หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ในการศึกษาองค์ประกอบความแปรปรวน(Variance components) มีความจำเป็นและมีความสำคัญเป็นอย่างยิ่งกรณีเป็นตัวแบบเชิงสุ่ม(Random model) ซึ่งเกิดขึ้นในกรณีที่ระดับของปัจจัย(Treatment) ในการทดลองถูกสุ่มมาจากระดับของปัจจัยทั้งหมดที่ทำการศึกษา และตัวแบบผสม(Mixed model) นั่นคือกรณีที่บางปัจจัย(Factor) มีการสุ่มมาบางระดับของปัจจัยที่ใช้ในการทดลอง บางปัจจัยนำทุกระดับของปัจจัยมาศึกษา

จากการศึกษาของ Graybill และ Hultquist(1961) ทำให้ทราบคุณสมบัติของตัวประมาณองค์ประกอบความแปรปรวนในรูปกำลังสองที่ไม่เอนเอียง และ มีความแปรปรวนต่ำที่สุดในบรรดาตัวประมาณที่ไม่เอนเอียง(Minimum variance quadratic unbiased) ซึ่งจากการศึกษาและสรุปผลของ Searle(1971) ทำให้ทราบว่าตัวประมาณที่มีคุณสมบัติดังกล่าวคือตัวประมาณจากการวิเคราะห์ความแปรปรวน(ANOVA) แต่ตัวประมาณดังกล่าวยังมีข้อด้อยที่สามารถให้ค่าตัวประมาณองค์ประกอบความแปรปรวนติดลบได้ ซึ่งทำให้ไม่สามารถแปลความหมายของตัวประมาณที่ติดลบเมื่อค่าพารามิเตอร์(Parameter) มีค่ามากกว่าหรือเท่ากับ 0 ซึ่งได้มีการเสนอแนวทางการแก้ไขสำหรับตัวประมาณ ที่มีค่าเป็นลบไว้หลายวิธีแต่วิธีที่นิยมนำมาใช้กันก็คือการใช้ค่า 0 แทนค่าลบ ซึ่งจะทำให้คุณสมบัติของตัวประมาณบางอย่างหายไป โดยเฉพาะคุณสมบัติความไม่เอนเอียง (Unbiasedness) ซึ่งตัวประมาณที่มีลักษณะสอดคล้องกับแนวคิดดังกล่าวซึ่งมีการรวบรวมวิธีพื้นฐานในการสร้างตัวประมาณที่แตกต่างกันไปดังนี้

Herbach(1959) ได้เสนอวิธีการประมาณองค์ประกอบความแปรปรวนแบบความควรจะเป็นสูงสุด ซึ่งเป็นการทำให้ฟังก์ชันความควรจะเป็นมีค่าสูงสุด(Maximum likelihood : ML) บนปริภูมิของพารามิเตอร์(Parameter space) ซึ่งมีค่าไม่เป็นลบ ทำให้ตัวประมาณที่ได้มีค่าไม่เป็นลบโดยพบว่าตัวประมาณความควรจะเป็นสูงสุด จะไม่คำนึงถึงผลของการประมาณผลกระทบคงที่ที่มีต่อขั้นแห่งความเป็นอิสระ(Degree of freedom) และยังพบอีกว่าตัวประมาณมีความยุ่งยากในการคำนวณเพราะค่าประมาณที่ได้มาจากการแก้สมการที่ไม่ใช่เชิงเส้น(Nonlinear equation) โดย

เฉพาะอย่างยิ่งกรณีข้อมูลไม่สมดุล(Unbalanced data) ซึ่งทำให้ไม่สามารถหาตัวประมาณที่มีรูปแบบที่ชัดเจนได้ ดังนั้นจึงมีการพยายามปรับปรุงตัวประมาณตัวใหม่ขึ้นมา

จากการศึกษาของ Thompson(1962) ผู้ใช้วิธีแบบความควรจะเป็นสูงสุดแบบมีข้อจำกัด (Restricted maximum likelihood : REML) โดยเป็นการหาค่าประมาณองค์ประกอบความแปรปรวนจากส่วนหนึ่งของฟังก์ชันความควรจะเป็นซึ่งเป็นส่วนของผลกระทบไม่คงที่เพื่อนำไปใช้ในการหาตัวประมาณ พบว่าภายใต้เงื่อนไขที่ทำให้ตัวประมาณองค์ประกอบความแปรปรวนมีค่ามากกว่าหรือเท่ากับ 0 ตัวประมาณองค์ประกอบความแปรปรวนทั้งที่มาจากปัจจัยที่ศึกษา($\hat{\sigma}_t^2$) รวมถึงองค์ประกอบความแปรปรวนที่มาจากความคลาดเคลื่อน($\hat{\sigma}_e^2$) ได้ค่าตัวประมาณเช่นเดียวกับวิธีวิเคราะห์ความแปรปรวน(ANOVA)

ในการศึกษาของ Rao(1971) ได้พบว่าตัวประมาณภายใต้เงื่อนไขว่าค่าสังเกตจะต้องมีการแจกแจงปกติ เพื่อหาตัวประมาณในรูปกำลังสองที่ไม่เอนเอียงและมีความแปรปรวนต่ำสุด (Mimimum quadratic unbiased estimator : MIVQUE)ในรูปทั่วไป ซึ่งจากการศึกษาของ Lamotte(1973) ทำให้ทราบว่ากรณีข้อมูลเป็นแบบสมดุล(Balanced data) ตัวประมาณองค์ประกอบความแปรปรวนจะมีรูปแบบที่แน่นอนและคือตัวประมาณวิธีวิเคราะห์ความแปรปรวน

ต่อมา Kelly และ Mathew(1993) ได้พยายามพัฒนาตัวประมาณองค์ประกอบความแปรปรวนที่ให้ค่าความน่าจะเป็นที่ตัวประมาณเป็นลบมีค่าน้อย และเป็นตัวประมาณที่ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณต่ำกว่าตัวประมาณ ด้วยวิธีวิเคราะห์ความแปรปรวน(ANOVA) ซึ่งเป็นการศึกษาภายใต้การแปลงค่าสังเกต $\underset{\sim}{u} = \mathbf{Z}'\underset{\sim}{y}$ ทำให้ได้ตัวประมาณกำลังสองไม่แปรเปลี่ยน(Invariant quadratic estimator : IQE) ซึ่งมีการกำหนดให้ตัวประมาณมีค่าเป็น 0 เมื่อตัวประมาณที่ได้มีค่าติดลบ ซึ่งอาจเรียกว่าตัวประมาณกำสังสองที่ไม่แปรเปลี่ยนที่ไม่เป็นลบ(nonnegative IQE)

จากการศึกษาเพื่อหาตัวประมาณองค์ประกอบความแปรปรวนที่ผ่านมาพบว่าตัวประมาณได้แต่ละวิธีมีข้อดีและมีความเหมาะสม สำหรับประมาณค่าองค์ประกอบความแปรปรวนที่แตกต่างกัน ดังนั้นถ้าเราสามารถนำวิธีการประมาณต่างๆที่มีมาทำการเฉลี่ยถ่วงน้ำหนัก โดยให้น้ำหนักของความสำคัญของแต่ละวิธีที่แตกต่างกัน ภายใต้ข้อกำหนดว่าผลรวมของตัวถ่วงน้ำหนักของตัวประมาณที่ใช้ในการเฉลี่ยมีค่าเป็น 1 ดังนั้นถ้าสามารถหาตัวถ่วงน้ำหนักภายใต้หลักการที่เหมาะสม จะทำให้ได้ตัวประมาณตัวใหม่ซึ่งแม้อาจไม่ใช่ตัวประมาณที่ดีที่สุดเฉพาะกรณีแต่จะเป็นตัว

ประมาณที่สามารถประมาณองค์ประกอบความแปรปรวนได้ค่อนข้างดี สำหรับข้อมูลการทดลองที่ใช้กันทั่วไป

1.2 วัตถุประสงค์ของการวิจัย

การวิจัยในครั้งนี้มีการแบ่งวัตถุประสงค์ของการวิจัยเป็น 2 ข้อดังนี้

1.2.1 เพื่อหาวิธีการเฉลี่ยถ่วงน้ำหนักของตัวประมาณองค์ประกอบความแปรปรวนที่มีอยู่ได้อย่างเหมาะสมสำหรับข้อมูลเป็นแบบสมดุล(Balanced data) ของแผนการทดลองสุ่มสมบูรณ์ (Completely randomized design : CRD) โดยได้รับอิทธิพลเชิงสุ่ม(Random-effect)

1.2.2 ศึกษาและเปรียบเทียบตัวประมาณองค์ประกอบความแปรปรวน ของแผนการทดลองข้างต้นซึ่งมีวิธีการประมาณดังต่อไปนี้

1.2.2.1 ตัวประมาณแบบความควรจะเป็นสูงสุด(Maximum Likelihood: ML)

1.2.2.2 ตัวประมาณแบบความควรจะเป็นสูงสุดแบบมีข้อจำกัด(Restricted Maximum Likelihood : REML)

1.2.2.3 ตัวประมาณแบบกำลังสองไม่แปรเปลี่ยน(Invariance Quadratic Estimator: IQE)

1.2.2.4 ตัวประมาณภายใต้การเฉลี่ยถ่วงน้ำหนักโดยวิธีการให้ค่าเฉลี่ยถ่วงน้ำหนักที่แตกต่างกัน

1.3 สมมติฐานการวิจัย

มีตัวประมาณองค์ประกอบความแปรปรวนตัวใหม่ ภายใต้การเฉลี่ยถ่วงน้ำหนักที่สามารถประมาณค่าองค์ประกอบความแปรปรวนได้ค่อนข้างดี สำหรับข้อมูลการทดลอง จากแผนการทดลองสุ่มสมบูรณ์(CRD)

1.4 ข้อตกลงเบื้องต้น

สำหรับการวิจัยในครั้งนี้เป็นการศึกษาเฉพาะแผนการทดลองสุ่มสมบูรณ์(CRD) ซึ่งได้รับอิทธิพลเชิงสุ่ม(Random-effect) เฉพาะกรณีที่จำนวนค่าสังเกตในแต่ละระดับของปัจจัยเท่ากัน โดยมีตัวแบบดังนี้

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad (1.1)$$

เมื่อ μ คือ ค่าเฉลี่ยรวมของประชากร

τ_i คือ ผลกระทบของปัจจัยทดลองที่ i

ε_{ij} คือ ความคลาดเคลื่อนของค่าสังเกตที่ j ปัจจัยทดลองที่ i

เมื่อ i =1 ,2 , ... ,a j = 1 , 2 , . . . ,n

$$Cov(y_{ij}, y_{i'j'}) = \begin{cases} \sigma_\tau^2 + \sigma_e^2 & i = i' \text{ และ } j = j' \\ \sigma_\tau^2 & i = i' \text{ และ } j \neq j' \\ 0 & i \neq i' \end{cases}$$

โดย $\tau_i \sim N(0, \sigma_\tau^2)$

และ $\varepsilon_{ij} \sim N(0, \sigma_e^2)$

σ_τ^2 และ σ_e^2 คือ พารามิเตอร์องค์ประกอบของความแปรปรวน(Variance component parameter) ซึ่งต้องทำการประมาณต่อไป

จากตัวแบบดังสมการ (1.1) สามารถนำมาสร้างตารางวิเคราะห์ความแปรปรวน(ANOVA table) ได้ดังตารางที่ 1

ตารางที่ 1 แสดงการวิเคราะห์ความแปรปรวนสำหรับแผนการทดลองสุ่มสมบูรณ์เชิงสุ่ม เมื่อจำนวนค่าสังเกตที่ใช้ในแต่ละระดับของปัจจัยมีจำนวนเท่ากัน

สาเหตุของความแปรปรวน	ระดับชั้นของความอิสระ	ผลรวมกำลังสอง	ผลรวมกำลังสองเฉลี่ย	ค่าคาดหวังของผลรวมกำลังสองเฉลี่ย
ปัจจัยทดลอง	a-1	$SSTr = n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2$	$\frac{SSTr}{a-1}$	$n\sigma_\tau^2 + \sigma_e^2$
ความคลาดเคลื่อน	a(n-1)	$SSE = \sum_{i=1}^{a}\sum_{i=1}^{n}(y_{ij} - \bar{y}_{i.})^2$	$\frac{SSE}{a(n-1)}$	σ_e^2
รวม	an-1	$SST = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$		

โดยที่ y_{ij} คือค่าสังเกตที่ j ที่ได้รับปัจจัยในระดับที่ i

$\bar{y}_{i.}$ คือค่าเฉลี่ยของค่าสังเกตที่ได้รับปัจจัยระดับที่ i

โดยที่ $\bar{y}_{i.} = \frac{1}{n}\sum_{j=1}^{n} y_{ij}$

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

$\bar{y}_{..}$ คือค่าเฉลี่ยของค่าสังเกตทุกตัวในทุกระดับของปัจจัย

โดยที่ $\bar{y}_{..} = \frac{1}{an}\sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}$

a จำนวนระดับของปัจจัยที่ใช้ในการทดลอง

n จำนวนค่าสังเกตในแต่ละระดับของปัจจัย

1.5 ขอบเขตของการวิจัย

1.5.1 การกำหนดกรณีศึกษา

- จำนวนระดับของปัจจัยที่ใช้ในการทดลอง(a) 2 , 4 , 6 , 8 , 10,12
- จำนวนค่าสังเกตในแต่ละระดับของปัจจัย(n) 2 , 4 , 6

1.5.2 การกำหนดค่าพารามิเตอร์

- ค่าเฉลี่ยรวม(μ) 40
- ค่าองค์ประกอบความแปรปรวน ซึ่งในที่นี้กำหนดให้ มีความสัมพันธ์ในรูป

$$\frac{\sigma_\tau^2}{\sigma_e^2} = k$$

โดย กำหนดให้ k เป็น 0.1 , 0.5 , 1 , 4 , 9

และ กำหนด σ_e^2 เป็น 8

นั่นคือ $\sigma_\tau^2 = k \times \sigma_e^2$

1.5.3 การวิจัยในครั้งนี้สร้างแบบจำลองข้อมูลกรณีต่างๆ ตามที่ระบุไว้ในข้อ 1.5.1 และ 1.5.2 รวมทั้งสิ้น 6X3X5 = 90 กรณี โดยใช้เทคนิคมอนติคาร์โล(Monte carlo technique) เขียนด้วยโปรแกรมภาษา S-PLUS 2000 ซึ่งแต่ละกรณีศึกษามีการจำลองข้อมูลซ้ำ 10,000 รอบ โดยวิธีการตัดสินใจในการทำซ้ำได้กล่าวรายละเอียดไว้ในภาคผนวก ก

1.6 คำจำกัดความที่ใช้ในงานวิจัย

ข้อมูลสมดุล(Balanced data) หมายถึง ข้อมูลที่มีจำนวนค่าสังเกตในแต่ละระดับของปัจจัยเท่ากัน

แผนการทดลองแบบสุ่มสมบูรณ์(Completely randomized design : CRD) หมายถึง แผนการทดลองที่หน่วยทดลองเป็นหน่วยที่ได้มาจากการสุ่ม

1.7 ประโยชน์ที่คาดว่าจะได้รับ

1.7.1 สามารถหาวิธีถ่วงน้ำหนักตัวประมาณองค์ประกอบความแปรปรวน เพื่อให้ได้ตัวประมาณองค์ประกอบความแปรปรวนที่เหมาะสมตัวใหม่

1.7.2 สามารถเปรียบเทียบวิธีการประมาณองค์ประกอบความแปรปรวนของตัวประมาณด้วยวิธีพื้นฐาน (ML, REML และ IQE) รวมทั้งตัวประมาณภายใต้การเฉลี่ยถ่วงน้ำหนัก

1.7.3 สามารถใช้เป็นแนวทางในการพัฒนาหาตัวประมาณองค์ประกอบความแปรปรวนในแผนการทดลองอื่นต่อไป