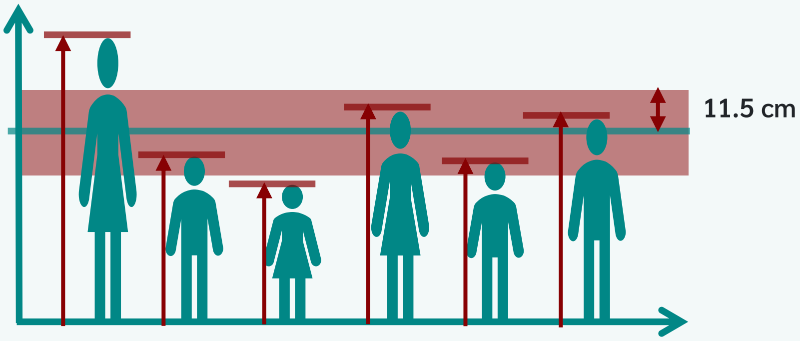
STANDARD DEVIATION

Playbook

Theory & Example



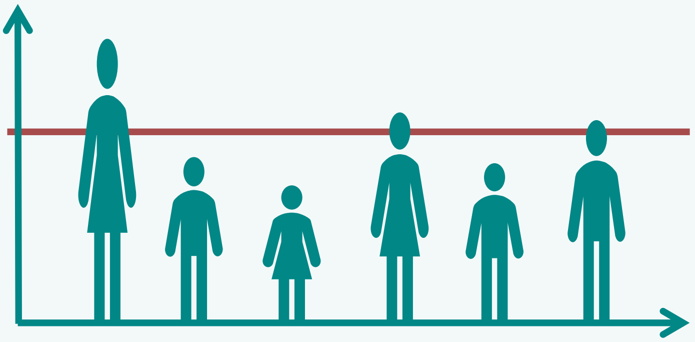
Author: Dr. Mathias Jesussek

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**.

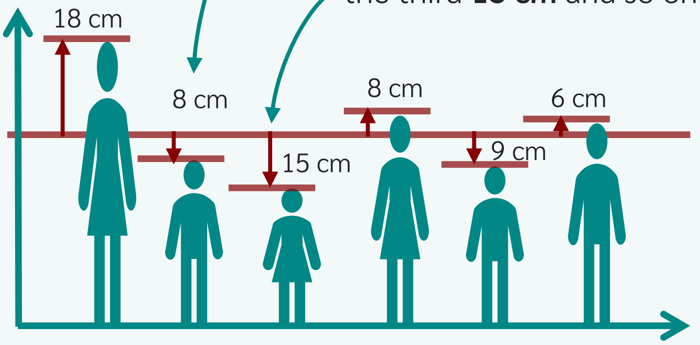
Let's say we get a mean value of **155 cm.**

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

So, we look at **the first person**, who deviates **18 cm** from the mean value,

the **second person** deviates **8 cm** from the mean value,

the third **15 cm** and so on.

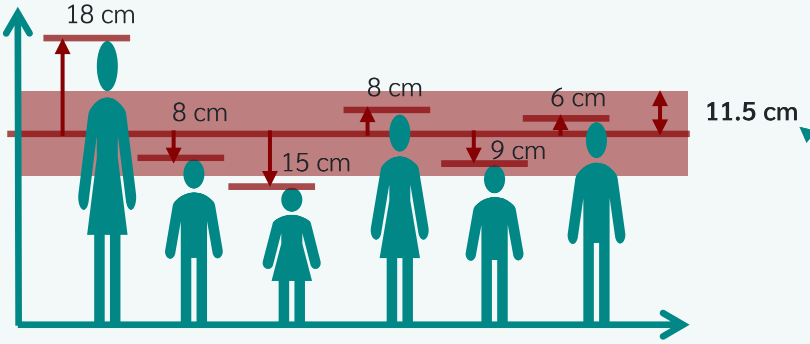


!

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.



In our example, the average deviation from the mean value is **11.5 cm.**

DATAtab

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

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

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

σ is the standard deviation

n is the number of persons

x_i is the size of each person

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

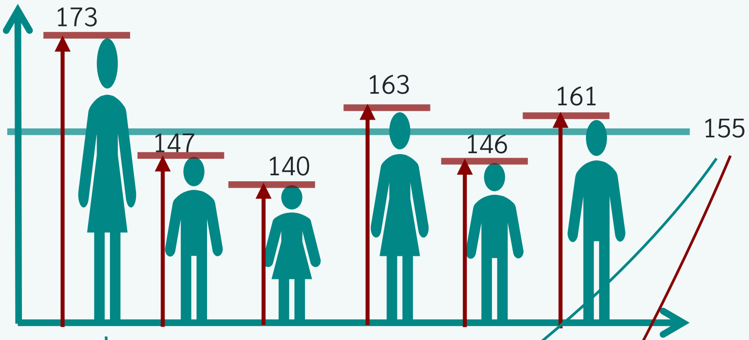
So, the **standard deviation** is

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**.

DATAtab



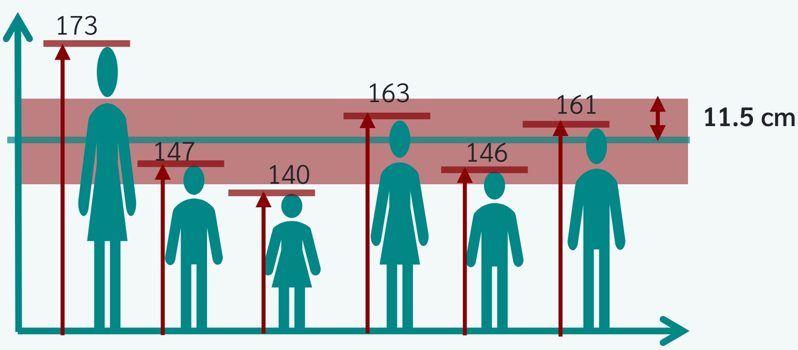
We **calculate** the size of the **first person** minus the mean and square that,

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

and so on until we arrive at the last person.

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

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



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

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

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.**

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

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

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

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

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

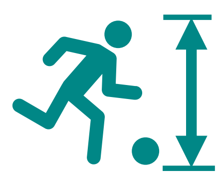
The **difference** is that in the first case we **divide by n**

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

Why are there two **different equations?**

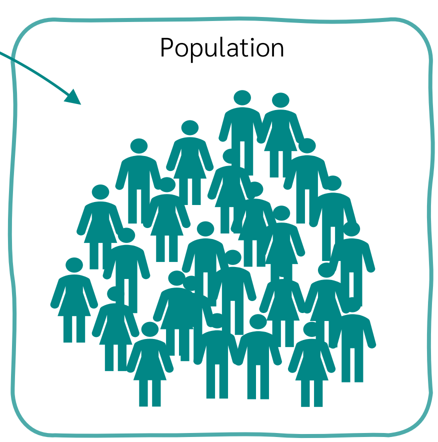
! Usually, we want to know the **standard deviation** of the **population**, for example, we want to know the **standard deviation** of the height **of all Austrian professional soccer players**.

Population

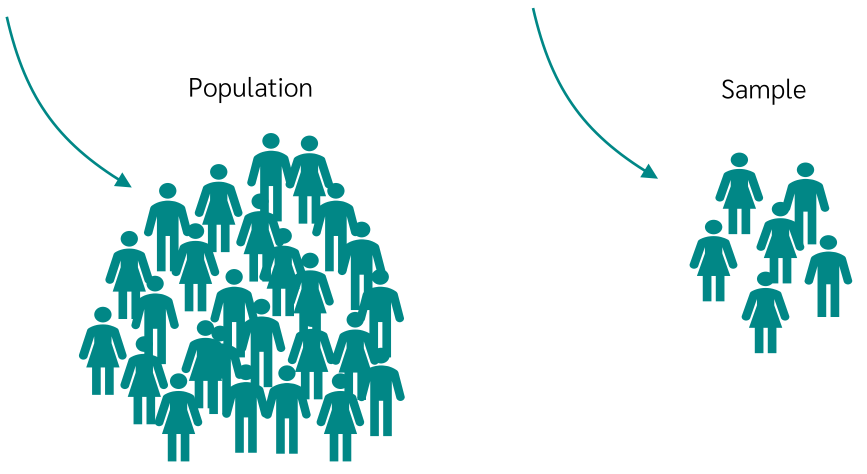


Now if we had the height of **really all** Austrian professional soccer players, we would take this **equation**, with **1 divided by n**

