VE401 Recitation Class Note13

Multiple Linear Regression

Chen Siyi
siyi.chen_chicy@sjtu.edu.cn

1 Basic Model

The multilinear model:

$$\mu_{Y|x_1,\ldots,x_p} = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

The polynomial model:

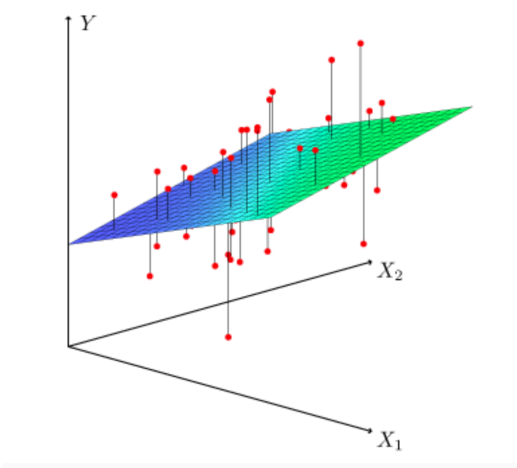
$$\mu_{Y|x} = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p$$

Generalization:

$$\boldsymbol{Y} = X\beta + \boldsymbol{E}$$

1.1 Least-Squares Estimation

$$\mathrm{SS_E} = \langle \boldsymbol{Y} - X\boldsymbol{b}, \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b} \rangle = \|\boldsymbol{Y}\|^2 - 2\langle X\boldsymbol{b}, \boldsymbol{Y} \rangle + \|X\boldsymbol{b}\|^2$$



$$\boldsymbol{b} = \left(X^T X\right)^{-1} X^T \boldsymbol{Y}$$

$$\widehat{\boldsymbol{Y}} = X\boldsymbol{b}$$

1.2 Sum-of-Squares Error Decomposition

$$\mathrm{SS_T} = \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2$$

Define $P := \frac{1}{n}\begin{pmatrix} 1 & 1 & \dots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix}$

Define $H := X\left(X^T X\right)^{-1} X^T$

$$\widehat{\boldsymbol{Y}} = H\boldsymbol{Y}$$

$$SS_E = \langle \boldsymbol{Y}, (1_n - H)\,\boldsymbol{Y} \rangle$$

$$\begin{aligned} \mathrm{SS_T} &= \langle \boldsymbol{Y}, (1_n - P)\,\boldsymbol{Y} \rangle \\ &= \underbrace{\langle \boldsymbol{Y}, (1_n - H)\,\boldsymbol{Y} \rangle}_{=SS_E} + \underbrace{\langle \boldsymbol{Y}, (H - P)\boldsymbol{Y} \rangle}_{=:SS_R} \end{aligned}$$

Further we get

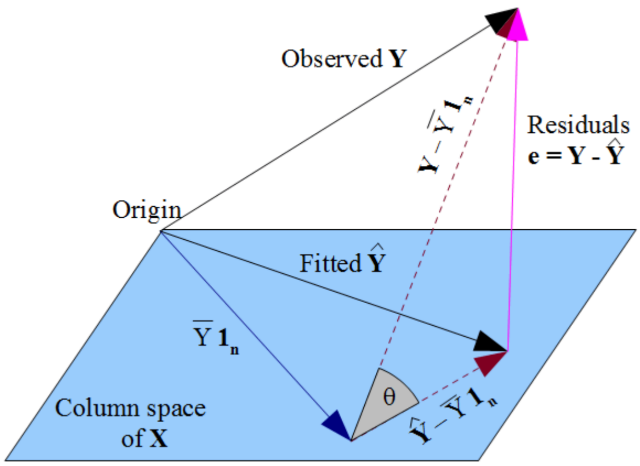
$$SS_R = \langle \boldsymbol{Y}, (H-P)\boldsymbol{Y} \rangle = \langle (H-P)\boldsymbol{Y}, (H-P)\boldsymbol{Y} \rangle$$

Summary:

$$\mathrm{SS_T} = \mathrm{SS_R} + \mathrm{SS_E}$$

$$\sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2 = \sum_{i=1}^{n} \left(\widehat{Y}_i - \bar{Y}\right)^2 + \sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i\right)^2$$

$$R^2 = \frac{\mathrm{SS_R}}{\mathrm{SS_T}}$$



1.3 Distribution of the Sum of Squares Error

Theorem:

1. $SS_E\ /\sigma^2$ follows a chi-squared distribution with n − p − 1 degrees of freedom.
2. If $\beta = (\beta_0, 0, \ldots, 0)$, then $SS_R\ /\sigma^2$ follows a chi-squared distribution with p degrees of freedom.
3. Furthermore, SS_E and SS_R are independent random variables.

Corollary:
The below estimator is unbiased for σ^2.

$$S^2 := \frac{\mathrm{SS_E}}{n - p - 1}$$

MLR in Practice

1. Find a quadratic model for the below data:

X	0.60	1.30	3.50	1.90	2.30
Y	0.33	1.35	2.30	0.95	1.25

2. Find a multilinear regression for the below data:

X1	0.60	1.30	3.50	1.90	2.30
X2	0.50	1.50	3.00	2.00	2.50
Y	0.33	1.35	2.30	0.95	1.25

2 Inferences on $\mu_{Y|\boldsymbol{x}_0}$

*Notice $Y \mid \boldsymbol{x}_0$ is not a vector.

2.1 Distribution of $\widehat{\mu}_{Y|\boldsymbol{x}_0}$

$\widehat{\mu}_{Y|\boldsymbol{x}_0}$ follows a normal distribution because:

$$\widehat{\mu}_{Y|\boldsymbol{x}_0} = \boldsymbol{x}_0^T \boldsymbol{b} = \boldsymbol{x}_0^T \left(X^T X\right)^{-1} X^T \boldsymbol{Y}$$

And

$$\mathrm{E}\left[\widehat{\mu}_{Y|\boldsymbol{x}_0}\right] = \mathrm{E}\left[\boldsymbol{x}_0^T \boldsymbol{b}\right] = \boldsymbol{x}_0^T \mathrm{E}[\boldsymbol{b}] = \boldsymbol{x}_0^T \beta = \mu_{Y|\boldsymbol{x}_{10},\ldots,\boldsymbol{x}_{p0}}$$

$$\mathrm{Var}\left[\widehat{\mu}_{Y|\boldsymbol{x}_0}\right] = \mathrm{Var}\left[\boldsymbol{x}_0^T \boldsymbol{b}\right] = x_0^T \,\mathrm{Var}[\boldsymbol{b}] \boldsymbol{x}_0 = \sigma^2 \boldsymbol{x}_0^T \left(X^T X\right)^{-1} \boldsymbol{x}_0$$

Therefore

$$T_{n-p-1} = \frac{\widehat{\mu}_{Y|\boldsymbol{x}_0} - \mu_{Y|\boldsymbol{x}_0}}{S\sqrt{\boldsymbol{x}_0 T \left(X^T X\right)^{-1} \boldsymbol{x}_0}}$$

2.2 Interval Estimation for $\mu_{Y|\boldsymbol{x}_0}$

$100(1-\alpha)\%$ confidence interval for $\mu_{Y|\boldsymbol{x}_0}$:

$$\mu_{Y|\boldsymbol{x}_0} = \widehat{\mu}_{Y|\boldsymbol{x}_0} \pm t_{\alpha/2,n-p-1} S \sqrt{\boldsymbol{x}_0^T \left(X^T X\right)^{-1} \boldsymbol{x}_0}$$

3 Predictions on $Y|\boldsymbol{x}_0$

Similarly as in SLR, we obtain $100(1-\alpha)\%$ prediction interval for $Y|\boldsymbol{x}_0$:

$$Y \mid \boldsymbol{x}_0 = \widehat{\mu}_{Y|\boldsymbol{x}_0} \pm t_{\alpha/2,n-p-1} S \sqrt{1 + \boldsymbol{x}_0^T \left(X^T X\right)^{-1} \boldsymbol{x}_0}$$

4 Inferences on β

4.1 Distribution of $\boldsymbol{b}$

Theorem:
The random vector $\boldsymbol{b}$ follows a normal distribution with mean $\boldsymbol{\beta}$ and variance-covariance matrix $\sigma^2\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}$.

Theorem:
The statistic $(np1)S^2/\sigma^2 = SS_E/\sigma^2$ is independent of $\boldsymbol{b}$.

Theorem:
The random vector $\boldsymbol{b}$ follows a normal distribution with mean $\boldsymbol{\beta}$ and variance-covariance matrix $\sigma^2\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}$.
Also then notice $\boldsymbol{b}$ is an unbiased estimator for $\boldsymbol{\beta}$.

Theorem:
The statistic $(np1)S^2/\sigma^2 = SS_E/\sigma^2$ is independent of $\boldsymbol{b}$.

Write

$$X^TX = \begin{pmatrix} \xi_{00} & * & \cdots & \cdots & * \\ * & \xi_{11} & \ddots & & * \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & * \\ * & \cdots & \cdots & * & \xi_{pp} \end{pmatrix}$$

Then

$$\text{Var}\left[B_i\right] = \xi_{ii}\sigma^2, \quad i = 0, \ldots, p$$

4.2 Interval Estimation for β

Statistic:

$$\frac{\widehat{\beta}_j - \beta_j}{S\sqrt{\xi_{jj}}} = T_{n-p-1} = \frac{\left(b_j - \beta_j\right)/(\sigma\sqrt{\xi_{jj}})}{\sqrt{(n-p-1)S^2/\sigma^2/(n-p-1)}}$$

CI for $\boldsymbol{\beta}$:

$$\beta_j = b_j \pm t_{\alpha/2,n-p-1}S\sqrt{\xi_{jj}}$$

5 Model Tests

5.1 F-Test for Significance of Regression

Let x_1, ... , x_p be the predictor variables in a multilinear model (26.1) for Y . Then

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$$

is rejected at significance level α if the test statistic

$$F_{p,n-p-1} = \frac{SS_R/p}{SS_E/(n-p-1)} = \frac{SS_R/p}{S^2}$$

satisfies $F_{p,n-p-1} > f_{\alpha,p,n-p-1}$

Notice you can also use R^2 to test, since

$$F_{p,n-p-1} = \frac{n-p-1}{p}\frac{R^2}{1-R^2}$$

5.2 T-Test for Model Sufficiency

Test whether a certain β_j is needed to be non-zero.

Suppose that a regression model using the parameters β_0,...,β_p is fitted to Y. Then for any j = 0,...,p

$$H_0 : \beta_j = 0$$

is rejected at significance level α if the test statistic

$$T_{n-p-1} = \frac{b_j}{S\sqrt{\xi_{jj}}}$$

satisfies $|T_{n-p-1}| > t_{\alpha/2,n-p-1}$.

5.3 General Partial F-Test for Model Sufficiency

Test whether a full model is needed.

A full model of p + 1 predictor variables:

$$\mu_{Y|x_1,\ldots,x_p} = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

A reduced model of m + 1 < p + 1 predictor variables:

$$\mu_{Y|x_1,\ldots,x_m} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \cdots + \tilde{\beta}_m x_m$$

H_0 : the reduced model is sufficient
is rejected at significance level α if the test statistic

$$F_{p-m,n-p-1} = \frac{n-p-1}{p-m} \frac{\mathrm{SS}_{\mathrm{E;reduced}} - \mathrm{SS}_{\mathrm{E;full}}}{\mathrm{SS}_{\mathrm{E;full}}}$$

satisfies $F_{p-m,n-p-1} > f_{\alpha,p-m,n-p-1}$.

The T-test for a single variable is equivalent to a partial F-Test when applied to a reduced model lacking only that single variable.

The F-test for significance of regression may be regarded as a partial F-test where the reduced model contains no regressors.

Partial F-Test in Practice

Find a linear model for the below data:

X	0.60	1.30	3.50	1.90	2.30
Y	0.33	1.35	2.30	0.95	1.25

Is the quadratic model find before sufficient?

6 Find the Right Model