Multiple Regression

Multiple regression estimates the outcomes (dependent variables) which may be affected by more than one control parameter (independent variables) or there may be more than one control parameter being changed at the same time.

An example is the two independent variables x and y and one dependent variable z in the linear relationship case:

$$z = a + bx + cy$$

For a given data set (x_1, y_1, z_1), (x_2, y_2, z_2), ..., (x_n, y_n, z_n), where $n \geq 3$, the best fitting curve $f(x)$ has the least square error, i.e.,

$$\Pi = \sum_{i=1}^{n} [z_i - f(x_i, y_i)]^2 = \sum_{i=1}^{n} [z_i - (a + bx_i + cy_i)]^2 = \min.$$

Please note that a, b, and c are unknown coefficients while all x_i, y_i, and z_i are given. To obtain the least square error, the unknown coefficients a, b, and c must yield zero first derivatives.

$$\begin{cases} \dfrac{\partial \Pi}{\partial a} = 2\displaystyle\sum_{i=1}^{n} [z_i - (a + bx_i + cy_i)] = 0 \\ \dfrac{\partial \Pi}{\partial b} = 2\displaystyle\sum_{i=1}^{n} x_i [z_i - (a + bx_i + cy_i)] = 0 \\ \dfrac{\partial \Pi}{\partial c} = 2\displaystyle\sum_{i=1}^{n} y_i [z_i - (a + bx_i + cy_i)] = 0 \end{cases}$$

Expanding the above equations, we have

$$\begin{cases} \displaystyle\sum_{i=1}^{n} z_i = a\sum_{i=1}^{n} 1 + b\sum_{i=1}^{n} x_i + c\sum_{i=1}^{n} y_i \\ \displaystyle\sum_{i=1}^{n} x_i z_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i y_i \\ \displaystyle\sum_{i=1}^{n} y_i z_i = a\sum_{i=1}^{n} y_i + b\sum_{i=1}^{n} x_i y_i + c\sum_{i=1}^{n} y_i^2 \end{cases}$$

The unknown coefficients a, b, and c can hence be obtained by solving the above linear equations.