|  |  |
| --- | --- |
| 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 | Nominal/Categorical |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Nominal/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 | Ratio |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Nominal |
| Fahrenheit Temperature | Ratio |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Interval |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Interval |
| Religious Preference | Ordinal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Interval |

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?

Possible outcomes = {HHH, HHT, HTH, THH, TTH, THT, HTT, TTT}

Getting two heads and one tail = { HHT, HTH, THH} = 3

Probability of getting two heads and one tail = 3/8 = 0.375

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

Total possible outcomes = 6^2 = 36

1. P(a) = 0/36 = 0
2. P(b) = 5/36 = 0.139 …….. {(1,1),(1,2),(1,3),(2,1),(2,2) }
3. P(c) = 6/36 = 0.167……….{(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)}

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?

Total number of balls = (2 + 3 + 2) = 7  
Then, n(S) = Number of ways of drawing 2 balls out of 7  
 =7C2​  
 =(7x6) ​/ (2x1)  
 =21  
Let E = Event of drawing 2 balls, none of which is blue.  
n(E)= Number of ways of drawing 2 balls out of (2 + 3) balls.  
 =5C2​  
 =(5x4)​ / (2x1)  
 =10  
probability = 10/21 = 0.476

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

Expected number of candies for a randomly selected child

=  1 \* 0.015  + 4\*0.20  + 3 \*0.65  + 5\*0.005  + 6 \*0.01  + 2 \* 0.12

= 0.015 + 0.8  + 1.95 + 0.025 + 0.06 + 0.24

=   3.09

Expected number of candies for a randomly selected child is 3.

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.

|  |  |  |  |
| --- | --- | --- | --- |
| Sr.no | Points | Score | Weigh |
| Mean | 3.59 | 3.21 | 17.84 |
| Median | 3.69 | 3.32 | 17.71 |
| Mode | 3.07 | 3.44 | 17.02 |
| Variance | 0.28 | 0.95 | 3.19 |
| Std. Deviation | 0.53 | 0.97 | 1.78 |
| Range | 2.17 | 3.91 | 8.4 |

Points – Variance is low data points are closely packed.

Score - Data for scores is negatively skewed (Mean<Median<Mode)

Weigh – Variance between values is high.

Data is positively skewed (Mean>Median>Mode)

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?

Probablity of selecting one patient = 1/9 = 0.111

Expected weight = sum of (probability of each patient\* weight value of each patient)

=0.11\*108+ 0.11\*110+ 0.11\*123+0.11\* 134+0.11\* 135+0.11\* 145+0.11\* 167+0.11\* 187+0.11\* 199

ie. 0.111\* 1308 = 145.188 pounds

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

**Cars speed and distance**

|  |  |  |
| --- | --- | --- |
|  | Skewness | Kurtosis |
| Car Speed (speed) | -0.11 | -0.50 |
| Distance (dist) | 0.80 | 0.40 |

Skewness for car speed is negatively skewed where as it is positively skewed for

Distance.

While negative kurtosis is seen in car speed. Which means the peak is flatter and distribution is more spread out.

Only fewer values are located near the mean and all other values are spread out along the tails of distribution

**SP and Weight(WT)**

|  |  |  |
| --- | --- | --- |
|  | Skewness | Kurtosis |
| SP | 1.61 | 2.97 |
| Weight (WT) | -0.61 | 0.95 |

The SP has a positive skewness and positive kurtosis.

Weight has negative skewness which means the data distribution has a longer tail on the left side and the outliers on left side are further away from the mean but the outliers on right side are closer to mean.

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



* The data is positively skewed
* Many outliers are present in the data
* 0-100 data bin has the highest frequency
* The mean is greater than the median
* The distribution is not symmetric

**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?

94% confidence interval – 143.57 to 256.42 pounds

98% confidence interval – 130.20 to 269.79 pounds

96% confidence interval – 138.38 to 261.61 pounds

**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.

Mean – 41

Median – 40.5

Variance – 25.5

Standard deviation – 5.05

1. What can we say about the student marks?

Majority students have performed well in the test as the marks are close to the overall mean. Barring some outliers most of the scores are consistent and symmetrically distributed.

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

Zero skewness is found if the mean = median ie. Distribution is symmetric.

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

If the mean is greater than the median, the distribution is positively skewed.

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

If the median is greater than the mean, the distribution is negatively skewed.

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

Positive values of kurtosis indicate that distribution is peaked and has thick tails.

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

Negative values of kurtosis indicate that a distribution is flat and has thin tails.

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



What can we say about the distribution of the data?

The data distribution is asymmetric.

What is nature of skewness of the data?

The data has negative skewness.

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

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.

IQR of boxplot 2 is greater than boxplot 1

Both plots have same median values

Boxplot 2 has a higher range than that of boxplot 1

Boxplot 2 has longer tails than 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

* 1. P(MPG>38) = 34.76%
  2. P(MPG<40) = 72.93%

c. P (20<MPG<50) = 89.89%

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

### MPG of cars does not follow normal distribution. The data is negatively skewed.

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

Dataset: wc-at.csv

Both Adipose tissue and Waist circumference does not follow normal distribution.

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

90% CI = 1.28

94% CI = 1.55

60% CI = 0.25

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

95% CI = 1.71

96% CI = 1.82

99% CI = 2.49

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

The probability of 18 selected bulbs having avg life of no more than 260 days is 67.83%