Q1) Identify the Data type for the Following:

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

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

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

Ans = 3/8

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

1. Equal to 1

Ans = 0

1. Less than or equal to 4

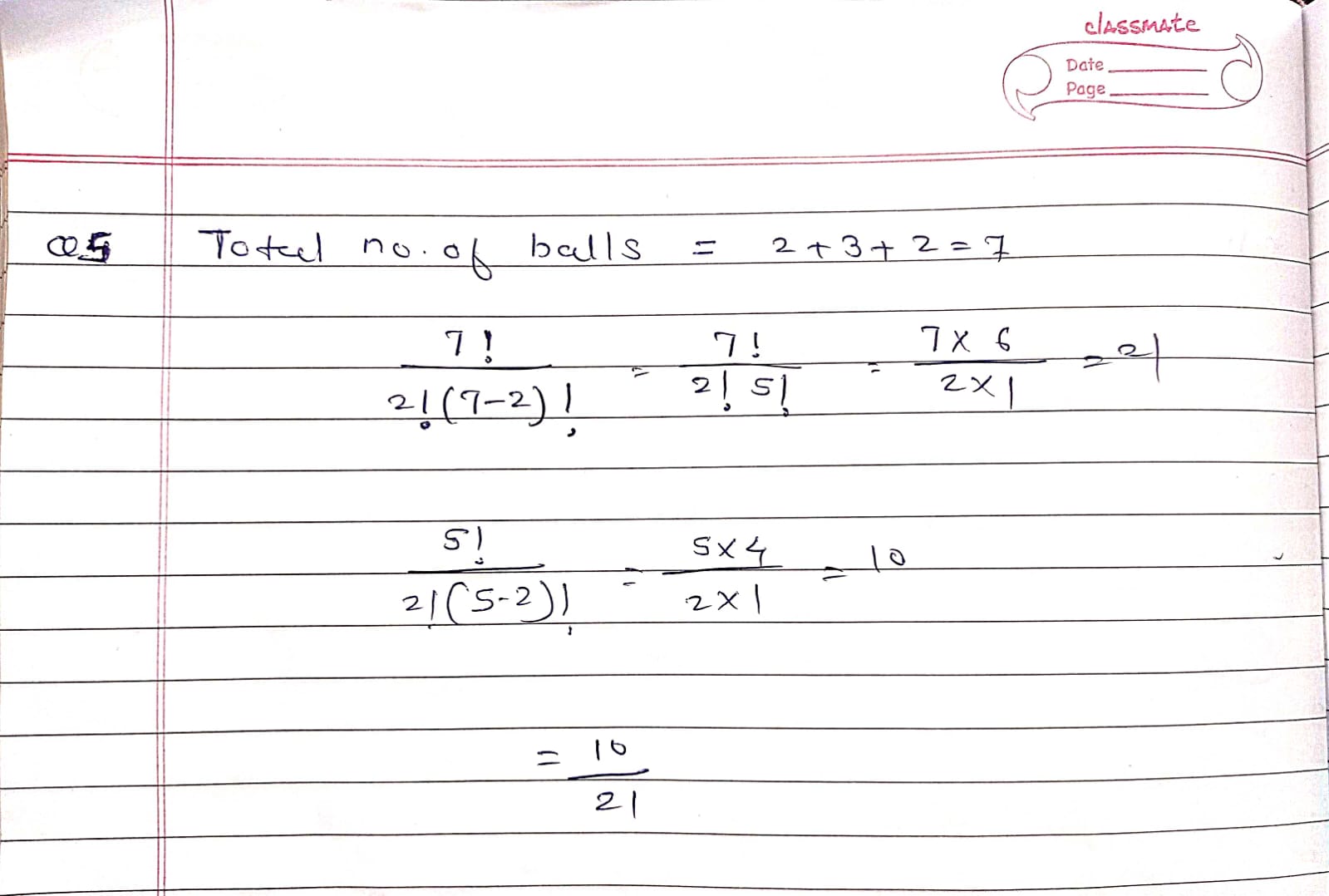
Ans = 1/6

1. Sum is divisible by 2 and 3

Ans = 1/6

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 = 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

