Homework #4

due Wednesday, October 24 at noon to my office

Instructions: Complete all problems and turn in a set of answers either to me or under my door (office 335L) by the assigned due date. Do ask me questions via email or in my office hours. Do work together. Do not copy answers from another student or turn in answers that are substantively identical. To clarify, if you work with another student, I would expect that your numerical answers would be quite close, but that your verbal explanations would be similar, but not identical, reflecting that you wrote up your answers independently. Show your work, and write out explanations for your answers. If you use Excel or a similar tool, write "According to Excel, ..." in your answer.

Problem 1 Fire damage in the United States amounts to billions of dollars each year, much of it insured. The time taken for firefighters to arrive at the fire is critical. This raises the question, should insurance companies lower premiums if the home to beinsured is close to a fire station? To help make a decision, a study was undertaken wherein a number of fires were investigated. The distance to the nearest fire station (in miles) and the percentage of fire damage were recorded (dataset Xr16-11). Determine the least squares regression line and interpret the coefficients.

Problem 2 An economist wanted to investigate the relationship between office rents (the dependent variable) and vacancy rates. Accordingly, he took a random sample of monthly office rents and the percentage of vacant office space in 30 different cities (dataset Xr16-16).

a. Determine the regression line.

b. Interpret the coefficients

Problem 3 Answer the following questions related to the 2008 General Social Survey (available on my website as the GSS2008 dataset).

a. Conduct an analysis of the relationship between income (INCOME) and age (AGE). Estimate with 95% confidence the average increase in income for each additional year of age. (hint: you may need to drop observations from the dataset for which you are missing information on either income or age)

b. Is there sufficient evidence to conclude that more educated people (EDUC) watch less television (TVHOURS)?

c. What does the dataset say about the relationship between education and income?

Problem 4 A computer dating service typically asks for various pieces of information such as height, weight, and income. One such service requests the length of index fingers. The only plausible reason for this request is to act as a proxy on height. Women in particular have often complained that men lie about their heights. If there is a strong relationship between heights and index fingers, the information can be used to "correct the false claims about heights. To test the relationship between the two variables, researchers gathered the heights and lengths of index fingers (in centimeters) of 121 students (dataset Xr16-107).

a. Using a computer, draw a scatter plot depicting the relationship between the two variables.

b. Is there sufficient evidence to infer that height and length of index fingers are linearly related?

c. Predict with 95% confidence the marginal increase in height associated with a 1cm increase in index finger length.

Problem 5 The admissions officer of a university is trying to decide which students to admit to the university. She believes that determinants of success include the standard variables — high school grades and SAT scores. However, she also believes that students who have participated in extracurricular activities are more likely to succeed than those who have not. To investigate this issue, she randomly sampled 100 fourth-year students and recorded the following variables (dataset Xr17-06):

- GPA for the first three years at the university (range is 0 to 12)
- GPA from high school (range is from 0 to 12)
- SAT score (range is 400 to 1600)
- Average number of hoursspent per week on extracurricular activities in the last year of high school

a. Develop a model that helps the admissions officer decide which students to admit and use the computer to estimates the coefficients in your model.

b. What is the coefficient of determination? Interpret its value.

c. Test the overall validity of the model (hint: see page 699. This is easy, but we did not cover it in class).

d. Test to determine whether each of the independent variables is linearly related to the dependent variable in the model.

Problem 6 Lotteries have become important sources of revenue for governments. Many people have criticized lotteries, however, referring to them as a rax on the poor and uneducated. In an examination of the issue, a random sample of 101 adults was asked how much they spend on lottery tickets and was interviewed about various socioeconomic variables. The purpose of the study is to test the following beliefs:

1. Relatively uneducated people spend more on lotteries than do relatively educated people
2. Older people buy more lottery tickets than do younger people
3. People with children spend more on lotteries than people with fewer children
4. Relatively poor people spend a greater proportion of their income on lotteries than do relatively rich people

The following data were recorded (dataset Xr17-13):

- Amount spent on lottery tickets as a percentage of total household income
- Number of years of education
- Age
- Number of children
- Personal income (in thousands of dollars)

a. Develop a multiple regression equation, formulated to test the beliefs enumerated above.

b. Estimate the coefficients of your equation using statistical software.

c. Is the model valid?

d. Test each of the beliefs 1-4 above. What conclusions can you draw?