Background

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

\begin{equation}
Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon
\end{equation}

and techniques for hypothesis testing concerning the $\beta_i$ (see pages 553, 587 in the text for convenient summaries). In this lab project, we will work through a “case-study” of how these methods might be used in a realistic statistical study.

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 each month’s daily mean temperature, in degrees Fahrenheit ($x_1$) and the cost per kilowatt-hour (in dollars) ($x_1$) 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:

\begin{equation}
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_2^2 + \epsilon
\end{equation}

The data they had to work with was as follows:

- With $x_2 = .08$ (i.e. 8 cents per kilowatt-hour):
  \[
  \begin{align*}
  x_1 \text{ (temp)} & \quad 31 \quad 34 \quad 39 \quad 42 \quad 47 \quad 56 \quad 62 \quad 66 \quad 68 \quad 71 \quad 75 \quad 78 \\
  y \text{ (demand)} & \quad 55 \quad 49 \quad 46 \quad 47 \quad 40 \quad 43 \quad 41 \quad 46 \quad 44 \quad 51 \quad 62 \quad 73
  \end{align*}
  \]

- With $x_2 = .1$ (i.e. 10 cents per kilowatt-hour):
  \[
  \begin{align*}
  x_1 \text{ (temp)} & \quad 32 \quad 36 \quad 39 \quad 42 \quad 48 \quad 56 \quad 62 \quad 66 \quad 68 \quad 72 \quad 75 \quad 79 \\
  y \text{ (demand)} & \quad 50 \quad 44 \quad 42 \quad 42 \quad 38 \quad 40 \quad 39 \quad 44 \quad 40 \quad 44 \quad 50 \quad 55
  \end{align*}
  \]

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

A) Start by computing the least squares estimators for the coefficients $\beta_i$ 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 11.12 in the text and the notes from 4/19 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 4/24, and consider the reduced model

\[(2) \quad Y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \varepsilon\]

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 4/24 and in section 11.14.

Assignment

One Maple worksheet from each lab group. Due by the end of the day on Friday, April 28 in Swords 335.