STAT 410 - Linear Regression
Lecture 5

Meng Li

Department of Statistics

Sep. 14, 2017
Multiple Regression Models

- Suppose that the yield in pounds of conversion in a chemical process depends on temperature $x_1$ and the catalyst concentration $x_2$.
- A multiple regression model that might describe this relationship is
  \[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon. \]  
  (1)
- This is a multiple linear regression model in two variables.
- In general, the multiple linear regression model with $k$ regressors is
  \[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon. \]  
  (2)
Examples of multiple regression models

- **Polynomial models:** \( y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_k x^k + \varepsilon \)
  - It becomes a multiple regression model if we let \( x_1 = x, x_2 = x^2, \ldots, x_k = x^k \).

- **Interaction effects:** \( y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon \)
  - It becomes a multiple regression model if we let \( x_3 = x_1 x_2 \) and \( \beta_3 = \beta_{12} \).

- **Nonlinear function with fixed basis expansion:** \( y = f(x) + \varepsilon \)
  - Where \( f(x) = \sum_{j=1}^{k} \beta_k \phi_k(x) \).
  - It becomes a multiple regression model if we let \( x_k = \phi_k(x) \).
  - There is a rich menu for \( \{ \phi_k(\cdot) : k \geq 1 \} \): wavelet basis, Fourier transformation, orthogonal polynomials, etc.

- In general, any regression model that is linear in the parameters \( \beta \)'s is a linear regression model, regardless of the shape of the surface that it generates. (\( \mathbf{V} \) and \( \mathbf{P} \) in SVP)
Model:

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon. \]  

(3)

Data: \((y_i; x_{i1}, \ldots, x_{ik})\) as shown in the above table.
- \(n\) — number of observations available
- \(k\) — number of regressor variables
- \(y_i\) — \(i\)th response or dependent variable
- \(x_{ij}\) — \(i\)th observation or level of regressor \(j\)

Sample regression model:

\[ y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \varepsilon_i. \]  

(4)
In matrix notation, model (4) becomes

\[ y = X \beta + \varepsilon, \quad (5) \]

where

\[
\begin{align*}
\mathbf{y} &= \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \\
\mathbf{X} &= \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}
\end{align*}
\]

\[
\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \\
\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}
\]

\[ \text{with dimensions: } \mathbf{y}, \mathbf{X}, \beta, \varepsilon \]
Least-squares estimator:

\[ \hat{\beta} = \arg\min_\beta \{ (y - X\beta)'(y - X\beta) \} = \arg\min_\beta \|y - X\beta\|^2. \]

The loss \( S(\beta) = (y - X\beta)'(y - X\beta) \) can be expressed as

\[ S(\beta) = y'y - \beta'X'y - y'X\beta + \beta'X'X\beta \]
\[ = y'y - 2\beta'X'y + \beta'X'X\beta. \]

The LS estimator satisfies that \( \frac{\partial S}{\partial \beta} = -2X'y + 2X'X\beta = 0. \)

This simplifies to

\[ X'X\hat{\beta} = X'y, \tag{6} \]

which are the so-called (least-squares) normal equations.

Thus, the LS estimator of \( \beta \) is

\[ \hat{\beta} = (X'X)^{-1}X'y, \tag{7} \]

provided that the inverse matrix \((X'X)^{-1}\) exists.
The dimension of $(X'X)$ is $(k + 1)$ by $(k + 1)$.

The inverse matrix $(X'X)^{-1}$ exists if the regressions $X$ are linearly independent, i.e., no column of $X$ is a linear combination of the other columns.

The vector of fitted values $\hat{y}_i$ corresponding to the observed values $y_i$ is

$$\hat{y} = X\hat{\beta} = X(X'X)^{-1}X'y.$$ 

The $n \times n$ matrix $H = X(X'X)^{-1}X'$ is called the hat matrix.

The residual vector can be conveniently written as

$$e = y - \hat{y} = y - Hy = (I - H)y.$$
The least squares fit is the **projection** of $y$ onto the span of $X$ (the estimation space), and the residual at the least squares solution is orthogonal to the span of $X$.

In the above figure, point $A$ denotes $y$, point $B$ is $X\beta$ for any $\beta$, and point $C$ is the least squares fit $X\hat{\beta}$.

The residual $e = y - \hat{y}$ is perpendicular to the span of $X$, i.e., $X'(y - X\hat{\beta}) = 0$ or $X'X\hat{\beta} = X'y$ — the normal equations.
Recall the model: \( y = X\beta + \epsilon \), where \( \epsilon_i \) is i.i.d. from a distribution that has mean 0 and variance \( \sigma^2 \).

- \( \hat{\beta} \) is unbiased, namely, \( E(\hat{\beta}) = \beta \).
- Variance matrix of \( \hat{\beta} \): \( \text{Var}(\hat{\beta}) = E\{(\hat{\beta} - E\hat{\beta})'(\hat{\beta} - E\hat{\beta})\} \).
- We can obtain that \( \text{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1} \).
- The LS estimator is the best linear unbiased estimator (BLUE) of \( \beta \) (the Gauss - Markov theorem).
- If we further assume \( \epsilon_i \)'s are normally distributed:
  - MLE is identical to LS estimator.
  - \( \hat{\beta} \) follows a **multivariate** normal distribution with mean \( \beta \) and covariance \( \sigma^2(X'X)^{-1} \).
- Similar to SLR, we estimate the variance component \( \sigma^2 \) by
  \[
  \hat{\sigma}^2 = \frac{SS_{res}}{n-p} = MS_{res},
  \]
  where \( p = k + 1 \) is the number of parameters in \( \beta \).
- \( \hat{\sigma}^2 \) is unbiased but is not the MLE.