บทที่ 3
วิธีดำเนินการวิจัย

ในการวิจัยครั้งนี้มีลักษณะเป็นการวิจัยเชิงทดลอง เพื่อทำการเปรียบเทียบประสิทธิภาพวิธีการทดสอบเทียบความกลมกลืนสำหรับตัวแบบการถดถอย ซึ่งพิจารณาจากค่าความผิดพลาดประเภทที่ 1 และค่าอำนาจการทดสอบของตัวสถิติทดสอบ 3 ตัว คือ

1. ตัวสถิติทดสอบเอฟ (F)
2. ตัวสถิติทดสอบ Kolmogorov-Smirnov (KS)
3. ตัวสถิติทดสอบ Cramer-von Mises (CvM)

โดยสถานการณ์ต่างๆ ที่ใช้ในการวิจัยนั้นถูกสร้างขึ้นมาจากการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล(Monte Carlo Simulation Technique) ซึ่งเขียนโปรแกรมคอมพิวเตอร์ด้วยภาษาฟอร์แทรน 77 (Fortran 77 Programming) บนเครื่องคอมพิวเตอร์ส่วนบุคคล(Personal Computer) รายละเอียดเกี่ยวกับขั้นตอนการทดลองและโปรแกรมที่ใช้ในการทดลองเป็นดังนี้

3.1 แผนการทดลอง

การวิจัยครั้งนี้ ทำการทดสอบเทียบความกลมกลืนสำหรับตัวแบบการถดถอย ด้วยสถิติทดสอบเอฟ สถิติทดสอบ KS และสถิติทดสอบ CvM โดยกำหนดสถานการณ์ต่างๆ ต่อไปนี้

3.1.1 ตัวแบบที่ใช้ในการศึกษามี 4 ตัวแบบ ดังนี้

ตัวแบบที่ 1 : ตัวแบบเชิงเส้น ที่มีตัวแปรอสระ 1 ตัว

$$Y_i = 2 + 5X_i + \varepsilon_i \quad \text{เมื่อ } 1 \leq i \leq n$$

ตัวแบบที่ 2 : ตัวแบบพหุนาม องศา 2

$$Y_i = 2 + 5X_i + \beta_2 X_i^2 + \varepsilon_i \quad \text{เมื่อ } 1 \leq i \leq n$$

กำหนดให้ β_2 เท่ากับ 1, 3 และ 5

ตัวแบบที่ 3 : ตัวแบบเชิงเส้นที่มีตัวแปรอิสระ 2 ตัว

$$Y_i = 2 + 5X_{1i} - 1X_{2i} + \varepsilon_i \quad \text{เมื่อ } 1 \leq i \leq n$$

ตัวแบบที่ 4 : ตัวแบบที่มีตัวแปรอิสระ 2 ตัว และมีผลกระทบร่วม(interaction)

$$Y_i = 2 + 5X_{1i} - 1X_{2i} + \beta_3 X_{1i} X_{2i} + \varepsilon_i \quad \text{เมื่อ } 1 \leq i \leq n$$

กำหนดให้ β_3 เท่ากับ 1, 3 และ 5

3.1.2 ข้อมูลที่ศึกษาหรือตัวแปรอิสระแบ่งออกเป็น 2 ประเภทคือ ข้อมูลที่มีค่าซ้ำกัน และข้อมูลที่มีค่าไม่ซ้ำกัน ซึ่งในกรณีที่ข้อมูลมีค่าไม่ซ้ำกันนั้นจะทำการพิจารณาเปรียบเทียบเพียง

ตัวสถิติทดสอบ KS และตัวสถิติทดสอบ CvM เท่านั้น เนื่องจากในการทดสอบสมมติฐานด้วยตัวสถิติทดสอบเอฟ ตัวแปรอิสระต้องมีค่าซ้ำกันจึงจะสามารถคำนวณค่าของตัวสถิติทดสอบได้

3.1.3 กำหนดขนาดตัวอย่างที่ใช้ในการวิจัย คือ 10, 15, 20, 25, 30, 50 และ 70 หน่วย

3.1.4 กำหนดระดับนัยสำคัญ(α)ในการทดสอบ 3 ระดับ คือ 0.01, 0.05 และ 0.1

3.1.5 ลักษณะการแจกแจงแจงของความคลาดเคลื่อนที่นำมาศึกษาวิจัยในครั้งนี้

ก. การแจกแจงแบบปกติ โดยสนใจศึกษาเมื่อพารามิเตอร์ $\mu = 0$ และ $\sigma^2 = 1, 2$ และ 3 ตามลำดับ

ข. การแจกแจงแบบลอกนอร์มอล โดยสนใจศึกษาเมื่อพารามิเตอร์ $\mu = 0$ และ $\sigma^2 = 0.25$, 1.0 และ 2.25 ตามลำดับ

สำหรับการกำหนดพารามิเตอร์ของแต่ละการแจกแจงสำหรับการผลิตข้อมูลของค่าความคลาดเคลื่อนในตัวแบบการถดถอยนั้น ผู้วิจัยจะพิจารณาค่าพารามิเตอร์ที่ให้ลักษณะรูปกราฟแตกต่างกัน เพื่อให้เกิดความหลากหลายของรูปแบบของค่าความคลาดเคลื่อนในแต่ละการแจกแจงต่างๆ ที่สนใจศึกษานั้น

3.2 ขั้นตอนในการวิจัย

แบ่งออกเป็น 4 ขั้นตอนดังนี้

3.2.1 การสร้างข้อมูลให้มีการแจกแจงและลักษณะตามที่กำหนดในแผนการทดลอง

3.2.2 การคำนวณค่าสถิติทดสอบ

3.2.3 การหาค่าวิกฤติสำหรับการทดสอบสมมติฐานด้วยตัวสถิติทดสอบ KS และตัวสถิติทดสอบ CvM

3.2.4 การคำนวณหาค่าความผิดพลาดประเภทที่ 1 และค่าอำนาจการทดสอบ

ในแต่ละขั้นตอนมีรายละเอียดดังต่อไปนี้

3.2.1 การสร้างข้อมูลให้มีการแจกแจงและลักษณะตามที่กำหนดในแผนการทดลอง

การสร้างข้อมูลตามลักษณะที่กำหนดในแผนการทดลองนั้นใช้การจำลองจากโปรแกรมภาษาฟอร์แทรน 77 กับเครื่องคอมพิวเตอร์ส่วนบุคคล โดยใช้เลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอในช่วง (0,1) เป็นพื้นฐานในการจำลอง เมื่อทำการผลิตค่าของตัวแปรอิสระ และค่าความคลาดเคลื่อนแล้ว จึงทำการสร้างข้อมูล (X,Y) ที่มีความสัมพันธ์กันดังนี้

สร้างชุดของตัวแปรอิสระและค่าความคลาดเคลื่อนโดยการจำลองจากการแจกแจงและตามขนาดที่กำหนดไว้ในแผนการทดลอง จากค่าสัมประสิทธิ์การถดถอยที่กำหนดไว้เป็นเบื้องต้น สามารถนำมาสร้างเป็นค่าของตัวแปรตามโดยการแทนค่าลงในตัวแบบ ซึ่งในการวิจัยจะพิจารณาตัวแบบที่ใช้ในการสร้างข้อมูลทั้งหมด 4 ตัวแบบดังนี้

ตัวแบบที่ 1 : $Y_i = 2 + 5X_i + \varepsilon_i$ เมื่อ $1 \leq i \leq n$

ตัวแบบที่ 2 : $Y_i = 2 + 5X_i + \beta_2 X_i^2 + \varepsilon_i$ เมื่อ $1 \leq i \leq n$

กำหนดให้ β_2 เท่ากับ 1, 3 และ 5

ตัวแบบที่ 3 : $Y_i = 2 + 5X_{1i} - 1X_{2i} + \varepsilon_i$ เมื่อ $1 \leq i \leq n$

ตัวแบบที่ 4 : $Y_i = 2 + 5X_{1i} - 1X_{2i} + \beta_3 X_{1i} X_{2i} + \varepsilon_i$ เมื่อ $1 \leq i \leq n$

กำหนดให้ β_3 เท่ากับ 1, 3 และ 5

จากตัวแบบข้างต้นจำแนกได้ว่าตัวแบบที่ 1 และตัวแบบที่ 3 เป็นตัวแบบที่สอดคล้องกับสมมติฐานว่าง(null hypothesis) จะใช้ในการผลิตข้อมูลเพื่อการหาค่าประมาณของค่าความผิดพลาดประเภทที่ 1 แต่สำหรับตัวแบบที่ 2 และ ตัวแบบที่ 4 จะเป็นตัวแบบที่สอดคล้องกับสมมติฐานทางเลือก(alternative hypothesis) ที่ใช้ผลิตข้อมูลเพื่อการคำนวณค่าประมาณของค่าอำนาจการทดสอบ โดยถ้าตัวสถิติทดสอบใดที่สามารถควบคุมค่าความผิดพลาดประเภทที่ 1 ได้ในตัวแบบที่ 1 ก็จะพิจารณาหาค่าอำนาจการทดสอบจากตัวแบบที่ 2 และหากตัวสถิติทดสอบใดที่สามารถควบคุมค่าความผิดพลาดประเภทที่ 1 ได้ในตัวแบบที่ 3 ก็จะพิจารณาหาค่าอำนาจการทดสอบจากตัวแบบที่ 4 ในแต่ละสถานการณ์ที่กำหนด ซึ่งจะใช้ในการพิจารณาเปรียบเทียบเพื่อคัดเลือกตัวสถิติทดสอบที่เหมาะสมในการทดสอบเทียบความกลมกลืนสำหรับตัวแบบการถดถอยต่อไป

3.2.2 การคำนวณค่าสถิติทดสอบ

ในการวิจัยครั้งนี้เป็นการเปรียบเทียบอำนาจการทดสอบของการทดสอบเทียบความกลมกลืนสำหรับตัวแบบการถดถอย ซึ่งจะทำการเปรียบเทียบจากการทดสอบเทียบความกลมกลืนด้วยตัวสถิติทดสอบเอฟ ตัวสถิติทดสอบ KS และตัวสถิติทดสอบ CvM ดังนั้นเมื่อทำการจำลองข้อมูลตามขนาด การแจกแจงและสร้างค่าตัวแปรดังที่กล่าวในขั้นตอนที่ 1 แล้วนำข้อมูลที่ได้ไปคำนวณค่าต่างๆ ตามสูตรของตัวสถิติทดสอบแต่ละวิธีที่เสนอในบทที่ 2 เมื่อได้ค่าสถิติทดสอบแต่ละตัวให้เปรียบเทียบกับค่าวิกฤติ โดยตัวสถิติทดสอบเอฟเปรียบเทียบกับค่าวิกฤติที่ได้จากตารางการแจกแจงแบบเอฟ(F Distribution) ตัวสถิติทดสอบ KS และตัวสถิติทดสอบ CvM จะทำการเปรียบเทียบกับค่าวิกฤติที่ได้จากวิธีการสุ่มซ้ำแบบบูตสแตรป ดังที่จะกล่าวในขั้นตอนต่อไป ซึ่งการยอมรับหรือการปฏิเสธสมมติฐานว่างนั้นให้ถือเกณฑ์ในบทที่ 2

3.2.3 การหาค่าวิกฤติสำหรับการทดสอบสมมติฐานด้วยตัวสถิติทดสอบ KS และตัวสถิติทดสอบ CvM

การหาค่าวิกฤติที่ระดับนัยสำคัญ (α) ต่างๆ ในการวิจัยครั้งนี้ใช้หลักการคิดเปอร์เซ็นไทล์ (Percentile) ที่ตำแหน่ง γ โดยที่ $\gamma = 100(1-\alpha)$ ซึ่งมีรายละเอียดดังต่อไปนี้

จากข้อมูลที่ผลิตขึ้นตามลักษณะที่กำหนดในแผนการทดลอง นำมาทำการสุ่มตัวอย่างแบบบูตสแตรป ซึ่งเป็นการสุ่มตัวอย่างซ้ำแบบใส่คืนให้ได้จำนวนเท่ากับขนาดตัวอย่างที่มีอยู่เดิม เป็นจำนวน 500 ชุด ทำการคำนวณหาค่าตัวสถิติทดสอบ 2 ตัว คือ ตัวสถิติทดสอบ KS และตัวสถิติทดสอบ CvM จากข้อมูลแต่ละชุดดังกล่าว เรียงลำดับค่าสถิติทดสอบของแต่ละวิธีที่ได้จากน้อยไปหามาก จะได้ค่าวิกฤติที่ระดับนัยสำคัญ 0.01, 0.05 และ 0.10 หรือค่าเปอร์เซ็นไทล์ตำแหน่งที่ 99, 95 และ 90 ตามลำดับ ซึ่งก็คือค่าสถิติทดสอบอันดับที่ 490, 475 และ 450 ตามลำดับ

3.2.4 การหาค่าความผิดพลาดประเภทที่ 1 และค่าอำนาจการทดสอบ

จากสมมติฐานว่างที่กำหนดไว้ว่าตัวแบบการถดถอยเป็นตัวแบบเชิงเส้นและข้อมูลที่ผลิตขึ้นตามสถานการณ์ต่างๆ ที่กล่าวไว้ในแผนการทดลอง เมื่อทำการทดสอบสมมติฐานผลสรุปที่ได้สามารถนำมาใช้ในการคำนวณค่าความผิดพลาดประเภทที่ 1 และค่าอำนาจการทดสอบได้ดังรายละเอียดต่อไปนี้

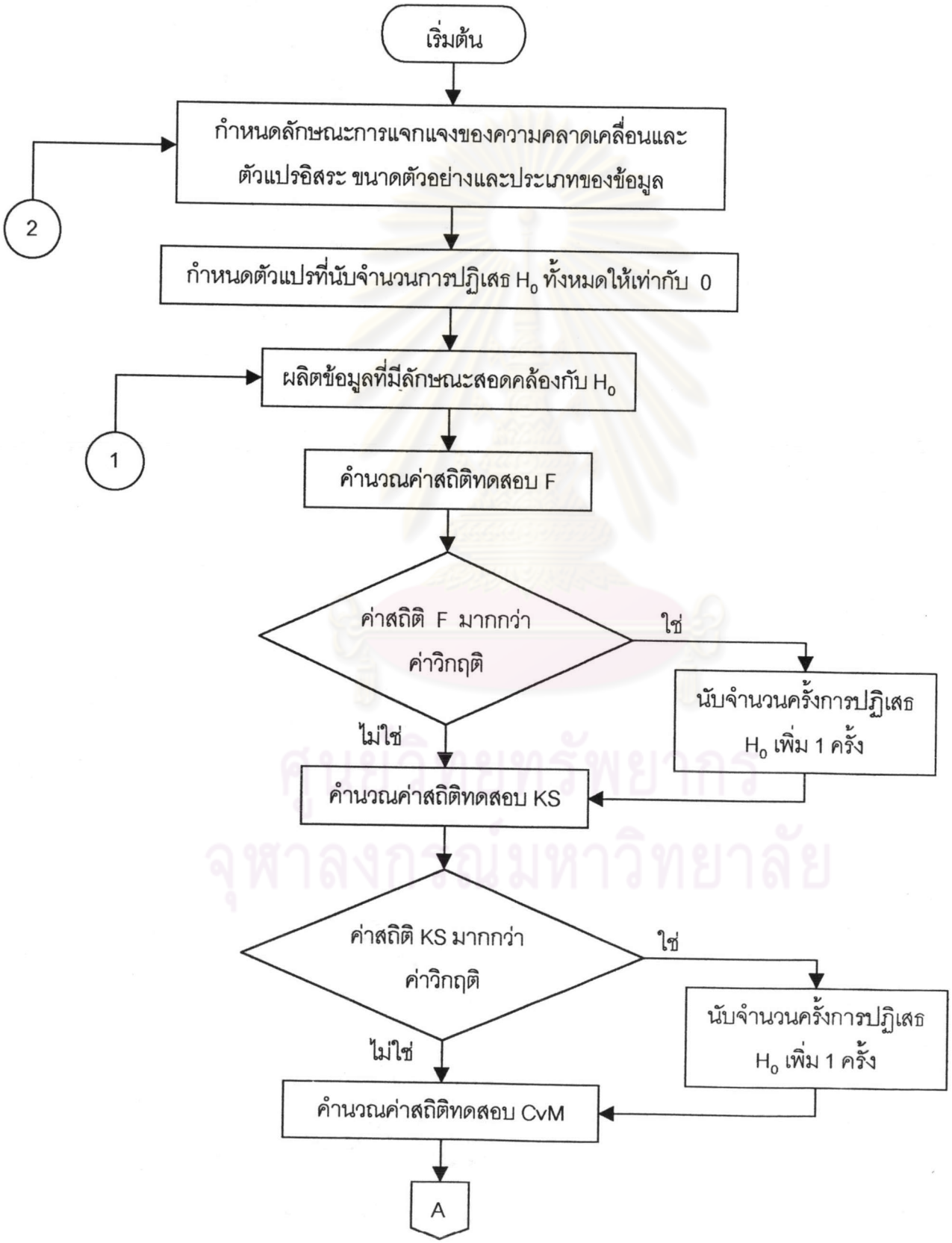
3.2.4.1 ในการดำเนินงานขั้นแรกจะทำการทดสอบสมมติฐานเพื่อคำนวณหาค่าประมาณของค่าความผิดพลาดประเภทที่ 1 จากสมมติฐานว่างที่ว่าตัวแบบที่พิจารณาเป็นตัวแบบเชิงเส้น เมื่อทำการผลิตข้อมูลจากตัวแบบตามสมมติฐานว่างและสถานการณ์ที่กำหนดไว้ในแผนการทดลองแล้ว นำมาคำนวณหาค่าตัวสถิติทดสอบทั้ง 3 ตัว คือ ตัวสถิติทดสอบเอฟ ตัวสถิติทดสอบ KS และตัวสถิติทดสอบ CvM และเปรียบเทียบกับค่าวิกฤติของแต่ละตัวสถิติทดสอบดังที่กล่าวไว้ในบทที่ 2 และหัวข้อก่อนหน้านี้ ทำซ้ำๆ กันเป็นจำนวน 1,000 ครั้ง ทำการนับจำนวนครั้งของการปฏิเสธสมมติฐานว่าง โดยสัดส่วนที่ได้จากการหารจำนวนครั้งที่ปฏิเสธสมมติฐานว่างด้วย 1,000 ซึ่งเป็นจำนวนครั้งทั้งหมดในการทดลองจะเป็นค่าประมาณของค่าความผิดพลาดประเภทที่ 1 จากนั้นทำการทดสอบแบบทวิ (Binomial Test) ภายใต้สมมติฐานว่าง $H_0 : \alpha \leq \alpha_0$ เมื่อ α_0 เป็นระดับนัยสำคัญที่กำหนด ดังที่กล่าวไว้ในบทที่ 1 พิจารณาว่าสถิติทดสอบใดที่ได้ผลการทดสอบเป็นยอมรับ H_0 แสดงว่าสามารถควบคุมความผิดพลาดประเภทที่ 1 ได้จึงนำมาทดสอบหาค่าอำนาจการทดสอบต่อไป

3.2.4.2 ในกรณีการหาค่าอำนาจการทดสอบนั้น จากสมมติฐานว่างที่ว่าตัวแบบที่พิจารณาเป็นตัวแบบเชิงเส้น เมื่อทำการผลิตข้อมูลตามสถานการณ์ที่กำหนดไว้ในแผนการทดลอง แต่ผลิตจากตัวแบบที่ไม่เป็นไปตามสมมติฐานว่างแล้ว นำมาคำนวณหาค่าตัวสถิติทดสอบทั้ง 3 ตัว คือ ตัวสถิติทดสอบเอฟ ตัวสถิติทดสอบ KS และตัวสถิติทดสอบ CvM และเปรียบเทียบกับค่าวิกฤติของแต่ละตัวสถิติทดสอบดังที่กล่าวไว้ในบทที่ 2 และหัวข้อก่อนหน้านี้ ทำซ้ำๆ กันเป็นจำนวน 1,000 ครั้ง ทำการนับจำนวนครั้งของการปฏิเสธสมมติฐานว่าง โดยสัดส่วนที่ได้จากการหารจำนวนครั้งที่ปฏิเสธสมมติฐานว่างด้วย 1,000 ซึ่งเป็นจำนวนครั้งทั้งหมดในการทดลองจะเป็นค่าประมาณของค่าอำนาจการทดสอบ พิจารณาเปรียบเทียบตัวสถิติทดสอบแต่ละตัวว่ามีค่าอำนาจการทดสอบสูงที่สุดในกรณีใดบ้าง และหากว่าตัวสถิติทดสอบตัวใดที่มีอำนาจการทดสอบสูงสุดจะถือว่าเป็นวิธีการที่เหมาะสมที่สุด ภายใต้ข้อตกลงเบื้องต้นของการวิจัยครั้งนี้

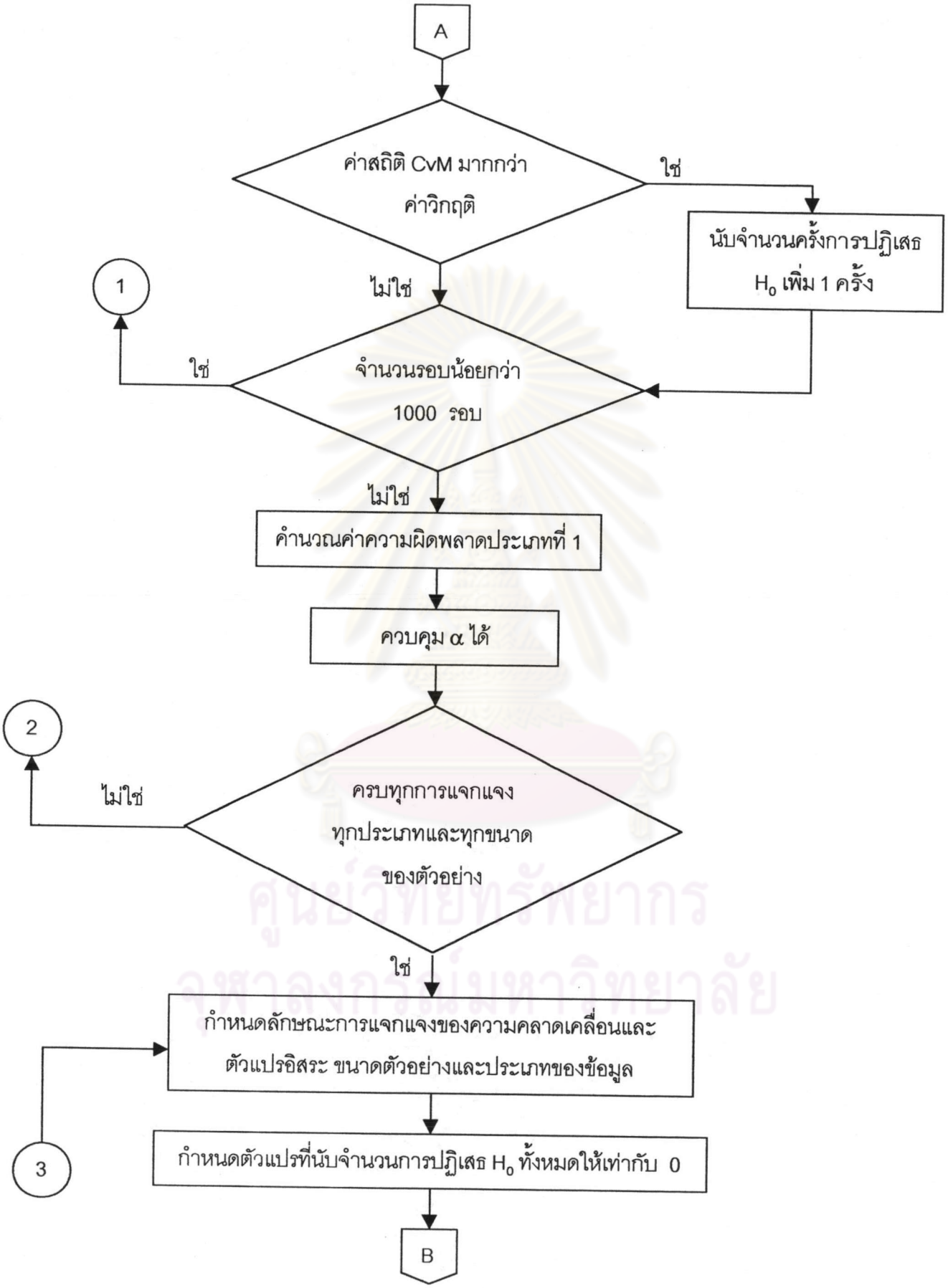
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

3.3 ขั้นตอนในการทำงานของโปรแกรม

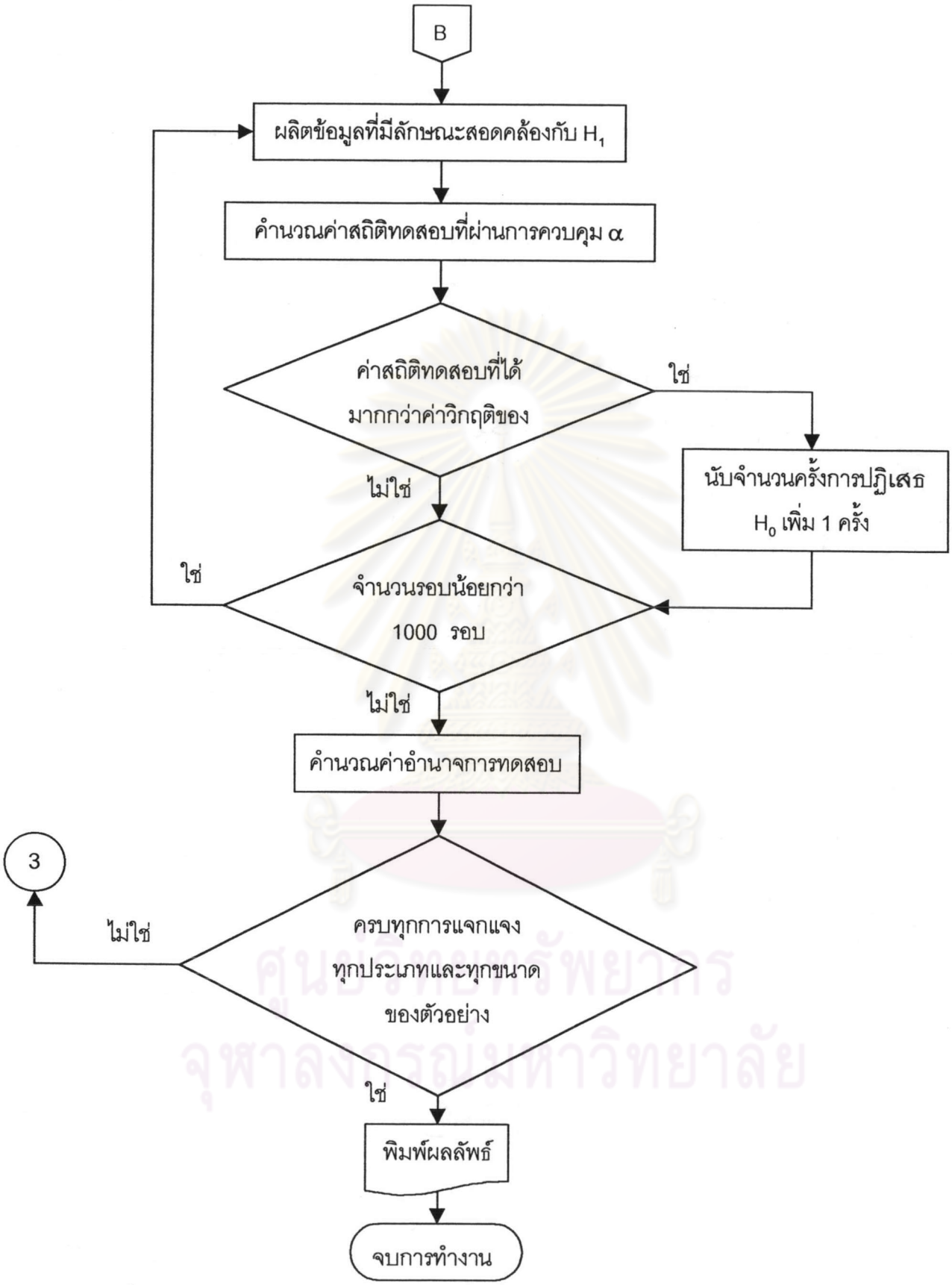
โปรแกรมที่สร้างขึ้นจากภาษาฟอร์แทรน 77 ใช้ในการเปรียบเทียบวิธีการทดสอบเทียบความกลมกลืนสำหรับตัวแบบการถดถอยได้นำเสนอไว้ในภาคผนวก ส่วนขั้นตอนในการทำงานของโปแกรมแสดงได้ดังแผนผังต่อไปนี้



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 3.1 แผนผังแสดงโปรแกรมคอมพิวเตอร์เปรียบเทียบประสิทธิภาพวิธีการทดสอบเทียบความกลมกลืนสำหรับตัวแบบการถดถอย

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย