บทที่ 2

ทฤษฎีที่เกี่ยวข้องกับการวิจัย

ในบทนี้จะกล่าวถึงการประมาณค่าสัมประสิทธิ์การถดถอยพหุแบบช่วงของตัวแบบการถดถอยเชิงเส้น ได้แก่วิธีการประมาณค่าแบบช่วงด้วยการแจกแจงที โดยใช้ตัวประมาณกำลังสองน้อยที่สุด และตัวประมาณริดจ์รีเกรสชัน และวิธีการประมาณค่าแบบช่วงด้วยวิธีบูตสแตรป โดยใช้ตัวประมาณริดจ์รีเกรสชัน ซึ่งมีรายละเอียดต่างๆ ดังต่อไปนี้

2.1 ตัวแบบทั่วไป

ตัวแบบทั่วไปของการวิเคราะห์การถดถอยพหุเชิงเส้นสามารถเขียนได้ดังนี้

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ : \\ : \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ : & : & \dots & : \\ : & : & \dots & : \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ : \\ : \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ : \\ : \\ \varepsilon_n \end{bmatrix}$$

โดยที่ $\underset{\sim}{y}$ เป็นเวกเตอร์ของตัวแปรตามขนาด $n \times 1$

X เป็นเมตริกซ์ของตัวแปรอิสระขนาด $n \times (p+1)$ และ $X'X$ มี full rank $= p+1$

$\underset{\sim}{\beta}$ เป็นเวกเตอร์ของสัมประสิทธิ์การถดถอยพหุขนาด $(p+1) \times 1$

$\underset{\sim}{\varepsilon}$ เป็นเวกเตอร์ของค่าความคลาดเคลื่อนที่เกิดขึ้นขนาด $n \times 1$

n เป็นจำนวนค่าสังเกต

p เป็นจำนวนตัวแปรอิสระ

โดยที่ $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$, $E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}') = \sigma^2 I_n$

2.2 วิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุแบบช่วง

วิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุแบบช่วง ในกรณีที่ตัวแปรอิสระเกิดพหุสัมพันธ์กันทั้ง 3 วิธี มีรายละเอียดดังต่อไปนี้

2.2.1 <u>การประมาณค่าแบบช่วงด้วยการแจกแจงที โดยใช้ตัวประมาณกำลังสองน้อยที่สุด</u>

การประมาณค่าสัมประสิทธิ์การถดถอยพหุ โดยใช้ตัวประมาณกำลังสองน้อยที่สุด

จากสมการ $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ เมื่อ $\underset{\sim}{\varepsilon} \sim N\left(\underset{\sim}{0}, \sigma^2 I_n\right)$ จะได้ว่าตัวประมาณกำลังสองน้อยที่สุดของ $\underset{\sim}{\beta}$ คือ $\hat{\underset{\sim}{\beta}}$ ที่ทำให้ผลรวมกำลังสองของความคลาดเคลื่อนกำลังสองมีค่าน้อยที่สุด

$$
\begin{aligned}
SSE &= \left(\hat{\underset{\sim}{\varepsilon}}'\hat{\underset{\sim}{\varepsilon}}\right) \\
&= \left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}\right)'\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}\right) \\
&= \left(\underset{\sim}{y}' - \hat{\underset{\sim}{\beta}}'X'\right)\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}\right) \\
&= \underset{\sim}{y}'\underset{\sim}{y} - \underset{\sim}{y}'X\hat{\underset{\sim}{\beta}} - \hat{\underset{\sim}{\beta}}'X'\underset{\sim}{y} + \hat{\underset{\sim}{\beta}}'X'X\hat{\underset{\sim}{\beta}} \\
&= \underset{\sim}{y}'\underset{\sim}{y} - 2\hat{\underset{\sim}{\beta}}'X'\underset{\sim}{y} + \hat{\underset{\sim}{\beta}}'X'X\hat{\underset{\sim}{\beta}}
\end{aligned}
$$

การหาผลรวมกำลังสองของความคลาดเคลื่อนมีค่าน้อยที่สุด ได้โดยการหาอนุพันธ์ของ SSE เทียบค่า β_i ; $i = 0,1,\ldots\ldots,p$ (p เป็นจำนวนตัวแปรอิสระ) แล้วกำหนดเท่ากับ 0

$$
\frac{\partial(SSE)}{\partial\hat{\underset{\sim}{\beta}}} = \frac{\partial}{\partial\hat{\underset{\sim}{\beta}}}\left(\underset{\sim}{y}'\underset{\sim}{y} - 2\hat{\underset{\sim}{\beta}}'X'\underset{\sim}{y} + \hat{\underset{\sim}{\beta}}'X'X\hat{\underset{\sim}{\beta}}\right)
$$

$$
\begin{aligned}
-2X'\underset{\sim}{y} + 2X'X\hat{\underset{\sim}{\beta}} &= \underset{\sim}{0} \\
(X'X)\hat{\underset{\sim}{\beta}} &= X'\underset{\sim}{y} \\
\hat{\underset{\sim}{\beta}} &= (X'X)^{-1}X'\underset{\sim}{y}
\end{aligned}
$$

และความแปรปรวนของตัวประมาณ $\hat{\underset{\sim}{\beta}}$ คือ

$$\begin{aligned} \mathrm{var}(\hat{\underset{\sim}{\beta}}) &= \mathrm{var}\left[(X'X)^{-1}X'\underset{\sim}{y}\right] \\ &= (X'X)^{-1}X'\mathrm{var}(\underset{\sim}{y})X(X'X)^{-1} \\ &= (X'X)^{-1}X'\sigma^2X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1} \end{aligned}$$

กำหนดในรูปแบบเมตริกซ์ดังนี้

$$(X'X)^{-1} = \begin{bmatrix} c_{00} & c_{10} & \cdots & c_{1p} \\ c_{01} & c_{11} & \cdots & c_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ c_{p0} & c_{p1} & \cdots & c_{pp} \end{bmatrix}$$

$\mathrm{var}(\beta_i) = \sigma^2(c_{ii})$; $i = 0,1,\ldots\ldots,p$ (p คือจำนวนตัวแปรอิสระ)

การประมาณค่าแบบช่วงด้วยการแจกแจงที โดยใช้ตัวประมาณกำลังสองน้อยที่สุด

$$T_i = \frac{\left(\hat{\beta}_i - \beta_i\right)}{S_{ii}} \quad ; \quad i = 0,1\ldots\ldots,p$$

$$P\left(-t_{(n-p-1),\alpha/2} < T_i < t_{(n-p-1),\alpha/2}\right) = 1-\alpha$$

$$P\left(-t_{(n-p-1),\alpha/2} < \frac{\left(\hat{\beta}_i - \beta_i\right)}{S_{ii}} < t_{(n-p-1),\alpha/2}\right) = 1-\alpha$$

$$P\left(\hat{\beta}_i - t_{(n-p-1),\alpha/2}S_{ii} < \beta_i < \hat{\beta}_i + t_{(n-p-1),\alpha/2}S_{ii}\right) = 1-\alpha$$

นั่นคือ ช่วงความเชื่อมั่น $100(1-\alpha)\%$ สำหรับ β_i ; $i = 1,2,\ldots\ldots,p$

ขีดจำกัดความเชื่อมั่นบน (PU) คือ $\hat{\beta}_i + t_{(n-p-1),\alpha/2}S_{ii}$

ขีดจำกัดความเชื่อมั่นล่าง (PL) คือ $\hat{\beta}_i - t_{(n-p-1),\alpha/2}S_{ii}$

โดยที่

$\hat{\beta}_i$ เป็นตัวประมาณโดยวิธีกำลังสองน้อยที่สุดของ β_i

$s_{ii} = \sqrt{\widehat{\mathrm{var}}(\hat{\beta}_i)}$ เป็นความคลาดเคลื่อนมาตรฐานของ $\hat{\beta}_i$

$$\widehat{\mathrm{var}(\beta_i)} = S^2(c_{ii})$$

$$S^2 = \hat{\sigma}^2 = \frac{1}{n-p-1}(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}})'(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}})$$

$$= \frac{1}{n-p-1}(\underset{\sim}{y}'\underset{\sim}{y} - 2\hat{\underset{\sim}{\beta}}'(X'X) + \hat{\underset{\sim}{\beta}}'(X'X)\hat{\underset{\sim}{\beta}})$$

p = จำนวนตัวแปรอิสระ

$t_{\alpha/2}$ เป็นเปอร์เซ็นไทล์ของการแจกแจงที (Percentiles of the t distribution)

2.2.2 <u>การประมาณค่าแบบช่วงด้วยการแจกแจงที โดยใช้ตัวประมาณริดจ์รีเกรสชัน</u>

การประมาณค่าสัมประสิทธิ์การถดถอยพหุ โดยใช้ตัวประมาณริดจ์รีเกรสชัน

Hoerl และ Kennard (1970) ได้ศึกษาวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุที่ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองไม่มากกว่าวิธีกำลังสองน้อยที่สุด เมื่อตัวแปรอิสระเกิดพหุสัมพันธ์กัน ซึ่งวิธีนี้เรียกว่าวิธีริดจ์รีเกรสชัน วิธีการนี้เป็นวิธีที่แก้ปัญหาการเกิดพหุสัมพันธ์โดยใช้หลักการดังนี้ ถ้า $|X'X|$ มีค่าเข้าใกล้ศูนย์ ซึ่งจะทำให้ค่า $(X'X)^{-1}$ หาค่าได้ยากหรือมีความคาดเคลื่อนมาก ดังนั้นวิธีการแก้ปัญหาโดยการทำให้ $(X'X)^{-1}$ มีค่าเพิ่มขึ้น กล่าวคือการบวกค่าคงที่ k ที่มากกว่าศูนย์เข้ากับสมาชิกทุกตัวบนเส้นทะแยงมุมของ $X'X$ ดังนั้นเราจะได้ค่าสัมประสิทธิ์การถดถอยพหุของวิธีริดจ์รีเกรสชัน ดังนี้

$$\hat{\underset{\sim}{\beta}}(R) = (X'X + kI)^{-1}X'\underset{\sim}{y}$$

และความแปรปรวนของ $\hat{\underset{\sim}{\beta}}(R)$ คือ

$$\mathrm{var}\left(\hat{\underset{\sim}{\beta}}(R)\right) = \mathrm{var}\left[(X'X + kI)^{-1}X'\underset{\sim}{y}\right]$$

$$= (X'X + kI)^{-1}X'\mathrm{var}(\underset{\sim}{y})X(X'X + kI)^{-1}$$

$$= \sigma^2(X'X + kI)^{-1}X'X(X'X + kI)^{-1}$$

กำหนดในรูปแบบเมตริกซ์ดังนี้

$$(X'X+kI)^{-1}X'X(X'X+kI)^{-1} = \begin{bmatrix} R_{00} & R_{10} \dots R_{1p} \\ R_{01} & R_{11} \dots R_{2p} \\ \vdots & \vdots \quad \dots \quad \vdots \\ R_{p0} & R_{p1} \dots R_{pp} \end{bmatrix}$$

$$\text{var}(\beta_i) = \sigma^2(R_{ii}) \quad ; i = 0,1,\dots\dots,p$$ (p คือจำนวนตัวแปรอิสระ)

<u>ตัวประมาณริดจ์ (Ridge Estimator)</u>

ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุด้วยวิธีริดจ์รีเกรสชันนั้นจะต้องกำหนดค่าคงที่ k ในการกำหนดค่า k นี้มีได้หลายวิธีด้วยกัน แต่ในการศึกษาครั้งนี้จะใช้วิธีการประมาณค่า k ด้วยกัน 2 วิธี ดังต่อไปนี้

1. วิธี Hoerl , Kernard and Baldwin (HKB)

วิธีนี้เสนอโดย Hoerl , Kernard and Baldwin (1975) ซึ่งแนะนำให้เลือกค่า k ดังนี้

$$k = \frac{pS^2}{\hat{\underset{\sim}{\beta}}'\hat{\underset{\sim}{\beta}}}$$

โดยที่

$$\hat{\underset{\sim}{\beta}} = (X'X)^{-1}X'\underset{\sim}{y}$$

$$S^2 = \hat{\sigma}^2 = \frac{1}{n-p-1}(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}})'(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}})$$

$$= \frac{1}{n-p-1}(\underset{\sim}{y}'\underset{\sim}{y}-2\hat{\underset{\sim}{\beta}}'(X'X)+\hat{\underset{\sim}{\beta}}'(X'X)\hat{\underset{\sim}{\beta}})$$

p = จำนวนตัวแปรอิสระ

2. วิธี Lawless and Wang (LW)

วิธีนี้เสนอโดย Lawless and Wang (1976) โดยกำหนดค่า k ดังนี้

$$k = \frac{pS^2}{\hat{\underset{\sim}{\beta}}'X'X\hat{\underset{\sim}{\beta}}}$$

โดยที่

$$\hat{\underset{\sim}{\beta}} = (X'X)^{-1}X'\underset{\sim}{y}$$

$$S^2 = \hat{\sigma}^2 = \frac{1}{n-p-1}(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}})'(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}})$$

$$= \frac{1}{n-p-1}(\underset{\sim}{y}'\underset{\sim}{y} - 2\hat{\underset{\sim}{\beta}}'(X'X) + \hat{\underset{\sim}{\beta}}'(X'X)\hat{\underset{\sim}{\beta}})$$

p = จำนวนตัวแปรอิสระ

การประมาณค่าแบบช่วงด้วยการแจกแจงที โดยใช้ตัวประมาณริดจ์รีเกรสชัน

$$T_i = \frac{\left(\hat{\beta}(R)_i - \beta_i\right)}{S_{ii}} \qquad ; \ i = 0,1,\ldots\ldots,p$$

$$P\left(-t_{(n-p-1),\alpha/2} < T_i < t_{(n-p-1),\alpha/2}\right) = 1-\alpha$$

$$P\left(-t_{(n-p-1),\alpha/2} < \frac{\left(\hat{\beta}(R)_i - \beta_i\right)}{S_{ii}} < t_{(n-p-1),\alpha/2}\right) = 1-\alpha$$

$$P\left(\hat{\beta}(R)_i - t_{(n-p-1),\alpha/2}S_{ii} < \beta_i < \hat{\beta}(R)_i + t_{(n-p-1),\alpha/2}S_{ii}\right) = 1-\alpha$$

นั่นคือช่วงความเชื่อมั่น $100(1-\alpha)\%$ สำหรับ β_i ; $i = 1,2,\ldots\ldots,p$

ขีดจำกัดความเชื่อมั่นบน (PU) คือ $\hat{\beta}(R)_i + t_{(n-p-1),\alpha/2}S_{ii}$

ขีดจำกัดความเชื่อมั่นล่าง (PL) คือ $\hat{\beta}(R)_i - t_{(n-p-1),\alpha/2}S_{ii}$

โดยที่

$\hat{\beta}(R)_i$ เป็นตัวประมาณโดยวิธีริดจ์ รีเกรสชันของ β_i

$S_{ii} = \sqrt{\hat{\mathrm{var}}\left(\hat{\beta}(R)_i\right)}$ เป็นความคลาดเคลื่อนมาตรฐานของ $\hat{\beta}(R)_i$

