

IN 540
Métodos Estadísticos
para Economía y Gestión

ANÁLISIS DE COMPONENTES PRINCIPALES

**Ref: “Introducción al Análisis Multivariante”
Angel Ganzo (2004)**

INTRODUCCIÓN

Con objeto de analizar la p -variables (correlacionadas) de la matriz de datos, se realiza una transformación de las variables originales en un nuevo conjunto de variables incorrelacionadas, mediante una rotación ortogonal en \mathbb{R}^p , que llamamos componentes o factores principales. Estas quedarán expresadas como combinación lineal de las originales, y se expresan en orden decreciente de importancia en cuanto a explicar la incidencia de cada componente principal en la descripción del problema.

El Análisis de Componentes Principales, ACP, (PCA en la literatura angloamericana), tiene su origen en los trabajos de Karl Pearson a principio de siglo, así como por Harold Hotelling, hacia 1930.

La técnica del ACP es adecuada cuando no se dispone de variables dependientes que permitan explicar el problema mediante una regresión múltiple, es decir que estamos ante una situación en que todas las variables, en principio, tienen la misma importancia, o bien que dicha importancia está enmascarada y es necesario ponerla de manifiesto.

El objetivo principal del análisis es averiguar cuantas variables, m , de entre las p , ($m < p$), explican mejor la variabilidad de los datos representados por las variables originales. Si ello es posible, podemos afirmar que la dimensionalidad del problema es menor que p . Por ejemplo si alguna de las variables originales están fuertemente correlacionadas con otras se pueden “agrupar” en una única variable (componente principal) expresada como combinación lineal de aquellas, y de esta manera se reduce la dimensión del problema.

Así pues, el ACP transforma un conjunto de variables correlacionadas en un nuevo conjunto de variables incorrelacionadas, donde la importancia de estas últimas vienen determinadas por la parte de varianza asociadas a ellas. La técnica ACP no requiere el uso de modelos probabilísticos, siendo, para nuestro caso, una técnica de tipo descriptivo.

Es conveniente realizar un análisis de componentes principales como estudio previo del análisis factorial.

PLANTEAMIENTO DEL PROBLEMA

Consideremos la nube de puntos en el espacio \mathbb{R}^p de variables tipificadas. Se trata de buscar direcciones u tales que $\overline{P_i A_i}$ sea lo más pequeño posible, Figura 1.

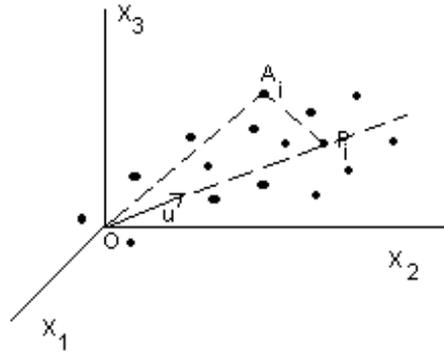


Figura 1

Como $\overline{OA_i}$ es constante cualquiera que sea la orientación del eje, será necesario que $\overline{OP_i}$ sea lo más grande posible, ya que $\overline{OA_i}^2 = \overline{OP_i}^2 + \overline{A_i P_i}^2$. Para tener en cuenta todos los puntos se toma la suma de los cuadrados de todos los puntos, es decir se pretende buscar aquella dirección u que maximiza el valor $\sum_{i=1}^n \overline{OP_i}^2$. Una dirección que cumple esta condición se llama *eje factorial* o *eje principal de inercia*, y es tal que proyectando la nube de puntos sobre él, estos se hallan muy separados o muy discriminados.

Se puede probar fácilmente que para una matriz de datos tipificados, la suma de cuadrados de las proyecciones es:

$$u^t R u, \quad \text{siendo } |u| = 1 \quad \text{o} \quad u^t u = 1$$

que es una forma cuadrática respecto de las componentes de $u = (u_1, u_2, \dots, u_p)$.

El máximo de esta función se determina por el método de los multiplicadores de Lagrange:

$$\begin{aligned} \varphi(\vec{u}) &= \varphi(u_1, u_2, \dots, u_p) = u^t R u - \lambda(u^t u - 1) = \\ &= u_1^2 + u_2^2 + \dots + u_p^2 + 2r_{12}u_1u_2 + \dots + 2r_{1p}u_1u_p + \dots + 2r_{p-1,p}u_{p-1}u_p - \lambda(u_1^2 + \dots + u_p^2 - 1) \end{aligned}$$

derivando φ respecto de u_1 , e igualando a cero, se obtiene:

$$\begin{aligned} \frac{\partial \varphi}{\partial u_1} &= 2u_1 + 2r_{12}u_2 + \dots + 2r_{1p}u_p - 2\lambda u_1 = 0 \\ (1 - \lambda)u_1 + r_{12}u_2 + \dots + r_{1p}u_p &= 0 \end{aligned}$$

realizando el mismo proceso respecto de u_2, \dots, u_p se obtiene el sistema de ecuaciones lineales:

$$\begin{aligned} (1-\lambda)u_1 + r_{12}u_2 + \dots + r_{1p}u_p &= 0 \\ r_{12}u_1 + (1-\lambda)u_2 + \dots + r_{1p}u_p &= 0 \\ \dots\dots\dots & \\ r_{1p}u_1 + r_{2p}u_2 + \dots + (1-\lambda)u_p &= 0 \end{aligned}$$

la condición de compatibilidad del sistema es que el determinante del sistema sea nulo:

$$\det \begin{pmatrix} 1-\lambda & r_{12} & \dots & r_{1p} \\ r_{12} & 1-\lambda & \dots & r_{2p} \\ \dots & \dots & \dots & \dots \\ r_{1p} & r_{2p} & \dots & r_{pp} \end{pmatrix} = 0$$

o bien:

$$\det(R-\lambda I)=0$$

siendo I la matriz identidad y R la matriz de correlaciones. Pero la ecuación anterior es la ecuación característica de la matriz R , por tanto los valores de λ para los que existen direcciones principales son los valores propios de R , y dichas direcciones principales son las dadas por los correspondientes vectores propios.

Si algún valor propio, λ_i , es raíz múltiple de la ecuación característica, entonces $\dim \text{Nuc}(A-\lambda_i) > 1$, habiendo arbitrariedad en la elección de la base de $\text{Nuc}(A-\lambda_i I)$, pero siempre pueden elegirse de manera que sean ortonormales. Estos factores explican evidentemente la misma inercia, dada por su valor propio.

Si algún valor propio fuese cero, entonces el factor correspondiente no presenta variabilidad, no aporta inercia a la inercia total de la nube de puntos o no explica ninguna parte de la inercia total, pudiendo prescindirse de dicho factor.

Hay que tener en cuenta que la matriz de correlaciones R es siempre diagonalizable por ser una matriz simétrica.

LA INERCIA DE LAS COMPONENTES PRINCIPALES

El problema consiste por tanto en diagonalizar la matriz de correlaciones R . Sean pues $\lambda_1, \lambda_2, \dots, \lambda_p$ los valores propios. Como la traza de R es invariante, se tiene:

$$\text{traza}R = \text{Inercia} = p = \sum_{i=1}^p \lambda_i$$

por tanto cada factor colabora a la inercia total en una cantidad igual a su valor propio.

Cada valor propio es la inercia de cada factor, siendo, en porcentaje, la inercia explicada por el factor u_k :

$$\frac{\lambda_k}{p} 100$$

mientras que la inercia explicada por los m ($< p$) primeros factores sería:

$$\frac{\sum_{i=1}^m \lambda_i}{p} 100$$

Evidentemente el factor de mayor valor propio será el que explique mayor inercia, y así sucesivamente. Por ese motivo es conveniente obtener los valores propios ordenados de mayor a menor:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$$

e incluso se puede dar una representación gráfica de su magnitud de forma individualizada, Figura 2, o de forma acumulada, Figura 3.

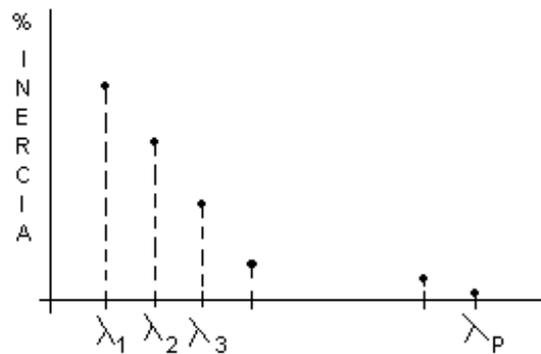


Figura 2

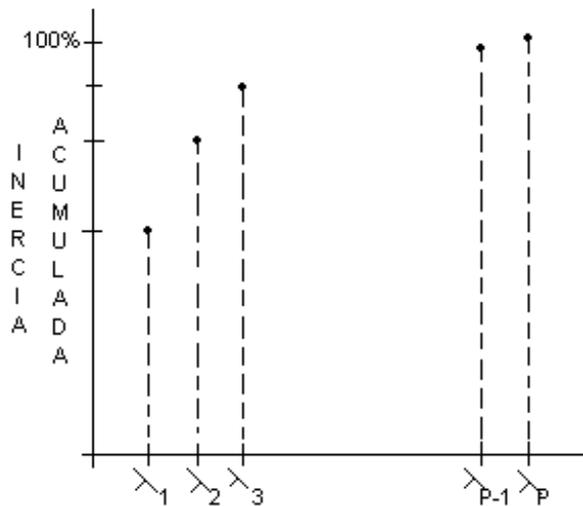


Figura 3

CALCULO DE LOS FACTORES PRINCIPALES. PROPIEDADES

Tal como venimos diciendo el problema consiste en diagonalizar la matriz R , a partir de lo cual obtendremos los vectores propios de R y sus correspondientes valores propios:

$$Ru = \lambda u$$

siendo u el vector propio asociado al valor propio λ .

Sea T la matriz cuyas columnas son las componentes de los vectores propios en la base inicial o matriz del cambio de base, y supongamos que la hemos elegido ortogonal, es decir: $T^{-1} = T^t$, y sea Λ la matriz diagonal de valores propios, entonces:

$$\Lambda = T^t RT$$

Si los vectores propios son:

$$\begin{aligned} u_1 &= (t_{11}, t_{21}, \dots, t_{p1}) \\ u_2 &= (t_{12}, t_{22}, \dots, t_{p2}) \\ &\dots \\ u_p &= (t_{1p}, t_{2p}, \dots, t_{pp}) \end{aligned}$$

entonces la matriz del cambio de base es:

$$T = \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1p} \\ t_{21} & t_{22} & \cdots & t_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ t_{p1} & t_{p2} & \cdots & t_{pp} \end{pmatrix}$$

En la base original, (e_1, \dots, e_p) , el individuo i -ésimo se expresa:

$$\overrightarrow{OA_i} = (z_{i1}, z_{i2}, \dots, z_{ip}) = z_{i1}e_1 + z_{i2}e_2 + \cdots + z_{ip}e_p$$

mientras que en las nuevas componentes será:

$$\begin{aligned} \overrightarrow{OA_i} &= (y_{i1}, y_{i2}, \dots, y_{ip}) = y_{i1}u_1 + y_{i2}u_2 + \cdots + y_{ip}u_p = \\ &= y_{i1}(t_{11}e_1 + \cdots + t_{p1}e_p) + \cdots + y_{ip}(t_{1p}e_1 + \cdots + t_{pp}e_p) = \\ &= (y_{i1}t_{11} + \cdots + y_{ip}t_{1p})e_1 + \cdots + (y_{i1}t_{p1} + \cdots + y_{ip}t_{pp})e_p \end{aligned}$$

por tanto:

$$\begin{aligned} z_{i1} &= y_{i1}t_{11} + y_{i2}t_{12} + \cdots + y_{ip}t_{1p} \\ z_{i2} &= y_{i1}t_{21} + y_{i2}t_{22} + \cdots + y_{ip}t_{2p} \\ &\dots\dots\dots \\ z_{ip} &= y_{i1}t_{p1} + y_{i2}t_{p2} + \cdots + y_{ip}t_{pp} \end{aligned}$$

que con la formulación matricial, será:

$$\begin{pmatrix} z_{i1} \\ z_{i2} \\ \vdots \\ z_{ip} \end{pmatrix} = T \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{ip} \end{pmatrix}$$

o bien, trasponiendo:

$$(z_{i1} \quad z_{i2} \quad \cdots \quad z_{ip}) = (y_{i1} \quad y_{i2} \quad \cdots \quad y_{ip})T^t$$

teniendo en cuenta que $T^{-1}=T^t$:

$$(y_{i1} \quad y_{i2} \quad \cdots \quad y_{ip}) = (z_{i1} \quad z_{i2} \quad \cdots \quad z_{ip})T$$

aplicando finalmente la relación anterior sobre cada individuo, $i=1,2,\dots,n$, se puede escribir:

$$Y = ZT$$

obteniéndose por tanto las coordenadas de los individuos en los nuevos ejes factoriales a partir de la matriz de datos tipificados.

Considerando las p variables: (Z_1, Z_2, \dots, Z_p) representativas de las columnas de la matriz de datos tipificados, y las p variables: (Y_1, Y_2, \dots, Y_p) representativas de las componentes principales, se tiene:

$$(Z_1, Z_2, \dots, Z_p) = (Y_1, Y_2, \dots, Y_p)T^t$$

o

$$(Y_1, Y_2, \dots, Y_p) = (Z_1, Z_2, \dots, Z_p)T$$

Teniendo en cuenta las expresiones anteriores, se cumple:

$$\bar{Y} = \bar{Z}T = 0$$

y

$$V_Y = T^t R T = \Lambda$$

por tanto los nuevos ejes principales están centrados, sus varianzas son los valores propios y sus covarianzas son nulas (están incorrelacionados). Es decir:

$$E(Y_i) = 0$$

$$\text{var}(Y_i) = \lambda_i = \frac{1}{n} \sum_{k=1}^n y_{ki}^2$$

$$\text{cov}(Y_i, Y_j) = 0 = \frac{1}{n} \sum_{k=1}^n y_{ki} y_{kj}$$

con $i, j=1, 2, \dots, p$.

Una propiedad importante es cuantificar el grado de interdependencia entre las variables originales y los nuevos ejes factoriales, pues ello nos permitirá interpretar las componentes principales. Para ello calculemos las covarianzas y correlaciones entre estos pares de variables:

$$\text{cov}(Y_j, Z_i) = \frac{1}{n} \sum_{k=1}^n y_{kj} z_{ki} = \frac{1}{n} \sum_{k=1}^n y_{kj} \sum_{r=1}^p y_{kr} t_{ir} = \sum_{r=1}^p t_{ir} \left(\frac{1}{n} \sum_{k=1}^n y_{kj} y_{kr} \right) = \sum_{r=1}^p t_{ir} \text{cov}(Y_j, Y_r)$$

pero teniendo en cuenta las relaciones anteriores resulta:

$$\text{cov}(Y_j, Z_i) = t_{ij} \lambda_j$$

y finalmente, dividiendo por las d.t.:

$$\rho_{ij} = \frac{t_{ij} \lambda_j}{\sqrt{\lambda_j \cdot 1}} = t_{ij} \sqrt{\lambda_j}$$

en forma de matriz:

