

Tema 5. Análisis de componentes principales

5.1. Introducción.

El análisis de componentes principales se concibe como una técnica de reducción de la dimensión, pues permite pasar de una gran cantidad de variables interrelacionadas a unas pocas componentes principales. El método consiste en buscar combinaciones lineales de las variables originales que representen lo mejor posible a la variabilidad presente en los datos. De este modo, con unas pocas combinaciones lineales, que serán las componentes principales, sería suficiente para entender la información contenida en los datos. Al mismo tiempo, la forma en que se construyen las componentes, y su relación con unas u otras variables originales, sirven para entender la estructura de correlación inherente a los datos. Por último, las componentes principales, que forman un vector aleatorio de dimensión menor, pueden ser empleadas en análisis estadísticos posteriores.

5.2. Descomposición de un vector aleatorio en sus componentes principales.

En esta sección se va a definir el concepto de componentes principales, y cómo se obtienen a partir de la matriz de covarianzas de un vector aleatorio. Estas ideas y procedimientos se aplican de igual modo tanto a un vector aleatorio como a un conjunto de datos. En este último caso, se puede pensar que el análisis se está aplicando a una distribución de probabilidad discreta equiprobable sobre los vectores observados. Por este motivo, esta sección se desarrolla sobre un vector aleatorio general, con la única condición de que exista la matriz de covarianzas.

Definición Sea $X = (X_1, \dots, X_p)'$ un vector aleatorio p -dimensional con vector de medias $\mu = E(X)$ y matriz de covarianzas $\Sigma = D(X) = Cov(X, X)$. Se define la primera componente principal de X como una variable aleatoria Z_1 tal que

$$Z_1 = v_1'X = v_{11}X_1 + \dots + v_{p1}X_p \quad \text{con} \quad v_1 = (v_{11}, \dots, v_{p1})' \in \mathbb{R}^p$$

$$Var(Z_1) = \max\{Var(v'X) : v \in \mathbb{R}^p, v'v = 1\}$$

La primera componente principal es una combinación lineal normalizada de las variables de X y, de entre todas las combinaciones lineales normalizadas, es la que tiene mayor varianza.

Nota. Debe observarse que $Var(\alpha v'X) = \alpha^2 Var(v'X)$ con $\alpha \in \mathbb{R}$. De modo que multiplicando por una constante podemos hacer la varianza tan grande o tan pequeña como queramos. Por ello es necesario normalizar, pidiendo que $\|v\| = \sqrt{v'v} = 1$.

Teorema 5.1 La primera componente principal de X adopta la forma

$$Z_1 = v_1'X$$

siendo λ_1 el mayor autovalor de $\Sigma = D(X)$ y v_1 un autovector asociado a λ_1 de norma uno ($v_1'v_1 = 1$). Además

$$\text{Var}(Z_1) = \lambda_1$$

DEMOSTRACIÓN

Consideremos $Z = v'X$. Entonces podemos calcular su varianza

$$\text{Var}(Z) = \text{Cov}(v'X, v'X) = v'\text{Cov}(X, X)v = v'\Sigma v$$

Nuestro problema consiste en

$$\begin{array}{ll} \text{Max} & v'\Sigma v \\ \text{sujeto a} & v'v = 1 \end{array}$$

Resolvemos este problema por el método de los multiplicadores de Lagrange.

$$L = v'\Sigma v - \lambda(v'v - 1)$$

Calculamos la matriz jacobiana de L como función de v e igualamos a cero

$$\frac{\partial L}{\partial v} = 2v'\Sigma - 2\lambda v' = 0$$

Equivalentemente

$$\Sigma v = \lambda v$$

Luego el vector v_1 que maximice la función sujeto a la restricción ha de ser un autovector de Σ , y su autovalor asociado es λ_1 . Multiplicando los dos términos de la ecuación por la izquierda por v_1' resulta

$$v_1'\Sigma v_1 = v_1'\lambda_1 v_1 = \lambda_1 v_1'v_1 = \lambda_1$$

Luego el autovalor λ_1 es la función objetivo en el máximo y por tanto, $\text{Var}(Z_1) = \lambda_1$.

Definición Se define la segunda componente principal de X como una variable aleatoria Z_2 tal que

$$Z_2 = v_2'X = v_{12}X_1 + \cdots + v_{p2}X_p \quad \text{con} \quad v_2 = (v_{12}, \dots, v_{p2})' \in \mathbb{R}^p$$

$$\text{Var}(Z_2) = \max\{\text{Var}(v'X) : v \in \mathbb{R}^p, v'v = 1, v'v_1 = 0\}$$

La segunda componente principal es otra combinación lineal de las variables de X y, de entre todas las combinaciones lineales formadas por vectores unitarios ortogonales a v_1 , es la que tiene mayor varianza.

Nótese que

$$\text{Cov}(v'X, Z_1) = \text{Cov}(v'X, v_1'X) = v'\text{Cov}(X, X)v_1 = v'\Sigma v_1$$

Enunciamos a continuación un resultado sencillo

Resultado. Sea $w \in \mathbb{R}^p$ un autovector de Σ asociado a un autovalor no nulo $\lambda \neq 0$, y sea $v \in \mathbb{R}^p$. Entonces

$$v'\Sigma w = 0 \iff v'w = 0$$

DEMOSTRACIÓN.

” \Leftarrow ”

$$v'\Sigma w = v'\lambda w = \lambda v'w = 0$$

” \Rightarrow ”

$$v'w = v'\frac{1}{\lambda}\Sigma w = \frac{1}{\lambda}v'\Sigma w = \frac{1}{\lambda}0 = 0$$

Por tanto, en aplicación de este resultado

$$Cov(v'X, Z_1) = 0 \iff v'v_1 = 0$$

Y hemos probado que la condición de ortogonalidad entre los vectores v_1 y v_2 es equivalente a la incorrelación de las componentes Z_1 y Z_2 . En definitiva, podríamos definir la segunda componente principal como la combinación lineal de las variables de X que tiene mayor varianza de entre las combinaciones lineales normalizadas e incorrelacionadas con la primera componente principal.

Respecto a la condición de que $\lambda_1 \neq 0$, hemos de recordar que $\lambda_1 = Var(Z_1)$. Por tanto si $\lambda_1 = Var(Z_1) = 0$, todos los demás autovalores serán también cero y Z_1 y las demás componentes serán variables degeneradas.

Teorema 5.2 *La segunda componente principal de X adopta la forma*

$$Z_2 = v'_2 X$$

siendo λ_2 el segundo mayor autovalor de $\Sigma = D(X)$ y v_2 un autovector asociado a λ_2 de norma uno ($v'_2 v_2 = 1$) y ortogonal a v_1 ($v'_1 v_2 = 0$). Además

$$Var(Z_2) = \lambda_2$$

DEMOSTRACIÓN

Procederemos de modo muy similar al caso anterior. Consideramos $Z = v'X$, cuya varianza es $Var(Z) = v'\Sigma v$. Nuestro problema consiste en

$$\begin{array}{ll} \text{Max} & v'\Sigma v \\ \text{sujeto a} & v'v = 1 \\ & v'_1 v = 0 \end{array}$$

Resolvemos este problema por el método de los multiplicadores de Lagrange.

$$L = v'\Sigma v - \lambda_2(v'v - 1) - \mu_2 v'_1 v$$

Derivamos respecto de v e igualamos a cero (momento en el cual podemos sustituir el valor v_2 para mejor comprensión):

$$\left(\frac{\partial L}{\partial v} /_{v=v_2} \right)' = 2\Sigma v_2 - 2\lambda_2 v_2 - \mu_2 v_1 = 0 \quad (5.1)$$

