

## Comparación de medias tomadas de a pares

CONDICION	Media	$n_i$	$s_i$
1	101.00	3	1.0000
2	102.00	3	1.7321
3	97.00	3	2.0000
4	93.00	2	1.4142
TOTAL	98.73	11	1.6036

Supongamos que hemos aplicado el test F y hemos rechazado la  $H_0$ .

¿Qué quiere decir la alternativa? Que no todas la medias son iguales pero, ¿cuáles son diferentes?

Cuando no se puede rechazar  $H_0$  generalmente el análisis termina ahí, pero cuando se rechaza generalmente el experimentador no se conforma con esa respuesta, sino que desea comparar las medias, frecuentemente de a pares, como para identificar cuáles son las que difieren.

Para ello deberíamos usar un método de comparaciones múltiples.

## Intervalo de confianza para la diferencia de dos medias.

Queremos comparar las medias de los grupos  $i$  y  $j$ .

**Empecemos por construir un IC para una sola diferencia  $\mu_i - \mu_j$**

El estimador puntual es  $\bar{X}_i - \bar{X}_j$

¿Cuál es su varianza? ¿Como se estima?

Puede demostrarse que

$$\left[ \bar{X}_i - \bar{X}_j - t_{n-k, \alpha/2} \sqrt{s_p^2 \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}; \bar{X}_i - \bar{X}_j + t_{n-k, \alpha/2} \sqrt{s_p^2 \left( \frac{1}{n_i} + \frac{1}{n_j} \right)} \right] \quad (4)$$

es un Intervalo de Confianza con nivel  $1-\alpha$ .

Si en vez de intervalo queremos estudiar la  $H_0: \mu_i = \mu_j$  también es fácil deducir un test bilateral por la relación entre ambos.

## ¿Se pueden calcular muchos IC o aplicar muchos tests?

¿Cuál es la crítica que se suele hacer a los IC “usando la distribución t” de la forma (4) y a los tests deducidos de estos intervalos?

Si, por ejemplo, tuviéramos 6 tratamientos y quisiéramos todas las comparaciones de a pares, el número de intervalos de confianza sería 15, si tuviéramos 4 tratamientos como en nuestro caso, sería 6 y si fueran 10, deberíamos realizar 45 comparaciones!!

¿Cuál será la probabilidad de que alguno no contenga al verdadero valor del parámetro?

Aunque no la podamos calcular exactamente, es evidente que esta probabilidad es mucho  $>$  que 0.05.

Por eso cuando uno planea de antemano hacer uno o muy pocos intervalos o tests puede usar (4), pero en caso contrario conviene utilizar un método de intervalos de confianza simultáneos.

**Veamos qué ocurre si tenemos dos parámetros:  $\theta_1$  y  $\theta_2$  y nos interesa preservar el nivel global  $1-\alpha$ .**

¿La probabilidad de que los dos intervalos sean correctos cuánto valdría si cada uno tiene nivel  $1-\alpha$ ?

$$P(\theta_1 \in [a_1(X), b_1(X)] \cap \theta_2 \in [a_2(X), b_2(X)]) = ?$$

$$\begin{aligned} P(\text{Error}) &= P(\theta_1 \notin [a_1(X), b_1(X)] \cup \theta_2 \notin [a_2(X), b_2(X)]) \\ &\leq P(\theta_1 \notin [a_1(X), b_1(X)]) + P(\theta_2 \notin [a_2(X), b_2(X)]) \\ &\leq 2\alpha \end{aligned}$$

$$P(\theta_1 \in [a_1(X), b_1(X)] \cap \theta_2 \in [a_2(X), b_2(X)]) \geq 1 - 2\alpha$$

Es decir, si  $\alpha=0.05$ , la cota es  $1-2\alpha=0.90$ .

Si en lugar de tener 2 parámetros tuviéramos  $m$ , la cota calculada sería

$$P\left(\bigcap \theta_i \in [a_i(X), b_i(X)]\right) \geq 1 - m\alpha$$

Por lo tanto, si  $m=6$  la cota es  $1-6*0.05=0.70$ , con lo cual el nivel baja mucho.

