MAT-269: Estadística descriptiva multivariada

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Datos multivariados:

Tenemos una muestra aleatoria $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ donde para cada observación se ha medido $p \geq 2$ variables (o características) de interés. Así $\boldsymbol{x}_i = (x_{i1}, \dots, x_{ip})^\top$ es vector p-dimensional.

Podemos disponer la información en una matriz de datos[1]

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1^\top \\ \boldsymbol{x}_2^\top \\ \vdots \\ \boldsymbol{x}_n^\top \end{pmatrix}.$$

Observación:

Por simplicidad asumiremos que $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ son variables aleatorias IID desde $F_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (con F_p común).

[1] $\boldsymbol{X} = (x_{ij})$, para $i = 1, \dots, n; j = 1, \dots, p$.

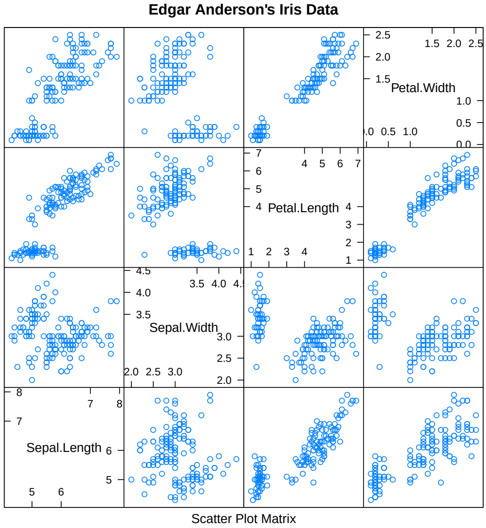
Datos de Flores Iris (Anderson, 1935; Fisher, 1936)

Iris setosa

Iris versicolor

Iris virginica

Datos de Flores Iris (Anderson, 1935; Fisher, 1936)



Datos de Flores Iris (Anderson, 1935; Fisher, 1936)

Datos observados:

Mediciones (cm) del largo y ancho de los sépalos y el largo y ancho de pétalos para 50 flores desde 3 especies de Iris (setosa, virginica y versicolor).

Objetivo:

- Obtener una función que permita discriminar entre especies.
- Usando las medidas de una flor, clasificarla apropiadamente.

Características del problema:

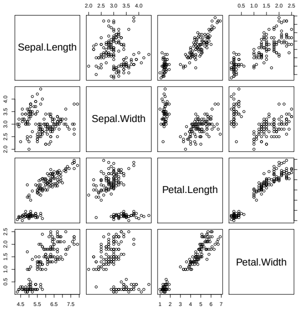
- El análisis exploratorio revela una separación evidente en 2 grupos.
- Técnicas más refinadas permiten identificar las 3 especies, p.ej.:
  - Análisis discriminante,
  - Técnicas de clasificación (Reconocimiento de patrones),
  - Aprendizaje de máquina (Máquinas de soporte vectorial, Data mining).

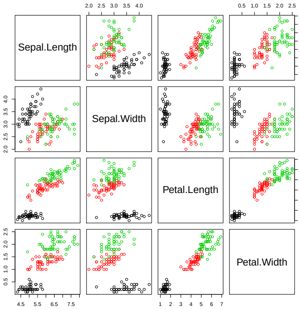
Conjunto de datos

```
> iris
    Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
1            5.1         3.5          1.4         0.2     setosa
2            4.9         3.0          1.4         0.2     setosa
3            4.7         3.2          1.3         0.2     setosa
4            4.6         3.1          1.5         0.2     setosa
5            5.0         3.6          1.4         0.2     setosa
6            5.4         3.9          1.7         0.4     setosa
7            4.6         3.4          1.4         0.3     setosa
8            5.0         3.4          1.5         0.2     setosa
9            4.4         2.9          1.4         0.2     setosa
10           4.9         3.1          1.5         0.1     setosa
11           5.4         3.7          1.5         0.2     setosa
12           4.8         3.4          1.6         0.2     setosa
13           4.8         3.0          1.4         0.1     setosa

...

148          6.5         3.0          5.2         2.0  virginica
149          6.2         3.4          5.4         2.3  virginica
150          5.9         3.0          5.1         1.8  virginica
```

Gráfico del conjunto de datos

```
x <- iris[,1:4]
pairs(x)                       # 1er panel
pairs(x, col = iris$Species) # colores representando 'especies'
```





Estadísticas de resumen

Análogamente a la media y covarianza $\overline{x}$ y s^2 para el caso unidimensional. Podemos definir sus contrapartes multivariadas como:

$$\overline{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i,$$

$$\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top.$$

En efecto, $\overline{\boldsymbol{x}} = (\overline{x}_1, \dots, \overline{x}_p)^\top$ y $\boldsymbol{S} = (s_{rs})$, donde

$$\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij},$$

$$s_{rs} = \frac{1}{n-1}\sum_{i=1}^{n} (x_{ir} - \overline{x}_r)(x_{is} - \overline{x}_s),$$

para $r, s = 1, \dots, p$.

Observación:

Algunas propiedades del vector de medias y la matriz de covarianza, surgen de escribir formas compactas que dependen de la matriz de datos $\boldsymbol{X} = (x_{ij}) \in \mathbb{R}^{n\times p}$. En efecto,

$$\overline{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i = \frac{1}{n}\boldsymbol{X}^\top\boldsymbol{1}_n.$$

Sea

$$\begin{aligned}\boldsymbol{Q} &= \sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top = \sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})\boldsymbol{x}_i^\top - \sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})\overline{\boldsymbol{x}}^\top \\ &= \sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^\top - \overline{\boldsymbol{x}}\sum_{i=1}^{n}\boldsymbol{x}_i^\top = \sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^\top - n\overline{\boldsymbol{x}}\,\overline{\boldsymbol{x}}^\top,\end{aligned}$$

pues $\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})\overline{\boldsymbol{x}}^\top = \boldsymbol{0}$.[2]

[2] No se hará distinción sobre el orden de las matrices de ceros.

Notando

$$\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top = \boldsymbol{X}^\top \boldsymbol{X},$$

sigue que

$$\begin{aligned} \boldsymbol{S} &= \frac{1}{n-1}\boldsymbol{Q} = \frac{1}{n-1}\Big\{\boldsymbol{X}^\top\boldsymbol{X} - n\Big(\frac{1}{n}\boldsymbol{X}^\top\mathbf{1}_n\Big)\Big(\frac{1}{n}\boldsymbol{X}^\top\mathbf{1}_n\Big)^\top\Big\} \\ &= \frac{1}{n-1}\Big(\boldsymbol{X}^\top\boldsymbol{X} - \frac{1}{n}\boldsymbol{X}^\top\mathbf{1}_n\mathbf{1}_n^\top\boldsymbol{X}\Big) = \frac{1}{n-1}\boldsymbol{X}^\top\boldsymbol{C}\boldsymbol{X} \end{aligned}$$

con $\boldsymbol{C} = \boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top$ la matriz de centrado. Es sencillo mostrar que

$$\boldsymbol{C}^\top = \boldsymbol{C}, \qquad \boldsymbol{C}^2 = \boldsymbol{C},$$

es decir $\boldsymbol{C}$ es matriz de proyección. Esto permite mostrar el siguiente resultado.

Resultado 1:

La matriz de covarianza $\boldsymbol{S}$, es semidefinida positiva.

Demostración:

Sea $\boldsymbol{a} \in \mathbb{R}^p$, vector no nulo. Tenemos que,

$$\begin{aligned}\boldsymbol{a}^\top \boldsymbol{S}\boldsymbol{a} &= \frac{1}{n-1}\boldsymbol{a}^\top \boldsymbol{X}^\top \boldsymbol{C}\boldsymbol{X}\boldsymbol{a} = \frac{1}{n-1}\boldsymbol{a}^\top \boldsymbol{X}^\top \boldsymbol{C}^2\boldsymbol{X}\boldsymbol{a} \\ &= \frac{1}{n-1}\boldsymbol{u}^\top \boldsymbol{u} \geq 0, \qquad \boldsymbol{u} = \boldsymbol{C}\boldsymbol{X}\boldsymbol{a},\end{aligned}$$

es decir, $\boldsymbol{S}$ es matriz semidefinida positiva.[3]

[3] $\boldsymbol{S}$ será definida positiva si $n \geq p+1$.