$$\begin{pmatrix} t_{11}\sqrt{\lambda_1} & t_{12}\sqrt{\lambda_2} & \cdots & t_{1p}\sqrt{\lambda_p} \\ t_{21}\sqrt{\lambda_1} & t_{22}\sqrt{\lambda_2} & \cdots & t_{2p}\sqrt{\lambda_p} \\ \vdots & \vdots & \cdots & \vdots \\ t_{p1}\sqrt{\lambda_1} & t_{p2}\sqrt{\lambda_2} & \cdots & t_{pp}\sqrt{\lambda_p} \end{pmatrix}$$

las columnas se refieren a las variables Y_j , mientras que las filas a las variables Z_i , por tanto en la intersección de columna con fila tenemos el coeficiente de correlación lineal entre ambas variables. Notese que la correlación entre ambas es tanto mayor cuanto mayor sea el valor propio (varianza) del factor Y_i , ya que es proporcional a:

$$d.t.(Y_i) = \sqrt{\lambda_i}$$

la matriz anterior recibe el nombre de matriz de factores o matriz de cargas factoriales (obsérvese que ya no es una matriz simétrica)

Interesa por tanto aquellos factores de alta d.t. pues serán los que están más fuertemente correlacionados con las variables originales. Se podrá prescindir de aquellos factores escasamente correlacionados con las originales, pudiendo considerarse como variables independientes.

Si un factor principal está muy correlacionado con una variable o grupo de variables originales, entonces ese factor explica por si solo a esa variable o grupo de variables originales. En caso contrario, dicho factor actuaría de forma independiente con relación a dicha variable o grupo de variables.

La matriz de cargas factoriales, que representamos por F , se puede escribir de la siguiente manera:

$$F = T\Lambda^{1/2} = \Lambda^{1/2}T^t$$

como fácilmente se comprueba, siendo $\Lambda^{1/2}$ la matriz diagonal de las desviaciones típicas ($\sqrt{\lambda_i}$). Teniendo en cuenta lo anterior la matriz de correlaciones R se puede expresar así:

$$R = T\Lambda T^t = T\Lambda^{1/2}\Lambda^{1/2}T^t = FF^t$$

Tipifiquemos los factores principales (considerando lo anterior) y designemos a estos por F_1, \dots, F_p :

$$F_i = \frac{Y_i - E(Y_i)}{d.t.(Y_i)} = \frac{Y_i - 0}{\sqrt{\lambda_i}} = \frac{1}{\sqrt{\lambda_i}} Y_i$$

ahora $E(F_i) = 0$, $\text{var}(F_i) = 1$ y $\text{cov}(F_i, F_j) = 0$. Hemos operado un cambio de escala en los factores principales al tipificarlos. Los llamaremos factores o componentes principales escalados.

Fácilmente se comprueba que:

$$\begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{pmatrix} = F \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_p \end{pmatrix}$$

o

$$(Z_1, Z_2, \dots, Z_p) = (F_1, F_2, \dots, F_p) F^t$$

y en componentes:

$$Z_i = t_{i1} \sqrt{\lambda_1} F_1 + t_{i2} \sqrt{\lambda_2} F_2 + \dots + t_{ip} \sqrt{\lambda_p} F_p, \quad i = 1, 2, \dots, p$$

de aquí se sigue que:

$$\text{var}(Z_i) = t_{i1}^2 \lambda_1 \text{var}(F_1) + t_{i2}^2 \lambda_2 \text{var}(F_2) + \dots + t_{ip}^2 \lambda_p \text{var}(F_p)$$

$$1 = t_{i1}^2 \lambda_1 + t_{i2}^2 \lambda_2 + \dots + t_{ip}^2 \lambda_p$$

esta es una propiedad relativa a las filas de la matriz de cargas factoriales, F : las filas de la matriz F son vectores unitarios.

CRITERIOS DE REPARTO DE LA INERCIA TOTAL Y REDUCCIÓN DE LA DIMENSIÓN.

