto	0.470	0.497	0.427	1.000					
V5. Calidad de producto	0.765	0.406	0.307	0.472	1.000				
V6. Profundidad de surtido	0.281	0.445	0.423	0.713	0.325	1.000			
V7. Anchura de surtido	0.354	0.490	0.471	0.719	0.378	0.724	1.000		
V8. Servicio dentro del establecimiento	0.242	0.719	0.733	0.428	0.240	0.311	0.435	1.000	
V9. Ambiente dentro del establecimiento	0.372	0.737	0.774	0.479	0.326	0.429	0.466	0.710	1.00

¿Están todos estos elementos separados en sus propiedades de valoración o están agrupados en áreas más generales de valoración?

1. Basado en Hair, Joseph; Anderson, Rolph; Thatam, Ronald & Black, William. “Análisis Multivariante”. Editorial Prentice Hall. 1999. España.

1.1 Objetivos del ACP

Según Diaz & Morales[2] el análisis de componentes principales tiene como objetivos, entre otros, los siguientes:

- Generar nuevas variables que expresen la información contenida en un conjunto de datos.
- Reducir la dimensión del espacio donde están inscritos los datos.
- Eliminar las variables (si es posible) que aporten poco al estudio del problema.
- Facilitar la interpretación de la información contenida en los datos.

El ACP tiene como propósito central la determinación de unos pocos factores (componentes principales) que retengan la mayor variabilidad contenida en los datos. Las nuevas variables poseen algunas características estadísticas "deseables", tales como independencia (bajo el supuesto de normalidad) y no correlación.

En el caso de no correlación entre las variables originales, el ACP no tiene mucho que hacer, pues las componentes se corresponderían con cada variable por orden de magnitud en la varianza; es decir, la primera componente coincide con la variable de mayor varianza, la segunda componente con la variable de segunda mayor varianza, y así sucesivamente.

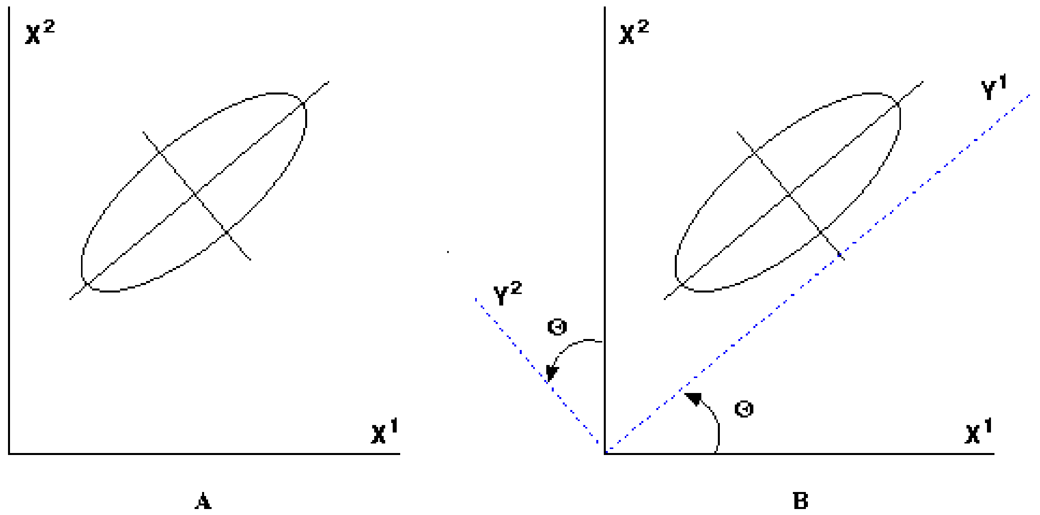
2. Basado en Díaz Monroy, Guillermo & Morales Rivera, Mario. "Análisis Estadístico de Datos Multivariados". Universidad Nacional de Colombia, Sede Bogotá, Facultad de Ciencias. 2012. Colombia.

II. Enfoque Matemático[3]

2.1 Introducción

Supongamos que representamos un conjunto de patrones bidimensionales que presentan cierto grado de correlación (Figura 1.A). En esta representación utilizamos por simplicidad una elipse en lugar de la nube de puntos. Si representamos estos datos en un nuevo espacio generado por las variables Y_1 e Y_2 (Figura 1.B), que se corresponden con los ejes de la elipse y consideramos únicamente Y_1, la proyección de los datos sobre este eje hace que su dispersión sea mayor que sobre cualquier otro eje (y en particular sobre cualquiera de los originales).

Figure 1: A) X_1 y X_2 están correlacionadas. B) Y_1 e Y_2 están no correlacionadas.



Si los datos están correlacionados en *P la dirección de máxima varianza en P es la del eje principal de la elipse que los caracteriza*. Este nuevo eje, Y_1, se calcula como una rotación de magnitud θ del primer eje de P, X_1. Si el nuevo sistema de coordenadas es ortogonal, el segundo eje, Y_2, se establece en base al segundo eje del sistema original, X_2, mediante una rotación de la misma magnitud, θ, que la aplicada al primero. En definitiva, la relación entre los nuevos ejes y los antiguos consiste en una relación lineal:

$$\begin{aligned} Y_1 &= W_{11}X_1 + W_{12}X_2 \\ Y_2 &= W_{21}X_1 + W_{22}X_2 \end{aligned} \tag{1}$$

que asume únicamente una rotación, manteniendo el origen de coordenadas común para los sistemas de coordenadas (X_1, X_2) e (Y_1, Y_2). Con esta transformación, la longitud del eje mayor Y_1 indicará el rango de los datos a lo largo de este nuevo eje, que será necesariamente mayor que sobre cualquiera de los ejes originales, X_1 ó X_2. En consecuencia, dependiendo de la relación entre los ejes originales, la mayor parte de la información contenida en el espacio *P* puede retenerse únicamente en el primer eje principal, Y_1, lo que implica una selección de características en el espacio transformado *P'*.

3 Basado en el material del curso "Reconocimiento de Formas" de Francisco José Cortijo Bon. Departamento de Ciencias de la Computación e Inteligencia Artificial. Escuela Técnica Superior de Ingeniería Informática. Universidad de Granada

El objetivo del análisis de componentes principales es transformar el espacio de representación *P* en un nuevo espacio, *P'*, en el que los datos estén **no correlacionados** (la matriz de covariancia en ese espacio será *diagonal*). Con otras palabras, *se trata de encontrar un nuevo conjunto de ejes ortogonales en el que la varianza de los datos sea máxima*. El objetivo final es reducir la dimensionalidad del problema una vez realizada la transformación.

Si se trabaja con datos de dimensionalidad mayor que 2 el procedimiento es similar: los nuevos ejes se definen secuencialmente de manera que un nuevo eje se define como aquel que es perpendicular al seleccionado anteriormente y su dirección es la de la máxima varianza de entre todos los ejes posibles.

La ecuación (1) puede reescribirse en términos matriciales como:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \tag{2}$$

y el objetivo será encontrar los coeficientes de transformación W_{ij} con las restricciones:

1. Los ejes que definen *P'* son ortogonales.
2. Los datos en *P'* están no correlacionados.

2.2 Transformación de componentes principales en espacios multidimensionales

En lo sucesivo, consideraremos espacios multidimensionales (p>2) para generalizar la solución al problema, que se plantea ahora como la búsqueda de una transformación lineal, *W*, de los datos originales en *P*, *X*, que da lugar a nuevas coordenadas *Y* en *P'* tal que

$$Y = W^T X \tag{3}$$

La ecuación (3) se puede escribir de manera explícita como

$$\begin{bmatrix} Y_1 \\ \cdot \\ \cdot \\ \cdot \\ Y_p \end{bmatrix} = \begin{bmatrix} W_{11} & \cdot & \cdot & \cdot & W_{1p} \\ \cdot & \cdot & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot \\ W_{p1} & \cdot & \cdot & \cdot & W_{pp} \end{bmatrix} \begin{bmatrix} X_1 \\ \cdot \\ \cdot \\ \cdot \\ X_p \end{bmatrix} \tag{4}$$

de forma que su desarrollo es:

$$
\begin{aligned}
Y_1 &= W_{11}X_1 + W_{12}X_2 + \dots + W_{1p}X_p \\
Y_2 &= W_{21}X_1 + W_{22}X_2 + \dots + W_{2p}X_p \\
&. \\
&. \\
&. \\
Y_p &= W_{p1}X_1 + W_{p2}X_2 + \dots + W_{pp}X_p
\end{aligned} \qquad (5)
$$

de donde se deduce que (4) una generalización de (2) y (5) es una generalización de (1). Para el cálculo de W se mantienen las restricciones:

1. Los ejes que definen P' son ortogonales.
2. Los datos en P' están no correlacionados.

Antes de abordar la metodología para la solución a este problema formalizaremos el significado de estas restricciones.

2.3 Sobre la ortogonalidad de los ejes

La transformación consiste, básicamente, en una rotación *rígida* o *solidaria* de los ejes de P tomando como referencia el origen de coordenadas. La consecuencia es que si los ejes de P' deben ser ortogonales, la distancia euclidiana entre el origen y los puntos se mantiene inalterada con esta transformación. Para que esto sea cierto, la matriz de transformación W debe ser *ortogonal*, esto es, que $W^{-1} = W^T$, por lo que

$$W^TW = WW^T = I \qquad (6)$$

De lo que se deduce que $W^TW - I = 0$

En definitiva, buscamos una matriz cuadrada pxp que sea ortogonal.

2.4 Sobre la no correlación en P'

Si las nuevas variables Y se calculan a partir de X mediante la transformación dada por la ecuación (3) puede demostrarse que

$$\mu_Y = E\{Y\} = E\{W^TX\} = W^TE\{X\} = W^T\mu_X \qquad (7)$$

$$\sum_Y = E\{(Y - \mu_Y)(Y-\mu_Y)^T\} = W^T\sum_X W \qquad (8)$$

Si los datos en P' deben estar no correlacionados, la matriz de covariancia en P', $\sum_Y$, debe ser *diagonal*.

2.5 Obtención de la matriz de transformación *W*

El problema puede formularse como un problema de maximización (varianza en *P'*) con restricciones (ortogonalidad de *W*). La técnica adecuada es la utilización de los *multiplicadores de Lagrange*, que puede plantearse como sigue.

Si el objetivo es maximizar una función $f\ (v_1, v_2, \ldots, v_p)$ con la condición $g(v_1, v_2, \ldots, v_p) = 0$ se puede construir una nueva función

$$F = f\ (v_1, v_2, \ldots, v_p) - \lambda\, g(v_1, v_2, \ldots, v_p)$$

y maximizar esta función sin restricciones.

En nuestro caso, se trata de maximizar la varianza en *P'*, por lo que $f = \sum_Y = W^T \sum_X W$ y la restricción es que $W^TW = I$, por lo que $g = W^TW - I = 0$. En definitiva, se trata de maximizar

$$F = W^T \sum\nolimits_X W - \lambda\ (W^TW - I) \tag{9}$$

y derivando respecto a *W*,

$$(\sum\nolimits_X - \lambda\, I)W = 0 \tag{10}$$

y se tratará de encontrar la solución al sistema de ecuaciones dado por (10). En definitiva, *W debe verificar que* $(\sum_X - \lambda\ I)W = 0$ *con objeto de que* $W^T \sum_X W = \sum_Y$ *sea máxima, sujeta a la restricción de que* $W^TW = I$.

Para que la ecuación (10) sea cierta solo pueden ocurrir dos casos:

1. Que $W = 0$ y en este caso la solución es *trivial* y no interesa.
2. Que $(\sum_X - \lambda\ I)$ sea singular (no invertible), esto es, que

$$| \sum\nolimits_X - \lambda\ | = 0. \tag{11}$$

La ecuación (11) es la ***ecuación característica*** de la matriz $\sum_X$ y su expresión es una ecuación polinómica de λ. Las soluciones a esta ecuación (los valores de λ) se conocen como los ***autovalores*** de

$\sum_X$. Cuando se sustituyen en (10), se calculan los vectores asociados a cada valor de λ, que se conocen como los ***autovectores*** de $\sum_X$.

En otras palabras, cada autovalor, λ_i, es solución a una ecuación del sistema $(\sum_X - \lambda\, I)W = 0$. Así, para cada ecuación, los parámetros de *W* asociados a la solución con λ_i es un autovector, ϕ_i. De esta manera, podemos expresar la matriz de transformación como un vector de p vectores columna (*W* es un vector de autovectores):

$$W = [\phi_i, \phi_2, \ldots, \phi_p] \tag{12}$$

Como $\sum_X$ es de orden *p x p*, tendrá *p* autovalores asociados, $\lambda_1, \lambda_2, \ldots, \lambda_p$ y como $\sum_X$ es simétrica, todos los autovalores serán reales.

2.6 Formalización del ACP

Como $\sum_X$ es simétrica, todos sus autovalores serán reales. Por otra parte, dado que $\sum_X$ es definida positiva, sus autovalores están ordenados:

$$\lambda_1 \geq \lambda_2 \geq \ldots, \geq \lambda_p \tag{13}$$

La matriz de covariancia $\sum_Y = W^T \sum_X W$ será una matriz *diagonal* formada por los *autovalores* de $\sum_X$:

$$\sum_Y = \begin{bmatrix} \lambda_1 & 0 & . & . & 0 \\ 0 & \lambda_2 & & & 0 \\ . & & . & & . \\ . & & & . & . \\ 0 & 0 & . & . & \lambda_p \end{bmatrix} \tag{14}$$

y los valores de la diagonal (*autovalores de* $\sum_X$) son las varianzas de los patrones en las respectivas coordenadas transformadas.

La matriz que contiene los coeficientes de la transformación, *W*, es la matriz de *autovectores de* $\sum_X$, asumiendo que *W* es *ortogonal*.

Cada autovalor λ_i tiene asociado un autovector ϕ_i y cada autovector define la dirección de un eje en el espacio transformado, *P'*. Dado que los autovalores están ordenados (por el valor de varianza en cada eje de *P'*), y que cada autovalor tiene asociado un autovector, podemos establecer un **orden** entre las variables transformadas de forma que:

Y_1: Primer eje en P' (*primera componente principal*).

La dirección de la máxima varianza de los patrones en P está determinada por este eje.

Y_2: Segundo eje en P' (*segunda componente principal*).

La dirección de la máxima varianza en P entre todos los ejes ortogonales a Y_1 está determinada por este eje.

… … … …

Con estas consideraciones, el ***teorema fundamental del análisis de componentes principales*** se enuncia como sigue:

Dado un conjunto de variables X_i (i = 1, 2,..., p) con matriz de covariancia $\sum_X$, no singular, siempre se puede derivar a partir de ellos un conjunto de variables no correlacionadas Y_i (i = 1, 2,..., p) mediante un conjunto de transformaciones lineales que corresponden a una rotación rígida cuya matriz de transformación W está formada, por columnas, por los p autovectores de $\sum_X$.

La matriz de covariancia del nuevo conjunto de variables, $\sum_Y$, es diagonal, y contiene los autovalores de $\sum_X$.

La transformación de componentes principales definida por (3) con la restricción de diagonalidad se conoce también como ***transformación de Karhunen-Loéwe*** o de ***Hotelling***.

A modo de resumen, y con una interpretación geométrica, los autovalores λ_i representan la varianza de los patrones en el espacio transformado y están relacionados con el *rango* de los patrones en cada uno de los ejes de este espacio mientras que los autovectores ϕ_i son vectores ortogonales que determinan la *dirección* de estos ejes.

Si se define la *varianza total* de un conjunto de datos multidimensionales como la suma de las varianzas asociadas a cada atributo, como las varianzas individuales se encuentran en la diagonal de la matriz de covariancia, el cálculo de la varianza global se reduce al cálculo de la *traza* de la matriz de covariancia.

Resulta evidente que si $\sum_Y$ es la matriz que contiene en su diagonal los autovalores $\lambda_1, \lambda_2,..., \lambda_p$ de $\sum_X$, entonces,

$$\mathrm{tr}(\textstyle\sum_Y) = \displaystyle\sum_{i=1}^{p} \lambda_i$$

Además, si la transformación de componentes principales preserva la varianza global, entonces,

$$\mathrm{tr}(\textstyle\sum_X) = \mathrm{tr}(\textstyle\sum_Y) = \displaystyle\sum_{i=1}^{p} \lambda_i \qquad (15)$$

2.7 Algoritmo de cálculo de análisis de componentes principales

El algoritmo de cálculo de *W* puede plantearse en 4 pasos:

1. Calcular la matriz de covariancia *global* $\sum_X$.

   Para este cálculo se utilizan todos los patrones *X*. En ningún caso se consideran *prototipos*, ya que no se tiene en cuenta la clase.

2. Calcular los autovalores de $\sum_X$, $\lambda_1, \lambda_2, \ldots, \lambda_p$

3. Calcular los autovectores[1] $\phi_1, \phi_2, \ldots, \phi_p$ asociados a $\lambda_1, \lambda_2, \ldots, \lambda_p$

4. Formar la matriz $W = [\phi_1, \phi_2, \ldots, \phi_p]$

Una vez formada la matriz de transformación *W* se procede a calcular los nuevos patrones, *Y*, a partir de cada *X*. Como $Y = W^T X$, (3) y la matriz *W* es la matriz formada por los autovectores de $\sum_X$ (12), sustituyendo en (4) obtenemos:

$$\begin{bmatrix} Y_1 \\ . \\ . \\ . \\ Y_p \end{bmatrix} = \begin{bmatrix} \phi_{11} & . & . & . & \phi_{1p} \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ \phi_{p1} & . & . & . & \phi_{pp} \end{bmatrix} \begin{bmatrix} X_1 \\ . \\ . \\ . \\ X_p \end{bmatrix}$$

o lo que es lo mismo,

$$\begin{aligned} Y_1 &= \phi_{11}X_1 + \phi_{12}X_2 + \ldots + \phi_{1p}X_p \\ Y_2 &= \phi_{21}X_1 + \phi_{22}X_2 + \ldots + \phi_{2p}X_p \\ &\vdots \\ Y_p &= \phi_{p1}X_1 + \phi_{p2}X_2 + \ldots + \phi_{pp}X_p \end{aligned}$$

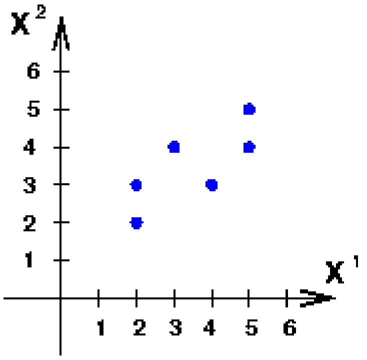
donde

$$\phi_1 = \begin{bmatrix} \phi_{11} \\ \phi_{12} \\ . \\ \phi_{1p} \end{bmatrix}, \ldots, \phi_p = \begin{bmatrix} \phi_{p1} \\ \phi_{p2} \\ . \\ \phi_{pp} \end{bmatrix}$$

2.8 Ejemplo de Aplicación Nº 1

Como ilustración, mostraremos cómo se aplica la transformación de componentes principales a un conjunto de datos que presenta cierta correlación. En la Figura 2 mostramos los 6 datos sobre los que se va a efectuar la transformación. Como se observa, las variables X_1 y X_2 presentan una correlación positiva.

