The Math of the Multiple Regression Equation

by Muqtasid Qureshi

for Advanced Linear Algebra

Abstract. This paper will go over the mathematical theory of the multiple regression equation, specifically for calculating the coefficients and the error terms. Prerequisites include a solid understanding of linear algebra and some familiarity with mathematical proofs.

1 Theorem

1.1 Dataset

For data analysis, a tabular (table-format) dataset is an m × n matrix with m rows (number of observations) and n columns (number of variables). The general format is given by

$$Dataset = \begin{array}{|c|c|c|c|c|} \hline Y & X_1 & X_2 & \cdots & X_r \\ \hline y_1 & x_{1,1} & x_{1,2} & \cdots & x_{1,r} \\ y_2 & x_{2,1} & x_{2,2} & \cdots & x_{2,r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_m & x_{m,1} & x_{m,2} & \cdots & x_{m,r} \\ \hline \end{array}$$

where:

- the Y column is the dependent variable
- each X column is an independent variable
- r = n - 1 = the number of independent variables
- the term $x_{m,1}$ means entry on mth row for the 1st x variable

Variables as Vectors Each column of a matrix can be represented as a vertical vector. This means that the dataset can also be represented as:

$$Dataset = \big[[Y], [X_1], [X_2], \cdots, [X_r]\big]$$

where:

- every variable is a vector within the dataset matrix

1.2 Equation - General Form

The general form of the regression equation is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_r X_r + \varepsilon \tag{1}$$

where :

- β_0 is the coefficient of the y-intercept
- all other $\beta's$ are the coefficient of the corresponding X variable
- ε is the error term of the regression

1.3 Equation - True Form

The true form of the regression equation is:

$$Y = X\beta + \varepsilon \tag{2}$$

where:

- X is the design matrix, which contains an all-one column for the y-intercept, and the x variables as the other columns
- β is a vector containing all coefficients of the regression

The expanded form of the regression equation is:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,r} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m,1} & x_{m,2} & \cdots & x_{m,r} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_r \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}$$

2 Proof

2.1 Beta Coefficients

The equation of the Beta Coefficient vector is:

$$\beta = (X^T X)^{-1} X^T Y \tag{3}$$

First, one must take the transpose of the design matrix and multiply it by the original design matrix:

$$X^T X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{1,1} & x_{2,1} & \cdots & x_{m,1} \\ x_{1,2} & x_{2,2} & \cdots & x_{m,2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,r} & x_{2,r} & \cdots & x_{m,r} \end{bmatrix} \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,r} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m,1} & x_{m,2} & \cdots & x_{m,r} \end{bmatrix}$$

$$X^T X = \begin{bmatrix} m & \sum x_{i,1} & \sum x_{i,2} & \cdots & \sum x_{i,r} \\ \sum x_{i,1} & \sum (x_{i,1})^2 & \sum x_{i,1} x_{i,2} & \cdots & \sum x_{i,1} x_{i,r} \\ \sum x_{i,2} & \sum x_{i,1} x_{i,2} & \sum (x_{i,2})^2 & \cdots & \sum x_{i,2} x_{i,r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_{i,r} & \sum x_{i,1} x_{i,r} & \sum x_{i,2} x_{i,r} & \cdots & \sum (x_{i,r})^2 \end{bmatrix}$$

Next, make an augmented matrix. The left hand side is the main matrix. The right hand side is the identity matrix of the same size (n × n).

$$[X^TX|I] = \left[\begin{array}{ccccc|ccccc} m & \sum x_{i,1} & \sum x_{i,2} & \cdots & \sum x_{i,r} & 1 & 0 & 0 & \cdots & 0 \\ \sum x_{i,1} & \sum (x_{i,1})^2 & \sum x_{i,1}\,x_{i,2} & \cdots & \sum x_{i,1}\,x_{i,r} & 0 & 1 & 0 & \cdots & 0 \\ \sum x_{i,2} & \sum x_{i,1}\,x_{i,2} & \sum (x_{i,2})^2 & \cdots & \sum x_{i,2}\,x_{i,r} & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \cdots & \cdots & \ddots & \cdots \\ \sum x_{i,r} & \sum x_{i,1}\,x_{i,r} & \sum x_{i,2}\,x_{i,r} & \cdots & \sum (x_{i,r})^2 & 0 & 0 & 0 & \cdots & 1 \end{array}\right]$$

Perform row operations until when the left hand side becomes an identity matrix. The row operations performed will also turn the right hand matrix into the inverse of the main matrix.

$$[I|(X^TX)^{-1}] = \left[\begin{array}{ccccc|ccccc} 1 & 0 & 0 & \cdots & 0 & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 1 & 0 & \cdots & 0 & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 1 & \cdots & 0 & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \ddots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & \cdots & \cdots & \cdots & \cdots & \cdots \end{array}\right]$$

Now, multiply the transpose of the design matrix, by the Y vector.

$$X^TY = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{1,1} & x_{2,1} & \cdots & x_{m,1} \\ x_{1,2} & x_{2,2} & \cdots & x_{m,2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,r} & x_{2,r} & \cdots & x_{m,r} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

$$X^TY = \begin{bmatrix} \sum y_i \\ \sum x_{i,1}\,y_i \\ \sum x_{i,2}\,y_i \\ \vdots \\ \sum x_{i,r}\,y_i \end{bmatrix}$$

Now, there is enough information to solve the β vector:

$$\beta = (X^TX)^{-1}(X^TY) = \cdots = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_r \end{bmatrix}$$

2.2 Predicted Y

The predicted Y vector, denoted by $\widehat{Y}$, contains estimations of the true values in Y.

$$\widehat{Y} = X\beta \tag{4}$$

To find predicted Y, multiply the design matrix by the beta vector.

$$\widehat{Y} = X\beta = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,r} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m,1} & x_{m,2} & \cdots & x_{m,r} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_r \end{bmatrix}$$

$$\widehat{Y} = X\beta = \begin{bmatrix} \beta_0 + \beta_1 x_{1,1} + \beta_2 x_{1,2} + \cdots + \beta_r x_{1,r} \\ \beta_0 + \beta_1 x_{2,1} + \beta_2 x_{2,2} + \cdots + \beta_r x_{2,r} \\ \vdots \\ \beta_0 + \beta_1 x_{m,1} + \beta_2 x_{m,2} + \cdots + \beta_r x_{m,r} \end{bmatrix}$$

$$\widehat{Y} = X\beta = \begin{bmatrix} \widehat{y_1} \\ \widehat{y_2} \\ \vdots \\ \widehat{y_m} \end{bmatrix}$$

Note that, since X$\beta = \widehat{Y}$, the regression equation can be rewritten as:

$$Y = \widehat{Y} + \varepsilon \tag{5}$$

The expanded form of this equation is:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} \widehat{y_1} \\ \widehat{y_2} \\ \vdots \\ \widehat{y_m} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}$$

2.3 Error Term

The error term is subtracting predicted Y values from the true Y values.

$$\varepsilon = Y - \widehat{Y} \tag{6}$$

The expanded form of this equation is:

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} - \begin{bmatrix} \widehat{y_1} \\ \widehat{y_2} \\ \vdots \\ \widehat{y_m} \end{bmatrix}$$

3 Example

To demonstrate, we will use a small example dataset.

3.1 Dataset

Note that in the table below, X_{2a} is linearly dependent on X_1. In statistics, this is called full multicollinearity. Due to this, the columns of X^TX won't be fully linearly independent, so it won't be invertible. This is a hindrance for finding the beta vector. (this is explained later in section 4)

Dataset =

Y	X_1	X_{2a}	X_{2b}
3	1	2	2
5	2	4	4
7	3	6	6
9	4	8	8
12	5	10	10
15	6	12	1

Therefore, we will use X_{2b} instead of X_{2a} for the design matrix. (note that in statistics, there is still some multicollinearity, or high correlation, between X_1 and X_{2b}, which means the beta coefficients may not be the most reliable. However, it can now be calculated from a pure mathematical perspective.)

3.2 Calculate the Beta Coefficients

First, find X^TX.

$$X^TX = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 4 & 6 & 8 & 10 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 4 \\ 1 & 3 & 6 \\ 1 & 4 & 8 \\ 1 & 5 & 10 \\ 1 & 6 & 1 \end{bmatrix}$$

$$X^TX = \begin{bmatrix} 6 & 21 & 31 \\ 21 & 91 & 116 \\ 31 & 116 & 221 \end{bmatrix}$$

Now, augment the matrix with an identity matrix on the right hand side.

$$[X^TX|I] = \left[\begin{array}{ccc|ccc} 6 & 21 & 31 & 1 & 0 & 0 \\ 21 & 91 & 116 & 0 & 1 & 0 \\ 31 & 116 & 221 & 0 & 0 & 1 \end{array}\right]$$

Next, perform row operations until there is an identity matrix on the left hand side.

$$R_1 = R_1 \div 6 \Rightarrow \left[\begin{array}{ccc|ccc} 1 & 7/2 & 31/6 & 1/6 & 0 & 0 \\ 21 & 91 & 116 & 0 & 1 & 0 \\ 31 & 116 & 221 & 0 & 0 & 1 \end{array}\right]$$

$$\begin{array}{c} R_2 = R_2 - 21 \times R_1 \Rightarrow \\ R_3 = R_3 - 31 \times R_1 \Rightarrow \end{array} \left[\begin{array}{ccc|ccc} 1 & 7/2 & 31/6 & 1/6 & 0 & 0 \\ 0 & 35/2 & 15/2 & -7/2 & 1 & 0 \\ 0 & 15/2 & 365/6 & -31/6 & 0 & 1 \end{array}\right]$$

$$R_2 = R_2 \div (35/2) \Rightarrow \left[\begin{array}{ccc|ccc} 1 & 7/2 & 31/6 & 1/6 & 0 & 0 \\ 0 & 1 & 3/7 & -1/5 & 2/35 & 0 \\ 0 & 15/2 & 365/6 & -31/6 & 0 & 1 \end{array}\right]$$

$$\begin{array}{c} R_1 = R_1 - (7/2) \times R_2 \Rightarrow \\ \\ R_3 = R_3 - (15/2) \times R_2 \Rightarrow \end{array} \left[\begin{array}{ccc|ccc} 1 & 0 & 11/3 & 13/15 & -1/5 & 0 \\ 0 & 1 & 3/7 & -1/5 & 2/35 & 0 \\ 0 & 0 & 1210/21 & -11/3 & -3/7 & 1 \end{array}\right]$$

$$R_3 = R_3 \div (1210/21) \Rightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 11/3 & 13/15 & -1/5 & 0 \\ 0 & 1 & 3/7 & -1/5 & 2/35 & 0 \\ 0 & 0 & 1 & -7/110 & -9/1210 & 21/1210 \end{array}\right]$$

$$\begin{array}{c} R_1 = R_1 - (11/3) \times R_3 \Rightarrow \\ R_2 = R_2 - (3/7) \times R_3 \Rightarrow \end{array} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 11/10 & -19/110 & -7/110 \\ 0 & 1 & 0 & -19/110 & 73/1210 & -9/1210 \\ 0 & 0 & 1 & -7/110 & -9/1210 & 21/1210 \end{array}\right]$$

Now, there is an identity matrix on the left-hand side. The matrix on the right hand side is the inverse of $X^T X$.

$$(X^T X)^{-1} = \begin{bmatrix} 11/10 & -19/110 & -7/110 \\ -19/110 & 73/1210 & -9/1210 \\ -7/110 & -9/1210 & 21/1210 \end{bmatrix}$$

Now, find $X^T Y$.

$$X^T Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 4 & 6 & 8 & 10 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 7 \\ 9 \\ 12 \\ 15 \end{bmatrix}$$

$$X^T Y = \begin{bmatrix} 51 \\ 220 \\ 275 \end{bmatrix}$$

Solve for the beta vector.

$$\beta = (X^T X)^{-1} X^T Y = \begin{bmatrix} 11/10 & -19/110 & -7/110 \\ -19/110 & 73/1210 & -9/1210 \\ -7/110 & -9/1210 & 21/1210 \end{bmatrix} \begin{bmatrix} 51 \\ 220 \\ 275 \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 3/5 \\ 133/55 \\ -6/55 \end{bmatrix} \approx \begin{bmatrix} 0.6 \\ 2.418 \\ -0.109 \end{bmatrix}$$

3.3 Predicted Y and Error term

Multiply design matrix by beta vector to find predicted Y.

$$X\beta = \widehat{Y} = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 4 \\ 1 & 3 & 6 \\ 1 & 4 & 8 \\ 1 & 5 & 10 \\ 1 & 6 & 1 \end{bmatrix} \begin{bmatrix} 3/5 \\ 133/55 \\ -6/55 \end{bmatrix}$$

$$\widehat{Y} = \begin{bmatrix} \widehat{y_1} \\ \widehat{y_2} \\ \widehat{y_3} \\ \widehat{y_4} \\ \widehat{y_5} \\ \widehat{y_6} \end{bmatrix} = \begin{bmatrix} 2.8 \\ 5 \\ 7.2 \\ 9.4 \\ 11.6 \\ 15 \end{bmatrix}$$

Now, solve for the error term.

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix} = Y - \widehat{Y} = \begin{bmatrix} 3 \\ 5 \\ 7 \\ 9 \\ 12 \\ 15 \end{bmatrix} - \begin{bmatrix} 2.8 \\ 5 \\ 7.2 \\ 9.4 \\ 11.6 \\ 15 \end{bmatrix}$$

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0 \\ -0.2 \\ -0.4 \\ 0.4 \\ 0 \end{bmatrix}$$

The equation is now:

$$Y \approx 0.6 + 2.418X_1 - 0.109X_2 \tag{7}$$

Note: the average error term is zero

4 Clarification on Linear Dependency

If we use X_{2a} for the design matrix:

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 4 & 6 & 8 & 10 & 12 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 4 \\ 1 & 3 & 6 \\ 1 & 4 & 8 \\ 1 & 5 & 10 \\ 1 & 6 & 12 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 6 & 21 & 42 \\ 21 & 91 & 182 \\ 42 & 182 & 364 \end{bmatrix}$$

It is evident that the 3rd column of $X^T X$ is 2 times the 2nd column. This is linear dependency, and as such it cannot correspond to an identity matrix of the same size. This is the danger of full multicollinearity.

5 Citations

Pennsylvania State University. 5.4 - A Matrix Formulation of the Multiple Regression Model — STAT 462. *Penn State Online*. https://online.stat.psu.edu/stat462/node/132/.

Montgomery, D. 3.2 - Estimation of the Model Parameters. *Introduction to Linear Regression Analysis* (4th ed., pp. 68–69).