**TAREA #1**

Martin Emilio Marin Gutiérrez

Facultad de Ciencias, Universidad Nacional de Colombia-Sede Medellín.

IAM.

Medellín, Antioquia.

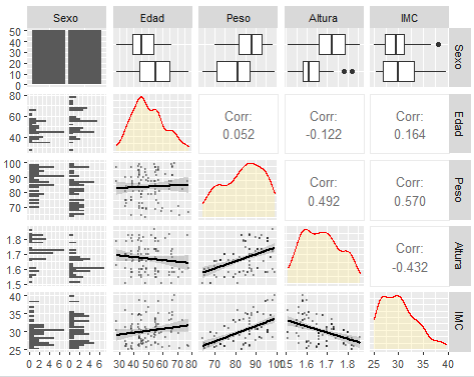
Marzo 2022

1. Realizar un resumen descriptivo de sus datos de la forma m ́as adecuada posible para

lo cu ́al se le sugiere que utilice ayudas como las siguientes:

(a) Ayudas gr ́afcas como por ejemplo: Diagramas de dispersi ́on de variables, Gr ́aficas

de cajas y bigotes o box-plot, histogramas, etc. etc.

gráfica 1

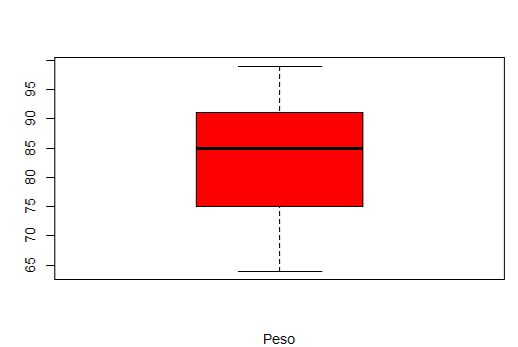
Nos apoyamos en Rstudio para poder hacer el análisis descriptivo. Como podemos observar en el gráfico hay una aparente relación lineal positiva entre el peso y el IMC, lo que nos dice que a mayor peso mayor IMC, esta misma tendencia parece reflejarse en entre la altura y el peso, es decir que a mayor altura mayor peso.

También podemos observar una correlación lineal negativa entre el IMC y la altura, lo que quiere decir que entre mayor sea la altura del encuestado, menor será su IMC

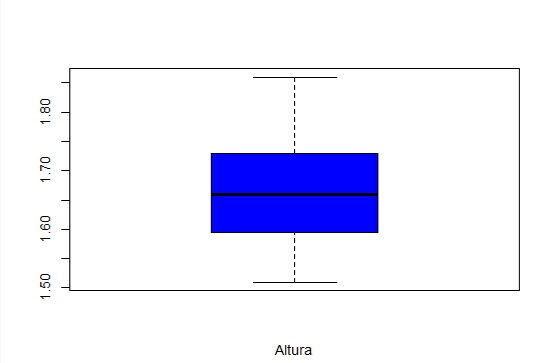
(b) Para 3 de las variables de su base, comente los aspecto m ́as relevantes de los

datos. Comente sobre medidas descriptivas relevantes de esas variables.

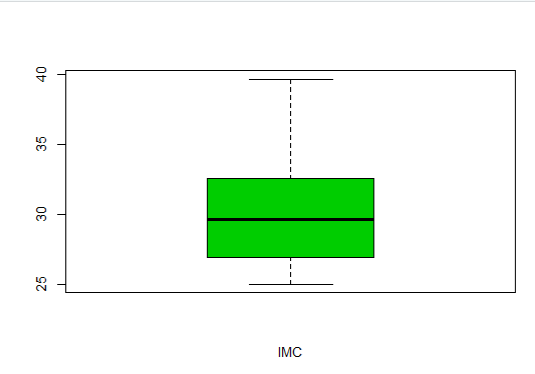
Trabajemos sobre las siguientes variables: Peso, Altura e IMC.

gráfica 2

En el caso de variable peso podemos observar que la media se encuentra aproximadamente en 83.6300 kilos y que no tenemos observaciones atípicas.

gráfica 3

En el caso de variable Altura podemos observar que la media se encuentra aproximadamente en 1.6688 metros y que no tenemos observaciones atípicas.

gráfica 4

En el caso de variable IMC podemos observar que la media se encuentra aproximadamente en 30.1054 y que no tenemos observaciones atípicas.

(c) Para las mismas 3 variables consideradas en el punto (b) realice an ́alisis d

descriptivos por categor ́ıas de una de sus variables cualitativas de sus datos.

-En el caso del peso podemos observar que al discriminarlo por sexo la media de en los hombres encuesados es mayor que en el caso de las mujeres y que que en el caso de las mujeres encuestadas hay una mayor dispercion e los datos(Lo podemos observar en el boxplot de la grafica 1)

-Para la altura podemos observar aspectos relevantes al hacer la misma discriminacion por sexo, podemos observar en el boxplot que en el caso de los hombres hay una mayor dispercion y tambien podemos ver que la media de altura en los hombres encuestados es mayor que la media de al aultra de las mijeres encuestadas.

-En el caso del IMC el bosxplot al discriminarlo por sexo no nos da mucha informacion, sus medias son parecidas y en el caso de las mujeres encuestadas podemos ver un dato atipico.

Igualmente pueda aca ayudarse de gráficos como box-plot.

2. Obtenga el vector de medias muestrales x, las matrices de varianzas-covarianzas

muestrales S y Sn y la matriz de correlaciones muestrales R de los datos. Comente

algunas entradas relevantes de dichos resultados. Pueden ayudarse con algunos

Gráficos apropiados.

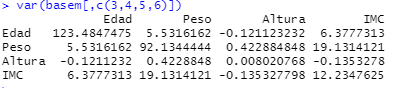
**Vector de medias**



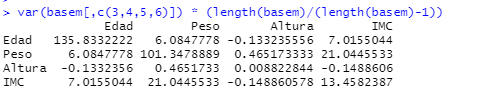
Con el apoyo de Rstudio concluimos que la edad promedio es de 49.9900 años, el peso promedio es 83.6300 kilos, un IMC promedio de 30.1054 y la altura promedio es de 1.6688 metros.

**Matriz de varianza-covarianza**

:



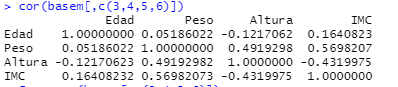
:



Lo que nos quiere decir que esta matriz de varianza-covarianza es que puede existir una cierta relación entre la edad y el peso según su covarianza=5.5316, tal cual como lo habíamos inferido en el análisis descriptivo, de la misma forma hay una aparente relación entre la altura y el peso según su covarianza=0.4228 y para el caso de el IMC con el peso también existe una aparente relación con una covarianza=7.0155.

3. Calcule la varianza total y la varianza generalizada de los datos. Comente:

Hallamos la ,matriz de correlación:



**Varianza total:**



Tomamos las diagonales de la matriz varianza-covarianza y las sumamos para hallar la varianza total=227.862

**Varianza generalizada:**



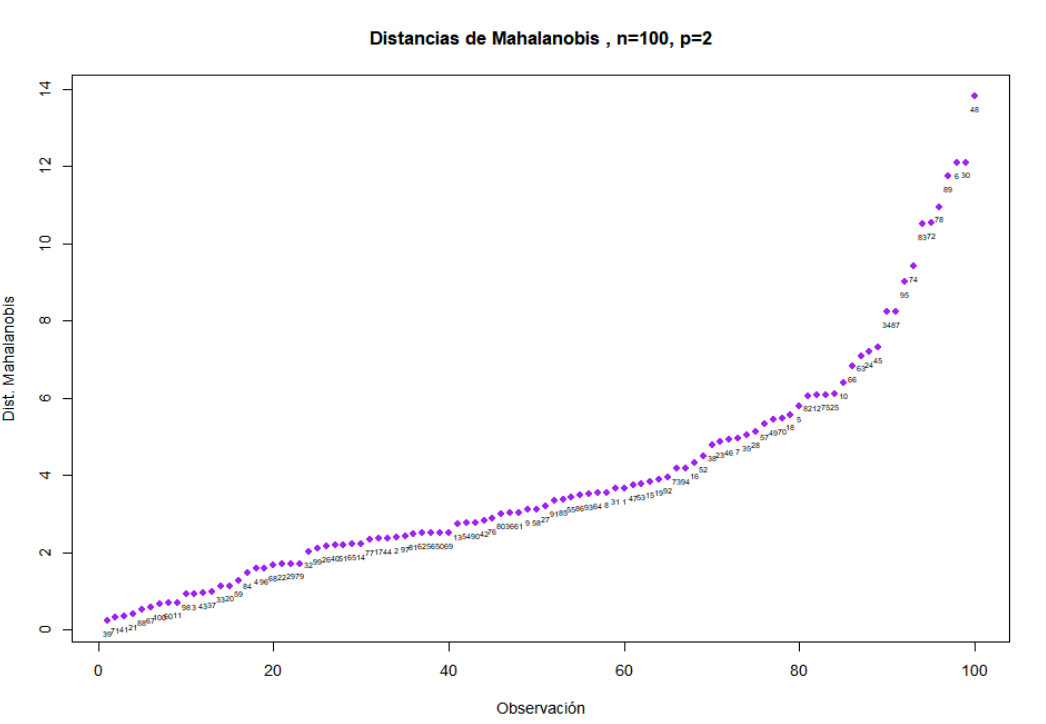
Por el valor que nos arroja la varianza generalizada podemos inferir que hay mucha dispersión en los datos.

4. Calcule la distancia de Mahalanobis entre cada dato u observación y el centroide

de los datos (vector de medias). ¿Qué observaciones están a mayor distancia? Qu ́e

¿Las observaciones están a menor distancia? Se podr ́a usar para esta matriz de datos la

distancia euclidiana? Por qué sí o porqué no?.

gráfica 5

Como podemos observar en el anterior gráfico entre las observaciones que se encuentra mayor distancia están:48, 30, 6 y 89; y en las que se encuentra menor distancia son:39, 71, 41 y 21.

En cuanto a las distancias euclídea no es conveniente,” ya que cuando las coordenadas representan medidas sujetas a fluctuaciones aleatorias de diferentes magnitudes, como lo es en este caso, es preferible ponderar las coordenadas de acuerdo a su variabilidad”(fragmento del capítulo 1 de las notas de clase).

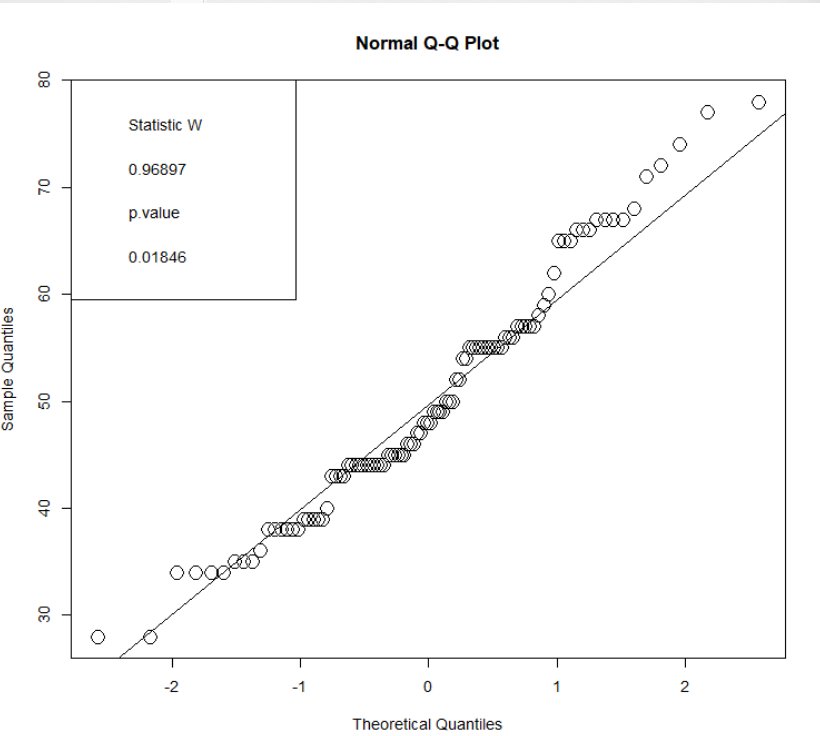
5. Para dos de las variables de su tabla de datos realice una evaluación de normalidad

univariada para cada una de ellas, siguiendo las distintas opciones que sean de su

conocimiento, como por ejemplo ayudas gráficas o pruebas formales que conozca.

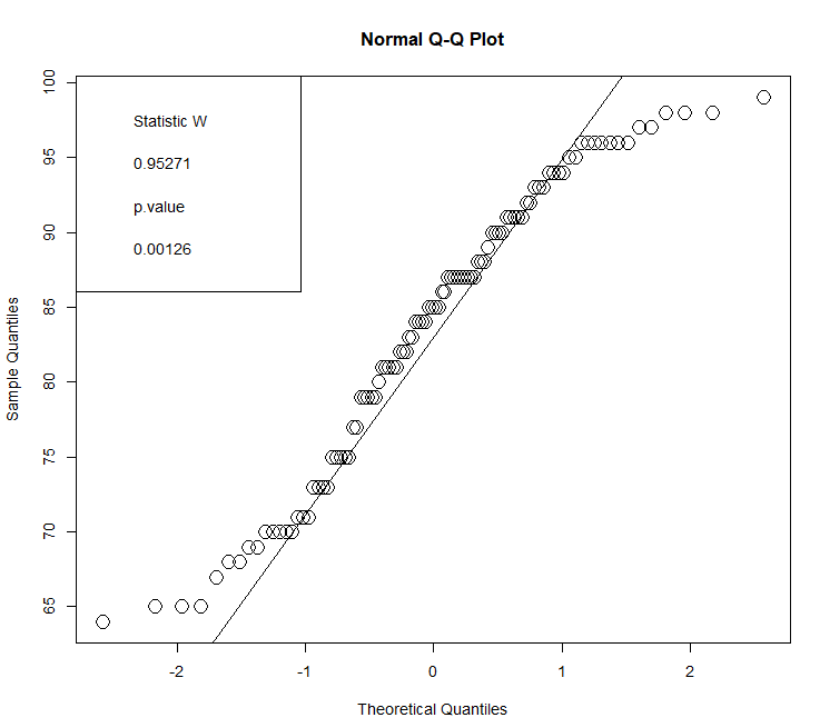
Comente los resultados de cada evaluación.

Realizamos una prueba de shapiro wilk a cada variable de nuestra base de datos, asi:

 gráfica 6

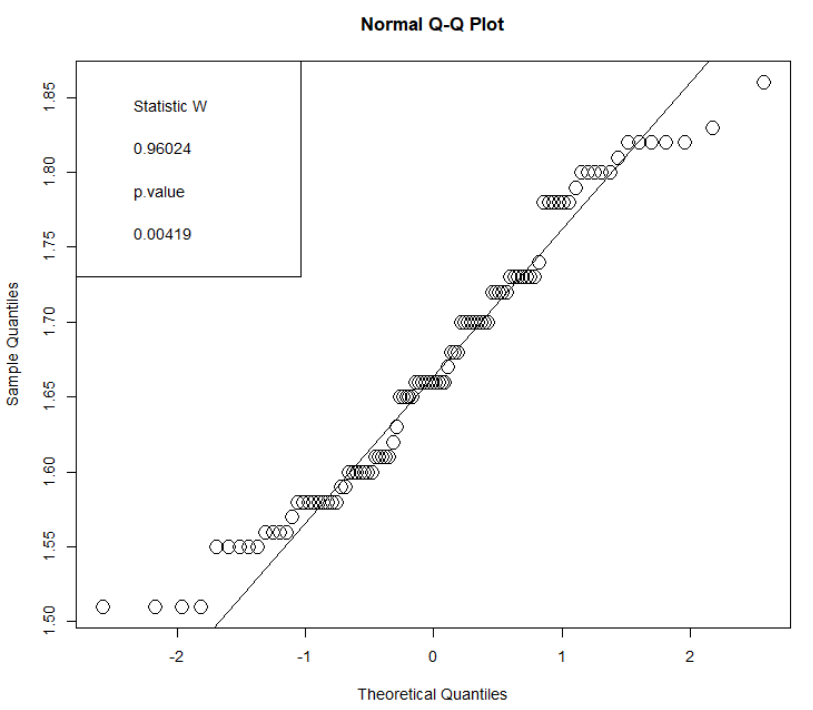
Tenemos que el p-valor es igual a 0.01846 y como p-valor. Por lo tanto se rechaza la hipótesis nula.

Podemos concluir que no sigue una distribución normal.

gráfica 7

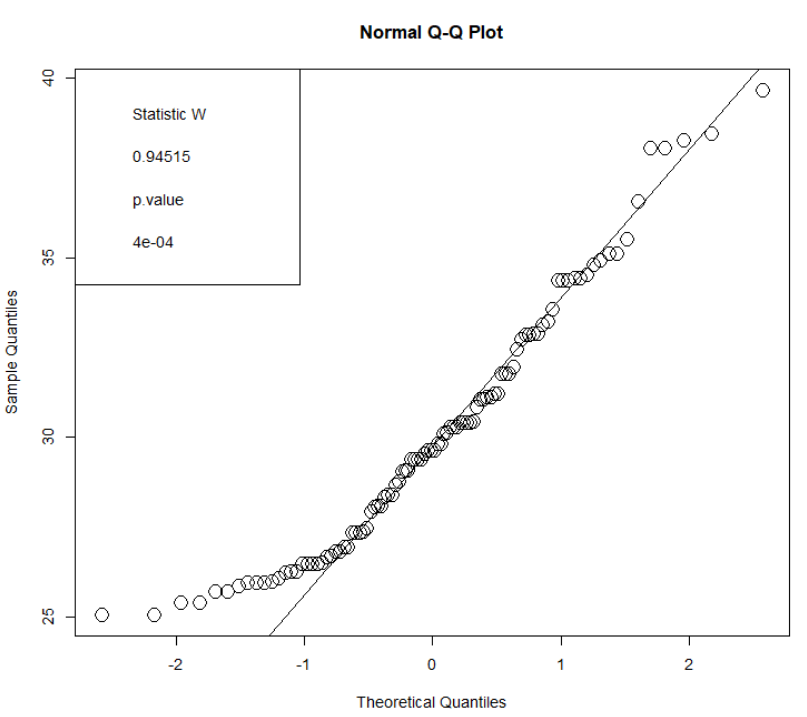
Tenemos que el p-valor es igual a 0.00126y como p-valor. Por lo tanto se rechaza la hipótesis nula.

Podemos concluir que no sigue una distribución normal.

gráfica 8

Tenemos que el p-valor es igual a 0.00419 y como p-valor. Por lo tanto se rechaza la hipótesis nula.

Podemos concluir que no sigue una distribución normal.

gráfica 9

Tenemos que el p-valor es igual a 4e-04 y como p-valor. Por lo tanto se rechaza la hipótesis nula.

Podemos concluir que no sigue una distribución normal.

**Bibliografía**

<https://jllopisperez.com/2017/02/22/base-de-datos-de-obesidad/comment-page-1/>

(Base de datos)