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



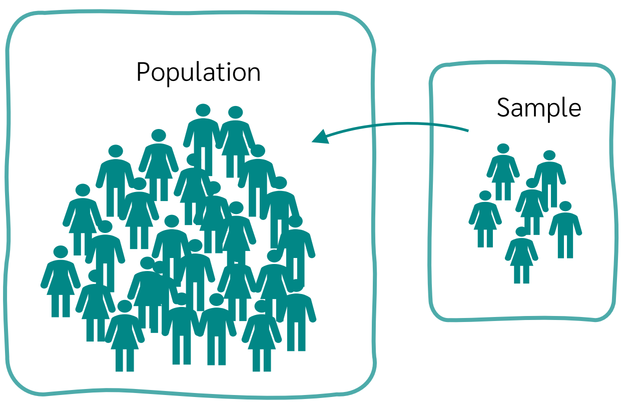
However, it is usually not possible **to survey the entire population**, so we draw a **sample.**



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

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

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



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**.

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

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

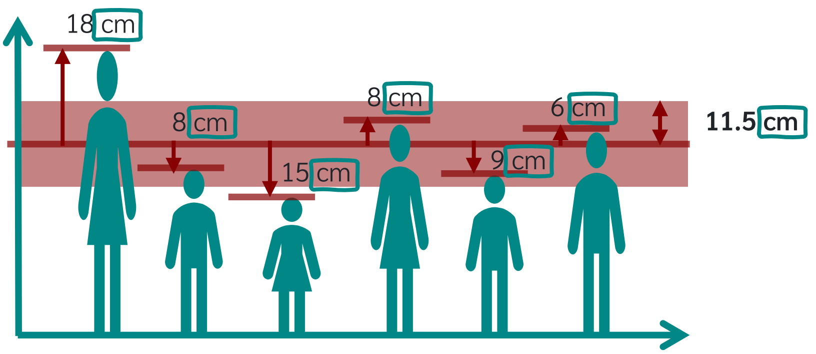
!

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.

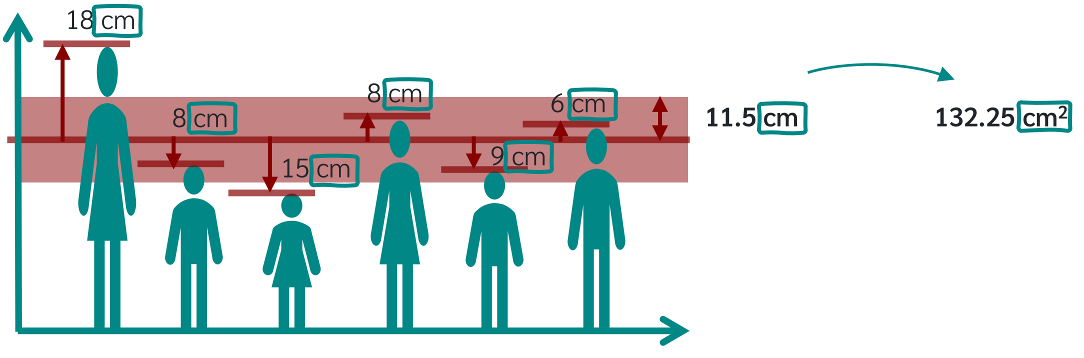
Because the **root** is taken, the **standard deviation** is always in the **same unit as the original data.**



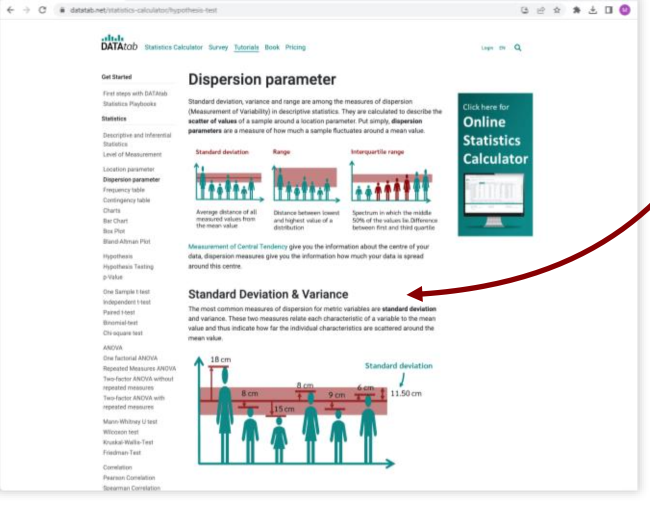
For this reason, it is **advisable** to always **use** the **standard deviation** to describe data, as this makes interpretation easier.

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

In our case **cm²**

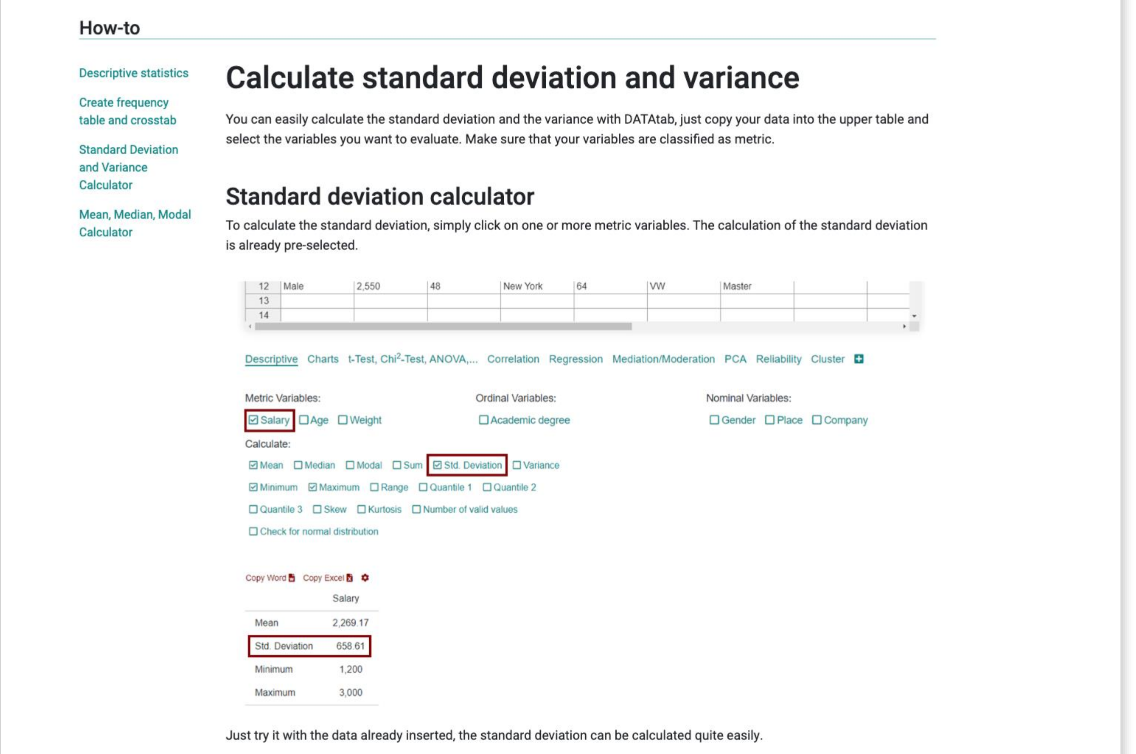


How do we calculate the **standard deviation** with **DATA*tab***?



More information about **standard daviation** and how to calculate it with **DATA*tab***

[GO TO DATA*tab*]



DATAtab

More about

STANDARD DEVIATION

on our website **datatab.net**

Feel free to share!