

Pearson Product-Moment Correlation

Explorable.com, Lyndsay T Wilson 103.1K reads

The Pearson Product-Moment Correlation is one of the measures of correlation which quantifies the strength as well as the direction of such relationship. It is usually denoted by the Greek letter ρ (rho).

In the study of relationships, two variables are said to be correlated if change in one variable is accompanied by change in the other - either in the same or opposite direction.



The banner features the Explorable logo at the top center. Below it, the text "Quiz Time!" is written in a cursive font. There are three quiz cards displayed in a row, each with a different image and title. The first card shows a pair of red roller skates on a wooden deck, titled "Quiz: Psychology 101 Part 2". The second card shows a fan of colorful pencils, also titled "Quiz: Psychology 101 Part 2". The third card shows a Ferris wheel at sunset, titled "Quiz: Flags in Europe". To the right of the cards is a link that says "See all quizzes =>".

When to Use This Statistic

This coefficient is used only when two conditions are satisfied:

1. the variables are in the interval or ratio scale of measurement ^[1]
2. a linear relationship ^[2] between the variables is suspected

Positive and Negative Correlation

The coefficient (ρ) is calculated as the ratio of covariance between the variables to the product of their standard deviations ^[3]. This formulation is very useful for two key reasons.

First, it tells us the **direction** of relationship. Once the coefficient is computed, $r > 0$ will indicate a positive relationship, $r < 0$ will indicate negative relationship while $r = 0$ indicates non existence of any relationship.

Second, it ensures (mathematically) that the numerical value of r ranges from -1.0 to +1.0. This enables us to get an idea of the **strength** of relationship - or rather the strength of linear relationship [2] between the variables. The closer the coefficient is to +1.0 or -1.0, the greater the strength of the linear relationship.

As a rule of thumb, the following guidelines are often useful (though many experts would disagree somewhat on the choice of boundaries).

Range of r

Value of r	Strength of relationship
-1.0 to -0.5 or 1.0 to 0.5	Strong
-0.5 to -0.3 or 0.3 to 0.5	Moderate
-0.3 to -0.1 or 0.1 to 0.3	Weak
-0.1 to 0.1	None or very weak

Properties of r

This measure of correlation has interesting properties:

1. It is independent of any units of measurement. For example, the r value between the highest day temperature (in Centigrade) and rainfall per day (in mm) is not expressed either in terms of centigrade or mm. This is because it is not expressing a quantity, but a relationship between quantities.
2. It is symmetric. This means that r between X and Y is exactly the same as r between Y and X.
3. Pearson's correlation coefficient is independent of change in origin and scale. This means that r between temperature (in Centigrade) and rainfall (in mm) would numerically be equal to r between temperature (in Fahrenheit) and rainfall (in cm).
- 4.

If the variables are truly independent of each other, then one would obtain $\rho = 0$. However, the converse is not true. In other words $\rho = 0$ does not imply that the variables are independent - it only indicates the non existence of a non-linear relationship [4]. You may also arrive at this result in error if your variables are not in interval or ratio scale of measurement.

Caveats and Warnings

While ρ is a powerful tool, it is a much abused one and hence has to be handled carefully.

1. People often forget or gloss over the fact that ρ is a measure of linear relationship. Consequently a small value of ρ is often interpreted to mean non existence of relationship when actually it only indicates non existence of a linear relationship or at best a very weak linear relationship.

Under such circumstances it is possible that a (possibly strong!) non linear relationship exists.

It's best to construct a scatter diagram to reveal any non linear relationships before firmly concluding the non existence of a relationship. If the scatter diagram points to a non linear relationship, an appropriate transformation can often attain linearity in which case ρ can be recomputed.

2. One has to be careful in interpreting the value of ρ , specifically when it makes no obvious sense to connect the variables in the first place.

For example, one could compute ρ between shoe size and intelligence, or height and income. Irrespective of the value of ρ , such a correlation makes no sense and is hence termed chance or non-sense correlation.

3. As with many related statistics, ρ should not be used to make claims about a cause and effect relationship [5]. Put differently, by examining the value of ρ , we can only conclude that variables X and Y are related.

However the same value of ρ does not tell us if X influences Y or the other way round - a fact that is of key importance in regression analysis [6].

Links

- [1] <https://explorable.com/measurement-scales>
- [2] <https://explorable.com/linear-relationship>
- [3] <https://explorable.com/calculate-standard-deviation>
- [4] <https://explorable.com/non-linear-relationship>
- [5] <https://explorable.com/cause-and-effect>
- [6] <https://explorable.com/linear-regression-analysis>