

APPENDIX TO PART III

Considerations on determining sample size to give desired power: (pp. 19, 26)

- The difference used in calculating sample size (i.e., the specific alternative used in calculating sample size) should be decided on the base of practical significance and/or "worst case scenario," depending on the consequences of decisions.
- Even when the goal is a hypothesis test, it may be wise to base the sample size on the width of a confidence interval rather than just ability to detect the desired difference: Even when power is large enough to detect a difference, the uncertainty, as displayed by the confidence interval, may still be too large to make the conclusions very credible to a knowledgeable reader.
- Determining sample size to give desired power and significance level will usually require some estimate of parameters such as variance, so will only be as good as these estimates.
 - These estimates usually need to be based on previous research, experience of experts in the field, or a pilot study.
 - In many cases, it may be wise to use a conservative estimate of variance (e.g., the upper bound of a confidence interval from a pilot study), or to do a sensitivity analysis to see how the sample size estimate depends on the parameter estimate. See Lenth (2001) for more details.
- Even when there is a good formula for power in terms of sample size, "inverting" the formula to get sample size from power is often not straightforward
 - This may require some clever approximation procedures.
 - Such procedures have been encoded into computer routines for many (not all) common tests.
 - See Russell Lenth's website or John C. Pezzullo's Interactive Statistics Pages for links to a number of online power and sample size calculators.
 - *Caution:* If you use software routines to calculate power, be sure it calculates *a priori* power, not *retrospective (or observed) power*. (See below)
- Good and Hardin (2006, p. 34) report that using the default settings for power and sample size calculations is a **common mistake** made by researchers.
- For *discrete* distributions, the "power function" (giving power as a function of sample size) is often saw-toothed in shape.

- A consequence is that software may not necessarily give the optimal sample size for the conditions specified.
- Good software for such power calculations will also output a graph of the power function, allowing the researcher to consider other sample sizes that might give be better than the default given by the software.

References for tests of equivalence (p. 30):

- Hoenig, John M. and Heisey, Dennis M. (2001
- Graphpad.com, Statistical Tests for Equivalence,
http://www.graphpad.com/library/biostatsspecial/article_182.htm
- Lauchenbruch, P. A. (2001)

Note regarding Cohen's d (p. 32):

Figure 1 of Browne (2010) This shows that, for the two-sample t-test, Cohen's classification of "large" d as 0.8 still gives substantial overlap between the two distributions being compared; d needs to be close to 4 to result in minimal overlap of the distributions.

REFERENCES:

Browne, Richard H. (2010). The t-Test p Value and Its Relationship to the Effect Size and $P(X>Y)$, *The American Statistician*, February 1, 2010, 64(1), p. 31

Claremont Graduate University WISE Project Statistical Power Demo,
http://wise.cgu.edu/powermod/power_applet.asp

Festing, Michael, Statistics and animals in biomedical research, *Significance* Volume 7 Issue 4 (December 2010),
<http://www.significancemagazine.org/details/magazine/879779/Statistics-and-animals-in-biomedical-research-.html>

Good and Hardin (2006 or 2010), *Common Errors in Statistics*, Wiley

Hoenig, John M. and Heisey, Dennis M. (2001), "The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis," *The American Statistician*, 55, 19-24

Kilkenny et al, (2010) Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. *PLoS Biology*, 8,
<http://www.plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.1000412>

Lauchenbruch, P. A. (2001), Equivalence Testing,
http://www.fda.gov/ohrms/dockets/ac/01/slides/3735s1_02_lachenbruch/index.htm

Lenth, Russell V. (2000), Two Sample-Size Practices that I Don't Recommend, comments from panel discussion at the 2000 Joint Statistical Meetings in Indianapolis, <http://www.stat.uiowa.edu/%7ErLenth/Power/2badHabits.pdf>

Lenth, Russell V. (2001) Some Practical Guidelines for Effective Sample Size Determination, *American Statistician*, 55(3), 187 – 193.

A discussion of many considerations in deciding on sample size. An early version and some related papers can be downloaded from his website (below)

Lenth, Russell, Power website, <http://www.stat.uiowa.edu/~rlenth/Power/> Has several online applets for calculating power, some advice on using the applets, and links to some papers on power.

Lipsey, MW (1990). *Design sensitivity: Statistical power for experimental research*. Newbury Park, CA: Sage.

Maxwell, S. E. and K Kelley (2011), Ethics and Sample Size Planning, Chapter 6 (pp. 159 - 183) in Panter, A. T. and S. K. Sterba, *Handbook of Ethics in Quantitative Methodology*, Routledge

Maxwell, S.E. (2004), The persistence of underpowered studies in psychological research: Causes, consequences, and remedies, *Psychological Methods* 9 (2), 147 - 163.

McClelland, Gary H. (2000) Increasing statistical power without increasing sample size, *American Psychologist* 55(8), 963 – 964

Rice Virtual Lab in Statistics, Robustness Simulation
http://onlinestatbook.com/stat_sim/robustness/index.html

Pezzullo, John C., Power, Sample Size and Experimental Design Calculations, *Interactive Statistics Pages*, <http://statpages.org/#Power>
Has links to online power calculations; I'd suggest trying Russell Lenth's page (above) first.

Wuensch, K. L. (1994). Evaluating the Relative Seriousness of Type I versus Type II Errors in Classical Hypothesis Testing,
<http://core.ecu.edu/psyc/wuenschk/StatHelp/Type-I-II-Errors.htm>

Wuensch, K.L. et al (2003), "Retrospective (Observed) Power Analysis, Stat Help website, <http://core.ecu.edu/psyc/wuenschk/stathelp/Power-Retrospective.htm>