Hoja 4 (c): Tests Chi-cuadrado con R

Estadística Computacional I. Grado en Estadística

Departamento de Estadística e Investigación Operativa. Universidad de Sevilla

Ejercicio 1

Bondad de ajuste. Comprobar si un dado es correcto a partir del número de veces que ha salido cada lado.

```
frecu <- c(22,21,22,27,22,36)
probs <- rep(1/6,6)
```

Solución

Utilizamos test Chi-Cuadrado

```
frecu <- c(22,21,22,27,22,36)
probs <- rep(1/6,6)

chisq.test(frecu, p = probs)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  frecu
## X-squared = 6.72, df = 5, p-value = 0.2423
```

Aceptamos que sigue ese modelo probabilístico (equiprobable).

Ejercicio 2

Por defecto se compara con la unif. discreta. En el siguiente ejemplo se trata de ver si en un texto las apariciones de las letras E,T,N,R,O se distribuyen según los valores conocidos en inglés.

```
x <-c(100,110,80,55,14)
probs <-c(29, 21, 17, 17, 16)/100
```

Solución

```
x <-c(100,110,80,55,14)
probs <-c(29, 21, 17, 17, 16)/100

chisq.test(x, p = probs)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  x
## X-squared = 55.395, df = 4, p-value = 2.685e-11
```

Se rechaza que la muestra siga el modelo teórico.

Ejercicio 3

En la siguiente simulación se ilustra la calidad de la aproximación. Se generan M muestras de tamaño n de una ley Uniforme discreta.

```
probabi<- c(0.03,0.25,0.45,0.27)
sum(probabi)
```

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

```
set.seed(12345)
n<-50 #tamaño muestral
n*probabi  #Se cumplen las condiciones
```

```
## [1]  1.5 12.5 22.5 13.5
```

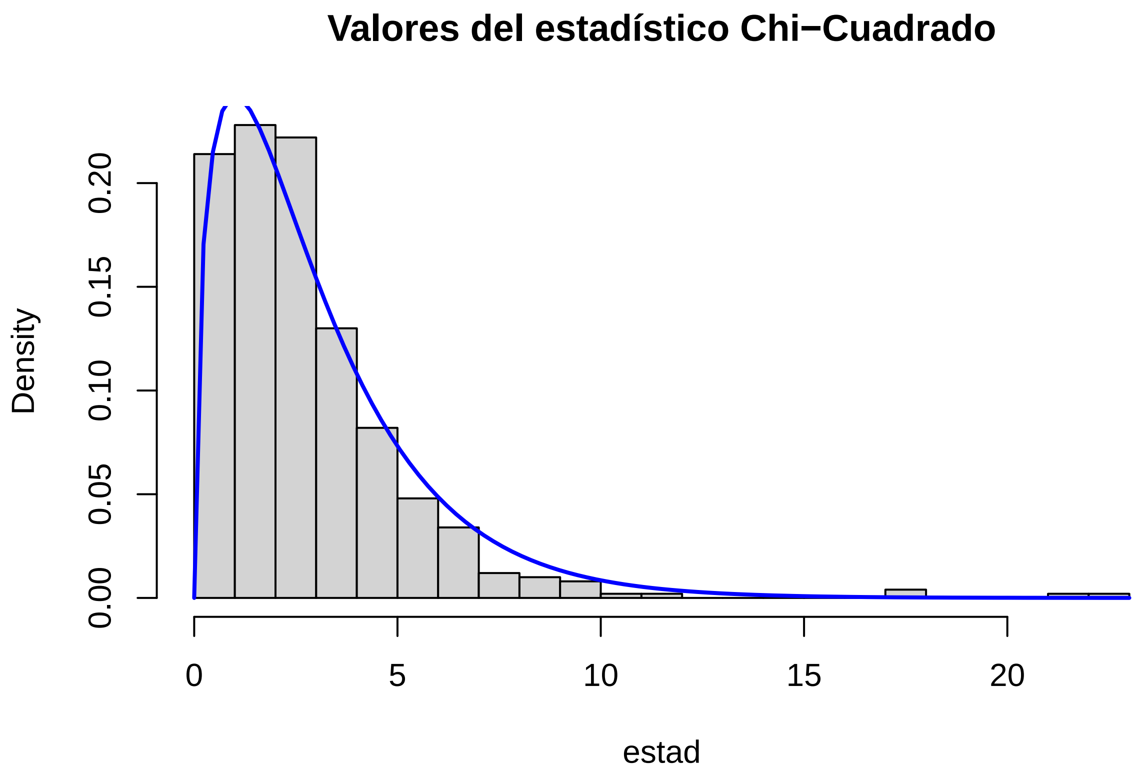
```
M<-500
estad<- numeric(M)
```

- Generar la muestra i, calcular la tabla y obtener el estadístico chi-cuadrado.
- Histograma del estadístico y densidad de la chi-cuadrado.

Solución

```
for (i in 1:M) {
  x = sample(1:4, size = n, prob = probabi, rep = T)
  res = c(sum(x==1),sum(x==2),sum(x==3),sum(x==4))
  chi = chisq.test(res, p = probabi)
  estad[i] = chi$statistic
}
```

```
graphics::hist(estad, breaks = 30, probability = TRUE,
     main = "Valores del estadístico Chi-Cuadrado")
graphics::curve(dchisq(x,length(probabi)-1),col = "blue",lwd = 2,
      add=TRUE)
```



Ejercicio 4

Tests de independencia en tablas de contingencia.

```
#save(TRABAJO,AUTOESTIMA,file="ej04.RData")
load("ej04.RData")
tabla<- table(TRABAJO,AUTOESTIMA)
tabla
```

```
##                         AUTOESTIMA
## TRABAJO                  Baja Media Alta
##   Actividad remunerada   90    65   91
##   Ama de Casa           101    76   42
```

Apartado a

Comprobación del p-valor y dibujar la distribución teórica, el cuantil y el valor observado.

Solución

La hipótesis nula es la independencia

```
resul = chisq.test(tabla)
resul
```

```
##
##  Pearson's Chi-squared test
##
## data:  tabla
## X-squared = 18.037, df = 2, p-value = 0.0001211
```

Rechazamos que las muestras sean independientes, es decir están relacionadas.

```r
resul$observed
```

```
##                        AUTOESTIMA
## TRABAJO                 Baja Media Alta
##   Actividad remunerada   90    65   91
##   Ama de Casa           101    76   42
```

```r
resul$expected
```

```
##                        AUTOESTIMA
## TRABAJO                      Baja    Media     Alta
##   Actividad remunerada 101.04516 74.59355 70.36129
##   Ama de Casa           89.95484 66.40645 62.63871
```

```r
resul$residuals
```

```
##                        AUTOESTIMA
## TRABAJO                      Baja     Media      Alta
##   Actividad remunerada -1.098789 -1.110781  2.460456
##   Ama de Casa           1.164554  1.177265 -2.607721
```

```r
sum(resul$residuals^2)
```

```
## [1] 18.03737
```

```r
sum((resul$observed-resul$expected)^2/(resul$expected))
```

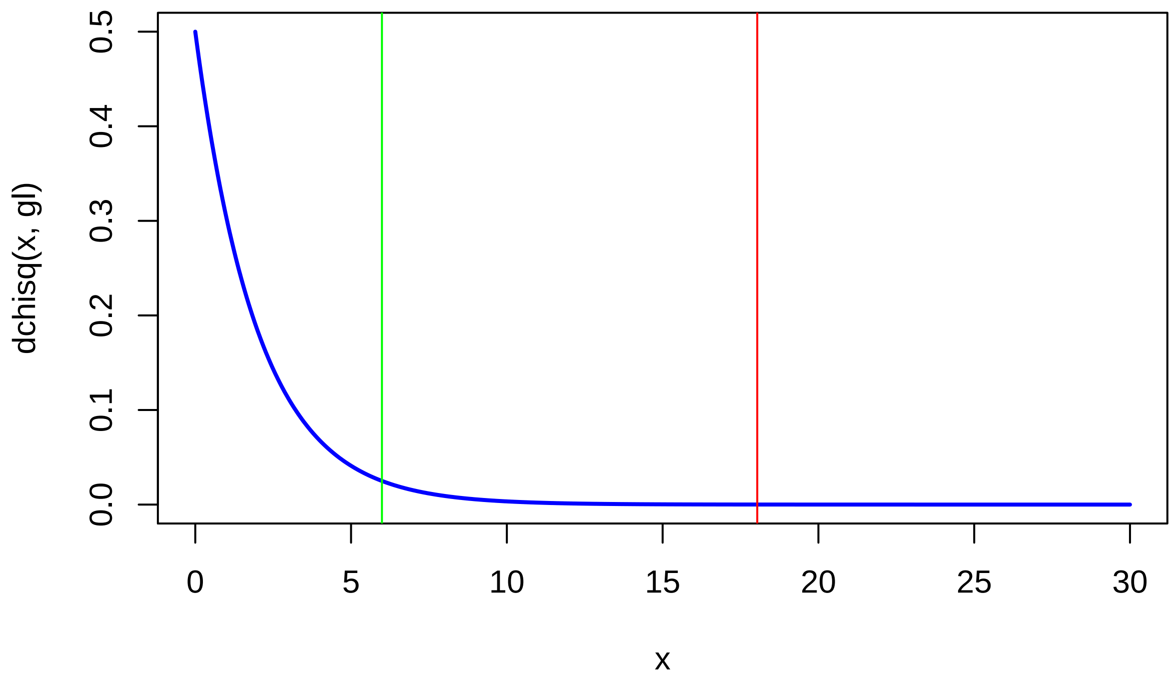
```
## [1] 18.03737
```

Para calcular el p-valor:

```r
nr = nrow(tabla)
nc = ncol(tabla)
gl = (nr-1)*(nc-1)
1-pchisq(resul$statistic, df = gl)
```

```
##    X-squared
## 0.0001211255
```

```r
curve(dchisq(x,gl),0,30,1000,lwd=2,col="blue")
abline(v=resul$statistic, col = "red")
abline(v = qchisq(0.95,gl), col = "green")
```



Ejercicio 5

Tests de independencia en tablas de contingencia (Uso de la librería vcd).

```
load("GSS.RData")
GSS
```

```
##       sex party count
## 1 mujeres   dem   279
## 2 hombres   dem   165
## 3 mujeres indep    73
## 4 hombres indep    47
## 5 mujeres   rep   225
## 6 hombres   rep   191
```

Solución

```
GSStab = xtabs(count ~ sex + party, data = GSS)
GSStab
```

```
##          party
## sex       dem indep rep
##   mujeres 279    73 225
##   hombres 165    47 191
```

```
chisq.test(GSStab)
```

```
##
##  Pearson's Chi-squared test
##
## data:  GSStab
## X-squared = 7.0095, df = 2, p-value = 0.03005
```

```r
library(vcd)
```

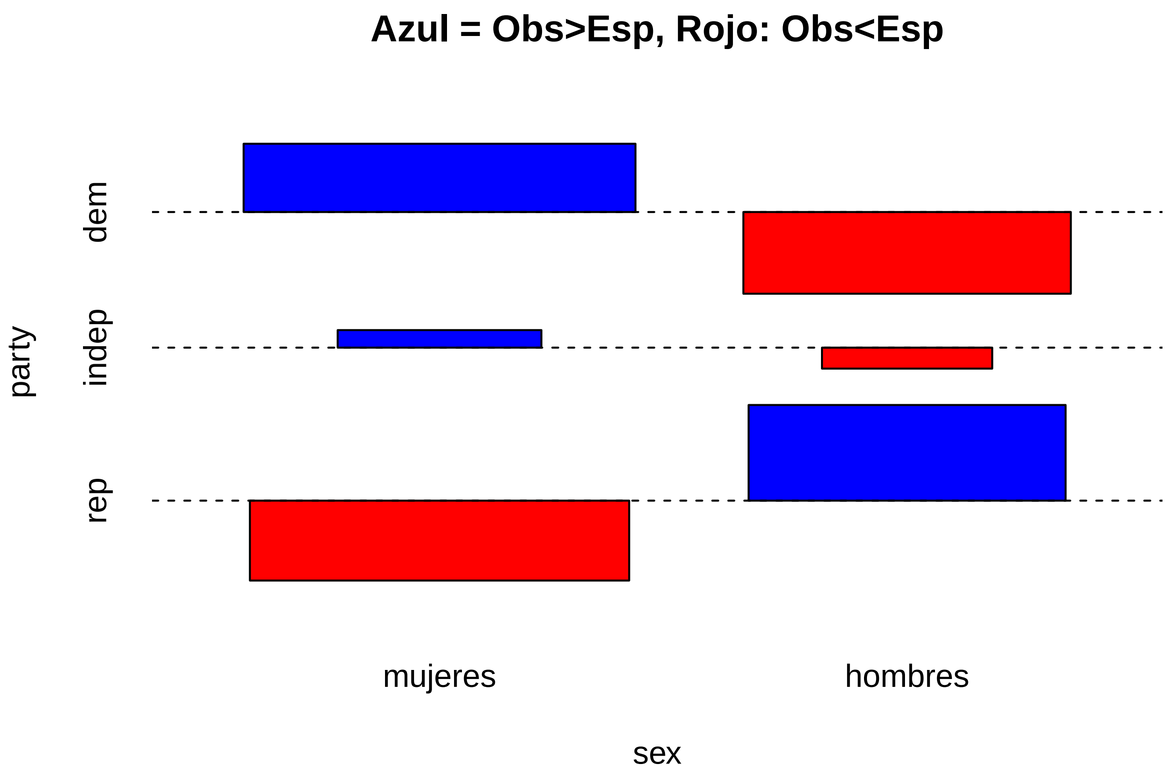
```
## Loading required package: grid
```

```r
vcd::assocstats(GSStab)
```

```
##                      X^2 df P(> X^2)
## Likelihood Ratio 7.0026  2 0.030158
## Pearson          7.0095  2 0.030054
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.084
## Cramer's V        : 0.085
```

```r
#CrossTable(GSStab)
```

```r
assocplot(GSStab,col = c("blue","red"),
          main = "Azul = Obs>Esp, Rojo: Obs<Esp")
```



Ejercicio 6

Una dama británica sostenía que era capaz de adivinar si en un té con leche se ha vertido antes el té o la leche.

Para ello se realizó un experimento donde se le pidió que lo adivinara para 8 tazas:

```
H0:P[dice leche=/leche]=P[dice té/leche]
H1:P[dice leche=/leche]>P[dice té/leche]
```

```r
pruebate <-
  matrix(c(
    3,1,1,3), nrow = 2,
    dimnames = list(Predice = c("Leche", "Té"),
```

```r
                     Verdad = c("Leche", "Té")))
pruebate
```

```
##          Verdad
## Predice Leche Té
##   Leche     3  1
##   Té        1  3
```

Solución

Probamos en primer lugar con el Test ChiCuadrado (a pesar de que las observadas no son mayores o iguales que 5):

```r
res = chisq.test(pruebate)
```

```
## Warning in chisq.test(pruebate): Chi-squared approximation may be incorrect
```

```r
res
```

```
## 
##  Pearson's Chi-squared test with Yates' continuity correction
## 
## data:  pruebate
## X-squared = 0.5, df = 1, p-value = 0.4795
```

```r
res$expected
```

```
##          Verdad
## Predice Leche Té
##   Leche     2  2
##   Té        2  2
```

Es más conveniente aplicar el test Exacto de Fisher:

```r
fisher.test(pruebate, alternative = "greater")
```

```
## 
##  Fisher's Exact Test for Count Data
## 
## data:  pruebate
## p-value = 0.2429
## alternative hypothesis: true odds ratio is greater than 1
## 95 percent confidence interval:
##  0.3135693       Inf
## sample estimates:
## odds ratio
##   6.408309
```

Ejercicio 7

Gafas y antecedentes.

```r
gafasante <-
  matrix(c(1, 8, 5, 2), nrow = 2,
         dimnames = list(Gafas = c("Sí", "No"),
                         Antecedentes = c("Sí", "No")))
gafasante
```

```
##       Antecedentes
```

```
## Gafas Sí No
##    Sí  1  5
##    No  8  2
```

Solución

Se quiere contrastar que H0: Variables categóricas independientes

```r
chisq.test(gafasante)
```

```
## Warning in chisq.test(gafasante): Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  gafasante
## X-squared = 3.8095, df = 1, p-value = 0.05096
```

También podría aplicarse el test exacto de Fisher si se plantea como un contraste de bondad de ajuste con variables categóricas con 2 modalidades.

```r
fisher.test(gafasante)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  gafasante
## p-value = 0.03497
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.0009525702 0.9912282442
## sample estimates:
## odds ratio
## 0.06464255
```

El `p-value = 0.03497` < 0.05, rechazamos la hipótesis y por tanto si existe comportamiento diferente (o están relacionadas).

Ejercicio 8

Test de McNemar (datos relacionados). Datos relacionados, por ejemplo antes-después.

Dos encuestas con un mes de separación, se pregunta a cada uno de los 1600 encuestados si aprueba o desaprueba a un gobernante.

```r
datos <- matrix(c(794, 86, 150, 570), nrow = 2,
       dimnames = list("Primera encuesta" = c("Aprueba", "Desaprueba"),
                       "segunda encuesta" = c("Aprueba", "Desaprueba")))
datos
```

```
##                 segunda encuesta
## Primera encuesta Aprueba Desaprueba
##       Aprueba        794        150
##       Desaprueba      86        570
```

Solución

Aplicamos el test de McNemar al considerar la comparación de dos muestras categóricas relacionadas

```r
mcnemar.test(datos)
```

```
##
##  McNemar's Chi-squared test with continuity correction
##
## data:  datos
## McNemar's chi-squared = 16.818, df = 1, p-value = 4.115e-05
```

El `p-value = 4.115e-05` < 0.05, por tanto rechazamos H0: "las dos muestras se comportan igual", es decir ha habido un cambio de opinión de una encuesta a otra.