Mathematics 376 – Probability and Statistics 2
Lab Project 4/Problem Set 9 – Regression Case-Studies
April 27 and 29, 2010

Background

In class, we have discussed the basic methods for obtaining the least-squares estimators for the coefficients β_i in simple linear models

$$(1) \qquad Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon$$

and techniques for hypothesis testing concerning the β_i (see page 585 for the case $k = 1$ and models of the form $Y = \beta_0 + \beta_1 x + \epsilon$, and page 618 in the text for the general case).

In this lab project, we will work through some "case-studies" of how these methods might be used.

From the Text

- Do problems 11.24 and 11.26 (using Maple is recommended!)

Additional Problem – The "Story"

Utility companies, which must plan the operation and expansion of electricity generating facilities, are vitally interested in predicting customer demand over both short and long periods of time. A short-term study was conducted to investigate the effect of

- x_1 = each month's *daily mean temperature, in degrees Fahrenheit*, and
- x_2 = the *cost per kilowatt-hour (in dollars)* per household.

The company officials expected the demand for electricity to rise in cold weather (due to heating), fall when the weather was moderate, then increase again in hot weather (due to air conditioning). They also expected demand to decrease as the cost increased. Because of the expectation about the dependence on x_1, a model just involving x_1 to the first power was thought to be inappropriate, so they set up a model of the form:

$$(1) \qquad Y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_2 x_1 + \beta_5 x_2 x_1^2 + \varepsilon$$

The data they had to work with was as follows:

- With $x_2 = .08$ (i.e. 8 cents per kilowatt-hour):

x_1 (*temp*)	31	34	39	42	47	56	62	66	68	71	75	78
y (*demand*)	55	49	46	47	40	43	41	46	44	51	62	73

- With $x_2 = .1$ (i.e. 10 cents per kilowatt-hour):

x_1 (*temp*)	32	36	39	42	48	56	62	66	68	72	75	79
y (*demand*)	50	44	42	42	38	40	39	44	40	44	50	55

(Note: The data were from different locations and months so the temperatures don't match up exactly.)

A) Start by computing the least squares estimators for the coefficients β_i in the model (1).
B) What does the model predict about the demand in a month with average temperature 50 degrees Fahrenheit, if the cost is .12 (12 cents per kilowatt-hour)? Report an estimated value, and a 95% confidence interval. (See Section 11.13 in the text and the notes from 4/26 for the needed formulas.)
C) Now, we want to consider a different question. Namely: *Did including the x_2 terms in (1) really give us any added predictive power?* Think of (1) as the *complete* model as in class on 4/26, and consider the *reduced model*

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \varepsilon \qquad (2)$$

where the x_2 terms have been removed (i.e. $\beta_3 = \beta_4 = \beta_5 = 0$). Is there sufficient evidence to indicate that the model (1) gives us a significantly better fit to the data? Test using the procedure outlined in class on 4/26 and in section 11.14.

Assignment

One Maple worksheet from each person. Due in class on Monday, May 3.