Tarea 1. Programa en R

Diana Laura Javier Garcia

9/9/2020

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

El objetivo de esta tarea es analizar los parámetros estadísticos de un grupo de datos ordenados en los campos, o variables, consumo, cilindrada (CC), potencia (CV), peso y aceleración.

Para elegir el par de variables de estudio al azar se generó un procedimiento en Excel el cual se explica en el documento anexo. De acuerdo con el procedimiento descrito se eligieron las variables consumo y aceleración.

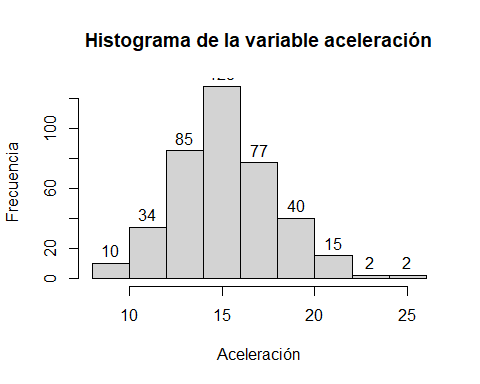
El código en R comienza con la lectura de los datos.

autos<-read.table("autos.txt",header = T)  
attach(autos)

# Parámetros estadísticos de la variable aceleración

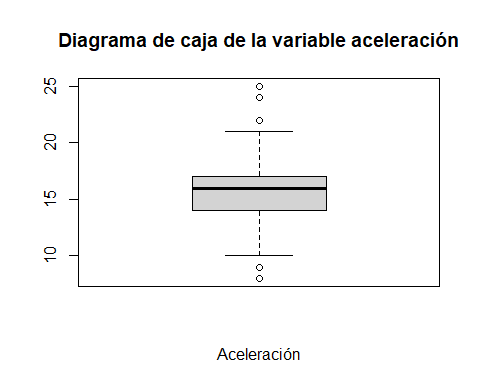
Para obtener el histograma se emplea el siguiente código:

hist(Acel, main = "Histograma de la variable aceleración", ylab = "Frecuencia", xlab = "Aceleración", labels = TRUE)



Mientras que para obtener el diagrama de caja se empleó la instrucción:

boxplot(Acel, main ="Diagrama de caja de la variable aceleración", xlab = "Aceleración", labels = TRUE)



Además, el programa nos permite conocer el resumen de los parámetros estadísticos con el siguiente código.

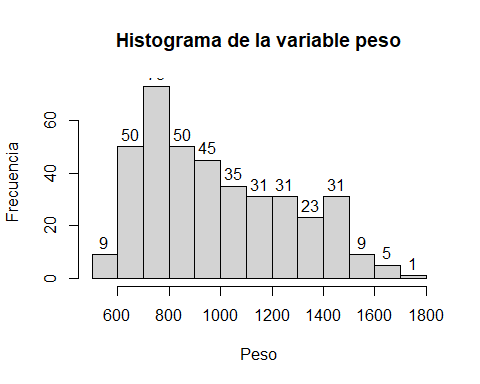
summary(Acel)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 8.00 14.00 16.00 15.67 17.00 25.00

# Parámetros estadísticos de la variable peso

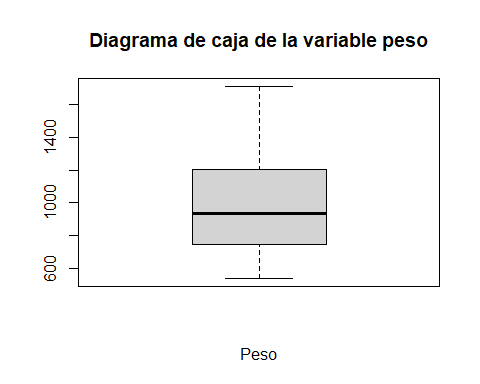
En esta sección se muestra el histograma de la variable peso, el cual se obtuvo con la siguiente instrucción:

hist(Peso, main = "Histograma de la variable peso", ylab = "Frecuencia", xlab = "Peso", labels = TRUE)



Por otra parte, se muestra el diagrama de caja de la variable peso, el cual se obtuvo con la siguiente instrucción:

boxplot(Peso, main = "Diagrama de caja de la variable peso", xlab = "Peso", labels = TRUE)



Además, el programa nos permite conocer el resumen de los parámetros estadísticos con el siguiente código.

summary(Peso)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 537.0 742.0 935.0 990.6 1203.0 1713.0

Finalmente, se muestra el resumen de todas las variables:

summary(autos)

## ï..Consumo CC CV Peso   
## Min. : 5.00 Min. :1114 Min. : 46.0 Min. : 537.0   
## 1st Qu.: 8.00 1st Qu.:1721 1st Qu.: 75.0 1st Qu.: 742.0   
## Median :10.00 Median :2474 Median : 93.0 Median : 935.0   
## Mean :11.22 Mean :3179 Mean :104.2 Mean : 990.6   
## 3rd Qu.:13.00 3rd Qu.:4293 3rd Qu.:125.0 3rd Qu.:1203.0   
## Max. :24.00 Max. :7456 Max. :230.0 Max. :1713.0   
## NA's :2   
## Acel   
## Min. : 8.00   
## 1st Qu.:14.00   
## Median :16.00   
## Mean :15.67   
## 3rd Qu.:17.00   
## Max. :25.00   
##

# Análisis de resultados

En la sección anterior se presentó el histograma y el diagrama de caja de las variables aceleración y peso. Con base en ello y en el resumen de datos se realizan las siguientes observaciones:

Aceleración:

1.- La aceleración tiene un promedio de .

2.- El primer y tercer cuartil tienen valores de 14 y 17, por lo tanto la diferencias de estos (conocido como rango intercuartílico) permite mostrar que los datos no presentan una dispersión importante.

3.- El diagrama de caja (bigotes) muestra que el valor mínimo y máximo de la aceleración es 8 y 25, por ello estos valores extremos deben ser investigados con importancia, ya que pueden aportar información importante.

4.- El primer cuartil indica que el primer de los datos se encuentran entre 8 y 14. Mientras que el de los datos se encuentran entre el valor mínimo de 8 y la mediana con valor de 16. Asimismo, el tercer cuartil que separa al de los datos ordenados de mayor a menor se encuentran entre 8 y 17.

5.- El histograma de la variable aceleración tiene una moda, con una distribución simétrica, es decir, presenta una distribución normal.

Peso

1.- Con base en los resultados presentados, se encontró que el peso promedio de los autos estudiados es de 990.6 kg.

2.- A partir del diagrama de caja se observa que existe una dispersión importante en los datos, sobre todo si se considera que el valor máximo (1713.0) se encuentra alejado de la mediana (935).

3.- El primer cuartil indica que el primer de los datos se encuentran entre 537.0 y 742.0. Mientras que el de los datos se encuentran entre el valor mínimo de 537.0 y la mediana con valor de 935.0. Asimismo, el tercer cuartil que separa al de los datos ordenados de mayor a menor se encuentran entre 537 y 1203.0.

4.- Con respecto a la distribución de los datos por frencuencias, se observa que los datos con mayor frencuencia se encuentran entre 600 y 1000, donde el intervalo entre 700 y 800 presenta la mayor frencuencia.

5.- Se observa además, que la distribución de los datos por intervalos tiene una moda, y presenta una distribución sesgada a la izquierda.

# Conclusiones

En este trabajo se lograron los objetivos establecidos. El programa R permitió realizar un análisis estadístico de los datos de aceleración y peso de varios autos. El análisis de los parámetros estadisticos (media, moda, mediana, primer cuartil, segundo cuartil, mínimo y máximo) así como del histograma y diagrama de caja permiten concluir lo siguiente:

Con respecto a los datos de aceleración se observó que presentan una distribución simétrica (normal), donde el promedio tiene un valor de . Debido a que la diferencia entre el tercer y primer cuartil tiene un valor de 3, se considera que los datos no presentan una dispersión importante.

Por otra parte, con respecto a los datos de peso se observó que estos presentan una distribución sesgada a la izquierda, lo cual sugiere que la mayoria de los autos estudiados tienen un peso bajo (menor a 1000 kg). Sin embargo, el diagrama de caja permitió mostrar que existen valores extremos comparados con la media, lo cual se puede observar gráficamente en los bigotes del diagrama.

De manera general, se observaron datos extremos, por ello es importante tomar en cuenta estos valores, ya que pueden aportar información relevante o bien, pueden ser mediciones erroneas.