|  |  |
| --- | --- |
| 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 | 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 | 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 | Nominal |
| Time on a Clock with Hands | Ratio |
| Number of Children | Nominal |
| 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 no. of outcomes = {HHH, HHT, HTT, TTT, THH, TTH, HTH, THT}= 8

Possibilities of getting two heads and one tail = {HHT, THH, HTH}= 3

Probability of two heads one tail= favourable events/total no. of events=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 no. of possible outcomes= 36

a) The sum of two dice can never be equal to 1

b) Favourable no. of events= {(1,1), (1,2), (1,3), (2,1), (2,2), (3,1)} = 6

Probability of getting the sum less than or equal to 4= 6/36=1/6

c) Favourable no. of events= {(1,5), (2,4), (3,3), (4,2), (5,1), (6,6)}

Probability of getting the sum divisible by 2 and 3= 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:** Total no. of possible outcomes= C (7,2) = 7! / (2! (7-2)!) = 21

Given 5 non blue balls in total i.e., (2 red + 3 green) balls.

No. of ways to draw two non-blue balls = C (5,2) = 5! / (2! (5-2)!) = 10

Therefore, Probability that 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

**Answer:** Expected no. of candies= (0.015\*1) +(0.20\*4) +(0.65\*3) + (0.005\*5) +(0.01\*6) + (0.120\*2) = 3.09

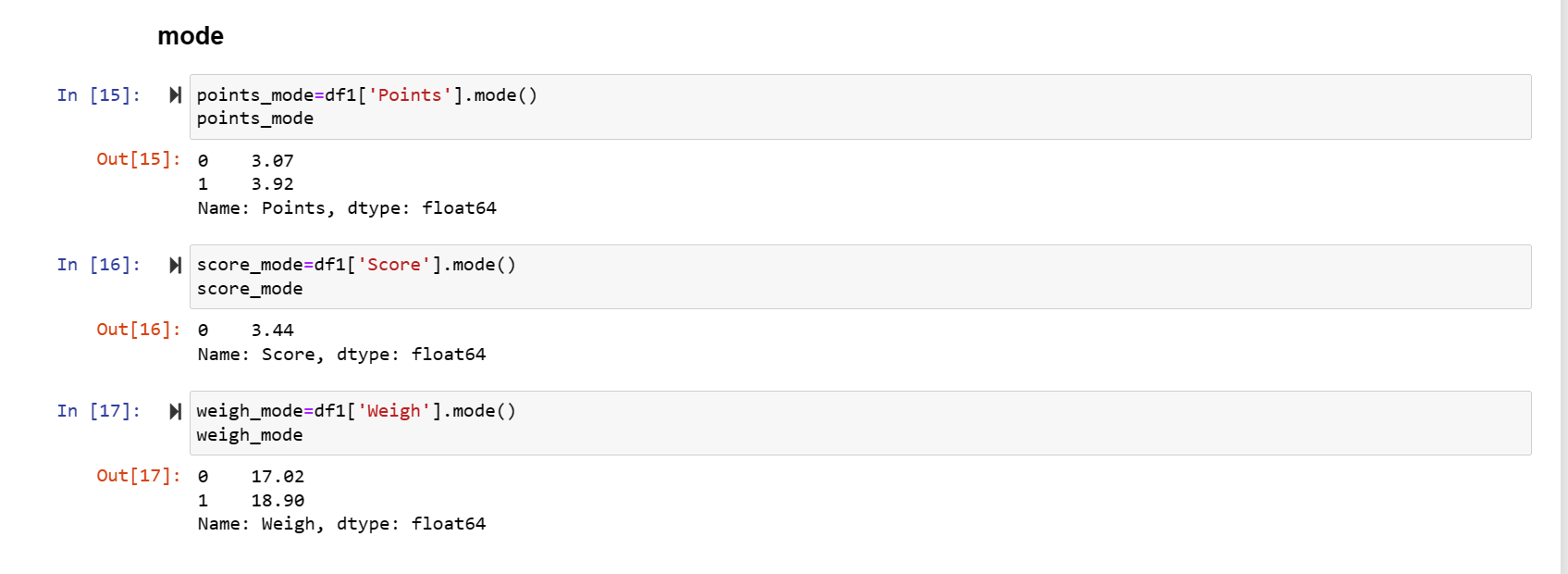
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points, Score, Weigh >

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

**Use Q7.csv file**



****

****

****

****

****

****

**Inference:**

**Points:**

\* The mean and median are close, which indicates symmetric distribution.

\* The mode includes 3.07 and 3.92, which indicates bimodal distribution.

\* The low variance and standard deviation indicates that the data points are close to the mean.

**Score:**

\* The mean and median are close, which indicates symmetric distribution.

\* The mode is 3.44, which indicates peak at this value.

\* The high variance and standard deviation indicate a wider spread of the data.

**Weigh:**

\* The mean and median are close, which indicates symmetric distribution.

\* The mode includes 17.02 and 18.90, which indicates bimodal distribution.

\* The low variance and standard deviation indicate that the data points are close to the mean.

**Range:**

\* Points have the moderate range with minimum at 2.76 and maximum at 4.93.

\* Score has a wider range from 1.51 to 5.42 which indicates variability.

\* weigh has a moderate range with the values ranging from 14.5 to 22.9

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

= 145.3333

So, the expected value of the weight of the randomly chosen patient is 145.333

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

**Cars speed and distance**

**Use Q9\_a.csv**

****

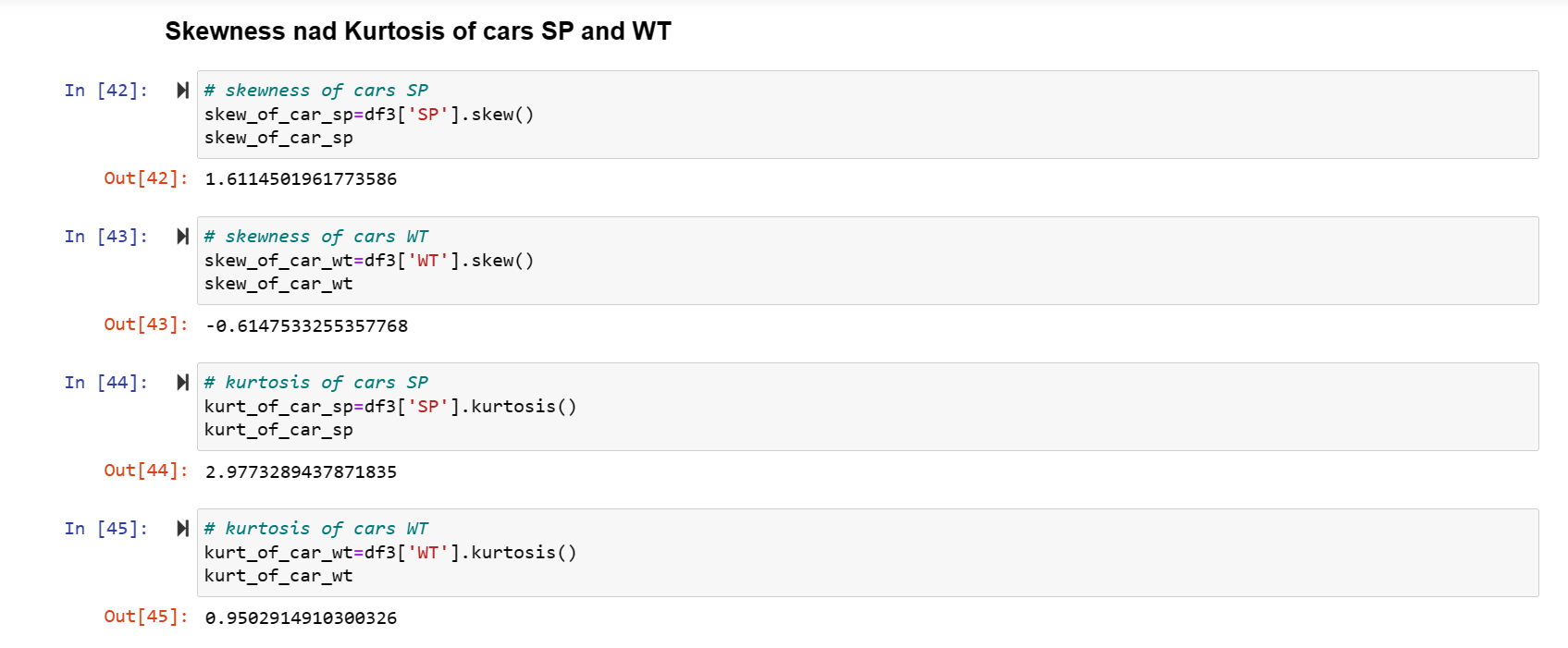
**Inference:**

**Cars speed:** The speed distribution is close to symmetric and the kurtosis indicates that it is less peaked than a normal distribution.

**Cars distance:** The distance distribution is positively skewed with right tail.

**SP and Weight (WT)**

**Use Q9\_b.csv**

****

**Inference:**

**Cars SP:** The SP distribution is positively skewed with right tail, and it has a higher kurtosis which indicates heavier tails and sharper peak.

**Cars WT:** The WT distribution is negatively skewed with left tail, and it has positive kurtosis which suggests heavier tails and sharper peak.

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

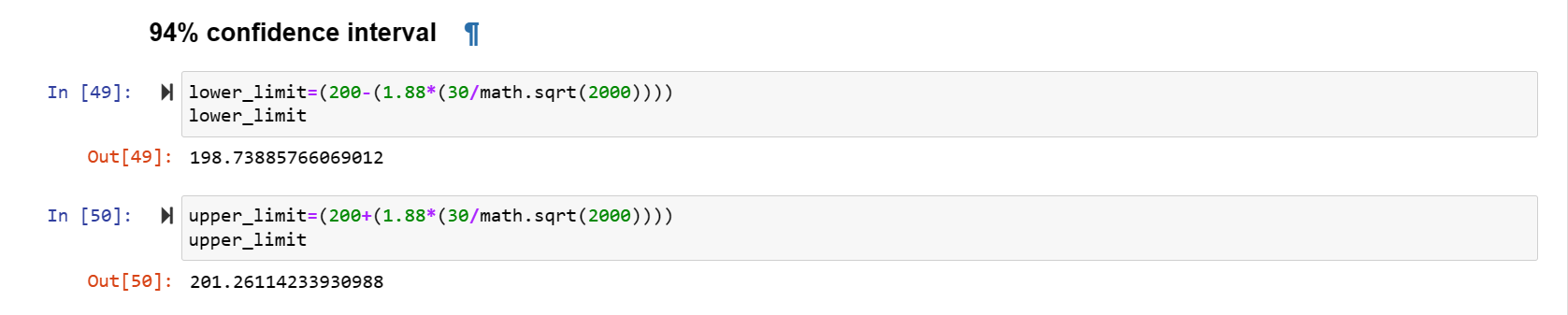


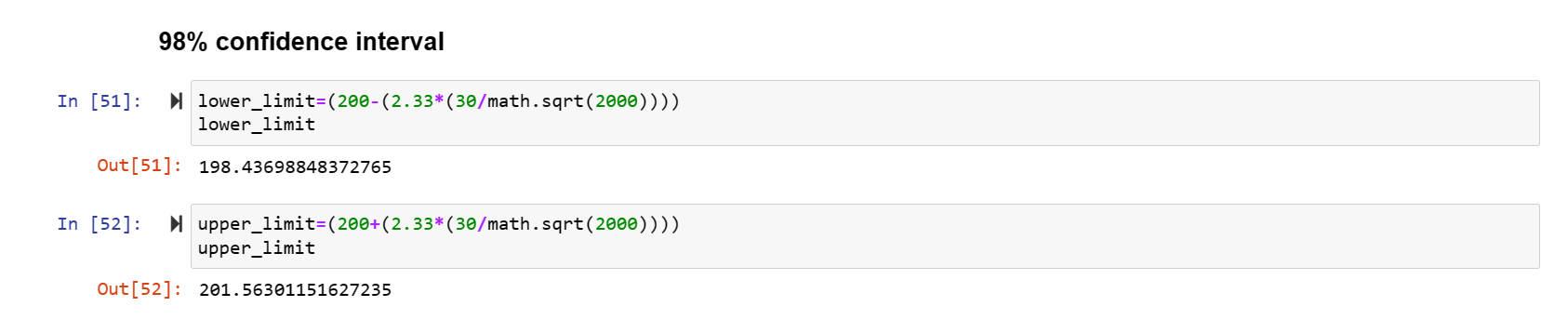
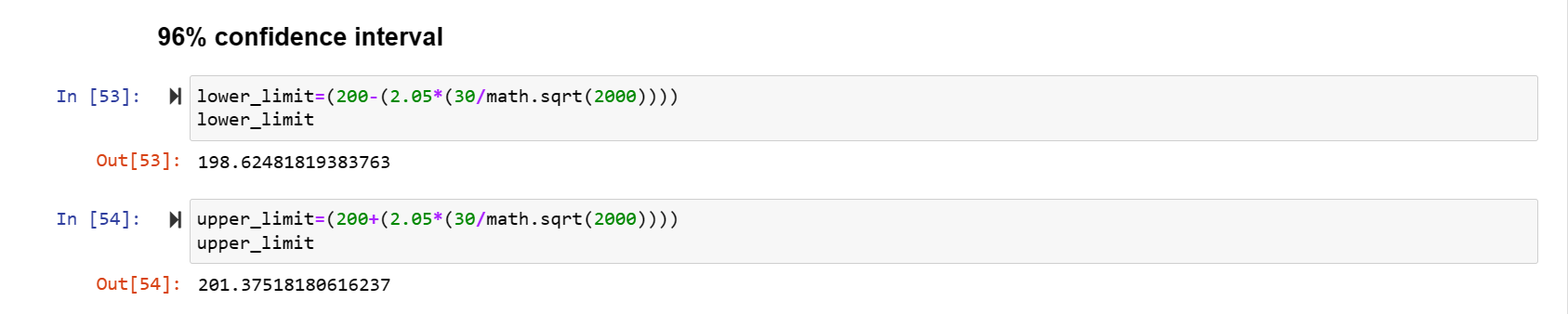
**Inference:** The frequency will be maximum when the checkpoint & weight ranges between 50 to 100, and the frequency will be minimum when the checkpoint & weight ranges between 350 to 400.



**Inference:** Median is near to the lower quartile (Q1), the boxplot is positively skewed because the intervals between the median and upper quartile (Q3) are maximum and outliers are present in the boxplot.

**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?

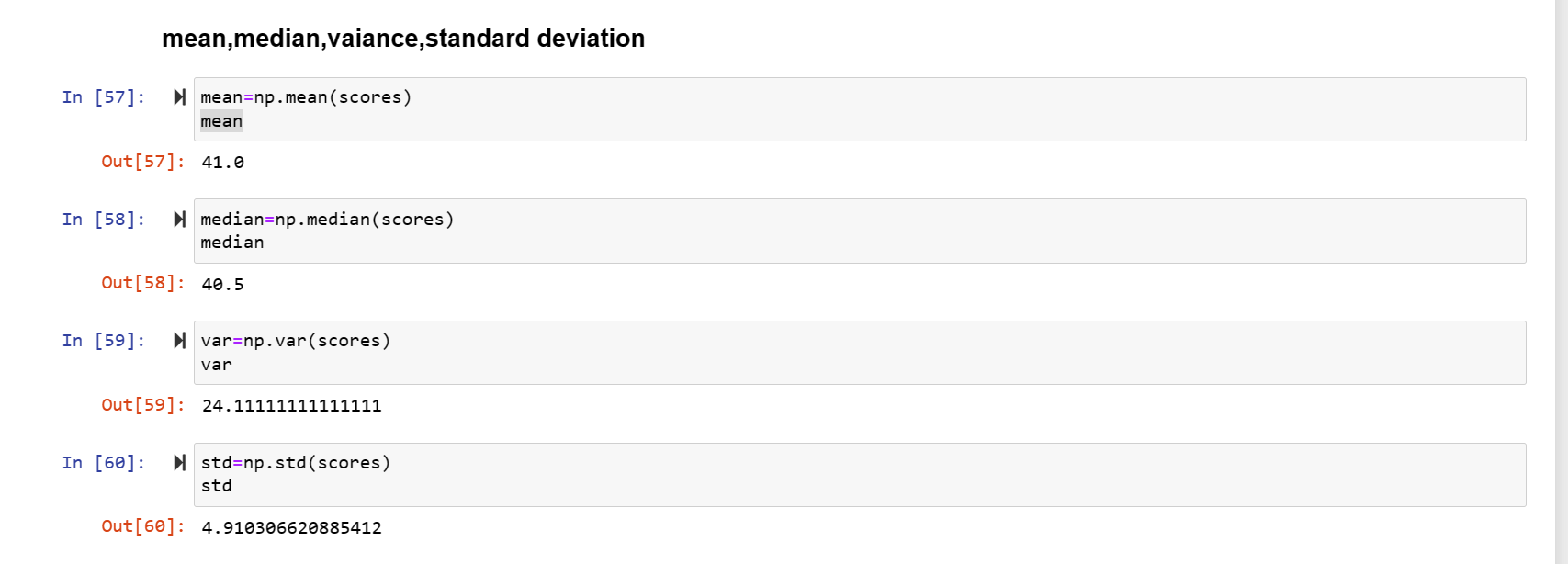


**Q12)** Below are the scores obtained by a student in tests

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?



**Answer:** The mean score is 41 which indicates the average performance, the median score is 40.5 which shows that half of the scores are below and half of them are above and there is some variability in the scores with some scores deviating from the mean.

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

**Answer:** Data is perfectly symmetrical and there is no skewness present.

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

**Answer:** The data is positively skewed and the tail on the right side is longer.

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

**Answer:** The data is negatively skewed and the tail on the left side is longer.

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

**Answer:** The data has heavy tails and it has more extreme values than the normal distribution.

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

**Answer:** The data has less tails and it has less extreme values than the normal distribution.

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



What can we say about the distribution of the data?

**Answer:** Median is 15.2(approx.), the interval between median and Q1 is greater than the interval between median and Q3.

What is nature of skewness of the data?

**Answer:** The distribution is negatively skewed.

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

**Answer:** IQR range is about 8.

Q19) Comment on the below Boxplot visualizations?



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

**Inference:**

**1)** The IQR range is 25 (approx.), the interval between median and Q1 is slightly greater than the interval between median and Q3 so the graph is positively skewed and no outliers are present.

**2)** The IQR range is 87.5 (approx.), the interval between median and Q1 is equal to the interval between the median and Q3 so the graph is symmetric and no outliers are present.

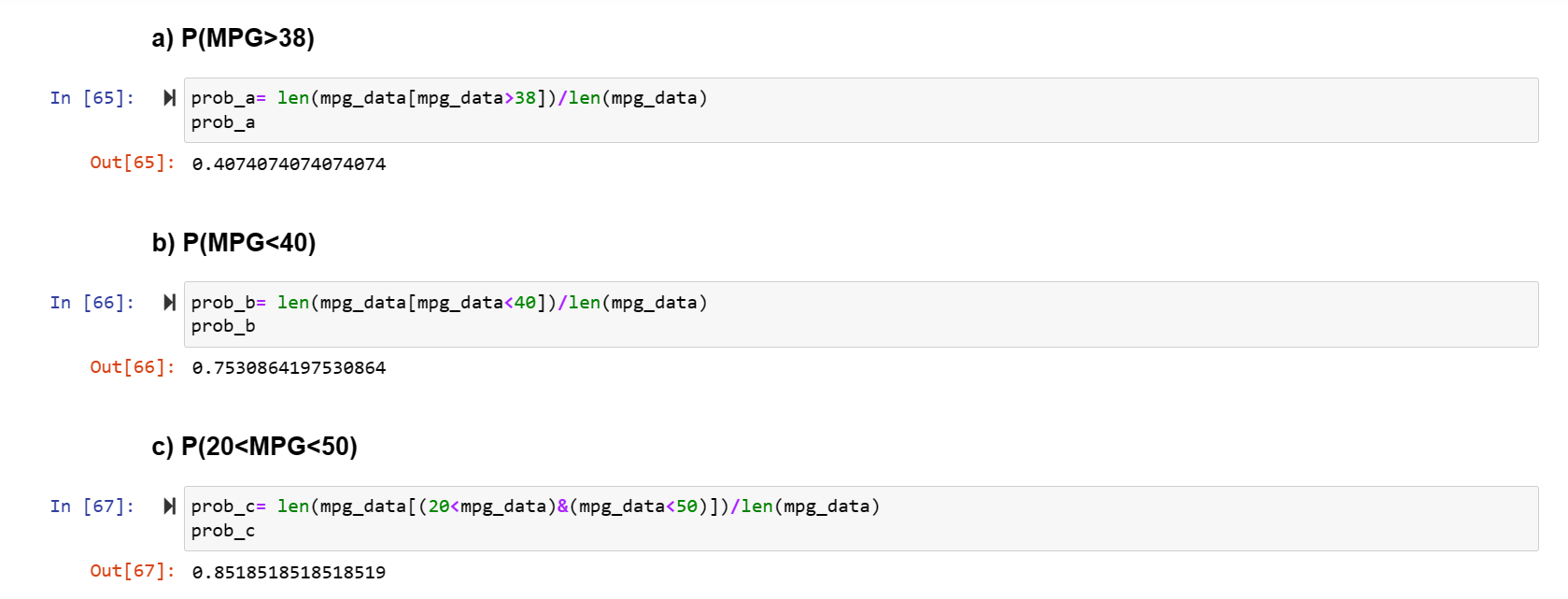
Q20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

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

MPG <- Cars$MPG

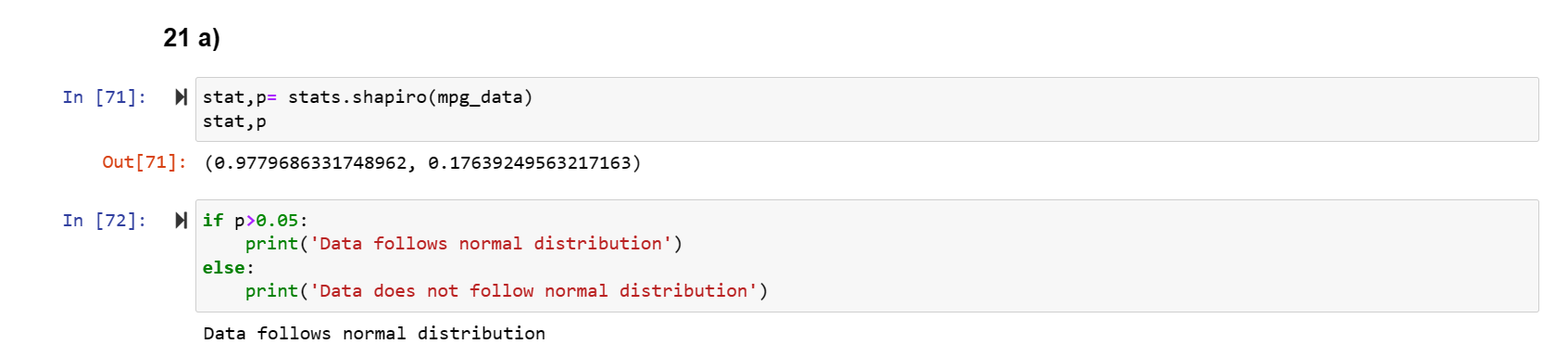
a. P(MPG>38) b. P(MPG<40) c. P (20<MPG<50)



Q21) Check whether the data follows normal distribution

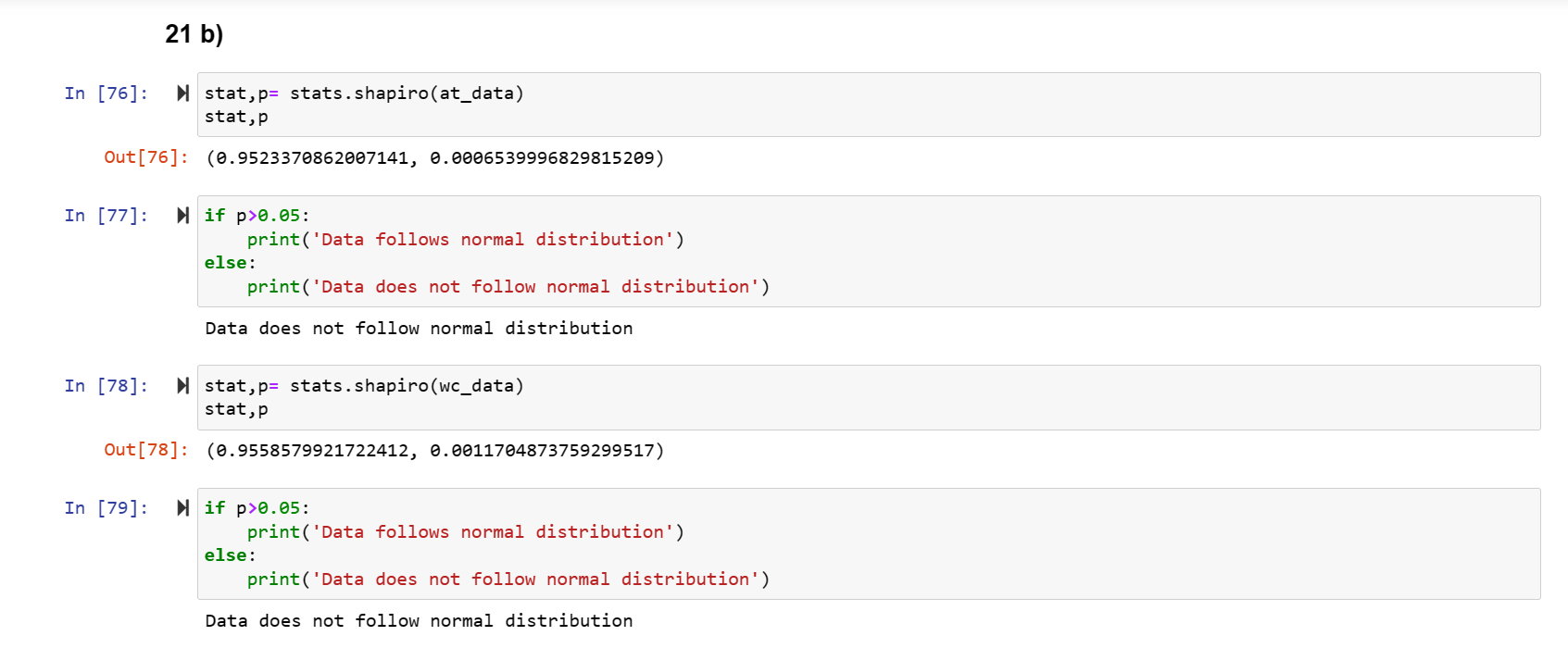
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv



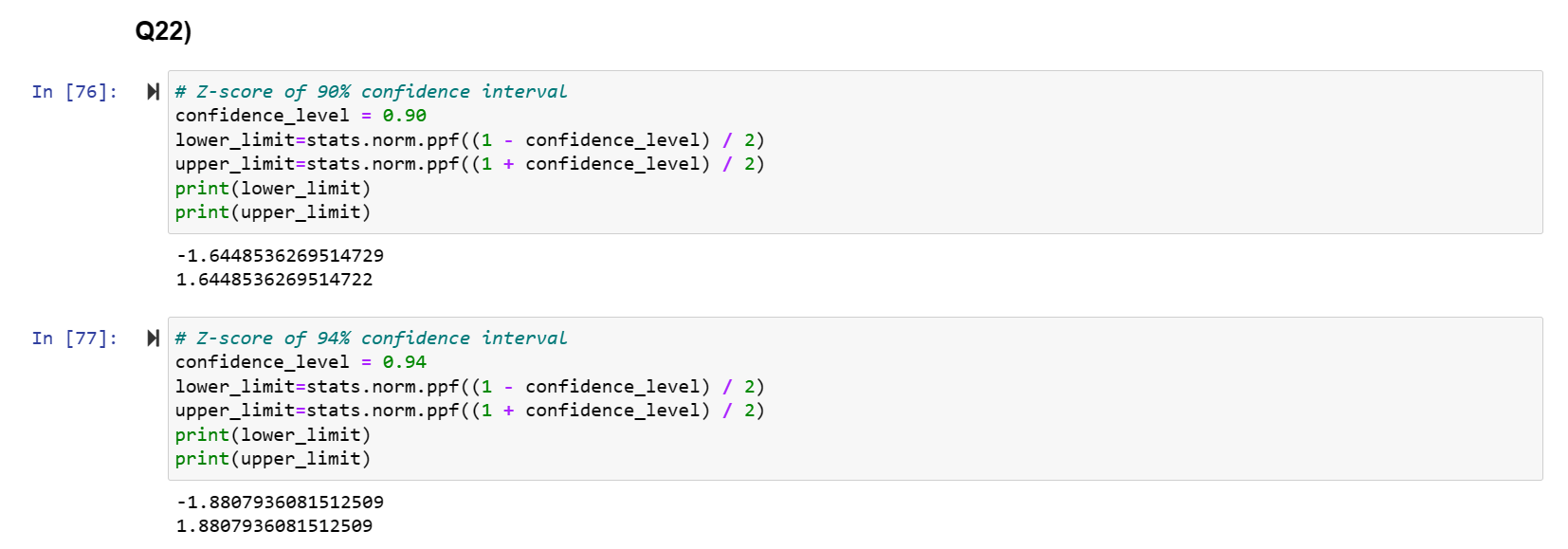
1. Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution

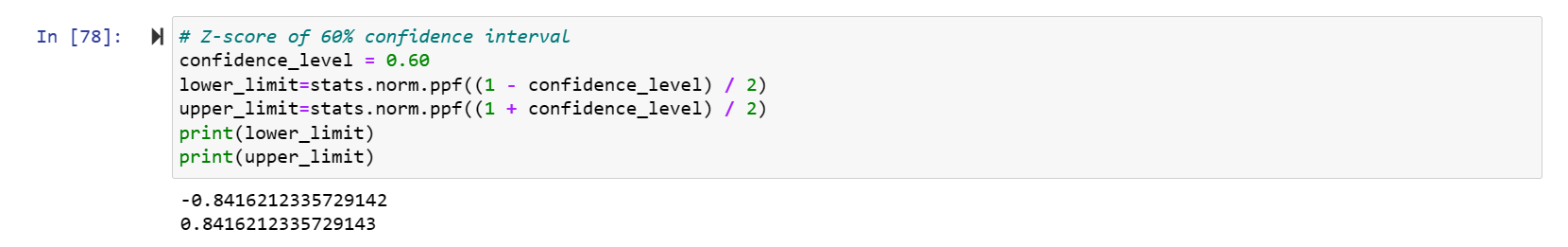
Dataset: wc-at.csv



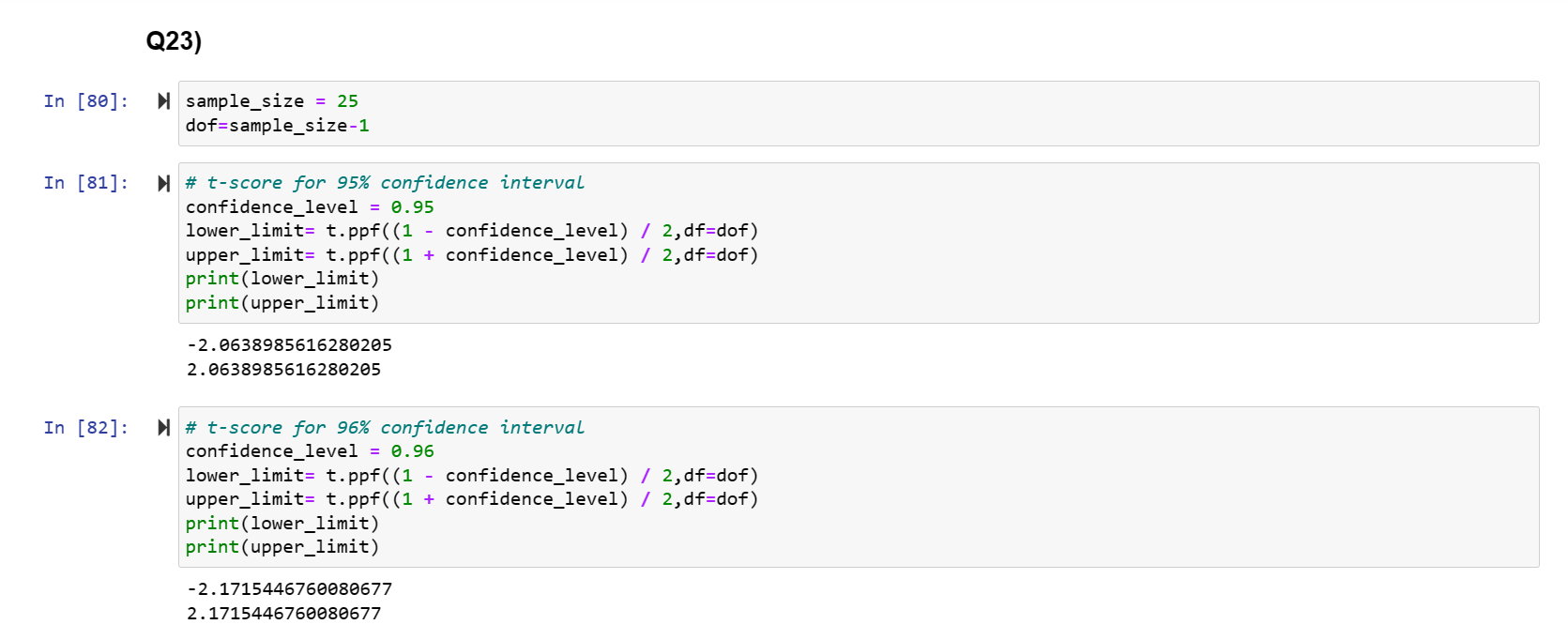
Q22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

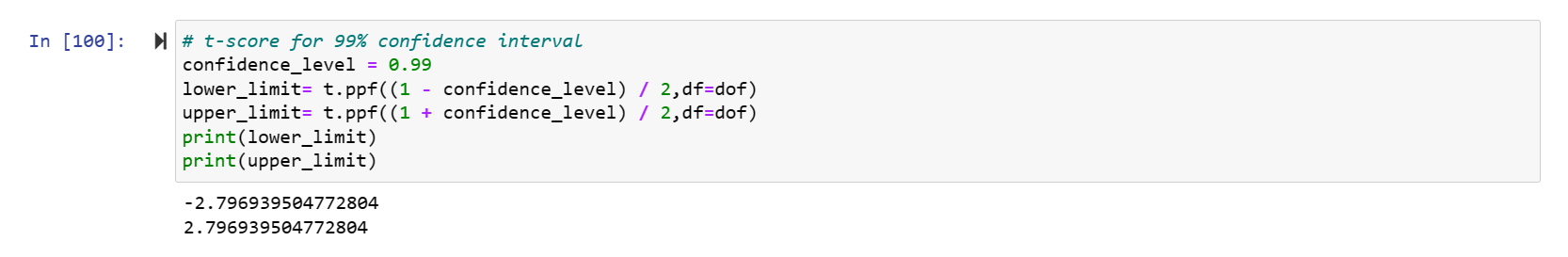
**Answer:**





Q23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25





Q24**)** 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

****