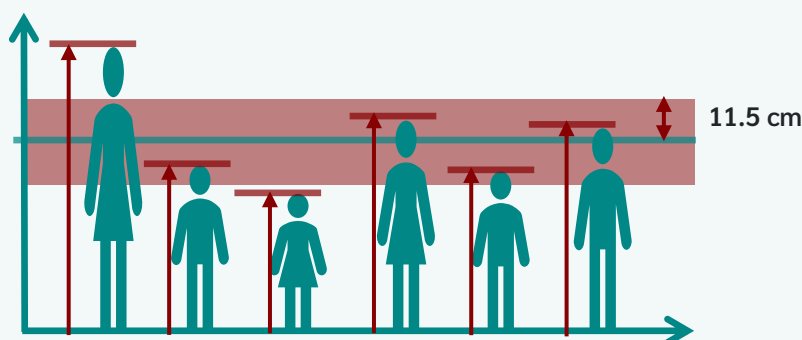


# STANDARD DEVIATION

## Playbook

Theory & Example



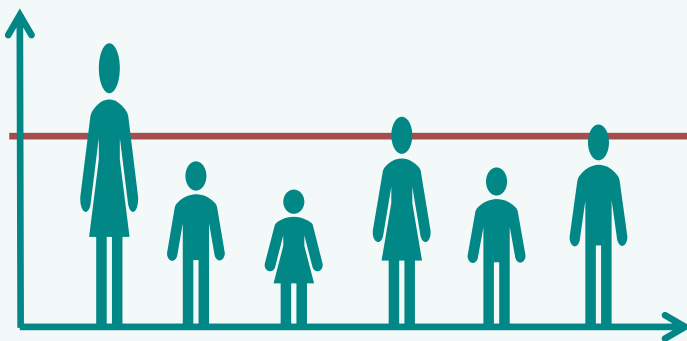


# What is the **standard deviation**?

The **standard deviation** is a **measure** that indicates how much data **scatter around the mean**!

## Example:

Let's say we **measured** the **height** of a small group of people.

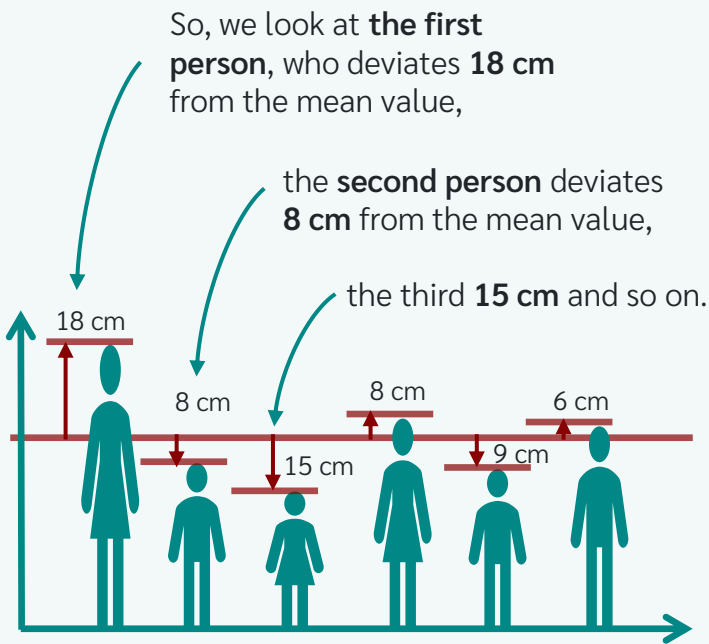


So first we need to **calculate the mean**.

We can get the **mean** simply by **summing the heights** of all individuals and **dividing it by the number of individuals**.

Let's say we get a mean value of **155 cm**.

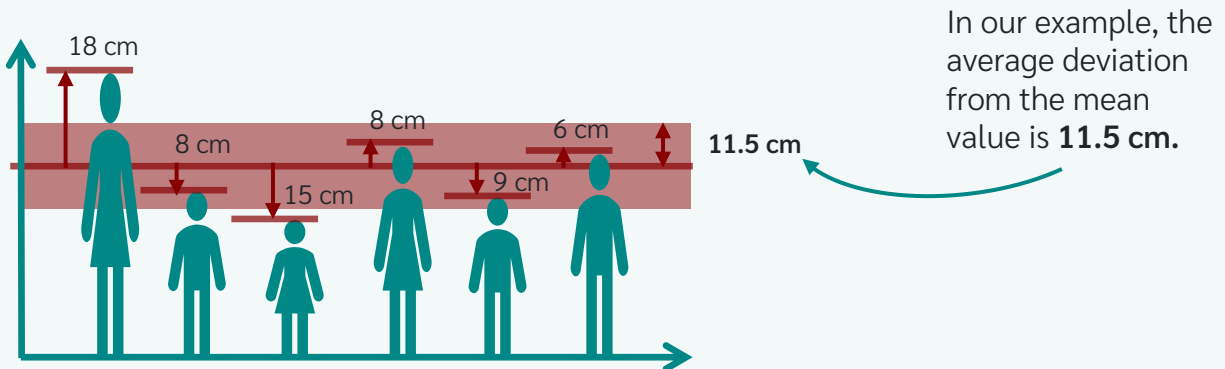
Now we want to know how **much each person** deviates from the **mean**.



! Simply said: People that are **very tall** or **very small** deviate more from the **mean**.

But we are **not interested** in the **deviation** of **each individual person** from the **mean value**, we want to know how much the **persons deviate** from the **mean value on average**.

➡ This is what the **standard deviation** tells us.





# How do we calculate the **standard deviation**?

To calculate the **standard deviation**, we can use this **equation**:

$\sigma$  is the standard deviation

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

$n$  is the number of persons

$x_i$  is the size of each person

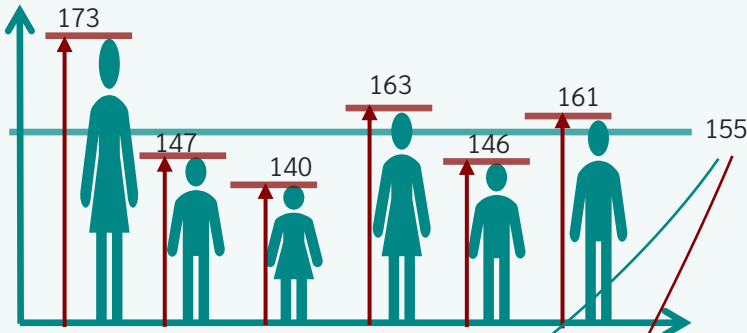
$\bar{x}$  is the mean value of all persons

So, the **standard deviation** is

the **root** of the **sum** of **squared deviations**

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

divided by the **number of values**.



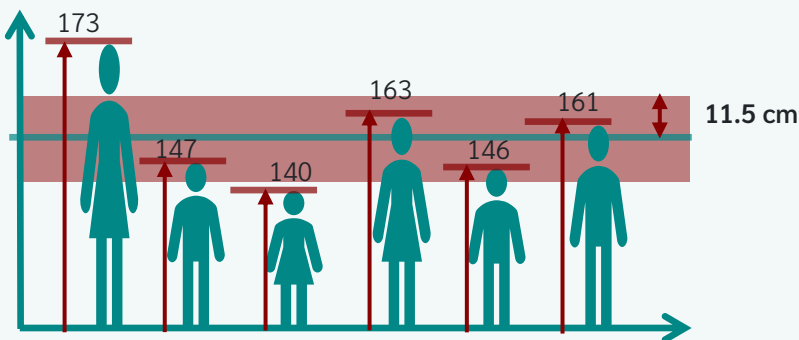
We **calculate** the size of the **first person** minus the mean and square that,

the size of the **second person** minus the **mean** and then **square** that

and so on until we arrive at the last person.

$$= \sqrt{\frac{(173 - 155)^2 + (147 - 155)^2 + \dots + (161 - 155)^2}{6}}$$

Then we **divide** this number by the **number of people**, so **6**.



The result is a **standard deviation of 11.5 cm**.

$$= \sqrt{\frac{(173 - 155)^2 + (147 - 155)^2 + \dots + (161 - 155)^2}{6}} = 11.5$$



For the average deviation, we would actually just **add up all deviations** and **divide** it by the **number of participants**, just like you calculate a mean value, right?

**Absolutely right, but** there are **different mean values**.

In the case of the standard deviation, it is not the **Arithmetic Mean** which is used, but the **Quadratic Mean**.

$$\bar{x}_{AM} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

$$\bar{x}_{QM} = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}}$$



And there are **two** slightly **different equations** for the **standard deviation**.

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

The **difference** is that in the first case we **divide by n**

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

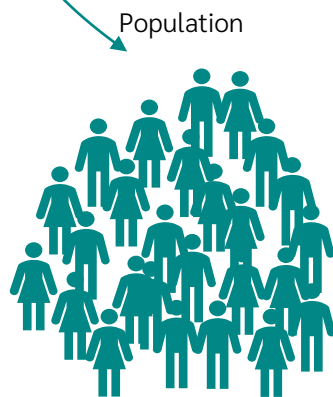
and in the second case **by n-1**.



# Why are there two different equations?

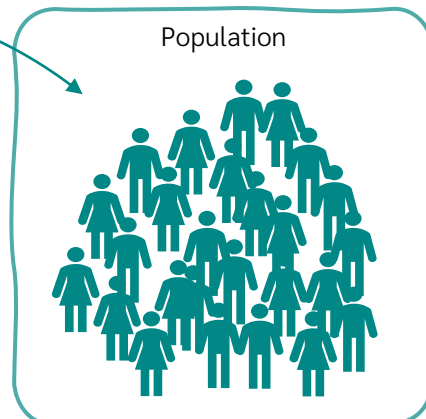


Usually, we want to know the **standard deviation** of the **population**, for example, we want to know the **standard deviation** of the height of **all Austrian professional soccer players**.



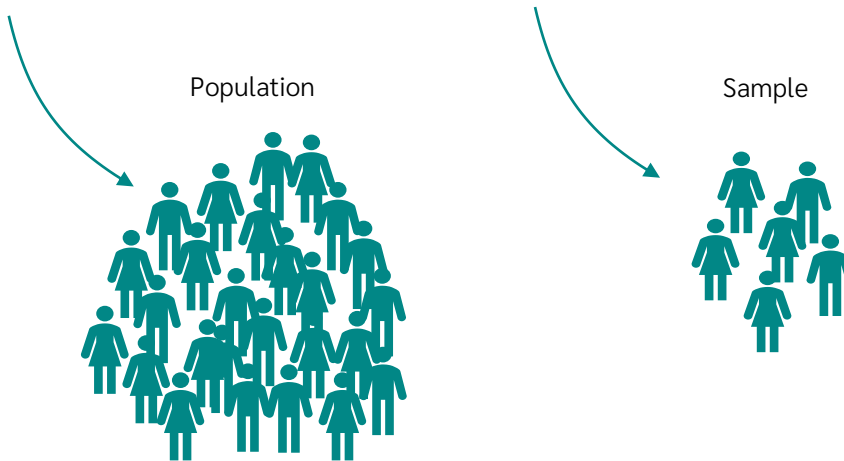
Now if we had the height of **really all** Austrian professional soccer players, we would take this **equation**, with **1** divided by **n**

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$





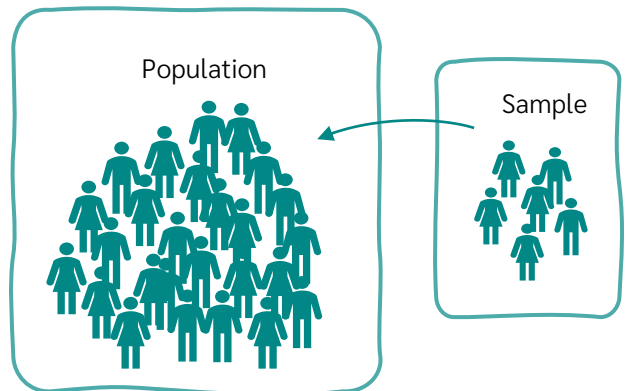
However, it is usually not possible to **survey the entire population**, so we draw a **sample**.



Then we use **the sample to estimate the standard deviation of the population**.

In that case you use this equation, with **n-1**.

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$



**To keep it simple:** If our survey doesn't cover the whole population, we always use this equation! Likewise, if we have conducted a clinical study, then we also use this equation to infer the population.





# What is the **difference** between the **standard deviation** and the **variance**?

As we now know, the **standard deviation** is the **quadratic mean** of the **distance** from the **mean**.

The **variance** now is the **squared standard deviation**.

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

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



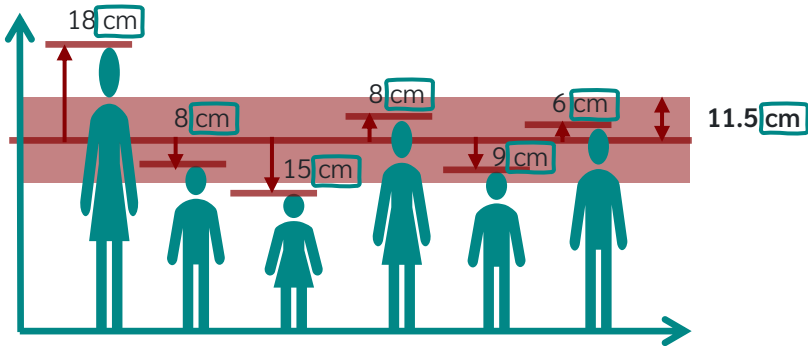
So, we have **one and the same equation!**

The only difference is that in order to calculate the **standard deviation** we take the **root**,

in order to calculate the **variance**, we don't.



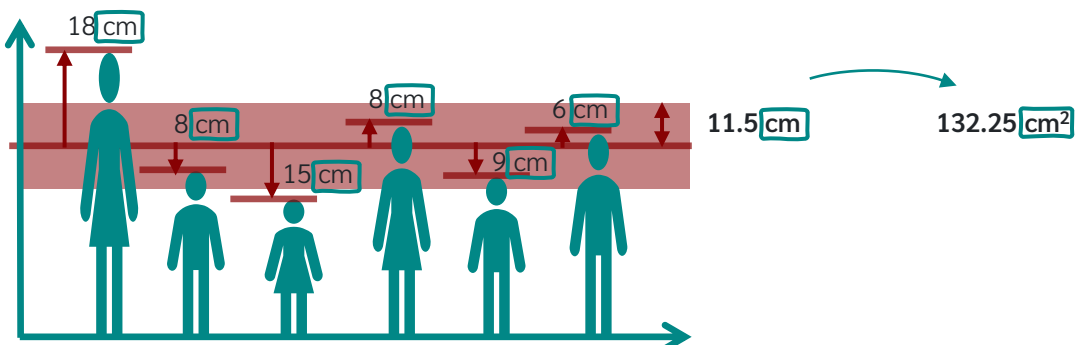
Because the **root** is taken, the **standard deviation** is always in the **same unit as the original data**.



For this reason, it is **advisable** to always **use** the **standard deviation** to describe data, as this makes interpretation easier.

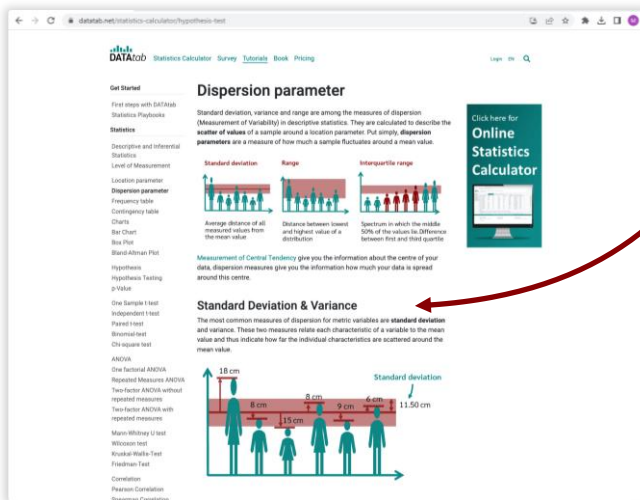
The **variance** is **more difficult to interpret** because the **unit** is the **square of the original unit**.

In our case  $\text{cm}^2$





# How do we calculate the standard deviation with DATAtab?



More information about **standard deviation** and how to calculate it with **DATAtab**

[GO TO DATAtab](#)

## How-to

### Descriptive statistics

Create frequency table and crosstab

Standard Deviation and Variance Calculator

Mean, Median, Modal Calculator

## Calculate standard deviation and variance

You can easily calculate the standard deviation and the variance with DATAtab, just copy your data into the upper table and select the variables you want to evaluate. Make sure that your variables are classified as metric.

### Standard deviation calculator

To calculate the standard deviation, simply click on one or more metric variables. The calculation of the standard deviation is already pre-selected.

12	Male	2,550	48	New York	64	VW	Master
13							
14							

Descriptive Charts t-Test, Chi<sup>2</sup>-Test, ANOVA, ... Correlation Regression Mediation/Moderation PCA Reliability Cluster

Metric Variables:  Salary  Age  Weight  
 Ordinal Variables:  Academic degree  
 Nominal Variables:  Gender  Place  Company

Calculate:  
 Mean  Median  Modal  Sum  Std. Deviation  Variance  
 Minimum  Maximum  Range  Quantile 1  Quantile 2  
 Quantile 3  Skew  Kurtosis  Number of valid values  
 Check for normal distribution

Copy Word Copy Excel

Salary	
Mean	2,269.17
Std. Deviation	658.61
Minimum	1,200
Maximum	3,000

Just try it with the data already inserted, the standard deviation can be calculated quite easily.

More about

# STANDARD DEVIATION

on our website [datatab.net](https://datatab.net)

Feel free  
to share!