

# Estadística Aplicada III

## Distribución normal multivariada

### Distribuciones muestrales

Jorge de la Vega Góngora

Departamento de Estadística,  
Instituto Tecnológico Autónomo de México

Semana 5



**ITAM**

# 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	$1.6384 \times 10^4$	$1.048576 \times 10^6$	$1.3421773 \times 10^8$	$3.4359738 \times 10^{10}$

- La dificultad entre escoger diferentes estadísticas de pruebas que pueden llevar a diferentes resultados.
- Revisaremos dos enfoques generales para construir pruebas de hipótesis:
    - 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: l1 y l2 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)
str(frets)
```

```
'data.frame': 25 obs. of 4 variables:
 $ l1: num  191 195 181 183 176 208 189 197 188 192 ...
 $ b1: num  155 149 148 153 144 157 150 159 152 150 ...
 $ l2: num  179 201 185 188 171 192 190 189 197 187 ...
 $ b2: num  145 152 149 149 142 152 149 152 159 151 ...
```

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)) # vector de medias
      l1      b1      l2      b2
185.72 151.12 183.84 149.24

(n <- dim(frets)[1])      # número de observaciones
[1] 25

(Sn <- ((n-1)/n)*var(frets)) # EMV de la varianza
      l1      b1      l2      b2
l1 91.4816 50.7536 66.8752 44.2672
b1 50.7536 52.1856 49.2592 33.6512
l2 66.8752 49.2592 96.7744 54.2784
b2 44.2672 33.6512 54.2784 43.2224
```

- Supongamos por un momento que las variables 11 y 12 son independientes (aunque claramente no lo son), y que ambas variables son normales con varianza de 100. 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 univariada para cada hipótesis nos daría estadísticas de prueba

$$z_1 = \frac{185.72 - 182}{10/\sqrt{25}} = 1.86 \text{ y } z_2 = \frac{183.84 - 182}{10/\sqrt{25}} = 0.92 \text{ con } p\text{-values } p_1 = 0.0628855 \text{ y}$$

## Ejemplo III

$p_2 = 0.3575728$ , por lo que ninguna de las dos hipótesis se rechaza al 5 % de significancia.

- Considerando ahora la hipótesis *bivariada* de que se den las dos condiciones simultáneamente:  $H_3 : H_1 \cap H_2$ , que será cierta si y sólo si ambas  $H_1$  y  $H_2$  son ciertas.
- Hay varias posibles formas de probar  $H_3$ :
  - 1 Acepta  $H_3$  si y sólo si ambas estadísticas  $z_1$  y  $z_2$  llevan a la aceptación de  $H_1$  y  $H_2$ .
  - 2 Si  $H_1$  y  $H_2$  son verdaderas,  $z_3 = (z_1 + z_2)/\sqrt{2} \sim \mathcal{N}(0, 1)$ . Como  $z_3 = 1.965757$ , el *p-value* es 0.04932, por lo que se rechaza  $H_3$ . En este caso, ¡la hipótesis nula multivariada se rechaza a pesar de que cada prueba en lo individual no fue significativa!
  - 3 Si  $H_1$  y  $H_2$  son verdaderas, 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  $H_3$ .
  - 4 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 más adelante.

Las pruebas anteriores, según la estadística de prueba que se utilice, tienen regiones de rechazo de diferente forma geométrica: 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 de una prueba LRT 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  $f(\mathbf{x}|\theta)$  y que se quiere contrastar las hipótesis

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

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

- También se puede utilizar la log-verosimilitud, a través de la siguiente transformación de la estadística LRT:

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

- En el caso particular de hipótesis simples  $\Omega_0 = \{\theta_0\}$  y  $\Omega_1 = \{\theta_1\}$ , el lema de Neyman-Pearson no da las propiedades óptimas de  $\lambda$ : es la prueba más poderosa de tamaño  $\alpha$ .
- En el caso de hipótesis compuestas, en general se favorece  $H_1$  cuando  $\lambda$  tiene un valor pequeño y  $H_0$  cuando  $\lambda$  es grande.

## Def

La estadística LRT de tamaño  $\alpha$  para probar  $H_0$  vs  $H_1$  tiene región de rechazo

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

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

- Una propiedad asintótica de la prueba LRT que es muy importante y se utiliza con frecuencia está dada por el siguiente teorema:<sup>1</sup>

## Teorema

Si  $\Omega_1 \subset \mathbb{R}^q$  y si  $\Omega_0 \subset \mathbb{R}^r$  es una subregión de  $\Omega_1$ , 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.

---

<sup>1</sup>no veremos la prueba, pero pueden consultar una guía a la prueba en el libro de Silvey (1970): *Statistical Inference* o en el libro de Cox y Hunkley (1974) *Theoretical Statistics*. Esta última usa la función score y la información de Fisher.

## Caso A: $H_0 : \mu = \mu_0$ , con $\Sigma_0$ conocida I

Para la hipótesis  $H_0 : \mu = \mu_0$ , con  $\Sigma_0$  conocida, evaluamos directo  $l_0$  y  $l_1$  y en ambos casos maximizamos.

- Bajo  $H_0$ ,  $l_0^* = l(\mu_0, \Sigma_0) = -\frac{n}{2} (\log |2\pi\Sigma_0| + \text{tr}(\Sigma_0^{-1}\mathbf{S}_n) + (\bar{x} - \mu_0)' \Sigma_0^{-1} (\bar{x} - \mu_0))$ .
- Como la hipótesis alternativa es  $H_1 : \mu \neq \mu_0$ , no impone restricciones sobre  $\mu$  excepto que no se evalúe en un punto, entonces  $l_1^*(\bar{\mathbf{x}}, \Sigma_0)$  es el máximo.
- De lo anterior y considerando que la dimensión del espacio parametral de  $H_0$  es  $r = 0$  (es un punto) y  $q = p$

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

Entonces  $-2 \log \lambda \sim \chi_{(p)}^2$  de manera exacta.

### **Ejemplo. [Hermanos, cont.]**

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

## Caso A: $H_0 : \mu = \mu_0$ , con $\Sigma_0$ conocida II

```
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

pchisq(LRT,2,lower.tail = F)

[1] 0.1161352
```

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:

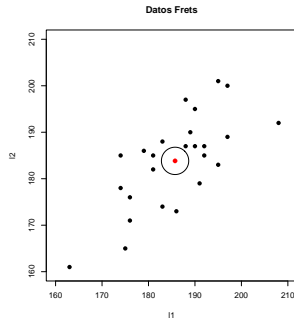
$$25(185.72 - \mu_1, 183.84 - \mu_3)' 0.01 \mathbf{I}_2 (185.72 - \mu_1, 183.84 - \mu_3) < 5.99$$

que es la región elíptica (en este caso circular, ya que la varianza es proporcional a la identidad):

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

# Caso A: $H_0 : \mu = \mu_0$ , con $\Sigma_0$ conocida III

```
library(ellipse)
par(pty = "s")
xbar <- colMeans(frets)[c(1,3)]
plot(frets[,c(1,3)], pch = 16, cex = 1.2, main = "Datos Frets", ylim=c(160,210),xlim=c(160,210))
lines(ellipse(0, centre = xbar, level = pt(23.96,1))) #encuentra el quantil t equivalente a la chi dada
points(xbar[1], xbar[2], col = "red", cex = 1.3, pch = 16)
```



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

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

- Bajo  $H_0$ , queremos maximizar  $l_0(\mu_0, \Sigma)$  como función de  $\Sigma$ .  
El exponente de la parte exponencial de  $L_0$  es

$$\begin{aligned} -\frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \mu_0)' \Sigma^{-1} (\mathbf{x}_i - \mu_0) &= -\frac{1}{2} \sum_{i=1}^n \text{tr}(\Sigma^{-1} (\mathbf{x}_i - \mu_0)(\mathbf{x}_i - \mu_0)') \\ &= -\frac{1}{2} \text{tr}(\Sigma^{-1} \mathbf{B}) \end{aligned}$$

donde  $\mathbf{B} = \sum_{i=1}^n (\mathbf{x}_i - \mu_0)(\mathbf{x}_i - \mu_0)'$ . Entonces la verosimilitud queda como

$$L(\mu_0, \Sigma) = \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left(-\frac{1}{2} \text{tr}(\Sigma^{-1} \mathbf{B})\right).$$

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

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

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

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

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

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

- Esta estadística se puede escribir como  $\lambda = \left( \frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|} \right)^{n/2}$  que se conoce como *la lambda de Wilks*.



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

- $\lambda$  está relacionada con la  $T^2$  de Hotelling del siguiente modo:

### Teorema (Relación entre $\lambda$ de Wilks y $T^2$ de Hotelling)

Sea  $\mathbf{x}_1, \dots, \mathbf{x}_n$  una muestra aleatoria  $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Entonces

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

Entonces la prueba basada en la lambda de Wilks es equivalente a una prueba basada en la  $T^2$  de Hotelling.

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

### Demostración.

Consideremos la matriz  $\mathbf{A}_{(p+1) \times (p+1)}$  dada por  $\mathbf{A} \begin{pmatrix} n\mathbf{S}_n & \sqrt{n}(\bar{\mathbf{x}} - \mu_0) \\ \sqrt{n}(\bar{\mathbf{x}} - \mu_0) & -1 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$

Usando propiedades de determinantes, (tarea)

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

$$= (-1)|n\mathbf{S}_n + n(\bar{\mathbf{x}} - \mu_0)(\bar{\mathbf{x}} - \mu_0)'| \quad (2)$$

$$= |n\mathbf{S}_n| \cdot |-1 - n(\bar{\mathbf{x}} - \mu_0)'(n\mathbf{S}_n)^{-1}(\bar{\mathbf{x}} - \mu_0)| \quad (3)$$

$$= (-1)|n\mathbf{S}_n| \left(1 + \frac{T^2}{n-1}\right), \quad (4)$$

ya que  $(\bar{\mathbf{x}} - \mu_0)'(\mathbf{S}_n)^{-1}(\bar{\mathbf{x}} - \mu_0) = \frac{1}{n-1}(\bar{\mathbf{x}} - \mu_0)'(\mathbf{S})^{-1}(\bar{\mathbf{x}} - \mu_0) = \frac{T^2}{n-1}$ . Entonces, sustituyendo en (2)  $n\mathbf{S}_n + n(\bar{\mathbf{x}} - \mu_0)(\bar{\mathbf{x}} - \mu_0)' = n\mathbf{S}_0$  donde  $n\mathbf{S}_0 = \sum_{i=1}^n (\mathbf{x}_i - \mu_0)(\mathbf{x}_i - \mu_0)'$  e igualando a (3):

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

y reacomodando los términos se obtiene el resultado deseado. □

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

## Ejemplo. [Hermanos, cont.]

Para el ejemplo considerado,  $(11, 12) \sim \mathcal{N}_2(\mu_0, \Sigma)$  con  $\mu_0 = (182, 182)'$  y  $\Sigma$  debe estimarse con el estimador máximo verosímil  $\frac{n}{n-1} \mathbf{S}$ . Podemos usar la estadística de Hotelling dada la relación con la estadística de Wilks. Entonces:

```
p <- 2
# La constante de la prueba resulta de despejar el valor de F(p,n-p)
prueba <- as.numeric(n*(n-p)/(p*(n-1))*(xbar[c(1,1)]-c(182,182)) %*% solve(25/24*Sn[c(1,3),c(1,3)]) %*% (xbar[c(1,2)]-c(182,182)))
prueba #valor de la estadística

[1] 1.528289

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

[1] 0.2381334
```



## 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 nula  $H_0 : \boldsymbol{\mu}' = (4, 50, 10)$  contra la alternativa  $H_a : \boldsymbol{\mu}' \neq (4, 50, 10)$  a un nivel de significancia del 10 %.

Bajo el supuesto de que estos datos son normales:

# 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)
n <- nrow(X0); p <- ncol(X0)
(xbar0 <- colMeans(X0))

      X1      X2      X3
4.640 45.400 9.965

(S0 <- var(X0))

      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(n*(t(xbar0 - mu0) %*% solve(S0) %*% (xbar0 - mu0))))

[1] 9.738773

pf((n-p)/(p*(n-1))*T20, p, n-p, lower.tail = F) # pvalue

[1] 0.06492834
```

La conclusión es que no se tiene evidencia para rechazar la hipótesis nula.

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

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

$$\begin{aligned} -2 \log \lambda &= 2(l_1^* - l_0^*) = n \text{tr}(|\Sigma_0^{-1} \mathbf{S}|) - n \log |\Sigma_0^{-1} \mathbf{S}| - np \\ &= n \cdot \text{tr}(\Sigma_0^{-1} \mathbf{S}) - n \log |\Sigma_0^{-1} \mathbf{S}| - np \\ &= np(a - \log(g) - 1) \end{aligned}$$

donde  $a$  es el promedio y  $g$  es la media geométrica de los valores propios de  $\Sigma_0^{-1} \mathbf{S}$ , de tal manera que  $\text{tr}(\Sigma_0^{-1} \mathbf{S}) = pa$  y  $|\Sigma_0^{-1} \mathbf{S}| = g^p$ .

En este caso, la estadística no tiene una distribución simple. Conviene usar (siempre que  $n$  y  $p$  sean adecuados) la distribución asintótica de  $-2 \log \lambda \sim \chi_m^2$  donde  $m = p(p+1)/2$ , el número de parámetros independientes en  $\Sigma$ .

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

## Ejemplo. [ Hermanos, cont.]

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

```
S0 <- 100*diag(2)
S1 <- Sn[c(1,3),c(1,3)] # Sn es el EMV de los datos
(lambdas <- eigen(solve(S0) %*% S1)$values)

[1] 1.6105554 0.2720046

a <- mean(lambdas); g <- exp(mean(log(lambdas)))
(prueba <- n*p*(a - log(g) - 1)) # valor de la estadística

[1] 21.23752

pchisq(prueba, p*(p+1)/2, lower.tail = F) #p-value de la distribución asintótica

[1] 0.001662748
```

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 $\mu$ desconocida III

```
S0 <- matrix(c(100,50,50,100), nrow = 2)
(lambdas <- eigen(solve(S0) %*% S1)$values)

[1] 1.0751163 0.5432943

a <- mean(lambdas)
g <- exp(mean(log(lambdas)))
prueba <- n*p*(a - log(g) - 1) # valor de la estadística
1 - pchisq(prueba, p*(p+1)/2) # p-value de la distribución asintótica, otra forma de cálculos

