Análisis Multivariado

ID 033521 - Clase 4901

Lina Maria Acosta Avena

Ciencia de Datos
Departamento de Matemáticas
Pontificia Universidad Javeriana

Semana 4(Parte 1): 05/08/24 – 10/08/24

Introducción

Recuerde que:

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \\ \vdots \\ X_p \end{bmatrix}_{p\times 1}$$

es un **vector aleatorio** $p-$variado con **fdpc**

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}(x_1, x_2, \ldots, x_p)$$

Observación:
las $X_i^{\prime}s$ son ***variables aleatorias*** *y $x_i^{\prime}s$* ***realizaciones*** *de X_i.*

Introducción

Sabemos que si

$$\mathbf{X} \sim \mathrm{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$$

su fdp es

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

y si

$$\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$$

es una **muestra aleatoria** que se extrae de esta población, entonces

$$f_{\mathbf{X}_j}(\mathbf{x}_j) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}_j-\boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_j-\boldsymbol{\mu})\right\}$$

Introducción

y por lo tanto, la **fdpc** de la muestra aleatoria queda dada por

$$f_{\mathbf{X}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \prod_{j=1}^{n} \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}_j - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}_j - \boldsymbol{\mu})\right\}$$

$$= \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{n}(\mathbf{x}_j - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}_j - \boldsymbol{\mu})\right\}$$

Introducción

y por lo tanto, la **fdpc** de la muestra aleatoria queda dada por

$$f_{\mathbf{X}}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) = \prod_{j=1}^{n} \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}_j - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}_j - \boldsymbol{\mu})\right\}$$

$$= \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{n}(\mathbf{x}_j - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}_j - \boldsymbol{\mu})\right\}$$

Cuando la **muestra** es **observada**

$$\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$$

la ecuación anterior es una **función** de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$, esto es

Introducción

$$f_{\mathbf{X}}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma} \middle| \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\right) = \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{n}(\mathbf{x}_j - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}_j - \boldsymbol{\mu})\right\}$$

$$= L\left(\boldsymbol{\mu}, \boldsymbol{\Sigma} \middle| \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\right)$$

y es llamada **función de verosimilitud**.

Nota: La función de verosimilitud suministra un orden natural de preferencia o plausibilidad entre los valores posibles de los parámetros basado en el valor observado de $\boldsymbol{X}$.

Introducción

Evidentemente que

- ✓ $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ son **desconocidos**
- ✓ el interés es **estimarlos**
- ✓ se quiere **hacer inferencia** sobre ellos.

Estimación de μ y $\boldsymbol{\Sigma}$

Estimación de $\boldsymbol{\mu}$

Los **estimadores de máxima verosimilitud** *(MLE, Maximum Likelihood Estimator)* de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ son aquellos que **maximizan** a

$$L\left(\boldsymbol{\mu}, \boldsymbol{\Sigma} \middle| \, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\right)$$

y están dados por

$$\widehat{\boldsymbol{\mu}} = \overline{\mathbf{x}}$$

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{j=1}^{n} \left(\mathbf{x}_j - \overline{\mathbf{x}}\right)\left(\mathbf{x}_j - \overline{\mathbf{x}}\right)^{\top} = \frac{n-1}{n} \mathbf{S}$$

donde

Estimación de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$

$$\overline{\mathbf{x}} = \begin{bmatrix} \overline{x}_1 \\ \overline{x}_2 \\ \vdots \\ \overline{x}_p \end{bmatrix}$$

es el **vector de media muestral** y

$$\mathbf{S} = \frac{1}{n-1} \sum_{j=1}^{n} (\mathbf{x}_j - \overline{\mathbf{x}}) (\mathbf{x}_j - \overline{\mathbf{x}})^\top$$

es la **matriz de varianzas y covarianzas muestral**.

Estimación de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$

Dos propiedades importantes de los MLE's:

1. Los estimadores de máxima verosimilitud (MLE) son **invariantes a transformaciones**, esto es, si $\widehat{\boldsymbol{\theta}}$ es el MLE de $\boldsymbol{\theta}$, entonces el MLE de $h(\theta)$ es

$$\widehat{h(\boldsymbol{\theta})} = h(\widehat{\boldsymbol{\theta}})$$

para $h(\cdot)$ una función continua de $\boldsymbol{\theta}$.

Por ejemplo, el MLE para $\boldsymbol{\mu}^\top \boldsymbol{\Sigma} \boldsymbol{\mu}$ es $\widehat{\boldsymbol{\mu}}^\top \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\mu}}$, donde $\widehat{\boldsymbol{\mu}}$ y $\widehat{\boldsymbol{\Sigma}}$ son los MLE's de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$, respectivamente.

Estimación de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$

2. Los estimadores de máxima verosimilitud (MLE) son **estadísticos suficientes**, esto es, toda la información que contiene la muestra sobre $\boldsymbol{\theta}$, esta contenida en $\widehat{\boldsymbol{\theta}}$. Así que, $\widehat{\boldsymbol{\mu}}$ y $\widehat{\boldsymbol{\Sigma}}$ contienen toda la información de $\mathbf{X}_{n\times p}$.

   Observación:

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} \exp\left\{ -\frac{1}{2}\sum_{j=1}^{n}(\mathbf{x}_j - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}_j - \boldsymbol{\mu}) \right\}$$

$$= \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} \exp\left\{ -\frac{1}{2}\mathrm{traza}\left[\boldsymbol{\Sigma}^{-1}(n-1)\,\mathbf{S} + n(\bar{\mathbf{x}} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu})\right] \right\}$$

$$= h(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) g\left(\bar{\mathbf{x}}, \mathbf{S}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$$

   Por teorema de factorización de Neyman - Pearson, $\overline{\boldsymbol{X}}$ y $\boldsymbol{S}$ son estadísticos sufientes para $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$, respectivamente.

Estimación de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$

Dada la **aleatoriedad** de la **muestra**, debemos estudiar el **comportamiento de las estadísticas** $\overline{\mathbf{X}}$ y $\mathbf{S}$, así que debemos estudiar su distribución que es llamada **distribución de muestreo**.

A continuación se presentaran algunos **resultados** importantes que serán útiles para hacer inferencias.

Estimación de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$

Distribución Muestral de $\overline{\mathbf{X}}$ y $\mathbf{S}$

Si $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ es una muestra aleatoria de una población $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Entonces,

1. $\overline{\mathbf{X}} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$
   i. $\mathrm{E}\left[\overline{\mathbf{X}}\right] = \boldsymbol{\mu}$, esto es, $\overline{\mathbf{X}}$ es un estimador **insesgado** para $\boldsymbol{\mu}$.
   ii. $\mathrm{Var}\left[\overline{\mathbf{X}}\right] = \frac{\boldsymbol{\Sigma}}{n}$.
2. $(n-1)\,\mathbf{S}$ tiene una distribución **Wishart**[a] con n-1 grados de libertad.

---

[a] una **generalización** de la distribución **chi-cuadrado**: si $\mathbf{Z}_j \overset{iid}{\sim} \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$, entonces $W = \sum_{j=1}^{n} \mathbf{Z}_j \mathbf{Z}_j^{\top} \sim W_{(n-1)}$.

