

# Multiple Linear Models

## Model and Model Assumptions

The MLR is a generalization of SLR for multiple predictors  $x_1, x_2, \dots, x_p$ . MLM becomes SLM when  $p = 1$ .

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon \quad (1)$$

Specifically, when applying the equation to a data set of  $n$  observations we will obtain  $n$  linear equation.

Data set

$y$	$x_1$	$x_2$	$\dots$	$x_p$
$y_1$	$x_{11}$	$x_{12}$	$\dots$	$x_{1p}$
$y_2$	$x_{21}$	$x_{22}$	$\dots$	$x_{2p}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$y_n$	$x_{n1}$	$x_{n2}$	$\dots$	$x_{np}$

Equations

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_p x_{1p} + \epsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_p x_{2p} + \epsilon_2 \\ &\dots \\ y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_p x_{np} + \epsilon_n \end{aligned}$$

These  $n$  equations can be written in a matrix form

$$\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \times \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_n \end{pmatrix}$$

Or

$$y = X\beta + \epsilon$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon \quad (2)$$

Let the error/residual of each approximation be  $e_i = y_i - \hat{y}_i$

To do inferential statistics, estimating the error of  $\hat{\beta}_1$  or finding confidence interval for  $\beta_0$ , we need to bring in a probability distribution. Assume that  $x_i$  is a known constant (non-random) and that

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad (3)$$

where

$$y =$$

$\epsilon_1, \epsilon_2, \dots, \epsilon_n \sim^{iid} N(0, \sigma)$ . From this equation, we see that  $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ .  $S_{xy}$  and  $S_{yy}$  are random,  $S_{xx}$  is non-random.

## Coefficient Estimators

Using the same approach as the SLM, we can estimate the coefficient parameters vector  $\beta$  using the least square method. The least squared estimators of  $\beta$  minimize the residual sum squares, which is a quadratic function of  $\beta$ .

The least squared estimators of  $\beta$ , denoted by  $\hat{\beta}$ , is

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

where  $X'$  denotes the transpose matrix of  $X$ .

## Goodness of Fit

We have the following formula

$$\sum_{i=1}^n (y_i - \bar{y}) = \sum_{i=1}^n (y_i - \hat{y}_i) + \sum_{i=1}^n (\hat{y}_i - \bar{y})$$

The quantity  $TSS = \sum_{i=1}^n (y_i - \bar{y})$  can be thought as the total amount of **information** (variance) in the response variable  $y$ . This quantity is decomposed into  $RSS = \sum_{i=1}^n (\hat{y}_i - \bar{y})$  which can be thought as the **information** explained by the regression models and  $RSS = \sum_{i=1}^n (y_i - \hat{y}_i)$ , which is the amount of **information** that can not be explained by the model (the sum of squared errors of the models).

This formula leads to the definition of  $R^2$

$$R^2 = 1 - \frac{RSS}{TSS}$$

Another way to view this formula is that TSS is actually the sum squared errors of the naive model  $y = \bar{y}$  (predicting  $y$  value by its average, disregarding  $x$ ). Thus,

$$R^2 = 1 - \frac{\text{Sum squared errors of Linear Model}}{\text{Sum squared errors of the naive Model}}$$

Thus,  $R^2$  also measure how good the linear model is when comparing with the naive model. This definition of  $R^2$  can be applied to measure the goodness of fit of other models, not just linear model.

To measure the goodness of fit of this linear approximation, we compare the approximation with a naive approximation, which is to approximate all  $y_i$  with their average,  $\bar{y}$ . The naive model is a special case of the linear model where  $\beta_1 = 0$ . In the other words, the naive models is

$$y = f_2(x) = \bar{y} \quad (4)$$

The sum square error of the naive models is  $\sum (y_i - \bar{y})^2$ . This quantity is also called the total sum square (SST or TSS). We use the following  $R^2$  to measure the goodness of fit of the linear model

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{SSE}{SST} \quad (5)$$

It's noticed that if the linear approximation fit the data perfectly ( $SSE = 0$ ),  $R^2 = 1$ .

### F-test and t-test

We are interested in testing a subset of  $q$   $\beta_i$  parameters are all zeros. This is equivalent to a subset of  $q$  predictors are not useful in regressing  $y$ . Since we can number the predictors (hence  $\beta_i$ ) however we want, we assume that these zeros  $\beta_i$  are the last  $q$  parameters, or

$$H_0 : \beta_{p-q+1} = \beta_{p-q} = \dots = \beta_p = 0$$

We can view this test in another way. We notice that the above  $H_0$  is equivalent to stating that the reduced model do not contain the last  $q$  predictors are better fit than the full model that contain all the predictors.

*Reduced Model:*

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_q + \epsilon$$

*Full Model:*

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon$$

A smaller p-value will support the full model and reject the reduced model. We use the following F statistics for this test

$$F = \frac{(RSS_0 - RSS_1)/q}{RSS_1/(n - p - 1)} \sim F_{q, n-p-1},$$

where  $RSS_0$  and  $RSS_1$  are the residual sum squares of the reduced and full model, respectively.

### t-test

$$H_0 : \beta_i = 0 \quad H_a : \beta_i \neq 0$$

We have that under  $H_0$

$$t = \frac{\hat{\beta}_1}{se(\hat{\beta}_i)} \sim t_{n-2}$$

### Practice Problems