

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**

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

Observación:

las X_i 's son **variables aleatorias** y x_i 's **realizaciones** de X_i .



Introducción

Sabemos que si

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

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

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

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

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



Introducción

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

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

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

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

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



Introducción

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

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 \mathbf{X} .



Introducción

Evidentemente que

- ✓ μ y Σ son **desconocidos**
- ✓ el interés es **estimarlos**
- ✓ se quiere **hacer inferencia** sobre ellos.



Estimación de μ y Σ

Estimación de μ

Los **estimadores de máxima verosimilitud** (*MLE, Maximum Likelihood Estimator*) de μ y Σ son aquellos que **maximizan** a

$$L(\mu, \Sigma | \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$$

y están dados por

$$\hat{\mu} = \bar{\mathbf{x}}$$

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

donde



Estimación de μ y Σ

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

es el **vector de media muestral** y

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

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



Estimación de μ y Σ

Dos propiedades importantes de los MLE's:

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

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

para $h(\cdot)$ una función continua de θ .

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



Estimación de μ y Σ

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

Observación:

$$\begin{aligned} L(\mu, \Sigma) &= \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp \left\{ -\frac{1}{2} \sum_{j=1}^n (\mathbf{x}_j - \mu)^\top \Sigma^{-1} (\mathbf{x}_j - \mu) \right\} \\ &= \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp \left\{ -\frac{1}{2} \text{traza} \left[\Sigma^{-1} (n-1) \mathbf{S} + n(\bar{\mathbf{x}} - \mu)^\top \Sigma^{-1} (\bar{\mathbf{x}} - \mu) \right] \right\} \\ &= h(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) g(\bar{\mathbf{x}}, \mathbf{S}, \mu, \Sigma) \end{aligned}$$

Por teorema de factorización de Neyman - Pearson, $\bar{\mathbf{X}}$ y \mathbf{S} son estadísticos suficientes para μ y Σ , respectivamente.



Estimación de μ y Σ

Dada la **aleatoriedad** de la **muestra**, debemos estudiar el **comportamiento de las estadísticas** \bar{X} y 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 μ y Σ

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

Si $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ es una muestra aleatoria de una población $N_p(\mu, \Sigma)$. Entonces,

1. $\bar{\mathbf{X}} \sim N_p(\mu, \Sigma/n)$
 - i. $E[\bar{\mathbf{X}}] = \mu$, esto es, $\bar{\mathbf{X}}$ es un estimador **insesgado** para μ .
 - ii. $\text{Var}[\bar{\mathbf{X}}] = \frac{\Sigma}{n}$.
2. $(n-1)\mathbf{S}$ tiene una distribución **Wishart**^a con $n-1$ grados de libertad.

^auna **generalización** de la distribución **chi-cuadrado**: si $\mathbf{Z}_j \stackrel{iid}{\sim} N_p(\mathbf{0}, \Sigma)$, entonces $W = \sum_{j=1}^n \mathbf{Z}_j \mathbf{Z}_j^\top \sim W_{(n-1)}$.

Estimación de μ y Σ

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

Estimación de μ y Σ

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

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



Estimación de μ y Σ

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

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

¿Y si $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ **no** proviene de una $N_p(\mu, \Sigma)$?



Estimación de μ y Σ

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

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

¿Y si $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ **no** proviene de una $N_p(\mu, \Sigma)$?

Recurrimos a la **teoría asintótica**!



Estimación de μ y Σ

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

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

1. Teorema del límite central (TLC):

$$\sqrt{n}(\bar{\mathbf{X}} - \mu) \stackrel{\text{aprox.}}{\sim} N_p(\mathbf{0}, \Sigma)$$

2. $\bar{\mathbf{X}}$ converge en probabilidad a μ , esto significa que con una alta probabilidad $\bar{\mathbf{X}}$ estará cerca a μ .

^aVariables con naturaleza diferente (discretas, continuas, mixtas)

Estimación de μ y Σ

3. **S converge en probabilidad a Σ** , esto significa que para tamaños de **muestras grandes**, **S se aproximará a Σ** .
4. Como **S converge en probabilidad a Σ** , entonces
 - i. $\sqrt{n}(\bar{\mathbf{X}} - \mu) \stackrel{\text{aprox.}}{\sim} N_p(\mathbf{0}, \mathbf{S})$
 - ii. $n(\bar{\mathbf{X}} - \mu)^\top \mathbf{S}^{-1} (\bar{\mathbf{X}} - \mu) \stackrel{\text{aprox.}}{\sim} \chi_p^2$

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 μ

Sea

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

una **muestra aleatoria** de una población $N_p(\mu, \Sigma)$.

Suponga que el **interés** es verificar

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

donde

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \qquad \mu_0 = \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \vdots \\ \mu_{0p} \end{bmatrix}$$

Inferencia sobre μ

donde μ_0 es **dado (conocido)**.

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

$$H_{0i} : \mu_i = \mu_{0i}$$

$$H_{1i} : \mu_i \neq \mu_{0i}$$

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

$$t_i^2 = \frac{(\bar{x}_i - \mu_{0i})^2}{s_{ii}/n} = n(\bar{x}_i - \mu_{0i})(s_{ii})^{-1}(\bar{x}_i - \mu_{0i}) \stackrel{H_0 \text{ verd}}{\sim} t_{(n-1)}$$



Inferencia sobre μ

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

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

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 , $i = 1, 2, \dots, p$.



Inferencia sobre μ

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 potasio (X_3).

Use un nivel de significancia del 10 % para probar

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

donde $\mu = [\mu_1, \mu_2, \mu_3]^\top$ y $\mu_0 = [4, 50, 10]^\top$.

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

Inferencia sobre μ

En el caso **univariado**:

$$H_{01} : \mu_1 = 4$$

$$H_{02} : \mu_2 = 50$$

$$H_{03} : \mu_3 = 10$$

$$H_{11} : \mu_1 \neq 4$$

$$H_{12} : \mu_2 \neq 50$$

$$H_{13} : \mu_3 \neq 10$$

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**:

$$H_{01} : \mu_1 = 4$$

$$H_{02} : \mu_2 = 50$$

$$H_{03} : \mu_3 = 10$$

$$H_{11} : \mu_1 \neq 4$$

$$H_{12} : \mu_2 \neq 50$$

$$H_{13} : \mu_3 \neq 10$$

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. El problema es que **pueden existir otros valores de μ que también 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 μ

En el caso **multivariado**, el problema es **determinar** si **(en conjunto)** el vector μ_0 dado **es un valor plausible** para μ :

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

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

$$T^2 = \sqrt{n} (\bar{\mathbf{X}} - \mu_0)^\top \mathbf{S}^{-1} \sqrt{n} (\bar{\mathbf{X}} - \mu_0) \stackrel{H_0 \text{ verd.}}{\sim} \frac{(n-1)p}{n-p} F_{p, n-p}$$

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



Inferencia sobre μ

Observación:

$$\begin{aligned} T^2 &= \sqrt{n} (\bar{\mathbf{X}} - \mu_0)^\top \mathbf{S}^{-1} \sqrt{n} (\bar{\mathbf{X}} - \mu_0) \\ &= \sqrt{n} (\bar{\mathbf{X}} - \mu_0)^\top \left[\frac{(n-1)\mathbf{S}}{n-1} \right]^{-1} \sqrt{n} (\bar{\mathbf{X}} - \mu_0) \\ &= [\mathbf{N}_p(0, \boldsymbol{\Sigma})]^\top \left[\frac{W_{p,n-1}}{n-1} \right]^{-1} \mathbf{N}_p(0, \boldsymbol{\Sigma}) \end{aligned}$$

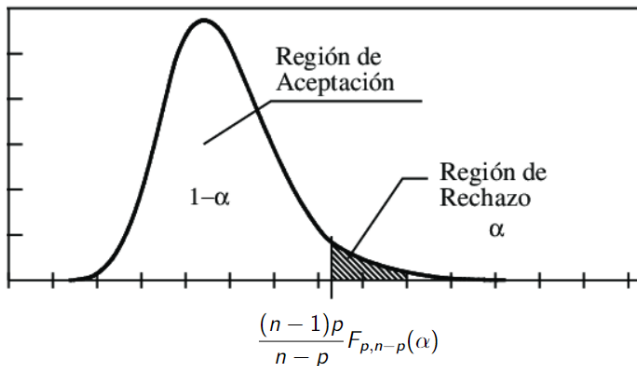
Recuerde que en el caso **univariado**

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



Inferencia sobre μ

$$T^2 = \sqrt{n} (\bar{\mathbf{X}} - \mu_0)^\top \mathbf{S}^{-1} \sqrt{n} (\bar{\mathbf{X}} - \mu_0) \stackrel{H_0 \text{ Verd}}{\sim} \frac{(n-1)p}{n-p} F_{p, n-p}$$



Inferencia sobre μ

Rechazamos H_0 a un nivel de significancia α si

$$T_{obs}^2 = \sqrt{n} (\bar{\mathbf{X}} - \mu_0)^\top \mathbf{S}^{-1} \sqrt{n} (\bar{\mathbf{X}} - \mu_0) > \frac{(n-1)p}{n-p} F_{p, n-p}(\alpha)$$

