**[Ejercicio Propuesto 3 (Grado en Psicología)](https://wpd.ugr.es/~bioestad/wp-content/uploads/TEMA2_AP22.pdf)**

**Un investigador estaba interesado en conocer la relación entre frecuencia cardíaca (FC) y percepción subjetiva del esfuerzo (PSE). Para ello realizó mediciones a 30 deportistas que practican diversos deportes (individuales o de equipo) a los que con un pulsómetro estableció su FC basal (medida por la mañana) y máxima (medida en una competición) y también midió la PSE tras el partido (en una escala muy alta, alta, media, baja y muy baja). Finalmente les preguntó la edad y el sexo de cada uno de ellos. Los datos de cada sujeto se muestran a continuación.**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Suj | Sexo | Edad | FCb | FCm | PSE | Deporte |  | Suj | Sexo | Edad | FCb | FCm | PSE | Deporte |
| 1 | 1 | 22 | 66 | 148 | 2 | 1 |  | 16 | 2 | 19 | 54 | 156 | 3 | 2 |
| 2 | 1 | 24 | 56 | 137 | 1 | 1 |  | 17 | 1 | 20 | 45 | 136 | 1 | 2 |
| 3 | 1 | 25 | 64 | 168 | 4 | 1 |  | 18 | 1 | 22 | 50 | 154 | 3 | 2 |
| 4 | 1 | 23 | 63 | 170 | 5 | 1 |  | 19 | 1 | 24 | 51 | 187 | 5 | 2 |
| 5 | 2 | 19 | 62 | 165 | 4 | 1 |  | 20 | 2 | 25 | 52 | 145 | 1 | 2 |
| 6 | 1 | 18 | 62 | 168 | 3 | 1 |  | 21 | 2 | 26 | 60 | 163 | 4 | 2 |
| 7 | 2 | 20 | 61 | 165 | 3 | 2 |  | 22 | 1 | 24 | 59 | 154 | 3 | 1 |
| 8 | 2 | 22 | 60 | 160 | 3 | 2 |  | 23 | 2 | 20 | 60 | 160 | 4 | 1 |
| 9 | 2 | 24 | 58 | 145 | 2 | 2 |  | 24 | 1 | 21 | 55 | 145 | 2 | 1 |
| 10 | 1 | 25 | 57 | 156 | 4 | 2 |  | 25 | 2 | 21 | 54 | 137 | 1 | 1 |
| 11 | 2 | 27 | 67 | 189 | 5 | 2 |  | 26 | 1 | 22 | 53 | 136 | 1 | 1 |
| 12 | 1 | 24 | 65 | 182 | 4 | 2 |  | 27 | 2 | 29 | 52 | 134 | 1 | 1 |
| 13 | 2 | 23 | 78 | 190 | 5 | 2 |  | 28 | 2 | 23 | 51 | 130 | 1 | 1 |
| 14 | 2 | 22 | 45 | 135 | 1 | 1 |  | 29 | 1 | 23 | 55 | 135 | 2 | 2 |
| 15 | 2 | 28 | 54 | 140 | 2 | 1 |  | 30 | 1 | 24 | 58 | 140 | 2 | 2 |

[Fichero de datos.txt](https://wpd.ugr.es/~bioestad/wp-content/uploads/propuesto3-2.txt)

Variables y códigos: Sexo (1: hombre; 2: mujer), Edad, FCb (Frecuencia cardíaca basal), FCm (Frecuencia cardíaca máxima), PSE (Percepción Subjetiva del Esfuerzo; 1: muy baja; 2: baja; 3: media; 4: alta y 5: muy alta), Deporte (1: individual; 2: de equipo).

**Se pide:**

**a) Crea el fichero de texto que recoja la información que muestran las tablas y guárdalo en el escritorio con el nombre *propuesto3.txt*. Desde *R* lee el fichero: propuesto3.txt y llámalo datos.**

**b) Realiza una tabla de frecuencias absolutas y otra de frecuencias relativas para la variable *PSE*. Almacena las tablas anteriores en dos variables y llámalas *abso* y *rela***

**c) Representa la variable *PSE* mediante un diagrama de barras y un diagrama de sectores. Incluye un título adecuado para cada gráfico y colorea las barras y los sectores de colores diferentes**

**d) Para la variable *Edad*, realiza un histograma y un diagrama de caja y bigotes considerando la opción *range* = 1.5. Incluye un título apropiado para cada gráfico y colorea las barras del histograma de color verde. ¿Existe algún valor atípico en esta variable? Reduce el valor del argumento *range* hasta 1. ¿Varían las conclusiones?**

**e) Realiza un resumen de la variable *FCb* mediante la orden *summary*. Comprueba que las medidas que proporciona *summary* coinciden con las medidas calculadas de forma individual usando su función específica**

**f) Calcula la edad media de los individuos y proporciona, al menos, dos medidas que indiquen la dispersión de esta variable**

**g) ¿Qué variable es más homogénea: *FCb* o *FCm*?**

**h) Estudia la asimetría y la curtosis de la variable *Edad*.**

**Solución**

1. **Crea el fichero de texto que recoja la información que muestran las tablas y guárdalo en el escritorio con el nombre *propuesto3.txt*. Desde *R* lee el fichero: propuesto3.txt y llámalo datos.**

**Recordar que cada vez que iniciemos R y tengamos instalado BrailerR hay que** **cargar el paquete BrailerR a nuestro espacio de trabajo mediante la sentencia**

> library(“BrailleR”)

Nos situamos en el directorio de trabajo mediante la orden:

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

Desde *R* leemos el fichero propuesto3.txt que se acaba de crear y almacenamos su contenido en una variable de nombre *datos*.

> datos = read.table("propuesto3.txt", header = TRUE)

> datos

 Suj Sexo Edad FCb FCm PSE Deporte

1 1 1 22 66 148 2 1

2 2 1 24 56 137 1 1

3 3 1 25 64 168 4 1

4 4 1 23 63 170 5 1

5 5 2 19 62 165 4 1

6 6 1 18 62 168 3 1

7 7 2 20 61 165 3 2

8 8 2 22 60 160 3 2

9 9 2 24 58 145 2 2

10 10 1 25 57 156 4 2

11 11 2 27 67 189 5 2

12 12 1 24 65 182 4 2

13 13 2 23 78 190 5 2

14 14 2 22 45 135 1 1

15 15 2 28 54 140 2 1

16 16 2 19 54 156 3 2

17 17 1 20 45 136 1 2

18 18 1 22 50 154 3 2

19 19 1 24 51 187 5 2

20 20 2 25 52 145 1 2

21 21 2 26 60 163 4 2

22 22 1 24 59 154 3 1

23 23 2 20 60 160 4 1

24 24 1 21 55 145 2 1

25 25 2 21 54 137 1 1

26 26 1 22 53 136 1 1

27 27 2 29 52 134 1 1

28 28 2 23 51 130 1 1

29 29 1 23 55 135 2 2

30 30 1 24 58 140 2 2

**b) Realiza una tabla de frecuencias absolutas y otra de frecuencias relativas para la variable *PSE*. Almacena las tablas anteriores en dos variables y llámalas *abso* y *rela***

> abso = table(datos$PSE)

> abso

1 2 3 4 5

8 6 6 6 4

> rela = prop.table(abso)

> rela

 1 2 3 4 5

0.2666667 0.2000000 0.2000000 0.2000000 0.1333333

1. **Representa la variable *PSE (***Percepción Subjetiva del Esfuerzo)**mediante un diagrama de barras y un diagrama de sectores. Incluye un título adecuado para cada gráfico y colorea las barras y los sectores de colores diferentes**

**y.bar = barplot(abso, col = c("yellow", "red"), main = "Diagrama de barras para la variable Percepción Subjetiva del Esfuerzo")**

****

**> VI(y.bar)**

**The summary of each variable is**

**V1: Min. 0.7 1st Qu. 1.9 Median 3.1 Mean 3.1 3rd Qu. 4.3 Max. 5.5**

**> pie(abso, col = c("light blue", "black"), main = "Diagrama de sectores para la variable Percepción Subjetiva del Esfuerzo")**



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

**d) Para la variable *Edad*, realiza un histograma y un diagrama de caja y bigotes considerando la opción *range* = 1.5. Incluye un título apropiado para cada gráfico y colorea las barras del histograma de color verde. ¿Existe algún valor atípico en esta variable? Reduce el valor del argumento *range* hasta 1. ¿Varían las conclusiones?**

**> y.hist =hist(datos$Edad, col = "green", main = "Histograma para la variable Edad", xlab = "Edades", ylab = "Frecuencia")**



> VI(y.hist)

This is a histogram, with the title: Histogram of datos$Edad

"datos$Edad" is marked on the x-axis.

Tick marks for the x-axis are at: 18, 20, 22, 24, 26, 28, and 30

There are a total of 30 elements for this variable.

Tick marks for the y-axis are at: 0, 2, 4, 6, 8, and 10

It has 6 bins with equal widths, starting at 18 and ending at 30 .

The mids and counts for the bins are:

mid = 19 count = 6

mid = 21 count = 7

mid = 23 count = 10

mid = 25 count = 4

mid = 27 count = 2

mid = 29 count = 1

**> y.box1 = boxplot(datos$Edad, main = "Cajas y bigotes para la variable Edad (range = 1.5)")**



Se observa que existe un dato más allá del bigote superior, dicho dato se considera un valor atípico.

**> VI(y.box1)**

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: 18, 20, 22, 24, 26, and 28

This variable has 30 values.

An outlier is marked at: 29

The whiskers extend to 18 and 28 from the ends of the box,

which are at 21 and 24

The median, 23 is 67 % from the lower end of the box to the upper end.

The upper whisker is 1.33 times the length of the lower whisker.

**> y.box2 = boxplot(datos$Edad, main = "Cajas y bigotes para la variable Edad (range = 1)")**



Se sigue observando un valor atípico

> VI( y.box2)

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: 18, 20, 22, 24, 26, and 28

This variable has 30 values.

An outlier is marked at: 29

The whiskers extend to 18 and 28 from the ends of the box,

which are at 21 and 24

The median, 23 is 67 % from the lower end of the box to the upper end.

The upper whisker is 1.33 times the length of the lower whisker.

**e) Realiza un resumen de la variable *FCb* mediante la orden *summary*. Comprueba que las medidas que proporciona *summary* coinciden con las medidas calculadas de forma individual usando su función específica**

**> summary(datos$FCb)**

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

 **45.00 53.25 57.50 57.57 61.75 78.00**

Calculemos cada medida por separado

> min(datos$FCb)

**[1] 45**

> max(datos$FCb)
**[1] 78**

> quantile(datos$FCb, probs = c(0.25, 0.75))
**25% 75%**

**53.25 61.75**

> mean(datos$FCb)
**[1] 57.56667**

> median(datos$FCb)
**[1] 57.5**

**f) Calcula la edad media de los individuos y proporciona, al menos, dos medidas que indiquen la dispersión de esta variable**

**> mean(datos$Edad)**

**[1] 22.96667**

La edad media de los individuos es casi 23 años

Como medidas de dispersión, se van a calcular la cuasi-varianza y el recorrido intercuartílico.

**> var(datos$Edad)**

**[1] 6.929885**

**> IQR(datos$Edad)**

**[1] 2.75**

**Nota:** Si la variable Edad tuviera valores faltantes entre sus observaciones. Este hecho tiene que ser indicado estableciendo como *TRUE* el valor del parámetro *na.rm. Es decir, las ordenes serían*

> mean(datos$Edad, na.rm = TRUE)

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

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

**g) ¿Qué variable es más homogénea: *FCb* o *FCm*?**

Para determinar la homogeneidad de una variable (o, lo que es lo mismo, la representatividad de su media), calculamos el coeficiente de variación de Pearson para cada una de ellas, el cual se define como el cociente entre la desviación típica y la media de la variable.

Vamos a comenzar con la variable FCb. En primer lugar, calculemos la media.

> media\_FCb = mean(datos$FCb)

> media\_FCb

**[1] 57.56667**

A continuación, obtenemos la desviación típica. Para ello, calcularemos en primer lugar la varianza de la variable Edad mediante la función var, que recordemos calcula la cuasi-varianza de una variable.

**> var\_FCb = 9/10 \* var(datos$FCb)**

**> var\_FCb**

**[1] 42.55966**

Una vez obtenida la varianza, la desviación típica se obtiene como su raíz cuadrada positiva.

**> dt\_FCb = sqrt(var\_FCb)**

**> dt\_FCb**

**[1] 6.523776**

Por último, calculamos el coeficiente de variación de la variable FCb. Es decir calculamos el cociente entre la desviación típica y la media y mostramos su valor.

**> CV\_FCb = dt\_FCb/media\_FCb**

**> CV\_FCb**

**[1] 0.1133256**

Repetimos el mismo proceso con la variable FCm

**> media\_FCm = mean(datos$FCm)**

**> media\_FCm**

**[1] 154.3333**

**> var\_FCm = 9/10 \* var(datos$FCm)**

**> var\_FCm**

**[1] 280.4483**

**> dt\_FCm = sqrt(var\_FCm)**

**> dt\_FCm**

**[1] 16.74659**

**> CV\_FCm = dt\_FCb/media\_FCm**

**> CV\_FCm**

**[1] 0.04227069**

La variable más homogénea es la variable FCm, ya que presenta un coeficiente de variación más próximo a 0.

**h) Estudia la asimetría y la curtosis de la variable *Edad*.**

Para determinar la simetría y la curtosis, es necesario instalar y cargar el paquete *e1071*. En nuestro caso ya está instalado por lo que sólo lo vamos a cargar

**> library("e1071")**

**> skewness(datos$Edad)**

**[1] 0.232235**

**> kurtosis(datos$Edad)**

**[1] -0.4008408**

**> y = datos$Edad**

**> y**

 **[1] 22 24 25 23 19 18 20 22 24 25 27 24 23 22 28 19 20 22 24 25 26 24 20 21 21**

**[26] 22 29 23 23 24**

[Muestra la siguiente salida en HTML](https://wpd.ugr.es/~bioestad/wp-content/uploads/Univariate-analysis-for-Y2.html)

**Editor de R**

library("BrailleR")

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

datos = read.table("propuesto3.txt", header = TRUE)

abso = table(datos$PSE)

abso

rela = prop.table(abso)

rela

y.bar = barplot(abso, col = c("yellow", "red"), main = "Diagrama de barras para la variable Percepción Subjetiva del Esfuerzo")

VI(y.bar)

pie(abso, col = c("light blue", "black"), main = "Diagrama de sectores para la variable Percepción Subjetiva del Esfuerzo")

y.hist =hist(datos$Edad, col = "green", main = "Histograma para la variable Edad", xlab = "Edades", ylab = "Frecuencia")

VI(y.hist)

y.box1 = boxplot(datos$Edad, main = "Cajas y bigotes para la variable Edad (range = 1.5)")

VI(y.box1)

 y.box2 = boxplot(datos$Edad, main = "Cajas y bigotes para la variable Edad (range = 1)")

VI( y.box2)

summary(datos$FCb)

min(datos$FCb)

max(datos$FCb)

quantile(datos$FCb, probs = c(0.25, 0.75))

mean(datos$FCb)

median(datos$FCb)

mean(datos$Edad)

var(datos$Edad)

IQR(datos$Edad)

media\_FCb = mean(datos$FCb)

media\_FCb

var\_FCb = 9/10 \* var(datos$FCb)

var\_FCb

dt\_FCb = sqrt(var\_FCb)

dt\_FCb

CV\_FCb = dt\_FCb/media\_FCb

CV\_FCb

media\_FCm = mean(datos$FCm)

media\_FCm

var\_FCm = 9/10 \* var(datos$FCm)

var\_FCm

dt\_FCm = sqrt(var\_FCm)

dt\_FCm

CV\_FCm = dt\_FCb/media\_FCm

CV\_FCm

library("e1071")

skewness(datos$Edad)

kurtosis(datos$Edad)

y = datos$Edad

y

UniDesc(y)