
Análisis de la varianza (ANOVA)

PID_00276234

Ferran Reverter



Universitat
Oberta
de Catalunya

Ferran Reverter

El encargo y la creación de este recurso de aprendizaje UOC han sido coordinados por la profesora: Teresa Sancho Vinuesa

Primera edición: septiembre 2020
© de esta edición, Fundació Universitat Oberta de Catalunya (FUOC)
Av. Tibidabo, 39-43, 08035 Barcelona
Autoría: Ferran Reverter
Producción: FUOC
Todos los derechos reservados

Ninguna parte de esta publicación, incluido el diseño general y la cubierta, puede ser copiada, reproducida, almacenada o transmitida de ninguna forma, ni por ningún medio, sea este eléctrico, mecánico, óptico, grabación, fotocopia, o cualquier otro, sin la previa autorización escrita del titular de los derechos.

Índice

1. Análisis de la varianza de un factor.....	5
1.1. Diseño completamente aleatorizado	6
1.2. Modelo estadístico	8
1.3. Estimación de los efectos	10
1.4. Mecánica del ANOVA	10
1.5. Tabla ANOVA	11
1.5.1. ¿Por qué funciona el ANOVA?	12
1.6. Contrastes	14
1.7. Contrastes múltiples	16
1.7.1. Test de Bonferroni	16
1.8. Comprobación de las suposiciones del modelo	18
1.8.1. Evaluando el cumplimiento de los supuestos de modelización	19
1.8.2. Evaluando la normalidad	21
1.8.3. Evaluando el cumplimiento de la constancia de la varianza (homocedasticidad)	22
1.8.4. Corrigiendo el incumplimiento de las condiciones	23
1.8.5. Formulación del modelo estadístico para analizar los datos del experimento anterior	25
1.8.6. Evaluación de las suposiciones de modelización (normalidad y homocedasticidad)	25
1.8.7. Transformación de la variable respuesta (número bacterias/cm ²) para proseguir con el análisis de los datos	27
1.8.8. Verificación de la adecuación	28
1.8.9. Evaluar si hay diferencias significativas entre los tratamientos	29
1.8.10. Estimación de los parámetros del modelo	29
1.8.11. ¿Se puede recomendar un tratamiento que mejore los tratamientos usuales?	30
1.9. Test de remuestreo para el ANOVA	31
1.10. Test no paramétrico basado en rangos	33
2. Análisis de la varianza de dos factores.....	35
2.1. Efectos principales e interacción	35
2.2. Parametrización del modelo lineal de dos factores	39
2.3. Análisis de la varianza en diseños balanceados	40
2.4. Diseños con efectos aleatorios	46
Bibliografía.....	49

1. Análisis de la varianza de un factor

Un experimento se caracteriza por las unidades experimentales bajo observación, los tratamientos a los que se verán expuestas, el modo como los tratamientos se asignan a las unidades y las mediciones de la variable respuesta en cada una de las unidades. Los estudios experimentales, a diferencia de los estudios observacionales, permiten la comparación de manera controlada de tratamientos sobre una variable respuesta de interés, siendo ésta aquella variable numérica que se pretende modelizar. En los estudios observacionales también hay tratamientos, pero la asignación de cada unidad experimental a un tratamiento no está controlada por el investigador.

Estudios experimentales se presentan a menudo en la investigación biomédica.

Los estudios observacionales son aquellos en los que los investigadores observan el efecto de un factor de riesgo, una prueba de diagnóstico, un tratamiento u otra intervención sin tratar de cambiar quién está o no expuesto a él. Los estudios de cohortes y los estudios de casos y controles son dos tipos de estudios observacionales.

1) Estudio de cohorte: para fines de investigación, una cohorte es cualquier grupo de personas que están vinculadas de alguna manera. Por ejemplo, una cohorte de nacimiento incluye a todas las personas nacidas dentro de un periodo de tiempo determinado. Los investigadores comparan lo que les sucede a los miembros de la cohorte que han estado expuestos a una variable particular con lo que les sucede a los otros miembros que no han estado expuestos.

2) Estudio de casos y controles: aquí los investigadores identifican a las personas con un problema de salud existente («casos») y un grupo similar sin el problema («controles») y luego los comparan con respecto a una exposición o exposiciones.

A veces, los estudios observacionales son el único modo como los investigadores pueden explorar ciertas preguntas. Por ejemplo, en un estudio de condiciones de seguridad en el trabajo, no sería ético diseñar un ensayo controlado aleatorio que exponga deliberadamente a los trabajadores a una situación potencialmente dañina. Sin embargo, los resultados de los estudios observacionales están, por su naturaleza, abiertos a controversias. Corren el riesgo de contener sesgos de confusión. Ejemplo: un estudio de cohorte podría encontrar que las personas que hacían yoga regularmente eran menos propensas a las

Ejemplo de estudio experimental

Comparar el efecto de tres tratamientos adelgazantes entre grupos de ratones de laboratorio con obesidad. Una variable respuesta de interés puede ser el porcentaje de tejido adiposo después del tratamiento. En un estudio experimental el equipo investigador puede garantizar la homogeneidad de los grupos que van a ser tratados, con el ánimo de reducir posibles causas que interfieran en la respuesta más allá del factor tratamiento.

enfermedades cardíacas que las que no lo practicaban. Pero el vínculo puede explicarse por el hecho de que las personas que realizan yoga también hacen más ejercicio y siguen dietas más saludables. En otras palabras, aunque una cohorte se define por una característica o exposición común, también pueden compartir otras características equívocas que afectan al resultado.

Un experimento es aleatorizado (*randomized*, en inglés) si el método de asignación de los tratamientos a las unidades experimentales implica un conocido esquema probabilístico. Como veremos más adelante, un experimento puede tener varias características aleatorias, además de la asignación de los tratamientos a las unidades. La aleatorización es uno de los elementos más importantes de un experimento bien diseñado. Es necesario distinguir entre un esquema casual y aleatorizado de asignación de los tratamientos. La introducción de más aleatoriedad en un experimento puede parecer una medida perversa. Después de todo, siempre estamos luchando contra el error aleatorio experimental. Sin embargo, la asignación aleatoria de tratamientos tiene dos consecuencias útiles:

- La aleatorización protege contra los factores de confusión.
- La aleatorización puede constituir la base para la inferencia.

La aleatorización se utiliza relativamente poco para la inferencia en la práctica, debido principalmente a los costes y las dificultades computacionales, pero en la actualidad su uso va en aumento, acorde con el auge de la computación. Sin embargo, el éxito de la aleatorización en la protección contra confusión es tan abrumador que es casi universalmente recomendada. Definimos la confusión como algo que ocurre cuando el efecto de un factor o tratamiento no puede ser distinguido del de otro factor o tratamiento.

1.1. Diseño completamente aleatorizado

El diseño completamente aleatorizado es el diseño más simple para comparar varios tratamientos, es decir, un factor con varios niveles. En general, en el ANOVA de un factor, los distintos tratamientos corresponden a los distintos niveles del factor.

Por ejemplo, en la comparación de un fármaco para reducir la tensión arterial podemos estar interesados en comparar el efecto del fármaco a cuatro concentraciones distintas, diríamos entonces que tenemos el factor fármaco a cuatro niveles.

Tenemos N unidades experimentales (ya sean personas, ratones, piezas...) y g tratamientos o niveles de un factor. El esquema de un experimento completamente aleatorizado consiste en:

- Seleccionar los tamaños muestrales n_1, n_2, \dots, n_g con $n_1 + n_2 + \dots + n_g = N$.
- Elegir n_1 unidades aleatoriamente para el tratamiento 1, elegir n_2 unidades aleatoriamente de entre las $N - n_1$ para el tratamiento 2, y así sucesivamente.

Este procedimiento da como resultado un diseño completamente aleatorizado; todas las disposiciones posibles de las N unidades en g grupos con tamaños n_1 hasta n_g son igualmente probables. En caso que $n_1 = n_2 = \dots = n_g = n$, diremos que el diseño es balanceado.

La aleatorización completa solo se refiere a la asignación de tratamientos a las unidades; en el diseño del experimento también se requieren la selección de los tratamientos, las unidades experimentales y la variable respuesta.

Ejemplo 1

Un factor importante en la selección de software para sistemas de procesamiento de textos y gestión de bases de datos es el tiempo requerido para aprender a manejar un sistema en particular. Con el fin de evaluar tres sistemas de gestión de bases de datos, una empresa diseñó una prueba para ver cuántas horas de capacitación se necesitaban para que los operadores fueran competentes en cada uno de los tres sistemas. Para ello, se seleccionaron dieciocho operadores, que se asignaron al azar cada uno de ellos a uno de los sistemas de software, por lo que quedaron seis operadores por sistema. Los tiempos se recopilan en la siguiente tabla.

Sistema 1	20	17	15	19	14	13
Sistema 2	18	17	14	29	13	12
Sistema 3	23	23	20	21	19	20

Empezamos con una descriptiva con la que visualizamos la distribución de los tiempos según el tipo de software.

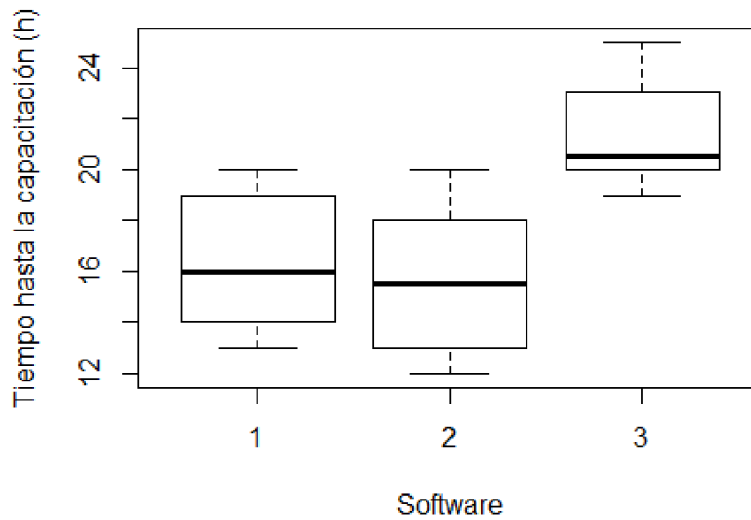
```
y<- c(20,17,15,19,14,13,
      18,17,14,20,13,12,
      23,25,20,21,19,20)
fact<-as.factor(c(rep(1,6),rep(2,6),rep(3,6)))
df<-data.frame(y,fact)
```

Obtenemos la media de los datos:

```
mean(y,data=df)
## [1] 17.77778
```

y con la función `tapply` obtendremos las medias dentro de cada tratamiento:

```
tapply(y,fact,mean,data=df)
##      1      2      3
## 16.33333 15.66667 21.33333
boxplot(y~fact,data=df,xlab="Software",ylab="Tiempo hasta la capacitación (h)")
```



Observamos, a nivel descriptivo, que aparentemente hay un efecto del factor, tipo de software, sobre el tiempo necesario para adquirir capacitación. El software tipo 3 requiere un esfuerzo mayor para adquirir capacitación. En cambio, parece que los tipos 1 y 2 requieren un esfuerzo en tiempo parecido. Es objetivo del diseño y análisis de experimentos verificar si estas diferencias que observamos son significativas y, en caso de serlo, determinar cuáles son los niveles significativamente distintos y obtener la estimación de los efectos sobre la variable respuesta.

1.2. Modelo estadístico

Empecemos con dos modelos básicos. Tenemos g tratamientos y N unidades experimentales. Denotemos por y_{ij} la respuesta j -ésima en el tratamiento i -ésimo. Por tanto, $i = 1, \dots, g$, y $j = 1, \dots, n_i$.

El modelo completo considera que cada tratamiento tiene su media μ_i . Además, se asume que el error aleatorio es independiente y distribuido normalmente con varianza constante, pero cada tratamiento con su propia media:

$$y_{ij} \sim N(\mu_i, \sigma^2)$$

Equivalentemente, podemos escribir el modelo como:

$$y_{ij} = \mu_i + e_{ij}$$

donde e_{ij} denotan los **errores** que son variables aleatorias independientes, con distribución normal de media cero y varianza σ^2 . Un segundo modelo corresponde a un modelo reducido. Este modelo asume que todos los tratamientos tienen la misma media, por lo que hay una única media. También considera que el error aleatorio es independiente y distribuido normalmente con varianza constante,

$$y_{ij} \sim N(\mu, \sigma^2)$$

Equivalentemente, podemos escribir el modelo como:

$$y_{ij} = \mu + e_{ij}$$

donde e_{ij} denotan los **errores** que son variables aleatorias independientes, con distribución normal de media cero y varianza σ^2 . Observad que el modelo reducido es un caso especial o restringido del modelo completo que corresponde al caso de que todas las μ_i son iguales.

Usualmente, la media de los tratamientos se expresa mediante:

$$\mu_i = \mu + \alpha_i$$

donde μ se denomina media general y α_i denota el efecto del tratamiento i -ésimo. Con esta formulación el modelo reducido corresponde al caso en el que todos los α_i son iguales, por ejemplo a cero. Esta formulación puede parecer ahora una complicación innecesaria pero más adelante, en el caso de los modelos factoriales generales, veremos que su uso es conveniente.

Dado que hay g medias μ_i para los g tratamientos, puede parecer que con la formulación de los efectos de los tratamientos hay un exceso de parámetros, $g + 1$, la μ y las α_i , con $i = 1, \dots, g$. No obstante, observamos que la diferencia entre efectos $\alpha_i - \alpha_j$ corresponderá a la diferencia entre las medias $\mu_i - \mu_j$, con independencia de cómo se defina μ .

La elección más usual para μ es la media de la media de los tratamientos:

$$\mu = \frac{1}{a} \sum_{i=1}^a \mu_i$$

Con esta elección, en un diseño balanceado, la suma de los efectos de los tratamientos es cero:

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

1.3. Estimación de los efectos

Un paso esencial en el modelado estadístico de los datos es la estimación de los parámetros del modelo. Resultan las siguientes estimaciones.

Media general

$$\hat{\mu} = \bar{y}_{..}$$

Media de los tratamientos

$$\hat{\mu}_i = \bar{y}_{i.}$$

Efectos

$$\hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..}$$

En el análisis estándar de un diseño completamente aleatorizado, estamos interesados en las respuestas medias de los grupos de tratamiento. Una manera obvia para comenzar es decidir si las medias son todas iguales, o si algunas de ellas son diferentes. En términos de modelos, nos preguntamos si los datos pueden ser adecuadamente descritos por el modelo de una sola media general ($y_{ij} = \mu + e_{ij}$, modelo reducido) o si necesitamos el modelo de medias separadas por cada tratamiento ($y_{ij} = \mu + \alpha_i + e_{ij}$, modelo completo).

1.4. Mecánica del ANOVA

A continuación vamos a considerar el modelo de un factor con a tratamientos y n réplicas en cada nivel, por lo tanto, los índices van $i = 1, \dots, a$ y $j = 1, \dots, n$. De acuerdo con lo anterior, el ANOVA de un factor resuelve el siguiente contraste de hipótesis

$$\begin{aligned} H_0: & \alpha_1 = \alpha_2 = \dots = \alpha_a = 0 \\ H_1: & \alpha_i \neq \alpha_j \text{ para algún } i \neq j \end{aligned}$$

Aceptar H_0 implica que el factor no es significativo; por el contrario, aceptar H_1 implica que el factor es significativo: al menos hay dos niveles cuyos efectos son significativamente distintos. En caso de aceptar H_1 deberemos encontrar entre qué niveles están las diferencias. El ANOVA responde solo a la pregunta de si hay o no diferencias, pero no responde entre qué niveles están tales diferencias; para ello deberemos usar otros test adicionales, como veremos más adelante.

Notación de puntos

En el contexto del ANOVA es habitual utilizar esta notación para denotar el índice sobre el cual se promedia. Por ejemplo, en el caso del ANOVA de un factor indicamos con y_{ij} la observación efectuada en la réplica j del nivel i del factor. Si queremos indicar la media dentro del nivel i del factor, deberemos promediar las observaciones en dicho nivel, es decir, sumar j desde 1 hasta n_i y dividir por n_i . Dado que estamos sumando sobre el índice j , la manera de denotar el índice sobre el que sumamos es con un punto, por tanto, indicaremos $y_{i.}$, y al tratarse de un promedio añadimos la barra, y resulta $\bar{y}_{i.}$. Pongamos otro ejemplo: la media general resultará de sumar para todos los niveles y para todas las réplicas, por tanto, dado que sumamos respecto de los dos índices, lo indicaremos con dos puntos, esto es, $y_{..}$. Y al tratarse de una media, añadiremos la barra, resulta, $\bar{y}_{..}$.

Estrictamente hablando, ANOVA es un procedimiento de cálculo para descomponer la variabilidad en un conjunto de datos en dos cantidades, una asociada a una estructura de datos con diferentes medias y otra cantidad residual. ANOVA parte de la igualdad

$$y_{ij} - \bar{y}_{..} = (\bar{y}_i - \bar{y}_{..}) + (y_{ij} - \bar{y}_i)$$

Elevando al cuadrado, resulta:

$$(y_{ij} - \bar{y}_{..})^2 = (\bar{y}_i - \bar{y}_{..})^2 + (y_{ij} - \bar{y}_i)^2 + 2(\bar{y}_i - \bar{y}_{..})(y_{ij} - \bar{y}_i)$$

Podemos sumar en los dos lados a través de j . Observamos que $\sum_{j=1}^n (y_{ij} - \bar{y}_i) = 0$, y obtenemos:

$$\sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = n(\bar{y}_i - \bar{y}_{..})^2 + \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2$$

Ahora, si sumamos en los dos lados a través de i obtenemos:

$$\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = n \sum_{i=1}^a (\bar{y}_i - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2$$

Y resulta la descomposición fundamental del ANOVA. A la izquierda, la suma de cuadrados total (SST) que se descompone, a la derecha, como la suma de cuadrados de los tratamientos (SSA) más la suma de cuadrados del error (SSE):

$$SST = SSA + SSE$$

Para un conjunto de datos dado (y por tanto la SST fijada), tener mayor diferencias entre las medias de grupos implica un mayor SSA que, a la vez, implica tener un menor SSE . Cabe señalar que es mediante la descomposición de la variabilidad total, de ahí el nombre de análisis de la varianza, como resolvemos un contraste de hipótesis que se plantea sobre medias.

1.5. Tabla ANOVA

Los cálculos ANOVA se resumen a modo de tabla con columnas para las fuentes de variabilidad, los grados de libertad asociados a las sumas de cuadrados (gl), las sumas de cuadrados (SS), los cuadrados medios (MS) y el valor del estadístico F.

Fuentes	gl	SS	MS	F
Tratamientos	$a - 1$	SSA	$MSA = SSA/(a - 1)$	$F = MSA/MSE$
Error	$N - a$	SSE	$MSE = SSE/(N - a)$	

1.5.1. ¿Por qué funciona el ANOVA?

El cuadrado medio del error es una variable aleatoria que depende del error aleatorio asociado a los datos. Si repitiéramos el experimento, conseguiríamos diferentes muestras de los errores aleatorios y, por lo tanto, diferentes evaluaciones en el muestreo del cuadrado medio del error. El valor esperado del cuadrado medio del error $E(MSE)$ se puede calcular teóricamente y se comprueba que coincide con la varianza del error σ^2 , es decir,

$$E(MSE) = \sigma^2$$

Cabe resaltar que el MSE nos proporciona una estimación centrada de la varianza σ^2 . Además, el cuadrado medio de los tratamientos es también una variable aleatoria, cuya esperanza puede demostrarse:

$$E(MSA) = \sigma^2 + \frac{n}{a-1} \sum_{i=1}^a \alpha_i^2$$

Podemos concluir que si la hipótesis nula H_0 es cierta ($\alpha_i = 0, i = 1, \dots, a$), ambas esperanzas, $E(MSE)$ y $E(MSA)$ coinciden. Es decir, las cantidades MSE y MSA serán estimadores centrados de σ^2 . Si es cierta la hipótesis nula H_0 , esperamos que MSE y MSA tengan valores similares, y por lo tanto

$$F = \frac{MSA}{MSE} \approx 1$$

Al contrario, si la hipótesis nula H_0 es falsa, hay algunos $\alpha_i \neq 0$, y la $E(MSA)$ será mayor que σ^2 . Por lo tanto, el cociente F tenderá a ser mayor que 1. Así, rechazaremos la hipótesis nula H_0 para valores suficientemente grandes de F . Teniendo en cuenta que bajo la hipótesis nula H_0 conocemos la distribución de referencia de la F , que es la de una F de Fisher con $a - 1$ y $N - a$ grados de libertad, nos es fácil determinar la región crítica del test fijado un nivel de significación.

Ejemplo 1 (continuación)

Recordemos el ejemplo anterior, donde se planteó un diseño de un factor para decidir usar un software para el procesamiento de textos y gestión de bases de datos. Con el fin de evaluar tres sistemas de gestión de bases de datos, una empresa diseñó una prueba para ver cuántas horas de capacitación se necesitaban para que los operadores sean competentes en cada uno de los tres sistemas.

Describir el diseño experimental asociado y formular el modelo estadístico correspondiente

La formulación del diseño experimental es:

$$y_{ij} = \mu + \alpha_i + e_{ij}, \text{ con } \sum_{i=1}^3 \alpha_i = 0$$

donde $i = 1, \dots, 3$ y $j = 1, \dots, 6$, y_{ij} es la observación para el nivel i y la réplica j , μ es la media general, α_i es el efecto del nivel i y e_{ij} es el error aleatorio. Asumimos que $e_{ij} \sim N(0, \sigma^2)$, donde σ^2 es constante para todos los niveles del factor.

Empezaremos creando un `data frame` con los datos del enunciado

```
y<- c(20,17,15,19,14,13,
      18,17,14,20,13,12,
      23,25,20,21,19,20)
fact<-as.factor(c(rep(1,6),rep(2,6),rep(3,6)))
df<-data.frame(y,fact)
str(df)
## 'data.frame': 18 obs. of 2 variables:
## $ y : num 20 17 15 19 14 13 18 17 14 20 ...
## $ fact: Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 2 2 2 2 ...
head(df)
##   y fact
## 1 20   1
## 2 17   1
## 3 15   1
## 4 19   1
## 5 14   1
## 6 13   1
round(mean(df$y),3)
## [1] 17.778
```

Es importante asegurarnos de que la variable que contiene el `factor` es de tipo factor, lo que podemos verificar con la función `str` de R.

Obtener la estimación de todos los parámetros del modelo

Próximamente comentaremos funciones R más directas para obtener la estimación de los parámetros de un modelo, pero por ahora nos conformaremos con el cálculo a partir de las fórmulas correspondientes.

La media general μ se estima con la media de los datos, esto es: $\hat{\mu} = \bar{y}_{..} = 17.778$.

Y los efectos de los tratamientos se estiman mediante $\hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..}$. Por lo tanto,

$$\hat{\alpha}_1 = \bar{y}_{1.} - \bar{y}_{..} = 16.333 - 17.778 = -1.445$$

$$\hat{\alpha}_2 = \bar{y}_{2.} - \bar{y}_{..} = 15.667 - 17.778 = -2.111$$

$$\hat{\alpha}_3 = \bar{y}_{3.} - \bar{y}_{..} = 21.333 - 17.778 = 3.555$$

Observamos que $\sum_{i=1}^3 \hat{\alpha}_i = 0$, como cabe esperar al tratarse de un diseño balanceado.

Naturalmente, aún falta la estimación de otro parámetro del modelo, la varianza del error, σ^2 . Si bien, a partir la fórmula

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^g \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2}{N - g}$$

podemos obtener una estimación de dicho parámetro, veremos a continuación la manera de obtenerla a partir de la tabla del análisis de la varianza.

Podemos preguntarnos si el factor tipo de software es significativo, en el sentido de que el tiempo necesario para la capacitación se ve afectada por dicho factor. Por lo tanto, planteamos el contraste de hipótesis

$$H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0$$

$$H_1: \alpha_i \neq \alpha_j \text{ para algún } i \neq j$$

Para calcular la tabla ANOVA asociada a este problema debemos primero ajustar el modelo mediante la función `lm`:

```
results<-lm(y~fact,data=df)
```

Podemos extraer la tabla ANOVA mediante:

```
taov<-anova(results)
taov
## Analysis of Variance Table
##
## Response: y
##          Df Sum Sq Mean Sq F value    Pr(>F)
## fact       2 115.11   57.556    7.5731 0.005326 **
## Residuals 15 114.00    7.600
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Obtenemos un p -valor 0.005, claramente inferior a un nivel de significación del 5 %. Por lo tanto, aceptamos la hipótesis alternativa y concluimos que el factor es significativo. Es decir, el tipo de software afecta al tiempo necesario para adquirir la capacitación. La tarea a continuación será determinar en qué niveles del factor están las diferencias.

Vale la pena recordar que la tabla ANOVA y en concreto el valor de MSE nos proporciona una estimación de la varianza del error $\hat{\sigma}^2 = 7.6$.

1.6. Contrastes

Formalmente, un contraste es una combinación lineal de las medias o

de los efectos $\sum_{i=1}^a w_i \alpha_i$, donde $\sum_{i=1}^a w_i = 0$. Denotaremos un contraste por $w(\{\alpha_i\})$.

Es probable que los contrastes más comunes sean las comparaciones por parejas, donde comparamos la respuesta media en un tratamiento con la respuesta media en un segundo tratamiento. Para una comparación por parejas, un coeficiente del contraste es 1, mientras que el segundo coeficiente del contraste es -1 , y todos los demás coeficientes de los contrastes son 0. Por ejemplo, en un diseño con cuatro niveles en el que $a = 4$, un contraste con coeficientes $(0, 1, 0, -1)$ compara los efectos de los tratamientos 2 y 4.

Planteamos contrastes sobre las medias observadas de los tratamientos o sobre los efectos para hacer inferencia sobre las medias o los efectos poblacionales.

Las medias muestrales de los grupos \bar{y}_i son un estimador centrado de las $\mu_i = \mu + \alpha_i$. Por tanto, cualquier combinación lineal de las medias muestrales de los tratamientos es un estimador centrado de la correspondiente combinación

lineal de las μ_i . En particular, un contraste definido en las medias muestrales es un estimador centrado del correspondiente contraste de las medias de los tratamientos. Por lo tanto,

$$E(w(\{\bar{y}_i\})) = w(\{\alpha_i\})$$

Por otro lado, la varianza de \bar{y}_i es $\frac{\sigma^2}{n}$, y las medias muestrales de los tratamientos son independientes. Por lo tanto resulta que la varianza de un contraste sobre las medias muestrales es:

$$\text{var}(w(\{\bar{y}_i\})) = \frac{\sigma^2}{n} \sum_{i=1}^a w_i^2$$

Usualmente no conocemos σ^2 , así que usamos la estimación que obtenemos de los cuadrados medios del error, MSE , de la tabla ANOVA.

Entonces, un intervalo de confianza para el contraste $w(\{\alpha_i\})$ de nivel de confianza $(1 - \epsilon)100$ % es de la forma:

$$w(\{\bar{y}_i\}) \pm t_{N-a, \epsilon/2} \sqrt{MSE/n} \sqrt{\sum_{i=1}^a w_i^2}$$

Este planteamiento también funciona en el caso de los test de hipótesis. Si tenemos una hipótesis nula $H_0: w(\{\alpha_i\}) = \delta$, usualmente interesa $\delta = 0$. Entonces, podemos usar un test basado en el estadístico t para resolver el test:

$$t = \frac{w(\{\bar{y}_i\}) - \delta}{\sqrt{MSE/n} \sqrt{\sum_{i=1}^a w_i^2}}$$

Bajo la hipótesis nula, el estadístico t sigue una distribución t de Student de $N - a$ grados de libertad.

En el caso concreto de las comparaciones por parejas, esto es,

$$\begin{aligned} H_0: & \alpha_i - \alpha_j = 0 \\ H_1: & \alpha_i - \alpha_j \neq 0 \end{aligned}$$

El estadístico t es de la forma:

$$t = \frac{(\bar{y}_i - \bar{y}_j) - 0}{\sqrt{MSE/n} \sqrt{2}}$$

Bajo la hipótesis nula, el estadístico t sigue una distribución t de Student de $N - a$ grados de libertad.

Con la práctica se concluye que la pareja de efectos α_i y α_j (o en equivalencia la pareja de medias μ_i y μ_j) son estadísticamente diferentes si

$$|\bar{y}_i - \bar{y}_j| > LSD$$

donde la cantidad LSD, denominada *Least Significant Difference*, viene dada por

$$LSD = t_{N-a, \epsilon/2} \sqrt{2MSE/n}$$

En este resumen hemos proporcionado las fórmulas para diseños balanceados, es decir, con el mismo número de réplicas, n , en cada tratamiento.

Ejemplo 1 (continuación)

Con la función `pairwise.t.test` del paquete `stat` podemos obtener los p -valores de todas las comparaciones por parejas.

```
library(stats)
pairwise.t.test(df$y, df$fact, p.adj=c("none"))
##
## Pairwise comparisons using t tests with pooled SD
##
## data: df$y and df$fact
##
## 1 2
## 2 0.6813 -
## 3 0.0067 0.0028
##
## P value adjustment method: none
```

Vemos que hay diferencias entre los niveles 1 y 2 con el nivel 3. No obstante, cabe remarcar que, en el caso de un diseño con a tratamientos, hay $a(a - 1)/2$ comparaciones por parejas de tratamientos distintas. Si estamos interesados en hacer todos los contrastes por parejas, hay que emplear técnicas de inferencia simultánea para ajustar los p -valores obtenidos mediante el test anterior.

1.7. Contrastes múltiples

1.7.1. Test de Bonferroni

Vamos a empezar con un ejemplo. Supongamos que tenemos un conjunto de contrastes de hipótesis que queremos testar simultáneamente. La primera idea que se nos puede ocurrir es realizar cada contraste por separado, usando un cierto nivel de significación ϵ .

A primera vista, esto no parece ser una mala idea. No obstante, consideremos que tuviéramos veinte hipótesis que testar y un nivel de significación del 5 %. ¿Cuál es la probabilidad de declarar al menos un resultado significativo por azar?

$$P(\text{al menos un resultado significativo}) = 1 - P(\text{ningún resultado significativo}) \\ = 1 - (1 - 0,05)^{20} \sim 0,64$$

Resulta, por lo tanto, que con veinte contrastes tenemos un 64 % de probabilidades de aceptar al menos un resultado significativo, incluso si todas las pruebas de hipótesis no son realmente significativas.

Los métodos para hacer frente a las comparaciones múltiples con frecuencia plantean ajustar el nivel de significación ϵ de alguna manera para que la probabilidad de observar al menos un resultado significativo debido al azar se mantenga por debajo del nivel de significación deseado ϵ .

El método de Bonferroni corrige el nivel de significación al punto de corte $\frac{\epsilon}{n}$, siendo n el número de contrastes que hay que realizar. Por ejemplo, en el ejemplo anterior, con $n = 20$ contrastes y nivel de significación $\epsilon = 0,05$, rechazaríamos hipótesis nulas en cuyo test el p -valor fuera inferior a $0,05/20 = 0,0025$.

La corrección de Bonferroni tiende a ser un poco conservadora. Por ejemplo, si calculamos la probabilidad de aceptar al menos un resultado significativo por azar cuando usamos la corrección, esto es

$$P(\text{al menos un resultado significativo}) = 1 - P(\text{ningún resultado significativo}) \\ = 1 - (1 - 0,0025)^{20} \sim 0,0488$$

En este ejemplo, estamos solo un poco por debajo del 0,05 deseado. En las aplicaciones prácticas, la corrección de Bonferroni podría ser extremadamente conservadora, lo que llevaría a una tasa alta de falsos negativos (mantener H_0 cuando es falsa).

Ejemplo 1 (continuación)

Seguimos con el ejemplo anterior. Realicemos las comparaciones múltiples con la corrección de Bonferroni. Usando el paquete `agricolae` y la función `LSD.test`. Vamos a comentar las dos opciones del argumento `group`. Empecemos con la opción `FALSE`.

```
library(agricolae)
## Warning: package 'agricolae' was built under R version 3.5.3
LSD.test(results,"fact",group=F,p.adj="bonferroni",console=T)
##
## Study: results ~ "fact"
##
## LSD t Test for y
## P value adjustment method: bonferroni
##
## Mean Square Error: 7.6
##
## fact, means and individual ( 95 %) CI
##
##          y      std r      LCL      UCL Min Max
## 1 16.33333 2.804758 6 13.93447 18.73220 13 20
## 2 15.66667 3.141125 6 13.26780 18.06553 12 20
## 3 21.33333 2.250926 6 18.93447 23.73220 19 25
##
## Alpha: 0.05 ; DF Error: 15
```

```
## Critical Value of t: 2.693739
##
## Comparison between treatments means
##
##      difference pvalue signif.      LCL      UCL
## 1 - 2  0.6666667 1.0000      -3.620810  4.9541430
## 1 - 3 -5.0000000 0.0202      * -9.287476 -0.7125237
## 2 - 3 -5.6666667 0.0085      ** -9.954143 -1.3791903
```

En la parte superior de los resultados, observamos los intervalos de confianza para la media de cada tratamiento. En la parte inferior encontramos listados los contrastes para todas las comparaciones por parejas y también los IC para las diferencias de medias. Los contrastes significativos, de acuerdo con la corrección de Bonferroni, aparecen marcados con asteriscos.

Si establecemos la opción TRUE en el argumento group, obtenemos un listado algo distinto en lo referente a las comparaciones por parejas.

```
LSD.test(results,"fact",group=T,p.adj="bonferroni",console=T)
##
## Study: results ~ "fact"
##
## LSD t Test for y
## P value adjustment method: bonferroni
##
## Mean Square Error: 7.6
##
## fact, means and individual ( 95 %) CI
##
##      y      std r      LCL      UCL Min Max
## 1 16.33333 2.804758 6 13.93447 18.73220 13 20
## 2 15.66667 3.141125 6 13.26780 18.06553 12 20
## 3 21.33333 2.250926 6 18.93447 23.73220 19 25
##
## Alpha: 0.05 ; DF Error: 15
## Critical Value of t: 2.693739
##
## Minimum Significant Difference: 4.287476
##
## Treatments with the same letter are not significantly different.
##
##      y groups
## 3 21.33333 a
## 1 16.33333 b
## 2 15.66667 b
```

En este caso, se hace uso de una codificación con letras. Aquellos grupos en los que no hay diferencias significativas se codifican con la misma letra. Se distinguen dos grupos, el grupo a con el tratamiento 3 y el grupo b con los tratamientos 1 y 2. Interpretamos de la manera siguiente los resultados. El tipo de software 3 requiere un tiempo de capacitación significativamente mayor que el de los tipos 1 y 2. No hay diferencias significativas en el tiempo de capacitación de los tipos 1 y 2.

1.8. Comprobación de las suposiciones del modelo

En resumen, mediante el ANOVA analizamos los datos experimentales comparando las respuestas medias en los distintos tratamientos y en caso de detectar diferencias significativas procedemos a un análisis más detallado basado en comparaciones por parejas u otros contrastes.

Todos estos procedimientos estadísticos se basan en la suposición de que los datos se ajustan al modelo:

$$y_{ij} = \mu + \alpha_i + e_{ij}$$

donde μ , α_i son parámetros (cantidades fijas pero desconocidas) y e_{ij} son variables aleatorias independientes y distribuidas según una normal de media 0 y varianza constante σ^2 . Por lo tanto, será necesario chequear si estas suposiciones se satisfacen razonablemente.

Las tres suposiciones básicas que debemos validar son que los errores e_{ij} son:

- incorrelacionados (independientes en caso de normalidad)
- distribuidos según una normal
- con varianza constante (no varía a través de los tratamientos)

La suposición de independencia es la más importante y difícil de corregir en caso de que no se cumpla. Es el diseño del experimento, una adecuada aleatorización en la asignación de los tratamientos y, en fin, cada paso del experimento, que debe asegurarnos que las observaciones son independientes. En este tema vamos a centrarnos en la validación de las otras dos suposiciones: normalidad y varianza constante. Veremos que la suposición de normalidad, aunque importante, puede relajarse en caso de muestras grandes. La condición de varianza constante (también llamada condición de homocedasticidad), si no se cumple, puede afectar a nuestras inferencias. No obstante, la falta de homocedasticidad puede corregirse en muchas situaciones.

A continuación se consideran distintos métodos de diagnóstico y posibles correcciones cuando los supuestos de modelización no se cumplen.

Hasta cierto punto podría parecer que «forzamos la máquina» porque vamos a estudiar una serie de transformaciones con el fin de que nuestros modelos lineales, con errores distribuidos normalmente, sean aplicables a muchos tipos de datos. Cabe decir que hay otras metodologías que podríamos usar en su lugar, por ejemplo: los modelos lineales generalizados, los métodos robustos, los métodos de permutaciones o los métodos no paramétricos basados en rangos. De hecho, para ciertos tipos de datos algunos de estos métodos alternativos pueden ser considerablemente más eficientes (por ejemplo, producir intervalos de confianza más cortos con el mismo nivel de confianza) que el modelo lineal basado en la distribución normal.

1.8.1. Evaluando el cumplimiento de los supuestos de modelización

Puede que las suposiciones de independencia, distribución normal de los errores con varianza constante no se cumplan por completo en datos reales. No obstante, los modelos lineales producen buenas inferencias, siempre que la desviación de los supuestos no sea demasiado grande. Por este motivo, pode-

mos utilizar métodos gráficos y descriptivos para evaluar el grado de cumplimiento de las suposiciones en lugar de testar. Aunque como veremos también disponemos de test estadísticos para tal fin.

Evaluamos el cumplimiento de las suposiciones sobre los errores aleatorios basándonos en los residuos. Los residuos r_{ij} vienen dados por la diferencia entre el valor observado y_{ij} y el valor ajustado según el modelo para dicha observación. En el caso del ANOVA de un factor, el valor ajustado para la observación y_{ij} es $\bar{y}_{i\cdot}$. Por lo tanto, para cada dato obtenemos un residuo dado por

$$r_{ij} = y_{ij} - (\hat{\mu} + \hat{\alpha}_i) = y_{ij} - \bar{y}_{i\cdot}$$

Los residuos r_{ij} son una muestra de la variable aleatoria e_{ij} . Según las suposiciones del modelo se asume que $e_{ij} \sim N(0, \sigma^2)$, por lo tanto será estudiando los residuos como podremos validar los supuestos del modelo.

Ejemplo 1 (continuación)

Nos preguntábamos si el factor «tipo de software» era significativo, en el sentido de que el tiempo de capacitación se ve afectado por dicho factor. Planteábamos el contraste de hipótesis

$$\begin{aligned} H_0: & \alpha_1 = \alpha_2 = \alpha_3 = 0 \\ H_1: & \alpha_i \neq \alpha_j \text{ para algún } i \neq j \end{aligned}$$

Para resolverlo calculábamos la tabla ANOVA:

```
results<-lm(y~fact,data=df)
```

Extrajimos la tabla ANOVA mediante:

```
taov<-anova(results)
taov
## Analysis of Variance Table
##
## Response: y
##          Df Sum Sq Mean Sq F value    Pr(>F)
## fact       2  115.11   57.556    7.5731 0.005326 **
## Residuals  15  114.00    7.600
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Obtuvimos un p -valor 0.005, claramente inferior a un nivel de significación del 5 %. Por lo tanto, aceptamos la hipótesis alternativa y concluimos que el factor es significativo.

Recordemos que las medias de cada grupo y_i son:

```
tapply(y, fact, mean, data=df)
##          1          2          3
## 16.33333 15.66667 21.33333
```

Para el cálculo de los residuos, recordemos que cada residuo se calcula mediante $r_{ij} = y_{ij} - \bar{y}_{i\cdot}$. Por ejemplo, el residuo de la observación correspondiente a la primera réplica del primer nivel, es decir, $i = 1$ y $j = 1$ será:

$$y_{11} - \bar{y}_{1\cdot} = 20 - 16.333 = 3.667$$

Podemos extraer con la función `residuals` todos los residuos del objeto, de tal manera que de cada observación obtenemos su residuo:

```
residuals(results)
```

##	1	2	3	4	5	6
##	3.6666667	0.6666667	-1.3333333	2.6666667	-2.3333333	-3.3333333
##	7	8	9	10	11	12
##	2.3333333	1.3333333	-1.6666667	4.3333333	-2.6666667	-3.6666667
##	13	14	15	16	17	18
##	1.6666667	3.6666667	-1.3333333	-0.3333333	-2.3333333	-1.3333333

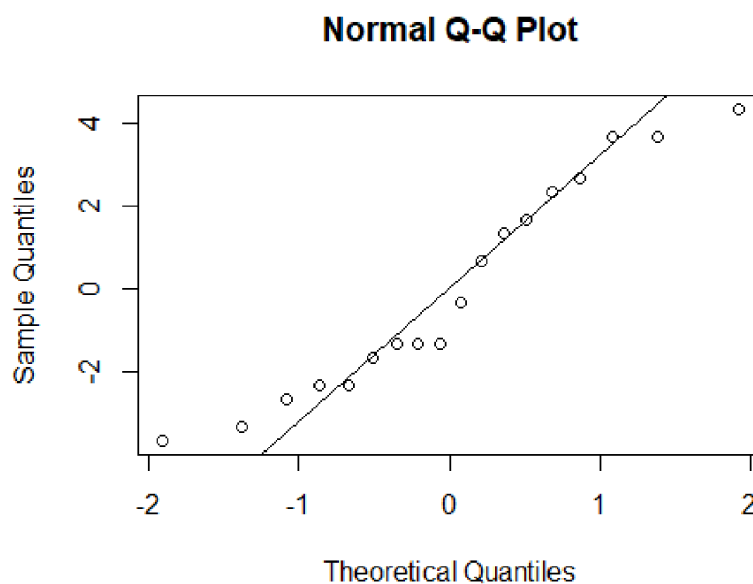
1.8.2. Evaluando la normalidad

Los qqplot son un procedimiento gráfico para la evaluación de la normalidad. Los qqplot son útiles en el diagnóstico de desviaciones de la normalidad porque permiten la comparación de los cuantiles de la distribución de los residuos de los datos con los cuantiles de una distribución normal estándar. Si se aprecian desviaciones de la linealidad, ello será indicio de que la normalidad no se cumple.

Ejemplo 1 (continuación)

Podemos obtener un qqplot de los residuos mediante la función `qqnorm`

```
qqnorm(residuals(results))
qqline(residuals(results))
```



Observamos que la mayoría de los residuos se ajustan a la recta, por lo que no hay evidencia en contra del supuesto de normalidad. Alternativamente, podemos obtener el qqplot utilizando la `plot.lm`, llamando al *slot 2*, es decir, `plot(results, which=2)`.

Adicionalmente, podemos contrastar la normalidad mediante el test de Shapiro-Wilk (donde la hipótesis nula es aquella que afirma que la distribución es normal):

```
shapiro.test(residuals(results))
##
## Shapiro-Wilk normality test
##
## data: residuals(results)
## W = 0.92953, p-value = 0.1905
```

A la vista del p -valor obtenido, mantenemos la hipótesis nula y podemos aceptar que la variable aleatoria e_{ij} sigue una distribución normal. Cabe señalar que debemos aplicar la función `shapiro.test` a los residuos, y observemos que será equivocado aplicarlo sobre las observaciones y_{ij} .

1.8.3. Evaluando el cumplimiento de la constancia de la varianza (homocedasticidad)

Evaluamos la constancia de la varianza mediante un *plot* de los residuos r_{ij} en el eje vertical sobre los valores ajustados (*fitted values*) en el eje horizontal. El valor ajustado de una observación es, en el ANOVA de un factor, la media del grupo al que corresponde la observación. Por lo tanto, para todas las observaciones que corresponden al nivel i , el valor ajustado es \bar{y}_i .

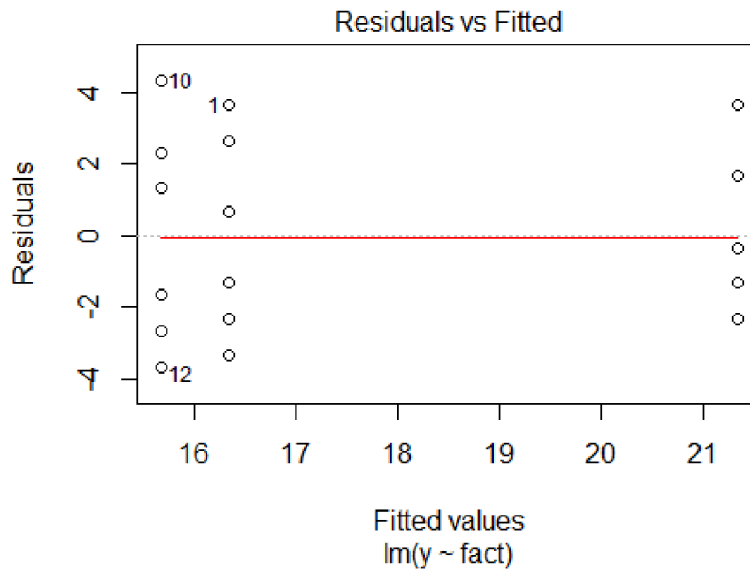
Por ejemplo, con los datos del ejemplo, todas las observaciones que corresponden al nivel 2, el valor ajustado es la media del nivel 2, $\bar{y}_2 = 15.667$.

En este tipo de gráfico de residuos aparecerán tiras verticales de puntos, una tira por cada nivel. Si la varianza es constante, la extensión vertical de las tiras será aproximadamente la misma. Si la varianza no es constante, aparecerá un patrón en la dispersión vertical de los residuos. Es decir, los niveles de tratamiento mostrarán una disposición de los residuos con distinta dispersión vertical.

Un patrón habitual cuando no se cumple la suposición de varianza constante es aquel que aparece en caso de que la varianza depende de la media del grupo. Por lo general, en estos casos hay un aumento de la varianza a medida que la media del grupo aumenta. En consecuencia, el gráfico de residuos muestra una apertura a la derecha en forma de embudo.

Sigamos analizando los datos del ejemplo. Podemos obtener un gráfico de residuos utilizando la función `plot.lm`, llamando al *slot* 1.

```
plot(results, which=1)
```



Observamos tres tiras verticales de puntos que están situadas en las medias de cada grupo. Como hemos dicho, estas corresponden a los valores ajustados de las observaciones. La disposición de los residuos muestra una dispersión parecida en cada tira. Por lo tanto, no se aprecia efecto de embudo.

Adicionalmente, podemos testar la homogeneidad de varianzas mediante el test de Bartlett, que es aplicable en caso de normalidad. La hipótesis nula es la de homogeneidad de varianzas a través de los niveles. Esto es

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$$

$$H_1 : \alpha_i \neq \alpha_j \text{ para algún } i \neq j$$

Sigamos analizando los datos del ejemplo. Evaluamos la homogeneidad de las varianzas.

```
bartlett.test(y~fact,data=df)
##
## Bartlett test of homogeneity of variances
##
## data: y by fact
## Bartlett's K-squared = 0.50494, df = 2, p-value = 0.7769
```

A la vista del p -valor, mantenemos la hipótesis nula y aceptamos que las varianzas son iguales.

1.8.4. Corrigiendo el incumplimiento de las condiciones

1) Corrigiendo la no normalidad

La no normalidad, sobre todo debida a la asimetría, a veces puede ser disminuida mediante la transformación de la variable respuesta a una escala diferente. La asimetría hacia la derecha se ve reducida por una raíz cuadrada, un logaritmo u otra transformación a una potencia menor que uno, mientras que la asimetría a la izquierda se ve reducida por un cuadrado, cubo u otra transformación a una potencia mayor que uno.

2) Corrigiendo la no constancia de la varianza

La forma habitual de solucionar la no constancia de la varianza es por la transformación de la variable respuesta. Para algunas distribuciones hay transformaciones estándar que igualan o estabilizan la varianza. Hay una teoría general de las transformaciones estabilizadoras de la varianza que se aplica a las distribuciones donde la varianza depende de la media.

Cuando no tenemos una distribución con una transformación conocida para la estabilización de la varianza (y por lo general no la tenemos), entonces se suele ensayar una familia de transformaciones de potencia. El método de Box-Cox establece un procedimiento para determinar a partir de los datos la transformación de potencia. La familia de transformaciones de Box-Cox vienen dadas por:

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda y^{\lambda-1}} & \lambda \neq 0 \\ \log(y) & \lambda = 0 \end{cases}$$

donde \bar{y} es la media geométrica de la variable respuesta.

La técnica de Box-Cox transforma los datos (variable respuesta) explorando un rango de valores de λ y realizando un ANOVA para cada transformación (es decir, para cada valor de λ). En cada ANOVA obtenemos una suma de cuadrados del error (SSE) que dependerá de la λ usada, $SSE(\lambda)$. La mejor transformación λ^* será aquella para la cual el $SSE(\lambda)$ alcanza el mínimo.

Para implementar la técnica de Box-Cox utilizaremos la función `boxcox` del paquete `MASS`.

Vamos a utilizar otro ejemplo en el que será necesario aplicar la técnica de Box-Cox.

Ejemplo 2

La vida útil de las carnes almacenadas es el tiempo que un corte previamente envasado es sano, nutritivo y vendible. Un paquete normal expuesto al aire ambiental tiene una vida aproximada de 48 horas, después de las cuales la carne comienza a deteriorarse por contaminación de microbios, degradación del color y encogimiento. El envase al vacío es efectivo para suprimir el desarrollo de microbios; sin embargo, continúan siendo un problema los otros aspectos.

Algunos estudios recientes sugieren las atmósferas controladas de gas como alternativa a los envases actuales. Dos atmósferas que prometen combinar la capacidad de suprimir el desarrollo de microbios con la conservación de las cualidades de la carne son: 1) dióxido de carbono puro (CO_2), y 2) mezclas de monóxido de carbono (CO), oxígeno (O) y nitrógeno (N). Según esta nueva información, un equipo investigador plantea la hipótesis de que alguna forma de atmósfera controlada proporcionaría un entorno más efectivo de envasado para el almacenamiento de carne.

El diseño del experimento desarrollado por el equipo investigador para evaluar la hipótesis incluye diversos tipos de envasados con: 1) aire del ambiente con un envasado comercial de plástico; 2) al vacío; 3) una mezcla de gases con 1 % CO , 40 % O y 59 % N , y 4) 100 % CO_2 . Los envases con el aire del ambiente y al vacío sirven como tratamientos de control, ya que ambos son estándares con cuya efectividad se puede comparar la de los nuevos envasados.

Se usa un diseño totalmente aleatorizado para el experimento. A cada conjunto de condiciones de envasado se le asignaron al azar cortes del mismo tamaño (75 g). Cada corte se envasa por separado en las condiciones asignadas.

Se evalúa la efectividad de cada tratamiento para suprimir el desarrollo bacteriano. Después de nueve días de almacenamiento a 4 C en una instalación normal, se midió el número de bacterias sicotrópicas en la carne. Las bacterias sicotrópicas se encuentran en la superficie de la carne y se asocian con la carne deteriorada. El crecimiento bacteriano se expresa como el número de bacterias (en millones) por centímetro cuadrado.

Atmósfera			
Comercial	45,7	9,5	63,1
Al vacío	0,2	0,7	0,6
Mezcla de gases	25,1	21,3	10,9
CO ₂	0,09	0,25	0,004

1.8.5. Formulación del modelo estadístico para analizar los datos del experimento anterior

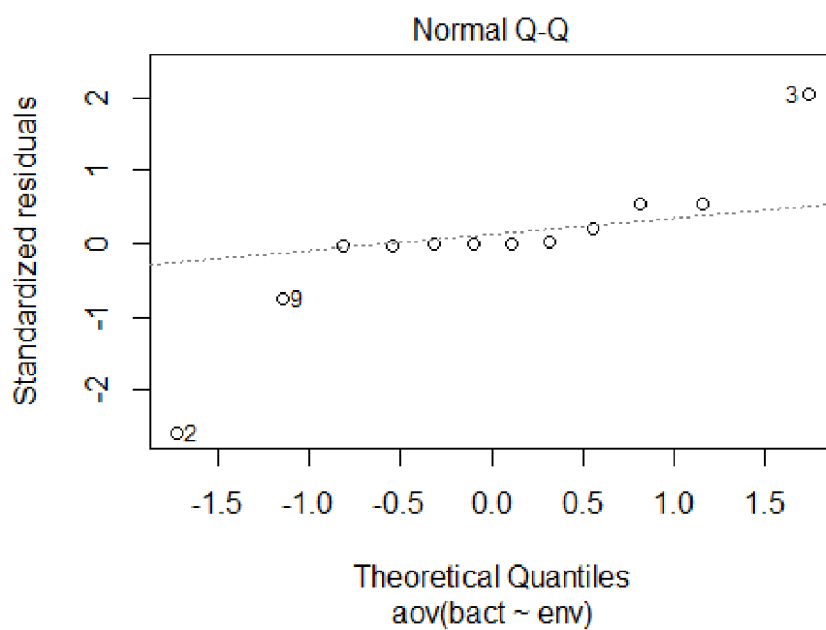
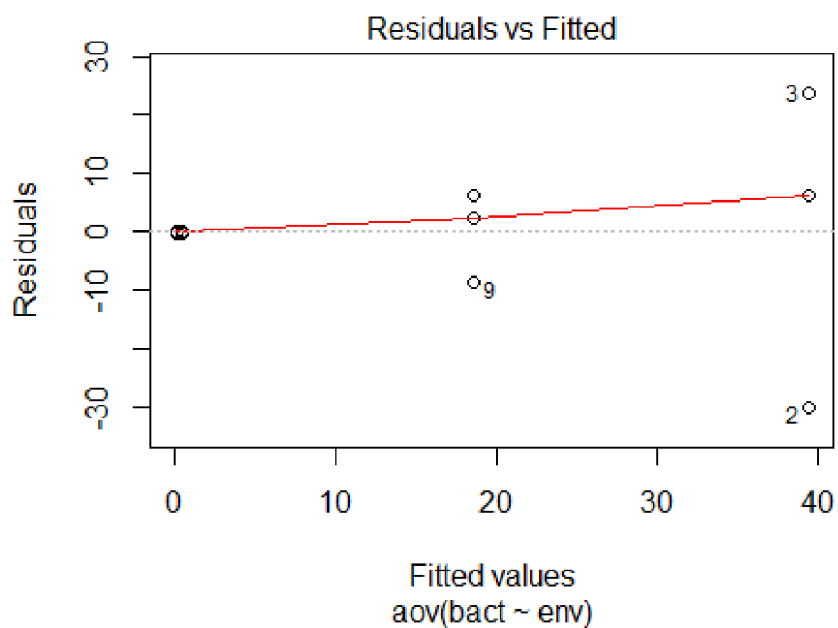
La variable repuesta y_{ij} es el número de bacterias sicotrópicas. El factor es el tratamiento, que tienen cuatro niveles (envasado comercial, al vacío, mezcla de gases y dióxido de carbono), y existen tres réplicas por nivel. Formulamos el modelo lineal

$$y_{ij} = \mu + \alpha_i + e_{ij}$$

con la restricción $\sum_{i=1}^4 \alpha_i = 0$. Asumimos que e_{ij} sigue la distribución $N(0, \sigma^2)$, donde σ^2 es constante a través de los grupos. Las observaciones son independientes.

1.8.6. Evaluación de las suposiciones de modelización (normalidad y homocedasticidad)

```
env<-c("Comercial", "Comercial", "Comercial", "Al vacío", "Al vacío",
      "Al vacío", "Mezcla de gases", "Mezcla de gases", "Mezcla de gases",
      "co2", "co2", "co2")
bact<-c(45.7, 9.5, 63.1, 0.2, 0.7, 0.6,
        25, 21, 10, 0.09, 0.25, 0.004)
df<-data.frame(bact, env)
results<-aov(bact~env, data=df)
plot(results, which=c(1, 2))
```



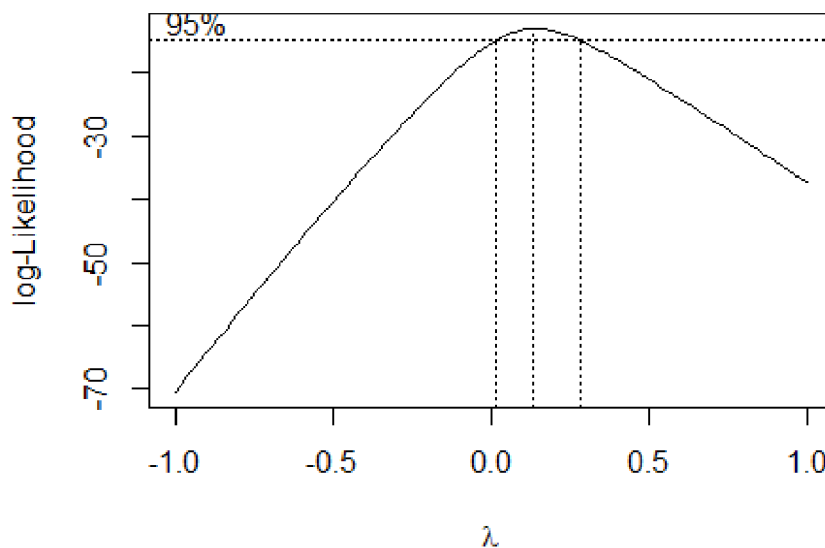
```
shapiro.test(residuals(results))
##
##  Shapiro-Wilk normality test
##
## data:  residuals(results)
## W = 0.81445, p-value = 0.01374
bartlett.test(bact~env,data=df)
##
```

```
## Bartlett test of homogeneity of variances
##
## data:  bact by env
## Bartlett's K-squared = 28.696, df = 3, p-value = 2.594e-06
```

A la vista de los resultados, **no** se cumplen los supuestos de modelización. Es necesario aplicar una transformación.

1.8.7. Transformación de la variable respuesta (número bacterias/cm²) para proseguir con el análisis de los datos

```
library(MASS)
## Warning: package 'MASS' was built under R version 3.5.3
bct<-boxcox(bact~env,lambda = seq(-1, 1, length = 10),data=df,plotit=T)
```

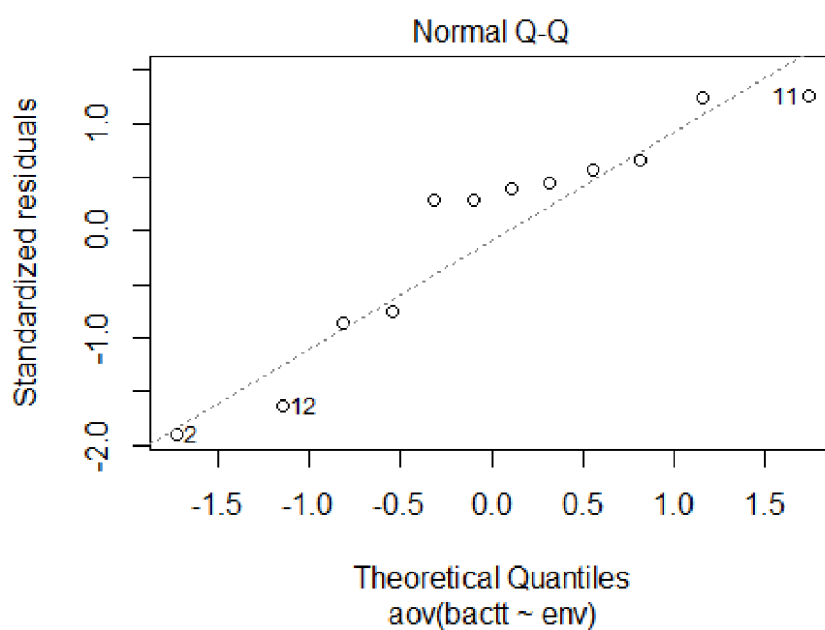
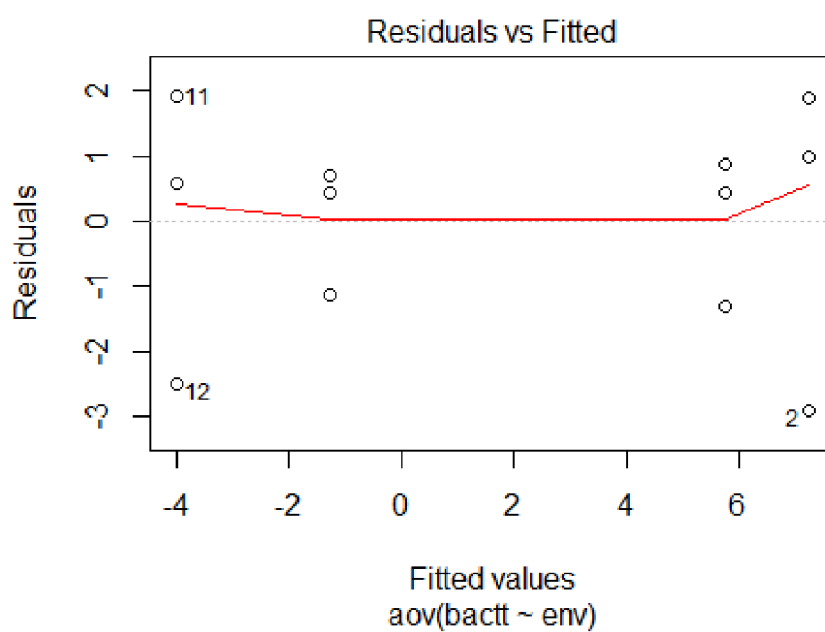


```
lambda<-bct$x[which.max(bct$y)]
lambda
## [1] 0.1313131
library(psych)
## Warning: package 'psych' was built under R version 3.5.3
gm<-geometric.mean(bact)
gm
## [1] 1.790726
df$bactt<-(bact^lambda-1)/(lambda*gm^(lambda-1))
```

Con ello hemos añadido una nueva variable al `data.frame` con la variable respuesta corregida.

1.8.8. Verificación de la adecuación

```
results<-aov(bactt~env,data=df)
plot(results,which=c(1,2))
```



```
shapiro.test(residuals(results))
##
##  Shapiro-Wilk normality test
##
## data:  residuals(results)
## W = 0.89049, p-value = 0.1196
bartlett.test(bactt~env,data=df)
##
##  Bartlett test of homogeneity of variances
##
## data:  bactt by env
## Bartlett's K-squared = 2.0049, df = 3, p-value = 0.5714
```

Una vez estimada la λ óptima. Claramente, la transformación corrige los incumplimientos anteriores.

1.8.9. Evaluar si hay diferencias significativas entre los tratamientos

Una vez transformada la variable, analizaremos la significación del factor envasado.

```
results<-aov(bactt~env,data=df)
anova(results)
## Analysis of Variance Table
##
## Response: bactt
##          Df Sum Sq Mean Sq F value    Pr(>F)
## env        3 265.500   88.500   25.401 0.0001927 ***
## Residuals   8  27.873    3.484
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Se observa que el factor tipo de envasado es significativo.

1.8.10. Estimación de los parámetros del modelo

```
model.tables(results,type="means")
## Tables of means
## Grand mean
##
## 1.932156
##
## env
## env
##          Al vacío          co2          Comercial Mezcla de gases
##          -1.268          -4.014           7.239           5.771
```

```

model.tables(results,type="effects")
## Tables of effects
##
##  env
##  env
##      Al vacío      co2      Comercial Mezcla de gases
##      -3.200      -5.946      5.307      3.839
anova(results)[2,3]
## [1] 3.484101

```

1.8.11. ¿Se puede recomendar un tratamiento que mejore los tratamientos usuales?

Deberemos realizar comparaciones múltiples.

```

library(agricolae)
LSD.test(results,"env",group=T,p.adj="bonferroni",console=T)
##
## Study: results ~ "env"
##
## LSD t Test for bactt
## P value adjustment method: bonferroni
##
## Mean Square Error:  3.484101
##
## env,  means and individual ( 95 %) CI
##
##      bactt      std r      LCL      UCL      Min
## Al vacío      -1.268048 0.9933341 3 -3.753153  1.217057 -2.406541
## co2          -4.013854 2.2643418 3 -6.498960 -1.528749 -6.514602
## Comercial      7.239082 2.5465278 3  4.753977  9.724188  4.345194
## Mezcla de gases 5.771443 1.1565648 3  3.286337  8.256548  4.459934
##
##      Max
## Al vacío      -0.578023
## co2          -2.102466
## Comercial      9.137538
## Mezcla de gases 6.645374
##
## Alpha: 0.05 ; DF Error: 8
## Critical Value of t: 3.478879
##
## Minimum Significant Difference: 5.301992
##
## Treatments with the same letter are not significantly different.
##
##      bactt groups
## Comercial      7.239082      a

```

```
## Mezcla de gases  5.771443      a
## Al vacío         -1.268048      b
## co2             -4.013854      b
HSD.test(results,"env",group=T,console=T)
##
## Study: results ~ "env"
##
## HSD Test for bactt
##
## Mean Square Error:  3.484101
##
## env,  means
##
##           bactt      std r      Min      Max
## Al vacío        -1.268048 0.9933341 3 -2.406541 -0.578023
## co2            -4.013854 2.2643418 3 -6.514602 -2.102466
## Comercial       7.239082 2.5465278 3  4.345194  9.137538
## Mezcla de gases  5.771443 1.1565648 3  4.459934  6.645374
##
## Alpha: 0.05 ; DF Error: 8
## Critical Value of Studentized Range: 4.52881
##
## Minimun Significant Difference: 4.88055
##
## Treatments with the same letter are not significantly different.
##
##           bactt groups
## Comercial       7.239082      a
## Mezcla de gases  5.771443      a
## Al vacío         -1.268048      b
## co2            -4.013854      b
```

Observamos que los tratamientos al vacío y con CO₂ forman un grupo homogéneo que provoca una reducción significativa del número de bacterias respecto a los otros tratamientos.

1.9. Test de remuestreo para el ANOVA

En ocasiones, la suposición de normalidad del término aleatorio no es plausible. En este caso, es imposible aplicar la prueba F en el contraste de la significación de los efectos del tratamiento. En tales casos, se pueden aplicar técnicas basadas en el remuestreo.

El procedimiento básico es el siguiente:

- Calcular el valor de F para los datos experimentales (llamémosle F_{obs}).

- Elegir B muestras, donde cada muestra es una permutación aleatoria de los datos originales.
- Para cada muestra, asignar las primeras observaciones n_1 al primer tratamiento, las siguientes observaciones n_2 al segundo tratamiento, y así sucesivamente.
- Para cada remuestreo, calcular F a partir de los datos.
- Si $F > F_{obs}$, incrementar un contador.
- Calcular p dividiendo el resultado en el contador por B después de completar el remuestreo.
- Si la hipótesis nula fuera cierta, p representa la probabilidad de obtener una F tan grande como la que obtuvimos con nuestros datos experimentales.
- Rechazar la hipótesis nula si $p < \alpha$, siendo α el nivel de significación.

Ejemplo 3

Se realizó un experimento para determinar el efecto de las burbujas de aire sobre la resistencia del asfalto. Para fines del experimento, las burbujas se controlan en tres niveles: bajo (2-4 %), medio (4-6 %) y alto (6-8 %). Los datos obtenidos aparecen en la tabla siguiente.

Burbujas de aire								
Bajo	106	90	103	90	79	88	92	95
Medio	80	69	94	91	70	83	87	83
Alto	78	80	62	69	76	85	69	85

¿Afectan de manera significativa los diferentes niveles de burbujas de aire en la resistencia del asfalto? Tomar $\alpha = 0.05$. Resolver el problema mediante un test de remuestreo.

Para implementar un test de remuestreo en R podemos cargar el paquete de R `coin` y llamar la función `oneway_test()`, y mediante el argumento `distribution=approximate(nresample=9999)` indicaremos que el remuestreo se basa en 9.999 permutaciones de las observaciones.

```
library(coin)
## Warning: package 'coin' was built under R version 3.5.3
## Loading required package: survival
## Warning: package 'survival' was built under R version 3.5.3
df<-data.frame(y=c(106,90,103,90,79,88,92,95,80,69,94,91,70,83,87,83,78,80,62,69,76,85,69,85),
               b= c(rep("1. Bajo",8), rep("2. Medio",8), rep("3. Alto",8) ) )
str(df)
## 'data.frame':   24 obs. of  2 variables:
## $ y: num  106 90 103 90 79 88 92 95 80 69 ...
## $ b: Factor w/ 3 levels "1. Bajo","2. Medio",...: 1 1 1 1 1 1 1 1 2 2 ...
set.seed(1234)
oneway_test(y~b,data=df,distribution=approximate(nresample=9999))
##
## Approximative K-Sample Fisher-Pitman Permutation Test
##
## data:  y by b (1. Bajo, 2. Medio, 3. Alto)
## chi-squared = 10.156, p-value = 0.0029
```


Resulta un p -valor de 0.0029, inferior al nivel de significación. Por lo tanto, rechazamos la igualdad de los efectos. Hay diferencias significativas de la resistencia del asfalto según el nivel de burbujas de aire.

1.10. Test no paramétrico basado en rangos

La prueba de Kruskal-Wallis no asume normalidad en los datos, en oposición al tradicional ANOVA. Sí asume, bajo la hipótesis nula, que los datos vienen de la misma distribución. Una forma común en que se viola este supuesto es con datos heterocedásticos.

Sean n_1, n_2, \dots, n_k los tamaños de cada una de los tratamientos y N el total de observaciones. Para el cálculo del estadístico de test se ordenan las n observaciones de menor a mayor y se les asignan rangos desde 1 hasta n . Si la hipótesis nula es cierta, es de esperar que el rango promedio sea aproximadamente igual para los k tratamientos; cuando dichos promedios sean muy diferentes, es un indicio de que H_0 es falsa.

El estadístico de prueba es:

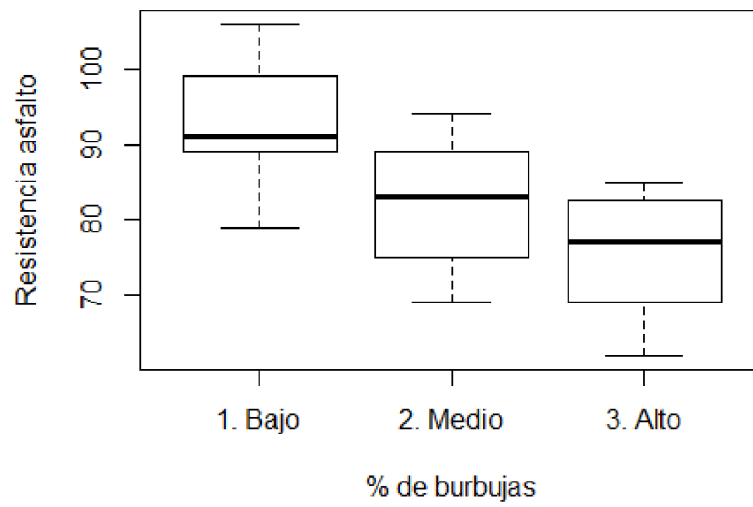
$$K = \frac{12}{N(N+1)} \left(\sum_{i=1}^k \frac{R_i^2}{n_i} \right) - 3(N+1)$$

donde, R_i es la suma de rangos de las observaciones correspondientes al grupo i . La distribución bajo la hipótesis nula del estadístico es aproximadamente χ^2_{k-1} , que requiere un mínimo de cinco observaciones por muestra.

Ejemplo 3 (continuación)

Analicemos los datos del ejemplo 3, en el que se realizó un experimento para determinar el efecto de las burbujas de aire sobre la resistencia del asfalto, con el test de Kruskal-Wallis.

```
kruskal.test(y~b,data=df)
##
##  Kruskal-Wallis rank sum test
##
## data:  y by b
## Kruskal-Wallis chi-squared = 10.578, df = 2, p-value = 0.005047
boxplot(y~b,data=df,xlab="% de burbujas", ylab="Resistencia asfalto")
```



Resulta un p -valor de 0.005, inferior al nivel de significación. Por lo tanto, aceptamos que existen diferencias significativas de la resistencia del asfalto según el nivel de burbujas de aire. El gráfico sugiere que la resistencia del asfalto decrece a medida que el porcentaje de burbujas de aire.

2. Análisis de la varianza de dos factores

Hasta ahora, hemos estado trabajando con diseños totalmente aleatorizados, donde los a tratamientos eran asignados a las unidades experimentales. Los tratamientos no tenían estructura, eran solo tratamientos.

Decimos que hay una estructura factorial de tratamientos cuando un experimento combina los niveles de dos o más factores para crear los tratamientos. Cuando los tratamientos son las combinaciones de los niveles de dos o más factores. No cambiamos la asignación aleatorizada; la diferencia es solo que ahora estamos considerando tratamientos que tienen una estructura factorial.

En particular, vamos a considerar diseños experimentales con dos factores. Además, nos centraremos en diseños balanceados, es decir, en experimentos con n observaciones en cada una de las combinaciones de los niveles de los factores.

2.1. Efectos principales e interacción

Utilizamos la notación y_{ijk} para indicar las respuestas en un diseño factorial de 2 factores. Con esta notación, y_{ijk} es la réplica k en la condición experimental correspondiente a la combinación del nivel i del factor A con el nivel j del factor B .

Podemos organizar los datos en forma de tabla.

	B1	B2	B3
A1	y_{111} \vdots y_{11n}	y_{121} \vdots y_{12n}	y_{131} \vdots y_{13n}
A2	y_{211} \vdots y_{21n}	y_{221} \vdots y_{22n}	y_{231} \vdots y_{23n}
A3	y_{311} \vdots y_{31n}	y_{321} \vdots y_{32n}	y_{331} \vdots y_{33n}

	B1	B2	B3
A4	y_{411} \vdots y_{41n}	y_{421} \vdots y_{42n}	y_{431} \vdots y_{43n}

Si olvidamos por un momento el factor B , y solo consideramos el factor A (filas en la tabla). Estaremos en un diseño de un factor. Podríamos (como en el tema anterior) calcular la variabilidad de la respuesta atribuible al factor A mediante la correspondiente suma de cuadrados SSA, considerando solo los niveles del

factor A . El valor esperado para la respuesta para la fila i es $\mu_i = \mu + \alpha_i$, $\sum_{i=1}^a \alpha_i = 0$.

Podemos revertir el papel de los factores A y B , y considerar el factor B (columnas en la tabla). Podríamos (como en el tema anterior) calcular la variabilidad de la respuesta atribuible al factor B mediante la correspondiente suma de cuadrados SSB, considerando solo los niveles de B . El valor esperado para

la respuesta para la columna j es $\mu_j = \mu + \beta_j$, $\sum_{j=1}^b \beta_j = 0$.

Los efectos α_i y β_j se llaman los efectos principales de los factores A y B , respectivamente. Los efectos principales del factor A miden la variabilidad debida solo a los niveles del factor A (filas en la matriz de datos). Los efectos principales del factor B miden la variabilidad debida solo a los niveles del factor B (columnas en la matriz de datos). Análogamente, las sumas SSA y SSB son las llamadas sumas de cuadrados de los efectos principales del factor A y del factor B , respectivamente.

La variabilidad explicada por los efectos principales es la variabilidad que ocurre a través de las filas o a través las columnas de la matriz de datos. Hay otra variabilidad por describir, que es la variabilidad que ocurre al cambiar filas y columnas simultáneamente. Llamamos a esta variabilidad la interacción entre los factores A y B , y la denotamos mediante SSAB.

El efecto principal de las filas nos dice cómo cambia la respuesta cuando nos movemos de una fila a otra, promediando en todas las columnas. El efecto principal de las columnas nos dice cómo la respuesta cambia cuando se pasa de una columna a otra, promediado en todas las filas.

La interacción nos dice cómo el cambio en la respuesta depende de columnas, cuando se mueve entre las filas, o cómo el cambio en respuesta depende de filas cuando se mueve entre las columnas. En otras palabras, la interacción entre los factores A y B nos dice el cambio en la respuesta media al pasar del nivel i_1 del factor A al nivel i_2 del factor A , dependiendo del nivel de factor B bajo consideración. No podemos decir simplemente que el cambiar de nivel

del factor *A* –esto conlleva un cierto cambio de la respuesta por una cantidad dada– es posible que tengamos una cantidad diferente de cambio según sea el nivel del factor *B*.

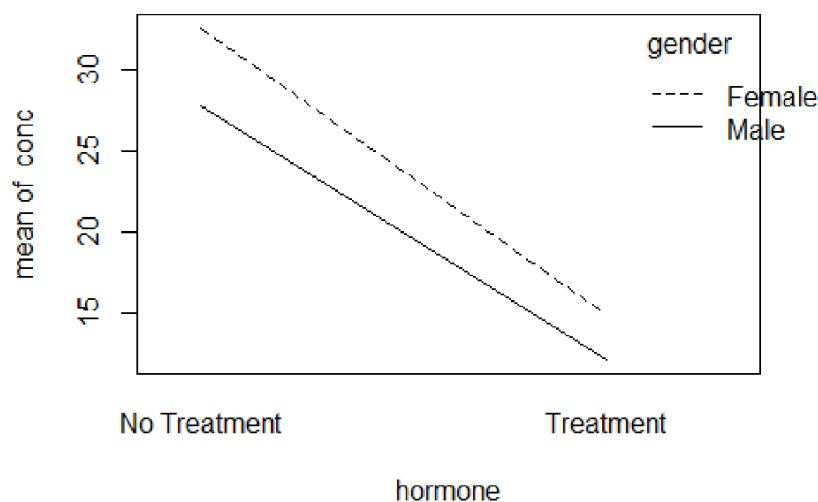
Ejemplo 4

Un equipo investigador analiza si existen diferencias en los niveles de calcio en plasma (mg Ca ml⁻¹) entre sexos en una especie de pájaros que recibieron o no un tratamiento hormonal.

```
conc<-c(16.5, 14.5, 39.1, 32.0,18.4, 11.0, 26.2, 23.8,12.7, 10.8, 21.3,
        28.8,14.0, 14.3, 35.8, 25.0,12.8, 10.0, 40.2, 29.3)
hormone<-rep(c("Treatment","Treatment","No Treatment","No Treatment"),5)
gender<-rep(c("Female","Male"),10)
df<-data.frame(conc,hormone,gender)
```

Podemos utilizar la función `interaction.plot` para estudiar la interacción de los factores principales.

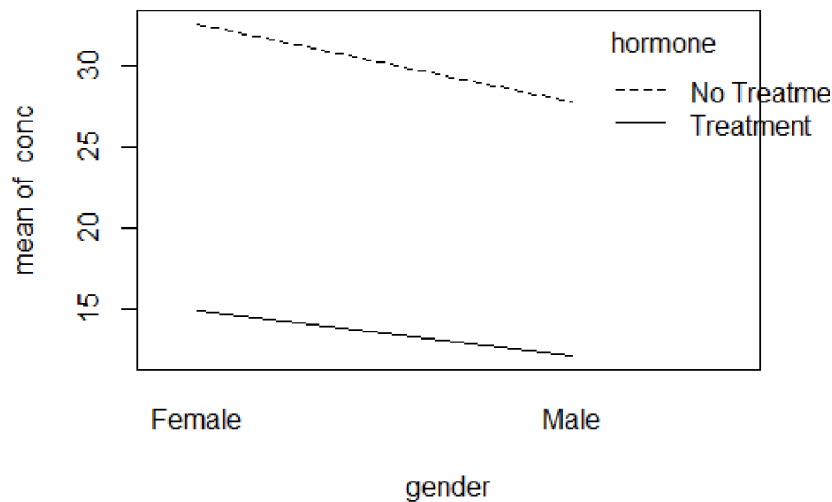
```
interaction.plot(hormone,gender,conc)
```



En el eje X aparecen los dos niveles del factor tratamiento, en el eje Y aparece la media de la variable respuesta (calcio en plasma). Estrictamente, el gráfico se basa en las cuatro medias para las cuatro condiciones experimentales, pero se unen mediante líneas las medias que corresponden al mismo nivel del factor que no está representado en el eje X. Vemos que tanto en machos como en hembras los niveles de calcio son mayores en el grupo no tratado con hormona. También observamos que las hembras tienen mayor calcio en plasma que los machos en cualquiera de los dos tratamientos. Constatamos que el descenso en el calcio en plasma entre sexos es aproximadamente el mismo en los dos tratamientos. Las líneas aparecen paralelas y diremos que no se aprecia interacción.

Podemos intercambiar los factores en la representación de la interacción. Llegamos a conclusiones similares.

```
interaction.plot(gender, hormone,conc)
```



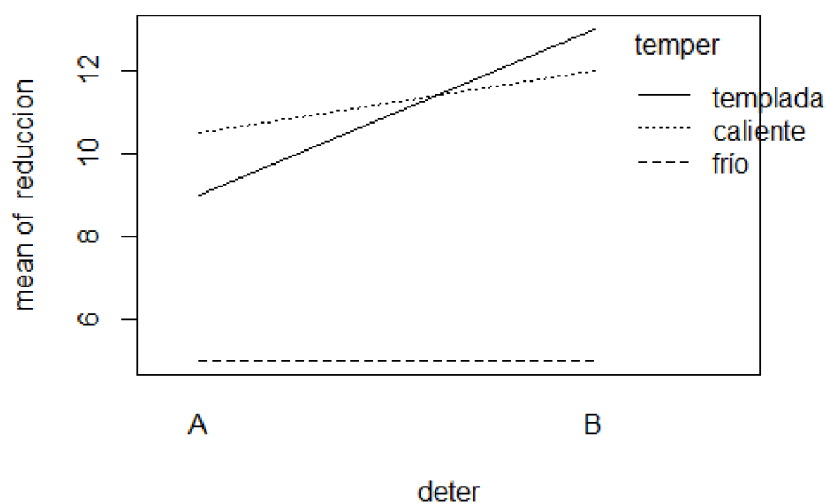
Ejemplo 5

Un departamento de control de calidad quiere comprobar si la marca del detergente para ropa usado y la temperatura de lavado afectan a la cantidad de suciedad eliminada de la ropa, cantidad medida en una escala conveniente, por ejemplo, número de manchas que desaparecen. Con este fin, se comparan dos marcas de detergente (A y B) y se fijan tres niveles de temperatura (frío, templada y caliente). A continuación, se divide la ropa al azar en seis montones con cuatro prendas por combinación de los niveles.

```
reduccion<-c(4,5,6,5,7,9,8,12,10,12,11,9,6,6,4,4,13,15,12,12,12,13,10,13)
temper<-rep(rep(c("frío", "templada", "caliente"),each=4),2)
deter<-rep(c("A","B"),each=12)
df<-data.frame(reduccion,temper,deter)
```

Representemos la interacción

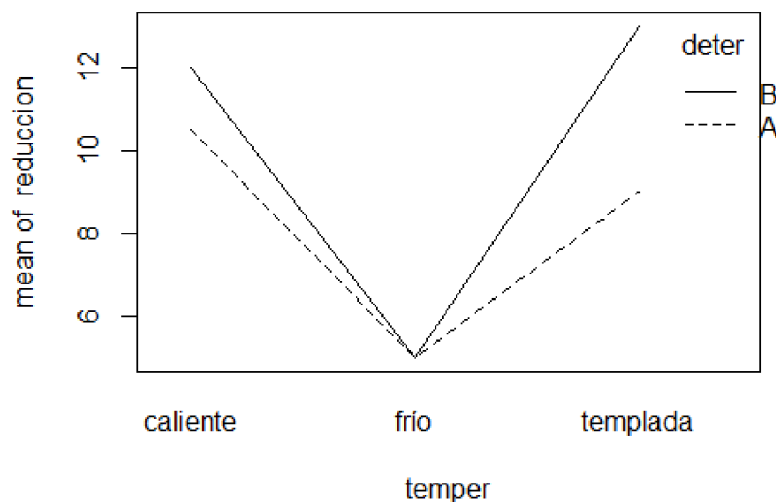
```
interaction.plot(deter, temper, reduccion)
```



Observamos que en este ejemplo hay interacción. Vemos que el detergente B, a temperatura templada, tiene mayor eficacia en la reducción de la suciedad. Las rectas no son paralelas.

A la misma conclusión llegamos si intercambiamos los factores en el gráfico de interacción.

```
interaction.plot(temper, deter, reduccion)
```



Observaciones

Fijaos en que el ejemplo 4 no se corresponde con un diseño completamente aleatorizado, ya que los niveles de uno de los factores (sexo) no han sido asignados aleatoriamente. El diseño ha sido el siguiente: primero se han elegido individuos dentro de cada sexo; posteriormente, se han asignado al azar los tratamientos entre los machos y lo mismo entre las hembras. Dicho de otra manera, se trata de un diseño aleatorizado dentro del factor sexo (bloques). Ello no impide analizar los datos de la misma manera que en un diseño completamente aleatorizado, pero sí que el efecto del factor que no se ha podido aleatorizar no puede ser atribuido únicamente a este factor, sino a otros que pudieran estar relacionados (confusores).

2.2. Parametrización del modelo lineal de dos factores

El diseño experimental de dos factores A y B , donde el factor A tiene a niveles, el factor B tiene b niveles y hay n unidades experimentales asignadas a cada combinación factor-nivel. El modelo es

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijk}$$

donde $i = 1, \dots, a$, $j = 1, \dots, b$, $k = 1, \dots, n$, y_{ijk} denota la respuesta k en el i -ésimo nivel de A y j -ésimo nivel de B , μ la media general, α_i el efecto del nivel i del factor principal A , β_j el efecto del nivel j del factor principal B , $\alpha\beta_{ij}$ la interacción del nivel i del factor A con el nivel j del factor B , e_{ijk} el error es independiente y normalmente distribuido con media cero y varianza σ^2 . Hay un total de $N = nab$ unidades experimentales (observaciones). No debemos con-

fundir $\alpha\beta_{ij}$ con el producto de α_i por β_j , ya que son ideas diferentes. El efecto de la interacción es una medida de lo lejos que los medias de los tratamientos difieren de la simple aditividad.

Hay ab diferentes medias de tratamientos, pero tenemos $1 + a + b + ab$ parámetros, así que tenemos sobreparametrización. Debemos elegir un conjunto de restricciones para que los efectos de los tratamientos estén bien definidos. Las restricciones usuales sobre los parámetros en los modelos de dos factores son:

$$\sum_{i=1}^a \alpha_i = 0, \sum_{j=1}^b \beta_j = 0, \sum_{i=1}^a \alpha\beta_{ij} = 0 \quad \forall j, \sum_{j=1}^b \alpha\beta_{ij} = 0 \quad \forall i$$

Las estimaciones de los parámetros son:

- $\hat{\mu} = \bar{y}_{...}$
- $\hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{...}$
- $\hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{...}$
- $\hat{\alpha}\beta_{ij} = \bar{y}_{ij.} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{...}$

2.3. Análisis de la varianza en diseños balanceados

El análisis de la varianza descansa en un algoritmo para la descomposición de la variabilidad de los datos. Hay una fuente de variabilidad para cada término en el modelo; en un análisis de dos factores estos son el factor A , el factor B , la interacción AB y el error. De tal manera que la suma de cuadrados totales se descompone en la suma de cuadrados de cada término, esto es,

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E$$

donde,

- $SS_A = bn \sum_i^a (\bar{y}_{i.} - \bar{y}_{...})^2$
- $SS_B = an \sum_j^b (\bar{y}_{.j} - \bar{y}_{...})^2$
- $SS_{AB} = n \sum_i^a \sum_j^b (\bar{y}_{ij.} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{...})^2$
- $SS_E = \sum_i^a \sum_j^b \sum_k^n (y_{ijk} - \bar{y}_{ij.})^2$

Ahora, la tabla ANOVA correspondiente al modelo de dos factores con interacción es de la siguiente forma:

Fuente	SS	gl	MS	F
A	SS_A	$a - 1$	$MS_A = SS_A / (a - 1)$	MS_A / MS_E
B	SS_B	$b - 1$	$MS_B = SS_B / (b - 1)$	MS_B / MS_E
AB	SS_{AB}	$(a - 1)(b - 1)$	$MS_{AB} = SS_{AB} / ((a - 1)(b - 1))$	MS_{AB} / MS_E
Error	SS_E	$(n - 1)ab$	$MS_E = SS_E / ((n - 1)ab)$	

Los test de hipótesis para inferir la significación de los factores requieren suposiciones sobre los errores. Asumimos que los errores e_{ijk} son independientes y normalmente distribuidos con la varianza constante en cada una de las condiciones experimentales.

Cuando las suposiciones son verdaderas, los cocientes F definidos en la tabla ANOVA son estadísticos válidos para resolver la significación de los factores.

En este sentido, para testar la significación del factor A , planteamos la hipótesis nula: $H_0: \alpha_1 = \dots = \alpha_a = 0$ frente a la hipótesis alternativa de que alguna/s α_i es/son diferente/s a cero. Para resolver el test anterior usamos el estadístico $F = MS_A / MS_E$ con $(a - 1)$ y $(n - 1)ab$ grados de libertad.

Análogamente, para testar la significación del factor B , planteamos la hipótesis nula $H_0: \beta_1 = \dots = \beta_b = 0$ frente a la hipótesis alternativa de que alguna/s β_j es/son diferente/s a cero. Para resolver el test anterior usamos el estadístico $F = MS_B / MS_E$ con $(b - 1)$ y $(n - 1)ab$ grados de libertad.

De manera similar, para testar la hipótesis de que todos los términos de interacción son nulos usamos el estadístico $F = MS_{AB} / MS_E$ con $(a - 1)(b - 1)$ y $(n - 1)ab$ grados de libertad.

Ejemplo 5 (continuación)

Ajustemos el modelo de dos factores con interacción

```
modelo <- aov(reduccion ~ temper + deter + temper:deter, data = df)
```

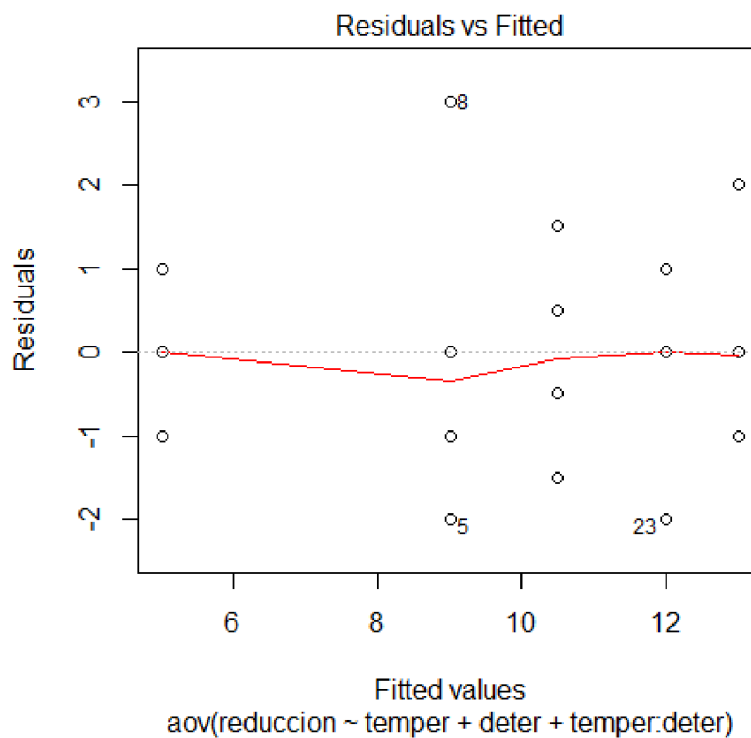
Fijaos en el uso del operador: para designar el término de la interacción dentro de la fórmula. También es posible usar el operador $*$, lo cual incluye tanto el término de interacción como los términos principales. Así, la anterior instrucción es equivalente a:

```
modelo <- aov(reduccion ~ temper*deter, data = df)
```

Validación de las suposiciones

Para la comprobación de la homocedasticidad, podemos hacer un gráfico de residuos frente a valores ajustados, y opcionalmente, un test de Bartlett para testar si la varianza en las seis condiciones experimentales es la misma. Para aplicar el test de Bartlett debemos crear una variable nueva mediante la función `interaction` con la que crear una variable factor con los cruces de los niveles de los dos factores.

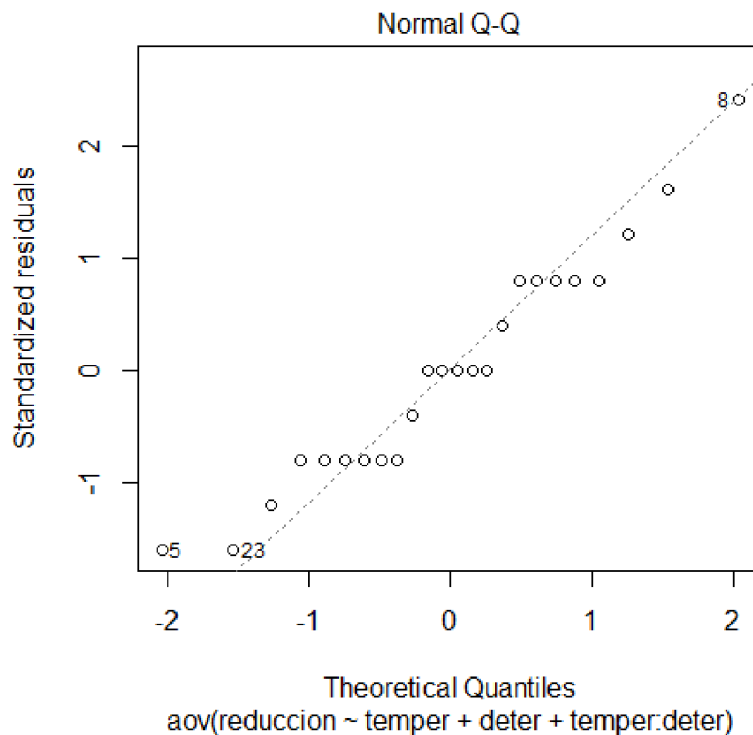
```
plot(modelo, which = 1)
```



```
condition <- with(df, interaction(temper, deter))
condition
## [1] frío.A      frío.A      frío.A      frío.A      templada.A templada.A
## [7] templada.A templada.A caliente.A caliente.A caliente.A caliente.A
## [13] frío.B      frío.B      frío.B      frío.B      templada.B templada.B
## [19] templada.B templada.B caliente.B caliente.B caliente.B caliente.B
## Levels: caliente.A frío.A templada.A caliente.B frío.B templada.B
bartlett.test(reduccion ~ condition, data = df)
##
## Bartlett test of homogeneity of variances
##
## data:  reduccion by condition
## Bartlett's K-squared = 2.665, df = 5, p-value = 0.7515
```

Vemos que el gráfico de residuos sugiere que podría haber una ligera falta de homocedasticidad. Parece que a valores mayores de la variable respuesta se aprecia mayor variabilidad de los residuos. No obstante, dado que el test de Bartlett no resulta significativo, aceptaremos mantener la homocedasticidad. Podemos comprobar la normalidad mediante un qqplot y el test de Shapiro.

```
plot(modelo, which = 2) # equivalente a un qqnorm; qqline
```



```
shapiro.test(residuals(modelo))
##
## Shapiro-Wilk normality test
##
## data: residuals(modelo)
## W = 0.95256, p-value = 0.3075
```

A la vista de la disposición de los cuantiles y del p -valor del test de Shapiro, aceptamos la hipótesis de normalidad.

Pasamos a resolver los contrastes de hipótesis de interés, a saber

$$H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0$$

$$H_1: \alpha_i \neq \alpha_j \text{ para algún } i \neq j$$

$$H_0: \beta_1 = \beta_2 = 0$$

$$H_1: \beta_1 \neq \beta_2$$

$$H_0: \alpha\beta_{11} = \alpha\beta_{12} = \dots = \alpha\beta_{32} = 0$$

$$H_1: \alpha\beta_{ij} \neq \alpha\beta_{i'j'} \text{ para algún } (i, j) \neq (i', j')$$

deberemos calcular la tabla ANOVA.

```
anova(modelo)
## Analysis of Variance Table
##
## Response: reduccion
##          Df Sum Sq Mean Sq F value    Pr(>F)
## temper      2 200.333  100.167  48.7297 5.44e-08 ***
## deter       1  20.167   20.167   9.8108 0.005758 **
## temper:deter 2   16.333    8.167   3.9730 0.037224 *
## Residuals   18   37.000    2.056
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Vemos que las F de los factores principales son significativas, y aceptamos por lo tanto que hay efecto de la temperatura y del detergente. También se aprecia interacción signi-

ficativa entre los niveles de los factores, lo que concuerda con el patrón de los gráficos de interacción.

Es interesante obtener la estimación de los parámetros del modelo. Por un lado, la estimación de la varianza del error resulta de la tabla ANOVA a partir de los MS_E , es por tanto, $\hat{\sigma}^2 = 2.056$.

Las medias por niveles y por condiciones experimentales son:

```
model.tables(modelo, type = "mean")
## Tables of means
## Grand mean
##
## 9.083333
##
## temper
## temper
## caliente      frío templada
##    11.25      5.00    11.00
##
## deter
## deter
##      A      B
## 8.167 10.000
##
## temper:deter
##      deter
## temper      A      B
## caliente 10.5 12.0
## frío      5.0  5.0
## templada  9.0 13.0
```

Y para obtener la estimación de los efectos principales e interacciones, haremos:

```
model.tables(modelo, type = "effects")
## Tables of effects
##
## temper
## temper
## caliente      frío templada
##    2.167    -4.083    1.917
##
## deter
## deter
##      A      B
## -0.9167  0.9167
##
## temper:deter
##      deter
## temper      A      B
## caliente 0.1667 -0.1667
## frío      0.9167 -0.9167
## templada -1.0833  1.0833
```

Dado que los factores principales han resultado significativos, deberemos realizar comparaciones por parejas. Por ejemplo, mediante el test de Tukey:

```
library(agricolae)
HSD.test(modelo, "temper", console = TRUE)
##
## Study: modelo ~ "temper"
##
## HSD Test for reduccion
##
## Mean Square Error:  2.055556
##
## temper, means
##
##      reduccion      std r Min Max
## caliente    11.25 1.4880476 8   9 13
## frío         5.00 0.9258201 8   4   6
## templada    11.00 2.7255406 8   7 15
```

```
##
## Alpha: 0.05 ; DF Error: 18
## Critical Value of Studentized Range: 3.609304
##
## Minimun Significant Difference: 1.829545
##
## Treatments with the same letter are not significantly different.
##
##      reduccion groups
## caliente      11.25      a
## templada      11.00      a
## frío          5.00      b
#HSD.test(modelo, "deter", console = TRUE)
#no sería necesario este paso pero nos va bien en la explicación
```

Detectamos un grupo homogéneo, formado por los niveles de temperatura caliente y templada, y un segundo grupo formado por el nivel frío.

En este ejemplo, dado que el término de interacción es significativo, para determinar la condición experimental óptima debemos hacer comparaciones múltiples de la interacción con el fin de determinar la condición experimental óptima.

```
library(agricolae)
modelolF <- aov(reduccion ~ condition)
HSD.test(modelolF, "condition", console = TRUE)
##
## Study: modelolF ~ "condition"
##
## HSD Test for reduccion
##
## Mean Square Error: 2.055556
##
## condition, means
##
##      reduccion      std r Min Max
## caliente.A      10.5 1.2909944 4 9 12
## caliente.B      12.0 1.4142136 4 10 13
## frío.A          5.0 0.8164966 4 4 6
## frío.B          5.0 1.1547005 4 4 6
## templada.A       9.0 2.1602469 4 7 12
## templada.B      13.0 1.4142136 4 12 15
##
## Alpha: 0.05 ; DF Error: 18
## Critical Value of Studentized Range: 4.49442
##
## Minimun Significant Difference: 3.221872
##
## Treatments with the same letter are not significantly different.
##
##      reduccion groups
## templada.B      13.0      a
## caliente.B      12.0     ab
## caliente.A      10.5     ab
## templada.A       9.0      b
## frío.A          5.0      c
## frío.B          5.0      c
```

Observamos que el HSD.test detecta tres grupos homogéneos: el grupo *a*, formado por los tratamientos templada.B, caliente.B y caliente.A; el grupo *b*, formado por los tratamientos caliente.B, caliente.A y templada.A, y el grupo *c*, formado por los tratamientos frío.A y frío.B. El tratamiento templada.B no es significativamente distinto de caliente.B y caliente.A, pero sí lo es de templada.A.

2.4. Diseños con efectos aleatorios

Los efectos aleatorios son otro enfoque para diseñar experimentos y modelar datos. Los efectos aleatorios son apropiados cuando los tratamientos son muestras aleatorias de una población de tratamientos potenciales.

Supongamos la siguiente situación experimental. Una empresa cuenta con cincuenta máquinas que fabrican cajas de cartón para conservas y quieren entender la variación en la fuerza de los cartones. Eligen diez máquinas al azar de las cincuenta y hacen cuarenta cartones en cada máquina, asignando al azar cuatrocientos lotes de materia prima para cartón a las diez máquinas. Los cartones resultantes se someten a pruebas de resistencia. Esta es un diseño aleatorizado, con diez tratamientos y cuatrocientas unidades.

Podemos plantear un modelo lineal de un factor

$$y_{ij} = \mu + \alpha_i + e_{ij}$$

Los parámetros μ , α_i se consideran constantes desconocidas, tales que la suma de los efectos es cero. El objetivo principal es hacer inferencia sobre los parámetros. Estos modelos se llaman de efectos fijos porque los efectos de los tratamientos se consideran fijos.

Los modelos de efectos fijos no son apropiados para nuestro ejemplo de los cartones. Sigue teniendo sentido descomponer la variabilidad de los datos en la media general, el efecto de los tratamientos y el error, pero la suposición de efectos fijos no tiene sentido por varias razones. Primero, nosotros estamos interesados en hacer inferencia sobre toda la población de máquinas y no solo sobre aquellas que hemos muestreado. Segundo, si planteásemos una replicación del experimento, tendríamos otro conjunto aleatorio de máquinas. Por lo tanto, la inferencia sobre, por ejemplo, el parámetro α_1 en el primer experimento no nos diría nada respecto al α_1 del segundo experimento, de hecho representarían probablemente a máquinas distintas de la población.

Necesitamos una nueva modelización. Los componentes del modelo son los mismos:

$$y_{ij} = \mu + A_i + e_{ij}$$

Asumimos que los e_{ij} son independientes con distribución normal de media cero y varianza σ^2 , como en el caso de los modelos de efectos fijos. En los modelos de efectos aleatorios consideramos que los efectos A_i son variables aleatorias con distribución normal de media cero y varianza σ_A^2 , y que e_{ij} y A_i

son variables aleatorias independientes (usamos letras latinas para indicar que son efectos aleatorios). En consecuencia, la variable respuesta y_{ij} tiene varianza $\sigma_A^2 + \sigma^2$; llamamos a σ_A^2 y σ^2 las componentes de la varianza.

Los parámetros de los modelos de efectos aleatorios son la media general μ , la varianza del error σ^2 y la varianza de los efectos σ_A^2 . Los efectos A_i son variables aleatorias, no parámetros. Para contrastar si el factor aleatorio es significativo, planteamos el contraste de hipótesis sobre σ_A^2 , esto es:

$$H_0: \sigma_A^2 = 0$$

$$H_1: \sigma_A^2 > 0$$

La tabla ANOVA en los modelos de un factor con efectos aleatorios es idéntica a la tabla ANOVA de un factor con efectos fijos. La tabla ANOVA tiene exactamente las mismas filas; factor, residuo y total, y las mismas columnas; Fuente de variación, suma de cuadrados, grados de libertad, cuadrados medios y F.

Para estimar las componentes de la varianza σ^2 y σ_A^2 , debemos tener en cuenta la esperanza matemática de los cuadrados medios del error y del factor A. Se verifica que

Factor	$E(MS)$
A	$\sigma^2 + n\sigma_A^2$
Error	σ^2

Por lo tanto, la estimación de las componentes de la varianza σ^2 y σ_A^2 se obtiene igualando los $E(MS)$ con los cuadrados medios respectivos en la tabla ANOVA. Esto es,

$$\sigma^2 + n\sigma_A^2 = MS_A$$

$$\sigma^2 = MS_E$$

Y, aplicando el método de los momentos para obtener las estimaciones, aislamos las sigmas

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

$$\hat{\sigma}_A^2 = \frac{MS_A - MS_E}{n}$$

Observaciones

Fijaos en que ahora las distintas observaciones de la variable respuesta y_{ij} ya no son todas ellas independientes. Una cosa es que los errores e_{ij} sean independientes y otra cosa es que lo sea la variable respuesta y_{ij} . Para ellos, fijémonos cuánto vale la covarianza para dos datos de un mismo nivel i :

$$\begin{aligned}
\text{cov}(y_{ij}, y_{ij}) &= \text{cov}(\mu + A_i + e_{ij}, \mu + A_i + e_{ij}) = \text{cov}(A_i + e_{ij}, A_i + e_{ij}) = \\
&= \text{cov}(A_i, A_i) + \text{cov}(A_i, e_{ij}) + \text{cov}(e_{ij}, A_i) + \text{cov}(e_{ij}, e_{ij}) = \\
&= \text{cov}(A_i, A_i) = \sigma_A^2 \text{ si la hipótesis nula no es cierta } \neq 0
\end{aligned}$$

Observaciones

Un experimento que tiene tanto factores con efectos fijos como factores con efectos aleatorios se dice que tiene efectos mixtos. La interacción de un efecto fijo y un efecto aleatorio debe ser tratada como un factor aleatorio, ya que una nueva muestra aleatoria de los niveles de factor también conducirá a una nueva muestra aleatoria de interacciones.

Bibliografía

Irizarry, R. A.; Love, M. I. (2015). *Data Analysis for the Life Sciences*. Leanpub.

Ugarte, M. D.; Militino, A. F.; Arnholt, R. (2015). *Probability and Statistics with R* (2.^a ed.). CRC Press / Chapman and Hall. ISBN: 9781466504394.

