# STAT 3008
# Applied Regression Analysis
## Tutorial 3: Multiple Linear Regression

LAI Chun Hei*

Department of Statistics, The Chinese University of Hong Kong

## 1 Model Assumption

In the last tutorial, the model assumes that the response $Y_{|X=x}$ only depends on one factor, $x$. However, our real world is much more complicated that many phenomena cannot be explained by only one factor. Therefore, we would often like to assume that the response are affected by multiple factors.

Therefore, to find out the relationship of our response with these factors, we have to use multiple linear regression. The model assumption is as follow.

**Multiple Linear Regression Model**

The model relating the factors are assumed to be:

$$y_i = Y_{|X=(x_{i1},x_{i2},\dots,x_{ip})} = \beta_0 + \beta_1 x_{i1} + +\beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i$$

where $\mathbb{E}(e_i) = 0$, $\text{Var}(e_i) = \sigma^2$ and $e_i$'s are i.i.d. Therefore, we have

$$\begin{aligned} \mathbb{E}[Y|X = (x_{i1}, x_{i2}, \dots, x_{ip})] &= \beta_0 + \beta_1 x_{i1} + +\beta_2 x_{i2} + \cdots + \beta_p x_{ip} \\ \text{Var}[Y|X = (x_{i1}, x_{i2}, \dots, x_{ip})] &= \sigma^2. \end{aligned}$$

For simplicity, the model is always express in matrix form:

$$\underline{Y} = X\,\underline{\beta} + \underline{e}$$

where the matrices are defined as

$$\underline{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \quad \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad \text{and} \quad \underline{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}.$$

Note that the $i$-th row of the matrix equation is actually the equation corresponding to the $i$-th response,

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i.$$

Also, the expectation and covariance matrix of $\underline{e}$ is given by

$$\mathbb{E}(\underline{e}) = \underline{0} \quad \text{and} \quad \text{Var}(\underline{e}) = \begin{bmatrix} \text{Var}(e_1) & \text{Cov}(e_1,e_2) & \cdots & \text{Cov}(e_1,e_n) \\ \text{Cov}(e_2,e_1) & \text{Var}(e_2) & \cdots & \text{Cov}(e_2,e_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(e_n,e_1) & \text{Cov}(e_n,e_2) & \cdots & \text{Var}(e_n) \end{bmatrix} = \sigma^2 \mathbb{I}$$

*s1155002282@sta.cuhk.edu.hk

**Remark 1.1.** The number of parameters to estimate is $p+1$, with $p$ coefficients, $\beta_1, \beta_2, \ldots, \beta_p$, and the intercept, $\beta_0$. ■

**Remark 1.2.** The error term is assumed to be independent as

$$\mathrm{Var}(\underline{e}) = \sigma^2 \mathbb{I} \qquad \Rightarrow \mathrm{Cov}(e_i, e_j) = 0 \quad , i \neq j.$$ ■

# 2 Some Operations of Matrix

In multidimensional analysis, the involvement of matrix simplifies the presentation of simultaneous equations. As in one dimensional cases, many operations such as differentiation and expectation. The definitions are given below.

## Matrix Differentiation

**Definition 2.1.** *Consider a vector* $\underline{\beta} = [\beta_1 \quad \beta_2 \quad \cdots \quad \beta_k]' \in \mathbb{R}^k$ *and a function* $f : \mathbb{R}^k \to \mathbb{R}$ *where*

$$f(\underline{\beta}) = f\left([\beta_1 \quad \beta_2 \quad \cdots \quad \beta_k]'\right),$$

*the derivative of* $f$ *with respect to* $\underline{\beta}$ *is given by*

$$\frac{\partial f}{\partial \underline{\beta}} = \begin{bmatrix} \frac{\partial f}{\partial \beta_1} & \frac{\partial f}{\partial \beta_2} & \cdots & \frac{\partial f}{\partial \beta_k} \end{bmatrix}'$$

**Lemma 2.1.** *Let* $\underline{\beta} \in \mathbb{R}^k$ *and* $\underline{c} \in \mathbb{R}^k$, *then we have*

$$\frac{\partial}{\partial \underline{\beta}} \underline{c}' \underline{\beta} = \frac{\partial}{\partial \underline{\beta}} \underline{\beta}\, \underline{c}' = \underline{c}$$

**Lemma 2.2.** *Let* $\underline{\beta} \in \mathbb{R}^k$ *and* $M \in \mathbb{R}^{k \times k}$, *then we have*

$$\frac{\partial}{\partial \underline{\beta}} \underline{\beta}' M \underline{\beta} = (M' + M) \underline{\beta}$$

## Trace of a Matrix

**Definition 2.2.** *Let* $A = [a_{ij}] \in \mathbb{R}^{n \times n}$, *then the trace of* $A$ *is defined as* $tr(A) = \sum_{i=1}^{n} a_{ii}$.

**Lemma 2.3.** *Let* $A$ *and* $B$ *are both* $n \times n$ *matrix, then we have the following properties*

1. $tr(A + B) = tr(A) + tr(B)$
2. $tr(AB) = tr(BA)$
3. $tr(\mathbb{E}(A)) = \mathbb{E}(tr(A))$

## Random Vector

**Definition 2.3.** *Let* $\underline{X} = [X_1 \quad X_2 \quad \cdots \quad X_n]'$ *where* $X_1, X_2, \ldots, X_n$ *are random variables, then the expectation and variance of* $\underline{X}$ *is defined as*

$$\mathbb{E}(\underline{X}) = \begin{bmatrix} \mathbb{E}(X_1) \\ \mathbb{E}(X_2) \\ \vdots \\ \mathbb{E}(X_n) \end{bmatrix} \quad \text{and} \quad \mathrm{Var}(\underline{X}) = \begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & Var(X_2) & \cdots & \mathrm{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \cdots & \mathrm{Var}(X_n) \end{bmatrix}$$

**Lemma 2.4.** *Let* $\underline{X}$ *be a random vector of size* $n$ *and* $A$ *be an* $n \times n$ *constant matrix, then*

$$\mathbb{E}(A \underline{X}) = A\, \mathbb{E}(\underline{X}) \qquad \text{and} \qquad \mathrm{Var}(A \underline{X}) = A\, \mathrm{Var}(\underline{X}) A'$$

# 3 Least Square Estimator

Similar as in simple linear regression, we want to find an estimate that minimises the distance in the response in the $p+1$ dimensional Euclidean space. The problem is equivalent to minimising

$$\mathrm{RSS}(\underline{\beta}) = (\underline{Y} - X\,\underline{\beta})'(\underline{Y} - X\,\underline{\beta}) = \underline{Y}'\,\underline{Y} - 2\,\underline{Y}'\,X\,\underline{\beta} + \underline{\beta}'\,X'X\,\underline{\beta}$$

By differentiation, we have the following result.

**Least Square Estimator**

The least square estimator of multiple linear regression is given by

$$\hat{\underline{\beta}} = (X'X)^{-1}X'\,\underline{Y}$$

$$\hat{\sigma^2} = \frac{\mathrm{RSS}(\hat{\underline{\beta}})}{n-(p+1)} = \frac{\underline{Y}'(I - X(X'X)^{-1}X')\,\underline{Y}}{n-(p+1)}$$

The expectation and variance of the estimators are

$$\mathbb{E}(\hat{\underline{\beta}}) = \underline{\beta}, \qquad \mathrm{Var}(\hat{\underline{\beta}}) = \sigma^2(X'X)^{-1} \quad \text{and} \quad \mathbb{E}(\hat{\sigma^2}) = \sigma^2.$$

Asymptotically, by central limit theorem, we know that

$$\hat{\underline{\beta}} \sim \mathcal{N}\left(\underline{\beta}, \sigma^2(X'X)^{-1}\right) \qquad \text{and} \qquad \frac{(n-p-1)\,\hat{\sigma^2}}{\sigma^2} \sim \chi^2(n-p-1).$$

**Remark 3.1.** Because of its frequency appearance in the context of multiple linear regression, the matrix $X(X'X)^{-1}X'$ is known as the hat matrix, denoted by $H$. ■

**Exercise 3.1.** Show that

$$\mathbb{E}(\hat{\underline{\beta}}) = \underline{\beta} \qquad \text{and} \qquad \mathrm{Var}(\hat{\underline{\beta}}) = \sigma^2(X'X)^{-1}.$$

**Exercise 3.2.** (2013 Fall Midterm #5) In searching for the estimates of the regression coefficient $\underline{\beta}$, we differentiate the RSS and solve for system of equations. Will there be more than one solutions? Will the solution be the maximiser of the RSS instead of minimiser?

**Exercise 3.3.** Show that $\mathbb{E}(\hat{\sigma^2}) = \sigma^2$.

**Exercise 3.4.** (2013 Fall Midterm #3) For the regression $\underline{Y} = X\,\underline{\beta} + \underline{e}$ where $\underline{e} \sim \mathcal{N}(\underline{0}, \sigma^2)$, let $\hat{y}_i = underline X_i\,\underline{\hat{\beta}}$ be the fitted value of the i-th observation, $i = 1, 2, \ldots, n$. Let $X$ be an $n \times p$ matrix.

1. Find $\mathbb{E}(\sum_{i=1}^{n} \hat{y_i}^2)$ in terms of $X$, $\underline{\beta}$, $p$ and $\sigma^2$.
2. Find $\mathbb{E}(\sum_{i=1}^{n} (\hat{y}_i - \bar{y})(y_i - \hat{y}_i))$.

# 4 Added-Variable Plot

In multiple regression, it is difficult to use a graph to illustrate the relationship between $\mathbb{E}[Y|X = (x_{i1}, x_{i2}, \ldots, x_{ip})]$ and $(x_{i1}, x_{i2}, \ldots, x_{ip})$, especially when we have $p > 2$. However, it is always easier to understand if we have graphs illustrating the meaning of the $\hat{\beta}_i$'s. The way of doing this is by a added-variable plot.

## Added-Variable Plot

For a regression model $\underline{Y} = X\,\underline{\beta} + \underline{e}$, it can be rewritten in the form

$$\underline{Y} = \underline{1}\beta_0 + \underline{X_1}\beta_1 + \cdots + \underline{X_i}\beta_i + \cdots + \underline{X_p}\beta_p + \underline{e}\,.$$

Now, if we want to now the effect of the $i$th factor, i.e. $\beta_i$, we first arrange the model in the following manner.

$$\underline{Y} = \underline{X_i}\beta_i + X_O\underline{\beta_O} + \underline{e}$$

where $X_O$ is a matrix grouping $\underline{X_1}, \underline{X_2}, \ldots, \underline{X_{i-1}}, \underline{X_{i+1}}, \ldots, \underline{X_p}$ and $\underline{\beta_O}$ is a vector, i.e.

$$X_O = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1(i-1)} & x_{1(i+1)} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2(i-1)} & x_{1(i+1)} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{n(i-1)} & x_{1(i+1)} & \cdots & x_{np} \end{bmatrix} \quad \text{and} \quad \underline{\beta_O} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{i-1} \\ \beta_{i+1} \\ \vdots \\ \beta_p \end{bmatrix}$$

Then, we can draw the added-variable plot following the process stated below:

1. Perform multiple linear regression on $\underline{Y}$ against $X_O$. Obtain the residuals $\hat{e}_{\underline{Y}\,|X_O}$.
2. Perform multiple linear regression on $\underline{X_i}$ against $X_O$. Obtain the residuals $\hat{e}_{\underline{X_i}|X_O}$.
3. Perform no-intercept simple linear regression on $\hat{e}_{\underline{Y}\,|X_O}$ against $\hat{e}_{\underline{X_i}|X_O}$.
4. The plot of $\hat{e}_{\underline{Y}\,|X_O}$ against $\hat{e}_{\underline{X_i}|X_O}$ is the added-variable plot.

**Proposition 4.1.** *The slope obtained from step 3 is in fact equal to $\hat{\beta}_i$ which we obtained from $\underline{\hat{\beta}} = (X'X)^{-1}X'\,\underline{Y}$.*

**Remark 4.1.** In reality, it is very often that different factors are correlated. Therefore, the $i$th factor may somehow depends on the other factors. The regression of $\underline{X_i}$ against $X_O$ investigate such a relationship. In general,

- $\hat{e}_{\underline{Y}\,|X_O}$ is the value of $Y$ which cannot be explained by $X_O$.
- $\hat{e}_{\underline{X_i}|X_O}$ is the value of the $i$th factor which cannot be explained by $X_O$. Such values somehow serves as a pure $i$th factor value.

Therefore, the regression of $\hat{e}_{\underline{Y}\,|X_O}$ against $\hat{e}_{\underline{X_i}|X_O}$ shows the pure relationship between the response and the $i$th predictor, net of the effect from other predictors. ■

**Remark 4.2.** In step 3, a no-intercept simple linear regression is needed. However, when plotting the added-variable plot, we can still apply the result from normal simple linear regression. The intercept is forced to be zero by the property of residuals.

Recall that in simple linear regression of $y$ against $x$, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\,\bar{x}$, here the response is $\hat{e}_{\underline{Y}\,|X_O}$ and the predictor is $\hat{e}_{\underline{X_i}|X_O}$. Therefore, we have the intercept (here, we use $\hat{\eta}_i$ for the parameters in this model, not to be confused with the model $\underline{Y} = X\,\underline{\beta} + \underline{e}$)

$$\hat{\eta}_0 = \frac{1}{n}\sum_{j=1}^{n}\left[\hat{e}_{\underline{Y}\,|X_O}\right]_j - \hat{\eta}_1\left(\frac{1}{n}\sum_{j=1}^{n}\left[\hat{e}_{\underline{X_i}|X_O}\right]_j\right) = \frac{1}{n}\cdot 0 - \hat{\eta}_1 \cdot 0 = 0.$$ ■

**Exercise 4.1.** This exercise proves Proposition 4.1.

1. Show that for the linear regression $\underline{Y} = X\,\underline{\beta} + \underline{e}$, let $H = X(X'X)^{-1}X$, we have
$$\underline{\hat{e}} = (\mathbb{I} - H)\,\underline{Y} \qquad \text{and} \qquad (\mathbb{I} - H)X = 0.$$

2. Show that for a linear regression $\underline{Y} = X\,\underline{\beta} + \underline{e}$, the $i$th coefficient $\hat{\beta}_i$ is equal to the coefficient $\hat{\eta_1}$ in the simple linear regression $\hat{e}_{\underline{Y}\,|X_O} = \eta_1 \hat{e}_{\underline{X_i}|X_O} + \tilde{e}$.

3. Show that, let $H_O = X_O(X_O'X_O)^{-1}X_O'$,
$$\hat{\beta}_i = \frac{\underline{X_i}(\mathbb{I} - H_O)\,\underline{Y}}{\underline{X_i}(\mathbb{I} - H_O)\underline{X_i}}.$$