# Calculating A T-Test For A Correlation

It's time to remind ourselves once more of correlations! Correlations are hugely important and useful numerics that are often reported and discussed in everyday life. So far in this class, though, we've only used correlations as descriptive statistics. That is, we've only calculated the correlation between our two sets of sample data, without making any generalizations about populations. In most cases, however, we're most interested in populations. We want to know if a relationship exists between two variables among everyone in the population, not just among the 50 or 100 people in our sample. What we're going to learn in this *Guide*, then, is how to take a correlation and use inferential statistics to make inferences about the correlation in the population. Note that this is very similar to what we've done in the past, except we're now trying to generalize a correlation ( $r_{xy}$ ) to the population, rather than, e.g., a sample mean ( $\bar{x}$ ).

So, how are we going to do this? Well, by using the *t*-test one last time! The good news is that, once you have your correlation coefficient, calculating the *t*-test for that correlation coefficient is simple. Let's take a look at the formula to illustrate:

t-test for Pearson's correlation

$$t_{r_{xy}} = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r^2xy}}$$

Now, let's assume that we're already comfortable with calculating the correlation coefficient ( $r_{xy}$ ). If you aren't, I encourage you to take a look at *Guide 03: Correlations* and/or review the videos/readings related to correlations. For now, though, we'll take the same data we used in our original correlation coefficient calculations. Here's the data:

| x | y |
|---|---|
| 13 | 6 |
| 9 | 1 |
| 5 | 15 |
| 5 | 5 |
| 2 | 14 |
| 9 | 2 |
| 6 | 15 |
| 7 | 4 |
| 8 | 2 |
| 3 | 8 |
| 2 | 7 |
| 1 | 12 |

If you review *Guide 03: Correlations*, you'll recall that we found a correlation between our $x$ and $y$ variables of $r_{xy} = -0.56$, which is a fairly strong correlation. In and of itself, this correlation only tells us that there is a fairly strong, negative correlation between our $x$ and $y$ variables **in this particular sample of 12 participants**. It does not tell us whether we have sufficient evidence to assume that there exists a correlation between our $x$ and $y$ variables **in the population** (i.e., across all people, assuming that "all people" is our population of interest). This is where the $t$-test comes in. Let's calculate the $t$-test statistic ($t_{r_{xy}}$) for this correlation coefficient ($r_{xy}$).

Looking back at the formula displayed on the first page of this *Guide*, you'll notice that we already have everything we need for the $t$-test statistic. In particular, we know that $r = -0.56$. We can easily square this value to get $r^2$, which comes out to $-0.56^2 = 0.3136$. Finally, we know that $n$, our sample size, is 12. Let's plug all of this into our formula:

$$t_{r_{xy}} = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r_{xy}^2}} = \frac{-0.56\sqrt{12-2}}{\sqrt{1-0.3136}} = \frac{-1.77}{0.828} = -2.14$$

And there you have it! The $t$-test statistic for this particular correlation coefficient is $t_{r_{xy}} = -2.14$. If you used statistical software to do this computation, the software would give you a $p$-value of $0.06$. Try to test yourself now—based on an alpha level ($\alpha$) of $0.05$, would we have sufficient evidence to suggest that there is a true negative correlation in the population? Would we *"reject the null hypothesis"* or *"fail to reject the null hypothesis"*? Feel free to message me Canvas if you aren't sure. :)