What is the relationship between mean and standard deviation, and can you provide an example?

Authored by

stats writer

June 27, 2024

The mean and standard deviation are two important measures of central tendency and variability in a set of data. The mean, also known as the arithmetic average, is the sum of all the values in a data set divided by the number of values. It represents the central point around which the data is distributed. The standard deviation, on the other hand, measures how much the data deviates from the mean. It is calculated by taking the square root of the average of the squared differences between each data point and the mean. In simpler terms, it shows how spread out the data is.

The relationship between the mean and standard deviation can be understood as follows: if the data is normally distributed, about 68% of the values will fall within one standard deviation of the mean, and about 95% will fall within two standard deviations. This means that a higher standard deviation indicates a larger spread of data, while a lower standard deviation indicates a more clustered set of data around the mean.

For example, let's consider the heights of 10 students in a class: 150cm, 155cm, 160cm, 165cm, 170cm, 175cm, 180cm, 185cm, 190cm, 195cm. The mean height is 175cm. If we calculate the standard deviation, we get a value of 14.14cm. This means that about 68% of the students have heights within 14.14cm of the mean, i.e. between 160.86cm and 189.14cm. Similarly, about 95% of the students have heights within 28.28cm of the mean, i.e. between 146.72cm and 203.28cm. This example illustrates the relationship between mean and standard deviation, and how they can be used to understand the distribution of data.

The Relationship Between Mean & Standard Deviation (With Example)

The mean represents the average value in a dataset.

It is calculated as:

Sample mean = Σxi / n

where:

Σ: A symbol that means "sum"xi: The ith observation in

a datasetn: The total number of observations in the dataset

The standard deviation represents how spread out the values are in a dataset relative to the mean.

It is calculated as:

Sample standard deviation = $\sqrt{\Sigma}$(xi - xbar)2 / (n-1)

where:

Σ: A symbol that means "sum"xi: The ith value in the samplexbar: The mean of the samplen: The sample size

Notice the relationship between the mean and standard deviation: The mean is used in the formula to calculate the standard deviation.

In fact, we can't calculate the standard deviation of a sample unless we know the sample mean.

The following example shows how to calculate the sample mean and sample standard deviation for a dataset in practice.

Example: Calculating the Mean & Standard Deviation for a Dataset

Suppose we have the following dataset that shows the points scored by 10 different basketball players:

Player	Points
Adam	22
Bob	14
Chad	15
Dean	18
Eric	19
Frank	8
George	9
Harold	34
Isaiah	30
John	7

We can calculate the sample mean of points scored by using the following formula:

Sample mean = Σxi / nSample mean = (22+14+15+18+19+8+9+34+30+7) / 10Sample mean = 17.6

Once we know the sample mean, we can the plug it into the formula to calculate the sample standard deviation:

Sample standard deviation = √Σ(xi - xbar)2 / (n-1)Sample standard deviation = √((22-17.6)2 + (14-17.6)2 +

(15-17.6)2 + (18-17.6)2 + (19-17.6)2 + (8-17.6)2 + (9-17.6)2 + (34-17.6)2 + (30-17.6)2 + (7-17.6)2) / (10-1)Sample standard deviation = 9.08

The sample standard deviation is 9.08. This represents the average distance between each points value and the sample mean of points.

It's helpful to know both the mean and the standard deviation of a dataset because each metric tells us something different.

The mean gives us an idea of where the "center" value of a dataset is located.

The standard deviation gives us an idea of how spread out the values are around the mean in a dataset. The higher the value for the standard deviation, the more spread out the values are in a sample.

By knowing both of these values, we can know a great deal about the distribution of values in a dataset.

The following tutorials provide additional information about the mean and standard deviation: