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

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

**ANS :**

S ={HHH,HHT,HTH,THH,TTT,TTH,THT,HTT}

We have to find AT least 2 HH & 1T

i.e HHT,HTH,THH

out of 8 possible outcomes,3 outcomes have two head &1tail

P = Favourable Outcomes/Total Outcomes

= 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 2 and 3

**ANS:**

a)Since there are no outcomes which correspond to a sum equal to 1)There fore in this case i.e, equal to 1 isProbability = No. of interested events/Total no of eventsProbability = 0/36=0

b)Less than or equal to 4 {(1,1) (1,2) (1,3) (2,1) (2,2) (3,1)}Probability = No. of interested events/Total no of eventsProbability = 6/36=1/6 (0.16666)

c) There are 5 rolls that produce 6, i.e., 1–5, 2–4, 3–3, 4–2, 5–1. This we have 6 of the 36 possible rolls that produce sums that are divisible by both 2 and 3. That is 16.67%

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

Total of ways that the balls can be picked out of 7 balls are 2 randomly [N(s)]

ncr = n!/(n-r)!(r)!

7c2 = (7\*6)/2 = 21

N(s) = 21

Total balls after picking the 2 balls out of 7 balls is 5 [N(e)]

ncr = n!/(n-r)!(r)!

5c2 = (5\*4)/2 = 10

N(e) = 10

There-fore probability = N (e)/N(s)

`i.e., **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**:

Child A:

Probability=0.015

Total event =1

Interested event/Expected candies =?

So formula:

Probability = No. of interested events/Total no of events

No. of interested events= Probability\* Total no of events

No. of interested events/ Expected candies(child A) =0.015\*1= 0.015

Similarly, for Child B

No. of interested events/ Expected candies (child B) =0.20\*4= 0.80

Similarly other= 1 \* 0.015 + 4\*0.20 + 3 \*0.65 + 5\*0.005 + 6 \*0.01 + 2 \* 0.12= 0.015 + 0.8 + 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**

**ANS :** .py file is aatched

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

1308/9 = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**ANS :** .py file is aatched

**SP and Weight(WT)**

**Use Q9\_b.csv**

**ANS :** .py file is attached

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



**ANS :** With the above histogram, we can conclude that it is a right skewed histogram and the skewness value will be always positive. In the interval 50-100 , we could see more number of chick weights.



**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?

**ANS :**

Sample n = 2000

Sample mean x = 200

Sample standard deviation = 30

94% Confidence Interval: (198.74, 201.26)

98% Confidence Interval: (198.44, 201.56)

96% Confidence Interval: (198.62, 201.38)

**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 : 1)** Mean: 41

Median: 40.5,

variance: 25.52941,

standard deviation: 5.052664

**2)**Mean > Median, This implies that the distributionis slightly skewed towards theright. Marks are not normally distributed. Person with mark 56 can be the outlier inour data. Many of the students are having 41 marks i.e. the modal value of our data

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

**ANS** : The distribution is perfectly symmetrical.

Skewness is 0.

Kurtosis is 3 (for normal distribution). If kurtosis is less than 3

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

**ANS :** Right skewed/ Positive skewness, most of the data will be lying on left side of the plot

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

**ANS :** left skewed/negative skewness, most of the data will be lying on the right side of the plot

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

**ANS :** sharp peak and less variation

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

**ANS :** Broad peak and more variation.

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



-What can we say about the distribution of the data?

**ANS :** The data is not symmetric. Mass of the datais concentrated towards the right side. Itis not a Normal Distribution

-What is nature of skewness of the data?

**ANS :** Negative Skewness

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

**ANS :** 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 :**

1) The median of the two boxplots are approximately260 & The boxplots are not skewed in +ve or –ve direction.

2) Outliers don't exist in both of the boxplots.

3)Comparing Boxplot\_1 (Red) has less variability,less variation, less standarddeviation, less range, less Inter-quartile-range value as compared toBoxplot\_2(blue)

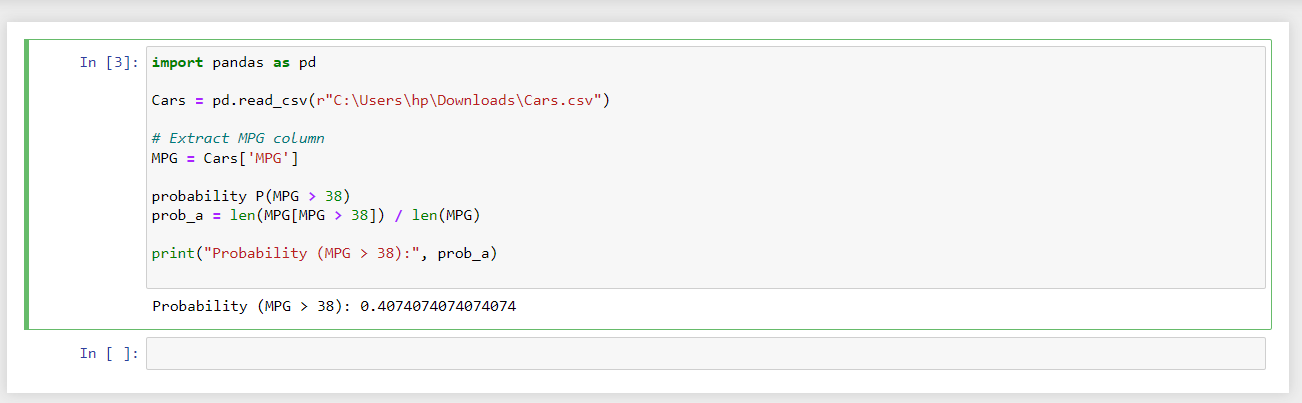
Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

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

MPG <- Cars$MPG

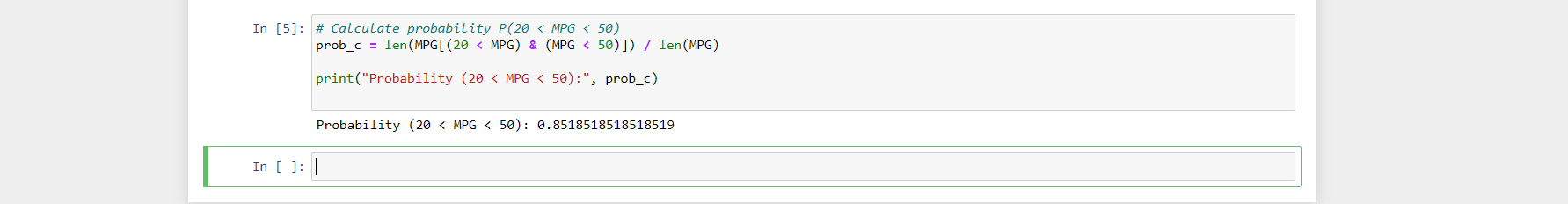
* 1. P(MPG>38)



* 1. P(MPG<40)



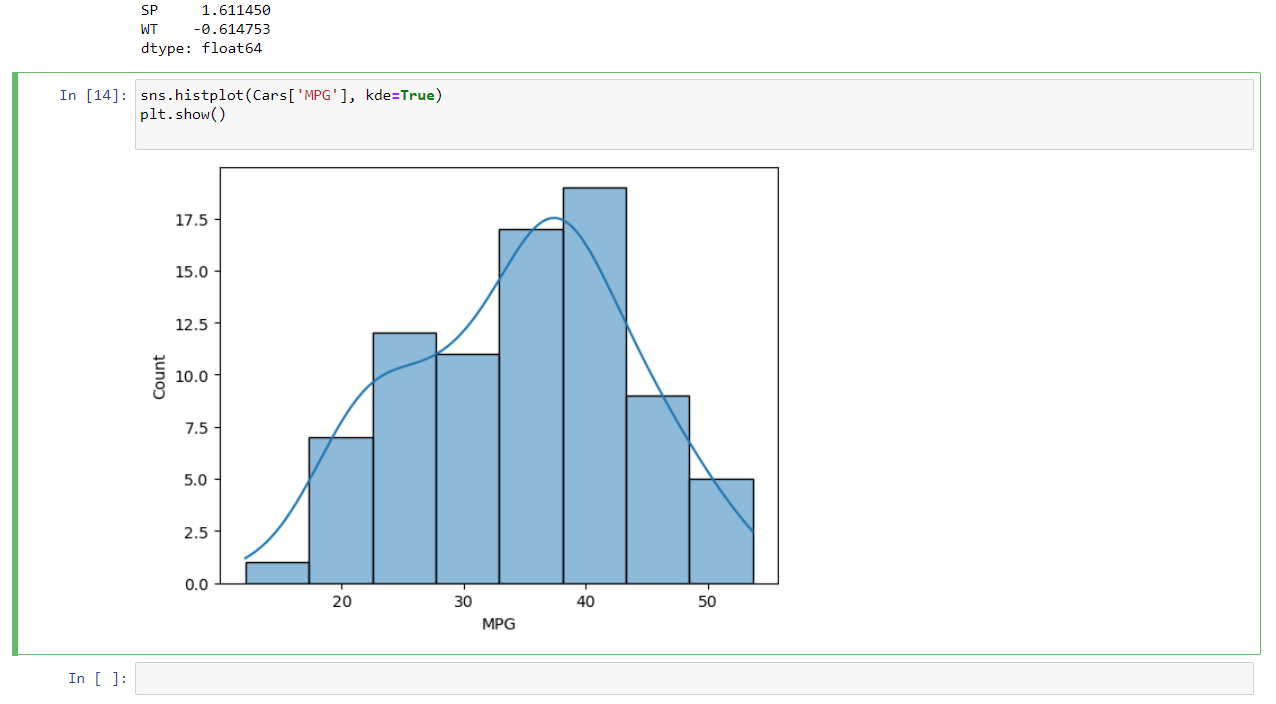
* 1. P (20<MPG<50)



Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

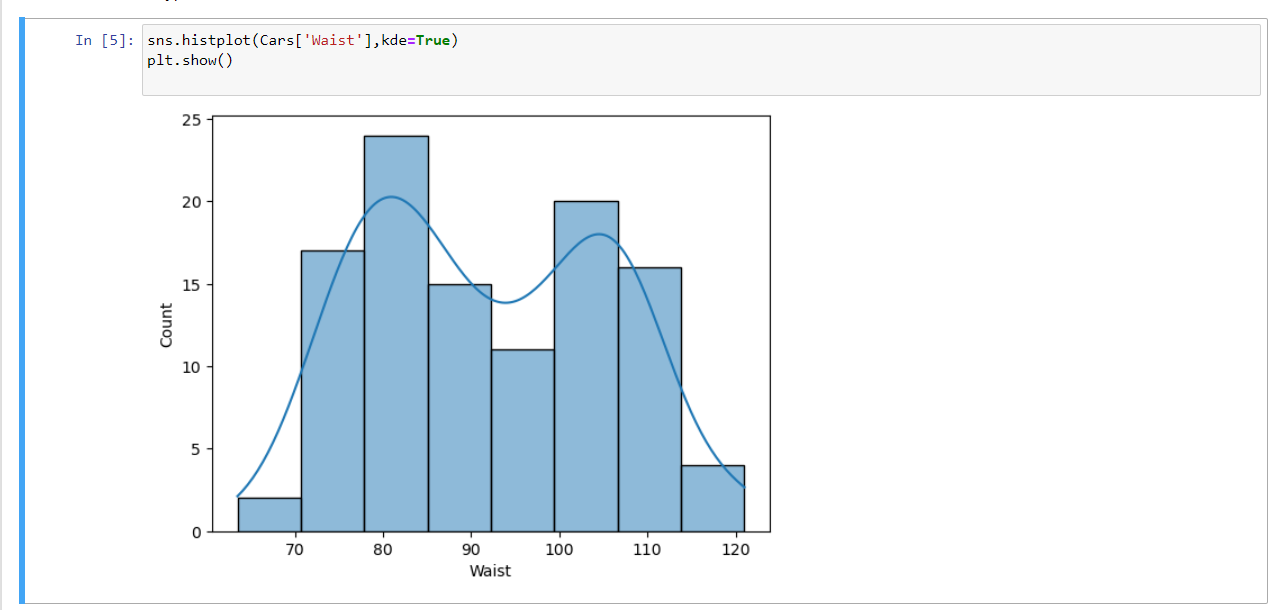
Dataset: Cars.csv ANS : MPG not follows Normal Distribution

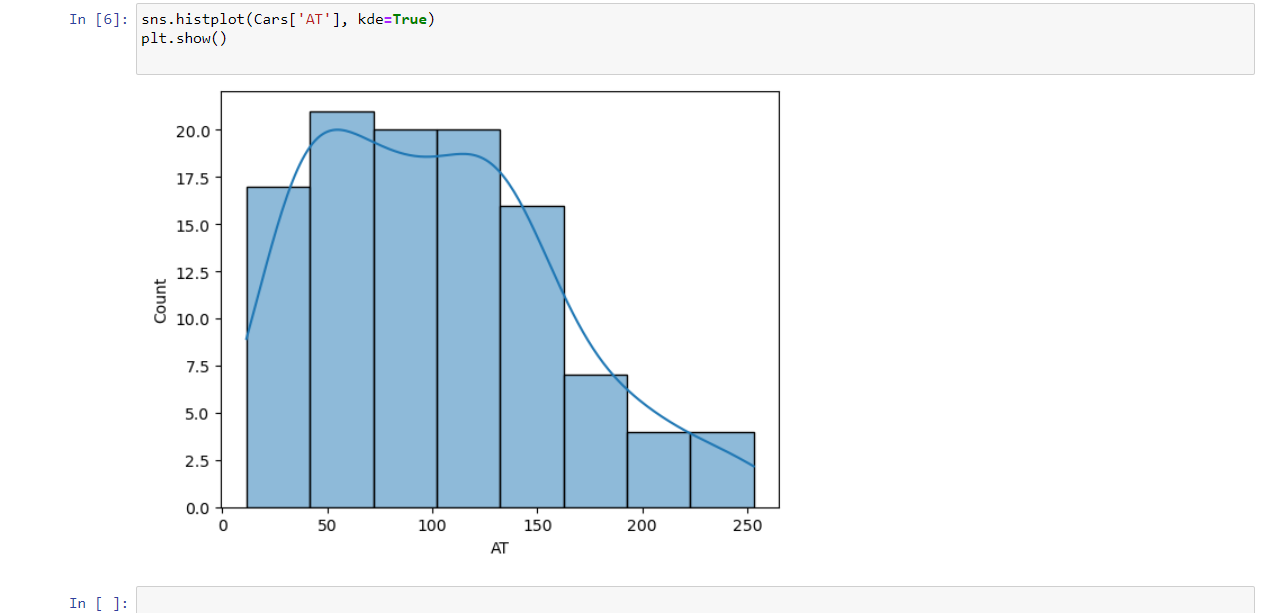


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

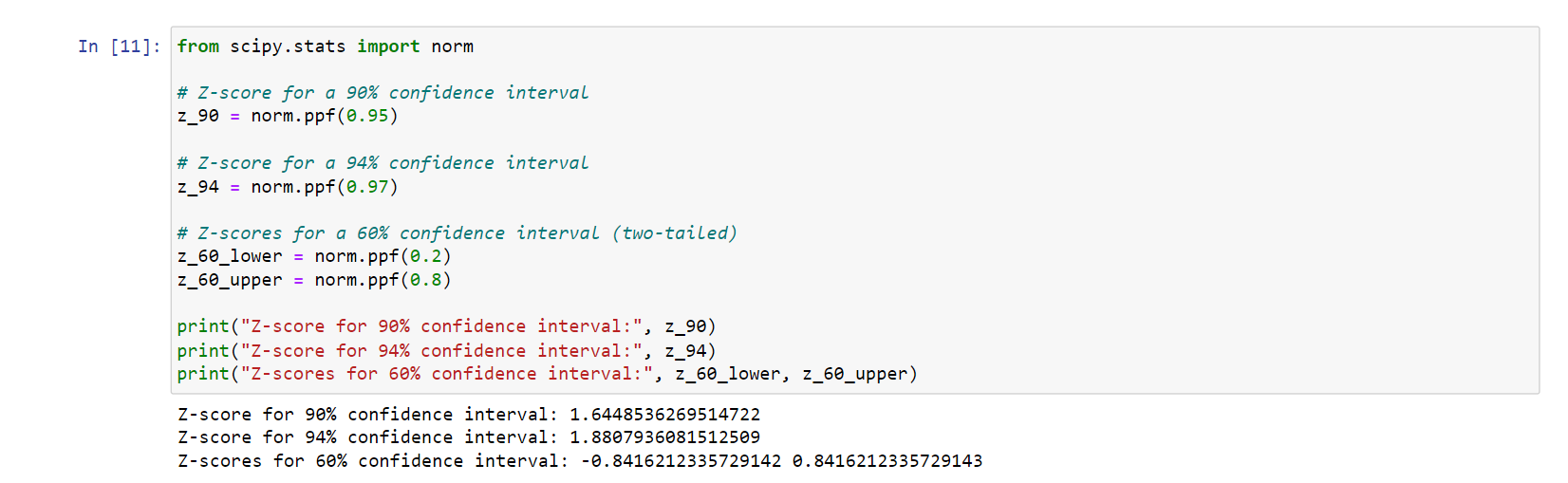






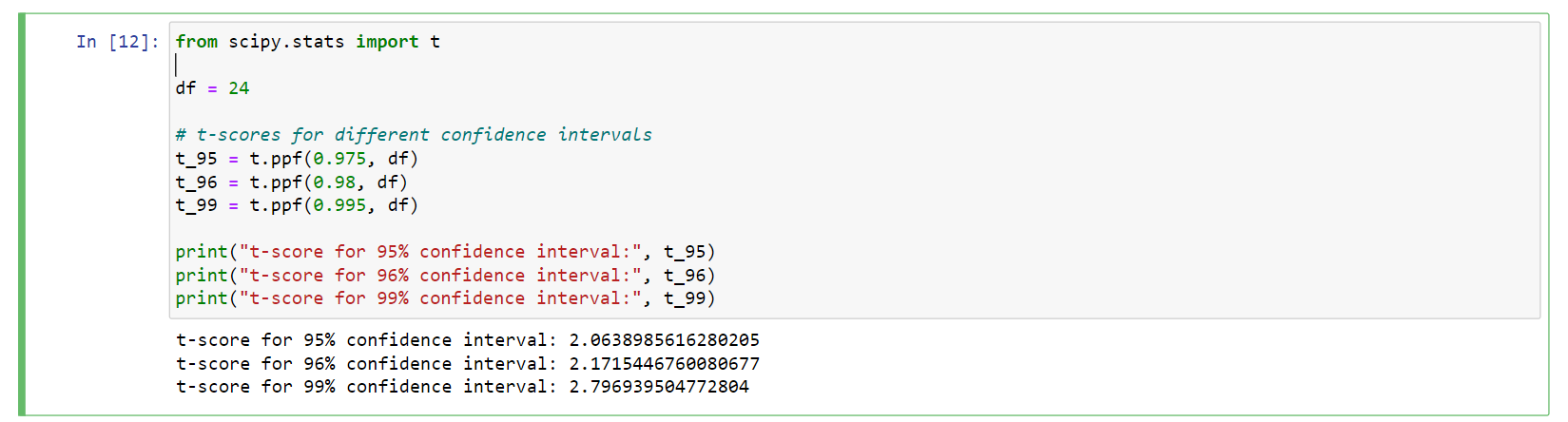
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

**ANS :**



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

**ANS:** calculate the t-scores for different confidence intervals with a sample size of 25 we use use the t-distribution table The t-score is based on the degrees of freedom, which is equal to the sample size minus 1 (df = 25 - 1 = 24).



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 :** # Given values

sample\_mean <- 260

population\_mean <- 270

sample\_sd <- 90

sample\_size <- 18

# Calculate t-score

t\_score <- (sample\_mean - population\_mean) / (sample\_sd / sqrt(sample\_size))

# Degrees of freedom

df <- sample\_size - 1

# Calculate probability using t-distribution

probability <- pt(t\_score, df)

# Display the result

probability

**The probability that 18 randomly selected bulbs would have an average life of no more than 260 days is: 0.3216725**