

Exercicis pràctic. Test Permutacions 17-18. Pearson correlation coefficient. Calculation on 'Law School' data

Jordi Ocaña, Sergi Civit,

28 de febrer de 2018

1 Pearson correlation coefficient

1.1 Enunciat de la Situació Experimental

The sample is about 82 law schools. Each faculty has the LSAT and GPA average rating of candidates for admission. LSAT (Law School Admission Test) is the score of tests made by accredited evaluation centers, independent of the university system. GPA (Grade Point Average) is a score that assesses the record of each student in their pre-university studies. The data will be organized into a data.frame of 82 rows for law schools and 2 columns for variables LSAT and GPA.

Possiblement, la hipòtesi a demostrar (alternativa) més lògica és que hi ha una dependència (correlació o el criteri de dependència que es vulgui utilitzar) positiva: si és que aquestes qualificacions tenen sentit, a millor expedient (GPA) seria d'esperar millor qualificació LSAT, i viceversa: $H_1: \rho > 0$

```
> lawSchool <- read.table(file="Law_School.txt", header=T)
> x <- lawSchool[,1]
> y <- lawSchool[,2]
> # coeficiente de correlación para los datos originales:
> r <- cor(x,y)
> r
```

```
[1] 0.7599979
```

Enumerar totes les possibles permutacions d'una de les columnes seria computacionalment complex:

```
> factorial(length(x))
```

```
[1] 4.753643e+122
```

1.2 Permutacions de Monte Carlo

És més viable un enfoc de test de permutaciones de Monte Carlo.

```
> # Permutació de la segunda columna:
> sample(y)

[1] 3.15 3.47 2.98 3.23 3.07 3.08 3.22 3.02 2.88 3.39 3.16 3.20 3.03 3.23 3.30
[16] 3.15 3.11 2.97 3.40 3.15 3.23 3.37 3.11 3.11 3.19 3.16 3.17 3.39 3.30 3.38
[31] 3.15 3.44 3.05 2.74 2.92 3.01 3.41 3.20 3.19 3.13 3.45 2.91 3.13 3.02 3.17
[46] 2.96 3.09 3.19 3.24 3.16 2.76 2.99 3.04 2.91 3.00 3.10 3.50 3.21 3.36 2.83
[61] 3.03 3.43 3.28 3.12 3.03 3.33 2.93 3.27 3.01 2.92 3.01 2.81 3.36 2.57 3.33
[76] 3.29 3.24 3.01 3.32 3.24 2.85 2.82

>
```

1.3 Estadistic Coef de correlació amb Permutacions de Monte Carlo

Fem SUBCONJUNT (19999) de TOTES LES REMOSTRES i els seus corresponents càlculs del coeficient de correlació

```
> # 'nperm' permutacions
> set.seed(2431)
> nperm <- 19999
> r.perm <- replicate(nperm, cor(x, sample(y)))
> #Veiem les 20 primeres
> r.perm[1:20]

[1] -0.13057553 0.09304960 0.09541940 -0.03713743 0.17130400 -0.07036551
[7] -0.10276415 0.04487481 -0.11764315 -0.09040730 0.11969299 0.07089190
[13] -0.11710148 -0.03388741 0.06075251 0.16165550 -0.07496970 -0.00135335
[19] -0.04438227 -0.10171467
```

1.4 Calculem EL P-VALOR

```
> # Significació de la correlació:
> p.valor <- (sum(r.perm > r) + 1) / (nperm + 1)
> p.valor

[1] 5e-05

> # p.valor significatiu per el test H0: estoc. indep. vs. H1: rho > 0
>
> # No importa respecte de quina columna permutem
>
> set.seed(2431)
> r.perm <- replicate(nperm, {
```

```

+     cor(sample(x),y)
+   }
+ )
> # Significació de la correlació:
> p.valor <- (sum(r.perm > r) + 1) / (nperm + 1)
> p.valor

[1] 5e-05

```

1.5 Estadístic MODIFICACIÓ Coef de correlació amb Permutacions de Monte Carlo

Es POT DUR A TERME UNA MODIFICACIÓ de l'estadístic correlació de Pearson

```

> # En lugar de la correlación se podría emplear la suma de productos
> # Estadístico sobre la muestra original:
> sumXY <- sum(x * y)
> # Permutaciones aleatorias:
> set.seed(2431)
> nperm <- 19999
> sumXY.perm <- replicate(nperm, {
+   sum(x * sample(y))
+ }
+ )
> p.valor <- (sum(sumXY.perm > sumXY) + 1) / (nperm + 1)
> p.valor

[1] 5e-05

```

1.6 Estadístic MODIFICACIÓ Coef de correlació amb Permutacions de Monte Carlo. Test BILATERAL

```

> # Per al test bilateral seria millor sumXY centrada:
>
> max.sumXY = sum(sort(x) * sort(y))
> min.sumXY = sum(sort(x) * sort(y, decreasing = TRUE))
> mean.sumXY = (min.sumXY + max.sumXY) / 2
> sumXY <- sum(x * y) - mean.sumXY
> # Permutacions aleatòries:
> set.seed(2431)
> nperm <- 19999
> sumXY.perm <- replicate(nperm, {
+   sum(x * sample(y)) - mean.sumXY
+ }
+ )

```

```
> p.valor <- (sum(abs(sumXY.perm) > abs(sumXY)) + 1) / (nperm + 1)
> p.valor

[1] 5e-05
>
```