

# FEIR 30: Comparación paramétrica de medias

Apuntes del curso FEIR3, curso 2014/15 actualizados. Última actualización: lunes 06 febrero 2017, 11:18:45

*Aurora González Vidal*

## Índice

<b>1. Comparaciones entre dos grupos</b>	<b>2</b>
1.1. Introducción . . . . .	2
1.2. $t$ -test . . . . .	2
1.2.1. $t$ -test dependiente ( <i>paired</i> ) . . . . .	3
1.2.2. $t$ -test independiente ( <i>unpaired</i> ) . . . . .	5
1.2.3. $t$ -test para una muestra . . . . .	6
1.3. Tamaño del efecto: . . . . .	6
<b>2. Comparaciones entre dos o más grupos</b>	<b>11</b>
2.1. ANOVA . . . . .	11
2.1.1. ANOVA de una vía (entre sujetos) . . . . .	12
2.1.2. El estadístico F. Valor de F (F value) . . . . .	16
2.1.3. Cómo comunicar los resultados de un ANOVA . . . . .	16
2.1.4. ANOVA de dos factores (de dos vías) . . . . .	16
2.1.5. Tipos de suma de cuadrados . . . . .	20
2.1.6. Efectos fijos vs efectos aleatorios . . . . .	20
2.1.7. ANOVA con medidas repetidas (within factors) . . . . .	23
2.1.8. Modelos mixtos ANOVA . . . . .	26
2.1.9. ANOVA y regresión lineal . . . . .	28
2.2. Después del ANOVA . . . . .	30
2.2.1. Contrastes <i>post hoc</i> o no planificados . . . . .	30
2.2.2. Comparaciones planificadas . . . . .	32
<b>Referencias y bibliografía</b>	<b>34</b>



# 1. Comparaciones entre dos grupos

## 1.1. Introducción

En los análisis estadísticos de datos procedentes de investigación es muy común que nos encontremos con la necesidad de comparar dos o más grupos.

Un investigador se puede encontrar con la necesidad de comparar la longitud del rabo de 3 tipos de ratones de laboratorio cuando son alimentados con dos dietas diferentes.

Otro, sin embargo, puede estar interesado en comparar la aceptación de la homosexualidad dependiendo de la edad: niños, adultos y ancianos.

Son muy diversas las áreas que necesitan analizar datos clasificados por grupos y los modelos que se pueden aplicar para resolver estas cuestiones serán el objeto de estudio.

### Test paramétricos frente a tests no paramétricos

Una de las las categorizaciones más importantes en estadística consiste en clasificar los problemas en paramétricos y no paramétricos. Paramétrico significa que *con unos pocos parámetros podemos describir bien la distribución* de la variable. Normalmente, para poder hacer esto hemos de suponer ciertas propiedades de la distribución: media, desviación típica...; y, con estos parámetros, debemos ser capaces de asumir que la distribución de la variable es de una determinada manera, generalmente una distribución **normal**. Hay otras condiciones iniciales para los test paramétricos (por ejemplo, homocedasticidad) que iremos discutiendo más adelante.

Esta condición de normalidad, asumir que la población de la que provienen los datos es normal, no implica necesariamente que los datos, la muestra, deba parecerse a una normal: sólo que podamos asumir que provienen de una población normal. Emplearemos tests como el test de Shapiro-Wilk, Kolmogorov-Smirnov..., analizaremos histogramas, Q-Q plots etc.

¿Cuándo emplear tests no paramétricos en lugar de los paramétricos?

- Con datos ordinales en general empleamos test no paramétricos: por ejemplo datos procedentes de escalas Likert (grado de satisfacción del 1 al 5, etc...)
- Cuando la normalidad de la población no pueda ser asumida.

Afortunadamente, muchos de los test paramétricos son bastante robustos a la falta de normalidad. También podemos optar por emplear transformaciones de datos para “alcanzar” la normalidad.

## 1.2. *t*-test

El más común de los test para comparar medias de dos grupos es el ***t*-test**.

Para poder emplear un *t*-test hemos de comprobar ciertos supuestos primero. Lo más importante es que las distribuciones de las poblaciones de las cuales muestreamos han de ser normales. Decimos distribución de las poblaciones, no de las muestras, que siempre hemos de suponerlas procedentes de un muestreo aleatorio simple. Si las muestras a comparar son suficientemente grandes podemos invocar el Teorema Central del Límite (TCL) y asumir la normalidad, pero en caso contrario, sí es importante comprobar la normalidad con muestras pequeñas donde el TCL no puede ser aplicado.

**Hay tres tipos de *t*-test:** *t*-test dependiente, *t*-test independiente y *t*-test para una muestra.

**A la hora de decidir cuál usar hay que tener en cuenta el diseño experimental que tengamos**, ya que no podemos emplear uno u otro a nuestro antojo.

Vamos a tomar estos datos simulados para todos los experimentos:



```
medida <- c( 3, 1, 2, 3, 1, 6, 2, 4, 1, 2, 6, 5, 1, 3, 5, 4, 2, 3, 4, 5 )
grupo  <- c( 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1 )
df     <- data.frame( grupo, medida )
head( df )

##      grupo medida
## 1      0      3
## 2      0      1
## 3      0      2
## 4      0      3
## 5      0      1
## 6      0      6
```

### 1.2.1. *t*-test dependiente (*paired*)

#### 1.2.1.1. Introducción

Se utiliza cuando tratamos de evaluar diferencias de medias entre dos grupos que están relacionados. Es decir, los mismos sujetos (personas, animales...) son medidos en dos condiciones diferentes, como por ejemplo de tiempo: antes y después. En este caso, primero calculamos la diferencia entre un grupo y otro (antes y después) y hacemos un ***t*-test para una muestra** (one-sample *t*-test). Estos test se denominan también ***t*-test para datos pareados**, diseño con factores intra sujetos (*within-subject design*) y diseños con medidas repetidas.

Un diseño apareado (*paired*) considera las diferencias individuales de cada observación, cuando hace la diferencia entre los dos grupos experimentales. Así que tiene ciertas ventajas sobre el *test independiente* que a continuación estudiaremos.

Si entendemos que la *variación sistemática* es la producida por el experimento (es decir, las diferencias entre grupos producidas por el hecho de ser sometidos a experimentos diferentes) y la *variación no-sistemática* es la debida a factores aleatorios que pueden existir entre las distintas condiciones experimentales (por ejemplo: personas que sin saber por qué tienen diferentes tolerancias a ciertas drogas). Así pues, con los diseños pareados :

- Se controla mejor la variación no sistemática porque se mide varias veces el mismo sujeto.
- Es más fácil descubrir los efectos de nuestra manipulación experimental.
- En diseños de medidas repetidas el ruido se mantiene en el mínimo.
- Los diseños de medidas repetidas tienen más potencia estadística, detectan diferencias que existen realmente más fácilmente que los diseños independientes.

#### 1.2.1.2. Supuestos

El *t*-test dependiente no nos exige homocedasticidad pero sí normalidad. Lo comprobamos para cada grupo. Para seleccionar unicamente las medidas del grupo 0 hacemos `df[df$grupo==0,2]`.

```
df[ df$grupo == 0, 2 ]
## [1] 3 1 2 3 1 6 2 4 1 2
```

El 2 es porque cogemos la segunda columna del conjunto de datos `df`. Al escribir el número 1 se selecciona la primera, que no es otra cosa que ceros

```
df[ df$grupo == 0, 1 ]
## [1] 0 0 0 0 0 0 0 0 0 0
```

El análisis:



```
shapiro.test( df[ df$grupo == 0, 2 ] )

##
##  Shapiro-Wilk normality test
##
## data:  df[df$grupo == 0, 2]
## W = 0.9, p-value = 0.08

shapiro.test( df[ df$grupo == 1, 2 ] )

##
##  Shapiro-Wilk normality test
##
## data:  df[df$grupo == 1, 2]
## W = 0.9, p-value = 0.7
```

Ambos p-valores 0.083 y 0.668 son mayores que 0.05 luego podemos aceptar la hipótesis nula, que es la de normalidad.

### 1.2.1.3. Hipótesis a contrastar

La hipótesis nula ( $H_0$ ) es “no hay diferencia entre las medias de dos poblaciones de las cuales tenemos dos muestras” y la alternativa “sí la hay”. Así, si fijamos la significación en 0.05, rechazaremos esta hipótesis nula si  $p < 0.05$  y concluiremos que aceptamos la hipótesis alternativa de que existen diferencias.

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Ejecutamos el  $t$ -test en R con la función `t.test()`.

Los dos primeros argumentos son los grupos a comparar y escribimos `paired=TRUE` porque son dependientes.

```
fitDep <- t.test( df[ df$grupo == 0, 2 ], df[ df$grupo == 1, 2 ], alternative= "two.sided",
  conf.level = 0.95, paired = TRUE )
```

El p-valor es 0,109 > 0,05 luego se acepta la hipótesis nula. El intervalo de confianza nos indica que la diferencia de medias está en el intervalo  $(-2,95, 0,35)$  con una confianza del 95 %.

**Nota:** la opción `alternative= "two.sided"` es la que viene por defecto e igualmente por defecto se calculan los intervalos de confianza de nivel 0.95 así que escribir

```
t.test( df[ df$grupo == 0, 2 ], df[ df$grupo == 1, 2 ], paired = TRUE )
```

será equivalente en este caso particular.

Cuando se desea contrastar si la media de un grupo es mayor que la media del otro, la hipótesis alternativa ( $H_1$ ) es: “la media del primer grupo es mayor que la del segundo” y la hipótesis nula es: “la media del primero es menor o igual que la del segundo”. Es decir,

$$H_0 : \mu_1 \leq \mu_2$$

$$H_1 : \mu_1 > \mu_2$$

En R se hará `t.test( x, y, alternative="greater", paired=TRUE )`.

Cuando, por el contrario, se desea contrastar si la media de un grupo es menor que la media del otro, la hipótesis alternativa ( $H_1$ ) es: “la media del primer grupo es menor que la del segundo” y la hipótesis nula es: “la media del primero es mayor o igual que la del segundo”. Es decir,

$$H_0 : \mu_1 \geq \mu_2$$

$$H_1 : \mu_1 < \mu_2$$

En R se hará `t.test( x,y,alternative="less",paired=TRUE )`.



### 1.2.2. *t*-test independiente (*unpaired*)

#### 1.2.2.1. Introducción

Se utiliza cuando los sujetos que pertenecen a cada grupo son distintos, es decir, a cada nivel del experimento son sometidos diferentes sujetos. En este caso se construye un modelo para cada grupo, se calcula su media y varianza y se ve cuándo hay o no hay diferencias entre los dos modelos mediante un *t*-test para muestra independientes.

#### 1.2.2.2. Supuestos

Este test, además de la hipótesis de normalidad, nos exige *homocedasticidad* (homogeneidad de las varianzas de las dos muestras a comparar).

Para comprobar si el modelo es homocedástico podemos utilizar un test de varianza por ejemplo `var.test()`.

```
var.test( df[ df$grupo == 0, 2 ], df[ df$grupo == 1, 2 ] )

##
## F test to compare two variances
##
## data: df[df$grupo == 0, 2] and df[df$grupo == 1, 2]
## F = 1, num df = 9, denom df = 9, p-value = 1
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.259 4.194
## sample estimates:
## ratio of variances
##                1.04
```

El p-valor es  $0.95 > 0.05$  luego se acepta la hipótesis nula, esto es, la igualdad de varianzas.

#### 1.2.2.3. Hipótesis a contrastar

Como antes, pueden ser

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2.$$

Así, si fijamos la significación en 0.05, rechazaremos esta hipótesis nula si  $p < 0.05$  y concluiremos que aceptamos la hipótesis alternativa de que *existen diferencias*.

Ejecutamos *t*-test en R con la función `t.test()` y elegiremos `var.equal=FALSE` o `var.equal= TRUE` dependiendo del resultado del test de homocedasticidad.

```
fitIndep <- t.test( df[ df$grupo == 0, 2 ], df[ df$grupo == 1, 2 ], var.equal = TRUE )
```

Como  $p = 0.079 > 0.05$  no hay una diferencia significativa.

De nuevo, las otras hipótesis pueden ser

$$H_0 : \mu_1 \leq \mu_2$$

$$H_1 : \mu_1 > \mu_2$$

se hará `t.test( x, y, alternative="greater" )` o

$$H_0 : \mu_1 \geq \mu_2$$

$$H_1 : \mu_1 < \mu_2$$

se hará `t.test( x, y, alternative="less" )`.



### 1.2.3. *t*-test para una muestra

Este es el caso en que vamos a contrastar la media de un conjunto de observaciones que pertenecen a un mismo grupo.

#### 1.2.3.1. Supuestos: normalidad

```
shapiro.test( df[ , 2 ] )

##
##  Shapiro-Wilk normality test
##
## data:  df[, 2]
## W = 0.9, p-value = 0.1
```

Aceptamos la hipótesis de normalidad ( $p=0.09553 > 0.05$  ).

#### 1.2.3.2. Hipótesis a contrastar

$$\begin{array}{lll} H_0 : \mu = \mu_0 & H_0 : \mu \leq \mu_0 & H_0 : \mu \geq \mu_0 \\ H_1 : \mu \neq \mu_0 & H_1 : \mu > \mu_0 & H_1 : \mu < \mu_0 \end{array}$$

```
t.test( df[ , 2 ] )

##
##  One Sample t-test
##
## data:  df[, 2]
## t = 8, df = 20, p-value = 7e-08
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  2.37 3.93
## sample estimates:
## mean of x
##      3.15
```

Es equivalente a

```
t.test( df[ , 2 ], mu = 0, conf.level = 0.95 )
```

### 1.3. Tamaño del efecto:

La idea básica que hemos seguido para obtener resultados se trata de generar una hipótesis nula y una hipótesis alternativa, encontrar un modelo estadístico para nuestros datos y evaluar este modelo con un test estadístico. Si la probabilidad de obtener nuestro test estadístico por azar es menor que 0.05 entonces aceptamos la hipótesis alternativa (rechazamos la hipótesis nula) y decimos normalmente que hay un *efecto significativo*, es decir, que los grupos reaccionan de forma diferente con respecto a los diferentes factores del experimento. **El experimento produce un efecto.**

Pero no nos engañemos; el hecho de que la probabilidad de que nuestro efecto sea azaroso sea pequeña (menor que 0.05) no implica que el efecto es importante.

Es decir, hemos encontrado que hay menos de un 5 % (el p valor) de probabilidad de que el efecto que estamos observando provenga del azar y hemos establecido que eso significa que no proviene del azar. Lo que pasa es que el p-valor depende del tamaño muestral y efectos muy pequeños e insignificantes pueden llegar a ser estadísticamente significantes porque se usan muchos sujetos en el experimento. Por esto, se necesita medir

de una forma estandarizada e independiente del tamaño muestral el *tamaño del efecto* (effect size o “es”). Es estandarizada para poder comparar los efectos de diferentes estudios que miden diferentes variables y que usan diversas escalas.

El tamaño del efecto puede ser medido según diferentes *métricas*. Las más usuales son la  $d$  de *Cohen*, la  $r$  de *Pearson* o *coeficiente de correlación de Pearson*, y el coeficiente *eta-cuadrado*.

Además deberíamos de calcular intervalos de confianza para esos tamaños de efectos, ya que su cálculo es simplemente un *estimador* del tamaño del efecto.

Una diferencia con respecto a los IC de medias es que estos IC pueden ser asimétricos (así el *effect size* no es la media de los extremos como ocurre con los IC de la *media*). Si este IC contiene al “0” concluiremos que el tratamiento no tiene ningún efecto, es decir, que no había diferencia significativa y la hipótesis nula no se rechaza, sino que se acepta.

(Lectura recomendada: Antonio Valera Espín, n.d.)

El IC no contiene el 0:

abs(es)	tamaño del IC	Lectura
pequeño	pequeño	El efecto aparentemente existe, pero estamos seguros de que es pequeño.
pequeño	grande	Sumamente improbable que ocurra esto.
grande	pequeño	El efecto aparentemente existe, y estamos seguros de que es grande.
grande	grande	El efecto aparentemente existe y puede ser grande, pero no estamos seguros de su tamaño.

El IC contiene el 0:

abs(es)	tamaño del IC	Lectura
pequeño	pequeño	No estamos seguros de que exista un efecto, en caso de existir es pequeño.
pequeño	grande	No estamos seguros de que exista un efecto. Necesitamos más datos (muestra).

abs(es)	tamaño del IC	Lectura
grande	pequeño	Sumamente improbable que ocurra esto.
grande	grande	No estamos seguros de que existan un efecto. Necesitamos más datos (muestra).

Un detalle es que el intervalo de confianza (IC) depende del tamaño muestral, así que aumentando el tamaño muestral podemos hacer mas estrecho el IC y llegar a conclusiones más fuertes. Pero eso sí, incrementar la muestra no contribuye a aumentar el tamaño del efecto (al contrario que ocurre con los p-valores). [?]

### Tamaño del efecto para un $t$ -test

Hay dos métricas disponibles para el tamaño del efecto (“es” de *effect size*) para un  $t$ -test: la  $d$  de Cohen (*Cohen’s d*), y la  $r$  de Pearson. Ambas son equivalentes y puedes pasar de  $d$  a  $r$  y viceversa. Aunque no podemos emplear la  $r$  de Pearson para los test dependientes, en este caso solo Cohen.

Dependiendo del área de conocimiento se consideran ciertos valores de  $d$  y  $r$  como efectos pequeños, medianos o grandes pero la sugerencia general suele ser la siguiente:

Efecto	Pequeño	Mediano	Grande
$d$ de Cohen	0.2	0.5	0.8
$r$ de Pearson	0.1	0.3	0.5

Efecto	$d$ de Cohen	$r$ de Pearson
$t$ - test dependiente ( <i>paired</i> )	$d = \frac{ M }{SD}$	No se puede emplear
$t$ - test independiente ( <i>unpaired</i> )	$d = \frac{ \mu_1 - \mu_2 }{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}}$	$r = \sqrt{\frac{t^2}{t^2 + gl}}$

**Nota:**  $M$  es la media de las diferencias y  $SD$  la desviación estándar de las diferencias.

#### 1.3.0.1. Cálculo del tamaño del efecto con R

Debemos cercionarnos de que en los ejemplos de  $t$ -test que hemos llevado a cabo hemos tomado la decisión correcta.

En el ejemplo del  $t$ -test dependiente decíamos que como  $p = 0,109 > 0,05$  se acepta la hipótesis nuula, es decir, no hay efecto significativo. Como no hay efecto, carece de sentido calcular su tamaño. De todas formas, lo haremos para comprobar que efectivamente el intervalo de confianza del mismo va a contener al 0.

Para el  $t$ -test dependiente lo hacemos con la  $d$  de Cohen.

```
s <- sd( df[ df$grupo == 1, 2 ] - df[ df$grupo == 0, 2 ] )
s
```



```
## [1] 2.31
M <- mean( df[ df$grupo == 1, 2 ] - df[ df$grupo == 0, 2 ] )
M
## [1] 1.3
d <- abs( M ) / s
d
```

```
## [1] 0.562
```

$d = 0.56$ , es decir, el tamaño del efecto es mediano.

Para calcular el IC del tamaño del efecto podemos emplear la función `ci.sm()` del paquete MBESS.

```
#install.packages( "MBESS" )
library( "MBESS" )
ci.sm( Mean = M, SD = s, N = 10, conf.level = 0.95 )
## [1] "The 0.95 confidence limits for the standardized mean are given as:"
## $Lower.Conf.Limit.Standardized.Mean
## [1] -0.121
##
## $Standardized.Mean
## [1] 0.562
##
## $Upper.Conf.Limit.Standardized.Mean
## [1] 1.22
```

Esta función no solo nos proporciona el intervalo de confianza  $IC = (-0.12, 1.22)$  sino que también devuelve la  $d$  de Cohen.

Efectivamente el intervalo de confianza contiene al “0” y la hipótesis nula no se rechaza tal y como habíamos concluido con el  $t$ -test.

Para la continuación del ejemplo del  $t$ -test independiente que arrojaba un  $p$ -valor  $= 0,079 > 0,05$  empleamos la función `tes()` del paquete *compute.es* para **calcular el tamaño del efecto**.

El modelo `fitIndep` devuelve una lista de valores. El primero de ellos es el **estadístico  $t$**  y su valor numérico se llama escribiendo `fitIndep[[1]]`. Los grados de libertad son el segundo valor y se llama escribiendo `fitIndep[[2]]`.

```
library( compute.es )
tes( fitIndep[[1]], 10, 10 )
## Mean Differences ES:
##
## d [ 95 %CI] = -0.83 [ -1.81 , 0.15 ]
## var(d) = 0.22
## p-value(d) = 0.09
## U3(d) = 20.3 %
## CLES(d) = 27.9 %
## Cliff's Delta = -0.44
##
## g [ 95 %CI] = -0.8 [ -1.73 , 0.14 ]
## var(g) = 0.2
## p-value(g) = 0.09
## U3(g) = 21.3 %
## CLES(g) = 28.7 %
```



```
##
## Correlation ES:
##
## r [ 95 %CI] = 0.4 [ -0.08 , 0.73 ]
## var(r) = 0.04
## p-value(r) = 0.1
##
## z [ 95 %CI] = 0.42 [ -0.08 , 0.93 ]
## var(z) = 0.06
## p-value(z) = 0.1
##
## Odds Ratio ES:
##
## OR [ 95 %CI] = 0.22 [ 0.04 , 1.31 ]
## p-value(OR) = 0.09
##
## Log OR [ 95 %CI] = -1.51 [ -3.28 , 0.27 ]
## var(lOR) = 0.71
## p-value(Log OR) = 0.09
##
## Other:
##
## NNT = -6.55
## Total N = 20
```

Donde podemos encontrar, entre otras medidas del tamaño del efecto, la  $r$  de Pearson = 0.4 y su intervalo de confianza (-0.08,0.73). Además de la  $r$  de Pearson esta función nos proporciona la  $d$  de Cohen y su intervalo de confianza para un  $t$ -test independiente.

Igualmente, podemos utilizar su fórmula para calcular la  $r$  de Pearson:

```
r<-function( t, gl ){
  sqrt( t ^ 2 / ( t ^ 2 + gl ) )
}
r( fitIndep[[1]], fitIndep[[2]] )

##      t
## 0.401
```

O también la  $d$  de Cohen para los  $t$  test independientes.

```
dCohen<-function( x, y ){
  lx <- length(x)- 1
  ly <- length(y)- 1
  md <- abs(mean(x) - mean(y))
  csd <- lx * var(x) + ly * var(y)
  csd <- csd/(lx + ly)
  csd <- sqrt(csd)
  cd <- md/csd
}
d <- dCohen( df$medida[ df$grupo == 1 ], df$medida[ df$grupo == 0 ] )

O también la función cohensD() del paquete lsr.

library(lsr)
cohensD( df$medida[ df$grupo == 1 ] , df$medida[ df$grupo == 0 ] )

## [1] 0.831
```



## 2. Comparaciones entre dos o más grupos

Vamos a generalizar el concepto de comparación entre dos grupos que hemos visto anteriormente a más grupos. Para ello consideraremos dos tipos de técnicas:

- Paramétricas: cuando la distribución de los datos de la muestra tiene cierta distribución particular usaremos el modelo ANOVA.
- No paramétricas: no será necesario suponer nada sobre la distribución de los datos. Estas pruebas se estudiarán en el apartado “*Contrastes no paramétricos: datos ordinales*” del próximo tema.

Borramos del entorno de trabajo el anterior conjunto de datos porque usaremos esta variable para un ejemplo general en el siguiente apartado.

### 2.1. ANOVA

El análisis de varianza, *ANOVA*, es una generalización de los *t*-test para dos muestras al caso de diseños con más de dos muestras.

A la variable categórica que define los grupos que deseamos comparar la llamamos *variable independiente* o *factor* y a la variable cuantitativa en la que deseamos comparar los grupos la llamamos *variable dependiente*.

Los factores o variables independientes pueden ser a su vez de dos tipos: que varían *entre sujetos* (*between subjects*) o que varían *dentro de los sujetos*, conocidos también como *intra sujetos* (*within subjects*).

Esta diferenciación será muy importante a la hora de elegir el modelo ANOVA a llevar a cabo, ya que hay varios.

- Los factores que varían entre sujetos (*between*) son los que no se miden dos (o más) veces para un mismo sujeto: la edad (un sujeto no puede tener dos edades diferentes), el género, la raza... o simplemente los que, según el experimento, solo afectan una vez a cada sujeto. A cada sujeto se le asigna un único valor para ese factor.
- Los factores que varían dentro de los sujetos (*within*) son los que se miden varias veces para el mismo sujeto. Por ejemplo: seguimiento del mismo tratamiento tomando muestras en varios momentos. Los momentos en los que se toman las muestras son niveles y el tiempo es el factor intra sujetos ya que para un mismo sujeto se repiten las mediciones. A cada sujeto se le asignan tantos valores de ese factor como niveles tenga éste.

¿Por qué usar un nuevo modelo diferente de los *t*-test?

Ya se ha comentado con anterioridad el desorbitado crecimiento del error de Tipo I al crecer las comparaciones<sup>1</sup>. Para solventar esto podemos aplicar el modelo ANOVA.

El conjunto de datos que utilizaremos para ilustrar todos los modelos ANOVA que vamos a estudiar es

```
df <- read.table( "files/ejemplo.csv", sep = ";", head = TRUE )
head( df )
```

```
##   id genero raza m0 m1 m3
## 1  1      1    3 35 25 16
## 2  2      2    1 37 23 12
## 3  3      2    1 36 22 14
## 4  4      1    2 34 21 13
## 5  5      2    3 60 43 22
## 6  6      1    3 54 46 26
```

<sup>1</sup>Si cada test tiene un nivel de significación  $\alpha = 0,05$  la probabilidad de no cometer error de Tipo I al hacer  $N$  comparaciones es  $0,95^N$ . Así, para 3 grupos hay un  $100(1 - (0,95)^3) = 14,3\%$  de probabilidad de cometer error de tipo I, no aceptar  $H_0$  cuando es verdadera, y conforme crece  $N$  lo hace la probabilidad de cometer este tipo de error.

donde se mide la efectividad de un tratamiento a un conjunto de enfermos de una enfermedad rara. Se realiza una prueba de falta de capacidad de raciocinio antes de ser expuestos a este tratamiento (mes cero), cuando se lleva un mes de tratamiento y cuando se llevan tres meses. A simple vista se aprecia una gran mejora. Además, se diferencian los sujetos por género y por raza. Podemos identificar los factores:

- género: varía entre sujetos (*between*).
- raza: varía entre sujetos (*between*).
- mes: varía dentro de los sujetos (*within*) y a su vez tiene 3 niveles: m0 (mes 0), m1 (mes 1) y m3 (mes 3).

En nuestro conjunto de datos, los factores vienen dados como variables numéricas donde por ejemplo, la edad 1 puede significar *adulto* y la edad 2 *anciano*. Sin embargo, los factores en R deben ser tratados como variables categóricas lo cual se consigue mediante la función `factor()`.

```
df$id      <- factor( df$id )
df$raza    <- factor( df$raza )
df$genero  <- factor( df$genero )
```

Si se desea comprobar que una variable es un factor se utiliza la función `is.factor()`.

```
is.factor( df$genero )
```

```
## [1] TRUE
```

### 2.1.1. ANOVA de una vía (entre sujetos)

El ANOVA de una vía (*one way ANOVA*), también conocido como *ANOVA de un factor*, examina la igualdad de las medias de la población para un resultado cuantitativo y una única variable categórica con dos o más niveles de tratamiento. Cada sujeto estará expuesto a un único nivel de este tratamiento. La hipótesis nula  $H_0$  es que no hay diferencia entre las medias y la alternativa,  $H_1$  que al menos una de las medias difiere del resto.

El hecho de verificar la hipótesis nula de que hay igualdad de medias entre grupos se puede interpretar como que las observaciones proceden de un único grupo cuya media y variabilidad es la misma que la de cualquiera de los grupos por separado. ANOVA es la forma más simple de mediar la variabilidad.

La cuestión es que si alguno de los grupos presenta unos valores que en media se alejan del resto, esto se apreciará en el contraste como una fuente extra de variabilidad no explicable por el azar. Nosotros nos dedicaremos a identificar cuándo se rechaza la hipótesis nula (cuando el p-valor es menor que 0.05) y cuándo se acepta (cuando es mayor). López (n.d.)

Vamos a tomar para este ANOVA de un factor la variable independiente *raza* y la variable dependiente, los datos, *m0* para inferir si el género influye en la prueba que hemos llevado a cabo.

#### 2.1.1.1. Supuestos y robustez de ANOVA

Hay una serie de premisas que se deben cumplir, ordenadas de mayor exigencia a menor, para poder llevar a cabo el modelo ANOVA de una vía:

##### 2.1.1.1.1. Independencia de las observaciones

Dos observaciones son independientes cuando una no depende de la otra. Es decir, el resultado de la primera observación no influye en el resultado de la siguiente.

Este supuesto es el más importante. Su violación va a invalidar las conclusiones del análisis porque produce errores en el cálculo de las varianzas y, por tanto, en los intervalos de confianza y las pruebas de hipótesis deducidas.



La falta de independencia se produce por ejemplo cuando se trabaja con variables aleatorias que se observan a lo largo del tiempo, por ello, si sucede hay que cercionarse de que el análisis es correcto aplicar este análisis o por el contrario la variable varía *dentro de los sujetos*. (Vilar, n.d.)

En otro caso, es debido a errores en el muestreo: efecto aprendizaje, descuidos, falta de aleatorización... o a la existencia de otros factores que también influyen en la variable respuesta, y no han sido tomados en consideración.

#### 2.1.1.1.2. Homocedasticidad: Homogeneidad de varianzas.

Tenemos que comprobar que las varianzas de los grupos son iguales. Para ello podemos emplear el test de Bartlett o el de Levene. Como ya se comentó en el módulo *20Introducción a los contrastes* en la sección “*Contrastes de homogeneidad de varianza*”, el test de Levene es menos sensible a la falta de normalidad que el de Bartlett. Sin embargo, si estamos seguros de que los datos provienen de una distribución normal, entonces el test de Bartlett es el mejor. Para ambos la hipótesis nula es la igualdad de varianzas entre grupos y la hipótesis alternativa es la no igualdad.

Para realizar el test de Bartlett la función que vamos a utilizar es `bartlett.test()` y para el test de Levene la función `LeveneTest()`.

Si falla la homocedasticidad, siempre que no haya grandes diferencias entre el número de observaciones en los distintos grupos, es decir, siempre que nuestro modelo sea balanceado el ANOVA sigue siendo fiable. Con otras palabras, ANOVA es robusto con respecto a esta condición.

Para nuestro conjunto:

```
fit2 <- bartlett.test( df$m0 ~ df$raza )
fit2

##
## Bartlett test of homogeneity of variances
##
## data: df$m0 by df$raza
## Bartlett's K-squared = 3, df = 2, p-value = 0.2
```

El p-valor = 0,182 > 0,05 luego aceptamos la hipótesis nula; los grupos son homocedásticos.

Veamos ahora cómo actuar cuando no se da igualdad de varianzas:

```
dfIS <- InsectSprays
head( dfIS )

##   count spray
## 1    10    A
## 2     7    A
## 3    20    A
## 4    14    A
## 5    14    A
## 6    12    A

fit3 <- bartlett.test( dfIS$count ~ dfIS$spray )
fit3

##
## Bartlett test of homogeneity of variances
##
## data: dfIS$count by dfIS$spray
## Bartlett's K-squared = 30, df = 5, p-value = 0.00009
```

El p-valor < 0,05 luego los grupos no son homocedásticos.



La alternativa que se suele emplear cuando el diseño no es homocedástico para seguir adelante con el análisis es la función `oneway.test()` que lleva a cabo el test de Welch (conocido también como ANOVA heterocedástica).

```
oneway.test( dfIS$count ~ dfIS$spray )

##
## One-way analysis of means (not assuming equal variances)
##
## data: dfIS$count and dfIS$spray
## F = 40, num df = 5, denom df = 30, p-value = 8e-12
```

se obtiene que se rechaza la hipótesis nula, esto es, hay diferencia entre grupos.

Otra opción es considerar las pruebas no paramétricas que se estudiarán en el módulo “50No Paramétricos”.

### 2.1.1.1.3. Normalidad

La distribución de los datos debe aproximarse a la de una normal. Ya hemos visto en el módulo 20 *Introducción a los contrastes* en la sección “*Contrastes de normalidad*” las herramientas para comprobarlo.

El contraste de ANOVA es robusto frente a la violación del supuesto de normalidad, es decir, si los datos no son normales todavía da buenos resultados.

Para muchos problemas de la biología la falta de normalidad se soluciona haciendo una transformación de los datos (por ejemplo, tomando logaritmos, en cuyo caso los datos seguirían una distribución lognormal).

```
fit4 <- shapiro.test( df$m0[df$raza == 1] )
fit5 <- shapiro.test( df$m0[df$raza == 2] )
fit55 <- shapiro.test( df$m0[df$raza == 3] )
```

En este caso, el p-valor = 0,1433 > 0,05 luego la distribución de los datos es normal.

### 2.1.1.2. ANOVA de una vía con R

Una vez comprobados los supuestos, queremos averiguar si la falta de raciocinio antes del tratamiento tiene que ver con la raza del paciente. Para ello llevaremos a cabo el análisis ANOVA de un factor.

Es muy importante el orden de los argumentos. El primero es siempre la variable dependiente (antes). Le sigue el símbolo tilde (~) y la variable independiente (raza). El último argumento de la función `aov` es el nombre del conjunto de datos donde se encuentran los datos que queremos analizar. Este formato será igual para todos los análisis ANOVA que se hagan con la función `aov`.

```
#ANOVA de una vía (entre sujetos)
# VI: m0
# VD: grupo
fit1E <- aov( m0 ~ raza, data=df )
summary( fit1E )

##           Df Sum Sq Mean Sq F value Pr(>F)
## raza       2      845      422   7.28 0.0062 **
## Residuals  15      870       58
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Vemos que el p-valor es 0,00178 < 0,05 luego se aprecia una diferencia de resultados para las distintas razas antes del tratamiento.



### 2.1.1.3. Tamaño del efecto ANOVA de 1 vía

Ya hemos comentado la importancia del tamaño del efecto. La medida más común para el ANOVA de una vía es *eta cuadrado* ( $\eta^2$ ) que es la razón entre la suma de cuadrados del efecto y la suma de cuadrados total (de todos los efectos y los residuos).

$$\eta^2 = \frac{SC_{efecto}}{SC_{total}}$$

Primero la realizamos a mano ya que mediante la función `summary()` se obtienen los datos necesarios:

$$\eta^2 = \frac{844,8}{844,8 + 870,3} = 0,492$$

y teniendo en cuenta que el tamaño para el ANOVA de una vía es

Efecto	Pequeño	Mediano	Grande
$\eta^2$	0.01	0.06	0.14

Luego en nuestro caso el efecto es grande.

Además, podemos calcular el intervalo de confianza para  $\eta^2$

```
ci.pvaf(F.value=7.28, df.1=2, df.2=15, N=18, conf.level=.95)
```

```
## $Lower.Limit.Proportion.of.Variance.Accounted.for
## [1] 0.0641
##
## $Probability.Less.Lower.Limit
## [1] 0.025
##
## $Upper.Limit.Proportion.of.Variance.Accounted.for
## [1] 0.671
##
## $Probability.Greater.Upper.Limit
## [1] 0.025
##
## $Actual.Coverage
## [1] 0.95
```

Es decir, hay un 95 % de probabilidad de que el verdadero valor de  $\eta^2$  esté en (0.064, 0.6) Clay (n.d.) y Watson (n.d.).

El valor de  $\eta^2$  puede ser obtenido también mediante la función `etaSquared()` del paquete `lsr` cuyos argumentos son el modelo tipo ANOVA y el tipo de sumas de cuadrados. Además, devuelve también el valor de  $\eta_p^2$  (eta cuadrado parcial), donde

$$\eta_p^2 = \frac{SC_{efecto}}{SC_{efecto} + SC_{residuos}}$$

```
library(lsr)
etaSquared( fit1E, type = 1, anova = FALSE )

##      eta.sq eta.sq.part
## raza  0.493      0.493
```



### 2.1.2. El estadístico F. Valor de F (F value)

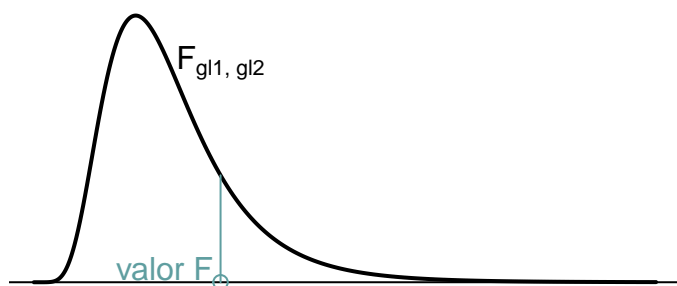
Además del p-valor, que es en lo que nos hemos estado fijando para aceptar o rechazar la hipótesis nula, vemos que la función `summary()` nos proporciona el valor de F. Este valor es a partir del cual se obtiene el p-valor.

El valor F es la división de la varianza entre grupos (en nuestro caso tenemos dos grupos: femenino y masculino) con la varianza dentro de los grupos, es decir, la división entre la varianza *explicada* entre la varianza *inexplicada*.

Se considera que la varianza entre grupos es la explicada porque es la que se debe a la variable independiente, es decir, se debe a que se supone que entre los grupos habrá ciertas diferencias. Sin embargo, la varianza dentro de los grupos es lo que se llama *error de varianza* o *varianza inexplicada* porque se supone que dentro del mismo grupo los resultados serían los mismos. Hall (n.d.)

Si en la población de la que proceden las muestras no hay diferencias reales entre los grupos, la varianza entre grupos será similar a la varianza dentro de los grupos y por tanto el cociente entre ambas estará cerca de 1. En el caso de que existan diferencias reales entre los grupos la varianza entre grupos será mayor que la varianza dentro de los grupos (el cociente entre ambas será mayor de 1). El estadístico que nos dice si las desviaciones respecto a ese valor de 1 son significativas es el estadístico F de Snedecor.

#### Función de densidad de la distribución F de Snedecor y valor F



Como vemos, a mayor valor de F menor p valor porque el área bajo la curva que queda a la derecha es menor conforme crece el valor de F.

En nuestro caso la probabilidad de que el estadístico F sea mayor que 5.802 (el p-valor) es menor que 0.05 luego se rechaza la hipótesis nula y concluimos que las medias no son iguales.

### 2.1.3. Cómo comunicar los resultados de un ANOVA

Cuando nos disponemos a hacer públicos o a comunicar los resultados de nuestro análisis tenemos que dar detalles del valor de F y de los grados de libertad con los que el estadístico F ha sido calculado. Así, los grados de libertad que se usan para estimar el valor de F son los grados de libertad para el efecto del modelo ( $glM = 2$ ) y para los residuos ( $glR = 15$ ). Por lo tanto, la forma correcta de expresar los resultados del modelo sería

Hay un efecto significativo de la raza en el test,  $F(2,15) = 7.28$ ,  $p < 0,05$ ,  $\eta^2 = 0.492$ .

(A. Field, Miles, & Field, 2012)

### 2.1.4. ANOVA de dos factores (de dos vías)

Éste análisis sirve para estudiar la relación entre una variable dependiente y dos variables independientes (dos factores). Cada factor puede tener a su vez varios niveles. El modelo ANOVA de dos vías no solo evalúa los efectos de las variables independientes sino que también puede evaluar los efectos de la interacción entre ellas



### 2.1.4.1. Supuestos

Son los mismos que en el de una vía. Para la **raza** ya los habíamos comprobado, nos falta para el **género**:

```
fit6 <- bartlett.test( df$m0 ~ df$genero )
fit6

##
## Bartlett test of homogeneity of variances
##
## data: df$m0 by df$genero
## Bartlett's K-squared = 0.05, df = 1, p-value = 0.8
```

Los p valores son  $> 0,05$  así que podemos seguir con el análisis ANOVA.

### 2.1.4.2. ANOVA de dos vías con R

La estructura del modelo es muy similar. Ahora aparecen los factores independientes multiplicados entre sí, lo que indica que su interacción será evaluada.

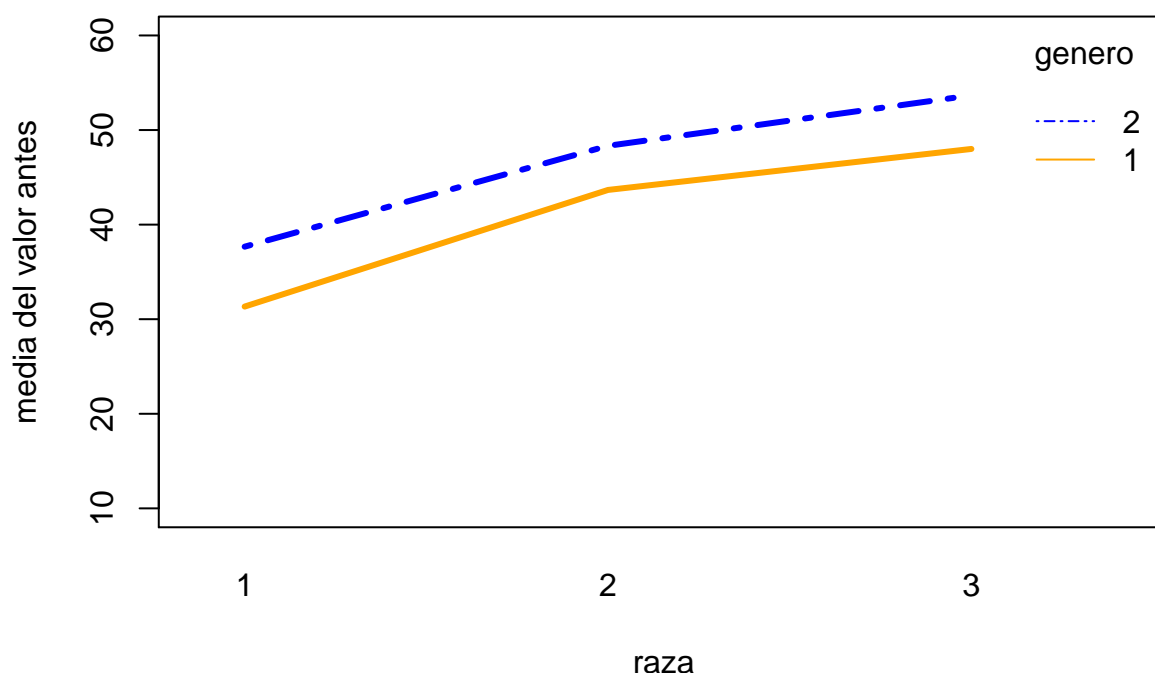
```
#ANOVA de dos vías (entre sujetos)
# VI: género (entre sujetos)
# VI: raza (entre sujetos)
# VD: joven
fit2E <- aov( m0 ~ raza * genero, data = df )
summary( fit2E )

##              Df Sum Sq Mean Sq F value Pr(>F)
## raza          2    845      422   6.95 0.0099 **
## genero         1    139      139   2.29 0.1565
## raza:genero    2         2         1   0.02 0.9828
## Residuals    12    729         61
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Así, el p-valor para la variable **raza** es menor que 0.05 por lo que ésta tiene un efecto significativo, el **género** no lo es y no hay interacción entre ambas porque el p-valor para la interacción es  $0,9828 > 0,05$ , esto es, se acepta la hipótesis nula.

Vamos a hacer una interpretación visual entre los efectos de los factores mediante un *interaction plot*.

```
interaction.plot( df$raza, df$genero, df$m0, ylim = c( 10, 60 ),
                  col = c( "orange", "blue" ), lty = c( 1, 12 ),
                  lwd = 3, ylab = "media del valor antes",
                  xlab = "raza", trace.label = "genero" )
```



El análisis entre **géneros** indica que la variable no tiene un efecto significativo, cosa que se aprecia en el *interaction plot* porque las líneas de ambos géneros están relativamente cerca. Sin embargo, la **raza** sí tiene un efecto significativo, ya que claramente las líneas crecen o decrecen conforme nos movemos entre las razas. Asimismo, no hay interacción entre **género** y **raza** porque las líneas del *interaction plot* van a la par, es decir, tienen la misma pendiente. Si hubiera interacción sería porque las líneas se cruzan o acabarían cruzándose al extenderlas.

**Nota:** el orden en el que multiplicamos los factores independientes *SÍ* altera el producto cuando los datos no son balanceados. Es decir, en nuestro caso el modelo

```
fit2E2 <- aov( m0 ~ genero * raza, data = df )
summary( fit2E2 )
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## genero      1    139      139    2.29 0.1565
## raza        2    845      422    6.95 0.0099 **
## genero:raza  2         2         1    0.02 0.9828
## Residuals   12    729        61
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

arroja los mismos resultados. Sin embargo, si los datos no son balanceados no es así. Vamos a modificar nuestro conjunto de datos usando solamente las 9 primeras filas de manera que ya no hay el mismo de personas de cada **raza** ni el mismo número de personas de cada **género**.

```
dfMOD <- df[ 1:11, ]
dfMOD
```

##	id	genero	raza	m0	m1	m3
## 1	1	1	3	35	25	16
## 2	2	2	1	37	23	12
## 3	3	2	1	36	22	14
## 4	4	1	2	34	21	13
## 5	5	2	3	60	43	22
## 6	6	1	3	54	46	26



```
## 7 7 1 2 50 46 23
## 8 8 2 2 60 47 25
## 9 9 2 3 48 35 20
## 10 10 2 2 47 32 19
## 11 11 1 1 30 20 11
```

Ahora volvemos a hacer el ANOVA de 2 vías:

```
fit2E3 <- aov( m0 ~ raza * genero, data = dfMOD )
summary( fit2E3 )
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## raza       2    442   221.2    2.38   0.19
## genero     1    241   240.7    2.59   0.17
## raza:genero 2     10     5.0    0.05   0.95
## Residuals  5    466    93.1
```

y cambiando el orden de los factores

```
fit2E4 <- aov( m0 ~ genero * raza, data = dfMOD )
summary( fit2E4 )
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## genero     1    149   149.3    1.60   0.26
## raza       2    534   266.9    2.87   0.15
## genero:raza 2     10     5.0    0.05   0.95
## Residuals  5    466    93.1
```

**¡Los resultados son diferentes!** ¿Está mal? *NO*. Solamente tenemos que hacer una buena interpretación de los resultados.

Lo que sucede es que la función `aov()` usa por defecto las sumas de cuadrados de tipo I. Esto significa que en `fit2E3` el valor de  $F = 2,376$  nos indica la variación explicada por la **raza** en la variable *m0* (medición antes de aplicar el tratamiento) sin tener en cuenta nada más y de ahí se obtiene el p-valor y, a continuación, la variable **género** tiene un valor  $F = 2,585$  que se interpreta de una forma algo diferente en el sentido de que como es la segunda variable incluida en el modelo mide la variación **adicional** que **género** explica pero que **raza** no explicaba.

En `fit2E4` vemos que los resultados son diferentes porque la variación en los resultados explicada únicamente por el **género** es 1,604 y la variación explicada por la **raza** tras tener en cuenta el **género** es 2,866. Si hacemos la interpretación correcta, todo será correcto. Sin embargo, la alternativa en el caso en que nos encontremos con un diseño no balanceado es utilizar la suma de cuadrados del Tipo III. Para ello, usaremos la función `Anova()` del paquete *car*. Esta función tiene como primer argumento un modelo lineal que se obtendrá mediante la función `lm()`.

```
options(contrasts = c("contr.sum", "contr.poly"))
fit2E5 <- lm( m0 ~ genero * raza, data = dfMOD )
library( "car" )
Anova( fit2E5, type = "III" )
```

```
## Anova Table (Type III tests)
##
## Response: m0
##           Sum Sq Df F value    Pr(>F)
## (Intercept) 19389  1  208.26 0.000029 ***
## genero       216  1    2.32    0.19
## raza        470  2    2.52    0.17
## genero:raza   10  2    0.05    0.95
## Residuals   466  5
```



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

Ahora sí se obtienen los mismos resultados independientemente del orden.

fit2E6 <- lm( m0 ~ raza * genero, data = dfMOD)
library( "car" )
Anova( fit2E2, type = "III" )

## Anova Table (Type III tests)
##
## Response: m0
##              Sum Sq Df F value    Pr(>F)
## (Intercept)   2945  1   48.46 0.000015 ***
## genero         60  1    0.99   0.339
## raza          449  2    3.69   0.056 .
## genero:raza     2  2    0.02   0.983
## Residuals     729 12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

bertolt (n.d.)
```

### 2.1.5. Tipos de suma de cuadrados

La suma de cuadrados se utiliza para descomponer la variabilidad. Si tenemos la fórmula

```
mod <- aov( y ~ A * B, data = datos )
```

- Tipo I: Se conoce como suma de cuadrados secuencial porque tiene en cuenta el orden de los factores a la hora del cálculo de la suma de cuadrados. Los efectos se ajustan para aquellos que aparecen antes en la fórmula. A no tiene en cuenta nada. B es ajustada teniendo en cuenta A y la interacción A\*B es ajustada teniendo en cuenta A y B.
- Tipo II: Se conoce como suma de cuadrados jerárquica porque los efectos se ajustan entre los que están al mismo nivel o en niveles inferiores, es decir, A se ajusta para B, B para A y A\*B para A y B.
- Tipo III: Se conoce como suma de cuadrados marginal porque tiene en cuenta todos los factores en todo momento. A se ajusta para B y para A \* B, B se ajusta para A y para A \* B y, por último, A \* B se ajusta para A y B.

Para datos balanceados, es decir, cada nivel tiene el mismo número de observaciones, serán equivalentes. Sin embargo, cuando no tengamos datos balanceados el adecuado es el tipo III como hemos explicado anteriormente. (Kabacoff, 2011)

### 2.1.6. Efectos fijos vs efectos aleatorios

Hasta ahora los factores han sido de *efectos fijos*. Esto ocurre cuando los niveles del factor son seleccionados específicamente por el experimentador ya que el interés del experimento se centra en conocer los efectos sobre la respuesta de estos niveles particulares. No es que en nuestra población haya ocurrido que hemos encontrado miembros de la raza 1, 2 y 3 sino que hemos buscado sujetos de dichas razas. Cuando ocurre lo contrario se dice que el factor es de *efectos aleatorios*.

Para llevar a cabo este análisis se utiliza la función `lme()` del paquete `nlme`. Esta función posee un argumento para la fórmula de factores fijos y otro para la de factores aleatorios. Veamos ejemplos que ilustran el modelo:

En 8 lugares de una isla (DBAN, LFAN, NSAN...) se eligen 4 parcelas cultivadas de maíz a las que se les aplica el mismo tratamiento.



```
#install.packages("DAAG")
library("DAAG")
dfA<-ant111b
head(dfA)

##   site parcel code island id plot trt ears harvwt
## 1 DBAN      I   58      1  3  3.0 111 43.5   5.16
## 2 LFAN      I   58      1 40  4.0 111 40.5   2.93
## 3 NSAN      I   58      1 186 5.5 111 20.0   1.73
## 4 ORAN      I   58      1 256 4.5 111 42.5   6.79
## 5 OVAN      I   58      1 220 3.5 111 31.5   3.25
## 6 TEAN      I   58      1  77 5.0 111 32.5   2.65
```

Hay dos niveles de variación aleatoria: lugares y parcelas dentro de los lugares. La variación entre lugares puede ser, por ejemplo, debido a variaciones en la elevación del terreno o proximidad a fuentes de agua. Entre parcelas se espera que estos factores no afecten pero sí lo hagan otros como la fertilidad del terreno o el drenaje.

- factores fijos: no hay
- factores aleatorios: lugar y parcela.

Si no solo incluimos en el modelo el lugar sino también las parcelas el modelo se satura porque contiene todo, es decir, hay tantos efectos aleatorios como puntos en el conjunto de datos, así que nos libramos del nivel inferior:

```
#install.packages("nlme")
library(nlme)
modA <- lme( fixed = harvwt ~ 1, random = ~1 |site, data= dfA)
summary(modA)

## Linear mixed-effects model fit by REML
## Data: dfA
## AIC BIC logLik
## 100 105 -47.2
##
## Random effects:
## Formula: ~1 | site
## (Intercept) Residual
## StdDev:      1.54      0.76
##
## Fixed effects: harvwt ~ 1
## Value Std.Error DF t-value p-value
## (Intercept) 4.29      0.56 24 7.66      0
##
## Standardized Within-Group Residuals:
## Min      Q1      Med      Q3      Max
## -2.0349 -0.5607  0.0706  0.7437  1.8864
##
## Number of Observations: 32
## Number of Groups: 8
```

2)

```
library("DAAG")
dfB<-science
head(dfB)

##   State PrivPub school class sex like Class
```



```
## 1  ACT  public      1      1  f      8      1.1
## 2  ACT  public      1      1  f      6      1.1
## 3  ACT  public      1      1  f      5      1.1
## 4  ACT  public      1      1  f      2      1.1
## 6  ACT  public      1      1  f      5      1.1
## 7  ACT  public      1      1  f      6      1.1
```

factores fijos: sexo y tipo (de escuela: pública y privada), factores aleatorios: escuela (dentro de las cuales se examinan ciertas clases). Variable dependiente: puntuación. modelo:

```
modB <- lme(fixed = like ~ sex + PrivPub, data = dfB, random = ~ 1 | school/class, na.action=na.omit)
summary(modB)
```

```
## Linear mixed-effects model fit by REML
## Data: dfB
##   AIC   BIC logLik
## 5561 5593 -2775
##
## Random effects:
## Formula: ~1 | school
##          (Intercept)
## StdDev:    0.000309
##
## Formula: ~1 | class %in% school
##          (Intercept) Residual
## StdDev:    0.566    1.75
##
## Fixed effects: like ~ sex + PrivPub
##              Value Std.Error   DF t-value p-value
## (Intercept)  5.02    0.0929 1316   54.1  0.0000
## sex1         -0.09    0.0491 1316   -1.9  0.0636
## PrivPub1     -0.21    0.0929   39   -2.2  0.0325
## Correlation:
##          (Intr) sex1
## sex1      -0.001
## PrivPub1  0.384  0.012
##
## Standardized Within-Group Residuals:
##   Min     Q1   Med     Q3    Max
## -2.693 -0.689  0.154  0.679  2.558
##
## Number of Observations: 1383
## Number of Groups:
##          school class %in% school
##              41              66
```

Otro más:

```
mod <- lme(fixed = yield ~ shade, random = ~ 1 | block/plot, data=kiwishade)
summary(mod)
```

```
## Linear mixed-effects model fit by REML
## Data: kiwishade
##   AIC   BIC logLik
## 269 281 -127
##
## Random effects:
```



```
## Formula: ~1 | block
##      (Intercept)
## StdDev:      2.02
##
## Formula: ~1 | plot %in% block
##      (Intercept) Residual
## StdDev:      1.48      3.49
##
## Fixed effects: yield ~ shade
##              Value Std.Error DF t-value p-value
## (Intercept)  96.5      1.34 36    72.0  0.0000
## shade1       3.7      1.14  6     3.2  0.0184
## shade2       6.7      1.14  6     5.9  0.0011
## shade3      -6.6      1.14  6    -5.8  0.0012
## Correlation:
##      (Intr) shade1 shade2
## shade1  0.000
## shade2  0.000 -0.333
## shade3  0.000 -0.333 -0.333
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -2.415 -0.598 -0.069  0.780  1.589
##
## Number of Observations: 48
## Number of Groups:
##      block plot %in% block
##              3              12
```

### 2.1.7. ANOVA con medidas repetidas (within factors)

Este modelo se utiliza cuando se toman medidas para el mismo sujeto en condiciones diferentes (es el equivalente al *t*-test dependiente o *paired*). Es decir, de una lista de sujetos a cada uno le corresponden varias medidas.

No podemos usar el ANOVA *estándar* porque se viola la hipótesis de independencia.

Una ventaja de este experimento es que tiene menor coste porque se necesitan menos sujetos ya que todos están expuestos a todas las condiciones del experimento. Además, se controla mucho mejor la variabilidad interna.

En nuestro caso de estudio, la variable que varía entre sujetos es el mes y tiene tres niveles: m0, m1 y m3. (Más info en: [explorable.com](http://explorable.com), n.d.)

#### 2.1.7.1. Supuestos

El único supuesto necesario es el de *esfericidad*, que quiere decir que las variancias de las diferencias entre todos los pares de medidas repetidas sean iguales. Para ello usamos la prueba de esfericidad de Mauchly. Si hay 3 niveles denominados A, B y C, entonces se tendrá esfericidad si la varianza de las diferencias entre A y B es igual a la varianza de las diferencias entre A y C, igual a la de entre B y C. (4) Para evaluar la esfericidad vamos a usar el test de Mauchly cuyas hipótesis son:

$H_0$  : hay esfericidad

$H_1$  : no hay esfericidad



Si los datos violan la condición de esfericidad, cosa que es bastante común, hay algunas correcciones que se pueden hacer para conseguir un valor de F adecuado.

Las que vamos a utilizar son la **corrección de Greenhouse–Geisser** y la **corrección de Huynh–Feldt**. Ambos nos conducen a un factor de corrección (que no es otra cosa que un estadístico) que se aplica a los grados de libertad usados para evaluar el valor de F obtenido. No entraremos en detalles sobre cómo se calculan pero diremos que la de **HF** se calcula a partir de la de **GG** con la intención de ser menos conservativa. (Para los curiosos: Baron, n.d.)

La corrección de Greenhouse-Geisser arroja un estadístico (número)  $\epsilon$ . Si  $k$  es el número de grupos que se testean, esta corrección está entre  $\frac{1}{k-1}$  y 1. Si  $\epsilon$  se sitúa más cerca de la cota inferior  $\frac{1}{k-1}$  entonces esta corrección no ha conseguido hacer las modificaciones necesarias para alcanzar la esfericidad y, por lo tanto, no podremos usar el p-valor obtenido. Por otra parte, si se excede de 0.75 debemos utilizar la corrección de Huynh–Feldt porque cuando  $\epsilon$  es grande, Greenhouse-Geisser tiende a realizar un análisis muy estricto. Cuando  $\epsilon$  es menor que 0.4 ambas correcciones indican que la violación de la esfericidad está afectando a los p valores.

### 2.1.7.2. Estructura de los datos

Si consideramos cuatro sujetos  $S1, \dots, S4$  y cuatro niveles  $n1, \dots, n4$  normalmente se nos proporcionarán los datos en una tabla como esta:

sujeto	n1	n2	n3	n4
S1	7	5	6	2
S2	1	0	3	0
S3	8	8	6	1
S4	4	3	1	2

Sin embargo, no es de esta forma como podremos realizar el análisis con R sino que hay que definir la variable entre sujetos y poner los niveles como una medida más, es decir:

sujeto	nivel	medida
S1	n1	7
S1	n2	1
S1	n3	8
S1	n4	4
S2	n1	5
S2	n2	0
S2	n3	8
...	...	...
S4	n4	2

Para hacer esta transformación del conjunto de datos utilizaremos la función `melt()` del paquete `reshape2` donde seleccionaremos las columnas que queremos mantener y las que queremos escribir como una sola, además del nombre de la nueva variable que definirán y el nombre de la variable dependiente.

```
library( reshape2 )
df2 <- melt( df, id = c( "id", "genero", "raza" ),
             measure = c( "m0", "m1", "m3" ),
             variable.name = "mes", value.name = "valor" )
head( df2 )

##   id genero raza mes valor
```



```
## 1 1      1 3 m0 35
## 2 2      2 1 m0 37
## 3 3      2 1 m0 36
## 4 4      1 2 m0 34
## 5 5      2 3 m0 60
## 6 6      1 3 m0 54
```

### 2.1.7.3. ANOVA de medidas repetidas ( un factor intra sujetos) con R

Hay al menos dos formas de realizar este análisis con R. Con aov el modelo es:

```
mod <- aov( variable dependiente ~ factor + Error( sujetos / factor ) )

# IV (dentroDe): mes
# DV:          valor
fit1D <- aov(valor ~ mes + Error(id / mes ), data = df2 )
summary( fit1D )

##
## Error: id
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals 17   3247      191
##
## Error: id:mes
##           Df Sum Sq Mean Sq F value Pr(>F)
## mes         2   5684    2842    142 <2e-16 ***
## Residuals 34    680      20
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Como el p-valor es menor que 0.05 se rechaza la hipótesis nula, es decir, el mes tiene un efecto sobre la variable dependiente.

El problema es que aquí nada nos asegura la esfericidad. Podríamos llevar a cabo el test de Mauchly, sin embargo, si ocurre que este es significativo nos veríamos obligados a calcular las correcciones, proceso que puede resultar tedioso.

Por esta razón, vamos a utilizar la función `ezANOVA()` del paquete `ez` que lleva a cabo, además del modelo ANOVA el análisis de supuestos previo:

```
library( ez )
options( contrasts = c( "contr.sum", "contr.poly" ) )
fit12 <- ezANOVA(data = df2, dv = .( valor ), wid = .( id ), within = .( mes ),
                type = 3)
fit12

## $ANOVA
##   Effect DFn DFd   F      p p<.05   ges
## 2     mes   2  34 142 3.09e-17 * 0.591
##
## $`Mauchly's Test for Sphericity`
##   Effect      W      p p<.05
## 2     mes 0.523 0.00563 *
##
## $`Sphericity Corrections`
##   Effect   GGe   p[GG] p[GG]<.05   HFe   p[HF] p[HF]<.05
## 2     mes 0.677 2.3e-12 * 0.715 6.12e-13 *

```



El test de Mauchly nos da un  $p\text{-valor} = 0.005 < 0.05$ , es decir, rechaza la hipótesis nula y no podemos asumir la esfericidad. Como  $\epsilon = 0.67 < 0.75$  usamos la corrección de Greenhouse–Geisser que nos dice que el  $p\text{-valor} = 2.298E10^{-12} < 0.05$ , esto es, que el mes tiene un efecto significativo.

Un modelo para resolver este tipo de problema en el que hay algún factor intra sujetos y que no precisa de esfericidad es el MANOVA. (A. Field et al., 2012)

### 2.1.8. Modelos mixtos ANOVA

Hasta ahora teníamos *ANOVA sin medidas repetidas* donde las medidas deben ser independientes y hay un sujeto para cada grupo y *ANOVA con medidas repetidas* donde cada sujeto se mide varias veces bajo diferentes condiciones. ¿Y si no se da una condición o la otra sino ambas? Utilizaremos *ANOVA mixto* (*mixed ANOVA*).

#### 2.1.8.1. ANOVA mixto con R

El modelo es:

```
mod <- aov( variable dependiente ~ los factores de intereses multiplicados entre sí +
            Error(Sujetos/factoresIntraSujetosMultiplicadosEntreSí)
          )
```

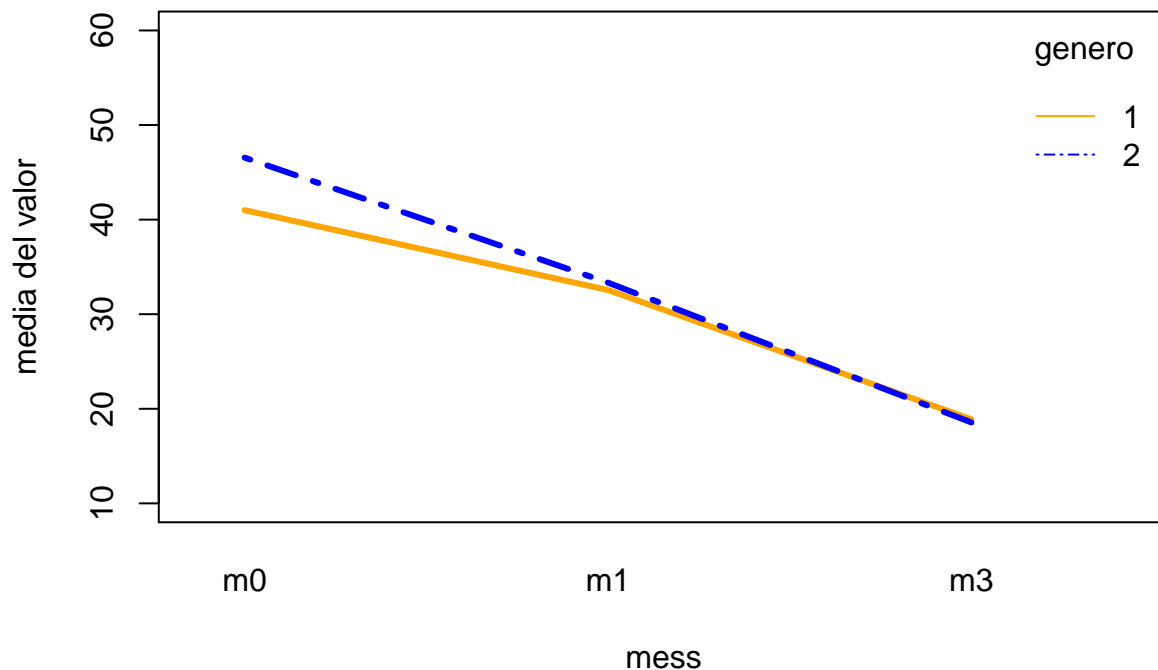
Si tenemos un factor entre sujetos y un factor intra sujetos:

```
# 2x2 mixed:
# IV between: género
# IV within: mes
# DV:        valor
fitM <- aov( valor ~ mes * genero + Error( id / mes ), data = df2 )
summary( fitM )

##
## Error: id
##           Df Sum Sq Mean Sq F value Pr(>F)
## genero      1     54      54    0.27   0.61
## Residuals  16    3193      200
##
## Error: id:mes
##           Df Sum Sq Mean Sq F value Pr(>F)
## mes         2    5684    2842  153.62 <2e-16 ***
## mes:genero  2      88      44    2.38   0.11
## Residuals  32     592       19
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Para interpretar los resultados recurriremos de nuevo a un *interaction.plot*

```
interaction.plot( df2$mes, df2$genero, df2$valor, ylim = c( 10, 60 ),
                  col = c( "orange", "blue" ), lty = c( 1, 12 ), lwd = 3,
                  ylab = "media del valor", xlab = "mes", trace.label = "genero")
```



El análisis entre **géneros** indica que el género no tiene un efecto significativo, cosa que se aprecia en el interaction plot porque las líneas de ambos géneros están muy juntas. El **mes** sí tiene un efecto significativo, es decir, los grupos cambian a través del tiempo (el valor va decreciendo). Por último no hay interacción entre **género** y **mes** porque los grupos cambian a través del tiempo pero de forma muy similar, casi paralelamente llegando a tocarse pero no a cruzarse.

Con `ezANOVA()`:

```
fit13 <- ezANOVA(data = df2, dv = .( valor ), wid = .( id ), between = .( género ), within = .( mes ),
  type = 3)
```

```
fit13
```

```
## $ANOVA
```

##	Effect	DFn	DFd	F	p	p<.05	ges
## 2	género	1	16	0.271	6.10e-01		0.0141
## 3	mes	2	32	153.615	3.93e-17	*	0.6003
## 4	género:mes	2	32	2.381	1.09e-01		0.0227

```
##
```

```
## $`Mauchly's Test for Sphericity`
```

##	Effect	W	p	p<.05
## 3	mes	0.402	0.00107	*
## 4	género:mes	0.402	0.00107	*

```
##
```

```
## $`Sphericity Corrections`
```

##	Effect	GGe	p[GG]	p[GG]<.05	HFe	p[HF]	p[HF]<.05
## 3	mes	0.626	1.63e-11	*	0.653	6.24e-12	*
## 4	género:mes	0.626	1.34e-01		0.653	1.32e-01	

Si ahora consideramos dos factores entre sujetos y un factor intra sujetos:

```
# 2x2 mixed:
# IV between: raza
# IV between: género
# IV within: mes
# DV: valor
```



```
fit2M <- aov( valor ~ mes * genero * raza +
              Error( id / mes ) , data = df2 )
summary( fit2M )

##
## Error: id
##           Df Sum Sq Mean Sq F value Pr(>F)
## genero      1     54      54    0.33  0.575
## raza        2    1217     608    3.74  0.055 .
## genero:raza  2      26      13    0.08  0.923
## Residuals   12    1950     163
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: id:mes
##           Df Sum Sq Mean Sq F value Pr(>F)
## mes         2    5684    2842  256.23 < 2e-16 ***
## mes:genero   2      88      44    3.97  0.03236 *
## mes:raza     4     297      74    6.69  0.00091 ***
## mes:genero:raza 4      29       7    0.65  0.63169
## Residuals   24     266      11
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### 2.1.8.2. Otros ANOVA mixtos:

- 2 factores intra sujetos (w1, w2)

```
fit2E <- aov(y ~ w1 * w2 + Error( sujetos/( w1 * w2 ) ), data = df )
```

- 1 factor entre sujetos ( b1 ) y 2 intra sujetos ( w1 y w2 )

```
fit1D2E <- aov(y ~ b1 * w1 * w2 + Error( sujetos / ( w1 * w2 ) ), data = df )
```

- 2 factores entre sujetos (b1 y b2) y 1 intra sujetos (w1)

```
fit2D1E <- aov( y ~ b1 * b2 * w1 + Error( sujetos / ( w1 ) ) )
```

R (n.d.)

### 2.1.9. ANOVA y regresión lineal

Vamos a ver con un ejemplo que no es que ANOVA sea un caso particular de regresión lineal múltiple sino que son lo mismo. Primero hacemos el análisis de varianza

```
fit1E <- aov( m0 ~ raza, data = df )
summary( fit1E )

##           Df Sum Sq Mean Sq F value Pr(>F)
## raza        2     845      422    7.28 0.0062 **
## Residuals   15     870       58
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

y mostramos las medias para cada raza

```
print( model.tables( fit1E, "means" ) )
```



```
## Tables of means
## Grand mean
##
## 43.8
##
## raza
## raza
##    1    2    3
## 34.5 46.0 50.8
```

Si ahora hacemos regresión lineal

```
summary( lm( m0 ~ factor( raza ), data = df ) )

##
## Call:
## lm(formula = m0 ~ factor(raza), data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.83  -3.33   1.25   3.79  14.00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      43.78       1.80   24.38 1.8e-13 ***
## factor(raza)1     -9.28       2.54   -3.65 0.0024 **
## factor(raza)2      2.22       2.54    0.88 0.3953
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.62 on 15 degrees of freedom
## Multiple R-squared:  0.493, Adjusted R-squared:  0.425
## F-statistic: 7.28 on 2 and 15 DF, p-value: 0.00617
```

podemos observar que el estadístico F (y en consecuencia el p-valor) es igual para ambos análisis.

En el modelo de regresión lineal lo que se está haciendo es representar los grupos mediante variables binarias. Lo hacemos con **raza**: como tiene 3 niveles se necesitan 2 variables binarias  $X_1$  y  $X_2$  definidas tal que así:

$$X_1 = \begin{cases} 0 & \text{si no es de raza 2} \\ 1 & \text{si es de raza 2} \end{cases}$$

$$X_2 = \begin{cases} 0 & \text{si no es de raza 3} \\ 1 & \text{si es de raza 3} \end{cases}$$

De esta forma hemos podido representar los 3 niveles del factor de la siguiente manera:

Grupo	$X_1$	$X_2$
1	0	0
2	1	0
3	0	1

El modelo de regresión lineal es  $y = b_0 + b_1X_1 + b_2X_2$  y, por los resultados de la regresión

$$y = 34,5 + 11,5X_1 + 16,33X_2.$$



Vamos a repasar los resultados para cada grupo:

**raza 1:**  $y = 34,5 + 11,5(0) + 16,33(0) = 34,5$

**raza 2:**  $y = 34,5 + 11,5(1) + 16,33(0) = 46$

**raza 3:**  $y = 34,5 + 11,5(0) + 16,33(1) = 50,83$

Que coinciden con las medias calculadas.

Así, ANOVA nos da cada media y el p-valor que dice si al menos dos son diferentes (de forma significativa) mientras que la regresión nos da solo una media (intercept) y las diferencias entre esta y las otras, pero los p-valores evalúan estas diferencias. Se obtiene la misma información con una estructura diferente. Factor (n.d.)

## 2.2. Después del ANOVA

Si al realizar el ANOVA se obtiene que el experimento es significativo, es decir, el p-valor es inferior a 0.05 se rechaza la hipótesis nula de igualdad de medias nos va a interesar, naturalmente, saber entre qué grupos se han producido diferencias. Para abordar la cuestión se tienen:

- Contrastes *post hoc* o no planificados: cuando no tenemos una idea previa de en qué grupos eran de esperar las mayores diferencias.
- Comparaciones planificadas: cuando tenemos alguna sospecha de dónde se pueden encontrar las diferencias.

### 2.2.1. Contrastes *post hoc* o no planificados

Los tests *post hoc* consisten en comparar los grupos de dos en dos. Es como hacer varios *t*-test consecutivos pero controlando el error de tipo I. Muchos de estos modelos son conservadores, es decir, reducen el error de tipo I a costa de aumentar la posibilidad de errores del tipo II, luego puede que hayan diferencias que no sean detectadas. Dependiendo de las características de nuestro modelo es aconsejable utilizar las una u otra.

El método de **Bonferroni** simplemente divide el nivel de confianza ( $\alpha = 0.05$ ) que queremos obtener por el número de test que hay que hacer. Este método se considera bastante conservador.

Si  $k$  es el número de grupos a comparar, Bonferroni impone a cada comparación un nivel de significación igual a  $\frac{\alpha}{k}$ , es decir, si hay 5 grupos este será  $\frac{0,05}{5} = 0,01$ .

Este ajuste se lleva a cabo usando el argumento `p.adj = "bonferroni"` en la función `pairwise.t.test()`.

```
fit1D <- aov(valor ~ mes + Error( id / mes), data = df2)
summary(fit1D)

##
## Error: id
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals 17   3247      191
##
## Error: id:mes
##           Df Sum Sq Mean Sq F value Pr(>F)
## mes         2   5684    2842    142 <2e-16 ***
## Residuals 34    680      20
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

pairwise.t.test(df2$valor, df2$mes, p.adj = "bonferroni", paired = TRUE)
```



```
##
## Pairwise comparisons using paired t tests
##
## data: df2$valor and df2$mes
##
##      m0      m1
## m1 3e-09 -
## m3 5e-10 2e-07
##
## P value adjustment method: bonferroni
```

Todos los p-valores son menores que 0.05, es decir, todas las razas difieren entre ellas.

El ajuste de **Holm** compara secuencialmente el p-valor con una significación que se reduce para cada test. Es decir, en nuestro caso tenemos 3 grupos, lo que quiere decir que el primer p-valor se compara con el nivel de significación  $\alpha = \frac{0,05}{3} = 0,017$ , el segundo con  $\alpha = \frac{0,05}{2} = 0,025$ , y el último con  $\alpha = \frac{0,05}{1} = 0,05$ .

En general, este método se considera superior al ajuste de Bonferroni y se utiliza escribiendo el argumento `p.adj = "holm"` en la función `pairwise.t.test()`.

```
pairwise.t.test( df2$valor, df2$mes, p.adj = "holm", paired = TRUE )
##
## Pairwise comparisons using paired t tests
##
## data: df2$valor and df2$mes
##
##      m0      m1
## m1 2e-09 -
## m3 5e-10 6e-08
##
## P value adjustment method: holm
```

Cuando haya más de 6 grupos evitaremos usar estos tests. Quick (n.d.)

Otro método que no es generalmente la opción más recomendada cuando tenemos otras a nuestra disposición es **Mínima diferencia significativa de Fisher** (*Fishers Least Significant Difference (LSD)*) pero en caso de que sea necesario aplicarlo utilizaremos la usando el argumento `p.adj = "none"` en la función `pairwise.t.test()`.

```
pairwise.t.test( df$m0, df$raza, p.adj = "none" )
##
## Pairwise comparisons using t tests with pooled SD
##
## data: df$m0 and df$raza
##
##      1      2
## 2 0.020 -
## 3 0.002 0.289
##
## P value adjustment method: none
```

Como vemos, las diferencias residen entre las razas 1 y 2 porque el p-valor es  $0.019 < 0.05$  y entre las razas 1 y 3 porque es también  $0.021 < 0.05$ .

Por último, cuando tenemos más de 6 grupos, el diseño es balanceado, es decir, todos los grupos tienen el mismo número de individuos y se da homocedasticidad usaremos **Tukey HSD** que compara cada grupo con los demás.



Recordamos que `fit1E` nos hacía ver que las medias entre todas las razas no son iguales.

Veamos dónde se encuentran esas diferencias:

```
fit1E <- aov( m0 ~ raza, data = df )
TukeyHSD( fit1E )

##    Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = m0 ~ raza, data = df)
##
## $raza
##      diff      lwr    upr p adj
## 2-1 11.50   0.0768 22.9 0.048
## 3-1 16.33   4.9102 27.8 0.006
## 3-2  4.83  -6.5898 16.3 0.529
```

De nuevo, las diferencias residen entre las razas 1 y 2 y entre las razas 1 y 3 porque sus p valores son inferiores a 0.05.

(muy buen video Tukey HSD y Fisher LSD: <https://www.youtube.com/watch?v=9iL8v6A-gi0>)

### 2.2.2. Comparaciones planificadas

Supongamos que hemos rechazado la hipótesis nula del análisis de varianza y queremos ver entre qué grupos hay diferencias pero tenemos una sospecha honesta de dónde se encuentran.

Esto ocurre por ejemplo cuando a un grupo de control se le aplica un placebo, una cierta cantidad de un medicamento y una cantidad superior pero que no esperamos que mejore mucho los resultados del segundo grupo. Posiblemente sospechemos que al que se le administra el placebo va a obtener resultados muy diferentes al segundo y al tercero pero no pensamos encontrar diferencias estos últimos.

Supongamos que nosotros sospechamos que la raza 1 va a diferir mucho de las razas 2 y 3 pero entre ellas no difieren significativamente.

Los contrastes se hacen de 2 en 2. En nuestro caso habrá 2 contrastes porque tenemos 3 grupos (siempre uno menos).

El primer contraste comparará la raza 1 con el grupo constituido por la raza 2 y 3 y el segundo contraste comparará las razas 2 y 3. Lo primero que vamos a hacer es ver si el modelo es balanceado. Podemos usar la función `summary()`

```
summary( df2 )

##      id      genero raza      mes      valor
## 1      : 3      1:27  1:18  m0:18  Min.    :11.0
## 2      : 3      2:27  2:18  m1:18  1st Qu.:21.2
## 3      : 3           3:18  m3:18  Median :30.5
## 4      : 3           Mean   :31.8
## 5      : 3           3rd Qu.:42.2
## 6      : 3           Max.   :60.0
## (Other):36
```

Lo que muestra que tenemos un diseño balanceado, con 6 replicaciones por **raza**. Para cada comparación necesitamos crear un vector de coeficientes y luego unirlos en una matriz mediante la función `cbind()`.

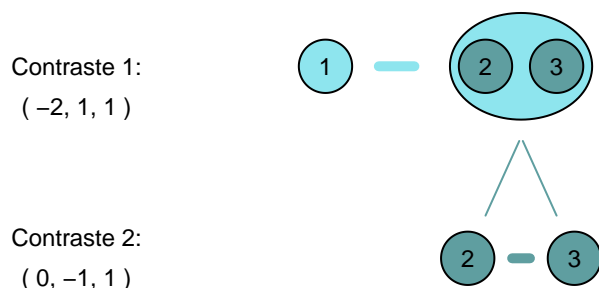
```
c1 <- c( -2,  1,  1 )
c2 <- c(  0, -1,  1 )
mat <- cbind( c1, c2 )
```



Los contrastes se realizan de dos en dos, es decir, uno de nuestros grupos será comparado a un conjunto comprendido por el resto de los grupos. Una vez que un grupo ha sido aisladamente comparado con un conjunto de grupos, el primero no vuelve a aparecer en los contrastes posteriores. En cada contraste se le asigna a cada grupo un peso que describe exactamente como será este contraste. Para asignar los factores se siguen las siguientes reglas:

- Si el grupo no interviene en el contraste se le asigna valor 0
- A los grupos que pertenecen a un conjunto se les dará un peso negativo y a los que pertenecen al otro se le darán pesos positivos.
- Para un contraste, los pesos asignados a cada grupo en cada conjunto debe ser igual al número de grupos en el conjunto opuesto. Es decir, en nuestro ejemplo primero comparamos el conjunto: raza 1 con el conjunto: raza 2 y raza 3. De forma que el grupo “raza 1” tendrá como valor -2 porque hay 2 grupos en el conjunto opuesto y los grupos “raza 2” y “raza 3” tendrán como valor 1 y 1 porque solo hay uno en el opuesto.
- Cada vector del contraste puede ser simplificado, es decir, si simplificamos dividiendo por algún divisor común el contraste seguirá siendo el mismo ( esto es, el vector (2,2,-2) será equivalente al vector (1,1,-1)). A. Field et al. (2012)

Lo podemos ver con un esquema:



A la variable independiente le tenemos que asignar la matriz del contraste mediante la función `contrasts()`.

```
contrasts( df$raza ) <- mat
```

Observamos en qué se ha transformado:

```
df$raza
## [1] 3 1 1 2 3 3 2 2 3 2 1 3 1 1 1 2 3 2
## attr("contrasts")
## c1 c2
## 1 -2 0
## 2 1 -1
## 3 1 1
## Levels: 1 2 3
```

Ahora el factor tiene dos contrastes atribuidos.

Por último, realizamos el análisis mediante la función `aov()` como habitualmente pero usamos la opción `split` cuando vamos a mostrar el `summary()`. Esta opción muestra la lista donde los resultados de los contrastes están guardados.

```
fit11<-aov( m0 ~ raza, data = df)
summary( fit11, split=list( raza = list( "C1" = 1, "C2" = 2 ) ) )

##           Df Sum Sq Mean Sq F value Pr(>F)
## raza      2    845     422    7.28 0.0062 **
## raza: C1   1    775     775   13.35 0.0024 **
```



```
## raza: C2 1 70 70 1.21 0.2891
## Residuals 15 870 58
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

que sugiere que hay diferencias significativas en el segundo contraste pero no en el primero.

También podemos usar la función

```
summary.lm( fit11 )
```

```
##
## Call:
## aov(formula = m0 ~ raza, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.83  -3.33   1.25   3.79  14.00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    43.78      1.80    24.38 1.8e-13 ***
## razac1         4.64      1.27     3.65 0.0024 **
## razac2         2.42      2.20     1.10 0.2891
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.62 on 15 degrees of freedom
## Multiple R-squared: 0.493, Adjusted R-squared: 0.425
## F-statistic: 7.28 on 2 and 15 DF, p-value: 0.00617
```



Volver al índice del curso Servicio de Apoyo a la Investigación, Universidad de Murcia

FEIR3

## Referencias y bibliografía

Antonio Valera Espín, J. S. M. (n.d.). Pruebas de significación y magnitud del efecto: Reflexiones y propuestas. Retrieved 1997, from [http://www.um.es/analesps/v13/v13\\_1/09-13-1.pdf](http://www.um.es/analesps/v13/v13_1/09-13-1.pdf)

Baron, J. (n.d.). Notes on the use of r for psychology experiments and questionnaires. Retrieved March 6, 2011, from <http://www.psych.upenn.edu/~baron/rpsych/rpsych.html>

bertolt. (n.d.). Obtaining the same anova results in r as in spss - the difficulties with type ii and type iii sums of squares. Retrieved May 24, 2008, from <http://myowelt.blogspot.com.es/2008/05/obtaining-same-anova-results-in-r-as-in.html>

Clay, R. (n.d.). Open science colaboration blog: Confidence intervals for effect sizes from noncentral distributions. Retrieved March 6, 2014, from <http://osc.centerforopenscience.org/2014/03/06/confidence%20intervals/>

explorable.com. (n.d.). Repeated measures anova. Retrieved May 25, 2009, from <https://explorable.com/>

repeated-measures-anova

Factor, T. A. (n.d.). Why anova and linear regression are the same analysis. Retrieved 2013, from <http://www.theanalysisfactor.com/why-anova-and-linear-regression-are-the-same-analysis/>

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using r* (1st edition.). Sage Publications Ltd.

Hall, R. (n.d.). Between-subject one-way anova description. Retrieved 1998, from <http://web.mst.edu/~psyworld/anovadescribe.htm>

Kabacoff, R. (2011). *R in action* (1st ed.). Shelter Island, NY: Manning Publications.

López, F. J. B. (n.d.). Apuntes y vídeos de bideoestadística. Retrieved 2014, from <http://www.bioestadistica.uma.es/baron/apuntes/>

Quick, J. M. (n.d.). R tutorial series anova pairwise comparison methods. Retrieved 2013, from <http://rtutorialseries.blogspot.com.es/2011/03/r-tutorial-series-anova-pairwise.html>

R, C. for. (n.d.). Cookbook for r. anova. Retrieved from [http://www.cookbook-r.com/Statistical\\_analysis/ANOVA/](http://www.cookbook-r.com/Statistical_analysis/ANOVA/)

Vilar, J. (n.d.). 4. chequeo y validación del modelo con un factor. independencia de los errores. Retrieved October 9, 2014, from [http://dm.udc.es/assignaturas/estadistica2/sec4\\_7.html](http://dm.udc.es/assignaturas/estadistica2/sec4_7.html)

Watson, P. (n.d.). Rules of thumb on magnitudes of effect sizes. Retrieved October 2, 2014, from <http://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/effectSize>