|  |  |
| --- | --- |
| 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) | 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 | 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) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| 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 :

Probabilty of tossing the coin and getting the result = 8 =HHT,HTH,THH,HHH,HTT,TTH,THT,TTT

Probability of getting 2 heads & 1 Tail = HHT, HTH,THH = 3

So the probability = 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

Ans :

When 2 dice are rolled possible outcomes = 62 = 36

1. Equal to 1 is 0
2. Less than or equal to 4 = 1+1, 1+2, 2+1,3+1,1+3,2+2 = 6 = 6/36 = 1/6
3. Sum is divisible by 2 & 3 = 1+5, 5+1, 2+4, 4+2,3+3,6+6 = 6 = 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?

Ans :

2R,3G,2B, = 7C2 = 7\*6/2\*1 = 21

Even of drawing 2 balls at random (2R+3G) = 5C2 = 5\*4/2\*1 = 10

P = 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: Total Candies = 21

Expected number of candies for randomly selected child = 0.015 \*1+0.20\*4+0.65\*3+0.005\*5+0.01\*6+0.120\*2 = 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.

**Mean**

**Points 3.596563**

**Score 3.217250**

**Weigh 17.848750**

**Median**

**Points 3.695**

**Score 3.325**

**Weigh 17.710**

**Mode**

**Points 3.07**

**Score 3.44**

**Weigh 17.02**

**Variance**

**Points 0.285881**

**Score 0.957379**

**Weigh 3.193166**

**Standard Deviation**

**Points 0.534679**

**Score 0.978457**

**Weigh 1.786943**

**Range**

**Points (4.93 – 2.76)**

**Score (3.21 – 5.42)**

**Weigh (17.84 – 22.90)**

**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?

**Expected value= Sum(Probabilty of weight X weight)**

**= 1/9\*108 +1/9\*110 +1/9\*123 +1/9\*134 +1/9\*145 +1/9\*167 +1/9\*187+1/9\*190**

**=145.33**

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans: For car ,skew value = -0.12**

**For car ,kurtosis value = -0.51**

**For dist, skew value = 0.81**

**For dist, kurtosis value = 0.41**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:**

**For SP ,skew value = 1.61**

**For SP ,kurtosis value = 2.98**

**For WT, skew value = -0.61**

**For WT, kurtosis value = 0.95**

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



**Ans: The skew is on the right side and tail is on the left side, since the skew is on the right side we are having the outliers on the right side and also the mean > median**



Ans: This boxplot is having many outlier on the Maximum side.

**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 : 94 % confidence interval= (198.73, 201.26)**

**98 % confidence interval= (198.43, 201.56)**

**96 % confidence interval= (198.62, 201.37)**

**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: mean = 41, median=40.5,variance=25.52,Standard Deviation=5.05**

**No outlier, & mean >median**

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

**Ans:no skewness, perfect uniformly distributed curve is obtained.**

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

**Ans: skewness is on the right side.**

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

**Ans: skewness is on the left side.**

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

**Ans: the curve is leptokurtic and curve is more peaked**

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

**Ans: curve will be flatter and broader**

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



What can we say about the distribution of the data?

**Ans: Box plot shows the mean is towards the higher value, and the curve is normally distributed**

What is nature of skewness of the data?

**Ans: The data is skewed to the left**

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

**Ans: 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.

**Ans: There are no outliers, and both box plot have same median, both are normally distributed. And there is no 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)

**Ans: MPG>38= 0.348**

**MPG<40= 0.729**

**MPG<40= 0.729**

**less\_than\_MPG\_20\_greater\_than\_MPG\_50= 0.013**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans:This data follows 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

**Ans:This data does not follows normal distribution**

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

**Ans:** **Z score of 60% confidence interval= -0.842**

**Z score of 90% confidence interval= -1.645**

**Z score of 94% confidence interval= -1.881**

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

**Ans:** **t score of 95% confidence interval= -2.064**

**t score of 96% confidence interval= -2.172**

**t score of 99% confidence interval= -2.797**

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: probability of bulbs with life of no more than 260 days 32.2 %**