Biostatistical Methods I (BIOS 5710)
Breheny

Assignment 3
Due: Thursday, September 22

1. 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 or false: A 75% confidence interval would contain the null hypothesis value.
2. 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?
3. There are 8 people in a club. One person makes a list of all the possible committees containing 2 members. Another person makes a list of all the possible committees containing 5 members. Who will have the bigger list?
4. In 2014, 32.2% of all babies born in the United States were delivered via cesarean section. Suppose that this is roughly the percentage of deliveries via C-section at the University of Iowa Hospitals & Clinics (UIHC) as well.
   (a) What is the probability that two out of the next eight babies delivered at UIHC will be born via C-section?
   (b) What is the probability that at least one out of the next eight babies delivered at UIHC will be born via C-section?
5. The binomial distribution makes three key assumptions. For each of the assumptions listed below, give an example of a study with a binary outcome for which the assumption would not hold. You do not have to use the same example for each assumption.
   (a) The number of trials is *fixed* (i.e., not random).
   (b) The probability of the outcome, p, is the same each time the process is repeated.

(c) The trials are *independent* (i.e., the outcome of one trial does not affect the outcome of the next trial).

6. Consider the following (incorrect!) argument: "Assumptions (b) and (c) from the previous problem are redundant. If the outcome of one trial affects the outcome of the next, then obviously the probabilities cannot be the same each time the process is repeated." For each of the following, give a counterexample involving balls and urns.

   (a) How could the trials be independent, but not have identical probabilities from outcome to outcome?

   (b) How could the trials have the same probability p, but not be independent?

7. {[vBFHL 6.3]} Our textbook provides the following data from Sweden regarding the gender of a child and the parity (number of times the mother has given birth):

	Order of birth							
Gender	1	2	3	4	5	6	7	Total
Males	2846	2554	2162	1667	1341	987	666	12,223
Females	2631	2361	1996	1676	1230	914	668	11,476
Total	5477	4915	4158	3343	2571	1901	1334	23,699

   (a) Using the combined data, find the p-value for testing the hypothesis that an offspring is equally likely to male and female.

   (b) Using the combined data, construct a 90% confidence interval for the probability that a child will be born female.

   (c) Repeat part (a) using only the data for fifth-born children.

   (d) Repeat part (b) using only the data for fifth-born children.

8. Generate 100 random variables following a binomial distribution with $n = 39$ and $p = 0.7$.

   (a) What fraction of the random variables are equal to 27?

   (b) From the binomial distribution, what is $P(X = 27)$?

   (c) What fraction of the random variables are less than or equal to 27?

   (d) From the binomial distribution, what is $P(X \leq 27)$?

   (e) Repeat (a) and (c), only this time generate 10,000 random variables.

9. Consider the function $f(x) = xe^x$. Solve for the value of x where $f(x) = 2$. A numerical solution is acceptable; if you manage to obtain a closed-form algebraic solution, you deserve great fame and fortune.