บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันการวางแผนการทดลอง (Experiment Designs) มีบทบาทในการทดลองทางด้านต่างๆมากมาย เช่น ด้านเกษตรกรรม ด้านอุตสาหกรรม และในห้องปฏิบัติการทางการแพทย์ เป็นต้น ทั้งนี้เนื่องจากการทดลองสำหรับตัวแบบข้ามกลุ่ม 2 ปัจจัย (Two-way Crossed Classification Model) สามารถทำการศึกษาได้ครั้งละหลายๆปัจจัย และใช้จำนวนหน่วยทดลองน้อยกว่าการศึกษาครั้งละ 1 ปัจจัย แต่ทำการทดลองหลายๆครั้ง อีกทั้งยังสามารถศึกษาผลกระทบร่วมระหว่างปัจจัยต่างๆได้

การประมาณค่าองค์ประกอบความแปรปรวน (Variance component) มีความสำคัญและจำเป็นอย่างมาก เพราะเป็นพารามิเตอร์ที่แสดงค่าความแปรปรวนของปัจจัยที่สนใจศึกษาในแผนแบบการทดลอง ซึ่งในการวิจัยครั้งนี้ทำการประมาณค่าพารามิเตอร์องค์ประกอบความแปรปรวนสำหรับตัวแบบข้ามกลุ่ม 2 ปัจจัย (Two-way Crossed Classification Model) เชิงสุ่ม โดยมีพารามิเตอร์แสดงค่าความแปรปรวน 4 ตัว คือความแปรปรวนของปัจจัย A ความแปรปรวนของปัจจัย B ความแปรปรวนของผลกระทบร่วมระหว่างปัจจัย A กับ ปัจจัย B และความแปรปรวนของความคลาดเคลื่อน

ในการวิจัยครั้งนี้จะใช้วิธีความควรจะเป็นสูงสุด (Maximum Likelihood Method) ซึ่งวิธีนี้จะพิจารณาว่าพารามิเตอร์ค่าองค์ประกอบความแปรปรวนเป็นค่าคงที่ที่ไม่ทราบค่าและต้องการประมาณค่า โดย Herbech (1995) ได้เสนอวิธีการประมาณค่าองค์ประกอบความแปรปรวนด้วยวิธีความควรจะเป็นสูงสุด ซึ่งเป็นการทำให้ฟังก์ชันความควรจะเป็นมีค่าสูงสุด (Maximum Likelihood : ML) บนปริภูมิพารามิเตอร์ (Parameter Space) ซึ่งมีค่าไม่เป็นลบ ทำให้ตัวประมาณที่ได้มีค่าไม่เป็นลบ โดยพบว่าตัวประมาณความควรจะเป็นสูงสุด จะไม่คำนึงถึงผลกระทบคงที่ที่มีต่อองศาความเป็นอิสระ (Degrees of Freedom) และยังพบว่าตัวประมาณมีความยุ่งยากในการคำนวณ เนื่องจากค่าประมาณที่ได้มาจากการแก้สมการที่ไม่ใช่เชิงเส้น (Nonlinear Equation)

วิธีความควรจะเป็นสูงสุดแบบมอนติคาร์โล (Monte Carlo Maximum Likelihood Method) เป็นการสร้างข้อมูลของตัวอย่างสุ่ม โดยทำการประมาณค่าพารามิเตอร์องค์ประกอบความแปรปรวนด้วยวิธีความควรจะเป็นสูงสุด แล้วใช้ประโยชน์จากคอมพิวเตอร์ในเรื่องการคำนวณ เพื่อการอนุมานเชิงสถิติเกี่ยวกับประชากรที่สนใจ และเนื่องจากในปัจจุบันเทคโนโลยีทางด้านคอมพิวเตอร์ได้มีการพัฒนาความสามารถและประสิทธิภาพของโปรแกรมที่อำนวยความ

สะดวกต่อการคำนวณ ส่งผลให้เกิดประโยชน์ต่อการพัฒนาแนวคิดใหม่ๆ ซึ่งวิธีความควรจะเป็นสูงสุดแบบมอนติคาร์โลเป็นแนวความคิดใหม่ที่นำวิธีมอนติคาร์โลมาใช้ในการประมาณค่าองค์ประกอบความแปรปรวน โดยจะทำการพิจารณาให้ค่าพารามิเตอร์ตั้งต้นในตัวแบบได้มาจากการสุ่มตัวอย่างและมีการกระทำซ้ำ เพื่อประมาณค่าพารามิเตอร์องค์ประกอบความแปรปรวนสำหรับตัวแบบข้ามกลุ่ม 2 ปัจจัยเชิงสุ่ม

