Correlation

Prerequisites
Values of the Pearson Correlation, Sampling Distribution of Pearson's r, Confidence Intervals

The computation of a confidence interval on the population value of Pearson's correlation (ρ) is complicated by the fact that the sampling distribution of r is not normally distributed. The solution lies with Fisher's z' transformation described in the section on the sampling distribution of Pearson's r. The steps in computing a confidence interval for ρ are:

1. Convert r to z'
2. Compute a confidence interval in terms of z'
3. Convert the confidence interval back to r.

Let's take the data from the case study Animal Research as an example. In this study, students were asked to rate the degree to which they thought animal research is wrong and the degree to which they thought is was necessary. As you might have expected, there was a negative relationship between these two variables: the more a student thinks animal research is wrong the less they think it is necessary. The correlation based on 34 observations is -0.654. The problem is to compute a 95% confidence interval on ρ based on this r of -.654.

The conversion of r to z' can be done using a table or calculator. The table contains only positive value of r, but that is not a problem. The value of z' associated with an r of 0.654 is 0.78. Therefore, the z' associated with an r of -0.654 is -0.78.

The sampling distribution of z' is approximately normally distributed and has a standard error of

$$\frac{1}{\sqrt{N-3}}$$

For this example, N = 34 and therefore the standard error is 0.180. The Z for a 95% confidence interval ($Z_{.95}$) is 1.96 as can be found using the normal distribution calculator (setting the shaded area to .95 and clicking on the "Between" button). The confidence interval is therefore computed as:

```
Lower limit = -0.78 - (1.96)(0.18)= -1.13
Upper limit = -0.78 + (1.96)(0.18)= -0.43
```

The final step is to convert the endpoints of the interval back to r using a table or calculator. The r associated with a z' of -1.13 is -0.81 and the r associated with a z' -0.43 is -0.40. Therefore, the population correlation (ρ) is likely to be between -0.81 and -0.40. The 95% confidence interval is:

$$-0.81 \leq \rho \leq -0.40$$

To calculate the 99% confidence interval, you use the Z for a 99% confidence interval of 2.58 as follows:

```
Lower limit = -0.775 - (2.58)(0.18)= -1.24
Upper limit = -0.775 + (2.58)(0.18)= -0.32
```

Converting back to r, the confidence interval is:

$$-0.84 \leq \rho \leq -0.31$$

Naturally, the 99% confidence interval is wider than the 95% confidence interval.