|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete data type |
| Results of rolling a dice | Discrete data type |
| Weight of a person | Continuous data type |
| Weight of Gold | Continuous data type |
| Distance between two places | Continuous data type |
| Length of a leaf | Continuous data type |
| Dog's weight | Continuous data type |
| Blue Color | Discrete data type |
| Number of kids | Discrete data type |
| Number of tickets in Indian railways | Discrete data type |
| Number of times married | Discrete data type |
| Gender (Male or Female) | Discrete data type |

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. Probability of two heads and one tail are 3/8.

When three coins are tossed up the total number of possible combinations are: 2\*2\*2= 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

Therefore, the Probability of getting two heads and one tails in the toss of three coins simultaneously is

P(Two Heads and One Tail) = Number of desired outcomes/Total no. of Outcomes

= 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

Ans.

Analysis: When two dice are rolled the total number of possible combinations are 6\*6=36

These combinations are (1,1), (1,2), (1,3), (1,4), (1,5), (1,6)

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

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

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

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

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

a) Answer : 0

The probability of the sum equal to 1 is 0 because they starts with (1,1).

b) Answer: 1/6

the sum is less than or equal to 4 the possible outcomes are (1,1),(1,2),(1,3),(2,1),(2,2),(3,1)

therefore probability = 6/36 = 1/6

c) Answer: 1/6

the sum is divisible by 2 and 3 the possible outcomes are (1,5), (2,4), (3,3), (4,2), (5,1), (6,6) therefore probability= 6/36 = 1/6

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. Probability that none of the balls drawn is blue =5C2/7C2=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 \* 0.01 + 2 \* 0.12

= 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24

= 3.090

Expected number of candies for a 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, Weight>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

Ans.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weight |
| Mean | 3.596563 | 3.21725 | 17.84875 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.92 | 3.44 | 17.02 |
| Standard deviation | 0.534679 | 0.978457 | 1.786943 |
| Range | 2.17 | 3.911 | 8.4 |

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 = ∑ (probability \* Value)

∑ P(x).E(x)

There are 9 patients

Probability of selecting each patient = 1/9

Expected Value = (1/9) (108) + (1/9) 110 + (1/9)123 + (1/9)134 + (1/9)135 + (1/9)145 + (1/9(167) + (1/9)187 + (1/9)199

= (1/9) (108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

= (1/9) (1308)

Therefore, Expected value of the weight = 145.33

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

**Ans.**

**Cars speed and distance**

|  |  |  |
| --- | --- | --- |
|  | Skewness | Kurtosis |
| Speed | -0.11751 | -0.50899 |
| Distance | 0.806895 | 0.405053 |

**Inferences:**

**For Speed** **-** Negative values of the skewness indicates that data is skewed towards left and we can say that left tail is long realtive than right. Distribution has a negative kurtosis and it is said to be platy kurtic, which means , that it has flatter peaks and thinner tails compared to normal distribution

**For Distance** - Positive Skewness and positive value of kurtosis observed,

In a positive skewness ,the tail of the distribution curve is longer on the right side. This means that the outliers of the distribution curve are further out towards right and closer to the mean on the left.

Positive value of kurtosis indicates that distribution is peaked and possesses thick tail and is known as Leptokurtic.

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | Skewness | Kurtosis |
| SP | 1.61145 | -0.61475 |
| Weight(WT) | 2.977329 | 0.950291 |

**For SP :** Postive skewness and negative Kurtosis observed .

In a positive skewness , the tail of the distribution curve is longer on the right side. This means that the outliers of the distribution curve are further out towards right and closer to the mean on the left.

Distribution has a negative kurtosis and it is said to be platy kurtic, which means , that it has flatter peaks and thinner tails compared to normal distribution.

**For Weight** : Postive skewness and positive Kurtosis observed.

In a positive skewness , the tail of the distribution curve is longer on the right side. This means that the outliers of the distribution curve are further out towards right and closer to the mean on the left.

Positive value of kurtosis indicates that distribution is peaked and possesses thick tail and is known as Leptokurtic.

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



Answer:

Histogram is right or positively skewed and the outliers are above the upper side of boxplot.

**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?

Ans.

Mean = 200 , SD= 30 , n =2000

1. 94% confidence interval at 201.26-198.73 & z value=1.88
2. 96% confidence interval at 201.38-198.61 & z value=2.053
3. 98% confidence interval at 201.57-198.42 & z value=2.32

**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.

|  |  |
| --- | --- |
| Mean | 41 |
| Median | 40.5 |
| Variance | 25.52941 |
| Standard Deviation | 5.052664 |

1. What can we say about the student marks?

Ans. Most of students marks at the range of between 35-45. It is positive skewness.

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

Sol: Data is Normalized and there is zero skewness.

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

Sol: Positive skewness implies mass of the distribution concentrated on Right side.

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

Sol: Negative skewness implies mass of the distribution concentrated on Left side.

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

Answer:

Positive values of kurtosis indicate that distribution is peaked and possesses thick tails. An extreme positive kurtosis indicates a distribution where more of the numbers are located in the tails of the distribution instead of around the mean.

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

Answer:

A distribution with a negative kurtosis value indicates that the distribution has lighter tails than the normal distribution. Platykurtic distributions have negative kurtosis values. A platykurtic distribution is flatter (less peaked) when compared with the normal distribution, with fewer values in its shorter (i.e., lighter, and thinner) tails.

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



What can we say about the distribution of the data?

Sol: Not normal distribution , the median is towards the higher value.

What is nature of skewness of the data?

Sol: Negative skewness

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

Sol: 18(Upper quartile)-10(Lower quartile) = 8

Q19) Comment on the below Boxplot visualizations?



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

Ans. Boxplot 1: Q1=250 Boxplot 2: Q1=225

Q3=275 Q2=300

Q2=265 (median) Q2=265 (median)

From both boxplots the median are same. There are no outliers.

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

Ans.

* 1. P(MPG>38)

Ans. 0.48

* 1. P(MPG<40)

Ans. 0.52

* 1. P (20<MPG<50)

Ans. 0.14

Q 21) Check whether the data follows normal distribution

Ans.

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans.

It follows as 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. AT and Waist not follows Normal distribution.

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

Ans. Z score of 90% = 1.645

Z score of 94% = 1.89

Z score of 60% = 0.84

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

Ans. t scores for 95% confidence interval = 2.06

96% confidence interval = 2.17

99% confidence interval =2.796

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. The p value is 0.52 .