บทที่ 3

วิธีดำเนินการวิจัย

ในการวิจัยครั้งนี้มีลักษณะเป็นการวิจัยเชิงทดลอง เพื่อทำการทดสอบเทียบความกลมกลืน (Goodness-of-Fit Test) สำหรับการแจกแจงไวบูลล์ และการแจกแจงกอมเพิรตซ์ กรณีขนาดตัวอย่างถูกตัดทิ้งประเภทที่ 2 จำนวนมาก สถิติทดสอบที่ใช้มี 3 วิธีด้วยกันคือ KS, K และ CVM ซึ่งสถานการณ์ต่าง ๆ ที่ใช้ในการวิจัยนั้นถูกสร้างขึ้นมาโดยอาศัยการจำลองโดยใช้เทคนิคการจำลองมอนติคาร์โล (Monte-Carlo Simulation Technique) โดยใช้เครื่องคอมพิวเตอร์ AMDAHL 5860 ในการเขียนโปรแกรมภาษาฟอร์แทรน (Fortran) ซึ่งจะกล่าวในรายละเอียดต่อไป นอกจากนี้ยังได้ทำการแสดงความสามารถในการควบคุมค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และการเปรียบเทียบค่าอำนาจการทดสอบของสถิติทดสอบ ดังได้กล่าวมาแล้ว มีขั้นตอนในการวิจัยดังต่อไปนี้

การจำลองโดยใช้เทคนิคมอนติคาร์โล

เทคนิคที่ใช้แก้ปัญหาในการคำนวณทางสถิตินั้นมีอยู่หลายวิธี วิธีการจำลองโดยใช้เทคนิคมอนติคาร์โล เป็นวิธีหนึ่งที่นิยมนำมาใช้แก้ปัญหากันอย่างแพร่หลายในปัจจุบัน ซึ่งหลักการของการจำลองโดยใช้เทคนิคดังกล่าว จะใช้เลขสุ่ม (Random Numbers) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา

ขั้นตอนของวิธีการจำลองด้วยเทคนิคมอนติคาร์โลที่ใช้กันในปัจจุบันแบ่งได้เป็น 3 ขั้นตอน ดังนี้คือ

1. การสร้างตัวเลขสุ่ม การใช้ตัวเลขสุ่มเป็นสิ่งที่สำคัญมากในเทคนิคนี้ ทั้งนี้เพราะว่าหลักการของการจำลองแบบมอนติคาร์โลนั้น จะใช้ตัวเลขสุ่มมาช่วยในการหาคำตอบของปัญหา โดยลักษณะของตัวเลขสุ่มที่นำมาใช้ จะมีการแจกแจงแบบสม่ำเสมอในช่วง (0,1) สำหรับวิธีการสร้างตัวเลขแบบสุ่ม มีผู้เสนอไว้หลายวิธี แต่วิธีที่ดีนั้นลักษณะของตัวเลขสุ่มที่ถูกสร้างขึ้นจะต้องมีการแจกแจงแบบสม่ำเสมอในช่วง (0,1) ตัวเลขสุ่มแต่ละตัวเป็นอิสระต่อกัน และมีช่วงยาวก่อนจะเกิดเลขสุ่มซ้ำ (มีวัฏจักรยาว)

2. การนำตัวเลขสุ่มมา ประยุกต์ใช้กับปัญหาที่ต้องการศึกษา ซึ่งขั้นตอนนี้ขึ้นอยู่กับลักษณะของปัญหา บางปัญหาอาจจะไม่ใช้ตัวเลขสุ่มโดยตรง แต่จะนำไปผลิตเลขสุ่มที่มีการแจกแจงแบบอื่นต่อไป

3. การทดลองกระทำ เมื่อนำตัวเลขสุ่มมาประยุกต์ให้เข้ากับปัญหาที่ต้องการศึกษาได้แล้ว ขั้นต่อไปคือ การทดลองโดยใช้กระบวนการของการสุ่ม (Random Process) มากระทำในลักษณะซ้ำ ๆ กันหลาย ๆ ครั้ง เพื่อหาคำตอบที่ต้องการ

แผนการทดลอง

การวิจัยครั้งนี้ทำการทดสอบเทียบความกลมกลืน สำหรับการแจกแจงไวบูลล์ และการแจกแจงกอมเพิรตซ์ เริ่มต้นจากการสร้างตารางแสดงค่าวิกฤตของทั้ง 2 การแจกแจง ที่ระดับนัยสำคัญ 6 ระดับ สถิติทดสอบ 3 วิธี ขนาดตัวอย่าง 4 ระดับ และเปอร์เซ็นต์การถูกตัดทิ้ง 3 ระดับ ซึ่งยกเว้นขนาดตัวอย่าง 100 และ 300 มีเปอร์เซ็นต์การถูกตัดทิ้งเพียง 2 ระดับ ขั้นที่สองจะทำการพิจารณาความสามารถในการควบคุมค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ในแต่ละสถานการณ์ และในขั้นสุดท้ายจะพิจารณาค่าอำนาจการทดสอบของสถิติทดสอบแต่ละวิธี

ขั้นตอนในการวิจัย

ขั้นตอนในการวิจัยแบ่งได้เป็น 3 ขั้นตอนดังนี้

1. การหาค่าวิกฤต

ในการหาค่าวิกฤตที่ระดับนัยสำคัญ (α) ต่าง ๆ ในการวิจัยครั้งนี้ใช้หลักการคิดเปอร์เซ็นต์ไทล์ (Percentile) ที่ตำแหน่ง γ โดยที่ $\gamma = (1 - \alpha)$ ซึ่งมีขั้นตอนดังต่อไปนี้

1.1 กำหนดลักษณะการแจกแจงของประชากรที่ตั้งเป็นสมมติฐานว่าง (H_0) ดังนี้

1.1.1 การแจกแจงไวบูลล์ ($\alpha = 3, \beta = 1$)

1.1.2 การแจกแจงกอมเพิรตซ์ ($B = 0.02, c = 20$)

1.2 กำหนดขนาดตัวอย่าง (Sample Size) มี 4 ระดับ คือ 100, 300, 500 และ 700

1.3 กำหนดเปอร์เซ็นต์การถูกตัดทิ้ง มี 3 ระดับ 90%, 95% และ 99% ยกเว้นที่ขนาดตัวอย่าง 100 และ 300 มี 2 ระดับ 90% และ 95%

1.4 สุ่มตัวอย่างจากที่กำหนดในข้อ 1.1 ถึงข้อ 1.3

1.5 คำนวณค่าสถิติทดสอบทั้ง 3 วิธี

1.5.1 สถิติทดสอบ KS

$$D^{+} = \max_{i \le r} \left\{ i/N - F(x_{i:N};\, \theta, \beta) \right\}$$

$$D^{-} = \max_{i \le r} \left\{ F(x_{i:N};\, \theta, \beta) - (i-1)/N \right\}$$

และ

$$D = \max (D^{+}, D^{-})$$

1.5.2 สถิติทดสอบ K

$$V = D^{+} + D^{-}$$

1.5.3 สถิติทดสอบ CVM

$$C = \sum_{i=1}^{r} \left[F(x_{i:N};\, \theta, \beta) - (i-0.5)/N \right]^{2} + 1/(12N)$$

1.6 ทำข้อ 1.4 ถึงข้อ 1.5 จำนวน 1,000 รอบ

1.7 เรียงลำดับค่าสถิติทดสอบแต่ละวิธีจากน้อยไปหามาก

1.8 คำนวณหาค่าวิกฤตที่ระดับนัยสำคัญ 0.25, 0.20, 0.15, 0.10, 0.05 และ 0.01 โดยคำนวณที่เปอร์เซ็นต์ไทล์ (γ) ที่ 0.75, 0.80, 0.85, 0.90, 0.95 และ 0.99 ตามลำดับมีวิธีคำนวณดังนี้

กำหนด m คือ จำนวนชุดข้อมูลเพื่อการหาค่าสถิติทดสอบ
(สำหรับการวิจัยครั้งนี้ m = 1000 ตามข้อ 1.6)

y_{γ} คือ ค่าที่ตำแหน่งเปอร์เซ็นต์ไทล์ที่ γ

สูตรที่ใช้

$$y_\gamma = (m + 1) * \gamma$$

ให้ t เป็นจำนวนเต็มที่มากที่สุดที่น้อยกว่าหรือเท่ากับ $(m + 1) * \gamma$

s เป็นค่าทศนิยมของค่า $y_\gamma = (m + 1) * \gamma$

หรือ $s = (m + 1) * \gamma - t$

ทำการประมาณค่าในช่วงแบบเชิงเส้น (Linear Interpolation) โดยใช้สูตร

$$y_\gamma = (1 - s) * y_t + s * y_{t+1}$$

1.9 ทำซ้ำข้อ 1.6 ถึงข้อ 1.8 เป็นจำนวน 100 รอบ

1.10 หาค่าเฉลี่ยที่แต่ละตำแหน่งเปอร์เซ็นต์ไทล์ (γ) และแต่ละวิธีของสถิติทดสอบ ค่าที่ได้ คือค่าวิกฤตที่ระดับนัยสำคัญ $\alpha = (1 - \gamma)$

2. การหาค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1

การหาค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 กระทำได้โดย การนับจำนวนครั้งของการปฏิเสธสมมติฐานว่าง หารด้วยจำนวนรอบทั้งหมด มีขั้นตอนการทำงานดังนี้

2.1 กำหนดลักษณะการแจกแจงของประชากร ในการหาค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 การแจกแจงของประชากรจะเป็นไปตามสมมติฐานว่าง ที่ตั้ง การแจกแจงของประชากรมีดังนี้

2.1.1 การแจกแจงไวบูลล์ ($\alpha = 3, \beta = 1$)

2.1.2 การแจกแจงกอมเพิรตซ์ ($B = 0.02, c = 20$)

2.2 กำหนดขนาดตัวอย่างมี 4 ระดับคือ 100, 300, 500 และ 700

2.3 กำหนดเปอร์เซ็นต์การถูกตัดทิ้ง มี 3 ระดับ 90%, 95% และ 99% ยกเว้นที่ขนาดตัวอย่าง 100 และ 300 มีเพียง 2 ระดับ 90% และ 95%

2.4 สุ่มตัวอย่างจากที่กำหนดในข้อ 2.1 ถึงข้อ 2.3

2.5 คำนวณค่าสถิติทดสอบ KS, K และ CVM

2.6 นำค่าสถิติทดสอบที่ได้จากข้อ 2.5 มาเปรียบเทียบกับค่าวิกฤต แล้วนับจำนวนครั้งที่ปฏิเสธสมมติฐานว่าง

2.7 ทำซ้ำข้อ 2.4 ถึงข้อ 2.6 จำนวน 2,000 รอบ

2.8 นำจำนวนครั้งที่ปฏิเสธสมมติฐาน ว่าง ทั้งหมด หารด้วยจำนวนรอบทั้งหมด 2,000 รอบ จะได้เป็นค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1

เมื่อได้ค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 แล้วก็ทำการเปรียบเทียบค่าที่ได้ในแต่ละสถานการณ์กับเกณฑ์ของ Coohran และเกณฑ์ของ Bradley พิจารณาว่าสถานการณ์ใดสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ และในสถานการณ์ใดที่ไม่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ ในการวิจัยนี้ ผู้วิจัยสนใจศึกษาอำนาจการทดสอบของสถิติทดสอบทั้ง 3 วิธีในทุกกรณีที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ ดังนั้น สถานการณ์ใดสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ไม่ว่าจะด้วยเกณฑ์ของ Coohran หรือเกณฑ์ของ Bradley จะถูกนำมาหาค่าอำนาจการทดสอบทั้งหมด

3. การหาค่าอำนาจการทดสอบ

การหาค่าอำนาจการทดสอบคล้ายกับการหาค่าความคลาดเคลื่อนประเภทที่ 1 ทำได้โดยการนับจำนวนครั้งของการปฏิเสธสมมติฐาน ว่าง หารด้วยจำนวนรอบทั้งหมด แต่ลักษณะการแจกแจงของประชากรเป็นการแจกแจงอื่นมิใช่การแจกแจงตามที่ตั้งไว้ในสมมติฐาน ว่าง มีขั้นตอนการทำงานดังนี้

3.1 กำหนดลักษณะการแจกแจงของประชากร ของสมมติฐาน ว่าง (H_0) มีดังนี้

3.1.1 การแจกแจงไวบูลล์ ($\alpha = 3, \beta = 1$)

3.1.2 การแจกแจงกอมเพิรตซ์ ($B = 0.02, c = 20$)

3.2 กำหนดลักษณะการแจกแจงของประชากร ของสมมติฐานแย้ง (H_1) มีดังนี้

3.2.1 การแจกแจงลอกนอร์มอล

3.2.2 การแจกแจงลอกโลจีสติก

3.2.3 การแจกแจงไวบูลล์ 3 พารามิเตอร์

3.2.4 การแจกแจงไค-สแควร์

3.3 กำหนดขนาดตัวอย่างมี 4 ระดับคือ 100, 300, 500 และ 700

3.4 กำหนดเปอร์เซ็นต์การถูกตัดทิ้ง มี 3 ระดับ 90%, 95% และ 99% ยกเว้นที่ขนาด 100 และ 300 มี 2 ระดับ คือ 90% และ 95%

3.5 สุ่มตัวอย่างจากที่กำหนดในข้อ 1. ถึงข้อ 3.

3.6 คำนวณค่าสถิติทดสอบ KS, K และ CVM

3.7 นำค่าสถิติทดสอบที่ได้จากข้อ 3.6 มาเปรียบเทียบกับค่าวิกฤต แล้วนับจำนวนครั้งที่ปฏิเสธสมมติฐาน ว่าง

3.8 ทำซ้ำข้อ 3.5 ถึงข้อ 3.7 จำนวน 2,000 รอบ

3.9 นำจำนวนครั้งที่ปฏิเสธสมมติฐาน ว่าง ทั้งหมดหารด้วยจำนวนรอบทั้งหมด 2,000 รอบ จะได้เป็นค่าอำนาจการทดสอบ

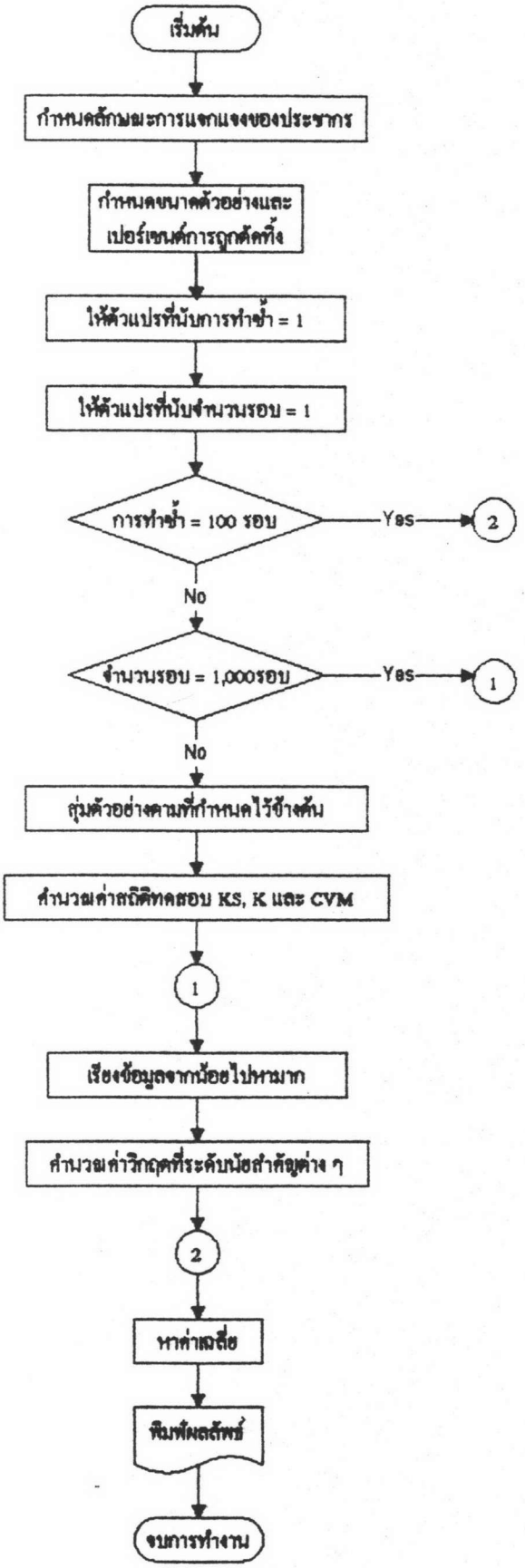
ขั้นตอนในการทำงานของโปรแกรม

สามารถเขียนผังงานแสดงขั้นตอนในการหาค่าวิกฤตได้ตามรูปที่ 3.1 และผังงานแสดงขั้นตอนในการหาค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และการหาค่าอำนาจการทดสอบ ได้ตามรูปที่ 3.2 ดังรายละเอียดต่อไปนี้

รูปที่ 3.1 แผนผังแสดงขั้นตอนการหาค่าวิกฤต

รูปที่ 3.2 แผนผังแสดงขั้นตอนการหาค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการทดสอบ

รูปที่ 3.1 แผนผังแสดงขั้นตอนการหาค่าวิกฤต



รูปที่ 3.2

แผนผังแสดงขั้นตอนการหาค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการทดสอบ