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

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

1. The probability of two heads and one tail when three coins are tossed is 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
4. a) The probability of sum is equal to 1 is 0/36

b) The probability of sum is lass than or equal to 4 is 6/36

c) The probability of sum is divisible by 2 and 3 is 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?

1. Total number of balls = 7

Number of balls except blue = 7-2 =5

The probability that none of the balls drawn is blue = Therefore, the probability is 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

1. The probability count of candies for children is =

Therefore, the probability is 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**

1. **POINTS**

Mean of Points is 3.5965

Median of Points is 3.6950

Mode of Points is 3.07, 3.92

Variance of Points is 0.2858

Std of Points is 0.5346

Range (2.7600, 4.9300)

**SCORE**

Mean of Score is 3.2172

Median of Score is 3.3250

Mode of Score is 3.44

Variance of Points is 0.9573

Std of Score is 0.9784

Range (1.5130, 5.4240)

**WEIGH**

Mean of Weigh is 17.8487

Median of Weigh is 17.7100

Mode of Weigh is 17.02, 18.90

Variance of Points is 3.1931

Std of Weigh is 1.7869

Range (14.5000, 22.9000)

**Inference:** As we observe data doesn’t consists null values. The data type of the first column is object value, and the remaining columns are float values.

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?

1. From the given data, 108, 110, 123, 134, 135, 145, 167, 187, 199

The probability weight of random patient is =

The expected value of the weight of a random patient is 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

1. **SPEED**

The Skewness of Cars speed is -0.1175

The Kurtosis of Cars speed is 0.8068

**DISTANCE**

The Skewness of Cars distance is -0.50899

The Kurtosis of Cars distance is 0.40505

**INFERENCE**

Hence, both cars speed and distance are positively skewed (mean > median) and Platykurtic (<3).

**SP and Weight(WT)**

**Use Q9\_b.csv**

1. **SP**

The Skewness of Cars SP is 1.6114

The Kurtosis of Cars SP is 2.9773

**WT**

The Skewness of Cars WT is -0.6147

The Kurtosis of Cars WT is 0.9502

**INFERENCE**

Hence, the SP of the car is positively skewed (mean > median) and Platykurtic (<3).

Hence, the weight of the car is negatively skewed (mean < median) and Platykurtic (<3).

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



1. **INFERENCE**

In the given histogram the mean is greater than the median.

So, it is positively skewed.

In the given boxplot, the distance between median and minimum is lesser than the distance between median and maximum. So, it is positively skewed.

**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?

1. Given, the sample size(n) = 2000

Mean of sample(x bar) = 200 pounds

Standard Deviation of sample(S) = 30 pounds

Confidence Intervals for 94% : (143.57619175546247, 256.42380824453755)

Confidence Intervals for 98% : (130.2095637787748, 269.7904362212252)

Confidence Intervals for 96% : (138.38753268104531, 261.61246731895466)

**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?
3. The Mean of scores of a student is 41.0

The Median of scores of a student is 40.5

The Variance of scores of a student is 24.1111

The Standard Deviation of scores of a student is 4.9103

According to student’s marks there is less variety of data. The mean is greater than the median. So, we can say that the data is positively skewed.

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

1. The skewness is 0, as the data is Symmetrical when mean and median are equal.

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

1. The Skewness is positive when the mean > median.

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

1. The Skewness is negative when the median > mean.

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

1. The positive kurtosis value indicates the data is Leptokurtic.

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

1. The negative kurtosis value indicates the data is Platykurtic.

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



What can we say about the distribution of the data?

1. Since the distance between minimum to median and maximum to median is not equal, the boxplot is Asymmetrical.

What is nature of skewness of the data?

1. The nature of skewness of the data is Negative.

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

1. The IQR of the data is Q3-Q1

Hence, 18 – 10 = 8.  
So, the IQR of the data is 8.

Q19) Comment on the below Boxplot visualizations?



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

1. **INFERENCE**

The medians of the both boxplots are same. So, there is no much difference between 1st and 2nd boxplot.

The IQR of 1st boxplot is less compared to the 2nd boxplot.

The 1st boxplot has small whisker range. So, 1st boxplot has the less0escattered data than the 2nd boxplot.

The 1st boxplot is positively skewed whether the 2nd boxplot is symmetrically skewed.

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)
  3. P (20<MPG<50)
     1. The probability of MPG of cars for the below cases

a) P(MPG > 38)

0.3475

b) P(MPG < 40)

0.7293

c) P(20 < MPG < 50)

0.8988

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

1. As the mean, median, and mode are not equal. The data is asymmetric and platykurtic. So, the MPG of cars doesn’t follow normal distribution.
2. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

1. As the mean, median, and mode are not equal. The data is asymmetric and platykurtic. So, the given data doesn’t follow normal distribution.

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

1. Z-score of 90% confidence Interval: 1.6448536269514722

Z-score of 94% confidence Interval: 1.8807936081512509

Z-score of 60% confidence Interval: 0.8416212335729143

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

1. The t-score values for the 95%,96% and 99% confidence intervals are:

T-score for 95% Confidence Interval is

2.0638985616280205

T-score for 96% Confidence Interval is

2.1715446760080677

T-score for 99% Confidence Interval is

2.796939504772804

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

1. Null Hypothesis(H0): The average life of a bulb is, x >= 260

Alternate Hypothesis(H1): The average life of a bulb is, x < 260

To calculate the t-score value, t-score=

Here, = Sample mean = 260

= Population mean = 270

= Sample standard deviation = 90

= Sample size = 18

Now the t score value is,

The value of t-score is -0.4714045207910317

The probability value of 18 randomly selected bulbs would have an average life of < 260 days is 0.32167253567098353.

Since, p\_value > 0.05, we accept null hypothesis.

So, the average life of a bulb is, x >= 260.