การเปรียบเทียบตัวแบบการถดถอยควอไซปัวซงและตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มากสำหรับข้อมูลนับที่ค่าความแปรปรวนมากกว่าค่าเฉลี่ย

A Comparison of Quasi-Poisson and Zero Inflated Negative Binomial Regression Models for Over-dispersion Count Data

นวพรรณ เชื้ออ่ำ*, บุญอ้อม โฉมที และ อภิญญา หิรัญวงษ์

Navapun Chuea-am*, Boonorm Chomtee and Apinya Hirunwong

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

Department Statistics, Faculty of Science, Kasetsart University

Received : 16 March 2018

Accepted : 7 June 2018

Published online : 12 June 2018

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบความเหมาะสมของตัวแบบการถดถอย 2 ตัวแบบ คือ ตัวแบบการถดถอยควอไซปัวซง (QP) และตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มาก (ZINB) สำหรับตัวแปรตามที่มีลักษณะเป็นข้อมูลนับและมีค่าความแปรปรวนมากกว่าค่าเฉลี่ย โดยการศึกษากับข้อมูลจริงที่ตัวแปรตาม คือ จำนวนผู้บาดเจ็บในการเกิดอุบัติเหตุแต่ละครั้ง ซึ่งแบ่งตามลักษณะของข้อมูลเป็น 3 กรณี คือ ตัวอย่างขนาดเล็ก (n =17) กลาง (n =32) และใหญ่ (n =56) โดยตัวอย่างขนาดเล็กมีค่าความน่าจะเป็นที่จะเกิดศูนย์ของตัวแปรตามเท่ากับ 0.25 และสำหรับตัวอย่างขนาดกลางและใหญ่มีค่าความน่าจะเป็นที่จะเกิดศูนย์ของตัวแปรตามเท่ากับ 0.50 มีตัวแปรอิสระ 3 ตัวแปร คือ ลักษณะบริเวณที่เกิดอุบัติเหตุ สาเหตุของการเกิดอุบัติเหตุและช่วงเวลาในการเกิดอุบัติเหตุ และข้อมูลจำลองที่ตัวแปรตามมีการแจกแจงทวินามเชิงลบที่มีศูนย์มาก โดยกำหนดค่าพารามิเตอร์การกระจาย (k) เท่ากับ 1.25 1.50 และ 1.75 ค่าความน่าจะเป็นที่จะเกิดศูนย์ (π) เท่ากับ 0.25 และ 0.50 ค่าเฉลี่ย (λ) เท่ากับ 1.4 2 และ 3 และกำหนดให้มีตัวแปรอิสระจำนวน 3 ตัวแปรโดยตัวแปรอิสระมีการแจกแจงแบร์นูลลีที่ความน่าเป็นของการเกิดความสำเร็จ (p)เท่ากับ 0.3 0.5 และ 0.8 ข้อมูลที่ศึกษา คือ ขนาดเล็ก (n = 15, 20) กลาง (n =30, 35) และใหญ่ (n =45, 50) เกณฑ์พิจารณาตัวแบบที่เหมาะสม คือ รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (AAE) โดยตัวแบบที่ให้ค่า RMSE และค่า AAE ต่ำกว่าแสดงว่าตัวแบบนั้นมีความเหมาะสมมากกว่า ผลการศึกษาพบว่า สำหรับข้อมูลจริงและข้อมูลจำลองที่ศึกษา เมื่อพิจารณาจากค่า RMSE และค่า AAE ตัวแบบการถดถอยควอไซปัวซงมีความเหมาะสมมากกว่าตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มากเกือบทุกกรณีที่ศึกษา

คำสำคัญ : ข้อมูลนับ, ค่าความแปรปรวนมากกว่าค่าเฉลี่ย, ตัวแบบการถดถอยควอไซปัวซง, ตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มาก

*Corresponding author. E-mail : navapunchueaam@gmail.com

Abstract

In this research study aimed to compare the appropriation of regression models between Quasi-Poisson (QP) and Zero inflated negative binomial (ZINB) which dependent variable was count data and the variance was greater than the mean. The dependent variable for the real data was the number of injured in each accident which there are three cases: small (n =17), medium (n =32) and large (n =56). The probabilities of zero event (π) were 0.25 for the small sample size and 0.50 for the medium and large sample sizes. In the real data set, there are three independent variables. For the simulation data, the dependent variable had Zero inflated negative binomial distribution. The dispersion parameter of the distribution (k) were 1.25, 1.50 and 1.75, the probability of zero events (π) were 0.25 and 0.50 and the mean (λ) were 1.4, 2 and 3. For the simulation data set, Three independent variables were determined with bernoulli distribution and the probability of success events (p) were 0.3, 0.5 and 0.8. The sample size of the simulation data (n) were small (n =15, 20), medium (n =30, 35) and large (n =45, 50). The criteria of model appropriation were root mean square error (RMSE) and absolute average error (AAE). The smaller values of RMSE or AAE indicate the better model. For the results based on RMSE and AAE, it is found that the Quasi-Poisson regression model was more appropriate than Zero inflated negative binomial regression model at almost case of study for both the real and simulation data.

Keywords : count data, over-dispersion, Quasi-Poisson Regression model, Zero inflated negative binomial regression model

บทนำ

ในปัจจุบันงานวิจัยทางด้านการแพทย์ วิทยาศาสตร์ สังคมศาสตร์ เศรษฐศาสตร์ ธุรกิจหรือวิศวกรรมศาสตร์ ได้นำตัวแบบการถดถอยมาใช้ในการตัดสินใจเพื่อหาความสัมพันธ์ระหว่างปัจจัยที่ต้องการศึกษาหรือเรียกว่า ตัวแปรตาม (Dependent variable) กับปัจจัยที่เกี่ยวข้องหรือเรียกว่า ตัวแปรอิสระ (Independent variable) ซึ่งในกรณีที่ตัวแปรตามเป็นตัวแปรสุ่มชนิดต่อเนื่องและตัวแปรอิสระเป็นตัวแปรเชิงปริมาณหรือเป็นตัวแปรเชิงกลุ่ม จะเรียกวิธีการวิเคราะห์นี้ว่า การวิเคราะห์การถดถอย (Regression analysis) แต่ในกรณีที่ตัวแปรตามเป็นตัวแปรสุ่มชนิดไม่ต่อเนื่องหรือเป็นจำนวนนับ (Count data) หากใช้การวิเคราะห์การถดถอยนั้นผลลัพธ์ที่ได้จะเกิดความเอนเอียง (Bias) เนื่องจากการวิเคราะห์การถดถอยเชิงเส้นมีค่าคาดหมายของตัวแปรตามเป็นได้ทั้งบวก ลบ และศูนย์ ในขณะที่ค่าคาดหมายของตัวแปรตามเป็นข้อมูลนับจะมีค่าไม่เป็นลบ (Non negative number)

ตัวแบบที่ใช้ในการวิเคราะห์ข้อมูลที่ตัวแปรตามเป็นจำนวนนับนั้นพบได้ในหลายสาขาวิชา และตัวแบบที่นิยมนำมาใช้คือ ตัวแบบการถดถอยปัวซง (Poisson regression model: P) โดยตัวแปรตาม (Y) มีการแจกแจงปัวซงซึ่งมีข้อสมมุติว่าค่าเฉลี่ยและค่าความแปรปรวนต้องมีค่าเท่ากัน นั่นคือ $E(Y)$ เท่ากับ $Var(Y)$ แต่ในความเป็นจริงมักจะพบว่าค่าความแปรปรวนอาจจะมีค่ามากกว่าหรือน้อยกว่าค่าเฉลี่ย โดยที่ถ้าค่าความแปรปรวนมากกว่าค่าเฉลี่ยจะเรียกว่า Over-dispersion หรือถ้าค่าความแปรปรวนน้อยกว่าค่าเฉลี่ยจะเรียกว่า Under-dispersion ซึ่งจากงานวิจัยที่ผ่านมาพบว่ากรณีที่ตัวแปรตาม

เป็นจำนวนนับมักจะเกิดปัญหาค่าความแปรปรวนมากกว่าค่าเฉลี่ย (Pornapraditpun, 2009) ซึ่งตัวแบบที่สามารถจัดการกับปัญหานี้ได้ คือ ตัวแบบการถดถอยควอไซปัวซง (Quasi-Poisson regression model: QP) (Potts and Elih, 2006; Sileshi, 2006; Ver Hoef and Boveng, 2007) เนื่องจากตัวแบบดังกล่าวมีพารามิเตอร์การกระจาย (Dispersion parameter) ที่สามารถจัดการกับปัญหาค่าความแปรปรวนมากกว่าค่าเฉลี่ยได้

นอกจากปัญหาค่าความแปรปรวนมากกว่าค่าเฉลี่ยแล้วในข้อมูลชุดเดียวกันนั้นบ่อยครั้งที่พบว่าตัวแปรตามมีค่าเป็นศูนย์จำนวนมาก (Charatiam, 2010) ซึ่งในการวิเคราะห์ข้อมูลอาจก่อให้เกิดปัญหาตัวแปรตามมีค่าเป็นศูนย์มาก (Zero-inflated) และตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มาก (Zero inflated negative binomial regression model: ZINB) (Yau *et al*., 2003; Nitchanpunsri, 2011; Hu *et al*., 2011; Batra *et al*., 2016; Yusuf *et al*., 2017) เป็นตัวแบบหนึ่งที่สามารถจัดการกับปัญหาค่าความแปรปรวนมากกว่าค่าเฉลี่ยและปัญหาที่ตัวแปรตามมีค่าเป็นศูนย์จำนวนมากได้

ดังนั้นผู้วิจัยจึงสนใจที่จะทำการเปรียบเทียบตัวแบบการถดถอย 2 ตัวแบบที่มีตัวแปรตามเป็นจำนวนนับและตัวแปรอิสระหลายตัว โดยตัวแบบที่ศึกษาและเปรียบเทียบ คือ ตัวแบบการถดถอยควอไซปัวซงและตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มาก เกณฑ์การพิจารณาตัวแบบที่เหมาะสม คือ รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root mean square error: RMSE) และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Absolute Average error: AAE) โดยตัวแบบที่มีค่า RMSE และ ค่า AAE ต่ำกว่าแสดงว่า ตัวแบบนั้นมีความเหมาะสมมากกว่า

วิธีดำเนินการวิจัย

การศึกษานี้จะศึกษาเปรียบเทียบตัวแบบการถดถอยควอไซปัวซงและตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มาก โดยใช้ข้อมูลจริงและข้อมูลจำลอง เพื่อหาความสัมพันธ์ระหว่างตัวแปรตามที่เป็นข้อมูลนับกับตัวแปรอิสระหลายตัวที่เป็นตัวแปรเชิงกลุ่ม โดยจะพิจารณาจากรากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (AAE) รายละเอียดของข้อมูลจริงและข้อมูลจำลองมีดังนี้

1. ข้อมูลจริงที่ใช้ในการศึกษา คือ ข้อมูลการเกิดอุบัติเหตุบนท้องถนนที่เป็นถนนทางหลวง ซึ่งมีรถสัญจรมาก และมีการเกิดอุบัติเหตุที่มีผู้บาดเจ็บเป็นจำนวนมาก โดยเก็บรวบรวมข้อมูลจากสำนักอำนวยความปลอดภัย กรมทางหลวง ตั้งแต่เดือนมกราคมถึงเดือนธันวาคม ปี พ.ศ. 2559 มีถนนทางหลวง 3 เส้น ดังนี้

   1. ถนนทางหลวงหมายเลข 0403 (สี่แยกกบินทร์บุรี – วังน้ำเขียว) จังหวัดปราจีนบุรี กิโลเมตรที่ 0208+660 ถึง 0219+000 เกิดอุบัติเหตุ 17 ครั้ง
   2. ถนนทางหลวงหมายเลข 0403 (สี่แยกกบินทร์บุรี – วังน้ำเขียว) จังหวัดปราจีนบุรี กิโลเมตรที่ 0166+571 ถึง 0208+660 เกิดอุบัติเหตุ 32 ครั้ง
   3. ถนนทางหลวงหมายเลข 0601 และถนนทางหลวงหมายเลข 0602 (ป่าเตียน-บ้านสิ้ว) จังหวัดจันทบุรี กิโลเมตรที่ 0287+900 ถึง 0324+300 เกิดอุบัติเหตุ 56 ครั้ง

เมื่อจำนวนการเกิดอุบัติเหตุ คือ ขนาดตัวอย่าง (n)

ข้อมูลจริงที่นำมาศึกษามีตัวแปรตามเป็นข้อมูลนับ 1 ตัวแปรและมีการแจกแจงทวินามเชิงลบที่มีศูนย์มากมีค่าความน่าจะเป็นที่จะเกิดศูนย์ของตัวอย่างขนาดเล็กเท่ากับ 0.25 และตัวอย่างขนาดกลางและใหญ่เท่ากับ 0.50 มีตัวแปรอิสระ 3

ตัวแปรที่มีการแจกแจงแบร์นูลลี โดยที่ตัวแปรอิสระแต่ละตัวเป็นตัวแปรเชิงกลุ่มที่มี 2 ค่า คือ การสนใจเหตุการณ์ที่เกิดขึ้นมีค่าเท่ากับ 1 และการไม่สนใจเหตุการณ์ที่เกิดขึ้นมีค่าเท่ากับ 0 โดยตัวแปรอิสระที่ 3 ตัวนี้ เป็นปัจจัยที่ทำให้เกิดอุบัติเหตุ (Pornapraditpun, 2009) ดังนี้

y คือ จำนวนผู้บาดเจ็บในการเกิดอุบัติเหตุแต่ละครั้ง

x_1 คือ ลักษณะบริเวณที่เกิดเหตุ (ทางตรง= 0, ทางโค้งหรือทางแยก = 1)

x_2 คือ สาเหตุของการเกิดอุบัติเหตุ (ตัดหน้ากระชั้นชิด มีสิ่งกีดขวาง= 0, ขับรถเร็ว เมาสุรา ฝ่าฝืนไฟจราจร หลับใน = 1)

x_3 คือ ช่วงเวลาในการเกิดอุบัติเหตุ (กลางวัน 06.00 – 18.00น. = 0, กลางคืน 18.01 – 05.59 น. = 1)

2. ข้อมูลจำลองที่ใช้ในการศึกษา จำลองด้วยโปรแกรม SAS เวอร์ชั่น 9.4 มีรายละเอียด ดังนี้

2.1 จำลองข้อมูลของตัวแปรอิสระ (X) 3 ตัวแปรที่มีการแจกแจงแบร์นูลลีที่ความน่าจะเป็นของการเกิดความสำเร็จเท่ากับ 0.3, 0.5 และ 0.8 โดยจะจำลองข้อมูลให้สอดคล้องกับข้อมูลจริงที่นำมาศึกษา (Pornapraditpun, 2009) ตามขนาดตัวอย่างที่กำหนด 6 ระดับแบ่งออกเป็น 3 ขนาด คือ เล็ก กลางและใหญ่ ดังนี้ ขนาดเล็ก (15, 20) ขนาดกลาง (30, 35) และขนาดใหญ่ (45, 50)

2.2 สร้างข้อมูลของตัวแปรตาม (Y) จากการแจกแจงทวินามเชิงลบที่มีศูนย์มากและกำหนดให้มีค่าความแปรปรวนมากกว่าค่าเฉลี่ย (Doyle, 2009; Charatiam, 2010) โดยใช้ข้อมูลของตัวแปรอิสระในขั้นตอนที่ 1 และกำหนดค่าพารามิเตอร์ค่าเฉลี่ย (λ) ค่าพารามิเตอร์การกระจาย (k) และค่าพารามิเตอร์ความน่าจะเป็นที่จะเกิดศูนย์ (π) ดังนี้

- ค่าพารามิเตอร์ค่าเฉลี่ย λ = 1.4, 2, 3
- ค่าพารามิเตอร์การกระจาย k = 1.25, 1.50, 1.75
- ค่าพารามิเตอร์ความน่าจะเป็นที่จะเกิดศูนย์ π = 0.25, 0.50

3. นำข้อมูลที่จำลองได้มาสร้างตัวแบบการถดถอยควอไซป์ซองและตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มาก

4. คำนวณค่า RMSE และ AAE ของตัวแบบ

5. สำหรับข้อมูลจำลองทำการทดลองซ้ำ 100 ครั้งสำหรับแต่ละสถานการณ์

6. คำนวณค่าเฉลี่ยของ RMSE และ ค่า AAE แล้วเปรียบเทียบตัวแบบที่ศึกษา โดยที่ตัวแบบที่มีค่า RMSE และ ค่า AAE ต่ำกว่าแสดงว่าตัวแบบนั้นมีความเหมาะสมมากกว่า

ตัวแบบการถดถอยและเกณฑ์ที่ใช้ในการศึกษา มีรายละเอียด ดังนี้

ตัวแบบการถดถอยเป็นการหาความสัมพันธ์ของตัวแปรตามและตัวแปรอิสระ สามารถเขียนความสัมพันธ์ได้ดังนี้

$$Y = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m) + \varepsilon \qquad (1)$$

โดยที่ Y คือ จำนวนครั้งของเหตุการณ์ที่สนใจ ณ บริเวณใดบริเวณหนึ่งหรือช่วงใดช่วงหนึ่ง

β_0 คือ ค่าคงตัวของตัวแบบ

$\beta_1, \beta_2, \cdots, \beta_m$ คือ ค่าพารามิเตอร์ของตัวแปรอิสระที่มีอิทธิพลต่อตัวแปรตาม

$X_1, X_2, \cdots, X_m$ คือ ตัวแปรอิสระจำนวน m ตัวแปร

ε คือ ค่าคลาดเคลื่อนของตัวแบบ

ตัวแบบการถดถอยควอไซปัวซง (Quasi-Poisson regression model)

ตัวแบบการถดถอยปัวซงเป็นตัวแบบที่ตัวแปรตามมีการแจกแจงปัวซงมีพารามิเตอร์ 1 ตัวคือพารามิเตอร์ค่าเฉลี่ย μ มีฟังก์ชันมวลน่าจะเป็น ดังนี้

$$P(Y_i) = \frac{e^{-\mu_i}\mu_i^{Y_i}}{Y_i!} \; ; Y_i = 0,1,2,\ldots ; i = 1,2,\ldots,n \tag{2}$$

ตัวแบบควอไซปัวซงเป็นตัวแบบที่ตัวแปรตามมีการแจกแจงปัวซงและมีการประมาณค่าด้วยวิธีกึ่งภาวะน่าจะเป็น (Quasi-likelihood) ซึ่งการประมาณค่าด้วยวิธีนี้จะใช้กับข้อมูลที่มีค่าความแปรปวนมากกว่าค่าเฉลี่ย โดยในการประมาณจะใช้ค่าเฉลี่ยและความแปรปรวนของฟังก์ชันของค่าเฉลี่ย (Wedderburn, 1974) มีพารามิเตอร์ 2 ตัว คือ พารามิเตอร์ค่าเฉลี่ย μ และพารามิเตอร์ Over-dispersion θ เมื่อให้ $Y_i \; ; i = 1,2,\ldots,n$ เป็นตัวแปรสุ่มที่มีค่าเฉลี่ยเท่ากับ μ_i และความแปรปรวนขึ้นอยู่กับฟังก์ชันของค่าเฉลี่ย หรือ $V(Y_i) \propto f(\mu_i)$ จะได้สมการที่ใช้ประมาณค่าเฉลี่ยและความแปรปรวนขึ้นอยู่กับฟังก์ชันค่าเฉลี่ย ดังสมการที่ (3)

$$\frac{\partial Q(Y_i;\mu_i)}{\partial \mu_i} = \frac{Y_i - \mu_i}{f(\mu_i)} \tag{3}$$

โดยมีค่าเฉลี่ยและความแปรปรวนของ Y_i คือ $E(Y_i) = \mu_i$ และ $Var(Y_i) = f(\mu_i) = \theta\mu_i \;\; (\theta > 1, \mu_i > 0)$

เมื่อ $Q(Y_i, \mu_i)$ คือ ฟังก์ชันกึ่งภาวะน่าจะเป็น (Quasi-likelihood function)

θ คือ พารามิเตอร์ Over-dispersion $(\theta > 1)$

ตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มาก (Zero inflated negative binomial regression model)

ให้ $Y_i \; ; i = 1,2,\ldots,n$ เป็นตัวแปรสุ่มที่มีการแจกแจงทวินามเชิงลบที่มีศูนย์มากมีพารามิเตอร์ 3 ตัวคือ พารามิเตอร์ค่าเฉลี่ย μ พารามิเตอร์การกระจาย k และพารามิเตอร์ของความน่าจะเป็นที่จะเกิดศูนย์ π หรือเขียนแทนด้วย $Y \sim ZINB(\mu, k, \pi)$ และฟังก์ชันมวลน่าจะเป็น คือ

$$Pr(Y_i = y_i) = \begin{cases} \pi_i + (1-\pi_i)\left[\dfrac{1/k}{(1/k)+\mu_i}\right]^{1/k} & ; y_i = 0 \\ (1-\pi_i)\left[\dfrac{\Gamma((1/k)+y_i)}{\Gamma(1/k)\,y_i!} \cdot \dfrac{k^{y_i}\mu_i^{y_i}}{(1+k\mu_i)^{y_i+\frac{1}{k}}}\right] & ; y_i > 0 \end{cases} \tag{4}$$

โดยมีค่าเฉลี่ยและความแปรปรวนของ Y_i คือ $E(Y_i) = (1-\pi_i)\mu_i$ และ $Var(Y_i) = (1-\pi_i)\mu_i(1+\pi_i\mu_i + k\mu_i)$

สมการของค่าประมาณพารามิเตอร์โดยวิธีภาวะน่าเป็นสูงสุดจะได้ฟังก์ชันล็อกภาวะน่าจะเป็น (Log likelihood: LL) ดังนี้

$$LL = \sum_{i=1}^{n} \left\{ I_{(y_i=0)} \log \left[\pi_i + (1-\pi_i) \left(\frac{1/k}{(1/k)+\mu_i} \right)^{1/k} \right] + I_{(y_i>0)} \log \Gamma \left(\frac{1}{k} + y_i \right) \right.$$
$$\left. + y_i \log k + y_i \log \mu_i - \log \Gamma \left(\frac{1}{k} \right) - \log y_i ! - \left(y_i + \frac{1}{k} \right) \log \left(1 + k\mu_i \right) \right\} \quad (5)$$

เกณฑ์ที่ใช้ในการเปรียบเทียบตัวแบบ

รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root mean square error: RMSE)

$$RMSE = \sqrt{\frac{1}{n} \left(\sum_{i=1}^{n} \left(\hat{Y}_i - Y_i \right)^2 \right)} \quad ; i = 1, \ldots, n \quad (6)$$

ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Absolute Average error: AAE)

$$AAE = \frac{1}{n} \sum_{i=1}^{n} \left(\left| \hat{Y}_i - Y_i \right| \right) \quad ; i = 1, \ldots, n \quad (7)$$

โดยที่ $\hat{Y}_i$ คือ ค่าประมาณของตัวแปรตามของค่าสังเกตที่ i

Y_i คือ ค่าสังเกตที่ i

n คือ ขนาดตัวอย่าง

ผลการวิจัยและวิจารณ์ผล

ผลการศึกษาจะแบ่งเป็น 2 ส่วน คือ ส่วนที่ 1 สำหรับข้อมูลจริงและส่วนที่ 2 สำหรับข้อมูลจำลอง มีรายละเอียดดังนี้

สำหรับข้อมูลจริงจะแบ่งเป็น 3 กรณี ดังนี้

กรณีที่ 1 เมื่อตัวอย่างขนาดเล็ก (n = 17)

กรณีที่ 2 เมื่อตัวอย่างขนาดกลาง (n = 32)

กรณีที่ 3 เมื่อตัวอย่างขนาดใหญ่ (n = 56)

สำหรับข้อมูลจำลองจะแบ่งเป็น 6 กรณี ดังนี้

กรณีที่ 1 คือ $k = 1.25$, $\pi = 0.25$, $n = 15, 20, 30, 35, 45, 50$

กรณีที่ 2 คือ $k = 1.50$, $\pi = 0.25$, $n = 15, 20, 30, 35, 45, 50$

กรณีที่ 3 คือ $k = 1.75$, $\pi = 0.25$, $n = 15, 20, 30, 35, 45, 50$

กรณีที่ 4 คือ $k = 1.25$, $\pi = 0.50$, $n = 15, 20, 30, 35, 45, 50$

กรณีที่ 5 คือ $k = 1.50$, $\pi = 0.50$, $n = 15, 20, 30, 35, 45, 50$

กรณีที่ 6 คือ $k = 1.75$, $\pi = 0.50$, $n = 15, 20, 30, 35, 45, 50$

แสดงผลการวิจัยดังตารางที่ 1 ถึงตารางที่ 7 ดังนี้

ส่วนที่ 1 การศึกษาสำหรับข้อมูลจริงแบ่งเป็น 3 กรณี มีรายละเอียดดังนี้

ตารางที่ 1 จำนวนการเกิดอุบัติเหตุและจำนวนผู้บาดเจ็บของตัวอย่างขนาดเล็ก กลางและใหญ่

ขนาดตัวอย่าง					
เล็ก		กลาง		ใหญ่	
จำนวนผู้บาดเจ็บ	จำนวนอุบัติเหตุ	จำนวนผู้บาดเจ็บ	จำนวนอุบัติเหตุ	จำนวนผู้บาดเจ็บ	จำนวนอุบัติเหตุ
0	5	0	17	0	27
1	5	1	6	1	11
2	2	2	4	2	7
3	1	3	3	3	6
5	1	5	2	6	4
9	1			10	1
13	1				
45	1				
รวม	17		32		56

ตารางที่ 2 การทดสอบภาวะสารูปดี ค่าความแปรปรวนและค่าเฉลี่ยของตัวอย่างขนาดเล็ก กลางและใหญ่

ขนาดตัวอย่าง	การทดสอบภาวะสารูปดี				ค่าความแปรปรวน	ค่าเฉลี่ย
	การแจกแจงปัวซง		การแจกแจงทวินามเชิงลบที่มีศูนย์มาก			
	ค่า Pearson χ^2	ค่า AIC	ค่า Pearson χ^2	ค่า AIC		
เล็ก	12.0796	176.87	**1.8766**	**93.6237**	112.1730	4.9412
กลาง	2.2338	106.8896	**1.3112**	**106.7538**	2.0303	1.0313
ใหญ่	2.2423	201.6330	**1.1732**	**193.1850**	4.1272	1.3750

กรณีที่ 1 เมื่อตัวอย่างขนาดเล็ก ใช้ข้อมูลถนนทางหลวงหมายเลข 0403 (สี่แยกกบินทร์บุรี - วังน้ำเขียว) จังหวัดปราจีนบุรี กิโลเมตรที่ 0208+660 ถึง 0219+000 เกิดอุบัติเหตุ 17 ครั้ง จากตารางที่ 1 จำนวนอุบัติเหตุประมาณ 29% ไม่มีผู้ได้รับบาดเจ็บ ($Y = 0$) และจากตารางที่ 2 พบว่า ข้อมูลมีความแปรปรวนมากกว่าค่าเฉลี่ยและเมื่อตรวจสอบการทดสอบ

ภาวะสารูปดีพบว่า จำนวนผู้บาดเจ็บมีการแจกแจงทวินามเชิงลบที่มีศูนย์มาก นำข้อมูลมาสร้างตัวแบบและทำการคัดเลือกตัวแบบที่เหมาะสม โดยดูจากรากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยและค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย ดังนี้

ตารางที่ 3 ค่า RMSE และ ค่า AAE ของจำนวนการเกิดอุบัติเหตุที่มีตัวอย่างขนาดเล็ก

ตัวแบบ	RMSE	AAE
QP	9.4281	1.64×10^{-13}
ZINB	9.7413	0.2963

จากตารางที่ 3 พบว่าตัวแบบการถดถอยควอไซปัวซงมีความเหมาะสมมากกว่าตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มากเนื่องจากมีค่า RMSE และ ค่า AAE ที่ต่ำกว่า ซึ่งจะได้สมการพยากรณ์ คือ

$$\hat{Y} = \exp\left(-1.6435 + 2.6190\left(X_1\right) - 0.2714\left(X_2\right) + 1.7554\left(X_3\right)\right) \quad (8)$$

กรณีที่ 2 เมื่อตัวอย่างขนาดกลาง ใช้ข้อมูลถนนทางหลวงหมายเลข 0403 (สี่แยกกบินทร์บุรี - วังน้ำเขียว) จังหวัดปราจีนบุรี กิโลเมตรที่ 0166+571 ถึง 0208+660 เกิดอุบัติเหตุ 32 ครั้ง จากตารางที่ 1 จำนวนอุบัติเหตุประมาณ 50% ไม่มีผู้ได้รับบาดเจ็บ $(Y = 0)$ และจากตารางที่ 2 พบว่า ข้อมูลมีความแปรปรวนมากกว่าค่าเฉลี่ยและเมื่อตรวจสอบการทดสอบภาวะสารูปดีพบว่า จำนวนผู้บาดเจ็บมีการแจกแจงทวินามเชิงลบที่มีศูนย์มาก นำข้อมูลมาสร้างตัวแบบและทำการคัดเลือกตัวแบบที่เหมาะสม โดยดูจากรากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยและค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย ดังนี้

ตารางที่ 4 ค่า RMSE และ ค่า AAE ของจำนวนการเกิดอุบัติเหตุที่มีตัวอย่างขนาดกลาง

ตัวแบบ	RMSE	AAE
QP	1.4175	3.88×10^{-11}
ZINB	1.4176	0.0007

จากตารางที่ 4 พบว่าตัวแบบการถดถอยควอไซปัวซงมีความเหมาะสมมากกว่าตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มากเนื่องจากมีค่า RMSE และ ค่า AAE ที่ต่ำกว่า ซึ่งจะได้สมการพยากรณ์ คือ

$$\hat{Y} = \exp\left(-0.0996 + 0.1901\left(X_1\right) + 0.1297\left(X_2\right) - 0.1337\left(X_3\right)\right) \quad (9)$$

กรณีที่ 3 เมื่อตัวอย่างขนาดใหญ่ ใช้ข้อมูลถนนทางหลวงหมายเลข 0601 และถนนทางหลวงหมายเลข 0602 (ป่าเตียน-บ้านสิ้ว) จังหวัดจันทบุรี กิโลเมตรที่ 0287+900 ถึง 0324+300 เกิดอุบัติเหตุ 56 ครั้ง จากตารางที่ 1 จำนวนอุบัติเหตุประมาณ 50% ไม่มีผู้ได้รับบาดเจ็บ $(Y = 0)$ และจากตารางที่ 2 พบว่า ข้อมูลมีความแปรปรวนมากกว่าค่าเฉลี่ยและเมื่อ

ตรวจสอบการทดสอบภาวะสารูปดีพบ จำนวนผู้บาดเจ็บว่ามีการแจกแจงทวินามเชิงลบที่มีศูนย์มาก นำข้อมูลมาสร้างตัวแบบและทำการคัดเลือกตัวแบบที่เหมาะสม โดยดูจากรากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยและค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย ดังนี้

ตารางที่ 5 ค่า RMSE และ ค่า AAE ของจำนวนการเกิดอุบัติเหตุที่มีตัวอย่างขนาดใหญ่

ตัวแบบ	RMSE	AAE
QP	1.6550	5.29×10^{-11}
ZINB	1.6560	0.0013

จากตารางที่ 5 พบว่าตัวแบบการถดถอยควอไซปัวซงมีความเหมาะสมมากกว่าตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มากเนื่องจากมีค่า RMSE และ ค่า AAE ที่ต่ำกว่า ซึ่งจะได้สมการพยากรณ์ คือ

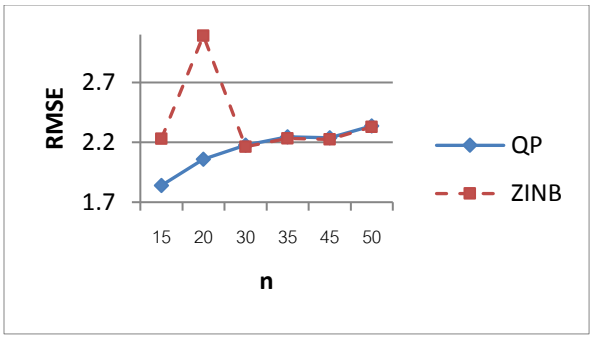
$$\hat{Y} = \exp\left(2.0464 + 0.2519\left(X_1\right) - 2.0853\left(X_2\right) + 0.2562\left(X_3\right)\right) \quad (10)$$

ส่วนที่ 2 การศึกษาสำหรับข้อมูลจำลองจะแบ่งเป็น 6 กรณี มีรายละเอียดดังนี้

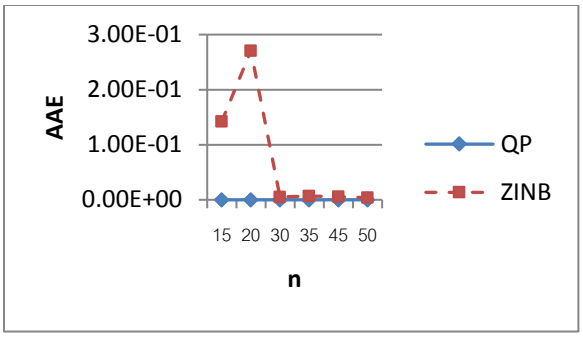
ตารางที่ 6 RMSE และ AAE กรณีมีค่าความน่าจะเป็นที่จะเกิดศูนย์ เท่ากับ 0.25

พารามิเตอร์ k	ตัวแบบ	เกณฑ์การวัด	ขนาดตัวอย่าง					
			เล็ก		กลาง		ใหญ่	
			15	20	30	35	45	50
1.25	QP	RMSE	**1.8395**	**2.0588**	2.1756	2.2449	2.2385	2.3375
		AAE	**1.7×10^{-7}**	**8.9×10^{-8}**	**3.6×10^{-8}**	**3.5×10^{-8}**	**1.9×10^{-8}**	**1.8×10^{-7}**
	ZINB	RMSE	2.2296	3.0918	**2.1635**	**2.2322**	**2.2246**	**2.3301**
		AAE	0.1422	0.2705	0.0049	0.0070	0.0057	0.0038
1.50	QP	RMSE	**2.0105**	**2.0682**	2.2045	**2.2910**	2.4323	2.4498
		AAE	**2.2×10^{-7}**	**6.8×10^{-8}**	**2.9×10^{-8}**	**2.8×10^{-8}**	**2.2×10^{-8}**	**1.3×10^{-8}**
	ZINB	RMSE	2.0997	3.7744	**2.1950**	2.7298	**2.4236**	**2.4460**
		AAE	0.0550	0.5640	0.0058	0.1258	0.0091	0.0016
1.75	QP	RMSE	**1.9385**	**2.1652**	**2.3184**	**2.2922**	2.4273	2.4614
		AAE	**1.7×10^{-7}**	**1.1×10^{-7}**	**5.5×10^{-8}**	**4.0×10^{-8}**	**2.8×10^{-8}**	**2.0×10^{-8}**
	ZINB	RMSE	2.1965	2.2338	2.3256	2.4016	**2.4193**	**2.4549**
		AAE	0.1221	0.0569	0.0128	0.0343	0.0152	0.0090

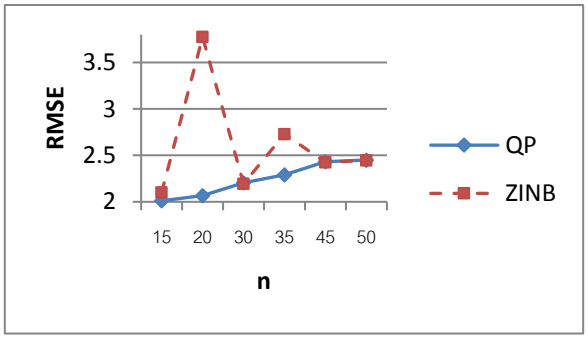
‘‘ตัวหนา หมายถึง ตัวแบบนั้นให้ค่าเกณฑ์การวัดที่ต่ำกว่า’’



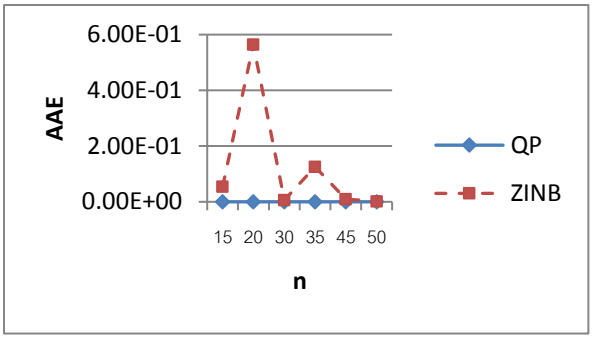
ภาพที่ 1 ค่า RMSE ของตัวแบบ QP และ ZINB ที่ $k = 1.25$ และ $\pi = 0.25$



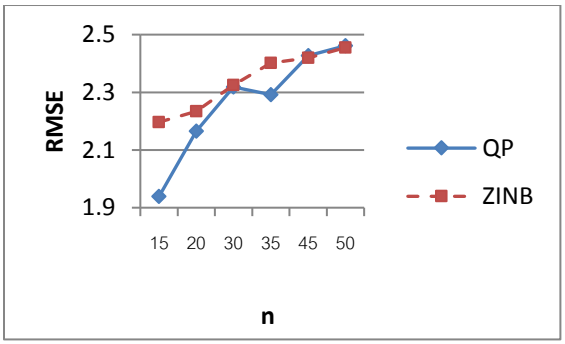
ภาพที่ 2 ค่า AAE ของตัวแบบ QP และ ZINB ที่ $k = 1.25$ และ $\pi = 0.25$



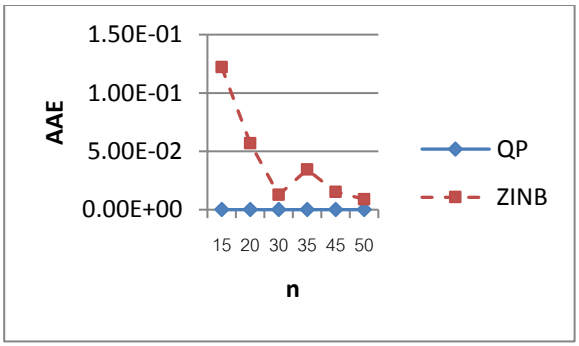
ภาพที่ 3 ค่า RMSE ของตัวแบบ QP และ ZINB ที่ $k = 1.50$ และ $\pi = 0.25$



ภาพที่ 4 ค่า AAE ของตัวแบบ QP และ ZINB ที่ $k = 1.50$ และ $\pi = 0.25$



ภาพที่ 5 ค่า RMSE ของตัวแบบ QP และ ZINB ที่ $k = 1.75$ และ $\pi = 0.25$



ภาพที่ 6 ค่า AAE ของตัวแบบ QP และ ZINB ที่ $k = 1.75$ และ $\pi = 0.25$

กรณีที่ 1 คือ กรณีที่มีค่าพารามิเตอร์การกระจาย (k) เท่ากับ 1.25 และค่าความน่าจะเป็นที่จะเกิดศูนย์ (π) เท่ากับ 0.25 และขนาดตัวอย่าง (n) เท่ากับ 15, 20, 30, 35, 45 และ 50

จากตารางที่ 6 ภาพที่ 1 และภาพที่ 2 เมื่อพิจารณารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยและค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย พบว่า ตัวอย่างขนาดเล็ก (15, 20) ตัวแบบการถดถอยควอไซปัวซงให้ค่า RMSE และค่า AAE ต่ำกว่าตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มาก ส่วนตัวอย่างขนาดกลาง (30, 35) และตัวอย่างขนาดใหญ่ (45, 50) ตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มากให้ค่า RMSE และค่า AAE ใกล้เคียงกับตัวแบบการถดถอยควอไซปัวซง

กรณีที่ 2 คือ กรณีที่มีค่าพารามิเตอร์การกระจาย (k) เท่ากับ 1.50 และค่าความน่าจะเป็นที่จะเกิดศูนย์ (π) เท่ากับ 0.25 และขนาดตัวอย่าง (n) เท่ากับ 15, 20, 30, 35, 45 และ 50

จากตารางที่ 6 และภาพที่ 3 เมื่อพิจารณารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย พบว่า ตัวอย่างขนาดเล็ก (15, 20) ตัวแบบการถดถอยควอไซปัวซงให้ค่า RMSE ต่ำกว่าตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มากสำหรับตัวอย่างขนาดกลาง (30, 35) พบว่า ตัวอย่างขนาด 30 ตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มากให้ค่า RMSE ต่ำกว่าตัวแบบการถดถอยควอไซปัวซง ส่วนตัวอย่างขนาด 35 ตัวแบบการถดถอยควอไซปัวซงให้ค่า RMSE ต่ำกว่าตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มากและตัวอย่างขนาดใหญ่ (45, 50) ตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มากให้ค่า RMSE ใกล้เคียงตัวแบบการถดถอยควอไซปัวซง

เมื่อพิจารณาจากค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยจากตารางที่ 6 และภาพที่ 4 พบว่า ตัวอย่างขนาดเล็ก (15, 20) และตัวอย่างขนาดกลาง (30, 35) ตัวแบบการถดถอยควอไซปัวซงให้ค่า AAE ต่ำกว่าตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มากและตัวอย่างขนาดใหญ่ (45, 50) ตัวแบบการถดถอยควอไซปัวซงมีค่า AAE ใกล้เคียงกับตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มาก

กรณีที่ 3 คือ กรณีที่มีค่าพารามิเตอร์การกระจาย (k) เท่ากับ 1.75 และค่าความน่าจะเป็นที่จะเกิดศูนย์ (π) เท่ากับ 0.25 และขนาดตัวอย่าง (n) เท่ากับ 15, 20, 30, 35, 45 และ 50

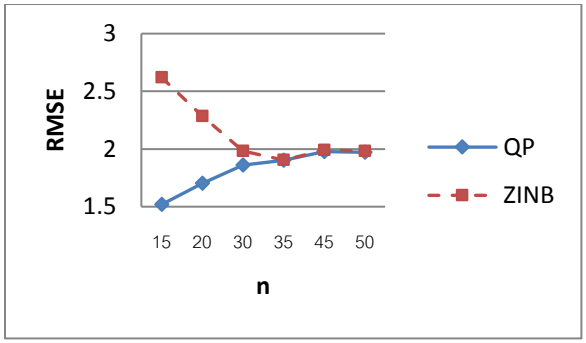
จากตารางที่ 6 ภาพที่ 5 เมื่อพิจารณารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย พบว่า ตัวอย่างขนาดเล็ก (15, 20) และตัวอย่างขนาดกลาง (30, 35) ตัวแบบการถดถอยควอไซปัวซงให้ค่า RMSE ต่ำกว่าตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มากและตัวอย่างขนาดใหญ่ (45, 50) ตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มากให้ค่า RMSE ใกล้เคียงกับตัวแบบการถดถอยควอไซปัวซง

เมื่อพิจารณาจากค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยจากตารางที่ 6 และภาพที่ 6 พบว่า ตัวแบบการถดถอยควอไซปัวซงมีค่า AAE ต่ำกว่าตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มากทุกขนาดตัวอย่าง

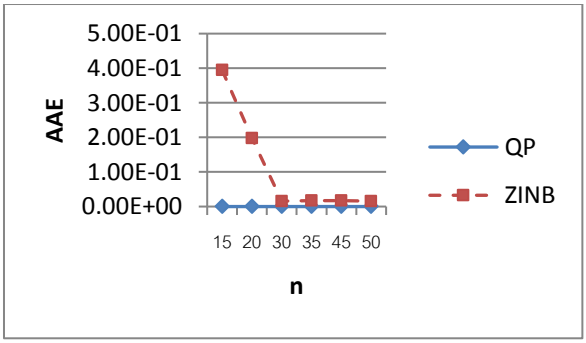
ตารางที่ 7 RMSE และ AAE กรณีมีค่าความน่าจะเป็นที่จะเกิดศูนย์เท่ากับ 0.50

พารามิเตอร์ k	ตัวแบบ	เกณฑ์การวัด	ขนาดตัวอย่าง					
			เล็ก		กลาง		ใหญ่	
			15	20	30	35	45	50
1.25	QP	RMSE	**1.5189**	**1.7011**	**1.8604**	**1.9032**	**1.9775**	**1.9706**
		AAE	**7.7×10^{-7}**	**3.4×10^{-7}**	**4.5×10^{-7}**	**4.3×10^{-8}**	**3.3×10^{-8}**	**3.1×10^{-8}**
	ZINB	RMSE	2.6198	2.2836	1.9814	1.9041	1.9914	1.9814
		AAE	0.3945	0.1974	0.0157	0.0167	0.0175	0.0157
1.50	QP	RMSE	**1.6427**	**1.8107**	**1.9842**	**1.9755**	**2.0848**	**2.0928**
		AAE	**7.3×10^{-7}**	**5.2×10^{-8}**	**9.9×10^{-8}**	**4.4×10^{-8}**	**3.7×10^{-8}**	**4.2×10^{-8}**
	ZINB	RMSE	2.3314	2.7327	2.4152	2.0114	2.0868	2.1028
		AAE	0.2264	0.2729	0.1333	0.0414	0.0050	0.0223
1.75	QP	RMSE	**1.5499**	**1.7826**	1.8576	**1.8981**	**1.9726**	**2.0195**
		AAE	**1.7×10^{-7}**	**1.8×10^{-7}**	**8.5×10^{-8}**	**2.5×10^{-7}**	**4.5×10^{-8}**	**3.9×10^{-8}**
	ZINB	RMSE	3.0940	2.6241	**1.8491**	1.9401	2.0018	2.0350
		AAE	0.5448	0.2590	0.0371	0.0265	0.0140	0.0261

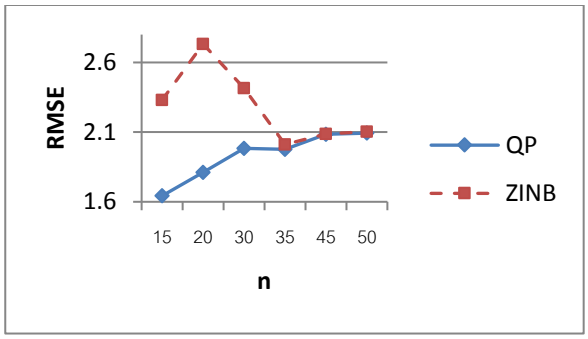
“ตัวหนา หมายถึง ตัวแบบนั้นให้ค่าเกณฑ์การวัดที่ต่ำกว่า”



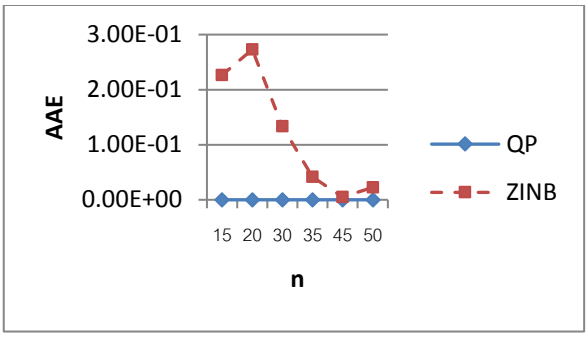
ภาพที่ 7 ค่า RMSE ของตัวแบบ QP และ ZINB ที่ k = 1.25 และ π = 0.50



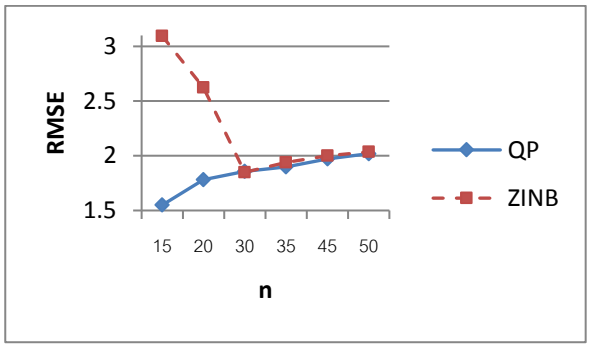
ภาพที่ 8 ค่า AAE ของตัวแบบ QP และ ZINB ที่ k = 1.25 และ π = 0.50



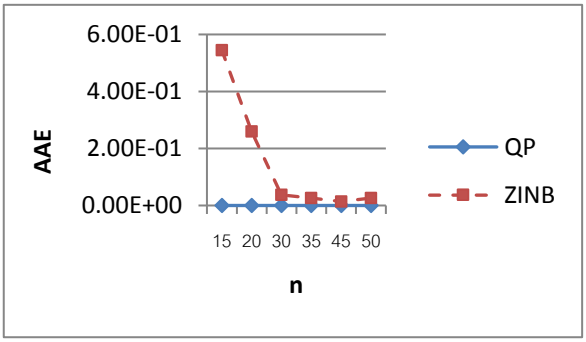
ภาพที่ 9 ค่า RMSE ของตัวแบบ QP และ ZINB ที่ $k = 1.50$ และ $\pi = 0.50$



ภาพที่ 10 ค่า AAE ของตัวแบบ QP และ ZINB ที่ $k = 1.50$ และ $\pi = 0.50$



ภาพที่ 11 ค่า RMSE ของตัวแบบ QP และ ZINB ที่ $k = 1.75$ และ $\pi = 0.50$



ภาพที่ 12 ค่า AAE ของตัวแบบ QP และ ZINB ที่ $k = 1.75$ และ $\pi = 0.50$

กรณีที่ 4 คือ กรณีที่มีค่าพารามิเตอร์การกระจาย (k) เท่ากับ 1.25 และค่าความน่าจะเป็นที่จะเกิดศูนย์ (π) เท่ากับ 0.50 และขนาดตัวอย่าง (n) เท่ากับ 15, 20, 30, 35, 45 และ 50

จากตารางที่ 7 และภาพที่ 7 เมื่อพิจารณารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบการถดถอย พบว่า ตัวอย่างขนาดเล็ก (15, 20) และตัวอย่างขนาดกลาง (30, 35) ตัวแบบการถดถอยควอไซป์วซงให้ค่า RMSE ต่ำกว่าตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มากและตัวอย่างขนาดใหญ่ (45, 50) ตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มากให้ค่า RMSE ใกล้เคียงกับตัวแบบการถดถอยควอไซป์วซง

เมื่อพิจารณาจากค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยจากตารางที่ 7 และภาพที่ 8 พบว่า ตัวแบบการถดถอยควอไซป์วซง มีค่า AAE ต่ำกว่าตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มากทุกขนาดตัวอย่าง

กรณีที่ 5 คือ กรณีที่มีค่าพารามิเตอร์การกระจาย (k) เท่ากับ 1.50 และค่าความน่าจะเป็นที่จะเกิดศูนย์ (π) เท่ากับ 0.50 และขนาดตัวอย่าง (n) เท่ากับ 15, 20, 30, 35, 45 และ 50

จากตารางที่ 7 และภาพที่ 9 เมื่อพิจารณารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบการถดถอย พบว่า ตัวอย่างขนาดเล็ก (15, 20) และตัวอย่างขนาดกลาง (30, 35) ตัวแบบการถดถอยควอไซป์วซงให้ค่า RMSE ต่ำกว่าตัวแบบ

การถดถอยทวินามเชิงลบที่มีศูนย์มากและตัวอย่างขนาดใหญ่ (45, 50) ตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มากให้ค่า RMSE ใกล้เคียงกับตัวแบบการถดถอยควอไซปัวซง

เมื่อพิจารณาจากค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยจากตารางที่ 7 และภาพที่ 10 พบว่า ตัวแบบการถดถอยควอไซปัวซงมีค่า AAE ต่ำกว่าตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มากทุกขนาดตัวอย่าง

กรณีที่ 6 คือ กรณีที่มีค่าพารามิเตอร์การกระจาย (k) เท่ากับ 1.75 และค่าความน่าจะเป็นที่จะเกิดศูนย์ (π) เท่ากับ 0.50 และขนาดตัวอย่าง (n) เท่ากับ 15, 20, 30, 35, 45 และ 50

จากตารางที่ 7 และภาพที่ 11 เมื่อพิจารณารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย พบว่า ตัวอย่างขนาดเล็ก (15, 20) ตัวแบบการถดถอยควอไซปัวซงให้ค่า RMSE ต่ำกว่าตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มาก ส่วนตัวอย่างขนาดกลาง (30, 35) และตัวอย่างขนาดใหญ่ (45, 50) ตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มากให้ค่า RMSE ใกล้เคียงกับตัวแบบการถดถอยควอไซปัวซง

เมื่อพิจารณาจากค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยจากตารางที่ 7 และภาพที่ 12 พบว่า ตัวแบบการถดถอยควอไซปัวซงมีค่า AAE ต่ำกว่าตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มากทุกขนาดตัวอย่าง

สรุปผลการวิจัย

ในที่นี้จะแบ่งการสรุปผลการวิจัยออกเป็น 2 ส่วน คือ ส่วนที่ 1 ผลการวิจัยสำหรับข้อมูลจริงและส่วนที่ 2 ผลการวิจัยสำหรับข้อมูลจำลอง

ส่วนที่ 1 ผลการวิจัยสำหรับข้อมูลจริง พบว่า เมื่อข้อมูลที่ศึกษามีค่าความแปรปรวนมากกว่าค่าเฉลี่ย สำหรับตัวอย่างขนาดเล็ก โดยที่ประมาณ 29% ของตัวแปรตามมีค่าเป็นศูนย์และสำหรับตัวอย่างขนาดกลางและใหญ่ โดยที่ประมาณ 50% ของตัวแปรตามมีค่าเป็นศูนย์ พบว่า เมื่อพิจารณาจากค่า RMSE และ AAE ตัวแบบการถดถอยควอไซปัวซงมีความเหมาะสมกับข้อมูลจริงสำหรับทุก ๆ ขนาดของข้อมูล ซึ่งสอดคล้องกับกรณีของตัวอย่างขนาดเล็กที่ π =0.25 และกรณีของตัวอย่างขนาดกลางและใหญ่ที่ π =0.50 โดยพบว่าตัวแบบการถดถอยควอไซปัวซงมีความเหมาะสมมากกว่าตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มาก

ส่วนที่ 2 ผลการวิจัยสำหรับข้อมูลจำลอง พบว่า โดยภาพรวม เมื่อพิจารณาจากค่า RMSE และ AAE ตัวแบบการถดถอยควอไซปัวซงมีความเหมาะสมมากกว่าตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มาก เมื่อพิจารณาค่า RMSE สำหรับตัวอย่างขนาดเล็ก ที่ π =0.25 เมื่อ k = 1.25, 1.50 และ 1.75 พบว่า ตัวแบบการถดถอยควอไซปัวซงมีความเหมาะสมมากกว่า สำหรับตัวอย่างขนาดกลางและใหญ่ที่ π =0.25 เมื่อ k =1.25, 1.50 และ 1.75 พบว่า ตัวแบบการถดถอยทวินามเชิงลบที่มีศูนย์มากมีความเหมาะสมมากกว่าและสำหรับตัวอย่างขนาดเล็ก กลางและใหญ่ที่ π =0.50 เมื่อ k =1.25, 1.50 และ 1.75 ตัวแบบการถดถอยควอไซปัวซงมีความเหมาะสมมากกว่า และเมื่อพิจารณาค่า AAE สำหรับตัวอย่างขนาดเล็ก กลางและใหญ่ที่ π =0.25 และ 0.50 เมื่อ k =1.25, 1.50 และ 1.75 พบว่า ตัวแบบการถดถอยควอไซปัวซงมีความเหมาะสมมากกว่า

เอกสารอ้างอิง

Batra, M., Shah, A.F., Rajput, P., & Shah, I.A. (2016). Comparison of linear and zero-inflated negative binomial regression models for appraisal of risk factors associated with dental caries. *J Indian Soc Pedod Prev Dent*, *34*(1), 71-75.

Charatiam, N. (2010). *Comparative Test Statistics for Zero-Inflated Generalized Poisson Regression Model against Generalized Poisson Regression Model in the Presence of Covariate Outliers.* Master of Science Thesis, Thammasat University. (in Thai)

Doyle, S.R. 2009. Examples of Computing Power for Zero-Inflated and Overdispersed Count Data. *Journal of Modern Applied Statistical Methods*, Vol. *8*(2), 360-376.

Hu, M.C., Pavlicova, M., & Nunes, E.V. (2011). Zero-inflated and Hurdle Models of Count Data with Extra Zeros: Examples form an HIV – Risk Reduction Intervention Trial. *Am J Drug Alcohol Abuse*, *37*(5), 367 – 375.

Nitchanpunsri, K. (2011). *Comparative Study of Model Fit for Generalized Linear Models: Zero-Inflated Distributions.* Master of Science Thesis, Thammasat University. (in Thai)

Pornapraditpun, S. (2009). *The Dispersion Estimation under Generalized Linear Model with Negative Binomial Distribution for Small Sample.* Master of Science Thesis, Thammasat University. (in Thai)

Potts, J.M., & Elih, J. (2006). Comparing species abundance models. *Ecological Modeling*, *199*, 153-163.

Sileshi, G. (2006). Selecting the right statistical model for analysis of insect count data by using information theoretic measures. *Bulletin of Entomological Research*, *96*, 479-488.

Ver Hoef, J.M., & Boveng, P.L. (2007). Quasi-Poisson VS. Negative binomial Regression: How Should We Model Overdispersed Count Data. *Ecology Society of America Journal*, *88*(1), 2766-2772.

Wedderburn, R.W.M. (1974). Quasi-likelihood function, Generalized linear models and the Gauss-Newton method. *Biometrika*, *61*, 439-447.

Yau, K.K.W., Wang, K., & Lee, A.H. (2003). Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. *Biometrical Journal*, *45*, 437-452.

Yusuf, O.B., Bello, T., & Gureje, O. (2017). Zero Inflated Poisson and Zero Inflated Negative Binomial Models with Application to Number of Falls in the Elderly. *Biostatistics and Biometrics Open Access Journal*, *1*(4), 1-7.