Identifying Outlier Observations in Linear - Circular Regression Model

Seyede Sedighe Azimi*, Mohammad Reza Farid Rohani
Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran

Received: June 2018 Accepted: Dec. 2018

Extended Abstract

Paper pages (99-108)

Introduction

In statistical analysis, the issues that statisticians focus on are identifying outlier observations. An outlier observation may occur due to a measurement error or a real value found in the study [2]. In either case, the detection of outlier observation is important because if this observation is obtained due to a measurement error, it can be ignored, but if this observation is a real value, then its identification can be useful in the future studies. Blessley et al. [3] identify outlier observations with the approach of removing each observation in each step and examining the effect of removing observation on coefficients estimation of linear regression models. Other studies in this area were also carried out by Beckman and Cook [4], Brent and Louis [5] and Montgomery and Pack [2].

One way to identify outlier observations in regression models, is to measure the difference between the response variables and their expected values under fitted model. This identification in circular regression, is possible by using of a circular distance.

According to the importance of identifying outlier observations in the linear-circular regression model, the Difference of Means Circular Error statistic that was introduced by Abuzaid et al. [1] is applied for outlier detection in linear-circular regression model.

Material and methods

In this paper, the Difference of Means Circular Error statistics is applied for outlier detection in linear-circular regression model and the cut-off points of this statistic are obtained by Monte Carlo simulations. In addition, the performance of this statistic is investigated with some simulation studies. Finally, this statistic is applied to identify outlier observations in speed and direction wind data set recorded at Mehrabad weather station in Tehran with parametric Bootstrap simulation method.

Results and discussion

In this paper, we obtained the cut-off points of the DMCE statistics in a linear-circular regression model using the Monte Carlo simulation method. These points were reduced to $(\kappa)n$ with the assumption of constant $(n)\kappa$. .Also, in simulation studies, the power of this statistic was obtained for large values of λ (contamination level) which was near one for various values of n and large values of κ. DMCE statistic is applied to identify outlier observations in real data set. The performance of this statistic is desirable in detecting outlier observations in real data.

Conclusion

The following conclusions were drawn from this research:

- Cut-off points of the DMCE statistics were reduced to $(\kappa)n$ with the assumption of constant $(n)\kappa$.
- In simulation studies, the power of DMCE statistic was near one for large values of λ (contamination level) and large values of κ.
- The performance of DMCE statistic is desirable in detecting outlier observations in real data.

Keywords: Linear-Circular regression model, Outlier observation, Difference of means, Circular Error.

*Corresponding author: sazimi04@gmail.com