OUTLIER DETECTION IN CIRCULAR DATA AND CIRCULAR-CIRCULAR REGRESSION MODEL

ADZHAR RAMBLI

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2011

OUTLIER DETECTION IN CIRCULAR DATA AND CIRCULAR-CIRCULAR REGRESSION MODEL

ADZHAR RAMBLI

DISSERTATION SUBMITTED IN FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE

INSTITUTE OF MATHEMATICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2011

ABSTRAK

Secara keseluruhan, kajian ini mengenai lanjutan daripada kaedah-kaedah berangka untuk mengesan nilai tersisih bagi taburan sampel bulatan dan model regresi bulatan. Perkara yang pertama adalah melihat kepada taburan yang berlainan untuk sampel bulatan dan model regressi bulatan. Parameter-parameter yang dapat menggambarkan setiap taburan bulatan akan dibincang dan ilustrasi menggunakan data yang sebenar. Bagi model regressi bulatan, kita akan mempertimbangkan model regressi bulatan DM yang dikemukakan oleh Down dan Mardia (2002). Kemudian kami menggunakan kaedah Anggaran Kebolehjadian Maksimum (MLE) untuk menganggarkan parameter-parameter yang didapati dalam model tersebut.

Perkara yang kedua adalah mempertimbangkan kehadiran nilai tersisih dalam data bulatan yang dijana daripada taburan keluarga "α-stable wrapped", iaitu taburan "wrapped normal". Empat kaedah-kaedah berangka statistik *C*, *M*, *D* and *A* dipertimbangkan untuk mengesan kehadiran nilai tersisih. Titik potongan bagi kesemua statistik diperolehi berdasarkan kajian simulasi. Kajian simulasi menunjukan bahawa statistik *A* mempunyai prestasi lebih baik berbanding statistik-statistik yang lain. Kajian simulasi, didapati bahawa titik potongan amat bergantung kepada ukuran penumpuan sampel bulatan. Kaedah-kaedah ini telah diaplikasikan kepada data arah angin di Kuantan dan telah berjaya mengesan sampel yang terletak jauh daripada sampel-sampel yang lain sebagai nilai tersisih.

Ketiganya adalah masalah untuk mengesan kewujudan nilai tersisih dengan menggunakan model regresi bulatan DM berdasarkan kepada dua jenis statistik; *COVRATIO* statistik dan *DMCEs* statistik. Titik potongan dan prestasi bagi kedua-dua statistik ini dikaji secara simulasi. Aplikasi bagi kedu-dua statistik ini dilakukan dengan

menggunakan data arah angin laut dan irama biologi sirkadian. Sekali lagi, kedua-dua statistik ini berjaya untuk mengenal pasti sepasang sampel sebagai nilai tersisih.

ABSTRACT

This study focuses on studying several numerical methods used to detect outliers in circular data and circular-circular regression models. Firstly, we will look at different distributions for circular variable and different regression models involving bivariate circular data. The parameters that describe each circular distribution will be explained in detail and illustrated using real data sets. As for the circular regression, we will consider the DM circular regression model proposed by Down and Mardia (2002). We employ the maximum likelihood estimation method to estimate the parameters of the model.

Secondly, we consider the occurrence of outliers in circular data generated from a family of α-stable wrapped distribution, that is, the wrapped normal distribution. Four numerical methods; the *C, M, D* and *A* statistics are considered to detect the outliers. The cut-off points of the statistics are obtained via simulation. In our case, we show through simulation that the *A* statistic performs better in detecting outliers than the other statistics. The methods are applied on the Kuantan wind direction data set and are able to detect observations further away from the rest as outliers.

Thirdly, we consider the problem of detecting a single influential observation in the DM circular regression based on two different statistics; the *COVRATIO* statistic and *DMCEs* statistic. The cut-off points and the performance of both procedures are studied via simulation. The application of the procedures is illustrated by the ocean wind direction data and circadian biological rhythm data respectively. Again, the procedures are able to identify outlying observation as possible influential observation.

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my supervisors, Assoc. Prof. Dr. Ibrahim bin Mohamed and Assoc. Prof. Dr. Abdul Ghapor bin Hussin for their guidance, invaluable help and encouragement throughout the period of my research.

Secondly, huge thanks go to my mother, brothers and sisters for their support and encouragement at all stages of this study. With their blessing and prayer, I am very grateful to eventually complete this study.

Next, I would like to gratefully acknowledge and the support from Assoc. Prof. Dr. Mohd Omar, Head of Department of the Institute of Mathematical Sciences (ISM), University of Malaya (UM) and all staff members of ISM, UM especially Puan Budiyah Yeop and En. Abd Malek Osmani. I would also like to thank the University of Malaya for the financial support.

Special thanks go to my friends; Ali, Safwati, Syuhada, Mardziah, Norli, Siti, and Amir for their assistance and support during the course of study.

CONTENTS

CHAPTER 5

CHAPTER 6

CHAPTER 7

LIST OF TABLES

LIST OF FIGURES