# การเปรียบเทียบวิธีบูตสแตรปในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นที่มีมิติสูง

นายภูวกร นิธิศนทีกุล



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2563

A COMPARISON OF BOOTSTRAP METHODS IN INTERVAL ESTIMATION OF HIGH-DIMENSIONAL REGRESSION COEFFICIENTS.

Mr. Phuwakorn Nithisnatheekun



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Statistics
Department of Statistics
FACULTY OF COMMERCE AND ACCOUNTANCY
Chulalongkorn University
Academic Year 2020

| | |
|---|---|
| หัวข้อวิทยานิพนธ์ | การเปรียบเทียบวิธีบูตสแตรปในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นที่มีมิติสูง |
| โดย | นายภูวกร นิธิศนทีกุล |
| สาขาวิชา | สถิติ |
| อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก | ผู้ช่วยศาสตราจารย์ ดร.วิฐรา พึ่งพาพงศ์ |

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

.................................................................. คณบดีคณะพาณิชยศาสตร์และการบัญชี

(รองศาสตราจารย์ ดร.วิเลิศ ภูริวัชร)

คณะกรรมการสอบวิทยานิพนธ์

.................................................................. ประธานกรรมการ

(รองศาสตราจารย์ ดร.เสกสรร เกียรติสุไพบูลย์)

.................................................................. อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ผู้ช่วยศาสตราจารย์ ดร.วิฐรา พึ่งพาพงศ์)

.................................................................. กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.อัครินทร์ ไพบูลย์พานิช)

.................................................................. กรรมการภายนอกมหาวิทยาลัย

(ผู้ช่วยศาสตราจารย์ ดร.ฐิตารีย์ รุ่งรัตน์เกษม)

# กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้เรื่อง “การเปรียบเทียบวิธีบูตสแตรปในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นที่มีมิติสูง” สำเร็จลุล่วงไปได้ด้วยดี เพราะได้รับความอนุเคราะห์ และช่วยเหลือจากผู้มีพระคุณหลายท่านด้วยกัน ทางผู้วิจัยจึงใคร่ขอขอบคุณในความช่วยเหลือต่าง ๆ ดังต่อไปนี้

ขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.วิฐรา พึ่งพาพงศ์ ที่กรุณารับเป็นอาจารย์ที่ปรึกษาวิทยานิพนธ์และคอยให้คำปรึกษาในเรื่องต่าง ๆ รวมถึงสละเวลาคอยติดตาจารย์ ดร.เสกสรร เกียรติสุไพบูลย์ ประธานกรรมการสอบวิทยานิพนธ์ และผู้ช่วยศาสตราจารย์ ดร.อัครินทร์ ไพบูลย์พานิช กรรมการสอบวิทยานิพนธ์ ที่ได้ให้ข้อเสนอแนะ และข้อคิดที่เป็นประโยชน์อย่างยิ่ง รวมทั้งชี้ให้เห็นถึงข้อผิดพลาดต่าง ๆ ซึ่งทำให้วิทยานิพนธ์ฉบับนี้สมบูรณ์ และมีประสิทธิภาพยิ่งขึ้น

ขอกราบขอบพระคุณ คุณพ่อ คุณแม่ และครอบครับที่คอยสนับสนุน เป็นกำลังใจ และติดตามความก้าวหน้าในการทำวิทยานิพนธ์มาโดยตลอด และขอขอบคุณเพื่อน ๆ พี่ ๆ ทุกคน ที่คอยให้กำลังใจให้ความช่วยเหลือ คำปรึกษา และข้อเสนอแนะต่าง ๆ ในการทำวิทยานิพนธ์ฉบับนี้

ภูวกร นิธิศนทีกุล

# สารบัญ

หน้า

# บทที่ 1
# บทนำ

## 1.1 ความเป็นมาและความสำคัญของปัญหา

จิตรณเวช (2558) การวิเคราะห์การถดถอยเป็นเครื่องมือที่ง่ายต่อการใช้งาน อีกทั้งยังเป็นเครื่องมือที่เป็นที่รู้จักอย่างแพร่หลาย เพราะสามารถใช้กับข้อมูลที่หลากหลาย (สินสมบูรณ์ทอง 2553) มีทฤษฎีทางสถิติอนุมานอย่างชัดเจน ทั้งการประมาณค่าด้วยวิธีกำลังสองน้อยสุด (Ordinary Least Squares Method) การประมาณช่วงความเชื่อมั่น การทดสอบสมมติฐานต่าง ๆ อย่างไรก็ตาม (พึ่งพาพงศ์ 2558) การวิเคราะห์การถดถอยยังมีข้อจำกัดสำหรับข้อมูลที่มีมิติสูง (High-dimensional Data) หรือข้อมูลที่จำนวนตัวแปรอิสระมากกกว่าจำนวนตัว อย่างกล่าวคือไม่สามารถหาค่าประมาณสัมประสิทธิ์การถดถอยจากวิธีกำลังสองน้อยสุดได้ นอกจากนี้ การอนุมานต่าง ๆ ที่เกี่ยวกับค่าสัมประสิทธิ์การถดถอยก็ไม่สามารถทำได้เช่นกัน ซึ่งได้แก่ การทดสอบสมมติฐานและการประมาณช่วงความเชื่อมั่น (พรดาเนินสวัสดิ์ 2560) ทั้งนี้การทดสอบสมมติฐานและการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยมีประโยชน์อย่างมากในการเข้าใจถึงความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระ

Tibshirani (1996) ได้เสนอวิธีแลซโซ ซึ่งสามารถใช้ได้กับข้อมูลที่มีมิติสูง ลดปัญหาตัวแปรอิสระที่มีความสัมพันธ์กันสูง (Multicollinearity) รวมทั้งยังสามารถประมาณค่าและทำการคัดเลือกตัวแปรได้พร้อม ๆ กัน โดยตัวแปรที่ไม่ได้คัดเลือกจะมีค่าสัมประสิทธิ์การถ ดถอยเท่ากับศูนย์ แต่ปัญหาของวิธีแลซโซคือ แลซโซสามารถเลือกตัวแปรเข้าได้มากสุดเป็นจำนวนของตัวแปรตามเท่านั้น ดังนั้นหากข้อมูลมีจำนวนตัวแปรอิสระมากกว่าขนาดของตัวอย่างเป็นจำนวนมาก ตัวแบบที่ได้จากวิธีแลซโซอาจจะไม่มีความเหมาะสม และในกรณีที่ตัวแปรอิสระที่มีความสัมพันธ์กันสูง วิธีแลซโซมีแนวโน้มที่จะเลือกตัวแปรเพียงตัวเดียวจากกลุ่มตัวแปรอิสระที่มีความสัมพันธ์กันสูงเข้าตัวแบบ โดยไม่สนใจว่าจะเป็นตัวแปรใดในกลุ่ม ดั้งนั้นวิธีแลซโซมีข้อจำกัดบางประการในการคัดเลือกตัวแปรเข้าสู่ตัวแบบ ที่ไม่มีความคงเส้นคงว่า (Consistency)

Zou (2006) นำเสนอแลซโซปรับได้ (Adaptive Lasso) ซึ่งช่วยแก้ปัญหาข้อจำกัดบางประการในการคัดเลือกตัวแปรเข้าสู่ตัวแบบ ที่ไม่มีความคงเส้นคงว่า (Consistency) ดังที่ได้กล่าวมาข้างต้น เนื่องจากตัวประมาณแลซโซมีการปรับค่าของตัวประมาณเท่ากันทุกตัว ดังนั้น Zou (2006) จึงเสนอวิธีการประมาณแบบแลซโซปรับได้ (Adaptive LASSO) ซึ่งปรับปรุงข้อจำกัดของตัวประมาณการถดถอยแบบ LASSO โดยถ่วงน้ำหนักของค่า $l_1$ (penalty function)

แม้ว่าวิธีแลซโซและแลซโซปรับได้สามารถแก้ไขข้อจำกัดของการวิเคราะห์การถดถอยในการวิเคราะห์ข้อมูลที่มีมิติสูงได้ อย่างไรก็ตาม วิธีการดังกล่าวให้ผลลัพธ์เป็นการประมาณค่าแบบจุดของสัมประสิทธิ์การถดถอยเท่านั้น ตามที่ได้กล่าวไว้ข้างต้นว่าปัญหาการถดถอยเชิงเส้นที่คนมักจะให้ความสนใจ คือการตรวจสอบตัวแปรอิสระแต่ละตัวว่ามีความสัมพันธ์กับตัวแปรตามหรือไม่ ดังนั้นผู้วิจัยจึงมีความสนใจในการตรวจสอบความสัมพันธ์ดังกล่าว

ในงานวิจัยนี้ทางผู้วิจัยจะใช้การสร้างช่วงความเชื่อมั่นของสัมประสิทธิ์การถดถอยในการตรวจสอบความสัมพันธ์ระหว่างตัวตัวแปรตามและแปรอิสระ (ลิสวัสด์ 2561) โดยทำการเปรียบเทียบช่วงความเชื่อมั่นจากวิธีบูตสแตรปในแบบต่าง ๆ (Chatterjee and Lahiri 2011) ซึ่งวิธีบูตสแตรปสามารถทำได้หลายวิธี ผู้วิจัยจึงได้นำเสนอ วิธีสุ่มตัวแปรตามและตัวแปรอิสระ วิธีสุ่มส่วนเหลือ และวิธีสุ่มค่าถ่วงน้ำหนัก (Rauch 2018) ได้นำเสนอการสร้างช่วงความเชื่อมั่นของสัมประสิทธิ์การถดถอยโดย Simple Percentile Bootstrap , Nonparametric Bias-corrected and Accelerated (BCa) Percentile Boostrap และNonparametric Bias-corrected (BC) Percentile Boostrap จึงสนใจที่จะศึกษาเป็นงานวิจัย

### 1.2 วัตถุประสงค์การวิจัย

เพื่อเปรียบเทียบประสิทธิภาพของการประมาณค่าสัมประสิทธิ์การถดถอยในสมการเชิงเส้นทั้ง 3 วิธี คือวิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มตัวแปรตามและตัวแปรอิสระ วิธีสุ่มส่วนเหลือ และวิธีสุ่มตัวค่าถ่วงน้ำหนัก



### 1.3 สมมติฐานการวิจัย

1.3.1 วิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นวิธีสุ่มตัวแปรตาม และตัวแปรอิสระ วิธีสุ่มส่วนเหลือ และวิธีสุ่มค่าถ่วงน้ำหนักน่าจะมีความเหมาะสมในการใช้งานที่แตกต่างกันตามลักษณะของข้อมูลที่จำลองขึ้น

1.3.2 วิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นวิธีสุ่มตัวแปรตาม และตัวแปรอิสระ วิธีสุ่มส่วนเหลือ และวิธีสุ่มค่าถ่วงน้ำหนัก ซึ่งวิธีสุ่มค่าถ่วงน้ำหนัก น่าจะมีประสิทธิภาพดีที่สุด

**1.4 ขอบเขตของการวิจัย**

1.4.1 ในงานวิจัยนี้จะใช้ผลลัพธ์ที่ได้จากการจำลองข้อมูลจาก

สมการถดถอยเชิงเส้นดังนี้

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1.1.1}$$

เมื่อ $\mathbf{y}$ คือเวกเตอร์ของตัวแปรตาม ขนาด $n \times 1$

$\mathbf{X}$ คือเมทริกซ์ของตัวแปรอิสระ ขนาด $n \times p$,

$\boldsymbol{\beta}$ คือเวกเตอร์พารามิเตอร์ที่ไม่ทราบค่าจริง หรือค่าสัมประสิทธิ์การถดถอย ขนาด $p \times 1$

$\boldsymbol{\varepsilon}$ คือเวกเตอร์ของความคลาดเคลื่อนเชิงสุ่ม ขนาด $n \times 1$

โดยกำหนดให้

$\mathbf{X_j} \sim N(\mathbf{0}, \mathbf{I_p})$ เมื่อ $j = 1, 2, 3, \dots, p$

$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I_n})$ เมื่อ $n = 1, 2, 3, \dots, 100$

ทั้งนี้ในการนำข้อมูลที่จำลองไปใช้ต่อในการวิเคราะห์การถดถอยเชิงเส้น จะถือว่าค่าของตัวแปรอิสระ เป็นค่าคงที่

1.4.2 ศึกษาภายใต้

1) ข้อมูลขนาดตัวอย่างมากกว่าจำนวนตัวแปรอิสระ (ข้อมูลที่มีมิติต่ำ)

อัตราส่วนระหว่างขนาดตัวอย่างและจำนวนตัวแปรอิสระ $n : p$ ที่ $100 : 20$

1) ข้อมูลที่ไม่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก (No small nonzero coefficients)

$$\boldsymbol{\beta} = \left[2, 4, -5, 1, -3, \underbrace{0, 0, 0, \dots, 0}_{15\text{ ตัว}}\right]^{\mathbf{T}}$$

2) ข้อมูลที่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก (Some small nonzero coefficients)

$$\boldsymbol{\beta} = \left[4, 0.1, -0.2, -4, -3, \underbrace{0, 0, 0, \dots, 0}_{15\text{ ตัว}}\right]^{\mathbf{T}}$$

2) ข้อมูลขนาดตัวอย่างน้อยกว่าจำนวนตัวแปรอิสระ (ข้อมูลที่มีมิติสูง)

อัตราส่วนระหว่างขนาดตัวอย่างและจำนวนตัวแปรอิสระ $n : p$ ที่ $100 : 500$

1) ข้อมูลที่ไม่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก (No small nonzero coefficients)

$$\boldsymbol{\beta} = \left[3, 12, 8, -5, -13, -4, 7, -11, 3, -6, 7, 9, -16, -3, -2, 5, 7, 4, 24, 4, \underbrace{0, 0, 0, \dots, 0}_{480\text{ ตัว}}\right]^{\mathbf{T}}$$

2) ข้อมูลที่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก (Some small nonzero coefficients)

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

1.4.3 ศึกษาภายใต้การกำหนดการทำซ้ำจำนวน 1,000 ครั้ง

**1.5 วิธีดำเนินการวิจัย**

1.5.1 ค้นคว้าเอกสาร ทฤษฎี และกรอบแนวคิดที่เกี่ยวข้อง

1.5.2 กำหนดค่าจำลอง

1) ขนาดตัวอย่าง

2) จำนวนตัวแปรอิสระ

1.5.3 จำลองข้อมูลตัวตัวแปรตามและแปรอิสระ

1.5.4 สุ่มบูตสแตรปแบบไม่ใช้พารามิเตอร์สำหรับการวิเคราะห์การถดถอยเชิงเส้น

1 ตัวแปรอิสระ

2 วิธีสุ่มส่วนเหลือ

3 วิธีสุ่มค่าถ่วงน้ำหนัก

1.5.5 สร้างช่วงความเชื่อมั่นของสัมประสิทธิ์การถดถอย

1 Simple Percentile Bootstrap

2 Nonparametric Bias-corrected and Accelerated (BCa) Percentile Boostrap

3 Nonparametric Bias-corrected (BC) Percentile Boostrap

1.5.6 เปรียบเทียบผลการวิเคราะห์และสรุปผล



**1.6 แนวทางการวิเคราะห์ข้อมูลและสถิติที่ใช้ในการวิเคราะห์**

แนวทางการวิเคราะห์ข้อมูลและสถิติที่ใช้ในการวิเคราะห์จะใช้การเปรียบเทียบค่าเฉลี่ยเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง ค่าเฉลี่ยความกว้างช่วงความเชื่อมั่น ค่าเฉลี่ยของอัตราผลบวกเทียม และค่าเฉลี่ยของอัตราผลลบเทียม

1.6.1. ค่าเฉลี่ยเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง

สำหรับค่าเฉลี่ยเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริงคำนวณโดยขั้นตอนต่อไปนี้ดังนี้

1) พิจารณาช่วงความเชื่อมั่นสำหรับแต่ละ $\beta_j$ ที่คำนวณได้จากการประมาณค่าโดยวิธีสุ่มต่าง ๆ ที่สามารถครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง

2) จะได้จำนวนครั้งที่ช่วงความเชื่อมั่นสำหรับแต่ละ $\beta_j$ ที่คำนวณได้จากการประมาณค่าโดยวิธีสุ่มต่าง ๆ ที่สามารครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง

3) นำจำนวนครั้งที่หาได้จากข้อ 2) คำนวณหาเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นสำหรับแต่ละ $\beta_j$ ที่คำนวณได้จากการประมาณค่าโดยวิธีสุ่มต่าง ๆ ที่สามารครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง

4) นำเปอร์เซ็นต์ จากข้อ 3) สำหรับแต่ละ $\beta_j$ คำนวณหาค่าเฉลี่ยเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นสำหรับแต่ละ $\beta_j$ ที่คำนวณได้จากการประมาณค่าโดยวิธีสุ่มต่าง ๆ ที่สามารครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง

1.6.2 ค่าเฉลี่ยความกว้างช่วงความเชื่อมั่น

สำหรับค่าเฉลี่ยความกว้างช่วงความเชื่อมั่นคำนวณโดยขั้นตอนต่อไปนี้ดังนี้

1) คำนวณความกว้างช่วงความเชื่อมั่นสำหรับแต่ละ $\beta_j$ ที่คำนวณได้จากการประมาณค่าโดยวิธีสุ่มต่าง ๆ คือ $\left[\underline{\beta_j^*}, \overline{\beta_j^*}\right]$ ได้จากสามการดั้งนี้

$$I_j = \overline{\beta_j^*} - \underline{\beta_j^*} \quad \text{เมื่อ } j = 1, 2, 3, \dots, p$$

2) นำค่าความกว้างช่วงความเชื่อมั่นที่คำนวณได้จากข้อที่ 2) หาค่าเฉลี่ย

1.6.3 ค่าเฉลี่ยของอัตราผลบวกเทียม (False positive rate)

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

สำหรับค่าเฉลี่ยของอัตราผลบวกเทียมคำนวณโดยขั้นตอนต่อไปนี้ดังนี้

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

2) คำนวณจำนวนที่ $0 \notin CI$

3) นำค่าที่ได้จากข้อที่ 1) และ 2) หาอัตราส่วนระหว่าง จำนวนที่ $\beta_j = 0$ และ $0 \notin CI$ ต่อจำนวนที่ $0 \notin CI$

4) นำอัตราส่วนที่ได้จากข้อ 3) คำนวณค่าเฉลี่ย

1.6.4 ค่าเฉลี่ยของอัตราผลลบเทียม (False negative rate)

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

สำหรับค่าเฉลี่ยของอัตราผลลบเทียมคำนวณโดยขั้นตอนต่อไปนี้ดังนี้

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

2) คำนวณจำนวนที่ $0 \in CI$

3) นำค่าที่ได้จากข้อที่ 1) และ 2) หาอัตราส่วนระหว่าง จำนวนที่ $\beta_j \neq 0$ และ $0 \in CI$ ต่อจำนวนที่ $0 \in CI$

4) นำอัตราส่วนที่ได้จากข้อ 3) คำนวณค่าเฉลี่ย

### 1.7 การนำเสนอข้อมูล

นำเสนอข้อมูลในรูปตาราง เพื่อตรวจสอบดูว่าวิธีสุ่มบูตสแตรปแบบไม่ใช้พารามิเตอร์สำหรับการวิเคราะห์การถดถอยเชิงเส้นแบบใดให้ผลดีกว่ากัน

# บทที่ 2
# เอกสารและงานวิจัยที่เกี่ยวข้อง

## 2.1. ตัวแบบการถดถอยเชิงเส้น (Linear Regression)

การวิเคราะห์การถดถอยเชิงเส้นเป็นการสร้างตัวแบบทางคณิตศาสตร์ เพื่อแสดงความสัมพันธ์ระหว่างตัวแปรอิสระ $\mathbf{X}$ และตัวแปรตาม $\mathbf{y}$ ซึ่งมีสมการถดถอยเชิงเส้นดังนี้

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2.1.1}$$

เมื่อ $\mathbf{y}$ คือเวกเตอร์ของตัวแปรตาม ขนาด $n \times 1$

$\mathbf{X}$ คือเมทริกซ์ของตัวแปรอิสระ ขนาด $n \times p$,

$\boldsymbol{\beta}$ คือเวกเตอร์พารามิเตอร์ที่ไม่ทราบค่าจริง หรือค่าสัมประสิทธิ์ความถดถอย ขนาด $p \times 1$

$\boldsymbol{\varepsilon}$ คือเวกเตอร์ของความคลาดเคลื่อนเชิงสุ่ม ขนาด $n \times 1$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{2.1.2}$$

โดยมีข้อสมมุติ $\boldsymbol{\varepsilon}$ มีการแจกแจงปกติมีค่าเฉลี่ย $\mathbf{0}$ และความแปรปรวน $\sigma^2\mathbf{I_n}$ นั่นคือ $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I_n})$

เมื่อ $\mathbf{0}$ คือเวกเตอร์ 0 ขนาด $n \times 1$

$\mathbf{I_n}$ คือเมทริกซ์เอกลักษณ์ขนาด $n \times n$

ดังนั้นจะได้ว่า $E(y|\mathbf{X}) = \mu_{\mathbf{y}|X} = \mathbf{X}\boldsymbol{\beta}$

## 2.2 การประมาณค่าสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยที่สุด (Ordinary Least Squares: OLS)



การวิเคราะห์การถดถอยเป็นการสร้างสมการแบบถดถอยภายใต้ตัวแบบที่กำหนดจากสมการ (2.1.1) เวกเตอร์ $\boldsymbol{\beta}$ เป็นเวกเตอร์ของพารามิเตอร์ที่ไม่ทราบค่าจริง จึงต้องใช้ข้อมูลจากตัวแปรตาม $\mathbf{y}$ และตัวแปรอิสระ $\mathbf{X}$ มาทำการประมาณค่าของพารามิเตอร์ $\boldsymbol{\beta}$ เมื่อแทนค่าประมาณพารามิเตอร์ลงในตัวแบบจะได้สมการถดถอย(2.2.1)

$$\hat{\mathbf{y}} = \hat{\boldsymbol{\beta}}_\mathbf{1}\mathbf{x_1} + \hat{\boldsymbol{\beta}}_\mathbf{2}\mathbf{x_2} + \cdots + \hat{\boldsymbol{\beta}}_\mathbf{p}\mathbf{x_p} \tag{2.2.1}$$

หรือเขียนในรูปเมทริกซ์

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \tag{2.2.2}$$

เมื่อ $\hat{\mathbf{y}}$ คือเวกเตอร์ของตัวแปรตาม ขนาด $n \times 1$

$\mathbf{X}$ คือเมทริกซ์ของตัวแปรอิสระ ขนาด $n \times p$,

$\widehat{\boldsymbol{\beta}}$ คือเวกเตอร์ของสัมประสิทธิ์การถดถอยที่ใช้ในการประมาณค่าเวกเตอร์ของพารามิเตอร์ของตัวแบบ ขนาด $p \times 1$

การหาค่าสัมประสิทธิ์การการถดถอยซึ่งใช้ในการประมาณพารามิเตอร์ของตัวแบบใช้วิธีกำลังสองน้อยที่สุด โดยมีหลักการที่พยายามหาค่าของ $\widehat{\boldsymbol{\beta}}$ ในพจน์ของตัวแปรตาม $\mathbf{y}$ และตัวแปรอิสระ $\mathbf{X}$ ที่ใช้ในตัวแบบ เพื่อที่จะทำให้ผลบวกของความคลาดเคลื่อนเชิงสุ่มกำลังสองมีค่าต่ำสุดนั่นคือ $\sum_{\mathbf{i=1}}^{\mathbf{n}} \boldsymbol{\varepsilon}^{\mathbf{2}}$ หรือ $\boldsymbol{\varepsilon}^{\mathbf{T}}\boldsymbol{\varepsilon}$ มีค่าต่ำสุด ซึ่งสามารถหาได้โดยการเขียน $\sum_{\mathbf{i=1}}^{\mathbf{n}} \boldsymbol{\varepsilon}^{\mathbf{2}}$ ในพจน์ของ $\boldsymbol{\beta}$ และหาอนุพันธ์ของ $\boldsymbol{\varepsilon}^{\mathbf{T}}\boldsymbol{\varepsilon}$ เทียบกับ $\boldsymbol{\beta}$ เมื่อ $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ แล้วกำหนดค่าอนุพันธ์ที่ได้ให้เท่ากับ 0 เพื่อแก้สมการหาค่าของ $\widehat{\boldsymbol{\beta}}$

จากตัวแบบในสมการ (2.1.1)
$$\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}. \tag{2.2.3}$$
ผลบวกของความคลาดเคลื่อนเชิงสุ่มกำลังสองเท่ากับสมการ(1.4.4)
$$\boldsymbol{\varepsilon}^{\mathbf{T}}\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathbf{T}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{y}^{\mathbf{T}}\mathbf{y} - \mathbf{2}\boldsymbol{\beta}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{y} + \boldsymbol{\beta}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\boldsymbol{\beta} \tag{2.2.4}$$
$\boldsymbol{\varepsilon}^{\mathbf{T}}\boldsymbol{\varepsilon}$ จะมีค่าต่ำสุดเมื่อ
$$\frac{\partial}{\partial \boldsymbol{\beta}}\boldsymbol{\varepsilon}^{\mathbf{T}}\boldsymbol{\varepsilon}\Big|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} = -2\mathbf{X}^{\mathbf{T}}\mathbf{y} + 2\mathbf{X}^{\mathbf{T}}\mathbf{X}\widehat{\boldsymbol{\beta}} = 0 \tag{2.2.5}$$
จากสมการ (2.2.5) เขียนได้เป็น
$$\mathbf{X}^{\mathbf{T}}\mathbf{y} = \mathbf{X}^{\mathbf{T}}\mathbf{X}\widehat{\boldsymbol{\beta}} \tag{2.2.6}$$
สมการใน (2.2.6) เรียกว่าสมการปกติ (normal equation) และเนื่องจากแรงก์ของเมทริกซ์ $\mathbf{X}$ เท่ากับ $p$ ทำให้เมทริกซ์ $\mathbf{X}^{\mathbf{T}}\mathbf{X}$ มีแรงก์เท่ากับ $p$ ดังนั้น เมทริกซ์ $\mathbf{X}^{\mathbf{T}}\mathbf{X}$ มีเมทริกซ์ผกผัน การแก้ระบบสมการได้ดังสมการที่ (2.2.7)
$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-\mathbf{1}}\mathbf{X}^{\mathbf{T}}\mathbf{y} \tag{2.2.7}$$
ดังนั้น $\widehat{\boldsymbol{\beta}}$ เป็นตัวประมาณค่ากำลังสองน้อยที่สุดของพารามิเตอร์ $\boldsymbol{\beta}$ ที่ทำให้ $\boldsymbol{\varepsilon}^{\mathbf{T}}\boldsymbol{\varepsilon}$ มีค่าต่ำที่สุด

### 2.3 การทดสอบสมมติฐานของพารามิเตอร์ของตัวแบบ

ในการวิเคราะห์การถดถอย เมื่อมีการประมาณพารามิเตอร์ของตัวแบบแล้ว จะได้ค่าของสัมประสิทธิ์การถดถอย ทำให้สามารถเขียนสมการถดถอยได้ อย่างไรก็ตามโดยทั่วไปเรามักจะสนใจด้วยว่า ตัวแปรอิสระแต่ละตัวมีความสัมพันธ์กับตัวแปรตามจริงหรือไม่ โดยการทดสอบสมมติฐานทางสถิติดังนี้

$$H_0: \beta_j = 0 \text{ vs } H_a: \beta_j \neq 0 \qquad ; j = 1, 2, \ldots, p \tag{2.3.1}$$

สำหรับสถิติที่ใช้ทดสอบคือ

$$t = \frac{\widehat{\beta}_j - 0}{\sqrt{\frac{MSE}{S_{xx}}}} \tag{2.3.2}$$

เมื่อ $MSE = \frac{SSE}{n-2}$

$\mathrm{SSE} = \sum_{j=1}^{n} e_j^2 = \sum_{j=1}^{n}(Y_j - \widehat{Y}_J)$

$S_{xx} = \sum_{j=1}^{n} X_j^2 - \frac{\left(\sum_{j=1}^{n} x_j\right)^2}{n}$

โดย หาก $|t| \le t_{\frac{\alpha}{2},n-p}$แล้วไม่ปฏิเสธสมมติฐานหลัก $(H_0)$

หาก $|t| > t_{\frac{\alpha}{2},n-p}$ แล้วปฏิเสธสมมติฐานหลัก $(H_0)$

**2.4 ช่วงความเชื่อมั่นสัมประสิทธิ์การถดถอย**

$\widehat{\beta_j}$ เป็นตัวประมาณพารามิเตอร์ $\beta_j$ โดยมีการแจกแจงแบบ $t_{n-p}$

ดังนั้น $100(1-\alpha)\%$ ช่วงความเชื่อมั่น $\beta_j$ เท่ากับสมการ (2.4.1)

$$\hat{\beta}_j = \pm t_{\frac{\alpha}{2},n-p}\,\mathrm{SE}(\hat{\beta}_j) \qquad ; j = 0, 1, 2, \dots, p \tag{2.4.1}$$

เมื่อ $\mathrm{SE}(\hat{\beta}_j)$ เท่ากับรากที่สองของสมาชิกบนเส้นทแยงมุมที่ j ของเมทริกซ์ $(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}$ คูณกับ $s^2$

นั่นคือ

$$\mathrm{SE}(\hat{\beta}_j) = \sqrt{s^2(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}} \tag{2.4.2}$$

โดยที่

$$s^2 = \frac{e^T e}{n-p} \tag{2.4.3}$$

ทั้งนี้ในการทดสอบสมมติฐาน $H_0: \beta_j = 0 \text{ vs } H_a: \beta_j \neq 0$ ที่ระดับนัยสำคัญ $\alpha$ สามารถใช้ช่วงความเชื่อมั่นที่ $(1-\alpha)\%$ ของ $\beta_j$ในการตรวจสอบได้ กล่าวคือ หากช่วงความเชื่อมั่นที่ $(1-\alpha)\%$ ของ $\beta$ ไม่สามารถครอบคลุมค่า 0 จะนำไปสู่การปฏิเสธ $H_0$ ที่ระดับนัยสำคัญ $\alpha$

**2.5 ตัวแบบการถดถอยแบบแลซโซ (Lasso Regression)**

ตัวแบบการถดถอยแบบแลซโซต้องการแก้ปัญหาความสัมพันธ์เชิงเส้นของตัวแปรอิสระ (Multicolinearity) และมีคุณสมบัติในการเลือกตัวแปรเข้าตัวแบบ (Variable Selection) ไปพร้อมกับการประมาณค่าสัมประสิทธิ์การถดถอย โดยการหาค่าสัมประสิทธิ์การถดถอยที่ทำให้ผลบวกระหว่างผลรวมความคลาดเคลื่อนกำลังสองรวมกับผลรวมกำลังสองของค่าสัมประสิทธิ์ถ่วงน้ำหนักมีค่าต่ำที่สุด พจน์แรกคือผลรวมความคลาดเคลื่อนกำลังสองนั้นเรียกว่า Loss Function สำหรับผลรวมของค่าสมบูรณ์ของสัมประสิทธิ์ถ่วงน้ำหนักนั้นเรียกว่า $l_1$ (Penalty Function) ซึ่ง Penalty Function ของวิธีแลซโซจะแตกต่างจาก Penalty Function ของตัวแบบการถดถอยแบบริดจ์จาก สมการ (2.5.1)

$$\hat{\boldsymbol{\beta}}_{\text{lasso}} = \underbrace{\operatorname{argmin}}_{\beta}\left(\underbrace{\sum_{i=1}^{n}(y_i - \mathbf{X_i}\boldsymbol{\beta})^2}_{\text{Loss Function}} + \underbrace{\lambda\sum_{j=1}^{p}|\beta_j|}_{\text{Penalty Function}}\right) \tag{2.5.1}$$

เมื่อ $\lambda$ คือพารามิเตอร์ปรับแต่ง (Tuning Parameter) ซึ่งควบคุมขนาดการหดตัว (Shrinkage) ของ $\widehat{\boldsymbol{\beta}}_{\text{lasso}}$

**2.6 ตัวแบบการถดถอยแลซโซปรับได้ (Adaptive LASSO Regression)**

ซู (Zou,2006) ได้กล่าวถึงสมบัติในการคัดเลือกตัวแปรอิสระและการประมาณค่าสัมประสิทธิ์การถดถอยของวิธีแลซโซว่าการคัดเลือกตัวแปรอิสระด้วยวิธีแลซโซอาจจะเป็นกระบวนการที่ไม่คงเส้นคงวาได้ เนื่องจากตัวประมาณแลซโซมีการปรับค่าของตัวประมาณเท่ากันทุกตัว ภายใต้ข้อจากัด $l_1$ (Penalty Function) ซึ่งบางทีอาจจะไม่เหมาะสม ดังนั้นซูจึงเสนอวิธีการ ประมาณแบบแลซโซปรับได้(Adaptive LASSO) ซึ่งปรับปรุงข้อจำกัดของตัวประมาณการถดถอยแบบ LASSO โดยถ่วงน้าหนักของค่า $l_1$ (Penalty Function) จากสมการ (2.6.1)

$$\widehat{\boldsymbol{\beta}}_{\text{adp.lasso}} = \underbrace{\operatorname{argmin}}_{\beta}\left(\underbrace{\sum_{i=1}^{n}(y_i - \mathbf{X}_i\boldsymbol{\beta})^2}_{\text{Loss Function}} + \underbrace{\lambda\sum_{j=1}^{p} w_j|\boldsymbol{\beta}_j|}_{\text{Penalty Function}}\right) \tag{2.6.1}$$

เมื่อ $\lambda$ คือพารามิเตอร์ปรับแต่ง (Tuning Parameter) ซึ่งควบคุมขนาดการหดตัว (Shrinkage) ของ $\widehat{\boldsymbol{\beta}}_{\text{adp.lasso}}$

$w_j$ คือค่าถ่วงน้ำหนักของ $|\beta_j|$

**2.7 วิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์สำหรับการวิเคราะห์การถดถอยเชิงเส้น**

1) วิธีสุ่มตัวแปรตามและตัวแปรอิสระ

วิธีสุ่มตัวแปรตาม และตัวแปรอิสระ จะเป็นการสุ่มจากกลุ่มตัวอย่างขนาด n แบบใส่กลับคืน กล่าวคือข้อมูลที่เกิดจากการสุ่มของกลุ่มตัวอย่างที่ได้มาในแต่ละการสุ่มแต่ละครั้งจะมีโอกาสได้ข้อมูลซ้ำ ซึ่งจะทำการสุ่ม 1,000 ครั้ง เท่ากับขนาดของกลุ่มตัวอย่าง จะเห็นว่าข้อมูลของกลุ่มตัวอย่างทั้งหมดในแต่ละข้อมูลจะให้ผลของการถ่วงน้ำหนักที่แตกต่างกัน ซึ่งในบางครั้งข้อมูลบางข้อมูลอาจจะไม่ได้ถูกสุ่มขึ้นมา โดยวิธีสุ่มตัวแปรตาม และตัวแปรอิสระ มีวิธีการสุ่มดังนี้

1) สุ่มตัวแปรตาม ($y_i$) และตัวแปรอิสระ ($x_i$)จากกลุ่มตัวอย่าง จะได้ $y_i^*$ และ $x_i^*$ โดยมีขนาดเท่ากับกลุ่มตัวอย่าง
2) นำ $y_i^*$ และ $x_i^*$ ประมาณค่าสัมประสิทธิ์ความถดถอย ($\widehat{\boldsymbol{\beta}}$)
3) ใช้ตัวอย่างของข้อมูลเดิมทำซ้ำตามขั้นตอน 1-2 จำนวน 1,000 ครั้ง
4) ได้ค่าสัมประสิทธิ์การถดถอยในแต่ละครั้งดังนี้

$$\widehat{\boldsymbol{\beta}}^* = [\widehat{\boldsymbol{\beta}}^{(1)}, \widehat{\boldsymbol{\beta}}^{(2)}, \widehat{\boldsymbol{\beta}}^{(3)}, \ldots, \widehat{\boldsymbol{\beta}}^{(1,000)}]$$

เมื่อ $\widehat{\boldsymbol{\beta}}^{(\mathbf{b})}$ คือสัมประสิทธิ์การถดถอยจากตัวอย่างบูตสแตรปครั้งที่ b

2) วิธีสุ่มส่วนเหลือ

วิธีสุ่มส่วนเหลือ เป็นการสุ่มส่วนเหลือที่ได้จากสมการถดถอยเชิงเส้น ที่มาจากกลุ่มตัวอย่างขนาด n แบบใส่กลับคืน กล่าวคือ มีโอกาสที่จะได้ส่วนเหลือที่ซ้ำกันซึ่งอาจจะทำให้ค่า $\hat{y}_i^*$ มีค่าเท่าเดิมหรือไม่เท่าเดิมก็ได้ แต่มีโอกาสค่อนข้างยากกว่า วิธีสุ่มตัวแปรตาม และตัวแปรอิสระ ที่จะได้$\hat{y}_i^*$ มีค่าเท่าเดิม และจะทำการสุ่ม 1,000 ครั้ง เท่ากับขนาดของกลุ่มตัวอย่าง โดยวิธีสุ่มส่วนเหลือ มีวิธีสุ่มดังนี้

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 (2.7.1)

$$e_i = y_i - \hat{y}_i \tag{2.7.1}$$

4) นำส่วนเหลือ$(e_i)$ ทำการสุ่มได้ $e_i^*$
5) นำ $e_i^*$ ย้อนกลับไปแทนลงในสมการที่ (2.7.2) จะได้ $y_i^*$

$$y_i^* = \hat{y}_i + e_i^* \tag{2.7.2}$$

6) นำ $y_i^*$ และ $x_i$ ประมาณค่าสัมประสิทธิ์ความถดถอย ($\widehat{\boldsymbol{\beta}}$)
7) นำตัวอย่างของข้อมูลเดิมทำซ้ำตามขั้นตอน 2-6 จำนวน 1,000 ครั้ง
8) ได้ค่าสัมประสิทธิ์การถดถอยในแต่ละครั้งดังนี้

$$\widehat{\boldsymbol{\beta}}^* = [\widehat{\boldsymbol{\beta}}^{(1)}, \widehat{\boldsymbol{\beta}}^{(2)}, \widehat{\boldsymbol{\beta}}^{(3)}, \dots, \widehat{\boldsymbol{\beta}}^{(1,000)}]$$

เมื่อ $\widehat{\boldsymbol{\beta}}^{(b)}$ คือสัมประสิทธิ์การถดถอยจากตัวอย่างบูตสแตรปครั้งที่ b

3) วิธีสุ่มค่าถ่วงน้ำหนัก



จาก Meeker et al. (2017) วิธีสุ่มตัวแปรตามและตัวแปรอิสระ และวิธีสุ่มส่วนเหลือมีโอกาสที่จะทำให้การสุ่มข้อมูลจากกลุ่มตัวอย่างขนาด n บางข้อมูลไม่ถูกนำมาใช้ในตัวแบบ ดังนั้นวิธีสุ่มค่าถ่วงน้ำหนัก จะช่วยแก้ปัญหาดังกล่าวโดยวิธีสุ่มค่าถ่วงน้ำหนัก มีวิธีการสุ่มดังนี้



1) สุ่มค่าถ่วงน้ำหนัก (Weight) การแจกแจงแบบเอกซ์โปเนนเชียล

   นั่นคือ $w_i \sim \text{Exp}(1)$

   จะได้ $w_1, w_2, \dots, w_p$ เมื่อ $j = 1, 2, 3, \dots, p$
2) นำ $y_i, x_i$ และน้ำหนักที่สุ่มได้ประมาณค่าสัมประสิทธิ์ความถดถอย ($\widehat{\boldsymbol{\beta}}$) โดยใช้ Weighted Regression
3) นำตัวอย่างของข้อมูลเดิมทำซ้ำตามขั้นตอน 1-2 จำนวน 1,000 ครั้ง
4) ได้ค่าสัมประสิทธิ์การถดถอยในแต่ละครั้งดังนี้

$$\widehat{\boldsymbol{\beta}}^* = [\widehat{\boldsymbol{\beta}}^{(1)}, \widehat{\boldsymbol{\beta}}^{(2)}, \widehat{\boldsymbol{\beta}}^{(3)}, \dots, \widehat{\boldsymbol{\beta}}^{(1,000)}]$$

เมื่อ $\widehat{\boldsymbol{\beta}}^{(b)}$ คือสัมประสิทธิ์การถดถอยจากตัวอย่างบูตสแตรปครั้งที่ b

### 2.8 วิธีการหาช่วงความเชื่อมั่นของสัมประสิทธิ์การถดถอยเชิงเส้นจากวิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ของสัมประสิทธิ์การถดถอย

จากค่าสัมประสิทธิ์การถดถอยด้วยวิธีบูตสแตรป $\hat{\boldsymbol{\beta}}^{*} = [\hat{\boldsymbol{\beta}}^{(1)}, \hat{\boldsymbol{\beta}}^{(2)}, \hat{\boldsymbol{\beta}}^{(3)}, \dots, \hat{\boldsymbol{\beta}}^{(b)}]$ Meeker et al. (2017) สามารถนำมาใช้หาช่วงความเชื่อมั่นของ $\boldsymbol{\beta}$ มีด้วยกันทั้งหมด 3 วิธี คือ Simple Percentile Bootstrap , Nonparametric Bias-corrected and Accelerated (BCa) Percentile Boostrap และNonparametric Bias-corrected (BC) Percentile

1) Simple Percentile Bootstrap

สำหรับ Simple Percentile Bootstrap การสุ่มตัวแปรตามและตัวแปรอิสระในวิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์กระทำซ้ำ 1,000 ครั้ง เป็นวิธีที่น่าสนใจเนื่องจากมีความง่ายต่อการใช้งานโดยมีวิธีดังนี้

1) จากที่ทำการบูตสแตรปเป็นจำนวน 1,000 ครั้ง สามารถหาช่วงความเชื่อมั่นของสัมประสิทธิ์การถดถอยเชิงเส้น โดย Simple Percentile Bootstrap กล่าวคือ ขอบเขตล่างสัมประสิทธิ์การถดถอยเชิงเส้น $(\underline{\beta_j})$ และขอบเขตบนสัมประสิทธิ์การถดถอยเชิงเส้น $(\overline{\beta_j})$ สามารถหาได้จากสมการ (2.8.1)

$$\left[\underline{\beta_j}, \overline{\beta_j}\right] = \left[\hat{\beta}_{j\ \frac{\alpha}{2}}^{\ *}, \hat{\beta}_{j\ \left(1-\frac{\alpha}{2}\right)}^{\ *}\right] \qquad (2.8.1)$$

เมื่อ $\hat{\beta}_q^{\ *}$ คือ เปอร์เซ็นต์ไทล์ที่ q ของ $\boldsymbol{\beta}$ ที่เกิดจากการบูตสแตรปเป็นจำนวน 1,000 ครั้งสำหรับสัมประสิทธิ์การถดถอย

2) Nonparametric Bias-corrected and Accelerated (BCa) Percentile Boostrap



จาก Simple Percentile Bootstrap เป็นวิธีที่น่าสนใจเนื่องจากมีความง่ายต่อการใช้งาน แต่ยังมีข้อจำกัด กล่าวคือในการทำบูตสแตรปสำหรับกลุ่มตัวอย่างขนาดใหญ่วิธี Simple Percentile Bootstrap ค่อนข้างใช้งานกับลักษณะข้อมูลนี้ได้ไม่ดีเท่ากับวิธี Nonparametric Bias-corrected and Accelerated (BCa) Percentile Boostrap

1) จากที่ทำการบูตสแตรปเป็นจำนวน 1,000 ครั้ง สามารถหาช่วงความเชื่อมั่นของสัมประสิทธิ์การถดถอยเชิงเส้นโดย Nonparametric Bias-corrected and Accelerated (BCa) Percentile Boostrap กล่าวคือ โดยขอบเขตล่างสัมประสิทธิ์การถดถอยเชิงเส้น $(\underline{\beta_j^*})$ และขอบเขตบนสัมประสิทธิ์การถดถอยเชิงเส้น $(\overline{\beta_j^*})$ สามารถหาได้จากสมการ (2.8.2)

$$\left[\underline{\beta_j^*}, \overline{\beta_j^*}\right] = \left[\hat{\beta}_{j\ \alpha_1}^{\ *}, \hat{\beta}_{j\ \alpha_2}^{\ *}\right] \qquad (2.8.2)$$

เมื่อ $\hat{\beta}_q{}^*$ คือเปอร์เซ็นต์ไทล์ที่ q ของ $\boldsymbol{\beta}$ ที่เกิดจากการบูตสแตรปเป็นจำนวน 1,000 ครั้งสำหรับสัมประสิทธิ์การถดถอย

$\alpha_1$ หาได้จากสมการ (2.8.3)

$$\alpha_1 = \Phi_{\mathrm{norm}}\left[z_{\hat{b}} + \frac{z_{\hat{b}} - z_{\left(1-\frac{\alpha}{2}\right)}}{1-\hat{a}\left[z_{\hat{b}} - z_{\left(1-\frac{\alpha}{2}\right)}\right]}\right] \quad (2.8.3)$$

$\alpha_2$ หาได้จากสมการ (2.8.4)

$$\alpha_2 = \Phi_{\mathrm{norm}}\left[z_{\hat{b}} + \frac{z_{\hat{b}} + z_{\left(1-\frac{\alpha}{2}\right)}}{1-\hat{a}\left[z_{\hat{b}} + z_{\left(1-\frac{\alpha}{2}\right)}\right]}\right] \quad (2.8.4)$$

เมื่อ $z_q$ คือเปอร์เซ็นต์ไทล์ที่ q ของการแจกแจงปกติมาตรฐาน

$\hat{b}$ คือสัดส่วนของจำนวนครั้งของการสุ่มบูตสแตรปของ $\hat{\boldsymbol{\beta}}^*$ ที่มีค่าน้อยกว่า $\hat{\boldsymbol{\beta}}$ ต่อจำนวนทั้งหมด

$\hat{a}$ คือความเร่งคงตัวหาได้จากสมการ (2.8.5)

$$\hat{a} = \frac{\sum_{i=1}^{n}\left(\hat{\beta}_{[.]} - \hat{\beta}^{(b)}\right)^3}{6\left[\sum_{i=1}^{n}\left(\hat{\beta}_{[.]} - \hat{\beta}^{(b)}\right)^2\right]^{\frac{3}{2}}} \quad (2.8.5)$$

เมื่อ

$$\hat{\beta}_{[.]} = \sum_{i=1}^{n}\frac{\hat{\beta}_{[i]}}{n} \quad (2.8.6)$$

เมื่อ $\hat{\boldsymbol{\beta}}^{(\mathbf{b})}$ คือสัมประสิทธิ์การถดถอยจากตัวอย่างบูตสแตรปครั้งที่ b

3) Nonparametric Bias-corrected (BC) Percentile Boostrap

1) จากที่ทำการบูตสแตรปเป็นจำนวน B ครั้ง สามารถหาช่วงความเชื่อมั่นของสัมประสิทธิ์การถดถอยเชิงเส้น โดย Nonparametric Bias-corrected (BC) Percentile Boostrap กล่าวคือขอบเขตล่างสัมประสิทธิ์การถดถอยเชิงเส้น $(\underline{\beta_j^*})$ และขอบเขตบนสัมประสิทธิ์การถดถอยเชิงเส้น $(\overline{\beta_j^*})$ สามารถหาได้จากสมการ (2.8.7) หลักการคำนวณเดียวกันกับ 2) โดย $\hat{a} = 0$

$$\left[\underline{\beta_j^*}, \overline{\beta_j^*}\right] = \left[\hat{\beta}_j{}^*{}_{\alpha_1}, \hat{\beta}_j{}^*{}_{\alpha_2}\right] \quad (2.8.7)$$

เมื่อ $\hat{\beta}_q{}^*$ คือเปอร์เซ็นต์ไทล์ที่ q ของ $\boldsymbol{\beta}$ ที่เกิดจากการบูตสแตรปเป็นจำนวน 1,000 ครั้งสำหรับสัมประสิทธิ์การถดถอย

$\alpha_1$ หาได้จากสมการ (2.8.8)

$$\alpha_1 = \Phi_{\mathrm{norm}}\left[2z_{\hat{b}} - z_{\left(1-\frac{\alpha}{2}\right)}\right] \quad (2.8.8)$$

$\alpha_2$ หาได้จากสมการ (2.8.9)

$$\alpha_2 = \Phi_{\mathrm{norm}}\left[2z_{\hat{b}} + z_{\left(1-\frac{\alpha}{2}\right)}\right] \quad (2.8.9)$$

เมื่อ $z_q$ คือเปอร์เซ็นต์ไทล์ที่ q ของการแจกแจงปกติมาตรฐาน

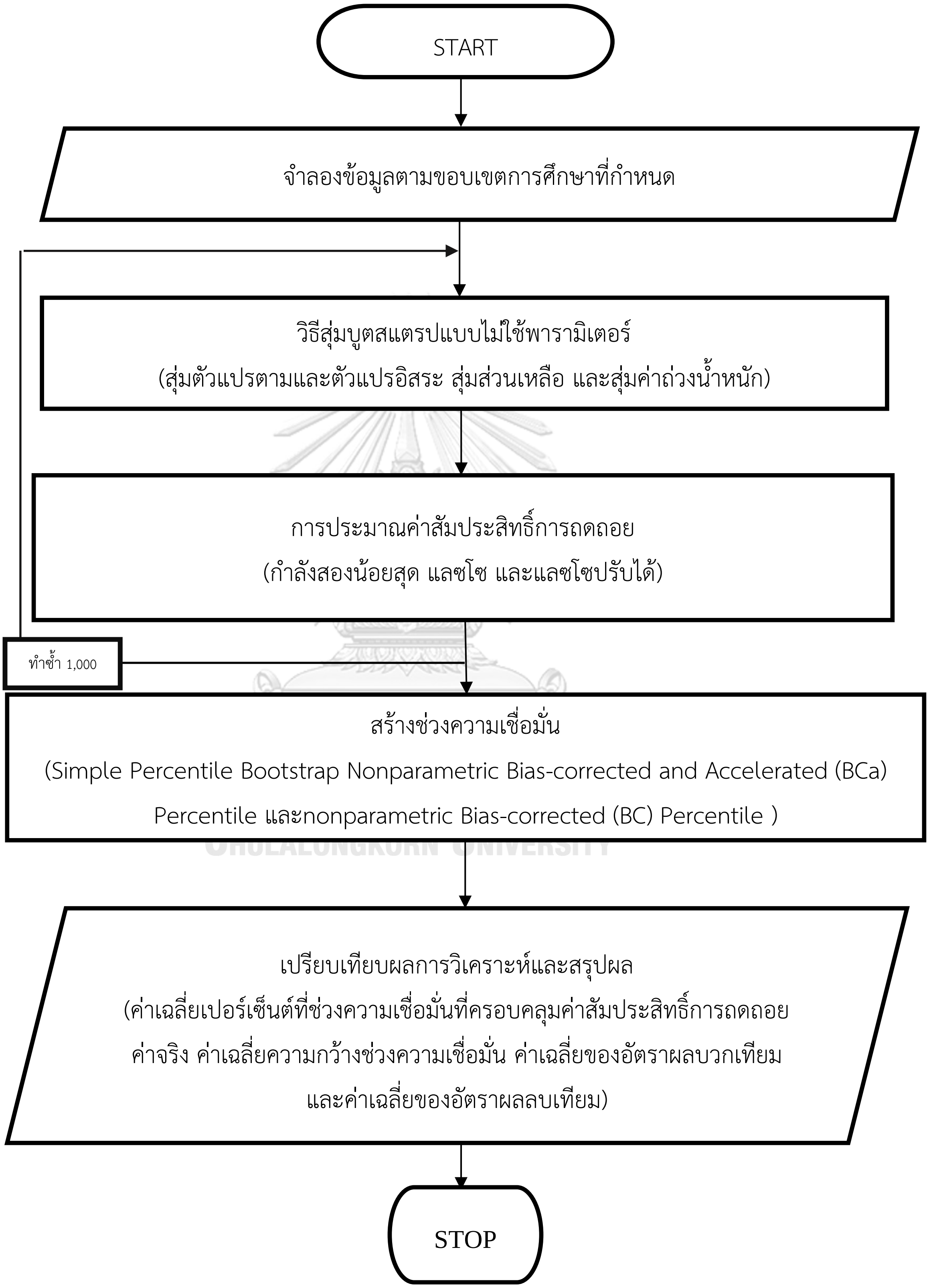
$\hat{b}$ คือสัดส่วนของจำนวนครั้งของการสุ่มบูตสแตรปของ $\widehat{\boldsymbol{\beta}}^*$ ที่มีค่าน้อยกว่า $\widehat{\boldsymbol{\beta}}$ ต่อจำนวนทั้งหมด

# บทที่ 3
# วิธีการดำเนินงานวิจัย

การวิจัยนี้เป็นการศึกษาเปรียบเทียบประสิทธิภาพของการประมาณค่าสัมประสิทธิ์การถดถอยในสมการเชิงเส้นทั้ง 3 วิธี คือวิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มตัวแปรตามและตัวแปรอิสระ วิธีสุ่มส่วนเหลือ และวิธีสุ่มค่าถ่วงน้ำหนัก โดยจะทำการประมาณค่าโดยกำลังสองน้อยที่สุด การประมาณค่าโดยแลซโซ และการประมาณค่าโดยแลซโซปรับได้ จากการวิเคราะห์จำลองข้อมูลทั้งในกรณีที่ข้อมูลขนาดตัวอย่างมากกว่าจำนวนตัวแปรอิสระหรือที่เรียกว่าข้อมูลที่มีมิติต่ำ และกรณีที่ข้อมูลขนาดตัวอย่างน้อยกว่าจำนวนตัวแปรอิสระหรือที่เรียกว่าข้อมูลที่มีมิติสูง ซึ่งในแต่ละกรณีแบ่งย่อยเป็นข้อมูลที่ไม่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก (No small nonzero coefficients) และข้อมูลที่มีค่าสัมประสิทธิ์การถดถอยไม่ใช่ศูนย์ขนาดเล็ก (Some small nonzero coefficients) และทำซ้ำ 1,000 ครั้ง ผ่านโปรแกรม R ในบทนี้จะกล่าวถึงแผนการดำเนินการวิจัย ขั้นตอนในการดำเนินวิจัย และขั้นตอนการทำงานของโปรแกรม ซึ่งมีรายละเอียดดังนี้

**3.1 วิธีการดำเนินงานวิจัย**

วิธีการดำเนินงานวิจัยอธิบายได้แผนภาพดังนี้



**ภาพที่** 3.1 แผนการดำเนินวิจัย

จากภาพที่ 3.1 อธิบายขั้นตอนวิธีการดำเนินงานวิจัยได้ดังนี้

1. กำหนดขอบเขตการศึกษา

1.1 กำหนดระดับความเชื่อมั่น 95%

1.2 $\mathbf{X}_j \sim \mathbf{N}(\mathbf{0}, \mathbf{I_p})$ เมื่อ $\mathrm{j} = 1, 2, \ldots, \mathrm{p}$

1.3 $\boldsymbol{\varepsilon} \sim \mathbf{N}(\mathbf{0}, \mathbf{I_n})$ เมื่อ $\mathrm{n} = 1, 2, \ldots, 100$

1.4 ค่าของตัวแปรอิสระเป็นค่าคงที่

1.5 กำหนด $\beta$ จริงดังนี้

1.5.1 ข้อมูลขนาดตัวอย่างมากกว่าจำนวนตัวแปรอิสระ (ข้อมูลที่มีมิติต่ำ) อัตราส่วนระหว่างขนาดตัวอย่างและจำนวนตัวแปรอิสระ $\mathrm{n} : \mathrm{p}$ ที่ $100 : 20$

1) ข้อมูลที่ไม่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก (No small nonzero coefficients)

$$\boldsymbol{\beta} = \Big[2, 4, -5, 1, -3, \underbrace{0, 0, 0, \ldots, 0}_{\mathbf{15}\text{ ตัว}}\Big]^{\mathbf{T}}$$

2) ข้อมูลที่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก (Some small nonzero coefficients)

$$\boldsymbol{\beta} = \Big[4, 0.1, -0.2, -4, -3, \underbrace{0, 0, 0, \ldots, 0}_{\mathbf{15}\text{ ตัว}}\Big]^{\mathbf{T}}$$

1.5.2 ข้อมูลขนาดตัวอย่างน้อยกว่าจำนวนตัวแปรอิสระ (ข้อมูลที่มีมิติสูง) อัตราส่วนระหว่างขนาดตัวอย่างและจำนวนตัวแปรอิสระ $\mathrm{n} : \mathrm{p}$ ที่ $100 : 500$

1) ข้อมูลที่ไม่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก (No small nonzero coefficients)

$$\boldsymbol{\beta} = \Big[3, 12, 8, -5, -13, -4, 7, -11, 3, -6, 7, 9, -16, -3, -2, 5, 7, 4, 24, 4, \underbrace{0, 0, 0, \ldots, 0}_{\mathbf{480}\text{ ตัว}}\Big]^{\mathbf{T}}$$

2) ข้อมูลที่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก (Some small nonzero coefficients)

$$\boldsymbol{\beta} = \Big[0.3, 3, -1, -0.65, 3, -3, 0.7, -1, 3, -0.6, 2, 1, -1, -3, -2, 5, 2, 4, 4, 4, \underbrace{0, \ldots, 0}_{\mathbf{480}\text{ ตัว}}\Big]^{\mathbf{T}}$$

2. จำลองการวิเคราะห์ข้อมูล

การจำลองการวิเคราะห์ข้อมูลมีขั้นตอนดังต่อไปนี้

2.1 สร้างข้อมูลของเมทริกซ์ตัวแปรอิสระ $\mathrm{n} \times \mathrm{p}$

สร้างข้อมูลของเมทริกซ์ตัวแปรอิสระโดยกำหนดให้ข้อมูลเมทริกซ์ตัวแปรอิสระ แต่ละสมาชิกในเมทริกซ์ มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1

2.2 สร้างเวกเตอร์ข้อมูลของความคลาดเคลื่อน

สร้างเวกเตอร์ข้อมูลของความคลาดเคลื่อนโดยกำหนดเวกเตอร์ข้อมูลของความคลาดเคลื่อนให้มีการแจกแจงปกติมีค่าเฉลี่ย 0 และความแปรปรวน $\mathbf{I_n}$

เมื่อ 0 คือเวกเตอร์ 0 ขนาด $\mathrm{n} \times 1$

$\mathbf{I_n}$ คือเมทริกซ์เอกลักษณ์ขนาด $\mathrm{n} \times \mathrm{n}$

2.3 สร้างข้อมูลของเวกเตอร์ตัวแปรตาม

สร้างข้อมูลของเวกเตอร์ตัวแปรตาม จากค่าเมทริกซ์ตัวแปรอิสระ เวกเตอร์ความคลาดที่สร้างมาจากข้างต้น และค่าสัมประสิทธิ์การถดถอยที่ถูกกำหนดจากขอบเขตของวิจัยตามสมการดังนี้

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

3. วิธีสุ่มบูตสแตรปแบบไม่ใช้พารามิเตอร์กระทำซ้ำ 1,000 ครั้ง

วิธีสุ่มบูตสแตรปแบบไม่ใช้พารามิเตอร์สามารถทำได้คือสุ่มตัวแปรตามและตัวแปรอิสระการสุ่มส่วนเหลือ และการสุ่มค่าถ่วงน้ำหนัก

*3.1 การสุ่มตัวแปรตามและตัวแปรอิสระ*

วิธีสุ่มตัวแปรตามและตัวแปรอิสระ จะเป็นการสุ่มจากกลุ่มตัวอย่างแบบใส่กลับคืน กล่าวคือข้อมูลที่เกิดจากการสุ่มของกลุ่มตัวอย่างที่ได้มาในแต่ละการสุ่มแต่ละครั้งจะมีโอกาสได้ข้อมูลซ้ำซึ่งจะทำการสุ่ม 1,000 ครั้ง เท่ากับขนาดของกลุ่มตัวอย่าง จะเห็นว่าข้อมูลของกลุ่มตัวอย่างทั้งหมดในแต่ละข้อมูลจะให้ผลของการถ่วงน้ำหนักที่แตกต่างกัน ซึ่งในบางครั้งข้อมูลบางข้อมูลอาจจะไม่ได้ถูกสุ่มขึ้นมา โดยวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีขั้นตอนดังนี้

1) สุ่มตัวแปรตาม ($\mathrm{y_i}$) และตัวแปรอิสระ ($\mathrm{x_i}$)จากกลุ่มตัวอย่าง จะได้ $\mathrm{y_i}^*$ และ $\mathrm{x_i}^*$ โดยมีขนาดเท่ากับกลุ่มตัวอย่าง

2) นำ $\mathrm{y_i}^*$ และ $\mathrm{x_i}^*$ ประมาณค่าสัมประสิทธิ์ความถดถอย ($\widehat{\boldsymbol{\beta}}$)

3) ใช้ตัวอย่างของข้อมูลเดิมทำซ้ำตามขั้นตอน 1-2 จำนวน 1,000 ครั้ง

4) ได้ค่าสัมประสิทธิ์การถดถอยในแต่ละครั้งดังนี้

$$\widehat{\boldsymbol{\beta}}^* =, \left[\widehat{\boldsymbol{\beta}}^{(1)}, \widehat{\boldsymbol{\beta}}^{(2)}, \widehat{\boldsymbol{\beta}}^{(3)}, \ldots, \widehat{\boldsymbol{\beta}}^{(1{,}000)}\right]$$

เมื่อ $\widehat{\boldsymbol{\beta}}^{(\mathbf{b})}$ คือสัมประสิทธิ์การถดถอยจากตัวอย่างบูตสแตรปครั้งที่ b

*3.2 การสุ่มส่วนเหลือ*

วิธีสุ่มส่วนเหลือเป็นการสุ่มส่วนเหลือที่ได้จากสมการถดถอยเชิงเส้นที่มาจากกลุ่มตัวอย่าแบบใส่กลับคืน กล่าวคือมีโอกาสที่จะได้ส่วนเหลือที่ซ้ำกันซึ่งอาจจะทำให้ค่า $\hat{y}_i^*$ มีค่าเท่าเดิมหรือไม่เท่าเดิมก็ได้ แต่มีโอกาสค่อนข้างยากกว่า วิธีสุ่มตัวแปรตามและตัวแปรอิสระที่จะได้ $\hat{y}_i^*$ มีค่าเท่าเดิมและจะทำการสุ่ม 1,000 ครั้งเท่ากับขนาดของกลุ่มตัวอย่าง โดยวิธีสุ่มส่วนเหลือมีขั้นตอนดังนี้

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 (4.2.1)

$$e_i = y_i - \hat{y}_i \tag{4.2.1}$$

4) นำส่วนเหลือ$(e_i)$ ทำการสุ่มได้ $e_i^*$

5) นำ $e_i^*$ ย้อนกลับไปแทนลงในสมการที่ (4.2.2) จะได้ $y_i^*$

$$y_i^* = \hat{y}_i + e_i^* \tag{4.2.2}$$

6) นำ $y_i^*$ และ $x_i$ ประมาณค่าสัมประสิทธิ์ความถดถอย $(\widehat{\boldsymbol{\beta}})$

7) นำตัวอย่างของข้อมูลเดิมทำซ้ำตามขั้นตอน 2-6 จำนวน 1,000 ครั้ง

8) ได้ค่าสัมประสิทธิ์การถดถอยในแต่ละครั้งดังนี้

$$\widehat{\boldsymbol{\beta}}^* =, [\widehat{\boldsymbol{\beta}}^{(1)}, \widehat{\boldsymbol{\beta}}^{(2)}, \widehat{\boldsymbol{\beta}}^{(3)}, \dots, \widehat{\boldsymbol{\beta}}^{(1,000)}]$$

เมื่อ $\widehat{\boldsymbol{\beta}}^{(\mathbf{b})}$ คือสัมประสิทธิ์การถดถอยจากตัวอย่างบูตสแตรปครั้งที่ b

*3.3 การสุ่มค่าถ่วงน้ำหนัก*

จาก Meeker et al. (2017) วิธีสุ่มตัวแปรตามและตัวแปรอิสระ และวิธีสุ่มส่วนเหลือมีโอกาสที่จะทำให้การสุ่มข้อมูลจากกลุ่มตัวอย่างบางข้อมูลไม่ถูกนำมาใช้ในตัวแบบ ดังนั้นวิธีสุ่มค่าถ่วงน้ำหนักจะช่วยแก้ปัญหาดังกล่าวโดยวิธีสุ่มค่าถ่วงน้ำหนัก มีขั้นตอนดังนี้

1) สุ่มค่าถ่วงน้ำหนักจากการแจกแจงแบบเอกซ์โปเนนเชียล

นั่นคือ $w_i \sim \mathrm{Exp}(1)$

จะได้ $w_1, w_2, \dots, w_p$ เมื่อ $j = 1, 2, \dots, p$

2) นำ $\mathbf{y_i}, \mathbf{x_i}$ และน้ำหนักที่สุ่มได้ประมาณค่าสัมประสิทธิ์ความถดถอย ($\widehat{\boldsymbol{\beta}}$) โดยใช้ Weighted Regression

3) นำตัวอย่างของข้อมูลเดิมทำซ้ำตามขั้นตอน 1-2 จำนวน 1,000 ครั้ง

4) ได้ค่าสัมประสิทธิ์การถดถอยในแต่ละครั้งดังนี้

$$\widehat{\boldsymbol{\beta}}^{*} =, [\widehat{\boldsymbol{\beta}}^{(1)}, \widehat{\boldsymbol{\beta}}^{(2)}, \widehat{\boldsymbol{\beta}}^{(3)}, \ldots, \widehat{\boldsymbol{\beta}}^{(1,000)}]$$

เมื่อ $\widehat{\boldsymbol{\beta}}^{(\mathbf{b})}$ คือสัมประสิทธิ์การถดถอยจากตัวอย่างบูตสแตรปครั้งที่ b

4. การประมาณค่าพารามิเตอร์

การประมาณค่าสัมประสิทธิ์การถดถอยจะใช้การประมาณค่าด้วยกันทั้งหมด 3 วิธี คือการประมาณค่าโดยกำลังสองน้อยที่สุด การประมาณค่าโดยแลซโซ และการประมาณค่าโดยแลซโซปรับได้ สำหรับข้อมูลที่มีมิติสูงจะไม่สามารถประมาณค่าโดยกำลังสองน้อยที่สุดได้

*4.1 การประมาณค่าโดยกำลังสองน้อยที่สุด*

*4.2 การประมาณค่าโดยแลซโซ*

*4.3 การประมาณค่าโดยแลซโซปรับได้*

5. วิธีสร้างช่วงความเชื่อมั่น

*5.1 Simple Percentile Bootstrap*

*5.2 Nonparametric Bias-corrected and Accelerated (BCa) Percentile*

*5.3 Nonparametric Bias-corrected (BC) Percentile*

6. การเปรียบเทียบผลการวิเคราะห์

*6.1 ค่าเฉลี่ยเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง*

*6.2 ค่าเฉลี่ยความกว้างช่วงความเชื่อมั่น*

*6.3 ค่าเฉลี่ยของอัตราผลบวกเทียม*

*6.4 ค่าเฉลี่ยของอัตราผลลบเทียม*



7. สรุปผลวิจัยในแต่ละสถานการณ์

ทำการสรุปจากการเปรียบเทียบสำหรับแต่ละวิธีการสุ่มบูตสแตรปแบบไม่ใช้พารามิเตอร์ว่าการสุ่มแบบใดมีประสิทธิภาพดีที่สุดในแต่ละสถานการณ์นั้น ๆ

# บทที่ 4
# ผลการวิเคราะห์ข้อมูล

ผลการวิจัยจะนำเสนอในรูปตารางดังนั้นเพื่อสะดวกและง่ายต่อการอภิปราย จะใช้สัญลักษณ์ต่อไปนี้แทนความหมายต่าง ๆ ดังนี้

random หมายถึง วิธีการสุ่ม

method of CI หมายถึง วิธีการสร้างช่วงความเชื่อมั่น

Type หมายถึง ชนิดของสัมประสิทธิ์การถดถอยที่ถูกกำหนด

n หมายถึง ขนาดตัวอย่าง

p หมายถึง จำนวนตัวแปรอิสระ

RXY หมายถึง วิธีสุ่มตัวแปรตามและตัวแปรอิสระ

RR หมายถึง วิธีสุ่มส่วนเหลือ

RW หมายถึง วิธีสุ่มค่าถ่วงน้ำหนัก

NSN หมายถึง ข้อมูลที่ไม่มีค่าสัมประสิทธิ์การถดถอยค่าน้อยที่ไม่ใช่ศูนย์ขนาดเล็ก (No small nonzero coefficients)

SSN หมายถึง ข้อมูลที่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก (Some small nonzero coefficients)

SPB หมายถึง วิธี Simple Percentile Bootstrap

BCaB หมายถึง วิธี Nonparametric Bias-corrected and Accelerated (BCa) Percentile Boostrap

BCB หมายถึง วิธี Nonparametric Bias-corrected (BC) Percentile Boostrap

OLS หมายถึง ประมาณค่าโดยกำลังสองน้อยสุด (Ordinary Least Squares Method)

LS หมายถึง ประมาณค่าโดยแลซโซ (Lasso)

ALS หมายถึง ประมาณค่าโดยแลซโซปรับได้ (Adaptive Lasso)

MEAN หมายถึง ค่าเฉลี่ย

SD หมายถึง ส่วนเบี่ยงเบนมาตรฐาน

ผลการวิเคราะห์จะแบ่งโดยจากการววิเคราะห์จำลองข้อมูลทั้งในกรณีที่ข้อมูลขนาดตัวอย่างมากกว่าจำนวนตัวแปรอิสระหรือเรียกว่าข้อมูลที่มีมิติต่ำ และกรณีที่ข้อมูลขนาดตัวอย่างน้อยกว่าจำนวนตัวแปรอิสระ หรือเรียกว่าข้อมูลที่มีมิติสูง

**4.1 ผลการวิเคราะห์ข้อมูลขนาดตัวอย่างมากกว่าจำนวนตัวแปรอิสระ (ข้อมูลที่มีมิติต่ำ)**

ผลการวิเคราะห์จากการวิเคราะห์จำลองข้อมูลในกรณีที่ข้อมูลขนาดตัวอย่างมากกว่าจำนวนตัวแปรอิสระ จะแบ่งย่อยซึ่งได้แก่ ข้อมูลที่ไม่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก และข้อมูลที่มีค่าสัมประสิทธิ์การถดถอยไม่ใช่ศูนย์ขนาดเล็ก

4.1.1 ข้อมูลที่ไม่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก (No small nonzero coefficients)

ผลการวิเคราะห์จากการวิเคราะห์จำลองข้อมูลที่ไม่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก (No small nonzero coefficients) จะสามารถเปรียบเทียบประสิทธิภาพวิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มตัวแปรตามและตัวแปรอิสระ วิธีสุ่มส่วนเหลือ และวิธีสุ่มตัวค่าถ่วงน้ำหนัก โดยใช้ค่าเฉลี่ยเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง ค่าเฉลี่ยความกว้างช่วงความเชื่อมั่น ค่าเฉลี่ยของอัตราผลบวกเทียม และค่าเฉลี่ยของอัตราผลลบเทียมดั้งนี้

1) การเปรียบเทียบค่าเฉลี่ยเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง

สำหรับการเปรียบเทียบค่าเฉลี่ยเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริงถ้ามีค่าเข้าใกล้ระดับความเชื่อมั่นที่ 95% วิธีสุ่มนั้นจะมีประสิทธิภาพดีที่สุด ผลการวิเคราะห์ดังตารางต่อไปนี้



| random | Type | method of CI | OLS | | LS | | ALS | |
|---|---|---|---|---|---|---|---|---|
| | | | MEAN | SD | MEAN | SD | MEAN | SD |
| RXY | NSN | SPB | **54.70** | 29.0838 | **57.05** | 31.0432 | **57.50** | 30.4635 |
| RR | | | 52.10 | 27.9987 | 56.50 | 32.3373 | 56.85 | 31.9773 |
| RW | | | 51.70 | 27.7908 | 56.95 | 31.4457 | 57.25 | 30.9151 |
| RXY | | BCaB | **55.10** | 29.3961 | 56.45 | 31.4248 | 57.05 | 31.6315 |
| RR | | | 51.30 | 27.9123 | **56.90** | 32.2457 | 57.20 | 32.2029 |
| RW | | | 51.45 | 27.7552 | 56.80 | 31.7305 | **57.35** | 31.9442 |
| RXY | | BCB | **54.90** | 29.1892 | 56.55 | 31.3637 | 57.10 | 31.5442 |
| RR | | | 52.60 | 28.2885 | **56.90** | 32.2457 | 57.20 | 32.2029 |

| RW | | | 51.55 | 28.0124 | 56.75 | 31.8168 | **57.35** | 31.9442 |
|---|---|---|---|---|---|---|---|---|

**ตารางที่ 4.1** ตารางการเปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริงสำหรับข้อมูลที่มีมิติต่ำไม่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก

จากตารางที่ 4.1 เลขตัวหนาในตารางแสดงถึงวิธีสุ่มที่ดีที่สุด ผลการวิเคราะห์ตารางที่ 4.1 เป็นดังนี้

การสร้างช่วงความเชื่อมั่นโดยวิธี Simple Percentile Bootstrap ซึ่งการประมาณค่าโดยกำลังสองน้อยสุด การประมาณค่าโดยแลซโซ และการประมาณค่าโดยแลซโซปรับได้ ผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected and Accelerated (BCa) Percentile Boostrap ซึ่งการประมาณค่าโดยกำลังสองน้อยสุดผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด การประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดย วิธี Nonparametric Bias-corrected (BC) Percentile Boostrap ซึ่งการประมาณค่าโดยกำลังสองน้อยสุดผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด การประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด

2) การเปรียบเทียบค่าเฉลี่ยความกว้างช่วงความเชื่อมั่น

สำหรับการเปรียบเทียบค่าเฉลี่ยความกว้างช่วงความเชื่อมั่นถ้ามีค่าน้อยที่สุดวิธีสุ่มนั้นจะมีประสิทธิภาพดีที่สุด ผลการวิเคราะห์ดังตารางต่อไปนี้

| random | Type | method of CI | OLS | | LS | | ALS | |
|---|---|---|---|---|---|---|---|---|
| | | | MEAN | SD | MEAN | SD | MEAN | SD |
| RXY | NSN | SPB | 0.4708 | 0.0659 | 0.2416 | 0.1401 | 0.1917 | 0.1826 |
| RR | | | 0.3904 | 0.0458 | **0.1661** | 0.1665 | 0.1581 | 0.1649 |
| RW | | | **0.3760** | 0.0532 | 0.1871 | 0.1716 | **0.1430** | 0.2018 |
| RXY | | BCaB | 0.4738 | 0.0668 | 0.2823 | 0.1413 | 0.2629 | 0.1666 |
| RR | | | 0.3780 | 0.0545 | **0.2419** | 0.1211 | 0.2339 | 0.1135 |
| RW | | | **0.3779** | 0.0545 | 0.2268 | 0.1663 | **0.1918** | 0.1871 |
| RXY | | BCB | 0.4709 | 0.0659 | 0.2683 | 0.1392 | 0.2562 | 0.1662 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| RR | | | 0.3901 | 0.0458 | 0.2353 | 0.1216 | 0.2280 | 0.1148 |
| RW | | | **0.3758** | 0.0537 | **0.2191** | 0.1652 | **0.1887** | 0.1870 |

**ตารางที่ 4.2** ตารางการเปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความกว้างช่วงความเชื่อมั่นสำหรับข้อมูลที่มีมิติต่ำไม่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก

จากตารางที่ 4.2 เลขตัวหนาในตารางแสดงถึงวิธีสุ่มที่ดีที่สุด ผลการวิเคราะห์ตารางที่ 4.2 เป็นดังนี้

การสร้างช่วงความเชื่อมั่นโดยวิธี Simple Percentile Bootstrap ซึ่งการประมาณค่าโดยกำลังสองน้อยสุด และการประมาณค่าโดยแลซโซปรับได้ ผลปรากฏว่าวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected and Accelerated (BCa) Percentile Boostrap ให้ผลเดียวกันกับ การสร้างช่วงความเชื่อมั่นโดยวิธี Simple Percentile Bootstrap

การสร้างช่วงความเชื่อมั่นโดย วิธี Nonparametric Bias-corrected (BC) Percentile Boostrap ซึ่งการประมาณค่าโดยกำลังสองน้อย แลซโซ และแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด

3) การเปรียบเทียบค่าเฉลี่ยของอัตราผลบวกเทียม (False positive rate)

สำหรับการเปรียบเทียบค่าเฉลี่ยของอัตราผลบวกเทียม (False positive rate) ถ้ามีค่าเข้าใกล้ 0 วิธีสุ่มนั้นจะมีประสิทธิภาพที่ดีสุด ผลการวิเคราะห์ดังตารางต่อไปนี้



| random | Type | method of CI | OLS | | LS | | ALS | |
|---|---|---|---|---|---|---|---|---|
| | | | MEAN | SD | MEAN | SD | MEAN | SD |
| RXY | NSN | SPB | **0.1168** | 0.1256 | 0.0033 | 0.0234 | **0.0000** | 0.0000 |
| RR | | | 0.2011 | 0.1410 | **0.0000** | 0.0000 | **0.0000** | 0.0000 |
| RW | | | 0.2166 | 0.1422 | **0.0000** | 0.0000 | **0.0000** | 0.0000 |
| RXY | | BCaB | **0.1004** | 0.1197 | 0.0195 | 0.0569 | **0.0000** | 0.0000 |
| RR | | | 0.2182 | 0.1496 | **0.0000** | 0.0000 | **0.0000** | 0.0000 |
| RW | | | 0.2161 | 0.1489 | 0.0083 | 0.0365 | **0.0000** | 0.0000 |
| RXY | | BCB | **0.1077** | 0.1249 | 0.0178 | 0.0550 | **0.0000** | 0.0000 |
| RR | | | 0.1861 | 0.1439 | **0.0000** | 0.0000 | **0.0000** | 0.0000 |
| RW | | | 0.2119 | 0.1445 | 0.6670 | 0.0328 | **0.0000** | 0.0000 |

**ตารางที่ 4.3** ตารางการเปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของอัตราผลบวกเทียม สำหรับข้อมูลที่มีมิติต่ำไม่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก

จากตารางที่ 4.3 เลขตัวหนาในตารางแสดงถึงวิธีสุ่มที่ดีที่สุด ผลการวิเคราะห์ตารางที่ 4.3 เป็นดังนี้

การสร้างช่วงความเชื่อมั่นโดยวิธี Simple Percentile Bootstrap ซึ่งการประมาณค่าโดยกำลังสองน้อยสุดผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด ประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มส่วนเหลือและวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด และประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มทั้ง 3 วิธีมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected and Accelerated (BCa) Percentile ซึ่งการประมาณค่าโดยกำลังสองน้อยสุดผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด การประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มทั้ง 3 วิธีมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected (BC) Percentile Boostrap ซึ่งการประมาณค่าโดยกำลังสองน้อยสุดผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด การประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยวิธีแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มทั้ง 3 วิธีมีประสิทธิภาพดีที่สุด

4) การเปรียบเทียบค่าเฉลี่ยของอัตราผลลบเทียม (False negative rate)

สำหรับการเปรียบเทียบค่าเฉลี่ยของอัตราผลลบเทียม (False negative rate) ถ้ามีค่าเข้าใกล้ 0 วิธีสุ่มนั้นจะมีประสิทธิภาพที่ดีสุดผลการวิเคราะห์ดังตารางต่อไปนี้



| random | Type | method of CI | OLS | | LS | | ALS | |
|---|---|---|---|---|---|---|---|---|
| | | | MEAN | SD | MEAN | SD | MEAN | SD |
| RXY | NSN | SPB | **0.0000** | 0.0000 | **0.0000** | 0.000 | **0.0000** | 0.0000 |
| RR | | | **0.0000** | 0.0000 | 0.0006 | 0.0062 | 0.0025 | 0.0123 |
| RW | | | **0.0000** | 0.0000 | **0.0000** | 0.0000 | **0.0000** | 0.0000 |
| RXY | | BCaB | **0.0000** | 0.0000 | **0.0000** | 0.0000 | **0.0018** | 0.0107 |
| RR | | | **0.0000** | 0.0000 | 0.0050 | 0.0170 | 0.0087 | 0.0217 |
| RW | | | **0.0000** | 0.0000 | 0.0018 | 0.0107 | 0.0100 | 0.0230 |
| RXY | | BCB | **0.0000** | 0.0000 | **0.0000** | 0.0000 | **0.0018** | 0.0107 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| RR | | | **0.0000** | 0.0000 | 0.0050 | 0.0170 | 0.0087 | 0.0217 |
| RW | | | **0.0000** | 0.0000 | 0.0188 | 0.0107 | 0.0100 | 0.0230 |

**ตารางที่ 4.4** ตารางการเปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของอัตราผลลบเทียมสำหรับข้อมูลที่มีมิติต่ำไม่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก

จากตารางที่ 4.4 เลขหนาในตารางแสดงถึงวิธีสุ่มที่ดีที่สุด ผลการวิเคราะห์ตารางที่ 4.4 เป็นดังนี้

จาการสร้างช่วงความเชื่อมั่นโดยวิธี Simple Percentile Bootstrap ซึ่งการประมาณค่าโดยกำลังสองน้อยสุดผลปรากฏว่าวิธีสุ่มทั้ง 3 วิธีมีประสิทธิภาพดีที่สุด การประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระ และวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระและวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected and Accelerated (BCa) Percentile ซึ่งการประมาณค่าโดยกำลังสองน้อยสุดผลปรากฏว่าวิธีสุ่มทั้ง 3 วิธีมีประสิทธิภาพดีที่สุด การประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด

จากการสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected (BC) Percentile Boostrap ซึ่งการประมาณค่าโดยกำลังสองน้อยสุดผลปรากฏว่าวิธีสุ่มทั้ง 3 วิธีมีประสิทธิภาพดีที่สุด การประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด

4.1.2 ข้อมูลที่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก (Some small nonzero coefficients)

ผลการวิเคราะห์จากการวิเคราะห์จำลองข้อมูลที่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก (Some small nonzero coefficients) จะสามารถเปรียบเทียบประสิทธิภาพวิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มตัวแปรตามและตัวแปรอิสระ วิธีสุ่มส่วนเหลือ และวิธีสุ่มตัวค่าถ่วงน้ำหนัก โดยใช้ค่าเฉลี่ยเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง ค่าเฉลี่ยความกว้างช่วงความเชื่อมั่น ค่าเฉลี่ยของอัตราผลบวกเทียม และค่าเฉลี่ยของอัตราผลลบเทียมดั้งนี้

1) การเปรียบเทียบค่าเฉลี่ยเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง

สำหรับการเปรียบเทียบค่าเฉลี่ยเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริงถ้ามีค่าเข้าใกล้ระดับความเชื่อมั่นที่ 95% วิธีสุ่มนั้นจะมีประสิทธิภาพดีที่สุด ผลการวิเคราะห์ดังตารางต่อไปนี้

| random | Type | method of CI | OLS | | LS | | ALS | |
|---|---|---|---|---|---|---|---|---|
| | | | MEAN | SD | MEAN | SD | MEAN | SD |
| RXY | SSN | SPB | **66.70** | 26.8762 | **66.30** | 32.7495 | **66.10** | 33.2239 |
| RR | | | 63.15 | 26.0405 | 63.80 | 36.9077 | 63.75 | 36.9915 |
| RW | | | 61.80 | 25.5299 | 64.80 | 35.1685 | 64.90 | 35.0251 |
| RXY | | BCaB | **66.60** | 27.0976 | **65.40** | 32.2449 | **66.25** | 33.2375 |
| RR | | | 61.25 | 25.7746 | 64.50 | 35.7354 | 64.50 | 35.8271 |
| RW | | | 61.20 | 25.7034 | 64.65 | 34.3059 | 65.50 | 34.2414 |
| RXY | | BCB | **66.60** | 27.1721 | **65.35** | 32.1867 | **66.20** | 33.2659 |
| RR | | | 62.95 | 26.3915 | 64.50 | 35.7354 | 64.50 | 35.8271 |
| RW | | | 61.40 | 25.6046 | 64.60 | 34.4779 | 65.50 | 34.2414 |

**ตารางที่ 4.5** ตารางการเปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริงสำหรับข้อมูลที่มีมิติต่ำมีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก



จากตารางที่ 4.5 เลขตัวหนาในตารางแสดงถึงวิธีสุ่มที่ดีที่สุด ผลการวิเคราะห์ตารางที่ 4.5 เป็นดังนี้

การสร้างช่วงความเชื่อมั่นโดยวิธี Simple Percentile Bootstrap ซึ่งประมาณค่าโดยกำลังสองน้อยสุด การประมาณค่าโดยแลซโซ และการประมาณค่าโดยแลซโซปรับได้ ผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected and Accelerated (BCa) Percentile Boostrap ซึ่งการประมาณค่าโดยกำลังสองน้อยสุด การประมาณค่าโดยแลซโซ และการประมาณค่าโดยแลซโซปรับได้ ผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดย วิธี Nonparametric Bias-corrected (BC) Percentile Boostrap ซึ่งการประมาณค่าโดยกำลังสองน้อยสุด การประมาณค่าโดยแลซโซ และการประมาณค่าโดยแลซโซปรับได้ ผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด

2) การเปรียบเทียบค่าเฉลี่ยความกว้างช่วงความเชื่อมั่น

สำหรับการเปรียบเทียบค่าเฉลี่ยความกว้างช่วงความเชื่อมั่นถ้ามีค่าน้อยที่สุดวิธีสุ่มนั้นจะมีประสิทธิภาพดีที่สุด ผลการวิเคราะห์ดังตารางต่อไปนี้

| random | Type | method of CI | OLS | | LS | | ALS | |
|---|---|---|---|---|---|---|---|---|
| | | | MEAN | SD | MEAN | SD | MEAN | SD |
| RXY | SSN | SPB | 0.4775 | 0.0675 | 0.1987 | 0.1337 | 0.1628 | 0.1815 |
| RR | | | 0.3973 | 0.0471 | **0.0893** | 0.1614 | **0.0831** | 0.1602 |
| RW | | | **0.3811** | 0.0550 | 0.1339 | 0.1656 | 0.1029 | 0.1881 |
| RXY | | BCaB | 0.4788 | 0.0682 | 0.2428 | 0.1368 | 0.2471 | 0.1712 |
| RR | | | **0.3820** | 0.0551 | 0.1911 | 0.1075 | 0.1868 | 0.1058 |
| RW | | | 0.3821 | 0.0552 | **0.1776** | 0.1536 | **0.1582** | 0.1766 |
| RXY | | BCB | 0.4775 | 0.0678 | 0.2387 | 0.1359 | 0.2453 | 0.1709 |
| RR | | | 0.3968 | 0.0472 | 0.1904 | 0.1078 | 0.1862 | 0.1060 |
| RW | | | **0.3810** | 0.0547 | **0.1761** | 0.1533 | **0.1576** | 0.1765 |

**ตารางที่ 4.6** ตารางการเปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความกว้างช่วงความเชื่อมั่นสำหรับข้อมูลที่มีมิติต่ำมีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก



จากตารางที่ 4.6 เลขตัวหนาในตารางแสดงถึงวิธีสุ่มที่ดีที่สุด ผลการวิเคราะห์ตารางที่ 4.6 เป็นดังนี้

การสร้างช่วงความเชื่อมั่นโดยวิธี Simple Percentile Bootstrap ซึ่งการประมาณค่าโดยกำลังสองน้อยสุดผลปรากฏว่าวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด การประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected and Accelerated (BCa) Percentile Boostrap ซึ่งการประมาณค่าโดยกำลังสองน้อยสุดผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด การประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด และวิธีการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดย วิธี Nonparametric Bias-corrected (BC) Percentile Boostrap ซึ่งการประมาณค่าโดยกำลังสองน้อยสุด การประมาณค่าโดยแลซโซ และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด

3) การเปรียบเทียบค่าเฉลี่ยของอัตราผลบวกเทียม (False positive rate)

สำหรับการเปรียบเทียบค่าเฉลี่ยของอัตราผลบวกเทียม (False positive rate) ถ้ามีค่าเข้าใกล้ 0 วิธีสุ่มนั้นจะมีประสิทธิภาพที่ดีสุด ผลการวิเคราะห์ดังตารางต่อไปนี้

| random | Type | method of CI | OLS | | LS | | ALS | |
|---|---|---|---|---|---|---|---|---|
| | | | MEAN | SD | MEAN | SD | MEAN | SD |
| RXY | SSN | SPB | **0.1298** | 0.1528 | 0.0075 | 0.0428 | **0.0000** | 0.0000 |
| RR | | | 0.2188 | 0.1618 | **0.0000** | 0.0000 | **0.0000** | 0.0000 |
| RW | | | 0.2418 | 0.1710 | 0.0025 | 0.0250 | **0.0000** | 0.0000 |
| RXY | | BCaB | 0.2314 | 0.1421 | 0.0245 | 0.0740 | 0.0050 | 0.0351 |
| RR | | | **0.2310** | 0.1690 | **0.0025** | 0.0250 | **0.0000** | 0.0000 |
| RW | | | 0.2314 | 0.1679 | 0.0045 | 0.0318 | **0.0000** | 0.0000 |
| RXY | | BCB | **0.1138** | 0.1452 | 0.0245 | 0.0740 | 0.0050 | 0.0351 |
| RR | | | 0.1977 | 0.1696 | **0.0025** | 0.0250 | **0.0000** | 0.0000 |
| RW | | | 0.2354 | 0.1608 | 0.0045 | 0.0319 | **0.0000** | 0.0000 |

**ตารางที่ 4.7** ตารางการเปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของอัตราผลบวกเทียม สำหรับข้อมูลที่มีมิติต่ำมีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก



จากตารางที่ 4.7 เลขตัวหนาในตารางแสดงถึงวิธีสุ่มที่ดีที่สุด ผลการวิเคราะห์ตารางที่ 4.7 เป็นดังนี้

การสร้างช่วงความเชื่อมั่นโดยวิธี Simple Percentile Bootstrap ซึ่งการประมาณค่าโดยกำลังสองน้อยสุดผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด การประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มทั้ง 3 วิธีมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected and Accelerated (BCa) Percentile ซึ่งการประมาณค่าโดยกำลังสองน้อยสุดผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด การประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด

และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มส่วนเหลือและวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected (BC) Percentile Boostrap ซึ่งการประมาณค่าโดยกำลังสองน้อยสุดผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด การประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มส่วนเหลือและวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด

4) การเปรียบเทียบค่าเฉลี่ยของอัตราผลลบเทียม (False negative rate)

การเปรียบเทียบสำหรับค่าเฉลี่ยของอัตราผลลบเทียม (False negative rate) ถ้ามีค่าเข้าใกล้ 0 วิธีสุ่มนั้นจะมีประสิทธิภาพที่ดีสุดผลการวิเคราะห์ดังตารางต่อไปนี้

| random | Type | method of CI | OLS | | LS | | ALS | |
|---|---|---|---|---|---|---|---|---|
| | | | MEAN | SD | MEAN | SD | MEAN | SD |
| RXY | SSN | SPB | 0.0946 | 0.0311 | **0.1173** | 0.0056 | **0.1176** | 0.0000 |
| RR | | | 0.0864 | 0.0363 | 0.1176 | 0.0000 | **0.1176** | 0.0000 |
| RW | | | **0.0835** | 0.0376 | 0.1177 | 0.0007 | **0.1176** | 0.0000 |
| RXY | | BCaB | 0.0904 | 0.0305 | **0.1051** | 0.0240 | 0.1166 | 0.0078 |
| RR | | | 0.0814 | 0.0375 | 0.1171 | 0.0055 | 0.1176 | 0.0000 |
| RW | | | **0.0808** | 0.0370 | 0.1105 | 0.0240 | **0.1165** | 0.0077 |
| RXY | | BCB | 0.0917 | 0.0305 | **0.1051** | 0.0240 | 0.1166 | 0.0078 |
| RR | | | 0.0855 | 0.0337 | 0.1171 | 0.0055 | 0.1176 | 0.0000 |
| RW | | | **0.0796** | 0.0375 | 0.1110 | 0.0179 | **0.1165** | 0.0077 |

**ตารางที่ 4.8** ตารางการเปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของอัตราผลลบเทียมสำหรับข้อมูลที่มีมิติต่ำมีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก

จากตารางที่ 4.8 เลขตัวหนาในตารางแสดงถึงวิธีสุ่มดีที่สุด ผลการวิเคราะห์ตารางที่ 4.8 เป็นดังนี้

การสร้างช่วงความเชื่อมั่นโดยวิธี Simple Percentile Bootstrap ซึ่งการประมาณค่าโดยกำลังสองน้อยสุดผลปรากฏว่าวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด การประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มทั้ง 3 วิธีมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected and Accelerated (BCa) Percentile ซึ่งการประมาณค่าโดยกำลังสองน้อยสุดผลปรากฏว่าวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด การประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected (BC) Percentile Boostrap ซึ่งการประมาณค่าโดยกำลังสองน้อยสุดผลปรากฏว่าวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด การประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด

**4.2 ผลการวิเคราะห์ข้อมูลขนาดตัวอย่างน้อยกว่าจำนวนตัวแปรอิสระ (ข้อมูลที่มีมิติสูง)**

4.2.1 ข้อมูลที่ไม่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก (No small nonzero coefficients)

ผลการวิเคราะห์จากการวิเคราะห์จำลองข้อมูลที่ไม่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก (No small nonzero coefficients) จะสามารถเปรียบเทียบประสิทธิภาพวิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มตัวแปรตามและตัวแปรอิสระ วิธีสุ่มส่วนเหลือ และวิธีสุ่มตัวค่าถ่วงน้ำหนัก โดยใช้ค่าเฉลี่ยเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง ค่าเฉลี่ยความกว้างช่วงความเชื่อมั่น ค่าเฉลี่ยของอัตราผลบวกเทียม และค่าเฉลี่ยของอัตราผลลบเทียมดั้งนี้

1) การเปรียบเทียบค่าเฉลี่ยเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง

สำหรับการเปรียบเทียบค่าเฉลี่ยเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริงถ้ามีค่าเข้าใกล้ระดับความเชื่อมั่นที่ 95% วิธีสุ่มนั้นจะมีประสิทธิภาพดีที่สุด ผลการวิเคราะห์ดังตารางต่อไปนี้

| random | Type | method of CI | OLS | | LS | | ALS | |
|---|---|---|---|---|---|---|---|---|
| | | | MEAN | SD | MEAN | SD | MEAN | SD |
| RXY | | | - | - | **95.2340** | 7.9516 | 94.0900 | 7.9889 |
| RR | NSN | SPB | - | - | 92.4660 | 7.8034 | 94.2240 | 8.0493 |
| RW | | | - | - | 93.2140 | 7.9020 | **95.2120** | 7.9709 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| RXY | | BCaB | - | - | **88.8060** | 7.2481 | **89.4220** | 7.5273 |
| RR | | | - | - | 87.3880 | 7.3517 | 88.9520 | 7.7301 |
| RW | | | - | - | 86.8420 | 7.3216 | 89.1680 | 7.3945 |
| RXY | | BCB | - | - | **88.8120** | 7.2505 | **89.4160** | 7.5389 |
| RR | | | - | - | 87.3640 | 7.3611 | 88.9400 | 7.7296 |
| RW | | | - | - | 86.8140 | 7.3124 | 89.1620 | 7.4011 |

**ตารางที่ 4.9** ตารางการเปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริงสำหรับข้อมูลที่มีมิติสูงไม่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก

จากตารางที่ 4.9 สำหรับข้อมูลที่มีมิติสูงการประมาณค่าโดยกำลังสองน้อยสุดจะไม่สามารถประมาณค่าสัมประสิทธิ์การถดถอยได้ และเลขตัวหนาในตารางแสดงถึงวิธีสุ่มที่ดีที่สุด ผลการวิเคราะห์ตารางที่ 4.9 เป็นดังนี้

การสร้างช่วงความเชื่อมั่นโดยวิธี Simple Percentile Bootstrap ซึ่งการประมาณค่าโดยแลซโซวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected and Accelerated (BCa) Percentile Boostrap ซึ่งการประมาณค่าโดยแลซโซ และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดย วิธี Nonparametric Bias-corrected (BC) Percentile Boostrap ซึ่งได้ใช้วิธีการประมาณค่าโดยแลซโซ และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด

2) การเปรียบเทียบค่าเฉลี่ยความกว้างช่วงความเชื่อมั่น

สำหรับการเปรียบเทียบค่าเฉลี่ยความกว้างช่วงความเชื่อมั่นถ้ามีค่าน้อยที่สุดวิธีสุ่มนั้นจะมีประสิทธิภาพดีที่สุด ผลการวิเคราะห์ดังตารางต่อไปนี้

| random | Type | method of CI | OLS | | LS | | ALS | |
|---|---|---|---|---|---|---|---|---|
| | | | MEAN | SD | MEAN | SD | MEAN | SD |
| RXY | NSN | SPB | - | - | 0.7516 | 0.4692 | 0.3500 | 0.0139 |
| RR | | | - | - | **0.1881** | 0.4454 | **0.2503** | 0.6081 |
| RW | | | - | - | 0.3534 | 0.7349 | 0.5647 | 1.3436 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| RXY | | BCaB | - | - | 1.7153 | 1.8203 | **0.4773** | 0.9623 |
| RR | | | - | - | **0.2690** | 0.4381 | 0.5521 | 0.6376 |
| RW | | | - | - | 0.4750 | 0.7593 | 1.0064 | 1.3747 |
| RXY | | BCB | - | - | 1.7152 | 1.8202 | **0.4772** | 0.9621 |
| RR | | | - | - | **0.2689** | 0.4378 | 0.5520 | 0.6374 |
| RW | | | - | - | 0.4748 | 0.7587 | 1.0063 | 1.3746 |

**ตารางที่ 4.10** ตารางการเปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความกว้างช่วงความเชื่อมั่นสำหรับข้อมูลที่มีมิติสูงไม่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก

จากตารางที่ 4.10 สำหรับข้อมูลที่มีมิติสูงการประมาณค่าโดยกำลังสองน้อยสุดจะไม่สามารถประมาณค่าสัมประสิทธิ์การถดถอยได้ และเลขตัวหนาในตารางแสดงถึงวิธีสุ่มที่ดีที่สุด ผลการวิเคราะห์ตารางที่ 4.10 เป็นดังนี้

การสร้างช่วงความเชื่อมั่นโดยวิธี Simple Percentile Bootstrap ซึ่งการประมาณค่าโดยแลซโซวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected and Accelerated (BCa) Percentile ซึ่งการประมาณค่าโดยแลซโซวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดย วิธี Nonparametric Bias-corrected (BC) Percentile Boostrap ซึ่งการประมาณค่าโดยแลซโซวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด

3) การเปรียบเทียบค่าเฉลี่ยของอัตราผลบวกเทียม (False positive rate)

สำหรับการเปรียบเทียบค่าเฉลี่ยของอัตราผลบวกเทียม (False positive rate) ถ้ามีค่าเข้าใกล้ 0 วิธีสุ่มนั้นจะมีประสิทธิภาพที่ดีสุด ผลการวิเคราะห์ดังตารางต่อไปนี้

| random | Type | method of CI | OLS | | LS | | ALS | |
|---|---|---|---|---|---|---|---|---|
| | | | MEAN | SD | MEAN | SD | MEAN | SD |
| RXY | NSN | SPB | - | - | **0.0000** | 0.0000 | **0.0000** | 0.0000 |
| RR | | | - | - | 0.1575 | 0.1042 | 0.0225 | 0.0508 |
| RW | | | - | - | 0.0601 | 0.0662 | 0.0050 | 0.0351 |
| RXY | | BCaB | - | - | 0.8091 | 0.0367 | **0.6549** | 0.0533 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| RR | | | - | - | **0.6494** | 0.0660 | 0.7310 | 0.0452 |
| RW | | | - | - | 0.7115 | 0.0603 | 0.8016 | 0.0375 |
| RXY | | BCB | - | - | 0.8090 | 0.0366 | **0.6552** | 0.0531 |
| RR | | | - | - | **0.6503** | 0.0655 | 0.7315 | 0.0453 |
| RW | | | - | - | 0.7120 | 0.0600 | 0.8018 | 0.0373 |

**ตารางที่ 4.11** ตารางการเปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของอัตราผลบวกเทียม สำหรับข้อมูลที่มีมิติสูงไม่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก

จากตารางที่ 4.11 สำหรับข้อมูลที่มีมิติสูงการประมาณค่าโดยกำลังสองน้อยสุดจะไม่สามารถประมาณค่าสัมประสิทธิ์การถดถอยได้ และเลขตัวหนาในตารางแสดงถึงวิธีสุ่มที่ดีที่สุด ผลการวิเคราะห์ตารางที่ 4.11 เป็นดังนี้

การสร้างช่วงความเชื่อมั่นโดยวิธี Simple Percentile Bootstrap ซึ่งการประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected and Accelerated (BCa) Percentile ซึ่งการประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected (BC) Percentile Boostrap ซึ่งการประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด

4) การเปรียบเทียบค่าเฉลี่ยของอัตราผลลบเทียม (False negative rate)

สำหรับการเปรียบเทียบค่าเฉลี่ยของอัตราผลลบเทียม (False negative rate) ถ้ามีค่าเข้าใกล้ 0 วิธีสุ่มนั้นจะมีประสิทธิภาพที่ดีสุด ผลการวิเคราะห์ดังตารางต่อไปนี้

| random | Type | method of CI | OLS | | LS | | ALS | |
|---|---|---|---|---|---|---|---|---|
| | | | MEAN | SD | MEAN | SD | MEAN | SD |
| RXY | NSN | SPB | - | - | 0.0321 | 0.0028 | **0.0207** | 0.0026 |
| RR | | | - | - | **0.0098** | 0.0081 | 0.0225 | 0.0035 |
| RW | | | - | - | 0.0130 | 0.0072 | 0.0320 | 0.0028 |
| RXY | | BCaB | - | - | 0.0276 | 0.0026 | **0.0177** | 0.0029 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| RR | | | - | - | **0.0090** | 0.0080 | 0.0219 | 0.0041 |
| RW | | | - | - | 0.0138 | 0.0066 | 0.0279 | 0.0024 |
| RXY | | BCB | - | - | 0.0276 | 0.0026 | **0.0177** | 0.0029 |
| RR | | | - | - | **0.0090** | 0.0080 | 0.0219 | 0.0041 |
| RW | | | - | - | 0.0138 | 0.0066 | 0.0279 | 0.0024 |

**ตารางที่ 4.12** ตารางการเปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของอัตราผลลบเทียม สำหรับข้อมูลที่มีมิติสูงไม่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก

จากตารางที่ 4.12 สำหรับข้อมูลที่มีมิติสูงการประมาณค่าโดยกำลังสองน้อยสุดจะไม่สามารถประมาณค่าสัมประสิทธิ์การถดถอยได้ และเลขตัวหนาในตารางแสดงถึงวิธีสุ่มที่ดีที่สุด ผลการวิเคราะห์ตารางที่ 4.12 เป็นดังนี้

การสร้างช่วงความเชื่อมั่นโดยวิธี Simple Percentile Bootstrap ซึ่งการประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected and Accelerated (BCa) Percentile ซึ่งการประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected (BC) Percentile Boostrap ซึ่งการประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด

4.2.2 ข้อมูลที่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก (Some small nonzero coefficients)

ผลการวิเคราะห์จากการวิเคราะห์จำลองข้อมูลที่มีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก (Some small nonzero coefficients) จะสามารถเปรียบเทียบประสิทธิภาพวิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มตัวแปรตามและตัวแปรอิสระ วิธีสุ่มส่วนเหลือ และวิธีสุ่มตัวค่าถ่วงน้ำหนัก โดยใช้ค่าเฉลี่ยเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง ค่าเฉลี่ยความกว้างช่วงความเชื่อมั่น ค่าเฉลี่ยของอัตราผลบวกเทียม และค่าเฉลี่ยของอัตราผลลบเทียมดั้งนี้

1) การเปรียบเทียบค่าเฉลี่ยเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง

สำหรับการเปรียบเทียบค่าเฉลี่ยเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริงถ้ามีค่าเข้าใกล้ระดับความเชื่อมั่นที่ 95% วิธีสุ่มนั้นจะมีประสิทธิภาพดีที่สุด ผลการวิเคราะห์ดังตารางต่อไปนี้

| random | Type | method of CI | OLS | | LS | | ALS | |
|---|---|---|---|---|---|---|---|---|
| | | | MEAN | SD | MEAN | SD | MEAN | SD |
| RXY | SSN | 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 |

**ตารางที่ 4.13** ตารางการเปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริงสำหรับข้อมูลที่มีมิติสูงมีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก



จากตารางที่ 4.13 สำหรับข้อมูลที่มีมิติสูงการประมาณค่าโดยกำลังสองน้อยสุดจะไม่สามารถประมาณค่าสัมประสิทธิ์การถดถอยได้ และเลขตัวหนาในตารางแสดงถึงวิธีสุ่มที่ดีที่สุด ผลการวิเคราะห์ตารางที่ 4.13 เป็นดังนี้

การสร้างช่วงความเชื่อมั่นโดยวิธี Simple Percentile Bootstrap ซึ่งการประมาณค่าโดยแลซโซวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected and Accelerated (BCa) Percentile Boostrap ซึ่งการประมาณค่าโดยแลซโซได้ผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดย วิธี Nonparametric Bias-corrected (BC) Percentile Boostrap ซึ่งการประมาณค่าโดยแลซโซการประมาณค่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด

2) การเปรียบเทียบค่าเฉลี่ยความกว้างช่วงความเชื่อมั่น

สำหรับการเปรียบเทียบค่าเฉลี่ยความกว้างช่วงความเชื่อมั่นถ้ามีค่าน้อยที่สุดวิธีสุ่มนั้นจะมีประสิทธิภาพดีที่สุด ผลการวิเคราะห์ดังตารางต่อไปนี้

| random | Type | method of CI | OLS | | LS | | ALS | |
|---|---|---|---|---|---|---|---|---|
| | | | MEAN | SD | MEAN | SD | MEAN | SD |
| RXY | SSN | SPB | - | - | 0.2393 | 0.4750 | 0.2502 | 0.5806 |
| RR | | | - | - | **0.0696** | 0.1534 | **0.1134** | 0.3920 |
| RW | | | - | - | 0.1030 | 0.2058 | 0.1754 | 0.4049 |
| RXY | | BCaB | - | - | 0.5493 | 0.5772 | 0.3577 | 0.6227 |
| RR | | | - | - | **0.1139** | 0.1508 | 0.4583 | 0.5247 |
| RW | | | - | - | 0.1364 | 0.2169 | **0.3155** | 0.4121 |
| RXY | | BCB | - | - | 0.5493 | 0.5772 | 0.3577 | 0.6226 |
| RR | | | - | - | **0.1138** | 0.1508 | 0.4583 | 0.5247 |
| RW | | | - | - | 0.1363 | 0.2168 | **0.3154** | 0.4120 |

ตารางที่ 4.14 ตารางการเปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความกว้างช่วงความเชื่อมั่นสำหรับข้อมูลที่มีมิติสูงมีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก

จากตารางที่ 4.14 สำหรับข้อมูลที่มีมิติสูงการประมาณค่าโดยกำลังสองน้อยสุดจะไม่สามารถประมาณค่าสัมประสิทธิ์การถดถอยได้ และเลขตัวหนาในตารางแสดงถึงวิธีสุ่มที่ดีที่สุด ผลการวิเคราะห์ตารางที่ 4.14 เป็นดังนี้

การสร้างช่วงความเชื่อมั่นโดยวิธี Simple Percentile Bootstrap ซึ่งการประมาณค่าโดยแลซโซวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected and Accelerated (BCa) Percentile ซึ่งการประมาณค่าโดยแลซโซวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected (BC) Percentile Boostrap ซึ่งการประมาณค่าโดยแลซโซวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด

3) การเปรียบเทียบค่าเฉลี่ยผลบวกเทียม (False positive rate)

สำหรับการเปรียบเทียบค่าเฉลี่ยผลบวกเทียม (False positive rate) ถ้ามีค่าเข้าใกล้ 0 วิธีสุ่มนั้นจะมีประสิทธิภาพที่ดีสุด ผลการวิเคราะห์ดังตารางต่อไปนี้

| random | Type | method of CI | OLS | | LS | | ALS | |
|---|---|---|---|---|---|---|---|---|
| | | | MEAN | SD | MEAN | SD | MEAN | SD |
| RXY | SSN | 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 |

**ตารางที่ 4.15** ตารางการเปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของอัตราผลบวกเทียมสำหรับข้อมูลที่มีมิติสูงมีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก



จากตารางที่ 4.15 สำหรับข้อมูลที่มีมิติสูงการประมาณค่าโดยกำลังสองน้อยสุดจะไม่สามารถประมาณค่าสัมประสิทธิ์การถดถอยได้ และเลขตัวหนาในตารางแสดงถึงวิธีสุ่มที่ดีที่สุด ผลการวิเคราะห์ตารางที่ 4.15 เป็นดังนี้

การคำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของอัตราผลบวกเทียมสำหรับข้อมูลที่มีมิติสูงมีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็กพบว่า การสร้างช่วงความเชื่อมั่นโดยวิธี Simple Percentile Bootstrap โดยการประมาณค่าโดยแลซโซสำหรับวิธีสุ่มตัวแปรตามและตัวแปรอิสระขนาด 100 ตัวอย่าง ผลปรากฏว่า มีบางตัวอย่างที่ทำให้ค่าของอัตราผลบวกเทียมไม่สามารถคำนวณค่าได้โดยมีค่าอัตราผลบวกเทียมเป็นศูนย์ส่วนศูนย์ ทางผู้วิจัยจะทำการตัดตัวอย่างนั้นทิ้งไปและคำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานที่ค่าไม่เป็น ศูนย์ส่วนศูนย์เท่านั้น และยังพบอีกว่าการสร้างช่วงความเชื่อมั่นโดยวิธี Simple Percentile Bootstrap โดยการประมาณค่าโดยแลซโซปรับได้

สำหรับวิธีสุ่มส่วนเหลือเกิดปัญหาเช่นเดียวกันกับการสร้างช่วงความเชื่อมั่นโดยวิธี Simple Percentile Bootstrap โดยการประมาณค่าโดยแลซโซสำหรับวิธีสุ่มตัวแปรตามและตัวแปรอิสระ

การสร้างช่วงความเชื่อมั่นโดยวิธี Simple Percentile Bootstrap ซึ่งการประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected and Accelerated (BCa) Percentile ซึ่งการประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected (BC) Percentile Boostrap ซึ่งการประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มส่วนเหลือมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด

4) การเปรียบเทียบค่าเฉลี่ยของอัตราผลลบเทียม (False negative rate)

สำหรับการเปรียบเทียบค่าเฉลี่ยของอัตราผลลบเทียม (False negative rate) ถ้ามีค่าเข้าใกล้ 0 วิธีสุ่มนั้นจะมีประสิทธิภาพที่ดีสุด ผลการวิเคราะห์ดังตารางต่อไปนี้

| random | Type | method | OLS | | LS | | ALS | |
|---|---|---|---|---|---|---|---|---|
| | | of CI | MEAN | SD | MEAN | SD | MEAN | SD |
| RXY | SSN | 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 |

**ตารางที่ 4.16** ตารางการเปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของอัตราผลลบเทียม สำหรับข้อมูลที่มีมิติสูงมีค่าสัมประสิทธิ์การถดถอยที่ไม่ใช่ศูนย์ขนาดเล็ก

จากตารางที่ 4.16 สำหรับข้อมูลที่มีมิติสูงการประมาณค่าโดยกำลังสองน้อยสุดจะไม่สามารถประมาณค่าสัมประสิทธิ์การถดถอยได้ และเลขตัวหนาในตารางแสดงถึงวิธีสุ่มที่ดีที่สุด ผลการวิเคราะห์ตารางที่ 4.16 เป็นดังนี้

การสร้างช่วงความเชื่อมั่นโดยวิธี Simple Percentile Bootstrap ซึ่งการประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected and Accelerated (BCa) Percentile ซึ่งการประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด

การสร้างช่วงความเชื่อมั่นโดยวิธี Nonparametric Bias-corrected (BC) Percentile Boostrap ซึ่งการประมาณค่าโดยแลซโซผลปรากฏว่าวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพดีที่สุด และการประมาณค่าโดยแลซโซปรับได้ผลปรากฏว่าวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพดีที่สุด

# บทที่ 5
# สรุปผลการวิจัย และข้อเสนอแนะ

## 5.1 สรุปผลการวิจัย

สำหรับสรุปผลการวิจัยจะทำการสรุปผลโดยแบ่งตามวิธีการเปรียบเทียบซึ่งได้แก่ ค่าเฉลี่ยเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง ค่าเฉลี่ยความกว้างช่วงความเชื่อมั่น ค่าเฉลี่ยของอัตราผลบวกเทียม และค่าเฉลี่ยของอัตราผลลบเทียมผลปรากฏว่า

1) การเปรียบเทียบโดยค่าเฉลี่ยเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง

สำหรับการเปรียบเทียบโดยค่าเฉลี่ยเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริงถ้ามีค่าเข้าใกล้ระดับความเชื่อมั่นที่ 95% วิธีสุ่มนั้นจะมีประสิทธิภาพดีที่สุด ผลสรุปได้ดังตารางต่อไปนี้

| random | n:p | method of CI | Type | OLS | | LS | | ALS | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | MEAN | SD | MEAN | SD | MEAN | SD |
| RXY | 100:20 | SPB | NSN | **54.70** | 29.0838 | **57.05** | 31.0432 | **57.50** | 30.4635 |
| RR | | | | 52.10 | 27.9987 | 56.50 | 32.3373 | 56.85 | 31.9773 |
| RW | | | | 51.70 | 27.7908 | 56.95 | 31.4457 | 57.25 | 30.9151 |
| RXY | | BCaB | | **55.10** | 29.3961 | 56.45 | 31.4248 | 57.05 | 31.6315 |
| RR | | | | 51.30 | 27.9123 | **56.90** | 32.2457 | 57.20 | 32.2029 |
| RW | | | | 51.45 | 27.7552 | 56.80 | 31.7305 | **57.35** | 31.9442 |
| RXY | | BCB | | **54.90** | 29.1892 | 56.55 | 31.3637 | 57.10 | 31.5442 |
| RR | | | | 52.60 | 28.2885 | **56.90** | 32.2457 | 57.20 | 32.2029 |
| RW | | | | 51.55 | 28.0124 | 56.75 | 31.8168 | **57.35** | 31.9442 |
| RXY | | SPB | SSN | **66.70** | 26.8762 | **66.30** | 32.7495 | **66.10** | 33.2239 |
| RR | | | | 63.15 | 26.0405 | 63.80 | 36.9077 | 63.75 | 36.9915 |
| RW | | | | 61.80 | 25.5299 | 64.80 | 35.1685 | 64.90 | 35.0251 |
| RXY | | BCaB | | **66.60** | 27.0976 | **65.40** | 32.2449 | **66.25** | 33.2375 |
| RR | | | | 61.25 | 25.7746 | 64.50 | 35.7354 | 64.50 | 35.8271 |
| RW | | | | 61.20 | 25.7034 | 64.65 | 34.3059 | 65.50 | 34.2414 |
| RXY | | BCB | | **66.60** | 27.1721 | **65.35** | 32.1867 | **66.20** | 33.2659 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| RR | | | | 62.95 | 26.3915 | 64.50 | 35.7354 | 64.50 | 35.8271 |
| RW | | | | 61.40 | 25.6046 | 64.60 | 34.4779 | 65.50 | 34.2414 |
| RXY | 100:500 | SPB | NSN | - | - | **95.23** | 7.9516 | 94.09 | 7.9889 |
| RR | | | | - | - | 92.46 | 7.8034 | 94.22 | 8.0493 |
| RW | | | | - | - | 93.21 | 7.9020 | **95.21** | 7.9709 |
| RXY | | BCaB | | - | - | **88.80** | 7.2481 | 89.42 | 7.5273 |
| RR | | | | - | - | 87.38 | 7.3517 | 88.95 | 7.7301 |
| RW | | | | - | - | 86.84 | 7.3216 | 89.16 | 7.3945 |
| RXY | | BCB | | - | - | **88.81** | 7.2505 | **89.41** | 7.5389 |
| RR | | | | - | - | 87.36 | 7.3611 | 88.94 | 7.7296 |
| RW | | | | - | - | 86.81 | 7.3124 | 89.16 | 7.4011 |
| RXY | | SPB | SSN | - | - | **95.33** | 7.7135 | 95.45 | 0.6735 |
| RR | | | | - | - | 93.09 | 7.8351 | 95.80 | 7.8782 |
| RW | | | | - | - | 93.24 | 7.7636 | **94.98** | 7.7549 |
| RXY | | BCaB | | - | - | **88.70** | 6.9260 | 88.74 | 6.9959 |
| RR | | | | - | - | 87.36 | 7.3610 | **91.47** | 7.2575 |
| RW | | | | - | - | 86.21 | 7.2234 | 88.49 | 7.1519 |
| RXY | | BCB | | - | - | **88.70** | 6.9254 | 88.74 | 6.9935 |
| RR | | | | - | - | 87.33 | 7.3577 | **91.47** | 7.2575 |
| RW | | | | - | - | 86.16 | 7.2192 | 88.48 | 7.1480 |

**ตารางที่ 5.1** ตารางการเปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตราฐานของค่าเปอร์เซ็นต์ที่ช่วงความเชื่อมั่นที่ครอบคลุมค่าสัมประสิทธิ์การถดถอยค่าจริง



จากตารางที่ 5.1 สำหรับข้อมูลที่มีมิติสูงการประมาณค่าโดยกำลังสองน้อยสุดจะไม่สามารถประมาณค่าสัมประสิทธิ์การถดถอยได้ และเลขตัวหนาในตารางแสดงถึงวิธีสุ่มที่ดีที่สุด ผลการวิเคราะห์ตารางที่ 5.1 ได้ดังนี้

1. การประมาณค่าโดยกำลังสองน้อยที่สุดพบว่า วิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพสูงที่สุด วิธีสุ่มส่วนเหลือรองลงมา และวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพต่ำที่สุด

2. การประมาณค่าโดยแลซโซพบว่า วิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพสูงที่สุด วิธีสุ่มส่วนเหลือรองลงมา และวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพต่ำที่สุด

3. การประมาณค่าโดยแลซโซปรับได้พบว่า วิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มตัวแปรตามและตัวแปรอิสระสูงที่สุด และวิธีสุ่มส่วนเหลือและวิธีสุ่มค่าถ่วงน้ำหนัก มีประสิทธิภาพรองลงมาและใกล้เคียงกัน

โดยรวมแล้ววิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพสูงที่สุด วิธีสุ่มส่วนเหลือรองลงมา และวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพต่ำที่สุด

2) การเปรียบเทียบค่าเฉลี่ยความกว้างช่วงความเชื่อมั่น

สำหรับการเปรียบเทียบค่าเฉลี่ยความกว้างช่วงความเชื่อมั่นถ้ามีค่าน้อยที่สุดวิธีสุ่มนั้นจะมีประสิทธิภาพดีที่สุด ผลสรุปได้ดังตารางต่อไปนี้

| random | n:p | method of CI | Type | OLS | | LS | | ALS | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | MEAN | SD | MEAN | SD | MEAN | SD |
| RXY | 100:20 | SPB | NSN | 0.4708 | 0.0659 | 0.2416 | 0.1401 | 0.1917 | 0.1826 |
| RR | | | | 0.3904 | 0.0458 | **0.1661** | 0.1665 | 0.1581 | 0.1649 |
| RW | | | | **0.3760** | 0.0532 | 0.1871 | 0.1716 | **0.143** | 0.2018 |
| RXY | | BCaB | | 0.4738 | 0.0668 | 0.2823 | 0.1413 | 0.2629 | 0.1666 |
| RR | | | | 0.3780 | 0.0545 | 0.2419 | 0.1211 | 0.2339 | 0.1135 |
| RW | | | | **0.3779** | 0.0545 | **0.2268** | 0.1663 | **0.1918** | 0.1871 |
| RXY | | BCB | | 0.4709 | 0.0659 | 0.2683 | 0.1392 | 0.2562 | 0.1662 |
| RR | | | | 0.3901 | 0.0458 | 0.2353 | 0.1216 | 0.2280 | 0.1148 |
| RW | | | | **0.3758** | 0.0537 | **0.2191** | 0.1652 | **0.1887** | 0.1870 |
| RXY | | SPB | SSN | 0.4775 | 0.0675 | 0.1987 | 0.1337 | 0.1628 | 0.1815 |
| RR | | | | 0.3973 | 0.0471 | **0.0893** | 0.1614 | **0.0831** | 0.1602 |
| RW | | | | **0.3811** | 0.0550 | 0.1339 | 0.1656 | 0.1029 | 0.1881 |
| RXY | | BCaB | | 0.4788 | 0.0682 | 0.2428 | 0.1360 | 0.2471 | 0.1712 |
| RR | | | | **0.3820** | 0.0551 | 0.1911 | 0.1075 | 0.1868 | 0.1058 |
| RW | | | | 0.3821 | 0.0552 | **0.1776** | 0.1536 | **0.1582** | 0.1766 |
| RXY | | BCB | | 0.4775 | 0.0678 | 0.2387 | 0.1359 | 0.2453 | 0.1709 |
| RR | | | | 0.3968 | 0.0472 | 0.1904 | 0.1078 | 0.1862 | 0.1060 |
| RW | | | | **0.3810** | 0.0547 | **0.1761** | 0.1533 | **0.1576** | 0.1765 |
| RXY | 100:500 | SPB | NSN | - | - | 0.7516 | 0.4692 | 0.3500 | 0.0139 |
| RR | | | | - | - | **0.1881** | 0.4454 | **0.2503** | 0.6081 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| RW | | | | - | - | 0.3534 | 0.7349 | 0.5647 | 1.3436 |
| RXY | | | | - | - | 1.7153 | 1.8203 | **0.4773** | 0.9623 |
| RR | | BCaB | | - | - | **0.2690** | 0.4381 | 0.5521 | 0.6376 |
| RW | | | | - | - | 0.4750 | 0.7593 | 1.0064 | 1.3747 |
| RXY | | | | - | - | 1.7152 | 1.8202 | **0.4772** | 0.9621 |
| RR | | BCB | | - | - | **0.2689** | 0.4378 | 0.5520 | 0.6374 |
| RW | | | | - | - | 0.4748 | 0.7587 | 1.0063 | 1.3746 |
| RXY | | | | - | - | 0.2393 | 0.4750 | 0.2502 | 0.5806 |
| RR | | SPB | | - | - | **0.0696** | 0.1534 | **0.1134** | 0.3920 |
| RW | | | | - | - | 0.1030 | 0.2058 | 0.1754 | 0.4049 |
| RXY | | | | - | - | 0.5493 | 0.5772 | 0.3577 | 0.6227 |
| RR | | BCaB | SSN | - | - | **0.113** | 0.1508 | 0.4583 | 0.5247 |
| RW | | | | - | - | 0.1364 | 0.2169 | **0.3155** | 0.4121 |
| RXY | | | | - | - | 0.5493 | 0.5772 | 0.3577 | 0.6226 |
| RR | | BCB | | - | - | **0.1138** | 0.1508 | 0.4583 | 0.5247 |
| RW | | | | - | - | 0.1363 | 0.2168 | **0.3154** | 0.4120 |

**ตารางที่ 5.2** ตารางการเปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตราฐานของค่าความกว้างช่วงความเชื่อมั่น

จากตารางที่ 5.2 สำหรับข้อมูลที่มีมิติสูงการประมาณค่าโดยกำลังสองน้อยสุดจะไม่สามารถประมาณค่าสัมประสิทธิ์การถดถอยได้ และเลขตัวหนาในตารางแสดงถึงวิธีสุ่มที่ดีที่สุด ผลการวิเคราะห์ตารางที่ 5.2 ได้ดังนี้

1. การประมาณค่าโดยกำลังสองน้อยที่สุดพบว่า วิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพสูงที่สุด วิธีสุ่มส่วนเหลือรองลงมา และวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพต่ำที่สุด

2. การประมาณค่าโดยแลซโซพบว่า วิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มส่วนเหลือประสิทธิภาพสูงที่สุด วิธีสุ่มค่าถ่วงน้ำหนักรองลงมา และวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพต่ำที่สุด

3. การประมาณค่าโดยแลซโซปรับได้พบว่า วิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพสูงที่สุด วิธีสุ่มส่วนเหลือรองลงมา และวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพต่ำที่สุด

