

## Ejercicio 1

Nos han seleccionado para una posición de Data Scientist en una empresa que fabrica radiadores para el sector de la automoción. Nos presentamos a nuestro nuevo puesto de trabajo con mucha ilusión y recibimos nuestro primer encargo a primera hora de la mañana, el cual proviene del departamento de calidad: determinar cuántas reclamaciones recibe nuestra compañía sobre la calidad de los radiadores que se fabrican. Disponemos de registros correspondientes a un periodo de 5 días, el cual muestra el siguiente número de reclamaciones:

- Lunes: 26
- Martes: 30
- Miércoles: 16
- Jueves: 32
- Viernes: 16

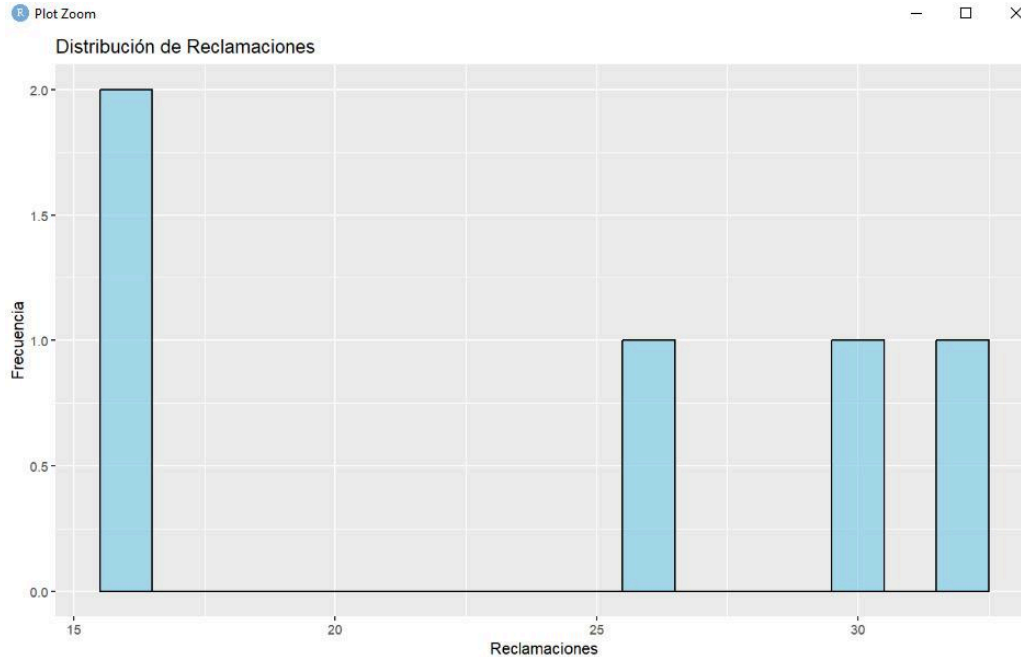
**Pregunta:** El encargo consiste en calcular: la media, la mediana, la moda, la frecuencia relativa de cada día, la varianza poblacional, la desviación estándar y dibujar la distribución normal de la variable. **NOTA:** En caso de no poder calcular alguno de ellos, se deberá explicar el porqué.

The screenshot shows the RStudio interface with the following content:

```
R 4.3.2 ~ /> reclamaciones <- c(26, 30, 16, 32, 16)
> media <- mean(reclamaciones)
> mediana <- median(reclamaciones)
> reclamaciones_ordenadas <- sort(reclamaciones)
Error: unexpected symbol in "reclamaciones_ordenadas <- sort"
> contar_repe <- table(reclamaciones_ordenadas)
> moda <- names(which.max(contar_repe))
> frecuencia_relativa_diaria <- table(reclamaciones) / length(reclamaciones)
> varianza_poblacional <- var(reclamaciones) * (length(reclamaciones)-1) / length(reclamaciones)
> desviacion_estandar <- sqrt(varianza_poblacional)
> library(ggplot2)
> histograma <- ggplot(data, aes(x =reclamaciones)) + geom_histogram(binwidth = 1, fill="skyblue", color="black", alpha(0,7))+labs(title="Distribución de reclamaciones", x="reclamaciones", y="Frecuencia")
Error: object 'histograma' not found
> data <- data.frame(reclamaciones = c(26, 30, 16, 32, 16))
> histograma <- ggplot(data, aes(x = reclamaciones)) +
+   geom_histogram(binwidth = 1, fill="skyblue", color="black", alpha=0.7) +
+   labs(title="Distribución de Reclamaciones", x="Reclamaciones", y="Frecuencia")
Error in `+` :
! Cannot use `+` with a single argument.
i Did you accidentally put `+` on a new line?
Run `rlang::last_trace()` to see where the error occurred.
> histograma <- ggplot(data, aes(x = reclamaciones)) +
+   geom_histogram(binwidth = 1, fill="skyblue", color="black", alpha=0.7) +
+   labs(title="Distribución de Reclamaciones", x="Reclamaciones", y="Frecuencia")
> print(histograma)
>
```

The Environment pane shows the following variables:

Variable	Value
data	5 obs. of 1 variable
histograma	List of 11
contar_repe	'table' int [1:4(1d)] 2 1 1 1
desviacion_es...	6.81175454637056
frecuencia_re...	'table' num [1:4(1d)] 0.4 0.2 0.2 0.2
media	24
mediana	26
moda	"16"
reclamaciones	num [1:5] 26 30 16 32 16
reclamaciones...	num [1:5] 16 16 26 30 32
varianza_pobl...	46.4



## Ejercicio 2

Nuestro encargo ha sido resuelto con un éxito absoluto, con lo cual recibimos felicitaciones y elogios por parte del departamento de calidad. ¡Nuestro primer día en la empresa ha sido todo un éxito! El día siguiente, nos vuelven a llamar del departamento de calidad para solicitarnos otro encargo. Tenemos que evaluar si hay relación entre la calidad de los radiadores fabricados en las 15 líneas de ensamblaje de la fábrica. Nos proporcionan la siguiente tabla donde apreciamos el identificador de cada línea, junto al promedio de la durabilidad en años de todos los radiadores fabricados en cada línea.

### Línea Durabilidad

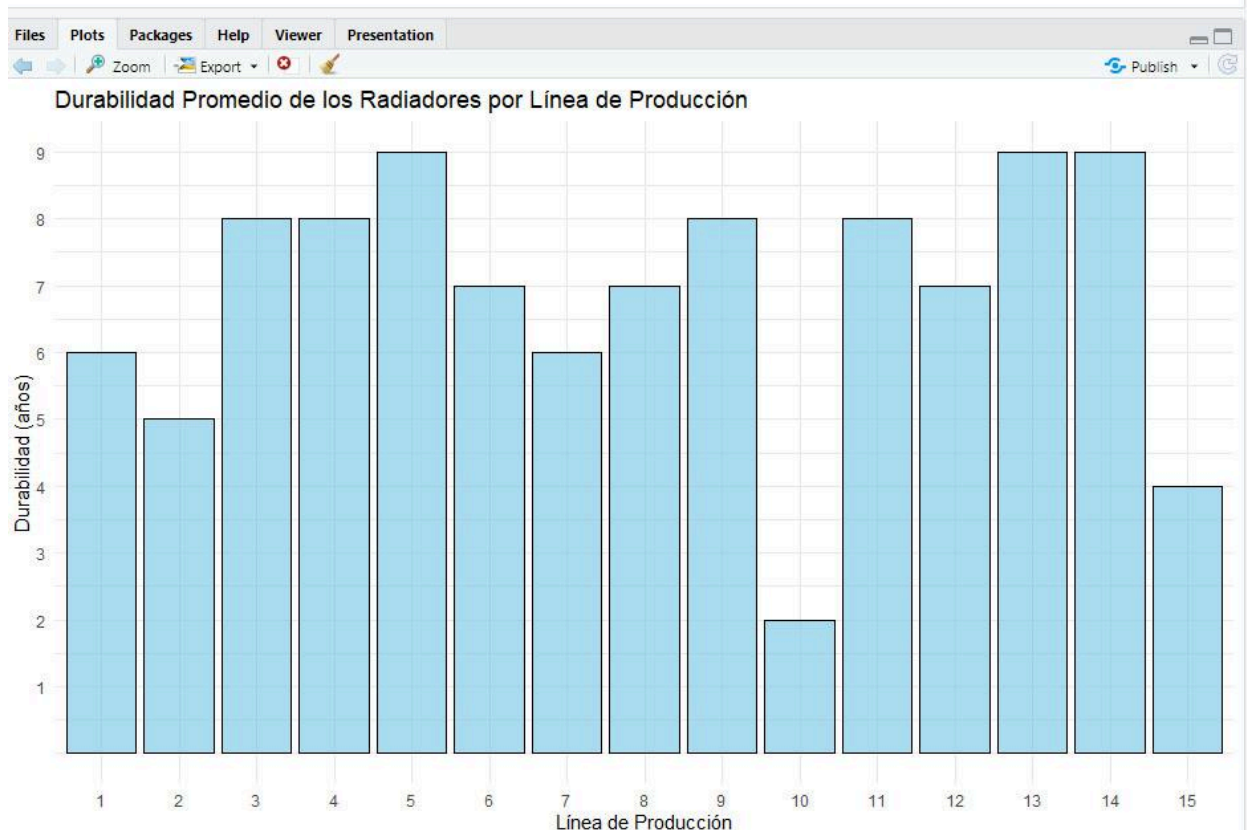
1	6
2	5
3	8
4	8
5	9
6	7
7	6
8	7
9	8
10	2
11	8
12	7
13	9
14	9
15	4

**Pregunta: 2.a. ¿Hay alguna línea de producción cuya calidad es excesivamente baja a comparación con las demás? ¿Cuál de ellas es?**

La décima línea tiene una calidad excesivamente baja se ve claramente en la gráfica..

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Console Terminal Background Jobs
R 4.3.2. ~/
> durabilidad <- c(6, 5, 8, 8, 9, 7, 6, 7, 8, 2, 8, 7, 9, 9, 4)
> mediana <- median(durabilidad)
> histograma <- ggplot(data, aes(x = factor(linea), y = durabilidad)) + geom_bar(stat = "identity", fill="skyblue", color="black", alpha=0.7) + labs(title="Durabilidad Promedio de los Radiadores por Línea de Producción", x="Línea de Producción", y="Durabilidad (años)") + theme_minimal()
Error in ggplot(data, aes(x = factor(linea), y = durabilidad)) :
  could not find function "ggplot"
> install.packages("ggplot2")
Error in install.packages : updating loaded packages
> library(ggplot2)
> histograma <- ggplot(data, aes(x = factor(linea), y = durabilidad)) + geom_bar(stat = "identity", fill="skyblue", color="black", alpha=0.7) + labs(title="Durabilidad Promedio de los Radiadores por Línea de Producción", x="Línea de Producción", y="Durabilidad (años)") + theme_minimal()
Error in `ggplot()`:
! `data` cannot be a function.
i Have you misspelled the `data` argument in `ggplot()`
Run `rlang::last_trace()` to see where the error occurred.
> durabilidad <- c(6, 5, 8, 8, 9, 7, 6, 7, 8, 2, 8, 7, 9, 9, 4)
> data <- data.frame(linea = 1:15,
+                   durabilidad = c(6, 5, 8, 8, 9, 7, 6, 7, 8, 2, 8, 7, 9, 9, 4))
>
> histograma <- ggplot(data, aes(x = factor(linea), y = durabilidad)) + geom_bar(stat = "identity", fill="skyblue", color="black", alpha=0.7) + scale_y_continuous(breaks = 1:10) + eje y de 1 a 10 labs(title="Durabilidad Promedio de los Radiadores por Línea de Producción", x="Línea de Producción", y="Durabilidad (años)") + theme_minimal()
Error: unexpected symbol in "histograma <- ggplot(data, aes(x = factor(linea), y = durabilidad)) + geom_bar(stat = "identity", fill="skyblue", color="black", alpha=0.7) + scale_y_continuous(breaks = 1:10) + eje y"
> histograma <- ggplot(data, aes(x = factor(linea), y = durabilidad)) +
+   geom_bar(stat = "identity", fill="skyblue", color="black", alpha=0.7) +
+   scale_y_continuous(breaks = 1:10) + # Esto especifica los 'breaks' en el eje y de 1 a 10
+   labs(title="Durabilidad Promedio de los Radiadores por Línea de Producción",
+         x="Línea de Producción",
+         y="Durabilidad (años)") +
+   theme_minimal()
> print(histograma)
> |
```

Environment		History	Connections	Tutorial
Project: (None)				
Import Dataset		179 MIB		
R		Global Environment		
Data				
data	15 obs. of 2 variables			
histograma	List of 11			
Values				
durabilidad	num [1:15] 6 5 8 8 9 7 6 7 8 2 ...			
mediana	7			



**2.b. ¿Cuál de los siguientes grupos ofrece la mayor calidad? ¿1-3, 4-6, o 7-9? Responde a ambas cuestiones mediante la representación gráfica, eligiendo aquellos gráficos que consideres más oportunos para cada una de las preguntas.**

Los promedios de durabilidad de cada grupo son:

$$\text{Grupo 1-3} = \frac{6+5+8}{3} = 6,33 \text{ años}$$

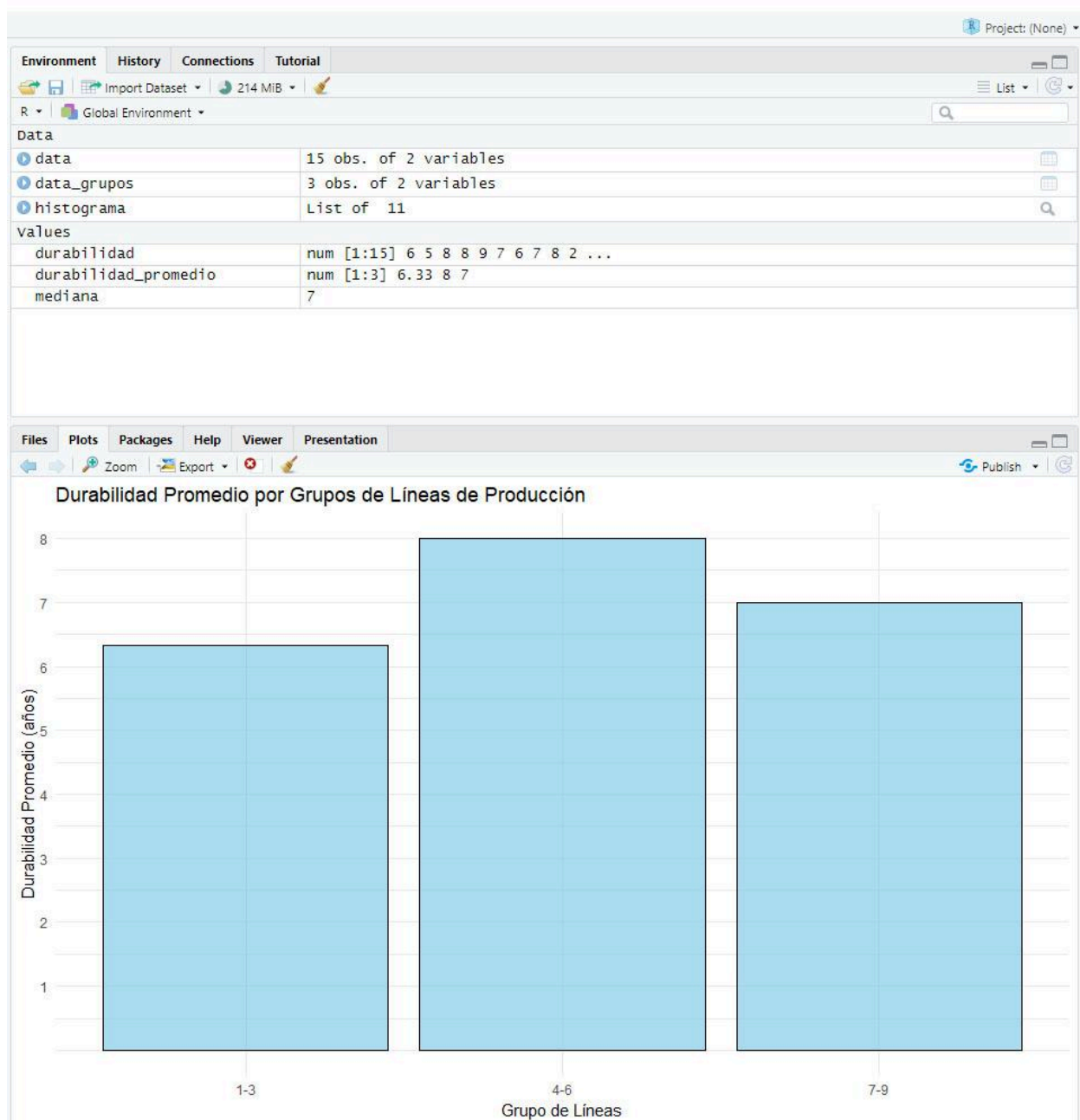
$$\text{Grupo 4-6} = \frac{7+6+8}{3} = 8 \text{ años}$$

$$\text{Grupo 7-9} = \frac{6+5+8}{3} = 7 \text{ años}$$

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Console Terminal Background Jobs
R 4.3.2 . ~/
> durabilidad <- c(6, 5, 8, 8, 9, 7, 6, 7, 8, 2, 8, 7, 9, 9, 4)
> mediana <- median(durabilidad)
> library(ggplot2)
> histograma <- ggplot(data, aes(x = factor(linea), y = durabilidad)) + geom_bar(stat = "identity", fill="skyblue", color="black", alpha=0.7) + scale_y_continuous(breaks = 1:10) + labs(title="Durabilidad Promedio de los Radiadores por Línea de Producción", x="Línea de Producción",y="Durabilidad (años)") + theme_minimal()
Error in `ggplot()`:
! `data` cannot be a function.
i Have you misspelled the `data` argument in `ggplot()`?
Run `rlang::last_trace()` to see where the error occurred.
> data <- data.frame(linea = 1:15, durabilidad = c(6, 5, 8, 8, 9, 7, 6, 7, 8, 2, 8, 7, 9, 9, 4))
> histograma <- ggplot(data, aes(x = factor(linea), y = durabilidad)) + geom_bar(stat = "identity", fill="skyblue", color="black", alpha=0.7) + scale_y_continuous(breaks = 1:10) + labs(title="Durabilidad Promedio de los Radiadores por Línea de Producción", x="Línea de Producción",y="Durabilidad (años)") + theme_minimal()
> print(histograma)
> durabilidad_promedio <- c(mean(data$durabilidad[1:3]),
+                           mean(data$durabilidad[4:6]),
+                           mean(data$durabilidad[7:9]))
> print(durabilidad_promedio)
[1] 6.333333 8.000000 7.000000
> data_grupos <- data.frame(grupo = c("1-3", "4-6", "7-9"),
+                           durabilidad_promedio = durabilidad_promedio)
> histograma <- ggplot(data_grupos, aes(x = grupo, y = durabilidad_promedio)) +
+   geom_bar(stat = "identity", fill="skyblue", color="black", alpha=0.7) +
+   scale_y_continuous(breaks = 1:10) + # Esto podría ajustarse si es necesario para el rango de datos
+   labs(title="Durabilidad Promedio por Grupos de Líneas de Producción",
+         x="Grupo de Líneas",
+         y="Durabilidad Promedio (años)") +
+   theme_minimal()
> print(histograma)
> |

```



El grupo de la línea **4.6** es el que ofrece mayor calidad.

**Ejercicio 3 A** continuación disponemos de la información sobre un estudio realizado en una empresa de mecanizados. En ella, se ha recogido la muestra de 6 operarios tipo A y de 5 operarios tipo B, para los cuales se ha anotado ls horas de dedicación a la formación interna de la empresa.

OP A	OP B
26	23
25	30
43	18
34	25
18	28
52	

**Pregunta:** Mediante la aplicación de la T-Student (t-test), se pide saber si deberíamos rechazar la  $H_0$  (hipótesis nula). ¿Qué conclusión podemos extraer sobre si existe una diferencia significativa entre las horas dedicadas de ambos grupos? Las dos hipótesis son:

**$H_0: \mu_1 - \mu_2 = 0$**  (No existe diferencia significativa entre las medias de los dos grupos)

**$H_1: \mu_1 - \mu_2 \neq 0$**  (Existe diferencia significativa entre las medias de los dos grupos)

Utilizamos el t-test para muestras independientes, ya que los dos grupos de operarios son distintos y no relacionados entre sí.

El t-test compara las medias de los dos grupos y evalúa si las diferencias observadas pueden considerarse estadísticamente significativas, dadas las variaciones dentro de cada grupo y el tamaño de las muestras.

The screenshot shows the RStudio interface with the following content:

```
R 4.3.2 ~ /> operarios_A <- c(26, 25, 43, 34, 18, 52)
> operarios_B <- c(23, 30, 18, 25, 28)
> test_result <- t.test(operarios_A, operarios_B)
> print(test_result)
```

welch Two Sample t-test

data: operarios\_A and operarios\_B  
t = 1.4726, df = 6.5433, p-value = 0.1873  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
-5.155702 21.555702  
sample estimates:  
mean of x mean of y  
33.0 24.8

The Environment pane shows the following data:

Variable	Value
test_result	List of 10
operarios_A	num [1:6] 26 25 43 34 18 52
operarios_B	num [1:5] 23 30 18 25 28

El valor de  $t$  es pequeño no podemos rechazar la hipótesis nula.

El valor  $p$  supera el umbral común de 0.05 utilizado para determinar la significancia estadística. No hay suficiente evidencia para rechazar la hipótesis nula.