บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการศึกษาและเปรียบเทียบวิธีการประมาณค่าองค์ประกอบความแปรปรวนสำหรับตัวแบบข้ามกลุ่ม 2 ปัจจัยเชิงสุ่ม ที่ใช้วิธีการสุ่มสมบูรณ์ในการจัดหน่วยทดลองให้กับปัจจัย เมื่อข้อมูลมีลักษณะสมดุล (Balanced data) ด้วยวิธีความควรจะเป็นสูงสุดและวิธีความควรจะเป็นสูงสุดแบบมอนติคาร์โล โดยทำการศึกษาในกรณีที่ความแปรปรวนมีการแจกแจงแบบปกติ การจำลองข้อมูลในแต่ละสถานการณ์จะใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Technique) โดยใช้โปรแกรม S-PLUS 2000 กับเครื่อง PC สำหรับในบทนี้จะกล่าวถึงวิธีดำเนินการวิจัย ซึ่งประกอบด้วย 3 ส่วนด้วยกันคือ

ส่วนที่ 1 การจำลองด้วยเทคนิคมอนติคาร์โล

ส่วนที่ 2 แผนการดำเนินการวิจัย

ส่วนที่ 3 ขั้นตอนในการดำเนินการวิจัย

3.1 การจำลองด้วยเทคนิคมอนติคาร์โล

เทคนิคมอนติคาร์โล เป็นเทคนิคที่ใช้ตัวเลขสุ่มมาช่วยในการแก้ไขและหาคำตอบของปัญหาที่ยังไม่แน่ใจในผลที่จะเกิดขึ้น โดยในการวิจัยครั้งนี้ได้ทำการสร้างข้อมูลจากการแจกแจงแบบปกติด้วยเทคนิคมอนติคาร์โล โดยใช้ฟังก์ชันที่มีอยู่ในโปรแกรมสำเร็จรูป S-PLUS 2000 คือ ฟังก์ชัน rnorm(n,mean,sd) สร้างโดย Kinderman และ Monahan* ซึ่งทำการสร้างข้อมูลด้วยการผลิตเลขสุ่ม (Random Number) ที่มีการแจกแจงสม่ำเสมอ (Uniform Distribution) ในช่วง (0,1) เป็นพื้นฐานในการสร้างเลขสุ่ม โดยเลขสุ่มที่ได้จะมีคุณสมบัติดังนี้

3.1.1 ตัวเลขที่ได้มีการกระจายความน่าจะเป็นแบบสม่ำเสมอและเป็นอิสระซึ่งกันและกัน

3.1.2 อนุกรมของตัวเลขที่ได้สามารถสร้างซ้ำเดิมได้

3.1.3 อนุกรมตัวเลขไม่ซ้ำเดิมในช่วงที่ต้องการตัวเลขแบบสุ่ม

3.1.4 ใช้เวลาสั้นในการสร้างตัวเลขแบบสุ่ม

3.1.5 ใช้หน่วยความจำของคอมพิวเตอร์น้อย

---

* Kinderman , A.J. and Monahan , J.F. (1997). "Computer generation of random variables using the ratio of uniform deriates." ACM Transaction on Mathematical Software. 3, pp.257-260

3.2 แผนการดำเนินการวิจัย

ในการวิจัยครั้งนี้กำหนดสถานการณ์ต่างๆ สำหรับทำการศึกษาและเปรียบเทียบการประมาณค่าองค์ประกอบความแปรปรวนสำหรับตัวแบบข้ามกลุ่ม 2 ปัจจัยเชิงสุ่ม ด้วยวิธีความควรจะเป็นสูงสุดและวิธีความควรจะเป็นสูงสุดแบบมอนติคาร์โล ดังนี้

3.2.1 ทำการทดลองสำหรับตัวแบบข้ามกลุ่ม 2 ปัจจัยเชิงสุ่ม ที่ใช้วิธีการสุ่มสมบูรณ์ในการจัดหน่วยทดลองให้กับปัจจัย

3.2.2 กำหนดระดับปัจจัยและขนาดของหน่วยทดลอง ดังนี้

3.2.2.1 ถ้า a = b = 2 แล้ว n = 2, 3, 4

3.2.2.2 ถ้า a = b = 3 แล้ว n = 2, 3, 4

3.2.2.3 ถ้า a = b = 4 แล้ว n = 2, 3, 4

3.2.3 กำหนดให้ค่าเฉลี่ยของประชากร (μ) เท่ากันทุกกลุ่มคือมีค่าเท่ากับ 40

3.2.4 การแจกแจงของความคลาดเคลื่อนที่ศึกษาในแผนการทดลองคือ การแจกแจงแบบปกติ

3.2.5 กำหนดให้ข้อมูลมีสัมประสิทธิ์ความแปรผัน (Coefficient of Variation : C.V.) ในระดับต่างๆกันดังนี้คือ 10% , 50% และ 90% และส่วนเบี่ยงเบนมาตรฐาน (σ) เท่ากับ 4, 20 และ 36 ตามลำดับ

3.2.6 กำหนดให้ค่าความแปรปรวนของปัจจัย A เท่ากับค่าความแปรปรวนของปัจจัย B เท่ากับค่าความแปรปรวนของผลกระทบร่วมของปัจจัย A กับปัจจัย B เท่ากับค่าความแปรปรวนของความคลาดเคลื่อนคูณกับค่าคงที่จำนวนเต็ม (h) นั่นคือ $\sigma_\alpha^2 = \sigma_\beta^2 = \sigma_\gamma^2 = h\sigma_\varepsilon^2$ โดยที่ h เป็นค่าจำนวนเต็มคงที่เท่ากับ 1, 2 และ 3

3.2.7 ทำการจำลองข้อมูล Y_{ijk} ที่สุ่มมาจากแต่ละประชากรที่มีการแจกแจงแบบปกติ โดยใช้ฟังก์ชันที่มีอยู่ในโปรแกรมสำเร็จรูป S-PLUS 2000 คือฟังก์ชัน rnorm(n,mean,sd)

เมื่อ n แทนขนาดหน่วยตัวอย่างที่ต้องการ

mean แทนค่าเฉลี่ยของประชากร (μ)

sd แทนส่วนเบี่ยงเบนมาตรฐานของประชากร (σ)

3.2.8 สำหรับการประมาณค่าองค์ประกอบความแปรปรวนด้วยวิธีความควรจะเป็นสูงสุดแบบมอนติคาร์โล กระทำการสุ่มตัวอย่างซ้ำ 400 ครั้ง

3.2.9 กำหนดการกระทำซ้ำในแต่ละสถานการณ์เป็น 500 ครั้ง

3.3 ขั้นตอนในการดำเนินการวิจัย

ขั้นตอนในการวิจัยแบ่งออกเป็น 5 ขั้นตอนดังนี้

3.3.1 สร้างการแจกแจงของความคลาดเคลื่อนตามที่กำหนดในแบบแผนการทดลอง

3.3.2 สร้างข้อมูลตามตัวแบบข้ามกลุ่ม 2 ปัจจัยเชิงสุ่ม ดังนี้

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

3.3.3 การกำหนดค่าสมมติเบื้องต้นสำหรับค่าองค์ประกอบความแปรปรวน

3.3.4 การคำนวณหาค่าประมาณองค์ประกอบความแปรปรวนของวิธีความควรจะเป็นสูงสุดและวิธีความควรจะเป็นสูงสุดแบบมอนติคาร์โล

3.3.5 การคำนวณหาค่าระยะทางมาหาลาโนบิสเฉลี่ยของวิธีความควรจะเป็นสูงสุดและวิธีความควรจะเป็นสูงสุดแบบมอนติคาร์โล เพื่อเป็นเกณฑ์ในการเปรียบเทียบ

ซึ่งรายละเอียดของแต่ละขั้นตอนเป็นดังนี้

3.3.1 สร้างการแจกแจงของความคลาดเคลื่อนตามที่กำหนดในแบบแผนการทดลอง

ในการวิจัยครั้งนี้สร้างความคลาดเคลื่อนให้การแจกแจงแบบปกติ โดยใช้ฟังก์ชันที่มีอยู่ในโปรแกรมสำเร็จรูป S-PLUS 2000 คือฟังก์ชัน rnorm(n,mean,sd)

เมื่อ n แทนขนาดหน่วยตัวอย่างที่ต้องการ

mean แทนค่าเฉลี่ยของประชากร (μ)

sd แทนส่วนเบี่ยงเบนมาตรฐานของประชากร (σ)

3.3.2 สร้างข้อมูลตามตัวแบบข้ามกลุ่ม 2 ปัจจัยเชิงสุ่ม

3.3.2.1 การสร้างตัวแปรสุ่ม α_i ดังนี้

$$\alpha_i \sim N(0,\sigma_\alpha^2)$$

3.3.2.2 การสร้างตัวแปรสุ่ม β_j ดังนี้

$$\beta_j \sim N(0,\sigma_\beta^2)$$

3.3.2.3 การสร้างตัวแปรสุ่ม γ_{ij} ดังนี้

$$\gamma_{ij} \sim N(0,\sigma_\gamma^2)$$

3.3.2.4 การสร้างตัวแปรสุ่ม ε_{ijk} ดังนี้

$$\varepsilon_{ijk} \sim N(0,\sigma_\varepsilon^2)$$

แล้วจึงสร้าง Y_{ijk} ตามตัวแบบต่อไปนี้

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

3.3.3 การกำหนดค่าสมมติเบื้องต้นสำหรับค่าองค์ประกอบความแปรปรวน

การกำหนดค่าสมมติเบื้องต้นสำหรับค่าองค์ประกอบความแปรปรวนของวิธีความควรจะเป็นสูงสุดและวิธีความควรจะเป็นสูงสุดแบบมอนติคาร์โล จะใช้ค่าสมมติเบื้องต้นเป็นค่าเดียวกัน ซึ่งได้มาจากวิธีวิเคราะห์ความแปรปรวน โดยจะทำการสร้างข้อมูล Y_{ijk} ให้เป็นไปตามข้อตกลงเบื้องต้น แล้วนำข้อมูลที่ได้มาหาค่าเฉลี่ยกำลังสอง ซึ่งเป็นสาเหตุของความแปรปรวน เพื่อนำไปคำนวณค่าประมาณ ดังนั้นสามารถหาค่าสมมติเบื้องต้นสำหรับค่าองค์ประกอบความแปรปรวนได้ดังนี้

ค่าสมมติเบื้องต้นของ σ_ε^2 คือ MSE

ค่าสมมติเบื้องต้นของ σ_γ^2 คือ $\dfrac{MSAB - MSE}{n}$

ค่าสมมติเบื้องต้นของ σ_α^2 คือ $\dfrac{MSA - MSAB}{bn}$

ค่าสมมติเบื้องต้นของ σ_β^2 คือ $\dfrac{MSB - MSAB}{an}$

3.3.4 การคำนวณหาค่าประมาณองค์ประกอบความแปรปรวนของวิธีความควรจะเป็นสูงสุดและวิธีความควรจะเป็นสูงสุดแบบมอนติคาร์โล

3.3.4.1 การคำนวณหาค่าประมาณองค์ประกอบความแปรปรวนของวิธีความควรจะเป็นสูงสุด

ฟังก์ชันความควรจะเป็น (Likelihood Function) ขององค์ประกอบความแปรปรวนสำหรับตัวแบบข้ามกลุ่ม 2 ปัจจัยเชิงสุ่ม คือ

$$L = L(\mu, V \mid \underset{\sim}{Y}) = \frac{\exp\left[-\frac{1}{2}\left(\underset{\sim}{Y} - \underset{\sim}{1}\mu\right)' V^{-1}\left(\underset{\sim}{Y} - \underset{\sim}{1}\mu\right)\right]}{(2\pi)^{\frac{N}{2}}|V|^{\frac{1}{2}}}$$

สมการลอการิทึมของฟังก์ชันความควรจะเป็น สามารถเขียนได้ดังต่อไปนี้

$$\begin{aligned} \ell &= \log\left\{ L(\mu, V \mid \underset{\sim}{Y}) \right\} \\ &= -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|V| - \frac{1}{2}\left(\underset{\sim}{Y} - \underset{\sim}{1}\mu\right)' V^{-1}\left(\underset{\sim}{Y} - \underset{\sim}{1}\mu\right) \end{aligned}$$

การหาตัวประมาณค่าพารามิเตอร์ที่ทำให้ ℓ มีค่ามากที่สุด สามารถทำได้โดยทำการหาอนุพันธ์ (Differentiate) เทียบกับพารามิเตอร์ $\mu, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\gamma^2, \sigma_\varepsilon^2$ แล้วให้ผลลัพธ์มีค่าเท่ากับ 0

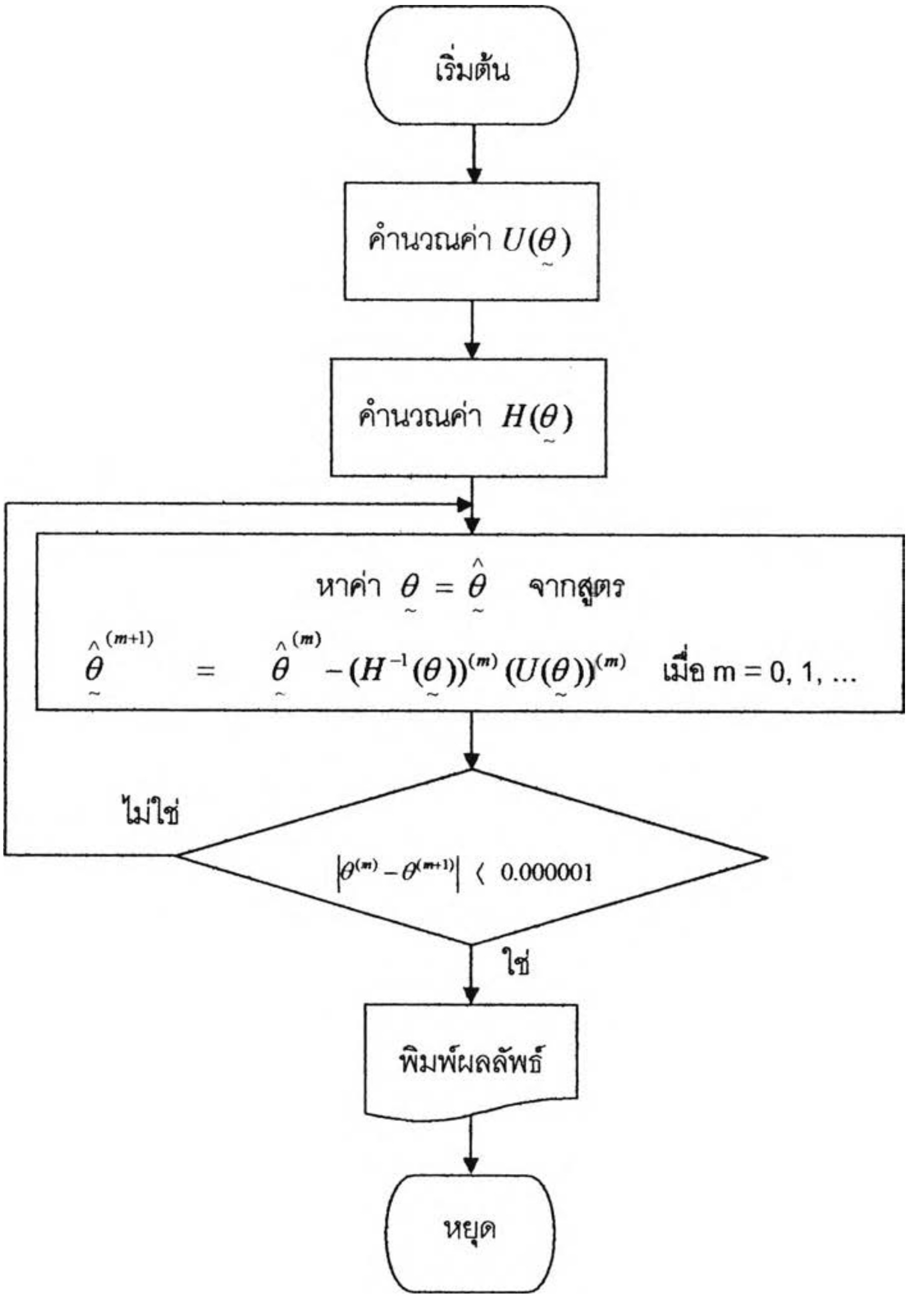
เนื่องจากการแก้สมการ เพื่อหาค่าประมาณของ μ นั้นสามารถแก้สมการได้ แต่ในการหาค่าประมาณของ $\sigma_\alpha^2, \sigma_\beta^2, \sigma_\gamma^2, \sigma_\varepsilon^2$ ไม่สามารถแก้สมการได้โดยตรง ดังนั้นจึงจำเป็นต้องอาศัยวิธีการประมาณเข้ามาช่วย ซึ่งวิธีการประมาณที่นำมาใช้คือ วิธี Newton-Raphson

วิธี Newton-Raphson จะเริ่มจากการหาอนุพันธ์ย่อย (Partial Derivatives) ในสมการลอการิทึมของฟังก์ชันความควรจะเป็น ℓ เทียบกับพารามิเตอร์ $\underset{\sim}{\theta}$ เรียกว่า Efficient Scores และจะนำ Efficient Scores มาเป็นสมาชิกของเวคเตอร์ $U(\underset{\sim}{\theta})$ และทำการหาอนุพันธ์ย่อยอันดับที่สอง (Second Partial Derivatives) ในสมการลอการิทึมของฟังก์ชันความควรจะเป็น ℓ เทียบกับพารามิเตอร์ $\underset{\sim}{\theta}$ และจะนำผลที่ได้มาเป็นสมาชิกของเมทริกซ์ $H(\underset{\sim}{\theta})$ โดยเรียกเมทริกซ์นี้ว่า Hassian Matrix แล้วทำการแทนค่า $U(\underset{\sim}{\theta})$ และ $H(\underset{\sim}{\theta})$ ลงในสมการข้างล่างนี้

$$\hat{\underset{\sim}{\theta}}^{(m+1)} = \hat{\underset{\sim}{\theta}}^{(m)} - (H^{-1}(\underset{\sim}{\theta}))^{(m)} (U(\underset{\sim}{\theta}))^{(m)}$$

เพื่อกระทำซ้ำจนได้ค่า $\hat{\underset{\sim}{\theta}}$ ที่มีค่าไม่แตกต่างกันมากและเป็นที่ยอมรับ ซึ่งขั้นตอนวิธี Newton-Raphson สามารถแสดงเป็นผังงานได้ดังนี้

รูปที่ 3.1 แผนผังแสดงขั้นตอนวิธี Newton-Raphson



3.3.4.2 การคำนวณหาค่าประมาณองค์ประกอบความแปรปรวนของวิธีความควรจะเป็นสูงสุดแบบมอนติคาร์โล

ขั้นตอนของวิธีความควรจะเป็นสูงสุดแบบมอนติคาร์โลสำหรับการประมาณค่าพารามิเตอร์องค์ประกอบความแปรปรวน โดยจะพิจารณาการสุ่มเป็นค่าพารามิเตอร์ตั้งต้นในตัวแบบข้ามกลุ่ม 2 ปัจจัยเชิงสุ่มและกระทำซ้ำ เพื่อประมาณค่าพารามิเตอร์องค์ประกอบความแปรปรวน มีขั้นตอนดังนี้

ขั้นที่ 1 นำค่าพารามิเตอร์ $\hat{\sigma}_\alpha^2,\hat{\sigma}_\beta^2,\hat{\sigma}_\gamma^2,\hat{\sigma}_\varepsilon^2$ ที่ได้จากวิธีความควรจะเป็นสูงสุด มาเป็นค่าพารามิเตอร์ตั้งต้นของตัวแบบข้ามกลุ่ม 2 ปัจจัยเชิงสุ่ม

ขั้นที่ 2 จำลองเซตข้อมูล $(y_1, y_2,\ldots,y_N)$ โดยกระทำซ้ำ 400 ครั้ง และประมาณค่าพารามิเตอร์ $\hat{\sigma}_\alpha^2,\hat{\sigma}_\beta^2,\hat{\sigma}_\gamma^2,\hat{\sigma}_\varepsilon^2$ ด้วยวิธีความควรจะเป็นสูงสุดในแต่ละเซตข้อมูลทุกรอบของการกระทำซ้ำ

ขั้นที่ 3 คำนวณหาค่าเฉลี่ยของตัวประมาณความควรจะเป็นสูงสุดแบบมอนติคาร์โล ของ $\hat{\sigma}_\alpha^2,\hat{\sigma}_\beta^2,\hat{\sigma}_\gamma^2,\hat{\sigma}_\varepsilon^2$ ที่ได้จากการกระทำซ้ำ

3.3.5 การคำนวณหาค่าระยะทางมาหาลาโนบิสเฉลี่ยของวิธีความควรจะเป็นสูงสุดและวิธีความควรจะเป็นสูงสุดแบบมอนติคาร์โล

ระยะทางมาหาลาโนบิสเฉลี่ยของวิธีความควรจะเป็นสูงสุด

$$Ma_{ML} = \frac{\sum_{i=1}^{m}\sqrt{\left[\underset{\sim ML}{\hat{\theta}} - \underset{\sim}{\theta}\right]' \Sigma_{ML}^{-1}\left[\underset{\sim ML}{\hat{\theta}} - \underset{\sim}{\theta}\right]}}{m}$$

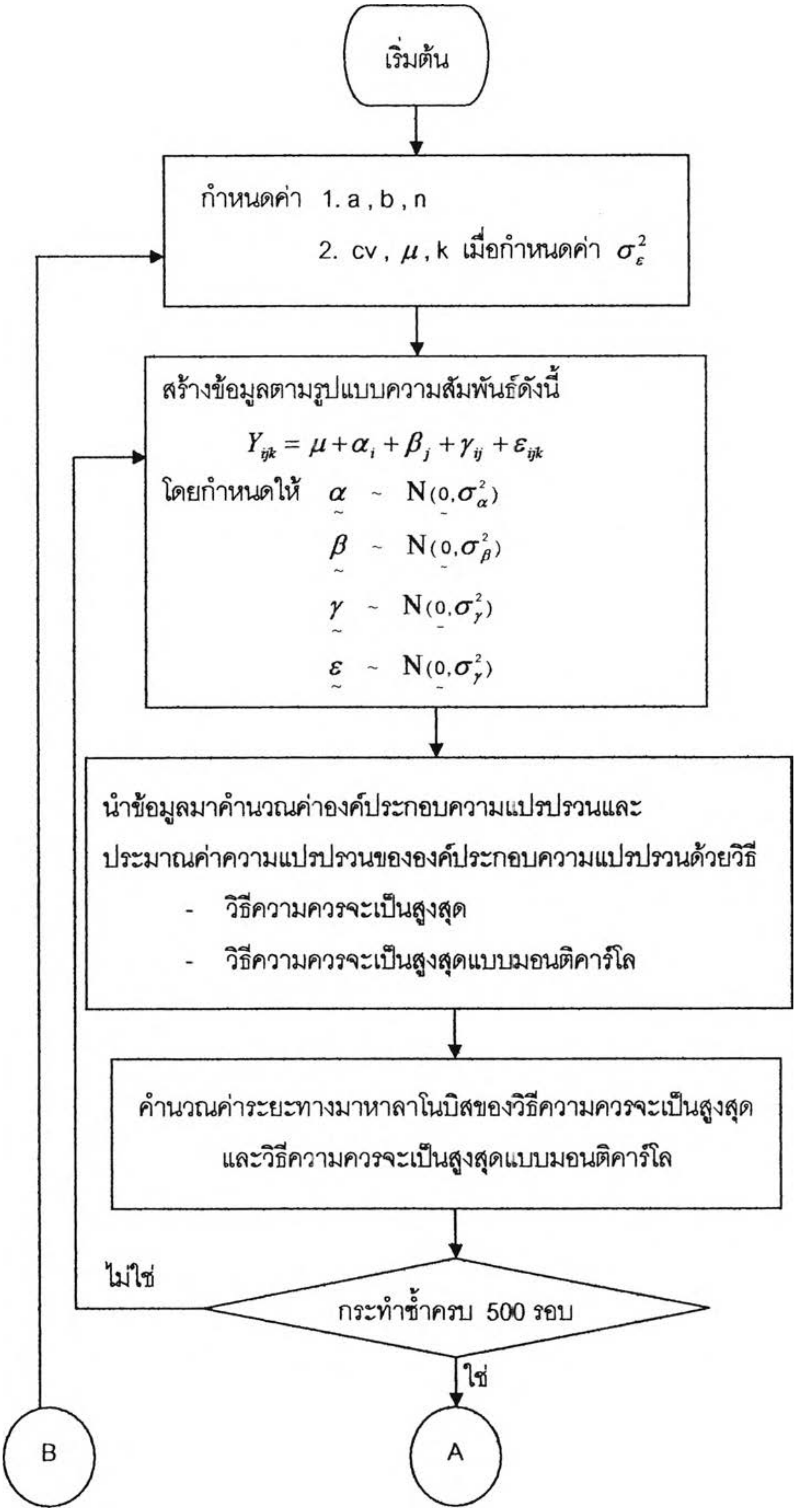
ระยะทางมาหาลาโนบิสเฉลี่ยของวิธีความควรจะเป็นสูงสุดแบบมอนติคาร์โล

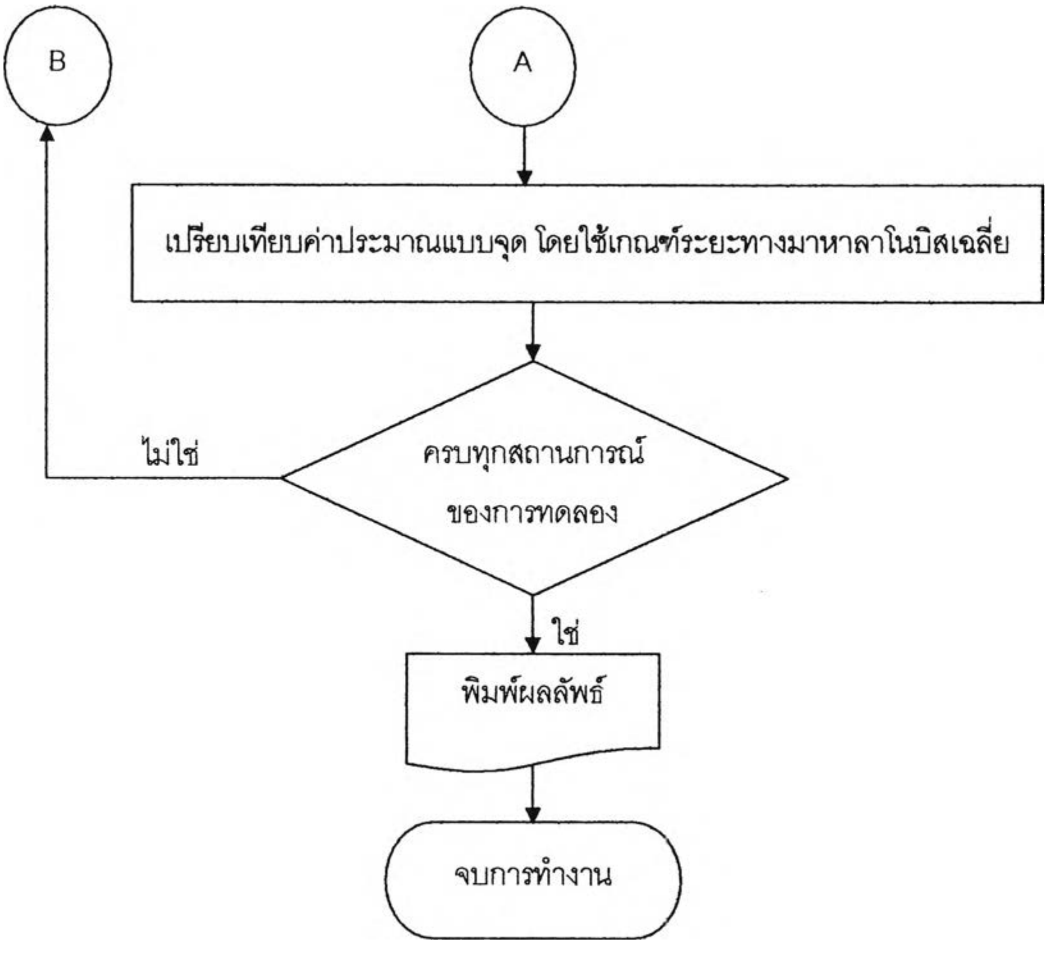
$$Ma_{MC} = \frac{\sum_{i=1}^{m}\sqrt{\left[\underset{\sim MC}{\hat{\theta}} - \underset{\sim}{\theta}\right]' \Sigma_{MC}^{-1}\left[\underset{\sim MC}{\hat{\theta}} - \underset{\sim}{\theta}\right]}}{m}$$

เมื่อ m คือ จำนวนครั้งของการกระทำซ้ำในแต่ละการทดลอง (ที่ทำให้ค่าประมาณมีค่าเป็นบวก) ซึ่งในการวิจัยครั้งนี้กำหนดให้ m มีค่าเท่ากับ 500

ดังนั้นถ้าวิธีการใดให้ค่าระยะทางมาหาลาโนบิสเฉลี่ยต่ำกว่าก็จะเป็นวิธีการประมาณที่เหมาะสมกว่าในภาพรวมของการประมาณ กล่าวคือค่าองค์ประกอบความแปรปรวนที่ได้โดยรวมแล้วมีค่าใกล้เคียงกับค่าจริงขององค์ประกอบความแปรปรวนมากกว่านั่นเอง

รูปที่ 3.2 แผนผังในการทำงานของโปรแกรม




B
A
เปรียบเทียบค่าประมาณแบบจุด โดยใช้เกณฑ์ระยะทางมาหาลาโนบิสเฉลี่ย
ไม่ใช่
ครบทุกสถานการณ์
ของการทดลอง
ใช่
พิมพ์ผลลัพธ์
จบการทำงาน