Evaluation Take Aways

Free Evaluation Technical Assistance from the Staff of ACET, Inc.

# Statistical Power

Statistical power, technically speaking, is the ability to detect a statistically significant effect from a sample when a statistically significant effect actually exists. In evaluation and research, generally we want to maximize power usually by increasing the sample size. Unfortunately, there is a cost to increasing sample sizes – it takes more money and more time. So in order to maximize power and minimize cost, we can use a power analysis to determine the ideal “zone” for a target sample size. In order to identify the ideal sample size we need to know two things: the effect size and sample size.

Effect size, appropriately named, refers to how big of an impact a variable or treatment has on outcomes. In general, the bigger the difference between the treatment and comparison groups, the larger the effect size. And the larger the actual effect we are trying to detect, the easier it will be to see its impact on outcome scores. For example, let’s say our population has two groups of elementary students, one using a specialized math program and the other using no specialized program. The effect size would be the difference in students’ math scores (our outcome) between the students who used the specialized math program (the treatment group) and the students who did not (the comparison group). Effect size is computed using this formula:

$$\frac{Mean\ of\ Treatment\ Group - Mean\ of\ Comparison\ Group}{Standard\ Deviation}$$

Sample size refers to the total number of cases within a sample. In the example, the sample size would be the total number of students being studied – those in the treatment group *plus* those in the comparison group. In general, the larger the sample size the more likely that the differences between groups are real and not due to chance. For example, if math scores were compared for one student using the specialized math program and one student who did not, any differences in their scores could be easily explained as random or due to chance. But if 50 students participated in both groups, any differences in the scores would likely be real and not due to some random event.

Effect size and sample size have an inverse relationship. When effect sizes are large they are easy to see, even with small sample sizes; but when effect sizes are small, they are a challenge to see and require large sample sizes. Power analysis incorporates this inverse relationship to estimate whether or not an evaluator or researcher will find statistically significant outcomes, and can do so both before and after data collection.

The table below shows effect size and the required sample for two levels of power. Three different effect sizes are displayed: large (0.80), moderate (0.50), and small (0.20). As can be seen in the table, as effect size decreases the required sample for adequate power increases. For example, with a moderate effect size of .50 and power at .95 the required sample size is 34, but with a small effect size of .20 with the same power level the required sample is 261.

| Effect Size | Power = .90 | Power = .95 |
|---|---|---|
| .80 | 7 | 8 |
| .50 | 27 | 34 |
| .20 | 207 | 261 |

For additional information about this or other ACET, Inc. resources, or for evaluation assistance, please contact:

ACET, Inc.
info@acetinc.com
952.922.1811

Suggested Citation:
ACET, Inc. (2013). *Statistical power*. Minneapolis, MN.