บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในการทดลองทุกประเภทมีหลักสำคัญข้อหนึ่งคือ ผู้ทดลองจะต้องพยายามใช้วิธีการต่าง ๆ เพื่อควบคุมหรือลดความแปรปรวนของหน่วยทดลอง (experimental unit) ให้มีค่าน้อยที่สุด ซึ่งการควบคุมหรือการลดความแปรปรวนดังกล่าวจะจำแนกได้เป็น 2 ประเภท คือ การควบคุมทางตรงและการควบคุมโดยใช้วิธีการทางสถิติ

การควบคุมความแปรปรวนทางตรงทำโดยอาศัยเทคนิคต่าง ๆ เช่น ใช้หน่วยทดลองและเลือกสภาพแวดล้อมที่สม่ำเสมอ วัดผลการทดลองให้ถูกต้อง ใช้แผนการทดลองที่เหมาะสม เป็นต้น การควบคุมความแปรปรวนทางตรงนั้นบางครั้งอาจไม่สามารถทำได้อย่างมีประสิทธิภาพ ดังนั้นผู้ทดลองควรจะใช้วิธีการทางสถิติประกอบการพิจารณา เช่น การวิเคราะห์ความแปรปรวนร่วม (analysis of covariance) เป็นต้น

การวิเคราะห์ความแปรปรวนร่วมเป็นวิธีที่ผสมผสานหลักการของการวิเคราะห์ความแปรปรวนกับการวิเคราะห์ความถดถอยเข้าด้วยกัน โดยที่แต่ละหน่วยทดลองจะต้องเก็บค่าตัวแปรตาม y (dependent variable) และตัวแปรร่วม x (covariate หรือ concomitant variables) ซึ่งตัวแปรร่วมจะเป็นตัวแปรที่วัดเพิ่มและใช้ปรับค่าตัวแปรตามเพื่อหาข้อสรุปอิทธิพลของทรีตเม้นต์ว่าแตกต่างกันหรือไม่

โดยทั่วไปวิธีการประมาณพารามิเตอร์ในการวิเคราะห์ความแปรปรวนร่วมจะใช้วิธีกำลังสองน้อยที่สุด (Least Square Method) ซึ่งจะได้ตัวประมาณที่ไม่เอนเอียง (unbiased estimator) ซึ่งมีความแปรปรวนต่ำสุดสม่ำเสมอในบรรดาตัวประมาณที่ไม่เอน-เอียง แต่ถ้าลักษณะการแจกแจงความคลาดเคลื่อนไม่เป็นไปตามข้อตกลงเบื้องต้น คือไม่เป็นการแจกแจงแบบปกติ เช่น เป็นการแจกแจงที่มีหางยาว หรือ มีการกระจายไปทางหางมากกว่าปกติ ซึ่งลักษณะการแจกแจงดังกล่าวจะทำให้ความคลาดเคลื่อนบางค่ามีค่าสูงมาก ๆ

หรือต่ำมาก ๆ โดยที่ความคลาดเคลื่อนที่เกิดขึ้นนี้จะทำให้ค่าตัวแปรตามบางค่าเป็นค่าผิดปกติ (outlier) ลักษณะดังกล่าวเกิดขึ้นมากทางด้านชีววิทยา เคมี และการแพทย์ ดังนั้นวิธีการประมาณพารามิเตอร์วิธีนี้อาจจะไม่เหมาะสม ผู้วิจัยอาจเลือกใช้วิธีการนอนพาราเมตริก (nonparametric method) ซึ่งใช้ประมาณค่าพารามิเตอร์ได้เช่นเดียวกันกับวิธีกำลังสองน้อยที่สุดโดยไม่จำเป็นต้องมีข้อตกลงเบื้องต้นเกี่ยวกับลักษณะการแจกแจงของความคลาดเคลื่อน

ในบรรดาวิธีการนอนพาราเมตริกได้มีผู้ศึกษาวิธีประมาณค่าส่วนเบี่ยงเบนมาตรฐาน (standard deviation) เมื่อไม่ทราบลักษณะการแจกแจงของประชากร โดยใช้เทคนิคของการสุ่มตัวอย่างซ้ำ (resampling) ซึ่งมีอยู่ด้วยกันหลายวิธี ได้แก่

- The Jackknife Method
- The Bootstrap Method
- Half - Sampling Method
- Subsampling Method
- Balanced Repeated Replication Method
- The Infinitesimal Jackknife Method
- Influence Function Techniques Method
- The Delta Method

