Power Analysis

Power analysis estimates the sample size needed to detect an effect of a given size with a given degree of confidence. Conversely, we can estimate the probability of detecting an effect of a given size with a given level of confidence and given sample size. This is really important *before you conduct an experiment* if you want to avoid wasting sweat and time. You can also conduct the same analyses after a study to find how much power you obtained.

The following four quantities are inter-related:

1. Sample size = the only measure for which you have direct control
2. significance level = P(Type I error) = probability of mistakenly finding an effect that does not exist (a false positive). Let's assume significance is prescribed ($p \leq 0.05$).
3. Effect size = many different measures exist; http://en.wikipedia.org/wiki/Effect_size. Below we use leading effect size measures for regressions and ANOVAs.
4. power = 1 - P(Type II error) = probability of finding an effect that is there

We can solve for any one of the above if we enter the other three. The overall goal is to minimize #2 and maximize #3 & #4. You can only do this by changing #1, or by picking a value of #3 to find what sample size is needed.

In all examples below, we work with our helicopter experiment.

1. First, install and load the `pwr` package in RStudio.

ANOVAs. Here we use Cohen's *d*, which is essentially the weighted sum of differences from the grand mean for each group, relative to the overall variance.

2. As a simple reminder of our results, ignore our factorial design, blocks, and steps, and simply compute a one-way ANOVA for Design effects on Time. For convenience, copter data are at:

   http://jenkins.cos.ucf.edu/wordpress/wp-content/uploads/copter-data-F16.csv

3. Run the following command, to s*olve for power* by leaving it blank - but **fill in:**

   `k` = the number of treatment groups,

   `n` = the intended sample size in each group,

   `f` = effect size (0.1, 0.25, and 0.4 represent small, medium, and large effect sizes, respectively).

   significance level = 0.05.

   ```
   pwr.anova.test(k = , n = , f = , sig.level = , power = )
   ```

   What are the odds that we will find an effect that is really there?

2. Now repeat that command but leave `n` blank and put in power=0.80. This tells us how many replicate helicopters we would need for power=0.80 (which is a common cutoff).

3. What happens if you back off power, say to 0.60? Would you be willing to let the odds of finding an effect that really exists = 50%? How low are you willing to go for your own research?

We ran the experiment with ~10 replicate helicopters per treatment; were our chances of detecting an effect reduced by that choice? What if we had used more replicates?

In your thesis you may struggle to get enough samples per treatment for power = 0.8. Very few experiments include dozens of replicates per treatment, unless the study system was highly variable and samples were likely to be lost.

So what can you do for your research? Think carefully to design an experiment most likely to clearly obtain differences between groups – clear differences increase your effect size. Clear differences come from consistent results per treatment and big differences between treatments. That is much easier in the long run than sampling 50 replicates per treatment! So in a range of 8 treatment settings (like the helicopters), controls and maximum settings help ensure strong differences. Intermediate treatments may be most interesting but uncertain in effect size.

Regressions. Here we use Cohen's f^2, which is simply R^2 / (1-R^2), where R^2 is the coefficient of determination for the regression (0-1 score for how close points are to the regression line). An important term is degrees of freedom (df; for a simple design, assume df = (number of samples – 1).

1. Again, use the `pwr` package to run the following command. *Solve for power* by leaving it blank but **fill in:**

   `u` = (the number of predictor variables - 1),

   v `=` (the total number of data points `-` 1),

   `f2` = effect size (0.02, 0.15, and 0.35 represent small, medium, and large effect sizes, respectively).

   significance level = 0.05.

   ```
   pwr.f2.test(u =, v = , f2 = , sig.level = , power = )
   ```

2. Now repeat that command but leave `v` blank and put in power=0.80. This tells us how many total helicopters we would need for power=0.80.

If you used the number of *designs* to calculate u, that is not really a regression. Try it for wing lengths instead.

The above power calculations can also be conducted after a study – by entering the needed values, including the observed significance level, and solving for power.

Other forms of power tests (e.g. *t* tests) exist with `pwr` – for a good summary see **http://www.statmethods.net/stats/power.html**.