

Overview

Logistic regression is used to model the relationship between a binary response variable and a set of predictor variables. It's used to estimate the probability of the response according to the various continuous and categorical predictors. The estimated probabilities can then be used to classify an unknown response into one of the two outcome levels, given a set of predictors.

For example, imagine you're studying the sinking of the Titanic, and you want to model the relationship between survival and passenger characteristics. That is, you want to find the probability of Titanic passengers surviving the crash given their ticket fare, age, and gender. You could use logistic regression to confirm or deny that the probability of survival is in fact higher for young female passengers traveling in top notch accommodations.

Instead of characterizing the relationship between response and predictors, imagine instead you work at a credit card company and you want to predict credit card fraud. You can use logistic regression to estimate the probability of a transaction as being fraudulent, given predictors such as time of day, type of transaction, and region of purchase. You could build a model on historical data and then score new transactions and classify them as fraudulent or not.

In this lesson, we'll first look for associations between predictors and a binary response using hypothesis tests. We'll then build a logistic regression model and discuss how to characterize the relationship between the response and predictors. Finally, we'll use logistic regression to build a model, or classifier, to predict unknown cases.

Statistics 1: Introduction to ANOVA, Regression, and Logistic Regression

Copyright © 2019 SAS Institute Inc., Cary, NC, USA. All rights reserved.

Close