|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | **Numeric-discrete** |
| Results of rolling a dice | **Numeric-discrete** |
| Weight of a person | **Numeric-continuous** |
| Weight of Gold | **Numeric-continuous** |
| Distance between two places | **Numeric-continuous** |
| Length of a leaf | **Numeric-continuous** |
| Dog's weight | **Numeric-continuous** |
| Blue Color | **Categorical-nominal** |
| Number of kids | **Numeric-discrete** |
| Number of tickets in Indian railways | **Numeric-discrete** |
| Number of times married | **Numeric-discrete** |
| Gender (Male or Female) | **Categorical-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) | **Interval** |
| Sales Figures | **Ratio** |
| Blood Group | **Nominal** |
| Time Of Day | **Interval** |
| Time on a Clock with Hands | **Interval** |
| Number of Children | **Ratio** |
| Religious Preference | **Nominal** |
| Barometer Pressure | **Interval** |
| SAT Scores | **Interval** |
| Years of Education | **Ratio** |

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

**Answer:-**

**Total cases: 8**

**Favorable cases: 3( HHT, HTH, THH )**

**Probability = 3/8**

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

**Answer: - Total cases**

1. **0**
2. **F=8 i.e. [( 1,3),(2,2),(3,1),(1,2),(2,1),(1,1)]**

**Probability = 6/36=1/6**

1. **F= (1,5),(2,4),(3,3),(4,2),(5,1),(6,6)**

**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?

**Answer :- (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

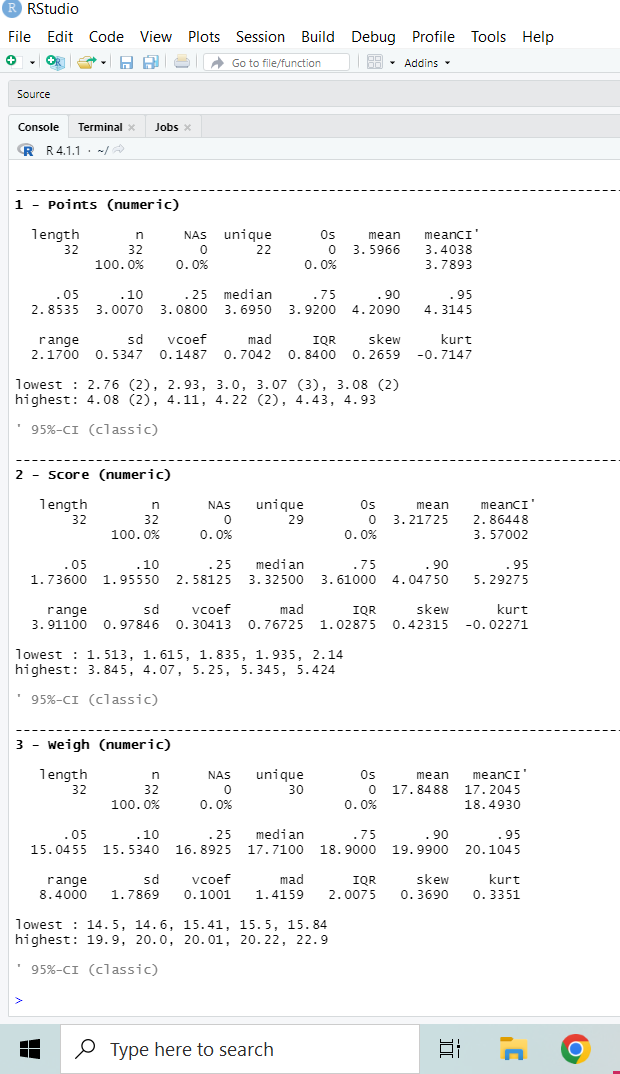
**Answer :- Expectation=(1\*0.015)+(4\*0.20)+(3\*0.65)+(5\*0.005)+(6\*0.01)+(2\*0.120)=3.09**

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

* For Points,Score,Weigh>

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

**Use Q7.csv file**

****

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Mean** | **Median** | **Mode** | **Variance** | **Standard deviation** | **Range** |
| **Points** | **3.60** | **3.695** | **3.07,3.92** | **0.29** | **0.53** | **2.170** |
| **Score** | **3.22** | **3.325** | **3.440** | **0.96** | **0.98** | **3.911** |
| **Weigh** | **17.85** | **17.710** | **18.90,17.02** | **3.19** | **1.79** | **8.400** |

**Data has fairly small standard deviation.**

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?

**Answer:-**

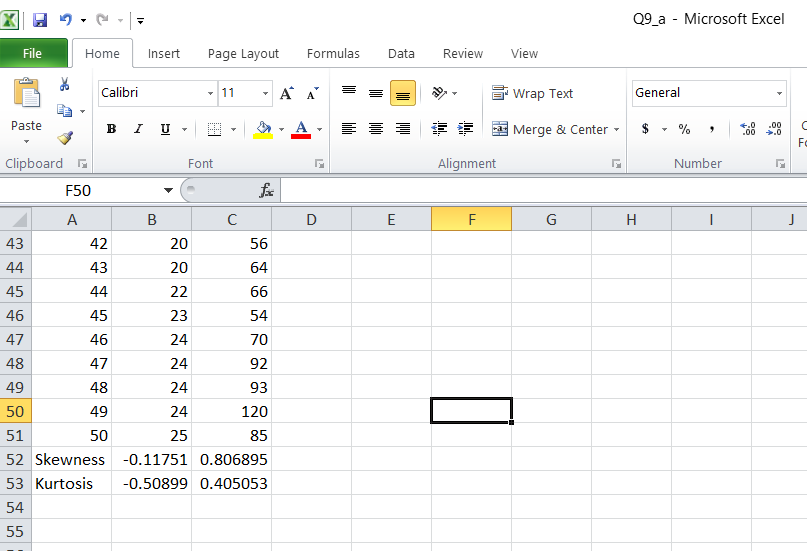
**Expectation =(108+110+123+134+135+145+167+187+199)/9**

**= 145.33 pounds**

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

**Cars speed and distance**

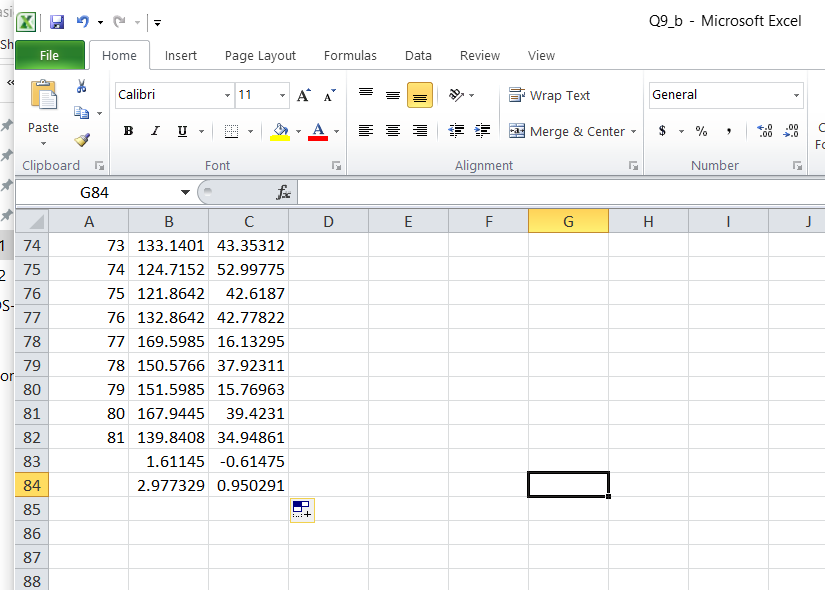
**Use Q9\_a.csv**

****

***speed***: skewness = -0.12, kurtosis= - 0.51 . The distribution is negatively skewed slightly and platykurtic and is close to normal distribution.

***distance***: skewness=0.81, kurtosis=0.41 . The distribution is positively skewed and leptokurtic and differs moderately from normal distribution.

**Use Q9\_b.csv**

****

***SP*:** skewness= 1.61, kurtosis= 2.98 . The distribution is positively skewed heavily and is leptokurtic and is different from normal distribution in a significant way.

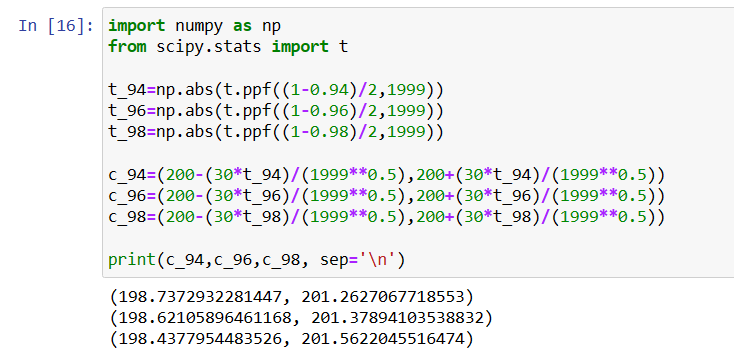
***WT*:** skewness= -0.61, kurtosis= 0.95 . The distribution is negatively skewed and is leptokurtic and differs from normal distribution moderately.

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



**Answer: - The data is positively skewed with a most of the values below the mean and a lot of outliers.**

**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 :-** 

**94% confidence interval=(198.74,201.26)**

**96% confidence interval=(198.62,201.56)**

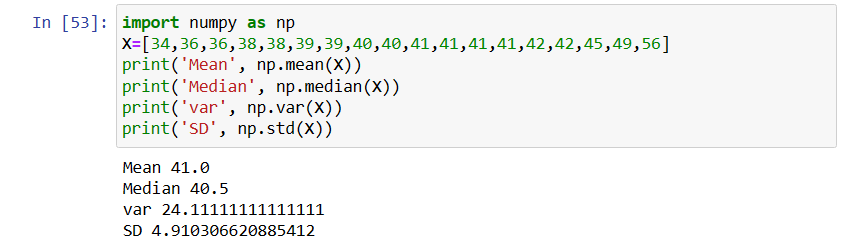
**98% confidence interval=(198.44,201.38)**

**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?

**Answer:-**



1. **Mean=41, median=40.5, variance=24.11, standard deviation=4.91**
2. **Mean and median are very close and the standard deviation is less.**

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

**Answer: - Symmetric**

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

**Answer: - Positively skewed**

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

**Answer: - Negatively skewed**

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

**Answer: - Leptokurtic - a steeper curve than normal distribution.**

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

**Answer: - Platykurtic - a flatter curve than normal distribution.**

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

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

**Answer:- The data is negatively skewed with large variance  
negatively skewed  
IQR=8**

Q19) Comment on the below Boxplot visualizations?



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

**Answer: - Median of both data is very close. The standard deviation and variance of boxplot 1 is much smaller than that of boxplot 2. Skewness of both boxplots is close to zero with mean also equal or close.**

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

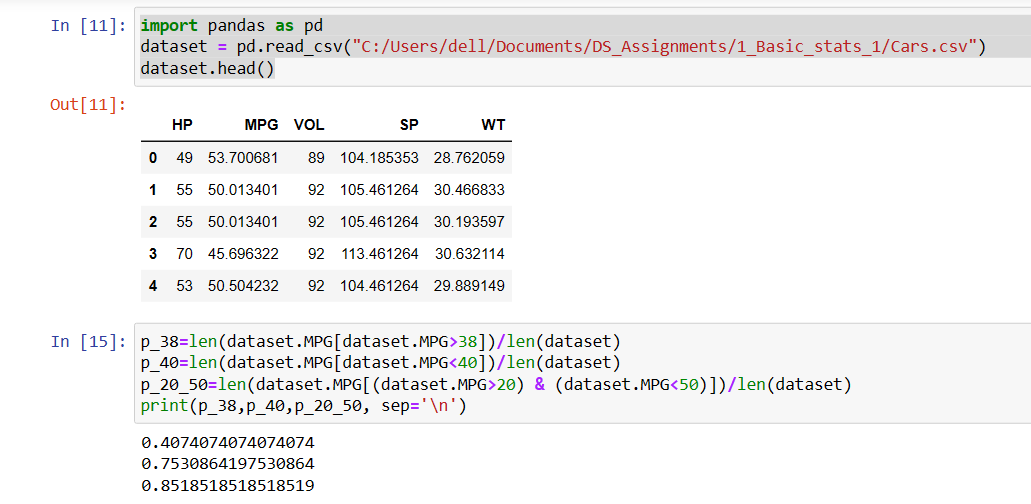
Data \_set: Cars.csv

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)

**Answer:-**

****

1. **0.41**
2. **0.75**
3. **0.85**

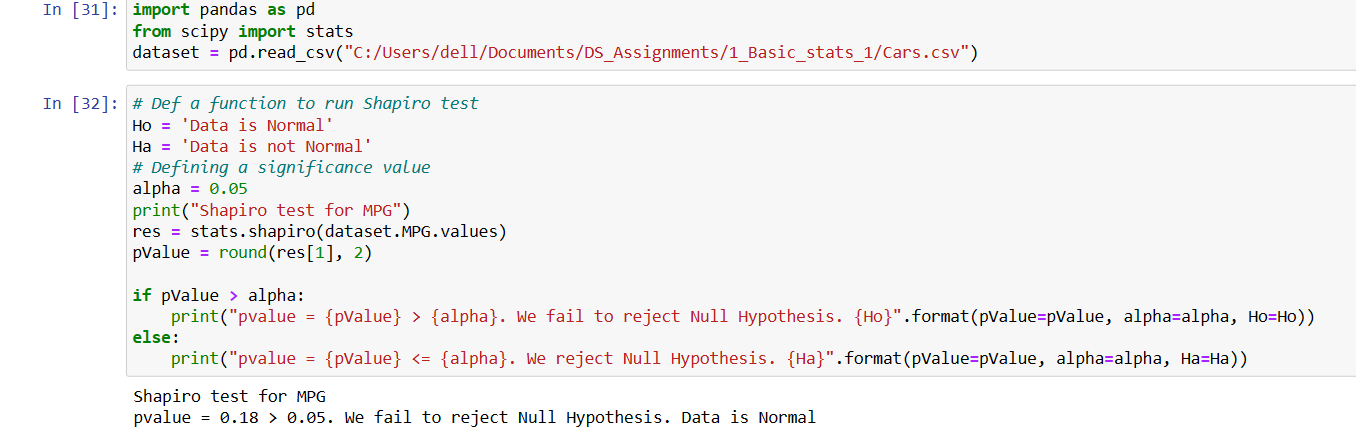
Q 21) Check whether the data follows normal distribution

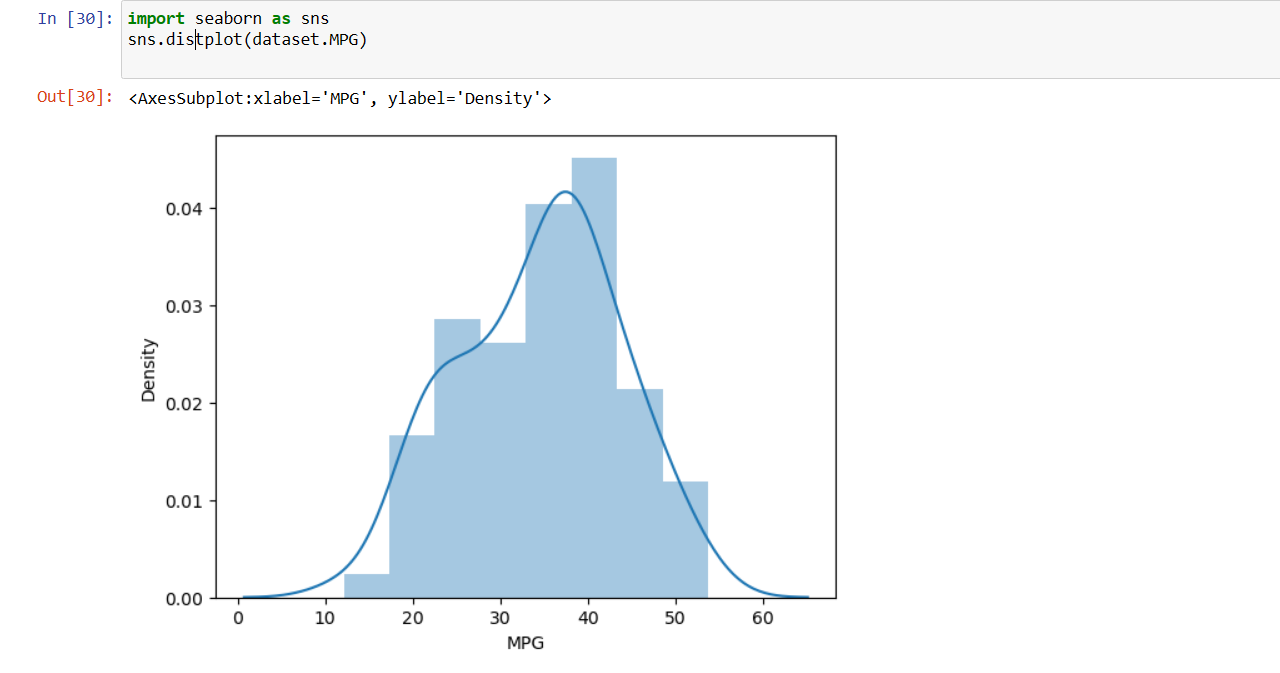
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Answer**:-

**Based on shapiro test and the density graph the data follows a normal distribution.**





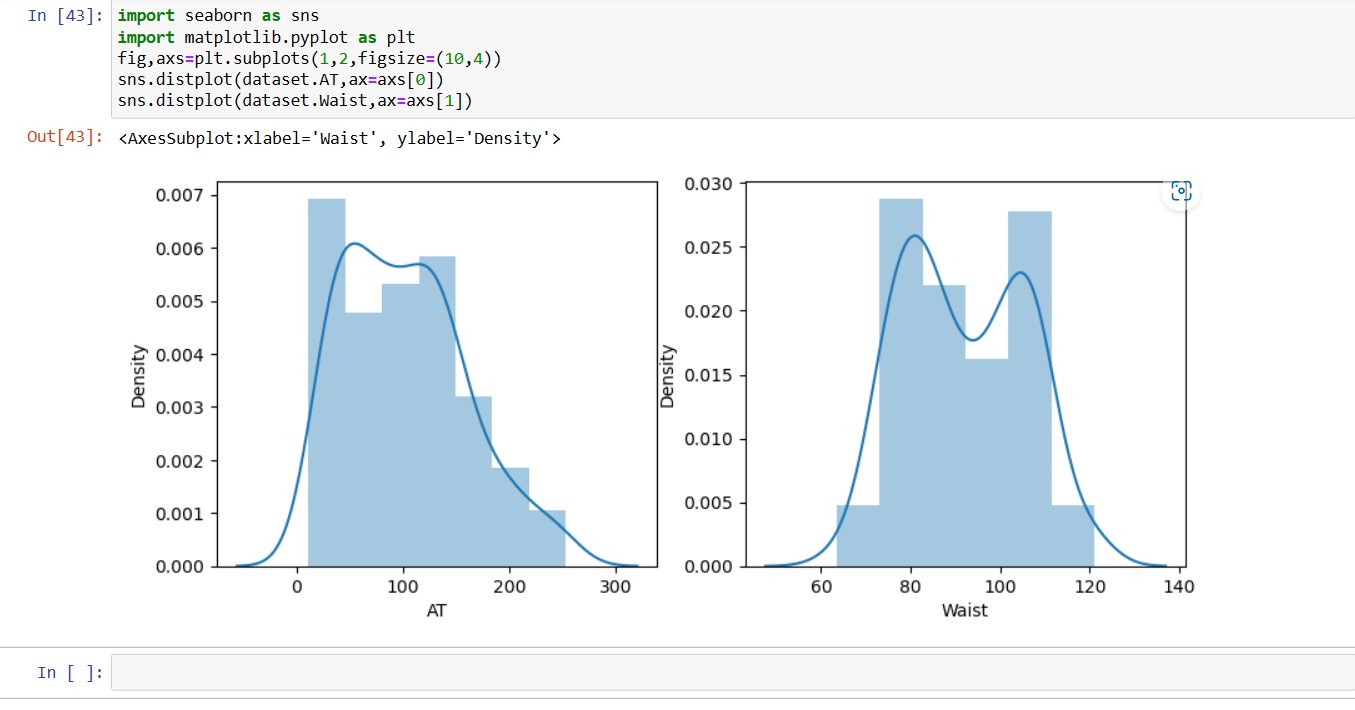
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

**Answer:-**

Based on shapiro test and the density graph the data does not follows a normal distribution.





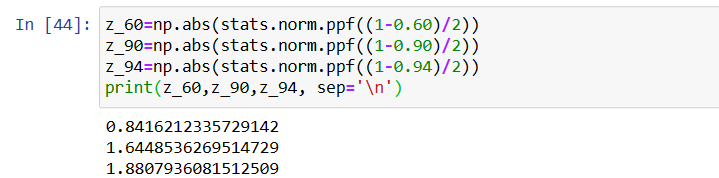
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

**Answer:-**

**60%: 0.842**

**90%: 1.645**

**94%: 1.88**

****

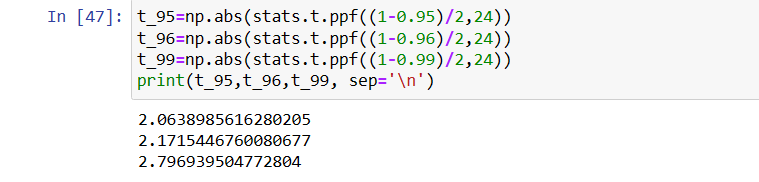
Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

**Answer:-**

**95%: 2.06**

**96%: 2.17**

**99%: 2.80**

****

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

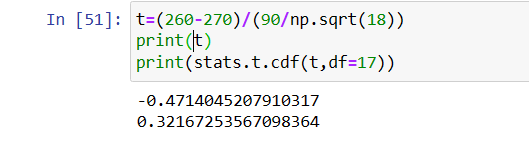
Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

**Answer:-**

**Probability=0.32**

****