Problemas de bondad de ajuste

1. Se lanzaron un par de dados 500 veces. En la taula siguiente se muestran las sumas que se obtuvieron. Provar la hipótesis de que los dados no estaban trucados, es decir, comprobar que el p-valor para aceptar que los dados no estan trucados no es pequeño.

Suma	Frecuencia
$2, 3, 4$	74
$5, 6$	120
7	83
$8, 9$	135
$10, 11, 12$	88

Solución

Ponemos las frecuencias empíricas en el vector `frecuencias.empíricas`:

```r
frecuencias.empiricas= c(74,120,83,135,88)
```

Calculamos las probabilidades teóricas suponiendo que los dados no están trucados:

$$
\begin{aligned}
P(S=2,3,4) &= P\{(1,1),(1,2),(2,1),(1,3),(3,1),(2,2)\} = \tfrac{6}{36} = \tfrac{1}{6} \approx 0.1667,\\
P(S=5,6) &= P\{(1,4),(4,1),(2,3),(3,2),(1,5),(5,1),(2,4),(4,2),(3,3)\} = \tfrac{9}{36} = \tfrac{1}{4} = 0.25,\\
P(S=7) &= P\{(1,6),(6,1),(2,5),(5,2),(3,4),(4,3)\} = \tfrac{6}{36} = \tfrac{1}{6} \approx 0.1667,\\
P(S=8,9) &= P\{(2,6),(6,2),(3,5),(5,3),(4,4),(3,6),(6,3),(4,5),(5,4)\} = \tfrac{9}{36} = \tfrac{1}{4} = 0.25,\\
P(S=10,11,12) &= P\{(4,6),(6,4),(5,5),(5,6),(6,5),(6,6)\} = \tfrac{6}{36} = \tfrac{1}{6} \approx 0.1667.
\end{aligned}
$$

```r
probabilidades.teóricas = c(1/6,1/4,1/6,1/4,1/6)
```

Por último realizamos el test χ^2:

```r
chisq.test(frecuencias.empiricas,p=probabilidades.teóricas)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  frecuencias.empiricas
## X-squared = 2.308, df = 4, p-value = 0.6793
```

Como el p-valor es grande, podemos concluir que no tenemos evidencias para rechazar que los dados no están trucados.

2. El 1972, el informe oficial dió la información siguiente sobre el número de días que fueron internados los enfermos en el hospital en el año 1971.

Número de días	Número de enfermos
1	89
2	152
3	105
$4-5$	165
$6-9$	221
$10-14$	124
$15-30$	106
31 o más	38

Probar la hipótesis que estos datos se obtuvieron de una distribución χ^2 con 4 grados de libertad.

Solución

Ponemos las frecuencias empíricas en el vector `frecuencias.empíricas`:

```
frecuencias.empíricas= c(89,152,105,165,221,124,106,38)
```

Como la distribución χ^2 es continua con valores en $(0,\infty)$, consideramos las clases siguientes:

Clases	Frecuencias empíricas
0-1.5	89
1.5-2.5	152
2.5-3.5	105
3.5-5.5	165
5.5-9.5	221
9.5-14.5	124
14.5-30.5	106
30.5-∞	38

Calculemos las probabilidades teóricas:

```
extremos.izquierda = c(0,1.5,2.5,3.5,5.5,9.5,14.5,30.5)
extremos.derecha=c(extremos.izquierda[-1],Inf)
probabilidades.teóricas = pchisq(extremos.derecha,4)-pchisq(extremos.izquierda,4)
```

Por último, realizamos el test χ^2:

```
chisq.test(frecuencias.empíricas,p=probabilidades.teóricas)
```

```
## Warning in chisq.test(frecuencias.empíricas, p = probabilidades.teóricas): Chi-
## squared approximation may be incorrect
```

```
##
##  Chi-squared test for given probabilities
##
## data:  frecuencias.empíricas
## X-squared = 374876, df = 7, p-value < 2.2e-16
```

R nos da un warning avisándonos que algunas celdas tienen frecuencias teóricas menores que 5. Averiguemos cuáles son estan celdas:

```
round(probabilidades.teóricas*sum(frecuencias.empíricas),2)
```

```
## [1] 173.36 182.01 166.76 238.15 189.98  43.89   5.86   0.00
```

Vemos que la última celda tiene una frecuencia teórica menor que 5. Juntamos las dos últimas celdas y volvemos a realizar el test χ^2:

```
frecuencias.empíricas2 = c(frecuencias.empíricas[-c(7,8)],sum(frecuencias.empíricas[7:8]))
probabilidades.teóricas2 = c(probabilidades.teóricas[-c(7,8)],sum(probabilidades.teóricas[7:8]))
chisq.test(frecuencias.empíricas2,p=probabilidades.teóricas2)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  frecuencias.empíricas2
## X-squared = 3499.7, df = 6, p-value < 2.2e-16
```

Como el p-valor es pequeño, concluimos que tenemos evidencias para afirmar que los datos no provienen de una distribución χ^2 con 4 grados de libertad.

3. Se realizó una prueba de inteligencia a 100 estudiantes. En la tabla siguiente se muestran las calificaciones obtenidas:

Calificación x	Frecuencia
$70 < x \leq 90$	8
$90 < x \leq 110$	38
$110 < x \leq 130$	45
$130 < x \leq 150$	9

   Podéis suponer que las calificaciones anteriores son una muestra aleatoria de las que tendrien todas las personas posibles que hicieran la prueba. Probar la hipótesis de que las calificaciones obtenidas por la población (conceptualmente infinita) estarían distribuidas normalmente.

Solución

Ponemos las frecuencias empíricas en el vector `frecuencias.empíricas`:

```
frecuencias.empíricas= c(8,38,45,9)
```

Como la distribución normal es continua con valores en $\mathbb{R}$, consideramos las clases siguientes:

Clases	Frecuencias empíricas
$-\infty$-90	8
90-110	38
110-130	45
130-∞	9

Antes de calcular las probabilidades téoricas, necesitamos estimar los valores μ y σ de la normal:

```
valores.prueba=rep(c(80,100,120,140),frecuencias.empíricas)
mu = mean(valores.prueba)
sigma = sd(valores.prueba)
```

Calculemos las probabilidades teóricas:

```
extremos.izquierda = c(-Inf,90,110,130)
extremos.derecha=c(extremos.izquierda[-1],Inf)
probabilidades.teóricas = pnorm(extremos.derecha,mu,sigma)-pnorm(extremos.izquierda,mu,sigma)
```

Por último, realizamos el test χ^2:

```
chisq.test(frecuencias.empíricas,p=probabilidades.teóricas)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  frecuencias.empíricas
## X-squared = 0.64549, df = 3, p-value = 0.8859
```

El p-valor que nos ha dado `R` es incorrecto ya que hemos tenido que estimar dos parámetros. Los grados de libertad correctos son: $4 - 1 - 2 = 1$:

```r
valor.estadístico=chisq.test(frecuencias.empíricas,p=probabilidades.teóricas)$statistic
p.valor.correcto = pchisq(valor.estadístico[[1]],1,lower.tail = FALSE)
p.valor.correcto
```

```
## [1] 0.4217309
```

Como el p-valor es grande, concluimos que no tenemos evidencias suficientes para rechazar que los datos no provengan de una distribución normal.

4. Consideremos la muestra aleatoria siguiente de una variable aleatoria X tal que $X(\Omega) = [0,1]$. Probar mediante el test χ^2, que podemos considerar que X sigue una distribución uniforme en $[0,1]$. (Considerar intervalos de amplitud 0.25.)

   0.479, 0.889, 0.216, 0.596, 0.359, 0.347, 0.646, 0.359, 0.991, 0.227
   0.774, 0.760, 0.448, 0.992, 0.742, 0.402, 0.049, 0.213, 0.296, 0.711

Solución

Calculemos las frecuencias empíricas considerando las clases siguientes:

Clases	Frecuencias empíricas
0-0.25	4
0.25-0.5	7
0.5-0.75	4
0.75-1	5

Calculemos las probabilidades teóricas:

```r
frecuencias.empiricas=c(4,7,4,5)
extremos.izquierda = c(0,0.25,0.5,0.75)
extremos.derecha=c(extremos.izquierda[-1],1)
probabilidades.teóricas = punif(extremos.derecha)-punif(extremos.izquierda)
```

Por último realizamos el test χ^2:

```r
chisq.test(frecuencias.empiricas,p=probabilidades.teóricas)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  frecuencias.empiricas
## X-squared = 1.2, df = 3, p-value = 0.753
```

Como el p-valor es grande, concluimos que no tenemos evidencias suficientes para rechazar que los datos no provengan de una distribución uniforme.

5. Sea X la variable aleatòria que tiene como función de densidad:

$$f(x) = \begin{cases} x+1, & \text{si } x \in [-1,0], \\ 1-x, & \text{si } x \in [0,1], \\ 0, & \text{en caso contrario.} \end{cases}$$

Comprobar mediante el test de la χ^2 que la muestra aleatòria simple siguiente tiene la misma distribución que X:

0.183,	0.647,	0.148,	-0.143,	-0.625,	0.858,	-0.177,	0.350,
-0.188,	-0.059,	0.845,	0.031,	-0.156,	0.564,	-0.235,	0.237,
0.294,	-0.257,	0.110,	0.478,	0.647,	0.276,	-0.528,	-0.075,
-0.498,	0.395,	-0.163,	-0.075,	-0.623,	0.053,	-0.647,	0.348,
-0.795,	-0.132,	-0.381,	-0.017,	-0.227,	0.277,	0.590,	-0.832

Solución

Consideremos las clases siguientes:

$$[-1,-0.75) \cup [-0.75,-0.5) \cup [-0.5,-0.25) \cup [-0.25,0) \cup [0,0.25) \cup [0.25,0.5) \cup [0.5,0.75) \cup [0.75,1].$$

Calculemos las frecuencias empíricas:

```
muestra.x = c(0.183,0.647,0.148,-0.143,-0.625,0.858,-0.177,0.350,-0.188,-0.059,
0.845,0.031,-0.156,0.564,-0.235,0.237,0.294,-0.257,0.110,0.478,
0.647,0.276,-0.528,-0.075,-0.498,0.395,-0.163,-0.075,-0.623,0.053,
-0.647,0.348,-0.795,-0.132,-0.381,-0.017,-0.227,0.277,0.590,-0.832)
extremos.izquierda = seq(from=-1,to=0.75,by=0.25)
extremos.derecha = c(extremos.izquierda[-1],1)
frecuencias.empiricas=c()
for (i in 1:length(extremos.izquierda)){
  frecuencias.empiricas=c(frecuencias.empiricas,length(muestra.x[muestra.x>=extremos.izquierda[i] &
                                        muestra.x< extremos.derecha[i]]))
}
```

La tabla de frecuencias empíricas será:

Clases	Frecuencias empíricas
$[-1,-0.75)$	2
$[-0.75,-0.5)$	4
$[-0.5,-0.25)$	3
$[-0.25,0)$	12
$[0,0.25)$	6
$[0.25,0.5)$	7
$[0.5,0.75)$	4
$[0.75,1]$	2

Calculemos a continuación las probabilidades teóricas calculando primero la función de distribución de la variable aleatoria X:

$$F(x)=\begin{cases} 0, & \text{si } x<-1,\\ \displaystyle\int_{-1}^{x}(t+1)\,dt=\left[\frac{(t+1)^2}{2}\right]_{-1}^{x}=\frac{(x+1)^2}{2}, & \text{si } x\in[-1,0],\\ \displaystyle\frac{(x+1)^2}{2}+\int_0^x(1-t)\,dt=\frac{1}{2}+\left[\frac{(1-t)^2}{2}\right]_x^0=\frac{1}{2}+\frac{1}{2}-\frac{(1-x)^2}{2}=-\frac{x^2}{2}+x+\frac{1}{2}, & \text{si } x\in[0,1],\\ 1, & \text{si } x>1. \end{cases}$$

Definamos la función anterior en `R`:

```
función.distribución= function(x){return(ifelse(x< -1,0,ifelse(x<=0,(x+1)^2/2,
                                                      ifelse(x<=1,-x^2/2+x+1/2,1))))}
```

Las probabilidades teóricas serán:

```
probabilidades.teóricas = función.distribución(extremos.derecha)-
            función.distribución(extremos.izquierda)
```

Apliquemos por último el test χ^2:

```
chisq.test(frecuencias.empiricas,p=probabilidades.teóricas)
```

```
## Warning in chisq.test(frecuencias.empiricas, p = probabilidades.teóricas): Chi-
## squared approximation may be incorrect
```

```
##
##  Chi-squared test for given probabilities
##
## data:  frecuencias.empiricas
## X-squared = 4.7848, df = 7, p-value = 0.6862
```

`R` nos avisa que hay clases con frecuencias teóricas menores que 5. Veamos qué clases son:

```
probabilidades.teóricas*sum(frecuencias.empiricas)
```

```
## [1] 1.25 3.75 6.25 8.75 8.75 6.25 3.75 1.25
```

Son las dos primeras y las dos últimas. Las juntamos y volvemos a realizar el test χ^2:

```
frecuencias.empiricas2 = c(sum(frecuencias.empiricas[1:2]),frecuencias.empiricas[3:6],
                           sum(frecuencias.empiricas[7:8]))
probabilidades.teóricas2 = c(sum(probabilidades.teóricas[1:2]),probabilidades.teóricas[3:6],
                           sum(probabilidades.teóricas[7:8]))
chisq.test(frecuencias.empiricas2,p=probabilidades.teóricas2)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  frecuencias.empiricas2
## X-squared = 4.2514, df = 5, p-value = 0.5138
```

Como el p-valor es grande, concluimos que no tenemos evidencias suficientes para rechazar que los datos anteriores no provengan de la variable aleatoria X.