บทที่ 3

การดำเนินการวิจัย

การวิจัยครั้งนี้มีลักษณะเป็นการวิจัยเชิงทดลอง เพื่อต้องการศึกษาและเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์หรือสัมประสิทธิ์การถดถอยในสมการถดถอยเชิงเส้นพหุ เมื่อข้อมูลมีค่าผิดปกติ โดยวิธีการประมาณค่าสัมประสิทธิ์การถดถอยที่นำมาศึกษาได้แก่

1. วิธีกำลังสองน้อยที่สุด (Ordinary Least Squares Method : OLS Method)
2. วิธีตัวประมาณ M (M-Estimator Method) เมื่อใช้เกณฑ์ความแกร่งของ Huber
3. วิธีตัวประมาณ M (M-Estimator Method) เมื่อใช้เกณฑ์ความแกร่งของ Tukey
4. วิธีตัวประมาณ Bounded-Influence (Bounded-Influence Estimator Method : BI Estimator Method) เมื่อใช้เกณฑ์ความแกร่งของ Huber
5. วิธีตัวประมาณ Bounded-Influence (Bounded-Influence Estimator Method : BI Estimator Method) เมื่อใช้เกณฑ์ความแกร่งของ Tukey

ผู้วิจัยจะทำการเปรียบเทียบความสามารถของการประมาณค่าสัมประสิทธิ์การถดถอยของวิธีการดังกล่าว โดยจะเปรียบเทียบจากค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของการประมาณค่าสัมประสิทธิ์การถดถอยจากทุกวิธี ซึ่งแต่ละวิธีการจะประมาณค่าสัมประสิทธิ์การถดถอยภายใต้สถานการณ์ต่าง ๆ ที่กำหนดให้มีขนาดตัวอย่าง 3 ขนาด คือ 20, 30 และ 50 และในกรณีที่ข้อมูลเกิดค่าผิดปกติแบบต่าง ๆ ซึ่งจำแนกตามตัวแปรที่เกิดค่าผิดปกติและตามตำแหน่งของการเกิดค่าผิดปกติ ดังนี้คือ

1. กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรตาม (ความคลาดเคลื่อน)
2. กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระและตัวแปรตาม (ความคลาดเคลื่อน) โดยข้อมูลในลักษณะนี้จะถูกแบ่งออกเป็น 3 กลุ่มตามชนิดของตัวแปรที่เกิดค่าผิดปกติ คือ

   2.1 กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระ x_1 และตัวแปรตาม

2.2 กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระ x_2 และตัวแปรตาม

2.3 กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระ x_1 ตัวแปรอิสระ x_2 และตัวแปรตาม

3. กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระและตัวแปรตาม (ความคลาดเคลื่อน) แบบกำหนดให้เกิด ณ ตำแหน่งเดียวกัน โดยข้อมูลในลักษณะนี้จะถูกแบ่งออกเป็น 3 กลุ่มตามชนิดของตัวแปรที่เกิดค่าผิดปกติ คือ

3.1 กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระ x_1 และตัวแปรตาม ณ ตำแหน่งเดียวกัน

3.2 กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระ x_2 และตัวแปรตาม ณ ตำแหน่งเดียวกัน

3.3 กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระ x_1 หรือตัวแปรอิสระ x_2 และตัวแปรตาม ณ ตำแหน่งเดียวกัน

เนื่องจากในการวิจัยครั้งนี้ ผู้วิจัยได้อาศัยเทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Technique) มาสร้างสถานการณ์ต่าง ๆ ดังนั้นตอนแรกผู้วิจัยจะกล่าวถึงวิธีการจำลองโดยใช้เทคนิคมอนติคาร์โล แล้วจึงแสดงรายละเอียดของขั้นตอนการวิจัยตามลำดับ ส่วนรายละเอียดของโปรแกรมที่ใช้ในการวิจัยจะแสดงไว้ในภาคผนวก ก. และ ข.

วิธีจำลองโดยใช้เทคนิคมอนติคาร์โล

เทคนิคที่ใช้แก้ปัญหาในการคำนวณทางคณิตศาสตร์นั้นมีอยู่หลายวิธี วิธีการจำลอง โดยใช้เทคนิคมอนติคาร์โล เป็นวิธีหนึ่งที่นิยมนำมาใช้แก้ปัญหากันอย่างแพร่หลายในปัจจุบัน ซึ่งหลักการของการจำลองโดยใช้เทคนิคดังกล่าวจะใช้เลขสุ่ม (Random Numbers) มาช่วยในการหาคำตอบของปัญหาที่ต้องศึกษา

ขั้นตอนของวิธีการจำลองด้วยเทคนิคมอนติคาร์โลที่ใช้กันในปัจจุบัน แบ่งได้เป็น 3 ขั้นตอนดังนี้คือ

1. การสร้างตัวเลขสุ่ม การใช้ตัวเลขสุ่มเป็นสิ่งที่สำคัญมากในเทคนิคนี้ ทั้งนี้เพราะว่าหลักการจำลองด้วยเทคนิคมอนติคาร์โลนั้น จะใช้ตัวเลขสุ่มมาช่วยในการหาคำตอบของปัญหา โดยลักษณะของตัวเลขสุ่มที่นำมาใช้จะมีการแจกแจงแบบสม่ำเสมอในช่วง (0,1)

สำหรับวิธีการสร้างตัวเลขสุ่ม มีผู้เสนอไว้หลายวิธีแต่วิธีที่ดีนั้นลักษณะของเลขสุ่มที่ถูกสร้างขึ้น จะต้องมีการแจกแจงแบบสม่ำเสมอในช่วง (0,1) ตัวเลขสุ่มแต่ละตัวเป็นอิสระต่อกัน และมีช่วงยาวก่อนจะเกิดเลขสุ่มซ้ำ (มีวัฏจักรยาว)

2. การนำตัวเลขสุ่มมาประยุกต์ใช้กับปัญหาที่ต้องศึกษา ซึ่งขั้นตอนนี้ขึ้นอยู่กับลักษณะของปัญหา บางปัญหาอาจจะไม่ใช้ตัวเลขสุ่มโดยตรง แต่จะนำไปผลิตเลขสุ่มที่มีการแจกแจงแบบอื่นต่อไป

3. การทดลองกระทำ เมื่อนำตัวเลขสุ่มมาประยุกต์ให้เข้ากับปัญหาที่ต้องการศึกษาได้แล้วขั้นต่อไปคือ การทดลองโดยใช้กระบวนการของการสุ่ม (Random Process) มา กระทำในลักษณะซ้ำ ๆ กันหลาย ๆครั้ง เพื่อหาคำตอบที่ต้องการ

แผนการทดลอง

การวิจัยครั้งนี้ต้องการประมาณค่าพารามิเตอร์หรือสัมประสิทธิ์การถดถอยในสมการถดถอยเชิงเส้นพหุ เมื่อข้อมูลมีค่าผิดปกติ โดยประมาณค่าสัมประสิทธิ์การถดถอยจากวิธีการ 5 วิธี ภายใต้สถานการณ์ต่าง ๆ ที่กำหนดขนาดตัวอย่าง 3 ขนาด คือ 20, 30 และ 50 และลักษณะที่ข้อมูลเกิดค่าผิดปกติในกรณีต่าง ๆ ดังต่อไปนี้

1. กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรตาม (ความคลาดเคลื่อน) เมื่อกำหนดให้ความคลาดเคลื่อนมีการแจกแจงแบบปกติปลอมปน ที่สเกลแฟคเตอร์เท่ากับ 3 และ 10 และมีอัตราส่วนปลอมปนของความคลาดเคลื่อน เท่ากับ 0.01, 0.05 และ 0.10

2. กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระและตัวแปรตาม (ความคลาด-เคลื่อน) และข้อมูลในลักษณะนี้จะถูกแบ่งออกเป็น 3 กลุ่ม ตามชนิดของตัวแปรที่เกิดค่าผิดปกติ คือ

2.1 กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระ x_1 และตัวแปรตาม เมื่อกำหนดให้ตัวแปรอิสระ x_1 เกิดค่าผิดปกติโดยมีระดับค่าผิดปกติ 2 ระดับ คือ ระดับปาน-กลางและระดับรุนแรง ซึ่งแต่ละระดับมีอัตราส่วนค่าผิดปกติเท่ากับ 0.05 และ 0.10 และความคลาดเคลื่อนมีการแจกแจงแบบปกติปลอมปน ที่สเกลแฟคเตอร์เท่ากับ 3 และ 10 และมีอัตราส่วนปลอมปนของความคลาดเคลื่อนเท่ากับ 0.01, 0.05 และ 0.10

2.2 กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระ x_2 และตัวแปรตาม เมื่อกำหนดให้ตัวแปรอิสระ x_2 เกิดค่าผิดปกติโดยมีระดับค่าผิดปกติ 2 ระดับ คือ ระดับปาน-

กลางและระดับรุนแรง ซึ่งแต่ละระดับมีอัตราส่วนค่าผิดปกติเท่ากับ 0.05 และ 0.10 และความคลาดเคลื่อนมีการแจกแจงแบบปกติปลอมปน ที่สเกลแฟคเตอร์เท่ากับ 3 และ 10 และมีอัตราส่วนปลอมปนของความคลาดเคลื่อนเท่ากับ 0.01, 0.05 และ 0.10

2.3 กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระ x_1 ตัวแปรอิสระ x_2 และตัวแปรตาม เมื่อกำหนดให้ตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 เกิดค่าผิดปกติ โดยแต่ละตัวแปรอิสระมีระดับค่าผิดปกติ 2 ระดับ คือ ระดับปานกลางและระดับรุนแรง ซึ่งแต่ละระดับมีอัตราส่วนค่าผิดปกติเท่ากับ 0.05 และ 0.10 และความคลาดเคลื่อนมีการแจกแจงแบบปกติ-ปลอมปน ที่สเกลแฟคเตอร์เท่ากับ 3 และ 10 และมีอัตราส่วนปลอมปนของความคลาดเคลื่อนเท่ากับ 0.01, 0.05 และ 0.10

3. กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระและตัวแปรตาม (ความคลาด-เคลื่อน) แบบกำหนดให้เกิดค่าผิดปกติ ณ ตำแหน่งเดียวกัน และข้อมูลในลักษณะนี้จะถูกแบ่งออกเป็น 3 กลุ่มตามชนิดของตัวแปรที่เกิดค่าผิดปกติ คือ

3.1 กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระ x_1 และตัวแปรตาม ณ ตำแหน่งเดียวกัน เมื่อกำหนดให้ตัวแปรอิสระ x_1 เกิดค่าผิดปกติโดยมีระดับค่าผิดปกติ 2 ระดับ คือ ระดับปานกลางและระดับรุนแรง ซึ่งแต่ละระดับมีอัตราส่วนค่าผิดปกติเท่ากับ 0.05 และ 0.10 และกำหนดให้ความคลาดเคลื่อนที่เกิดขึ้น ณ ตำแหน่งที่มีค่าของตัวแปรอิสระ x_1 ผิดปกติ มีการแจกแจงแบบปกติ ที่มีความแปรปรวนเท่ากับ 9 และ 100 ส่วนความคลาดเคลื่อนที่เกิดขึ้น ณ หน่วยตัวอย่างที่มีค่าของตัวแปรอิสระ x_1 ปกติ มีการแจกแจงแบบปกติ ที่มีความแปรปรวนเท่ากับ 1

3.2 กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระ x_2 และตัวแปรตาม ณ ตำแหน่งเดียวกัน เมื่อกำหนดให้ตัวแปรอิสระ x_2 เกิดค่าผิดปกติโดยมีระดับค่าผิดปกติ 2 ระดับ คือ ระดับปานกลางและระดับรุนแรง ซึ่งแต่ละระดับมีอัตราส่วนค่าผิดปกติเท่ากับ 0.05 และ 0.10 และกำหนดให้ความคลาดเคลื่อนที่เกิดขึ้น ณ ตำแหน่งที่มีค่าของตัวแปร-อิสระ x_2 ผิดปกติ มีการแจกแจงแบบปกติ ที่มีความแปรปรวนเท่ากับ 9 และ 100 ส่วนความคลาดเคลื่อนที่เกิดขึ้น ณ หน่วยตัวอย่างที่มีค่าของตัวแปรอิสระ x_2 ปกติ มีการแจก-แจงแบบปกติ ที่มีความแปรปรวนเท่ากับ 1

3.3 กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระ x_1 หรือตัวแปรอิสระ x_2 และตัวแปรตาม ณ ตำแหน่งเดียวกัน เมื่อกำหนดให้ตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 เกิดค่าผิดปกติโดยที่แต่ละตัวแปรอิสระมีระดับค่าผิดปกติ 2 ระดับ คือ ระดับปานกลางและระดับรุนแรง ซึ่งแต่ละระดับมีอัตราส่วนค่าผิดปกติเท่ากับ 0.05 และ 0.10 และกำหนดให้

ความคลาดเคลื่อนที่เกิดขึ้น ณ ตำแหน่งที่มีค่าของตัวแปรอิสระ x_1 หรือตัวแปรอิสระ x_2 ผิดปกติ มีการแจกแจงแบบปกติ ที่มีความแปรปรวนเท่ากับ 9 และ 100 ส่วนความคลาดเคลื่อนที่เกิดขึ้น ณ ตำแหน่งที่มีค่าของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 ปกติ มีการแจกแจงแบบปกติ ที่มีความแปรปรวนเท่ากับ 1

จากนั้นทำการเปรียบเทียบค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของการประมาณค่าสัมประสิทธิ์การถดถอยของทุกวิธี เพื่อหาวิธีที่เหมาะสมสำหรับการประมาณค่าสัมประสิทธิ์การถดถอยในแต่ละสถานการณ์

ขั้นตอนในการวิจัย

สมการถดถอยที่ใช้ในการวิจัยครั้งนี้จะใช้สมการถดถอยเชิงเส้นพหุ (Multiple Linear Regression Equation) โดยมีรูปแบบสมการดังนี้

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \qquad , i = 1, 2, \ldots , n$$

เมื่อ y_i คือ ตัวแปรตาม

x_{1i}, x_{2i} คือ ตัวแปรอิสระตัวที่ 1 และตัวแปรอิสระตัวที่ 2

β_b คือ พารามิเตอร์หรือสัมประสิทธิ์การถดถอยตัวที่ b , b = 0, 1, 2

ε_i คือ ความคลาดเคลื่อน

n คือ จำนวนค่าสังเกตทั้งหมด

และผู้วิจัยจะการจำลองการทดลองตามสถานการณ์ต่าง ๆ โดยสร้างโปรแกรมคอมพิวเตอร์ด้วยภาษาฟอร์แทรน (FORTRAN 77) โดยใช้กับเครื่องคอมพิวเตอร์ AMDAHL 5850 เพื่อสร้างข้อมูลให้เป็นไปตามการวิจัย โดยขั้นตอนการวิจัยจะถูกแบ่งเป็น 4 ขั้นตอน ดังนี้

1. สร้างข้อมูลของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 ตามลักษณะของการเกิดค่าผิดปกติที่กำหนด

2. สร้างข้อมูลของความคลาดเคลื่อน (ε) ตามลักษณะของการเกิดค่าผิดปกติที่กำหนด

3. สร้างข้อมูลตัวแปรตาม (y)

4. ประมาณค่าพารามิเตอร์หรือสัมประสิทธิ์การถดถอยจากวิธีการ 5 วิธี

5. คำนวณค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของการประมาณค่าสัมประสิทธิ์การถดถอย

ซึ่งแต่ละขั้นตอนมีรายละเอียดดังต่อไปนี้

1. สร้างข้อมูลของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 ตามลักษณะการเกิดค่าผิดปกติที่กำหนด ดังนี้

1.1 กรณีที่ข้อมูลของตัวแปรอิสระมีค่าปกติ

ผู้วิจัยจะสร้างข้อมูลของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 จากการแจกแจงแบบปกติ คือ $x_{1i} \sim N(15,5)$ และ $x_{2i} \sim N(2,0.5)$ แล้วนำค่าของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 มาตรวจสอบค่าผิดปกติโดยใช้กราฟแบบ Box และ Whisker และถ้าพบว่าค่าของตัวแปรอิสระ x_1 ทุกค่าอยู่ในช่วง $(Q_{11} - 1.5\,(IQR_1)\,,\; Q_{31} + 1.5\,(IQR_1))$

เมื่อ Q_{11} คือ ค่าควอไทล์ที่ 1 ของตัวแปรอิสระ x_1

Q_{31} คือ ค่าควอไทล์ที่ 3 ของตัวแปรอิสระ x_1

และ IQR_1 เท่ากับ $Q_{31} - Q_{11}$

และค่าของตัวแปรอิสระ x_2 ทุกค่าอยู่ในช่วง $(Q_{12} - 1.5\,(IQR_2)\,,\; Q_{32} + 1.5\,(IQR_2))$

เมื่อ Q_{12} คือ ค่าควอไทล์ที่ 1 ของตัวแปรอิสระ x_2

Q_{32} คือ ค่าควอไทล์ที่ 3 ของตัวแปรอิสระ x_2

และ IQR_2 เท่ากับ $Q_{32} - Q_{12}$

จะถือว่าข้อมูลของตัวแปรอิสระมีค่าปกติ

1.2 กรณีที่ข้อมูลของตัวแปรอิสระมีค่าผิดปกติ

เมื่อกำหนดให้มีระดับค่าผิดปกติ 2 ระดับ คือ ระดับปานกลางและระดับรุนแรง ตามเงื่อนไขของการตรวจสอบค่าผิดปกติโดยใช้กราฟแบบ Box และ Whisker และในแต่ละระดับค่าผิดปกติจะกำหนดให้มีอัตราส่วนค่าผิดปกติเท่ากับ 0.05 และ 0.10 โดยหลักของการสร้างจะเริ่มจากการสร้างข้อมูลของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 ที่มีค่าปกติตามข้อ 1.1 จากนั้นนำข้อมูลของตัวแปรอิสระตัวที่ต้องการสร้างให้มีค่าผิดปกติมาเรียงลำดับค่าจากน้อยไปมาก แล้วทำการแทนที่ค่าของตัวแปรอิสระที่มากที่สุดด้วยค่าค่าหนึ่งที่ทำให้ค่าของตัวแปรอิสระที่มากที่สุดมีค่ามากขึ้นกว่าเดิม หรือแทนที่ค่าของตัวแปรอิสระที่น้อยที่สุดด้วยค่าค่าหนึ่งที่ทำให้ที่ค่าของตัวแปรอิสระที่น้อยที่สุดมีค่าน้อยลงกว่าเดิม เพื่อให้ค่าเหล่านี้กลายเป็นค่าผิดปกติในระดับที่ต้องการและมีจำนวนของค่าผิดปกติตามที่กำหนดโดย

จำนวนค่าผิดปกติของแต่ละขนาดตัวอย่างที่ศึกษามีดังนี้

ขนาดตัวอย่าง	อัตราส่วนการเกิดค่าผิดปกติ	
	0.05	0.10
20	1	2
30	2	3
50	3	5

และขั้นตอนของการสร้างข้อมูลที่มีค่าผิดปกติในตัวแปรอิสระ มีดังนี้

ขั้นที่ 1 สร้างข้อมูลของตัวแปรอิสระ x_1 และตัวแปรอิสระ x_2 ให้มีค่าปกติ ตามข้อ 1.1

ขั้นที่ 2 นำข้อมูลของตัวแปรอิสระตัวที่ต้องการสร้างให้มีค่าผิด-ปกติมาเรียงลำดับค่าจากน้อยไปมาก ได้แก่ D_1 , D_2 , D_3 , ... , D_{n-2} , D_{n-1} , D_n

เมื่อ D_1 คือ ค่าของตัวแปรอิสระที่มีค่าน้อยที่สุด

D_n คือ ค่าของตัวแปรอิสระที่มีค่ามากที่สุด

ขั้นที่ 3 แทนค่าของตัวแปรอิสระที่มากที่สุดด้วยค่าค่าหนึ่ง ซึ่งทำให้ มีค่ามากขึ้นกว่าเดิม หรือแทนค่าของตัวแปรอิสระที่น้อยที่สุดด้วยค่าค่าหนึ่ง ซึ่งทำให้มีค่าน้อยลงกว่าเดิม เพื่อให้ได้ค่าผิดปกติในระดับที่กำหนดและมีจำนวนตามที่ต้องการ

จำนวนค่าผิดปกติ	กำหนดค่าผิดปกติ
1	$D_1 = Q_1 - L\,(IQR)$
2	$D_1 = Q_1 - L\,(IQR)$ และ $D_n = Q_3 + L\,(IQR)$
3	$D_1 = Q_1 - L\,(IQR)$, $D_n = Q_3 + L\,(IQR)$ และ $D_2 = D_1(0.98)$
5	$D_1 = Q_1 - L\,(IQR)$, $D_n = Q_3 + L\,(IQR)$, $D_2 = D_1(0.98)$, $D_{n-1} = D_n(0.98)$ และ $D_{n-2} = D_n(0.995)$

เมื่อ D_1 คือ ค่าของตัวแปรอิสระที่มีค่าน้อยที่สุด

D_n คือ ค่าของตัวแปรอิสระที่มีค่ามากที่สุด

Q_1 คือ ค่าควอไทล์ที่ 1 ของตัวแปรอิสระ

Q_3 คือ ค่าควอไทล์ที่ 3 ของตัวแปรอิสระ

IQR เท่ากับ $Q_3 - Q_1$

และ L เท่ากับ 2 เมื่อต้องการค่าผิดปกติระดับปานกลาง (เนื่องจากจะได้ค่าที่อยู่กึ่งกลางของช่วง $(Q_1 - 3\,(IQR),\ Q_1 - 1.5\,(IQR))$ หรือ $(Q_3 + 1.5\,(IQR),\ Q_3 + 3\,(IQR))$ ซึ่งค่าที่อยู่ในช่วงเหล่านี้จะเป็นค่าผิดปกติระดับปานกลางตามเงื่อนไขการตรวจสอบค่าผิดปกติของวิธีกราฟแบบ Box และ Whisker)

L เท่ากับ 5 เมื่อต้องการค่าผิดปกติระดับรุนแรง (เนื่องจากจะได้ค่าที่น้อยกว่าค่าของ Q_1 -3 (IQR) หรือมากกว่าค่าของ $Q_3 + 3$ (IQR) ซึ่งจะเป็นค่าผิดปกติระดับรุนแรงตามเงื่อนไขการตรวจสอบค่าผิดปกติของวิธีกราฟแบบ Box และ Whisker)

2. สร้างข้อมูลของความคลาดเคลื่อน (ε) ตามลักษณะของการเกิดค่าผิดปกติที่กำหนด

2.1 กรณีที่ข้อมูลของตัวแปรตามมีค่าผิดปกติ

ผู้วิจัยจะทำการสร้างข้อมูลของ ε จากการแจกแจงแบบปกติปลอมปน คือ

$$\varepsilon_i \sim CN(PE, C_\varepsilon^2)$$

เมื่อ PE คือ อัตราส่วนปลอมปนของ ε ในที่นี้ PE = 0.01, 0.05 และ 0.10

และ C_ε คือ สเกลแฟคเตอร์ของ ε ในที่นี้ C_ε = 3 และ 10

2.2 กรณีที่ข้อมูลของตัวแปรตามและตัวแปรอิสระเกิดค่าผิดปกติ ณ ตำแหน่งเดียวกัน

ในการสร้างข้อมูลขั้นแรกจะสร้างข้อมูลของตัวแปรอิสระให้ค่าผิดปกติตามข้อ 1.2 และกำหนดให้ความคลาดเคลื่อน (ε) ที่เกิดขึ้น ณ ตำแหน่งที่มีค่าของตัวแปร-อิสระ ผิดปกติ มีการแจกแจงแบบปกติ คือ

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$

เมื่อ σ_ε^2 คือ ความแปรปรวนของ ε ในที่นี้ σ_ε^2 = 9 และ 100 ส่วนความคลาดเคลื่อน (ε) ที่เกิดขึ้น ณ ตำแหน่งที่มีค่าของตัวแปรอิสระปกติ มีการแจกแจงแบบปกติ คือ

$$\varepsilon_i \sim N(0,1)$$

และการสร้างข้อมูลของตัวแปรอิสระและ ε ตามข้อ 1. และข้อ 2. ผู้วิจัยจะแสดงรายละเอียดของการผลิตเลขสุ่มและโปรแกรมที่ใช้สร้างไว้ในภาคผนวก ก. และ ข.

3. สร้างข้อมูลตัวแปรตาม (y)

สร้างค่าตัวแปรตาม y_i ตามรูปแบบความสัมพันธ์ดังนี้

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \quad , i = 1, 2, \ldots , n$$

เมื่อ β_0 , β_1 และ β_2 คือ พารามิเตอร์หรือสัมประสิทธิ์การถดถอย ซึ่งในการวิจัยครั้งนี้ กำหนด $\beta_0 = 1$, $\beta_1 = 2$ และ $\beta_2 = 5$

4. ประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีการ 5 วิธีคือ

4.1 วิธีกำลังสองน้อยที่สุด

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

4.2 วิธีตัวประมาณ M เมื่อใช้เกณฑ์ความแกร่งของ Huber ซึ่งจะใช้การคำนวณตามวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักซ้ำหลายรอบ (The Iteratively Reweighted Least Squares Procedure : IRLS Procedure) โดยมีขั้นตอนดังนี้

ขั้นที่ 1 ทำการประมาณค่าสัมประสิทธิ์การถดถอยเริ่มต้น โดยวิธี OLS เมื่อกำหนด $k = 0$ ซึ่ง k คือ จำนวนรอบของการทำซ้ำ

$$\hat{\beta}^{(0)} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

ขั้นที่ 2 เข้าสู่การทำซ้ำรอบที่ k+1 เพื่อหาค่า $e_i^{(k+1)}$ และ $\hat{\sigma}^{(k+1)}$ เมื่อ $k = 0, 1, 2, \ldots$ และ $i = 1, 2, \ldots , n$

$$e_i^{(k+1)} = y_i - \mathbf{x}_i^T \hat{\beta}^{(k)}$$

และ

$$\hat{\sigma}^{(k+1)} = 1.4826 \ \underset{i}{\mathrm{med}} \left(\left| e_i^{(k+1)} - \underset{j}{\mathrm{med}} \left(e_j^{(k+1)}\right) \right| \right)$$

ขั้นที่ 3 การทำซ้ำรอบที่ k+1 เพื่อหา $w_i^{(k+1)}$ เมื่อใช้เกณฑ์ความแกร่งของ Huber และ $i = 1, 2, \ldots, n$

$$w_i^{(k+1)} = \begin{cases} \dfrac{-1.345}{\left(e_i^{(k+1)}/\hat{\sigma}^{(k+1)}\right)} & , \dfrac{e_i^{(k+1)}}{\hat{\sigma}^{(k+1)}} < -1.345 \\ 1 & , -1.345 \le \dfrac{e_i^{(k+1)}}{\hat{\sigma}^{(k+1)}} \le 1.345 \\ \dfrac{1.345}{\left(e_i^{(k+1)}/\hat{\sigma}^{(k+1)}\right)} & , \dfrac{e_i^{(k+1)}}{\hat{\sigma}^{(k+1)}} > 1.345 \end{cases}$$

ขั้นที่ 4 การทำซ้ำรอบที่ k+1 เพื่อหา $\hat{\beta}^{(k+1)}$

$$\hat{\beta}^{(k+1)} = (\mathbf{X}^T W^{(k+1)} \mathbf{X})^{-1} \mathbf{X}^T W^{(k+1)} \mathbf{y}$$

เมื่อ

$$W^{(k+1)} = \begin{bmatrix} w_1^{(k+1)} & 0 & \cdots & 0 \\ 0 & w_2^{(k+1)} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & w_n^{(k+1)} \end{bmatrix}_{n \times n}$$

ขั้นที่ 5 คำนวณค่าสัมบูรณ์ของผลต่างระหว่างค่าประมาณสัมประสิทธิ์การถดถอยรอบที่ k กับค่าประมาณสัมประสิทธิ์การถดถอยรอบที่ k+1 ของสัมประสิทธิ์การถดถอยทุกค่า

$$P_b = \left| \hat{\beta}_b^{(k+1)} - \hat{\beta}_b^{(k)} \right| \quad , b = 0, 1, 2$$

ขั้นที่ 6 ถ้า P_b อย่างน้อย 1 ค่า มากกว่า 0.001 แล้วจะไปทำขั้นที่ 2 ถึง ขั้นที่ 6 ซ้ำ ๆ จนกระทั่ง P_b ทุกค่า น้อยกว่าหรือเท่ากับ 0.001 แล้วจะได้ค่าประมาณสัมประสิทธิ์การถดถอยรอบที่ k+1

4.3 วิธีตัวประมาณ M เมื่อใช้เกณฑ์ความแกร่งของ Tukey ซึ่งจะใช้การคำนวณตามวิธี IRLS โดยมีขั้นตอนดังนี้

ขั้นที่ 1 การประมาณค่าสัมประสิทธิ์การถดถอยเริ่มต้น โดยวิธี OLS เมื่อกำหนด $k = 0$ ซึ่ง k คือ จำนวนรอบของการทำซ้ำ

$$\hat{\beta}^{(0)} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

ขั้นที่ 2 เข้าสู่การทำซ้ำรอบที่ $k+1$ เพื่อหาค่า $e_i^{(k+1)}$ และ $\hat{\sigma}^{(k+1)}$ เมื่อ $k = 0, 1, 2, \ldots$ และ $i = 1, 2, \ldots, n$

$$e_i^{(k+1)} = y_i - \mathbf{x}_i^T\hat{\beta}^{(k)}$$

และ

$$\hat{\sigma}^{(k+1)} = 1.4826 \underset{i}{\text{med}} \left(\left| e_i^{(k+1)} - \underset{j}{\text{med}} (e_j^{(k+1)}) \right| \right)$$

ขั้นที่ 3 การทำซ้ำรอบที่ $k+1$ เพื่อหา $w_i^{(k+1)}$ เมื่อใช้เกณฑ์ความแกร่งของ Tukey และ $i = 1, 2, \ldots, n$

$$w_i^{(k+1)} = \begin{cases} \left[1 - \left(\dfrac{e_i^{(k+1)}}{\hat{\sigma}^{(k+1)}} \dfrac{1}{4.685}\right)^2\right]^2 & , \left|\dfrac{e_i^{(k+1)}}{\hat{\sigma}^{(k+1)}}\right| \leq 4.685 \\ 0 & , \left|\dfrac{e_i^{(k+1)}}{\hat{\sigma}^{(k+1)}}\right| > 4.685 \end{cases}$$

ขั้นที่ 4 การทำซ้ำรอบที่ $k+1$ เพื่อหา $\hat{\beta}^{(k+1)}$

$$\hat{\beta}^{(k+1)} = (\mathbf{X}^T\mathbf{W}^{(k+1)}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{(k+1)}\mathbf{y}$$

เมื่อ

$$\mathbf{W}^{(k+1)} = \begin{bmatrix} w_1^{(k+1)} & 0 & \cdots & 0 \\ 0 & w_2^{(k+1)} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & w_n^{(k+1)} \end{bmatrix}_{n \times n}$$

ขั้นที่ 5 คำนวณค่าสัมบูรณ์ของผลต่างระหว่างค่าประมาณสัมประสิทธิ์การถดถอยรอบที่ k กับค่าประมาณสัมประสิทธิ์การถดถอยรอบที่ k+1 ของสัมประสิทธิ์การถดถอยทุกค่า

$$P_b = \left|\hat{\beta}_b^{(k+1)} - \hat{\beta}_b^{(k)}\right| \quad , b = 0, 1, 2$$

ขั้นที่ 6 ถ้า P_b อย่างน้อย 1 ค่า มากกว่า 0.001 แล้วจะไปทำ ขั้นที่ 2 ถึง ขั้นที่ 6 ซ้ำ ๆ จนกระทั่ง P_b ทุกค่า น้อยกว่าหรือเท่ากับ 0.001 แล้วจะได้ค่าประมาณสัมประสิทธิ์การถดถอยรอบที่ k+1

4.4 วิธีตัวประมาณ BI เมื่อใช้เกณฑ์ความแกร่งของ Huber ซึ่งจะใช้การคำนวณตามวิธี IRLS โดยมีขั้นตอนดังนี้

ขั้นที่ 1 การประมาณค่าสัมประสิทธิ์การถดถอยเริ่มต้น โดยวิธี OLS เมื่อกำหนด $k = 0$ ซึ่ง k คือ จำนวนรอบของการทำซ้ำ

$$\hat{\beta}^{(0)} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

และ

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$$

$$= \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n2} & \cdots & h_{nn} \end{bmatrix}_{n\times n}$$

ขั้นที่ 2 เข้าสู่การทำซ้ำรอบที่ k+1 , $k = 0, 1, 2, \ldots$ เพื่อหาค่า $e_i^{(k+1)}$, $\hat{\sigma}^{(k+1)}$ และ $DFFITS_i^{(k+1)}$ เมื่อ $i = 1, 2, \ldots, n$

$$e_i^{(k+1)} = y_i - \mathbf{x}_i^T\hat{\beta}^{(k)}$$

$$\hat{\sigma}^{(k+1)} = 1.4826 \ \underset{i}{\text{med}} \left(\left| e_i^{(k+1)} - \underset{j}{\text{med}} \left(e_j^{(k+1)}\right) \right| \right)$$

และ

$$\mathrm{DFFITS}_i^{(k+1)} = \frac{e_i^{(k+1)}}{\hat{\sigma}^{(k+1)}} \cdot \frac{\sqrt{h_{ii}}}{(1-h_{ii})}$$

ขั้นที่ 3 การทำซ้ำรอบที่ k+1 เพื่อหา $w_i^{(k+1)}$ เมื่อใช้เกณฑ์ความแกร่งของ Huber และ i = 1, 2, ... , n

$$w_i^{(k+1)} = \begin{cases} \dfrac{-1.345}{(\mathrm{DFFITS}_i^{(k+1)})} &, \mathrm{DFFITS}_i^{(k+1)} < -1.345 \\ 1 &, -1.345 \le \mathrm{DFFITS}_i^{(k+1)} \le 1.345 \\ \dfrac{1.345}{(\mathrm{DFFITS}_i^{(k+1)})} &, \mathrm{DFFITS}_i^{(k+1)} > 1.345 \end{cases}$$

