بسم الله الرحمن الرحيم

قال تعالى :

ـهِ وَبِرَحْمَ((تِهِ فَبِذَلِكَ فَلْيَفْرَحُوا هُوَ خَيْرٌ

مِّمَّا يَجْمَعُونَ (58)))

سورة يونس الايه (58)

Dedicated to

The memory of my father.

My most dear mother.

My wife.

My brothers and sisters.

ACKNOWLEDGMENTS

I would like to express my gratitude to my supervisor Porf. Z. A. EL- Beshir for his guidance and advice through out my research.

Thanks are also due to my family for the support and encouragement I Received form it.

Mr. Fakhr eldin Ismail also deserve thanks for his valuable help in Writing and executing the necessary computer programmers.

Finally my thanks extend to all those friends and colleagues who Provided helps of various kinds.

المستخلص

يهد ف هذا البحث لدراسة مشكلة اكتشاف القيم الشاذة في البيانات ذات البعد الواحد عندما يكون توزيعها ملتويا لليمين .

إقترحت في الاطروحه طريقه لامعلميه جديده تتضمن قاعده محافظه نسبيا لا كتشاف القيم الشاذه.

تمت مقارنة آداء الطريقه المقترحه من خلال دراسة محاكاه مع أربعه من الطرق اللا معلميه شائعة الاستخدام. وقد أجريت المقارنه بإفتراض ثلاثة توزيعات للبيانات وهي توزيعات مربع كاى وبيتا والتوزيع الطبيعي اللوغاريتم .

وقد استخدمت أحجام عينات صغيره ومتوسطه وكبيره في التجربه.

وقد أكدت دراسة المحاكاه تفوق الطريقه المقترحه على الطرق ألاخرى عند محاولة اكتشاف القيم الشاذه في الطرف الايمن من التوزيع .

ويتزايد هذا التفوق مع زيادة حجم العينه , أما أداؤها في الطرف ألايسر فهو ضعيف نسبيا.

Abstract

This research aimed at investigating the problem of outlier detection in univariate data when the distribution is skewed to the right .

A new nonparametric method is proposed that involves a more conservative rule of detection. The performance of the suggested method is compared through a simulation experiment, to that of four widely used nonparametric methods.

The comparison is made under three distributions , namely the chi-square , beta and log normal distribution. Small, medium and large sample size are used in the experiment.

The result of the simulation, confcrmed the superiority of the Proposed method to the other methods when testing for outliers in the right tail of the distribution .

This superiority increase with increase in sample size. The Performance of the method in the left tail is relatively poor.

Contents

List of Tables

List of Symbols

MD^2 =Mahalanobis distance

CD^2 = Cook's squared distance

$DFFITS_{(1)}$= The number of standard deviations from the i^{th} observation

$DFBETAS$= A diagnostic measure based on the change in the j^{th} regression coefficient of $\beta_{j(\iota)}$.

COVRATIO's= Determinant of a covariance matrix of the estimates after deleting the i^{th} observation

AIC = Applied Akaike Criterion

h_{ii} = is the i^{th} diagonal element of $X^l(X^l X)^{-1}X$

$\bar{x}$= Sample mean

SD= Standard deviation

G=**Grubs test**

r=Dixon test

R= **Extreme Studentized Deviate**

MAD= Median Absolute Deviation

MAD_e=1.483×MAD

GLDs= Generalized Lambda Distributions

Z_i= **Z-SCORE**

Q_1 = Lower quartile

Q_2= Median

Q_3= Upper quartile

IQR= Inter quartile range

MC= Medcouple

L= Lower critical value

U= Upper critical value

MCD= Minimum covariance determinant

MVE= Minimum volume ellipsoid

M=M-estimator

$|X_{LL}|$=Largest absolute value

$|X_{SL}|$= Smallest absolute value

b= Deference between $|X_{LL}|$ and $|X_{SL}|$