|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continuous |
| Weight of Gold | Continuous |
| Distance between two places | Continuous |
| Length of a leaf | Continuous |
| Dog's weight | Continuous |
| Blue Color | Discrete |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Discrete |

Q1) Identify the Data type for the Following:

Q2) Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

|  |  |
| --- | --- |
| Data | Data Type |
| Gender | Nominal |
| High School Class Ranking | Ordinal |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Ratio |

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?

**ANS =** Three coins are tossed the total number of possible combinations are,

[HHH,HHT,HTH,THH,TTT,TTH,THT,HTT],

The number of combinations that have two heads and one tail are, 3 i.e

[HHT,HTH,THH]

P(two heads and one tails) = Number of desired outcomes/Total number of

possible combinations

= 3/8 = 0.375

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1
2. Less than or equal to 4
3. Sum is divisible by 2 and 3

**ANS** = Two dice are rolled the total number of possible outcomes are,

[(1,1)(1,2)(1,3)(1,4)(1,5)(1,6)(2,1)(2,2)(2,3)(2,4)(2,5)(2,6)(3,1)(3,2)

(3,3)(3,4)(3,5)(3,6)(4,1)(4,2)(4,3)(4,4)(4,5)(4,6)(5,1)(5,2)(5,3)(5,4)

(5,5)(5,6)(6,1)(6,2)(6,3)(6,4)(6,5)(6,6)].

1. Equal 1

ANS = Zero.

1. Less than or equal to 4

ANS = 6/36 = 1/6 = 0.166

1. Sum is divisible by 2 and 3

ANS = 6/36 = 1/6 = 0.166

Q5) A bag contains 2 red, 3 green and 2 blue balls. Two balls are drawn at random. What is the probability that none of the balls drawn is blue?

**ANS =** Total number of ball is (2+3+2=7).

So the probability of first ball not being blue is 5/7.

The probability of second ball not being blue assuming that 4/6.

The probability that neither ball drawn blue is (5/7) \* (4/6) = 20/42 = 10/21.

Q6) Calculate the Expected number of candies for a randomly selected child

Below are the probabilities of count of candies for children (ignoring the nature of the child-Generalized view)

|  |  |  |
| --- | --- | --- |
| CHILD | Candies count | Probability |
| A | 1 | 0.015 |
| B | 4 | 0.20 |
| C | 3 | 0.65 |
| D | 5 | 0.005 |
| E | 6 | 0.01 |
| F | 2 | 0.120 |

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

**ANS** =

Expected number of candies for a randomly selected child

= 1\*0.015 + 4\*0.20 + 3\*0.65 + 5\*0.005 + 6\*0.01 + 2\*0.120

= 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24

= 3.09

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

**ANS =**

Mean - Points = 3.596 , Score = 3.217 , Weigh = 17.848

Median - Points = 3.695 , Score = 3.325 , Weigh = 17.710

Mode - Points = 3.07 , Score = 3.44 , Weigh = 17.02

Variance - Points = 0.285 , Score = 0.957 , Weigh = 3.193

Standard Deviation - Points = 0.534 , Score = 0.978 , Weigh = 1.786

Range[MIN–MAX] - Points = [2.760- 4.930], Score = [1.513- 5.424], Weigh = [14.500- 22.900]

Q8) Calculate Expected Value for the problem below

1. The weights (X) of patients at a clinic (in pounds), are

108, 110, 123, 134, 135, 145, 167, 187, 199

Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

**ANS** = There are 9 patients so the probability of selecting each patient is 1/9,

Expected value = (1/9) (108+110+123+134+135+145+167+187+ 199)

= (1/9) (1308)

= 145.33

Expected value of the weight of the patient is 145.33

**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

**ANS =**

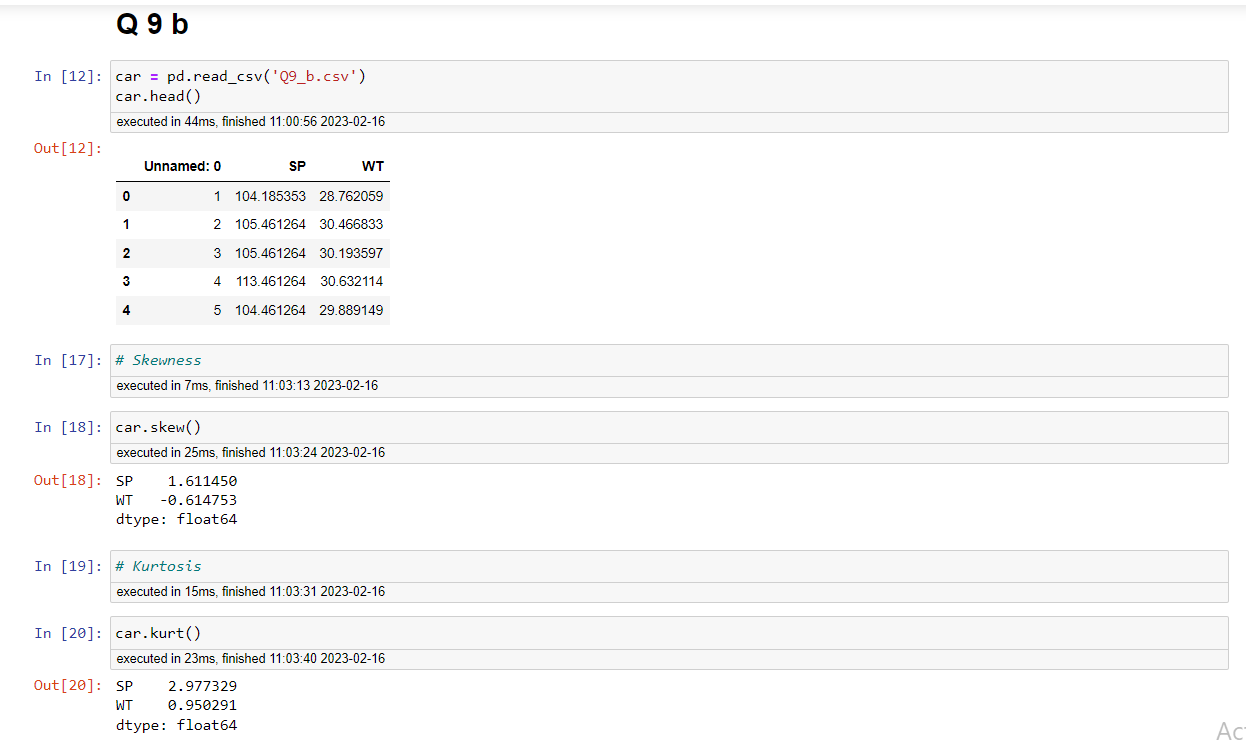


****

**SP and Weight(WT)**

**Use Q9\_b.csv**

**ANS =**



**Q10) Draw inferences about the following boxplot & histogram**



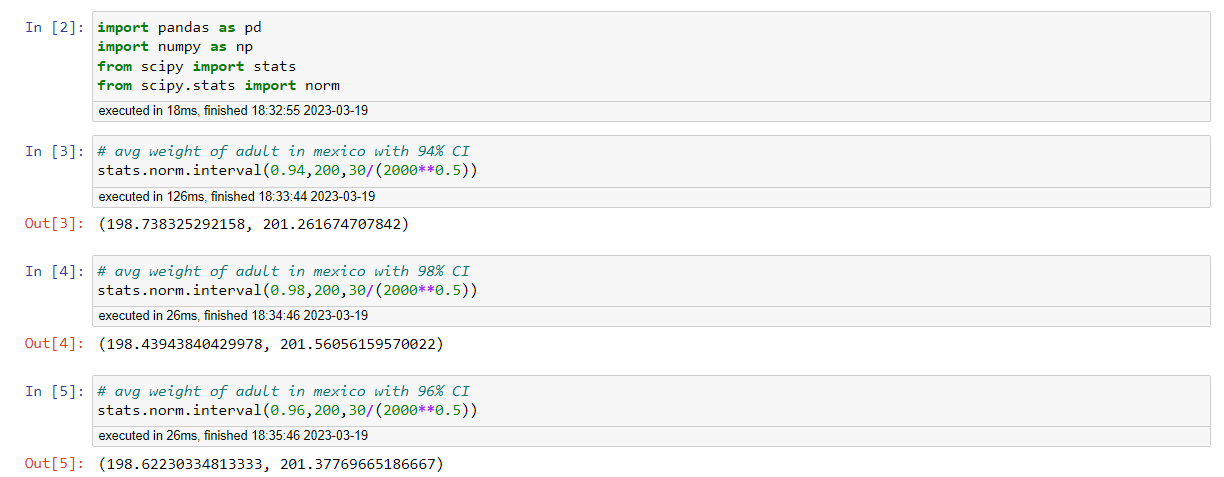
**ANS =** The above histogram is right skewed, the right tail is greater than left side and the mean > madian.



**ANS =** The above boxplot has an outliers on its upper extream

**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

**ANS =**



**Q12)** Below are the scores obtained by a student in tests

**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

1. Find mean, median, variance, standard deviation.

**ANS** = For Mean,

34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56/18

= 738/18 = 41

For Median, 40+41/2 = 81/2 = 40.5

For Variance, 434/17 = 25.52

For Standard Deviation, Square root of Variance(25.52) = 5.05

1. What can we say about the student marks?

**ANS =** We don’t have any outliers and by compairing mean and median

the data has slightly positive skewed.

Q13) What is the nature of skewness when mean, median of data are equal?

**ANS** = It states that data distribution is symmetrical and the distribution has zero skewness.

Q14) What is the nature of skewness when mean > median ?

**ANS** = When mean > median then the data distribution is positive skewed and the right tail is longer than left tail.

Q15) What is the nature of skewness when median > mean?

**ANS** = When median > mean then the data distribution is negative skewed and the left tail is longer than right tail.

Q16) What does positive kurtosis value indicates for a data ?

**ANS** = Positive kurtosis value indicates the peakness of the data and thick tail.

Q17) What does negative kurtosis value indicates for a data?

**ANS** = Negative kurtosis value indicates that lighter tail than normal distribution.

Q18) Answer the below questions using the below boxplot visualization.



What can we say about the distribution of the data?

**ANS** = The above data is not normally distributed.

What is nature of skewness of the data?

**ANS** = The skewness of the data is negative skewed.

What will be the IQR of the data (approximately)?

**ANS** = IQR=Upper quartile (Q3) - Lower quartile(Q1) = 18-10 = 8  
  
  
Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

**ANS =** Boxplot 1 & Boxplot 2 have no outliers. Both the boxplot shares the same median which is approximately in a range of 250 to 275. Both the boxplot are normally distributed.

Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

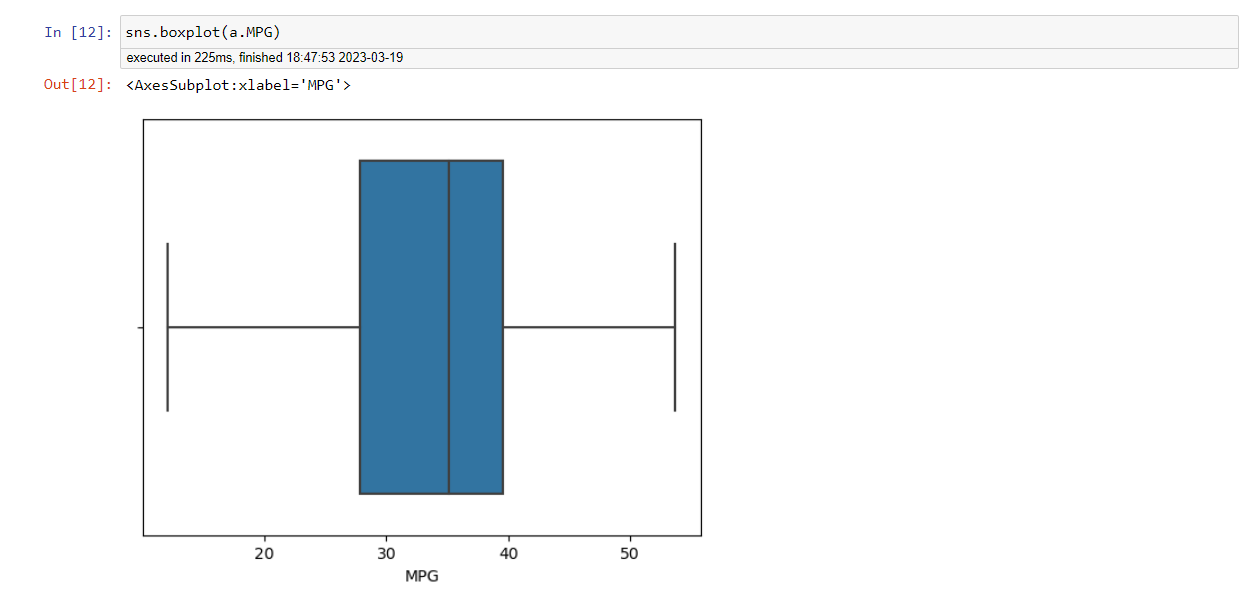
MPG <- Cars$MPG

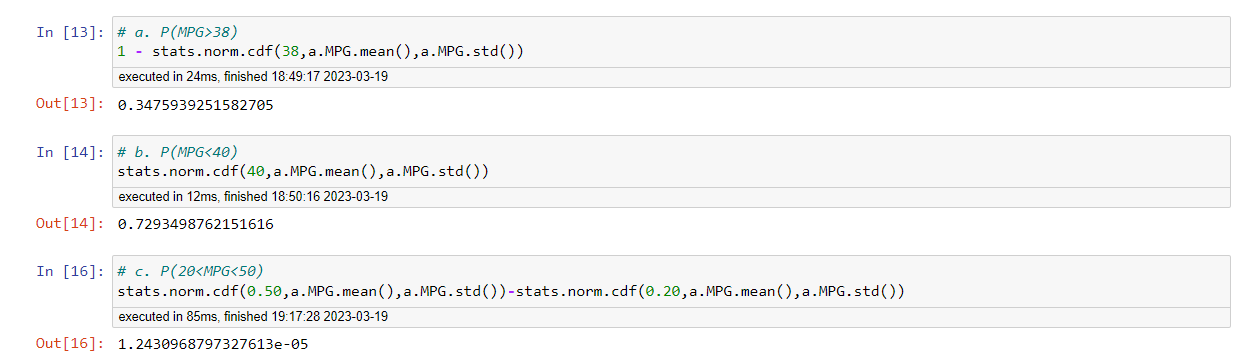
* 1. P(MPG>38)
  2. P(MPG<40)

c. P (20<MPG<50)

**ANS =**





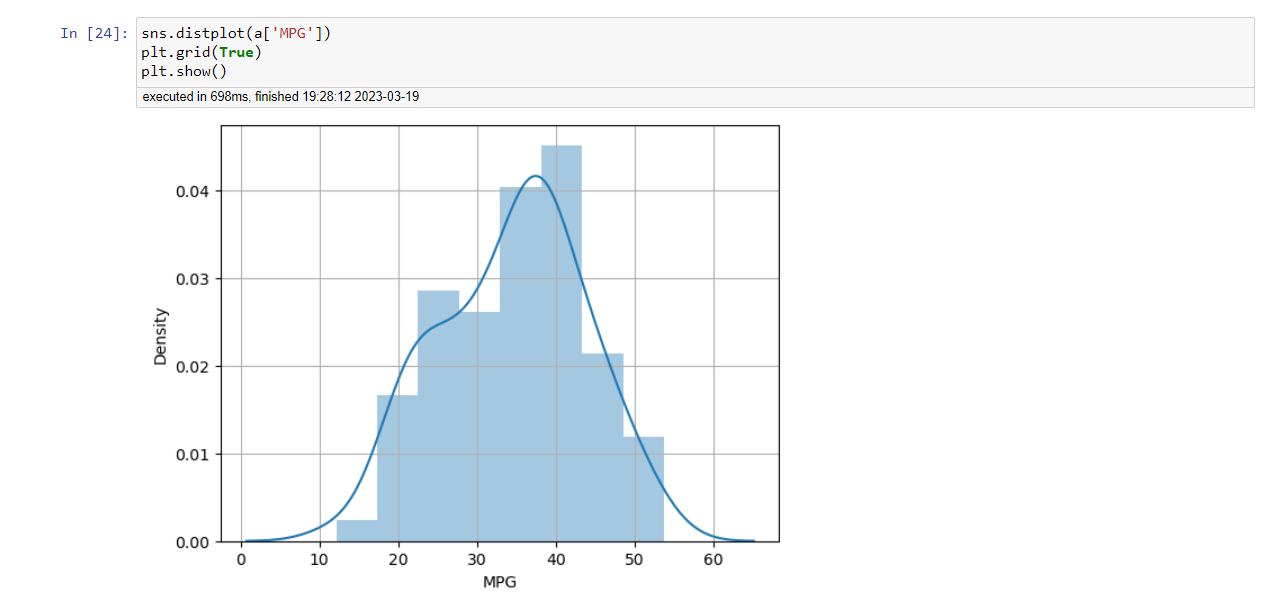
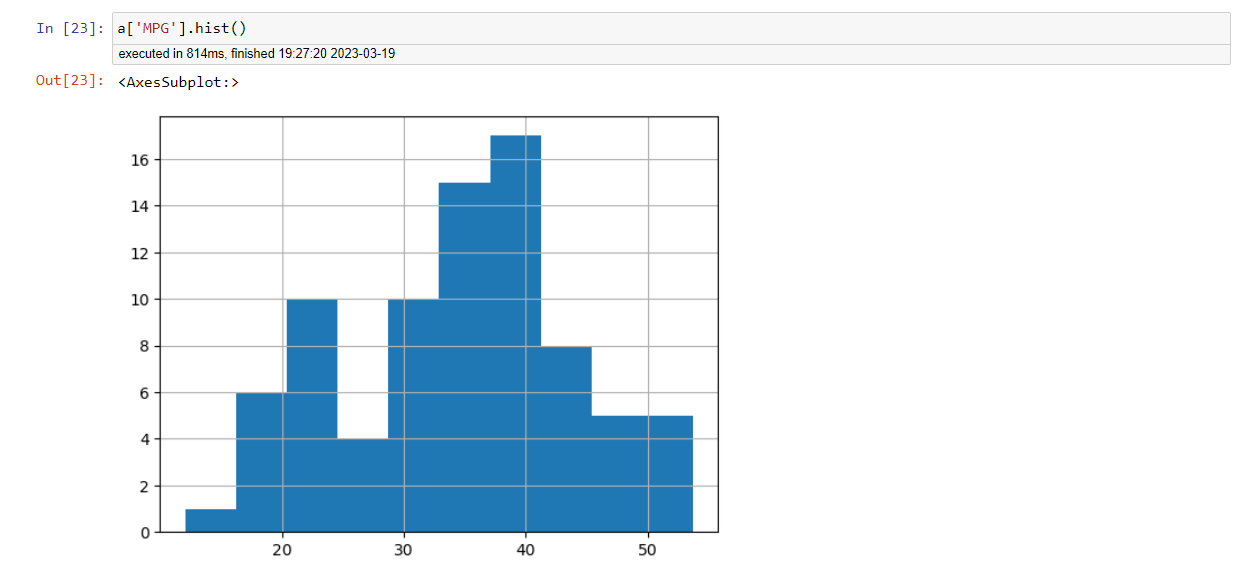


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**ANS =**



From above plot and value we can say that data is fairly normally distributed.

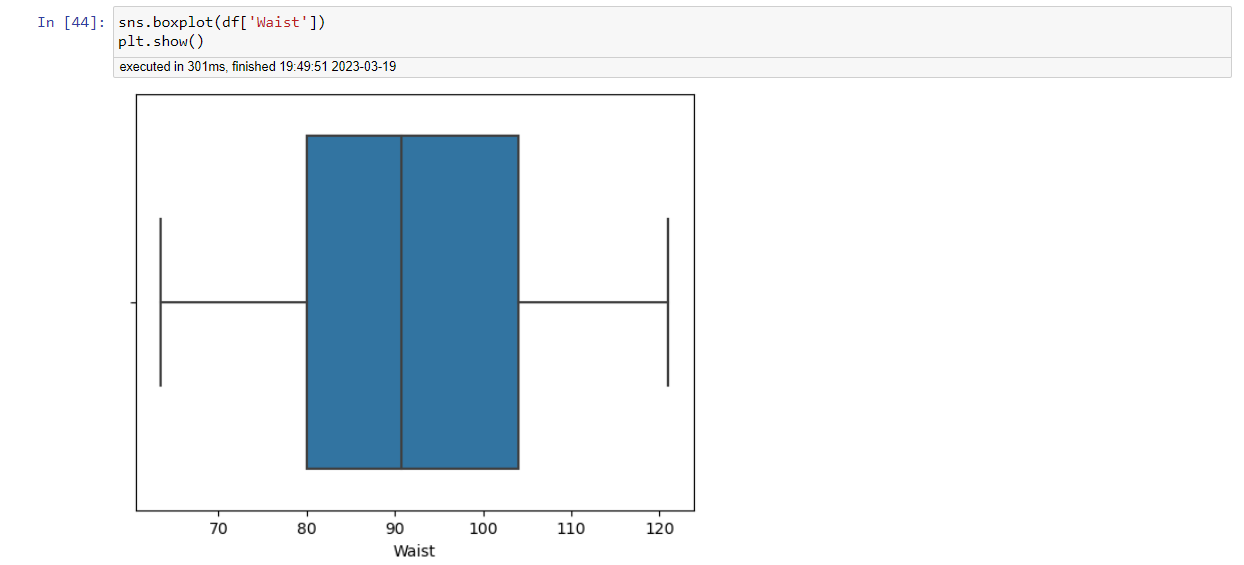
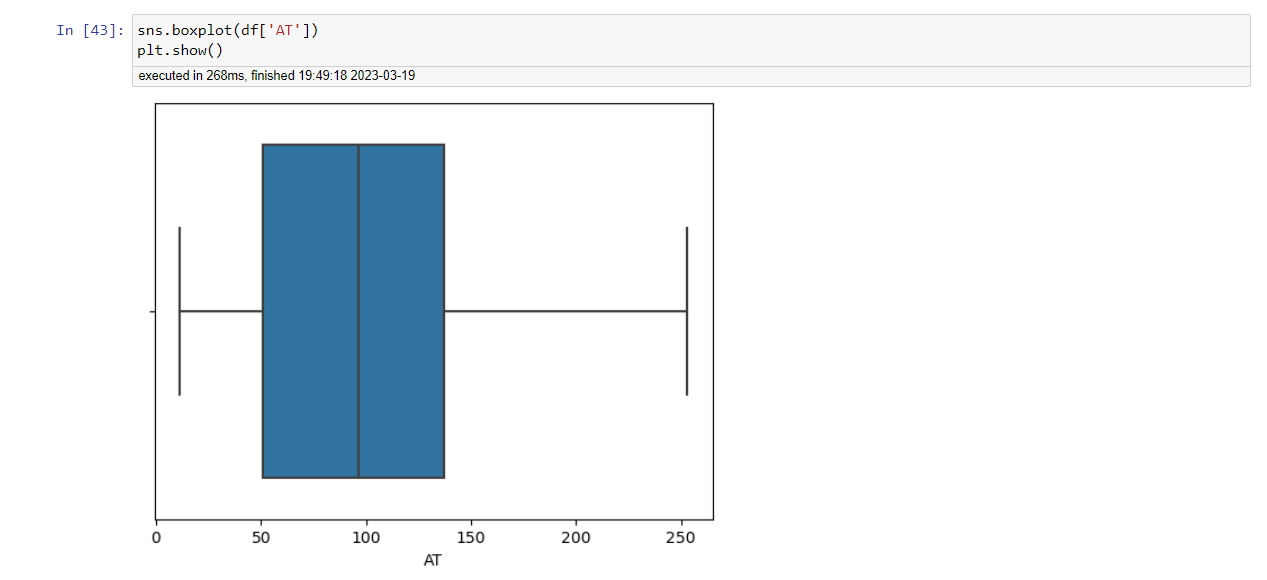
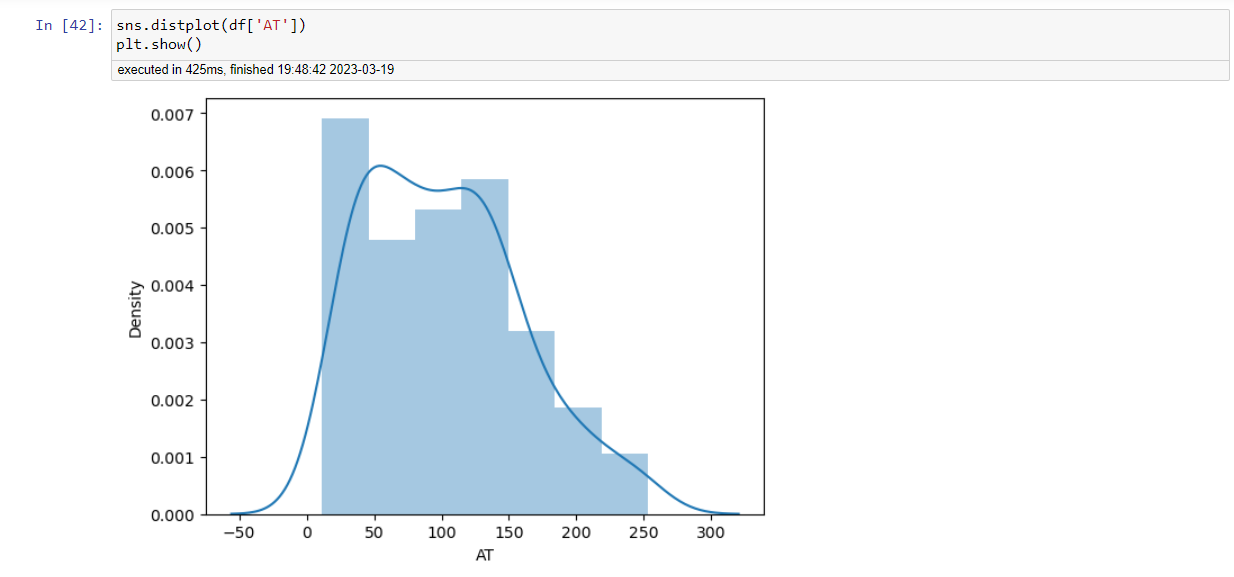
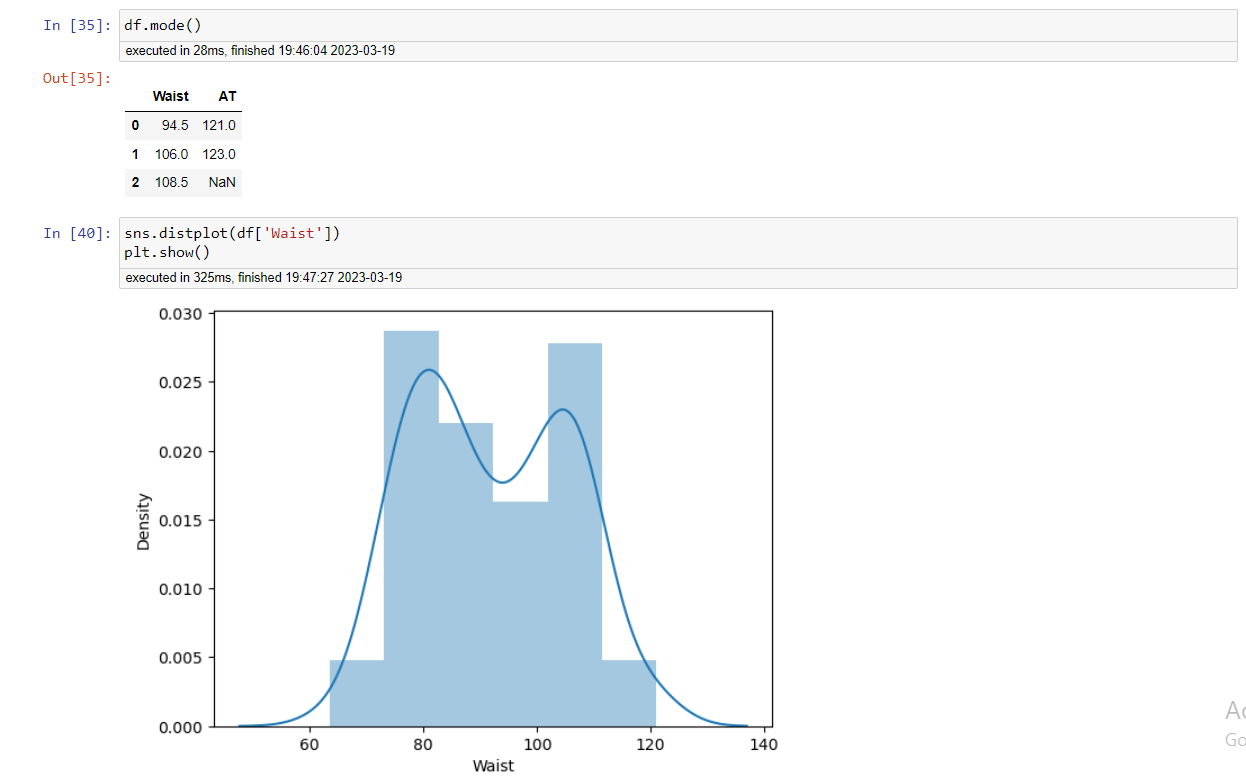
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

**ANS =**

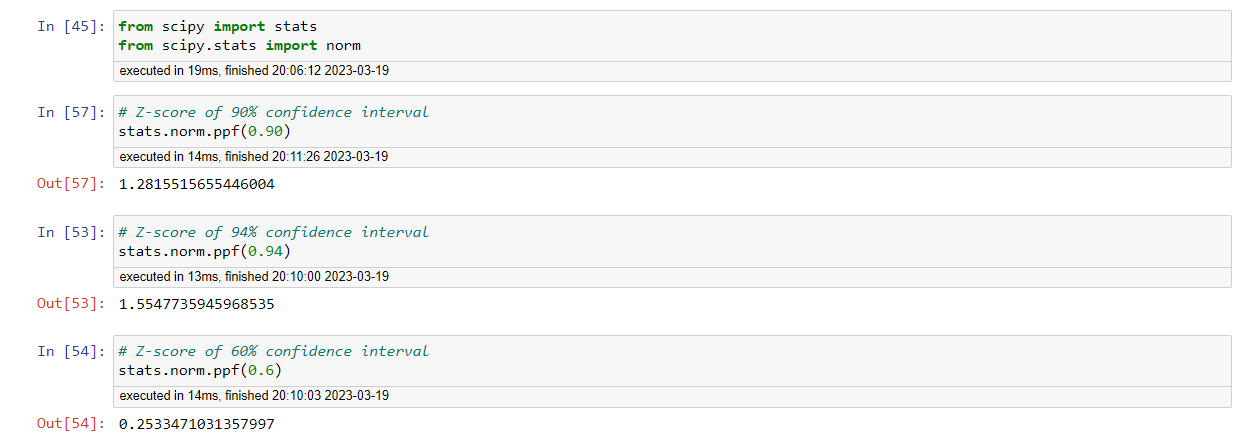
From below plot and numbers we can say that in Adipose Tissue, mean > median, right whisker is longer than left whisker so the data is positively skewed.

From below plot and numbers we can say that in Waist Circumference mean > median, both the whisker having same length, median is slightly shifted towards the left so the data is fairly symmetrical.



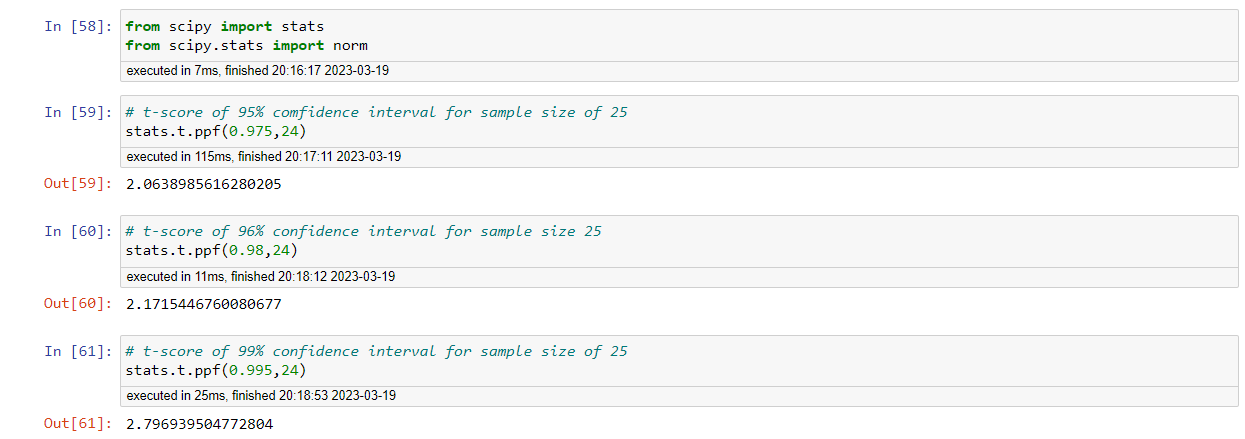
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

**ANS =**



Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

**ANS =**



Q 24**)** A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

**ANS =**

