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

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

Ans:

Total possible outcomes = [HHH, TTT, HHT, HTH, THH, TTH, THT, HTT]

P(getting 2H and 1T)=P(HHT)+P(HTH)+P(THH)

= **=**

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

Ans:

1. When 2 dice are rolled, their sum cannot be equal to 1. The minimum sum of numbers that can occur is (1,1). So the probability of getting a sum (equal to 1) is 0

b) Total possible outcomes : (1,1) (1,2) (1,3) (3,1) (2,1) (2,2)

Total outcomes : 6\*6 =36

P(sum<=4) =  **=**

c) Total outcomes: 6\*6 =36

**Total possible outcomes=(1,5) (2,4) (3,3) (6,6)**

**Therefore, Probability = 4/36 = 0.111**

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 no of balls in bag=2(red)+3(green)+2(Blue) = **7**

Number of non-blue balls = 2+3 = **5**

Probability = (Number of non-blue balls) / (Total no of balls in bag)

Number of ways to draw 2 non blue ball = C(5,2) = 5!/(2!(5-2)!) = **10**

Total outcomes =C(7,2) = 7!/(2! (7-2)!) = **21**

Therefore, 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

**Ans:**

**Expected number of candidates=(Candies count) \* (Probablity)**

**Child A = (1\*0.015) = 0.015**

**Child B = (4\*0.20) = 0.8**

**Child C = (3\*0.65) = 1.95**

**Child D = (5\*0.005) = 0.025**

**Child E = (6\*0.01) = 0.06**

**Child F = (2\*0.120) = 0.24**

**Therefore, expected number of candies for a randomly selected child**

**= (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**

**Python 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?

**Python file**

**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**

**Python file**

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



**1: Data is right skewed (Positively Skewed).**

**2: Most of the Chicks Weight lie in the range (50-100), followed by (100-150) and (150-200).**



**1: Data is Positively Skewed.**

**2: Data has Outliers**

**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?

**Python file**

**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?

**Python file**

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

**Ans: Symmetrical Distribution**

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

**Ans: Positive Skewed OR Right Skewed**

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

**Ans: Negetive Skewed OR Left Skewed**

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

**Ans: Leptokurtic Distribution (High Peak on central part of data)**

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

**Ans: Platykurtic Distribution (Flat peak on the central part of data)**

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



What can we say about the distribution of the data?

**Data is asymmetric. Most of the data is on the right side.**

What is nature of skewness of the data?

**Positively Skewed**

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

Upper quartile=18

Lower quartile=10  
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.

**Both the boxplot is normally distributed and has no outliers. Both has a median around 262.5. The only difference is first graph has a short range compared to second one.**

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)
  3. P (20<MPG<50)

**Python file**

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

**Python file**

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

**Python file**

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

**Python file**

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

**Python file**