



SOCIETY OF ACTUARIES

Article from:

Product Matters!

October 2013 – Issue 87

Danger in Predictive Models for Underwriting

by Chris Davis



Chris Davis, FSA, MAAA, is Vice President at Lewis & Ellis Inc in Overland Park, Kan. He can be reached at cdavis@lewisellis.com.

The purpose of this article is to alert the reader to a potential danger in using predictive models as part of the underwriting process. The danger lies in the possibility of unintentional illegal discrimination. Actuarial presentations tend to explain predictive modeling in various levels of detail based solely on the mathematics involved but fail to adequately address the concept of prohibited variables. There are other legal dangers such as privacy concerns, but this article will focus solely on the danger of illegal discrimination. It should be kept in mind that the author is not a lawyer, and the views expressed herein are solely those of the author, not necessarily those of his employer.

Most legal actions concerning illegal discrimination in the underwriting process have been associated with industrial insurance and race. Applications from decades ago contained a race question. Later applications omitted the race question but used questions about socio-economic status. In some cases, courts concluded that underwriting on the basis of socio-economic status was a proxy for race.

Predictive modeling is a process of creating a statistical model that tries to best predict the most likely outcome based on the values of predictor variables. In particular, it can be used to identify those insureds or policies that have attributes that cause life companies to take a certain action.

Some examples of possible uses of predictive models in the life insurance industry are in the areas of marketing, underwriting, fraud detection, modeling for asset adequacy analysis, and conservation of policies by identifying those likely to lapse.

Predictive modeling focuses on the statistical differences implied by certain predictor variables. Our society has laws that prohibit the use of certain variables in the underwriting of life insurance policies. Examples include race and, in some cases, sex. Predictive models can easily exclude prohibited variables. The concern is that other variables, separately or in combination, included in the model could possibly be deemed by a court to be a proxy for a prohibited variable. When there is a statistical difference among values of a prohibited variable, e.g., males have higher mortality rates than females, the model is likely to include some variables that distinguish this difference.

For example, the predictive model excludes all prohibited variables but may use variables that describe personal habits and geographic location of residence in the underwriting process. The life insurance company intends to not use a prohibited variable in the underwriting process and is unaware of the variables actually used, or at least does not realize a possibility that they are, in effect, using a proxy for a prohibited variable. Variables that describe the applicant's magazine subscriptions could be considered to be a proxy for a prohibited variable if the targeted audience of the magazines is highly correlated with a particular value of the prohibited variable.

Consider the example in which the company sets up the underwriting process and takes the position that the predictive model is used solely to classify the applicant as to whether or not more information is needed. The company claims that this results in no adverse action



for any applicant and that the model could only result in a positive action by streamlining the underwriting process. This could be debatable.

Suppose we have two values of a prohibited variable. Those characterized by a specific value will be referred to as group A, and those characterized by the other value will be referred to as group B. The statistical technique called discriminant analysis can be used to make a statistical determination as to whether one group is being charged more than another for life insurance.

Discriminant analysis distinguishes between two or more groups of data based on a set of input variables. One input variable is a classification factor that identifies which group each observation belongs to. It distinguishes by constructing discriminant functions that are linear combinations of the variables. The objective of the analysis is to be able to describe observed cases mathematically in a manner that separates them into groups as well as possible. In constructing the discriminant functions, the procedure allows inclusion of all of the variables or a stepwise selection procedure that includes only those variables that are statistically significant discriminators amongst the groups. Statistical summaries and tests of significance for the number of discriminant functions needed are performed.

Is either group A or B being charged more than the other group? Note that it is possible that the group with the lower mortality could be charged more and be able to claim illegal discrimination. It seems plausible to determine the existence of illegal discrimination based on the premiums (or mortality charges) determined rather than the input variables used. All insureds considered should be included rather than just those actually issued a policy as discrimination could have been an unintentional and unknowing factor in an applicant's decision to decline the policy.

There is a need to correct for all variables that are well-accepted for distinguishing premium rates and upon which the premiums do vary, such as age, tobacco usage, and the effects of banding and policy fees. If the variable can help in distinguishing between groups

A and B, then a separate analysis should be performed for each value of that variable; otherwise, that variable should be used as an input variable to the discriminant analysis. The other input variables are the classification factor that identifies group A or B and the premium rate.

If the discriminant analysis shows that an applicant can be correctly classified into group A or B with a very high probability, then there is a proxy to the prohibited variable being used. This can be interpreted as strong evidence of illegal discrimination based on this prohibited variable in the form of a proxy to it.

Companies should bring the concern of some of their variables possibly serving as a proxy to an illegal discriminatory variable to the attention of their legal counsel before using predictive modeling as a part of the underwriting function. ■