Summary

Descriptive vs Inferential Statistics

Descriptive statistics involves describing the data and summarizing it in a meaningful way, it helps in organizing the data and presenting it in a meaningful way. Inferential statistics involves making inferences about the data, it helps in drawing conclusions.

Mean, Median, and Mode

Mean: The average of all the given data points. It is the sum of the data points divided by the number of data points. It is a measure of central tendency.

Median: The middle of the data points ordered ascendingly. It is a measure of central tendency.

Mode: The data point that appears most frequently in a data set is the mode. If all data points appear an equal number of times, then there is no mode.

Variance

It is a measure of how the data is dispersed, how much they vary from the mean. It is calculated by dividing the sum of each data point minus the mean squared by the number of data points minus one.

Standard Deviation

It is a measure of how the data is dispersed, how much they vary from the mean. It is the square root of the variance.

Interquartile range (IQR)

Quartiles: Q1,Q2,Q3:

Q2: is the actual median.

Q1: Median of the lower half of the data, the data located to the left of the median if it were ordered ascendingly.

Q3: Median of the upper half of the data, the data located to the right of the median if it were ordered ascendingly.

Quartiles divide data into four groups each of which is 25% of the whole data set. These divisions help in providing insights into central tendency.

The interquartile range (IQR) is Q3-Q1 which is where the middle 50% of the data lies. It helps in identifying if there are any outliers.

Outliers

Outliers are data points that significantly differ from the rest of the data in a dataset. A data point is an outlier if it lies outside of Q1-1.5*IQR to Q3+1.5IQR.