## Intervalos de confianza simultáneos (concepto general, no sólo para el análisis de varianza de un factor)

¿Cuál es la **definición** de IC para un parámetro  $\theta$ ?

Recordemos que si  $X=(X_1, X_2, \dots, X_n)$  es la muestra observada, un intervalo  $[a(X), b(X)]$  es un IC para  $\theta$  con nivel  $1-\alpha$  si

$$P( a(X) \leq \theta \leq b(X) ) = 1-\alpha$$

Ahora deseamos calcular IC para cada uno de los parámetros  $\theta_j$  (digamos  $j=1, \dots, m$ ). Se dice que el intervalo  $[a_j(X), b_j(Y)]$  es un IC para  $\theta_j$  calculado por un método simultáneo si

$$P \left( \bigcap_{j=1}^m [a_j(X) \leq \theta_j \leq b_j(X)] \right) \geq 1-\alpha$$

(5)

o sea que la probabilidad de que todos los IC sean correctos (contengan al verdadero valor del parámetro) es  $\geq 1-\alpha$ .

La probabilidad de que alguno sea incorrecto es  $\leq \alpha$ .

## Método de Bonferroni.

Un método muy general (para cualquier modelo) para obtener intervalos de confianza simultáneos es calcular cada uno de ellos con nivel  $1-\alpha/m$ , donde  $m$  es el número de IC que se desea calcular.

Este método tiene la ventaja de ser muy simple y muy general, pero sólo se usa en la práctica si  $m$  es pequeño, porque para valores moderados de  $m$  da IC de gran longitud.

Para el caso particular del análisis de la varianza de un factor, basta usar (4), pero reemplazando  $t_{n-k,\alpha/2}$  por  $t_{n-k,\alpha/2m}$  donde  $m$  es el número de IC que se desea calcular:

$$\left[ \bar{X}_i - \bar{X}_j - t_{n-k,\alpha/2m} \sqrt{s_p^2 \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}; \bar{X}_i - \bar{X}_j + t_{n-k,\alpha/2m} \sqrt{s_p^2 \left( \frac{1}{n_i} + \frac{1}{n_j} \right)} \right]$$

## Método de Tukey.

Los intervalos de Tukey son similares a los dados, pero reemplazando  $t_{n-k,\alpha/2}$  por el valor  $q_{k,n-k,\alpha} / \sqrt{2}$ , resultando

$$\bar{X}_i - \bar{X}_j \pm \frac{1}{\sqrt{2}} q_{k,n-k,\alpha} \sqrt{S_p^2 \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}$$

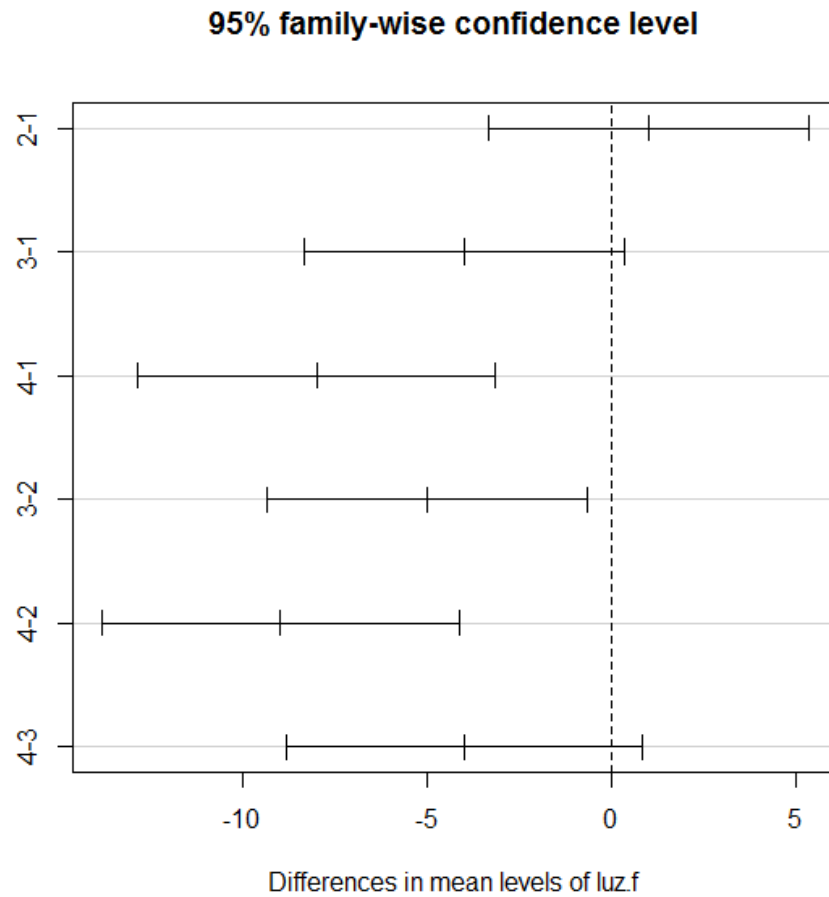
donde los valores "q" están tabulados y corresponden a la distribución estudiada por Tukey, llamada distribución del "**rango studentizado**" de k variables normales independientes. El  $\sqrt{2}$  que aparece se debe simplemente a como se construyó la tabla.

Se basa en la distribución de

$$Q = \max \left| \frac{\bar{X}_i - \bar{X}_j - (\mu_i - \mu_j)}{\sqrt{S_p^2 \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}} \right|$$



```
FLUOR.tuk<-TukeyHSD(salida,"luz.f",ordered=FALSE,conf.level=0.95)  
plot(FLUOR.tuk)
```



Para el caso originalmente pensado por Tukey en el que los tamaños de muestras son iguales ( $n_1=n_2=\dots=n_I$ ), este método hace que se cumpla el  $=$  en vez del  $\geq$  en (5) cuando se realizan todas las comparaciones de a pares.

El método de Tukey es óptimo (da IC de la menor longitud posible) cuando se desea calcular IC para todos los pares posibles y los  $n_j$ 's son iguales.

Para el caso en que los tamaños de muestras no son iguales, se demostró que sigue valiendo (5) pero con " $>$ ". En este caso el método se conoce también como "método de Tukey-Kramer".

**Tests simultáneos:** son los derivados de IC simultáneos. Tienen la propiedad de que la probabilidad de cometer algún error tipo I es menor o igual que  $\alpha$ .

## Comparación de los métodos considerados

Si se desea calcular un IC o aplicar un test para una sola diferencia de medias elegidas **a priori**, evidentemente el método de elección es el basado en la distribución t.

Si son unos pocos, elegidos a priori conviene usar Bonferroni. Si se hacen muchas comparaciones de a pares (o algunas elegidas a posteriori, que es "igual que hacer muchas") conviene usar Tukey pues da intervalos de menor longitud que Bonferroni.

Para elegir entre Bonferroni y Tukey, no es "trampa" elegir el método que da IC de menor longitud. No se necesita hacer las cuentas del IC para elegir el método: basta comparar quien es menor entre los valores de la tabla de "t" y de la tabla de "q", es decir entre  $t_{n-k, \alpha/2m}$  y  $q_{k, n-k, \alpha} / \sqrt{2}$ .

## Tabla de ANOVA

```
fluo<-read.table("C:\\Users\\Ana\\estadisticaQ\\2012\\ejemplo_anova.txt",header=T)
fluo
attach(fluo)
luz.f<- factor(LUZ)   ¿qué quiere decir esta instrucción?

salida<- aov(FLUOR~luz.f)
anova(salida)
```

## Analysis of Variance Table

Response: FLUOR

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
luz.f	3	122.18	40.727	5.838	0.001676 **
Residuals	7	18.00	2.571		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

```
names(salida)
```

```
[1] "coefficients" "residuals" "effects" "rank" "fitted.values"  
[6] "assign" "qr" "df.residual" "contrasts" "xlevels"  
[11] "call" "terms" "model"
```