

Statistics Assignment 2

1. How can we figure out what the interquartile range is?
2. What exactly is the value of the 5-number theory?
3. What is the relationship between standard deviation and variance?
4. What does the difference between variance and standard deviation mean?
5. When is it appropriate to refer to a skewed data distribution?

1. i) Order the data from least to greatest

ii) Find the median

iii) Calculate the median of both the lower and upper half of the data

iv) The IQR is the difference between the upper and lower medians.

2. There are a variety of descriptive statistics. Numbers such as the mean, median, mode, skewness, kurtosis, standard deviation, first quartile and third quartile, to name a few each tell us something about our data. Rather than looking at these descriptive statistics individually sometimes combining them help to give us a complete picture. With this end in mind, the 5 number summary is a convenient way to combine five descriptive statistics.

It is clear that there are to be 5 number's in our summary, but which 5? The

number chosen are to help us know the center of our data, as well as how spread out the data points are. With this in mind the 5 number summary consists of the following.

The minimum- this is the smallest value in our dataset

The 1st quartile- this number is denoted as Q_1 and 25% of our data falls below the 1st quartile.

The median- this is the midway point of the data. 50% of all data falls below the median.

The 3rd quartile- this number is denoted by Q_3 and 75% of our data falls

below the 3rd quartile.

The maximum- This is the largest value in our dataset.

The mean and standard deviation also be used together to convey the center and the spread of a set of data. However, both of these statistics are susceptible to outliers. The median, first quartiles and third quartiles are not as heavily influenced by outliers.

5 number summaries can be compared to one another. We will find that two sets with similar means and standard deviation may have very difference 5 number summaries. To easily compare two 5 number summaries at a glance, we can use a box plot or box and whiskers graph.

3. Standard deviation is the spread of a group of numbers from the mean .

The variance measures the average degree to which each point differs from the mean. While standard deviation is the square root of the variance, variance is the average of all data points within a group.

4. Variance- it can simply be defined as the numerical value, which describes how variable the observations are. Variance is nothing but the average taken out of the squared deviation. Variance is expressed in squared units. It is mathematically denoted as σ^2 . Variance is a perfect indicator of the individuals spread out in a group.

Standard Deviation- It can simply be defined as the observation that get

measured through dispersion within a dataset. It is defined as the root of the mean square deviation. It is expressed in the same units of data available. It is mathematically denoted as σ . Standard deviation is the perfect indicator of the observations in a data set.

5. Skewed data is cumbersome and common. It is often desirable to transform skewed data and to convert into values between 0 to 1. Standard functions used for such conversions include normalization, the sigmoid, log, cube root, the hyperbolic tangent.

.

iNeuron