

Análisis de la varianza de un factor

El **test t de 2 muestras** se aplica cuando se quieren comparar las medias de dos poblaciones con **distribuciones normales** con **varianzas iguales** y se observan **muestras independientes** para cada población. Ahora consideraremos una generalización para el caso en que se quieren comparar las medias de tres o más poblaciones.

Ejemplo: En la tabla siguiente se muestran los resultados obtenidos en una investigación acerca de la estabilidad de un reactivo fluorescente en diferentes condiciones de almacenamiento.

Se conservaron tres muestras en cada una de 4 condiciones. Supongamos (porque a veces puede ocurrir) que para una de las condiciones, la medición no pudo realizarse o se detectó una falla y fue eliminada. Los datos observados son:

Condiciones	Mediciones observadas (señales de fluorescencia)	Media Muestral
Recientemente preparada	102 100 101	101
Una hora en la oscuridad	101 101 104	102
Una hora con luz tenue	97 95 99	97
Una hora con luz brillante	92 94	93

Mirando los promedios muestrales se ven diferencias y nos preguntamos si las condiciones de almacenamiento no influyeron sobre la fluorescencia de las muestras (ésta será nuestra H_0), ¿cuál es la probabilidad de que por simple azar se observen diferencias de esta magnitud entre las medias muestrales?

Para generalizar el problema e introducir notación, podemos pensar que observamos k muestras (en el ejemplo $k=4$). Suponemos el siguiente modelo:

Modelo de k muestras normales independientes con varianzas iguales.

Muestra 1: $X_{11}, X_{12}, \dots, X_{1n_1}$ v. a. i.i.d $N(\mu_1, \sigma^2)$

.....

Muestra i : $X_{i1}, X_{i2}, \dots, X_{in_i}$ v. a. i.i.d $N(\mu_i, \sigma^2)$

.....

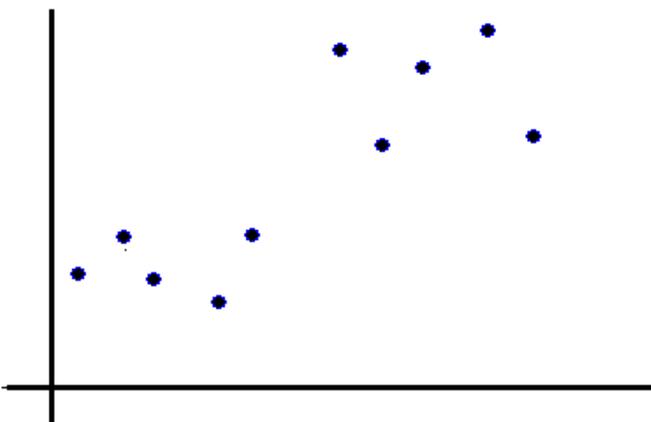
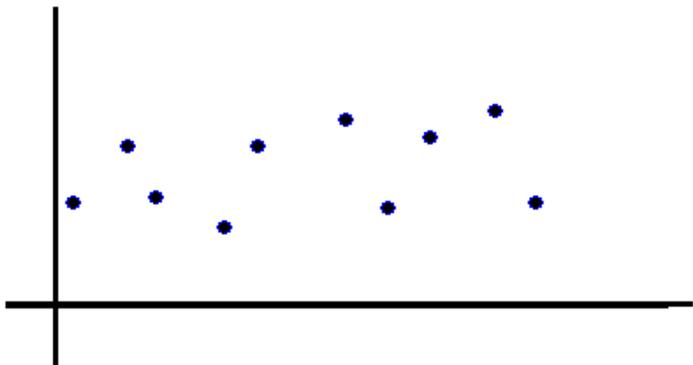
Muestra k : $X_{k1}, X_{k2}, \dots, X_{kn_k}$ v. a. i.i.d $N(\mu_k, \sigma^2)$

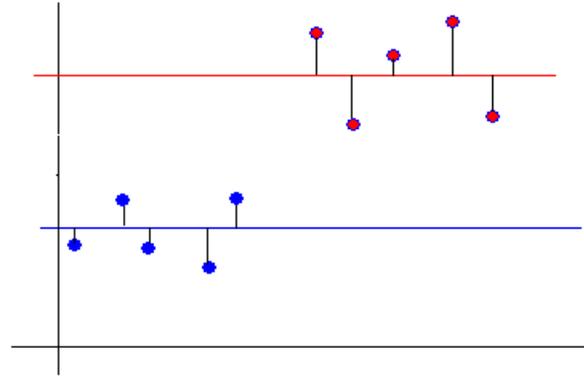
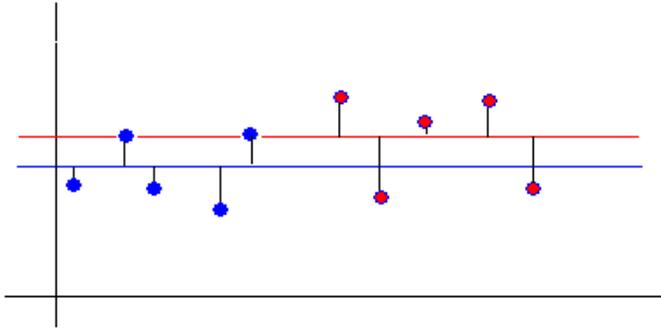
y asumimos que **las v. a. de una muestra son independientes de las v. a. de otra muestra.**

Llamaremos \bar{X}_i y S_i^2 a la **media y la varianza muestrales** de la muestra $i = 1, 2, \dots, k$.

Vamos a testear:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k \quad \text{vs.} \quad H_1: \text{existen } i \text{ y } j \text{ para los cuales } \mu_i \neq \mu_j$$





Parece natural proponer un **estimador de σ^2** basado en un **promedio ponderado de las varianzas de cada muestra S_i^2** , tal como se hacemos con el **S_p^2** cuando comparamos dos muestras. Se puede demostrar que el mejor estimador insesgado de σ^2 bajo el modelo anterior es:

$$S_p^2 = \frac{(n_1 - 1) * S_1^2 + \dots + (n_k - 1) * S_k^2}{n_1 + \dots + n_k - k} = \frac{\sum_{i=1}^k (n_i - 1) * S_i^2}{n - k} = \frac{SS_W}{n - k} \quad (1)$$

En la última expresión hemos llamado

$$n = \sum_{i=1}^k n_i$$

al **número total** de observaciones.

Bajo la hipótesis nula:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k$$

todas las observaciones tienen la misma distribución.

Llamemos

$$\bar{X} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij}}{n} = \frac{\sum_{i=1}^k n_i \bar{X}_i}{n}$$

a la [media general](#) de todas las observaciones.

El estadístico para el test óptimo para este problema, tiene al estimador de la varianza dado por (1) en el denominador y una medida de las diferencias (similar a la variancia) entre las medias de las distintas muestras en el numerador. Esta medida es:

$$\frac{\sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2}{k-1} = \frac{SS_B}{k-1} \quad (2)$$

El estadístico del test se obtiene dividiendo (2) sobre (1):

$$F = \frac{\left(\sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2 \right) / (k - 1)}{S_p^2} = \frac{SS_B / k - 1}{SS_W / n - k} \quad (3)$$

Si H_0 fuera cierta, el denominador y el numerador serían parecidos, por lo tanto el cociente sería cercano a 1.

Si las medias poblacionales no son todas iguales, como vimos en el gráfico, el numerador tiende a ser mayor que el denominador y por lo tanto, el cociente será mayor a 1.

Test F :

1er. paso: Cálculo el estadístico

$$F = \frac{\left(\sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2 \right) / (k-1)}{S_p^2} = \frac{SS_B / k - 1}{SS_W / n - k}$$

Nota: Si $H_0: \mu_1 = \mu_2 = \dots = \mu_k$ es cierta, este estadístico tiene distribución F con **$k-1$** grados de libertad en el **numerador** y **$n-k$** grados de libertad en el **denominador**.

¿De dónde surgen los grados de libertad? Se puede demostrar, que si se satisfacen los supuestos del Análisis de Varianza que hemos realizado, entonces:

Bajo H_0 :

$$\text{i) } \frac{SS_W}{\sigma^2} \sim \chi_{n-k}^2 \quad \text{ii) } \frac{SS_B}{\sigma^2} \sim \chi_{k-1}^2 \quad \text{y además son independientes.}$$

2do. paso: Si $F > F_{k-1, n-k; \alpha}$, rechazamos H_0 .

Con frecuencia los resultados del Análisis de Varianza se presentan una tabla como la que sigue:

Análisis de Varianza					
Fuente	SS	gl	MS	F	Prob > F
Between	SSB	k-1	MSB = SSB/k-1	MSB/MSW	
Within	SSW	n-k	MSW = SSW/n-k		
Total	SST	n-1	MST = SST/n-1		

Veamos como quedaría en nuestro ejemplo:

Fuente	gl	SS	MS	F	P
BETWEEN	3	122.182	40.7273	15.84	0.0017
WITHIN	7	18.000	2.57143		
TOTAL	10	140.182			

Notemos que el p-valor es 0.0017, por lo tanto rechazamos a los niveles habituales.

En particular, rechazamos la hipótesis $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ al nivel 0.01, es decir las medias de la fluorescencia difieren significativamente a este nivel. O dicho de otro modo: concluimos que la media de la fluorescencia depende de las condiciones de almacenamiento.

La pregunta ahora es: ¿cuáles son las que difieren?

Comentarios sobre la “tabla del análisis de la varianza”.

Se puede demostrar que vale la siguiente igualdad:

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2 = \sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$$

En la expresión anterior aparecen tres “sumas de cuadrados”:

“suma de cuadrados entre grupos” (**SS_B: Between**)

“suma de cuadrados dentro de grupos” (**SS_W: Within**)

“suma de cuadrados total” (**SS_T: Total**)

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	15.838	0.001676 **
Residuals	7	18.00	2.571		

Suposiciones del modelo. Diagnóstico.

El test F ha sido deducido bajo el supuesto de que las k muestras aleatorias tienen distribución normal, con igual varianza y son independientes. Cuando el tamaño de la muestra de cada grupo es grande, el test F es válido en forma aproximada (el valor p calculado es aproximado) aunque la variable no tenga distribución exactamente normal.

En la práctica no es esperable que el modelo se cumpla exactamente, pero sí en forma aproximada. Al igual que con el test t , hay que analizar los datos para detectar si el modelo es aproximadamente cierto o si en cambio es falso.

Boxplots Paralelos

Cuando hay una cantidad suficiente de observaciones se pueden realizar boxplots paralelos de las observaciones originales por tratamiento.

En el presente ejemplo, hay solo 3 y hasta 2 observaciones por casilla, con lo cual no parece muy razonable este gráfico. En su lugar podemos realizar un boxplot de los residuos todos juntos.

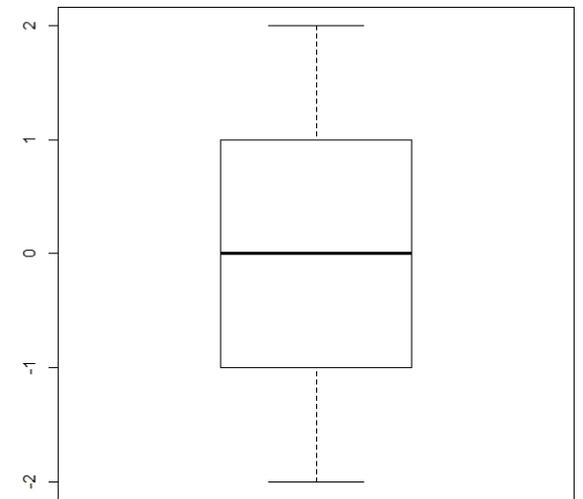
Para cada observación, el residuo r_{ij} se calcula como:

$$r_{ij} = X_{ij} - \bar{X}_i$$

El siguiente gráfico muestra el boxplot correspondiente a los residuos del ejemplo de fluorescencia:

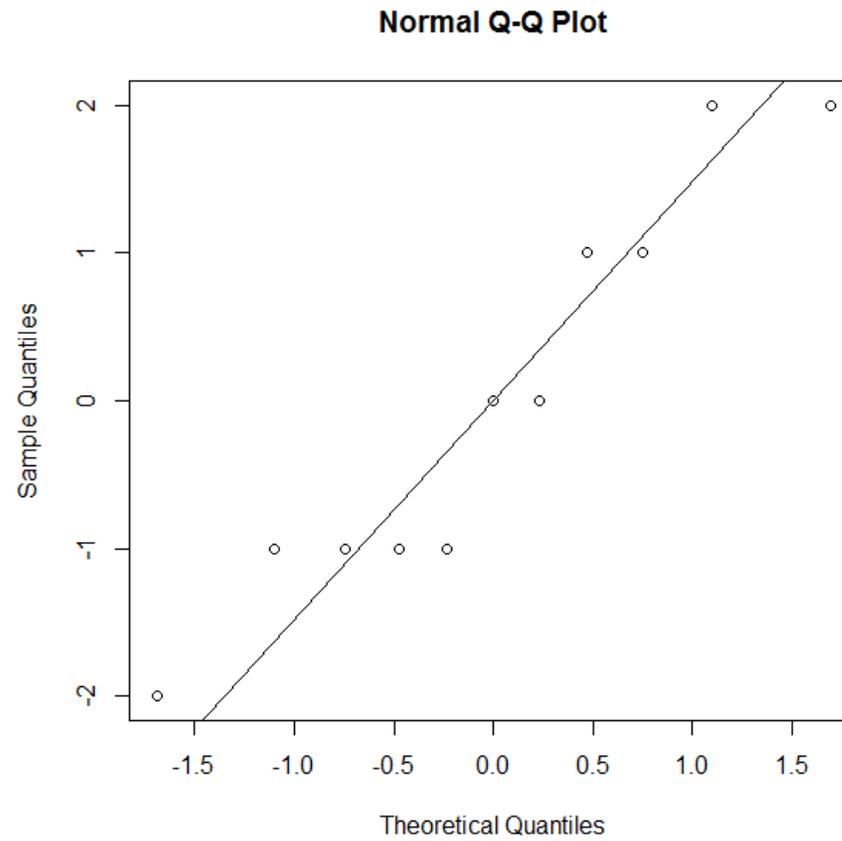
```
boxplot(salida$res)
```

Los residuos parecen tener una distribución simétrica y no se observan datos atípicos, por lo que no parece haber importantes apartamientos de la normalidad.



QQ-plot y Test de Shapiro-Wilk en nuestro ejemplo

```
qqnorm(salida$res)  
qqline(salida$res)
```



```
shapiro.test(salida$res)
```

Shapiro-Wilk normality test

```
data: salida$res  
W = 0.9081, p-value = 0.2315
```

