|  |  |
| --- | --- |
| 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 | Interval |
| Length of a leaf | Continuous |
| Dog's weight | Continuous |
| Blue Color | Nominal |
| 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 | Ratio |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Nominal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Nominal |
| Number of Children | Ratio |
| Religious Preference | Ordinal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Ratio |

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

Ans: When three coins are tossed the total number of possible combinations are 2³ = 8. These combinations are HHH, HHT, HTH, THH, TTH, THT, HTT, TTT.

The number of combinations which have two heads and one tail are HHT, HTH, TTH which makes them 3 in number.

Probability of Tow head and 1 tail=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

Answer:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Dice rolls** | **1** | **2** | **3** | **4** | **5** | **6** |
| **1** | 2 | 3 | 4 | 5 | 6 | 7 |
| **2** | 3 | 4 | 5 | 6 | 7 | 8 |
| **3** | 4 | 5 | 6 | 7 | 8 | 9 |
| **4** | 5 | 6 | 7 | 8 | 9 | 10 |
| **5** | 6 | 7 | 8 | 9 | 10 | 11 |
| **6** | 7 | 8 | 9 | 10 | 11 | 12 |

The probability that sum is

1. Equal to 1=0
2. Less than or equal to 4= 5/36=0.13
3. Sum is divisible by 2 and 3= 6/36

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:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Balls Drawns** | **Red** | **Red** | **Green** | **Green** | **Green** | **Blue** | **Blue** |
| **Red** | RedRed | RedRed | RedGreen | RedGreen | RedGreen | RedBlue | RedBlue |
| **Red** | RedRed | RedRed | RedGreen | RedGreen | RedGreen | RedBlue | RedBlue |
| **Green** | GreenRed | GreenRed | GreenGreen | GreenGreen | GreenGreen | GreenBlue | GreenBlue |
| **Green** | GreenRed | GreenRed | GreenGreen | GreenGreen | GreenGreen | GreenBlue | GreenBlue |
| **Green** | GreenRed | GreenRed | GreenGreen | GreenGreen | GreenGreen | GreenBlue | GreenBlue |
| **Blue** | BlueRed | BlueRed | BlueGreen | BlueGreen | BlueGreen | BlueBlue | BlueBlue |
| **Blue** | BlueRed | BlueRed | BlueGreen | BlueGreen | BlueGreen | BlueBlue | BlueBlue |

**25/49** is the probability that none of the drawn balls are blue.

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 value for Randomly Selected Child=[(1\*0.015)+(4\*0.20)+(3\*0.65)+(5\*0.0005)+(6\*0.01)+(2\*0.12)]

Expected value for Randomly Selected Child =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**: Comments I have mentioned in the python file Separately for Values and for Visualizations

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: The Expected value can be thought of as the “Average” value Attained by the random variable

Hence : Expected Value of the Randomly Chosen patient is patient =

∑(108, 110, 123, 134, 135, 145, 167, 187, 199)\9

=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**

**Answer:** Comments I have mentioned in the python file Separately for Values and for Visualizations

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



**Inferences:** on the Basis of above Histogram it is found that the data is Right Skewed and most of data points are in Range of 0-200



**Inferences:** On the basis of above Boxplot it is found that the data has more than 5 outliers which needs to be treated and data is Right Skewed as Whisker line is towards it if Boxplot changed to Horizontal View.

**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: Weight at 94% Confidence is: [143.5762 256.4238]

Weight at 96% Confidence is: [138.3875 261.6125]

Weight at 98% Confidence is: [130.2096 269.7904]

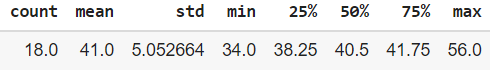
**Calculation part given in python File.**

**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?

Answer:



Most of the Students have scored in between 34 to 46

Variance is 25.529

Data is more dense in this range.

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

**Answer**: When Mean and Median is Equal it means that Data has symmetric distribution with 2 Modes And when Mean, Median and Mode is equal it means Zero Skewness and implies a Symmetric Distribution.

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

**Answer:** If Mean > Median it means it is Right Skewed Distribution.

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

**Answer:** If Median > Mean it means it is Left Skewed Distribution.

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

**Answer**:  Positive values of kurtosis indicate that a distribution is peaked and possess thick tail.

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

**Answer:** Negative kurtosis Indicate that distribution is flat and has thin tails.

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



What can we say about the distribution of the data?

**Answer:** Most of the data points are in between 10-18

Lower Quartile Whisker line is extended more than Upper Quartile whisker line.

There’s no outliers in the data set as per above Diagram,

What is nature of skewness of the data?

**Answer:** Data set is left Skewed Distributed.

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

**Answer:** IQR can be Calculated as Upper Quartile- Lower Quartile

Which is Approx =18.2-10.2

**IQR=8**

Q19) Comment on the below Boxplot visualizations?

A



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

**Answer: from** Above Boxplots We can say that boxplot 1 has less number of data points than boxplot 2 and Boxplot 1 spread of Distribution is also less than boxplot 2 However both of them have approx. same Median which is 262.5 Approx.

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)=0.34759
  2. P(MPG<40)=0.7293
  3. P (20<MPG<50)=0.89886

Calculations Performed 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

**Answer:** MPG for cars doesn’t follow 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

**Answer:** Distribution plot for AT and Waist drawn in python file and it doesn’t follow Normal Distribution.

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

**Answer:**

Z Scores of 90% CI is: 1.28155

Z Scores of 94% CI is: 1.55477

Z Scores of 60% CI is: 0.25334

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

**Answer:**

t Scores of 95% CI is: 1.7108

t Scores of 96% CI is: 1.8280

t Scores of 99% CI is: 2.4921

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:

Probability of Average light Bulb No More than 260 days is 32%.

P value is 0.32