[1] 0.5851205
```

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  $H_{0\mathbf{a}}$  basada en la estadística  $z_{\mathbf{a}}$  es de la forma

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

- Como  $H_0$  es verdadera si y sólo si *cada*  $H_{0\mathbf{a}}$  es verdadera 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:

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

# Pruebas tipo unión-intersección II

- En el ejemplo considerado,  $H_0$  se acepta si y sólo si  $z_{\mathbf{a}}^2 \leq c^2$  para toda  $z_{\mathbf{a}}$  o equivalentemente,  $H_0$  se acepta 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  $z_{\mathbf{a}}^2$ .
- En el ejemplo dado,  $z_{\mathbf{a}}^2 = \frac{y_{\mathbf{a}}^2}{\mathbf{a}'\mathbf{a}} = \frac{\mathbf{a}'\mathbf{xx}'\mathbf{a}}{\mathbf{a}'\mathbf{a}}$ . Este cociente de formas cuadráticas se maximiza<sup>2</sup> como función de  $\mathbf{a}$  cuando  $\mathbf{a} = \mathbf{x}$ , por lo que

$$\max_{\mathbf{a}} z_{\mathbf{a}}^2 = \frac{\mathbf{x}'\mathbf{xx}'\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \mathbf{x}'\mathbf{x}$$

que se distribuye bajo  $H_0$  como  $\chi_p^2$ .

- Una ventaja de UIT sobre LRT es que si  $H_0$  se rechaza, podemos indagar cuál(es) región(es)  $R_{\mathbf{a}}$  es (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 a continuación.

---

<sup>2</sup>Ver los teoremas de maximización de formas cuadrática

## UIT Caso A: $H_0 : \mu = \mu_0$ , con $\Sigma_0$ conocida I

- Considerando toda la muestra, sea  $\mathbf{X}$  la matriz  $n \times p$  con renglones dadas por las observaciones  $\mathbf{x}_i$ . Entonces  $\mathbf{y}_a = \mathbf{X}\mathbf{a}$  es un vector donde cada componente tiene distribución  $\mathcal{N}(\mathbf{a}'\mu_0, \mathbf{a}'\Sigma_0\mathbf{a})$ .
- Tomando  $\bar{y} = \mathbf{a}'\bar{\mathbf{x}}$  podemos tomar como estadística de prueba a

$$z_{\mathbf{a}}^2 = \frac{n\mathbf{a}'(\bar{\mathbf{x}} - \mu_0)(\bar{\mathbf{x}} - \mu_0)'\mathbf{a}}{\mathbf{a}'\Sigma_0\mathbf{a}}.$$

- Usando la unión de las regiones de rechazo y el lema de optimización de cociente de formas cuadráticas, rechazamos para valores grandes del máximo:

$$\max_{\mathbf{a}} z_{\mathbf{a}}^2 = n(\bar{\mathbf{x}} - \mu_0)'\Sigma_0^{-1}(\bar{\mathbf{x}} - \mu_0) \sim \chi_p^2$$

- Entonces para este caso, el resultado es equivalente al de LRT basado en la chi-cuadrada.

## UIT Caso B: $H_0 : \mu = \mu_0$ , con $\Sigma$ desconocida I

- Bajo esta hipótesis otra vez tenemos que cada componente de  $\mathbf{y}_a = \mathbf{X}\mathbf{a}$  tiene distribución  $\sim \mathcal{N}(\mathbf{a}'\mu_0, \mathbf{a}'\Sigma\mathbf{a})$ , pero en este caso la varianza de cada componente es  $\sigma_y^2 = \mathbf{a}'\Sigma\mathbf{a}$  y 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  $H_{0a}$  es  $t_a = \frac{\bar{y} - \mathbf{a}'\mu_0}{s_y/\sqrt{n-1}}$ . Pero noten que se puede escribir como  $t_a^2 = (n-1) \frac{\mathbf{a}'(\bar{\mathbf{x}} - \mu_0)(\bar{\mathbf{x}} - \mu_0)'\mathbf{a}}{\mathbf{a}'\mathbf{S}\mathbf{a}}$ , por lo que:

$$\max_a t_a^2 = (n-1)(\bar{\mathbf{x}} - \mu_0)'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \mu_0)$$

que es la  $T^2$  de Hotelling.

- Entonces también para este caso,  $\text{UIT} \equiv \text{LRT}$ .

## UIT Caso C: $H_0 : \Sigma = \Sigma_0$ , con $\mu$ desconocida I

- Queremos ver si  $\sigma_y^2 = \mathbf{a}'\Sigma_0\mathbf{a}$ . De nuevo, un estimador de  $\sigma_y^2$  es  $s_y^2 = \mathbf{a}'\mathbf{S}\mathbf{a}$ .
- Sea  $u_{\mathbf{a}} = \frac{s_y^2}{\mathbf{a}'\Sigma_0\mathbf{a}} = \frac{\mathbf{a}'\mathbf{S}\mathbf{a}}{\mathbf{a}'\Sigma_0\mathbf{a}}$  y se rechaza  $H_{0\mathbf{a}}$  si estamos en las colas de la distribución, es decir si  $u_{\mathbf{a}} < c_{1\mathbf{a}}$  o  $u_{\mathbf{a}} > c_{2\mathbf{a}}$ , donde  $c_{1\mathbf{a}}$  y  $c_{2\mathbf{a}}$  se escogen para hacer la prueba de tamaño  $\alpha$ . Se consideran ambos posibles casos: el cociente es o muy grande o muy chico (y por lo tanto no cercano a 1).
- Pero como  $\max_{\mathbf{a}} u_{\mathbf{a}} = \lambda_1(\Sigma_0^{-1}\mathbf{S})$  y  $\min_{\mathbf{a}} u_{\mathbf{a}} = \lambda_p(\Sigma_0^{-1}\mathbf{S})$ , donde  $\lambda_1$  y  $\lambda_p$  son los valores propios más grande y más chico de  $\Sigma_0^{-1}\mathbf{S}$ , entonces la región crítica es de la forma:

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

Para calcular la distribución de la prueba, necesitamos conocer la distribución conjunta de los valores propios de la matriz  $\Sigma_0^{-1}\mathbf{S}$ . Sin embargo, la distribución es muy complicada y no hay tablas para esta prueba.

- Entonces, las pruebas UIT y LRT son diferentes en este caso, aunque ambas estadísticas dependen de los valores propios de  $\Sigma_0^{-1}\mathbf{S}$ .

## Regiones e intervalos de confianza

# Afirmaciones de confianza sobre la media I

Dado un vector aleatorio  $\mathbf{x}$ , dos preguntas:

- 1 ¿Qué relación tienen las conclusiones marginales que se puedan tomar con la estructura de dependencia conjunta de las diferentes variables?
- 2 ¿Qué podemos decir para los componentes de su vector de medias  $\mu$  de manera individual?

Trataremos de responder ambas preguntas. Para la primera pregunta, consideremos la región conjunta.

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

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



# Afirmaciones de confianza sobre la media II

donde estas regiones toman en cuenta la dependencia de las variables en  $\mathbf{x}$  a través de  $\mathbf{S}$ . 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 de los vectores propios  $\mathbf{e}_i$ , donde  $\mathbf{S}\mathbf{e}_i = \lambda_i \mathbf{e}_i$ ,  $i = 1, \dots, p$ .

## Ejemplo. [Función de cálculo]

Hagamos una función para calcular los ejes de los elipsoides