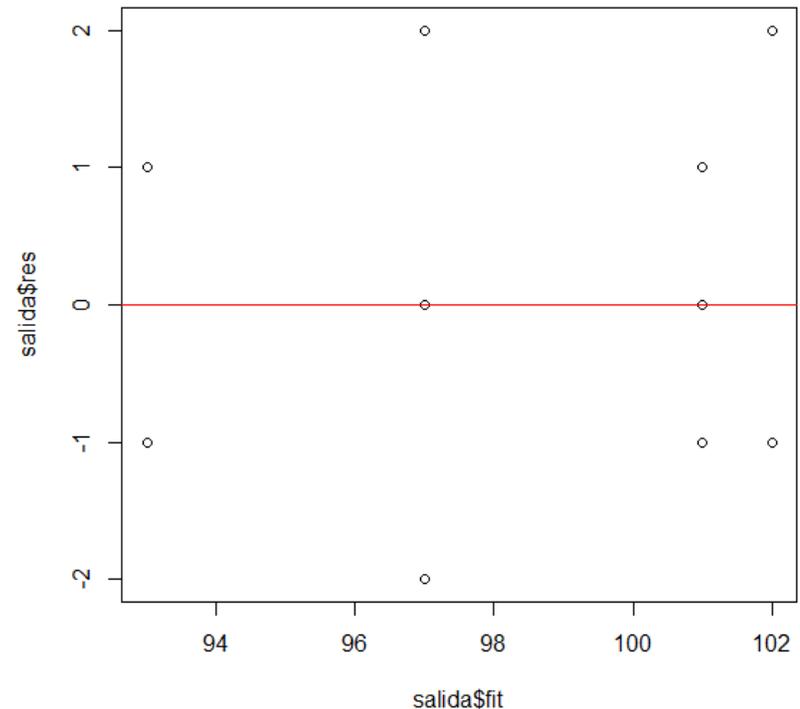
En nuestro ejemplo el estadístico del test de Shapiro-Wilk es 0.9081 y el p-valor correspondiente es de 0.2315, **con lo cual no rechazamos el supuesto de normalidad.**

Tests para estudiar si las varianzas son iguales

Para estudiar la suposición de igualdad de varianzas podemos graficar y también se pueden realizar algunos tests.

Respecto del gráfico podemos considerar un scatter-plot o diagrama de dispersión de los promedios muestrales versus los residuos. En el ejemplo de Fluorescencia resultaría:

Se observan algunas diferencias en la dispersión de los residuos, pero no parece haber grandes apartamientos del supuesto de homoscedasticidad en este caso. Sin embargo, deberíamos aplicar un test para chequear este supuesto.



Respecto de tests existen algunas alternativas.

Consideremos el modelo

$$X_{ij} \sim N(\mu_i, \sigma_i^2) \quad (i=1, \dots, k; j=1, \dots, n_i) \text{ independientes}$$

y la hipótesis que nos interesa testear es

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$$

Hay varios tests. El más antiguo es el **test de Bartlett**. Se basa en un estadístico que tiene distribución aproximadamente χ^2_{k-1} bajo H_0 .

Si hay k muestras con tamaño n_i y varianzas de las muestras S_i^2 , como en nuestro problema, entonces estadístico de prueba de Bartlett, que se basa en una escala logarítmica, es:

$$X^2 = \frac{(n - k) \ln(S_p^2) - \sum_{i=1}^k (n_i - 1) \ln(S_i^2)}{1 + \frac{1}{3(k-1)} \left(\sum_{i=1}^k \left(\frac{1}{n_i - 1} \right) - \frac{1}{n - k} \right)}$$

El numerador tiende a dar valores grandes cuando las varianzas muestrales difieren mucho, por lo tanto se rechaza la hipótesis nula de igualdad de varianzas cuando el estadístico es grande.

La zona de rechazo es

$$X^2 > \chi_{k-1, \alpha}^2$$

```
bartlett.test(FLUOR,luz.f)
```

Bartlett test of homogeneity of variances

```
data:  FLUOR and luz.f
```

```
Bartlett's K-squared = 0.7515, df = 3, p-value = 0.861
```

En nuestro ejemplo el estadístico del test de Bartlett es 0.7515 con un p-valor de 0.861, por lo tanto **no rechazamos el supuesto de homogeneidad de varianzas**

Sin embargo, este test tiene una alta sensibilidad a la falta de normalidad. Por esta razón, es necesario disponer de alguna alternativa más resistente a la falta de normalidad.

Un test que es poco sensible a la falta de normalidad es el **test de Modificado de Levene**. Para aplicarlo, primero se calculan

$$d_{ij} = | X_{ij} - \tilde{X}_i |$$

donde \tilde{X}_i denota la mediana del tratamiento i . Luego se calcula el estadístico F del análisis de un factor a los d_{ij} .

