บทที่ 3

วิธีดำเนินการวิจัย

ในการศึกษาการวิเคราะห์ความถดถอยเมื่อตัวแปรตามบางค่ามีค่าขาดหายได้ใช้เทคนิคการซึมเลทส์ร้างข้อมูล เพื่อให้ได้ข้อมูลที่มีคุณสมบัติตรงตามที่ต้องการ โดยได้ใช้โปรแกรมภาษาฟอร์แทรน (FORTRAN 77) กับเครื่อง UNIVAC 1100/60 สำหรับขั้นตอนการวิจัยมีดังต่อไปนี้

3.1 วิธีการศึกษา

ในการศึกษาครั้งนี้ ต้องการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ ในการวิเคราะห์ความถดถอยเมื่อตัวแปรตามบางค่ามีค่าขาดหายของวิธีการกำลังสองต่ำสุด, วิธีการของมิลเลอร์, วิธีการของบัคเลย์และเจมส์ ซึ่งจะใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) เป็นตัวเปรียบเทียบ ตัวประมาณที่ได้จากวิธีการใดให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด จะถือว่าเป็นตัวประมาณที่ดีที่สุด ในการศึกษาครั้งนี้ ได้ทำการศึกษาเมื่อขนาดตัวอย่างมีค่าเป็น 10, 20, 50, 60, 100 และ 150 ซึ่งในแต่ละขนาดตัวอย่างจะให้เกิดเปอร์เซ็นต์เฉลี่ยของค่าขาดหายเป็น 5%, 10%, 15%, 20%, 25%, 30% 50% 60% 70% 80% และในแต่ละขนาดตัวอย่างจะให้ค่าความคลาดเคลื่อนมีการแจกแจงเป็นแบบปกติที่มีค่าเฉลี่ยเป็น 0 และค่าความแปรปรวนเป็น 16 และให้ข้อมูลที่ขาดหายมีการแจกแจงเป็นแบบยูนิฟอร์ม แบบแกมม่า แบบปกติ และเป็นเชิงเส้นกับค่าความคลาดเคลื่อน สำหรับวิธีการดำเนินการศึกษามีดังต่อไปนี้

1. สร้างข้อมูล X_i จากตัวเลขสุ่มในช่วง (0, 1) โดยใช้โปรแกรมย่อย RANDU
2. สร้างค่าความคลาดเคลื่อน ε ให้มีการแจกแจงเป็นแบบปกติที่มีค่าเฉลี่ยเป็น 0 และค่าความแปรปรวนเป็น 16 โดยใช้โปรแกรมย่อย NORMAL
3. กำหนดค่าพารามิเตอร์ $\alpha = 30$, $\beta = 20$
4. สร้างค่าที่ไม่ขาดหาย จากความสัมพันธ์

$$T_i = \alpha + \beta X_i + \varepsilon_i$$

5. สร้างค่าขาดหาย C_i ให้มีการแจกแจงแบบยูนิฟอร์ม โดยใช้โปรแกรมย่อย UNIFOR แบบแกมม่า โดยใช้โปรแกรมย่อย GAMMA แบบปกติ โดยใช้โปรแกรมย่อย NORMAL แบบเชิงเส้นกับค่าความคลาดเคลื่อนซึ่งสร้างจาก $C_i = \alpha_1 + \beta_1 X_i + \varepsilon_i$ เมื่อ ε มีการ

แจกแจงแบบปกติ ที่มีค่าเฉลี่ย 0 และค่าความแปรปรวน σ^2

6. สร้างค่า Y_i จาก $Y_i = \min(T_i, C_i)$ และให้ δ_i มีค่าสัมพันธ์กับ Y_i ดังนี้

$$\delta_i = \begin{cases} 1 & \text{ถ้า } Y_i = T_i \text{ เป็นข้อมูลที่ไม่ขาดหาย} \\ 0 & \text{ถ้า } Y_i = C_i \text{ เป็นข้อมูลที่ขาดหาย} \end{cases}$$

7. หาค่าเปอร์เซ็นต์ไทล์ของค่าขาดหายให้ได้เป็น 5% 10% 15% 20% 25% 30% 50% 60% 70% และ 80%

8. คำนวณค่าประมาณด้วยวิธีการกำลังสองต่ำสุด วิธีการของมิลเลอร์ วิธีการของบัคเลย์และเจมส์

9. คำนวณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

3.2 การสร้างข้อมูลสำหรับใช้ในการศึกษา

3.2.1 การสร้างตัวเลขสุ่ม (Random Number) ในการสร้างลักษณะการแจกแจงแบบต่าง ๆ นั้นจะต้องอาศัยตัวเลขสุ่ม ซึ่งมีการแจกแจงแบบสม่ำเสมอในช่วง (0, 1) เป็นพื้นฐานในการสร้าง Shanon (1975:352-356) ได้เสนอวิธีการสร้างเลขสุ่มดังนี้

1. เลือกเลขคี่บางตัวซึ่งน้อยกว่า 9 หลักเป็นค่าเริ่มต้น

2. คูณตัวเลขที่กำหนดเป็นค่าเริ่มต้นด้วยค่า a ซึ่งเป็นเลขจำนวนเต็มอย่างน้อย 5 หลัก (ค่าของ a โดยปกติจะเท่ากับ $2^{B/2} \pm 3$ เมื่อ B คือ จำนวน bit ของเครื่องคอมพิวเตอร์ สำหรับเครื่อง 32 bit เช่น IBM 370/3031 ค่า a จะเท่ากับ 65539 สำหรับเครื่อง 36 bit เช่น UNIVAC ค่า a จะเท่ากับ 262147)

3. บวกผลคูณในขั้นตอนที่ 2 ด้วยค่า m+1 (ค่าของ m โดยปกติจะเท่ากับ 2^{B-1} สำหรับเครื่อง 32 bit ค่า m จะเท่ากับ 2147483647 สำหรับเครื่อง 36 bit ค่า m จะเท่ากับ 34359738367)

4. คูณผลลัพธ์ในขั้นตอนที่ 3 ด้วยเศษที่มีค่า 1/m

5. จากขั้นตอนที่ 4 ก็จะได้ค่าตัวเลขสุ่มที่มีค่าในช่วง (0, 1)

6. กำหนดค่าเริ่มต้นใหม่ ให้มีค่าเท่ากับผลบวกในขั้นตอนที่ 3

7. กระทำซ้ำ ๆ กันจากขั้นตอนที่ 2 ถึง 6 จนกระทั่งได้ค่าตัวเลขสุ่มครบตามต้องการ

สำหรับในการศึกษาครั้งนี้ ได้ใช้เครื่องคอมพิวเตอร์ขนาด 36 bit โปรแกรมย่อยสำหรับการสร้างเลขสุ่มสามารถแสดงได้ดังนี้

```
      SUBROUTINE  RANDU (IX, IY, RN)
      IY  =  IX * 262147
      IF (IY)  3, 4, 4
3     IY  =  IY + 34359738367 + 1
4     RN  =  IY
      RN  =  RN/34359738367
      IX  =  IY
      RETURN
      END
```

3.2.2 การสร้างการแจกแจงแบบยูนิฟอร์มในช่วง (A, B)

การสร้างตัวเลขสุ่มที่มีการแจกแจงแบบยูนิฟอร์มในช่วง (A, B) นั้นใช้วิธี Inverse Transformation ซึ่งแสดงได้ดังนี้

พิจารณาฟังก์ชั่นความน่าจะเป็นของการแจกแจงแบบยูนิฟอร์ม

$$f(x) = \frac{1}{b-a} \qquad ; a < x < b$$

หาค่าฟังก์ชั่นการแจกแจงสะสม F(X)

$$F(X) = \int_a^x f(x)\,dx$$

$$= \int_a^x \frac{1}{b-a}\,dx$$

$$= \frac{1}{b-a} x \Big|_a^x$$

$$= \frac{x-a}{b-a}$$

$$X = bF(X) - aF(X) + a$$

$$= a + (b-a)F(X)$$

เนื่องจาก F(X) มีการแจกแจงแบบยูนิฟอร์มในช่วง (0, 1)(Gibbon : 1971 : 23) ดังนั้น F(X) ก็คือค่า RN จากโปรแกรมย่อย RANDU ซึ่งโปรแกรมย่อย RANDU นี้ ใช้สร้างตัวเลขสุ่มที่มีการแจกแจงแบบยูนิฟอร์มในช่วง (0, 1) ดังนั้น โปรแกรมย่อยซึ่งใช้สร้างการแจกแจงแบบยูนิฟอร์มในช่วง (A, B) แสดงได้ดังนี้

```
SUBROUTINE  UNIFOR (A, B, XU)
CALL  RANDU (IX, IY, RN)
XU  =  A + (B-A) * RN
RETURN
END
```

3.2.3 การสร้างการแจกแจงแบบปกติ

วิธีการสร้างการแจกแจงแบบปกติ ในการศึกษาครั้งนี้จะใช้วิธีการของ Gauss ซึ่งเป็นวิธีสร้างการแจกแจงปกติที่มีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น 1 ส่วนค่าเฉลี่ยและค่าความแปรปรวนอื่น จะใช้วิธีการแปลงข้อมูลในรูป X = AMEAN + (SIGMA)X โดยที่ AMEAN คือ ค่าเฉลี่ย และ $(SIGMA)^2$ คือ ค่าความแปรปรวนที่ต้องการ

พิจารณาจากสูตร

$$X = \frac{\sum_{i=1}^{k} RN_i - \frac{K}{2}}{\frac{K}{12}}$$

โดยที่ X เป็นตัวเลขสุ่มที่มีการแจกแจงแบบปกติมาตรฐานที่มีค่าเฉลี่ย 0 และค่าความแปรปรวน 1

RN_i เป็นตัวเลขสุ่มที่มีการแจกแจงแบบยูนิฟอร์ม ในช่วง (0, 1) จากโปรแกรมย่อย RANDU

K เป็นจำนวนค่าของ RN_i ที่ใช้ในการประมาณค่า X

โดยปกติแล้ว ตัวเลขสุ่ม X จะมีค่าเข้าใกล้เลขสุ่มที่มีการแจกแจงแบบปกติ เมื่อค่าของ K ใหญ่มากคือเข้าใกล้ค่าอนันต์ (Infinity) แต่เพื่อให้เครื่องลดเวลาในการคำนวณ (Execution) จะใช้ K = 12 ดังนั้น จากสูตรในการคำนวณข้างต้นจะได้สูตรใหม่เป็น

$$X = \sum_{i=1}^{12} RN_i - 6.0$$

เพื่อให้ตัวเลขสุ่ม X มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยและค่าความแปรปรวนตามที่ต้องการ จะแปลงค่าตัวเลขสุ่ม X ใหม่ดังนี้

$$X' = X + S + AM$$

เมื่อ S เป็นส่วนเบี่ยงเบนมาตรฐานตามที่ต้องการ

AM เป็นค่าเฉลี่ยตามที่ต้องการ

โปรแกรมย่อยที่ใช้สร้างการแจกแจงแบบปกติ แสดงได้ดังนี้

```
SUBROUTINE  NORMAL (AMEAN, SIGMA, XN)
A  =  0
DØ  10   I  =  1, 12
CALL RANDU (IX, IY, RN)
A  =  A + RN
10   CONTINUE
XN  =  (A-6.0) * SIGMA + AMEAN
RETURN
END
```

3.2.4 การสร้างการแจกแจงแบบแกมม่า

การแจกแจงแบบแกมม่า เป็นการแจกแจงซึ่งมีฟังก์ชั่นความน่าจะเป็นดังนี้

$$f(x) = \frac{1}{\Gamma(\beta)} x^{\beta-1} \frac{e^{-x/\alpha}}{\alpha^{\beta}} \quad ; 0 < x < \infty \quad ; \alpha, \beta > 0$$

การสร้างตัวแปรสุ่มที่มีการแจกแจงแบบแกมม่า จะใช้วิธี Inverse Transformation ซึ่งสามารถแสดงเป็นโปรแกรมย่อยได้ดังนี้

```
      SUBROUTINE  GAMMA (ALPHA, BETA, XG)
      XG  =  0.0
      V   =  1.0
    5 CALL RANDU (IX, IY, RN)
      V   =  V *  RN
      IF ( BETA .EQ. 1.0) GOTO 10
      BETA  =  BETA  - 1.0
      GOTO 5
   10 XG   =  ALPHA * ALOG (V)
      RETURN
      END
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย