Capítulo 1

Análisis Descriptivo Multivariante

1.1. Cálculo de Estadígrafos Multivariantes

1. Se dispone de 3 indicadores económicos x_1, x_2 y x_3, que se miden en cuatro paises, con los resultados siguientes:

x_1	x_2	x_3
2	3	-1
1	5	-2
2	2	1
2	3	1

   a) Calcular el vector de medias $\bar{x}$, la matriz de covarianzas S_x y la matriz de correlaciones R.

   b) Calcular el vector de medias utilizando las definiciones dadas en clases.

   c) Calcular la matriz de covarianzas utilizando todas las definiciones dadas en clases.

   d) Calcular la matriz de correlaciones utilizando la matriz de covarianzas.

   e) Estandarizar los datos, es decir construir la matriz Z.

   f) Utilizando la matriz Z hallada, calcular el vector de medias y la matriz de covarianzas.

g) A partir de los tres indicadores económicos x_1, x_2 y x_3 definidos inicialmente, se construyen dos nuevos indicadores:

$$\begin{aligned} y_1 &= (1/3)x_1 + (1/3)x_2 + (1/3)x_3 \\ y_2 &= x_1 - 0.5x_2 - 0.5x_3 \end{aligned}$$

Calcular el vector de medias para $y' = (y_1, y_2)$, su matriz de covarianzas y la matriz de correlaciones.

2. Considere la matriz de datos X que recoge $n = 6$ observaciones de un vector aleatorio $x = (x_1, x_2, x_3)'$, como se muestra a continuación:

x_1	x_2	x_3
-2	1	4
3	0	-1
5	1	2
-1	3	6
2	-7	4
-1	0	-1

a) Calcular el vector de medias $\bar{x}$ y la matriz de covarianzas muestrales S_x.

b) Calcular la matriz de covarianzas muestrales de los datos estandarizados a media cero y varianza unidad.

c) Sea el vector aleatorio $y = (y_1, y_2)'$, donde $y_1 = -x_1 + 2x_2 - x_3$ y $y_2 = x_1 + x_2$. Calcular el vector de medias $\bar{y}$ y la matriz de covarianzas muestrales S_y de la matriz de datos Y.

d) Calcular la matriz de observaciones Y mediante una operación matricial en la que aparezca la matriz de datos X

e) Calcular la matriz de covarianzas S_z del vector aleatorio $Z = (z_1, z_2)$, donde $z_1 = y_1/\sqrt{6}$ y $z_2 = y_2/\sqrt{2}$.

f) Calcular las matrices de correlaciones de X, Y y Z y de la matriz de datos estandarizada obtenida en el inciso b).

1.2. Aspectos Teóricos

1. Dada una muestra aleatoria de n vectores p-dimensionales $x_1, x_2, \ldots, x_n$, considere la combinación lineal $y_i = a_1x_{i1} + \ldots + a_px_{ip} = a'x_i$, donde $i = 1, \ldots, n$, donde $a = (a_1, \ldots, a_n)'$ es un vector fijo. Mostrar que:

   a) el vector de medias está definido como: $\bar{y} = a'\bar{x}$

   b) la matriz de covarianzas está dado por: $S_y = a'Sa$.

2. Sea A un matriz fija de dimensión $m \times p$, bajo la transformación:

$$y_i = Ax_i + b, \qquad i = 1, \ldots, n.$$

   mostrar que:

   $i)$ $\bar{y} = A\bar{x} + b$ $\qquad$ $ii)$ $S_y = AS_xA'$

   donde $\bar{x}$, S_x son el vector de medias y la matriz de covarianzas de la muestra aleatoria compuesta por los vectores p-dimensionales $x_1, x_2, \ldots, x_n$.

3. Sea

$$S(a) = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)(x_i - a)'$$

   donde la matriz de covarianzas S está definida cuando $x = a$. Mostrar que:

   a) $S(a) = S + (\bar{x} - a)(\bar{x} - a)'$

   b) $\min_a |S(a)| = |S|$

   c) $\min_a tr\, S(a) = tr\, S$

4. Sea $M = X'X$ donde X es una matriz de datos. Mostrar que:

   i) $m_{ii} = x_i'x_i = n(s_{ii} + \bar{x}_i^2)$

   ii) $m_{ij} = x_i'x_j = n(s_{ij} + \bar{x}_i\bar{x}_j)$

5. Dada una muestra aleatoria de n vectores p-dimensionales $x_1, x_2, \ldots, x_n$, considere la transformación $y_i = D^{-1}(x_i - \bar{x})$ con $i = 1, \ldots, n$ donde $D = diag(s_{ii})$. Mostrar que:

   a) la transformación puede escribirse como $Y = HXD^{-1}$, donde $Y' = (y_1, \ldots, y_n)$.

b) $\bar{y} = 0$ y $S_y = R$, utilizando el hecho de que $Y'1 = 0$ y $Y'HY = D^{-1}X'HXD^{-1}$.

6. Mostrar que la transformación de Mahalanobis definida por: $z_i = S^{-1/2}(x_i - \bar{x})$ para $i = 1, \ldots, n$, puede ser escrita como: $Z = HXS^{-1/2}$ donde $Z' = (z_1, \ldots, x_n)$. Asi mismo, mostrar que $\bar{z} = 0$ y $S_z = I$.

7. Sea $u_{pq} = M_{pq}/s_1^p s_2^q$ donde

$$M_{pq} = \frac{1}{n}\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^p (x_{i2} - \bar{x}_2)^q$$

Mostrar que:

$$\begin{aligned} b_{1,2} &= (1-r^2)^{-3}[u_{30}^2 + u_{03}^2 + 3(1+2r^2)(u_{12}^2 + u_{21}^2) - 2r^3 u_{30}u_{03} \\ &\qquad +6r\{u_{30}(ru_{12} - u_{21}) + u_{03}(ru_{21} - u_{12}) - (2+r^2)u_{12}u_{21}\}] \end{aligned}$$

y

$$b_{2,2} = (1-r^2)^{-2}[u_{40} + u_{04} + 2u_{22} + 4r(ru_{22} - u_{13} - u_{31})]$$

Por lo tanto, para $s_1 = s_2 = 1$, $r = 0$, mostrar que:

$$b_{1,2} = M_{30}^2 + M_{03}^2 + 3M_{12}^2 + 3M_{21}^2$$

y

$$b_{2,2} = M_{40} + M_{04} + 2M_{22}$$

Así $b_{1,2}$ acumula los efectos de M_{21}, M_{12}, M_{03}, M_{30} mientras que $b_{2,2}$ acumula los efectos de M_{22}, M_{04} y M_{40}.