Figure 2: Patrones *X* en el espacio original *P*



X_i
(2, 2)
(2, 3)
(3, 4)
(4, 3)
(5, 4)
(5, 5)

1. Cálculo de $\sum_X$

El vector medio μ_X y la matriz de covariancia $\sum_X$ se calculan a partir de los datos, obteniendo:

$$\mu_X = \begin{pmatrix} 3.50 \\ 3.50 \end{pmatrix} \qquad \sum_X = \begin{pmatrix} 1.9 & 1.1 \\ 1.1 & 1.1 \end{pmatrix}$$

2. Cálculo de los autovalores de $\sum_X$

Como p = 2 habrán dos autovalores asociados a $\sum_X$: λ_1 y λ_2.

Serán las soluciones a la ecuación (11) ($|\sum_X - \lambda I| = 0$) y las soluciones son: $\lambda_1 = 2.67$ y $\lambda_2 = 0.33$

3. Cálculo de los autovectores ϕ_1 y ϕ_2 asociados a λ_1 y λ_2

El autovector asociado a $\lambda_1 = 2.67$ es

$$\phi_1 = \begin{bmatrix} 0.82 \\ 0.57 \end{bmatrix}$$

El autovector asociado a $\lambda_2 = 0.33$ es

$$\phi_2 = \begin{bmatrix} -0.57 \\ 0.82 \end{bmatrix}$$

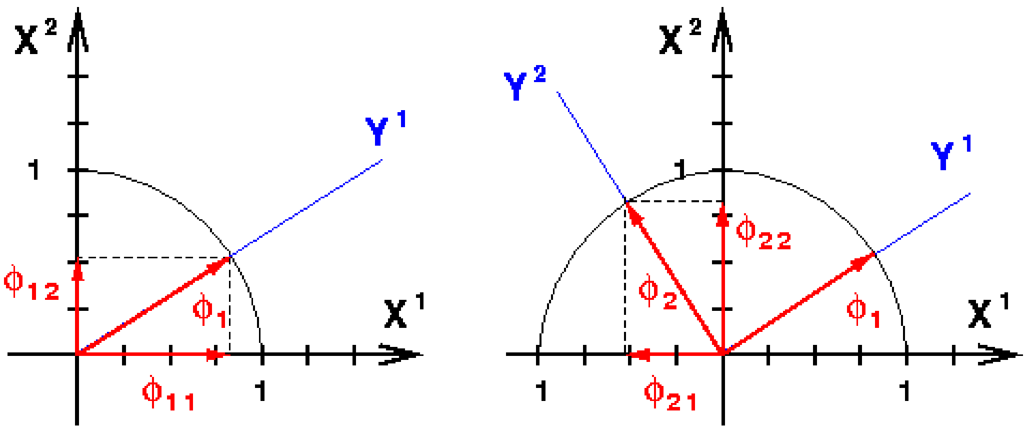
Los autovectores que se acaban de calcular están normalizados. Esto implica que son de longitud 1. Efectivamente, para ambos autovectores

$\phi_{11}^2 + \phi_{12}^2 = 0.82^2 + 0.57^2 = 0.67 + 0.33 = 1$

$\phi_{21}^2 + \phi_{22}^2 = - 0.57^2 + 0.82^2 = 0.33 + 0.67 = 1$

Las componentes de un autovector indican la dirección de los nuevos ejes respecto al sistema de coordenadas original. La interpretación geométrica del nuevo sistema de coordenadas (Y_1, Y_2) respecto al original (X_1, X_2) en base a los autovectores ϕ 1 y ϕ 2 se detalla en la Figura 3.

Figure 3: Los autovectores determinan el nuevo sistema de coordenadas



4. Formar la matriz de transformación *W*

La matriz de transformación es una matriz cuadrada 2 x 2 cuyas columnas son los autovectores y ϕ 2:

$$W = [\phi_1 \quad \phi_2] = \begin{bmatrix} 0.82 & -0.57 \\ 0.57 & 0.82 \end{bmatrix}$$

Finalmente, se procede a la transformación de coordenadas para expresar los patrones X en las coordenadas del nuevo espacio. La transformación viene dada por la ecuación (3) ($Y = W^T X$), o sea,

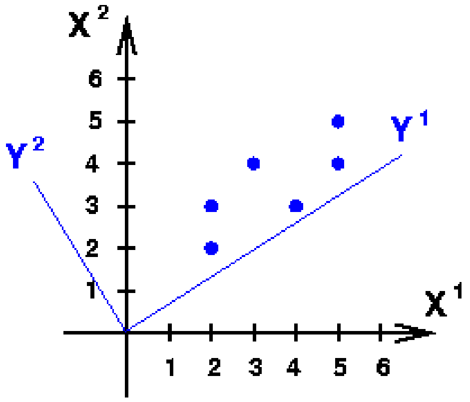
$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} 0.82 & 0.57 \\ -0.57 & 0.82 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

Expresado en término del sistema de ecuaciones es:

$Y_1 = 0.82\ X_1 + 0.57\ X_2$
$Y_2 = -0.57\ X_1 + 0.82\ X_2$

Si aplicamos esta transformación a los datos, el resultado se muestra en la Figura 4.

Figure 4: 6 Datos en dos sistemas de coordenadas



X_i	Y_i
(2, 2)	(2.78, 0.50)
(2, 3)	(3.35, 1.32)
(3, 4)	(4.74, 1.57)
(4, 3)	(4.99, 0.18)
(5, 4)	(6.38, 0.43)
(5, 5)	(6.95, 1.25)

Consideraciones finales

1) La matriz de covariancia en $\sum_Y$ es diagonal y contiene los autovalores asociados a $\sum_X$. En este caso,

$$\sum_Y = \begin{bmatrix} 2.67 & 0 \\ 0 & 0.33 \end{bmatrix}$$

Este hecho queda patente de manera gráfica en la Figura 4 si consideramos el nuevo sistema de coordenadas (Y_1, Y_2). Al comparar las matrices de covariancia en P y P' observamos que:

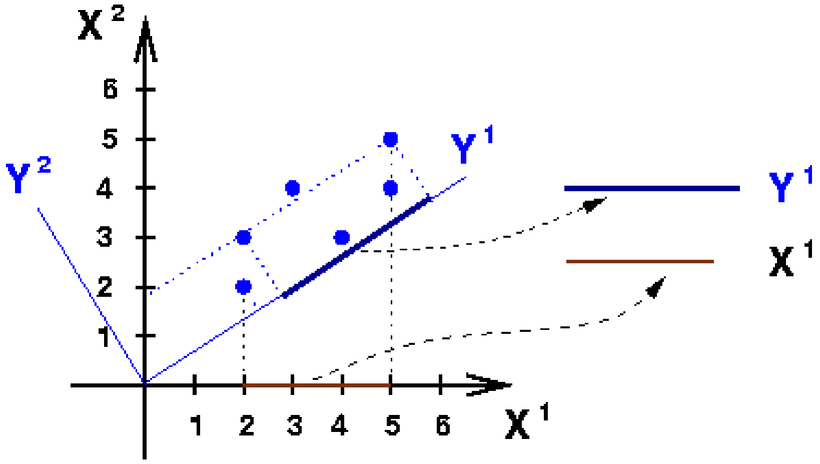
1. Las variables Y_1 e Y_2 están no correlacionadas ($\rho_{Y_1Y_2}$ = 0) mientras que las variables X_1 y X_2 están (fuertemente) correlacionadas:

$$\rho_{X1,X2)} = \frac{\sum_{12X}}{\sqrt{Var(X_1)}\sqrt{Var(X_2)}} = \frac{1.1}{\sqrt{1.9}\sqrt{1.1}} = 0.76$$

2. La transformación aplicada ha tenido el efecto de maximizar la varianza. La varianza en el primer eje principal, Y_1, es 2.67, bastante mayor que en X_1, 1.9. Además, no existe ningún otro eje en el que haya una varianza mayor.

   De manera gráfica puede verse en la Figura 5 en la que hemos proyectado los datos con menor y mayor valor de la variable X_1 sobre los ejes X_1 e Y_1.

Figure 5: Rango de los datos en los ejes originales y transformados



3. La transformación preserva la varianza global:

$$tr(\sum_X) = 1.9 + 1.1 = 3 \;\; y \;\; tr(\sum_Y) = \Sigma\lambda_i = 2.67 + 0.33 = 3$$

2) $\sum_Y$ también puede calcularse como se indica en (8) ($\sum_Y = W^T \sum_X W$).

$$W^T \sum_X W = \begin{bmatrix} 0.82 & 0.57 \\ -0.57 & 0.82 \end{bmatrix} \begin{bmatrix} 1.9 & 1.1 \\ 1.1 & 1.1 \end{bmatrix} \begin{bmatrix} 0.82 & -0.57 \\ 0.57 & 0.82 \end{bmatrix} = \begin{bmatrix} 2.67 & 0 \\ 0 & 0.33 \end{bmatrix} = \sum_Y$$

3) La matriz de transformación indicada en (10) debe ser ortogonal.

Como ya indicamos en (6), W es una matriz ortogonal si $W^TW = WW^T = I$. En este caso,

$$W^T W = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Resultados con el MINITAB

Análisis de componente principal: X1; X2

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

Valor propio   2,6705   0,3295
Proporción      0,890    0,110
Acumulada       0,890    1,000


Variable     PC1      PC2
X1         0,819   -0,574
X2         0,574    0,819
```

III. Enfoque Estadístico[4]

3.1 Introducción

El análisis de componentes principales tiene como objeto transformar un conjunto de variables, a las que se denominan variables originales, en un nuevo conjunto de variables denominadas componentes principales. Estas ultimas se caracterizan por estar no correlacionadas entre si.

En muchas ocasiones el investigador se enfrenta a situaciones en las que para analizar un fenómeno, dispone de información de muchas variables que están correlacionadas entre si en mayor o menor grado. Estas correlaciones son como un velo que impiden evaluar adecuadamente el papel que juega cada variable en el fenómeno estudiado. El ACP permite pasar a un nuevo conjunto de variables -las componentes principales- que gozan de la ventaja de estar no correlacionadas entre si y que, además, pueden ordenarse de acuerdo con la información que llevan incorporada.

Como medida de la cantidad de información incorporada en una componente se utiliza su variancia. Es decir, cuanto mayor sea su variancia mayor es la información que lleva incorporada dicha componente. Por esta razón se selecciona como primera componente aquella que tenga mayor variancia, mientras que, por el contrario, la ultima es la de menor variancia

3.2 Propiedades en ACP con las variables originales

Las componentes principales satisfacen las siguientes propiedades:

$$Var(Y_i) = \lambda_i$$

$$Cov(Y_i, Y_j) = C_{ij} = 0, \, para \, i \neq j$$

$$\rho_{Y_i X_k} = \frac{e_{ik}\sqrt{\lambda_i}}{\sqrt{Var(X_k)}}$$

Si las variables $X_1,.....,X_p$ tienen variancias correspondientes $Var(X_1), Var(X_2), ..., Var(X_p)$ entonces:

$$Var(X_1) + Var(X_2) + ... + Var(X_p) = \sum_{i=1}^{p} Var(X_i) = \lambda_1 + \lambda_2 + ... + \lambda_p = \sum_{i=1}^{p} Var(Y_i)$$

3. Basado en el Uriel, Ezequiel & Aldas, Joaquín. “Análisis Multivariante Aplicado. Aplicaciones al marketing, investigación de mercados, economía, dirección de empresas y turismo”. Editorial Thomson. 2005. España.

3.3 Propiedades en ACP con las variables estandarizadas

En general, la extracción de componentes principales se efectúa sobre variables estandarizadas para evitar problemas derivados de escala.

Si p variables están estandarizadas, la suma de las variancias es igual a p, ya que la variancia de una variable estandarizada es por definición igual a 1.

Cuando se estandarizan los datos, la matriz de variancia-covariancia es igual a la matriz de correlación. Es decir, que para aplicar ACP sobre datos estandarizados es suficiente trabajar con la matriz de correlación.

Si: $Z_i = \frac{(X_i - \mu_i)}{\sqrt{Var(X_i)}}$, las componentes principales satisfacen las siguientes propiedades:

$$\sum_{i=1}^{p} Var(Y_i) = \lambda_1 + .\lambda_2 + ... + \lambda_p = \sum_{i=1}^{p} Var(Z_i) = p$$

$$\rho_{Y_i Z_k} = e_{ik}\sqrt{\lambda_i}$$

3.3 Ejemplo de Aplicación Nº 2

1	2	3	4	5	6
X1	X2	Z1	Z2	Score ACP 1	Score ACP 2
775104	23795	1.25752	0.55653	1.28273	0.495674
775218	58778	1.25838	1.92752	2.25277	-0.473152
700963	1531	0.69751	-0.316	0.26977	0.716657
674063	-12756	0.49432	-0.87591	-0.26982	0.9689
631003	14729	0.16908	0.20123	0.26185	-0.022738
537744	9059	-0.53534	-0.02098	-0.39337	-0.363711
489155	12541	-0.90235	0.11548	-0.5564	-0.719717
448465	13495	-1.20969	0.15287	-0.74729	-0.96348
445853	-34824	-1.22942	-1.74076	-2.10023	0.361567

A) Estadísticas descriptivas: X1; X2

```
Variable  N   Media     Varianza
X1        9  608619  17527605543
X2        9    9594    651098064
```

B) Covarianzas: X1; X2

```
             X1           X2
X1  17527605543
X2   1844599328    651098064
```

C) Correlaciones: X1; X2

```
Correlación de Pearson de X1 y X2 = 0,546
```

D) Covarianzas: Z1; Z2

```
         Z1       Z2
Z1  1,00000
Z2  0,54603  1,00000
```

E) 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
```

F) Análisis de componente principal: Z1; Z2

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

Valor propio  1,5460  0,4540
Proporción     0,773   0,227
Acumulada      0,773   1,000


Variable    PC1     PC2
Z1        0,707   0,707
Z2        0,707  -0,707
```

G) Estadísticas descriptivas: Puntuacion 1; Puntuacion 2

```
Variable     N  Media  Varianza
Puntuacion 1 9  0,000     1,546
Puntuacion 2 9  0,000     0,454
```

H) Covarianzas: Puntuacion 1; Puntuacion 2

```
             Puntuacion 1  Puntuacion 2
Puntuacion 1    1,546031
Puntuacion 2    0,000000     0,453969
```

I) Correlaciones: Puntuacion 1; Puntuacion 2