Multiplicando por la izquierda por v_1' resulta

$$2v_1'\Sigma v_2 - 2\lambda_2 v_1'v_2 - \mu_2 v_1'v_1 = 0$$

Las restricciones imponen que $v_1'v_2 = 0$. Por la nota, de ello deducimos que $v_1'\Sigma v_2 = 0$. Y v_1 ya verificaba $v_1'v_1 = 1$. Entonces la ecuación anterior se reduce a $\mu_2 = 0$. Y al sustituir ese valor en la ecuación (5.1) se obtiene

$$\Sigma v_2 = \lambda_2 v_2$$

Razonando igual que antes, el vector v_2 que maximice la función sujeto a las restricciones ha de ser un autovector de Σ , y su autovalor asociado es λ_2 . La restricción de ortogonalidad respecto de v_1 nos obliga a tomar el segundo mayor autovalor de Σ . Como antes, $\text{Var}(Z_2) = v_2'\Sigma v_2 = \lambda_2$.

Podríamos continuar este proceso extrayendo las componentes principales de X mediante los autovalores de la matriz de covarianzas Σ y la base ortonormal de autovectores asociados. De este modo, obtenemos p componentes principales.

Definición Se definen las p componentes principales de X como las variables aleatorias (Z_1, \dots, Z_p) tales que

$$\begin{aligned} Z_1 &= v_1'X, \dots, Z_p = v_p'X & v_1, \dots, v_p &\in \mathbb{R}^p \\ \text{Var}(Z_1) &= \max\{\text{Var}(v'X) : v \in \mathbb{R}^p, v'v = 1\} \\ \text{Var}(Z_2) &= \max\{\text{Var}(v'X) : v \in \mathbb{R}^p, v'v = 1, v_1'v = 0\} \\ &\vdots & \vdots \\ \text{Var}(Z_j) &= \max\{\text{Var}(v'X) : v \in \mathbb{R}^p, v'v = 1, v_1'v = 0, \dots, v_{j-1}'v = 0\} \\ &\vdots & \vdots \\ \text{Var}(Z_p) &= \max\{\text{Var}(v'X) : v \in \mathbb{R}^p, v'v = 1, v_1'v = 0, \dots, v_{p-1}'v = 0\} \end{aligned}$$

Teorema 5.3 Las p componentes principales de X adoptan la forma

$$Z_j = v_j'X \quad j \in \{1, \dots, p\}$$

siendo $\lambda_1 \geq \dots \geq \lambda_p \geq 0$ los p autovalores ordenados de $\Sigma = D(X)$ y v_1, \dots, v_p sus autovectores asociados normalizados, esto es, $\{v_1, \dots, v_p\}$ es una base ortonormal de autovectores. Además las componentes son incorreladas $\text{Cov}(Z_j, Z_k) = 0$ si $j \neq k$ y

$$\text{Var}(Z_j) = \lambda_j \quad j \in \{1, \dots, p\}$$

Este teorema nos permite expresar

$$Z = V'X$$

siendo $V = (v_1, \dots, v_p)$ la matriz cuyas columnas son los autovectores de Σ . Entonces

$$\text{Cov}(Z, Z) = \text{Cov}(V'X, V'X) = V'\text{Cov}(X, X)V = V'\Sigma V$$

Pero la matriz de covarianzas de las componentes principales es diagonal y contiene los autovalores de Σ en la diagonal. Luego, el proceso de extracción de las componentes principales se reduce a la diagonalización de la matriz de covarianzas del vector aleatorio X .

Ejemplo 5.1 Sea $X = (X_1, X_2)'$ un vector aleatorio con vector de medias $\mu = E(X) = (4, 2)'$ y matriz de covarianzas

$$\Sigma = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

Obtener la descomposición del vector X en sus componentes principales.

Ejemplo 5.2 Se ha examinado a 25 alumnos, aspirantes a ingresar en la Facultad de Matemáticas, de 5 materias diferentes: Geometría Diferencial (cuyo resultado se almacena en la variable *geodif*), Análisis Complejo (*ancompl*), Álgebra (*alg*), Análisis Real (*anreal*) y Estadística (*estad*). Las puntuaciones obtenidas figuran en la tabla siguiente:

<i>geodif</i>	<i>ancompl</i>	<i>alg</i>	<i>anreal</i>	<i>estad</i>
36	58	43	36	37
62	54	50	46	52
31	42	41	40	29
76	78	69	66	81
46	56	52	56	40
12	42	38	38	28
39	46	51	54	41
30	51	54	52	32
22	32	43	28	22
9	40	47	30	24
32	49	54	37	52
40	62	51	40	49
64	75	70	66	63
36	38	58	62	62
24	46	44	55	49
50	50	54	52	51
42	42	52	38	50
2	35	32	22	16
56	53	42	40	32
59	72	70	66	62
28	50	50	42	63
19	46	49	40	30
36	56	56	54	52
54	57	59	62	58
14	35	38	29	20

El objetivo de este estudio es obtener un ranking global de alumnos para la entrada en la Facultad de Matemáticas, a través de una puntuación global, extraída como cierta combinación lineal de las calificaciones en las cinco materias examinadas.

Análisis de componentes principales en lenguaje R. El comando básico para realizar un análisis de componentes principales en R es *princomp*. También se pueden obtener resultados similares con el comando *prcomp*. La diferencia principal entre uno y otro es que *princomp* diagonaliza la matriz S , mientras que *prcomp* diagonaliza la matriz S_c . Esto modifica los autovalores

en la proporción $n/(n-1)$, pero no supone ningún cambio en los autovectores. Hemos optado por el comando *princomp*. A continuación se muestran las instrucciones que permiten realizar un análisis de componentes principales para el ejemplo, utilizando R.

```

dat=read.table("ejemplo5.2.txt",header=TRUE) # Lectura de los datos
test.pca=princomp(dat) # Comando R que ejecuta el análisis de componentes principales
test.pca # Salida básica del objeto, que muestra las desviaciones típicas de las componentes,
# que son las raíces cuadradas de los autovalores
summary(test.pca) # A las desviaciones típicas de las componentes les añade
# la proporción de varianza explicada y sus valores acumulados
test.pca$sdev**2 # De este modo podemos obtener las varianzas de las componentes,
# que son los autovalores
screplot(test.pca) # Representación de los autovalores (gráfico de sedimentación)
windows()
screplot(test.pca,type="lines") # El mismo gráfico con líneas
loadings(test.pca) # Proporciona los coeficientes de las componentes (autovectores)
windows()
barplot(loadings(test.pca),beside=TRUE) # Representación de los coeficientes
test.pca$scores # Puntuaciones de los individuos en las componentes
windows()
biplot(test.pca) # Comando que genera el biplot
# Anexo: Cálculos directos
n=nrow(dat)
s=cov(dat)*(n-1)/n
auto=eigen(s)
lambda=auto$values
windows()
plot(lambda,type="l")
v=auto$vectors

```

5.3. Propiedades geométricas de las componentes principales

Esta sección contiene dos propiedades geométricas interesantes, que cumplen las componentes principales, cuando se comparan con el modelo de regresión. En primer lugar veremos que las componentes principales también son aproximaciones mínimo-cuadráticas del vector aleatorio, y en segundo lugar veremos que en el caso de un vector bivalente (X_1, X_2) , la primera componente principal describe una recta que pasa entre las dos rectas de regresión lineal simple, una de X_2 sobre X_1 y la otra de X_1 sobre X_2 .

Las componentes principales como mejores aproximaciones mínimo-cuadráticas

Recordemos que la componente principal es la variable aleatoria unidimensional $Z_1 = v_1'X$. Al proyectar el vector X sobre la recta que pasa por el origen con la dirección v_1 , se obtiene el vector aleatorio

$$Z_1^m = v_1 Z_1 = v_1 v_1' X$$

que coincide con la representación de la primera componente Z_1 en \mathbb{R}^p a lo largo de la dirección v_1 .

Observamos que este vector aleatorio está contenido en un espacio de dimensión uno. De hecho, su matriz de covarianzas es

$$\text{Cov}(Z_1^m, Z_1^m) = \text{Cov}(v_1 Z_1, v_1 Z_1) = v_1 \text{Cov}(Z_1, Z_1) v_1' = \text{Var}(Z_1) v_1 v_1' = \lambda_1 v_1 v_1'$$

la cual claramente es una matriz de rango uno y no es más que el primer sumando de la descomposición espectral de Σ .