Si la hipótesis $H: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$ es cierta y los n_i “no son muy pequeños”, el estadístico tiene distribución aproximadamente F con $k-1$ y $n-k$ grados de libertad. Esto permite aplicar un test aproximado de la hipótesis de igualdad de varianzas. Rechazamos la igualdad de varianzas si el estadístico toma un valor muy grande. Se puede hacer a mano:

```
medians<-tapply(FLUOR,luz.f,median)
abs.dif<- abs(FLUOR-mediands[luz.f])
summary(aov(abs.dif~luz.f))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
luz.f	3	0.667	0.2222	0.167	0.916
Residuals	7	9.333	1.3333		

Como el p-valor = 0.916, no rechazamos el supuesto de homoscedasticidad.

También se puede hacer con

```
library(lawstat)  
levene.test(FLUOR,luz.f)
```

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 las 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? ¿Cómo 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.

En nuestro ejemplo, el intervalo de confianza de nivel 0.95 para $\mu_1 - \mu_2$ basado en la muestra sería

$$\left[\bar{X}_1 - \bar{X}_2 - t_{7,0.025} \sqrt{s_p^2 \left(\frac{1}{3} + \frac{1}{3} \right)}; \bar{X}_1 - \bar{X}_2 + t_{7,0.025} \sqrt{s_p^2 \left(\frac{1}{3} + \frac{1}{3} \right)} \right]$$

¿Cómo lo calculamos efectivamente con el R?

```
#percentil bajo la t con 7 grados de libertad
qt(0.975,7)
[1] 2.364624
```

```
# 2.571 es el valor de  $s_p^2$  obtenido con esta muestra y corresponde al valor que se
encuentra en Mean Sq para Residuals en la salida de ANOVA del R
```

```
# semilongitud del intervalo de confianza de nivel individual 0.05
2.364624*sqrt(2.571)*sqrt(2/3)
[1] 3.095
```

```
# límite superior del intervalo de confianza de nivel individual 0.05
101-102+2.364624*sqrt(2.571)*sqrt(2/3)
[1] 2.095762
```

```
# límite inferior del intervalo de confianza de nivel individual 0.05
101-102-2.364624*sqrt(2.571)*sqrt(2/3)
[1] -4.095762
```

¿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: θ_1 y θ_2 y nos interesa preservar el nivel global $1-\alpha$. En nuestro caso podrían ser, por ejemplo, $\theta_1=\mu_1 - \mu_2$ y $\theta_2=\mu_3 - \mu_4$

¿La probabilidad de que los dos intervalos sean correctos cuánto valdría si cada uno tiene nivel $1-\alpha^*$? Denotemos \vec{X} (ojo! es una flecha arriba) a la muestra

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

Calculemos la probabilidad de su complemento:

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

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

Es decir, si $\alpha^*=0.05$, la cota inferior es $1-2\alpha^*=0.90$. Es decir que pasamos de intervalos con un nivel individual de 0.95 a intervalos con nivel simultáneo más bajo, ya que solo podemos garantizar que el nivel global sea mayor o igual a 0.90.

Si en lugar de tener 2 parámetros tuviéramos m , la cota calculada sería más baja aún y estaría dada por

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

Por lo tanto , si $m=6$ la cota es $1-6*0.05=0.70$, con lo cual el nivel global puede bajar 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 θ ?

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

$$P\left(\theta \in [a(\vec{X}), b(\vec{X})]\right) = P\left(a(\vec{X}) \leq \theta \leq b(\vec{X})\right) \geq 1 - \alpha$$

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

$$P\left(\bigcap_{j=1}^m (a_j(\vec{X}) \leq \theta_j \leq b_j(\vec{X}))\right) \geq 1 - \alpha \quad (5)$$

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. Vimos que si tuviéramos m parámetros y para cada uno de ellos un intervalo de nivel α^* , tendríamos

$$P\left(\bigcap \theta_i \in [a_i(\vec{X}), b_i(\vec{X})]\right) \geq 1 - m\alpha^* \quad (6)$$

Luego, si quisiéramos, por ejemplo, que la cota inferior fuera 0.95 (=1-0.05), tendríamos que

$$1 - m\alpha^* = 0.95 \Rightarrow m\alpha^* = 0.05 \Rightarrow \alpha^* = 0.05 / m$$

Es decir, que cada intervalo debería tener nivel $1 - \alpha^* = 1 - \frac{0.05}{m}$ para lograr

que la cota inferior en (6) sea 0.95. En general, si deseamos una cota inferior igual a $1-\alpha$ con este criterio, cada intervalo debería tener nivel $1-\alpha/m$ (es decir $\alpha^*=\alpha/m$). Esto da origen al método de Bonferroni.

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 para que cada intervalo individual tenga nivel $1-\alpha/m$ debemos reemplazar $t_{n-k,\alpha/2}$ por $t_{n-k,\alpha/2m}$, donde m es el número de IC que se desea calcular:

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

En nuestro ejemplo, el intervalo de confianza de nivel simultáneo 0.95 para $\mu_1 - \mu_2$ basado en la muestra sería

$$\left[\bar{X}_1 - \bar{X}_2 - t_{7, \frac{0.05}{12}} \sqrt{s_p^2 \left(\frac{1}{3} + \frac{1}{3} \right)}; \bar{X}_1 - \bar{X}_2 + t_{7, \frac{0.05}{12}} \sqrt{s_p^2 \left(\frac{1}{3} + \frac{1}{3} \right)} \right]$$

¿Cómo lo calculamos efectivamente con el R?

```
#percentil bajo la t con 7 grados de libertad Cada intervalo tiene nivel individual
1-0.05/6
qt(1-(0.05/12),7)
[1] 3.635807
```

```
# 2.571 es el valor de  $s_p^2$  obtenido con esta muestra y corresponde al valor que se
encuentra en Mean Sq para Residuals en la salida de ANOVA del R
```

```
# semilongitud del intervalo de confianza de nivel individual 1-0.05/6=0.9917
3.635807*sqrt(2.571)*sqrt(2/3)
[1] 4.759992
```

```
# límite superior del intervalo de confianza de nivel individual 0.05
101-102+2.364624*sqrt(2.571)*sqrt(2/3)
[1] 3.759992
```

```
# límite inferior del intervalo de confianza de nivel individual 0.05
101-102-2.364624*sqrt(2.571)*sqrt(2/3)
[1] -5.759992
```

¿Cómo es la longitud de este intervalo respecto del que obtuvimos antes? ¿Qué nivel individual tiene cada uno?

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

```
fluo  
  FLUOR LUZ  
1    102   1  
2    100   1  
3    101   1  
4    101   2  
5    101   2  
6    104   2  
7     97   3  
8     95   3  
9     99   3  
10    92   4  
11    94   4
```

```
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	15.838	0.001676 **
Residuals	7	18.00	2.571		

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

```
bb<-pairw.anova(y = FLUOR, x = luz.f, method = "bonf")
plot(bb,type="2")
plot(bb,type="1")
```

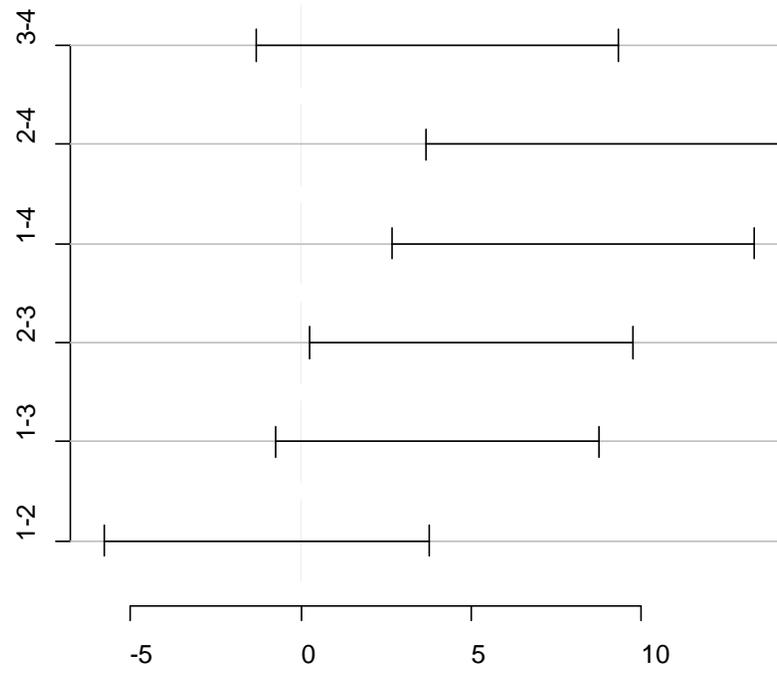
```
library(asbio)
bb<-pairw.anova(y = FLUOR, x = luz.f, method = "bonf")
plot(bb,type="2")
```

```
> bb
```

