|  |  |
| --- | --- |
| 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 | Interval |
| 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) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Nominal |

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

Ans) Three possible outcomes are (HHT),(HTH)&(THH)

Each outcome has a probability of (1/2)^2\*(1/2)=1/8

Since there are three possible outcomes the total probability is:

P(2 heads and 1 tail) = P(HHT) + P(HTH) + P(THH)

= 1/8 + 1/8 + 1/8

= 3/8

= **0.375**

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

1. Equal to 1

Ans) When two dice are rolled, the minimum sum that can be obtained is 2, and

the maximum sum is 12. The sum cannot be 1 because the lowest value on a

single die is 1, and the sum of 1’s is 2.

Therefore, the probability of obtaining a sum of 1 when two dice are rolled is

0, as it is not possible outcome.

1. Less than or equal to 4

Ans) Total no. of outcomes = 36

Expected outcomes = 6

Probability = 6/36 = **0.17 .**

outcomes:

(1,1),(1,2),(2,1),(1,3),(3,1)

Since each die has 6 sides, there are 6 possible outcomes for each die.

As we are rolling two dice, the total number of possible outcomes is 6\*6=36.

Therefore, the probability of obtaining a sum less than equal to 4 is:

P(sum <= 4) = Number of favourable outcome/ Total number of possible

outcomes = 5/36 which is **0.1389.**

1. Sum is divisible by 2 and 3

Ans) The number divisible by 2 are 2,4,6,8,10,12. The number divisible by 3 are

3,6,9,12.

Now, we have to find common numbers in there two sets. The number that

are divisible by 2 & 3 are 6 & 12.

we calculate the total number of outcomes when two dice are rolled. Each

die has 6 sides, so there are 6 x 6 = 36 possible outcomes.

Out of these 36 possible outcomes, there are 2 favorable outcomes

(6 and 12).

Therefore, the probability that the sum of two dice rolls is divisible by both

2 and 3 is:

P(divisible by 2 and 3) = Favorable outcomes / Total outcomes

= 2 / 36

= 1 / 18

Hence, the probability is 1/18 or approximately **0.0556.**

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= (2 + 3 + 2)= 7  
 Let S be the sample space  
 Then, n(S) = Number of ways of drawing 2 balls out of 7

n(S)=7C2

n(S)=(7×6)/(2×1)

n(S)=21

Let E = Event of 2 balls, none of which is blue  
 ∴ n(E) = Number of ways of drawing 2 balls out of (2 + 3) balls

n(E)=5C2n(E)=(5×4)/(2×1)

n(E)=10

∴P(E)=n(E)/n(S)=**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) 1\*0.015 + 4\*0.20 + 3\*0.65 + 5\*0.005 + 6\*0.01 + 2\*0.120

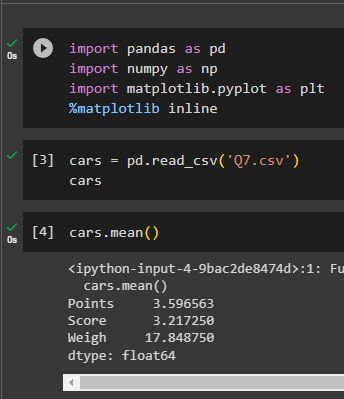
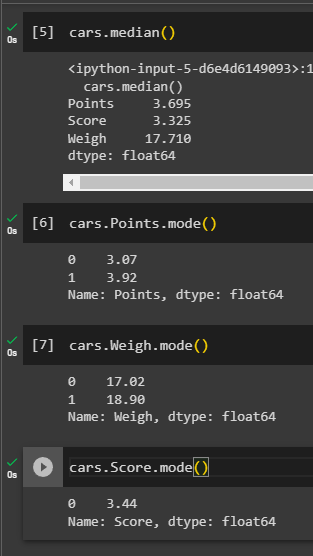
Expected no. of candies for a randomly selected child = **3.09**

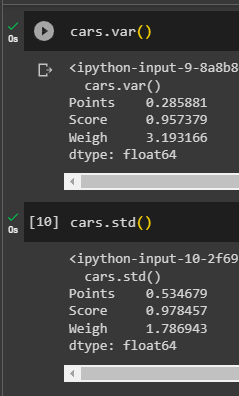
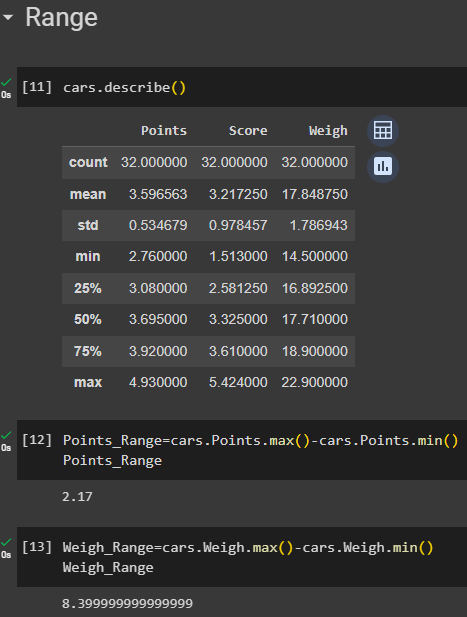
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* 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**

** **

** **

****

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) Expected value = Mean of given data

108+110+123+134+135+145+167+187+199/ 9

Hence mean = 1308/9

= **145.33**

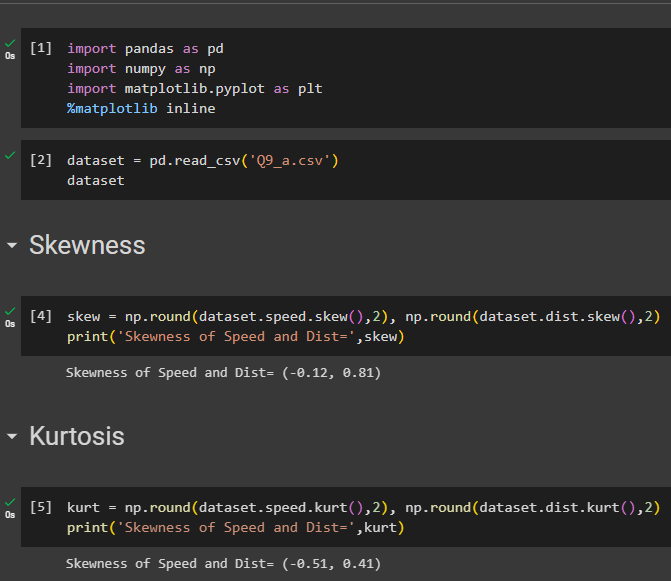
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

**Answer:** Skewness of Speed & Distance**=(-0.12,0.81)**

Kurtosis of Speed & Distance**=(-0.51,0.41)**

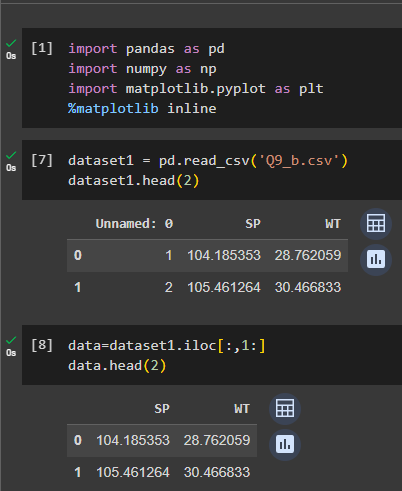
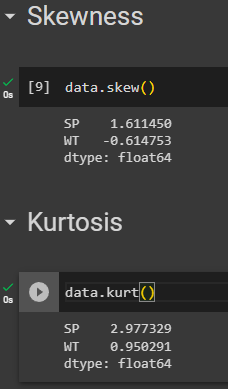
****

**SP and Weight(WT)**

**Use Q9\_b.csv**

**ANSWER**: Skewness of SP & WT is 1.611 & -0.614

Kurtosis of SP & WT is 2.977 & 0.950.

** **

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

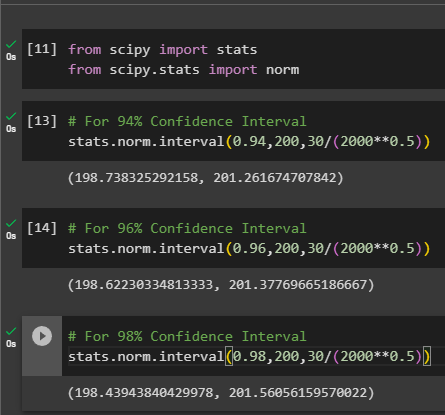


**Answer :**  Given histogram is right skewed, which means that there are lesser number of concentration of chic weight in the 300 – 400 category.

In the given boxplot, median is less than mean, which clearly mean that it has right skewed distribution. Also, there are some outliers on the upperside of the boxplot.

**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:**

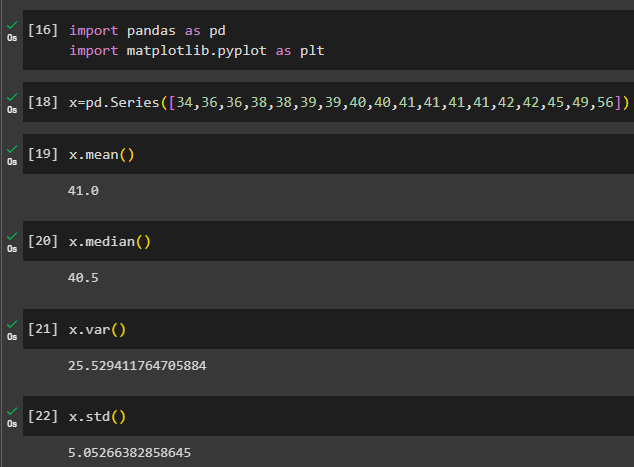
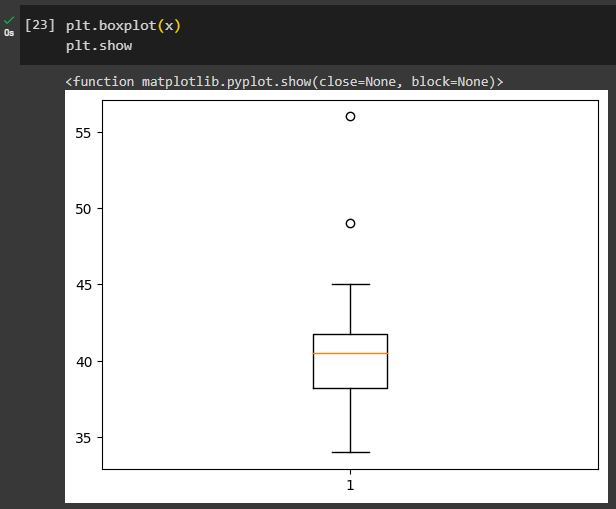


**Q12)** Below are the scores obtained by a student in tests

**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:**

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

**Answer** : If the mean and median of data are equal, this means that the distribution is symmetric. And such distribution has zero skewness.

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

**Answer** : If mean is greater than median, then the nature will be positively skewed.

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

**Answer** : If median is greater than mean, then the nature will be negatively skewed.

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

**Answer** : Positive values of kurtosis indicates that distribution is peaked and has thick tail.

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

**Answer** : 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:** The above boxplot is following normal distribution, the median is towards higher value.

What is nature of skewness of the data?

**Answer:** The data is left skewed, the whisker range of minimum value is greater than maximum value.

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

**Answer:** IQR=Q3-Q1(Upper quantile – Lower quantile)

IQR=18-10

**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:**

* Both the boxplot has no outliers.
* Both the boxplots are normally distributed, hence no skewness can be seen.
* Both the boxplots shares the same median.

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

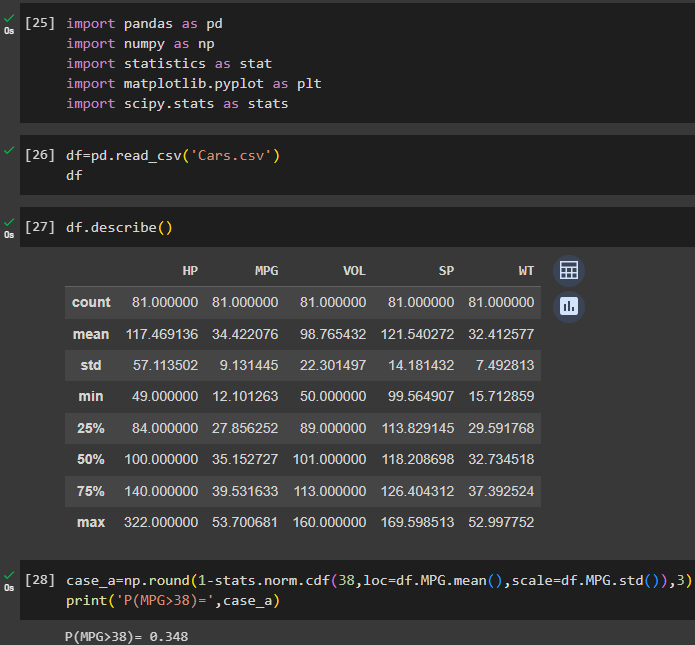
Data \_set: Cars.csv

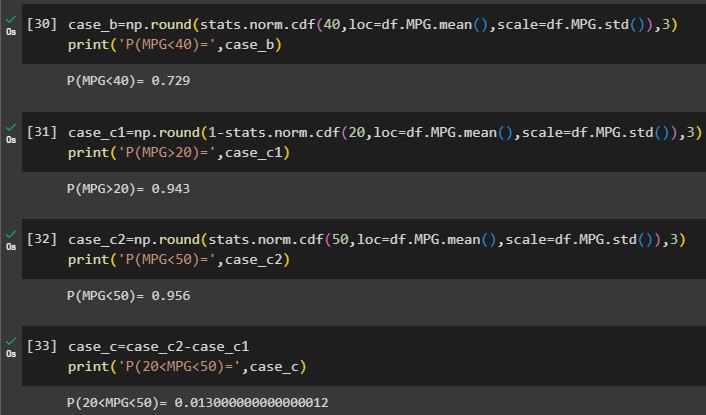
Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

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

c. P (20<MPG<50)

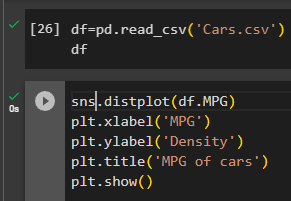
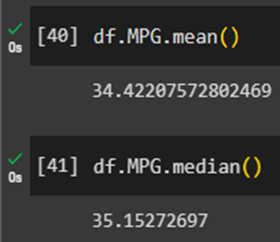


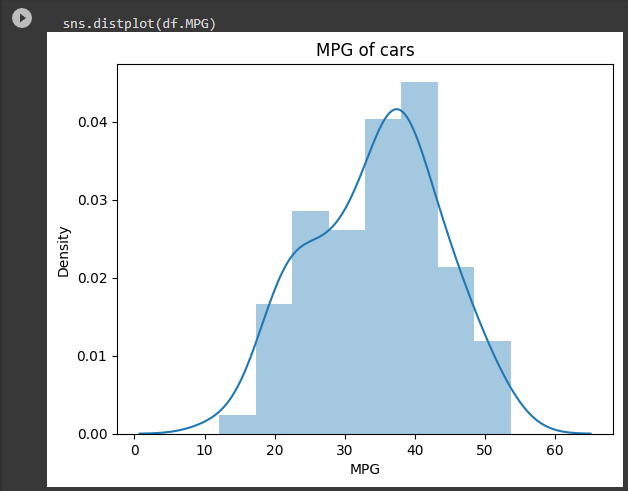


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

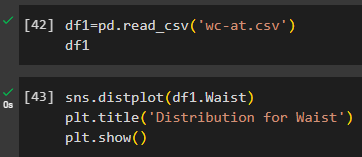
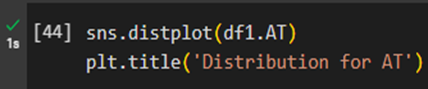
Dataset: Cars.csv

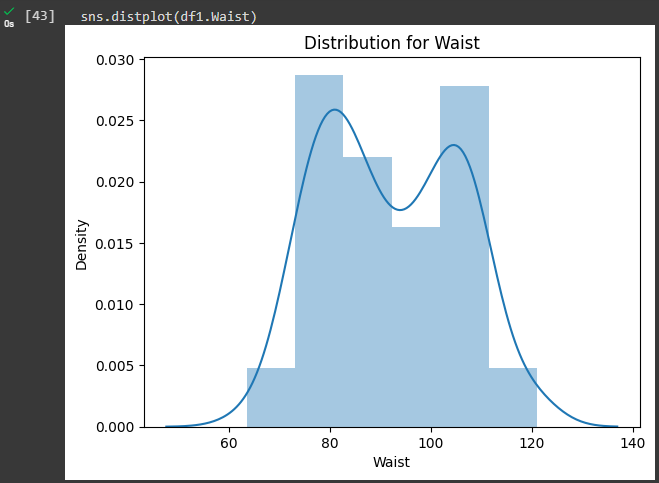
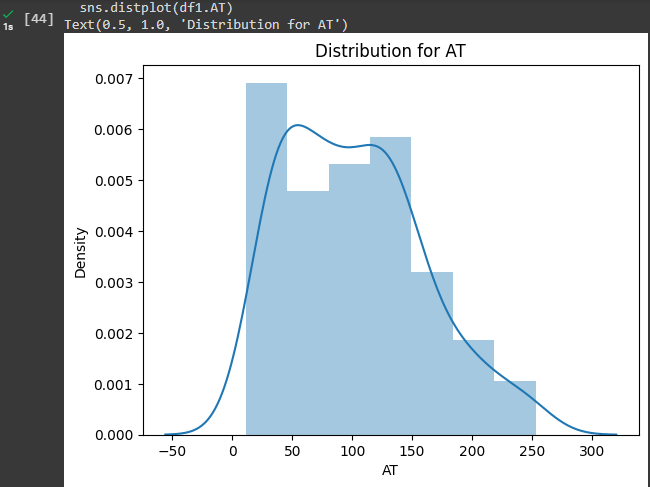
 

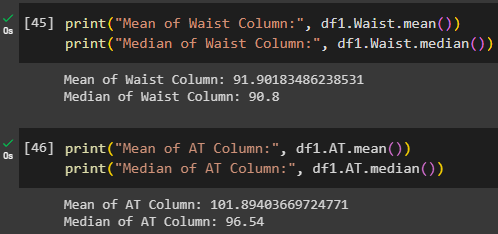


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

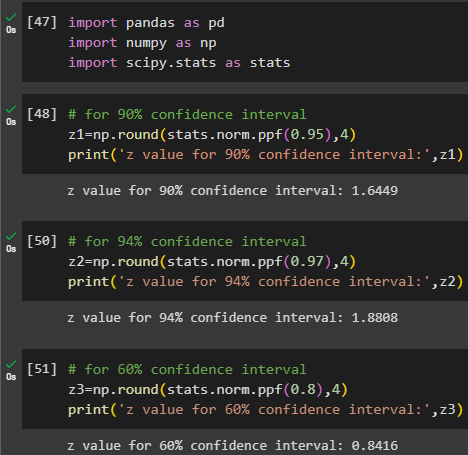
Dataset: wc-at.csv

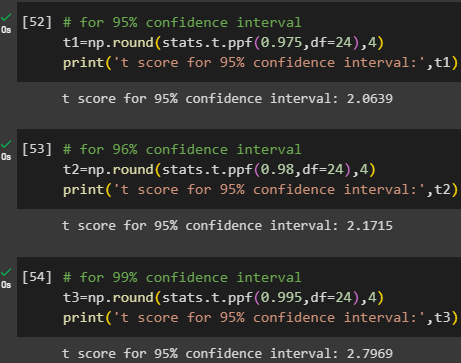
 



Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval



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

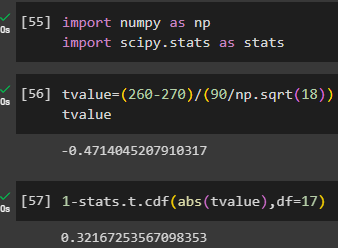


Q 24**)** A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs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

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom



Probability of life of a bulb not more than 260 days is **32%**.