Contraste de aleatoriedad. Test de Rachas

Contraste de aleatoriedad. Test de Rachas

El procedimiento Prueba de Rachas contrasta la aleatoriedad de un conjunto de observaciones de una variable continua. Para ello, el test de rachas cuenta las cadenas de valores consecutivos que presenta la variable por encima y por debajo de un determinado punto de corte. Cada uno de estas cadenas recibe el nombre de racha (de ahí el nombre del contraste). Un número muy elevado o muy reducido de rachas apuntarán hacia la no aleatoriedad de los datos que componen la muestra.

Una racha es una secuencia de observaciones similares, una sucesión de símbolos idénticos consecutivos. Ejemplo: $++---+--++++---$ (6 rachas). Una muestra con un número excesivamente grande o excesivamente pequeño de rachas sugiere que la muestra no es aleatoria

```r
library(Rcmdr)
```

```
Warning: package 'Rcmdr' was built under R version 4.1.3
```

```
Loading required package: splines
```

```
Loading required package: RcmdrMisc
```

```
Warning: package 'RcmdrMisc' was built under R version 4.1.3
```

```
Loading required package: car
```

```
Warning: package 'car' was built under R version 4.1.3
```

```
Loading required package: carData
```

```
Warning: package 'carData' was built under R version 4.1.3

Loading required package: sandwich

Warning: package 'sandwich' was built under R version 4.1.3

Loading required package: effects

Warning: package 'effects' was built under R version 4.1.3

lattice theme set by effectsTheme()
See ?effectsTheme for details.

La interfaz R-Commander sólo funciona en sesiones interactivas



Attaching package: 'Rcmdr'

The following object is masked from 'package:base':

    errorCondition
```

Supuesto Práctico 13

Se realiza un estudio sobre el tiempo en horas de un tipo determinado de escáner antes de la primera avería. Se ha observado una muestra de 10 escáner y se ha anotado el tiempo de funcionamiento en horas: 18.21; 2.36; 17.3; 16.6; 4.70; 3.63; 15.56; 7.35; 9.78; 14.69. A un nivel de significación del 5%, ¿se puede considerar aleatoriedad en la muestra?

Solución Formulamos el contraste que debemos resolver.

$H_0 \equiv$ Los datos de la muestra son aleatorios

$H_1 \equiv$ Los datos de la muestra no son aleatorios

Lo primero que vamos a hacer es crear un fichero de texto con los datos del problema con la siguiente estructura:

```
> library(randtests, pos=18)
```

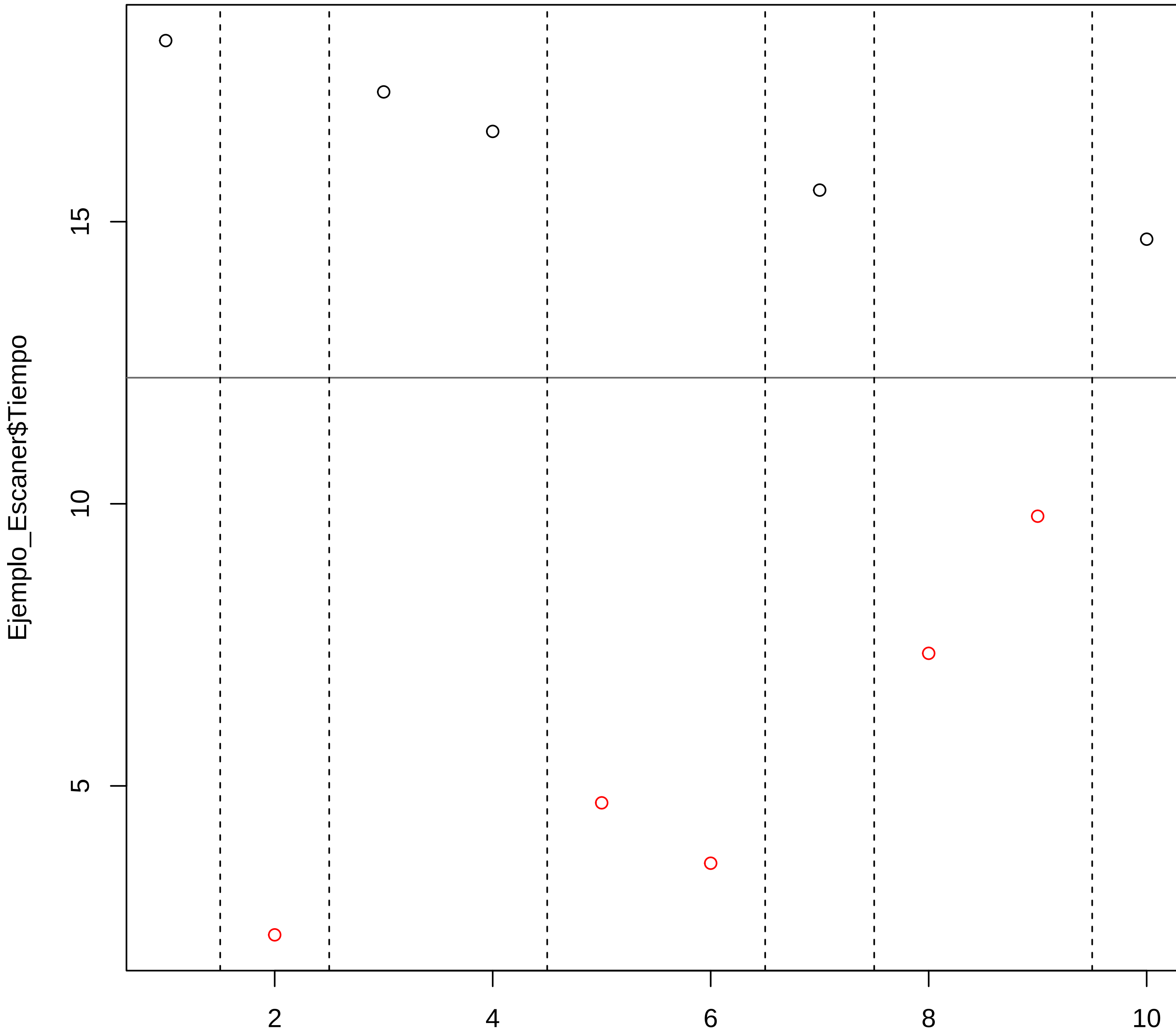
```
Warning: package 'randtests' was built under R version 4.1.3
```

A continuación, instalamos y cargamos el paquete randtests. Una vez hecho esto, cargamos el fichero de datos creado: Datos/Importar datos/desde archivo de texto, portapapeles o URL

```
> Ejemplo_Escaner <-
+   read.table("C:/Users/user/OneDrive/Paquete R/PRACTICAS-S10/supuesto13.txt",
+    header=TRUE, stringsAsFactors=TRUE, sep="", na.strings="NA", dec=".",
+   strip.white=TRUE)
```

A continuación, escribimos en R Script la siguiente orden para llamar a la función runs.test.

```
> runs.test(Ejemplo_Escaner$Tiempo,alternative="two.sided",threshold=median(Ejemplo_Escane
```


Ejemplo_Escaner$Tiempo
5
10
15
2
4
6
8
10

```
    Runs Test

data:  Ejemplo_Escaner$Tiempo
statistic = 0.67082, runs = 7, n1 = 5, n2 = 5, n = 10, p-value = 0.5023
alternative hypothesis: nonrandomness
```

Según los resultados del test de rachas, se han encontrado 7 rachas (runs), que vienen separadas por líneas discontinuas verticales. Hay 5 valores por encima de la mediana (n_1), marcados en negro, y otros 5 valores por debajo de la mediana (n_2), marcados en rojo.

El p-valor asociado al contraste es 0.5023 superior a 0.05, por lo que no es posible rechazar la hipótesis nula. Por tanto, podemos concluir que los datos de la muestra son aleatorios.

Contraste sobre bondad de ajuste: Procedimiento Prueba de Kolmogorov-Smirnov

Mediante el contraste de bondad de ajuste de Kolmogorv-Smirnov se prueba si los datos de una muestra proceden, o no, de una determinada distribución de probabilidad. Lo que se hace es comparar la función de distribución acumulada que se calcula a partir de los datos de la muestra con la función de distribución acumulada teórica de la distribución con la que se compara.

El contraste de hipótesis que se plantea es el siguiente:

$H_0 \equiv$ Los datos de la muestra proceden de la distribución de probabilidad

$H_1 \equiv$ Los datos de la muestra no proceden de la distribución de probabilidad

Supuesto Práctico 14

Las puntuaciones de 10 individuos en una prueba de una oposición han sido las siguientes: 41.81, 40.30, 40.20, 37.14, 39.29, 38.79, 40.73, 39.26, 35.74, 41.65. ¿Puede suponerse, a un nivel de significación del 5% que dichas puntuaciones se ajustan a una distribución normal de media 40 y desviación típica 3?

Solución El contraste de hipótesis que se plantea es el siguiente:

$H_0 \equiv$ Los datos de la muestra proceden de una distribución N(40,3)

$H_1 \equiv$ Los datos de la muestra no proceden de de una distribución N(40,3)

Lo primero que vamos a hacer es crear un fichero de texto con los datos del problema con la siguiente estructura:

La variable a estudiar debe aparecer en la primera fila entre comillas, y a continuación se introducen los valores numéricos que nos da el enunciado en columna y sin entrecomillar, ya que estamos trabajando con una variable cuantitativa.

```
> Ejemplo_Oposicion <-
+   read.table("C:/Users/user/OneDrive/Paquete R/PRACTICAS-S10/supuesto14.txt",
+   header=TRUE, stringsAsFactors=TRUE, sep="", na.strings="NA", dec=".",
+   strip.white=TRUE)
```

A continuación, se resuelve el contraste mediante una llamada a la función ks.test. Debemos tener en cuenta que la distribución de comparación es la distribución normal (por tanto, el argumento y tomará el valor pnorm) de media igual a 40 y desviación típica igual a 3.

```
> ks.test(Ejemplo_Oposicion$Puntuaciones,y=pnorm,40,3,alternative="two.sided")
```

```
    One-sample Kolmogorov-Smirnov test

data:  Ejemplo_Oposicion$Puntuaciones
D = 0.27314, p-value = 0.3752
alternative hypothesis: two-sided
```

En este caso, el valor del estadístico de contraste es 0.27314 y el p-valor asociado al contraste es 0.3752. Como el p-valor es superior a 0.05 no podemos rechazar la hipótesis nula, por lo que concluimos que los datos de la muestra proceden de una distribución normal de media 40 y de desviación típica 3.

Pruebas para dos muestras independientes

El procedimiento Pruebas para dos muestras independientes compara dos grupos de casos existentes en una variable y comprueba si provienen de la misma población (homogeneidad). Estos contrastes, son la alternativa no paramétrica de los tests basados en el t de Student, Al igual que con el test de Student, se tienen dos grupos de observaciones independientes y se compara si proceden de la misma población.

Supuesto Práctico 15

En unos grandes almacenes se realiza un estudio sobre el rendimiento de ventas de los vendedores. Para ello, se observa durante 10 días el número de ventas de dos vendedores:

Vendedor A: 10 40 60 15 70 90 30 32 22 13 Vendedor B: 45 60 35 30 30 15 50 20 32 9

Contrastar, considerando un nivel de significación del 5%, si los rendimientos medianos de ambos vendedores pueden asumirse iguales.

Solución Comenzamos creando el archivo de datos de ventas de los dos vendedores:

Para resolver este contraste debemos tener en cuenta que los datos proceden de muestras independientes, que el valor de la diferencia entre las medianas que se pretende comprobar es 0 y que la hipótesis alternativa del contraste es del tipo "distinto de".

En primer lugar, cargamos el fichero seleccionando: Datos/Importar datos/desde archivo de texto, portapapeles o URL…

Se muestra una ventana en la que introducimos el nombre que queremos asignarle al conjunto de datos con el que vamos a trabajar; en nuestro caso, escribiremos Ejemplo_Ventas. El resto de opciones las dejamos por defecto, ya que el archivo de texto que hemos creado cumple con todas ellas y pulsamos Aceptar

Para resolver el contraste planteado, seleccionamos en el menú: Estadísticos/Test no paramétricos/Test de Wilcoxon para dos muestras

```
> Ejemplo_Ventas <-
+   read.table("C:/Users/user/OneDrive/Paquete R/PRACTICAS-S10/supuesto15.txt",
+   header=TRUE, stringsAsFactors=TRUE, sep="", na.strings="NA", dec=".",
+   strip.white=TRUE)
```

En Datos tenemos que realizar dos selecciones: por un lado, elegir la variable "Vendedor" en la ventana de Grupos para que las comparaciones las haga entre los dos vendedores, y en la ventana Variable explicada seleccionamos la variable "Frecuencia" dónde aparecían las ventas de cada vendedor por cada uno de los 10 días.

En la pestaña Opciones, las opciones que vienen seleccionadas por defecto son las que necesitamos para resolver nuestro problema, excepto el Tipo de prueba. Si dejamos la opción "Por defecto" nos aplica el corrector por continuidad que, en nuestro caso, no vamos a aplicar. Por lo tanto, seleccionamos la opción "Exacto" y pulsamos Aceptar.

Test de Wilcoxon: Frecuencia ~ Vendedor

```
> Tapply(Frecuencia ~ Vendedor, median, na.action=na.omit,
+   data=Ejemplo_Ventas) # medians by group
```

```
 A  B 
31 31 
```

```
> wilcox.test(Frecuencia ~ Vendedor, alternative='two.sided', exact=TRUE,
+   correct=FALSE, data=Ejemplo_Ventas)
```

```
Warning in wilcox.test.default(x = c(10L, 40L, 60L, 15L, 70L, 90L, 30L, : cannot
compute exact p-value with ties


    Wilcoxon rank sum test

data:  Frecuencia by Vendedor
W = 52.5, p-value = 0.8497
alternative hypothesis: true location shift is not equal to 0
```

En este caso, el p-valor asociado al contraste es, aproximadamente, 0.8497. Como este p-valor es mayor que 0.05 no se puede rechazar la hipótesis nula, considerando un nivel de significación del 5%. Por tanto, concluimos que las medianas de las ventas de ambos vendedores pueden asumirse iguales.

Pruebas para dos muestras relacionadas

Esta prueba es similar a la anterior, con la salvedad de que ahora se supone que los datos de las muestras están relacionados, es decir, no son independientes.

Supuesto Práctico 16

En un encinar de Navarra se pretende comprobar si un tratamiento ayuda a disminuir el nivel de húmedas de las hojas de las encinas. Para ello, se realiza un estudio a 10 encinas, en las que se seleccionan aleatoriamente 10 hojas y se registra el nivel de humedad de las hojas antes y después del tratamiento.

Suponiendo un nivel de significación del 5%, ¿Puede suponerse efectivo el tratamiento?

Solución Comenzamos creando el archivo de datos del nivel de humedad de las hojas de las encinas:

En primer lugar, cargamos el fichero seleccionando: Datos/Importar datos/desde archivo de texto, portapapeles o URL…

En la ventana resultante introducimos el nombre que queremos asignarle al conjunto de datos con el que vamos a trabajar; en nuestro caso, escribiremos Ejemplo_Encinas. El resto de opciones las dejamos por defecto, ya que el archivo de texto que hemos creado cumple con todas ellas.

Pulsamos Aceptar y se abre una ventana para que seleccionemos el archivo de texto que hemos creado y guardado anteriormente en nuestro ordenador. Cuando abrimos el archivo, podemos ver que en Conjunto de datos aparece el nombre que le hemos asignado a nuestro conjunto de datos.

Para resolver el contraste planteado, seleccionamos en el menú: Estadísticos/Test no paramétricos/Test de Wilcoxon para muestras pareadas…

```
> Ejemplo_Encinas <-
+   read.table("C:/Users/user/OneDrive/Paquete R/PRACTICAS-S10/supuesto16.txt",
+   header=TRUE, stringsAsFactors=TRUE, sep="", na.strings="NA", dec=".",
+   strip.white=TRUE)
```

En Datos tenemos que realizar dos selecciones: por un lado, elegir la variable que nos indica la humedad antes del tratamiento "Antes" en la ventana Primera variable, y la que nos indica la humedad después del tratamiento "Después" en la ventana Segunda variable.

En la pestaña Opciones, necesitamos seleccionar la Hipótesis alternativa mayor que (>), en-Tipo de prueba seleccionamos la opción "Exacto" para que no nos aplique el corrector por continuidad y pulsamos Aceptar.

Test de Wilcoxon para datos pareados Antes, Despues

```
> with(Ejemplo_Encinas, median(Antes - Despues, na.rm=TRUE))
```

```
[1] 3.05
```

```
>   # median difference
> with(Ejemplo_Encinas, wilcox.test(Antes, Despues, alternative='greater',
+   exact=TRUE, paired=TRUE))
```

```

    Wilcoxon signed rank exact test

data:  Antes and Despues
V = 49, p-value = 0.01367
alternative hypothesis: true location shift is greater than 0
```

En este ejemplo, el p-valor asociado al contraste es 0.01367, inferior a 0.05, por lo que se rechaza la hipótesis nula considerando un nivel de significación del 5%. Esto quiere decir que el tratamiento utilizado es efectivo para reducir el nivel de humedad de las hojas de las encinas.

Ejercicios

Ejercicios Guiados

Ejercicio Guiado1

En la base de datos universidad.txt tenemos algunas variables de interés medidas para dos grupos de alumnos, dependiendo del turno de clase en el que se encuentren. El turno de mañana se define como A y el de tarde como B. Los datos se muestran en la siguiente tabla:

Se pide:

a) Contrastar al 99% si el cociente intelectual medio es 100, sabiendo que la varianza poblacional es igual a 3

En este caso nos encontramos ante un contraste de hipótesis sobre la media de una población normal con varianza conocida, por lo que hay que calcularlo mediante código

Introducimos en R los datos relativos al nivel de significación y la varianza poblacional de la variable que proporciona el enunciado. Y calculamos el valor del estadístico y comprobamos si se cumple o no la condición de rechazo.

```
> universidad <-
+   read.table("C:/Users/user/OneDrive/Paquete R/PRACTICAS-S10/universidad.txt",
+    header=TRUE, stringsAsFactors=TRUE, sep="", na.strings="NA", dec=".",
+   strip.white=TRUE)
```

```
> alpha<-0.01
> varianza<-3
> mu0<-100
> n<-nrow(universidad)
> media<-mean(universidad$C.I.)
> z0<-(media- mu0)/(sqrt(varianza)/sqrt(n))
> z0
```

```
[1] 1.549193
```

Seleccionamos todas las sentencias y pulsamos Ejecutar

```
> cuantil<-qnorm(1-alpha/2)
> cuantil
```

```
[1] 2.575829
```

Como $\mid Z \mid -\alpha < Z_{1-\alpha/2}$ no tenemos evidencia muestral para rechazar la hipótesis nula, por lo que podemos considerar que el coeficiente intelectual es 100

b) Contrastar al 95% de confianza si el cociente intelectual medio puede considerarse 120

En este caso nos encontramos ante un contraste de hipótesis sobre la media de una población normal con varianza desconocida

Test t para una muestra: C.I.

```
> with(universidad, (t.test(C.I., alternative='two.sided', mu=120,
+   conf.level=.95)))
```

```

	One Sample t-test

data:  C.I.
t = -14.68, df = 4, p-value = 0.0001253
alternative hypothesis: true mean is not equal to 120
95 percent confidence interval:
  97.64442 104.75558
sample estimates:
mean of x
    101.2
```

La resolución del contraste se hará basándonos en el p-valor (p-value = 0.0001253)

El p-valor es una probabilidad (oscila, por lo tanto, entre 0 y 1).

- Si el p-valor es mayor que el nivel de significación, no rechazamos la hipótesis nula.
- Si el p-valor es menor que el nivel de significación, rechazamos la hipótesis nula en favor de la hipótesis alternativa.

c) Realizar un contraste de hipótesis a un nivel de confianza del 98% para la diferencia de medias del cociente intelectual entre el grupo A y B. ¿Puede suponerse que el cociente intelectual medio entre ambos grupos es igual?

En este caso nos encontramos ante contraste de hipótesis para la diferencia de medias en dos poblaciones normales independientes.

Test F de dos varianzas: C.I. Grupo

```
> Tapply(C.I. ~ Grupo, var, na.action=na.omit, data=universidad)
```

```
       A        B
12.33333  8.00000
```

```
>   # variances by group
> var.test(C.I. ~ Grupo, alternative='two.sided', conf.level=.98,
+   data=universidad)
```

```

    F test to compare two variances

data:  C.I. by Grupo
F = 1.5417, num df = 2, denom df = 1, p-value = 0.9897
alternative hypothesis: true ratio of variances is not equal to 1
98 percent confidence interval:
 3.083642e-04 1.518580e+02
sample estimates:
ratio of variances
          1.541667
```

Fijándonos en el p-valor vemos que $0.9897 \geq 0.02$, por lo que no tenemos suficiente evidencia muestral para rechazar , es decir, podemos decir que el cociente intelectual medio del grupo A y B son iguales.

d) Contrasta al 90% de confianza si la proporción de alumnos del grupo A es 0.5.

Test de proporciones para una muestra: Grupo

```
> local({
+   .Table <- xtabs(~ Grupo , data= universidad )
+   cat("\nFrequency counts (test is for first level):\n")
+   print(.Table)
```

```
+   prop.test(rbind(.Table), alternative='two.sided', p=.5, conf.level=.90,
+   correct=FALSE)
+ })
```

```
Frequency counts (test is for first level):
Grupo
A B
3 2

Warning in prop.test(rbind(.Table), alternative = "two.sided", p = 0.5, : Chi-
squared approximation may be incorrect


    1-sample proportions test without continuity correction

data:  rbind(.Table), null probability 0.5
X-squared = 0.2, df = 1, p-value = 0.6547
alternative hypothesis: true p is not equal to 0.5
90 percent confidence interval:
 0.2724832 0.8572935
sample estimates:
  p
0.6
```

Por lo que si nos fijamos en el p-valor vemos que es $0.6547 \geq 0.1$, por lo que no tenemos suficiente evidencia muestral para rechazar la hipótesis nula, es decir, la proporción de alumnos del grupo A puede considerarse igual a 0.5

e) Contrasta al 90% de confianza si la proporción de alumnos del grupo B es 0.5.

Dado que la hipótesis que se ha planteado se ha hecho sobre el grupo B es necesario hacer una recodificación de la variable

```
> universidad <- within(universidad, {
+   Grup_rec <- Recode(Grupo, '"A"="2A"; "B"="1B"', as.factor=TRUE,
+   to.value="=", interval=":", separator=";")
+ })
```

Una vez recodificada la variable, pasamos a calcular el contraste de hipótesis planteado

Test de proporciones para una muestra: Grup_rec

```
> local({
+   .Table <- xtabs(~ Grup_rec , data= universidad )
+   cat("\nFrequency counts (test is for first level):\n")
+   print(.Table)
+   prop.test(rbind(.Table), alternative='two.sided', p=.5, conf.level=.90,
+   correct=FALSE)
+ })
```

```
Frequency counts (test is for first level):
Grup_rec
1B 2A
 2  3
```

```
Warning in prop.test(rbind(.Table), alternative = "two.sided", p = 0.5, : Chi-
squared approximation may be incorrect
```

```
    1-sample proportions test without continuity correction

data:  rbind(.Table), null probability 0.5
X-squared = 0.2, df = 1, p-value = 0.6547
alternative hypothesis: true p is not equal to 0.5
90 percent confidence interval:
 0.1427065 0.7275168
sample estimates:
  p
0.4
```

Observando el p-valor vemos que es $0.6547 \geq 0.1$, por lo que no tenemos suficiente evidencia muestral para rechazar la hipótesis nula, es decir, la proporción de alumnos del grupo B puede considerarse igual a 0.5

f) Obtener un contraste de hipótesis a un nivel de confianza del 93% para la diferencia entre la proporción de alumnos en el grupo A y B que tienen clase de estadística.

En este caso nos encontramos ante un contraste de hipótesis para la diferencia de dos proporciones

```
> library(abind, pos=19)
```

Test de proporciones para dos muestras: Grupo, Estadistica

```
> local({  .Table <- xtabs(~Grupo+Estadistica, data=universidad)
+   cat("\nPercentage table:\n")
+   print(rowPercents(.Table))
+   prop.test(.Table, alternative='two.sided', conf.level=.93, correct=FALSE)
+ })
```

```
Percentage table:
     Estadistica
Grupo   No   Si Total Count
    A 33.3 66.7   100     3
    B 50.0 50.0   100     2
```

```
Warning in prop.test(.Table, alternative = "two.sided", conf.level = 0.93, :
Chi-squared approximation may be incorrect
```

```
    2-sample test for equality of proportions without continuity
    correction

data:  .Table
X-squared = 0.13889, df = 1, p-value = 0.7094
alternative hypothesis: two.sided
93 percent confidence interval:
 -0.9750999  0.6417665
sample estimates:
   prop 1    prop 2
0.3333333 0.5000000
```

Por último el programa devuelve la salida, y concretamente nos fijamos en el valor del $p-valor = 0.7094 \geq 0.07$, por lo que no rechazamos la hipótesis nula y concluimos que las proporciones de alumnos en ambos grupos que tienen clase de estadística coinciden

Ejercicio Guiado 2

En un hospital se elige una muestra de pacientes y se les mide la tasa cardíaca por la mañana y a última hora de la tarde. Estudiar mediante un contraste de hipótesis al 99% si, por término medio, la tasa cardíaca es igual por la mañana y a última hora de la tarde.

Solución: En primer lugar para trabajar con R_Commander escribimos la siguiente sentencia en R

```
> tcardiaca <-
+   read.table("C:/Users/user/OneDrive/Paquete R/PRACTICAS-S10/tcardiaca.txt",
+   header=TRUE, stringsAsFactors=TRUE, sep="", na.strings="NA", dec=".",
+   strip.white=TRUE)
```

Test t para datos emparejados TCM, TCT

```
> with(tcardiaca, (t.test(TCM, TCT, alternative='two.sided', conf.level=.99,
+   paired=TRUE)))
```

```

	Paired t-test

data:  TCM and TCT
t = -1.6236, df = 4, p-value = 0.1798
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 -18.411312   8.811312
sample estimates:
mean of the differences 
                   -4.8 
```

Observando el $p-valor = 0.1798 \geq 0.01$ podemos concluir que no tenemos suficiente evidencia muestral para rechazar la hipótesis nula, por lo que puede asumirse que la diferencia entre dichas medias es 0, o dicho de otro modo, que ambas tasas cardíacas son iguales.

Ejercicio Guiado 3

El ayuntamiento quiere comprobar si la ocupación del metro se produce en la misma proporción durante todos los días de la semana. Para ello, se registra el número de usuarios durante una semana cualquiera.

Contrastar, a un nivel de significación del 5%, si la hipótesis de la directora del hospital puede suponerse cierta. ¿Puede asumirse que las proporciones usuarios que utilizan el metro de lunes a domingo son (0.17, 0.17, 0.17, 0.17, 0.17, 0.10, 0.05)?

Solución: En primer lugar para trabajar con R_Commander escribimos la siguiente sentencia en R

```
> #library(Rcmdr)
```

A continuación, cargamos el fichero seleccionando: Datos/Importar datos/desde archivo de texto, portapapeles o URL

```
> Ejemplo_Metro <-
+   read.table("C:/Users/user/OneDrive/Paquete R/PRACTICAS-S10/ejemplo_metro.txt",
+    header=TRUE, stringsAsFactors=TRUE, sep="", na.strings="NA", dec=".",
+   strip.white=TRUE)
```

Para transformar la tabla de frecuencias en un conjunto de datos (data.frame) con el que R-Commander pueda trabajar hay que escribir las siguientes instrucciones en la ventana R Script, seleccionar ambas a la vez y darle a Ejecutar:

```
> P<-rep(Ejemplo_Metro$Dias,Ejemplo_Metro$Frecuencia)
> Afluencia_diaria<-data.frame(P)
```

Con esto conseguimos que la variable P contenga los datos en forma de lista y que el nuevo archivo de datos con el que vamos a trabajar se llame Afluencia_diaria.

Para visualizar el nuevo conjunto de datos, pulsamos en el botón Conjunto de datos y seleccionamos Afluencia_diaria. Una vez seleccionado, si pulsamos en el botón Visualizar conjunto de datos podemos comprobar el nuevo formato de los datos.

Frecuencias: P

```
> local({
+   .Table <- with(Afluencia_diaria, table(P))
+   cat("\ncounts:\n")
+   print(.Table)
+   cat("\npercentages:\n")
+   print(round(100*.Table/sum(.Table), 2))
+   .Probs <- c(0.142857142857143,0.142857142857143,0.142857142857143,
+   0.142857142857143,0.142857142857143,0.142857142857143,0.142857142857143)
+   chisq.test(.Table, p=.Probs)
+ })
```

```
counts:
P
  Domingo     Jueves      Lunes     Martes Miercoles     Sabado    Viernes
       10         33         32         29        35         26         29
```

```
percentages:
P
  Domingo     Jueves      Lunes     Martes Miercoles     Sabado    Viernes
     5.15      17.01      16.49      14.95      18.04      13.40      14.95


    Chi-squared test for given probabilities

data:  .Table
X-squared = 15.134, df = 6, p-value = 0.01924
```

l estadístico de contraste, que sigue una distribución Chi-cuadrado, toma el valor 15.134. Los grados de libertad de la distribución Chi-cuadrado para este ejemplo son 6. El p-valor asociado al contraste es menor que 0.05 por lo que, considerando un nivel de significación del 5%, se rechaza la hipótesis nula. Es decir, se concluye que la ocupación del metro no se produce en la misma proporción todos los días de la semana.

Para comprobar si podemos asumir que las proporciones de ocupación correspondientes a cada día de la semana (de Lunes a Domingo) son (0.17, 0.17, 0.17, 0.17, 0.17, 0.10, 0.05), seguimos los mismos pasos, pero teniendo en cuenta que, ahora, tenemos que introducir los valores de las nuevas proporciones consideradas.

```
> local({
+   .Table <- with(Afluencia_diaria, table(P))
+   cat("\ncounts:\n")
+   print(.Table)
+   cat("\npercentages:\n")
+   print(round(100*.Table/sum(.Table), 2))
+   .Probs <- c(0.0513950073421439,0.174743024963289,0.174743024963289,
+   0.146842878120411,0.174743024963289,0.102790014684288,0.174743024963289)
+   chisq.test(.Table, p=.Probs)
+ })
```

```
counts:
P
  Domingo     Jueves      Lunes     Martes Miercoles     Sabado    Viernes
       10         33         32         29         35         26         29

percentages:
P
  Domingo     Jueves      Lunes     Martes Miercoles     Sabado    Viernes
     5.15      17.01      16.49      14.95      18.04      13.40      14.95
```

	Levocular	Ambiocular	Dextrocular	
Zurdo	34	62	28	124
Ambidextro	27	28	20	75
Dextro	57	105	52	214
	118	195	100	413

```
    Chi-squared test for given probabilities

data:  .Table
X-squared = 2.7245, df = 6, p-value = 0.8425
```

En este caso, el valor del estadístico de contraste es 3.3681. El p-valor asociado es 0.7614 que, al ser superior a 0.05, nos indica que no se puede rechazar la hipótesis nula. Esto equivale a decir que, a un nivel de significación del 5%, puede suponerse que la ocupación del metro se produce según los valores de las proporciones consideradas.

Ejercicio Guiado 4

El resultado de un estudio de relación entre el dominio de la vista y el predominio de la mano viene dado en la siguiente tabla:

Contrastar, a un nivel de significación del 5%, si el dominio de la vista influye en el predominio de la mano.

Solución:

A continuación introducimos los datos en R-Commander. Para ello, tenemos que crear un fichero de texto como el que aparece en la Imagen.

```
> Ejemplo_VistaMano <-
+   read.table("C:/Users/user/OneDrive/Paquete R/PRACTICAS-S10/Ejemplo_VistaMano.txt",
+    header=TRUE, stringsAsFactors=TRUE, sep="", na.strings="NA", dec=".",
+   strip.white=TRUE)
```

Para transformar la tabla de contingencia en un conjunto de datos (data.frame) con el que R-Commander pueda trabajar hay que escribir las siguientes instrucciones en la ventana R Script, seleccionar las tres a la vez y darle a Ejecutar:

```
> P<-rep(Ejemplo_VistaMano$Vista,Ejemplo_VistaMano$Frecuencia)
> Q<-rep(Ejemplo_VistaMano$Mano,Ejemplo_VistaMano$Frecuencia)
> Vista_Mano<-data.frame(P,Q)
```

Una vez seleccionado, si pulsamos en el botón Visualizar conjunto de datos podemos comprobar que la tabla de frecuencias se ha transformado en un listado de datos.

El contraste de hipótesis que se debe resolver es:

$H_0 \equiv$ El dominio de la vista y el predominio de la mano son independientes

$H_1 \equiv$ El dominio de la vista y el predominio de la mano no son independientes

Tablas de contingencia de dos vías P, Q

```
> local({
+   .Table <- xtabs(~P+Q, data=Vista_Mano)
+   cat("\nFrequency table:\n")
+   print(.Table)
+   .Test <- chisq.test(.Table, correct=FALSE)
+   print(.Test)
+ })
```

```
Frequency table:
            Q
P             Ambiocular Dextrocular Levolular
  Ambidextro         28          20        27
  Dextro            105          52        57
  Zurdo              62          28        34

    Pearson's Chi-squared test

data:  .Table
X-squared = 4.0205, df = 4, p-value = 0.4032
```

Como podemos ver, el estadístico de contraste, que sigue una distribución Chi-Cuadrado con 4 grados de libertad, toma el valor 4.0205. El p-valor asociado al contraste es 0.4032. Como este p-valor es mayor que 0.05, no podemos rechazar la hipótesis nula, por lo que concluimos que el dominio de la vista y el predominio de la mano son variables independientes.

Ejercicio Guiado 5

Se quiere comprobar si la proporción de hombres y mujeres en la Faculta de Educación de la Universidad de Granada es la misma o no. Para ello, se selecciona una muestra aleatoria de estudiantes de la facultad, de los cuales 1218 son hombres y 3733 son mujeres. A un nivel de significación del 5%, ¿puede asumirse cierta la igualdad en el número de hombres y mujeres estudiantes?

Solución: Comencemos planteando las hipótesis del contraste. En este caso, se quiere probar la igualdad de hombres y de mujeres en la Faculta de Educación de la Universidad de Granada. Para ello, es posible plantear el contraste de hipótesis de dos formas distintas. Por un lado, se puede contrastar si la proporción de hombres es de 0.5 (en cuyo caso la proporción de mujeres será también 0.5 y habrá equidad entre ambos géneros) frente a que esta proporción es distinta de 0.5. Pero, alternativamente, se puede contrastar si la proporción de mujeres es de 0.5 (lo que implica que la proporción de hombre será, igualmente, de 0.5 y habrá equidad entre géneros) frente a que esta proporción es distinta de 0.5.

```
> Ejemplo_Facultad <-
+   read.table("C:/Users/user/OneDrive/Paquete R/PRACTICAS-S10/Ejemplo_Facultad.txt",
+    header=TRUE, stringsAsFactors=TRUE, sep="", na.strings="NA", dec=".",
+   strip.white=TRUE)
```

```
Warning in read.table("C:/Users/user/OneDrive/Paquete R/PRACTICAS-S10/
Ejemplo_Facultad.txt", : incomplete final line found by readTableHeader on 'C:/
Users/user/OneDrive/Paquete R/PRACTICAS-S10/Ejemplo_Facultad.txt'
```

Para transformar la tabla de frecuencias en un conjunto de datos (data.frame) con el que R pueda trabajar hay que escribir las siguientes instrucciones en la ventana R Script, seleccionar ambas a la vez y darle a Ejecutar

```
> P<-rep(Ejemplo_Facultad$Sexo,Ejemplo_Facultad$Frecuencia)
> Sexo_Facultad<-data.frame(P)
```

Para visualizar el conjunto de datos en forma de lista deberemos pulsar en el botón Conjunto de datos y seleccionar el nuevo conjunto de datos creado en forma de lista, al que hemos llamado Sexo_Facultad (observar la segunda instrucción). Una vez seleccionado, si pulsamos en el botón Visualizar conjunto de datos podemos comprobar que la tabla de frecuencias se ha transformado en un listado de datos.

Para resolver el contraste planteado, seleccionamos en el menú: Estadísticos/ Proporciones/Test de proporciones para una muestra

Test de proporciones para una muestra: P

```
> local({
+   .Table <- xtabs(~ P , data= Sexo_Facultad )
+   cat("\nFrequency counts (test is for first level):\n")
+   print(.Table)
```

```
+   binom.test(rbind(.Table), alternative='two.sided', p=.5, conf.level=.95)
+ })
```

```
Frequency counts (test is for first level):
P
Hombre  Mujer
  1218   3733


    Exact binomial test

data:  rbind(.Table)
number of successes = 1218, number of trials = 4951, p-value < 2.2e-16
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.2340658 0.2582562
sample estimates:
probability of success
             0.2460109
```

En primer lugar, se muestra la tabla de frecuencias. A continuación, los datos de entrada que se han usado para resolver el contraste (1218 hombres de 4951 estudiantes muestreados) así como el tipo de hipótesis alternativa (distinto de) y la proporción que se ha usado como referente para el contraste (0.5).

También se muestra un p-valor, que es el que nos ayuda a resolver el contraste. En este caso, el p-valor es menor que 0.05, por lo que podemos rechazar la hipótesis nula, asumiendo que la proporción de hombres en la Facultad de Educación de la Universidad de Granada no es de 0.5. Consecuentemente, la proporción de mujeres tampoco puede considerarse igual a 0.5 y puede concluirse que el número de hombres y mujeres en la facultad no es el mismo.

Por último, en la salida se incluye un intervalo de confianza al nivel de confianza indicado (95% por defecto), para la proporción de hombres en el municipio. Este intervalo es (0.234, 0.258). Como era de esperar, la proporción hombres que cursan estudios en la Facultad de Educación es inferior al de mujeres.

Ejercicio Guiado 6

Se realiza un estudio sobre el tiempo de vida en meses de la batería de un dispositivo electrónico. Se ha observado una muestra de 10 dispositivos y se ha anotado el tiempo de vida en meses:

18; 24; 17; 16; 14; 13; 15; 27; 9; 14. A un nivel de significación del 5%, ¿se puede considerar aleatoriedad en la muestra?

Solución: Formulamos el contraste que debemos resolver.

$H_0 \equiv$ Los datos de la muestra son aleatorios $H_1 \equiv$ Los datos de la muestra no son aleatorios

cargamos el fichero de datos creado: Datos/Importar datos/desde archivo de texto, portapapeles o URL

Se muestra la siguiente ventana en la cual vamos a introducir el nombre que queremos asignarle al conjunto de datos con el que vamos a trabajar; en nuestro caso, escribiremos Ejemplo_Dispositivos. El resto de opciones las dejamos por defecto, ya que el archivo de texto que hemos creado cumple con todas ellas.

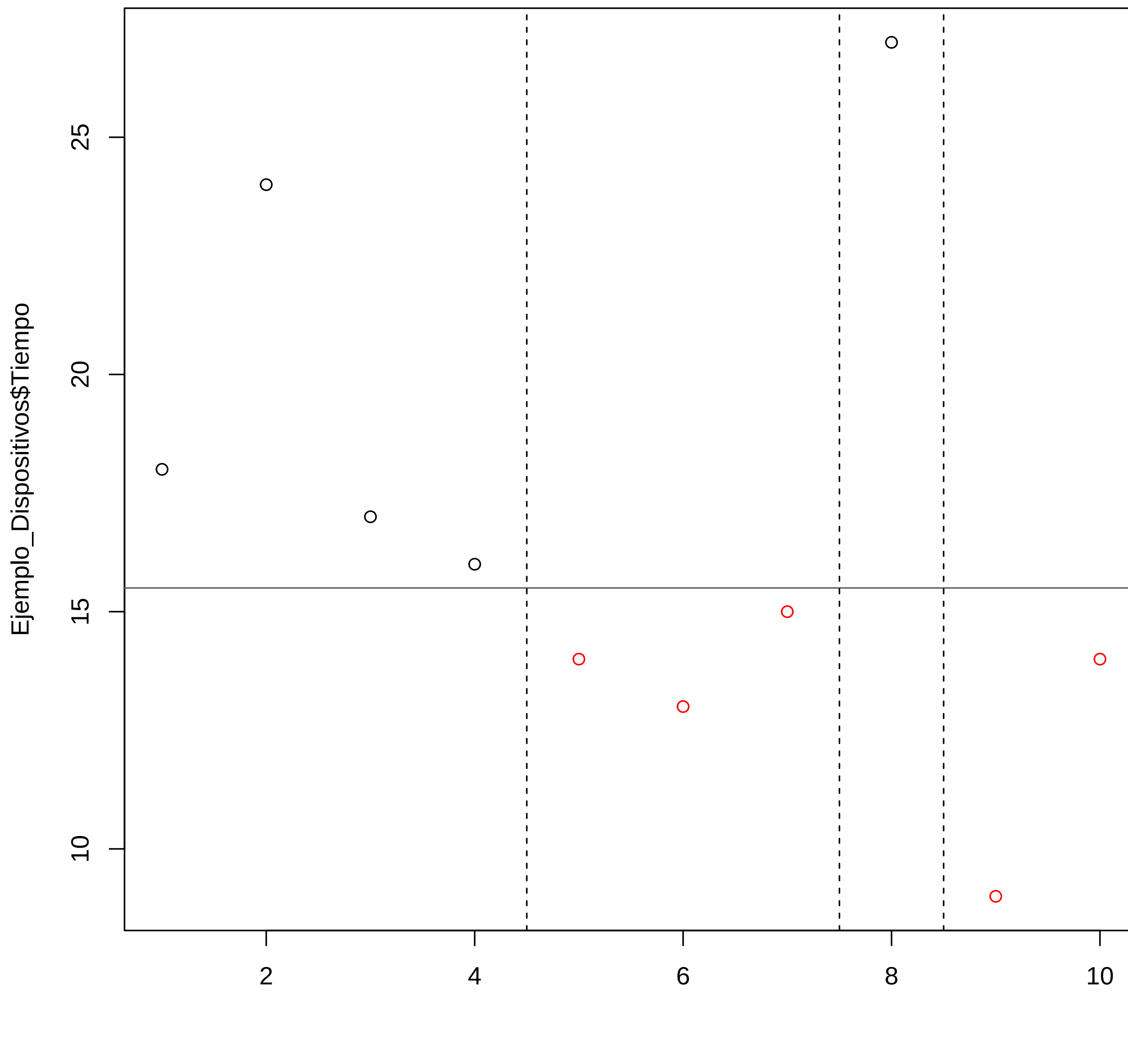
cargamos el fichero de datos creado: Datos/Importar datos/desde archivo de texto, portapapeles o URL…

Se muestra la siguiente ventana en la cual vamos a introducir el nombre que queremos asignarle al conjunto de datos con el que vamos a trabajar; en nuestro caso, escribiremos Ejemplo_Dispositivos. El resto de opciones las dejamos por defecto, ya que el archivo de texto que hemos creado cumple con todas ellas.

```
> Ejemplo_Dispositivos <-
+   read.table("C:/Users/user/OneDrive/Paquete R/PRACTICAS-S10/Ejemplo_Dispositivo.txt",
+    header=TRUE, stringsAsFactors=TRUE, sep="", na.strings="NA", dec=".",
+   strip.white=TRUE)
```

A continuación, escribimos en R Script la siguiente orden para llamar a la función runs.test

```
> randtests::runs.test (Ejemplo_Dispositivos$Tiempo, alternative="two.sided",threshold=med
```


Ejemplo_Dispositivos$Tiempo
10
15
20
25
2
4
6
8
10

```
	Runs Test

data:  Ejemplo_Dispositivos$Tiempo
statistic = -1.3416, runs = 4, n1 = 5, n2 = 5, n = 10, p-value = 0.1797
alternative hypothesis: nonrandomness
```

Según los resultados del test de rachas, se han encontrado 4 rachas (runs), que vienen separadas por líneas discontinuas verticales. Hay 5 valores por encima de la mediana (n1), marcados en negro, y otros 5 valores por debajo de la mediana (n2), marcados en rojo.

El p-valor asociado al contraste es 0.1797 superior a 0.05, por lo que no es posible rechazar la hipótesis nula. Por tanto, podemos concluir que los datos de la muestra son aleatorios

Ejercicio Guiado 7

Las puntuaciones de selectividad de 10 estudiantes han sido las siguientes: 10.81, 13.30, 4.20, 7.14, 9.29, 8.79, 4.73, 9.26, 5.74, 4.65. ¿Puede suponerse, a un nivel de significación del 5%, que dichas calificaciones se ajustan a una distribución normal de media 7 y desviación típica 2.5?

Solución: El contraste de hipótesis que se plantea es el siguiente:

$H_{0} $ Los datos de la muestra proceden de una distribución N(7,2.5)

$H_{1} $ Los datos de la muestra no proceden de una distribución N(7,2.5)

Expresión 45: Contraste de hipótesis para el Ejercicio Guiado 7

Lo primero que vamos a hacer es crear un fichero de texto con los datos del problema con la siguiente estructura:

```
> Ejemplo_Selectividad <-
+   read.table("C:/Users/user/OneDrive/Paquete R/PRACTICAS-S10/Ejemplo_Selectividad.txt",
+    header=TRUE, stringsAsFactors=TRUE, sep="", na.strings="NA", dec=".",
+   strip.white=TRUE)
```

A continuación, escribimos en R Script la siguiente orden para llamar a la función ks.test. Debemos tener en cuenta que la distribución de comparación es la distribución normal (por tanto, el argumento y tomará el valor pnorm) de media igual a 7 y desviación típica igual a 2.5.

```
> ks.test(Ejemplo_Selectividad$Calificaciones,y=pnorm,7,2.5,alternative="two.sided")
```

```
    One-sample Kolmogorov-Smirnov test

data:  Ejemplo_Selectividad$Calificaciones
D = 0.263, p-value = 0.421
alternative hypothesis: two-sided
```

En este caso, el valor del estadístico de contraste es 0.263 y el p-valor asociado al contraste es 0.421. Como el p-valor es superior a 0.05 no podemos rechazar la hipótesis nula, por lo que concluimos que los datos de la muestra proceden de una distribución normal de media 7 y de desviación típica 2.5.

Ejercicio Guiado 8

En unos grandes almacenes se realiza un estudio sobre el tiempo de cobro de las cajeras por cada cliente. Para ello, se observa el número de minutos que tardan en completar el cobro de los productos de dos cajeras a diez clientes cada uno:

```
> Ejemplo_Cajeras <-
+   read.table("C:/Users/user/OneDrive/Paquete R/PRACTICAS-S10/Ejemplo_Cajeras.txt",
+    header=TRUE, stringsAsFactors=TRUE, sep="", na.strings="NA", dec=".",
+   strip.white=TRUE)
```

Para resolver el contraste planteado, seleccionamos en el menú: Estadísticos/ Test no paramétricos/Test de Wilcoxon para dos muestras

Test de Wilcoxon: Frecuencia ~ Cajera

```
> Tapply(Frecuencia ~ Cajera, median, na.action=na.omit,
+   data=Ejemplo_Cajeras) # medians by group
```

```
  A   B
4.5 3.5
```

```
> wilcox.test(Frecuencia ~ Cajera, alternative='two.sided', exact=TRUE,
+   correct=FALSE, data=Ejemplo_Cajeras)
```

```
Warning in wilcox.test.default(x = c(10L, 4L, 6L, 5L, 7L, 9L, 3L, 3L, 2L, :
cannot compute exact p-value with ties
```

```
	Wilcoxon rank sum test

data:  Frecuencia by Cajera
W = 61.5, p-value = 0.3768
alternative hypothesis: true location shift is not equal to 0
```

En este caso, el p-valor asociado al contraste es 0.3768. Como este p-valor es mayor que 0.05 no se puede rechazar la hipótesis nula, considerando un nivel de significación del 5%. Por tanto, concluimos que las medianas de los tiempos de ambas cajeras pueden asumirse iguales.

Ejercicio Guiado 9

En una clínica se pretende comprobar la eficacia de un tratamiento de pérdida de peso. Para ello, se realiza un estudio a 10 voluntarios, y se registra el peso antes y después del tratamiento.

```
> Ejemplo_Tratamiento <-
+   read.table("C:/Users/user/OneDrive/Paquete R/PRACTICAS-S10/Ejemplo_Tratamiento.txt",
+    header=TRUE, stringsAsFactors=TRUE, sep="", na.strings="NA", dec=".",
+   strip.white=TRUE)
```

Solución:

Para resolver el contraste planteado, seleccionamos en el menú: Estadísticos/ Test no paramétricos/Test de Wilcoxon para muestras pareadas

- En Datos tenemos que realizar dos selecciones: por un lado, elegir la variable que nos indica el peso antes del tratamiento "Antes" en la ventana Primera variable, y la que nos indica el peso después del tratamiento "Después" en la ventana Segunda variable.
- En la pestaña Opciones, necesitamos seleccionar la Hipótesis alternativa mayor que (>), en Tipo de prueba seleccionamos la opción "Exacto" para que no nos aplique el corrector por continuidad y pulsamos Aceptar.

Test de Wilcoxon para datos pareados Antes, Despues

```
> with(Ejemplo_Tratamiento, median(Antes - Despues, na.rm=TRUE))
```

```
[1] 4.4
```

```
>   # median difference
> with(Ejemplo_Tratamiento, wilcox.test(Antes, Despues, alternative='greater',
+    exact=TRUE, paired=TRUE))
```

```
	Wilcoxon signed rank exact test

data:  Antes and Despues
V = 55, p-value = 0.0009766
alternative hypothesis: true location shift is greater than 0
```

En este ejemplo, el p-valor asociado al contraste es 0.0009766, inferior a 0.05, por lo que se rechaza la hipótesis nula considerando un nivel de significación del 5%. Esto quiere decir que el tratamiento utilizado es efectivo para reducir el peso en los pacientes.