|  |  |
| --- | --- |
| 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 data |
| High School Class Ranking | Ordinal data |
| Celsius Temperature | Ratio data |
| Weight | Ratio data |
| Hair Color | Nominal data |
| Socioeconomic Status | Nominal data |
| Fahrenheit Temperature | Ratio data |
| Height | Ratio data |
| Type of living accommodation | Nominal data |
| Level of Agreement | Ordinal data |
| IQ(Intelligence Scale) | Ordinal data |
| Sales Figures | Ratio data |
| Blood Group | Nominal data |
| Time Of Day | Interval data |
| Time on a Clock with Hands | Interval data |
| Number of Children | Interval data |
| Religious Preference | Nominal data |
| Barometer Pressure | Ratio data |
| SAT Scores | Interval data |
| Years of Education | Ratio data |

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

Ans:- The probability of getting that two heads and one tail is 3/8

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

1. Equal to 1

🡪Zero

1. Less than or equal to 4

🡪6/36

1. Sum is divisible by 2 and 3

🡪6/36

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:- 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

Ans:- 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**

Ans:- **Point Score Weight**

Mean- 3.59 3.21 17.84

Median- 3.69 3.25 17.710

Mode- 3.89 3.54 17.43

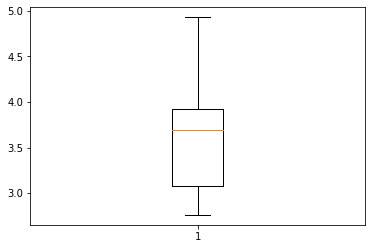
Variance- 0.28 0.95 3.19

Standard Deviation- 0.53 0.97 1.78

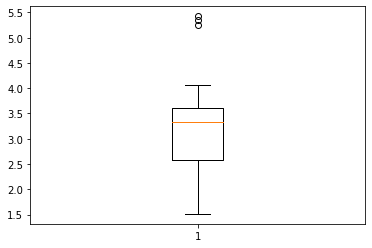
Max, Min- 4.93, 2.76 5.42, 1.51 22.9, 14.5

Range- 2.17 3.91 8.39

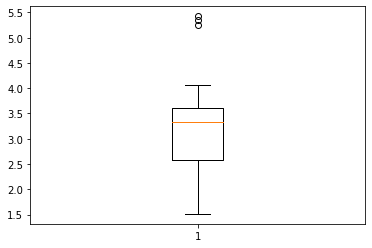
Boxplot of Points



Box plot of Score



Box plot of Weigh



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:- 145.33

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

**Cars speed and distance**

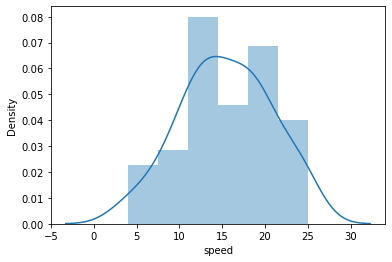
**Use Q9\_a.csv**

Ans**:- Cars speed Distance**

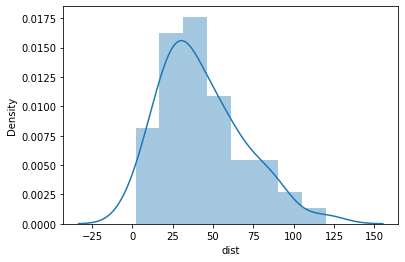
Skewness -0.1175 0.8068

Kurtosis -0.5089 0.4050

Cars speed inferences



Distance inferences



**SP and Weight(WT)**

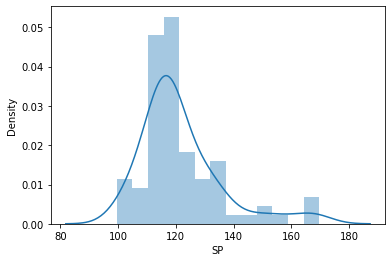
**Use Q9\_b.csv**

**Ans:- SP WT**

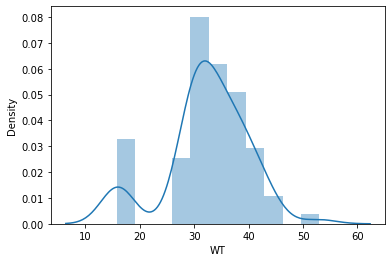
**Skewness**  1.611 -0.614

**Kurtosis**  2.977 0.950

**SP inferences**

****

**WT inferences**

****

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



Ans:- The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.

Ans:- The boxplot has outliers on the maximum side.

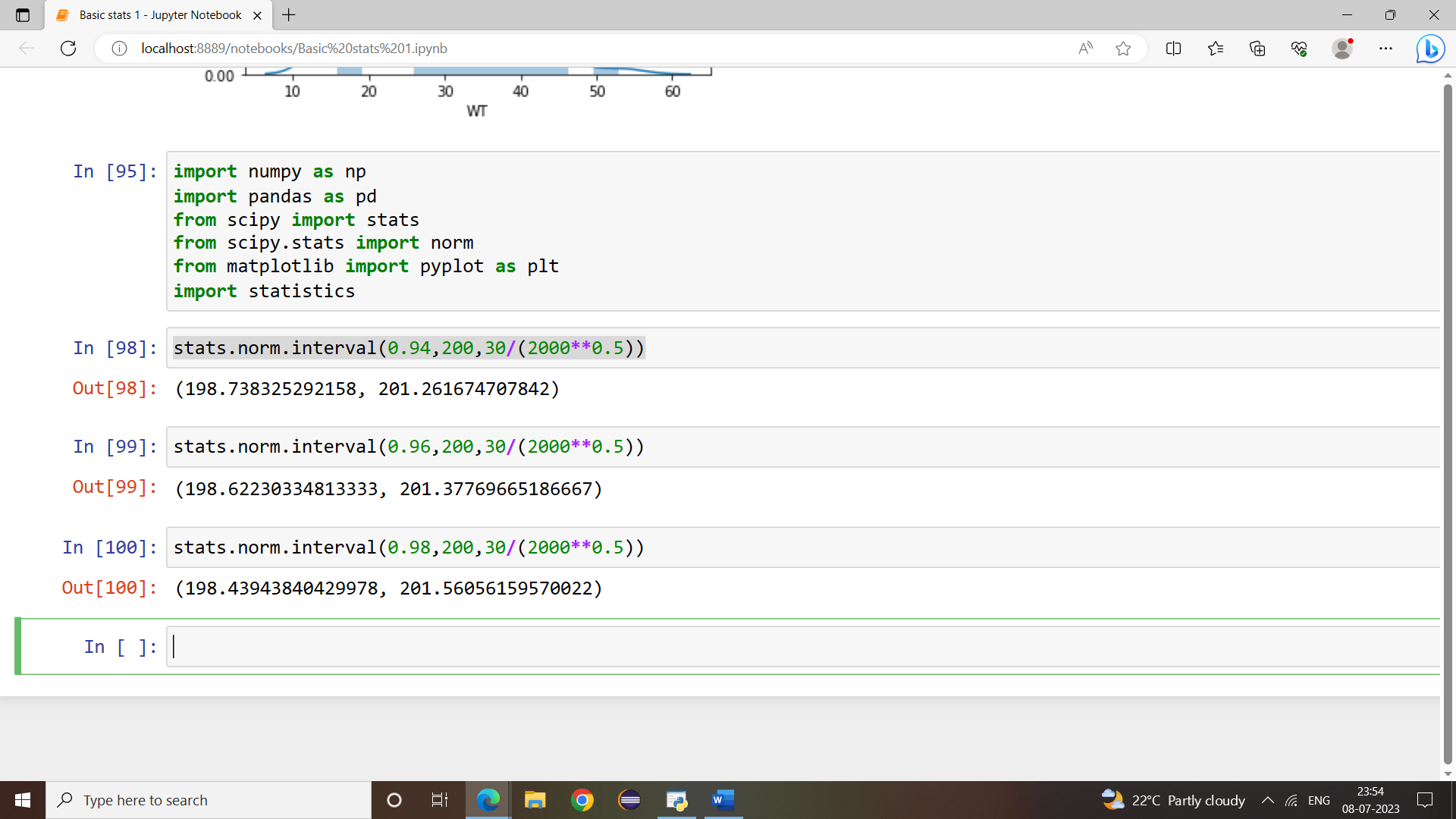
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

Ans:- Range Z value

CI 94% 198.7383 201.2616

CI 96% 198.6223 201.37776

CI 98% 198.4394 201.5605



**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?
3. Ans:-

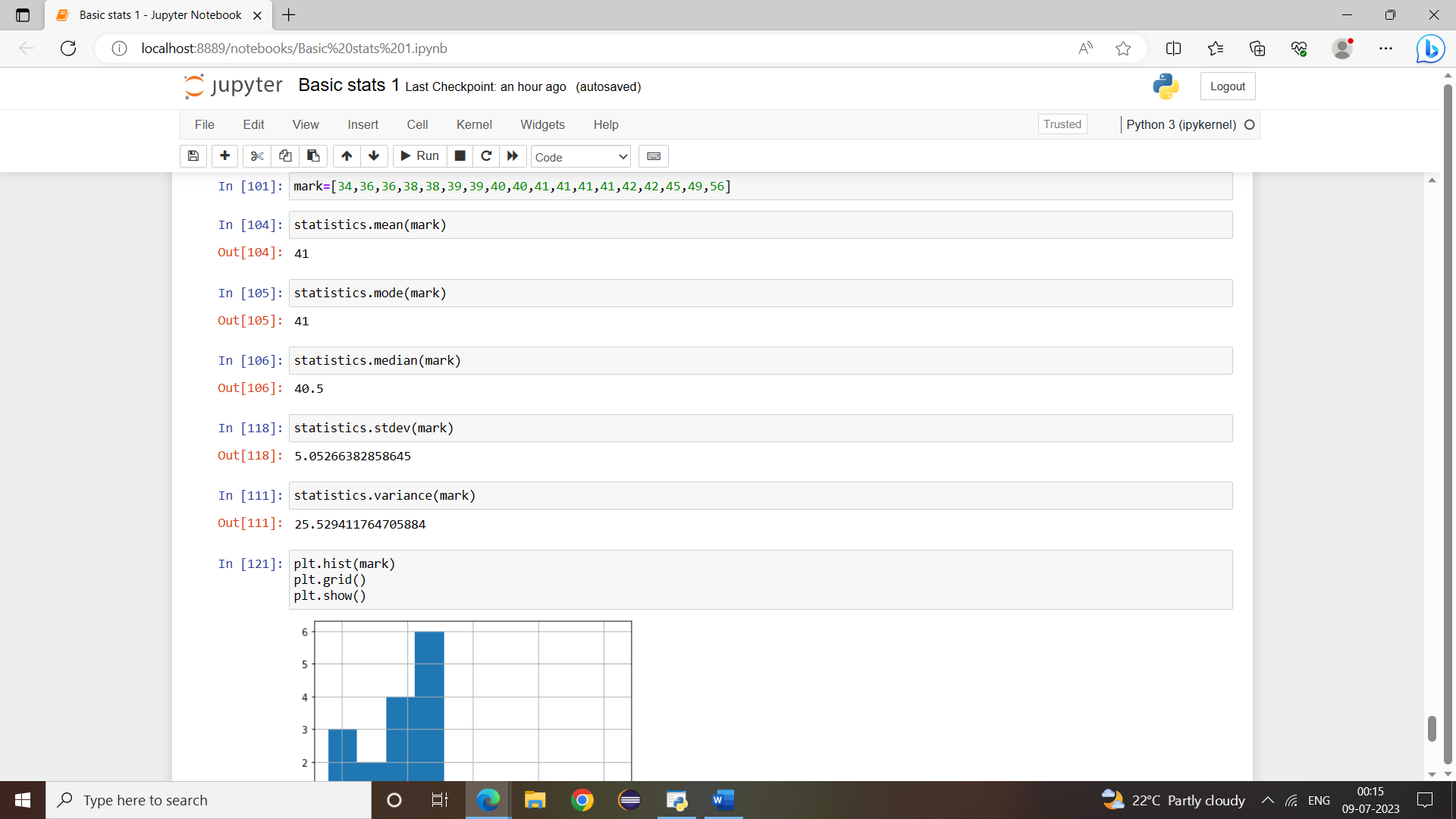
Mean:41

Median:40.5

Mode:41

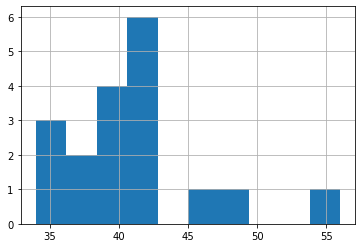
Standard deviation:5.0526

Variance:25.5294

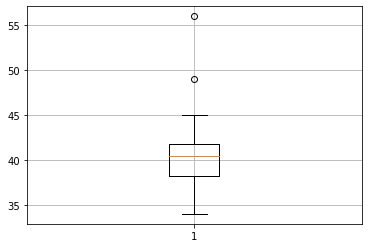


1. Ans:- Mass of students marks is b/w 41-42

There are two outlier 49,56



Skewness(1.52) is positive because mass of marks in left side of the plot.



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

Ans:- Data is normalized and there is no skewness.

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

Ans:- Negative skewness implies mass of the distribution concentrated on right side.

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

Ans:- Positive skewness implies mass of the distribution concentrated on left side.

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

Ans:- Positive kurtosis value indicates that thinned peak and wider tails.

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

Ans:- Negative kurtosis value indicates that wider peak and thinner tails.

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



What can we say about the distribution of the data?

Ans:- Not normally distributed

What is nature of skewness of the data?

Ans:-Negative skewness

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

Ans:- Approximately 10-18  
  
  
Q19) Comment on the below Boxplot visualizations?



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

Ans:- First there are no outliers. Second both the box plot shares the same median that is approximately in arrange between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

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)

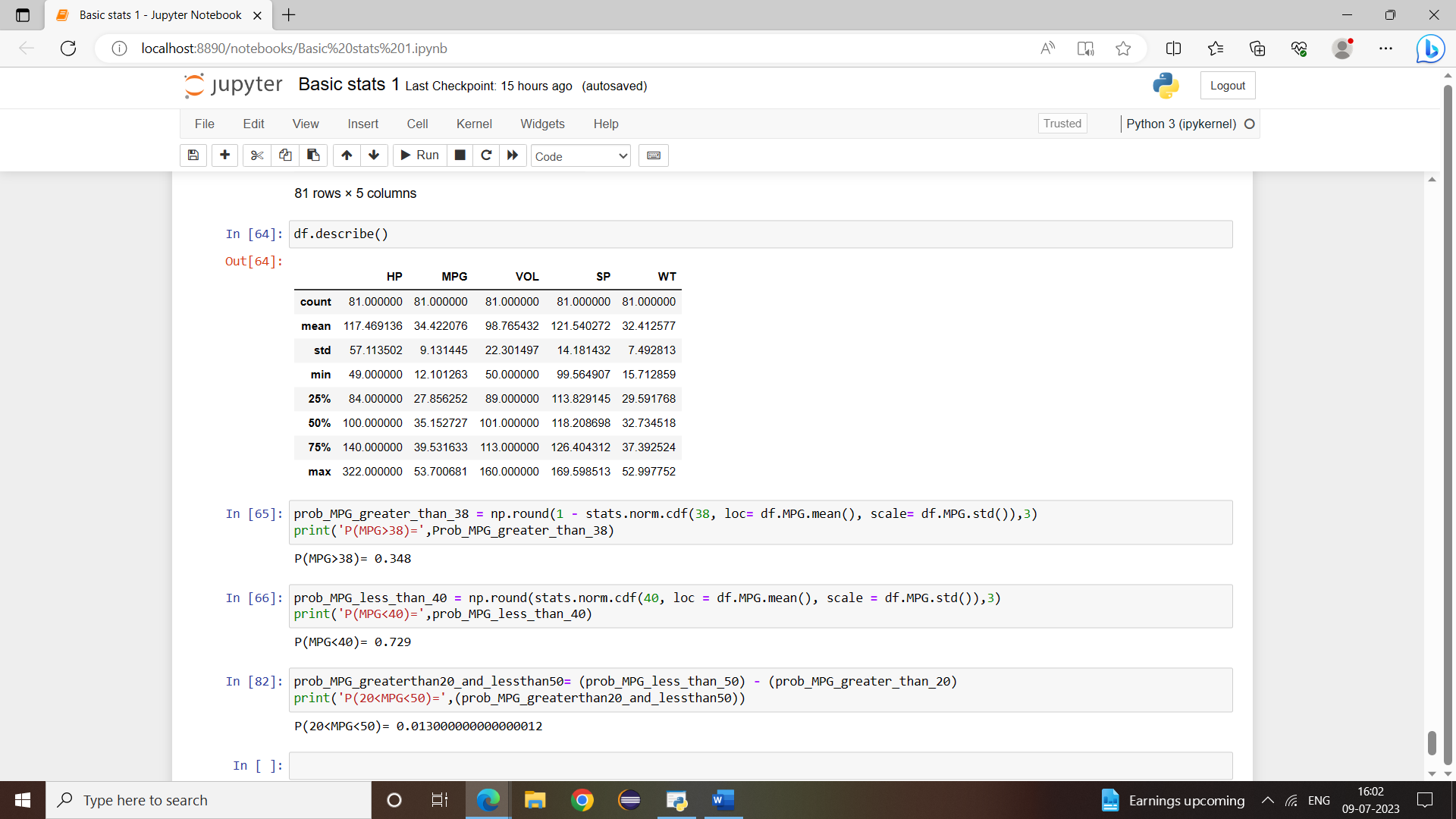
Ans:- 1-pnorm(38,34,422,9.13144)=0.3475908

* 1. P(MPG<40)

Ans:- pnorm(40,34,422,9.13144)=0.7293527

* 1. P (20<MPG<50)

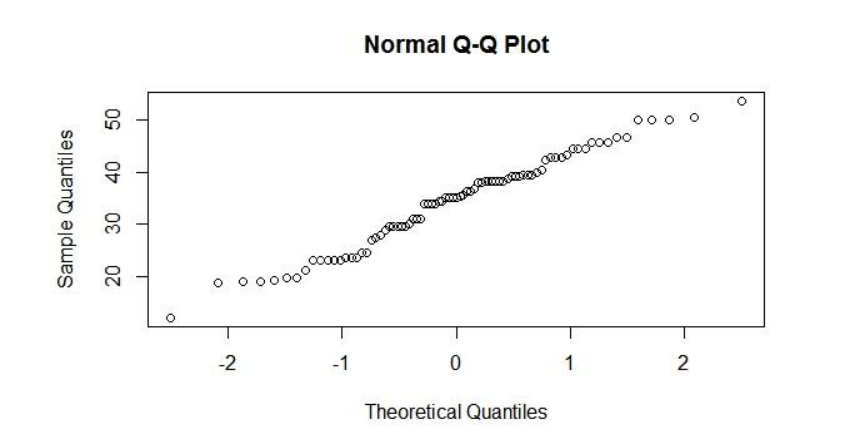
Ans:-pnorm(50,34.422,9.13144)-(1-pnormm(20,34.422,9.13144)=0.01311818



Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

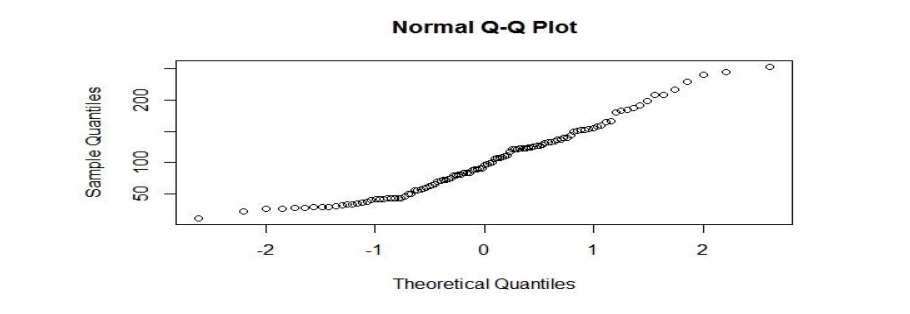
Dataset: Cars.csv



Distributed normally

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

Dataset: wc-at.csv



Does not follow normal distribution

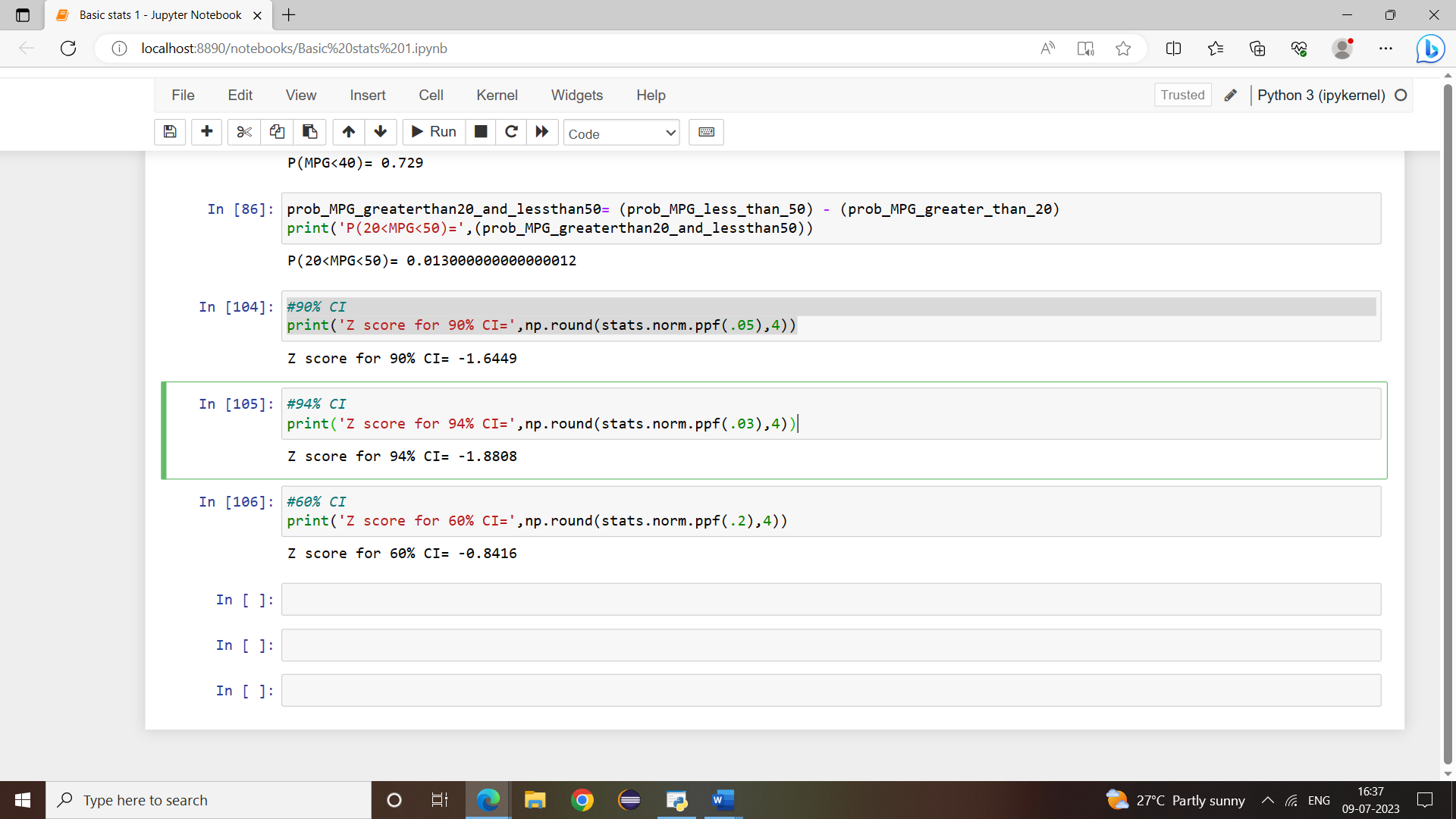
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

Ans:- CI Z score

90% -1.6449

94% -1.8808

60% -0.8416



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

Ans:- CI T score

99% 2.79694

96% 2.17154

95% 2.06389

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

