

Pruebas de hipótesis con dos muestras

Ejemplo 1

En R se utiliza la función `var.test()` para realizar una prueba F de comparación de las varianzas de dos muestras

```
var.test(x, y, ratio = 1, alternative = "two.sided", conf.level = 0.95)
```

Detalles: `x` y `y` son los vectores que contienen las dos muestras. El argumento opcional `ratio` es la hipótesis nula; el valor por defecto es `ratio = 1`, si no se especifica. El argumento opcional `alternative` da la hipótesis alternativa para la prueba. El valor por defecto para la alternativa es "dos lados", siendo las otras opciones posibles "menor" o "mayor". El argumento opcional `conf.level` da el nivel de confianza que se utilizará en la prueba; el valor por defecto de 0,95 equivale a $\alpha = 0.05$.

Entre los pocos linajes de reptiles que surgieron de la extinción del Mesozoico se encuentran los actuales. Uno de ellos, el tuátara, *Sphenodon punctatum*, de Nueva Zelanda, es el único superviviente de un grupo que, por lo demás, desapareció hace 100 millones de años. A continuación, se indican las masas (en g) de muestras aleatorias de tuátara macho adulto procedentes de dos islotes del Estrecho de Cook. ¿Es la variabilidad de la masa de los machos adultos es diferente en los dos islotes?

	Locación A	Locación B
510	790	650
773	440	600
836	435	600
505	815	575
765	460	452
780	690	320
235		660

La pregunta puede formularse como el siguiente par de hipótesis:

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

$$H_a: \sigma_A^2 \neq \sigma_B^2$$

Obtener los datos en <http://waveland.com/Glover-Mitchell/Example07-1.txt>.

El término NA, "no disponible", indica que no hay datos en esa posición de la tabla. (Las entradas en blanco no se permiten en los marcos de datos; cada fila debe tener el mismo número de entradas). Para realizar una prueba F con estos datos, utilice `var.test()` con los valores predeterminados `ratio = 1` `alternativa = dos. lados`, y `nivel conf.= 0.95` o `a = 0.05`.

```

> var.test(data.Ex07.1$A, data.Ex07.1$B)

##
## F test to compare two variances
##
## data: data.Ex07.1$A and data.Ex07.1$B
## F = 2.5173, num df = 12, denom df = 6, p-value = 0.2661
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.469106 9.385360
## sample estimates:
## ratio of variances
## 2.51734

```

La función `var.test()` proporciona dos formas de responder si las varianzas de las poblaciones son significativamente diferentes. El valor P es 0,226, que es mayor que $\alpha = 0,05$, por lo que no hay pruebas suficientes para rechazar la hipótesis nula de que las varianzas son iguales.

De forma equivalente, `var.test()` proporciona un intervalo de confianza para el valor de F. Como es habitual, si el valor hipotetizado (aquí esperamos que la razón F sea 1) cae dentro del intervalo de confianza, entonces se retiene la hipótesis nula. En caso contrario, se rechaza la hipótesis nula. En este ejemplo, el intervalo de confianza del 95 por ciento [0,469, 9,385] contiene el valor esperado de F de 1. Así que la hipótesis nula no se rechaza.

Ejemplo 2

En una planta de procesamiento de frutas, la máquina que se utiliza normalmente para envasar fresas congeladas produce paquetes con una media de 250 g/caja. Se está estudiando una nueva máquina que procesará mucho más rápido, pero que puede producir resultados más variables. Para investigar esta preocupación, el supervisor de control de calidad midió el contenido de 50 cajas producidas por cada máquina y encontró $s_O^2 = 25 \text{ g}^2$ y $s_N^2 = 64 \text{ g}^2$. A partir de sus resultados ¿se confirman sus sospechas?

Las hipótesis son:

$$H_0: \sigma_N^2 \leq \sigma_O^2 \quad E(F) \leq 1$$

$$H_a: \sigma_N^2 > \sigma_O^2 \quad E(F) > 1$$

Realizando una prueba F para determinar si las varianzas son iguales. Como no se dan los datos originales, no se puede utilizar la función `var.test()`. Este problema debía hacerse a mano. Sin embargo, sea creado una función llamada `f.test2()` que puede utilizarse aquí.

`f.test2(sx, nx, sy, ny, alternativa = "dos caras", nivel de confianza = 0,95)`
 Detalles: `sx` y `sy` son las desviaciones estándar de las dos muestras y `nx` y `ny` son los tamaños de muestra correspondientes. El argumento opcional `alternativa` da la hipótesis alternativa para la prueba. La alternativa por defecto es "dos caras" y las otras opciones posibles son "menor" o "mayor". El argumento opcional `conf.level` da el nivel de confianza que se utilizará en la prueba; el valor por defecto de 0,95 equivale a $\alpha = 0,05$.

Observe que la prueba utiliza las desviaciones estándar de las muestras como entradas, no las varianzas.

Descargue la función utilizando el comando `source()`. La función sólo necesita ser descargar una vez por sesión

En este ejemplo, la hipótesis alternativa es igual mayor debido a $H_a: \sigma_N^2 > \sigma_O^2$ or $E(F) > 1$. Adicionalmente, con el nivel de confianza = 0.95.

```
> source("http://waveland.com/Glover-Mitchell/f.test2.txt")

## Downloaded: f.test2( ).

> f.test2(sx = sqrt(64), nx = 50, sy = sqrt(25), ny = 50, alternative = "greater")

##
## F test to compare two variances
##
## data:  sx = 8, nx = 50; and sy = 5, ny = 50
## F = 2.56, num df = 49, denom df = 49, p-value = 0.0006496
## alternative hypothesis: true ratio of variances is greater than 1
## 95 percent confidence interval:
##  1.59274      Inf
## sample estimates:
## ratio of variances
##                2.56
```

`f.test2()` proporciona las mismas dos formas que `var.test()` para responder si las varianzas son significativamente diferentes. El valor P es aproximadamente 0,00065, que es mucho más pequeño que $\alpha = 0,05$, por lo que hay una fuerte evidencia para rechazar la hipótesis nula de que la varianza de la nueva máquina es menor o igual que la varianza de la máquina antigua. De forma equivalente, `f.test2()` proporciona el intervalo de confianza para el valor de F para la hipótesis alternativa especificada. En este ejemplo, el intervalo de confianza del 95% $[1,593, \infty)$ no contiene el valor F esperado de 1. Por lo tanto, el valor F esperado de 1. Por tanto, se rechaza la hipótesis nula.

Ejemplo 3

Volviendo a las tuataras neozelandesas del ejemplo 7.1, nos preguntamos ahora ¿Difiere la masa media de los machos adultos entre la localidad A y la localidad B?

Utiliza los `data.Ex07.1` leídos anteriormente. Las muestras son independientes (no emparejadas), la primera observación en el lugar A no tiene ninguna relación especial con ninguna observación en el lugar B. La hipótesis de igualdad de varianzas se probó en el ejemplo 7.1 y se aceptó H_o se aceptó. Las nuevas hipótesis sobre las medias son

$$H_0: \mu_A = \mu_B$$

$$H_a: \mu_A \neq \mu_B$$

que conduce a una prueba de dos caras con el valor predeterminado $\mu_1 - \mu_2 = 0$. Utilice el nivel `conf.level = 0,95` o $\alpha = 0,05$. El ejemplo 1 mostró que las varianzas de las dos poblaciones deben considerarse iguales. Así que establezca `var.equal = TRUE`. La prueba t sobre estos datos, utilizando las opciones descritas, viene dada por

```

> t.test(data.Ex07.1$A, data.Ex07.1$B, var.equal = TRUE)

##
## Two Sample t-test
##
## data: data.Ex07.1$A and data.Ex07.1$B
## t = 0.8217, df = 18, p-value = 0.422
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -104.297 238.297
## sample estimates:
## mean of x mean of y
## 618 551

```

El valor P es 0,422, que es mayor que $\alpha = 0,05$, por lo que no hay pruebas suficientes para rechazar la hipótesis nula de que las medias son iguales. De forma equivalente, el intervalo de confianza del 95 por ciento es $[-104,297, 238,297]$. El valor de la hipótesis es $\mu_1 - \mu_2 = 0$, que se encuentra dentro de este intervalo. Por tanto, no se rechaza la hipótesis nula. Siga suponiendo que las medias de las masas de las dos poblaciones son iguales.

