Section 2

Dataset:
Student Name: [Adam, Bella, Charlie, David, Emily, Frank, Grace, Harry, Isabella, Jack]
Score: [56, 78, 90, 82, 60, 95, 72, 88, 80, 76]

Problem: Calculate the mean score of the class. What does this tell you about the overall performance of the class?

Answer:

Mean = Sum of all the observations / No. of observations

Sum of the observations = 777
Total no of observations = 10

Therefore, Mean = 777/10

Mean = 77.7

The overall performance of the class is good with the average of 77.7, which means that most students score is clustered around this value. Although, some students also occupy extreme values which makes this average slightly inconsistent of any further concreate conclusion.

Problem: Calculate the median score of the class. How does it compare to the mean and what does this comparison tell you about the distribution of scores?

Answer:

To calculate the Median, we arrange the score in an order of magnitude and find the middle occupying value.

Ascending order : 56, 60, 72, 76, 78, 80, 82, 88, 90, 95

Since the no. of observation is even, we take the average of the middle two observations.

78 + 80 / 2 = 158 / 2 = 79

So, Median = 79 and Mean = 77.7

Since the median is closer to the mean, it concludes the symmetric distribution of the observations. Which means that there is a uniform distribution of scores and very few extreme values.

Problem: Calculate the mode of the scores. What does this tell you about the most common score in the class?

Answer:

Mode is most frequent occurring value in the observation.

Here, 56, 78, 90, 82, 60, 95, 72, 88, 80, 76

As, there no value repeating itself more than once, hence the following data has no mode.

No mode signifies, the data is relatively uniform and evenly distributed across different points.

Problem: Based on the mean, median, and mode you calculated, describe the performance of the class.

Answer:

1. The class performance when considered mean and median signifies the decent performance of the class.
2. It also says that the data is not heavily skewed to any of the sides but is symmetric as mean and median are close to each other.
3. There is no mode, concluding the absence of any peaks in the data.
4. The scores are uniformly distributed with no repetitions and lacks a strong central tendency.

Problem: What could be the possible implications if the mode is significantly lower or higher than the mean and median?

Answer:

Case 1: Mode is significantly lower than the mean and median.

> In this case, it can be interpreted as the more than one score is less than the central tendency making the data distribution skewed and not uniform, which in this case would be Positively Skewed.
>
> It tells us, more concentration of scores at values less than mean and median. Positive skewed data shows relatively poor performance of students compared to uniform distribution of no mode.

Case 2: Mode is significantly higher than the mean and median.

> In this case, more than one scores are more than the mean and median. Which signifies the skewed distribution of data in Negative Skewed fashion.
>
> Here, the this can be good sign, as more than one student are performing well in the exams, more than average.

The mode when having two values suggests that two peaks exist in the data, which results in two subgroups in the data. The distribution is not uniform but has two separate patters in it, each with separate range of values.

Multimodal on the other hand concludes of having more than two frequently occurring values in the data set. Each peak maybe symmetrical or maybe not. The multimodal represents multiple subgroups inside the data with each subgroup having its own pattern and characteristics. These subgroups may represent different conditions can be useful in better interpretation of the data.

Here, we see Mode giving completely different approach of interpreting the data, which can be insightful.
The mean is sensitive to the outliers and the median returns the middle value but mode is helpful with the categorical data. It helps in better understanding of the subgroups which cannot be identified by mean and median.
But, mode does not take into account the outliers and the range of values.
In a case where the median and mode are over the higher sides but of at least one repeated occurrence is over the lower side than the mean and median can result into the positively skewed distribution, which can give completely different interpretation of the data than mean and mode which showing uniform distribution and vice versa.