Q1) Identify the Data type for the Following:

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

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)	Ratio
Sales Figures	Ratio
Blood Group	Nominal
Time Of Day	Ratio
Time on a Clock with Hands	Ordinal
Number of Children	Ratio

Religious Preference	Ordinal
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?

Solution:

When three coins are tossed the possible outcomes are:

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

For getting two heads and one tail ={HHT.HTH,THH}

Probability =(no. of favorable outcome/total no. of possible outcomes)

=3/8

So, the probability that two heads and one tail is 3/8.

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

Solution:

Two dice are rolled

S={ (1,1), (1,2), (1,3), (1,4), (1,5), (1,6),

(2,1), (2,2), (2,3), (2,4), (2,5), (2,6),

(3,1), (3,2), (3,3), (3,4), (3,5), (3,6),

(4,1), (4,2), (4,3), (4,4), (4,5), (4,6),

(5,1), (5,2), (5,3), (5,4), (5,5), (5,6),

(6,1), (6,2), (6,3), (6,4), (6,5), (6,6) }

a) **Equal to 1:**

There is no combination that gives a sum of 1 with two six-sided dice.

Probability = 0

b) **Less than or equal to 4:**

Possible combinations: (1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (3, 1)

Total possible outcomes = 36

Probability = Number of favorable outcomes / Total possible outcomes = 6 / 36 = 1/6

c) **Sum is divisible by 2 and 3:**

Combinations with sums divisible by 2:

(1, 1), (1, 3), (1, 5), (2, 2), (2, 4), (2, 6), (3, 1), (3, 3), (3, 5), (4, 2), (4, 4), (4, 6), (5, 1), (5, 3), (5, 5), (6, 2), (6, 4), (6, 6)

Combinations with sums divisible by 3:

(1, 2), (1, 5), (2, 1), (2, 4), (3, 3), (3, 6), (4, 2), (4, 5), (5, 1), (5, 4), (6, 3), (6, 6)

Combinations with sums divisible by both 2 and 3: (1, 5), (3, 3), (2, 4), (4, 2), (5, 1), (6, 6)

Total possible outcomes= 36

Probability = Number of favorable outcomes / Total possible outcomes =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?
Solution :

The total number of balls in the bag is 2 (red) + 3 (green) + 2 (blue) = 7 balls.

1. **Probability of drawing the first non-blue ball:**

Number of non-blue balls: 2 (red) + 3 (green) = 5

Probability of the first draw being non-blue: 5/7

2. **Probability of drawing the second non-blue ball (after the first non-blue ball has been drawn):**

Number of non-blue balls remaining:

4 (2 red + 2 green, as one non-blue ball has already been drawn)

Probability of the second draw being non-blue: 4/6=2/3

Now, to find the probability of both events occurring (drawing two non-blue balls), multiply the probabilities:

P(None are blue)= P(First non-blue) × P(Second non blue)

P (None are blue)=5/7×2/3

P (None are blue)=10/21 P

So, the probability that none of the balls drawn is blue is 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

Solution :

The formula for expected value (mean) is given by:

$E(X) = \sum xi \cdot P(X=xi)$

where is the candies count, and $P(X=xi)$ is the probability of that candies count.

For the provided data:

$E(X)=(1\cdot0.015)+(4\cdot0.20)+(3\cdot0.65)+(5\cdot0.005)+(6\cdot0.01)+(2\cdot0.120)$

$E(X)=0.015+0.80+1.95+0.025+0.06+0.$

$E\ (X)=3.13$

So, the expected number of candies for a randomly selected child is 3.13.

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:

	Points	**Score**	**Weigh**
Total	**115.09**	**102.952**	**571.16**
Mean	**3.596563**	**3.21725**	**17.84875**
Median	**3.695**	**3.325**	**17.71**
Mode	**3.92**	**3.44**	**17.02**
Variance	**0.276948**	**0.927461**	**3.09338**
S.D.	**0.526258**	**0.963048**	**1.758801**
Max	**4.93**	**5.424**	**22.9**
Min	**2.76**	**1.513**	**14.5**

Mean value are closer for both ‘Point’ and ‘Score’.

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?

Solution:

$E(X)= 1/n * \sum xi$

where *xi* is the weight of the i-th patient, and *n* is the total number of patients.

Let's calculate the expected value for the given weights:

$E(X)=108+110+123+134+135+145+167+187+199$

$E(X)=1308/9$

$E(X)\approx145.33$

So, the expected value of the weight of a randomly chosen patient is approximately 145.33 pounds.

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

Cars speed and distance

Use Q9_a.csv

Solution:

	Speed	**Distance**
Skewness	**-0.11751**	**0.806895**

Kurtosis	**-0.50899**	**0.405053**

Speed:

Here the value of skewness is -ve since the data is left skewed.

Most of the values lies towards the right part of the plot.

Distance:

Here the value of skewness is +ve hence the data is right skewed.

Most of the data lies towards the left side of the plot

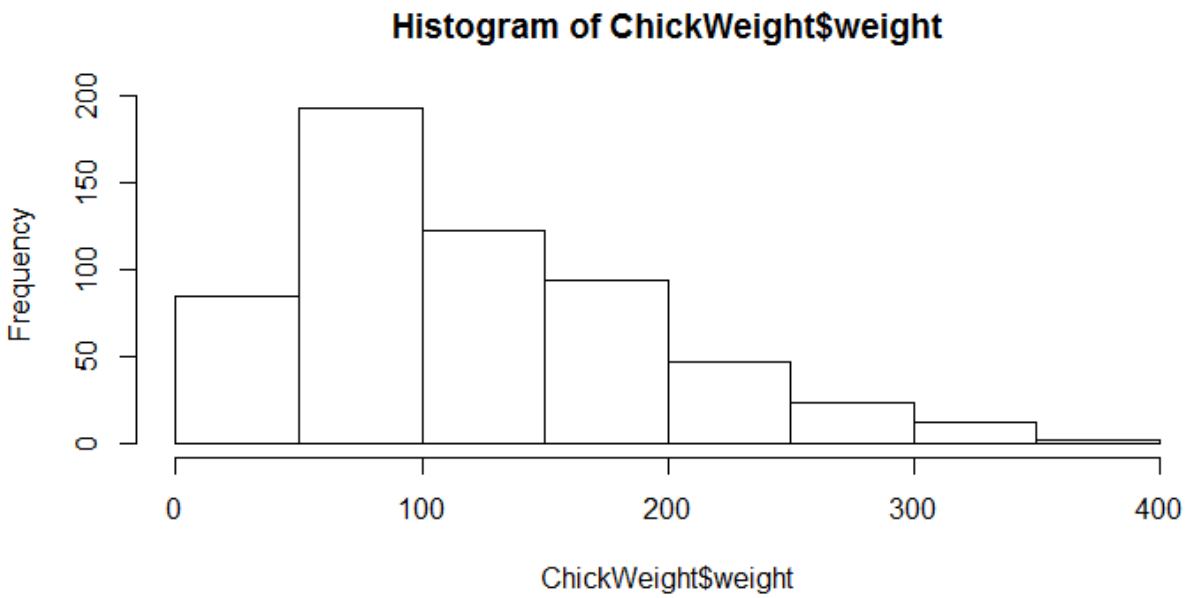
SP and Weight(WT)

Use Q9_b.csv

Solution:

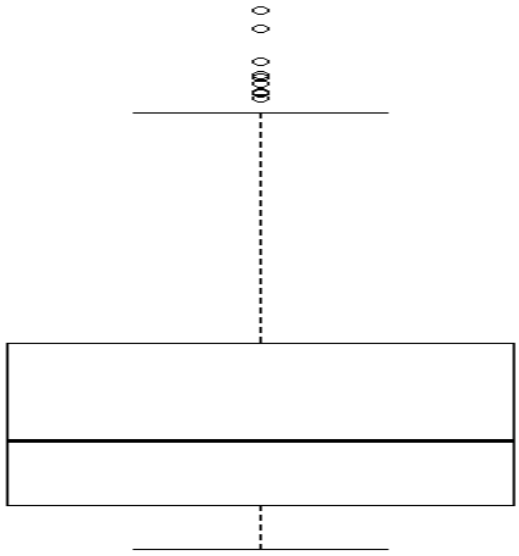
	SP	**WT**
Skewness	**1.6114502**	**-0.61475**
Kurtosis	**2.97732894**	**0.950291**

Q10) Draw inferences about the following boxplot & histogram



Solution:

Here we can see that the major Chick weights fall in the catagory of 50-100g(measures in x) as the maximum which is 200.The minimum weights have a frequency if less than or equal to 5. The plot is Right skewed which show that there is lesser concentration of chick weights in the 300-400gram category . The expected value should be above 46.45



Solution: Median is less than mean right skewed and we have outlier on the upperside of box plot and there is less data points between Q1 and bottom point.

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 weight 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?

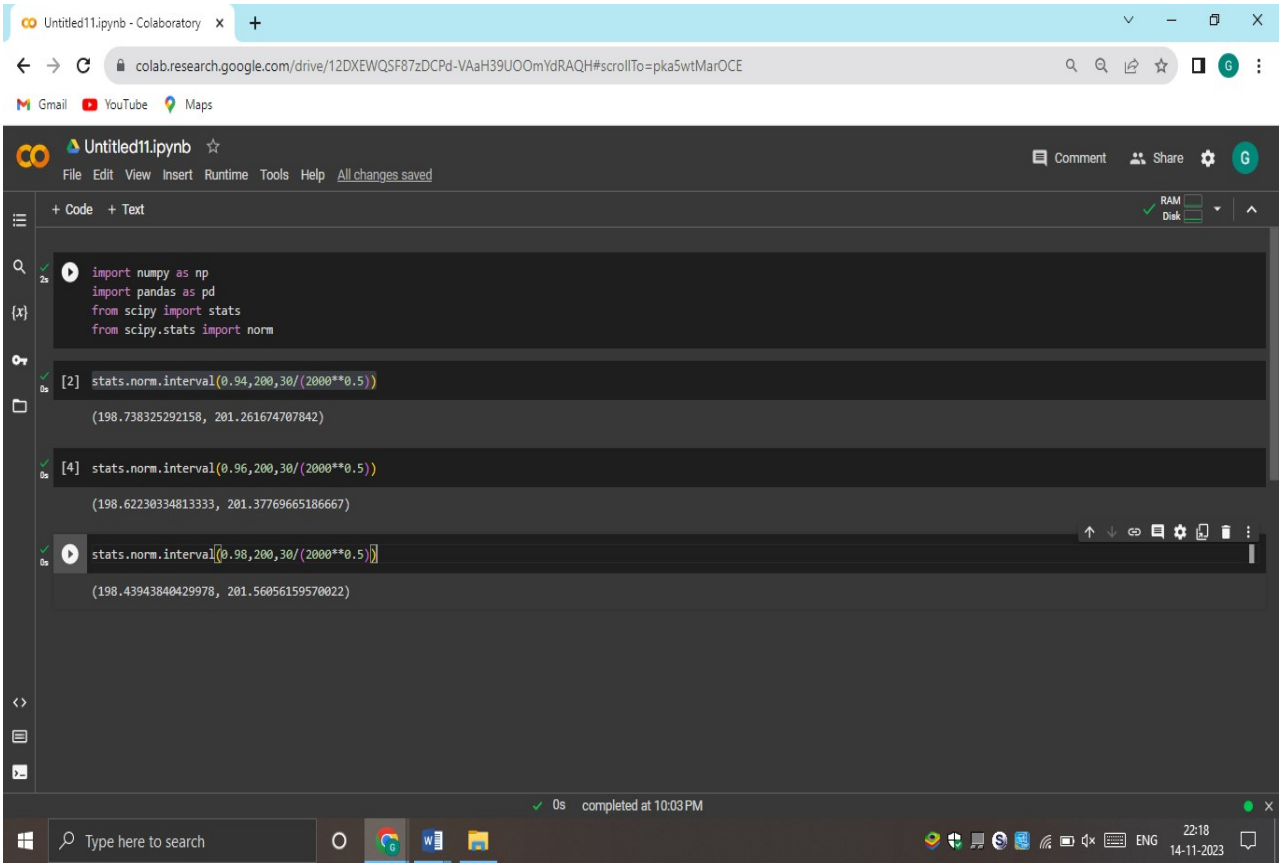
Solution:

Confidence Interval=Sample Mean±(Critical Value×Sample Size Standard Deviation)

The critical values for various confidence levels are as follows:

- For a 94% confidence interval, the critical z-value is approximately 1.88.
- For a 98% confidence interval, the critical z-value is approximately 2.33.
- For a 96% confidence interval, the critical z-value is approximately 1.96.

Now, let's plug in the values:



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?

Solution:

Let's start by calculating the requested statistics for the given test scores:

Test Scores: 34, 36, 36, 38, 38, 39, 39, 40, 40, 41, 41, 41, 41, 42, 42, 45, 49, 56

Mean	41
Median	40.5
Mode	41
Variance	24.11111

Standard deviation	4.910307

- **Mean (Average):** The average score is around 41 .
- **Median:** The median score is 40.5 .
- **Variance:** The variance is approximately 24.1111.
- **Standard Deviation:** The standard deviation is approximately 4.910307.

Interpretation:

1. The mean and median give us an idea of the central tendency of the scores. In this case, the mean is slightly greater than the median, indicating a slightly Right-skewed distribution.
2. The variance and standard deviation measure the spread or dispersion of the scores. A higher standard deviation suggests more variability in the scores.

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

Solution:

When the mean and median of a dataset are equal, it indicates that the distribution of the data is approximately symmetric. In a perfectly symmetric distribution, the mean and median coincide at the center of the distribution.

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

Solution:

When the mean is greater than the median, it indicates a right-skewed or positively skewed distribution. This means that there is a longer or fatter tail on

the right side of the distribution, and the majority of the data points are concentrated on the left side.

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

Solution:

When the median is greater than the mean, it indicates a left-skewed or negatively skewed distribution. This implies that there is a longer or fatter tail on the left side of the distribution, and the majority of the data points are concentrated on the right side.

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

Solution:

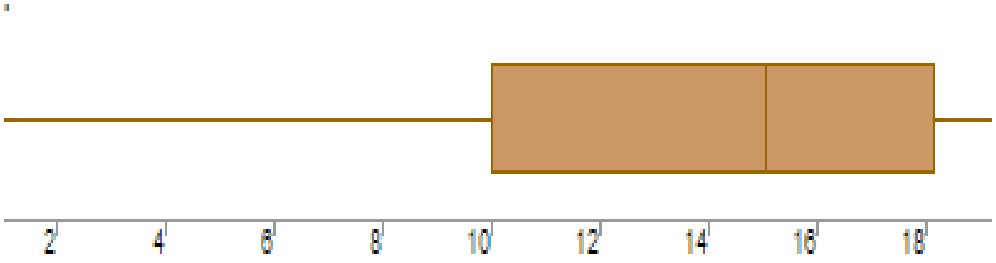
A positive kurtosis value indicates that the dataset has heavier tails and a more peaked or pronounced central peak than a normal distribution. It suggests that there are more extreme values (outliers) than would be expected in a normal distribution. This kind of distribution is referred to as leptokurtic.

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

Solution:

A negative kurtosis value indicates that the dataset has lighter tails and a flatter central peak than a normal distribution. It suggests that there are fewer extreme values than would be expected in a normal distribution. This type of distribution is referred to as platykurtic.

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



What can we say about the distribution of the data?

The majority of the data points are concentrated on the right side of the distribution with a long tail extending to the left. This means that the mean is less than the median, which is less than the mode.

What is nature of skewness of the data?

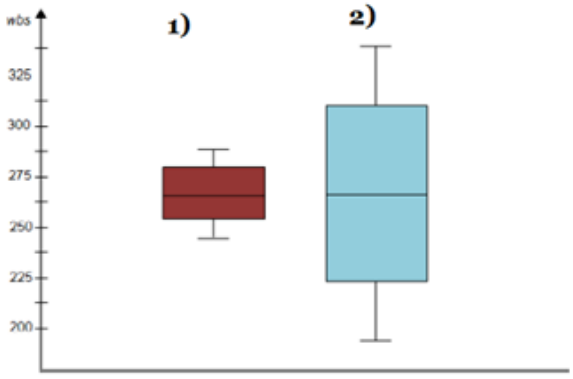
The data is Negatively-skewed.

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

The approximately Interqurtile range of the data is 8.

Q19) Comment on the below Boxplot visualizations?

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



Solution:

Here when we compare box plot 1 with box plot 2 we can say that the data in box plot 1 is widely spread. Here the main inference is that since the data range varies high in box plot 2 it is hard to make a prediction in box plot 2. The median in the 2 box plots are equal. And the data spread in both of them are symmetrical.

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)

Solution

P(MPG>38)

= mean(MPG)=34.42208

= sd(MPG)=9.131445

= 1 – pnorm(38,mean(MPG),sd(MPG))

= 0.330

= 33%

P(MPG<40)

=pnorm(40,mean(MPG),sd(MPG))

=0.7293499

=72.3%

P (20<MPG<50)

=pnorm(50,mean(MPG),sd(MPG)) – pnorm(20,mean(MPG),sd(MPG))

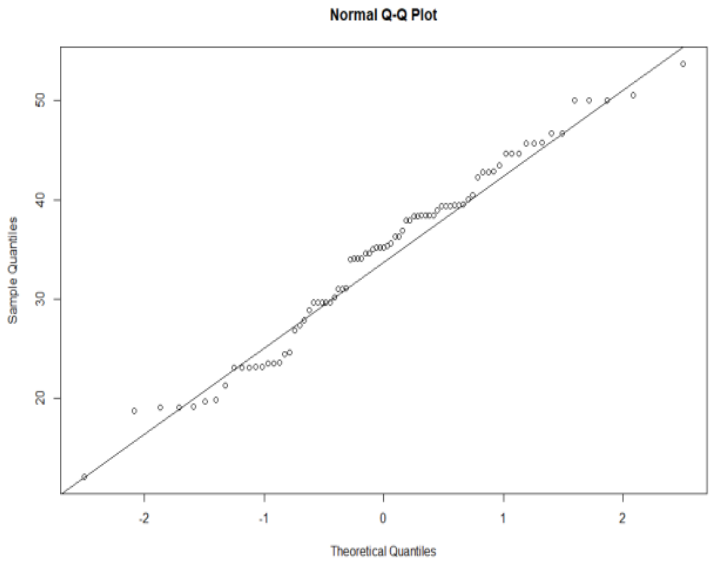
=0.955 -0.057

=0.8988689

=89.88%

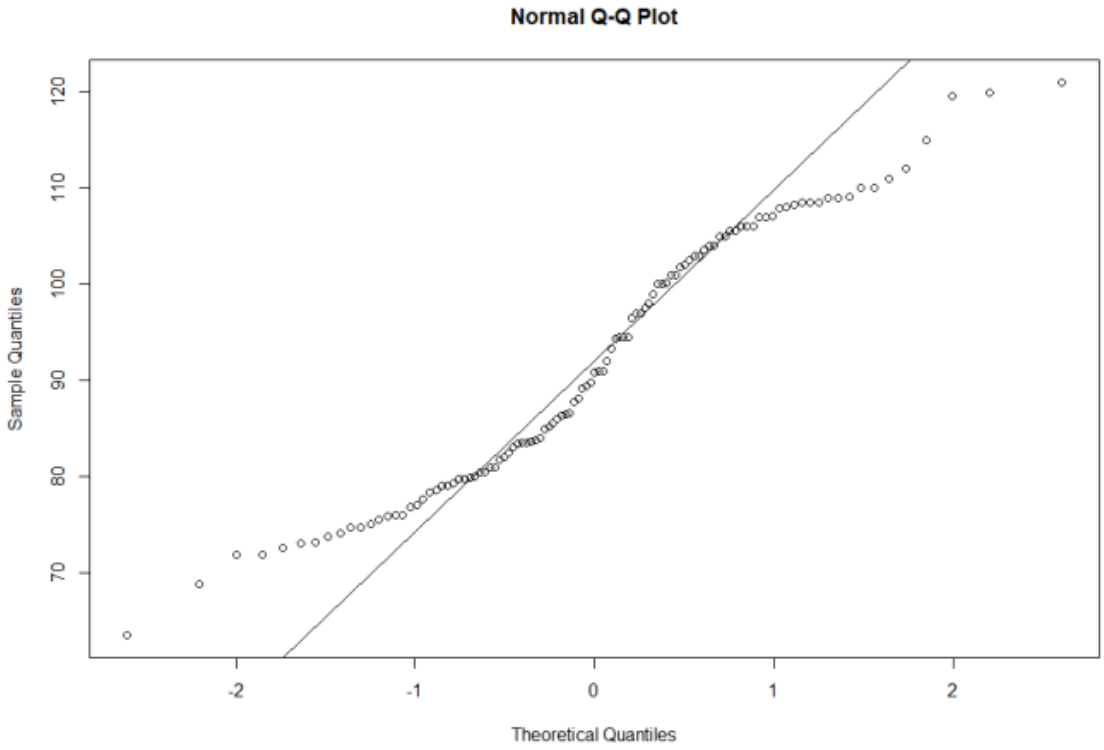
Q 21) Check whether the data follows normal distribution

a) Check whether the MPG of Cars follows Normal Distribution
Dataset: Cars.csv



When we plot check the qqnorm and qqline we can almost get a straight line thus the data is normalized.

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



This data set is not normal because the data points follows an abnormal curve.

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

Solution:

Z scores
=90%
= 95+2.5
=97.5
=qnorm(0.975)
=1.96
94%
= 94+4

=97
=qnorm(0.97)
=1.88
60%
= 60 + 20
= 80
= qnorm(0.80)
= 0.841

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

Solution:
TSCORE CALCULATION
T((1,alpha),(n-1))
Here n = 25
n-1 = 24
Hence t score values will be:
95%
= qt(0.975,24)
= 2.063899
96%
= qt(0.98,24)
=2.171545
99%
= qt(0.995,24)
= 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

Solution:

Sample size = 18 = n

Sample mean = 260

days = x

Sample standard deviation = s = 90days

=260 – 270/90/SQRT(18)

=-10/9.487

=-1.054