|  |  |
| --- | --- |
| 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 |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | 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 | Interval |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| 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 | Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Soln:

Sample Space: 23=8

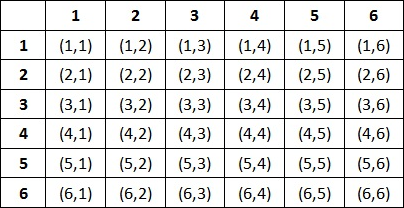
No. of Events E={HHT,HTH,THH}=3

Required 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

Soln:



Total No. of outcomes: 62=36

(a). P(X=1) = 0

(b). E (X=less than or equal to 4) = {(2,2),(3,1),(1,3),(1,1),(1,2),(2,1)}=6

P (X=less than or equal to 4) = 6/36=1/6.

(c). E (X= Sum is divisible by 2 and 3) ={(3,3),(2,4),(4,2),(6,6),(1,5),(5,1)}=6

P(X= Sum is divisible by 2 and 3)=6/36=1/6

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?

Soln: Total No. of Events = =21

Interested events=

Probability that none of the balls is blue =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

Soln.:

Expected number = E(x)

E(x)=1\*0.015+4\*0.2+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**

**Answers:**

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: Probability of selecting each patient=1/9

E(X)= 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

= 145.33

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

Cars speed and distance

**Use Q9\_a.csv**

**Answer: a.** Speed

Skewness = -0.1175

Kurtosis = -0.5089

Distance

Skewness = 0.8068

Kurtosis = 0.4050

**SP and Weight (WT)**

**Use Q9\_b.csv**

**Answer: b.** SP

Skewness = 1.6114 Kurtosis = 2.9773

WT

Skewness = -0.6147 Kurtosis = 0.9502

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



Ans:

1. Majority of the chicks has weight in range 50-100, followed by 100-150 and 150-200
2. The data is positively skewed.



Ans:

1. Data has lots of outliers towards upper extreme.

**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.

Ans: Mean=41

Median=40.5

Variance=25.53

Standard Deviation=5.0526

1. What can we say about the student marks?

Ans: Mean > Median, implies that the distribution is slightly skewed towards right. No Outliers are present.

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

Ans.: Skewness=0, perfectly symmetric bell-shaped curve.

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

Ans.: Skewness= Positive, data is distributed more on left.

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

Ans.: Skewness= Negative, data is distributed more on right.

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

Ans.: Peakedness (sharp peak) and less variance.

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

Ans.: Less peakedness (broad peak) and more variance.

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



What can we say about the distribution of the data?

Ans.: The data is not symmetric. Data is more concentrated towards right side.

What is nature of skewness of the data?

Ans.: Skewness = Negative

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

Ans.: IQR Data is 8 (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.

Ans.:

1. Data is normally distributed. No outliers. Median around 265.5. Comparatively, the first boxplot has less range.
2. Data is normally distributed. No outliers. Median around 265.5. Comparatively, the second boxplot has more range.

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

**Answer:** MPG of cars is not Normal Distribution.

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

Dataset: wc-at.csv

**Answer:**  Adipose tissue (AT) and Waist circumference do not follow Normal Distrbution.

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