บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ในงานวิจัยที่ใช้ตัวแบบการถดถอยเชิงเส้นเชิงเดียว (simple regression model) ซึ่งมีรูปแบบ

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \qquad i = 1,2,...,n$$

โดยที่ Y_i คือ ตัวแปรตาม

β_0, β_1 คือ พารามิเตอร์ของการถดถอย

X_i คือ ตัวแปรอิสระ ที่กำหนดเป็นค่าคงที่

ε_i คือ ความคลาดเคลื่อนสุ่ม ซึ่งเป็นอิสระต่อกัน และมีการแจกแจง $N(0,\sigma^2)$

n คือ ขนาดตัวอย่าง

วัตถุประสงค์หนึ่งที่ศึกษาวิจัยกันมากคือ การประมาณค่าพารามิเตอร์ของการถดถอย $\underline{\beta} = (\beta_0, \beta_1)'$ เพื่อนำไปพยากรณ์ค่าตัวแปรตาม และวิธีหนึ่งที่ใช้ในการประมาณค่าพารามิเตอร์ของการถดถอยคือวิธีความควรจะเป็นสูงสุด ซึ่งมีวิธีการคือหาค่าพารามิเตอร์ของการถดถอย $\underline{\beta}$ เพื่อให้ได้ฟังก์ชันความหนาแน่นของ $\underline{Y} = (Y_1, Y_2, ..., Y_n)'$ มีค่าสูงสุด แต่วิธีนี้มิได้นำเอาข้อมูลเกี่ยวกับค่าพารามิเตอร์ของการถดถอยที่มีอยู่มาวิเคราะห์ร่วมในการประมาณด้วย ซึ่งในหลายกรณีผู้วิจัยจะมีข้อมูลเกี่ยวกับค่าพารามิเตอร์ของการถดถอยอยู่บ้างและต้องการข้อมูลที่ได้จากการทดลองเพิ่มขึ้นมาวิเคราะห์ร่วมกัน ซึ่งกรรมวิธีวิเคราะห์ดังกล่าวนี้ คือ วิธีการแบบเบส์ (Bayesian Method)

วิธีเบส์ มีแนวความคิดโดยการนำเอาส่วนประกอบที่สำคัญ 3 ส่วนมาวิเคราะห์เพื่อการประมาณค่าพารามิเตอร์ที่ต้องการ นั่นคือ ฟังก์ชันความควรจะเป็น (likelihood function) ซึ่งเป็นข้อมูลที่ได้จากการทดลอง การแจกแจงก่อน (prior distribution) เป็นข้อมูลในอดีต หรือข้อมูลเพิ่มเติมเกี่ยวกับค่าพารามิเตอร์ของการถดถอย และการแจกแจงภายหลัง (posterior distribution) เป็นข้อมูลที่เป็นผลมาจากฟังก์ชันความควรจะเป็นและการแจกแจงก่อน โดยอาศัยความสัมพันธ์ที่ว่า การแจกแจงภายหลังแปรผันตาม การแจกแจงก่อนคูณกับฟังก์ชันความควรจะเป็น

ดังนั้นวิธีเบส์ เป็นการนำเอาข้อมูลในอดีตที่มีอยู่แล้วมาใช้ประโยชน์ร่วมกับข้อมูลใหม่เพื่อ

การประมาณค่าพารามิเตอร์ ซึ่งคาดว่าน่าจะให้ตัวประมาณพารามิเตอร์ที่ใกล้เคียงกับค่าพารามิเตอร์จริงมากขึ้น

งานวิจัยนี้จะศึกษาการประมาณค่าพารามิเตอร์ของการถดถอยของตัวแบบการถดถอยเชิงเส้นเชิงเดียว เมื่อฟังก์ชันความควรจะเป็นของ $\underline{Y}$ ซึ่งอยู่ในรูปของ $f(\underline{Y}\,|\,\underline{\beta})$ มีการแจกแจงปกติและล็อกนอร์มัล ด้วยวิธีความควรจะเป็นสูงสุด กับวิธีเบส์กรณีการแจกแจงก่อนของค่าพารามิเตอร์ของการถดถอยเป็นแบบปกติสองตัวแปร ซึ่งอยู่ในรูปของ $h(\underline{\beta})$ โดยมีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของการประมาณค่าพารามิเตอร์ของการถดถอย ในตัวแบบการถดถอยเชิงเส้นเชิงเดียว ภายใต้สถานการณ์ต่าง ๆ