โดยรวมแล้ววิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มส่วนเหลือประสิทธิภาพสูงที่สุด วิธีสุ่มค่าถ่วงน้ำหนักรองลงมา และวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพต่ำที่สุด

3) การเปรียบเทียบค่าเฉลี่ยของอัตราผลบวกเทียม (False positive rate)

สำหรับการเปรียบเทียบค่าเฉลี่ยของอัตราผลบวกเทียม (False positive rate) ถ้ามีค่าเข้าใกล้ 0 วิธีสุ่มนั้นจะมีประสิทธิภาพที่ดีสุด ผลสรุปได้ดังตารางต่อไปนี้

| random | n:p | method of CI | Type | OLS | | LS | | ALS | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | MEAN | SD | MEAN | SD | MEAN | SD |
| RXY | 100:20 | SPB | NSN | **0.0000** | 0.0000 | **0.0000** | 0.0000 | **0.0000** | 0.0000 |
| RR | | | | **0.0000** | 0.0000 | 0.0006 | 0.0062 | 0.0025 | 0.0123 |
| RW | | | | **0.0000** | 0.0000 | **0.0000** | 0.0000 | **0.0000** | 0.0000 |
| RXY | | BCaB | | **0.0000** | 0.0000 | **0.0000** | 0.0000 | **0.0018** | 0.0107 |
| RR | | | | **0.0000** | 0.0000 | 0.0050 | 0.0170 | 0.0087 | 0.0217 |
| RW | | | | **0.0000** | 0.0000 | 0.0018 | 0.0107 | 0.0100 | 0.0230 |
| RXY | | BCB | | **0.0000** | 0.0000 | **0.0000** | 0.0000 | **0.0018** | 0.0107 |
| RR | | | | **0.0000** | 0.0000 | 0.0050 | 0.0170 | 0.008 | 0.0217 |
| RW | | | | **0.0000** | 0.0000 | 0.0018 | 0.0107 | 0.0100 | 0.0230 |
| RXY | | SPB | SSN | 0.0946 | 0.0311 | **0.1173** | 0.0056 | **0.1176** | 0.0000 |
| RR | | | | 0.0864 | 0.0363 | 0.1176 | 0.0000 | **0.1176** | 0.0000 |
| RW | | | | **0.0835** | 0.0376 | 0.1177 | 0.0007 | **0.1176** | 0.0000 |
| RXY | | BcaB | | 0.0904 | 0.0305 | **0.1051** | 0.0240 | 0.1166 | 0.0078 |
| RR | | | | 0.0814 | 0.0375 | 0.1171 | 0.0055 | 0.1176 | 0.0000 |
| RW | | | | **0.0808** | 0.0370 | 0.1105 | 0.0240 | **0.1165** | 0.0077 |
| RXY | | BCB | | 0.0917 | 0.0305 | **0.1051** | 0.0240 | 0.1166 | 0.0078 |
| RR | | | | 0.0855 | 0.0337 | 0.1171 | 0.0055 | 0.1176 | 0.0000 |
| RW | | | | **0.0796** | 0.0375 | 0.1110 | 0.0179 | **0.1165** | 0.0077 |
| RXY | 100:500 | SPB | NSN | - | - | 0.0321 | 0.0028 | **0.0207** | 0.0026 |
| RR | | | | - | - | **0.0098** | 0.0081 | 0.0225 | 0.0035 |
| RW | | | | - | - | 0.0130 | 0.0072 | 0.0320 | 0.0028 |
| RXY | | BcaB | | - | - | 0.0276 | 0.0026 | **0.0177** | 0.0029 |
| RR | | | | - | - | **0.0090** | 0.0080 | 0.0219 | 0.0041 |
| RW | | | | - | - | 0.0138 | 0.0066 | 0.0279 | 0.0024 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| RXY | | | | - | - | 0.0276 | 0.0026 | **0.0177** | 0.0029 |
| RR | | BCB | | - | - | **0.0090** | 0.0080 | 0.0219 | 0.0041 |
| RW | | | | - | - | 0.0138 | 0.0066 | 0.0279 | 0.0024 |
| RXY | | | | - | - | 0.0323 | 0.0037 | 0.0329 | 0.0024 |
| RR | | SPB | | - | - | 0.0136 | 0.0055 | 0.0374 | 0.0027 |
| RW | | | | - | - | **0.0134** | 0.0048 | **0.0290** | 0.0031 |
| RXY | | | | - | - | 0.0291 | 0.0040 | **0.0240** | 0.0029 |
| RR | | BcaB | SSN | - | - | 0.0106 | 0.0056 | 0.0341 | 0.0026 |
| RW | | | | - | - | **0.0098** | 0.0058 | 0.0261 | 0.0038 |
| RXY | | | | - | - | 0.0292 | 0.0040 | **0.0240** | 0.0029 |
| RR | | BCB | | - | - | 0.0106 | 0.0056 | 0.0341 | 0.0026 |
| RW | | | | - | - | **0.0099** | 0.0058 | 0.0261 | 0.0038 |

**ตารางที่ 5.3** ตารางการเปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของอัตราผลบวกเทียม

จากตารางที่ 5.3 สำหรับข้อมูลที่มีมิติสูงการประมาณค่าโดยกำลังสองน้อยสุดจะไม่สามารถประมาณค่าสัมประสิทธิ์การถดถอยได้ และเลขตัวหนาในตารางแสดงถึงวิธีสุ่มที่ดีที่สุด ผลการวิเคราะห์ตารางที่ 5.3 ได้ดังนี้

1. การประมาณค่าโดยกำลังสองน้อยที่สุดพบว่า วิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพสูงที่สุด วิธีสุ่มส่วนเหลือรองลงมา และวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพต่ำที่สุด



2. การประมาณค่าโดยแลซโซพบว่า วิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพสูงที่สุด วิธีสุ่มค่าถ่วงน้ำหนักรองลงมา และวิธีสุ่มส่วนเหลือมีประสิทธิภาพต่ำที่สุด

3. การประมาณค่าโดยแลซโซปรับได้พบว่า วิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพสูงที่สุด วิธีสุ่มส่วนเหลือรองลงมา และวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพต่ำที่สุด

โดยรวมแล้ววิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มตัวแปรตามและตัวแปรอิสระมีประสิทธิภาพสูงที่สุด วิธีสุ่มส่วนเหลือรองลงมา และวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพต่ำที่สุด

4) การเปรียบเทียบค่าเฉลี่ยของอัตราผลลบเทียม (False negative rate)

สำหรับการเปรียบเทียบค่าเฉลี่ยของอัตราผลลบเทียม (False negative rate) ถ้ามีค่าเข้าใกล้ 0 วิธีสุ่มนั้นจะมีประสิทธิภาพที่ดีสุด ผลสรุปได้ดังตารางต่อไปนี้

| random | n:p | method of CI | Type | OLS | | LS | | ALS | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | MEAN | SD | MEAN | SD | MEAN | SD |
| RXY | 100:20 | SPB | NSN | **0.1168** | 0.1256 | 0.0033 | 0.0234 | **0.0000** | 0.0000 |
| RR | | | | 0.2011 | 0.1410 | **0.0000** | 0.0000 | **0.0000** | 0.0000 |
| RW | | | | 0.2166 | 0.1422 | **0.0000** | 0.0000 | **0.0000** | 0.0000 |
| RXY | | BcaB | | **0.1004** | 0.1197 | 0.0195 | 0.0569 | **0.0000** | 0.0000 |
| RR | | | | 0.2182 | 0.1496 | **0.0000** | 0.0000 | **0.0000** | 0.0000 |
| RW | | | | 0.2161 | 0.1489 | 0.0083 | 0.0365 | **0.0000** | 0.0000 |
| RXY | | BCB | | **0.1077** | 0.1249 | 0.0178 | 0.0550 | **0.0000** | 0.0000 |
| RR | | | | 0.1861 | 0.1439 | **0.0000** | 0.0000 | **0.0000** | 0.0000 |
| RW | | | | 0.2119 | 0.1445 | 0.0066 | 0.0328 | **0.0000** | 0.0000 |
| RXY | | SPB | SSN | **0.1298** | 0.1528 | 0.0075 | 0.0428 | **0.0000** | 0.0000 |
| RR | | | | 0.2188 | 0.1618 | **0.0000** | 0.0000 | **0.0000** | 0.0000 |
| RW | | | | 0.2418 | 0.1710 | 0.0025 | 0.0250 | **0.0000** | 0.0000 |
| RXY | | BcaB | | 0.2314 | 0.1421 | 0.0245 | 0.0740 | 0.0050 | 0.0351 |
| RR | | | | **0.2310** | 0.1690 | **0.0025** | 0.0250 | **0.0000** | 0.0000 |
| RW | | | | 0.2314 | 0.1679 | 0.0045 | 0.0318 | **0.0000** | 0.0000 |
| RXY | | BCB | | **0.1138** | 0.1452 | 0.0245 | 0.0740 | 0.0050 | 0.0351 |
| RR | | | | 0.1977 | 0.1696 | **0.0025** | 0.0250 | **0.0000** | 0.0000 |
| RW | | | | 0.2354 | 0.1608 | 0.0045 | 0.0319 | **0.0000** | 0.0000 |
| RXY | 100:500 | SPB | NSN | - | - | **0.0000** | 0.0000 | **0.0000** | 0.0000 |
| RR | | | | - | - | 0.1575 | 0.1042 | 0.0225 | 0.0508 |
| RW | | | | - | - | 0.0601 | 0.0662 | 0.0050 | 0.0351 |
| RXY | | BcaB | | - | - | 0.8091 | 0.0367 | **0.6549** | 0.0533 |
| RR | | | | - | - | **0.6494** | 0.0660 | 0.7310 | 0.0452 |
| RW | | | | - | - | 0.7115 | 0.0603 | 0.8016 | 0.0375 |
| RXY | | BCB | | - | - | 0.8090 | 0.0366 | **0.6552** | 0.0531 |
| RR | | | | - | - | **0.6503** | 0.0655 | 0.7315 | 0.0453 |
| RW | | | | - | - | 0.7120 | 0.0600 | 0.8018 | 0.0373 |
| RXY | | SPB | SSN | - | - | **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 |

**ตารางที่ 5.4** ตารางการเปรียบเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของอัตราผลลบเทียม

จากตารางที่ 5.4 สำหรับข้อมูลที่มีมิติสูงการประมาณค่าโดยกำลังสองน้อยสุดจะไม่สามารถประมาณค่าสัมประสิทธิ์การถดถอยได้ และเลขตัวหนาในตารางแสดงถึงวิธีสุ่มที่ดีที่สุด ผลการวิเคราะห์ตารางที่ 5.4 ได้ดังนี้

1. การประมาณค่าโดยกำลังสองน้อยที่สุดพบว่า วิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มแปรตามและตัวแปรอิสระมีประสิทธิภาพสูงที่สุด วิธีสุ่มส่วนเหลือรองลงมา และวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพต่ำที่สุด

2. การประมาณค่าโดยแลซโซพบว่า วิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มส่วนเหลือมีประสิทธิภาพสูงที่สุด วิธีสุ่มตัวแปรตามและตัวแปรอิสระรองลงมา และวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพต่ำที่สุด

3. การประมาณค่าโดยแลซโซปรับได้พบว่า วิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มทั้ง 3 วิธี มีประสิทธิภาพใกล้เคียงกัน



โดยรวมแล้วไม่สามารถสรุปได้ว่าวิธีบูตสแตรปแบบไม่ใช้พารามิเตอร์ในการประมาณช่วงความเชื่อมั่นของค่าสัมประสิทธิ์การถดถอยเชิงเส้นโดยวิธีสุ่มวิธีใดมีประสิทธิภาพดีที่สุด แต่สามารถสรุปได้ว่าวิธีสุ่มค่าถ่วงน้ำหนักมีประสิทธิภาพต่ำที่สุด

## 5.2 ข้อเสนอแนะ

1) การวิจัยครั้งนี้ได้ทำการเปรียบเทียบ วิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้น สำหรับการทำวิจัยในครั้งต่อไป อาจทำการศึกษาการประมาณค่าสมการในรูปแบบอื่น เช่น พหุนาม สมการเอ็กซ์โปเนนเชียล หรือในรูปแบบอื่น ๆ

2) การวิจัยครั้งนี้ได้ทำการศึกษาภายใต้ความคลาดเคลื่อนที่มีการแจกแจงแบบปกติเท่านั้น ในการวิจัยครั้งต่อไปอาจจะศึกษาภายใต้ความคลาดเคลื่อนที่มีการแจกแจงในรูปแบบอื่น เช่น แบบโลจิสติก แบบดับเบิ้ลเอ็กซ์โปรเนนเชียล หรือในรูปแบบอื่น ๆ

# บรรณานุกรม


Chatterjee, A. and S. N. Lahiri (2011). "Bootstrapping lasso estimators." Journal of the American Statistical Association **106**(494): 608-625.

Rauch, G. (2018). "W. Q. Meeker, G. J. Hahn, L. A. Escobar. Statistical Intervals—A Guide for Practitioners and Researchers. (2017) Hoboken, NJ: John Wiley & Sons., 648 pages, ISBN 978-0-471-68717-7." Biometrical Journal **60**(5): 1027-1027.

Tibshirani, R. (1996). "Regression Shrinkage and Selection via the Lasso." Journal of the Royal Statistical Society. Series B (Methodological) **58**(1): 267-288.

[We propose a new method for estimation in linear models. The `lasso' minimizes the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a constant. Because of the nature of this constraint it tends to produce some coefficients that are exactly 0 and hence gives interpretable models. Our simulation studies suggest that the lasso enjoys some of the favourable properties of both subset selection and ridge regression. It produces interpretable models like subset selection and exhibits the stability of ridge regression. There is also an interesting relationship with recent work in adaptive function estimation by Donoho and Johnstone. The lasso idea is quite general and can be applied in a variety of statistical models: extensions to generalized regression models and tree-based models are briefly described.]

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

จิตรถเวช, จ. (2558). การวิเคราะห์การถดถอย.

พรดาเนินสวัสดิ์, พ. (2560). การเปรียบเทียบประสิทธิภาพของการประมาณค่าพารามิเตอร์ด้วยวิธีแลซโซในการถดถอยเชิงเส้นที่มีมิติสูง. สาขาวิชาสถิติประยุกต์ แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต

ภาควิชาสถิติ, มหาวิทยาลัยศิลปากร. บัณฑิตวิทยาลัย.

พึ่งพาพงศ์, ว. (2558). "บทวิเคราะห์วิธีวิเคราะห์การถดถอยเชิงเส้นสำหรับข้อมูลที่มีมิติสูง. ." วารสารวิทยาศาสตร์และเทคโนโลยี **2**: 211-233.

ลิสวัสด์, ช. ช. แ. (2561). การเปรียบเทียบประสิทธิภาพของการวิธีวิเคราะห์การถดถอยแบบบริดจ์ แลซโซ และแลซโซแบบปรับปรุงในตัวแบบการถดถอยปัวซงภายใต้ข้อมูลที่มีมิติสูงแบบบางเบา และตัวแปรอิสระที่มีความสัมพันธ์สูง. สาขาวิชาคณิตศาสตร์และสถิติ
คณะวิทยาศาสตร์และเทคโนโลย, มหาวิทยาลัยธรรมศาสตร์. ทยาศาสตรมหาบัณฑิต.

สินสมบูรณ์ทอง, ส. (2553). การวิเคราะห์เชิงสถิติ.

# ภาคผนวก

การวิจัยครั้งนี้ใช้โปรแกรม R เวอร์ชั่น 1.3.1073 ในการจำลองข้อมูลและคำนวณค่าต่าง ๆ

**โปรแกรมประมาณช่วงความเชื่อมั่นค่าสมประสิทธิ์การถดถอยเชิงเส้นทุกสถานการณ์สำหรับข้อมูลที่มีมิติต่ำ (No small nonzero)**

**กำหนดค่า**

$$\boldsymbol{\beta} = \left[2, 4, -5, 1, -3, \underbrace{0, 0, 0, \dots, 0}_{15 \text{ ตัว}}\right]^{\mathbf{T}}$$

```
start<-Sys.time()
beta <- c(2,4,-5,1,-3,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
p<-length(beta)+1
B<-1000
ndatasets<-100
# Initiate matrix to store coefficients for 3 bootstrap methods
boot1<-matrix(rep(0, B*p), nrow=B)
boot2<-matrix(rep(0, B*p), nrow=B)
boot3<-matrix(rep(0, B*p), nrow=B)
boot1_lasso<-matrix(rep(0, B*p), nrow=B)
boot2_lasso<-matrix(rep(0, B*p), nrow=B)
boot3_lasso<-matrix(rep(0, B*p), nrow=B)
boot1_adaptlasso<-matrix(rep(0, B*p), nrow=B)
boot2_adaptlasso<-matrix(rep(0, B*p), nrow=B)
boot3_adaptlasso<-matrix(rep(0, B*p), nrow=B)

alpha<-0.05
simple_boot1_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot1_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot2_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot2_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot3_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot3_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot1_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
```

```
BCa_boot1_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot2_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot2_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot3_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot3_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot1_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot1_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot2_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot2_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot3_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot3_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BootstrpT_boot1_lb<-matrix(rep(0,ndatasets*(p-1)), nrow=ndatasets)
BootstrpT_boot1_ub<-matrix(rep(0,ndatasets*(p-1)), nrow=ndatasets)
BootstrpT_boot2_lb<-matrix(rep(0,ndatasets*(p-1)), nrow=ndatasets)
BootstrpT_boot2_ub<-matrix(rep(0,ndatasets*(p-1)), nrow=ndatasets)
BootstrpT_boot3_lb<-matrix(rep(0,ndatasets*(p-1)), nrow=ndatasets)
BootstrpT_boot3_ub<-matrix(rep(0,ndatasets*(p-1)), nrow=ndatasets)
se_boot1<-matrix(rep(0, B*(p-1)), nrow=B)
se_boot2<-matrix(rep(0, B*(p-1)), nrow=B)
se_boot3<-matrix(rep(0, B*(p-1)), nrow=B)
z_boot1<-matrix(rep(0, B*(p-1)), nrow=B)
z_boot2<-matrix(rep(0, B*(p-1)), nrow=B)
z_boot3<-matrix(rep(0, B*(p-1)), nrow=B)

simple_boot1_lasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot1_lasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot2_lasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot2_lasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot3_lasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot3_lasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot1_lasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
```

```
BCa_boot1_lasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot2_lasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot2_lasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot3_lasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot3_lasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot1_lasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot1_lasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot2_lasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot2_lasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot3_lasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot3_lasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BootstrpT_boot1_lasso_lb<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
BootstrpT_boot1_lasso_ub<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
BootstrpT_boot2_lasso_lb<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
BootstrpT_boot2_lasso_ub<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
BootstrpT_boot3_lasso_lb<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
BootstrpT_boot3_lasso_ub<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
se_boot1_lasso<-matrix(rep(0, B*(p-1)), nrow=B)
se_boot2_lasso<-matrix(rep(0, B*(p-1)), nrow=B)
se_boot3_lasso<-matrix(rep(0, B*(p-1)), nrow=B)
z_boot1_lasso<-matrix(rep(0, B*(p-1)), nrow=B)
z_boot2_lasso<-matrix(rep(0, B*(p-1)), nrow=B)
z_boot3_lasso<-matrix(rep(0, B*(p-1)), nrow=B)
se_boot1_lasso_forBT<-matrix(rep(0, B*5), nrow=B)
se_boot2_lasso_forBT<-matrix(rep(0, B*5), nrow=B)
se_boot3_lasso_forBT<-matrix(rep(0, B*5), nrow=B)
z_boot1_lasso_forBT<-matrix(rep(0, B*5), nrow=B)
z_boot2_lasso_forBT<-matrix(rep(0, B*5), nrow=B)
z_boot3_lasso_forBT<-matrix(rep(0, B*5), nrow=B)

simple_boot1_adaptlasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
```

```
simple_boot1_adaptlasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot2_adaptlasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot2_adaptlasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot3_adaptlasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot3_adaptlasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot1_adaptlasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot1_adaptlasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot2_adaptlasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot2_adaptlasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot3_adaptlasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot3_adaptlasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot1_adaptlasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot1_adaptlasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot2_adaptlasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot2_adaptlasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot3_adaptlasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot3_adaptlasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BootstrpT_boot1_adaptlasso_lb<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
BootstrpT_boot1_adaptlasso_ub<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
BootstrpT_boot2_adaptlasso_lb<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
BootstrpT_boot2_adaptlasso_ub<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
BootstrpT_boot3_adaptlasso_lb<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
BootstrpT_boot3_adaptlasso_ub<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
se_boot1_adaptlasso<-matrix(rep(0, B*(p-1)), nrow=B)
se_boot2_adaptlasso<-matrix(rep(0, B*(p-1)), nrow=B)
se_boot3_adaptlasso<-matrix(rep(0, B*(p-1)), nrow=B)
z_boot1_adaptlasso<-matrix(rep(0, B*(p-1)), nrow=B)
z_boot2_adaptlasso<-matrix(rep(0, B*(p-1)), nrow=B)
z_boot3_adaptlasso<-matrix(rep(0, B*(p-1)), nrow=B)
for (ds in 1:ndatasets)
{
```

```
#OLS
xmatrix <- matrix(rnorm(100*20, mean=0, sd=1), 100, 20)
ymatrix <- xmatrix%*%beta+matrix(rnorm(100), ncol=1)
df<-as.data.frame(cbind(ymatrix, xmatrix))
names(df)[1]<-"y"
fit.model2 <- lm(y~., data=df)
residual<-resid(fit.model2)

#Lasso
lcvmodel2<-cv.glmnet(x=xmatrix, y=ymatrix, family="gaussian", type.measure="mse")
lcoef2<-as.numeric(coef(lcvmodel2))
y_hat_lasso <- predict(lcvmodel2, xmatrix)
residual_lasso<-ymatrix-y_hat_lasso
#Adaptlasso
w_adaptlasso2<-(1/abs(betaest2_without_intercept))
alcvmodel2<-cv.glmnet(x=xmatrix, y=ymatrix, family="gaussian",
type.measure="mse", penalty.factor=w_adaptlasso2)
alcoef2<-as.numeric(coef(alcvmodel2))
y_hat_adaptlasso <- predict(alcvmodel2, xmatrix)
residual_adaptlasso<-ymatrix-y_hat_adaptlasso
for (i in 1:B)
{
  #  Bootstrap Method 1
  sample_index = sample.int(nrow(xmatrix), nrow(ymatrix), replace=TRUE)
  new_data <- df[sample_index, ]
  xmatrix_sample<-data.matrix(new_data[,-1])
  ymatrix_sample<-data.matrix(new_data[,1])
  fit.model1 <- lm(y~., data=new_data)
  summary.model1 <- summary(fit.model1)
  coef1 <- summary.model1$coefficients
  betaest1 <- coef1[,"Estimate"]
```

```
    boot1[i,]<-betaest1
    SE_beta1<-coef1[,"Std. Error"]
    dfSE_beta1<-as.data.frame(SE_beta1)
    dfbetaest1<-as.data.frame(betaest1)
    betaest1_without_intercept<-dfbetaest1[-1,]
    SE_beta1_without_intercept<-dfSE_beta1[-1,]
    z1_boot1=(beta-betaest1_without_intercept)/( SE_beta1_without_intercept)
    z_boot1[i,]<-z1_boot1
    se_boot1[i,]<-SE_beta1_without_intercept

    # Bootstrap Method 2
    residual_index = sample(residual,size=100, replace=TRUE)
    dfnew_ranresi<-as.data.frame(cbind(ymatrix,
xmatrix,e=residual,y_hat=fitted(fit.model2),e_random=residual_index))
    y_hat<-dfnew_ranresi[,23]
    residual_ran<-dfnew_ranresi[,24]
    y_hat_star<-y_hat+residual_ran
    dfnewfinal<-as.data.frame(cbind(ymatrix,
xmatrix,e=residual,y_hat=fitted(fit.model2),e_random=residual_index,y_star=y_hat_st
ar))
    dfforfit<-as.data.frame(cbind(ystar=y_hat_star,xmatrix))
    names(dfforfit)[1]<-"ystar"
    fit.modelnew <- lm(ystar~., data=dfforfit)
    summary.model2 <- summary(fit.modelnew)
    coef2 <- summary.model2$coefficients
    betaest2 <- coef2[,"Estimate"]
    boot2[i,]<-betaest2
    SE_beta2<-coef2[,"Std. Error"]
    dfSE_beta2<-as.data.frame(SE_beta2)
    dfbetaest2<-as.data.frame(betaest2)
    betaest2_without_intercept<-dfbetaest2[-1,]
```

```
SE_beta2_without_intercept<-dfSE_beta2[-1,]
z2_boot2=(beta-betaest2_without_intercept)/( SE_beta2_without_intercept)
z_boot2[i,]<-z2_boot2
se_boot2[i,]<-SE_beta2_without_intercept

# Bootstrap Method 3 - Random weight
w<-rexp(100,1)
#fit.model <- lm.wfit(xmatrix,ymatrix,w)
# OLS
fit.model3<-lm(y~.,weights=w, data=df)
summary.model3 <- summary(fit.model3)
coef3 <- summary.model3$coefficients
betaest3 <- coef3[,"Estimate"]
boot3[i,]<-betaest3
SE_beta3<-coef3[,"Std. Error"]
dfSE_beta3<-as.data.frame(SE_beta3)
dfbetaest3<-as.data.frame(betaest3)
betaest3_without_intercept<-dfbetaest3[-1,]
SE_beta3_without_intercept<-dfSE_beta3[-1,]
z2_boot3=(beta-betaest3_without_intercept)/( SE_beta3_without_intercept)
z_boot3[i,]<-z2_boot3
se_boot3[i,]<-SE_beta3_without_intercept

# Lasso
# Bootstrap Method 1
library(glmnet)
lcvmodel1<-cv.glmnet(x=xmatrix_sample, y=ymatrix_sample, family="gaussian",
type.measure="mse")
lcoef1<-as.numeric(coef(lcvmodel1))
boot1_lasso[i,]<-lcoef1
SE_beta1_lasso_without_intercept<-sqrt((beta-lcoef1[-1])^2/(ndatasets-1))
```

```
    z1_lasso_boot1=(beta[1:5]-lcoef1[2:6])/( SE_beta1_lasso_without_intercept[1:5])
    z_boot1_lasso_forBT[i,]<-z1_lasso_boot1
    se_boot3_lasso[i,]<-SE_beta3_without_intercept
    # Bootstrap Method 2
    residual_lasso_index = sample(residual_lasso,size=100, replace=TRUE)
    y_hat_lasso_star<-y_hat_lasso+residual_lasso_index
    lcvmodel2new<-cv.glmnet(x=xmatrix, y=y_hat_lasso_star, family="gaussian",
type.measure="mse")
    lcoef2new<-as.numeric(coef(lcvmodel2new))
    boot2_lasso[i,]<-lcoef2new
    # Bootstrap Method 3
    lcvmodel3<-cv.glmnet(x=xmatrix, y=ymatrix, family="gaussian",
type.measure="mse", weights=w)
    lcoef3<-as.numeric(coef(lcvmodel3))
    boot3_lasso[i,]<-lcoef3
    # Adaptive Lasso
    # Bootstrap Method 1
    # For n>p
    # 1. Use beta from OLS
    # 2. Define weight w=1/abs(beta_OLS)
    w_adaptlasso1<-(1/abs(betaest1_without_intercept))
    alcvmodel1<-cv.glmnet(x=xmatrix_sample, y=ymatrix_sample, family="gaussian",
type.measure="mse", penalty.factor=w_adaptlasso1)
    alcoef1<-as.numeric(coef(alcvmodel1))
    boot1_adaptlasso[i,]<-alcoef1
    # Bootstrap Method 2
    residual_adaptlasso_index = sample(residual_adaptlasso,size=100, replace=TRUE)
    y_hat_adaptlasso_star<-y_hat_adaptlasso+residual_adaptlasso_index
    alcvmodel2new<-cv.glmnet(x=xmatrix, y=y_hat_adaptlasso_star, family="gaussian",
type.measure="mse")
    alcoef2new<-as.numeric(coef(alcvmodel2new))
```

```
    boot2_adaptlasso[i,]<-alcoef2new
    # Bootstrap Method 3
    # 1. Use beta from Weighted_OLS (with weights=wt)
    # 2. Define weight w=1/abs(beta_Weighted_OLS)
    w_adaptlasso3<-1/abs(betaest3_without_intercept)
    alcvmodel3<-cv.glmnet(x=xmatrix, y=ymatrix, family="gaussian",
type.measure="mse", penalty.factor=w_adaptlasso3, weights=w)
    alcoef3<-as.numeric(coef(alcvmodel3))
    boot3_adaptlasso[i,]<-alcoef3
    print("ds: ")
    print(ds)
    print(" i : ")
    print(i)
  }

  # Code for getting CI for each method
  #OLS
  # CI-simple
  simple_boot1_lb[ds,]<-apply(boot1, 2, quantile, probs=alpha/2)
  simple_boot1_ub[ds,]<-apply(boot1, 2, quantile, probs=1-alpha/2)
  simple_boot2_lb[ds,]<-apply(boot2, 2, quantile, probs=alpha/2)
  simple_boot2_ub[ds,]<-apply(boot2, 2, quantile, probs=1-alpha/2)
  simple_boot3_lb[ds,]<-apply(boot3, 2, quantile, probs=alpha/2)
  simple_boot3_ub[ds,]<-apply(boot3, 2, quantile, probs=1-alpha/2)
  # CI-BCa
  #BCa boot1
  jackknif_sample_mean_boot1<-as.data.frame(apply(boot1, 2, mean))
  a_BCa_boot1<-sum((jackknif_sample_mean_boot1-
boot1)^3)/(6*((sum((jackknif_sample_mean_boot1-boot1)^2))^1.5))
  b_BCa_boot1<-(sum(boot1<beta))/(B*p)
  Zb_BCa_boot1<-qnorm(b_BCa_boot1)
```

```
  Z<-qnorm(1-alpha/2)
  alpha1_BCa_boot1<-pnorm(Zb_BCa_boot1+((Zb_BCa_boot1-Z)/(1-
(a_BCa_boot1*(Zb_BCa_boot1-Z)))))
  alpha2_BCa_boot1<-pnorm(Zb_BCa_boot1+((Zb_BCa_boot1+Z)/(1-
(a_BCa_boot1*(Zb_BCa_boot1+Z)))))
  BCa_boot1_lb[ds,]<-apply(boot1, 2, quantile, probs=alpha1_BCa_boot1)
  BCa_boot1_ub[ds,]<-apply(boot1, 2, quantile, probs=alpha2_BCa_boot1)
  #BCa boot2
  jackknif_sample_mean_boot2<-as.data.frame(apply(boot2, 2, mean))
  a_BCa_boot2<-sum((jackknif_sample_mean_boot2-
boot2)^3)/(6*((sum((jackknif_sample_mean_boot2-boot2)^2))^1.5))
  b_BCa_boot2<-(sum(boot2<beta))/(B*p)
  Zb_BCa_boot2<-qnorm(b_BCa_boot2)
  Z<-qnorm(1-alpha/2)
  alpha1_BCa_boot2<-pnorm(Zb_BCa_boot2+((Zb_BCa_boot2-Z)/(1-
(a_BCa_boot2*(Zb_BCa_boot2-Z)))))
  alpha2_BCa_boot2<-pnorm(Zb_BCa_boot2+((Zb_BCa_boot2+Z)/(1-
(a_BCa_boot2*(Zb_BCa_boot2+Z)))))
  BCa_boot2_lb[ds,]<-apply(boot3, 2, quantile, probs=alpha1_BCa_boot2)
  BCa_boot2_ub[ds,]<-apply(boot3, 2, quantile, probs=alpha2_BCa_boot2)
  #BCa boot3
  jackknif_sample_mean_boot3<-as.data.frame(apply(boot3, 2, mean))
  a_BCa_boot3<-sum((jackknif_sample_mean_boot3-
boot3)^3)/(6*((sum((jackknif_sample_mean_boot3-boot3)^2))^1.5))
  b_BCa_boot3<-(sum(boot3<beta))/(B*p)
  Zb_BCa_boot3<-qnorm(b_BCa_boot3)
  Z<-qnorm(1-alpha/2)
  alpha1_BCa_boot3<-pnorm(Zb_BCa_boot3+((Zb_BCa_boot3-Z)/(1-
(a_BCa_boot3*(Zb_BCa_boot3-Z)))))
  alpha2_BCa_boot3<-pnorm(Zb_BCa_boot3+((Zb_BCa_boot3+Z)/(1-
(a_BCa_boot3*(Zb_BCa_boot3+Z)))))
```

```
BCa_boot3_lb[ds,]<-apply(boot3, 2, quantile, probs=alpha1_BCa_boot3)
BCa_boot3_ub[ds,]<-apply(boot3, 2, quantile, probs=alpha2_BCa_boot3)
#CI-BC
#BC boot1
alpha1_BC_boot1<-pnorm((2*Zb_BCa_boot1)-Z)
alpha2_BC_boot1<-pnorm((2*Zb_BCa_boot1)+Z)
BC_boot1_lb[ds,]<-apply(boot1, 2, quantile, probs=alpha1_BC_boot1)
BC_boot1_ub[ds,]<-apply(boot1, 2, quantile, probs=alpha2_BC_boot1)
#BC boot2
alpha1_BC_boot2<-pnorm((2*Zb_BCa_boot2)-Z)
alpha2_BC_boot2<-pnorm((2*Zb_BCa_boot2)+Z)
BC_boot2_lb[ds,]<-apply(boot2, 2, quantile, probs=alpha1_BC_boot2)
BC_boot2_ub[ds,]<-apply(boot2, 2, quantile, probs=alpha2_BC_boot2)
#BC boot3
alpha1_BC_boot3<-pnorm((2*Zb_BCa_boot3)-Z)
alpha2_BC_boot3<-pnorm((2*Zb_BCa_boot3)+Z)
BC_boot3_lb[ds,]<-apply(boot3, 2, quantile, probs=alpha1_BC_boot3)
BC_boot3_ub[ds,]<-apply(boot3, 2, quantile, probs=alpha2_BC_boot3)
#CI-Bootstrap-t
#Bootstrap boot1
Z1_star_lb<-apply(z_boot1, 2, quantile, probs=alpha/2)
Z1_star_ub<-apply(z_boot1, 2, quantile, probs=1-(alpha/2))
Z1_lb<-beta+(Z1_star_lb*SE_beta1_without_intercept)
Z1_ub<-beta+(Z1_star_ub*SE_beta1_without_intercept)
BootstrpT_boot1_lb[ds,]<-Z1_lb
BootstrpT_boot1_ub[ds,]<-Z1_ub
#Bootstrap boot2
Z2_star_lb<-apply(z_boot2, 2, quantile, probs=alpha/2)
Z2_star_ub<-apply(z_boot2, 2, quantile, probs=1-(alpha/2))
Z2_lb<-beta+(Z2_star_lb*SE_beta2_without_intercept)
Z2_ub<-beta+(Z2_star_ub*SE_beta2_without_intercept)
```

```
BootstrpT_boot2_lb[ds,]<-Z2_lb
BootstrpT_boot2_ub[ds,]<-Z2_ub
#Bootstrap boot3
Z3_star_lb<-apply(z_boot3, 2, quantile, probs=alpha/2)
Z3_star_ub<-apply(z_boot3, 2, quantile, probs=1-(alpha/2))
Z3_lb<-beta+(Z3_star_lb*SE_beta3_without_intercept)
Z3_ub<-beta+(Z3_star_ub*SE_beta3_without_intercept)
BootstrpT_boot3_lb[ds,]<-Z3_lb
BootstrpT_boot3_ub[ds,]<-Z3_ub

#Lasso
# CI-simple
simple_boot1_lasso_lb[ds,]<-apply(boot1_lasso, 2, quantile, probs=alpha/2)
simple_boot1_lasso_ub[ds,]<-apply(boot1_lasso, 2, quantile, probs=1-alpha/2)
simple_boot2_lasso_lb[ds,]<-apply(boot2_lasso, 2, quantile, probs=alpha/2)
simple_boot2_lasso_ub[ds,]<-apply(boot2_lasso, 2, quantile, probs=1-alpha/2)
simple_boot3_lasso_lb[ds,]<-apply(boot3_lasso, 2, quantile, probs=alpha/2)
simple_boot3_lasso_ub[ds,]<-apply(boot3_lasso, 2, quantile, probs=1-alpha/2)
#BCa boot1
jackknif_sample_mean_boot1_lasso<-as.data.frame(apply(boot1_lasso, 2, mean))
a_BCa_boot1_lasso<-sum((jackknif_sample_mean_boot1_lasso-
boot1_lasso)^3)/(6*((sum((jackknif_sample_mean_boot1_lasso-boot1_lasso)^2))^1.5))
b_BCa_boot1_lasso<-(sum(boot1_lasso[,-1]<beta))/(B*p)
Zb_BCa_boot1_lasso<-qnorm(b_BCa_boot1_lasso)
Z<-qnorm(1-alpha/2)
alpha1_BCa_boot1_lasso<-pnorm(Zb_BCa_boot1_lasso+((Zb_BCa_boot1_lasso-
Z)/(1-(a_BCa_boot1_lasso*(Zb_BCa_boot1_lasso-Z)))))
alpha2_BCa_boot1_lasso<-
pnorm(Zb_BCa_boot1_lasso+((Zb_BCa_boot1_lasso+Z)/(1-
(a_BCa_boot1_lasso*(Zb_BCa_boot1_lasso+Z)))))
```

```
  BCa_boot1_lasso_lb[ds,]<-apply(boot1_lasso, 2, quantile,
probs=alpha1_BCa_boot1_lasso)
  BCa_boot1_lasso_ub[ds,]<-apply(boot1_lasso, 2, quantile,
probs=alpha2_BCa_boot1_lasso)
  #BCa boot2
  jackknif_sample_mean_boot2_lasso<-as.data.frame(apply(boot2_lasso, 2, mean))
  a_BCa_boot2_lasso<-sum((jackknif_sample_mean_boot2_lasso-
boot2_lasso)^3)/(6*((sum((jackknif_sample_mean_boot2_lasso-boot2_lasso)^2))^1.5))
  b_BCa_boot2_lasso<-(sum(boot2_lasso[,-1]<beta))/(B*p)
  Zb_BCa_boot2_lasso<-qnorm(b_BCa_boot2_lasso)
  Z<-qnorm(1-alpha/2)
  alpha1_BCa_boot2_lasso<-pnorm(Zb_BCa_boot2_lasso+((Zb_BCa_boot2_lasso-
Z)/(1-(a_BCa_boot2_lasso*(Zb_BCa_boot2_lasso-Z)))))
  alpha2_BCa_boot2_lasso<-
pnorm(Zb_BCa_boot2_lasso+((Zb_BCa_boot2_lasso+Z)/(1-
(a_BCa_boot2_lasso*(Zb_BCa_boot2_lasso+Z)))))
  BCa_boot2_lasso_lb[ds,]<-apply(boot2_lasso, 2, quantile,
probs=alpha1_BCa_boot2_lasso)
  BCa_boot2_lasso_ub[ds,]<-apply(boot2_lasso, 2, quantile,
probs=alpha2_BCa_boot2_lasso)
  #BCa boot3
  jackknif_sample_mean_boot3_lasso<-as.data.frame(apply(boot3_lasso, 2, mean))
  a_BCa_boot3_lasso<-sum((jackknif_sample_mean_boot3_lasso-
boot3_lasso)^3)/(6*((sum((jackknif_sample_mean_boot3_lasso-boot3_lasso)^2))^1.5))
  b_BCa_boot3_lasso<-(sum(boot3_lasso[,-1]<beta))/(B*p)
  Zb_BCa_boot3_lasso<-qnorm(b_BCa_boot3_lasso)
  Z<-qnorm(1-alpha/2)
  alpha1_BCa_boot3_lasso<-pnorm(Zb_BCa_boot3_lasso+((Zb_BCa_boot3_lasso-
Z)/(1-(a_BCa_boot3_lasso*(Zb_BCa_boot3_lasso-Z)))))
```

```
  alpha2_BCa_boot3_lasso<-
pnorm(Zb_BCa_boot3_lasso+((Zb_BCa_boot3_lasso+Z)/(1-
(a_BCa_boot3_lasso*(Zb_BCa_boot3_lasso+Z)))))
  BCa_boot3_lasso_lb[ds,]<-apply(boot3_lasso, 2, quantile,
probs=alpha1_BCa_boot3_lasso)
  BCa_boot3_lasso_ub[ds,]<-apply(boot3_lasso, 2, quantile,
probs=alpha2_BCa_boot3_lasso)
  #BC boot1
  alpha1_BC_boot1_lasso<-pnorm((2*Zb_BCa_boot1_lasso)-Z)
  alpha2_BC_boot1_lasso<-pnorm((2*Zb_BCa_boot1_lasso)+Z)
  BC_boot1_lasso_lb[ds,]<-apply(boot1_lasso, 2, quantile,
probs=alpha1_BC_boot1_lasso)
  BC_boot1_lasso_ub[ds,]<-apply(boot1_lasso, 2, quantile,
probs=alpha2_BC_boot1_lasso)
  #BC boot2
  alpha1_BC_boot2_lasso<-pnorm((2*Zb_BCa_boot2_lasso)-Z)
  alpha2_BC_boot2_lasso<-pnorm((2*Zb_BCa_boot2_lasso)+Z)
  BC_boot2_lasso_lb[ds,]<-apply(boot2_lasso, 2, quantile,
probs=alpha1_BC_boot2_lasso)
  BC_boot2_lasso_ub[ds,]<-apply(boot2_lasso, 2, quantile,
probs=alpha2_BC_boot2_lasso)
  #BC boot3
  alpha1_BC_boot3_lasso<-pnorm((2*Zb_BCa_boot3_lasso)-Z)
  alpha2_BC_boot3_lasso<-pnorm((2*Zb_BCa_boot3_lasso)+Z)
  BC_boot3_lasso_lb[ds,]<-apply(boot3_lasso, 2, quantile,
probs=alpha1_BC_boot3_lasso)
  BC_boot3_lasso_ub[ds,]<-apply(boot3_lasso, 2, quantile,
probs=alpha2_BC_boot3_lasso)

  #AdaptLasso
  # CI-simple
```

```
  simple_boot1_adaptlasso_lb[ds,]<-apply(boot1_adaptlasso, 2, quantile,
probs=alpha/2)
  simple_boot1_adaptlasso_ub[ds,]<-apply(boot1_adaptlasso, 2, quantile, probs=1-
alpha/2)
  simple_boot2_adaptlasso_lb[ds,]<-apply(boot2_adaptlasso, 2, quantile,
probs=alpha/2)
  simple_boot2_adaptlasso_ub[ds,]<-apply(boot2_adaptlasso, 2, quantile, probs=1-
alpha/2)
  simple_boot3_adaptlasso_lb[ds,]<-apply(boot3_adaptlasso, 2, quantile,
probs=alpha/2)
  simple_boot3_adaptlasso_ub[ds,]<-apply(boot3_adaptlasso, 2, quantile, probs=1-
alpha/2)
  #BCa boot1
  jackknif_sample_mean_boot1_adaptlasso<-as.data.frame(apply(boot1_adaptlasso,
2, mean))
  a_BCa_boot1_adaptlasso<-sum((jackknif_sample_mean_boot1_adaptlasso-
boot1_adaptlasso)^3)/(6*((sum((jackknif_sample_mean_boot1_adaptlasso-
boot1_adaptlasso)^2))^1.5))
  b_BCa_boot1_adaptlasso<-(sum(boot1_adaptlasso[,-1]<beta))/(B*p)
  Zb_BCa_boot1_adaptlasso<-qnorm(b_BCa_boot1_adaptlasso)
  Z<-qnorm(1-alpha/2)
  alpha1_BCa_boot1_adaptlasso<-
pnorm(Zb_BCa_boot1_adaptlasso+((Zb_BCa_boot1_adaptlasso-Z)/(1-
(a_BCa_boot1_adaptlasso*(Zb_BCa_boot1_adaptlasso-Z)))))
  alpha2_BCa_boot1_adaptlasso<-
pnorm(Zb_BCa_boot1_adaptlasso+((Zb_BCa_boot1_adaptlasso+Z)/(1-
(a_BCa_boot1_adaptlasso*(Zb_BCa_boot1_adaptlasso+Z)))))
  BCa_boot1_adaptlasso_lb[ds,]<-apply(boot1_adaptlasso, 2, quantile,
probs=alpha1_BCa_boot1_adaptlasso)
  BCa_boot1_adaptlasso_ub[ds,]<-apply(boot1_adaptlasso, 2, quantile,
probs=alpha2_BCa_boot1_adaptlasso)
```

```
  #BCa boot2
  jackknif_sample_mean_boot2_adaptlasso<-as.data.frame(apply(boot2_adaptlasso,
2, mean))
  a_BCa_boot2_adaptlasso<-sum((jackknif_sample_mean_boot2_adaptlasso-
boot2_adaptlasso)^3)/(6*((sum((jackknif_sample_mean_boot2_adaptlasso-
boot2_adaptlasso)^2))^1.5))
  b_BCa_boot2_adaptlasso<-(sum(boot2_adaptlasso[,-1]<beta))/(B*p)
  Zb_BCa_boot2_adaptlasso<-qnorm(b_BCa_boot2_adaptlasso)
  Z<-qnorm(1-alpha/2)
  alpha1_BCa_boot2_adaptlasso<-
pnorm(Zb_BCa_boot2_adaptlasso+((Zb_BCa_boot2_adaptlasso-Z)/(1-
(a_BCa_boot2_adaptlasso*(Zb_BCa_boot2_adaptlasso-Z)))))
  alpha2_BCa_boot2_adaptlasso<-
pnorm(Zb_BCa_boot2_adaptlasso+((Zb_BCa_boot2_adaptlasso+Z)/(1-
(a_BCa_boot2_adaptlasso*(Zb_BCa_boot2_adaptlasso+Z)))))
  BCa_boot2_adaptlasso_lb[ds,]<-apply(boot2_adaptlasso, 2, quantile,
probs=alpha1_BCa_boot2_adaptlasso)
  BCa_boot2_adaptlasso_ub[ds,]<-apply(boot2_adaptlasso, 2, quantile,
probs=alpha2_BCa_boot2_adaptlasso)
  #BCa boot3
  jackknif_sample_mean_boot3_adaptlasso<-as.data.frame(apply(boot3_adaptlasso,
2, mean))
  a_BCa_boot3_adaptlasso<-sum((jackknif_sample_mean_boot3_adaptlasso-
boot3_adaptlasso)^3)/(6*((sum((jackknif_sample_mean_boot3_adaptlasso-
boot3_adaptlasso)^2))^1.5))
  b_BCa_boot3_adaptlasso<-(sum(boot3_adaptlasso[,-1]<beta))/(B*p)
  Zb_BCa_boot3_adaptlasso<-qnorm(b_BCa_boot3_adaptlasso)
  Z<-qnorm(1-alpha/2)
  alpha1_BCa_boot3_adaptlasso<-
pnorm(Zb_BCa_boot3_adaptlasso+((Zb_BCa_boot3_adaptlasso-Z)/(1-
(a_BCa_boot3_adaptlasso*(Zb_BCa_boot3_adaptlasso-Z)))))
```

```
  alpha2_BCa_boot3_adaptlasso<-
pnorm(Zb_BCa_boot3_adaptlasso+((Zb_BCa_boot3_adaptlasso+Z)/(1-
(a_BCa_boot3_adaptlasso*(Zb_BCa_boot3_adaptlasso+Z)))))
  BCa_boot3_adaptlasso_lb[ds,]<-apply(boot3_adaptlasso, 2, quantile,
probs=alpha1_BCa_boot3_adaptlasso)
  BCa_boot3_adaptlasso_ub[ds,]<-apply(boot3_adaptlasso, 2, quantile,
probs=alpha2_BCa_boot3_adaptlasso)
  #BC boot1
  alpha1_BC_boot1_adaptlasso<-pnorm((2*Zb_BCa_boot1_adaptlasso)-Z)
  alpha2_BC_boot1_adaptlasso<-pnorm((2*Zb_BCa_boot1_adaptlasso)+Z)
  BC_boot1_adaptlasso_lb[ds,]<-apply(boot1_adaptlasso, 2, quantile,
probs=alpha1_BC_boot1_adaptlasso)
  BC_boot1_adaptlasso_ub[ds,]<-apply(boot1_adaptlasso, 2, quantile,
probs=alpha2_BC_boot1_adaptlasso)
  #BC boot2
  alpha1_BC_boot2_adaptlasso<-pnorm((2*Zb_BCa_boot2_adaptlasso)-Z)
  alpha2_BC_boot2_adaptlasso<-pnorm((2*Zb_BCa_boot2_adaptlasso)+Z)
  BC_boot2_adaptlasso_lb[ds,]<-apply(boot2_adaptlasso, 2, quantile,
probs=alpha1_BC_boot2_adaptlasso)
  BC_boot2_adaptlasso_ub[ds,]<-apply(boot2_adaptlasso, 2, quantile,
probs=alpha2_BC_boot2_adaptlasso)
  #BC boot3
  alpha1_BC_boot3_adaptlasso<-pnorm((2*Zb_BCa_boot3_adaptlasso)-Z)
  alpha2_BC_boot3_adaptlasso<-pnorm((2*Zb_BCa_boot3_adaptlasso)+Z)
  BC_boot3_adaptlasso_lb[ds,]<-apply(boot3_adaptlasso, 2, quantile,
probs=alpha1_BC_boot3_adaptlasso)
  BC_boot3_adaptlasso_ub[ds,]<-apply(boot3_adaptlasso, 2, quantile,
probs=alpha2_BC_boot3_adaptlasso)
   }
stop<-Sys.time()
time <-stop-start
```

**โปรแกรมประมาณช่วงความเชื่อมั่นค่าสัมประสิทธิ์การถดถอยเชิงเส้นทุกสถานการณ์สำหรับข้อมูลที่มีมิติสูง (Some small nonzero)**

**กำหนดค่า**

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

```
start<-Sys.time()
beta<-c(0.3,3,-1,-0.65,3,-3,0.7,-1,3,-0.6,2,1,-1,-3,-
2,5,2,4,4,4,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
```

```
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
p<-length(beta)+1
B<-1000
ndatasets<-100
# Initiate matrix to store coefficients for 3 bootstrap methods
boot1<-matrix(rep(0, B*p), nrow=B)
boot2<-matrix(rep(0, B*p), nrow=B)
boot3<-matrix(rep(0, B*p), nrow=B)
boot1_lasso<-matrix(rep(0, B*p), nrow=B)
boot2_lasso<-matrix(rep(0, B*p), nrow=B)
boot3_lasso<-matrix(rep(0, B*p), nrow=B)
boot1_adaptlasso<-matrix(rep(0, B*p), nrow=B)
boot2_adaptlasso<-matrix(rep(0, B*p), nrow=B)
boot3_adaptlasso<-matrix(rep(0, B*p), nrow=B)
alpha<-0.05
simple_boot1_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot1_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot2_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot2_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot3_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot3_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot1_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot1_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot2_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot2_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot3_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot3_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot1_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
```

```
BC_boot1_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot2_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot2_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot3_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot3_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BootstrpT_boot1_lb<-matrix(rep(0,ndatasets*(p-1)), nrow=ndatasets)
BootstrpT_boot1_ub<-matrix(rep(0,ndatasets*(p-1)), nrow=ndatasets)
BootstrpT_boot2_lb<-matrix(rep(0,ndatasets*(p-1)), nrow=ndatasets)
BootstrpT_boot2_ub<-matrix(rep(0,ndatasets*(p-1)), nrow=ndatasets)
BootstrpT_boot3_lb<-matrix(rep(0,ndatasets*(p-1)), nrow=ndatasets)
BootstrpT_boot3_ub<-matrix(rep(0,ndatasets*(p-1)), nrow=ndatasets)
se_boot1<-matrix(rep(0, B*(p-1)), nrow=B)
se_boot2<-matrix(rep(0, B*(p-1)), nrow=B)
se_boot3<-matrix(rep(0, B*(p-1)), nrow=B)
z_boot1<-matrix(rep(0, B*(p-1)), nrow=B)
z_boot2<-matrix(rep(0, B*(p-1)), nrow=B)
z_boot3<-matrix(rep(0, B*(p-1)), nrow=B)

simple_boot1_lasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot1_lasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot2_lasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot2_lasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot3_lasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot3_lasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot1_lasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot1_lasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot2_lasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot2_lasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot3_lasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot3_lasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot1_lasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
```

```
BC_boot1_lasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot2_lasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot2_lasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot3_lasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot3_lasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BootstrpT_boot1_lasso_lb<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
BootstrpT_boot1_lasso_ub<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
BootstrpT_boot2_lasso_lb<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
BootstrpT_boot2_lasso_ub<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
BootstrpT_boot3_lasso_lb<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
BootstrpT_boot3_lasso_ub<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
se_boot1_lasso<-matrix(rep(0, B*(p-1)), nrow=B)
se_boot2_lasso<-matrix(rep(0, B*(p-1)), nrow=B)
se_boot3_lasso<-matrix(rep(0, B*(p-1)), nrow=B)
z_boot1_lasso<-matrix(rep(0, B*(p-1)), nrow=B)
z_boot2_lasso<-matrix(rep(0, B*(p-1)), nrow=B)
z_boot3_lasso<-matrix(rep(0, B*(p-1)), nrow=B)
se_boot1_lasso_forBT<-matrix(rep(0, B*(p-1)), nrow=B)
se_boot2_lasso_forBT<-matrix(rep(0, B*(p-1)), nrow=B)
se_boot3_lasso_forBT<-matrix(rep(0, B*(p-1)), nrow=B)
z_boot1_lasso_forBT<-matrix(rep(0, B*(p-1)), nrow=B)
z_boot2_lasso_forBT<-matrix(rep(0, B*(p-1)), nrow=B)
z_boot3_lasso_forBT<-matrix(rep(0, B*(p-1)), nrow=B)

simple_boot1_adaptlasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot1_adaptlasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot2_adaptlasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot2_adaptlasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot3_adaptlasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
simple_boot3_adaptlasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot1_adaptlasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
```

```
BCa_boot1_adaptlasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot2_adaptlasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot2_adaptlasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot3_adaptlasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BCa_boot3_adaptlasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot1_adaptlasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot1_adaptlasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot2_adaptlasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot2_adaptlasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot3_adaptlasso_lb<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BC_boot3_adaptlasso_ub<-matrix(rep(0,ndatasets*p), nrow=ndatasets)
BootstrpT_boot1_adaptlasso_lb<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
BootstrpT_boot1_adaptlasso_ub<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
BootstrpT_boot2_adaptlasso_lb<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
BootstrpT_boot2_adaptlasso_ub<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
BootstrpT_boot3_adaptlasso_lb<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
BootstrpT_boot3_adaptlasso_ub<-matrix(rep(0,ndatasets*5), nrow=ndatasets)
se_boot1_adaptlasso<-matrix(rep(0, B*(p-1)), nrow=B)
se_boot2_adaptlasso<-matrix(rep(0, B*(p-1)), nrow=B)
se_boot3_adaptlasso<-matrix(rep(0, B*(p-1)), nrow=B)
z_boot1_adaptlasso<-matrix(rep(0, B*(p-1)), nrow=B)
z_boot2_adaptlasso<-matrix(rep(0, B*(p-1)), nrow=B)
z_boot3_adaptlasso<-matrix(rep(0, B*(p-1)), nrow=B)
  for (ds in 1:ndatasets)
{
  #Generate X Y
  xmatrix <- matrix(rnorm(ndatasets*(p-1), mean=0, sd=1), ndatasets, p-1)
  ymatrix <- xmatrix%*%beta+matrix(rnorm(ndatasets), ncol=1)
  df<-as.data.frame(cbind(ymatrix, xmatrix))
  names(df)[1]<-"y"
```

```
  #Lasso
  library(glmnet)
  lcvmodel2<-cv.glmnet(x=xmatrix, y=ymatrix, family="gaussian", type.measure="mse")
  lcoef2<-as.numeric(coef(lcvmodel2))
  y_hat_lasso <- predict(lcvmodel2, xmatrix)
  residual_lasso<-ymatrix-y_hat_lasso

  #Adaptlasso
  wt2<-1/abs(ridgebeta_without_intercept2)
  alcvmodel2<-cv.glmnet(x=xmatrix, y=ymatrix, family="gaussian",
type.measure="mse", penalty.factor=wt2)
  alcoef2<-as.numeric(coef(alcvmodel2))
  y_hat_adaptlasso <- predict(alcvmodel2, xmatrix)


  for (i in 1:B)
  {

    #RIDGE

    ridgecvmodel2<-cv.glmnet(x=xmatrix, y=ymatrix, family="gaussian",
type.measure="mse", alpha=0)
    ridgebeta2<-as.numeric(coef(ridgecvmodel2))
    ridgebeta_without_intercept2<-ridgebeta2[-1]
    #  Bootstrap Method 1
    sample_index = sample.int(nrow(xmatrix), nrow(ymatrix), replace=TRUE)
    new_data <- df[sample_index, ]
    xmatrix_sample<-data.matrix(new_data[,-1])
    ymatrix_sample<-data.matrix(new_data[,1])

    #lasso
```

```
    lcvmodel1<-cv.glmnet(x=xmatrix_sample, y=ymatrix_sample, family="gaussian",
type.measure="mse")
    lcoef1<-as.numeric(coef(lcvmodel1))
    boot1_lasso[i,]<-lcoef1
    SE_beta1_lasso_without_intercept<-sqrt((beta-lcoef1[-1])^2/(ndatasets-1))
    z1_lasso_boot1=(beta[1:20]-lcoef1[2:21])/( SE_beta1_lasso_without_intercept[1:20])
    z_boot1_lasso_forBT[i,]<-z1_lasso_boot1
    #se_boot3_lasso[i,]<-SE_beta3_without_intercept

    #Bootstrap Method 2
    residual_lasso_index = sample(residual_lasso,size=100, replace=TRUE)
    y_hat_lasso_star<-y_hat_lasso+residual_lasso_index
    lcvmodel2new<-cv.glmnet(x=xmatrix, y=y_hat_lasso_star, family="gaussian",
type.measure="mse")
    lcoef2new<-as.numeric(coef(lcvmodel2new))
    boot2_lasso[i,]<-lcoef2new

    # Bootstrap Method 3
    w<-rexp(100,1)
    lcvmodel3<-cv.glmnet(x=xmatrix, y=ymatrix, family="gaussian",
type.measure="mse", weights=w)
    lcoef3<-as.numeric(coef(lcvmodel3))
    boot3_lasso[i,]<-lcoef3


    # Adaptive Lasso
    # Bootstrap Method 1
    # 1. Use beta from Weighted_Ridge
    library(MASS)
    ridgecvmodel1<-cv.glmnet(x=xmatrix, y=ymatrix, family="gaussian",
type.measure="mse", alpha=0)
```

```
ridgebeta1<-as.numeric(coef(ridgecvmodel1))
ridgebeta_without_intercept1<-ridgebeta1[-1]
wt1<-1/abs(ridgebeta_without_intercept1)
alcvmodel1<-cv.glmnet(x=xmatrix_sample, y=ymatrix_sample, family="gaussian",
type.measure="mse", penalty.factor=wt)
alcoef1<-as.numeric(coef(alcvmodel1))
boot1_adaptlasso[i,]<-alcoef1

# Bootstrap Method 2
residual_adaptlasso_index = sample(residual_adaptlasso,size=100, replace=TRUE)
y_hat_adaptlasso_star<-y_hat_adaptlasso+residual_adaptlasso_index
alcvmodel2new<-cv.glmnet(x=xmatrix, y=y_hat_adaptlasso_star, family="gaussian",
type.measure="mse")
alcoef2new<-as.numeric(coef(alcvmodel2new))
boot2_adaptlasso[i,]<-alcoef2new
# Bootstrap Method 3
# 1. Use beta from Weighted_Ridge
ridgecvmodel3<-cv.glmnet(x=xmatrix, y=ymatrix, family="gaussian",
type.measure="mse", alpha=0, weights=w)
ridgebeta3<-as.numeric(coef(ridgecvmodel3))
ridgebeta_without_intercept3<-ridgebeta3[-1]
# 2. Define weight w=1/abs(beta_Weighted_Ridge)
wt3<-1/abs(ridgebeta_without_intercept3)
alcvmodel3<-cv.glmnet(x=xmatrix, y=ymatrix, family="gaussian",
type.measure="mse", penalty.factor=wt3, weights=w)
alcoef3<-as.numeric(coef(alcvmodel3))
boot3_adaptlasso[i,]<-alcoef3
print("ds: ")
print(ds)
print(" i : ")
print(i)
```

```
    }

# Code for getting CI for each method
#Lasso
# CI-simple
simple_boot1_lasso_lb[ds,]<-apply(boot1_lasso, 2, quantile, probs=alpha/2)
simple_boot1_lasso_ub[ds,]<-apply(boot1_lasso, 2, quantile, probs=1-alpha/2)
simple_boot2_lasso_lb[ds,]<-apply(boot2_lasso, 2, quantile, probs=alpha/2)
simple_boot2_lasso_ub[ds,]<-apply(boot2_lasso, 2, quantile, probs=1-alpha/2)
simple_boot3_lasso_lb[ds,]<-apply(boot3_lasso, 2, quantile, probs=alpha/2)
simple_boot3_lasso_ub[ds,]<-apply(boot3_lasso, 2, quantile, probs=1-alpha/2)
#BCa boot1
jackknif_sample_mean_boot1_lasso<-as.data.frame(apply(boot1_lasso, 2, mean))
a_BCa_boot1_lasso<-sum((jackknif_sample_mean_boot1_lasso-
boot1_lasso)^3)/(6*((sum((jackknif_sample_mean_boot1_lasso-boot1_lasso)^2))^1.5))
b_BCa_boot1_lasso<-(sum(boot1_lasso[,-1]<beta))/(B*p)
Zb_BCa_boot1_lasso<-qnorm(b_BCa_boot1_lasso)
Z<-qnorm(1-alpha/2)
alpha1_BCa_boot1_lasso<-pnorm(Zb_BCa_boot1_lasso+((Zb_BCa_boot1_lasso-Z)/(1-
(a_BCa_boot1_lasso*(Zb_BCa_boot1_lasso-Z)))))
alpha2_BCa_boot1_lasso<-pnorm(Zb_BCa_boot1_lasso+((Zb_BCa_boot1_lasso+Z)/(1-
(a_BCa_boot1_lasso*(Zb_BCa_boot1_lasso+Z)))))
BCa_boot1_lasso_lb[ds,]<-apply(boot1_lasso, 2, quantile,
probs=alpha1_BCa_boot1_lasso)
BCa_boot1_lasso_ub[ds,]<-apply(boot1_lasso, 2, quantile,
probs=alpha2_BCa_boot1_lasso)
#BCa boot2
jackknif_sample_mean_boot2_lasso<-as.data.frame(apply(boot2_lasso, 2, mean))
a_BCa_boot2_lasso<-sum((jackknif_sample_mean_boot2_lasso-
boot2_lasso)^3)/(6*((sum((jackknif_sample_mean_boot2_lasso-boot2_lasso)^2))^1.5))
```

```
b_BCa_boot2_lasso<-(sum(boot2_lasso[,-1]<beta))/(B*p)
Zb_BCa_boot2_lasso<-qnorm(b_BCa_boot2_lasso)
Z<-qnorm(1-alpha/2)
alpha1_BCa_boot2_lasso<-pnorm(Zb_BCa_boot2_lasso+((Zb_BCa_boot2_lasso-Z)/(1-
(a_BCa_boot2_lasso*(Zb_BCa_boot2_lasso-Z)))))
alpha2_BCa_boot2_lasso<-pnorm(Zb_BCa_boot2_lasso+((Zb_BCa_boot2_lasso+Z)/(1-
(a_BCa_boot2_lasso*(Zb_BCa_boot2_lasso+Z)))))
BCa_boot2_lasso_lb[ds,]<-apply(boot2_lasso, 2, quantile,
probs=alpha1_BCa_boot2_lasso)
BCa_boot2_lasso_ub[ds,]<-apply(boot2_lasso, 2, quantile,
probs=alpha2_BCa_boot2_lasso)
#BCa boot3
jackknif_sample_mean_boot3_lasso<-as.data.frame(apply(boot3_lasso, 2, mean))
a_BCa_boot3_lasso<-sum((jackknif_sample_mean_boot3_lasso-
boot3_lasso)^3)/(6*((sum((jackknif_sample_mean_boot3_lasso-boot3_lasso)^2))^1.5))
b_BCa_boot3_lasso<-(sum(boot3_lasso[,-1]<beta))/(B*p)
Zb_BCa_boot3_lasso<-qnorm(b_BCa_boot3_lasso)
Z<-qnorm(1-alpha/2)
alpha1_BCa_boot3_lasso<-pnorm(Zb_BCa_boot3_lasso+((Zb_BCa_boot3_lasso-Z)/(1-
(a_BCa_boot3_lasso*(Zb_BCa_boot3_lasso-Z)))))
alpha2_BCa_boot3_lasso<-pnorm(Zb_BCa_boot3_lasso+((Zb_BCa_boot3_lasso+Z)/(1-
(a_BCa_boot3_lasso*(Zb_BCa_boot3_lasso+Z)))))
BCa_boot3_lasso_lb[ds,]<-apply(boot3_lasso, 2, quantile,
probs=alpha1_BCa_boot3_lasso)
BCa_boot3_lasso_ub[ds,]<-apply(boot3_lasso, 2, quantile,
probs=alpha2_BCa_boot3_lasso)
#BC boot1
alpha1_BC_boot1_lasso<-pnorm((2*Zb_BCa_boot1_lasso)-Z)
alpha2_BC_boot1_lasso<-pnorm((2*Zb_BCa_boot1_lasso)+Z)
BC_boot1_lasso_lb[ds,]<-apply(boot1_lasso, 2, quantile,
probs=alpha1_BC_boot1_lasso)
```

```
BC_boot1_lasso_ub[ds,]<-apply(boot1_lasso, 2, quantile,
probs=alpha2_BC_boot1_lasso)
#BC boot2
alpha1_BC_boot2_lasso<-pnorm((2*Zb_BCa_boot2_lasso)-Z)
alpha2_BC_boot2_lasso<-pnorm((2*Zb_BCa_boot2_lasso)+Z)
BC_boot2_lasso_lb[ds,]<-apply(boot2_lasso, 2, quantile,
probs=alpha1_BC_boot2_lasso)
BC_boot2_lasso_ub[ds,]<-apply(boot2_lasso, 2, quantile,
probs=alpha2_BC_boot2_lasso)
#BC boot3
alpha1_BC_boot3_lasso<-pnorm((2*Zb_BCa_boot3_lasso)-Z)
alpha2_BC_boot3_lasso<-pnorm((2*Zb_BCa_boot3_lasso)+Z)
BC_boot3_lasso_lb[ds,]<-apply(boot3_lasso, 2, quantile,
probs=alpha1_BC_boot3_lasso)
BC_boot3_lasso_ub[ds,]<-apply(boot3_lasso, 2, quantile,
probs=alpha2_BC_boot3_lasso)

#AdaptLasso
# CI-simple
simple_boot1_adaptlasso_lb[ds,]<-apply(boot1_adaptlasso, 2, quantile,
probs=alpha/2)
simple_boot1_adaptlasso_ub[ds,]<-apply(boot1_adaptlasso, 2, quantile, probs=1-
alpha/2)
simple_boot2_adaptlasso_lb[ds,]<-apply(boot2_adaptlasso, 2, quantile,
probs=alpha/2)
simple_boot2_adaptlasso_ub[ds,]<-apply(boot2_adaptlasso, 2, quantile, probs=1-
alpha/2)
simple_boot3_adaptlasso_lb[ds,]<-apply(boot3_adaptlasso, 2, quantile,
probs=alpha/2)
simple_boot3_adaptlasso_ub[ds,]<-apply(boot3_adaptlasso, 2, quantile, probs=1-
alpha/2)
```

```
#BCa boot1
jackknif_sample_mean_boot1_adaptlasso<-as.data.frame(apply(boot1_adaptlasso, 2,
mean))
a_BCa_boot1_adaptlasso<-sum((jackknif_sample_mean_boot1_adaptlasso-
boot1_adaptlasso)^3)/(6*((sum((jackknif_sample_mean_boot1_adaptlasso-
boot1_adaptlasso)^2))^1.5))
b_BCa_boot1_adaptlasso<-(sum(boot1_adaptlasso[,-1]<beta))/(B*p)
Zb_BCa_boot1_adaptlasso<-qnorm(b_BCa_boot1_adaptlasso)
Z<-qnorm(1-alpha/2)
alpha1_BCa_boot1_adaptlasso<-
pnorm(Zb_BCa_boot1_adaptlasso+((Zb_BCa_boot1_adaptlasso-Z)/(1-
(a_BCa_boot1_adaptlasso*(Zb_BCa_boot1_adaptlasso-Z)))))
alpha2_BCa_boot1_adaptlasso<-
pnorm(Zb_BCa_boot1_adaptlasso+((Zb_BCa_boot1_adaptlasso+Z)/(1-
(a_BCa_boot1_adaptlasso*(Zb_BCa_boot1_adaptlasso+Z)))))
BCa_boot1_adaptlasso_lb[ds,]<-apply(boot1_adaptlasso, 2, quantile,
probs=alpha1_BCa_boot1_adaptlasso)
BCa_boot1_adaptlasso_ub[ds,]<-apply(boot1_adaptlasso, 2, quantile,
probs=alpha2_BCa_boot1_adaptlasso)
#BCa boot2
jackknif_sample_mean_boot2_adaptlasso<-as.data.frame(apply(boot2_adaptlasso, 2,
mean))
a_BCa_boot2_adaptlasso<-sum((jackknif_sample_mean_boot2_adaptlasso-
boot2_adaptlasso)^3)/(6*((sum((jackknif_sample_mean_boot2_adaptlasso-
boot2_adaptlasso)^2))^1.5))
b_BCa_boot2_adaptlasso<-(sum(boot2_adaptlasso[,-1]<beta))/(B*p)
Zb_BCa_boot2_adaptlasso<-qnorm(b_BCa_boot2_adaptlasso)
Z<-qnorm(1-alpha/2)
alpha1_BCa_boot2_adaptlasso<-
pnorm(Zb_BCa_boot2_adaptlasso+((Zb_BCa_boot2_adaptlasso-Z)/(1-
(a_BCa_boot2_adaptlasso*(Zb_BCa_boot2_adaptlasso-Z)))))
```

```
alpha2_BCa_boot2_adaptlasso<-
pnorm(Zb_BCa_boot2_adaptlasso+((Zb_BCa_boot2_adaptlasso+Z)/(1-
(a_BCa_boot2_adaptlasso*(Zb_BCa_boot2_adaptlasso+Z)))))
BCa_boot2_adaptlasso_lb[ds,]<-apply(boot2_adaptlasso, 2, quantile,
probs=alpha1_BCa_boot2_adaptlasso)
BCa_boot2_adaptlasso_ub[ds,]<-apply(boot2_adaptlasso, 2, quantile,
probs=alpha2_BCa_boot2_adaptlasso)
#BCa boot3
jackknif_sample_mean_boot3_adaptlasso<-as.data.frame(apply(boot3_adaptlasso, 2,
mean))
a_BCa_boot3_adaptlasso<-sum((jackknif_sample_mean_boot3_adaptlasso-
boot3_adaptlasso)^3)/(6*((sum((jackknif_sample_mean_boot3_adaptlasso-
boot3_adaptlasso)^2))^1.5))
b_BCa_boot3_adaptlasso<-(sum(boot3_adaptlasso[,-1]<beta))/(B*p)
Zb_BCa_boot3_adaptlasso<-qnorm(b_BCa_boot3_adaptlasso)
Z<-qnorm(1-alpha/2)
alpha1_BCa_boot3_adaptlasso<-
pnorm(Zb_BCa_boot3_adaptlasso+((Zb_BCa_boot3_adaptlasso-Z)/(1-
(a_BCa_boot3_adaptlasso*(Zb_BCa_boot3_adaptlasso-Z)))))
alpha2_BCa_boot3_adaptlasso<-
pnorm(Zb_BCa_boot3_adaptlasso+((Zb_BCa_boot3_adaptlasso+Z)/(1-
(a_BCa_boot3_adaptlasso*(Zb_BCa_boot3_adaptlasso+Z)))))
BCa_boot3_adaptlasso_lb[ds,]<-apply(boot3_adaptlasso, 2, quantile,
probs=alpha1_BCa_boot3_adaptlasso)
BCa_boot3_adaptlasso_ub[ds,]<-apply(boot3_adaptlasso, 2, quantile,
probs=alpha2_BCa_boot3_adaptlasso)
#BC boot1
alpha1_BC_boot1_adaptlasso<-pnorm((2*Zb_BCa_boot1_adaptlasso)-Z)
alpha2_BC_boot1_adaptlasso<-pnorm((2*Zb_BCa_boot1_adaptlasso)+Z)
BC_boot1_adaptlasso_lb[ds,]<-apply(boot1_adaptlasso, 2, quantile,
probs=alpha1_BC_boot1_adaptlasso)
```

```
BC_boot1_adaptlasso_ub[ds,]<-apply(boot1_adaptlasso, 2, quantile,
probs=alpha2_BC_boot1_adaptlasso)
#BC boot2
alpha1_BC_boot2_adaptlasso<-pnorm((2*Zb_BCa_boot2_adaptlasso)-Z)
alpha2_BC_boot2_adaptlasso<-pnorm((2*Zb_BCa_boot2_adaptlasso)+Z)
BC_boot2_adaptlasso_lb[ds,]<-apply(boot2_adaptlasso, 2, quantile,
probs=alpha1_BC_boot2_adaptlasso)
BC_boot2_adaptlasso_ub[ds,]<-apply(boot2_adaptlasso, 2, quantile,
probs=alpha2_BC_boot2_adaptlasso)
#BC boot3
alpha1_BC_boot3_adaptlasso<-pnorm((2*Zb_BCa_boot3_adaptlasso)-Z)
alpha2_BC_boot3_adaptlasso<-pnorm((2*Zb_BCa_boot3_adaptlasso)+Z)
BC_boot3_adaptlasso_lb[ds,]<-apply(boot3_adaptlasso, 2, quantile,
probs=alpha1_BC_boot3_adaptlasso)
BC_boot3_adaptlasso_ub[ds,]<-apply(boot3_adaptlasso, 2, quantile,
probs=alpha2_BC_boot3_adaptlasso)
}
stop<-Sys.time()
time <-stop-start
```

## ประวัติผู้เขียน

| | |
|---|---|
| **ชื่อ-สกุล** | นายภูวกร นิธิศนทีกุล |
| **วัน เดือน ปี เกิด** | 28 พฤษภาคม 2538 |
| **สถานที่เกิด** | พิจิตร |
| **วุฒิการศึกษา** | ระดับมัธยมศึกษาตอนปลาย โรงเรียนยุพราชวิทยาลัย |
| | ปริญญาตรี คณะวิทยาศาสตร์ สาขาคณิตศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย |
| **ที่อยู่ปัจจุบัน** | 43 หมู่ที่4 ตำบลโพธิ์ไทรงาม อำเภอบึงนาราง จังหวัดพิจิตร 66130 |