บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในการศึกษาการวางแผนการทดลองแบบสุ่มในบล็อคสมบูรณ์(Randomized Complete Block Design) เรียกโดยย่อว่า RCBD ใช้กับงานทดลองที่พบว่า นอกจากวิธีการทดลองแล้วยังมีปัจจัยที่ทราบสาเหตุอื่นอีก 1 ปัจจัยที่จะส่งผลกระทบต่อการทดลอง ซึ่งจะมีผลทำให้การทดลองไม่สม่ำเสมอกัน ปัจจัยเหล่านี้ ได้แก่ หน่วยทดลองมีความแตกต่าง วิธีปฏิบัติระหว่างการทดลองมีความแตกต่าง การวางแผนการทดลองแบบสุ่มในบล็อคสมบูรณ์ ยังแบ่งออกเป็นอีก 2 ลักษณะคือ 1) การวางแผนการทดลองแบบมีซ้ำ 2) การวางแผนการทดลองแบบไม่มีซ้ำ การวางแผนการทดลองแบบสุ่มในบล็อคสมบูรณ์ แบบไม่มีซ้ำเมื่อมีค่าสังเกตสูญหายอิทธิพลระหว่างบล็อคกับวิธีการทดลองจะแยกไม่ออก ทำให้ความน่าเชื่อถือของแผนแบบลดลง และไม่สามารถวิเคราะห์ข้อมูลตามวิธีการของการวางแผนการทดลองแบบสุ่มในบล็อคสมบูรณ์ ตามปกติได้ ส่วนกรณีมีซ้ำหากมีการทำซ้ำ 1 ครั้งหรือมากกว่า เมื่อมีค่าสังเกตสูญหายเราสามารถตัดค่าสังเกตบางค่าทิ้งไปได้ โดยยังสามารถแยกอิทธิพลระหว่างบล็อคกับวิธีการทดลองได้ สำหรับในการวิจัยครั้งนี้ ผู้วิจัยสนใจศึกษาการประมาณค่าสังเกตสูญหายในการวางแผนการทดลองแบบสุ่มในบล็อคสมบูรณ์ แบบไม่มีซ้ำ

การวิเคราะห์ข้อมูลเมื่อมีค่าสังเกตสูญหายทำได้หลายวิธี แต่ที่ใช้กันอยู่มี 2 วิธีคือ

1. หาผลวิเคราะห์ข้อมูลโดยประมาณ ซึ่งวิธีนี้จะต้องประมาณค่าสังเกตที่สูญหายขึ้นมาจากนั้นวิเคราะห์ความแปรปรวนไปตามขั้นตอนปกติ

2. วิเคราะห์ข้อมูลที่แท้จริงแม้ว่าข้อมูลจะไม่ใช่เชิงตั้งฉาก

วิธีวิเคราะห์ข้อมูลโดยการประมาณค่าสังเกตสูญหายสามารถทำได้หลายวิธีแต่ในที่นี้จะศึกษาเปรียบเทียบวิธีการประมาณเพียง 3 วิธี คือ

1. วิธีประมาณค่าสูญหายโดยวิธีกำลังสองน้อยสุด (least square method)
2. วิธีประมาณค่าสูญหายโดยวิธี EM algorithm (Expectation Maximization)
3. วิธีประมาณค่าสูญหายโดยวิธี Imputation Method

วิธีแรกคือวิธีกำลังสองน้อยสุด วิธีนี้เป็นการประมาณค่าที่ทำให้ผลรวมกำลังสองของความคลาดเคลื่อนมีค่าน้อยที่สุด วิธีที่สองคือ วิธี EM algorithm วิธีนี้อาศัยฟังก์ชันไลลิฮูดเข้ามาช่วยในการประมาณค่าพารามิเตอร์ และหาค่าตัวสถิติเพียงพอ ซึ่งทำให้การวิเคราะห์ง่ายขึ้น และวิธีที่สามคือวิธี Imputation Method วิธีนี้เป็นการสร้างชุดข้อมูลขึ้นมาใหม่จากชุดที่มีอยู่เดิมและนำค่าความแปรปรวนมาประกอบในการพิจารณาเลือกค่าประมาณ

จากหลักการดังกล่าวข้างต้น ผู้วิจัยจึงสนใจที่จะศึกษาและเปรียบเทียบวิธีการประมาณค่าสังเกตที่สูญหายทั้ง 3 วิธี

1.2 วัตถุประสงค์ของการวิจัย

เพื่อศึกษาเปรียบเทียบวิธีการประมาณค่าสังเกตที่สูญหายโดยใช้วิธีการประมาณค่า 3 วิธีคือ

1. วิธีประมาณค่าสูญหายโดยวิธีกำลังสองน้อยสุด (least square method)
2. วิธีประมาณค่าสูญหายโดยวิธี EM algorithm (Expectation Maximization)
3. วิธีประมาณค่าสูญหายโดยวิธี Imputation Method

1.3 สมมติฐานของการวิจัย

จากการศึกษาวิจัยครั้งนี้ การประมาณค่าสูญหายในการวางแผนการทดลองด้วย วิธี Imputation Method จะให้ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน MSE(Mean Square Error) ต่ำสุดในทุกวิธีที่ทำการศึกษา

