บทที่ 1

บทนำ

1.1 หลักการ ทฤษฎี และเหตุผล

การวางแผนการทดลอง หมายถึงกำหนดการทดลองเพื่อให้ได้ความรู้เกี่ยวกับเรื่องที่ต้องการศึกษาให้มากที่สุด และแผนการทดลองนั้นจะต้องไม่ซับซ้อน และต้องมีประสิทธิภาพที่เชื่อถือได้ ทั้งนี้เพื่อประหยัดเวลา ทุนทรัพย์ และบุคลากรตลอดจนวัตถุดิบที่จะใช้ในการทดลองด้วย ซึ่งแผนการทดลองนั้นมีหลายแบบ แต่แผนการทดลองแบบการสุ่มอย่างสมบูรณ์ (Completely Randomized Design) ซึ่งเป็นแผนการทดลองที่เราสุ่มหน่วยทดลองให้ได้รับวิธีปฏิบัติใดๆ เป็นไปโดยสุ่ม (Random assignment) และกระทำต่อทุกหน่วยทดลอง เป็นแบบแผนที่ง่ายและสะดวก มีข้อดีหลายอย่างคือให้ค่าองศาแห่งความอิสระ (Degree of freedom) ของค่ากำลังสองของความคลาดเคลื่อน (SSE) สูงสุด และถึงแม้ในแต่ละวิธีปฏิบัติจะมีจำนวนซ้ำไม่เท่ากันก็ไม่ทำให้การวิเคราะห์มีความซับซ้อนแต่อย่างใด นอกจากนี้สิ่งสำคัญที่เราต้องคำนึงถึงในการวางแผนการทดลองคือค่าความคลาดเคลื่อนของการทดลอง (Experimental error) หมายถึงค่าความแตกต่างของค่าสังเกตที่ได้จากหน่วยทดลองต่างๆ ซึ่งได้รับวิธีปฏิบัติ (treatment) เดียวกัน โดยสาเหตุสำคัญที่ทำให้เกิดความคลาดเคลื่อนนี้มี 2 ประการคือ

- ความแตกต่างในสิ่งทดลองที่มีมาก่อน (Inherent variability) เช่น นักเรียนแต่ละคนย่อมมีความแตกต่างในเรื่องประสบการณ์เดิม ระดับสติปัญญา (IQ) และอื่นๆ หรือแม้แต่แปลงผักแต่ละแปลง ถึงแม้ว่าได้รับวิธีปฏิบัติเดียวกัน แต่ให้ผลการเจริญเติบโตของพืชต่างกัน เพราะความสมบูรณ์ของดินแต่ละแปลงไม่เหมือนกัน

- ความแปรปรวนจากการทดลอง (Extraneous variability) คือความแปรปรวนเนื่องจากการขาดความเป็นเอกลักษณ์ของการกระทำทางกายภาพต่อสิ่งที่ทดลอง เช่นวิธีแรกสอนตอนเช้า วิธีหลังสอนตอนบ่าย ย่อมแตกต่างกันโดยธรรมชาติ หรือไม่ได้ใช้คนสอนคนเดียวกัน

จากปัญหาข้างต้น เราสามารถควบคุมและลดความแปรปรวนที่เกิดจากสิ่งทดลองที่มีมาก่อน (Inherent variability) โดยใช้เทคนิคการวิเคราะห์ความแปรปรวนร่วม (Analysis of Covariance) ซึ่งเป็นวิธีการหนึ่งที่นำมาใช้ในการวิเคราะห์ความแปรปรวนและการทดสอบ สมมติฐานของความ

แตกต่างระหว่างอิทธิพลของวิธีปฏิบัติ (treatment) โดยส่วนใหญ่จะพบในการวิเคราะห์ข้อมูลทางด้านชีววิทยา ด้านการแพทย์ และด้านการเกษตร เทคนิคดังกล่าวนี้เป็นการปรับอิทธิพลของตัวแปรบกวนที่เกิดจากสิ่งทดลองที่มีมาก่อนซึ่งเราไม่สามารถควบคุมได้ ตัวแปรดังกล่าวแทนด้วย X และเรียกว่าตัวแปรนี้ว่าตัวแปรร่วมหรือตัวแปรเสริมสัมพันธ์ (covariate or concomitant variable) โดยที่ตัวแปรดังกล่าวต้องมีความสัมพันธ์เชิงเส้นกับตัวแปรของค่าตอบสนอง (response) คือ Y ซึ่งเป็นค่าสังเกตที่ได้จากการให้วิธีปฏิบัติ เพื่อหาข้อสรุปเกี่ยวกับอิทธิพลของวิธีปฏิบัติ จึงได้ว่าวิธีการนี้เป็นการรวมทั้งการวิเคราะห์ความแปรปรวนและการวิเคราะห์ความถดถอย ซึ่งมีข้อตกลงเบื้องต้นดังนี้

- วิธีสุ่ม (Randomization) หมายถึงการเลือกหน่วยทดลองจากประชากรต้องเป็นไปโดยสุ่ม และการเลือกหน่วยทดลองให้กับวิธีปฏิบัติก็ต้องเป็นไปโดยสุ่มและอิสระซึ่งกันและกัน
- สัมประสิทธิ์การถดถอยในกลุ่มประชากรต้องเป็นเอกพันธ์ สำหรับวิธีปฏิบัติ t ระดับ (Homogeneity of Within-Group Regression Slopes)
- ตัวแปรร่วม X เป็นตัวแปรอิสระ และถือว่าไม่ได้รับอิทธิพลของวิธีปฏิบัติ (Statistical Independence of Covariate and Treatment)
- ค่าตัวแปรร่วมจะเป็นค่าที่กำหนด และการวัดไม่มีความคลาดเคลื่อน (Fixed Covariate Value Measured without Error)
- การถดถอยของตัวแปรตาม Y ที่ขึ้นกับตัวแปรร่วม X ต้องเป็นเชิงเส้น (Linearity of Within-Group Regression)
- ค่าความคลาดเคลื่อนของการทดลองเกิดขึ้นโดยการสุ่มและเป็นอิสระซึ่งกันและกัน มีการแจกแจงปกติโดยมีความแปรปรวนเท่ากัน (Normality and Homogeneity of Variance of Condition y Scores)
- ตัวแบบหุ่นแบบบวก (Additive Model) หมายถึงอิทธิพลของวิธีปฏิบัติและสิ่งแวดล้อมเป็นแบบบวก

ในทางปฏิบัติผู้วิจัยมักจะประสบปัญหาที่ค่าสังเกตบางค่ามีค่าสูงหรือต่ำมากอย่างผิดปกติไปจากข้อมูลส่วนใหญ่ ซึ่งเรียกว่าข้อมูลผิดปกติ (Outlier) อาจเนื่องมาจากความคลาดเคลื่อนจากการทดลอง เช่น การวัดหรือบันทึกค่าข้อมูลที่ผิดพลาด และการใช้เครื่องมือเครื่องใช้ที่คุณภาพต่ำ สิ่งเหล่านี้จะเป็นสาเหตุทำให้การแจกแจงของความคลาดเคลื่อนเปลี่ยนแปลงไปจากการแจกแจงปกติ นอกจากนี้ข้อมูล X และ Y อาจจะไม่มีความสัมพันธ์เชิงเส้น ซึ่งอาจทำให้วิธีการประมาณพารามิเตอร์แบบวิธีกำลังสองน้อยที่สุด (Least Square method) ซึ่งจะให้ค่าประมาณที่ไม่เอนเอียงที่มีความแปรปรวนต่ำสุด (Uniform Minimum Variance Unbiased Estimator) ในบรรดาตัวประมาณ

ที่ไม่มีความเอนเอียง ให้ผลการทดสอบที่คลาดเคลื่อนไปจากความเป็นจริง นั่นคืออาจจะทำให้เกิดค่าความคลาดเคลื่อนประเภทที่ 1 (Type I error) มีค่าสูง

