Introduction to Biostatistics (BIOS 4120)
Breheny

Assignment 2
Due: Tuesday, February 10

1. A drug has been developed that may reduce a person's cholesterol. Investigators are interested in estimating the amount by which the drug will reduce cholesterol, and in calculating a confidence interval. For each of the following, say whether the change will cause the confidence interval to get wider or get narrower:

   (a) The investigators decide to enroll more people in the study

   (b) More sophisticated lab techniques are used, allowing for more accurate measurement of cholesterol

   (c) The investigators want a 99% confidence interval instead of a 95% confidence interval

2. Which statistical procedure tells you more about the clinical significance of a study? (i) confidence intervals (ii) hypothesis tests

3. If one analyzes the clofibrate study as it was randomized, the p-value is 0.51.

   (a) True or false: There is a 51% probability that the null hypothesis is true.

   (b) True, false, or cannot be determined: A 75% confidence interval for the effectiveness of the drug would contain the null hypothesis value.

4. Read the short article "That Confounded p-value", by Lang *et al.*, available on the course website under "Articles".

   (a) The authors say that the information conveyed by p-values is "confounded." What do they mean by that?

   (b) On the second page of the article, the authors remark, "No one could infer the [confidence intervals] from the p-values. Given the [confidence intervals], no one needs these p-values." Would you agree with this sentiment? (There isn't necessarily a right or wrong answer here. You could probably make a valid argument either way; I just want to hear your opinion along with a reasonable justification)

   (c) The authors indicate a desire "to ban the reporting of all p-values from *Epidemiology*." Does this mean that they want to ban all statistical analysis?

5. A common tool in laboratory medicine and biology is the microarray, a device that can measure the expression levels of thousands of genes at once. So, for example, an investigator might collect samples from normal subjects and subjects with cancer in the hopes of finding genes that are significantly associated with cancer. The investigator is therefore testing a null hypothesis for every gene on the array. Suppose that a given array has 2,000 genes, of which 20 are truly associated with cancer. Suppose further that the investigator's hypothesis tests have a Type I error rate of 5% and a Type II error rate of 20%.

(a) Out of the 2,000 hypothesis tests that the investigator carries out, how many are type I errors?

(b) How many are type II errors?

(c) How many times did the investigator correctly reject the null hypothesis?

(d) What was the investigator's false discovery rate?

(e) If, for each gene, a 95% confidence interval was calculated for the association between the gene and cancer status, how many of those confidence intervals would contain the true association for that gene?

6. A study of sexual bias in admissions was conducted by the Graduate Division at the University of California, Berkeley. Admissions results from the six largest majors are listed below (university policy does not allow the departments to be identified by name):

	Men		Women	
Major	Applicants	Admitted	Applicants	Admitted
A	825	512	108	89
B	560	353	25	17
C	325	120	593	202
D	417	138	375	131
E	191	53	393	94
F	373	22	341	24

(a) From the table above, calculate the overall percentage of men and the overall percentage of women who were admitted.

(b) Does your answer from part (a) suggest sexual bias? If so, against whom?

(c) Create a table listing the percentage of men and women who were admitted, broken down by department.

(d) Does the table you made in part (c) indicate sexual bias? If so, against whom?

(e) Construct a weighted average of the percentage of male and female applicants who were admitted, controlling for the effect of department (*i.e.*, report one number for men and one number for women).

(f) Does your answer from part (e) indicate sexual bias? If so, against whom?

(g) The purpose of the study was to answer the question, "Were admissions biased on the basis of sex"? The analyses above indicate different answers to this question. Which of the analyses in (b), (d), and (f) do you feel best answers this question? Why?