STATISTICS 265 – Spring 2018 – Homework 1

Handed out: Tuesday April 17, 2018
Due: Tuesday May 1, 2018

NOTE: There is some computing required for this assignment. You can use any software or programming language that you like. A sample R program that does the exact (and simulated) p-value calculation for the $n = 6$ honey experiment in Chapter 5 is provided at the end of the assignment. Please email or come see me if you need help.

1. **Completely randomized study - Fisher's test:** The Table below gives data from a randomized experiment carried out around 1970 in Southern Florida to evaluate cloud seeding (a technique that involves injecting clouds with a chemical to increase rainfall). On each of 52 days that were suitable for cloud seeding, a random mechanism was used to decide whether to seed or not. A plane flew through the clouds each suitable day (both seeding and non-seeding days). The experimenter set up the plane to inject or not depending on the random outcome. The plane's pilot did not know whether the seeding mechanism was loaded to inject or not. Precipitation was measured as total rain volume (in acre-feet) resulting from the clouds in the 24 hours following the plane's run. The data:

   Seeded days (n=26): 2745.6, 274.7, 115.3, 1697.1, 274.7, 92.4, 1656.4, 255.0, 40.6, 978.0, 242.5, 32.7, 703.4, 200.7, 31.4, 489.1, 198.6, 17.5, 430.0, 129.6, 7.7, 334.1, 119.0, 4.1, 302.8, 118.3

   Unseeded days (n=26): 1202.6, 87.0, 26.0, 830.1, 81.2, 24.4, 372.4, 68.5, 21.4, 345.5, 47.3, 17.3, 321.2, 41.1, 11.5, 244.3, 36.6, 4.9, 163.0, 29.0, 4.9, 147.8, 28.6, 1.0, 95.0, 26.3

   Data are available on the course website

   `http://www.ics.uci.edu/~sternh/courses/265/`

   as either a .Rdata file or a comma delimited data file (.csv). You can load the .Rdata file (if you have R installed) by clicking on it, or by saving it to your machine and using the load command in R (as follows)

   ```
   load("z:\\HAL\\Courses\\Stat265\\cloudseed.Rdata")
   ```

   where my course directory should be replaced by the location where you have the data stored. You can read the .csv file with the read.csv command in R

   ```
   a <- read.csv("z:\\HAL\\Courses\\Stat265\\cloudseed.csv")
   ```

   (a) Why did the plane fly through the cloud even on non-seeding days? Why were the pilots not informed as to whether the cloud seeding mechanism was set to inject or not?

   (b) Suppose we want to test Fisher's sharp null hypothesis of no treatment effect. Use the difference in sample means as a test statistic and use simulation to approximate the p-value. (Generate 1000 simulated random assignments under the null hypothesis, compute the test statistic for each, and determine the proportion of trials with a larger outcome than the observed statistic.) See the sample program below for help.

   (c) How accurate is the simulation approximation? Explain by referring to the expected variability in a sample proportion of this type.

   (d) Report the p-value you would obtain using a traditional t-test to compare the two group means. Is this related to the p-value in (a)? Explain.

   (e) The test statistic used above is sensitive to a change in the mean amount of rainfall. Suppose that you expected the treatment to increase variability of rainfall but not the mean. Develop a test statistic that would be sensitive to this alternative hypothesis and obtain an estimate of the p-value using this test statistic.

   (f) The wide range of data values suggest that the impact of cloud seeding may be multiplicative rather than additive. Repeat the test once more using the difference in the mean of the log rainfall as a test statistic. What p-value do you obtain?

   (g) What conclusion do you draw about these data?

2. **Completely randomized study - Neyman's test.** A randomized study was carried out in 12 first-grade classrooms spread across one school district to investigate the effectiveness of a new reading curriculum. Six of the classrooms were randomly assigned to use the new curriculum while the other six used the traditional curriculum. Average scores for each class on a standardized test were measured at the end of the year. The data:

   new curriculum: 57.6, 43.8, 66.5, 70.9, 72.7, 80.8
   old curriculum: 29.3, 39.3, 51.1, 59.9, 62.2, 63.5

   (a) A colleague of the experimenter notes that the sample size would have been larger if she had used individual student exam scores as the responses. Explain why the original design is superior.

   (b) Use Neyman's procedure to estimate the average treatment effect in this set of classrooms. Give an approximate 95% confidence interval for the average treatment effect. Is there evidence to reject the null hypothesis of no treatment effect?

   (c) Students in the district were all given a pretest at the start of the year. For the new curriculum group the average pretest scores (in the same order as the classes listed above) were 46.6, 63.2, 65.4, 66.0, 73.4, 84.8. For the old curriculum group the average pretest scores were: 50.0, 55.9, 55.8, 69.8, 81.1, 91.5. Based on these pretest scores do you believe that treatments were randomly assigned? Why or why not?

   (d) One way to use the pretest scores is to focus on changes (post - pre) rather than just looking at the end of year scores. Repeat part (b) using the change in score. How does your answer change?

   (e) Another way to use the pretest scores is to use a regression model, regressing end of year score on pretest score and treatment group. What does the regression model tell you about the treatment effect? (In R you can carry out a regression using the "lm" (linear model) command. If you have created variables called "pre", "post","trt", then the regression is run by the first command below and the typical summary table is produced by the second command below.)

   ```
   regmodel <- lm(post ~ pre + trt)
   summary(regmodel)
   ```

   (f) Compare the results of the previous two analyses. Is each valid for this randomized study? What are the advantages and disadvantages of each approach?

3. **Theory.** Consider the education experiment in the previous problem.

   (a) Write the test statistic for part (b) (the difference in sample means) as an explicit function of the potential outcomes and the assignment indicators for the 12 units.

   (b) Show that under complete randomization the statistic is unbiased for the average treatment effect in the 12 units.

   (c) Write the test statistic for part (d) (the difference in the sample mean gain (post - pre) scores) as an explicit function of the potential outcomes, the covariate (pretest scores) and the assignment indicators for the 12 units.

   (d) Show that under complete randomization this statistic is unbiased for the average treatment effect in the 12 units.

   (e) If both are unbiased (they should be), then why are the two statistics different in part (b) and (d) of the previous problem?

4. **Project.** There is a final project expected for this class. You can work alone or in a group of two. You will be expected to present a twenty-minute talk on your project (to be delivered during Week 10 and Finals Week) and produce a 5-7 page writeup.

   There is considerable flexibility when it comes to topic. You can: (1) focus on a methodological topic that we don't cover (see below); (2) review the literature in a controversial area of application where causal arguments play a role; or (3) carry out an analysis (or reanalysis) of data from an observational study using methods in the class. For this assignment your responsibility is to identify your partner (if any), list one (or more) possible topics, and write a paragraph on what you are thinking about for the topic(s). (I say one or more topics in case you want feedback on more than one idea.)

   In terms of methodology – possible topics include

   - failure of SUTVA (e.g., peer effects)
   - studies with more than 2 treatments
   - sequential treatment problems (repeated treatments per subject)
   - regression discontinuity designs
   - artificial/synthetic controls
   - causes of effects

For application areas virtually anything is possible including effects of global warming, SAT coaching, education vouchers, etc. For your own analysis – you/we would need to find a data set like the ones used in Imbens/Rubin or elsewehere.

Sample R programs for (1) and (2)

```
#
# PROGRAM TO COMPUTE RANDOMIZATION DISTRIBUTION FOR HONEY DATA IN CHAPTER 5
#
# set data with "nt" treatment obs followed by "nt" control obs
n <- 6
yobs <- c(3,5,0,4,0,1)
nt <- n/2
#
# compute observed test statistic
sumy <- sum(yobs)
tobs <- sum(yobs[1:nt])/nt - sum(yobs[(nt+1):n])/nt
#
# set up vector to hold randomization distribution of test stat
out <- rep(0,choose(n,nt))
cnt <- 0
#
# loop to identify unique randomizations
for (i1 in (1:(nt+1))) {
   for (i2 in ((i1+1):(nt+2))) {
      for (i3 in ((i2+1):n)) {
#
# compute test statistic for randomization and store
         cnt <- cnt + 1
         trtsum <- sum(yobs[c(i1,i2,i3)])
         out[cnt] <- trtsum/nt - (sumy-trtsum)/nt
         }}}
#
# compute p-value
pval <- sum(abs(out) >= abs(tobs))/choose(n,nt)

#
# SIMULATION VERSION
# uses the R sample command to choose the sample
n <- 6
yobs <- c(3, 5, 0, 4, 0, 1)
nt <- n/2
sumy <- sum(yobs)
tobs <- sum(yobs[1:nt])/nt - sum(yobs[(nt+1):n])/nt
outsim <- rep(0,1000)
for (i in (1:1000)) {
   ysamp <- sample(yobs,nt)
   trtsum <- sum(ysamp)
   outsim[i] <- trtsum/nt - (sumy-trtsum)/nt
   }
pvalsim <- sum(abs(outsim) >= abs(tobs))/1000
```