|  |  |
| --- | --- |
| 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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ (Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| 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) P = **0**

b) P = 6/36 = **1/6**

c) P = 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?

Ans: - P = **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: - Expected number of candies for a randomly selected child = **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: - 1) Mode for Points = 3.07 & 3.92 (Bimodal)

2) Mode for Score = 3.44

3) Mode for Weigh = 17.02 & 18.90 (Bimodal)

For Points data, Mean < Median hence indicate Negative Skewness. Also, there is no outliers in Points data.

For Score data, Mean < Median hence indicate Negative Skewness. There are two outliers in Score data.

For Weigh data, Mean > Median hence indicate Positive Skewness. There is only one outlier in Weigh data.

(For rest of the answer’s please find the attached R file named “Ans\_Q7”.)

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 of the Weight of that patient = **145.3333**

{Expected Value = (1/9)108 + (1/9)110 + (1/9)123 + (1/9)134 + (1/9)135 + (1/9)145 + (1/9(167) + (1/9)187 + (1/9)199 = 145.3333}

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

a) Car’s speed and distance

Use Q9\_a.csv

Ans: -

1) For car’s speed data, Skewness = -0.1139548 hence indicate Negatively Skewed data. Also, there is no outliers in speed data.

2) For car’s distance data, Skewness = 0.7824835 hence indicate Positively Skewed data. There is one outlier in distance data.

(For rest of the answer’s please find the attached R file named “Ans\_Q9\_a”.)

b) SP and Weight (WT)

Use Q9\_b.csv

Ans: -

1) For car’s speed data, Skewness = 1.581454 hence indicate Positively Skewed data.

2) For car’s weight data, Skewness = -0.6033099 hence indicate Negatively Skewed data.

(For rest of the answer’s please find the attached R file named “Ans\_Q9\_b”.)

Q10) Draw inferences about the following boxplot & histogram



Ans: -

1) Given Histogram shows asymmetric distributed data which indicating the Positive skewed or right-skewed data.

2) Given Boxplot shows asymmetric distributed data which indicating the Positive skewed or right-skewed data, also from boxplot there are 7 outliers in this data which are making this data more asymmetric.

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: -

1. Confidence Interval for 94% = (134.8507, 265.1492)
2. Confidence Interval for 98% = (122.6512, 277.3487)
3. Confidence Interval for 96% = (130.1535, 269.8464)

(Refer the attached IPYNB.file named “Ans\_CI”.)

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. Please find the attached R file named “Ans\_Q12”.
2. For student scores data, Mean > Median hence indicate Positive Skewness or right-skewed data.

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

Ans: - If mean, median of data are equal then it shows the normal distribution or the data is symmetrically distributed. And the skewness is equal to zero.

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

Ans: - If mean > median then it shows that the data is asymmetrically distributed and mass of the distribution concentrated on left side. Which indicates the data is Positively skewed or right-skewed.

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

Ans: - If median > mean then it shows that the data is asymmetrically distributed and mass of the distribution concentrated on right side. Which indicates the data is Negatively skewed or left-skewed.

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

Ans: - Positive values of kurtosis indicate that distribution is peaked and possesses thick tails.

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

Ans: - Negative values of kurtosis indicate that distribution is flatter (less peaked) and possesses lower tails.

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



a) What can we say about the distribution of the data?

Ans: - Data is asymmetrically distributed.

b) What is nature of skewness of the data?

Ans: - Negatively skewed or left-skewed.

c) What will be the IQR of the data (approximately)?   
Ans: - IQR = 18-10 = 8

Q19) Comment on the below Boxplot visualizations?



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

Ans: -

* The medians for boxplot–1 and boxplot-2 are at the same level.
* The values for all the four quartile groups for boxplot–1 is comparatively lesser the quartile groups value of boxplot-2.
* The IQR for boxplot-2 is greater than the IQR of boxplot-1.

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) = 0.3475
  2. P(MPG<40) = 0.7293
  3. P (20<MPG<50) = 0.8988

(Refer the attached IPYNB.file named “Ans\_Q20”.)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

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

Dataset: wc-at.csv

Ans: -

1. Mean for Standardize MPG = 0.0

Standard Deviation for Standardize MPG = 1.0

Hence MPG data from Cars.csv dataset follows **Normal Distribution**.

1. Mean for Standardize Waist = -0.0

Standard Deviation for Standardize Waist = 1.0

Hence Waist Circumference (Waist) from wc-at data set follows **Normal Distribution**.

Mean for Standardize AT = -0.0

Standard Deviation for Standardize AT = 1.0

Hence Adipose Tissue (AT) from wc-at data set follows **Normal Distribution**.

(Refer the attached IPYNB.file named “Ans\_Q21”.)

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

Ans: -

1. Z scores of 90% confidence interval = 1.64
2. Z scores of 94% confidence interval = 1.88
3. Z scores of 60% confidence interval = 0.84

(Refer the attached IPYNB.file named “Ans\_CI”.)

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

Ans: -

1. t scores of 95% confidence interval = 2.06
2. t scores of 95% confidence interval = 2.17
3. t scores of 95% confidence interval = 2.79

(Refer the attached IPYNB.file named “Ans\_CI”.)

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-score = -0.4714,

df = 17,

P(X<260) = 0.6783

P(X>260) = 0.3216