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

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 | Nominal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Interval |

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

Ans. The combinations that result in two heads and one tail are:

1. H, H, T

2. H, T, H

3. T, H, H

So, there are three favorable outcomes.

The probability of getting two heads and one tail can be calculated using the formula:

(Two Heads and One Tail)=Number of Favorable OutcomesTotal Number of OutcomesP(Two Heads and One Tail)=Total Number of OutcomesNumber of Favorable Outcomes

(Two Heads and One Tail)=38P(Two Heads and One Tail)=83

So, the probability of getting two heads and one tail is 3883 or 37.5%.

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. When two dice are rolled, there are \(6 \times 6 = 36\) possible outcomes since each die has 6 sides.

a) Probability that the sum is equal to 1:

There is only one way to get a sum of 1, which is when both dice show a 1. So, the probability is:

[P(text{Sum = 1}) = frac{text{Number of ways to get a sum of 1}}{text{Total number of outcomes}} = frac{1}{36}]

b) Probability that the sum is less than or equal to 4:

The possible combinations for sums less than or equal to 4 are:

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

(1, 4), (2, 3), (3, 2), (4, 1), (2, 4), (4, 2), (3, 3)]

So, there are 13 favorable outcomes.

[P(\text{Sum} \leq 4) = \frac{\text{Number of favorable outcomes}}{\text{Total number of outcomes}} = \frac{13}{36}\]

c) Probability that the sum is divisible by both 2 and 3:

In order for the sum to be divisible by both 2 and 3, it must be divisible by 6. The possible combinations for sums that are divisible by 6 are:

\[

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

\]

So, there are 5 favorable outcomes.

\[P(\text{Sum divisible by 2 and 3}) = \frac{\text{Number of favorable outcomes}}{\text{Total number of outcomes}} = \frac{5}{36}\]

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?

Solution:

There are a total of 7 balls in the bag (2 red + 3 green + 2 blue). You want to draw 2 balls without getting any blue ones, so we can think about this problem in two steps:

1. Probability of not drawing a blue ball on the first draw: There are 5 non-blue balls (2 red + 3 green), so the probability of picking one on the first draw is 5/7.
2. Probability of not drawing a blue ball on the second draw: After taking out the first ball, there are 6 balls left. Since the first draw was non-blue, there are still 5 non-blue balls remaining. So, the probability of picking another non-blue ball on the second draw is 5/6.

Now, to get the overall probability of not drawing any blue balls, we need to multiply these two probabilities together:

(Probability of not blue on draw 1) \* (Probability of not blue on draw 2) = (5/7) \* (5/6) = 25/42

Therefore, the probability of drawing two balls without getting any blue ones is 25/42, which is approximately 0.5952 or 59.52%.

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

The expected number of candies for a randomly selected child (often denoted as *E*(*X*) or *μ*) can be calculated by multiplying each candy count by its corresponding probability and then summing up these products for all children. The formula for the expected value is:

*E*(*X*)=∑*i*​(*Xi*​⋅*P*(*Xi*​))

where *Xi*​ is the candy count for child *i* and *P*(*Xi*​) is the probability of child *i* having that candy count.

Let's calculate it for the given data:

E(X)=(1⋅0.015)+(4⋅0.20)+(3⋅0.65)+(5⋅0.005)+(6⋅0.01)+(2⋅0.120)*E*(*X*)=(1⋅0.015)+(4⋅0.20)+(3⋅0.65)+(5⋅0.005)+(6⋅0.01)+(2⋅0.120)

E(X)=0.015+0.80+1.95+0.025+0.06+0.24*E*(*X*)=0.015+0.80+1.95+0.025+0.06+0.24

E(X)=3.14*E*(*X*)=3.14

Therefore, the expected number of candies for a randomly selected child is 3.14.

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**

Ans.

Q8) Calculate Expected Value for the problem below

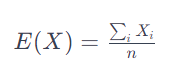
1. The weights (X) of patients at a clinic (in pounds), are

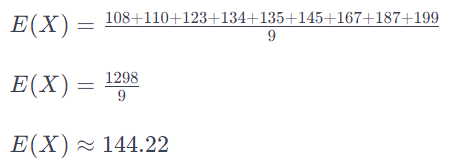
108, 110, 123, 134, 135, 145, 167, 187, 199

Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

Solution:

To calculate the expected value (mean) for the weights of the patients, you add up all the weights and then divide by the number of patients. The formula for the expected value ( (E(X))) is:





Therefore, the expected value of the weight of a randomly chosen patient is approximately 144.22 pounds.

**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**



**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.
2. What can we say about the student marks?

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

Ans: When the mean, median, and mode of a distribution are equal, it indicates that the distribution is symmetric. In a symmetric distribution, the data is evenly distributed on both sides of the center, and there is a balance between the left and right tails.

For a perfectly symmetric distribution, the skewness is zero. Skewness is a measure of the asymmetry of a probability distribution. When the mean, median, and mode are equal, the distribution is perfectly centered, and there is no skewness present. This is characteristic of a symmetric distribution where the tails on both sides of the central point are mirror images of each other.

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

When the mean is greater than the median, it suggests that the distribution of the data is right-skewed or positively skewed. In a right-skewed distribution, the tail on the right-hand side is longer or fatter than the left-hand side. This means that there are some relatively large values that are pulling the mean to the right.

In a right-skewed distribution:

- The mean is greater than the median.

- The median is closer to the left tail.

- The right tail is longer.

The skewness value is positive in a right-skewed distribution, indicating the direction and degree of skewness. A positive skewness value means that the distribution is skewed to the right.

It's important to note that skewness provides information about the shape of the distribution but doesn't provide details about specific outliers or extreme values. The skewness value helps in understanding the general asymmetry of the distribution.

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

When the median is greater than the mean, it suggests that the distribution of the data is left-skewed or negatively skewed. In a left-skewed distribution, the tail on the left-hand side is longer or fatter than the right-hand side. This indicates that there are some relatively small values that are pulling the mean to the left.

In a left-skewed distribution:

- The median is greater than the mean.

- The median is closer to the right tail.

- The left tail is longer.

The skewness value is negative in a left-skewed distribution, indicating the direction and degree of skewness. A negative skewness value means that the distribution is skewed to the left.

Again, it's important to note that skewness provides information about the shape of the distribution but doesn't provide details about specific outliers or extreme values. The skewness value helps in understanding the general asymmetry of the distribution.

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

Positive kurtosis indicates that the distribution of a dataset has heavier tails and a sharper peak (more peaked or leptokurtic) compared to a normal distribution. Kurtosis is a measure of the "tailedness" or the degree of peakedness of a distribution.

In a distribution with positive kurtosis:

- The tails of the distribution are fatter than those of a normal distribution.

- There is an increased probability of extreme values (outliers) in the distribution.

- The peak of the distribution is higher and sharper than that of a normal distribution.

Positive kurtosis can suggest that a dataset has more observations in the tails and potentially more outliers or extreme values than would be expected in a normal distribution. It is an indication that the distribution has a more pronounced central peak and heavier tails, which can result from factors such as high variability or kurtotic behavior in the underlying data.

It's important to note that interpreting kurtosis requires considering other measures of central tendency and dispersion, and the context of the specific dataset being analyzed.

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

A negative kurtosis value indicates that the distribution of a dataset has lighter tails and a flatter peak compared to a normal distribution. This type of distribution is referred to as platykurtic. In a platykurtic distribution:

- The tails of the distribution are thinner than those of a normal distribution.

- There is a decreased probability of extreme values (outliers) in the distribution.

- The peak of the distribution is lower and broader than that of a normal distribution.

Negative kurtosis suggests that the distribution has fewer observations in the tails and is less prone to extreme values or outliers compared to a normal distribution. It indicates a flatter and more spread-out shape in the tails.

In summary, negative kurtosis typically means that the distribution has less mass in the tails and a less pronounced peak than a normal distribution. It is important to interpret kurtosis in the context of the specific dataset and consider other measures of central tendency and dispersion for a comprehensive understanding of the data's distribution.

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

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

Q19) Comment on the below Boxplot visualizations?



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

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