1.2 วัตถุประสงค์ของการวิจัย

ในงานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของวิธีที่นำมาใช้ในการประมาณค่าพารามิเตอร์ของการถดถอยของตัวแบบการถดถอยเชิงเส้นเชิงเดียว เมื่อฟังก์ชันความควรจะเป็นของตัวแปรตาม ซึ่งอยู่ในรูปของ $f(\underline{Y}\,|\,\underline{\beta})$ มีการแจกแจงปกติ และล็อกนอร์มอล และกำหนดให้มีการนำเอาการแจกแจกของพารามิเตอร์การถดถอยที่เป็นปกติสองตัวแปรมาร่วมพิจารณาด้วย ซึ่งอยู่ในรูปของ $h(\underline{\beta})$ ซึ่งวิธีประมาณค่าพารามิเตอร์ของการถดถอยมี 2 วิธีดังนี้

1.2.1 วิธีความควรจะเป็นสูงสุด (Maximum Likelihood Method)

1.2.2 วิธีเบส์ (Bayesian Method)

1.3 สมมติฐานของการวิจัย

สมมติฐานของการวิจัยนี้คือ การประมาณค่าพารามิเตอร์ของการถดถอยของตัวแบบการถดถอยเชิงเส้นเชิงเดียว เมื่อฟังก์ชันความควรจะเป็นของตัวแปรตามเป็นแบบปกติ ด้วยวิธีเบส์กรณีการแจกแจงก่อนของค่าพารามิเตอร์การถดถอยเป็นปกติ 2 ตัวแปร จะให้ตัวประมาณของพารามิเตอร์ของการถดถอยมีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำกว่าตัวประมาณที่ได้จากวิธีความควรจะเป็นสูงสุด และกรณีที่ใช้ฟังก์ชันความควรจะเป็นของตัวแปรตามเป็นแบบล็อกนอร์มัล

1.4 ขอบเขตของการวิจัย

งานวิจัยนี้มีขอบเขตการทดลองดังนี้

(1) ตัวแบบที่ใช้ในการวิจัย

ตัวแบบการถดถอยเชิงเส้นในงานวิจัยนี้กำหนดเป็นแบบเชิงเส้นเชิงเดียว

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \qquad i = 1,2,\ldots,n$$

โดยที่ Y_i คือ ตัวแปรตาม

β_0, β_1 คือ พารามิเตอร์ของการถดถอย

X_i คือ ตัวแปรอิสระ ที่กำหนดเป็นค่าคงที่

ε_i คือ ความคลาดเคลื่อนสุ่ม ซึ่งเป็นอิสระต่อกัน และมีการแจกแจง $N(0, \sigma^2)$

n คือ ขนาดตัวอย่าง

(2) ค่าของตัวแปรอิสระ X_i เป็นค่าคงที่ใด ๆ ในที่นี้ X_i มาจากการแจกแจงปกติ ที่มีค่าเฉลี่ย เท่ากับ 50 และค่าส่วนเบี่ยงเบนมาตรฐาน เท่ากับ 10

(3) กำหนดการแจกแจงของความคลาดเคลื่อนสุ่ม ε_i เป็นแบบปกติ ที่มีค่าเฉลี่ย $E(\varepsilon_i)$ เท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐาน $\sigma = SD(\varepsilon_i)$ โดยสมมติว่าทราบค่า ให้เท่ากับ 1.0 3.0 5.0 7.0 และ 9.0

(4) กำหนดค่าพารามิเตอร์การถดถอย $\underline{\beta} = (\beta_0, \beta_1)'$ ที่ใช้ในตัวแบบข้อ (1) เท่ากับ $\underline{\beta} = (1.0, 1.0)'$

(5) ในงานวิจัยนี้ได้ศึกษาการแจกแจงของ Y สองกรณี คือ

(5.1) $Y_i \sim N(\underline{\beta}' \underline{X}_i, \sigma^2)$, $\underline{X}_i' = (1, X_i)$, $\underline{\beta}' = (\beta_0, \beta_1)$

สำหรับวิธีความควรจะเป็นสูงสุด และวิธีเบส์ จากตัวแบบในข้อ (1) ได้ว่า $Y_1, \ldots, Y_n$ เป็นค่าที่เป็นอิสระต่อกันและมีการแจกแจงปกติเหมือนกัน ที่มีค่าเฉลี่ย $E(Y_i)$ เท่ากับ $\beta_0 + \beta_1 X_i$ และส่วนเบี่ยงเบนมาตรฐาน $SD(Y_i) = SD(\varepsilon_i)$ มีค่า 1.0 3.0 5.0 7.0 และ 9.0

(5.2) $Y_i \sim LN(\underline{\beta}' \underline{X}_i, \sigma^2)$, $\underline{X}_i' = (1, X_i)$, $\underline{\beta}' = (\beta_0, \beta_1)$

สำหรับวิธีความควรจะเป็นสูงสุด และวิธีเบส์ จากตัวแบบในข้อ (1) ได้ว่า $Y_1, \ldots, Y_n$ เป็นค่าที่เป็นอิสระต่อกันและมีการแจกแจงล็อกนอร์มัลเหมือนกัน และ Y_i มีค่าเฉลี่ย $E(Y_i)$ เท่ากับ

$\beta_0 + \beta_1 X_i$ และส่วนเบี่ยงเบนมาตรฐาน $SD(Y_i) \;=\; SD(\varepsilon_i)$ มีค่า 1.0 3.0 5.0 7.0 และ 9.0

(6) กำหนดการแจกแจงก่อนของพารามิเตอร์ของการถดถอยของวิธีเบส์ $\underline{\beta} = (\beta_0, \beta_1)'$ เป็น $N(\underline{\mu}, \underline{V})$ ซึ่งอยู่ในรูปของฟังก์ชันความหนาแน่นก่อน $h(\underline{\beta})$ เขียนได้ดังนี้

$$h(\underline{\beta}) \;=\; \frac{1}{(2\pi)|\underline{V}|^{1/2}} \exp\left[-\frac{1}{2}(\underline{\beta}-\underline{\mu})'\underline{V}^{-1}(\underline{\beta}-\underline{\mu})\right],$$

$$-\infty < \beta_i < \infty, \quad i = 0,1$$

โดยที่ $\underline{\mu}$ คือ เวกเตอร์ค่าเฉลี่ย ของค่าพารามิเตอร์ $\underline{\beta} = (\beta_0, \beta_1)'$ หรือ

$$\underline{\mu} = \quad E(\underline{\beta}) = \begin{bmatrix} \mu_0 \\ \mu_1 \end{bmatrix}, \quad \mu_i = E(\beta_i), \quad i = 0,1$$

ในงานวิจัยนี้กำหนด $\mu_0 = 0.5,\ 1.0,\ 2.0$ $\quad \mu_1 = 1.0\ ,2.0,\ 3.0$

และ $\underline{V}$ คือ เมทริกซ์ความแปรปรวนร่วม (covariance matrix) ของ β_0 และ β_1

$$\underline{V} \;=\; \begin{bmatrix} \sigma_{00} & \sigma_{01} \\ \sigma_{10} & \sigma_{11} \end{bmatrix}, \quad \sigma_{ij} = Cov(\beta_i, \beta_j) \quad ,i,j = 0,1$$

โดยที่ $\sigma_{ii} = \sigma_i^2, \quad i = 0,1$

กำหนดสัมประสิทธิ์ความแปรผันของพารามิเตอร์ β_0 คือ $C.V.(\beta_0)$ และ β_1 คือ $C.V.(\beta_1)$ มี 3 ระดับคือ ระดับต่ำ กลาง สูง ซึ่งกำหนดค่าเป็น 0.6 1.3 และ 1.8 ตามลำดับ และกำหนดค่าสหสัมพันธ์ของพารามิเตอร์ ρ ระหว่างพารามิเตอร์ β_0 และ β_1 แบ่งเป็น 3 ระดับคือ

ระดับต่ำ ค่า ρ มีค่า -0.3 และ -0.1
ระดับปานกลาง ค่า ρ มีค่า 0.5 และ 0.7
ระดับสูง ค่า ρ มีค่า 0.9

