

EDPSY 604
Significance Tests of Correlation Coefficients

Significance tests of correlation coefficients are generally very similar to tests of means. One difference is that the sampling distribution of r is more complex than sampling distributions for means or mean differences. The shape of the sampling distribution of r tends slowly towards the normal distribution as N increases. With an n of 3 or more, a t -test can be used to test whether a given observed r is significantly different than a hypothesized population correlation (ρ , rho) of zero:

$$t = \frac{r_{xy} \sqrt{N-2}}{\sqrt{1-r_{xy}^2}}, \text{ with } N-2 \text{ df.}$$

When $\rho \neq 0$ the sampling distribution of r is very skewed. This problem is minimized through use of Fisher's r to Z transformation. Fisher's transform is appropriate when $N \geq 10$. The transformation is calculated as:

$$Z = \frac{1}{2} \log_e \left(\frac{1+r_{xy}}{1-r_{xy}} \right)$$

but it is usually simpler to use tables like the attached. The sampling distribution of Z is approximately normal. Goodness of the approximation to the normal increases the smaller the absolute value of ρ and the larger the sample size.

Given that the variance of the sampling distribution of $Z = 1 / (N - 3)$, a z -test statistic can be calculated by taking the root of the variance to produce a standard error which is used as the denominator of the test. The numerator is analogous to that for a one sample z -test of means and tests whether the observed transformed correlation Z differs significantly from a hypothesized population value $Zeta$.

$$z = \frac{Z - Zeta}{\sqrt{1 / (N - 3)}}$$

The obtained z -test value is referred to a normal distribution to determine significance. The test requires the assumption of bivariate normality.

For example, given an $N = 100$, $r = +0.35$, and the null hypothesis that $\rho = +0.50$. Using the table of Fisher's r to Z transforms, the obtained r of .35 is equivalent to a Z of .3654 and the hypothesized population value of .50 is equivalent to a transformed value of $Zeta = .5493$. The test statistic then is:

$$z = \frac{Z - Zeta}{\sqrt{1/(N-3)}} = \frac{.3654 - .5493}{\sqrt{1/(100-3)}} = -1.81$$

The critical value from the normal distribution is +/- 1.96, so the observed correlation does not differ significantly from the hypothesized value of .50.

Hypotheses regarding differences between two correlation coefficients can also be tested:

$$z = \frac{(Z_1 - Z_2) - (Zeta_1 - Zeta_2)}{\sqrt{[1/(N_1-3)] + [1/(N_2-3)]}}$$

Of course, most commonly the difference in population Zetas is hypothesized as zero under the null, so the second term in the numerator is zero. It is also assumed that the two samples are independent and each is bivariate normal. The test should only be applied when $N \geq 10$. For example, if the researcher wanted to test whether the previous correlation of +0.35 was significantly different than a correlation of +0.25 in a second sample of 75 subjects, the following test is performed. First r 's are transformed to Z 's: $Z_1 = .3654$, $Z_2 = .2554$. Then:

$$z = \frac{(.3654 - .2554) - (0)}{\sqrt{[1/(100-3)] + [1/(75-3)]}} = -0.7071,$$

which is not significant at $\alpha = .05$ and a critical z of +/-1.96. Using the same approach, confidence intervals around an obtained correlation coefficient can also be constructed. For example, for the correlation of .35, a 95% confidence interval for ρ is:

$$Z - z(\alpha/2)\sqrt{1/(N-3)} \leq Zeta \leq Z + z(\alpha/2)\sqrt{1/(N-3)}$$

Do note the difference between Z , the transform of r , and z the critical value of z in the normal distribution. In our example of an r of +0.35 with $N = 100$:

$$.3654 - 1.96 \sqrt{1/(100-3)} \leq Zeta \leq .3654 + 1.96 \sqrt{1/(100-3)}$$

So, the interval in Z transform is: $.1664 \leq Zeta \leq .5644$

And using the r to Z transformation table to convert back to r, the CI 95% is:

$$.165 \leq \text{Rho} \leq .511$$

So, the probability is .95 that the true value of the population correlation coefficient, Rho, falls in the range from +.165 to +.511 given our sample point estimate of +.35.

The same procedures can also be applied to test the significance of partial or semipartial correlation coefficients. The only change is that the standard error of the coefficient must reflect additional reduction in degrees of freedom associated with additional predictors in the equation:

$$z = \frac{Z - \text{Zeta}}{\sqrt{1 / (N - 3 - p - 2)}}$$

Where p = the total number of variables in the equation. In analogous fashion, the formulae presented above can be adapted to test the difference between partials or semipartial or to construct confidence intervals around a point estimate of a partial or semipartial coefficient.