ดังนั้นผู้วิจัยจึงสนใจศึกษาเปรียบเทียบประสิทธิภาพของวิธีการทดสอบในกรณีที่การแจกแจงของความคลาดเคลื่อนเป็นแบบเบ้ โดย**วิธีกำลังสองน้อยที่สุด** (Ordinary Least Square Method) **วิธีตัวประมาณเอ็ม** (M-Estimation Method) ซึ่งมีการนำเสนอเป็นครั้งแรก โดย Huber เมื่อปี ค.ศ. 1964 วิธีนี้จะให้ตัวประมาณที่มีความแกร่ง (robust) และ **วิธีการแปลงข้อมูลเป็นลำดับ** (Rank Transformation method) ซึ่งเป็นวิธีนอนพาราเมตริกที่นำมาใช้ในกรณีข้อมูลเป็นอันดับหรือข้อมูลไม่มีการแจกแจงแบบปกติและมีความแปรปรวนไม่เท่ากัน

1.2 สมมติฐานของการวิจัย

1.2.1 ในการวิเคราะห์ความแปรปรวนร่วมที่มีความคลาดเคลื่อนแบบเบ้ วิธีตัวประมาณเอ็มที่ใช้เกณฑ์ความแกร่งของแรมเซย์ (Ramsay) จะให้อำนาจการทดสอบสูงสุด

1.2.2 จำนวนซ้ำ $n \geq 15$ ทำให้อำนาจการทดสอบของวิธีการประมาณแบบกำลังสองน้อยที่สุดมีค่าใกล้เคียงกับวิธีตัวประมาณเอ็มที่ใช้เกณฑ์ความแกร่งของแรมเซย์

1.3 วัตถุประสงค์ของการวิจัย

1.3.1 ศึกษาวิธีการจำลองข้อมูลโดยใช้เทคนิคมอนติคาร์โลเพื่อสร้างเลขสุ่มและนำมาประยุกต์ใช้เพื่อแก้ปัญหาที่ต้องการศึกษา

1.3.2 เปรียบเทียบอำนาจทดสอบเพื่อหาวิธีการที่เหมาะสมในการวิเคราะห์ความแปรปรวนร่วมที่มีความคลาดเคลื่อนแบบเบ้ ด้วยการพิจารณาวิธีกำลังน้อยที่สุดกับวิธีตัวประมาณเอ็มที่ใช้เกณฑ์ความแกร่งของแรมเซย์ และวิธีการแปลงข้อมูลเป็นลำดับ

1.3.3 พิจารณาหาจำนวนซ้ำ (n) ที่ทำให้อำนาจการทดสอบของวิธีการประมาณแบบกำลังสองน้อยที่สุดมีค่าใกล้เคียงกับวิธีที่ให้อำนาจการทดสอบสูงสุด

1.4 ประโยชน์ที่จะได้รับจากการวิจัยเชิงทฤษฎี

เพื่อให้ผู้วิจัยสามารถเลือกใช้วิธีการทดสอบและขนาดตัวอย่าง ในการวิเคราะห์ความแปรปรวนร่วมได้อย่างมีประสิทธิภาพ เมื่อลักษณะการแจกแจงของความคลาดเคลื่อนเป็นแบบเบ้

1.5 ขอบเขต และวิธีการวิจัย

1.5.1 สร้างโปรแกรมโดยใช้โปรแกรมภาษาเทอร์โบปาสคาล เพื่อจำลองข้อมูลสุ่มในสถานการณ์ต่างๆ

1.5.2 กำหนดกลุ่มตัวอย่างหรือจำนวนวิธีปฏิบัติเท่ากับ 3,5 และ 7

1.5.3 เพื่อให้ตัวแปรอิสระ X มีความคล้ายธรรมชาติ จึงกำหนดให้ $X \sim N(\mu,\sigma^2)$

1.5.4 กำหนดขนาดตัวอย่างโดยใช้แต่ละกลุ่มเท่ากัน ดังแสดงในตารางที่ 1.1

ตารางที่ 1.1 แสดงขนาดตัวอย่างที่ใช้ในการวิจัย

จำนวนวิธีปฏิบัติ	จำนวนตัวแปรร่วม	ขนาดตัวอย่างแต่ละกลุ่ม
3	2	5 10 15
	4	5 10 15
5	2	5 10 15
	4	5 10 15
7	2	5 10 15
	4	5 10 15

มหาวิทยาลัยเชียงใหม่ Chiang Mai University

1.5.5 กำหนดให้ค่าเฉลี่ยแต่ละกลุ่มเท่ากันคือ $\mu = 15$

1.5.6 ลักษณะการแจกแจงของความคลาดเคลื่อนที่นำมาทดสอบ โดยมีค่าเฉลี่ยเท่ากับ μ และความแปรปรวนเท่ากับ σ^2 ทุกรูปแบบการศึกษาโดยมีการแจกแจงดังนี้

ก. การแจกแจงแบบลอกนอร์มอล (Lognormal distribution)

ผู้วิจัยสนใจ $\mu = 0$ และ $\sigma = 0.5$ และ 1.0 ตามลำดับ

ข. การแจกแจงแบบไคสแควร์ (Chi-square distribution)

ผู้วิจัยสนใจเมื่อ $\nu = 1, 2$ และ 3 ตามลำดับ

1.5.7 พิจารณาระดับนัยสำคัญ 2 ระดับคือ $\alpha = 0.01$ และ 0.05

1.5.8 การจำลองข้อมูลโดยใช้เทคนิคมอนติคาร์โล โดยกระทำซ้ำๆ กัน ไม่น้อยกว่า 1000 ครั้ง ในแต่ละสถานการณ์

1.5.9 ตัวแบบ (model) ที่ใช้เป็นแบบอิทธิพลกำหนด (fixed effect model) ในแผนแบบทดลองแบบสุ่มตลอด (Completely Randomized Design) เนื่องจากการวิจัยนี้ผู้วิจัยไม่ได้คำนึงที่จะสรุปผลไปถึงวิธีปฏิบัติอื่นๆ และไม่ต้องการให้เกิดความแปรปรวนอื่นเนื่องจากการสุ่มวิธีปฏิบัติ ตัวแบบที่ใช้คือ

$$Y_{ij} = \mu + \tau_j + \sum_{k=1}^{q} \beta_k (X_{ijk} - \bar{X}_{..k}) + \varepsilon_{ij} \quad ; \quad i = 1,2,\ldots,n_j$$

$$j = 1,2,\ldots,p$$

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 ค่าสัมประสิทธิ์การถดถอยของตัวแปรร่วมที่ k (partial regression coefficient with covariate k^{th})

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

1.6 คำจำกัดความของคำต่างๆ ที่ใช้ในการวิจัย

วิธีปฏิบัติ หมายถึง ปัจจัยที่ใช้ศึกษาทดลอง

ตัวแปรตาม (Dependent variable ; y) หมายถึง ค่าของหน่วยทดลองที่วัดได้หลังจากที่หน่วยทดลองได้รับอิทธิพลของวิธีปฏิบัติแล้ว

ตัวแปรร่วม (Concomitant variable ; x) หมายถึง ค่าของหน่วยทดลองที่วัดได้ก่อนที่จะได้รับอิทธิพลของวิธีปฏิบัติ ซึ่งเป็นค่าของตัวแปรที่แฝงมากับหน่วยทดลอง

ความน่าจะเป็นของความผิดพลาดประเภทที่ 1 (Type I error) คือ ความน่าจะเป็นของความผิดพลาดที่เกิดจากการปฏิเสธสมมติฐานว่าง (H_0) เมื่อสมมติฐานว่างเป็นจริง ใช้แทนด้วยสัญลักษณ์ α

ความน่าจะเป็นของความผิดพลาดประเภทที่ 2 (Type II error) คือ ความน่าจะเป็นของความผิดพลาดที่เกิดจากการยอมรับสมมติฐานว่าง (H_0) เมื่อสมมติฐานว่างไม่จริง ใช้แทนด้วยสัญลักษณ์ β

อำนาจการทดสอบ (Power of the test) คือ ค่าความน่าจะเป็นที่จะปฏิเสธสมมติฐานว่าง (H_0) เมื่อสมมติฐานว่างไม่จริง ซึ่งมีค่าเท่ากับ $1-\beta$

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