Comparison of Confidence Interval in the OLS and Ridge Linear Regression Model: A Comparative study via Simulation

Sultana Mubarika Rahman Chowdhury[1], B.M Golam Kibria[2], Zoran Bursac[1]

ARTICLE HISTORY
Compiled August 28, 2024

[1] Biosatistics Department, Florida International University, Miami, Fl, USA
[2] Department of Mathematics and Statistics, Florida International University, Miami, Fl, USA

ABSTRACT
Write down the abstract here

KEYWORDS
Multiple Linear Regression; Ridge Regression; Multicolinearity

1. Introduction

Multiple linear regression maps the relationship between two or more predictors and dependent variable to a linear equation. The aim is to predict the response variable using the independent variables. If X is a n×p full rank matrix of predictors and Y is a n × 1 vector of response variables the multiple linear regression can be explained as,

$$Y = X \times \beta + \epsilon,$$

where, β is an p × 1 unknown regression paramaters and ϵ is the n × 1 vector of error with mean zero and equal variance.Ordinary Least Square (OLS) method is commonly used to estimate the unknown regression parameter. The OLS estimate in case of linear regression model is defined as follows,

$$\hat{\beta} = (X^T X)^{-1} X^T y.$$

One of the key assumption of the widely used multiple linear regression model is that the predictors need to be independent of each other. Violating the assumption of indepen-

CONTACT: Sultana Mubarika Rahman Chowdhury. Email: schow034@fiu.edu.

dence results in an issue known as multicollinearity. Multicollinearity occurs when two or more independent variables are highly correlated with each other. It causes standard error of the OLS coefficient to increase resulting in wide confidence intervals and less reliable results. Identifying and interpreting significant predictors also becomes difficult. As the field of study progressed, researchers developed a number of methods to address presence of multicollinearity in data. Ridge regression (Hoerl and Kennard 1970), Lasso,

Ridge regression is one of the methods proposed by Hoerl and Kennard (1970) that mitigates the effects of multicollinearity by introducing an amount of bias in to the model.

para about differnt techniques,

para about ridge

para about CI

para about Aim

Note:

2. Statistical Methodology

2.1. *Ridge Regression Estimators*

2.2. *Confidence Interval*

2.3. *CI for Ridge regression*

3. Simulation Study

4. Application

5. Discussion and Conclusion

Bibliography

Hoerl, Arthur E, and Robert W Kennard. 1970. "Ridge Regression: Biased Estimation for Nonorthogonal Problems." *Technometrics* 12 (1): 55–67.