Multiple Linear Regression Analysis With Constraints

written by Juan Cervantes | October 4, 2024

Multiple linear regression is a statistical technique used to model the relationship between a dependent variable and two or more independent variables. It extends simple linear regression by allowing multiple factors to influence the outcome, making it useful for predicting or understanding the effect of several variables simultaneously. The model takes the form

$$
y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n
$$

where y is the dependent variable, $x_1, x_2, \dots, x_n$ are the independent variables, and $b_0, b_1, \dots, b_n$ are the coefficients to be estimated.

With constraints

Here's how to approach the problem for multiple linear regression analysis with constraints:

Suppose the linear model is of the form $y = b_1x_1 + b_2x_2 + b_3x_3$, subject to the constraint $b_1 + b_2 + b_3 = 1$.

To work with this constraint, we can re-express b_3 as $b_3 = 1 - b_1 - b_2$. This means the model can be written as:

$$
y = x_3 + b_1 \cdot x_1 + b_2 \cdot x_2 + (1 - b_1 - b_2) \cdot x_3
$$

Simplifying, we get:

$$
y - x_3 = b_1 \cdot (x_1 - x_3) + b_2 \cdot (x_2 - x_3)
$$

Now, define new variables:
$$
y' = y - x_3, \quad x_1' = x_1 - x_3, \quad x_2' = x_2 - x_3
$$
With these transformations, you can perform a standard linear regression using y' as the dependent variable and x_1' and x_2' as the independent variables.

Note: Excel has a built-in tool called 'Solver' that might seem applicable for maximizing the R^2 value under the constraint $b_1 + b_2 + b_3 = 1$. However, this approach won't work effectively for regression problems because Solver is designed for linear programming, not regression analysis. Instead, it's best to use the transformed variables for regression directly.