

Estadística Aplicada III

Pruebas de hipótesis

Jorge de la Vega Góngora

Departamento de Estadística,
Instituto Tecnológico Autónomo de México

Semana 5: 12/14 de septiembre de 2018

Estimación

Resumen sobre estimación I

- Hemos visto que los estimadores máximo verosímiles para μ y Σ de una muestra $\mathbf{x}_1, \dots, \mathbf{x}_n \sim \mathcal{N}_p(\mu, \Sigma)$ están dados por:
 - ▶ $\hat{\mu} = \bar{\mathbf{x}} \sim \mathcal{N}_p\left(\mu, \frac{1}{n}\Sigma\right)$
 - ▶ $\hat{\Sigma} = \mathbf{S}_n = \left(\frac{n-1}{n}\right)\mathbf{S}$ y $n\hat{\Sigma} = (n-1)\mathbf{S} \sim W_{n-1}((n-1)\mathbf{S}|\Sigma)^1$, donde $\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$.
 - ▶ $\bar{\mathbf{x}}$ y \mathbf{S} son independientes.
- Las siguientes propiedades de la distribución Wishart serán utilizadas en algunos casos más adelante:
 - ① La distribución Wishart con q grados de libertad es la correspondiente a $\sum_{i=1}^q \mathbf{Z}_i \mathbf{Z}_i'$ donde las \mathbf{Z}_i son independientes $\mathcal{N}_p(\mathbf{0}, \Sigma)$. (equivalente a χ^2 multivariada).
 - ② Si $\mathbf{S}_1 \sim W_n(\mathbf{S}_1|\Sigma)$, $\mathbf{S}_2 \sim W_m(\mathbf{S}_2|\Sigma)$ y $\mathbf{S}_1 \perp\!\!\!\perp \mathbf{S}_2$, entonces $\mathbf{S} = \mathbf{S}_1 + \mathbf{S}_2 \sim W_{n+m}(\mathbf{S}|\Sigma)$.
 - ③ Si $\mathbf{S} \sim W_n(\mathbf{S}|\Sigma)$ y $\mathbf{C}_{l \times p} \neq \mathbf{0}$ entonces $\mathbf{B} = \mathbf{C} \mathbf{S} \mathbf{C}' \sim W_l(\mathbf{B}|\mathbf{C} \Sigma \mathbf{C}')$
- Adicionalmente tenemos los siguientes resultados sobre formas cuadráticas (noten que hay resultados con valores poblacionales y otros con valores muestrales):
 - ▶ Si $\mathbf{x} \sim \mathcal{N}_p(\mu, \Sigma)$ entonces $(\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu) \sim \chi^2_{(p)}$

Resumen sobre estimación II

- ▶ Si μ_0 es un vector fijo, la estadística

$T^2 = n(\bar{\mathbf{x}} - \mu_0)' \mathbf{S}^{-1} (\bar{\mathbf{x}} - \mu_0) \sim \frac{(n-1)p}{n-p} F_{p, n-p}$. Esta estadística es llamada *T² de Hotelling*. Esta estadística combina una normal multivariada con una matriz Wishart, y es análoga a la distribución *t* univariada elevada al cuadrado, que coincide con una distribución *F*.

- Para la estadística T^2 de Hotelling, una de sus principales características es que es invariante bajo transformaciones afines $\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{d}$, donde \mathbf{C} no sea singular. Esto es fácil de demostrar considerando las propiedades de la media y varianza de \mathbf{y} a partir de la de \mathbf{x} .
- Los resultados de la Ley de los grandes números se pueden aplicar a estos estimadores: cuando $n \rightarrow \infty$:
 - ▶ $\bar{\mathbf{x}} \xrightarrow{p} \mu$
 - ▶ $\mathbf{S} \xrightarrow{p} \Sigma$
- Finalmente, por el Teorema del límite central,
 - ▶ $\sqrt{n}(\bar{\mathbf{x}} - \mu) \stackrel{d}{\sim} \mathcal{N}_p(\mathbf{0}, \Sigma)$ donde $n >> p$.
 - ▶ $(\bar{\mathbf{x}} - \mu)' \mathbf{S}^{-1} (\bar{\mathbf{x}} - \mu) \stackrel{d}{\sim} \chi^2_{(p)}$ para $n - p$ grande.

¹ $W_{n-1}(\cdot | \Sigma)$ es la distribución Wishart con $n - 1$ grados de libertad. La densidad Wishart con m grados de libertad para una matriz $\mathbf{A}_{k \times k}$ definida positiva está dada por:

$$w_m(\mathbf{A} | \Sigma) = \frac{|\mathbf{A}|^{\frac{m-p-1}{2}} e^{-\text{tr}[\mathbf{A}\Sigma^{-1}]/2}}{2^{pm/2} \pi^{p(p-1)/4} |\Sigma|^{m/2} \prod_{i=1}^p \Gamma\left(\frac{1}{2}(m+1-i)\right)}$$

Ejemplos I

- ① Con la matriz de datos $\mathbf{X} = \begin{bmatrix} 2 & 12 \\ 8 & 9 \\ 6 & 9 \\ 8 & 10 \end{bmatrix}$, calculen T^2 para la prueba $H_0 : \mu' = (7, 11)$, suponiendo que los datos son muestra de una distribución binormal.
- ② ¿Cuál es la distribución de T^2 ?
- ③ ¿Cómo se distribuye $3S$?

Solución.

```
X <- matrix(c(2,8,6,8,12,9,9,6), nrow=4)
n <- dim(X)[1]
mu <- c(7,11)
S <- var(X)
S

[,1] [,2]
[1,]     8    -6
[2,]    -6     6

xbar <- colMeans(X)
xbar

[1] 6 9

T2 <- as.numeric(n * t(xbar-mu) %*% solve(S) %*% (xbar-mu))
T2

[1] 20.66667
```

Ejemplos II

La distribución de $T^2 \sim \frac{2(4-1)}{4-2} F_{2,4-2} = 3F_{2,2}$. La distribución de la matriz $3\mathbf{S}$ es Wishart con 3 grados de libertad.



- ④ Verifiquen que pasa con T^2 si cada caso de la matriz \mathbf{X} se sustituye por $\mathbf{Cx}_i + \mathbf{d}$, donde $\mathbf{C} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ y $\mathbf{d} = c(1, 1)'$.

Solución.

```
C <- matrix(c(1,1,-1,1), nrow=2)
Y <- X %*% t(C) + c(1,1)
ybar <- colMeans(Y)
S2 <- var(Y)
muy <- C%*%mu + c(1,1)
T3 <- as.numeric(n*t(ybar-muy) %*% solve(S2) %*% (ybar-muy)))
T3

[1] 20.66667
```

Vimos que la distribución no cambia bajo transformaciones afines, por las propiedades de las medias y las varianzas.



Requerimientos de optimización matricial

Un recordatorio rápido I

- Durante el curso en diversas partes utilizaremos los siguientes resultados de optimización de cocientes de formas cuadráticas, que parten en todos los casos de la desigualdad de Cauchy-Schwarz. Revisaremos los resultados para tenerlos presentes.

