

Práctico 10

Ejercicio 2: Correlación de pearson

```
# Cargo los datos del práctico
datos <- c(74.3, 74.1, 75.4, 67.4, 69.3, 70.5, 70.1, 69.9, 68.7, 70.3, 70.7, 71.1, 74.4, 70.2)
# Creo un nuevo vector y almaceno los datos ordenados
datos_ordenados <- sort(datos)
# El largo del vector es el n
n <- length(datos)
```

Calculamos S utilizando $S = \sum_{k=1}^n (R_k - k)^2$. Por ejemplo para el sumando de $n=1$ va a ser $(1 - 12)^2$. El resultado de eso lo guardo en un vector `vector_S` y al final sumo todos sus elementos para obtener S .

```
vector_S <- c()
for (i in 1:14) {
  vector_S <- c(vector_S, (i-rank(datos)[i])^2)
}
S <- sum(vector_S)
```

El valor de $S = 490$. Ahora con el resultado de S , calculo el $\rho_S = 1 - \frac{6S}{n(n^2-1)}$

```
rho <- 1-((6*S)/(n*(n^2-1)))
z1 <- sqrt(n-1)*rho
```

El resultado de $\rho_S = -0.077$. El siguiente paso es hallar la región crítica para $\alpha = 0.05$.

Como tenemos un $n < 30$ no podemos utilizar $\sqrt{n-1}\rho_S \approx \mathcal{N}(0,1)$. Si fuese el caso tendríamos $\sqrt{n-1}\rho_S = Z_{1-\frac{\alpha}{2}}$ y este valor $Z_{1-\frac{\alpha}{2}}$ se puede obtener mediante R utilizando `qnorm(1 - $\frac{\alpha}{2}$)`.

Pero si no se utilizaría esto, usaríamos la tabla de los valores críticos con un $n = 14$ y $\alpha = 0.01$. Así obtendríamos un valor C tal que

$$RC = \{|\rho_s| > c\} = 0.01$$

En la tabla buscamos el valor c antes descripto y obtenemos $c = 0.497$.

Podemos ver que $RC = [-1; -0.497] \cup [1; 0.497]$, en nuestro caso que obtuvimos un $\rho_S = -0.077$ podemos decir que este **no** cae en la región crítica, por lo tanto no se rechaza la hipótesis de que los datos son *i.i.d*

Test de spearman auspiciado por R.

```
cor.test(datos,datos_ordenados,method="spearman")
```

```
##
## Spearman's rank correlation rho
##
## data:  datos and datos_ordenados
```

```
## S = 490, p-value = 0.7966
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.07692308
```

Ejercicio 3

```
datos3 <- c(22.67, 21.66, 16.31, 15.95, 15.15, 17.88, 23.17, 24.85, 15.17, 23.19, 21.21, 20.60,
  ↪ 17.44, 23.33, 17.63, 22.54, 21.60, 17.14, 21.02, 21.05)
# Valor del test de R
cor.test(datos3, sort(datos3), method="spearman")
```

```
##
## Spearman's rank correlation rho
##
## data:  datos3 and sort(datos3)
## S = 1244, p-value = 0.7869
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.06466165
```

```
# Y si pruebo en hacerlo a mano:
datos_ordenados3 <- sort(datos3)
n3 <- length(datos3)
vector_S3 <- c()
for (i in 1:n3) {
  vector_S3 <- c(vector_S3, (i-rank(datos3)[i])^2)
}
S3 <- sum(vector_S3)
rho3 <- 1-((6*S)/(n3*(n3^2-1)))
```

Ejercicio 4

Parte b

Estudiar la aleatoriedad de las muestras

```
muestraA <- c(1.52, 2.92, 4.44, 4.24, 1.72, 3.70, 3.64, 4.82, 2.12, 2.08)
muestraB <- c(2.08, 3.03, 0.80, 0.96, 2.71, 2.39, 3.07, 2.87, 0.33, 1.76)
n4 <- length(muestraA)
#Para cada muestra por separado
vector_S4a <- c()
vector_S4b <- c()
for (i in 1:n4) {
  vector_S4a <- c(vector_S4a, (i-rank(muestraA)[i])^2)
  vector_S4b <- c(vector_S4b, (i-rank(muestraB)[i])^2)
}
S4a <- sum(vector_S4a)
S4b <- sum(vector_S4b)
rho4a <- 1-((6*S4a)/(n4*(n4^2-1)))
rho4b <- 1-((6*S4b)/(n4*(n4^2-1)))
```

$$RC = \{|\rho_s| > c\} = 0.01$$

Buscamos este valor c en la tabla y obtenemos que $c = 0.618$. Por lo tanto $RC = [-1; -0.618] \cup [1; 0.618]$ Vemos que $\rho_S^a = 0.091$ y que $\rho_S^b = -0.091$ vemos que ambos valores de ρ_S no caen en la región crítica, por lo que no se rechaza H_0 , es decir, ambas muestras son *i.i.d.*

Ahora utilizando `cor.test()`

```
cor.test(muestraA,sort(muestraA),method="spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: muestraA and sort(muestraA)
## S = 150, p-value = 0.8114
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.09090909
```

```
cor.test(muestraB,sort(muestraB),method="spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: muestraB and sort(muestraB)
## S = 180, p-value = 0.8114
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.09090909
```

Parte b

Pasamos a testear si las muestras son independientes entre sí.

```
# Para dos muestras
vector_S4 <- c()
for (i in 1:n4) {
  vector_S4 <- c(vector_S4, (rank(muestraA)[i]-rank(muestraB)[i])^2)
}
S4 <- sum(vector_S4)
rho4 <- 1-((6*S4)/(n4*(n4^2-1)))
```

Ahora queda calcular la región crítica, al tener un $n < 30$ tenemos que usar la tabla. Es el mismo procedimiento que usamos para calcular la región crítica para una muestra.

$$RC = \{|\rho_s| > c\} = 0.01$$

Buscamos este valor c en la tabla y obtenemos que $c = 0.618$. Por lo tanto $RC = [-1; -0.618] \cup [1; 0.618]$ $\rho_S = 0.018$ vemos que no cae en la región crítica, por lo que no se rechaza la hipótesis nula. Las muestras presentan independencia.

Ahora calculamos la independencia en R, también utilizando `cor.test`

```
cor.test(muestraA, muestraB, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: muestraA and muestraB
## S = 162, p-value = 0.9728
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.01818182
```

Ejercicio 5

Se hizo en teórico-práctico. # Ejercicio 6

```
datos6 <- as.table(matrix(c(15,20,25,35,10,5), ncol=3))
chisq.test(datos6)
```

```
##
## Pearson's Chi-squared test
##
## data: datos6
## X-squared = 3.1647, df = 2, p-value = 0.2055
```

```
qchisq(0.95,2)
```

```
## [1] 5.991465
```

Sobre el p-valor

Si el p-valor da un numero grande quiere decir que ocupa mucha área y que, por lo tanto, tiene mayor prbabilidad. En conclusión a $p\text{-valor} > \alpha$ no rechazo.