$\hat{\mathrm{var}}(\hat{\beta}(R)_i) = S^2R_{ii}$

$$S^2 = \hat{\sigma}^2 = \frac{1}{n-p-1}(\underset{\sim}{y}-X\underset{\sim}{\hat{\beta}})'(\underset{\sim}{y}-X\underset{\sim}{\hat{\beta}})$$
$$= \frac{1}{n-p-1}(\underset{\sim}{y}'\underset{\sim}{y}-2\underset{\sim}{\hat{\beta}}'(X'X)+\underset{\sim}{\hat{\beta}}'(X'X)\underset{\sim}{\hat{\beta}})$$

p = จำนวนตัวแปรอิสระ

$t_{\alpha/2}$ เป็นเปอร์เซ็นไทล์ของการแจกแจงที

2.2.3 <u>การประมาณค่าแบบช่วงด้วยวิธีบูตสแตรป โดยใช้ตัวประมาณริดจ์รีเกรสชัน</u>

วิธีบูตสแตรป (Bootstrap)

วิธีการหาตัวประมาณของพารามิเตอร์วิธีนี้เป็นวิธีที่เสนอขึ้นโดย แบรดเลย์ เอฟรอน (Bradley Efron) ในปี ค.ศ. 1979 โดยมีหลักเกณฑ์ดังนี้คือ จากข้อมูลที่เก็บรวบรวมมาจะทำการสุ่มตัวอย่างแบบใส่คืน (with replacement) โดยสุ่มเป็นจำนวนเท่ากับจำนวนขนาดตัวอย่างหรือข้อมูลที่มีอยู่ เพื่อสร้างข้อมูลชุดใหม่แล้วนำมาใช้ในการประมาณค่าพารามิเตอร์ที่สนใจ

สำหรับการหาตัวประมาณด้วยวิธีนี้ มีขั้นตอนการทำงานดังนี้

1. สร้างเลขสุ่มที่มีค่าอยู่ระหว่าง 0 กับ 1 เพื่อนำไปใช้ในการสุ่มตัวอย่างแบบใส่คืน

2. จากตัวอย่างที่สุ่มได้แต่ละชุด นำมาหาค่าประมาณของพารามิเตอร์

รายละเอียดของวิธีการสุ่มตัวอย่างแบบใส่คืนและรายละเอียดต่างๆ จะเสนอไว้ในภาคผนวก ก.

การประมาณค่าสัมประสิทธิ์การถดถอยพหุด้วยวิธีบูตสแตรป

จากตัวแบบ

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

ความคลาดเคลื่อนเป็นตัวแปรสุ่มแบบต่อเนื่องที่มีการแจกแจงแบบเดียวกัน และเป็นอิสระซึ่งกันและกัน(identically independent distribution) นั่นคือ $\underset{\sim}{\varepsilon} \sim F$, $i = 1,2,\ldots\ldots,n$ เมื่อ F เป็นการแจกแจงความน่าจะเป็นที่ไม่ทราบ

การหาตัวประมาณกระทำตามขั้นตอนดังต่อไปนี้

จากข้อมูล X และ $\underset{\sim}{y}$ ที่ได้จะนำมาหาค่าพารามิเตอร์

1. คำนวณหาค่า $\hat{\underset{\sim}{\beta}}(R)$ โดยวิธีริดจ์รีเกรสชัน

$$\hat{\underset{\sim}{\beta}}(R) = (X'X + kI)^{-1} X'\underset{\sim}{y}$$

2. จากค่า $\hat{\underset{\sim}{\beta}}(R)$ ที่ได้โดยวิธีริดจ์รีเกรสชันนำมาหา $\hat{\underset{\sim}{\varepsilon}} = \underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}(R)$

3. จากค่า $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots\ldots, \hat{\varepsilon}_n$ ทำการสุ่มตัวอย่างขนาด n แบบใส่คืนได้ $\hat{\varepsilon}_1^*, \hat{\varepsilon}_2^*, \ldots\ldots, \hat{\varepsilon}_n^*$ เมื่อ $\hat{\varepsilon}_j^*$ คือตัวอย่างที่สุ่มได้ตัวที่ j จาก $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots\ldots, \hat{\varepsilon}_n$

4. สร้างรูปแบบ (model) ใหม่

$$\underset{\sim}{y}^* = X\hat{\underset{\sim}{\beta}}(R) + \underset{\sim}{\varepsilon}^*$$

5. นำค่า $\underset{\sim}{y}^*$ และ X คำนวณหาค่า $\hat{\beta}^*(R)_i$ โดยวิธีริดจ์รีเกรสชัน

6. กระทำตามขั้นตอนในข้อ 3 - 5 ซ้ำกันเท่ากับจำนวนครั้งที่ต้องการบูตสแตรป

<u>การหาจำนวนครั้งของการทำบูตสแตรป ในแต่ละขนาดตัวอย่างในการวิจัยนี้</u>

จากการประมาณค่าแบบช่วงด้วยวิธีบูตสแตรป โดยใช้ตัวประมาณริดจ์รีเกรสชัน ซึ่งมีสูตรการประมาณดังต่อไปนี้

$$u_i = S_{ii} z_{1-\alpha/2} - S_{ii} \frac{\left(\hat{p}^*\left(z_{1-\alpha/2}\right) - (1-\alpha/2)\right)}{\varphi\left(z_{\alpha/2}\right)}$$

$$v_i = S_{ii} z_{1-\alpha/2} + S_{ii} \frac{\left(\hat{p}^*\left(z_{\alpha/2}\right) - \alpha/2\right)}{\varphi\left(z_{\alpha/2}\right)} \quad ; i = 1,2,\ldots,p$$

โดยที่ $\hat{p}^*\left(z_{1-\alpha/2}\right)$ และ $\hat{p}^*\left(z_{\alpha/2}\right)$ ประมาณได้จากวิธีการบูตสแตรปดังนี้

$$\hat{p}^*(z) = \frac{\#\{t^* \le z\}}{B}$$

เมื่อ B คือ จำนวนครั้งของการทำบูตสแตรป

ดังนั้นผู้วิจัยได้ทำการหาจำนวนครั้งของการทำบูตสแตรปในแต่ละขนาดตัวอย่าง ดังขั้นตอนต่อไปนี้

ในการวิจัยครั้งนี้ผู้วิจัยได้ทำการศึกษาที่ตัวแบบเชิงเส้นของตัวแปรอิสระ 3 ตัวดังนี้

$$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \beta_3 x_{3j} + \varepsilon_j \quad ; \; j = 1,2,\ldots\ldots,n$$

1. ต้องกำหนดจำนวนครั้งของการทำบูตสแตรปเริ่มต้นไว้ก่อน
2. เราจะทำการประมาณค่า $\hat{p}^*(z)$ ด้วยวิธีบูตสแตรปแล้วนั้นจะได้ค่า $\hat{p}^*(z)$ ของแต่ละค่า β_i ; $i = 1,2,3$
3. นำค่า $\hat{p}^*(z)$ ทั้ง 3 ค่ามาทำการหาค่าเฉลี่ย
4. พิจารณาดูว่าค่าเฉลี่ยของค่า $\hat{p}^*(z)$ มีค่าที่คงที่แล้วหรือยัง ถ้าคงที่แล้วไปขั้นตอนที่ 6 ถ้าไม่คงที่ไปยังขั้นตอนที่ 5
5. เพิ่มจำนวนครั้งของการทำบูตสแตรป และกลับไปยังขั้นตอนที่ 1
6. นำจำนวนครั้งของการบูตสแตรปที่ได้นำไปใช้ในการวิจัยของแต่ละขนาดตัวอย่าง

จากขั้นตอนดังกล่าวข้างต้น ผู้วิจัยได้ทำการทดลองเพื่อจำนวนครั้งของการทำบูตสแตรปในแต่ละขนาดตัวอย่าง ได้ผลสรุปดังต่อไปนี้

- ขนาดตัวอย่างเท่ากับ 15 จำนวนครั้งของการทำบูตสแตรปเท่ากับ 300 ครั้ง
- ขนาดตัวอย่างเท่ากับ 30 จำนวนครั้งของการทำบูตสแตรปเท่ากับ 500 ครั้ง
- ขนาดตัวอย่างเท่ากับ 40 จำนวนครั้งของการทำบูตสแตรปเท่ากับ 600 ครั้ง
- ขนาดตัวอย่างเท่ากับ 50 จำนวนครั้งของการทำบูตสแตรปเท่ากับ 700 ครั้ง

การประมาณค่าแบบช่วงด้วยวิธีบูตสแตรป โดยใช้ตัวประมาณริดจ์รีเกรสชัน

ในการประมาณช่วงความเชื่อมั่นที่ Rayner (1989) ได้ทำพัฒนาขึ้นนั้นเป็นวิธีการหาช่วงความเชื่อมั่นในกรณีของการประมาณริดจ์รีเกรสชัน โดยการรวมเทคนิคบูตสแตรป และวิธีการ Edgeworth Expansion

การประมาณช่วงความเชื่อมั่นของ β_i โดยวิธีประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีริดจ์รีเกรสชัน $\hat{\beta}(R)_i$ มีช่วงความเชื่อมั่น $100(1-\alpha)\%$ สำหรับ β_i ; $i = 1,2,\ldots..,p$

ขีดจำกัดความเชื่อมั่นบน (PU) คือ $\hat{\beta}(R)_i + v_i$

ขีดจำกัดความเชื่อมั่นล่าง (PL) คือ $\hat{\beta}(R)_i - u_i$

$$u_i = S_{ii} z_{1-\alpha/2} - S_{ii} \frac{\left(\hat{p}^*\left(z_{1-\alpha/2}\right) - (1-\alpha/2)\right)}{\varphi\left(z_{\alpha/2}\right)}$$

$$v_i = S_{ii} z_{1-\alpha/2} + S_{ii} \frac{\left(\hat{p}^*\left(z_{\alpha/2}\right) - \alpha/2\right)}{\varphi\left(z_{\alpha/2}\right)} \qquad ; i = 1,2,\ldots,p$$

โดยที่

$\hat{\beta}(R)_i$ เป็นตัวประมาณโดยวิธีริดจ์รีเกรสชันของ β_i

$S_{ii} = \sqrt{\hat{\text{var}}\left(\hat{\beta}(R)_i\right)}$ เป็นความคลาดเคลื่อนมาตรฐานของ $\hat{\beta}(R)_i$

$\hat{\text{var}}(\hat{\beta}(R)_i) = S^2 R_{ii}$

$$S^2 = \hat{\sigma}^2 = \frac{1}{n-p-1}(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}})'(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}})$$

$$= \frac{1}{n-p-1}(\underset{\sim}{y}'\underset{\sim}{y} - 2\hat{\underset{\sim}{\beta}}'(X'X) + \hat{\underset{\sim}{\beta}}'(X'X)\hat{\underset{\sim}{\beta}})$$

