Computing Effect Size
Part I

The following example borrows heavily from a paper presented by Robert Coe to the Annual Conference of the British Educational Research Association at the University of Exeter, in September, 2002.

A Sample Problem

An experiment was conducted to determine whether children learn better in the morning or afternoon. A group of 38 children were randomly assigned to either a morning or afternoon reading group. Each group listened to a story on tape, and was then asked to answer 20 questions about the story. The morning group, which listened to the story at 9 am, averaged a score of 15.2 correct answers, while the afternoon group, listening to the story at 3 pm averaged 17.9 correct answers. The difference between the two means is 2.7, but what does this tell us?

One way to understand the magnitude of the difference between these two groups is to compute the effect size. This is simply the *standardized* mean difference between the two groups. The idea of using a standardized measure is important because it means the value is expressed in terms of standard deviation units, which allows us to make direct comparisons between different things. This means, for instance, that we can directly compare the results of different trials, or different treatments, in order to evaluate outcomes. *Standardized* variables are often referred to as *z*-scores, and are easily computed by subtracting the mean from an individual score and dividing the result by the standard deviation. The formula is stated as:

$$z = \frac{X - \bar{X}}{S_X}$$

Computing Effect Size

To compute the effect size for the example above, we will use a formula that is very similar. In fact, we are simply going to subtract the mean for one group from the mean for the other group, and divide by the standard deviation. We could write this formula as:

$$Effect\ Size = \frac{Mean\ for\ pm\ group - Mean\ for\ am\ group}{Standard\ Deviation}$$

Note, if you are using a control group, then you would subtract the mean for the control group from the mean of the experimental group. If it is not possible to define a control group, then simply subtract the mean for one group from the mean of the other. In our example, we are going to subtract the mean of 15.2 for the morning group from the mean of 17.9 for the afternoon group, and divide by the standard deviation. But, which standard deviation? Therein lies a small wrinkle. Ideally, you would use the standard deviation for the population. However, since the standard deviation for the population is seldom known, we must find a way to estimate it. We are offered the following alternatives:

Cohen's *d* — If the standard deviations for each group are roughly similar, then use the *pooled* standard deviation.

Glass's Δ — If the standard deviations for each group differ, then it is likely that the assumption of homogeneity of variance has been violated, and pooling the standard deviations is not appropriate. In this case, the preference is to use the standard deviation for the control group, because it is presumed that the standard deviation for the control group is "untainted by the effects of the treatment and will therefore more closely reflect the population standard deviation" (Ellis, 2010, p. 10)

Hedge's *g* — Finally, if the groups are dissimilar in size, the preferable treatment is to weight each individual standard deviation by its sample size, then proceed to pool them.

For our sample exercise, we will presume the individual groups have relatively similar standard deviations, and our pooled standard deviation works out to be 3.3. Given this last bit of information, we can proceed to compute the effect size as follows:

$$Effect\ size = \frac{Mean\ for\ pm\ group - Mean\ for\ am\ group}{Standard\ Deviation_{Pooled}}$$

$$Effect\ size = \frac{17.9 - 15.2}{3.3}$$

$$Effect\ size = 0.8$$

Interpreting Effect Size

What does 0.8 mean? It is actually quite significant. One way of interpreting this statistic is to say that the "score of the average [child] in the [afternoon group] is 0.8 standard deviation units above the average [child] in the [morning group]" (Coe, 2002, p. 2). We could also state that the average child in the afternoon group scored higher than 79% of the children in the morning group. These statements are quite strong, and convey much more meaning to colleagues than mean differences and *p* values alone.

Other common interpretations of effect size rely on benchmarks established by Cohen (1969). He established the following scale to estimate effect size:

An effect size of 0.2 is *small*
An effect size of 0.5 is *medium*
An effect size of 0.8 is *large*

Effect Sizes for Other Statistical Tests

Obviously, this discussion of effect size pertains to situations where differences between groups are being compared on a continuous outcome. Other statistical tests, such as group comparisons with dichotomous outcomes, regression, ANOVA, chi-square, etc., have their own particular means of estimating effect sizes. These will each be explored individually, in future essays.

Equations for Cohen's *d*, Glass's Δ, and Hedge's *g*

As discussed, you will need to make a decision about whether to use a pooled standard deviation, the deviation for a control group, or a weighted pooled standard deviation. Once you have made that determination, then select the appropriate formula below.

Using a Pooled Standard Deviation

$$SD_{Pooled} = \sqrt{\frac{\sum(X_A - \bar{X}_A)^2 + \sum(X_B - \bar{X}_B)^2}{N_A + N_B - 2}}$$

Where: SD_{Pooled} is the value we are solving for
The formula in the numerator is the same as the formula for the sum of the squares for each group
N_A is the sample size for group A
N_B is the sample size for group B

Using the Standard Deviation for the Control Group

This requires no additional computation. If your standard deviations demonstrate a high degree of variability, then simply use the standard deviation for the control group.

Using a Weighted Pooled Standard Deviation

$$SD^*_{Pooled} = \sqrt{\frac{(N_A - 1)SD_A^2 + (N_B - 1)SD_B^2}{N_A + N_B - 2}}$$

Where: SD^*_{Pooled} indicates a weighted pooled standard deviation
SD^2_A is the squared standard deviation for group A
SD^2_B is the squared standard deviation for group B
N_A is the sample size for group A
N_B is the sample size for group B

References

Coe, R. (2002). *It's the effect size, stupid: What effect size is and why it is important*. Paper presented at the Annual Conference of the British Educational Research Association, University of Exeter, England.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Ellis, P. D. (2010). *The essential guide to effect sizes, statistical power, meta-analysis, and the interpretation of research results*. New York: Cambridge University Press.