บทที่ 3

วิธีการดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการศึกษาเปรียบเทียบวิธีการประมาณพารามิเตอร์ในการวิเคราะห์ความแปรปรวนร่วม เมื่อความผิดพลาดมีการแจกแจงแบบหางยาว ระหว่างวิธีกำลังสองน้อยที่สุด (least square method) กับวิธีของตัวประมาณเอ็ม (M-estimator) ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay และตัวประมาณสเกลความคลาดเคลื่อนแบบ median absolute - deviation โดยจะศึกษาประสิทธิภาพของการวิเคราะห์ด้วยการพิจารณาความสามารถในการในการควบคุมความน่าจะเป็นของความผิดพลาดประเภทที่ 1 (P(Type I Error)) และเปรียบเทียบอำนาจการทดสอบ สำหรับการวิจัยครั้งนี้จะใช้วิธีจำลองสถานการณ์ขึ้นมา โดยใช้เทคนิคมอนติคาร์โลชิมูเลชันโดยใช้โปรแกรมภาษาฟอร์แทรน 77 (FORTRAN 77) ด้วยเครื่องคอมพิวเตอร์ AMDAHL 5860 ซึ่งหลักการของวิธีมอนติคาร์โลชิมูเลชันเป็นการนำตัวเลขสุ่ม (random number) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษาตามขั้นตอนดังนี้

1. การสร้างตัวเลขสุ่ม ลักษณะของตัวเลขสุ่มจะมีการแจกแจงแบบสม่ำเสมอ (uniform distribution)ในช่วง(0,1) และเป็นอิสระต่อกัน

2. การประยุกต์ปัญหาที่ต้องการศึกษามาใช้กับตัวเลขสุ่ม ซึ่งในขั้นตอนนี้จะขึ้นอยู่กับลักษณะของปัญหา

3. การทดลองกระทำ เมื่อประยุกต์ปัญหาให้ใช้กับตัวเลขสุ่มได้แล้วก็ทำการทดลองซ้ำๆกัน (replication) เพื่อหาคำตอบของปัญหา

สำหรับการวิจัยครั้งนี้ได้กำหนดแผนการดำเนินการ, วิธีการและขั้นตอนการวิจัยดังต่อไปนี้

3.1 แผนการดำเนินการวิจัย

3.1.1 กำหนดให้ความคลาดเคลื่อนมีการแจกแจงแบบหางยาวมีค่าเฉลี่ย = 0 ความแปรปรวน = 625 โดยใช้การแจกแจงแบบโลจิสติค (logistic distribution) และการ

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

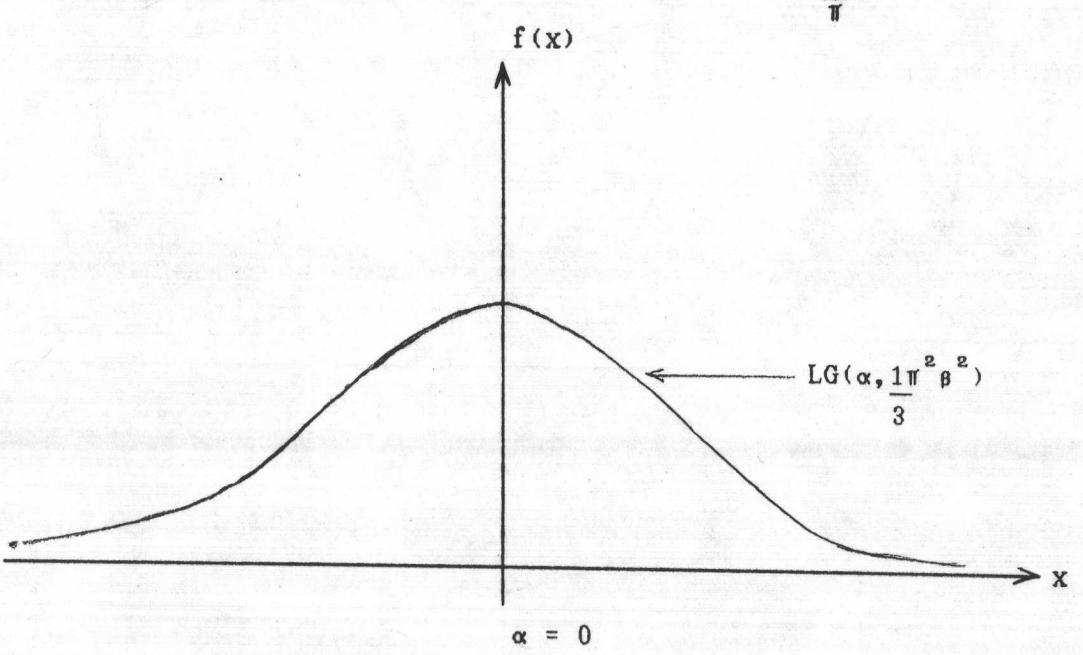
แจกแจงแบบปกติปลอมปน (scale-contaminated normal distribution)

ก. <u>การแจกแจงแบบโลจิสติค</u>

ฟังก์ชันความน่าจะเป็นอยู่ในรูปของ

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

การวิจัยครั้งนี้สนใจศึกษาเมื่อพารามิเตอร์ $\alpha = 0$ และ $\beta = \frac{\sqrt{3}\,\sigma}{\pi}$



<u>รูปที่ 3.1</u> แสดงเส้นโค้งของการแจกแจงแบบโลจิสติค ซึ่งใช้ $\alpha = 0$ และ $\beta = \frac{\sqrt{3}\,\sigma}{\pi}$

โปรแกรมย่อยที่ใช้คือ LOGIS ซึ่งแสดงไว้ในภาคผนวก ข และผลลัพธ์ที่ได้จากโปรแกรมคือค่า EX โดยการใช้คำสั่ง CALL LOGIS(AMEANO,SDO,E(I,J)) เมื่อ AMEANO เป็นค่าเฉลี่ย และ SDO เป็น ความแปรปรวน

ข. การแจกแจงแบบปกติปลอมปน

ฟังก์ชันความน่าจะเป็นอยู่ในรูปของ

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

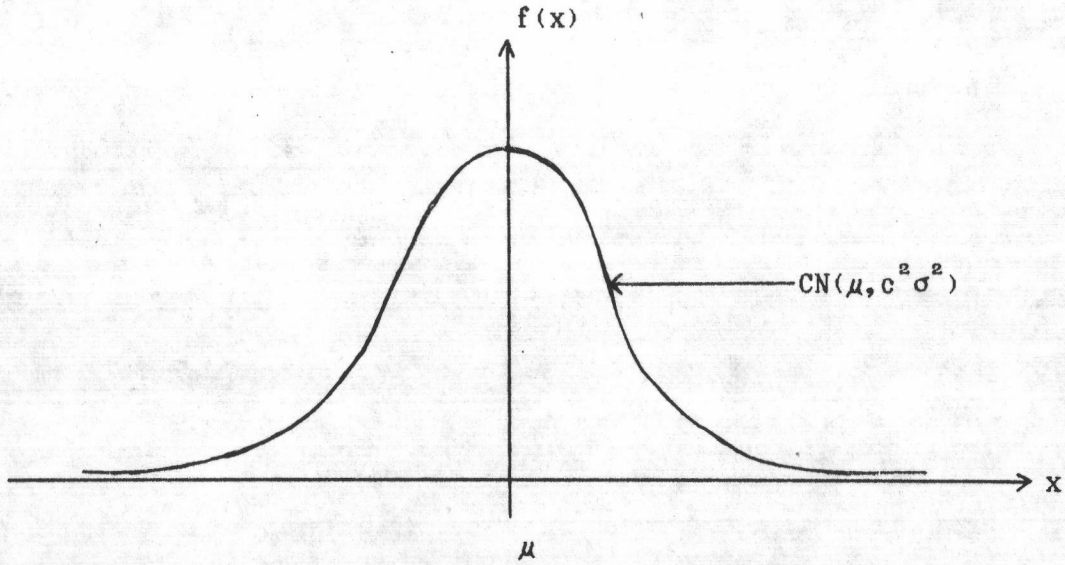
เป็นการแจกแจงซึ่งสร้างขึ้นเพื่อให้มีลักษณะหางยาว โดยสร้างเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอในช่วง 0 ถึง 1 ถ้าหากตัวเลขสุ่มที่ได้มีค่าอยู่ในช่วง 0 ถึง 1-p จะสร้างตัวแปรสุ่มที่มีการแจกแจงแบบปกติมีค่าเฉลี่ย μ และความแปรปรวน σ^2 แต่ถ้าตัวแปรสุ่มมีค่าอยู่ในช่วง 1-p ถึง 1 จะสร้างตัวแปรสุ่มที่มีการแจกแจงแบบปกติมีค่าเฉลี่ย μ และความแปรปรวน $c^2\sigma^2$ ซึ่ง p คือเปอร์เซ็นต์การปลอมปน และ c คือสเกลแฟคเตอร์