Pontificia Universidad JAVERIANA Bogotá

Estimación de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$

3. $\overline{\mathbf{X}}$ y $\mathbf{S}$ son estadísticamente **independientes**.
4. $\mathrm{E}\left[\frac{n-1}{n}\mathbf{S}\right] = \boldsymbol{\Sigma}$, es decir, $\frac{n-1}{n}\mathbf{S}$ es un estimador **insesgado** para $\boldsymbol{\Sigma}$.
5. $\overline{\mathbf{X}}$ y $\mathbf{S}$ son estimadores insesgados para $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ con matriz de covarianza mínima, es decir, son los **estimadores insesgados más precisos**.

Estimación de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$

3. $\overline{\mathbf{X}}$ y $\mathbf{S}$ son estadísticamente **independientes**.
4. $\mathrm{E}\left[\frac{n-1}{n}\mathbf{S}\right] = \boldsymbol{\Sigma}$, es decir, $\frac{n-1}{n}\mathbf{S}$ es un estimador **insesgado** para $\boldsymbol{\Sigma}$.
5. $\overline{\mathbf{X}}$ y $\mathbf{S}$ son estimadores insesgados para $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ con matriz de covarianza mínima, es decir, son los **estimadores insesgados más precisos**.

Estos resultados son **válidos** bajo **normal multivariada**.

Estimación de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$

3. $\overline{\mathbf{X}}$ y $\mathbf{S}$ son estadísticamente **independientes**.
4. $\mathrm{E}\left[\frac{n-1}{n}\mathbf{S}\right] = \boldsymbol{\Sigma}$, es decir, $\frac{n-1}{n}\mathbf{S}$ es un estimador **insesgado** para $\boldsymbol{\Sigma}$.
5. $\overline{\mathbf{X}}$ y $\mathbf{S}$ son estimadores insesgados para $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ con matriz de covarianza mínima, es decir, son los **estimadores insesgados más precisos**.

Estos resultados son **válidos** bajo **normal multivariada**.

¿Y si $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ **no proviene** de una $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$?

Estimación de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$

3. $\overline{\mathbf{X}}$ y $\mathbf{S}$ son estadísticamente **independientes**.
4. $\mathrm{E}\left[\frac{n-1}{n}\mathbf{S}\right] = \boldsymbol{\Sigma}$, es decir, $\frac{n-1}{n}\mathbf{S}$ es un estimador **insesgado** para $\boldsymbol{\Sigma}$.
5. $\overline{\mathbf{X}}$ y $\mathbf{S}$ son estimadores insesgados para $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ con matriz de covarianza mínima, es decir, son los **estimadores insesgados más precisos**.

Estos resultados son **válidos** bajo **normal multivariada**.

¿Y si $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ **no proviene** de una $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$?

Recurrimos a la **teoría asintótica**!

Estimación de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$

Distribución Asintótica de $\overline{\mathbf{X}}$ y $\mathbf{S}$

Si $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ es una muestra aleatoria de una **población** (**cualquiera**[a]) $p-$variada con vector de medias $\boldsymbol{\mu}$ y matriz de varianzas y covarianzas $\boldsymbol{\Sigma}$ definida positiva. Entonces para $\underline{n > p}$ **suficientemente grande** tenemos que:

1. **Teorema del límite central** (TLC):

$$\sqrt{n}\left(\overline{\mathbf{X}} - \boldsymbol{\mu}\right) \overset{\text{aprox.}}{\sim} \mathrm{N}_p\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$$

2. $\overline{\mathbf{X}}$ **converge en probabilidad** a $\boldsymbol{\mu}$, esto significa que con una **alta probabilidad** $\overline{\mathbf{X}}$ **estará cerca** a $\boldsymbol{\mu}$.

[a] Variables con naturaleza diferente (discretas, continuas, mixtas)

Estimación de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$

3. **S converge en probabilidad** a $\boldsymbol{\Sigma}$, esto significa que para tamaños de **muestras grandes**, **S** se **aproximará** a $\boldsymbol{\Sigma}$.
4. Como **S converge en probabilidad** a $\boldsymbol{\Sigma}$, entonces
   i. $\sqrt{n}\left(\overline{\mathbf{X}} - \boldsymbol{\mu}\right) \overset{\text{aprox.}}{\sim} \mathrm{N}_p\left(\mathbf{0}, \mathbf{S}\right)$
   ii. $n\left(\overline{\mathbf{X}} - \boldsymbol{\mu}\right)^\top \mathbf{S}^{-1}\left(\overline{\mathbf{X}} - \boldsymbol{\mu}\right) \overset{\text{aprox.}}{\sim} \chi^2_p$

Observación: no se requiere de normalidad multivariada para que exista la convergencia. Solamente se necesita que exista el vector de medias poblacional y la matriz de covarianzas poblacional.

Inferencia Sobre μ

Inferencia Sobre $\boldsymbol{\mu}$

Sea

$$\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$$

una **muestra aleatoria** de una población $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Suponga que el **interés** es verificar

$$\begin{aligned} \mathrm{H}_0 &: \boldsymbol{\mu} = \boldsymbol{\mu}_0 \\ \mathrm{H}_1 &: \boldsymbol{\mu} \neq \boldsymbol{\mu}_0 \end{aligned}$$

donde

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \qquad \boldsymbol{\mu}_0 = \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \vdots \\ \mu_{0p} \end{bmatrix}$$

donde $\boldsymbol{\mu}_0$ es **dado (conocido)**.

Usted podría pensar en hacer **pruebas** t **univariadas** (para cada media), esto es

$$\begin{aligned} \mathrm{H}_{0i} &: \mu_i = \mu_{0i} \\ \mathrm{H}_{1i} &: \mu_i \neq \mu_{0i} \end{aligned}$$

$i = 1, 2, \ldots, p$, cuya **estadística de prueba** está dada por

$$t_i^2 = \frac{(\overline{x}_i - \mu_{0i})^2}{s_{ii}/n} = n(\overline{x}_i - \mu_{0i})(s_{ii})^{-1}(\overline{x}_i - \mu_{0i}) \overset{\mathrm{H}_0\, verd}{\sim} t_{(n-1)}$$

Inferencia Sobre $\boldsymbol{\mu}$

y cuyo criterio de decisión es **rechazar** H_{0i} a un **nivel de significancia** α si

$$\frac{(\overline{x}_i - \mu_{0i})^2}{s_{ii}/n} = n(\overline{x}_i - \mu_{0i})(s_{ii})^{-1}(\overline{x}_i - \mu_{0i}) > t_{(n-1)}(\alpha/2)$$

donde $t_{(n-1)}(\alpha/2)$ es el **cuantil** $\alpha/2$ **superior** de la distribución **t-student** con $n-1$ grados de libertad.

Si H_0 **no es rechazada**, se concluye que μ_{0i} es **un valor plausible para** μ_i, $i = 1, 2, \ldots, p$.

Inferencia Sobre $\boldsymbol{\mu}$

Example

Los datos del archivo `Mujeres.xlsx` contiene la información de la transpiración de 20 mujeres saludables. Se midieron las siguientes 3 componentes: tasa de sudoración (X_1), contenido de sodio (X_2) y contenido de potacio (X_3).

