

# MAT-269: Sesión 2, Estadística descriptiva

**Felipe Osorio**

[fosorios.mat.utfsm.cl](mailto:fosorios.mat.utfsm.cl)

Departamento de Matemática, UTFSM



## Datos multivariados:

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

Podemos disponer la información en una **matriz de datos**<sup>1</sup>

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

## Observación:

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

---

<sup>1</sup> $\mathbf{X} = (x_{ij})$ , para  $i = 1, \dots, n; j = 1, \dots, p$ .

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

*Iris setosa*



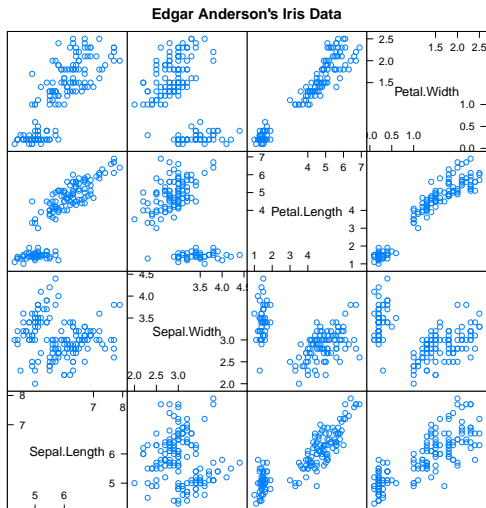
*Iris versicolor*



*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).



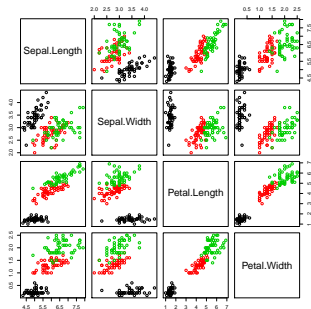
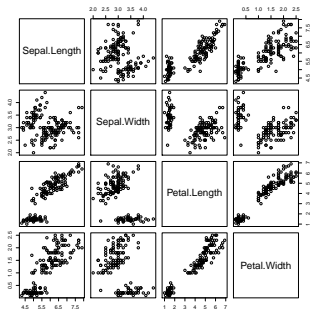
## Conjunto de datos

```
> 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)
pairs(x, col = iris$Species) # 1er panel
pairs(x, col = iris$Species) # colores representando 'especies'
```



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

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i,$$

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

En efecto,  $\bar{\mathbf{x}} = (\bar{x}_1, \dots, \bar{x}_p)^\top$  y  $\mathbf{S} = (s_{jk})$ , donde

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

$$s_{jk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k),$$

para  $j, k = 1, \dots, p$ .





## Observación:

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

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

Sea

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

pues  $\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})\bar{\mathbf{x}}^\top = \mathbf{0}$ .<sup>2</sup>

---

<sup>2</sup>No se hará distinción sobre el *orden* de las matrices de ceros.



Notando

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

sigue que

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

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

$$\mathbf{C}^\top = \mathbf{C}, \quad \mathbf{C}^2 = \mathbf{C},$$

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



## Resultado:

La matriz de covarianza  $S$ , es semidefinida positiva.

## Demostración:

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

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

es decir,  $S$  es matriz semidefinida positiva.<sup>3</sup>

---

<sup>3</sup> $S$  será definida positiva si  $n \geq p + 1$ .



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

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

donde

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

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

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

$$\mathbf{R} = \mathbf{D}^{-1/2} \mathbf{S} \mathbf{D}^{-1/2}, \quad \mathbf{S} = \mathbf{D}^{1/2} \mathbf{R} \mathbf{D}^{1/2}.$$



## 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:

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

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

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

mientras que

$$\mathbf{y}_i - \bar{\mathbf{y}} = \mathbf{A}\mathbf{x}_i + \mathbf{b} - \mathbf{A}\bar{\mathbf{x}} - \mathbf{b} = \mathbf{A}(\mathbf{x}_i - \bar{\mathbf{x}}).$$

De este modo,

$$\begin{aligned} S_Y &= \frac{1}{n-1} \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^\top = \frac{1}{n-1} \sum_{i=1}^n \mathbf{A}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top \mathbf{A}^\top \\ &= \frac{1}{n-1} \mathbf{A} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top \mathbf{A}^\top = \mathbf{A}S_X \mathbf{A}^\top. \end{aligned}$$



## Transformación de Mahalanobis

En particular, para la transformación,

$$\mathbf{z}_i = \mathbf{S}^{-1/2}(\mathbf{x}_i - \bar{\mathbf{x}}), \quad i = 1, \dots, n,$$

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

$$\bar{\mathbf{z}} = \mathbf{0}, \quad \text{y} \quad \mathbf{S}_Z = \mathbf{I}_p.$$

### Observación:

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

- ▶ descomposición **Cholesky**.
- ▶ descomposición **espectral**.<sup>4</sup>

---

<sup>4</sup>Este procedimiento no es recomendado.





Suponga  $\mathbf{S} > 0$  y considere la [descomposición espectral](#)

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

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

$$\mathbf{z}_i = \mathbf{U}\mathbf{\Lambda}^{-1/2}\mathbf{U}^\top(\mathbf{x}_i - \bar{\mathbf{x}}), \quad i = 1, \dots, n. \quad (1)$$

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

$$\mathbf{z}_i = \mathbf{G}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}), \quad i = 1, \dots, n. \quad (2)$$

---

<sup>5</sup>aún otra alternativa es considerar  $\mathbf{S}^{1/2} = \mathbf{U}\mathbf{\Lambda}^{1/2}$  de ahí que  $\mathbf{S}^{-1/2} = \mathbf{\Lambda}^{-1/2}\mathbf{U}^\top$ .



### Observación:

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

$$\mathbf{y}_i = U^\top (\mathbf{x}_i - \bar{\mathbf{x}}), \quad i = 1, \dots, n. \quad (3)$$

Entonces, es fácil notar que

$$\bar{\mathbf{y}} = \mathbf{0},$$

mientras que

$$S_Y = U^\top S U = U^\top U \Lambda U^\top U = \Lambda.$$

Además,

$$\text{tr } S_Y = \text{tr } \Lambda = \sum_{j=1}^p \lambda_j,$$

$$|S_Y| = |\Lambda| = \prod_{j=1}^p \lambda_j.$$

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



## Definición (Distancia de Mahalanobis):

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

$$D_i = \{(\mathbf{x}_i - \bar{\mathbf{x}})^\top \mathbf{S}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})\}^{1/2}, \quad i = 1, \dots, n,$$

como la distancia de la observación  $i$ -ésima hacia el “centro” de los datos,  $\bar{\mathbf{x}}$  ponderada por la matriz de covarianza.

## Observación:

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

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

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



Usando

$$g_{ij} = (\mathbf{x}_i - \bar{\mathbf{x}})^\top \mathbf{S}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}), \quad i, j = 1, \dots, n.$$

Mardia (1970)<sup>6</sup> definió medidas de sesgo y curtosis multivariadas, dadas por

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

respectivamente.

Bajo normalidad, debemos tener:

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

*Observación:*

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

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

---

<sup>6</sup>Mardia (1970) *Biometrika* **57**, 519-530.



