

THE MULTIPLE LINEAR REGRESSION MODEL

Notation:

p predictors x_1, x_2, \dots, x_p

$k-1$ *non-constant* terms u_1, u_2, \dots, u_{k-1}

Each u_j is a function of x_1, x_2, \dots, x_p : $u_j = u_j(x_1, x_2, \dots, x_p)$

For convenience, we often set $u_0 = 1$ (constant function/term)

The Basic Multiple Linear Regression Model: Two assumptions:

1. $E(Y|\underline{x}) = \eta_0 + \eta_1 u_1 + \dots + \eta_{k-1} u_{k-1}$ (Linear Mean Function)
2. $\text{Var}(Y|\underline{x}) = \sigma^2$ (Or: $\text{Var}(Y|\underline{u}) = \sigma^2$) (Constant Variance)

Assumption (1) in vector notation:

$$\underline{u} = \begin{bmatrix} u_0 \\ u_1 \\ \cdot \\ \cdot \\ u_{k-1} \end{bmatrix} = \begin{bmatrix} 1 \\ u_1 \\ \cdot \\ \cdot \\ u_{k-1} \end{bmatrix}, \quad \underline{\eta} = \begin{bmatrix} \eta_0 \\ \eta_1 \\ \cdot \\ \cdot \\ \eta_{k-1} \end{bmatrix}$$

Then $\underline{\eta}^T = [\eta_0 \ \eta_1 \ \dots \ \eta_{k-1}]$ and

$\underline{\eta}^T \underline{u} = \eta_0 + \eta_1 u_1 + \dots + \eta_{k-1} u_{k-1}$,
so (1) becomes:

$$(1') E(Y|\underline{x}) = \underline{\eta}^T \underline{u}$$

If we have data with i^{th} observation $x_{i1}, x_{i2}, \dots, x_{ip}, y_i$, recall

$$\underline{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \cdot \\ \cdot \\ x_{ip} \end{bmatrix} = [x_{i1}, x_{i2}, \dots, x_{ip}]^T$$

Define similarly

$u_{ij} = u_j(x_{i1}, x_{i2}, \dots, x_{ip})$ = the value of the j^{th} term for the i^{th} observation, and

$$\underline{u}_i = \begin{bmatrix} u_{i0} \\ u_{i1} \\ \cdot \\ \cdot \\ u_{i,k-1} \end{bmatrix}$$

So in particular, the model says

$$E(Y|\underline{x}_i) = \underline{\eta}^T \underline{u}_i$$

Estimation of Parameters: Analogously to the case of simple linear regression, consider functions ("hyperplanes") of the form

$$y = h_0 + h_1 u_1 + \dots + h_{k-1} u_{k-1} = \underline{h}^T \underline{u}.$$

The *least squares estimate* of $\underline{\eta}$ is the vector

$$\hat{\underline{\eta}} = \begin{bmatrix} \hat{\eta}_0 \\ \hat{\eta}_1 \\ \mathbf{M} \\ \hat{\eta}_{k-1} \end{bmatrix}$$

that minimizes the "objective function"

$$RSS(\underline{h}) = \sum_{i=1}^n (y_i - \underline{h}^T \underline{u}_i)^2$$

Recall: In simple linear regression, the solution had $SXX = \sum_{i=1}^n (x_i - \bar{x})^2$ in the

denominator. So the formula won't work if all x_i 's = \bar{x} . In this case, there is not a unique solution to the least squares problem. (Draw a picture in the case $n = 2$!)

In multiple regression: There is a unique solution $\hat{\underline{\eta}}$ *provided:*

- i) $k < n$ (the number of terms is less than the number of observations)
- ii) no u_j is (as a function) a linear combination of the other u_j 's

When (ii) is violated, we say there is (*strict*) *multicollinearity*.

If there is a unique solution, it is called the *ordinary least squares (OLS) estimate* of the (vector of) coefficients.