**[Ejercicio Propuesto 2 (Grado en Psicología)](https://wpd.ugr.es/~bioestad/wp-content/uploads/TEMA2_AP21-1.pdf)**

**Un estudio con 30 sujetos se investigaron las siguientes variables sexo (1: hombre; 2: mujer), Edad, Estado civil (1: soltero; 2: casado; 3: otro) y Cociente Intelectual (CI).**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Suj** | **sexo** | **edad** | **EC** | **CI** |  | **Suj** | **Sexo** | **edad** | **EC** | **CI** |
| 1 | 1 | 16 | 1 | 100 |  | 16 | 1 | 15 | 1 | 101 |
| 2 | 2 | 17 | 1 | 103 |  | 17 | 1 | 16 | 1 | 102 |
| 3 | 1 | 18 | 1 | 107 |  | 18 | 1 | 16 | 1 | 102 |
| 4 | 2 | 19 | 1 | 110 |  | 19 | 1 | 17 | 1 | 103 |
| 5 | 1 | 16 | 1 | 99 |  | 20 | 1 | 17 | 1 | 103 |
| 6 | 2 | 16 | 1 | 104 |  | 21 | 1 | 20 | 3 | 104 |
| 7 | 1 | 17 | 1 | 104 |  | 22 | 1 | 19 | 2 | 102 |
| 8 | 2 | 17 | 1 | 99 |  | 23 | 1 | 18 | 1 | 100 |
| 9 | 1 | 18 | 1 | 100 |  | 24 | 2 | 16 | 1 | 105 |
| 10 | 2 | 18 | 1 | 100 |  | 25 | 2 | 15 | 1 | 109 |
| 11 | 1 | 18 | 1 | 101 |  | 26 | 2 | 14 | 1 | 110 |
| 12 | 2 | 19 | 2 | 103 |  | 27 | 2 | 15 | 1 | 108 |
| 13 | 1 | 20 | 2 | 107 |  | 28 | 2 | 16 | 1 | 105 |
| 14 | 2 | 21 | 2 | 110 |  | 29 | 2 | 17 | 1 | 104 |
| 15 | 1 | 14 | 1 | 99 |  | 30 | 2 | 21 | 2 | 103 |

[Fichero de datos.txt](https://wpd.ugr.es/~bioestad/wp-content/uploads/propuesto3-1.txt)

**Se pide:**

1. **Defina las variables del estudio, su nivel de medida, su tipo y sus modalidades y almacena el contenido de la tabla en las variables correspondientes seleccionando el tipo de variable adecuado en cada caso.**
2. **Realice un gráfico para las variables EC y edad.**
3. **Calcule las características de la variable edad.**

**Solución:**

En primer lugar, si no tenemos instalado el paquete BrailerR, lo instalamos mediante la orden:

> install.packages("BrailleR")

**Recordar que cada vez que iniciemos R hay que** **conectar la biblioteca a nuestro espacio de trabajo mediante la sentencia**

> library(“BrailleR”)

Creamos un fichero de texto que recoja la información que muestran las tablas y lo guardamos en el escritorio con el nombre *propuesto2.txt*.

Nos situamos en el directorio de trabajo mediante la orden:

> setwd("C:/Users/Usuario/Desktop/Datos")

Desde *R* leemos el fichero que se acaba de crear y almacenamos su contenido en una variable de nombre *datos*.

> datos = read.table("propuesto2.txt", header = TRUE)

> datos

Suj sexo edad EC CI

1 1 1 16 1 100

2 2 2 17 1 103

3 3 1 18 1 107

4 4 2 19 1 110

5 5 1 16 1 99

6 6 2 16 1 104

7 7 1 17 1 104

8 8 2 17 1 99

9 9 1 18 1 100

10 10 2 18 1 100

11 11 1 18 1 101

12 12 2 19 2 103

13 13 1 20 2 107

14 14 2 21 2 110

15 15 1 14 1 99

16 16 1 15 1 101

17 17 1 16 1 102

18 18 1 16 1 102

19 19 1 17 1 103

20 20 1 17 1 103

21 21 1 20 3 104

22 22 1 19 2 102

23 23 1 18 1 100

24 24 2 16 1 105

25 25 2 15 1 109

26 26 2 14 1 110

27 27 2 15 1 108

28 28 2 16 1 105

29 29 2 17 1 104

30 30 2 21 2 103

**Nota: También se puede introducir los datos definiendo cada variable mediante vectores y factores. Por ejemplo**

Las variables **Sexo y Estados civil** son variables cualitativas sin un orden específico entre sus categorías, de modo que creamos un ***factor no ordenado*** para almacenar la información de esta variable.

Es resto de las variables se almacena cada una en un ***vector*** de números reales.

Nosotros vamos a introducir los datos mediante la orden read.table

Una vez introducidos los datos, creamos las tablas de frecuencia*.* Para ello, se emplea la función ***table*** o la función ***prop.table***, dependiendo de si la tabla muestra las frecuencias absolutas o las frecuencias relativas.

> table(x)  # para frecuencias absolutas

> prop.table(tab)   # para las frecuencias relativas

**> tabla\_sexo = table(datos$sexo)**

**> tabla\_sexo**

**1 2**

**16 14**

**> prop.table(tabla\_sexo)**

**1 2**

**0.5333333 0.4666667**

**> tabla\_edad = table(datos$edad)**

**> tabla\_edad**

**14 15 16 17 18 19 20 21**

**2 3 7 6 5 3 2 2**

**> prop.table(tabla\_edad)**

**14 15 16 17 18 19 20**

**0.06666667 0.10000000 0.23333333 0.20000000 0.16666667 0.10000000 0.06666667**

**21**

**0.06666667**

**> tabla\_EC = table(datos$EC)**

**> tabla\_EC**

**1 2 3**

**24 5 1**

**> prop.table(tabla\_EC)**

**1 2 3**

**0.80000000 0.16666667 0.03333333**

**> tabla\_CI = table(datos$CI)**

**> tabla\_CI**

**99 100 101 102 103 104 105 107 108 109 110**

**3 4 2 3 5 4 2 2 1 1 3**

**> prop.table(tabla\_CI)**

**99 100 101 102 103 104 105**

**0.10000000 0.13333333 0.06666667 0.10000000 0.16666667 0.13333333 0.06666667**

**107 108 109 110**

**0.06666667 0.03333333 0.03333333 0.10000000**

**b) Realice un gráfico para las variables EC y edad**

**Gráfico de sectores**

pie(table(datos$EC), col = c("yellow", "white", "green"), main = "Diagrama de sectores para la variable Estado Civil")



**Nota: El gráfico diagrama de sectores no está implementado en BrailerR.**

**Gráfico de barras**

> barplot(table(datos$EC), col=c("yellow", "white", "green"), xlab="Color", ylab="Frecuencias absolutas",main ="Diagrama de barras para la variable Estado Civil")

> y.bar = barplot(table(datos$EC), col=c("yellow", "white", "green"), xlab="Color", ylab="Frecuencias absolutas",main ="Diagrama de barras para la variable Estado civil")

Se muestra el siguiente gráfico de barras para la variable Estado civil:



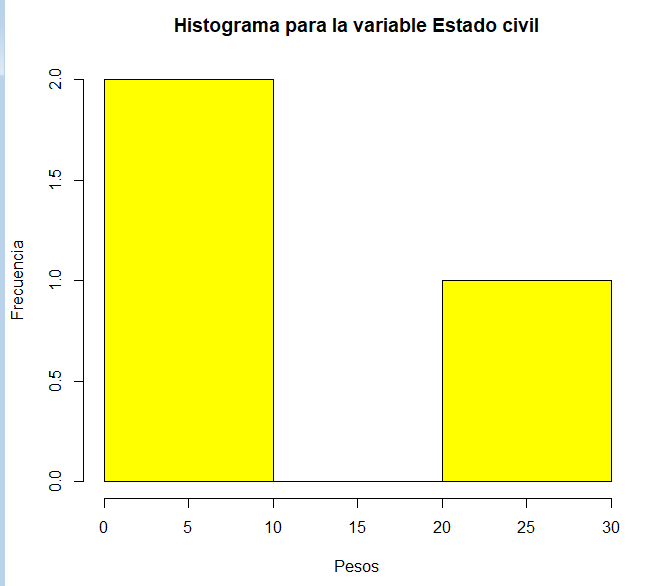
> VI(y.bar)

The summary of each variable is

V1: Min. 0.7 1st Qu. 1.3 Median 1.9 Mean 1.9 3rd Qu. 2.5 Max. 3.1

**Histograma**

y.hist = hist(table(datos$EC), col = "yellow", main = "Histograma para la variable Estado civil", xlab="Pesos", ylab="Frecuencia")



> VI(y.hist)

This is a histogram, with the title: Histogram of table(datos$EC)

"table(datos$EC)" is marked on the x-axis.

Tick marks for the x-axis are at: 0, 5, 10, 15, 20, 25, and 30

There are a total of 3 elements for this variable.

Tick marks for the y-axis are at: 0, 0.5, 1, 1.5, and 2

It has 3 bins with equal widths, starting at 0 and ending at 30 .

The mids and counts for the bins are:

mid = 5 count = 2

mid = 15 count = 0

mid = 25 count = 1

**Gráfico de tallos y hojas**

**> stem(table(datos$edad))**

**The decimal point is at the |**

**2 | 00000**

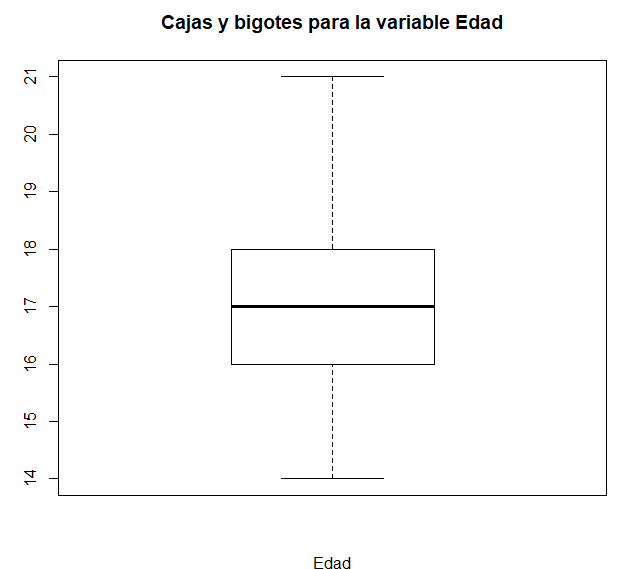
**4 | 0**

**6 | 00**

**Gráfico de caja con bigotes**

y.box = boxplot(datos$edad, xlab="Edad", main = "Cajas y bigotes para la variable Edad")

Se muestra el siguiente boxplot (Cajas y bigotes) de la variable Edad



**> VI(y.box)**

**This graph has a boxplot printed vertically**

**with the title:**

**"" appears on the x-axis.**

**"" appears on the y-axis.**

**Tick marks for the y-axis are at: 14, 15, 16, 17, 18, 19, 20, and 21**

**This variable has 30 values.**

**There are no outliers marked for this variable**

**The whiskers extend to 14 and 21 from the ends of the box,**

**which are at 16 and 18**

**The median, 17 is 50 % from the lower end of the box to the upper end.**

**The upper whisker is 1.5 times the length of the lower whisker.**

1. **Calcule las características de la variable edad.**

### ****Medidas de posición central****

**> mean(datos$edad)**

**[1] 17.2**

**> median (datos$edad)**

**[1] 17**

> min(datos$edad)

[1] 14

> max(datos$edad)

[1] 21

En cuanto a la moda, ***R*** no tiene implementada ninguna función que la calcule.

### ****> quantile(datos$edad, probs = c(0.25, 0.75))****

### ****25% 75%****

### ****16 18****

Mediante la función ***summary***, aplicada a una variable cuantitativa, R devuelve el mínimo, el máximo, la media, la mediana y los cuartiles primero y tercero de la variable.

### ****> summary(datos$edad)****

### ****Min. 1st Qu. Median Mean 3rd Qu. Max.****

### ****14.0 16.0 17.0 17.2 18.0 21.0****

### ****Medidas de dispersión:****

### Entre las medidas de dispersión más utilizadas se encuentran la ****cuasi-varianza****, la ****cuasi-desviación típica**** y el ****rango intercuartílico****, que en *R* se calculan a través de las funciones *var*, *sd* e *IQR*, respectivamente.

**> var(datos$edad, na.rm = TRUE)**

**[1] 3.544828**

**> sd(datos$edad, na.rm = TRUE)**

**[1] 1.882771**

**> IQR(datos$edad, na.rm = TRUE)**

[1] 2

### ****Medidas de forma****

Para determinar la simetría y la curtosis, es necesario instalar y cargar el paquete *e1071*.

> install.packages("e1071") # cuidado con las comillas (tecla")

> library("e1071")

**> skewness(datos$edad)**

**[1] 0.3200432**

**> kurtosis(datos$edad)**

**[1] -0.697412**

**Nota: A diferencia de la instalación, la carga de los paquetes es necesaria cada vez que se inicia una nueva sesión de R**. (mediante la orden library(“”)

### ****Análisis exploratorio unidimensional con BrailerR****

> y = datos$edad

> y

[1] 16 17 18 19 16 16 17 17 18 18 18 19 20 21 14 15 16 16 17 17 20 19 18 16 15

[26] 14 15 16 17 21

> UniDesc(y)

[Muestra la siguiente salida en HTML.](https://wpd.ugr.es/~bioestad/wp-content/uploads/Univariate-analysis-for-Y.html)

Esta salida muestra un resumen de medidas estadísticas básicas, gráficos invariantes básicos, contrastes y gráficos de normalidad y contrastes de curtosis y simetrías.