|  |  |
| --- | --- |
| 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 | Categorical |
| 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 | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Nominal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ordinal |
| Years of Education | Ordinal |

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

Ans:

The coin has 2 possibility one H and T, as coins has toss 3 time’s

So,the total no of outcome =2x2x2=8

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

Two heads and one tail outcomes = {HHT,HTH,THH}=3  
Probability = 3/8 = 0.375

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

1. Equal to 1

Ans: 0 probability

1. Less than or equal to 4

Ans: (1,1), (1,2), (1,3), (2,1), (2,2), (3,1)

1/6 probability

1. Sum is divisible by 2 and 3

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

1/6 probability around 1.667

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:

The total number of ways to draw 2 balls from 7 is given by the combination formula, which is:

C(7,2) = (7!)/(2!(7-2)!) = 21

This means that there are 21 possible pairs of balls that can be drawn.

To find the probability that none of the balls drawn is blue, we need to count the number of ways to draw 2 balls from the 5 non-blue balls,

which is: C(5,2) = (5!)/(2!(5-2)!) = 10

So, out of the 21 possible pairs, 10 pairs do not contain any blue balls. Therefore, the probability of drawing two balls without getting any blue ball is 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: Expected value=sum of(Candies count x Probability)  
Expected number of candies = (1 x 0.015) + (4 x 0.20) + (3 x 0.65) + (5 x 0.005) + (6 x 0.01) + (2 x 0.12)

Simplifying this expression, we get: Expected number of candies = 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24 Expected number of candies = 3.085

Therefore, the expected number of candies for a randomly selected child is 3.085.

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**

**Ans:**

**Mean:** df['Points'].mean() , df.Points.mean() 3.59

**Median:** data.Points.median()- 3.695

**Mode:** data.point.mode()

**Variance:**

**Standard Deviation:**

**Range:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Count | 32.00 | 32.00 | 32.00 |
| Mean | 3.596 | 3.217 | 17.84 |
| median | 3.695 | 3.325 | 17.71 |
| Mode | 6.99 | 3.44 | 35.92 |
| Variance | 0.28 | 0.95 | 3.19 |
| S.D | 0.534 | 0.978 | 1.786 |
| Range | 2.14 | 2.21 | 5.05 |

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?

Solve:

Expected Value = ∑(probability\*value)

There are total 9 patient probability = 1/9

Expected Value= (1/9\*108)+(1/9\*110)+(1/9\*123)+(1/9\*134)+(1/9\*135)+(1/9\*145)+(1/9\*167)+(1/9\*187)+(1/9\*199)

=1/9(1308)

=145.33

**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**

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **SP** | **1.61** | **2.97** |
| **WT** | **-0.614** | **0.950** |

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



**1)Data is positively Skewes**

**2)Majority Of Chicks Has Weight in range of 50-100**

**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?

Solve:

To calculate the confidence intervals

CI = X̄ ± Z\*(σ/√n)

σ = 30

we calculate the Z-scores

For 94% confidence interval, the Z-score is 1.88

For 98% confidence interval, the Z-score is 2.33

For 96% confidence interval, the Z-score is 1.

we can calculate the confidence intervals:

For the 94% confidence interval:

CI = 200 ± 1.88\*(30/√2000) CI = 200 ± 2.14 CI = (197.86, 202.14)

For the 98% confidence interval:

CI = 200 ± 2.33\*(30/√2000) CI = 200 ± 2.69 CI = (197.31, 202.69)

For the 96% confidence interval:

CI = 200 ± 1.96\*(30/√2000) CI = 200 ± 2.26 CI = (197.74, 202.26)

**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.

ANS = Mean=41

Median =40.5

Variance=25.5291

SD=5.0526

1. What can we say about the student marks?

Ans: 1)Data has outlier

2)Not Normally Distributed

3)Majority Of Students Got Marks Between 35-45

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

ANS = Zero Skew, perfectly symmetric curve.

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

ANS = Positive Skew,Data Distributed More On left.

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

ANS = Negative Skew,Data Distributed More On Right.

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

ANS=Distribution Is Peaked And There Are Thick Tail

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

ANS=Wider peak on central part of data.

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 Symmetric.

What is nature of skewness of the data?

ANS : Skewness = Negative

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 :

1. Data is Normally Distrubuted.No Outliers.Center around 263 and first one has less range.
2. Data is Normally Distributed.No Outliers.Center around 263 and second one has more range compare to first one

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)

ANS : 0.48

* 1. P(MPG<40)

ANS : 0.753

* 1. 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

ANS : MPG 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 : Both AT And Waist Dose Not Follows The Normal Distribution.

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

ANS :

|  |  |
| --- | --- |
| 90% | 1.28 |
| 94% | 1.75 |
| 60% | 0.28 |

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

|  |  |
| --- | --- |
| 95% | 1.71 |
| 96% | 1.82 |
| 99% | 2.49 |

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

Degree Of Freedom = 17

P(t) = 0.31