# บทที่ 1

# บทนำ

## 1.1 ที่มาและความสำคัญของปัญหา

ในการวางแผนการทดลอง หลักการจัดบล็อคนับว่ามีความสำคัญมากอย่างหนึ่ง ในกรณีที่มีทรีทเมนต์ (Treatment) จำนวนมาก ๆ ทำให้ขนาดของบล็อคใหญ่มากเกินไป ซึ่งเป็นสาเหตุให้ภายในบล็อคไม่สม่ำเสมอ ซึ่งหลักสำคัญในการจัดบล็อคก็คือภายในบล็อคเดียวกันต้องมีความสม่ำเสมอและต่างบล็อคกันต้องแตกต่างกันมากที่สุด จากเหตุผลนี้ทำให้ผู้ทดลองไม่สามารถใช้แผนการทดลองแบบบล็อคสมบูรณ์ได้อย่างถูกต้องและมีประสิทธิภาพ จึงจำเป็นต้องใช้แผนการทดลองแบบบล็อคไม่สมบูรณ์ (แต่ละบล็อคมีไม่ครบทุกสิ่งทดลอง) เพื่อให้ขนาดของบล็อคเหมาะสม โดยทั่วไปจำนวนสิ่งทดลองต่อบล็อคจะเท่ากันทุกบล็อคและทรีทเมนต์ (Treatment) แต่ละคู่ปรากฏร่วมกันในบล็อคเป็นจำนวนครั้งเท่ากัน แผนการทดลองประเภทนี้ เรียกว่า แผนการทดลองแบบบล็อคไม่สมบูรณ์สมดุลย์ (Balanced Incomplete Block Design : BIB) ซึ่งมีหลักการจัดบล็อคและทรีทเมนต์มาจากแผนการทดลองแบบบล็อคสมบูรณ์ (Randomized Complete Block Design : RBD) และแผนการทดลองแบบลาตินสแควร์ (Latin Square Design : LS) เป็นส่วนใหญ่

แผนการทดลองแบบบล็อคไม่สมบูรณ์สมดุลย์ (Balanced Incomplete Block Design : BIB) แผนการทดลองนี้ เป็นแผนแบบการทดลองที่มีการจัดให้ทรีทเมนต์แต่ละคู่ปรากฏร่วมกันในบล็อคเป็นจำนวนครั้งเท่ากัน มีหลักเกณฑ์ในการพิจารณาดังนี้

1. ในแต่ละบล็อคไม่สมบูรณ์ประกอบด้วยหน่วยทดลองเท่ากัน

2. แต่ละทรีทเมนต์ จะปรากฏในการทดลองเป็นจำนวนครั้งเท่ากัน (จำนวนซ้ำเท่ากัน)

3. ทรีทเมนต์แต่ละคู่ปรากฏร่วมกันในบล็อคเป็นจำนวนครั้งเท่ากัน ซึ่งจำนวนครั้งหรือ $\lambda$ หาได้จากสูตร $\lambda = r(k-1)/(t-1)$

4. จำนวนหน่วยทดลองทั้งหมดหาได้จาก

$$tr = bk$$

โดยที่

t = จำนวนทรีทเมนต์ (treatment)

b = จำนวนบล็อค (block)

k = จำนวนทรีทเมนต์หรือจำนวนหน่วยทดลองต่อบล็อค (k<t)

r = จำนวนซ้ำ

$\lambda$ = จำนวนครั้งที่ทรีทเมนต์แต่ละคู่ปรากฏร่วมกันในบล็อค

ในงานทดลองบางครั้งข้อมูลที่รวบรวมมาเพื่อทำการวิเคราะห์อาจมีค่าสูญหายหรือใช้ไม่ได้บ้าง ซึ่งไม่สามารถตามไปเก็บเพิ่มเติมได้ เช่น สัตว์ทดลองอาจป่วยตาย ซึ่งสาเหตุมิใช่เนื่องจากทรีทเมนต์ (treatment) หรือหมู่ป่าลงกินข้าวโพดในแปลงทดลอง เป็นต้น ซึ่งปัญหาการสูญหายของข้อมูล อาจมีผลกระทบต่อผลสรุปของการวิเคราะห์ ถ้าต้องตัดข้อมูลนั้นทิ้งไป หรือบางครั้งเราไม่สามารถตัดข้อมูลนั้น ๆ ทิ้งได้ เช่น ในแผนการทดลองแบบบล็อคไม่สมบูรณ์สมดุลย์ (BIB) เพราะจากข้อกำหนดของแผนการทดลองนี้ถ้าเราตัดข้อมูลทิ้ง เราจะไม่สามารถวิเคราะห์ข้อมูลตามแผนการทดลองที่วางไว้แต่ต้นได้ เพราะเราไม่สามารถหาค่า r, $\lambda$ ได้ ดังนั้นจึงต้องหาวิธีการแก้ปัญหาข้อมูลสูญหาย เพื่อให้ได้ผลสรุปจากการวิเคราะห์ข้อมูลชุดนั้น ๆ อย่างถูกต้อง

สำหรับปัญหาข้อมูลสูญหายในแผนการทดลองแบบบล็อคไม่สมบูรณ์สมดุลย์นั้น เมื่อ ค.ศ. 1940 CORNISH, E.A. ได้เสนอวิธีประมาณค่าข้อมูลที่สูญหาย ดังนี้

1. กรณีข้อมูลสูญหาย 1 ค่า

$$Y_{ij} = \frac{tr(k-1)\,B_j + k(t-1)\,Q_i - (t-1)\,Q_i}{(k-1)\,[tr(k-1) - k(t-1)]} \qquad \text{.....(1.1)}$$

โดยที่ $Q_i = kT_i - B_{(i)}$ = ค่า Q สำหรับทรีทเมนต์ที่มีค่าสูญหาย

$T_i$ = ผลรวมทรีทเมนต์ (treatment) ที่ i

$B_{(i)}$ = ผลรวมของทุกบล็อคที่สิ่งทดลองที่ i ปรากฏ

$B_j$ = ผลรวมค่าสังเกตในบล็อคที่มีค่าสูญหาย

$Y_{ij}$ = ค่าประมาณของข้อมูลตัวที่สูญหาย

$Q_i$ = ผลรวมของ Q สำหรับทรีทเมนต์ในบล็อคเดียวกันกับค่าสูญหาย

2. <u>เมื่อมีข้อมูลสูญหายจำนวนหลายค่า</u>

ให้แทนข้อมูลที่สูญหายด้วยตัวแปร เช่น x, y, z, เป็นต้น แล้วคำนวณหาค่าประมาณข้อมูลที่สูญหายตัวแรกก่อน แล้วนำค่าที่ได้ไปแทนที่ข้อมูลตัวนั้น แล้วคำนวณหาค่าประมาณข้อมูลสูญหายตัวที่ 2 ตามสูตร (1.1) ทำเช่นนี้ครบตามจำนวนข้อมูลที่สูญหาย ถือว่าครบ 1 รอบแล้วทำรอบที่ 2 , 3 ... ต่อไปอีก จนกว่าค่าประมาณข้อมูลที่สูญหายแต่ละค่าจะคงที่ไม่เปลี่ยนแปลงจากรอบที่ผ่านมา หรือเปลี่ยนแปลงน้อยมาก แล้วใช้ค่าเหล่านี้เป็นค่าประมาณข้อมูลที่สูญหาย ทำการวิเคราะห์ข้อมูลตามปกติ แต่องศาความเป็นอิสระของทั้งหมดและของความคลาดเคลื่อนจะลดลงเท่ากับจำนวนข้อมูลที่สูญหาย

จากวิธีการของ CORNISH, E.A. นั้นจะเห็นว่าในทางปฏิบัตินั้นยุ่งยากเสียเวลามากและในบล็อคหนึ่ง ๆ จะต้องมีข้อมูลอย่างน้อย 1 ค่า ดังนั้นถ้าหากเกิดปัญหาข้อมูลสูญหายทั้งบล็อคขึ้นวิธีของ CORNISH, E.A. จะไม่เป็นวิธีที่เหมาะสมผู้วิจัยจึงสนใจที่จะศึกษาเปรียบเทียบวิธีวิเคราะห์เมื่อข้อมูลสูญหายทั้งบล็อคในแผนการทดลองแบบบล็อคไม่สมบูรณ์สมดุลย์ 3 วิธี คือ

วิธีที่ 1 วิธีของ P.D. PURI

วิธีที่ 2 วิธีของ G.N. WILKINSON

วิธีที่ 3 วิธีประมาณค่าสูญหายโดยให้ค่าความคลาดเคลื่อนกำลังสองต่ำสุด

โดยพิจารณาจากค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MEAN ABSOLUTE ERROR:MAE) และอำนาจการทดสอบ

$$\text{MEAN ABSOLUTE ERROR} = \sum_{i=1}^{m} \frac{|MSE_i - MSE_{ij}|}{m}$$

$MSE_i$ : MEAN SQUARE ERROR ตัวที่ i เมื่อข้อมูลปกติ

$MSE_{ij}$ : MEAN SQUARE ERROR ตัวที่ i เมื่อแก้ปัญหาข้อมูลสูญหายด้วยวิธี j

$j = 1,2,3$

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อเปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) จากการวิเคราะห์ทั้ง 3 วิธี

2. เพื่อเปรียบเทียบอำนาจการทดสอบของทั้ง 3 วิธีว่าแตกต่างกันหรือไม่

1.3 สมมติฐานของการวิจัย

1. ภายใต้ลักษณะการแจกแจงของความคลาดเคลื่อนต่าง ๆ กัน วิธีที่ 3 จะให้ค่าอำนาจการทดสอบสูงสุด

2. ภายใต้ลักษณะการแจกแจงของความคลาดเคลื่อนต่าง ๆ กันวิธีที่ 3 จะให้ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยต่ำสุด

1.4 ข้อตกลงเบื้องต้น

1. การวิจัยครั้งนี้ถือว่าค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยเป็นเกณฑ์สำคัญที่ใช้เปรียบเทียบวิธีวิเคราะห์ทั้ง 3 วิธี และอำนาจการทดสอบเป็นเกณฑ์ต่อไปที่พิจารณา (อันดับที่ 2)

2. การสูญหายของข้อมูลเป็นไปโดยสุ่มและหายเพียง 1 บล็อคเท่านั้น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

1.5 ขอบเขตการวิจัย

1. ตัวแบบ (Model) ที่ศึกษาเป็นแบบอิทธิพลกำหนด (Fixed Effect Model) คือ

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

$\mu$ หมายถึง ค่าเฉลี่ยประชากร

$\tau_i$ หมายถึง อิทธิพลของทรีทเมนต์ (TREATMENT EFFECT) ที่ i และเป็นอิทธิพลคงที่

$\beta_j$ หมายถึง อิทธิพลของบล็อค (BLOCK EFFICT) ที่ j และเป็นอิทธิพลคงที่เช่นกัน

$\varepsilon_{ij}$ หมายถึง ความคลาดเคลื่อนจากทรีทเมนต์ที่ i บล็อค j

2. ศึกษาความคลาดเคลื่อนสัมบูรณ์เฉลี่ยของทั้ง 3 วิธี เมื่อความคลาดเคลื่อนมีการแจกแจงแบบปกติแบบโลจิสติค แบบดับเบิ้ลเอ็กซ์โปเนนเชียลและแบบปกติปลอมปน

3. ศึกษาอำนาจการทดสอบของทั้ง 3 วิธี เมื่อความคลาดเคลื่อนมีการแจกแจงแบบต่าง ๆ ในข้อ 2

4. กำหนดให้ค่าเฉลี่ยของความคลาดเคลื่อนเท่ากับ 0 และความแปรปรวนเป็น 10 25 50 75

5. ในกรณีการแจกแจงแบบปกติปลอมปน จะศึกษาเมื่อเปอร์เซ็นต์การปลอมปนเป็น 5% 10% และ 25% สำหรับสเกลแฟคเตอร์ (Scale factor) มี 2 ระดับคือ 3 และ 10

6. ศึกษาอำนาจการทดสอบ ณ ระดับนัยสำคัญ 0.01 0.05 และ 0.10

7. แผนการทดลองที่ศึกษาแบ่งตามค่า t, b, k, r, $\lambda$ ดังนี้

| t | k | r | b | λ |
|---|---|---|---|---|
| 4 | 3 | 3 | 4 | 2 |
| 6 | 3 | 5 | 10 | 2 |
| 7 | 3 | 3 | 7 | 1 |
| 7 | 4 | 4 | 7 | 2 |
| 9 | 3 | 4 | 12 | 1 |
| 10 | 4 | 6 | 15 | 2 |
| 10 | 5 | 9 | 18 | 4 |
| 10 | 6 | 9 | 15 | 5 |

