|  |  |
| --- | --- |
| 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 | Discrete- Nominal |
| High School Class Ranking | Discrete- Nominal |
| Celsius Temperature | Continuous- Interval |
| Weight | Continuous- Ratio |
| Hair Color | Discrete- Ratio |
| Socioeconomic Status | Continuous- Interval |
| Fahrenheit Temperature | Continuous- Ratio |
| Height | Continuous- Ratio |
| Type of living accommodation | Discrete- Ordinal |
| Level of Agreement | Discrete- Ordinal |
| IQ(Intelligence Scale) | Discrete- Ordinal |
| Sales Figures | Discrete- Ordinal |
| Blood Group | Discrete- Ratio |
| Time Of Day | Continuous- Interval |
| Time on a Clock with Hands | Continuous- Interval |
| Number of Children | Discrete- Interval |
| Religious Preference | Discrete- Ratio |
| Barometer Pressure | Discrete- Interval |
| SAT Scores | Discrete- Ratio |
| Years of Education | Discrete- Nominal |

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

Sol: Sample= HHH, HHT, HTH, THH, TTT, TTH, THT, HTT

Total Sample= 8

Prob = P ((HHT) + (HTH) + (THH)) / total sample

= 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

Sol: a) No outcome i.e., 0/36 Probability is 0

b) (1,3), (3,1), (1,2) = 3/36 = 0.083

c) 6/36 = 1/6 = 0.17

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?

Sol: (2R, 3G, 2B) = 7 balls in bag

P of 1st ball not being blue is 5/7

P of 2nd ball not being blue is 4/6

5/7 \* 4/6 = 10/21 = 0.48

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

Sol: 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

= 3.09

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points, Score, Weight

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

Use Q7.csv file

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?

Sol: Expected Value = ∑ (probability  \* Value)

there are 9 patients

Probability of selecting each patient = 1/9

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

= (1/9) \* (108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

= (1/9) \* (1308)

= 145.33

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



Sol: Histogram- Noticed a long tail towards right so it is highly skewed at right.

Boxplot- Noticed a outlier at upper end.

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?

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

Sol: If mean and median of the data is equal then the skewness is **ZERO.**

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

Sol: The mean of **positively skewed** data will be greater than median.

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

Sol: The mean of **negatively skewed** data will be greater than median.

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

Sol: Distribution of data is peaked and possesses thick tails.

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

Sol: Data has lighter tail than the normal distribution.

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



What can we say about the distribution of the data?

Sol: Data is not normally distributed as the median is towards the higher side.

What is nature of skewness of the data?

Sol: Whisker range of minimum is greater than maximum, the data is skewed towards left.

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

Sol: IQR= Q3 – Q1 = 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.

Sol: Data has no outliers, both data shares the same mean approx. in range of 260 to 275, data is normally distributed with zero skewness.

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