

## Case-based Social Statistics II

CSSS 321

Professor: Mark S. Handcock

### Homework 8

Due Friday, June 6, 2002

*Problems to be handed in:*

- 1) This question returns to the HeadStart type program considered in Homework 6. This 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 available on the course website under “Data”.

Using the “Simple Linear Regression” command in Datatools, refit the model for income in terms of the length of training. Click the button that saves the residuals (into the variable `var3`). Using the “Scatter Plot” command, plot the residuals against the length of training.

Interpret the Scatterplot. Do the assumptions of linear regression appear to hold here?

- 2) You run an apple farm in Yakima and obtain the following data containing information on crop yields, total rainfall and average temperature for this arid region of Washington over eight years.

variable	units	value							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
time	year	1992	1993	1994	1995	1996	1997	1999	1999
yield	bushels/acre	60	50	70	70	80	50	60	40
rainfall	inches	8	10	11	10	9	9	12	11
temperature	fahrenheit	56	47	53	53	56	47	44	44

The data are available on the course website under “Data”.

- a) Using the “Scatterplot” command in Datatools, plot the yield against rainfall, and yield against temperature. Does it appear that these variables are linearly related?
- b) Using the “Multiple Linear Regression” command, fit a regression model for yield based on rainfall. Give an interpretation of the slope. Does it represent an intuitively reasonable model for the relationship between rainfall and yield. Why?
- c) Using the “Multiple Linear Regression” command, regress yield on both rainfall and temperature. That is, fit the model

$$\text{YIELD} = \text{CONSTANT} + \text{RAINFALL} + \text{TEMP}$$

Note that the sign of the estimate for the slope of rainfall is opposite to that of (a).

The vice-president of the company ask you to explain this phenomenon in simple terms. What do you say?

- 3) If the only possible values of a explanatory variable (independent variable) are 1 and 0, then that variable is called a “indicator.” The '1' usually represents the presence of some attribute and '0' represents the absence of the attribute. The regression coefficient in such a case is usually interpreted as the added response associated with the presence of the attribute. In this problem we will explore the use of a indicator variable.

The set of values listed below are the price-per-pound (PRICE) for twenty beef steers. Also given are the actual total weight (SIZE) and the USDA grade (PRIME). This are coded so that PRIME=0 for “choice grade” steers and PRIME=1 for “prime grade.” The data are available on the course website under “Data”.

PRICE	SIZE	PRIME
-----	-----	-----
36.75	736	0
35.50	506	0
28.00	394	0
31.00	419	0
32.75	504	0
35.25	599	0
31.50	442	0
33.50	505	0
32.50	512	0
31.25	400	0
33.50	339	1
36.00	517	1
33.00	396	1
30.50	382	1
36.00	479	1
35.25	515	1
37.50	606	1
30.50	336	1
33.00	425	1
34.75	498	1

- Find the regression equation for PRICE on both SIZE and PRIME. Note the regression coefficient of PRIME and give an interpretation.
- Calculate a 95% confidence interval for the price-per-pound difference between choice and prime steaks of identical weight
- Test the hypothesis, 99% confidence, that there is no price-per-pound difference between choice and prime steaks of identical weight