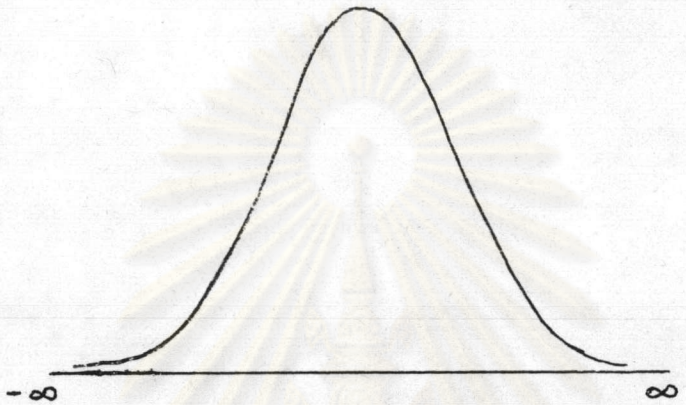
8. ในการวิจัยครั้งนี้ จำลองการทดลองขึ้นโดยใช้เทคนิคมอนติคาร์โลซึ่งเลขสุ่มจากเครื่องคอมพิวเตอร์ IBM 370/3031 ซึ่งศึกษาเมื่อความคลาดเคลื่อนมีการแจกแจงแบบปกติ โลจิสติค ดับเบิ้ลเอ็กซ์โปเนนเชียล และแบบปกติปลอมปน โดยมีค่าฟังก์ชันความน่าจะเป็นค่าคาดหวัง ค่าความแปรปรวนของการแจกแจงแต่ละรูปแบบเป็นดังนี้

8.1 <u>การแจกแจงแบบปกติ</u> (Normal Distribution)

ฟังก์ชันความน่าจะเป็นคือ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

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



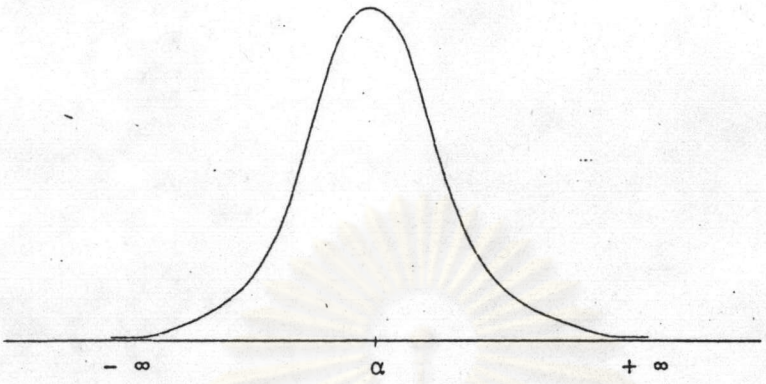

ค่าคาดหวัง $E(x) = \mu$

ความแปรปรวน $V(x) = \sigma^2$

8.2 <u>การแจกแจงแบบโลจิสติค</u> (Logistic Distribution)

ฟังก์ชันความน่าจะเป็นคือ

$$f(x) = \frac{1}{\beta} \frac{e^{-\left(\frac{x-\alpha}{\beta}\right)}}{\left[1 + e^{-\left(\frac{x-\alpha}{\beta}\right)}\right]^2}, \quad -\infty < x < \infty \qquad -\infty < \alpha < \infty \qquad \beta > 0$$



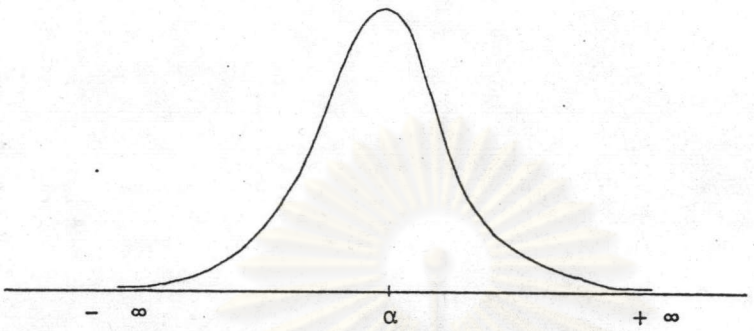

ค่าคาดหวัง $E(x) = \alpha$

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

8.3 <u>การแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียล</u> (Double Exponential Distribution)

$$f(x) = \frac{1}{2\beta} \cdot e^{-\frac{|x-\alpha|}{\beta}}, \quad -\infty < x < \infty, \quad -\infty < \alpha < \infty, \quad \beta > 0$$

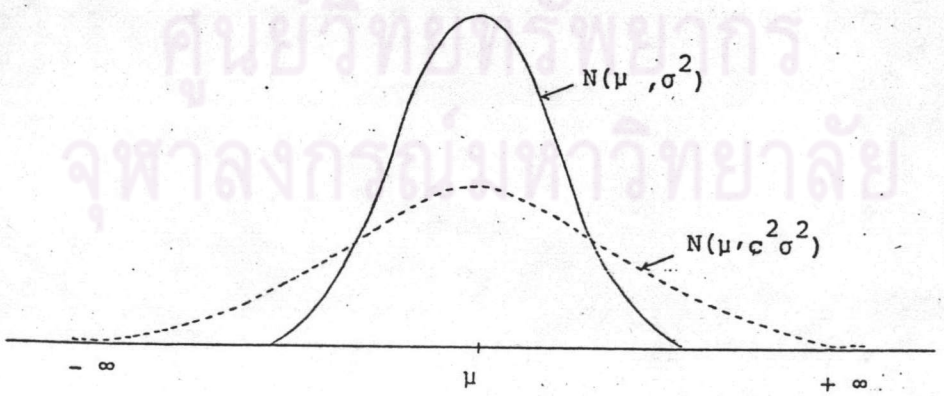
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย




ค่าคาดหวัง $E(x) = \alpha$

ค่าความแปรปรวน $V(x) = 2\beta^2$

8.4 <u>การแจกแจงแบบปกติปลอมปน</u> (Scale Contaminated Normal Distribution)




ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ลักษณะการแจกแจงแบบปกติปลอมปนที่พิจารณาในวิทยานิพนธ์นี้ เป็นการแจกแจงที่แปลงมาจากการแจกแจงแบบปกติ ซึ่งมีฟังก์ชันการแปลงดังนี้

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

หมายความว่าค่า x จะมาจากการแจกแจง $N(\mu , \sigma^2)$ ด้วยความน่าจะเป็น $(1 - p)$ และจากการแจกแจง $N(\mu , c^2\sigma^2)$ ด้วยความน่าจะเป็น p

$\mu$ และ $\sigma^2$ เป็นพารามิเตอร์ที่กำหนดค่าเฉลี่ย และค่าความแปรปรวนของความคลาดเคลื่อน

p และ c เป็นค่าที่กำหนดสัดส่วนการปลอมปนและสเกลแฟคเตอร์

9. การจำลองการทดลองจะกระทำซ้ำกัน 1,000 ครั้ง ในแต่ละสถานการณ์ของการทดลอง

## 1.6 ประโยชน์ที่คาดว่าจะได้รับ

เพื่อเป็นแนวทางให้ผู้วิจัยได้เลือกใช้วิธีวิเคราะห์ที่เหมาะสมเมื่อเกิดปัญหาข้อมูลสูญหายทั้งบล็อคในแผนการทดลองแบบบล็อคไม่สมบูรณ์สมดุลย์

## 1.7 คำจำกัดความ

1.7.1 ความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error;MSE) ความผันแปรระหว่างค่าสังเกตที่ได้รับทรีทเมนต์เดียวกัน

1.7.2 ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error) ค่าสัมบูรณ์ของค่าประมาณที่แตกต่างไปจากค่าจริงเฉลี่ย ในวิทยานิพนธ์ฉบับนี้นำค่า MSE มาเปรียบเทียบ ดังนั้นค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย จึงเป็นค่าสัมบูรณ์ของ MSE ของวิธีต่าง ๆ ทั้ง 3 วิธี ที่แตกต่างไปจากค่า MSE ของกรณีข้อมูลปกติเฉลี่ย

1.7.3 ความคลาดเคลื่อนประเภทที่ 1 (Type I Error) คือการปฏิเสธสมมติฐานว่าง (Null Hypothesis) เมื่อสมมติฐานว่างนั้นเป็นจริง

1.7.4 ความคลาดเคลื่อนประเภทที่ 2 (Type II Error) คือการยอมรับสมมติฐานว่าง เมื่อสมมติฐานว่างนั้นเป็นเท็จ

1.7.5 อำนาจการทดสอบ (Power of the test) หมายถึง ความน่าจะเป็น (Probability) ที่จะปฏิเสธสมมติฐานว่าง เมื่อสมมติฐานว่างนั้นเป็นเท็จจะมีค่าเท่ากับ $1 - \beta$ เมื่อ $\beta$ คือความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 2 (Probability of Type II Error) (Kirk, 1969:555, Minium 1978:364)

1.7.6 องศาแห่งความเป็นอิสระ (Degree of freedom ; df) หมายถึง จำนวนข้อมูลที่เป็นอิสระลบด้วยจำนวนพารามิเตอร์

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย