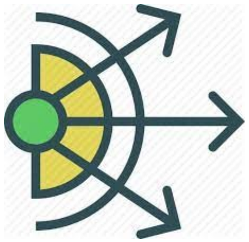
Measure of Dispersion



Tushar B. Kute,
http://tusharkute.com

Dispersion

- Dispersion refers to measures of how spread out our data is.
- Typically they're statistics for which values near zero signify not spread out at all and for which large values (whatever that means) signify very spread out.

Dispersion - Types

- Absolute Deviation from Mean
- Variance
- Standard Deviation
- Range
- Quartiles
- Skewness
- Kurtosis

Reference: https://towardsdatascience.com

Mean Absolute Deviation

- The Absolute Deviation from Mean, also called Mean Absolute Deviation (MAD), describe the variation in the data set, in sense that it tells the average absolute distance of each data point in the set.
- It is calculated as,

$$Mean\ \ Absolute\ \ Deviation\ = \frac{1}{N}\sum_{i=1}^{N}\left|X_i - \overline{X}\right|$$

Variance

- In statistics, the variance is a measure of how far individual (numeric) values in a dataset are from the mean or average value.
- The variance is often used to quantify spread or dispersion. Spread is a characteristic of a sample or population that describes how much variability there is in it.
- A high variance tells us that the values in our dataset are far from their mean. So, our data will have high levels of variability.
- On the other hand, a low variance tells us that the values are quite close to the mean. In this case, the data will have low levels of variability.

Variance

- To calculate the variance in a dataset, we first need to find the difference between each individual value and the mean. The variance is the average of the squares of those differences. We can express the variance with the following math expression:

$$\sigma^2 = \frac{1}{n}\sum_{i=0}^{n-1}(x_i - \mu)^2$$

- In this equation, xi stands for individual values or observations in a dataset. μ stands for the mean or average of those values. n is the number of values in the dataset.
- The term xi - μ is called the deviation from the mean. So, the variance is the mean of square deviations. That's why we denoted it as σ^2.

Variance

Say we have a dataset [3, 5, 2, 7, 1, 3]. To find its variance, we need to calculate the mean which is:

$$(3 + 5 + 2 + 7 + 1 + 3)/6 = 3.5$$

Then, we need to calculate the sum of the square deviation from the mean of all the observations. Here's how:

$$(3 - 3.5)^2 + (5 - 3.5)^2 + (2 - 3.5)^2 + (7 - 3.5)^2 + (1 - 3.5)^2 + (3 - 3.5)^2 = 23.5$$

To find the variance, we just need to divide this result by the number of observations like this:

$$23.5/6 = 3.916666667$$

Variance

- That's all. The variance of our data is 3.916666667. The variance is difficult to understand and interpret, particularly how strange its units are.
- For example, if the observations in our dataset are measured in pounds, then the variance will be measured in square pounds.
- So, we can say that the observations are, on average, 3.916666667 square pounds far from the mean 3.5.
- Fortunately, the standard deviation comes to fix this problem

Standard Deviation

- Standard deviation is a number used to tell how measurements for a group are spread out from the average (mean or expected value).
- A low standard deviation means that most of the numbers are close to the average, while a high standard deviation means that the numbers are more spread out.
- The reported margin of error is usually twice the standard deviation. Scientists commonly report the standard deviation of numbers from the average number in experiments.

tusharkute.com

Standard Deviation

- The standard deviation measures the amount of variation or dispersion of a set of numeric values.
- Standard deviation is the square root of variance σ2 and is denoted as σ.
- So, if we want to calculate the standard deviation, then all we just have to do is to take the square root of the variance as follows:

$$Std\ Deviation = \sqrt{Variance} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2}$$

Standard Deviation

- Again, we need to distinguish between the population standard deviation, which is the square root of the population variance ($\sigma 2$) and the sample standard deviation, which is the square root of the sample variance ($S2$).
- We'll denote the sample standard deviation as S:

$$S=\sqrt{S2}$$

tusharkute.com

Standard Deviation

- There are six steps for finding the standard deviation:
  - List each score and find their mean.
  - Subtract the mean from each score to get the deviation from the mean.
  - Square each of these deviations.
  - Add up all of the squared deviations.
  - Divide the sum of the squared deviations by N – 1.
  - Find the square root of the number you found.

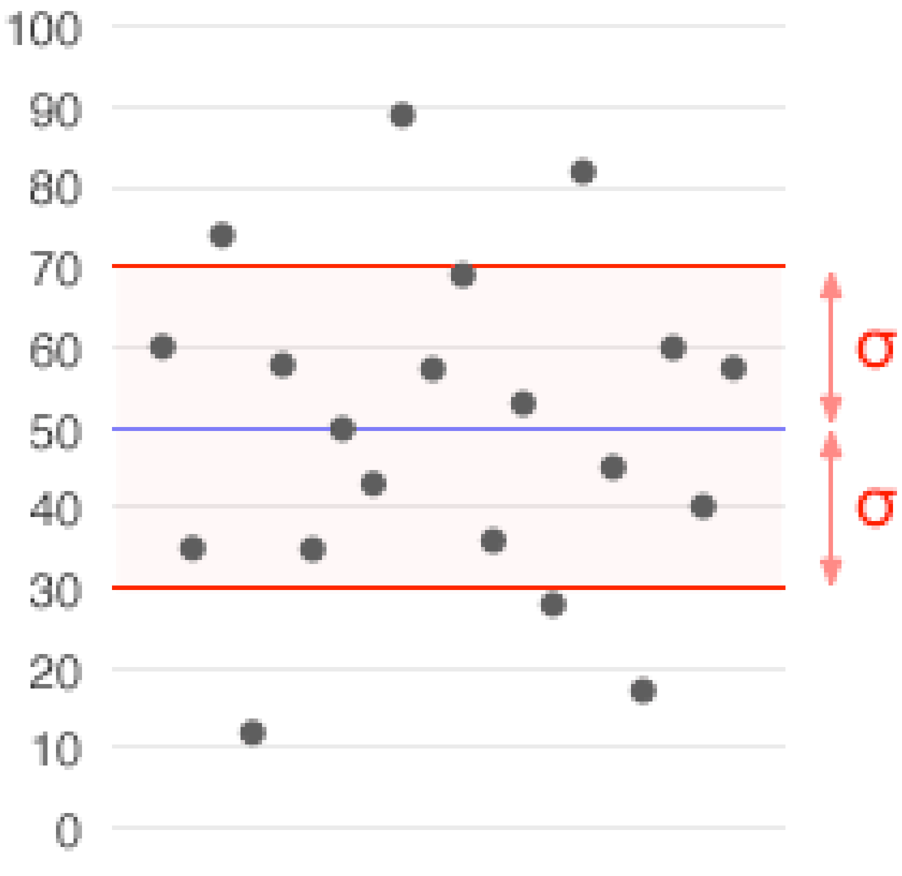
Standard Deviation

- Low values of standard deviation tell us that individual values are closer to the mean.
- High values, on the other hand, tell us that individual observations are far away from the mean of the data.
- Values that are within one standard deviation of the mean can be thought of as fairly typical, whereas values that are three or more standard deviations away from the mean can be considered much more atypical. They're also known as outliers.

Standard Deviation : Example

- The average height for grown men in the United States is 70", with a standard deviation of 3".
- A standard deviation of 3” means that most men (about 68%, assuming a normal distribution) have a height 3" taller to 3” shorter than the average (67"–73") — one standard deviation.
- Almost all men (about 95%) have a height 6” taller to 6” shorter than the average (64"–76") — two standard deviations. Three standard deviations include all the numbers for 99.7% of the sample population being studied. This is true if the distribution is normal (bell-shaped).

Standard Deviation



Standard Deviation

If we're trying to estimate the standard deviation of the population using a sample of data, then we'll be better served using **n - 1** degrees of freedom. Here's a math expression that we typically use to estimate the population variance:

$$\sigma_x = \sqrt{\frac{\sum_{i=0}^{n-1} (x_i - \mu_x)^2}{n-1}}$$

Note that this is the square root of the sample variance with **n - 1** degrees of freedom. This is equivalent to say:

$$S_{n-1} = \sqrt{S_{n-1}^2}$$

tusharkute.com

Range

- Range is the difference between the Maximum value and the Minimum value in the data set.
- It is given as,

$$\text{range} = \text{maximum} - \text{minimum}$$

tusharkute.com

Thank you

This presentation is created using LibreOffice Impress 5.1.6.2, can be used freely as per GNU General Public License

@mitu_skillologies

/mITuSkillologies

@mitu_group

/company/mitu-skillologies

MITUSkillologies

Web Resources
https://mitu.co.in
http://tusharkute.com

contact@mitu.co.in

tushar@tusharkute.com