บทที่ 5

สรุปผลการวิจัย และข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบประสิทธิภาพของวิธีที่นำมาใช้ในการประมาณค่าพารามิเตอร์ของการถดถอยของตัวแบบการถดถอยเชิงเส้นเชิงเดียว เมื่อฟังก์ชันความควรจะเป็นของตัวแปรตาม ซึ่งอยู่ในรูปของ $f(\underline{Y} \mid \underline{\beta})$ มีการแจกแจงปกติ และล็อกนอร์มัล เมื่อกำหนดการแจกแจงก่อนของค่าพารามิเตอร์การถดถอยเป็นแบบปกติสองตัวแปร ซึ่งอยู่ในรูปของ $h(\underline{\beta})$ ซึ่งวิธีประมาณค่าพารามิเตอร์ของการถดถอยมี 2 วิธีดังนี้

1. วิธีความควรจะเป็นสูงสุด (Maximum Likelihood Method (MLE))
2. วิธีเบส์ (Bayesian Method (BAYES))

โดยวิธีความควรจะเป็นสูงสุด เป็นวิธีที่พิจารณา ฟังก์ชันความควรจะเป็น โดยอาศัยหลักการที่ว่าตัวประมาณพารามิเตอร์การถดถอย ที่ทำให้ฟังก์ชันความควรจะเป็นมีค่ามากที่สุด เป็นตัวประมาณความควรจะเป็นสูงสุด ส่วนวิธีเบส์ เป็นวิธีที่นำเอาการแจกแจงก่อนของพารามิเตอร์การถดถอย ซึ่งในที่นี้ กำหนดเป็นการแจกแจงแบบปกติ 2 ตัวแปร มาพิจารณาร่วมในการประมาณค่าพารามิเตอร์การถดถอยด้วย ทั้งนี้ได้ศึกษาและเปรียบเทียบว่าวิธีการใดเป็นวิธีที่เหมาะสมที่สุด ซึ่งมีการกำหนดสถานการณ์ต่าง ๆ ในงานวิจัยครั้งนี้ ไว้ดังนี้

1. ขนาดตัวอย่างเท่ากับ 10 20 30 50 70 และ 90
2. ตัวแปรอิสระเป็นค่าคงที่ซึ่งจำลองจากประชากรที่มีการแจกแจงปกติด้วยค่าเฉลี่ยเท่ากับ 50 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10
3. ความคลาดเคลื่อนสุ่มมีการแจกแจงปกติซึ่งมีค่าเฉลี่ย $E(\varepsilon_i)$ เท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐาน $SD(\varepsilon_i)$ กำหนดเท่ากับ 1.0 3.0 5.0 7.0 และ 9.0
4. กำหนดค่าพารามิเตอร์การถดถอย $\underline{\beta} = (\beta_0, \beta_1)'$ เท่ากับ $\underline{\beta} = (1.0, 1.0)'$
5. กำหนดให้การแจกแจงก่อนของพารามิเตอร์การถดถอย $\underline{\beta} = (\beta_0, \beta_1)'$ เป็นแบบปกติสองตัวแปร
6. กำหนดค่าสัมประสิทธิ์ความแปรผันของพารามิเตอร์ β_0 คือ $C.V.(\beta_0)$ และ β_1 คือ $C.V.(\beta_1)$ ของการแจกแจงก่อนของพารามิเตอร์ $\underline{\beta}$ มี 3 ระดับคือ ระดับต่ำ กลาง สูง ซึ่งงานวิจัยนี้กำหนดค่าเป็น 0.6 1.3 และ 1.8 ตามลำดับและกำหนดให้ค่าเฉลี่ยของการแจกแจงก่อนของพารามิเตอร์ $\underline{\mu} = \begin{pmatrix}\mu_0\\ \mu_1\end{pmatrix} = \begin{pmatrix}0.5\\ 1.0\end{pmatrix}, \begin{pmatrix}1.0\\ 2.0\end{pmatrix}, \begin{pmatrix}2.0\\ 3.0\end{pmatrix}$ ค่าความแปรปรวนของพารามิเตอร์ β_0

คือ σ_1^2 และความแปรปรวนของพารามิเตอร์ β_1 คือ σ_2^2 มีค่าสอดคล้องกับสัมประสิทธิ์ความแปรผันของพารามิเตอร์ $C.V.(\beta_0)$ และ $C.V.(\beta_1)$ 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 –0.1

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 –0.3

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

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

เกณฑ์ที่ใช้ในการประเมินว่าวิธีการประมาณพารามิเตอร์การถดถอยเชิงเส้นเชิงเดียว เมื่อตัวแปรตามมีการแจกแจงปกติ และล็อกนอร์มัล วิธีใดมีประสิทธิภาพในการประมาณมากที่สุด จะพิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ย (Average Mean Squares Error (AMSE)) ผลการวิจัยมีข้อสรุปดังนี้

5.1 สรุปผลการวิจัย

กรณีที่ตัวแปรตามมีการแจกแจงแบบล็อกนอร์มัลและแปลงเป็นการแจกแจงแบบปกติจะให้ผลสรุปที่เหมือนกับกรณีที่ตัวแปรตามมีการแจกแจงแบบปกติ ซึ่งสรุปผลตามกรณีของสหสัมพันธ์และความแปรปรวนของพารามิเตอร์ β_0 และ β_1 ได้ดังนี้

1. ค่า ρ = -0.3 ถึง -0.1 σ_1^2 = 0.09, 0.36, 1.44 และ σ_2^2 = 0.36, 1.69, 1.44, 3.24, 6.76,

15.21 เมื่อ $\mu_0 = 0.5,1.0$ และ $\mu_1 = 1.0,2.0$ ที่ $SD(\varepsilon_i)$ = 1.0 ทุกค่า n และที่ $SD(\varepsilon_i)$ = 3.0, 5.0 , 7.0 ค่า n = 70 – 90 และที่ $SD(\varepsilon_i)$ = 9.0 ทุกค่า n พบว่า MLE ให้ประสิทธิภาพในการประมาณดีที่สุด ส่วนที่ $SD(\varepsilon_i)$ = 3.0, 5.0, 7.0 ค่า n = 10 - 50 พบว่า BAYES ให้ประสิทธิภาพในการประมาณดีที่สุด แต่เมื่อ $\mu_0 = 2.0$, $\mu_1 = 3.0$ ที่ $SD(\varepsilon_i)$ = 1.0 ทุกค่า n และที่ $SD(\varepsilon_i)$ = 3.0, 5.0 , 7.0 ค่า n = 50 – 90 และที่ $SD(\varepsilon_i)$ = 9.0 ทุกค่า n พบว่า MLE ให้ประสิทธิภาพในการประมาณดีที่สุด ส่วนที่ $SD(\varepsilon_i)$= 3.0, 5.0, 7.0 ค่า n=10 - 30 พบว่า BAYES ให้ประสิทธิภาพในการประมาณดีที่สุด

2. ค่า ρ = 0.5 ถึง 0.7 σ_1^2 = 0.09, 0.4225, 0.36, 1.44, 1.69, 6.76 และ σ_2^2 = 3.24, 12.96, 29.16 เมื่อ $\mu_0 = 0.5,1.0,2.0$ และ $\mu_1 = 1.0,2.0,3.0$ ที่ $SD(\varepsilon_i)$ = 1.0, 3.0 ทุกค่า n และที่ $SD(\varepsilon_i)$ = 5.0, 7.0 ค่า n = 30 - 90 และที่ $SD(\varepsilon_i)$ = 9.0 ค่า n = 70 - 90 พบว่า MLE ให้ประสิทธิภาพในการประมาณดีที่สุด ส่วนที่ $SD(\varepsilon_i)$ = 5.0, 7.0 เมื่อ n = 10 - 20 และที่ $SD(\varepsilon_i)$ = 9.0 ค่า n = 10 - 50 พบว่า BAYES ให้ประสิทธิภาพในการประมาณดีที่สุด

3. ค่า ρ = 0.9 σ_1^2 = 0.81, 3.24, 12.96 และ σ_2^2 = 3.24, 12.96, 29.16 ที่ $SD(\varepsilon_i)$ = 1.0, 3.0, 5.0, 7.0 ทุกค่า n และที่ $SD(\varepsilon_i)$ = 9.0 ค่า n = 30 - 90 พบว่า MLE ให้ประสิทธิภาพในการประมาณดีที่สุด ส่วนที่ $SD(\varepsilon_i)$ = 9.0 เมื่อ n = 10 – 20 พบว่า BAYES ให้ประสิทธิภาพในการประมาณดีที่สุด

4. ค่า AMSE มีแนวโน้มเพิ่มขึ้นเมื่อ $SD(\varepsilon_i)$, σ_1^2 , σ_2^2 หรือ ρ เพิ่มขึ้น แต่มีแนวโน้มลดลงเมื่อ n เพิ่มขึ้น

5.2 ข้อเสนอแนะ

ผลการวิจัยครั้งนี้จะเสนอแนะเป็น 2 ด้าน คือ

5.2.1 ด้านการนำไปใช้ประโยชน์

เพื่อเป็นแนวทางในการเลือกวิธีการประมาณพารามิเตอร์การถดถอยเชิงเส้นเชิงเดียว เมื่อตัวแปรตามมีการแจกแจงแบบปกติ และล็อกนอร์มัล ให้เหมาะสมในแต่ละสถานการณ์

งานวิจัยนี้ศึกษา 2 กรณี คือ กรณีที่การแจกแจงของตัวแปรตามมีการแจกแจงปกติ และกรณีที่มีการแจกแจงล็อกนอร์มัล ทั้งนี้ โดยอาศัยหลักการแปลงตัวแปรสุ่มแบบล็อกนอร์มัลให้เป็นแบบปกติก่อนการประมาณค่าพารามิเตอร์ จะได้ผลสรุปที่เหมือนกัน ดังนั้น ในทางปฏิบัติของการวิเคราะห์การถดถอยเชิงเส้นเชิงเดียว ผู้วิเคราะห์มีข้อมูลตัวแปรอิสระและตัวแปรตามแล้วนำมา

วิเคราะห์การแจกแจงของตัวแปรตามว่ามีการแจกแจงแบบปกติหรือไม่ อาจใช้วิธีการทดสอบ χ^2 – goodness of fit หรือ Kolmogorov – Sminov ถ้าผลการทดสอบเป็นการแจกแจงแบบปกติก็ใช้ข้อมูลตัวแปรอิสระ และตัวแปรตามนั้นเลย ถ้าผลการทดสอบไม่เป็นการแจกแจงแบบปกติ สามารถใช้วิธีการแปลงข้อมูลให้เป็นการแจกแจงแบบปกติ และแทนตัวแปรตามด้วยสูตรการแปลงที่ได้ ทั้งนี้อาจใช้วิธีการเพิ่มจำนวนตัวอย่างก็ได้ ต่อมานำข้อมูลของตัวแปรอิสระและตัวแปรตามที่ได้จากสูตรการแปลงข้อมูลแล้วไปประมาณค่าพารามิเตอร์

กรณีที่มีข้อมูลแรกเริ่มของพารามิเตอร์ ถ้าความแปรปรวนของพารามิเตอร์มีค่าสอดคล้องกับสัมประสิทธิ์ความแปรผันของการแจกแจงก่อนระดับสูงคือค่าตั้งแต่ 1.4 ค่าสหสัมพันธ์ระหว่างพารามิเตอร์มีค่าระดับสูงในช่วง 5.8 – 1.0 หรือ (-1.0) - (-5.8) หรือใช้ขนาดตัวอย่างตั้งแต่ 46 ขึ้นไป ควรประมาณพารามิเตอร์ด้วยวิธีความควรจะเป็นสูงสุด แต่ถ้าข้อมูลก่อนของพารามิเตอร์มีค่าสัมประสิทธิ์ความแปรผันระดับมีค่าน้อยกว่า 1.4 ระดับสหสัมพันธ์ระหว่างพารามิเตอร์มีค่าระดับต่ำในช่วง 0.0 – 5.7 หรือ 0.0 – (-5.7) หรือใช้ขนาดตัวอย่างน้อยกว่า 46 ควรประมาณพารามิเตอร์ด้วยวิธีเบส์

กรณีที่ไม่มีข้อมูลแรกเริ่มของพารามิเตอร์ ควรประมาณพารามิเตอร์ด้วยวิธีความควรจะเป็นสูงสุด แต่ควรใช้ขนาดตัวอย่างตั้งแต่ 46 จะให้ประสิทธิภาพในการประมาณดีที่สุด

5.2.2 ด้านการศึกษาวิจัย

เพื่อเป็นแนวทางในการศึกษาเพิ่มเติม เพื่อเป็นการขยายผลการวิจัยออกไปให้เกิดประโยชน์มากยิ่งขึ้น ในกรณีต่าง ๆ ดังนี้

1. ควรทำการศึกษาในกรณีเมื่อจำนวนตัวแปรอิสระมากกว่า 1 ตัว
2. ควรศึกษาในกรณีที่ทำการแปลงตัวแปรตามจากการแจกแจงอื่น ๆ
3. ควรทำการศึกษาเมื่อการแจกแจงก่อนของพารามิเตอร์การถดถอยเป็นการแจกแจงร่วมแบบอื่น ๆ