

Classification of the Insurance Sector with Logistic Regression

BAHADTIN RUZGAR*, NURSEL SELVER RUZGAR**

*Banking and Insurance School, Actuaries Department, Marmara University
Goztepe Campus, Kadikoy, 34722, Istanbul, TURKEY
bruzgar@marmara.edu.tr, <http://mimoza.marmara.edu.tr/~bruzgar/>

**Vocational School of Social Sciences, Marmara University
Beyazit Campus, Eminonu, Istanbul, TURKEY
nruzgar@marmara.edu.tr

Abstract: - In cases when categorical results such as successful-unsuccessful, ill-not ill, good-fair-bad are obtained especially as a result of evaluation of data, the logistic regression is a rather suitable statistical method. In this study, the data of 53 companies that are active in the insurance sector in Turkey for 2004, 2005 and 2006 years were evaluated by using logistic regression method. However, as the data were not sufficient for all the insurance companies, 12 insurance companies were eliminated from the evaluation. 41 companies used for the analysis were divided into two groups depending on their activity area. 17 companies were evaluated by using the data on individual accident, health and life branches and 24 companies by using data on fire, transportation, engineering, agriculture, all-risks, obligatory traffic, obligatory highway transportation, individual accident and other accident and health branches.

Key-Words: - Logistic regression, classification, discriminant analysis, insurance companies classification

1 Introduction

The researchers or model designers always endeavor to convert the data they obtain from real events or experiments to functional structures by means of various models. Though to establish mathematical models is rather difficult, they ensure very beneficial information. Classification of the data used in models constitutes the very important part of the statistical analysis but it is widely used by various science branches mainly in health. Some studies about logistic regression are; the comparison of the mobile nursing system that was established between the years of 1977-1985 in America to restrict the health expenses to the former system was examined with multi logistic regression analysis [1]. In America, the data obtained from extraordinary events such as wars, elections, political crisis and epidemic diseases were used to determine differences between the periods when such events occurred and other periods by means of logistic regression [2]. Binary logistic regression was used to calculate the retirement age of people depending on age, sex, economical and social statuses [3]. Between the years of 1980-1995, the data of bankrupted American companies were examined, 237 of the bankrupted companies were handled as samples in 1992 and the financial and non-financial values of their final bankruptcy resolutions were examined with logistic analysis and their classifications were tried to be estimated [4]. Again, in respect to health insurance, the health insurance classification of insured and uninsured low-income children in America between the years of 1995-1999 and the classification of uninsured ones

according their sex, age and economical status were made by using the logistic regression model [5]. The national health researches of the Australian households were made by using 2001 data and their passing rate to private health insurance and the reasons of such passing such as economical, social and health factors were examined by means of multi logistic regression analysis [6]. In classification of car accidents in America, the logistic regression analysis was made by using the variables such as literacy rate, economical status and sex [7]. 57 big parent companies that were very important for the Japan economy between the years of 1998-2001 were classified as "financially under stress" and "peaceful" [8]. To determine whether the Treatment Center established for purpose of treating visual disabled or blind people in order they could find new jobs was beneficial for such people, logistic modeling was used [9]. Between the years of 1980-2004, the disability risk and disability risk insurances were examined in America and a classification was made by using the logistic regression according to workability limits, non-workability situations and the need to get health aid of people who retired for reason of a physical disability [10]. In another study, the logistic regression was used in determination and classification of car insurance tariffs of insurance companies [11]. When theoretical studies related to the logistic regression are viewed, it can be determined that the widely used logistic regression models and development of coefficient estimation methods have caused logistic regression models to be examined in a more detailed manner. In the logistic