|  |  |
| --- | --- |
| 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 |
| 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 | Ordinal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Total Possible Outcomes = 2^3 = 8

outcomes where two heads and one tail are obtained = 3

Probability = 3/8 = 0.375

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 Least possible sum while rolling 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 obtained = 6

Probability = 6/36 = 1/6 = 0.16

1. Sum is divisible by 2 and 3

Total Possible Outcomes = 6^2 = 36

Outcomes where Sum is divisible by 2 and 3 obtained = 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 draw 2 balls from the 5 non-blue 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 = Summation 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 = **Σ** P(x) \* E(x)

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

Expected Value = 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**

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



From the above Histogram we can say

Frequency on the Y-axis ranges from 0 to 200

Chickweight on the X-axis ranges from 0 to 400

The Highest Frequency is about 200 of Chickweight around 50 to 100

And the Data in the Histogram is Right Skeweed i.e Maximum of data lies on the

left side and very few observations lies on the right side

And the Data between 350 and 400 are Outliers



From the above Boxplot we can say

The Lower limit is 0

The Upper limit is 350 approximately

The Box(of Q1 ,median ,Q3) which contains 50% of data is know as InterQuartile

Range(IQR)

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

The observations above the Upper limit are consider 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.

Ans : **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

**Variance** =

Step1 : Find the mean:

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

Step2 : Calculating the sum of squared deviations from the mean for Each Number:

(34 - 41)^2 = 49

(36 - 41)^2 = 25

(36 - 41)^2 = 25

(38 - 41)^2 = 9

(38 - 41)^2 = 9

(39 - 41)^2 = 4

(39 - 41)^2 = 4

(40 - 41)^2 = 1

(40 - 41)^2 = 1

(41 - 41)^2 = 0

(41 - 41)^2 = 0

(41 - 41)^2 = 0

(41 - 41)^2 = 0

(42 - 41)^2 = 1

(42 - 41)^2 = 1

(45 - 41)^2 = 16

(49 - 41)^2 = 64

(56 - 41)^2 = 225

Next we add all the squared deviations:

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

Variance = 434 / (18 - 1)

= 434 / 17

= 25.52

**Standard Deviation** =

Square root of Variance = Square root of 25.52

= 5.051

Top of Form

1. What can we say about the student marks?

Here Student Marks Ranges Between 34 and 56

Maximum marks obtained are 56

Minimum marks obtained are 34

And the Average Marks Obtained by the Students are 41

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

Skewness = 0, i.e Data is having Normal Distribution

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

Data is Right Skeweed

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

Data is Left Skeweed

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

Indicates Sharper Peak and Fatter Tails (k>3)

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

Indicates Fatter Peak and Thinner Tails (k<3)

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



What can we say about the distribution of the data?

The Distribution of Data in the above Boxplot is Right side Distribution means

Maximum of data lies between 10 and 18 (Left Skeweed)

What is nature of skewness of the data?

Left Skeweed because Maximum Data 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 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.

Answer: This show that the blue colour barplot is more **positive symmetric** in nature. The red colour is having high peakness with **Wider quartile range.**

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

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 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