|  |  |
| --- | --- |
| 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 | Nominal |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Nominal |

**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) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Ordinal |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Ratio |

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

**Ans:** when three coins are tossed sample space is given as

S={HHH,HHT,HTH,THH,HTT,TTH,THT,TTT}

n(S)=8

probability of two heads and one tail

A={HHT,HTH,THH}

p(A)=3/8

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

**Ans:** when we roll two dice possible outcomes are 6\*6=36 possibilities

1. **Equal to 1:**when two dice are rolled,the minimum sum possible is 2.Therefore,its impossible to get a sum of 1. So the probability is 0.
2. **Less than or equal to 4:to** find the sum of less than or equal to 4,the possible outcomes are as follows;

Sum of 2:(1,1)

Sum of 3:(1,2),(2,1)

Sum of 4:(1,3),(2,2),(3,1)

There are total of 6 possible outcomes

6/36=1/6

1. **Sum is divisible by 2 and 3:**The sum tht are divisible by both 2 and 3 are those that are divisible by their least common multiple,which is 6.

Sum of 6:(1,5),(2,4),(3,3),(4,2),(5,1)

Sum of 12:(6,6)

Soo there are total 6 outcomes

Probability =6/36=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:** total number of balla=(2+3+2)=7

Lets S be the sample space

n(S)=number of way of drawn 2 balls out of 7=7C2= 7!/2!(7-2)!=7\*6/2=21

number of way choose 2 out of 5=5C2=5!/2!(5-2)!=5\*4/2=10

to find probability of none ball is blue

probability=number of ways to drawn non-blue ball/total number of ways to drawn 2 balls

=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**

E(x)=1\*p(1)+2\*p(2)+3\*p(3)+4\*p(4)+5\*p(5)+6\*p(6)

E(x)=0.015+0.24+1.95+0.8+.025+0.06

E(x)=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**

****

Here mean and median values are quite similar to each other hence the distribution of data is symmetric.Range of Weight column is larger than all other two columns.

**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: X={108,110,123,134,135,145,167,187,199}

Xi=1308

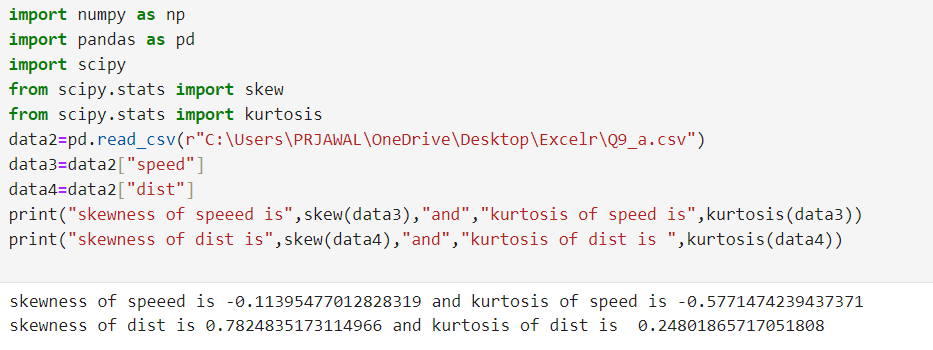
Xi/n

E(x)=145.33

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

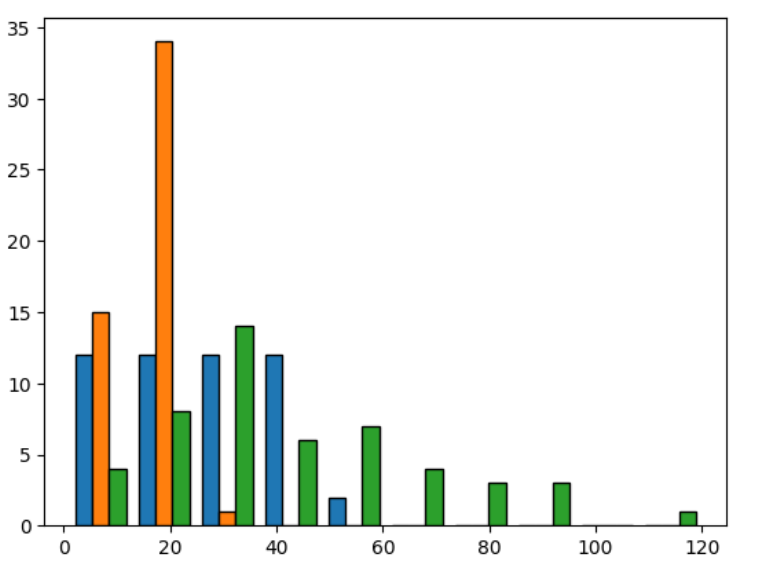
**Cars speed and distance**

**Use Q9\_a.csv**

****

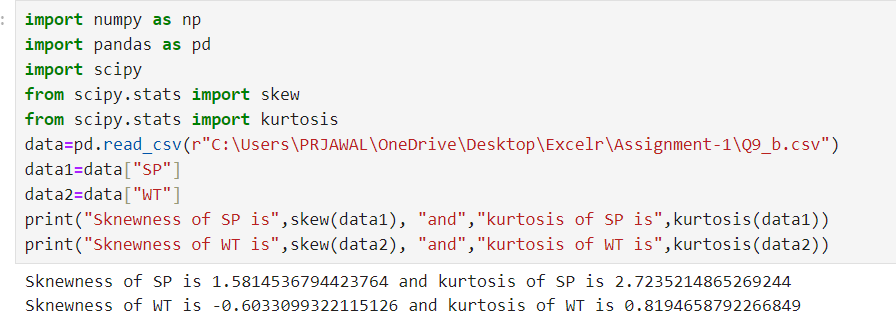
**Skewness**:Both "speed" and "distance" have positive skewness, indicating that their distributions are skewed towards the right (i.e., they have a tail on the right side).The skewness of "speed" (0.919) is slightly higher than that of "distance" (0.782), indicating a relatively more pronounced skewness in the "speed" data.

**Kurtosis:**Both "speed" and "distance" have positive kurtosis, indicating that their distributions have heavier tails and are more peaked than a normal distribution.The kurtosis values for both "speed" and "distance" are close to 0, indicating that their distributions are approximately mesokurtic

****

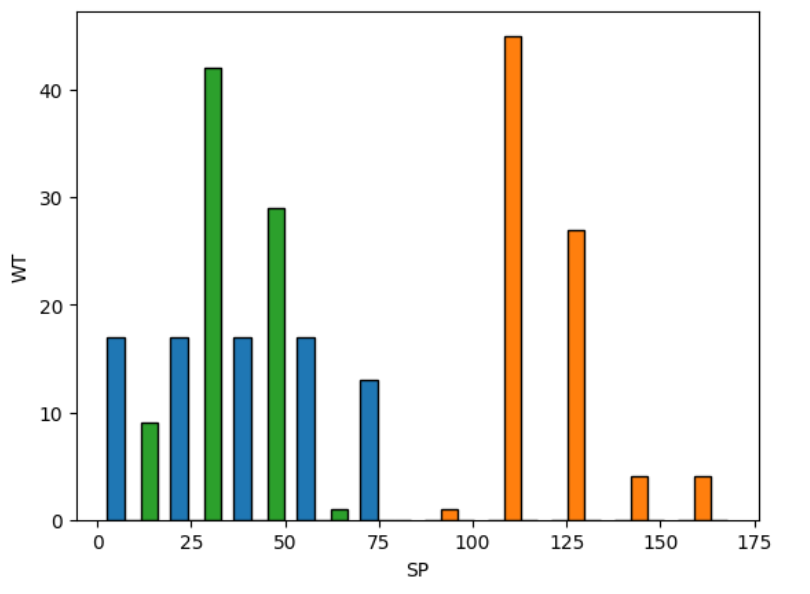
**SP and Weight(WT)**

**Use Q9\_b.csv**

****

**Skewness:**The skewness of the “SP”(speed) data is 1.58,indicating a moderate positive skewness,meaning the data is skewed towards higher values.The skewness of the “WT”(Weight) data is -0.60,indicating a very negative skewness,suggesting a nearly symmetric distribution.

**Kurtosis:**The kurtosis of the “SP”(Speed)data is 2.72,indicating a leptokurtic distribution,meaning it has heavy tails and a sharp peak.The kurtosis of the “WT”(weight)data is 0.81,indicating a platykurtic distribution,meaning it has lighter tails and a flatter peak compared to a normal distribution.



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



The above histogram shows that it is not a normal distribution as it is right skewed.There might be chance of present of outlier in chickWeight as we can see there is very a smaller number of datapoint in right bin of 350 to 400.central tendency of data is between 50-1000.most datapoints fall under range of 50-100 where between 300-400 very a smaller number of datapoints are present.



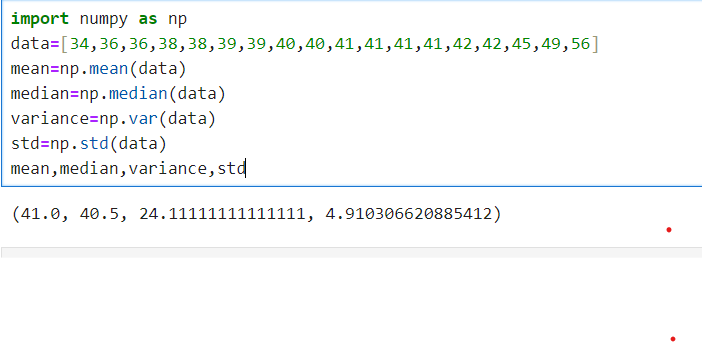
The above boxplot shows that the data is more concentrated in the lower half.The IQR is smaller so we can say that there is less variability in datapoints.Given distribution of data is asymmetric as it has longer whisker than other.There are more than one outlier are present.

**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.**



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

Ans: The average students mean is about 41 and median is 40.5, The variance is approximately 24.11 and the standard deviation is approximately 4.91. This indicates that the scores have some variability around the mean.

Based on these calculations, we can say that the student's marks are clustered around the mean with some variability, as indicated by the standard deviation. The majority of scores fall within a certain range around the mean, with a few outliers on the higher end.

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

**Ans**: When the mean and median of data are equal, the skewness is zero, indicating symmetry in the distribution.

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

**Ans**: When the mean is greater than the median, it suggests that the right tail of the distribution is longer than the left tail. This indicates positive skewness, meaning the distribution is skewed to the right.

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

**Ans:** When the median is greater than the mean, it suggests that the left tail of the distribution is longer than the right tail. This indicates negative skewness, meaning the distribution is skewed to the left**.**

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

**Ans:** A positive kurtosis value indicates that the data has heavier tails and a sharper peak than a normal distribution. It means the distribution has more outliers or extreme values than would be expected under a normal distribution. This type of distribution is called leptokurtic.

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

**Ans:** A negative kurtosis value indicates that the data has lighter tails and a flatter peak than a normal distribution. It means the distribution has fewer outliers or extreme values than would be expected under a normal distribution. This type of distribution is called platykurtic.

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



**What can we say about the distribution of the data?**

**Ans**: Given data is left skewed and had median value of approx.most values lie netween 10 to 18 data points.More data is concentrated oneside.

**What is nature of skewness of the data?**

**Ans:**Given data is left skewed

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

**Ans:**IQR of given boxplot is 10 to 18

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 has less variance compare to boxplot 2 since we don’t have any idea about how many datapoints are present in both boxplot.median point of both boxplot is equal.data is normally distributed and there is no skewness present.

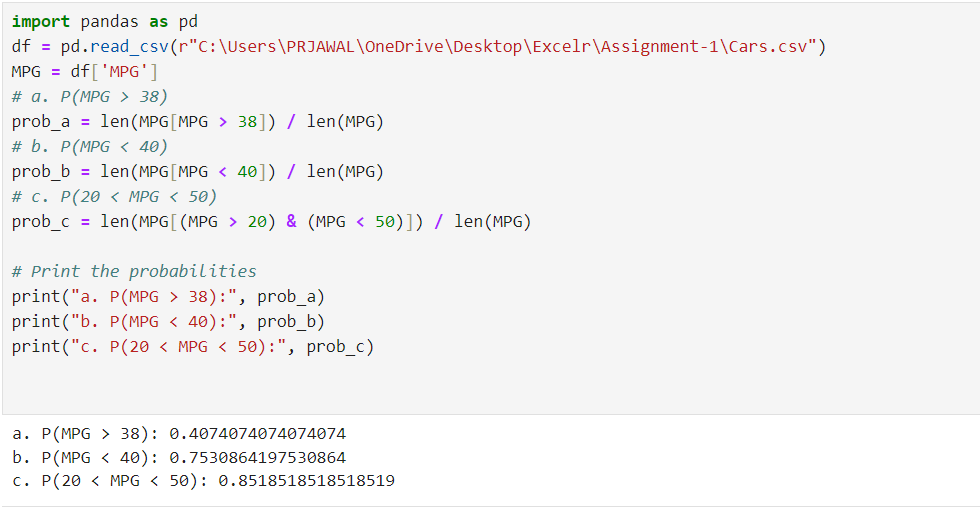
**Q 20) Calculate probability from the given dataset for the below cases**

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)

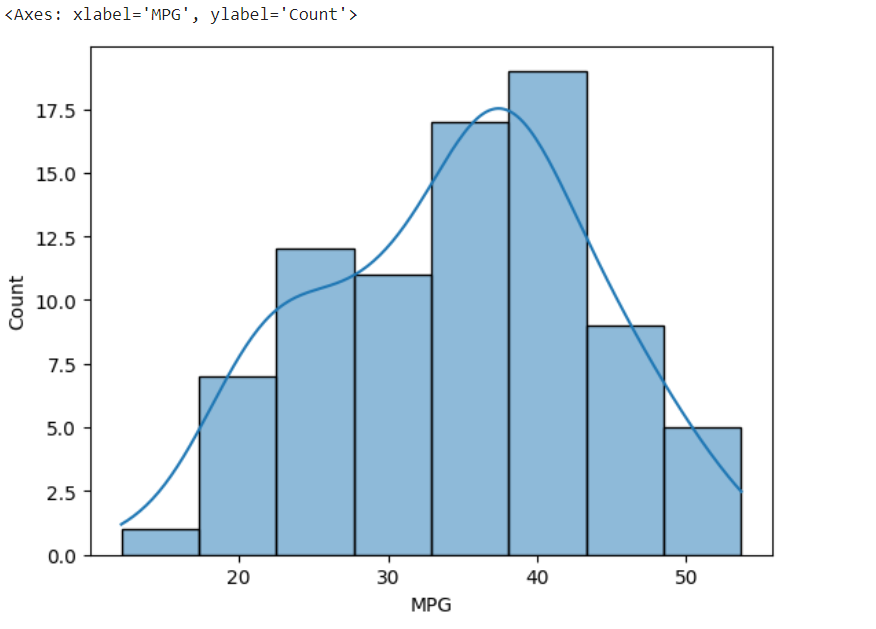


**Q 21) Check whether the data follows normal distribution**

1. **Check whether the MPG of Cars follows Normal Distribution**

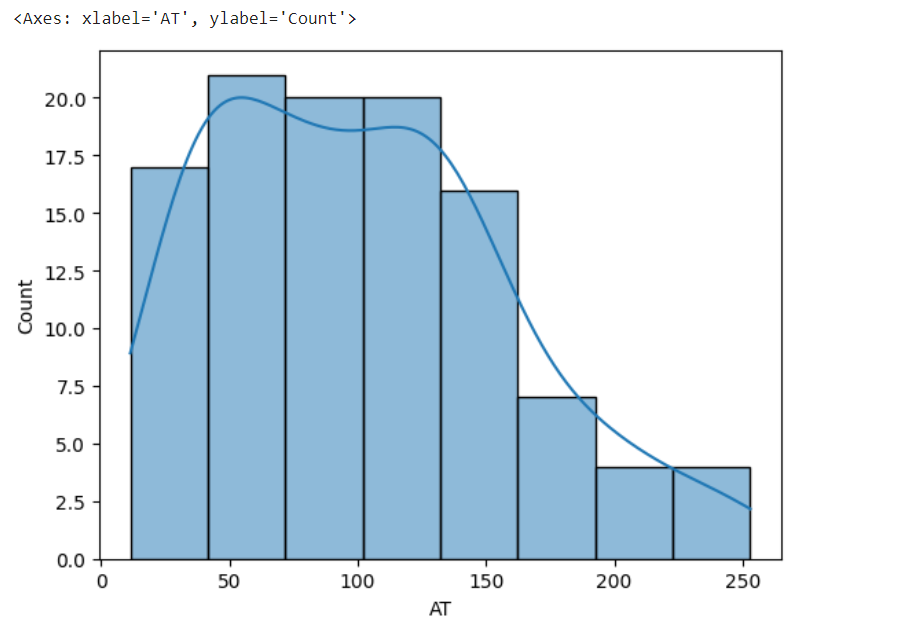
**Dataset: Cars.csv**

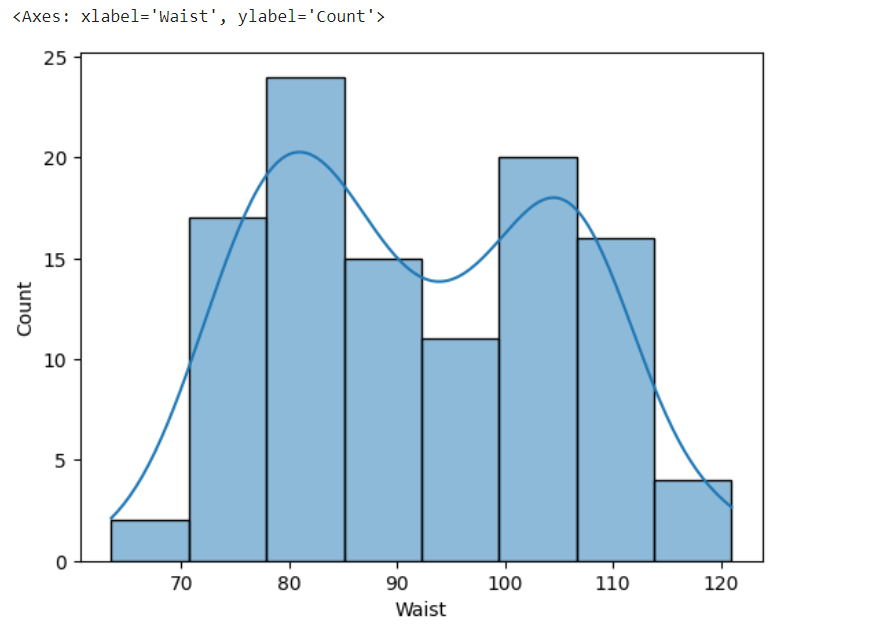
**Ans:MPG of car is slightly left skewed.**

****

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

**Dataset: wc-at.csv**

AT is right skewed distribution**.**

****

Waist is not a normal distribution.

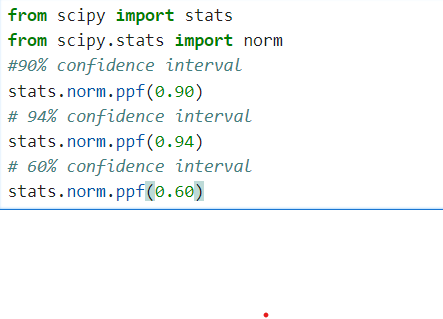
**Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval**

**Ans:**

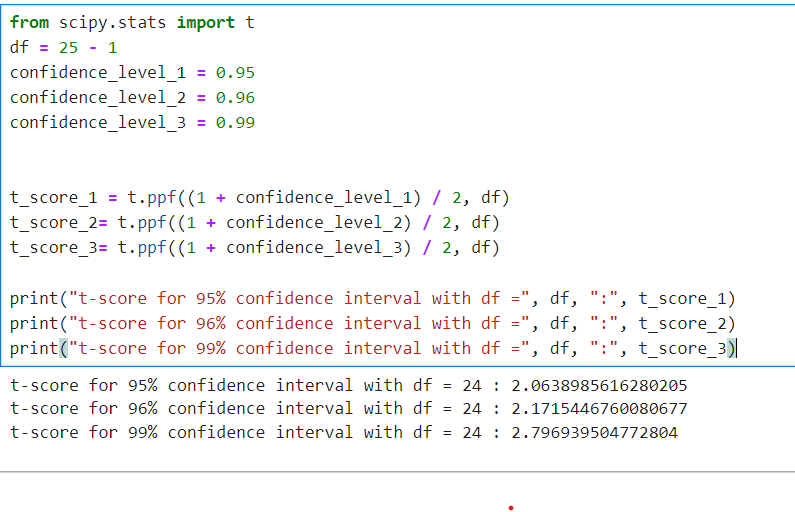
90%=1.281551565544

94%=1.55477359459685

60%=0.2533471031358



**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: Given;

Sample mean=260 , population mean=270 ,std=90 , n=18

T-Score =

=

=

=

=

= -0.471404521

Degree of freedom = n – 1 = 18 – 1 = 17

The probability t < -0.471 with 17 degree of freedom, the t-value is less than the t-value obtained.

The probability of the bulbs lasting less than 260 days on average of 0.3218 assuming the mean life of the bulb is 300 days