|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | descriptive |
| Results of rolling a dice | descriptive |
| Weight of a person | continues |
| Weight of Gold | Continues |
| Distance between two places | continues |
| Length of a leaf | continues |
| Dog's weight | continues |
| Blue Color | descriptive |
| Number of kids | descriptive |
| Number of tickets in Indian railways | descriptive |
| Number of times married | descriptive |
| Gender (Male or Female) | Descriptive(nominal) |

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 | 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: -**

The possible outcomes are:

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

Probability= (Number of Favorable Outcomes)/(Total Number of Possible Outcomes)

Probability = 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

When rolling two dice, each die has 6 possible outcomes (1 to 6).

let's analyze each case:

**Ans: -**

a) Equal to 1:

Probability = (Number of Favorable Outcomes) / (Total Number of Possible Outcomes)

Probability = 0 / 36

b) Less than or equal to 4:

The combinations that result in a sum less than or equal to 4 are: (1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (3, 1).

Probability = (Number of Favorable Outcomes) / (Total Number of Possible Outcomes)

Probability = 6 / 36 = 1 / 6

1. Sum is divisible by 2 and 3:

(1,3),(2,2),(3,1),(1,1),(1,2),(2,1) = 6 Outcomes,

Probability = (Number of Favorable Outcomes) / (Total Number of Possible Outcomes)

Probability = 6/ 36 = 1 / 6

So, the probabilities are:

a) Probability of sum equal to 1: 1/36

b) Probability of sum less than or equal to 4: 1/6

c) Probability of sum divisible by 2 and 3: 1/12

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: -** First, let's calculate the total number of ways to choose 2 balls out of the 7 in the bag:

Total ways to choose 2 balls = C(7, 2) = 21

Now, let's find the number of ways to choose 2 balls such that none of them is blue. There are 5 non-blue balls (2 red and 3 green), and you want to choose 2 out of these 5:

Ways to choose 2 non-blue balls = C(5, 2) = 10

Now, calculate the probability:

Probability of choosing 2 non-blue balls = (Number of ways to choose 2 non-blue balls) / (Total number of ways to choose 2 balls)

Probability = 10 / 21

So, the probability that none of the balls drawn is blue is 10/21 or approximately 0.4762 .

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: -**

To calculate the expected number of candies for a randomly selected child, you can use the following formula:

Expected Value = Σ (X \* P(X))

Where:

- X represents the possible values (candies count in this case).

- P(X) represents the probability associated with each value.

Let's calculate it step by step:

Child A:

- X (candies count) = 1

- P(X) = 0.015

Child B:

- X (candies count) = 4

- P(X) = 0.20

Child C:

- X (candies count) = 3

- P(X) = 0.65

Child D:

- X (candies count) = 5

- P(X) = 0.005

Child E:

- X (candies count) = 6

- P(X) = 0.01

Child F:

- X (candies count) = 2

- P(X) = 0.120

Now, calculate the expected value:

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

Expected Value = 0.015 + 0.80 + 1.95 + 0.025 + 0.06 + 0.24

Expected Value = 3.105

So, the expected number of candies for a randomly selected child is 3.105 candies.

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: -** Let's calculate the Mean, Median, Mode, Variance, Standard Deviation, and Range for the given dataset, which appears to be automotive data with three variables: Points, Score, and Weight.

Mean (Average):

Mean Points = (Sum of Points) / (Total Number of Data Points)

Mean Score = (Sum of Score) / (Total Number of Data Points)

Mean Weight = (Sum of Weight) / (Total Number of Data Points)

Mean Points = 3.5969

Mean Score = 3.2172

Mean Weight = 17.8488

Median (Middle Value):

To calculate the median, we'll first arrange the data in ascending order for each variable and then find the middle value. Since there are 32 data points (an even number), the median is the average of the 16th and 17th values when sorted.

Points median= 3.695

Score median= 3.325

Weigh median= 17.710

Mode (Most Frequent Value):

The mode is the value that appears most frequently in the dataset. It can be different for each variable. To find the mode, we need to count the occurrences of each value.

- For Points 3.92

- For Score 3.44

- For Weight 18.90

Variance:

To calculate the variance, we'll find the average of the squared differences between each data point and the mean for each variable.

Variance Points = 0.285881

Variance Score = 0.957379

Variance Weight = 3.193166

Standard Deviation:

The standard deviation is the square root of the variance for each variable.

Standard Deviation Points = 0.534679

Standard Deviation Score = 0.978457

Standard Deviation Weight = 1.786943

Range:

The range is the difference between the maximum and minimum values for each variable.

Range Points = Max(4.93) - Min(2.76) = 2.17

Range Score = Max(5.424) - Min(1.513) = 3.911

Range Weight = Max(22.9) - Min(14.5) = 8.4

Comments/Inferences:

1. The mean Points value is approximately 3.597, indicating that, on average, the Points variable is around this value.

2. The mean Score value is approximately 3.217, indicating that, on average, the Score variable is around this value.

3. The mean Weight value is approximately 17.849, indicating that, on average, the Weight variable is around this value.

4. The medians provide an alternative measure of central tendency. The medians for Points = 3.695, Score = 3.325, Weigh=17.710, respectively.

5. The standard deviations for Points, Score, and Weight indicate the degree of spread or dispersion of the data around the mean. Lower standard deviations suggest less variability, while higher standard deviations suggest more variability.

6. The range for Weight is the largest, indicating the greatest spread among the three variables. This suggests that there is more variability in the Weight data compared to Points and Score.

7. The lack of a clear mode suggests that there are no values that significantly dominate the dataset for Points, Score, and Weight.

8. To draw more specific inferences or insights from this data, it would be helpful to understand the context and purpose of collecting this information.

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: -** To calculate the expected value (also known as the mean) of the weights of patients at the clinic, you need to sum up all the weights and then divide by the total number of patients.

The weights of the patients are: 108, 110, 123, 134, 135, 145, 167, 187, 199.

Expected Value (Mean) = (Sum of all weights) / (Total number of patients)

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

Expected Value = 1308 / 9

Expected Value ≈ 145.33 (rounded to two decimal places)

So, the expected value of the weight of a randomly chosen patient is approximately 145.33 pounds.

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

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

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



**Ans: -** - The data is right skewed.

- There are outliers at upper side

The most of the data points are Concerted in the range 50-100 with frequency 200. And least range of weight is 400 Somewhere around 0-10. Around 0-10.

So the expected value the above distribution is 75.

Skewness -we can notice a long tail towards right so it is heavily right skewed.

Median is less than mean right skewed and we have outlier on the upper side of the box plot and there is 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?

**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) Mean: 41.0

Median: 40.5

Variance: 24.11111111111111

std deviation: 4.910306620885412.

2) Mass of students marks between 38-42.

Skewness (1.52) is positive because mass of marks in left side of plot.

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

**Ans: -** Data is normalized and there is no skewness. Symmetrical

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

**Ans: -** Negative Skewness implies mass of the Distribution concentrated on right side. Right Skewed

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

**Ans: -** Positive Skewness implies mass of the Distribution concentrated on left side. Left Skewed.

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

**Ans: -** Positive kurtosis value indicates that thinner peak and wider tails. The data is normally distributed and kurtosis value is 0.

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

**Ans: -** Negative kurtosis value indicates that wider peak and thinner tails. The distribution of the data has lighter tails and a flatter peaks 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: -** Not normally distributed

What is nature of skewness of the data?

**Ans: -** Negative skewness

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

Q19) Comment on the below Boxplot visualizations?



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

**Ans: -** By observing both the plots whisker’s level is high in boxplot 2.Mean and median are equal hence distribution is symmetrical.

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

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

Dataset: wc-at.csv

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

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

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