

Statistics 111 Homework 2 - Due in recitation on Friday, October 3rd

1. IPS exercise 2.20, p. 98 [Foot Deformities] (Data on website: `foot.txt`)
2. IPS exercise 2.38, p. 106 [Fidelity Funds] (Data on website: `fidelity.txt`)
3. IPS exercise 2.68, p. 123 [Foot Deformities II] (Data on website: `foot.txt`)
Note for this question: answer only parts (a) and (b)
4. IPS exercise 2.136, p. 161 [Fruits/Smoking] (Data on website: `smokefruit.txt`)
5. IPS exercise 2.148, p. 165 [Running Speed].

Note for this question: do not answer part (c) as given in book, but rather answer the following question:

- (c) Would you recommend using the regression line from part (b) for prediction of running speed? Why or why not?

6. In a survey of 988 men aged 18-24, the regression equation for predicting weight from height was:

$$\text{weight (in pounds)} = -167 + (4.7) (\text{height in inches})$$

What is the intercept and what is its interpretation? In this example, does the intercept have a meaningful interpretation? What is the slope and what is its interpretation?

7. Based on the regression equation in the previous question, what is the predicted weight of a newborn infant boy measuring 14 inches? Comment on the appropriateness of this prediction based on the regression equation.
8. A designer of a new automobile is concerned about gasoline mileage. Data is gathered (on website in the file `cardata.txt`) about the gasoline mileage for current car models along with several other variables. Make three scatterplots, all with *MPG City* as the Y variable but with three different X variables: *weight*, *horsepower* and *displacement*. Which among these three X variables appears to have the strongest relationship with *MPG City*? Why?
9. In a study of high school students, a correlation of .34 was found between hours spent per week doing homework and scores on standardized aptitude tests. For *each* of the following statements, explain *briefly* whether it is supported by this data.
 - a. Students who do well on standardized tests generally tend to study more.
 - b. Students who study more generally tend to do better on standardized tests.
 - c. If you study more, you will do better on standardized tests.
 - d. 34% of student's performance on standardized tests can be attributed to the time he or she spends doing homework.
 - e. There is a strong relationship between hours spent per week doing homework and scores on standardized aptitude tests.

10. The dataset `csdata.txt` contains education information on 224 computer science students. The columns of interest are GPA (grade point average), SATM (SAT mathematics score) and GENDER (male=1, female=2). The other columns are not of interest for this problem. We are interested in comparing the SAT mathematics scores and grade point averages of female students with those of male students. Make two sets of side-by-side boxplots to carry out these comparisons and write a brief (one or two sentences) discussion of the male-female comparisons.