|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete Data Type |
| 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 | Nominal Data Type |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| 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 Scale |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Ordinal |
| Time Of Day | Interval data type |
| Time on a Clock with Hands | Ratio Scale |
| Number of Children | Discreate |
| Religious Preference | Nominal |
| Barometer Pressure | Interval Data Type |
| SAT Scores | Interval |
| Years of Education | Discreate |

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

1. Three coins are tosssed

No of possible outcomes={(T,T,T),(H,H,H),(T,T,H),(T,H,T),(H,H,T),(H,T,H),(T,H,H),(H,T,T)}

Probability(P)=no.of possible outcomes/total no of outcomes

P=3/8

Output=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

A).a)

Zero, Because the minimum sum of two dice are 1+1 =2.So,there is no chance to getting sum of two diece is 1.

(b) :- The possible of getting sum is 2 has 1 chance (1,1)

- The possible of getting sum is 3 has 2 chance (1,2),(2,1)

- The possible of getting sum is 4 has 1 chance (1,3),(2,2),(3,1)

Total No. of Possiblities = 36

Sum is less than or equal to 4:-(1+2+3)/36=0.166666 (1/6)

©:-The probability of getting sum is divisble by 2 and 3 .6 and 12 are the only divisble by 2 and 3.The possible ways to get a sum of 6 are (1,5),(2,4,(3,3),(4,2),(5,1) and sum of 12 are (6,6) . The total number of possible outcomes is 36.

Sum is divisible by 2 and 3 = (5+1)/36

= 1/6 (0.6666)

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?

1. Total no. Of balls drawn =21{(2+3+2)=7c2

Probability that none of the balls drawn is blue =(None of the balls are blue /no.of balls drawn)

=20/42

=10/21

Output = 0.476190

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 = (X1\*probablity(X1)+(X2\*probablity(X2)=…...

= (1\*0.015)+(4\*0.20)+(3\*0.65)+(5\*0.005)+(6\*0.01)+(2\*0.120) = 3.09

Expected Numbers of Candies= 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?

A). To calculate the expected value of the weight for a randomly chosen patient from the given dataset,we need to sum up all the weights and divide the sum by the total numbers of patients

Expected value=X\*(probability of X)

The weights of the patient are probability of(X) is 108,110, 123, 134, 135, 145, 167, 187, 199

X= sum of observations/no.of obseravtions

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

Output= 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**

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

**Answer: This histogram shows of the data of ChickWeight$weight and the histogram shows that this data is right skewened or positive skewned because the data is Heavily Tailed**

**. In the boxplot inference that the shows that there is so many outliers .the box plot shows that data is ‘Negative skewed data’**



**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 Excel file.

**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?

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

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

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

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

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

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

From 12 to 18 All In Python File.

Q19) Comment on the below Boxplot visualizations?



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

1. 1) In the boxplot (1) the data move towards the ‘Left skewed data’

The variance is also less ,compared to the boxplot(2)

2). In the boxplot(2) the data move towards the ‘Normal data’

The data is plotted at the median

The variance is also very high ,compared to the boxplot(1)

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)

In Python File

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

In Python File.

1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

In Python File.

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

1. We need to calculate the Z-scores for 90% confidence interval from the Z-score table

90% confidence interval =1.645 (from Z-score table we will get these values)

60% confidence interval =0.842 (from Z-score table we will get these values)

94% confidence interval for that we need to find Area

Area= 1+ confidence interval/2 =(1 + 94 /100) /2

= 1 + 0.94 / 2

=0.9700

We need to find the 0.9700 in the Z- score table

=1.8+0.08

OUTPUT = 1.88 (approximately)

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

1. Sample size(n)=25

Degrees of freedom =sample size-1=24

From the t-score table,we need to find the 95% confidence interval

95% confidence interval =2.064

99% confidence interval = 2.797

96% confidence interval = 2.171

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 à pt(tscore,df)

df à degrees of freedom

In Python File.