บทที่ 3

วิธีดำเนินการวิจัย

ในการประมาณค่าพารามิเตอร์ ในสมการถดถอยเชิงเส้นพหุ เมื่อตัวแปรตามบางค่ามีค่าขาดหาย ได้ใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Technique) สร้างข้อมูลเพื่อให้ได้ข้อมูลที่มีคุณสมบัติตรงตามที่ต้องการ

3.1 วิธีจำลองโดยใช้เทคนิคมอนติการ์โล

เทคนิคมอนติการ์โล เป็นวิธีการหนึ่งที่นิยมนำมาใช้กันอย่างแพร่หลายในปัจจุบัน โดยจะอาศัยเลขสุ่ม (Random Numbers) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา โดยมีขั้นตอนดังนี้

3.1.1 สร้างตัวเลขสุ่ม ที่มีการแจกแจงแบบสม่ำเสมอในช่วง (0,1) ซึ่งตัวเลขสุ่มแต่ละตัวเป็นอิสระต่อกันและมีช่วงยาวก่อนจะเกิดเลขสุ่มซ้ำ White และ Schmidt (1975) ได้เสนอวิธีการสร้างเลขสุ่มดังนี้

1. เลือกเลขคี่บางตัวซึ่งน้อยกว่า 9 หลัก เป็นค่าเริ่มต้น
2. คูณตัวเลขที่กำหนดเป็นค่าเริ่มต้นด้วยค่า a ซึ่งเป็นเลขจำนวนเต็มอย่างน้อย 5 หลัก (โดยปกติแล้ว a จะเท่ากับ $2^{b/2} \pm 3$ เมื่อ B คือจำนวน bit ของเครื่องคอมพิวเตอร์ สำหรับเครื่อง 32 bit a จะเท่ากับ 65539 สำหรับเครื่อง 36 bit ค่า a จะเท่ากับ 262147)
3. บวกผลคูณในขั้นที่ 2 ด้วยค่า m+1 (ค่าของ m โดยปกติจะเท่ากับ 2^{B-1} สำหรับเครื่อง 32 bit m จะเท่ากับ 2147483647 สำหรับเครื่อง 36 bit m จะเท่ากับ 34359738367)
4. คูณผลลัพธ์ในขั้นที่ 3 ด้วย เศษที่มีค่า $\frac{1}{m}$
5. จากขั้นที่ 4 ก็จะได้ค่าตัวเลขสุ่มที่มีค่าในช่วง (0,1)
6. กำหนดค่าเริ่มต้นใหม่ ให้มีค่าเท่ากับผลบวกในขั้นตอนที่ 3
7. กระทำซ้ำ ๆ กันจากขั้นตอนที่ 2 ถึง 6 จนกระทั่งได้ตัวเลขสุ่มครบตามจำนวนที่ต้องการ

3.1.2 การนำตัวเลขสุ่มมาประยุกต์ใช้กับปัญหาที่ต้องการศึกษา ขั้นตอนนี้ขึ้นอยู่กับลักษณะของปัญหา บางปัญหาอาจจะไม่ใช้เลขสุ่มโดยตรง แต่จะนำไปผลิตเลขสุ่มที่มีการแจกแจงแบบอื่นต่อไป

3.1.3 การทดลองกระทำ เมื่อนำตัวเลขสุ่มมาประยุกต์ให้เข้ากับปัญหาที่ต้องการศึกษาได้แล้ว ขั้นตอนต่อไปคือการทดลองโดยใช้กระบวนการของการสุ่มมากระทำในลักษณะซ้ำ ๆ กันหลาย ๆ ครั้ง เพื่อหาคำตอบที่ต้องการ

3.2 แผนการทดลอง

ในการวิจัยครั้งนี้ ต้องการประมาณค่าพารามิเตอร์ ในสมการถดถอยเชิงเส้นพหุ เมื่อค่าสังเกตของตัวแปรตาม เป็นค่าที่ถูกตัดทิ้งทางขวาและเป็นการตัดทิ้งประเภทที่ 1 โดยประมาณค่าพารามิเตอร์ด้วยวิธีการสมิธ วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด และวิธีการโมดิไฟ-แอคชัวเรียล ซึ่งจะใช้ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองต่ำสุด (RMSE) เป็นตัวเปรียบเทียบ ตัวประมาณที่ได้จากวิธีการใด ให้ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองต่ำสุด จะถือว่าเป็นตัวประมาณที่ดีที่สุด ในการศึกษาครั้งนี้ได้ทำการศึกษาเมื่อขนาดตัวอย่างมีค่าเป็น 10 20 30 50 และ 100 ซึ่งในแต่ละขนาดตัวอย่างจะให้เกิดเปอร์เซนต์เฉลี่ยของค่าที่ถูกตัดทิ้งเป็น 0.1,0.2,0.3,0.4,0.5 รวมทั้งสิ้น 25 สถานการณ์
รายละเอียดของแผนทดลองมีดังนี้

