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

**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 | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Ordinal |
| SAT Scores | Interval |
| Years of Education | Interval |

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

**Head and Tail = 2x2x2 = 8 Combinations.**

T T T = No

T T H = No

T H T = No

T H H = Yes

H T T = No

H T H = Yes

H H T = Yes

H H H = No

**Only 2 out of 8 Probability = 3/8.**

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

1. **Equal to 1**

Ans: a) = 0.

1. **Less than or equal to 4**

Ans: b) 4 = 6/36, 1/6

1. **Sum is divisible by 2 and 3**

Ans: c) (1,5) (2,4) (3,3) (4,2) (5,1) (6,6) then probability = 6/36, 1/6

2Dice =

**[(**1,1) (1,2) (1,3) (1,4) (1,5) (1,6) = 3 combination (1,1) (1,2) (1,3**)]**

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

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

(4,1) (4,2) (4,3) (4,4) (4,5) (4,6) = 0000

(5,1) (5,2) (5,3) (5,4) (5,4) (5,6) = 000

(6,1) (6,2) (6,3) (6,4) (6,5) (6,6) = 000

**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: Red balls = 2

Green balls = 3

Blue balls = 2

Total number of balls is 7.

Probability of getting number of blue balls in 1st drawn = p1 = 5/7

Probability of getting number of blue balls in 2nd drawn = p2 = 4/6 = 2/3

Probability that none of the balls drown is blue = p1 x p2 = 5/7 x 2/3 = 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** |

**Ans:** Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

Child C - probability of having 3 candies = 0.65

Child D - probability of having 5 candies = 0.005

Child E - probability of having 6 candies = 0.01

Child F - probability of having 2 candies = 0.120

= 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

**The Expected number of candies for randomly selected child is = 3.09**

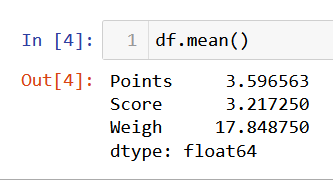
**Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset**

* **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**

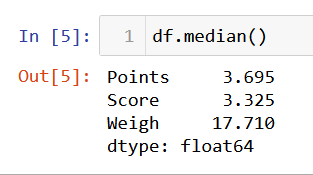
**Mean**.



Mean:

**Points = 3.596563, Score = 3.217250, Weigh = 17.848750.**

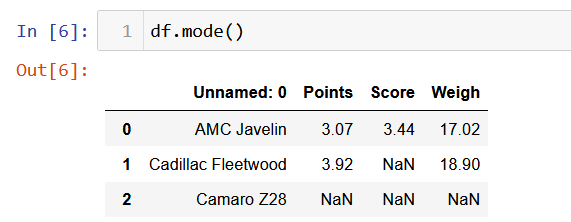
**Median.**

****

Median**:**

**Points = 3.695, Score = 3.325, Weigh = 17.710.**

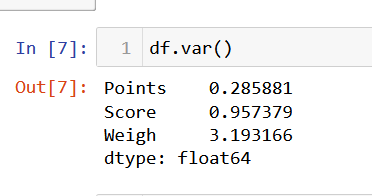
Mode**:**

****

**Mode:**

**Points = 3.07, Score = 3.44, Weigh = 17.02.**

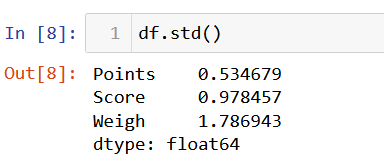
Variance:



**Variance:**

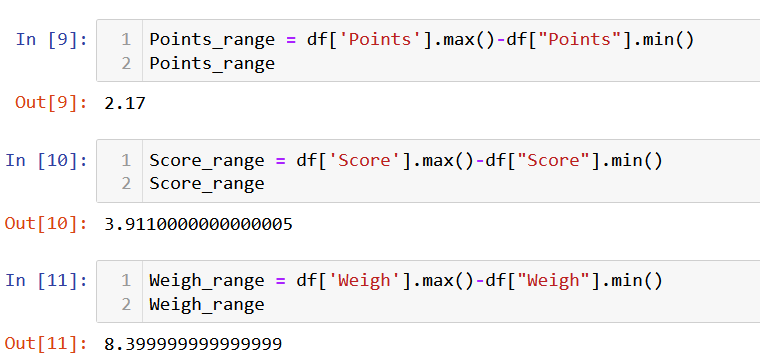
**Points = 0.285881, Score = 0.957379, Weigh = 3.193166.**

Standard Deviation:



**Standard Deviation:**

**Points = 0.534679, Score = 0.978457, Weigh = 1.786943.**

****

**Range:**

**Points = 2.17, Score = 3.9110000000000005, Weigh = 8.399999999999999.**

**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

P(x)Probability of selecting each patient is 1/9.

Expected Value = ∑ (probability \* Value)

 ∑ P(x). E(x)

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

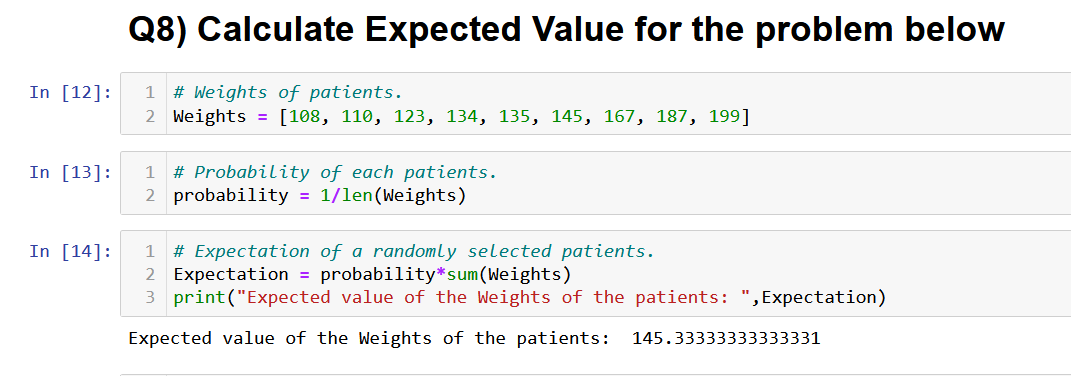
P(x) 1/9 1/9   1/9 1/9   1/9   1/9   1/9   1/9 1/9

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

= (1/9) (1308)

= 145.33

Expected Value of the Weight of the patient = 145.33

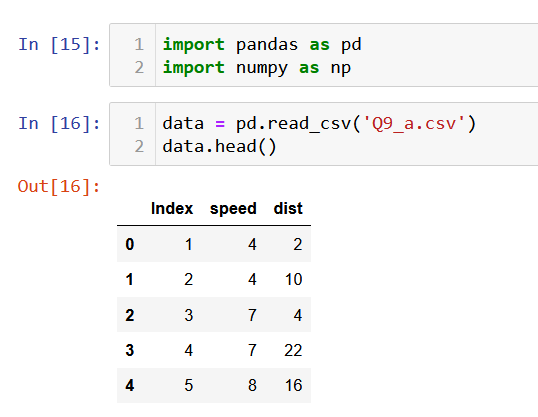


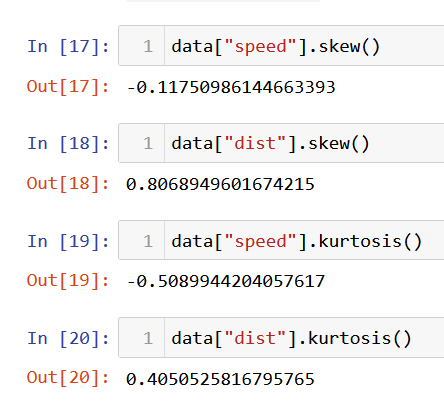
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans**:



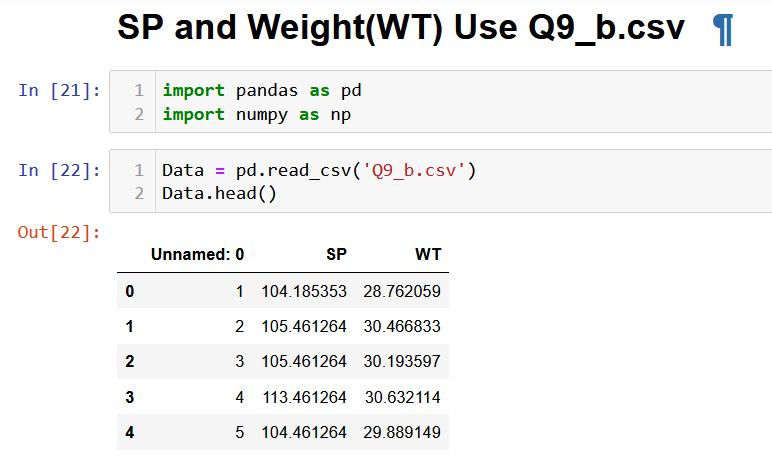


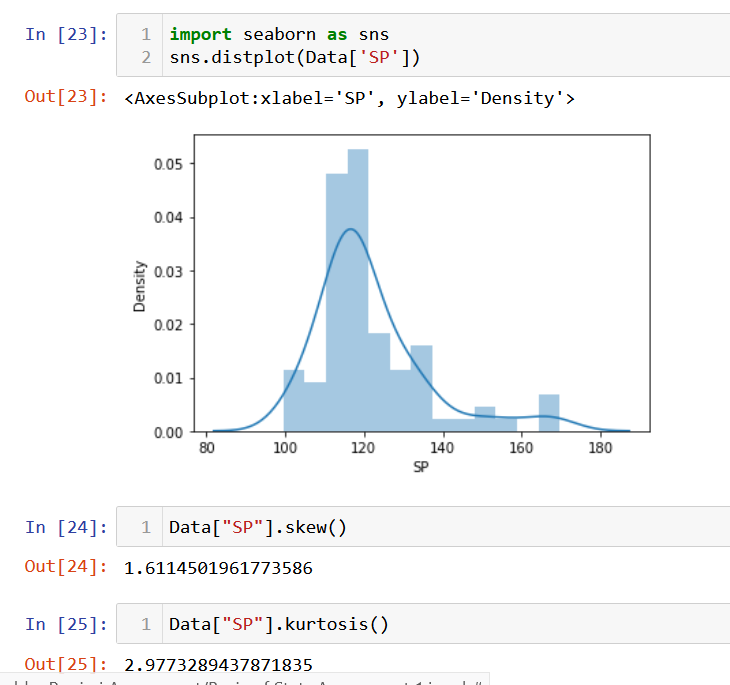
Reference:

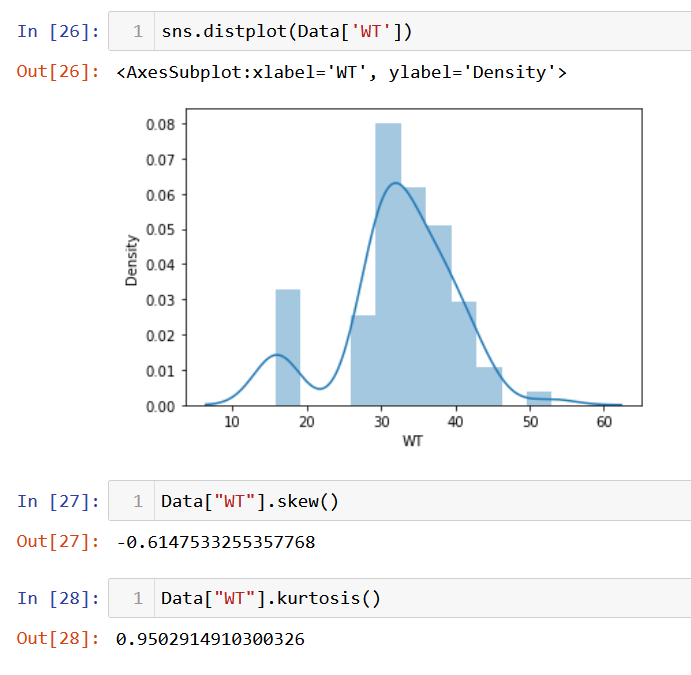
* The skewness of speed is negative it shows that the data are negatively skewed or left skewed distribution.
* 2. Mean<Median<Mode
* 3. The kurtosis of speed is negative which means that it is platykurtic in which more data locate near the mean and fewer outliers in the dataset.
* 4. The kurtosis of speed is positive which means that it is leptokurtic in which their outliers are exist.
* 5. Mean>Median>Mode.

**SP and Weight(WT) Use Q9\_b.csv**

**Ans:**

****

****

****

**Reference:**

* The skewness of SP is positive it shows that the data are negatively skewed or right skewed distribution.
* Mean>Median>Mode.
* The skewness of distance is negative which signifies it a left skewed Distribution.
* The kurtosis of speed is positive which means that it is Leptokurtic in which their outliers are exist.
* Mean<Median<Mode.

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



Ans:

Referance:

* The above histogram shows that the values are positively skewed.
* The thin tail present towards the right side.
* Skewness>0.
* Mean>Median>Mode.



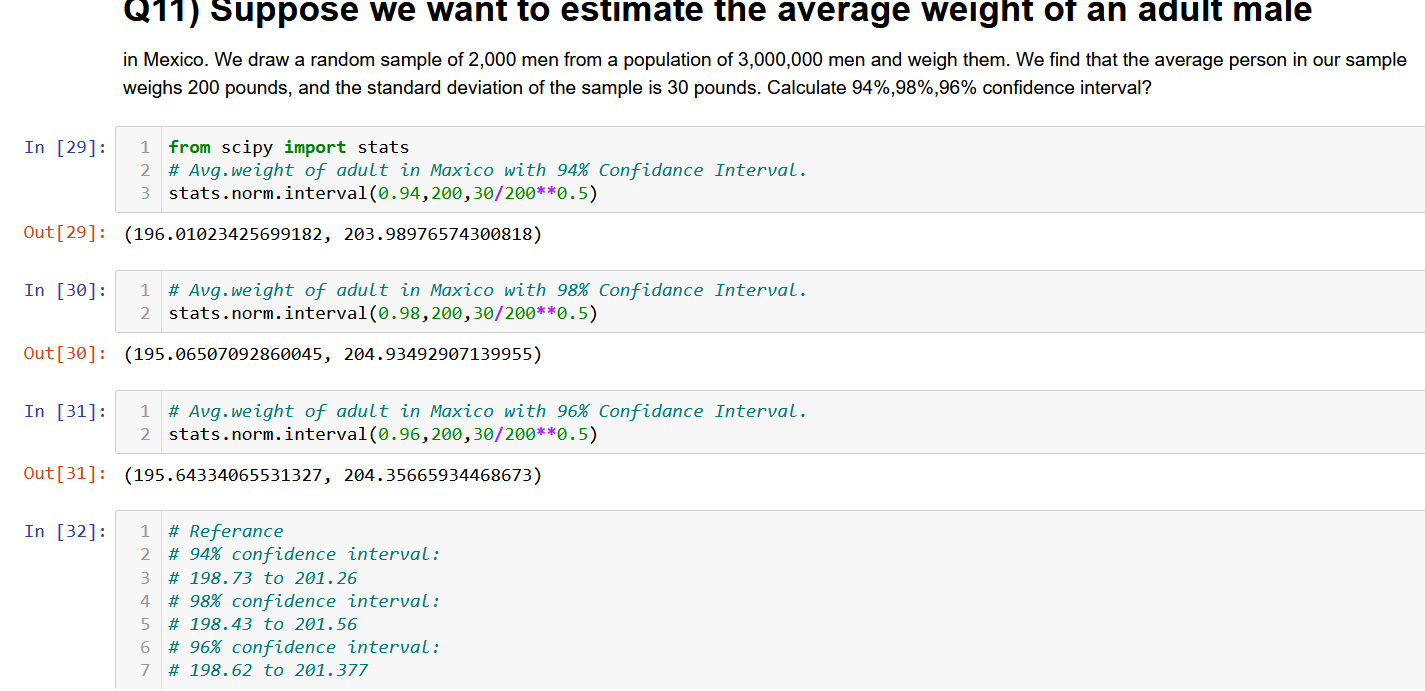
Ans:

Referance:

* The box plot shows that greater data are present towards the left.
* The data are positively skewed or right skewed.
* The data contains outliers in the positive direction which should be avoided.

**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?**

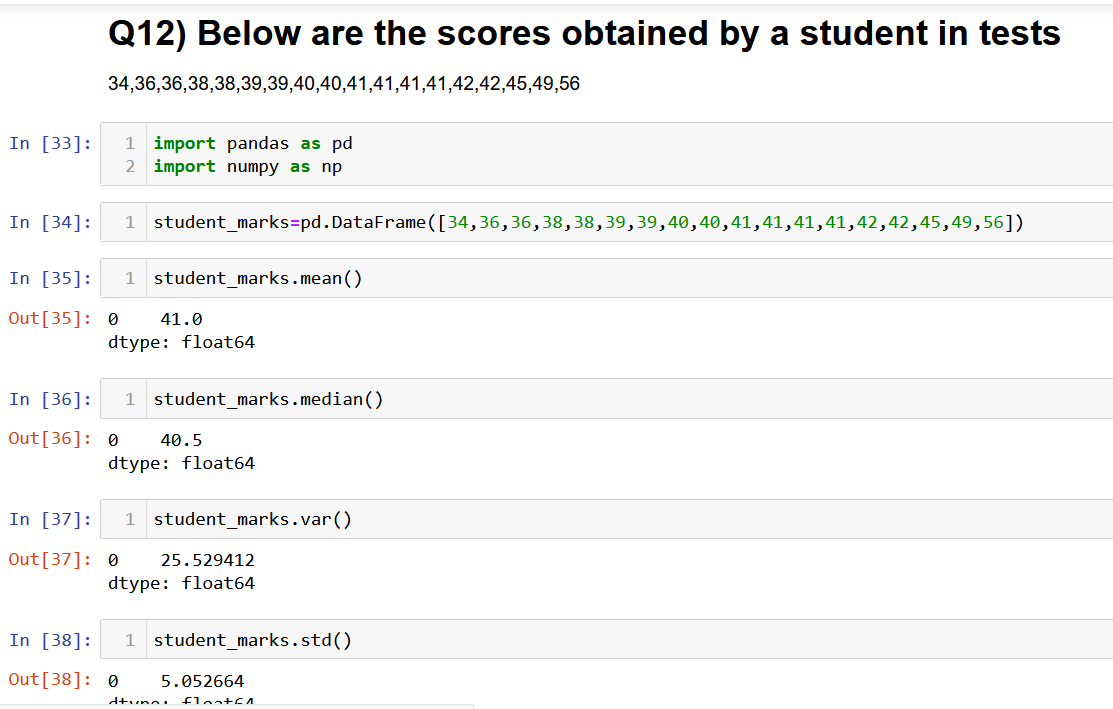
Ans:

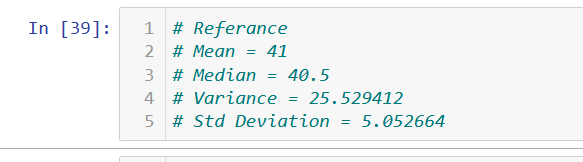


**Q12) Below are the scores obtained by a student in tests**

**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.**





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

Ans:

1. The average marks of student in test are 41.

2. The range of the marks of the students are 56 to 34.

3. Most of the student secured 41 marks.

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

Ans: No Skewness.

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

Ans: Positive skewness

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

Ans: Positive skewness.

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

Ans: In positive kurtosis (>3), the distribution is peaked and has thick tail which

indicates Most of the value in the distribution located in tail rather than around the mean.

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

Ans: In negative (<3), the distribution is flat and has thin tail which indicates there are fewer in its shorter tail.

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



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

**Ans:**

* 1. The distribution of data is not normally distribution.
* 2. Most of the value in left so it is left skewed.
* 3. The median value of the distribution is 15.
* 4. 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.

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

**Ans:** IQR = QR3 – QR 1

= 18-10 = 8

**Q19) Comment on the below Boxplot visualizations?**



**Ans**:

The given box plot is normally distributed, both have median around 262.5

and they have no outliers.

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

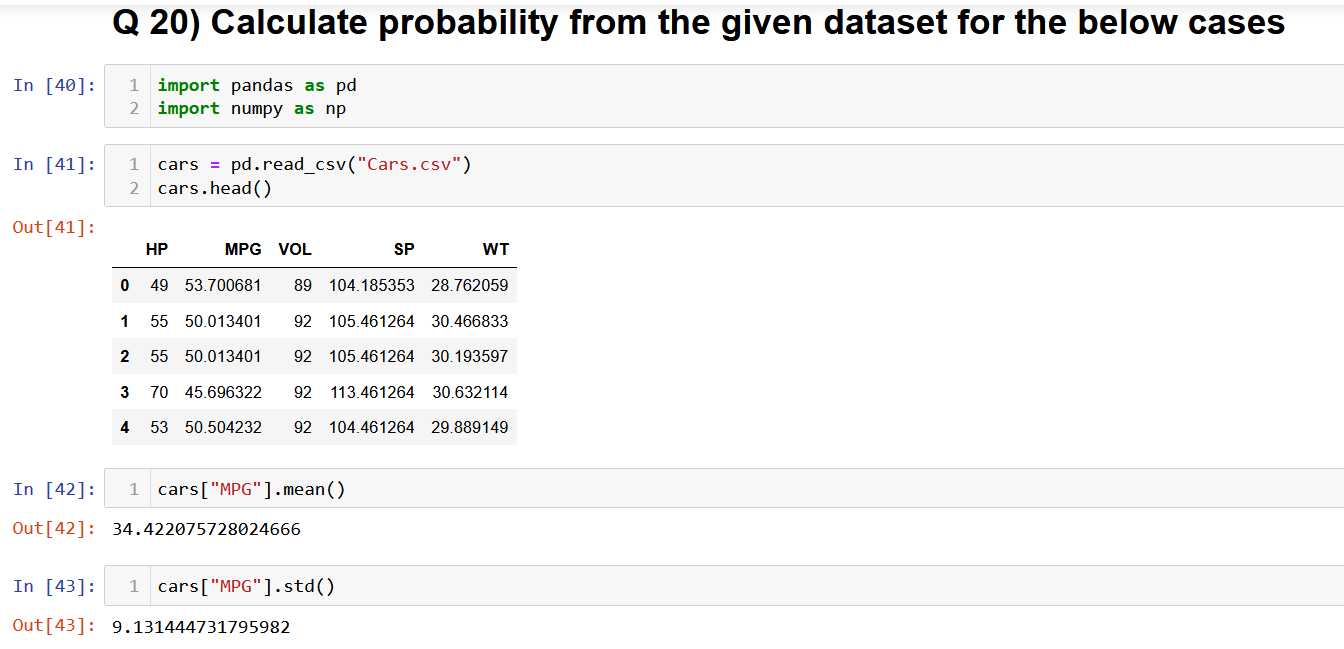
Data \_set: Cars.csv

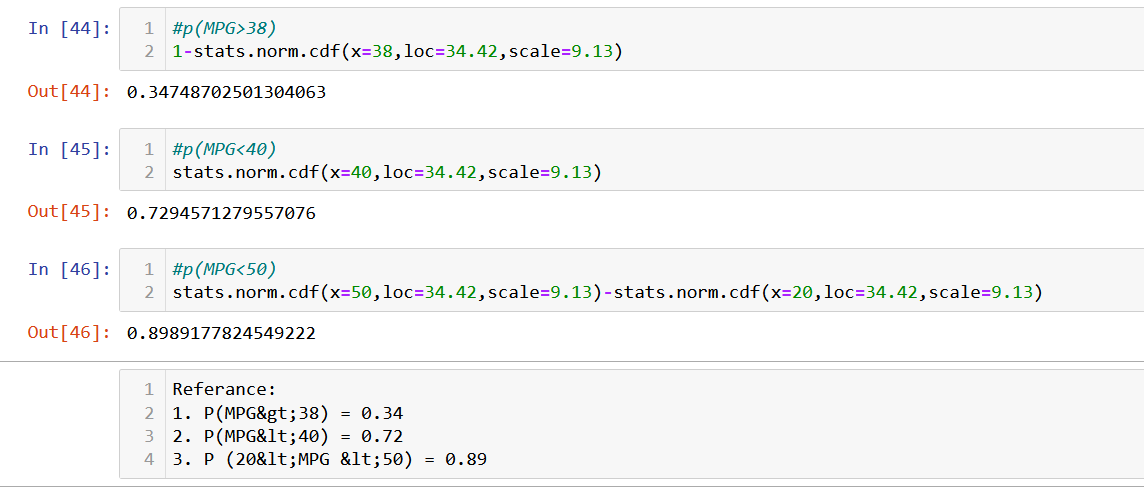
Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

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

c. P (20<MPG<50)



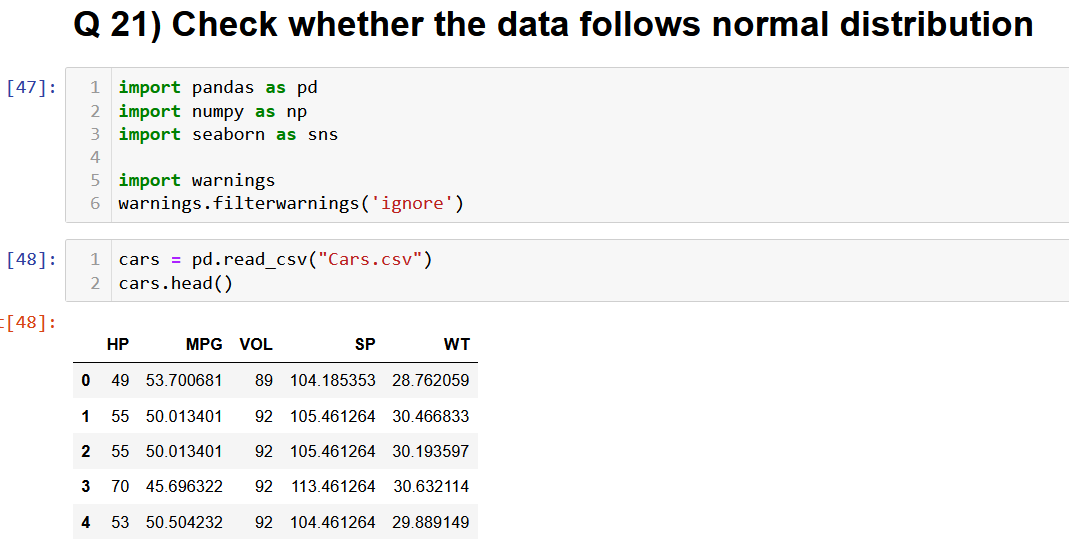


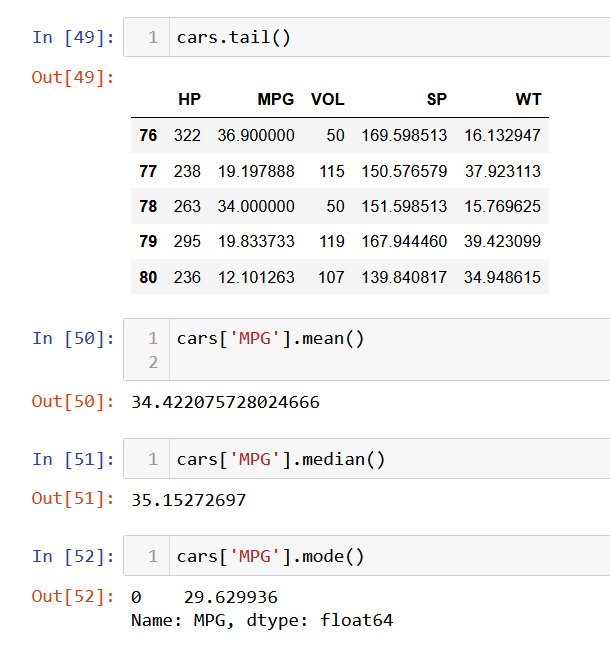
**Q 21) Check whether the data follows normal distribution**

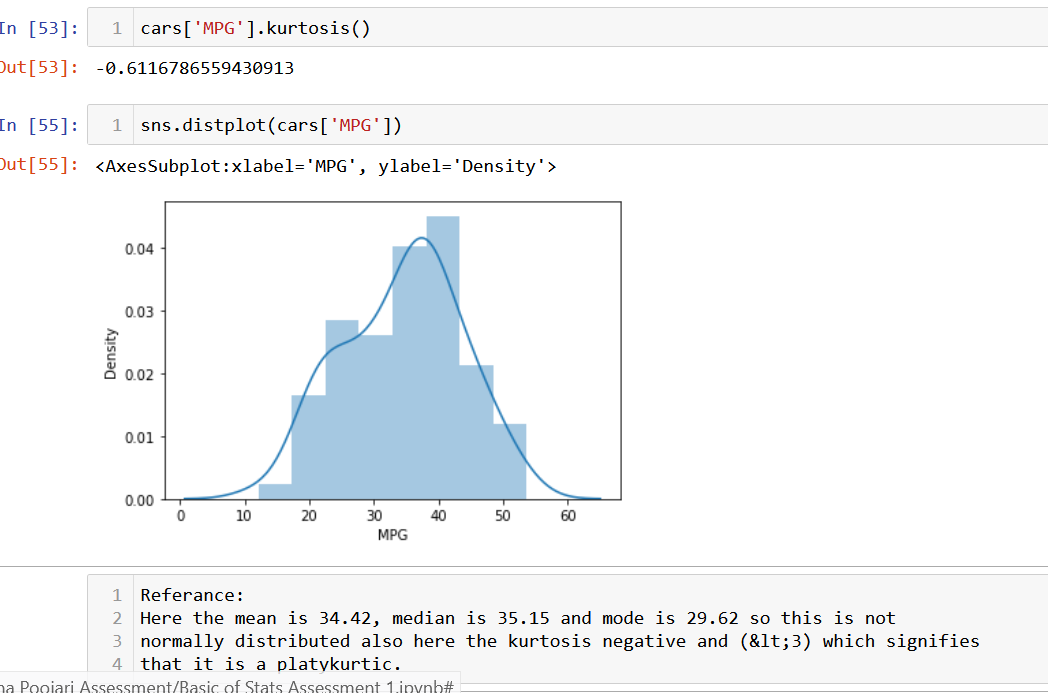
1. **Check whether the MPG of Cars follows Normal Distribution**

**Dataset: Cars.csv**

**Ans:**

****

****

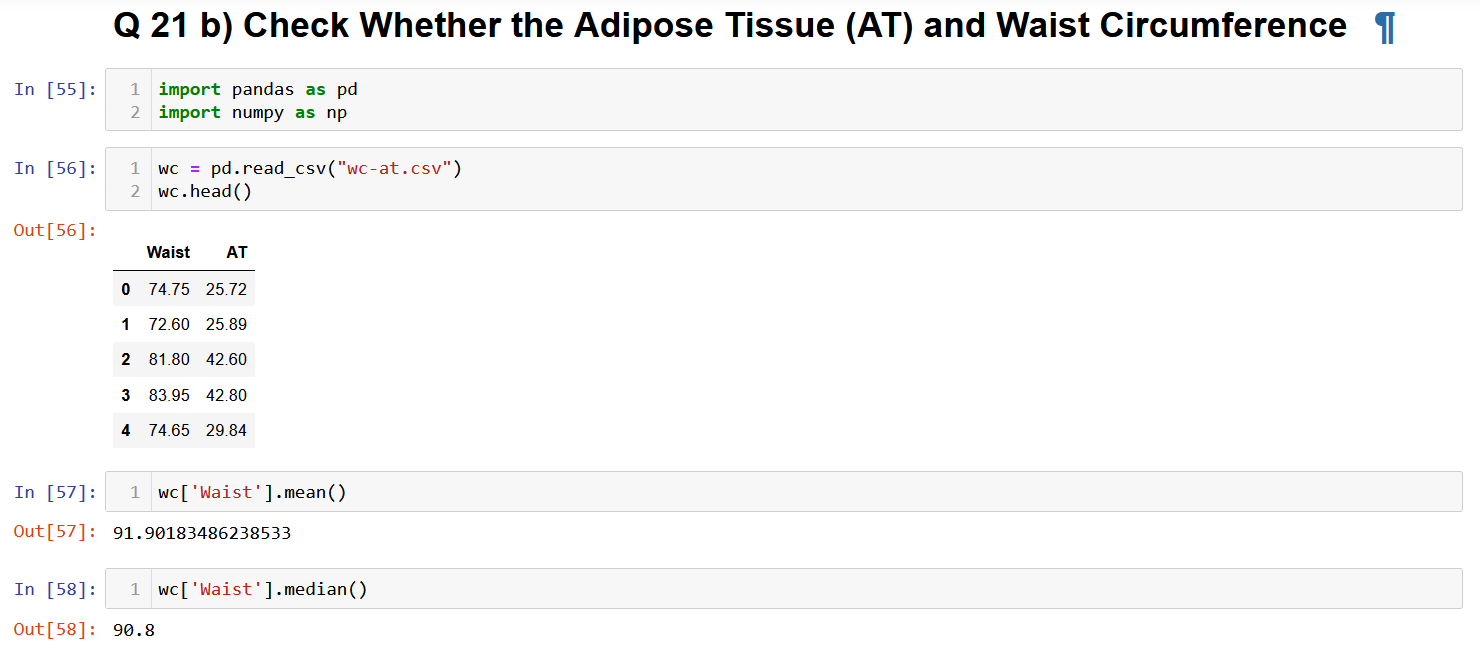
****

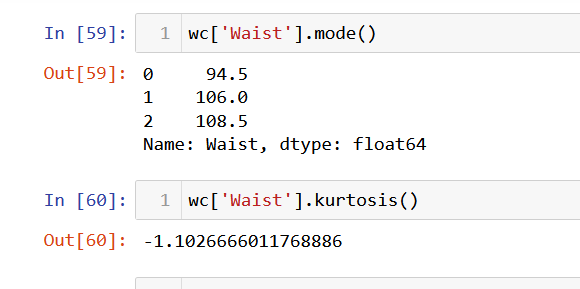
**b) Check Whether the Adipose Tissue (AT) and Waist Circumference**

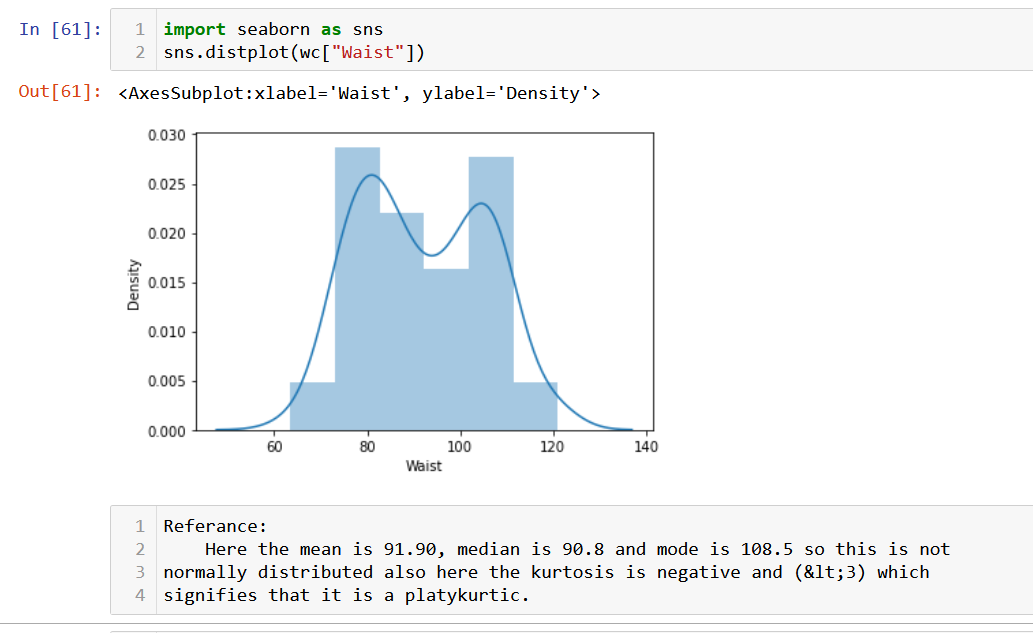
**(Waist) from Wc-at data set follows Normal Distribution**

**Dataset: wc-at.csv**

**Ans:**

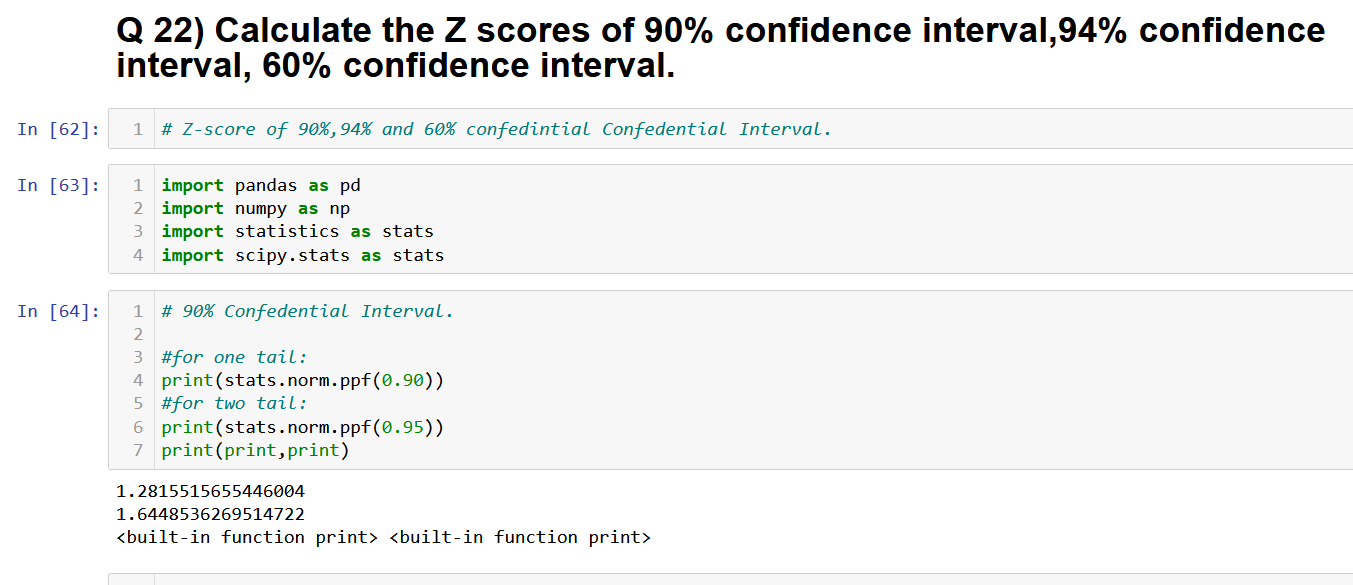
****

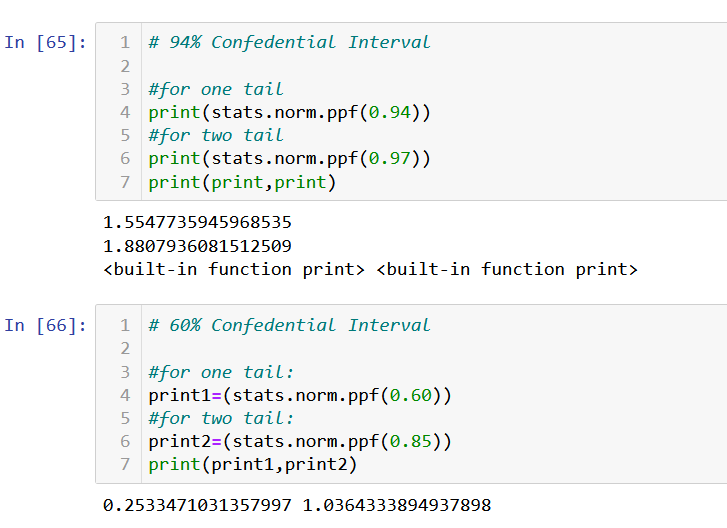
****

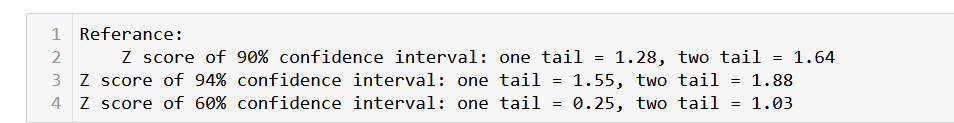
****

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

**Ans:**

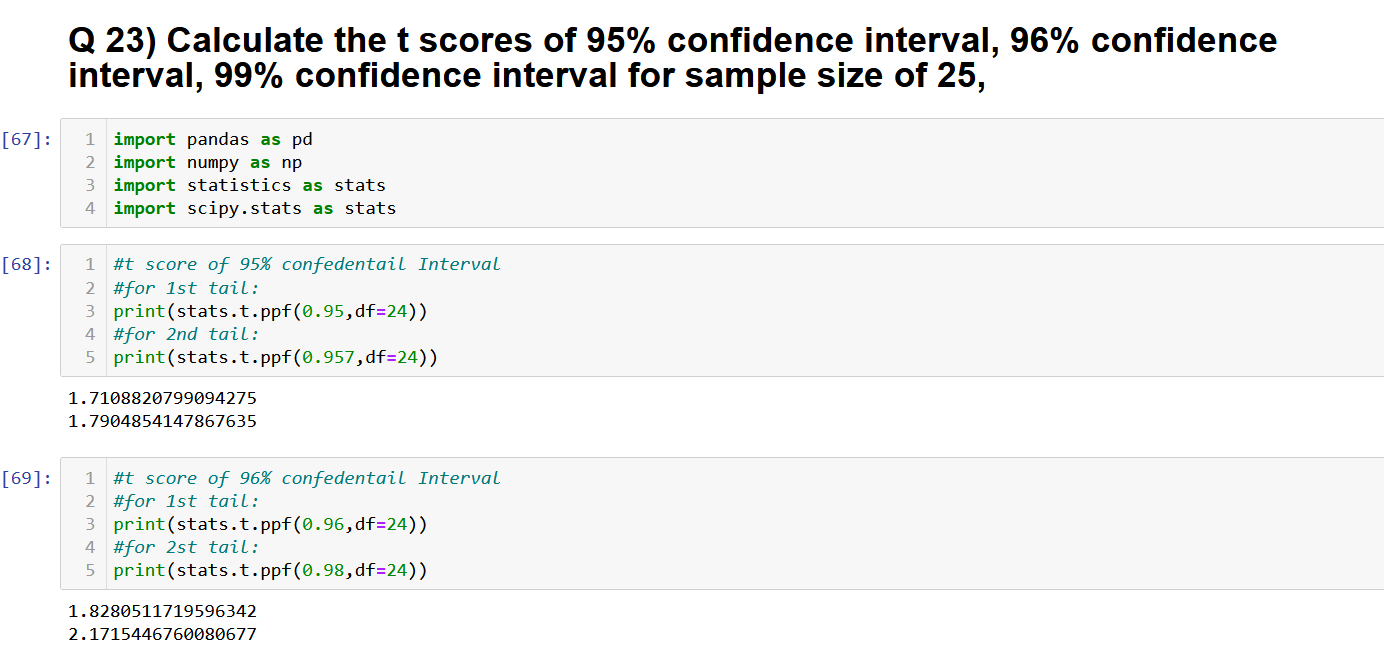
****

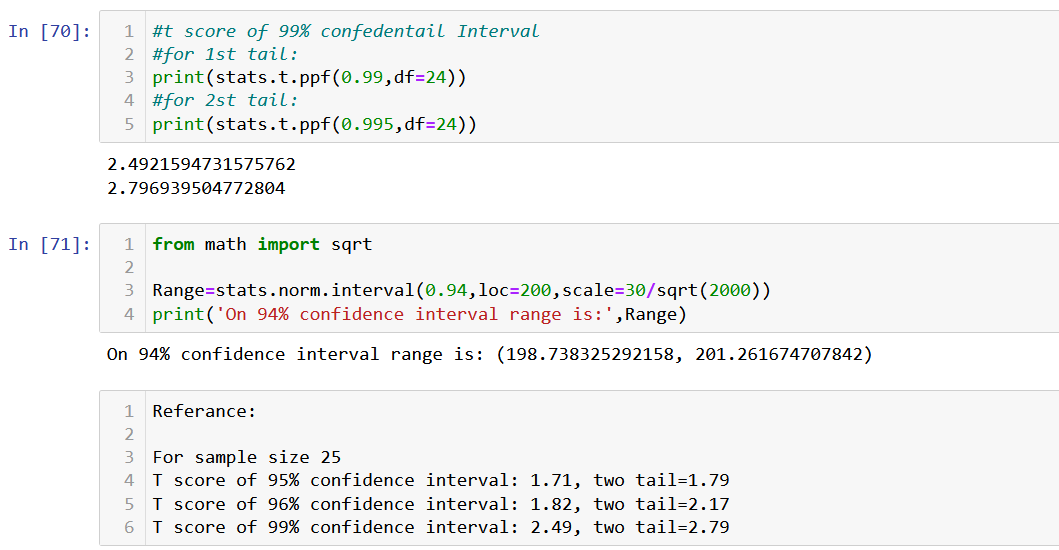
****

****

**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**

Ans:

