|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete Countable |
| Results of rolling a dice | Discrete Countable |
| Weight of a person | Continous Ratio |
| Weight of Gold | Continous Ratio |
| Distance between two places | Continous Ratio |
| Length of a leaf | Continous Ratio |
| Dog's weight | Continous Ratio |
| Blue Color | Nominal |
| Number of kids | Discrete Countable |
| Number of tickets in Indian railways | Discrete Countable |
| Number of times married | Discrete Countable |
| 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 | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| 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?

**Ans:**

The total number of possible outcomes when tossing three coins 2^3=8

*P*(Two heads and one tail)=*P*(HHT)+*P*(HTH)+*P*(THH)=1/8+1/8+1/8=3/8

Therefore, the probability of obtaining two heads and one tail when three coins are tossed is **3/8**.

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1 - The minimum value on each die is 1, the minimum sum we can get is 2 (1+1). Therefore, the probability of getting a sum of 1 is **0**, because it is not possible.
2. Less than or equal to 4 - The total number of possible outcomes when rolling two dice is 6×6=36.

the total number of favorable outcomes is 1 (for sum = 2) + 2 (for sum = 3) + 3 (for sum = 4) = 6.

*P*(Sum≤4)=6/36=**1/6**

1. Sum is divisible by 2 and 3 - The number of outcomes where the sum is divisible by 6 (since 6 is the least common multiple of 2 and 3).

The possible outcomes where the sum is divisible by 6 are: (1, 5), (2, 4), (3, 3), (4, 2), (5, 1).

So, the total number of favorable outcomes is 5.

Therefore, the probability that the sum is divisible by both 2 and 3 is: *P*(Sum is divisible by 2 and 3)= **5/36**

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 number of ways to draw 2 balls out of 7: 7!/ 2!(7−2)!= 21

There are 5 balls that are not blue (2 red and 3 green).

Number of ways to draw 2 balls out of 5 non-blue balls: 5!/ 2!(5−2)!= 10

Therefore, the probability that none of the balls drawn is blue is: Probability=Number of ways to draw 2 nonblue balls/Total number of ways to draw 2 balls= **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:

we need to multiply each candy count by its corresponding probability and then sum up these values.

*E*=∑*i*​ *xi*​×*Pi*​

We can calculate the expected number of candies as:  
E=1×0.015+4×0.20+3×0.65+5×0.005+6×0.01+2×0.120

*E*=3.135

Therefore, the expected number of candies for a randomly selected child is **3.135.**

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?

Ans: To calculate the expected value (mean) of the weights of patients at the clinic, we need to find the mean of the given data.

*E*(*X*)=(108+110+123+134+135+145+167+187+199)/9= 146.44

Therefore, the expected value of the weight of a patient chosen at random is **146.44** pounds

**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**



Ans: Here we can see that the major Chick weights fall in the catogory of 50-100g(measures in x) as the maximum which is 200.The minimum weights have afrequency if less than or equal to 5.

The plot is Right sqewed which show that there is lesser concentration ofchick weights in the 300-400gram category .



Ans: Mean>Median, so right skewed and we have outlier on the upperside ofbox plot and there is less data points between Q1 and bottom point.

**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:

The formula for the confidence interval of the mean when the population standard deviation is unknown:

Confidence Interval = x’(+/-)z(s/Sqrt(n))

x’ is the sample mean (200 pounds)

S is the sample standard deviation (30 pounds)

n is the sample size (2000 men)

z is the z-score corresponding to the desired confidence level

For a 94% confidence interval, the z-score is approximately 1.88.

For a 98% confidence interval, the z-score is approximately 2.33.

For a 96% confidence interval, the z-score is approximately 2.05.

1. 94% Confidence Interval:

200(+/-)1.88(30/sqrt(2000))= (**198.738,201.262)**

2 **.** 98% confidence interval:

200(+/-)2.33(30/sqrt(2000))= (**198.436,201.564**)

3. 96% confidence interval:

200(+/-)2.05(30/sqrt(2000))= (**198.627,201.373**)

Therefore, the 94% confidence interval is (198.738, 201.262), the 98% confidence interval is (198.436, 201.564), and the 96% confidence interval is (198.627, 201.373) for the average weight of adult males in Mexico.

**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?

Ans: The distribution is symmetrical

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

Ans: The distribution is right-skewed or positively skewed

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

Ans: The distribution is left-skewed or negatively skewed

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

Ans: A positive kurtosis value indicates that the distribution has heavier tails and a sharper peak than a normal distribution. It suggests that the data has more outliers or extreme values than would be expected in a normal distribution.

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

A negative kurtosis value indicates that the distribution has lighter tails and a flatter peak than a normal distribution. It suggests that the data has fewer outliers or extreme values than would be expected in a normal distribution.

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



What can we say about the distribution of the data?

Ans: The data is not actually equally distributed across the plane. There might beoutliers influencing the data . Median of the data is 14.7(app x)

25 percent of the data lies between 0-10

50 percent of the data lies between 10-18

25 percent of the data lies after 18-20 appx

What is nature of skewness of the data?

Ans: The data will be left skewed since whisker length on the upper quadrant ishigher than the data on the lower quadrant. Median will be greater than themean since data is left skewed.

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

Ans: IQR is the inter quartile range.

Here Q1 = 10

Q2 = 14.7

Q3 = 18

IQR = Q3 –Q1 = **8**(approx)

Q19) Comment on the below Boxplot visualizations?



Ans: Here there is a representation of 2 box plots in which box plot 2) is highlydistributed across the plane and 1) is slightly less distributed.(variance)

Whiskers in these diagrams also show this.100% of the data is spread acrossvalues from 350 in 2 whereas its spread in range 250-290 app x in 1

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

Ans: Here when we compare box plot 1 with box plot 2 we can say that the data in boxplot 1 is widely spread. Here the main inference is that since the data range varieshigh in box plot 2 it is hard to make a prediction in box plot 2. The median in the 2box plots are equal. And the data spread in both of them are symmetrical.

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

Ans:

Using a standard normal distribution table or calculator, the z-score for a 90% confidence interval is approximately **1.645**

The z-score for a 94% confidence interval is approximately **1.88**.

z-score for a 60% confidence interval is approximately **0.84**.

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

ANS: TSCORE CALCULATION

T((1,alpha),(n-1))

Here n = 25

n-1 = 24

95%= qt(0.975,24)= **2.063899**

96% = qt(0.98,24)=**2.171545**

99% = qt(0.995,24)= **2.79694**

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