228:105, Statistical Methods and Computing
Instructor: Costas
Spring 2004, Homework 3

Due: Fri, 02/13 in class
Please put your name at the top of your homework, and list the names of any classmates
with whom you collaborated.

1. For textbook problem 4.27, the data are on the course web page as “winefat.dat.”
   (a) Use SAS to produce the scatterplot for part (a).
   (b) Answer parts (b) and (c).
   (c) Use SAS to compute the correlation coefficient \( r \) between alcohol from wine
       and head-disease death rate.
   (d) In the data, wine consumption is measured as yearly liters of alcohol from drink-
       ing wine, per person. If it had instead been measured in yearly ounces per person,
       how would the value of \( r \) have been affected?

2. Investors ask about the relationship between returns on investments in the U.S. and
   on investments overseas. The data file “stockdat” gives the total returns on U.S.
   and overseas common stocks over a 26 year period. (The total return is change in
   price plus any dividends paid, converted into U.S. dollars. Both returns are averages
   over many individual stocks. Use SAS to get the numbers required for answering
   the following questions.
   (a) Make a scatterplot suitable for predicting overseas returns from U.S. returns
       (Use SAS).
   (b) Find the correlation and \( r^2 \). Describe the relationship between U.S. and overseas
       returns in words, using \( r \) and \( r^2 \) to make your description more precise.
   (c) Find the least squares regression line of overseas returns on U.S. returns. Show
       this line superimposed on the scatterplot.
   (d) In 1997, the return on U.S. stocks was 33.4%. Use the regression line to predict
       the return on overseas stocks. (You may either calculate this by hand or use
       SAS output.) The actual overseas return was 2.1%. Are you confident that
       predictions using the regression line will be quite accurate? Why?
   (e) Circle the point that has the largest residual (either positive or negative). What
       year is this? Are there any points that seem likely to be very influential?
   (f) Find the 25th, median, 75th summaries for both U.S. and overseas returns. Make boxplots
       of data to compare the two distributions.
   (g) Were returns generally higher in the U.S. or overseas during this period? Explain
       your answer.
   (h) Were returns more volatile (that is, more variable) in the U.S. or overseas during
       this period? Explain your answer.

3. Textbook problem 5.10,
   Use SAS, enter the data yourself as part of the data step using the “data=…” state-
   ment. Note that you need a semicolon after “data=…” and another semicolon by
   itself on the line after the last row in the list of data. Answer all parts of the question,
   and include relevant SAS output.

   data shop ;
   input year pop ;
data lines ;
   1935 22.1
   1940 30.6
   1945 24.4
   1950 23.0
   1955 19.9
   1960 16.6
   1965 12.4
   1970 9.7
   1975 8.9
   1980 7.2
   ;

4. Textbook problems 4.19, 4.31, 4.32, 5.12, 5.25, 5.32