|  |  |
| --- | --- |
| 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 | Continuous |
| Number of times married | Discrete |
| Gender (Male or Female) | Continuous |

Q1) Identify the Data type for the Following:

Subject :

Discrete Data Type: Numerical type of data that includes whole, concrete numbers with specific and fixed data values determined by counting.

Continuous Data Type: data that can be measured on an infinite scale, It can take any value between two numbers.

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

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

Answer:

Total number of possible combinations = 23 = 8

The combinations are: HHH, HHT, HTH, THH, TTH, THT, HTT, TTT.

Number of combinations that have two heads and one tail = 3, i.e., HHT, HTH, TTH

Probability (two heads and one tail)= Number of possible outcomes

i.e., 3/8= 0.375

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 2and 3

Answer:

Total possible outcome =62 i.e., 36

1. Sum equal to 1**= 0**

*Minimum possible sum= (1,1)=2*

*P(1)= 0/36=0*

b) Sum is less than or equals to 4: **= 3** *i.e. (1,3)(2,2)(3,1)*

probability =3/36 =1/12

Correction: 6/36 =1/6

c) Sum is divisible by 2and 3

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

**The probability of an event is given as:**

**Probability = number of favorable outcomes**

**number of possible outcomes**

Thus the probability that sum is divisible by 2 and 3 is  **5/36 = 0.138**

Correction: 6/36 =1/6

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:

we use the combination formula, as follows.

nCr = n! ­­­

r! (n-r)!

! represents a **factorial.**

Total no.of balls = 2+3+2 i.e., 7

n(S) = Total no of ways to draw 2 balls out of 7

S= sample space

Replace the above formula: n= 7 and r=2

7C2 = 7 x 6 x 5 x 4 x 3 x 2 x 1

2 x 1 (7-2 )!

7C2 = 7 x 6 x 5 x 4 x 3 x 2 x 1

2 x 1 (5 x 4 x 3 x 2 x 1 )

7C2 = 5040 = 21

240

Therefore , let “E” is the event of drawing 2 balls, none of them are blue

n(E) =Total no of ways to draw 2 balls out of 5 (i.e., 2 Red and 3 Green) n=5, r=2.

5C2 = 5 x 4 x 3 x 2 x 1

2 x 1 (5-2 )!

5C2 = 5 x 4 x 3 x 2 x 1

2 x 1 (3 x 2 x 1 )

5C2 = 120 = 10

12

P(E) = n(E) = **10**  = **0.476**

N(S) **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

=  1x0.015  + 4x0.20  + 3x0.65  + 5x0.005  + 6x0.01  + 2 x0.12

= 0.015 + 0.8 +1.95 + 0.025 + 0.06 + 0.24

=       **3.09**

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

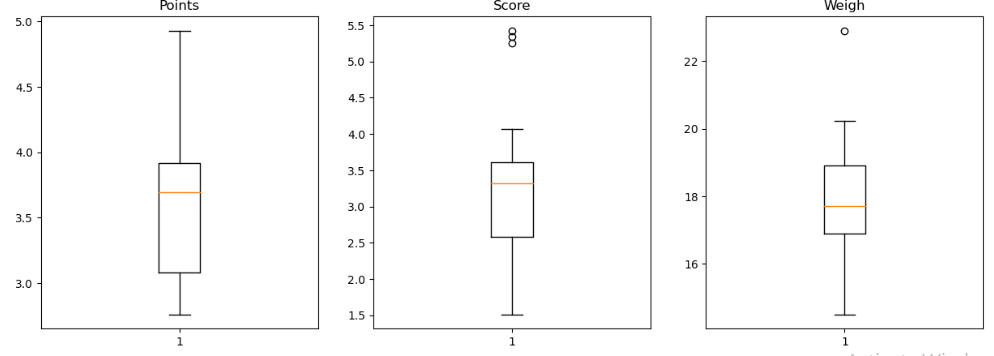
* For Points,Score,Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

Answer: Solved using Jupyter Note Book (Python)- File attached

|  |  |  |  |
| --- | --- | --- | --- |
| Statistics Parameter | Points | Score | Weigh |
| Mean | 3.596563 | 3.217250 | 17.848750 |
| Median | 3.695 | 3.325 | 17.710 |
| Mode | 0 3.07  1 3.92 | 0 3.44 | 0 17.02  1 18.90 |
| Variance | 0.285881 | 0.957379 | 3.193166 |
| Standard Deviation | 0.534679 | 0.978457 | 1.786943 |
| Range | 2.17 | 3.91 | 8.4 |



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:

Mean value of the weights shall be considered for the expected value of the weight of the patient, assuming that one of the patients is chosen at random

Manual

Sum= 108 + 110 + 123 +134+ 135+ 145+ 167+ 187+ 199 = 1308

Total Weights = 9

Mean= 1308 /9

Mean= 145.3

Python Code

In [ ]#Import python libraries

import statistics

In [ ]print(statistics.mean([108, 110, 123, 134, 135, 145, 167, 187, 199]))

145.33333333333334

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

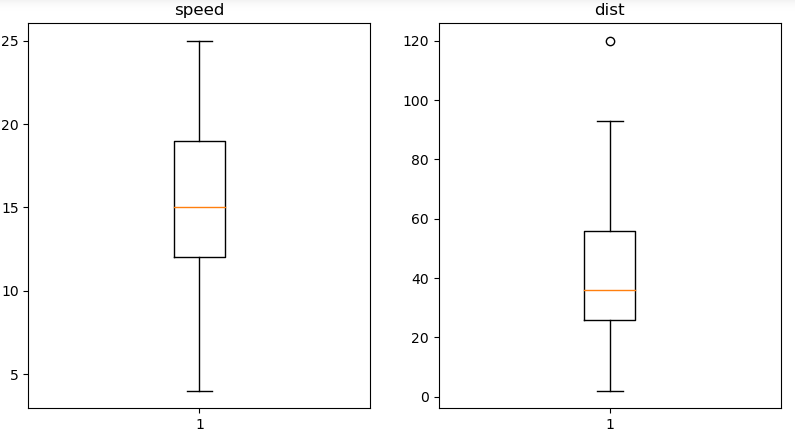
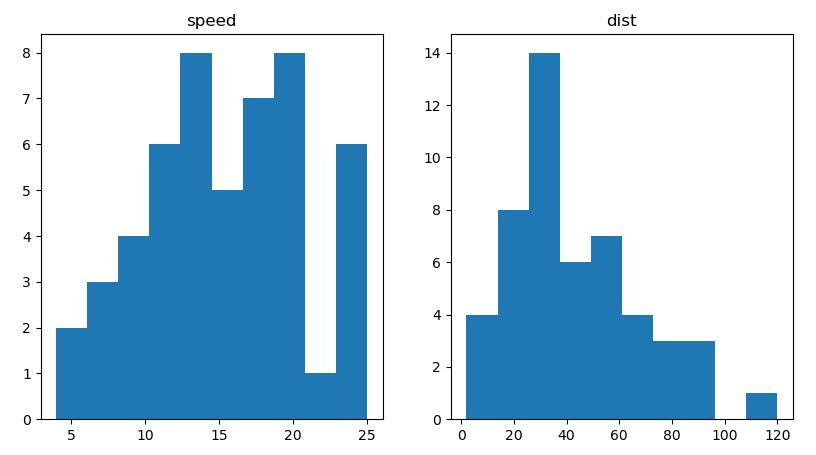
**Cars speed and distance**

**Use Q9\_a.csv**

Answer: Solved using Jupyter Note Book (Python)- File attached

|  |  |  |
| --- | --- | --- |
|  | **Car Speed** | **Distance** |
| **Skewness** | -0.117510 | 0.806895 |
| **kurtosis** | -0.508994 | 0.405053 |

**Histogram Box Plot**

****

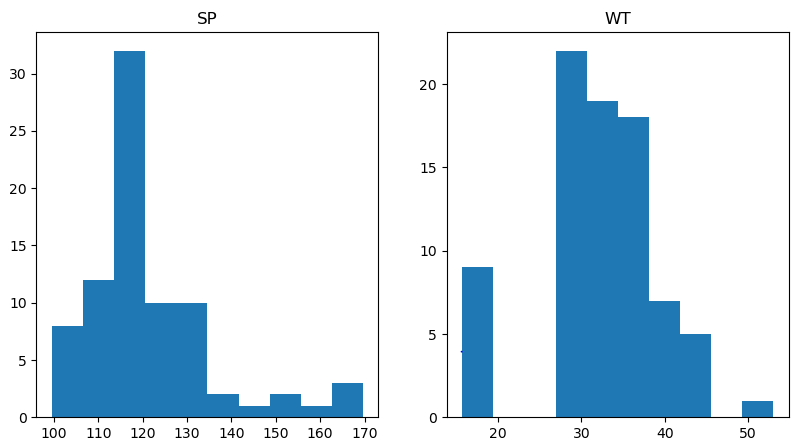
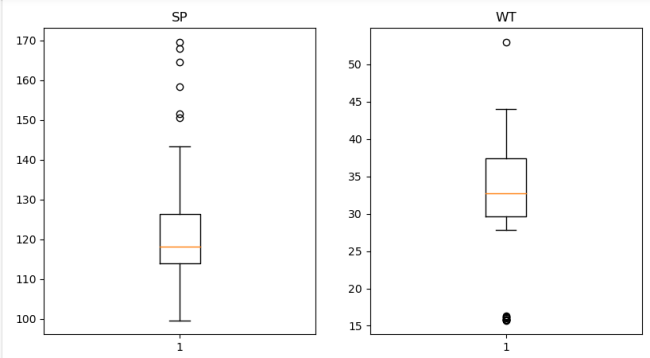
**SP and Weight(WT)**

**Use Q9\_b.csv**

Answer: Solved using Jupyter Note Book (Python)- File attached

|  |  |  |
| --- | --- | --- |
|  | **SP** | **Weight (WT)** |
| **Skewness** | 1.611450 | -0.614753 |
| **kurtosis** | 2.977329 | 0.950291 |

**Histogram Box Plot**

** **

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



Answer:

Chick weight data is right skewed or positively skewed.

More than 50% Chick Weight is between 50 to 150.

Most of the chick weight is between 50 to 100.



Answer:

data is right skewed or positively skewed (outliers on top)

**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?

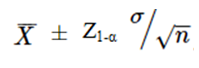
Answer:

Sample Population = 2,000

Total Population = 3,000,000

Average weight i.e., sample mean = 200 Pounds

Standard deviation = 30 pounds



= Mean of sample (200)

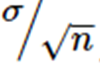


σ = standard deviation(30)

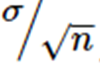
n=Sample size/no.of data points (observation)(2000)

1-α = confidence level(94%, 98% , 96%)

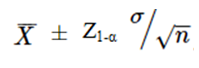
Z= Z score or Z table

= 30/(2000)

= 30/44.72

=0.6708

Manual Calculation:



|  |  |  |  |
| --- | --- | --- | --- |
|  | 94% | 98% | 96% |
| Z values 94%, 98%, 96%  Py Code: Stats.norm.ppf(CIi.e., 0.97, 0.99, 0.98) | 1.88 | 2.32 | 2.05 |
| Lower Limit | 200-1.88\* 0.6708  =200-1.261104  =198.7389 | 200-2.05\* 0.6708  =200-1.556256  =198.4437 | 200-2.05\* 0.6708  =200+1.37514  = 198.6249 |
| Upper Limii | 200+1.88\* 0.6708  =200+1.261104  =201.2611 | 200+2.05\* 0.6708  =200+1.556256  =201.5563 | 200+2.05\* 0.6708  =200+1.37514  =201.3751 |

Python Codes: **Solved using Jupyter Note Book (Python)- File attached**

*# Avg. weight of Adult in Mexico with 94% CI*

stats**.**norm**.**interval(0.94,200,30**/**(2000**\*\***0.5))

*(*(198.738325292158, 201.261674707842)

*# Avg. weight of Adult in Mexico with 98% CI*

stats**.**norm**.**interval(0.98,200,30**/**(2000**\*\***0.5))

(198.43943840429978, 201.56056159570022)

*# Avg. weight of Adult in Mexico with 96% CI*

stats**.**norm**.**interval(0.96,200,30**/**(2000**\*\***0.5))

(198.62230334813333, 201.37769665186667)

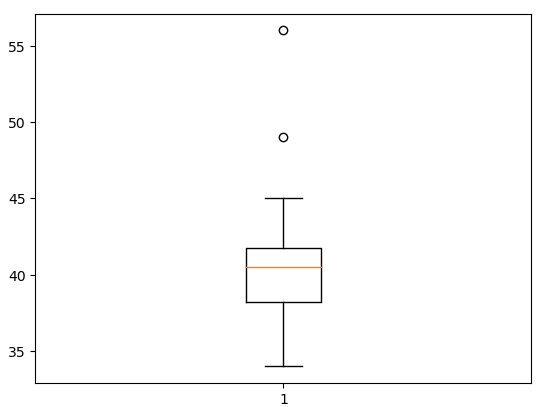
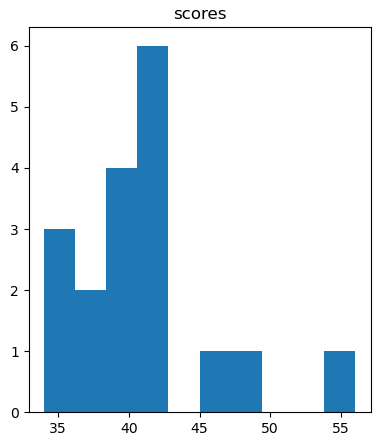
**Q12)**Below are the scores obtained by a student in tests

**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?

Answer: Solved using Jupyter Note Book (Python)- File attached

|  |  |  |  |
| --- | --- | --- | --- |
| mean | median | variance | standard deviation |
| 41.0 | 40.5 | 25.53 | 5.05 |

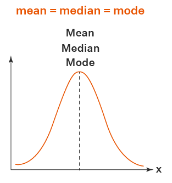
 

Maximum number of students marks between 38 -42.

Skewnessis positive because mass of marks in left side of plot

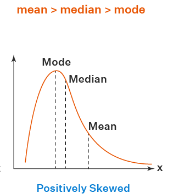
Q13) What is the nature of skewness when mean, median of data are equal?

Answer: Symmetric



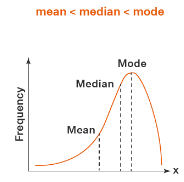
Q14) What is the nature of skewness when mean >median ?

Answer: Positively Skewed



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

Answer: Negatively Skewed

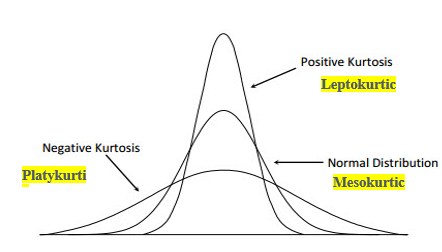


Q16) What does positive kurtosis value indicates for adata ?

Answer: Positive values of kurtosis indicate that distribution is peaked and possesses thick tails. An extreme positive kurtosis indicates a distribution where more of the numbers are located in the tails of the distribution instead of around the mean.

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

Answer: A distribution with a negative kurtosis value indicates that the distribution has lighter tails than the normal distribution.



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



What can we say about the distribution of the data?

Answer: It is not the Normal Distribution

What is nature of skewness of the data?

Answer:Left side Skewed or Negative Skewness

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

Median= 15

Upper Quartile (UQ)= 18

Lower Quartile(LQ)=10

IQR = UQ-LQ

IQR= 8

Q19) Comment on the below Boxplot visualizations?



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

Answer:

There are no outliers.

Median is same for both the box plots i.e., approximately in a range between 275 to 250.

Distribution is normal/symmetric.

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

Data \_set: Cars.csv

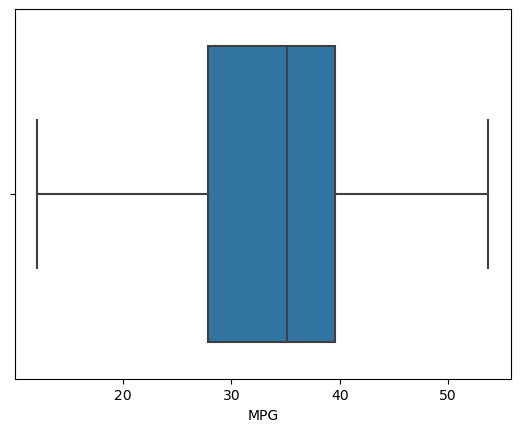
Calculate the probability of MPG ofCars for the below cases.

MPG<- Cars$MPG

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

c. P (20<MPG<50)

Answer: **Solved using Jupyter Note Book (Python)- File attached**



1. P(MPG>38)

1-stats.norm.cdf(38,cars.MPG.mean(),cars.MPG.std())

Ans: 0.3475939251582705

1. P(MPG<40)

stats.norm.cdf(40,cars.MPG.mean(),cars.MPG.std())

Ans: 0.7293498762151616

c. P (20<MPG<50)

stats.norm.cdf(0.50,cars.MPG.mean(),cars.MPG.std())-stats.norm.cdf(0.20,cars.MPG.mean(),cars.MPG.std())

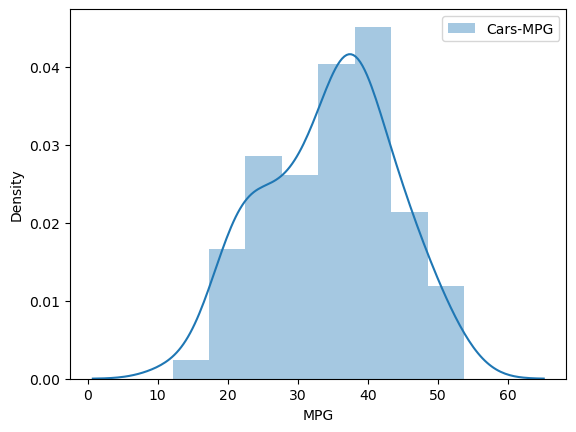
Ans: 1.2430968797327613e-05

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer: **Solved using Jupyter Note Book (Python)- File attached**

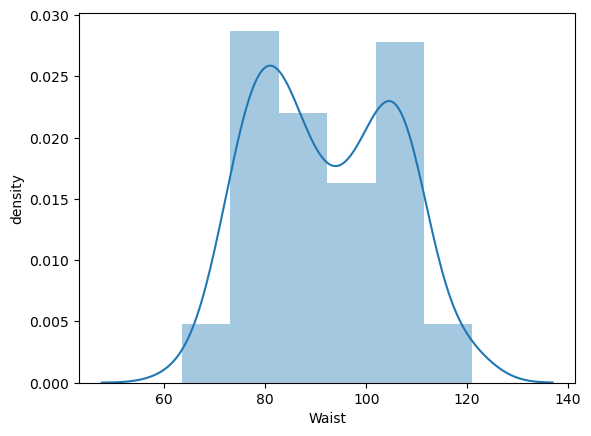
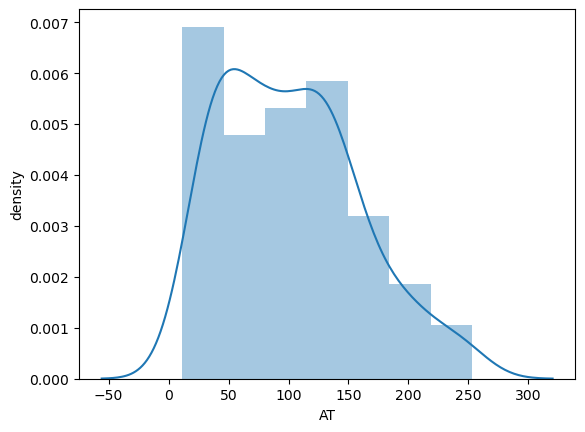


Data is Normally Distributed

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

Dataset: wc-at.csv

Answer: **Solved using Jupyter Note Book (Python)- File attached**

Data is Normally Distributed for both AT and Waist Circumference

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

Answers:

|  |  |  |
| --- | --- | --- |
| **Confidence Interval** | **Python Code** | **Z scores** |
| 90 | stats.norm.ppf(.95) | 1.645 |
| 94 | stats.norm.ppf(.97) | 1.881 |
| 60 | stats.norm.ppf(.80) | 0.841 |

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

Answers:

|  |  |  |
| --- | --- | --- |
| **Confidence Interval** | **Python Code** | **t scores** |
| 95 | stats.t.ppf(0.975,24) | 2.064 |
| 96 | stats.t.ppf(0.98,24) | 2.172 |
| 99 | stats.t.ppf(0.995,24) | 2.80 |

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

Answers: **Solved using Jupyter Note Book (Python)- File attached**

Claimed Average Life span of Bulb= 270 days

Sample bulbs= 18

Sampled bulb average Life span= 260 days

SD= 90 days

**Assume Null Hypothesis is: Ho = Avg life of Bulb >= 260 days**

**Alternate Hypothesis is: Ha = Avg life of Bulb < 260 days**

Answer is Avg life of Bulb < 260 days

P< α

0.32167411684460556 < 0.05