

# **Inferencia Estadística Multivariada**

Raúl Alberto Pérez  
raperez1@unal.edu.co

Profesor Asociado - Escuela de Estadística  
Universidad Nacional de Colombia, Sede Medellín

Semestre 2022-01

# Introducción

En este tema realizaremos tareas de inferencia sobre el vector de medias ( $\underline{\mu}$ ) y la matriz de varianzas covarianzas ( $\Sigma$ ) de una población normal multivariada, en base a una muestra aleatoria simple extraída de ella.

También se tratarán problemas que involucren a varias poblaciones.

Muchos procedimientos resultarán ser extensiones naturales de los métodos ya conocidos para poblaciones normales univariadas, mientras que en algún caso surgirán problemas nuevos, por ejemplo, comparación entre componentes del vector de medias o cuestiones de inferencia simultánea.

## Inferencia Estadística para la Media ( $\mu$ ): Caso Univariado

Desde el punto de vista de pruebas de hipótesis (PH), el problema se puede formular como una prueba de hipótesis que compiten:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_a : \mu \neq \mu_0 \end{cases}$$

Si  $X_1, X_2, \dots, X_n$  es una m.a de una población normal, el estadístico de prueba es:

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1} , \quad \text{con}$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{y} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Rechazamos  $H_0$ , ie. rechazamos que  $\mu_0$  es un valor viable de  $\mu$ , si  $|t|$ -excede un punto porcentual especificado de la distribución  $t$  con  $(n-1)$  grados de libertad.

Ahora, Rechazar  $H_0$  cuando  $|t|$ -es grande, es equivalente a rechazar  $H_0$  si **su cuadrado** también es grande, ie: rechazamos  $H_0$  si:

$$t^2 = \frac{(\bar{X} - \mu_0)^2}{S^2/n} = n(\bar{X} - \mu_0)(S^2)^{-1}(\bar{X} - \mu_0), \quad \text{es grande.}$$

Lo anterior es el cuadrado de la Distancia entre la media muestral  $\bar{X}$  y el valor propuesto  $\mu_0$ .

Las unidades de esta distancia son expresadas en términos de desviaciones estándar estimadas de  $\bar{X}$ , ie. en términos de  $S/\sqrt{n}$ .

Una vez que  $\bar{X}$  y  $S^2$  son observados, la prueba se convierte en:

**Rechazar  $H_0$ , en favor de  $H_1$ , a un nivel de significancia del  $\alpha\%$ , si**

$$t_0^2 = (\bar{x} - \mu_0)(s^2/n)^{-1}(\bar{x} - \mu_0) > t_{1-\alpha/2; n-1}^2$$

donde,  $t_{1-\alpha/2;n-1}$ -es el percentil  $(1 - \alpha/2)100\%$  de la distribución  $t$ -student con  $(n - 1)$ -grados de libertad.

Si  $H_0$ -no es rechazada, concluimos que  $\mu_0$  es un valor viable para la media poblacional normal ( $\mu$ ) de donde se extrajo la muestra.

**Pregunta** ¿Existen otros valores de  $\mu$  que sean también consistentes con los datos observados? Respuesta: Si.

Siempre existe un conjunto de valores viables para una media poblacional normal ( $\mu$ ).

A partir de la correspondencia bien-conocida entre la Región de Aceptación para la P.H:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_a : \mu \neq \mu_0 \end{cases}$$

y los intervalos de confianza para  $\mu$ , se tiene que:

$$\left\{ \text{No-Rechazamos } H_0 : \mu = \mu_0 \text{ al nivel } \alpha \right\} \iff \left| \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \right| \leq t_{1-\alpha/2; n-1}$$

lo cual equivale a que:

$$\left\{ \mu_0 : \text{caé en el I.C del } (1 - \alpha)100\% : \bar{x} \pm t_{\alpha/2; n-1} \frac{s}{\sqrt{n}} \right\}$$

$$\bar{x} - t_{\alpha/2; n-1} \frac{s}{\sqrt{n}} \leq \mu_0 \leq \bar{x} + t_{\alpha/2; n-1} \frac{s}{\sqrt{n}}$$

Este intervalo consiste de todos aquellos valores  $\mu_0$ , que no serían rechazados al nivel  $\alpha$ -para probar:  $H_0 : \mu = \mu_0$ .

Antes de que la muestra aleatoria sea seleccionada, el I.C anterior es un Intervalo-Aleatorio, debido a que los extremos finales dependen de las vs.as  $\bar{X}$  y  $S$ , ie.

$$\bar{X} - t_{\alpha/2;n-1} \frac{S}{\sqrt{n}} \leq \mu_0 \leq \bar{X} + t_{\alpha/2;n-1} \frac{S}{\sqrt{n}}.$$

La probabilidad de que este Intervalo-Aleatorio contenga a  $\mu$  es  $(1 - \alpha)$ .

Entre un número grande de tales intervalos-aleatorios independientes, aproximadamente el  $(1 - \alpha)100\%$  de ellos contendrán a  $\mu$ .

## Inferencia Estadística para la Media ( $\mu$ ): Caso Multivariado

### Pruebas de Hipótesis para $\underline{\mu}$ cuando $\Sigma$ es Desconocida (Normal)

Ahora se considerará el problema de determinar si un vector dado  $\underline{\mu}_0$  es un vector-viable para el vector de medias  $\underline{\mu}$  de una población Normal-Multivariada (con  $\Sigma$ -desconocida).

Se procederá de manera análoga al caso univariado.

Ahora se tendrá la distancia al cuadrado dada por:

$$T^2 = (\bar{\mathbf{x}} - \underline{\mu}_0)^t \left( \frac{1}{n} \mathbf{S} \right)^{-1} (\bar{\mathbf{x}} - \underline{\mu}_0) = n(\bar{\mathbf{x}} - \underline{\mu}_0)^t \mathbf{S}^{-1} (\bar{\mathbf{x}} - \underline{\mu}_0)$$

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \quad , \quad \mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^t \quad \text{y} \quad \underline{\mu}_0 = \begin{bmatrix} \mu_{10} \\ \mu_{20} \\ \vdots \\ \mu_{p0} \end{bmatrix}$$



La estadística  $T^2$ -se llama **Estadística  $T^2$ -de Hotelling**, en honor a Harold Hotelling, un pionero en Análisis Multivariado, quien fue el primero en obtener su distribución muestral.

Aquí,  $\frac{\mathbf{S}}{n}$ -es la matriz de varianzas-covarianzas estimada de  $\bar{\mathbf{x}}$ , pues

$$Var[\bar{\mathbf{x}}] = \frac{\mathbf{\Sigma}}{n} \quad ie. \quad \widehat{Var}[\bar{\mathbf{x}}] = \frac{\hat{\mathbf{\Sigma}}}{n} = \frac{\mathbf{S}}{n}$$

Si la distancia estadística observada  $T^2$  es mu grande, ie.  $\bar{\mathbf{x}}$ -está demasiado lejos de  $\underline{\mu}_0$ , y la hipótesis  $H_0 : \underline{\mu} = \underline{\mu}_0$  es rechazada.

Pero en este caso las tablas de puntos porcentuales de la distribución de la  $T^2$ -de Hotelling, no son necesarias para las pruebas formales de hipótesis.

Se utilizará el siguiente resultado:

$$T^2 \sim \frac{(n-1)p}{(n-p)} F_{p,n-p} = kF, \quad \text{con: } k = \frac{(n-1)p}{(n-p)}$$

o equivalentemente:  $\frac{(n-p)}{(n-1)p} T^2 = \frac{1}{k} T^2 \sim F_{p,n-p}$ , donde,  $F_{p,n-p}$  denota una v.a con distribución  $F$  con  $p$  y  $n-p$  grados de libertad respectivamente.

**En Resumen:** Sea  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$  una m.a de una distribución  $N_p(\underline{\mu}, \underline{\Sigma})$ , ( $\underline{\Sigma}$ -desconocido), entonces:

$$\begin{aligned} P \left[ T^2 > \frac{(n-1)p}{(n-p)} F_{\alpha;p,n-p} \right] &= P[T^2 > kF] \\ &= P \left[ n(\bar{\underline{x}} - \underline{\mu})^t \underline{S}^{-1} (\bar{\underline{x}} - \underline{\mu}) > kF \right] = \alpha \end{aligned}$$

cualesquiera sean los valores de  $\underline{\mu}$  y  $\underline{\Sigma}$ .

Al nivel de significancia del  $\alpha\%$ , rechazamos  $H_0 : \underline{\mu} = \underline{\mu}_0$ , en favor de:  $H_a : \underline{\mu} \neq \underline{\mu}_0$ , si el valor de:

$$T_0^2 = n(\bar{\mathbf{x}} - \underline{\mu}_0)^t \mathbf{S}^{-1}(\bar{\mathbf{x}} - \underline{\mu}_0) > kF = \frac{(n-1)p}{(n-p)} F_{\alpha;p,n-p},$$

o equivalentemente, rechazamos  $H_0$  si:

$$F_0 = \frac{(n-p)}{(n-1)p} T^2 = \frac{1}{k} T_0^2 > F_{\alpha;p,n-p},$$

en caso contrario no-rechazamos  $H_0$ .

**Ejemplo-1** Considere la siguiente matriz de datos asociada a una ma de tamaño  $n = 3$  de una población normal bi-variada con vector de medias  $\underline{\mu}$  y matriz de Var-Cov  $\mathbf{\Sigma}$ -desconocida, ie.  
 $\underline{x}_i \sim N_2(\underline{\mu}, \mathbf{\Sigma})$ , para  $i = 1, 2, 3$ :

$$\mathbf{X} = \begin{bmatrix} 6 & 9 \\ 10 & 6 \\ 8 & 3 \end{bmatrix}$$

Realizar la siguiente prueba de hipótesis:

$$\begin{cases} H_0 : \underline{\mu} = \underline{\mu}_0 \\ H_a : \underline{\mu} \neq \underline{\mu}_0 \end{cases} \iff \begin{cases} H_0 : \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \mu_{10} \\ \mu_{20} \end{bmatrix} \\ H_a : \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \neq \begin{bmatrix} \mu_{10} \\ \mu_{20} \end{bmatrix} \end{cases} \iff \begin{cases} H_0 : \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 9 \\ 5 \end{bmatrix} \\ H_a : \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \neq \begin{bmatrix} 9 \\ 5 \end{bmatrix} \end{cases}$$

De la matriz de datos se encuentra que:

$$\begin{aligned}\bar{\mathbf{x}} &= \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i = \frac{1}{3} \left[ \begin{pmatrix} 6 \\ 9 \end{pmatrix} + \begin{pmatrix} 10 \\ 6 \end{pmatrix} + \begin{pmatrix} 8 \\ 3 \end{pmatrix} \right] = \frac{1}{3} \begin{bmatrix} 6 + 10 + 8 \\ 9 + 6 + 3 \end{bmatrix} \\ &= \frac{1}{3} \begin{bmatrix} 24 \\ 18 \end{bmatrix} = \begin{bmatrix} 8 \\ 6 \end{bmatrix} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix}\end{aligned}$$

Similarmente se tiene que:

$$s_{11} = \frac{(6-8)^2 + (10-8)^2 + (8-8)^2}{2} = 4$$

$$s_{22} = \frac{(9-6)^2 + (6-6)^2 + (3-6)^2}{2} = 9$$

$$s_{12} = \frac{(6-8)(9-6) + (10-8)(6-6) + (8-8)(3-6)}{2} = -3$$

es decir que:

$$\mathbf{S} = \begin{bmatrix} 4 & -3 \\ -3 & 9 \end{bmatrix}, \text{ luego : } \mathbf{S}^{-1} = \frac{1}{27} \begin{bmatrix} 9 & 3 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{9} \\ \frac{1}{9} & \frac{4}{27} \end{bmatrix}$$

Ahora la estadística de prueba calculada es:

$$\begin{aligned} T_0^2 &= n(\bar{\mathbf{x}} - \underline{\mu}_0)^t \mathbf{S}^{-1} (\bar{\mathbf{x}} - \underline{\mu}_0) = 3 \left[ \begin{pmatrix} 8 \\ 6 \end{pmatrix} - \begin{pmatrix} 9 \\ 5 \end{pmatrix} \right]^t \begin{bmatrix} \frac{1}{3} & \frac{1}{9} \\ \frac{1}{9} & \frac{4}{27} \end{bmatrix} \left[ \begin{pmatrix} 8 \\ 6 \end{pmatrix} - \begin{pmatrix} 9 \\ 5 \end{pmatrix} \right] \\ &= 3 \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & \frac{1}{9} \\ \frac{1}{9} & \frac{4}{27} \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \frac{7}{9} \end{aligned}$$

Antes de que la muestra es seleccionada, la estadística  $T^2$  tiene la siguiente distribución:

$$T^2 \sim \frac{(n-1)p}{(n-p)} F_{p, n-p} = kF = \frac{(3-1)2}{(3-2)} F_{2, 3-1} = 4F_{2, 1}$$

Con un nivel de significancia de  $\alpha = 0.05$  se tiene que:

$F_{\alpha; 2, 1} = F_{0.05; 2, 1} = 200$ , de donde

$$kF = 4F_{\alpha; 2, 1} = 4F_{0.05; 2, 1} = 4 \times 200 = 800,$$

y por lo tanto como:  $T_0^2 = \frac{7}{9} < 800 = kF_{cal}$ , luego no se rechaza  $H_0$ . **equivalentemente:**

$$F_0 = \frac{1}{k} T_0^2 = \frac{1}{4} (7/9) = 0.19444 < 200 = F_{tabla}, \text{ luego: No Rechazamos } H_0$$

## El Ejemplo Anterior en R, mediante una función creada por Usuario

```
datos<-data.frame(x1=c(6,10,8),x2=c(9,6,3) )  
datos
```

```
##    x1 x2  
## 1   6  9  
## 2  10  6  
## 3   8  3
```

```
mu_0<-c(9,5)  
res_mu0<-HT2_mu0(datos,mu_0,0.05)
```



##	T2	K	F0	df1	df2
##	"0.777778"	"4"	"0.194444"	"2"	"1"

##	Media	F_Tabla	Valor_p
##	"8, 6"	"199.5"	"0.848528"

$T_0^2$	$K$	$F_0$	$df_1$	$df_2$	$F_{Tabla}$	Valor- $p$
0.777778	4	0.194444	2	1	199.5	0.848528

## Resultados de esta PH utilizando la función HotellingsT2 del paquete ICSNP del R.

En R existen varias funciones en distintos paquetes o librerías, las cuales se utilizan para realizar este tipo de pruebas de hipótesis. Se recomienda leer muy bien las ayudas que existen sobre estas funciones para utilizarlas de manera adecuada y definir de forma apropiada sus respectivos argumentos.

```
mu_0<-c(9,5)
HotellingsT2(datos, mu = mu_0)
```

```
##
## Hotelling's one sample T2-test
##
## data:  datos
## T.2 = 0.19444, df1 = 2, df2 = 1, p-value = 0.8485
## alternative hypothesis: true location is not equal to c(9,5)
```

# Resultados de esta PH Utilizando la función T2.test del paquete rrcov del R.

```
mu_0<-c(9,5)
T2.test(datos, mu = mu_0)
```

```
##
## One-sample Hotelling test
##
## data:  datos
## T2 = 0.77778, F = 0.19444, df1 = 2, df2 = 1, p-value = 0.8485
## alternative hypothesis: true mean vector is not equal to (9, 5)'
##
## sample estimates:
##                x1 x2
## mean x-vector  8  6
```

## Pruebas de Hipótesis para $\underline{\mu}$ cuando $\Sigma$ es Conocida (Población Normal)

Sea  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$  una m.a de una distribución normal  $p$ -variada con  $\underline{\mu}$ -desconocida y  $\Sigma$ -conocida, ie.  $\underline{x}_i \sim N_p(\underline{\mu}, \Sigma)$ .

Se desea contrastar las hipótesis: 
$$\begin{cases} H_0 : \underline{\mu} = \underline{\mu}_0 \\ H_a : \underline{\mu} \neq \underline{\mu}_0 \end{cases}$$

El estadístico de prueba utilizado en esta situación es:

$$\chi_0^2 = n(\bar{\underline{x}} - \underline{\mu}_0)^t \Sigma^{-1} (\bar{\underline{x}} - \underline{\mu}_0) \underset{\text{Bajo } H_0}{\sim} \chi_{(p)}^2$$

La regla de decisión es: Rechazar  $H_0$  si  $\chi_0^2 > \chi_{\alpha;p}^2$ , con  $\chi_{\alpha;p}^2$ -cuantil  $\alpha$ -superior de la distribución chi-cuadrado con  $p$ -grados de libertad.

## Pruebas de Hipótesis para $\underline{\mu}$ en Muestra Grande

Sea  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$  una m.a de una población  $p$ -variada con vector de medias  $\underline{\mu}$ -desconocida y matriz de varianzas covarianzas  $\Sigma$ -desconocida.

Se desea contrastar las hipótesis: 
$$\begin{cases} H_0 : \underline{\mu} = \underline{\mu}_0 \\ H_a : \underline{\mu} \neq \underline{\mu}_0 \end{cases}$$

El estadístico de prueba utilizado en esta situación es:

$$\chi_0^2 = n(\bar{\underline{x}} - \underline{\mu}_0)^t \mathbf{S}^{-1}(\bar{\underline{x}} - \underline{\mu}_0) \underset{\text{Bajo } H_0}{\sim} \chi_{(p)}^2$$

La regla de decisión es: Rechazar  $H_0$  si  $\chi_0^2 > \chi_{\alpha; p}^2$ .

