|  |  |
| --- | --- |
| 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 | Ratio |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Ordinal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Answer: When a Coin is tossed three time then number of outcome is= 23 =8

{HHH,TTT,HTT,THT,TTH,THH,HTH,HHT}

Probability of getting two heads and one tail = Interested outcome /possible outcomes

=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

Answer: The probability of rolling a Dice two times = 62 = 36

a)Zero

b){(1,1),(1,2),(1,3),(2,1),(3,1),(2,2)}

Probability of getting Less than or equal to 4 =6/36 =1/6

C){(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)}

Probability of 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?

Answer: The total number of balls={g,g,g,r,r,b,b}=7

Probability of first ball drawn at random that none of the ball is blue=5/7

Probability of drawing the second ball twice at random that not blue ={(r,r),(r,g),(g,g),(g,r)}=4/6

Probability than none of the balls drawn at random=5/7 \*4/6

=20/42 =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

Answer: Expected number of candies

=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,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?

Answer: Given 108,110,123,134,135,145,167,187,199

Total numbers of patients is 9

Probability of selecting each patient at random=1/9

Expected value of the Weight of that patient

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

=1/9(1308) =145.3

**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**



Answer: The histogram peak has right skew and tail is on right. Mean>Median. Because we have outliers on the higher side.



Answer: In this box plot median is less than mean so it is right skewed and we have outlier on the upper side of the box plot

**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.

Answer:34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56

1. Mean=41
2. Median=40.5

C) Variance = 5.05

d)Standard Deviation = 25.5

1. What can we say about the student marks?

Answer: Mean is greater than median so, the data is slightly skewed towards right

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

Answer: When mean and median of data are equal then the nature of skewness is Zero

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

Answer: When mean is greater than median then the nature of skewness is positive in nature

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

Answer: when median is greater than mean then the nature of skewness is negative in nature

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

Answer: Positive kurtosis (leptokurtic kurtosis>3) value indicates that the data values are located near the mean and more data values are located on the tail.

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

Answer: Negative kurtosis (platykurtic kurtosis<3) value indicates that the data values are located near the mean and less data values are located on the tail.

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



What can we say about the distribution of the data?

Answer: The data is not actually equally distributed across the plane. Median of the data is 14.6(appr)

25 percent of the data is lies between 0-10

50 percent of the data is lies between 10-18

25 percent of the data is lies between 18-20(appr)

What is nature of skewness of the data?

Answer: Nature of skewness of the data is Negative skew in nature.

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

Answer: Q3=18 and Q1=10

IQR (Interquartile Range) = 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.

Answer: Here we have to compare boxplot 1 with respect to boxplot2. Here the main inference is box polt2 data range varies is high than the box plot1. The median in the two box plots are equal.

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 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

Answer: Adipose Tissue(AT) and Waist Circumference(Waist) from wc-at data does not follows Normal distribution

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