
EJERCICIOS RESUELTOS

8.1.- Las longitudes en micras de 25 grietas medidas en una pieza de hormigón están en el archivo [grietas](#).

- 1) Agrupar los datos en los siguientes intervalos:
[20,40), [40,60), [60,80), [80,100), [100,120).
- 2) Construir una tabla de frecuencias en la que figuren las siguientes columnas: Intervalo / Marca de clase / Frecuencia absoluta / Frecuencia absoluta acumulada / Frecuencia relativa / Frecuencia relativa acumulada.

SOLUCIÓN 8.1

8.2.- Los valores del archivo [terrenos](#) son los precios, en miles de euros, de 35 terrenos rurales vendidos en Bizkaia en el año 2006.

Se pide:

- 1º) Histograma.
- 2º) Diagrama boxplot.
- 3º) Detectar datos atípicos.
- 4º) Media, mediana, desviación típica y RIQ antes y después de eliminar los datos atípicos, si existen.

SOLUCIÓN 8.2

8.3.- Utilizar los datos **energy** del paquete **ISwR** para comparar el gasto de energía entre delgados (*lean*) y obesos (*obese*).

SOLUCIÓN 8.3

8.4.- Utilizar los datos **uscereal** del paquete **MASS** para calcular la recta de regresión de la variable cuantitativa *grasas* (**fat**) sobre la variable cuantitativa *calorías* (**calories**). Repetir la regresión tras eliminar un punto atípico.

SOLUCIÓN 8.4

SOLUCIÓN 8.1

```
> datos<-read.table("C:\\grietas.txt")
> grietas<-datos$V1
> intervalos<-c(20,40,60,80,100,120)
> grietas.int<-cut(grietas,intervalos,right=F)
> grietas.int
[1] [40,60) [60,80) [80,100) [80,100) [60,80)
[60,80) [60,80)
.....
[40,60) [60,80)
[22] [20,40) [60,80) [60,80) [40,60)
Levels: [20,40) [40,60) [60,80) [80,100) [100,120)
> Int<-levels(grietas.int)
> xi<-c(30,50,70,90,100,110)
> table(grietas.int)
grietas.int
 [20,40) [40,60) [60,80) [80,100) [100,120)
       2       5       10       6       2
> a<-matrix(table(grietas.int))
      [,1]
[1,]  2
[2,]  5
[3,] 10
[4,]  6
[5,]  2
```

En la instrucción anterior hemos creado una matriz con los valores de las frecuencias absolutas. Con la instrucción siguiente construimos el vector Frec.abs seleccionando los correspondientes elementos de la matriz (1ª columna y todas las filas).

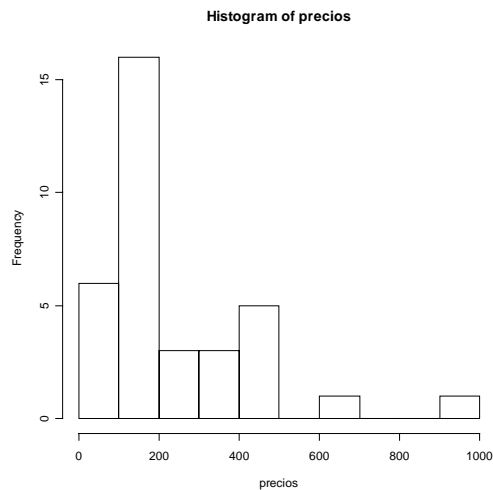
```
> Fr.abs<-c(a[,1])
> Fr.abs
[1] 2 5 10 6 2
> sum(Fr.abs)
[1] 25
> Fr.rel<-Fr.abs/25
> Fr.rel
[1] 0.08 0.20 0.40 0.24 0.08
> Fr.abs.ac<-cumsum(Fr.abs)
> Fr.rel.ac<-cumsum(Fr.rel)
> data.frame(Int,xi,Fr.abs,Fr.abs.ac,Fr.rel,Fr.rel.ac)
      Int xi Fr.abs Fr.abs.ac Fr.rel F.rel.ac
1 [20,40) 30     2         2  0.08   0.08
2 [40,60) 50     5         7  0.20   0.28
3 [60,80) 70    10        17  0.40   0.68
4 [80,100) 90     6        23  0.24   0.92
5 [100,120) 110    2        25  0.08   1.00
```

SOLUCIÓN 8.2

```

> datos<-read.table("C:\\terrenos.txt")
> precios<-datos$V1
> precios
 [1] 115 232 181 161 155 137 165 171 139 130 406 69
 [2] 171 135 135 132 88 410 87
 [20] 90 123 157 345 323 411 334 80 87 235 198 450
 [30] 223 602 415 950
> hist(precios)

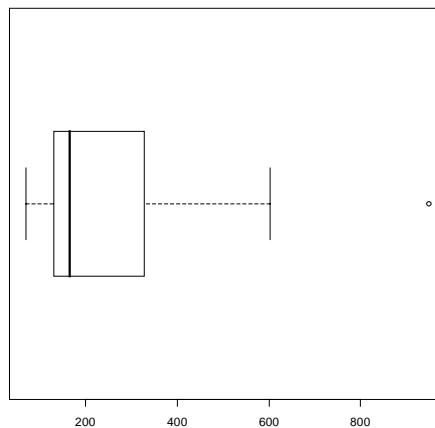
```



```

> boxplot(precios, horizontal=T)

```



```

> boxplot.stats(precios)
 $stats
 [1] 69.0 131.0 165.0 328.5 602.0
 $n
 [1] 35
 $conf
 [1] 112.2539 217.7461
 $out
 [1] 950

```

Ahora vamos a construir un nuevo vector, `precios.nuevo`, que contiene todos los valores del vector `precios` excepto el dato atípico. Para ello

elegimos dentro de este vector todos los valores que sean distintos ($i \neq$) de 950, el único valor atípico:

```
> precios.nuevo<-precios[precios!=950]
> precios.nuevo
 [1] 115 232 181 161 155 137 165 171 139 130 406 69
171 135 135 132 88 410 87
[20] 90 123 157 345 323 411 334 80 87 235 198 450
223 602 415
> length(precios);length(precios.nuevo)
[1] 35
[1] 34
>mean(precios);median(precios);sqrt(34/35)*sd(precios);
quantile(precios,0.75)-quantile(precios,0.25)
[1] 235.4857
[1] 165
[1] 177.7582
 75%
197.5
>mean(precios.nuevo);median(precios.nuevo);sqrt(33/34)*
sd(precios.nuevo);quantile(precios.nuevo,0.75)-
quantile(precios.nuevo,0.25)
[1] 214.4706
[1] 163
[1] 130.6525
 75%
170.5
```

SOLUCIÓN 8.3

```
> library(ISwR)
> data(energy)
> attach(energy)
> energy
  expend stature
1     9.21  obese
2     7.53   lean
3     7.48   lean
4     8.08   lean
5     8.09   lean
6    10.15   lean
7     8.40   lean
8    10.88   lean
9     6.13   lean
10    7.90   lean
11   11.51  obese
12   12.79  obese
13    7.05   lean
14   11.85  obese
15    9.97  obese
16    7.48   lean
```

```

17  8.79  obese
18  9.69  obese
19  9.68  obese
20  7.58   lean
21  9.19  obese
22  8.11   lean

```

A continuación vamos a separar el vector `expend` en dos vectores de acuerdo al valor que toma el factor `stature` en cada uno de los 22 elementos:

```

> expend.lean<-expend[stature=="lean"]
> expend.obese<-expend[stature=="obese"]
> expend.lean
 [1]  7.53  7.48  8.08  8.09 10.15  8.40 10.88  6.13
 [2]  7.90  7.05  7.48  7.58
 [13]  8.11
> expend.obese
 [1]  9.21 11.51 12.79 11.85  9.97  8.79  9.69  9.68
 [2]  9.19

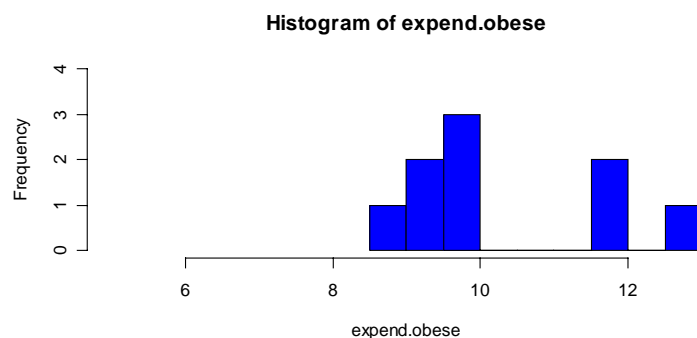
```

Vamos ahora a dibujar los histogramas (con determinados parámetros de ejes, color, etc.) de las dos variables en un mismo gráfico, para lo que debemos hacer previamente una partición del gráfico en dos filas y una columna:

```

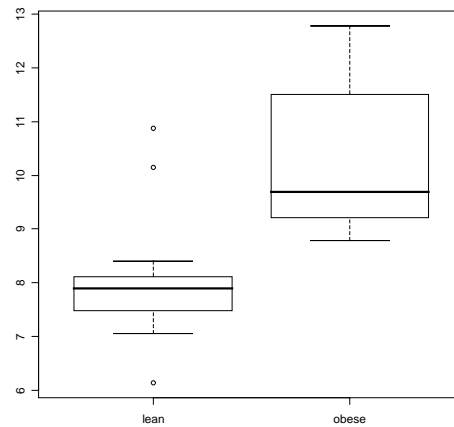
> par(mfrow=c(2,1))
> hist(expend.lean,breaks=10,xlim=c(5,13),ylim=c(0,4),
col="red")
> hist(expend.obese,breaks=10,xlim=c(5,13),ylim=c(0,4),
col="blue")

```



Por último vamos a generar un gráfico boxplot por grupos, volviendo previamente a una ventana en el gráfico:

```
> par(mfrow=c(1,1))
> boxplot(expend~stature)
```



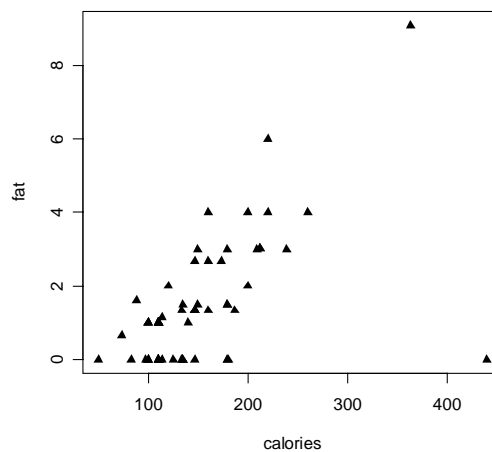
Alternativamente podríamos haber hecho lo siguiente para obtener el mismo gráfico anterior:

```
> boxplot(expend.lean,expend.obese)
```

SOLUCIÓN 8.4

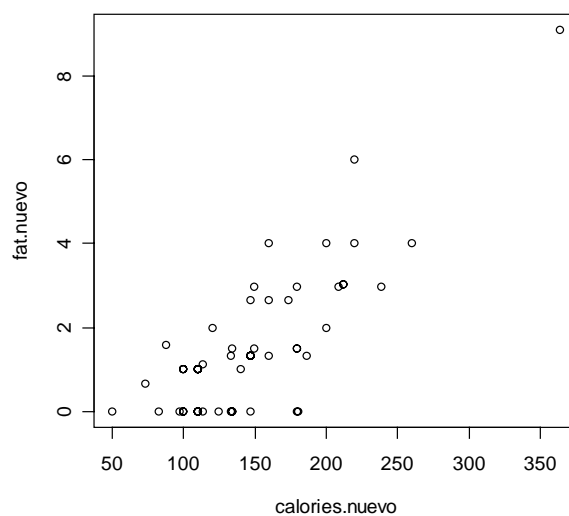
```
> library(MASS)
> data(UScereal)
> attach(UScereal)
> names(UScereal)
[1] "mfr"           "calories"      "protein"       "fat"
"sodium"        "fibre"
[7] "carbo"        "sugars"        "shelf"         "potassium"
"vitamins"
```

```
> plot(calories,fat,pch=17)
```



En el gráfico se observa un valor atípico en el ángulo inferior derecho. Para detectarlo escribimos la instrucción `identify`, pulsamos Enter y después accionamos el botón izquierdo del ratón al colocar éste sobre el punto en el gráfico. Así se obtiene el índice (número de orden) que corresponde al punto buscado.

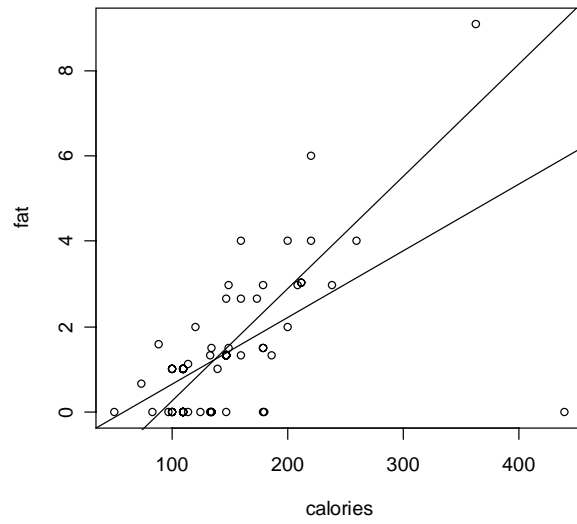
```
> identify(calories,fat,n=1)
[1] 31
> calories.nuevo<-calories[-31] #Eliminamos el dato atípico
> fat.nuevo<-fat[-31] #Eliminamos el dato atípico
> plot(calories.nuevo,fat.nuevo)
```



```
> lm(fat~calories)
Call:
lm(formula = fat ~ calories)
Coefficients:
(Intercept)    calories
  -0.90472      0.01558

> lm(fat.nuevo~calories.nuevo)
Call:
lm(formula = fat.nuevo ~ calories.nuevo)
Coefficients:
(Intercept)  calories.nuevo
  -2.36711      0.02631

> plot(calories,fat)
> abline(-0.90472,0.01558)
> abline(-2.36711,0.02631)
```



Como se ve en este último gráfico las rectas de regresión con o sin el punto atípico varían considerablemente.