

VARIABLE SELECTION IN REGRESSION USING- R

D. N. Kashid
Department of Statistics,
Shivaji University, Kolhapur
Dnkashid_in@yahoo.com

Abstract:

Variable selection problem is one of the important problems in regression analysis. Over the years, several variable selection methods are proposed in the literature and some frequently used methods are Mallows's Cp-statistic, Forward and Backward, Stepwise selection method etc. All these methods assume that the error distribution is normal and present software packages offer some of these methods for variable selection in regression. It is well known that in the absence of normality or absence of linearity assumption or outlier(s) presence in the data, the classical subset selection methods perform poorly. Such situations demand alternative approaches.

In the last decade, a few methods are developed in the literature based on different situation mentioned above. Ronchetti and Staudte (1994) have proposed robust version of Mallows's Cp called RCp for outlier data. Kashid and Kulkarni (2002, 2003) suggested variable selection techniques to deal the situation mentioned above.

Since these methods are computationally intensive, so it is difficult to select a set of variables without using the software. The implementation of these methods is possible by using R-software. In this article, we exploit use of R in variable selection problem in regression.

References:

- Kashid and Kulkarni (2002), A More General Criterion for Subset Selection in Multiple Linear Regression. *Communication in Statistics-Theory & Method*, 31(5), 795-811.
- Mallows's (1973), Some Comments on Cp. *Technometrics*, 15, 661-665.
- Ronchetti and Statutdte (1994), A Robust version of Mallows Cp. *JASA*, 89(246), 550-559.