

22S:105, Statistical Methods and Computing
 Instructor: Cowles
 Spring 2006, Homework 4

Due: Fri., 02/17 in class

Please put your name at the top of your homework, and list the names of any classmates with whom you collaborated.

1. Investors ask about the relationship between returns on investments in the U.S. and on investments overseas. The data file "stocks.dat" 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) Find the least-squares regression equation of overseas returns on U.S. returns.
 - (b) Produce a scatterplot with the regression line superimposed on it.
 - (c) Produce a residual plot.
 - (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) In both of your plots, circle the point that has the largest residual (either positive or negative). What year is this?
 - (f) Are there any points that seem likely to be very influential? If so, identify them in your scatterplot and residual plot.

2. Textbook problem 5.10.

Use SAS. Enter the data yourself as part of the data step using the "datalines" statement as shown below. Note that you need a semicolon after "datalines" 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 farm ;
input year pop ;
datalines ;
1935 32.1
1940 30.5
1945 24.4
1950 23.0
1955 19.1
1960 15.6
1965 12.4
1970 9.7
```

```
1975 8.9
1980 7.2
;
```

3. Textbook problems: 5.12, 5.25, 5.32 I.40, I.44, 7.3, 7.11, 7.18, 7.20, 7.41