บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การวิเคราะห์ความแปรปรวนร่วมเป็นวิธีที่ผสมผสานระหว่างการวิเคราะห์ ความถดถอย และการวิเคราะห์ความแปรปรวน โดยการนำค่าตัวแปรร่วม X มาปรับค่าตัวแปรตาม Y วิธีการที่ใช้ในการประมาณพารามิเตอร์ของการวิเคราะห์ความแปรปรวนร่วม จะใช้วิธีกำลังสองน้อยที่สุด (ORDINARY LEAST SQUARE) เพราะจะทำให้ค่าที่ประมาณได้ใกล้ชิดกับค่าที่แท้จริงมากที่สุด แต่การประมาณพารามิเตอร์โดยวิธีกำลังสองน้อยที่สุด จะได้ค่าที่ถูกต้องแม่นยำหรือไม่ขึ้นอยู่กับว่าข้อมูลที่นำมาวิเคราะห์มีคุณสมบัติตรงตามข้อตกลงเบื้องต้นในการวิเคราะห์ความแปรปรวนร่วมหรือไม่

ข้อตกลงเบื้องต้นในการวิเคราะห์ความแปรปรวนร่วมของวิธีกำลังสองน้อยที่สุด คือ

1. การเลือกหน่วยทดลองจากประชากรต้องเป็นแบบสุ่ม และอิสระต่อกัน
2. การแจกแจงความคลาดเคลื่อนต้องมีการแจกแจงแบบปกติ
3. การถดถอยของตัวแปรตาม y ที่ขึ้นกับตัวแปรร่วม x ต้องเป็นเชิงเส้น

ในกรณีข้อมูลไม่เป็นไปตามข้อตกลงเบื้องต้นการใช้วิธีประมาณพารามิเตอร์โดยวิธีกำลังสองน้อยที่สุดจึงไม่เหมาะสม ประกอบกับได้มีผู้ทำการวิจัยศึกษา วิธีประมาณพารามิเตอร์ในการวิเคราะห์ความแปรปรวนร่วมเมื่อการแจกแจงของความคลาดเคลื่อนเป็นแบบหางยาวกว่าปกติไว้ดังนี้

- มนูญ ศรีวิรัตน์ ศึกษาการเปรียบเทียบการประมาณพารามิเตอร์ระหว่างวิธีกำลังสองน้อยที่สุด กับวิธีบูตสแตรป ผลสรุปโดยส่วนใหญ่วิธีบูตสแตรปเป็นตัวประมาณที่ดีกว่า
- กรรณิการ์ อุโฆษกุล ศึกษาเปรียบเทียบวิธีการประมาณพารามิเตอร์ระหว่างวิธีกำลังสองน้อยที่สุด กับวิธีตัวประมาณเอ็ม ผลสรุปโดยส่วนใหญ่วิธีตัวประมาณเอ็มเป็นตัวประมาณที่ดีกว่า

ด้วยเหตุนี้เองผู้วิจัยจึงสนใจการเปรียบเทียบการประมาณพารามิเตอร์ระหว่างวิธีบูตสแตรป และวิธีตัวประมาณเอ็ม ในกรณีการแจกแจงความคลาดเคลื่อนไม่เป็นไปตามข้อตกลง

เบื้องต้น โดยจะศึกษาเมื่อมีการแจกแจงแบบหางยาว ว่าวิธีใดจะให้ผลการวิเคราะห์ความแปรปรวน-ร่วมมีประสิทธิภาพมากกว่ากัน

วัตถุประสงค์การวิจัย

เพื่อเปรียบเทียบประสิทธิภาพวิธีการประมาณพารามิเตอร์ในการวิเคราะห์ความแปร--ปรวนร่วมระหว่างวิธีตัวประมาณเอ็ม และวิธีบูตสแตรป เมื่อการแจกแจงความคลาดเคลื่อนไม่เป็นแบบปกติ คือมีการแจกแจงแบบหางยาว โดยใช้เกณฑ์การตัดสินใจจากความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดประเภทที่ 1 และอำนาจการทดสอบ

สมมติฐานของการวิจัย

การวิเคราะห์ความแปรปรวนร่วมเมื่อการแจกแจงความคลาดเคลื่อนไม่เป็นแบบปกติคือมีการแจกแจงเป็นแบบหางยาว การประมาณพารามิเตอร์ด้วยวิธีตัวประมาณเอ็มจะให้อำนาจการทดสอบสูงกว่าวิธีบูตสแตรป

ข้อตกลงเบื้องต้น

ในการวิจัยจะพิจารณาอำนาจการทดสอบในกรณีที่สามารถควบคุมความน่าจะเป็นของความผิดพลาดประเภทที่1 ได้เท่านั้น โดยใช้เกณฑ์ของ BRADLEY
การวิจัยครั้งนี้พิจารณาตัวแบบ (Model)

$$y_{ij} = \mu + \tau_i + \sum_{k=1}^{q} \beta_k (x_{ijk} - \bar{x}_{..k}) + \varepsilon_{ij} \quad , \; i = 1, 2, 3, \ldots, p$$

$$j = 1, 2, 3, \ldots, n_i$$

$$k = 1, 2, 3, \ldots, q$$

โดยกำหนดให้

- p เป็นจำนวนวิธีปฏิบัติ(Treatment)
- n_i เป็นขนาดตัวอย่างในแต่ละวิธีปฏิบัติที่ i
- q เป็นจำนวนตัวแปรร่วม
- y_{ij} หมายถึง ค่าสังเกตของตัวแปรตามจากหน่วยทดลองที่ j วิธีปฏิบัติที่ i

μ หมายถึง ค่าเฉลี่ยทั้งหมดของ y_i

τ_i หมายถึง อิทธิพลของวิธีปฏิบัติที่ i

β_k หมายถึง ค่าสัมประสิทธิ์ความถดถอยของตัวแปรร่วมที่ k

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

ε_{ij} หมายถึง ค่าคลาดเคลื่อนของการทดลองจากหน่วยทดลองที่ j วิธีปฏิบัติที่ i

ขอบเขตของการวิจัย

1. ใช้ตัวแบบ (Model) เป็นอิทธิพลกำหนด (Fixed Effect) ในแผนการทดลองแบบสุ่มตลอด (Completely Randomized Design)

2 กำหนดตัวแปรร่วม x = 1, 3, 5

3. กำหนดจำนวนวิธีปฏิบัติ 3, 5, 7

4. กำหนดขนาดตัวอย่างแต่ละวิธีปฏิบัติ 5, 10, 20, และ 30

5. กำหนดให้ค่าเฉลี่ยของแต่ละกลุ่มเท่ากันคือ $\mu = 100$

6. ลักษณะการแจกแจงความคลาดเคลื่อนที่นำมาศึกษามีค่าเฉลี่ย (μ) = 0 และความแปรปรวน (σ^2) = 100

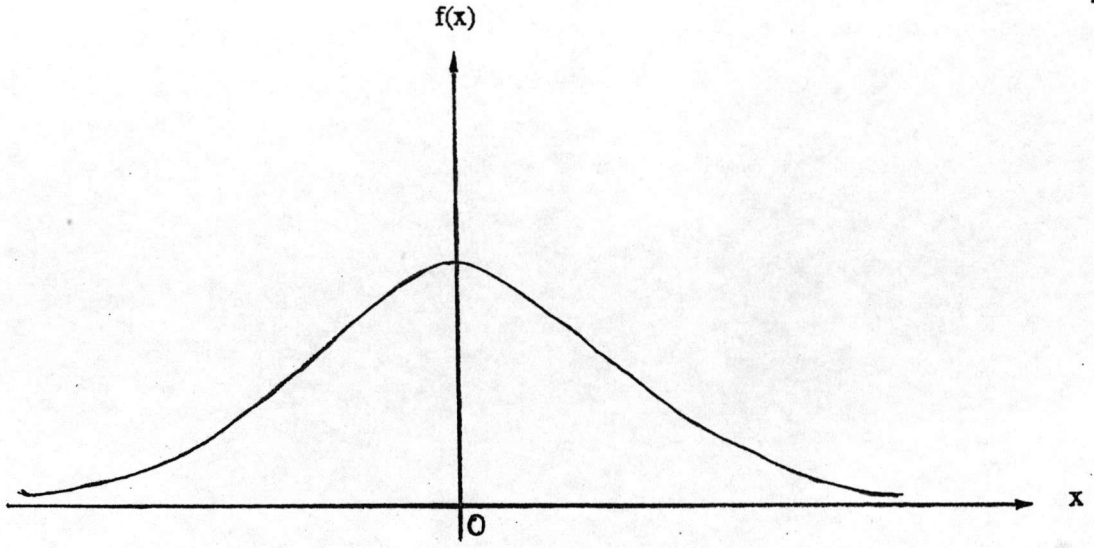
การแจกแจงที่นำมาศึกษามี 3 การแจกแจงมีดังนี้

ก. การแจกแจงแบบโลจิสติก (Logistic Distribution)

ฟังก์ชันความหนาแน่นอยู่ในรูป

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

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



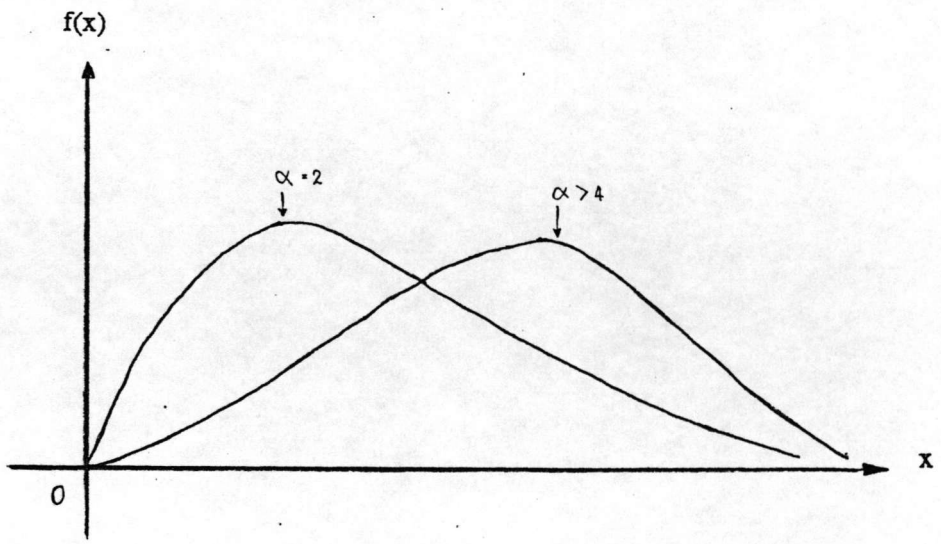
จะได้ว่าค่าคาดหวัง $E(x) = \alpha$ และความแปรปรวน $V(x) = \frac{1}{3}\pi^2\beta^2$

ข. การแจกแจงแบบแกมมา (Gamma Distribution)

ฟังก์ชันความหนาแน่นอยู่ในรูป

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \quad ; \quad \beta > 0$$

ผู้วิจัยสนใจศึกษาเมื่อ $\alpha > 4$ รูปการแจกแจงเป็นแบบเบ้ซ้าย และ $\alpha = 2$ รูปการแจกแจงเป็นแบบเบ้ขวาดังนี้



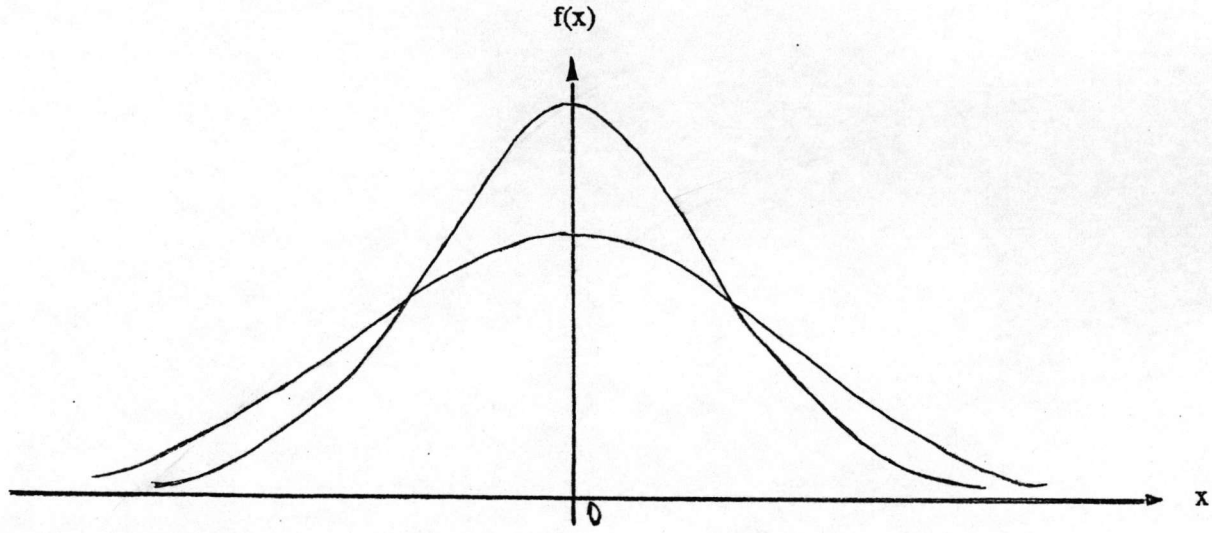
ค่าคาดหวัง $E(x) = \frac{\alpha}{\beta}$ ความแปรปรวน $V(x) = \frac{\alpha}{\beta^2}$

ค. การแจกแจงแบบปกติปลอมปน (Scale Contaminate Normal Distribution)

การแจกแจงแบบปกติปลอมปนเป็นการแจกแจงที่แปลงมาจากการแจกแจงแบบปกติ ซึ่งมีฟังก์ชันความน่าจะเป็น ดังนี้

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

โดยที่ x มาจากการแจกแจงแบบปกติที่มีค่าเฉลี่ย μ ความแปรปรวน σ^2 ด้วยความน่าจะเป็น 1-P และมาจากการแจกแจงแบบปกติที่มีค่าเฉลี่ย μ ความแปรปรวน $c^2\sigma^2$ ด้วยความน่าจะเป็น P สามารถเขียนรูปได้ดังนี้



เมื่อ C และ P เป็นค่าคงที่ C คือสเกลแฟคเตอร์ (Scale Factor) ในการวิจัยครั้งนี้จะใช้ C = 3, 5, 10, 20 และ P คือ เปอร์เซนต์การปลอมปน (Percent of Contamination) โดยใช้ P = 5, 10, 15, 20, 25

7. ระดับนัยสำคัญที่ใช้ในการวิจัยใช้ 2 ระดับ คือ α = 0.01 และ 0.05

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

1. ศึกษาวิธีการประมาณพารามิเตอร์ในการวิเคราะห์ความแปรปรวนร่วม เขียนโปรแกรมจำลองข้อมูลให้มีการแจกแจงตามฟังก์ชั่นที่กำหนด 3 กรณีคือ การแจกแจงแบบโลจิสติก การแจกแจงแบบแกมมา และการแจกแจงแบบปกติปลอมปนโดยใช้เทคนิคมอนติคาร์โล (Monticarlo Technique) ทำการทดลองซ้ำ ๆ กัน 1,000 ครั้งในแต่ละสถานการณ์ จากนั้นนำข้อมูลที่ได้ไปประมาณหาค่าพารามิเตอร์ในการวิเคราะห์ความแปรปรวนร่วม 2 วิธีดังนี้

ก) วิธีตัวประมาณเอ็ม

ข) วิธีบูตสแตรป

2. ศึกษาเปรียบเทียบความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดประเภทที่ 1 และอำนาจการทดสอบ ของวิธีประมาณพารามิเตอร์ในการวิเคราะห์ความแปรปรวนร่วมระหว่างวิธีบูตสแตรปและวิธีตัวประมาณเอ็ม

ประโยชน์ที่คาดว่าจะได้รับ

เพื่อเป็นแนวทางในการเลือกใช้วิธีการประมาณค่าพารามิเตอร์ในการวิเคราะห์ความแปรปรวนร่วมเมื่อลักษณะการแจกแจงความคลาดเคลื่อนไม่เป็นแบบปกติ

ความหมายของคำต่างๆ ที่ใช้ในการวิจัย

1. การแจกแจงแบบหางยาว คือการแจกแจงที่ข้อมูลมีการกระจายไปสู่หางมากกว่าปกติ มีผลทำให้ฟังก์ชั่นมีค่าเข้าใกล้ศูนย์ (0) ในอัตราที่ช้า

2. วิธีปฏิบัติ(Treatment) คือวิธีการหนึ่ง ๆ ที่ใช้ในการทดลอง

3. ตัวแปรร่วม หรือตัวแปรอิสระ คือค่าของข้อมูลที่วัดได้ก่อนที่หน่วยทดลองจะได้รับอิทธิพลของวิธีปฏิบัติ

4. ตัวแปรตาม คือค่าของข้อมูลที่วัดได้หลังจากได้รับอิทธิพลของตัวแปรอิสระและวิธีปฏิบัติแล้ว

5. ความผิดพลาดประเภทที่ 1 (Type I Error) คือความผิดพลาดที่เกิดจากการปฏิเสธสมมติฐานว่าง (H_0) เมื่อสมมติฐานว่างเป็นจริง

6. ความผิดพลาดประเภทที่ 2 (Type II Error) คือความผิดพลาดที่เกิดจากการยอมรับสมมติฐานว่าง (H_0) เมื่อสมมติฐานว่างเป็นเท็จ

7. อำนาจการทดสอบ (Power of the Test) คือความน่าจะเป็นของการปฏิเสธสมมติ-ฐานว่าง (H_0) เมื่อสมมติฐานว่าง เป็นเท็จ