STAT 440 - FORECASTING
Inference on Correlations

Testing the null hypothesis H_0: $\rho = 0$

The sample correlation coefficient *r* has standard error

$$s.e.(r) = \sqrt{\frac{1-r^2}{n-2}} \text{ (when } \rho=0\text{).}$$

The test statistic $t = \frac{obs-ex}{sd} = \frac{r-0}{s.e.(r)}$ has Student's *t* distribution with n-2 df.

Confidence intervals on ρ, or tests of the null hypothesis that ρ is equal to a nonzero value

A transformation (due to Fisher):

$$z_r = \frac{1}{2}\ln\frac{1+r}{1-r}$$

is approximately normal, with mean $\frac{1}{2}\ln\frac{1+\rho}{1-\rho}$ and variance 1/(n-3).

This can be back-transformed by

$$r = \frac{e^{2z}-1}{e^{2z}+1}.$$

Fisher's z-transformation allows hypothesis tests of the form H_0: ρ = # (testing whether the correlation is equal to some prespecified value.

The transformation also allows hypothesis tests of the form H_0: $\rho_{Gp1} = \rho_{Gp2}$ (testing whether two separate groups have the same correlation between two variables). Note that here the two correlations in question are independent of one another.

If, however, the two correlations being tested are dependent on one another (for example, testing whether A is more closely correlated with B or with C), we need to make an adjustment to the test statistic.

There are several ways of doing this. Steiger (1980) outlines various procedures, indicating which ones work well (and which others work not so well). What follows is derived from his article (reference below).

Denote the two (population) correlations in question by ρ_{AB} and ρ_{AC}, and the corresponding sample values by r_{AB} and r_{AC}.

"It can be shown that" the difference between the two Fisher z-transformations of the correlations is asymptotically normal, with variance

$$Var(z_{AB} - z_{AC}) = \frac{2 - 2\cdot\xi}{n-3}$$

where

$$\xi = \frac{\rho_{BC}\cdot(1-\rho_{AB}^{\ 2} - \rho_{AC}^{\ 2}) - \tfrac{1}{2}(\rho_{AB}\cdot\rho_{AC})\cdot(1-\rho_{AB}^{\ 2} - \rho_{AC}^{\ 2} - \rho_{BC}^{\ 2})}{(1-\rho_{AB}^{\ 2})\cdot(1-\rho_{AC}^{\ 2})}$$

Note that if the two correlations are assumed equal, then in practice a pooled estimate

$$r_{pooled} = \tfrac{1}{2}\cdot(r_{AB} + r_{AC})$$

may be used as an estimate of both ρ_{AB} and ρ_{AC}.

REFERENCE: Steiger, James H., "Tests for Comparing Elements of a Correlation Matrix," *Psychological Bulletin*, vol. 87, pp. 245-251, 1980.