โดยที่แต่ละวิธีมีแนวความคิดพื้นฐานคล้ายกันคือ เมื่อเก็บรวบรวมข้อมูลได้แล้วจะทำการสุ่มตัวอย่างซ้ำ ซึ่งพบว่าวิธีบูตสแตรป (Bootstrap Method) เป็นวิธีที่ให้ผลดีที่สุด นอกจากนี้วิธีบูตสแตรปสามารถนำไปใช้ในการประมาณค่าพารามิเตอร์อื่น ๆ ที่สนใจเมื่อไม่-ทราบลักษณะการแจกแจงของประชากร

ในปี ค.ศ.1979 แบรดเลย์ เอฟรอน (Bradley Efron) ได้ศึกษาและนำเอาวิธีบูตสแตรปมาใช้ในการประมาณค่าต่าง ๆ ที่ไม่สามารถหาได้โดยตรงในทางพาราเมตริก (parametric) เช่น ค่าส่วนเบี่ยงเบนมาตรฐานของค่าสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient) และค่าส่วนเบี่ยงเบนมาตรฐานของสัมประสิทธิ์ความถดถอยเชิงเส้นในกรณีที่ไม่ทราบลักษณะการแจกแจงของความคลาดเคลื่อน ซึ่งวิธีบูตสแตรปนี้มีแนวความคิดมาจากวิธี Jackknife ของเควโนอีล (Quenouille) และตูกี้ (Turkey)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

เพราะฉะนั้นผู้วิจัยจึงสนใจศึกษาและนำวิธีบูตสแตรปมาใช้ประมาณค่าพารามิเตอร์ในการวิเคราะห์ความแปรปรวนร่วม เมื่อลักษณะการแจกแจงของความคลาดเคลื่อนเป็นแบบหางยาว หรือ กระจายไปทางหางมากกว่าปกติ

1.2 วัตถุประสงค์ของการวิจัย

เพื่อเปรียบเทียบประสิทธิภาพวิธีประมาณพารามิเตอร์ระหว่างวิธีกำลังสองน้อยที่สุดกับวิธีบูตสแตรปในการวิเคราะห์ความแปรปรวนร่วม เมื่อลักษณะการแจกแจงของความคลาดเคลื่อนไม่เป็นไปตามข้อตกลงเบื้องต้น กล่าวคือเป็นการแจกแจงแบบหางยาว หรือกระจายไปทางหางมากกว่าปกติ โดยพิจารณาจากความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดประเภทที่ 1 (P(type I error)) และอำนาจการทดสอบ (power of the test)

1.3 สมมติฐานของการวิจัย

ในการวิเคราะห์ความแปรปรวนร่วมเมื่อลักษณะการแจกแจงของความคลาดเคลื่อนเป็นแบบหางยาว* วิธีการประมาณพารามิเตอร์ด้วยวิธีบูตสแตรปจะให้อำนาจการทดสอบสูงกว่าวิธีกำลังสองน้อยที่สุด

---

* โดยทั่วไปจะใช้ค่าความโด่ง (kurtosis,K) ในการพิจารณาลักษณะการแจกแจงของประชากรแบบสมมาตร ถ้าค่า $K > 4.5$ จะเป็นการแจกแจงแบบหางยาวกว่าปกติ ในกรณีที่ค่า $K < 2.5$ จะเป็นการแจกแจงแบบหางสั้นกว่าปกติ และ $2.5 \leq K \leq 4.5$ จะเป็นการแจกแจงที่เข้าใกล้แบบปกติ ส่วน $K=3$ จะเป็นการแจกแจงแบบปกติ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

1.4 ข้อตกลงเบื้องต้น

1.4.1 พิจารณาตัวแบบ (model)

$$y_{ij} = \mu + \tau_i + \sum_{k=1}^{q} \beta_k (x_{ijk} - \bar{x}_{..k}) + \varepsilon_{ij}$$

$i = 1,2,3,...p$

$j = 1,2,3,...n_i$

$k = 1,2,3,...q$

โดยที่

p เป็นจำนวนทรีตเม้นต์

n_i เป็นขนาดตัวอย่างในทรีตเม้นต์ที่ i

q เป็นจำนวนตัวแปรร่วม

y_{ij} หมายถึง ค่าสังเกตของตัวแปรตามจากหน่วยทดลองที่ j ทรีตเม้นต์ที่ i

μ หมายถึง ค่าเฉลี่ยทั้งหมดของ y

τ_i หมายถึง อิทธิพลของทรีตเม้นต์ที่ i

β_k หมายถึง ค่าสัมประสิทธิ์ความถดถอยของตัวแปรร่วมที่ k

$\bar{x}_{..k}$ หมายถึง ค่าเฉลี่ยของตัวแปรร่วมที่ k

ε_{ij} หมายถึง ค่าความคลาดเคลื่อนของการทดลองจากหน่วยทดลองที่ j ทรีตเม้นต์ที่ i

1.4.2 วิธีสุ่ม (Randomization) หมายถึงการเลือกหน่วยทดลองจากประชากรต้องเป็นไปโดยสุ่ม และการเลือกหน่วยทดลองให้กับทรีตเม้นต์ต้องเป็นไปโดยสุ่มและเป็นอิสระซึ่งกันและกัน

1.4.3 τ_i เป็นอิทธิพลของทรีตเม้นต์ (treatment effect) ที่ i โดยที่ $\sum_{i=1}^{p} \tau_i = 0$

1.4.4 ความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรร่วมเป็นแบบเชิงเส้น

1.4.5 ค่าสัมประสิทธิ์ความถดถอยของแต่ละทรีตเม้นต์มีค่าเท่ากัน

1.4.6 ตัวแปรร่วม x เป็นตัวแปรอิสระ ถือว่าไม่ได้รับอิทธิพลของทรีตเม้นต์และการวัดไม่มีความคลาดเคลื่อน

1.4.7 อิทธิพลของทรีตเม้นต์และสิ่งแวดล้อมอื่น ๆ เป็นแบบบวก (Additive)

1.4.8 ความคลาดเคลื่อนเป็นตัวแปรสุ่มแบบต่อเนื่องที่มีการแจกแจงแบบเดียวกันและเป็นอิสระซึ่งกันและกัน มีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น σ^2

1.5 ขอบเขตของการวิจัย

1.5.1 ตัวแบบเป็นอิทธิพลกำหนด (fixed effect) ในแผนแบบการทดลองแบบสุ่มตลอด (completely randomized design)

1.5.2 กำหนดจำนวนตัวแปรร่วม x = 1,3,5

1.5.3 กำหนดจำนวนทรีตเม้นต์ = 3,5,7

1.5.4 กำหนดขนาดตัวอย่างแต่ละทรีตเม้นต์ ดังนี้

จำนวนทรีตเม้นต์	จำนวนตัวแปรร่วม	ขนาดตัวอย่างในแต่ละทรีตเม้นต์
3	1	6
	1	8
	1	10
	1	12
3	3	6
	3	8
	3	10
	3	12
3	5	6
	5	8
	5	10
	5	12

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จำนวนทรีตเม้นต์	จำนวนตัวแปรร่วม	ขนาดตัวอย่างในแต่ละทรีตเม้นต์
5	1	4
	1	6
	1	8
5	3	4
	3	6
	3	8
5	5	4
	5	6
	5	8
7	1	4
	1	6
	1	8
7	3	4
	3	6
	3	8
7	5	4
	5	6
	5	8

1.5.5 ลักษณะการแจกแจงความคลาดเคลื่อนที่นำมาศึกษา มีค่าเฉลี่ย (μ) = 0 และความแปรปรวน (σ^2) =100 และ 400 ทุกรูปแบบ * โดยมีการแจกแจงดังนี้

---

* การวิจัยครั้งนี้กำหนดค่าความแปรปรวนดังกล่าว เนื่องจากการทดลองที่จำนวนทรีตเม้นต์เท่ากับ 3 จำนวนตัวแปรร่วมเท่ากับ 1 และ 3 ขนาดตัวอย่างในแต่ละทรีตเม้นต์เท่ากับ 6 และที่จำนวนทรีตเม้นต์เท่ากับ 5 จำนวนตัวแปรร่วมเท่ากับ 1 ขนาดตัวอย่างในแต่ละทรีตเม้นต์เท่ากับ 4 พบว่าค่าความแปรปรวนน้อยกว่า 400($\sigma^2 < 400$) มีผลต่อค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1 และอำนาจการทดสอบ ทุกรูปแบบการแจกแจง