Ejemplo 4

Una ornitóloga que estudiaba varias poblaciones de urraca australiana, *Gymnorhina tibicen*, capturó con redes de niebla 25 adultos de una zona rural aislada y 36 de una zona urbana de picnic. Midió la longitud total del cuerpo en centímetros e informó de lo siguiente estadísticas resumidas

	Rural	Urbana
\bar{X}	38 cm	35 cm
S	4 cm	3 cm
N	25	36

Dado que los excursionistas suelen alimentar a las urracas, se piensa que la población urbana podría estar algo atrofiada debido al consumo de alimentos procesados en lugar de naturales. Analiza completamente los datos dados para ver si apoyan esta hipótesis

Las muestras son independientes, pero ¿proceden de poblaciones con la misma varianza? Realice una prueba F con $\alpha = 0,05$ para responder a esta pregunta. Las hipótesis son

$$H_0: \sigma_r^2 = \sigma_u^2$$

$$H_a: \sigma_r^2 \neq \sigma_u^2$$

Como no se dan los datos originales, no se puede utilizar la función `var.test()`. En su lugar, utilice `f.test2()` como en el ejemplo 2 con `sx = 4`, `nx = 25`, `sy = 3` y `ny = 36`.

```

> f.test2(sx = 4, nx = 25, sy = 3, ny = 36)

```

```
##
## F test to compare two variances
##
## data:  sx = 4, nx = 25; and sy = 3, ny = 36
## F = 1.7778, num df = 24, denom df = 35, p-value = 0.1182
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.862306 3.863724
## sample estimates:
## ratio of variances
##          1.77778
```

El valor P es 0,118, que es mayor que $\alpha = 0,05$. Alternativamente, observe que el valor esperado de F de 1 está en el intervalo de confianza. La hipótesis nula de que las dos varianzas son iguales no se puede rechazar.

Ahora se puede abordar la cuestión del tamaño medio. ¿Son las urracas más pequeñas en las zonas urbanas que en las rurales?

$$H_0: \mu_r \leq \mu_u$$

$$H_a: \mu_r > \mu_u$$

Esto requiere una prueba t de una cola porque anticipamos una desviación de la igualdad en una dirección particular

R no proporciona una función para llevar a cabo una prueba t cuando no se dan los datos originales. Se Puede utilizar `t.test2()` con argumentos adicionales para realizar el análisis

`t.test2(mx, sx, nx, my, sy, ny, mu = 0, alternative = "two.sided", paired=FALSE, var.igual = FALSE, nivel.conf.=0.95)`
 Detalles: mx, sx y nx son la media, la desviación estándar y el tamaño de la muestra para la muestra x y my, sy y ny son los valores correspondientes para la muestra y. Todos los argumentos, excepto mx, sx y nx, son opcionales. Si my, sy y ny no están presentes, se lleva a cabo una prueba de una sola muestra. El argumento opcional mu es el valor esperado de la diferencia de las medias para una prueba de dos muestras (o el valor esperado de la media en una prueba de una muestra). El valor por defecto es mu = 0. El argumento opcional alternative es la hipótesis alternativa. El valor por defecto es "dos lados", siendo las otras las otras opciones posibles son "menor" o "mayor". El argumento paired indica si los datos están emparejados o no y por defecto es FALSE o datos no emparejados, si no se especifica. El argumento var.equal especifica si las varianzas asociadas a x e y son iguales o desiguales, por defecto FALSE o desigual, si no se especifica. Por último, el argumento conf.level da el nivel de confianza que se utilizará en la prueba t; el valor por defecto de 0,95 es equivale a $\alpha = 0,05$.

```
> source("http://waveland.com/Glover-Mitchell/t.test2.txt")
## Downloaded: t.test2( ).
```

La media, la desviación estándar y el tamaño de la muestra para la primera muestra son $mx = 38$, $sx = 4,0$ y $nx = 25$, respectivamente. Del mismo modo, para la segunda muestra $my = 35$, $sy = 3,0$ y $ny = 36$. La hipótesis alternativa es que la media de la muestra rural x es mayor que la de la muestra urbana y. La prueba F anterior indica que `var.equal = TRUE`.

```
> t.test2(mx = 38, sx = 4.0, nx= 25, my = 35, sy = 3.0, ny = 36,
+ alternative = "greater", var.equal = TRUE)
```

```
##
```

```
## Two Sample t-test
```

```
##
```

```
## data: mx = 38, sx = 4, nx = 25; and my = 35, sy = 3, ny = 36
```

```
## t = 3.3478, df = 59, p-value = 0.0007113
```

```
## alternative hypothesis: true difference in means is greater than 0
```

```
## 95 percent confidence interval:
```

```
## 1.50253      Inf
```

```
## sample estimates:
```

```
## mean of x mean of y
```

```
##      38      35
```

El valor P de 0,0007, que es mucho menor que $\alpha = 0,05$, indica que debe rechazarse la hipótesis nula. De forma equivalente, el intervalo de confianza del 95 por ciento para $\mu_1 - \mu_2$ es $[1,503, \infty)$. Este intervalo no incluye 0, por lo que la hipótesis nula debe ser rechazada. Hay pruebas de que la masa media de las urracas rurales es mayor que la de las urracas urbanas.

Ejemplo 5

Como parte de un estudio más amplio sobre los efectos de las anfetaminas, un farmacólogo quiso investigar la afirmación de que las anfetaminas reducen el consumo total de agua. Inyectó a 15 ratas de laboratorio estándar con una dosis adecuada de anfetamina y a 10 con una solución salina como control. Durante las siguientes 24 horas midió la cantidad de agua consumida por cada rata y expresó los resultados en ml/kg de peso corporal:

	Anfetaminas	Salina
\bar{X}	15	10
S	115	135
N	40	15

¿La anfetamina suprime significativamente el consumo de agua? Utilice un nivel alfa de 0,05 para cualquier prueba de hipótesis.

En primer lugar, realice una prueba F preliminar con $\alpha = 0,05$ para las varianzas

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

$$H_a: \sigma_A^2 \neq \sigma_S^2$$

```
> f.test2(sx = 40, nx = 15, sy = 15, ny = 10)

##
## F test to compare two variances
##
## data:  sx = 40, nx = 15; and sy = 15, ny = 10
## F = 7.1111, num df = 14, denom df = 9, p-value = 0.005652
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  1.87235 22.82169
## sample estimates:
## ratio of variances
##      7.1111
```

El diminuto valor P de 0,0057 indica que debe rechazarse la hipótesis nula. Hay pruebas sólidas de que las dos varianzas no son iguales. Las hipótesis para las medias son:

$$H_0: \mu_A \geq \mu_S$$

$$H_a: \mu_A < \mu_S$$

Como sólo se proporcionan los datos de resumen, utilice la función `t.test2()` para realizar el análisis. Los resultados de la prueba F indican que `var.equal = FALSE`. Al revisar las hipótesis de las medias, `alternative = "less"`.

```
> t.test2(mx = 115, sx = 40, nx = 15, my = 135, sy = 15, ny = 10,
+ alternative = "less", var.equal = FALSE)

##
## Welch Two Sample t-test
##
## data:  mx = 115, sx = 40, nx = 15; and my = 135, sy = 15, ny = 10
## t = -1.7598, df = 19.2, p-value = 0.04719
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##  -Inf -0.35863
## sample estimates:
## mean of x mean of y
##      115      135
```

Como el valor P es 0,047 es menor que $\alpha = 0,05$, se rechaza H_0 . Los datos apoyan la afirmación de que las anfetaminas suprimen significativamente el consumo de agua.

Ejemplo 6

Los triglicéridos son componentes de la sangre que se cree que desempeñan un papel en la enfermedad arterial coronaria. Se realizó un experimento para comprobar si el ejercicio regular podía reducir los niveles de triglicéridos. Los investigadores midieron la concentración de triglicéridos en el suero sanguíneo de 7 voluntarios varones antes y después de participar en un programa de ejercicio de 10 semanas.

Los resultados se muestran a continuación. Obsérvese la variación de los niveles de una persona a otra, tanto antes como después. Pero está claro un patrón general de reducción después del ejercicio.

Una pregunta obvia es: "¿Cuánto redujo el ejercicio los niveles de triglicéridos?" Esto es pedir una estimación puntual o un intervalo de confianza para la diferencia media

Antes	0.87	1.13	3.14	2.14	2.98	1.18	1.60
Después	0.57	1.03	1.47	1.43	1.20	1.09	1.51

Datos: <http://waveland.com/Glover-Mitchell/sign.txt>.

```
> source("http://waveland.com/Glover-Mitchell/sign.txt")

## Downloaded: sign.test( ).

> data.Ex07.11 <- read.table("http://waveland.com/Glover-Mitchell/Example07-11.txt"
+ header = TRUE)
> data.Ex07.11

##   Before After
## 1    0.87  0.57
## 2    1.13  1.03
## 3    3.14  1.47
## 4    2.14  1.43
## 5    2.98  1.20
## 6    1.18  1.09
## 7    1.60  1.51
```

En este caso se requiere una prueba de cola derecha, $H_a: MBef > MAft$, porque los investigadores están probando si el ejercicio regular reduce los niveles de triglicéridos. Por tanto, establezca la $alternative = "greater"$. La hipótesis nula es $H_0: MBef \leq MAft$, los niveles de triglicéridos antes del ejercicio no son mayores que después. Utilice el `sign.test()` con los valores predeterminados $\mu = 0$ y $conf.level = 0,95$ o $\alpha = 0,05$. Cuando se especifican dos archivos de datos, `sign. Test()` realiza automáticamente una prueba emparejados.

```
> sign.test(data.Ex07.11$Before, data.Ex07.11$After, alternative = "greater")

##
## Paired-samples Sign-Test
##
## data:  x = data.Ex07.11$Before and y = data.Ex07.11$After
## s = 7, p-value = 0.007813
## alternative hypothesis: true median difference of x - y is greater than 0
## 95 percent confidence interval:
##  0.09  Inf
## sample estimates:
## median of x - y
##              0.3
```

El diminuto valor P de 0,0078 es inferior a $\alpha = 0,05$, por lo que hay pocas pruebas que apoyen la hipótesis nula. Hay pruebas sólidas de que hay una caída en la mediana de los niveles de triglicéridos después del ejercicio

Ejemplos de ANOVA

Ejemplo 7

Una gerontóloga que investigaba varios aspectos del proceso de envejecimiento quería comprobar si mantenerse "delgado y cascarrabias", es decir, por debajo del peso corporal normal, alargaría la vida. Asignó aleatoriamente a ratas recién nacidas de una línea altamente endogámica una de las tres dietas: (1) acceso ilimitado a la comida, (2) el 90% de la cantidad de comida que una rata de ese tamaño comería normalmente, o (3) el 80% de la cantidad de comida que una rata de ese tamaño comería normalmente. Mantuvo a las ratas con las tres dietas durante toda su vida y registró su esperanza de vida (en años). ¿Hay pruebas de que la dieta afectó a la duración de la vida en este estudio?

Ilimitado	90% dieta	80% dieta
2.5	2.7	3.1
3.1	3.1	2.9
2.3	2.9	3.8
1.9	3.7	3.9
2.4	3.5	4.0

```
> data.Ex08.1 <- read.table("http://waveland.com/Glover-Mitchell/Example08-1.txt",
+ header = TRUE)
> data.Ex08.1
```

```
##    Lifespan    Diet
## 1      2.5 Unlimited
## 2      3.1 Unlimited
## 3      2.3 Unlimited
## 4      1.9 Unlimited
## 5      2.4 Unlimited
## 6      2.7   90%Diet
## 7      3.1   90%Diet
## 8      2.9   90%Diet
## 9      3.7   90%Diet
## 10     3.5   90%Diet
## 11     3.1   80%Diet
```

```
## 12     2.9   80%Diet
## 13     3.8   80%Diet
## 14     3.9   80%Diet
## 15     4.0   80%Diet
```

Tenga en cuenta que el formato de los datos difiere del texto. Cada fila consiste en una respuesta (Vida útil) a un nivel de tratamiento (Dieta). Realizamos el ANOVA con el comando `aov()` y ponemos el resultado en la tabla `aov.Ex08.1`. Tenga en cuenta que estamos realizando un análisis de la varianza de la duración de la vida por los factores de la dieta. Los argumentos de `aov()` lo reflejan utilizando los nombres de las cabeceras de las columnas correspondientes en el marco de datos `data.Ex08.1`.

```
> aov.Ex08.1<- aov(Lifespan ~ Diet, data = data.Ex08.1)
```

Para ver realmente los resultados del ANOVA, utilice el comando `summary()`.

```
> summary(aov.Ex08.1)

##           Df Sum Sq Mean Sq F value Pr(>F)
## Diet         2   3.15   1.573     7.7 0.0071 **
## Residuals    12   2.45   0.204
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Observe los dos asteriscos ** al final de la fila de la Dieta en la tabla. Utilizando los códigos de significación en la salida, esto significa que la prueba F es significativa al nivel 0,01. De hecho, el valor P se indica en la tabla como $\text{Pr}(>F) = 0,00707$. Para ver las medias por tratamiento, utilice la función `tapply()`. Este comando se utiliza para aplicar una función (en este caso, la media) a grupos (en este caso, el tipo de dieta) dentro de una tabla o marco de datos. En otras palabras, `tapply()` puede considerarse como una abreviatura de "aplicar a la tabla".

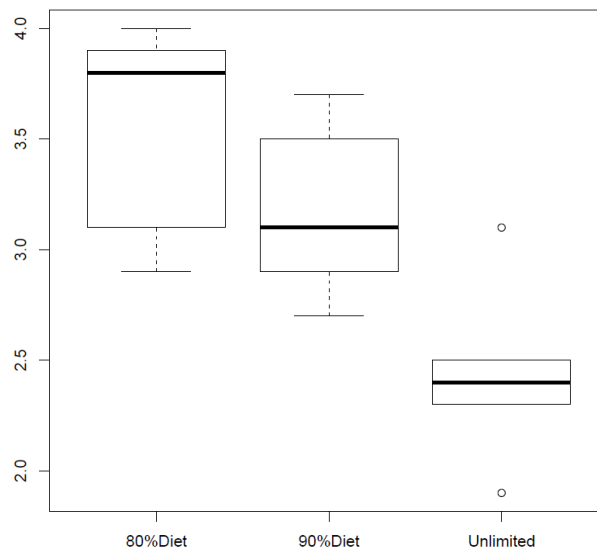
`tapply(x, g, FUN = función)`
 Detalles: x es típicamente un vector (columna) de datos y g es una lista de factores correspondientes de la misma longitud que x. FUN es la función que se aplicará a cada grupo de celdas definidas por las variables categóricas en g.

```
> tapply(data.Ex08.1$Lifespan, data.Ex08.1$Diet, mean)

##      80%Diet      90%Diet Unlimited
##         3.54         3.18         2.44
```

El gráfico de caja de los datos proporciona una indicación de si alguno de los tratamientos (dietas) es significativamente diferente. Puede que no nos tomemos el tiempo de hacer esto a mano, pero R lo hace fácil.

```
> boxplot(Lifespan ~ Diet, data = data.Ex08.1)
```



La prueba F fue significativa (véase la tabla ANOVA anterior) y el gráfico de caja indica que hay una diferencia entre algunas medias de vida. Para llevar a cabo las pruebas t de Bonferroni-Holm t para localizar cualquier diferencia en las medias de vida, utilice

```
prueba.t.pairwise(x, g, método.p.adjust = "holm", nivel.conf = 0.95)
```

Detalles: x son los datos de medición o respuesta, g contiene los niveles de tratamiento correspondientes, y p.adjust.method = "holm" ajusta los valores P de la secuencia de pruebas según el método seleccionado. La abreviatura p.adj puede utilizarse en lugar de p.adjust.method. El argumento opcional conf.level da el nivel de confianza que se utilizará en la prueba; el valor por defecto de 0,95 equivale a $\alpha = 0,05$. Hay argumentos adicionales para esta función que no son necesarios en este momento.

En este ejemplo, data.Ex08.1\$Lifespan es el dato de respuesta y data.Ex08.1\$Diet contiene el tratamiento correspondiente (grupo).

```
> pairwise.t.test(data.Ex08.1$Lifespan, data.Ex08.1$Diet, p.adj = "holm")

##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  data.Ex08.1$Lifespan and data.Ex08.1$Diet
##
##           80%Diet 90%Diet
## 90%Diet    0.232   -
## Unlimited 0.007   0.047
##
## P value adjustment method: holm
```

El resultado de pairwise.t.test() se presenta en forma de tabla. La entrada en la columna 80%Diet y la fila 90%Diet es 0,232. Este es el valor P para la prueba de la diferencia de medias, por lo que no es significativo al nivel $\alpha = 0,05$. Se aplican observaciones similares a las demás entradas. La tabla indica que la dieta ilimitada es diferente de las dietas del 80% y del 90%; los valores P correspondientes son 0,007 y 0,047, respectivamente. Este último es apenas significativo.

Ejemplo 8

Un endocrinólogo que estudia los efectos genéticos y ambientales en la producción de insulina del tejido pancreático crió cinco camadas de ratones experimentales. A los 2 meses de edad sacrificó a los ratones, diseccionó el tejido pancreático y trató las muestras de tejido con una solución de glucosa. La cantidad de insulina liberada por estas muestras se registró en pg/ml. ¿Hay diferencias significativas en la liberación de insulina entre las camadas?

Camada				
1	2	3	4	5
9	2	3	4	8
7	6	5	10	10
5	7	9	9	12
5	11	10	8	13
3	5	6	10	11

En este ejemplo, las camadas son una muestra aleatoria de camadas de ratones. El objetivo del estudio es determinar si existe una variabilidad significativa entre camadas que pueda atribuirse a diferencias genéticas entre ellas. Las camadas individuales no son de interés. El endocrinólogo sólo está interesado en una afirmación general de que existe una variabilidad significativa entre camadas (un

componente genético significativo) o no. Este tipo de estudio se analiza como efectos aleatorios o Modelo II ANOVA. Para este experimento las hipótesis son

H_0 : No hay una variabilidad significativa entre las camadas, $\sigma_{camada}^2 = 0$

H_a : Existe una importante variabilidad entre las camadas, $\sigma_{camada}^2 > 0$

Los análisis preliminares de los ANOVAs del Modelo I y del Modelo II son idénticos, pero los ANOVAs del Modelo II no requieren la separación de medias. Para realizar el análisis, lea y anote al menos algunos de los datos en <http://waveland.com/Glover-Mitchell/Example08-3.txt>.

```
> data.Ex08.3 <- read.table("http://waveland.com/Glover-Mitchell/Example08-3.txt",
+ header = TRUE)
> head(data.Ex08.3, n = 2)

##    Insulin Litter
## 1         9      I
## 2         7      I

> tail(data.Ex08.3, n = 2)

##    Insulin Litter
## 24        13      V
## 25        11      V
```

Las dos columnas son la medida de la respuesta (insulina) y los niveles de tratamiento (camada). Realice el ANOVA utilizando el comando `aov()` y ponga el resultado en la tabla `aov.Ex08.3`. Utilice el comando `summary()` para imprimir la tabla ANOVA.

```
> aov.Ex08.3 <- aov(Insulin ~ Litter, data = data.Ex08.3)
> summary(aov.Ex08.3)

##              Df Sum Sq Mean Sq F value Pr(>F)
## Litter         4   83.8   20.96    3.07  0.04 *
## Residuals     20  136.4    6.82
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

El asterisco * al final de la fila Litter en la tabla ANOVA significa que la prueba F es significativa al nivel 0,05. El valor P es de 0,04. Por lo tanto, existe una variabilidad significativa entre las camadas. El Modelo II ANOVA, un estudio de efectos aleatorios, tiene una sola pregunta en mente que se responde con la prueba F global. Las técnicas de separación de medias no se realizan junto con este tipo de ANOVA.

Ejemplo 9

Los valores de la siguiente tabla son las anchuras máximas medidas de cráneos egipcios masculinos de diferentes épocas (basados en datos de *The Ancient Races of the Thebaid*, de Thomson y Randall Maciver, 1905, Clarendon Press). Los cambios en la forma de la cabeza a lo largo del tiempo sugieren

que se produjo un mestizaje con poblaciones inmigrantes. Utilice un nivel de significación de 0,05 para probar la afirmación de que las diferentes épocas no tienen toda la misma media.

4000 BCE	1850 BCE	150 CE
131	129	128
138	134	138
125	136	136
129	137	139
132	137	141
135	129	142
132	136	137
134	138	145
138	134	137

```
> data.Ex08.4 <- read.table("http://waveland.com/Glover-Mitchell/Example08-4.txt",
+ header = TRUE)
> tail(data.Ex08.4, n = 3)

##      Breadth  Era
## 25      137 150CE
## 26      145 150CE
## 27      137 150CE
```

Realice el ANOVA de la amplitud por Era utilizando `aov()`.

```
> aov.Ex08.4 <- aov(Breadth ~ Era, data = data.Ex08.4)
> summary(aov.Ex08.4)

##              Df Sum Sq Mean Sq F value Pr(>F)
## Era              2    139    69.4    4.05  0.031 *
## Residuals       24    411    17.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

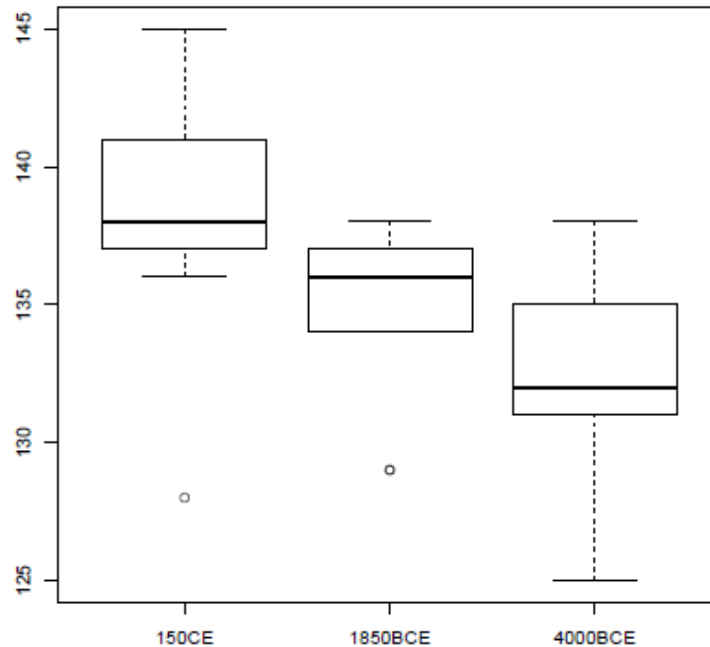
El sencillo * al final de la fila de la Era en la tabla significa que la prueba F es significativa a el nivel 0,05. El valor P es 0,031. Para aplicar la función media a cada uno de los tratamientos, utilice la función `tapply()`.

```
> tapply(data.Ex08.4$Breadth, data.Ex08.4$Era, mean)

##      150CE 1850BCE 4000BCE
## 138.111 134.444 132.667
```

Para crear un diagrama de caja de la Anchura por Era utilice

```
> boxplot(Breadth ~ Era, data = data.Ex08.4)
```



Dado que la prueba F fue significativa, realice las pruebas t por pares de Bonferroni-Holm para localizar las diferencias en la media de la amplitud de las distintas épocas. Utilice la prueba `pairwise.t()` con el ajuste de probabilidad establecido en "holm".

```
> pairwise.t.test(data.Ex08.4$Breadth, data.Ex08.4$Era, p.adj = "holm")

##
## Pairwise comparisons using t tests with pooled SD
##
## data: data.Ex08.4$Breadth and data.Ex08.4$Era
##
##      150CE 1850BCE
## 1850BCE 0.14  -
## 4000BCE 0.03 0.37
##
## P value adjustment method: holm
```

La entrada en la columna 150CE y la fila 1850BCE es 0,14, que es el valor P para la prueba de la diferencia de medias. Esta diferencia no es significativa al nivel $\alpha = 0,05$. Observaciones similares se aplican a las demás entradas. La tabla indica que sólo las épocas de 150 EC y 4000 BCE tienen amplitudes medias significativamente diferentes.

Como alternativa, las comparaciones pareadas pueden realizarse mediante la prueba de Tukey con `TukeyHSD()` aplicada a los resultados del ANOVA, `aov.Ex08.4`.

`TukeyHSD(fit, ordered = TRUE, conf.level = 0.95)`
 Detalles: `fit` es la salida de la función `aov()` (ANOVA) realizada anteriormente en el marco de datos correspondiente. El argumento `ordered` especifica si se ordenan las medias en el análisis. El valor por defecto es `FALSE`, pero para ajustarse más al procedimiento del texto, utilice `ordered = TRUE`. El argumento opcional `conf.level` da el nivel de confianza que se utilizará en la prueba; el valor por defecto de 0,95 equivale a $\alpha = 0,05$.

```

> TukeyHSD(aov.Ex08.4, ordered = TRUE)

## Tukey multiple comparisons of means
## 95% family-wise confidence level
## factor levels have been ordered
##
## Fit: aov(formula = Breadth ~ Era, data = data.Ex08.4)
##
## $Era
##              diff      lwr      upr    p adj
## 1850BCE-4000BCE 1.77778 -3.09455  6.65010 0.638624
## 150CE-4000BCE   5.44444  0.57212 10.31677 0.026465
## 150CE-1850BCE   3.66667 -1.20566  8.53899 0.166403

```

La tabla de resultados proporciona un intervalo de confianza del 95% para la diferencia de cada par de medias y el valor P correspondiente. Sólo las épocas de 150 a.C. y 4000 a.C. tienen anchuras medias significativamente diferentes. El resultado es el mismo que con las pruebas t de Bonferroni-Holm.

Ejemplo 10

Los murciélagos de la fruta son importantes agentes de dispersión de muchas especies de higos. Por ejemplo, algunas especies de murciélagos frugívoros recorren distancias superiores a los 10 km por noche y, por tanto, son capaces de dispersar semillas de higo a grandes distancias. Investigadores de Filipinas recogieron "fragmentos" de higo no consumidos generados cuando los murciélagos se alimentaban, masas de higo masticado pero expulsado ("ejecta") y masas fecales ("splats") bajo cinco higueras, *Ficus chrysalis*, en las montañas de la isla de Negros en Filipinas. Se realizaron pruebas de germinación de las semillas en estas masas y se registraron los porcentajes de semillas que germinaron. ¿Hubo una diferencia significativa en los porcentajes de germinación de las semillas de higo según la forma en que el murciélago interactuó con las semillas? (Basado en las ideas de: Utzurrum, R y P. Heideman. 1991. Differential ingestion of viable vs nonviable *Ficus* seeds by fruit bats. *Biotropica* 23(3): 311-312.)

Fragments	Ejecta	Splats
47	3	88
50	20	90
55	46	91
68	54	91
	76	92
	87	96

```

> data.Ex08.5 <- read.table("http://waveland.com/Glover-Mitchell/Example08-5.txt",
+ header = TRUE)
> head(data.Ex08.5, n = 2)

##   Rate Treatment
## 1   47   Fragment
## 2   50   Fragment

> tail(data.Ex08.5, n = 2)

##   Rate Treatment
## 15   92     Splat
## 16   96     Splat

```

Las dos columnas son la medida de respuesta o tasa de germinación y los niveles de tratamiento. Realice la prueba de Kruskal-Wallis mediante el comando `kruskal.test()`. El análisis consiste en determinar si la tasa de germinación varía según el tratamiento. Los argumentos del comando `kruskal.test()` son los nombres de las cabeceras de las columnas correspondientes de los datos tabla `datos.Ex08.5`

```
kruskal.test(y g, datos = mydataframe)
```

Detalles: y es la columna de medidas de respuesta (datos) y g es la columna correspondiente a los niveles de tratamiento (grupos) que se utilizaron. El argumento `data = mydataframe` es el marco de datos que contiene los datos (medidas con niveles de tratamiento correspondientes) que se van a analizar.

```

> kruskal.test(Rate ~ Treatment, data = data.Ex08.5)

##
##   Kruskal-Wallis rank sum test
##
## data:  Rate by Treatment
## Kruskal-Wallis chi-squared = 10.6775, df = 2, p-value = 0.004802

```

El pequeño valor P (inferior a 0,05) indica un resultado significativo. La cuestión ahora es saber qué tratamientos difieren significativamente. Es necesario realizar comparaciones por pares. Antes de llevar a cabo las comparaciones, veamos los rangos correspondientes a las tasas de germinación. Esto se consigue aplicando la función `rank()` a `data.Ex08.5$Rate`

Añada los rangos a los datos originales creando una nueva columna en la tabla `data.Ex08.5` utilizando `data.Ex08.5["Rank"]` y luego coloque los datos de rango en la columna.


```
> data.Ex08.5["Rank"] <- rank(data.Ex08.5$Rate)
> data.Ex08.5
```

```
##      Rate Treatment Rank
## 1      47   Fragment  4.0
## 2      50   Fragment  5.0
## 3      55   Fragment  7.0
## 4      68   Fragment  8.0
## 5       3    Ejecta  1.0
## 6      87    Ejecta 10.0
## 7      20    Ejecta  2.0
## 8      76    Ejecta  9.0
## 9      46    Ejecta  3.0
## 10     54    Ejecta  6.0
## 11     88     Splat 11.0
## 12     90     Splat 12.0
## 13     91     Splat 13.5
## 14     91     Splat 13.5
## 15     92     Splat 15.0
## 16     96     Splat 16.0
```

Calcule las medias de los rangos por Tratamiento utilizando el comando `tapply()`:

```
> tapply(data.Ex08.5$Rank, data.Ex08.5$Treatment, mean)

##      Ejecta Fragment      Splat
## 5.16667  6.00000 13.50000
```

Parece que el índice de germinación de Splats puede ser diferente de los otros dos índices.

Aunque R no proporciona una función para llevar a cabo las comparaciones pareadas descritas en el texto, proporcionamos `pairwise.kruskal.test()` para esta tarea. Utilice la función `source()` para descargarla

```
pairwise.kruskal.test(x, g)
```

Detalles: x son los datos de medición o respuesta y g es el vector de niveles de niveles de tratamiento (grupos). La prueba es de dos caras.

En este ejemplo, `data.Ex08.5$Rate` es la respuesta medida y `data.Ex08.5$Treatment` contiene los niveles de tratamiento.

```

> source("http://waveland.com/Glover-Mitchell/pairwise.kruskal.txt")
> pairwise.kruskal.test(data.Ex08.5$Rate, data.Ex08.5$Treatment)

## Dunn's Multiple Comparisons for Kruskal-Wallis test (adjusted for any ties)
##
## Data: data.Ex08.5$Rate and data.Ex08.5$Treatment
##
## p values (compare directly to alpha)
##
##           Ejecta Fragment
## Fragment 1.000000      NA
## Splat    0.007242 0.043786

```

El resultado de `pairwise.kruskal.test()` se presenta en forma de tabla. La entrada en la columna Ejecta y la fila Splat es 0,0072. Este es el valor P para la prueba de la diferencia en las medias de estos dos tratamientos, y es significativo al nivel $\alpha = 0,05$. Observaciones similares se aplican a las demás entradas. La tabla indica que la media de la tasa de germinación del grupo Splat difiere significativamente de las otras dos.

Ejemplos de Análisis de dos factores

Ejemplo 11

Una forma muy eficaz de capturar salmones comercialmente es mediante redes de enmalle. Tradicionalmente, los pescadores de redes de enmalle cuelgan largas redes de monofilamento en vertical sobre los caladeros. Cuando los salmones se encuentran con estas redes, se enredan en la malla por sus cubiertas branquiales y mueren rápidamente. Por desgracia, muchas otras especies también quedan atrapadas en estas redes. En particular, las aves marinas buceadoras, como los móridos comunes y las alcas rinocerontes, son presas de las redes de enmalle. Además, las focas y los delfines pueden quedar atrapados fácilmente en las redes de monofilamento, casi invisibles. Las aves y mamíferos marinos capturados y muertos en las redes de enmalle se denominan eufemísticamente "capturas accesorias" en la industria pesquera. Supongamos que, en el marco de un proyecto de conservación sobre la reducción de las capturas accesorias, se utilizan tres técnicas modificadas en los barcos que pescan salmón con redes de enmalle.

- Redes de enmalle con cubierta blanca (los 7 pies superiores de la red son de malla blanca que es considerablemente más fácil de ver en la columna de agua).
- Las redes de enmalle de cubierta blanca con pingers que emiten un molesto pitido similar al de los relojes con alarma. que emiten los relojes con alarma.
- Redes de enmalle no modificadas con pingers.
- Redes de enmalle no modificadas (método estándar).

En la siguiente tabla se indica la cantidad de salmones capturados por barco y día con cada una de estas técnicas.

Catch of salmon (100 kg/boat-day)				
Day	I	II	III	IV
1	23	21	24	25
2	16	14	19	17
3	19	17	23	21
4	11	12	15	16
5	41	30	40	42
6	32	20	37	35
7	9	11	19	17
8	10	14	19	16

¿Existen diferencias significativas en la cantidad de salmones capturados por barco por día utilizando estas técnicas?

Las hipótesis son:

$$H_0: \mu_I = \mu_{II} = \mu_{III} = \mu_{IV}$$

H_a : Al menos un para de μ_i es no igual

```
> data.Ex09.1 <- read.table("http://waveland.com/Glover-Mitchell/Example09-1.txt",
+ header = TRUE)
> tail(data.Ex09.1, n = 5)
```

##	Day	Technique	Catch
## 28	G	IV	17
## 29	H	I	10
## 30	H	II	14
## 31	H	III	19
## 32	H	IV	16

Los 32 valores de datos parecen estar ahí. Hay tres columnas: los bloques (Día), el tratamiento (Técnica) y la medida de respuesta (Captura). Realice el ANOVA con `aov()` y ponga el resultado en la tabla `aov.Ex09.1`. El análisis se centra en si la Captura varía en función de los factores Técnica de pesca y Día. Los argumentos de `aov()` son los nombres de las cabeceras de las columnas correspondientes de la tabla de datos `data.Ex09.1`. Muestre los resultados con el comando `summary()`. Observe cómo la ecuación del modelo en la función `aov()` expresa la Captura como una combinación de Técnica + Día.

```

> aov.Ex09.1 <- aov(Catch ~ Technique + Day, data = data.Ex09.1)
> summary(aov.Ex09.1)

##              Df Sum Sq Mean Sq F value    Pr(>F)
## Technique      3    259      86    10.2 0.00023 ***
## Day            7   2267     324    38.4 1.4e-10 ***
## Residuals     21    177       8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

La prueba F para la diferencia de medias de las Técnicas de pesca es significativa al nivel $\alpha = 0,001$. Para determinar qué medias son diferentes, realice comparaciones múltiples de las medias utilizando la función `TukeyHSD()`. Los datos proceden de los resultados del ANOVA `aov.Ex09.1`. Para separar las medias por Técnica, incluya "Técnica" como argumento de `TukeyHSD()`. Establecer el argumento `ordenado = TRUE` garantiza que en cada comparación la media más pequeña se reste de la más grande haciendo que la diferencia no sea negativa, lo que simplifica el análisis.

```

> TukeyHSD(aov.Ex09.1, "Technique", ordered = TRUE) # Note the quotes

##      Tukey multiple comparisons of means
##      95% family-wise confidence level
##      factor levels have been ordered
##
## Fit: aov(formula = Catch ~ Technique + Day, data = data.Ex09.1)
##
## $Technique
##           diff           lwr          upr      p adj
## I-II      2.750 -1.297872  6.79787 0.260763
## IV-II     6.250  2.202128 10.29787 0.001656
## III-II    7.125  3.077128 11.17287 0.000405
## IV-I      3.500 -0.547872  7.54787 0.105854
## III-I     4.375  0.327128  8.42287 0.031096
## III-IV    0.875 -3.172872  4.92287 0.930127

```

Tres de las diferencias tienen valores P inferiores a $\alpha = 0,05$. La conclusión es que la Técnica II es significativamente diferente tanto de la III como de la IV y que la Técnica I es significativamente diferente de la III. La tabla de resultados también proporciona un intervalo de confianza del 95% para la diferencia de cada par de medias. Como es habitual, sólo cuando el intervalo de confianza no contiene 0 la diferencia entre las medias es significativa. Podemos resumir estos resultados utilizando la notación de superíndice. Cualquier media con el mismo superíndice no es significativamente diferente.

Los resultados indican que cualquiera de los dos métodos para evitar las capturas accesorias (redes de enmalle de cabeza blanca sin emisores de ultrasonidos o emisores de ultrasonidos solos) no reduce significativamente las capturas de salmón. Esto indica que los pescadores podrían utilizar las redes de enmalle con tapa blanca o los emisores de ultrasonidos (pero no ambos) para reducir las "capturas accesorias" sin que disminuyan significativamente sus capturas de salmón.

Ejemplo 12

En un intento de encontrar los métodos más eficaces para el adiestramiento de perros de compañía para discapacitados físicos, se realizó un experimento para comparar tres regímenes de adiestramiento diferentes en combinación con tres sistemas de recompensa distintos. Todos los animales del estudio eran labradores y tenían entre 6 y 8 meses de edad al inicio del experimento. Los perros fueron asignados al azar a una combinación de régimen de adiestramiento y sistema de recompensa. Al final de un periodo de entrenamiento de 10 semanas, los perros fueron sometidos a una prueba estandarizada para medir su capacidad de funcionar como perros de compañía para los discapacitados visuales. Los resultados de la prueba se indican a continuación:

Reward	Training regime		
	I	II	III
Praise	45	51	52
	69	50	18
	53	62	25
	51	68	32
Tangible	54	53	51
	72	63	59
	69	67	47
	66	70	42
Praise and tangible	91	69	66
	87	73	68
	89	77	70
	91	74	64

El científico del comportamiento que dirige este estudio quiere saber qué régimen de entrenamiento, si lo hay, es el mejor y qué sistema de recompensa, si lo hay, es el mejor. Además, está buscando interacciones entre los regímenes de entrenamiento y los sistemas de recompensa. Estas interacciones pueden ser positivas (sinergia) o negativas (interferencia).

Lea el archivo de datos <http://waveland.com/Glover-Mitchell/Example09.2.txt> y enumere algunos de los datos para asegurarse de su exactitud.

```
> data.Ex09.2 <- read.table("http://waveland.com/Glover-Mitchell/Example09-2.txt",
+ header = TRUE)
> tail(data.Ex09.2)

##           Reward Training Score
## 31 PraiseAndTangible      II    77
## 32 PraiseAndTangible      II    74
## 33 PraiseAndTangible     III    66
## 34 PraiseAndTangible     III    68
## 35 PraiseAndTangible     III    70
## 36 PraiseAndTangible     III    64
```

Los 36 valores de los datos parecen estar presentes. Hay tres columnas: los dos factores diferentes (Recompensa y Entrenamiento) y la medida de respuesta (Puntuación). Realice el ANOVA con `aov()` y ponga el resultado en la tabla `aov.Ex09.2`. Observe la sintaxis: Para indicar que la puntuación depende posiblemente de una interacción entre el sistema de recompensa y el método de entrenamiento, utilizamos un asterisco `*` en lugar de `+` en el comando `aov()`: `Score ~ Reward * Training`.

```
> aov.Ex09.2 <- aov(Score ~ Reward * Training, data = data.Ex09.2)
> summary(aov.Ex09.2)

##           Df Sum Sq Mean Sq F value Pr(>F)
## Reward      2   4968    2484   38.14 1.4e-08 ***
## Training    2   2671    1335   20.50 3.8e-06 ***
## Reward:Training 4    583     146    2.24 0.091 .
## Residuals   27   1759      65
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Primero se investigan las interacciones. Las hipótesis son

$$H_0: (\alpha\beta)_{ij} = 0 \text{ para toda } i, j$$

$$H_a: (\alpha\beta)_{ij} \neq 0 \text{ para alguna } i, j$$

La estadística de la prueba es

$$F_{AxB} = F_{\text{Reward} \times \text{Training}}$$

Esto se encuentra en la tercera fila de la tabla ANOVA, denominada `Reward:Training`. $F_{AB} = 2,24$ con un valor P correspondiente de 0,091. Esto supera el 0,05, por lo que no tenemos pruebas de interacción aquí. Aceptamos H_0 .

A continuación, se comprueban los efectos principales de los regímenes de formación. Las hipótesis son

$$H_0: \mu_I = \mu_{II} = \mu_{III}$$

$$H_a: \text{al menos un par de } \mu \text{ no son iguales}$$

En la segunda fila de la tabla ANOVA, vemos que la estadística de la prueba es

$$F_B = F_{Training} = 20.50$$

pruebas de que al menos algunos regímenes de Formación son significativamente diferentes.

Por último, probamos los efectos principales para los sistemas de Reward. Las hipótesis son:

$$H_0: \mu_P = \mu_T = \mu_{P+T}$$

$$H_a: \text{al menos un par } \mu \text{ no son iguales}$$

En la primera fila de la tabla ANOVA, la estadística de la prueba es

$$F_A = F_{Reward} = 38.14$$

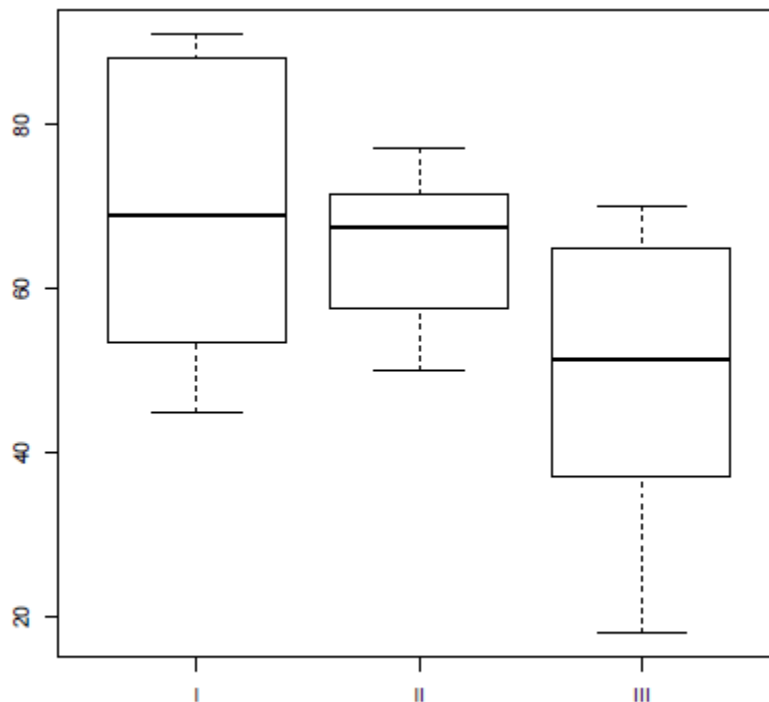
con un valor P correspondiente de casi 0. Tenemos pruebas de que al menos algunos sistemas de recompensa son significativamente diferentes.

Técnicas de separación de medias. Separe los efectos principales de los regímenes de entrenamiento mediante la prueba de Tukey. Enumere las puntuaciones medias por sistema de recompensa y haga un gráfico de caja de estos datos.

```
> tapply(data.Ex09.2$Score, data.Ex09.2$Training, mean)
```

```
##      I      II     III  
## 69.75 64.75 49.50
```

```
> boxplot(data.Ex09.2$Score~data.Ex09.2$Training)
```



El gráfico de caja indica que los resultados del método III pueden ser diferentes a los de los otros dos métodos. Confirme esto con la prueba de Tukey. Recuerde incluir el argumento "Formación", y por conveniencia establezca ordenado = TRUE.

```
> TukeyHSD(aov.Ex09.2, "Training", ordered = TRUE)

## Tukey multiple comparisons of means
## 95% family-wise confidence level
## factor levels have been ordered
##
## Fit: aov(formula = Score ~ Reward * Training, data = data.Ex09.2)
##
## $Training
##      diff      lwr      upr    p adj
## II-III 15.25  7.0811 23.4189 0.000237
## I-III  20.25 12.0811 28.4189 0.000004
## I-II   5.00 -3.1689 13.1689 0.298744
```

Utilizando los valores P, los regímenes de formación I y II no son significativamente diferentes, pero son superiores al régimen III.

III	II	I
49.5 ^a	64.8 ^b	69.8 ^b

Realizando un análisis similar para los sistemas de recompensa, determine los medios:

```
> tapply(data.Ex09.2$Score, data.Ex09.2$Reward, mean)

##      Praise PraiseAndTangible      Tangible
##      48.0000      76.5833      59.4167
```

```
> TukeyHSD(aov.Ex09.2, "Reward", ordered = TRUE)

## Tukey multiple comparisons of means
## 95% family-wise confidence level
## factor levels have been ordered
##
## Fit: aov(formula = Score ~ Reward * Training, data = data.Ex09.2)
##
## $Reward
##      diff      lwr      upr    p adj
## Tangible-Praise      11.4167  3.24777 19.5856 0.004919
## PraiseAndTangible-Praise 28.5833 20.41444 36.7522 0.000000
## PraiseAndTangible-Tangible 17.1667  8.99777 25.3356 0.000050
```

La prueba de Tukey indica que los tres sistemas de recompensa son significativamente diferentes entre sí, ya que todos los valores P son muy inferiores a 0,05. El gran estadístico F (FA = 38,14) fue un primer indicio de que las diferencias aquí eran drásticas. Esto se resume como

P	T	P+T
48.0 ^a	59.4 ^b	76.6 ^c

Por último, concluimos que, aunque el elogio más las recompensas tangibles es lo que más aumenta las puntuaciones de las pruebas, el régimen de entrenamiento I o II es mejor que el III y no se indica ninguna interacción entre el régimen de entrenamiento y el sistema de recompensas.

Ejemplo 13

Como parte de un estudio mucho más amplio sobre la salud neonatal en Estados Unidos, se registraron los pesos al nacer de 36 recién nacidos a término en kilogramos. Los principales factores de interés fueron la edad materna y la condición de fumadora de la madre. Analice los datos siguientes de forma adecuada, asumiendo la normalidad de los pesos al nacer.

Edad de la madre	Condición de fumadora de la madre		
	No fumadora	Fumadora moderada	Fumadora frecuente
Menores < 25 años	3.63	3.35	3.23
	3.45	3.30	3.05
	3.40	3.31	3.18
	3.51	3.20	3.40
	3.60	3.25	3.19
	3.55	3.43	3.15
	$\bar{X}_{11} = 3.523$	$\bar{X}_{21} = 3.307$	$\bar{X}_{31} = 3.200$
Mayores > 25 años	3.61	3.08	3.05
	3.49	3.10	3.00
	3.58	3.37	3.18
	3.67	3.28	3.20
	3.40	3.19	2.99
	3.18	3.31	3.25
	$\bar{X}_{12} = 3.488$	$\bar{X}_{22} = 3.222$	$\bar{X}_{32} = 3.112$
$\sum_i \sum_j \sum_k X_{ijk}^2 = 395.336$		$\sum_i \sum_j \sum_k X_{ijk} = T \dots = 119.110$	

Leer el archivo de datos <http://waveland.com/Glover-Mitchell/Problem09.16.txt> y enumerar algunos de los datos para garantizar la exactitud

```
> data.Prob09.16 <- read.table("http://waveland.com/Glover-Mitchell/Problem09-16.txt",
+ header = TRUE)
> tail(data.Prob09.16, n = 3)

##      Age SmokingStatus Weight
## 34 Older           Heavy   3.20
## 35 Older           Heavy   2.99
## 36 Older           Heavy   3.25
```

Hay tres columnas: los dos factores diferentes (Edad y Condición de fumador) y la medida de respuesta (Peso). Realice el ANOVA con `aov()` y ponga el resultado en la tabla `aov.Prob09.16`. Observe la sintaxis: Para indicar que el peso depende posiblemente de una interacción Edad y Estado de fumador.

```
> aov.Prob09.16 <- aov(Weight ~ Age * SmokingStatus, data = data.Prob09.16)
> summary(aov.Prob09.16)

##              Df Sum Sq Mean Sq F value    Pr(>F)
## Age              1   0.043    0.043     3.04   0.091 .
## SmokingStatus    2   0.771    0.385    27.02 1.9e-07 ***
## Age:SmokingStatus 2   0.005    0.003     0.19   0.830
## Residuals       30   0.428    0.014
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Prueba de interacción. Prueba $H_0: (\alpha\beta)_{ij} = 0$ para todo i, j $H_a: (\alpha\beta)_{ij} \neq 0$ para alguno i, j la prueba es:

$$F_{AxB} = F_{agexSmokingStatus}$$

Esto se encuentra en la tercera fila de la tabla ANOVA etiquetada como Age:SmokingStatus. $F_{AxB} = 0,19$ con un valor P correspondiente de 0,830. Esto supera el 0,05, por lo que no tenemos pruebas de interacción.

Comprobar el estado del fumador $H_0: \mu_{NS} = \mu_{MS} = \mu_{HS}$, $H_a: \text{Al menos un par de } \mu \text{ es diferente}$ Desde $F_{SmokingStatus} = 27.02$ tiene un valor P ínfimo, rechaza H_0 . Las medias de los pesos al nacer son significativamente diferentes para al menos dos de los niveles de la conducta de fumar

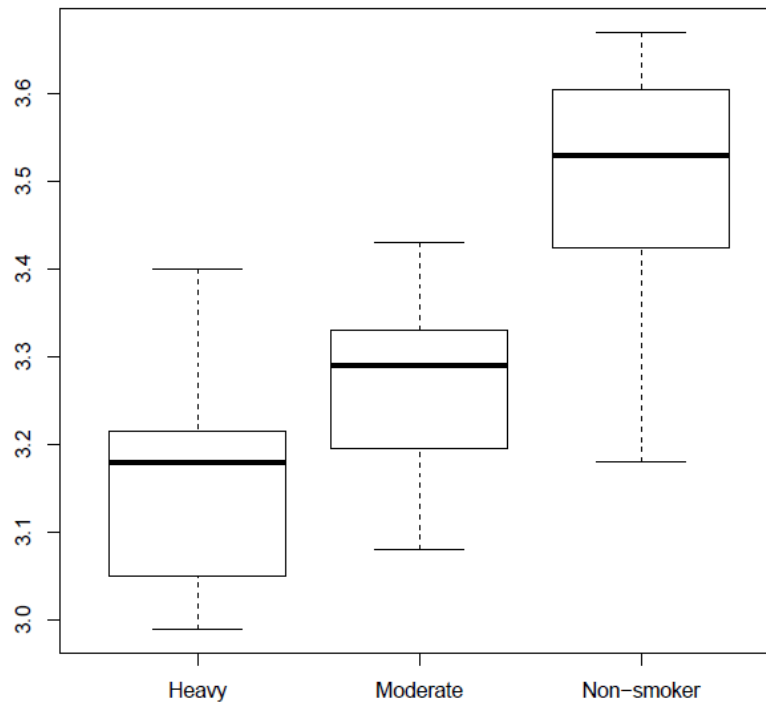
Comprueba la edad de las madres $H_0: \mu_Y = \mu_O$, $H_a: \mu_Y \neq \mu_O$. $F_{Age} = 3.04$ tiene un valor $P > 0,05$, se acepta H_0 . El peso medio al nacer no es significativamente diferente para las madres más jóvenes y las mayores madres más jóvenes y mayores. La separación de medias no es necesaria.

Separe los efectos principales de SmokingStatus mediante la prueba de Tukey. Enumere las puntuaciones medias según el estado de fumador y haga un gráfico de caja de estos datos.

```
> tapply(data.Prob09.16$Weight, data.Prob09.16$SmokingStatus, mean)

##      Heavy  Moderate Non-smoker
##      3.15583   3.26417   3.50583
```

```
> boxplot(data.Prob09.16$Weight ~ data.Prob09.16$SmokingStatus)
```



El gráfico de cajas indica que los no fumadores pueden ser diferentes a los fumadores. Confirme esto con la prueba de Tukey. Recuerde incluir el argumento "SmokingStatus", y por conveniencia establezca `ordered = TRUE`.

```
> TukeyHSD(aov.Prob09.16, "SmokingStatus", ordered = TRUE)

## Tukey multiple comparisons of means
## 95% family-wise confidence level
## factor levels have been ordered
##
## Fit: aov(formula = Weight ~ Age * SmokingStatus, data = data.Prob09.16)
##
## $SmokingStatus
##           diff      lwr      upr    p adj
## Moderate-Heavy  0.108333 -0.0118396 0.228506 0.083596
## Non-smoker-Heavy 0.350000  0.2298271 0.470173 0.000000
## Non-smoker-Moderate 0.241667  0.1214938 0.361840 0.000076
```

Los resultados muestran que los no fumadores difieren de los fumadores moderados o de los fumadores empedernidos, mientras que la diferencia entre los fumadores moderados y los empedernidos no llega a ser significativa (el intervalo de confianza apenas contiene 0). Esta es una respuesta ligeramente diferente a la que se da en la parte posterior del texto. Las pruebas de Tukey son más conservadoras que las de Bonferroni-Holm.

Referencias

[1] K. Mitchell y T. Glover, *An Introduction to biostatistics using R*.