บทที่ 3

วิธีดำเนินการวิจัย

การประมาณค่าพารามิเตอร์ ในสมการถดถอยเชิงเส้นอย่างง่าย เมื่อค่าสังเกตของตัวแปรตามมีค่าที่ถูกตัดทิ้งทางขวา โดยประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองต่ำสุด วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด และวิธีการของบัคเลย์และเจมส์นั้น วิธีแรก เป็นการประมาณค่าพารามิเตอร์เพียงรอบเดียว ก็จะได้ค่าประมาณพารามิเตอร์ วิธีที่สอง เป็นการประมาณค่าพารามิเตอร์ และนำค่าประมาณพารามิเตอร์มาประมาณค่าสังเกตที่ถูกตัดทิ้ง โดยใช้อัตราการสูญเสีย (Failure Rate) และมีการกระทำวนซ้ำ (Iterative) จนกระทั่งได้ค่าประมาณพารามิเตอร์ ส่วนวิธีที่สาม ประมาณค่าพารามิเตอร์และประมาณฟังก์ชันการแจกแจงของค่าความคลาดเคลื่อน F โดยใช้ตัวประมาณพีแอล เพื่อหาค่าถ่วงน้ำหนัก และนำมาประมาณค่าสังเกตที่ถูกตัดทิ้ง วิธีนี้มีการกระทำวนซ้ำจนกระทั่งได้ค่าประมาณพารามิเตอร์ ในแต่ละวิธีการจะประมาณพารามิเตอร์ ภายใต้สถานการณ์เมื่อขนาดตัวอย่างมี 5 ขนาด คือ 10, 15, 30, 50 และ 70 สัดส่วนของข้อมูลที่ไม่ถูกตัดทิ้งมี 4 ระดับ คือ 10%, 20%, 30% และ 40% และการแจกแจงของค่าที่ถูกตัดทิ้ง 4 แบบ คือ แบบสม่ำเสมอ แบบปกติตัดปลายทางซ้าย แบบไวบูลล์ตัดปลายทางซ้าย และแบบเชิงเส้น โดยที่ในแต่ละสถานการณ์มีการทำซ้ำ 1,000 ครั้ง

การวิจัยครั้งนี้ใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Technique) สร้างสถานการณ์ต่าง ๆ ดังนั้น จะขอเริ่มด้วยการกล่าวถึงวิธีการจำลองโดยใช้เทคนิคมอนติคาร์โล แล้วจึงแสดงรายละเอียดของขั้นตอนการวิจัย และโปรแกรมที่ใช้ในการวิจัยตามลำดับ

3.1 <u>วิธีจำลองโดยใช้เทคนิคมอนติคาร์โล</u>

เทคนิคที่ใช้แก้ปัญหาในการคำนวณทางคณิตศาสตร์นั้นมีอยู่หลายวิธี วิธีการจำลองโดยใช้เทคนิคมอนติคาร์โลเป็นวิธีหนึ่งที่นิยมนำมาใช้แก้ปัญหากันอย่างแพร่หลายในปัจจุบัน ซึ่งหลักการ

ของการจำลองโดยใช้เทคนิคดังกล่าว จะใช้เลขสุ่ม (Random Numbers) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา

ขั้นตอนของวิธีการจำลองด้วยเทคนิคมอนติคาร์โลที่ใช้กันในปัจจุบัน แบ่งได้เป็น 3 ขั้นตอนดังนี้คือ

3.1.1 การสร้างตัวเลขสุ่ม การใช้ตัวเลขสุ่มเป็นสิ่งที่สำคัญมากในเทคนิคนี้ ทั้งนี้เพราะว่าหลักการของการจำลองแบบมอนติคาร์โลนั้น จะใช้ตัวเลขสุ่มมาช่วยในการหาคำตอบของปัญหา โดยลักษณะของตัวเลขสุ่มที่นำมาใช้ จะมีการแจกแจงแบบสม่ำเสมอในช่วง (0,1) สำหรับวิธีการสร้างตัวเลขสุ่ม มีผู้เสนอไว้หลายวิธี แต่วิธีที่ดีนั้นลักษณะของเลขสุ่มที่ถูกสร้างขึ้นจะต้องมีการแจกแจงแบบสม่ำเสมอในช่วง (0,1) ตัวเลขสุ่มแต่ละตัวเป็นอิสระต่อกัน และมีช่วงยาวก่อนจะเกิดเลขสุ่มซ้ำ (มีวัฏจักรยาว)

