

Statistics and probability

COURSE: STATISTICS AND PROB... > UNIT 3

Lesson 5: Variance and standard deviation of a sample

- Sample variance
- Sample standard deviation and bias
- Variance
Not started
- Sample and population standard ...
Not started
- Population and sample standard ...**

Population and sample standard deviation review

Google Classroom

Population and sample standard deviation

Standard deviation measures the spread of a data distribution. It measures the typical distance between each data point and the mean.

The formula we use for standard deviation depends on whether the data is being considered a population of its own, or the data is a sample representing a larger population.

- If the data is being considered a population on its own, we divide by the number of data points, N .
- If the data is a sample from a larger population, we divide by one fewer than the number of data points in the sample, $n - 1$.

Population standard deviation:

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

Sample standard deviation:

$$s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

The steps in each formula are all the same except for one—we divide by one less than the number of data points when dealing with sample data.

We'll go through each formula step by step in the examples below.

Why we divide by $n - 1$ is a pretty complex concept. If you want to learn more about the intuition behind this topic, check out [this video](#).

Population standard deviation

Here's the formula again for population standard deviation:

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

Here's how to calculate population standard deviation:

Step 1: Calculate the mean of the data—this is μ in the formula.

Step 2: Subtract the mean from each data point. These differences are called deviations. Data points below the mean will have negative deviations, and data points above the mean will have positive deviations.

Step 3: Square each deviation to make it positive.

Step 4: Add the squared deviations together.

Step 5: Divide the sum by the number of data points in the population. The result is called the variance.

Step 6: Take the square root of the variance to get the standard deviation.

Example: Population standard deviation

Four friends were comparing their scores on a recent essay.

Calculate the standard deviation of their scores:

6, 2, 3, 1

Step 1: Find the mean.

$$\mu = \frac{6 + 2 + 3 + 1}{4} = \frac{12}{4} = 3$$

The mean is 3 points.

Step 2: Subtract the mean from each score.

Score: x_i	Deviation: $(x_i - \mu)$
6	$6 - 3 = 3$
2	$2 - 3 = -1$
3	$3 - 3 = 0$
1	$1 - 3 = -2$

Step 3: Square each deviation.

Score: x_i	Deviation: $(x_i - \mu)$	Squared deviation: $(x_i - \mu)^2$
6	$6 - 3 = 3$	$(3)^2 = 9$
2	$2 - 3 = -1$	$(-1)^2 = 1$
3	$3 - 3 = 0$	$(0)^2 = 0$
1	$1 - 3 = -2$	$(-2)^2 = 4$

Step 4: Add the squared deviations.

$$9 + 1 + 0 + 4 = 14$$

Step 5: Divide the sum by the number of scores.

$$\frac{14}{4} = 3.5$$

Step 6: Take the square root of the result from Step 5.

$$\sqrt{3.5} \approx 1.87$$

The standard deviation is approximately 1.87.

Want to learn more about population standard deviation? Check out [this video](#).

Want to practice some problems like this? Check out this exercise on [standard deviation of a population](#).

Sample standard deviation

Here's the formula again for sample standard deviation:

$$s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

Here's how to calculate sample standard deviation:

Step 1: Calculate the mean of the data—this is \bar{x} in the formula.

Step 2: Subtract the mean from each data point. These differences are called deviations. Data points below the mean will have negative deviations, and data points above the mean will have positive deviations.

Step 3: Square each deviation to make it positive.

Step 4: Add the squared deviations together.

Step 5: Divide the sum by one less than the number of data points in the sample. The result is called the variance.

Step 6: Take the square root of the variance to get the standard deviation.

Example: Sample standard deviation

A sample of 4 students was taken to see how many pencils they were carrying.

Calculate the sample standard deviation of their responses:

2, 2, 5, 7

Step 1: Find the mean.

$$\bar{x} = \frac{2 + 2 + 5 + 7}{4} = \frac{16}{4} = 4$$

The sample mean is 4 pencils.

Step 2: Subtract the mean from each score.

Pencils: x_i	Deviation: $(x_i - \mu)$
2	$2 - 4 = -2$
2	$2 - 4 = -2$
5	$5 - 4 = 1$
7	$7 - 4 = 3$

Step 3: Square each deviation.

Pencils: x_i	Deviation: $(x_i - \bar{x})$	Squared deviation: $(x_i - \bar{x})^2$
2	$2 - 4 = -2$	$(-2)^2 = 4$
2	$2 - 4 = -2$	$(-2)^2 = 4$
5	$5 - 4 = 1$	$(1)^2 = 1$
7	$7 - 4 = 3$	$(3)^2 = 9$

Step 4: Add the squared deviations.

$$4 + 4 + 1 + 9 = 18$$

Step 5: Divide the sum by one less than the number of data points.

$$\frac{18}{4 - 1} = \frac{18}{3} = 6$$

Step 6: Take the square root of the result from Step 5.

$$\sqrt{6} \approx 2.45$$

The sample standard deviation is approximately 2.45

Want to learn more about sample standard deviation? Check out [this video](#).

Want to practice some problems like this? Check out this exercise on [sample and population standard deviation](#).

[Questions](#) [Tips & Thanks](#)

Want to join the conversation?

To get started, your account must be at least 3 days old, have a verified email address, and have at least 5,000 energy points.

Sort by: Top Voted ▾

 neha.yargal 7 years ago

how to identify that the problem is sample problem or population problem?

[Answer](#) [Comment](#)

(31 votes) ▲ Upvote ▼ Downvote ⚡ Flag more ▾

 Jonathon 7 years ago

Great question! It depends on why you are calculating the standard deviation. In the case of sampling, you are randomly selecting a set of data points for the purpose of **estimating** the true values for mean, standard deviation, etc. In the case of a population problem you are collecting data points from 100% of the subjects you wish to study.

[1 comment](#)

(49 votes) ▲ Upvote ▼ Downvote ⚡ Flag more ▾

 tamjrab 7 years ago

Why standard deviation is a better measure of the diversity in age than the mean?

[Answer](#) [1 comment](#)

(9 votes) ▲ Upvote ▼ Downvote ⚡ Flag more ▾

 Andrea Rizzi 7 years ago

I'll try to give you a quick example that I hope will clarify this. If you picked three people with ages 49, 50, 51, and then other three people with ages 15, 50, 85, you can understand easily that the ages are more "diverse" in the second case. In the first case people are all around 50, while in the second you have a young, a middle-aged, and an old person.

However, in both cases the average is 50! The average cannot pick on this diversity, and in fact it doesn't measure diversity at all, only central tendency. On the other hand, the standard deviation turns out to be 0.8, and 28.6 respectively, and correctly assigns greater "diversity" to the second case. Hope this helps!

[1 comment](#)

(70 votes) ▲ Upvote ▼ Downvote ⚡ Flag more ▾

 Alfonso Parrado 6 years ago

Why do we have to subtract 1 from the total number of individuals when we're dealing with a sample instead of a population? I know how to calculate the sample standard deviation, but I want to know the underlying reason why the formula has that tiny variation

[Answer](#) [2 comments](#)

(17 votes) ▲ Upvote ▼ Downvote ⚡ Flag more ▾

 Pedro IPF 7 years ago

If the sample has about 70% or 80% of the population, should I still use the "n-1" rules?? Or i just divided by n?

[Answer](#) [1 comment](#)

(6 votes) ▲ Upvote ▼ Downvote ⚡ Flag more ▾

 ragetactic27 4 years ago

this is why I hate both love and hate stats. how can you effectively tell whether you need to use a sample or the whole population? (this seems to be the most asked question)

[Answer](#) • [1 comment](#)

(4 votes) ▲ Upvote ▼ Downvote ⚡ Flag more ▾



Saivishnu Tulugu 4 years ago

You have to look at the hints in the question. With popn. you will usually see words like all, true, or whole. For sample, words will be like a representative, sample, this group, etc.

[Comment](#)

(5 votes) ▲ Upvote ▼ Downvote ⚡ Flag more ▾



23altdelana 3 years ago

If a problem is giving you all the grades in both classes from the same test, when you compare those, would you use the standard deviation for population or sample?

[Answer](#) • [Comment](#)

(4 votes) ▲ Upvote ▼ Downvote ⚡ Flag more ▾



Kailie Krombos 3 years ago

If you are assessing ALL of the grades, you will use the population formula to calculate the standard deviation.

A way to remember the difference is that a sample is only a group, a part of a whole. The population is referring to the entire set. So when you are receiving data from the ENTIRE population, you can be confident in using the population formula. If you are only given data from a PART of the group, you know to use the sample formula. I hope this helps!

[Comment](#)

(4 votes) ▲ Upvote ▼ Downvote ⚡ Flag more ▾



Izzah Nabilah 6 years ago

Can i know what the difference between the $(\sum(x-\mu)^2)/N$ formula and $[(\sum x^2) - (\sum x)^2]/N$ formula. How can i know which one im suppose to use ?

[Answer](#) • [Comment](#)

(5 votes) ▲ Upvote ▼ Downvote ⚡ Flag more ▾



021490 3 years ago

How do I find the standard deviation if I am only given the sample size and the sample mean?

[Answer](#) • [Comment](#)

(4 votes) ▲ Upvote ▼ Downvote ⚡ Flag more ▾



RyanYang14 9 months ago

I don't think you can since there's not enough information given

[Comment](#)

(1 vote) ▲ Upvote ▼ Downvote ⚡ Flag more ▾



Evelyn Lutz 5 years ago

is The standard deviation for a sample is most likely larger than the standard deviation of the population?

[Answer](#) • [2 comments](#)

(3 votes) ▲ Upvote ▼ Downvote ⚡ Flag more ▾



Bryanna McGlinchey 4 years ago

For the population standard deviation equation, instead of doing mu for the mean, I learned the bar x for the mean is that the same thing basically? If so, then why use mu for population and bar x for sample?

[Answer](#) • [1 comment](#)

(3 votes) ▲ Upvote ▼ Downvote ⚡ Flag more ▾

Show more comments

Up next: Quiz 3