

ชื่อวิทยานิพนธ์	กระบวนการวิเคราะห์ข้อมูลทวินามแบบลบที่มีความแปรปรวนเป็นฟังก์ชันพหุนามกำลังสองของค่าเฉลี่ย
ผู้เขียน	นางสาวจิรภัทร ภูขำวิญทอง
สาขาวิชา	คณิตศาสตร์และสถิติ
ปีการศึกษา	2549

บทคัดย่อ

ตัวแบบถดถอยปัวซอง หรือตัวแบบลอกเชิงเส้น เป็นตัวแบบพื้นฐานสำหรับข้อมูลแบบนับที่มีข้อสมมติเบื้องต้นที่จำเป็นต้องตรวจสอบ คือการเท่ากันของความแปรปรวนและค่าเฉลี่ย ในทางปฏิบัติมักพบว่าข้อมูลแบบนับจะมีความแปรปรวนมากกว่าค่าเฉลี่ย หรือเกิดปรากฏการณ์ที่เรียกว่า Overdispersion ดังนั้นตัวแบบถดถอยปัวซองจึงไม่เหมาะสมกับข้อมูล ตัวแบบที่มักใช้แทนตัวแบบถดถอยปัวซองคือ ตัวแบบทวินามแบบลบ (Negative Binomial (NB) model) ตัวแบบ NB มีหลายรูปแบบจำแนกตามความสัมพันธ์ระหว่างความแปรปรวนและค่าเฉลี่ย แต่ที่นิยมใช้มากที่สุดคือตัวแบบ Linear mean – variance และ Quadratic mean – variance NB (NB1 และ NB2 ตามลำดับ)

วิทยานิพนธ์ฉบับนี้กล่าวถึงการวิเคราะห์ Overdispersed Poisson Counts โดยใช้ตัวแบบ NB2 ซึ่งประกอบด้วย การหาค่าประมาณความควรจะเป็นสูงสุดของสัมประสิทธิ์การถดถอยสำหรับค่าเฉลี่ยของข้อมูลแบบนับ และของ Overdispersion parameter ตรวจสอบ Overdispersion ในตัวแบบถดถอยปัวซองโดยใช้ Likelihood ratios test, Wald test และ Score Test เลือกตัวแบบ NB2 ที่ดีที่สุดโดยใช้ค่า Akaike information criteria (AIC) และวินิจฉัยตัวแบบที่ได้โดย Half normal plot with simulate envelope เราดำเนินการกระบวนการเหล่านี้โดยเขียนชุดคำสั่งสำหรับใช้ในโปรแกรม R และตรวจสอบความถูกต้องกับผลลัพธ์ที่ปรากฏในเอกสารที่ตรวจสอบ

นอกจากศึกษากระบวนการต่าง ๆ ที่กล่าวมาข้างต้นแล้ว ผู้วิจัยได้พัฒนา Robust standard error ของตัวประมาณค่าของ NB2 พร้อมทั้งทำการจำลองข้อมูล (Simulation study) เพื่อเปรียบเทียบระหว่าง Robust standard error กับ Asymptotic standard error ของตัวประมาณค่าที่ได้จากกระบวนการประมาณความควรจะเป็นสูงสุด จากการศึกษาได้ว่า ถ้าข้อมูลแบบนับนั้นเป็นข้อมูลจากการแจกแจง NB2 แล้ว Robust standard error กับ Asymptotic standard error จะมีค่าใกล้เคียงกัน ดังนั้นจึงสามารถสรุปได้ว่า Robust standard error ของตัวแบบ NB2 เป็นสถิติตัวหนึ่งที่

สามารถใช้ตรวจสอบว่าตัวแบบ NB2 เป็นตัวแบบที่ถูกต้องสำหรับ Overdispersed Poisson counts นั้น หรือไม่

จากการประยุกต์กระบวนการทั้งหลายที่กล่าวมาข้างต้นกับข้อมูล NB2 จำนวน 2 ชุด ได้แก่ ข้อมูลจำนวนรอยตำหนิบนผ้าที่มีขนาดต่าง ๆ กันที่ผลิตจากโรงงานแห่งหนึ่ง และข้อมูลจำนวนวันขาดเรียนของนักเรียนที่เป็นชนพื้นเมืองและชนผิวขาวของโรงเรียนแห่งหนึ่งในประเทศออสเตรเลีย พบว่า Robust standard error และ Asymptotic standard error มีค่าใกล้เคียงกัน นั่นคือ ตัวแบบ NB2 เหมาะสมกับข้อมูลทั้ง 2 ชุด

Thesis Title	Performance of Quadratic Mean – Variance Negative Binomial Models
Author	Miss Jirapat Phookwantong
Major Program	Mathematics and Statistics
Academic Year	2006

ABSTRACT

Poisson Regression or log-linear models are basically applied for modeling count data. There is a strong assumption for Poisson Regression to be checked, in particular the equality of the mean and the variance. In practice, count data usually encounter overdispersion: the variance exceeds the mean. These overdispersed Poisson counts indicate that the Poisson regression is no longer suitable. Then the Negative Binomial (NB) model is the parametric model for choice. There are various forms of NB distributions defined by the relationship between the mean and variance. The most commonly used are linear mean-variance NB (NB1) and quadratic mean-variance NB (NB2).

This thesis describes the procedure of NB2 modelling including maximum likelihood estimation for regression coefficients and the overdispersion parameter; test statistics such as the Likelihood ratio test, the Wald test, and score tests for checking overdispersion in Poisson regression; model selection using Akaike information criteria (AIC); checking adequacy of the model using half normal plots with a simulated envelope.

Apart from studying the NB2 modelling, we developed robust standard errors for the estimated parameters, and conducted a small simulation study to compare the robust standard errors and the asymptotic ones. Based upon our analysis, we found that the robust standard errors for NB2 model can be used to check whether the NB2 is the right model for studied overdispersed counts.

For illustration, we applied the presented procedure to two sets of NB2 data available in the statistical modeling literature, namely one about analyzing the faults in rolls of fabric in a textile factory and the other about the absence from school of aboriginal and white children in Australia. It's shown that NB2 model is appropriate with both data.