



การเปรียบเทียบวิธีบูตสเตรปในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์ การถดถอยเชิงเส้นที่มีมิติสูง

A Comparison of Bootstrap Methods in the Interval Estimation of High-dimensional Regression Coefficient

ภูวกร นิธิสนธิกุล* และ วิฐรา พึ่งพาพงศ์

Phuwakorn Nithisnatheekun* and Vitara Pungpapong

สถิติศาสตรมหาบัณฑิต สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย กรุงเทพฯ ประเทศไทย

The Degree of Master of Science Program in Statistics Department of Statistics Faculty of Commerce and Accountancy

Chulalongkorn University, Bangkok, Thailand

**Corresponding author, E-mail: n.phuwakorn@gmail.com*

บทคัดย่อ

งานวิจัยฉบับนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบช่วงความเชื่อมั่นสำหรับค่าสัมประสิทธิ์การถดถอย โดยแนวทางบูตสเตรปที่แตกต่างกัน (1) วิธีสุ่มตัวแปรตามและตัวแปรอิสระ (2) วิธีสุ่มส่วนเหลือ และ (3) วิธีสุ่มค่าถ่วงน้ำหนัก ผู้วิจัยได้จำลองชุดข้อมูลขนาดมิติสูงขึ้น และ นำมาวิเคราะห์เปรียบเทียบด้วยวิธีบูตสเตรปที่แตกต่างกัน 3 วิธี โดยการวัดค่าเฉลี่ยเปอร์เซ็นต์ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง ค่าเฉลี่ยอัตราผลบวกเทียม และค่าเฉลี่ยอัตราผลลบเทียม ระหว่าง 1,000 ข้อมูล การวิเคราะห์ของเราพบว่าบูตสเตรปที่ใช้สุ่มตัวแปรตามและตัวแปรอิสระดีที่สุด ในแง่ของทั้งค่าเฉลี่ยเปอร์เซ็นต์ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริงและค่าเฉลี่ยอัตราผลบวกเทียม และบูตสเตรปที่ใช้สุ่มค่าถ่วงน้ำหนักดีที่สุด ในแง่ของค่าเฉลี่ยอัตราผลลบเทียม

คำสำคัญ: บูตสเตรป ข้อมูลมิติสูง การวิเคราะห์การถดถอย ช่วงความเชื่อมั่น แลชโซ แลชโซปรับได้

Abstract

The objective of this research was to study and compare confidence intervals for regression coefficients using three different bootstrap approaches: (1) random dependent variables and independent variables, (2) random residual, and (3) random weight. The high-dimensional datasets were simulated, and the performance of the three methods were compared. The researchers figured out an average confidence interval in percentage covering the true value of



regression coefficient, average false positive rates, and average false negative rates among 1,000 datasets. The research found that the bootstrap using random dependent variables and independent variables was the best method in terms of confidence interval percentages covering the true value of regression coefficient and average false positive rates. The bootstrap using random weight was the best method in terms of average false negative rates.

Keywords: *Bootstrap, High-Dimensional, Regression, Coefficients Interval, Lasso, Adaptive Lasso*

1. บทนำ

การวิเคราะห์การถดถอย (จิราวัลย์ จิตรถเวช, 2558) เป็นเครื่องมือที่รู้จักอย่างแพร่หลายและง่ายต่อการใช้งาน เพราะสามารถใช้กับข้อมูลที่หลากหลาย (สายชล สันสมบูรณ์ทอง, 2553) มีทฤษฎีทางสถิติอนุมานอย่างชัดเจน ทั้งการประมาณค่าด้วยวิธีกำลังสองน้อยสุด การประมาณช่วงความเชื่อมั่น การทดสอบสมมติฐานต่าง ๆ อย่างไรก็ตาม (จิราวัลย์ จิตรถเวช, 2558) การวิเคราะห์การถดถอยยังมีข้อจำกัดสำหรับข้อมูลที่มีมิติสูง หรือข้อมูลที่จำนวนตัวแปรอิสระมากกว่าจำนวนตัวแปรตาม ทำให้ไม่สามารถหาค่าประมาณสัมประสิทธิ์การถดถอยจากวิธีกำลังสองน้อยสุดได้ Tibshirani (1996) ได้เสนอวิธีแลซโซ ซึ่งสามารถใช้ได้กับข้อมูลที่มีมิติสูง ลดปัญหาตัวแปรอิสระที่มีความสัมพันธ์กันสูง รวมทั้งยังสามารถประมาณค่าและทำการคัดเลือกตัวแปรได้พร้อม ๆ กัน โดยตัวแปรที่ไม่ได้คัดเลือกจะมีค่าสัมประสิทธิ์การถดถอยเท่ากับศูนย์ แต่ปัญหาของวิธีแลซโซคือ แลซโซสามารถเลือกตัวแปรเข้าได้มากที่สุดจำนวน n ตัวเท่านั้น ดังนั้น หากข้อมูลมีจำนวนตัวแปรอิสระมากกว่าขนาดของตัวอย่างเป็นจำนวนมาก ๆ ตัวแปรที่ได้จากวิธีแลซโซอาจจะไม่มีความเหมาะสม และในกรณีที่ตัวแปรอิสระที่มีความสัมพันธ์กันสูง วิธีแลซโซมีแนวโน้มที่จะเลือกตัวแปรเพียงตัวเดียว จากกลุ่มตัวแปรอิสระที่มีความสัมพันธ์กันสูงเข้าตัวแบบ โดยไม่สนใจว่าจะเป็นตัวแปรใดซึ่งเกิดปัญหาการคัดเลือกตัวแปรเข้าสู่ตัวแบบ ที่ไม่มีความคงเส้นคงวา Zou (2006) จึงได้นำเสนอแลซโซปรับได้ซึ่งช่วยแก้ปัญหาในการคัดเลือกตัวแปรเข้าสู่ตัวแบบ ที่ไม่มีความคงเส้นคงวา ดังกล่าว เนื่องจากตัวประมาณแลซโซมีการปรับค่าของตัวประมาณเท่ากันทุกตัว โดยถ่วงน้ำหนักของค่าฟังก์ชันเพนัลตี้ (penalty function)

แม้ว่าวิธีแลซโซและแลซโซปรับได้สามารถแก้ไขข้อจำกัดของการวิเคราะห์การถดถอยในการวิเคราะห์ข้อมูลที่มีมิติสูงได้ อย่างไรก็ตาม วิธีการดังกล่าวให้ผลลัพธ์เป็นการประมาณค่าแบบจุดของสัมประสิทธิ์การถดถอยเท่านั้น เรามักจะให้ความสนใจในการตรวจสอบตัวแปรอิสระแต่ละตัวว่ามีความสัมพันธ์กับตัวแปรตามหรือไม่ ดังนั้นผู้วิจัยจึงมีความสนใจในการตรวจสอบความสัมพันธ์ โดยใช้การสร้างช่วงความเชื่อมั่นของสัมประสิทธิ์การถดถอยและทำการเปรียบเทียบช่วงความเชื่อมั่นจากวิธีบูตสตรอปในแบบต่าง ๆ ซึ่งวิธีบูตสตรอปสามารถทำได้หลายวิธี ผู้วิจัยจึงได้นำเสนอ วิธีสุ่มตัวแปรตามและตัวแปรอิสระจะเป็นการสุ่มจากกลุ่มตัวอย่างแบบใส่กลับคืน ซึ่งจะทำให้การสุ่มแต่ละครั้งจะมีโอกาสได้ข้อมูลซ้ำ ส่งผลให้ข้อมูลบางตัวอาจจะไม่ได้ถูกสุ่มขึ้นมา วิธีสุ่มส่วนเหลือให้ผลการสุ่มเช่นเดียวกับวิธีสุ่มตัวแปรตามและตัวแปรอิสระ แต่มีโอกาสน้อยกว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระ จากปัญหาดังกล่าววิธีสุ่มค่าถ่วงน้ำหนักจะช่วยแก้ไขปัญหที่เกิดขึ้น Meeker, Hahn and Escobar (2017) ได้นำเสนอการสร้างช่วงความ



เชื่อมั่นของสัมประสิทธิ์การถดถอยโดย Simple Percentile Bootstrap , Nonparametric Bias-corrected and Accelerated (BCa) Percentile Bootstrap และ Nonparametric Bias-corrected (BC) Percentile Bootstrap จึงสนใจที่จะศึกษาเป็นงานวิจัย

2. วัตถุประสงค์

เพื่อเปรียบเทียบเปรียบเทียบประสิทธิภาพของวิธีบูตสเตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มตัวแปรตามและตัวแปรอิสระ วิธีสุ่มส่วนเหลือ และวิธีสุ่มตัวค่าถ่วงน้ำหนัก

3. วิธีดำเนินงานวิจัย

3.1 สร้างข้อมูลที่มีอัตราส่วนระหว่างขนาดตัวอย่างและจำนวนตัวแปรอิสระ $n : p$ ที่ 100 : 500 จากสมการดังนี้

$$Y = X\beta + \varepsilon$$

โดยที่ $X_i \sim N(0, I_p)$, $\varepsilon \sim N(0, I_n)$ และ

$$\beta = \left[0.3, 3, -1, -0.65, 3, -3, 0.7, -1, 3, -0.6, 2, 1, -1, -3, -2, 5, 2, 4, 4, 4, \underbrace{0, 0, 0, \dots, 0}_{480 \text{ ตัว}} \right]^T$$

3.2 วิธีสุ่มบูตสเตรปแบบไม่ใช้พารามิเตอร์กระทำซ้ำ 1,000 ครั้ง

3.2.1 การสุ่มตัวแปรตามและตัวแปรอิสระ

- 1) สุ่มตัวแปรตาม (y_i) และตัวแปรอิสระ (x_i) จากกลุ่มตัวอย่าง จะได้ y_i^* และ x_i^* โดยมีขนาดเท่ากับกลุ่มตัวอย่าง
- 2) นำ y_i^* และ x_i^* ประมาณค่าสัมประสิทธิ์ความถดถอย (β)
- 3) ใช้ตัวอย่างของข้อมูลเดิมทำซ้ำตามขั้นตอน 1-2 จำนวน B ครั้ง
- 4) ได้ค่าสัมประสิทธิ์การถดถอยในแต่ละครั้งดังนี้

$$\beta^* = [\beta^{(1)}, \beta^{(2)}, \beta^{(3)}, \dots, \beta^{(b)}]$$

เมื่อ $\beta^{(b)}$ คือสัมประสิทธิ์การถดถอยจากตัวอย่างบูตสเตรปครั้งที่ b

3.2.2 การสุ่มส่วนเหลือ

- 1) จากกลุ่มตัวอย่างตัวแปรตาม (y_i) และตัวแปรอิสระ (x_i) สร้างสมการถดถอยเชิงเส้น
- 2) นำ x_i แทนลงในสมการถดถอยเชิงเส้นที่ได้จากข้อ 1 จะได้ค่าทำนาย ($\hat{y}$)
- 3) นำค่าสังเกต (y_i) และค่าทำนาย ($\hat{y}_i$) คำนวณส่วนเหลือ (e_i) จากสมการ (3.2.1)

$$e_i = y_i - \hat{y}_i \quad (3.2.1)$$

- 4) นำส่วนเหลือ (e_i) ทำการสุ่มได้ e_i^*



5) นำ e_i ย้อนกลับไปแทนลงในสมการที่ (3.2.2) จะได้ y_i^*

$$y_i^* = \hat{y}_i + e_i \quad (3.2.2)$$

6) นำ y_i^* และ x_i ประมาณค่าสัมประสิทธิ์ความถดถอย (β)

7) นำตัวอย่างของข้อมูลเดิมทำซ้ำตามขั้นตอน 2-6 จำนวน B ครั้ง

8) ได้ค่าสัมประสิทธิ์การถดถอยในแต่ละครั้งดังนี้

$$\beta^* = [\beta^{(1)}, \beta^{(2)}, \beta^{(3)}, \dots, \beta^{(b)}]$$

เมื่อ $\beta^{(b)}$ คือสัมประสิทธิ์การถดถอยจากตัวอย่างบูตสเตรปครั้งที่ b

3.2.3 การสุ่มค่าถ่วงน้ำหนัก

1) สุ่มค่าถ่วงน้ำหนักจากการแจกแจงแบบเอกซ์โปเนนเชียล

นั่นคือ $w_i \sim \text{Exp}(1)$

จะได้ $w_1, w_2, \dots, w_n$

2) นำ y_i, x_i และน้ำหนักที่สุ่มได้ประมาณค่าสัมประสิทธิ์ความถดถอย (β) โดยใช้ Weighted Regression

3) นำตัวอย่างของข้อมูลเดิมทำซ้ำตามขั้นตอน 1-2 จำนวน B ครั้ง

4) ได้ค่าสัมประสิทธิ์การถดถอยในแต่ละครั้งดังนี้

$$\beta^* = [\beta^{(1)}, \beta^{(2)}, \beta^{(3)}, \dots, \beta^{(b)}]$$

เมื่อ $\beta^{(b)}$ คือสัมประสิทธิ์การถดถอยจากตัวอย่างบูตสเตรปครั้งที่ b

3.3 การประมาณค่าพารามิเตอร์

3.3.1 การประมาณค่าโดยกำลังสองน้อยที่สุด

3.3.2 การประมาณค่าโดยแลชโซ

3.3.3 การประมาณค่าโดยแลชโซปรับได้

3.4 วิธีสร้างช่วงความเชื่อมั่นที่ 95%

จากที่ทำการบูตสเตรปเป็นจำนวน B ครั้ง สามารถหาช่วงความเชื่อมั่นของสัมประสิทธิ์การถดถอยเชิงเส้น โดยขอบเขตล่างสัมประสิทธิ์การถดถอยเชิงเส้น (β_j^*) และขอบเขตบนสัมประสิทธิ์การถดถอยเชิงเส้น ($\bar{\beta}_j^*$) ดังนี้

3.4.1 Simple Percentile Bootstrap

สามารถหาได้จากสมการ (3.4.1)

$$[\underline{\beta}_j^*, \bar{\beta}_j^*] = [\beta_j^*_{(\frac{\alpha}{2})}, \beta_j^*_{(1-\frac{\alpha}{2})}] \quad (3.4.1)$$

เมื่อ β_q^* คือ เปอร์เซนต์ไทล์ที่ q ของ β ที่เกิดจากการบูตสเตรปเป็นจำนวน B ครั้งสำหรับ สัมประสิทธิ์การถดถอย (β)



3.4.2 Nonparametric Bias-corrected and Accelerated (BCa) Percentile

สามารถหาได้จากสมการ (3.4.2)

$$[\underline{\beta}_j^*, \overline{\beta}_j^*] = [\beta_{\alpha_1}^*, \beta_{\alpha_2}^*] \quad (3.4.2)$$

เมื่อ β_q^* คือเปอร์เซ็นต์ไทล์ที่ q ของ β ที่เกิดจากการบูตสเตรปเป็นจำนวน B ครั้งสำหรับ
สัมประสิทธิ์การถดถอย (β)

α_1 หาได้จากสมการ (3.4.3)

$$\alpha_1 = \Phi_{\text{norm}} \left[z_b + \frac{z_b - z_{(1-\frac{\alpha}{2})}}{1 - \hat{a} \left[z_b - z_{(1-\frac{\alpha}{2})} \right]} \right] \quad (3.4.3)$$

α_2 หาได้จากสมการ (3.4.4)

$$\alpha_2 = \Phi_{\text{norm}} \left[z_b + \frac{z_b + z_{(1-\frac{\alpha}{2})}}{1 - \hat{a} \left[z_b + z_{(1-\frac{\alpha}{2})} \right]} \right] \quad (3.4.4)$$

เมื่อ z_q คือเปอร์เซ็นต์ไทล์ที่ q ของการแจกแจงปกติมาตรฐาน

b คือสัดส่วนของจำนวนครั้งของการสุ่มบูตสเตรปของ β^* ที่มีค่าน้อยกว่า β ต่อจำนวน
ทั้งหมด

$\hat{a}$ คือความเร่งคงตัวหาได้จากสมการ (3.4.5)

$$\hat{a} = \frac{\sum_{i=1}^n (\beta_{[i]} - \beta^{(b)})^3}{6 \left[\sum_{i=1}^n (\beta_{[i]} - \beta^{(b)})^2 \right]^{\frac{3}{2}}} \quad (3.4.5)$$

$$\text{เมื่อ } \beta_{[i]} = \sum_{j=1}^n \frac{\beta_{[ij]}}{n} \quad (3.4.6)$$

เมื่อ $\beta^{(b)}$ คือสัมประสิทธิ์การถดถอยจากตัวอย่างบูตสเตรปครั้งที่ b

3.4.3 Nonparametric Bias-corrected (BC) Percentile

สามารถหาได้จากสมการ (3.4.7) หลักการคำนวณเดียวกันกับ 3.4.2 โดย $\hat{a} = 0$

$$[\underline{\beta}_j^*, \overline{\beta}_j^*] = [\beta_{\alpha_1}^*, \beta_{\alpha_2}^*] \quad (3.4.7)$$

เมื่อ β_q^* คือเปอร์เซ็นต์ไทล์ที่ q ของ β ที่เกิดจากการบูตสเตรปเป็นจำนวน B ครั้งสำหรับ
สัมประสิทธิ์การถดถอย (β)

α_1 หาได้จากสมการ (3.4.8)

$$\alpha_1 = \Phi_{\text{norm}} \left[2z_b - z_{(1-\frac{\alpha}{2})} \right] \quad (3.4.8)$$

α_2 หาได้จากสมการ (3.4.9)

$$\alpha_2 = \Phi_{\text{norm}} \left[2z_b + z_{(1-\frac{\alpha}{2})} \right] \quad (3.4.9)$$

เมื่อ z_q คือเปอร์เซ็นต์ไทล์ที่ q ของการแจกแจงปกติมาตรฐาน

b คือสัดส่วนของจำนวนครั้งของการสุ่มบูตสเตรปของ β^* ที่มีค่าน้อยกว่า β ต่อจำนวน
ทั้งหมด



3.5 การเปรียบเทียบผลการวิเคราะห์

3.5.1 ค่าเฉลี่ยเปอร์เซ็นต์ในช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง

1) พิจารณช่วงความเชื่อมั่นสำหรับแต่ละ β_j ที่คำนวณได้จากการประมาณค่าโดยวิธีกลุ่มต่าง ๆ ที่สามารถครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง

2) จะได้จำนวนครั้งที่ช่วงความเชื่อมั่นสำหรับแต่ละ β_j ที่คำนวณได้จากการประมาณค่าโดยวิธีกลุ่มต่าง ๆ ที่สามารถครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง

3) นำจำนวนครั้งที่หาได้จากข้อ 2) คำนวณหาเปอร์เซ็นต์ในช่วงความเชื่อมั่นสำหรับแต่ละ β_j ที่คำนวณได้จากการประมาณค่าโดยวิธีกลุ่มต่าง ๆ ที่สามารถครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง

4) นำเปอร์เซ็นต์ จากข้อ 3) สำหรับแต่ละ β_j คำนวณหาค่าเฉลี่ยเปอร์เซ็นต์ในช่วงความเชื่อมั่นสำหรับแต่ละ β_j ที่คำนวณได้จากการประมาณค่าโดยวิธีกลุ่มต่าง ๆ ที่สามารถครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง

3.5.2 ค่าเฉลี่ยอัตราผลบวกเทียม

$$H_0: \beta_j = 0 \text{ vs } H_a: \beta_j \neq 0$$

สำหรับค่าเฉลี่ยของอัตราผลบวกเทียมคำนวณโดยขั้นตอนต่อไปนี้ดังนี้

1) จำนวนจำนวนที่ $\beta_j = 0$ และ $0 \notin CI$

2) จำนวนจำนวนที่ $0 \notin CI$

3) นำค่าที่ได้หาอัตราส่วนระหว่างข้อที่ 1) และ 2) หาอัตราส่วนระหว่าง

4) นำอัตราส่วนที่ได้จากข้อ 3) คำนวณค่าเฉลี่ย

3.5.3 ค่าเฉลี่ยอัตราผลลบเทียม

$$H_0: \beta_j = 0 \text{ vs } H_a: \beta_j \neq 0$$

สำหรับค่าเฉลี่ยอัตราผลลบเทียมคำนวณโดยขั้นตอนต่อไปนี้ดังนี้

1) จำนวนจำนวนที่ $\beta_j \neq 0$ และ $0 \in CI$

2) จำนวนจำนวนที่ $0 \in CI$

3) นำค่าที่ได้หาอัตราส่วนระหว่างข้อที่ 1) และ 2)

4) นำอัตราส่วนที่ได้จากข้อ 3) คำนวณค่าเฉลี่ย

3.6 สรุปผลวิจัยในแต่ละสถานการณ์

4.ผลการวิจัยและวิเคราะห์

ผลการวิจัยจะนำเสนอในรูปแบบตารางดังนั้นเพื่อสะดวกและง่ายต่อการอภิปราย จะใช้สัญลักษณ์ต่อไปนี้แทนความหมายต่าง ๆ ดังนี้

random	หมายถึง วิธีกลุ่ม
method of CI	หมายถึง วิธีการสร้างช่วงความเชื่อมั่น
RXY	หมายถึง วิธีกลุ่มตัวแปรตามและตัวแปรอิสระ



RR	หมายถึง วิธีสุ่มส่วนเหลือ
RW	หมายถึง วิธีสุ่มค่าถ่วงน้ำหนัก
SPB	หมายถึง วิธี Simple Percentile Bootstrap
BCaB	หมายถึง วิธี Nonparametric Bias-corrected and Accelerated (BCa) Percentile Bootstrap
BCB	หมายถึง วิธี Nonparametric Bias-corrected (BC) Percentile Bootstrap
ALS	หมายถึง ประมาณค่าโดยแลซโซปรับได้ (Adaptive Lasso)
MEAN	หมายถึง ค่าเฉลี่ย
SD	หมายถึง ส่วนเบี่ยงเบนมาตรฐาน

การแสดงตัวเลขในตารางเลขตัวหนาในตารางแสดงถึงวิธีสุ่มที่ดีที่สุด

1) การเปรียบเทียบค่าเฉลี่ยเปอร์เซ็นต์ในช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง

สำหรับการเปรียบเทียบค่าเฉลี่ยเปอร์เซ็นต์ในช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริงถ้ามีค่าเข้าใกล้ระดับความเชื่อมั่นที่ 95% วิธีสุ่มนั้นจะมีประสิทธิภาพดีที่สุด ผลการวิเคราะห์ดังตารางต่อไปนี้

ตารางที่ 1 ตารางการเปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของเปอร์เซ็นต์ในช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง

random	method of CI	LS		ALS	
		MEAN	SD	MEAN	SD
RXY	SPB	95.336	7.7135	95.450	0.6735
RR		93.090	7.8351	95.806	7.8782
RW		93.246	7.7636	94.986	7.7549
RXY	BCaB	88.704	6.9260	88.748	6.9959
RR		87.368	7.3610	91.478	7.2575
RW		86.210	7.2234	88.494	7.1519
RXY	BCB	88.702	6.9254	88.744	6.9935
RR		87.338	7.3577	91.478	7.2575
RW		86.162	7.2192	88.488	7.1480

จากตารางที่ 1 ผลการวิเคราะห์เป็นดังนี้

ในการสร้างช่วงความเชื่อมั่นด้วยวิธี Simple Percentile Bootstrap ร่วมกับการประมาณค่าด้วยแลซโซ พบว่าวิธีบูตสเตรปที่มีประสิทธิภาพดีที่สุด คือ วิธีสุ่มตัวแปรตามและตัวแปรอิสระ และการประมาณค่าโดยแลซโซปรับได้พบว่าวิธีบูตสเตรปที่มีประสิทธิภาพดีที่สุด คือ วิธีสุ่มค่าถ่วงน้ำหนัก

ในการสร้างช่วงความเชื่อมั่นด้วยวิธี Nonparametric Bias-corrected and Accelerated (BCa) Percentile Bootstrap ร่วมกับการประมาณค่าด้วยแลซโซ พบว่าวิธีบูตสเตรปที่มีประสิทธิภาพดีที่สุด คือ วิธีสุ่มตัวแปรตามและตัว



แปรอิสระมี และการประมาณค่าโดยเลขโซปรับได้ พบว่า วิธีบูตสเตรปที่มีประสิทธิภาพดีที่สุด คือ วิธีสุ่มส่วนเหลือ ซึ่งผลการวิเคราะห์ได้ผลเดียวกันกับการสร้างช่วงความเชื่อมั่น โดย วิธี Nonparametric Bias-corrected (BC) Percentile Bootstrap

2) การเปรียบเทียบค่าเฉลี่ยผลบวกเทียม (False positive rate)

สำหรับการเปรียบเทียบค่าเฉลี่ยผลบวกเทียม (False positive rate) ถ้ามีค่าเข้าใกล้ 0 วิธีสุ่มนั้นจะมีประสิทธิภาพที่ดีที่สุด ผลการวิเคราะห์ดังตารางต่อไปนี้

ตารางที่ 2 ตารางการเปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานอัตราผลบวกเทียม

random	method of CI	LS		ALS	
		MEAN	SD	MEAN	SD
RXY	SPB	0.0051	0.0507	0.0050	0.0285
RR		0.1098	0.0947	0.0000	0.0000
RW		0.0716	0.0667	0.0070	0.0362
RXY	BCaB	0.8345	0.0455	0.7706	0.0367
RR		0.6634	0.0468	0.8450	0.0527
RW		0.6942	0.0530	0.7994	0.0454
RXY	BCB	0.8348	0.0453	0.7706	0.0366
RR		0.6644	0.0469	0.8450	0.0527
RW		0.6960	0.0523	0.7995	0.0454

จากตารางที่ 2 ผลการวิเคราะห์เป็นดังนี้

การคำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานอัตราผลบวกเทียมสำหรับข้อมูลที่มีมิติสูงมีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็กพบว่า การสร้างช่วงความเชื่อมั่นโดยวิธี Simple Percentile Bootstrap โดยการประมาณค่าโดยเลขโซสำหรับวิธีสุ่มตัวแปรตามและตัวแปรอิสระขนาด 100 ตัวอย่าง ผลปรากฏว่า มีบางตัวอย่างที่ทำให้ค่าอัตราผลบวกเทียมไม่สามารถคำนวณค่าได้โดยมีอัตราผลบวกเทียมเป็นศูนย์ส่วนศูนย์ ทางผู้วิจัยจะทำการตัดตัวอย่างนั้นทิ้งไปและคำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานที่ค่าไม่เป็น ศูนย์ส่วนศูนย์เท่านั้น และยังพบอีกว่าการสร้างช่วงความเชื่อมั่นโดยวิธี Simple Percentile Bootstrap โดยการประมาณค่าโดยเลขโซปรับได้สำหรับวิธีสุ่มส่วนเหลือเกิดปัญหาเช่นเดียวกันกับการสร้างช่วงความเชื่อมั่นโดยวิธี Simple Percentile Bootstrap โดยการประมาณค่าโดยเลขโซสำหรับวิธีสุ่มตัวแปรตามและตัวแปรอิสระ

ในการสร้างช่วงความเชื่อมั่นด้วยวิธี Simple Percentile Bootstrap ร่วมกับการประมาณค่าด้วยเลขโซ พบว่า วิธีบูตสเตรปที่มีประสิทธิภาพดีที่สุด คือ วิธีสุ่มตัวแปรตามและตัวแปรอิสระ และการประมาณค่าโดยเลขโซปรับได้ พบว่า วิธีบูตสเตรปที่มีประสิทธิภาพดีที่สุด คือ วิธีสุ่มส่วนเหลือ



ในการสร้างช่วงความเชื่อมั่นด้วยวิธี Nonparametric Bias-corrected and Accelerated (BCa) Percentile ร่วมกับการประมาณค่าด้วยเลขโซ พบว่า วิธีบูตสเตรปที่มีประสิทธิภาพดีที่สุด คือ วิธีสุ่มส่วนเหลือ และการประมาณค่าด้วยเลขโซปรับได้ พบว่า วิธีบูตสเตรปที่มีประสิทธิภาพดีที่สุด คือ วิธีสุ่มตัวแปรตามและตัวแปรอิสระ ซึ่งผลการวิเคราะห์ได้ผลเดียวกันกับการสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected (BC) Percentile Bootstrap

3) การเปรียบเทียบค่าเฉลี่ยอัตราผลลบเทียม (False negative rate)

สำหรับการเปรียบเทียบค่าเฉลี่ยอัตราผลลบเทียม (False negative rate) ถ้ามีค่าเข้าใกล้ 0 วิธีสุ่มนั้นจะมีประสิทธิภาพที่ดีที่สุด ผลการวิเคราะห์ดังตารางต่อไปนี้

ตารางที่ 3 ตารางการเปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานอัตราผลลบเทียม

random	method of CI	LS		ALS	
		MEAN	SD	MEAN	SD
RXY	SPB	0.0323	0.0037	0.0329	0.0024
RR		0.0136	0.0055	0.0374	0.0027
RW		0.0134	0.0048	0.0290	0.0031
RXY	BCaB	0.0291	0.0040	0.0240	0.0029
RR		0.0106	0.0056	0.0341	0.0026
RW		0.0098	0.0058	0.0261	0.0038
RXY	BCB	0.0292	0.0040	0.0240	0.0029
RR		0.0106	0.0056	0.0341	0.0026
RW		0.0099	0.0058	0.0261	0.0038

จากตารางที่ 3 ผลการวิเคราะห์เป็นดังนี้

ในการสร้างช่วงความเชื่อมั่นด้วยวิธี Simple Percentile Bootstrap ร่วมกับการประมาณค่าด้วยเลขโซและเลขโซปรับได้ พบว่า วิธีบูตสเตรปที่มีประสิทธิภาพดีที่สุด คือ วิธีสุ่มค่าถ่วงน้ำหนัก

ในการสร้างช่วงความเชื่อมั่นด้วยวิธี Nonparametric Bias-corrected and Accelerated (BCa) Percentile ร่วมกับการประมาณค่าด้วยเลขโซ พบว่า วิธีบูตสเตรปที่มีประสิทธิภาพดีที่สุด คือ วิธีสุ่มค่าถ่วงน้ำหนัก และการประมาณค่าด้วยเลขโซปรับได้ พบว่า วิธีบูตสเตรปที่มีประสิทธิภาพดีที่สุด คือ วิธีสุ่มตัวแปรตามและตัวแปรอิสระ ซึ่งผลการวิเคราะห์ได้ผลเดียวกันกับการสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected (BC) Percentile Bootstrap



5. สรุปผลการศึกษา

สำหรับสรุปผลการวิจัยจะทำการสรุปผลโดยแบ่งตามวิธีการเปรียบเทียบซึ่งได้แก่ ค่าเฉลี่ยเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง ค่าเฉลี่ยความกว้างช่วงความเชื่อมั่น ค่าเฉลี่ยอัตราผลบวกเทียม และค่าเฉลี่ยอัตราผลลบเทียมผลปรากฏว่า

5.1 การเปรียบเทียบโดยค่าเฉลี่ยเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง

โดยรวมแล้ววิธีนุตสเตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพสูงที่สุด วิธีสุ่มส่วนเหลือรองลงมา และวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพต่ำที่สุด

5.2 การเปรียบเทียบค่าเฉลี่ยอัตราผลบวกเทียม (False positive rate)

โดยรวมแล้ววิธีนุตสเตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพสูงที่สุด วิธีสุ่มส่วนเหลือรองลงมา และวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพต่ำที่สุด

5.3 การเปรียบเทียบค่าเฉลี่ยอัตราผลลบเทียม (False negative rate)

โดยรวมแล้วไม่สามารถสรุปได้ว่าวิธีนุตสเตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพสูงที่สุด วิธีสุ่มตัวแปรตามและตัวแปรอิสระรองลงมา และวิธีสุ่มส่วนเหลือมีประสิทธิภาพต่ำที่สุด

6. กิตติกรรมประกาศ

วิจัยฉบับนี้ สำเร็จลุล่วงไปได้ด้วยดี เพราะได้รับความอนุเคราะห์ และช่วยเหลือจากผู้มีพระคุณหลายท่านด้วยกัน ทางผู้วิจัยจึงใคร่ขอขอบคุณในความช่วยเหลือต่าง ๆ ดังต่อไปนี้ ขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร. วิฐรา พึ่งพาพงศ์ ที่กรุณารับเป็นอาจารย์ที่ปรึกษาและคอยให้คำปรึกษาในเรื่องต่าง ๆ รวมถึงสละเวลาคอยติดตามความก้าวหน้า เสนอแนะ และชี้ให้เห็นปัญหาและข้อผิดพลาดต่าง ๆ จนทำให้วิจัยฉบับนี้สำเร็จลุล่วงอย่างสมบูรณ์และมีประสิทธิภาพ สามารถนำไปเผยแพร่ และนำเสนอได้ และขอกราบขอบพระคุณ คุณพ่อ คุณแม่ และครอบครัวที่คอยสนับสนุน เป็นกำลังใจ และติดตามความก้าวหน้าในการทำวิจัยมาโดยตลอด และขอขอบคุณเพื่อน ๆ พี่ ๆ ทุกคน ที่คอยให้กำลังใจ ให้ความช่วยเหลือ คำปรึกษา และข้อเสนอแนะต่าง ๆ ในการทำวิจัยฉบับนี้

7. เอกสารอ้างอิง

จิราวัลย์ จิตรถเวช (2558). *การวิเคราะห์การถดถอย*. กรุงเทพฯ:โครงการส่งเสริมและพัฒนาเอกสารวิชาการ สถาบันบัณฑิตพัฒนบริหารศาสตร์.



วิฐรา พึ่งพาพงศ์. (2558). บทวิเคราะห์วิธีวิเคราะห์การถดถอยเชิงเส้นสำหรับข้อมูลที่มีมิติสูง. *วารสารวิทยาศาสตร์และเทคโนโลยี*, 23(2), 211-223.

สายชล สันสมบูรณ์ทอง. (2553). *การวิเคราะห์สถิติ*. กรุงเทพฯ: สาขาวิชาสถิติประยุกต์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

Meeker, W, Hahn, G and Escobar, L. (2017). *Statistical Intervals A Guide for Practitioners and Researchers*. (2nd ed).

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476), 1418-1429.