|  |  |
| --- | --- |
| 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 | Nominal |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Nominal |

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) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| 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:-** Total Outcomes = 8, Favourable outcomes = 3

Probability = Favourable outcomes/Total Outcomes

= 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 (0.167)**
3. Sum is divisible by 2 and 3 = **5/36 (0.128)**

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 :- **No. of ways to take 2 balls from bag = 7C2 = 21**

**No. of ways to take 2 non-blue balls = 5C2 = 10**

**Now, probability = 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 :- 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**

**Points Score Weigh**

**Mean -**  3.597 3.217 17.849

**Median -**  3.695 3.325 17.71

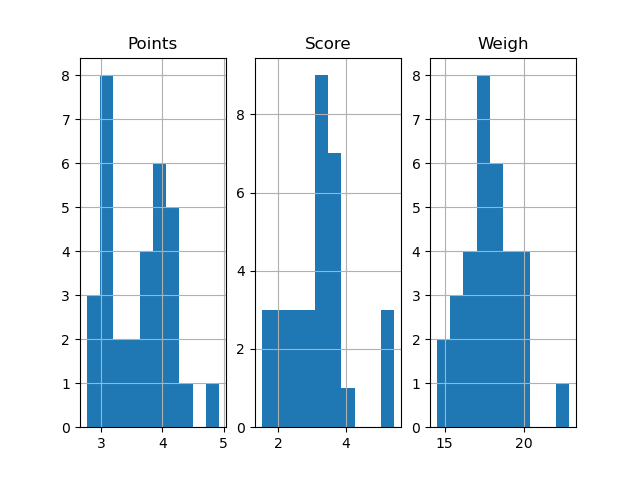
**Mode -**  3.07/3.92 3.44 17.02/18.90

**Standard Deviation -** 0.535 0.978 1.787

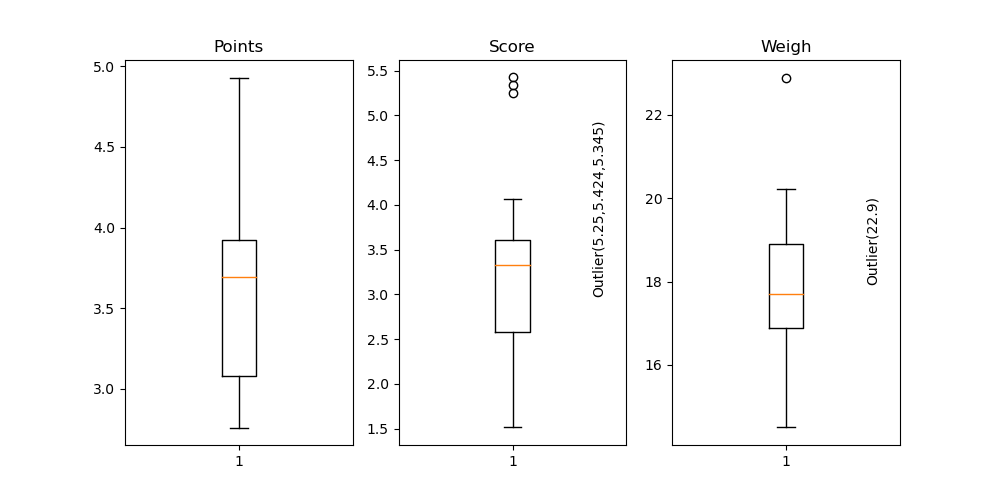
**Variance -**  0.286 0.957 3.193

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

**Range -**  2.17 3.911 8.399

**Histogram Representation of Q7 data:-**

**Box plot representation of Q7 data:-**

****

**Inference :-**

1. **Both Points , Weigh column are Bi-modal in nature.**
2. **Both Score , Weigh column contains outliers (extreme max) means there are some specific cars that scored much higher(5.25,5.424,5.345) than normal distribution, and one car have higher(22.9) weight than the other.**

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:- The Expected Value of the Weight of that patient** = **145.88**

**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**

**Car Speed**  **Distance**

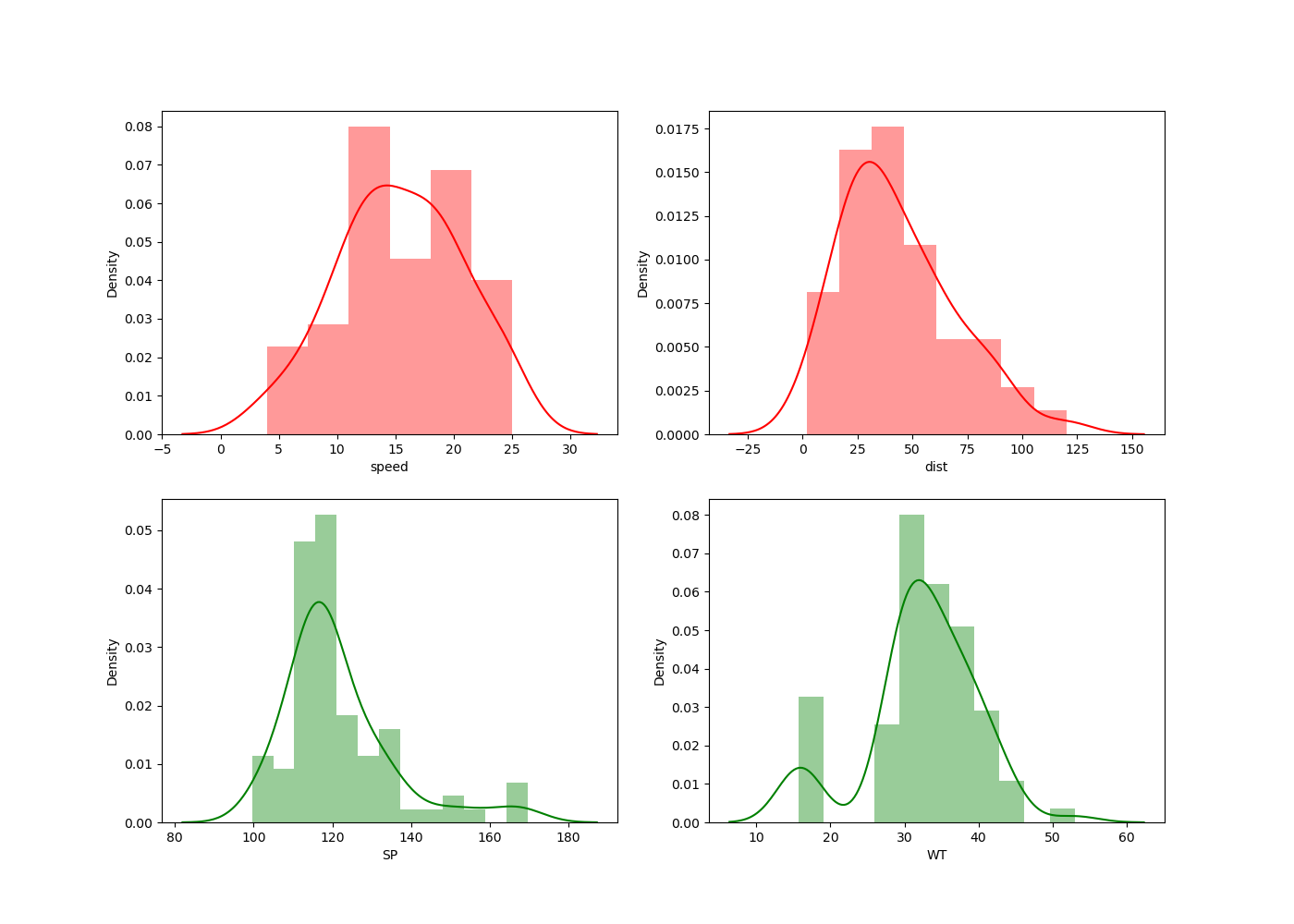
**Skewness :**  - **0.118 0.807**

**Kurtosis : - 0.508 0.405**

**SP**  **WT**

**Skewness :**  **1.611 - 0.615**

**Kurtosis : 2.977 0.95**

****

**Inference :-**

1. **If the distribution is negative skew(left skew) than the mass of that distribution is situated on right side, vice-versa. This implies that distribution is asymmetric in nature.**
2. **Kurtosis refers to “peakedness”, negative kurtosis means the distribution has wider peak and shorter tail. Opposite(positive kurtosis) is also true.**

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



**Ans:** **The distribution is right skew and tail is on right. Mean > Median. We have outliers on the higher side. Most of the mass of the distribution is on left 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?

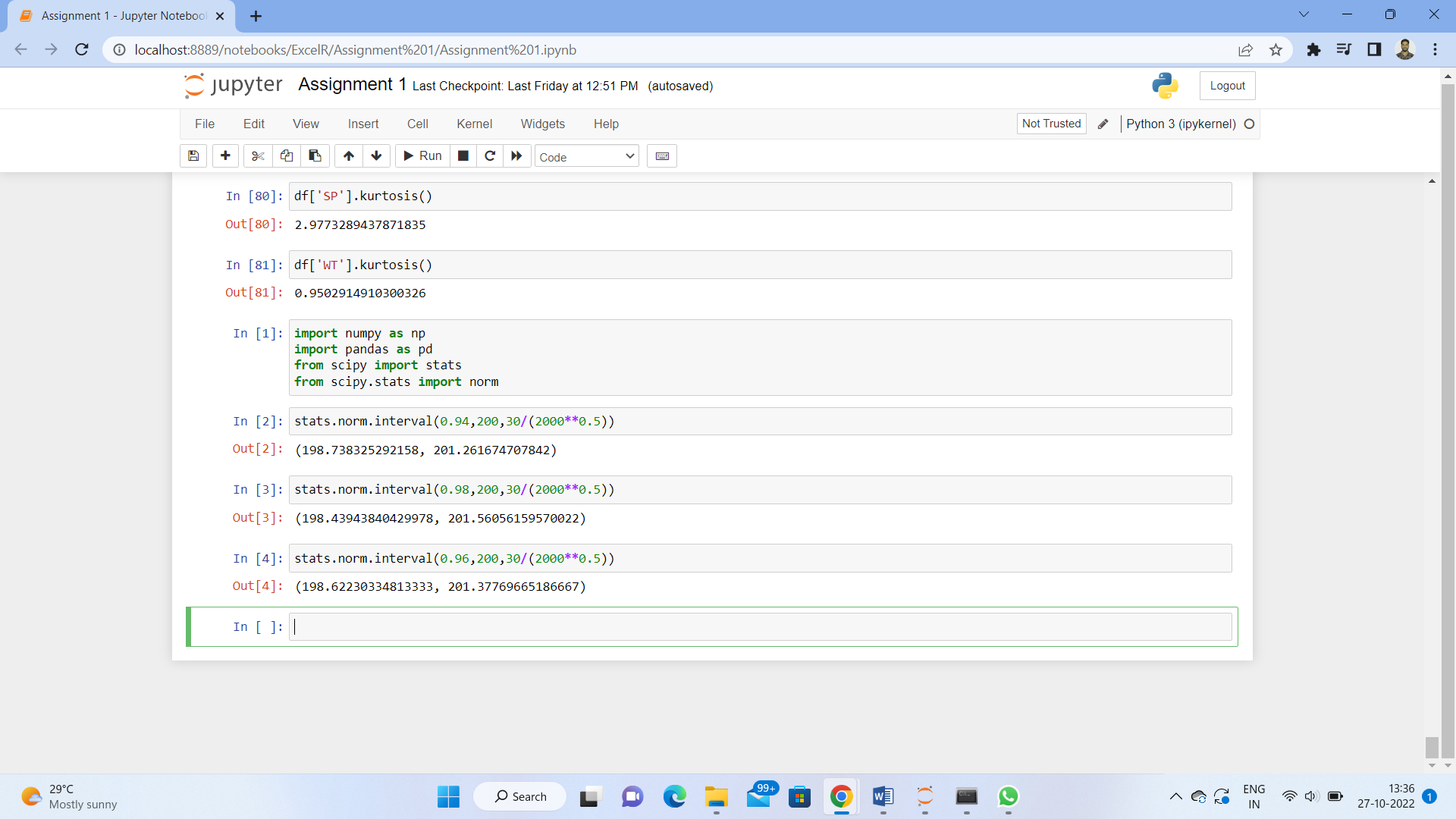
**Solution** =

**Confidence Interval**  **Range**

**94%**  **[198.738 , 201.262]**

**96%** **[198.622 , 201.378]**

**98%**  **[198.439 , 201.561]**



**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.

**Mean = 41**

**Median = 40.5**

**Variance = 25.54**

**Standard Deviation = 5.05**

1. What can we say about the student marks?

**Ans : On an average student get in between 40-42 marks. The distribution contain two oulier (49,56).**

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

**Ans : Data is normalize(Symmetry) and there is no Skewness.**

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

**Ans : The data is skew to the right (positive skew). The mass of the distribution is concentrated on left side of the mean.It has longer tail on right side.**

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

**Ans : The data is skew to the left (negative skew). The mass of the distribution Is concentrated on right side of the mean. It has longer tail 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 : Data is not normally distributed. It has large variation among greater values.**

What is nature of skewness of the data?

**Ans : Data is left skewed as have long whisker on lower side.**

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

**Ans : IQR will be (18 - 10) = 8 (approx)**

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 260 to 270 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

**Mean (MPG) = 34.422 Std(MPG) = 9.131**

* 1. P(MPG>38)

**stats.norm.sf(38, 34.422, 9.131) = 0.348 (approx)**

* 1. P(MPG<40)

**stats.norm.cdf(40, 34.422, 9.131) = 0.729 (approx)**

* 1. P (20<MPG<50)

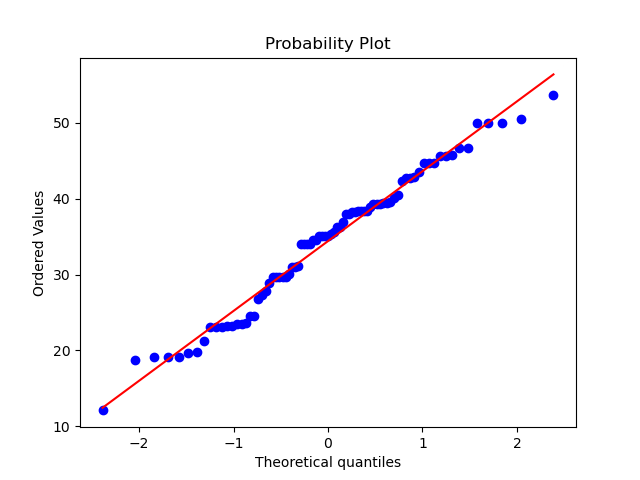
**= stats.norm.cdf(50, 34.422, 9.131) - stats.norm.cdf(20, 34.422, 9.131)**

**= 0.899 (approx)**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

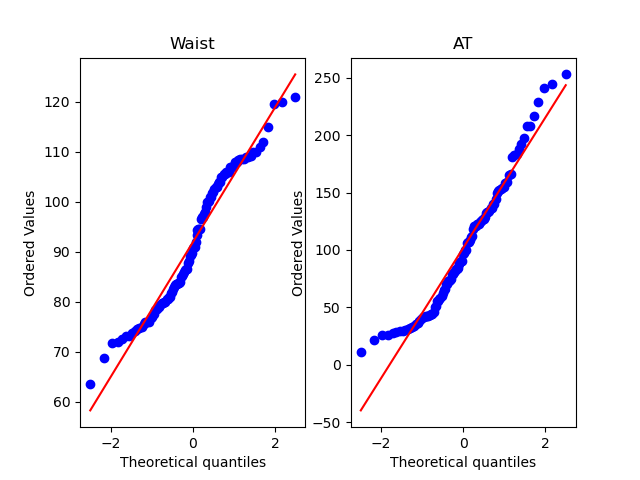
Dataset: Cars.csv



**Ans : From Q-Q Plot, points fall approximately along a straight line, the data follows a normal distribution.**

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

Dataset: wc-at.csv



**From shapiro - wilk Test : p-value for both is less than 0.05 (level)**

1. **value(Waist) = 0.0012 (approx)**

**P-value(AT) = 0.00065 (approx)**

**Conclusion :- from QQ plot and shapiro-wilk test , both distribution**

**does not follow normal distribution.**

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

**Ans :**

**Confidence Interval**  **Z Scores**

**60% - 0.8416**

**90% - 1.6449**

**94% - 1.8808**

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

**Ans : Probability that 18 randomly selected bulbs would have an average life of no more than 260 days.**

**P(x<260) = 0.322 (approx)**