$\varepsilon \sim N(0, \sigma^2)$

หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยในครั้งนี้ต้องการศึกษาเปรียบเทียบประสิทธิภาพของตัวประมาณ 4 วิธีคือ วิธีกำลังสองต่ำสุด วิธีของบราวน์และมูด้ วิธีของเซ็นและทิลล์ และวิธีของซีเวอร์ โดยจะศึกษาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณในแต่ละวิธี แล้วทำการเปรียบเทียบในลักษณะประสิทธิภาพสัมพัทธ์ของตัวประมาณในแต่ละวิธีกับตัวประมาณกำลังสองต่ำสุด พร้อมทั้งศึกษาเปรียบเทียบสถิติทดสอบในการทดสอบพารามิเตอร์ β_1 ของวิธี 4 วิธีดังกล่าวด้วย และศึกษาเปรียบเทียบสถิติทดสอบในการทดสอบพารามิเตอร์ β_0 , β_1 ของอีก 3 วิธีคือ วิธีกำลังสองต่ำสุด วิธีของบราวน์และมูด้ และวิธีของแลนคาสเตอร์และเควด โดยจะศึกษาอำนาจของการทดสอบ และความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของสถิติทดสอบทุกวิธีดังกล่าว เมื่อความคลาดเคลื่อน (ε) มีการแจกแจงแบบยูนิฟอร์ม แบบโลจิสติค แบบดับเบิ้ลเอ็กซ์โปเนนเชียล และแบบปกติปลอมปน ทั้งนี้เนื่องจากการแจกแจงแบบต่าง ๆ ดังกล่าวเป็นการแจกแจงที่มีลักษณะการกระจายไปทางหางมาก หรือมีหางยาวกว่าปกติ ซึ่งเป็นลักษณะที่สนใจศึกษา ยกเว้นการแจกแจงแบบยูนิฟอร์ม ที่ศึกษาเนื่องจากต้องการทราบผลสรุปกรณีเมื่อ ε มีการแจกแจงไม่เป็นทั้งแบบปกติ และลักษณะการกระจายไปทางหางมาก หรือหางยาวกว่าปกติ ส่วนรูปแบบของการแจกแจงแบบปกติปลอมปน จะทำการศึกษาเมื่อมีเปอร์เซ็นต์การปลอมปนเป็น 1%, 5%, 10 %, และ 25% สำหรับสเกลแฟคเตอร์ 2 ระดับคือ 3 และ 10 ซึ่งถ้าสเกลแฟคเตอร์นี้มีค่าสูงจะทำให้เกิดค่าผิดปกติได้ และจากการศึกษาที่ผ่านมาสเกลแฟคเตอร์ที่มีค่าน้อยกว่า 3 จะมีโอกาสของการเกิดค่าผิดปกติน้อย ส่วนสเกลแฟคเตอร์ที่มีค่ามากกว่า 10 จะมีโอกาสของการเกิดค่าผิดปกติมาก จึงทำการศึกษาเพียง 2 ระดับดังกล่าว ขนาดตัวอย่างที่ใช้ในการศึกษาครั้งนี้มี 4 ขนาด คือ 10, 15, 20 และ 50 และเนื่องจากช่วงห่างระหว่างค่าของตัวแปรอิสระ (x) มีอิทธิพลต่อวิธีการต่าง ๆ ดังกล่าว จึงศึกษาทั้งกรณีที่ช่วงห่างระหว่าง x มีค่าเท่ากันและไม่เท่ากัน ทั้งนี้เทคนิคที่ใช้ในการหาค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจของการทดสอบ คือ วิธีมอนติคาร์โล ซึ่งเป็นเทคนิคอย่างหนึ่งที่ใช้แก้ปัญหาในการคำนวณทางคณิตศาสตร์ โดยเฉพาะกรณีที่ไม่สามารถหาคำตอบโดยวิธีทางทฤษฎีได้

เนื่องจากวิธีมอนติคาร์โลเป็นเทคนิคที่ใช้ในการวิจัยครั้งนี้ ดังนั้นในตอนแรกของบทจะกล่าวถึงวิธีมอนติคาร์โลก่อน แล้วจึงแสดงรายละเอียดของแผนการทดลอง ขั้นตอนการวิจัย และโปรแกรมที่ใช้ในการวิจัยตามลำดับ ซึ่งรายละเอียดต่าง ๆ เป็นดังนี้

3.1 วิธีมอนติคาร์โล (Monte Carlo Method)

เทคนิคที่ใช้สำหรับแก้ปัญหาในการคำนวณทางคณิตศาสตร์นั้นมีอยู่หลายวิธี วิธีมอนติคาร์โลเป็นวิธีหนึ่งที่จะใช้แก้ปัญหาได้ และเป็นวิธีที่นิยมใช้กันอย่างแพร่หลายในปัจจุบัน แฮมเมอร์สเลย์และแฮนสคอมบ์ (Hammersley and Handscomb 1964 : 2) กล่าวว่า วิธีมอนติคาร์โลเป็นสาขาหนึ่งของคณิตศาสตร์เชิงทดลอง ซึ่งหลักการของวิธีมอนติคาร์โลนั้น จะใช้ตัวเลขสุ่ม (Random Number) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา

ในการวิจัยครั้งนี้จะใช้เทคนิคมอนติคาร์โลดังกล่าว ในการสร้างข้อมูลที่มีลักษณะการแจกแจงตามที่ต้องการ ซึ่งขั้นตอนของวิธีมอนติคาร์โลที่ใช้กันอยู่ในปัจจุบัน แบ่งได้เป็น 3 ขั้นตอนดังนี้

3.1.1 การสร้างตัวเลขสุ่ม การใช้ตัวเลขสุ่มเป็นสิ่งที่สำคัญมากในวิธีมอนติคาร์โล ทั้งนี้เพราะว่า หลักการของวิธีมอนติคาร์โลนั้นจะใช้ตัวเลขสุ่มมาช่วยในการหาคำตอบของปัญหา ลักษณะของตัวเลขสุ่มจะมีการแจกแจงแบบยูนิฟอร์มในช่วง (0,1) สำหรับวิธีการสร้างตัวเลขสุ่มมีผู้เสนอไว้หลายวิธี แต่วิธีที่ดีนั้นลักษณะของตัวเลขสุ่มที่เกิดขึ้นจะต้องมีการแจกแจงแบบยูนิฟอร์มในช่วง (0,1) และเป็นอิสระกัน

3.1.2 การประยุกต์ปัญหาที่ต้องการศึกษามาใช้กับตัวเลขสุ่ม ซึ่งขั้นตอนนี้ขึ้นอยู่กับลักษณะของปัญหาที่ต้องการศึกษา บางปัญหาอาจจะไม่ใช้ตัวเลขสุ่มโดยตรง แต่อาจจะมีขั้นตอนอื่นอีกหลาย ๆ ขั้นตอน ซึ่งขั้นตอนเหล่านี้มีบางขั้นตอนที่ต้องใช้ตัวเลขสุ่ม

3.1.3 การทดลองกระทำ เมื่อประยุกต์ปัญหาให้ใช้กับตัวเลขสุ่มได้แล้ว ขั้นตอนต่อไปก็คือ การทดลองโดยใช้กระบวนการของการสุ่ม (Random Process) มากระทำในลักษณะที่ซ้ำ ๆ กัน เพื่อหาคำตอบของปัญหาที่ต้องการศึกษา

3.2 แผนการทดลอง

ในการวิจัยครั้งนี้ กำหนดสถานการณ์ต่าง ๆ สำหรับศึกษาเปรียบเทียบประสิทธิภาพของตัวประมาณ อำนาจของการทดสอบและความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของวิธีต่าง ๆ ดังกล่าว โดยสุ่มตัวอย่างจากประชากรที่มีการแจกแจงแบบเดียวกัน ซึ่งการแจกแจงที่สนใจศึกษาคือ การแจกแจงแบบยูนิฟอร์ม แบบโลจิสติค แบบดับเบิ้ลเอ็กซ์โปเนนเชียล และแบบปกติปลอมปน

สำหรับการแจกแจงแบบปกติปลอมปน กำหนดเปอร์เซ็นต์การปลอมปนเป็น 1%, 5%, 10% และ 25% สำหรับสเกลแฟคเตอร์ 2 ระดับคือ 3 และ 10 ซึ่งแสดงรายละเอียดดังตาราง 3.1

ตารางที่ 3.1 แสดงค่าสเกลแฟคเตอร์และเปอร์เซ็นต์การปลอมปนทั้งหมดที่ใช้ในการวิจัยครั้งนี้

(3,1)	(3,5)	(3,10)	(3,25)
(10,1)	(10,5)	(10,10)	(10,25)

สำหรับค่าสเกลแฟคเตอร์ที่น้อยกว่า 3 นั้น จะไม่ถือว่ามีอิทธิพลต่อการทำให้เกิดการแจกแจงที่มีลักษณะกระจายไปทางหางมาก หรือทำให้เกิดค่าผิดปกติแต่อย่างใด จึงไม่ได้สนใจศึกษาไว้ในการวิจัยครั้งนี้

3.3 ขั้นตอนในการวิจัย

ขั้นตอนในการวิจัย แบ่งออกเป็น 5 ขั้นตอน คือ

1. การสร้างโปรแกรมย่อย สำหรับสร้างการแจกแจงของความคลาดเคลื่อน (ε) ตามที่กำหนด
2. การสร้างข้อมูล (x , y) ที่มีความสัมพันธ์กันเชิงเส้นตรง
3. การหาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ
4. การหาค่าอำนาจของการทดสอบและความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 สำหรับการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ β_1

5. การหาค่าอำนาจของการทดสอบและความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 สำหรับการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ β_0 , β_1

ซึ่งรายละเอียดสำหรับแต่ละขั้นตอนเป็นดังนี้

3.1.1 <u>การสร้างโปรแกรมย่อยสำหรับสร้างการแจกแจงของความคลาดเคลื่อน (ε) ตามที่กำหนด</u>

การสร้างลักษณะการแจกแจงของความคลาดเคลื่อนทุกรูปแบบ ตามที่กำหนดไว้ในแผนการทดลองนั้น ใช้โปรแกรมภาษาฟอร์แทรนโฟ (Fortran IV) โดยใช้กับเครื่อง IBM 370/3031 ซึ่งการสร้างลักษณะการแจกแจงแบบต่าง ๆ นั้น จะต้องใช้ตัวเลขสุ่ม ซึ่งมีการแจกแจงแบบยูนิฟอร์มในช่วง (0,1) เป็นพื้นฐานในการสร้าง สำหรับโปรแกรมที่ใช้สร้างตัวเลขสุ่มในการวิจัยครั้งนี้ ใช้วิธีที่ไวท์และชมิดท์ (White and Schmidt 1975 : 421) เสนอไว้ ซึ่งรายละเอียดแสดงไว้ในภาคผนวก ก ส่วนรายละเอียดในการสร้างการแจกแจงแบบต่าง ๆ เป็นดังนี้

3.3.1.1 <u>การแจกแจงแบบยูนิฟอร์ม</u>

โปรแกรมย่อยที่ใช้ในการสร้างการแจกแจงแบบยูนิฟอร์ม ใช้วิธี Inverse Transformation ซึ่งรายละเอียดแสดงไว้ในภาคผนวก ก การใช้โปรแกรมย่อยนี้ใช้คำสั่ง CALL UNIFRM (A,B,X) โดยค่า A และ B เป็นค่าพารามิเตอร์ที่กำหนดขึ้นเพื่อให้ได้ค่าเฉลี่ย และความแปรปรวนตามที่กำหนด ค่า A และ B นี้จะถูกส่งมาจากโปรแกรมหลัก ส่วนผลลัพธ์ที่ได้คือ X ซึ่งเป็นตัวแปรสุ่มที่มีการแจกแจงแบบยูนิฟอร์มในช่วง(A , B) ที่มีค่าเฉลี่ยเป็น $\frac{(A+B)}{2}$ และความแปรปรวนเป็น $\frac{(B-A)^2}{2}$

3.3.1.2 <u>การแจกแจงแบบโลจิสติค</u>

โปรแกรมย่อยที่ใช้ในการสร้างการแจกแจงแบบโลจิสติค ใช้วิธี Inverse Transformation ซึ่งรายละเอียดแสดงไว้ในภาคผนวก ก การใช้โปรแกรมย่อยนี้ใช้คำสั่ง CALL LOGIST (ALPHA, BETA, X) โดยค่า ALPHA, BETA เป็นค่าพารามิเตอร์ที่กำหนดขึ้นเพื่อให้ได้ค่าเฉลี่ย และความแปรปรวนตามกำหนด ค่า ALPHA และ BETA นี้ จะถูกส่งมาจากโปรแกรมหลัก ส่วนผลลัพธ์ที่ได้คือ X ซึ่งเป็นตัวแปรสุ่มที่มีการแจกแจงแบบโลจิสติค ที่มีค่าเฉลี่ยเป็น ALPHA และความแปรปรวน $\frac{1}{3}\pi^2$ (BETA)^2

3.3.1.3 การแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียล

โปรแกรมย่อยที่ใช้ในการสร้างการแจกแจงแบบดับเบิ้ลเอ็กซ์-โปเนนเชียล ใช้วิธี Inverse Transformation ซึ่งรายละเอียดแสดงไว้ในภาคผนวก ก การใช้โปรแกรมย่อยนี้ใช้คำสั่ง CALL DOUBLE (ALPHA, BETA, X) โดยค่า ALPHA, BETA เป็นค่าพารามิเตอร์ที่กำหนดขึ้นเพื่อให้ได้ค่าเฉลี่ย และความแปรปรวนตามที่กำหนด ค่า ALPHA และ BETA นี้ จะถูกส่งมาจากโปรแกรมหลัก ส่วนผลสัพธ์ที่ได้คือ X ซึ่งเป็นตัวแปรสุ่มที่มีการแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียล ที่มีค่าเฉลี่ยเป็น ALPHA และความแปรปรวนเป็น $2(BETA)^2$

3.3.1.4 การแจกแจงแบบปกติปลอมปน

โปรแกรมย่อยที่ใช้ในการสร้างการแจกแจงแบบปกติปลอม-ปนนั้น ใช้วิธีแปลงข้อมูลจากการแจกแจงแบบปกติ โดยพิจารณาการแจกแจงแบบปกติปลอมปนของตัวแปร X ซึ่งมาจาก $F = (1-p)\ N(\mu, \sigma^2) + pN(\mu, c^2\sigma^2)$ เมื่อ p เป็นเปอร์เซ็นต์การปลอมปน c เป็นสเกลแฟคเตอร์ที่จะทำให้เกิดค่าผิดปกติ ($c > 0$) หมายความว่าตัวแปร X จะมาจากการแจกแจงแบบ $N(\mu, \sigma^2)$ ด้วยความน่าจะเป็น $(1-p)$ และมาจากการแจกแจง $N(\mu, c^2\sigma^2)$ ด้วยความน่าจะเป็น p สำหรับรายละเอียดแสดงไว้ในภาคผนวก ก การใช้โปรแกรมย่อยนี้ใช้คำสั่ง CALL SCNRML (C,P,SMEAN, SIGMA, X) ค่า c,p เป็นค่ากำหนดสเกลแฟคเตอร์ และเปอร์เซ็นต์การปลอมปน ส่วน SMEAN และ $(SIGMA)^2$ เป็นค่ากำหนดค่าเฉลี่ย และความแปรปรวน ซึ่งค่า c,p, SMEAN, SIGMA นี้ จะถูกส่งมาจากโปรแกรมหลัก ส่วนผลสัพธ์ที่ได้คือ X ซึ่งเป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติปลอมปน ที่มีค่าเฉลี่ยเป็น SMEAN และความแปรปรวนเป็น $(SIGMA)^2$

สำหรับโปรแกรมย่อยที่ใช้ในการสร้างการแจกแจงแบบปกติในการวิจัยครั้งนี้ ใช้วิธีของ Gauss ซึ่งเป็นวิธีที่สร้างการแจกแจงแบบปกติ ซึ่งมีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น 1 ส่วนค่าเฉลี่ยและความแปรปรวนค่าอื่น จะใช้วิธีแปลง-ข้อมูลในรูป $X = SMEAN + (SIGMA)\ X$ โดย SMEAN และ $(SIGMA)^2$ คือ ค่าเฉลี่ยและความแปรปรวนที่ต้องการ สำหรับรายละเอียดแสดงไว้ในภาคผนวก ก การใช้โปรแกรมย่อยนี้ใช้คำสั่ง CALL NORMAL (SMEAN, SIGMA, X) ค่า SMEAN, $(SIGMA)^2$ เป็นค่าพารามิเตอร์ ซึ่งกำหนดค่าเฉลี่ย และความแปรปรวน ซึ่งจะถูกส่งมาจากโปรแกรมหลัก ส่วน

ผลสัพธ์คือค่า x ซึ่งเป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น SMEAN และความแปรปรวนเป็น $(SIGMA)^2$

3.3.2 <u>การสร้างข้อมูล (x,y) ที่มีความสัมพันธ์กันเชิงเส้นตรง</u>

การสร้างข้อมูล (x,y) ที่มีความสัมพันธ์กันเชิงเส้นตรงนั้น ในการวิจัยครั้งนี้จะทำการสร้างค่า x ซึ่งเป็นค่าคงที่ก่อน แล้วจึงสร้างค่า y ที่มีความสัมพันธ์เชิงเส้นตรงกับค่า x ตามรูปแบบดังนี้คือ $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ เมื่อ β_0, β_1 เป็นค่าพารามิเตอร์ที่ถูกกำหนดขึ้นมา และ ε_i เป็นความคลาดเคลื่อนที่มีรูปแบบการแจกแจงเป็นแบบต่าง ๆ ดังที่กล่าวมาแล้วใน 3.3.1 สำหรับค่า x นั้นมีวิธีการสร้าง 2 แบบคือ แบบแรกสร้างค่า x จากการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น 30 และความแปรปรวนเป็น 100 กรณีนี้จะได้ช่วงห่างระหว่าง x มีค่าไม่เท่ากัน ส่วนอีกแบบจะใช้วิธีกำหนดค่า x ขึ้นมา เพื่อให้ช่วงระหว่าง x มีค่าเท่ากัน ซึ่งในการสร้างข้อมูลนั้นจะเริ่มจากการกำหนดขนาดของตัวอย่างที่ต้องการศึกษา พารามิเตอร์ β_0, β_1 ค่าเฉลี่ย ค่าความแปรปรวนและลักษณะการแจกแจงของ ε แล้วสร้างค่าคงที่ x แบบใดแบบหนึ่งดังกล่าว จากนั้นก็จะใช้คำสั่งสุ่มตัวอย่างเพื่อสร้าง ε ที่มีลักษณะการแจกแจงเป็นแบบต่าง ๆ ตามที่ต้องการศึกษา แล้วจึงสร้างค่า y ตามรูปแบบความสัมพันธ์ดังกล่าว

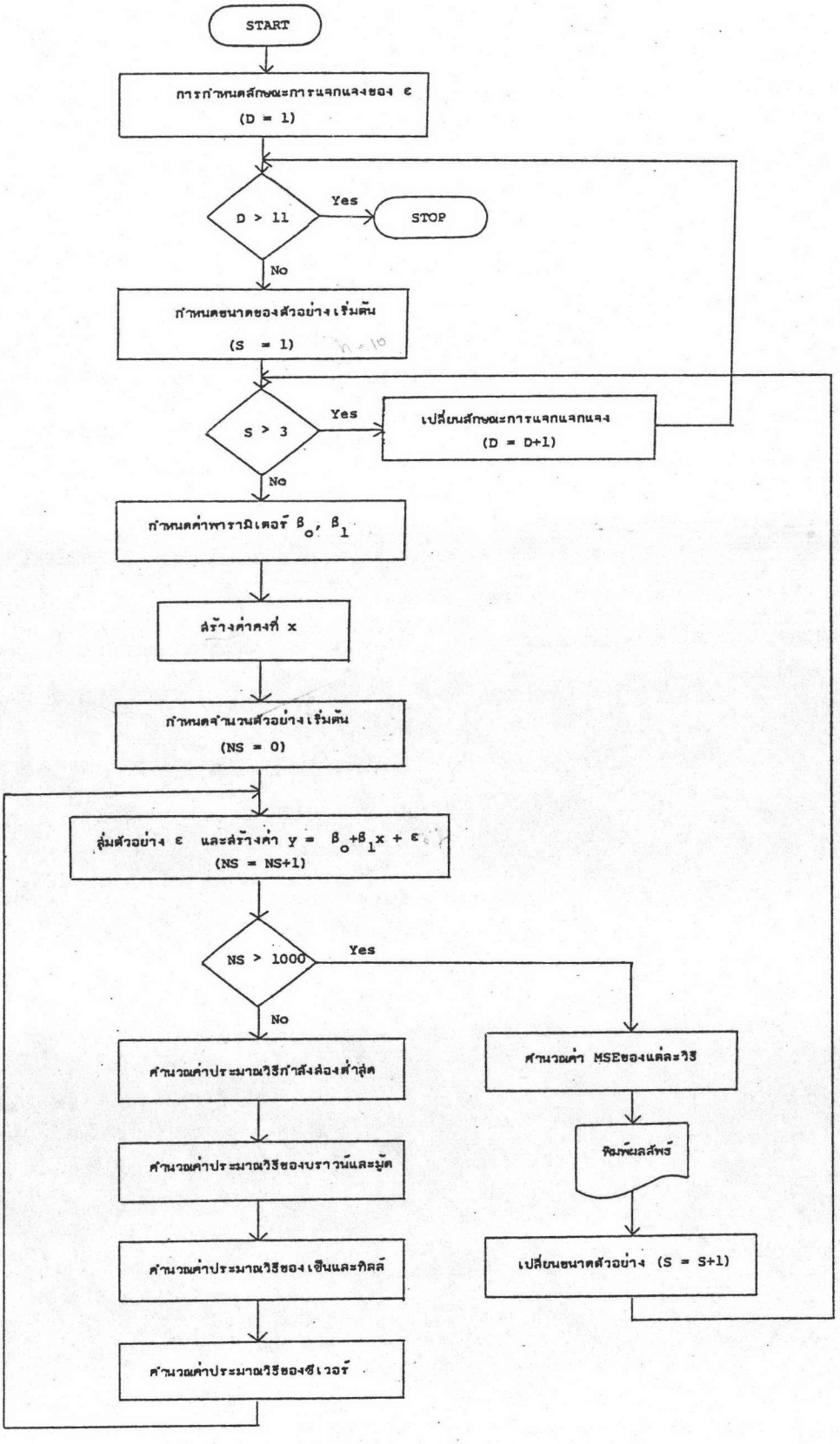
<u>หมายเหตุ</u> ในการสร้างค่าคงที่ x กรณีที่มีช่วงห่างไม่เท่ากัน สร้างจากการแจกแจงแบบปกติ เพื่อให้เกิดค่าที่เป็นธรรมชาติ และจากการทดลองกระทำที่ขนาดตัวอย่าง 10 แล้วพบว่า ค่าเฉลี่ยและค่าความแปรปรวนไม่มีผลต่อค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ค่าความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจของการทดสอบ อย่างไรก็ตามที่กำหนดค่าเฉลี่ยเป็น 30 และค่าความแปรปรวนเป็น 100 <u>เพื่อให้ได้ค่าที่ไม่ติดลบ และมีค่าห่างกันพอควร</u>

3.3.3 <u>การหาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ</u>

เมื่อสร้างข้อมูล (x,y) ที่มีความสัมพันธ์กันเชิงเส้นตรงตามรูปแบบที่ต้องการได้แล้ว ขั้นต่อไปคือ การทดลองเพื่อหาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ โดยจะต้องหาค่าประมาณของพารามิเตอร์ทุกวิธีก่อนคือ วิธีกำลังสองต่ำสุด จะต้องหาค่าประมาณ $\hat{\beta}_0, \hat{\beta}_1$ ที่ทำให้ ผลบวกของค่าคลาดเคลื่อนกำลังสอง ($\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$)

มีค่าต่ำสุด วิธีของบราวน์และมูด จะต้องหาค่าประมาณ $\hat{\beta}_0, \hat{\beta}_1$ ที่ทำให้มัธยฐานของค่าคลาดเคลื่อนของข้อมูล 2 กลุ่มที่แบ่งโดยมัธยฐานของ x มีค่าเป็น 0 วิธีของเซ็นและทิลล์ จะต้องหาค่าความชันย่อย $\binom{n}{2}$ ค่า แล้วเรียงค่าความชันย่อยจากน้อยไปมาก จะได้ค่าประมาณของพารามิเตอร์ β_1 คือมัธยฐานของความชันย่อยนั้น สำหรับวิธีของธีเวอร์ จะต้องหาความชันย่อย $\binom{n}{2}$ ค่าเช่นเดียวกับวิธีของเซ็นและทิลล์ แต่วิธีนี้จะมีค่าถ่วงน้ำหนักในการได้ความชันย่อยนั้น ๆ มาเกี่ยวข้องด้วย ซึ่งจะได้ค่าประมาณของพารามิเตอร์ β_1 คือ มัธยฐานของความน่าจะเป็นในการเกิดความชันย่อยนั้น เมื่อคำนวณค่าประมาณของพารามิเตอร์ครบทุกวิธีแล้ว ก็จะนำค่าประมาณที่ได้ลบออกจากค่าพารามิเตอร์ แล้วยกกำลังสองบวกสะสมไว้ จากนั้นก็ย้อนกลับไปสุ่มตัวอย่างชุดใหม่ ซึ่งยังคงใช้ค่าคงที่ x ชุดเดิม จนกระทั่งครบ 1,000 ครั้ง แล้วคำนวณความคลาดเคลื่อนกำลังสองเฉลี่ย จากนั้นจะเปลี่ยนขนาดตัวอย่างจนครบทุกรูปแบบที่ต้องการศึกษา โดยในแต่ละขนาดตัวอย่างจะสุ่มตัวอย่างซ้ำกัน 1,000 ครั้ง เมื่อได้ขนาดตัวอย่างครบทุกค่าแล้ว ขั้นต่อไปก็จะเปลี่ยนลักษณะการแจกแจงของความคลาดเคลื่อน (ε) จนครบทุกการแจกแจง โดยแต่ละการแจกแจงของความคลาดเคลื่อนจะใช้ขนาดตัวอย่างครบทุกรูปแบบ และแต่ละขนาดของตัวอย่างจะคำนวณความคลาดเคลื่อนกำลังสองเฉลี่ยของแต่ละวิธี ซึ่งสรุปเป็นผังงานได้ดังรูป 3.1

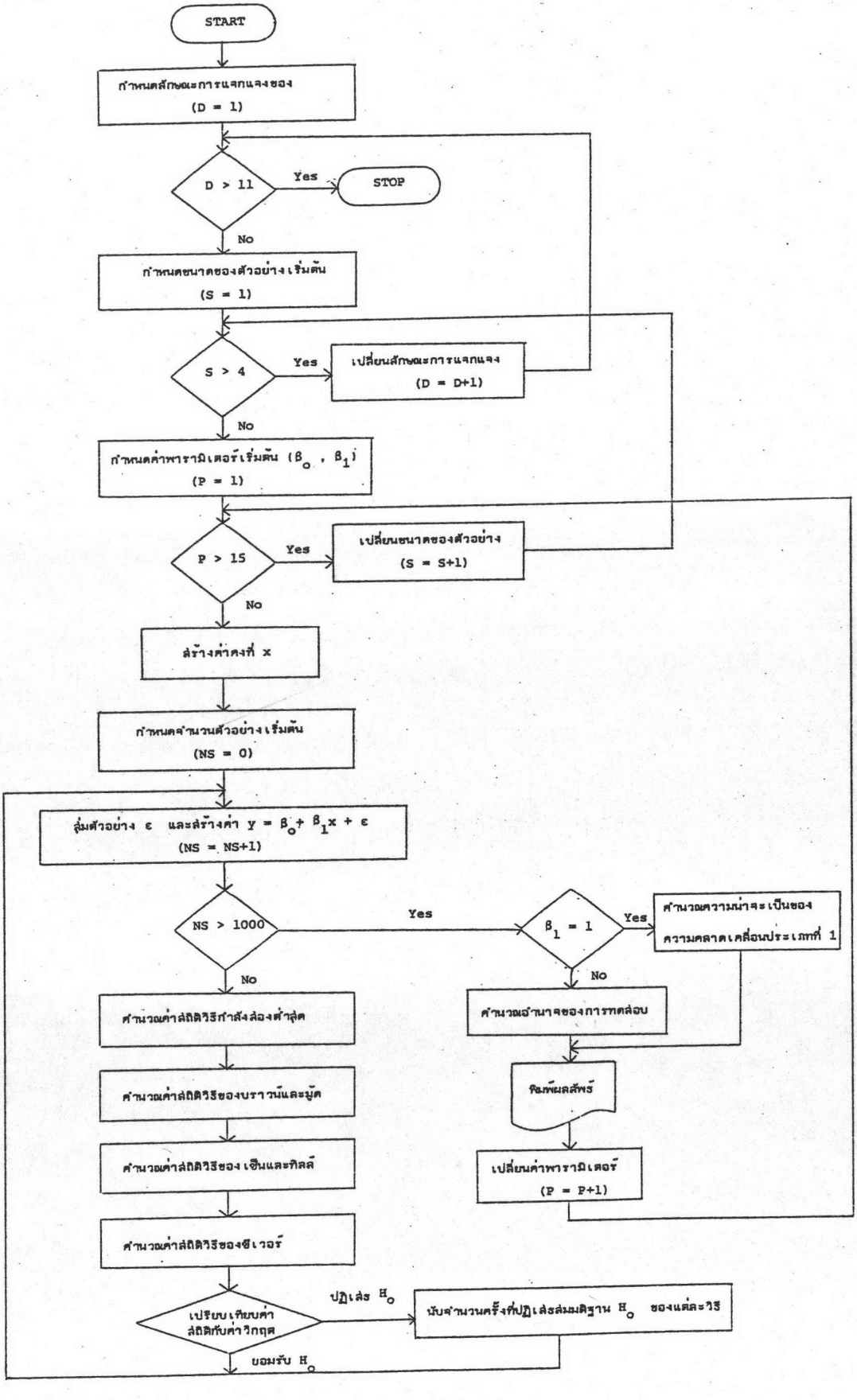
รูปที่ 3.1 แสดงผังงานสำหรับหาค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของตัวประมาณ 4 วิธี



3.3.4 การหาค่าอำนาจของการทดสอบและความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 สำหรับการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ β_1

ในการวิจัยครั้งนี้ จะเขียนโปรแกรมคำนวณค่าอำนาจของการทดสอบ และความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 แยกจากการหาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ ดังนั้นเมื่อสร้างข้อมูล (x , y) ที่มีความสัมพันธ์กันเชิงเส้นตรงตามรูปแบบที่ต้องการได้แล้ว ขั้นต่อไปคือ การทดลอง เพื่อหาค่าอำนาจของการทดสอบ และความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 โดยจะต้องคำนวณค่าสถิติของตัวสถิติทดสอบทุกวิธีก่อน คือ วิธีกำลังสองต่ำสุด จะต้องหาค่าประมาณ $\hat{\beta}_1$ ก่อน แล้วนำค่าพารามิเตอร์ β_{10} ลบออกจากค่าประมาณ $\hat{\beta}_1$ จากนั้นหารด้วยส่วนเบี่ยงเบนมาตรฐานของตัวประมาณนั้นก็จะได้ค่าสถิติทดสอบ วิธีของบราวน์และมูด้ จะต้องแบ่งข้อมูลออกเป็น 2 กลุ่มก่อน โดยใช้ค่ามัธยฐานของ x เป็นตัวแบ่ง แล้วนับจุดที่อยู่เหนือเส้น $y_i = a + \beta_{10}x_i$ และอยู่ทางซ้ายของมัธยฐานของ x จากนั้นจะนำค่าที่นับได้นี้ไปคำนวณค่าสถิติทดสอบ วิธีของเซ็นและทีลล์ จะพิจารณาจากสถิติของเคนดาลล์ระหว่าง x_i กับ $z_i = y_i - \beta_{10}x_i$ สำหรับวิธีของซีเวอร์ จะพิจารณาจากสถิติอันดับถ่วงน้ำหนัก ซึ่งคล้ายกับสถิติของเคนดาลล์ แต่จะมีค่าถ่วงน้ำหนักมาเกี่ยวข้องด้วย เมื่อคำนวณค่าสถิติครบทุกวิธีแล้ว ก็จะนำค่าสถิติมาเปรียบเทียบกับค่าวิกฤตเพื่อจะได้ตัดสินใจว่าจะปฏิเสธหรือยอมรับสมมติฐาน H_o ในกรณีที่ปฏิเสธสมมติฐาน H_o ให้นับจำนวนครั้งที่ปฏิเสธด้วย จากนั้นก็ย้อนกลับไปสุ่มตัวอย่างชุดใหม่จนกระทั่งครบ 1,000 ครั้ง แล้วคำนวณค่าความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 เมื่อพารามิเตอร์ β_1 มีค่าเป็น 1 หรือค่าอำนาจของการทดสอบ เมื่อพารามิเตอร์ β_1 ไม่เท่ากับ 1 จากนั้นก็จะเปลี่ยนค่าพารามิเตอร์ β_1 จนกระทั่งครบทุกค่าตามที่ต้องการ โดยในแต่ละค่าของพารามิเตอร์ β_1 จะสุ่มตัวอย่างซ้ำ ๆ กัน 1,000 ครั้ง เมื่อค่าพารามิเตอร์ β_1 ครบทุกค่าแล้ว ขั้นต่อไปจะเปลี่ยนขนาดของตัวอย่างจนครบทุกรูปแบบที่ต้องการศึกษา ในแต่ละขนาดตัวอย่างจะคำนวณค่าอำนาจของการทดสอบและความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ทุกค่าของพารามิเตอร์ β_1 ดังกล่าวเมื่อขนาดของตัวอย่างครบทุกรูปแบบแล้ว ขั้นต่อไปก็จะเปลี่ยนลักษณะการแจกแจงของความคลาดเคลื่อน (ε) จนครบทุกการแจกแจง โดยแต่ละการแจกแจงของความคลาดเคลื่อนจะใช้ขนาดตัวอย่างครบทุกรูปแบบ และในแต่ละขนาดตัวอย่างจะคำนวณความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 และอำนาจของการทดสอบจนครบทุกค่าของพารามิเตอร์ β_1 ที่ต้องการศึกษา ซึ่งสรุปเป็นผังงานได้ดังรูป 3.2

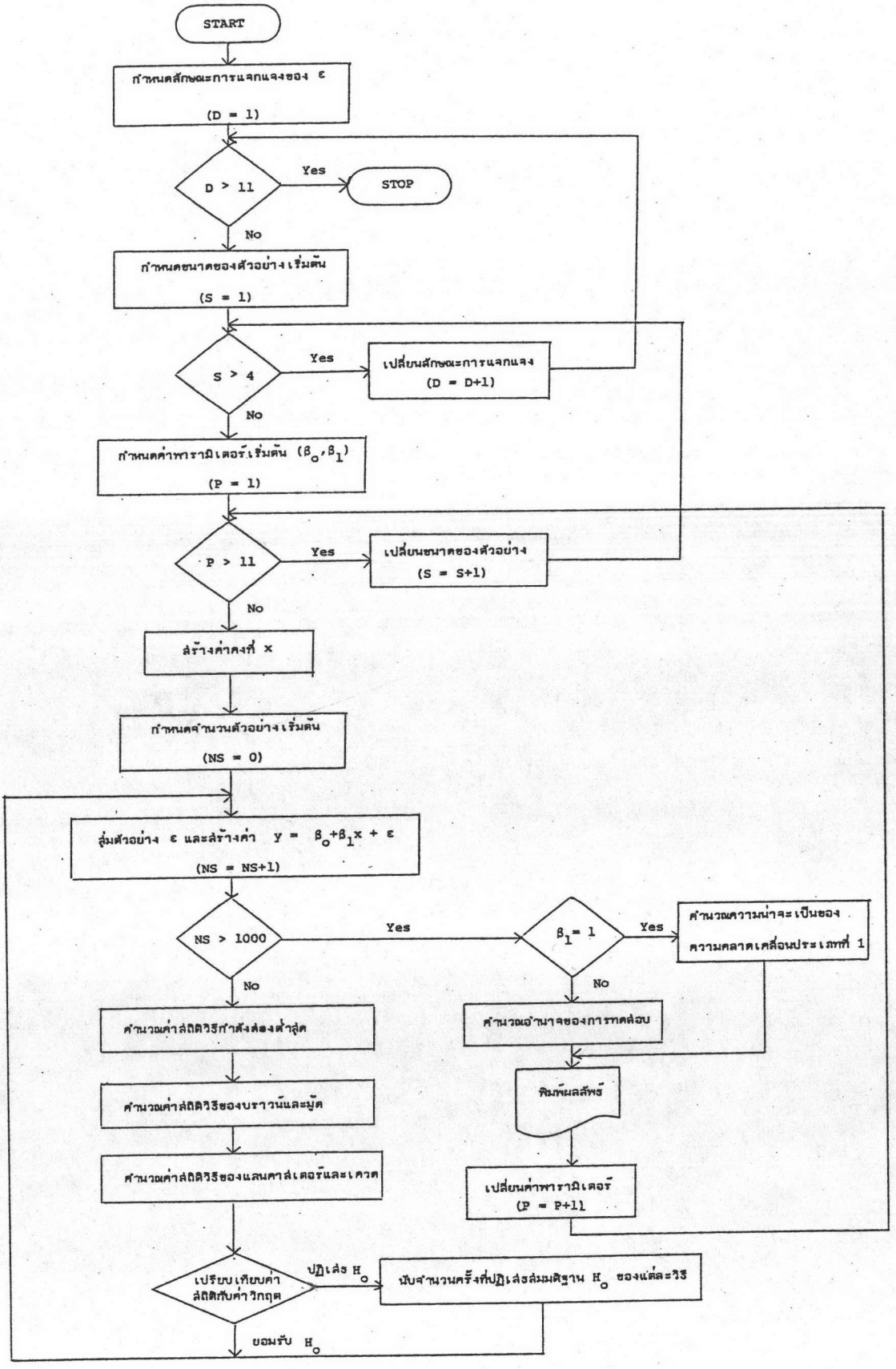
รูปที่3.2 แสดงผังงานสำหรับการหาค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และอำนาจของการทดสอบ 4 วิธี ในการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ β_1



3.3.5 การหาค่าอำนาจของการทดล่อบและความน่าจะเป็นที่จะเกิดความคลาด-เคลื่อนประเภทที่ 1 สำหรับการทดล่อบสมมติฐานเกี่ยวกับพารามิเตอร์ β_0, β_1

ในการวิจัยครั้งนี้ จะเขียนโปรแกรมคำนวณค่าอำนาจของการทดล่อบ และความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 แยกจากการหาค่าความคลาดเคลื่อนกำลังล่องเฉลี่ยของตัวประมาณ ดังนั้นเมื่อล่ร้างข้อมูล (x ,y) ที่มีความล่มพันธ์กันเชิงเล้นตรงตามรูปแบบที่ต้องการได้แล้ว ขั้นต่อไปคือ การทดลองเพื่อหาค่าอำนาจของการทดล่อบและความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 โดยจะต้องคำนวณค่าล่ถิติของตัวล่ถิติทดล่อบทุกวิธีก่อน คือ วิธีกำลังล่องต่ำล่ด จะต้องหาค่าประมาณ $\hat{\beta}_0$ และ $\hat{\beta}_1$ ก่อน แล้วจึงคำนวณล่ถิติทดล่อบเอฟ วิธีของบราวน์และมูด้ จะต้องแบ่งข้อมูลออกเป็น 2 กลุ่มก่อน โดยใช้ค่ามัธยฐานของ x เป็นตัวแบ่ง แล้วนับจุดที่อยู่เหนือเล้น $Y_i = \beta_{00} + \beta_{10}x_i$ และอยู่ทางซ้ายของมัธยฐานของ x และนับจุดที่อยู่เหนือเล้นเดียวกันนี้ และอยู่ทางขวาของมัธยฐานของ x จากนั้นจะนำค่าที่นับได้นี้ไปคำนวณค่าล่ถิติทดล่อบ สำหรับวิธีของแสนคาล่เตอร์และเควด จะต้องหาล่ถิติของเคนดาลล์ ระหว่าง x_i กับ $R_i = y_i - \beta_{00} - \beta_{10}x_i$ และล่ถิติเครื่องหมายก่อน แล้วนำค่าล่ถิติทั้งล่องมารวมกัน ก็จะได้ค่าล่ถิติทดล่อบของวิธีนี้ เมื่อคำนวณค่าล่ถิติครบทุกวิธีแล้ว ก็จะนำค่าล่ถิติมาเปรียบเทียบกับค่าวิกฤต เพื่อจะได้ตัดสินใจว่าจะปฏิเล่ธหรือยอมรับล่มมติฐาน H_o ในกรณีที่ปฏิเล่ธล่มมติฐาน H_o ให้นับจำนวนครั้งที่ปฏิเล่ธด้วย จากนั้นก็ย้อนกลับไปสุ่มตัวอย่างชุดใหม่จนกระทั่งครบ 1,000 ครั้ง แล้วคำนวณค่าความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 เมื่อพารามิเตอร์ β_1 มีค่าเป็น 1 หรือค่าอำนาจของการทดล่อบ เมื่อพารามิเตอร์ β_1 ไม่เท่ากับ 1 จากนั้นก็จะเปลี่ยนค่าพารามิเตอร์ β_1 จนกระทั่งครบทุกค่าตามที่ต้องการ โดยในแต่ละค่าของพารามิเตอร์ β_1 จะสุ่มตัวอย่างซ้ำ ๆ กัน 1,000 ครั้ง เมื่อค่าพารามิเตอร์ β_1 ครบทุกค่าแล้ว ขั้นต่อไปจะเปลี่ยนขนาดของตัวอย่างจนครบทุกรูปแบบที่ต้องการศึกษา ในแต่ละขนาดตัวอย่างจะคำนวณค่าอำนาจของการทดล่อบ และความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 ทุกค่าของพารามิเตอร์ β_1 ดังกล่าว เมื่อขนาดของตัวอย่างครบทุกรูปแบบแล้ว ขั้นต่อไปก็จะเปลี่ยนลักษณะการแจกแจงของความคลาดเคลื่อน (ε) จนครบทุกการแจกแจง โดยแต่ละการแจกแจงของความคลาดเคลื่อน จะใช้ขนาดตัวอย่างครบทุกรูปแบบ และในแต่ละขนาดตัวอย่างจะคำนวณความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 และอำนาจของการทดล่อบจนครบทุกค่าของพารามิเตอร์ β_1 ที่ต้องการศึกษา ซึ่งล่รุปเป็นผังงานได้ดังรูป 3.3

รูปที่ 3.3 แสดงผังงานสำหรับการหาค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และอำนาจของการทดสอบ 3 วิธี ในการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ β_0 , β_1



3.4 โปรแกรมที่ใช้ในการวิจัย

โปรแกรมที่ใช้ในการวิจัยทั้งหมดเขียนด้วยภาษาฟอร์แทรนโฟ ซึ่งแสดงรายละเอียดไว้ในภาคผนวก ข โดยลักษณะการทำงานของโปรแกรม แบ่งเป็น 3 ลักษณะ ดังตาราง 3.2

ตารางที่ 3.2 แสดงลักษณะการทำงานของโปรแกรมทั้งหมดที่ใช้ในการวิจัย

ลำดับที่ของโปรแกรม	ลักษณะการทำงาน	โปรแกรมย่อยที่เรียกใช้
1	คำนวณค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ 4 วิธี เมื่อพารามิเตอร์ $\beta_0 = 50$, $\beta_1 = 1$ และขนาดของตัวอย่างเป็น 10,15 และ 20 โดยรูปแบบการแจกแจงของความคลาดเคลื่อน (ε) เป็นแบบยูนิฟอร์ม แบบโลจิสติค แบบดับเบิ้ลเอ็กซ์โปเนนเชียล และแบบปกติปลอมปน ทั้งกรณีที่ช่วงห่างระหว่าง x เท่ากัน และไม่เท่ากัน	โปรแกรมสร้างตัวเลขสุ่ม โปรแกรมสร้างลักษณะการแจกแจงของ ε โปรแกรมจัดเรียงอันดับ โปรแกรมหาค่ามัธยฐาน
2	คำนวณค่าอำนาจของการทดสอบ หรือความคลาดเคลื่อนประเภทที่ 1 ในการทดสอบพารามิเตอร์ β_1 ของสถิติทดสอบ 4 วิธี เมื่อพารามิเตอร์ $\beta_o = 50$, β_1 =.1, .3,.5,.6,.7,.8,.9,1.0,1.1,1.2 1.3,1.4,1.5,1.7 และ 1.9 และขนาดตัวอย่างเป็น 10,15,20และ50 โดยศึกษาเมื่อกำหนดรูปแบบการแจกแจงของ ε และค่าคงที่ x เหมือนโปรแกรม 1	เหมือนโปรแกรม 1

ลำดับที่ของโปรแกรม	ลักษณะการทำงาน	โปรแกรมย่อยที่เรียกใช้
3	คำนวณค่าอำนาจของการทดล่อบ หรือความคลาดเคลื่อนประเภทที่ 1 ในการทดล่อบพารามิเตอร์ β_0, β_1 ของล่ถิติทดล่อบ 3 วิธี เมื่อพารามิเตอร์ $\beta_o = 50$, $\beta_1 = .5,.6,.7,.8$ .9,1.0,1.1,1.2,1.3,1.4 และ 1.5 โดยศึกษาเมื่อกำหนดขนาดตัวอย่าง รูปแบบการการแจกแจงของ ε และค่าคงที่ x เหมือนโปรแกรม 2	เหมือนโปแกรม 1