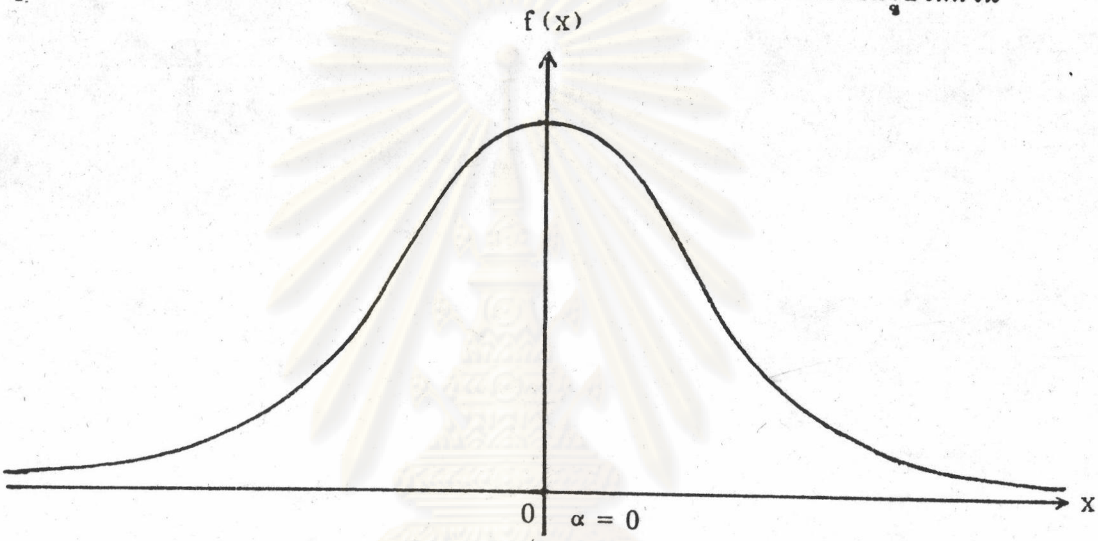
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ก. <u>การแจกแจงแบบโลจิสติค</u> (logistic distribution)

ฟังก์ชันความหนาแน่นอยู่ในรูปของ

$$f(x) = \begin{cases} \dfrac{1}{\beta} \dfrac{e^{\frac{-(x-\alpha)}{\beta}}}{[1 + e^{\frac{-(x-\alpha)}{\beta}}]^2} & , \beta>0 \\ 0 & , \text{อื่น ๆ} \end{cases}$$

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



จะได้ว่า ค่าคาดหวัง $E(x) = \alpha$ และความแปรปรวน $Var(x) = \frac{1}{3}\pi^2\beta^2$

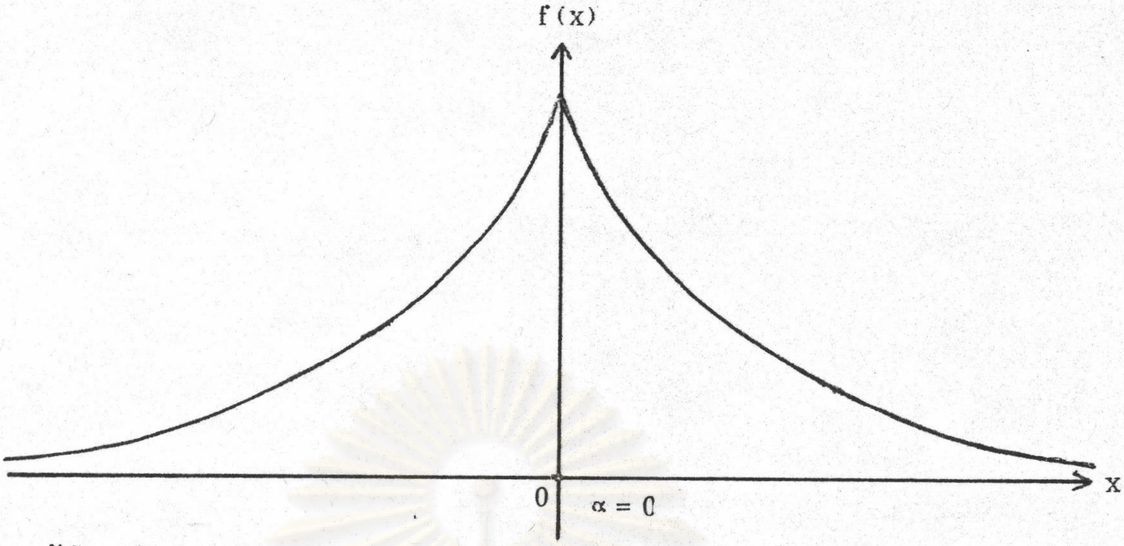
ข. <u>การแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียล</u> (double exponential distribution)

ฟังก์ชันความหนาแน่นอยู่ในรูปของ

$$f(x) = \begin{cases} \dfrac{1}{2\beta} \cdot e^{-\left|\frac{x-\alpha}{\beta}\right|} & , \beta>0 \\ 0 & , \text{อื่น ๆ} \end{cases}$$

ผู้วิจัยสนใจศึกษาเมื่อ $\alpha = 0,\ \beta = \sigma / \sqrt{2}$ ซึ่งสามารถเขียนรูปได้ดังนี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



จะได้ว่า ค่าคาดหวัง $E(x) = \alpha$ และความแปรปรวน $Var(x) = 2\beta^2$

ค. <u>การแจกแจงแบบปกติปลอมปน</u> (scale contaminated normal distribution)

การแจกแจงแบบปกติปลอมปนเป็นการแจกแจงที่แปลงมาจากการแจกแจงแบบปกติ ซึ่งมีฟังก์ชันการแปลงเป็นดังนี้

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

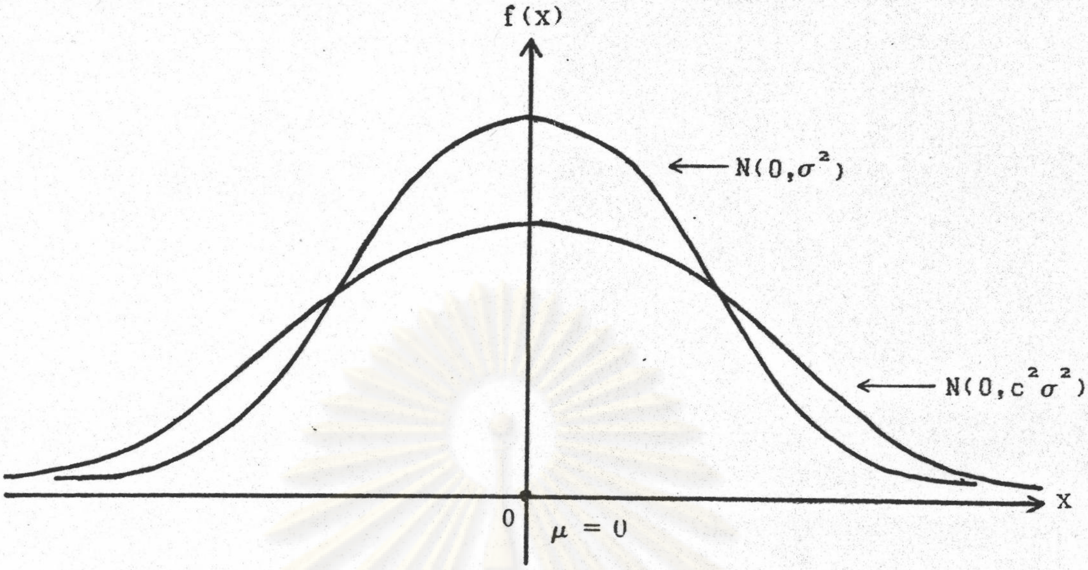
เมื่อ c คือสเกลแฟคเตอร์ (scale factor) จะใช้ c = 3,10 * และ p คือเปอร์เซ็นต์ของการปลอมปน (percent of contamination) ซึ่งใช้ p= 5,10,25

หมายความว่าค่า x จะมาจากการแจกแจงแบบ $N(0,\sigma^2)$ ด้วยความน่าจะเป็น 1-p และจะมาจากการแจกแจงแบบ $N(0,c^2\sigma^2)$ ด้วยความน่าจะเป็น p ผู้วิจัยสนใจศึกษาเมื่อกำหนดค่าเฉลี่ย = 0 และความแปรปรวน = σ^2

---

* สเกลแฟคเตอร์ หมายถึง ค่าที่ปรับให้ข้อมูลมีการกระจายมากขึ้น โดยที่สเกลแฟคเตอร์มีค่ามากกว่า 10 จะมีโอกาสของการเกิดค่าผิดปกติมาก ถ้ามีค่าน้อยกว่า 3 จะมีโอกาสของการเกิดค่าผิดปกติน้อย

ซึ่งสามารถเขียนรูปได้ดังนี้



1.5.6 กำหนดระดับนัยสำคัญ 2 ระดับ คือ α = 0.01 และ 0.05

1.6 ประโยชน์ที่คาดว่าจะได้รับ

เพื่อเป็นแนวทางให้ผู้วิจัยสามารถเลือกใช้วิธีประมาณพารามิเตอร์ในการวิเคราะห์ความแปรปรวนร่วมได้อย่างมีประสิทธิภาพเมื่อลักษณะการแจกแจงของความคลาดเคลื่อนเป็นแบบหางยาว หรือกระจายไปทางหางมากกว่าปกติ

1.7 วิธีดำเนินการวิจัย

1.7.1 ศึกษาวิธีการประมาณพารามิเตอร์ในการวิเคราะห์ความแปรปรวนร่วมและเขียนโปรแกรมจำลองค่าสังเกตของตัวแปรในตัวแบบที่ต้องการศึกษา รวมทั้งโปรแกรมสำหรับการคำนวณค่าประมาณพารามิเตอร์ของแต่ละวิธีดังนี้

ก) วิธีกำลังสองน้อยที่สุด

ข) วิธีบูตสแตรป

1.7.2 ศึกษาเปรียบเทียบความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดประเภทที่ 1 และอำนาจการทดสอบ ของวิธีการประมาณพารามิเตอร์

ในการวิเคราะห์ความแปรปรวนร่วม ระหว่างวิธีกำลังสองน้อยที่สุดกับวิธีบูตสแตรป ข้อมูลที่ใช้ในการวิจัยครั้งนี้ได้จากการจำลองโดยเทคนิคมอนติคาร์โล (Monte Carlo Technique) จะกระทำซ้ำ 1000 ครั้งในแต่ละสถานการณ์

1.8 คำจำกัดความของค่าต่างๆ ที่ใช้ในการวิจัย

1.8.1 ความผิดพลาดประเภทที่ 1 (Type I error) คือความผิดพลาดที่เกิดจากการปฏิเสธสมมติฐานว่าง เมื่อสมมติฐานว่างเป็นจริง

1.8.2 ความผิดพลาดประเภทที่ 2 (Type II error) คือความผิดพลาดที่เกิดจากการยอมรับสมมติฐานว่าง เมื่อสมมติฐานว่างนั้นไม่จริง

1.8.3 อำนาจการทดสอบ (Power of the test) คือความน่าจะเป็นที่จะปฏิเสธสมมติฐานว่าง เมื่อสมมติฐานว่างนั้นไม่จริง

1.8.4 ทรีตเม้นต์ (Treatment) หรือสิ่งทดลอง คือปัจจัยหนึ่ง ๆ วิธีการหนึ่ง หรือบางครั้งคือ ระดับหนึ่ง ๆ ของปัจจัยที่ใช้ศึกษาทดลอง เช่น พันธุ์ต่าง ๆ ของพืช

1.8.5 หน่วยทดลอง (Experimental unit) คือหน่วยที่เล็กที่สุดของวัสดุหรือวัตถุในการทดลอง ที่จะได้รับแต่ละสิ่งทดลอง (treatment) ต่าง ๆ และนำมาศึกษาข้อมูล

1.8.6 ตัวแปรตาม y (Dependent variable) คือค่าของหน่วยทดลองที่วัดได้หลังจากที่หน่วยทดลองได้รับอิทธิพลของทรีตเม้นต์

1.8.7 ตัวแปรร่วม x (Covariate หรือ Concomitant variables) คือค่าของหน่วยทดลองที่วัดได้ก่อนที่หน่วยทดลองจะได้รับอิทธิพลของทรีตเม้นต์ ซึ่งเป็นค่าของตัวแปรที่แฝงมากับหน่วยทดลอง

1.8.8 ค่าผิดปกติ (Outlier) คือ ค่าสังเกตที่มีค่ามากหรือน้อยกว่าค่าสังเกตอื่น ๆ อย่างผิดปกติ

1.8.9 ค่าอัตราส่วนของค่าแตกต่างอำนาจการทดสอบ (Ratio of Different of Power of the test (RDP)) ระหว่างวิธีกำลังสองน้อยที่สุดกับวิธีบูตสแตรป

$$
RDP = \begin{cases} \left[\dfrac{P_{OLS} - P_B}{P_B}\right] \times 100 & , \; P_B < P_{OLS} \\[2ex] \left[\dfrac{P_B - P_{OLS}}{P_{OLS}}\right] \times 100 & , \; P_{OLS} < P_B \end{cases}
$$

เมื่อ P_{OLS} แทน อำนาจการทดสอบของวิธีประมาณพารามิเตอร์วิธีกำลังสองน้อยที่สุด

P_B แทน อำนาจการทดสอบของวิธีประมาณพารามิเตอร์วิธีบูตสแตรป

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย