|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Continuous |
| 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 | Continuous |
| Number of tickets in Indian railways | Continuous |
| Number of times married | Continuous |
| 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 | Ration |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ordinal |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Ordinal |
| 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**= If three coins are tossed; there are 2\*2\*2 = 8 possible outcomes. Those are

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

So, three outcomes give two heads and a tail (HHT, HTH, THH).

Therefore, the probability of getting two heads and a tail is 3/8 =0.375=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** = a) 0/36 = 0

b)6/36 = 0.1666 = 16.66%

c)24/36 = 0.666 = 66.66%

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**= There are 7 balls originally with 2 of them blue so the probability of the first ball not being blue is 5/7. This leaves 6 balls with 2 blue. The probability of the second ball not being blue assuming that the first wasn’t is 4/6. The probability that neither ball drawn was blue is (5/7)\*(4/6)=20/42=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 = 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.

**ANS =**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Mean** | **Median** | **Mode** | **Variance** | **Standard Deviation** | **Range** |
| Points | 3.59656 | 3.695 | 3.92 | 0.28588 | 0.5262 | 2.17 |
| Score | 3.21725 | 3.325 | 3.44 | 0.9274 | 0.9630 | 3.91 |
| Weigh | 17.8488 | 17.71 | 17.02 | 3.0933 | 1.7588 | 8.40 |

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 = 145**

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

**Cars speed and distance**

**ANS =**

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **Index** | **-3.8E-17** | **-1.2** |
| **Speed** | **-0.117** | **-0.5099** |
| **Dist** | **0.8068** | **0.4050** |

**SP and Weight(WT)**

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



**ANS**= From the above histogram we can conclude that the data is normally distributed and it is positively skewed as the tail is shifted to the right. The data is moderately skewed. The mean is greater than the median.



**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**=

94% confidence interval = (198.73 – 201.26)

96% confidence interval = (198.62 – 201.37)

98% confidence interval = (198.43 – 201.56)

**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.

**ANS=**

Mean = 41

Median = 40.5

Variance = 24.888

Standard deviation = 4.98

1. What can we say about the student marks?

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

**ANS** = If the mean and median of the data are equal then it is symmetrical and normally distributed with zero skewness.

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

**ANS** = When the mean > median, the distribution is right or positively skewed.

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

**ANS** =When the median > mean, the distribution is left or negatively skewed.

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

**ANS** = A distribution with a positive kurtosis value indicates that the distribution has heavier tails than the normal distribution.

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

**ANS** = A distribution with a 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?

**ANS=**First 25% of the data lies between 2-10

50% of the data is ≥10 to 18

Median is approximately at point 15

Last 25% of the data ≤18

What is nature of skewness of the data?

**ANS=**Distribution of the data is negatively or left skewed.

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

**ANS=**  IQR ≈ 8

Q19) Comment on the below Boxplot visualizations?



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

**ANS=**

1. The data is normally distributed in both the boxplots.
2. The data in the boxplot 2 is widely distributed than boxplot 1.
3. In boxplot 1, the lower extreme is approx. 245 and the higher extreme is approx. 285 and for boxplot 2 the lower and higher extreme values are 225&315 respectively.
4. The median value for both the boxplot is approx. same.
5. There are no outliers present in any boxplot.
6. The IQR of boxplot 1 is 40.
7. The IQR of boxplot 2 is 90.

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

**ANS=**

* 1. P(MPG>38) = 0.33 =33%
  2. P(MPG<40) = 0.2546 = 25.46%

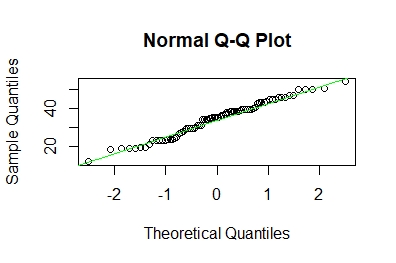
c. P (20<MPG<50) = 0.901 = 90.1%

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

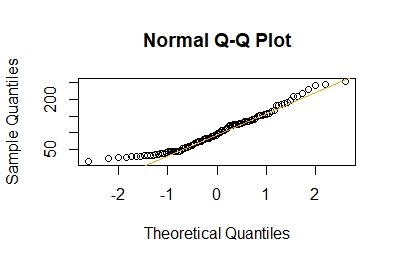
**ANS=**  As we can see in the attached plots, MPG column of Cars data set follows normal distribution.



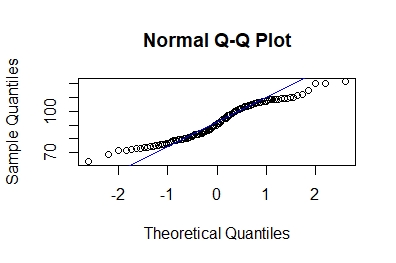
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

**ANS= Q-Q plot of Adipose Tissue =** It follows normal distribution as can see a straight line in the below plot.



**Q-Q plot of Waist Circumference =**  It follows normal distribution as can see a straight line in the below plot.



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

**ANS=** To calculate z score we can use the (formula = 1+CL/2)

**z score of 60% CI = 0.84**

**z score of 90% CI = 1.64**

**z score of 94% CI = 1.88**

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

**ANS=** T score of 95% CI = 2.064

T score of 99%CI = 2.797

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 =**

t - Statistics for the data is given as follows;

x = Mean of the sample of bulbs = 260

μ = Population mean = 270

s = Standard deviation of the sample = 90

n = Number of items in the sample = 18

t=-10/23.23

t = - 0.471

Degrees of freedom is n – 1 = 18-1= 17

In R code,

pt (-0.471,17)=0.321

Probability is 32%