Teorema (Desigualdad de Cauchy-Schwarz)

Sean $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$. Entonces

$$(\mathbf{x}'\mathbf{y})^2 \leq (\mathbf{x}'\mathbf{x})(\mathbf{y}'\mathbf{y})$$

y la igualdad se cumple si y sólo si $\mathbf{x} = c\mathbf{y}$ para alguna $c \in \mathbb{R}$

- El primer paso es extender esta desigualdad para incorporar una matriz definida positiva.

Teorema (Desigualdad de Cauchy-Schwarz extendida)

Sean $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$ y $\mathbf{B}_{p \times p} > 0$ (definida positiva). Entonces:

$$(\mathbf{x}'\mathbf{y})^2 \leq (\mathbf{x}'\mathbf{B}\mathbf{x})(\mathbf{y}'\mathbf{B}^{-1}\mathbf{y})$$

y la igualdad se da si y sólo si $\mathbf{x} = c\mathbf{B}^{-1}\mathbf{y}$ para alguna $c \in \mathbb{R}$.

Un recordatorio rápido II

Demostración.

Básicamente tenemos que escribir $\mathbf{B} = \mathbf{B}^{1/2}\mathbf{B}^{1/2}$ usando la descomposición espectral. Entonces

$$\mathbf{x}'\mathbf{y} = \mathbf{x}'\mathbf{I}\mathbf{y} = \mathbf{x}'\mathbf{B}^{1/2}\mathbf{B}^{-1/2}\mathbf{y} = (\mathbf{B}^{1/2}\mathbf{x})'(\mathbf{B}^{-1/2}\mathbf{y})$$

y el resultado sigue por la desigualdad de Cauchy-Schwarz.

□

- El siguiente lema es el primer resultado sobre optimización del cociente de formas cuadráticas.

Lema

Sea $\mathbf{B}_{p \times p} > 0$ y $\mathbf{a} \in \mathbb{R}^p$. Entonces para todo vector $\mathbf{x} \neq \mathbf{0}$ en \mathbb{R}^p

$$\max_{\mathbf{x} \neq 0} \frac{(\mathbf{x}'\mathbf{a})^2}{\mathbf{x}'\mathbf{B}\mathbf{x}} = \mathbf{a}'\mathbf{B}^{-1}\mathbf{a}$$

donde el máximo se alcanza cuando $\mathbf{x} = c\mathbf{B}^{-1}\mathbf{a}$ para alguna $c \neq 0$.

Un recordatorio rápido III

Demostración.

Por el teorema extendido de Cauchy-Schwarz $(\mathbf{x}'\mathbf{a})^2 \leq (\mathbf{x}'\mathbf{B}\mathbf{x})(\mathbf{a}'\mathbf{B}^{-1}\mathbf{a})$. Además como $\mathbf{x} \neq \mathbf{0}$ y $\mathbf{B} > 0$, $\mathbf{x}'\mathbf{B}\mathbf{x} > 0$. Entonces:

$$\frac{(\mathbf{x}'\mathbf{a})^2}{\mathbf{x}'\mathbf{B}\mathbf{x}} \leq \mathbf{a}'\mathbf{B}^{-1}\mathbf{a}$$

Como la igualdad se alcanza con $\mathbf{x} = c\mathbf{B}^{-1}\mathbf{a}$, está claro que:

$$\max_{\mathbf{x} \neq \mathbf{0}} \frac{(\mathbf{x}'\mathbf{a})^2}{\mathbf{x}'\mathbf{B}\mathbf{x}} = \mathbf{a}'\mathbf{B}^{-1}\mathbf{a}$$

□

Un recordatorio rápido IV

- Por ultimo, el siguiente teorema relaciona los óptimos con los eigenvalores.

Teorema (Maximización de formas cuadráticas para puntos en la esfera unitaria)

Sea $\mathbf{B}_{p \times p} > \mathbf{0}$ con eigenvalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ y correspondientes eigenvectores normalizados $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p\}$. Entonces:

$$\max_{\mathbf{x} \neq 0} \frac{\mathbf{x}' \mathbf{B} \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \lambda_1 \quad \text{que se alcanza cuando } \mathbf{x} = \mathbf{e}_1$$

$$\min_{\mathbf{x} \neq 0} \frac{\mathbf{x}' \mathbf{B} \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \lambda_p \quad \text{que se alcanza cuando } \mathbf{x} = \mathbf{e}_p$$

$$\max_{\mathbf{x} \perp \mathbf{e}_1, \dots, \mathbf{e}_k} \frac{\mathbf{x}' \mathbf{B} \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \lambda_{k+1} \quad \text{que se alcanza cuando } \mathbf{x} = \mathbf{e}_{k+1}, \quad k = 1, 2, \dots, p-1$$

Demostración.

Un recordatorio rápido V

Supongamos que $\mathbf{B} = \mathbf{P}\Lambda\mathbf{P}'$ es la descomposición espectral de tal manera que $\mathbf{B}^{1/2} = \mathbf{P}\Lambda^{1/2}\mathbf{P}'$, donde la matriz \mathbf{P} cumple con la condición $\mathbf{P}'\mathbf{P} = \mathbf{P}\mathbf{P}' = \mathbf{I}$. Sea $\mathbf{y} = \mathbf{P}'\mathbf{x}$. Entonces, si $\mathbf{x} \neq 0, \mathbf{y} \neq 0$ y

$$\begin{aligned}\frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} &= \frac{\mathbf{x}'\mathbf{B}^{1/2}\mathbf{B}^{1/2}\mathbf{x}}{\mathbf{x}'\mathbf{P}\mathbf{P}'\mathbf{x}} &=& \frac{\mathbf{x}'\mathbf{P}\Lambda^{1/2}\mathbf{P}'\mathbf{P}\Lambda^{1/2}\mathbf{P}'\mathbf{x}}{\mathbf{y}'\mathbf{y}} \\&=& \frac{\mathbf{y}'\Lambda\mathbf{y}}{\mathbf{y}'\mathbf{y}} \\&=& \frac{\sum_{i=1}^p \lambda_i y_i^2}{\sum_{i=1}^p y_i^2} \\&\leq& \lambda_1 \frac{\sum_{i=1}^p y_i^2}{\sum_{i=1}^p y_i^2} = \lambda_1\end{aligned}$$

Si se toma $\mathbf{x} = \mathbf{e}_1$ entonces $\mathbf{y} = \mathbf{P}'\mathbf{e}_1 = \mathbf{e}_1$, entonces $\frac{\mathbf{e}_1'\mathbf{B}\mathbf{e}_1}{\mathbf{e}_1'\mathbf{e}_1} = \mathbf{e}_1'\mathbf{B}\mathbf{e}_1 = \lambda_1$.