จากหลักการดังกล่าวข้างต้น จึงทำให้ผู้วิจัยสนใจทำการศึกษาและเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์องค์ประกอบความแปรปรวนสำหรับตัวแบบข้ามกลุ่ม 2 ปัจจัยเชิงสุ่มด้วยวิธีความควรจะเป็นสูงสุดและวิธีความควรจะเป็นสูงสุดแบบมอนติคาร์โลว่าวิธีการใดจะให้ค่าประมาณองค์ประกอบความแปรปรวนใกล้เคียงกับค่าจริงของพารามิเตอร์องค์ประกอบความแปรปรวนมากที่สุด

1.2 วัตถุประสงค์ของการวิจัย

เพื่อศึกษาและเปรียบเทียบการประมาณค่าองค์ประกอบความแปรปรวน สำหรับตัวแบบข้ามกลุ่ม 2 ปัจจัย (Two-way Crossed Classification Model) เชิงสุ่ม ในกรณีที่ใช้วิธีการสุ่มสมบูรณ์ 2 วิธีคือ

1.2.1 วิธีความควรจะเป็นสูงสุด (Maximum Likelihood Method)

1.2.2 วิธีความควรจะเป็นสูงสุดแบบมอนติคาร์โล (Monte Carlo Maximum Likelihood Method)

1.3 สมมติฐานของการวิจัย

การประมาณค่าองค์ประกอบความแปรปรวนด้วยวิธีความควรจะเป็นสูงสุดแบบมอนติคาร์โล (Monte Carlo Maximum Likelihood Method) จะให้ค่าประมาณที่มีค่าใกล้เคียงกับค่าจริงขององค์ประกอบความแปรปรวนมากกว่าการประมาณค่าองค์ประกอบความแปรปรวนด้วยวิธีความควรจะเป็นสูงสุด(Maximum Likelihood Method)

1.4 ข้อตกลงเบื้องต้น

1.4.1 ทำการศึกษาเฉพาะตัวแบบข้ามกลุ่ม 2 ปัจจัย (Two-way Crossed Classification Model) ซึ่งได้รับอิทธิพลเชิงสุ่ม (Random Effect) โดยที่ปัจจัย A มี a ระดับ ปัจจัย B มี b ระดับ และจัดให้ทุกระดับของปัจจัย A ปรากฏอยู่ในทุกระดับของปัจจัย B ซึ่งทำให้ได้ปัจจัยทดลองผสม $a \times b$ วิธี และในแต่ละวิธีจะมีหน่วยทดลองเท่ากันคือ n หน่วย และใช้วิธีการสุ่ม

สมบูรณ์ (Complete Randomized Design : CRD) ในการจัดหน่วยทดลองให้กับปัจจัย โดยมีตัวแบบดังนี้

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \qquad \text{โดยที่} \quad i=1,\ldots,a \quad j=1,\ldots,b \quad k=1,\ldots,n$$

โดยที่ Y_{ijk} แทนค่าสังเกตที่ k ระดับที่ i ของปัจจัย A และระดับที่ j ของปัจจัย B

μ แทนค่าเฉลี่ยรวม

α_i แทนผลกระทบระดับที่ i ของปัจจัย A

β_j แทนผลกระทบระดับที่ j ของปัจจัย B

γ_{ij} แทนผลกระทบร่วมระดับที่ i ของปัจจัย A และระดับที่ j ของปัจจัย B

ε_{ijk} แทนความคลาดเคลื่อนของค่าสังเกตที่ k ระดับที่ i ของปัจจัย A และระดับที่ j ของปัจจัย B

1.4.2 สมมติให้ μ เป็นพารามิเตอร์ที่ไม่ทราบค่า และพารามิเตอร์ α_i, β_j, γ_{ij} และ ε_{ijk} เป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติและเป็นอิสระซึ่งกันและกัน * โดยที่

$$\alpha_i \sim \text{nid}(0, \sigma_\alpha^2)$$
$$\beta_j \sim \text{nid}(0, \sigma_\beta^2)$$
$$\gamma_{ij} \sim \text{nid}(0, \sigma_\gamma^2)$$
$$\varepsilon_{ijk} \sim \text{nid}(0, \sigma_\varepsilon^2)$$

ค่าเฉลี่ยและความแปรปรวนของข้อมูล Y_{ijk} คือ

$$E(Y_{ijk}) = \mu$$

$$Var(Y_{ijk}) = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2 + \sigma_\varepsilon^2$$

จะได้ว่า $Y_{ijk} \sim N(\mu, \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2 + \sigma_\varepsilon^2)$

ดังนั้น $\bar{Y}_{...} \sim N(\mu, \frac{1}{abn}(bn\sigma_\alpha^2 + an\sigma_\beta^2 + n\sigma_\gamma^2 + \sigma_\varepsilon^2))$

---

* เป็นอิสระที่มีเงื่อนไข (Conditional Independence) จากการกำหนด $\sigma_\alpha^2 = \sigma_\beta^2 = \sigma_\gamma^2 = h\sigma_\varepsilon^2$ จะได้ว่า $\sigma_\alpha^2, \sigma_\beta^2$ และ σ_γ^2 เป็นอิสระซึ่งกันและกันเมื่อกำหนด σ_ε^2

$$COV(Y_{ijk}, Y_{i'j'k'}) = \begin{cases} \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2 + \sigma_\varepsilon^2 & \text{สำหรับ } i = i', j = j', k = k' \\ \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2 & \text{สำหรับ } i = i', j = j', k \neq k' \\ \sigma_\alpha^2 & \text{สำหรับ } i = i', j \neq j' \\ \sigma_\beta^2 & \text{สำหรับ } i \neq i', j = j' \\ 0 & \text{สำหรับ } i \neq i', j \neq j' \end{cases}$$

โดยเรียก $\sigma_\alpha^2, \sigma_\beta^2, \sigma_\gamma^2$ และ σ_ε^2 ว่าพารามิเตอร์องค์ประกอบความแปรปรวน (Variance-Components Parameter)

ตารางที่ 1.1 ตารางวิเคราะห์ความแปรปรวนสำหรับตัวแบบข้ามกลุ่ม 2 ปัจจัยเชิงสุ่ม

สาเหตุ	องศาความเป็นอิสระ	ผลรวมกำลังสอง	ผลรวมกำลังสองเฉลี่ย	ค่าคาดหวังของผลรวมกำลังสองเฉลี่ย
ปัจจัย A	(a-1)	$SSA = bn(\sum \bar{y}_{i..} - \bar{y}_{...})^2$	$MSA = \frac{SSA}{(a-1)}$	$\sigma_\varepsilon^2 + n\sigma_\gamma^2 + bn\sigma_\alpha^2$
ปัจจัย B	(b-1)	$SSB = an(\sum \bar{y}_{.j.} - \bar{y}_{...})^2$	$MSB = \frac{SSB}{(b-1)}$	$\sigma_\varepsilon^2 + n\sigma_\gamma^2 + an\sigma_\beta^2$
ผลกระทบร่วม	(a-1)(b-1)	$SSAB = n\sum\sum(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$	$MSAB = \frac{SSAB}{(a-1)(b-1)}$	$\sigma_\varepsilon^2 + n\sigma_\gamma^2$
ความคลาดเคลื่อน	ab(n-1)	$SSE = \sum\sum\sum(y_{ijk} - \bar{y}_{ij.})^2$	$MSE = \frac{SSE}{ab(n-1)}$	σ_ε^2
รวม	abn-1	$SST = \sum\sum\sum(y_{ijk} - \bar{y}_{...})^2$		

โดยให้ค่าต่างๆที่อยู่ในตารางเป็นดังนี้

y_{ijk} แทนค่าสังเกตตัวที่ k ในระดับที่ i ของปัจจัย A และในระดับที่ j ของปัจจัย B

$\bar{y}_{...}$ แทนค่าเฉลี่ยของค่าสังเกตทุกตัวและในทุกระดับของปัจจัย A และในทุกระดับของปัจจัย B

โดยที่ $\bar{y}_{...} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}}{abn}$

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

$\bar{y}_{i..}$ แทนค่าเฉลี่ยของค่าสังเกตทุกตัวและในทุกระดับของปัจจัย B และในระดับที่ i ของปัจจัย A

โดยที่ $\bar{y}_{i..} = \dfrac{\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}}{bn}$

$\bar{y}_{.j.}$ แทนค่าเฉลี่ยของค่าสังเกตทุกตัวและในทุกระดับของปัจจัย A และในระดับที่ j ของปัจจัย B

โดยที่ $\bar{y}_{.j.} = \dfrac{\sum_{i=1}^{a}\sum_{k=1}^{n} y_{ijk}}{an}$

$\bar{y}_{ij.}$ แทนค่าเฉลี่ยของค่าสังเกตทุกตัวในระดับที่ j ของปัจจัย B และในระดับที่ i ของปัจจัย A

โดยที่ $\bar{y}_{ij.} = \dfrac{\sum_{k=1}^{n} y_{ijk}}{n}$

1.5 ขอบเขตของการวิจัย

1.5.1 ศึกษาวิธีการประมาณค่าองค์ประกอบความแปรปรวนสำหรับตัวแบบข้ามกลุ่ม 2 ปัจจัยเชิงสุ่มด้วยวิธีความควรจะเป็นสูงสุดและวิธีความควรจะเป็นสูงสุดแบบมอนติคาร์โล

1.5.2 ตัวแบบเป็นปัจจัยเชิงสุ่ม (Random Effect Model) ในแผนแบบของการทดลองสำหรับตัวแบบข้ามกลุ่ม 2 ปัจจัย โดยจัดหน่วยทดลองให้กับปัจจัยด้วยวิธีการสุ่มสมบูรณ์ (Complete Randomized Design : CRD)

1.5.3 ระดับปัจจัย และขนาดหน่วยทดลองที่ใช้คือ (a,b,n) มีค่าดังตารางข้างล่างนี้

(2,2,2)	(2,2,3)	(2,2,4)
(3,3,2)	(3,3,3)	(3,3,4)
(4,4,2)	(4,4,3)	(4,4,4)

1.5.4 กำหนดพารามิเตอร์ดังนี้

- ค่าเฉลี่ยรวม(μ) = 40*
- ค่าความแปรปรวนของความคลาดเคลื่อนของการทดลอง (σ_ε^2) โดยมีหลักการ ดังนี้

เนื่องจาก $C.V.(Y_{ijk}) = \frac{SD(Y_{ijk})}{\mu} = \frac{\sqrt{\sigma_\alpha^2+\sigma_\beta^2+\sigma_\gamma^2+\sigma_\varepsilon^2}}{\mu}$

ในการศึกษาครั้งนี้กำหนดให้ $\sigma_\alpha^2 = \sigma_\beta^2 = \sigma_\gamma^2 = h\sigma_\varepsilon^2$ **

จะได้ว่า $C.V.(Y_{ijk}) = \frac{\sqrt{h\sigma_\alpha^2+h\sigma_\beta^2+h\sigma_\gamma^2+\sigma_\varepsilon^2}}{\mu} = \frac{\sqrt{(3h+1)\sigma_\varepsilon^2}}{\mu}$

ดังนั้น $\sigma_\varepsilon^2 = \frac{(C.V.\times\mu)^2}{3h+1}$

เมื่อ h เป็นค่าจำนวนเต็มคงที่ โดยกำหนดให้มีค่าเท่ากับ 1 , 2 และ 3

1.5.5 กำหนดให้ข้อมูลมีค่าสัมประสิทธิ์ความแปรผัน (Coefficient of Variation : C.V.) ในระดับต่างๆกันคือ 10% , 50% และ 90% จะได้ส่วนเบี่ยงเบนมาตรฐาน(σ) เท่ากับ 4 , 20 และ 36 ตามลำดับ

1.5.6 ในการวิจัยครั้งนี้สร้างแบบการจำลองข้อมูลโดยใช้เทคนิคมอนติคาร์โล (Monte Carlo Simulation) ซึ่งทำการจำลองข้อมูล Y_{ijk} ที่สุ่มมาจากแต่ละประชากรที่มีการแจกแจงแบบปกติ (Normal Distribution) โดยใช้ฟังก์ชันสำเร็จรูปที่มีอยู่ในโปรแกรม S-PLUS 2000 คือ ฟังก์ชัน morm(n,mean,sd)

เมื่อ n แทนขนาดหน่วยตัวอย่างที่ต้องการ

mean แทนค่าเฉลี่ยของประชากร (μ)

sd แทนส่วนเบี่ยงเบนมาตรฐานของประชากร(σ)

---

* ในการวิจัยครั้งนี้ การกำหนดค่า μ ณ ระดับต่างๆจะไม่ทำให้ผลที่ได้เปลี่ยนแปลงไป ดังนั้นการวิจัยครั้งนี้จึงกำหนดให้ μ เท่ากับ 40

** ในทางปฏิบัติ การกำหนดให้ $\sigma_\alpha^2 = \sigma_\beta^2 = \sigma_\gamma^2 = h\sigma_\varepsilon^2$ ไม่จำเป็นเสมอไป แต่ในการวิจัยครั้งนี้กำหนดให้มีค่าดังกล่าวเพื่อความสะดวกในการวิจัย และทำให้การจำลองข้อมูลที่นำมาใช้ในการศึกษามีหลักเกณฑ์มากขึ้น

โดยในแต่ละสถานการณ์ของการทดลองจะกระทำซ้ำ 500 รอบ และการสุ่มตัวอย่างในวิธีความควรจะเป็นสูงสุดแบบมอนติคาร์โลจะกระทำซ้ำ 400 รอบ

1.6 คำจำกัดความที่ใช้ในการวิจัย

1.6.1 หน่วยทดลอง (Experiment Unit) หมายถึง สิ่งหนึ่งหรือกลุ่มหนึ่งของวัตถุที่นำมาใช้ในการทดลอง ซึ่งหน่วยทดลองแต่ละหน่วยจะได้รับสิ่งทดลองที่นำมาศึกษา และนำผลที่ได้มาหาวิเคราะห์หาข้อสรุป

1.6.2 ปัจจัย (Factor) หมายถึง ตัวแปรที่กำลังศึกษาว่ามีความสัมพันธ์หรือมีผลกระทบต่อตัวแปรตาม

1.6.3 ระดับของปัจจัย (Factor Level) หมายถึง ประเภทต่างๆของปัจจัยหรือตัวแปรอิสระที่กำลังศึกษาอยู่

1.6.4 แผนการทดลองแบบสุ่มสมบูรณ์ (Complete Randomized Design : CRD) หมายถึง แผนการทดลองที่หน่วยทดลองเป็นหน่วยที่ได้มาจากการสุ่ม

1.6.5 สัมประสิทธิ์ความแปรผัน (Coefficient of Variation : C.V.) หมายถึง ค่าที่ใช้วัดการกระจายของข้อมูล โดยคิดในรูปสัมประสิทธิ์การกระจาย ซึ่งไม่มีหน่วยและคำนวณค่าออกมาเป็นเปอร์เซนต์

1.6.6 ระยะทางมาหาลาโนบิส (Mahalanobis Distance) หมายถึง ค่าที่ใช้ในการเปรียบเทียบองค์ประกอบความแปรปรวนที่ศึกษากับค่าจริงขององค์ประกอบความแปรปรวนแต่ละตัว โดยทำการเปรียบเทียบขนาดและทำการถ่วงน้ำหนักด้วยความแปรปรวนร่วม

1.6.7 ข้อมูลสมดุล (Balanced data) หมายถึง ข้อมูลที่มีจำนวนค่าสังเกตในแต่ละระดับของปัจจัยเท่ากัน

1.7 ประโยชน์ที่คาดว่าจะได้รับ

1.7.1 สามารถใช้วิธีความควรจะเป็นสูงสุดแบบมอนติคาร์โลประมาณค่าองค์ประกอบความแปรปรวนในแผนแบบการทดลองสำหรับตัวแบบข้ามกลุ่ม 2 ปัจจัยเชิงสุ่ม

1.7.2 สามารถเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์องค์ประกอบความแปรปรวนระหว่างวิธีความควรจะเป็นสูงสุดกับวิธีความควรจะเป็นสูงสุดแบบมอนติคาร์โล ว่าวิธีการใดให้ค่าประมาณพารามิเตอร์องค์ประกอบความแปรปรวนได้ดีกว่ากัน

1.7.3 เพื่อเป็นแนวทางในการศึกษาการประมาณค่าพารามิเตอร์องค์ประกอบความแปรปรวนสำหรับแผนแบบการทดลองของตัวแบบอื่นๆ

1.8 วิธีดำเนินการวิจัย

1.8.1 ศึกษาการประมาณค่าพารามิเตอร์องค์ประกอบความแปรปรวนสำหรับตัวแบบข้ามกลุ่ม 2 ปัจจัยเชิงสุ่ม 2 วิธีคือ

1.8.1.1 วิธีความควรจะเป็นสูงสุด

1.8.1.2 วิธีความควรจะเป็นสูงสุดแบบมอนติคาร์โล

1.8.2 ศึกษาการเขียนโปรแกรมจำลองค่าสังเกตในตัวแบบข้ามกลุ่ม 2 ปัจจัยเชิงสุ่ม และเขียนโปรแกรมการประมาณค่าองค์ประกอบความแปรปรวนของทั้ง 2 วิธี

1.8.3 จำลองข้อมูลตามขอบเขตที่ต้องการศึกษา

1.8.4 ศึกษาและเปรียบเทียบวิธีการประมาณค่าองค์ประกอบความแปรปรวนของทั้ง 2 วิธี โดยใช้ระยะทางมาหาลาโนบิสเฉลี่ย