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

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon. \tag{1}$$

- This is a multiple linear regression model in two variables.
- In general, the multiple linear regression model with k regressors is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon. \tag{2}$$

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 β's is a linear regression model, regardless of the shape of the surface that it generates. (**V** and **P** in SVP)

Data and Notation

		Regressors			
Observation, i	Response, y	x_1	x_2	$\dots$	x_k
1	y_1	x_{11}	x_{12}	$\dots$	x_{1k}
2	y_2	x_{21}	x_{22}	$\dots$	x_{2k}
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
n	y_n	x_{n1}	x_{n2}	$\dots$	x_{nk}

- Model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon. \tag{3}$$

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

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i. \tag{4}$$

Matrix notation

In matrix notation, model (4) becomes

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{5}$$

where

$$\mathbf{y} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{n\times 1}, \ \mathbf{X} = \underbrace{\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}}_{n\times(k+1)}, \ \boldsymbol{\beta} = \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}}_{(k+1)\times 1}, \ \boldsymbol{\varepsilon} = \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}}_{n\times 1}.$$

LS estimators

- Least-squares estimator:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 .$$

- The loss $S(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ can be expressed as

$$\begin{aligned} S(\boldsymbol{\beta}) &= \mathbf{y}'\mathbf{y} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}. \end{aligned}$$

- The LS estimator satisfies that $\frac{\partial S}{\partial \boldsymbol{\beta}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = 0.$
- This simplifies to

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}, \tag{6}$$

  which are the so-called (least-squares) *normal equations*.
- Thus, the LS estimator of $\boldsymbol{\beta}$ is

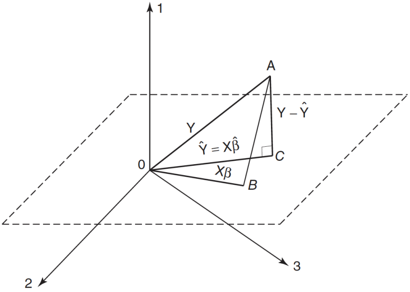
$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \tag{7}$$

  provided that the inverse matrix $(\mathbf{X}'\mathbf{X})^{-1}$ exists.

LS estimators

- The dimension of $(\mathbf{X}'\mathbf{X})$ is $(k+1)$ by $(k+1)$.
- The inverse matrix $(\mathbf{X}'\mathbf{X})^{-1}$ exists if the regressions $\mathbf{X}$ are linearly independent, i.e., no column of $\mathbf{X}$ is a linear combination of the other columns.
- The vector of fitted values $\hat{y}_i$ corresponding to the observed values y_i is
$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$
- The $n \times n$ matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is called the **hat matrix.**
- The residual vector can be conveniently written as
$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{H}\mathbf{y} = (\mathbf{I} - \mathbf{H})\mathbf{y}.$$

Geometric interpretation



- The least squares fit is the **projection** of $\mathbf{y}$ onto the span of $\mathbf{X}$ (the estimation space), and the residual at the least squares solution is orthogonal to the span of $\mathbf{X}$.
- In the above figure, point A denotes $\mathbf{y}$, point B is $\mathbf{X}\boldsymbol{\beta}$ for any $\boldsymbol{\beta}$, and point C is the least squares fit $\mathbf{X}\hat{\boldsymbol{\beta}}$.
- The residual $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$ is perpendicular to the span of $\mathbf{X}$, i.e., $\mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = 0$ or $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ — the normal equations.

Properties of LS estimator

Recall the model: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where ε_i is $i.i.d.$ from a distribution that has mean 0 and variance σ^2.

- $\hat{\boldsymbol{\beta}}$ is unbiased, namely, $\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$.
- Variance matrix of $\hat{\boldsymbol{\beta}}$: $\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \mathrm{E}\{(\hat{\boldsymbol{\beta}} - \mathrm{E}\hat{\boldsymbol{\beta}})'(\hat{\boldsymbol{\beta}} - \mathrm{E}\hat{\boldsymbol{\beta}})\}$.
- We can obtain that $\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.
- The LS estimator is the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ (the Gauss - Markov theorem).
- If we further assume ε_i's are normally distributed:
  - MLE is identical to LS estimator.
  - $\hat{\boldsymbol{\beta}}$ follows a **multivariate** normal distribution with mean $\boldsymbol{\beta}$ and covariance $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.
- Similar to SLR, we estimate the variance component σ^2 by

$$\widehat{\sigma^2} = \frac{SS_{res}}{n-p} = MS_{res},$$

  where $p = k+1$ is the number of parameters in $\boldsymbol{\beta}$.
- $\widehat{\sigma^2}$ is unbiased but is not the MLE.