95% Bonferroni confidence intervals

	Diff	Lower	Upper	Decision	Adj. p-value
mu1-mu2	-1	-5.76039	3.76039	FTR H0	1
mu1-mu3	4	-0.76039	8.76039	FTR H0	0.110709
mu2-mu3	5	0.23961	9.76039	Reject H0	0.039314
mu1-mu4	8	2.67772	13.32228	Reject H0	0.005645
mu2-mu4	9	3.67772	14.32228	Reject H0	0.00281
mu3-mu4	4	-1.32228	9.32228	FTR H0	0.175393

95% Bonferroni confidence i



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

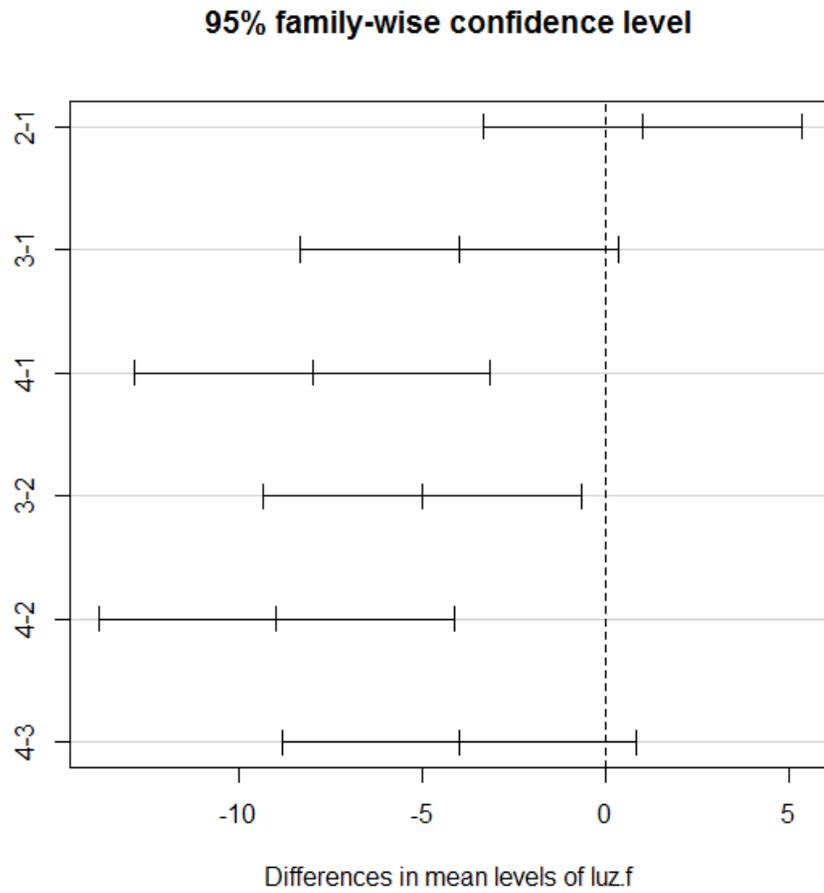
$$\bar{X}_i - \bar{X}_j \pm \frac{1}{\sqrt{2}} q_{k,n-k,\alpha} \sqrt{S_p^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

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 α .

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"
```