1 CONTRASTES DE HIPÓTESIS

Contrastes de hipótesis paramétricos
Supuesto Práctico 1
Con el fin de estudiar el número medio de flexiones continuadas que pueden realizar sus alumnos, un profesor de educación física somete a 75 de ellos, elegidos aleatoriamente, a una prueba. El número de flexiones realizado por cada alumno, así como su sexo y si realizan o no deporte fuera del horario escolar se muestran en el fichero Flexiones.txt.

```r
datos<-read.table("Flexiones.txt", header=TRUE)
datos

##    Flexiones Sexo Deporte
## 1         60    H       0
## 2         41    H       0
## 3         53    M       1
## 4         53    M       0
## 5         41    H       0
## 6         56    H       0
## 7         50    H       0
## 8         53    M       1
## 9         50    M       1
## 10        48    M       0
## 11        50    M       1
## 12        48    M       1
## 13        56    H       0
## 14        52    M       1
## 15        54    M       0
## 16        50    H       1
## 17        50    H       0
## 18        54    H       0
## 19        52    H       1
## 20        48    H       0
## 21        48    H       1
## 22        35    M       1
## 23        50    M       1
## 24        41    M       1
## 25        56    M       1
## 26        52    M       1
## 27        56    M       0
## 28        54    H       1
## 29        53    H       0
## 30        53    M       0
## 31        53    H       0
## 32        41    M       1
```

```
## 33          48    M        0
## 34          50    H        1
## 35          50    M        1
## 36          52    H        0
## 37          53    M        0
## 38          35    H        0
## 39          35    H        0
## 40          54    M        0
## 41          46    M        1
## 42          48    H        0
## 43          50    M        0
## 44          48    H        0
## 45          41    M        0
## 46          48    M        1
## 47          60    H        1
## 48          53    M        0
## 49          54    M        1
## 50          56    H        1
## 51          50    H        1
## 52          41    H        0
## 53          60    M        1
## 54          60    M        1
## 55          54    H        0
## 56          54    H        0
## 57          53    H        0
## 58          35    M        0
## 59          54    H        0
## 60          48    M        0
## 61          50    H        0
## 62          54    H        0
## 63          54    H        0
## 64          53    H        0
## 65          52    H        0
## 66          50    H        0
## 67          52    H        0
## 68          48    H        1
## 69          46    H        1
## 70          53    H        0
## 71          50    H        0
## 72          35    H        0
## 73          50    H        1
## 74          60    M        1
## 75          50    H        0
```

Una vez hecho esto, introducimos en R el nivel de significación que proporciona el enunciado.

```r
alpha<-0.05
```

A continuación, calculamos el valor del estadístico de contraste.

```r
media<-mean(datos$Flexiones)
mu_0<-55
varianza<-7.5
n<-nrow(datos)
Z<-(media-mu_0)/(sqrt(varianza)/sqrt(n))
Z

## [1] -15.47408
```

Y también el valor crítico, que en este caso coincide con $Z_{1-\alpha/2}$, el cuantil $1-\alpha/2$ de una distribución normal de media 0 y varianza 1.

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

## [1] 1.959964
```

Contrastes de hipótesis para la media de una población normal con Varianza desconocida

Supuesto Práctico 2

Considerando nuevamente el conjunto de datos que se ha presentado en el Supuesto práctico1, relativo al número de flexiones y el sexo de los alumnos. Contrastar a un nivel de significación del 2% la hipótesis de que el número medio de flexiones realizada por los alumnos es de 50. Suponer en este caso que el número de flexiones se distribuye según una normal de varianza desconocida. El fichero es Flexiones.txt.

```r
t.test(datos$Flexiones,alternative="two.sided",mu=50, conf.level=0.98)

##
##  One Sample t-test
##
## data:  datos$Flexiones
## t = 0.15451, df = 74, p-value = 0.8776
## alternative hypothesis: true mean is not equal to 50
## 98 percent confidence interval:
##  48.46512 51.74822
## sample estimates:
## mean of x
##  50.10667
```

Supuesto Práctico 3

Considerando nuevamente el conjunto de datos que se ha presentado en el Supuesto práctico1, relativo al número de flexiones y el sexo de los alumnos. Contrastar a un nivel de confianza del 95%, si la proporción de alumnos varones es mayor o igual que 0.5 frente a que dicha proporción es menor. El fichero es Flexiones.txt.

```
table(datos$Sexo)

##
##  H  M
## 43 32
```

De los 75 estudiantes que conforman la muestra, 43 son chicos. Por lo que la llamada a prop.test sería la siguiente:

```
prop.test(43,75,p=0.5,alternative = "less",conf.level = 0.95)

##
##  1-sample proportions test with continuity correction
##
## data:  43 out of 75, null probability 0.5
## X-squared = 1.3333, df = 1, p-value = 0.8759
## alternative hypothesis: true p is less than 0.5
## 95 percent confidence interval:
##  0.0000000 0.6693525
## sample estimates:
##         p
## 0.5733333
```

El valor del estadístico de contraste es 1.3333, con un p-valor de 0.8759. Como el p-valor es mayor que el nivel de significación, que es 0.05, no rechazamos la hipótesis de que la proporción de alumnos es mayor o igual que 0.5.

Supuesto Práctico 4

Continuando con los datos relativos a las flexiones realizadas por un grupo de estudiantes y asumiendo que las flexiones que realizan los chicos y las que realizan las chicas se distribuyen según sendas distribuciones normales con medias y varianzas desconocidas, contrastar a un nivel de significación del 5% si las varianzas poblacionales de ambas distribuciones pueden asumirse iguales.

Lo primero que tenemos que hacer para aplicar la función var.test es separar en dos variables los datos relativos a las flexiones realizadas por los chicos y por las chicas.

```
Flexiones.chicos<-datos$Flexiones[datos$sexo=="H"]
Flexiones.chicas<-datos$Flexiones[datos$sexo=="M"]
```

A continuación, utilizamos la función var.test

```
var.test(Flexiones.chicos,Flexiones.chicas,alternative="two.sided",conf.level=0.95)

## Error in var.test.default(Flexiones.chicos, Flexiones.chicas, alternative
= "two.sided", :  not enough 'x' observations
```

Supuesto Práctico 5

En vista de los resultados obtenidos en el Supuesto Práctico 4, y suponiendo que el número de flexiones que realizan los alumnos y las alumnas se distribuyen de acuerdo a variables normales de medias y varianzas desconocidas, ¿puede suponerse, a un nivel de significación del 5%, que el número medio de flexiones que realizan los chicos y las chicas es igual?

```
datos<-read.table("Flexiones.txt", header=TRUE)
Flexiones.chicos<-datos$Flexiones[datos$sexo=="H"]
Flexiones.chicas<-datos$Flexiones[datos$sexo=="M"]

t.test(Flexiones.chicos,Flexiones.chicas,alternative="two.sided",mu =0, var.equal=TRUE)

## Error in t.test.default(Flexiones.chicos, Flexiones.chicas, alternative
= "two.sided", :  not enough 'x' observations
```

Supuesto Práctico 6

Para estudiar los efectos de un programa de control de peso, el profesor de educación física selecciona aleatoriamente a 6 alumnos y se les toma nota de sus pesos antes y después de pasar por el programa.

```
Antes<-c(72,73.5,70,71.5,76,80.5)
Despues<-c(73,74.5,74,74.5,75,80)
```

¿Puede suponerse, a un nivel de significación del 5%, que el programa para el control de peso es efectivo? O, dicho de otra forma, ¿el peso medio de los alumnos antes de someterse al programa es igual al peso medio tras el programa?

```
t.test(Antes,Despues,alternative="two.sided",mu=0,paired=TRUE)

##
##  Paired t-test
##
```

```
## data:  Antes and Despues
## t = -1.5759, df = 5, p-value = 0.1759
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.2889862  0.7889862
## sample estimates:
## mean of the differences
##                   -1.25
```

Supuesto Práctico 7

Retomando el conjunto de datos relativo a las flexiones que realizan un grupo de estudiantes, contrastar, a un nivel de significación del 8% si la proporción de alumnos y de alumnas que practican deporte pueden considerarse iguales.

En primer lugar, utilicemos el comando table para determinar cuántos chicos y cuantas chicas practican deporte.

```
table(datos$Sexo,datos$Deporte)

##
##      0  1
##   H 32 11
##   M 13 19
```

En total, 11 de los 43 y 19 de las 32 chicas muestreados practican deporte fuera del horario escolar. Vamos a crear dos vectores con esta información: en uno indicaremos el total de chicos y chicas que practican deporte y en el otro el total de chicos y chicas en la muestra.

```
vector_Deporte<-c(11,19)
vector_Sexo<-c(43,32)

prop.test(vector_Deporte,vector_Sexo,alternative = "two.sided",conf.level = 0.92)

##
##  2-sample test for equality of proportions with continuity correction
##
## data:  vector_Deporte out of vector_Sexo
## X-squared = 7.3787, df = 1, p-value = 0.0066
## alternative hypothesis: two.sided
## 92 percent confidence interval:
##  -0.5566881 -0.1191840
## sample estimates:
##   prop 1   prop 2
## 0.255814 0.593750
```

Día de la semana	Número de ingresos
Lunes	78
Martes	90
Miércoles	94
Jueves	89
Viernes	110
Sábado	84
Domingo	44

Según la salida de la función prop.test, el p-valor asociado al contraste es 0.0066, que al ser menor que el nivel de significación (0.08), nos lleva a concluir que las proporciones de chicos y chicas que hacen deporte no coinciden.

La directora de un hospital quiere comprobar si los ingresos en el hospital se producen en la misma proporción durante todos los días de la semana. Para ello, se anota el número de ingresos durante una semana cualquiera. Los datos se recogen en la siguiente tabla:

Contrastar, a un nivel de significación del 5%, si la hipótesis de la directora del hospital puede suponerse cierta. ¿Puede asumirse que las proporciones de ingresos de lunes a domingo son (0.15, 0.15, 0.15, 0.15, 0.20, 0.15, 0.05)?

Solución

```
frecuencias<-c(78,90,94,89,110,84,44)

chisq.test(frecuencias)

##
##  Chi-squared test for given probabilities
##
## data:  frecuencias
## X-squared = 29.389, df = 6, p-value = 5.135e-05
```

El estadístico de contraste, que sigue una distribución chi-cuadrado, toma el valor 29.389. 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 los ingresos hospitalarios no se producen en la misma proporción todos los días de la semana.

Para comprobar si el vector (0.15, 0.15, 0.15, 0.15, 0.20, 0.15, 0.05) puede considerarse como el vector de proporciones de ingresos hospitalarios durante los 7 días de la semana, creamos un vector en R que recoja estos valores:

Cara dado	Frecuencia
1	116
2	120
3	115
4	120
5	125
6	124

```
proporciones<-c(0.15, 0.15, 0.15, 0.15, 0.20, 0.15, 0.05)
```

Volvemos a llamar a la función chisq.test incluyendo como argumento el vector que acabamos de definir.

```
chisq.test(frecuencias,p=proporciones)

##
##  Chi-squared test for given probabilities
##
## data:  frecuencias
## X-squared = 9.5286, df = 6, p-value = 0.146
```

En este caso, el valor del estadístico de contraste es 9.5286. El p-valor asociado es 0.146 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 los ingresos hospitalarios se producen según los valores que se recogen en el vector proporciones.

Supuesto Práctico 9
Lanzamos un dado 720 veces y obtenemos los resultados que se muestran en la tabla.

Contrastar la hipótesis de que el dado está bien construido.

Solución
Comencemos introduciendo en R las frecuencias con las que aparecen los valores del dado.

```
frecuencias<-c(116,129,115,120,125,124)
```

Que el dado esté bien construido equivale a decir que todos sus valores aparecen en la misma proporción. Por tanto, el contraste de hipótesis que se debe resolver es el siguiente:

$H_0 \equiv$ Los valores del dado aparecen en la misma proporción

$H_1 \equiv$ Los valores del dado no aparecen en la misma proporción

Para resolver este contraste de hipótesis se utiliza la función chisq.test, que recibe como argumento el vector de frecuencias.

```
chisq.test(frecuencias)

##
##  Chi-squared test for given probabilities
##
## data:  frecuencias
## X-squared = 1.2305, df = 5, p-value = 0.9419
```