جامعة الموصل

كلية علوم الحاسوب والرياضيات

تطوير مقدرات جديدة لمعلمة انحدار الحرف في نموذج الانحدار الخطي المتعدد

محمد قاسم يحيى الاوجار

اطروحة دكتوراه

إحصاء

بإشراف

الاستاذ الدكتور

موفق محمد توفيق القصاب

كلية التربية – جامعة تيشك العالمية

اربيل – العراق

1441 هـ

2020 م

المستخلص

إنّ تعدد العلاقة الخطية (multicollinearity) بين المتغيرات التوضيحية في الانحدار الخطي المتعدد تعتبر من المشاكل المهمة في مجال تحليل الانحدار ، تناولت هذه الدراسة هذه المشكلة واسبابها وآثارها وانواعها وطرق الكشف عنها وكذلك التعرف على اساليب معالجتها، حيث تم اختيار احد انواع المعالجة وهو طريقة انحدار الحرف (Ridge Regression) ويوجد الكثير من الاساليب لهذه الطريقة استخدمت من قبل العديد من الباحثين غير انها اقتصرت على ايجاد معلمة الحرف k كعدد ثابت ، تم اقتراح طريقة جديدة لإيجاد معلمة الحرف k ، عندما تكون معلمة الحرف k مصفوفة وكحالة خاصة منها هو المتجه ، تختلف الطريقة المقترحة بانها تستخدم الجذور المميزة λ_i لمصفوفة الارتباطات $X'X$ بعد ان كانت الطرق السابقة تستخدم معلمات الانحدار $\underline{\beta}$ لطريقة المربعات الصغرى في تقدير معلمة انحدار الحرف k . تناول الجانب التطبيقي دراسة تطبيقية على بيانات حقيقية تم اجراء دراسات عليها بطرق سابقة وتم تطبيق الطريقة المقترحة الجديدة عليها واجراء المقارنة وظهر من النتائج كفاءة الطريقة المقترحة على الطرق السابقة.

و كذلك استخدمت طريقة مونت كارلو للمحاكاة لتوليد اربعة عينات مختلفة تحوي بيانات عشوائية لمتغيرات توضيحية تتوزع توزيع طبيعي وهذه العينات تحوي بين متغيراتها التوضيحية تعدد علاقة خطية وبثلاثة درجات ارتباط مختلفة وطبقت عليها طريقة المربعات الصغرى وبعض طرق انحدار الحرف السابقة والطريقة المقترحة لانحدار الحرف . تم إجراء مقارنة بين جميع هذه الطرق فأظهرت النتائج مرة اخرى كفاءة الطريقة المقترحة وذلك بأن هذه الطريقة تملك أقل متوسط مربعات خطأ . كذلك اظهرت النتائج ان معلمة الحرف K كمتجه (vector) التي تم الحصول عليها تمتلك اقل متوسط مربعات خطأ ويدل ذلك الحصول على افضل نموذج لمعلمات الانحدار واقل تباينات لها .

University of Mosul

College of Computer Sciences and Mathematics

Developing New Estimators for the Ridge Regression Parameter in Multiple Linear Regression Model

Mohammed Qasim Yahya Al-Awjar

Ph.D. Thesis

Statistics

Supervised By

Prof.

Dr. Mowafaq M. Al-Kassab

Department of Mathematics Education

Faculty of Education

Tishk International University

Erbil , Iraq

1441 A.H. **2020 A.D.**

Abstract

The multiplicity of linear relationship (multicollinearity) between the explanatory variables in multiple linear regression is considered as an important problem in the field of regression analysis, this study examined this problem, its causes, effects, types and methods of detection, as well as identifying the methods of its treatment, where one of the types of treatment was chosen, which is the method of (Ridge Regression). Since there are many methods for this method used by many researchers that are limited to finding the parameter of the ridge k as a fixed number, a new method has been proposed to find the parameter k:

when k is fixed , vector and matrix, the proposed method varies , it uses the Eigen Values of the correlation matrix X 'X whereas the previous methods used the regression parameters $\underline{\beta}$ of the least squares method to estimate the k-regression parameter. An applied study on real data were conducted and the new proposed method was applied to it, and the comparison was conducted , the results showed the efficiency of the proposed method compared with other methods previously used by other researchers.

Also, Monte Carlo simulation method was used to generate four different samples containing random data for ten explanatory variables having standard normal distribution and these samples contain , between the explanatory variables three different degrees of correlation . These data applied to the least squares method and some of the previous ridge regression methods in addition to the proposed method . A comparison was made between all these methods, and the results showed once again the efficiency of the proposed method, when the parameter k is in the case of the vector , and in all samples where it appears that it has the lowest mean squared errors, and thus we can obtain the best model for the regression parameters and the least variations thereof.