Effect Size

Rick Balkin, Ph.D., LPC
Department of Counseling
Texas A&M University-Commerce
Rick_balkin@tamu-commerce.edu

Statistical vs. Practical Significance

- *Statistical significance* refers to the probability that the rejection of the null hypothesis occurred outside the realm of chance (alpha level).
- *Practical significance* refers to the meaningfulness of the differences, by specifying the magnitude of the differences between the means or the strength of the association between the independent variable(s) and the dependent variable.

Practical significance: why we need it.

- A school counselor wants to compare a set of scores on the SAT to the national norm. The population has a mean of 500 and a standard deviation of 100.

Practical significance: why we need it.

If the school counselor has 25 students in the sample with a mean ($\bar{X}$) of 520, then the z-test would be conducted as follows:

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{520 - 500}{\frac{100}{\sqrt{25}}} = \frac{20}{20} = 1.00$$

With an alpha level of .05 (non -directional) and z_{crit} = 1.96, there is no statistic ally significant difference between the sample group and the population ($z = 1.00$, $p > .05$).

Practical significance: why we need it.

Now, take the same scores, but increase the sample size to 100.

$$z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{520 - 500}{\frac{100}{\sqrt{100}}} = \frac{20}{10} = 2.00$$

With an alpha level of .05 (non -directional) and z_{crit} = 1.96, there is a statistically significant difference between the sample group and the population ($z = 2.00$, $p < .05$). The observed value is greater than the critical value (2.00> 1.96).

Balkin, R. S.(2008).

Practical significance: why we need it.

- Did you notice that with the smaller sample size you did not have statistical significance but with the larger sample size you did?
- Although the *magnitude* of the mean differences did not change, the interpretation of the results changed strictly based on the increase in sample size. When sample size increased, the error decreased.

Practical significance: why we need it.

- Thus, statistically significant differences are more likely to occur when large samples are utilized.
- Nearly any null hypothesis can be rejected when a large enough sample is attained.

Practical significance: why we need it.

- Practical significance is important because it addresses the magnitude of a treatment effect without the complication of sample size, thereby providing more meaningful information that has usefulness to practitioners and researchers (Kirk, 1995).

Practical significance: why we need it.

- The following procedures are utilized to provide measures of effect size to determine practical significance.
- Currently, statistical packages do not compute Cohen's *d* or Cohen's *f*, which measure effect size in standard deviation units. However, they are relatively simple computations.

Practical significance: why we need it.

- The reporting of practical significance is very important when reporting results and mandatory in many social science journals.
- "For the reader to fully understand the importance of your findings, it is almost always necessary to include some index of effect size or strength of relationship in your Results section" (APA, 2001, p. 25).

Cohen's d

- Cohen's *d* is used to determine the effect size for the differences between two groups, such as in a *t*-test or pairwise comparisons (i.e. Tukey post hoc), and is expressed in standard deviation units.
- Cohen (1988) created the following categories to interpret *d*:
  - Small = .2
  - Medium = .5
  - Large = .8

Cohen's d

$$\text{Cohen's } d = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{\sqrt{s^2_{error}}}$$

or

$$\sqrt{MS}$$

The numerator value is the difference between two group means. The denominator is the error term, which can be expressed in one of three ways.

Computing *Cohen's d*

Table 2.

Tukey post hoc analysis

Group Comparisons		Mean Difference	*p*	*d*
1	2	-3.00	0.0815	1.65
	3	-1.00	0.8215	0.55
	4	3.60 *	0.0301	1.97
2	3	2.00	0.3393	1.09
	4	6.60*	0.0002	3.62
3	4	4.60*	0.0052	2.52

$*p < .05$

From the ANOVA example in the notepack, the first group comparison would be computed as follows:

$$\frac{|-3|}{\sqrt{3.325}} = 1.65$$

The -3 came from subtraction of the means from groups 1 and 2 in the ANOVA example.

Understanding *Cohen's d*

- So, if *d* = 1.65, then the difference between the groups is 1.65 standard deviation units.
- This would be considered a very large effect size, as it is greater than .8.

Cohen's *f*

- Cohen's *f* also expresses effect size in standard deviation units, but does so for two or more groups.
- When conducting an ANOVA, Cohen's *f* can be computed to determine the practical significance in the differences among the groups.

Cohen's *f*

- Like the ANOVA, the Cohen's *f* will identify the magnitude of the differences among the groups, but it will not explain differences between specific groups.
- To identify differences between specific groups, a Tukey post hoc analysis followed by Cohen's *d* for each pairwise comparison would be necessary.

Cohen's *f*

- Cohen (1988) created the following categories to interpret *f*:
  - Small = .10
  - Medium = .25
  - Large = .40

Computing Cohen's *f*

$$f = \sqrt{\frac{\sum(\mu_j - \mu)^2}{(J)MS_{error}}} = \sqrt{\frac{[(6-6.1)^2 + (9-6.1)^2 + (7-6.1)^2 + (2.4-6.1)^2]}{(4)3.325}}$$

$$= \sqrt{\frac{(.01 + 8.41 + .81 + 13.69)}{13.30}} = 1.31$$

So, a large effect size was found among the four groups with an effect size of approximately 1.31 standard deviations.

Balkin, R. S.(2008).

η^2

Omega squared ω^2 and Eta-squared η^2

- Practical significance is not always measured in standard deviation units and may be expressed in variance units.
- There are mathematical relationships between effect sizes expressed in standard deviation units and strengths of association expressed in variance units.

Omega squared ω^2 and Eta-squared η^2

- However, when conducting parametric statistics, in which the focus of the study is on group differences, it is best practice to express effect size in standard deviation units as it better compliments the descriptive data, such as means and standard deviations.
- As a rule of thumb, Cohen's *d* and Cohen's *f* may be more informative for ANOVA. However, many statistical packages provide measures of strength of association, especially η^2 and ω^2, and so they are widely used.

Omega squared ω^2 and Eta-squared η^2

- Cohen (1988) created the following categories to interpret strength of association:
  - Small = .02
  - Medium = .13
  - Large = .26

Eta-squared: η^2

- Eta-squared refers to strength of association between the independent variable(s) and the dependent variable.
- It indicates the amount of variance accounted for in the dependent variable by the independent variable(s).
- If the strength of association is weak, or low, the independent variable(s) have less meaning/relevance to the dependent variable.

Eta-squared: η²

$$\eta^2 = \frac{SS_B}{SS_{TOT}} = \frac{114.6}{167.8} = .68$$

Similar to Cohen's *f*, .683 is a very large effect size. The IV accounts for 63% of the variance in the DV.

Balkin, R. S.(2008).

Omega squared: ω²

The computation of ω^2 also uses terms from the ANOVA computation:

$$\omega^2 = \frac{SS_B - (j-1)(MS_W)}{SS_{TOT} + MS_W}$$

where SS_{TOT} is the sum of $SS_B + SS_W$ and j is the number of groups.

Omega squared: ω^2

$$\omega^2 = \frac{114.6 - (4-1)(3.325)}{167.8 + 3.325} = \frac{104.625}{171.125} = .61$$

From this statistic, we can conclude that the four groups of students account for 61% of the variance in self-efficacy scores.

Effect size summary

- Keep in mind, effect size is always computed when a statistical test is conducted.
- Even if your *F*-test is not significant, you should still report a Cohen's *F*. However, since a TMC is not conducted for a non-significant test, no further analysis is necessary.