Problem Set 3

Research Design for Causal Inference
Due: April 21, 2015

Part I – Gerber & Green Chapter 3 Exercises

Excercise 1

Exercise 5

Exercise 6 *(Bonus: verify/implement G & G's calculation of the 95% CI for this estimate using statistical software.)*

Excercise 12

Part II – The Lady Tasting Tea

In a very famous example that effectively invented Randomization Inference, Sir R.A. Fisher's (1935) provided the following description of a Lady Tasting Tea:

> A lady declares that by tasting a cup of tea made with milk she can discrimate whether the milk or the tea infusion was first added to the cup.

To test The Lady's claim, he proposes to make six cups of tea – three of them "milk first" and three of them "tea first" – and to present the six cups to The Lady in random order. The Lady knows that this will be the study design, so she knows she will receive three cups of each type and that the order of the presentation will be random.

Part a Suppose that The Lady correctly identifies whether milk or tea was added first for all six out of six cups presented to her in the experiment. We call the milk-first cups the treatment ($A_i = 1$) and The Lady's guess $Y_i = 1$ when she guesses milk-first. Use the number of correctly identified cups as the test statistic, formally represented by the following equation:

$$\sum_i A_i Y_i + (1 - A_i)(1 - Y_i) \tag{1}$$

What is the p-value for a test under the "sharp null hypothesis" that The Lady has no ability to discriminate?

> **Hint**: Use your knowledge of the research design to calculate the total number of potential outcomes for the experiment and then to calculate the likelihood that The Lady would produce 6 out of 6 correct answers if she had been guessing randomly).

Part b Imagine that The Lady makes one mistake and therefore correctly identifies whether milk or tea was added first for four out of the six cups (revisit the study design if you're confused as to why one mistake yields two wrong answers). What is the p-value for a test under the sharp null hypothesis that she has no ability to discriminate under these conditions?

Part c Now, imagine that instead of using "fixed margins" for the randomization (i.e. requiring three milk-first cups as well as three tea-first cups) you were to conduct Fisher's experiment using "binomial randomization" (i.e. randomly determining whether to add milk or tea first *for each cup*). For each cup, we'll say that there is a probability $p = 1/2$ that it is milk-first and a $1-p$ probability that it is tea-first. In this case, The Lady does not know the value of p, but she does know that the cups have been assigned under binomial randomization. If, under these conditions, The Lady makes one mistake, what is the p-value for a test under the sharp null hypothesis of no effects?

> **Hint**: To do this, consider that under binomial randomization in an experiment with N trials, the probability, Pr, that the number of "successes", Y (in this case, correctly identified cups of tea), is equal to some value y is:
>
> $$Pr(Y=y) = \binom{N}{y} p^y (1-p)^{N-y} \tag{2}$$
>
> Read $\binom{N}{y}$ as "N choose y." Note that if you want to calculate $\binom{N}{y}$ you can probably do so easily by hand for small numbers, but as the numbers grow you may want to use the `choose()` function in R, a scientific calculator, or Wolfram Alpha to handle the calculations.

Part III – Key concepts for next class

- Covariates.
- Covariate adjustment.
- The difference between an estimator and an estimate.
- Sampling distribution of an estimator.

- Covariate imbalance, how to test for it, and the implications if you find it.
- Comparison of difference-in-means, regression, and randomization inference.
- Block randomization as an alternative to post-hoc covariate adjustment.