```
lim.region.confianza <- function(X, alpha = 0.05){  
  # X es el conjunto de datos  
  p <- dim(X)[2]; n <- dim(X)[1]  
  xbar <- colMeans(X)  
  S <- var(X)  
  vp <- eigen(S)  
  v <- sqrt(vp$values*p*(n-p)/(n*(n-p))*qf(alpha,p,n-p,lower.tail = F))  
  return(list(lims = v, vals = vp$values, vec = vp$vectors))  
}
```

# Afirmaciones de confianza sobre la media III

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

```
lim.region.confianza(X0)

$lims
[1] 9.8043391 1.4741006 0.7899615

$vals
[1] 200.462464 4.531591 1.301392

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



## ***Ejemplo. [Ejemplo 5.3, p. 221]***

Con los datos de radiación del ejemplo 4.10 (Tabla 4.1, p.180) del libro de J&W, con las transformaciones para normalizar (sacando raíz cuarta de cada columna):

# Afirmaciones de confianza sobre la media IV

```
X <- data.frame(x1 = readLines("https://raw.githubusercontent.com/jvega68/EA3/master/datos/J%26W/T4-1.DAT"),
               x2 = readLines("https://raw.githubusercontent.com/jvega68/EA3/master/datos/J%26W/T4-5.DAT"))
X <- apply(X, 2, function(x) as.numeric(x)^(.25))
lim.region.confianza(X)
```

\$lims

[1] 0.06345367 0.02050405

\$vals

[1] 0.026163638 0.002731895

\$vec

	[,1]	[,2]
[1,]	0.7041574	-0.7100439
[2,]	0.7100439	0.7041574



# Afirmaciones simultáneas de confianza de la media I

- Para lograr una confianza colectiva de  $1 - \alpha$  para *cualquier* elección de  $\mathbf{a}$ , aplicamos el enfoque UIT. Si  $\mathbf{y}_a = \mathbf{X}\mathbf{a}$ , cada componente  $y_i \sim \mathcal{N}(\mathbf{a}'\mathbf{x}_i, \mathbf{a}'\Sigma\mathbf{a})$ . Se pueden obtener intervalos de confianza simultáneos para  $\mathbf{a}'\boldsymbol{\mu}$  para diferentes  $\mathbf{a}$ .
- Con  $\mathbf{a}$  fija, un intervalo de confianza de  $100(1 - \alpha) \%$  está dado por:

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

- Para encontrar el límite de la región de aceptación-rechazo  $c^2$  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:

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

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

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

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

- Estos intervalos para cada media coinciden con las proyecciones de los límites que contienen a la elipse: metemos la elipse en una región rectangular que se ciñe a la elipse.
- Los intervalos de confianza simultáneos  $T^2$  son ideales para ‘husmear’, ya que el nivel de confianza no cambia independientemente de la  $\mathbf{a}$  tomada.

# Afirmaciones simultáneas de confianza de la media III

## Ejemplo. [Función de cálculo]

Hacemos una función que haga los cálculos por nosotros:

```
int.simultaneos <- function(X,a, alfa = 0.05){  
  # X es la matriz de datos, a es el vector combinación lineal  
  p <- dim(X)[2]; n <- dim(X)[1]  
  xbar <- colMeans(X)  
  S <- var(X)  
  lims <- sum(a*xbar) + c(-1,1)*as.numeric(sqrt(p*(n-1)/(n-p)*qf(1-alfa,p,n-p)*t(a) %*% S %*% a/n))  
  return(list(media = sum(a*xbar), limites = lims))  
}
```



## Ejemplo. [Transpiración, cont.]

En el ejemplo de los datos de transpiración, los intervalos simultáneos para cada componente del vector de medias están dados por:

# Afirmaciones simultáneas de confianza de la media IV

```
int.simultaneos(X0, a = c(1,0,0))
```

```
$media  
[1] 4.64
```

```
$limites  
[1] 3.397768 5.882232
```

```
int.simultaneos(X0, a = c(0,1,0))
```

```
$media  
[1] 45.4
```

```
$limites  
[1] 35.05241 55.74759
```

```
int.simultaneos(X0, a = c(0,0,1))
```

```
$media  
[1] 9.965
```

```
$limites  
[1] 8.570664 11.359336
```



***Ejemplo. [Ej. 5.5, pag. 226]***



# Afirmaciones simultáneas de confianza de la media VI

```
X <- read.delim("https://raw.githubusercontent.com/jvega68/EA3/master/datos/J&W/T5_2.DAT", sep = ",", header=F)
int.simultaneos(X, a = c(1,0,0))
```

```
$media
[1] 526.5862
```

```
$limites
[1] 502.9994 550.1730
```

```
int.simultaneos(X, a = c(0,1,0))
```

```
$media
[1] 54.68966
```

```
$limites
[1] 51.21484 58.16447
```

```
int.simultaneos(X, a = c(0,0,1))
```

```
$media
[1] 25.12644
```

```
$limites
[1] 23.63855 26.61432
```



## Ejemplo. [3]

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  $\mathbf{a} = (1, -1, 0)$ . Este vector usualmente se llama *contraste*.

```
int.simultaneos(X0, a = c(1,-1,0))

$media
[1] -40.76

$limites
[1] -50.65376 -30.86624

int.simultaneos(X, a = c(0, 1, -1)) # calculo ejemplo 5.5 p.229

$media
[1] 29.56322

$limites
[1] 26.43153 32.69491
```

# Afirmaciones simultáneas de confianza de la media VIII



# Afirmaciones de confianza sobre la media: marginales I

Ahora consideremos el caso de las afirmaciones marginales o individuales de confianza.

- Con el mismo enfoque, Con  $\mathbf{a}$  fija, un intervalo de confianza de  $100(1 - \alpha) \%$  está dado por:

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

- Eligiendo diferentes coeficientes  $\mathbf{a} = \mathbf{e}_1, \mathbf{e}_2$ , podemos obtener los diferentes hipótesis para cada componente de  $\boldsymbol{\mu}$  y cada uno tendrá confianza  $1 - \alpha$ .
- Pero la confianza conjunta asociada a todas las afirmaciones tomadas no tiene confianza  $1 - \alpha$ .** Usualmente los intervalos individuales son más cerrados o cortos, y por eso tienen una confianza menor considerados simultáneamente.
- Lo cerrado de los intervalos individuales comparados con los simultáneos dependerá de los valores de  $n$ ,  $p$  y del nivel de significancia  $\alpha$ .

# Afirmaciones de confianza sobre la media: marginales II

- Se consideran importantes en el sentido de que contienen la mejor información marginal para componentes específicos del vector media. Esto da una mejor estimación marginal.

