|  |  |
| --- | --- |
| 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 | Continuous |
| Number of times married | Discrete |
| Gender (Male or Female) | Qualitative |

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

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

Total no of possibilities = 23 = 8

Total no sample {(HHH), (HHT), (HTH), (THH),

(HTT), (TTH), (THT), (TTT)}

The number of combination which have 2 head and 1 tail are: HHT, HTH, TTH = 3/8

OUTTCOME: The probability of getting 2 head & 1 tails in the toss of three coins simultaneously is 3/8 or 3.075

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

The sample space = 62 = 36

1. The sum is equal to 1 is zero because they starts with (1,1)

other than in the dice we are not having zero

1. The total possibilities of sum being less than or equal to 4 = 6

The probability is = 6/36 = **1/6**

The probability that the sum is less than or equal to 4 is **0.166 (or) 16.6%**

1. The total possible outcomes divisible by 2 and 3 = 24

The probability is possible outcome/total outcome = **24/36=2/3**

The probability that the Sum is divisible by 2 and 3 = **0.6666 (or) 66.7%**

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?

**Answer: -**

The total no of balls= (2+3+2) = 7

The probability of the first ball is not blue = 5/7(“blue ball \ total no of balls”)

The probability of the second ball is not blue = 4/6(“blue ball \ total no of balls”)

probability that none of the balls drawn is blue

**P = (5/7) \* (4/6)**

**= 20/42**

**= 10/21**

**= 0.476**

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

**Answer:-**

Expected value = ∑probability\*candies

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

**= 3.09**

The answer is 3.09

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

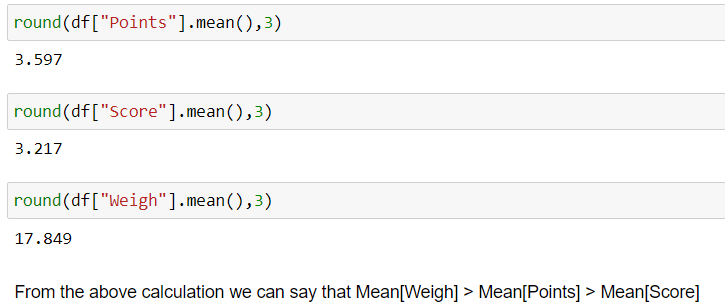
* For Points,Score,Weigh>

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

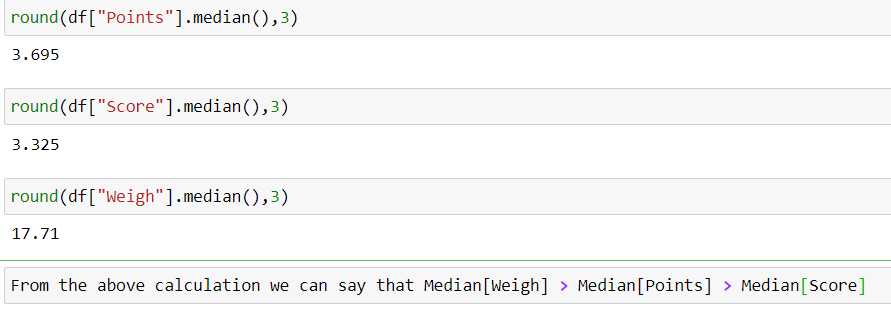
**Use Q7.csv file**

**Answer:-**

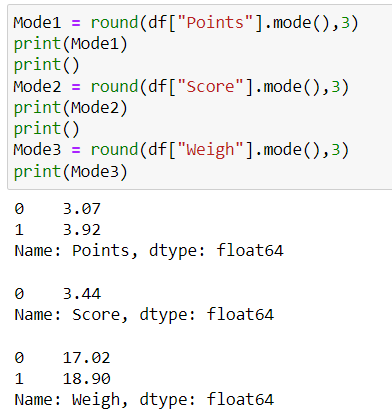
**Mean Calculation:**

****

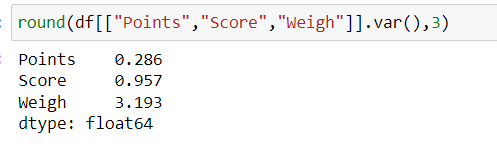
**Median Calculation:**

****

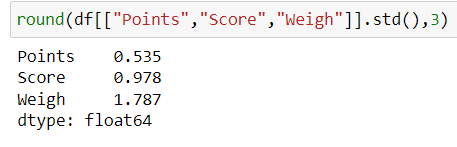
**Mode Calculation:**

****

**Variance Calculation:**

****

**Standard Deviation:**

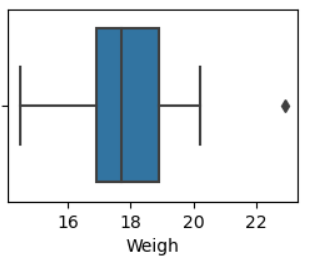
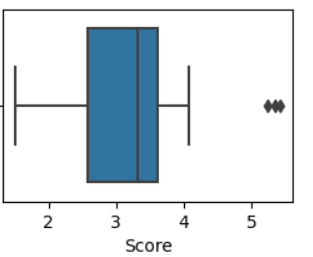
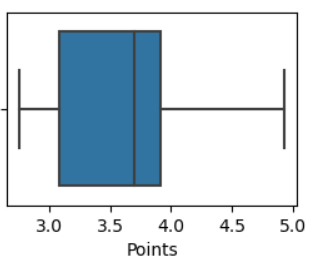
****

**Range:**

****

**Inference:**

1. **For Weigh Mean > Median that means data is slightly right skewed.**
2. **Except Points Score and Weigh have some outliers**

****

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:-

Expected value = ∑probability\*weights

= (1/9) \* [108+110+123+134+135+145+167+187+199)

=1308/9

**= 145.33**

Expected value of the weight of that patient = 145.33

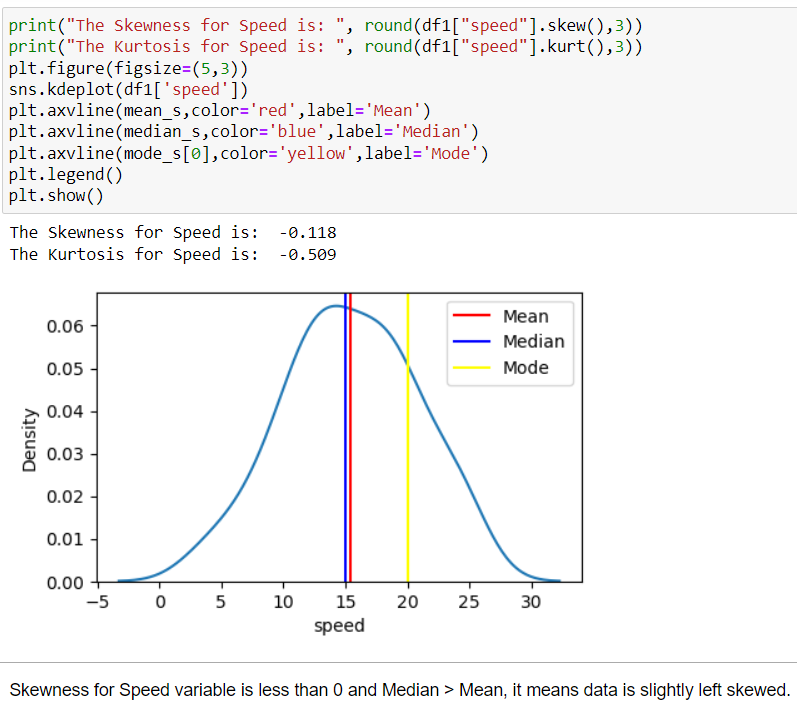
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

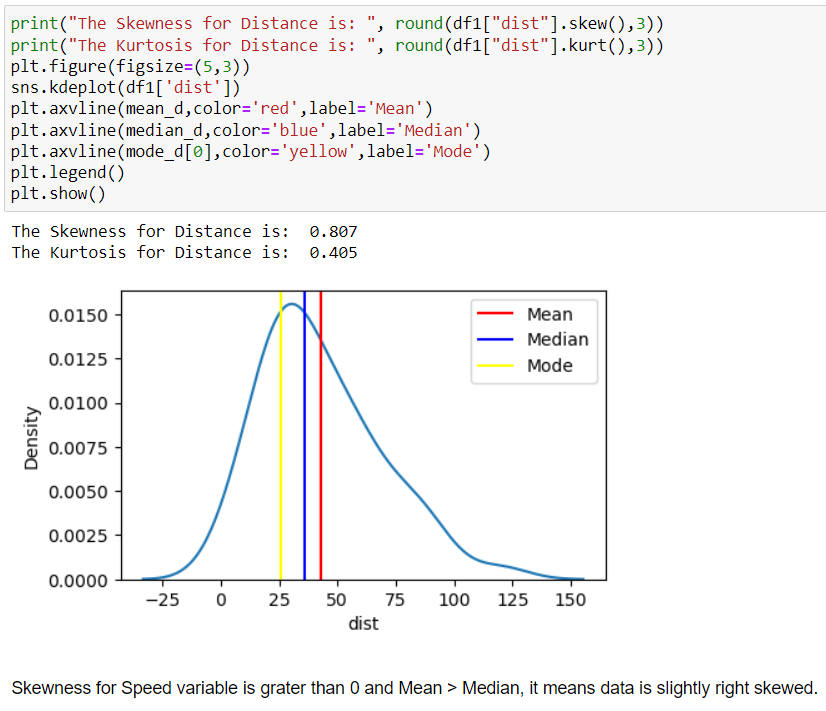
**Answer:**

Skewness and Kurtosis for Car Speed:

****

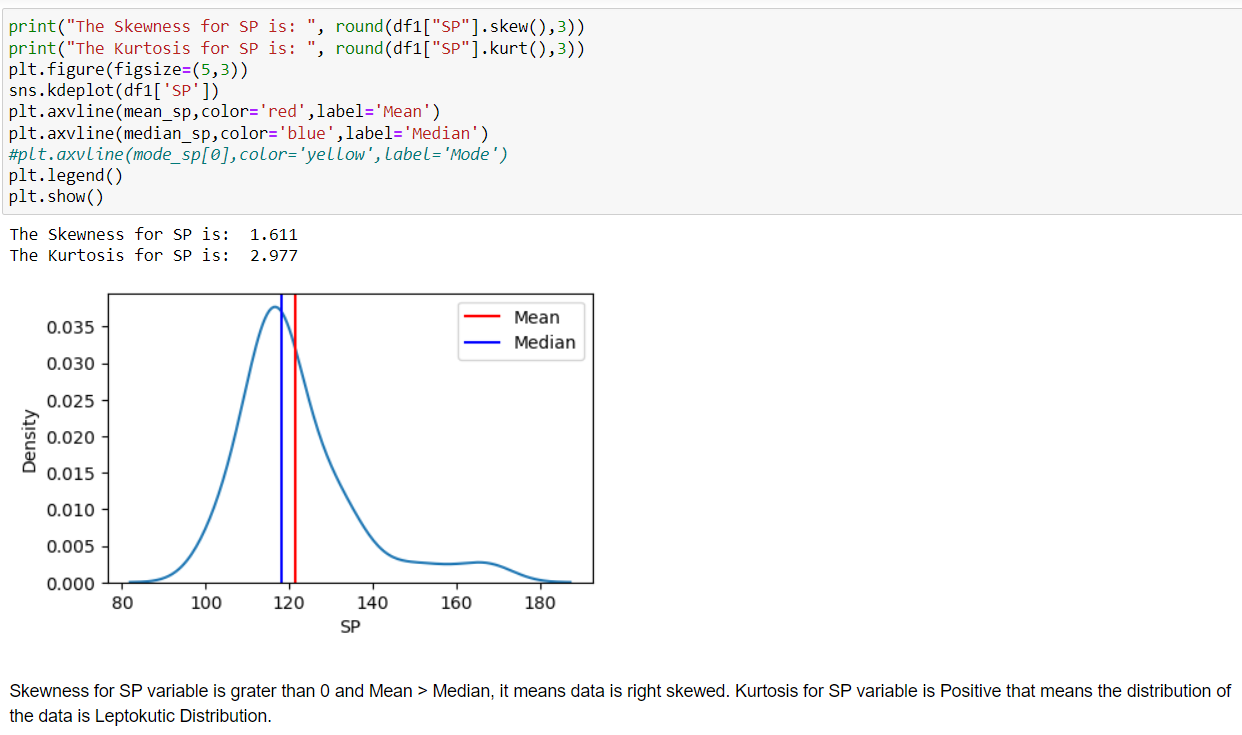
Platykurtic Distribution

Skewness and Kurtosis for Car Distance:

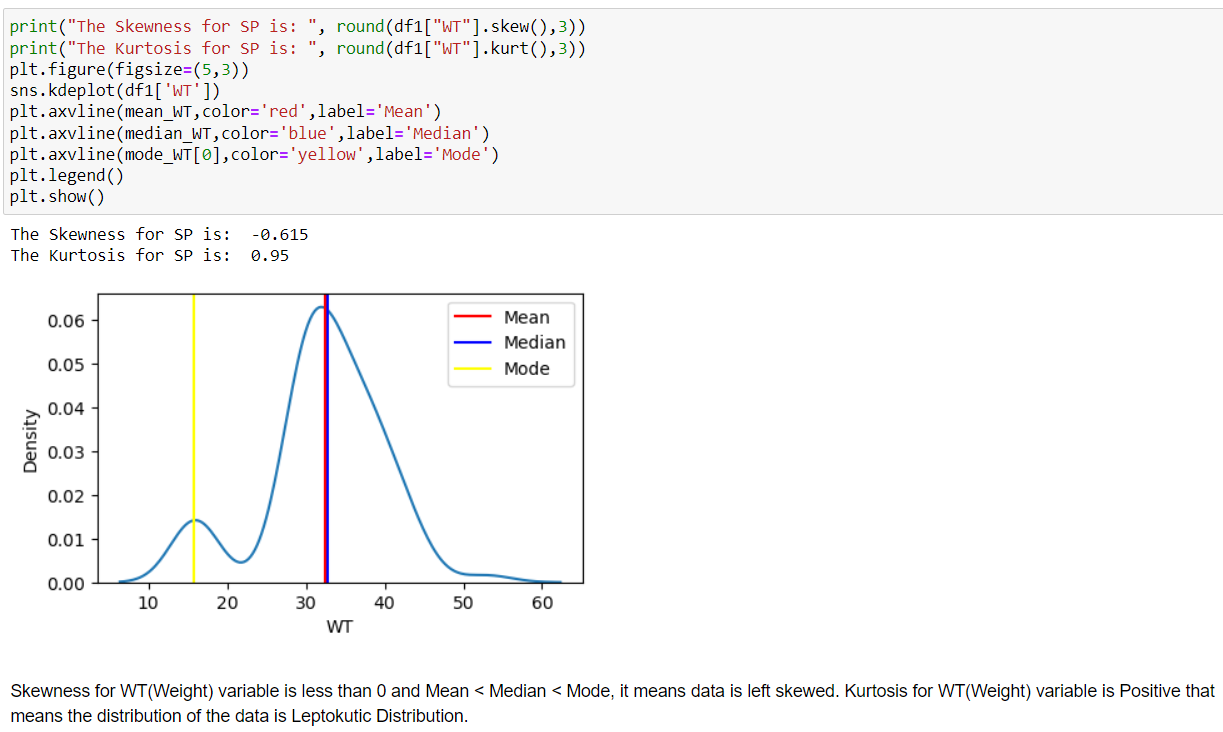
****

**Use Q9\_b.csv**

Skewness and Kurtosis for Car SP:



Skewness and Kurtosis for Car Weight:



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

Ans :- From the below histogram we can say that data is rightly skewed, comparatively frequency of 400 is very less to other bins

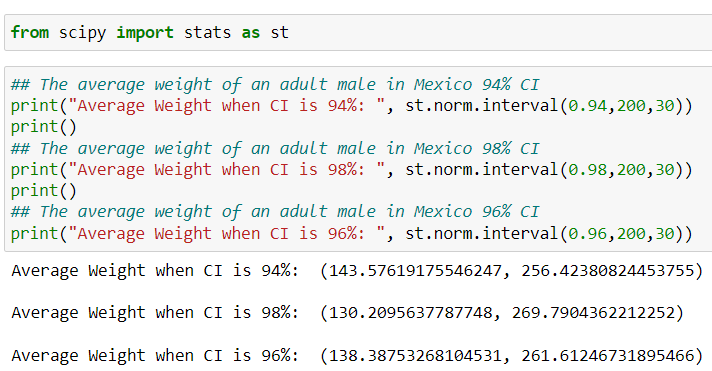


From the below boxplot we can say that outliers are present in above the upper whiskers and we can mean > median because median line lies towards the 25% of the data



**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?

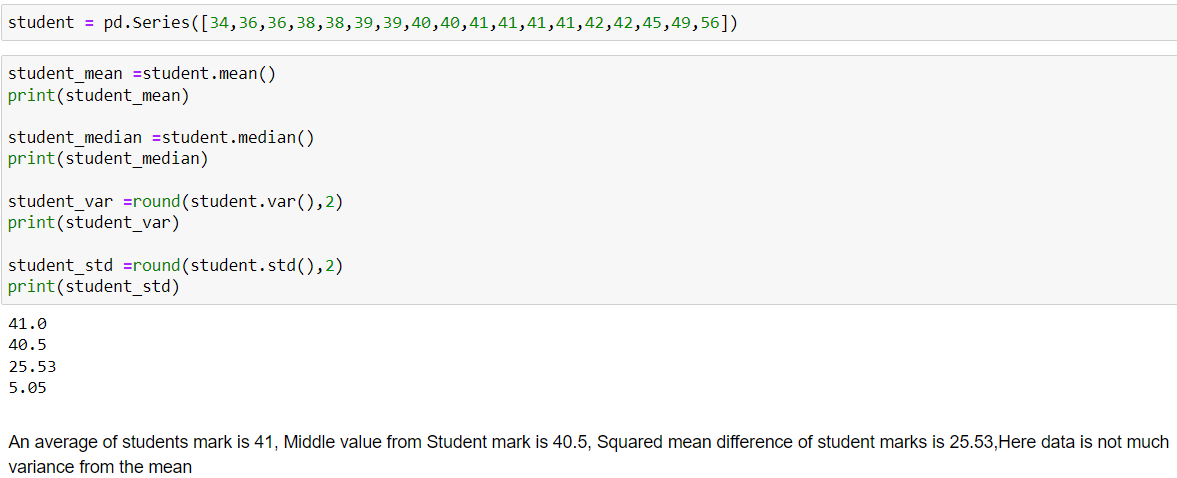
**Answer:**



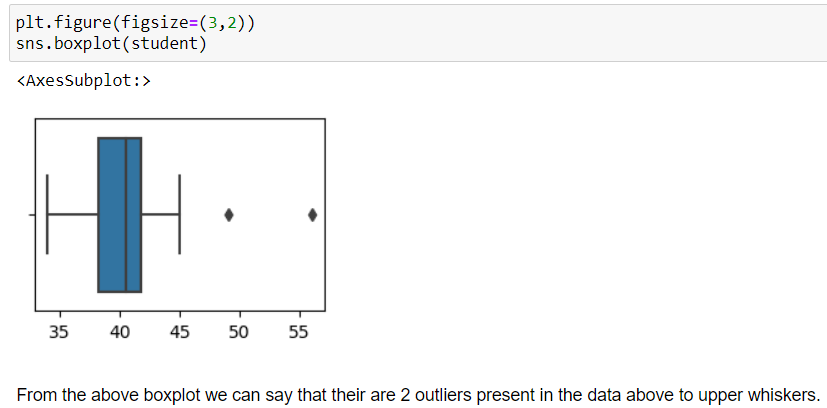
**Q12)** Below are the scores obtained by a student in tests

**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.



1. What can we say about the student marks?



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

Ans:-If mean and median of data are equal in skewness that means data is normally distributed

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

Ans :- If mean > median the nature of skewness is Right Skewed

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

Ans :-If median > mean the nature of skewness is Left Skewed

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

Ans :- Positive kurtosis value indicates for a data as Leptokurtic that means peakedness is high compare to others two skewness type

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

Ans :- Negative kurtosis value indicates for a data as Platykurtic that means peakedness is low compare to others two skewness types

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



What can we say about the distribution of the data?

Ans:-

* The distribution of this data is not Normally distributed
* Most of the values in left so it is left skewed
* The median value of the distribution is 15
* The Lower and Upper quartile are 10 and 18

What is nature of skewness of the data?

Ans: - Most of the values are skewed towards left side so the nature of the skewness is left skewed means median > mean

What will be the IQR of the data (approximately)?   
Ans:-

IQR = inter quartile range

Formula= upper quartile-lower quartile

upper quartile=18

lower quartile= 10

IQR = 18-10

IQR =8

Q19) Comment on the below Boxplot visualizations?



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

Ans: -

* The median of both the boxplots is the same 🡪(262.5)
* Both Plots shown above has equal distribution of data above and below the median and apparently, it’s a normal distribution.
* Kurtosis is negative in 1st plot as it has thinner tails and positive in the 2ndplot.

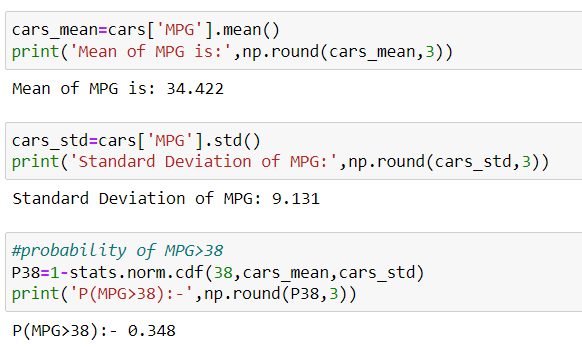
Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

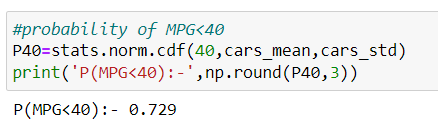
Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

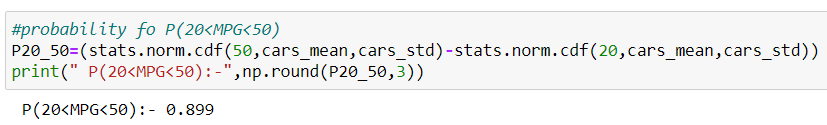
* 1. P(MPG>38)



* 1. P(MPG<40)



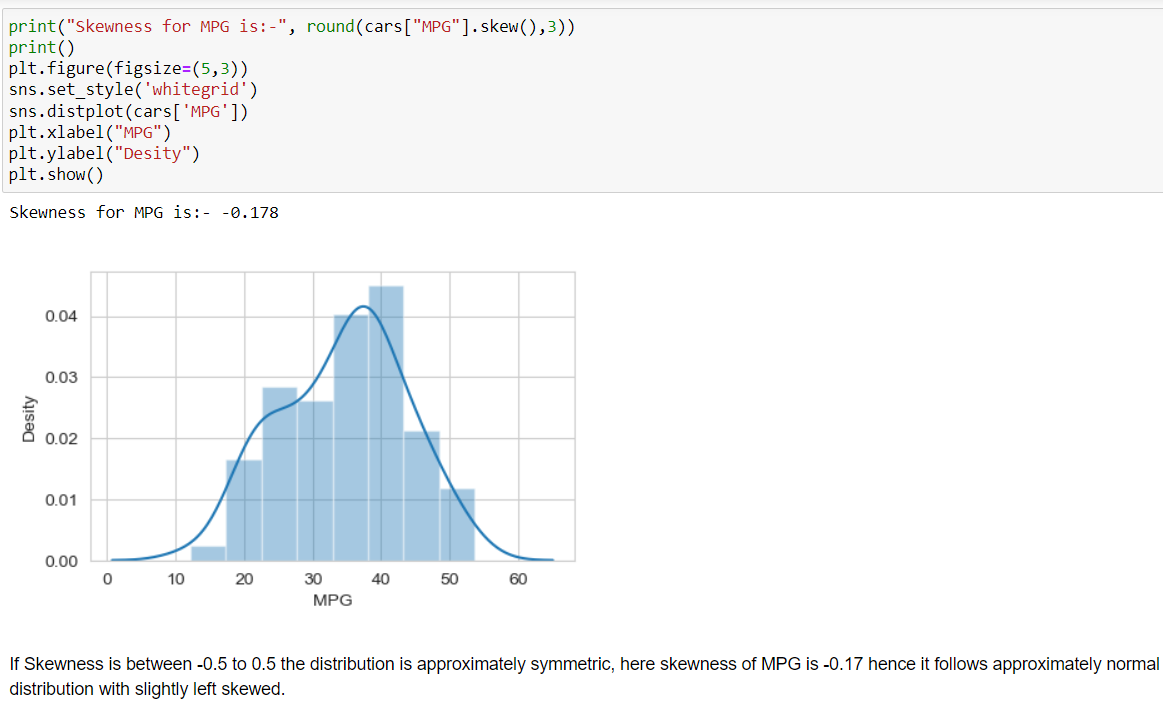
c. P (20<MPG<50)



Q 21) Check whether the data follows normal distribution

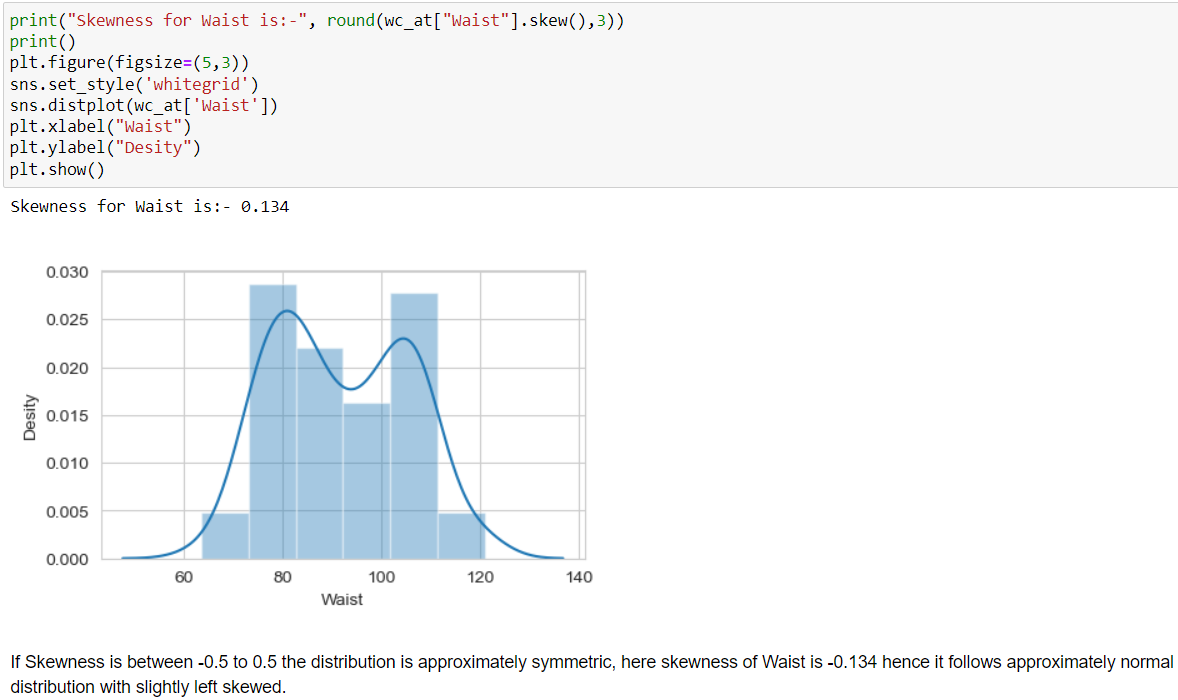
1. Check whether the MPG of Cars follows Normal Distribution

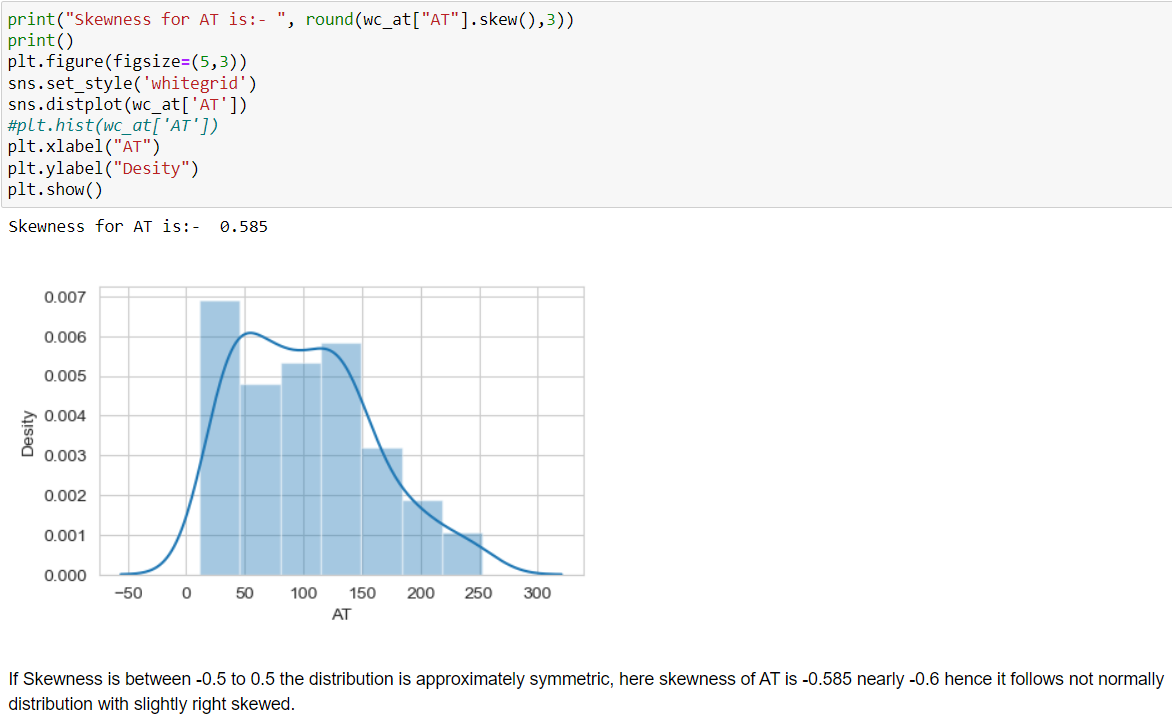
Dataset: Cars.csv



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

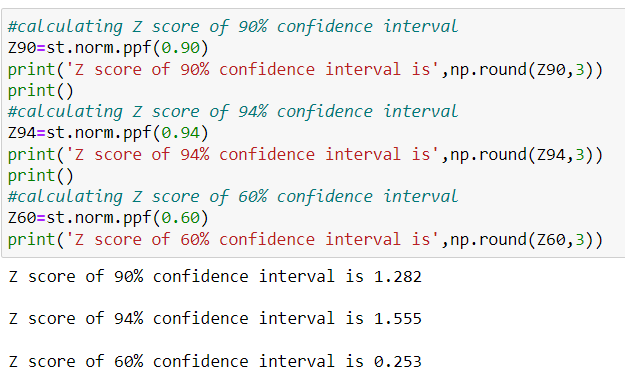
Dataset: wc-at.csv





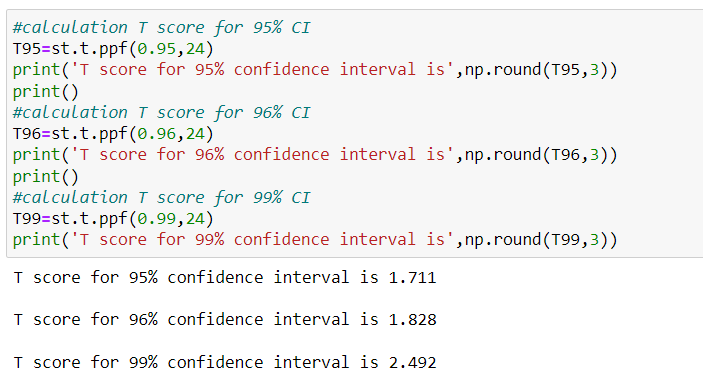
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

**Answer:**



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

**Answer:**



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

**Answer:**

