Postwork 4

Orlando Antonio Aguilar Puerto Mario Alberto Encinas Cardona
Emanuel Flores Martínez Andrés Benjamín Sánchez Alvarado

16/01/2021

Desarrollo

Ahora investigarás la dependencia o independencia del número de goles anotados por el equipo de casa y el número de goles anotados por el equipo visitante mediante un procedimiento denominado bootstrap, revisa bibliografía en internet para que tengas nociones de este desarrollo.

1.Ya hemos estimado las probabilidades conjuntas de que el equipo de casa anote X=x goles (x=0,1,. . . ,8), y el equipo visitante anote Y=y goles (y=0,1,. . . ,6), en un partido. Obtén una tabla de cocientes al dividir estas probabilidades conjuntas por el producto de las probabilidades marginales correspondientes.

Librerías

```
library(knitr)
library(ggplot2)
library(dplyr)
library(rsample)
library(normtest)
require(nortest)
```

```
base<-basepostwork2
base<-base[,c("FTHG","FTAG")] # extraemos las dos columnas de interés
df.home<-base$FTHG
df.visit<-base$FTAG


prob.home<-prop.table(table(base$FTHG)) # Probabilidades marginales del equipo de casa
prob.visit<-prop.table(table(base$FTAG)) # Probabilidades marginales del equipo visitante.
conjunta<-prop.table(table(base$FTHG,base$FTAG)) # Probabilidad conjunta

cocientes<-conjunta
```

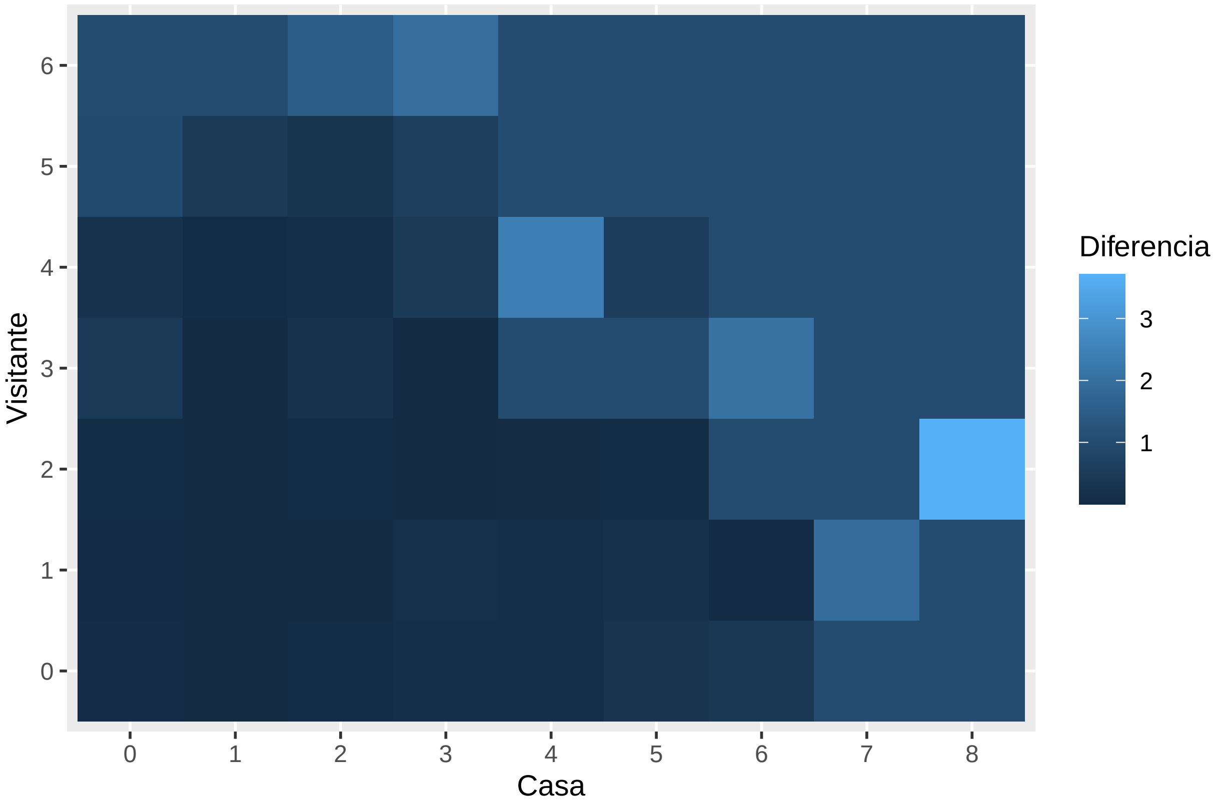
Calculamos la tabla de cocientes:

```
for(i in 1:9){ for(j in 1:7){
cocientes[i,j]<-cocientes[i,j]/(prob.home[i]*prob.visit[j]) }}
```

Table 1: Casa/Visitante

	0	1	2	3	4	5	6
0	0.9547829	1.0200350	0.9243724	1.4570907	0.7821612	1.9554031	0.00000
1	1.0060639	1.0318952	0.9850885	0.9859033	0.9261516	0.5556910	0.00000
2	0.9351621	1.0341495	1.0847107	0.7862903	1.1363636	0.6818182	2.50000
3	1.1327151	0.8493073	1.0304752	1.0055444	0.5397727	1.6193182	2.96875
4	1.1371571	0.8814433	0.9421488	0.0000000	3.4545455	0.0000000	0.00000

	0	1	2	3	4	5	6
5	1.2922240	0.8013121	1.0706236	0.0000000	1.5702479	0.0000000	0.00000
6	1.4214464	0.9793814	0.0000000	3.0645161	0.0000000	0.0000000	0.00000
7	0.0000000	2.9381443	0.0000000	0.0000000	0.0000000	0.0000000	0.00000
8	0.0000000	0.0000000	4.7107438	0.0000000	0.0000000	0.0000000	0.00000



Heatmap de diferencias absolutas entre cocientes y 1; entre más obscuro más cercano a 1

2. Mediante un procedimiento de boostrap, obtén más cocientes similares a los obtenidos en la tabla del punto anterior. Esto para tener una idea de las distribuciones de la cual vienen los cocientes en la tabla anterior. Menciona en cuáles casos le parece razonable suponer que los cocientes de la tabla en el punto 1, son iguales a 1 (en tal caso tendríamos independencia de las variables aleatorias X y Y).

Observando la tabla de distribución conjunta, decidimos reducirla para solo tomar los valores de x,y=0,1,2,3,4, ya que las demás parejas de resultados son muy poco frecuentes.Se considerarán como atípicos.

```
conjunta<-conjunta[1:5,1:5]
```

Table 2: Distribución conjunta reducida: Casa/Visitante

	0	1	2	3	4
0	0.0780702	0.0807018	0.0456140	0.0184211	0.0052632
1	0.1157895	0.1149123	0.0684211	0.0175439	0.0087719
2	0.0877193	0.0938596	0.0614035	0.0114035	0.0087719
3	0.0447368	0.0324561	0.0245614	0.0061404	0.0017544
4	0.0140351	0.0105263	0.0070175	0.0000000	0.0035088

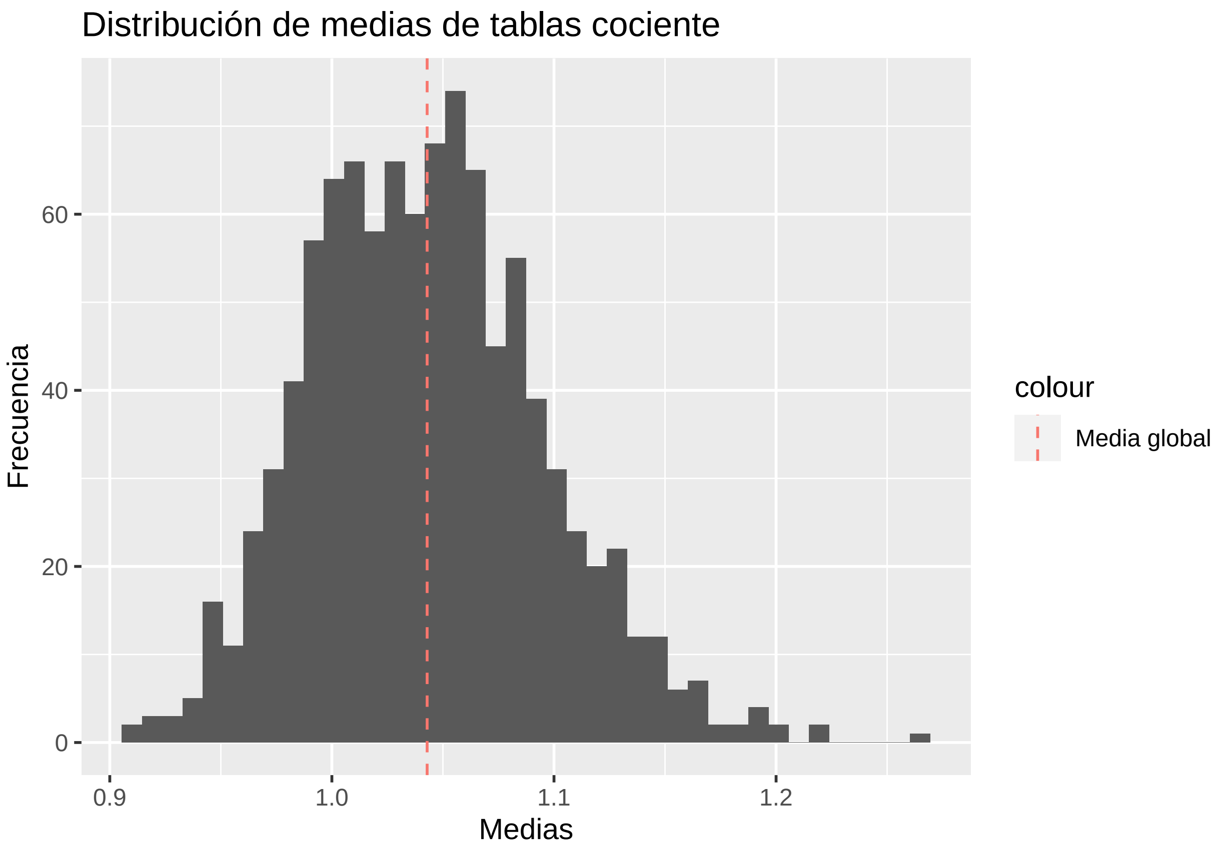
```
cocientes_list<-list() # Creamos una lista para guardar cada table de cocientes.
cocientes<-matrix(0,5,5) # Redefinimos la tabla como una matriz de 5x5
medias<-0 # Creamos un vector para guardar las medias de cada tabla de cocientes.
```

Calculamos 1000 tablas de cocientes por medio de un reemuestreo con reemplazo

```
for(n in 1:1000){

resample<-sample(1:nrow(base),nrow(base),replace = TRUE)


conjunta<-prop.table(table(df.home[resample],df.visit[resample]))

conjunta<-conjunta[1:5,1:5]

prob.home<-prop.table(table(df.home[resample]))
prob.home<-prob.home[1:5]

prob.visit<-prop.table(table(df.visit[resample]))
prob.visit<-prob.visit[1:5]

colnames(cocientes)<-c(0,1,2,3,4)
rownames(cocientes)<-c(0,1,2,3,4)

# Calculamos la tabla de cocientes
for(i in 1:length(prob.home)){
  for(j in 1:length(prob.visit)){

    cocientes[i,j]<-conjunta[i,j]/(prob.home[i]*prob.visit[j])

  }
}
medias[n]<-mean(cocientes)
cocientes_list[[n]]<-cocientes
}
```

Para determinar la independencia de las variables, tomaremos el promedio de la tabla de cocientes como estadístico.

```
medias<-as.data.frame(medias)
ggplot(medias) + geom_histogram(aes(x=medias),bins=40) + geom_vline(aes(xintercept = mean(medias),color=
```



Para finalmente determinar si las variables son independientes, recurriremos a una prueba de hipótesis.

Realizamos pruebas de normalidad de nuestra poblacion de medias para corroborar los supuestos:

```
norm.tests<-0
ad<-ad.test(medias$medias)
li<-lillie.test(medias$medias)
pear<-pearson.test(medias$medias)
sw<-shapiro.test(medias$medias)
ks<-ks.test(medias$medias,'pnorm')

norm.tests<-c(ad$p.value,li$p.value,pear$p.value,sw$p.value,ks$p.value)



names(norm.tests)<-c("Anderson Darling","Lilliefors","Pearson","Shapiro'Wilks","Kolmogorov-Smirnof")
norm.tests<-as.data.frame(norm.tests)
names(norm.tests)<-"p-valores"
kable(norm.tests,caption="p-valores de pruebas de normalidad") #Todas las pruebas rechazan normalidad,
```

Table 3: p-valores de pruebas de normalidad

	p-valores
Anderson Darling	0.0003386
Lilliefors	0.0194170
Pearson	0.1840457
Shapiro'Wilks	0.0000124

	p-valores
Kolmogorov-Smirnof	0.0000000

Considerando una significancia del 5%, todas las pruebas rechazan la normalidad de las medias, ya que todos los p-valores son menores a .05.

Por lo tanto, se realiza la prueba no paramétrica de Wilcoxon, ya que no necesita de supuestos de distribución y además es muy útil cuando se tienen datos con distribución simétrica.

```
wilcox.test(medias$medias,alternative = "two.sided",mu=1)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  medias$medias
## V = 440869, p-value < 2.2e-16
## alternative hypothesis: true location is not equal to 1
```

Con un p-valor muy pequeño, se rechaza la hipotesis de que nuestra media es igual a 1, por lo que rechazamos la independencia.

Esto coincide con la naturaleza de las variables. Durante un partido, influye mucho la marcación del un equipo con el del otro. Si un equipo lleva una ventaja considerable (de 2 o mas goles), la estrategia más común es jugar a la defensiva y no seguir acumulando puntos.Otro ejemplo es la moral; cuando un equipo va perdiendo por goleada, es común que su moral baje y con ello el rendimiento general del equipo, lo que termina en más anotaciones por parte del equipo con ventaja.