

Guía docente. Unidad 8 - 2da parte

Modelos jerárquicos con dos factores

En algunas situaciones experimentales donde intervienen dos o más factores no es posible cruzar todos los niveles de un factor con los niveles de otro factor sino que presentan una *jerarquía* donde cada nivel de uno de ellos se cruza solamente con algunos niveles del otro. En este caso se dirá que los dos factores no están cruzados sino que están *anidados*. A este tipo de diseño se denomina, sinónimamente, diseño jerárquico o diseño anidado, *nested* en inglés. Es importante saber qué factor está anidado y cuál es el principal, el orden es importante.

Un ejemplo de situación experimental en que no es posible cruzar los dos factores por razones prácticas sería la siguiente:

- Se desea comparar 3 tipos de pesticidas en la producción de árboles fruteros. Se divide el terreno en 9 parcelas homogéneas y se asignan al azar los 3 fertilizantes a cada parcela (3 parcelas por fertilizante). Como el subministro de los pesticidas se realiza vía aérea con avionetas, no es posible experimentar los 3 pesticidas dentro de cada parcela sino que se suministra el mismo pesticida a toda la parcela. Finalmente se eligen 5 árboles por parcela y se mide la producción de cada árbol en Kg.

En este ejemplo el factor parcela está anidado a pesticida y los árboles serían las réplicas ($n = 5$).

La ecuación del modelo se escribe de la siguiente manera:

$$y_{ijk} = \mu + \alpha_i + B_{j(i)} + e_{ijk}$$

donde $i = 1, 2, 3$ son los niveles del factor A (fertilizante), $j = 1, 2, 3$ son los niveles del factor B (parcela) para cada nivel del factor A y $k = 1, \dots, 5$ son las réplicas (muestras dentro de cada nivel del factor B). Destacaremos que el rango de j no va de 1 hasta 9, sino solamente hasta 3 y también el uso del paréntesis en el subíndice del efecto del factor anidado.

Hay que remarcar especialmente que no se puede introducir la interacción aunque tengamos réplicas en todas las celdas observables: **la interacción entre dos factores que no están cruzados no se puede incluir en el modelo.**

En este ejemplo el factor B (el que está anidado) es de efectos aleatorios mientras que el factor A es de efectos fijos. Ésta es la situación más común, pero son teóricamente posibles las 4 combinaciones, es decir, (A fijo - B aleatorio), (A aleatorio - B aleatorio), (A aleatorio - B fijo), (A fijo - B fijo). La tercera de ellas, (A aleatorio - B fijo), aunque puede estudiarse matemáticamente es inverosímil que se produzca en la práctica.

Para la obtención de los cocientes F será necesario disponer de las fórmulas de los E(MS) para cada modelo. Las fórmulas cuando el experimento es **balanceado** se derivan de las reglas de Bennett-Franklin. Las tablas para las combinaciones fijo-aleatorio y aleatorio-aleatorio, son:

- Tabla 1: A fijo, B aleatorio, anidado a A

Factor	E(MS)
A	$\sigma^2 + n\sigma_{B(A)}^2 + \frac{nb}{a-1} \sum \alpha_i^2$
B(A)	$\sigma^2 + n\sigma_{B(A)}^2$
Error	σ^2

- Tabla 2: A aleatorio, B aleatorio, anidado a A

Factor	E(MS)
A	$\sigma^2 + n\sigma_{B(A)}^2 + nb\sigma_A^2$
B(A)	$\sigma^2 + n\sigma_{B(A)}^2$
Error	σ^2

Veamos algunos ejemplos.

Ejemplo 1

Se quiere ensayar el efecto de dos fármacos anti-depresivos sobre pacientes que manifiestan síntomas graves de esta enfermedad. Se escogen 4 hospitales, de forma que cada hospital ensaya sólo una droga. Dentro de cada hospital se escogen 5 pacientes al azar. Los médicos miden el grado de efectividad del fármaco de acuerdo a una variable que recoge en un solo valor diferentes indicadores de mejoría entre el estado antes de iniciar el tratamiento y después del administración del fármaco. Se puede considerar que esta variable tiene distribución normal. Los datos obtenidos son:

Fármaco A		Fármaco B	
Hospital 1	Hospital 2	Hospital 1	Hospital 2
15	18	30	35
12	16	28	33
18	20	32	37
13	17	31	34
14	19	29	36

Los factores son:

- Fármaco: de efectos fijos con dos niveles
- Hospital: de efectos aleatorios, anidado a fármaco y con cuatro niveles en total o 2 niveles por fármaco

Los pacientes de cada hospital serían las réplicas ($n = 5$). Destacaremos que en la tabla se etiquetan cuatro columnas pero solamente dos títulos diferentes: *hospital 1* y *hospital 2*. Algunas veces la presentación de las tablas de datos pueden confundir. En este ejemplo en realidad el hospital 1 asignado al fármaco A es diferente del hospital 1 del asignado al B y, análogamente, el hospital 2 que aparece etiquetando en las columnas 2 y 4 se refiere a dos hospitales diferentes.

Modelo

$$y_{ijk} = \mu + \alpha_i + B_{j(i)} + e_{ijk}$$

$i = 1, 2$ fármaco, $j = 1, 2$ hospital, $k = 1, \dots, 5$.

Restricciones

$$\sum_{i=1}^2 \alpha_i = 0$$

Suposiciones

$B_{j(i)} \sim N(0, \sigma_B^2)$, $e_{ijk} \sim N(0, \sigma^2)$ variables aleatorias independientes

Entrada de datos

```
y <- c(15, 18, 30, 35, 12, 16, 28, 33,
      18, 20, 32, 37, 13, 17, 31, 34,
      14, 19, 29, 36)

hospital <- factor(rep(1:2, 10))
farmaco <- factor(rep(1:2, each=2, 5))
```

Los contrastes de interés son:

$$\begin{aligned}H_0 : \quad & \alpha_1 = \dots = \alpha_a = 0 \\H_1 : \quad & \alpha_i \neq \alpha_j \text{ para algún } i \neq j \\H_0 : \quad & \sigma_{B(A)}^2 = 0 \\H_1 : \quad & \sigma_{B(A)}^2 > 0\end{aligned}$$

Con la tabla de E(MS) anterior (Tabla 1), aplicando la regla comentada en unidades anteriores de explorar qué fuentes de variación coinciden en sus esperanzas de cuadrados medios bajo las hipótesis nulas de interés resultan estos dos cocientes F correspondientes a cada contraste:

$$F_A = \frac{MS_A}{MS_{B(A)}}$$

y

$$F_B = \frac{MS_B}{MS_E}$$

Tabla ANOVA por defecto

La tabla por defecto que resulta de la función `anova()` **no** proporciona los valores F correctos ya que en todos interviene MSE en el denominador, en consecuencia, tampoco son correctos los pvalores. Es necesario modificar la tabla de acuerdo con las fórmulas antes indicadas. Utilizamos la sintaxis `%in%` para indicar el anidamiento.

```
modelo <- aov(y ~ farmaco + hospital%in%farmaco)
tanova <- anova(modelo)
tanova
```

```
## Analysis of Variance Table
##
## Response: y
##              Df Sum Sq Mean Sq F value    Pr(>F)
## farmaco          1 1328.5  1328.45  415.141 7.184e-13 ***
## farmaco:hospital  2   94.9    47.45   14.828 0.0002275 ***
## Residuals       16   51.2     3.20
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Esta tabla es **incorrecta**, el cociente F para el fármaco es:

```
FA <- tanova[1,3]/tanova[2,3]
pvalueA <- pf(FA,tanova[1,1],tanova[2,1],lower.tail = F)
tanova[1,4] <- FA
tanova[1,5] <- pvalueA
```

no hace falta corregir FB <- tanova[2,3]/tanova[3,3] ya era correcto la división por MSE

Y obtenemos finalmente la tabla anova correcta:

```
tanova
```

```
## Analysis of Variance Table
##
## Response: y
##              Df Sum Sq Mean Sq F value    Pr(>F)
## farmaco          1 1328.5  1328.45   27.997 0.0339119 *
## farmaco:hospital  2   94.9    47.45   14.828 0.0002275 ***
## Residuals       16   51.2     3.20
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Aunque el factor fijo fármaco es significativo, no son necesarias las comparaciones múltiples pues sólo hay dos niveles.

Estimemos las componentes de la varianza aplicando el método de los momentos a la Tabla 1 de E(MS):

$$\hat{\sigma}^2 = MS_E = 3.20$$

$$\hat{\sigma}_{B(A)}^2 = \frac{MS_{B(A)} - \hat{\sigma}^2}{n} = \frac{47.45 - 3.20}{5} = 8.85$$

Ejemplo 2

En una práctica universitaria se quiere estudiar la variabilidad de la tensión arterial entre distintas camadas de ratas después de suministrar un fármaco hipotensor a la hembra unas horas antes del parto. Se organizan tres grupos de estudiantes, cada grupo escoge del estabulario 4 hembras al azar (12 hembras en total) que están en la fase final de su gestación. Una vez suministrado el fármaco, miden 3 ratas escogidas al azar de cada camada mediante un sensor no invasivo de presión sanguínea.

Cada grupo reporta por tanto 12 mediciones donde las camadas están anidadas al grupo y con 3 ratas=réplicas por celda experimental.

Los datos se muestran a continuación:

	Grupo 1			Grupo 2			Grupo 3		
Camada 1	124.0	118.4	118.7	120.6	120.1	119.8	120.1	120.3	119.9
Camada 2	118.7	119.0	118.2	117.1	117.9	117.9	118.3	117.3	117.2
Camada 3	117.2	117.9	118.3	119.5	119.6	120.3	121.5	121.2	121.1
Camada 4	121.3	121.2	121.2	121.2	120.4	120.9	122.8	122.5	121.8

Aunque la tabla etiqueta las filas con “Camada 1” hasta “Camada 4”, no tenemos 4 camadas sino 12, ya que las camadas son diferentes en cada grupo.

Los factores son:

- Grupo: de efectos aleatorios puesto que si se repitiera el experimento serían otros estudiantes. 3 niveles.
- Camada: de efectos aleatorios anidado a grupo, 4 por niveles por grupo, 12 niveles en total

Los ratones son una fuente de variabilidad, pero al tener una sola observación de cada ratón su efecto lo recoge el error experimental.

Modelo

$$y_{ijk} = \mu + A_i + B_{j(i)} + e_{ijk}$$

$i = 1, \dots, 3$ grupo, $j = 1, \dots, 4$ camada, $k = 1, \dots, 3$ ratón .

Restricciones

No hay, los dos factores son aleatorios

Suposiciones

$A_i \sim N(0, \sigma_A^2)$, $B_{j(i)} \sim N(0, \sigma_{B(A)}^2)$, $e_{ijk} \sim N(0, \sigma^2)$ independientes

Los contrastes de interés son:

$$\begin{aligned} H_0 : \quad & \sigma_A^2 = 0 \\ H_1 : \quad & \sigma_A^2 > 0 \end{aligned}$$

$$\begin{aligned} H_0 : \quad & \sigma_{B(A)}^2 = 0 \\ H_1 : \quad & \sigma_{B(A)}^2 > 0 \end{aligned}$$

Como anteriormente, la Tabla 2 de E(MS) nos proporciona implícitamente los cocientes F correspondientes a cada contraste que en este caso coinciden con los del ejemplo 1:

$$F_A = \frac{MS_A}{MS_{B(A)}}$$

y

$$F_B = \frac{MS_{B(A)}}{MS_E}$$

respectivamente.

Entrada de datos

```
y <- 115 + c(
9, 3.4, 3.7, 5.6, 5.1, 4.8, 5.1, 5.3, 4.9,
3.7, 4, 3.2, 2.1, 2.9, 2.9, 3.3, 2.3, 2.2,
2.2, 2.9, 3.3, 4.5, 4.6, 5.3, 6.5, 6.2, 6.1,
6.3, 6.2, 6.2, 6.2, 5.4, 5.9, 7.8, 7.5, 6.8)

camada <- factor(rep(1:4, each=9))
grupo <- factor(rep(1:3, each=3, 4))
```

Tabla anova por defecto

```
modelo2 <- aov(y ~ grupo + camada%in%grupo)
tanova2 <- anova(modelo2)
tanova2
```

```
## Analysis of Variance Table
##
## Response: y
##          Df Sum Sq Mean Sq F value    Pr(>F)
## grupo      2  4.865   2.4325   2.4640    0.1063
## grupo:camada 9 77.032   8.5591   8.6699 1.137e-05 ***
## Residuals  24 23.693   0.9872
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Los cocientes F correctos son

```
FA2 <- tanova2[1,3]/tanova2[2,3]
FA2
```

```
## [1] 0.2842013
```

```
# no hace falta corregir FB <- tanova[2,3]/tanova[3,3] ya es el cociente correcto
```

```
pf(FA2,tanova2[1,1],tanova2[2,1],lower.tail = F)
```

```
## [1] 0.7591259
```

A partir de la Tabla 2 se estiman las componentes de la varianza:

$$\hat{\sigma}^2 = MS_E = 0.9872222$$

$$\hat{\sigma}_{B(A)}^2 = \frac{MS_{B(A)} - \hat{\sigma}^2}{n} = \frac{8.5590741 - 0.9872222}{3} = 2.5239506$$

$$\hat{\sigma}_A^2 = \frac{MS_A - n\hat{\sigma}_{B(A)}^2 - \hat{\sigma}^2}{nb} = \frac{MS_A - MS_{B(A)}}{nb} = \frac{2.4325 - 8.5590741}{12} = -0.5105478$$

que asignamos a cero al estimar una varianza negativa. Finalmente, el porcentaje que representa la variabilidad entre camadas sobre el total

$$\frac{\hat{\sigma}_{B(A)}^2}{\hat{\sigma}_A^2 + \hat{\sigma}_{B(A)}^2 + \hat{\sigma}^2} \times 100 = 71.88\%$$

que es el porcentaje más relevante de la variabilidad total. En cambio los grupos de estudiantes proporcionan resultados muy concordantes al no presentar significación el factor grupo.

Modelos jerárquicos con tres factores

Ciertas condiciones experimentales dan lugar a diseños que contienen combinaciones de factores que están cruzados mientras que otros están anidados.

Vamos a considerar un diseño con tres factores, un factor A fijo con a niveles cruzado con un factor B fijo con b niveles, y un tercer factor C con c niveles aleatorio que está cruzado con A pero anidado a B . El modelo es:

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + C_{k(j)} + \alpha C_{ik(j)} + e_{ijkl}$$

$i = 1, \dots, a$, $j = 1, \dots, b$, $k = 1, \dots, c$, $l = 1, \dots, n$. Donde α_i , es el efecto fijo para el factor A, β_j es el efecto fijo del factor B, $\alpha\beta_{ij}$, es el efecto fijo para la interacción A con B, $C_{k(j)}$ es el efecto aleatorio de C anidado a B, $\alpha C_{ik(j)}$ es el efecto aleatorio para la interacción A con C, e_{ijkl} es el error experimental aleatorio. Las suposiciones para los efectos son coincidentes con las indicadas en unidades anteriores referentes a los modelos aleatorios, los mixtos y los anidados de dos factores. Adoptamos el modelo restringido para la interacción AC.

Ejemplo 3

Una empresa farmacéutica está en fase de desarrollo de un fármaco y le interesa evaluar el efecto de **tres dosis** distintas del fármaco en la mejoría de los pacientes. Se quieren ensayar además simultáneamente **tres formas** de tratamiento: la primera, suministrar el medicamento continuadamente durante 4 semanas, la segunda, administrar continuadamente durante 12 semanas y la tercera, también de duración total 12 semanas, pero alternando cada dos semanas la administración con un periodo de descanso de una semana. Disponen de **6 equipos médicos** de diferentes hospitales que colaboran en el estudio. Para simplificar la logística hospitalaria, cada equipo médico evaluará la eficacia de las tres dosis pero solamente **una forma** de administración. Cada equipo medico seleccionó 6 enfermos al azar y asignó al azar la dosis de tratamiento por grupos de dos, es decir, dos pacientes por cada nivel de dosis. Los médicos miden el grado de efectividad de la dosis de fármaco de acuerdo a una variable continua que refleja la mejora entre el estado antes de iniciar el tratamiento y después del tratamiento.

	Forma 1		Forma 2		Forma 3	
	Equipo 1	Equipo 2	Equipo 3	Equipo 4	Equipo 5	Equipo 6
Dosis 1	41.2	41.2	39.8	41.5	41.9	45.5
	42.6	41.4	40.3	43.0	42.7	44.7
Dosis 2	135.7	143.0	132.4	134.4	137.4	141.1
	136.8	143.3	130.3	130.0	135.2	139.1
Dosis 3	163.2	181.4	173.6	174.9	166.6	175.0
	163.3	180.3	173.9	175.6	165.5	172.0

Podemos analizar estos datos de acuerdo con el modelo anterior. El factor fijo *dosis* ($a=3$), que está cruzado con el factor fijo *forma* ($b=3$), el factor *equipo médico* ($c=2$) es aleatorio (por cierto, es muy habitual considerar los equipos humanos una muestra de un colectivo mucho mayor). El equipo está anidado a forma pero cruzado con dosis. Para cada dosis y cada equipo médico se analizan dos pacientes ($n=2$). Se puede estimar la interacción entre dosis y forma y entre dosis y equipo médico, pero no la interacción entre forma y equipo médico pues este último está anidado a forma.

Las hipótesis a contrastar serán:

- Para el factor fijo dosis, A:

$$\begin{aligned} H_0 : & \alpha_1 = \alpha_2 = \alpha_3 = 0 \\ H_1 : & \text{alguna } \alpha_i \neq 0 \end{aligned}$$

- Para el factor fijo forma, B:

$$\begin{aligned} H_0 : & \beta_1 = \beta_2 = \beta_3 = 0 \\ H_1 : & \text{alguna } \beta_i \neq 0 \end{aligned}$$

- Para la interacción dosis \times forma, AB:

$$\begin{aligned} H_0 : & \alpha\beta_{1,1} = \alpha\beta_{1,2} = \dots \alpha\beta_{3,3} = 0 \\ H_1 : & \text{alguna } \alpha\beta_{i,j} \neq 0 \end{aligned}$$

- Para el factor aleatorio equipo médico anidado a forma, C(B):

$$\begin{aligned} H_0 : & \sigma_C^2 = 0 \\ H_1 : & \sigma_C^2 > 0 \end{aligned}$$

- Para la interacción concentración \times equipo médico anidado a forma, AC(B):

$$\begin{aligned} H_0 : & \sigma_{AC}^2 = 0 \\ H_1 : & \sigma_{AC}^2 > 0 \end{aligned}$$

Entrada de datos

```
resp <- c(41.2, 41.2, 39.8, 41.5, 41.9, 45.5,
          42.6, 41.4, 40.3, 43.0, 42.7, 44.7,
          135.7, 143.0, 132.4, 134.4, 137.4, 141.1,
          136.8, 143.3, 130.3, 130.0, 135.2, 139.1,
          163.2, 181.4, 173.6, 174.9, 166.6, 175.0,
          163.3, 180.3, 173.9, 175.6, 165.5, 172.0)
dosis <- as.factor(rep(1:3, each=12))
equipo <- as.factor(rep(1:2, 9))
forma <- as.factor(rep(rep(1:3, each=2), 6))
mydata <- data.frame(resp, dosis, equipo, forma)
summary(mydata)
```

```
##      resp      dosis equipo forma
## Min.   : 39.80   1:12   1:18   1:12
## 1st Qu.: 42.92   2:12   2:18   2:12
## Median :136.25   3:12           3:12
## Mean   :116.94
## 3rd Qu.:165.78
## Max.   :181.40
```

Para no alargar esta guía asumiremos el cumplimiento de las condiciones de modelización y pasamos a calcular la tabla anova por defecto.

Nota cómo se codifican los niveles del factor anidado **equipo** a pesar de que hay 6 equipos distintos el código sus códigos van de 1 a 2, coincidiendo con el rango del subíndice k .

Aplicando el algoritmo de Bennett y Franklin se obtienen las esperanzas de los cuadrados medios (EMS)

Fuente	E(MS)
A	$\sigma^2 + n\sigma_{AC}^2 + \frac{bcn}{a-1} \sum \alpha_i^2$
B	$\sigma^2 + an\sigma_C^2 + \frac{acn}{b-1} \sum \beta_j^2$
AB	$\sigma^2 + n\sigma_{AC}^2 + \frac{cn}{(a-1)(b-1)} \sum \sum \beta_{ij}^2$
C(B)	$\sigma^2 + an\sigma_C^2$
AC(B)	$\sigma^2 + n\sigma_{AC}^2$
Error	σ^2

La tabla de las esperanzas de los cuadrados medios E(MS) permite conocer los términos que deberemos considerar en el cálculo de los cocientes F para resolver los contrastes del diseño.

Si queremos resolver el contraste para el factor fijo dosis (A), es decir, la hipótesis nula $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$ o de manera compacta $\theta_A^2 = 0$, el cociente F es de la forma:

$$F_A = \frac{MSA}{MSAC}$$

Si queremos resolver el contraste para el factor fijo forma (B), es decir, la hipótesis nula $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ el cociente F es de la forma:

$$F_B = \frac{MSB}{MSC}$$

Si queremos resolver el contraste para la interacción dosis \times forma (AB), el cociente F es de la forma:

$$F_{AB} = \frac{MSAB}{MSAC}$$

Si queremos resolver el contraste para el factor aleatorio equipo médico (C(B)), es decir, la hipótesis nula $H_0 : \sigma_{CB}^2 = 0$, el cociente F es de la forma:

$$F_C = \frac{MSC(B)}{MSE}$$

Si queremos resolver el contraste para la interacción concentración \times equipo médico (AC(B)), es decir, la hipótesis nula $H_0 : \sigma_{AC}^2 = 0$, el cociente F es de la forma:

$$F_{AC} = \frac{MSAC(B)}{MSE}$$

Asumiremos el cumplimiento de las condiciones de modelización y pasamos a calcular la tabla anova por defecto

```
res.aov<-aov(resp ~ dosis + forma + dosis:forma + equipo%in%forma +
             dosis:(equipo%in%forma),data=mydata)
tAnova<-anova(res.aov)
tAnova
```

```
## Analysis of Variance Table
##
## Response: resp
##              Df Sum Sq Mean Sq    F value    Pr(>F)
## dosis          2 108264    54132 37693.3290 < 2.2e-16 ***
## forma          2     25      12    8.6613 0.002317 **
## dosis:forma     4     176     44   30.7073 8.057e-08 ***
## forma:equipo    3     263     88   61.0689 1.240e-09 ***
## dosis:forma:equipo 6     180     30   20.9153 3.330e-07 ***
## Residuals     18      26      1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Esta tabla por defecto no proporciona los valores del estadístico F correctos para algunas filas y en consecuencia tampoco los pvalores. Es necesario modificar la tabla de acuerdo con las fórmulas antes indicadas.

```
FA    <- tAnova[1,3]/tAnova[5,3]
pvalA <- pf(FA,tAnova[1,1],tAnova[5,1],lower.tail = F)
FB    <- tAnova[2,3]/tAnova[4,3]
pvalB <- pf(FB,tAnova[2,1],tAnova[4,1],lower.tail = F)
FAB   <- tAnova[3,3]/tAnova[5,3]
```

```
pvalAB <- pf(FAB,tAnova[3,1],tAnova[5,1],lower.tail = F)
# FAC <-tAnova[5,3]/tAnova[6,3] no es necesario
tAnova[1,4] <- FA
tAnova[1,5] <- pvalA
tAnova[2,4] <- FB
tAnova[2,5] <- pvalB
tAnova[3,4] <- FAB
tAnova[3,5] <- pvalAB
tAnova
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: resp
```

```
##          Df Sum Sq Mean Sq  F value    Pr(>F)
## dosis      2 108264   54132 1802.1909 4.59e-09 ***
## forma      2     25      12    0.1418   0.8733
## dosis:forma  4    176      44    1.4682   0.3206
## forma:equipo  3    263      88   61.0689 1.24e-09 ***
## dosis:forma:equipo  6    180      30  20.9153 3.33e-07 ***
## Residuals   18     26       1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

En consecuencia sólo resultan significativos el factor **dosis**, el factor **equipo** y la interacción con el **equipo** y **forma**. Según a la tabla la **forma** no resulta significativa ni como factor principal ni interactuando con **dosis**. En la tabla ANOVA el término etiquetado como **forma:equipo** no es una interacción, R esta reportando las cantidades asociadas a **equipo** %in% **forma** o, si se prefiere, como **equipo(forma)** utilizando la terminología más habitual de los modelos anidados.

Comparaciones múltiples

Visto que el factor **dosis** es significativo nos podemos preguntar entre qué dosis hay diferencias. Veamos, en este diseño, cuál es la esperanza y la varianza de la variable respuesta y de qué depende:

$$E(y_{ijkl}) = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$$

Así, en promedio, la variable respuesta sólo depende de los dos factores fijos más su interacción. Como en este caso, en la parte fija del modelo, solamente es significativo el factor principal tiene sentido preguntarnos entre qué dosis hay diferencias sin tener en cuenta ni el factor **equipo** ni la interacción **dosis:equipo**.

Por otra parte

$$Var(y_{ijkl}) = \sigma_C^2 + \sigma_{AC}^2 + \sigma^2$$

La varianza, en cambio, es la suma de la varianza entre médicos, de la varianza de la interacción y de la varianza residual.

Veamos ahora cuál es la distribución (media y varianza) de la diferencia de medias muestrales entre el nivel 1 y 2 de la dosis. Hay que tener en cuenta que las restricciones habituales del modelo ANOVA (por ejemplo $\sum \beta_j = 0$) implican que se cancelen todos los términos β_j y $\alpha\beta_{ij}$ al promediar las observaciones dentro de cualquier nivel del factor A.

$$\bar{y}_{1\dots} = \mu + \alpha_1 + \frac{\sum_{j=1}^b \sum_{k=1}^c C_{k(j)}}{bc} + \frac{\sum_{j=1}^b \sum_{k=1}^c \alpha C_{1k(j)}}{bc} + \bar{\epsilon}_{1\dots}$$

$$\bar{y}_{2...} = \mu + \alpha_2 + \frac{\sum_{j=1}^b \sum_{k=1}^c C_{k(j)}}{bc} + \frac{\sum_{j=1}^b \sum_{k=1}^c \alpha C_{2k(j)}}{bc} + \bar{\epsilon}_{2...}$$

$$\bar{y}_{1...} - \bar{y}_{2...} = \alpha_1 - \alpha_2 + \frac{\sum_{j=1}^b \sum_{k=1}^c \alpha C_{1k(j)}}{bc} - \frac{\sum_{j=1}^b \sum_{k=1}^c \alpha C_{2k(j)}}{bc} + \bar{\epsilon}_{1...} - \bar{\epsilon}_{2...}$$

$$E(\bar{y}_{1...} - \bar{y}_{2...}) = \alpha_1 - \alpha_2$$

$$Var(\bar{y}_{1...} - \bar{y}_{2...}) = \frac{2\sigma_{AC}^2}{bc} + \frac{2\sigma^2}{bcn} = \frac{2}{bcn} (n\sigma_{AC}^2 + \sigma^2)$$

El término cuya esperanza de su cuadrado medio es $n\sigma_{AC}^2 + \sigma^2$ corresponde a la interacción `dosis:equipo`. Así, tenemos que especificar en la función `HSD.test` (si usamos la técnica de Tukey), que el término de error es `dosis:equipo`, pero cuidado con la notación de R por que la designa como `dosis:forma:equipo`.

```
library(agricolae)
HSD.test(mydata$resp, mydata$dosis,
         MSerror=tAnova['dosis:forma:equipo', 'Mean Sq'],
         DFerror=tAnova['dosis:forma:equipo', 'Df'],
         console=TRUE)

##
## Study: mydata$resp ~ mydata$dosis
##
## HSD Test for mydata$resp
##
## Mean Square Error:  30.03667
##
## mydata$dosis,  means
##
##   mydata.resp      std  r   Min   Max
## 1      42.1500 1.671961 12   39.8  45.5
## 2     136.5583 4.487449 12  130.0 143.3
## 3     172.1083 6.165809 12  163.2 181.4
##
## Alpha: 0.05 ; DF Error: 6
## Critical Value of Studentized Range: 4.339195
##
## Minimum Significant Difference: 6.865062
##
## Treatments with the same letter are not significantly different.
##
##   mydata$resp groups
## 3      172.1083      a
## 2      136.5583      b
## 1       42.1500      c
```

Vemos que las tres dosis presentan diferencias significativas entre sí y que la que proporciona una media más alta es la dosis=3, mientras que la de una media más baja es dosis=1.

Ejemplo 4

Se quiere comparar el grado de contaminación de tres zonas prefijadas con distintas características. En cada zona se eligen 3 puntos al azar. Para medir la contaminación en cada uno de los puntos, se usan 2 aparatos distintos en cada punto y se toman dos medidas para cada aparato. Debido a como están fabricados los aparatos de medición sólo se pueden tomar dos medidas y una vez tomadas quedan inhabilitados. Los datos en número de partículas finas por metro cúbico de aire fueron los siguientes:

	Punto 1		Punto 2		Punto 3	
	Aparato 1	Aparato 2	Aparato 1	Aparato 2	Aparato 1	Aparato 2
Zona 1	4.85, 4.99	6.43, 6.03	3.36, 3.19	4.24, 4.55	3.76, 3.72	4.51, 4.27
Zona 2	4.88, 5.10	5.85, 5.45	3.86, 3.50	4.63, 4.71	3.28, 3.41	4.70, 4.78
Zona 2	1.41, 1.34	2.70, 2.37	0.01, 0.15	1.31, 2.04	0.71, 0.72	1.16, 1.17

Destaquemos que las **zonas** son prefijadas, es decir, se trata de un factor de efectos **fijos** ya que nos interesa comparar aquellas tres zonas en concreto. Por otro lado, los **puntos**, al ser elegidos al azar, forman un factor de efectos **aleatorios** que está **anidado** a zona ya que los tres puntos de una zona son distintos a los de otra zona. Finalmente, los **aparatos** también son un factor de efectos **aleatorios**. El motivo es que, al no poderse tomar más de dos medidas por aparato, si repitiéramos el experimento tendríamos que usar otros aparatos distintos. Además, el factor aparato está **anidado** a punto: por el mismo motivo que antes un mismo aparato no puede ser usado en distintos puntos.

Así pues tenemos un diseño de tres factores:

- A: zona
- B(A): punto, aleatorio, anidado a zona
- C(B(A)): aparato, aleatorio, anidado a punto.

$$y_{ijkl} = \mu + \alpha_i + B_{j(i)} + C_{k(j(i))} + \epsilon_{ijkl}$$

$i = 1, 2, 3$ (zonas), $j = 1, 2, 3$ (puntos), $k = 1, 2$ (aparatos), $l = 1, 2$ (réplicas), con $\sum_{i=1}^3 \alpha_i = 0$ y $B_{j(i)} \sim N(0, \sigma_B^2)$, $C_{k(j(i))} \sim N(0, \sigma_C^2)$ y $\epsilon_{ijkl} \sim N(0, \sigma^2)$ independientes.

Como los tres factores están anidados en dos niveles de jerarquía no puede resolverse ninguna interacción entre ellos.

Entrada de datos

```
resp <- c(4.85, 4.99, 6.43, 6.03, 3.36, 3.19, 4.24, 4.55, 3.76, 3.72, 4.51, 4.27,
          4.88, 5.10, 5.85, 5.45, 3.86, 3.50, 4.63, 4.71, 3.28, 3.41, 4.70, 4.78,
          1.41, 1.34, 2.70, 2.37, 0.01, 0.15, 1.31, 2.04, 0.71, 0.72, 1.16, 1.17)
zona   <- factor(rep(1:3, each=12))
punto  <- factor(rep(rep(1:3, each=4), 3))
aparato <- factor(rep(rep(1:2, each=2), 9))
```

Aplicando las reglas de Bennett y Franklin obtenemos las esperanzas de los cuadrados medios (EMS). En el apéndice pueden consultarse los detalles

Fuente	E(MS)
A	$\sigma^2 + n\sigma_{C(B(A))}^2 + nc\sigma_{B(A)}^2 + \frac{nbc}{a-1} \sum \alpha_i^2$
B(A)	$\sigma^2 + n\sigma_{C(B(A))}^2 + nc\sigma_{B(A)}^2$
C(B(A))	$\sigma^2 + n\sigma_{C(B(A))}^2$
Error	σ^2

Obtengamos la tabla anova por defecto para posteriormente modificar los cocientes F.

```

modelo <- aov(resp ~ zona + punto%in%zona + aparato%in%(punto%in%zona))
tAnova <- anova(modelo)
tAnova

```

```

## Analysis of Variance Table
##
## Response: resp
##
##           Df Sum Sq Mean Sq  F value    Pr(>F)
## zona       2 84.221  42.111 1084.393 < 2.2e-16 ***
## zona:punto  6 14.022   2.337   60.179 6.185e-11 ***
## zona:punto:aparato 9 10.847   1.205   31.035 3.640e-09 ***
## Residuals  18  0.699   0.039
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Modifiquemos las Fs.

```

FA <- tAnova[1,3]/tAnova[2,3]
pvalA <- pf(FA, tAnova[1,1], tAnova[2,1], lower.tail = F)
FBinA <- tAnova[2,3]/tAnova[3,3]
pvalBinA <- pf(FBinA, tAnova[2,1], tAnova[3,1], lower.tail = F)
# no hace falta corregir FCinB <- tAnova[3,3]/tAnova[4,3], el cociente por defecto es correcto
tAnova[1,4] <- FA
tAnova[1,5] <- pvalA
tAnova[2,4] <- FBinA
tAnova[2,5] <- pvalBinA
tAnova

```

```

## Analysis of Variance Table
##
## Response: resp
##
##           Df Sum Sq Mean Sq F value    Pr(>F)
## zona       2 84.221  42.111 18.0194 0.002907 **
## zona:punto  6 14.022   2.337  1.9391 0.178581
## zona:punto:aparato 9 10.847   1.205 31.0353 3.64e-09 ***
## Residuals  18  0.699   0.039
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Los paréntesis que se utilizan en la nomenclatura de la fórmula del modelo no tienen correspondencia literal en el listado con R, por ejemplo `zona:punto:aparato` en el listado se refiere a `aparato(punto(zona))`.

La tabla corregida muestra diferencias significativas entre zonas. Por otra parte, hay variabilidad significativamente superior a cero entre aparatos pero no entre puntos de la misma zona.

Comparaciones múltiples

Como hemos visto diferencias significativas entre zonas nos podemos preguntar entre qué zonas. En este punto vamos a proceder de la misma forma que en el ejemplo 1 para averiguar cuál es el término de error para la diferencia de medias muestrales entre un par de zonas:

$$y_{ijkl} = \mu + \alpha_i + B_{j(i)} + C_{k(j(i))} + \epsilon_{ijkl}$$

$$\bar{y}_{1\dots} = \mu + \alpha_1 + \frac{\sum_{j=1}^b B_{j(1)}}{b} + \frac{\sum_{j=1}^b \sum_{k=1}^c C_{1(j(k))}}{bc} + \bar{\epsilon}_{1jkl}$$

$$\bar{y}_{2...} = \mu + \alpha_2 + \frac{\sum_{j=1}^b B_{j(2)}}{b} + \frac{\sum_{j=1}^b \sum_{k=1}^c C_{2(j(k))}}{bc} + \bar{\epsilon}_{2jkl}$$

$$\bar{y}_{1...} - \bar{y}_{2...} = \alpha_1 - \alpha_2 + \frac{\sum_{j=1}^b B_{j(1)}}{b} - \frac{\sum_{j=1}^b B_{j(2)}}{b} + \frac{\sum_{j=1}^b \sum_{k=1}^c C_{1(j(k))}}{bc} - \frac{\sum_{j=1}^b \sum_{k=1}^c C_{2(j(k))}}{bc} + \bar{\epsilon}_{1jkl} - \bar{\epsilon}_{2jkl}$$

$$E(\bar{y}_{1...} - \bar{y}_{2...}) = \alpha_1 - \alpha_2$$

$$Var(\bar{y}_{1...} - \bar{y}_{2...}) = \frac{2\sigma_B^2}{b} + \frac{2\sigma_C^2}{bc} + \frac{2\sigma^2}{bcn} = \frac{2}{bcn} (cn\sigma_B^2 + n\sigma_C^2 + \sigma^2)$$

Así pues, el término de error para las comparaciones múltiples corresponderá al factor punto.

```
HSD.test(resp,zona, MSerror=tAnova[2,3],
          DFEror=tAnova[2,1], console=TRUE)
```

```
##
## Study: resp ~ zona
##
## HSD Test for resp
##
## Mean Square Error:  2.336961
##
## zona,  means
##
##      resp      std  r  Min  Max
## 1 4.491667 0.9872997 12 3.19 6.43
## 2 4.512500 0.8253388 12 3.28 5.85
## 3 1.257500 0.8175476 12 0.01 2.70
##
## Alpha: 0.05 ; DF Error: 6
## Critical Value of Studentized Range: 4.339195
##
## Minimum Significant Difference: 1.914892
##
## Treatments with the same letter are not significantly different.
##
##      resp groups
## 2 4.512500      a
## 1 4.491667      a
## 3 1.257500      b
```

Tanto las zonas 1 como la 2 presentan diferencias significativas respecto a la 3, pero no entre sí. Estas dos zonas son las que tienen más contaminación mientras que la zona 3 es la que tiene menos.

Ejemplo 5

Se diseña un estudio para comparar tres fármacos que disminuyen la tensión arterial y, además, del eventual efecto del sexo del paciente. Para ello se eligieron a 18 pacientes hipertensos de cada sexo. A cada una de las 18 mujeres participantes se les asignó uno de los tres tratamientos al azar de tal forma que cada fármaco se ensayó en 6 de ellas, y, de la misma forma, se asignaron los hombres a los fármacos. Pasados tres meses se les midió la tensión arterial sistólica (TAS). Por limitaciones operativas para cada combinación de fármaco y sexo se eligieron 3 días al azar para medir en cada fecha la tensión de solamente 2 pacientes.

	Hombres			Mujeres		
	Dia 1	Dia 2	Dia 3	Dia 1	Dia 2	Dia 3
Fármaco A	186.1, 188.3	182.5, 182.9	183.8, 182.5	175.6, 175.1	170.5, 171.6	172.5, 172.4
Fármaco B	180.1, 179.3	176.5, 177.3	175.9, 174.5	182.3, 181.1	176.8, 177.1	175.9, 176.3
Fármaco C	180.8, 181.0	175.0, 174.8	175.4, 174.3	178.3, 179.9	175.8, 178.2	177.3, 177.3

En este diseño intervienen el factor fármaco y el factor sexo. Además, como el fármaco se experimenta en participantes de ambos sexos, estos dos factores estarán cruzados y podemos resolver en el modelo la interacción. De hecho interesa comprobar si la interacción es significativa, y si fuera este el caso, la diferencia entre los tres fármacos dependería del sexo. Por otra parte, al tomar fechas al azar para medir la tensión hay que tener en cuenta el factor día. Este factor es de efectos aleatorios, al contrario que los factores fármaco y sexo que son de efectos fijos. El factor día está anidado a la combinación de sexo y fármaco, muchas veces se abrevia esta situación indicando que esta *anidado a su interacción*.

Así el modelo será:

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + C_{k(ij)} + \epsilon_{ijkl}$$

con $i = 1, 2, 3$ (fármacos), $j = 1, 2$ (sexo), $k = 1, 2, 3$ (día) y $l = 1, 2$ (réplicas), $\sum_{i=1}^a \alpha_i = 0$, $\sum_{j=1}^b \beta_j = 0$, $\sum_{i=1}^a \alpha\beta_{ij} = 0 \quad \forall j$, $\sum_{j=1}^b \alpha\beta_{ij} = 0 \quad \forall i$, $C_{k(ij)} \sim N(0, \sigma_C^2)$ y $\epsilon_{ijkl} \sim N(0, \sigma^2)$ independientes.

Entrada de datos

```
resp <- c(186.1, 188.3, 182.5, 182.9, 183.8, 182.5, 175.6, 175.1, 170.5, 171.6, 172.5, 172.4,
          180.1, 179.3, 176.5, 177.3, 175.9, 174.5, 182.3, 181.1, 176.8, 177.1, 175.9, 176.3,
          180.8, 181.0, 175.0, 174.8, 175.4, 174.3, 178.3, 179.9, 175.8, 178.2, 177.3, 177.3)
farmaco <- factor(rep(1:3, each=12), labels=LETTERS[1:3])
sexo <- factor(rep(rep(1:2, each=6), 3), labels=c("Hombres", "Mujeres"))
dia <- as.factor(rep(rep(1:3, each=2), 6))
```

Aplicando las reglas de Bennett y Franklin obtenemos las esperanzas de los cuadrados medios (EMS). En el apéndice pueden consultarse los detalles

Fuente	E(MS)
A	$\sigma^2 + n\sigma_C^2 + bcn\theta_A^2$
B	$\sigma^2 + n\sigma_C^2 + acn\theta_B^2$
AB	$\sigma^2 + n\sigma_C^2 + cn\theta_{AB}^2$
C	$\sigma^2 + n\sigma_C^2$
Error	σ^2

Obtengamos la tabla anova por defecto.

```
modelo <- aov(resp ~ farmaco + sexo + farmaco:sexo + dia%in%(farmaco:sexo))
tAnova <- anova(modelo)
tAnova
```

```
## Analysis of Variance Table
##
## Response: resp
##           Df Sum Sq Mean Sq F value    Pr(>F)
## farmaco      2  10.722    5.361   8.5018 0.002514 **
## sexo         1  90.250   90.250 143.1278 5.299e-10 ***
## farmaco:sexo  2 305.052 152.526 241.8912 9.836e-14 ***
## farmaco:sexo:dia 12 154.457  12.871  20.4128 4.810e-08 ***
## Residuals    18  11.350    0.631
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

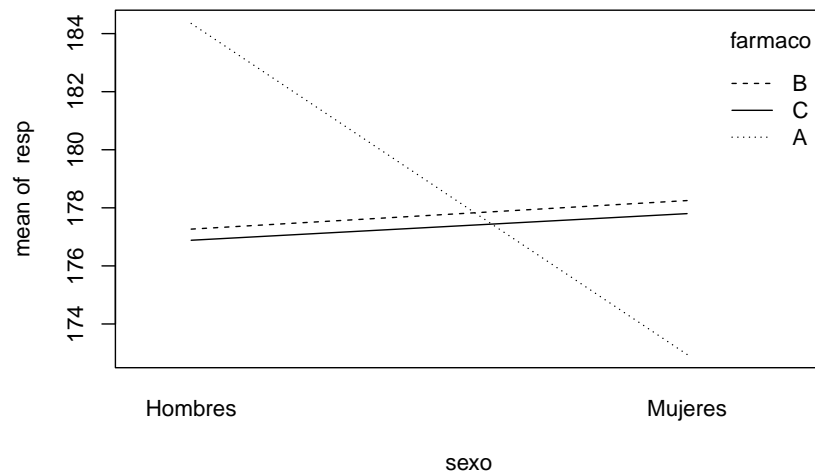
Los cocientes F se modifican de acuerdo con los cuadrados medios esperados de la tabla anterior

```
FA      <- tAnova[1,3]/tAnova[4,3]
pvalA   <- pf(FA,tAnova[1,1],tAnova[4,1],lower.tail = F)
FB      <- tAnova[2,3]/tAnova[4,3]
pvalB   <- pf(FB,tAnova[2,1],tAnova[4,1],lower.tail = F)
FAB     <- tAnova[3,3]/tAnova[4,3]
pvalAB  <- pf(FAB,tAnova[3,1],tAnova[4,1],lower.tail = F)
tAnova[1,4] <- FA
tAnova[1,5] <- pvalA
tAnova[2,4] <- FB
tAnova[2,5] <- pvalB
tAnova[3,4] <- FAB
tAnova[3,5] <- pvalAB
tAnova
```

```
## Analysis of Variance Table
##
## Response: resp
##           Df Sum Sq Mean Sq F value    Pr(>F)
## farmaco      2  10.722    5.361   0.4165 0.668531
## sexo         1  90.250   90.250   7.0117 0.021259 *
## farmaco:sexo  2 305.052 152.526 11.8500 0.001442 **
## farmaco:sexo:dia 12 154.457  12.871  20.4128 4.81e-08 ***
## Residuals    18  11.350    0.631
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

La interacción sexo-fármaco es significativa con lo cual no se pueden interpretar el efecto del fármaco y del sexo por separado. Puede ayudar a la interpretación de este resultado el siguiente gráfico de interacción:

```
interaction.plot(sexo, farmaco, resp)
```

Los fármacos B y C parecen muy similares tanto en hombres como en mujeres. Para los hombres los mejores fármacos serían indistintamente el C y el B, mientras que para las mujeres parece que el mejor fármaco es el A. Recordad que como la variable respuesta es la tensión arterial sistólica, cuanto más bajo su valor, mejor.

Podríamos efectuar comparaciones múltiples para ver qué fármacos son mejores para cada sexo. Así, se trata de comparar las distintas combinaciones de sexo y fármaco. Visto así, se trataría de un diseño de un factor cuyos niveles son las 6 combinaciones de sexo y fármaco y un factor anidado a éste que sería el participante. Este tipo de diseños se han visto anteriormente, donde vimos que el término de error era el factor anidado, en este caso el participante.

```
comb <- interaction(farmaco, sexo)
modelo2F <- aov(resp ~ comb + dia%in%comb)
tAnova2F <- anova(modelo2F)
tAnova2F
```

```
## Analysis of Variance Table
##
## Response: resp
##          Df Sum Sq Mean Sq F value    Pr(>F)
## comb         5 406.02   81.205 128.783 1.933e-13 ***
## comb:dia     12 154.46   12.871  20.413 4.810e-08 ***
## Residuals    18  11.35    0.631
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Modificamos el cociente F de acuerdo a la tabla 1 de la sección de modelos jerárquicos de dos factores.

```
FA <- tAnova2F[1,3]/tAnova2F[2,3]
pvalueA <- pf(FA,tAnova2F[1,1],tAnova2F[2,1],lower.tail = F)
# FB <- tAnova2F[2,3]/tAnova2F[3,3] # ya era correcto
tAnova2F[1,4]<-FA
tAnova2F[1,5]<-pvalueA
tAnova2F
```

```
## Analysis of Variance Table
##
## Response: resp
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## comb          5 406.02  81.205   6.3089 0.00429 **
## comb:dia     12 154.46  12.871  20.4128 4.81e-08 ***
## Residuals    18  11.35   0.631
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

rescomp <- HSD.test(resp, comb,
                    MSError=tAnova2F[2,3],
                    DFEror=tAnova2F[2,1], console=TRUE,
                    group=FALSE)

##
## Study: resp ~ comb
##
## HSD Test for resp
##
## Mean Square Error:  12.87139
##
## comb,  means
##
##           resp      std r   Min   Max
## A.Hombres 184.3500 2.362837 6 182.5 188.3
## A.Mujeres 172.9500 1.998750 6 170.5 175.6
## B.Hombres 177.2667 2.110608 6 174.5 180.1
## B.Mujeres 178.2500 2.730385 6 175.9 182.3
## C.Hombres 176.8833 3.132039 6 174.3 181.0
## C.Mujeres 177.8000 1.365284 6 175.8 179.9
##
## Alpha: 0.05 ; DF Error: 12
## Critical Value of Studentized Range: 4.750231
##
## Comparison between treatments means
##
##           difference pvalue signif.      LCL      UCL
## A.Hombres - A.Mujeres 11.400000 0.0015    **   4.4425218 18.357478
## A.Hombres - B.Hombres  7.083333 0.0452     *   0.1258552 14.040811
## A.Hombres - B.Mujeres  6.100000 0.0990     .  -0.8574782 13.057478
## A.Hombres - C.Hombres  7.466667 0.0331     *   0.5091885 14.424145
## A.Hombres - C.Mujeres  6.550000 0.0694     .  -0.4074782 13.507478
## A.Mujeres - B.Hombres -4.316667 0.3555  -11.2741448  2.640811
## A.Mujeres - B.Mujeres -5.300000 0.1817  -12.2574782  1.657478
## A.Mujeres - C.Hombres -3.933333 0.4469  -10.8908115  3.024145
## A.Mujeres - C.Mujeres -4.850000 0.2502  -11.8074782  2.107478
## B.Hombres - B.Mujeres -0.983333 0.9962   -7.9408115  5.974145
## B.Hombres - C.Hombres  0.383333 1.0000   -6.5741448  7.340811
## B.Hombres - C.Mujeres -0.533333 0.9998   -7.4908115  6.424145
## B.Mujeres - C.Hombres  1.366667 0.9833   -5.5908115  8.324145
## B.Mujeres - C.Mujeres  0.450000 0.9999   -6.5074782  7.407478
## C.Hombres - C.Mujeres -0.916667 0.9973   -7.8741448  6.040811
```

Destacamos el uso del argumento `group=FALSE` con el que se muestran todas las comparaciones dos a dos. De todas ellas sólo nos interesan aquellas que comparan dos fármacos distintos dentro del mismo sexo.

- Hombres

```
library(knitr)
tabla <- rescomp$comparison
rn <- rownames(tabla)
x <- kable(tabla[grepl("[A,B,C]\\..Hombres - [A,B,C]\\..Hombres",rn)],,
           "latex",booktabs=T,align='c',
           digits = 4)%>%kable_styling(position = "center",latex_options = "hold_position")
column_spec(x,column=c(1,2,4), border_right=T)
```

	difference	pvalue	signif.	LCL	UCL
A.Hombres - B.Hombres	7.0833	0.0452	*	0.1259	14.0408
A.Hombres - C.Hombres	7.4667	0.0331	*	0.5092	14.4241
B.Hombres - C.Hombres	0.3833	1.0000		-6.5741	7.3408

Comprobamos que el fármaco A es significativamente distinto de los otros dos, mientras que B y C no presentan diferencias significativas entre ellos.

- Mujeres

```
tabla <- rescomp$comparison
rn <- rownames(tabla)
x <- kable(tabla[grepl("[A,B,C]\\..Mujeres - [A,B,C]\\..Mujeres",rn)],,
           "latex",booktabs=T,align='c',
           digits = 4)%>%kable_styling(position = "center",latex_options = "hold_position")
column_spec(x,column=c(1,2,4), border_right=T)
```

	difference	pvalue	signif.	LCL	UCL
A.Mujeres - B.Mujeres	-5.30	0.1817		-12.2575	1.6575
A.Mujeres - C.Mujeres	-4.85	0.2502		-11.8075	2.1075
B.Mujeres - C.Mujeres	0.45	0.9999		-6.5075	7.4075

En las mujeres, en cambio, los fármacos A y C no difieren entre ellos.

Esquema de posibles diseños con tres factores

En un diseño con tres factores existen sólo 4 posibilidades de relación jerárquica entre ellos:

- A, B, C todos cruzados
- A, B(A), C: ejemplo 3
- A, B(A), C(B(A)): ejemplo 4
- A, B, C(AB): ejemplo 5
- A, B(A), C(A)

Apéndice: obtención de la tabla E(MS) mediante Bennett i Franklin

La confección de las tablas de las esperanzas de los cuadrados medios a partir del algoritmo de Bennett i Franklin está fuera de nuestro temario. Las incluimos aquí meramente como material de referencia para aquellos que quieran estudiar los detalles de como aplicar el algoritmo. Ahora bien, nuestro temario sí que incluye saber deducir los **cocientes F** oportunos a cada caso una vez se ha proporcionado ya calculada la tabla de esperanzas.

Ejemplo 1

Apliquemos el algoritmo de Bennett-Franklin para mostrar su desarrollo en la obtención de la Tabla 1 para el caso A fijo, B aleatorio, anidado a A, antes indicada.

1. Se escribe el modelo lineal

$$y_{ijk} = \mu + \alpha_i + B_{j(i)} + e_{k(ij)}$$

2. Se construye una tabla de dos factores con:
 - (a) Una fila para cada término del modelo, excluyendo a μ , etiquetada con el término del modelo
 - (b) Una columna para cada subíndice usado en el modelo.
3. En la parte superior de cada subíndice de columna se escribe el número de niveles del factor correspondiente y se escribe 'R' si el factor es aleatorio y 'F' si es fijo.
4. Se añade una columna cuyos elementos son la componente fija o aleatoria de la varianza para el efecto representado por esa fila de la tabla. Los términos correspondientes a los efectos fijos se suelen simplificar con la notación θ^2 .

		F	R	R	
		a	b	n	
Fuente		i	j	k	Componente
A	α_i				θ_A^2
B(A)	$B_{j(i)}$				$\sigma_{B(A)}^2$
Error	$e_{k(ij)}$				σ^2

5. Para cada fila, si no aparecen los subíndices de la columna en el efecto de la fila, se introduce el número de niveles correspondientes al subíndice.

		F	R	R	
		a	b	n	
Fuente		i	j	k	Componente
A	α_i		b	n	θ_A^2
B(A)	$B_{j(i)}$			n	$\sigma_{B(A)}^2$
Error	$e_{k(ij)}$				σ^2

6. Si un subíndice está entre paréntesis en el efecto de la fila, se coloca un 1 en celdas bajo esos subíndices en el paréntesis.

		F	R	R	
		a	b	n	
Fuente		i	j	k	Componente
A	α_i		b	n	θ_A^2
B(A)	$B_{j(i)}$	1		n	$\sigma_{B(A)}^2$
Error	$e_{k(ij)}$	1	1		σ^2

7. a. Para cada renglón, si ahora el subíndice concuerda con el subíndice de la columna, se introduce un 0 si la columna representa un factor fijo F y existe una componente fija de la varianza para el efecto representado por el renglón.
- b. Se coloca un 1 en el resto de las celdas.

		F	R	R	
		a	b	n	
Fuente		i	j	k	Componente
A	α_i	0	b	n	θ_A^2
B(A)	$B_{j(i)}$	1	1	n	$\sigma_{B(A)}^2$
Error	$e_{k(ij)}$	1	1	1	σ^2

8. Para determinar el cuadrado medio esperado para una fuente específica de variación:
 - a. De las componentes de la varianza restantes se incluyen sólo aquellas cuyos términos correspondientes en el modelo incluyen subíndices del efecto en consideración.
 - b. Se ocultan las columnas que contienen subíndices sin paréntesis para el efecto en consideración.
 - c. El coeficiente para cada componente en la $E(CM)$ es el producto de las columnas restantes del renglón de ese efecto.
9. Obtenemos las expresiones para los $E(CM)$ de cada fuente.

		F	R	R		
		a	b	n		
Fuente		i	j	k	Componente	E(CM)
A	α_i	0	b	n	θ_A^2	$\sigma^2 + n\sigma_{B(A)}^2 + bn\theta_A^2$
B(A)	$B_{j(i)}$	1	1	n	$\sigma_{B(A)}^2$	$\sigma^2 + n\sigma_{B(A)}^2$
Error	$e_{k(ij)}$	1	1	1	σ^2	σ^2

Ejemplo 4

La tabla de esperanzas desglosada a continuación corresponde a la del ejemplo 4, con tres factores anidados A, B(A), C(B(A))

		F	R	R	R		
		a	b	c	n		
Fuente		i	j	k	l	Componente	E(MS)
A	α_i	0	b	c	n	θ_A^2	$\sigma^2 + n\sigma_{C(B(A))}^2 + nc\sigma_{B(A)}^2 + bcn\theta_A^2$
B(A)	$B_{j(i)}$	1	1	c	n	$\sigma_{B(A)}^2$	$\sigma^2 + n\sigma_{C(B(A))}^2 + nc\sigma_{B(A)}^2$
C(B(A))	$C_{k(j(i))}$	1	1	1	n	$\sigma_{C(B(A))}^2$	$\sigma^2 + n\sigma_{C(B(A))}^2$
Error	$e_{l(ijk)}$	1	1	1	1	σ^2	σ^2

Ejemplo 5

Finalmente la tabla de esperanzas a continuación corresponde a la del ejemplo 5, con tres factores anidados A, B, C(AB), los dos primeros fijos, el último aleatorio

		F	F	R	R		
		a	b	c	n		
Fuente		i	j	k	l	Componente	E(MS)
A	α_i	0	b	c	n	θ_A^2	$\sigma^2 + n\sigma_C^2 + bcn\theta_A^2$
B	β_j	a	0	c	n	θ_B^2	$\sigma^2 + n\sigma_C^2 + acn\theta_B^2$
AB	$(\alpha\beta)_{ij}$	0	0	c	n	θ_{AB}^2	$\sigma^2 + n\sigma_C^2 + acn\theta_{AB}^2$
C	$C_{k(ij)}$	1	1	1	n	σ_C^2	$\sigma^2 + n\sigma_C^2$
Error	$e_{l(ijk)}$	1	1	1	1	σ^2	σ^2

Materiales de estudio

Tema-Fuente	Libro Oehlert
8.1 a 8.4	Páginas de 279 a 288
8.7	Páginas de 298 a 304

Actividades complementarias

Resolver la actividad no evaluable de la unidad 8 publicada en el campus de la asignatura.