**Reviewing correlations**

1. ***Notation***

: This statement says that is a *statistic* representing the sample correlation coefficient, which is a *point estimator* of the population correlation coefficient, .

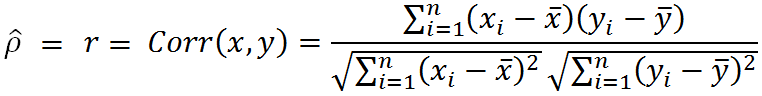
: This statement says that ,the value of sample correlation coefficient in our sample, is a *point estimate* of the population correlation coefficient, .

Take home point: the statistic (capitalized) is an estimator and the value of the statistic (lower-case) is an estimate.

Also, the notation is the same *regardless* of whether we are using Pearson or Spearman correlation coefficients, but we should always specify whether we’re talking about Pearson *r* or Spearman *r*. If no such specification is made, it is typically assumed that we are dealing with Pearson correlations.

1. ***Pearson Correlation Coefficient***

The formula:

****

Values of *r* range between (perfect negative correlation) to (perfect positive correlation). A value of indicates that there’s no (linear) association between and.

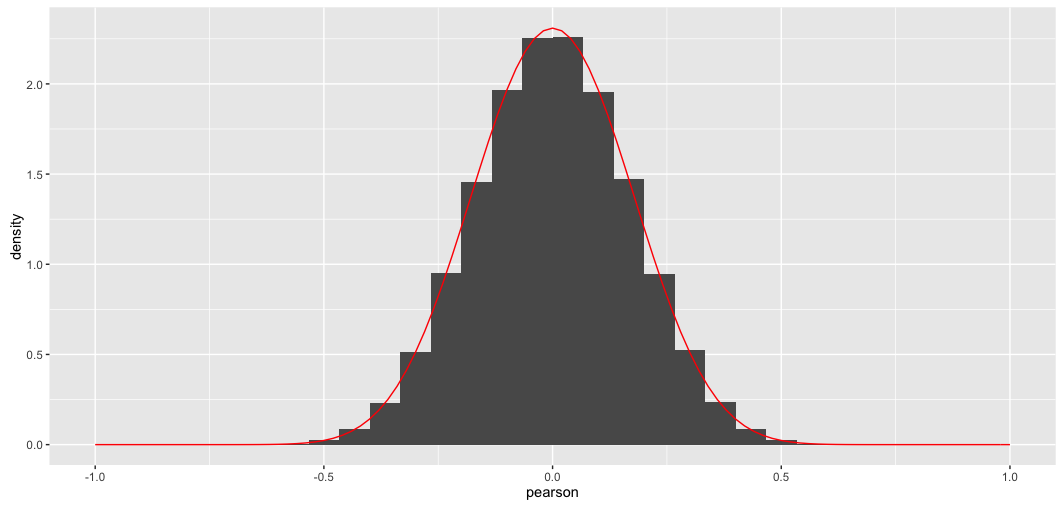
1. ***Hypothesis tests (Pearson Correlation Coefficients)***

In general, we want to do the following hypothesis test:

vs.

Let’s assume for a moment that is true and that the population correlation coefficient . If we were to have many samples (each of size *n*) of variables *x* and *y*, and calculate the sample correlation *r* for each pair of samples, we could look at the distribution of the *r*’s, which might look like the graph below ([source](https://i.stack.imgur.com/fsMAY.png)). We should note a few things:

* The distribution of *r* looks like a bell-shaped curve (this is true if two variables are normal)
* The distribution is centered around 0, which is the expected value of the statistic *R* if is true. That is, .
* The spread of the distribution will depend on the sample size. The greater the sample size, the lower the variance.



Specifically, we know that when is true, the quantity has a *t*-distribution with degrees of freedom, where *n* is the sample size. We also know that when is true, , and .

Plugging everything in, we get:

From this, we can also calculate the *p-value*, the probability of getting a value of *r* at least as large as the one we obtain in our sample if is true (i.e., if , the actual value of the correlation coefficient in the population, is ).

1. ***Example***

Slide 30 presents a sample of children. For each child, we have two variables: age and vocabulary (# of words the child knows). We can calculate the sample Pearson correlation coefficient in the R software using the *rcorr* command in the library *Hmisc* (or in Excel, using the *correl* or *pearson* functions). In our sample, .

We can also run the hypothesis test, where and :

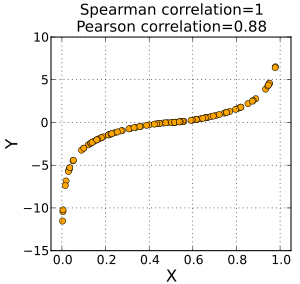
The *rcorr* command in R automatically gives us a *p-value*. We can also calculate it in Excel using the formula , where is the value of the *t*-statistic calculated above, is the degrees of freedom (), and is the number of tails, corresponding to the formulation of the alternative hypothesis . (If the form of the alternative hypothesis were , we would have a one-tailed test).

The *p-value* is computed to be , indicating that if is true and the population correlation coefficient between age and vocabulary is , then the probability of us observing a sample (of size ) where the correlation coefficient is at least is very low (i.e., ). That is, it’s very unlikely that this sample comes from a population where . Also, because the *p-value* , we reject for .

1. ***Spearman correlation coefficients***

Remember that Spearman correlations are simply Pearson correlations between ranked values of *x*’s and *y*’s. Thus, hypothesis tests for Spearman correlations will be essentially the same as the hypothesis tests for Pearson correlations.

A Spearman correlation of 1 results when the two variables being compared are monotonically related, even if their relationship is not linear. This means that all data-points with greater *x*-values than that of a given data point will have greater *y*-values as well. In contrast, this does not give a perfect Pearson correlation.

[](https://en.wikipedia.org/wiki/File:Spearman_fig1.svg)

Source: <https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient>.