

Logistic Regression Analysis

Kelly Wahl

Director of Assessment and Data Analysis

Loyola Marymount University

Automobile insurance companies create statistical models to represent the chance that prospective clients will avoid being involved in an accident, considering a variety of variables, which might include demographics, previous driving habits, and driver training courses attended. Using this technique, I can create a logistic regression model to “predict” whether students will demonstrate a variety of target outcomes, and I include personal characteristics, academic indicators, and other variables, such as participation in campus programs, use of services, or data points collected on a variety of survey instruments. Potential target outcomes include learning a body of knowledge, mastering a particular skill, or acquiring (or strengthening) a personal value.

When considering the results of a model that accounts for as much variance as possible (i.e., a model that explains why students in some groups demonstrate the target outcome while others aren't, taking into account as many characteristics as possible), the research isolates a Baseline Group, and then percentages are assigned to each characteristic (or experience, or other descriptor) to determine how adding that trait to an individual's profile would affect the predicted probability of the target outcome for that individual. The model strives to predict the target outcome for every group that can be described by each permutation of the characteristics considered, and as a consequence, it estimates the probability of the target outcome for each individual in these groups.

For example, in a recent logistic regression analysis I performed for an academic program on campus, the Baseline Group, which is described in the constant of the statistical model's equation, is female, is non-Catholic, has a local area code for their permanent residence phone number, is white, is below the median on high school GPA, SAT Verbal, and SAT Math, has a declared major, and is not in the academic program in question. This Baseline Group's probability of demonstrating the target outcome is 87.7%. Only students submitting both high school GPAs and SAT scores can be included in the model.

If we changed the Baseline Group's religion to Catholic, the target outcome probability for the group would rise to 89.4% (87.7% plus 1.7%). If we made them both Catholic and male, this probability would drop to 87.6% (87.7% plus 1.7% minus 1.8%). Whenever we add the characteristic “participation in the academic program” to the group, we raise the group's target outcome probability by 1.9%.

In models such as this one, with limited information to be associated with the target outcome, the variance the model can describe is relatively small (typically under 10.0%); however, the model does suggest the comparative impact of personal characteristics, success in prior school work, and participation in the academic program. Who the students are and what they are capable of doing (as measured by the demographics and the SAT scores, respectively) seemed to have an influence on the outcome. Most demographics, however, seem to have a smaller positive impact (or a greater negative impact) than both how well the students did in high school and whether they participate in the academic program.