

Statistics 111 Homework 2 - Due in recitation on Friday, February 15th

1. IPS exercise 2.12, p. 118 [Foot Deformities] (Data on website: `foot.txt`)
2. IPS exercise 2.26, p. 129 [Fidelity Funds] (Data on website: `fidelity.txt`)
3. IPS exercise 2.36, p. 131 [Investment]

4. CANCELLED

5. IPS exercise 2.106, p. 186 [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. IPS exercise 2.112, p. 188 [Regression to the mean]
7. 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?

8. 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.
9. 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?
10. With the data from the previous question, calculate the correlation between *MPG City* and each of the three variables: *weight*, *horsepower* and *displacement*. Generally, what does the correlation measure?

11. 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. CANCELLED**
- e. There is a strong relationship between hours spent per week doing homework and scores on standardized aptitude tests.

12. The dataset `csdata.txt` contains education information on 224 computer science students (described in the book's data appendix on page D-2). 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.