3.1.2 การนำตัวเลขสุ่มมาประยุกต์ใช้กับปัญหาที่ต้องการศึกษา ซึ่งขั้นตอนนี้ขึ้นอยู่กับลักษณะของปัญหา บางปัญหาอาจจะไม่ใช้ตัวเลขสุ่มโดยตรง แต่จะนำไปผลิตเลขสุ่มที่มีการแจกแจงแบบอื่นต่อไป

3.1.3 การทดลองกระทำ เมื่อนำตัวเลขสุ่มมาประยุกต์ให้เข้ากับปัญหาที่ต้องการศึกษาได้แล้วขั้นต่อไปคือ การทดลองโดยใช้กระบวนการของการสุ่ม (Random Process) มากระทำในลักษณะซ้ำ ๆ กันหลาย ๆ ครั้ง เพื่อหาคำตอบที่ต้องการ

3.2 แผนการทดลอง

การวิจัยครั้งนี้ต้องการประมาณค่าพารามิเตอร์ ในสมการถดถอยเชิงเส้นอย่างง่ายเมื่อค่าสังเกตของตัวแปรตามเป็นค่าที่ถูกตัดทิ้งทางขวา การประมาณค่าพารามิเตอร์ในสมการถดถอยเชิงเส้นอย่างง่าย ทั้ง 3 วิธี ใช้วิเคราะห์กับข้อมูลค่าสังเกตของตัวแปรตามมีค่าที่ถูกตัดทิ้ง ทำการประมาณค่าพารามิเตอร์จากข้อมูลทั้งหมดด้วยวิธีการ 3 วิธี ณ สัดส่วนของการถูกตัดทิ้ง 4 ระดับ ขนาดตัวอย่าง 5 ระดับ และการแจกแจงของค่าถูกตัดทิ้ง 4 แบบ รวมทั้งสิ้น 80 สถานการณ์ และค่าพารามิเตอร์ที่ประมาณได้ในแต่ละวิธีการ นำมาประมาณค่าสังเกตของตัวแปรตามด้วยวิธีใดจะดีกว่าจะทำการเปรียบเทียบค่าความคลาดเคลื่อนระหว่างค่าประมาณของตัวแปรตามกับค่าจริง ในรูปของค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ของทั้ง

3 วิธี เพื่อหาวิธีที่ดีที่สุดในแต่ละสถานการณ์ต่อไป รายละเอียดแผนการทดลองมีดังนี้

1. ค่าสังเกตที่ไม่ถูกตัดทิ้ง $T_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, $i = 1,\ldots,n$, n คือ จำนวนค่าสังเกตที่ไม่ถูกตัดทิ้ง โดยที่ $X_i \sim N(20,60)$ และ $\varepsilon_i \sim N(0,16)$ กำหนด $\beta_0 = 2$ และ $\beta_1 = 1$

2. ค่าสังเกตที่ถูกตัดทิ้ง C_i , $i = 1,\ldots,m$, m คือ จำนวนค่าสังเกตที่ถูกตัดทิ้ง และ C_i มีการแจกแจงแบบสม่ำเสมอ แบบปกติตัดปลายทางซ้าย แบบไวบูลล์ตัดปลายทางซ้าย และ $C_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ โดยที่ $X_i \sim N(20,60)$ และ $\varepsilon_i \sim N(0,25)$ กำหนด $\beta_0 = 2$ และ $\beta_1 = 1$

3. ขนาดตัวอย่าง 5 ขนาด คือ 10, 15, 30, 50 และ 70

4. สัดส่วนของข้อมูลที่ถูกตัดทิ้ง 4 ระดับ คือ 10%, 20%, 30% และ 40% ของขนาดตัวอย่าง

3.3 ขั้นตอนในการวิจัย

T_i เป็นตัวแปรสุ่มที่เกิดจากความสัมพันธ์เชิงเส้น $T_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, $i = 1,\ldots,n+m$

C_i , $i = 1,\ldots,n+m$ มีการแจกแจงที่เหมือนกันและเป็นอิสระกัน ดังนั้น C_i เป็นอิสระกับ T_i เมื่อมีการตัดทิ้งแบบสุ่ม ค่าสังเกต Y_i ได้จากนิยาม

$$Y_i = \min(T_i, C_i)$$

$$\delta_i = \begin{cases} 1 & \text{ถ้า } T_i \leq C_i \quad \text{(เป็นค่าสังเกตที่ไม่ถูกตัดทิ้ง)} \\ 0 & \text{ถ้า } T_i > C_i \quad \text{(เป็นค่าสังเกตที่ถูกตัดทิ้ง)} \end{cases}$$

การจำลองข้อมูลเพื่อให้เกิดค่าสังเกตที่ไม่ถูกตัดทิ้ง และเกิดค่าสังเกตที่ถูกตัดทิ้งทางขวา แสดงไว้ในขั้นตอนของการวิจัย ดังต่อไปนี้

<u>**ขั้นตอนของการวิจัย**</u> แบ่งเป็น 5 ขั้นตอนหลัก คือ

1. จำลองตัวแปรอิสระ X_i และค่าความคลาดเคลื่อน ε_i จากการแจกแจงปกติ $N(\mu,\sigma^2)$ และจำลอง T_i จากความสัมพันธ์เชิงเส้น

$$T_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad , \quad i = 1,\ldots,n+m$$

$n+m$ คือ จำนวนค่าสังเกตทั้งหมดทั้งที่ไม่ถูกตัดทิ้งและถูกตัดทิ้ง

2. จำลองค่าสังเกตที่ถูกตัดทิ้ง C_i , $i = 1,\ldots,n+m$

3. หาค่าสังเกตของตัวแปรตาม Y_i , $i = 1,\ldots,n+m$

4. นำค่า Y_i, X_i หาค่าประมาณพารามิเตอร์ จากสมการถดถอยเชิงเส้นอย่างง่าย ด้วยวิธีการทั้ง 3 วิธี จะได้ค่าประมาณพารามิเตอร์ของแต่ละวิธีการ

5. หาค่าความคลาดเคลื่อนจากการประมาณค่าตัวแปรตาม และทำการเปรียบเทียบ

ซึ่งแต่ละขั้นตอนมีรายละเอียดดังนี้

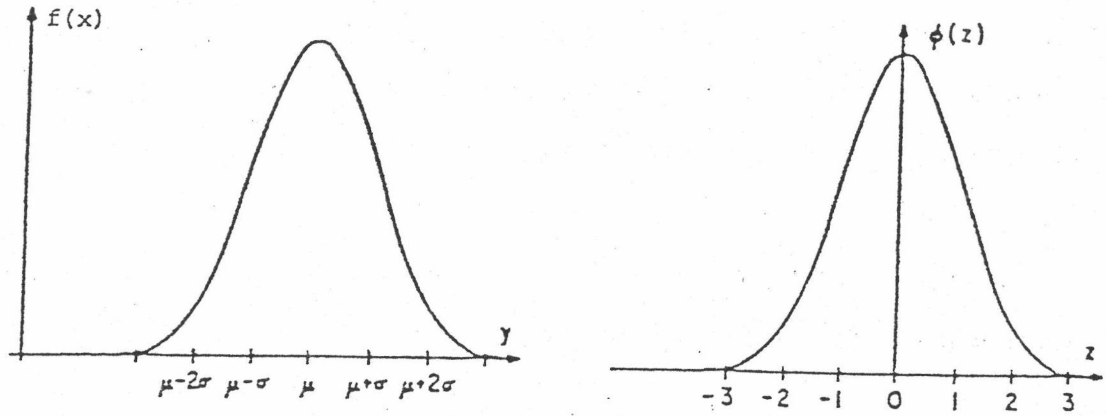
3.3.1 การจำลองตัวแปรอิสระ X_i และค่าความคลาดเคลื่อน ε_i จากการแจกแจงแบบปกติ โดยที่ X_i มีค่าเฉลี่ยเป็น 20 มีค่าความแปรปรวนเป็น 60 และ ε_i มีค่าเฉลี่ย 0 มีค่าความแปรปรวนเป็น 16 การจำลอง X_i และ ε_i ทำได้โดยวิธี Box-Muller รายละเอียดแสดงในภาคผนวก ก

จำลอง $T_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ โดยที่ β_i, $i= 0,1$ เป็นพารามิเตอร์ที่ไม่ทราบค่า (ในการวิจัยครั้งนี้กำหนด $\beta_0 = 2$, $\beta_1 = 1$)

3.3.2 จำลองค่าสังเกตที่ถูกตัดทิ้งทางขวา C_i เมื่อ C_i มีการแจกแจงแบบ

1. แบบสม่ำเสมอช่วง (a,b) (รายละเอียดในภาคผนวก ก)
2. แบบปกติตัดปลายทางซ้าย (C_0,μ,σ^2) (รายละเอียดในภาคผนวก ก)
3. แบบไวบูลล์ตัดปลายทางซ้าย (C_0,α,β) (รายละเอียดในภาคผนวก ก)
4. $C_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ โดยที่ X_i และ ε_i มีแจกแจงแบบปกติ X_i มีค่าเฉลี่ยเป็น 20 มีค่าความแปรปรวนเป็น 60 และ ε_i มีค่าเฉลี่ยเป็น 0 มีค่าความแปรปรวนเป็น 25 จำลอง X_i และ ε_i โดยวิธี Box-Muller รายละเอียดแสดงในภาคผนวก ก กำหนดค่า $\beta_0 = 2$, $\beta_1 = 1$

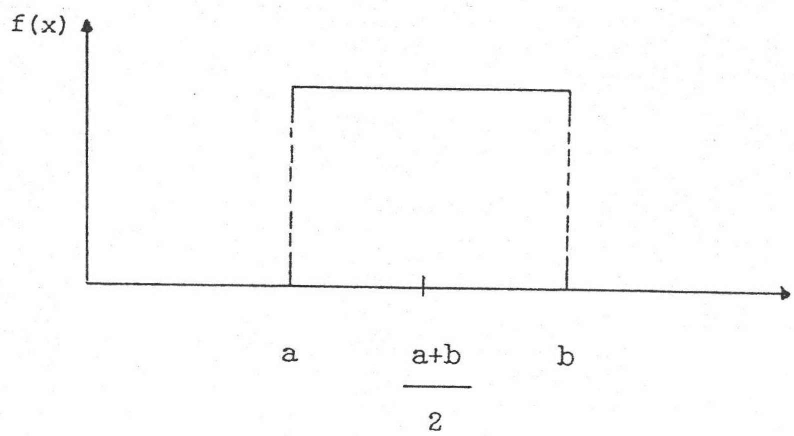
รูปที่ 3.1, 3.2, 3.3, และ 3.4 แสดงการแจกแจงรูปแบบต่าง ๆ ที่ใช้ในการวิจัยครั้งนี้



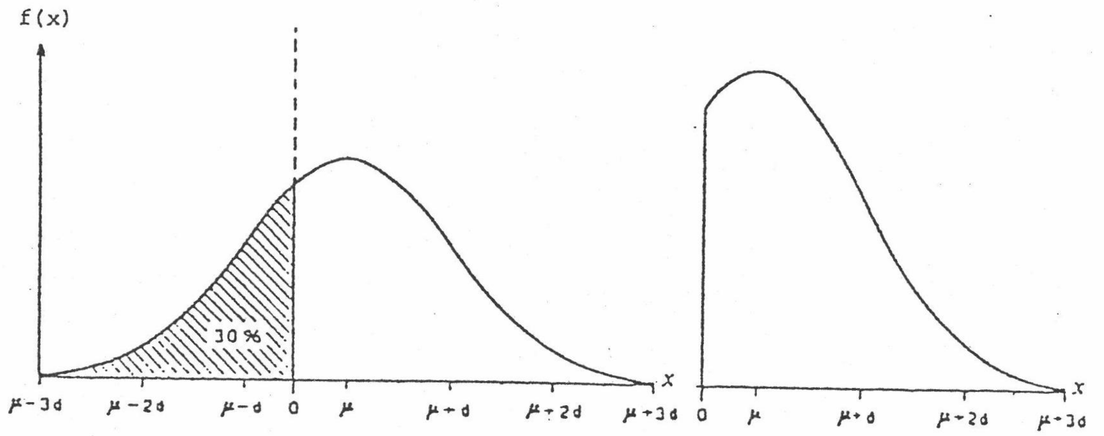
ฟังก์ชันความหนาแน่นของการแจกแจงแบบปกติ

ฟังก์ชันความหนาแน่นของการแจกแจงแบบปกติมาตรฐาน

รูปที่ 3.1 แสดงการแจกแจงปกติ และแบบปกติมาตรฐาน



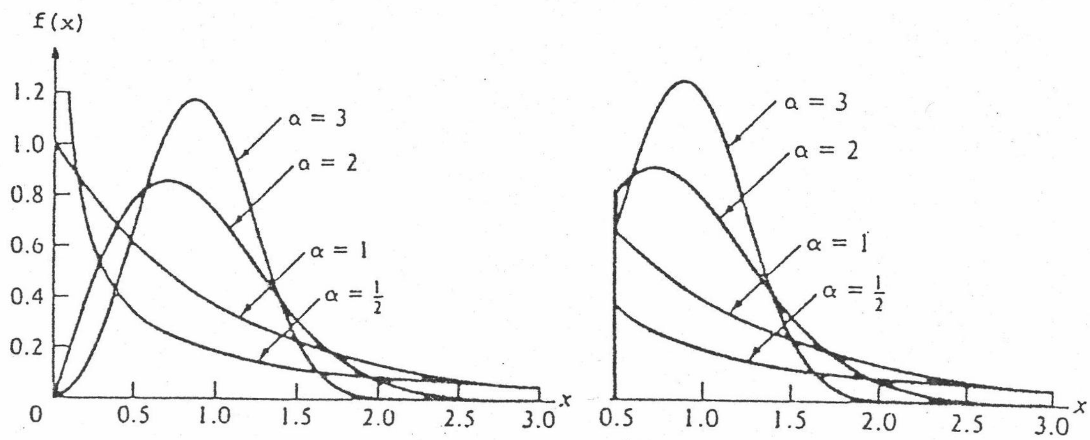
รูปที่ 3.2 แสดงการแจกแจงแบบสม่ำเสมอ



โค้งการแจกแจงและโค้งการแจกแจงสะสมของการแจกแจงแบบปกติ

และแบบปกติตัดปลายทางซ้าย ด้วยสัดส่วนการตัดเป็น 30%

รูปที่ 3.3 แสดงการแจกแจงแบบปกติ และแบบปกติตัดปลายทางซ้าย



โค้งการแจกแจงและโค้งการแจกแจงสะสมของการแจกแจงแบบไวบูลล์

และแบบไวบูลล์ตัดปลายทางซ้าย ด้วยสัดส่วนการตัดเป็น 17%

รูปที่ 3.4 แสดงการแจกแจงแบบไวบูลล์ และแบบไวบูลล์ตัดปลายทางซ้าย

3.3.3 หาค่าสังเกตของตัวแปรตาม Y_i , $i = 1,...,n+m$ $n+m$ คือ ขนาดตัวอย่าง n คือ จำนวนค่าสังเกตที่ไม่ถูกตัดทิ้ง m คือ จำนวนค่าสังเกตที่ถูกตัดทิ้ง

$$Y_i = \begin{cases} T_i & \text{ถ้า } T_i \leqslant C_i \quad \text{(ไม่ถูกตัดทิ้ง)} \\ C_i & \text{ถ้า } T_i > C_i \quad \text{(ถูกตัดทิ้ง)} \end{cases}$$

$$\delta_i = \begin{cases} 1 & \text{ถ้า } T_i \leqslant C_i \\ 0 & \text{ถ้า } T_i > C_i \end{cases}$$

การหาค่าสังเกตของตัวแปรตาม Y_i ทำได้โดยการจำลองค่า T_i และ C_i ทีละคู่ นำค่า T_i และ C_i มาเปรียบเทียบกัน จะได้ค่า $Y_i = T_i$ เมื่อ $T_i \leqslant C_i$ ให้ $\delta_i = 1$ และนับเป็นจำนวนค่าสังเกตที่ไม่ถูกตัดทิ้ง และจะได้ค่า $Y_i = C_i$ เมื่อ $T_i > C_i$ ให้ $\delta_i = 0$ และนับเป็นจำนวนค่าสังเกตที่ถูกตัดทิ้ง จะทำการจำลองเช่นนี้จนกระทั่งได้ค่าสังเกตที่ไม่ถูกตัดทิ้งและค่าสังเกตที่ถูกตัดทิ้งครบตามจำนวน n และ m ที่กำหนด

3.3.4 นำค่า Y_i และคู่อันดับ X_i , $i=1,...,n+m$ คำนวณค่าประมาณพารามิเตอร์ จากสมการถดถอยเชิงเส้นอย่างง่าย ด้วยวิธีกำลังสองต่ำสุด วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด และวิธีการของบัคเลย์และเจมส์ เนื่องจากวิธีกำลังสองต่ำสุด และวิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด สามารถหาค่าประมาณพารามิเตอร์ได้ (Convergence) ทุก ๆ สถานการณ์ที่ศึกษา แต่วิธีการของบัคเลย์และเจมส์ซึ่งใช้วิธีการกระทำวนซ้ำจะมีบางชุดของข้อมูล เช่น ข้อมูลที่มีสัดส่วนของค่าที่ถูกตัดทิ้งจำนวนมาก ค่าประมาณของพารามิเตอร์อาจจะมีค่าแกว่งระหว่าง 2 ค่า หรือไม่สามารถหาค่าประมาณพารามิเตอร์ได้ ดังนั้น เมื่อค่าประมาณพารามิเตอร์มีค่าแกว่งระหว่าง 2 ค่า ให้หาค่าเฉลี่ยของ 2 ค่านั้นเป็นค่าประมาณพารามิเตอร์ หรือตรวจสอบว่าถ้าค่าประมาณพารามิเตอร์ของรอบปัจจุบันกับรอบที่แล้วต่างกัน 0.001 ถือว่าค่าประมาณพารามิเตอร์หาค่าได้ ถ้าไม่เข้าข่ายทั้ง 2 กรณีจะทิ้งชุดข้อมูลนั้น และทำการจำลองข้อมูลที่หาค่าประมาณพารามิเตอร์ได้ทั้ง 3 วิธีการ จำนวน 1,000 ชุดข้อมูล ในแต่ละสถานการณ์ การหาค่าประมาณพารามิเตอร์แต่ละวิธีการมีรายละเอียดดังนี้

3.3.4.1 วิธีกำลังสองต่ำสุด มีขั้นตอนดังนี้

ประมาณค่าพารามิเตอร์ $\hat{\underset{\sim}{\beta}}$ โดยการหาอนุพันธ์ของผลบวกกำลังสองของความคลาดเคลื่อนเทียบกับ $\hat{\underset{\sim}{\beta}}$ แล้วกำหนดสมการให้เท่ากับ $\underset{\sim}{0}$

$$\frac{\partial}{\partial \hat{\underset{\sim}{\beta}}} (\underset{\sim}{Y}'\underset{\sim}{Y} - 2\hat{\underset{\sim}{\beta}}'X'\underset{\sim}{Y} + \hat{\underset{\sim}{\beta}}'X'X\hat{\underset{\sim}{\beta}}) = \underset{\sim}{0}$$

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

3.3.4.2 วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด มีขั้นตอนดังนี้

ฟังก์ชั่นภาวะน่าจะเป็นของค่าสังเกตที่ไม่ถูกตัดทิ้งและถูกตัดทิ้ง เมื่อมีการตัดทิ้งแบบสุ่ม แสดงได้ดังนี้

$$L = \sigma^{-n} \prod_{i=1}^{n} f(z_i) \prod_{n+1}^{n+m} S(z_i)$$

$$= \frac{1}{\sigma^{n}} \frac{1}{(2\pi)^{(n+m)/2}} \exp\left(\frac{-1}{2\sigma^2} \sum_{i=1}^{n+m} (y_i - \mu)^2 \right)$$

1. ประมาณค่าพารามิเตอร์ β_j โดยหาอนุพันธ์บางส่วนของ ล็อกของภาวะน่าจะเป็น เทียบกับ β_j และให้สมการอนุพันธ์บางส่วนเท่ากับ 0

$$\ln L = -n \ln \sigma - \frac{(n+m)\ln(2\pi)}{2} - \frac{1}{2\sigma^2} \sum_{i=1}^{n+m} (y_i^2 - 2y_i \sum_{j=0}^{1} \beta_j x_{ij} + (\sum_{j=0}^{1} \beta_j x_{ij})^2)$$

$$\frac{\partial \ln L}{\partial \beta_j} = \frac{-1}{2\sigma^2} \sum_{i=1}^{n+m} (-2y_i x_{ij} + 2\sum_{j=0}^{1} \beta_j x_{ij}^2)$$

$$= \frac{1}{\sigma^2} \sum_{i=1}^{n+m} (w_i - \mu_i) x_{ij}$$

โดยที่ $w_i = \begin{cases} y_i & , (i = 1,\ldots,n) \\ \mu_i + \sigma\, h(z_i) & , (i = n+1,\ldots,n+m) \end{cases}$

2. ประมาณค่าพารามิเตอร์ σ โดยหาอนุพันธ์บางส่วนของ ล็อกของภาวะน่าจะเป็น เทียบกับ σ และให้สมการอนุพันธ์บางส่วนเท่ากับ 0

$$\ln L = -n \ln \sigma - \frac{(n+m)\ln(2\pi)}{2} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu_i)^2 - \frac{1}{2\sigma^2}\sum_{n+1}^{n+m}(y_i-\mu_i)^2$$

$$\frac{\partial \ln L}{\partial \sigma} = \frac{-n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(y_i-\mu_i)^2 + \frac{1}{\sigma^3}\sum_{n+1}^{n+m}(y_i-\mu_i)^2$$

$$\sigma^2 = \sum_{i=1}^{n}(y_i-\mu_i)^2 \Big/ \left(n - \sum_{n+1}^{n+m} z_i\, h(z_i) \right)$$

3. ประมาณค่าสังเกตที่ถูกตัดทิ้งได้ด้วย $E(Y_i|Y_i>c_i,\beta,\sigma) = w_i$

$$w_i = \mu_i + \sigma\, h(z_i)$$

วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด ใช้วิธีการกระทำวนซ้ำ จนกระทั่งได้ค่าประมาณพารามิเตอร์

3.3.4.3 วิธีการของบัคเลย์และเจมส์ มีขั้นตอนดังนี้

1. เฉพาะข้อมูลที่ไม่ถูกตัดทิ้ง ประมาณค่าพารามิเตอร์เริ่มต้น $\hat{\alpha}$, $\hat{\beta}$ โดยวิธีกำลังสองต่ำสุด

$$\hat{\beta} = \frac{\Sigma_{uc}\; y_i(x_i - \bar{x}^{uc})}{\Sigma_{uc}\;(x_i - \bar{x})^2}$$

$$\hat{\alpha} = \bar{y}^{uc} - \hat{\beta}\bar{x}^{uc}$$

2. ประมาณ $\hat{F}$ โดยใช้ ตัวประมาณพีแอล

$$\hat{F} = 1 - \prod_{i:e(\hat{\beta})\leq e}\left[\frac{n-i}{n-i+1}\right]$$

$$= 1 - \hat{S}(e_i,\hat{\beta})$$

ซึ่งได้จากการหาลำดับที่ของความคลาดเคลื่อนบางส่วน (Partial Residuals) โดยคำนวณที่ $\hat{\alpha} = 0$, $\hat{e}_i(0,\hat{\beta}) = y_i - \hat{\beta}x_i$, $\hat{e}_i \leq e_i$, $\hat{e}_1 < \hat{e}_2 < \hat{e}_n$

3. นำ $\hat{F}$ มาหาค่าถ่วงน้ำหนัก $W(e_i,\hat{\beta})$

4. ประมาณค่าสังเกตที่ถูกตัดทิ้งด้วย $E(Y_i | Y_i > c_i, \beta x_i) = \bar{y}(\beta)$

$$\bar{y}(\hat{\beta}) = \hat{\beta}x_i + \Sigma_{uc} \frac{W_{ik}(e,\hat{\beta})\ (y_k - \hat{\beta}x_i)}{\{ 1 - \hat{F}(c_i - \hat{\beta}x_i) \}}$$

5. ประมาณค่าพารามิเตอร์ ตามวิธีการของบัคเลย์และเจมส์ ได้ดังนี้

$$\hat{\beta}_{BJ} = \frac{\{ \Sigma_{uc}\ y_i(x_i - \bar{x}) + \Sigma_c\ \bar{y}_i(\hat{\beta})\ (x_i - \bar{x}) \}}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\hat{\alpha}_{BJ} = n^{-1}\ \{ \Sigma_{uc}\ y_i + \Sigma_c\ \bar{y}_i(\hat{\beta}) \} - \hat{\beta}_{BJ}\bar{x}$$

วิธีการของบัคเลย์และเจมส์ใช้วิธีกการกระทำวนซ้ำจนกระทั่งได้ค่าประมาณพารามิเตอร์

3.3.5 หาค่าความคลาดเคลื่อนจากการประมาณค่าตัวแปรตาม นำค่าประมาณพารามิเตอร์มาประมาณค่าของตัวแปรตาม $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$ และนำมาเปรียบเทียบกับค่าจริง y_i เพื่อคำนวณหาค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) เนื่องจากในการทดลองได้จำลองข้อมูลซ้ำ ๆ กัน จำนวน 1,000 ครั้ง ในแต่ละสถานการณ์ให้ j แทนรอบที่ทำซ้ำ j=1,...,1000 ดังนั้น การหาค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองหาได้จากสูตรดังนี้

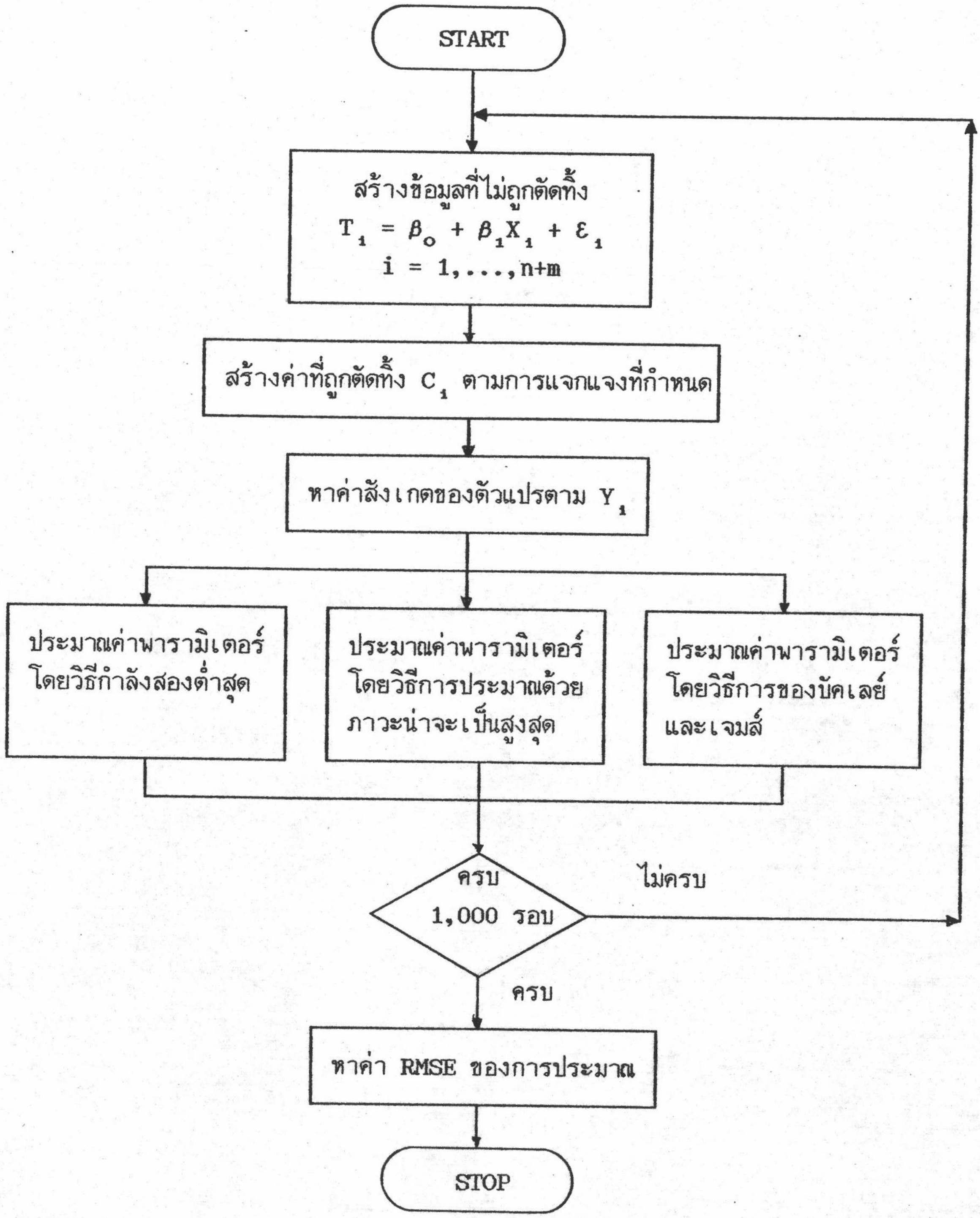
$$MSE_j = \sum_{i=1}^{n} (\ y_i - \hat{y}_i\ )^2 \ /\ n$$

$$RMSE_j = \sqrt{MSE_j}$$

$$RMSE = \sum_{j=1}^{1000} RMSE_j \ /\ 1000$$

จากนั้นจะนำค่า RMSE ของการประมาณค่าตัวแปรตาม ของวิธีการทั้ง 3 วิธี มาเปรียบเทียบเพื่อหาว่าวิธีการใดให้ค่า RMSE ของการประมาณค่าตัวแปรตามต่ำที่สุดจะเป็นวิธีที่ประมาณค่าพารามิเตอร์ของแต่ละสถานการณ์ได้ดีที่สุด

การคำนวณค่า RMSE ของการประมาณ ของทั้ง 3 วิธี ในการทดลองที่ขนาดตัวอย่างหนึ่ง ๆ จะเปลี่ยนค่าสัดส่วนของข้อมูลที่ถูกตัดทิ้งเป็น 4 ระดับ และจะเปลี่ยนการแจกแจงของค่าสังเกตที่ถูกตัดทิ้งเป็น 4 แบบ โดยแต่ละสถานการณ์จะทดลองซ้ำ ๆ จำนวน 1,000 ครั้ง จนครบทุกสถานการณ์ ซึ่งขั้นตอนของการทดลองดังกล่าวนี้ สรุปเป็นผังงานได้ดังรูปที่ 3.5



รูปที่ 3.5 แสดงผังงานสำหรับหาค่าความคลาดเคลื่อนจากการประมาณด้วยวิธีการทั้ง 3 วิธี

3.4 โปรแกรมที่ใช้ในการวิจัย

โปรแกรมที่ใช้ในการวิจัยทั้งหมด เขียนด้วยภาษาฟอร์แทรน (FORTRAN) โดยใช้กับเครื่อง AMDAHL 5860 ซึ่งในแต่ละสถานการณ์ของการทดลอง ลักษณะการทำงานของโปรแกรมจะเหมือนกัน สำหรับรายละเอียดของโปรแกรมจะแสดงไว้ในภาคผนวก ข ซึ่งจะเป็นโปรแกรมการทำงานของแต่ละวิธีการ คือ วิธีกำลังสองต่ำสุด วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด และวิธีการของบัคเลย์และเจมส์