1. ค่าสังเกตที่ถูกตัดทิ้งและไม่ถูกตัดทิ้ง $Y_i = \beta X_i + \varepsilon_i$; i = 1,..,nm
โดยที่ $X \sim N(0, 25)$ และ $\varepsilon \sim N(0, 25)$ กำหนด $\beta = 1$
2. ขนาดตัวอย่าง 5 ขนาด คือ 10,20,30,50 และ 100
3. สัดส่วนของข้อมูลที่ถูกตัดทิ้ง 10 ระดับ คือ 0.1,0.2,0.3,0.4,0.5

3.3 ขั้นตอนในการวิจัย

tc เป็นค่าที่กำหนดไว้ล่วงหน้า

T_i เป็นตัวแปรสุ่มที่เกิดจากความสัมพันธ์เชิงเส้น $T_i = \beta X_i + \varepsilon_i$; i=1,..,nm
และค่าสังเกต T_i มีการแจกแจงที่เหมือนกันและเป็นอิสระกัน ดังนั้นค่าสังเกตุ Y_i ได้จาก

$$Y_i = \begin{cases} T_i & \text{ถ้า } T_i \leq tc \text{ (ค่าสังเกตที่ไม่ถูกตัดทิ้ง)} \\ tc & \text{ถ้า } T_i > tc \text{ (ค่าสังเกตที่ถูกตัดทิ้ง)} \end{cases}$$

$$\delta_i = \begin{cases} 1 & \text{กรณีที่ไม่ถูกตัดทิ้ง} \\ 0 & \text{กรณีที่ถูกตัดทิ้ง} \end{cases}$$

การจำลองข้อมูลเพื่อให้เกิดค่าสังเกตที่ถูกตัดทิ้งและไม่ถูกตัดทิ้ง แสดงได้ดังนี้

1. กำหนดค่า tc ดังนี้

จากความสัมพันธ์เชิงเส้น $Y= \beta X + \varepsilon$ หาค่าสังเกตตัวที่ 75 เปอร์เซนต์ (Y_{75}) ให้ค่าที่จุดนี้เป็น tc ของกลุ่มที่ 1 จากนั้นหาค่า tc แบบเดิมอีกให้ครบ 1000 กลุ่ม นำค่า tc ทั้ง 1000 กลุ่ม นี้มาหาค่าเฉลี่ย และใช้ค่าเฉลี่ยที่ได้นี้เป็นค่าที่กำหนดไว้ล่วงหน้า

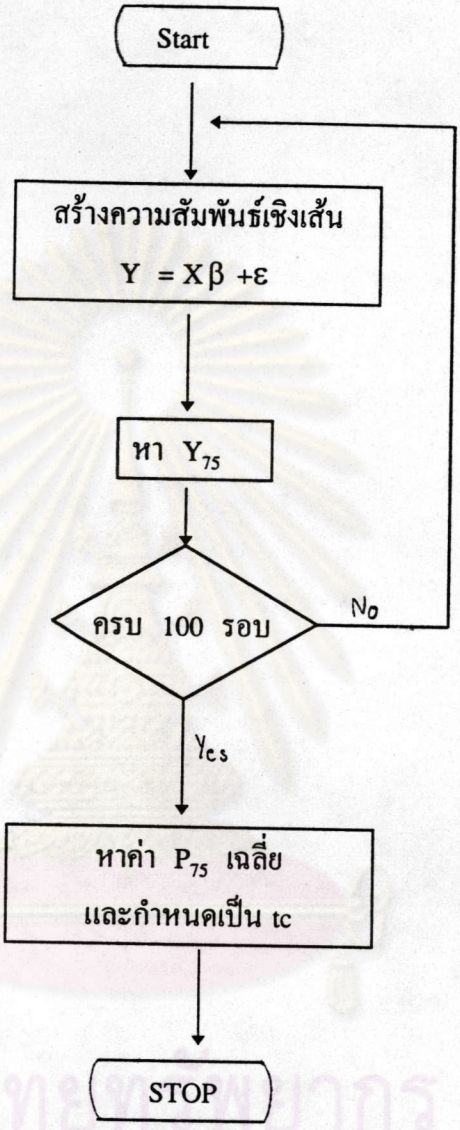
2. จำลองตัวแปรอิสระ X และค่าความคลาดเคลื่อน ε_i จากการแจกแจงปกติ N(0, 25)และจำลอง T จากความสัมพันธ์เชิงเส้น $T = \beta X + \varepsilon$

3. หาค่าสังเกตของตัวแปรตาม Y โดยนำค่า T เปรียบเทียบกับ tc ถ้า $T \leq tc$ ค่าสังเกตที่ได้นี้จะเป็นค่าสังเกตที่ไม่ถูกตัดทิ้ง ให้ $\delta = 1$ ถ้า $T > tc$ จะเป็นค่าสังเกตที่ถูกตัดทิ้ง ให้ $\delta = 0$ ทำการจำลองเช่นนี้จนกระทั่งได้ค่าสังเกตที่ไม่ถูกตัดทิ้งและค่าสังเกตที่ถูกตัดทิ้ง ครบตามจำนวนที่กำหนด

4. นำค่า Y,X หาค่าประมาณพารามิเตอร์ จากสมการถดถอยเชิงเส้นพหุ ด้วยวิธีการทั้ง 3 วิธี และทำการจำลองข้อมูลที่ประมาณพารามิเตอร์ได้ทั้ง 3 วิธี จำนวน 800 ชุดข้อมูลในแต่ละสถานการณ์

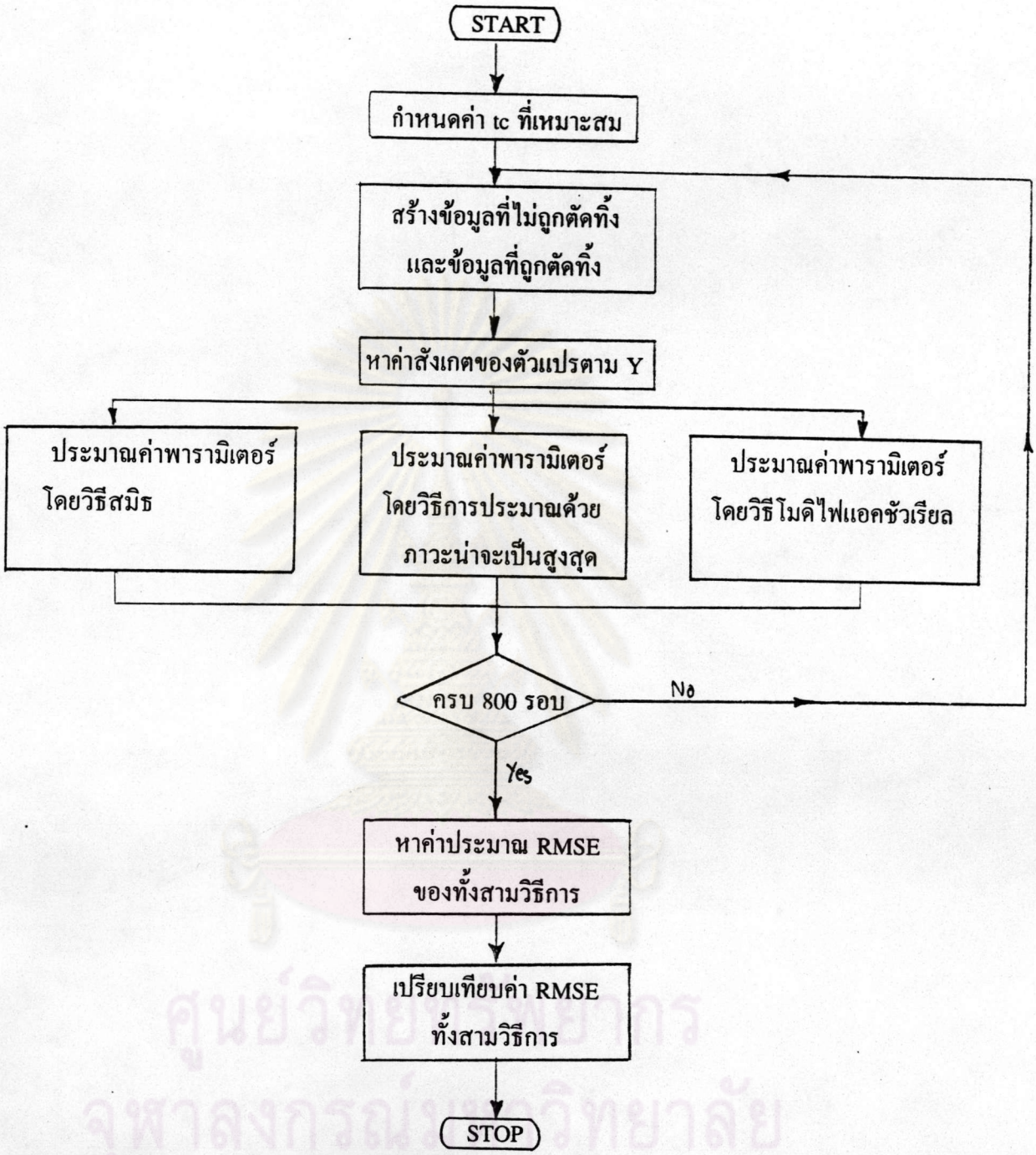
5. หาค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ทั้ง 3 วิธี (เนื่องจากในการทดลองได้จำลองข้อมูลซ้ำ ๆ กัน จำนวน 800 ครั้ง ในแต่ละสถานการณ์) เปรียบเทียบค่า RMSE ของทั้ง 3 วิธี วิธีการใดให้ค่า RMSE ต่ำที่สุด ถือว่าเป็นวิธีที่ ประมาณค่าพารามิเตอร์ของแต่ละสถานการณ์ได้ดีที่สุด

ซึ่งขั้นตอนของการวิจัยดังกล่าวนี้ สามารถสรุปเป็นผังงานได้ดังรูป



รูปที่ 3.1 แสดงผังงานสำหรับหาค่า tc

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



รูป 3.2 แสดงผังงานสำหรับหาค่าความคลาดเคลื่อนจากการประมาณทั้ง 3 วิธีการ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย