Q1) Identify the Data type for the Following:

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

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	Interval Data
Weight	Ratio Data
Hair Color	Nominal Data
Socioeconomic Status	Ordinal Data
Fahrenheit Temperature	Interval Data
Height	Ratio Data
Type of living accommodation	Nominal Data
Level of Agreement	Ordinal Data
IQ(Intelligence Scale)	Interval 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	Ordinal Data
Religious Preference	Nominal Data

Barometer Pressure	Interval 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.3/8

Explanation: – (HHH, HHT, HTH, THH, TTH, THT, HTT, TTT)

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

 a) Equal to 1
 b) Less than or equal to 4
 c) Sum is divisible by 2 and 3

Ans. (a) 0

Explanation–

 ➢ Total possible outcomes $= 6^2 = 36$
 ➢ A favorable outcomes (sum equal to 1) $= 0$ (i.e. not possible that sum always exceed to 1)
 ➢ Required probability $= 0/36 = 0$

Ans. (b) 1/6

Explanation–

 ➢ Favorable outcome (sum equal to 4) $=3$ [i.e.(1,3) (2,2) (3,1) (1,1) (1,2) (2,1)]
   $= 6$
 ➢ Total Outcome $= 36$
 ➢ Probability = 6/36
           $=1/6$

Ans. 5/36

Explanation–

> ➢ Favorable outcomes = (1,5), (3,3) (4,2) (5,1) (6,6) = 5
> ➢ Total outcomes = 36
> ➢ Probability = 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

Explanation–

> ➢ Total number of balls in the bag = (2+3+2) = 7 Balls
> ➢ Ans Finding two balls out of 7 balls is $^7c_2$
> ➢ ^nc_r = n / r(n-r) = $^7c_2$ = 7 / 2 (7–2) = 7*6*5 / 2*5 = 21
> ➢ Total number of values of blue balls = 7–2 = 5
> ➢ Using that formula, 5 / 2 (5–2)
> ➢ 5*4*3 / 2*3 = 10
> ➢ Probability = FO/TO = 21/10

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 is 3.09

Explanation–

➢ 1*0.015+4*0.20+3*0.65+5*0.005+6*0.01+2*0.120
➢ 0.015+0.80+1.95+0.025+0.06+0.24
  3.090

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

Solution:–

1) **Points**
   Mean = 3.596
   Median = 3.695
   Mode = 3.07
   Variance = 0.285
   Standard Deviation = 0.534
   Range = 2.17

2) **Score**
   Mean = 3.217

Median = 3.325

Mode = 3.44

Variance = 0.957379

Standard Deviation = 0.978457

Range = 3.911

3) Weigh.
   Mean = 17.848
   Median = 17.71

   Mode = 17.02

   Variance = 3.193166

   Standard Deviation = 1.786943

   Range = 8.4

➢ **Assumes**:–
➢ 1)The given data of points, scores, weights are shifted towards left and its tail on the right side.
➢ 2)It is a positively/right-skewed distribution.
   3) Points dataset is a bimodal because it is having two modes and score, weight are unimodal because it having single mode

Q8) Calculate Expected Value for the problem below

   a) 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. Expected Value of the Weight of that patient = 145.33

Explanation.

➢ Expected Value = $\sum$ (probability * Value) = $\sum P(x) * E(x)$

there are 9 patients

- ➢ Probability of selecting each patient P(x) = 1/9
- ➢ E(x) = {108, 110, 123, 134, 135, 145, 167, 187, 199}
- ➢ 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

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**

A) Index, Speed and Distance
- ➢ Skewness of Index is: 0.0
- ➢ Skewness of Speed is: -0.11395477012828319
- ➢ Skewness of Distance is: 0.7824835173114966
- ➢ Kurtosis of Index is: -1.2009603841536614
- ➢ Kurtosis of Speed is: -0.5771474239437371
- ➢ Kurtosis of Distance is: 0.24801865717051808

Inference:

The given data of speed are shifted towards right and its tail on the left side, it is a negatively/left-skewed distribution
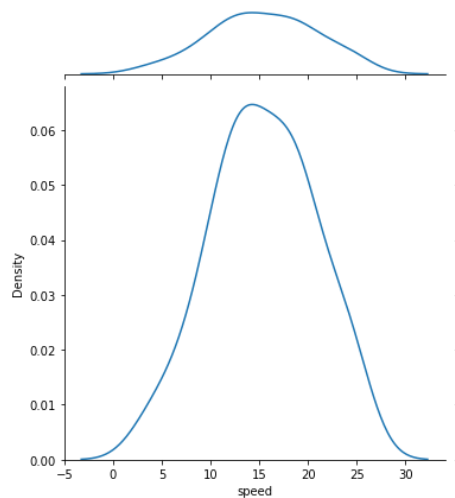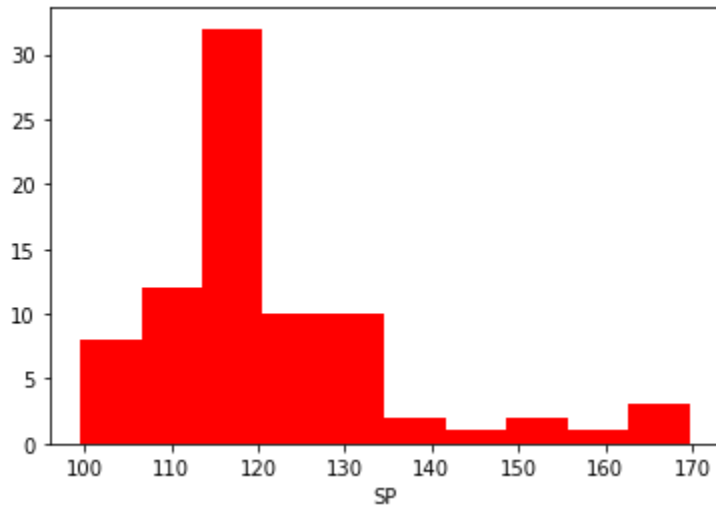
The given data of dist are shifted towards left and its tail on the right side, it is a po sitively/right-skewed distribution.
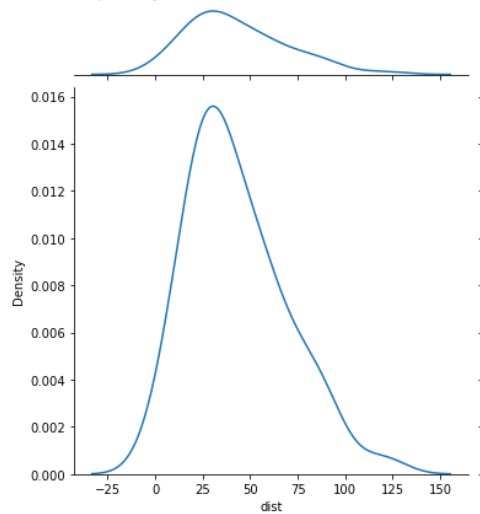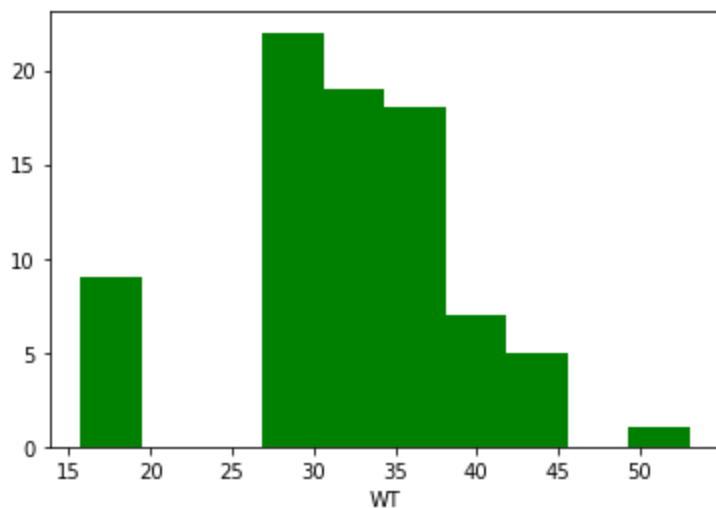

B) SP and Weight (WT)
  ➢ Skewness of SP is: 1.5814536794423764
  ➢ Skewness of Weight is: -0.6033099322115126
  ➢ Kurtosis of SP is: 2.7235214865269244
 ➢ Kurtosis of Weight is: 0.8194658792266849

Inference:

Most data points of SP are shifted towards left and its tail on the right side, it is a positively/right-skewed distribution

Most data points of WT are shifted towards right and its tail on the left side, it is a negatively/left-skewed distribution.

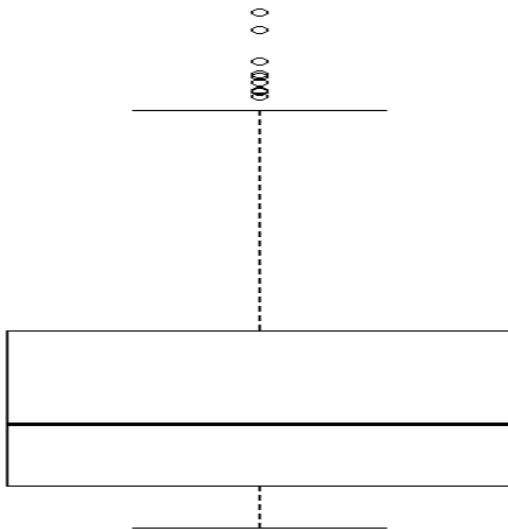**Q10) Draw inferences about the following boxplot & histogram**

Histogram of ChickWeight$weight

ChickWeight$weight
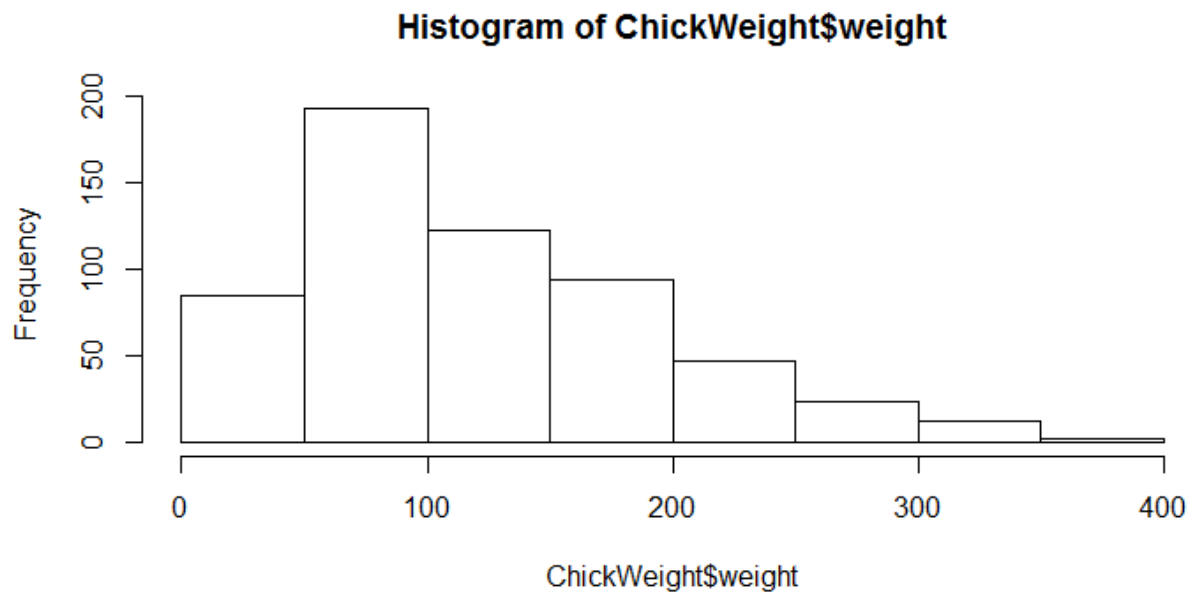
Explanation:–

Histogram:–

- ➤ Chick weight data is right skewed /positively skewed.
- ➤ More than 50% Chick Weight is between 50 to 150.

Most of the chick weight is between 50 to 100.

Boxplot:–

> ➢ The data is right skewed/positively skewed.
There are outliers 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?

Explanation :–

> ➢ **Confidence Interval for 94% is: (198.738, 201.261).**
> ➢ **Confidence Interval for 98% is: (198.439, 201.560).**
> ➢ **Confidence Interval for 96% is: (198.622, 201.376).**

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?

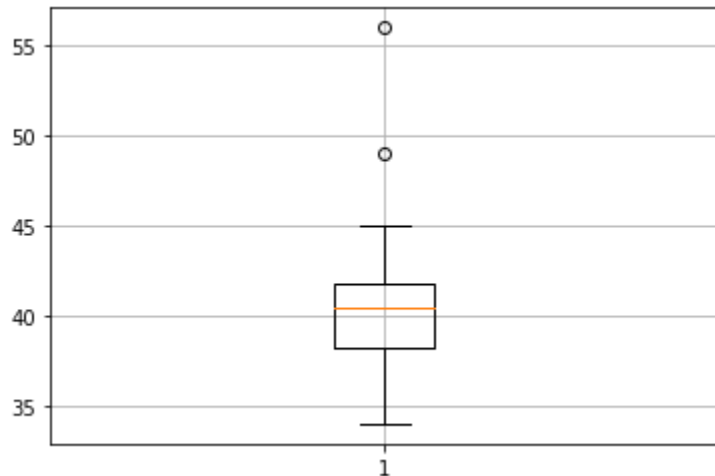**Ans.** Mean=41.0, Median= 40.5, Variance= 24.111,Std Dev = 4.91

Explanation:–

1)

> ➢ Mean =
> 34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56 / 18
> $= 738/18 = 41.0$

> ➢ Median = (40+41)/2
> $= 81/2 = 40.5$

> Variance $= (49+25+25+9+9+4+4+1+1+0+0+0+0+1+1+16+64+225)/18$
> $= 434/18 = 24.111$

Standard Deviation = 4.910

From above plot we can say that mean of marks of student is 41 which is slightly greater than median. Most of the students got marks in between 40-42, there are two outlier 49,56.

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

Ans. When mean, median of data are equal there is no skewness. The distribution is both symmetric and unimode

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

Ans. If the mean is greater than the median, the distribution is positively skewed.

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

**Ans**. If the mean is less than the median, the distribution is negatively skewed.

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

Ans. Positive values of kurtosis indicates that distribution is peaked and possesses thick tails.

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

Ans. Negative values of kurtosis indicates that a distribution is flat and has thin tails.

Q18) Answer the below questions using the below boxplot visualization.
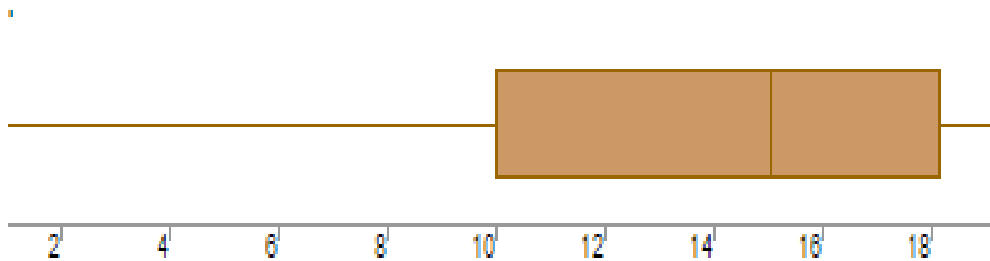
What can we say about the distribution of the data?

Ans. The peak of the given boxplot or the most of data points are situated towards right side and the tail is at left side of boxplot.

What is nature of skewness of the data?

Ans. Negative Skewness.

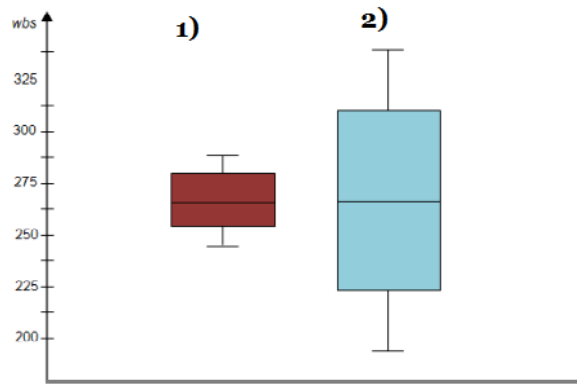What will be the IQR of the data (approximately)?

Ans. 8.

Explanation :–

IQR = upper quartile – lower quartile

= 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.

➢ By observing both the boxplots there is no skewness but whisker level is high in boxplot 2 as compared to boxplot 1.
➢ The IQR of the first boxplot is between 255 to 280.
➢ The IQR of the second boxplot is between 225 to 315.

And there is no outliers in both the boxplots

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

a. P(MPG>38)
b. P(MPG<40)
c. P (20<MPG<50)

Ans.

- ➢ Probability for MPG>38 is: 0.3475939251582705.
- ➢ Probability for MPG<40 is: 0.7293498762151616.
- ➢ Probability for 20<MPG<50 is: -0.013116469610523339.

Q 21) Check whether the data follows normal distribution
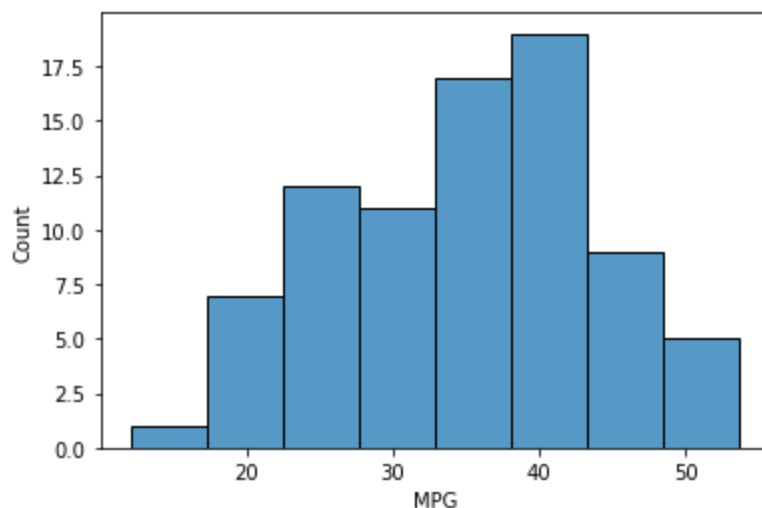 a) Check whether the MPG of Cars follows Normal Distribution
      Dataset: Cars.csv

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

Explanation:–

The data set is not a normal distribution, it is right skewed / positively skewed

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

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

     Dataset: wc-at.csv

Explanation:–
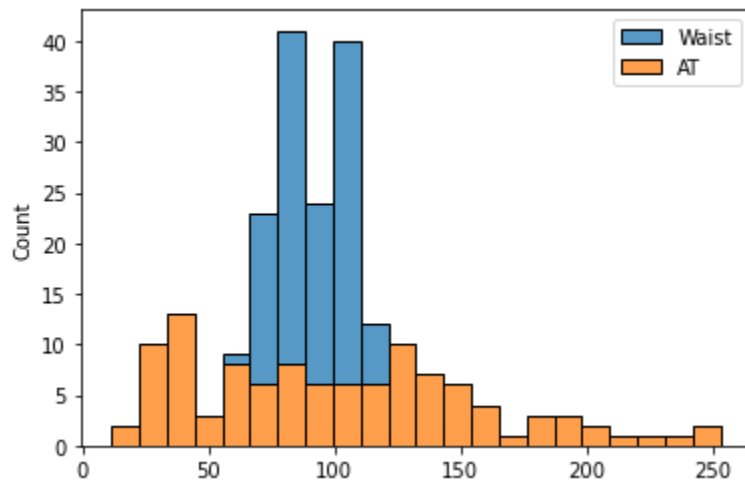
Both the Adipose Tissue (AT) and Waist Circumference (Waist) data set do not follow the normal distribution approximately (as mean and median of both the data are approximately different)

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

Ans.

- ➢ 90% confidence interval Z-score: 1.64
- ➢ 94% confidence interval Z-score: 1.55
- ➢ 60% confidence interval Z-score: 0.25

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

Ans.

> ➢ 95% confidence interval t-score: 2.06
> ➢ 96% confidence interval t-score: 2.17
> ➢ 99% confidence interval t-score: 2.8

Q 24)   A Government company claims that an average light bulb lasts <mark>270</mark> days. A researcher randomly selects <mark>18</mark> bulbs for testing. The sampled bulbs last an average of <mark>260</mark> days, with a standard deviation of <mark>90</mark> days. If the CEO's claim were true, what is the probability that <mark>18</mark> randomly selected bulbs would have an average life of no more than <mark>260</mark> days

Hint:

  rcode → pt(tscore,df)

 df → degrees of freedom

Ans.

> ➢ T_Score is -0.4714045207910317.

P_value is: 0.32167411684460556.