ANÁLISIS DE COMPONENTES PRINCIPALES

El reconocimiento de patrones desde un punto de vista estadístico es la selección o extracción de características.

Esta selección se refiere a procesos donde el espacio de los datos es transformado a un espacio de características que en teoría conserva la misma dimensión del espacio de los datos originales.

Sin embargo, esta transformación se puede representar mediante un conjunto *efectivo y reducido* de características reteniendo la mayor cantidad de información contenida internamente en los datos originales. Es decir, hay una *reducción de dimensionalidad* del espacio de los datos originales. Esto es básicamente el propósito de esta técnica.

Es quizás la más vieja y mejor conocida de las técnicas multivariantes.

Fue iniciada por Pearson (1901) en el área de la biología y fue desarrollada por Hotelling (1903) en el área de la psicometría.

La varianza total juega un papel importante en esta técnica. Mayor varianza presente en las características reducidas, mayor información retenida de los datos originales.

Los datos originales están organizados en observaciones con un conjunto de variables.

GEOMETRÍA

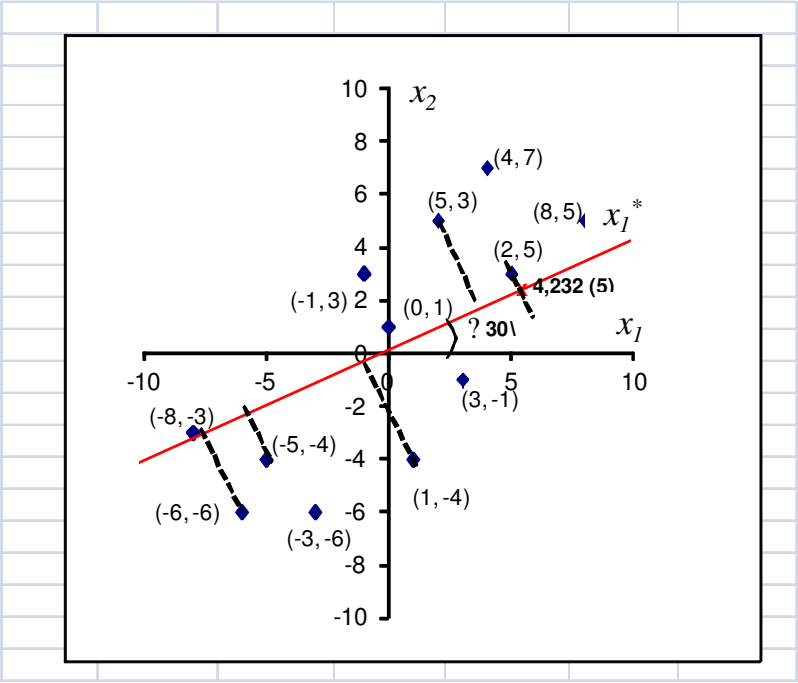
Siguiendo el ejemplo de la tabla de Subhash en el Capítulo 4, se puede obtener una idea mediante un ejemplo básico con dos variables solamente y 12 observaciones que serán movidas desde su plano original hasta un plano donde se maximice su varianza. Para ello lo resolveremos en varios pasos.

Se prepara el conjunto de datos, en este caso un conjunto de doce observaciones de las variables x_1 y x_2. Se estandarizan a una distribución normal *N(0,1)* con media en cero y varianza a la unidad, según la tabla más abajo. Se representan gráficamente en un par de ejes x_1 y x_2.

		x_1			x_2	
#	Datos Originales	Media Cero	Datos Estandarizados	Datos Originales	Media Cero	Datos Estandarizados
1	16	8	1,739	8	5	1,137
2	12	4	0,869	10	7	1,592
3	13	5	1,087	6	3	0,682
4	11	3	0,652	2	-1	-0,227
5	10	2	0,435	8	5	1,137
6	9	1	0,217	-1	-4	-0,910
7	8	0	0,000	4	1	0,227
8	7	-1	-0,217	6	3	0,682
9	5	-3	-0,652	-3	-6	-1,365
10	3	-5	-1,087	-1	-4	-0,910
11	2	-6	-1,304	-3	-6	-1,365
12	0	-8	-1,739	0	-3	-0,682
Media	8	0	0	3	0	0
Varianza	23,09	23,09	1	21,09	21,09	1
Desv. Est.	4,601	4,601	1	4,397	4,4	1

En la tabla anterior se observa que la varianza total de ambas variables (x_1 y x_2) es de 44,18, aportando el 52,2% la variable x_1 y el resto (47,8%) la variable x_2.

De igual modo asumimos un eje arbitrario de rotación θ (30°) para una de las variables. Por ejemplo x_1^*.



Así, el ángulo entre x_1 y x_1^* es π/6 (30°), entonces la proyección de la observación 5, de medias (2,5) en este nuevo eje sería 4,232 y de igual modo, el resto de los pares de medias de (x_1,x_2) tendrán su propia proyección en x_1^* de acuerdo a la función de proyección dada por $x_1^* = x_1 \cos\theta + x_2 sen\theta$. De este modo, quedaría conformado por

θ	#	1	2	3	4	5	6	7	8	9	10	11	12
30°	x_1^*	9,428	6,964	5,830	2,098	4,232	-1,134	0,500	0,634	-5,598	-6,330	-8,196	-8,428

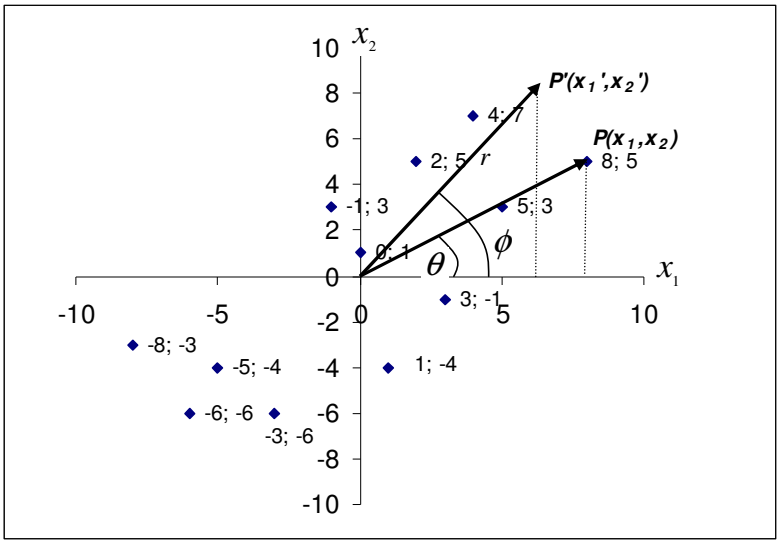
Para el otro eje, x_2^*, se tienen las siguientes proyecciones para el conjunto total de observaciones.

θ	#	1	2	3	4	5	6	7	8	9	10	11	12
30°	x_2^*	0,330	4,062	0,098	-2,366	3,330	-3,964	0,866	3,098	-3,696	-0,964	-2,196	1,402

Esta expresión se desprende del siguiente procedimiento:

Si por ejemplo *r* denota la longitud del vector definido entre el punto *P* y el origen y donde los valores de x_1, x_2, x_1' y x_2' viene dados por

$$x_1 = r\cos\theta,\ x_2 = r\text{sen}\theta \text{ y } x'_1 = r\cos(\theta+\varphi),\ x'_2 = r\text{sen}(\theta+\varphi)$$



En términos generales, para un espacio de orden *p*, se puede definir un nuevo conjunto ortogonal de vectores x_i tal que:

- Las coordenadas de las observaciones con respecto a cada uno de los ejes da los valores para las nuevas variables, los nuevos ejes o las variables se llaman *componentes principales* y los valores correspondientes son el score de los componentes.
- Cada nueva variable resulta entonces ser una combinación lineal de las variables originales.
- La primera nueva variable contabiliza el mayor aporte de varianza en los datos. La segunda contabiliza el segundo mayor aporte de varianza y así sucesivamente para las *p-2* variables restantes del espacio de orden *p*.
- Las nuevas son totalmente independientes. Es decir no están correlacionadas.

ENFOQUE ANALÍTICO

Sea **X**, una matriz de orden *p*-dimensional conformada por *p* variables con N observaciones para cada variable.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & . & . & x_{1p} \\ x_{21} & x_{22} & x_{23} & . & . & x_{2p} \\ x_{31} & x_{32} & x_{33} & . & . & x_{3p} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ x_{N1} & x_{N2} & x_{N3} & . & . & x_{Np} \end{bmatrix}.$$

Mediante Componentes Principales (ACP), el conjunto total de *p* variables podría ser reducido a un nuevo conjunto *enteramente independiente* de nuevas variables (algunas veces conocidas como variables latentes) expresada en un matriz resultante Z de orden *k*-dimensional.

$$\mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} & z_{13} & . & . & z_{1k} \\ z_{21} & z_{22} & z_{23} & . & . & z_{2k} \\ z_{31} & z_{32} & z_{33} & . & . & z_{3k} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ z_{N1} & z_{N2} & z_{N3} & . & . & z_{Nk} \end{bmatrix}$$

El nuevo conjunto de variables define un espacio *k*-dimensional mucho más reducido que el original donde $k<<p$ del cual se le puede hacer algunas consideraciones:

- Los valores entre sí de las nuevas variables presentes en la matriz **Z** no están correlacionados. Es decir *las nuevas variables son completamente independientes*.
- A pesar de que la dimensión de la matriz **Z** es igual a la de la matriz **X**, su utilidad práctica preferiblemente se reduce a una matriz de dimensión mucho menor ($k<<p$).

