Scribbr

Pearson Correlation Coefficient (r) | Guide & Examples

Published on May 13, 2022 by Shaun Turney. Revised on February 10, 2024.

The **Pearson correlation coefficient (*r*)** is the most common way of measuring a linear correlation. It is a number between –1 and 1 that measures the strength and direction of the relationship between two variables.

Pearson correlation coefficient (*r*)	Correlation type	Interpretation	Example
Between 0 and 1	Positive correlation	When one variable changes, the other variable changes in the **same direction**.	Baby length & weight: The longer the baby, the heavier their weight.
0	No correlation	There is **no relationship** between the variables.	Car price & width of windshield wipers: The price of a car is not related to the width of its windshield wipers.

Table of contents

Pearson coefficient (*r*)	**Correlation**	**Interpretation**	**Example**
Between 0 and −1	Negative correlation	When one variable changes, the other variable changes in the **opposite direction**.	Elevation & air pressure: The higher the elevation, the lower the air pressure.

What is the Pearson correlation coefficient?

The Pearson correlation coefficient (*r*) is the most widely used correlation coefficient and is known by many names:

- Pearson's *r*
- Bivariate correlation
- Pearson product-moment correlation coefficient (PPMCC)
- The correlation coefficient

The Pearson correlation coefficient is a descriptive statistic.

Table of contents

Specifically, it describes the strength and direction of the linear relationship between two quantitative variables.

Although interpretations of the relationship strength (also known as effect size) vary between disciplines, the table below gives general rules of thumb:

Pearson correlation coefficient (*r*) value	Strength	Direction
Greater than .5	Strong	Positive
Between .3 and .5	Moderate	Positive
Between 0 and .3	Weak	Positive
0	None	None
Between 0 and –.3	Weak	Negative
Between –.3 and –.5	Moderate	Negative
Less than –.5	Strong	Negative

The Pearson correlation coefficient is also an inferential statistic, meaning that it can be used to test statistical hypotheses. Specifically, we can test whether there is a significant relationship between two variables.

Table of contents

Here's why students love Scribbr's proofreading services

Excellent

Rated 4.6 / 5 based on 13,101 reviews on

Trustpilot

Invited

Isabella, 2 hours ago

Having citations copied and pasted...

Read more

Invited

Discover proofreading & editing

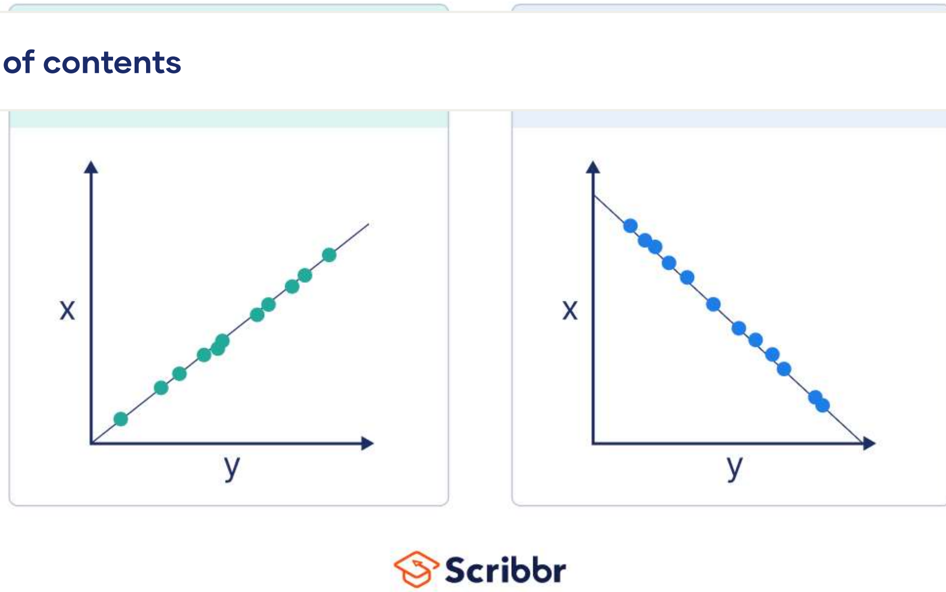
Visualizing the Pearson correlation coefficient

Another way to think of the Pearson correlation coefficient (*r*) is as a measure of how close the observations are to a line of best fit.

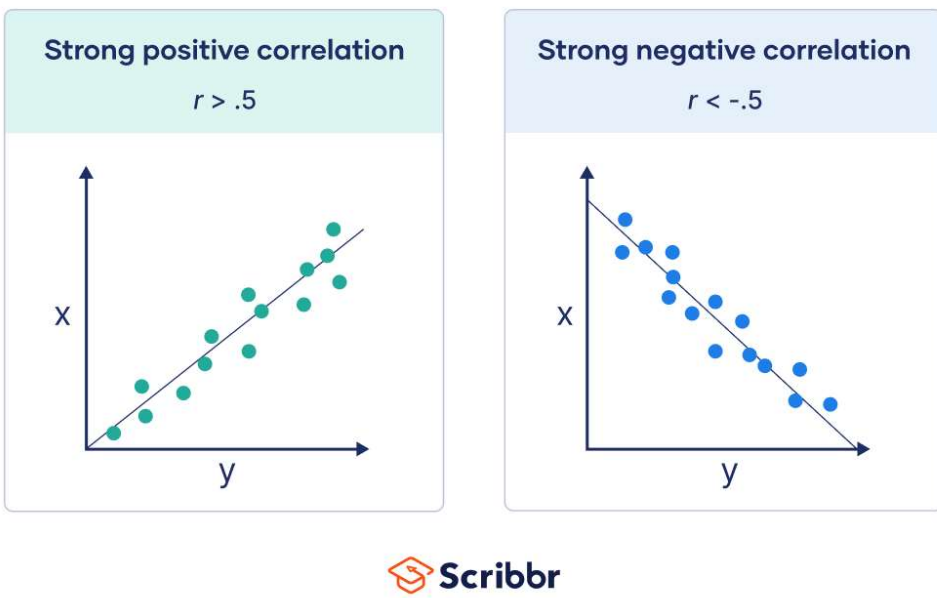
The Pearson correlation coefficient also tells you whether the slope of the line of best fit is negative or positive. When the slope is negative, *r* is negative. When the slope is positive, *r* is positive.

When *r* is 1 or −1, all the points fall exactly on the line of best fit:

Table of contents

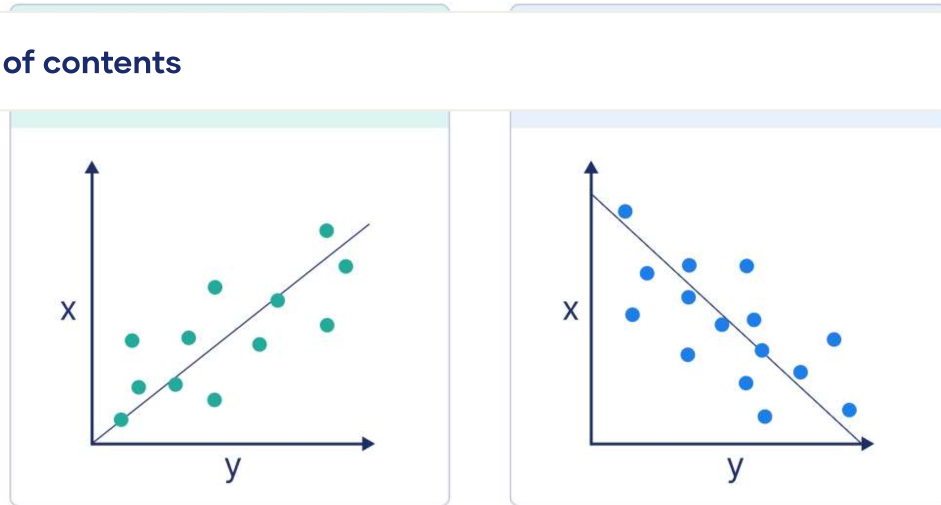


When *r* is greater than .5 or less than –.5, the points are close to the line of best fit:

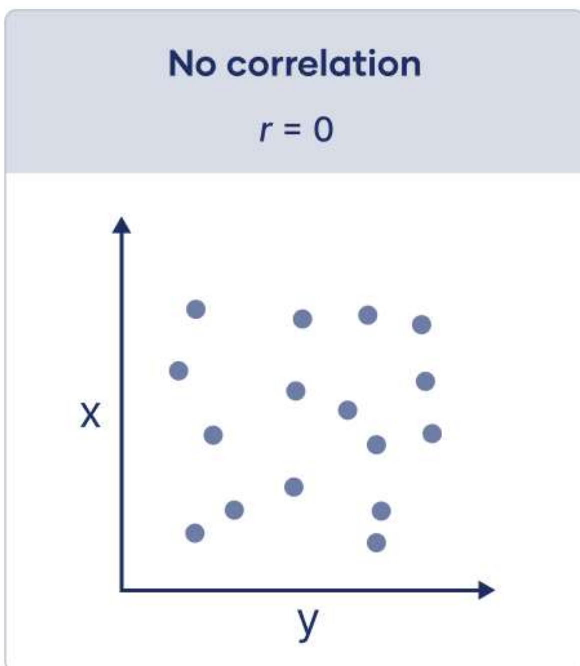


When *r* is between 0 and .3 or between 0 and –.3, the points are far from the line of best fit:

Table of contents



When *r* is 0, a line of best fit is not helpful in describing the relationship between the variables:



When to use the Pearson correlation

Table of contents

The Pearson correlation coefficient (*r*) is one of several correlation coefficients that you need to choose between when you want to measure a correlation. The Pearson correlation coefficient is a good choice when **all** of the following are true:

- **Both variables are quantitative:** You will need to use a different method if either of the variables is qualitative.
- **The variables are normally distributed:** You can create a histogram of each variable to verify whether the distributions are approximately normal. It's not a problem if the variables are a little non-normal.
- **The data have no outliers:** Outliers are observations that don't follow the same patterns as the rest of the data. A scatterplot is one way to check for outliers—look for points that are far away from the others.
- **The relationship is linear:** "Linear" means that the relationship between the two variables can be described reasonably well by a straight line. You can use a scatterplot to check whether the relationship between two variables is linear.

Pearson vs. Spearman's rank correlation coefficients

Spearman's rank correlation coefficient is another widely used correlation coefficient. It's a better choice than the Pearson correlation coefficient when **one or more** of the following is true:

- The variables are ordinal.
- The variables aren't normally distributed.
- The data includes outliers.
- The relationship between the variables is non-linear **and** monotonic.

Calculating the Pearson correlation

Table of contents

Below is a formula for calculating the Pearson correlation coefficient (*r*):

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

The formula is easy to use when you follow the step-by-step guide below. You can also use software such as R or Excel to calculate the Pearson correlation coefficient for you.

Example: Dataset

Imagine that you're studying the relationship between newborns' weight and length. You have the weights and lengths of the 10 babies born last month at your local hospital. After you convert the imperial measurements to metric, you enter the data in a table:

Weight (kg)	Length (cm)
3.63	53.1
3.02	49.7
3.82	48.4
3.42	54.2
3.59	54.9
2.87	43.7
3.03	47.2
3.46	45.2
3.36	54.4

Table of contents

Step 1: Calculate the sums of *x* and *y*

Start by renaming the variables to "*x*" and "*y*." It doesn't matter which variable is called *x* and which is called *y*—the formula will give the same answer either way.

Next, add up the values of *x* and *y*. (In the formula, this step is indicated by the Σ symbol, which means "take the sum of".)

Example: Calculating the sums of x and y

Weight = x

Length = y

$\Sigma x = 3.63 + 3.02 + 3.82 + 3.42 + 3.59 + 2.87 + 3.03 + 3.46 + 3.36 + 3.30$

$\Sigma x = 33.5$

$\Sigma y = 53.1 + 49.7 + 48.4 + 54.2 + 54.9 + 43.7 + 47.2 + 45.2 + 54.4 + 50.4$

$\Sigma y = 501.2$

Step 2: Calculate x^2 and y^2 and their sums

Create two new columns that contain the squares of *x* and *y*. Take the sums of the new columns.

Example: Calculating x^2 and y^2 and their sums

Table of contents

x	y	x^2	y^2
3.02	49.7	9.12	2 470.1
3.82	48.4	14.59	2 342.6
3.42	54.2	11.7	2 937.6
3.59	54.9	12.89	3 014
2.87	43.7	8.24	1 909.7
3.03	47.2	9.18	2 227.8
3.46	45.2	11.97	2 043
3.36	54.4	11.29	2 959.4
3.3	50.4	10.89	2 540.2

Σx^2 = 13.18 + 9.12 + 14.59 + 11.70 + 12.89 + 8.24 + 9.18 + 11.97 + 11.29 + 10.89

Σx^2 = 113.05

Σy^2 = 2 819.6 + 2 470.1 + 2 342.6 + 2 937.6 + 3 014.0 + 1 909.7 + 2 227.8 + 2 043.0 + 2 959.4 + 2 540.2

Σy^2 = 25 264

Step 3: Calculate the cross product and its sum

In a final column, multiply together x and y (this is called the cross product). Take the sum of the new column.

Example: Calculating the cross product and its sum

Table of contents

3.02	49.7	9.12	2 470.1	150.1
3.82	48.4	14.59	2 342.6	184.9
3.42	54.2	11.7	2 937.6	185.4
3.59	54.9	12.89	3 014	197.1
2.87	43.7	8.24	1 909.7	125.4
3.03	47.2	9.18	2 227.8	143
3.46	45.2	11.97	2 043	156.4
3.36	54.4	11.29	2 959.4	182.8
3.3	50.4	10.89	2 540.2	166.3

Σ*xy* = 192.8 + 150.1 + 184.9 + 185.4 + 197.1 + 125.4 + 143.0 + 156.4 + 182.8 + 166.3

Σ*xy* = 1 684.2

Step 4: Calculate *r*

Use the formula and the numbers you calculated in the previous steps to find *r*.

Example: Calculating *r*

$n = 10$

$\sum x = 33.5$

$\sum y = 501.2$

$\sum x^2 = 113.05$

$$\sum y^2 = 25\,264$$

Table of contents

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

$$r = \frac{10\sum 1\,684.2 - (33.5)(501.2)}{\sqrt{[(10)(113.05) - (33.5)^2][(10)(25\,264) - (501.2)^2]}}$$

$$r = \frac{16\,842 - 16\,790.2)}{\sqrt{[1\,130.5 - 1\,122.25][252\,640 - 251\,201.4]}}$$

$$r = \frac{51.8}{\sqrt{11\,868.45}}$$

$$r = 0.47$$

Prevent plagiarism. Run a free check.

Try for free

Testing for the significance of the Pearson correlation coefficient

The Pearson correlation coefficient can also be used to test whether the relationship between two variables is significant.

The Pearson correlation of the sample is *r*. It is an estimate of rho (*ρ*), the Pearson correlation of the population. Knowing *r* and *n* (the sample size), we can infer whether *ρ* is significantly different from 0.

- Null hypothesis (H_0): $\rho = 0$

To test the hypotheses, you can either use software like R or Stata or you can follow the three steps below.

Step 1: Calculate the *t* value

Calculate the *t* value (a test statistic) using this formula:

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

Example: Calculating the *t* value

The weight and length of 10 newborns has a Pearson correlation coefficient of .47. Since we know that n = 10 and r = .47, we can calculate the *t* value:

$$t = \frac{0.47}{\sqrt{\frac{1-(0.47)^2}{10-2}}}$$

$$t = \frac{0.47}{\sqrt{\frac{1-0.22}{8}}}$$

$$t = \frac{0.47}{\sqrt{0.0975}}$$

$$t = 1.506$$

Step 2: Find the critical value of *t*

You can find the critical value of *t* (t^*) in a *t* table. To use the table, you need to know three things:

- **The degrees of freedom (*df*):** For Pearson correlation tests, the formula is $df = n - 2$.
- **Significance level (α):** By convention, the significance level is usually .05.

- **One-tailed or two-tailed:** Most often, two-tailed is an

Table of contents

Example: Finding the critical value of *t*

For a two-tailed test of significance at α = .05 and *df* = 8, the critical value of *t* (*t**) is 2.306.

Step 3: Compare the *t* value to the critical value

Determine if the absolute *t* value is greater than the critical value of *t*. "Absolute" means that if the *t* value is negative you should ignore the minus sign.

Example: Comparing the *t* value to the critical value of *t* (*t)**

t = 1.506

t^* = 2.306

The *t* value is less than the critical value of *t*.

Step 4: Decide whether to reject the null hypothesis

- If the *t* value is **greater** than the critical value, then the relationship is statistically significant ($p < \alpha$). The data allows you to reject the null hypothesis and provides support for the alternative hypothesis.
- If the *t* value is **less** than the critical value, then the relationship is not statistically significant ($p > \alpha$). The data doesn't allow you to reject the null hypothesis and doesn't provide support for the alternative hypothesis.

Example: Deciding whether to reject the null hypothesis

For the correlation between weight and height in a sample of 10 newborns, the *t* value is less than the critical value of *t*.

Therefore, we **don't reject** the null hypothesis that the

Table of contents

There is no significant relationship between weight and height ($p > .05$).

(Note that a sample size of 10 is very small. It's possible that you would find a significant relationship if you increased the sample size.)

Reporting the Pearson correlation coefficient

If you decide to include a Pearson correlation (*r*) in your paper or thesis, you should report it in your results section. You can follow these rules if you want to report statistics in APA Style:

- You don't need to provide a reference or formula since the Pearson correlation coefficient is a commonly used statistic.
- You should italicize *r* when reporting its value.
- You shouldn't include a leading zero (a zero before the decimal point) since the Pearson correlation coefficient can't be greater than one or less than negative one.
- You should provide two significant digits after the decimal point.

When Pearson's correlation coefficient is used as an inferential statistic (to test whether the relationship is significant), *r* is reported alongside its degrees of freedom and *p* value. The degrees of freedom are reported in parentheses beside *r*.

Example: Reporting the Pearson correlation coefficient in APA Style

Newborns' weight and length were moderately correlated, although the relationship was not statistically significant,