|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Ordinal |
| Weight of a person | Ratio |
| Weight of Gold | Ratio |
| Distance between two places | Ratio |
| Length of a leaf | Ratio |
| Dog's weight | Ratio |
| Blue Color | Nominal |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Ordinal |
| 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 | 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 | Interval |
| Blood Group | Nominal |
| Time Of Day | ordinal |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ordinal |
| Religious Preference | nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Interval |

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

**Ans** = When Three coins are tossed possible outcomes are,

HHH, HHT, HTH, THH, TTH, THT, HTT, TTT.

The number of combinations which have two heads and one tail are 3

So,

Probability (2H and 1T) = No. of favourable outcomes / Total No of outcomes

p = 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 two dice are rolled simultaneously,total no of possible outcomes is 36

n(s) = 36

a] let E1 be the event that sum is equal to 1 which is 0 because they starts with

(1,1) likewise . other than in the dice we are not having zero.

p(E1) =n(E1) / n(S) = 0/36 = 0

b] let E2 be the event that sum is less than or equal to 4 is 6

(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)

p(E2) =n(E2) / n(S) = 6/36 = 0.1666

c] let E3 be the event that sum is divisible by 2 and 3 is 6

(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)

p(E3) =n(E3) / n(S) = 6/36 = 0.166

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** = Total no of balls = 2 + 3 +2 = 7 balls

From these 7 we have to drawn 2 balls

S=Two balls are drawn at random

n(S) = 7C2= 7x6/2x1 = 42/2 = 21

A= None of the balls are drawn is blue

n(A) = 5C2= 5x4/2x1 =20/2 = 10

p(A) = n(A)/n(S) = 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** = let X be the number of candies,expected number of candies E(X)

So the expected value is the sum of: [(each of the possible outcomes) × (the probability of the outcome occurring)]

E(X) =∑ xi f(xi) #f(x) is the respective probability of that candies

No of expected candies from child A = 1 x 0.015=0.015

No of expected candies from child B = 4 x 0.20=0.8

No of expected candies from child C = 3 x 0.65=1.95

No of expected candies from child D = 5 x 0.005=0.025

No of expected candies from child E = 6 x 0.01=0.06

No of expected candies from child F = 2 x 0.120=0.24

E(X) = 0.015+0.8+1.95+0.025+0.06+0.24=3.09

**Ans = 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** -In jupyter notebook

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** = 145.333

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

**Cars speed and distance**

**Use Q9\_a.csv**- ans in jupyter notebook

**SP and Weight(WT)**

**Use Q9\_b.csv-** ans in jupyter notebook

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



-From the above histogram it is clear that the distribution is right skewed because long tail extends to the right while most of the data are concentrate on left side





Box plot provide basic information about a distribution

For example,a distribution with positive skew would have a longer value in the positive direction than in the negative direction.A larger mean than median would also indicate a positive skew. It has maximum observation in upper quartile,lots of outlier in upper side,many of the details of a distribution are not revealed in a 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?

-In jupyter notebook

**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 In jupyter notebook

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

**Ans** = It is symmetrical distribution and  value of coefficient of skewness is zero

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

**Ans** = Positive skewness (often referred to as Right-Skewed) implies mass of the distribution concentrated on left side

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

**Ans** = Negative skewness (often referred to as Left-Skewed) implies mass of the distribution concentrated on right side

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

**Ans** = It is Leptokurtic type kurtosis indicates more values in the distribution tails and more values close to the mean (i.e. sharply peaked with heavy tails)

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

**Ans** = It is Platykurtic type kurtosis indicates fewer values in the tails and fewer values close to the mean (i.e. the curve has a flat peak and has more dispersed scores with lighter tails).

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



What can we say about the distribution of the data?

-Not normally distributed

What is nature of skewness of the data?

-Negative skewness

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

-10-18

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

Ans = the first box is wider than the second box because the width of the boxes are proportional to the number of subject in the distribution data

In first boxplot the IQR (Interquartile range) is in between 252 to 275 and in second boxplot the IQR is in between 220 to 312 approximately,

Median is common in both boxplot, there is no outlier is given in above plot

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 in jupyter notebook

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

Ans = In Jupyter notebook

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

interval, 60% confidence interval

Ans = In Jupyter notebook

Q 23) Calculate the t scores of 95% confidence interval, 96% confidence

interval, 99% confidence interval for sample size of 25

Ans = In Jupyter notebook

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 = in jupyter notebook