

# Análisis Multivariado 2 - Práctica 1

## 1. Normal Multivariada Singular

1. Consideremos el modelo de regresión lineal

$$y_i = \beta_0 + \sum_{j=1}^{p-1} x_{i,j} + \epsilon_i \quad , \quad 1 \leq i \leq n$$

Es decir,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

donde  $\mathbf{y} = (y_1, \dots, y_n)^T$ ,  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{p-1})^T$ ,  $\mathbf{X} \in \mathbb{R}^{n \times p}$  de rango  $p < n$  tiene su primer columna con unos. Supongamos que  $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ .

Sea el vector de residuos  $\mathbf{e} = (\mathbf{I}_n - \mathbf{P})\mathbf{y}$ , donde  $\mathbf{P} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$  es la matriz de proyección. Mostrar que

- COV( $\mathbf{e}$ ) es singular.
- $\mathbf{e}$  tiene distribución normal multivariada singular.
- $\mathbf{1}^T \mathbf{e} = 0$ .

2. Sea  $\mathbf{x} \sim N_3(\mathbf{0}, \boldsymbol{\Sigma})$  con

$$\boldsymbol{\Sigma} = \begin{pmatrix} 3 & -2 & 1 \\ -2 & 2 & -2 \\ 1 & -2 & 3 \end{pmatrix}$$

a) Muestre que los autovectores de  $\boldsymbol{\Sigma}$  son:

$$\boldsymbol{\beta}_1 = (1, -1, 1)^T / \sqrt{3}$$

$$\boldsymbol{\beta}_2 = (1, 0, -1)^T / \sqrt{2}$$

$$\boldsymbol{\beta}_3 = (1, 2, 1)^T / \sqrt{6}$$

Deduzca que  $\mathbf{x}$  tiene distribución multivariada singular.

b) Pruebe que  $\mathbb{P}(x_1 + 2x_2 + x_3 = 0) = 1$ .

c) Calcular

$$\mathbb{P}(x_1 - 6x_2 - x_3 > 3)$$

de dos maneras diferentes:

- mostrando que  $y = x_1 - 6x_2 - x_3 \sim N(0, 76)$  a partir de la distribución de  $\mathbf{x}$ ,
- mostrando que  $y = x_1 - 6x_2 - x_3 = 2\sqrt{3}v_1 + \sqrt{2}v_2 - 2\sqrt{6}v_3$  y usando la distribución de  $(v_1, v_2, v_3)$ ,  $v_j = \boldsymbol{\beta}_j^T \mathbf{x}$ .

d) Calcular

$$\mathbb{P}\left(\frac{x_1^2}{2} - \frac{7x_2^2}{2} + \frac{x_3^2}{2} - 5x_1x_2 - 5x_2x_3 + x_1x_3 > 2\right)$$

**Ayuda:** Calcular  $v_1^2/6 + v_2^2/2 + 6v_3^2$  donde  $v_j = \boldsymbol{\beta}_j^T \mathbf{x}$ .

## 2. Biplots

1. En la tabla 1.1 figuran los datos de las mediciones de huesos y dientes de ratones campestres (de la especie *Microtus*). Consideremos solamente las 43 ratas del grupo 1.
  - a) Consideremos solamente las primeras 3 variables ( $Y_1, Y_2, Y_3$ ) que son el ancho de los molares superiores izquierdos 1, 2 y 3 respectivamente, medidos en mm./1000, para las 43 ratas del grupo 1.
    - 1) Hallar las componentes principales muestrales y los autovalores de  $S$ .
    - 2) Hallar los porcentajes de la variabilidad total explicados por la primera y por las dos primeras componentes, e interpretarlas en función de las variables originales.
    - 3) Realizar un biplot.
  - b) Realizar un biplot con todas las variables.
2. Repetir el análisis del ejercicio 1 de esta sección pero agregar a los datos las mediciones del grupo 2.
3. En la tabla 1.2 figuran los datos correspondientes a 91 países y 6 variables. Las variables son indicadoras de desarrollo.

<b>tasanat</b>	tasa de natalidad cada 1000 habitantes
<b>tasamor</b>	tasa de mortalidad cada 1000 habitantes
<b>mortinf</b>	mortalidad infantil (por debajo de un año)
<b>esphom</b>	esperanza de vida en hombres
<b>espmuj</b>	esperanza de vida en mujeres
<b>pnb</b>	producto bruto nacional per cápita

Cuadro 1: Variables indicadores de desarrollo.

- a) Realizar un biplot de los datos.
- b) Realizar un biplot de los datos transformados por logaritmos.