

OLS, Probit, Logit, Logistic Regression and Discriminant Analysis

Dr. J.C. Sharma*

In research it is a common situation where one has to predict a dependent variable from one or more independent variables. Since the type of independent and dependent variables have different levels of measurement viz; nominal (categorical), ordinal (categorical or ranking or rating), interval and ratio levels, the technique to be used for fitting such models will also vary. The researcher must use an appropriate technique which will suit to the type of data (variables).

Examples of categorical variables are 'user vs nonuser of online purchase', 'defaulter vs non-defaulter of loan', 'low motivated vs high motivated' etc. The examples of ordinal variables are 'rating scores in Likert scale', 'degree of job satisfaction' etc. The examples of interval and ratio scale variables are family size, number of days of absence from work, expenditure on medical care, rate of interest etc. Depending on a combination of dependent and independent variables one has to use appropriate statistical technique of regression analysis. Various techniques like Ordinary Least Square Regression (OLS), Logistic regression (Logit), Probit regression and discriminant analysis are available for use. A brief introduction to these techniques is given below.

Ordinary least-squares (OLS) regression, a commonly used technique, is a generalized linear modeling technique that is used to predict a single response variable (dependent variable) which has been recorded on at least an interval scale. The technique may be applied to single or multiple explanatory (independent) variables and also categorical explanatory variables that have been appropriately coded (called as dummy variables).

Probit Analysis is a specialized regression model of binomial response variables. Probit regression, also called a probit model, is used to predict dichotomous or binary outcome (dependent) variables. In the probit model, the inverse standard normal distribution of the probability is modeled as a linear combination of the predictors (independent variables).

Logistic regression, or logit regression, is a type of probabilistic classification model. It is also used to predict a binary response from a binary predictor used for predicting the outcome of a categorical dependent variable (i.e., a class label) based on one or more predictor variables (features). That is, it is used in estimating the parameters of a qualitative response model. Logistic regression measures the relationship between a categorical dependent variable and one or more independent variables, which are usually (but not necessarily) continuous, by using probability scores as the predicted values of the dependent variable. As such it treats the same set of problems as probit by using similar techniques.

* Executive Editor, BVIMSR's Journal of Management Research (BJMR), Mumbai

Discriminant function analysis is a statistical analysis to predict a categorical dependent variable (called a grouping variable) by one or more continuous or binary independent variables (called predictor variables). It is different from an ANOVA or MANOVA which is used to predict one (ANOVA) or multiple (MANOVA) continuous dependent variables by one or more independent categorical variables. Discriminant function analysis is useful in determining whether a set of variables is effective in predicting category membership.

There are many variants in all the above techniques of analysis which can be fitted to any given situation. These techniques are available in popular statistical soft-wares like SPSS, SAS and STATA etc.