Permutation test for Correlation and Slope; Spearman's Rank Correlation

For paired data (X_i, Y_i), the data may be sampled either by 1) Bivariate sampling, or 2) Fixed-X sampling.

Under bivariate sampling, the population correlation coefficient

$$\rho = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \, \sigma_Y}$$

is a measure of linear association between X and Y, and is estimated by the sample correlation coefficient r, also known as the Pearson correlation coefficient. If the (X, Y) pairs are random samples from a bivariate normal population then we have a t statistic,

$$t_{corr} = \sqrt{\frac{n-2}{1-r^2}}\, r,$$

that follows a t distribution $n-2$ with degrees of freedom under $H_0 : \rho = 0$.

The linear regression model is $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, where the ε_i's are i.i.d. random variables with mean 0 and finite variance. The interpretation of the model under the two sampling schemes differs. The least-squares estimates of β_0 and β_1 are the values that minimize the sum of squared errors, or SSE. If the ε_i's are normally distributed then we can test the null hypothesis $H_0 : \beta_1 = 0$ with a t statistic:

$$t_{slope} = \sqrt{\frac{\sum(X_i - \overline{X})^2}{MSE}}\, \widehat{\beta}_1$$

that follows a t distribution $n-2$ with degrees of freedom under H_0.

The least-squares estimate of slope and the Pearson correlation are related via: $\widehat{\beta}_1 = r\ (S_Y/S_X)$, and it can be shown that $t_{corr} = t_{slope}$.

The permutation test for population correlation or slope

The permutation distribution consists of the $n!$ possible ways to permute the Y's among the X's.

Alternative statistics: S_{XY} is enough, a large sample approximation to the permutation distribution of r under H_0 has $E(r) = 0$ and $Var(r) = 1/(n-1)$. Thus, $Z = r\sqrt{n-1}$ is approximately standard normal in large samples.

Spearman Rank Correlation

The Spearman correlation, r_S, is the Pearson correlation of the pairs $(R(X_i), R(Y_i))$.

Tests for Spearman correlation: Permutation tests, Table A12, and a large sample approximation is that $Z = r_S\sqrt{n-1}$ is approximately standard normal in large samples.