Introduction to Biostatistics (BIOS 4120)
Breheny

Assignment 1
Due Tuesday, January 30

1. One study on slavery in America estimated that "11.9% of slaves were skilled craftsmen." This estimate was based on the records of thirty plantations in Plaquemines Parish, Louisiana. The study is making a generalization about a population based on a sample.

   (a) In this study, what is the population of interest?

   (b) In this study, what does the sample consist of?

   (c) Could the study's generalization be biased? If so, what type of bias would this be?

2. A Dutch study of rheumatoid arthritis found that patients who experienced more pain were more likely to respond to a health survey. In their questionnaire, 2% of respondents said that they experience no pain associated with their condition.

   (a) Is the "2%" number an estimate or a parameter of interest?

   (b) Is the "2%" number a biased estimate? If so, what type of bias is present?

   (c) Is the true percentage of individuals with rheumatoid arthritis who suffer no pain likely to be less than 2%, greater than 2%, or equal to 2%?

3. This question involves an article by Abramowicz *et al.* (2008) on the epidemiology of ultrasound [Link]. Read the section on the second page titled "Low Birth Weight."

   (a) The authors discuss the use of animal studies to make inferences about humans. What sampling concept that we discussed in class does this violate?

   (b) Are the observational studies and controlled experiments in agreement?

   (c) Do the authors seem to place more emphasis on the observational studies or the randomized controlled experiments?

   (d) It is not explicitly stated in the article, but were the controlled experiments randomized controlled, double-blind experiments?

4. In 1975, The National Institutes of Health conducted a randomized controlled double-blind experiment to determine whether vitamin C is effective at treating the common cold. Subjects who developed a cold were prescribed capsules, to be taken six per day for the first five days of the cold. The subjects were randomly assigned to receive either vitamin C capsules or a placebo. Almost all participants adhered to the protocol: 99% took at least four capsules per day.

   (a) The average duration of cold for the vitamin C group was 6.5 days, while the average duration in the placebo group was 7.1 days. Does this provide evidence that vitamin C is effective at treating colds?

   (b) Could the results described in part (a) be affected by confounding?

(c) Investigators later discovered that the blinding of subjects failed in this study (vitamin C has a characteristic sour taste that the placebo lacked; this was noticed by 42% of subjects). How does this new information affect your conclusion from part (a)?

(d) If we restrict analysis only to those patients who remained blinded, the average duration of cold for the vitamin C group was 6.7 days, while the average duration in the placebo group was 6.3 days. Does this analysis provide evidence that vitamin C is effective at treating colds?

(e) Are there any potential problems with the analysis in part (d), in which we exclude certain subjects from the analysis after they have been randomized to a group?

(f) Based on all of the information you have been provided with in this question, do you think this experiment provides evidence that vitamin C is effective at treating colds? Or do you think that the study is inconclusive?

5. A 1979 randomized trial was performed comparing a surgical treatment for angina pectoris (chest pains due to obstruction of the coronary arteries) to a placebo (non-surgical medical management). In the study, 6 of the patients randomized to receive surgery died before they could be operated on. There was a subsequent debate over how best to analyze the data. Some favored approach A, in which all patients were analyzed as they were randomized. Others favored approach B, in which those 6 patients were excluded on the grounds that they never actually received the surgical treatment that they were randomized to. Which approach do you think is better? Why?

6. In class, we discussed the polio and clofibrate studies, both which had problems with "non-compliance", in which subjects did not want the treatment assigned to them. In the polio study, we were able to sidestep the problem of non-compliance by only studying the compliant group. Why is this not an option in the clofibrate study?

7. True or False: In a hypothesis test, the null hypothesis can be summarized as "nothing is going on besides chance variation."

8. The null hypothesis is a hypothesis about (i) the sample (ii) the population

9. Suppose that a scientist carries out 100 hypothesis tests. Unbeknownst to her, in all 100 cases, the null hypothesis is true. If she uses a cutoff of $p < 0.15$ to reject the null, about how many mistakes would you expect that she makes?