```
Correlación de Pearson de Puntuacion 1 y Puntuancion 2 = 0,000
```

3.4 Número de componentes a retener

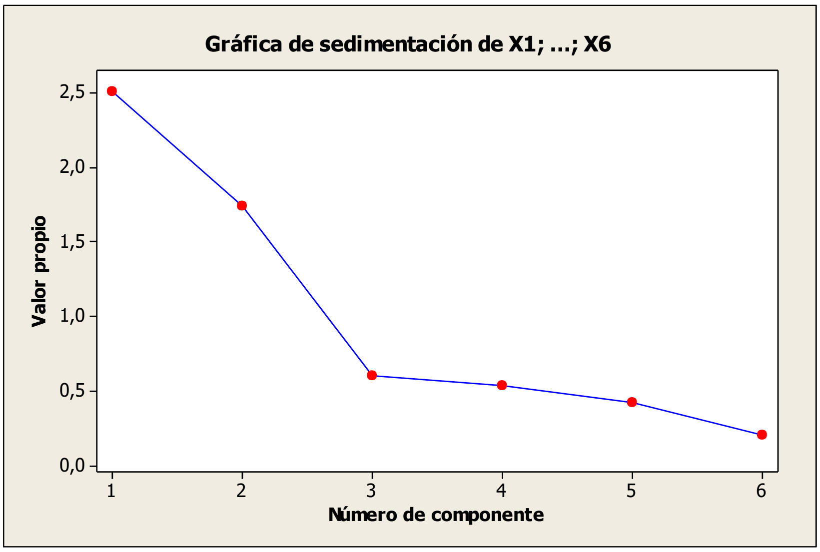
En general, el objetivo de la aplicación del ACP es reducir las dimensiones de las variables originales, pasando de p variables originales a m p componentes principales
El problema que se plantea es como fijar m, o, dicho de otra forma, que número de componentes se deben retener.

Criterio de la media aritmética
Se debe seleccionar aquellas componentes cuyo autovalor excede de la media de los autovalores.

Caso	Criterio
Cuando se trabajan con las variables originales	$\lambda_h > \bar{\lambda} = \frac{\sum_{i=1}^{p} \lambda_i}{p}$
Cuando se trabajan con las variables estandarizadas	$\lambda_h > 1$

Criterio del grafico de sedimentación (Scree Plot)

Se obtiene al representar en ordenadas los autovalores y en abscisas el numero de la componente en orden decreciente. Uniendo todos los puntos se obtiene una figura que, en general, se parece al perfil de una montaña con una pendiente fuerte hasta llegar a la base, formada por una meseta con una ligera inclinación. Continuando con el símil de la montaña, en esa meseta es donde se acumulan los guijarros caídos desde la cumbre, es decir, donde se sedimentan.



De acuerdo con este criterio, se retienen todas aquellas componentes previas a la zona de sedimentación.

Retención de variables

Si se retiene un numero determinado de componentes, que hacer si alguna variable esta correlacionada muy débilmente con cada una de las componentes retenidas. Si se plantea un caso de este tipo, seria conveniente suprimir dicha variable del conjunto de variables originales, ya que no estaría representada por las componentes retenidas. Ahora bien, si se considera que la variable a suprimir juega un papel esencial en la investigación, entonces se deberían retener componentes adicionales en el caso de que algunas de ellas estuvieran correlaciones de forma importante con la variable a eliminar.

3.5 Ejemplo de Aplicación Nº 3[5]

HATCO es una empresa fabricante de maquinaria industrial que ha pasado una encuesta a los jefes de compras de las empresas que adquieren sus productos, los cuales han valorado su satisfacción con HATCO respecto a seis atributos determinantes de su servicio.

De forma más detallada, las seis variables que miden la percepción que tienen de HATCO sus clientes son las siguientes:

- X1: **Rapidez del servicio**. Tiempo que tarda en servirse el pedido una vez que éste ha sido confirmado.
- X2: **Nivel de precios**. Valoración sobre el precio que se carga respecto a otros suministradores.
- X3: **Flexibilidad de precios**. Voluntad de los vendedores de HATCO de negociar el precio en todo tipo de compras.
- X4: **Imagen del fabricante**. Imagen global de HATCO.
- X5: **Imagen de los vendedores**. Imagen global de la fuerza de ventas de HATCO.
- X6: **Calidad del producto**. Nivel de calidad percibida de los productos de HATCO.

Todas estas variables se han medido mediante una escala constituida por una línea de diez centímetros donde en los extremos aparecen las palabras "muy mala" y "excelente".

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

Estadísticas descriptivas: X1; X2; X3; X4; X5; X6

Variable	N	Media	Desv.Est.	Varianza
X1	100	3,515	1,321	1,744
X2	100	2,364	1,196	1,430
X3	100	7,894	1,387	1,922
X4	100	5,248	1,131	1,280
X5	100	2,6660	0,7704	0,5936
X6	100	6,971	1,585	2,513

Correlaciones: X1; X2; X3; X4; X5; X6

	X1	X2	X3	X4	X5
X2	-0,349				
	0,000				
X3	0,509	-0,487			
	0,000	0,000			
X4	0,050	0,272	-0,116		
	0,618	0,006	0,250		
X5	0,078	0,184	-0,035	0,788	
	0,442	0,066	0,728	0,000	
X6	-0,483	0,470	-0,448	0,200	0,176
	0,000	0,000	0,000	0,046	0,080

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

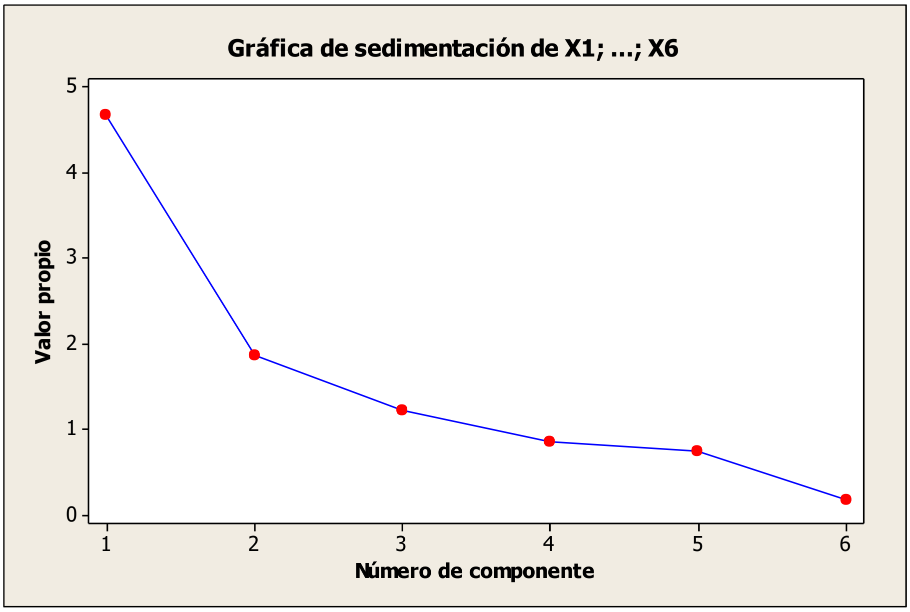
Resultado del MINITAB usando la matriz de covariancia

Análisis de componente principal: X1; X2; X3; X4; X5; X6

Análisis de los valores y vectores propios de la matriz de covarianza

Valor propio	4,6733	1,8550	1,2079	0,8433	0,7383	0,1651
Proporción	0,493	0,196	0,127	0,089	0,078	0,017
Acumulada	0,493	0,688	0,816	0,905	0,983	1,000

Variable	PC1	PC2	PC3	PC4	PC5	PC6
X1	-0,440	0,417	-0,004	-0,691	-0,393	-0,000
X2	0,388	0,177	0,272	-0,529	0,680	-0,037
X3	-0,496	0,173	-0,625	0,101	0,568	0,034
X4	0,138	0,734	0,161	0,366	-0,022	0,531
X5	0,068	0,463	0,044	0,251	-0,027	-0,846
X6	0,622	0,108	-0,712	-0,188	-0,242	0,024



Puntuacion COV 1	Puntuacion COV 2
-0,9447919	8,13070368
3,65705153	9,97393665
3,80834425	9,86400614
1,50245427	8,76732045
-2,85827572	12,6611866
.	.
.	.
.	.

Correlaciones entre las Variables y los Componentes

	PC1	PC2
X1	-0.720	0.430
X2	0.702	0.202
X3	-0.773	0.170
X4	0.265	0.884
X5	0.192	0.818
X6	0.848	0.093

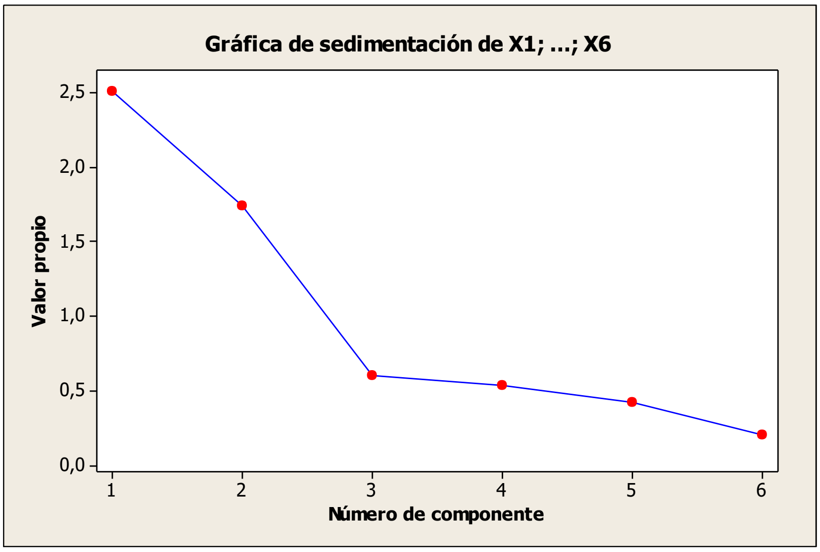
Resultado del MINITAB usando la matriz de correlación

Análisis de componente principal: X1; X2; X3; X4; X5; X6

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

Valor propio	2,5127	1,7397	0,5976	0,5300	0,4156	0,2044
Proporción	0,419	0,290	0,100	0,088	0,069	0,034
Acumulada	0,419	0,709	0,808	0,897	0,966	1,000

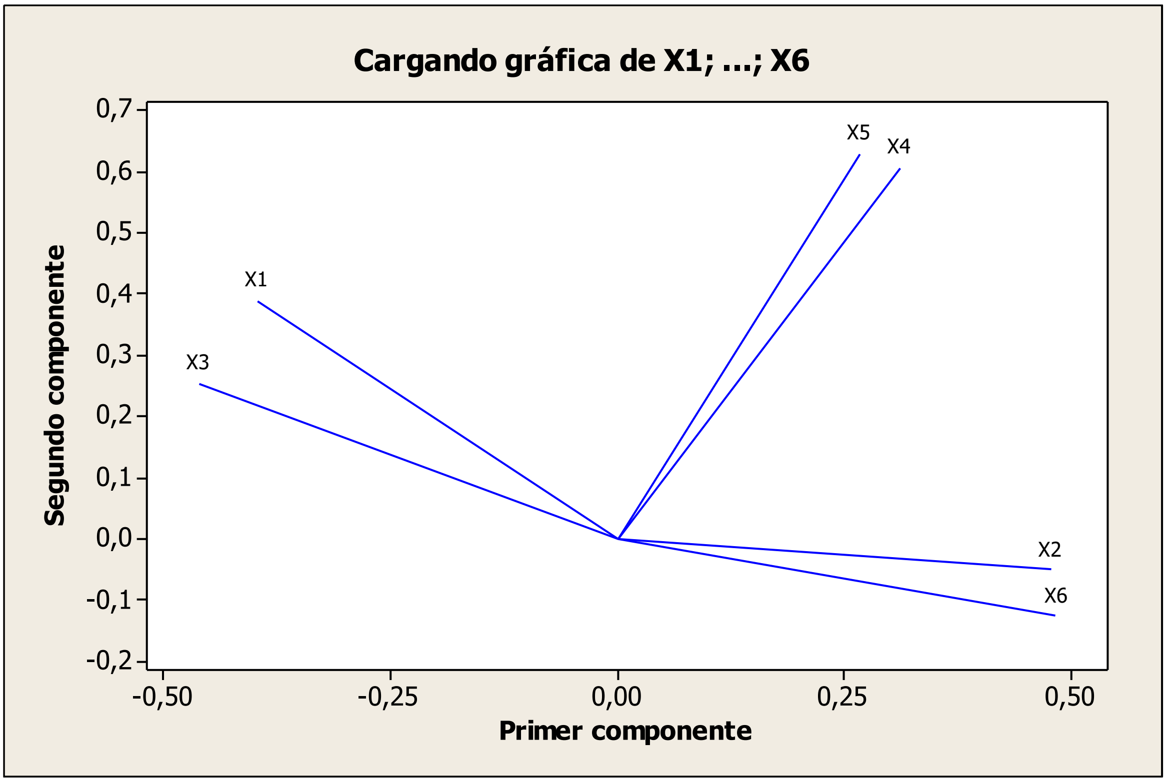
Variable	PC1	PC2
X1	-0,396	0,390
X2	0,478	-0,052
X3	-0,461	0,254
X4	0,311	0,606
X5	0,267	0,631
X6	0,484	-0,127



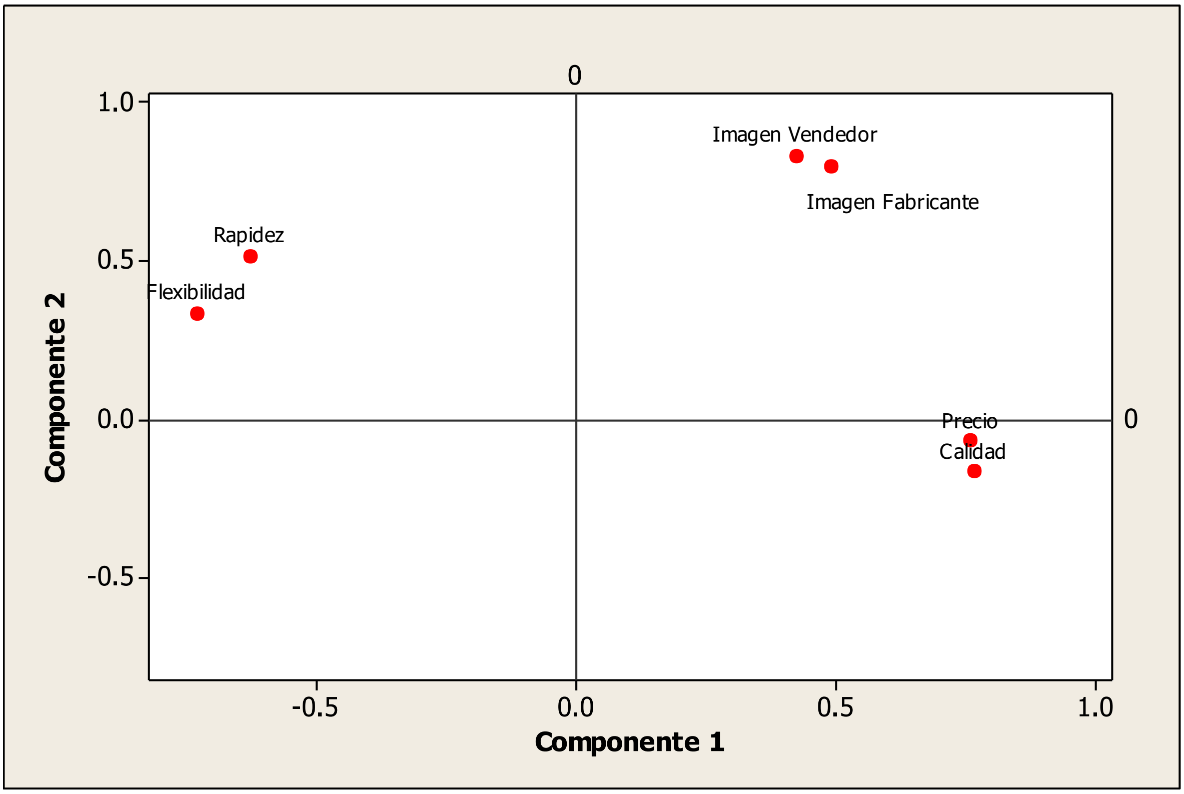
Puntuacion CORR 1	Puntuacion CORR 2
-1,33421996	-0,30285262
2,56865743	0,87678834
2,49209568	-0,22595742
0,26756719	-0,34407798
-1,27815258	4,25646285
.	.
.	.
.	.

Correlaciones entre las Variables y los Componentes

Variable	**PC1**	**PC2**
X1 (Rapidez)	-0.627	0.514
X2 (Precios)	0.758	-0.068
X3 (Flexibilidad)	-0.730	0.335
X4 (Imagen del fabricante)	0.494	0.799
X5 (Imagen del vendedor)	0.424	0.832
X6 (Calidad)	0.767	-0.167



Cargando gráfica de X1; ...; X6
Segundo componente
Primer componente
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
-0,2
-0,50
-0,25
0,00
0,25
0,50
X1
X2
X3
X4
X5
X6



Componente 2
Componente 1
1.0
0.5
0.0
-0.5
-0.5
0.0
0.5
1.0
0
0
Imagen Vendedor
Imagen Fabricante
Rapidez
Flexibilidad
Precio
Calidad

Resultado del R usando la matriz de correlación

