وزارة التعليم العالي والبحث العلمي

جامعة الموصل

كلية علوم الحاسوب والرياضيات

قسم الإحصاء والمعلوماتية

بعض طرائق التقليص لمقدرات الانحدار اللوجستي
(دراسة مقارنة)

أوس عطاالله مطلوب

رسالة ماجستير

الإحصاء والمعلوماتية

بإشراف الأستاذ المساعد

د. زكريا يحيى الجمّال

١٤٣٨هـ ٢٠١٧م

المستخلص

يعد نموذج الانحدار اللوجستي أحد أهم نماذج الانحدار اللوغاريتمية الخطية، وهو الأداة التي يتم من خلالها نمذجة المتغير المعتمد عندما تكون قيم ذلك التابع على شكل قيم ثنائية وصفية. قد يتعرض هذا النموذج الى مشاكل، من اهم هذه المشاكل هو مشكلة التعدد الخطي التي تحدث نتيجة ارتباط عالي وتلازم خطي بين متغيرين أو أكثر من المتغيرات التوضيحية (المستقلة) الداخلة في بناءه مما يؤثر بشكل سلبي على عملية تقدير معلمات النموذج. تهدف هذه الدراسة إلى استعراض ومقارنة طرائق تقدير معلمات نموذج الانحدار اللوجستي عندما يعاني من مشكلة التعدد الخطي شبه التام عبر طرائق عدة بالإضافة الى اقتراح طريقة جديدة للتقدير. تم استخدام أسلوب مونت – كارلو في المحاكاة لتوليد بيانات تتبع نموذج الانحدار اللوجستي وتعاني بذات الوقت من مشكلة التعدد الخطي شبه التام تبعا لعوامل مختلفة كقيمة معامل الارتباط البسيط وحجم العينة وعدد المتغيرات التوضيحية، فقد أظهرت نتائج المحاكاة بالاعتماد على متوسط مربعات الخطأ ودقة التصنيف بوضعهما معياراً للمقارنة تفوق الطريقة المقترحة في تقدير معلمات نموذج الانحدار اللوجستي عندما يعاني من مشكلة التعدد الخطي مقارنةَ بطرائق التقدير الأخرى، فضلاً عن عدم تأثرها بالتغير في كافة عوامل المحاكاة المختلفة. اضافة الى ذلك، تم تطبيق الطريقة المقترحة على بيانات حقيقية تم جمعها من قبل (عبد و رشيد، البحث قيد النشر) لمرضى مصابين بالجلطة القلبية في مستشفى مدينة الطب وجاءت النتائج التي تم الحصول عليها مطابقة الى نتائج تجارب المحاكاة بتفوق الطريقة المقترحة ايضا.

The Ministry of Higher Education

and Scientific Research

University of Mosul

College of Computer Sciences and Mathematics

Some Shrinkage Methods for Logistic Regression Estimators (Comparative Study)

Aws Ataallah Matloob

M.Sc./Thesis

Statistics

Supervised by

Assistant Prof.

Dr. Zakariya Yahya Algamal

2017 A.D. **1438 A.H.**

Abstract

Logistic regression is considered as a powerful discriminative method and as an important model among linear logarithmic regression models. Logistic regression model is used to modeling the response variable when it takes binary qualitative values. In some of the practical situations of logistic regression, explanatory variables are correlated which leads to the problem of multicollinearity. It is known that multicollinearity affects the variance of maximum likelihood estimator negatively. Consequently unstable estimates can be obtained. In this thesis, we explored and compared several methods that can estimate the logistic regression parameters when the multicollinearity is presence. Furthermore, a new estimate method is proposed. A Monte Carlo study is used to show the effectiveness of the proposed methods over other used methods. In addition, a real data application which is related to the heart attack is illustrated to show the benefits of our proposed method. The simulation and real data application results reveal that the proposed method is quite efficient and feasible for estimating the logistic regression model when multicollinearity is exist in terms of mean squared error and classification accuracy.