



Módulo 4. Parte II. Introducción a la Estadística Descriptiva con R

Solución de los Ejercicios

Autor: Raquel Dormido Canto

Actualizado Enero 2023

Contenido

EJERCICIO 1.	3
-------------------	---

EJERCICIO 1.

Disponemos de la siguiente información sobre los tipos de coches que se venden en dos ciudades distintas, Madrid y París.

Madrid			
Peso	Velocidad Máxima	Consumo	Marca
830.3	165	1.1	Seat
860.6	153	1.0	Renault
880.8	148	1.2	Mercedes
1050.5	169	1.6	Mercedes
1050.5	190	1.8	Mercedes
1080	195	1.9	Mercedes
1100.0	155	11.5	Seat
1100.0	176	NA	Renault
1110	188	2.3	Mercedes
1120	176	1.9	Seat
1130	186	2.4	Renault
1140	178	2.1	Seat
1140	178	2.05	Mercedes
1170	162	1.9	Mercedes
1200	176	2.2	Renault
1290	174	3.3	Renault
1290	200	2.1	Seat

Tabla 1: Tabla de datos del Ejercicio 1

París			
Peso	Velocidad Máxima	Consumo	Marca
1330.5	198	2.7	Renault
1370.6	163	2.5	Mercedes
1380	147	2.4	Mercedes
1400	179	NA	Renault
1420.7	184	3.1	Renault
1450	170	3.6	Mercedes
1600	166	3.8	Mercedes
1630	177	4.2	Mercedes
1730.5	186	5.5	Seat
1750	189	5.5	Seat
1790.8	184	5.8	Renault
1800	184	NA	Mercedes
1800.2	184	5.1	Seat
2600.1	200	7.7	Mercedes

Tabla 2: Tabla de datos del Ejercicio 1

Se pide:

- Crear dos estructuras de datos denominadas Madrid y Paris que contengan la información que aparece en las Tablas 1 y 2.
- Representar la variable Marca mediante un diagrama de sectores en cada ciudad. Incluir un título descriptivo en cada gráfico.
- Representar la variable Velocidad Máxima mediante un histograma en cada ciudad.
- ¿Existe algún dato atípico en la variable Peso en la ciudad de Madrid? ¿Y en París?
- Determinar cuál es el valor máximo del 30% de los menores Pesos de los coches de Madrid. Calcular también el valor mínimo del 26% de las Velocidades Máximas mayores de los coches de París.
- ¿En qué región presentan los coches una Velocidad Máxima media mayor?
- Estudia la asimetría y la curtosis de la variable Peso en Madrid e interpreta el resultado.

Solución Ejercicio 1

- a) Crear dos estructuras de datos denominadas Madrid y Paris que contengan la información que aparece en las Tablas 1 y 2.

```
#Estructuras de datos
Peso<- c(830.3,860.6,880.8,1050.5,1050.5,1080,1100.0,1100.0,1110,1120,
1130,1140,1140,1170,1200,1290,1290)
Velmax<-c(165,153,148,169,190,195,155,176,188,176,186,178,178,162,176,
174,200)
Consumo<-c(1.1,1.0,1.2,1.6,1.8,1.9,11.5,NA,2.3,1.9,2.4,2.1,2.05,1.9,2.
2,3.3,2.1)
Marca<-c("Seat","Renault","Mercedes","Mercedes","Mercedes","Mercedes",
"Seat","Renault","Mercedes","Seat","Renault","Seat","Mercedes","Merced
es","Renault","Renault","Seat")
Madrid<- data.frame (Peso, Velmax, Consumo, Marca)

Peso2<-c(1330.5,1370.6,1380,1400,1420.7,1450,1600,1630,1730.5,1750,179
0.8,1800,1800.2,2600.1)
Velmax2<-c(198,163,147,179,184,170,166,177,186,189,184,184,184,200)
Consumo2<-c(2.7,2.5,2.4,NA,3.1,3.6,3.8,4.2,5.5,5.5,5.8,NA,5.1,7.7)
Marca2<-c("Renault","Mercedes","Mercedes","Renault","Renault","Mercede
s","Mercedes","Mercedes","Seat","Seat","Renault","Mercedes","Seat","Me
rcedes")
Paris<- data.frame (Peso2, Velmax2, Consumo2, Marca2)
```

- b) Representar la variable Marca mediante un diagrama de sectores en cada ciudad. Incluir un título descriptivo en cada gráfico.

```
#Gráfico por sectores de la variable cualitativa Marca
#Madrid
tabla_marcas_madrid<-table(Madrid$Marca)
pie(tabla_marcas_madrid,col=rainbow(6),main=c("Variedad de Marcas de c
oche de Madrid"))
#Paris
tabla_marcas_paris<-table(Paris$Marca)
pie(tabla_marcas_paris,col=rainbow(6),main=c("Variedad de Marcas de co
che de Paris"))
```

Variedad de Marcas de coche de Madrid

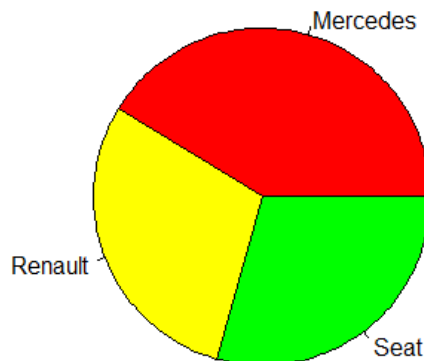


Diagrama de sectores de Marcas de Madrid

Variedad de Marcas de coche de Paris

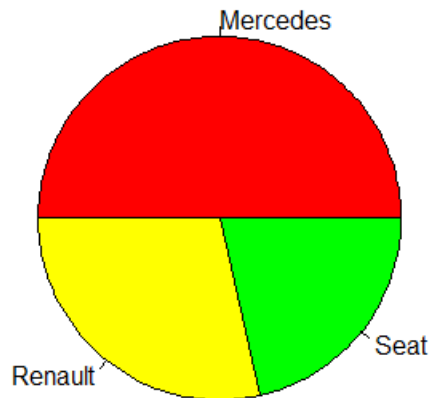


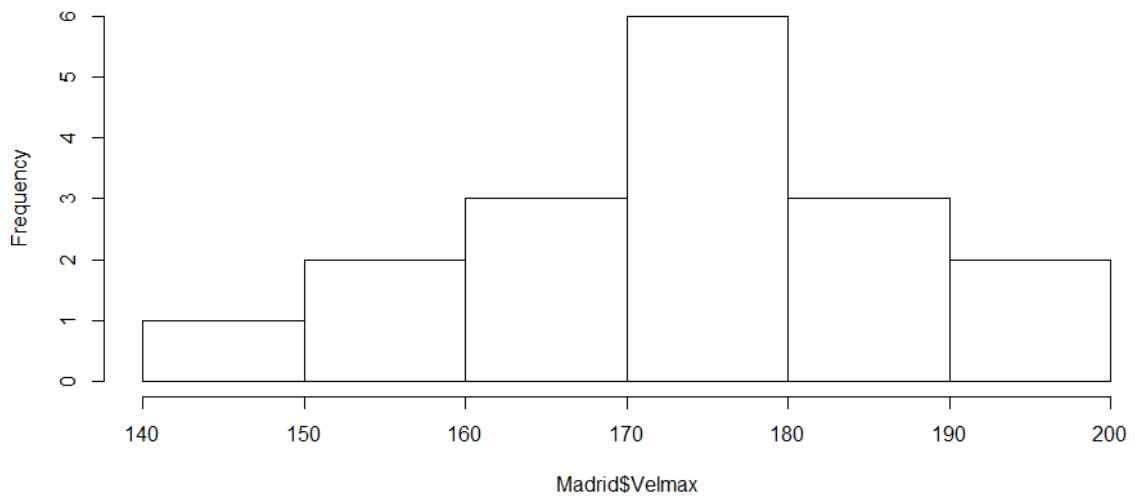
Diagrama de sectores de Marcas de Paris

c) Representar la variable Velocidad Máxima mediante un histograma en cada ciudad.

```
#Histograma Velocidad máxima Madrid  
hist(Madrid$Velmax,main="Distribución de coches de Madrid según su velocidad máxima")
```

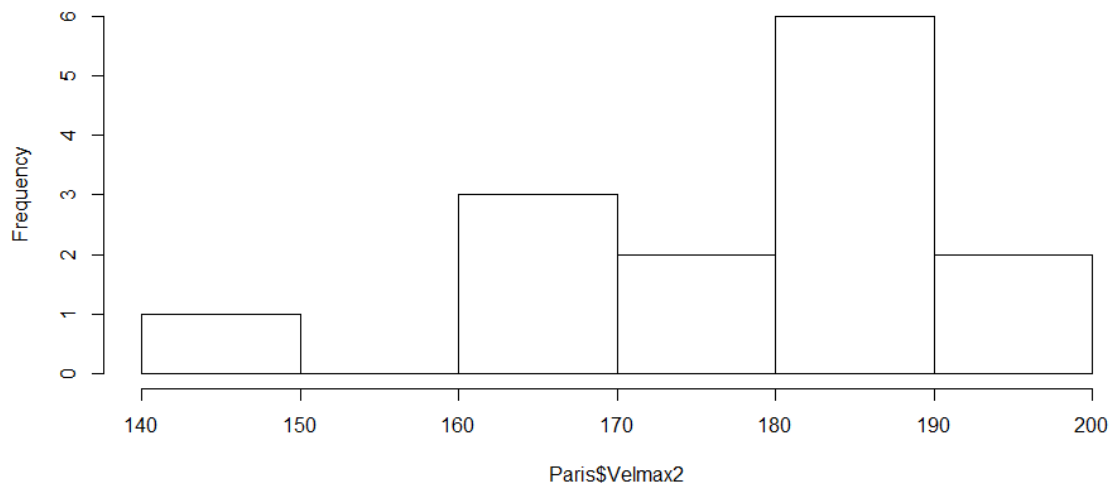
```
#Histograma Velocidad máxima Paris  
hist(Paris$Velmax2,main="Distribución de coches de Paris según su velocidad máxima")
```

Distribución de coches de Madrid según su velocidad máxima



Histograma de velocidad máxima de Madrid

Distribución de coches de París según su velocidad máxima

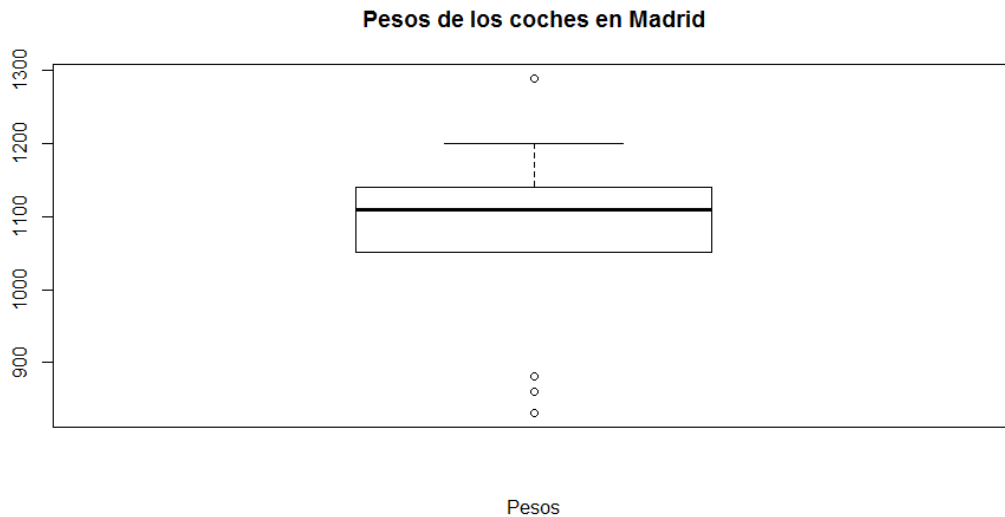


Histograma de velocidad máxima de París

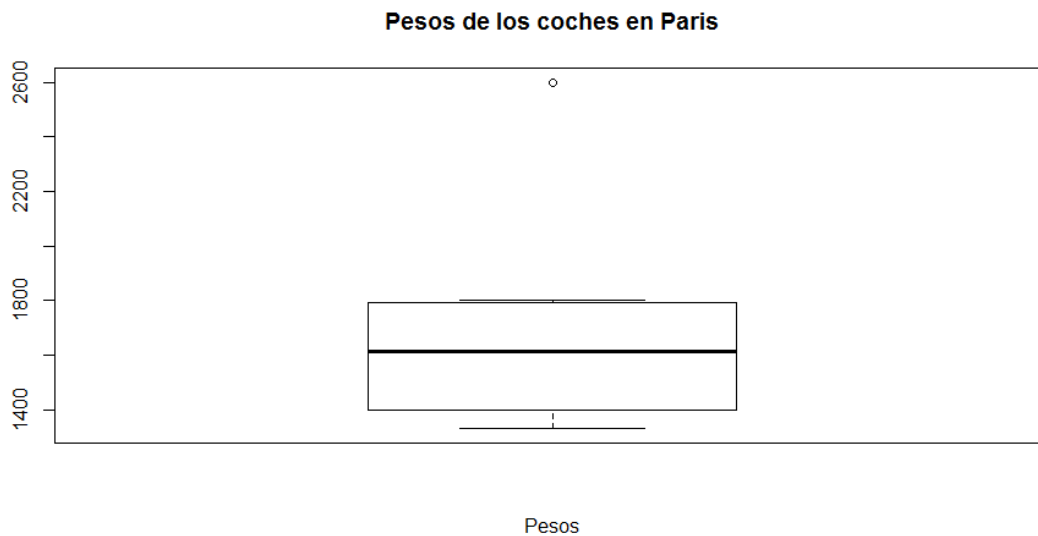
d) ¿Existe algún dato atípico en la variable Peso en la ciudad de Madrid? ¿Y en París?

Para responder a esta pregunta, vamos a calcular los gráficos de caja y bigotes para la variable Peso en ambas ciudades.

```
#Gráficos de caja y bigotes para la variable Peso en ambas ciudades.  
boxplot(Madrid$Peso, xlab="Pesos", main = "Pesos de los coches en Madrid")  
boxplot(Paris$Peso2, xlab="Pesos", main = "Pesos de los coches en París")
```



Caja y bigotes de Peso (Madrid)



Caja y bigotes de Peso (París)

Hay cuatro valores atípicos (datos más allá de los extremos de los bigotes) para la variable Peso de Madrid y uno para la variable Peso de París.

- e) Determinar cuál es el valor máximo del 30% de los menores Pesos de los coches de Madrid. Calcular también el valor mínimo del 25% de las Velocidades Máximas mayores de los coches de París.

Los valores que nos están pidiendo son el percentil 30 de la variable Peso en Madrid y el percentil 75 de la variable Velocidad Máxima en París, respectivamente.

```
#Percentil 30 de la variable Peso en Madrid  
quantile(Madrid$Peso, probs=c(.30))  
#Percentil 75 de la variable Velocidad Máxima en París  
quantile(Paris$Velmax2, probs=c(.75))
```

Percentil 30 de Peso en Madrid:

```
> quantile(Madrid$Peso, probs=c(.30))  
30%  
1074.1
```

Percentil 75 de Velocidad Máxima en París:

```
> quantile(Paris$Velmax2, probs=c(.75))  
75%  
185.5
```

f) ¿En qué región presentan los coches una Velocidad Máxima media mayor?

A la vista de las siguiente medias la Velocidad Máxima media es mayor en París:

```
> mean(Madrid$Velmax)  
[1] 174.6471  
> mean(Paris$Velmax2)  
[1] 179.3571
```

g) Estudia la asimetría y la curtosis de la Peso en Madrid e interpreta el resultado

```
> skewness(Madrid$Peso, na.rm = TRUE)  
[1] -0.5598267  
> kurtosis(Madrid$Peso, na.rm = TRUE)  
[1] -0.4451126
```

Como el coeficiente de asimetría es menor que 0, concluiremos que la distribución de la variable Peso en Madrid es asimétrica a la izquierda. Igualmente, dado que el coeficiente de curtosis es también negativo, la distribución de esta variable es menos apuntada que la distribución normal es decir, platicúrtica.