SOCIETY OF ACTUARIES

Article from:

ARCH 2013.1 Proceedings

August 1- 4, 2012

Michael V. Loginov, Emily Marlow, Victoria Potruch

Predictive Modeling in Healthcare Costs using Regression Techniques

Michael V. Loginov[1], Emily Marlow[2] and Victoria Potruch[3]

UC Santa Barbara, Santa Barbara, USA
[1]mloginov@umail.ucsb.edu
[2]ecmarlow@umail.ucsb.edu
[3]vpotruch@gmail.com

After President Obama signed the Patient Protection and Affordable Care Act, as well as the Health Care and Education Reconciliation Act of 2010, the United States healthcare system became subject to extensive changes in patient eligibility requirements. The Congressional Budget Office (2010) projected that these acts will reduce the number of uninsured nonelderly people by about 32 million, with only 23 million nonelderly citizens remaining uninsured by 2019. Using data from an insurance provider, we investigated healthcare expenditures and applied variable selection to produce models for predicting the cost that insurers will spend on covered members. Statistical techniques such as data transformation, linear regression, and least angle regression were used to develop the models. We pinpointed the effect of high cost factors and developed a method of predicting a member's cost to the insurer by analyzing claims cost data. In this project, we used a dataset provided by a private insurance company which contained demographic, cost, and diagnostic data. Our analysis allowed us to determine which factors drove costs up and how these factors will affect the total cost for future years. We used linear regression and least angle regression (LAR) techniques to develop the models. We also implemented the special modification least absolute shrinkage and selection operator (lasso) to reduce the number of independent variables for our predictive model. These statistical techniques were performed using the open source statistical software R, including the lars package that implements LAR and lasso. Using Lars, we picked 13 key predictors for our model for determining the cost of insuring an individual, while maintaining an adjusted R^2 of 0.4 which is consistent with the models published by the Society of Actuaries.