|  |  |
| --- | --- |
| 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) | Interval |
| Sales Figures | Ratio, Interval |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal-Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ordinal |

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

Ans: {HHH, HHT, HTT, HTH, THH, TTH, THT, TTT}

i.e. 3/8 = 0.375

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1

Is Zero, because dice not having value zero instead starts with p(1,1)

1. Less than or equal to 4

Favorable outcomes = (1,3), (2,2), (3,1), (1,1), (1,2), (2,1)

i.e., P = 6/36 = 1/6= 0.166

1. Sum is divisible by 2 and 3

Favourable outcomes = (1 ,5), (3, 3), (4, 2), (5, 1), (6, 6)

i.e., P = 5/36

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:

Favorable outcomes = 21 (for total 7 balls) among 2 are blue so favorable outcomes for neither balls are drawn blue will be

P = (5/7) \* (4/6) = (5x4) / (7x6) = 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:

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\*0.120)

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

= 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**

File attached

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 from above probability is 1/9.

Hence,

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

= 11.88+12.1+13.53+14.74+14.85+15.95+18.37+20.57+21.89

= 143.88 would be the Expected value of the weight of that patient.

**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**

**File attached**

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



1) Chick weight data is positively skewed.

2) Most of the chick weight is between 50 to 130 frequencies.

3) Most of the chick weight is under 50 to 200.



1. There are outliers in the data at upper extreme.
2. Data is rightly skewed.

**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: attached

**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: attached

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

Ans:

When mean = median then the skewness will be equal to 0. i.e. Normal Distribution.

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

Ans:

The data will be Positively skewed.

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

Ans:

The data will be negatively skewed.

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

Ans:

A Positive kurtosis tell that data distribution is peaked and Heavy tails with more data.

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

Ans:

A Negative kurtosis tell that data distribution is flatter peak and lighter data in the tail.

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



What can we say about the distribution of the data?

Ans: As on the y-axis it plots from 2 to 18 , it shows 25% of the data 1st quartile is

Around 10 , where in 75% of the data is 18 and median which is 2nd quartile is around 14.8.

What is nature of skewness of the data?

Ans: Data is negatively skewed as median is close to upper whisker of the box.

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

Ans: Q3 = 18.1 , Q1 = 10 Hence, Q3 – Q1 = 18.1-10 = 8.1 = IQR.

Q19) Comment on the below Boxplot visualizations?



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

Ans: Box plot 1 and 2 have different IQR but Normally distributed as median for both is in center of both whiskers.

BP1’s min value is higher than the BP2 min value and also BP1’s max value is lower than the BP2’s max value.

BP1 covers lowest IQR wherein, BP2 covering Higher IQR. However median for both is equal.

Based on the observation, even if both Box plots have equal median and normally distributed. But, since Box plot 2 covering higher range of data is most accurate.

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)

Ans: attached

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: attached

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

Ans: attached

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

Ans: attached

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: attached