ขั้นที่ 4 การทำซ้ำรอบที่ k+1 เพื่อหา $\hat{\beta}^{(k+1)}$

$$\hat{\beta}^{(k+1)} = (\mathbf{x}^T W^{(k+1)} \mathbf{x})^{-1} \mathbf{x}^T W^{(k+1)} \mathbf{y}$$

เมื่อ

$$W^{(k+1)} = \begin{bmatrix} w_1^{(k+1)} & 0 & \dots & 0 \\ 0 & w_2^{(k+1)} & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & w_n^{(k+1)} \end{bmatrix}_{n \times n}$$

ขั้นที่ 5 คำนวณค่าสัมบูรณ์ของผลต่างระหว่างค่าประมาณสัมประสิทธิ์การถดถอยรอบที่ k กับค่าประมาณสัมประสิทธิ์การถดถอยรอบที่ k+1 ของสัมประสิทธิ์การถดถอยทุกค่า

$$P_b = \left| \hat{\beta}_b^{(k+1)} - \hat{\beta}_b^{(k)} \right| \quad , b = 0, 1, 2$$

ขั้นที่ 6 ถ้า P_b อย่างน้อย 1 ค่า มากกว่า 0.001 แล้วจะไปทำขั้นที่ 2 ถึง ขั้นที่ 6 ซ้ำ ๆ จนกระทั่ง P_b ทุกค่า น้อยกว่าหรือเท่ากับ 0.001 แล้วจะได้ค่าประมาณสัมประสิทธิ์การถดถอยรอบที่ k+1

4.5 วิธีตัวประมาณ BI เมื่อใช้เกณฑ์ความแกร่งของ Tukey ซึ่งจะใช้การคำนวณตามวิธี IRLS โดยมีขั้นตอนดังนี้

ขั้นที่ 1 การประมาณค่าสัมประสิทธิ์การถดถอยเริ่มต้น โดยวิธี OLS เมื่อกำหนด $k = 0$ ซึ่ง k คือ จำนวนรอบของการทำซ้ำ

$$\hat{\beta}^{(0)} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

และ

$$\mathrm{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n2} & \cdots & h_{nn} \end{bmatrix}_{n\times n}$$

ขั้นที่ 2 เข้าสู่การทำซ้ำรอบที่ k+1 , $k = 0, 1, 2, \ldots$ เพื่อหาค่า $e_i^{(k+1)}$, $\hat{\sigma}^{(k+1)}$ และ $\mathrm{DFFITS}_i^{(k+1)}$ เมื่อ $i = 1, 2, \ldots, n$

$$e_i^{(k+1)} = y_i - \mathbf{x}_i^T \hat{\beta}^{(k)}$$

$$\hat{\sigma}^{(k+1)} = 1.4826 \ \underset{i}{\mathrm{med}} \left(\left| e_i^{(k+1)} - \underset{j}{\mathrm{med}} \left(e_j^{(k+1)}\right) \right| \right)$$

และ

$$\mathrm{DFFITS}_i^{(k+1)} = \frac{e_i^{(k+1)}}{\hat{\sigma}^{(k+1)}} \cdot \frac{\sqrt{h_{ii}}}{(1-h_{ii})}$$

ขั้นที่ 3 การทำซ้ำรอบที่ k+1 เพื่อหา $w_i^{(k+1)}$ เมื่อใช้เกณฑ์ความแกร่งของ Tukey และ $i = 1, 2, \ldots, n$

$$w_i^{(k+1)} = \begin{cases} \left[1 - \left(\dfrac{\mathrm{DFFITS}_i^{(k+1)}}{4.685}\right)^2\right]^2 & , \left|\mathrm{DFFITS}_i^{(k+1)}\right| \le 4.685 \\ 0 & , \left|\mathrm{DFFITS}_i^{(k+1)}\right| > 4.685 \end{cases}$$

ขั้นที่ 4 การทำซ้ำรอบที่ k+1 เพื่อหา $\hat{\beta}^{(k+1)}$

$$\hat{\beta}^{(k+1)} = (\mathbf{x}^T W^{(k+1)} \mathbf{x})^{-1} \mathbf{x}^T W^{(k+1)} \mathbf{y}$$

เมื่อ

$$W^{(k+1)} = \begin{bmatrix} w_1^{(k+1)} & 0 & \cdots & 0 \\ 0 & w_2^{(k+1)} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & w_n^{(k+1)} \end{bmatrix}_{n \times n}$$

ขั้นที่ 5 คำนวณค่าสัมบูรณ์ของผลต่างระหว่างค่าประมาณสัมประสิทธิ์การถดถอยรอบที่ k กับค่าประมาณสัมประสิทธิ์การถดถอยรอบที่ k+1 ของสัมประสิทธิ์การถดถอยทุกค่า

$$P_b = \left| \hat{\beta}_b^{(k+1)} - \hat{\beta}_b^{(k)} \right| \quad , b = 0, 1, 2$$

ขั้นที่ 6 ถ้า P_b อย่างน้อย 1 ค่า มากกว่า 0.001 แล้วจะไปทำขั้นที่ 2 ถึง ขั้นที่ 6 ซ้ำ ๆ จนกระทั่ง P_b ทุกค่า น้อยกว่าหรือเท่ากับ 0.001 แล้วจะได้ค่าประมาณสัมประสิทธิ์การถดถอยรอบที่ k+1

5. คำนวณค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของการประมาณค่าสัมประสิทธิ์การถดถอยทั้ง 5 วิธี ภายใต้สถานการณ์ต่าง ๆ โดยมีการทำซ้ำ 500 ครั้ง ดังนี้

$$\text{MSE} = \frac{\sum_{b=0}^{p} \left[\sum_{t=1}^{500} \frac{(\beta_{bt} - \hat{\beta}_{bt})^2}{500} \right]}{p+1}$$

เมื่อ β_{bt} คือ ค่าสัมประสิทธิ์การถดถอยที่กำหนด ตัวที่ b ของการทำซ้ำรอบที่ t

$\hat{\beta}_{bt}$ คือ ค่าประมาณสัมประสิทธิ์การถดถอย ตัวที่ b ของการทำซ้ำรอบที่ t

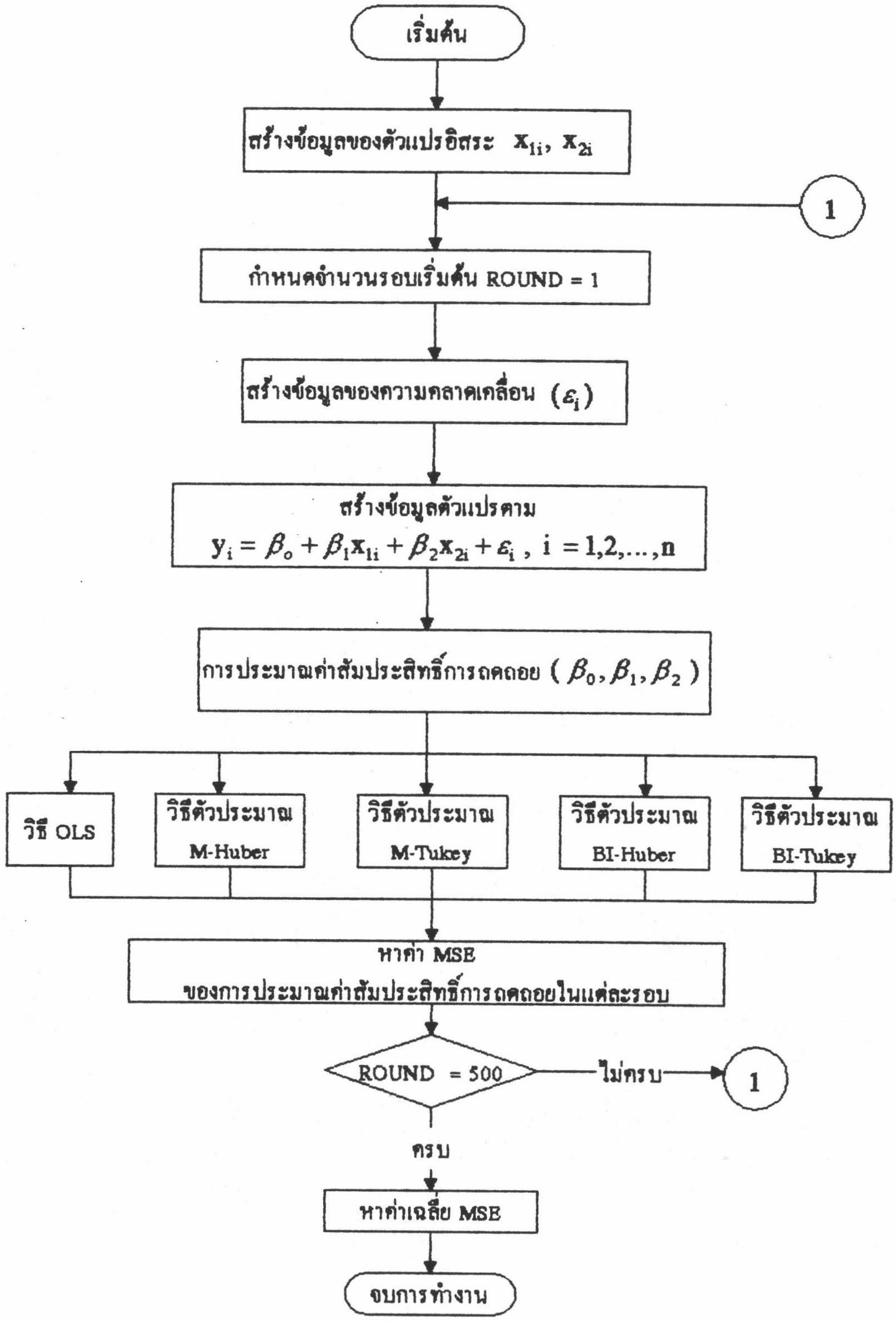
p คือ จำนวนตัวแปรอิสระ

t คือ จำนวนรอบในการทำการทดลองซ้ำ

ในการคำนวณค่า MSE ของการประมาณค่าสัมประสิทธิ์การถดถอยทั้ง 5 วิธี ภายใต้สถานการณ์ที่สร้างให้เกิดข้อมูลเกิดค่าผิดปกติแบบต่าง ๆ ตามที่กำหนด โดยใน

ทุกสถานการณ์ ผู้วิจัยจะทำการทดลองทำซ้ำ 500 ครั้งจนครบทุกสถานการณ์ ซึ่งขั้นตอนของการทดลองถูกสรุปเป็นผังงานได้ดังรูปที่ 3.1

รูปที่ ธ.1 แสดงผังงานสำหรับหาค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของการประมาณค่าสัมประสิทธิ์การถดถอยของทั้ง 5 วิธี



โปรแกรมที่ใช้ในการวิจัย

โปรแกรมที่ใช้ในการวิจัยทั้งหมด เขียนด้วยภาษาฟอร์แทรน (FORTRAN) โดยใช้กับเครื่อง AMDAHL 5860 ซึ่งในแต่ละสถานการณ์ของการทดลอง ลักษณะของการทำงานของโปรแกรมจะเหมือนกัน สำหรับรายละเอียดของโปรแกรมจะแสดงไว้ในภาคผนวก ข ซึ่งเป็นโปรแกรมการทำงานของแต่ละวิธี