|  |  |
| --- | --- |
| 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 | Nominal |
| 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 | 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?

Answer:

H=heads

T=tails

Possible events with equal probability:

HHH, HHT, HTH , HTT, THH, THT, TTH, TTT

Number with 2 heads: 3

Total number: 8

From the definition of probability, the number you are looking for is 3/8=0.375=37.5%

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:

1. The sum is equal to 1 is zero because they starts with (1,1)….likewise. other than in the dice we are not having zero.
2. The sum is equal to 4 the possible outcomes are (1,3), (2,2), (3,1) therefore n(b)=3/36=1/12.
3. When two dice are rolled, sample space is given as:

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

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

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

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

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

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

Therefore,

Total number of possible outcomes=36

Favorable outcomes=sum is divisible by 2 and 3

Sum should be divisible by both 2 and 3

Favorable outcomes=(1,5), (3,3),(4,2), (5,1), (6,6)

Therefore,

Number of favorable outcomes=5

Probability=number of favorable outcomes/number of possible outcomes

Probability=5/36

Thus the probability that sum is divisible by 2 and 3 id 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?

Answer:

Total number of balls = (2+3+2) = 7

Let S be the sample space

Then, n(S) = Number of ways of drawing 2 balls out of 7

n(S)=7C2

n(S)=(7\*6)/(2\*1)

n(S)=21

Let E=Event of 2 balls, none of which is blue

n(E)=Number of ways of drawing 2 balls out of (2+3) balls

n(E)=5C2

n(E)=(5\*4)/(2\*1)

n(E)=10

P(E)=n(E)/n(S)=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 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

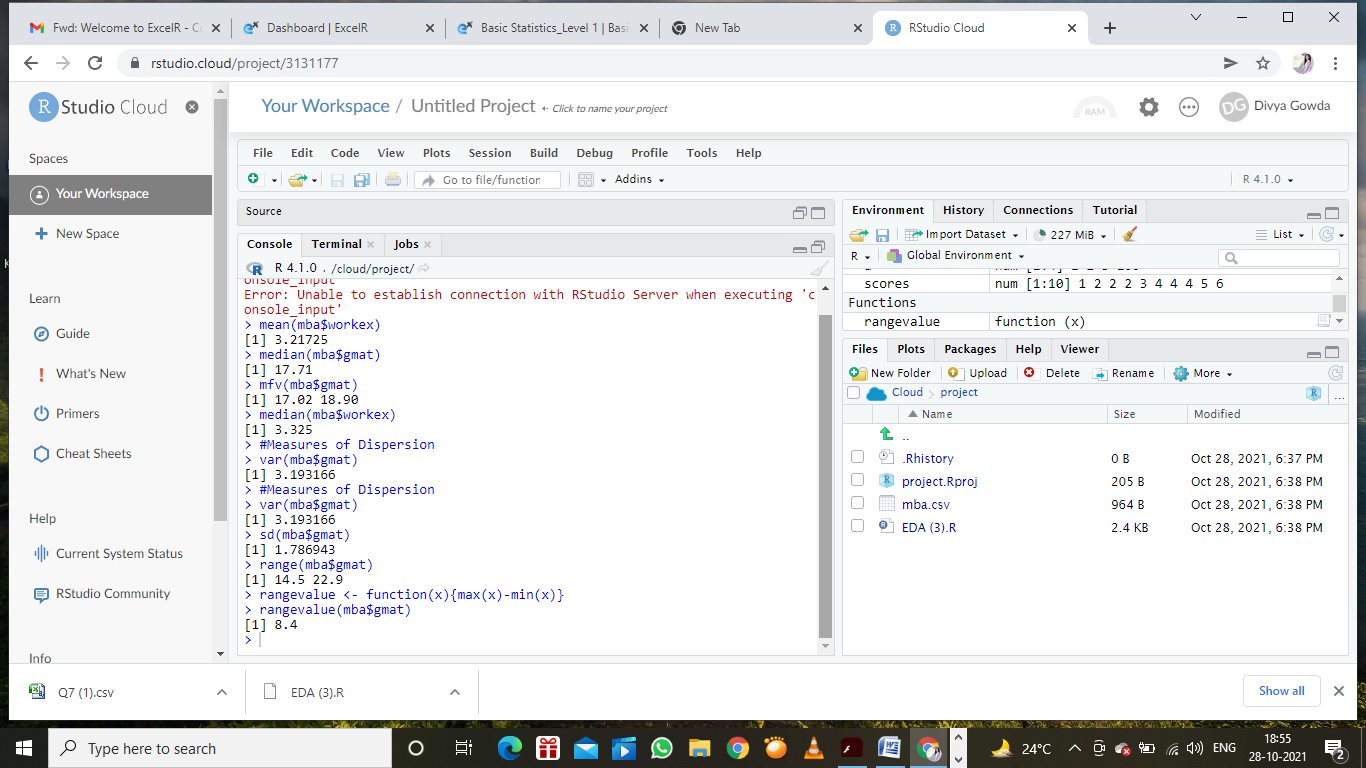
=3.09

Expected number of candies for a randomly selected child = 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**

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:

The weights(X) of patients at a clinic(in pounds), are 108,110,123,134,135,145,167,187,199

One of the patients is chosen at random.

To find expected value

Expected value=sum(probability\*value)

Sump(x).E(x)

There are 9 patients

Probability of selecting each patient=1/9

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

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

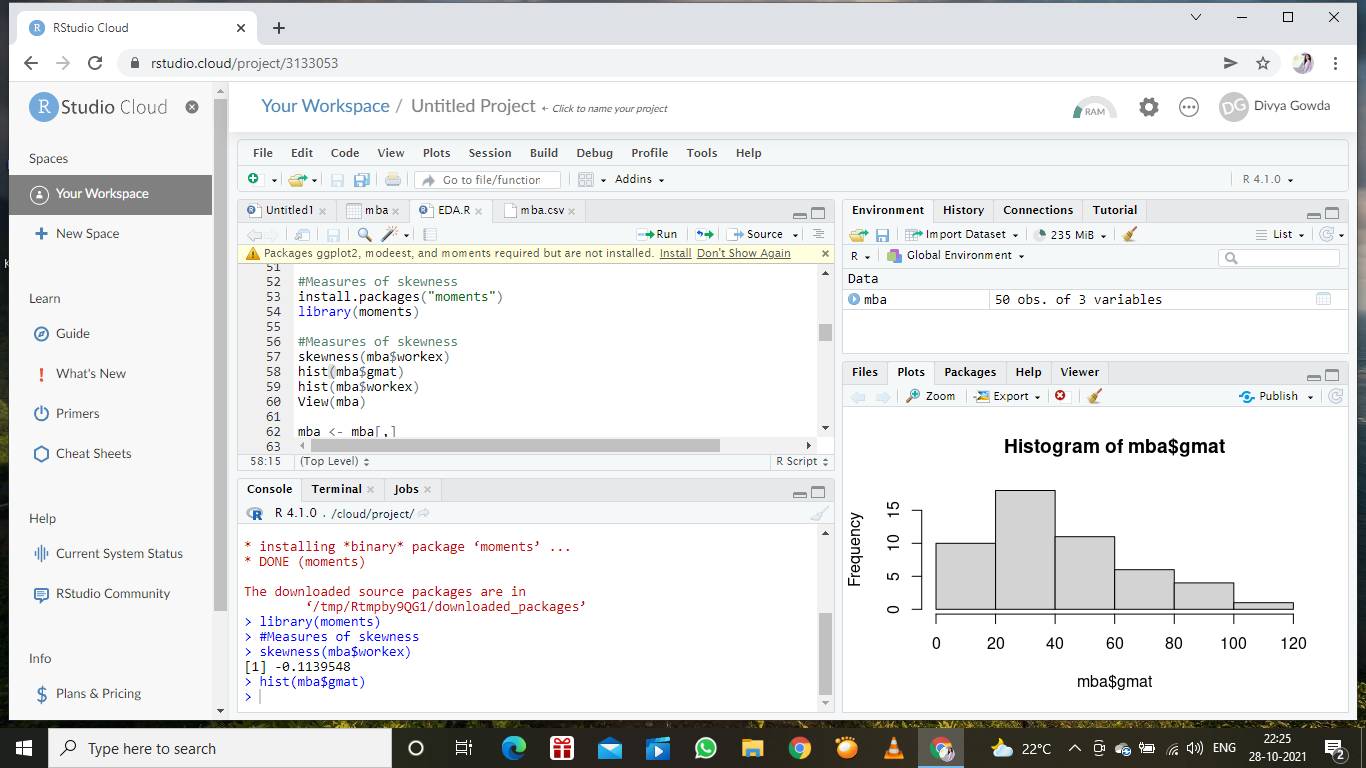
=(1/9)(108+110+123+134+135+145+167+187+199)

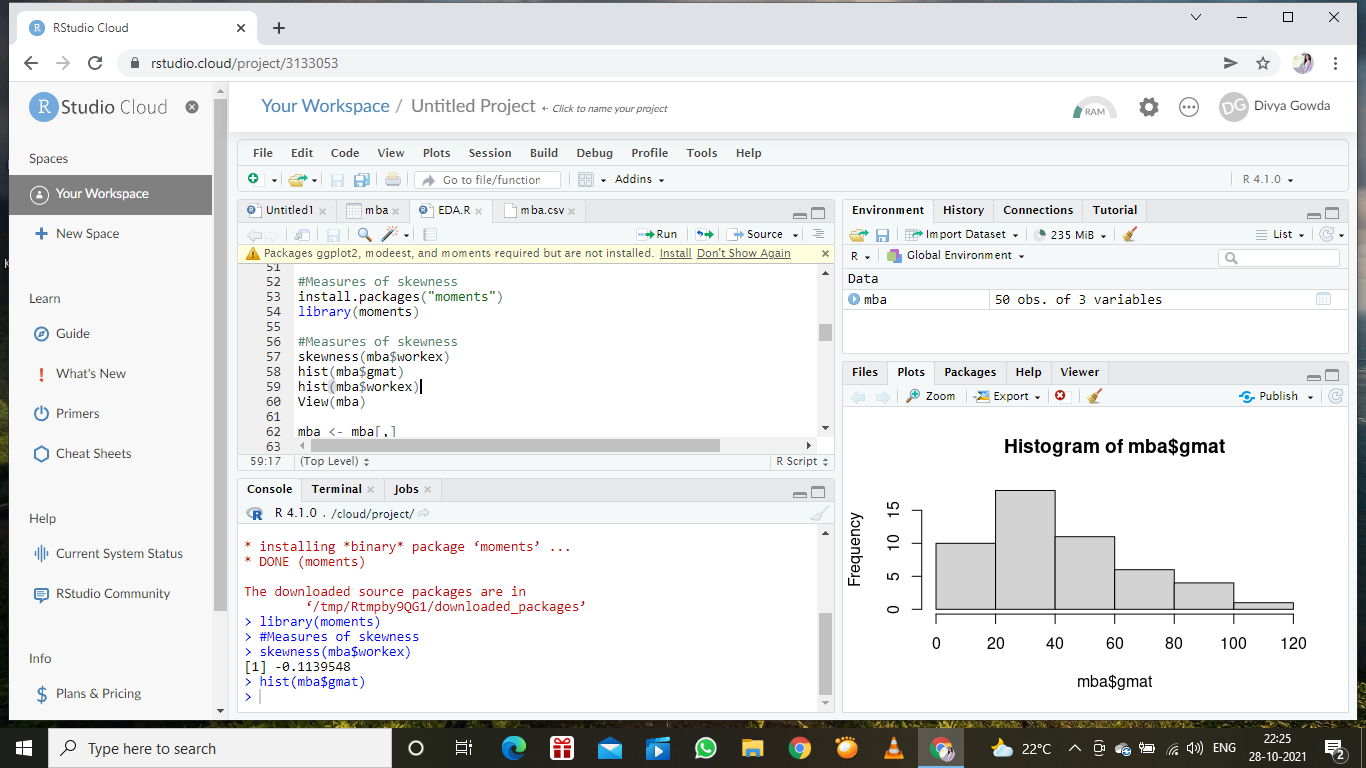
=(1/9)(1308)

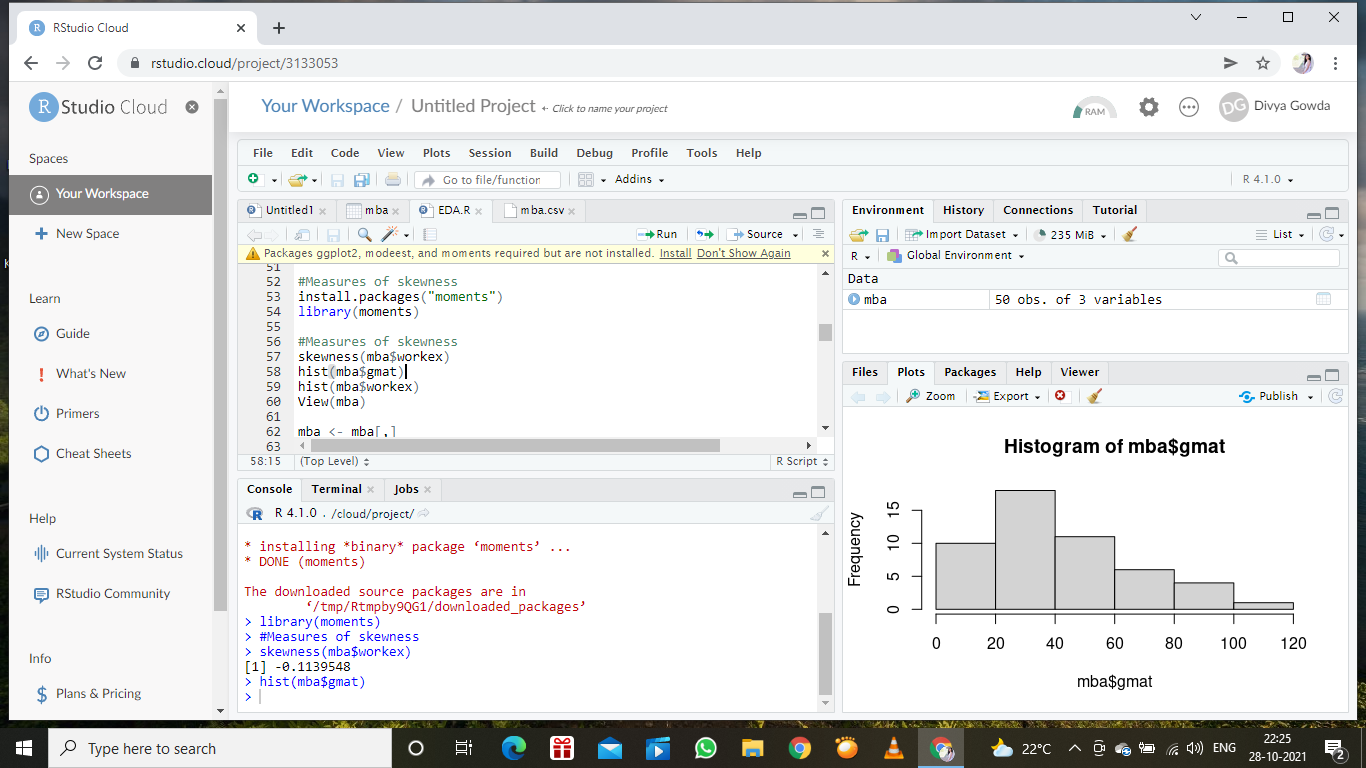
=145.33

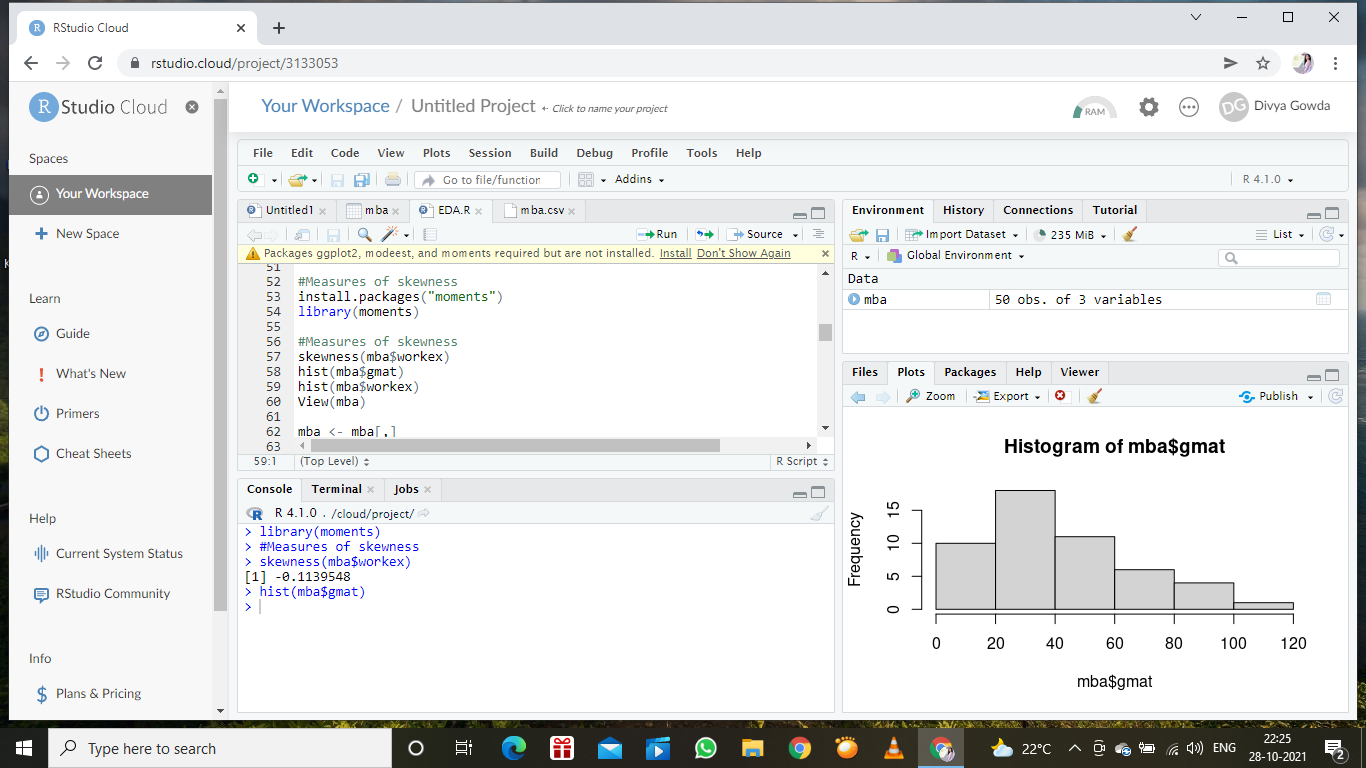
Expected value of the weight of that patient=145.33

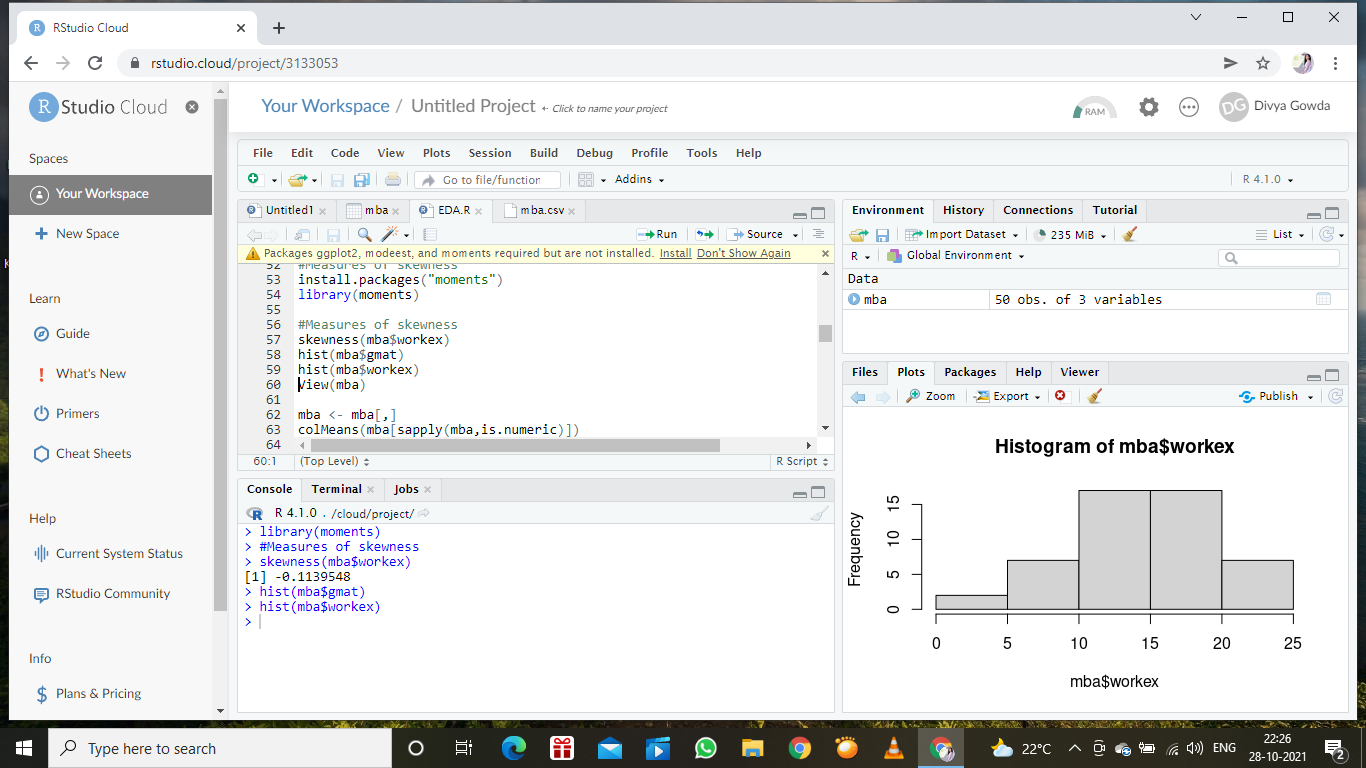
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

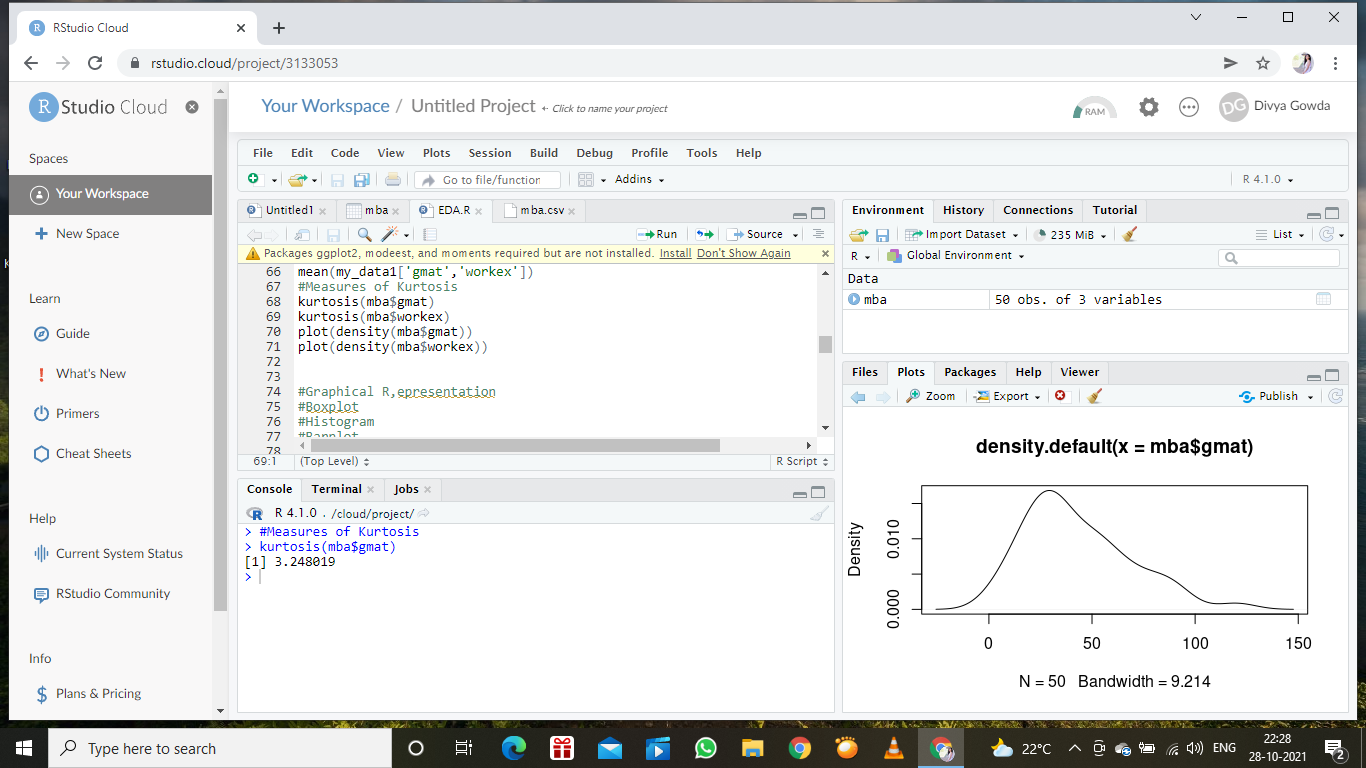
**Cars speed and distance Use Q9\_a.csv**

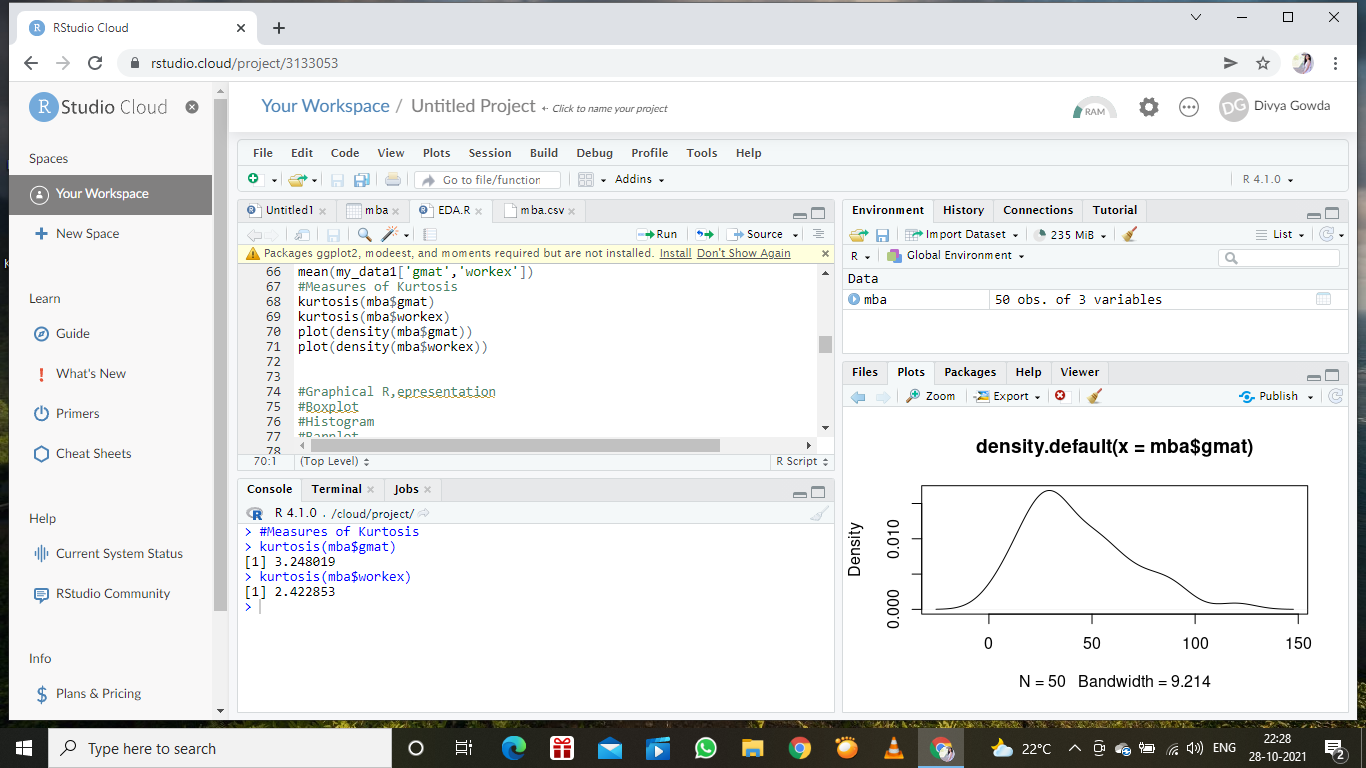
****

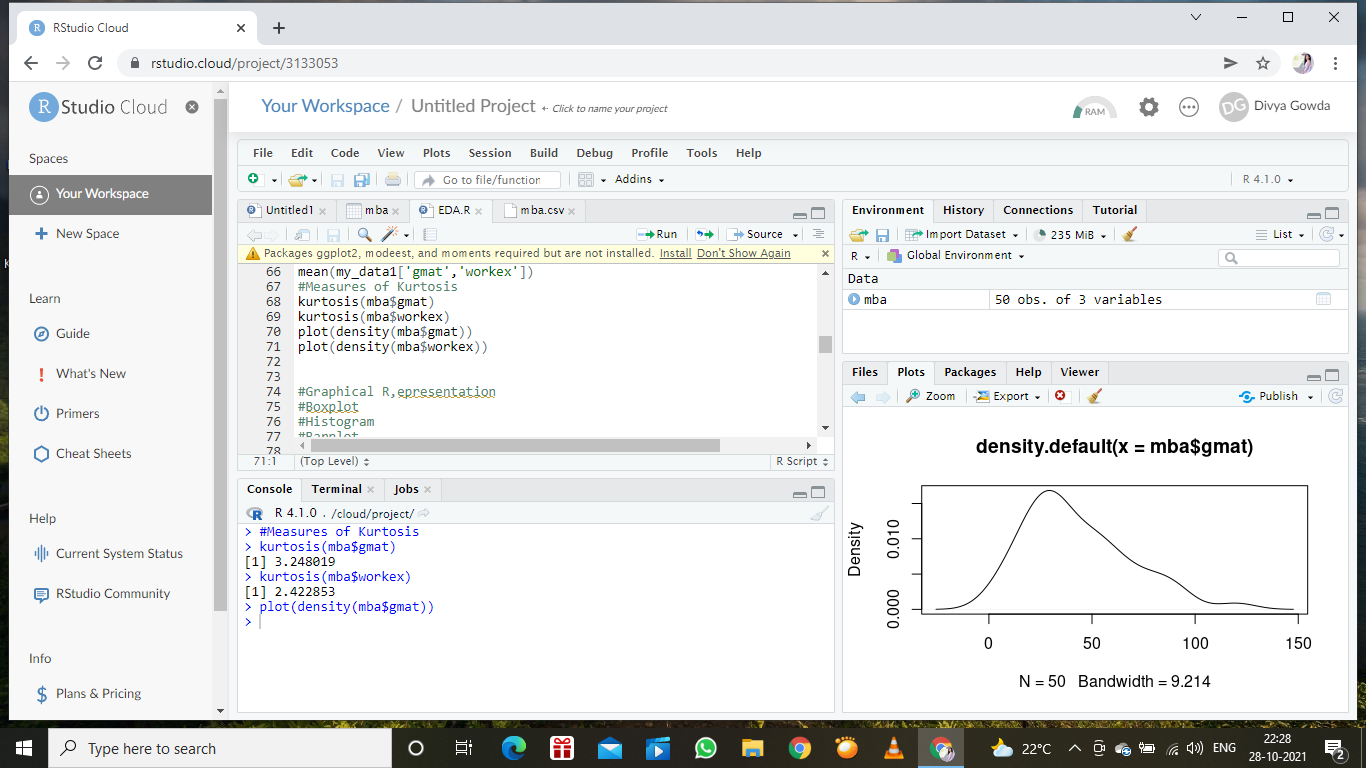
****

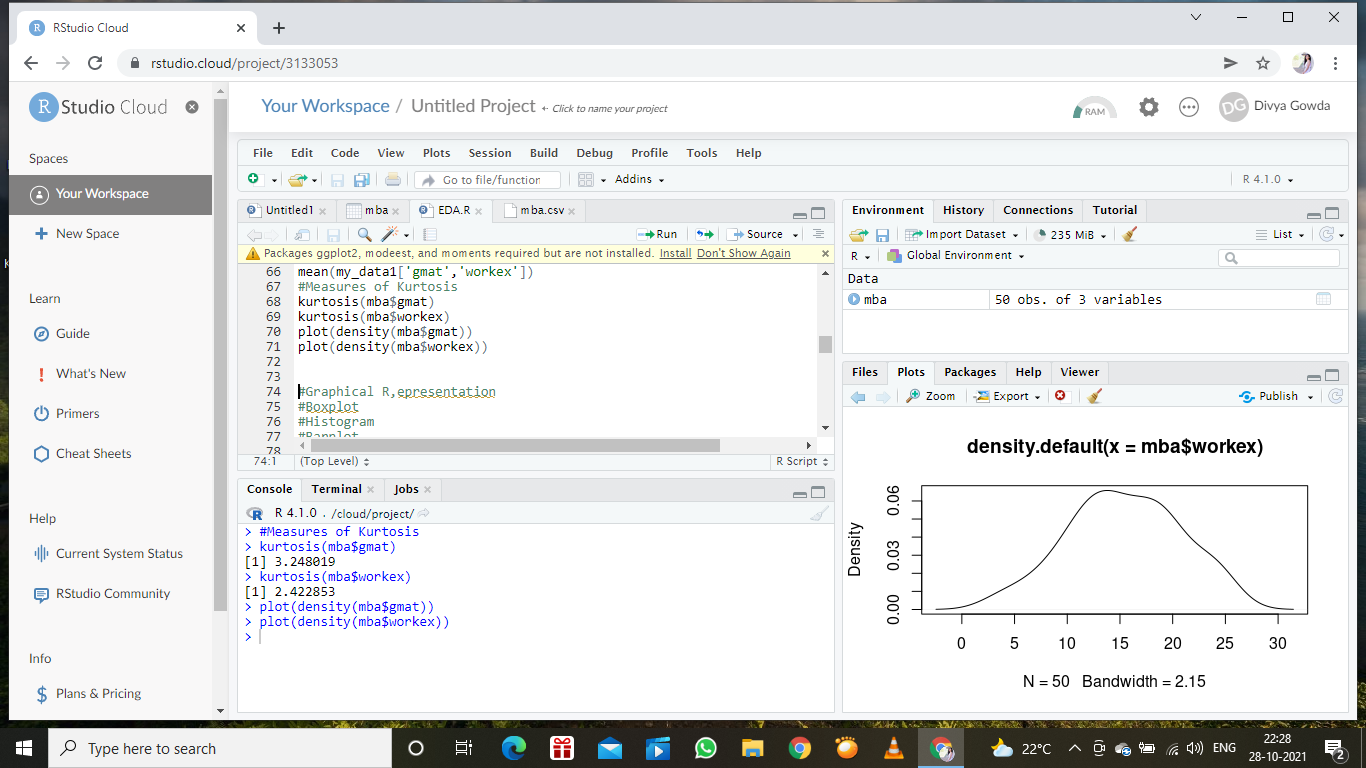
****

****

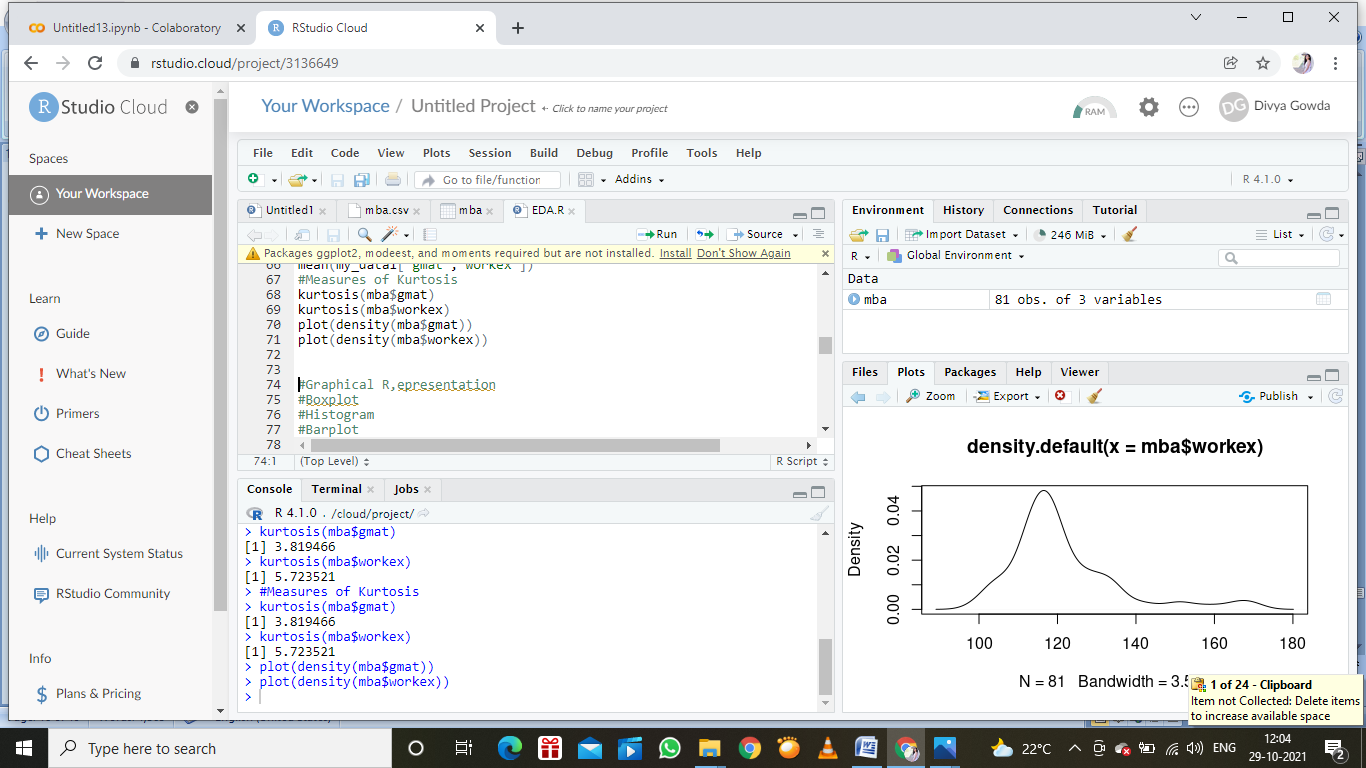
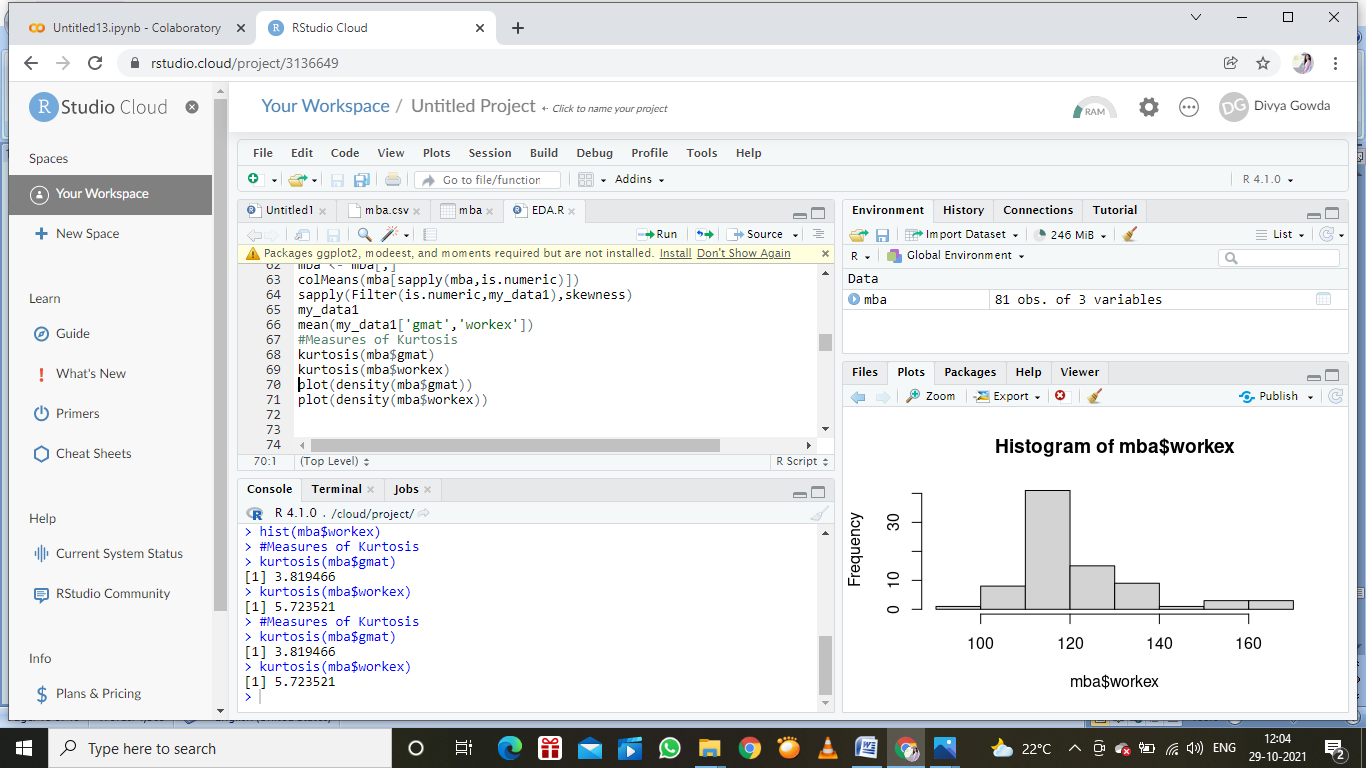
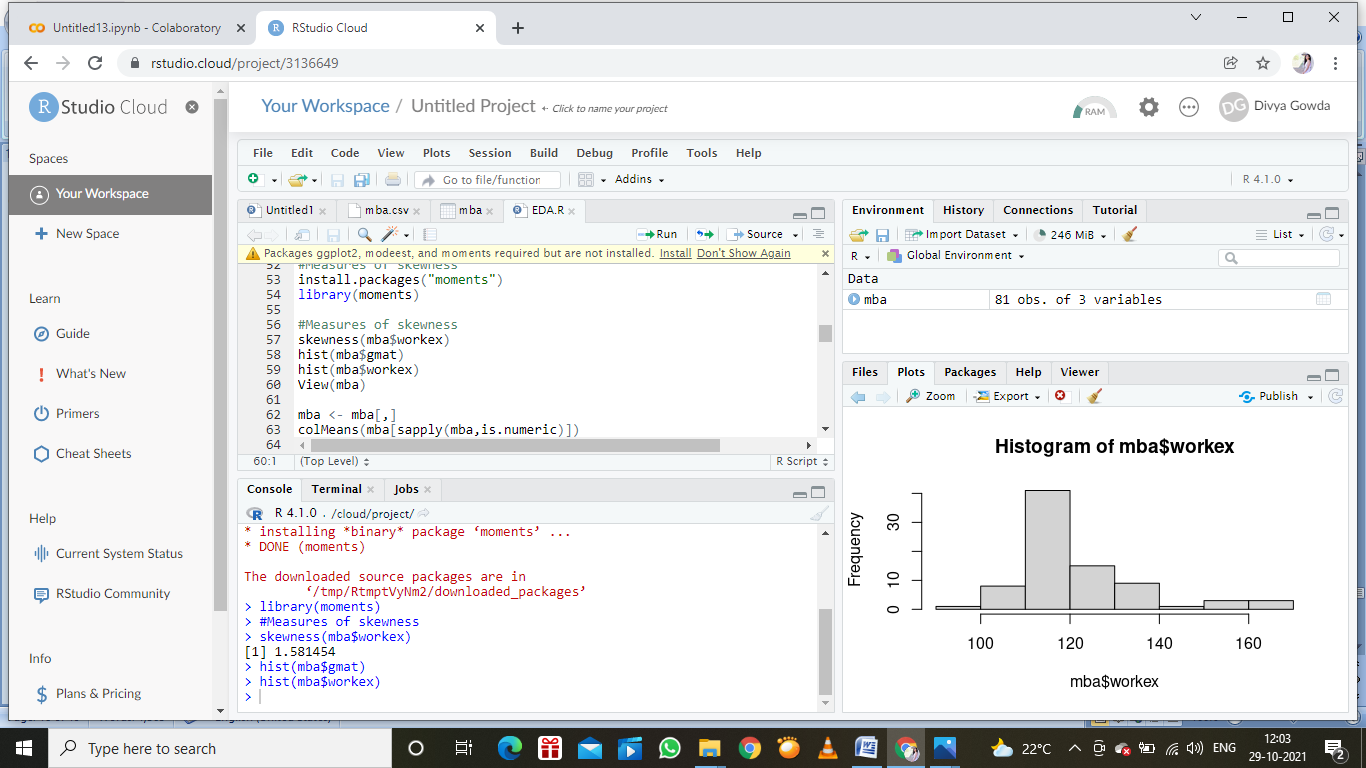
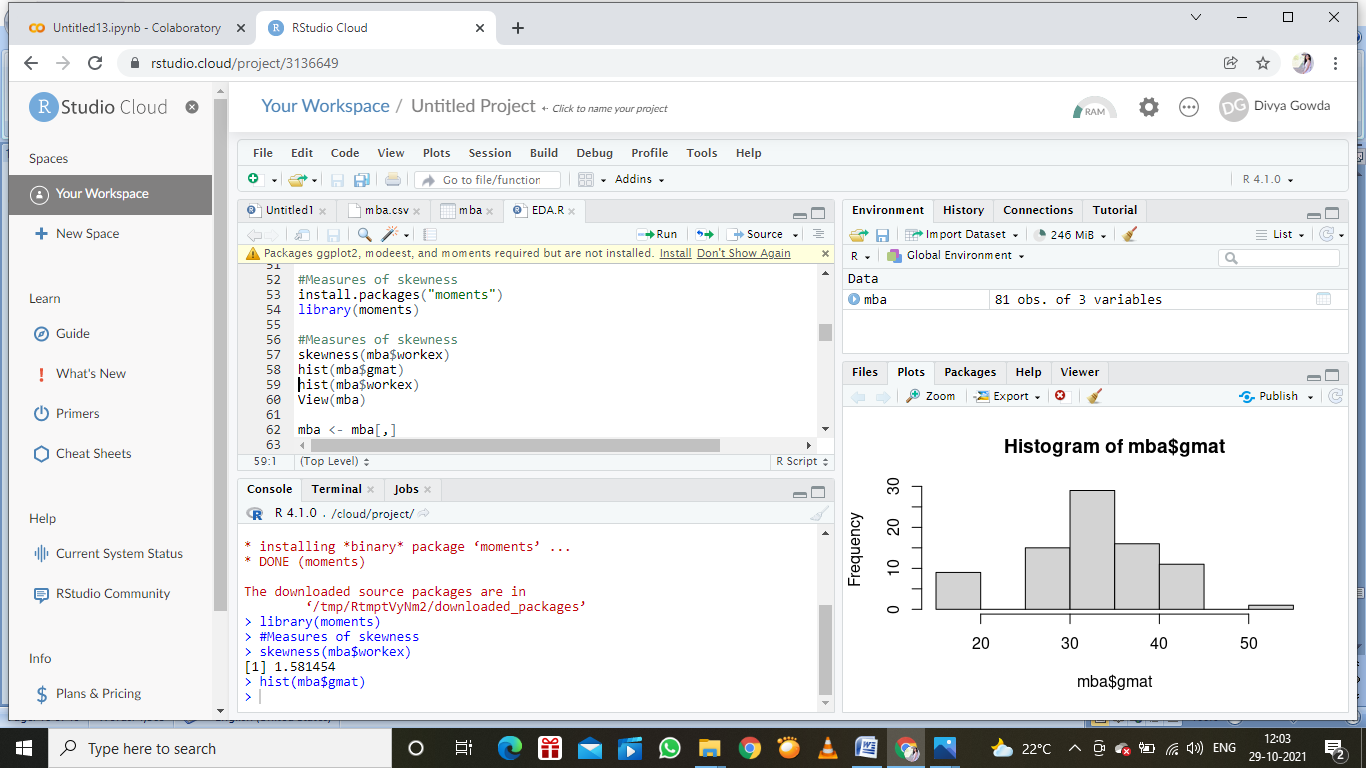
****

****

****

****

**SP and Weight(WT)**

**Use Q9\_b.csv**

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

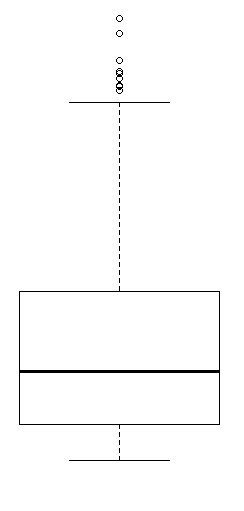


Answer:

-Chick weight data is right skewed or positivity skewed-Yes

-More than 50% chick weight is between 50 to 150-Yes

-Most of the chick weight is between 50 to 100-Yes



Answer:

-The data is right skewed.

-There are outlier at upper 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?

Answer:

1. CI=94%( Mu naught-1,04<x< Mu naught+1,04)

2. CI=94%( Mu naught-1,04<x< Mu naught+1,04)

3. CI=96%( Mu naught-1,75<x< Mu naught+1,75)

Sample size n=3000000

Sample mean x=200

Standard deviation s=30

From z-table values of z(c):

CI 94% Confidence level a=6% a=0,06 z(c)=1,55

CI 98% Confidence level a=2% a=0,02 z(c)=2,05

CI 96% Confidence level a=4% a=0,04 z(c)=1,75

MOE=z(c)\*sigma/square root of n

1. MOE=1,55\*30/square root of 2000 MOE=1,04
2. MOE=2,05\*30/square root of 2000 MOE=1,38
3. MOE=1,75\*30/square root of 2000 MOE=1,17

Then CI

1. CI=94% (Mu naught-MOE<x<Mu naught-MOE)

CI=(Mu naught-1,04<x<Mu naught+1,04)

1. CI=98%

CI=(Mu naught-2,04<x<Mu naught+2,05)

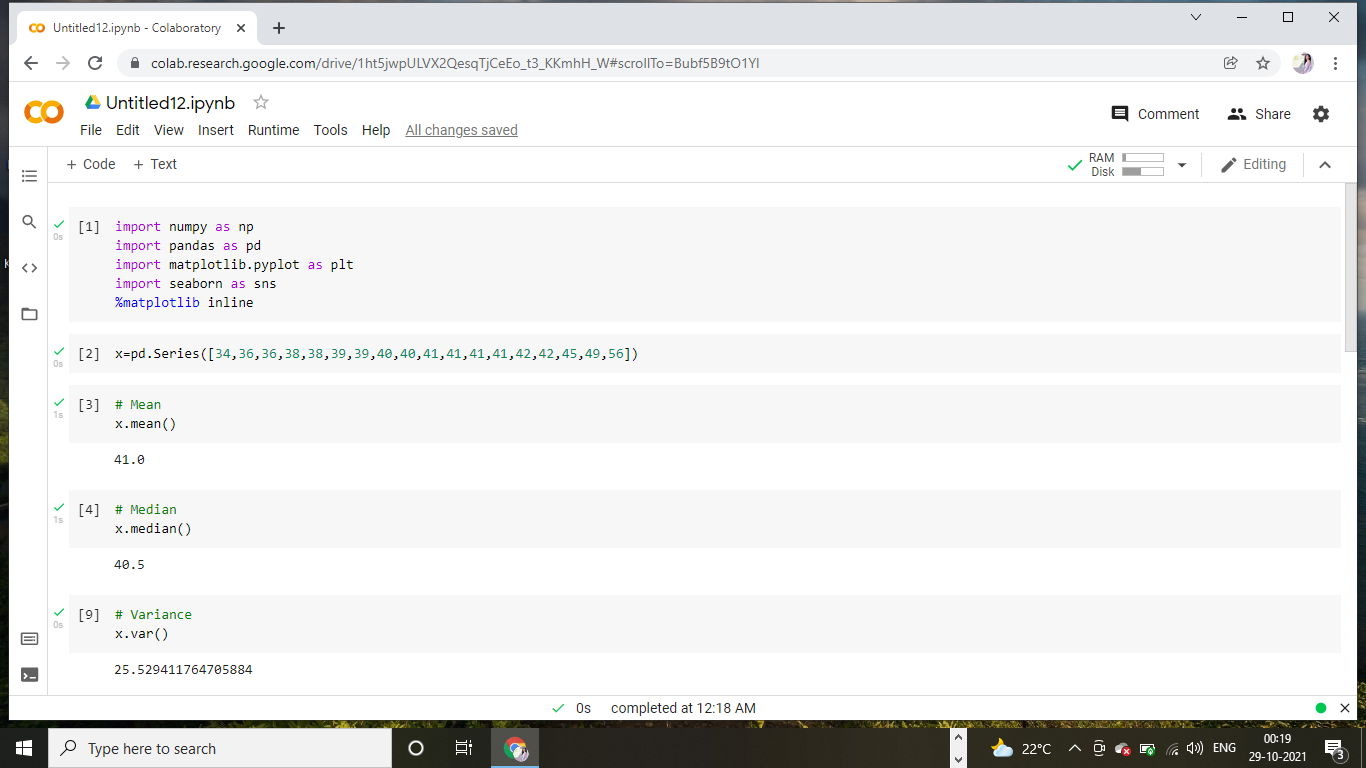
1. CI=96%

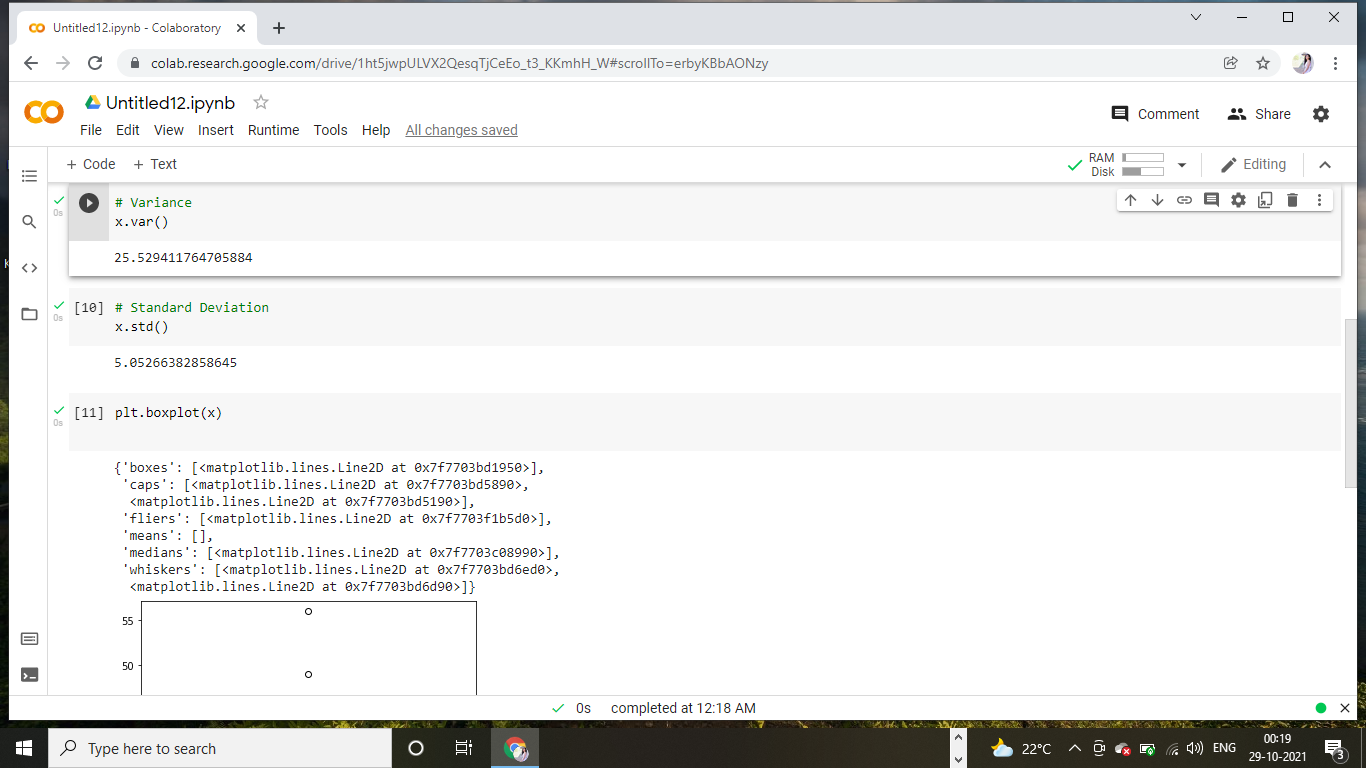
CI=(Mu naught-1,75<x<Mu naught+1,75)

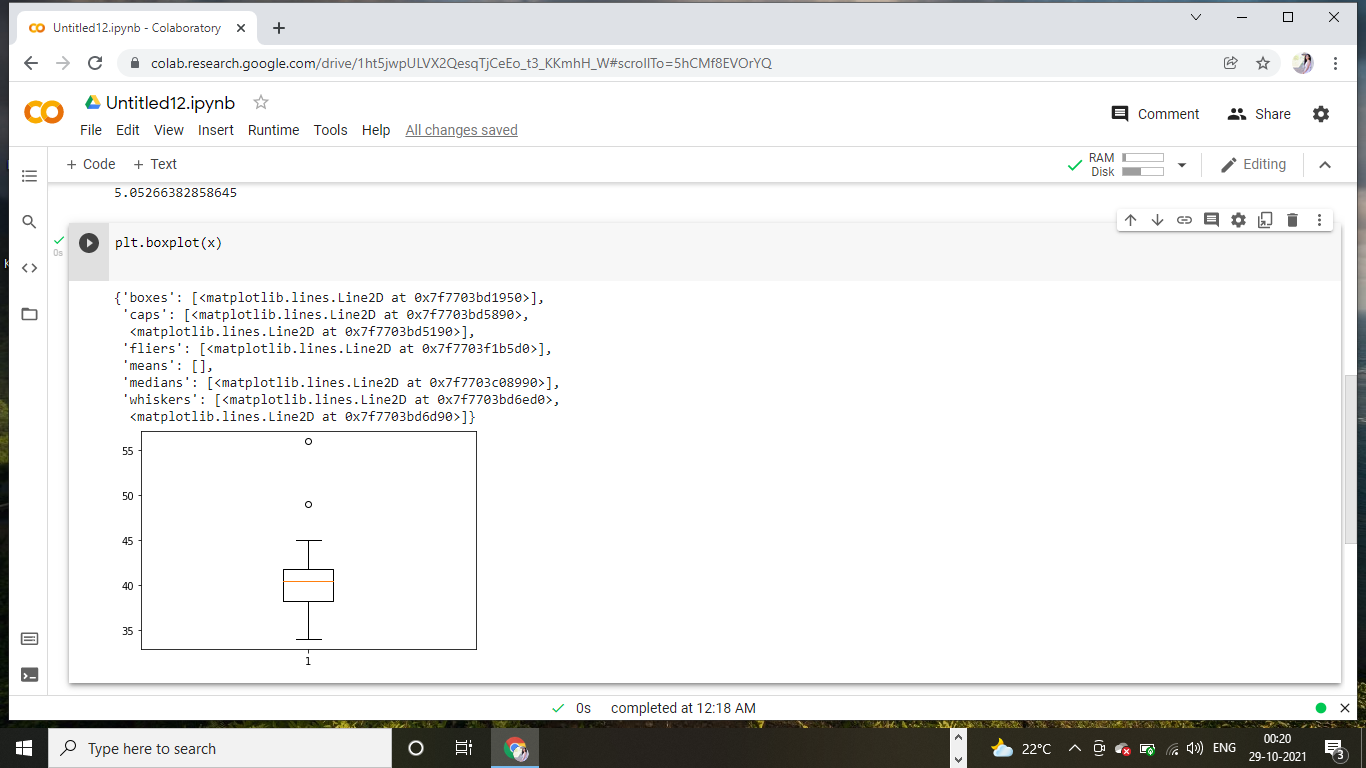
**Q12)**Below are the scores obtained by a student in tests

**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?







Inference: 1. There are 2Outliars in students marks:49 and 56

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

Answer: The distribution is symmetric ,the mean=median=mode. then the mean is equal to the median, and the distribution has zero skewness.

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

Answer: The distribution is positive.

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

Answer: The distribution is negative.

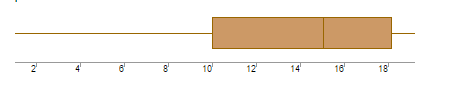
Q16) What does positive kurtosis value indicates for a data ?

Answer: Positive values of kurtosis indicate that distribution is peaked and possesses thick tails. (kurtosis>3)

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

Answer: Negative values of kurtosis indicate that a distribution is flat and has thin tails.

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



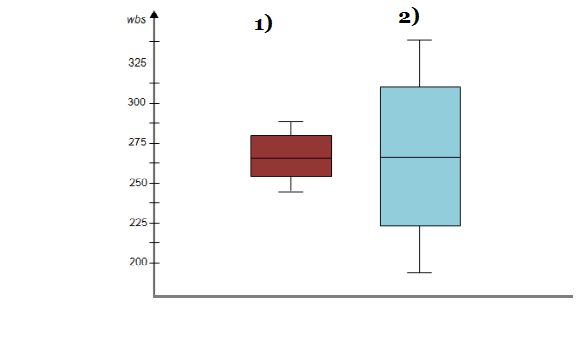
What can we say about the distribution of the data?

Answer: Not normally distributed.

What is nature of skewness of the data?

Answer: Negative skewness.

What will be the IQR of the data (approximately)?   
Answer: 10-18.  
  
Q19) Comment on the below Boxplot visualizations?



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

Answer: The box plot 1 designed with range=3, The second one range is=1.5

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

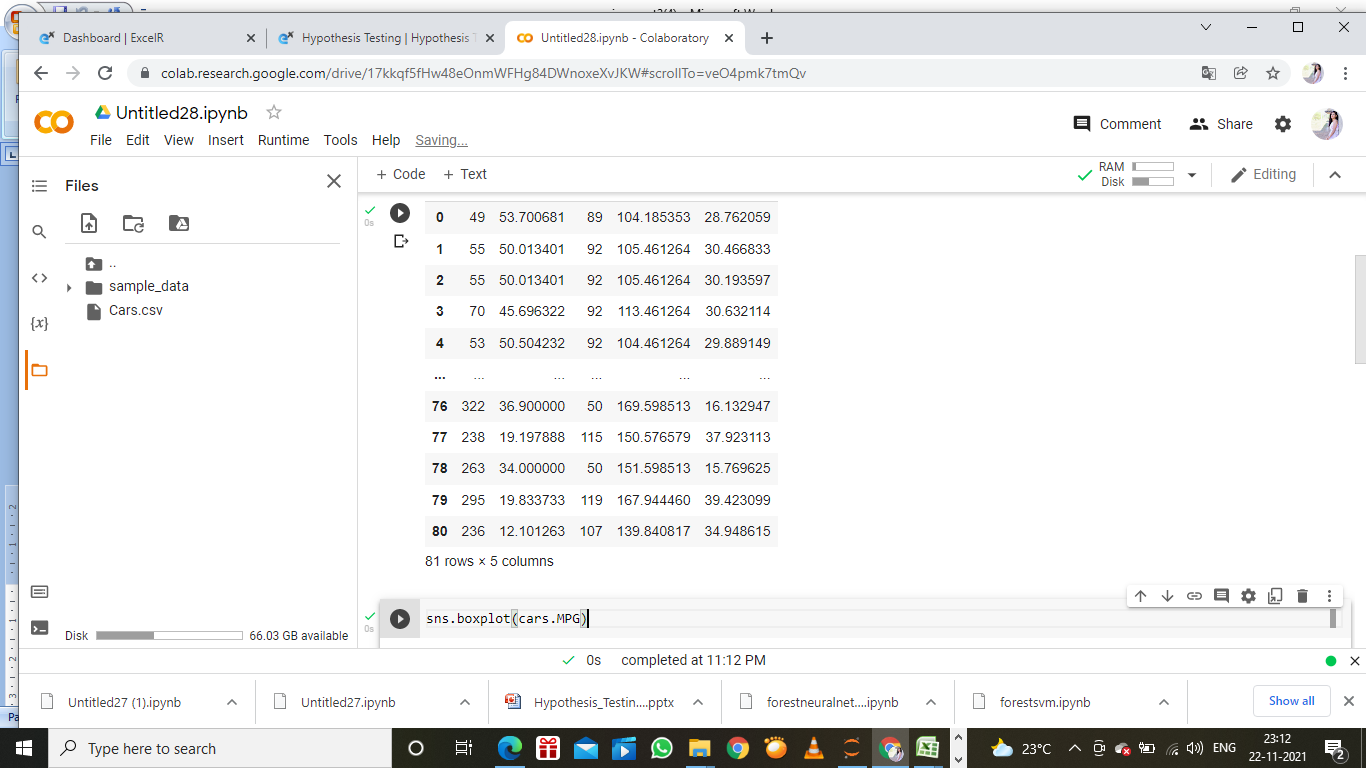
Data \_set: Cars.csv

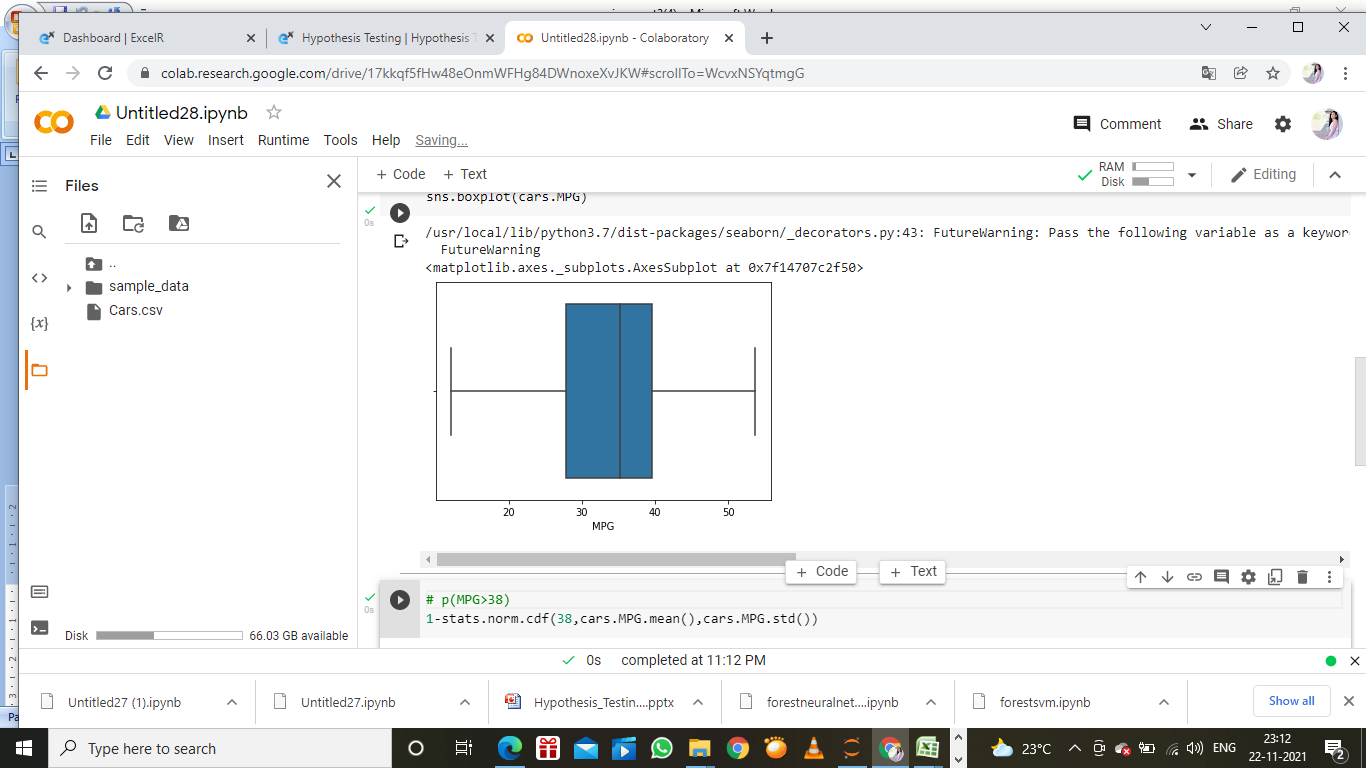
Calculate the probability of MPG ofCars for the below cases.

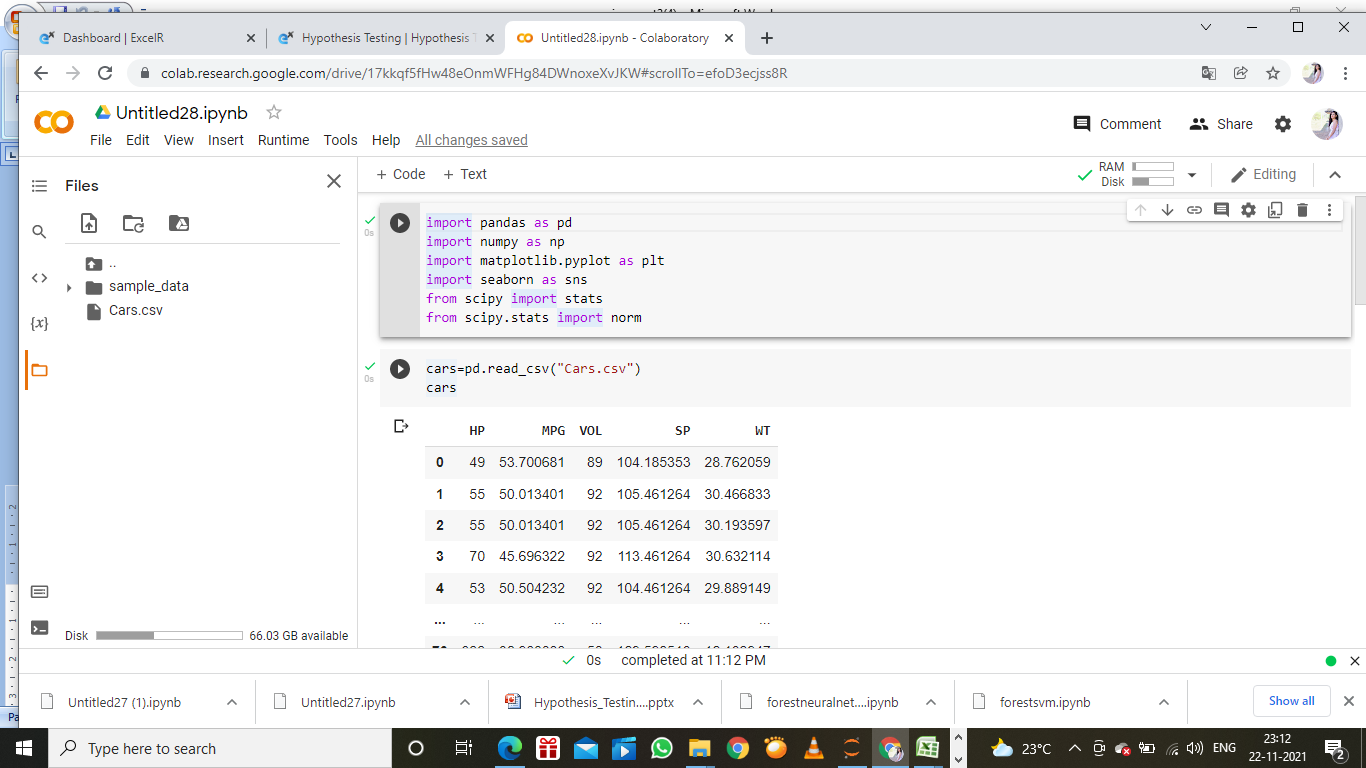
MPG<- Cars$MPG

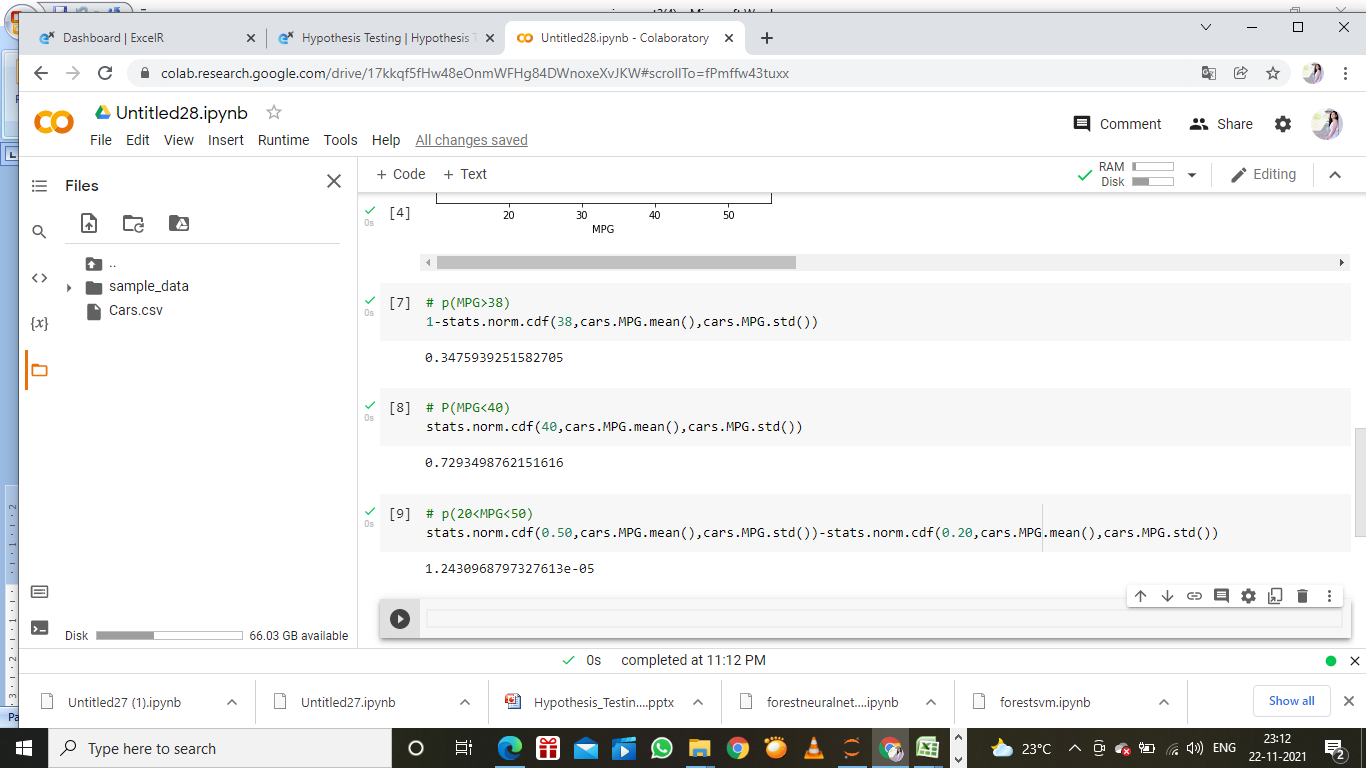
* 1. P(MPG>38)
  2. P(MPG<40)

c. P (20<MPG<50)





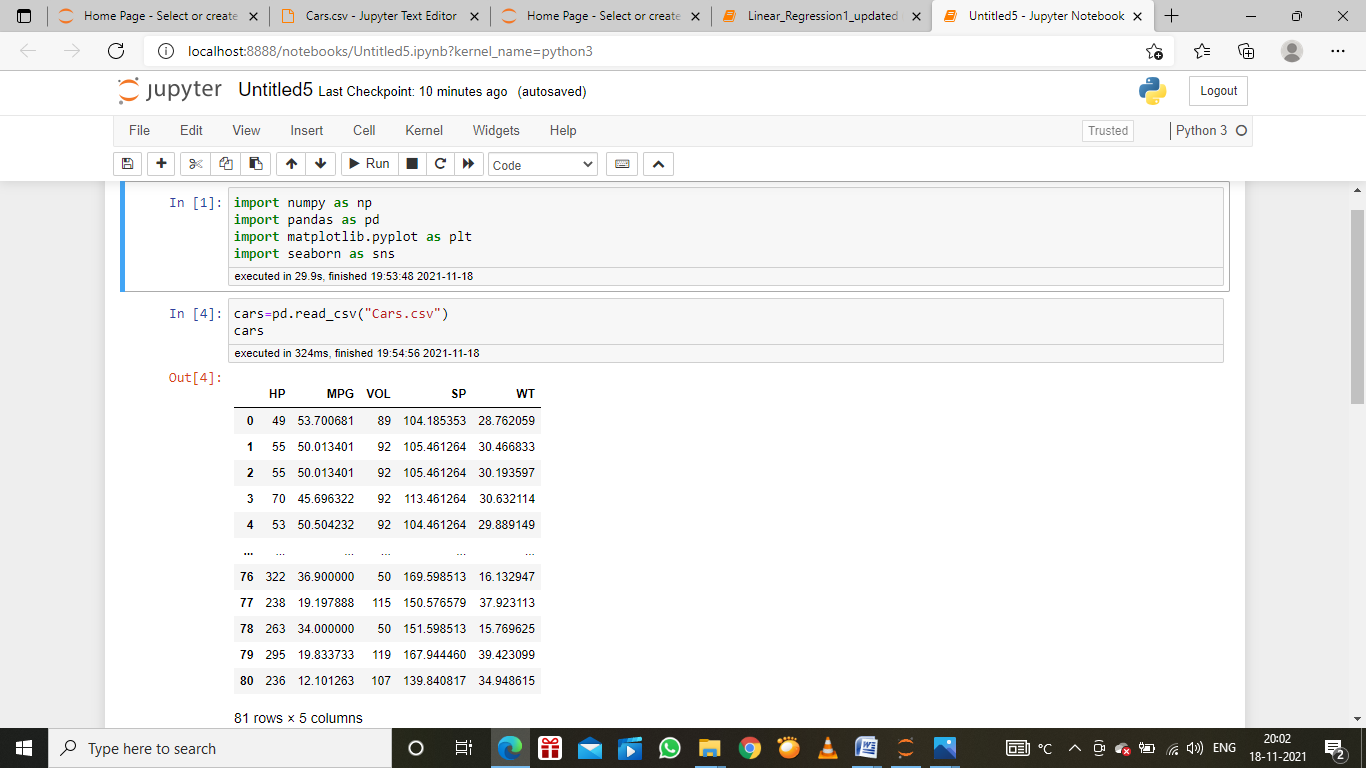


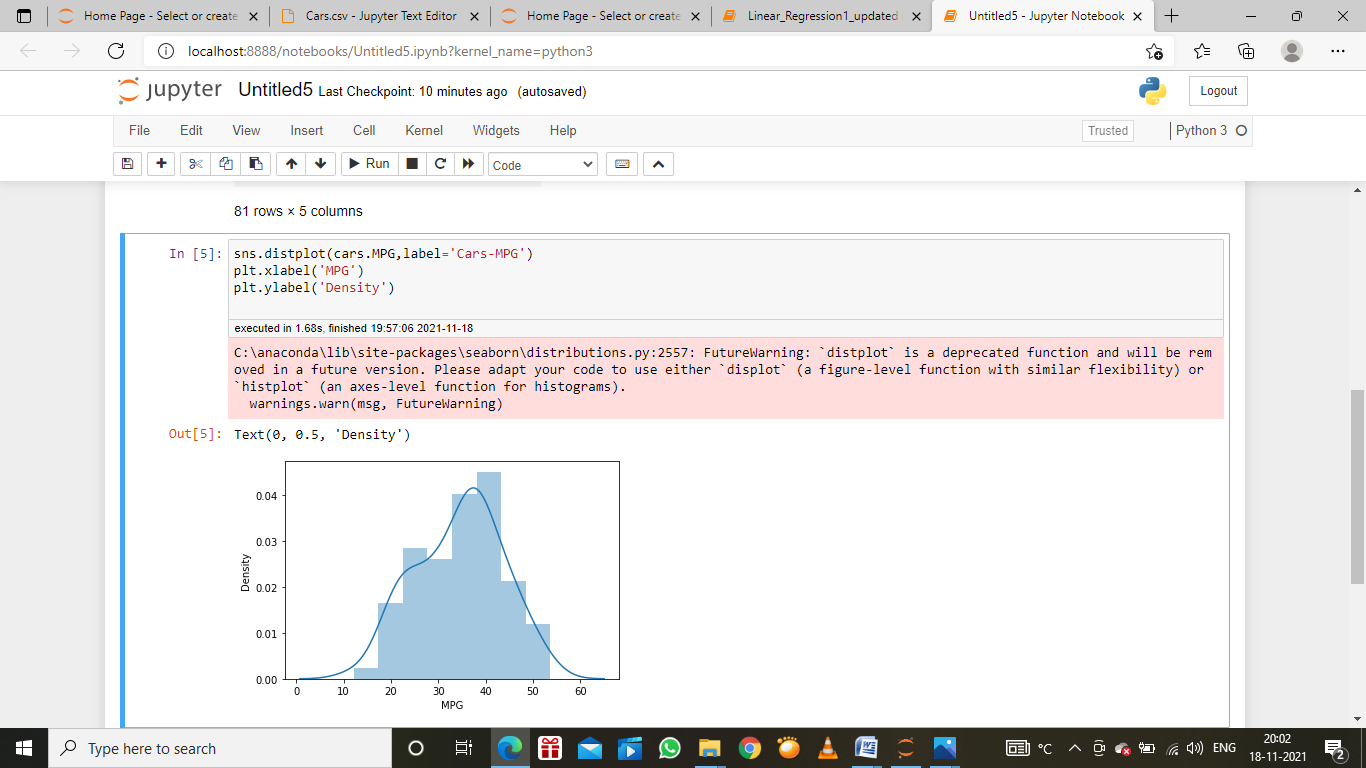


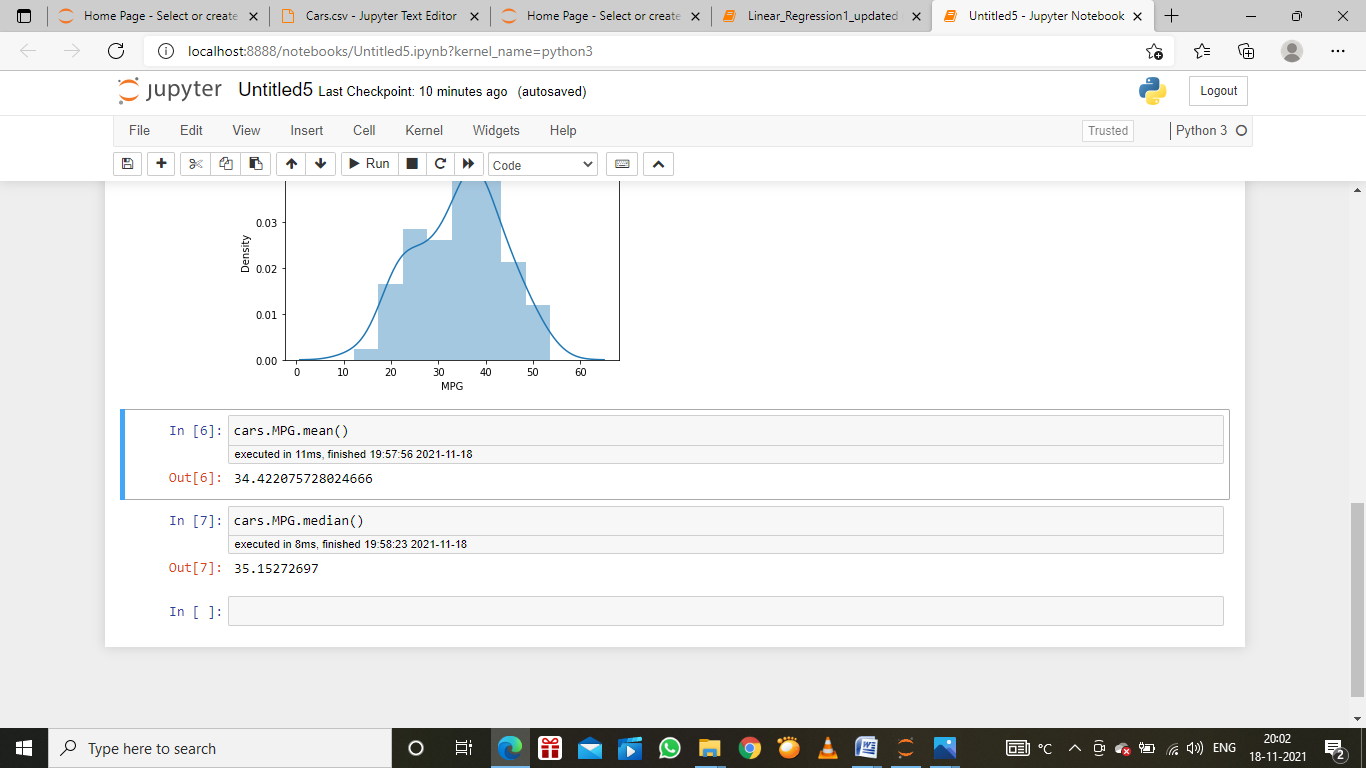
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

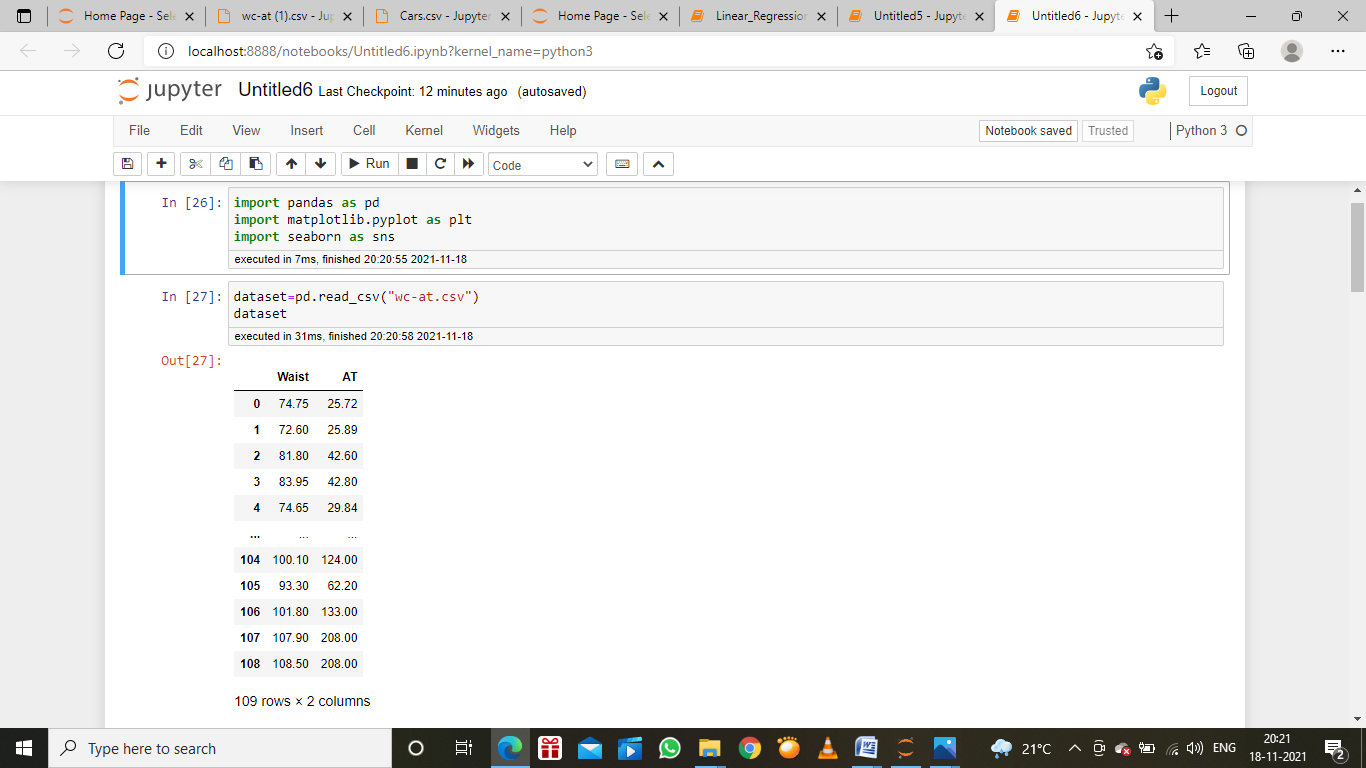


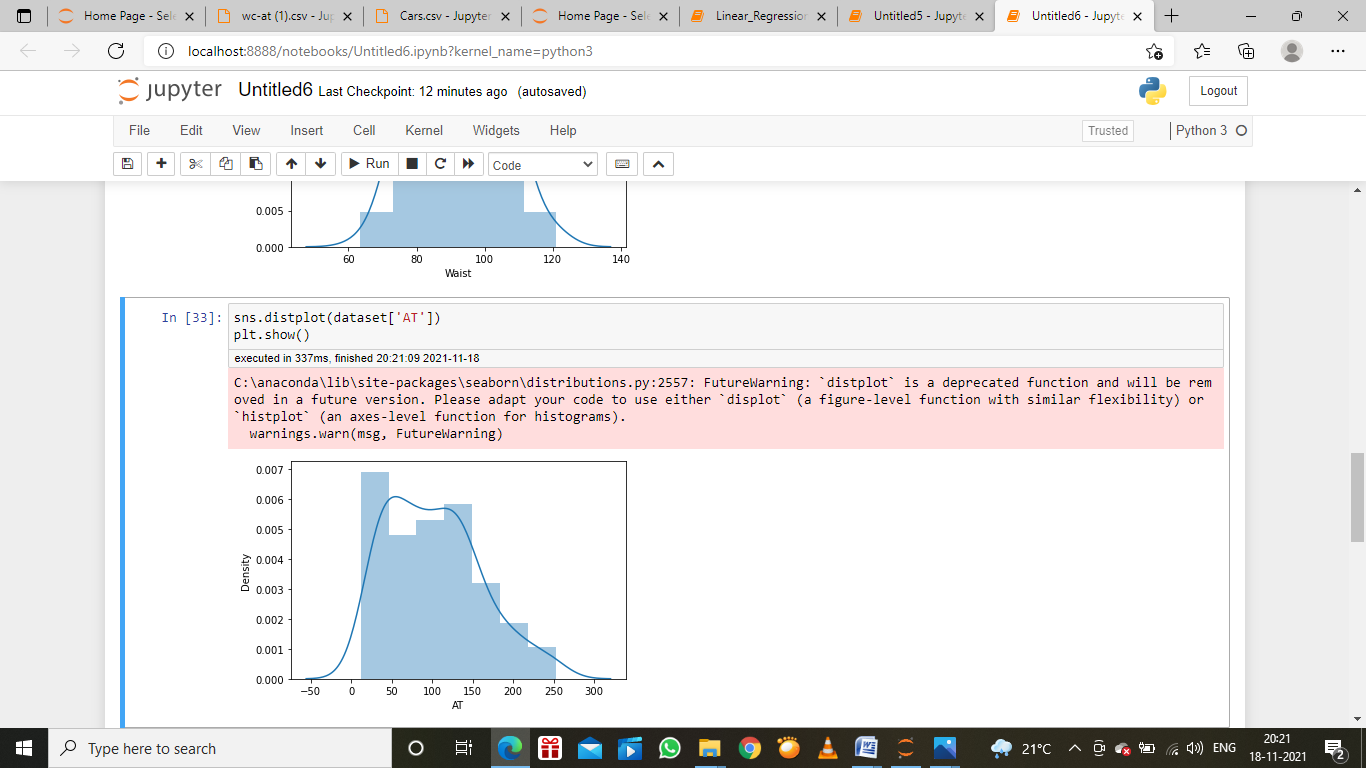
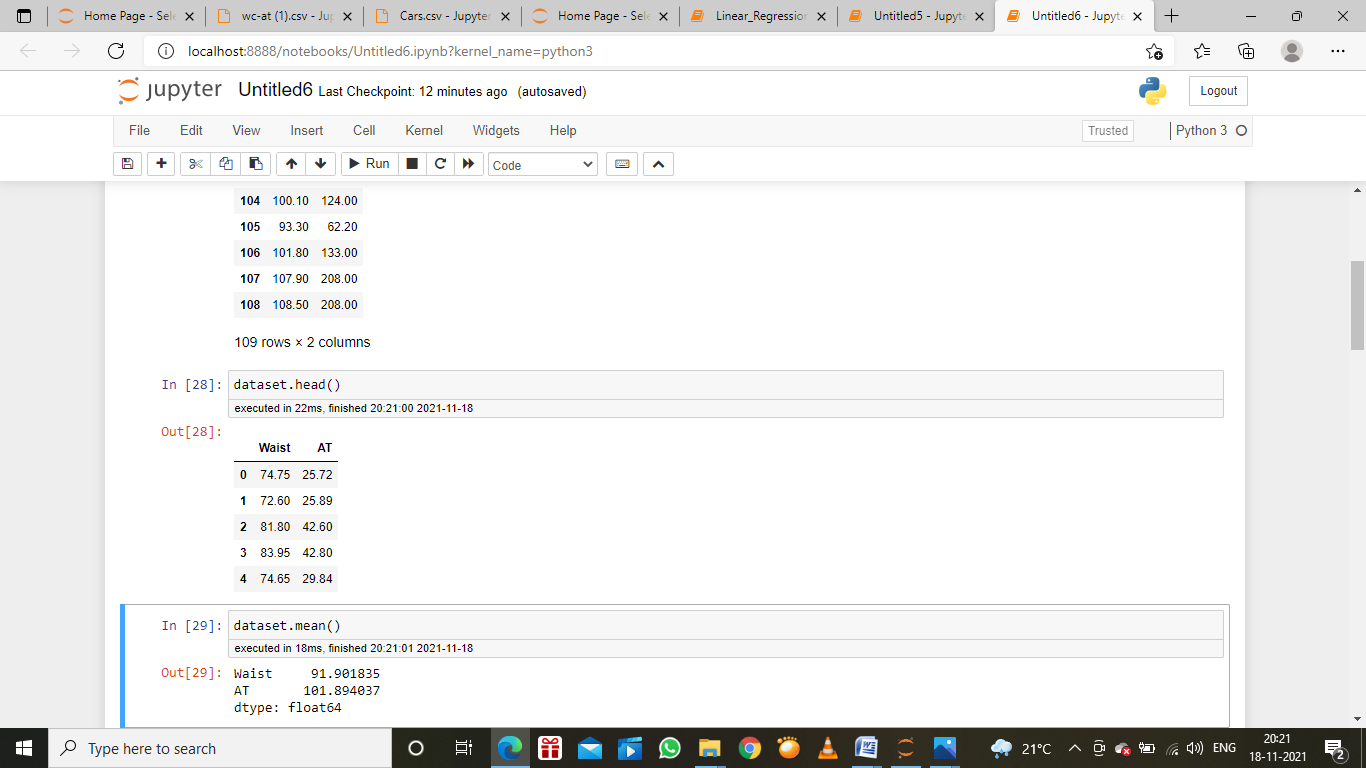
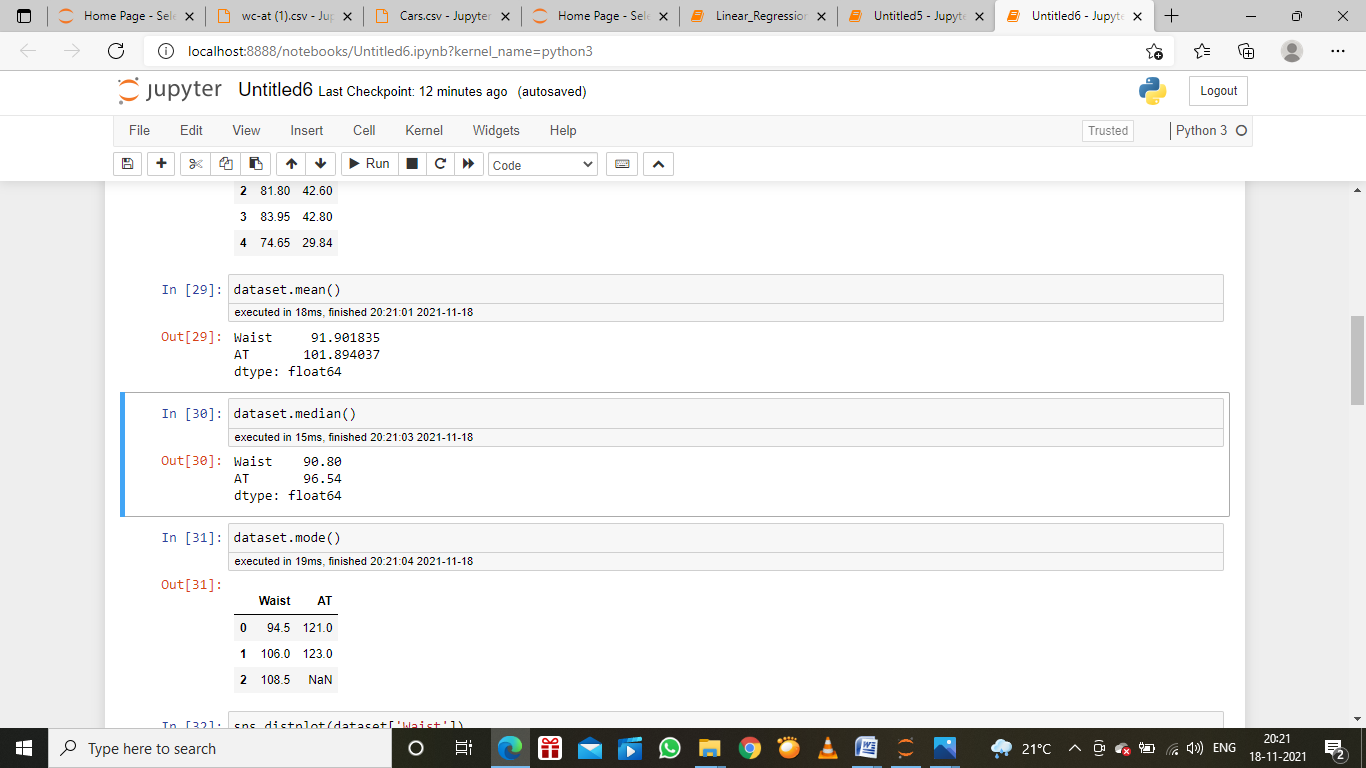


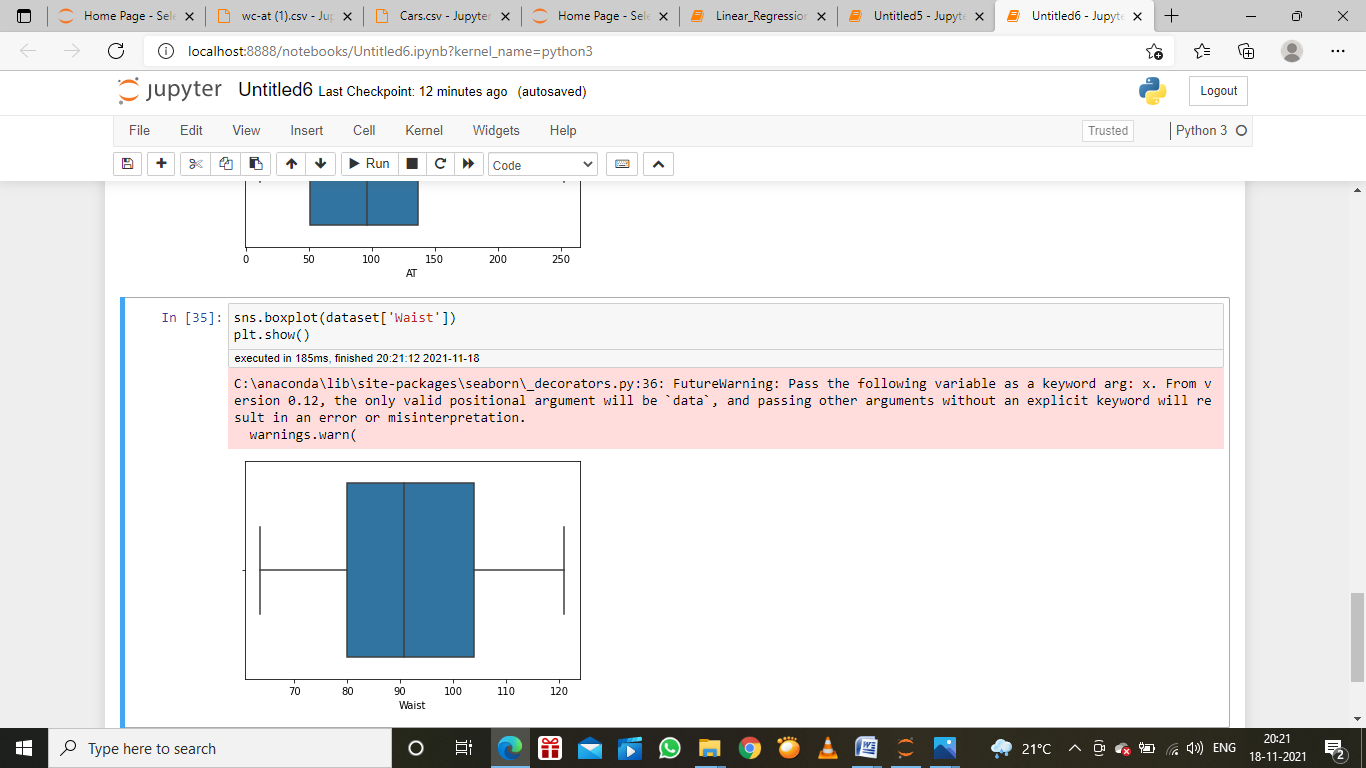
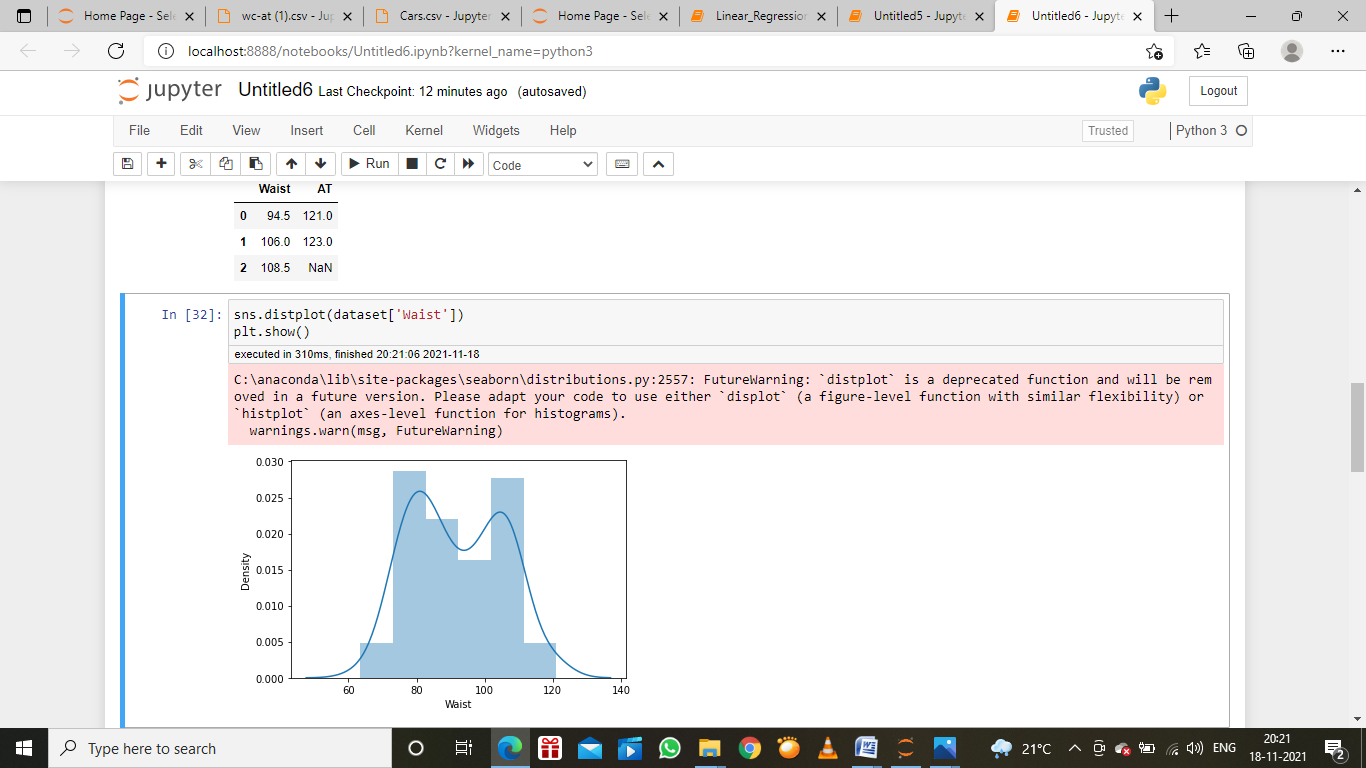
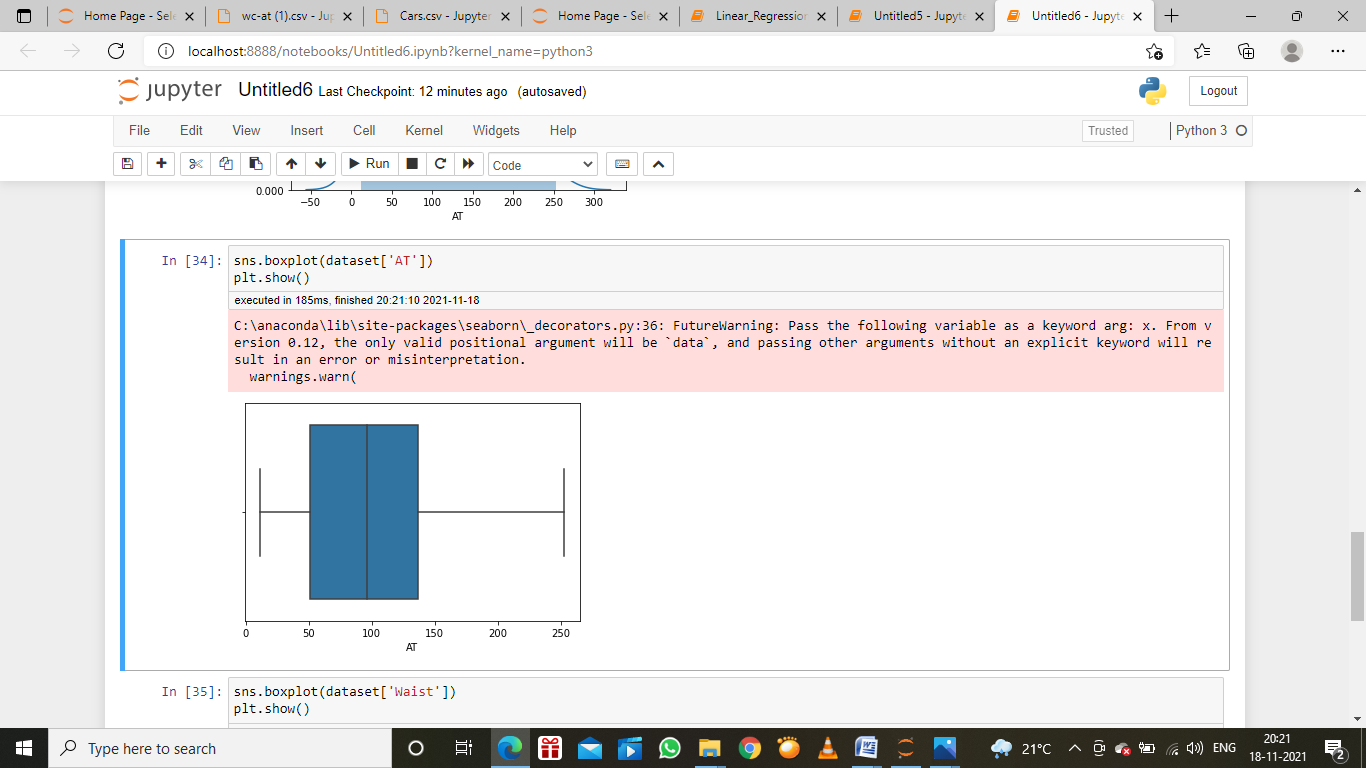


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv







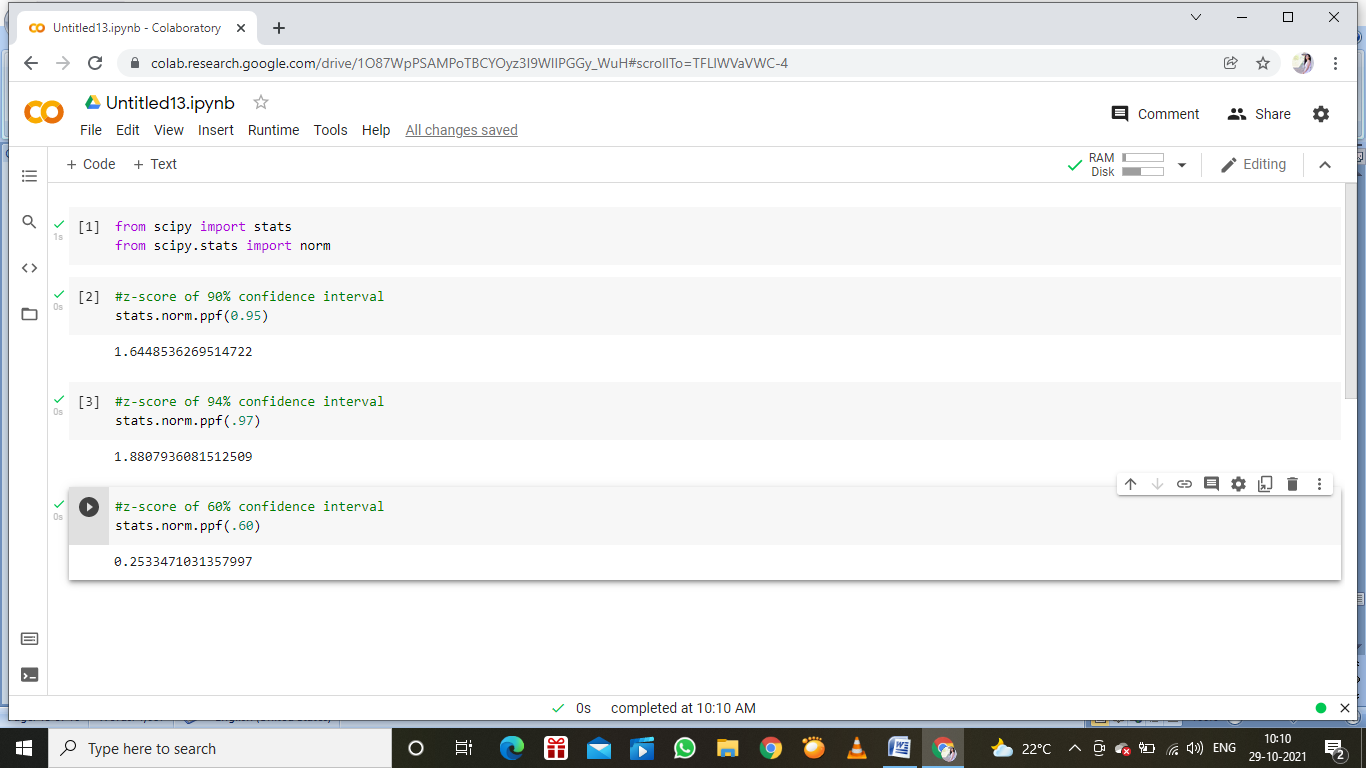
Q 22) Calculate the Z scoresof 90% confidence interval,94% confidence interval, 60% confidence interval

Answer:

90%=1.645

94%=1.880

60%=0.253



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

Answer: To compute the 95% confidence interval, start by computing the mean and standard error: M=(2+3+4+5+6+9)/5=5. Sigma M==1.118. z.95 can be found using the normal distribution calculator and specifying that the shaded area is 0.95. and indicating that you want the area to be between the cutoff points

Confidence level z

0.90 1.645

0.92 1.75

0.95 1.96

0.96 2.05

With a 90 percent confidence interval, you have a 10 percent chance of being wrong. A 99 percentage confidence interval would be wider than a 95 percent confidence interval.

Q 24**)**A Government companyclaims 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

Answer:

t-statistics for the data is given as follows:

t=\dfrac{x-\mu}{\frac{s}{\sqrt}}t=

n

s

x-mu

x=mean of the sample of bulbs=260

mu=population mean=270

s=standard deviation of the sample=90

n=number of items in the sample=18

t=\dfrac{260-270}{\frac{90}{\sqrt 18}}t=

1

8

90

260-270

t=\dfrac{-10}{\frac{30}{\sqrt2}}t=

2

30

-10

t=\dfrac{-1\times\sqrt2}{3}t=

3

-1

2

t=-0.471

For probability calculation, the number of degree of freedom is n -1,so here you need the t-distribution with 17 degree of freedom.

The probability that t<-0.471 with 17 degree of freedom assuming the population mean is true, the t-value is less than the t-value obtained with 17 degree of freedom and a t score of -0.471,the probability of the bulbs lasting less than 260 days on average of 0.3218 assuming the mean life of the bulbs is 300 days.