1.4 ข้อตกลงเบื้องต้น

1.4.1 ในการศึกษาครั้งนี้ศึกษาการวางแผนการทดลองแบบสุ่มในบล็อคสมบูรณ์กรณีอิทธิพลกำหนด ซึ่งมีแบบจำลองเชิงบวกของค่าสังเกตแต่ละค่าคือ

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad i = 1,2,...,t \quad j = 1,2,...,b$$

เมื่อ Y_{ij} = ค่าสังเกตจากหน่วยทดลองที่ได้รับวิธีการทดลองที่ i และอยู่ใน บล็อคที่ j โดยที่ $Y_{ij} \sim NID(\mu_{ij}, \sigma^2)$

μ = ค่าเฉลี่ยของประชากร

τ_i = อิทธิพลของวิธีการทดลองที่ i

β_j = อิทธิพลของบล็อคที่ j

ε_{ij} = ความคลาดเคลื่อนของค่าสังเกตจากหน่วยทดลองที่อยู่ในบล็อคที่ j และได้รับวิธีการทดลองที่ i โดยที่ $\varepsilon_{ij} \sim NID(0, \sigma^2)$

1.5 ขอบเขตของการวิจัย

1.5.1 ศึกษาเปรียบเทียบวิธีประมาณค่าสังเกตสูญหาย โดยวิธีกำลังสองน้อยสุด วิธี EM algorithm และวิธี Imputation Method

1.5.2 กำหนดจำนวนวิธีการทดลองมีขนาด 3, 4 และ 5 วิธีการทดลอง จำนวนบล็อคในการทดลองมีขนาด 2, 4 และ 6 บล็อค

1.5.3 กำหนดให้ข้อมูลมีการแจกแจงของความคลาดเคลื่อนเป็นแบบปกติ (normal distribution) โดยมีฟังก์ชันความหนาแน่นอยู่ในรูปของ

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\tfrac{1}{2\sigma^2}(x-\mu)^2\} \quad , \quad -\infty < x < \infty$$

ค่าคาดหวัง $E(X) = \mu$

ความแปรปรวน $V(X) = \sigma^2$

1.5.4 กำหนดให้จำนวนข้อมูลหาย 10%, 20% และ 30% อย่างสุ่ม

1.5.5 กำหนดให้จำนวนข้อมูลสูญหายโดยฟังก์ชันการแจกแจงแบบยูนิฟอร์มในโปรแกรม S-plus 2000

1.5.6 กำหนดให้ข้อมูลมีสัมประสิทธิ์ความแปรผัน (Coefficient of variation: C.V.(y_{ij})) ในระดับต่าง ๆ กัน คือ 5%, 25% และ 45% และค่าเฉลี่ย (μ) เท่ากับ 40

จาก

$$c.v.(y_{ij}) = \frac{S.D.(y_{ij})}{\mu} = \frac{\sqrt{\sigma_\tau^2 + \sigma_\beta^2 + \sigma_\varepsilon^2}}{\mu}$$

กำหนดให้

$\sigma_\tau^2 = \sigma_\beta^2 = h\sigma_\varepsilon^{2^*}$ โดยที่ h เป็นค่าจำนวนเต็มคงที่เท่ากับ 1, 2 และ 3

นั่นคือ

$$c.v.(y_{ij}) = \frac{\sqrt{h\sigma_\varepsilon^2 + h\sigma_\varepsilon^2 + \sigma_\varepsilon^2}}{\mu} = \frac{\sigma_\varepsilon\sqrt{2h+1}}{\mu}$$

ดังนั้นจะได้ว่า

$$\sigma_\varepsilon^2 = \left(\frac{c.v.(y_{ij})\mu}{\sqrt{2h+1}}\right)^2$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

โดยมีความแปรปรวน $\sigma^2_{\gamma_{ij}}$ เท่ากับ 4, 100 และ 324 ตามลำดับ

1.5.7 ในการวิจัยครั้งนี้สร้างแบบจำลองข้อมูลโดยใช้เทคนิคมอนติคาร์โลซีมูเลชั่น (Monte Carlo Simulation Technique) เขียนด้วยโปรแกรม S-plus 2000 จำลองในแต่ละการทดลอง 200 ครั้ง

1.6 เกณฑ์การตัดสินใจ

เกณฑ์การพิจารณาว่า วิธีการประมาณค่าสูญหายใด จะให้ค่าประมาณที่ดีจะพิจารณาโดยเปรียบเทียบค่าจริงที่สูญหายและค่าที่ประมาณขึ้นใหม่ ด้วยค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน MSE (Mean Square Error) วิธีใดให้ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนต่ำกว่าเป็นวิธีที่ดีกว่า

1.7 ประโยชน์ที่คาดว่าจะได้รับ

ผลจากการวิจัย จะเป็นแนวทางในการเลือกวิธีการประมาณค่าสูญหายในการวางแผนการทดลองแบบ สุ่มในบล็อคสมบูรณ์ เพื่อให้ได้ผลสรุปจากการวางแผนการทดลองมีความน่าเชื่อถือมากที่สุด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

---

*การกำหนดให้ $\sigma^2_\tau = \sigma^2_\beta = h\sigma^2_\varepsilon$ เพื่อสะดวกต่อการวิจัย และทำให้การจำลองแบบข้อมูลที่นำมาใช้ในการศึกษามีหลักเกณฑ์มากขึ้น