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

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 | **Nominal** |
| 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?

**Ans:** **3/8**

There are three coins. Each coin is having 0.5 chance of getting head & 0.5 chance of getting tail. So, the total number of outcomes is 2^3 i.e. 8. Out of that, ways to get two heads and one tail are (H,H,T) (H,T,H) (T,H,H) i.e, 3. So, probability that two heads and one tail obtained are 3/8.

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

**Ans:** Probability that sum is

1. Equal to 1 - **0**
2. Less than or equal to 4 - 6/36 = **1/6**

(1,1) (1,2) (1,3) (2,1) (2,2) (3,1)

1. Sum is divisible by 2 and 3 - 6/36 = **1/6**

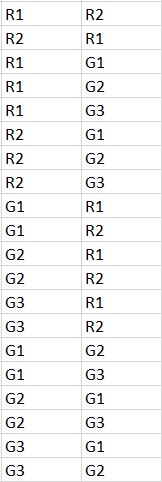
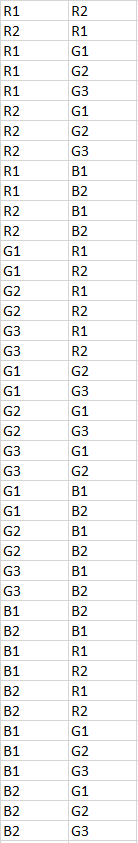
(1,5) (2,4) (3,3) (4,2) (5,1) (6,6)

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?

**Ans:** **10/21**

There are total 7 balls. Out of that 5 balls are not blue. Number of favorable outcomes i.e. none of the balls drawn is blue is 20. Total number of outcomes (including blue ball) is 42.

Number of favorable outcomes Total number of outcomes



Therefore,

Probability = Number of Favorable Outcomes / Total Number of Outcomes

= 20 / 42

= **10 / 21**

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

**Ans:** Expected number of candies for a randomly selected child

= 1 \* 0.015 + 4 \* 0.20 + 3 \* 0.65 + 5 \* 0.005 + 6 \* .01 + 2 \* 0.120

= 0.015 + 0.80 + 1.95 + 0.025 + 0.06 + 0.240

= **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.



|  |  |  |  |
| --- | --- | --- | --- |
|  | **Points** | **Score** | **Weigh** |
| **Mean** | 3.596563 | 3.21725 | 17.84875 |
| **Median** | 3.695 | 3.325 | 17.71 |
| **Mode** | 3.07 | 3.44 | 17.02 |
| 3.92 |  | 18.9 |
| **Variance** | 0.285881 | 0.957379 | 3.193166 |
| **Standard Deviation** | 0.534679 | 0.978457 | 1.786943 |
| **Range** | 2.17 | 3.911 | 8.4 |

**Ans:**

1. For all 3 columns mean ≠ median ≠ mode, so we can say that distribution is not symmetrical.
2. For two columns i.e. ‘Points’ and ‘Score’, mean < median < mode, so the distribution is skewed to the left i.e. negatively skewed.
3. Two columns i.e. ‘Score’ & ‘Weigh’ contains outliers.

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?

**Ans:** Expected Value of the Weight of a person = **145.33** Pounds

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

**Cars speed and distance**



|  |  |  |
| --- | --- | --- |
|  | **speed** | **dist** |
| **Skewness** | -0.113955 | 0.782484 |
| **Kurtosis** | -0.508994 | 0.405053 |

**Ans:**

Both the columns are skewed in nature means non-symmetric.

‘Speed’ is negatively skewed. Mass of the distribution is concentrated on the right. As the skewness is -0.113955 i.e. between -0.5 and 0.5, the data are fairly symmetrical.

‘Dist’ is positively skewed. Mass of the distribution is concentrated on the left. As the skewness is 0.782484 i.e. between 0.5 and 1, the data are moderately skewed.

Kurtosis of ‘speed’ is negative, so the distribution is light tails.

For ‘dist’, it is positive, so the distribution has heavier tails.

**SP and Weight(WT)**



|  |  |  |
| --- | --- | --- |
|  | **SP** | **WT** |
| **Skewness** | 1.581454 | -0.60331 |
| **Kurtosis** | 2.977329 | 0.950291 |

**Ans:**

Both the columns are skewed in nature means non-symmetric.

‘SP’ is positively skewed. Mass of the distribution is concentrated on the left. As the skewness is 1.581454 i.e. greater than 1, the data highly skewed.

‘WT’ is negatively skewed. Mass of the distribution is concentrated on the right. As the skewness is -0.60331 i.e. between -1 and -0.5, the data are moderately skewed.

Kurtosis of ‘SP’ as well ‘WT’ is positive, so the distribution has heavier tails. There are higher peaks in smaller width for both the columns.

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



**Ans:** This data is positively skewed. Many chicken are having weight between 50-100 gm. Chickens with weight more than 300 are very less.

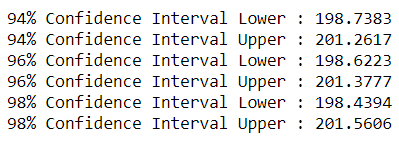


**Ans:** Data represented by boxplot is positively skewed. There are some outliers in data.

**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 ?

|  |  |  |  |
| --- | --- | --- | --- |
| **Confidence Interval for** | **94%** | **98%** | **96%** |
| **Lower** | 198.7383 | 198.4394 | 198.6223 |
| **Upper** | 201.2617 | 201.5606 | 201.3777 |

**Ans:**

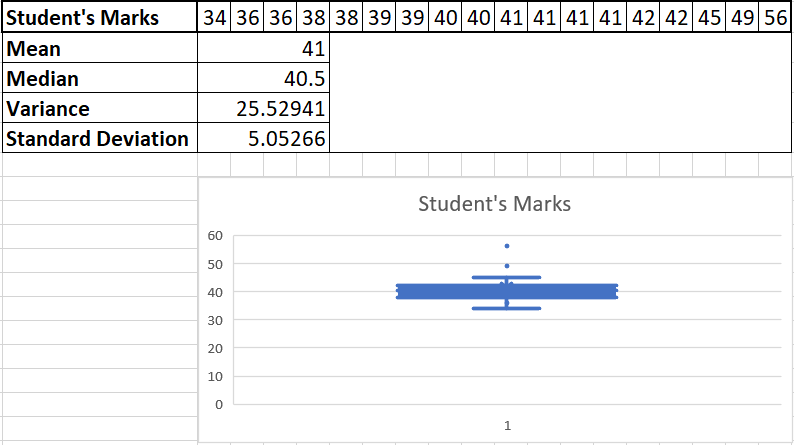


**Q12)** Below are the scores obtained by a student in tests

**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:**



1. What can we say about the student marks?

**Ans:** Marks are not normally distributed. There are some outliers in the data. Four students are having 41 mark i.e. modal value of the student’s marks.

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

**Ans:** In case of mean and median of the data are equal, there will be a perfectly symmetrical distribution. In other words, there won’t be any skewness i.e. skewness = 0.

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

**Ans:** In case of mean > median, data will be positively skewed.

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

**Ans:** In case of median > mean, data will be negatively skewed.

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

**Ans:** A distribution with a positive kurtosis value indicates that the distribution has heavier tails and a sharper peak than the normal distribution.

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

**Ans:** A distribution with a negative kurtosis value indicates that the distribution has lighter tails and a flatter peak than the normal distribution.

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



What can we say about the distribution of the data?

**Ans:** Data is not symmetric. Most of the data is concentrated towards right side.

What is nature of skewness of the data?

**Ans:** The data is negatively skewed.

What will be the IQR of the data (approximately)?   
**Ans:** Approximate IQR of the data is 9 (18-10).

Q19) Comment on the below Boxplot visualizations?



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

**Ans:** Data in both boxplots are symmetrically distributed. There is no outlier in both the boxplot. Both seems to have same median. Boxplot1 has less variability, less variation, less standard deviation, less range, less interquartile range value as compared to boxplot2.

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)

**Ans:** a. P(MPG>38) = 0.407407

b. P(MPG<40) = 0.753086

c. P (20<MPG<50) = 0.851852

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans:** MPG of Cars follows 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

**Ans:** Both of the columns from wc-at data set doesn’t follows normal distribution.

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

(1-.9)/2 = .1/2 = .05

1-.05=.95

**Ans:** Z score of 90% confidence interval = 1.644854

Z score of 94% confidence interval = 1.880794

Z score of 60% confidence interval = 0.841621

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

**Ans:** t score of 95% confidence interval = 2.063899

t score of 96% confidence interval = 2.171545

t score of 99% confidence interval = 2.79694

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

**Ans:** t-score = -0.4714

Degree of Freedom = 17

P(t) = 0.321673