La inercia que explica cada factor es numéricamente igual a su valor propio, que expresada en porcentaje de la inercia total es:

$$\frac{\lambda_i}{p} 100$$

siendo λ_i el valor propio correspondiente Al factor propio Y_i .

Si los valores propios son $\lambda_1 > \dots > \lambda_p$, cada factor explica distinta proporción de la inercia total de manera que cuantos más factores retenemos mejor será la calidad de la representación, pero entonces no simplificamos el problema ya que retenemos todos los factores.

Es preciso por tanto definir un criterio para fijar el número de factores a retener. Existen varios criterios que enumeramos a continuación:

- i) retener aquellos factores cuyos valores propios sean mayores que 1.
 - ii) retener aquellos factores cuyos valores propios sean superiores a un valor fijo previamente fijado por el investigador.
 - iii) retener un número fijo de factores.
-

-
- iv) retener aquellos factores que expliquen una proporción predeterminada de la inercia total (por ejemplo, un 75% o más).

Si hemos retenido m factores el porcentaje de inercia explicada por dichos factores vale:

$$\frac{\sum_{i=1}^m \lambda_i}{p} 100$$

Al retener m factores, la parte de varianza de la variable Z_i contenida en dichos factores es:

$$h_i^2 = t_{i1}^2 \lambda_1 + t_{i2}^2 \lambda_2 + \dots + t_{im}^2 \lambda_m \leq 1$$

se llama comunalidad de la variable Z_i y se interpreta como una medida de la calidad de la representación de dicha variable Z_i por los m primeros factores. Evidentemente la representación será tanto mejor cuanto más se aproxime a la unidad

Con la reducción de la dimensión, los m primeros columnas de la matriz T , que forman la matriz T_m , engendran un nuevo subespacio de representación de la nube de puntos, siendo las nuevas coordenadas de los individuos:

$$Y_m = ZT_m$$

valiendo ahora la inercia:

$$\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^m y_{ik}^2 = \sum_{k=1}^m \left(\frac{1}{n} \sum_{i=1}^n y_{ik}^2 \right) = \sum_{k=1}^m \lambda_k$$

que es la parte de variabilidad explicada por dichos factores.

Si no hay correlación entre las p variables iniciales, entonces:

$$R = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

y los factores principales son las variables original y todas explican la misma cantidad de inercia, no siendo posible la reducción.

ROTACIÓN DE LAS COMPONENTES

Los factores principales obtenidos tal como hemos indicado anteriormente pueden tener interpretaciones difíciles de comprender, por lo que se procede a tomar otras soluciones para hacerlos más interpretables a base de rotar la solución inicial obtenida. Son las soluciones rotadas o factores rotados. Sólo vamos a considerar

rotaciones ortogonales, pues de esta manera los nuevos ejes siguen estando incorrelacionados entre sí, mantienen las comunanlidades (la capacidad conjunta de cada factor para retener la información de cada variable), sin embargo se altera las correlaciones entre factores y variables, Así como el porcentaje de inercia condensada por cada factor. Después de la rotación hay que calcular la nueva matriz de factores que contiene las correlaciones entre los factores rotados y las variables originales, que se obtiene multiplicando la matriz de factores obtenida antes de la rotación por la matriz de correlaciones entre los factores rotados y no rotados, que algunos llaman matriz de transformación de los factores.

Para efectuar una rotación ortogonal de ejes se aplican dos criterios, denominados rotación quartimax y rotación varimax.

La rotación quartimax o criterio quartimax tiene por objeto determinar la transformación ortogonal que transforma la matriz de factores en otra de manera que la varianza de los cuadrados de las cargas factoriales es máxima, recayendo el énfasis del método en la simplificación por filas.

La rotación varimax o criterio varimax hace énfasis en la simplificación de las columnas o factores de la matriz de factores (Kaiser, 1958) con el fin de satisfacer la sencillez de interpretación, maximizándose suma de varianzas de los cuadrados de las cargas factoriales de cada factor.