## Ejemplo. [Función de cálculo]

Hacemos una función que haga los cálculos por nosotros:

```
int.ind <- function(X, a, alfa = 0.05){  
  # X es la matriz de datos, a es el vector combinación lineal  
  p <- dim(X)[2]; n <- dim(X)[1]  
  xbar <- colMeans(X)  
  S <- var(X)  
  lims <- sum(a*xbar) + c(-1,1)*abs(qt(1-alfa/2,n-1))*sqrt(as.numeric(t(a) %*% S %*% a)/n)  
  return(lims)  
}
```



## Ejemplo. [Transpiración, cont.]

En el ejemplo de los datos de transpiración, los intervalos marginales para cada componente del vector de medias están dados por:

# Afirmaciones de confianza sobre la media: marginales III

```
int.ind(X0, a = c(1,0,0))
```

```
[1] 3.84584 5.43416
```

```
int.ind(X0, a = c(0,1,0))
```

```
[1] 38.78478 52.01522
```

```
int.ind(X0, a = c(0,0,1))
```

```
[1] 9.073601 10.856399
```



# Método de Bonferroni para comparaciones múltiples I

- El método de Bonferroni pretende mejorar la estimación del intervalo, considerando que no todas las combinaciones lineales  $\mathbf{a}'\bar{\mathbf{x}}$  son relevantes o de interés.
- Es un punto de vista intermedio entre los intervalos individuales y los simultáneos, que busca obtener intervalos más precisos. En muchos contextos se preferirán los intervalos bonferronizados a los simultáneos.
- Este método se basa en la *desigualdad de Bonferroni*:

$$P\left(\bigcap_{i=1}^n C_i\right) \geq \sum_{i=1}^n P(C_i) - (n - 1)$$

# Método de Bonferroni para comparaciones múltiples II

- Ex-ante, se requiere hacer afirmaciones de  $k$  combinaciones lineales, para cada una de las cuales, hay una región de confianza  $C_i$  para  $H_i$  y supongamos que  $P(C_i) = 1 - \alpha_i$ . Entonces:

$$P\left(\bigcap_{i=1}^k C_i\right) = 1 - P\left(\bigcup_{i=1}^k C_i^c\right) \geq 1 - \sum_{i=1}^k P(C_i^c) = 1 - \sum_{i=1}^k (1 - P(C_i)) = 1 - \sum_{i=1}^k \alpha_i$$

por lo tanto  $P\left(\bigcap_{i=1}^k C_i\right) \geq 1 - \sum_{i=1}^k \alpha_i$ .

- Entonces se puede buscar controlar las  $\alpha'_i$ s y el margen de significancia para las pruebas importantes. Suponiendo que las  $p$  medias valieran lo mismo, típicamente se puede hacer  $\alpha_i = \alpha/p$ : los intervalos *bonferronizados* se obtienen tomando el percentil  $\alpha/2p$  de una  $t_{n-1}$ .



# Método de Bonferroni para comparaciones múltiples III

- Los intervalos bonferronizados son de la forma

$$\bar{x}_i \pm t_{n-1, \alpha/(2p)} \sqrt{s_{ii}/n} \quad i = 1, \dots, p$$

## Ejemplo. [Función de cálculo]

Hacemos una función que haga los cálculos por nosotros:

```
int.bonferronizados <- function(X, alfa = 0.05){  
  # X es la matriz de catos, a es el vector combinación lineal  
  p <- dim(X)[2]; n <- dim(X)[1]  
  xbar <- colMeans(X)  
  S <- var(X)  
  lims <- list(NULL)  
  for(i in 1:p) lims[[i]] <- xbar[i] + c(-1,1)*abs(qt(1-alfa/(2*p),n-1))*sqrt(diag(S)[i]/n)  
  return(list(media = xbar, limites = lims))  
}
```



## Ejemplo. [Transpiración, cont.]

En el ejemplo de los datos de transpiración, los intervalos bonferranizados para cada componente del vector de medias están dados por:

# Método de Bonferroni para comparaciones múltiples IV

```
int.bonferronizados(X0)

$media
      X1      X2      X3
4.640 45.400  9.965

$limites
$limites[[1]]
[1] 3.643952 5.636048

$limites[[2]]
[1] 37.10308 53.69692

$limites[[3]]
[1] 8.846992 11.083008
```



- Los intervalos marginales son más precisos que los simultáneos para una elección específica de **a**.
- Los intervalos simultáneos consideran un nivel de confianza fijo de  $1 - \alpha$  independientemente del contraste considerado.
- Los intervalos bonferronizados representan un compromiso entre los dos casos anteriores.
- Usualmente se consideran los intervalos simultáneos de manera inicial, para hacer un *screening* de los datos, pero en caso de que se requiera más precisión sobre algún valor específico, se tiene que considerar los intervalos marginales.