Lecture # 37

Prof. John W. Sutherland

Nov. 28, 2005
Linear Regression

![Linear Regression graph]

- Dependent variable (y)
- Independent variable (x)
- Scatter plot with data points
**Modeling**

To describe the data above, propose the model:

\[ y = B_0 + B_1 x + \varepsilon \]

Fitted model will then be

\[ \hat{y} = b_0 + b_1 x \]

Want to select values for \( b_0 \) & \( b_1 \) that minimize

\[ n = 6 \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]
Modeling (Cont.)

Define

\[ S(b_0, b_1) = \sum_{i=1}^{n=6} (y_i - \hat{y}_i)^2 \]

the model residual Sum of Squares.

Minimize

\[ S(b_0, b_1) = \sum_{i=1}^{n=6} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n=6} (y_i - b_0 - b_1 x_1)^2 \]
Modeling (Cont.)

To find minimum $S$, take partial derivatives of $S$ with respect to $b_0$ & $b_1$, set these equal to zero, and solve for $b_0$ & $b_1$

$$\frac{\partial}{\partial b_0} S(b_0, b_1) = 2\sum(y_i - b_0 - b_1x_i)(-1) = 0$$

$$\frac{\partial}{\partial b_1} S(b_0, b_1) = 2\sum(y_i - b_0 - b_1x_i)(-x_i) = 0$$
Modeling (Cont.)

\[-\sum y_i + \sum b_o + \sum b_1 x_i = 0\]

\[-\sum x_i y_i + \sum b_o x_i + \sum b_1 x_i^2 = 0\]

Simplifying, we obtain:

\[n b_0 + b_1 \sum x_i = \sum y_i\]

\[b_0 \sum x_i + b_1 \sum x_i^2 = \sum x_i y_i\]
Modeling (Cont.)

These two equations are known as “Normal Equations”.

The values of $b_0$ & $b_1$ that satisfy the Normal Equations are the least squares estimates -- they are the values that give a minimum $S$. 
Matrix Form

\[
\begin{bmatrix}
N & \sum x_i \\
\sum x_i & \sum x_i^2 
\end{bmatrix}
\begin{bmatrix}
b_0 \\
b_1 
\end{bmatrix} =
\begin{bmatrix}
\sum y_i \\
\sum x_i y_i 
\end{bmatrix}
\]

\[
\begin{bmatrix}
b_0^* \\
b_1^* 
\end{bmatrix} = \text{Least Squares Estimates}
\]

\[
= \begin{bmatrix}
n & \sum x \\
\sum x & \sum x^2 
\end{bmatrix}^{-1}
\begin{bmatrix}
\sum y \\
\sum xy 
\end{bmatrix}
\]
Matrix Form (Cont.)

\[
\begin{cases}
  b_0^* \\
  b_1^*
\end{cases}
= \frac{1}{n \Sigma x^2 - (\Sigma x)^2} \begin{bmatrix}
  \Sigma x^2 & -\Sigma x \\
  -\Sigma x & n
\end{bmatrix} \begin{bmatrix}
  \Sigma y \\
  \Sigma xy
\end{bmatrix}
\]

\[
= \frac{1}{n \Sigma x^2 - (\Sigma x)^2} \begin{bmatrix}
  \Sigma x^2 \Sigma y - \Sigma x \Sigma xy \\
  -\Sigma x \Sigma y + n \Sigma xy
\end{bmatrix}
\]
Matrix Form (Cont.)

\[
\begin{bmatrix} b_0^* \\ b_1^* \end{bmatrix} = \begin{bmatrix} 1.4667 \\ 0.9143 \end{bmatrix}
\]

\(b_0^*, b_1^*\) are the values of \(b_0 \& b_1\) that minimize \(S\), the Residual Sum of Squares.

\(b_0^* = \hat{B}_0 = \) an estimate of \(B_0\)

\(b_1^* = \hat{B}_1 = \) an estimate of \(B_1\)
Fitted Line
Matrix Approach

\[ \vec{y} : \text{Vector of Observations} = \begin{bmatrix} 2 \\ 3 \\ 5 \\ 5 \\ 7 \\ 6 \end{bmatrix}, \]

\[ \vec{x} : \text{Matrix of Independent Variables, i.e.,} \]

\[ \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \end{bmatrix} \]

the Design Matrix =
Matrix Approach (Cont)

\( \hat{y} = \text{Vector of Predictions} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} = x\hat{b} \)

b coefficients = \[
\begin{bmatrix}
 b_0 \\
 b_1 
\end{bmatrix} \]
Matrix Approach (Cont.)

\[ e \approx \text{Vector of Prediction Errors} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \]

\[ e \approx y - \hat{y} \]

Want to Min \( e^T e \) or Min \( (y - \hat{y})^T (y - \hat{y}) \)
Matrix Approach (Cont.)

Take derivative with respect to b’s and set = 0

\[-\tilde{x}^T(\tilde{y} - \tilde{x}\tilde{z}) = 0 = -\tilde{x}^T\tilde{y} + (x^T x)\tilde{z}\]

\[(x^T x)\tilde{z} = x^T \tilde{y}\]
Matrix Approach (Cont.)

Therefore, 

\[
b \sim = (x^T x)^{-1} x^T y \]

It is analogous to

\[
\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma y \\ \Sigma xy \end{bmatrix}
\]
Matrix Approach (Cont.)

Re-run experiments several times

\[
\begin{bmatrix}
  b_0 \\
  b_1
\end{bmatrix} = \begin{bmatrix}
  1.4667 \\
  0.9143
\end{bmatrix},
\begin{bmatrix}
  b_0 \\
  b_1
\end{bmatrix} = \begin{bmatrix}
  1.5309 \\
  0.9741
\end{bmatrix},
\begin{bmatrix}
  b_0 \\
  b_1
\end{bmatrix} = \begin{bmatrix}
  1.5512 \\
  1.0134
\end{bmatrix}
\]

If true model is \( y = B_0 + B_1 x + \varepsilon \)

Then \( E(b_0) = B_0 \), \( E(b_1) = B_1 \), \( E[b] = \bar{B} \)
Matrix Approach (Cont.)

\[
Var(\hat{b}) = (\hat{x}^T \hat{x})^{-1} \sigma_y^2
\]

where, \( \sigma_y^2 \) describes the experimental error variation in the y’s (\( \sigma_\varepsilon^2 \)).
Matrix Approach (Cont.)

For our example, \( Var(b) = \begin{bmatrix} \text{Var}(b_0) & \text{Cov}(b_0, b_1) \\ \text{Cov}(b_0, b_1) & \text{Var}(b_1) \end{bmatrix} \).

If \( \sigma_y^2 \) (or \( \sigma_\varepsilon^2 \)) is unknown, we can estimate it with

\[
\hat{s}^2 = \frac{(y - \hat{y})^T (y - \hat{y})}{(n - \# \text{ of parameters})} = \frac{e^T e}{n - p} = \frac{S_{res}}{v}
\]
Matrix Approach (Cont.)

For the example,

\[ s_y^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n - 2} = 0.67619 \]

\[ \text{Var}(\hat{b}) = (x^T x)^{-1} s_y^2 = \begin{bmatrix} 0.586 & -0.135 \\ -0.135 & 0.039 \end{bmatrix} \]

\[ s_{b_0}^2 = 0.586, \quad s_{b_0} = 0.767 \quad \text{standard error of } b_0 \]

\[ s_{b_1}^2 = 0.039, \quad s_{b_1} = 0.197 \quad \text{standard error of } b_1 \]
Dynamic Systems

- Many processes have dynamic characteristics -- data are produced as a result of dynamic behavior within the process
  - Chemical processes
  - Vibrating systems
  - ?

Process Data, X’s
More on Dynamic Systems

- Because of our experience with differential equations and vibrations, we tend to think of dynamic behavior as in the figure below.
Common Cause Variability

- With the addition of process noise, however, we often see behavior like that below.
Time Series Analysis

• For situations like that shown in the previous figure, we can use time series analysis to extract information about the process.

• From a time series model we can “back out” information about the unknown underlying system dynamics.

• Simple autoregressive model [ AR(1) ]

\[ X_i = \phi X_{i-1} + a_i \]
Interpreting Phi

\[ \phi = 0.9 \]
Interpreting Phi

$\phi = -0.7$
Interpreting Phi

\[ \phi = 1.02 \]

Behavior is unstable because \( \phi > 1 \)
Let's say that we have a series of data such as that below.
Finding $\phi$

Let’s apply this model to data: $X_i = \phi X_{i-1} + a_i$

Same form as $y_i = \beta_1 x_i + \varepsilon_i$

We now know how to estimate the value for $\beta_1$ (called estimate $b_1$)

$$S = \sum_{i=1}^{n} (X_i - \hat{X}_i)^2 = \sum_{i=1}^{n} (X_i - \phi X_{i-1})^2$$

$$\frac{dS}{d\phi} = 2\sum(X_i - \phi X_{i-1})(-1) = 0$$
Finding $\phi$

$$\hat{\phi} = \frac{\sum_{i=2}^{n} X_i X_{i-1}}{\sum_{i=2}^{n} X_{i-1}^2}$$

For data shown previously,

$$\sum_{i=2}^{n} X_i X_{i-1} = 725.71$$
$$\sum_{i=2}^{n} X_{i-1}^2 = 1562.15$$

$$\hat{\phi} = 0.465$$

So, $\hat{X}_i = \hat{\phi} X_{i-1}$ -- also can be thought of as a forecast
Backshift Operator

Define backshift operator as

\[ X_{t-1} = BX_t \]

In general,

\[ X_{t-j} = B^j X_t \]

Previous equation: \( X_i = \phi X_{i-1} + a_i \), can be rewritten as

\[
\begin{align*}
X_i &= \phi X_{i-1} + a_i \\
X_i &= \phi BX_i + a_i \\
X_i(1 - \phi B) &= a_i \\
X_i &= a_i/(1 - \phi B)
\end{align*}
\]

-- denominator is characteristic eqn
More on AR(1)

\[ X_i = \phi X_{i-1} + a_i \], or since \( X_{i-1} = \phi X_{i-2} + a_{i-1} \)

\[ X_i = \phi(\phi X_{i-2} + a_{i-1}) + a_i \text{ or } X_i = \phi^2 X_{i-2} + \phi a_{i-1} + a_i \]

or

\[ X_i = \sum_{j=0}^{k} \phi^j a_{i-j} \text{ Note similar form to EWMA} \]
AR(2) Model

AR(2) model:

\[ X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + a_t \]

or

\[ X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = a_t \]

\[ a_t \sim \text{NID}(0, \sigma^2_a) \]