Use un nivel de significancia del 10 % para probar

$$\begin{aligned} \mathrm{H}_0 &: \boldsymbol{\mu} = \boldsymbol{\mu}_0 \\ \mathrm{H}_1 &: \boldsymbol{\mu} \neq \boldsymbol{\mu}_0 \end{aligned}$$

donde $\boldsymbol{\mu} = [\mu_1,\ \mu_2,\ \mu_3]^\top$ y $\boldsymbol{\mu}_0 = [4,\ 50,\ 10]^\top$.

Observación: Example 5.2 de Johnson and Wichern (2013), Applied Multivariate Statistical Analysis, pp. 214

Inferencia Sobre μ

En el caso **univariado**:

$$\begin{array}{lll} H_{01}: \mu_1 = 4 & H_{02}: \mu_2 = 50 & H_{03}: \mu_3 = 10 \\ H_{11}: \mu_1 \neq 4 & H_{12}: \mu_2 \neq 50 & H_{13}: \mu_3 \neq 10 \end{array}$$

tenemos

$$p - \text{valor}_1 = 0.108, \quad p - \text{valor}_2 = 0.1619, \quad p - \text{valor}_3 = 0.9354$$

Observe que **en todos los casos** $p - \text{valor} < \alpha$, así que para un nivel de significancia del 10 % **se concluye** que los **valores** $4, 50, 10$ **son plausibles** para μ_1, μ_2 y μ_3, respectivamente.

Inferencia Sobre μ

En el caso **univariado**:

$$\begin{array}{lll} \mathrm{H}_{01} : \mu_1 = 4 & \mathrm{H}_{02} : \mu_2 = 50 & \mathrm{H}_{03} : \mu_3 = 10 \\ \mathrm{H}_{11} : \mu_1 \neq 4 & \mathrm{H}_{12} : \mu_2 \neq 50 & \mathrm{H}_{13} : \mu_3 \neq 10 \end{array}$$

tenemos

$$p - \mathrm{valor}_1 = 0.108, \quad p - \mathrm{valor}_2 = 0.1619, \quad p - \mathrm{valor}_3 = 0.9354$$

Observe que **en todos los casos** $p - \mathrm{valor} < \alpha$, así que para un nivel de significancia del $10\,\%$ **se concluye** que los **valores** $4, 50, 10$ **son plausibles** para μ_1, μ_2 y μ_3, respectivamente. El problema es que **pueden existir otros valores** de μ **que tambén pueden ser consistentes con los datos**.

Inferencia Sobre μ

```
# ----- Example 5.2 del libro pp 214 ----- #
require(tidyverse)
Mujeres <- read_excel("D:/Desktop/Mujeres.xlsx")
X<-Mujeres %>% as.data.frame()

# ----- PH univariadas ----- #
t.test(X$Azucar,mu=4,conf.level = 0.90)
t.test(X$Sodio,mu=50,conf.level = 0.90)
t.test(X$Potacio,mu=10,conf.level = 0.90)
```

Inferencia Sobre $\boldsymbol{\mu}$

En el caso **multivariado**, el problema es **determinar** si (**en conjunto**) el vector $\boldsymbol{\mu}_0$ dado **es un valor plausible** para $\boldsymbol{\mu}$:

$$\begin{aligned} \mathrm{H}_0 &: \boldsymbol{\mu} = \boldsymbol{\mu}_0 \\ \mathrm{H}_1 &: \boldsymbol{\mu} \neq \boldsymbol{\mu}_0 \end{aligned}$$

Para evaluar esas hipótesis, la **estadística de prueba** está dada por

$$T^2 = \sqrt{n}\left(\overline{\mathbf{X}} - \boldsymbol{\mu}_0\right)^\top \mathbf{S}^{-1} \sqrt{n}\left(\overline{\mathbf{X}} - \boldsymbol{\mu}_0\right) \overset{\mathrm{H}_0 \text{ verd.}}{\sim} \frac{(n-1)p}{n-p} F_{p,n-p}$$

La cual es conocida como **Estadística** T^2 **de Hotelling** y es una **generalización** de la **prueba** t **univariada**.

Inferencia Sobre $\boldsymbol{\mu}$

Observación:

$$T^2 = \sqrt{n}\left(\overline{\mathbf{X}} - \boldsymbol{\mu}_0\right)^\top \mathbf{S}^{-1} \sqrt{n}\left(\overline{\mathbf{X}} - \boldsymbol{\mu}_0\right)$$

$$= \sqrt{n}\left(\overline{\mathbf{X}} - \boldsymbol{\mu}_0\right)^\top \left[\frac{(n-1)\,\mathbf{S}}{n-1}\right]^{-1} \sqrt{n}\left(\overline{\mathbf{X}} - \boldsymbol{\mu}_0\right)$$

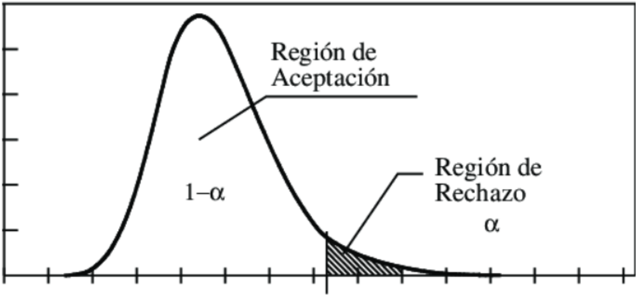
$$= \left[\mathrm{N}_p\left(0, \boldsymbol{\Sigma}\right)\right]^\top \left[\frac{W_{p,n-1}}{n-1}\right]^{-1} \mathrm{N}_p\left(0, \boldsymbol{\Sigma}\right)$$

Recuerde que en el caso **univariado**

$$t^2 = \frac{(\overline{x} - \mu_0)^2}{s^2/n} = n(\overline{x} - \mu_0)(s^2)^{-1}(\overline{x} - \mu_0)$$

Inferencia Sobre $\boldsymbol{\mu}$

$$T^2 = \sqrt{n}\left(\overline{\mathbf{X}} - \boldsymbol{\mu}_0\right)^\top \mathbf{S}^{-1} \sqrt{n}\left(\overline{\mathbf{X}} - \boldsymbol{\mu}_0\right) \overset{\mathrm{H}_0\,Verd}{\sim} \frac{(n-1)p}{n-p} F_{p,n-p}$$



$$\frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)$$

Inferencia Sobre $\boldsymbol{\mu}$

Rechazamos H_0 a un **nivel de significancia** α si

$$T^2_{obs} = \sqrt{n}\left(\overline{\mathbf{X}} - \boldsymbol{\mu}_0\right)^\top \mathbf{S}^{-1}\sqrt{n}\left(\overline{\mathbf{X}} - \boldsymbol{\mu}_0\right) > \frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)$$

donde $F_{p,n-p}(\alpha)$ es el cuantil $\alpha-$superior de la distribución $F_{p,n-p}$.