## Pruebas de Hipótesis para $\mu$ n-Grande en R

##		X1	X2	X3	Grupos
## 1		9.6838	8.7045	4.0509	1
## 2		5.9703	4.5724	8.3056	1
## 3		1.5094	6.3666	7.5676	1
## 4		.	.	.	<NA>
## 5		.	.	.	<NA>
## 6		.	.	.	<NA>
## 98		7.8864	6.0692	6.445	3
## 99		4.1332	4.5396	6.7288	3
## 100		4.4509	7.6449	8.2376	3

## Resultados de la PH usando la función de usuario HT2\_mu0\_ngrande()

Para este se creo una función de usuario llamada:  
HT2\_mu0\_ngrande la cual se utiliza a continuación.

```
mu_0<-c(0,0,0)  
res_mu0nggrande<-HT2_mu0_ngrande(grupo1[,1:3], mu_0, 0.05)
```

$\chi_0^2$	$df$	$\chi_{Tabla}$	Valor- $p$
307.064	3	7.81473	0

# Resultados utilizando la Función HotellingsT2 del R

Igualmente, también se puede usar la función HotellingsT2 del R de la siguiente forma.

```
mu_0<-c(0,0,0)
HotellingsT2(grupo1[,1:3],mu=mu_0,test="chi")
```

```
##
## Hotelling's one sample T2-test
##
## data: grupo1[, 1:3]
## T.2 = 307.06, df = 3, p-value < 2.2e-16
## alternative hypothesis: true location is not equal to c(0,0,0)
```



# Medias para cada Grupo

```
descriptivos<-resumen_xgrupos(datos[1:3],datos[4])
descriptivos[1]
```

```
## $Medias
##   Grupos      X1      X2      X3
## 1      1 4.504767 4.987719 5.178689
## 2      2 4.928179 5.102939 4.947521
## 3      3 5.690567 5.277362 4.760589
```

```
mu_0<-c(4.5,5,5)
HotellingsT2(grupo1[,1:3],mu=mu_0,test="chi")
```

```
##
## Hotelling's one sample T2-test
##
## data: grupo1[, 1:3]
## T.2 = 0.22328, df = 3, p-value = 0.9737
## alternative hypothesis: true location is not equal to c(4.5,5,5)
```

## Con la función de Usuario HT2\_mu0\_ngrande

```
mu_0<-c(4.5,5,5)  
res_mu0ngrande<-HT2_mu0_ngrande(grupo1[,1:3], mu_0, 0.05)
```

$\chi_0^2$	$df$	$\chi_{Tabla}$	Valor- $p$
0.223285	3	7.81473	0.973746

## PH para igualdad de vectores de medias de dos Poblaciones

$\underline{\mu}_1, \underline{\mu}_2$

Sea  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$  una m.a de una población normal  $p$ -variada con vector de medias  $\underline{\mu}_1$ -desconocida y matriz de varianzas-covarianzas  $\Sigma_1$ , ie.  $\underline{x}_i \sim N_p(\underline{\mu}_1, \Sigma_1)$  y sean  $\underline{y}_1, \underline{y}_2, \dots, \underline{y}_m$  una m.a de una población normal  $p$ -variada con vector de medias  $\underline{\mu}_2$ -desconocida y matriz de varianzas covarianzas  $\Sigma_2$ , ie.  $\underline{y}_i \sim N_p(\underline{\mu}_2, \Sigma_2)$ . Ambas m.a son independientes entre si.

❶  $\Sigma_1 = \Sigma_2 = \Sigma$ -Desconocida

Se desea contrastar las hipótesis:

$$\begin{cases} H_0 : \underline{\mu}_1 - \underline{\mu}_2 = \underline{\delta}_0 \\ H_a : \underline{\mu}_1 - \underline{\mu}_2 \neq \underline{\delta}_0 \end{cases}$$

En este caso, se usa como estimador de  $\Sigma$  a la varianza ponderada dada por:

$$S_p = \hat{\Sigma} = \frac{(n-1)S_1 + (m-1)S_2}{n+m-2},$$

Bajo  $H_0$ -cierto, el estadístico de prueba es:

$$T^2 = \frac{nm}{n+m} (\bar{\mathbf{x}} - \bar{\mathbf{y}} - \underline{\delta}_0)^t \mathbf{S}_p^{-1} (\bar{\mathbf{x}} - \bar{\mathbf{y}} - \underline{\delta}_0) \sim \frac{(n+m-2)p}{n+m-p-1} F_{p; n+m-p-1} = k$$

$$\text{con: } k = \frac{(n+m-2)p}{n+m-p-1}$$

Rechazamos  $H_0$  si:

$$T_0^2 > kF = \frac{(n+m-2)p}{n+m-p-1} F_{\alpha : p, n+m-p-1}$$

O equivalentemente, rechazamos  $H_0$  si:

$$F_0 = \frac{n+m-p-1}{(n+m-2)p} T_0^2 = \frac{1}{k} T_0^2 > F_{\text{tabla}} = F_{\alpha : p, n+m-p-1}$$

**Ejemplo-2** Cuatro pruebas psicológicas fueron aplicadas a 32 hombres y 32 mujeres. Las variables consideradas en dicha prueba fueron:  $X_1$ : Inconsistencias pictóricas,  $X_2$ : Reconocimiento de herramientas,  $X_3$ : Forma de emplear el papel y  $X_4$ : Vocabulario. Se pretende establecer si la respuesta media para las cuatro variables es similar en hombres y mujeres. El experimento fue llevado a cabo de tal manera que las observaciones para ambos grupos fueron independientes. Se asume además que cada grupo de personas es una m.a de una distribución normal 4-variada, con vectores de media  $\underline{\mu}_H$  y  $\underline{\mu}_M$  respectivamente, y matriz de varianzas-covarianzas  $\Sigma$ -común pero desconocida.

Se desea probar las hipótesis:

$$\begin{cases} H_0 : \underline{\mu}_H = \underline{\mu}_M \\ H_a : \underline{\mu}_H \neq \underline{\mu}_M \end{cases} \iff \begin{cases} H_0 : \underline{\mu}_H - \underline{\mu}_M = \underline{0} \\ H_a : \underline{\mu}_H - \underline{\mu}_M \neq \underline{0} \end{cases}, \quad \underline{\delta}_0 = \underline{0}$$

La información recopilada permite obtener los siguientes cálculos resúmenes:

$$n = 32, \quad \bar{\mathbf{x}}_H = \begin{bmatrix} 15.97 \\ 15.91 \\ 27.19 \\ 22.75 \end{bmatrix}, \quad \mathbf{S}_H = \begin{bmatrix} 5.192 & 4.545 & 6.522 & 5.250 \\ & 13.18 & 6.760 & 6.266 \\ & & 28.67 & 14.47 \\ & & & 16.65 \end{bmatrix}$$

$$m = 32, \quad \bar{\mathbf{x}}_M = \begin{bmatrix} 12.34 \\ 13.91 \\ 16.59 \\ 21.94 \end{bmatrix}, \quad \mathbf{S}_M = \begin{bmatrix} 9.136 & 7.549 & 5.531 & 4.151 \\ & 18.60 & 5.446 & 5.466 \\ & & 13.55 & 13.55 \\ & & & 28.00 \end{bmatrix}$$

Usando estos resultados se tiene que la matriz de var-cov ponderada es:

$$S_p = \hat{\Sigma} = \frac{(n-1)S_1 + (m-1)S_2}{n+m-2} = \begin{bmatrix} 7.164 & 6.047 & 6.027 & 4.701 \\ & 15.89 & 6.103 & 5.866 \\ & & 21.11 & 14.01 \\ & & & 22.325 \end{bmatrix}$$

Bajo  $H_0$ -cierto, el estadístico de prueba es:

$$T_0^2 = \left[ \begin{pmatrix} 15.97 \\ 15.91 \\ 27.19 \\ 22.75 \end{pmatrix} - \begin{pmatrix} 12.34 \\ 13.91 \\ 16.59 \\ 21.94 \end{pmatrix} \right]^t$$

$$\left[ \left( \frac{1}{32} + \frac{1}{32} \right) \begin{pmatrix} 7.164 & 6.047 & 6.027 & 4.701 \\ & 15.89 & 6.103 & 5.866 \\ & & 21.11 & 14.01 \\ & & & 22.325 \end{pmatrix} \right]^{-1} \left[ \begin{pmatrix} 15.97 \\ 15.91 \\ 27.19 \\ 22.75 \end{pmatrix} - \begin{pmatrix} 12.34 \\ 13.91 \\ 16.59 \\ 21.94 \end{pmatrix} \right]$$

$$= 136.67$$

Usando  $\alpha = 0.05$ , se tiene que:

$$F_{\text{tabla}} = F_{\alpha} ; p, n+m-p-1 = F_{0.05} ; 4, 32+32-4-1 = F_{0.05} ; 4, 59 = 2.53,$$

luego:

$$kF = \frac{(n+m-2)p}{n+m-p-1} F_{\alpha ; p, n+m-p-1} = \frac{(62)4}{59} F_{0.05 ; 4, 59} = 10.635,$$

de donde, como:

$$T_0^2 = 136.67 > 10.635 = kF = \frac{(n+m-2)p}{n+m-p-1} F_{\alpha ; p, n+m-p-1},$$

entonces se rechaza  $H_0$  y se concluye que los resultados medios son diferentes para hombres y mujeres a un nivel de significancia del 5%.

equivalentemente:

$$F_0 = \frac{1}{k} T_0^2 = \frac{59}{(62)4} (136.67) > 2.53 = F_{\text{tabla}}$$



## Otro Ejemplo Usando R: $\Sigma_1 = \Sigma_2 = \Sigma$ -Desconocida, Pob. Normal

Resultados usando la función de usuario de Usuario  
HT2\_sigmas\_iguales()

```
x<-subset(datos1,datos1$Grupos==1)
x<-x[,1:3]
y<-subset(datos1,datos1$Grupos==2)
y<-y[,1:3]
mu_0<-c(0,0,0)
res2p<-HT2_sigmas_iguales(x,y,mu_0,0.05)
```

$T_0^2$	$K$	$F_0$	$df_1$	$df_2$	$F_{Tabla}$	Valor- $p$
0.387667	3.11765	0.124346	3	51	2.78623	0.945298

# Resultados Utilizando la función HotellingsT2 del paquete ICSNP del R.

Existen varias funciones en distintos paquetes del R que se utilizan para esta prueba de hipótesis cuando  $\Sigma_1 = \Sigma_2 = \Sigma$ -Desconocida, Pob. Normal.

```
res_1<-HotellingsT2(x,y)
res_1
```

```
##
## Hotelling's two sample T2-test
##
## data:  x and y
## T.2 = 0.12435, df1 = 3, df2 = 51, p-value = 0.9453
## alternative hypothesis: true location difference is not equal to c(0,0,0)
```

Estadístico= $F_0 \sim F$	$gl_{num}$	$gl_{den}$	p-Valor
0.12435	3	51	0.9453

## Resultados Utilizando la función `T2.test` del paquete `rrcov` del R.

Two-sample Hotelling test

data: x and y  $T2 = 0.38767$ ,  $F = 0.12435$ ,  $df1 = 3$ ,  $df2 = 51$ ,  $p\text{-value} = 0.9453$   
alternative hypothesis: true difference in mean vectors is not equal to (0,0,0) sample  
estimates: X1 X2 X3 mean x-vector 4.504767 4.987719 5.178689 mean y-vector  
4.928179 5.102939 4.947521

$T2$	$F$	$df_1$	$df_2$	$Valor - p$
0.38767	0.12435	3	51	0.9453

# Resultados Utilizando la función `hotelling.test` del paquete `Hotelling` del R.

```
resh<-hotelling.test(x, y)  
resh
```

```
## Test stat: 0.38767  
## Numerator df: 3  
## Denominator df: 51  
## P-value: 0.9453
```

$T^2$	$1/K$	$df_1$	$df_2$	$n_x$	$n_y$	$p$
0.387666735927368	0.320754716981132	3	51	27	28	3

## 2 $\Sigma_1 \neq \Sigma_2$ -Desconocida

Se desea contrastar las hipótesis:

$$\begin{cases} H_0 : \underline{\mu}_1 - \underline{\mu}_2 = \underline{\delta}_0 \\ H_a : \underline{\mu}_1 - \underline{\mu}_2 \neq \underline{\delta}_0 \end{cases}$$

Bajo  $H_0$ -cierto, el estadístico de prueba es:

$$T^2 = (\bar{\mathbf{x}} - \bar{\mathbf{y}} - \underline{\delta}_0)^t \left[ \frac{S_1}{n} + \frac{S_2}{m} \right]^{-1} (\bar{\mathbf{x}} - \bar{\mathbf{y}} - \underline{\delta}_0) \sim \frac{vp}{v-p+1} F_{p; v-p+1} = kF$$

$$\text{con: } k = \frac{vp}{v-p+1} \quad \text{y} \quad v = \frac{\text{tr}(S_e) + [\text{tr}(S_e)]^2}{\sum_{i=1}^2 \frac{1}{n_i-1} \left\{ \text{tr}(V_i) + [\text{tr}(V_i)]^2 \right\}}$$

$$V_i = \frac{S_i}{n_i} \quad \text{y} \quad S_e = V_1 + V_2 = \frac{S_1}{n} + \frac{S_2}{m}.$$

Rechazamos  $H_0$  si:  $T_0^2 > kF$    ó    $F_0 = \frac{1}{k} T_0^2 > F_{\text{tabla}}$

**Ejemplo-3** Se compararon dos tipos de suelos, uno de los cuales tiene un tipo de bacteria y el otro no. Las variables aleatorias de interés que fueron medidas son:  $X_1$ : PH del suelo,  $X_2$ : Cantidad de fosfato y  $X_3$ : Contenido de nitrógeno.

Los resultados de las mediciones se muestran en la siguiente tabla. Se desea verificar si hay similitud en ambos suelos, en relación con los vectores de medias asociados a estas tres variables medidas. ¿Cuál es la conclusión usando un  $\alpha = 0.05$ ?  $n = 13$  y  $m = 10$ .

$X_1$	$X_2$	$X_3$	$X_1$	$X_2$	$X_3$	$X_1$	$X_2$	$X_3$
8.0	60	58	8.3	57	60	5.7	42	14
8.0	156	68	7.0	94	43	6.2	40	23
8.0	90	37	8.5	86	40	6.4	49	18
6.1	44	27	8.4	52	48	5.8	31	17
7.4	207	31	7.9	146	52	6.4	31	19
7.4	120	32	6.2	49	30	5.4	62	26
8.4	65	43	5.6	31	23	5.4	42	16
8.1	237	45	5.8	42	22			

Se desea probar las hipótesis:

$$\begin{cases} H_0 : \underline{\mu}_1 = \underline{\mu}_2 \\ H_a : \underline{\mu}_1 \neq \underline{\mu}_2 \end{cases} \iff \begin{cases} H_0 : \underline{\mu}_1 - \underline{\mu}_2 = \underline{\mathbf{0}} \\ H_a : \underline{\mu}_1 - \underline{\mu}_2 \neq \underline{\mathbf{0}} \end{cases}$$

es decir que:  $\underline{\delta}_0 = \underline{\mathbf{0}}$ .

Realizando los cálculos respectivos, se obtiene que bajo  $H_0$ -cierto, el estadístico de prueba es:

$$T^2 = (\underline{\bar{x}} - \underline{\bar{y}} - \underline{\delta}_0)^t \left[ \frac{S_1}{n} + \frac{S_2}{m} \right]^{-1} (\underline{\bar{x}} - \underline{\bar{y}} - \underline{\delta}_0) = 96.8178$$

Después de ciertos cálculos, se obtiene que el valor de  $v$  es:

$$v = \frac{tr(S_e) + [tr(S_e)]^2}{\sum_{i=1}^2 \frac{1}{n_i-1} \left\{ tr(V_i) + [tr(V_i)]^2 \right\}} = 12.874 \approx 13$$

Para  $\alpha = 0.05$  se tiene que:

$$F_{\text{tabla}} = F_{\alpha : p, v-p+1} = F_{0.05; 3, 13-3+1} = F_{0.05; 3, 11} = 3.59$$

y

$$kF = \frac{vp}{v-p+1} F_{\alpha : p, v-p+1} = \frac{13(3)}{11} \times 3.59 = 3.54 \times 3.59 = 12.7083,$$

$$\text{como } T_0^2 = 96.18 > 12.73 = kF_{\text{tabla}},$$

luego rechazamos  $H_0$  y se concluye que las características medias son diferentes para ambos tipos de suelo, aun nivel de significancia del 5%.

equivalentemente:

$$F_0 = \frac{1}{k} T_0^2 = \frac{11}{13(3)} (96.18) = 27.1276 > 3.59 = F_{\text{tabla}}$$



## Usando R. $\Sigma_1 \neq \Sigma_2$ -Desconocidas, Pob. Normal

```
##      x1  x2 x3
## 1    8  60 58
## 2    8 156 68
## 3    8  90 37
```

```
##      x1 x2 x3
## 1 6.2 49 30
## 2 5.6 31 23
## 3 5.8 42 22
```

Resultados usando la función de usuario HT2\_sigmas\_diferentes() (**Aproximación de: Nel and Van Der Merwe-1986**).

$T_2$	$v$	$K = vp/(v - p + 1)$	$F_0$	$df_1$	$df_2$	$F_{Tabla}$	Valor- $p$
96.8178	13	3.54545	27.3076	3	11	3.58743	2.14103e-05

Resultados usando la función de usuario HT2\_sigmas\_diferentes\_texto\_guia() (**Aproximación de: Krishnamoorthy and Yu-2004**)

$T_2$	$v$	$K = vp/(v - p + 1)$	$F_0$	$df_1$	$df_2$	$F_{Tabla}$	Valor- $p$
96.8178	70	3.08824	31.3505	3	68	2.7395	7.69607e-13

## Otro Ejemplo Usando R. $\Sigma_1 \neq \Sigma_2$ -Desconocidas, Pob. Normal

Resultados usando la función de usuario

HT2\_sigmas\_diferentes() (**Aproximación de: Nel and Van Der Merwe-1986**).

```
mu_0<-c(0,0,0)
res_sigmasdif<-res2p<-HT2_sigmas_diferentes(x,y,mu_0,0.05)
```

$T_2$	$\nu$	$K = \nu p / (\nu - p + 1)$	$F_0$	$df_1$	$df_2$	$F_{Tabla}$	Valor- $p$
0.386428	53	3.11765	0.123949	3	51	2.78623	0.94554

## Otro Ejemplo Usando R. $\Sigma_1 \neq \Sigma_2$ -Desconocidas, Pob. Normal

Resultados usando la función de usuario

HT2\_sigmas\_diferentes\_texto\_guia() (**Aproximación de: Krishnamoorthy and Yu-2004**)

Para este caso el  $v$  se calcula como sigue:

$$v = \frac{p + p^2}{\frac{1}{n_1} \left\{ \text{tr}[(V_1 S_e^{-1})^2] + [\text{tr}(V_1 S_e^{-1})]^2 \right\} + \frac{1}{n_2} \left\{ \text{tr}[(V_2 S_e^{-1})^2] + [\text{tr}(V_2 S_e^{-1})]^2 \right\}}$$

```
mu_0<-c(0,0,0)
res_sigmasdif1<-res2p_vp<-HT2_sigmas_diferentes_texto_guia(x,y,mu_0,0.05)
```

$T_2$	$v$	$K = vp/(v - p + 1)$	$F_0$	$df_1$	$df_2$	$F_{Tabla}$	Valor- $p$
0.386428	211	3.02871	0.127589	3	209	2.6478	0.943665

### 3 $\Sigma_1 = \Sigma_2 = \Sigma$ -Conocida

Se desea contrastar las hipótesis: 
$$\begin{cases} H_0 : \underline{\mu}_1 - \underline{\mu}_2 = \underline{\delta}_0 \\ H_a : \underline{\mu}_1 - \underline{\mu}_2 \neq \underline{\delta}_0 \end{cases}$$

Como ambas m.a provienen de poblaciones normales  $p$ -variadas, entonces:

$$\underline{\bar{x}} \sim N_p \left( \underline{\mu}_1, \frac{\Sigma}{n} \right) \quad y \quad \underline{\bar{y}} \sim N_p \left( \underline{\mu}_2, \frac{\Sigma}{m} \right)$$

y como además, ambas m.a provienen de poblaciones independientes, luego:

$$(\underline{\bar{x}} - \underline{\bar{y}}) \sim N_p \left( \underline{\mu}_1 - \underline{\mu}_2, \frac{\Sigma}{n} + \frac{\Sigma}{m} \right) \quad \text{es decir,}$$

$$Var[\underline{\bar{x}} - \underline{\bar{y}}] = \left( \frac{1}{n} + \frac{1}{m} \right) \Sigma = \frac{n+m}{nm} \Sigma.$$

Luego, bajo  $H_0 : \underline{\mu}_1 - \underline{\mu}_2 = \underline{\delta}_0$ -cierto, es decir:  $\underline{\mu}_1 - \underline{\mu}_2 = \underline{\delta}_0$ , se tiene que el estadístico de prueba:

$$\chi_0^2 = \left( \frac{nm}{n+m} \right) (\bar{\mathbf{x}} - \bar{\mathbf{y}} - \underline{\delta}_0)^t \mathbf{\Sigma}^{-1} (\bar{\mathbf{x}} - \bar{\mathbf{y}} - \underline{\delta}_0) \sim \chi_{(p)}^2.$$

La regla de decisión es: Rechazamos  $H_0$  si  $\chi_0^2 > \chi_{\alpha;p}^2$ .

**Ejemplo-4** Los Biólogos **Grojan** y **Wirth** (1981) descubrieron dos nuevas especies de insectos **Ameroheleafasciata** (AF) y **Apseudofasciata** (APF). Puesto que las especies son similares en apariencia, resulta útil para el biólogo estar en capacidad de clasificar un espécimen como AF o APF basado en características externas que son fáciles de medir. Entre alguna de las características que distinguen los AF de los APF, los biólogos reportan medidas de la longitud de las antenas (X) y la longitud de las alas (Y), ambas medidas en milímetros, de nueve insectos AF y de seis insectos APF.

Suponga que las observaciones bivariadas para ambas poblaciones de insectos, provienen de distribuciones normales bivariadas, con igual matriz de covarianzas.

Sea  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_9$  una m.a que representa las observaciones bivariadas para la población de insectos  $AF$ , ie.  $\underline{x}_i \sim N_p(\underline{\mu}_{AF}, \Sigma)$  para  $i = 1, 2, \dots, 9$ . Análogamente, Sea  $\underline{y}_1, \underline{y}_2, \dots, \underline{y}_6$  una m.a que representa las observaciones bivariadas para la población de insectos  $APF$ , ie.  $\underline{y}_i \sim N_p(\underline{\mu}_{APF}, \Sigma)$  para  $i = 1, 2, \dots, 6$ .

Se asume que ambas m.a son independientes entre si.

ESPECIE	X	Y
AF	1.38	1.64
AF	1.40	1.20
AF	1.24	1.72
AF	1.36	1.74
AF	1.38	1.82
AF	1.48	1.82
AF	1.54	1.82
AF	1.38	1.90
AF	1.56	2.08

Además se asume que:

$$\Sigma_1 = \Sigma_2 = \Sigma = \begin{bmatrix} 0.022 & 0.006 \\ 0.006 & 0.043 \end{bmatrix}$$

Se desea contrastar las hipótesis:

$$\begin{cases} H_0 : \mu_{AF} = \mu_{APF} \\ H_a : \mu_{AF} \neq \mu_{APF} \end{cases} \iff \begin{cases} H_0 : \mu_{AF} - \mu_{APF} = \underline{0} \\ H_a : \mu_{AF} - \mu_{APF} \neq \underline{0} \end{cases} .$$

De los datos muestrales se tiene que:

$$\bar{\underline{x}} = \begin{bmatrix} 1.41 \\ 1.80 \end{bmatrix} \quad \text{y} \quad \bar{\underline{y}} = \begin{bmatrix} 1.23 \\ 1.93 \end{bmatrix}$$



luego bajo  $H_0$ -cierto, el estadístico de prueba es:

$$\begin{aligned}\chi_0^2 &= \left( \frac{nm}{n+m} \right) (\bar{\mathbf{x}} - \bar{\mathbf{y}})^t \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \bar{\mathbf{y}}) \\&= \left( \frac{9 \times 6}{9+6} \right) [1.41 - 1.23 \quad 1.80 - 1.93] \begin{bmatrix} 0.022 & 0.006 \\ 0.006 & 0.043 \end{bmatrix}^{-1} \begin{bmatrix} 1.41 - 1.23 \\ 1.80 - 1.93 \end{bmatrix} \\&= \frac{54}{15} [0.18 \quad -0.13] \begin{bmatrix} 48.62 & -7.29 \\ -7.29 & 24.31 \end{bmatrix} \begin{bmatrix} 0.18 \\ -0.13 \end{bmatrix} = \frac{54}{15} \times 2.33 = 8.39\end{aligned}$$

Para  $\alpha = 0.05$ , se tiene que:  $\chi_{\alpha;p}^2 = \chi_{0.05;2}^2 = 5.99$ , de donde, como  $\chi_0^2 > \chi_{\alpha;p}^2$ , luego se rechaza  $H_0$  y se concluye que la información registrada indica que los vectores de medias para la longitud de alas y antenas de las dos especies de insectos son diferentes a un nivel de significancia del 5%.

## Pruebas de Hipótesis Acerca de dos vectores de Medias Poblacionales $\underline{\mu}_1$ y $\underline{\mu}_2$ , para muestras grandes

En este caso se desconoce la distribución de la cual provienen ambas muestras. Se supone que los tamaños de ambas m.a son grandes,  $n - p$  y  $m - p$  grandes.

Sea  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$  una m.a de una población  $p$ -variada con vector de medias  $\underline{\mu}_1$ -desconocida y matriz de varianzas-covarianzas  $\Sigma_1$  y sean  $\underline{y}_1, \underline{y}_2, \dots, \underline{y}_m$  una m.a de una población  $p$ -variada con vector de medias  $\underline{\mu}_2$ -desconocida y matriz de varianzas-covarianzas  $\Sigma_2$ . Ambas m.a son independientes entre si.

①  $\Sigma_1 = \Sigma_2 = \Sigma$ -Desconocida

Se usa como estimador de  $\Sigma$  a  $S_P$ , ie.

$$\hat{\Sigma} = S_P = \frac{(n-1)S_1 + (m-1)S_2}{n+m-2}$$

y el estadístico de prueba es:

$$\chi_0^2 = \frac{nm}{n+m} (\bar{\mathbf{x}} - \bar{\mathbf{y}} - \underline{\gamma}_0)^t \mathbf{S}_P^{-1} (\bar{\mathbf{x}} - \bar{\mathbf{y}} - \underline{\gamma}_0) \sim \chi_p^2.$$

Rechazamos  $H_0$  si  $\chi_0^2 > \chi_{\alpha,p}$ .

## 2 $\Sigma_1 \neq \Sigma_2$ -Desconocidas

Se usa como estimador de  $\Sigma$  a  $S_e$ , ie.

$$\hat{\Sigma} = S_e = \frac{S_1}{n} + \frac{S_2}{m}$$

y el estadístico de prueba es:

$$\chi_0^2 = (\bar{\mathbf{x}} - \bar{\mathbf{y}} - \underline{\gamma}_0)^t \mathbf{S}_e^{-1} (\bar{\mathbf{x}} - \bar{\mathbf{y}} - \underline{\gamma}_0) \sim \chi_p^2.$$

Rechazamos  $H_0$  si  $T_0^2 > \chi_{\alpha,p}$ .

### 3 $\Sigma_1 = \Sigma_2 = \Sigma$ -Conocida

Se desea contrastar las hipótesis:

$$\begin{cases} H_0 : \underline{\mu}_1 - \underline{\mu}_2 = \underline{\gamma}_0 \\ H_a : \underline{\mu}_1 - \underline{\mu}_2 \neq \underline{\gamma}_0 \end{cases}$$

En este caso se utiliza como estadístico de prueba a:

$$\chi_0^2 = \frac{nm}{n+m} (\underline{\bar{x}} - \underline{\bar{y}} - \underline{\gamma}_0)^t \Sigma^{-1} (\underline{\bar{x}} - \underline{\bar{y}} - \underline{\gamma}_0) \sim \chi_p^2$$

Rechazamos  $H_0$  si  $\chi_0^2 > \chi_{\alpha,p}^2$ .

## Ejemplos Usando R. $\Sigma_1 = \Sigma_2 = \Sigma$ -Desconocida para $n$ -grande

Resultados usando la función de usuario

HT2\_sigmas\_iguales\_ngrande()

$\chi_0^2$	$df$	$\chi_{Tabla}$	Valor- $p$
0.387667	3	7.81473	0.942778

## Resultados Utilizando la función HotellingsT2 del paquete ICSNP del R:

```
res_1A<-HotellingsT2(x,y,test="chi")  
res_1A
```

```
##  
## Hotelling's two sample T2-test  
##  
## data: x and y  
## T.2 = 0.38767, df = 3, p-value = 0.9428  
## alternative hypothesis: true location difference is not equal to c(0,0,0)
```

Estadístico= $\chi_0 \sim \chi^2$	gl	p-Valor
0.38767	3	0.94278

## Pruebas de Hipótesis Acerca de dos vectores de Medias Poblacionales $\underline{\mu}_1$ y $\underline{\mu}_2$ , Observaciones Pareadas

Sea  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$  y  $\underline{y}_1, \underline{y}_2, \dots, \underline{y}_n$  dos m.a de una población  $p$ -variada con vectores de medias  $\underline{\mu}_x, \underline{\mu}_y$  desconocidos y matriz de varianzas-covarianzas  $\Sigma$ -Desconocida.

Suponga además que  $Cov[\underline{x}_i, \underline{y}_i] = \Sigma \neq \mathbf{0}, i = 1, 2, \dots, n$ , ie. las dos m.a son correladas (Muestras no independientes o dependientes).

Para contrastar las hipótesis: 
$$\begin{cases} H_0 : \underline{\mu}_x - \underline{\mu}_y = \underline{0} \\ H_a : \underline{\mu}_x - \underline{\mu}_y \neq \underline{0} \end{cases}$$

Se trabaja con las diferencias de cada par de observaciones multivariadas, definidas como:

$$\underline{D}_i = \underline{x}_i - \underline{y}_i.$$



Se asume que estas  $n$ -diferencias tienen una distribución normal-multivariada con vector de medias  $\underline{0}$  y matriz de Var-Cov dada por  $\Sigma$ .

La hipótesis a probar, es equivalente a probar: 
$$\begin{cases} H_0 : \underline{\mu}_D = \underline{0} \\ H_0 : \underline{\mu}_D \neq \underline{0} \end{cases}$$

Bajo  $H_0$ -cierta el estadístico de prueba a usar es:

$$T^2 = n\overline{\underline{D}}^t \underline{S}_D^{-1} \overline{\underline{D}} \sim \frac{(n-1)p}{n-p} F_{p,n-p}, \text{ donde :}$$

$$\overline{\underline{D}} = \frac{1}{n} \sum_{i=1}^n \underline{D}_i \quad \text{y} \quad \underline{S}_D = \frac{1}{n-1} \sum_{i=1}^n (\underline{D}_i - \overline{\underline{D}})(\underline{D}_i - \overline{\underline{D}})^t,$$

son el vector de medias y la matriz de Var-Cov de las diferencias muestrales.

Se rechaza  $H_0$  si:  $T_0^2 > \frac{(n-1)p}{n-p} F_{\alpha;p,n-p}.$

**Ejemplo-5:** Se desea comparar dos tipos de esmalte para la resistencia a la corrosión, 15 piezas de tubería fueron cubiertas con cada tipo de esmalte. Dos tuberías, cada una con un esmalte diferente, se enterraron y se dejaron durante el mismo período de tiempo en 15 lugares distintos; esto corresponde a un par de observaciones en condiciones similares, excepto por el tipo de cubrimiento. El efecto por la corrosión fue medido a través de dos variables:  $X_1$ -Profundidad máxima de la picadura por corrosión (milésimas de pulgadas) y  $X_2$ -Número de picaduras por corrosión. Los datos y las respectivas diferencias aparecen en la siguiente tabla:

La hipótesis a probar es: 
$$\begin{cases} H_0 : \underline{\mu}_D = \underline{0} \\ H_0 : \underline{\mu}_D \neq \underline{0} \end{cases}$$

donde  $\underline{\mu}_D$ -representa el vector de medias para las diferencias.

Localidad	Esmalte-1		Esmalte-2		Diferencia	
	$X_1$	$X_2$	$Y_1$	$Y_2$	$D_{E1}$	$D_{E2}$
1	73	31	51	35	22	-4
2	43	19	41	14	2	5
3	47	22	43	19	4	3
4	53	26	41	29	42	-3
5	58	36	47	34	11	2
6	47	30	32	26	15	4
7	52	29	24	19	28	10
8	38	36	43	37	-5	-1
9	61	34	53	24	8	10
10	56	33	52	27	4	6
11	56	19	57	14	-1	5
12	34	19	44	19	-10	0
13	55	26	57	30	-2	-4
14	65	15	40	7	25	8
15	75	18	68	13	7	5

Los resultados muestrales son:

$$\underline{d} = \begin{bmatrix} 8.000 \\ 3.067 \end{bmatrix}, \quad \mathbf{s}_D = \begin{bmatrix} 121.571 & 17.071 \\ 17.171 & 21.781 \end{bmatrix}$$

Bajo  $H_0$ -cierta, el estadístico de prueba es:

$$T_0^2 = (15) \begin{bmatrix} 8.000 & 3.067 \end{bmatrix} \begin{bmatrix} 121.571 & 17.071 \\ 17.171 & 21.781 \end{bmatrix}^{-1} \begin{bmatrix} 8.000 \\ 3.067 \end{bmatrix} = 10.815$$

Para  $\alpha = 0.05$ , se tiene que  $F_{\alpha;p,n-p} = F_{0.05;2,13} = 3.806$ , de donde,

$$\frac{(n-1)p}{n-p} F_{\alpha;p,n-p} = \frac{14(2)}{13} \times 3.806 = 8.198,$$

luego, como  $T_0^2 > 8.198$  entonces se rechaza  $H_0$  y se concluye que la evidencia muestral no apoya la hipótesis nula, ie. que hay diferencia entre los dos tipos de esmalte, con un nivel de significancia del 5%.

## PH de Contrastes para el vector de medias Poblacional $\underline{\mu}$ , de una $N_p(\underline{\mu}, \Sigma)$

### a) $\Sigma$ -desconocida

Sea  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$  una m.a de una población normal  $p$ -variada con vector de medias  $\underline{\mu}$ -desconocido y matriz de varianzas-covarianzas  $\Sigma$ - desconocida, ie.  $\underline{x}_i \sim N_p(\underline{\mu}, \Sigma)$ .

Sea  $C_{k \times p}$ -una matriz de constantes.  $C$ -contiene los coeficientes para  $k$ -combinaciones lineales simultáneas de las componentes de  $\underline{\mu} = (\mu_1, \mu_2, \dots, \mu_p)^t$ , es decir:

$$C_{\underline{\mu}} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{k1} & c_{k2} & \cdots & c_{kp} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \begin{bmatrix} c_{11}\mu_1 + c_{12}\mu_2 + \cdots + c_{1p}\mu_p \\ c_{21}\mu_1 + c_{22}\mu_2 + \cdots + c_{2p}\mu_p \\ \vdots \\ c_{k1}\mu_1 + c_{k2}\mu_2 + \cdots + c_{kp}\mu_p \end{bmatrix}_{k \times 1}$$

Se desea contrastar las hipótesis:

$$\begin{cases} H_0 : C\underline{\mu} = \underline{\gamma} \\ H_0 : C\underline{\mu} \neq \underline{\gamma} \end{cases}, \quad \text{con} \quad \underline{\gamma} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_k \end{pmatrix} - \text{Vector de Constantes.}$$

Un estimador insesgado para  $C\underline{\mu}$  es:  $C\underline{\bar{x}}$ , el cual tiene la siguiente distribución:

$$C\underline{\bar{x}} \sim N_k \left( C\underline{\mu}, C\underline{\Sigma}_{\bar{x}} C^t \right), \quad \text{es decir :}$$

$$C\underline{\bar{x}} \sim N_k \left( C\underline{\mu}, \frac{1}{n} C\underline{\Sigma} C^t \right), \quad \text{pues :} \quad \underline{\Sigma}_{\bar{x}} = \frac{\underline{\Sigma}}{n}.$$

Como  $\underline{\Sigma}$ -es desconocida se usa la estadística de prueba:

$$T_0^2 = n(C\underline{\bar{x}} - \underline{\gamma})^t [C\underline{S} C^t]^{-1} (C\underline{\bar{x}} - \underline{\gamma}) \sim \frac{(n-1)k}{n-k} F_{k, n-k} = cF$$

Se rechaza  $H_0$  si:  $T_0^2 > cF$  ó  $F_0 = \frac{1}{c} T_0^2 > F_{\text{tabla}}, \quad c = \frac{(n-1)k}{n-k}$

b)  $\Sigma$ -es conocida

En este caso se utiliza:  $\chi_0^2 = n(C\bar{\mathbf{x}} - \underline{\gamma})^t [C\mathbf{\Sigma}C^t]^{-1} (C\bar{\mathbf{x}} - \underline{\gamma}) \sim \chi_k^2$ .

Se rechaza  $H_0$  : si  $\chi_0^2 > \chi_{\alpha;k}^2$ .

**Ejemplo-6:** Los datos contenidos en la siguiente tabla, corresponden a los pesos (en gramos) del corcho encontrado en muestras tomadas de 28 árboles cultivados en una parcela experimental. Estas muestras fueron tomadas en 4-direcciones: Norte (N), este (E), sur (S) y oeste (O). En este caso se tienen 4-variables, que corresponden a las mediciones tomadas en cada dirección. Se quiere verificar si los pesos medios de corcho son iguales en la dirección norte-sur y en la dirección este-oeste. Se asume que estas cuatro mediciones tienen una distribución normal 4-variada, con vector de medias  $\underline{\mu} = (\mu_N, \mu_E, \mu_S, \mu_O)$  y matriz de varianzas-covarianzas  $\mathbf{\Sigma}$ , ambas cantidades desconocidas. Es decir se desea probar:  $H_0 : \mu_1 = \mu_3 \text{ y } \mu_2 = \mu_4$ .

N	E	S	O	N	E	S	O
$X_1$	$X_2$	$X_3$	$X_4$	$X_1$	$X_2$	$X_3$	$X_4$
72	66	76	77	91	79	100	75
60	53	66	63	56	68	47	50
56	57	64	58	79	65	70	61
41	29	36	38	81	80	68	58
32	32	35	36	78	55	67	60
30	35	34	26	46	38	37	38
39	39	31	27	39	35	34	37
42	43	31	25	32	30	30	32
37	40	31	25	60	50	67	54
33	29	27	36	35	37	48	39
32	30	34	28	39	36	39	31
63	45	74	63	50	34	37	40
54	46	60	52	43	37	39	50
47	51	52	43	48	54	57	43



Esta hipótesis se puede expresar como:

$$\begin{cases} H_0 : C\mu = \underline{\gamma} \\ H_0 : C\mu \neq \underline{\gamma} \end{cases}, \quad \text{con} \quad \underline{\gamma} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

$$\text{y con : } C = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}, \quad \text{y} \quad \underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix}.$$

Después de realizar algunos cálculos previos, se obtiene el vector de medias y la matriz de var-cov muestrales dadas por:

$$\bar{\underline{x}} = \begin{bmatrix} 50.535 \\ 46.179 \\ 49.679 \\ 45.179 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 290.41 & 233.75 & 288.44 & 226.27 \\ & 219.93 & 229.06 & 171.37 \\ & & 350.00 & 259.54 \\ & & & 226.00 \end{bmatrix}$$

De igual forma se tiene que:

$$\mathbf{C}\bar{\mathbf{x}} = \begin{bmatrix} 0.857 \\ 1.00 \end{bmatrix}, \quad \mathbf{CSC}^t = \begin{bmatrix} 63.53 & 27.96 \\ 27.96 & 103.19 \end{bmatrix}$$

Luego, bajo  $H_0$ -cierto y con:  $\underline{\gamma} = \underline{0}$ , se tiene que el estadístico de prueba es:

$$T_0^2 = n(\mathbf{C}\bar{\mathbf{x}} - \underline{\gamma})^t [\mathbf{CSC}^t]^{-1} (\mathbf{C}\bar{\mathbf{x}} - \underline{\gamma}) \sim \frac{(n-1)k}{n-k} F_{k,n-k}$$

$$T_0^2 = 28 \begin{bmatrix} 0.857 & 1.00 \end{bmatrix} \begin{bmatrix} 63.53 & 27.96 \\ 27.96 & 103.19 \end{bmatrix}^{-1} \begin{bmatrix} 0.857 \\ 1.00 \end{bmatrix} = 28(0.0158) = 0.443$$

Para  $\alpha = 0.05$ , se tiene:  $F_{\alpha; k, n-k} = F_{0.05; 2, 26} = 3.369$ . Ahora,

$$cF_{\text{tabla}} = \frac{(n-1)k}{n-k} F_{\alpha; k, n-k} = \frac{(28-1)2}{28-2} \times 3.69 = \frac{27}{13} \times 3.369 = 6.7$$

Luego, como  $T_0^2 = 0.4426 < 6.7 = cF_{\text{tabla}}$ , entonces no se rechaza a  $H_0$  y se concluye que la evidencia muestral parece ser coherente con la hipótesis planteada a un nivel de significancia del 5%, es decir, el contenido medio de corcho en la direcciones indicadas no son significativamente diferentes.

## Prueba de Razón de Verosimilitud ( $\Sigma$ -Desconocida)

Sea  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$  una m.a de una distribución  $N_p(\underline{\mu}, \underline{\Sigma})$ , donde tanto  $\underline{\mu}$  como  $\underline{\Sigma}$  son desconocidas.

### Estadística de razón de verosimilitud

Se desea contrastar las hipótesis:

$$\begin{cases} H_0 : \underline{\mu} = \underline{\mu}_0 \\ H_a : \underline{\mu} \neq \underline{\mu}_0 \end{cases} \iff \begin{cases} H_0 : \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \begin{bmatrix} \mu_{10} \\ \mu_{20} \\ \vdots \\ \mu_{0p} \end{bmatrix} \\ v.s \quad H_a : \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \neq \begin{bmatrix} \mu_{10} \\ \mu_{20} \\ \vdots \\ \mu_{0p} \end{bmatrix} \end{cases}$$

Ya se tiene una prueba para esto basada en la estadística  $T^2$ -de Hotelling, ahora se construirá otra prueba que será equivalente a la  $T^2$ -de Hotelling, pero que tiene algunas propiedades estadísticas deseables.

La estadística de Razón de Verosimilitud se define como sigue:

$$\begin{aligned}\lambda &:= \frac{\underset{\underline{\mu}, \underline{\Sigma}}{\text{Máx}} L(\underline{\mu}_0, \underline{\Sigma})}{\underset{\underline{\mu}, \underline{\Sigma}}{\text{Máx}} L(\underline{\mu}, \underline{\Sigma})} = \frac{\text{Máximo de L-Restringida}}{\text{Máximo de L-No Restringida}} \\&= \frac{\text{Máximo de L-Bajo } H_0\text{-Cierta}}{\text{Máximo de L-General}} \\&= \frac{\text{Máximo de L-Sobre el Espacio de Parámetros Restringido}}{\text{Máximo de L-Sobre todo el espacio de parámetros}}\end{aligned}$$

donde,  $L$ -es la función de Verosimilitud de los Datos.

La Estadística  $\lambda$  toma valores entre cero y uno, ie.  $0 \leq \lambda \leq 1$ .

Para valores grandes de  $\lambda$ , ie.  $\lambda$ -ceranos a 1, No se Rechaza  $H_0$ , mientras que para valores pequeños se Rechaza  $H_0$ .

Ahora se hallará una expresión para la Estadística de Razón de Verosimilitud y su respectiva distribución de probabilidad.

Primero recordemos que la función de verosimilitud asociada a una m.a de una distribución normal p-variada es:

$$\begin{aligned} L(\underline{\mu}, \underline{\Sigma}) &= L(\underline{\mu}, \underline{\Sigma} \mid \underline{x}_1, \underline{x}_1, \dots, \underline{x}_n) \\ &= \frac{1}{(2\pi)^{np/2}} \frac{1}{|\underline{\Sigma}|^{n/2}} e^{-\frac{1}{2} \sum_{j=1}^n (\underline{x}_j - \underline{\mu})' \underline{\Sigma}^{-1} (\underline{x}_j - \underline{\mu})} \end{aligned}$$

{ y que los estimadores de Máxima-Verosimilitud de  $\underline{\mu}$  y  $\underline{\Sigma}$  son, respectivamente: }  $\hat{\underline{\mu}} = \bar{\underline{x}}$  y

$$\hat{\underline{\Sigma}} = \underline{S}_n = \frac{1}{n} \sum_{i=1}^n (\underline{x}_i - \hat{\underline{\mu}})(\underline{x}_i - \hat{\underline{\mu}})^t = \frac{1}{n} \sum_{i=1}^n (\underline{x}_i - \bar{\underline{x}})(\underline{x}_i - \bar{\underline{x}})^t = \frac{(n-1)}{n} \underline{S}$$

Luego, bajo  $H_0 : \underline{\mu} = \underline{\mu}_0$ -Cierta, se tiene que el MLE de  $\underline{\Sigma}$  es:

$$\hat{\underline{\Sigma}}_0 = \frac{1}{n} \sum_{i=1}^n (\underline{x}_i - \underline{\mu}_0)(\underline{x}_i - \underline{\mu}_0)^t$$

luego, reemplazando las expresiones anteriores en la definición de  $\lambda$  se tiene que:

$$\begin{aligned}
 \lambda &:= \frac{\underset{\underline{\mu}, \underline{\Sigma}}{\text{Máx}} L(\underline{\mu}_0, \underline{\Sigma})}{\underset{\underline{\mu}, \underline{\Sigma}}{\text{Máx}} L(\underline{\mu}, \underline{\Sigma})} = \frac{\underset{\hat{\underline{\mu}}, \hat{\underline{\Sigma}}}{\text{Máx}} L(\underline{\mu}_0, \hat{\underline{\Sigma}}_0)}{\underset{\hat{\underline{\mu}}, \hat{\underline{\Sigma}}}{\text{Máx}} L(\hat{\underline{\mu}}, \hat{\underline{\Sigma}})} = \frac{\underset{\hat{\underline{\mu}}, \hat{\underline{\Sigma}}}{\text{Máx}} L(\underline{\mu}_0, \hat{\underline{\Sigma}}_0)}{\underset{\hat{\underline{\mu}}, \hat{\underline{\Sigma}}}{\text{Máx}} L(\underline{\bar{x}}, \hat{\underline{\Sigma}})} \\
 &= \frac{\underset{\hat{\underline{\mu}}, \hat{\underline{\Sigma}}}{\text{Máx}} \frac{1}{(2\pi)^{np/2}} \frac{1}{|\hat{\underline{\Sigma}}_0|^{n/2}} e^{-\frac{1}{2} \sum_{j=1}^n (x_j - \underline{\mu}_0)' \hat{\underline{\Sigma}}_0^{-1} (x_j - \underline{\mu}_0)}}{\underset{\hat{\underline{\mu}}, \hat{\underline{\Sigma}}}{\text{Máx}} \frac{1}{(2\pi)^{np/2}} \frac{1}{|\hat{\underline{\Sigma}}|^{n/2}} e^{-\frac{1}{2} \sum_{j=1}^n (x_j - \underline{\bar{x}})' \hat{\underline{\Sigma}}^{-1} (x_j - \underline{\bar{x}})}} \\
 &= \frac{\underset{\hat{\underline{\mu}}, \hat{\underline{\Sigma}}}{\text{Máx}} \frac{1}{|\hat{\underline{\Sigma}}_0|^{n/2}} e^{-\frac{1}{2} \sum_{j=1}^n (x_j - \underline{\mu}_0)' \hat{\underline{\Sigma}}_0^{-1} (x_j - \underline{\mu}_0)}}{\underset{\hat{\underline{\mu}}, \hat{\underline{\Sigma}}}{\text{Máx}} \frac{1}{|\hat{\underline{\Sigma}}|^{n/2}} e^{-\frac{1}{2} \sum_{j=1}^n (x_j - \underline{\bar{x}})' \hat{\underline{\Sigma}}^{-1} (x_j - \underline{\bar{x}})}} \\
 &= \frac{\underset{\hat{\underline{\mu}}, \hat{\underline{\Sigma}}}{\text{Máx}} \frac{1}{|\hat{\underline{\Sigma}}_0|^{n/2}} e^{-\frac{1}{2} \text{tr} \left\{ \hat{\underline{\Sigma}}_0^{-1} \sum_{j=1}^n \left[ (x_j - \underline{\mu}_0)(x_j - \underline{\mu}_0)' \right] \right\}}}{\underset{\hat{\underline{\mu}}, \hat{\underline{\Sigma}}}{\text{Máx}} \frac{1}{|\hat{\underline{\Sigma}}|^{n/2}} e^{-\frac{1}{2} \text{tr} \left\{ \hat{\underline{\Sigma}}^{-1} \sum_{j=1}^n \left[ (x_j - \underline{\bar{x}})(x_j - \underline{\bar{x}})' \right] \right\}}} = \frac{\frac{1}{|\hat{\underline{\Sigma}}_0|^{n/2}} e^{-\frac{np}{2}}}{\frac{1}{|\hat{\underline{\Sigma}}|^{n/2}} e^{-\frac{np}{2}}} = \frac{|\hat{\underline{\Sigma}}|^{n/2}}{|\hat{\underline{\Sigma}}_0|^{n/2}}
 \end{aligned}$$

## En Resumen:

$$\lambda = \frac{|\hat{\Sigma}|^{n/2}}{|\hat{\Sigma}_0|^{n/2}} = \frac{|\frac{1}{n} \sum_{i=1}^n (\underline{x}_i - \bar{\underline{x}})(\underline{x}_i - \bar{\underline{x}})^t|^{n/2}}{|\frac{1}{n} \sum_{i=1}^n (\underline{x}_i - \underline{\mu}_0)(\underline{x}_i - \underline{\mu}_0)^t|^{n/2}}$$

$$= \frac{|\sum_{i=1}^n (\underline{x}_i - \bar{\underline{x}})(\underline{x}_i - \bar{\underline{x}})^t|^{n/2}}{|\sum_{i=1}^n (\underline{x}_i - \underline{\mu}_0)(\underline{x}_i - \underline{\mu}_0)^t|^{n/2}}$$

$$\lambda^{2/n} = \frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|}$$

Haciendo,  $A = \sum_{i=1}^n (\underline{x}_i - \bar{\underline{x}})(\underline{x}_i - \bar{\underline{x}})^t$ , se tiene que:

$$\sum_{i=1}^n (\underline{x}_i - \underline{\mu}_0)(\underline{x}_i - \underline{\mu}_0)^t = \sum_{i=1}^n (\underline{x}_i - \bar{\underline{x}})(\underline{x}_i - \bar{\underline{x}})^t + n(\bar{\underline{x}} - \underline{\mu}_0)(\bar{\underline{x}} - \underline{\mu}_0)^t$$

$$= A + n(\bar{\underline{x}} - \underline{\mu}_0)(\bar{\underline{x}} - \underline{\mu}_0)^t$$

es decir, que

$$\lambda = \frac{|A|^{n/2}}{|A + n(\underline{\bar{x}} - \underline{\mu}_0)(\underline{\bar{x}} - \underline{\mu}_0)^t|^{n/2}}$$

de donde,

$$\begin{aligned}\lambda^{2/n} &= \frac{|A|}{|A + n(\underline{\bar{x}} - \underline{\mu}_0)(\underline{\bar{x}} - \underline{\mu}_0)^t|} \\ &= \frac{|A|}{|A + \underbrace{\sqrt{n}(\underline{\bar{x}} - \underline{\mu}_0)}_{\underline{b}} \underbrace{\sqrt{n}(\underline{\bar{x}} - \underline{\mu}_0)^t}_{\underline{b}^t}|} \\ \lambda &= \frac{|A|}{|A + \underline{b}\underline{b}^t|}\end{aligned}$$



## Propiedad de Determinantes:

$$|A + \underline{b}\underline{b}^t| = |A| \left(1 + \underline{b}^t A^{-1} \underline{b}\right) = |A| + |A| \underline{b}^t A^{-1} \underline{b}$$

Usando esta propiedad con:

$$A = \sum_{i=1}^n (\underline{x}_i - \bar{\underline{x}})(\underline{x}_i - \bar{\underline{x}})^t = (n-1)\mathbf{S}$$

y

$$\underline{b} = \sqrt{n}(\bar{\underline{x}} - \underline{\mu}_0)$$

se tiene que:

$$\begin{aligned}
 & |A + \underbrace{\sqrt{n}(\bar{\mathbf{x}} - \underline{\mu}_0)}_{\underline{b}} \underbrace{\sqrt{n}(\bar{\mathbf{x}} - \underline{\mu}_0)^t}_{\underline{b}}| \\
 &= |A| + |A| \left\{ \left[ \sqrt{n}(\bar{\mathbf{x}} - \underline{\mu}_0) \right]^t [(n-1)\mathbf{S}]^{-1} \left[ \sqrt{n}(\bar{\mathbf{x}} - \underline{\mu}_0) \right] \right\} \\
 &= |A| + |A| \left( \frac{1}{n-1} \right) \underbrace{n(\bar{\mathbf{x}} - \underline{\mu}_0)^t \mathbf{S}^{-1} (\bar{\mathbf{x}} - \underline{\mu}_0)} \\
 &= |A| + |A| \left( \frac{1}{n-1} \right) T_0^2
 \end{aligned}$$

$$\begin{aligned}
\text{Luego, } \lambda^{2/n} &= \frac{|A|}{|A + n(\bar{\mathbf{x}} - \underline{\mu}_0)(\bar{\mathbf{x}} - \underline{\mu}_0)^t|} \\
&= \frac{|A|}{|A| + |A| \left(\frac{1}{n-1}\right) n(\bar{\mathbf{x}} - \underline{\mu}_0)^t \mathbf{S}^{-1}(\bar{\mathbf{x}} - \underline{\mu}_0)} \\
&= \frac{1}{1 + \left(\frac{1}{n-1}\right) n(\bar{\mathbf{x}} - \underline{\mu}_0)^t \mathbf{S}^{-1}(\bar{\mathbf{x}} - \underline{\mu}_0)} \\
&= \frac{1}{1 + \left(\frac{1}{n-1}\right) T_0^2} = \frac{1}{1 + \frac{T_0^2}{n-1}}
\end{aligned}$$

$$\text{A la Estadística: } \lambda^{2/n} = \frac{1}{1 + \frac{T_0^2}{n-1}} = \frac{n-1}{(n-1) + T_0^2}$$

se le llama **Estadística Lambda de Wilks**.

De lo anterior se tiene que:

$$\lambda = \left[ \frac{n-1}{(n-1) + T_0^2} \right]^{n/2}$$

$$\text{y por tanto: } T_0^2 = (n-1) \left[ \frac{1 - \lambda_0^{2/n}}{\lambda_0^{2/n}} \right]$$

De donde:

Rechazar  $H_0$  si  $\lambda$ -Es pequeño,  $\iff T_0^2$  - Es Grande.

En la prueba de RV se rechaza  $H_0$  si

$$\lambda = \left( \frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|} \right)^{n/2} < C_\alpha$$

donde  $C_\alpha$ - es el percentil  $\alpha$  de la distribución de  $\lambda$ , llamada Distribución Lambda de Wilks.

Pero no es necesario usar la distribución de  $\lambda$ , debido a la relación que existe entre el estadístico  $\lambda$ -de Wilks y el estadístico  $T^2$ -de Hotelling.

Una Expresión para  $T^2$ -sin necesidad de hallar  $\mathbf{S}^{-1}$  es:

$$T^2 = n(\bar{\mathbf{x}} - \underline{\mu}_0)^t \mathbf{S}^{-1} (\bar{\mathbf{x}} - \underline{\mu}_0)$$

$$T^2 = (n - 1) \left[ \frac{|\hat{\underline{\Sigma}}_0|}{|\hat{\underline{\Sigma}}|} - 1 \right]$$

$$= (n - 1) \left[ \frac{|\sum_{i=1}^n (\mathbf{x}_i - \underline{\mu}_0)(\mathbf{x}_i - \underline{\mu}_0)^t|}{|\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^t|} - 1 \right]$$

Los métodos de verosimilitud, producen Estadísticos de prueba que se reducen a la familia de Estadísticas  $F$  y a la familias de estadísticas  $t$  en el caso univariado.

## Método de Máxima-Verosimilitud Generalizado

Sea  $\underline{\theta}$  el vector de todos los parámetros poblacionales desconocidos, de una distribución multivariada. Sea  $\underline{\Theta}$  el espacio del vector de parámetros  $\underline{\theta}$ , ie.  $\underline{\theta} \in \underline{\Theta}$ , con  $v = \dim(\underline{\Theta})$ . Sea  $L(\underline{\theta})$  la función de verosimilitud obtenida al evaluar la f.d.p conjunta de  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$  en sus valores observados  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ .

Para la prueba de hipótesis dada

$$\text{por: } \begin{cases} H_0 : \underline{\theta} \in \underline{\Theta}_0 \\ H_a : \underline{\theta} \in \underline{\Theta}_0^c = \underline{\Theta} - \underline{\Theta}_0 \end{cases}$$

La estadística de Razon de Verosimilitud es:  $\lambda := \frac{\underset{\underline{\theta} \in \underline{\Theta}_0}{\text{Máx}} L(\underline{\theta})}{\underset{\underline{\theta} \in \underline{\Theta}}{\text{Máx}} L(\underline{\theta})}$

$H_0$ -es rechazada cuando  $\lambda$ -es pequeña.

**Teorema:** Cuando  $n$ -es grande, bajo  $H_0$ -cierto, se cumple que:

$$-2\text{Log}\lambda = -2\text{Log}\left(\frac{\underset{\underline{\theta} \in \underline{\Theta}_0}{\text{Máx}} L(\underline{\theta})}{\underset{\underline{\theta} \in \underline{\Theta}}{\text{Máx}} L(\underline{\theta})}\right) \sim \chi^2_{v-v_0=v^*},$$

con,  $v_0 = \dim(\underline{\Theta}_0)$ , ie.  $v^* = \dim(\underline{\Theta}) - \dim(\underline{\Theta}_0) = v - v_0$ .

$v_0$ -número de parámetros desconocidos bajo  $H_0$ -cierta.

**NOTA:** El Test de RV tiene la mayor potencia, ie.

$$1 - \beta = P\left[\text{Rechazar } H_0 \mid H_0 \text{ es Falsa}\right],$$

entre todas las pruebas con el mismo nivel de significancia  $\alpha$ , ie.

$$\alpha = P\left[\text{Rechazar } H_0 \mid H_0 \text{ es V}\right]$$

**Ejemplo-7:** Para el caso de una distribución Normal  $p$ -variada, ie.  $N_p(\underline{\mu}, \underline{\Sigma})$ , se tiene que:

$$\underline{\theta}^t = [\mu_1, \dots, \mu_p, \sigma_{11}, \dots, \sigma_{1p}, \sigma_{21}, \dots, \sigma_{2p}, \dots, \sigma_{p1}, \dots, \sigma_{pp}]$$

$\underline{\Theta}$ -consiste un un espacio  $p$ -dimensional (de  $p$ -medias), combinado con un espacio  $p(p+1)/2$ -dimensional (de varianzas-covarianzas,  $p$ -varianzas y  $\binom{p}{2}$ -covarianzas), ie. que:

$$v = \dim(\underline{\Theta}) = p + \frac{p(p+1)}{2} = \frac{p(p+3)}{2}$$

Cuando  $H_0 : \underline{\mu} = \underline{\mu}_0$  es cierto y  $\underline{\Sigma}$  es no especificada, ie. cuando  $\underline{\mu}$  es conocida por el valor de  $\underline{\mu}_0$  y  $\underline{\Sigma}$  es desconocida,  $\underline{\theta}$  caerá en un subconjunto  $\underline{\Theta}_0$  de  $\underline{\Theta}$ , ie. Bajo  $H_0$ -cierto:  $\underline{\theta} \in \underline{\Theta}_0 \subset \underline{\Theta}$ .

En este caso:

$$v_0 = \dim(\underline{\Theta}_0) = 0 + \frac{p(p+1)}{2} = \frac{p(p+1)}{2}$$



## Región de Confianza:

En esta situación se tiene que el conjunto de todos los  $\underline{\mu}$  para los cuales se cumple la desigualdad:

$$n(\underline{\bar{x}} - \underline{\mu})^t \underline{\Sigma}^{-1} (\underline{\bar{x}} - \underline{\mu}) \leq \chi_{\alpha;p}^2,$$

se denomina una Región de Confianza del  $(1 - \alpha)100\%$  para el vector de medias poblacionales  $\underline{\mu}$ .

La expresión anterior representa el interior y la superficie de un elipsoide con centro en  $\underline{\mu} = \underline{\bar{x}}$  y cuya forma, tamaño e inclinación, dependen de  $\underline{\Sigma}$  y de  $\chi_{\alpha;p}^2$ , particularmente de los valores y vectores propios de  $\underline{\Sigma}$ .

Por ejemplo, si  $\underline{\Sigma} = I_p$  entonces la región de confianza es una esfera.

Los valores de  $\underline{\mu}_0$  dentro de la región de confianza apoyan la hipótesis nula  $H_0 : \underline{\mu} = \underline{\mu}_0$ .

# Inferencia para la Matriz de Covarianza

Sea  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$  una m.a de una población normal  $p$ -variada con vector de medias  $\underline{\mu}$ -desconocida y matriz de varianzas-covarianzas  $\Sigma$ -Desconocida, ie.  $\underline{x}_i \sim N_p(\underline{\mu}, \Sigma)$ .

Se tiene interés en la siguiente PH:

$$\begin{cases} H_0 : \Sigma = \Sigma_0 \\ H_a : \Sigma \neq \Sigma_0 \end{cases}$$

donde,  $\Sigma_0$ -es una matriz de valores fijos conocida para  $\Sigma$  (usualmente planteada por experiencia previa).

Dependiendo de la forma particular de  $\Sigma_0$ , existen distintos nombres para la PH asociada:

- ①  $\Sigma_0$ -Cualquier matriz de valor fijo. Prueba general.
- ②  $\Sigma_0 = \Delta$ -Diagonal. Prueba de independencia de variables.
- ③  $\Sigma_0 = \sigma^2 I_p$ . Prueba de homocedasticidad e independencia.
- ④  $\Sigma_0 = I_p$ . Prueba de Esfericidad, ie. variables con varianzas unitarias e incorreladas.
- ⑤  $\Sigma_0 = B_m + \sigma^2 I_p$ , con  $B$ -de rango  $m < p$ . Prueba de homocedasticidad e independencia parcial. Con  $m = 0$ -se tiene la prueba de homocedasticidad e independencia.
- ⑥  $\Sigma_0 = B_m + I_p$ , con  $B$ -de rango  $m < p$ . Prueba de Esfericidad parcial. Con  $m = 0$ -se tiene la Prueba de Esfericidad.

La estadística de Razón de Verosimilitud para la PH considerada está dado por:

$$\lambda := \frac{\underset{\underline{\mu}}{\text{Máx}} L(\underline{\mu}, \mathbf{\Sigma}_0)}{\underset{\underline{\mu}, \mathbf{\Sigma}}{\text{Máx}} L(\underline{\mu}, \mathbf{\Sigma})} = \frac{\text{Máximo de L-Restingida}}{\text{Máximo de L-No Restingida}}$$

$$= \frac{\text{Máximo de L-Bajo } H_0\text{-Cierta}}{\text{Máximo de L-General}}$$

Luego de realizar los cálculos y simplificaciones necesarias se obtiene que:

$$\lambda = \left[ \left( \frac{n-1}{n} \right)^p \frac{|\mathbf{S}|}{|\mathbf{\Sigma}_0|} \right]^{\frac{n}{2}} \text{Exp} \left\{ -\frac{1}{2} \left[ (n-1) \text{tr}(\mathbf{S}\mathbf{\Sigma}_0^{-1}) - np \right] \right\}$$

haciendo:  $\frac{n-1}{n} \approx 1$ , ie.  $n-1 \approx n = v$ , ie.  $v = n-1 = n$ , se tiene:

$$\lambda = \frac{|\mathbf{S}|^{\frac{v}{2}}}{|\mathbf{\Sigma}_0|^{\frac{v}{2}}} \text{Exp} \left\{ -\frac{1}{2} \left[ v \text{tr}(\mathbf{S}\mathbf{\Sigma}_0^{-1}) - vp \right] \right\}$$

y haciendo  $\lambda^* = -2\log\lambda$ , se tiene que:

$$\lambda^* = v \left[ \text{Log}|\mathbf{\Sigma}_0| - \text{Log}|\mathbf{S}| + \text{tr}(\mathbf{S}\mathbf{\Sigma}_0^{-1}) - p \right]$$

Bajo  $H_0$ -cierta, se tiene que:

$$\lambda^* \sim \chi_k^2, \quad \text{para } n-1 \text{ grande}$$

con  $k = p + \frac{p(p+1)}{2} - p = \frac{p(p+1)}{2}$ , ie.

$k$ =Parámetros en  $\Theta$  menos parámetros en  $\Theta_0$  (ie. menos  $p$ ).

Rechazamos  $H_0$  si.

$$\lambda^* > \chi_{\alpha}^2 ; k$$

Ahora, otra forma alterna de  $\lambda^*$  es como sigue:

Si  $\lambda_1, \lambda_2, \dots, \lambda_p$  son los valores propios de  $\mathbf{S}\mathbf{\Sigma}_0^{-1}$ , entonces se sabe que:

$$\text{tr}(\mathbf{S}\mathbf{\Sigma}_0^{-1}) = \sum_{i=1}^p \lambda_i$$

y usando propiedades de determinantes se tiene que:

$$\text{Log}|\mathbf{\Sigma}_0| - \text{Log}|\mathbf{S}| = -\text{Log}|\mathbf{S}\mathbf{\Sigma}_0^{-1}| = -\text{Log}\left(\prod_{i=1}^p \lambda_i\right),$$

luego,

$$\lambda^* = v \left[ \text{Log}|\mathbf{\Sigma}_0| - \text{Log}|\mathbf{S}| + \text{tr}(\mathbf{S}\mathbf{\Sigma}_0^{-1}) - p \right]$$

$$= v \left[ -\text{Log} \left( \prod_{i=1}^p \lambda_i \right) + \sum_{i=1}^p \lambda_i - p \right]$$

$$\lambda^* = v \left[ \sum_{i=1}^p [\lambda_i - \text{Log} \lambda_i] - p \right] \sim \chi_k^2$$

y rechazamos  $H_0$  si  $\lambda^* > \chi_{\alpha}^2 ; k$

con  $k = \frac{p(p+1)}{2}$ .

Una Modificación para  $\lambda^*$ -fue propuesta por Bartlet, (**para el caso de muestras pequeñas**) la cual es: {

$$\lambda_1^* = \left\{ 1 - \frac{1}{6(n-1)} \left[ 2p + 1 - \frac{2}{p+1} \right] \right\} \lambda^* \sim \chi_k^2$$

} es decir,

$$\lambda_1^* = c\lambda^* \sim \chi_k^2$$

con

$$c = 1 - \frac{1}{6(n-1)} \left[ 2p + 1 - \frac{2}{p+1} \right]$$

que puede usarse para tamaños de muestras moderadamente pequeños.



**EJEMPLO:** Se tomaron 20 sujetos y se les midió los tiempos de reacción ante un estímulo en centésimas de segundo. A cada individuo se le midieron estos tiempos en 3 intervalos de tiempos distintos. Se asume que estas mediciones tienen una distribución  $N_3(\underline{\mu}, \underline{\Sigma})$ . Pruebe la hipótesis:

$$H_0 : \underline{\Sigma} = \begin{bmatrix} 4 & 3 & 2 \\ 3 & 6 & 5 \\ 2 & 5 & 10 \end{bmatrix} \quad v.s \quad H_a : \underline{\Sigma} \neq \begin{bmatrix} 4 & 3 & 2 \\ 3 & 6 & 5 \\ 2 & 5 & 10 \end{bmatrix}$$

**Solución:** Como

$$\underline{\Sigma}_0 = \begin{bmatrix} 4 & 3 & 2 \\ 3 & 6 & 5 \\ 2 & 5 & 10 \end{bmatrix}, \quad \text{luego} \quad \underline{\Sigma}_0^{-1} = \begin{bmatrix} 0.41 & -0.23 & 0.03 \\ -0.23 & 0.42 & -0.16 \\ 0.03 & -0.16 & 0.17 \end{bmatrix}$$

$$\mathbf{S}\boldsymbol{\Sigma}_0^{-1} = \begin{bmatrix} 0.85 & -0.01 & 0.03 \\ -0.58 & 1.68 & -0.08 \\ -0.41 & 0.72 & 0.68 \end{bmatrix}.$$

Los valores propios de  $\mathbf{S}\boldsymbol{\Sigma}_0^{-1}$  son:  $\lambda_1 = 1.61$ ,  $\lambda_2 = 0.87$  y  $\lambda_3 = 0.73$ , luego:  $\text{tr}(\mathbf{S}\boldsymbol{\Sigma}_0^{-1}) = \sum \lambda_i = 3.2216$ ,  $|\boldsymbol{\Sigma}_0| = 86$ ,  $|\mathbf{S}| = 88.6355$ , de donde:

$$\lambda^* = v \left[ \text{Log}|\boldsymbol{\Sigma}_0| - \text{Log}|\mathbf{S}| + \text{tr}(\mathbf{S}\boldsymbol{\Sigma}_0^{-1}) - p \right]$$

$$= 20 [\text{Log}(86) - \text{Log}(88.6355) + 3.2216 - 3]$$

$$\lambda^* = 3.63$$

$$\begin{aligned}\lambda_1^* &= \left\{ 1 - \frac{1}{6(n-1)} \left[ 2p + 1 - \frac{2}{p+1} \right] \right\} \lambda^* \\ &= \left\{ 1 - \frac{1}{6(19)} \left[ 2(3) + 1 - \frac{2}{3+1} \right] \right\} 3.63\end{aligned}$$

$$\lambda_1^* = 3.42$$

Para  $\alpha = 0.05$ , se tiene que:  $\chi_{\alpha;\nu} = \chi_{0.05;6} = 12.592$ .

En este caso,  $\nu = p(p+1)/2 = 3(4)/2 = 6$ , y como:

$$\lambda_1^* = 3.42 < 12.592 = \chi_{\alpha;\nu}^2,$$

entonces, no se rechaza  $H_0$  y se concluye que la evidencia muestral apoya la hipótesis:

$$H_0 : \mathbf{\Sigma} = \begin{bmatrix} 4 & 3 & 2 \\ 3 & 6 & 5 \\ 2 & 5 & 10 \end{bmatrix},$$

a un nivel de significancia del 5%.

## Dos o más Matrices de Covarianzas

Recordar que uno de los supuestos cuando se comparan dos o más vectores de medias, es que las respectivas matrices de Var-Cov asociadas a cada población diferente, sean iguales. Un test muy común para probar la igualdad de matrices de Var-Cov es el [M-Test de Box](#).

Suponga que se tienen  $g$ -poblaciones diferentes, con matrices de Var-Cov asociadas dadas por:  $\Sigma_1, \Sigma_2, \dots, \Sigma_g$ , respectivamente. Se desea contrastar las hipótesis:

$$\begin{cases} H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_g = \Sigma \\ H_a : \Sigma_i \neq \Sigma_j \text{ p.a } i, j \end{cases}$$

donde  $\Sigma$ -es una matriz de Var-Cov común.

Si se tienen  $g$ -muestras aleatorias, una para cada población, de tamaños  $n_1, n_2, \dots, n_g$ , respectivamente y además se asume que cada muestra proviene de una población con distribución normal  $p$ -variada, entonces el **estadístico de prueba de razón de verosimilitud para esta PH** es:

$$\lambda = \prod_{i=1}^g \left( \frac{|\mathbf{S}_i|}{|\mathbf{S}_p|} \right)^{\frac{n_i-1}{2}},$$

donde,  $\mathbf{S}_i$ -es la matriz de Var-Cov muestral asociada a la m.a de la  $i$ -ésima población,  $i = 1, 2, \dots, g$  y

$$\mathbf{S}_p = \frac{1}{\sum_{i=1}^g (n_i - 1)} \left[ \sum_{i=1}^g (n_i - 1) \mathbf{S}_i \right]$$

haciendo,  $\{v_i = n_i - 1\}$  y  $\{v = \sum_{i=1}^g v_i = \sum_{i=1}^g (n_i - 1)\}$ , se tiene que:

$$\mathbf{S}_p = \frac{1}{v} \left[ \sum_{i=1}^g v_i \mathbf{S}_i \right]$$

La **Estadística M de Box** se define como:

$M = -2\text{Log}\lambda \sim \chi^2$ , (n-grande).

$$\begin{aligned} M &= \left[ \sum_{i=1}^g (n_i - 1) \right] \text{Log}|\mathbf{S}_p| - \sum_{i=1}^g (n_i - 1) \text{Log}|\mathbf{S}_i| \\ &= v \text{Log}|\mathbf{S}_p| - \sum_{i=1}^g v_i \text{Log}|\mathbf{S}_i| \end{aligned}$$

Bajo  $H_0$ -cierto, se espera que las matrices de Var-Cov muestrales no sean muy diferentes, en cuyo caso, el valor de  $\lambda$  estaría cerca a uno y por lo tanto  $M$ -sería pequeño.

Ahora, sea

**EJEMPLO:** En el departamento de salud y servicios sociales de Wisconsin se realizó un estudio para investigar el efecto de la propiedad o la certificación ( o ambas) sobre los costos. Cuatro costos fueron seleccionados para el análisis; estos fueron calculados diariamente por paciente y fueron medidos en horas por paciente diario. Las variables fueron:  $X_1$ -Costo de la enfermería,  $X_2$ -Costo de alimentación,  $X_3$ -Costo de operación y mantenimiento y  $X_4$ -Costo de administración y lavandería. Se registraron un total de  $n = 516$ -observaciones de cada una de las cuatro variables, separadas previamente en tres grupos de interés: Privados ( $n_1 = 271$ ), Públicos ( $n_2 = 138$ ) y Gubernamentales ( $n_3 = 107$ ). Se asume que el vector  $\underline{x} = (X_1, X_2, X_3, X_4)^t$  tiene una distribución  $N_4(\underline{\mu}_i, \underline{\Sigma}_i)$ , para  $i = 1, 2, 3$ . Se desea probar la hipótesis:

$$\begin{cases} H_0 : \underline{\Sigma}_1 = \underline{\Sigma}_2 = \underline{\Sigma}_3 = \underline{\Sigma} \\ H_a : \underline{\Sigma}_i \neq \underline{\Sigma}_j \text{ p.a } i \neq j = 1, 2, 3 \end{cases}$$



Los resultados muestrales obtenidos a partir de los datos fueron:

$i = 1$  Privados

$$n_1 = 271, \quad \bar{\mathbf{x}}_1 = \begin{pmatrix} 2.066 \\ 0.480 \\ 0.082 \\ 0.36 \end{pmatrix}, \quad S_1 = \begin{pmatrix} 0.291 & & & \\ -0.01 & 0.011 & & \\ 0.02 & 0.000 & 0.001 & \\ 0.010 & 0.003 & 0.000 & 0.10 \end{pmatrix}$$

$i = 2$  Públicos

$$n_2 = 138, \quad \bar{\mathbf{x}}_1 = \begin{pmatrix} 2.167 \\ 0.596 \\ 0.124 \\ 0.418 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 0.561 & & & \\ 0.011 & 0.025 & & \\ 0.001 & 0.004 & 0.005 & \\ 0.037 & 0.007 & 0.002 & 0.019 \end{pmatrix}$$

$i = 3$  Gubernamentales

$$n_3 = 107, \quad \bar{\mathbf{x}}_1 = \begin{pmatrix} 2.273 \\ 0.521 \\ 0.125 \\ 0.383 \end{pmatrix}, \quad S_3 = \begin{pmatrix} 0.261 & & & \\ 0.030 & 0.017 & & \\ 0.003 & 0.000 & 0.004 & \\ 0.018 & 0.006 & 0.001 & 0.013 \end{pmatrix}$$

Con la información anterior se obtiene:

$$|\mathbf{S}_1| = 2.783 \times 10^{-8}, \quad \text{Log}|\mathbf{S}_1| = -17.397$$

$$|\mathbf{S}_2| = 89.539 \times 10^{-8}, \quad \text{Log}|\mathbf{S}_2| = -13.926$$

$$|\mathbf{S}_3| = 14.579 \times 10^{-8}, \quad \text{Log}|\mathbf{S}_3| = -15.741$$

$$|\mathbf{S}_p| = 17.398 \times 10^{-8}, \quad \text{Log}|\mathbf{S}_1| = -15.564$$

$$u = \left[ \frac{1}{270} + \frac{1}{137} + \frac{1}{106} - \frac{1}{270 + 137 + 106} \right] \left[ \frac{2(4)^2 + 3(4) - 1}{6(4 + 1)(3 - 1)} \right] = 0.0133$$

$$= (270 + 137 + 106)(-15.564) - [270(-17.397) + 137(-13.926) + 106(-15.741)] = 289.3$$

$$C = (1 - u)M = (1 - 0.0133)289.3 = 285.5$$

Ahora, para  $\alpha = 0.05$ , se tiene que:

$$\chi_{\alpha;k}^2 = \chi_{0.05; \frac{p(p+1)}{2}(g-1)}^2 = \chi_{0.05; 20}^2 = 31.34,$$

Como  $C > \chi^2_{\alpha;k}$ , entonces se rechaza  $H_0$  y se concluye que las matrices de Var-Cov asociadas a los vectores de costos para las tres poblaciones consideradas son diferentes a un nivel de significancia del 5%.

# REGIONES DE CONFIANZA Y COMPARACIONES SIMULTÁNEAS ENTRE LAS COMPONENTES DEL VECTOR DE MEDIAS $\underline{\mu}$

Extensión del concepto de intervalo de confianza univariado para un parámetro al caso de una región de confianza multivariada.

Considere el vector  $\underline{\theta}$  de parámetros desconocidos de una población, y sea  $\Theta$  su espacio paramétrico.

Una región de confianza para  $\underline{\theta}$  es un conjunto que contiene los posibles valores de  $\Theta$ .

Esta región está determinada por los datos, y será denotada por

$R(\mathbf{X})$  , donde

$$\mathbf{X} = \begin{bmatrix} \underline{X}_1 \\ \underline{X}_1 \\ \vdots \\ \underline{X}_n \end{bmatrix} \quad \text{es la matriz de datos}$$

Se dice que  $R(\mathbf{X})$  es una región con  $(1 - \alpha)\%$  de confianza para  $\underline{\mu}$  si, antes de obtener la muestra se tiene que:

$$P\left[R(\mathbf{X}) - \text{Contenga al verdadero valor de } \underline{\theta}\right] = 1 - \alpha$$

**Construcción de una región de confianza para  $\underline{\mu}$  cuando la población tiene distribución  $N_p(\underline{\mu}, \underline{\Sigma})$ , con  $\underline{\mu}$  y  $\underline{\Sigma}$  desconocidos**

Bajo normalidad multivariada se tiene que:

$$T^2 = n(\bar{\mathbf{x}} - \underline{\mu})^t \mathbf{S}^{-1}(\bar{\mathbf{x}} - \underline{\mu}) \sim \frac{(n-1)p}{(n-p)} F_{p, n-p} = kF,$$

se tiene que antes de obtener la muestra,

$$P \left[ n(\bar{\mathbf{x}} - \underline{\mu})^t \mathbf{S}^{-1}(\bar{\mathbf{x}} - \underline{\mu}) \leq \frac{(n-1)p}{(n-p)} F_{\alpha; p, n-p} \right] = 1 - \alpha.$$

Lo anterior quiere decir que:

el vector de medias  $\bar{\mathbf{x}}$  está dentro de una distancia de

$$\sqrt{\frac{(n-1)p}{(n-p)} F_{\alpha; p, n-p}} = \sqrt{kF} \quad \text{de } \underline{\mu} \quad \text{con probabilidad de } (1 - \alpha) .$$

**Definición:** Una región de confianza del  $(1 - \alpha)100\%$  para el vector de medias poblacionales  $\underline{\mu}$  de una población normal  $p$ -dimensional, es el conjunto de todos los valores  $\underline{\mu}$  que satisfacen:

$$n(\underline{\bar{x}} - \underline{\mu})^t \mathbf{S}^{-1}(\underline{\bar{x}} - \underline{\mu}) = T^2 \leq c^2 = kF = \frac{(n-1)p}{(n-p)} F_{\alpha; p, n-p}$$

es decir,

$$n(\underline{\bar{x}} - \underline{\mu})^t \mathbf{S}^{-1}(\underline{\bar{x}} - \underline{\mu}) \leq c^2$$

con

$$c^2 = \frac{(n-1)p}{(n-p)} F_{\alpha; p, n-p} = kF.$$

y en este caso, geométricamente la región es una elipsoide centrada en  $\underline{\bar{x}}$ .

Para determinar si algún  $\underline{\mu}_0$  cualesquiera caé dentro de la región de confianza (ie. si  $\underline{\mu}_0$  es un valor plausible o viable para  $\underline{\mu}$ ), se necesita calcular la distancia cuadrada generalizada dada por:

$$T_0^2 = n(\bar{\mathbf{x}} - \underline{\mu}_0)^t \mathbf{S}^{-1}(\bar{\mathbf{x}} - \underline{\mu}_0)$$

y compararla con el número percentil:

$$c^2 = kF = \frac{(n-1)p}{(n-p)} F_{\alpha; p, n-p},$$

si dicha distancia cuadrática es mayor que  $c^2$ , entonces  $\underline{\mu}_0$  no estará en la región de confianza.



Debido a que lo anterior, es el análogo a probar:

$$\begin{cases} H_0 : \underline{\mu} = \underline{\mu}_0 \\ H_a : \underline{\mu} \neq \underline{\mu}_0 \end{cases}$$

se observa que la región de confianza consiste de todos aquellos vectores  $\underline{\mu}_0$  para los cuales la prueba  $T^2$ -de Hotelling no rechaza a  $H_0$  en favor de  $H_a$  a un nivel de significancia de  $\alpha$ .

Para el caso  $p = 2$ , los ejes de la región de confianza o Elipse de confianza y sus respectivas longitudes relativas, son determinados a partir de los eigen-valores y eigen-vectores de  $\mathbf{S}$ .

## Las semi-longitudes y direcciones del elipsoide de confianza

$$n(\bar{\mathbf{x}} - \underline{\boldsymbol{\mu}})^t \mathbf{S}^{-1}(\bar{\mathbf{x}} - \underline{\boldsymbol{\mu}}) \leq c^2 = \frac{(n-1)p}{(n-p)} F_{\alpha; p, n-p},$$

están dadas por:

$$\sqrt{\lambda_i} \frac{c}{\sqrt{n}} = \sqrt{\lambda_i} \frac{1}{\sqrt{n}} \sqrt{\frac{(n-1)p}{(n-p)} F_{\alpha; p, n-p}}$$

unidades medidas a lo largo de los eigen-vectores  $\mathbf{e}_i$ .

Iniciando en el centro  $\bar{\mathbf{x}}$ , los ejes del elipsoide de confianza son:

$$\pm \sqrt{\lambda_i} \frac{1}{\sqrt{n}} \sqrt{\frac{(n-1)p}{(n-p)} F_{\alpha; p, n-p}} \mathbf{e}_i$$

con  $\mathbf{S}\mathbf{e}_i = \lambda_i \mathbf{e}_i$  para  $i = 1, 2, \dots, p$ .

Los cocientes entre las  $\lambda_i$  ayudan a identificar las longitudes relativas de cada par de ejes.

**EJEMPLO:**  $n = 42$ -datos de radiación de hornos microondas y  $p = 2$ -variables.

Considere los datos de radiación de hornos microondas usados anteriormente, donde:

$$X_1 = \sqrt[4]{\text{Radiación medida con puerta cerrada}}$$

$$X_2 = \sqrt[4]{\text{Radiación medida con puerta Abierta}}$$

Para los  $n = 42$  pares de datos  $(x_1, x_2)$  se tienen los siguientes resúmenes descriptivos:

$$\bar{\mathbf{x}} = \begin{bmatrix} 0.564 \\ 40.603 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 0.0144 & 0.0117 \\ 0.0117 & 0.0146 \end{bmatrix} \quad \text{y} \quad \mathbf{S}^{-1} = \begin{bmatrix} 203.018 & -163.391 \\ -163.391 & 200.228 \end{bmatrix}$$

Con estas cantidades, la elipse del 95% de confianza para  $\underline{\mu}$  contiene todos los valores  $(\mu_1, \mu_2)$  que satisfacen:

$$\begin{aligned} 42 \begin{bmatrix} 0.564 - \mu_1 \\ 0.603 - \mu_2 \end{bmatrix}^t \begin{bmatrix} 203.018 & -163.391 \\ -163.391 & 200.228 \end{bmatrix} \begin{bmatrix} 0.564 - \mu_1 \\ 0.603 - \mu_2 \end{bmatrix} &\leq \frac{41(2)}{40} F_{0.05;2,40} \\ &= \frac{41(2)}{40} 3.23 \\ &= 6.62 \end{aligned}$$

o equivalentemente:

$$42(203.018)(0.564 - \mu_1)^2 + 42(200.228)(0.603 - \mu_2)^2 - 84(163.391)(0.564 - \mu_1)(0.603 - \mu_2) \leq 6.62$$

¿Cae el vector:

$$\underline{\mu} = \begin{bmatrix} 0.562 \\ 0.589 \end{bmatrix}$$

dentro de la elipse del 95% de confianza?

Utilizando la elipse anterior donde se reemplazan los valores de  $\mu_1$  y  $\mu_2$  por los valores dados en  $\underline{\mu}$ , se obtiene:

$$1.30 < 6.62$$

y por lo tanto el vector

$$\underline{\mu} = \begin{bmatrix} 0.562 \\ 0.589 \end{bmatrix}$$

si cae dentro de la región de confianza del 95%.

Equivalentemente, en la prueba de:

$$H_0 : \underline{\mu} = \begin{bmatrix} 0.562 \\ 0.589 \end{bmatrix} \quad \text{contra:} \quad H_1 : \underline{\mu} \neq \begin{bmatrix} 0.562 \\ 0.589 \end{bmatrix}$$

un nivel de significancia de  $\alpha = 0.05$ , no se rechaza  $H_0$ .

## Ahora pasemos a graficar la Región de Confinaza obtenida

El centro de la elipse es:

$$\bar{\mathbf{x}} = \begin{bmatrix} 0.564 \\ 0.603 \end{bmatrix}$$

y las semi-longitudes de los ejes mayor y menor son, respectivamente,

$$\sqrt{\lambda_1} \frac{1}{\sqrt{n}} \sqrt{\frac{(n-1)p}{(n-p)} F_{\alpha; p, n-p}} = \sqrt{0.026} \frac{1}{\sqrt{42}} \sqrt{\frac{(41)^2}{(40)} 3.323} = 0.064$$

y

$$\sqrt{\lambda_2} \frac{1}{\sqrt{n}} \sqrt{\frac{(n-1)p}{(n-p)} F_{\alpha; p, n-p}} = \sqrt{0.002} \frac{1}{\sqrt{42}} \sqrt{\frac{(41)^2}{(40)} 3.323} = 0.018$$

Los ejes caen a lo largo de:

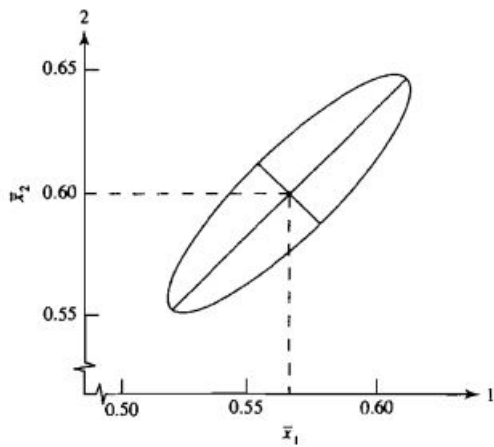
$$\underline{e}_1 = \begin{bmatrix} 0.704 \\ 0.710 \end{bmatrix} \quad \text{y} \quad \underline{e}_2 = \begin{bmatrix} -0.710 \\ 0.704 \end{bmatrix}$$

cuando estos vectores son graficados desde el origen:

$\underline{\bar{x}} = \begin{bmatrix} 0.564 \\ 0.603 \end{bmatrix}$ . Ahora, una indicación de la elongación de la elipse de confianza está dada por el cociente entre las longitudes del eje mayor y del eje menor, ie.

$$\frac{2\sqrt{\lambda_1} \frac{1}{\sqrt{n}} \sqrt{\frac{(n-1)p}{(n-p)} F_{\alpha; p, n-p}}}{2\sqrt{\lambda_2} \frac{1}{\sqrt{n}} \sqrt{\frac{(n-1)p}{(n-p)} F_{\alpha; p, n-p}}} = \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}} = \frac{\sqrt{0.026}}{\sqrt{0.002}} = \frac{0.161}{0.045} = 3.6$$

Lo que indica que el eje mayor es 3.6 veces el eje menor.





# INTERVALOS DE CONFIANZA SIMULTÁNEOS PARA LAS COMPONENTES DEL VECTOR DE MEDIAS $\underline{\mu}$

Mientras que las regiones de confianza proporcionan un conocimiento conjunto sobre los valores plausibles para  $\underline{\mu}$ , generalmente las conclusiones de un estudio **también incluyen 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 sus medias bajo una probabilidad especificada.

Esta clase de intervalos son llamados **Intervalos de Confianza Simultáneos.**

Al hacer lo anterior, se adopta la actitud de que toda las afirmaciones de confianza separadas o individuales **deben ser**

Se inicia considerando afirmaciones de confianza simultáneas que están intimamente relacionadas con la región de confianza basada en la estadística  $T^2$ -de Hotelling.

Sea  $\underline{\mathbf{x}} \sim N_p(\underline{\boldsymbol{\mu}}, \boldsymbol{\Sigma})$  y considere la c.l dada por:

$$Z = a_1X_1 + a_2X_2 + \cdots + a_pX_p = \underline{\mathbf{a}}^t \underline{\mathbf{x}}.$$

Claramente, se tiene que:

$$\mu_Z = E[Z] = \underline{\mathbf{a}}^t \underline{\boldsymbol{\mu}} \quad \text{y} \quad \sigma_Z^2 = \text{Var}[Z] = \underline{\mathbf{a}}^t \boldsymbol{\Sigma} \underline{\mathbf{a}}.$$

Además,

$$Z \sim N\left(\underline{\mathbf{a}}^t \underline{\boldsymbol{\mu}}, \underline{\mathbf{a}}^t \boldsymbol{\Sigma} \underline{\mathbf{a}}\right)$$

Si se tiene disponible una m.a  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$  de la distribución  $N_p(\underline{\mu}, \underline{\Sigma})$ , se tendrá también una m.a correspondiente de las  $Z$ 's, tomando las respectivas combinaciones lineales, ie.

$$Z_j = a_1 X_{j1} + a_2 X_{j2} + \dots + a_p X_{jp}, \quad \text{para } j = 1, 2, \dots, n.$$

La media y la varianza muestral de los valores observados de las  $Z$ 's: es decir de los datos:  $z_1, z_2, \dots, z_n$ , están dados por:

$$\overline{Z} = \underline{a}^t \underline{\bar{x}} \quad \text{y} \quad S_Z^2 = \underline{a}^t \underline{S} \underline{a},$$

donde,  $\underline{\bar{x}}$  y  $\underline{S}$  son el vector de medias muestrales y la matriz de Var-Cov muestrales de las  $\underline{x}_i$ , para  $i = 1, 2, \dots, n$ .

Los Intervalos de Confianza Simultáneos pueden ser desarrollados considerando los Intervalos de Confianza para  $\underline{a}^t \underline{\mu}$ , para cualesquier elección de  $\underline{a}$ .

Para  $\underline{a}$ -fijo y  $\sigma_Z^2$ -desconocida, Un I.C del  $(1 - \alpha)100\%$  para  $\mu_Z = \underline{a}^t \underline{\mu}$  está basado en la  $t$ -de Student dada por:

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

lo cual lleva al I.C para  $\mu_Z = \underline{a}^t \underline{\mu}$  dado por:

$$\bar{Z} - t_{\alpha/2; n-1} \frac{S_Z}{\sqrt{n}} \leq \mu_Z \leq \bar{Z} + t_{\alpha/2; n-1} \frac{S_Z}{\sqrt{n}}$$

o equivalentemente:

$$\underline{a}^t \bar{\mathbf{X}} - t_{\alpha/2; n-1} \frac{\sqrt{\underline{a}^t \mathbf{S} \underline{a}}}{\sqrt{n}} \leq \mu_Z \leq \underline{a}^t \bar{\mathbf{X}} + t_{\alpha/2; n-1} \frac{\sqrt{\underline{a}^t \mathbf{S} \underline{a}}}{\sqrt{n}}$$

ie.

$$\underline{a}^t \bar{\mathbf{X}} \pm t_{\alpha/2; n-1} \frac{\sqrt{\underline{a}^t \mathbf{S} \underline{a}}}{\sqrt{n}}$$

El intervalo anterior puede ser interpretado como una afirmación acerca de las componentes del vector de medias poblacionales  $\underline{\mu}$ ,

por ejemplo, con

$$\underline{a} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \text{ se tiene que: } \mu_Z = \underline{a}^t \underline{\mu} = \mu_1,$$

el anterior I.C es un intervalo de confianza usual para una media poblacional normal, en este caso:

$$\underline{a}^t \underline{S} \underline{a} = s_{11} = \text{Var}[X_1], \quad \underline{a}^t \underline{\bar{x}} = \bar{X}_1, \text{ de donde, el I.C es:}$$

$$\bar{X}_1 \pm t_{\alpha/2; n-1} \frac{\sqrt{s_{11}}}{\sqrt{n}}$$

De la misma forma, eligiendo el adecuado vector de constantes  $\underline{a}$  se pueden construir los intervalos de confianza para todas las componentes  $\mu_i$  del vector  $\underline{\mu}$ .

Sin embargo la confianza  $1 - \alpha$ -asociada a cada intervalo de confianza es individual y no conjunta, y lo que se quiere es asociar una confianza " colectiva " del  $1 - \alpha$  a todos los intervalos de confianza que pueden ser generados por las diferentes elecciones de  $\underline{a}$ .

¿ Cómo asociar un coeficiente de confianza " coletivo" del  $(1 - \alpha)$  para los I.C que pueden ser generados mediante todas las elecciones de  $\underline{a}$  ?

Para lograr lo anterior, se debe pagar un precio:

los intervalos simultáneos son más amplios (es decir, menos precisos) que el intervalo de confianza individual para una sola elección de  $\underline{a}$ .

**TEOREMA:** Sea  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$  un m.a de una población  $N_p(\underline{\mu}, \underline{\Sigma})$ , con  $\underline{\Sigma}$ -definida positiva, entonces, Simultáneamente para todo  $\underline{a}$ , el I.C dado por:

$$\underline{a}^t \underline{\bar{x}} \pm \underbrace{\sqrt{\frac{(n-1)p}{(n-p)} F_{\alpha;p,n-p}}}_{\text{}} \frac{\sqrt{\underline{a}^t \underline{S} \underline{a}}}{\sqrt{n}}$$

A diferencia de:

$$\underline{a}^t \underline{\bar{x}} \pm \underbrace{t_{\alpha/2;n-1}}_{\text{}} \frac{\sqrt{\underline{a}^t \underline{S} \underline{a}}}{\sqrt{n}}$$

contendrá a  $\underline{a}^t \underline{\mu}$  con probabilidad de  $(1 - \alpha)$ .

A estos I.C simultáneos, se les llaman **Intervalos- $T^2$** , ya que la probabilidad de cobertura de dichos intervalos se determina con la distribución  $T^2$ -de Hotelling.



## Justificación del teorema anterior:

Dado un conjunto de datos  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ , y una elección específica de  $\underline{a}$ , el intervalo de confianza para  $\underline{a}^t \underline{\mu}$  es aquel conjunto de valores que satisfacen:

$$|t| = \left| \frac{\sqrt{n} (\underline{a}^t \underline{\bar{x}} - \underline{a}^t \underline{\mu})}{\sqrt{\underline{a}^t \underline{S} \underline{a}}} \right| \leq t_{n-1}(\alpha/2)$$

o equivalentemente,

$$t^2 = \left( \frac{\sqrt{n} (\underline{a}^t \underline{\bar{x}} - \underline{a}^t \underline{\mu})}{\sqrt{\underline{a}^t \underline{S} \underline{a}}} \right)^2 = \frac{n (\underline{a}^t (\underline{\bar{x}} - \underline{\mu}))^2}{\underline{a}^t \underline{S} \underline{a}} \leq t_{n-1}^2(\alpha/2)$$

Cuando los intervalos son desarrollados para muchas elecciones de  $\underline{a}$ , parece razonable esperar que la constante para el intervalo individual,  $t_{n-1}^2(\alpha/2)$ , sea reemplazada por un valor mayor,  $c^2$ , que produzca una confianza simultánea deseada.

Sea

$$t^2(\underline{a}) = \frac{n \left( \underline{a}^t (\bar{\mathbf{x}} - \underline{\mu}) \right)^2}{\underline{a}^t \mathbf{S} \underline{a}}$$

entonces, se desea obtener un valor  $c^2$  tal que:

$$P \left[ t^2(\underline{a}) \leq c^2 \right] = 1 - \alpha, \quad \forall \underline{a} \in \mathbb{R}^p$$

Esto es equivalente a obtener un valor  $c^2$  tal que:

$$P \left[ \max_{\underline{a}} t^2(\underline{a}) \leq c^2 \right] = 1 - \alpha, \quad \forall \underline{a} \in \mathbb{R}^p$$

Para hallar dicho valor de  $c^2$ -se utiliza el siguiente resultado:

**Lema:** Sea  $\mathbf{B}$  una matriz definida positiva y  $\underline{d}$  un vector dado cualquiera, entonces para un vector  $\underline{x}_{p \times 1}$  se cumple que:

$$\max_{\underline{x}} \frac{(\underline{x}^t \underline{d})^2}{\underline{x}^t \mathbf{B} \underline{x}} = \underline{d}^t \mathbf{B}^{-1} \underline{d}$$

y el máximo se obtiene cuando:  $\underline{x} = k \mathbf{B}^{-1} \underline{d}$ , con  $k \in \mathbb{R}$  y  $k \neq 0$ .

Aplicando este teorema con:  $\underline{x} = \underline{a}$  ,  $\underline{d} = (\bar{x} - \underline{\mu})$  y  $\mathbf{B} = \mathbf{S}$ , se obtiene que:

$$\begin{aligned} \max_{\underline{a}} t^2(\underline{a}) &= \max_{\underline{a}} \frac{n \left( \underline{a}^t (\bar{x} - \underline{\mu}) \right)^2}{\underline{a}^t \mathbf{S} \underline{a}} \\ &= n \left[ \max_{\underline{a}} \frac{\left( \underline{a}^t (\bar{x} - \underline{\mu}) \right)^2}{\underline{a}^t \mathbf{S} \underline{a}} \right] \\ &= n (\bar{x} - \underline{\mu})^t \mathbf{S}^{-1} (\bar{x} - \underline{\mu}) = T^2 \end{aligned}$$

y el máximo ocurre en un  $\underline{a}$  que es proporcional a:  $\mathbf{S}^{-1}(\bar{\mathbf{x}} - \underline{\mu})$ .

y usando el resultado de que:

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

se tiene que antes de obtener la muestra,

$$P \left[ n(\bar{\mathbf{x}} - \underline{\boldsymbol{\mu}})^t \mathbf{S}^{-1} (\bar{\mathbf{x}} - \underline{\boldsymbol{\mu}}) \leq \frac{(n-1)p}{(n-p)} F_{\alpha; p, n-p} \right] = 1 - \alpha.$$

y de lo anterior se tiene que, si  $\underline{\mathbf{x}}_1, \underline{\mathbf{x}}_2, \dots, \underline{\mathbf{x}}_n$  es una muestra aleatoria de una población  $N_p(\underline{\boldsymbol{\mu}}, \mathbf{\Sigma})$ , entonces, **simultáneamente**  $\forall \underline{\mathbf{a}} \in \mathbb{R}^p$ , el intervalo:

$$\underline{\mathbf{a}}^t \bar{\mathbf{x}} \pm \underbrace{\sqrt{\frac{(n-1)p}{(n-p)} F_{\alpha; p, n-p}}}_{\text{}} \frac{\sqrt{\underline{\mathbf{a}}^t \mathbf{S} \underline{\mathbf{a}}}}{\sqrt{n}}$$

contendrá a  $\underline{\mathbf{a}}^t \underline{\boldsymbol{\mu}}$  con probabilidad  $1 - \alpha$ .

A diferencia de:

$$\underline{\mathbf{a}}^t \bar{\mathbf{x}} \pm \underbrace{t_{\alpha/2; n-1}}_{\text{}} \frac{\sqrt{\underline{\mathbf{a}}^t \mathbf{S} \underline{\mathbf{a}}}}{\sqrt{n}}$$

que lo contiene para un  $\underline{\mathbf{a}}$ -particular o individual.

**Ejemplo-1:** Con las elecciones sucesiva de  $\underline{a}$  dadas por:

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

para los intervalos- $T^2$ , se tiene que **todos los siguientes intervalos son simultáneamente verdaderos con un nivel de conf. del  $(1-\alpha)$ :**

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

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

$$\vdots$$

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

**Similarmente**, sin modificar el nivel de confianza de  $(1 - \alpha)$ , se pueden hacer afirmaciones acerca de la diferencia de medias  $\mu_i - \mu_k$ , correspondientes a:  $\underline{a}^t = (0, \dots, a_i, \dots, a_k, \dots, 0)$  con  $a_i = 1$  y  $a_k = -1$ .

En este caso se tiene que:  $\underline{a}^t \mathbf{S} \underline{a} = s_{ii} - 2s_{ik} + s_{kk}$ , y el I.C del  $(1 - \alpha)100\%$  para la diferencias de medias  $(\mu_i - \mu_k)$  es:

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

**Nota:** Los I.C  $T^2$ -simultáneos para las componentes individuales de un vector de medias  $\underline{\mu}$ , son las **sombras o proyecciones** de la **elipsoide de confianza** sobre los ejes componentes.

Los intervalos simultáneos  $T^2$  son útiles para examinar datos.

El coeficiente de confianza  $1 - \alpha$  no cambia para cualquier elección de  $\underline{a}$ . Por tanto, se pueden evaluar las combinaciones lineales de las componentes  $\mu_i$  que ameriten inspección basados sobre un examen de los datos.



## Ejemplo-2: Datos de radiación.

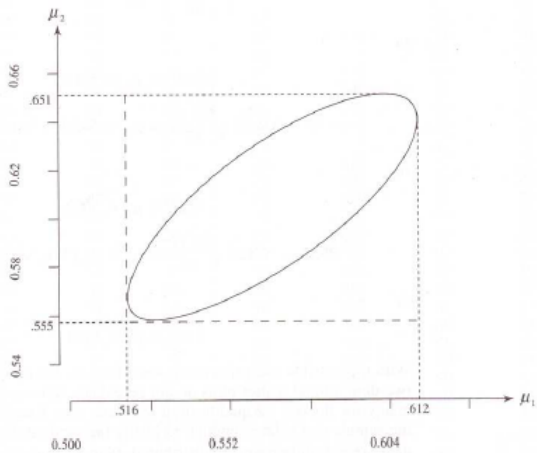
En el ejemplo anterior se construyó una elipse del 95% de confianza para las medias de la raíz cuarta de las variables de radiación a puerta cerrada y a puerta abierta de hornos microondas. Los intervalos simultáneos  $T^2$  del 95% de confianza para las medias componentes son.

$$\left( \bar{x}_1 - \sqrt{\frac{(n-1)p}{(n-p)}} F_{\alpha;p,n-p} \sqrt{\frac{s_{11}}{n}} , \bar{x}_1 + \sqrt{\frac{(n-1)p}{(n-p)}} F_{\alpha;p,n-p} \sqrt{\frac{s_{11}}{n}} \right)$$
$$\left( 0.564 - \sqrt{\frac{(41)^2}{40}} 3.23 \sqrt{\frac{0.0144}{42}} , 0.564 + \sqrt{\frac{(41)^2}{40}} 3.23 \sqrt{\frac{0.0144}{42}} \right)$$
$$= (0.516 , 0.612)$$

y para  $\underline{\mu}_2$

$$\left( \bar{x}_2 - \sqrt{\frac{(n-1)p}{(n-p)} F_{\alpha;p,n-p}} \sqrt{\frac{s_{22}}{n}} , \bar{x}_2 + \sqrt{\frac{(n-1)p}{(n-p)} F_{\alpha;p,n-p}} \sqrt{\frac{s_{22}}{n}} \right)$$
$$\left( 0.603 - \sqrt{\frac{(41)^2}{40}} 3.23 \sqrt{\frac{0.0144}{42}} , 0.603 + \sqrt{\frac{(41)^2}{40}} 3.23 \sqrt{\frac{0.0144}{42}} \right)$$
$$= (0.555 , 0.651)$$

El siguiente gráfico muestra la elipse del 95% de confianza del ejemplo anterior y los intervalos simultáneos del 95% de confianza, los cuales son sombras o proyecciones de la elipse sobre los ejes de las medias componentes.



**Ejemplo-3:** Construcción de intervalos simultáneos y elipses.

Las notas obtenidas por  $n = 87$  estudiantes en la prueba CLEP (College Level Examination Program),  $X_1$  = ciencias sociales e historia, y en la prueba CQT (College Qualification Test),  $X_2$  = aptitud verbal y  $X_3$  = ciencias, están dadas en la siguiente tabla.

De los datos se tiene que:

$$\bar{\mathbf{x}} = \begin{bmatrix} 527.74 \\ 54.69 \\ 25.13 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 5691.34 & 600.51 & 215.25 \\ 600.51 & 126.05 & 23.37 \\ 215.25 & 23.37 & 23.11 \end{bmatrix}$$

Individuo	$X_1$	$X_2$	$X_3$	Individuo	$X_1$	$X_2$	$X_3$
1	468	41	26	45	494	41	24
2	428	39	26	46	541	47	25
3	514	53	21	47	362	36	17
4	547	67	33	48	408	28	17
5	614	61	27	49	594	68	23
6	501	67	29	50	501	25	26
7	421	46	22	51	687	75	33
8	527	50	23	52	633	52	31
9	527	55	19	53	647	67	29
10	620	72	32	54	647	65	34
11	587	63	31	55	614	59	25
12	541	59	19	56	633	65	28
13	561	53	26	57	448	55	24
14	468	62	20	58	408	51	19
15	614	65	28	59	441	35	22
16	527	48	21	60	435	60	20
17	507	32	27	61	501	54	21
18	580	64	21	62	507	42	24
19	507	59	21	63	620	71	36
20	521	54	23	64	415	52	20
21	574	52	25	65	554	69	30
22	587	64	31	66	348	28	18
23	488	51	27	67	468	49	25
24	488	62	18	68	507	54	26
25	587	56	26	69	527	47	31
26	421	38	16	70	527	47	26
27	481	52	26	71	435	50	28
28	428	40	19	72	660	70	25
29	640	65	25	73	733	73	33
30	574	61	28	74	507	45	28
31	547	64	27	75	527	62	29
32	580	64	28	76	428	37	19
33	494	53	26	77	481	48	23
34	554	51	21	78	507	64	19
35	647	58	23	79	527	66	23
36	507	65	22	80	488	41	28
37	454	52	28	81	607	69	28
38	427	57	21	82	561	59	34
39	521	66	26	83	614	70	23
40	468	57	14	84	527	49	30
41	587	55	30	85	474	41	16
42	507	61	31	86	441	47	26
43	574	54	31	87	607	67	32
44	507	53	23				

Los intervalos del 95% de confianza para las medias componentes son:

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

$$\left( 527.74 - \sqrt{\frac{(86)3}{84} F_{3,84}(0.05)} \sqrt{\frac{5691.34}{87}} , 527.74 + \sqrt{\frac{(86)3}{84} F_{3,84}(0.05)} \sqrt{\frac{5691.34}{87}} \right)$$

$$\mu_1 \in (504.45 , 551.03)$$

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

$$\left( 54.69 - \sqrt{\frac{(86)3}{84} F_{3,84}(0.05)} \sqrt{\frac{126.05}{87}} , 54.69 + \sqrt{\frac{(86)3}{84} F_{3,84}(0.05)} \sqrt{\frac{126.05}{87}} \right)$$

$$\mu_2 \in (51.22 , 58.16)$$

$$\left( \bar{x}_3 - \sqrt{\frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{33}}{n}} , \bar{x}_3 + \sqrt{\frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{33}}{n}} \right)$$

$$\left( 25.13 - \sqrt{\frac{(86)3}{84} F_{3,84}(0.05)} \sqrt{\frac{23.11}{87}} , 25.13 + \sqrt{\frac{(86)3}{84} F_{3,84}(0.05)} \sqrt{\frac{23.11}{87}} \right)$$

$$\mu_3 \in (23.65 , 26.61)$$

Con la posible excepción de las notas sobre aptitud verbal, los gráficos  $Q - Q$  y los diagramas de dispersión para cada par de variables no rebelan serios alejamientos de la normalidad. Además, como lo veremos, aunque los datos no sean normalmente distribuidos, la metodología se puede justificar dado que la muestra es lo suficientemente grande.

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 95%.

Los intervalos obtenidos anteriormente pueden ser más amplios de lo necesario, debido a que, con la misma confianza, podemos hacer inferencias para cualquier combinación lineal de las componentes de medias, tales como las diferencias de medias.

Por ejemplo, con  $\underline{a} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$  el intervalo para  $\mu_2 - \mu_3$ , tiene los siguientes límites:

$$(\bar{x}_2 - \bar{x}_3) \pm \sqrt{\frac{(n-1)p}{(n-p)} F_{p, n-p}(\alpha)} \sqrt{\frac{s_{22} + s_{33} - 2s_{23}}{n}}$$
$$(54.69 - 25.13) \pm \sqrt{8.29} \sqrt{\frac{126.05 + 23.11 - 2(23.37)}{87}} = 29.56 \pm 3.12$$

Por tanto, (26.44, 32.68) es un intervalo del 95% de confianza para  $\mu_2 - \mu_3$ . Similarmente, se pueden construir intervalos simultáneos para las otras diferencias de medias.



Finalmente, se pueden construir elipses del 95% de confianza para cada par de medias y los mismos intervalos simultáneos 95% de confianza se mantienen. Por ejemplo, **para el par de medias  $(\mu_2, \mu_3)$** , se tiene que:

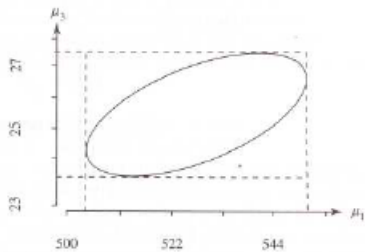
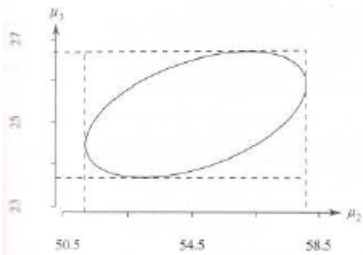
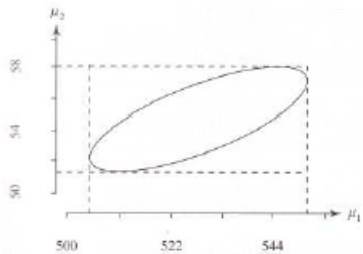
$$n \begin{bmatrix} \bar{x}_2 - \mu_2 \\ \bar{x}_3 - \mu_3 \end{bmatrix}^t \begin{bmatrix} s_{22} & s_{23} \\ s_{23} & s_{33} \end{bmatrix}^{-1} \begin{bmatrix} \bar{x}_2 - \mu_2 \\ \bar{x}_3 - \mu_3 \end{bmatrix} \leq \frac{(n-1)p}{(n-p)} F_{p, n-p}(\alpha)$$

$$87 \begin{bmatrix} 54.69 - \mu_2 \\ 25.13 - \mu_3 \end{bmatrix}^t \begin{bmatrix} 126.05 & 23.37 \\ 23.37 & 23.11 \end{bmatrix}^{-1} \begin{bmatrix} 54.69 - \mu_2 \\ 25.13 - \mu_3 \end{bmatrix} \leq \frac{(86)3}{84} F_{2,84}(0.05)$$

o equivalentemente:

$$0.849(54.69 - \mu_2)^2 + 4.633(25.13 - \mu_3)^2 - 2(54.69 - \mu_2)(25.13 - \mu_3) \leq 8.29$$

La siguiente gráfica presenta la elipse anterior del 95% de confianza y las elipses para los otros 2-pares de medias, junto con los intervalos generados por las sombras o proyecciones de las elipses sobre los ejes de medias, los cuales son los respectivos intervalos simultáneos  $T^2$ .



# Comparación de los intervalos de confianza simultáneos con los intervalos univariados o uno a la vez

Una aproximación alternativa para construir intervalos de confianza es considerar las componentes de  $\underline{\mu}$  una a una con  $\underline{a}^t = (0, \dots, a_i, \dots, 0)$  con  $a_i = 1$ .

Esta aproximación ignora la estructura de covarianza de las  $p$  variables y conduce a los intervalos dados por:

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

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

$\vdots$

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

Aunque, antes de muestrear, el  $i$ -ésimo intervalo tiene  $1 - \alpha$  de probabilidad de contener a  $\mu_i$ . Sin embargo, en general no se puede asegurar cuál es la probabilidad de que todos los intervalos contengan a sus respectivos  $\mu_i$ . Como se ha señalado, esta probabilidad no es  $1 - \alpha$ .

**Ejemplo-4:** Considere el caso especial donde las observaciones proceden de una distribución normal conjunta, con

$$\Sigma = \begin{bmatrix} \sigma_{11} & 0 & \cdots & 0 \\ & \sigma_{22} & \cdots & 0 \\ & & \ddots & \\ & & & \sigma_{pp} \end{bmatrix}$$

es decir, las variables son independientes.

Sea  $I_i$  el intervalo aleatorio para  $\mu_i$  con probabilidad  $(1 - \alpha)$ .

Debido a la independencia de las variables, se tiene que antes de que la muestra sea seleccionada se cumple que:

$$\begin{aligned} &P\left[\text{Todos los t-intervalos contengan las } \mu_i\right] \\ &= P\left[\mu_1 \in I_1, \mu_2 \in I_2, \dots, \mu_p \in I_p\right] \\ &= \prod_{i=1}^p P\left[\mu_i \in I_i\right] \\ &= (1 - \alpha)(1 - \alpha) \cdots (1 - \alpha) \\ &= (1 - \alpha)^p \end{aligned}$$

de donde por ejemplo, si  $(1 - \alpha) = 0.95$  y  $p = 6$ , entonces:

$$P[\text{Todos los t-intervalos contengan las } \mu_i] = (0.95)^6 = 0.74.$$

Para garantizar una probabilidad de  $1 - \alpha$  de que todos los intervalos contengan simultáneamente la componente de media respectiva, los intervalos individuales deberían ser más amplios que los intervalos separados basados en la  $t$ , y **el qué tan ancho**, depende tanto de  $p$  como de  $n$  al igual que de  $(1 - \alpha)$ .

**Por ejem:** Para  $(1 - \alpha) = 0.95$ ,  $n = 15$  y  $p = 4$ , el **factor multiplicador** de  $\sqrt{\frac{s_{ii}}{n}}$ , para los interv. separados (o individuales) es:

$$t_{\alpha/2;n-1} = t_{0.025,14} = 2.145$$

y para los intervalos simultáneos  $T^2$  es:

$$\sqrt{\frac{(n-1)p}{n-p} F_{\alpha;p,n-p}} = \sqrt{\frac{(14)4}{11} F_{0.05;4,11}} = \sqrt{\frac{56}{11} (3.36)} = 4.14$$

de donde en este caso, se tiene que los I.C simultáneos  $T^2$  son:

$$\frac{(4.14 - 2.145)}{2.145} 100\% = 93\%$$

**más anchos**, que aquellos intervalos derivados usando la  $t$  (o individuales).

En general, el ancho de los intervalos simultáneos  $T^2$ , relativo al ancho de los  $t$ -intervalos individuales, **crece para  $n$ -fijo cuando  $p$ -crece**; y **decrece para  $p$ -fijo cuando  $n$ -crece**. (Ver la siguiente tabla del libro de Richard A. Johnson, Sixth Edition, Pág. 231).

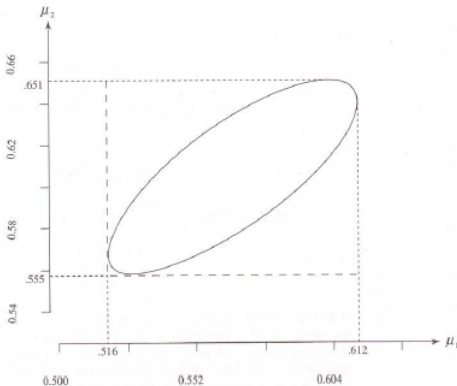
$n$	$t_{n-1}(.025)$	$\sqrt{\frac{(n-1)p}{(n-p)} F_{p, n-p}(.05)}$	
		$p = 4$	$p = 10$
15	2.145	4.14	11.52
25	2.064	3.60	6.39
50	2.010	3.31	5.05
100	1.970	3.19	4.61
$\infty$	1.960	3.08	4.28



La comparación de la tabla anterior es un poco injusta debido a que para  $n$  fijo, el nivel de confianza asociada a un conjunto de intervalos  $T^2$  es  $1 - \alpha$ , mientras que la confianza asociada a un conjunto de intervalos  $t$ , para el mismo  $n$ , puede ser menor que  $1 - \alpha$ , como vimos anteriormente.

Los intervalos  $t$  son demasiado cortos para mantener un nivel de confianza global para los intervalos separados sobre, por ejemplo, las  $p$  medias. Sin embargo, algunas veces se les considera la mejor información posible sobre una media, si es la única inferencia que nos interesa. Además, 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  $T^2$ .

Los intervalos  $T^2$  son demasiado amplios si se aplican solamente a las  $p$  medias componentes. Para ver porqué, considere la elipse de confianza y los intervalos simultáneos del ejemplo anterior:



Si  $\mu_1$  cae en su intervalo  $T^2$  y  $\mu_2$  cae en su intervalo  $T^2$ , entonces  $(\mu_1, \mu_2)$  cae en el rectángulo formado por estos dos intervalos.

Este rectángulo contiene la elipse de confianza y más. La elipse de confianza es más pequeña, pero tiene una probabilidad  $1 - \alpha$  de contener al vector  $\underline{\mu} = [\mu_1 \ \mu_2]^t$ .

En consecuencia, la probabilidad de contener la dos medias individuales  $\mu_1$  y  $\mu_2$  será mayor que  $1 - \alpha$  para el rectángulo formado por los intervalos  $T^2$ .

Este resultado conduce a considerar una segunda aproximación para hacer comparaciones simultáneas.

# Método de Bonferroni para Comparaciones Múltiples

Frecuentemente la atención está restringida a un número pequeño de intervalos de confianza.

En estos casos la metodología de los intervalos  $T^2$  puede conducir a intervalos demasiado amplios.

Si un número pequeño  $m$  de componentes componentes de medias específicas  $\mu_i$  o de combinaciones lineales de la forma  $\underline{a}^t \underline{\mu} = a_1\mu_1 + a_2\mu_2 + \cdots + a_p\mu_p$  son de interés, se pueden construir I.C simultáneos mas cortos (más precisos) que los intervalos  $T^2$ .

El método alternativo para múltiples comparaciones, se llama: Método de Bonferroni, debido a que ha sido desarrollado a partir de la Desigualdad de Bonferroni.

Suponga que antes de recolectar los datos, se quieren construir intervalos de confianza simultáneos para  $m$ -combinaciones lineales de la forma:

$$\underline{a_1^t \underline{\mu}}, \quad \underline{a_2^t \underline{\mu}}, \quad \cdots, \quad \underline{a_m^t \underline{\mu}}$$

Sea  $C_i$  el intervalo para  $\underline{a_i^t \underline{\mu}}$ , con:

$$P\left[C_i \text{ contenga a: } \underline{a_i^t \underline{\mu}}\right] = 1 - \alpha_i, \quad \text{para } i = 1, 2, \dots, m$$

entonces,

$$\begin{aligned} & P[\underline{a_1^t \mu} \in C_1, \underline{a_2^t \mu} \in C_2, \dots, \underline{a_m^t \mu} \in C_m] \\ &= P\left[\bigcap_{i=1}^m \underline{a_i^t \mu} \in C_i\right] = 1 - P\left[\left(\bigcap_{i=1}^m \underline{a_i^t \mu} \in C_i\right)^c\right] \\ &= 1 - P\left[\bigcup_{i=1}^m \underline{a_i^t \mu} \notin C_i\right] \geq 1 - \sum_{i=1}^m P[\underline{a_i^t \mu} \notin C_i] \\ &= 1 - \sum_{i=1}^m \left(1 - P[\underline{a_i^t \mu} \in C_i]\right) \\ &= 1 - (1 - (1 - \alpha_1) + 1 - (1 - \alpha_2) + \dots + 1 - (1 - \alpha_m)) \\ &= 1 - (\alpha_1 + \alpha_2 + \dots + \alpha_m) = 1 - \sum_{i=1}^m \alpha_i \end{aligned}$$

es decir,

$$P\left[\underline{a_1^t\mu} \in C_1, \underline{a_2^t\mu} \in C_2, \dots, \underline{a_m^t\mu} \in C_m\right] \geq 1 - (\alpha_1 + \alpha_2 + \dots + \alpha_m)$$

La expresión anterior es un caso especial de la desigualdad de Bonferroni y permite al investigador controlar el error global  $\alpha_1 + \alpha_2 + \dots + \alpha_m$ , sin tener en cuenta la estructura de correlación del sistema de variables aleatorias.

También tiene la flexibilidad de controlar la tasa de error global de un grupo de intervalos importantes y balancearla con otra elección de los intervalos menos importantes.

## Intervalos simultáneos para un grupo restringido de componentes $\mu_i$ .

Considere los intervalos  $t$  individuales:

$$\bar{x}_i \pm t_{n-1}(\alpha_i/2)\sqrt{\frac{s_{ii}}{n}}, \quad \text{para } i = 1, 2, \dots, m$$

con  $\alpha_i = \alpha/m$ .

Ahora, puesto que,

$$P\left[\mu_i \in \left(\bar{x}_i - t_{n-1}(\alpha_i/2)\sqrt{\frac{s_{ii}}{n}}, \bar{x}_i + t_{n-1}(\alpha_i/2)\sqrt{\frac{s_{ii}}{n}}\right)\right] = 1 - \alpha_i = 1 - \alpha/m$$

para  $i = 1, 2, \dots, m$ ,



entonces por la anterior desigualdad se tiene que:

$$P\left[\mu_i \in \left(\bar{\mathbf{x}}_i - t_{n-1}(\alpha_i/2)\sqrt{\frac{S_{ii}}{n}}, \bar{\mathbf{x}}_i + t_{n-1}(\alpha_i/2)\sqrt{\frac{S_{ii}}{n}}\right) \text{ para todo } i\right]$$

$$\geq 1 - \sum_{i=1}^m \alpha_i$$

$$= 1 - \left(\frac{\alpha}{m} + \frac{\alpha}{m} + \cdots + \frac{\alpha}{m}\right)$$

$$= 1 - \alpha$$

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

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

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

$\vdots$

$\vdots$

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

**Ejemplo-6:** Construcción de intervalos de Bonferroni y comparación con los intervalos  $T^2$ .

Considere de nuevo los datos de radiación de hornos microondas. Los intervalos simultáneos de Bonferroni del 95% de confianza para las medias  $\mu_1$  y  $\mu_2$  de las raíces cuartas de la radiación a puerta cerrada y a puerta abierta de hornos microondas son:

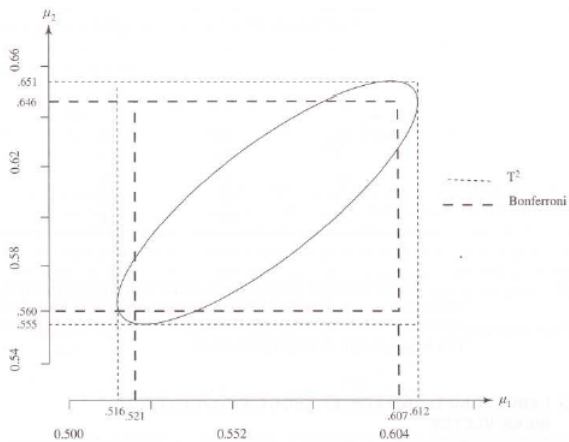
$$\bar{x}_1 \pm t_{n-1} \left( \frac{\alpha}{2p} \right) \sqrt{\frac{s_{11}}{n}} = 0.564 \pm t_{41} \left( \frac{0.05}{2(2)} \right) \sqrt{\frac{0.0144}{42}}, \quad \text{ie. } 0.521 \leq \mu_1 \leq 0.607$$

$$\bar{x}_2 \pm t_{n-1} \left( \frac{\alpha}{2p} \right) \sqrt{\frac{s_{33}}{n}} = 0.603 \pm t_{41} \left( \frac{0.05}{2(2)} \right) \sqrt{\frac{0.0146}{42}}, \quad \text{ie. } 0.560 \leq \mu_1 \leq 0.646$$

El siguiente gráfico muestra los intervalos  $T^2$  del 95% de confianza para  $\mu_1$  y  $\mu_2$  junto con los intervalos simultáneos de Bonferroni del 95% de confianza.

Para cada componente, el intervalo de Bonferroni cae dentro del correspondiente intervalo  $T^2$ . Por tanto, la región rectangular formada por los intervalos de Bonferroni cae dentro la región rectangular formada por los intervalos  $T^2$ .

Por otro lado, la región (o elipse) de confianza del 95% de para el vector  $\underline{\mu}$  proporciona todos los valores plausibles para los pares  $(\mu_1, \mu_2)$  cuando se tiene en cuenta la correlación entre las variables medidas.



Los intervalos de Bonferroni para combinaciones lineales  $\underline{a}^t \underline{\mu}$  y los análogos intervalos  $T^2$ , tienen la forma general,

$$\underline{a}^t \underline{\bar{x}} \pm (\text{Valor Crítico}) \sqrt{\frac{\underline{a}^t \mathbf{S} \underline{a}}{n}}$$

Por tanto, para cada grupo de  $m$  intervalos simultáneos de Bonferroni,

$$\frac{\text{Longitud del Intervalo de Bonferroni}}{\text{Longitud del Intervalo } T^2} = \frac{t_{n-1}(\alpha/2m)}{\sqrt{\frac{(n-1)p}{n-p}} F_{p, n-p}(\alpha)}$$

lo cual que no depende de las cantidades aleatorias  $\underline{\bar{x}}$  y  $\mathbf{S}$ . Para un número pequeño  $m$  de funciones específicas  $\underline{a}^t \underline{\mu}$ , los intervalos de Bonferroni siempre serán más cortos.

La siguiente tabla muestra cuánto más cortos son para valores seleccionados de  $n$  y  $p$ .

Longitud Int. Bonferroni/Longitud Int.  $T^2$ ,  $1-\alpha=.95$

$n$	$m = p$		
	2	4	10
15	.88	.69	.29
25	.90	.75	.48
50	.91	.78	.58
100	.91	.80	.62
$\infty$	.91	.81	.66

Se observa que el método de Bonferroni proporciona los intervalos más cortos cuando  $m = p$ . Debido a que los intervalos simultáneos de Bonferroni son más fáciles de aplicar y son relativamente más cortos para hacer inferencia, en la práctica, generalmente son preferidos.

## Observación final:

Finalmente, la independencia de las observaciones multivariadas es crucial, y los resultados vistos basados sobre este supuesto pueden estar muy equivocados si las observaciones son dependientes.

**Lectura:** Inferencias para muestras grandes sobre proporciones, Johnson y Wichern (2007), Cap. 5, pág. 234.

**Ver librería ICSNP** para la prueba usando el estadístico  $T^2$  de Hotelling. Estudiar el siguiente ejemplo en esa librería.

```
Library(ICSNP)
```

```
data(pulmonary)
```

```
HotellingsT2(pulmonary)
```

```
HotellingsT2(pulmonary, mu = c(0,0,2), test = "chi")
```

**Otras librerías:** Hotelling, rrcov.