Introduction to Biostatistics (171:161)
Breheny

Assignment 8
Due: Tuesday, April 1

1. According to an article in the *American Journal of Public Health*, the distribution of birth weights in a certain population is approximately normal with mean 3500 grams and standard deviation 430 grams.

   (a) What is the probability that a newborn's weight will be less than 3200 grams?

   (b) Suppose we take a sample of 9 newborns. What is the probability that their average weight will be less than 3200 grams?

   (c) In the aforementioned sample of 9 newborns, how many newborns would you expect to weigh under 3200 grams?

   (d) What is the probability that our sample of 9 newborns will contain exactly 3 newborns who weigh less than 3200 grams?

   (e) Suppose we take 5 samples of 9 newborns. What is the probability that at least one of the sample averages will be less than 3200 grams?

   (f) How large must our sample be in order to ensure a 95% probability that the sample mean will be within 50 grams of the population mean?

2. ⌨ In lab, we looked at random 25-person subsamples of the women in the NHANES study. On the course website is a data set called `nhanes-subsamples.txt`. Each column of the data set contains 1,000 sample means of triglyceride levels calculated from 1,000 randomly drawn subsamples. For each column, however, the number of women in those subsamples was different. For each of the three columns (A, B, and C), how large was the sample size? All of the sample sizes I used are multiples of ten, so please select your answers from $\{10, 20, 30, \ldots\}$. Also, please describe how you came to your answers.

3. ⌨ / ✎ A common test for association between a trait and a gene in genetics is the transmission disequilibrium test. It relies on finding (i.e., sampling) parent-child pairs in which the child has the trait of interest and the parent is heterozygous for the gene of interest (i.e., has one copy of each version of the gene). The parent is equally likely to pass on either copy, so if there is no link between the trait and the gene, we would expect 50% of the children to have version "A" and the other 50% to have version "B". However, because we have systematically sampled only children with the trait (any children without the trait are not included in the study), if version "A" causes a child to be more likely to develop the trait of interest, we would expect to find a higher proportion of version "A" in the children than version "B". In other words, if the two are associated, the "transmission" of the gene is distorted away from "equilibrium", hence the name of the test.

   In a 1989 study reported in the journal *Genetic epidemiology*, data was collected for 124 such parent-child pairs, in which the offspring had Type I diabetes and the parent was heterozygous for 5'FP (a flanking polymorphism adjacent to the insulin gene on chromosome 11). Among the children, 78 received the "class 1" version of 5'FP from their parent, while the other 46 did not.

(a) Carry out the transmission disequilibrium test based on the exact distribution of the number of offspring who received the "class 1" version of 5'FP. What is the probability of seeing a distortion as extreme or more extreme than the 78/46 split we saw in the data, if there really is no link between 5'FP and Type I diabetes?

(b) Carry out the transmission disequilibrium test based on the central limit theorem approximation for the number of offspring who received the "class 1" version of 5'FP. As in part (a), what is the probability of seeing a distortion as extreme or more extreme than the one observed if 5'FP and Type I diabetes truly are independent?

(c) In the *Genetic epidemiology* article, the authors report the results from the approximate test in (b) and state that "an exact binomial test can be used, if desired, instead". Expand on their advice. Does the difference between the approximate and exact test matter in this study? If not, when might it matter?

4. In assignment 6, we analyzed data from a crossover study in which 12 out of 14 individuals had lower cholesterol on an oat bran diet than a corn flake diet.

(a) Conduct a z-test of the null hypothesis that eating an oat bran diet has no effect on cholesterol (compared to a corn flake diet).

(b) Construct a 95% confidence level based on the central limit theorem approximation for the proportion of people who would lower their cholesterol by switching from a corn flake diet to an oat bran diet.

(c) Compare your answers for (a)-(b) to the exact results you got in assignment 6. Why don't the approximate/exact results agree as well here as they did in Problem 3 (the transmission disequilibrium test)?

5. (a) For the Student's curve with 5 degrees of freedom, find the area outside $\pm$ 2

(b) For the Student's curve with 15 degrees of freedom, find the values that contain the middle 95% of the area.

6. True or false:

(a) To create a confidence interval for a sample with 5 observations, we would use the Student's curve with 5 degrees of freedom.

(b) The area outside $\pm$ 2 for the Student's curve with 10 degrees of freedom will be larger than the area outside $\pm$ 2 for the Student's curve with 50 degrees of freedom.

(c) The z-test and t-test will be in closer agreement when n is small than when n is large.

(d) When performing a t-test, I need to worry more about the distribution of the data when n is small than when n is large.

7. (a) A paper in *Pediatrics* reported on a sample of 10 infants receiving a certain type of antacid to treat digestive disorders. The antacids contained aluminum, and physicians were concerned about the levels of aluminum in the plasma of these infants. In the sample, the mean aluminum level was 37.2 μg/l, with standard deviation 7.13 μg/l. Calculate a 95% confidence interval for the average plasma aluminum level of infants taking these antacids.

(b) The average plasma level of aluminum in the general population of infants is between 4 and 5 μg/l. What would you conclude about the levels of aluminum in infants taking these antacids versus the average levels in healthy infants?