|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | District |
| Results of rolling a dice | District |
| 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 | District |
| Number of kids | District |
| Number of tickets in Indian railways | District |
| Number of times married | District |
| Gender (Male or Female) | District |

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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| 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?

Ans- s=8, a=3

P(one tail two head)=a/s=3/8

=0.375

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

Total event=36

1. Equal to 1 -p=0
2. Less than or equal to 4 -p=6/36=0.166
3. Sum is divisible by 2 and 3 -p=6/36=0.166

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- total ball=7,

No of way drawing two ball out of 7= (7\*6) / (1\*2) =21

And from them ball is not blue 7-2=5

From five ball 2 are taken randomly = (5\*4) / (1\*2) =10

probability that none of the balls drawn is blue =10/21

p=0.476

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- Expected number of candies for a randomly selected child

(if child is randomly selected it will got candies)

=1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120

=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-

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Mean | Median | Mode | Variance | SD | Range |
| Points | 3.596 | 3.695 | 3.93 | 0.285 | 0.526 | 2.17 |
| Score | 3.217 | 3.325 | 3.44 | 0.957 | 0.963 | 3.911 |
| weight | 17.84 | 17.71 | 17.02 | 3.193 | 1.758 | 8.4 |

From the above values of points, score, weight of mean, median, mode is likely same So when mean=median=mode then distribution is normal distribution.

Therefore from above values distribution is likely normal distribution.

**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= P(selecting each person)=1/9

Expected Value of the Weight of that patient

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

=145.33 pounds

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans-**

**Skewness**

**Car speed =** **from scipy.stats import skew**

**skew(df['speed'],axis=0,bias=True)**

**=-0.113**

**Distance= skew(df['dist'],axis=0,bias=True)**

**=0.782**

**Kurtosis-**

**speed =** **from scipy.stats import kurtosis**

**###### Kurtosis(df['speed'],axis=0,bias=True)**

**= -0.577**

**Distance= kurtosis(df['dist'],axis=0,bias=True)**

**=0.248**

**####SP and Weight(WT)**

**Use Q9\_b.csv**

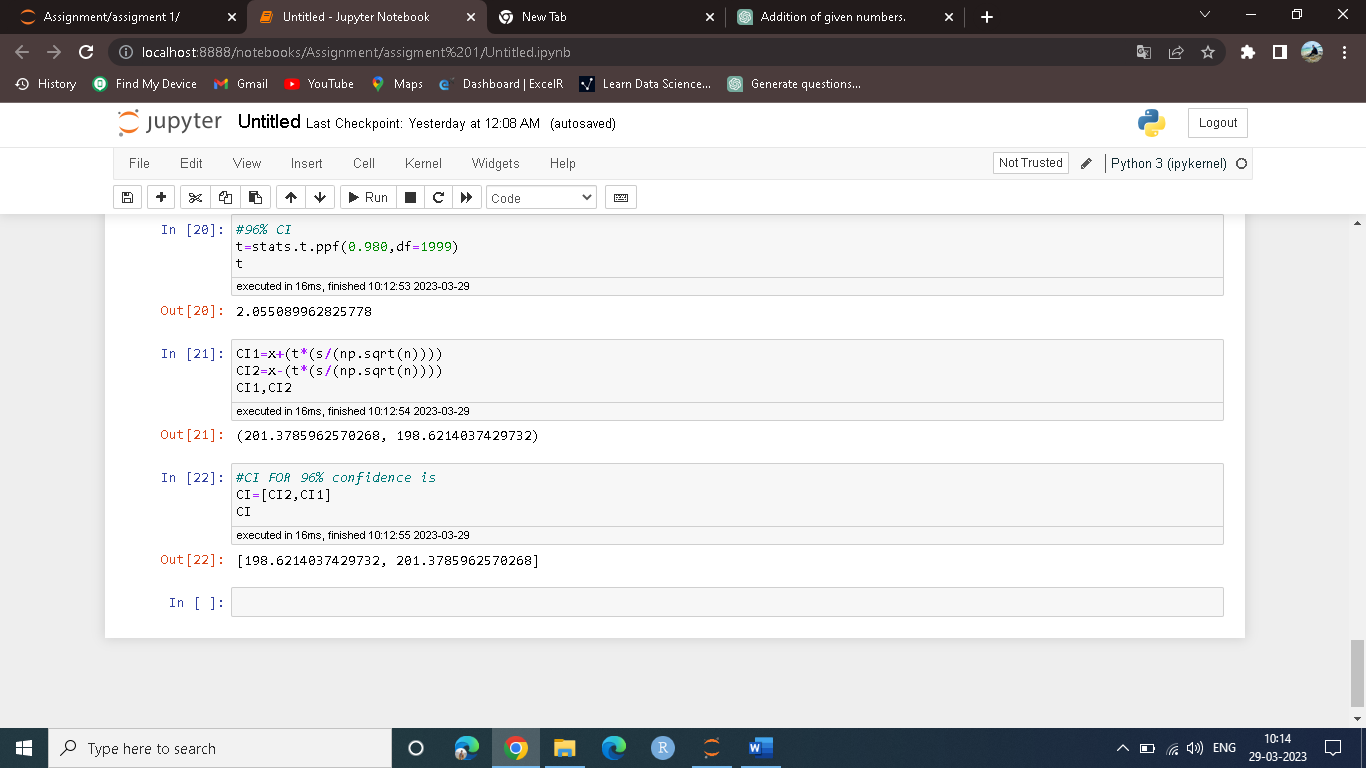
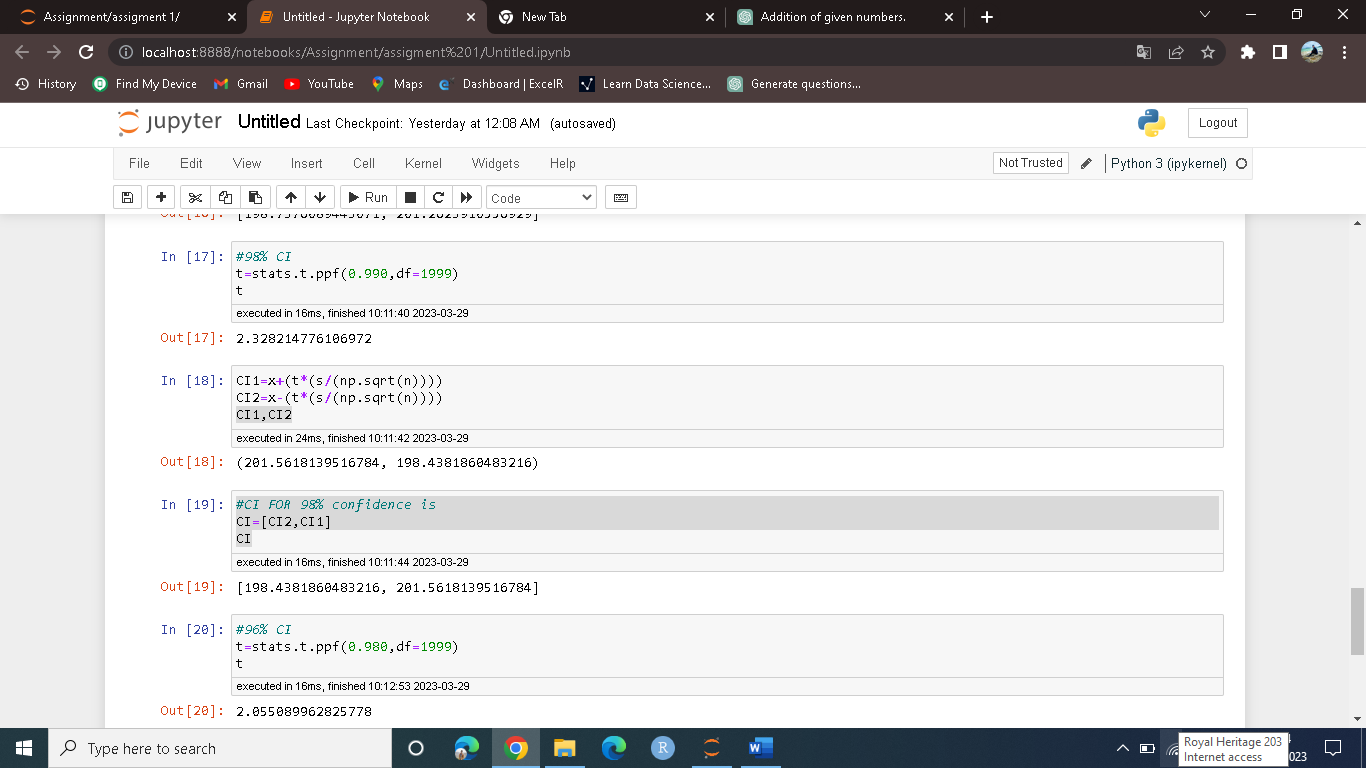
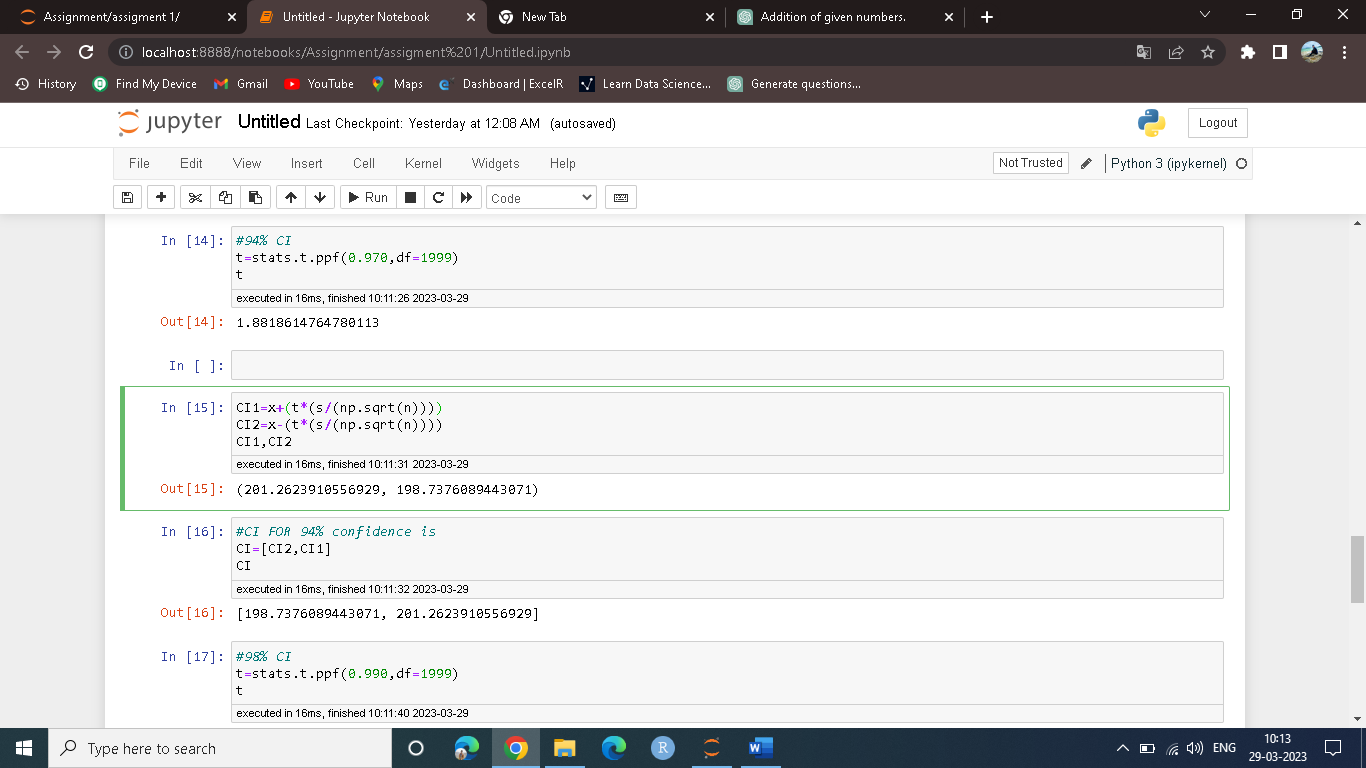
**Q10) Draw inferences about the following boxplot & histogram**



Ans- Histogram is positively Skew. We can predict that there is outlier above upper extreme. Most of data point is present in between 50-150 wt .

-From the box plot we see that the outlier is actually present above upper extreme and mean>median.

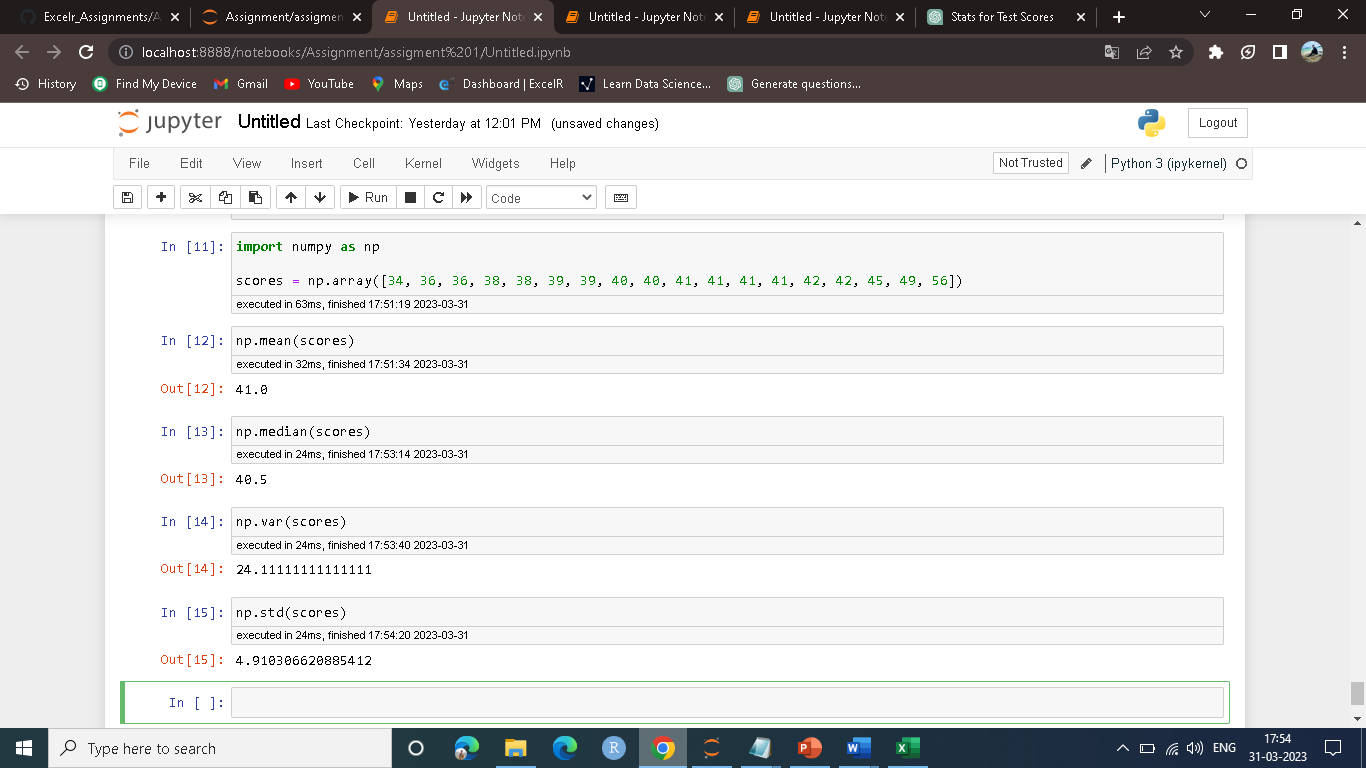
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate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

Ans- 

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

Ans- Its normally distributed .skewness zero

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

Ans-Its positive skewness, right skew

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

Ans-Its negative skewness, left skew

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

Ans-Positive kurtosis value indicate distribution is peaked and have thick tail.

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

Ans-Distribution is flatter than normal curve.

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



What can we say about the distribution of the data?

Ans – data is -ve skew or left skew. Here mean<median

What is nature of skewness of the data?

data is -ve skew or left skew

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

Ans-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.

Ans- Data of boxplot1 is closely distributed and boxplot2 data is somehow widely distributed boxplot1 and boxplot2 is seems to be like normally distributed.

Both boxplot is same app 262.

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

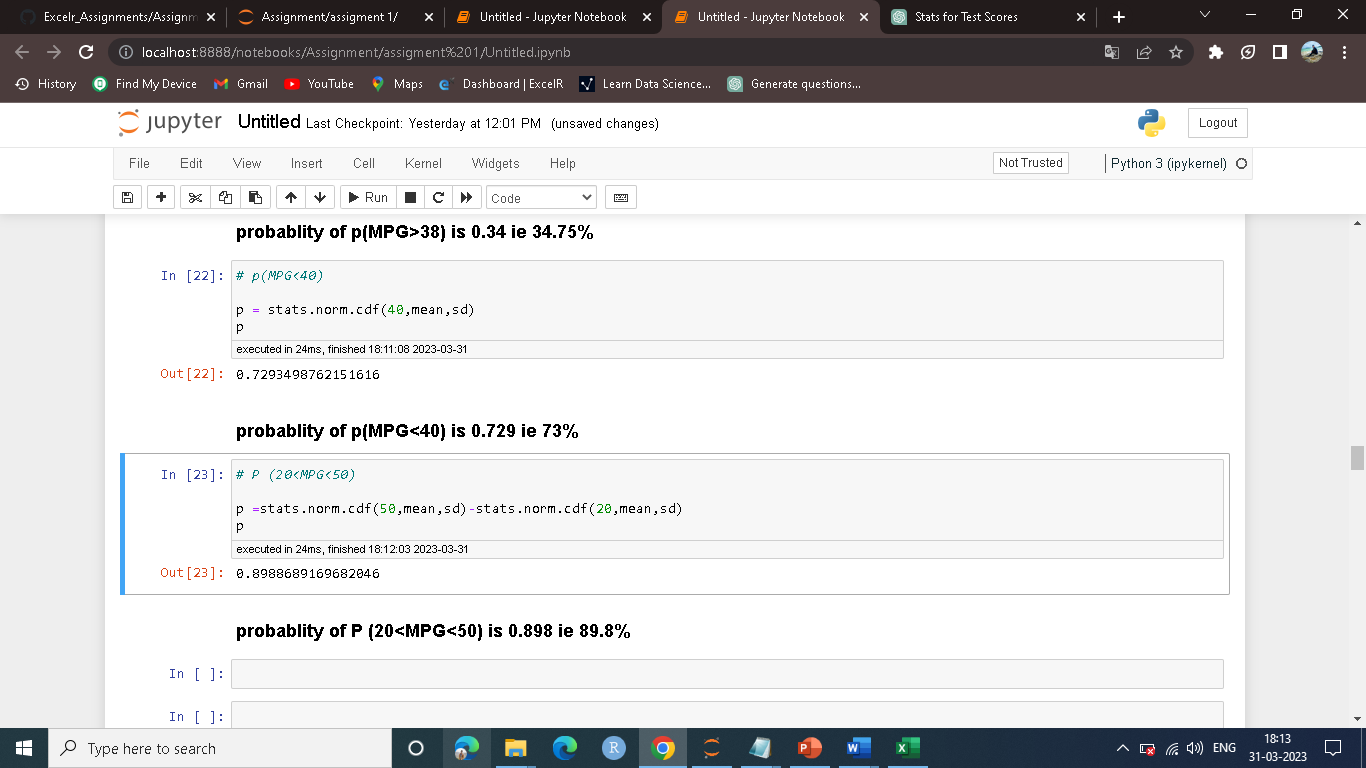
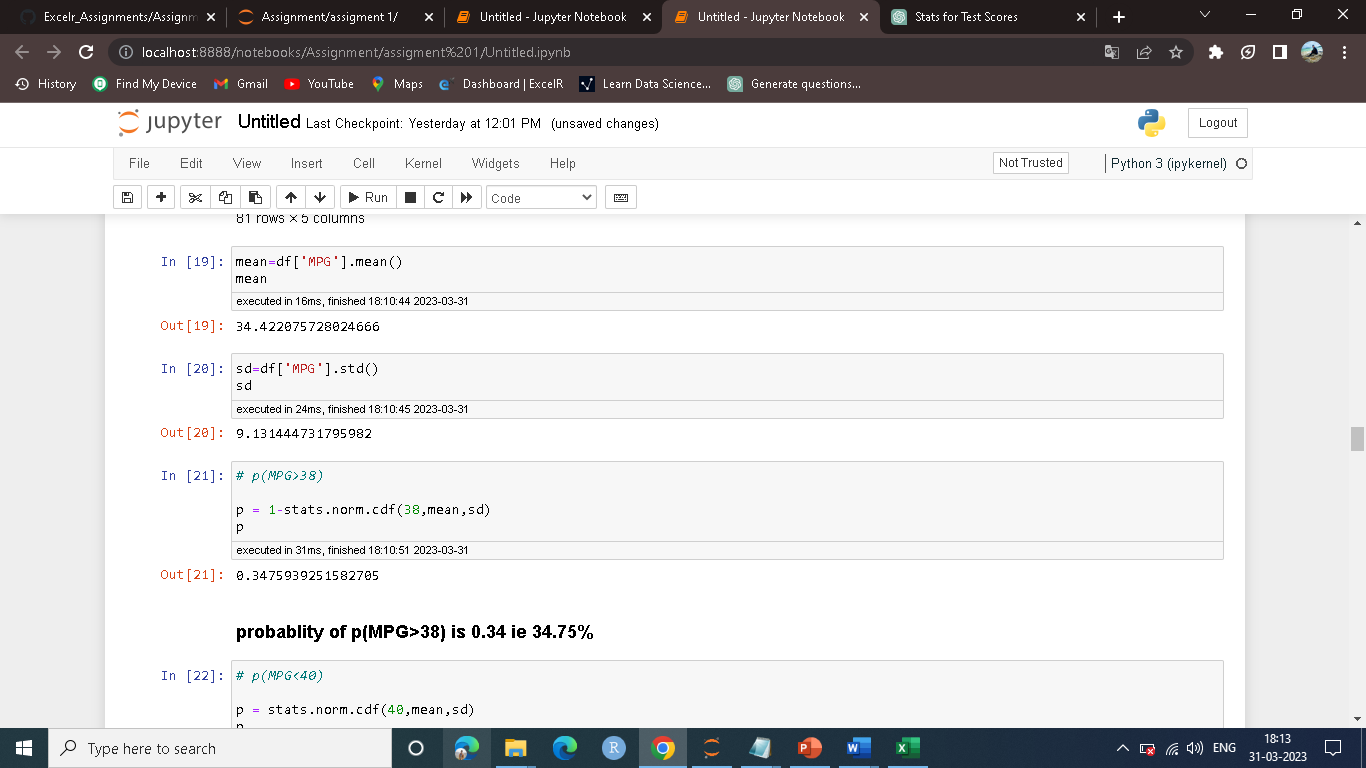
Data \_set: Cars.csv

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

MPG <- Cars$MPG

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

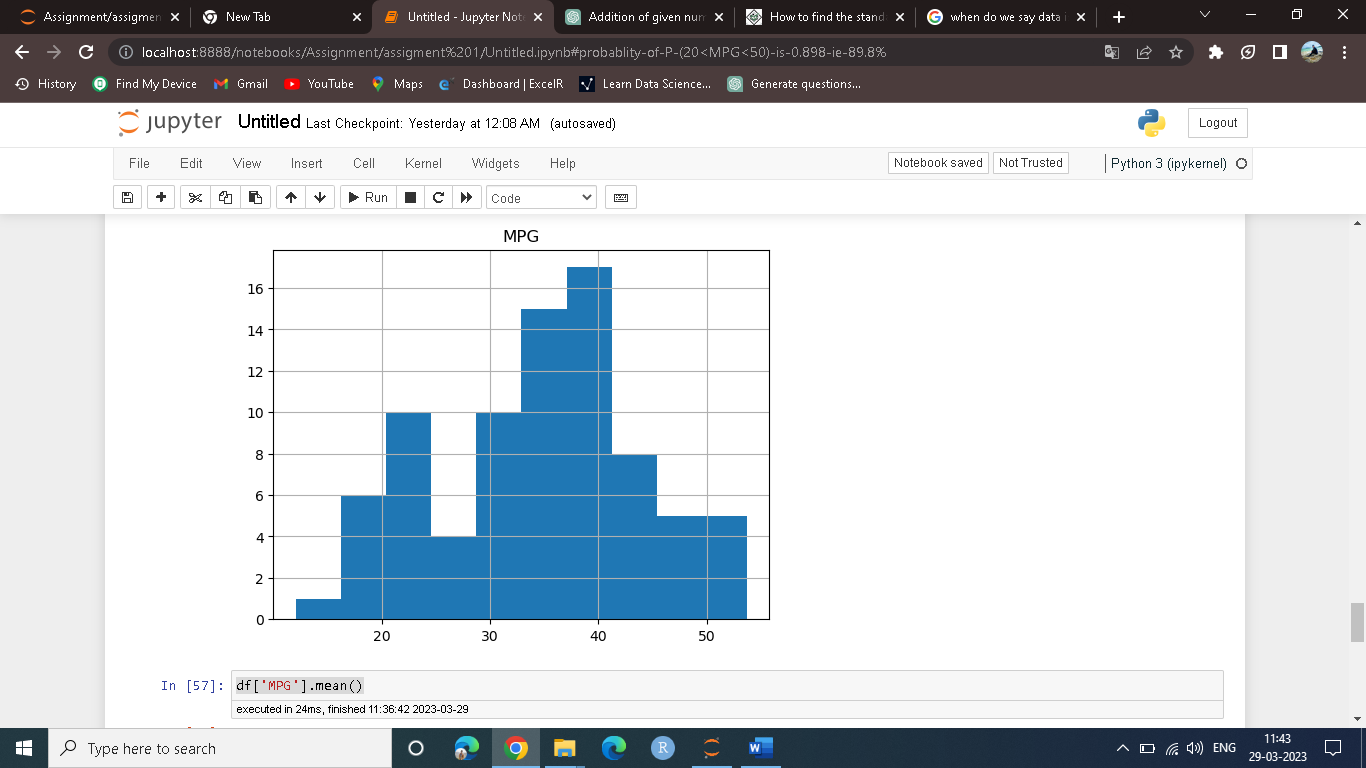
c. P (20<MPG<50)



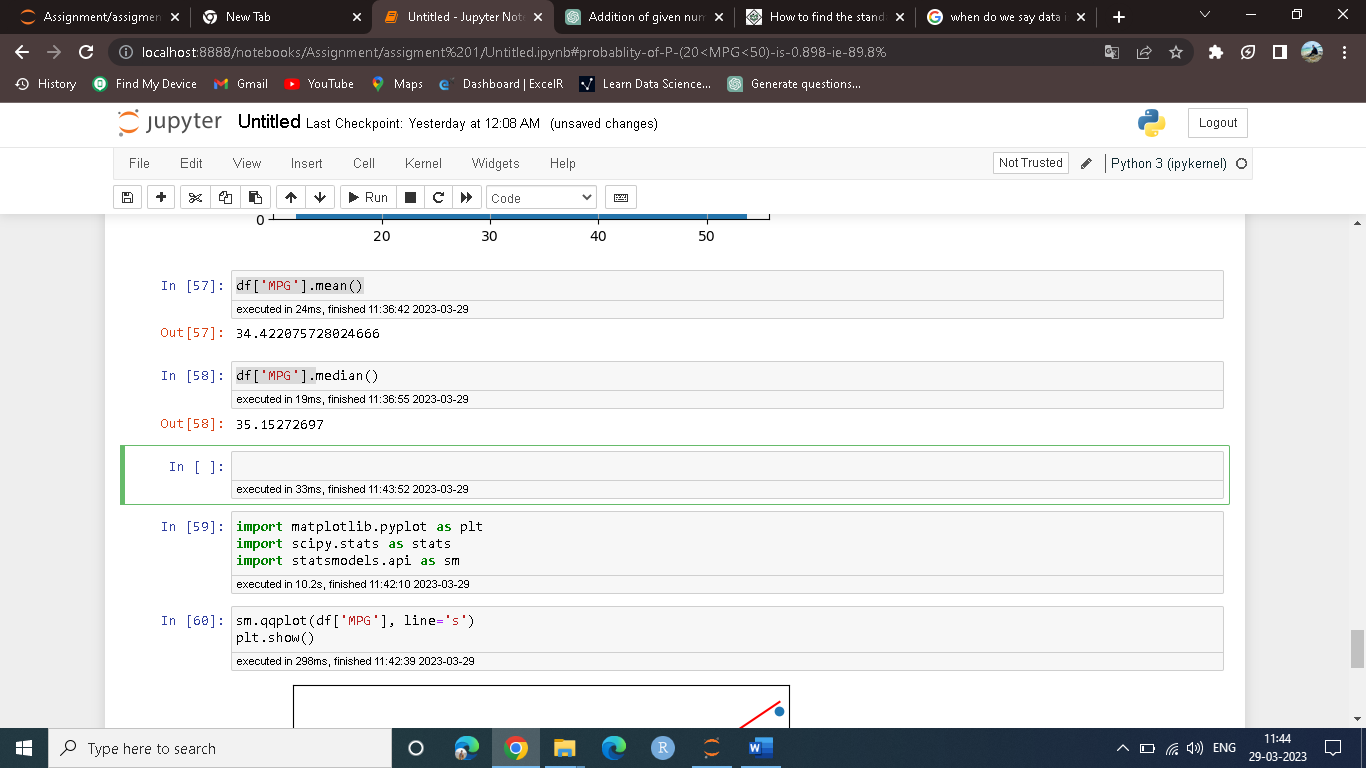
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

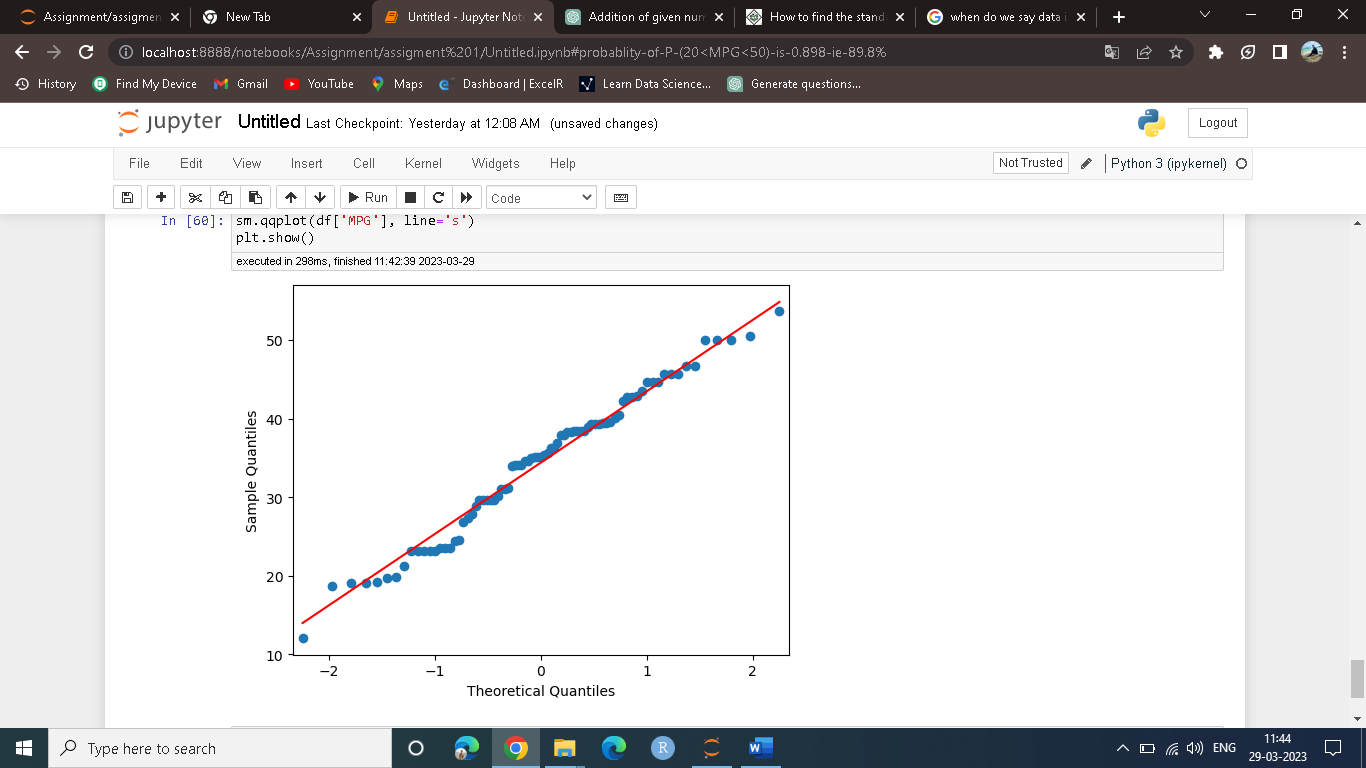
Dataset: Cars.csv



1)From above Histogram we can say that the data is more likely normally distributed. Pdf make a Bell like curve shape.



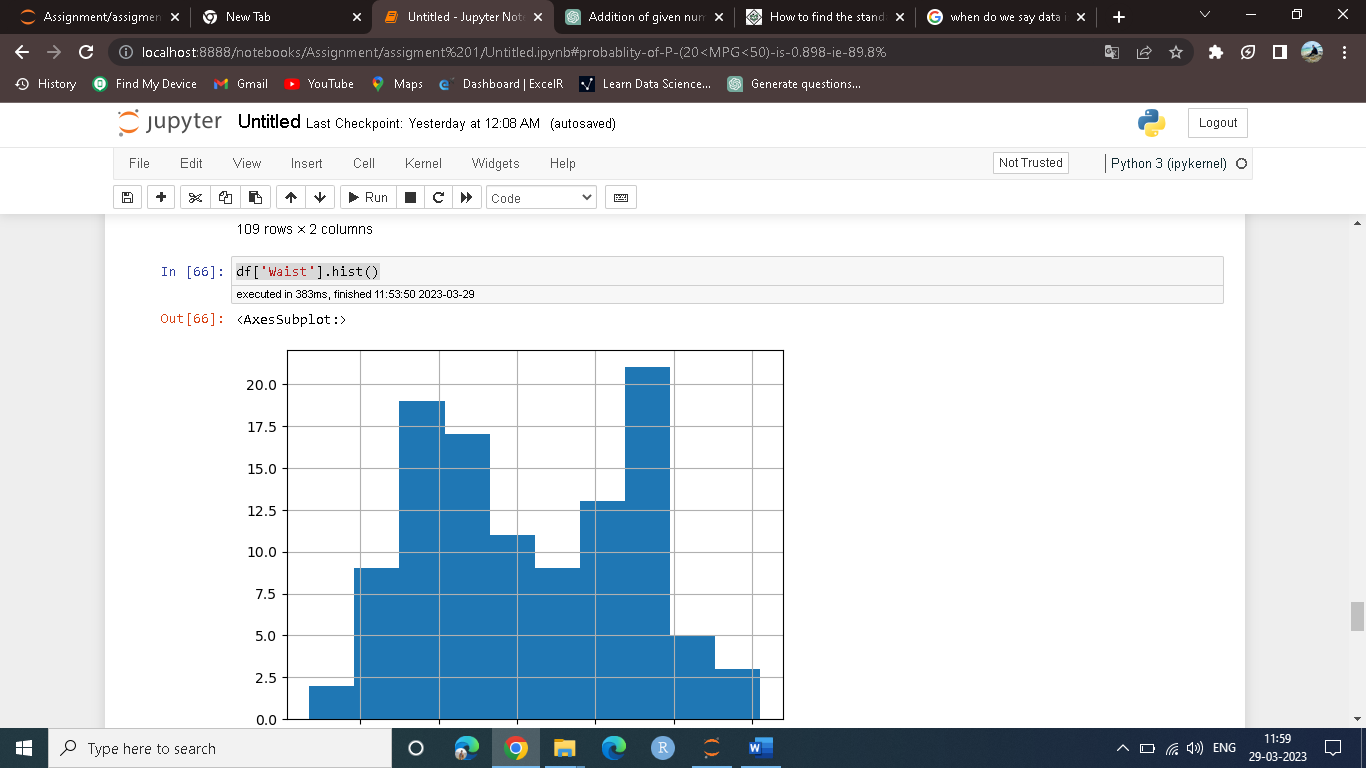
2)Here mean and median value of MPG is appx. Equal then also we say that data is ND.



3)The line in normal probability graph make straight and point are on the line so is ND.

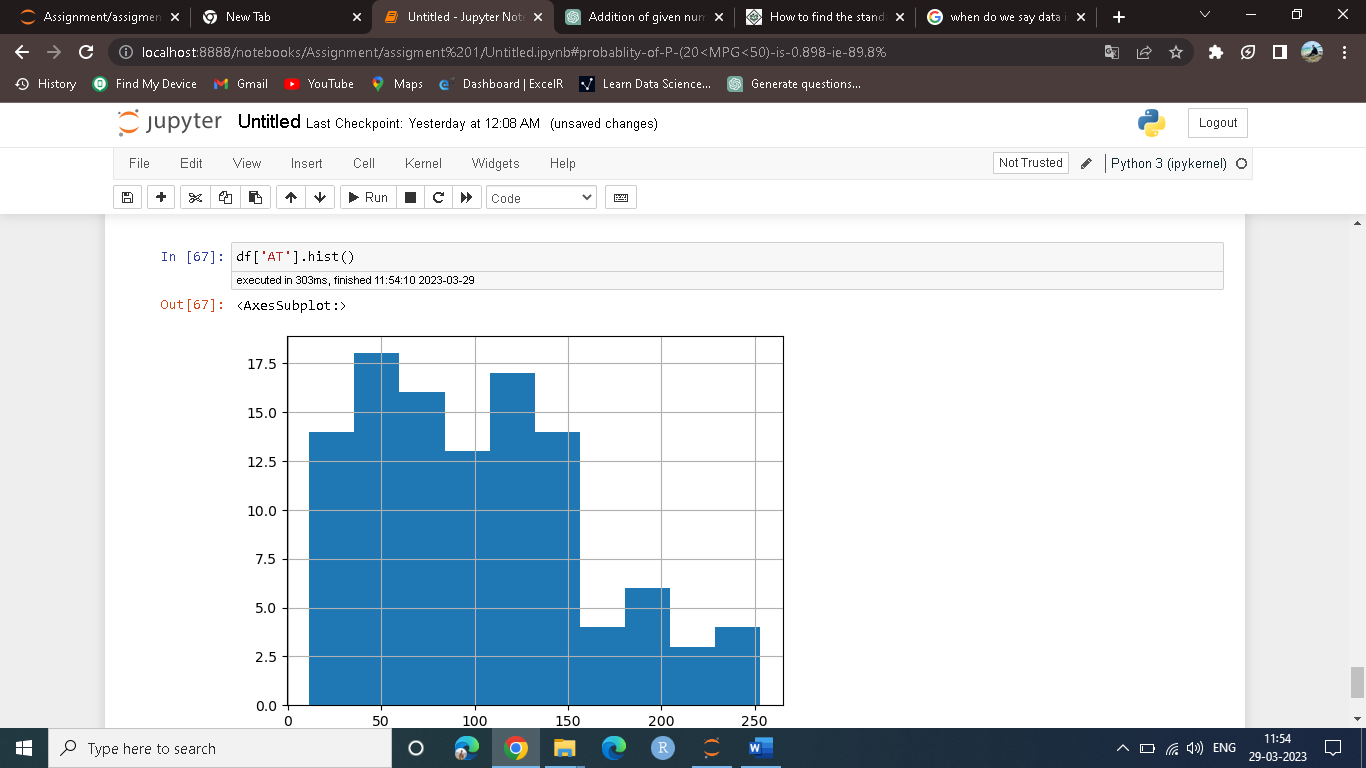
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv



1)For waist histogram is like bell shaped so is like ND.

And mean and median value is also nearly equal.



2)for AT histogram is look like Positive skew so we can not say that the AT is normally distributed.

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

CI=90% Z score-1.64

CI=94% Z score-1.88

CI=60% 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

Ans - CI=95% t score-2.063

CI=96% t score-2.17

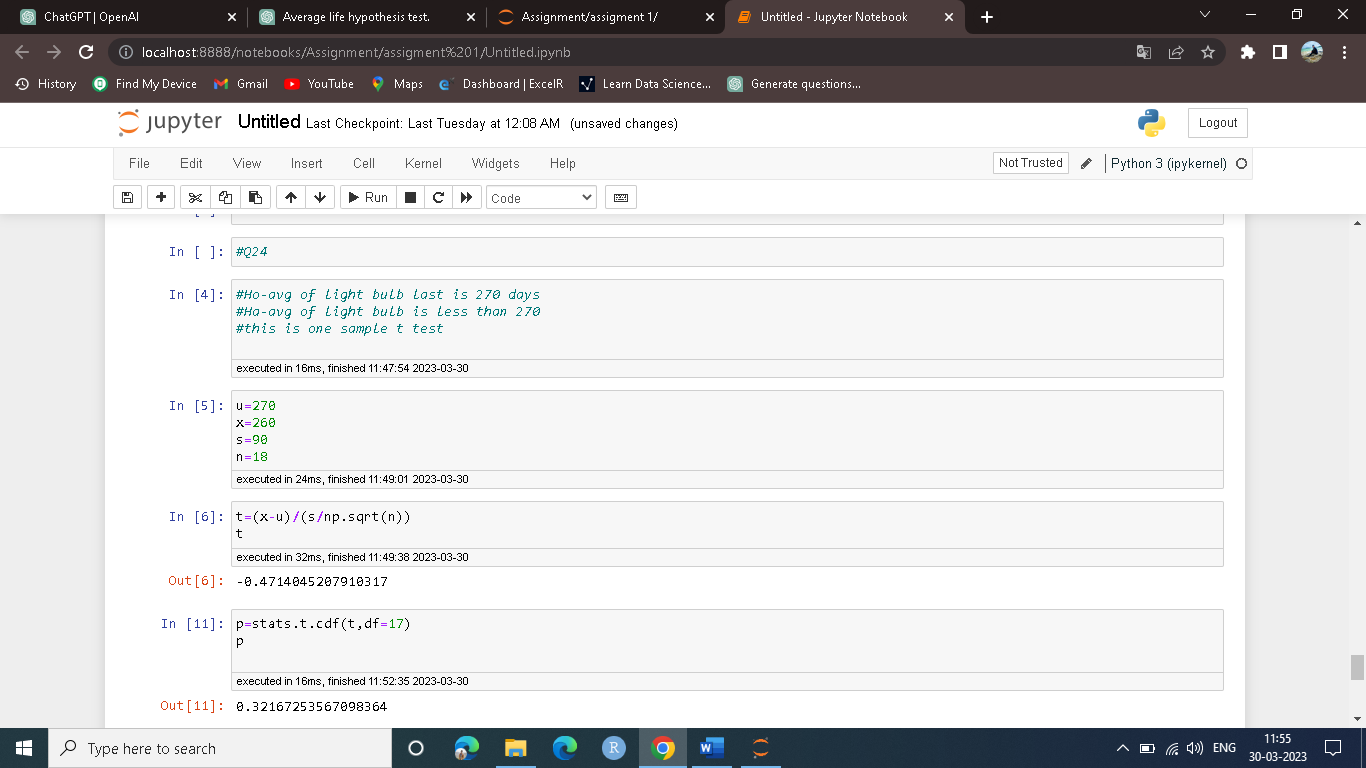
CI=99% t score-2.796

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



-Here p value = 0.32 is greater than alpha 0.05

So, we say that fail to reject null hypothesis.

Those there is no evidence to fail claim of CEO that avg life of light bulb is 270.