Lecture Notes for Theory of Linear Models
STAT 443/851

Lecture 1

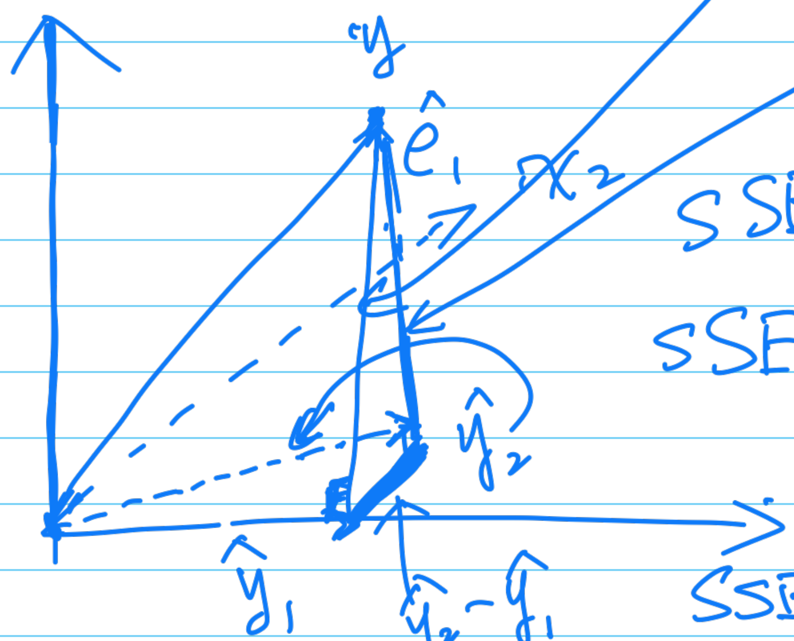
- Overview of STAT 443/851
- Linear Statistical Models
  - Multiple Linear Regression
  - ANOVA

Longhai Li
University of Saskatchewan

overview

$e_1 = y - \hat{y}_1$

$e_2 = y - \hat{y}_2$



$SSE_1 = \|y - \hat{y}_1\|^2$

$SSE_2 = \|y - \hat{y}_2\|^2$

$SSE_1 - SSE_2$

$= \|\hat{y}_2 - \hat{y}_1\|^2$

predicted values of y

$H_0: y \sim x_1$ (reduced)

$H_1: y \sim x_1 + x_2$ (full)

F test is based on $SSE_1 - SSE_2$

suppose we have observation on Y & X_j:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ \vdots \\ Y_n \end{pmatrix} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1,p-1} \\ 1 & X_{21} & & X_{2,p-1} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & & X_{n,p-1} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{p-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix}$$

$$\underset{n\times 1}{y} = \underset{n\times p}{X}\ \underset{p\times 1}{\beta} + \underset{n\times 1}{\varepsilon}$$

$$\varepsilon \sim N_n(0, \sigma^2 I_n)$$

$$\Leftrightarrow \varepsilon_1, \ldots, \varepsilon_n \overset{iid}{\sim} N(0, \sigma^2)$$

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & 1 \end{pmatrix}, \quad \text{identity}$$

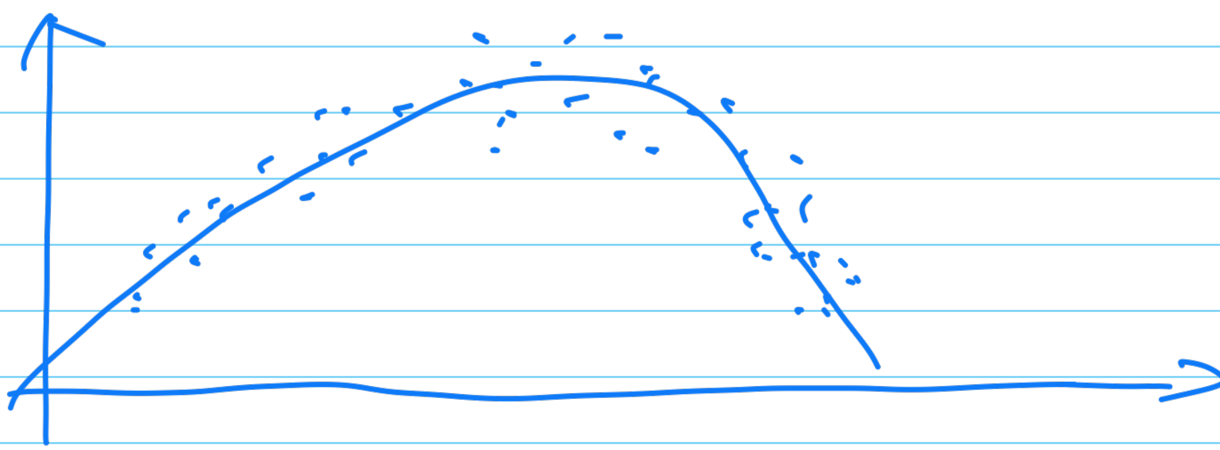
$$Y_1 = \beta_0 + \beta_1 \cdot X_{11} + \cdots + \beta_p X_{1p} + \varepsilon_1$$

polynomial regression

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_{p-1} x_i^{p-1} + \varepsilon_i$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{p-1} \\ 1 & x_2 & x_2^2 & & x_2^{p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & & x_n^{p-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$y \quad = \quad X\beta \quad + \quad \varepsilon$$



One-way ANOVA for categorical X_j

G1	G2	G3
Y_{11} Y_{12}	Y_{21} Y_{22}	Y_{31} Y_{32}

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

Cell means models

$$
\begin{array}{cccccc}
y & X & \boxed{x_1} & x_2 & x_3 & \\
\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \\ Y_{31} \\ Y_{32} \end{bmatrix} &
\begin{bmatrix} G1 \\ G1 \\ G2 \\ G2 \\ G3 \\ G3 \end{bmatrix} &
\multicolumn{3}{c}{\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}} &
\begin{pmatrix} \beta_0 \\ \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} \\
y & & & X & &
\end{array}
$$

$$\underset{6\times 1}{y} = \underset{6\times 3}{X} \cdot \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} + \underset{6\times 1}{\varepsilon}$$

$$y = \underline{X \cdot \beta} + \varepsilon$$

Analysis of Covariance

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & 1 & 0 \\ 1 & x_{21} & 1 & 0 \\ \vdots & \vdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & 0 & 1 \end{pmatrix} \beta + \varepsilon$$

Continuous · Dummy variables for a categorical variable!

$$\underset{n\times 1}{y} = X\beta + \varepsilon$$

$$\underset{p\times 1}{\hat\beta} = (X'X)^{-1}X'y \quad \text{— LS}$$

$$\underset{n\times 1}{\hat y} = X\hat\beta = X(X'X)^{-1}X'y$$

$$= P_X y$$

$$\underset{n\times 1}{\hat e} = y - \hat y \qquad SSE = \|\hat e\|^2$$

$$R^2 = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST}, \quad SST = \sum (y_i - \bar y)^2$$