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. If you run a power analysis *before you conduct an experiment*, you can avoid wasting years of sweat and time. But that first requires an estimate of variation – such as from preliminary data (a reason to run a preliminary study) or even someone else's study. You can also conduct the same analyses after a study to find how much power you obtained (and hope it was enough).

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. Let's assume significance is prescribed ($p \leq 0.05$).
3. **Effect size** = many different measures exist; see 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 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, work with our helicopter experiment (number of treatment groups = 4, number of helicopters per treatment = 10).

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

<u>ANOVAs</u>. Here we use Cohen's *d*, which is essentially weighted differences, or Σ(group mean - grand mean), relative to the overall variance. We ignore the factorial design for now, and simply analyze Design differences as a more simple ANOVA.

2. Run the following command, to s*olve for power* by leaving it blank but **fill in** `k` =the number of treatment groups, `n` = the sample size in each group, and `f` = effect size (0.1, 0.25, and 0.4 represent small, medium, and large effect sizes, respectively). Assume significance = 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?
3. 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.
4. 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 used 5 replicate helicopters per treatment. Some years we used 10. Depending on your own study, you may struggle to get that many replicates. **So what can you do for your research?** Think carefully to design an experiment most likely to obtain clear differences between groups, because clear differences increase your *effect size*. Clear differences come from consistent results per treatment and big treatment effects. That is much easier in the long run than sampling 50 replicates per treatment! So in a range of four treatment settings (like the helicopters), controls and maximum settings (e.g., 100% mortality) help anchor the ends of the responses. Then choose intermediate treatments that are more interesting but uncertain in effect size.

<u>Regressions</u>. 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 – for now, assume df = (number of samples – 1).

1. Again, using the `pwr` package, run the following command, to s*olve 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). Assume significance = 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.

This analysis treats designs as predictor variables in a regression – not quite correct, but you should get the idea.

The above power calculations can 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**.