$z_{\alpha/2}$ และ $z_{1-\alpha/2}$ เป็นเปอร์เซ็นไทล์ของการแจกแจงแบบปกติมาตรฐาน

$\varphi(z)$ เป็นฟังก์ชันความหนาแน่นของการแจกแจงปกติมาตรฐาน โดยที่

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \qquad ; -\infty < z < \infty$$

การหาค่า $\hat{p}^*(z)$ โดยนำวิธีการของบูตสแตรปเข้ามาช่วย ซึ่งมีขั้นตอนดังต่อไปนี้

1. จากค่า $\hat{\underset{\sim}{\beta}}(R)$ ที่ได้โดยวิธีริดจ์รีเกรสชันนำมาหา $\hat{\underset{\sim}{\varepsilon}} = \underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}(R)$
2. จากค่า $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots\ldots, \hat{\varepsilon}_n$ ทำการสุ่มตัวอย่างขนาด n แบบใส่คืนได้ $\hat{\varepsilon}_1^*, \hat{\varepsilon}_2^*, \ldots\ldots, \hat{\varepsilon}_n^*$
   เมื่อ $\hat{\varepsilon}_j^*$ คือตัวอย่างที่สุ่มได้ตัวที่ j จาก $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots\ldots, \hat{\varepsilon}_n$
3. สร้างรูปแบบ (model) ใหม่

$$\underset{\sim}{y}^* = X\hat{\underset{\sim}{\beta}}(R) + \underset{\sim}{\varepsilon}^*$$

4. นำค่า $\underset{\sim}{y}^{*}$ และ X คำนวณหาค่า $\hat{\beta}^{*}(R)_i$ โดยวิธีริดจ์รีเกรสชัน

5. สร้างตัวสถิติ

$$T_i^{*} = \frac{\left(\hat{\beta}^{*}(R)_i - \hat{\beta}(R)_i\right)}{S_{ii}^{*}}$$

โดย

$\hat{\beta}^{*}(R)_i$ เป็นตัวประมาณด้วยริดจ์รีเกรสชันของ β_i^{*}

$\hat{\underset{\sim}{\beta}}^{*}(R) = (X'X + kI)^{-1}X'\underset{\sim}{y}^{*}$

$S_{ii}^{*} = \sqrt{\widehat{\mathrm{var}}\left(\hat{\beta}^{*}(R)_i\right)}$ เป็นความคลาดเคลื่อนมาตรฐานของ $\hat{\beta}_i^{*}$

$\widehat{\mathrm{var}}(\hat{\beta}^{*}(R)_i) = S^{*2}R_{ii}$

$$S^{*2} = \hat{\sigma}^{*2} = \frac{1}{n-p-1}(\underset{\sim}{y}^{*} - X\hat{\underset{\sim}{\beta}}^{*})'(\underset{\sim}{y}^{*} - X\hat{\underset{\sim}{\beta}}^{*})$$

$$= \frac{1}{n-p-1}(\underset{\sim}{y}^{*'}\underset{\sim}{y}^{*} - 2\hat{\underset{\sim}{\beta}}^{*'}(X'X) + \hat{\underset{\sim}{\beta}}^{*'}(X'X)\hat{\underset{\sim}{\beta}}^{*})$$

p = จำนวนตัวแปรอิสระ

6. กระทำตามขั้นตอนที่ 2 - 5 ซ้ำกันเท่ากับจำนวนครั้งของการทำบูตสแตรปในแต่ละขนาดตัวอย่าง แล้วนำมาหาค่า $\hat{p}^{*}(z)$ ดังสมการข้างล่าง

$$\hat{p}^{*}(z) = \frac{\#\left\{t^{*} \le z\right\}}{B}$$

โดย $\hat{p}^{*}(z)$ เป็นตัวประมาณของ $p^{*}(z) = P\left(T_i^{*} \le z\right)$

B จำนวนครั้งของการทำบูตสแตรป

z แทน $z_{\alpha/2}$ และ $z_{1-\alpha/2}$ เป็นเปอร์เซ็นไทล์ของการแจกแจงแบบปกติมาตรฐาน

2.3 เกณฑ์ที่ใช้ในการเปรียบเทียบวิธีการประมาณค่าแบบช่วง

การเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุแบบช่วง ในกรณีที่ตัวแปรอิสระเกิดพหุสัมพันธ์กันทั้ง 3 วิธีนี้ จะทำการเปรียบเทียบค่าระดับความเชื่อมั่นของช่วงความเชื่อมั่น และค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่คำนวณได้จากแต่ละวิธีการประมาณ ที่ระดับความเชื่อมั่น 3 ระดับ คือ 90% , 95% และ 99% ในแต่ละสถานการณ์ของการทดลอง ทำการทดลองซ้ำ 500 ครั้ง การเปรียบเทียบค่าระดับความเชื่อมั่นของช่วงความเชื่อมั่น จะทำการเปรียบเทียบระหว่างค่าระดับความเชื่อมั่นของช่วงความเชื่อมั่นที่คำนวณได้จากแต่ละวิธีการประมาณกับค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด ในการตรวจสอบว่าวิธีการประมาณใดให้ค่าระดับความเชื่อมั่นจากการทดลองไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดได้หรือไม่นั้น ผู้วิจัยจะอาศัยการทดสอบสมมติฐานโดยใช้ตัวสถิติ Z ดังนี้

$$H_0 : p \ge p^*$$
$$H_1 : p < p^*$$

$$-Z_\alpha < \frac{\hat{P} - p^*}{\left(p^*(1-p^*)/n\right)^{1/2}} < 1$$

$$p^* - Z_\alpha\left(p^*(1-p^*)/n\right)^{1/2} < \hat{P} < 1$$

เพราะฉะนั้น ช่วงในยอมรับสมมติฐานหลักคือ

$$\left(p^* - Z_\alpha\left(p^*(1-p^*)/n\right)^{1/2} \;,\; 1\right)$$

โดยที่ α คือระดับนัยสำคัญหรือ Type I error ที่กำหนดในการทดสอบ
ในการวิจัยครั้งนี้กำหนดระดับนัยสำคัญของการทดสอบเท่ากับ 0.05

1. ค่าสัมประสิทธิ์ความเชื่อมั่น 90%

$$H_0 : p \ge 0.90$$
$$H_1 : p < 0.90$$

จะได้

$$\left(0.90 - 1.645(0.90(0.1)/500)^{1/2} \;,\; 1\right)$$
$$\left(0.8779 \;,\; 1\right)$$

2. ค่าสัมประสิทธิ์ความเชื่อมั่น 95%

$$H_0 : p \geq 0.95$$
$$H_1 : p < 0.95$$

จะได้

$$\left(0.95 - 1.645(0.95(0.05)/500)^{1/2} , 1\right)$$
$$\left(0.934 , 1\right)$$

3. ค่าสัมประสิทธิ์ความเชื่อมั่น 99%

$$H_0 : p \geq 0.99$$
$$H_1 : p < 0.99$$

จะได้

$$\left(0.99 - 1.645(0.99(0.01)/500)^{1/2} , 1\right)$$
$$\left(0.9827 , 1\right)$$

ในการวิจัยครั้งนี้มีพารามิเตอร์ที่ต้องประมาณ 3 ตัว เพราะฉะนั้นค่าระดับความเชื่อมั่นของช่วงความเชื่อมั่นของแต่ละวิธีการประมาณในแต่ละสถานการณ์จะมี 3 ค่า โดยจะนำค่าระดับความเชื่อมั่นของช่วงความเชื่อมั่นทั้ง 3 ค่ามาหาค่าเฉลี่ย แล้วจึงนำค่าระดับความเชื่อมั่นของช่วงความเชื่อมั่นเฉลี่ยมาเปรียบเทียบกับค่า 0.8779 , 0.934 และ 0.9827 ที่สัมประสิทธิ์ความเชื่อมั่น 90% , 95% และ 99% ตามลำดับ ถ้าวิธีการประมาณใดให้ค่าระดับความเชื่อมั่นของช่วงความเชื่อมั่นเฉลี่ยจากการทดลองไม่ต่ำกว่าค่าดังกล่าวจะถือว่าวิธีประมาณนั้นให้ค่าระดับความเชื่อมั่นจากการทดลองไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดในสถานการณ์นั้น ๆ ขั้นต่อไปคือพิจารณาเปรียบเทียบค่าความยาวเฉลี่ยของช่วงความเชื่อมั่น ในการวิจัยครั้งนี้มีพารามิเตอร์ที่ประมาณ 3 ตัว เพราะฉะนั้นค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นของแต่ละวิธีในแต่ละสถานการณ์จะมี 3 ค่า และจะนำค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นทั้ง 3 ค่ามาหาค่าเฉลี่ยแล้วนำค่าเฉลี่ยของค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นมาทำการเปรียบเทียบว่าวิธีการประมาณใดให้ค่าเฉลี่ยของค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นต่ำที่สุด และจะถือว่าวิธีการประมาณนั้นเหมาะสมที่สุด ทั้งนี้ในการเปรียบเทียบค่าเฉลี่ยของค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นจะเปรียบเทียบเฉพาะในกรณีที่วิธีการประมาณนั้นให้ค่าระดับความเชื่อมั่นจากการทดลองไม่ต่ำกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดเท่านั้น

2.4 ผลงานวิจัยที่เกี่ยวข้อง

ในหัวข้อนี้กล่าวถึงผลงานวิจัยที่เกี่ยวข้องกับการประมาณค่าสัมประสิทธิ์การถดถอยพหุ กรณีที่ตัวแปรอิสระเกิดพหุสัมพันธ์ ซึ่งนำเสนอพอเป็นสังเขปดังนี้

2.5.1 ในปี ค.ศ. 1995 Crivelliet AL. ทำการวิจัยเรื่อง "CONFIDENCE INTERVALS IN RIDGE REGRESSION BY BOOTSTRAPPING THE DEPENDENT VARIABLE : A SIMULATION STYDY" ได้ทำการทดลองเชิงวิจัยเพื่อเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยพหุแบบช่วง ในกรณีที่ตัวแปรอิสระเกิดพหุสัมพันธ์ด้วยวิธีการประมาณที่เปรียบเทียบ 4 วิธี คือ

1. การประมาณค่าแบบช่วง โดยใช้ตัวประมาณกำลังสองน้อยที่สุด (OLS)
2. การประมาณค่าแบบช่วงด้วยการแจกแจงที โดยใช้ตัวประมาณริดจ์รีเกรสชัน (HORT)
3. การประมาณค่าแบบช่วงด้วยการแจกแจงปกติ โดยใช้ตัวประมาณริดจ์รีเกรสชัน (HORN)
4. การประมาณค่าแบบช่วงด้วยวิธีบูตสแตรป โดยใช้ตัวประมาณริดจ์รีเกรสชัน (HORB)

<u>ขอบเขตของการวิจัย</u>

- มีลักษณะการแจกแจงของความคลาดเคลื่อนมีการแจกแจงแบบปกติ ด้วยค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 0.01 , 0.1 และ 1.0
- ขนาดตัวอย่างเท่ากับ 15 และ 25
- ระดับความสัมพันธ์ของตัวแปรอิสระ 3 ระดับ คือ มีระดับความสัมพันธ์ต่ำ , ระดับความสัมพันธ์สูง , ระดับความสัมพันธ์สูงมาก
- สัมประสิทธิ์ความเชื่อมั่นที่กำหนดเท่ากับ 95%

<u>สรุปผลการวิจัย</u>

- โดยส่วนใหญ่แล้ว วิธี HORB นั้นจะให้ค่าระดับความเชื่อมั่นของช่วงความเชื่อมั่นจากการทดลองสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด (0.95) ส่วนวิธีอื่นๆ คือ OLS , HORT , HORN ให้ค่าระดับความเชื่อมั่นจากการทดลองต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด 0.95

- วิธี HORB นั้นให้ค่าระดับความเชื่อมั่นจากการทดลองสูงกว่าค่าระดับความเชื่อมั่นจากการทดลองที่ได้จาก วิธี OLS , HORT , HORN แต่วิธี HORB นั้นจะให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นสูงกว่า วิธี HORT , HORN ในทุกสถานการณ์ และจะมีค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นต่ำกว่าวิธี OLS ในกรณีที่ค่าความแปรปรวนเท่ากับ 0.1 และ 1.0 ในทุกระดับความสัมพันธ์ของตัวแปรอิสระ และในกรณีที่ค่าความแปรปรวนเท่ากับ 0.01 วิธี HORB จะให้ค่า

ความยาวเฉลี่ยของช่วงความเชื่อมั่นต่ำกว่าวิธี OLS เมื่อระดับความสัมพันธ์ของตัวแปรอิสระมีความสัมพันธ์สูงมาก

2.5.2 นางสาวชันยากร ต้นชลขันธ์ : การเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยพหุโดยวิธีกำลังสองน้อยที่สุด วิธีริดจ์รีเกรสชัน และวิธีที่ใช้หลักการของริดจ์และสไตน์ ในกรณีที่เกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ

เปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยพหุโดยวิธีกำลังสองน้อยที่สุด (OLS) วิธีริดจ์รีเกรสชัน (RR) และวิธีที่ใช้หลักการของริดจ์และสไตน์ (RS) เกณฑ์การเปรียบเทียบคือ เปอร์เซ็นต์อัตราส่วนของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง

<u>ขอบเขตของการวิจัย</u>

- มีลักษณะการแจกแจงของความคลาดเคลื่อน

1. การแจกแจงแบบปกติ ด้วยค่าเฉลี่ยเท่ากับ 1 และความแปรปรวนเท่ากับ 0.05 , 0.1 และ 0.15

2. การแจกแจงแบบปกติปลอมปน สเกลเฟคเตอร์=3,10% การปลอมปน= 5 และ 10

3. การแจกแจงลอกนอร์มอล ด้วยค่าเฉลี่ยเท่ากับ 1 และความแปรปรวนเท่ากับ 0.05 , 0.3 และ 0.7

- ขนาดตัวอย่างเท่ากับ 30 , 50 และ 100

- ระดับความสัมพันธ์ของตัวแปรอิสระ 3 ระดับคือ ระดับต่ำ ,ระดับปานกลาง , ระดับสูง

<u>สรุปผลการวิจัย</u>

- กรณีที่ความคลาดเคลื่อนมีการแจกแจงแบบปกติและปกติปลอมปนพบว่าวิธี RR จะให้ผลดีที่สุดโดยส่วนใหญ่สำหรับทุกการแจกแจง ส่วนวิธี RS จะให้ผลดีที่สุดในกรณีที่ความคลาดเคลื่อนมีการแจกแจงแบบปกติ โดยระดับพหุสัมพันธ์ในระดับต่ำ และระดับปานกลาง ซึ่งในกรณีนี้ความแปรปรวนเท่ากับ 0.05 และขนาดตัวอย่างเท่ากับ 100 นอกจากนั้นวิธี RS จะให้ผลดีที่สุดสำหรับการแจกแจงปกติปลอมปน โดยที่ระดับพหุสัมพันธ์ในระดับต่ำ และระดับปานกลาง ซึ่งในกรณีนี้ความแปรปรวนเท่ากับ 0.05 และขนาดตัวอย่างเท่ากับ 100 สเกลเฟคเตอร์เท่ากับ 3 และเปอร์เซ็นต์การปลอมปนเท่ากับ 5 และ 10