ในการทำวิจัยครั้งนี้จะกำหนดเปอร์เซนต์การปลอมปน และสเกลแฟคเตอร์ต่างๆเพื่อให้มีรูปร่างของการแจกแจงต่างๆกันดังแสดงในตารางที่ 3.1

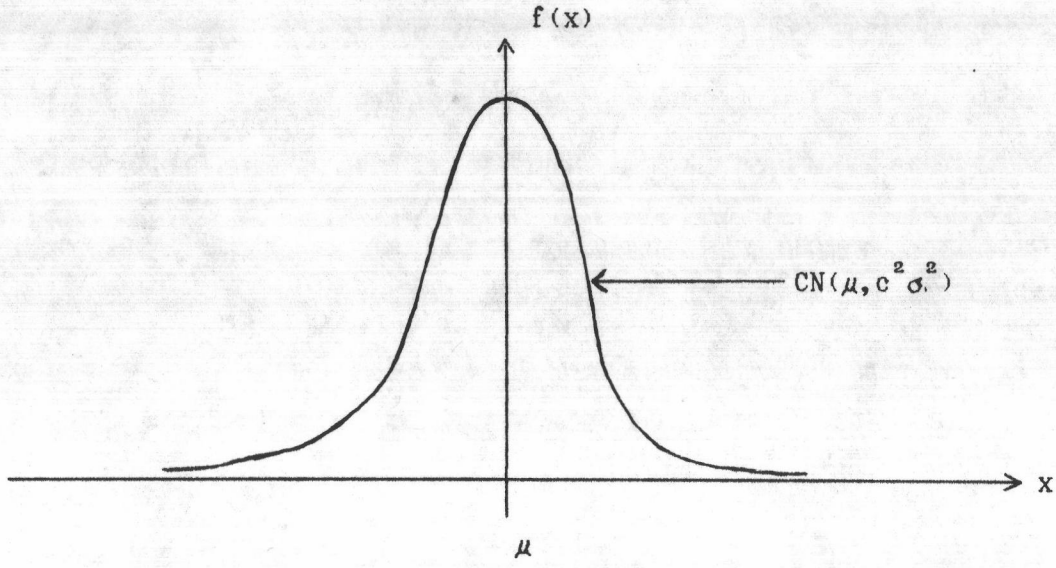
ตารางที่ 3.1 แสดงค่าสเกลแฟคเตอร์และเปอร์เซนต์การปลอมปน

c \ p	5	10	20	30
5	(5, 5)	(5, 10)	(5, 20)	(5, 30)
10	(10, 5)	(10, 10)	(10, 20)	(10, 30)
15	(15, 5)	(15, 10)	(15, 20)	(15, 30)

