وزارة التعليم العالي والبحث العلمي
جامعة الموصل
كلية علوم الحاسوب والرياضيات
قسم الإحصاء والمعلوماتية

نماذج الانحدار لبيانات العد في ظل وجود مشكلة زيادة أو نقصان التشتت

رسالة مقدمة

الى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل

كجزء من متطلبات نيل شهادة ماجستير علوم في الإحصاء

من قبل

محمد خالد محمد نوري الجادر

بإشراف

الأستاذ الدكتور

زكريا يحيى نوري

١٤٤٤هـ ٢٠٢٢م

المستخلص

يلقى موضوع تحليل الانحدار اهتماماً واسعاً وواضحاً في معظم الدراسات. ويعد أنموذج انحدار بواسون أحد أهم نماذج الانحدار اللوغاريتمية الخطية، وهو الأداة التي تتم من خلالها نمذجة متغير الاستجابة عندما تكون قيم ذلك المتغير على شكل قيم قابلة للعد. إن كفاءة استخدام هذا الأنموذج تعتمد على توفر فرض تساوي التوقع الرياضي والتباين. وكغيره من سائر نماذج الانحدار، فان هذا الشرط قد لا يتوفر مما يؤثر سلباً على دقة الأنموذج وبساطته في تفسير النتائج.

تهدف هذه الدراسة إلى استعراض ومقارنة نماذج انحدار بيانات العد في حالتي زيادة التشتت ونقصان التشتت عبر طريقة الإمكان باستخدام المحاكاة والبيانات الحقيقة. وقد تم استخدام أسلوب مونت – كارلو في المحاكاة لتوليد بيانات تتبع أنموذج انحدار بواسون تبعاً لعوامل مختلفة كحجم العينة، وعدد المتغيرات، وقيمة معلمة التشتت. كما تم الاعتماد على عدة معايير للمقارنة هي AIC,BIC,Deviance. وقد أظهرت نتائج المحاكاة تفوق أنموذج انحدار ثنائي الحدين السالب مقارنةً بنماذج انحدار أخرى عندما يكون هناك زيادة للتشتت، في حين أظهرت نتائج المحاكاة تفوق انموذج انحدار بواسون المعمم مقارنةَ بالنماذج المستخدمة الاخرى عندما يكون هناك نقصان التشتت. فضلا عن ذلك، طبقت النماذج المستخدمة على بيانات حقيقية تم جمعها من مصابين بمرض العجز الكلوي المزمن والذين يتعالجون بالغسيل الكلوي المستمر، وقد شّخصت حالة المرضى من أطباء مختصين بالتعاون مع مستشفى ابن سينا التعليمي – وحدة الكلية الاصطناعية. وقد بينت النتائج تفوق أنموذج انحدار ثنائي الحدين السالب على باقي النماذج الأخرى بسبب وجود مشكلة زيادة التشتت.

Ministry of Higher Education and
Scientific Research
University of Mosul
College of Computer Science and
Mathematics
Department of Statistics and Informatics

Regression Models for count data with overdispersion or underdispersion Issue

A Thesis Submitted to the Council of the College of
Computer Science and Mathematics
University of Mosul
As a Partial Fulfillment of Requirements
for the Degree of Master of Science
in
Statistics

By

Mohammed Khalid MohammedNouri Aljader

Supervised by

Professor
Dr. Zakariya Yahya Nouri

2022 A.D. 1444 A.H.

Abstract

Regression analysis is great of interest in several studies. The Poisson regression model is one of the most important models of linear logarithmic regression, the tool by which the dependent variable is modeled when the values of that variable are in the form of count values. The efficiency of the Poisson regression model is depending on the equalization of the mean and variance of the data. However, the model may contain the overdispersion or underdispersion problem, which negatively affects the accuracy of the model and its simplicity in interpreting the results. The aim of this study is to review and compare the methods of modeling the count data under the overdispersion or underdispersion problem using simulations and real data. The Monte-Carlo method is used in simulations to generate data following the Poisson regression model according to several factors, such as sample size and the dispersion factor. In terms of prediction accuracy, simulation results demonstrated that the negative binomial regression model has better results comparing with other regression model when the dispersion factor exist. Furthermore, generalized Poisson regression model yields better performance compared to other regression models when the underdispersion is exist. In addition, the result of the real data application, which is related to the patients with chronic renal insufficiency, reveals that the negative binomial regression model method is quite efficient and feasible for estimating the count data.