

What is the Standard Deviation in Statistics?

Authored by
stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the Standard Deviation in Statistics?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=159051>

The Standard Deviation is a commonly used measure of variability or spread in a set of data in statistics. It measures how much the data values deviate from the mean, which is the average of all the values. A lower standard deviation indicates that the data points are closer to the mean, while a higher standard deviation indicates that the data points are more spread out. It is an important tool for understanding the distribution of data and identifying any outliers or unusual values. The standard deviation is calculated by taking the square root of the average of the squared differences between each data point and the mean. It is often used in conjunction with the mean to provide a more comprehensive understanding of a data set.

Statistics - Standard Deviation

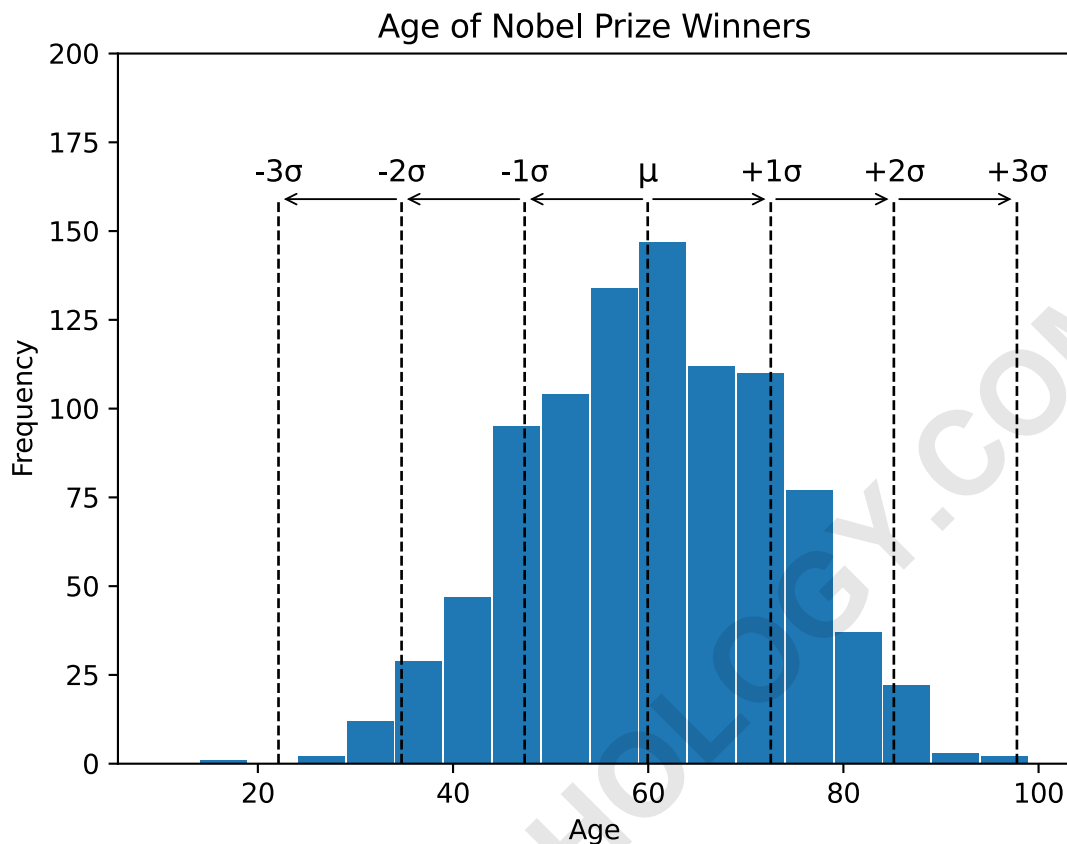
Standard deviation is the most commonly used measure of variation, which describes how spread out the data is.

Standard Deviation

Standard deviation (σ) measures how far a 'typical' observation is from the average of the data (μ).

Standard deviation is important for many statistical methods.

Here is a histogram of the age of all 934 Nobel Prize winners up to the year 2020, showing **standard deviations**:



Each dotted line in the histogram shows a shift of one extra standard deviation.

If the data is **normally distributed**:

Roughly 68.3% of the data is within 1 standard deviation of the average (from $\mu-1\sigma$ to $\mu+1\sigma$)

Roughly 95.5% of the data is within 2 standard deviations of the average (from $\mu-2\sigma$ to $\mu+2\sigma$)

Roughly 99.7% of the data is within 3 standard deviations of the average (from $\mu-3\sigma$ to $\mu+3\sigma$)

Note: A **normal** distribution has a "bell" shape and spreads out equally on both sides.

Calculating the Standard Deviation

You can calculate the standard deviation for both the **population** and the **sample**.

The formulas are **almost** the same and uses different symbols to refer to the standard deviation ((σ)) and **sample** standard deviation ((s)).

Calculating the **standard deviation** ((σ)) is done with this formula:

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

Calculating the **sample standard deviation** (s) is done with this formula:

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

n is the total number of observations.

\sum is the symbol for adding together a list of numbers.

(x_i) is the list of values in the data: (x_1, x_2, x_3, \dots)

μ is the population mean and (\bar{x}) is the sample mean (average value).

$(x_i - \mu)$ and $(x_i - \bar{x})$ are the differences between the values of the observations (x_i) and the mean.

Each difference is squared and added together.

Then the sum is divided by n or $(n - 1)$ and then we find the square root.

Using these 4 example values for calculating the **population standard deviation**:

4, 11, 7, 14

We must first find the mean:

$$\mu = \frac{\sum x_i}{n} = \frac{4 + 11 + 7 + 14}{4} = \frac{36}{4} = \underline{9}$$

Then we find the difference between each value and the mean $(x_i - \mu)$:

$$(4 - 9 ; := -5) (11 - 9 = 2) (7 - 9 ; := -2) (14 - 9 = 5)$$

Each value is then squared, or multiplied with itself $(x_i - \mu)^2$:

$$((-5)^2 = (-5)(-5) = 25) (2^2 ; ; ; ; ; = 2*2 ; ; ; ; ; = 4) ((-2)^2 = (-2)(-2) = 4) (5^2 ; ; ; ; ; = 5*5 ; ; ; ; ; = 25)$$

All of the squared differences are then added together $(\sum (x_i - \mu)^2)$:

$$(25 + 4 + 4 + 25 = 58)$$

Then the sum is divided by the total number of observations, n :

$$\frac{58}{4} = 14.5$$

Finally, we take the square root of this number:

($\sqrt{14.5}$ approx 3.81)

So, the standard deviation of the example values is roughly: (3.81)

Calculating the Standard Deviation with Programming

The standard deviation can easily be calculated with many programming languages.

Using software and programming to calculate statistics is more common for bigger sets of data, as calculating by hand becomes difficult.

Population Standard Deviation

Example

With Python use the NumPy library `std()` method to find the standard deviation of the values 4,11,7,14:

```
import numpy
values =
x = numpy.std(values)
print(x)
```

Example

Use an R formula to find the standard deviation of the values 4,11,7,14:

```
values <- c(4,7,11,14)
sqrt(mean((values-mean(values))^2))
```

Sample Standard Deviation

Example

With Python use the NumPy library `std()` method to find the **sample** standard deviation of the values 4,11,7,14:

```
import numpy
```

```
values =
```

```
x = numpy.std(values, ddof=1)
```

```
print(x)
```

Example

Use the R `sd()` function to find the **sample** standard deviation of the values 4,11,7,14:

```
values <- c(4,7,11,14)
```

```
sd(values)
```

Statistics Symbol Reference

| Symbol | Description |
|---------------|---|
| (σ) | Population standard deviation. Pronounced 'sigma'. |
| (s) | Sample standard deviation. |
| (μ) | The population mean. Pronounced 'mu'. |
| (\bar{x}) | The sample mean. Pronounced 'x-bar'. |
| (\sum) | The summation operator, 'capital sigma'. |
| (x) | The variable 'x' we are calculating the average for. |
| (i) | The index 'i' of the variable 'x'. This identifies each observation for a variable. |
| (n) | The number of observations. |

★+1 W3schools PathfinderTrack your progress - it's free!

Log in

Sign Up