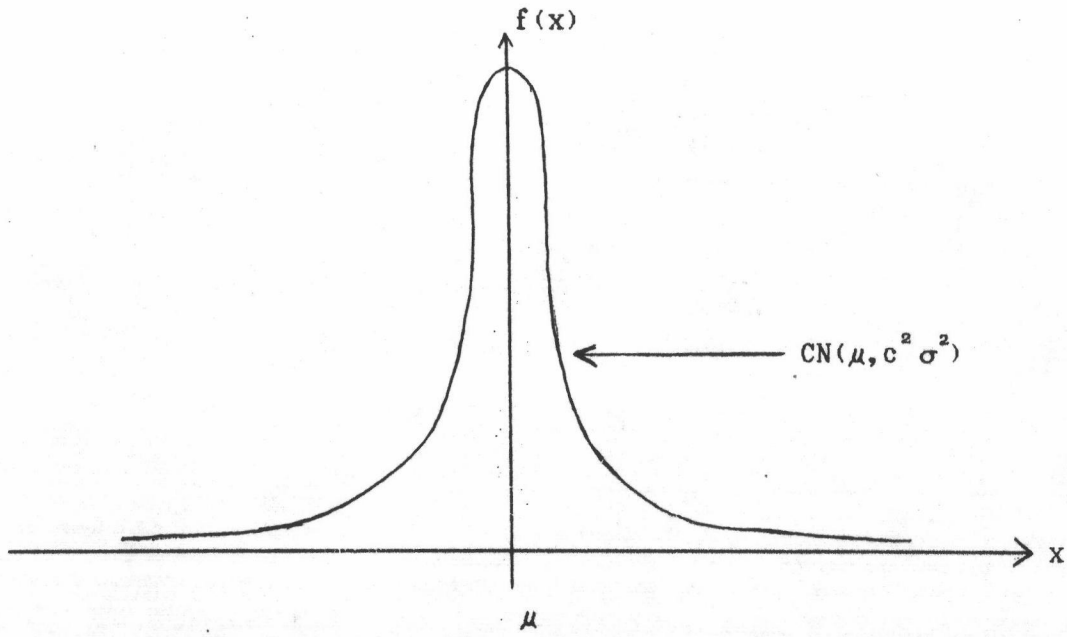
โปรแกรมย่อยที่ใช้คือ SCAL ซึ่งแสดงไว้ในภาคผนวก ข โดยใช้คำสั่ง CALL SCNML(CS,PS,AMEANO,SDO,E(I,J)) สร้างโดยการใช้วิธีแปลงข้อมูลจากการแจกแจงปกติ ผลลัพธ์ E(I,J) ได้มาจากการใช้คำสั่ง CALL NORMAL(DMEAN,SD,EX2) ด้วยความน่าจะเป็น (1-ps) เมื่อ DMEAN, SD, PS เป็นค่าเฉลี่ย ความแปรปรวน และสัดส่วนของการปลอมปน ตามลำดับและได้มาจากการใช้คำสั่ง CALL NORMAL(DMEAN,CSD, EX1) ด้วยความน่าจะเป็น ps เมื่อ CSD เป็นผลคูณระหว่าง CS กับ SD โดยที่ CS เป็นสเกลแฟคเตอร์



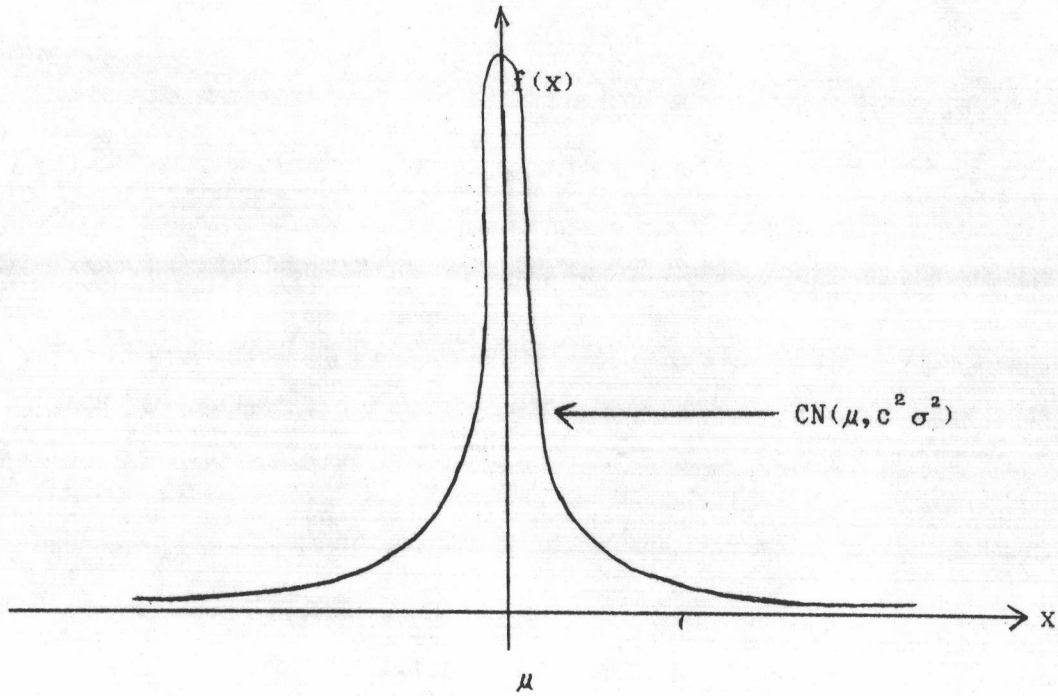
รูปที่ 3.2 เส้นโค้งของการแจกแจงแบบปกติปลอมปน ซึ่งใช้สเกลแฟคเตอร์เท่ากับ 5 เปอร์เซนต์ของการปลอมปนเท่ากับ 5



รูปที่ 3.3 เส้นโค้งของการแจกแจงแบบปกติปลอมปน ซึ่งใช้สเกลแฟคเตอร์เท่ากับ 5 เปอร์เซนต์ของการปลอมปนเท่ากับ 30



รูปที่ 3.4 เส้นโค้งของการแจกแจงแบบปกติปลอมปน ซึ่งใช้สเกลแฟคเตอร์เท่ากับ 15 เปอร์เซนต์ของการปลอมปนเท่ากับ 5



รูปที่ 3.5 เส้นโค้งของการแจกแจงแบบปกติปลอมปน ซึ่งใช้สเกลแฟคเตอร์เท่ากับ 15 เปอร์เซนต์ของการปลอมปนเท่ากับ 30

3.1.2 กำหนดขนาดตัวอย่างในแต่ละวิธีปฏิบัติ (treatment)

ก. จำนวนวิธีปฏิบัติเท่ากับ 3, 5 และ 7 จะใช้จำนวนตัวแปรร่วมเท่ากับ 2 และขนาดตัวอย่างเท่ากับ 5 และ 15

ข. จำนวนวิธีปฏิบัติเท่ากับ 3, 5 และ 7 จะใช้จำนวนตัวแปรร่วมเท่ากับ 5 และขนาดตัวอย่างเท่ากับ 10 และ 20

3.1.3 การหาค่าอำนาจการทดสอบจะทำที่ระยะห่างของค่าเฉลี่ยของประชากร 2 ชุด (δ) เท่ากับ $\max\left|\mu_i-\mu_j\right|$ ซึ่ง $i \neq j$ เป็นจำนวนเท่าต่างๆของความแปรปรวน นั่นคือ ทำการศึกษาที่ δ เท่ากับ 2σ

3.2 ขั้นตอนการดำเนินการวิจัย

การดำเนินการวิจัยมีขั้นตอนที่สำคัญเรียงตามลำดับดังต่อไปนี้

ก. <u>การสร้างข้อมูล</u>

จะสร้างตัวแปรตาม (y) และตัวแปรร่วม (x) เมื่อการแจกแจงของความคลาดเคลื่อนเป็นแบบหางยาวกว่าการแจกแจงแบบปกติ โดยตัวแปรตาม (y) สร้างมาจากตัวแปรร่วม (x), อิทธิพลของวิธีปฏิบัติ (treatment effect) และค่าความผิดพลาดที่มีลักษณะการแจกแจงตามที่ต้องการศึกษา โดยให้ตัวแปรตามมีความสัมพันธ์เชิงเส้น ตามตัวแบบดังนี้

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

1. สร้างค่าความคลาดเคลื่อน (ε_{ij}) ให้มีการแจกแจงตามที่กำหนด
2. สร้างตัวแปรร่วม (x) ให้มีการแจกแจงแบบปกติ มีค่าเฉลี่ย = μ และความแปรปรวน = σ^2
3. ทำการปรับค่าตัวแปรร่วม (x) โดยการลบด้วยค่าเฉลี่ย
4. กำหนดพารามิเตอร์ μ และ β ซึ่งเป็นค่าคงที่ขึ้นมา
5. สร้างอิทธิพลของวิธีปฏิบัติ โดยพิจารณาจาก $\sum_{j=1}^{p} \tau_j = 0$

ข. นำข้อมูลที่ได้มาสร้างตัวแบบใหม่ให้อยู่ในรูปของสมการถดถอย โดยอาศัยตัวแปร dummy เป็นตัวบ่งชี้อิทธิพลของวิธีปฏิบัติดังนี้

$$D_{i1} = \begin{cases} 1 & \text{ถ้าเป็นวิธีปฏิบัติที่ 1} \\ 0 & \text{อื่นๆ} \end{cases}$$

$$D_{i2} = \begin{cases} 1 & \text{ถ้าเป็นวิธีปฏิบัติที่ 2} \\ 0 & \text{อื่นๆ} \end{cases}$$

.
.
.

$$D_{ip-1} = \begin{cases} 1 & \text{ถ้าเป็นวิธีปฏิบัติที่ p-1} \\ 0 & \text{อื่นๆ} \end{cases}$$

ตัวแบบเต็มรูป

$$y_{ij} = \alpha_0 + \alpha_1 D_{i1} + \alpha_2 D_{i2} + \ldots + \alpha_{p-1} D_{ip-1} + \sum_{k=1}^{t} \beta_k (x_{ijk} - \bar{x}_{..k}) + \varepsilon_{ij}$$

ตัวแบบลดรูป

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

ค. ประมาณสัมประสิทธิ์การถดถอยพหุของตัวแบบเต็มรูป และตัวแบบลดรูป ด้วยวิธีกำลังสองน้อยที่สุด , วิธีของตัวประมาณเอ็ม เมื่อใช้เกณฑ์ความแกร่งของ Ramsay และสเกลความคลาดเคลื่อนแบบ (MAD) median absolute deviation

ง. คำนวณค่าผลบวกกำลังสองของความผิดพลาดทั้งตัวแบบเต็มรูปและตัวแบบลดรูป

ของทั้งสองวิธี เพื่อนำมาหาค่าตัวสถิติ F

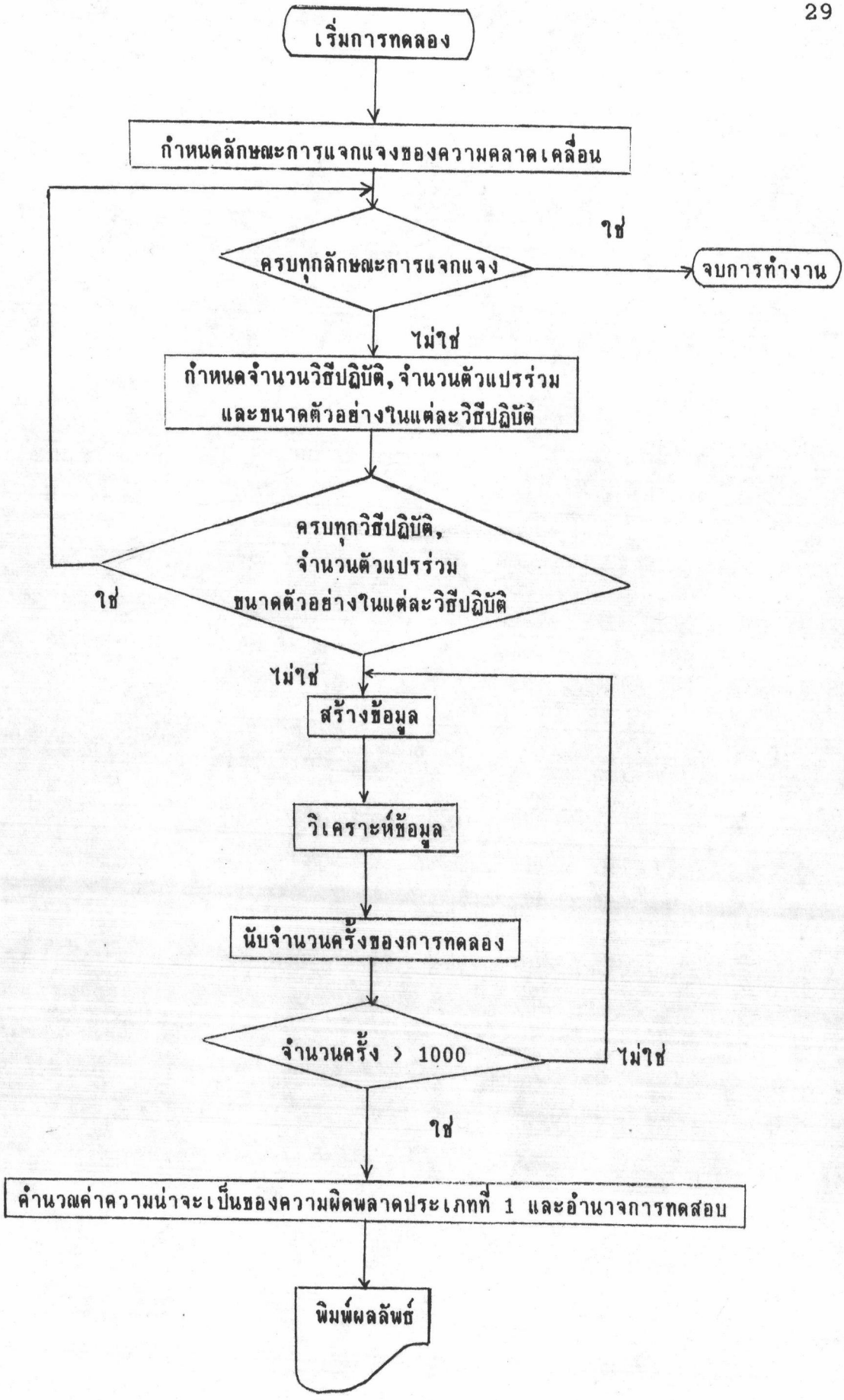
จ. เปรียบเทียบค่าตัวสถิติทดสอบที่ได้ทั้ง 2 วิธี กับค่าวิกฤต (critical value) ซึ่งได้จากตารางเอฟ ณ ระดับนัยสำคัญ 2 ระดับคือ $\alpha = 0.01$ และ 0.05 โดยจะปฏิเสธสมมุติฐานเมื่อค่าตัวสถิติทดสอบมากกว่าค่าวิกฤต

ฉ. กระทำการทดลองจะทำซ้ำๆ กันไม่น้อยกว่า 1000 ครั้ง ต่อ 1 สถานการณ์ที่กำหนดขึ้นจากการจัดหมู่ (combination) ของลักษณะการแจกแจง, จำนวนวิธีปฏิบัติ, จำนวนตัวแปรอิสระ และขนาดตัวอย่าง และทำการนับจำนวนครั้งของการปฏิเสธสมมุติฐาน

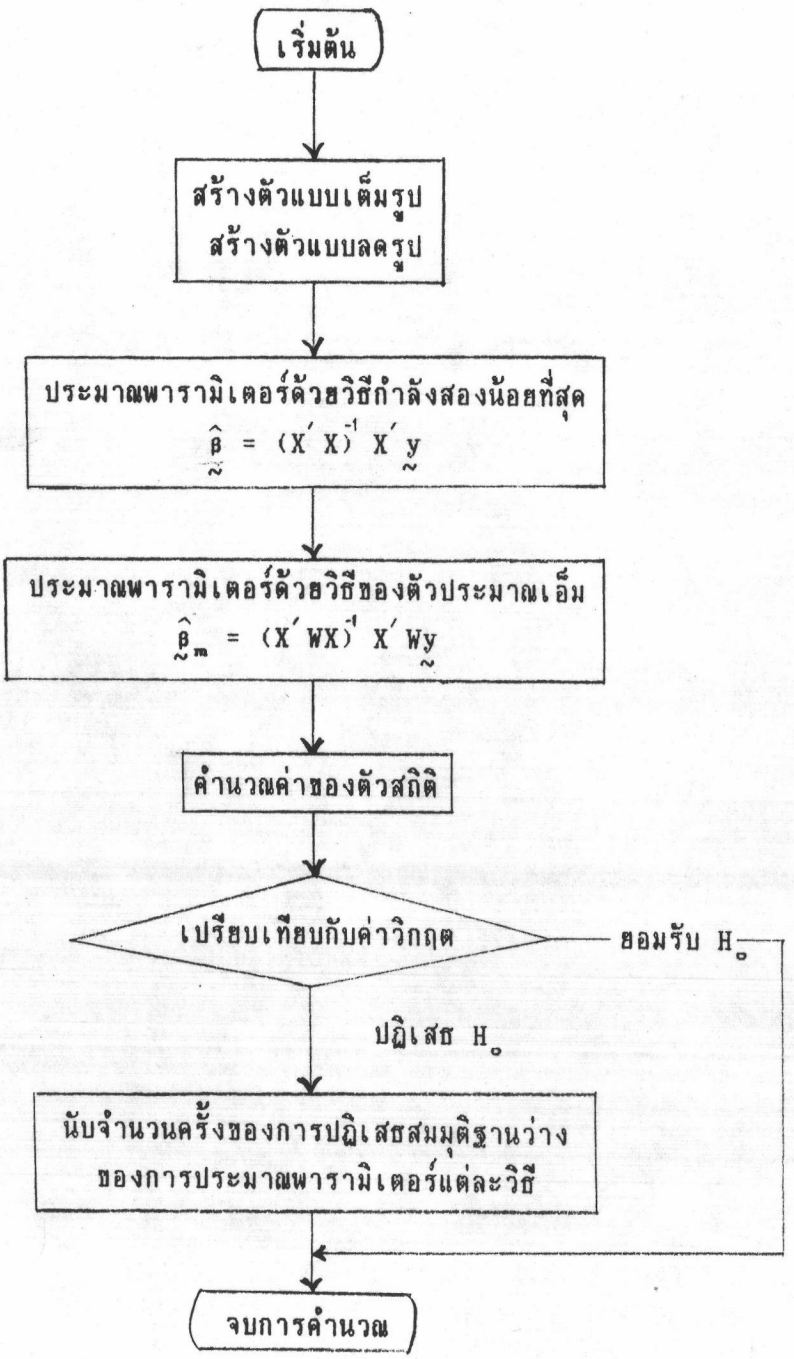
1. การหาความน่าจะเป็นของความผิดพลาดประเภทที่ 1 จะทำในกรณีที่อิทธิพลของวิธีปฏิบัติเป็น 0 นั่นคือ $\tau_1 = \tau_2 = \ldots\ldots = \tau_p = 0$ โดยคำนวณจากจำนวนครั้งของการปฏิเสธสมมุติฐานหารด้วยจำนวนครั้งที่ใช้ทดสอบ

2. คำนวณหาอำนาจการทดสอบ ทำเฉพาะกรณีที่ตัวสถิติที่ได้จากการประมาณพารามิเตอร์ทั้งสองวิธีสามารถควบคุมความน่าจะเป็นของความผิดพลาดประเภทที่ 1 ได้เท่านั้น โดยสร้างข้อมูลให้อิทธิพลของวิธีปฏิบัติ (τ_j) มีค่าอื่นๆ และ $\sum_{j=1}^{p} \tau_j = 0$

ขั้นตอนในการทำงานของโปรแกรมฟอร์แทรนที่ใช้ในการประมาณพารามิเตอร์ การคำนวณความน่าจะเป็นของความผิดพลาดประเภทที่ 1 และการหาค่าอำนาจการทดสอบสามารถเขียนให้อยู่ในรูปแผนผังการทำงานได้ดังนี้


เริ่มการทดลอง
กำหนดลักษณะการแจกแจงของความคลาดเคลื่อน
ครบทุกลักษณะการแจกแจง
ใช่
จบการทำงาน
ไม่ใช่
กำหนดจำนวนวิธีปฏิบัติ, จำนวนตัวแปรร่วม
และขนาดตัวอย่างในแต่ละวิธีปฏิบัติ
ครบทุกวิธีปฏิบัติ,
จำนวนตัวแปรร่วม
ขนาดตัวอย่างในแต่ละวิธีปฏิบัติ
ใช่
ไม่ใช่
สร้างข้อมูล
วิเคราะห์ข้อมูล
นับจำนวนครั้งของการทดลอง
จำนวนครั้ง > 1000
ไม่ใช่
ใช่
คำนวณค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1 และอำนาจการทดสอบ
พิมพ์ผลลัพธ์

แผนผังโปรแกรมการวิเคราะห์ข้อมูลซึ่งแสดงการประมาณค่าพารามิเตอร์ในการวิเคราะห์ความแปรปรวนร่วมโดยวิธีกำลังสองน้อยที่สุดและวิธีของตัวประมาณเอ็มและทดสอบสมมุติฐาน ณ ระดับนัยสำคัญ α = .01 และ .05



แผนผังโปรแกรมการสร้างข้อมูล