บทที่ 3

วิธีดำเนินการวิจัย

ข้อมูลที่ใช้ในการวิจัยครั้งนี้ได้จากการจำลองขึ้นโดยใช้โปรแกรมภาษาฟอร์แทรน 77 กับเครื่องคอมพิวเตอร์ โดยมีขั้นตอนแผนการทดลองและโปรแกรมที่ใช้ในการศึกษาดังต่อไปนี้

3.1 แผนการทดลอง

ในการวิจัยครั้งนี้กำหนดสถานการณ์ต่าง ๆ ที่ต้องการศึกษาดังนี้

3.1.1 เลือกกลุ่มตัวอย่างจากประชากรโดยกำหนดให้ประชากรมีการแจกแจงดังต่อไปนี้

ก. ประชากรมีการแจกแจงแบบปกติ โดยที่พารามิเตอร์ μ เท่ากับ 0 และ σ^2 เท่ากับ 1

ข. ประชากรมีการแจกแจงแบบปกติปลอมปน โดยที่พารามิเตอร์ μ เท่ากับ 0 และ σ^2 เท่ากับ 1

ค. ประชากรมีการแจกแจงแบบลอกนอร์มอล โดยที่พารามิเตอร์ μ เท่ากับ 0 และ σ^2 เท่ากับ 1 ตามลำดับ

3.1.2 ในทุกการแจกแจงของประชากร จะศึกษาในกรณีที่มีขนาดตัวอย่างเท่ากับ 10 , 30 และ 50

3.1.3 ระดับความสัมพันธ์ของตัวแปรอิสระ

เนื่องจากวิธีการริดจ์รีเกรสชั่นและวิธีการลาเท้นรูทธ์รีเกรสชั่นเป็นวิธีการหนึ่งในการวิเคราะห์สัมประสิทธิ์การถดถอยพหุ กรณีที่ตัวแปรอิสระมีพหุสัมพันธ์กัน ซึ่งในการวิจัยนี้ ผู้วิจัยได้ทำการจำลองข้อมูลให้เกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระในระดับต่าง ๆ 9 ระดับ คือ

[0.11-0.20] , [0.21-0.30] , ... , [0.91-1.00]

3.2 ขั้นตอนในการวิจัย

ขั้นตอนในการวิจัยมีดังนี้ คือ

1. การสร้างโปรแกรมย่อย (subroutine) สำหรับค่าความคลาดเคลื่อนตามที่ต้องการศึกษา

2. การสร้างตัวแปรอิสระ (x) ให้มีความสัมพันธ์ในระดับต่าง ๆ จะใช้หลักการของมัลตินอร์มอล (Multinormal) โดยความสัมพันธ์ของตัวแปรอิสระแต่ละคู่ที่ได้ในแต่ละรอบของการจำลองข้อมูลจะมีค่าใกล้เคียงกัน

3. การสร้างข้อมูลของตัวแปรตาม (y) และตัวแปรอิสระ (x) เมื่อการแจกแจงของค่าความคลาดเคลื่อนเป็นแบบปกติ ตัวแปรตาม (y) จะสร้างจากตัวแปรอิสระ (x) ซึ่งเป็นค่าคงที่ และค่าความคลาดเคลื่อนที่มีการแจกแจงตามที่ต้องการศึกษา โดยที่ตัวแปรตาม (y) มีความสัมพันธ์เชิงเส้นตรงกับตัวแปรอิสระ (x) สำหรับกรณีที่การแจกแจงของค่าความคลาดเคลื่อนเป็นแบบลอกนอร์มอล จะสร้างตัวแปรตาม (y) ให้มีการแจกแจงตามที่ต้องการศึกษา

4. การหาค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยพหุโดยใช้วิธีการริดจ์รีเกรสชั่นและวิธีการลาเท้นรูทธ์รีเกรสชั่น

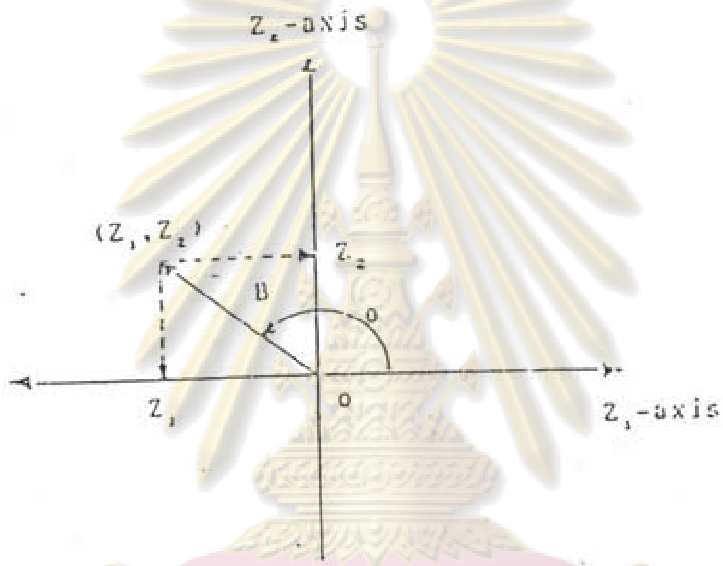
สำหรับรายละเอียดของแต่ละขั้นตอนมีดังนี้

3.2.1 การสร้างการแจกแจงของค่าความคลาดเคลื่อนตามที่ต้องการศึกษา

การสร้างค่าความคลาดเคลื่อนให้มีลักษณะการแจกแจงตามที่ต้องการศึกษานั้น ใช้โปรแกรมภาษาฟอร์แทรน 77 กับเครื่องคอมพิวเตอร์ โดยที่การสร้างลักษณะการแจกแจงต่าง ๆ จะต้องใช้เลขสุ่ม (Random number) ซึ่งมีการแจกแจงแบบสม่ำเสมอในช่วง (0,1) เป็นพื้นฐาน สำหรับรายละเอียดในการสร้างการแจกแจงต่าง ๆ มีดังนี้

3.2.1.1 การผลิตเลขสุ่มที่มีการแจกแจงแบบปกติ

การผลิตเลขสุ่มที่มีการแจกแจงแบบปกติใช้วิธีของ Box และ Muller (1978) โดยผลิตเลขสุ่มที่มีการแจกแจงแบบปกติมาตรฐาน ที่มีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น 1 พร้อม ๆ กัน 2 ค่าดังนี้



จากรูปจะได้ว่า

(1) $$Z_1 = B\cos\theta$$

(2) $$Z_2 = B\sin\theta$$

โดยที่ $B^2 = Z_1^2 + Z_2^2$ มีการแจกแจงแบบไคสแควร์ด้วยระดับความเป็นอิสระเท่ากับ 2 ซึ่งเทียบเท่ากับการแจกแจงแบบเอ็กซ์โปเนนเชียลที่มีค่าเฉลี่ยเท่ากับ 2 ดังนั้นจะได้รัศมี B มีค่าดังนี้

(3) $$B = (-2\ln R)^{1/2}$$

โดยที่ R เป็นเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอ

จากการสมมาตรของการแจกแจงแบบปกติ จะได้ว่า θ มีการแจกแจงสม่ำเสมอระหว่าง 0 กับ 2π เรเดียนและรัศมี ρ กับ θ เป็นอิสระซึ่งกันและกัน จาก (1), (2) และ (3)

เราสามารถสร้างเลขสุ่มที่มีการแจกแจงแบบปกติมาตรฐานจากเลขสุ่ม 2 ชุด คือ R_1 และ R_2 กล่าวคือ

$$Z_1 = (-2\ln R_1)^{1/2} \cos(2\pi R_2)$$
$$Z_2 = (-2\ln R_1)^{1/2} \sin(2\pi R_2)$$

ฟังก์ชันสำหรับการจำลองแบบประชากรที่มีการแจกแจงเป็นแบบปกติที่มีค่าเฉลี่ย μ และมีส่วนเบี่ยงเบนมาตรฐาน σ คือ SUBROUTINE NORMAL ซึ่งได้จาก

$$\text{NORMAL} = \mu + Z_1$$

หรือ

$$\text{NORMAL} = \mu + Z_2$$

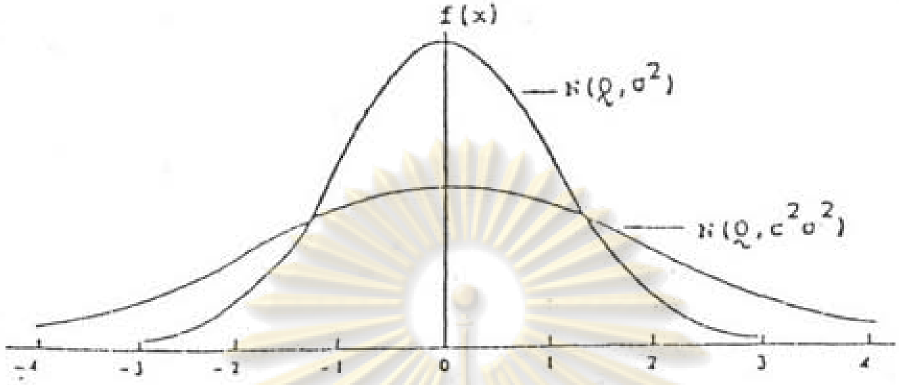
โดยได้แสดง SUBROUTINE NORMAL ไว้ในภาคผนวก ค.

3.2.1.2 การสร้างเลขสุ่มที่มีการแจกแจงแบบปกติปลอมปน

การสร้างเลขสุ่มที่มีการแจกแจงปกติปลอมปนที่มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานตามที่กำหนด จะใช้วิธีที่ Ramsay (ค.ศ. 1977) เสนอไว้ โดยการสร้างการแจกแจงที่มีการแปลงมาจากแบบปกติที่มีฟังก์ชันการแจกแจงอยู่ในรูปของ

$$F(x) = (1-p)N(0,\sigma^2) + pN(0,c^2\sigma^2)$$

เมื่อ c คือสเกลแฟคเตอร์ (scale factor) และ p คือเปอร์เซนต์การปลอมปน (percent of contamination)



รูปแสดงเส้นโค้งการแจกแจงแบบปกติปลอมปน

ตัวแปรสุ่ม x มาจากการแจกแจงแบบ $N(\mu, \sigma^2)$ ด้วยความน่าจะเป็น 1-p และการแจกแจง $N(\mu, c^2 \sigma^2)$ ด้วยความน่าจะเป็น p โดยที่ μ และ σ^2 เป็นค่าเฉลี่ยและความแปรปรวนตามลำดับสำหรับคำสั่งในการสร้างตัวแปรให้มีการแจกแจงแบบปกติปลอมปนได้แสดงไว้ในภาคผนวก

3.2.1.3 การสร้างเลขสุ่มที่มีการแจกแจงแบบลอกนอร์มอล

การแจกแจงแบบลอกนอร์มอลมีฟังก์ชันความหนาแน่นอยู่ในรูปของ

$$f(x) = \begin{cases} \dfrac{1}{x\sigma\sqrt{2\pi}} \exp\left[\dfrac{-(\ln x - \mu)^2}{2\sigma^2}\right] & ; \quad x>0 \ , \ \sigma>0 \\ & \quad -\alpha<\mu<+\alpha \\ 0 & ; \quad \text{อื่น ๆ} \end{cases}$$

เมื่อ μ และ σ^2 เป็นค่าเฉลี่ยและความแปรปรวนของ y โดยที่ $y = \ln x$ แล้ว y จะมีการแจกแจงแบบปกติโดยมี $\exp(\sigma^2)$ เป็น scale parameter และ μ เป็น shape parameter สำหรับค่าเฉลี่ยและความแปรปรวนของการแจกแจงแบบลอกนอร์มอล คือ $\exp(\mu+\sigma^2/2)$ และ $\exp(2\mu+\sigma^2)[\exp(\sigma^2)^{-1}]$ ตามลำดับ สำหรับการสร้างเลขสุ่มที่มีการแจกแจงแบบลอกนอร์มอลได้จาก exponential ของ SUBROUTINE NORMAL เมื่อ DMEAN

เป็นค่าเฉลี่ยและ SIGMA เป็นส่วนเบี่ยงเบนมาตรฐานของการแจกแจงแบบปกติ ดังได้แสดงไว้ใน ภาคผนวก ค.

3.2.2 การสร้างตัวแปรอิสระ x ให้มีความสัมพันธ์ในระดับต่าง ๆ

จากฟังก์ชั่นความหนาแน่นเมื่อเกิดตัวแปรอิสระ k ตัว ซึ่งมีรูปแบบการแจกแจง ดังนี้

$$f(x) = (1/(2\pi)^{m/2}\sqrt{|\Sigma|})*e^{-1/2(\underset{\sim}{x}-\underset{\sim}{\mu})' \Sigma^{-1} (\underset{\sim}{x}-\underset{\sim}{\mu})}$$

$$\text{โดยที่ } \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2m} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_{mm} \end{bmatrix}$$

และ m หมายถึง จำนวนตัวแปรอิสระคูณกับจำนวนตัวอย่าง

เมื่อ $\sigma_{ij} = \rho_{ij}(\sigma_i \sigma_j)$

ρ_{ij} คือ ตัวเลขสุ่มที่อยู่ในช่วงระดับความสัมพันธ์ที่กำหนด

ทั้งนี้ Σ จะเป็นเมตริกซ์สมมาตรและมีสมาชิกทุกตัวเป็นบวก

กำหนดให้ $\Sigma = C'C$ เมื่อ C เป็นเมตริกซ์เชิงสามเหลี่ยมล่าง(Lower triangular matrix)

$$C = \begin{bmatrix} C_{11} & 0 & \cdots & 0 \\ C_{21} & C_{22} & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ C_{m1} & C_{m2} & \cdots & C_{mm} \end{bmatrix}$$

หาค่า C โดยใช้วิธีถอดรากกำลังสอง (Square root method) ซึ่งสามารถคำนวณได้ดังนี้

1) กำหนดให้ $a = \sqrt{\sigma_{11}}$

$C_{i1} = \sigma_{i1}/a$ เมื่อ $i = 1, 2, \ldots, m$

2) $C_{ii} = \left(\sigma_{ii} - \sum_{j=1}^{i-1} C_{ij}^2 \right)^{1/2}$ เมื่อ $i = 2, 3, \ldots, m$

3) $C_{ij} = \left(\sigma_{ij} - \sum_{k=1}^{j-1} C_{ik}C_{jk} \right)/C_{jj}$ เมื่อ $i = 2, 3, \ldots, m$

$j = 1, 2, 3, \ldots, i-1$

นั่นคือ จำลองข้อมูลของตัวแปรอิสระ x_1 จาก

$$x_1 = C_{11}Z_1 + \mu_1 \qquad (1)$$

$$V(x_1) = \sigma_{11} = C_{11}^2 \text{ และ } C_{11} = \sqrt{\sigma_{11}}$$

นั่นคือ จำลองข้อมูลของตัวแปรอิสระ x_2 จาก

$$x_2 = C_{21}Z_1 + C_{22}Z_2 + \mu_2 \qquad (2)$$

$$V(x_2) = \sigma_{22} = V(C_{21}Z_1 + C_{22}Z_2)$$

$$= C_{21}^2 + C_{22}^2 \qquad (3)$$

จาก (1) และ (2) จะได้

$$E[(x_1 - \mu_1)(x_2 - \mu_2)] = \sigma_{12} = E[C_{11}Z_1(C_{21}Z_1 + C_{22}Z_2)] \qquad (4)$$

$$= C_{11}C_{21} + C_{11}C_{21}E(Z_1Z_2) = C_{11}C_{21}$$

นั่นคือ $C_{21} = \sigma_{12}/C_{11}$

แทนค่า C_{21} ใน (3)

$$C_{22} = (\sigma_{22} - C_{21}^2)$$

ในทำนองเดียวกันสามารถหา C_{31} C_{32} และ C_{33} จาก

$$C_{31} = \sigma_{13} / C_{11}$$
$$C_{32} = (\sigma_{32} - C_{21}C_{31})/C_{22}$$
$$C_{33} = [\sigma_{33} - (C_{31}^2 + C_{32}^2)]^{1/2}$$

จากสมการ $\underset{\sim}{x} = C \underset{\sim}{Z} + \underset{\sim}{\mu}$ (4)

ในที่นี้ μ เท่ากับ 0 และ $Z = (Z_1 , Z_2 , \ldots , Z_m)'$ เมื่อ $Z_i \sim N(0,1)$ หรือ $\underset{\sim}{Z} \sim N(0,I)$ จากสมการที่ (4) สามารถจำลองข้อมูลตัวแปรอิสระ x ให้มีความสัมพันธ์ในระดับต่าง ๆ ตามที่ต้องการได้

สำหรับในการวิจัยครั้งนี้ผู้วิจัยได้กำหนดให้เมตริกซ์ C เป็นเมตริกซ์ที่คงที่ หลังจากนั้นจึงจำลองเมตริกซ์ Z ซึ่งมีการแจกแจงแบบปกติ มาคูณกับเมตริกซ์ C ในแต่ละรอบของการทดลอง (การวิจัยครั้งนี้ทำการทดลอง 1,000 รอบในแต่ละสถานการณ์) ผลลัพธ์ที่ได้คือ เมตริกซ์ X จากขั้นตอนดังกล่าว พบว่า เมตริกซ์ X ที่ได้ในแต่ละรอบ จะมีระดับความสัมพันธ์ของแต่ละคู่อยู่ในช่วงระดับความสัมพันธ์ที่กำหนด นอกจากนี้ระดับความสัมพันธ์ของ X แต่ละคู่ที่ได้ในแต่ละรอบของการทดลองจะมีค่าที่ใกล้เคียงกัน

อนึ่ง ผู้วิจัยได้ทำการทดลองเปลี่ยนค่าเมตริกซ์ C และทำการคำนวณข้อมูลตัวแปรอิสระ X ตามขั้นตอนข้างต้น พบว่า การเปลี่ยนเมตริกซ์ C ไม่มีผลต่อผลสรุปของการวิจัยในทุกสถานการณ์

3.2.3 การสร้างข้อมูลให้มีความสัมพันธ์เชิงเส้นตรง

3.2.3.1 ถ้าการแจกแจงของค่าความคลาดเคลื่อนเป็นแบบปกติและปกติปลอมปน จะสร้างข้อมูลของตัวแปรอิสระ x ซึ่งเป็นค่าคงที่ดังได้กล่าวมาแล้ว จากนั้นจึงสร้างตัวแปร y ที่มีความสัมพันธ์เชิงเส้นกับ x ให้มีลักษณะการแจกแจงตามความคลาดเคลื่อนที่กำหนดรูปแบบของความสัมพันธ์เชิงเส้น คือ $\underset{\sim}{y} = x\underset{\sim}{\beta} + \underset{\sim}{\epsilon}$ เมื่อ $\underset{\sim}{\beta}$ เป็นค่าคงที่ที่กำหนดขึ้น (ในที่นี้กำหนดค่า $\underset{\sim}{\beta} = \underset{\sim}{1}$)

สำหรับการสร้าง y นั้นจะเริ่มจากการกำหนดจำนวนตัวแปรอิสระและขนาดตัวอย่างที่ต้องการศึกษา ค่าเฉลี่ยและความแปรปรวน ลักษณะการแจกแจงของค่าความคลาดเคลื่อนพร้อมทั้งค่า $\underset{\sim}{\beta}$ ที่กำหนดขึ้น

3.2.3.2 ถ้าการแจกแจงของค่าความคลาดเคลื่อนเป็นแบบลอกนอร์มอล จะสร้างตัวแปรอิสระ x ซึ่งเป็นค่าคงที่และมีความสัมพันธ์กันดังที่กล่าวมาในตอนต้น แล้วจึงสร้างตัวแปรตาม y ให้มีการแจกแจงเป็นแบบลอกนอร์มอลตามต้องการ

3.3 การประมาณค่าสัมประสิทธิ์การถดถอยพหุโดยวิธีริดจ์รีเกรสชั่น

การประมาณค่าสัมประสิทธิ์การถดถอยพหุโดยวิธีริดจ์รีเกรสชั่น มีขั้นตอนดังต่อไปนี้

3.3.1 คำนวณหาค่าตัวประมาณริดจ์รีเกรสชั่น ตามวิธีการของ Hoerl , Kennard and Baldwin (HKB) และวิธี Lawless and Wang (LW)

3.3.2 นำค่า k ของแต่ละวิธีนี้ไปปรับกับค่าในแนวทางทแยงมุมของเมตริกซ์ $x'x$ แล้ว หาค่า $\hat{\underset{\sim}{\beta}}(k)$ ของแต่ละวิธีออกมา

3.3.3 นำค่า k และ $\hat{\underset{\sim}{\beta}}(k)$ ของแต่ละวิธีมาหาค่า MSE

3.4 การประมาณค่าสัมประสิทธิ์การถดถอยพหุโดยวิธีลาเท้นรูทรีเกรสชั่น

การประมาณค่าสัมประสิทธิ์การถดถอยโดยวิธีลาเท้นรูทรีเกรสชั่น มีขั้นตอนดังต่อไปนี้

3.4.1 คำนวณหาค่าลาเท้นรูท และลาเท้นเวกเตอร์ ของเมตริกซ์ $A'A$ โดยเรียกใช้ SUBROUTINE LATENT

3.4.2 นำค่าลาเท้นรูท และลาเท้นเวกเตอร์ ไปปรับกับสูตรที่แสดงไว้ในตอนต้น จากนั้นจะหาค่า $\hat{\underset{\sim}{\beta}}_{lr}$ และ MSE ของตัวประมาณ $\hat{\underset{\sim}{\beta}}_{lr}$

3.5 การหาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยพหุโดยวิธีริดจ์รีเกรสชั่นและวิธีลาเท้นรูทธ์รีเกรสชั่น

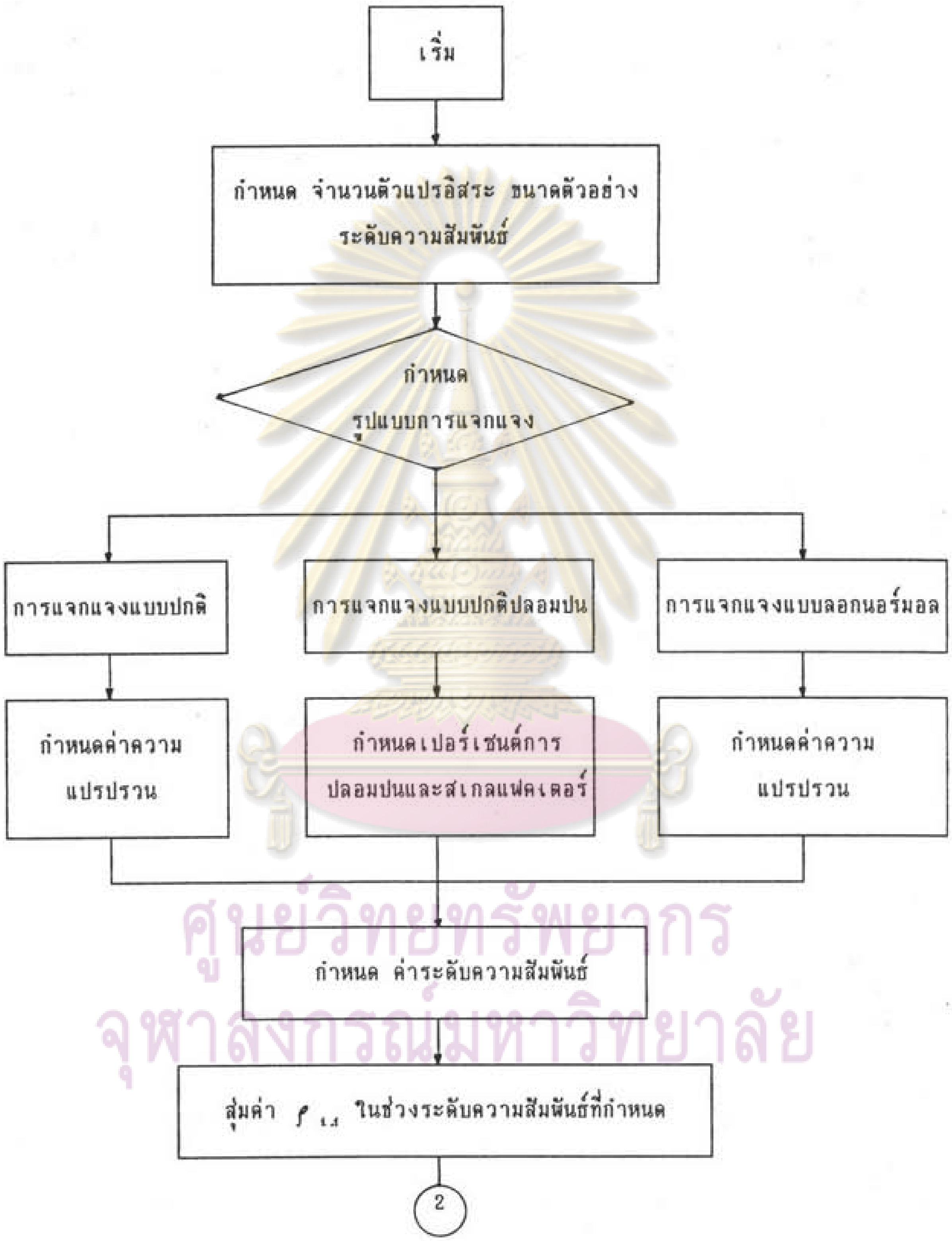
3.5.1 คำนวณค่า MSE($\hat{\underset{\sim}{\beta}}_{\text{รร}}$) ของแต่ละวิธีและ MSE($\hat{\underset{\sim}{\beta}}_{\text{ลร}}$)

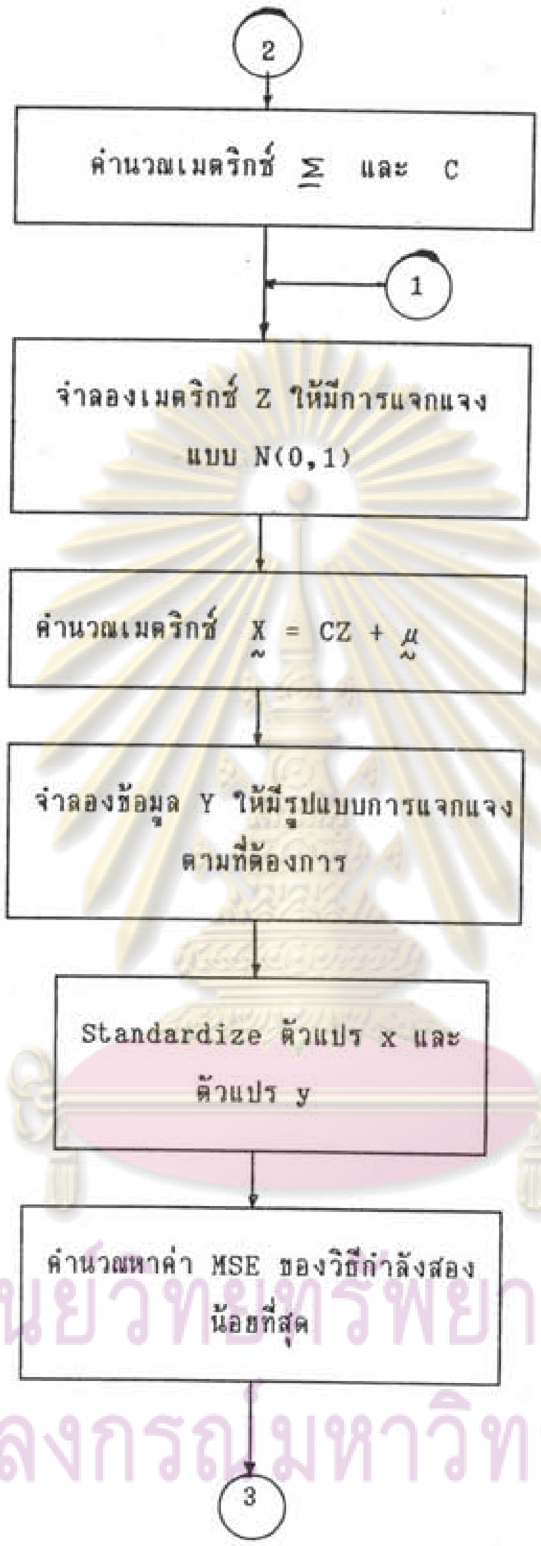
3.5.2 เนื่องจากการวิจัยครั้งนี้กระทำซ้ำ 1,000 รอบในแต่ละสถานการณ์ ดังนั้นจึงคำนวณค่าเฉลี่ยของ MSE ของแต่ละวิธี นั่นคือ

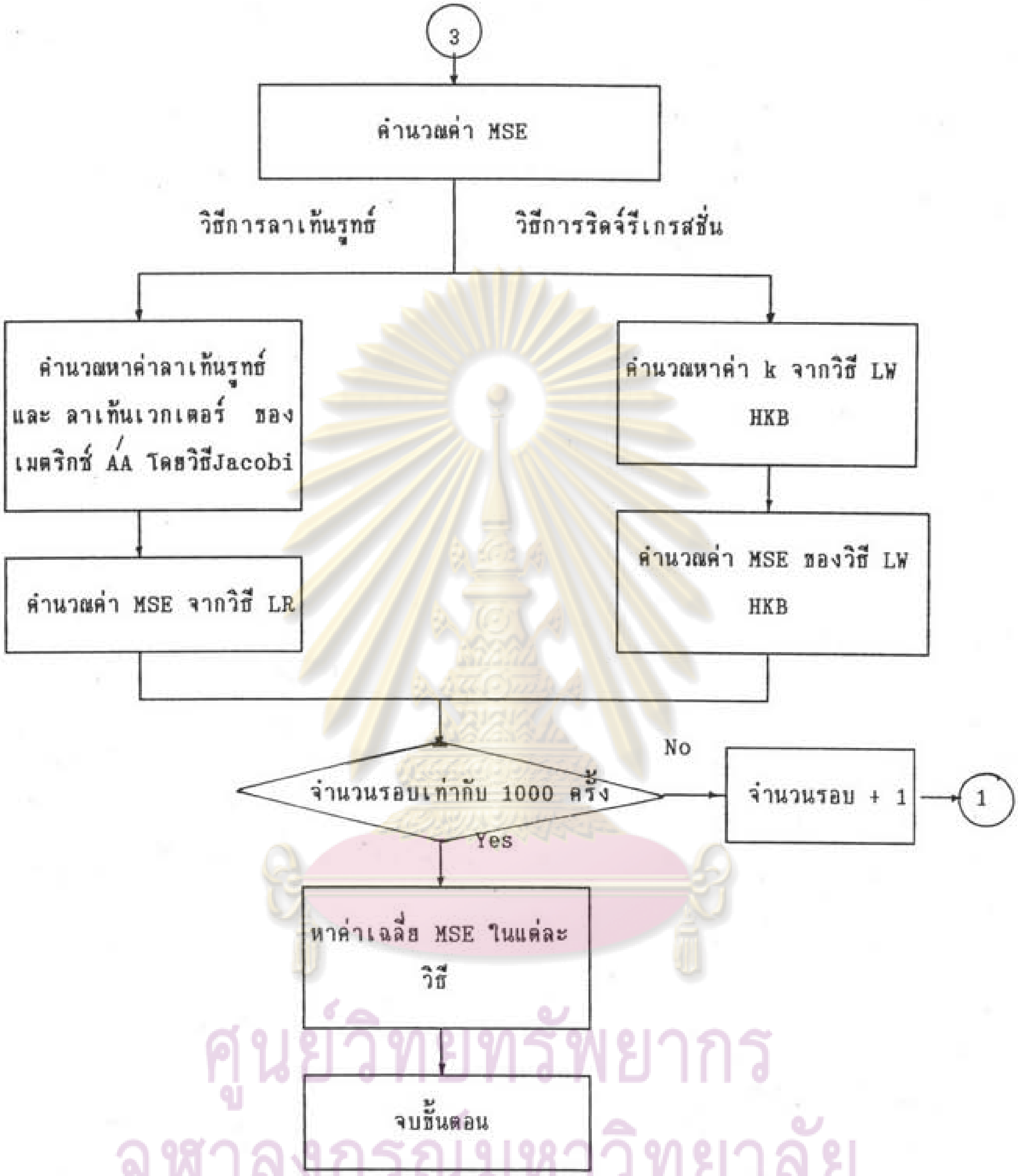
$$AMSE = 1/1000 \sum_{i=1}^{1000} MSE(\hat{\underset{\sim}{\beta}})$$

โดยที่ AMSE คือค่าเฉลี่ยของ MSE($\hat{\underset{\sim}{\beta}}$) ใน 1000 ครั้ง

แผนภาพแสดงขั้นตอนการหา Mean Square Error ของตัวประมาณในแต่ละวิธี




2
คำนวณเมตริกซ์ Σ และ C
1
จำลองเมตริกซ์ Z ให้มีการแจกแจงแบบ N(0,1)
คำนวณเมตริกซ์ X = CZ + μ
จำลองข้อมูล Y ให้มีรูปแบบการแจกแจงตามที่ต้องการ
Standardize ตัวแปร x และตัวแปร y
คำนวณหาค่า MSE ของวิธีกำลังสองน้อยที่สุด
3


3
คำนวณค่า MSE
วิธีการลาเท้นรูทธ์
วิธีการริดจ์รีเกรสชั่น
คำนวณหาค่าลาเท้นรูทธ์ และ ลาเท้นเวกเตอร์ ของ เมตริกซ์ A'A โดยวิธีJacobi
คำนวณค่า MSE จากวิธี LR
คำนวณหาค่า k จากวิธี LW HKB
คำนวณค่า MSE ของวิธี LW HKB
จำนวนรอบเท่ากับ 1000 ครั้ง
No
Yes
จำนวนรอบ + 1
1
หาค่าเฉลี่ย MSE ในแต่ละวิธี
จบขั้นตอน