

REGRESSION ANALYSIS

Regression analysis is a technique *used for the modeling and analysis of numerical data consisting of values of a dependent variable (response variable) and of one or more independent variables (explanatory variables)*. The dependent variable in the regression equation is modeled as a function of the independent variables, corresponding parameters ("constants"), and an error term. *The error term is treated as a random variable*. It represents unexplained variation in the dependent variable. The parameters are estimated so as to give a "best fit" of the data. Most commonly the best fit is evaluated by using the least squares method, but other criteria have also been used.

Data modeling can be used without there being any knowledge about the underlying processes that have generated the data, in this case the model is an empirical model. Moreover, in modelling, knowledge of the probability distribution of the errors is not required. *Regression analysis requires assumptions to be made regarding probability distribution of the errors*. Statistical tests are made on the basis of these assumptions. In regression analysis the term "model" embraces both the function used to model the data and the assumptions concerning probability distributions.

Regression can be used for prediction (including forecasting of time-series data), inference, hypothesis testing, and modeling of causal relationships. These uses of regression rely heavily on the underlying assumptions being satisfied. Regression analysis has been criticized as being misused for these purposes in many cases where the appropriate assumptions cannot be verified to hold. One factor contributing to the misuse of regression is that it can take considerably more skill to critique a model than to fit a model.