```
###################################################################
# Ejemplo N°2 - HATCO con cinco variables                         #
# Análisis de Componentes Principales usando la libreria ade4     #
###################################################################

# Librerias
library(ade4)
library(FactoMineR)

# Lectura de datos
datos=read.delim("hatco-acp.txt")
str(datos)
```

```
'data.frame':   100 obs. of  7 variables:
 $ id: int  1 2 3 4 5 6 7 8 9 10 ...
 $ x1: num  4.1 1.8 3.4 2.7 6 1.9 4.6 1.3 5.5 4 ...
 $ x2: num  0.6 3 5.2 1 0.9 3.3 2.4 4.2 1.6 3.5 ...
 $ x3: num  6.9 6.3 5.7 7.1 9.6 7.9 9.5 6.2 9.4 6.5 ...
 $ x4: num  4.7 6.6 6 5.9 7.8 4.8 6.6 5.1 4.7 6 ...
 $ x5: num  2.35 4 2.7 2.3 4.6 1.9 4.5 2.2 3 3.2 ...
 $ x6: num  5.2 8.4 8.2 7.8 4.5 9.7 7.6 6.9 7.6 8.7 ...
```

```
# No considerar la primera columna Id
datos=datos[,-1]

str(datos)
```

```
'data.frame':   100 obs. of  6 variables:
 $ x1: num  4.1 1.8 3.4 2.7 6 1.9 4.6 1.3 5.5 4 ...
 $ x2: num  0.6 3 5.2 1 0.9 3.3 2.4 4.2 1.6 3.5 ...
 $ x3: num  6.9 6.3 5.7 7.1 9.6 7.9 9.5 6.2 9.4 6.5 ...
 $ x4: num  4.7 6.6 6 5.9 7.8 4.8 6.6 5.1 4.7 6 ...
 $ x5: num  2.35 4 2.7 2.3 4.6 1.9 4.5 2.2 3 3.2 ...
 $ x6: num  5.2 8.4 8.2 7.8 4.5 9.7 7.6 6.9 7.6 8.7 ...
```

```
library(ade4)

# Analisis de Componentes Principales
acp=dudi.pca(datos,scannf=FALSE,nf=ncol(datos))

str(acp)

# Valores propios
inertia.dudi(acp)
```

```
    inertia      cum     ratio
1 2.5131167 2.513117 0.4188528
2 1.7396167 4.252733 0.7087889
3 0.5975188 4.850252 0.8083754
4 0.5297714 5.380024 0.8966706
5 0.4156729 5.795697 0.9659494
6 0.2043035 6.000000 1.0000000
```

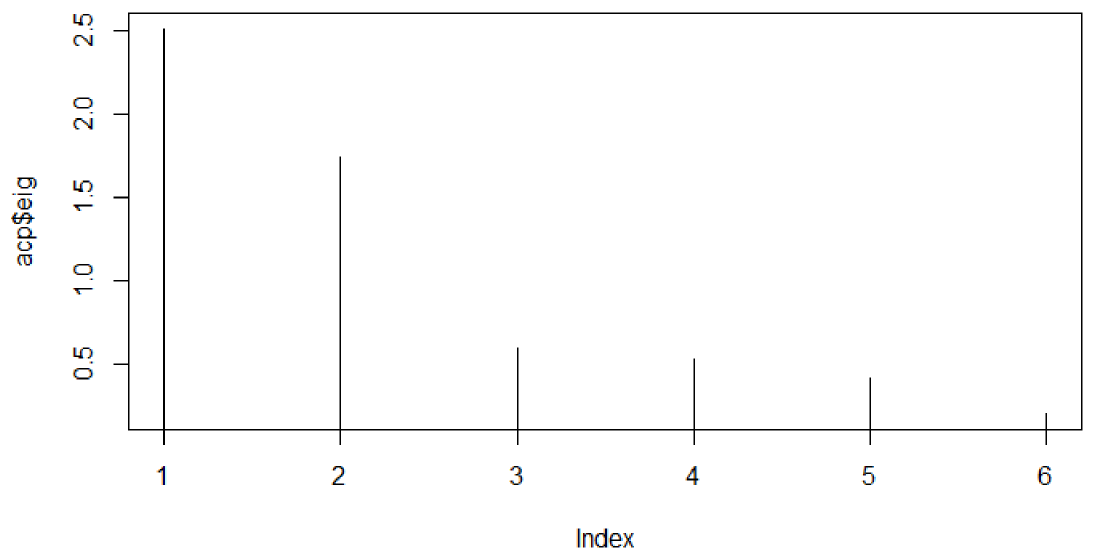
Vectores propios
acp$c1

