Assignment-No.2

December 6, 2020

1 Probabilty & Statistics

[2]:
```
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline

import numpy as np
import pandas as pd
```

[31]:
```
url = "data/nhanes_2015_2016.csv"
da = pd.read_csv(url)
```

The data which we are going to analyse is taken from **NHANES** data set. We are going to give the analysis of the Age which is given by `RIDAGEYR` in the dataset.

1.1 Mean

In order to find mean of the age we can use the bultin function of `NUMPY` called `np.mean`.

[33]:
```
mean_age=np.mean(da.RIDAGEYR)
print("Mean of the AGE  = " , mean_age)
```

```
Mean of the AGE is =  48.05231037489102
```

1.2 Median

In order to find mean of the age we can use the bultin function of `NUMPY` called `np.median`.

[34]:
```
median_age=np.median(da.RIDAGEYR)
print("Median of the AGE = ", median_age)
```

```
Median of the AGE =  48.0
```

1.3 Q1

In order to find mean of the age we can use the bultin function of `NUMPY` called `np.quantile`. As Q3 is .25 of the data so we will specfiy it in the parameter.

[35]:
```
q1_age=np.quantile(da.RIDAGEYR, .25)
print("Q1 of the Age =  ", q1_age)
```

```
Q1 of the Age =   32.0
```

1.4 Q3

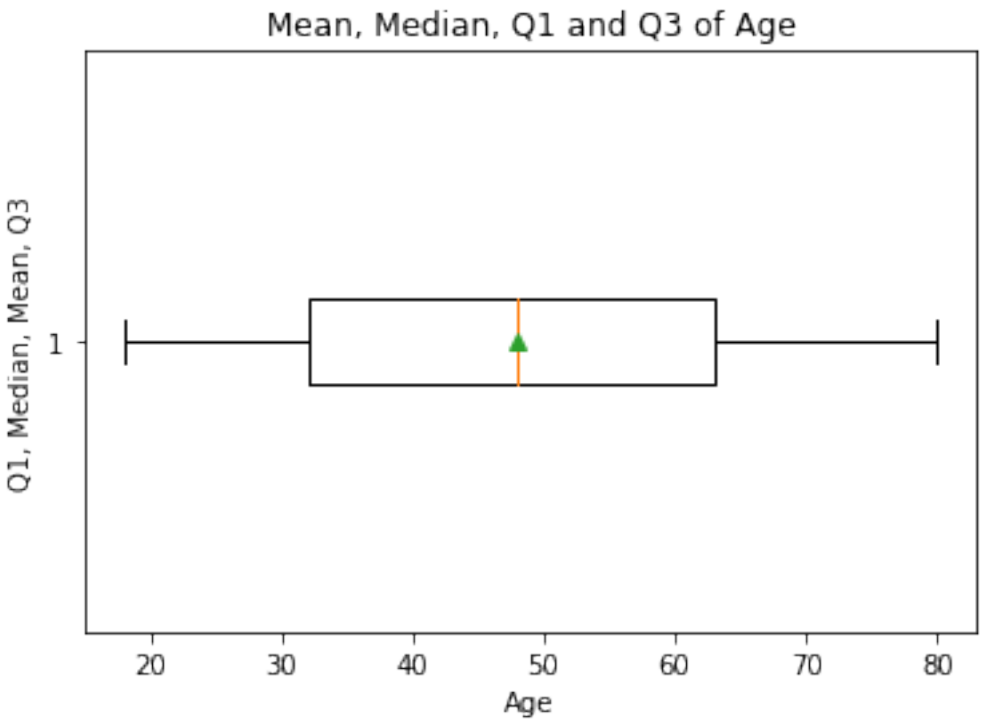
In order to find mean of the age we can use the bultin function of `NUMPY` called `np.quantile`. As Q3 is .75 of the data so we will specfiy it in the parameter.

[36]:
```
q3_age=np.quantile(da.RIDAGEYR, .75)
print("Q3 of the Age =  ", q3_age)
```

```
Q3 of the Age =   63.0
```

2 Box Plot of the three values.

[42]:
```
plt.boxplot(da.RIDAGEYR, showmeans=True, showfliers=True, vert=False)
plt.title("Mean, Median, Q1 and Q3 of Age")
plt.xlabel("Age")
plt.ylabel("Q1, Median, Mean, Q3")
plt.show()
```



In the above diagram the very first line before `20` shows the **minimum** value. The second line(vertical) shows the **First Quantile** value which `32`. The third line highlighted in **red** shows the `median` while the arrow shows `mean` value. The fourth line (3rd line of box) shows the **Third Quantile**.