La matriz de correlación entre las p variables de interés, es dada por:

$$\boldsymbol{R} = (r_{ij}),$$

donde

$$\begin{aligned} r_{jk} &= \frac{\sum_{i=1}^{n}(x_{ij}-\overline{x}_j)(x_{ik}-\overline{x}_k)}{\sqrt{\sum_{i=1}^{n}(x_{ij}-\overline{x}_j)^2\sum_{i=1}^{n}(x_{ik}-\overline{x}_k)^2}} \\ &= \frac{s_{jk}}{\sqrt{s_{jj}s_{kk}}}, \end{aligned}$$

para $j,k = 1,\dots,p$, con $\boldsymbol{S} = (s_{jk})$.

Sea, $\boldsymbol{D} = \operatorname{diag}(s_{11}, s_{22}, \dots, s_{pp})$. Así, podemos escribir

$$\boldsymbol{R} = \boldsymbol{D}^{-1/2}\boldsymbol{S}\boldsymbol{D}^{-1/2}, \qquad \boldsymbol{S} = \boldsymbol{D}^{1/2}\boldsymbol{R}\boldsymbol{D}^{1/2}.$$

Estadísticas de resumen: Datos Iris

```
> x <- iris[,1:4]

# cálculo de estadísticas de resumen multivariadas
> xbar <- apply(x, 2, mean)
> S <- cov(x)
> R <- cor(x)

# salida
> xbar
Sepal.Length  Sepal.Width Petal.Length  Petal.Width
      5.8433       3.0573       3.7580       1.1993

> S
             Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length       0.6857     -0.0424         1.27       0.516
Sepal.Width       -0.0424      0.1900        -0.33      -0.122
Petal.Length       1.2743     -0.3297         3.12       1.296
Petal.Width        0.5163     -0.1216         1.30       0.581

# alternativamente podemos hacer
> xbar <- colMeans(x)
> R <- cov2cor(S)
> z <- cov.wt(x, cor = TRUE, method = "unbiased")
```

Estadísticas de resumen: Datos Iris

```
  # 'cov.wt' entrega una 'lista'
  > z <- cov.wt(x, cor = TRUE, method = "unbiased")
  > z
$cov
             Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length    0.6856935  -0.0424340    1.2743154   0.5162707
Sepal.Width    -0.0424340   0.1899794   -0.3296564  -0.1216394
Petal.Length    1.2743154  -0.3296564    3.1162779   1.2956094
Petal.Width     0.5162707  -0.1216394    1.2956094   0.5810063

$center
Sepal.Length  Sepal.Width Petal.Length  Petal.Width
    5.843333     3.057333     3.758000     1.199333

$n.obs
[1] 150

$cor
             Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length    1.0000000  -0.1175698    0.8717538   0.8179411
Sepal.Width    -0.1175698   1.0000000   -0.4284401  -0.3661259
Petal.Length    0.8717538  -0.4284401    1.0000000   0.9628654
Petal.Width     0.8179411  -0.3661259    0.9628654   1.0000000
```

Transformaciones lineales

Considere:

$$\boldsymbol{y}_i = \boldsymbol{A}\boldsymbol{x}_i + \boldsymbol{b}, \qquad i = 1, \dots, n,$$

donde $\boldsymbol{A} \in \mathbb{R}^{q\times p}$ y $\boldsymbol{b} \in \mathbb{R}^p$. Entonces,

$$\overline{\boldsymbol{y}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{y}_i = \boldsymbol{A}\,\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i + \boldsymbol{b} = \boldsymbol{A}\overline{\boldsymbol{x}} + \boldsymbol{b},$$

mientras que

$$\boldsymbol{y}_i - \overline{\boldsymbol{y}} = \boldsymbol{A}\boldsymbol{x}_i + \boldsymbol{b} - \boldsymbol{A}\overline{\boldsymbol{x}} - \boldsymbol{b} = \boldsymbol{A}(\boldsymbol{x}_i - \overline{\boldsymbol{x}}).$$

De este modo,

$$\begin{aligned}\boldsymbol{S}_Y &= \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{y}_i - \overline{\boldsymbol{y}})(\boldsymbol{y}_i - \overline{\boldsymbol{y}})^\top = \frac{1}{n-1}\sum_{i=1}^{n}\boldsymbol{A}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top\boldsymbol{A}^\top \\ &= \frac{1}{n-1}\boldsymbol{A}\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top\boldsymbol{A}^\top = \boldsymbol{A}\boldsymbol{S}_X\boldsymbol{A}^\top.\end{aligned}$$