```
          CS1         CS2         CS3        CS4         CS5         CS6
x1  0.3955782 -0.38979714  0.50720784  0.2712569  0.59972722 -0.03260799
x2 -0.4785076  0.05152613  0.69879999  0.3384002 -0.39921358 -0.07853632
x3  0.4604792 -0.25457311 -0.25579305  0.6130807 -0.52641191  0.06887653
x4 -0.3115565 -0.60546529 -0.04024493 -0.1536064 -0.04571997  0.71346997
x5 -0.2676730 -0.63077805 -0.19253817 -0.1126846 -0.05718850 -0.69096378
x6 -0.4835195  0.12699058 -0.38768298  0.6322534  0.44551705  0.03942657
```

Correlaciones entre las variables y los componentes
acp$co

```
        Comp1       Comp2       Comp3       Comp4       Comp5       Comp6
x1  0.6271027 -0.51412111  0.39206831  0.19743538  0.38666034 -0.01473879
x2 -0.7585691  0.06796015  0.54016778  0.24630594 -0.25738378 -0.03549837
x3  0.7299890 -0.33576801 -0.19772634  0.44623319 -0.33939198  0.03113215
x4 -0.4939048 -0.79857560 -0.03110906 -0.11180304 -0.02947690  0.32248801
x5 -0.4243370 -0.83196175 -0.14883073 -0.08201792 -0.03687097 -0.31231522
x6 -0.7665143  0.16749363 -0.29967638  0.46018812  0.28723687  0.01782078
```

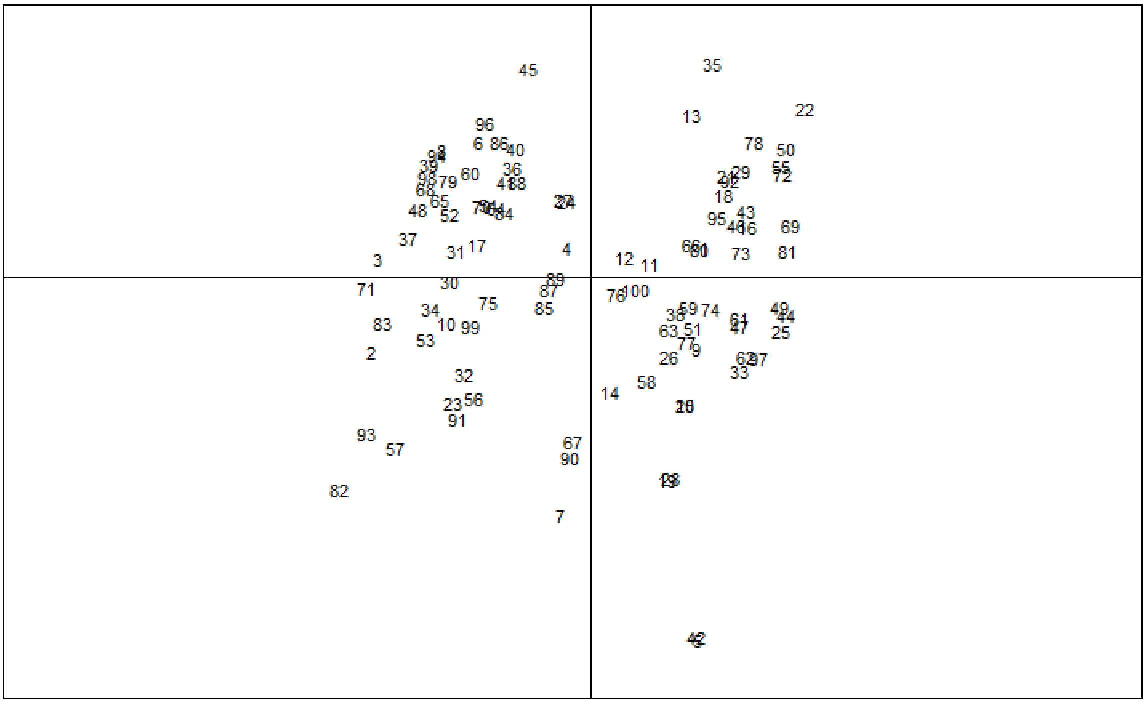
# Grafica de Valores propios
plot(acp$eig,type="h")



Scores o Puntuaciones de cada individuo
acp$li

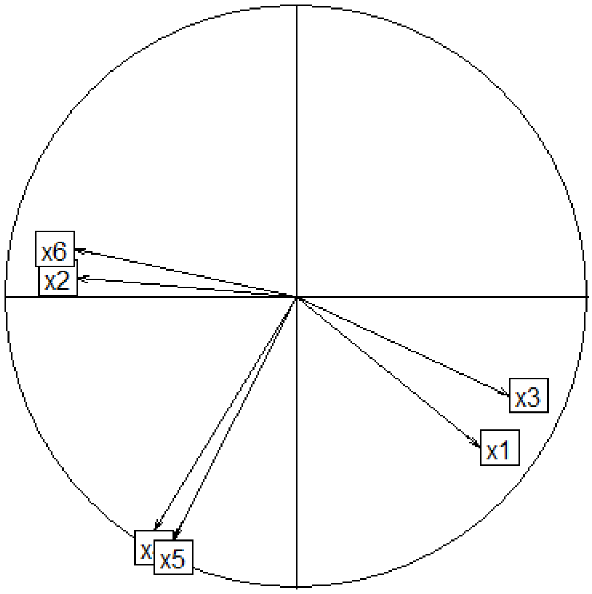
```
        Axis1       Axis2       Axis3       Axis4        Axis5        Axis6
1   1.3585305  0.345191265 -0.09195343 -1.41151293  0.783771230 -0.054961760
2  -2.5822265 -0.879518018 -0.72746921 -0.68926911 -0.138676440 -0.388993758
3  -2.5045616  0.227923493  1.69064038  0.19289660  0.147061116  0.182326260
4  -0.2687168  0.345943358 -1.10383849 -0.61202830  0.623680613  0.833936750
5   1.2822237 -4.278243325 -0.18677005 -0.76833138  0.028511822 -0.067456079
6  -1.3059876  1.608883233 -0.53718295  1.20308218 -0.207317072  0.452686341
7  -0.3592074 -2.802293223 -0.52143919  0.74597540 -0.143799975 -0.729663066
8  -1.7454567  1.505968351  0.67725205 -0.62764745 -0.959902188  0.173087725
9   1.2495874 -0.829596395 -0.18085726  1.13944579  0.762592056 -0.556647469
10 -1.7009623 -0.542373931  0.52687656  0.31565103  0.790023452 -0.118464588
.
.
.
```

# Grafica de individuos sobre el primer plano de componentes
s.label(acp$li,clabel=0.7,grid=FALSE,boxes=FALSE)



# Grafica de individuos sobre los componentes 2 y 3
s.label(acp$li,xax=2,yax=3,clabel=0.7,grid=FALSE,boxes=FALSE)

# Grafica de circulo de correlaciones
s.corcircle(acp$co,grid=FALSE)



# Grafica de individuos sobre el primer plano con biplot
s.label(acp$li,clabel=0.7,grid=FALSE,boxes=FALSE)
s.corcircle(acp$co,grid=FALSE,add=TRUE,clabel=0.7)

# Almacena los resultados de los scores en un archivo CSV
write.csv(acp$li,"hatco-scores.csv")

# Almacena los datos y los resultados de los scores en un archivo CSV
datosc=cbind(datos,acp$li)
datosc
str(datosc)
write.csv(datosc,"hatco-resultados.csv")

# Almacena los resultados de las correlaciones en un archivo CSV
write.csv(acp$co,"hatco-correlaciones.csv")

Análisis de Componentes Principales usando la función princomp()

# Lectura de datos
datos=read.delim("hatco-acp.txt")
str(datos)

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

# Estadisticas descriptivas
summary(datos)

```
       x1              x2              x3               x4              x5              x6        
 Min.   :0.000   Min.   :0.200   Min.   : 5.000   Min.   :2.500   Min.   :1.100   Min.   : 3.700  
 1st Qu.:2.500   1st Qu.:1.475   1st Qu.: 6.700   1st Qu.:4.575   1st Qu.:2.200   1st Qu.: 5.800  
 Median :3.400   Median :2.150   Median : 8.050   Median :5.000   Median :2.600   Median : 7.150  
 Mean   :3.515   Mean   :2.364   Mean   : 7.894   Mean   :5.248   Mean   :2.666   Mean   : 6.971  
 3rd Qu.:4.600   3rd Qu.:3.225   3rd Qu.: 9.100   3rd Qu.:6.000   3rd Qu.:3.000   3rd Qu.: 8.325  
 Max.   :6.100   Max.   :5.400   Max.   :10.000   Max.   :8.200   Max.   :4.600   Max.   :10.000  
```

sapply(datos,mean)

```
    x1     x2     x3     x4     x5     x6 
3.5150 2.3640 7.8940 5.2480 2.6655 6.9710 
```

sapply(datos,var)

```
       x1        x2        x3        x4        x5        x6 
1.7443182 1.4296000 1.9223879 1.2800970 0.5938735 2.5129889 
```

cov(datos)

