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

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

**Ans** – 3/8 = 0.375

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

1. Equal to 1 ----- Zero
2. Less than or equal to 4 -----1/6
3. Sum is divisible by 2 and 3 ----- 5/36

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.

**Ans**

Points Score Weigh

Mean - 3.596 3.217 17.848

Median - 3.695 3.325 17.71

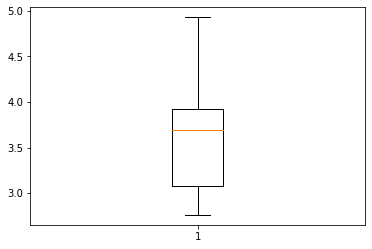
Mode - 3.891 3.54 17.43

Standard Deviation - 0.534 0.978 1.786

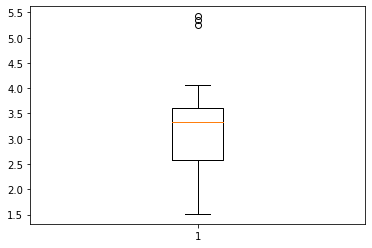
Variance - 0.285 0.957 3.19

Min, Max - 2.76 , 4.93 1.513,5.424 14.5,22.9

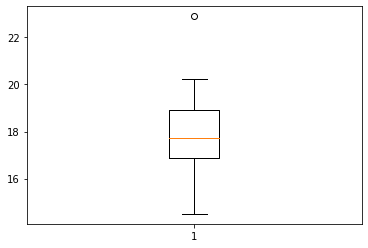
Range - 2.17 3.911 8.399



Box plot of Points



Box Plot of Score



Box Plot of Weigh

**Use Q7.csv file**

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**

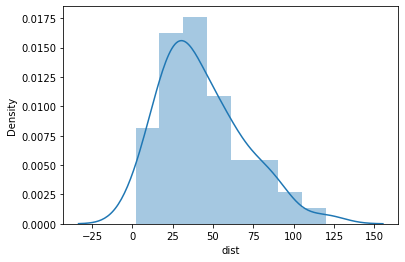
**Cars speed and distance**

**Use Q9\_a.csv**

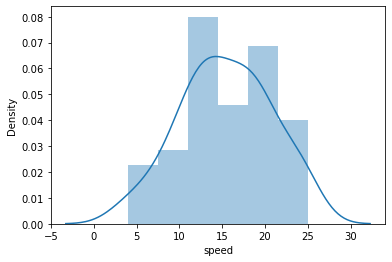
Car Speed Distance

Skewness -0.11 3 0.782

Kurtosis -0.508 0.405



Distance Skew



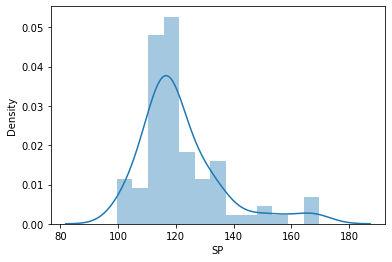
Car speed Skewness

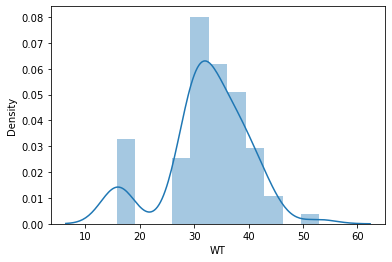
**SP and Weight(WT)**

SP WT

Skewness - 1.5814 -0.6033

Kurtosis - 2.9773 0.9502





Use Q9\_b.csv

**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?

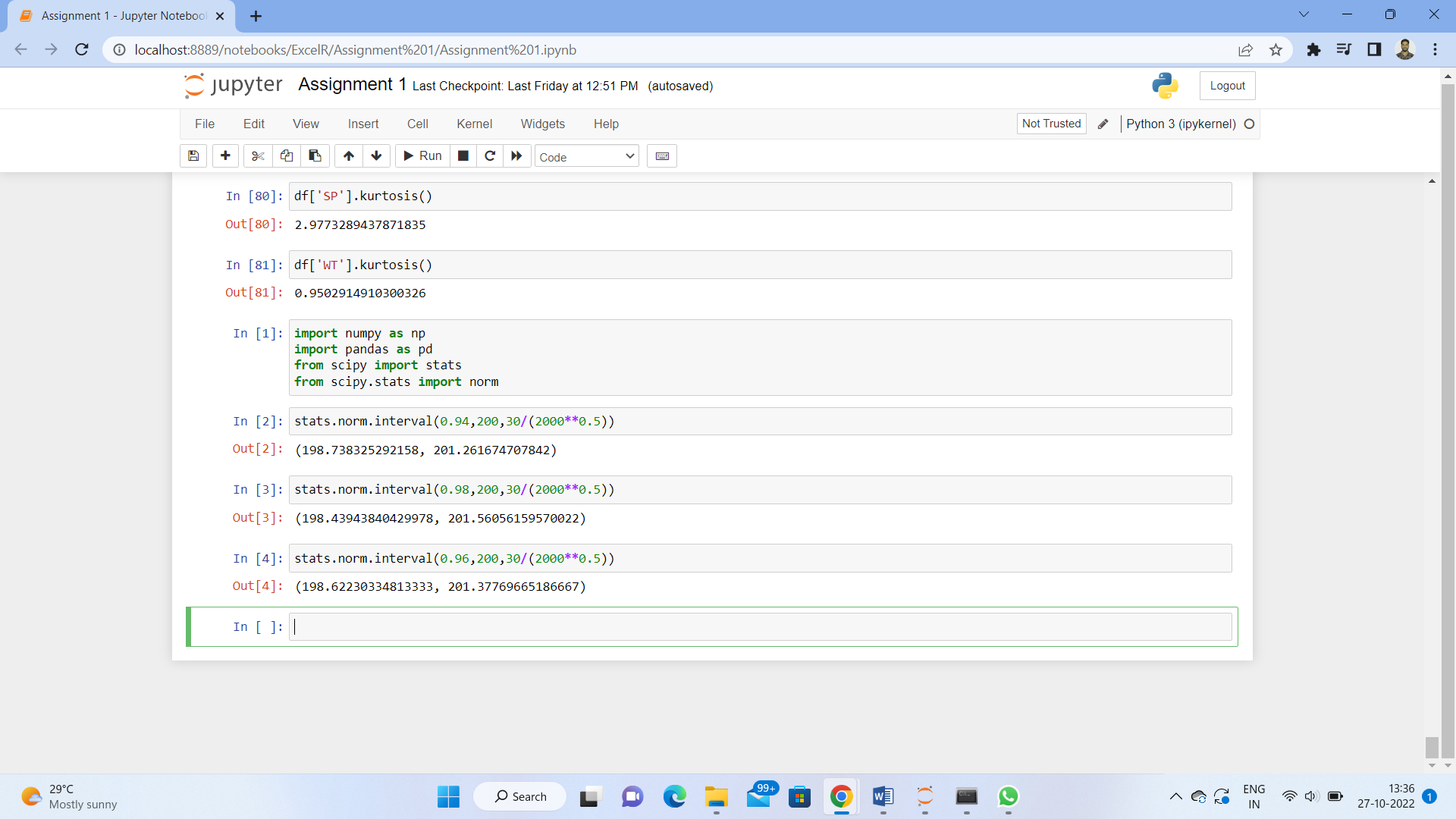
**Solution** =

Confidence Interval Z value Range

Confidence interval94% 201.2616 198.74,201.26

Confidence interval96% 201.5605 198.62,201.38

Confidence interval98% 201.3776 198.43,201.56



**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?

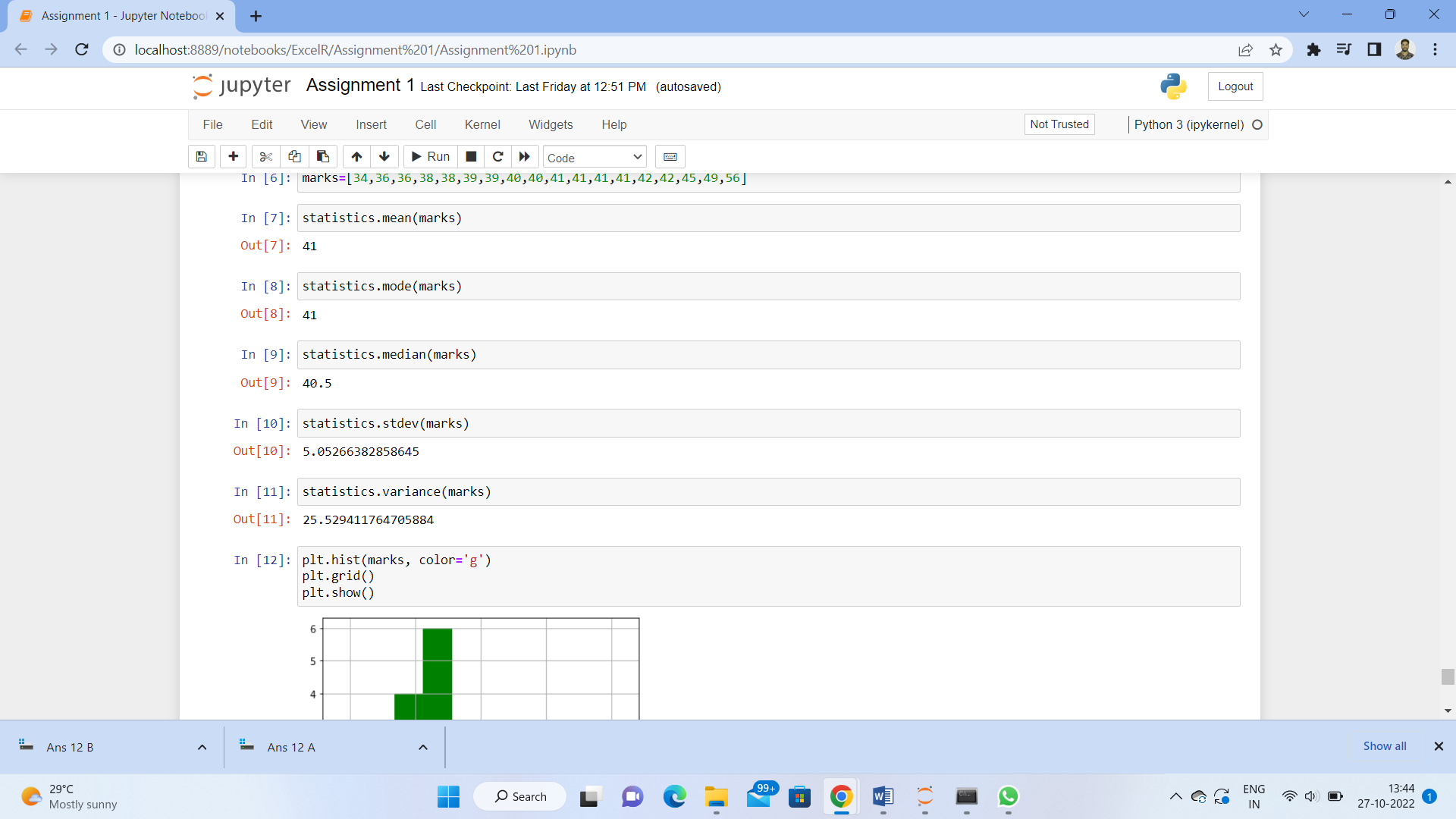
**Solution**=

Mean 41

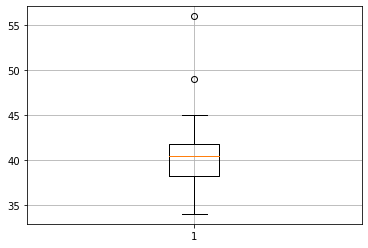
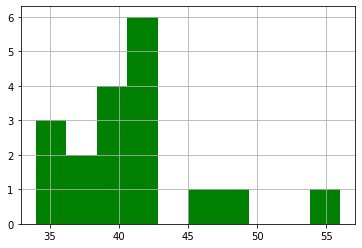
Median 40.5

Variance 25.52

Standard Deviation 5.05664

2)Mass of students marks is between 41-42. There are two outlier 49,56

Skewness(1.52) is positive because mass of marks in left side of 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 thr 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?

-Not normally distributed

What is nature of skewness of the data?

-Negative Skewness

What will be the IQR of the data (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 a range 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)

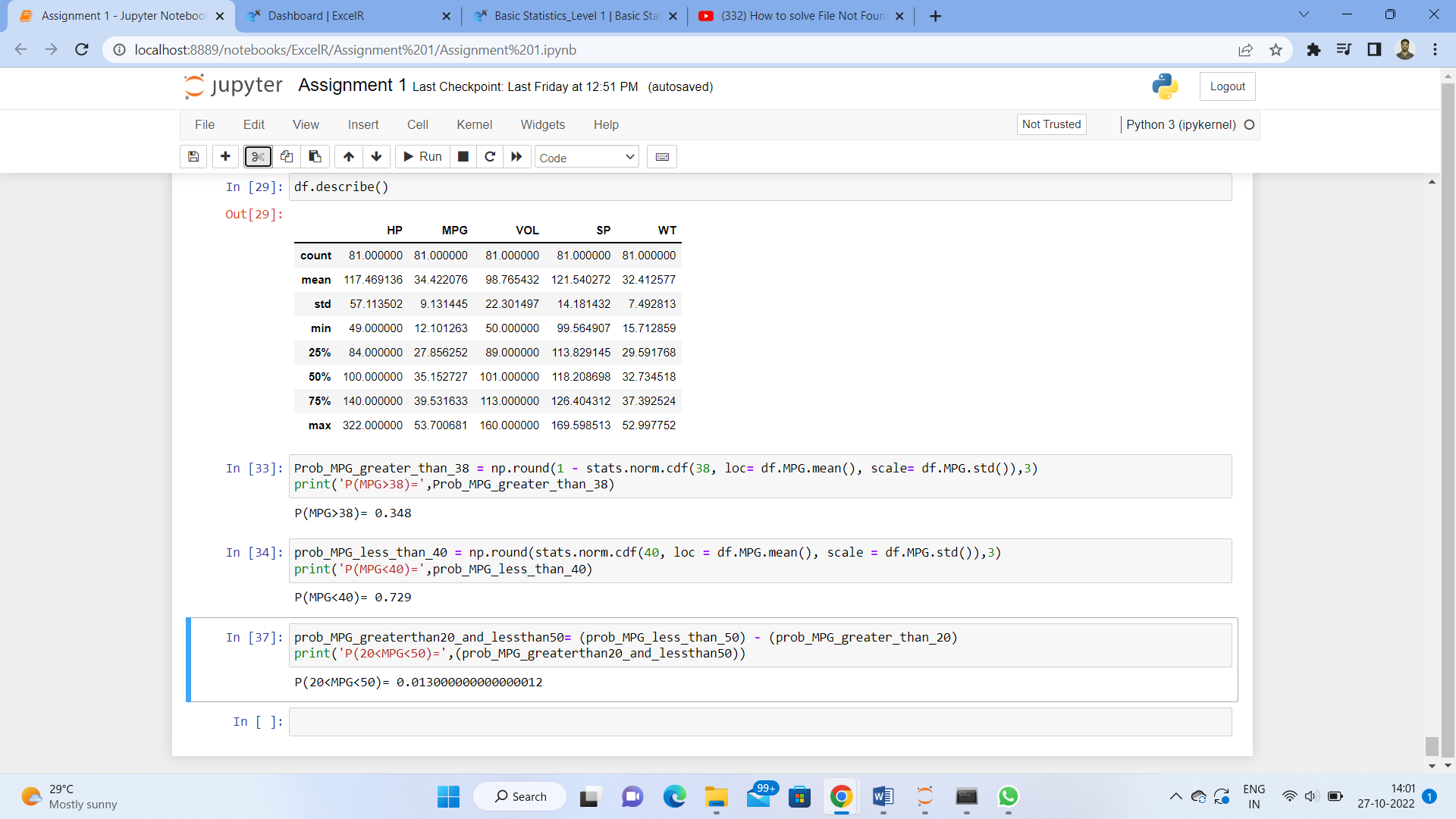
1-pnorm(38,34.422,9.13144) = 0.3475908

* 1. P(MPG<40)

Pnorm(40,34.422,9.13144) = 0.7293527

* 1. P (20<MPG<50)

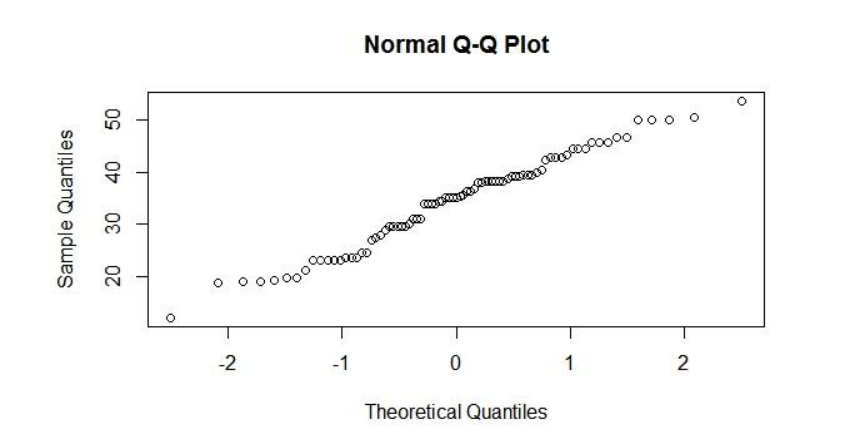
Pnorm(50,34.422,9.13144) – (1-pnormm(20,34.422,9.13144)) = 0.01311818



Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

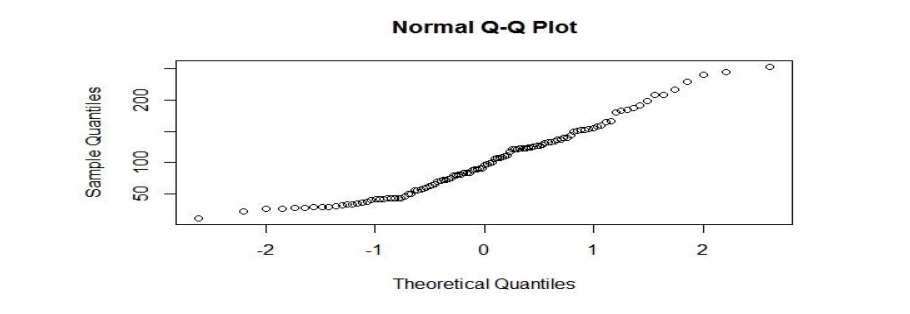
Dataset: Cars.csv



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

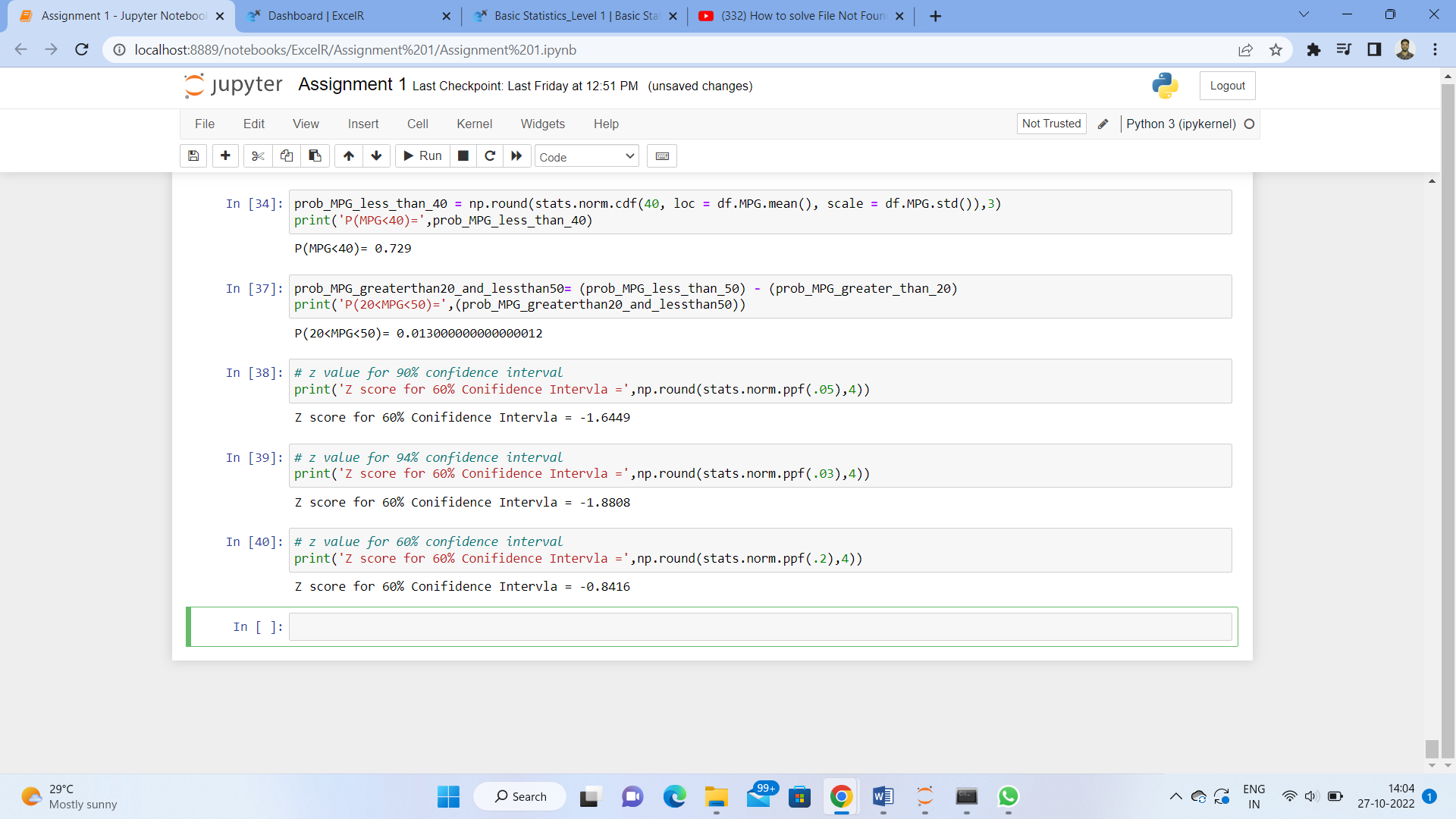
**Solution-**

Confidence Interval Z Scores

60% -0.8416212

90% -1.6448544

94% -1.880794



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

**Solution-**

Confidence interval T scores

95% 2.063899

96% 2.171545

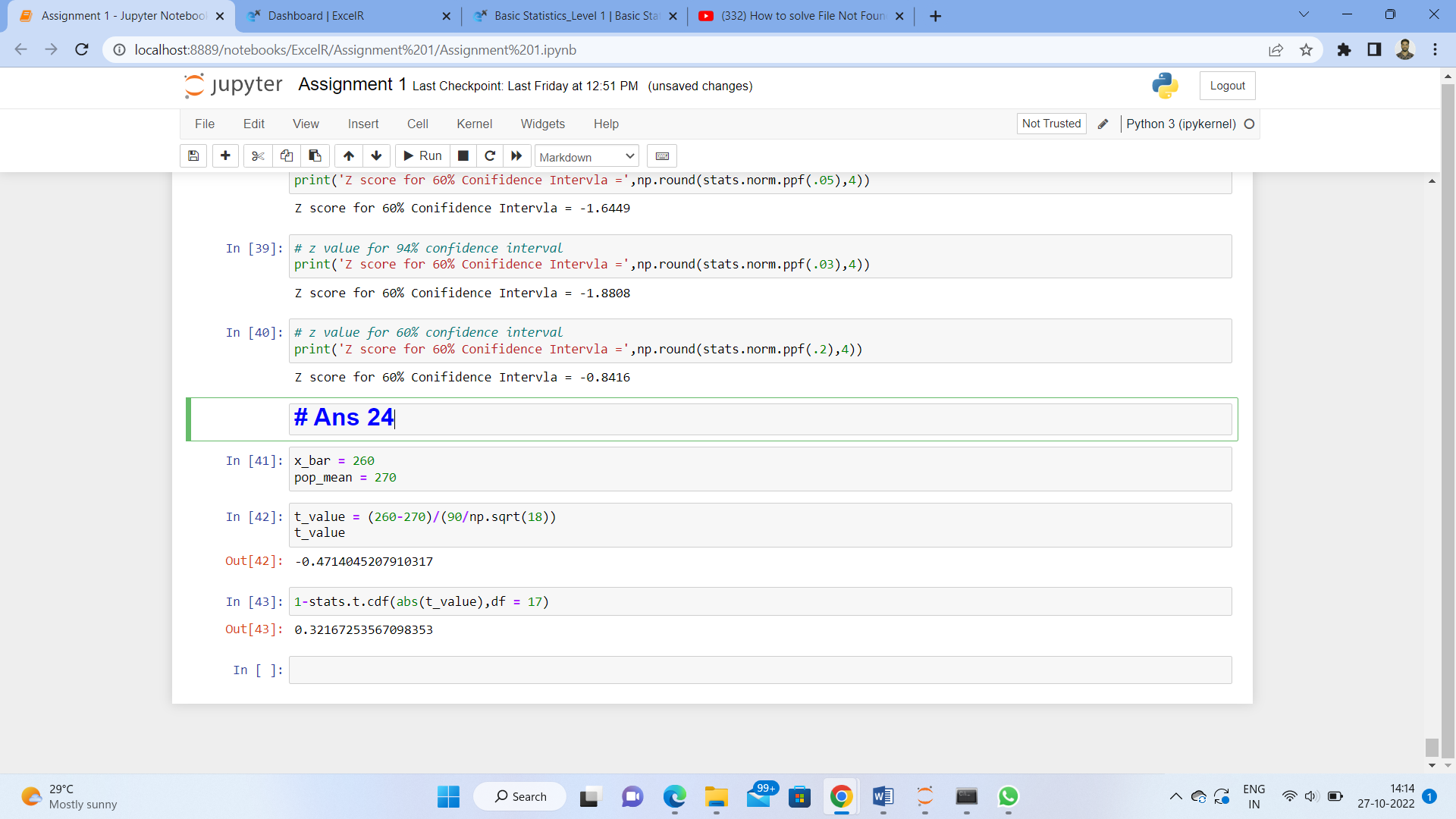
99% 2.79694

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

****