บทที่ 3

การดำเนินการวิจัย

การวิจัยเรื่องนี้มีลักษณะเป็นการวิจัยเชิงทดลอง ซึ่งจำลองขึ้นด้วยการทำงานของเครื่องคอมพิวเตอร์ โดยใช้เทคนิคมอนติคาร์โลซิมูเลชั่น เพื่อหาผลสรุปในการเปรียบเทียบอำนาจการทดสอบการแจกแจงแบบปกติของตัวสถิตินอนพาราเมตริก 6 ตัว

หลักสำคัญของเทคนิคมอนติคาร์โลซิมูเลชั่นคือ การใช้เลขสุ่ม (Random Number) มาช่วยหาคำตอบที่ต้องการศึกษา ขั้นตอนของวิธีมอนติคาร์โลแบ่งได้เป็น 3 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 สร้างเลขสุ่ม การใช้เลขสุ่มเป็นสิ่งสำคัญมากในวิธีมอนติคาร์โล ทั้งนี้เพราะว่าหลักการของวิธีมอนติคาร์โลนั้นจะใช้เลขสุ่มมาช่วยในการหาคำตอบของปัญหา แต่วิธีการสร้างเลขสุ่มนั้นมีหลายวิธี แต่วิธีที่ดีนั้นจะให้ลักษณะของเลขสุ่มมีการแจกแจงแบบยูนิฟอร์มในช่วง (0,1) และเป็นอิสระกัน

ขั้นตอนที่ 2 ประยุกต์ปัญหาที่ต้องการศึกษามาใช้กับเลขสุ่ม ซึ่งขั้นตอนนี้ขึ้นอยู่กับลักษณะของปัญหาที่ศึกษา บางปัญหาอาจจะใช้เลขสุ่มได้โดยตรง ในขณะที่บางปัญหาอาจต้องใช้ขั้นตอนอื่นอีกหลายขั้นตอน โดยที่มีการใช้ตัวเลขสุ่มในบางขั้นตอนเท่านั้น

ขั้นตอนที่ 3 ทดลองกระทำ เมื่อประยุกต์ปัญหาที่สนใจให้ใช้เลขสุ่มได้แล้ว ขั้นต่อไปก็คือ การทดลอง โดยใช้กระบวนการของการสุ่ม (Random Process) มากระทำในลักษณะซ้ำ ๆ กัน (Replication) เพื่อหาคำตอบของปัญหาที่ต้องการศึกษา

การวางแผนการทดลอง

กำหนดสถานการณ์ต่าง ๆ สำหรับหาค่าความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 และอำนาจการทดสอบของตัวสถิติทั้ง 6 ตัวดังนี้

1. สุ่มตัวอย่างจากประชากร โดยกำหนดให้ประชากรมีการแจกแจงเป็น

   1.1) การแจกแจงแบบปกติ

1.2) การแจกแจงแบบเบ้ ความเบ้และความโด่งที่ใช้ในการศึกษาครั้งนี้ คือ ความเบ้เท่ากับ 0.0(0.25)1.00 ความโด่งเท่ากับ 2.0(0.4)6.0

ทำการศึกษาโดยแบ่งเป็นกลุ่มตามความเบ้และความโด่ง โดยอาศัยเกณฑ์ของ Shapiro, Wilk และ Chen (1968) ดังนี้

ลักษณะการแจกแจง	ความเบ้	ความโด่ง
Near Normal	0.0	2.8(0.4)4.4
	0.25	2.8(0.4)4.4
Symmetric Short-Tailed	0.0	2.0, 2.4
Symmetric Long-Tailed	0.0	4.8(0.4)6.0
Asymmetric Short-Tailed	0.50	2.4, 2.8
	0.75	2.8
Asymmetric Long-Tailed	0.50	3.2(0.4)6.0
	0.75	3.2(0.4)6.0
	1.00	3.6(0.4)6.0

2. กำหนดขนาดตัวอย่าง(Sample Size) 5 ระดับ คือ 10 20 30 50 และ 100

3. กำหนดพารามิเตอร์(Parameter) μ คือ ค่าเฉลี่ยเลขคณิตของประชากรในแผนการทดลองครั้งนี้เท่ากับ 100 และ σ^2 คือ ความแปรปรวนของประชากรเท่ากับ 100 กรณีที่ค่าเฉลี่ยและความแปรปรวนของประชากรมีค่าอื่น พบว่าผลการทดสอบสมมติฐานที่คำนวณได้มีค่าเท่ากัน ไม่ว่าจะทำการศึกษา ณ.จุดที่ค่าเฉลี่ยและความแปรปรวนมีค่าใด ๆ ก็ตาม

วิธีการทดลอง

เขียนโปรแกรมคอมพิวเตอร์ภาษาฟอร์แทรน(Fortran) โดยใช้กับเครื่องคอมพิวเตอร์ AMDAHL 5850 เพื่อสร้างข้อมูลให้เป็นไปตามแผนการทดลอง และคำนวณค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 พร้อมทั้งคำนวณค่าอำนาจการทดสอบของตัวสถิติที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ ซึ่งวิธีการทดลองจะแบ่งออกเป็น 2 ขั้นตอนดังต่อไปนี้

1. สร้างรูปแบบการแจกแจงของประชากรให้เป็นไปตามแผนการทดลองตามลำดับขั้นต่อไปนี้

1.1 สร้างตัวเลขสุ่มที่มีการแจกแจงแบบยูนิฟอร์ม (Uniformly Distributed Random Number)

ใช้วิธีการสร้างตัวเลขสุ่มแบบ Multiplicative Congruential Method โดยคำนวณจากสมการ

$$x_{i+1} = x_i \cdot a \quad (\text{Modulo } m) \qquad \ldots\ldots\ldots\ldots(1)$$

เมื่อ x_i เป็นเลขคล้ายสุ่มตัวที่ i

x_{i+1} เป็นเลขคล้ายสุ่มตัวที่ i+1

a เป็นตัวคูณคงที่ (Constant Multiplier)

Modulo m หมายความว่า ค่า $(x_i \cdot a)$ ถูกหารด้วย m จนกระทั่งเหลือเศษน้อยกว่าค่า m เลขที่เหลือเศษจึงเป็นเลขคล้ายสุ่มตัวต่อไปคือ x_{i+1}

วิธีการสร้างตัวเลขสุ่มแบบนี้เริ่มต้นโดยกำหนดค่าเริ่มต้น x เรียกว่า Initial Value หรือ Seed จากการใช้สมการที่ (1) จะได้เลขคล้ายสุ่มที่เป็นเลขจำนวนเต็มค่าใดค่าหนึ่งในช่วง 0,1,...,m-1 หลังจากนั้นแล้วจะได้เลขคล้ายสุ่มชุดเดิมอีก ซึ่งคาบของเลขคล้ายสุ่มที่ได้มีค่าไม่เกิน m (โดยอาจมีค่าน้อยกว่า m ก็ได้ เมื่อเลือกค่า a และ x ไม่เหมาะสม) การเลือกค่า m, a และ x จึงมีความสำคัญในการผลิตเลขคล้ายสุ่มที่มีคาบใกล้เคียงกับค่า m มากที่สุด

Lehmer ได้ทดลองเลือกใช้ค่า m, a และ x ที่จับคู่ต่าง ๆ กันเพื่อใช้ผลิตเลขคล้ายสุ่มตามสมการที่ (1) พบว่าถ้าเลือก x เป็นเลขคี่ และ $m = 2^r$ (เมื่อ r>2) และ $a = 8k+3$ (เมื่อ k เป็นจำนวนเต็มบวกใด ๆ) จะได้คาบของเลขคล้ายสุ่มมากที่สุด

และเท่ากับ 2^{r-2} วิธีการต่อไปนี้เป็นวิธีการเลือกค่าพารามิเตอร์ทั้ง 3 ตัว เพื่อจะได้กลุ่มของเลขคล้ายสุ่มที่ดี

จากหลักการดังกล่าวจะได้ตัวเลขสุ่มที่มีการแจกแจงแบบยูนิฟอร์มในช่วง (0,1) ซึ่งเขียนเป็นโปรแกรมด้วยภาษาฟอร์แทรนได้ดังนี้

```
      SUBROUTINE RANDUM(IX, IY, RN)
      IY = IX * 16807
      IF (IY) 3,4,4
    3 IY = IY + 2147483647 +1
    4 RN = IY
      RN = RN * 0.46566E-9
      IX = IY
      RETURN
      END
```

เมื่อ IX คือ ค่าเริ่มต้น

IY คือ เลขสุ่มตัวถัดไปที่คำนวณได้จากค่าเริ่มต้น

RN คือ เลขสุ่มที่มีการแจกแจงแบบยูนิฟอร์มในช่วง (0,1)

1.2 <u>แปลงตัวเลขสุ่มให้มีการแจกแจงแบบปกติ</u> โดยใช้เทคนิคการแปลงโดยตรงจากสมการ

$$\phi(x) = \int_{-\alpha}^{\alpha} (2\pi)^{-1/2} \exp(-x^2/2)\, dx \quad , -\alpha < x < \alpha$$

Box และ Muller (1958) ได้สร้างเลขสุ่มที่มีการแจกแจงแบบปกติมาตรฐานพร้อม ๆ กัน 2 ค่าที่เป็นอิสระต่อกัน โดยใช้ตัวผลิต (Generator) z_1 และ z_2 ดังนี้

$$z_1 = (-2\ln R_1)^{1/2}\cos(2\pi R_2)$$

$$z_1 = (-2\ln R_1)^{1/2}\sin(2\pi R_2)$$

R_1 และ R_2 เป็นตัวเลขสุ่มที่สร้างจากโปรแกรมย่อย RANDUM เมื่อได้เลขสุ่มที่มีการแจกแจงแบบปกติมาตรฐานแล้ว ทำการแปลงค่าเลขสุ่มดังกล่าวโดยใช้ฟังก์ชัน

$$z_1' = \mu + \sigma z_1$$

$$z_2' = \mu + \sigma z_2$$

จะได้ว่า z_1' และ z_2' มีการแจกแจงแบบปกติมีค่าเฉลี่ย μ และความแปรปรวน σ^2

โปรแกรมย่อยที่ใช้ในการสร้างตัวเลขสุ่มให้มีการแจกแจงแบบปกติ มีค่าเฉลี่ย μ และความแปรปรวน σ^2 คือ FUNCTION NORMAL(EX, STD) ซึ่งเขียนได้ดังนี้

```
      FUNCTION NORMAL(EX, STD)
      REAL NORMAL
      COMMON IX
      DATA K/0/, PI/3.14159/
      IF (K.EQ.1) GOTO 10
           CALL RANDUM(IX, IY, RN)
                RONE = RN
           CALL RANDUM(IX, IY, RN)
                RTWO = RN
       ZONE = SQRT(-2.*ALOG(RONE))*COS(2.*PI*RTWO)
       ZTWO = SQRT(-2.*ALOG(RONE))*SIN(2.*PI*RTWO)
            NORMAL = ZONE * STD + EX
       K = 1
       RETURN
   10       NORMAL = ZTWO * STD + EX
       K = 0
```

```
      RETURN
      END
```

1.3 <u>แปลงตัวเลขสุ่มให้มีการแจกแจงแบบเบ้</u> โปรแกรมย่อยที่ใช้ในการสร้างการแจกแจงแบบเบ้นั้น ใช้การแปลงข้อมูลของ Ramberg และ Schineiser ซึ่งเรียกว่า Generalized Lambda Distribution (GLD) การสร้างการแจกแจงแบบ GLD นั้นใช้การแปลงข้อมูลในลักษณะดังนี้

$$x = \lambda_1 + (p^{\lambda_3} - (1-p)^{\lambda_4})/\lambda_2 \qquad ,0 < p < 1 \quad .....(2)$$

เมื่อ p = ตัวเลขสุ่ม (RN) มีค่าอยู่ระหว่าง 0 ถึง 1

λ_1 = พารามิเตอร์ที่กำหนดตำแหน่ง

λ_2 = พารามิเตอร์ที่กำหนดสเกล

λ_3, λ_4 = พารามิเตอร์ที่กำหนดรูปแบบการแจกแจง

ในการวิจัยครั้งนี้จะสร้างการแจกแจงแบบ GLD ซึ่งมีความเบ้และความโด่งตามต้องการ โดยมีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น 1 ก่อน ถ้าต้องการให้ข้อมูลมีค่าเฉลี่ย μ และความแปรปรวน σ^2 จะต้องแปลงค่า λ_1 และ λ_2 ในตารางของ Ramberg ก่อนที่จะแทนในสมการที่ (2) ดังนี้คือ

$$\lambda_1(EX, STD) = \lambda_1(0,1)\,\sigma + \mu$$

$$\lambda_2(EX, STD) = \lambda_2(0,1)/\sigma$$

ส่วน λ_3 และ λ_4 จะกำหนดค่าต่าง ๆ กันตามความเบ้และความโด่งของการแจกแจงที่ต้องการ

โปรแกรมย่อยที่ใช้ในการสร้างตัวเลขสุ่มให้มีการแจกแจงแบบเบ้ คือ FUNCTION SKEWED(RLM1, RLM2, RLM3, RLM4, EX, STD) ซึ่งเขียนได้ดังนี้

```
      FUNCTION SKEWED(RLM1, RLM2, RLM3, RLM4, EX, STD)
      COMMON IX
          CALL RANDUM(IX, IY, RN)
      R1 = RLM3 * ALOG(RN)
      R2 = RLM4 * ALOG(1-RN)
```

```
      RX1 = EXP(R1)
      RX2 = EXP(R2)
   X1 = RLM1 + (RX1 - RX2)/RLM2
      SKEWED = EX + STD * X1
   RETURN
   END
```

2. คำนวณความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนชนิดที่ 1 หรือค่าอำนาจการทดสอบของตัวสถิติทั้ง 6 ตัว ตามขั้นตอนต่อไปนี้

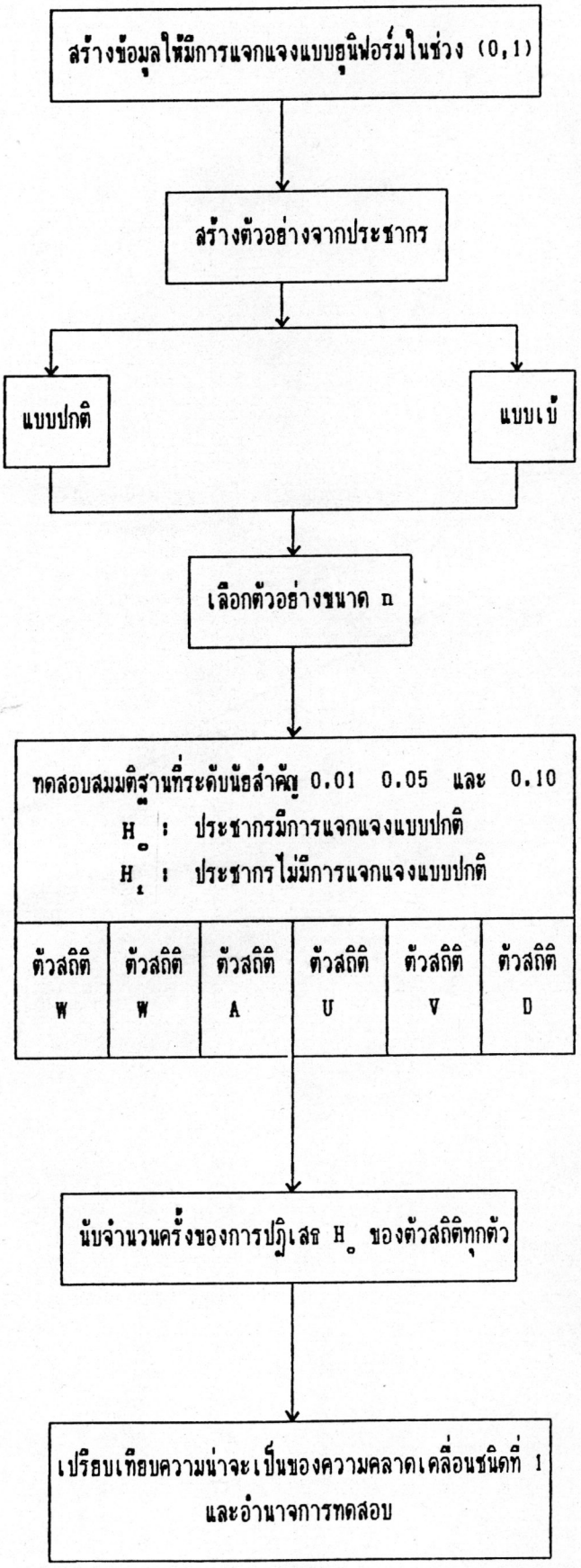
2.1 เขียนคำสั่งให้คอมพิวเตอร์อ่านค่าตัวแปรต่าง ๆ ที่ใช้เป็นตัวกำหนดสถานการณ์ในแผนการทดลอง ได้แก่ ขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนของประชากรที่ต้องการ

2.2 สร้างข้อมูลให้มีการแจกแจงตามแผนการทดลองที่กำหนดไว้ แล้วนำข้อมูลที่ได้มาคำนวณค่าสถิติ และนำค่าสถิติที่ได้มาเปรียบเทียบกับค่าวิกฤตเพื่อจะได้ตัดสินใจว่าจะยอมรับหรือปฏิเสธสมมติฐานว่าง ในกรณีที่ปฏิเสธสมมติฐานว่างให้นับจำนวนไว้ด้วย ทำการคำนวณและนับจำนวนครั้งของการปฏิเสธสมมติฐานว่างจนครบทุกตัวสถิติทดสอบ จากนั้นก็ย้อนกลับไปสุ่มตัวอย่างชุดใหม่จนครบ 1,000 ครั้ง

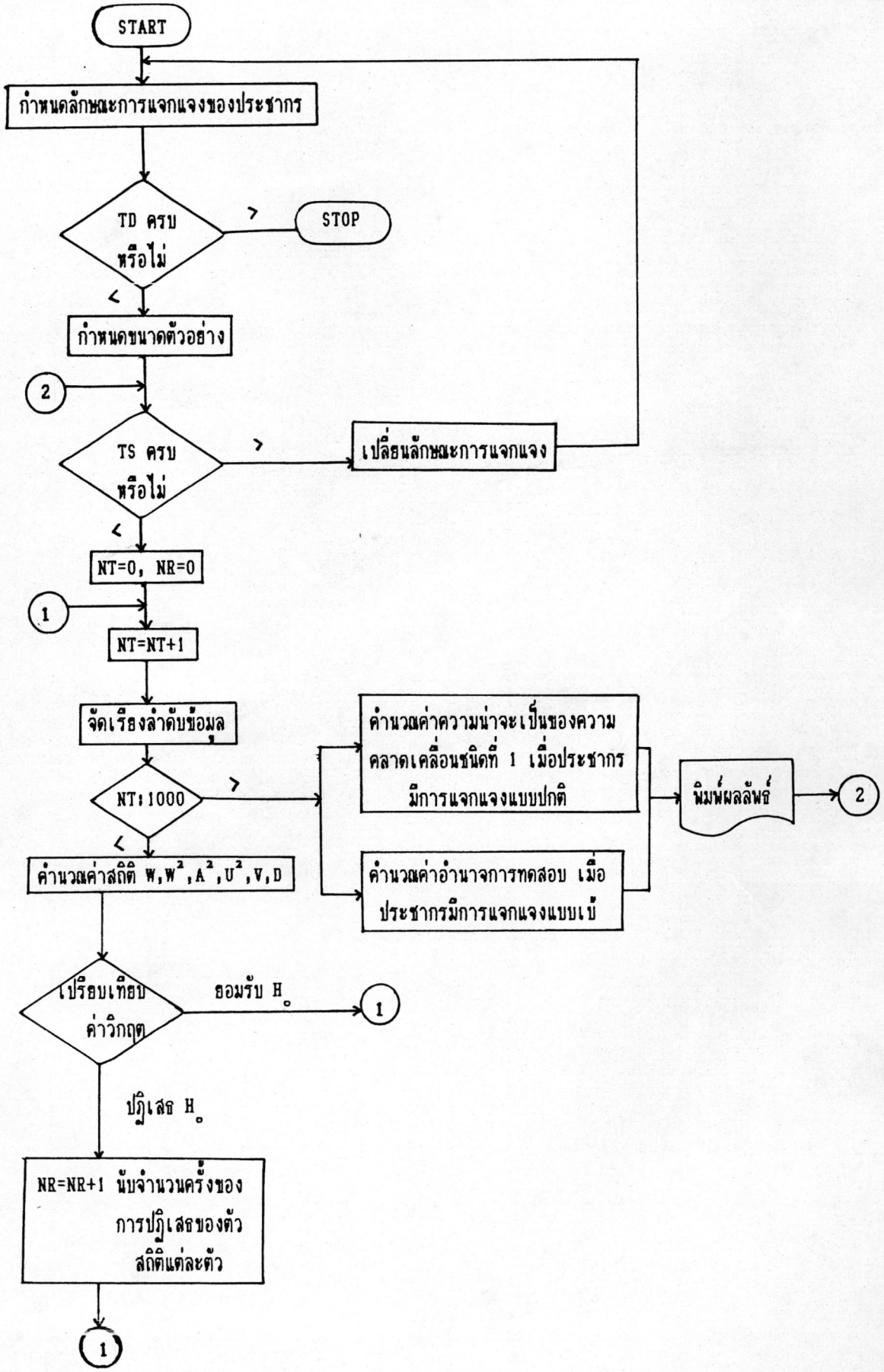
2.3 คำนวณความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนชนิดที่ 1 เมื่อประชากรมีการแจกแจงแบบปกติ เมื่อทำครบวงจรแล้วต่อไปก็เปลี่ยนขนาดตัวอย่างจนครบทุกรูปแบบที่ต้องการศึกษา จากนั้นก็คำนวณค่าอำนาจการทดสอบเมื่อประชากรมีการแจกแจงแบบเบ้ โดยทำให้ครบทุกค่าของความเบ้ ความโด่ง และขนาดตัวอย่างที่ต้องการศึกษา ซึ่งสามารถแสดงเป็นผังงานได้ดังนี้

i17229724

<u>แผนผังแสดงขั้นตอนการดำเนินการวิจัย</u>



แผนผังโปรแกรมคอมพิวเตอร์ที่คำนวณความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 และอำนาจการทดสอบ



เมื่อ TD แทน จำนวนการแจกแจงทั้งหมด

TS แทน จำนวนขนาดตัวอย่างทั้งหมด

NT แทน จำนวนรอบของการทำงาน

NR แทน จำนวนครั้งของการปฏิเสธ