|  |  |
| --- | --- |
| 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 | Categorical |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Categorical |

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

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) 0

b) 1/6

c) 2/3

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 – 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 – 0.015

Child B – 0.80

Child C – 1.95

Child D – 0.025

Child E – 0.06

Child F – 0.240

Expected no of candies for a randomly selected child = 0.015+0.80+1.95+0.025+0.06+0.240 = 3.09 ~ 3

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

* For Points,Score,Weight

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



Ans – Mean – Points : 3.5 , Score : 3.21, Weight : 17.84

Median – Points : 3.69, Score : 3.32, Weight : 17.71

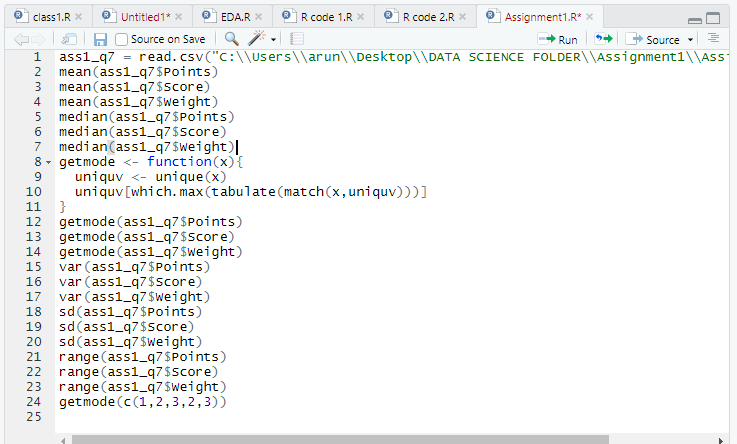
Mode – Points : 3.92, Score : 3.44, Weight : 17.02

Variance – Points : 0.285, Score : 0.957, Weight : 3.19

SD – Points : 0.534, Score : 0.978, weight : 1.786

Range – Points : 2.76 – 4.93, Score : 1.53 – 5.42, Weight : 14.5 – 22.9

Here mean = median = mode for points, score and weight which means distribution is symmetric with zero skewness.



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 : 145.33

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

**Cars speed and distance**



Ans – Skewness - Speed : -0.11 , Distance : 0.78

Kurtosis – Speed : 2.42, Distance : 3.24

Form skewness we can infer that there is little asymmetry in speed data in the left side and little asymmetry in distance in the right side

From kurtosis we can infer that the data has high number of outliers in distance as compared to speed

**SP and Weight(WT)**



**Ans** – Skewness – SP : 1.58 , WT : -0.603

Kurtosis – SP : 5.73, WT : 3.81

SP is positively skewed and is asymmetric towards right side and WT is negatively skewed and is asymmetric towards left side

SP has more number of outliers as compared to WT

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



Ans – It is positively skewed and many otliers are present in right side of data. Here mean is greater than the median.



Ans – Here the data is positively skewed with high number of outliers in the upper quartile range.

**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 – a) 94% = [198.75,201.25]

1. 98% = [198.45,201.55]
2. 96% = [198.625,201.375]

**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 = 41, median = 40.5, variance = 25.52, sd = 5.05

Here mean is approx. equal to median which means data is mostly symmetric

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

Ans – This means the distribution has zero skewness

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

Ans – Positive skewness

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

Ans – Negative skewness

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

Ans – It indicates the peakedness of the distribution

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

Ans – It means the distribution is flat

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



What can we say about the distribution of the data?

Ans – Here as per the boxplot, the distribution is negatively skewed and the data in first 25th percentile is having high standard deviation as compared to the data beyond of 75th percentile.

What is nature of skewness of the data?

Ans – Negatively skewed

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

Ans - 8

Q19) Comment on the below Boxplot visualizations?



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

Ans – Both the boxplots are near about symmetrical but the data in the first plot are very close to median as compared to the second plot and boxplot 1 has low standard deviation as compared to boxplot2.

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)

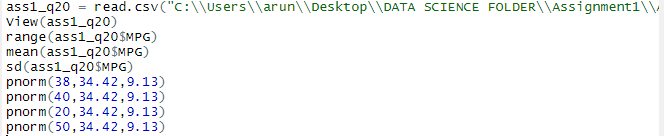
c. P (20<MPG<50)

Ans - a) P(MPG<38) = 0.65

P(MPG>38)= 1 – 0.65 = 0.35

b) P(MPG<40) = 0.72

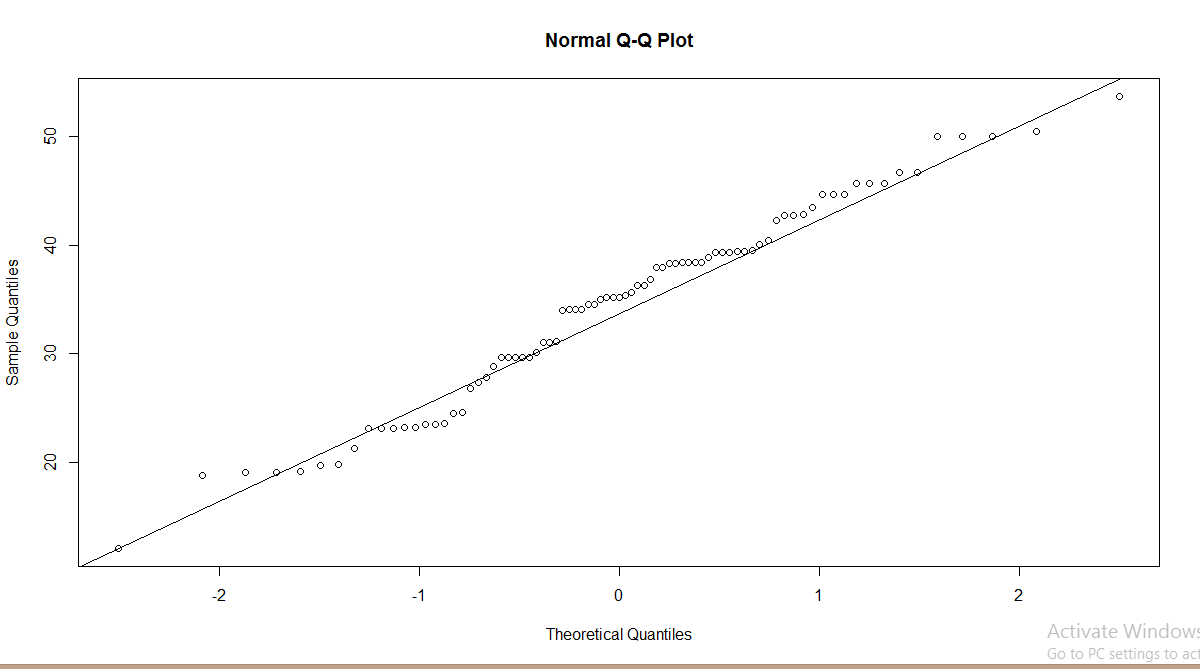
c) P(20<MPG<50) = 0.95 – 0.57 = 0.38



Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

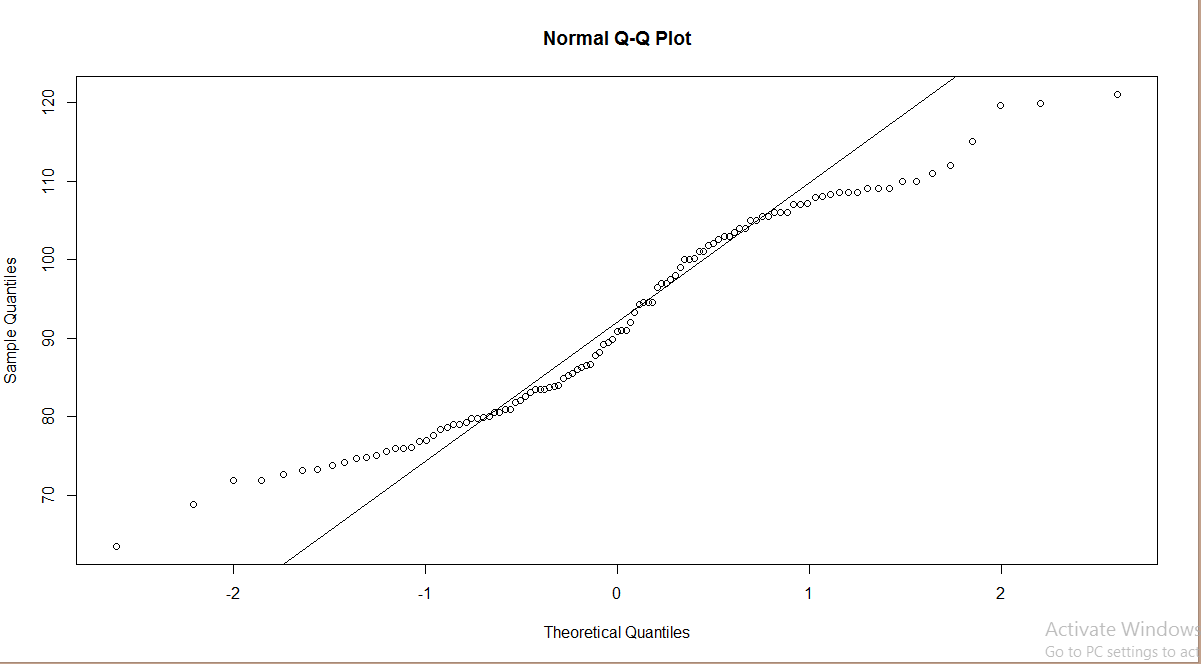
Dataset: Cars.csv



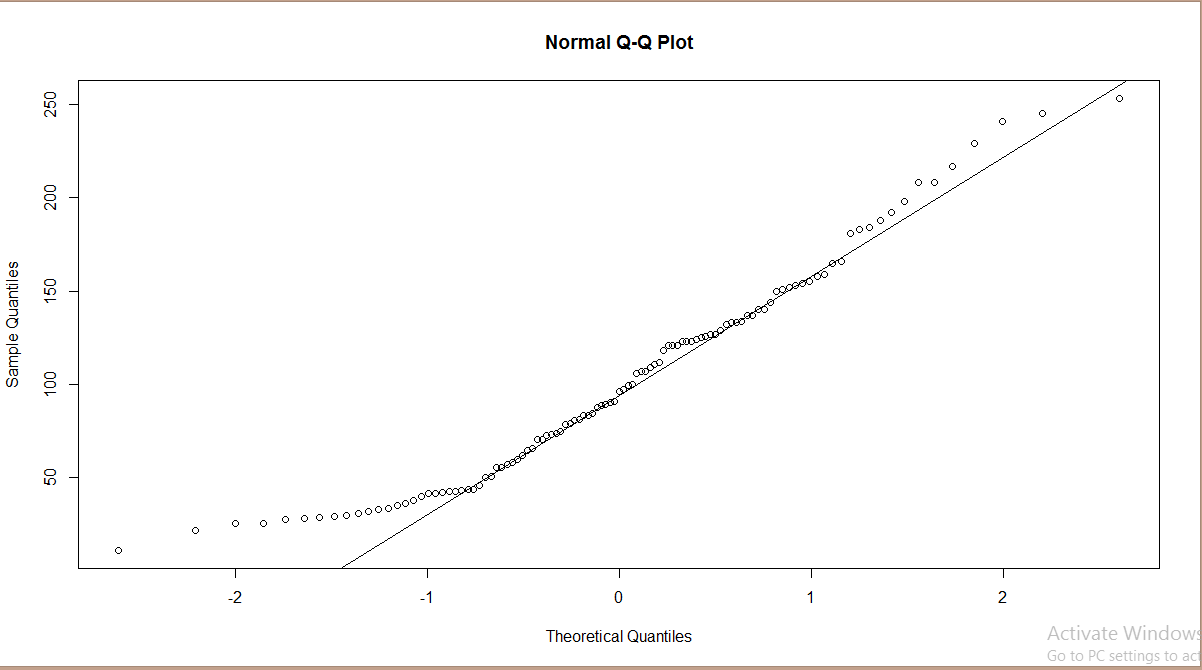
Here most of the data points are either on or near the qqline, so it can be considered that it follows the normal distribution.

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

Dataset: wc-at.csv



For waist, it is clear that most of the data points are away from qqline, hence it does not follow the normal distribution



For AT, data points are away from qqline, hence it does not follow normal distribution

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

Ans – Z(90%) = 1.64 = qnorm(0.95)

Z(94%) = 1.88 = qnorm(0.97)

Z(60%) = 0.67 = qnorm(0.75)

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

Ans – qt(0.975,24) = 2.06

qt(0.98,24) = 2.17

qt(0.995,24) = 2.79

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 – here we will use the formula to calculate t where sample mean, population mean, sample standard deviation, n is given

t= (260 - 270)/(90/sqrt(18)) = -0.47

pt(-0.47,18) = 0.3219