## Distancia de Mahalanobis: Datos Iris

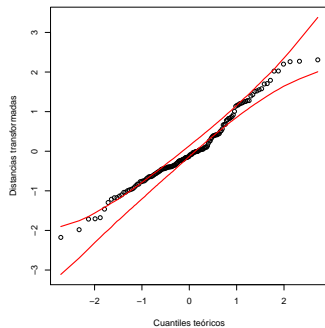
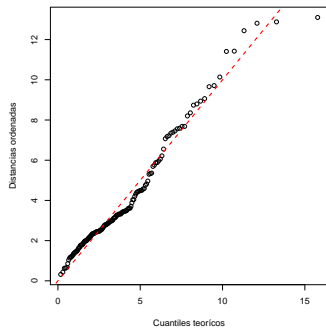
```
# número de obs y variables
> nobs <- nrow(x) # 150 obs
> p <- ncol(x)    # 4 variables

# distancias de Mahalanobis y kurtosis
> D2 <- mahalanobis(x, xbar, S)
> kurtosis <- mean(D2^2)
> kurtosis
[1] 23.42418
> 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 teóricos",
+ ylab = "Distancias ordenadas")
> abline(c(0,1), col = "red", lwd = 2, lty = 2)
```



# Distancia de Mahalanobis: Datos Iris



## Un algoritmo online para calcular $\bar{x}$ y $S$ (Clarke, 1971)<sup>7</sup>

Considere

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i.$$

De este modo,

$$\begin{aligned}\bar{x}_n &= \frac{1}{n} \left( \sum_{i=1}^{n-1} x_i + x_n \right) = \frac{1}{n} \left( (n-1)\bar{x}_{n-1} + x_n \right) \\ &= \frac{1}{n} \left( n\bar{x}_{n-1} - \bar{x}_{n-1} + x_n \right) \\ &= \bar{x}_{n-1} + \frac{\delta_n}{n},\end{aligned}\tag{4}$$

con  $\delta_n = x_n - \bar{x}_{n-1}$ .

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

---

<sup>7</sup>Clarke (1971) *Applied Statistics* 20, 206-209.



## Un algoritmo online para calcular $\bar{x}$ y $S$ (Clarke, 1971)

Sea

$$Q_n = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}_n)(\mathbf{x}_i - \bar{\mathbf{x}}_n)^\top.$$

Es fácil notar que (Tarea):

$$Q_n = Q_{n-1} + \left(1 - \frac{1}{n}\right) \delta_n \delta_n^\top. \quad (5)$$

con  $\delta_n = \mathbf{x}_n - \bar{\mathbf{x}}_{n-1}$ .

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





## Cálculo de la varianza muestral. Algoritmo online (1-paso)<sup>8</sup>

---

**Algoritmo AS 41:** Promedio y matriz de covarianza muestral.

---

**Entrada:** Matriz de datos  $X^T = (x_1, \dots, x_n)$ .

**Salida :** Promedio y matriz de covarianza,  $\bar{x}$  y  $S$ .

```
1 begin
2    $M \leftarrow x_1$ 
3    $Q \leftarrow 0$ 
4   for  $i = 2$  to  $n$  do
5      $\delta \leftarrow x_i - M$ 
6      $M \leftarrow M + \frac{1}{i}\delta$ 
7      $Q \leftarrow Q + \left(1 - \frac{1}{i}\right)\delta\delta^T$ 
8   end
9    $\bar{x} \leftarrow M$ 
10   $S \leftarrow \frac{1}{n-1}Q$ 
11 end
```

---

<sup>8</sup>Algoritmo implementado en el paquete [heavy](#).



## Propiedades básicas de los momentos muestrales

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

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

mientras que

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

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

Además

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



Ahora

$$\begin{aligned} E(\mathbf{Q}) &= \sum_{i=1}^n E(\mathbf{y}_i \mathbf{y}_i^\top) - n E(\bar{\mathbf{y}} \bar{\mathbf{y}}^\top) \\ &= \sum_{i=1}^n \text{Cov}(\mathbf{y}_i) - n \text{Cov}(\bar{\mathbf{y}}) \\ &= n\boldsymbol{\Sigma} - n\left(\frac{1}{n}\boldsymbol{\Sigma}\right) = (n-1)\boldsymbol{\Sigma}. \end{aligned}$$

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

*Observación:*

Evidentemente

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \mathbf{Q} = \left(\frac{n-1}{n}\right) \mathbf{S},$$

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