งานวิจัยนี้ ศึกษากรณีตามค่าความแปรปรวนของพารามิเตอร์ และระดับสหสัมพันธ์ของพารามิเตอร์ดังนี้

กรณีที่ 1 σ_1^2 มีค่าสอดคล้องกับสัมประสิทธิ์ความแปรผันของพารามิเตอร์ β_0 ณ ระดับต่ำที่ค่า 0.6 และ σ_2^2 มีค่าสอดคล้องกับสัมประสิทธิ์ความแปรผันของพารามิเตอร์ β_1 ณ ระดับต่ำที่ค่า 0.6, ค่า ρ อยู่ในระดับต่ำที่ค่า –0.3

กรณีที่ 2 σ_1^2 มีค่าสอดคล้องกับสัมประสิทธิ์ความแปรผันของพารามิเตอร์ β_0 ณ ระดับต่ำที่ค่า 0.6 และ σ_2^2 มีค่าสอดคล้องกับสัมประสิทธิ์ความแปรผันของพารามิเตอร์ β_1 ณ ระดับปานกลางที่ค่า 1.3, ค่า ρ อยู่ในระดับต่ำที่ค่า –0.1

กรณีที่ 3 σ_1^2 มีค่าสอดคล้องกับสัมประสิทธิ์ความแปรผันของพารามิเตอร์ β_0 ณ ระดับต่ำที่ค่า 0.6 และ σ_2^2 มีค่าสอดคล้องกับสัมประสิทธิ์ความแปรผันของพารามิเตอร์ β_1 ณ ระดับสูงที่ค่า 1.8, ค่า ρ อยู่ในระดับสูงที่ค่า 0.5

กรณีที่ 4 σ_1^2 มีค่าสอดคล้องกับสัมประสิทธิ์ความแปรผันของพารามิเตอร์ β_0 ณ ระดับปานกลางที่ค่า 1.3 และ σ_2^2 มีค่าสอดคล้องกับสัมประสิทธิ์ความแปรผันของพารามิเตอร์ β_1 ณ ระดับสูงที่ค่า 1.8, ค่า ρ อยู่ในระดับสูงที่ค่า 0.7

กรณีที่ 5 σ_1^2 มีค่าสอดคล้องกับสัมประสิทธิ์ความแปรผันของพารามิเตอร์ β_0 ณ ระดับสูงที่ค่า 1.8 และ σ_2^2 มีค่าสอดคล้องกับสัมประสิทธิ์ความแปรผันของพารามิเตอร์ β_1 ณ ระดับสูงที่ค่า 1.8, ค่า ρ อยู่ในระดับปานกลางที่ค่า 0.9

จาก
$$\sigma_i = (C.V.(\beta_i)) \times \mu(\beta_i) \quad i = 0,1$$

และ
$$\rho = \frac{Cov(\beta_i, \beta_j)}{\sqrt{(\sigma_0^2)(\sigma_1^2)}}$$

เพราะฉะนั้น
$$Cov(\beta_i, \beta_j) = \rho\sqrt{(\sigma_0^2)(\sigma_1^2)}$$

สามารถ เขียนเป็นกรณีของเมทริกซ์ความแปรปรวนร่วมของพารามิเตอร์ คือ

(1) $\mu_0 = 0.5,\ \mu_1 = 1.0$ ดังนั้น σ_1^2 มีค่า 0.09 0.4225 และ 0.81 ตามลำดับ และ σ_2^2 มีค่า 0.36 1.69 และ 3.24 ตามลำดับ ดังนั้น

$$\underline{V} = \begin{bmatrix} 0.09 & -0.017 \\ -0.017 & 0.36 \end{bmatrix}, \begin{bmatrix} 0.09 & -0.039 \\ -0.039 & 1.69 \end{bmatrix}, \begin{bmatrix} 0.09 & 0.27 \\ 0.27 & 3.24 \end{bmatrix}, \begin{bmatrix} 0.4225 & 0.819 \\ 0.819 & 3.24 \end{bmatrix}, \begin{bmatrix} 0.81 & 1.458 \\ 1.458 & 3.24 \end{bmatrix}$$

(2) $\mu_0 = 1.0,\ \mu_1 = 2.0$ ดังนั้น σ_1^2 มีค่า 0.36 1.69 และ 3.24 ตามลำดับ และ σ_2^2 มีค่า 1.44 6.76 และ 12.96 ตามลำดับ ดังนั้น

$$\underline{V} = \begin{bmatrix} 0.36 & -0.216 \\ -0.216 & 1.44 \end{bmatrix}, \begin{bmatrix} 0.36 & -0.156 \\ -0.156 & 6.76 \end{bmatrix}, \begin{bmatrix} 0.36 & 1.08 \\ 1.08 & 12.96 \end{bmatrix}, \begin{bmatrix} 1.69 & 3.276 \\ 3.276 & 12.96 \end{bmatrix}, \begin{bmatrix} 3.24 & 5.832 \\ 5.832 & 12.96 \end{bmatrix}$$

(3) $\mu_0 = 2.0,\ \mu_1 = 3.0$ ดังนั้น σ_1^2 มีค่า 1.44 6.76 และ 12.96 ตามลำดับ และ σ_2^2 มีค่า 3.24 15.21 และ 29.16 ตามลำดับ ดังนั้น

$$\underline{V} = \begin{bmatrix} 1.44 & -0.648 \\ -0.648 & 3.24 \end{bmatrix}, \begin{bmatrix} 1.44 & -0.468 \\ -0.468 & 15.21 \end{bmatrix}, \begin{bmatrix} 1.44 & 3.24 \\ 3.24 & 29.16 \end{bmatrix}, \begin{bmatrix} 6.76 & 9.828 \\ 9.828 & 29.16 \end{bmatrix}, \begin{bmatrix} 12.96 & 17.496 \\ 17.496 & 29.16 \end{bmatrix}$$

(7) ขนาดตัวอย่างที่ใช้คือ 10 20 30 50 70 และ 90

1.5 เกณฑ์การตัดสินใจ

เกณฑ์ที่ใช้ในการพิจารณาประสิทธิภาพของตัวประมาณพารามิเตอร์การถดถอยเชิงเส้นเชิงเดียวของวิธีเบส์ และวิธีประมาณความควรจะเป็นสูงสุด คือ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ย (Average Mean Square Error : AMSE) โดยตัวประมาณที่ดีที่สุดจะให้ค่า AMSE ต่ำที่สุด โดยค่า AMSE มีสูตรการคำนวณดังนี้

$$MSE_i = \frac{1}{500}\sum_{j=1}^{500}\left(\theta_{ij} - \beta_i\right)^2$$

$$AMSE = \frac{1}{k}\sum_{i=1}^{k} MSE_i$$

เมื่อ β_i คือ พารามิเตอร์การถดถอยเชิงเส้นเชิงเดียวตัวที่ $i,\ i = 0,1$

θ_{ij} คือ ค่าประมาณพารามิเตอร์การถดถอยเชิงเส้นเชิงเดียวตัวที่ i จากการประมาณ ครั้งที่ j ซึ่ง $\theta_{ij} = \hat{\beta}_{ij}$ สำหรับตัวประมาณความควรจะเป็น และ $\theta_{ij} = \beta_{ij}^{*}$ สำหรับตัวประมาณเบส์, $i = 0,1;\ j = 1,2,\ldots,500$

k คือ จำนวนพารามิเตอร์การถดถอย ซึ่งเท่ากับ 2

MSE_i คือ ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณตัวที่ i สำหรับพารามิเตอร์การถดถอยตัวที่ $i,\ i = 0,1$

1.6 ประโยชน์ที่คาดว่าจะได้รับ

ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัยครั้งนี้ คือ

(1) ทราบประสิทธิภาพของตัวประมาณของแต่ละวิธี ภายใต้สถานการณ์ต่าง ๆ

(2) ผลที่ได้จากการวิจัยจะเป็นแนวทางในการเลือกใช้วิธีในการประมาณค่าพารามิเตอร์การถดถอยเชิงเส้นเชิงเดียว ในแต่ละสถานการณ์

(3) เพื่อใช้เป็นแนวทางการศึกษาตัวแบบเชิงเส้นพหุคูณ และการใช้วิธีการประมาณอื่นต่อไป