Introduction to Biostatistics (BIOS:4120)
Breheny

Assignment 3
Due: Tuesday, February 15

1. ⌨ The course web page contains data from a British population health study carried out in Whickham from 1972-1974. Twenty years later, a follow-up study was conducted. The data set contains the smoking status and age (categorized for the sake of simplicity) of 1,314 women from the original study, along with whether the woman was alive or not when the twenty-year follow-up was conducted.

   (a) What percent of smokers were alive 20 years later?

   (b) What percent of nonsmokers were alive 20 years later?

   (c) Based on your answers from (a) and (b), does smoking appear to be beneficial or harmful?

   (d) Based on this sample, describe the relationship between smoking and age in the early 1970s in Whickham.

   (e) Based on your answers from (c) and (d), describe how a confounding relationship might be present.

   (f) If you adjust for the possible confounder of age by carrying out comparisons in more homogeneous subgroups, does smoking appear beneficial or harmful (back up your conclusion with numbers).

   (g) Make a bar chart depicting the number of women who were/were not alive vs. smoking status.

   (h) Make a bar chart depicting the number of women who were/were not alive vs. smoking status, conditioning on age.

2. (a) Ten people in a room have an average height of 5 feet, 6 inches. An 11th person, who is 6 feet, 6 inches, enters the room. Find the average height of all 11 people.

   (b) Fifty people in a room have an average height of 5 feet, 6 inches. A 51st person, who is 6 feet, 6 inches, enters the room. Find the average height of all 51 people. Compare this answer with your answer from (a).

   (c) In part (b), how tall would the 51st person have to have been in order to raise the average height by an inch?

3. (a) A study is conducted concerning the age of college freshmen. Would you expect the standard deviation to be about a month, about a year, or about five years?

   (b) A study is conducted concerning the income of adults in the United States. Would you expect the mean and the median incomes to be the same? Why or why not?

   (c) A study is conducted concerning the GPA of graduate students. Would you expect the correlation between students' undergraduate GPA and their graduate GPA to be closest to -1, -0.5, 0, 0.5, or 1?

4. (a) Suppose the governor of Iowa proposes to cut the salary of all state employees by $100 per month. What will this do to the average monthly salary of state employees?

   (b) What will it do to the standard deviation?

(c) What will the correlation be between the pre- and post-cut salaries?

(d) *Extra credit:* Suppose instead that the governor proposes to cut salaries by 5%. How does that change your answers to (a)-(c)? Be specific (*i.e.*, instead of "it will go down", answer with "it will go down by 10%" or whatever the effect happens to be).

5. ⌨ The course web page contains a data set concerning the effects of lead exposure on the neurological development of children for a group of children in El Paso, Texas. The data set consists of 57 children who lived within 1 mile of a lead smelter and 67 children who lived at least 1 mile away from the smelter. All children were given an IQ test and the results recorded.

(a) Make a table listing the mean, median, and standard deviation for each group.

(b) Create a histogram for each group.

(c) Create box plots comparing the distribution of IQ between the two groups. Does living close to the smelter appear to be harmful to the neurological development of children?