Por otra parte, $\mathbf{y} = \mathbf{P}'\mathbf{x} \implies \mathbf{P}\mathbf{y} = \mathbf{x} = \sum_{i=1}^p y_i \mathbf{e}_i$. Si $\mathbf{x} \perp \mathbf{e}_1, \dots, \mathbf{e}_k$ entonces

$$0 = \mathbf{e}_i'\mathbf{x} = y_1 \mathbf{e}_1'\mathbf{e}_1 + y_2 \mathbf{e}_2'\mathbf{e}_1 + \cdots + y_p \mathbf{e}_p'\mathbf{e}_1 = y_i, \quad i \leq k$$

Un recordatorio rápido VI

Por lo tanto el cociente se puede escribir como:

$$\frac{\mathbf{x}' \mathbf{B} \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \frac{\sum_{i=k+1}^p \lambda_i y_i^2}{\sum_{i=k+1}^p y_i^2}$$

Si tomamos $y_{k+1} = 1$ y el resto de las $y_{k+i} = 0$ para $i = 2, \dots, p$, se obtiene el máximo establecido.

□

Ejemplo

Sea $\mathbf{B} = \begin{pmatrix} 2.2 & 0.4 \\ 0.4 & 2.8 \end{pmatrix}$. consideremos como función de \mathbf{x} la forma cuadrática

$$g(\mathbf{x}) = \frac{\mathbf{x}' \mathbf{B} \mathbf{x}}{\mathbf{x}' \mathbf{x}}$$

Los eigenvalores de \mathbf{B} son $\lambda_1 = 3$ y $\lambda_2 = 2$ y la matriz

$\mathbf{P} = \begin{pmatrix} 0.4472136 & -0.8944272 \\ 0.8944272 & 0.4472136 \end{pmatrix}$ es la matriz de eigenvectores. Las direcciones hacia

donde se maximiza y minimiza respectivamente la función $g(\mathbf{x})$ son precisamente las columnas de \mathbf{P} , que se muestran en la gráfica con las flechas.

Adicionalmente se generan una muestra de puntos para calcular el valor de la función g y ver que precisamente en la dirección de esos dos vectores es donde se tiene el valor más alto y el más bajo respectivamente.

Un recordatorio rápido VII

```
options(scipen=5) #control de expansión en notación científica
set.seed(100)
B <- matrix(c(2.2,.4,.4,2.8),nrow=2) #matriz ejemplo

#Define la función cociente forma cuadrática
g <- function(x,Q){ as.numeric((t(x) %*% Q %*% x)/sum(x*x)) }

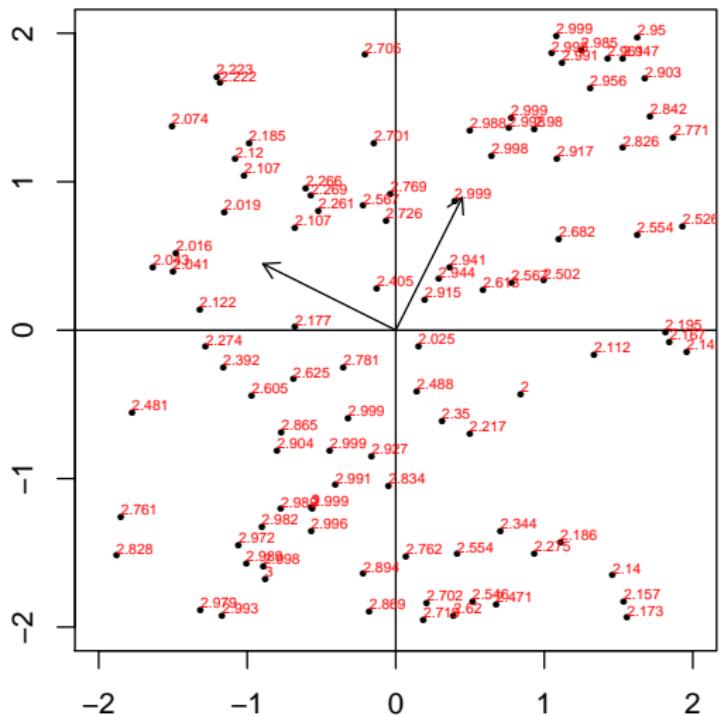
#genera algunas direcciones al azr en el cuadro [-2,2]^2
vectores <- cbind(runif(100,-2,2),runif(100,-2,2))
#calcula la función cuadrática en esas direcciones
z <- apply(vectores, 1, g, Q = B)
z

[1] 2.865464 2.605354 2.702268 2.481084 2.405223 2.725723 2.985216 2.261115 2.719030 2.121670 2.988370 2.82576
[14] 2.991407 2.997557 2.471272 2.221621 2.268857 2.999396 2.997854 2.488158 2.000051 2.024597 2.502406 2.99894
[27] 2.998698 2.947164 2.914689 2.997919 2.834432 2.842367 2.265531 2.195287 2.566529 2.173319 2.273766 2.54612
[40] 2.015677 2.106951 2.139769 2.185648 2.955779 2.554457 2.768516 2.991348 2.157182 2.993137 2.988717 2.17655
[53] 2.972192 2.982185 2.941364 2.184734 2.074136 2.120140 2.620404 2.018893 2.701188 2.612524 2.166557 2.34415
[66] 2.995829 2.868977 2.566814 2.107218 2.998822 2.781370 2.625233 2.943850 2.770820 2.998349 2.217282 2.96128
[79] 2.111565 2.043335 2.926591 2.998736 2.902600 2.525851 2.761411 2.350119 2.275301 2.988704 2.904123 2.98037
[92] 2.392441 2.999654 2.704568 2.554010 2.998603 2.762132 2.040846 2.828235 2.917087

#Ahora calcula los valores propios, grafica las direcciones de los vectores propios y muestra las direcciones si

lambdas <- eigen(B)$values
P <- eigen(B)$vectors
par(pty = "s") #gráfica cuadrada
plot(c(-2,2),c(-2,2),type="n",xlim=c(-2,2),ylim=c(-2,2),xlab="",ylab "") #gráfica vacía
abline(h=0)
abline(v=0)
arrows(0,0,P[,1],P[,1], length = 0.1)
arrows(0,0,P[,2],P[,2], length = 0.1)
points(vectores, pch = 16, cex = 0.5)
text(vectores, labels = round(z,3), adj=c(0,-0.2),cex = 0.5,col="red")
```

Un recordatorio rápido VIII



Pruebas de hipótesis

Tipos de pruebas

- En estadística multivariada hay dos fuentes de problemas para pruebas de hipótesis:

- El número enorme de hipótesis que se pueden probar. Para una normal multivariada, por ejemplo, hay un total de p parámetros para la media y $p + (p - 1) + \dots + 1 = \frac{p(p+1)}{2}$ para la matriz de varianzas, con un total de $p + \frac{p(p+1)}{2} = p(p+3)/2$ parámetros, de los que se pueden formar $2^{p(p+3)/2}$ subconjuntos a probar.

p	2	3	4	5	6	7
parámetros $\frac{p(p+3)}{2}$	5	9	14	20	27	35
subconjuntos 2^p	32	512	16384	1048576	134217728	3.4359738×10^{10}

- La dificultad entre escoger diferentes estadísticas de pruebas que pueden llevar a diferentes resultados.

- Revisaremos dos enfoques generales:

- ▶ Pruebas de razón de verosimilitud: LRT (likelihood ratio tests).
- ▶ Pruebas de unión-intersección: UIT (union-intersection tests).

En algunos casos, los dos enfoques coinciden en las pruebas, pero no en todos los casos y se pueden obtener diferentes pruebas.

Ejemplo I

Consideren medidas sobre el primer y segundo hijo en una muestra de 25 familias (Frets, 1921). las variables son: 11 y 12 corresponden a la longitud de la cabeza del primer y segundo hijo respectivamente, y b1 y b2 es la amplitud de la cabeza.

```
library(boot) #sólo para tomar los datos
data(frets)
frets$ll

[1] 191 195 181 183 176 208 189 197 188 192 179 183 174 190 188 163 195 186 181 175 192 174 176 197 190

frets$ll2

[1] 179 201 185 188 171 192 190 189 197 187 186 174 185 195 187 161 183 173 182 165 185 178 176 200 187

frets$b1

[1] 155 149 148 153 144 157 150 159 152 150 158 147 150 159 151 137 155 153 145 140 154 143 139 167 163

frets$b2

[1] 145 152 149 149 142 152 149 152 159 151 148 147 152 157 158 130 158 148 146 137 152 147 143 158 150
```

En este ejemplo, podemos hacer una prueba para cada media, para cada varianza y para cada covarianza. Con 4 variables tenemos 14 parámetros a probar y hay 16,384 posibles subconjuntos a formar combinando estos parámetros.

Ejemplo II

- Los estimadores máximo verosímiles para multinormalidad están dados por:

```
xbar <- colMeans(frets); n <- dim(frets)[1]; Sn <- ((n-1)/n)*var(frets)  
xbar  
11      b1      12      b2  
185.72 151.12 183.84 149.24  
  
Sn  
  
11      b1      12      b2  
11 91.4816 50.7536 66.8752 44.2672  
b1 50.7536 52.1856 49.2592 33.6512  
12 66.8752 49.2592 96.7744 54.2784  
b2 44.2672 33.6512 54.2784 43.2224
```

- Supongamos por un momento que 11 y 12 son independientes (aunque claramente no lo son), y que cada variable tiene es normal con varianza de 100, y que se desea probar las dos hipótesis de que ambas medias son 182:

$$H_1 : 11 \sim \mathcal{N}(182, 100), \quad H_2 : 12 \sim \mathcal{N}(182, 100)$$

La prueba usual para cada hipótesis nos daría $z_1 = \frac{185.72 - 182}{10/\sqrt{25}} = 1.86$ y $z_2 = \frac{183.84 - 182}{10/\sqrt{25}} = 0.92$ con *p-values* $p_1 = 0.0628855$ y $p_2 = 0.3575728$, por lo que ninguna de las dos se rechaza al 5% de significancia.

Ejemplo III

- Consideremos ahora la hipótesis *bivariada* de que se den las dos condiciones simultáneamente: $H_3 : H_1 \cap H_2$, la cual será cierta si y sólo si ambas H_1 y H_2 son ciertas.
- Hay varias formas de probar H_3 :
 - Acepta H_3 si y sólo si ambas estadísticas z_1 y z_2 llevan a la aceptación de H_1 y H_2 .
 - Si H_1 y H_2 son ciertas, $z_3 = (z_1 + z_2)/\sqrt{2} \sim \mathcal{N}(0, 1)$. Como $z_3 = 1.965757$, el *p-value* es 0.04932, por lo que se rechaza la hipótesis nula. En este caso, ¡la hipótesis nula multivariada se rechaza a pesar de que cada prueba en lo individual no fue significativa!
 - Si H_1 y H_2 son ciertas, entonces $z_4 = z_1^2 + z_2^2 \sim \chi_2^2$. En este caso, $z_4 = 4.306$ y el *p-value* es 0.1161352, por lo que la prueba no es significativa, por lo que no se rechaza la hipótesis nula.
 - Otra opción es considerar la correlación entre z_1 y z_2 , lo que correspondería a usar la distancia de Mahalanobis, y por lo tanto una región de aceptación elíptica. Este caso se verá en detalle.

Las pruebas anteriores según la estadística pueden tener regiones de rechazo de diferente forma: rectangular, lineal, circular, y elíptica. La forma de la región de aceptación puede llevar a resultados contradictorios.

Pruebas de razón de verosimilitud (LRT)

Pruebas de razón de verosimilitud I

- El objetivo es maximizar la verosimilitud bajo la hipótesis nula y también bajo la alternativa, y comparar las verosimilitudes a través de su cociente.

Def

Sea $\mathbf{x}_1, \dots, \mathbf{x}_n$ una muestra aleatoria de $f(\mathbf{x}|\boldsymbol{\theta})$ y que se quiere contrastar las hipótesis

$$H_0 : \boldsymbol{\theta} \in \Omega_0 \quad \text{vs.} \quad H_1 : \boldsymbol{\theta} \in \Omega_1$$

Entonces la estadística de razón de verosimilitud (LRT) para probar H_0 vs H_1 se define como

$$\lambda(\mathbf{x}) = L_0^*/L_1^*$$

donde $L_i^* = \max\{L(\boldsymbol{\theta}) | \boldsymbol{\theta} \in \Omega_i\}$ y $L(\boldsymbol{\theta})$ es la función de verosimilitud.

- También se puede utilizar la log-verosimilitud, lo que nos daría la estadística:

$$-2 \log \lambda(\mathbf{x}) = 2(l_1^* - l_0^*)$$

- En el caso particular de que $\Omega_0 = \{\boldsymbol{\theta}_0\}$ y $\Omega_1 = \{\boldsymbol{\theta}_1\}$, el lema de Neyman-Pearson no da las propiedades óptimas de λ : es la prueba más poderosa de tamaño α .

Pruebas de razón de verosimilitud II

- En el caso de hipótesis compuestas, en general se favorece H_1 cuando λ tiene un valor pequeño y H_0 cuando λ es grande.

Def

La LRT de tamaño α para probar H_0 vs H_1 tiene región de rechazo

$$R_c = \{\mathbf{x} | \lambda(\mathbf{x}) < c\}$$

donde c es tal que $\sup_{\boldsymbol{\theta} \in \Omega_0} P_{\boldsymbol{\theta}}(\mathbf{x} \in R_c) = \alpha$.

Distribución asintótica de LRT

- Una propiedad asintótica de la prueba LRT que es muy importante y se utiliza con frecuencia está dada por el siguiente teorema²

Teorema

Si $\Omega_1 \subset \mathbb{R}^q$ y si $\Omega_0 \subset \mathbb{R}^r$ es una subregión de Ω_1 , entonces bajo condiciones adecuadas de regularidad, para $\theta \in \Omega_0$, cuando $n \rightarrow \infty$

$$-2 \log \lambda \sim \chi_{q-r}^2$$

- En lo que sigue, veremos tres casos de aplicación de obtención de la prueba LRT en el caso de la normal multivariada.

²no veremos la prueba, pero pueden consultar una guía a la prueba en el libro de Silvey (1970): *Statistical Inference*.

Caso A: $H_0 : \mu = \mu_0$, con Σ_0 conocida I

Para la hipótesis $H_0 : \mu = \mu_0$, con Σ_0 conocida, evaluamos directo l_0 y l_1 y en ambos casos maximizamos.

- Bajo H_0 , $l_0^* = l(\mu_0, \Sigma_0) = -\frac{n}{2} (\log |2\pi\Sigma_0| + \text{tr}(\Sigma_0^{-1}\mathbf{S}) + (\bar{x} - \mu_0)' \Sigma_0^{-1} (\bar{x} - \mu_0))$
- Como H_1 no impone restricciones sobre μ , entonces $l_1^*(\bar{x}, \Sigma_0)$ es el máximo.
- De lo anterior

$$-2 \log \lambda = 2(l_1^* - l_0^*) = n(\bar{x} - \mu_0)' \Sigma_0^{-1} (\bar{x} - \mu_0)$$

Entonces $-2 \log \lambda \sim \chi_p^2$ de manera exacta.

Ejemplo

Para el ejemplo considerado, $(11, 12) \sim \mathcal{N}_2(\mu_0, \Sigma_0)$ con $\mu_0 = (182, 182)'$ y $\Sigma_0 = 100\mathbf{I}_2$. Entonces:

```
LRT <- as.numeric(n*(xbar[c(1,3)]-c(182,182)) %*% solve(100*diag(1,2)) %*% (xbar[c(1,3)]-c(182,182)))  
LRT  
  
[1] 4.306  
  
1- pchisq(LRT, 2)  
  
[1] 0.1161352
```

Caso A: $H_0 : \mu = \mu_0$, con Σ_0 conocida II

Entonces no se tiene evidencia para rechazar la hipótesis de que ambas medias sean iguales a 182. Una región de 95 % de confianza, con $\chi^2_{.95,2} = 5.99$ satisface la ecuación:

$$25(185.72 - \mu_1, 183.84 - \mu_3)'0.01\mathbf{I}_2(185.72 - \mu_1, 183.84 - \mu_3) < 5.99$$

que es la región circular:

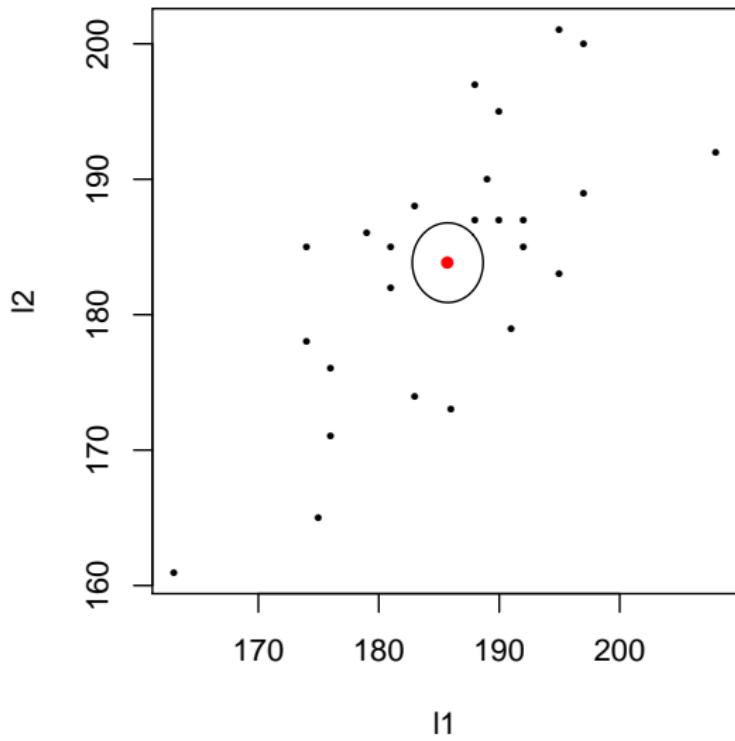
$$(185.72 - \mu_1)^2 + (183.84 - \mu_3)^2 < 23.96$$

```
library(ellipse)

Attaching package: 'ellipse'
The following object is masked from 'package:graphics':
  pairs

par(pty="s")
xbar <- colMeans(frets)[c(1,3)]
plot(frets[,c(1,3)],pch=16,cex=0.5)
lines(ellipse(0,centre = xbar, level = pt(23.96,1))) #encuentra el quantilt t equivalente a la chi dada
points(xbar[1],xbar[2],col="red",cex=0.9,pch=16)
```

Caso A: $H_0 : \mu = \mu_0$, con Σ_0 conocida III



Caso B: $H_0 : \mu = \mu_0$, con Σ desconocida I

Este caso corresponde a una hipótesis compuesta. Para la hipótesis $H_0 : \mu = \mu_0$, con Σ desconocida, se tiene que estimar Σ bajo H_0 y H_1 .

- Bajo H_0 , queremos maximizar $L(\mu_0, \Sigma)$ como función de Σ . El exponente de la verosimilitud es

$$\begin{aligned} -\frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_0) &= -\frac{1}{2} \sum_{i=1}^n \text{tr}(\boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_0)(\mathbf{x}_i - \boldsymbol{\mu}_0)') \\ &= -\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{B}) \end{aligned}$$

donde $\mathbf{B} = \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu}_0)(\mathbf{x}_i - \boldsymbol{\mu}_0)'$. Entonces la verosimilitud queda como

$$L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}) = \frac{1}{2\pi^{np/2} |\boldsymbol{\Sigma}|^{n/2}} \exp\left(-\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{B})\right).$$

Por el lema de maximización que vimos la clase anterior, tomando \mathbf{B} y $b = n/2$, el máximo se alcanza en

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \mathbf{B}$$

Entonces $l_0^* = -\frac{np}{2} \log(2\pi) - \frac{n}{2} \log |\frac{1}{n} \mathbf{B}| - \frac{np}{2}$

Caso B: $H_0 : \mu = \mu_0$, con Σ desconocida II

- Bajo H_1 : no tenemos restricciones así que los estimadores son los usuales de máxima verosimilitud: $\hat{\mu} = \bar{x}$ y $\hat{\Sigma} = S_n$.
- De lo anterior

$$-2 \log \lambda = 2(l_1^* - l_0^*) = -n \log |S_n| + n \log \left| \frac{1}{n} B \right| = n \log \left(\frac{\hat{\Sigma}_0}{\hat{\Sigma}} \right)$$

- Esta estadística se puede escribir como $\lambda^{2/n} = \frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|}$ que se conoce como *la lambda de Wilks*.
- λ está relacionada con la T^2 de Hotelling del siguiente modo:

Teorema (Relación entre λ y T^2)

Sea x_1, \dots, x_n una muestra aleatoria $\mathcal{N}_p(\mu, \Sigma)$. Entonces

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

Caso B: $H_0 : \mu = \mu_0$, con Σ desconocida III

Demostración.

Consideremos la matriz $\mathbf{A}_{(p+1) \times (p+1)}$ dada por

$$\mathbf{A} \begin{pmatrix} \mathbf{S}_n & \sqrt{n}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \\ \sqrt{n}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)' & -1 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$$

Usando propiedades de determinantes, (tarea)

$$|\mathbf{A}| = |\mathbf{A}_{22}| |\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}| = |\mathbf{A}_{11}| |\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}| \quad (1)$$

$$= (-1) |\mathbf{S}_n + n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'| \quad (2)$$

$$= |\mathbf{S}_n| |-1 - n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'(\mathbf{S}_n)^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)| \quad (3)$$

Entonces, sustituyendo en (2) $\mathbf{S}_n + n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)' = \mathbf{S}_0$ donde $\mathbf{S}_0 = \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu}_0)(\mathbf{x}_i - \boldsymbol{\mu}_0)'$ e igualando a (3):

$$-|\mathbf{S}_0| = |\mathbf{S}_n| (-1) \left(1 + \frac{T^2}{n-1} \right)$$

y reacomodando los términos se obtiene el resultado deseado. □

Caso B: $H_0 : \mu = \mu_0$, con Σ desconocida IV

Ejemplo

Para el ejemplo considerado, $(11, 12) \sim \mathcal{N}_2(\mu_0, \Sigma)$ con $\mu_0 = (182, 182)'$ y Σ debe estimarse. Entonces:

```
p <- 2
prueba <- as.numeric((n-p)/p*(xbar[c(1,1)]-c(182,182)) %*% solve(25/24*Sx[c(1,3),c(1,3)]) %*% (xbar[c(1,2)]-c(182,182))
prueba #valor de la estadística

[1] 1.467157

1- pf(prueba,p,n-p) #p-value de la distribución F

[1] 0.2513681
```



Ejemplo 2 I

Los siguientes datos corresponden a la transpiración de 20 mujeres sanas.
 X_1 =tasa de sudoración. X_2 = contenido de sodio y X_3 =contenido de potasio.

Se quiere probar la hipótesis $H_0 : \mu' = (4, 50, 10)$ contra la nula

$H_a : \mu' \neq (4, 50, 10)$ a un nivel de significancia del 10%.

Bajo el supuesto de que estos datos son normales (se puede verificar aplicando las técnicas vistas en el laboratorio pasado)

Ejemplo 2 II

```
X0 <- data.frame( X1=c(3.7,5.7,3.8,3.2,3.1,4.6,2.4,7.2,6.7,5.4,3.9,4.5,3.5,4.5,1.5,8.5,4.5,6.5,4.1,5.5),  
X2=c(48.5,65.1,47.2,53.2,55.5,36.1,24.8,33.1,47.4,54.1,36.9,58.8,27.8,40.2,13.5,  
56.4,71.6,52.8,44.1,40.9),  
X3=c(9.3,8,10.9,12,9.7,7.9,14,7.6,8.5,11.3,12.7,12.3,9.8,8.4,10.1,7.1,8.2,10.9,11.2,9.4))  
mu0 <- c(4,50,10)  
xbar0 <- colMeans(X0)  
xbar0  
  
X1 X2 X3  
4.640 45.400 9.965  
  
S0 <- var(X0)  
S0  
  
X1 X2 X3  
X1 2.879368 10.0100 -1.809053  
X2 10.010000 199.7884 -5.640000  
X3 -1.809053 -5.6400 3.627658  
  
T20 <- as.numeric(nrow(X0) * (t(xbar0-mu0) %*% solve(S0) %*% (xbar0-mu0)))  
T20  
  
[1] 9.738773  
  
#pvalue  
pf(((nrow(X0)-ncol(X0)) / (nrow(X0)-1)) * T20, ncol(X0), nrow(X0)-ncol(X0), lower.tail=F)  
  
[1] 0.001007584
```

La conclusión es que se rechaza la hipótesis nula.

Caso C: $H_0 : \Sigma = \Sigma_0$, con μ desconocida I

- Bajo H_0 los estimadores máximo verosímiles son $\hat{\mu} = \bar{x}$ y $\Sigma = \Sigma_0$.
- Bajo H_1 $\hat{\mu} = \bar{x}$ y $\hat{\Sigma} = \frac{n-1}{n} \mathbf{S}$
- Entonces: $l_0^* = l(\bar{x}, \Sigma_0) = -\frac{n}{2} \log |2\pi \Sigma_0| - \frac{n}{2} \text{tr}(\Sigma_0^{-1} \mathbf{S})$ y
 $l_1^* = l(\bar{x}, \mathbf{S}) = \frac{n}{2} \log |2\pi \mathbf{S}| - \frac{np}{2}$
- De lo anterior

$$\begin{aligned}-2 \log \lambda &= 2(l_1^* - l_0^*) = n \text{tr}(|\Sigma_0^{-1} \mathbf{S}|) - n \log |\Sigma_0^{-1} \mathbf{S}| - np \\&= n \left(\sum_{i=1}^p \lambda_i - \sum_{i=1}^p \log(\lambda_i) - p \right)\end{aligned}$$

donde $\lambda_i \in \text{eigen}\{\Sigma_0^{-1} \mathbf{S}\}$. En este caso, la estadística no tiene una distribución simple. Conviene usar (con n y p adecuados) la distribución asintótica de $-2 \log \lambda \sim \chi_m^2$ donde $m = p(p+1)/2$, el número de parámetros independientes en Σ .

Caso C: $H_0 : \Sigma = \Sigma_0$, con μ desconocida II

Ejemplo

Para el ejemplo considerado, $(11, 12) \sim \mathcal{N}_2(\mu, \Sigma)$ y suponiendo la hipótesis $\Sigma_0 = 100\mathbf{I}_2$. Entonces:

```
Sigma0 <- 100*diag(2)
lambda <- eigen(solve(Sigma0) %*% Sn[c(1, 3), c(1, 3)])$values
lambda

[1] 1.6105554 0.2720046

prueba <- n*(sum(lambda)-sum(log(lambda))-p)
prueba #valor de la estadística

[1] 17.69793

1- pchisq(prueba, p*(p+1)/2) #p-value de la distribución asintótica

[1] 0.0005076659
```

La hipótesis claramente se rechaza. Esto lo sabíamos desde el principio, ya que originalmente los datos están altamente correlacionados (porque son hermanos). Podemos probar una varianza que esté más cercana a la realidad, por ejemplo

$H_0 : \Sigma = \begin{pmatrix} 100 & 50 \\ 50 & 100 \end{pmatrix}$. En este caso:

Caso C: $H_0 : \Sigma = \Sigma_0$, con μ desconocida III

```
Sigma0 <- matrix(c(100,50,50,100), nrow=2)
lambda <- eigen(solve(Sigma0) %*% c(1,3), c(1,3))$values
lambda

[1] 1.0751163 0.5432943

prueba <- n*(sum(lambda)-sum(log(lambda))-p)
prueba #valor de la estadística

[1] 3.902146

1-pchisq(prueba, p*(p+1)/2) #p-value de la distribución asintótica

[1] 0.2722263
```

Como se esperaba, la hipótesis no se rechaza.



Pruebas tipo unión-intersección (UIT)

Pruebas tipo unión-intersección I

Sea $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{I})$ y $\mathbf{a} \in \mathbb{R}^p$. Sabemos que $y_{\mathbf{a}} = \mathbf{a}'\mathbf{x} \sim \mathcal{N}(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\mathbf{a})$. Supongamos que queremos probar $H_0 : \boldsymbol{\mu} = \mathbf{0}$.

Bajo H_0 , $y_{\mathbf{a}} \sim \mathcal{N}(0, \mathbf{a}'\mathbf{a}) \quad \forall \mathbf{a} \in \mathbb{R}^p$. Entonces podemos escribir:

$$H_0 = \bigcap_{\mathbf{a}} H_{0\mathbf{a}}$$

Para una \mathbf{a} fija, podemos utilizar $z_{\mathbf{a}} = \frac{y_{\mathbf{a}}}{\sqrt{\mathbf{a}'\mathbf{a}}} \sim \mathcal{N}(0, 1)$. La región de rechazo para cada $H_{0\mathbf{a}}$ basados en $z_{\mathbf{a}}$ es de la forma

$$R_{\mathbf{a}} = \{z_{\mathbf{a}} : |z_{\mathbf{a}}| > c\} = \{z_{\mathbf{a}} : z_{\mathbf{a}}^2 > c^2\}$$

Como H_0 es verdadera si y sólo si *cada* $H_{0\mathbf{a}}$ es cierta para $\mathbf{a} \in \mathbb{R}^p$. Entonces podemos rechazar H_0 si *alguna* $H_{0\mathbf{a}}$ se rechaza, por lo que la región de rechazo de la prueba es de la forma:

$$R = \bigcup_{\mathbf{a}} R_{\mathbf{a}}$$

En el ejemplo considerado, H_0 no se rechaza si y sólo si $\max_{\mathbf{a}} z_{\mathbf{a}}^2 < c^2$.

- En general, la prueba UIT lleva a una minimización o maximización de una estadística función de $z_{\mathbf{a}}^2$.

Pruebas tipo unión-intersección II

- En el ejemplo dado, $z_a^2 = \frac{y_a}{a'a} = \frac{a'xx'a}{a'a}$. Este cociente de formas cuadráticas se maximiza como función de a cuando $a = x$, por lo que

$$\max_a z_a^2 = \frac{x'xx'x}{x'x} = x'x$$

que se distribuye bajo H_0 como χ_p^2 .

- Una ventaja de UIT sobre LRT es que si H_0 se rechaza, podemos indagar cuál/es región/es R_a es o son las responsables.
- Para comparar UIT a LRT, lo haremos en los casos A, B y C anteriores.
- Una de las aplicaciones más importantes de las pruebas de unión-intersección son los **intervalos de confianza simultáneos**, que se verán después.

Caso A: $H_0: \mu = \mu_0$, con Σ_0 conocida I

- Considerando toda la muestra, sea \mathbf{X} la matriz $n \times p$ con renglones dadas por las observaciones \mathbf{x}_i . Entonces $\mathbf{y}_a = \mathbf{X}\mathbf{a} \sim \mathcal{N}_p(\mathbf{a}'\boldsymbol{\mu}_0, \mathbf{a}'\boldsymbol{\Sigma}_0\mathbf{a})$ es el vector de las combinaciones lineales para cada caso en la muestra.
- Podemos tomar como estadística para la prueba, basada en la media de las observaciones de las combinaciones lineales $\bar{y} = \mathbf{a}'\bar{\mathbf{x}}$ a $z_a^2 = \frac{n\mathbf{a}'(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'a}{a'\boldsymbol{\Sigma}_0a}$. Usando la unión de las regiones de rechazo, Rechazamos para valores grandes del máximo:
$$\max_{\mathbf{a}} z_{\mathbf{a}}^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}_0^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \sim \chi_p^2$$
- Entonces para este caso, el resultado es equivalente al de LRT basado en la ji-cuadrada.

Caso B: $H_0 : \mu = \mu_0$, con Σ desconocida I

- Bajo esta H_0 otra vez $\mathbf{y}_\mathbf{a} = \mathbf{X}\mathbf{a} \sim \mathcal{N}_p(\mathbf{a}'\boldsymbol{\mu}_0, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$, pero en este caso la varianza $\sigma_y^2 = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}$ se tiene que estimar. Un estimador obvio para la varianza es:

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{a}'\mathbf{x}_i - \mathbf{a}'\bar{\mathbf{x}})^2 = \mathbf{a}'\mathbf{S}\mathbf{a}$$

- La estadística de prueba para $H_{0\mathbf{a}}$ es $t_\mathbf{a} = \frac{\bar{y} - \mathbf{a}'\boldsymbol{\mu}_0}{s_y / \sqrt{n-1}}$. Pero noten que $t_\mathbf{a}^2 = (n-1) \frac{\mathbf{a}'(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)' \mathbf{a}}{\mathbf{a}'\mathbf{S}\mathbf{a}}$, por lo que:

$$\max_{\mathbf{a}} t_\mathbf{a}^2 = (n-1)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)' \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)$$

que es la T^2 de Hotelling.

- Entonces también para este caso, UIT \equiv LRT.

Caso C: $H_0 : \Sigma = \Sigma_0$, con μ desconocida I

- Queremos ver si $\sigma_y^2 = \mathbf{a}'\Sigma_0\mathbf{a}$. De nuevo, un estimador de σ_y^2 es $s_y^2 = \mathbf{a}'\mathbf{S}\mathbf{a}$.
- Sea $u_{\mathbf{a}} = \frac{ns_y^2}{\mathbf{a}'\Sigma_0\mathbf{a}} = \frac{n\mathbf{a}'\mathbf{S}\mathbf{a}}{\mathbf{a}'\Sigma_0\mathbf{a}}$ y se rechaza H_0 si $u_{\mathbf{a}} < c_{1a}$ o $u_{\mathbf{a}} > c_{2a}$, donde c_{1a} y c_{2a} se escogen para hacer la prueba de tamaño α . Se consideran ambos posibles casos: el cociente es o muy grande o muy chico (y por lo tanto no cercano a 1).
- Pero como $\max_{\mathbf{a}} u_{\mathbf{a}} = n\lambda_1(\Sigma_0^{-1}\mathbf{S})$ y $\min_{\mathbf{a}} u_{\mathbf{a}} = n\lambda_p(\Sigma_0^{-1}\mathbf{S})$ entonces la región crítica es de la forma:

$$R_{UIT} = \{\lambda_p(\Sigma_0^{-1}\mathbf{S}) < c_1 \text{ o } \lambda_1(\Sigma_0^{-1}\mathbf{S}) > c_2\}$$

Sin embargo, la distribución conjunta de los eigenvalores de $\Sigma_0^{-1}\mathbf{S}$ es muy complicada. En este caso, las pruebas UIT y LRT son diferentes aunque ambas estadísticas dependen de los valores de $\Sigma_0^{-1}\mathbf{S}$.

Intervalos de confianza simultáneos

Afirmaciones de confianza sobre la media I

Dado un vector aleatorio \mathbf{x} , ¿qué podemos decir para los componentes de su vector de medias $\boldsymbol{\mu}$ de manera marginal? ¿y qué relación tendrían las conclusiones marginales que se puedan tomar con la estructura de dependencia conjunta de las diferentes variables?

Consideremos primero el caso conjunto a través de regiones de confianza.

- hemos visto que en las pruebas LRT y UIT las regiones de confianza de tamaño $1 - \alpha$ para $\boldsymbol{\mu}$ son elipsoides dadas por la ecuación:

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

donde estas regiones toman en cuenta la dependencia de las variables en \mathbf{x} . Los ejes del elipsoide están centrados en $\bar{\mathbf{x}}$ y tienen longitud

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

en dirección \mathbf{e}_i donde $\mathbf{S}\mathbf{e}_i = \lambda_i \mathbf{e}_i$, $i = 1, \dots, p$

Afirmaciones de confianza sobre la media II

Ejemplo 1

Para los datos de transpiración que vimos antes, tenemos los siguientes datos:

```
xbar0 #centro de la elipsoide

    X1      X2      X3
4.640 45.400  9.965

S0

    X1      X2      X3
X1  2.879368 10.0100 -1.809053
X2 10.010000 199.7884 -5.640000
X3 -1.809053 -5.6400  3.627658

# Obtenemos la descomposición de S0
lambdas <- eigen(S0)$values
#A continuación se calculan las medias longitudes de cada dirección en R^3 dadas por los eigenvectores
eigen(S0)$vectors

[,1]      [,2]      [,3]
[1,] -0.05084144 -0.57370364  0.81748351
[2,] -0.99828352  0.05302042 -0.02487655
[3,]  0.02907156  0.81734508  0.57541452

p <- 3
n <- 20
for(i in 1:3) print( sqrt(lambdas[i] * p * (n-1) / (n*(n-p)) * qf(.05,p,n-p,lower.tail=F) ) )

[1] 10.36503
[1] 1.558402
[1] 0.835138
```

Afirmaciones de confianza sobre la media III



Afirmaciones de confianza sobre la media: marginales I

Ahora consideremos el caso de las afirmaciones marginales de confianza

- Para afirmaciones marginales o individuales de los componentes de μ , se considera que *todas las afirmaciones que se hagan deben cumplirse simultáneamente con una alta probabilidad*.
- Aquí aplicamos el enfoque UIT. Si $y_a = Xa$ cada componente $y_i \sim \mathcal{N}(a'x_i, a'\Sigma a)$
- Se pueden obtener intervalos de confianza simultáneos para $a'\mu$ para diferentes a . Con a fija, un intervalo de confianza de $100(1 - \alpha)\%$ está dado por:

$$a'\bar{x} - t_{n-1, \alpha/2} \frac{\sqrt{a'Sa}}{\sqrt{n}} \leq a'\mu \leq a'x + t_{n-1, \alpha/2} \frac{\sqrt{a'Sa}}{\sqrt{n}}$$

- Eligiendo diferentes coeficientes a , podemos probar diferentes hipótesis para cada componente de μ y cada uno tendrá confianza $1 - \alpha$. **Pero la confianza asociada a todas las afirmaciones tomadas de manera conjunta no tiene confianza $1 - \alpha$.**

Afirmaciones de confianza sobre la media: marginales II

Ejemplo

En el ejemplo de transpiración, los intervalos marginales para cada media están dados por:

```
xbar0

      X1      X2      X3
4.640 45.400  9.965

for(i in 1:3)print(xbar0[i]+c(-1,1)*qt(.025,n-1,lower.tail=F)*sqrt(S0[i,i]/n))

[1] 3.84584 5.43416
[1] 38.78478 52.01522
[1] 9.073601 10.856399

#Ancho de los intervalos:
for(i in 1:3)print(2*qt(.025,n-1,lower.tail=F)*sqrt(S0[i,i]/n))

[1] 1.588319
[1] 13.23044
[1] 1.782799
```



Afirmaciones simultáneas de confianza de la media I

- Para lograr una confianza colectiva para *todas* las elecciones de \mathbf{a} de $1 - \alpha$, los intervalos marginales serán menos precisos que los considerados para una elección específica de \mathbf{a} .
- Para encontrar la c^2 tal que $t^2 = \frac{n(\mathbf{a}'\bar{\mathbf{x}} - \mathbf{a}'\boldsymbol{\mu})^2}{\mathbf{a}'\mathbf{S}\mathbf{a}} \leq c^2$, podemos hacer c^2 el máximo de las t^2 sobre todas las posibles \mathbf{a} :

$$\max_{\mathbf{a}} t^2 = \max_{\mathbf{a}} \frac{n(\mathbf{a}'(\bar{\mathbf{x}} - \boldsymbol{\mu}))^2}{\mathbf{a}'\mathbf{S}\mathbf{a}}$$

que utilizando el lema de maximización que vimos antes, nos da:

$$\max_{\mathbf{a}} \frac{n(\mathbf{a}'(\bar{\mathbf{x}} - \boldsymbol{\mu}))^2}{\mathbf{a}'\mathbf{S}\mathbf{a}} = n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) = T^2$$

y el máximo se alcanza cuando \mathbf{a} es proporcional a $\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu})$.

- Entonces, simultáneamente para toda \mathbf{a} , el intervalo

$$\left(\mathbf{a}'\bar{\mathbf{x}} - \sqrt{\frac{p(n-1)}{n-p} F_{p,n-p,\alpha} \mathbf{a}'\mathbf{S}\mathbf{a}}, \mathbf{a}'\bar{\mathbf{x}} + \sqrt{\frac{p(n-1)}{n-p} F_{p,n-p,\alpha} \mathbf{a}'\mathbf{S}\mathbf{a}} \right)$$

contiene a $\mathbf{a}'\boldsymbol{\mu}$ con probabilidad $1 - \alpha$.

Afirmaciones simultáneas de confianza de la media II

Ejemplo 1

Otra vez, en el ejemplo de transpiración, los intervalos simultáneos de tamaño $1 - \alpha$ están dados por:

```
c <- sqrt(p * (n-1) / (n-p) * qf(.05,p,n-p,lower.tail=F))
xbar0

      X1      X2      X3
4.640 45.400 9.965

for(i in 1:3)print(xbar0[i]+c(-1,1)*c*sqrt(S0[i,i]/n))

[1] 3.397768 5.882232
[1] 35.05241 55.74759
[1] 8.570664 11.359336

#Ancho de los intervalos:
for(i in 1:3)print(2*c*sqrt(S0[i,i]/n))

[1] 2.484464
[1] 20.69518
[1] 2.788671
```

Afirmaciones simultáneas de confianza de la media III



Ejemplo 2

Usando el enfoque UIT, podemos calcular diferencias de medias, sólo tomando la combinación lineal adecuada. El ancho de la constante no cambia, sólo cambia la combinación lineal a tomar. Por ejemplo, para encontrar un intervalo de confianza para la diferencia de medias $\bar{y}_1 - \bar{y}_2$, tomamos $a = (1, -1, 0)$. Este vector usualmente se llama *contraste*.

```
a <- c(1, -1, 0)
as.numeric(t(a) %*% xbar0) #valor de y1-y2

[1] -40.76

as.numeric(t(a) %*% xbar0) + c(-1, 1)*c*sqrt(as.numeric(t(a) %*% S0 %*% a)) #intervalo para la diferencia

[1] -85.006237  3.486237
```