- La entrada principal al ACP es la matriz de covarianza. Sin embargo, las variables originales podrían ser estandarizadas (media cero y varianza uno) para eliminar el efecto de la varianza relativa de las variables originales. En este caso, se puede sustituir como entrada a la matriz de correlación por la matriz de covarianza. Esta sustitución es útil para eliminar las altas varianzas generadas por las variables involucradas con diferente escala en las unidades de medida.
- La traza de la matriz de correlación es igual a la varianza total de las variables transformadas.
- Los autovectores de la matriz de covarianza o de correlación definen los nuevos ejes en el espacio *k*-dimensional.
- La cantidad de componentes principales es igual a la cantidad de variables originales consideradas en el ACP. Es decir, son *p* componentes.
- Cada z_j es una variable transformada de las x_i(s) variables originales. Siendo *i=1,2,3….,p.* Estas nuevas variables contienen los valores de las variables transformadas.
- ACP captura solamente linealidad entre las variables. Por eso se le conoce como una técnica que transforma las variables originales mediante un método de combinación lineal. Es decir las z_i son la combinación lineal de los componentes y las variables originales x_i.
- Los primeros componentes obtenidos mediante ACP explican la mayor cantidad de la varianza total de las variables originales. Es decir, agrupa la mayor cantidad de información que puedan suministrar las variables originales.

Habiendo definido **X** y **Z**, con *p* variables (originales y transformadas) y *N* observaciones, consideremos adicionalmente la matriz de covarianza $\sum$ de la matriz **X**; entonces ACP permite calcular mediante la matriz $\sum$ un nuevo conjunto de *p* variables no correlacionadas (*z*) tal que ellas sean combinación lineal de las variables originales.

De este modo, para cualquier observación *i* en **X** dado por $\lfloor x_1 \quad x_2 \quad x_3 \quad . \quad . \quad x_p \rfloor$, existe una función lineal $\mathbf{Z = a'_i x,}$ para todo *i=1…,N*. Es decir,

$z_j = a_{i1}x_1 + a_{i2}x_2 + a_{i3}x_3 + ... + a_{ip}x_p$, *j=1,…,P, i=1,…N*, donde:

- a'_i es un vector transpuesto de pesos o parámetros del i-ésimo componente principal.
- $\mathbf{a}_j$ es un vector de pesos o parámetros del el *j-ésimo* componente principal o *autovector* de $\sum$ para formar la combinación lineal de las *p* variables originales.
- $\mathbf{a}_j$ tiene que ser ortogonal y ortonormal. Es decir $a'_i a_j = 0$ y $a'_i a_i = 1$.
- z_j se corresponde con los valores nuevos de las variables originales.
- La varianza del *j-ésimo* componente es
  $V(z_j) = var(\mathbf{a'}_i\mathbf{x}) = \mathbf{a'}_i\sum \mathbf{a'}_i$, para todo *j=1,…,P, i=1,…N.*
- Las varianzas resultantes para cada componente son decrecientes y en estricto orden: $V(z_1) > V(z_2) > V(z_3) >\ldots V(z_p)$. Su suma representa la variación total de las variables originales.

En general, el objetivo es entonces el de encontrar los componentes principales $\mathbf{a}_j$ tal que $\mathbf{a'}_j\sum \mathbf{a}_j$ es un máximo sujeto a que $\mathbf{a'}_j\mathbf{a}_j = 1$, donde *j=1,…P*. Usando Lagrange para el *j-ésimo* componente, se tiene:

$$L = \mathbf{a'}_j\sum \mathbf{a}_j - \lambda_j(\mathbf{a'}_j\mathbf{a}_j - 1),$$

λ_j es el multiplicador y la parcial con respecto al componente es

$$\frac{\partial L}{\partial \mathbf{a}_j} = 2\Sigma\mathbf{a}_j - 2\lambda_j\mathbf{a}_j.$$