Task4-notes

Measures of spread

Measures of spread describe how similar or varied the set of observed values are for a particular variable (data item). Measures of spread include the range, quartiles and the interquartile range, variance and standard deviation.

Range

The range is the difference between the smallest value and the largest value in a dataset.

Quartiles

Quartiles divide an ordered dataset into four equal parts, and refer to the values of the point between the quarters. A dataset may also be divided into quintiles (five equal parts) or deciles (ten equal parts).

The lower quartile (Q1) is the point between the lowest 25% of values and the highest 75% of values. It is also called the 25th percentile.

The second quartile (Q2) is the middle of the data set. It is also called the 50th percentile, or the median.

The upper quartile (Q3) is the point between the lowest 75% and highest 25% of values. It is also called the 75th percentile.

Interquartile range

The interquartile range (IQR) is the difference between the upper (Q3) and lower (Q1) quartiles, and describes the middle 50% of values when ordered from lowest to highest. The IQR is often seen as a better measure of spread than the range as it is not affected by outliers.

Variance and standard deviation

The variance and the standard deviation are measures of the spread of the data around the mean. They summarise how close each observed data value is to the mean value.

The standard deviation of a normal distribution enables us to calculate confidence intervals. In a normal distribution, about 68% of the values

are within one standard deviation either side of the mean and about 95% of the scores are within two standard deviations of the mean.

outlier

An outlier is a piece of data that is an abnormal distance from other points. In other words, it's data that lies **outside the other values** in the set.it is an extremely high or extremely low data point relative to the nearest data point and the rest of the neighboring co-existing values in a data graph or dataset you're working with.

An outlier has to satisfy either of the following two conditions:

outlier < Q1 - 1.5(IQR)
outlier > Q3 + 1.5(IQR)

Inferential Statistics

Inferential statistics is a branch of statistics that makes the use of various analytical tools to draw inferences about the population data from sample data.

Inferential statistics help to draw conclusions about the population

Inferential statistics helps to develop a good understanding of the population data by analyzing the samples obtained from it. It helps in making generalizations about the population by using various analytical tests and tools. In order to pick out random samples that will represent the population accurately many sampling techniques are used.

Types of Inferential Statistics

Inferential statistics can be classified into hypothesis testing and regression analysis