|  |  |
| --- | --- |
| 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 | 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 | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| 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?

Ans:- Total Possible outcome =8

HHH,HHT,HTT,THT,TTH,HTH,THH,TTT

P(two head and one tail) = 3/8 = 37.5%

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:- Total number of outcomes = 36

a) Total number of outcomes equal to 1 = 0

b) Number of outcomes less than or equal to 4 is 6. Hence, 6/36 = 1/6 = 16.66%

c) Number of outcomes sum is divisible by 2 & 3 is 6. Hence, 6/36=1/6 = 16.66%

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 balls are 7.

Total number of ways to draw 2 balls = 7!/2!(7-2)! = 7\*6/2\*1 = 21

Total number of ways to draw 2 non-blue balls = 5!/2!(5-2)! = 5\*4/2\*1 = 10

P = Total number of ways to draw 2 non-blue balls = 10/21

Total number of ways to draw 2 balls

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:- The expected number of candies for a randomly selected child is 3.09

E(x) = (1\*0.015) + (4\*0.20) + (3\*0.65) + (5\*0.005) + (6\*0.01) + (2\*0.120)

E(x) = 3.09

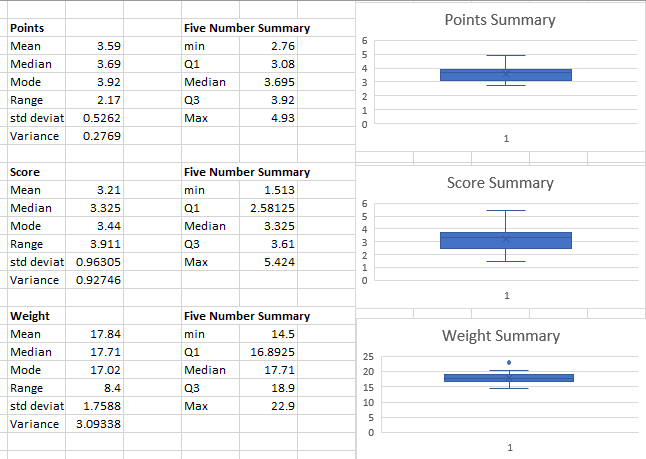
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* 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**

**Ans:-**



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:- E(X) = 1/9\*(1308)

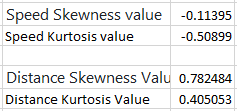
E(X) = 145.33

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

**Cars speed and distance**

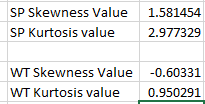
**Use Q9\_a.csv**

**Ans)**

****

**SP and Weight(WT)**

**Use Q9\_b.csv**

****

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



Ans) A) The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.

B) Median is less than mean right skewed and we have outlier on the upper side of box plot and there are 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 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:- Since we are dealing with a large sample size (n>30), we can you Z-distribution instead of T-distribution.

****

**Sample Mean = 200 pounds**

**Sample Standard Deviation = 30 pounds**

**Sample size = 2000 men**

**Z= (1 + CL)/2**

**Z1 = 1+.94/2 =0.97 = 1.88**

**Z2 = 1+.98/2 = 0.99 = 2.33**

**Z3= 1 +.96/2 = 0.98 = 1.96**

**For a 94% confidence level:**

**Confidence Interval=200±(1.88×0.6708) = 200 ± 1.261**

**For a 98% confidence level:**

**Confidence Interval=200±(2.33×0.6708) = 200 ± 1.563**

**For a 96% confidence level:**

**Confidence Interval=200±(1.96×0.6708) = 200 ± 1.315**

**For the 94% confidence level: (198.739, 201.261)**

**For the 98% confidence level: (198.437, 201.563)**

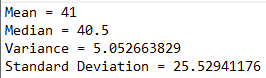
**For the 96% confidence level: (198.685, 201.315)**

**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)** Find mean, median, variance, standard deviation.

****

1. What can we say about student marks?

Ans 2) we don’t have outliers and the data is slightly skewed towards right because mean is greater than median.

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

Ans) Symmetric distribution with zero skewness. When the mean and median are equal, it implies that the data are symmetrically distributed around the central value, meaning there is no skewness.

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

Ans) Skewness and tail is towards Right

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

Ans) Skewness and tail is towards left

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

Ans) A positive kurtosis value indicates that the distribution of data has heavier tails and a sharper peak compared to a normal distribution. This could indicate the presence of outliers or a non-normal distribution with more extreme values.

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

Ans) A negative kurtosis value indicates that the distribution of data has lighter tails and a flatter peak compared to a normal distribution. This could indicate a distribution that is more spread out and less concentrated around the mean, with fewer extreme values.

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



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

The above Boxplot is not normally distributed, Median is towards the higher value.

There might be outliers influencing the data.

What is nature of skewness of the data? 🡪 The data is skewed towards left. since whisker length on the upper quadrant is higher than the data on the lower quadrant. Median will be greater than the mean since data is left skewed

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

IQR is the inter quartile range.

Here

Q1 = 10

Q3 = 18

IQR = Q3 – Q1 = 8(approx)

Q19) Comment on the below Boxplot visualizations?



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

Ans)  First there are no outliers. Second both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

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)
  3. P(20<MPG<50)

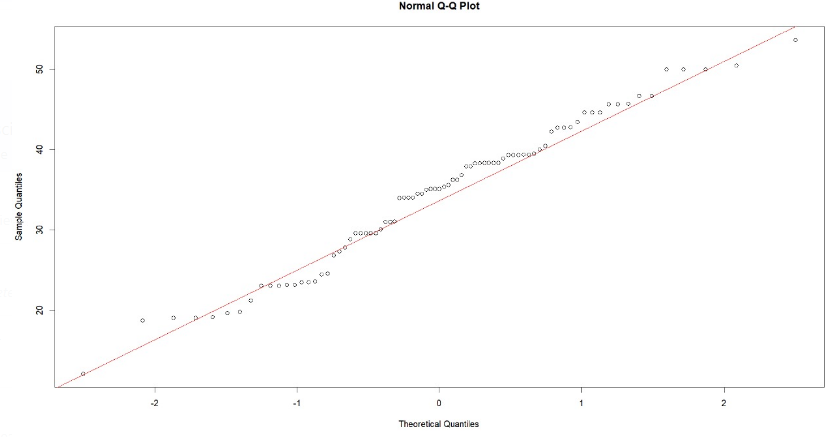
Ans)

1. P(MPG>38) = 34.7%
2. P(MPG<40) = 72.9%
3. P(20<MPG<50) = 88.3%

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

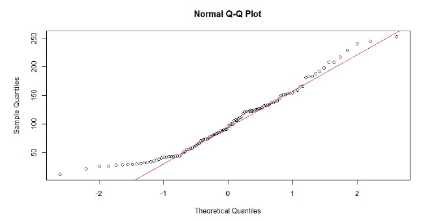
Dataset: Cars.csv

Ans) 

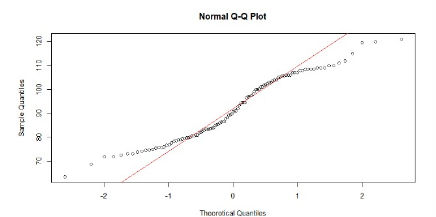
we can almost get a straight-line thus the data is normalized.

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

Dataset: wc-at.csv



Majority of the data points lie on the qqline hence Normal Distribution.



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

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

**Z= (1 + CL)/2**

For 90%

Z= (1+.9)/2 = 0.95=1.65

For 94%

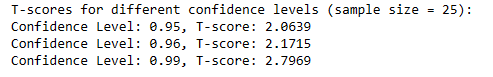
Z= (1+.94)/2 = 0.97=1.88

For 60%

Z=(1+.6)/2=0.8 = 0.84

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

Ans)



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) Sample Standard Deviation = 90 days

Sample Bulbs (n)= 18

Sample Means (x) = 260 days

Population Means = 270 days

df = n-1



T = 260 – 270/90/(18\*0.5)

T = -0.4714

Probability = 32%