Centramos el vector Z_1^m en el vector de medias de X , $\mu = E(X)$.

$$Z_1^{mc} = Z_1^m - E(Z_1^m) + E(X) = \mu + v_1 Z_1^c$$

siendo $Z_1^c = v_1'(X - \mu)$ la primera componente centrada.

Este vector aleatorio Z_1^{mc} es el resultado de proyectar el vector X sobre la recta que pasa por el vector de medias μ y tiene dirección v_1 .

Teorema 5.4 *La recta que pasa por el vector de medias $\mu = E(X)$ y tiene dirección v_1 es la que hace mínima la distancia cuadrática del vector aleatorio X a la recta, definiendo esta distancia cuadrática como la media del cuadrado de la distancia euclídea usual (distancia ortogonal o menor distancia) del vector aleatorio X a la recta.*

DEMOSTRACIÓN. Omitiremos la demostración de este teorema.

Observación. Este resultado debe relacionarse con la recta de regresión, que hace mínima la media (suma en el caso de una muestra) de la distancia vertical a la recta. En este caso consideramos la distancia ortogonal (la menor distancia), pues no hay variable respuesta ni variable explicativa, sino que todas las variables son tratadas de igual modo.

Posición relativa en el plano de la primera componente principal respecto a las rectas de regresión

Resultado. Sea (X, Y) un vector aleatorio en el plano. Al representar en el plano XY la primera componente principal, junto con las dos rectas de regresión de Y sobre X y de X sobre Y , obtenemos tres rectas que pasan por el vector de medias, y la primera componente principal forma una recta "menos horizontal" que la recta de regresión de Y sobre X , pero "más horizontal" que la recta de regresión de X sobre Y . (Se omite la demostración de este resultado).

Ejemplo 5.3 *Se ha obtenido la muestra siguiente de un vector aleatorio (X, Y) : $(-3, 7)$, $(1, 4)$, $(0, 6)$, $(3, 5)$, $(1, 9)$, $(7, 7)$, $(6, 9)$, $(8, 10)$, $(7, 12)$, $(10, 11)$.*

1. Realizar un análisis de componentes principales sobre esta muestra.
2. Representar un diagrama de dispersión, junto con la primera componente principal y las dos rectas de regresión.

A continuación incluimos el código en lenguaje R que permite realizar las tareas propuestas en este ejemplo.

```
x=c(-3,1,0,3,1,7,6,8,7,10) # Introducción de los datos
y=c(7,4,6,5,9,7,9,10,12,11)
plot(x,y) # Diagrama de dispersión de los datos
mx=mean(x) # Media de x
my=mean(y) # Media de y
points(mx,my,pch=18) # Representación del vector de medias
S=cov(cbind(x,y)) # Cálculo de la matriz de covarianzas
byx=S[1,2]/S[1,1] # Pendiente de la recta de regresión de Y sobre X
bxy=S[2,2]/S[1,2] # Pendiente de la recta de regresión de X sobre Y
ex=c(-2,10)
ryx=my+byx*(ex-mx)
lines(ex,ryx,col="blue") # Representación de la recta de regresión de Y sobre X
rxy=my+bxy*(ex-mx)
lines(ex,rxy,col="red") # Representación de la recta de regresión de X sobre Y
v=eigen(S)$vectors
bv=v[2,1]/v[1,1] # Pendiente de la primera componente principal
rv=my+bv*(ex-mx)
lines(ex,rv,col="green") # Representación de la primera componente principal
```

5.4. Descomposición de la variabilidad, proporción de varianza explicada y criterios para la reducción de dimensión

Según acabamos de ver, desde un punto de vista matricial podemos contemplar el proceso de extracción de las componentes principales como la diagonalización de la matriz de covarianzas

$$\Lambda = V'\Sigma V$$

siendo $\Sigma = Cov(X, X)$ la matriz de covarianzas del vector aleatorio original, $\Lambda = Cov(Z, Z)$ la matriz de covarianzas de las componentes (que es diagonal) y $V = (v_1 \dots, v_p)$ la matriz ortogonal cuyas columnas son los autovectores, que a su vez constituyen los coeficientes que definen las componentes.

