Multiple Regression Analysis

Multiple regression analysis is a powerful technique used for predicting the unknown value of a variable from the known value of two or more variables- also called the predictors.

More precisely, multiple regression analysis helps us to predict the value of Y for given values of X_1, X_2, …, X_k.

> For example the yield of rice per acre depends upon quality of seed, fertility of soil, fertilizer used, temperature, rainfall. If one is interested to study the joint affect of all these variables on rice yield, one can use this technique.
>
> An additional advantage of this technique is it also enables us to study the individual influence of these variables on yield.

Dependent and Independent Variables

By multiple regression, we mean models with just one dependent and two or more independent (exploratory) variables. The variable whose value is to be predicted is known as the dependent variable [1] and the ones whose known values are used for prediction are known independent [2] (exploratory) variables.

The Multiple Regression Model

In general, the multiple regression equation of Y on X_1, X_2, …, X_k is given by:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + …………………… + b_k X_k$$

Interpreting Regression Coefficients

Here b_0 is the intercept and b_1, b_2, b_3, …, b_k are analogous to the slope in <u>linear regression</u> [3] equation and are also called regression coefficients. They can be interpreted the same way as slope. Thus if b_i = 2.5, it would indicates that Y will increase by 2.5 units if X_i increased by 1 unit.

The appropriateness of the multiple regression model as a whole can be tested by the F-test in the <u>ANOVA</u> [4] table. A significant F indicates a linear relationship between Y and at least one of the X's.

How Good Is the Regression?

Once a multiple regression equation has been constructed, one can check how good it is (in terms of predictive ability) by examining the coefficient of determination (R2). R2 always lies between 0 and 1.

R_2 - coefficient of determination

All software provides it whenever regression procedure is run. The closer R_2 is to 1, the better is the model and its prediction.

A related question is whether the independent variables individually influence the dependent variable significantly. Statistically, it is equivalent to <u>testing</u> [5] the <u>null hypothesis</u> [6] that the relevant regression coefficient is zero.

This can be done using t-test. If the t-test of a regression coefficient is significant, it indicates that the variable is in question influences Y <u>significantly</u> [7] while controlling for other independent explanatory variables.

Assumptions

Multiple regression technique does not test whether data are <u>linear</u> [8]. On the contrary, it proceeds by assuming that the relationship between the Y and each of X_i's is linear. Hence as a rule, it is prudent to always look at the scatter plots of (Y, X_i), i= 1, 2,…,k. If any plot suggests <u>non linearity</u> [9], one may use a suitable transformation to attain linearity.

Another important assumption is non existence of multicollinearity- the independent variables are not related among themselves. At a very basic level, this can be tested by computing the correlation coefficient between each pair of independent variables.

Other assumptions include those of homoscedasticity and normality.

Multiple regression analysis is used when one is interested in predicting a continuous dependent variable from a number of independent variables. If dependent variable is dichotomous, then logistic regression should be used.

Source URL: https://explorable.com/multiple-regression-analysis

Links
[1] https://explorable.com/dependent-variable
[2] https://explorable.com/independent-variable
[3] https://explorable.com/linear-regression-analysis
[4] https://explorable.com/anova
[5] https://explorable.com/hypothesis-testing
[6] https://explorable.com/null-hypothesis
[7] https://explorable.com/statistically-significant-results
[8] https://explorable.com/linear-relationship
[9] https://explorable.com/non-linear-relationship