บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การวิเคราะห์ความแปรปรวนร่วม (analysis of covariance) เป็นวิธีการหนึ่งที่ถูกนำมาใช้ในการวิเคราะห์ความแปรปรวนและการทดสอบสมมุติฐานของความแตกต่างระหว่างอิทธิพลของวิธีปฏิบัติ (treatment) ซึ่งส่วนใหญ่จะพบในการวิเคราะห์ข้อมูลทางด้านชีววิทยา ด้านการแพทย์ และด้านการเกษตร โดยนำหลักการของการวิเคราะห์ความแปรปรวนผสมผสานกับการวิเคราะห์ความถดถอย เพื่อลดความแปรปรวนให้กับหน่วยทดลอง (experimental unit) วิธีการก็คือทำการวัดค่าตัวแปรร่วม x (covariate หรือ concomitant variables) แล้วนำมาปรับค่าตัวแปรตาม y (dependent variable) ซึ่งเป็นค่าสังเกตที่ได้จากการให้วิธีปฏิบัติ เพื่อหาข้อสรุปเกี่ยวกับอิทธิพลของวิธีปฏิบัติ และผลการวิเคราะห์จะให้ข้อสรุปที่ถูกต้องแม่นยำก็ต่อเมื่อข้อมูลที่นำมาวิเคราะห์มีคุณสมบัติตามข้อตกลงเบื้องต้นของการวิเคราะห์ความแปรปรวนร่วมเท่านั้น ซึ่งข้อตกลงดังกล่าวได้แก่

1. วิธีสุ่ม(Randomization) หมายถึงการเลือกหน่วยทดลองจากประชากรต้องเป็นไปโดยสุ่ม และการเลือกหน่วยทดลองให้กับวิธีปฏิบัติก็ต้องเป็นไปโดยสุ่มและอิสระซึ่งกันและกัน

2. Homogeneity of Within-Group Regression Slopes หมายถึง สัมประสิทธิ์การถดถอยในกลุ่มประชากรต้องเป็นเอกพันธ์* สำหรับวิธีปฏิบัติ t ระดับ

3. Statistical Independence of Covariate and Treatment หมายถึงตัวแปรร่วม x เป็นตัวแปรอิสระ และถือว่าไม่ได้รับอิทธิพลของวิธีปฏิบัติ

---

* เอกพันธ์ หมายถึง การทดสอบสัมประสิทธิ์การถดถอยสำหรับวิธีปฏิบัติ t ระดับแล้วต้องไม่มีความแตกต่างกัน หรือแตกต่างกันอย่างไม่มีนัยสำคัญ

4. Fixed Covariate Value Measured without Error หมายถึงค่าตัวแปรร่วมจะเป็นค่าที่กำหนด และการวัดไม่มีความคลาดเคลื่อน

5. Linearity of Within-Group Regression การถดถอยของตัวแปรตาม y ที่ขึ้นกับตัวแปรร่วม x ต้องเป็นเชิงเส้น

6. Normality and Homogeneity of Variance of Condition y Scores ค่าความคลาดเคลื่อนของการทดลองเกิดขึ้นโดยสุ่มและเป็นอิสระซึ่งกันและกัน มีการแจกแจงแบบปกติโดยมีความแปรปรวนเท่ากัน

7. ตัวแบบหุ่นแบบบวก (Additive Model) หมายถึงอิทธิพลของวิธีปฏิบัติและสิ่งแวดล้อมอื่น ๆ เป็นแบบบวก

ในทางปฏิบัติผู้วิจัยมักประสบปัญหาที่ค่าสังเกตบางค่ามีค่าสูงหรือต่ำมากอย่างผิดปกติไปจากข้อมูลส่วนใหญ่ ซึ่งจะเรียกว่า **ข้อมูลผิดปกติ** (outlier) อาจเนื่องมาจากความคลาดเคลื่อนจากการวัด (measurement error) ได้แก่การบันทึกข้อมูล การใช้เครื่องมือเครื่องใช้ที่มีคุณภาพต่ำ ความไม่สม่ำเสมอเกี่ยวกับวิธีการในการทดลอง ความคลาดเคลื่อนอันเนื่องมาจากการปฏิบัติการ (execute error)ได้แก่ การลงรหัส การเจาะบัตร เป็นต้น สิ่งเหล่านี้จะเป็นสาเหตุทำให้การแจกแจงของความคลาดเคลื่อนเปลี่ยนแปลงไปจากการแจกแจงแบบปกติ ซึ่งจะส่งผลกระทบให้ตัวแปรตาม y มีการแจกแจงเปลี่ยนไปด้วย ดังนั้นหากผู้วิจัยยอมฝ่าฝืนข้อตกลงเบื้องต้น และทำการวิเคราะห์ความแปรปรวนร่วมต่อไป จะส่งผลให้ประสิทธิภาพในการวิเคราะห์ลดลง

โดยทั่วไปแล้วในการวิเคราะห์ความแปรปรวนร่วม จะใช้การประมาณพารามิเตอร์แบบวิธีกำลังสองน้อยที่สุด (Least Square method) ซึ่งจะได้ตัวประมาณที่ไม่เอนเอียงที่มีความแปรปรวนต่ำสุดสม่ำเสมอ (Uniformly Minimum Variance Unbiased Estimator) ในบรรดาตัวประมาณที่ไม่เอนเอียงเมื่อข้อมูลเป็นไปตามข้อตกลงเบื้องต้น แต่ถ้าข้อมูลขาดคุณสมบัติของข้อตกลงเบื้องต้น เช่น ข้อมูลมีการแจกแจงแบบหางยาวกว่าปกติ วิธีนี้อาจจะไม่เหมาะสมจึงได้มีผู้คิดวิธีประมาณพารามิเตอร์ขึ้นมาเรียกว่าวิธีตัวประมาณเอ็ม (M-Estimator method) ซึ่งมีการนำเสนอเป็นครั้งแรก โดย Huber เมื่อปี ค.ศ. 1964 วิธีนี้จะให้ตัวประมาณที่มีความแกร่งด้วยเหตุนี้เองผู้วิจัยจึงทำการเปรียบเทียบวิธีการประมาณพารามิเตอร์ทั้ง 2 วิธีนี้ว่าวิธีการใดจะส่งผลให้การวิเคราะห์ความแปรปรวนร่วมมีประสิทธิภาพมากกว่ากัน

1.2 วัตถุประสงค์ของการวิจัย

เพื่อเปรียบเทียบอำนาจการทดสอบในการวิเคราะห์ความแปรปรวนร่วม เมื่อความคลาดเคลื่อนมีการแจกแจงแบบหางยาวกว่าปกติ ด้วยการพิจารณาวิธีการประมาณพารามิเตอร์โดยใช้วิธีการประมาณแบบกำลังสองน้อยที่สุด (OLS) กับวิธีตัวประมาณเอ็มที่ใช้เกณฑ์ความแกร่งของ Ramsay และตัวประมาณสเกลความคลาดเคลื่อนแบบ MAD(median absolute - deviation)

1.3 สมมุติฐานของการวิจัย

ในการวิเคราะห์ความแปรปรวนร่วมเมื่อความคลาดเคลื่อนมีการแจกแจงแบบหางยาว-กว่าปกติการประมาณพารามิเตอร์ด้วยวิธีตัวประมาณเอ็มจะให้อำนาจการทดสอบสูงกว่าวิธีกำลังสอง-น้อยที่สุด

1.4 ข้อตกลงเบื้องต้น

การวิจัยครั้งนี้ถือว่าความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดประเภทที่ 1 และอำนาจการทดสอบเป็นดัชนีสำคัญในการใช้เป็นเกณฑ์วินิจฉัยวิธีการประมาณพารามิเตอร์ของการวิเคราะห์ความแปรปรวนร่วม

1.5 ขอบเขตการวิจัย

1.5.1 ตัวแบบ (model) ที่ใช้เป็นแบบอิทธิพลกำหนด(fixed effect model) ในแผนแบบการทดลองแบบสุ่มตลอด (completely randomized design) เนื่องจากการวิจัยครั้งนี้ผู้วิจัยไม่ได้คำนึงที่จะสรุปผลไปถึงวิธีปฏิบัติอื่นๆ และไม่ต้องการให้เกิดความแปรปรวนอันเนื่องมาจากการสุ่มวิธีปฏิบัติ

$$Y_{ij} = \mu + \tau_j + \sum_{k=1}^{t}\beta_k(x_{ijk} - \bar{x}_{..k}) + \varepsilon_{ij} \quad ; \quad \begin{matrix} i=1,2,\ldots,n_j \\ j=1,2,\ldots,p \end{matrix}$$

โดยที่ p หมายถึง จำนวนวิธีปฏิบัติ

q หมายถึง จำนวนตัวแปรร่วม

n_j หมายถึง ขนาดตัวอย่างในแต่ละวิธีปฏิบัติที่ j

μ หมายถึง ค่าเฉลี่ยทั้งหมดของ y (grand population mean of all y scores)

τ_j หมายถึง อิทธิพลของวิธีปฏิบัติที่ j (effect of the j^{th} treatment)

β_k หมายถึง ค่าสัมประสิทธิการถดถอยของตัวแปรร่วมที่ k (partial regression coefficient associated with covariate k^{th})

$\bar{x}_{..k}$ หมายถึง ค่าเฉลี่ยของตัวแปรร่วมที่ k (mean of covariate k^{th})

ε_{ij} หมายถึง ค่าความคลาดเคลื่อนของค่าสังเกตจากค่าเฉลี่ยที่ปรับแล้ว (deviation of observation from adjust mean of group)

1.5.2 กำหนดกลุ่มตัวอย่าง หรือ จำนวนวิธีปฏิบัติ = 3,5,7

1.5.3 กำหนดจำนวนตัวแปรร่วม x = 2,5 โดยที่แต่ละตัวมีค่าเป็นช่วงไม่เท่ากัน เพื่อให้ได้ตัวแปรที่คล้ายธรรมชาติ จึงกำหนดให้ $x \sim N(\mu, \sigma^2)$

1.5.4 กำหนดขนาดตัวอย่างโดยใช้แต่ละกลุ่มเท่ากัน ดังแสดงในตารางที่ 1.1

ตารางที่ 1.1 แสดงขนาดตัวอย่างที่ใช้ในการวิจัย

จำนวนวิธีปฏิบัติ	จำนวนตัวแปรร่วม	ขนาดตัวอย่างแต่ละกลุ่ม
3	2	5 15
	5	10 20
5	2	5 15
	5	10 20
7	2	5 15
	5	10 20

1.5.5 กำหนดให้ค่าเฉลี่ยของแต่ละกลุ่มเท่ากันคือ $\mu = 100$

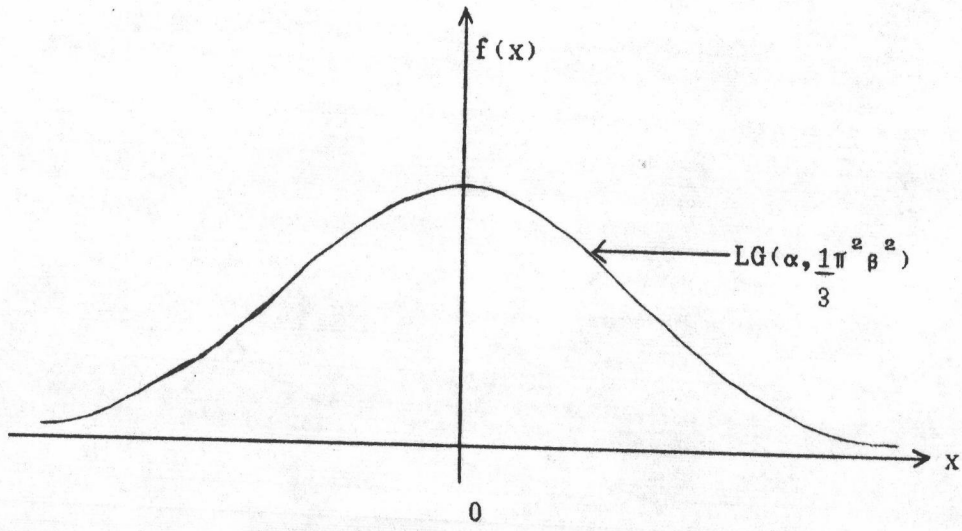
1.5.6 ลักษณะการแจกแจงของความคลาดเคลื่อนที่นำมาทดสอบ โดยมีค่าเฉลี่ย $\mu = 0$ และความแปรปรวน $= \sigma^2$ ทุกรูปแบบการศึกษาโดยมีการแจกแจงดังนี้

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ก. การแจกแจงแบบโลจิสติค (Logistic distribution)

$$f(x) = \begin{cases} \frac{1}{\beta} \cdot \frac{\exp\{-(x-\alpha)/\beta\}}{\{1+\exp(x-\alpha)/\beta\}^2} & , \ \beta > 0 \\ 0 & , \ \text{อื่น ๆ} \end{cases}$$

ผู้วิจัยสนใจศึกษาเมื่อ $\alpha = 0$ $\beta = \sqrt{3} \times \sigma / \pi$ ซึ่งสามารถเขียนรูปได้ดังนี้



รูปที่ 1.1 แสดงเส้นโค้งของการแจกแจงแบบโลจิสติค

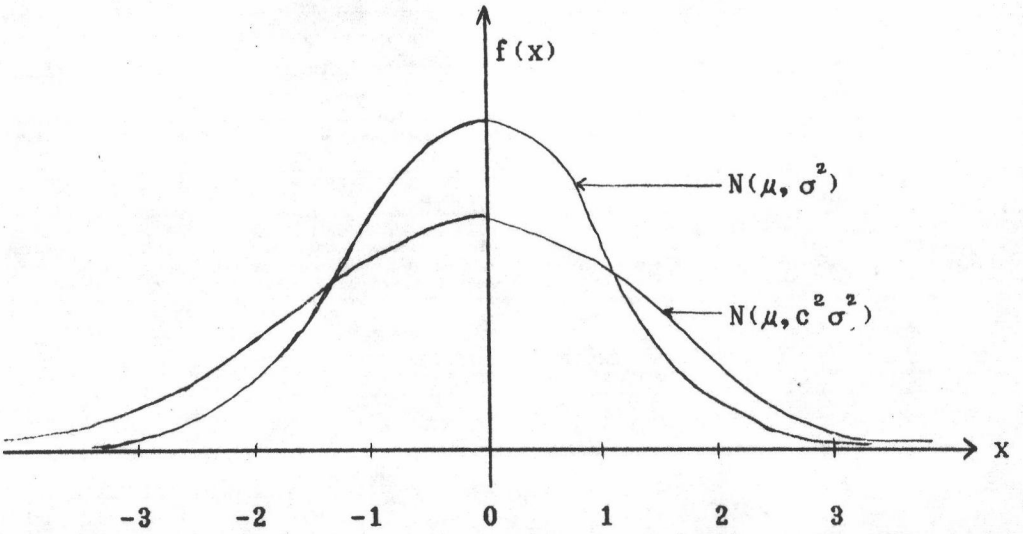
ค่าคาดหวัง $E(x) = \alpha$

ค่าความแปรปรวน $Var(x) = \frac{1}{3}\pi^2\beta^2$

ข. การแจกแจงแบบปกติปลอมปน (Scale-contaminated normal distribution)

ฟังก์ชันความน่าจะเป็นคือ

$$F = (1-p)N(\mu,\sigma^2) + pN(\mu,c^2\sigma^2) \quad ; \quad c>0$$



รูปที่ 1.2 แสดงเส้นโค้งของการแจกแจงแบบปกติปลอมปน

หมายความว่าตัวแปรสุ่มมาจากการแจกแจงแบบ $N(\mu,\sigma^2)$ ด้วยความน่าจะเป็น 1-p และมาจากการแจกแจงแบบ $N(\mu,c^2\sigma^2)$ ด้วยความน่าจะเป็น p

เมื่อ μ และ σ^2 เป็นพารามิเตอร์ที่กำหนดค่าเฉลี่ยและความแปรปรวนของความคลาดเคลื่อน

p และ c เป็นค่าคงที่ (fixed constant) ที่กำหนดเปอร์เซ็นต์การปลอมปนและสเกลแฟคเตอร์

1.5.7 พิจารณาระดับนัยสำคัญ 2 ระดับคือ α =.01 และ .05

1.5.8 การจำลองจะกระทำซ้ำ ๆ กันไม่น้อยกว่า 1000 ครั้ง ในแต่ละสถาณการณ์ของการทดลอง

1.6 คำจำกัดความของคำต่าง ๆที่ใช้ในการวิจัย

วิธีปฏิบัติ หมายถึง ปัจจัยที่ใช้ศึกษาทดลอง

ตัวแปรตาม (Dependent variable ; y) หมายถึงค่าของหน่วยทดลองที่วัดได้หลังจากที่หน่วยทดลองได้รับอิทธิพลของวิธีปฏิบัติแล้ว

ตัวแปรร่วม (Concomitant variable ; x) หมายถึงค่าของหน่วยทดลองที่วัดได้ก่อนที่จะได้รับอิทธิพลของวิธีปฏิบัติ ซึ่งเป็นค่าของตัวแปรที่แฝงมากับหน่วยทดลอง

ความน่าจะเป็นของความผิดพลาดประเภทที่ 1 (P(Type I error)) คือความน่าจะเป็นของความผิดพลาดที่เกิดจากการปฏิเสธสมมุติฐานว่าง (Ho) เมื่อสมมุติฐานว่างเป็นจริง ใช้แทนด้วยสัญญลักษณ์ α

ความน่าจะเป็นของความผิดพลาดประเภทที่ 2 (P(Type II error)) คือความน่าจะเป็นของความผิดพลาดที่เกิดจากการยอมรับสมมุติฐานว่าง (Ho) เมื่อสมมุติฐานว่างไม่จริง ใช้แทนด้วยสัญญลักษณ์ β

อำนาจการทดสอบ (Power of the test) คือค่าความน่าจะเป็นที่จะปฏิเสธสมมุติฐานว่าง (Ho) เมื่อสมมุติฐานว่างไม่จริงซึ่งมีค่าเท่ากับ $1-\beta$

การแจกแจงแบบหางยาวกว่าปกติ (Long-tailed distribution) คือการแจกแจงซึ่งมีฟังก์ชันการแจกแจงความน่าจะเป็น $f(x)$ เข้าใกล้ 0 ในอัตราที่ช้ากว่าฟังก์ชันการแจกแจงความน่าจะเป็นแบบปกติเมื่อ x เข้าใกล้ $-\infty$ หรือ $+\infty$

อิทธิพลกำหนด หมายถึง การกำหนดวิธีปฏิบัติเพื่อเปรียบเทียบอิทธิพล โดยที่วิธีปฏิบัติมิได้มีการสุ่มขึ้นมา เพื่อให้ τ_j เป็นส่วนเบี่ยงเบนของวิธีปฏิบัติจากค่าเฉลี่ย และ $\sum_{j=1} \tau_j = 0$

1.7 ประโยชน์ของการวิจัย

เพื่อให้ผู้วิจัยสามารถเลือกใช้วิธีการประมาณพารามิเตอร์ ในการวิเคราะห์ความแปรปรวนร่วมได้อย่างมีประสิทธิภาพ เมื่อลักษณะการแจกแจงของความคลาดเคลื่อนเป็นแบบหางยาวกว่าปกติ