Multiplicando por V por la izquierda y por V' por la derecha, podemos expresar

$$\Sigma = V\Lambda V' = \sum_{j=1}^p \lambda_j v_j v_j'$$

que se conoce como la descomposición espectral de la matriz Σ .

Sabemos que

$$rango(\Sigma) = rango(\Lambda) = \text{"número de autovalores no nulos"}$$

A su vez también recordamos que el rango de la matriz de covarianzas coincide con la dimensión del menor espacio lineal que contiene al vector aleatorio (con probabilidad uno).

Así si el rango es k , el vector aleatorio estaría contenido en un espacio lineal de dimensión k , los k primeros autovalores son no nulos y los $(p - k)$ últimos son cero, y las $(p - k)$ últimas componentes tienen varianzas cero. Se puede por tanto reducir la dimensión hasta k sin perder variabilidad. Sólo estaríamos cambiando el sistema de coordenadas.

Consideremos como medida global de variabilidad del vector aleatorio, la traza de su matriz de covarianzas. Nótese que no es más que la suma de las varianzas de las variables del vector, pero además tiene esta justificación como medida global de variabilidad

$$traza(\Sigma) = E(\|X - E(X)\|^2)$$

Es la media de las desviaciones cuadráticas respecto a la media y por tanto es una extensión natural de la varianza de una variable aleatoria.

De la descomposición espectral se deduce

$$traza(\Sigma) = \sum_{j=1}^p \lambda_j traza(v_j v_j') = \sum_{j=1}^p \lambda_j$$

La última igualdad es consecuencia de que los vectores v_j son de norma uno.

Si extraemos una componente principal, obtenemos la variable aleatoria unidimensional

$$Z_1 = v_1' X$$

Si en lugar de recoger todo el vector aleatorio, sólo aportamos Z_1 reduciendo la información a una variable unidimensional; junto con la simplificación, se produce una reducción de variabilidad, que pasa a ser

$$Var(Z_1) = \lambda_1$$

Decimos que el cociente

$$\frac{\lambda_1}{\lambda_1 + \dots + \lambda_p}$$

es la proporción de variabilidad explicada por la primera componente principal.

Si en lugar de una única componente principal extraemos r componentes resulta

$$\frac{\lambda_1 + \dots + \lambda_r}{\lambda_1 + \dots + \lambda_r + \lambda_{r+1} + \dots + \lambda_p}$$

es la proporción de variabilidad explicada por las r primeras componentes principales.

Debemos decidir entre la simplificación que supone la reducción de la dimensión y la pérdida de información resultante de la variabilidad no explicada. Como criterios para tomar esta decisión, se suelen emplear los siguientes:

- Criterio de la varianza explicada. Consiste en retener el número de componentes que conjuntamente expliquen una proporción de varianza establecida, habitualmente un 90% o 95% del total.
- Criterio del valor propio (Kaiser). Retener sólo las componentes cuyos autovalores sean mayores que la media, esto es, mayores que $\text{traza}(\Sigma)/p$. Estas componentes tendrían más varianza que la que corresponde en media a una variable, mientras que las demás componentes tienen menos.
- Gráfico de sedimentación (*screeplot*). Representar en un gráfico los valores propios $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ en orden decreciente, y buscar un "codo" en el gráfico, entendiendo por codo un punto a partir del cual los valores propios son claramente más pequeños que los anteriores, y muy similares entre sí.
- Retener un número preestablecido de componentes principales. Por ejemplo, es costumbre retener dos componentes, pues se pueden representar fácilmente en el plano.

5.5. Las componentes principales y los cambios de escala

Un problema importante del análisis de componentes principales es que sus resultados dependen de la escala de medida de las variables originales, X . Así, si se cambia de escala una variable (dejando las demás fijas) las componentes principales se modifican, cambiando incluso la posible interpretación de las mismas. En concreto, si se aumenta la escala de una variable, ésta verá incrementada su varianza y su aportación proporcional a la varianza total, atrayendo de este modo a la primera componente principal, que se asemejará a esta variable.

Claro está que si se realiza el mismo cambio de escala en todas las variables, entonces los resultados del análisis de componentes principales se mantendrán idénticos.

Este problema se puede solventar de dos maneras:

- Midiendo todas las variables en la misma escala (siempre que sean de la misma naturaleza).
- Aplicando el análisis de componentes principales a las variables estandarizadas. Esto último equivale a trabajar con la matriz de correlaciones, en lugar de la matriz de covarianzas.

5.6. Interpretación de las componentes principales

Para interpretar las componentes principales, Z , es preciso relacionarlas con las variables originales, X . El instrumento natural para establecer esta relación consiste en calcular la covarianza o la correlación entre las componentes y las variables originales X . Lo obtenemos a continuación.

$$\text{Cov}(X, Z_k) = \text{Cov}(X, v'_k X) = \text{Cov}(X, X)v_k = \Sigma v_k = \lambda_k v_k$$

La notación matricial nos permite calcularlas para todas a la vez

$$\text{Cov}(X, Z) = \text{Cov}(X, V'X) = \text{Cov}(X, X)V = \Sigma V = V\Lambda$$

En particular para una variable X_j y una componente Z_k :

$$\text{Cov}(X_j, Z_k) = \lambda_k v_{jk}$$

Se trata por tanto de la varianza de la componente Z_k multiplicada por el coeficiente v_{jk} , que acompaña a X_j en la construcción de Z_k . El coeficiente de correlación será

$$\text{Corr}(X_j, Z_k) = \frac{\text{Cov}(X_j, Z_k)}{\sqrt{\text{Var}(X_j)\text{Var}(Z_k)}} = \frac{v_{jk}}{\sigma_j} \sqrt{\lambda_k}$$

En notación matricial

$$\text{Corr}(X, Z) = \text{diag}(1/\sigma_1, \dots, 1/\sigma_p) V \Lambda \text{diag}(1/\sqrt{\lambda_1}, \dots, 1/\sqrt{\lambda_p}) = \text{diag}(1/\sigma_1, \dots, 1/\sigma_p) V \Lambda^{1/2}$$

Se interpretará cada componente principal en función de las variables originales que estén correlacionadas con dicha componente. Asimismo, el signo de cada correlación permite valorar qué variables van en el sentido de la componente y cuáles en sentido opuesto.

5.7. Representación simultánea de individuos y variables: el biplot

El *biplot* es una representación gráfica simultánea de los individuos (mediante puntos) y las variables (mediante flechas), en un mismo sistema de coordenadas bidimensional construido en base a las dos primeras componentes principales. Permite interpretar el significado de las componentes (la primera en el eje horizontal y la segunda en el eje vertical) en base a las direcciones de las flechas.

A su vez, se valoran como parecidos los individuos cuyos puntos están próximos en el biplot. De igual modo tendrán correlación positiva las variables con flechas semejantes. Asimismo, los individuos que se encuentran en la dirección de cierta flecha tendrán observaciones altas en la variable representada por la flecha.

El biplot fue introducido por Gabriel (1971).

Bibliografía.

- Anderson, T.W. (2003). *An introduction to multivariate statistical analysis*. Wiley.
- Gabriel, K.R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, **58**, 453–467.
- Johnson, R.A. y Wichern, D.W. (1982). *Applied multivariate statistical analysis*. Prentice-Hall.
- Mardia, K.V., Kent, J.T. y Bibby, J.M. (1979). *Multivariate analysis*. Academic Press.
- Peña, D. (2002). *Análisis de datos multivariantes*. McGraw-Hill.
- Pérez, C. (2004). *Técnicas de análisis multivariante de datos*. Pearson Educación, S.A.
- Seber, G.A.F. (1984). *Multivariate observations*. Wiley.