Regression model

A regression model is a statistical model that is used to predict a continuous outcome variable (also known as the dependent variable) based on one or more predictor variables (also known as independent variables, covariates, or features). Regression models are used in many fields, including economics, finance, biology, and engineering, to understand the relationship between variables and make predictions. There are several types of regression models, including linear regression, logistic regression, and polynomial regression, among others. The choice of regression model depends on the nature of the data and the research question being addressed.

Choosing the correct linear regression model can be difficult. After all, the world and how it works is complex. Trying to model it with only a sample doesn't make it any easier.

It starts when a researcher wants to mathematically describe the relationship between some predictors and the response variable. The research team tasked to investigate typically measures many variables but includes only some of them in the model. The analysts try to eliminate the variables that are not related and include only those with a true relationship. Along the way, the analysts consider many possible models.

They strive to achieve a Goldilocks balance with the number of predictors they include.

Too few: An underspecified model tends to produce biased estimates.

Too many: An overspecified model tends to have less precise estimates.

Just right: A model with the correct terms has no bias and the most precise estimates.

Statistical Methods for Finding the Best Regression Model

For a good regression model, you want to include the variables that you are specifically testing along with other variables that affect the response in order to avoid biased results.

see Generalized additive mixed model