So, 1\*0.015 + 4\*0.20 + 3\*0.65 + 5\*0.005 + 6\*0.01 + 2\*0.120 = 3.09

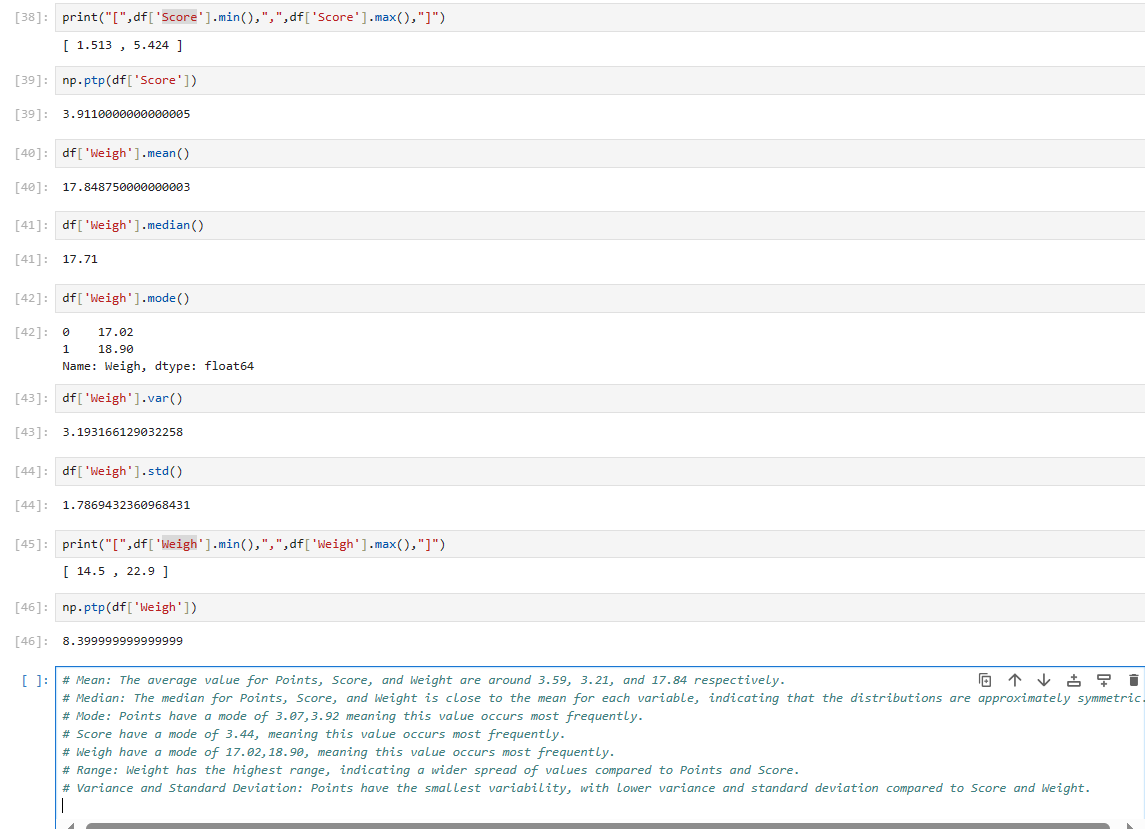
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points, Score, Weight

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

**Use Q7.csv file**

****

****

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?

108+110+123+134+135+145+167+187+199 / 9 = 145.33

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

**Cars speed and distance**

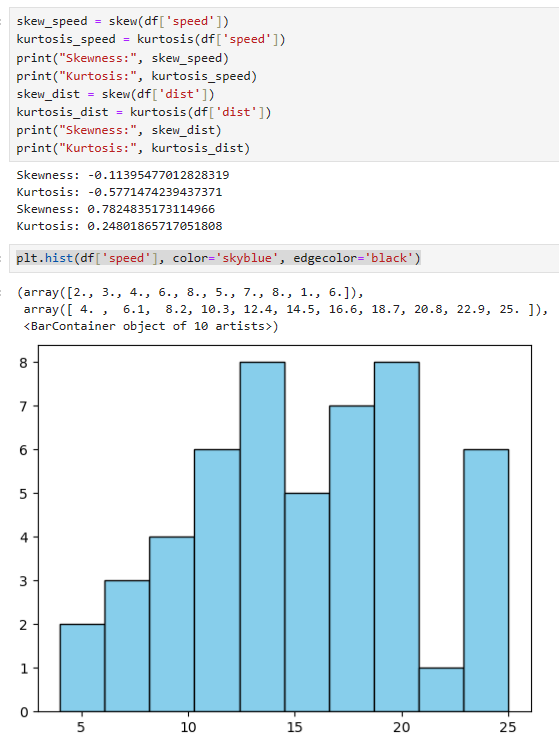
**Use Q9\_a.csv**

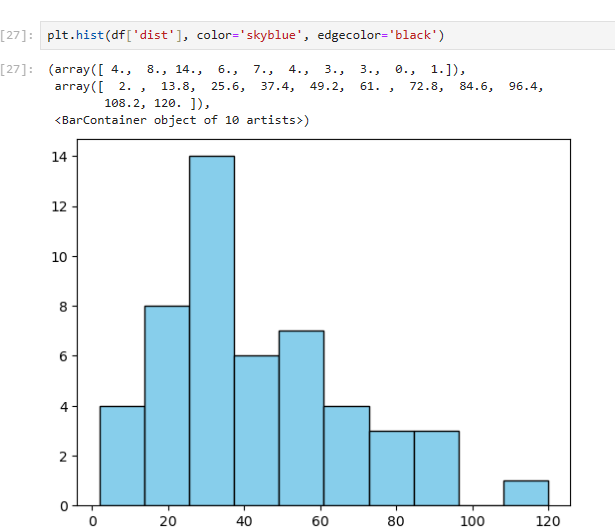
Skewness of speed: -0.11395477012828319

Kurtosis of speed: -0.5771474239437371

Skewness dist: 0.7824835173114966

Kurtosis dist: 0.24801865717051808

****

****

**SP and Weight(WT)**

**Use Q9\_b.csv**

Skewness WT: -0.6033099322115126

Kurtosis WT: 0.8194658792266849

Skewness SP: 1.5814536794423764

Kurtosis SP: 2.7235214865269244

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



Histogram shows that it is positively skwed data, that mean we can conclude dataset have some outliers

Boxplot shows that it has outlier at upper site

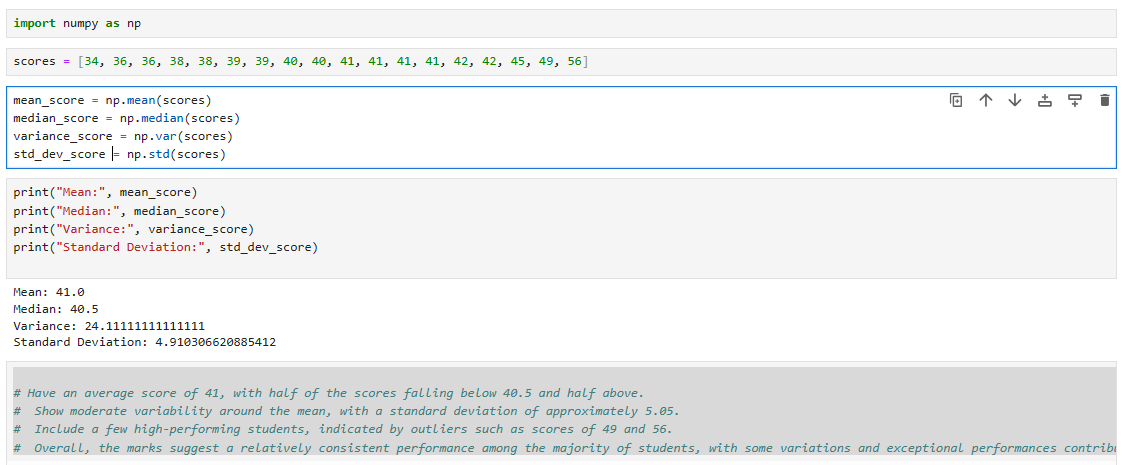
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate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



**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.
2. What can we say about the student marks?



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

When the mean and median of a dataset are equal, it implies that the distribution is symmetric around the center. There is no skewness present in the distribution.

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

When the mean of a dataset is greater than the median, it suggests that there is a right-skewness or positive skewness in the distribution.

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

When the median of a dataset is greater than the mean, it suggests that there is a left-skewness or negative skewness in the distribution.

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

positive kurtosis indicates that there are more extreme values in the dataset than would be expected in a normal distribution.

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

negative kurtosis indicates that there are fewer extreme values in the dataset than would be expected in a normal distribution.

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



What can we say about the distribution of the data?

Median is closer to the upper quartile so it suggests positive skewness.

What is nature of skewness of the data?

Positively Skewed

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

IQR= UQ – LQ

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.

Boxplot 1 and Boxplot 2 having same Q2 that is nothing but median =25

Boxplot 1 and Boxplot 2 Q2 is present at exact center point that show there is no skewness present

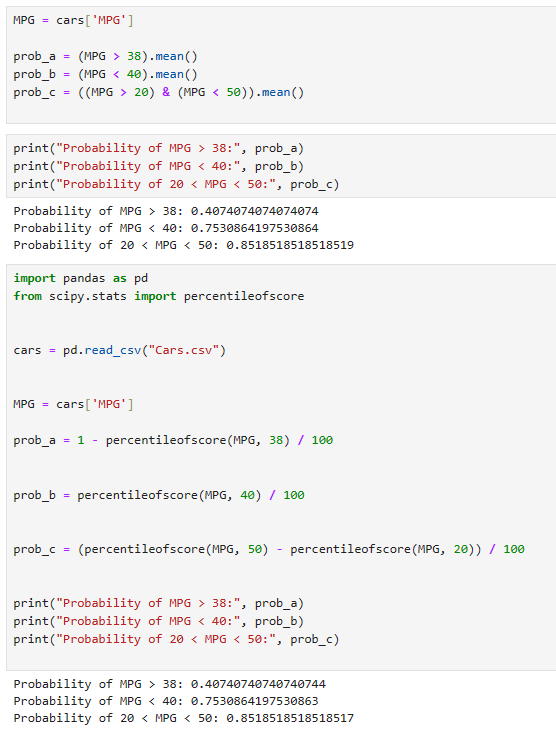
Q 20) Calculate probability from the given dataset for the below cases

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)
  3. P (20<MPG<50)



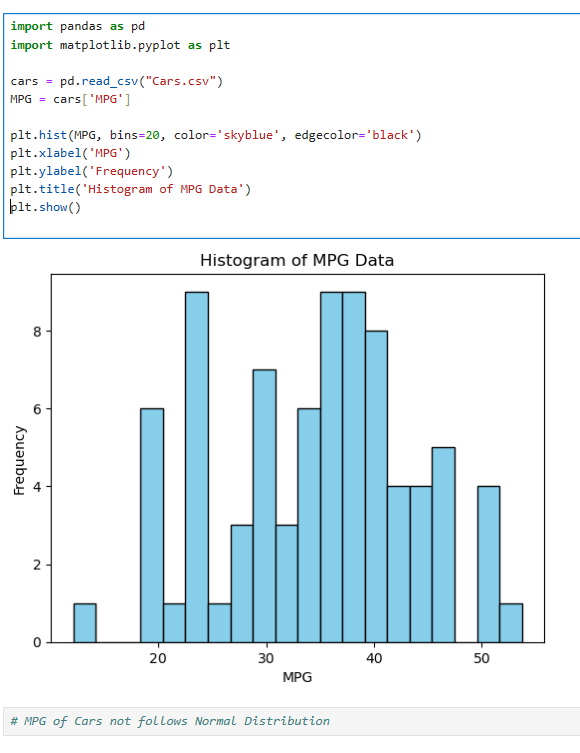
Q 21) Check whether the data follows normal distribution

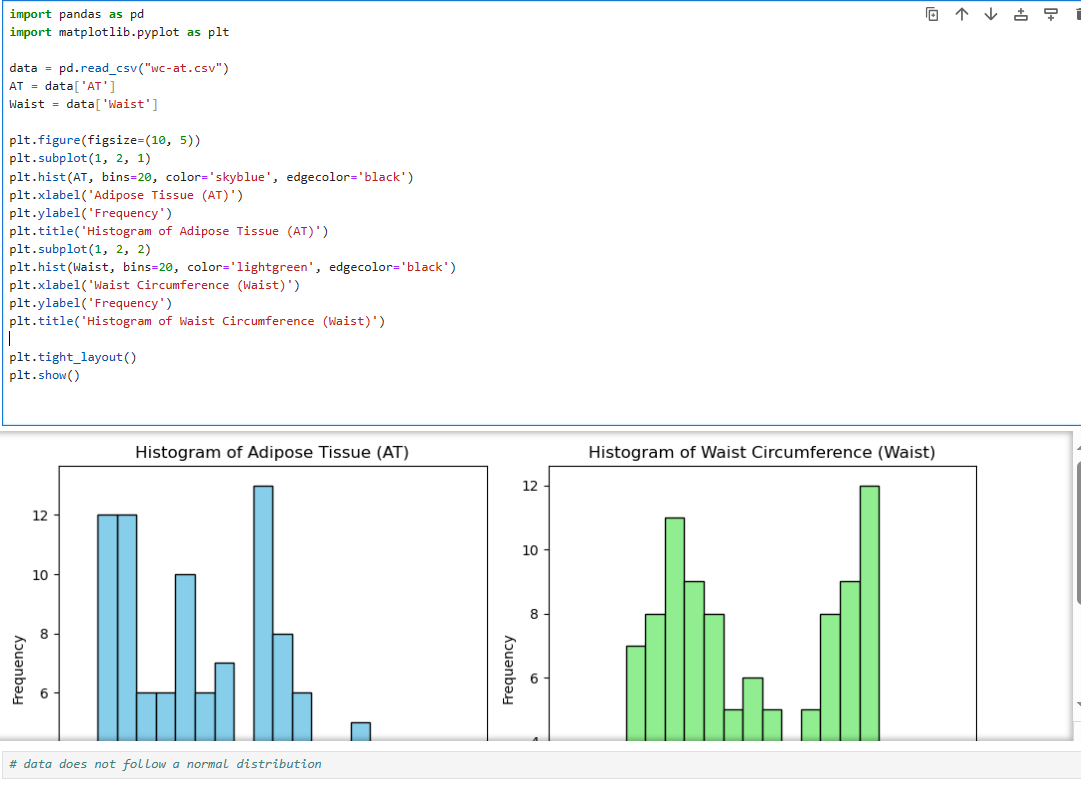
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

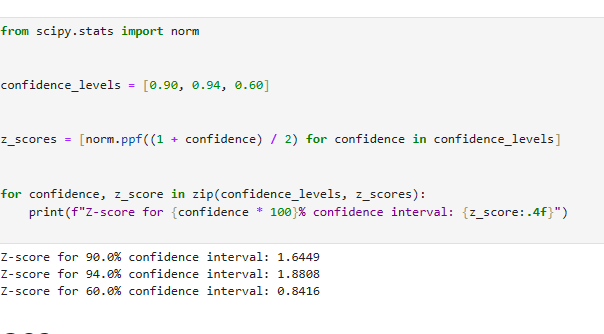
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

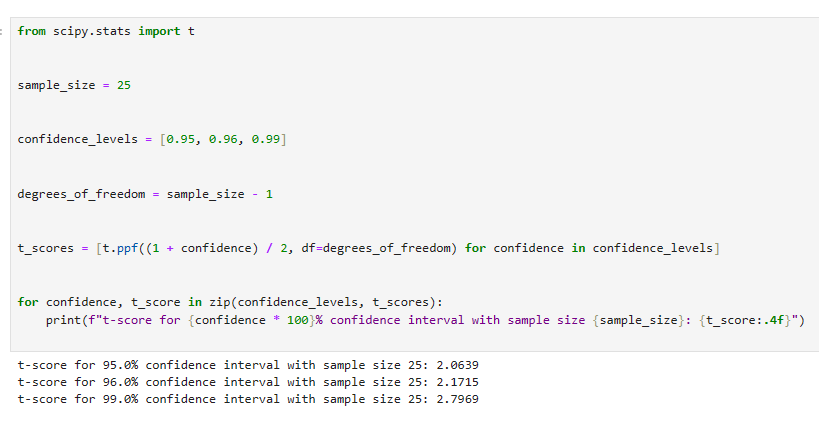




Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval



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



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

