# Multiple Linear Regression Model with Normal Errors

$n \times p$ covariate matrix, first column is a vector of 1's

$$\underset{n\times 1}{\underset{\sim}{y}} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \qquad \text{equivalent to} \qquad y_i = \beta_0 + \sum_{k=1}^{p-1} \beta_k X_{ik} + \varepsilon_i$$

response vector

* $E\{\underset{\sim}{\varepsilon}\} = \underset{\sim}{0}$ equivalent to $E\{\varepsilon_i\} = 0$
* $\sigma^2\{\underset{\sim}{\varepsilon}\} = \sigma^2 I_n$ equivalent to $\sigma^2\{\varepsilon_i\} = \sigma^2$, $\sigma\{\varepsilon_i, \varepsilon_j\} = 0$, $i \neq j$
* $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

$$\Rightarrow E\{\underset{\sim}{y}\} = X\underset{\sim}{\beta}, \quad \sigma^2\{\underset{\sim}{y}\} = \sigma^2 I_n$$

Estimating $\underset{\sim}{\beta}$
- least squares estimation (doesn't use our normality assumption)
- maximum likelihood estimation (does use normality assumption)

both give same $\underset{\sim}{b}$, estimate of $\underset{\sim}{\beta}$

least squares estimates minimize

$$Q = \sum_{i=1}^{n} \left(Y_i - \beta_0 - \sum_{k=1}^{p-1} \beta_k X_{ik}\right)^2$$

Define $\underset{\sim}{b} = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{pmatrix}$ to be least squares estimates of $\underset{\sim}{\beta}$

satisfy $(X^T X)\underset{\sim}{b} = X^T \underset{\sim}{y}$

$$\Longrightarrow \underset{\sim}{b} = (X'X)^{-1} X' \underset{\sim}{y}$$

maximizes

$$\prod_{i=1}^{n} p(y_i \mid x, \underset{\sim}{\beta}) = L(\underset{\sim}{\beta}, \sigma^2) =$$

$$\frac{1}{\sqrt{2\pi\sigma^2}^{\,n}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(Y_i - \beta_0 - \sum_{k=1}^{p-1}\beta_k X_{ik}\right)^2\right\}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-1}{2\sigma^2} Q\right\}$$

Understanding $(X^T X)\underset{\sim}{b} = X^T \underset{\sim}{y}$

$$X'y = \begin{pmatrix} 1 & \cdots & 1 \\ x_{11} & \cdots & x_{n1} \\ \vdots & & \\ x_{1p-1} & \cdots & x_{np-1} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$Q = \sum_{i=1}^{n} \left( y_i - \beta_0 - \sum_{k=1}^{p-1} \beta_k X_{ik} \right)^2$$

$$= \sum_{i=1}^{n} y_i^2 - 2y_i \left( \beta_0 + \sum_{k=1}^{p-1} \beta_k X_{ik} \right) + \left( \beta_0 + \sum_{k=1}^{p-1} \beta_k X_{ik} \right)^2$$

$$= \underset{\sim}{y}'\underset{\sim}{y} - 2\underset{\sim}{y}' X \underset{\sim}{\beta} + \underset{\sim}{\beta} X' X \underset{\sim}{\beta}$$

Take derivatives with respect to each $\beta_k$ and set equal to zero...

$$-2y'X + 2X'X\underset{\sim}{\beta} = \underset{\sim}{0} \implies (X'X)\underset{\sim}{b} = X'\underset{\sim}{y}$$

$$\cancel{(X'X)^{-1}(X'X)}\underset{\sim}{b} = (X'X)^{-1}X'\underset{\sim}{y}$$

$$\underset{\sim}{b} = (X'X)^{-1}X'\underset{\sim}{y}$$

Once we have $\underset{\sim}{b} = (X'X)^{-1}X'\underset{\sim}{y}$, can talk about:

* fitted values $\hat{Y}_i = b_0 + \sum_{k=1}^{p-1} b_k X_{ik}$ ] estimated regression function

equivalent to

$$\underset{\sim}{\hat{Y}} = \begin{pmatrix} \hat{Y}_1 \\ \vdots \\ \hat{Y}_n \end{pmatrix} = X\underset{\sim}{b}$$

* residuals

$$e_i = Y_i - \hat{Y}_i$$
$$= Y_i - b_0 - \sum_{k=1}^{p-1} b_k X_{ik}$$

equivalent to

$$\underset{\sim}{e} = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = \underset{\sim}{Y} - \underset{\sim}{\hat{Y}} = \underset{\sim}{Y} - X\underset{\sim}{b}$$
$$= \underset{\sim}{Y} - X(X'X)^{-1}X'\underset{\sim}{y}$$
$$= (I - \underbrace{X(X'X)^{-1}X'}_{H,\ \text{hat matrix}})\underset{\sim}{y} \qquad (I\underset{\sim}{y} = \underset{\sim}{y})$$

$$\underset{\sim}{e} = (I - H)\underset{\sim}{y}$$

$$\sigma^2\{\underset{\sim}{e}\} = \sigma^2(I - H)$$

* estimator of $\sigma^2$, $s^2$

$$s^2 = \sum_{i=1}^{n} e_i^2 / n - p = MSE$$

ANOVA for Multiple Linear Regression

$$SSTO = SSE + SSR$$

$$\sum_{i=1}^{n}(y_i - \bar{Y})^2 = \sum_{i=1}^{n}(\underbrace{y_i - \hat{Y}_i}_{e_i})^2 + \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$$

n-1 degrees of freedom — n-p degrees of freedom — p-1 degrees of freedom

everything from simple linear regression setting carries through, we know how to compute $\hat{Y}_i$'s, $e_i$'s, etc.

$$MSE = \frac{SSE}{n-p} \qquad MSR = \frac{SSR}{p-1}$$

$$E\{MSE\} = \sigma^2 \qquad E\{MSR\} = \sigma^2 + K$$

K: nonnegative

only equal to zero if $\beta_1 = \beta_2 = \dots = \beta_{p-1} = 0$

$E\{MSE\} = E\{MSR\}$ if and only if $\beta_1 = \dots = \beta_{p-1} = 0$

# F Test For Multiple Linear Regression

Test null $H_0: \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0$ $\}$ $\beta_k = 0$ for $k = 1, \dots, p-1$

versul alternative $H_a$: at least one $\beta_k \neq 0$ $k = 1, \dots, p-1$

Use $F^* = \dfrac{MSR}{MSE}$ as test statistic

Decision rule for level-$\alpha$ test is

conclude $H_0$ if $F^* \leq F(1-\alpha; p-1, n-p)$

conclude $H_a$ otherwise