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

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

SOLUTION:

**Tossing of the coin is Nominal**

**S = {HHH, HHT, HTH, THH, TTH, THT, HTT, TTT}**

**n(S) = 8**

**The probability that two heads and one tail are obtained = Event (A)**

**Event (A) - {HHT, HTH, THH} = n (A) = 3**

**P(A) = n (A) / n (S) = 3/8 = 0.375**

**% = P(A)/100% = 37.5%**

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

SOLUTION:

1. N(S) = 36

Event A = NULL EVENT, There is no instance where the Event (A) is taking place.

N (A) = 0

P(A) = N(A)/N(S)

**P (A) = 0**

1. Event B where the probability that sum is <= 4

N(B) = 6

**P(B) = N(B)/N(S) = 6/36 = 1/6**

1. Event C where the probability that sum is divisible by 2 & 3

N(C) = 30

**P(C) = 30/36 = 5/6**

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?

SOLUTION: A bag contains 2 red, 3 green and 2 blue balls. Two balls are drawn at random = 7c2 = 21 Combinations.

To Calculate the combinations of balls drawn use nCr

Note: If using fx-991ES Calculator use ***Shift + division Operator****.*

Let A be the event where no one of the balls drawn is Blue

So total no.of balls are = 3+2 = 5

N(A) = 5c2 = 10

**P(A) = 10/21 = 0.47**

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

Child C - probability of having 3 candies = 0.65

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**

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?

**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**



**Histogram: It is a positive skewed or right skewed data distribution.**

**Boxplot: The data is denser at the lower range of value. The variability of the lower quartile is much small than the that of the upper quartile which is indicated by the asymmetric whiskers. There are 7 outliers in the dataset.**



**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?

**The data distribution is symmetric when mean and median are equal and the distribution has zero skewness.**

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

**When mean is greater than median then the distribution is positively skewed.**

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

**When median is greater than mean then the distribution is negatively skewed.**

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

**Positive kurtosis value indicates that the distribution is peaked and possess thick tails.**

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

**Negative kurtosis value indicates that the distribution is flat and possess 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 spread over a wide range which is interpreted through the width of the boxplot. The middle quartile represents the midpoint of the data which is close to 15.5 approx which represents the whole data distribution. The value of the midian is closer to the quartile of the data spectrum.**

What is nature of skewness of the data?

**The data negatively skewed.**

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

**8.2**

Q19) Comment on the below Boxplot visualizations?



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

**Boxplot 1: The box plot is comparatively shorter which suggests that the data spread within a concise range of values.**

**Boxplot 2: The data is spread across a wide range of values indicating variance in the data.**

**The medians are at the same level but the data distribution is completely different.**

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