Inferencia Sobre $\boldsymbol{\mu}$

Para los datos de `Mujeres.xlsx` tenemos que

$$T^2 = 9.6644 > 8.1726 = \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)$$

Por lo tanto, se **rechaza**

$$\mathrm{H}_0 : \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 50 \\ 10 \end{bmatrix}$$

a un **nivel de significancia** del 10 %. Luego, hay suficiente evidencia en la muestra para concluir que $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$

Inferencia Sobre $\boldsymbol{\mu}$

Observaciones:

- *Con las pruebas **univariadas** los valores dados fueron plausibles (**no se rechazó** H_0), mientras que en la **multivariada** no lo fueron (**se rechazó** H_0).*
- H_0 *será rechazada si una o más de las componentes difiere mucho de los valores hipotéticos $\boldsymbol{\mu}_0$. En este punto no se tiene idea de cuales de estos valores hipotéticos no son soportados por los datos.*
- *Para que la prueba sea válida es necesario que los datos provengan de una distribución normal multivariada.*

Inferencia Sobre μ

```
# --- PH multivariadas --- #
T2<-function(mu0,alpha,n,p){
     Xbarra<-colMeans(X)
     S<-cov(X)
     InvS<-solve(S)
     DifMed<-Xbarra - mu0
     T2<-n%*%t(DifMed)%*%InvS%*%DifMed
     return(T2)
                         }

   T2(c(4,50,10),0.10,20,3)
         [,1]
[1,] 9.738773
```

Inferencia Sobre μ

```
# --- Valor critico

alpha=0.10; p=3; n=20
qf<-qf(alpha,p,n-p,lower.tail = F)
round(qf,4)
[1] 2.4374

Vc<- (((n-1)*p)/(n-p))*qf
round(Vc,4)
[1] 8.1726
```

Inferencia Sobre μ

Recuerde que en el caso **univariado**, cuando la muestra aleatoria provenía de una población con fdp $f_X(x|\theta)$, $\theta \in \Theta$, y el interés era probar

$$\begin{aligned} \mathrm{H}_0 &: \ \theta \in \Theta_0 \\ \mathrm{H}_1 &: \ \theta \in \Theta_0^c \end{aligned}$$

donde $\Theta_0 \subset \Theta$ y Θ_0^c es el complemento de Θ_0, usabamos la **prueba de razón de verosimilitud** *(LRT, Likelihood Ratio Test)*:

$$\Lambda(\mathbf{x}) = \frac{\sup_{\Theta_0} L(\theta|\,\mathbf{x})}{\sup_{\Theta} L(\theta|\,\mathbf{x})}$$

que **proporciona estadísticas que se reducen a los estadísticos** F y t.

Inferencia Sobre μ

La **LRT** tiene varias **propiedades óptimas** para **muestras** razonablemente **grandes**. En el caso de **normalidad multivariada** el **método** es particularmente **conveniente**.

Inferencia Sobre $\boldsymbol{\mu}$

La **LRT** tiene varias **propiedades óptimas** para **muestras** razonablemente **grandes**. En el caso de **normalidad multivariada** el **método** es particularmente **conveniente**.

Suponga $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ es una muestra aleatoria de una $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y que se quiere probar

$$\begin{aligned} \mathrm{H}_0 &: \boldsymbol{\mu} = \boldsymbol{\mu}_0 \\ \mathrm{H}_1 &: \boldsymbol{\mu} \neq \boldsymbol{\mu}_0 \end{aligned}$$

La **estadística de la razón de verosimilitud** queda dada por

$$\Lambda = \frac{\text{máx}_{\boldsymbol{\Sigma}} L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})}{\text{máx}_{\boldsymbol{\mu},\boldsymbol{\Sigma}} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})} = \left[\frac{\left|\widehat{\boldsymbol{\Sigma}}\right|}{\left|\widehat{\boldsymbol{\Sigma}}_0\right|} \right]^{n/2}$$

Inferencia Sobre $\boldsymbol{\mu}$

donde

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{x}_j - \overline{\mathbf{x}})(\mathbf{x}_j - \overline{\mathbf{x}})^\top$$

$$\widehat{\boldsymbol{\Sigma}}_0 = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{x}_j - \boldsymbol{\mu}_0)(\mathbf{x}_j - \boldsymbol{\mu}_0)^\top$$

Cuando n es **grande** y bajo condiciones de regularidad,

$$-2\ln\Lambda \overset{\mathrm{H}_0 \text{ verd.}}{\sim} \chi^2_{(v-v_0)}$$

donde

$$v - v_0 = \dim(\boldsymbol{\Theta}) - \dim(\boldsymbol{\Theta_0}) = \textbf{No. de restricciones impuestas,}$$

siendo $\mathbf{\Theta}$ el **espacio paramétrico** de

$$\boldsymbol{\theta}^\top = [\mu_1\, \mu_2\, \cdots\, \mu_p\, , \sigma_{11}\, \sigma_{12}\, \ldots, \sigma_{1p}\, \sigma_{22}, \ldots\, \sigma_{2p}, \ldots, \sigma_{p-1,1} \ldots, \sigma_{pp}]$$

Observación: $\mathbf{\Theta}$ *tiene dimensión* $p + p(p+1)/2$.

La estadística

$$\Lambda^{2/n} = \frac{\left|\widehat{\mathbf{\Sigma}}\right|}{\left|\widehat{\mathbf{\Sigma}}_0\right|} = \left[1 + \frac{T^2}{n-1}\right]^{-1}$$

es llamada **Lambda de Wilks**. Entonces, se **rechaza** H_0 para **valores pequeños** de Λ o para **valores** de T^2 **grandes**.

Inferencia Sobre μ

Recuerde que las **pruebas son evaluadas/comparadas con base en la función potencia**:

$$\beta(\theta) = P_\theta[\mathbf{X} \in \mathcal{R}]$$

$$= \begin{cases} \text{Prob. Error. Tipo I} & , \text{si } \theta \in \Theta_0 \\ \\ 1 - \text{Prob. Error. Tipo II} & , \text{si } \theta \in \Theta_0^c \end{cases}$$

$$= \begin{cases} \alpha & , \text{si } \theta \in \Theta_0 \\ 1 - \beta & , \text{si } \theta \in \Theta_0^c \end{cases}$$

donde $\mathcal{R}$ es la región de rechazo de H_0.

Inferencia Sobre μ

Lo ideal es que

$$P_\theta[\mathbf{X} \in \mathcal{R}] = \begin{cases} 0 & \text{, si } \theta \in \Theta_0 \\ 1 & \text{, si } \theta \in \Theta_0^c \end{cases}$$

esto es

- **niveles de significancia**

$$\alpha = P\,[\,\text{Prob. Error. Tipo I}\,]$$

$$= P\,[\text{Rechazar}\, \mathsf{H}_0 \text{ cuando } \mathsf{H}_0 \text{ verdadera}]$$

pequeños (cercanos a 0).

Inferencia Sobre μ

- **altas potencias** (cercanas a 1)

$$1-\beta = 1 - P\,[\,\text{Prob. Error. Tipo II}\,]$$

$$= P[\text{Rechazar}\, \mathsf{H}_0 \text{ cuando } \mathsf{H}_0 \text{ falsa}]$$

Inferencia Sobre μ

- **altas potencias** (cercanas a 1)

$$1 - \beta = 1 - P\,[\,\text{Prob. Error. Tipo II}\,]$$

$$= P[\text{Rechazar}\,\mathsf{H}_0 \text{ cuando } \mathsf{H}_0 \text{ falsa}]$$

En general,

- en el caso **uniparamétrico**, la **LRT** es la **más potente** entre todas aquellas que tienen el mismo nivel de significancia.
- en el caso **multivariado**, la prueba **LRT** tiene **mayor potencia** (promedio) posible cuando el tamaño muestral es grande.

Inferencia Sobre $\boldsymbol{\mu}$

En la situación **multivariada**, **generalmente no es suficiente** con probar la hipótesis

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0,$$

donde todas las componentes del vector de medias son especificadas bajo la hipótesis nula. Con frecuencia **es preferible encontrar regiones** de μ **con valores que son plausibles a la luz de los datos observados**. Estas regiones son llamadas **regiones de confianza**.

Inferencia Sobre $\boldsymbol{\mu}$

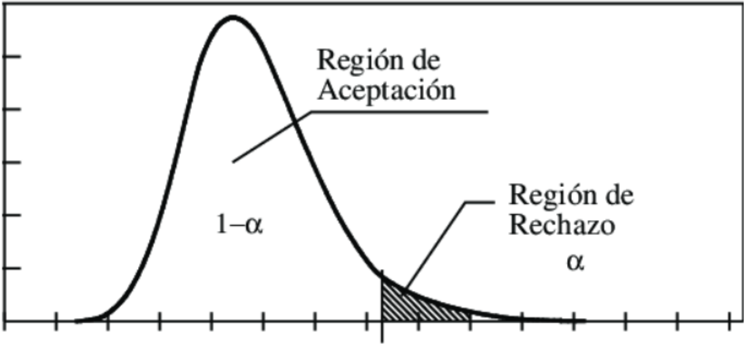
Una región de confianza para un vector de parámetro $\boldsymbol{\theta}$, es un **conjunto que contiene los posibles valores** de $\boldsymbol{\theta}$ y está determinada por los datos.

Región de Confianza

Se dice que $R(\mathbf{X})$ es una región con $(1-\alpha)100\,\%$ de confianza para $\boldsymbol{\theta}$, **antes de obtener la muestra**

$$P\left[R(\mathbf{X})\,\text{contenga al verdadero}\,\boldsymbol{\theta}\right] = 1-\alpha$$

Recuerde que



$$\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)$$

Inferencia Sobre $\boldsymbol{\mu}$

Para **construir una región de confianza** para $\boldsymbol{\mu}$ recuerde que:
Si $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ es una muestra aleatoria de una población $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$T^2 = n\left(\overline{\mathbf{X}} - \boldsymbol{\mu}\right)^\top \mathbf{S}^{-1}\left(\overline{\mathbf{X}} - \boldsymbol{\mu}\right) \sim \frac{(n-1)p}{n-p} F_{p,n-p}$$

Entonces, **antes de obtener la muestra**

$$P\left[n\left(\overline{\mathbf{X}} - \boldsymbol{\mu}\right)^\top \mathbf{S}^{-1}\left(\overline{\mathbf{X}} - \boldsymbol{\mu}\right) \leq \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)\right] = 1 - \alpha$$

Inferencia Sobre $\boldsymbol{\mu}$

Luego, para **una muestra dada**, el **conjunto de valores** de $\boldsymbol{\mu}$ que **satisfacen**

$$n\left(\bar{\mathbf{x}} - \boldsymbol{\mu}\right)^{\top} \mathbf{S}^{-1}\left(\bar{\mathbf{x}} - \boldsymbol{\mu}\right) \leq \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)$$

definen una región para $\boldsymbol{\mu}$ con $(1-\alpha)100\,\%$ de confianza.

Observación:

- *Geométricamente la **región** es una **elipsoide** centrada en $\bar{x}$.*
- *La **región** $(1-\alpha)100\,\%$ de confianza para $\boldsymbol{\mu}$ **contiene todos los posibles valores** $\boldsymbol{\mu}_0$ para los cuales la prueba T^2 **no rechaza** $\mathrm{H}_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ a un nivel de significancia α.*

Inferencia Sobre $\boldsymbol{\mu}$

Entonces, si **queremos determinar si un vector** μ_0 cualquiera **cae dentro de la región de confianza**, esto es, **si queremos verificar si** μ_0 **es un valor plausible para** μ, **comparamos la distancia generalizada**

$$n\left(\overline{\mathbf{x}}-\boldsymbol{\mu}_0\right)^{\top}\mathbf{S}^{-1}\left(\overline{\mathbf{x}}-\boldsymbol{\mu}_0\right)$$

con el cuantil

$$\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)$$

Si

$$n\left(\overline{\mathbf{x}}-\boldsymbol{\mu}_0\right)^{\top}\mathbf{S}^{-1}\left(\overline{\mathbf{x}}-\boldsymbol{\mu}_0\right)\leq\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)$$

$\boldsymbol{\mu}_0$ CAE en la región de confianza.

Inferencia Sobre $\boldsymbol{\mu}$

Para los datos de `Mujeres.xlsx` tenemos que

$$\overline{x} = \begin{bmatrix} 4.64 \\ 45.40 \\ 9.96 \end{bmatrix} \qquad \mathbf{S} = \begin{bmatrix} 2.879 & 10.010 & -1.809 \\ 10.010 & 199.788 & -5.640 \end{bmatrix}$$

$$\mathbf{S}^{-1} = \begin{bmatrix} 10.010 & 199.788 & -5.640 \\ -0.022 & 0.006 & -0.002 \\ 0.258 & -0.002 & 0.402 \end{bmatrix}$$

luego, la **elipse del 95 % de confianza** para

$$\boldsymbol{\mu} = [\mu_1, \mu_2, \mu_3]^\top$$

contiene todos los valores μ_1, μ_2, μ_3 que **satisfacen**

Inferencia Sobre $\boldsymbol{\mu}$

$$20[4.64-\mu_1, 45.40-\mu_2, 9.96-\mu_3]\begin{bmatrix}10.010 & 199.788 & -5.640\\ -0.022 & 0.006 & -0.002\\ 0.258 & -0.002 & 0.402\end{bmatrix}\begin{bmatrix}4.64-\mu_1\\ 45.40-\mu_2\\ 9.96-\mu_3\end{bmatrix} \leq \frac{19*3}{17}(2.44) = 8.174$$

¿Cae $\boldsymbol{\mu}_0 = [4, 50, 10]^\top$ **dentro de la elipse** del 90 % de confianza?.

$$20[4.64-\mu_1, 45.40-\mu_2, 9.96-\mu_3]\begin{bmatrix}10.010 & 199.788 & -5.640\\ -0.022 & 0.006 & -0.002\\ 0.258 & -0.002 & 0.402\end{bmatrix}\begin{bmatrix}4.64-\mu_1\\ 45.40-\mu_2\\ 9.96-\mu_3\end{bmatrix} \leq \frac{19*3}{17}(2.44) = 8.174$$

¿Cae $\boldsymbol{\mu}_0 = [4, 50, 10]^\top$ **dentro de la elipse** del 90 % de confianza?.

Para saber, simplemente, **remplazamos** $\mu_1 = 4$, $\mu_2 = 50$, $\mu_3 = 10$ en la **elipse anterior** y obtenemos 9.6644. Evidentemente,

$$9.6644 \not\leq 8.174$$

es decir, el **vector** $\boldsymbol{\mu}_0 = [4, 50, 10]^\top$ NO cae dentro de la región y por lo tanto se rechaza H_0. En consecuencia, $\boldsymbol{\mu}_0 = [4, 50, 10]^\top$ NO es un valor plausible para $\boldsymbol{\mu}$.

Inferencia Sobre μ

Las regiones de confianza proporcionan un conocimiento conjunto sobre los valores plausibles para μ. Generalmente, las **conclusiones en los estudios también suelen incluir aseveraciones sobre las componentes individuales del vector de medias**. Para hacerlo, necesitamos **construir intervalos individuales** para cada componente del vector de media de forma tal que simultáneamente cada uno de ellos contenga a su media bajo una probabilidad especificada. Esta clase de intervalos son llamados **intervalos de confianza simultáneos**.

Inferencia Sobre $\boldsymbol{\mu}$

Para **construir los intervalos simultáneos** tenga en cuenta que

$$\mu_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}_{1\times p} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}_{p\times 1} = \mathbf{a}^\top \boldsymbol{\mu}$$

$$\vdots$$

$$\mu_p = \begin{bmatrix} 0 & 0 & \cdots & 1 \end{bmatrix}_{1\times p} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}_{p\times 1} = \mathbf{a}^\top \boldsymbol{\mu}$$

Observación: $\mathbf{a}^\top$ *es un* ***vector*** *(fila)* ***canónico****.*

Inferencia Sobre $\boldsymbol{\mu}$

La $i-$**ésima media** puede ser escrita por

$$\mu_i = \begin{bmatrix} 0 & 0 & \cdots & 1 & 0 & \cdots 0 \end{bmatrix}_{1\times p} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_i \\ \vdots \\ \mu_p \end{bmatrix}_{p\times 1} = \mathbf{a}^\top \boldsymbol{\mu}$$

La $i-$**ésima media** puede ser escrita por

$$\mu_i = \begin{bmatrix} 0 & 0 & \cdots & 1 & 0 & \cdots 0 \end{bmatrix}_{1\times p} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_i \\ \vdots \\ \mu_p \end{bmatrix}_{p\times 1} = \mathbf{a}^\top \boldsymbol{\mu}$$

Considere que $\mathbf{X} \sim \mathrm{N}_p\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ y que

$$Z = a_1X_1 + a_2X_2 + \cdots + a_pX_p = \mathbf{a}^\top \mathbf{X}$$

Inferencia Sobre $\boldsymbol{\mu}$

por resultado visto antes sigue que

$$Z \sim N_1\left(\underbrace{\mathbf{a}^\top \boldsymbol{\mu}}_{\mu_Z}, \underbrace{\mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a}}_{\sigma_Z^2}\right)$$

Ahora bien, si $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ es una muestra aleatoria de $\mathbf{X}$,

$$Z_j = a_1 X_{j1} + a_2 X_{j2} + \cdots + a_p X_{jp} = \mathbf{a}^\top \mathbf{X}_j$$

y por lo tanto,

$$\overline{Z} = \mathbf{a}^\top \overline{\mathbf{X}} \qquad ; \qquad S_Z^2 = \mathbf{a}^\top \mathbf{S}\, \mathbf{a}$$

Si $\sigma_Z^2 = \mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a}$ es **deconocido**

$$t = \frac{\overline{Z} - \mu_Z}{S_Z/\sqrt{n}} = \sqrt{n}\frac{(\mathbf{a}^\top \overline{\mathbf{X}} - \mathbf{a}^\top \boldsymbol{\mu})}{\sqrt{\mathbf{a}^\top \mathbf{S}\, \mathbf{a}}} \sim t_{n-1}$$

Inferencia Sobre $\boldsymbol{\mu}$

Haciendo el mismo procedimiento para los **intervalos univariados** tenemos

$$\bar{z} \pm t_{(n-1)}(\alpha/2)\frac{S_z}{\sqrt{n}}$$

equivalentemente

$$\mathbf{a}^\top \bar{\mathbf{x}} \pm t_{(n-1)}(\alpha/2)\frac{\sqrt{\mathbf{a}^\top\,\mathbf{S}\,\mathbf{a}}}{\sqrt{n}}$$

Este intervalo puede ser interpretado como una afirmación sobre las componentes del vector $\boldsymbol{\mu}$.

Observación: $\boldsymbol{a}^\top\,\boldsymbol{S}\,\boldsymbol{a} = s_{ii}$.

Inferencia sobre $\boldsymbol{\mu}$

Por lo tanto,

$$\bar{x}_1 - t_{(n-1)}(\alpha/2)\sqrt{\frac{s_{11}}{n}} \leq \mu_1 \leq \bar{x}_1 + t_{(n-1)}(\alpha/2)\sqrt{\frac{s_{11}}{n}}$$

$$\bar{x}_2 - t_{(n-1)}(\alpha/2)\sqrt{\frac{s_{22}}{n}} \leq \mu_2 \leq \bar{x}_2 + t_{(n-1)}(\alpha/2)\sqrt{\frac{s_{22}}{n}}$$

$$\vdots$$

$$\bar{x}_p - t_{(n-1)}(\alpha/2)\sqrt{\frac{s_{pp}}{n}} \leq \mu_p \leq \bar{x}_p + t_{(n-1)}(\alpha/2)\sqrt{\frac{s_{pp}}{n}}$$

Inferencia sobre $\boldsymbol{\mu}$

Observación: Estos intervalos ignoran la estructura de covarianza de las p variables. Considere que $\mathbf{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *donde*

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \qquad \boldsymbol{\Sigma} \begin{bmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{pp} \end{bmatrix}$$

Observe que se está asumiendo que las variables son independientes. Entonces, si I_i denota al intervalo aleatorio para μ_i:

$$P\left[\mu_1 \in I_1, \mu_2 \in I_2, \ldots, \mu_p \in I_p\right] = \prod_{i=1}^{p} (1-\alpha) = (1-\alpha)^p$$

Inferencia sobre μ

Los tres intervalos para las medias de los datos de `Mujeres.xlsx`:

```
# ----- IC simultaneos ----- #
IC1<-t.test(X$Azucar,mu=4,conf.level = 0.90)
IC2<-t.test(X$Sodio,mu=50,conf.level = 0.90)
IC3<-t.test(X$Potacio,mu=10,conf.level = 0.90)

L1<-IC1$conf.int[2]-IC1$conf.int[1]
L2<-IC2$conf.int[2]-IC2$conf.int[1]
L3<-IC3$conf.int[2]-IC3$conf.int[1]
```

Inferencia sobre $\boldsymbol{\mu}$

Variable	Intervalo	Longitud del IC
X_1	$[3.983912\,,\ 5.296088]$	1.3122
X_2	$[39.93489\,,\ 50.86511]$	10.9302
X_3	$[9.228578\,,\ 10.701422]$	1.4728

Observe que todos los intervalos contienen el respectivo valor de μ_0. En este caso

$$\begin{aligned} P\left[\mu_1 \in I_1, \mu_2 \in I_2, \mu_3 \in I_3\right] &= \prod_{i=1}^{3}(1-0.10) \\ &= (1-0.10)^3 \\ &= 0.729 \end{aligned}$$

la "confianza conjunta" disminuyó.

Inferencia sobre μ

Observación: La ***confianza*** $1-\alpha$ *asociada a cada intervalo de confianza es individual y no conjunta, y lo que* ***se quiere*** *es asociar un* ***confianza "colectiva"*** *de* $1-\alpha$ ***a todos*** *los intervalos de confianza que pueden ser generados por las diferentes elecciones de* ***a****. Al hacer esto, los* ***intervalos simultáneos son más amplios*** *(es decir,* ***menos precisos****) que el intervalo para una sola elección de* ***a***.

Inferencia sobre $\boldsymbol{\mu}$

Observación: La ***confianza*** $1-\alpha$ *asociada a cada intervalo de confianza es individual y no conjunta, y lo que* ***se quiere*** *es asociar un* ***confianza "colectiva"*** *de* $1-\alpha$ ***a todos*** *los intervalos de confianza que pueden ser generados por las diferentes elecciones de* $\boldsymbol{a}$*. Al hacer esto, los* ***intervalos simultáneos son más amplios*** *(es decir,* ***menos precisos****) que el intervalo para una sola elección de* $\boldsymbol{a}$.

¿Que hacemos?

Inferencia sobre $\boldsymbol{\mu}$

*Observación: La **confianza** $1-\alpha$ asociada a cada intervalo de confianza es individual y no conjunta, y lo que **se quiere** es asociar un **confianza "colectiva"** de $1-\alpha$ **a todos** los intervalos de confianza que pueden ser generados por las diferentes elecciones de **a**. Al hacer esto, los **intervalos simultáneos son más amplios** (es decir, **menos precisos**) que el intervalo para una sola elección de **a**.*

¿Que hacemos?

Intervalos simultáneos basados en la estadística T^2 de Hotelling.

Inferencia sobre $\boldsymbol{\mu}$

Para una muestra dada $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ y para todo $\mathbf{a} \in \mathbb{R}^p$ el intervalo:

$$\mathbf{a}^\top \overline{\mathbf{X}} \pm \sqrt{\frac{(n-1)p}{(n-p)n} F_{p,n-p}(\alpha) \mathbf{a}^\top \mathbf{S}\, \mathbf{a}}$$

contendrá $\mathbf{a}^\top \boldsymbol{\mu}$ con probabilidad $1-\alpha$. Estos intervalos simultáneos son llamados **intervalos** T^2.

Las sucesivas elecciones de $\mathbf{a}$:

$$\mathbf{a} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{a} = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \cdots, \mathbf{a} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

para los intervalos T^2, permite concluir que

Inferencia sobre $\boldsymbol{\mu}$

$$\bar{x}_1 - \sqrt{\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{11}}{n}} \leq \mu_1 \leq \bar{x}_1 - \sqrt{\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{11}}{n}}$$

$$\bar{x}_2 - \sqrt{\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{22}}{n}} \leq \mu_2 \leq \bar{x}_2 - \sqrt{\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{22}}{n}}$$

$$\vdots$$

$$\bar{x}_p - \sqrt{\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{pp}}{n}} \leq \mu_p \leq \bar{x}_p - \sqrt{\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{pp}}{n}}$$

Tienen una **confianza simultanea** de $1-\alpha$, la cuál no cambia para cualquier elección de **a**.

Inferencia sobre μ

```
# ----- IC simultaneos T2 ----- #
ICT2<-function(x,alpha,p,n){
                    qf<-qf(alpha,p,n-p,lower.tail = F)
                    Vc<- (((n-1)*p)/(n-p))*qf
                    xbar<-mean(x)
                    sii<-var(x)
                    linf<-xbar - sqrt(Vc*sii/n)
                    lsup<-xbar + sqrt(Vc*sii/n)
                    IC<-c(linf,lsup)
                    return(IC)
}
round(ICT2(X$Azucar,0.10,3,20),4)
```

Inferencia sobre $\boldsymbol{\mu}$

Para los datos de `Mujeres.xlsx`:

	Intervalo	**Longitud del IC**
μ_1	[3.5553 , 5.7247]	2.1694
μ_2	[36.3646 , 54.4354]	18.0708
μ_3	[8.7475 , 11.1825]	2.1435

Observe que todos los intervalos contienen el respectivo valor de μ_0. Además que la **longitud** de éstos es **superior** a los **intervalos simultaneos**.

Observación: Los intervalos simultáneos T^2 son más amplios que los intervalos univariados debido a que los 3 intervalos son válidos para una misma confianza del 90 %.

Inferencia sobre $\boldsymbol{\mu}$

Si **comparamos** las dos **estructuras** en **los dos intervalos**:

$$\bar{x}_1 \pm t_{(n-1)}(\alpha/2)\frac{\sqrt{s_{11}}}{n} \qquad ; \qquad \bar{x}_1 \pm \sqrt{\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{11}}{n}}$$

vemos que **difieren** en sus **valores críticos**:

$$t_{(n-1)}(\alpha/2) \qquad ; \qquad \sqrt{\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)}$$

Note que para $1-\alpha = 0.90$, $n = 20$ y $p = 3$:

$$t_{(n-1)}(\alpha/2) = 1.7291 \qquad ; \qquad \sqrt{\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)} = 8.1726$$

Inferencia sobre μ

por lo tanto, **los intervalos simultaneos** T^2 son

$$\left|\frac{8.1726 - 1.7291}{1.7291}\right| \times 100\,\% = 372.6505\,\%$$

más amplios que los intervalos **simultaneos** derivados de la t.

Inferencia sobre μ

por lo tanto, **los intervalos simultaneos** T^2 son

$$\left|\frac{8.1726 - 1.7291}{1.7291}\right| \times 100\,\% = 372.6505\,\%$$

más amplios que los intervalos **simultaneos** derivados de la t.

En general,

n	$t_{n-1}(0.025)$	$\sqrt{\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)}$	
		$p=4$	$p=10$
15	2.145	4.14	11.52
25	2.064	3.60	6.39
50	2.010	3.31	5.05
100	1.970	3.19	4.61
∞	1.96	3.08	4.28

Inferencia sobre μ

De acuerdo con la tabla anterior[1], la **amplitud** de los intervalos T^2 con respecto a los intervalos t,

- incrementa para n fijo cuando p incrementa,
- disminuye para p fijo a medida que n incrementa.

[1] Table 5.3 de Johnson and Wichern (2013), Applied Multivariate Statistical Analysis, pp. 231

Inferencia sobre μ

De acuerdo con la tabla anterior[1], la **amplitud** de los intervalos T^2 con respecto a los intervalos t,

- incrementa para n fijo cuando p incrementa,
- disminuye para p fijo a medida que n incrementa.

Existe **una alernativa** a los intervalos de confianza simultáneos T^2 que proporciona **mayor precisión (más cortos)**, éstos son llamados **intervalos de Bonferroni**.

[1] Table 5.3 de Johnson and Wichern (2013), Applied Multivariate Statistical Analysis, pp. 231

Inferencia sobre $\boldsymbol{\mu}$

Suponga que **antes de obtener los datos**, se quieren **construir intervalos simultáneos** para

$$\mathbf{a}_1^\top\boldsymbol{\mu}, \mathbf{a}_2^\top\boldsymbol{\mu}, \ldots, \mathbf{a}_m^\top\boldsymbol{\mu}$$

y que C_i denota al intervalo $(1-\alpha_i)100\,\%$ de confianza que contiene $\mathbf{a}_i^\top\boldsymbol{\mu}$, $i = 1, \ldots, m$, esto es

$$P\left[C_i \,\text{contiene}\; \mathbf{a}_i^\top\boldsymbol{\mu}\right] = 1-\alpha_i \qquad i = 1, 2, \ldots, m$$

Luego

$$\begin{aligned} P\left[\text{todos los } C_i \text{ contengan}\, \mathbf{a}_i^\top\boldsymbol{\mu}\right] &= P\left[\mathbf{a}_1^\top\boldsymbol{\mu} \in C_1, \ldots, \mathbf{a}_m^\top\boldsymbol{\mu} \in C_m\right] \\ &= 1 - P[\text{al menos un } C_i \text{ no contiene } \mathbf{a}_i^\top\boldsymbol{\mu}] \end{aligned}$$

Inferencia sobre $\boldsymbol{\mu}$

$$
\begin{aligned}
P\left[\text{todos los } C_i \text{ contengan } \mathbf{a}_i^\top \boldsymbol{\mu}\right] &\geq 1 - \sum_{i=1}^{m} P\left[\mathbf{a}_i^\top \boldsymbol{\mu} \notin C_i\right] \\
&= 1 - \sum_{i=1}^{m} \alpha_i \\
&= 1 - (\alpha_1 + \cdots + \alpha_m)
\end{aligned}
$$

Observación: $\alpha_1 + \cdots + \alpha_m$ *le permite al investigador* ***controlar la tasa de error global*** *independiente de la estructura de covarianza.*

Denotando por $\alpha_i = \frac{\alpha}{m}$ $i = 1, \ldots, m$, tenemos

$$P\left[\text{todos los } C_i \text{ contengan } \mathbf{a}_i^\top \boldsymbol{\mu}\right] \geq 1 - \underbrace{\left(\frac{\alpha_1}{m} + \cdots + \frac{\alpha_m}{m}\right)}_{\text{m - términos}}$$

$$= 1 - \alpha$$

Por tanto, con un nivel de confianza global mayor o igual a $1 - \alpha$, los siguientes intervalos simultáneos son válidos para las $m = p$ medias:

Inferencia sobre $\boldsymbol{\mu}$

$$\overline{x}_1 - t_{(n-1)}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{11}}{n}} \leq \mu_1 \leq \overline{x}_1 + t_{(n-1)}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{11}}{n}}$$

$$\overline{x}_2 - t_{(n-1)}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{22}}{n}} \leq \mu_2 \leq \overline{x}_2 + t_{(n-1)}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{22}}{n}}$$

$$\vdots$$

$$\overline{x}_p - t_{(n-1)}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{pp}}{n}} \leq \mu_p \leq \overline{x}_p + t_{(n-1)}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{pp}}{n}}$$

Inferencia sobre μ

Para los datos de `Mujeres.lxsx` los **intervalos de Bonferroni** son

	Intervalo	**Longitud**
μ_1	[3.77 , 5.51]	1.74
μ_2	[38.15 , 52.65]	14.5
μ_3	[8.99 , 10.94]	1.95

Observe que estos intervalos son **más cortos** que los simultáneos T^2.

[2] Table 5.4 de Johnson and Wichern (2013), Applied Multivariate Statistical Analysis, pp. 234

Inferencia sobre $\boldsymbol{\mu}$

Para los datos de `Mujeres.lxsx` los **intervalos de Bonferroni** son

	Intervalo	**Longitud**
μ_1	[3.77 , 5.51]	1.74
μ_2	[38.15 , 52.65]	14.5
μ_3	[8.99 , 10.94]	1.95

Observe que estos intervalos son **más cortos** que los simultáneos T^2.

En la siguiente tabla [2] se muestra la proporción

$$\frac{\text{Cuantil Int. Simul. Bonferroni}}{\text{Cuantil Int. Simul.}\, T^2} = \frac{t_{(n-1)}\left(\frac{\alpha}{2p}\right)}{\sqrt{\frac{(n-1)p}{(n-p)n} F_{p,n-p}(\alpha)}}$$

[2]Table 5.4 de Johnson and Wichern (2013), Applied Multivariate Statistical Analysis, pp. 234

Inferencia sobre μ

en que éstos son **más cortos** que los intervalos **simultáneos** T^2 para algunos valores de n y p y $1 - \alpha = 0.95$.

n	$m = p$		
	2	4	10
15	0.88	0.69	0.29
25	0.90	0.75	0.48
50	0.91	0.78	0.58
100	0.91	0.80	0.62
∞	0.91	0.81	0.66

Como los intervalos simultáneos de **Bonferroni** son más fáciles de aplicar y son relativamente **más cortos** para hacer inferencia, **en la práctica**, generalmente **son preferidos**.

Inferencia sobre μ

```
# ----- Intervalos Bonferroni ----- #
ICB<-function(x,alpha,p,n){
  xbarra<-mean(x)
  sii<-var(x)
  tc<-qt(alpha/(2*p),n-1,lower.tail = F)
  error<-tc*sqrt(sii/n)
  linf<-xbarra - error
  lsup<-xbarra + error
  ICB<-c(linf,lsup)
  return(ICB)
                        }
ICB<-ICB(X$Potacio,0.10,3,20)
l<-round(ICB[2] - ICB[1], 2)
```

Inferencia sobre $\boldsymbol{\mu}$

Lectura: Sección 5.5 de Johnson and Wichern (2013), Applied Multivariate Statistical Analysis, pp. 234.

Inferencia sobre $\boldsymbol{\mu}$

Lectura: Sección 5.5 de Johnson and Wichern (2013), Applied Multivariate Statistical Analysis, pp. 234.

Nota: algunos investigadores piensan que si se calculan los intervalos separados solamente cuando la prueba T^2 rechaza la hipótesis nula, ellos pueden representar con más precisión la información sobre las medias que los intervalos T^2.

Inferencia sobre $\boldsymbol{\mu}$

Como en los **intervalos simultaneos** T^2 el **coeficiente de confianza** $1-\alpha$ **no cambia** para cualquier elección de **a**, se pueden evaluar las **combinaciones lineales** de las componentes μ_i que **ameriten inspección** basados sobre un examen de los datos.

Los **intervalos** T^2 para las **diferencias** $\mu_i - \mu_k$ están dados por

$$\bar{x}_i - \bar{x}_k \pm \sqrt{\frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{ii} + s_{kk} - 2s_{ik}}{n}}$$

Observación: Éstos intervalos son para la ***misma población****.*

Inferencia sobre μ

Para los datos de `Mujeres.xlsx` tenemos

	Intervalo
$\mu_1 - \mu_2$	[-49.40 , -32.12]
$\mu_1 - \mu_3$	[-7.36 , -3.29]
$\mu_2 - \mu_3$	[26.07 , 44.80]

Es claro que:

- $\mu_1 - \mu_2 < 0 \Longrightarrow \mu_1 < \mu_2$
- $\mu_1 - \mu_3 < 0 \Longrightarrow \mu_1 < \mu_3$
- $\mu_2 - \mu_3 > 0 \Longrightarrow \mu_2 < \mu_3$

Inferencia sobre μ

```
# ----- IC simultaneos T2 para comparaciones de medias
ICT2<-function(xi,xk,alpha,n,p){
  qf<-qf(alpha,p,n-p,lower.tail = F)
  vc<- (((n-1)*p)/(n-p))*qf
  xibar<-mean(xi)
  xkbar<-mean(xk)
  sii<-var(xi)
  skk<-var(xk)
  sik<-cov(xi,xk)
  DE<-sqrt((sii + skk - 2*sik)/n)
  linf<-xibar-xkbar - sqrt(vc)*DE
  lsup<-xibar-xkbar + sqrt(vc)*DE
  IC<-c(linf,lsup)
  return(IC)
}
```