บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ทำการศึกษาเปรียบเทียบประสิทธิภาพของตัวสถิติที่ใช้สำหรับการคัดกรองการแจกแจงแบบโลจิสติกจากการแจกแจงแบบปกติ ศึกษาเปรียบเทียบคุณสมบัติทั่วไปในเรื่องของความสัมพันธ์กับการแจกแจงแบบอื่นๆ ของการแจกแจงแบบโลจิสติก และการแจกแจงแบบปกติ และศึกษาผลกระทบของการประมาณค่าสัมประสิทธิ์การถดถอย เมื่อความคลาดเคลื่อนมีการแจกแจงแบบโลจิสติก โดยใช้โปรแกรม R เวอร์ชัน 2.9.2 ในการจำลองข้อมูล ในบทนี้จะกล่าวถึงแผนการดำเนินการวิจัย ขั้นตอนในการดำเนินการวิจัย และขั้นตอนการทำงานของโปรแกรม ซึ่งมีรายละเอียด ดังนี้

3.1 แผนการดำเนินการวิจัย

การวิจัยครั้งนี้ ได้กำหนดสถานการณ์ต่างๆในการศึกษาแต่ละส่วน ดังนี้

1. ตัวสถิติที่นำมาใช้เปรียบเทียบประสิทธิภาพการคัดกรองการแจกแจงแบบโลจิสติกจากการแจกแจงแบบปกติ มี 6 ตัวสถิติ ดังนี้
   - ตัวสถิติ Kolmogorov-Smirnov
   - ตัวสถิติ Shapiro Wilk
   - ตัวสถิติ Anderson Darling
   - ตัวสถิติ Lilliefors
   - ตัวสถิติ Cramer Von Mises
   - ตัวสถิติ Chi-square
2. ในการประมาณค่าสัมประสิทธิ์การถดถอย กำหนดให้ตัวแปรอิสระ (X) มีการแจกแจงแบบ $U(-100,100)$, $N(0,1111.56)$ และ $Exp(0.023)$
3. ในการประมาณค่าสัมประสิทธิ์การถดถอย ความคลาดเคลื่อน (ε) มีการแจกแจงแบบ $L\left(0,\sqrt{3}/\pi\right)$
4. กำหนดค่า β_i เริ่มต้นในการประมาณค่าสัมประสิทธิ์การถดถอย ดังนี้

   กรณีการถดถอยอย่างง่าย $\beta_0 = 1$ และ $\beta_1 = 0.5, 1, 2$

   กรณีการถดถอยเชิงพหุ $\beta_0 = 1$, $\beta_1 = 1, 2$ และ $\beta_2 = 1$
5. ในกรณีตัวแปรอิสระ X_1, X_2 ไม่อิสระ กำหนดให้มีความสัมพันธ์กัน 10%, 30% และ 50%

6. ระดับนัยสำคัญ (α) ที่ใช้ในการทดสอบสมมติฐาน คือ 0.01, 0.05, 0.10
7. ขนาดตัวอย่าง (n_j) ที่ใช้ในศึกษาเท่ากับ 10, 25, 50 และ 100
8. การจำลองข้อมูลในแต่ละสถานการณ์จะกระทำซ้ำ 5000 รอบ

3.2 ขั้นตอนในการดำเนินการวิจัย

การดำเนินการวิจัย แบ่งการศึกษาออกเป็น 3 ส่วน คือ การเปรียบเทียบประสิทธิภาพของตัวสถิติที่ใช้สำหรับการคัดกรองการแจกแจงแบบโลจิสติกจากการแจกแจงแบบปกติ การศึกษาเปรียบเทียบคุณสมบัติทั่วไปในเรื่องของความสัมพันธ์กับการแจกแจงแบบอื่นๆ ของการแจกแจงแบบโลจิสติก และการแจกแจงแบบปกติ และการศึกษาผลกระทบของการประมาณค่าสัมประสิทธิ์การถดถอย เมื่อความคลาดเคลื่อนมีการแจกแจงแบบโลจิสติก ซึ่งในแต่ละส่วนมีขั้นตอนในการดำเนินการวิจัย ดังนี้

1. การทดสอบความสามารถของตัวสถิติที่ใช้คัดกรองข้อมูลที่มีการแจกแจงแบบโลจิสติกจากการแจกแจงแบบปกติ
   1.1 สร้างข้อมูลที่มีการแจกแจง $N(0,1)$ และ $L\left(0,\sqrt{3}/\pi\right)$ ให้มีจำนวนตามขนาดตัวอย่าง (n_j) ที่กำหนด
   1.2 คำนวณค่าตัวสถิติที่ใช้สำหรับคัดกรองการแจกแจงแบบโลจิสติกจากการแจกแจงแบบปกติทั้ง 6 ตัวสถิติ
   1.3 นำค่าที่คำนวณได้ มาทดสอบสมมติฐานว่าข้อมูลมีการแจกแจงแบบปกติหรือไม่
   1.4 นับจำนวนครั้งของการปฏิเสธสมมติฐานของแต่ละการทดสอบ
   1.5 ทำซ้ำข้อ 1.1 – 1.4 จำนวน 5,000 รอบ
   1.6 หาค่าความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่ 1 (α) และหาค่ากำลังการทดสอบ ดังนี้

   สมมติฐานการทดสอบ H_0 : ข้อมูลมีการแจกแจงแบบปกติ
   H_1 : ข้อมูลไม่ได้มีการแจกแจงแบบปกติ

   ความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่ 1 (α)
   = P (ปฏิเสธ H_0 | H_0 จริง)
   = (จำนวนครั้งของการปฏิเสธ H_0 เมื่อ H_0เป็นจริง) / 5,000

   กำลังการทดสอบ (1 - β)
   = P(ปฏิเสธ H_0 | H_0 เท็จ)

= (จำนวนครั้งของการปฏิเสธ H_0 เมื่อ H_0เป็นเท็จ) / 5,000

2. การศึกษา และเปรียบเทียบคุณสมบัติทั่วไปของการแจกแจงแบบปกติ และการแจกแจงแบบโลจิสติก ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1

2.1 ทดสอบว่าการแจกแจงแบบโลจิสติกยกกำลังสองมีการแจกแจงไคสแควร์หรือไม่

2.1.1 สร้างข้อมูลที่มีการแจกแจง $L(0,\sqrt{3}/\pi)$ ให้มีจำนวนตามขนาดตัวอย่าง (n_j) ที่กำหนด

2.1.2 คำนวณผลรวมของข้อมูลยกกำลังสอง $\left(\sum_{i=1}^{n} X_i^2\right)$

2.1.3 นำค่าที่คำนวณได้ มาทดสอบสมมติฐานว่าข้อมูลมีการแจกแจงแบบไคสแควร์หรือไม่

2.1.4 นับจำนวนครั้งของการปฏิเสธสมมติฐานของแต่ละการทดสอบ

2.1.5 ทำซ้ำข้อ 2.1.1 – 2.1.4 จำนวน 5,000 รอบ

2.1.6 หาค่ากำลังการทดสอบ ดังนี้

สมมติฐานการทดสอบ H_0 : ข้อมูลมีการแจกแจงแบบไคสแควร์ ที่ df = n

H_1 : ข้อมูลไม่ได้มีการแจกแจงแบบไคสแควร์ ที่ df = n

กำลังการทดสอบ (1 - β)

= P(ปฏิเสธ H_0 | H_0 เท็จ)

= (จำนวนครั้งของการปฏิเสธ H_0 เมื่อ H_0เป็นเท็จ) / 5,000

2.2 ทดสอบว่าข้อมูลที่มีการแจกแจงแบบโลจิสติกเมื่อนำค่าเฉลี่ยหารด้วยส่วนเบี่ยงเบนมาตรฐานส่วนรากที่สองของจำนวนข้อมูลมีการแจกแจงแบบทีหรือไม่

2.2.1 สร้างข้อมูลที่มีการแจกแจง $L(0,\sqrt{3}/\pi)$ ให้มีจำนวนตามขนาดตัวอย่าง (n_j) ที่กำหนด

2.2.2 คำนวณค่า $\frac{\bar{X}}{s/\sqrt{n_j}}$

2.2.3 นำค่าที่คำนวณได้ มาทดสอบสมมติฐานว่าข้อมูลมีการแจกแจงแบบทีหรือไม่

2.2.4 นับจำนวนครั้งของการปฏิเสธสมมติฐานของแต่ละการทดสอบ

2.2.5 ทำซ้ำข้อ 2.2.1 – 2.2.4 จำนวน 5,000 รอบ

2.2.6 หาค่ากำลังการทดสอบ ดังนี้

สมมติฐานการทดสอบ H_0 : ข้อมูลมีการแจกแจงแบบที ที่ df = n_j-1

H_1 : ข้อมูลไม่ได้มีการแจกแจงแบบที ที่ df = n_j-1

กำลังการทดสอบ (1 - β)

= P(ปฏิเสธ H_0 | H_0 เท็จ)

= (จำนวนครั้งของการปฏิเสธ H_0 เมื่อ H_0เป็นเท็จ) / 5,000

2.3 ทดสอบว่าข้อมูลที่มีการแจกแจงแบบโลจิสติกเมื่อนำค่าความแปรปรวนของตัวอย่าง 2 ชุด มาหารกัน มีการแจกแจงแบบเอฟหรือไม่ กรณีข้อมูลเริ่มต้นมี 2 ชุด

2.3.1 สร้างข้อมูลที่มีการแจกแจง $N(0,1)$ 2 ชุด กำหนดเป็น ชุด A, B และ $L(0,\sqrt{3}/\pi)$ 2 ชุด กำหนดเป็น ชุด C, D ให้มีจำนวนตามขนาดตัวอย่าง (n_j) ที่กำหนด

2.3.2 คำนวณค่า $\frac{s_A^2}{s_D^2}$, $\frac{s_C^2}{s_B^2}$, $\frac{s_C^2}{s_D^2}$

2.3.3 นำค่าที่คำนวณได้ มาทดสอบสมมติฐานว่าข้อมูลมีการแจกแจงแบบเอฟหรือไม่

2.3.4 นับจำนวนครั้งของการปฏิเสธสมมติฐานของแต่ละการทดสอบ

2.3.5 ทำซ้ำข้อ 2.3.1 – 2.3.4 จำนวน 5,000 รอบ

2.3.6 หาค่ากำลังการทดสอบ ดังนี้

สมมติฐานการทดสอบ

H_0 : ข้อมูลมีการแจกแจงแบบเอฟ ที่ $df1 = n_1 - 1, \quad df2 = n_2 - 1$

H_1 : ข้อมูลไม่ได้มีการแจกแจงแบบเอฟ ที่ $df1 = n_1 - 1, \quad df2 = n_2 - 1$

กำลังการทดสอบ (1 - β)

= P(ปฏิเสธ H_0 | H_0 เท็จ)

= (จำนวนครั้งของการปฏิเสธ H_0 เมื่อ H_0เป็นเท็จ) / 5,000

2.4 ทดสอบว่าข้อมูลที่มีการแจกแจงแบบโลจิสติกเมื่อนำค่าความแปรปรวนของตัวอย่าง 2 ชุด มาหารกัน มีการแจกแจงแบบเอฟหรือไม่ กรณีข้อมูลเริ่มต้นจากข้อมูลชุดเดียว

2.4.1 สร้างข้อมูลที่มีการแจกแจง $L(0,\sqrt{3}/\pi)$ ให้มีจำนวนตามขนาดตัวอย่าง (n_j) ที่กำหนด

2.4.2 นำข้อมูลที่สร้างขึ้นมาแบ่งเป็น 2 ชุดเท่า ๆ กัน

2.4.3 คำนวณค่า $\frac{s_1^2}{s_2^2}$ เก็บค่าไว้

2.4.4 ทำซ้ำข้อ 2.4.2 – 2.4.3 จำนวน n_j รอบ

2.4.5 นำค่าข้อมูล $\frac{s_1^2}{s_2^2}$ ที่เก็บไว้ n_j ค่า มาทดสอบสมมติฐานว่าข้อมูลมีการแจกแจงแบบเอฟหรือไม่

2.4.6 นับจำนวนครั้งของการปฏิเสธสมมติฐานของแต่ละการทดสอบ

2.4.7 ทำซ้ำข้อ 2.3.1 – 2.3.6 จำนวน 5,000 รอบ

2.4.8 หาค่ากำลังการทดสอบ ดังนี้

สมมติฐานการทดสอบ

H_0 : ข้อมูลมีการแจกแจงแบบเอฟ ที่ $df1 = n_1 - 1, \quad df2 = n_2 - 1$

H_1 : ข้อมูลไม่ได้มีการแจกแจงแบบเอฟ ที่ $df1 = n_1 - 1, \quad df2 = n_2 - 1$

กำลังการทดสอบ (1 - β)

= P(ปฏิเสธ H_0 | H_0 เท็จ)

= (จำนวนครั้งของการปฏิเสธ H_0 เมื่อ H_0เป็นเท็จ) / 5,000

3. การศึกษาผลกระทบของการประมาณค่าสัมประสิทธิ์การถดถอย เมื่อความคลาดเคลื่อนมีการแจกแจงแบบโลจิสติก

3.1 ทดลองประมาณค่าสัมประสิทธิ์การถดถอยในกรณีต่างๆ โดยความคลาดเคลื่อนได้จากการสร้างข้อมูลโดยตรง

3.1.1 กรณีการถดถอยอย่างง่าย

3.1.1.1 สร้างความคลาดเคลื่อนที่มีการแจกแจง $L(0, \sqrt{3}/\pi)$ และสร้างตัวแปรอิสระ (X) ให้มีการแจกแจงและขนาดตัวอย่าง (n_j) ที่กำหนด

3.1.1.2 คำนวณค่าตัวแปรตาม (Y) จากสมการ $Y = \beta_0 + \beta_1 X + \varepsilon$

3.1.1.3 ประมาณค่า β_0, β_1 ด้วยวิธี OLS แล้วแทนด้วยสัญลักษณ์ b_0, b_1 ตามลำดับ ทำการทดสอบว่า $\beta_i = \beta_{i0}$ โดยที่ β_{i0} เป็นค่าตั้งต้นของ β_i ที่กำหนดไว้ (ภายใต้เงื่อนไข ε มีการแจกแจงแบบปกติ)

3.1.1.4 นับจำนวนครั้งของการปฏิเสธสมมติฐาน

3.1.1.5 ทำซ้ำข้อ 3.1.1.1 – 3.1.1.4 จำนวน 5,000 รอบ

3.1.1.6 หาค่าความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่ 1 ดังนี้

สมมติฐานการทดสอบ

$H_0 : \beta_i = \beta_{i0}$

$H_1 : \beta_i \neq \beta_{i0}$ โดยที่ β_{i0} เป็นค่าตั้งต้นของ β_i ที่กำหนดไว้

ความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่ 1 (α)

= P (ปฏิเสธ H_0 | H_0 จริง)

= (จำนวนครั้งของการปฏิเสธ H_0 เมื่อ H_0เป็นจริง) / 5,000

3.1.1.7 ทำการทดสอบว่า β_i ที่เก็บค่าไว้มีการแจกแจงแบบปกติหลายตัวแปรหรือไม่ โดยทดสอบว่าผลรวมเชิงเส้น $\underset{\sim}{C}_k'\underset{\sim}{\beta}$ มีการแจกแจงแบบปกติหรือไม่ เมื่อ $\underset{\sim}{C}_1=\begin{bmatrix}1\\1\end{bmatrix}$, $\underset{\sim}{C}_2=\begin{bmatrix}1\\10\end{bmatrix}$ และ $\underset{\sim}{C}_3=\begin{bmatrix}10\\1\end{bmatrix}$

3.1.2 กรณีการถดถอยเชิงพหุ ตัวแปรอิสระ (X_1, X_2) ไม่มีความสัมพันธ์กัน

3.1.2.1 สร้างความคลาดเคลื่อนที่มีการแจกแจง $L\left(0,\sqrt{3}/\pi\right)$ และสร้างตัวแปรอิสระ (X_1, X_2) ให้มีการแจกแจงและขนาดตัวอย่าง (n_j) ที่กำหนด

3.1.2.2 คำนวณค่าตัวแปรตาม (Y) จากสมการ $Y=\beta_0+\beta_1X_1+\beta_2X_2+\varepsilon$

3.1.2.3 ประมาณค่า β_0, β_1, β_2 ด้วยวิธี OLS แล้วแทนด้วยสัญลักษณ์ b_0, b_1, b_2 ตามลำดับ ทำการทดสอบว่า $\beta_i = \beta_{i0}$ โดยที่ β_{i0} เป็นค่าตั้งต้นของ β_i ที่กำหนดไว้ (ภายใต้เงื่อนไข ε มีการแจกแจงแบบปกติ)

3.1.2.4 นับจำนวนครั้งของการปฏิเสธสมมติฐาน

3.1.2.5 ทำซ้ำข้อ 3.1.2.1 – 3.1.2.4 จำนวน 5,000 รอบ

3.1.2.6 หาค่าความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่ 1 ดังนี้

สมมติฐานการทดสอบ

$H_0 : \beta_i = \beta_{i0}$

$H_1 : \beta_i \neq \beta_{i0}$ โดยที่ β_{i0} เป็นค่าตั้งต้นของ β_i ที่กำหนดไว้

ความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่ 1 (α)

= P (ปฏิเสธ H_0 | H_0 จริง)

= (จำนวนครั้งของการปฏิเสธ H_0 เมื่อ H_0เป็นจริง) / 5,000

3.1.2.7 ทำการทดสอบว่า β_i ที่เก็บค่าไว้มีการแจกแจงแบบปกติหลายตัวแปรหรือไม่ โดยทดสอบว่าผลรวมเชิงเส้น $\underset{\sim}{C}_k'\underset{\sim}{\beta}$ มีการแจกแจงแบบปกติหรือไม่ เมื่อ $\underset{\sim}{C}_1=\begin{bmatrix}1\\1\\1\end{bmatrix}$, $\underset{\sim}{C}_2=\begin{bmatrix}0.1\\1\\10\end{bmatrix}$ และ $\underset{\sim}{C}_3=\begin{bmatrix}10\\1\\0.1\end{bmatrix}$

3.1.3 กรณีการถดถอยเชิงพหุ ตัวแปรอิสระ (X_1, X_2) มีความสัมพันธ์กัน

3.1.3.1 สร้างความคลาดเคลื่อนที่มีการแจกแจง $L\left(0,\sqrt{3}/\pi\right)$

3.1.3.2 สร้าง Z_1, Z_2, Z_3 ให้มีการแจกแจง และพารามิเตอร์ที่ทำให้ X_1, X_2 สัมพันธ์กัน 10%, 30% และ 50% ขนาดตัวอย่าง (n_j) เมื่อกำหนด $X_1 = Z_1 + Z_2$ และ $X_2 = Z_2 + Z_3$

3.1.3.3 คำนวณค่าตัวแปรตาม (Y) จากสมการ $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

3.1.3.4 ประมาณค่า β_0, β_1, β_2 ด้วยวิธี OLS แล้วแทนด้วยสัญลักษณ์ b_0, b_1, b_2 ตามลำดับ ทำการทดสอบว่า $\beta_i = \beta_{i0}$ โดยที่ β_{i0} เป็นค่าตั้งต้นของ β_i ที่กำหนดไว้ (ภายใต้เงื่อนไข ε มีการแจกแจงแบบปกติ)

3.1.3.5 นับจำนวนครั้งของการปฏิเสธสมมติฐาน

3.1.3.6 ทำซ้ำข้อ 3.1.3.1 – 3.1.3.5 จำนวน 5,000 รอบ

3.1.3.7 หาค่าความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่ 1 ดังนี้

สมมติฐานการทดสอบ

$H_0 : \beta_i = \beta_{i0}$

$H_1 : \beta_i \neq \beta_{i0}$ โดยที่ β_{i0} เป็นค่าตั้งต้นของ β_i ที่กำหนดไว้

ความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่ 1 (α)

= P (ปฏิเสธ H_0 | H_0 จริง)

= (จำนวนครั้งของการปฏิเสธ H_0 เมื่อ H_0เป็นจริง) / 5,000

3.1.3.8 ทำการทดสอบว่า β_i ที่เก็บค่าไว้มีการแจกแจงแบบปกติหลายตัวแปรหรือไม่ โดยทดสอบว่าผลรวมเชิงเส้น $\underset{\sim}{C}_k'\underset{\sim}{\beta}$ มีการแจกแจงแบบปกติหรือไม่ เมื่อ $\underset{\sim}{C}_1 = \begin{bmatrix}1\\1\\1\end{bmatrix}$, $\underset{\sim}{C}_2 = \begin{bmatrix}0.1\\1\\10\end{bmatrix}$ และ $\underset{\sim}{C}_3 = \begin{bmatrix}10\\1\\0.1\end{bmatrix}$

3.2 ทดลองประมาณค่าสัมประสิทธิ์การถดถอยในกรณีต่างๆ โดยคัดกรองความคลาดเคลื่อนจากการแจกแจงแบบปกติก่อนนำไปใช้

3.2.1 กรณีการถดถอยอย่างง่าย

3.2.1.1 สร้างความคลาดเคลื่อนที่มีการแจกแจง $L\left(0,\sqrt{3}/\pi\right)$ นำไปคัดกรองโดยใช้ตัวสถิติ Shapiro Wilk และคัดกรองจากความสัมพันธ์กับตัวสถิติเอฟก่อนนำไปใช้

3.2.1.2 สร้างตัวแปรอิสระ (X) ให้มีการแจกแจงและขนาดตัวอย่าง (n_j) ที่กำหนด

3.2.1.3 คำนวณค่าตัวแปรตาม (Y) จากสมการ $Y = \beta_0 + \beta_1 X + \varepsilon$

สำนักงานคณะกรรมการวิจัยแห่งชาติ ห้องสมุดงานวิจัย

3.2.1.4 ประมาณค่า β_0, β_1 ด้วยวิธี OLS แล้วแทนด้วยสัญลักษณ์ b_0, b_1 ตามลำดับ ทำการทดสอบว่า $\beta_i = \beta_{i0}$ โดยที่ β_{i0} เป็นค่าตั้งต้นของ β_i ที่กำหนดไว้ (ภายใต้เงื่อนไข ε มีการแจกแจงแบบปกติ)

3.2.1.5 นับจำนวนครั้งของการปฏิเสธสมมติฐาน

3.2.1.6 ทำซ้ำข้อ 3.2.1.1 – 3.2.1.5 จำนวน 5,000 รอบ

3.2.1.7 หาค่าความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่ 1 ดังนี้

สมมติฐานการทดสอบ

$H_0 : \beta_i = \beta_{i0}$

$H_1 : \beta_i \neq \beta_{i0}$ โดยที่ β_{i0} เป็นค่าตั้งต้นของ β_i ที่กำหนดไว้

ความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่ 1 (α)

= P (ปฏิเสธ H_0 | H_0 จริง)

= (จำนวนครั้งของการปฏิเสธ H_0 เมื่อ H_0เป็นจริง) / 5,000

3.2.1.8 ทำการทดสอบว่า β_i ที่เก็บค่าไว้มีการแจกแจงแบบปกติหลายตัวแปรหรือไม่ โดยทดสอบว่าผลรวมเชิงเส้น $\underset{\sim}{C}_k'\underset{\sim}{\beta}$ มีการแจกแจงแบบปกติหรือไม่ เมื่อ $\underset{\sim}{C}_1 = \begin{bmatrix}1\\1\end{bmatrix}$, $\underset{\sim}{C}_2 = \begin{bmatrix}1\\10\end{bmatrix}$ และ $\underset{\sim}{C}_3 = \begin{bmatrix}10\\1\end{bmatrix}$

3.2.2 กรณีการถดถอยเชิงพหุ ตัวแปรอิสระ (X_1, X_2) ไม่มีความสัมพันธ์กัน

3.2.2.1 สร้างความคลาดเคลื่อนที่มีการแจกแจง $L\left(0,\sqrt{3}/\pi\right)$ นำไปคัดกรองโดยใช้ตัวสถิติ Shapiro Wilk และคัดกรองจากความสัมพันธ์กับตัวสถิติเอฟก่อนนำไปใช้

3.2.2.2 สร้างตัวแปรอิสระ (X_1, X_2) ให้มีการแจกแจงและขนาดตัวอย่าง (n_j) ที่กำหนด

3.2.2.3 คำนวณค่าตัวแปรตาม (Y) จากสมการ $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

3.2.2.4 ประมาณค่า β_0, β_1, β_2 ด้วยวิธี OLS แล้วแทนด้วยสัญลักษณ์ b_0, b_1, b_2 ตามลำดับ ทำการทดสอบว่า $\beta_i = \beta_{i0}$ โดยที่ β_{i0} เป็นค่าตั้งต้นของ β_i ที่กำหนดไว้ (ภายใต้เงื่อนไข ε มีการแจกแจงแบบปกติ)

3.2.2.5 นับจำนวนครั้งของการปฏิเสธสมมติฐาน

3.2.2.6 ทำซ้ำข้อ 3.2.2.1 – 3.2.2.5 จำนวน 5,000 รอบ

3.2.2.7 หาค่าความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่ 1 ดังนี้

สมมติฐานการทดสอบ

$H_0 : \beta_i = \beta_{i0}$

$H_1 : \beta_i \neq \beta_{i0}$ โดยที่ β_{i0} เป็นค่าตั้งต้นของ β_i ที่กำหนดไว้

ความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่ 1 (α)

= P (ปฏิเสธ H_0 | H_0 จริง)

= (จำนวนครั้งของการปฏิเสธ H_0 เมื่อ H_0เป็นจริง) / 5,000

3.2.2.8 ทำการทดสอบว่า β_i ที่เก็บค่าไว้มีการแจกแจงแบบปกติหลายตัวแปรหรือไม่ โดยทดสอบว่าผลรวมเชิงเส้น $\underset{\sim}{C}_k'\underset{\sim}{\beta}$ มีการแจกแจงแบบปกติหรือไม่ เมื่อ $\underset{\sim}{C}_1 = \begin{bmatrix}1\\1\\1\end{bmatrix}$, $\underset{\sim}{C}_2 = \begin{bmatrix}0.1\\1\\10\end{bmatrix}$ และ $\underset{\sim}{C}_3 = \begin{bmatrix}10\\1\\0.1\end{bmatrix}$

3.2.3 กรณีการถดถอยเชิงพหุ ตัวแปรอิสระ (X_1, X_2) มีความสัมพันธ์กัน

3.2.3.1 สร้างความคลาดเคลื่อนที่มีการแจกแจง $L(0,\sqrt{3}/\pi)$ นำไปคัดกรองโดยใช้ตัวสถิติ Shapiro Wilk และคัดกรองจากความสัมพันธ์กับตัวสถิติเอฟก่อนนำไปใช้

3.2.3.2 สร้าง Z_1 , Z_2 , Z_3 ให้มีการแจกแจง และพารามิเตอร์ที่ทำให้ X_1, X_2 สัมพันธ์กัน 10%, 30% และ 50% ขนาดตัวอย่าง (n_j) เมื่อกำหนด $X_1 = Z_1 + Z_2$ และ $X_2 = Z_2 + Z_3$

3.2.3.3 คำนวณค่าตัวแปรตาม (Y) จากสมการ $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

3.2.3.4 ประมาณค่า β_0, β_1, β_2 ด้วยวิธี OLS แล้วแทนด้วยสัญลักษณ์ b_0, b_1, b_2 ตามลำดับ ทำการทดสอบว่า $\beta_i = \beta_{i0}$ โดยที่ β_{i0} เป็นค่าตั้งต้นของ β_i ที่กำหนดไว้ (ภายใต้เงื่อนไข ε มีการแจกแจงแบบปกติ)

3.2.3.5 นับจำนวนครั้งของการปฏิเสธสมมติฐาน

3.2.3.6 ทำซ้ำข้อ 3.2.3.1 – 3.2.3.5 จำนวน 5,000 รอบ

3.2.3.7 หาค่าความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่ 1 ดังนี้

สมมติฐานการทดสอบ

$H_0 : \beta_i = \beta_{i0}$

$H_1 : \beta_i \neq \beta_{i0}$ โดยที่ β_{i0} เป็นค่าตั้งต้นของ β_i ที่กำหนดไว้

ความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่ 1 (α)

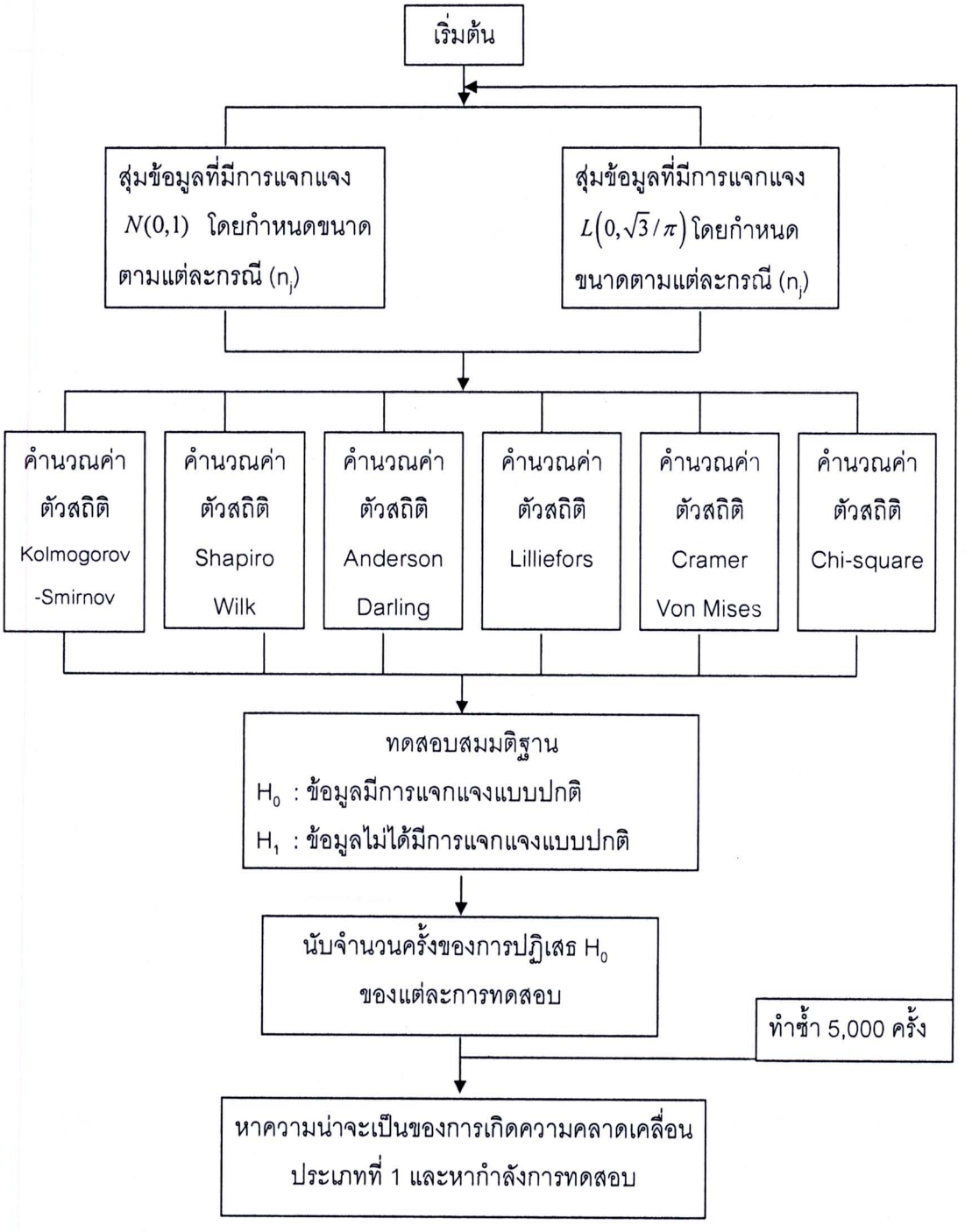
= P (ปฏิเสธ H_0 | H_0 จริง)

= (จำนวนครั้งของการปฏิเสธ H_0 เมื่อ H_0เป็นจริง) / 5,000

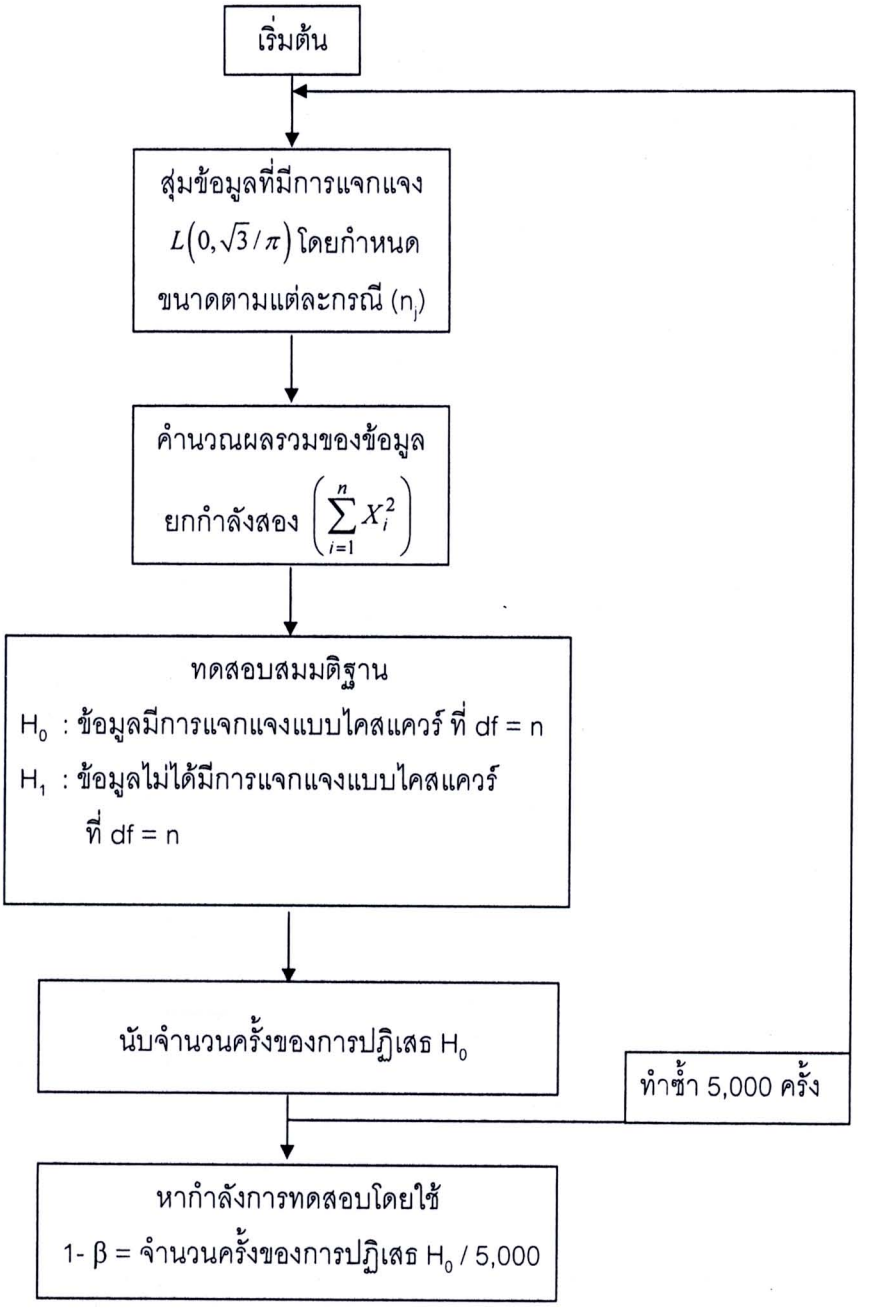
3.2.3.8 ทำการทดสอบว่า β_i ที่เก็บค่าไว้มีการแจกแจงแบบปกติหลายตัวแปร หรือไม่ โดยทดสอบว่าผลรวมเชิงเส้น $\underset{\sim}{C}_k'\underset{\sim}{\beta}$ มีการแจกแจงแบบปกติ หรือไม่ เมื่อ $\underset{\sim}{C}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\underset{\sim}{C}_2 = \begin{bmatrix} 0.1 \\ 1 \\ 10 \end{bmatrix}$ และ $\underset{\sim}{C}_3 = \begin{bmatrix} 10 \\ 1 \\ 0.1 \end{bmatrix}$

3.3 ขั้นตอนการทำงานของโปรแกรม

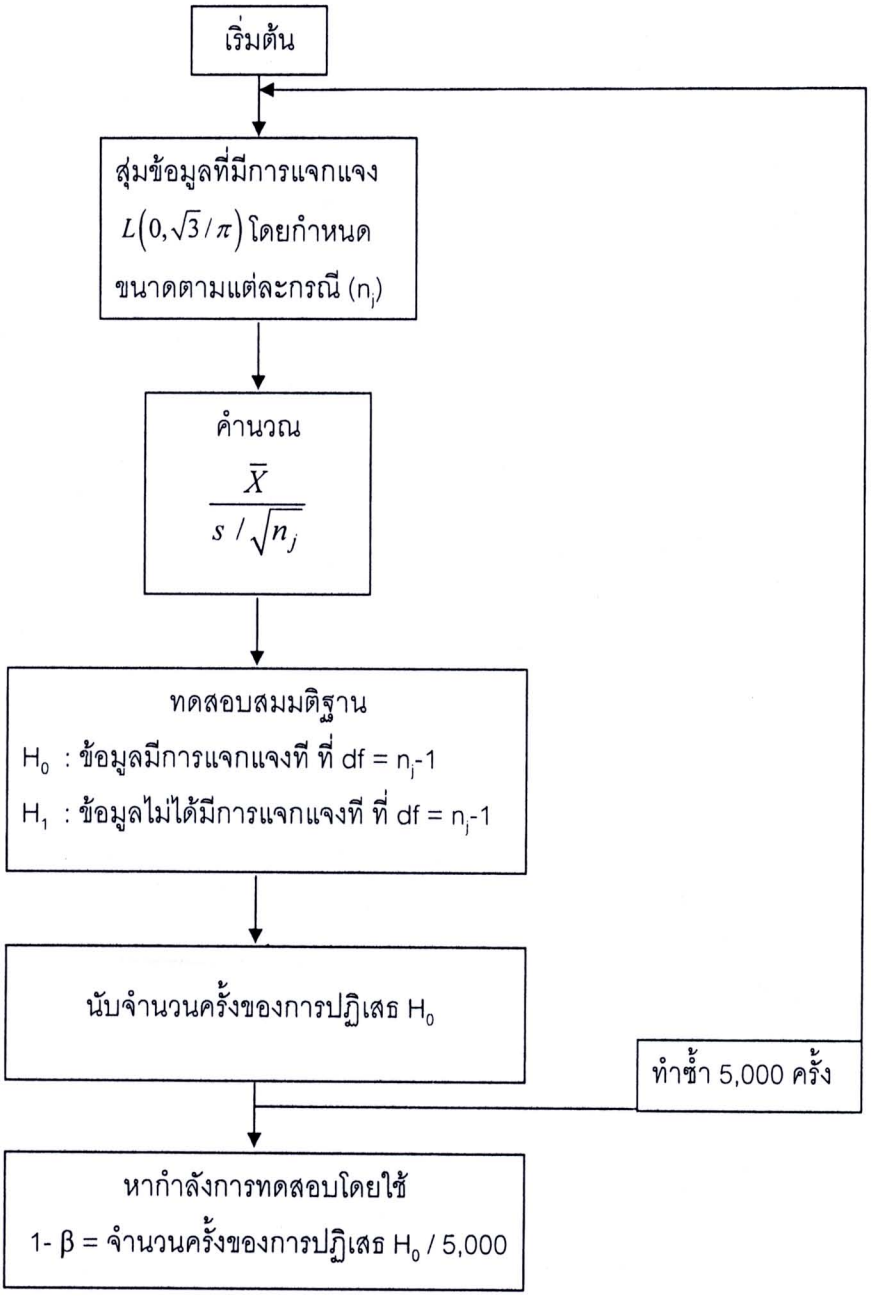
ภาพที่ 3.1 แสดงขั้นตอนการทำงานของโปรแกรมในการเปรียบเทียบประสิทธิภาพของตัวสถิติที่ใช้สำหรับการคัดกรองการแจกแจงแบบโลจิสติกจากการแจกแจงแบบปกติ



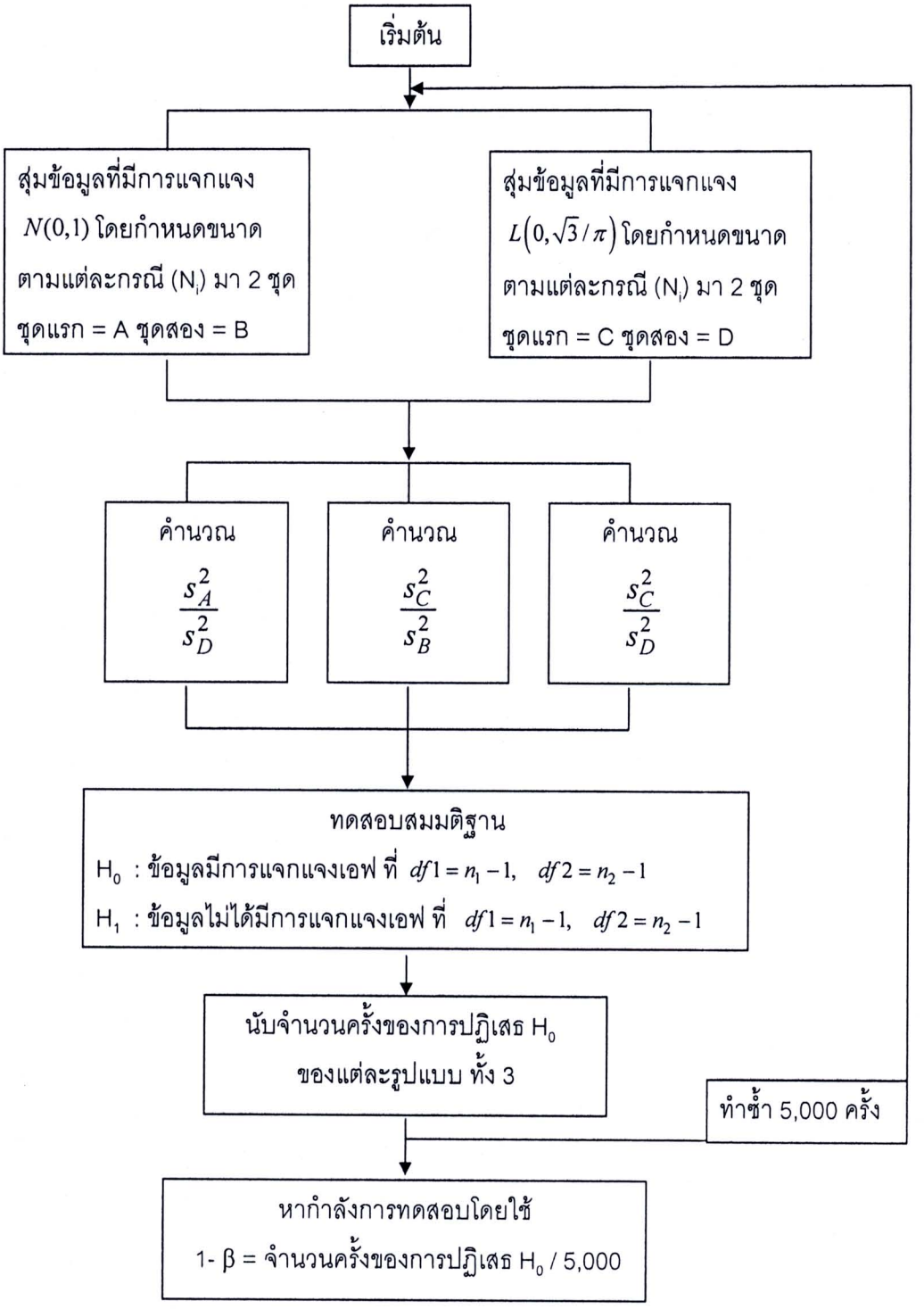
ภาพที่ 3.2 แสดงขั้นตอนการทำงานของโปรแกรมในการทดสอบความสัมพันธ์ของการแจกแจงแบบโลจิสติกกับการแจกแจงแบบไคสแควร์



