บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในการอนุมานข้อมูลทางสถิติมีการวิเคราะห์ข้อมูลได้หลายวิธี เช่นการประมาณค่าและการทดสอบสมมติฐาน องค์ประกอบความแปรปรวน (Variance Components) เป็นพารามิเตอร์ที่น่าสนใจมากตัวหนึ่งในแผนแบบการทดลอง (Design Experiment) เมื่อเราสนใจที่จะศึกษาความคลาดเคลื่อนที่เกิดจากปัจจัยต่าง ๆ นอกเหนือไปจากความคลาดเคลื่อนที่เกิดจากการทดลองซึ่งเราไม่สามารถควบคุมได้ โดยปรกติการคำนวณองค์ประกอบความแปรปรวนโดยใช้ตัวแบบเต็มรูป (Full Model) จะให้ผลรวดเร็วกว่าและใช้เวลาในการคำนวณไม่มากนัก แต่การประมาณองค์ประกอบความแปรปรวนโดยใช้วิธีการเฉลี่ยตัวแบบ (Model Averaging) เป็นวิธีการที่น่าสนใจมากกว่า โดยทำการลดพารามิเตอร์ของตัวแบบเต็มรูปทีละตัวเพื่อให้ได้ตัวแบบที่เป็นไปได้ทั้งหมด จากนั้นจึงทำการประมาณองค์ประกอบความแปรปรวนของแต่ละปัจจัยตามตัวแบบที่ได้แล้วนำองค์ประกอบความแปรปรวนของแต่ละปัจจัยที่เกิดจากสาเหตุความแปรปรวนเดียวกันมาทำการเฉลี่ย วัตถุประสงค์ก็เพื่อพัฒนาให้เกิดแนวความคิดและวิธีการใหม่ ๆ โดยที่เราคาดคะเนว่าการประมาณองค์ประกอบความแปรปรวนโดยใช้วิธีการเฉลี่ยตัวแบบจะทำให้เกิดความคลาดเคลื่อนในการประมาณองค์ประกอบความแปรปรวนโดยรวมที่ต่ำกว่าการประมาณองค์ประกอบความแปรปรวนโดยใช้ตัวแบบเต็มรูปเพียงตัวแบบเดียว ถึงแม้ว่าวิธีการคำนวณจะค่อนข้างซับซ้อนมากกว่าแต่ในปัจจุบันเทคโนโลยีทางด้านคอมพิวเตอร์ได้มีการพัฒนาความสามารถ และประสิทธิภาพของโปรแกรมทางด้านซอฟแวร์ที่เอื้ออำนวยต่อการการคำนวณ ส่งผลให้เกิดประโยชน์ต่อการพัฒนาแนวความคิดใหม่ ๆ นอกจากนี้ยังสามารถนำความรู้ที่ได้ไปประยุกต์ใช้กับแผนแบบการทดลอง อื่น ๆ ได้อีกจึงทำให้ความรู้ทางวิชาการถูกพัฒนาขึ้นเรื่อยๆ

โดยทั่วไปแผนแบบการทดลองจะแบ่งออกเป็น 3 ลักษณะคือ 1) ตัวแบบคงที่ (Fixed Model) สำหรับระดับปัจจัยที่ศึกษาเป็นปัจจัยคงที่ (Fixed Factor) ซึ่งมาจากระดับของประชากรที่สนใจทั้งหมด หรือระดับปัจจัยถูกระบุอย่างชัดเจนโดยผู้ทำการทดลองโดยที่ระดับปัจจัยนั้นไม่มีการเปลี่ยนแปลงไม่ว่าจะทำการทดลองเมื่อใด ในการอนุมานเราสามารถสรุปผลได้เพียงระดับปัจจัยที่เรานำมาศึกษาเท่านั้น 2) ตัวแบบเชิงสุ่ม (Random Model) สำหรับกรณีที่ระดับปัจจัยที่นำมา

ศึกษาเป็นปัจจัยสุ่ม (Random Factor) ที่ถูกสุ่มมาจากระดับของประชากรทั้งหมด ในการอนุมานเราสามารถสรุปผลถึงระดับปัจจัยทั้งหมดที่เป็นไปได้ 3) ตัวแบบผสม (Mixed Model) มีทั้งตัวแบบคงที่และตัวแบบเชิงสุ่มอยู่ในแผนแบบการทดลองเดียวกัน ในการศึกษาครั้งนี้สนใจแผนแบบการทดลองจตุรัสละตินตัวแบบเชิงสุ่มที่ไม่มีการทำซ้ำ ซึ่งมีปัจจัยที่สนใจหนึ่งอย่างคือปัจจัยวิธีการทดลอง และบล็อกด้วยปัจจัยแบ่งบล็อกทั้งสองปัจจัยคือ ปัจจัยแบ่งบล็อกปัจจัยแรกตามแถว และปัจจัยแบ่งบล็อกปัจจัยสองตามสดมภ์ และในแต่ละปัจจัยเป็นอิสระจากกันโดยปัจจัยวิธีการทดลองจะปรากฏขึ้นเพียงครั้งเดียวตามแถวและสดมภ์ของปัจจัยแบ่งบล็อก ดังนั้นแผนแบบการทดลองจตุรัสละตินที่ศึกษาจะมีองค์ประกอบความแปรปรวน 4 ตัว คือ องค์ประกอบความแปรปรวนเนื่องจากปัจจัยวิธีการทดลอง ปัจจัยแบ่งบล็อกปัจจัยแรกตามแถว ปัจจัยแบ่งบล็อกปัจจัยสองตามสดมภ์ และความคลาดเคลื่อน

1.2 วัตถุประสงค์ของการวิจัย

เพื่อศึกษาและเปรียบเทียบการประมาณองค์ประกอบความแปรปรวนสำหรับตัวแบบจตุรัสละติน 2 แบบ คือ

1.2.1 การประมาณค่าโดยใช้ตัวแบบเต็มรูป (Full Model Estimation)

1.2.2 การประมาณค่าโดยใช้วิธีการเฉลี่ยตัวแบบ (Model Averaging Estimation)

1.3 สมมติฐานของการวิจัย

การประมาณองค์ประกอบความแปรปรวนโดยใช้วิธีการเฉลี่ยตัวแบบสำหรับตัวแบบจตุรัสละตินจะให้ค่าประมาณใกล้เคียงค่าจริงขององค์ประกอบความแปรปรวนมากกว่าการประมาณองค์ประกอบความแปรปรวนโดยใช้ตัวแบบเต็มรูป

1.4 ข้อตกลงเบื้องต้น

1.4.1 ศึกษาภายใต้ตัวแบบจตุรัสละติน สมมติว่าปัจจัยที่ศึกษามี 3 อย่างคือ ปัจจัยแบ่งบล็อกปัจจัยแรกมี p ระดับ ปัจจัยแบ่งบล็อกปัจจัยสองตามสดมภ์มี p ระดับ และปัจจัยวิธีการทดลองมี p ระดับ ปัจจัยวิธีการทดลองปรากฏขึ้นเพียงครั้งเดียวตามแถวและสดมภ์ของปัจจัยแบ่งบล็อก ดังนั้น จะมีค่าสังเกตในการทดลองหนึ่ง ๆ เท่ากับ $p \times p$ ข้อมูล จะได้ตัวแบบสำหรับ Y_{ijk} ซึ่งคือค่าสังเกตที่ i ของปัจจัยวิธีการทดลอง บล็อกที่ j ของปัจจัยแบ่งบล็อกปัจจัยแรกตามแถว บล็อกที่ k ของปัจจัยแบ่งบล็อกปัจจัยสองตามสดมภ์ดังนี้คือ

$$Y_{ijk} = \mu + \tau_i + \alpha_j + \beta_k + \varepsilon_{ijk} \quad ; \quad i, j, k = 1, \ldots, p$$

โดยที่ Y_{ijk} คือ ค่าสังเกตระดับที่ i ของปัจจัยวิธีการทดลอง ระดับที่ j ของปัจจัยแบ่งบล็อกปัจจัยแรกตามแถว และระดับที่ k ของปัจจัยปัจจัยแบ่งบล็อกปัจจัยสองตามสดมภ์ และ p เป็นจำนวนระดับของปัจจัยวิธีการทดลอง เท่ากับจำนวนระดับของปัจจัยแบ่งบล็อกทั้งสองปัจจัย ซึ่งค่าสังเกต Y_{ijk} ประกอบด้วยส่วนประกอบ 5 ส่วน ดังนี้

1) μ คือ ค่าเฉลี่ยรวมของประชากร
2) τ_i คือ ผลกระทบระดับที่ i ของปัจจัยวิธีการทดลอง
3) α_j คือ ผลกระทบระดับที่ j ของปัจจัยแบ่งบล็อกปัจจัยแรกตามแถว
4) β_k คือ ผลกระทบระดับที่ k ของปัจจัยแบ่งบล็อกปัจจัยสองตามสดมภ์
5) ε_{ijk} คือ ความคลาดเคลื่อนระดับที่ i ของปัจจัยวิธีการทดลอง ระดับที่ j ของปัจจัยแบ่งบล็อกปัจจัยแรกตามแถว และระดับที่ k ของปัจจัยปัจจัยแบ่งบล็อกปัจจัยสองตามสดมภ์

1.4.2 สมมติให้ μ เป็นพารามิเตอร์ที่ไม่ทราบค่า และ $\tau_i, \alpha_j, \beta_k, \varepsilon_{ijk}$ เป็นตัวแปรสุ่มมีการแจกแจงแบบปกติและเป็นอิสระกันดังนี้ $\tau_i \sim N(0, \sigma_\tau^2)$, $\alpha_j \sim N(0, \sigma_\alpha^2)$,

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

ดังนั้น

นั่นคือ $E\left(Y_{ijk}\right) = \mu$, $Var\left(Y_{ijk}\right) = \sigma_{\tau}^{2} + \sigma_{\alpha}^{2} + \sigma_{\beta}^{2} + \sigma_{\varepsilon}^{2}$

$$Y_{ijk} \sim \mathrm{N}\left(\mu, \sigma_{\tau}^{2} + \sigma_{\alpha}^{2} + \sigma_{\beta}^{2} + \sigma_{\varepsilon}^{2}\right)$$

$$\mathrm{Cov}\left(Y_{ijk}, Y_{i'j'k'}\right) = \begin{cases} \sigma_{\tau}^{2} + \sigma_{\alpha}^{2} + \sigma_{\beta}^{2} + \sigma_{\varepsilon}^{2} & , i = i', j = j', k = k' \\ \sigma_{\tau}^{2} & , i \neq i', j = j', k = k' \\ \sigma_{\alpha}^{2} & , i = i', j \neq j', k = k' \\ \sigma_{\beta}^{2} & , i = i', j = j', k \neq k' \\ 0 & , i \neq i', j \neq j', k \neq k' \end{cases}$$

ซึ่ง $\sigma_{\tau}^{2}, \sigma_{\alpha}^{2}, \sigma_{\beta}^{2}$ และ σ_{ε}^{2} นี้เรียกว่า พารามิเตอร์องค์ประกอบความแปรปรวน (Variance Components Parameter) ที่ต้องการประมาณ ซึ่งมีตารางวิเคราะห์ความแปรปรวนของตัวแบบจตุรัสละตินดังนี้

ตารางที่ 1 แสดงการวิเคราะห์ความแปรปรวนสำหรับตัวแบบจตุรัสละติน

สาเหตุของความแปรปรวน	องศาอิสระ	ผลรวมกำลังสอง	ผลรวมกำลังสองเฉลี่ย	ค่าคาดหวังของผลรวมกลังสองเฉลี่ย
ปัจจัยวิธีการทดลอง	$(p-1)$	$SST = (1/p)\sum_{j=1}^{p}(\sum_{k=1}^{p} y_{ijk})^2 - (\sum_{j=1}^{p}\sum_{k=1}^{p} y_{ijk})^2/p^2$	$MST = \frac{SST}{p-1}$	$\sigma_\tau^2 = \sigma_\varepsilon^2 + p\sigma_\tau^2$
ปัจจัยแบ่งบล็อกปัจจัยแรก	$(p-1)$	$SSA = (1/p)\sum_{k=1}^{p}(\sum_{j=1}^{p} y_{ijk})^2 - (\sum_{k=1}^{p}\sum_{j=1}^{p} y_{ijk})^2/p^2$	$MSA = \frac{SSA}{p-1}$	$\sigma_\alpha^2 = \sigma_\varepsilon^2 + p\sigma_\alpha^2$
ปัจจัยแบ่งบล็อกปัจจัยที่สอง	$(p-1)$	$SSB = (1/p)\sum_{i=1}^{p}(\sum_{j=1}^{p} y_{ijk})^2 - (\sum_{i=1}^{p}\sum_{j=1}^{p} y_{ijk})^2/p^2$	$MSB = \frac{SSB}{p-1}$	$\sigma_\beta^2 = \sigma_\varepsilon^2 + p\sigma_\beta^2$
ความคลาดเคลื่อน	$(p-1)(p-2)$	$SSE = \sum_{j=1}^{p}\sum_{k=1}^{p}(y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + 2\bar{y}_{...})^2$	$MSE = \frac{SSE}{(p-1)(p-2)}$	σ_ε^2
รวม	p^2-1	$SSY = \sum_{j=1}^{p}\sum_{k=1}^{p}(y_{ijk})^2 - (\sum_{j=1}^{p}\sum_{k=1}^{p} y_{ijk})^2/p^2$		$\sigma_{total}^2 = \sigma_\tau^2 + \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\varepsilon^2$

โดยให้ค่าต่าง ๆ ที่อยู่ในตารางเป็นดังนี้

y_{ijk} = ค่าสังเกตระดับที่ i ของปัจจัยวิธีการทดลอง ระดับที่ j ของปัจจัยแบ่งบล็อกปัจจัยแรกตามแถว และระดับที่ k ของปัจจัยปัจจัยแบ่งบล็อกปัจจัยสองตามสดมภ์

$\bar{y}_{...}$ = ค่าเฉลี่ยของค่าสังเกตทุกตัวและในทุกระดับของปัจจัยแบ่งบล็อกปัจจัยแรกตามแถว และในทุกระดับของปัจจัยแบ่งบล็อกปัจจัยสองตามสดมภ์

$$= \frac{\sum_{j=1}^{p}\sum_{k=1}^{p} y_{ijk}}{p^2}$$

$\bar{y}_{i..}$ = ค่าเฉลี่ยของค่าสังเกตทุกตัวในทุกระดับของปัจจัยแบ่งบล็อกปัจจัยแรกตามแถว หรือปัจจัยแบ่งบล็อกปัจจัยสองตามสดมภ์ ในระดับที่ i ของปัจจัยวิธีการทดลอง

$$= \frac{\sum_{j=1}^{p} y_{ijk}}{p} = \frac{\sum_{k=1}^{p} y_{ijk}}{p}$$

$\bar{y}_{.j.}$ = ค่าเฉลี่ยของค่าสังเกตทุกตัวในทุกระดับของปัจจัยแบ่งบล็อกปัจจัยสองตามสดมภ์ ในระดับที่ j ของปัจจัยแบ่งบล็อกปัจจัยแรกตามแถว

$$= \frac{\sum_{k=1}^{p} y_{ijk}}{p}$$

$\bar{y}_{..k}$ = ค่าเฉลี่ยของค่าสังเกตทุกตัวในทุกระดับของปัจจัยแบ่งบล็อกปัจจัยแรกตามแถว ในระดับที่ k ของปัจจัยปัจจัยแบ่งบล็อกปัจจัยสองตามสดมภ์

$$= \frac{\sum_{j=1}^{p} y_{ijk}}{p}$$

1.5 ข้อจำกัดของการวิจัย

1.5.2 กำหนดระดับของปัจจัยคือ $i, j, k = 1, \ldots, p$ โดย $p = 3, 4, 5$ ดังนั้น ขนาดตัวอย่างที่ใช้ในแต่ละระดับของปัจจัยสำหรับตัวแบบจตุรัสละตินเท่ากับ 9, 16 และ 25 ตามลำดับ

1.5.2 กำหนดให้ข้อมูลมีค่าสัมประสิทธิ์ความแปรผัน (Coefficient of variation: $C.V.\left(Y_{ijk}\right)$) ในระดับต่าง ๆ กัน คือ 5%, 15%, 25%, 35% , 45% และ 55% และค่าเฉลี่ย (μ) เท่ากับ 40

จาก

$$C.V.\left(Y_{ijk}\right) = \frac{S.D.\left(Y_{ijk}\right)}{\mu} = \frac{\sqrt{\sigma_\tau^2 + \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\varepsilon^2}}{\mu}$$

กำหนดให้

$\sigma_\tau^2 = \sigma_\alpha^2 = \sigma_\beta^2 = m\sigma_\varepsilon^{2*}$ โดยที่ m เป็นค่าจำนวนเต็มคงที่เท่ากับ 0.5, 1.0, 1.5, 2.0, 2.5 และ 3.0

นั่นคือ

$$C.V.\left(Y_{ijk}\right) = \frac{\sqrt{m\sigma_\varepsilon^2 + m\sigma_\varepsilon^2 + m\sigma_\varepsilon^2 + \sigma_\varepsilon^2}}{\mu} = \frac{\sigma_\varepsilon\sqrt{3m+1}}{\mu}$$

ดังนั้นจะได้ว่า

$$\sigma_\varepsilon^2 = \left(\frac{C.V.\left(Y_{ijk}\right) \times \mu}{\sqrt{3m+1}}\right)^2$$

โดยมีความแปรปรวน $\sigma_{Y_{ijk}}^2$ เท่ากับ 4, 36 , 100 , 196 , 324 และ 484 ตามลำดับ

---

*การกำหนดให้ $\sigma_\tau^2 = \sigma_\alpha^2 = \sigma_\beta^2 = m\sigma_\varepsilon^2$ เพื่อสะดวกต่อการวิจัย และทำให้การจำลองแบบข้อมูลที่นำมาใช้ในการศึกษามีหลักเกณฑ์มากขึ้น

1.5.3 ในการวิจัยครั้งนี้สร้างแบบจำลองข้อมูลโดยใช้เทคนิคมอนติคาร์โลซิมูเลชั่น (Monte Carlo Simulation Technique) เป็นเทคนิคจำลองข้อมูลโดยใช้ตัวเลขสุ่ม เขียนด้วยโปรแกรมภาษา S-plus 2000 ทำการทดลองจนกว่าระยะทางยุคลิดเฉลี่ยลู่เข้าสู่ค่าคงที่จึงจะหยุดทำการทดลอง ซึ่งหมายความว่า เมื่อระยะทางยุคลิดของจำนวนการทดลองในรอบนั้นมีค่าแตกต่างจากระยะทางยุคลิดเฉลี่ยของจำนวนการทดลองรอบก่อนหน้านั้นน้อยกว่าหรือเท่ากับ 0.001 จึงหยุดทำการทดลอง ($\left|\overline{E}u_N - \overline{E}u_{N-1}\right| \leq 0.001$ โดยที่ N คือจำนวนการทดลองที่ทำให้ค่าประมาณองค์ประกอบความแปรปรวนแบบจุดมีค่าเป็นบวก)

1.6 คำจำกัดความที่ใช้ในการวิจัย

ระยะทางยุคลิด (Euclidean Distance) หมายถึง ความคลาดเคลื่อนระหว่างค่าประมาณของเวกเตอร์องค์ประกอบความแปรปรวนที่ศึกษากับค่าจริงของเวกเตอร์องค์ประกอบความแปรปรวน

1.7 ประโยชน์ที่คาดว่าจะได้รับ

1.7.1 สามารถใช้วิธีการเฉลี่ยตัวแบบประมาณองค์ประกอบความแปรปรวนสำหรับแผนแบบการทดลองจตุลัสละตินได้

1.7.2 สามารถเปรียบเทียบวิธีการประมาณองค์ประกอบความแปรปรวนโดยใช้ตัวแบบเต็มรูปและวิธีการเฉลี่ยตัวแบบได้ว่าแต่ละกรณีวิธีการใดให้ค่าประมาณที่ดีกว่าในเชิงสถิติ

1.7.3 เพื่อใช้เป็นแนวทางในการศึกษาการประมาณค่าองค์ประกอบความแปรปรวนสำหรับแผนแบบการทดลองอื่นต่อไป