บทความวิจัย

การเปรียบเทียบประสิทธิภาพวิธีการคัดเลือกตัวแปรอิสระสำหรับตัวแบบการถดถอยไวบูลไม่ต่อเนื่อง

มณฑิรา ดวงสาพล* และ ปวีณ์กร มิ่งเชื้อ
สาขาวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

* ผู้นิพนธ์ประสานงาน โทรศัพท์ 0 2564 4440–59 ต่อ 2101 อีเมล: monthira@mathstat.sci.tu.ac.th DOI: 10.14416/j.kmutnb.2023.03.007
รับเมื่อ 29 กันยายน 2565 แก้ไขเมื่อ 15 ธันวาคม 2565 ตอบรับเมื่อ 12 มกราคม 2566 เผยแพร่ออนไลน์ 28 มีนาคม 2566

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อนำเสนอประสิทธิภาพการคัดเลือกตัวแปรอิสระของตัวแบบการถดถอยไวบูลไม่ต่อเนื่องทั้งหมด 4 วิธี ได้แก่ วิธีการถดถอยทีละขั้น วิธีบูตสแทร็ปทีละขั้น วิธีแบบเบส์ภายใต้การแจกแจงก่อน คือ การแจกแจงปรกติ และวิธีแบบเบส์ภายใต้การแจกแจงก่อน คือ การแจกแจงลาปลาซ โดยเปรียบเทียบประสิทธิภาพของวิธีการคัดเลือกตัวแปรด้วยอัตราความสำเร็จ รวมถึงได้ศึกษาประสิทธิภาพของการประมาณค่าพารามิเตอร์ด้วยค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยจากการจำลองด้วยเทคนิคมอนติคาร์โลภายใต้สถานการณ์ที่ตัวแปรอิสระมีความสัมพันธ์เชิงเส้นพหุ และไม่มีความสัมพันธ์เชิงเส้นพหุ ตัวแบบผ่านฟังก์ชันเชื่อมโยงแบบล็อก-ล็อกและลอจิต และลักษณะของข้อมูลตัวแปรตามมีการกระจายต่ำกว่าเกณฑ์ และการกระจายเกินเกณฑ์ นอกจากนั้นผู้วิจัยยังนำทั้ง 4 วิธีมาประยุกต์ใช้กับข้อมูลจริง ผลจากการจำลองโดยสรุปพบว่า โดยส่วนใหญ่วิธีบูตสแทร็ปทีละขั้นให้ประสิทธิภาพดีที่สุด วิธีแบบเบส์จะมีประสิทธิภาพรองลงมา และให้ประสิทธิภาพดีที่สุดเมื่อขนาดตัวอย่างเท่ากับ 100 ส่วนวิธีการถดถอยทีละขั้นจะมีประสิทธิภาพน้อยที่สุดแต่จะมีประสิทธิภาพมากขึ้นเมื่อขนาดตัวอย่างมากขึ้น และผลจากการประยุกต์ใช้กับข้อมูลจริงพบว่า วิธีบูตสแทร็ปทีละขั้นให้ประสิทธิภาพดีที่สุด

คำสำคัญ: การประมาณค่าแบบเบส์ วิธีบูตสแทร็ป ข้อมูลเชิงนับ ดีเวียนซ์ วิธีการถดถอยทีละขั้น

การอ้างอิงบทความ: มณฑิรา ดวงสาพล และ ปวีณ์กร มิ่งเชื้อ, “การเปรียบเทียบประสิทธิภาพวิธีการคัดเลือกตัวแปรอิสระสำหรับตัวแบบการถดถอยไวบูลไม่ต่อเนื่อง,” *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, 2566, doi: 10.14416/j.kmutnb.2023.03.007.

Research Article

Efficiency Comparison of Independent Variables Selection Methods for Discrete Weibull Regression Model

Monthira Duangsaphon* and Paweekon Mingchue
Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Rangsit Campus, Pathum Thani, Thailand

* Corresponding Author, Tel. 0 2564 4440–59 Ext. 2101, E–mail: monthira@mathstat.sci.tu.ac.th DOI: 10.14416/j.kmutnb.2023.03.007
Received 29 September 2022; Revised 15 November 2022; Accepted 12 January 2023; Published online: 28 March 2023

Abstract

This research aimed to propose the performance of independent variables selection of discrete Weibull regression model with four different methods, namely the stepwise regression, the stepwise bootstrap and the Bayesian method based on Normal and Laplace prior distributions. The comparison among methods was conducted in terms of the Success Rate (SR) as well as the Mean of the Mean Square Error (MMSE) that studied the performance of the parameter estimation via the Monte Caro simulation technique. The explanatory variables were generated by both multicollinearity and no multicollinearity. The model was constructed by log-log and logit link functions. The response variable was considered for under-dispersion and over-dispersion data. Moreover, we apply four methods with real data. The findings show that results from the simulation study, the stepwise bootstrap method presents the best performance in most all the cases. The both schemes of Bayesian method present the second performance and present the best performance for sample size is 100. The stepwise regression presents the lowest performance but satisfactory performance when sample size increases. For real data, the stepwise bootstrap method presents the best performance.

Keywords: Bayesian Estimation, Bootstrap Method, Count Data, Deviance, Stepwise Regression Method

Please cite this article in press as: M. Duangsaphon and P. Mingchue, "Efficiency comparison of independent variables selection methods for discrete weibull regression model," *The Journal of KMUTNB*, 2023 (in Thai), doi: 10.14416/j.kmutnb.2023.03.007

1. บทนำ

ข้อมูลเชิงนับ (Count Data) พบได้ในงานหลากหลายด้าน ไม่ว่าจะเป็นด้านการแพทย์และสาธารณสุข ด้านวิศวกรรม ด้านประกันภัย และด้านสังคมศาสตร์ เป็นต้น ข้อมูลเชิงนับอาจมีรูปแบบการแจกแจงที่หลากหลายทางสถิติที่สามารถนำไปสร้างตัวแบบการวิเคราะห์การถดถอย (Regression Analysis) ได้ เช่น ตัวแบบการถดถอยปัวซง (Poisson Regression Model) เป็นตัวแบบที่เหมาะสมกับข้อมูลที่มีค่าความแปรปรวนเท่ากับค่าเฉลี่ย (Equi-Dispersion) [1] ตัวแบบการถดถอยทวินามเชิงลบ (Negative Binomial Regression Model) เป็นตัวแบบที่เหมาะสมกับข้อมูลที่มีการกระจายเกินเกณฑ์ (Over-Dispersion) [2] ตัวแบบการถดถอยคอนเวย์แมกซ์เวลล์ปัวซง (Conway-Maxwell-Poisson Regression Model) เป็นตัวแบบที่เหมาะสมกับข้อมูลที่มีการกระจายต่ำกว่าเกณฑ์ (Under-Dispersion) [2] และตัวแบบการถดถอยไวบูลไม่ต่อเนื่อง (Discrete Weibull Regression Model) ที่เหมาะสมทั้งข้อมูลที่มีการกระจายต่ำกว่าเกณฑ์และการกระจายเกินเกณฑ์ [3], [4] ซึ่งสิ่งที่โดดเด่นในตัวแบบ คือ มีพารามิเตอร์ที่สามารถอธิบายลักษณะการกระจายของข้อมูลตัวแปรตามได้ รวมถึงเป็นทางเลือกในการหาตัวแบบการถดถอยที่มีประสิทธิภาพสำหรับการประยุกต์ใช้กับข้อมูลจริง นอกจากนี้ในการประยุกต์ใช้กับข้อมูลจริง หากเราสามารถประเมินได้ว่าข้อมูลตัวแปรตามนั้นมีลักษณะการกระจายแบบใด ในเบื้องต้นเราจะสามารถเลือกตัวแบบการถดถอยที่เหมาะสมกับลักษณะการกระจายข้อมูลตัวแปรตามได้ จากนั้นจึงเลือกตัวแบบการถดถอยที่ดีที่สุดจากสถิติที่ใช้คัดเลือกตัวแบบต่อไป

มีงานวิจัยที่ได้นำเสนอเกี่ยวกับตัวแบบการถดถอยไวบูลไม่ต่อเนื่อง เช่น Klakattawi และคณะ[4] ได้เสนอตัวแบบการถดถอยการกระจายที่ปรับตัวอย่างง่ายสำหรับข้อมูลเชิงนับ โดยแสดงให้เห็นว่าตัวแบบการถดถอยไวบูลไม่ต่อเนื่องสามารถที่จะปรับตัวอย่างง่ายกับการกระจายตัวประเภทต่างๆ และ Yoo [5] ได้นำเสนอการประยุกต์ใช้ตัวแบบการถดถอยไวบูลไม่ต่อเนื่องกับข้อมูลจำนวนนับที่มีการสูญหายด้วยวิธี Multiple Imputation นอกจากนั้นยังมีงานวิจัยที่ได้นำเสนอตัวแบบการถดถอยไวบูลไม่ต่อเนื่องด้วยวิธีการแบบเบส์ [6], [7]

การวิเคราะห์ตัวแบบการถดถอยให้มีประสิทธิภาพขึ้นอยู่กับการเลือกตัวแบบ และการคัดเลือกตัวแปรอิสระที่มีอิทธิพลต่อตัวแปรตามที่เหมาะสม หากไม่มีการคัดเลือกตัวแปรอาจจะมีตัวแปรอิสระจำนวนมากเกินความจำเป็น รวมถึงอาจจะเกิดปัญหาความสัมพันธ์เชิงเส้นพหุระหว่างตัวแปรอิสระ (Multicollinearity) ซึ่งจะทำให้ค่าประมาณสัมประสิทธิ์การถดถอย และค่าพยากรณ์ที่ได้มีความคลาดเคลื่อนสูง การคัดเลือกตัวแปรด้วยวิธีการถดถอยทีละขั้น (Stepwise Regression Method) เป็นวิธีที่ใช้กันอย่างแพร่หลาย Sakate และคณะ [8] ได้นำเสนอการคัดเลือกเซตย่อยของตัวแปรอิสระของการถดถอยปัวซง โดยการใช้เกณฑ์สถิติที่เป็นฟังก์ชันของค่าดีเวียนซ์ (Deviance) บนพื้นฐานของกระบวนการคัดเลือกตัวแปรแบบทีละขั้น ส่วนการนำเทคนิคบูตสแทร็ป (Bootstrap Method) [9] มาประยุกต์ใช้ในประมาณค่าพารามิเตอร์ในตัวแบบต่างๆ เพื่อคัดเลือกตัวแปรเมื่อเกิดปัญหาความสัมพันธ์เชิงเส้นพหุ จะให้ประสิทธิภาพที่ดีภายใต้ระดับความสัมพันธ์ของตัวแปรอิสระ และขนาดตัวอย่างที่แตกต่างกัน [10]–[12] นอกจากนี้วิธีการแบบเบส์เป็นวิธีหนึ่งที่นำมาประยุกต์ใช้ในการประมาณค่าพารามิเตอร์สำหรับตัวแบบการถดถอย Haselimashhadi และคณะ [6] ได้นำเสนอการคัดเลือกตัวแปรอิสระสำหรับตัวแบบการถดถอยไวบูลไม่ต่อเนื่องด้วยวิธีการแบบเบส์จากการจำลองข้อมูลตัวแปรอิสระไม่มีความสัมพันธ์เชิงเส้นพหุ ดังนั้นในการวิเคราะห์การถดถอย วิธีการประมาณค่าพารามิเตอร์ที่มีประสิทธิภาพจะส่งผลต่อการคัดเลือกตัวแปรที่เหมาะสมทั้งนี้ขึ้นอยู่กับปัจจัยต่างๆ ในแต่ละตัวแบบ เช่น ขนาดตัวอย่าง จำนวนตัวแปรอิสระ และระดับความสัมพันธ์ของตัวแปรอิสระ

ในงานวิจัยนี้ได้นำเสนอวิธีการคัดเลือกตัวแปรอิสระของตัวแบบการถดถอยไวบูลไม่ต่อเนื่องด้วยวิธีการถดถอยทีละขั้น วิธีบูตสแทร็ปทีละขั้น และวิธีแบบเบส์ ทั้งในกรณีที่มีความสัมพันธ์เชิงเส้นพหุ และไม่มีความสัมพันธ์เชิงเส้นพหุระหว่างตัวแปรอิสระ และจะพิจารณาตามลักษณะของ

ข้อมูลตัวแปรตามที่มีการกระจายต่ำกว่าเกณฑ์ และการกระจายเกินเกณฑ์ ในวิธีการถดถอยทีละขั้นและวิธีบูตสแทร็ปทีละขั้นจะใช้เกณฑ์ตัวสถิติที่เป็นฟังก์ชันของค่าดีเวียนซ์บนพื้นฐานของกระบวนการคัดเลือกตัวแปร สำหรับวิธีแบบเบส์จะคัดเลือกตัวแปรด้วยการสร้างช่วงความเชื่อมั่นของพารามิเตอร์ (Confidence Interval) เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพวิธีการคัดเลือกตัวแปรคืออัตราความสำเร็จ (Success Rate; SR) รวมถึงผู้วิจัยได้เปรียบเทียบประสิทธิภาพของการประมาณค่าพารามิเตอร์ร่วมด้วยโดยใช้ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean of Mean Square Error; MMSE)

2. วัสดุ อุปกรณ์และวิธีการวิจัย

2.1 การแจงแจงไวบูลไม่ต่อเนื่อง

Nakagawa และ Osaki [13] ได้นำเสนอการแจกแจงไวบูลไม่ต่อเนื่อง (Discrete Weibull Distribution) ซึ่งกำหนดให้ Y เป็นตัวแปรสุ่มที่มีการแจกแจงไวบูลไม่ต่อเนื่องที่มีพารามิเตอร์ q และ β ฟังก์ชันมวลความน่าจะเป็นแสดงดังสมการที่ (1) นี้

$$f(y)=\begin{cases} q^{y^{\beta}}-q^{(y+1)^{\beta}} & ; y=0,1,2,\ldots \\ 0 & ; y \text{ มีค่าอื่นๆ} \end{cases} \qquad (1)$$

เมื่อ $0<q<1$ และ $\beta>0$ เป็นพารามิเตอร์บ่งรูปร่าง (Shape Parameter) โดย q คือ ความน่าจะเป็นที่ Y จะมีค่ามากกว่าศูนย์ซึ่งจะส่งผลต่อจำนวนศูนย์ที่เกิดขึ้นและ β จะส่งผลต่อความเบ้ของการแจกแจง ถ้า $0<\beta\leq 1$ ที่ทุกๆ ค่า q ข้อมูลจะมีการกระจายเกินเกณฑ์ ถ้า $1<\beta<2$ ข้อมูลจะมีทั้งการกระจายเกินเกณฑ์ หรือต่ำกว่าเกณฑ์โดยขึ้นอยู่กับ q และถ้า $\beta\geq 2$ ที่ทุกๆ ค่า q ข้อมูลจะมีการกระจายต่ำกว่าเกณฑ์ [3]

2.2 ตัวแบบการถดถอยไวบูลไม่ต่อเนื่อง

การวิเคราะห์การถดถอยเป็นวิธีการทางสถิติที่ใช้ศึกษาความสัมพันธ์ระหว่างตัวแปรตาม (Response Variable) และตัวแปรอิสระ (Independent Variables) โดยมีจุดมุ่งหมายเพื่อที่จะอธิบายหรือทำนายตัวแปรตาม ในงานวิจัยนี้จะกล่าวถึงตัวแบบการถดถอยไวบูลไม่ต่อเนื่อง ซึ่งตัวแปรตาม Y มีการแจกแจงไวบูลไม่ต่อเนื่องและพารามิเตอร์ $q=q(\mathbf{x}_i)$ มีความสัมพันธ์กับตัวแปรอิสระ $\mathbf{x}_i$ ผ่านฟังก์ชันเชื่อมโยง การแจกแจงแบบมีเงื่อนไขของตัวแปรตามเมื่อทราบตัวแปรอิสระแสดงดังสมการที่ (2)

$$f\left(y_i \mid \mathbf{x}_i\right)=q(\mathbf{x}_i)^{y_i^{\beta}}-q(\mathbf{x}_i)^{(y_i+1)^{\beta}}\,;\, y_i=0,1,2,\ldots \qquad (2)$$

เมื่อ y_i คือ ตัวแปรตาม และ $\mathbf{x}_i=(1\ x_{i1}\ \cdots\ x_{ip})$ คือ เวกเตอร์ของตัวแปรอิสระที่ค่าสังเกตที่ i ขนาด $1\times(p+1)$, $i=1,2,,...,n$, p คือ จำนวนตัวแปรอิสระ และ n คือ ขนาดตัวอย่าง

กำหนดให้ $\boldsymbol{\alpha}=(\alpha_0\quad \alpha_1\quad \cdots\quad \alpha_p)'$ คือ เวกเตอร์ของสัมประสิทธิ์การถดถอยขนาด $(p+1)\times 1$

2.2.1 ตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงล็อก-ล็อก (Log-Log Link Function)

ตัวแบบผ่านฟังก์ชันเชื่อมโยงล็อก-ล็อกแสดงดังสมการที่ (3) และ (4)

$$\log\left(-\log\left(q(\mathbf{x}_i)\right)\right)=\mathbf{x}_i\boldsymbol{\alpha} \qquad (3)$$

$$q(\mathbf{x}_i)=e^{-e^{\mathbf{x}_i\boldsymbol{\alpha}}} \qquad (4)$$

ดังนั้นฟังก์ชันภาวะน่าจะเป็นแสดงดังสมการที่ (5)

$$L(\boldsymbol{\alpha},\beta \mid \mathbf{x}_1,\ldots,\mathbf{x}_n,y_1,\ldots,y_n) =\prod_{i=1}^{n}\left[(e^{-e^{\mathbf{x}_i\boldsymbol{\alpha}}})^{y_i^{\beta}}-(e^{-e^{\mathbf{x}_i\boldsymbol{\alpha}}})^{(y_i+1)^{\beta}}\right] \qquad (5)$$

2.2.2 ตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงลอจิต (Logit Link Function)

ตัวแบบผ่านฟังก์ชันเชื่อมโยงลอจิตแสดงดังสมการที่ (6) และ (7)

$$logit\left(q(\mathbf{x}_i)\right)=\log\left(\frac{q(\mathbf{x}_i)}{1-q(\mathbf{x}_i)}\right)=\mathbf{x}_i\boldsymbol{\alpha} \qquad (6)$$

$$q(\mathbf{x}_i) = \frac{e^{\mathbf{x}_i \boldsymbol{\alpha}}}{1+e^{\mathbf{x}_i \boldsymbol{\alpha}}} \quad (7)$$

ดังนั้นฟังก์ชันภาวะน่าจะเป็นแสดงดังสมการที่ (8)

$$L(\boldsymbol{\alpha}, \beta \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_n, y_1, \ldots, y_n) = \prod_{i=1}^{n} \left[\left(\frac{e^{\mathbf{x}_i \boldsymbol{\alpha}}}{1+e^{\mathbf{x}_i \boldsymbol{\alpha}}} \right)^{y_i^{\beta}} - \left(\frac{e^{\mathbf{x}_i \boldsymbol{\alpha}}}{1+e^{\mathbf{x}_i \boldsymbol{\alpha}}} \right)^{(y_i+1)^{\beta}} \right] \quad (8)$$

2.3 การคัดเลือกตัวแปร

2.3.1 วิธีการถดถอยทีละขั้น (Stepwise Regression Method)

ในหัวข้อนี้จะกล่าวถึงขั้นตอนของวิธีการถดถอยทีละขั้น [8] กำหนดให้ A เป็นเซตย่อยของตัวแปรอิสระที่อยู่ในตัวแบบ p คือ จำนวนพารามิเตอร์ของตัวแบบเซตย่อย A และ k คือ จำนวนพารามิเตอร์ของตัวแบบเต็ม (Full Model) ส่วน $\hat{\boldsymbol{\alpha}}_A$ และ $\hat{\boldsymbol{\alpha}}$ คือ ตัวประมาณสัมประสิทธิ์การถดถอยจากตัวแบบของเซตย่อย A และตัวแบบเต็มตามลำดับ $\hat{q}$ และ $\hat{\beta}$ คือ ตัวประมาณพารามิเตอร์ q และ β ตามลำดับ โดยทุกตัวประมาณคำนวณจากวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation; MLE) นอกจากนี้ กำหนดให้

$L(\hat{\boldsymbol{\alpha}}_A, \hat{\beta})$ คือ ฟังก์ชันภาวะน่าจะเป็นของตัวแบบภายใต้ตัวประมาณสัมประสิทธิ์การถดถอย $\hat{\boldsymbol{\alpha}}_A$ และ $\hat{\beta}$

$L(\hat{\boldsymbol{\alpha}}, \hat{\beta})$ คือ ฟังก์ชันภาวะน่าจะเป็นของตัวแบบภายใต้ตัวประมาณสัมประสิทธิ์การถดถอย $\hat{\boldsymbol{\alpha}}$ และ $\hat{\beta}$

$L(\hat{q}, \hat{\beta})$ คือ ฟังก์ชันภาวะน่าจะเป็นของตัวแบบภายใต้ตัวประมาณที่ถูกตัวประมาณด้วย n พารามิเตอร์ $q_1, q_2, \ldots, q_n$ และ β ซึ่งไม่เกี่ยวข้องกับตัวประมาณสัมประสิทธิ์การถดถอย โดยจะแทน $(\hat{q}_i, \hat{\beta})$ ด้วย $(\hat{q}, \hat{\beta})$

ตัวสถิติที่ใช้ประกอบการคัดเลือกตัวแปรอิสระแสดงดังสมการที่ (9)

$$D_p = D(\hat{\boldsymbol{\alpha}}_A, \hat{\beta}) - D(\hat{\boldsymbol{\alpha}}, \hat{\beta}) - (k-2p) \quad (9)$$

เมื่อ $D(\hat{\boldsymbol{\alpha}}_A, \hat{\beta})$ คือ ดีเวียนซ์ (Deviance) ของตัวแบบเซตย่อย A ซึ่งมีรูปแบบดังนี้

$$D(\hat{\boldsymbol{\alpha}}_A, \hat{\beta}) = -2\left(\ln L(\hat{\boldsymbol{\alpha}}_A, \hat{\beta}) - \ln L(\hat{q}, \hat{\beta})\right)$$

และ $D(\hat{\boldsymbol{\alpha}}, \hat{\beta})$ คือ ดีเวียนซ์ของตัวแบบเต็มซึ่งมีรูปแบบดังนี้

$$D(\hat{\boldsymbol{\alpha}}, \hat{\beta}) = -2\left(\ln L(\hat{\boldsymbol{\alpha}}, \hat{\beta}) - \ln L(\hat{q}, \hat{\beta})\right)$$

เกณฑ์การนำตัวแปรเข้าและนำตัวแปรออกเป็นดังนี้

1) เกณฑ์การนำตัวแปรเข้า (Addition Criterion)

ให้ $D(\hat{\boldsymbol{\alpha}}_p, \hat{\beta})$ คือ ค่าดีเวียนซ์ของตัวแบบเซตย่อยที่ประกอบด้วยตัวแปรอิสระ $p-1$ ตัว และ $D(\hat{\boldsymbol{\alpha}}_{p+1}, \hat{\beta})$ คือ ค่าดีเวียนซ์ที่น้อยที่สุดซึ่งคำนวณได้จากการเพิ่มตัวแปรหนึ่งเข้าสู่เซตย่อย คำนวณค่า R_1 แสดงดังสมการที่ (10)

$$R_1 = \frac{D(\hat{\boldsymbol{\alpha}}_p, \hat{\beta}) - D(\hat{\boldsymbol{\alpha}}_{p+1}, \hat{\beta})}{D(\hat{\boldsymbol{\alpha}}_{p+1}, \hat{\beta}) \big/ (n-p-2)} \quad (10)$$

เปรียบเทียบ R_1 กับ $F_{1,n-p-2,\alpha}$ ซึ่งก็คือควอนไทล์บนที่ α (The Upper α^{th} Quantile) ของการแจกแจงเอฟที่มีองศาเสรีเท่ากับ 1 และ $n-p-2$ หากค่า R_1 มากกว่า $F_{1,n-p-2,\alpha}$ แล้วตัวแปรดังกล่าวจะถูกนำเข้าสู่เซตย่อย

2) เกณฑ์การนำตัวแปรออก (Deletion Criterion)

ให้ $D(\hat{\boldsymbol{\alpha}}_p, \hat{\beta})$ คือ ค่าดีเวียนซ์ของตัวแบบเซตย่อยที่ประกอบด้วยตัวแปรอิสระ p ตัวที่ได้จากการนำตัวแปรเข้าก่อนหน้า และ $D(\hat{\boldsymbol{\alpha}}_{p-1}, \hat{\beta})$ คือ ค่าดีเวียนซ์ที่น้อยที่สุดซึ่งคำนวณได้หลังจากนำตัวแปรหนึ่งโดยตัดแต่ละตัวออกจากเซตย่อยจากขั้นตอนการนำตัวแปรเข้าก่อนหน้า คำนวณค่า R_2 แสดงดังสมการที่ (11)

$$R_2 = \frac{D(\hat{\boldsymbol{\alpha}}_{p-1}, \hat{\beta}) - D(\hat{\boldsymbol{\alpha}}_p, \hat{\beta})}{D(\hat{\boldsymbol{\alpha}}_p, \hat{\beta}) \big/ (n-p-2)} \quad (11)$$

เปรียบเทียบ R_2 กับ $F_{1,n-p-2,\alpha}$ หากค่า R_2 น้อยกว่า $F_{1,n-p-2,\alpha}$ แล้วตัวแปรดังกล่าวจะถูกนำออกจากตัวแบบเซตย่อย

วิธีการถดถอยทีละขั้นมีขั้นตอนการนำตัวแปรเข้า และการนำตัวแปรออกดังนี้

ขั้นตอนที่ 1 สร้างสมการถดถอย Y ด้วยตัวแปรอิสระ X_i ทุกตัว และคำนวณหาค่า $D(\hat{\boldsymbol{\alpha}}, \hat{\beta})$

ขั้นตอนที่ 2 สร้างสมการถดถอย Y ด้วยตัวแปรอิสระ X_i แต่ละตัว คำนวณหาค่าสถิติ D_p จากนั้นนำตัวแปรอิสระ X_i ที่มีค่าสถิติ D_p ต่ำที่สุดเข้าสู่ตัวแบบ ดังนั้น $A = \{X_i\}$

ขั้นตอนที่ 3 สร้างสมการถดถอย Y ด้วยตัวแปรอิสระ X_i ที่ถูกนำเข้าก่อนหน้ากับตัวแปรอิสระ X_j ที่เหลือทีละตัว คำนวณหาค่าสถิติ D_p แล้วเลือกเซตย่อยตัวแปรอิสระที่มีค่าสถิติ D_p ต่ำที่สุด หลังจากนั้นคำนวณเกณฑ์การนำเข้า R_1 หาก $R_1 > F_{1,\, n-p-2,\, \alpha}$ จะนำเข้าตัวแปรอิสระ X_j เข้าสู่เซตย่อย A ดังนั้น $A = \{X_i, X_j\}$

ขั้นตอนที่ 4 หากจำนวนตัวแปรอิสระในการคัดเลือกมีมากกว่า 2 ตัวแปร จะทำเหมือนในขั้นตอนที่ 3

ขั้นตอนที่ 5 สร้างสมการถดถอย Y ด้วยตัวแปรอิสระที่ตัดตัวแปรหนึ่งโดยตัดแต่ละตัวออกจากเซตย่อยของตัวแปรจากขั้นตอนการนำตัวแปรเข้าก่อนหน้า คำนวณหาค่าสถิติ D_p แล้วเลือกเซตย่อยของตัวแปรอิสระที่มีค่าสถิติ D_p ต่ำที่สุด หลังจากนั้นคำนวณเกณฑ์การนำออก R_2 หาก $R_2 < F_{1, n-p-2, \alpha}$ จะนำตัวแปรนั้นออกจากเซตย่อย

ขั้นตอนที่ 6 ทำซ้ำขั้นตอนที่ 4–5 จนกระทั่งไม่สามารถนำตัวแปรอิสระเข้าตัวแบบได้

2.3.2 วิธีบูตสแทร็ปทีละขั้น (Stepwise Bootstrap Method)

ในงานวิจัยนี้ประยุกต์ใช้แนวคิดของ Sauerbrei และคณะ [11] และ Ekman [12] ในการคัดเลือกตัวแปรด้วยวิธีบูตสแทร็ปโดยมีขั้นตอนดังนี้

ขั้นตอนที่ 1 สร้างตัวอย่างบูตสแทร็ปด้วยการสุ่มซ้ำค่าสังเกตจากข้อมูลตั้งต้นโดยทำการสุ่มแบบคืนที่ด้วยความน่าจะเป็น $\frac{1}{n}$ นั่นคือ $\underset{\sim}{z}_i^* = (y_i^*, \underset{\sim}{x}_i^*), i = 1, 2, \ldots, n$

ขั้นตอนที่ 2 ใช้วิธีการถดถอยทีละขั้นในการคัดเลือกตัวแปรจากตัวอย่างบูตสแทร็ป

ขั้นตอนที่ 3 สร้างเวกเตอร์เก็บตัวแปรอิสระที่ถูกคัดเลือกเข้าสู่ตัวแบบในแต่ละครั้ง

ขั้นตอนที่ 4 ทำซ้ำขั้นตอนที่ 1–3 เท่ากับจำนวนรอบบูตสแทร็ป

ขั้นตอนที่ 5 หากตัวแปรอิสระใดมีความถี่มากกว่าหรือเท่ากับ 80 เปอร์เซ็นต์ของจำนวนรอบบูตสแทร็ปจะนำตัวแปรอิสระดังกล่าวเข้าสู่ตัวแบบ

2.3.3 วิธีแบบเบส์ (Bayesian Method)

ตัวแบบการถดถอยไวบูลไม่ต่อเนื่องไม่มีการแจกแจงก่อนสังยุค (Conjugate Prior) ดังนั้นในงานวิจัยนี้จะพิจารณาการแจกแจงก่อนของพารามิเตอร์สัมประสิทธิ์การถดถอย $\boldsymbol{\alpha}$ ภายใต้ค่าพารามิเตอร์ที่เป็นไปได้ของ $\boldsymbol{\alpha}$ ซึ่งก็คือจำนวนจริง ดังนั้นจึงเลือกใช้การแจกแจงก่อน (Prior Distribution) ของ α_i เป็นการแจกแจงปรกติ (Normal Distribution) และการแจกแจงลาปลาซ (Laplace Distribution) ส่วนการแจกแจงก่อนของพารามิเตอร์บ่งรูปร่าง β คือ การแจกแจงเอกรูปที่ไม่ให้สารสนเทศ (Uniform Non-Informative Prior Distribution) รวมถึงพิจารณาภายใต้ข้อสมมติความเป็นอิสระต่อกันของพารามิเตอร์

การแจกแจงภายหลัง (Posterior Distribution) ของพารามิเตอร์ $\boldsymbol{\alpha}$ และ β แสดงดังสมการที่ (12)

$$p(\boldsymbol{\alpha}, \beta \mid \underset{\sim}{x}, \underset{\sim}{y}) \propto L(\boldsymbol{\alpha}, \beta \mid \underset{\sim}{x}, \underset{\sim}{y}) \times p(\boldsymbol{\alpha}) \times p(\beta) \quad (12)$$

ตัวประมาณแบบเบส์คำนวณได้จากค่าคาดหมายของพารามิเตอร์ และภายใต้การแจกแจงภายหลังตามขอบ (Marginal Posterior Distribution) ของแต่ละพารามิเตอร์ ซึ่งจากการแจกแจงภายหลังข้างต้นไม่สามารถจัดการแจกแจงภายหลังตามขอบของพารามิเตอร์ $\boldsymbol{\alpha}$ และ β ให้อยู่ในรูปแบบการแจกแจงใดๆ ได้ ดังนั้นในงานวิจัยนี้จึงประยุกต์ใช้กระบวนการเมโทโพลิส-แฮสติงส์แบบปรับตัว (Adaptive Metropolis-Hastings Procedure) เพื่อประมาณค่าพารามิเตอร์แบบเบส์ [6], [14], [15] โดยมีขั้นตอนดังนี้

ขั้นตอนที่ 1 กำหนดค่าเริ่มต้นโดยใช้ค่าประมาณพารามิเตอร์จากวิธี MLE

ขั้นตอนที่ 2 กำหนดการแจกแจงนำเสนอ (Proposal Distribution) $g(\cdot)$ คือ การแจกแจงปรกติหลายตัวแปรที่มีเวกเตอร์ค่าเฉลี่ยมีสมาชิกเท่ากับศูนย์และเมทริกซ์ความแปรปรวนร่วมเท่ากับเมทริกซ์ผกผันสารสนเทศของฟิชเชอร์ และกำหนดให้ $\pi = (\boldsymbol{\alpha}, \beta)$

ขั้นตอนที่ 3 เลือกตัวอย่างสุ่ม π_l จากการแจกแจงนำเสนอจากรอบมอลติคาร์โลที่ l

ขั้นตอนที่ 4 คำนวณค่าความน่าจะเป็นในการยอมรับ

$$r = \min\left(1, \frac{L(\pi_l \mid \underset{\sim}{x}, \underset{\sim}{y}) p(\pi_l) g(\pi_{l-1} \mid \pi_l)}{L(\pi_{l-1} \mid \underset{\sim}{x}, \underset{\sim}{y}) p(\pi_{l-1}) g(\pi_l \mid \pi_{l-1})}\right)$$

เมื่อ $L(\pi_l \mid \underset{\sim}{x}, \underset{\sim}{y})$ คือ ฟังก์ชันภาวะน่าจะเป็นของการแจกแจงไวบูลไม่ต่อเนื่อง และ $p(\cdot)$ คือ การแจกแจงก่อน

ขั้นตอนที่ 5 จะยอมรับตัวอย่างสุ่ม π_l จากการแจกแจงนำเสนอด้วยความน่าจะเป็นเท่ากับ r

ขั้นตอนที่ 6 ตามหลักการของเมโทโพลิส-แฮสติงส์แบบปรับตัวจะต้องอัพเดทความแปรปรวนร่วมของการแจกแจงนำเสนอจากการคำนวณความแปรปรวนร่วมตัวอย่างของโซ่มาร์คอฟ

ขั้นตอนที่ 7 ตัดค่าตัวอย่างสุ่มช่วงแรก (Burn-in Period) จำนวน L_0 ค่าออกจากการจำลองมอลติคาร์โลทั้งหมด L รอบเพื่อที่จะให้โซ่มาร์คอฟเข้าใกล้กระบวนการคงที่ก่อน (Convergence to Stationary) จากนั้นคำนวณตัวประมาณจากฐานนิยมของตัวประมาณทั้ง $L - L_0$ รอบ และช่วงความเชื่อมั่นสร้างจาก 95 เปอร์เซ็นต์ Highest Posterior Density (HPD) Interval [6]

2.4 การศึกษาโดยใช้การจำลอง

ในงานวิจัยนี้ใช้โปรแกรม R จำลองด้วยเทคนิคมอนติคาร์โลกระทำซ้ำ 1,000 รอบเพื่อสร้างตัวแบบการถดถอยโดยกำหนดสถานการณ์ต่างๆ เป็นดังนี้

1) กำหนดขนาดตัวอย่างที่ใช้ในการศึกษาเท่ากับ 100, 250, 400 และ 550

2) กรณีที่ตัวแปรอิสระไม่มีความสัมพันธ์กัน สร้างตัวแปรอิสระจำนวน 4 ตัวแปรจากการแจกแจงปรกติมาตรฐาน การแจกแจงเอกรูป การแจกแจงแบร์นูลลีและการแจกแจงเอกรูป ตามลำดับดังนี้ $X_1 \sim N(0,1)$, $X_2 \sim U(-1.5,1.5)$, $X_3 \sim Bernoulli(0.4)$ และ $X_4 \sim U(0,1)$

3) กรณีที่ตัวแปรอิสระมีความสัมพันธ์กัน สร้างตัวแปรอิสระ X_1, X_2 และ X_4 จากการแจกแจงปรกติมาตรฐานและการแจกแจงเอกรูป ตามลำดับดังนี้

$X_1 \sim N(0,1)$, $X_2 \sim U(-1.5,1.5)$ และ $X_4 \sim U(0,1)$ กำหนดตัวแปรอิสระ X_1 และ X_3 มีการแจกแจงปรกติมาตรฐานสองตัวแปรโดยมีสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร X_1 และ X_3 เท่ากับ 0.2, 0.5 และ 0.8

4) กำหนดเวกเตอร์สัมประสิทธิ์การถดถอย $\boldsymbol{\alpha}$ และพารามิเตอร์ β เพื่อจำลองตัวแปรตามผ่านฟังก์ชันเชื่อมโยงดังนี้

กรณีที่ 1 ข้อมูลมีการกระจายต่ำกว่าเกณฑ์และตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงล็อก-ล็อก
กำหนด $\boldsymbol{\alpha} = (-0.6,0.6,0.6,-0.6,0)$ และ $\beta = 3$

กรณีที่ 2 ข้อมูลมีการกระจายเกินเกณฑ์และตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงล็อก-ล็อก
กำหนด $\boldsymbol{\alpha} = (0.6,-0.6,0.6,-0.6,0)$ และ $\beta = 0.9$

กรณีที่ 3 ข้อมูลมีการกระจายต่ำกว่าเกณฑ์และตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงลอจิต
กำหนด $\boldsymbol{\alpha} = (0.8,0.8,0.8,0.8,0)$ และ $\beta = 3$

กรณีที่ 4 ข้อมูลมีการกระจายเกินเกณฑ์และตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงลอจิต
กำหนด $\boldsymbol{\alpha} = (0.6,-0.6,0.6,-0.6,0)$ และ $\beta = 0.9$

5) คัดเลือกตัวแปรสำหรับตัวแบบการถดถอยไวบูลไม่ต่อเนื่องด้วยวิธีการถดถอยทีละขั้น วิธีบูตสแทร็ปทีละขั้นและวิธีแบบเบส์ โดยวิธีบูตสแทร็ปทีละขั้นจะทำซ้ำจำนวน 50 ครั้ง และรอบการจำลองมอลติคาร์โลในวิธีแบบเบส์จะทำซ้ำ 25,000 รอบ และ Burn-in Period เท่ากับ 25 เปอร์เซ็นต์ของ 25,000 รอบการจำลองมอลติคาร์โลในวิธีแบบเบส์

6) เกณฑ์ที่ใช้เปรียบเทียบประสิทธิภาพ

6.1) อัตราความสำเร็จ (SR) คือ ค่าประสิทธิภาพการคัดเลือกตัวแปร ซึ่งก็คืออัตราของจำนวนครั้งที่การคัดเลือกตัวแปร พบตัวแปรอิสระตัวที่ j ที่ตัวประมาณสัมประสิทธิ์การถดถอย $\hat{\alpha}_j$, $j = 1,2,3$ ไม่เท่ากับศูนย์ตรงตามที่ได้กำหนดเวกเตอร์สัมประสิทธิ์การถดถอยซึ่งคำนวณได้จากสูตร $\sum_{i=1}^{m} I_i / m$

เมื่อ $I_i = \begin{cases} 1;\ \hat{\alpha}_j \neq 0,\ j = 1,2,3 \\ 0;\ \text{กรณีอื่นๆ} \end{cases}$

และ m คือ จำนวนรอบกระทำซ้ำมอนติคาร์โล

6.2) ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (MMSE) คือ ค่าที่ใช้วัดประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ซึ่งมีสูตรดังนี้

$$MMSE = \sum_{i=1}^{m} \left(\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta} \right)' \left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta} \right) \Big/ k \right) \Big/ m$$

เมื่อ $\hat{\boldsymbol{\theta}}_i$ คือ เวกเตอร์ค่าประมาณพารามิเตอร์รอบที่ i , $\boldsymbol{\theta}$ คือ เวกเตอร์พารามิเตอร์แท้จริงตามที่กำหนด m คือ จำนวนรอบกระทำซ้ำมอนติคาร์โล และ k คือ จำนวนพารามิเตอร์ของตัวแบบเต็ม

2.5 การประยุกต์ใช้กับข้อมูลจริง

ในงานวิจัยนี้ได้ประยุกต์ใช้ข้อมูลจำนวน 2 ชุด ได้แก่ ชุดที่ 1 คือ ข้อมูลจำนวนครั้งการใช้เครื่องพ่นยาโรคหืดระหว่างวันที่อยู่ในโรงเรียน Kunsberg School of National Jewish Health เมืองเดนเวอร์ รัฐโคโลราโด ประเทศสหรัฐอเมริกา สำหรับผู้ป่วยเด็กโรคหืดที่มีอายุระหว่าง 6–13 ปี ใน ค.ศ. 2002–2003 จำนวน 5,209 ค่าสังเกต [16] ซึ่งข้อมูลตัวแปรตาม คือ จำนวนครั้งการใช้เครื่องพ่นยาโรคหืด ตัวแปรอิสระคือการออกกำลังกายของผู้ป่วยเด็กในโรงเรียนซึ่งเป็นข้อมูลจำแนกประเภทคือมีการออกกำลังกายและไม่มีออกกำลังกาย (X_1) ร้อยละของค่าความชื้น (X_2) อุณหภูมิเฉลี่ย (X_3) และค่า PM2.5 ในช่วงเช้า (X_4) และชุดที่ 2 คือ ข้อมูลจำนวนครั้งการเข้าพบแพทย์ของผู้ป่วยประเทศสหรัฐอเมริกา ค.ศ. 1986 จำนวน 485 ราย ข้อมูลจาก Ecdat R package ชุดข้อมูล Doctor [17] ซึ่งข้อมูลตัวแปรตาม คือ จำนวนครั้งการเข้าพบแพทย์ของผู้ป่วย ตัวแปรอิสระ คือ จำนวนเด็กในครัวเรือน (X_1) ค่าการเข้าถึงการดูแลด้านสุขภาพ (X_2) และค่าสถานะทางสุขภาพ (X_3)

ผู้วิจัยจะทดสอบการแจกแจงไวบูลไม่ต่อเนื่องของข้อมูลตัวแปรตามด้วยสถิติทดสอบคอลโมโกรอฟ-สมีร์นอฟ (Kolmogorov-Smirnov Test) ที่ระดับนัยสำคัญ 0.05 โดยสมมติฐานว่างของการทดสอบ คือ ข้อมูลมีการแจกแจงไวบูลไม่ต่อเนื่อง คำนวณค่าเฉลี่ยและความแปรปรวน รวมถึงค่าประมาณพารามิเตอร์ β เพื่อพิจารณาลักษณะการกระจายของข้อมูล จากนั้นสร้างเมทริกซ์สหสัมพันธ์ของตัวแปรอิสระและทดสอบความสัมพันธ์ระหว่างตัวแปรอิสระด้วยสถิติทดสอบสหสัมพันธ์เพียร์สัน ที่ระดับนัยสำคัญ 0.05 รวมถึงสร้างตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงล็อก-ล็อกและลอจิต ทำการคัดเลือกตัวแปรและใช้เกณฑ์สารสนเทศของอะกะอิเกะ (AIC) และเกณฑ์สารสนเทศของเบส์ (BIC) คัดเลือกตัวแบบที่เหมาะสม [3], [6]

3. ผลการทดลอง

ในส่วนนี้ผู้วิจัยได้เสนอผลการวิจัยจากการจำลองในรูปแบบตารางดังตารางที่ 1–4 ตามลักษณะการกระจายข้อมูลและฟังก์ชันเชื่อมโยง โดยบรรทัดแรกแสดงค่า SR และบรรทัดที่สองแสดง $MMSE$ โดยผู้วิจัยได้ทำตัวเข้มสำหรับวิธีที่ให้ประสิทธิภาพมากที่สุด โดยค่า SR เข้าใกล้ค่าหนึ่งมากที่สุดแสดงว่าวิธีการคัดเลือกตัวแปรมีประสิทธิภาพมากที่สุด ส่วนค่า $MMSE$ ที่มีค่าน้อยที่สุดแสดงว่าวิธีการประมาณค่าพารามิเตอร์มีประสิทธิภาพมากที่สุดซึ่งจะใช้สัญลักษณ์ต่อไปนี้แทนความหมายต่างๆ ดังนี้ n แทนขนาดตัวอย่าง ρ แทนระดับความสัมพันธ์ SW แทนวิธีการถดถอยทีละขั้น BootSW แทนวิธีบูตสแทร็ปทีละขั้น BayesN แทนวิธีแบบเบส์ที่พารามิเตอร์สัมประสิทธิ์การถดถอยมีการแจกแจงก่อนเป็นการแจกแจงปรกติ และ BayesL แทนวิธีแบบเบส์ที่พารามิเตอร์สัมประสิทธิ์การถดถอยมีการแจกแจงก่อนเป็นการแจกแจงลาปลาซ นอกจากนั้นในตารางที่ 5–8 เป็นผลการวิจัยที่ได้ประยุกต์ใช้กับข้อมูล

3.1 ผลจากการจำลอง

ผลการจำลองเพื่อเปรียบเทียบประสิทธิภาพวิธีการคัดเลือกตัวแปรด้วยค่า SR พบว่า เมื่อขนาดตัวอย่างเท่ากับ 250, 400 และ 550 วิธี BootSW จะมีประสิทธิภาพมากที่สุดเกือบทุกกรณี ยกเว้นกรณีที่ขนาดตัวอย่างเท่ากับ 250 ตัวแปรอิสระมีความสัมพันธ์กันที่ระดับความสัมพันธ์เท่ากับ 0.8 เมื่อตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงล็อก-ล็อก ทั้งข้อมูลมีการกระจายต่ำกว่าเกณฑ์และเกินเกณฑ์ และข้อมูลมีการกระจายเกินเกณฑ์และตัวแบบการถดถอยผ่าน

ฟังก์ชันเชื่อมโยงลอจิตวิธี BayesL จะมีประสิทธิภาพมากที่สุดดังแสดงในตารางที่ 1, 2 และ 4 ตามลำดับ รวมถึงกรณีที่ตัวแปรอิสระไม่มีความสัมพันธ์กันในข้อมูลมีการกระจายเกินเกณฑ์และตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงลอจิตวิธี BayesN จะมีประสิทธิภาพมากที่สุดดังแสดงในตารางที่ 4

ส่วนที่ขนาดตัวอย่างเท่ากับ 100 ในตารางที่ 1 เมื่อสถานการณ์ที่กำหนดให้ข้อมูลมีการกระจายต่ำกว่าเกณฑ์และตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงล็อก-ล็อก ในกรณีที่ตัวแปรอิสระไม่มีความสัมพันธ์กัน วิธี SW จะมีประสิทธิภาพมากที่สุด ในขณะที่กรณีที่ตัวแปรอิสระมีความสัมพันธ์กัน ที่ระดับความสัมพันธ์ 0.2, 0.5 และ 0.8 วิธี BayesN จะมีประสิทธิภาพมากที่สุด ในตารางที่ 2 เป็นสถานการณ์ที่กำหนดให้ข้อมูลมีการกระจายเกินเกณฑ์และตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงล็อก-ล็อก ในกรณีที่ตัวแปรอิสระไม่มีความสัมพันธ์กันและตัวแปรอิสระมีความสัมพันธ์กันที่ระดับความสัมพันธ์ 0.8 และ วิธี SW จะมีประสิทธิภาพมากที่สุด ส่วนกรณีที่ตัวแปรอิสระมีความสัมพันธ์กันที่ระดับความสัมพันธ์ 0.2 และ 0.5 วิธี BayesL และ BayesN จะมีประสิทธิภาพมากที่สุดตามลำดับ ในตารางที่ 3 เป็นสถานการณ์ที่กำหนดให้ข้อมูลมีการกระจายต่ำกว่าเกณฑ์ และตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงลอจิต ในกรณีที่ตัวแปรอิสระไม่มีความสัมพันธ์กันและตัวแปรอิสระมีความสัมพันธ์กันที่ระดับความสัมพันธ์ 0.8 และ วิธี BayesN จะมีประสิทธิภาพมากที่สุด ส่วนกรณีที่ตัวแปรอิสระมีความสัมพันธ์กันที่ระดับความสัมพันธ์ 0.2 และ 0.5 วิธี BayesL จะมีประสิทธิภาพมากที่สุด และในตารางที่ 4 เป็นสถานการณ์ที่กำหนดให้ข้อมูลมีการกระจายเกินเกณฑ์และตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงลอจิต ในกรณีที่ตัวแปรอิสระไม่มีความสัมพันธ์กันและตัวแปรอิสระมีความสัมพันธ์กันที่ระดับความสัมพันธ์ 0.8 และวิธี SW จะมีประสิทธิภาพมากที่สุด ส่วนกรณีที่ตัวแปรอิสระมีความสัมพันธ์กันที่ระดับความสัมพันธ์ 0.2 และ 0.5 วิธี BayesL และ BayesN จะมีประสิทธิภาพมากที่สุดตามลำดับ

นอกจากนี้ผลการจำลองเพื่อเปรียบเทียบประสิทธิภาพการประมาณค่าพารามิเตอร์โดยพิจารณาจากค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (MMSE) เป็นดังนี้ เกือบทุกสถานการณ์วิธี BootSW มีประสิทธิภาพมากที่สุด ยกเว้นที่ขนาดตัวอย่างเท่ากับ 100 วิธี SW มีประสิทธิภาพมากที่สุดเมื่อที่ตัวแปรอิสระไม่มีความสัมพันธ์กันภายใต้ข้อมูลมีการกระจายต่ำกว่าเกณฑ์และตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงล็อก-ล็อก และข้อมูลมีการกระจายเกินเกณฑ์และตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงลอจิต รวมถึงวิธี BayesN มีประสิทธิภาพมากที่สุดเมื่อตัวแปรอิสระไม่มีความสัมพันธ์กันภายใต้ข้อมูลมีการกระจายต่ำกว่าเกณฑ์และตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงลอจิต รวมถึงที่ระดับความสัมพันธ์ 0.8 ภายใต้ข้อมูลมีการกระจายทั้งต่ำกว่าเกณฑ์และเกินเกณฑ์ที่ตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงล็อก-ล็อก นอกจากนั้นที่ขนาดตัวอย่างเท่ากับ 250 วิธี BayesL มีประสิทธิภาพมากที่สุดคือกรณีที่ตัวแปรอิสระไม่มีความสัมพันธ์กันภายใต้ข้อมูลมีการกระจายต่ำกว่าเกณฑ์และตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงล็อก-ล็อก

ตารางที่ 1 ค่า SR และ MMSE เมื่อข้อมูลมีการกระจายต่ำกว่าเกณฑ์และตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงล็อก-ล็อก

ρ	n	วิธีการ			
		SW	BootSW	BayesN	BayesL
0	100	**0.709**	0.481	0.691	0.651
		0.098000	0.10125	0.11864	0.11868
	250	0.825	**0.912**	0.906	0.911
		0.03188	**0.02707**	0.034077	0.03275
	400	0.850	**0.964**	0.922	0.943
		0.01828	**0.01501**	0.01948	0.01858
	550	0.851	**0.964**	0.912	0.927
		0.01285	**0.01040**	0.01372	0.01340
0.2	100	0.850	0.891	**0.933**	0.931
		0.09141	**0.076011**	0.11691	0.116971
	250	0.870	**0.966**	0.927	0.950
		0.02712	**0.02183**	0.03070	0.02860
	400	0.866	**0.974**	0.933	0.949
		0.01601	**0.01254**	0.01743	0.01569
	550	0.862	**0.980**	0.916	0.925
		0.01112	**0.00839**	0.01221	0.01155

ตารางที่ 1 ค่า SR และ MMSE เมื่อข้อมูลมีการกระจายต่ำกว่าเกณฑ์และตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงล็อก-ล็อก (ต่อ)

ρ	n	วิธีการ			
		SW	BootSW	BayesN	BayesL
0.5	100	0.817	0.780	**0.897**	0.890
		0.09863	**0.08765**	0.12241	0.11563
	250	0.850	**0.966**	0.921	0.937
		0.03043	**0.02380**	0.03298	0.03081
	400	0.846	**0.969**	0.920	0.931
		0.01775	**0.01394**	0.02008	0.01777
	550	0.832	**0.966**	0.898	0.912
		0.01214	**0.00905**	0.01300	0.01252
0.8	100	0.382	0.108	**0.692**	0.607
		0.15056	0.16017	**0.14111**	0.14348
	250	0.715	0.739	0.915	**0.926**
		0.05038	0.05283	0.03608	**0.03468**
	400	0.845	**0.959**	0.937	0.943
		0.02067	**0.01675**	0.01902	0.01882
	550	0.825	**0.975**	0.927	0.938
		0.01444	**0.01118**	0.01456	0.01481

ตารางที่ 2 ค่า SR และ MMSE เมื่อข้อมูลมีการกระจายเกินเกณฑ์และตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงล็อก-ล็อก

ρ	n	วิธีการ			
		SW	BootSW	BayesN	BayesL
0	100	**0.711**	0.537	0.673	0.618
		0.07127	**0.07095**	0.07540	0.07901
	250	0.859	**0.935**	0.934	0.930
		0.01978	**0.01617**	0.01939	0.01937
	400	0.851	**0.965**	0.935	0.947
		0.01121	**0.00833**	0.01042	0.01027
	550	0.835	**0.971**	0.914	0.936
		0.00864	**0.00589**	0.00858	0.00754
0.2	100	0.816	0.861	0.906	**0.910**
		0.07500	**0.05502**	0.06520	0.06591
	250	0.844	**0.976**	0.934	0.956
		0.01972	**0.01270**	0.01871	0.01682
	400	0.843	**0.975**	0.930	0.948
		0.01190	**0.00751**	0.01089	0.01004

ตารางที่ 2 ค่า SR และ MMSE เมื่อข้อมูลมีการกระจายเกินเกณฑ์และตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงล็อก-ล็อก (ต่อ)

ρ	n	วิธีการ			
		SW	BootSW	BayesN	BayesL
0.2 (ต่อ)	550	0.856	**0.963**	0.920	0.918
		0.00808	**0.00560**	0.00792	0.00830
0.5	100	0.776	0.728	**0.863**	0.850
		0.08146	**0.06553**	0.06995	0.07175
	250	0.824	**0.951**	0.923	0.941
		0.02270	**0.01544**	0.01956	0.01813
	400	0.826	**0.955**	0.893	0.919
		0.01292	**0.00856**	0.01281	0.01128
	550	0.838	**0.957**	0.899	0.910
		0.00857	**0.00581**	0.00892	0.00845
0.8	100	**0.646**	0.330	0.624	0.577
		0.09483	0.11144	**0.08601**	0.09083
	250	0.833	0.904	0.899	**0.922**
		0.02486	**0.02069**	0.02414	0.02256
	400	0.819	**0.944**	0.912	0.926
		0.01392	**0.00987**	0.01293	0.01261
	550	0.796	**0.931**	0.867	0.900
		0.01031	**0.00698**	0.01096	0.00890

ตารางที่ 3 ค่า SR และ MMSE เมื่อข้อมูลมีการกระจายต่ำกว่าเกณฑ์และตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงลอจิต

ρ	n	วิธีการ			
		SW	BootSW	BayesN	BayesL
0	100	0.778	0.668	**0.817**	0.786
		0.11555	0.11403	**0.11043**	0.12332
	250	0.857	**0.955**	0.935	0.949
		0.03482	**0.02919**	0.03400	0.03361
	400	0.839	**0.971**	0.901	0.924
		0.02144	**0.01625**	0.02109	0.02014
	550	0.848	**0.977**	0.905	0.931
		0.01464	**0.01117**	0.01696	0.01474
0.2	100	0.848	0.941	0.944	**0.960**
		0.10166	**0.07586**	0.09562	0.10297
	250	0.844	**0.966**	0.931	0.946
		0.03481	**0.02604**	0.03210	0.03182

ตารางที่ 3 ค่า SR และ MMSE เมื่อข้อมูลมีการกระจายต่ำกว่าเกณฑ์และตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงลอจิต (ต่อ)

ρ	n	วิธีการ SW	BootSW	BayesN	BayesL
0.2 (ต่อ)	400	0.845	**0.970**	0.915	0.941
		0.01992	**0.01456**	0.01967	0.01786
	550	0.861	**0.974**	0.915	0.924
		0.01387	**0.01038**	0.01431	0.01385
0.5	100	0.829	0.899	0.927	**0.933**
		0.11165	**0.08913**	0.10227	0.11034
	250	0.854	**0.972**	0.937	0.957
		0.03521	**0.02682**	0.03362	0.03287
	400	0.854	**0.971**	0.927	0.943
		0.01978	**0.01496**	0.02024	0.01828
	550	0.812	**0.966**	0.901	0.925
		0.01537	**0.01087**	0.01443	0.01392
0.8	100	0.754	0.570	**0.827**	0.765
		0.14012	0.15724	**0.12339**	0.14154
	250	0.859	**0.963**	0.939	0.955
		0.03983	**0.03244**	0.03683	0.03582
	400	0.859	**0.971**	0.924	0.942
		0.02259	**0.01760**	0.02223	0.02110
	550	0.846	**0.967**	0.926	0.934
		0.01642	**0.01299**	0.01669	0.01686

ตารางที่ 4 ค่า SR และ MMSE เมื่อข้อมูลมีการกระจายเกินเกณฑ์และตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงลอจิต

ρ	n	วิธีการ SW	BootSW	BayesN	BayesL
0	100	**0.636**	0.392	0.526	0.434
		0.08057	0.08324	0.08679	0.09450
	250	0.839	0.863	**0.885**	0.880
		0.02418	**0.02311**	0.02783	0.02967
	400	0.856	**0.959**	0.933	0.935
		0.01407	**0.01074**	0.01366	0.01392
	550	0.857	**0.974**	0.930	0.931
		0.01010	**0.00728**	0.00969	0.01047

ตารางที่ 4 ค่า SR และ MMSE เมื่อข้อมูลมีการกระจายเกินเกณฑ์และตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงลอจิต (ต่อ)

ρ	n	วิธีการ SW	BootSW	BayesN	BayesL
0.2	100	0.833	0.861	0.911	**0.913**
		0.05126	**0.03675**	0.05258	0.05493
	250	0.834	**0.963**	0.920	0.953
		0.01948	**0.01252**	0.02093	0.01882
	400	0.856	**0.967**	0.936	0.943
		0.01067	**0.00720**	0.00988	0.00999
	550	0.842	**0.972**	0.923	0.931
		0.00773	**0.00484**	0.00790	0.00755
0.5	100	0.821	0.791	**0.887**	0.879
		0.05770	**0.04729**	0.05835	0.06045
	250	0.840	**0.958**	0.931	0.942
		0.02014	**0.01373**	0.02040	0.02009
	400	0.846	**0.961**	0.919	0.946
		0.01183	**0.00823**	0.01196	0.01113
	550	0.832	**0.969**	0.910	0.924
		0.00870	**0.00552**	0.00860	0.00822
0.8	100	**0.661**	0.324	0.605	0.539
		0.08025	0.10310	**0.08070**	0.08689
	250	0.851	0.935	0.929	**0.941**
		0.02186	**0.01783**	0.02370	0.02299
	400	0.864	**0.976**	0.938	0.940
		0.01241	**0.00882**	0.01209	0.01290
	550	0.829	**0.960**	0.917	0.930
		0.01046	**0.00753**	0.01012	0.00968

ตารางที่ 5 เมทริกซ์สหสัมพันธ์และ p-value ข้อมูลชุดที่ 1

	X_1	X_2	X_3	X_4
X_1	1			
X_2	–0.0315 (0.0229)	1		
X_3	0.0026 (0.8500)	–0.5801 (<0.0000)	1	
X_4	–0.0306 (0.0273)	0.1908 (<0.0000)	–0.1034 (<0.0000)	1

3.2 ผลจากการประยุกต์ใช้ข้อมูลจริง

3.2.1 จากข้อมูลชุดที่ 1 ข้อมูลจำนวนครั้งการใช้เครื่องพ่นยาโรคหืด จากการทดสอบการแจกแจงด้วยสถิติทดสอบคอลโมโกรอฟ-สมีร์นอฟพบว่า ค่าสถิติ D = 0.02227 และค่า p-value = 0.151 สรุปได้ว่าข้อมูลมีการแจกแจงไวบูลไม่ต่อเนื่อง มีค่าเฉลี่ยเท่ากับ 1.2705 ค่าความแปรปรวนเท่ากับ 0.8433 และค่าประมาณพารามิเตอร์ β จากทั้ง 4 วิธีได้ค่าประมาณ 2.3 ซึ่งมีค่ามากกว่า 2 จึงสามารถสรุปได้ว่าเป็นข้อมูลที่มีการกระจายต่ำกว่าเกณฑ์ [3] เมทริกซ์สหสัมพันธ์ระหว่างตัวแปรอิสระทั้ง 4 ตัวแปร และจากสถิติทดสอบสหสัมพันธ์เพียร์สัน จากตารางที่ 5 พบว่าตัวแปรอิสระ X_1 และ X_3 ไม่มีความสัมพันธ์กัน ส่วนตัวแปรอิสระคู่อื่นนั้นมีความสัมพันธ์กัน ที่ระดับนัยสำคัญ 0.05 จากผลการคัดเลือกตัวแปรแสดงในรูปแบบตารางที่ 6 พบว่า ทั้งสองฟังก์ชันเชื่อมโยงวิธี SW จะคัดเลือกตัวแปรอิสระได้ทั้ง 4 ตัวแปร ในขณะที่วิธี BootSW BayesN และBayesL จะคัดเลือกตัวแปรได้ 3 ตัวแปร คือ ตัวแปรการออกกำลังกายที่โรงเรียนของเด็กนักเรียน ร้อยละค่าความชื้นและอุณหภูมิเฉลี่ย เมื่อพิจารณาที่ค่า AIC พบว่า วิธี SW และ BootSW ให้ค่าที่ใกล้เคียงกัน และต่ำกว่าทั้ง 2 วิธีแบบเบส์ ในขณะที่ค่า BIC วิธี BootSW จะให้ค่าที่ต่ำที่สุด

3.2.2 จากข้อมูลชุดที่ 2 ข้อมูลจำนวนครั้งการเข้าพบแพทย์ จากการทดสอบการแจกแจงด้วยสถิติทดสอบคอลโมโกรอฟ-สมีร์นอฟพบว่า ค่าสถิติ D = 0.02886 และค่า p-value = 0.9876 สรุปได้ว่าข้อมูลมีการแจกแจงไวบูลไม่ต่อเนื่อง มีค่าเฉลี่ยเท่ากับ 1.6103 และค่าความแปรปรวนเท่ากับ 11.2011 และค่าประมาณพารามิเตอร์ β จากทั้ง 4 วิธี ได้ค่าประมาณ 0.78 จึงสามารถสรุปได้ว่าเป็นข้อมูลที่มีการกระจายเกินเกณฑ์ [3] เมทริกซ์สหสัมพันธ์ระหว่างตัวแปรอิสระทั้ง 3 ตัวแปรและจากสถิติทดสอบสหสัมพันธ์เพียร์สัน จากตารางที่ 7 พบว่า ตัวแปรอิสระไม่มีความสัมพันธ์กันที่ระดับนัยสำคัญ 0.05 จากผลการคัดเลือกตัวแปรแสดงในรูปแบบตารางที่ 8 พบว่า ทั้ง 4 วิธีคัดเลือกตัวแปรอิสระได้ตัวแปรอิสระ 2 ตัวแปร เช่นเดียวกันนั่นคือจำนวนเด็กในครัวเรือนและค่าสถานะทางสุขภาพ เมื่อพิจารณาที่ค่า AIC และ BIC พบว่า วิธี SW และ BootSW ให้ค่าที่ต่ำที่สุด

ตารางที่ 6 ค่าประมาณพารามิเตอร์ AIC และ BIC ชุดที่ 1

Method	SW	BootSW	BayesN	BayesL
Log–log Link				
Intercept	–2.0981	–2.1297	–2.0851	2.0465
X_1	1.3843	1.3846	1.3808	1.3804
X_2	0.3169	0.2856	0.2833	0.2895
X_3	0.4710	0.4672	0.4155	0.4221
X_4	–0.0259			
$\hat{\beta}$	2.3035	2.3027	2.3095	2.3040
AIC	12,522.56	12,523.27	12,524.58	12,523.83
BIC	12,561.91	12,556.06	12,563.93	12,563.17
Logit Link				
Intercept	2.0673	2.1011	2.0487	2.0071
X_1	–1.6516	–1.6523	–1.6459	–1.6442
X_2	–0.3618	–0.3303	–0.3239	–0.3352
X_3	–0.5506	–0.5472	–0.4585	–0.4196
X_4	0.0270			
$\hat{\beta}$	2.3033	2.3025	2.3017	2.3060
AIC	12,522.36	12,522.71	12,532.46	12,524.08
BIC	12,561.71	12,555.50	12,571.81	12,563.43

ตารางที่ 7 เมทริกซ์สหสัมพันธ์และ p-value ข้อมูลชุดที่ 2

	X_1	X_2	X_3
X_1	1		
X_2	–0.0162 (0.7226)	1	
X_3	0.0040 (0.9306)	–0.0638 (0.1608)	1

ตารางที่ 8 ค่าประมาณพารามิเตอร์ AIC และ BIC ชุดที่ 2

Method	SW	BootSW	BayesN	BayesL
Log–log Link				
Intercept	–0.5977	–0.5977	–0.4819	–0.5019
X_1	0.1047	0.1047	0.1072	0.0933
X_3	–0.2174	–0.2174	–0.2122	–0.2098
$\hat{\beta}$	0.7810	0.7810	0.7801	0.7780

ตารางที่ 8 ค่าประมาณพารามิเตอร์ AIC และ BIC ชุดที่ 2 (ต่อ)

Method	SW	BootSW	BayesN	BayesL
AIC	1,575.02	1,575.02	1,576.06	1,576.11
BIC	**1,591.75**	**1,591.75**	1,596.98	1,597.03
Logit Link				
Intercept	0.2984	0.2984		
X_1	–0.1427	–0.1427	–0.1563	–0.1384
X_3	0.2772	0.2772	0.2885	0.2833
$\hat{\beta}$	0.7801	0.7801	0.7759	0.7768
AIC	1,576.63	1,576.63	1,579.07	1,577.19
BIC	1,593.37	1,593.37	1,599.99	1,598.11

4. อภิปรายผลและสรุป

ในงานวิจัยนี้ได้ศึกษา และเปรียบเทียบประสิทธิภาพวิธีการคัดเลือกตัวแปรของตัวแบบการถดถอยไวบูลไม่ต่อเนื่องผ่านฟังก์ชันเชื่อมโยงล็อก-ล็อก และฟังก์ชันเชื่อมโยงลอจิตด้วยวิธีการถดถอยทีละขั้น วิธีบูตสแทร็ปทีละขั้น และวิธีแบบเบส์ภายใต้การจำลองข้อมูลทั้งสถานการณ์ที่ข้อมูลตัวแปรตามมีลักษณะการกระจายต่ำกว่าเกณฑ์ และการกระจายเกินเกณฑ์ ตัวแปรอิสระเท่ากับ 4 ตัวแปรทั้งกรณีที่มีความสัมพันธ์เชิงเส้นพหุ และไม่มีความสัมพันธ์เชิงเส้นพหุพบว่า เมื่อขนาดตัวอย่างเท่ากับ 100 วิธีการประมาณแบบเบส์ทั้งสองการแจกแจงก่อนจะมีประสิทธิภาพมากที่สุดเกือบทุกกรณี แต่จะมีบางกรณีที่วิธีการถดถอยทีละขั้นให้ประสิทธิภาพดีกว่า เมื่อขนาดตัวอย่างเท่ากับ 250, 400 และ 550 วิธีบูตสแทร็ปทีละขั้นจะมีประสิทธิภาพมากที่สุดเกือบทุกกรณี ซึ่งจะมีบางกรณีที่วิธีแบบเบส์ให้ประสิทธิภาพดีกว่า ส่วนการเปรียบเทียบประสิทธิภาพการประมาณค่าพารามิเตอร์ วิธีบูตสแทร็ปทีละขั้นมีประสิทธิภาพมากที่สุดเกือบทุกกรณียกเว้นกรณีที่ขนาดตัวอย่างเท่ากับ 100 และ 250 ซึ่งจะมีบางกรณีที่วิธีการถดถอยทีละขั้นและวิธีการประมาณค่าแบบเบส์ให้ประสิทธิภาพดีกว่า

ผลการวิจัยข้างต้นสรุปได้ว่าการสร้างตัวแบบการถดถอยไวบูลไม่ต่อเนื่องเมื่อขนาดตัวอย่างเท่ากับ 250, 400 และ 550 วิธีบูตสแทร็ปทีละขั้นเป็นวิธีที่มีประสิทธิภาพในการคัดเลือกตัวแปร และประมาณค่าพารามิเตอร์มากที่สุด ซึ่งเป็นวิธีที่สอดคล้องกับงานวิจัยของ Ekman [12] แต่หากขนาดตัวอย่างเท่ากับ 100 วิธีการคัดเลือกตัวแปรและประมาณค่าพารามิเตอร์ที่มีประสิทธิภาพมากที่สุดอาจจะต้องพิจารณาจากระดับความสัมพันธ์ของตัวแปรอิสระ และลักษณะการกระจายข้อมูลของตัวแปรตาม นอกจากนั้นเมื่อประยุกต์ใช้กับข้อมูลจริงพบว่า มีความสอดคล้องกับสถานการณ์จากการจำลอง คือ วิธีบูตสแทร็ปทีละขั้นมีประสิทธิภาพมากที่สุด ในงานวิจัยนี้ได้ศึกษาเชิงจำลองเพียงจำนวนตัวแปรอิสระเท่ากับ 4 ตัวแปร และขนาดตัวอย่างเท่ากับ 100, 250, 400 และ 550 ซึ่งในการสร้างตัวแบบที่มีจำนวนตัวแปรอิสระที่มากกว่านี้ จากขนาดตัวอย่างที่กล่าวมาอาจจะส่งผลต่อประสิทธิภาพการประมาณค่าพารามิเตอร์และการคัดเลือกตัวแปร ดังนั้นอาจจะต้องเพิ่มขนาดตัวอย่างนอกเหนือจากการศึกษานี้ รวมถึงเราอาจจะทดสอบการกระจายข้อมูลด้วยสถิติทดสอบเพื่อสนับสนุนผลสรุปเกี่ยวกับลักษณะการกระจายข้อมูลตัวแปรตามนอกเหนือจากงานวิจัยนี้

การคัดเลือกตัวแปรด้วยวิธีแบบเบส์จะมีประสิทธิภาพในกรณีที่ขนาดตัวอย่างเท่ากับ 100 และตัวแปรอิสระมีความสัมพันธ์กัน ยิ่งไปกว่านั้นเมื่อขนาดตัวอย่างเพิ่มขึ้นวิธีแบบเบส์ให้อัตราความสำเร็จที่สูง แสดงให้เห็นว่าเมื่อขนาดตัวอย่างมากขึ้นวิธีแบบเบส์ยังคงมีประสิทธิภาพ นอกจากนั้นการคำนวณหาค่าประมาณ และการหาช่วงความเชื่อมั่นของพารามิเตอร์สามารถทำได้รวดเร็วกว่าวิธีบูตสแทร็ปทีละขั้น นอกจากนี้ในการประยุกต์ใช้กับข้อมูลจริงที่ขนาดตัวอย่างน้อยกว่า 100 อาจจะเกิดความคลาดเคลื่อนสูงในการประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุดที่จะส่งผลต่อการคัดเลือกตัวแปรในวิธีการถดถอยทีละขั้นและวิธีบูตสแทร็ปทีละขั้น ในขณะที่วิธีแบบเบส์นั้นสามารถหาค่าประมาณได้ สำหรับการคัดเลือกตัวแปรด้วยวิธีการถดถอยทีละขั้นอาจจะมีประสิทธิภาพมากขึ้นเมื่อขนาดตัวอย่างมากเพียงพอ

เอกสารอ้างอิง

[1] A. C. Cameron and P. K. Trivedi, *Regression*

Analysis of Count Data. Cambridge University Press, Cambridge: 2013.

[2] K. F. Sellers and G. Shmueli, "A flexible regression model for count data," *The Annals of Applied Statistics*, vol. 4, no. 2, pp. 943–961, 2010.

[3] H. S. Klakattawi, "Discrete Weibull regression model for count data," Ph.D dissertation, Department of Mathematics and Computing College of Engineering, Design and Physical Sciences, Brunel University, London, UK, 2017.

[4] H. S. Klakattawi, V. Vinciotti, and K. Yu, "A simple and adaptive dispersion regression model for count data," *Entropy*, vol. 20, no. 2, pp. 142, 2018.

[5] H. Yoo, "Application of discrete Weibull regression model with multiple Imputation," *Communications for Statistical Applications and Methods*, vol. 26, no. 3, pp. 325–336, 2019.

[6] H. Haselimashhadi, V. Vinciotti, and K. Yu, "A novel Bayesian regression model for counts with an application to health data," *Journal of Applied Statistics*, vol. 45, no. 6, pp. 1085–1105, 2018.

[7] O. S. Adesina, A. S. Onanaye, and D. M. Okewole, "Bayesian Optimization for parameter of Discrete Weibull Regression," *Journal of Advances in Mathematics and Computer Science*, vol. 34, no. 6, pp. 1–13, 2020.

[8] D. M. Sakate, D. N. Kashid, and D. T. Shirhe, "Subset selection in poisson regression," *Journal of Statistical Theory and Practice*, vol. 5, no. 2, pp. 207–219, 2011.

[9] B. Efron, "Bootstrap methods: another look at the Jackknife," *The Annals of Statistics*, vol. 7, no. 1, pp. 1–26, 1979.

[10] N. Sudjai and M. Duangsaphon, "Liu-type logistic regression coefficient estimation with multicollinearity problem by using the bootstrapping method," *Science, Engineering and Health Studies*, vol. 14, no. 3, pp. 203–214, 2020 (in Thai).

[11] W. Saurerbrei and M. Schumacher, "A bootstrap resampling procedure for model building: Application to the cox regression model," *Statistics in Medicine*, vol. 11, no. 16, pp. 2093–2109, 1992.

[12] A. Ekman, "Variable Selection for the Cox proportional hazards model: A simulation study comparing the stepwise, lasso and bootstrap approach" M.S. thesis, Department of Mathematics and Mathematical Statistics, Umea University, 2011.

[13] T. Nakagawa and S. Osaki, "The discrete Weibull distribution," *IEEE Transaction on Reliability*, vol. 24, no. 5, pp. 300–301, 1975.

[14] W. K. Hastings, "Monte Carlo sampling methods using markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[15] H. Haario, E. Saksman, and J. Tamminen, "An adaptive Metropolis algorithm," *Bernoulli*, vol. 7, no. 2, pp. 223–242, 2001.

[16] G. K. Grunwald, S. L. Bruce, L. Jiang, M. Strand, and N. Rabinovitch, "A statistical model for under- or overdispersed clustered and longitudinal count data," *Biometrical Journal*, vol. 53, no. 4, pp. 578–594, 2011.

[17] S. Gurmu, "Semiparametric estimation of hurdle regression models with an application to medicaid utilization," *Journal of Applied Econometrics*, vol. 12, no, 3, pp. 225–242, 1997.