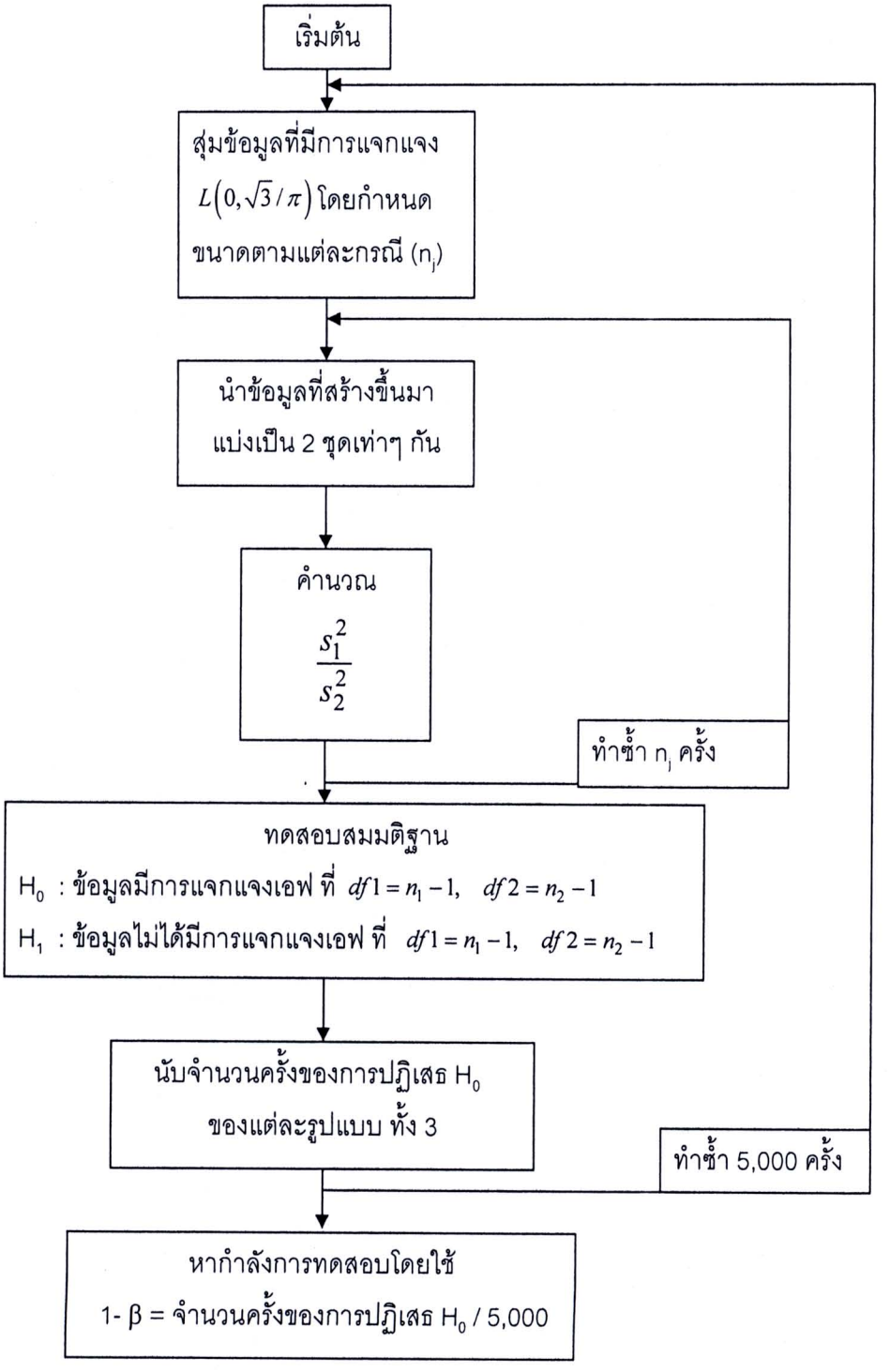
ภาพที่ 3.3 แสดงขั้นตอนการทำงานของโปรแกรมในการทดสอบความสัมพันธ์ของการแจกแจงแบบโลจิสติกกับการแจกแจงแบบที



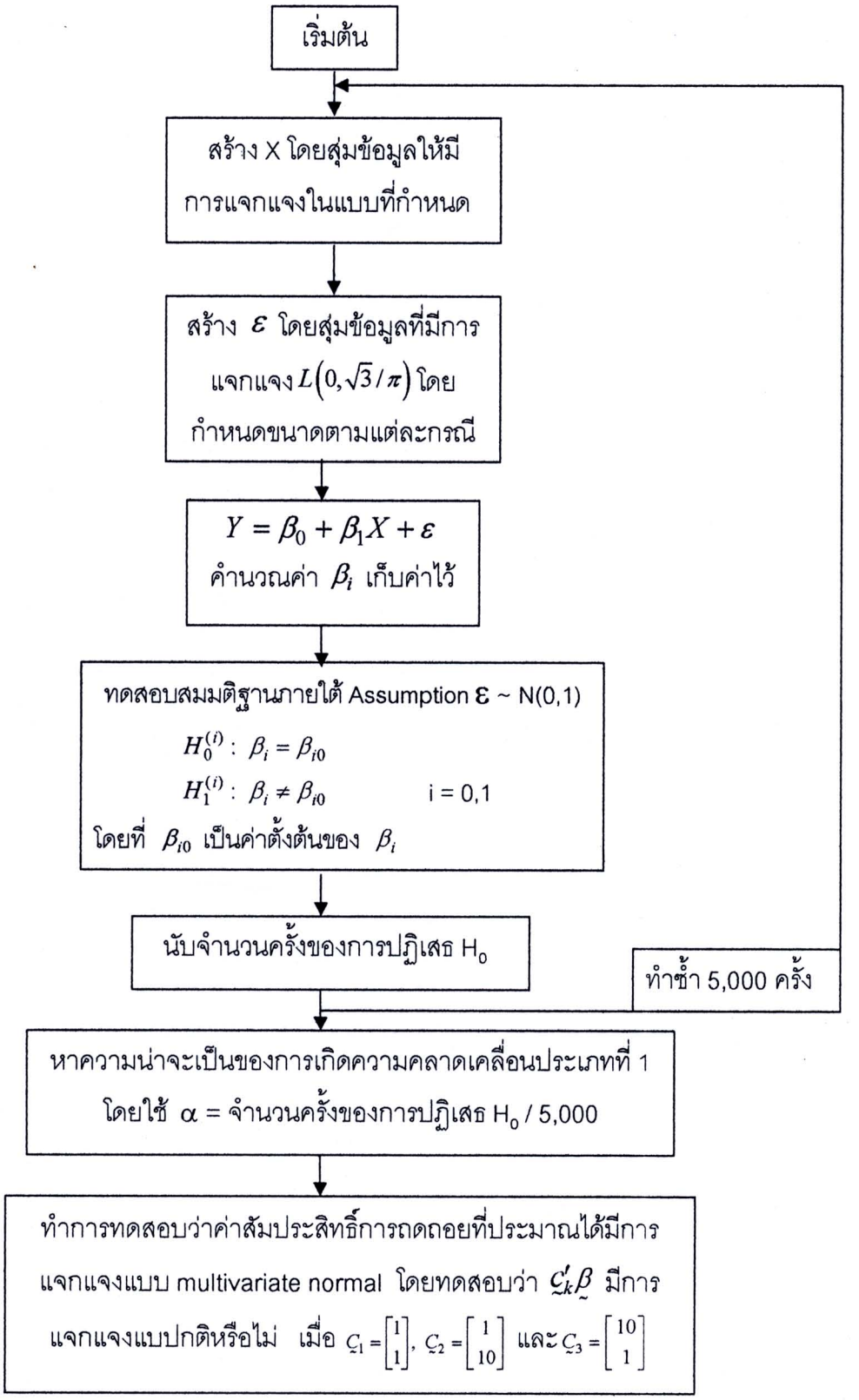
ภาพที่ 3.4 แสดงขั้นตอนการทำงานของโปรแกรมในการทดสอบความสัมพันธ์ของการแจกแจงแบบโลจิสติกกับการแจกแจงแบบเอฟ กรณีข้อมูลเริ่มต้นมี 2 ชุด



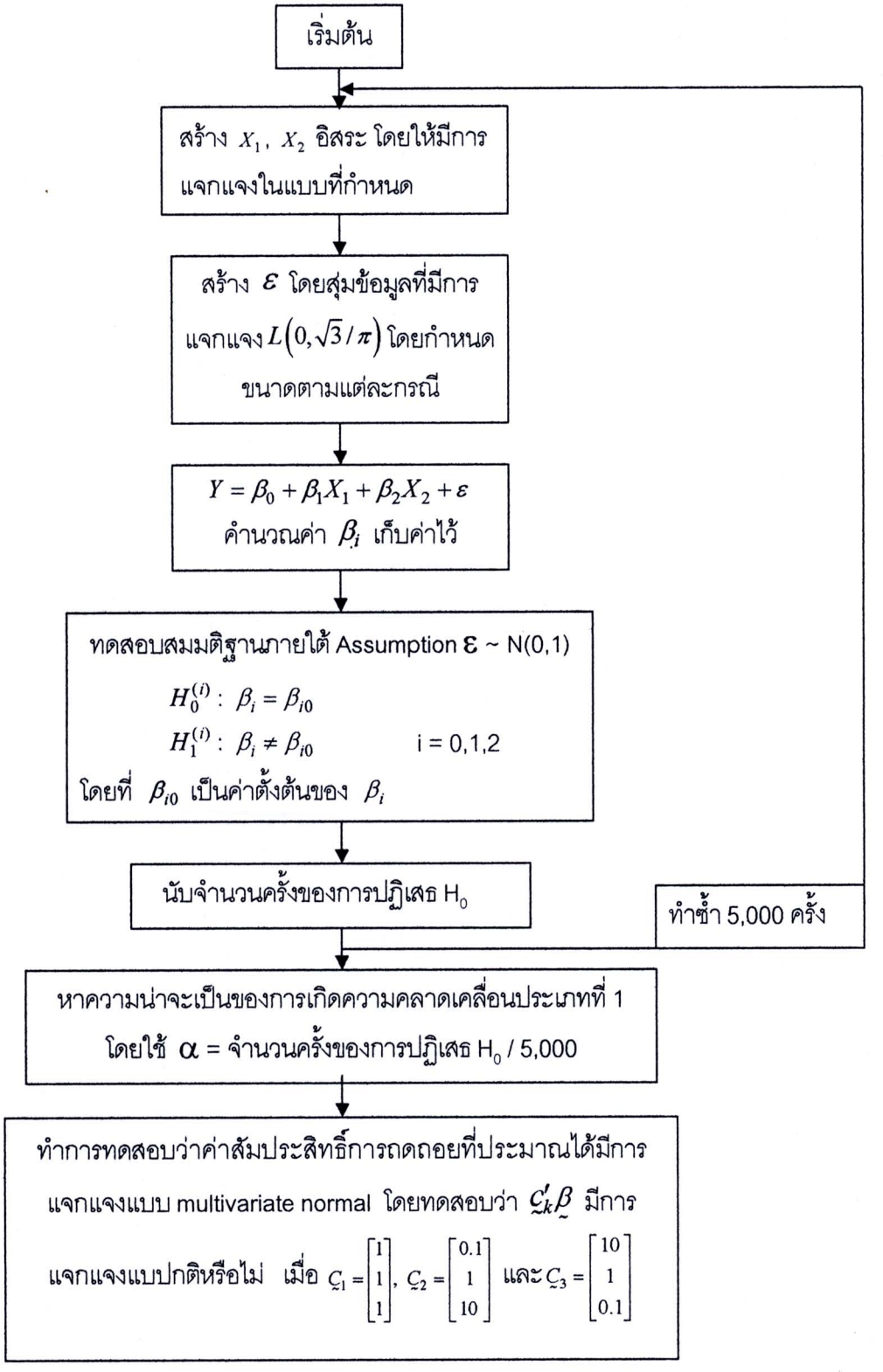
ภาพที่ 3.5 แสดงขั้นตอนการทำงานของโปรแกรมในการทดสอบความสัมพันธ์ของการแจกแจงแบบโลจิสติกกับการแจกแจงแบบเอฟ กรณีข้อมูลเริ่มต้นจากข้อมูลชุดเดียว



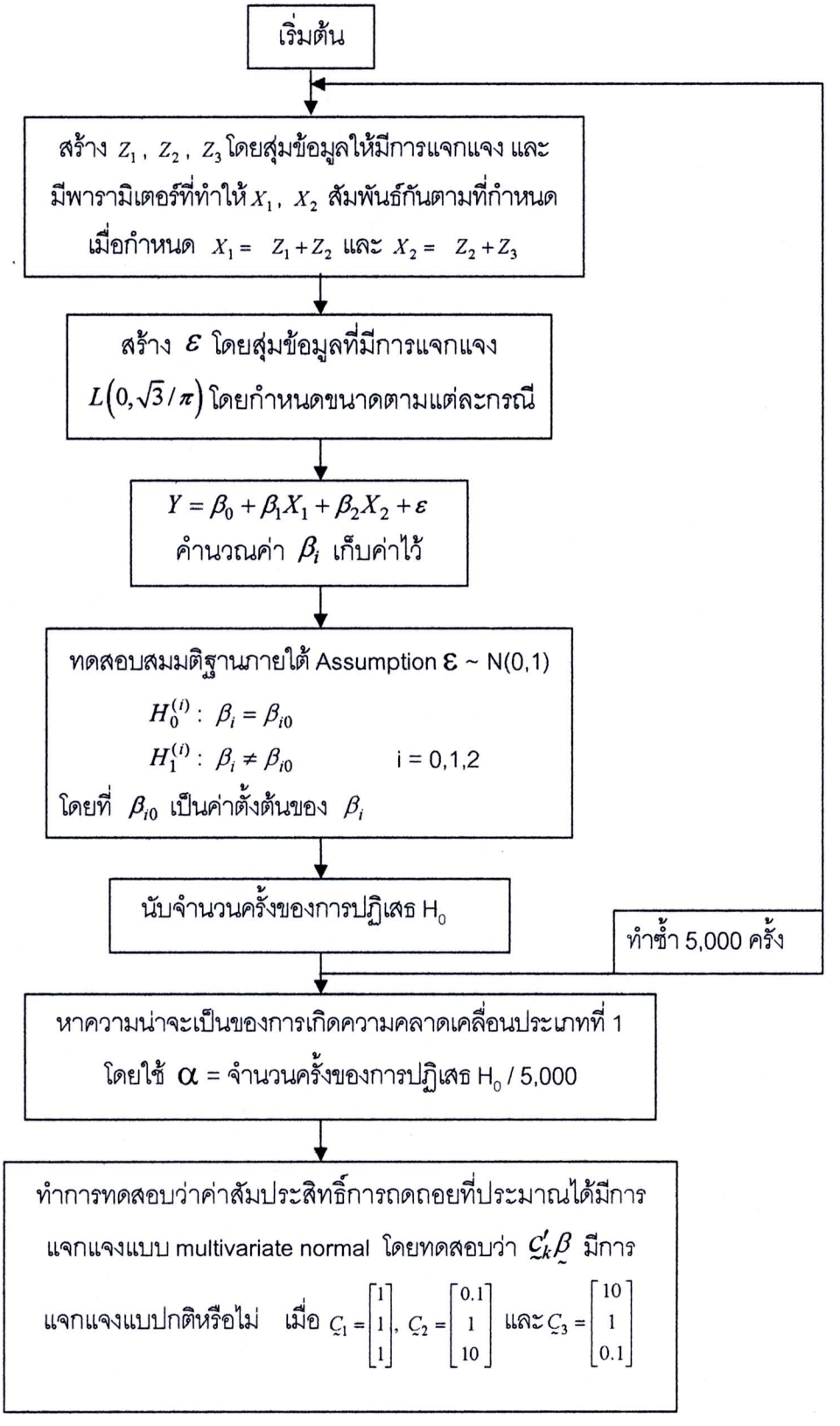
ภาพที่ 3.6 แสดงขั้นตอนการทำงานของโปรแกรมในการศึกษาผลกระทบของการประมาณค่าสัมประสิทธิ์การถดถอย โดยความคลาดเคลื่อนได้จากการสร้างข้อมูลโดยตรง กรณีการถดถอยอย่างง่าย



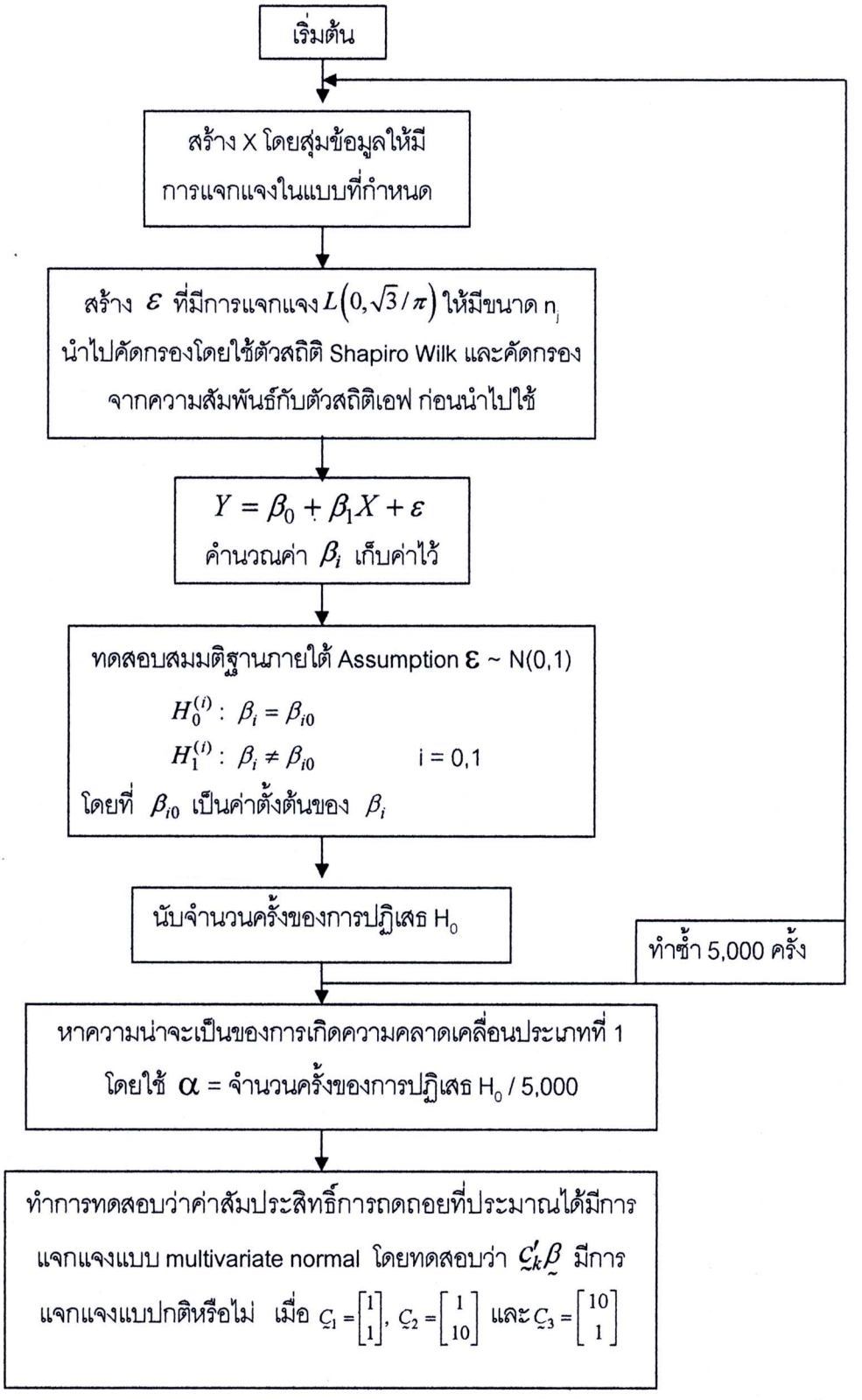
ภาพที่ 3.7 แสดงขั้นตอนการทำงานของโปรแกรมในการศึกษาผลกระทบของการประมาณค่าสัมประสิทธิ์การถดถอย โดยความคลาดเคลื่อนได้จากการสร้างข้อมูลโดยตรง กรณีการถดถอยเชิงพหุ เมื่อ X_1, X_2 อิสระ



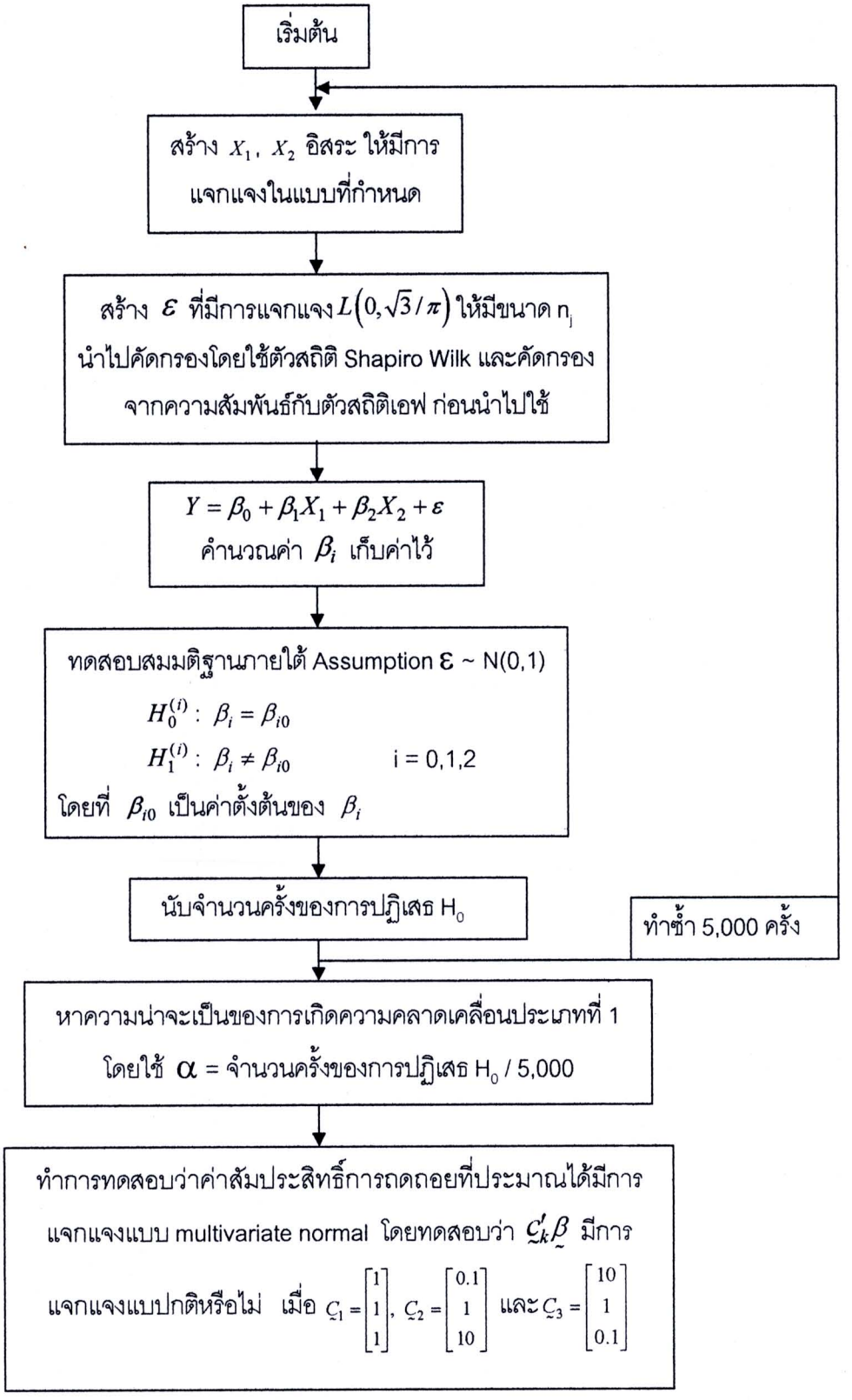
ภาพที่ 3.8 แสดงขั้นตอนการทำงานของโปรแกรมในการศึกษาผลกระทบของการประมาณค่าสัมประสิทธิ์การถดถอย โดยความคลาดเคลื่อนได้จากการสร้างข้อมูลโดยตรง กรณีการถดถอยเชิงพหุ เมื่อ X_1, X_2 มีความสัมพันธ์กัน



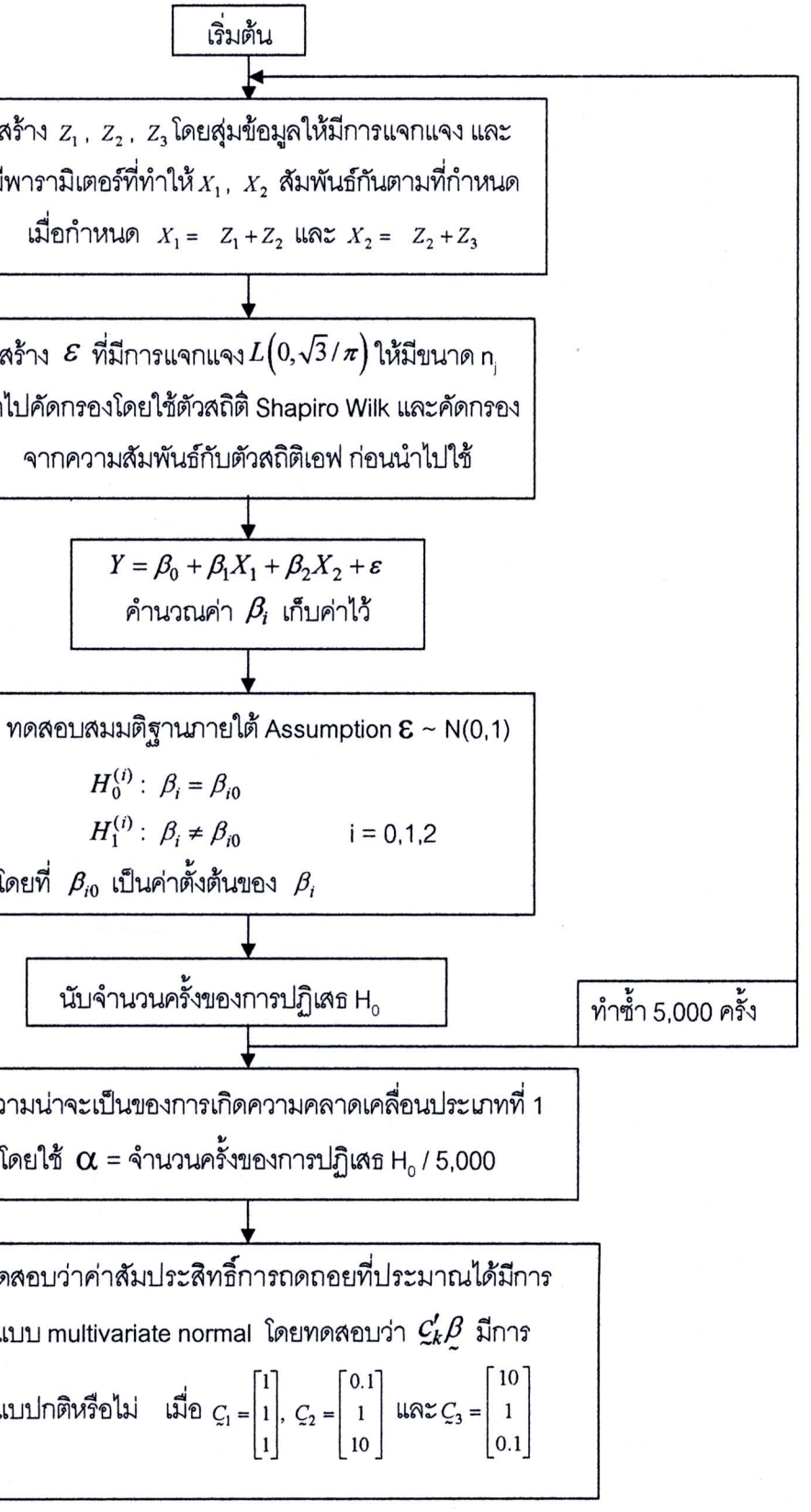
ภาพที่ 3.9 แสดงขั้นตอนการทำงานของโปรแกรมในการศึกษาผลกระทบของการประมาณค่าสัมประสิทธิ์การถดถอย โดยคัดกรองความคลาดเคลื่อนจากการแจกแจงแบบปกติก่อนนำไปใช้ กรณีการถดถอยอย่างง่าย



ภาพที่ 3.10 แสดงขั้นตอนการทำงานของโปรแกรมในการศึกษาผลกระทบของการประมาณค่าสัมประสิทธิ์การถดถอย โดยคัดกรองความคลาดเคลื่อนจากการแจกแจงแบบปกติก่อนนำไปใช้กรณีการถดถอยเชิงพหุ เมื่อ X_1, X_2 อิสระ



ภาพที่ 3.11 แสดงขั้นตอนการทำงานของโปรแกรมในการศึกษาผลกระทบของการประมาณค่าสัมประสิทธิ์การถดถอย โดยคัดกรองความคลาดเคลื่อนจากการแจกแจงแบบปกติก่อนนำไปใช้กรณีการถดถอยเชิงพหุ เมื่อ X_1, X_2 มีความสัมพันธ์กัน



สำนักงานคณะกรรมการวิจัยแห่งชาติ ห้องสมุดงานวิจัย