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

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

**Answer:**

If three coins are tossed then sample space is

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

n(s)=8

Let A be the event that,

probability that two heads and one tail are obtained is given by

A={HHT,HTH,THH,}

n(A)=3

P(A)=n(A)/n(S)=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:**

If two dice are rolled then sample space is given by

S={ (1, 1) (1, 2) (1, 3) (1, 4) (1, 5) (1, 6) ,(2, 1) (2, 2) (2, 3) (2, 4) (2, 5) (2, 6), (3, 1) (3, 2) (3, 3) (3, 4) (3, 5) (3, 6), (4, 1) (4, 2) (4, 3) (4, 4) (4, 5) (4, 6),(5, 1) (5, 2) (5, 3) (5, 4) (5, 5) (5, 6), (6, 1) (6, 2) (6, 3) (6, 4) (6, 5) (6, 6)}

n(S)=36

1. A be the event that sum is equal to 1 and is given by

A={} -------------A is null event

n(A)=0

P(A)=0

b)B be the event that sum is Less than or equal to 4 and is given by

B={(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)}

n(B)=6

P(B)=n(B)/n(S)= 6/36=1/6

C)C be the event that Sum is divisible by 2 and 3 and is given by

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

n(C)=24

P(A)=n(C)/n(S)= 24/36=2/3

---------------------------------------------------------------------------------------------------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:**

Total Number of Balls in bag =N=7

We have to draw two balls at random

Then the number of ways 2 balls drawn from 7 balls is given by

n(ῼ)= 7 C2= 21

Let A be the event that none of the ball drawn is blue

The number of ways 2 balls drawn from 2 red, 3 green is given by

n(A)= 5C2= 10

Hence P(A)= 10/21= 0.4761

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

Expected number of candies for a randomly selected child is given by

E(x)= ∑ x.P(x)

= 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,Weight

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:

Expected value of X = E(x) = = = 145.3333

-------------------------------------------------------------------------------------------------------------

**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 nature of the above histogram is right skewed because most of the data points are clustered at the right side of the histogram. The peak value in the data shows the most common value for the weight.**



**Interepretation:** Box plot displays the distribution of data based on five number summary which involves minimum of the data , Lower(First) quartile, median, Upper(Third) quartile and maximum score. The points outside the whiskers of the boxplot are the outliers.Median divides the box plot in two parts. In the above boxplot the median is closer to the bottom of the box and if the whisker is shorter on the lower end of the box, then the distribution is positively skewed. Hence the above boxplot shows that the data is positively skewed.

-------------------------------------------------------------------------------------------------------------

**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?

Answer: Given that

-----------------------------------------------------------------------------------

**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?

Answer: If the mean, median and mode of the data are equal then we say that it is symmetric.

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

Answer: If the mean of the data is greater than median then it is positively skewed.

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

Answer: If the mean of the data is less than median then we can say that it is negatively skewed.

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

Answer: Positive value of the kurtosis indicates that the distribution is leptokurtic.

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

Answer: Negative value of the kurtosis indicates that the distribution is platykurtic.

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



What can we say about the distribution of the data?

Answer: Here we can see that the median is closer to the bottom of the box and whisker is shorter on the lower end of the box hence the distribution is positively skewed.

What is nature of skewness of the data?

Answer: The distribution of the data is positively skewed.

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

Answer:

Q1=10

Q3=18

IQR= 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:** From the above two boxplot we can say that median of the both boxplot are same. In the both the boxplots medians at the middle of the box and the whiskers are about the sane on both sides of the box then distribution is symmetric. Spread of the first boxplot is less than the spread of the second boxplot. Average of the first boxplot is less than the second boxplot.

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

Hypothesis:

H0: meu=60

Hint:

Sd=90 days

Meu=260

Xbar=270

N=18

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom