|  |  |
| --- | --- |
| 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 scale** |
| Weight | **Ratio** |
| Hair Color | **Nominal** |
| Socioeconomic Status | **Ordinal** |
| Fahrenheit Temperature | **Interval scale** |
| 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 scale** |
| Time on a Clock with Hands | **Interval scale** |
| Number of Children | **Ratio** |
| Religious Preference | **Nominal** |
| Barometer Pressure | Ratio |
| SAT Scores | Interval scale |
| Years of Education | Ratio |

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

Answer: 3/8=0375

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

Answers: a)0

b)3/36=0.0833

c)6/36=1/6=0.169

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: 10/21=0.476

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: 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**

**Answer: please refer the attached”** **ASSIGNMENT-1 [ 7th question(Q7.csv)].ipynb”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: 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Answer: please refer the attached”** **ASSIGNMENT-1[9th question(Q9(a).csv)].ipynb”file**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Answer: please refer the attached”** **ASSIGNMENT1[9th question(Q9- b.csv)].ipynb”file**

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



Answer: The data in the histogram shows the different trends of chick weight. The data in the histogram is right skewed as the “frequency” is diminishing towards right. The maximum number of chicks weight ranges between 50 and 100.

The box plot is also right skewed as the box is towards the left. There are outliers in the given data set which are represented as circles out side the box plot. These may be inspected before analysis.

**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, Mean (μ)=200, Standard deviation (σ)=30

For calculating 94% confidence interval, Zscore = 1.555

Max value x= μ+σz = 200 + 30 x 1.555 = 246.65 pounds

Min value x= μ-σz = 200 - 30 x 1.555 = 153.35 pounds

Similarly, For calculating 98% confidence interval, Zscore = 2.326

(Max value, Min value) = (269.78,130.22)

,For calculating 96% confidence interval, Zscore = 2.053

(Max value, Min value) = (261.5,138.5)

**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?

Answer: please refer the attached “ASSIGNMENT1[12 th question)].ipynb”file

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

Answer: perfect symmetrical skewness

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

Answer: Right Skewed

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

Answer: Left skewed

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

Answer: Distribution is peaked

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

Answer: Distribution if flat

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



What can we say about the distribution of the data?

Answer: The distribution is not symmetric and the data is left skewed

What is nature of skewness of the data?

Answer: Left skewed

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

Answer: 8

Q19) Comment on the below Boxplot visualizations?



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

Answer: The two box plots are symmetrical following normal distribution. But the range of values differ from each other. The mean and median of the box plots are same(approx.). But the variance of box plot-2 is higher than the box plot 1. We can observe that there are no outliers in both the data sets.

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) Answer: 0.652406
  2. P(MPG<40) Answer: 0.72934

c. P (20<MPG<50) Answer:0.9551-0.0571=0.8988

Answer: please refer the attached “ASSIGNMENT1[20 th question)].ipynb”file

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer: please refer the attached “ASSIGNMENT1[21(a) th question)].ipynb”file

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

Dataset: wc-at.csv

Answer: please refer the attached “ASSIGNMENT1[21(b)th question)].ipynb”file

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

Answer: please refer the attached “ASSIGNMENT1[22 nd question)].ipynb” file

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

Answer: please refer the attached “ASSIGNMET1[23 rd question)].ipynb” file

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

Answer: please refer the attached “ASSIGNMET1[24 th question].ipynb”file