

14.5: Hypotheses

As we've been learning, Pearson's correlation coefficient, r , tells us about the strength and direction of the linear relationship between two variables. This is the basis of our research hypothesis. We perform a hypothesis test of the "significance of the correlation coefficient" to decide whether the linear relationship in the sample data is strong enough to use to model the relationship in the population; the null hypothesis is that there is no relationship between the variables.

Research Hypothesis

Pearson's r measures the strength and direction of a linear relationship between two quantitative variables; what does this look like as a research hypothesis? As with all inferential statistics, the sample data are used to compute r , the correlation coefficient for the sample. If we had data for the entire population, we could find the population correlation coefficient. But because we have only sample data, we cannot calculate the population correlation coefficient. The sample correlation coefficient, r , is our estimate of the unknown population correlation coefficient.

- The symbol for the population correlation coefficient is ρ , the Greek letter "rho."
- ρ = population correlation coefficient (unknown)
- r = sample correlation coefficient (known; calculated from sample data)

If the test concludes that the correlation coefficient is significantly different from zero, we say that the correlation coefficient is "significant," and can state that there is sufficient evidence to conclude that there is a significant linear relationship between x and y . We can use the regression line to model the linear relationship between x and y in the population. We'll talk about the regression line much more in the next chapter. For now, just know that it's a line that tries to be as close as possible to all of the data points. The closer the dots are to this straight line, the stronger the correlation. The conclusion means that there is a significant linear relationship between x and y . This leads to the research hypothesis of:

*Pearson's Correlation research hypothesis: There is a (positive or negative) **linear** relationship between one quantitative variable and another quantitative variable (name the variables).*

With research hypotheses for mean differences, we specified which mean(s) would be bigger than which other means. With a correlation, we specify whether the two variables vary together (positive correlation: both go up or both down down) or vary in opposite directions (negative correlation: as one variable increases, the other variable decreases). With all hypothesis, it's important to include the names of the variables and what was measured. In Dr. Navarro's example the name of the variable was what was measured (Dani's Sleep, Baby's Sleep, Dani's Grumpiness). This is nice because in the prior examples comparing mean differences, we had to identify the IV, the levels of the IV (the groups), and the DV (what was measured). With correlations, the IV is usually one variable (sometimes called the predictor) and the DV is usually the other variable (sometimes called the outcome). But be careful! These names are starting to suggest that one variable (the predicting IV) causes the changes in the other variable (outcome DV), but we learned earlier (Correlation versus Causation [Part 1](#) and [Part 2](#)) that correlations just show a linear relationship, not whether one variable *caused changes* in the other variable.

Null Hypothesis

If the test concludes that the correlation coefficient is not significantly different from zero (it is close to zero), we say that correlation coefficient is "not significant," and state that there is insufficient evidence to conclude that there is a significant linear relationship between x and y . This conclusion means that there is not a significant linear relationship between x and y . This leads to the null hypothesis of:

*Pearson's Correlation research hypothesis: There is **no linear** relationship between one quantitative variable and another quantitative variable (name the variables).*

We CANNOT use the regression line to model a linear relationship between x and y in the population.

Making the Decision

Null hypothesis significance testing lets us decide whether the value of the population correlation coefficient ρ is "close to zero," meaning that there is no linear relationship between the two variable in the population; when one variable changes, we know

nothing about the other variables changes. We reject this null hypothesis based on the sample correlation coefficient r and the sample size n .

The [Table of Critical Values of \$r\$](#) page (or found through the [Common Critical Value Table](#) page) is used to give you a good idea of whether the computed value of r is significant or not. Compare the absolute value of your calculated r to the appropriate critical value in the table. If calculated r is bigger than the critical value then the correlation coefficient is significant. If r is significant, then you may use the line for prediction.

With null hypothesis significance testing, we either retain the null hypothesis (we don't think that there is a linear relationship between the two variables) or we reject the null hypothesis (we think that there is a linear relationship between the two variables). Again, rejecting the null hypothesis does not mean that the data automatically support the research hypothesis. To support the research hypothesis, the correlation has to be in the direction (positive or negative) that we predicted in our research hypothesis.

✓ Example 14.5.1

What do you think the results sentence will look like?

Solution

The statistical sentence would be: $r(\underline{df}) = \underline{r\text{-calc}}$, $p \underline{\quad} .05$

In which everything underlined is replaced.

Now that we know a little bit about what correlational analyses shows, let's look at the actual formulas to calculate Pearson's r .

Contributors and Attributions

- Barbara Illowsky and Susan Dean (De Anza College) with many other contributing authors. Content produced by OpenStax College is licensed under a Creative Commons Attribution License 4.0 license. Download for free at <http://cnx.org/contents/30189442-699...b91b9de@18.114>.
- [Dr. MO \(Taft College\)](#)

This page titled [14.5: Hypotheses](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [Michelle Oja](#).

- [12.5: Testing the Significance of the Correlation Coefficient](#) by OpenStax is licensed [CC BY 4.0](#). Original source: <https://openstax.org/details/books/introductory-statistics>.