Q1) Identify the Data type for the Following:

|  |  |
| --- | --- |
| 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 | Continuous |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Discrete |

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) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ordinal |
| Years of Education | Ordinal |

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

If we tossed three coins we will get

Total Possible Outcomes = 2^3 = 8

Oucomes where two heads and one tail are obtained = 3

Probability = 3/8 =0.375

Example:- H H T

H T H

T H H

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

1. Equal to 1

It is not possible to get the Sum = 1, because the Least possible sum while rolling the two dices is 2.

1. Less than or equal to 4

Total Possible Outcomes = 6^2 = 36

Outcomes where Less than or equal to 4 is obtained = 6

Probability = 6/36 = 1/6 =0.16.

1. Sum is divisible by 2and 3

Total possible Outcomes = 6^2 =36

Outcomes where Sum is divisible by 2 and 3 is obatined = 6

Probability = 6/36 = 1/6 =0.16

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?

Total no.of ways to draw 2 Balls from 7 Balls = 7! / 2! \* (7-2)!

= 5040 / 2 \* 120 =21

The no.of ways to drawn 2 Balls from 5 non-balls = 5! / 2! \* (5-2)! = 10

Probability = 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

Expected Value = Summarization of no.of Candies count \* Probability

Expected number of candies = (1 \* 0.015) + (4 \* 0.20) + (3 \* 0.65) + (5 \* 0.005) + (6 \* 0.01) + (2 \* 0.120)

Expected number of candies = 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24

Expected number of candies = 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**

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?

Probability of Choosing 1 patient out of 9 patients P(x) is 1/9

Expected Value = ΣP(x) \* E(x)

E(x) = 108 + 110 + 123 + 134 + 135 + 145 + 167 +187 +199 = 1308

Epected Value = 1/9 \* 1308

= 145.33

**9) 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**



From the above Histogram we can say that :-

This Histogram is Right Skewed, I can expect the Skewness should be “Positive” ,So most of your observations are lies on the Left hand side, very few observations are lies on the Right hand side.

In the ChickWeight the data between 350 to 400 are Outliers



This is one of the technique to identify the outliers.

Boxplot is represent the data from other than Histogram

For the above Boxplot we can say that:-

The Lower Limit is 0.

The Upper Limit is 350 approximately.

The Box (Q1, median (means the middle line in the box), Q3) which contains

50% of data is known as “Inter-Quartile Range (IQR)”.

Here Approximately Q1 = 50, Q2 = 200 and Median = 125

The observations above the Upper Limit are considered as Outliers.

**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?

**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.

Step:1:- Finding Mean and Median for given data

Mean= 34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56/18 = 41

Median =

40+41/2 =40.5

Step:2:- To Calculate Variance we use squared mean and also calculating Standard Deviation.

Variance=

(34-41)\*2 = (-7)^2 = 49

(36-41)\*2 = (-5)^2 = 25

(36-41)\*2 = (-5)^2 = 25

(38-41)\*2 = (-3)^2 = 9

(38-41)\*2 = (-3)^2 = 9

(39-41)\*2 = (-2)^2 = 4

(39-41)\*2 = (-2)^2 = 4

(40-41)\*2 = (-1)^2 = 1

(40-41)\*2 = (-1)^2 = 1

(41-41)\*2 = (0)^2 = 0

(41-41)\*2 = (0)^2 = 0

(41-41)\*2 = (0)^2 = 0

(41-41)\*2 = (0)^2 = 0

(42-41)\*2 = (1)^2 = 1

(42-41)\*2 = (1)^2 = 1

(45-41)\*2 = (4)^2 = 16

(49-41)\*2 = (8)^2 = 64

(56-41)\*2 = (15)^2 = 225

**Variance:**

Now adding all the new variance data calculated on mean to get actual Variance and Standard Deviation

49+25+25+9+9+4+4+1+1+0+0+0+0+1+1+16+64+225=434

Variance = 434/(n-1)

= 434/(18-1)

= 434/17

= 25.52

**Standard Deviation:**

Standard Deviation = Square root of Variance 25.52

= 5.05

1. What can we say about the student marks?

Given that Students Marks between 34 to 56.

Maximum marks obtained are 56 and

Minimum marks obtained are 34 and

Finally the Average Marks obtained by the Students are 41.

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

Ans: When mean=median then it is Perfectly Symmetric i.e, Normal

i.e, Normal Distribution the skewness = 0.

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

Ans: mean >median = Right Skewed.

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

Ans: median > mean = Left Skewed.

Q16) What does positive kurtosis value indicates for adata ?

Ans: Positive Kurtosis indicates sharper peak(k>3) it has fatter tails.

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

Ans: Negative Kurtosis Indicates fatter peak(k<3) it has thinner tails.

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



What can we say about the distribution of the data?

The Dsitribution of Data in the above Boxplot is Right side Distribution, it

Means the Maximum of data lies between 10 to 18 and it is Left Skewed

What is nature of skewness of the data?

Left Skeweed because Maximum of Data is presented on the Right Side of the

Boxplot i.e; Negative Skewness

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

IQR = Q3 – Q1 (here IQR stands for Inter Quartile Range)

Here Q3 = 18 and Q1 = 10

Q3 – Q1 = 18 – 10 = 8

IQR = 8

Q19) Comment on the below Boxplot visualizations?



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

It shows that the Boxplot 2 barplot is more positve symmetric in nature and,

The Boxplot1 is having high peakness with Wider quatile range.

Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

Calculate the probability of MPG ofCars for the below cases.

MPG<- Cars$MPG

* 1. P(MPG>38)
  2. P(MPG<40)
  3. 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 scoresof 90% confidence interval,94% confidence interval, 60% confidence interval

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

Q 24**)**A Government companyclaims 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