

Soluciones control junio 19-20

MAT III GIN2 Ricardo Alberich

11/6/2020

Problema 5

La siguiente tabla de contingencia se ha obtenido a partir de los datos del juego FIFA 2019 de EA Sports. En ella se relacionan las variables **Club** con **Nivel**. El *nivel* de los jugadores se representa por un número entre 0 y 100 y ha sido dividido en tres grupos: (0, 50], (50, 65] y (65, 100]. En este ejercicio se trata de decidir si ambas variables son independientes realizando un test de χ^2 sobre los valores de la tabla.

```
data=read.csv("./data.csv",stringsAsFactors = FALSE)
clubs=c("FC Barcelona", "Real Madrid", "Manchester City","Paris Saint-Germain")
data2=data[data$Club%in% clubs,c("Club", "Name","Penalties", "Marking", "Interceptions")]
nivel=cut(data2$Penalties,breaks=c(0,50,65,100))
tabla_nivel=table(data2$Club,nivel)
tabla_nivel
```

```
##              nivel
##              (0,50] (50,65] (65,100]
## FC Barcelona      12      9      12
## Manchester City   13     10     10
## Paris Saint-Germain 12     10      8
## Real Madrid       16      9      8
```

La tabla de valores esperados calculada a partir de los valores anteriores es:

```
      (0,50] (50,65] (65,100]
[1,]      X   9.720930 9.720930
[2,] 13.55814      X   9.720930
[3,] 12.32558 8.837209 8.837209
[4,] 13.55814 9.720930 9.720930
```

Y el resultado del test de χ^2 es:

Pearson's Chi-squared test

```
data:  tabla_nivel
X-squared = 1.8447, df = X, p-value = XXXXX
% «» = % chisq.test(tabla_nivel) % @
```

Se pide:

- Completar los valores que faltan (marcados con X) en la tabla de valores esperados. (1 punto.)
- Calcular el p-valor del contraste y decidir si se puede aceptar o no la hipótesis de independencia entre las variables. (1 punto.)

Solución

```
frec_obs=tabla_nivel
frec_obs
```

```
##              nivel
##      (0,50] (50,65] (65,100]
## FC Barcelona      12      9      12
## Manchester City    13     10     10
## Paris Saint-Germain 12     10      8
## Real Madrid       16      9      8
```

```
nclub=rowSums(tabla_nivel)
nnivel=colSums(tabla_nivel)
n=sum(tabla_nivel)
frec_esp=nclub%*%t(nnivel)/n
frec_esp
```

```
##      (0,50] (50,65] (65,100]
## [1,] 13.55814 9.720930 9.720930
## [2,] 13.55814 9.720930 9.720930
## [3,] 12.32558 8.837209 8.837209
## [4,] 13.55814 9.720930 9.720930
```

Para calcular el estadístico manualmente

```
X2=sum((frec_obs-frec_esp)^2/frec_esp)
X2
```

```
## [1] 1.844687
```

y el p -valor del contraste

```
1-pchisq(X2,(3-1)*(4-1))
```

```
## [1] 0.933414
```

```
pchisq(X2,(3-1)*(4-1),lower.tail =FALSE)
```

```
## [1] 0.933414
```

```
chisq.test(frec_obs)
```

```
##
## Pearson's Chi-squared test
##
## data:  frec_obs
## X-squared = 1.8447, df = 6, p-value = 0.9334
```