```
            x1         x2          x3          x4          x5         x6
x1  1.74431818 -0.5514747  0.93261616  0.07533333  0.07880556 -1.0104697
x2 -0.55147475  1.4296000 -0.80769293  0.36821010  0.17076566  0.8903596
x3  0.93261616 -0.8076929  1.92238788 -0.18213333 -0.03717879 -0.9849232
x4  0.07533333  0.3682101 -0.18213333  1.28009697  0.68717778  0.3586788
x5  0.07880556  0.1707657 -0.03717879  0.68717778  0.59387348  0.2157571
x6 -1.01046970  0.8903596 -0.98492323  0.35867879  0.21575707  2.5129889
```

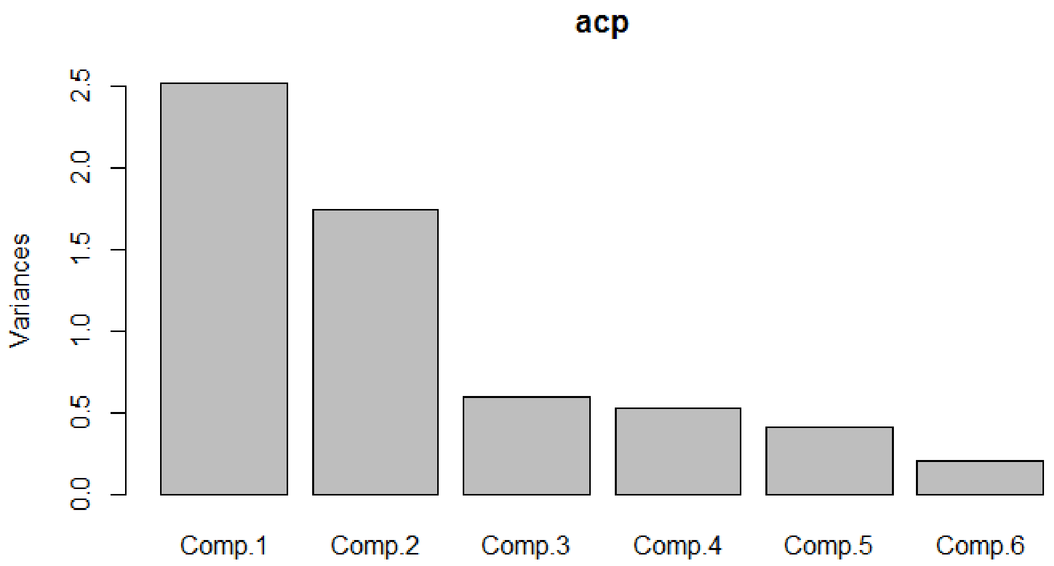
cor(datos)

```
            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
```

#Analisis de componentes principales desde la matriz de correlaciones
acp<-princomp(datos,cor=TRUE)
summary(acp)

```
Importance of components:
                          Comp.1    Comp.2     Comp.3     Comp.4     Comp.5     Comp.6
Standard deviation     1.5852813 1.3189453 0.77299340 0.72785395 0.64472701 0.45199942
Proportion of Variance 0.4188528 0.2899361 0.09958647 0.08829523 0.06927882 0.03405058
Cumulative Proportion  0.4188528 0.7087889 0.80837537 0.89667060 0.96594942 1.00000000
```

# grafico scree
plot(acp)



la desviacion estandar de cada componente principal
es decir la raiz de los valores propios de la matriz
acp$sdev

```
   Comp.1    Comp.2    Comp.3    Comp.4    Comp.5    Comp.6 
1.5852813 1.3189453 0.7729934 0.7278540 0.6447270 0.4519994 
```

Matriz con los vectores propios
acp$loadings

```
Loadings:
   Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
x1  0.396 -0.390  0.507  0.271  0.600       
x2 -0.479          0.699  0.338 -0.399       
x3  0.460 -0.255 -0.256  0.613 -0.526       
x4 -0.312 -0.605        -0.154         0.713
x5 -0.268 -0.631 -0.193 -0.113        -0.691
x6 -0.484  0.127 -0.388  0.632  0.446       

               Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
SS loadings     1.000  1.000  1.000  1.000  1.000  1.000
Proportion Var  0.167  0.167  0.167  0.167  0.167  0.167
Cumulative Var  0.167  0.333  0.500  0.667  0.833  1.000
```

la media de las variables originales con la que se centran
las observaciones
acp$center

numero de observaciones
acp$n.obs

las coordenadas factoriales
acp$scores

```
         Comp.1       Comp.2      Comp.3      Comp.4       Comp.5       Comp.6
 [1,]  1.3585305  0.345191265 -0.09195343 -1.41151293  0.783771230 -0.054961760
 [2,] -2.5822265 -0.879518018 -0.72746921 -0.68926911 -0.138676440 -0.388993758
 [3,] -2.5045616  0.227923493  1.69064038  0.19289660  0.147061116  0.182326260
 [4,] -0.2687168  0.345943358 -1.10383849 -0.61202830  0.623680613  0.833936750
 [5,]  1.2822237 -4.278243325 -0.18677005 -0.76833138  0.028511822 -0.067456079
 [6,] -1.3059876  1.608883233 -0.53718295  1.20308218 -0.207317072  0.452686341
 [7,] -0.3592074 -2.802293223 -0.52143919  0.74597540 -0.143799975 -0.729663066
 [8,] -1.7454567  1.505968351  0.67725205 -0.62764745 -0.959902188  0.173087725
 [9,]  1.2495874 -0.829596395 -0.18085726  1.13944579  0.762592056 -0.556647469
[10,] -1.7009623 -0.542373931  0.52687656  0.31565103  0.790023452 -0.118464588
```

#biplots
par(mfrow=c(2,2))

# primer plano factorial
biplot(acp)

#segundo plano factorial
biplot(acp,choices = c(1,3))

# otro plano factorial
biplot(acp,choices = c(2,3))
par(mfrow=c(1,1))

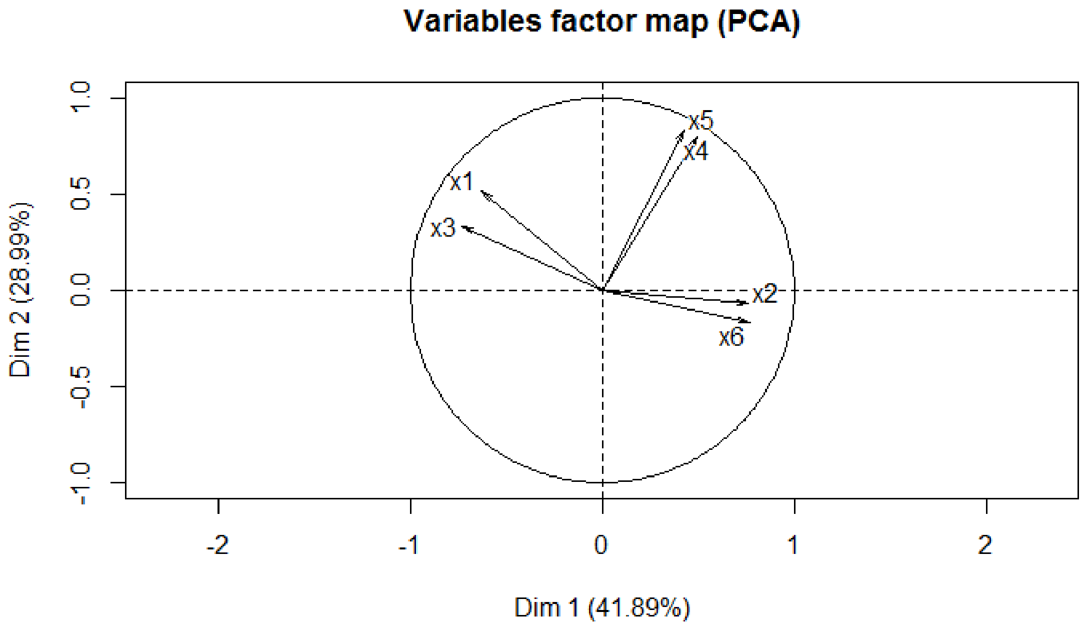
Análisis de Componentes Principales usando la libreria FactoMineR

# Lectura de datos
datos=read.delim("hatco-acp.txt")
str(datos)

# No considerar la primera columna Id
datos=datos[,-1]

str(datos)

library(FactoMineR)
resul=PCA(datos,ncp=2)



resul$eig

```
       eigenvalue percentage of variance cumulative percentage of variance
comp 1  2.5131167              41.885279                          41.88528
comp 2  1.7396167              28.993612                          70.87889
comp 3  0.5975188               9.958647                          80.83754
comp 4  0.5297714               8.829523                          89.66706
comp 5  0.4156729               6.927882                          96.59494
comp 6  0.2043035               3.405058                         100.00000
```

resulvarcor

```
        Dim.1       Dim.2
x1 -0.6271027  0.51412111
x2  0.7585691 -0.06796015
x3 -0.7299890  0.33576801
x4  0.4939048  0.79857560
x5  0.4243370  0.83196175
x6  0.7665143 -0.16749363
```

Conclusiones

Analizando las correlaciones y el gráfico el componente 1 viene explicado por las variables:

X1	Rapidez del servicio	signo -
X3	Flexibilidad de precios	signo -
X2	Nivel de precios	signo +
X6	Calidad del producto	signo +

Los signos indican que los grupos de variables se mueven de manera contraria, es decir, que cuando la calidad del producto y el precio disminuyen, la rapidez del servicio y la flexibilidad de precios aumentan. Podemos considerar que este componente está midiendo los factores principales del producto, su calidad y el precio, pudiendo interpretarlo como **relación calidad-precio**.

Por su parte, el segundo componente viene explicado por las variables:

X4	imagen del fabricante	signo +
X5	imagen de la fuerza de ventas	signo +

que permiten interpretar al componente como **imagen** de la empresa.

En síntesis, la percepción que tienen los clientes de HATCO, puede medirse mediante muchas variables (X1 a X6), pero podemos afirmar que, sin perder demasiada información, esta percepción se sintetiza en dos grandes ejes: la relación calidad-precio de sus productos y la imagen de la empresa.