

## Práctica 3. Relación de Problemas de Descriptiva.

Antonio Martín Andrés, Juan de Dios Luna del Castillo, Christian J. Acal González



**UNIVERSIDAD  
DE GRANADA**

Todo el material para el conjunto de actividades de este curso ha sido elaborado y es propiedad intelectual del grupo **BioestadísticaR** formado por:

Antonio Martín Andrés  
Juan de Dios Luna del Castillo,  
Pedro Femia Marzo,  
Miguel Ángel Montero Alonso,  
Christian José Acal González,  
Pedro María Carmona Sáez,  
Juan Manuel Melchor Rodríguez,  
José Luis Romero Béjar,  
Manuela Expósito Ruíz,  
Juan Antonio Villatoro García.

Todos los integrantes del grupo han participado en todas las actividades, en su elección, construcción, correcciones o en su edición final, no obstante, en cada una de ellas, aparecerán uno o más nombres correspondientes a las personas que han tenido la máxima responsabilidad de su elaboración junto al grupo de **BioestadísticaR**.

Todos los materiales están protegidos por la Licencia Creative Commons **CC BY-NC-ND** que permite "descargar las obras y compartirlas con otras personas, siempre que se reconozca su autoría, pero no se pueden cambiar de ninguna manera ni se pueden utilizar comercialmente".

## Práctica 3. Relación de Problemas de Descriptiva

Antonio Martín Andrés, Juan de Dios Luna del Castillo y Christian J. Acal González

### Ejercicio 3.1

Un experimento, que se realiza con 60 estudiantes, consiste en la medición de la concentración de sodio en el sudor. Las determinaciones debían redondearse al número entero más cercano, expresado en mEq/l, siendo los resultados los siguientes:

46; 29; 35; 61; 54; 37; 53; 57; 52; 51; 43; 67; 66; 31; 53; 51; 48; 59; 55; 47; 76; 49; 59; 50; 65; 41; 60; 51; 43; 82; 63; 58; 43; 61; 73; 38; 71; 47; 47; 60; 69; 53; 51; 39; 66; 53; 56; 72; 75; 52; 63; 57; 54; 77; 59; 36; 45; 63; 67; 44.

- Calcular la media, varianza y mediana de estos datos;
- Hacer una representación gráfica.

Resolver este problema con la calculadora o con R.

### Solución Manual

Considerando la muestra de  $n = 60$ , se obtiene:  $\sum_{i=1}^{60} x_i = 3283$  y  $\sum_{i=1}^{60} x_i^2 = 187973$ .

A partir de ahí,  $\bar{x} = \frac{\sum x_i}{n} = \frac{3283}{60} = 54.7$

$$s^2 = \frac{1}{n-1} \left( \sum x_i^2 - \left( \frac{(\sum x_i)^2}{n} \right) \right) = \frac{1}{60-1} \left( 187973 - \frac{3283^2}{60} \right) = 141.325144,$$

$$s = \sqrt{s^2} = 11.88$$

La mediana se calcula ordenando los datos de menor a mayor y promediando los valores que están en la posición 30 y la 31 y se obtiene 53.5.

Aunque no se piden, de la misma manera, el primer cuartil sería el que está en el punto medio de los valores en la posición 1 y la 16,  $c_1 = 47$ , el segundo cuartil sería la mediana y el tercer cuartil sería el punto medio de los valores que están en la posición 45 y la 46,  $c_3 = 63$ .

La gráfica queda a elección del alumno pero la más apropiada sería un histograma.

### Solución con R

```
#Introducción de datos
datos=c(46,29,35,61,54,37,53,57,52,51,43,67,66,31,53,51,48,59,55,47,76,49,59,50,65,41,60,
51,43,82,63,58,43,61,73,38,71,47,47,60,69,53,51,39,66,53,56,72,75,52,63,57,54,77,
59,36,45,63,67,44)
```

```
#Aunque en R no es necesario, a continuación se obtienen los pasos previos  
sum(datos)
```

```
## [1] 3283
```

```
sum(datos^2)
```

```
## [1] 187973
```

```
#Media  
mean(datos)
```

```
## [1] 54.71667
```

```
#Varianza  
var(datos)
```

```
## [1] 141.3251
```

```
#Desviación típica  
sd(datos)
```

```
## [1] 11.88803
```

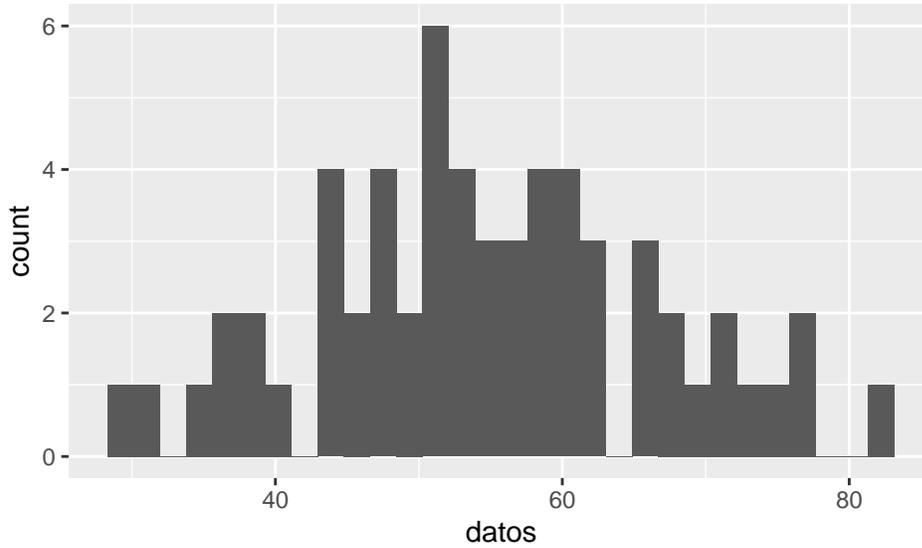
```
#Mediana  
median(datos)
```

```
## [1] 53.5
```

```
#Resumen de las medidas  
library(summarytools)  
descr(datos,stats=c("mean","sd","q1","med","q3"))
```

```
## Descriptive Statistics  
## datos  
## N: 60  
##  
##           datos  
## -----  
##      Mean  54.72  
##   Std.Dev  11.89  
##        Q1  47.00  
##     Median  53.50  
##        Q3  63.00
```

```
library(ggplot2)  
ggplot(as.data.frame(datos), aes(x=datos)) + geom_histogram()
```



### Ejercicio 3.2

La variabilidad de las medidas de glucosa en sangre es una variable que refleja muy bien el control de la diabetes en pacientes de Tipo II. Esa variabilidad se mide, en pacientes de los Centros de Salud mediante el coeficiente de variación de 10 medidas de tal variable hechas en 2 semanas. Como se asocia fuertemente esa medida de variabilidad con evolución y complicaciones de la enfermedad se ha establecido el valor del CV del 36% para indicar que hay un descontrol que debe ser controlado.

Conocidos estos hechos se trata de completar la siguiente tabla para 5 pacientes señalando los que podríamos considerar cuales no tienen el control metabólico de la enfermedad.

P	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10
1	149	243	155	232	139	149	183	135	142	132
2	232	30	103	210	128	121	128	99	127	82
3	188	246	205	151	33	184	186	205	161	154
4	84	159	104	147	158	126	172	143	141	118
5	64	95	115	118	161	274	94	156	191	170

### Solución Manual

La tabla con los resultados aparece a continuación

P	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	xmedia	sd	cv	Control(N/S)
1	149	243	155	232	139	149	183	135	142	132	165.9	40.41	24.4	S
2	232	30	103	210	128	121	128	99	127	82	126	58.44	46.4	N
3	188	246	205	151	33	184	186	205	161	154	171.3	56.23	32.8	S
4	84	159	104	147	158	126	172	143	141	118	135.2	27.18	20.1	S
5	64	95	115	118	161	274	94	156	191	170	143.8	60.59	42.1	N

La media, desviación típica y cv se calculan con las expresiones conocidas. Así para el primer paciente:

- $\sum x_i = 1659$  y  $\sum x_i^2 = 289923$
- Media:  $\bar{x} = \frac{\sum x_i}{n} = \frac{1659}{10} = 165.9$
- Varianza:

$$s^2 = \frac{1}{n-1} \left( \sum x_i^2 - \left( \frac{(\sum x_i)^2}{n} \right) \right) = \frac{1}{10-1} \left( 289923 - \frac{1659^2}{10} \right) = 1632.77$$

- Desviación típica:  $s = \sqrt{s^2} = \sqrt{1632.77} = 40.41$
- Coeficiente de variación:  $cv = \frac{s}{\bar{x}} = \frac{40.41}{165.9} \times 100 = 24.4$

## Solución con R

```
#Introducción de datos
p1=c(149,243,155,232,139,149,183,135,142,132)
p2=c(232,30,103,210,128,121,128,99,127,82)
p3=c(188,246,205,151,33,184,186,205,161,154)
p4=c(84,159,104,147,158,126,172,143,141,118)
p5=c(64,95,115,118,161,274,94,156,191,170)

#Medias
media_p1=mean(p1)
media_p2=mean(p2)
media_p3=mean(p3)
media_p4=mean(p4)
media_p5=mean(p5)
medias=c(media_p1,media_p2,media_p3,media_p4,media_p5)

#Desviaciones típicas
sd_p1=sd(p1)
sd_p2=sd(p2)
sd_p3=sd(p3)
sd_p4=sd(p4)
sd_p5=sd(p5)
desviaciones=c(sd_p1,sd_p2,sd_p3,sd_p4,sd_p5)

#Coeficientes de variacion
cv=(desviaciones/medias)*100

#Controlado
control=cv<36
control[control==TRUE]="S"
control[control==FALSE]="N"

resultados=data.frame(medias,desviaciones,cv,control)
rownames(resultados)=c("P1","P2","P3","P4","P5")
resultados
```

##	medias	desviaciones	cv	control
## P1	165.9	40.40751	24.35654	S
## P2	126.0	58.43895	46.38012	N
## P3	171.3	56.22583	32.82302	S
## P4	135.2	27.18169	20.10480	S
## P5	143.8	60.58566	42.13189	N