The Multiple Linear Regression Model

Let's review the simple linear regression model.

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Y is the response variable, X is the predictor variable, β_0 is the intercept parameter, β_1 is the slope parameter, and ε is the error term representing the variation of Y around the line of best fit. The regression line is the mean of Y at any given X, which equals $\beta_0 + \beta_1 X$.

The model for multiple regression is similar. When you have two predictor variables, you model the relationship of the three variables - three dimensions - with a two-dimensional plane. Let's look at a model with two predictors.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

Y is the response variable, X_1 and X_2 are the predictor variables, ε is the error term, and β_0, β_1, and β_2 are unknown parameters.

β_0 is the y-intercept, and has the same meaning as the intercept in a simple linear regression. It's the value of Y when the predictors are equal to 0. The slopes in a multiple regression have a somewhat different interpretation than they did when there was only one predictor. The slopes or regression coefficients describe the average change in Y for a one-unit change in X, that is, β_1 is the average change in Y for a one-unit change in X_1, holding X_2 constant, and β_2 is the average change in Y for a one-unit change in X_2, holding X_1 constant.

If there's no relationship between Y, X_1, and X_2, that is, the slopes β_1 and β_2 equal 0, the model is a horizontal plane passing through the point where Y equals β_0.

When there is a linear relationship between Y, X_1, and X_2, the model is a sloping plane. In this case, X_1, X_2, or both affect Y, so the plane tilts.

In a multiple regression model, you model the response variable, Y, as a linear function of the k predictor variables, X_1 through X_k. You investigate the relationship among the k predictors and the response using a k dimensional surface for prediction. The model has k + 1 parameters, the regression coefficient slopes, and the intercept.

You can also use linear regression to model non-linear relationships with the response variable by adding polynomials, such as squared or cubed terms, or you can add interactions to your model. If the polynomial model has the predictors X_1, X_1^2, X_2, and X_2^2, but it's still a linear model despite the exponents on the predictor variables. That's because these polynomial models are linear in the parameters.