Transformación de Mahalanobis

En particular, para la transformación,

$$\boldsymbol{z}_i = \boldsymbol{S}^{-1/2}(\boldsymbol{x}_i - \overline{\boldsymbol{x}}), \qquad i = 1, \dots, n,$$

donde $\boldsymbol{S} = \boldsymbol{S}^{1/2}\boldsymbol{S}^{1/2}$ con $\boldsymbol{S}^{1/2}$ un factor raíz cuadrada de $\boldsymbol{S}$, sigue que

$$\overline{\boldsymbol{z}} = \boldsymbol{0}, \qquad \text{y} \qquad \boldsymbol{S}_Z = \boldsymbol{I}_p.$$

Observación:

En la práctica podemos considerar los siguientes métodos para obtener $\boldsymbol{S}^{-1/2}$:

- descomposición Cholesky.
- descomposición espectral.[4]

[4] Este procedimiento no es recomendado.

Suponga $\boldsymbol{S} > 0$ y considere la descomposición espectral

$$\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top,$$

donde $\boldsymbol{U}$ es matriz ortogonal y $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_p)$ con $\lambda_1 \geq \cdots \geq \lambda_p > 0$ son los valores propios de $\boldsymbol{S}$. De este modo podemos considerar $\boldsymbol{S}^{-1/2} = \boldsymbol{U}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}^\top$,[5] con $\boldsymbol{\Lambda}^{-1/2} = \operatorname{diag}(\lambda_1^{-1/2}, \dots, \lambda_p^{-1/2})$. Esto lleva a la transformación

$$\boldsymbol{z}_i = \boldsymbol{U}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}^\top(\boldsymbol{x}_i - \overline{\boldsymbol{x}}), \qquad i = 1, \dots, n. \tag{1}$$

Usando la descomposición Cholesky tenemos $\boldsymbol{S} = \boldsymbol{G}\boldsymbol{G}^\top$, con $\boldsymbol{G}$ matriz triangular inferior. En este caso, $\boldsymbol{S}^{-1} = (\boldsymbol{G}\boldsymbol{G}^\top)^{-1} = \boldsymbol{G}^{-\top}\boldsymbol{G}^{-1}$ y hacemos

$$\boldsymbol{z}_i = \boldsymbol{G}^{-1}(\boldsymbol{x}_i - \overline{\boldsymbol{x}}), \qquad i = 1, \dots, n. \tag{2}$$

[5] aún otra alternativa es considerar $\boldsymbol{S}^{1/2} = \boldsymbol{U}\boldsymbol{\Lambda}^{1/2}$ de ahí que $\boldsymbol{S}^{-1/2} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}^\top$.

Observación:

Si consideramos $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$ y hacemos

$$\boldsymbol{y}_i = \boldsymbol{U}^\top(\boldsymbol{x}_i - \overline{\boldsymbol{x}}), \qquad i = 1, \dots, n. \tag{3}$$

Entonces, es fácil notar que

$$\overline{\boldsymbol{y}} = \boldsymbol{0},$$

mientras que

$$\boldsymbol{S}_Y = \boldsymbol{U}^\top\boldsymbol{S}\boldsymbol{U} = \boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{\Lambda}.$$

Además,

$$\operatorname{tr}\boldsymbol{S}_Y = \operatorname{tr}\boldsymbol{\Lambda} = \sum_{j=1}^p \lambda_j,$$

$$|\boldsymbol{S}_Y| = |\boldsymbol{\Lambda}| = \prod_{j=1}^p \lambda_j.$$

La transformación en (3) surge en análisis de componentes principales.

Definición 1 (Distancia de Mahalanobis):

Considere una muestra de n observaciones $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$. De este modo, la distancia de Mahalanobis es dada por

$$D_i = \{(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top \boldsymbol{S}^{-1}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})\}^{1/2}, \qquad i = 1, \dots, n,$$

como la distancia de la i-ésima observación hacia el "centro" de los datos, $\overline{\boldsymbol{x}}$ ponderada por la matriz de covarianza.

Observación:

Note que las distancias D_i pueden ser calculadas de forma bastante eficiente usando (2), como:

$$D_i = \{(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top \boldsymbol{S}^{-1}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})\}^{1/2} = (\boldsymbol{z}_i^\top \boldsymbol{z}_i)^{1/2},$$

es más, $\boldsymbol{z}_i$ es obtenido como solución del sistema $\boldsymbol{G}\boldsymbol{z}_i = \boldsymbol{x}_i - \overline{\boldsymbol{x}}$ para $i = 1, \dots, n$.

Usando

$$g_{ij} = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top \boldsymbol{S}^{-1}(\boldsymbol{x}_j - \overline{\boldsymbol{x}}), \qquad i,j = 1,\dots,n.$$

Mardia (1970)[6] definió medidas de sesgo y curtosis multivariadas, dadas por

$$b_{1p} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}^3, \qquad b_{2p} = \frac{1}{n}\sum_{i=1}^{n} g_{ii}^2,$$

respectivamente.

Bajo normalidad, debemos tener:

$$b_{1p} = 0, \qquad b_{2p} = p(p+2).$$

Observación:

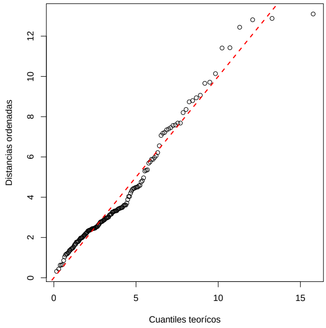
Las estadísticas b_{1p} y b_{2p} son invariantes bajo transformaciones afín:

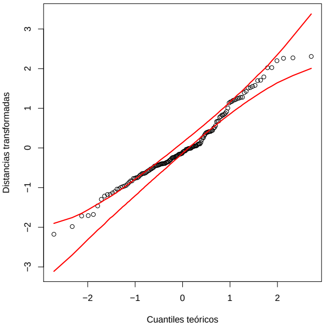
$$\boldsymbol{y}_i = \boldsymbol{A}\boldsymbol{x}_i + \boldsymbol{b}.$$

[6] Biometrika **57**, 519-530.

```
  # número de obs y variables
  > nobs <- nrow(x) # 150 obs
  > p <- ncol(x)   # 4 variables

  > library(fastmatrix) # https://faosorios.github.io/fastmatrix/

  # distancias de Mahalanobis, sesgo y kurtosis
  > D2 <- Mahalanobis(x, xbar, S)
  > skewness(x)
[1] 2.69722
  > kurtosis(x)
[1] 23.73966
attr(,"excess")
[1] -0.2603421
  > p * (p + 2) # valor de kurtosis bajo normalidad
[1] 24

  # QQ-plot de distancias de Mahalanobis
  > qqplot(qchisq(ppoints(nobs), df = p), D2,
  +  xlab = "Cuantiles teorícos",
  +  ylab = "Distancias ordenadas")
  > abline(c(0,1), col = "red", lwd = 2, lty = 2)
```

Distancia de Mahalanobis: Datos Iris





Un algoritmo online para calcular $\overline{x}$ y S (Clarke, 1971)[7]

Considere

$$\overline{\boldsymbol{x}}_n = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i.$$

De este modo,

$$\begin{aligned}
\overline{\boldsymbol{x}}_n &= \frac{1}{n}\Big(\sum_{i=1}^{n-1}\boldsymbol{x}_i + \boldsymbol{x}_n\Big) = \frac{1}{n}\Big((n-1)\overline{\boldsymbol{x}}_{n-1} + \boldsymbol{x}_n\Big) \\
&= \frac{1}{n}\big(n\overline{\boldsymbol{x}}_{n-1} - \overline{\boldsymbol{x}}_{n-1} + \boldsymbol{x}_n\big) \\
&= \overline{\boldsymbol{x}}_{n-1} + \frac{\boldsymbol{\delta}_n}{n}, \qquad (4)
\end{aligned}$$

con $\boldsymbol{\delta}_n = \boldsymbol{x}_n - \overline{\boldsymbol{x}}_{n-1}$.

Ecuación en (4) corresponde a un algoritmo recursivo para el cálculo del promedio.

[7] Applied Statistics **20**, 206-209.

Sea

$$\boldsymbol{Q}_n = \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_n)(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_n)^\top.$$

Es fácil notar que (Tarea):

$$\boldsymbol{Q}_n = \boldsymbol{Q}_{n-1} + \Big(1 - \frac{1}{n}\Big)\boldsymbol{\delta}_n\boldsymbol{\delta}_n^\top. \qquad (5)$$

con $\boldsymbol{\delta}_n = \boldsymbol{x}_n - \overline{\boldsymbol{x}}_{n-1}$.

Ecuaciones (4) y (5) llevan al siguiente algoritmo.

Cálculo de la varianza muestral. Algoritmo online (1-paso)[8]

Algoritmo AS 41: Promedio y matriz de covarianza muestral.

Entrada: Matriz de datos $\boldsymbol{X}^\top = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$.
Salida : Promedio y matriz de covarianza, $\overline{\boldsymbol{x}}$ y $\boldsymbol{S}$.

begin
- $\boldsymbol{M} \leftarrow \boldsymbol{x}_1$
- $\boldsymbol{Q} \leftarrow \boldsymbol{0}$
- **for** $i = 2$ **to** n **do**
  - $\boldsymbol{\delta} \leftarrow \boldsymbol{x}_i - \boldsymbol{M}$
  - $\boldsymbol{M} \leftarrow \boldsymbol{M} + \frac{1}{i}\boldsymbol{\delta}$
  - $\boldsymbol{Q} \leftarrow \boldsymbol{Q} + \left(1 - \frac{1}{i}\right)\boldsymbol{\delta}\boldsymbol{\delta}^\top$
- **end**
- $\overline{\boldsymbol{x}} \leftarrow \boldsymbol{M}$
- $\boldsymbol{S} \leftarrow \frac{1}{n-1}\boldsymbol{Q}$

end

[8] Algoritmo implementado en la función cov.weighted del paquete `fastmatrix`.

Propiedades básicas de los momentos muestrales

Suponga que $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ son una muestra aleatoria, tal que $\mathsf{E}(\boldsymbol{x}_i) = \boldsymbol{\mu}$ y $\mathsf{Cov}(\boldsymbol{x}_i) = \boldsymbol{\Sigma}$. De este modo,

$$\mathsf{E}(\overline{\boldsymbol{x}}) = \frac{1}{n} \sum_{i=1}^{n} \mathsf{E}(\boldsymbol{x}_i) = \boldsymbol{\mu},$$

mientras que

$$\mathsf{Cov}(\overline{\boldsymbol{x}}) = \frac{1}{n^2} \sum_{i=1}^{n} \mathsf{Cov}(\boldsymbol{x}_i) = \frac{1}{n} \boldsymbol{\Sigma}.$$

Sea $\boldsymbol{y}_i = \boldsymbol{x}_i - \boldsymbol{\mu}$, así $\overline{\boldsymbol{y}} = \overline{\boldsymbol{x}} - \boldsymbol{\mu}$ y $\mathsf{E}(\boldsymbol{y}_i) = \boldsymbol{0}$, $\mathsf{Cov}(\boldsymbol{y}_i) = \boldsymbol{\Sigma}$, para $i = 1, \dots, n$.

Además

$$\begin{aligned} \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top &= \sum_{i=1}^{n} (\boldsymbol{y}_i - \overline{\boldsymbol{y}})(\boldsymbol{y}_i - \overline{\boldsymbol{y}})^\top \\ &= \sum_{i=1}^{n} \boldsymbol{y}_i \boldsymbol{y}_i^\top - n \overline{\boldsymbol{y}}\, \overline{\boldsymbol{y}}^\top \end{aligned}$$

Ahora

$$\begin{aligned}
\mathsf{E}(\boldsymbol{Q}) &= \sum_{i=1}^{n} \mathsf{E}(\boldsymbol{y}_i \boldsymbol{y}_i^\top) - n\,\mathsf{E}(\overline{\boldsymbol{y}}\,\overline{\boldsymbol{y}}^\top) \\
&= \sum_{i=1}^{n} \mathsf{Cov}(\boldsymbol{y}_i) - n\,\mathsf{Cov}(\overline{\boldsymbol{y}}) \\
&= n\boldsymbol{\Sigma} - n\Big(\frac{1}{n}\boldsymbol{\Sigma}\Big) = (n-1)\boldsymbol{\Sigma}.
\end{aligned}$$

Es decir, $\overline{\boldsymbol{x}}$ y $\boldsymbol{S}$ son estimadores insesgados de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$, respectivamente.

Observación:

Evidentemente

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{Q} = \Big(\frac{n-1}{n}\Big)\boldsymbol{S},$$

es un estimador sesgado para $\boldsymbol{\Sigma}$.