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

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 | Interval |
| Celsius Temperature | Interval |
| Weight | Interval |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Interval |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| 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?

Solution *:* The probability of two heads and one tail is =3/8 =0.375

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

Solution:

1. If two dice are rolled, then total possible cases = 36

The probability of the sum equal to 1 = 0

As the minimum sum is 2 with the outcome of (1,1)

1. The probability of the sum less than or equal to 4 = 6/36 =1/36

The possible outcomes are (1,1) (1,2) (2,1) (2,2) (1,3) (3,1)

c) the probability of sum is divisible by 2 and 3 = 6/36 = 1/6

The possible outcomes are (1,5) (2,6) (3,3) (4,2) (5,1) (6,6)

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?

Solution: The probability of none of the balls drawn is blue = 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

Solution: The expected number of candies for a randomly selected child

= (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

= 3.09

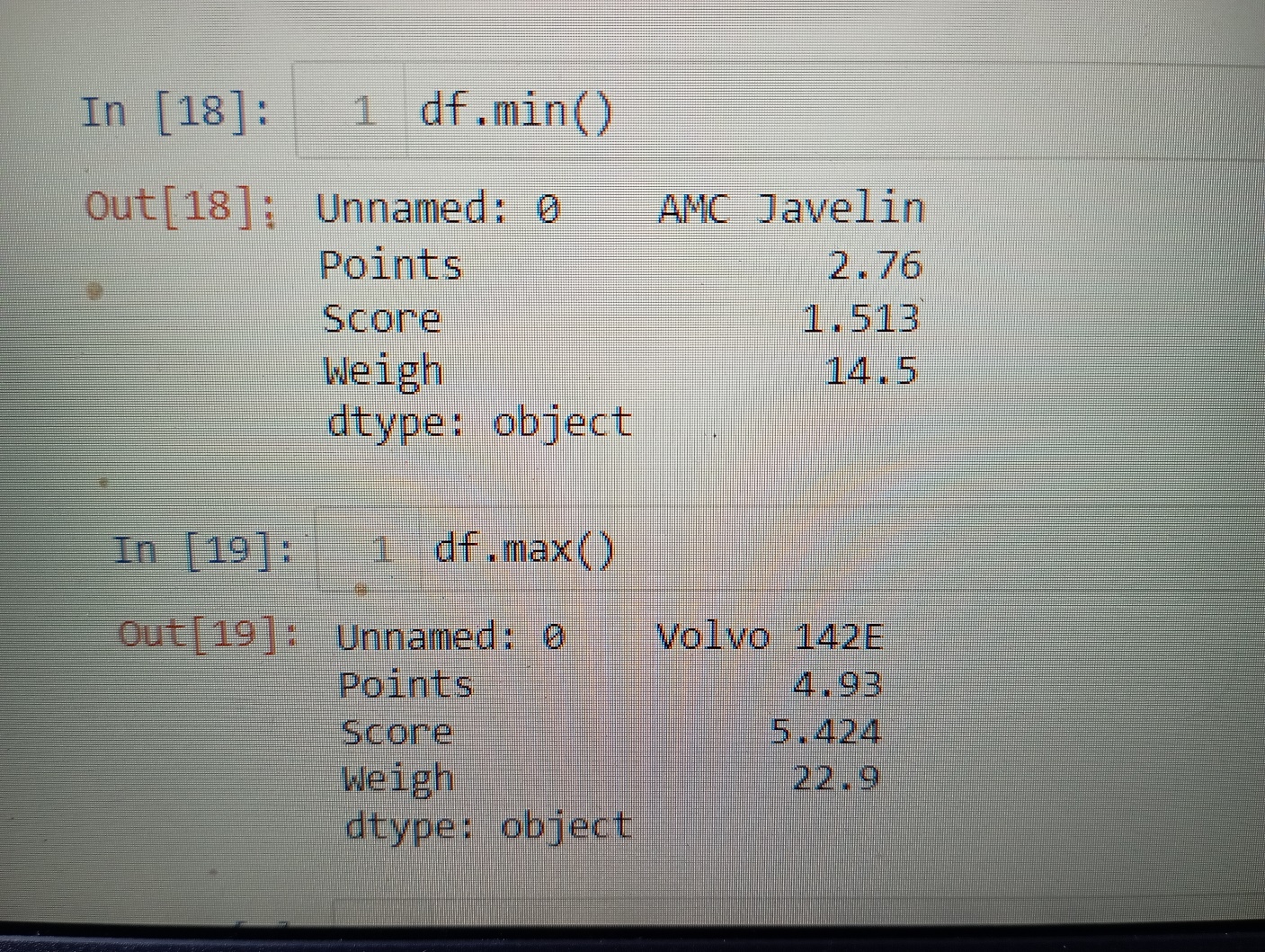
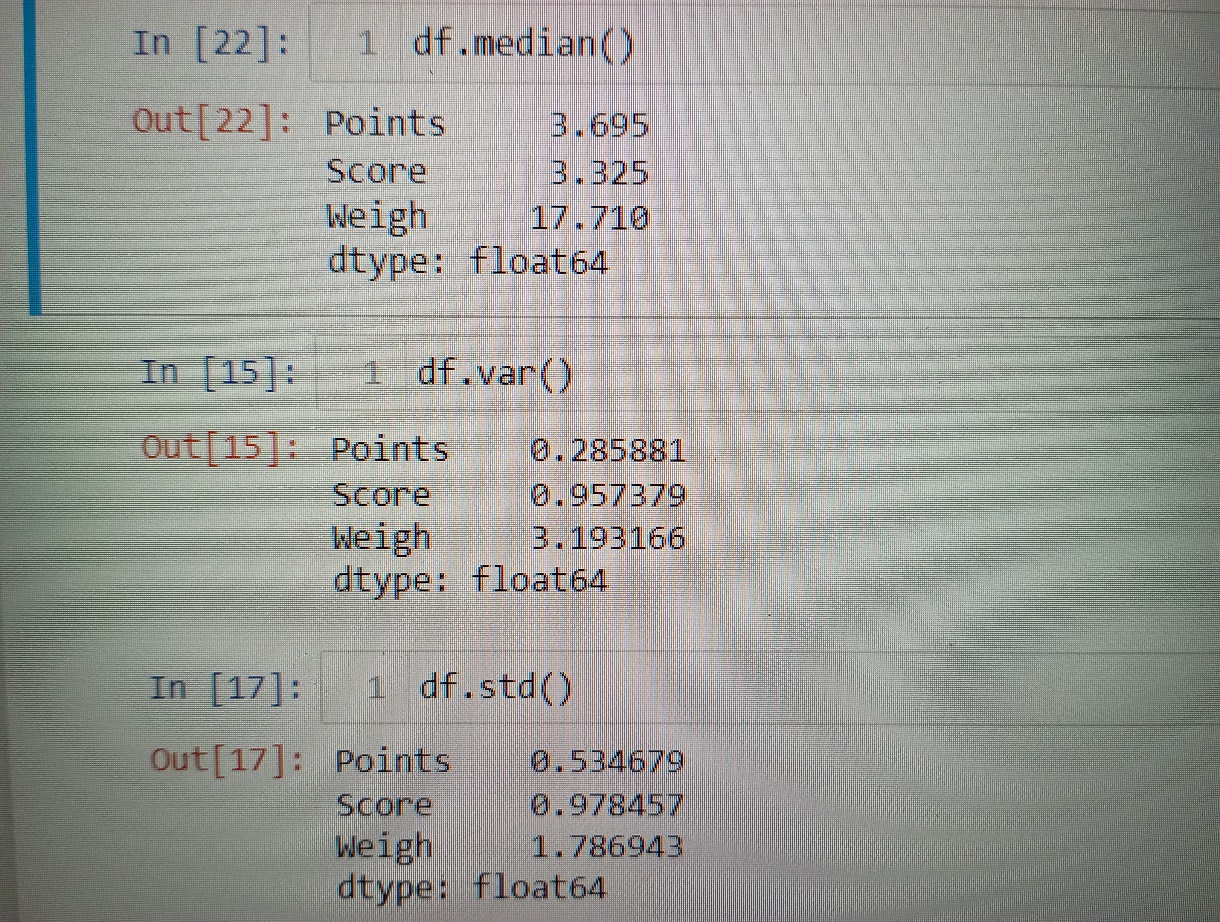
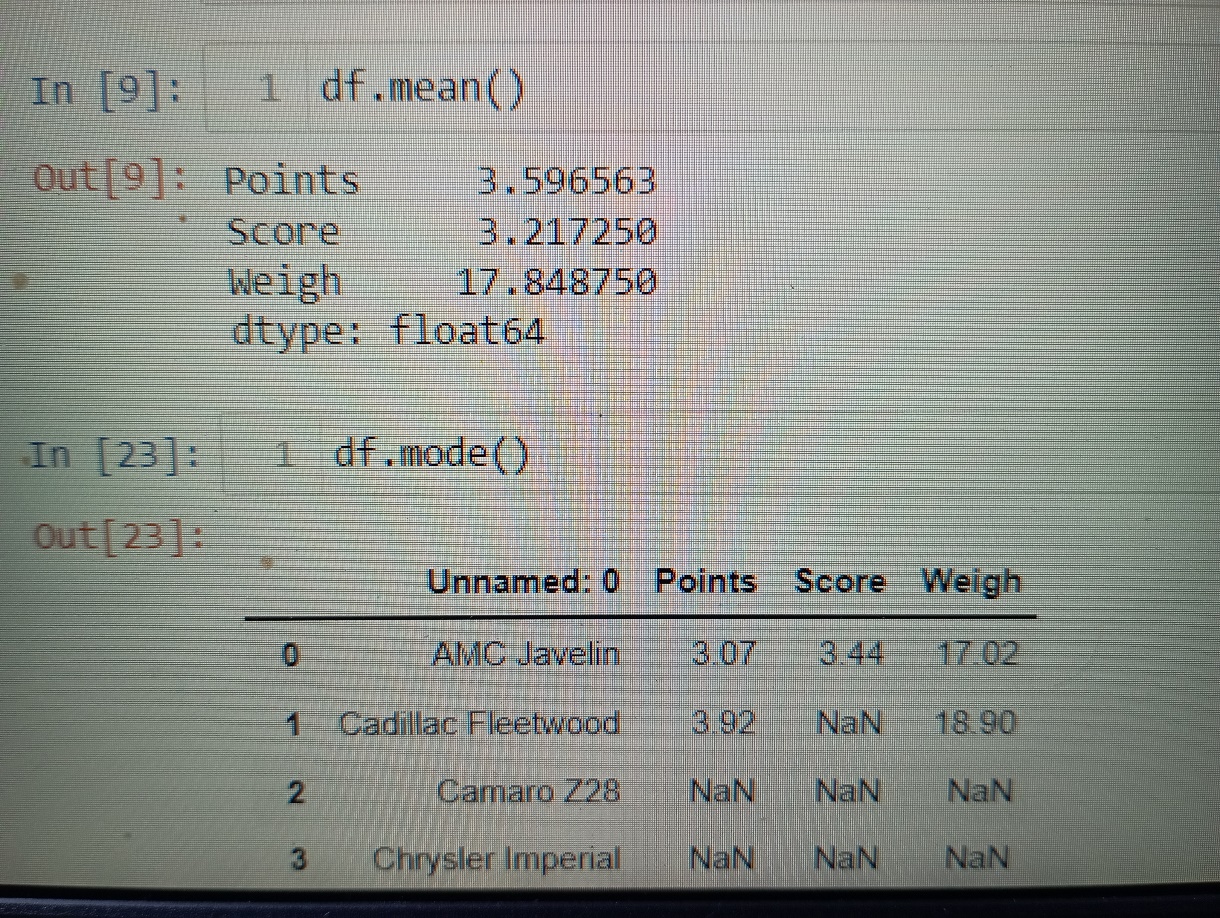
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* 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**

**Solution : By using Python**

****

* Mean and median are approx. equal, so there are no outliers on this data set.
* The data set is normally distributed as there is no outliers.
* Skewness is zero as dataset is normally distributed.

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?

Solution:

X 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

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

= 145.3333

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

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

Solution:

1. Speed

Skewness = -0.11751

Kurtosis = -0.50899

Distance

Skewness = 0.806895

Kurtosis = 0.405053

Inferences:

* Variable speed is negatively skewed data because the concentration of data is on the left side.
* Variable distance is positively skewed data because the concentration of the data is on the right side.
* Variable speed is negative kurtosis because the distribution is flat.
* Variable distance is positive kurtosis because the distribution is peaked.

1. SP

Skewness = 1.61145

Kurtosis = 2.977329

Weight (WT)

Skewness = -0.61479

Kurtosis = 0.950291

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



Solution:

Inference

* The above histogram has positive outliers.
* The above boxplot suggest that the distribution has a lot of outliers towards the upwards extreme.

**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?

Solution:

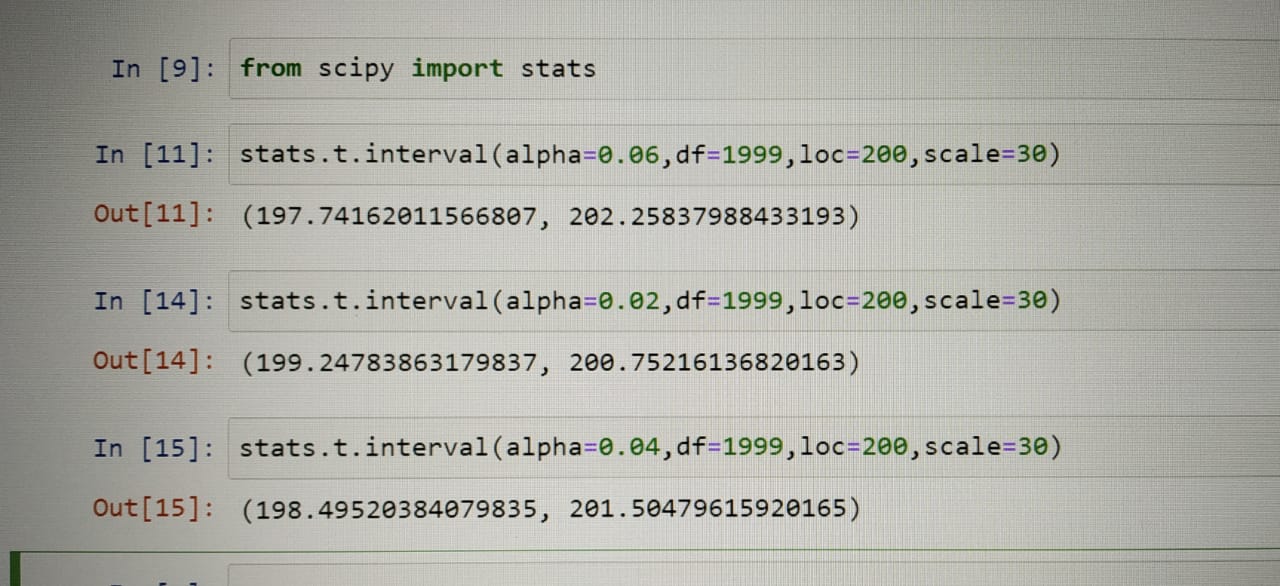
N = 30,00,000

n = 2000

µ = 200

s = 30

By using python,



**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?

Solution:

Mean = 41

Median = 40.5

Variance = 25.52941

Standard deviation = 5.052664

Mean > Median. This implies that the distribution is slightly skewed towards the right.

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

Solution: If Mean = Median = Mode, the distribution has skewness and is symmetric.

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

Solution: If mean is greater than median, the distribution is positively skewed.

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

Solution: If median is greater than mean, the distribution is negatively skewed.

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

Solution: It has high peakness and less variation.

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

Solution: It has less peakness and more variation.

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



What can we say about the distribution of the data?

Solution: Distribution is not normal according to the figure.

What is nature of skewness of the data?

Solution: Distribution is negatively skewed according to the figure.

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

Solution:

Q3 = 18

Q1 = 10

IQR = Q3 – Q1

= 18 – 10

= 8

Q19) Comment on the below Boxplot visualizations?



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

Solution:

1. Both the boxplot 1 and boxplot 2 are normally distributed.
2. Both boxplot 1 and boxplot 2 are symmetric in nature.

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)

Solution:

a. P(MPG>38) = 0.34759

b. P(MPG<40) = 0.72935

c. P (20<MPG<50) = 0.89886

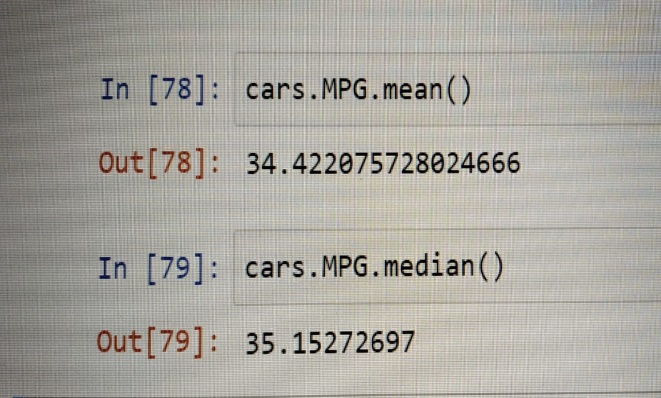
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Solution:

By using python,



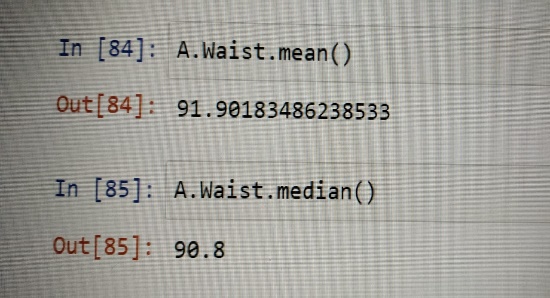
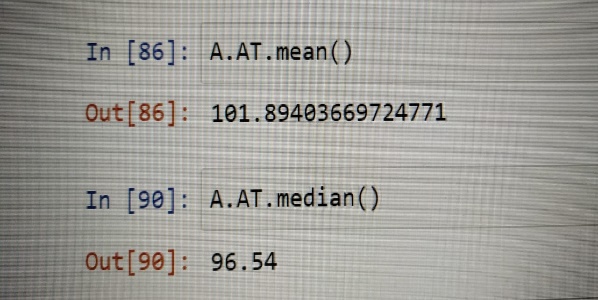
Since, the mean and median is approximately same them MPG of cars follows normal distribution.

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

Dataset: wc-at.csv

Solution:

By using python,

Since, mean and median is approximately same for waist. Then waist follows normal distribution.

Since, mean is greater than median for Adipose Tissue (AT). Hence, it follows normal distribution.

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

Solution:

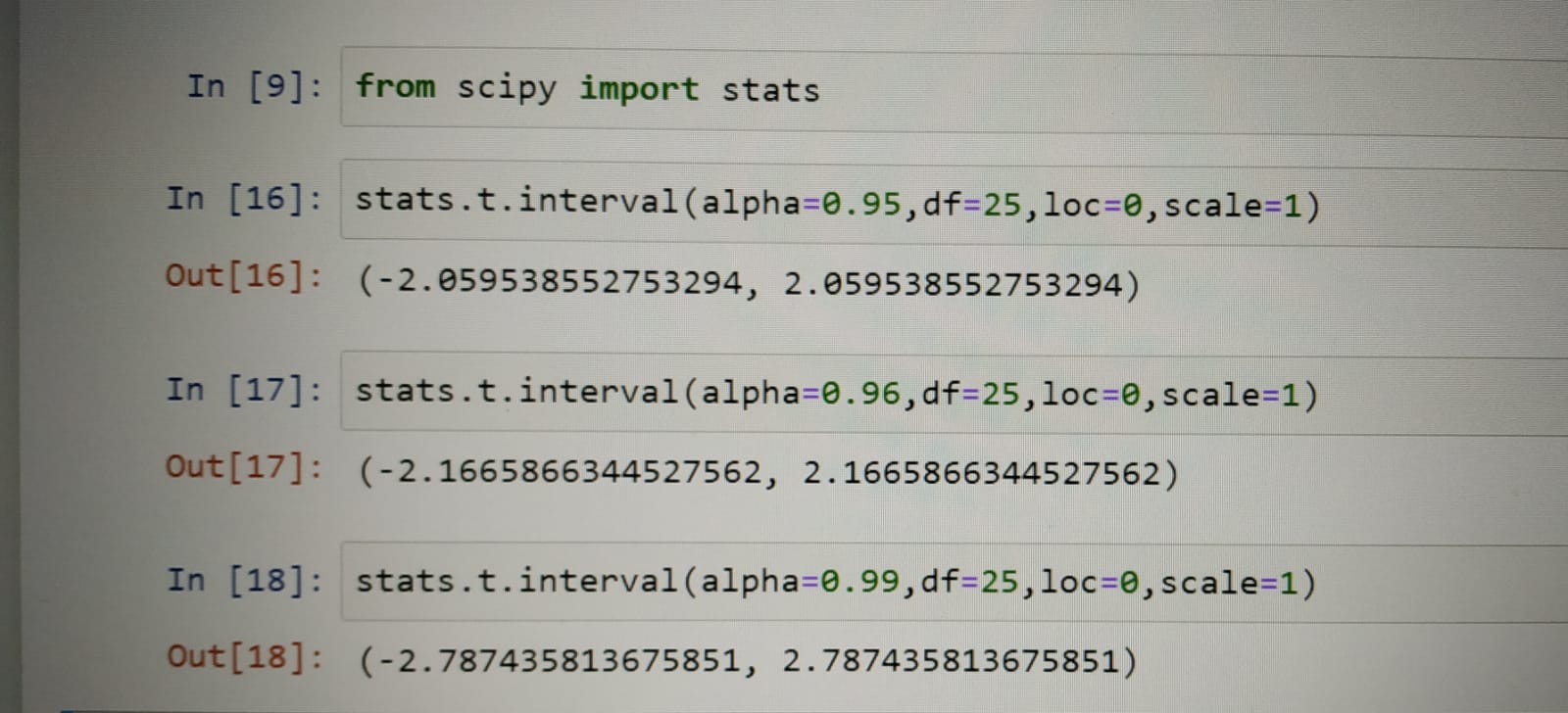
90% confidence interval, (0.95) Z score = 1.65

94% confidence interval, (0.97) Z score = 1.89

60% confidence interval, (0.8) Z score = 0.84

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

Solution: By using python,



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

Solution:

Population mean = µ = 270

Number of bulbs in a sample = n = 18

Sample mean = x = 260

Standard deviation of the sample = s = 90

t – statistics for the data is given as

t = )

t = (260 – 270)/ (90/√18)

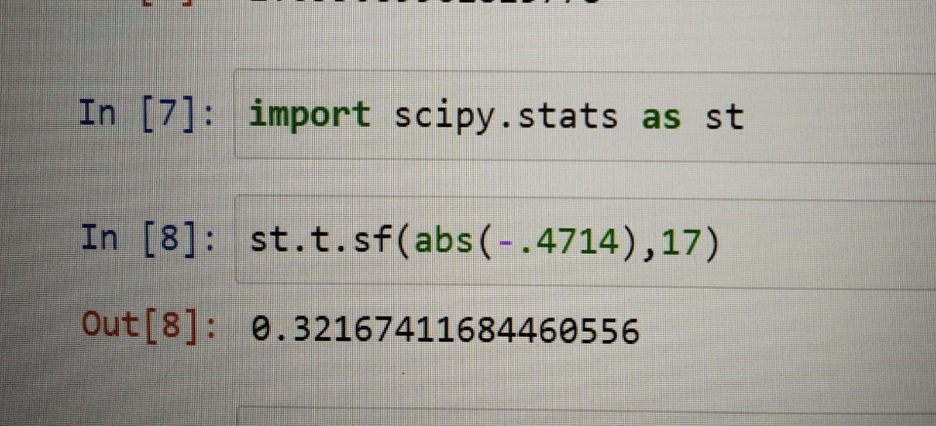
t = (-10)/ (90/3√2)

t = (-10)/ (30/√2)

t = -0.4714

degree of freedom = n-1 = 18 – 1 = 17

By using python,



18 randomly selected bulbs would have an average life of no more than 260 days = 0.3216 = 32.16%