Case-based Social Statistics II
CSSS 322
Professor: Mark S. Handcock

Homework 6

Due Friday, May 24, 2002

Problems to be handed in:

1) Suppose the regression model:

$$Y = \beta_0 + \beta_1 X$$

holds with $\beta_0 = 100, \beta_1 = 20,$ and $\sigma^2 = 25$. An observation of Y is made at $X = 5$.

a) Suppose the regression model holds without the normality assumption on the error term ϵ_i. This was assumption (iv) in the notes for Lecture 4.0. Can you state the exact probability that Y will fall between 195 and 205? Explain briefly.

b) Suppose now that the regression model holds with the normality assumption on the error term ϵ_i. Can you now state the exact probability that Y will fall between 195 and 205? If so state it.

2) A junk mail recycling company wants to promote its service. Suppose that four levels of direct mail advertisements were randomly assigned to four similar residential blocks in Greenwich village, resulting in the following people signing up for the companies service:

Advertisements (mailings/year)	Signups (signups per 100 residents/year)
1	70
2	70
4	80
5	100

a) Calculate, by hand, the regression line of the signups against mailings

b) Graph the four points and the regression line. Check that the line fits the data reasonably well.

c) Use the regression line to predict:

i) The signups if 3 mailings/year (mailings per year) were sent.

ii) The signups if 4 mailings/year were sent.

iii) The increase in signups for every 1 mailing/year increase in mailings.

Show these three on the graph.

3) A HeadStart type program aims to improve the job-prospects of participants. In seeking to determine just how influential the HeadStart training is, the review committee of a program has collected data over the previous year on the weekly income on participants, and the number of months they were trained within the program. The data are given below:

TRAINING	INCOME
3.0	50
5.0	250
7.0	700
6.0	450
6.5	600
8.0	1000
3.5	75
4.0	150
4.5	200
6.5	550
7.0	750
7.5	800
7.5	900
8.5	1100
7.0	600

a) The data are available on the course website under "Data".

b) Using the "Scatterplot" command in Datatools, plot the data. Does it appear that the length of training and income are linearly related?

c) Below is fit of the linear regression model of income based on training time.

```
LEAST SQUARES LINEAR REGRESSION OF INCOME  Income (thousands per month)

PREDICTOR
VARIABLES    COEFFICIENT    STD ERROR     STUDENT'S T       P
---------    -----------    ---------     -----------    ------
CONSTANT       -644.951       72.8973        -8.85       0.0000
TRAINING        195.074       11.5381        16.91       0.0000

R-SQUARED           0.9565      RESIDUAL MEAN SQUARE (MSE)    5404.98
ADJUSTED R-SQUARED  0.9532      STANDARD ERROR OF ESTIMATE    73.5186

SOURCE        DF       SS           MS          F        P
----------   ---   ----------   ----------   -----   ------
REGRESSION     1   1.545E+06    1.545E+06   285.84   0.0000
RESIDUAL      13     70264.8      5404.98
TOTAL         14   1.615E+06

CASES INCLUDED 15   MISSING CASES 0
```

Fit the model in DataTools using the "Simple Linear Regression" command and hand in a copy of the output. You can use the above to make sure you are doing it right.

d) Give an economic interpretation of the coefficient of training, $\widehat{\beta}_1$, in the model.

e) If the sign of the slope were negative, what would that say about the relationship between training and sales?

f) Give an economic interpretation of the intercept, $\widehat{\beta}_0$, in the model. What does the value of the intercept tell us?

g) Calculate a 99% confidence interval for β_0. Do you think that there will be no income if there was no training?

h) Calculate a 95% confidence interval for β_1.