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



Inferencia sobre μ

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**

$$H_0 : \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 50 \\ 10 \end{bmatrix}$$

a un **nivel de significancia** del 10%. Luego, hay suficiente evidencia en la muestra para concluir que $\mu \neq \mu_0$



Inferencia sobre μ

```
# ----- Inferencia (PH) sobre mu ----- #  
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

Concluyendo con el **p-valor**,

$$P \left[T^2 \geq \frac{n-p}{(n-1)p} T_{obs}^2 \right] = P \left[T^2 \geq \frac{20-3}{(20-1)3} 9.66 \right]$$
$$= 0.0650824$$

```
pvalue<-pf((17/(3*19))*9.73,3,17,lower.tail = F)
> pvalue
[1] 0.0650824
```



Inferencia sobre μ

En R usamos:

```
# ----- Paquete MVTests ----- #  
require(MVTests)  
OneSampleHT2(X, mu0=c(4,50,10), alpha = 0.10)  
  
# ----- Paquete ICSNP ----- #  
require(ICSNP)  
HotellingsT2(X, mu=c(4,50,10))
```



Inferencia sobre μ

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 μ_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

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

$$H_0 : \theta \in \Theta_0$$

$$H_1 : \theta \in \Theta_0^c$$

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 μ

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

Suponga $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ es una muestra aleatoria de una $N_p(\mu, \Sigma)$ y que se quiere probar

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

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

$$\Lambda = \frac{\max_{\Sigma} L(\mu_0, \Sigma)}{\max_{\mu, \Sigma} L(\mu, \Sigma)} = \left[\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|} \right]^{n/2}$$



Inferencia sobre μ

donde

$$\hat{\Sigma} = \frac{1}{n} \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^\top$$
$$\hat{\Sigma}_0 = \frac{1}{n} \sum_{j=1}^n (\mathbf{x}_j - \mu_0)(\mathbf{x}_j - \mu_0)^\top$$

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

$$-2 \ln \Lambda \stackrel{H_0 \text{ verd.}}{\sim} \chi^2_{(v-v_0)}$$

donde

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



Inferencia sobre μ

siendo Θ el **espacio paramétrico** de

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

Observación: Θ tiene dimensión $p + p(p + 1)/2$.

La estadística

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

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**

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

$$= P[\text{Rechazar } H_0 \text{ cuando } H_0 \text{ verdadera}]$$

pequeños (ceranos a 0).



Inferencia sobre μ

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

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

$$= P[\text{Rechazar } H_0 \text{ cuando } H_0 \text{ falsa}]$$

Inferencia sobre μ

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

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

$$= P[\text{Rechazar } H_0 \text{ cuando } H_0 \text{ falsa}]$$

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 μ

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

$$H_0 : \mu = \mu_0,$$

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 μ

Una región de confianza para un vector de parámetro θ , es un **conjunto que contiene los posibles valores** de θ y está determinada por los datos.

Región de Confianza

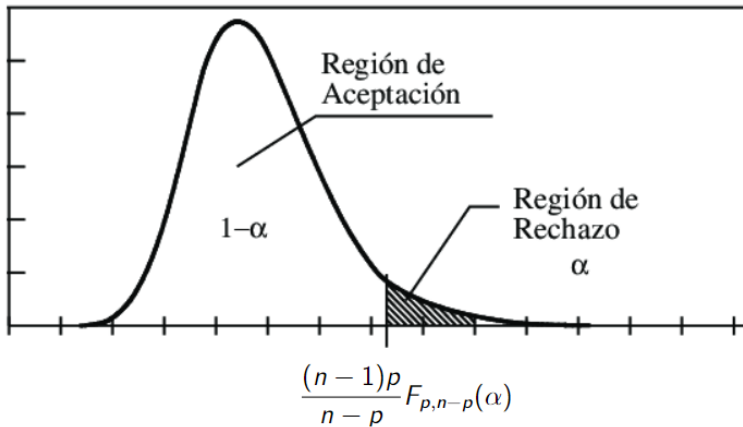
Se dice que $R(\mathbf{X})$ es una región con $(1 - \alpha)100\%$ de confianza para θ , **antes de obtener la muestra**

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



Inferencia sobre μ

Recuerde que



Inferencia sobre μ

Para **construir una región de confianza** para μ recuerde que:
Si $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ es una muestra aleatoria de una población $N_p(\mu, \Sigma)$

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

Entonces, **antes de obtener la muestra**

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



Inferencia sobre μ

Luego, para **una muestra dada**, el **conjunto de valores** de μ que **satisfacen**

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

definen una **región** para μ con $(1 - \alpha)100\%$ de **confianza**.

Observación:

- Geométricamente la **región** es una **elipsoide** centrada en $\bar{\mathbf{x}}$.
- La **región** $(1 - \alpha)100\%$ de confianza para μ **contiene todos los posibles valores** μ_0 para los cuales la prueba T^2 **no rechaza** $H_0 : \mu = \mu_0$ a un nivel de significancia α .



Inferencia sobre μ

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**

$$n(\bar{\mathbf{x}} - \mu_0)^\top \mathbf{S}^{-1} (\bar{\mathbf{x}} - \mu_0)$$

con el cuantil

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

Si

$$n(\bar{\mathbf{x}} - \mu_0)^\top \mathbf{S}^{-1} (\bar{\mathbf{x}} - \mu_0) \leq \frac{(n-1)p}{n-p} F_{p, n-p}(\alpha)$$

μ_0 CAE en la región de confianza.



Inferencia sobre μ

Para los datos de Mujeres.xlsx tenemos que

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

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

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

$$\mu = [\mu_1, \mu_2, \mu_3]^T$$

contiene todos los valores μ_1, μ_2, μ_3 que satisfacen



Inferencia sobre μ

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

¿Cae $\mu_0 = [4, 50, 10]^T$ dentro de la elipse del 90 % de confianza?.

Inferencia sobre μ

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

¿Cae $\mu_0 = [4, 50, 10]^T$ **dentro de la elipse** del 90 % de confianza?.

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

$$9.6644 \not\leq 8.174$$

es decir, el **vector** $\mu_0 = [4, 50, 10]^T$ **NO cae dentro de la región** y por lo tanto **se rechaza** H_0 . En consecuencia, $\mu_0 = [4, 50, 10]^T$ **NO es un valor plausible** para μ .



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 μ

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 μ

La i —ésima **media** puede ser escrita por

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

Inferencia sobre μ

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 $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y que

$$Z = a_1 X_1 + a_2 X_2 + \cdots + a_p X_p = \mathbf{a}^\top \mathbf{X}$$



Inferencia sobre μ

por resultado visto antes sigue que

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

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

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

y por lo tanto,

$$\bar{Z} = \mathbf{a}^\top \bar{\mathbf{X}} \quad ; \quad S_Z^2 = \mathbf{a}^\top \mathbf{S} \mathbf{a}$$

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

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



Inferencia sobre μ

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 μ .

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



Inferencia sobre μ

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 μ

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

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

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

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



Inferencia sobre μ

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

```
# ----- IC Univariados ----- #  
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)  
  
# ----- Longitud  
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 μ

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

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

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 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**.

¿Que hacemos?



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**.

¿Que hacemos?

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



Inferencia sobre μ

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

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

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

Las sucesivas elecciones de \mathbf{a} :

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

para los intervalos T^2 , permite concluir que



Inferencia sobre μ

$$\begin{aligned}\bar{x}_1 - \sqrt{\frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{11}}{n}} &\leq \mu_1 \leq \bar{x}_1 + \sqrt{\frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{11}}{n}} \\ \bar{x}_2 - \sqrt{\frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{22}}{n}} &\leq \mu_2 \leq \bar{x}_2 + \sqrt{\frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{22}}{n}} \\ &\vdots \\ \bar{x}_p - \sqrt{\frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{pp}}{n}} &\leq \mu_p \leq \bar{x}_p + \sqrt{\frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{pp}}{n}}\end{aligned}$$

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

Inferencia sobre μ

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**.

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 μ

En R usamos:

```
require(MVTests)  
OneSampleHT2(X, mu0=c(4,50,10), alpha = 0.10)
```

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 μ

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

$$\bar{x}_i \pm t_{(n-1)}(\alpha/2) \frac{\sqrt{S_{ii}}}{n} \quad ; \quad \bar{x}_i \pm \sqrt{\frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)} \sqrt{\frac{S_{ii}}{n}}$$

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

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

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

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



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 anterior¹, 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.

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

Inferencia sobre μ

De acuerdo con la tabla anterior¹, 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 alternativa** a los intervalos de confianza simultáneos T^2 que proporciona **mayor precisión (más cortos)**, éstos son llamados **intervalos de Bonferroni**.

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

Inferencia sobre μ

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

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

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

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

Luego

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



Inferencia sobre μ

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

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



Inferencia sobre μ

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

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

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 μ

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

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

\vdots

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



Inferencia sobre μ

Para los datos de Mujeres.1xlsx 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 .

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



Inferencia sobre μ

Para los datos de Mujeres.1xxs 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

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

²Table 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

```
# ----- IC simultaneos 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 μ

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



Inferencia sobre μ

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 μ

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: Estos 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:

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

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)
}
```