Tarea 1 Metodos no Paramétricos
github:

Rudy Miranda

abril, 2023

Índice

1. Introducción **1**

2. Análisis Exploratorio **1**

2.1. Variables Categoricas 1

2.2. Variables Cuantitativas 2

3. Test de Normalidad Variable Capitalización **2**

3.1. Test de Kolmogorov-Smirnov 2

4. Test de Normalidad Variable Tiempo **3**

5. Capitalización Mediana de Empresas Nacionales **3**

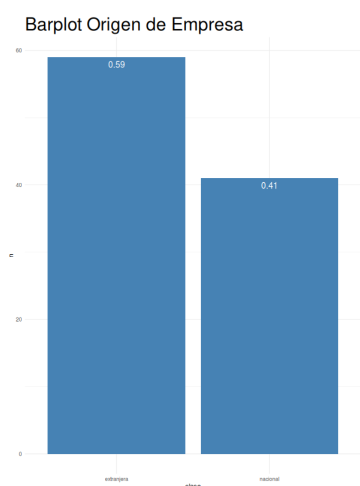
1. Introducción

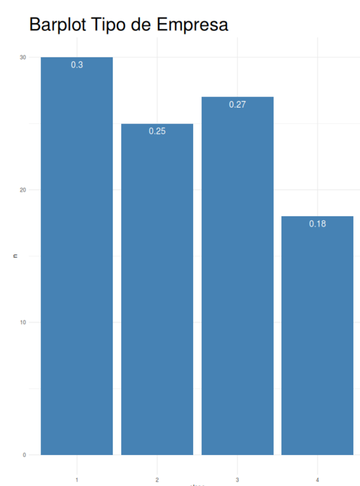
Se nos presenta una pequena base de datos referente a 100 empresas, las cuales presentan 4 variables cada una; 2 cualitativas (origen y tipo), y las restantes cuantitativas (tiempo y capitazación).

Luego de un análisis exploratorio realizaremos dos pruebas no paramétricas. La primera sera la prueba de Kolmogorov-Smirnov para el análisis de normalidad, seguido de una prueba de signos para una mediana propuesta.

2. Análisis Exploratorio

2.1. Variables Categoricas

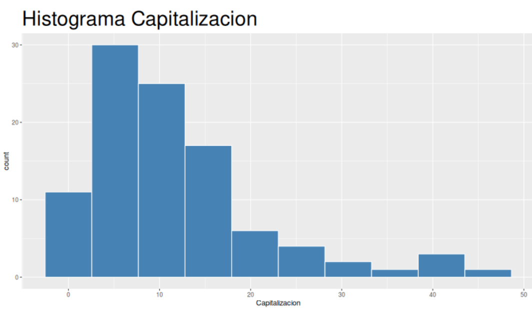


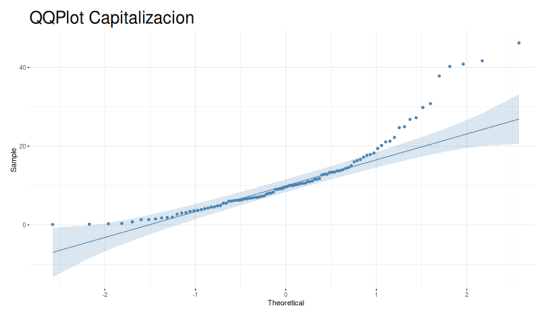


2.2. Variables Cuantitativas

	Tiempo	Capitalización
mean	3.176	11.695
sd	3.821	9.658
min	0	0.1
max	18.4	46.18
skewness	1.898	1.587
kurtosis	6.61	5.543

3. Test de Normalidad Variable Capitalización





En primera instancia notamos que tanto por la kurtosis y coeficiente de asimetría, ademas del histograma y el *QQplot*, que no estamos en presente de una variable aleatoria que se distribuya normal.

Para formalizar esta afirmación la respaldaremos el test de normalidad no paramétrico de Kolmogorov-Smirnov.

3.1. Test de Kolmogorov-Smirnov

Proponemos las hipótesis

$$H_0 : X_{\text{Capitalización}} \sim N(\mu, \sigma^2) \quad \text{vs} \quad H_1 : X_{\text{Capitalización}} \nsim N(\mu, \sigma^2)$$

el estadístico de prueba en este caso es

$$D = \max_{1 \leq i \leq n} = \left\{ D^+, D^- \right\} \tag{1}$$

Donde los

$$D^+ = \left| \frac{i}{n} - F_0(x_i) \right| \tag{2}$$

$$D^- = \left| F_0(x_i) - \frac{i-1}{n} \right| \tag{3}$$

nuestra region de rechazo para D seria

$$]D_\alpha, +\infty[\tag{4}$$

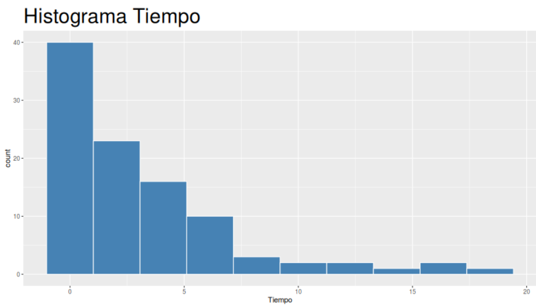
Donde, para un nivel de confianza del 0.95, y nuestra cantidad de datos (100)

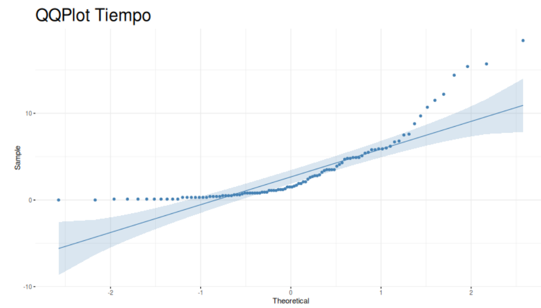
$$D_\alpha = \frac{C_\alpha = 0{,}895}{K(n) = \sqrt{100} - 0{,}01 + \frac{0{,}85}{\sqrt{100}}} \tag{5}$$

$$= 0{,}089 \tag{6}$$

Con la función nativa de R, *ks.test*, obtenemos que el valor del estadístico $D = 0{,}146$. Por lo anterior rechazamos H_0, ya que pertenece a la region de rechazo.

4. Test de Normalidad Variable Tiempo





En este caso seremos mas breves, ya que el procedimiento es el mismo que con la variable anterior.

Proponemos las siguientes hipótesis

$$H_0 : X_{\text{Tiempo}} \sim N(\mu, \sigma^2) \;\; \text{vs} \;\; H_1 : X_{\text{Tiempo}} \not\sim N(\mu, \sigma^2)$$

En este caso el valor de nuetro estadístico $D = 0{,}203$, y el valor critico $D_\alpha = 0{,}089$. Nuevamente se rechaza la hipotesis nula, al ser $D > D_\alpha$.

5. Capitalización Mediana de Empresas Nacionales

Proponemos las hipótesis

$$H_0 : m = 10{,}5 \text{ vs } H_1 : m \neq 10{,}5$$

donde m corresponde a la mediana poblacional.

El test a usar sera el de los signos, en el cual el estadístico de prueba es

$$r = \max_{1 \leq i \leq n} \left\{ r^+, r^- \right\} \tag{7}$$

Donde

$$\begin{align} r^+ &= \text{cantidad de observación por sobre la mediana propuesta} \tag{8} \\ r^- &= \text{cantidad de observación bajo la mediana propuesta} \tag{9} \\ n &= \text{cantidad de observaciones}(41) \tag{10} \end{align}$$

un inconveniente con este test es que require descartar las observaciones iguales a la mediana propuesta, pero no es el caso de nuestros datos, por lo que no debemos disminuir nuetro tamano muestral.

Al calcular r^+ y r^- (18 y 23 respectivamente), obtenemos el valor de nuestro estadístico $r = 23$.

Finalmente podemos conocer el valor-p, ya que sabemos que $r \sim B(41, 0{,}5)$, entonces

$$p = 2 * P(r > 23) = 0{,}53 \tag{11}$$

Dado que $p > \alpha = 0{,}05$, no hay evidencia suficiente para rechazar la hipótesis nula.