Journal of Applied Probability and Statistics
Vol. 6, No. 1&2, pp. 47-60
ISOSS Publications 2012

A COMPARATIVE STUDY FOR THE USE OF ROC ANALYSIS FOR VARIABLE FOR VARIABLE SELECTION IN MEDICAL DATA

Christos Koukouvinos
Department of Mathematics, National Technical University of Athens, Zografou, Athens, 15773, Greece.
Email: ckoukouv@math.ntua.gr

Christina Parpoula
Department of Mathematics, National Technical University of Athens, Zografou, Athens, 15773, Greece.
Email: parpoula.ch@gmail.com

Abstract

In health studies, many potential factors are usually introduced to determine an outcome variable. In our health study, different statistical methods are applied to analyze trauma annual data. The Trauma data set used here was collected in an annual registry conducted during the period $01/01/2005 - 31/12/2005$ by the Hellenic Trauma and Emergency Surgery Society involving 30 General Hospitals in Greece. The dataset consists of 8862 observations and 92 factors that include demographic, transport and intrahospital data used to detect possible risk factors of death. The statistical methods employed in this work are the nonconcave penalized likelihood methods, SCAD, LASSO, and Hard, the maximum likelihood estimation method, the generalized linear logistic regression, the best subset variable selection, and the Receiver Operating Characteristic (ROC) analysis. The performance, the pros, and cons of these various statistical techniques for identifying the significant variables in large medical datasets are discussed.

Keywords: variable selection, maximum likelihood estimation, nonconcave penalized likelihood, modified BIC (mBIC), ROC, AUC, Trauma.

2000 Mathematics Subject Classification: 62J05, 62P10, 62-07.