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

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

Ans: 3/8

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) Equal to 1=0

b) less than or equal to 4=5/36

c)sum is divisible by 2 and 3 =6/36

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 : 5c2/7c2 = 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 \* 0.12

= 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.



Ans:

|  |  |  |
| --- | --- | --- |
| Mean | | |
| 3.596563 | 3.21725 | 16.46 |
|  |  |  |
| Median | | |
| 3.695 | 3.325 | 16.46 |
|  |  |  |
| Mode | | |
| 3.92 | 3.44 | 16.46 |

The above values are for Mean median and mode for the following table.

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: Mean = 145.33

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

**Cars speed and distance**

Ans: 1)Sknewness of speed = -0.1139548

Since skewness is negative it is left skewed.

2) Kurtosis of speed is 2.422853

Since Kurtosis is Positive peak-edness is high

1)Sknewness of speed = 0.7824835

Since skewness is positive it is right skewed.

2) Kurtosis of speed is 3.248019

Since Kurtosis is Positive peak-edness is high



**SP and Weight(WT)**



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



Ans: From the above boxplot and histogram we can make the following inferences:

1)There are a number of **outliers** which needs to be treated if we have to perform analysis

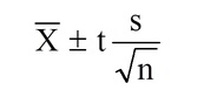
2) The distribution is skewed towards the right side hence is **Right Skewed.**

3) The highest frequency of weight is in the range **50-100**.

4) The data is normally distributed.

**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: The formula for t distribution is



Hence we can find the values for 94,98,96% confidence intevals as follows

1)200= 201.262 or 198.738 at 94%

2) 200= 201.561 or 198.438 at 98%

3) 200= 201.3784 or 198.621 at 94%

**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 | median | variance | standard deviation |
|  |  |  |  |
| 41 | 40.5 | 24.11111 | 4.910307 |

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

Ans: The skewnesss value would be **zero** and the data would be **symmetric**

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

Ans: If the mean is greater than the median, the distribution is **positively or right** 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 value of kutorsis indicates the data has a **Sharp peak**

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

Ans: Negativr kurtosis value indicates that the data does not have a **Sharp peak**

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



What can we say about the distribution of the data?

Ans: Data is not distributed normally , there is a sknewness factor

What is nature of skewness of the data?

Ans: Left skewed

What will be the IQR of the data (approximately)?   
Ans: IQR=Q3-Q1 =18-9=9

Q19) Comment on the below Boxplot visualizations?



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

Ans:The IQR of boxplot 1 w.r.t to boxplot 2 is very small .

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)

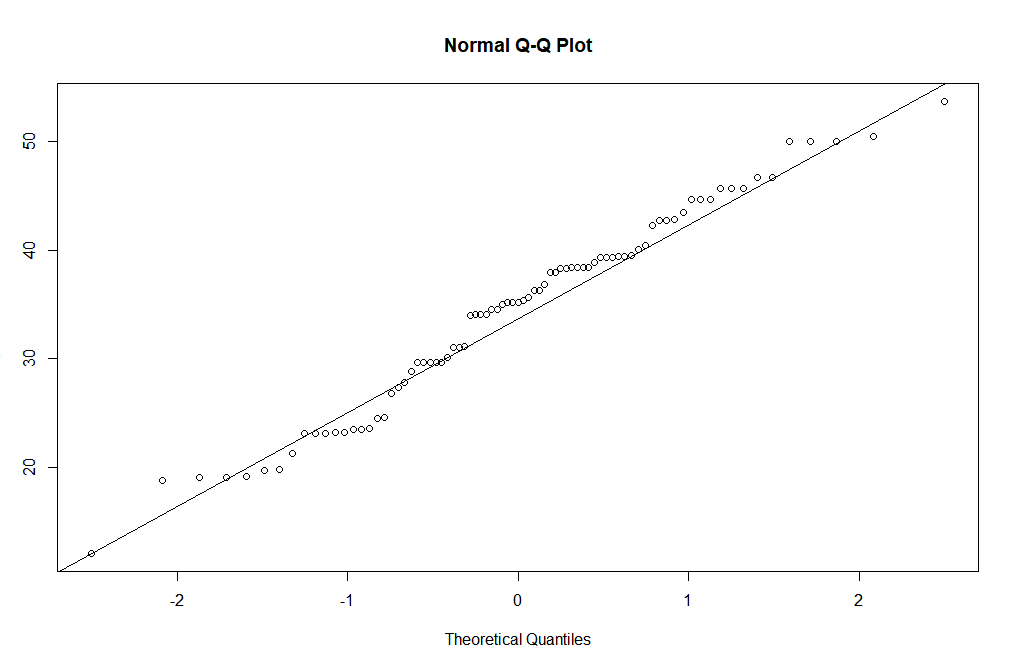
c. P (20<MPG<50)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

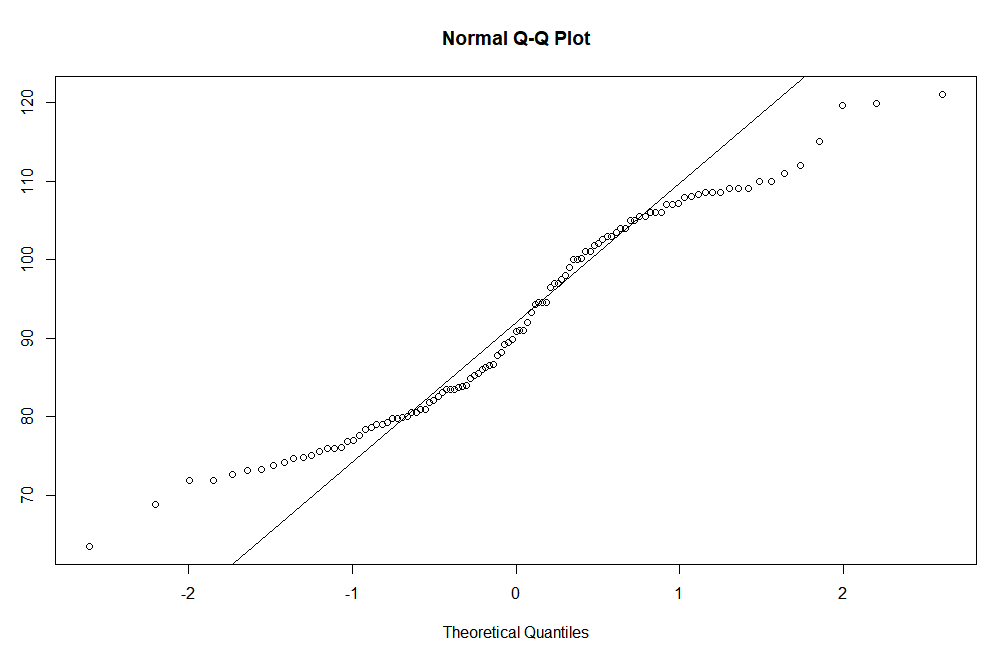
Ans: Yes the data follows normal distribution.



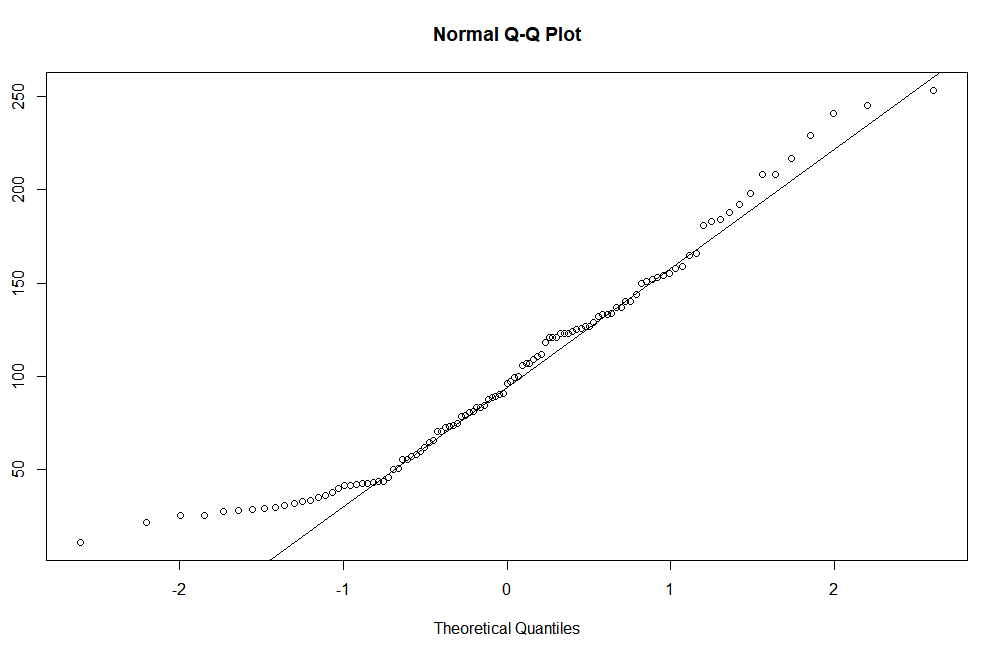
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Ans : Yes both of these follow normal distribution



Qqnorm of waist



Qqnorm of Adipose tissue

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

Ans: Z score at 90%= 1.644854

Z score at 94%= 1.880794

Z score at 60%= 0.8416212

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

Ans: T score at 95%= 2.063899

T score at 96%= 2.171545

T score at 99%= 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