บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยในครั้งนี้สิ่งที่ต้องการศึกษาคือ การตรวจสอบความเหมาะสมของตัวแบบถดถอยเชิงเส้นโดยการแบ่งข้อมูลด้วยวิธีดูเพล็กซ์ ซึ่งประกอบด้วย การแบ่งข้อมูลด้วยวิธีดูเพล็กซ์ การประมาณค่าสัมประสิทธิ์ถดถอยเชิงเส้นโดยวิธีกำลังสองน้อยที่สุด การประมาณค่ารากที่ P ของสัดส่วนระหว่างดีเทอร์มิแนนท์ของตัวแปรอิสระที่ได้จากข้อมูลที่ใช้ในการประมาณค่ากับดีเทอร์มิแนนท์ของตัวแปรอิสระที่ได้จากข้อมูลที่ใช้ในการพยากรณ์ และศึกษาความแกร่งของการทดสอบเช่า ซึ่งจะพิจารณาจากความน่าจะเป็นของความผิดพลาดประเภทที่ 1 โดยจะศึกษาเมื่อความคลาดเคลื่อนมีการแจกแจงแบบ โลจิสติค ดับเบิ้ลเอ็กซ์โปเนนเชียล และปกติปลอมปน ทั้งนี้เนื่องจากการแจกแจงแบบต่าง ๆ ดังกล่าวเป็นการแจกแจงที่มีลักษณะการกระจายไปทางหางมากหรือมีหางยาวกว่าปกติ ซึ่งเป็นลักษณะที่สนใจศึกษา สำหรับรูปแบบของการแจกแจงแบบปกติปลอมปน จะศึกษาเมื่อเปอร์เซ็นต์การปลอมปนเป็น 1% 5% 10% และ 25% สเกลแฟคเตอร์ที่ใช้มี 2 ระดับคือ 3 และ 10 ซึ่งถ้าสเกลแฟคเตอร์มีค่าสูงจะทำให้เกิดค่าผิดปกติและจากการทดลองกระทำในการวิจัยครั้งนี้พบว่า ถ้าสเกลแฟคเตอร์ที่มีค่าน้อยกว่า 3 จะมีโอกาสของการเกิดค่าผิดปกติน้อย ส่วนสเกลแฟคเตอร์ที่มีค่ามากกว่า 10 จะมีโอกาสของการเกิดค่าผิดปกติมาก จึงทำการศึกษาเพียง 2 ระดับดังกล่าว ขนาดตัวอย่างที่สนใจศึกษาในการวิจัยครั้งนี้คือ 30 40 50 60 70 80 90 และ 100 ส่วนจำนวนตัวแปรอิสระที่สนใจศึกษาคือ 1 2 4 6 8 และ 10 สำหรับการวิจัยครั้งนี้จะใช้วิธีมอนติคาร์โลในการสร้างข้อมูลให้มีสภาพการแจกแจงตามที่ต้องการ

เนื่องจากมอนติคาร์โลเป็นเทคนิคที่ใช้ในการวิจัยนี้ ดังนั้นในตอนแรกของบทจะกล่าวถึงวิธีมอนติคาร์โลก่อน แล้วจึงแสดงรายละเอียดของแผนการทดลอง ขั้นตอนการวิจัย และโปรแกรมที่ใช้ในการวิจัยตามลำดับ ซึ่งรายละเอียดต่าง ๆ เป็นดังนี้

3.1 วิธีมอนติคาร์โล (Monte Carlo Method)

เทคนิคในการจำลองตัวแบบทางคณิตศาสตร์มีอยู่หลายวิธี วิธีมอนติคาร์โลก็เป็นวิธีหนึ่งที่นิยมใช้กันอย่างแพร่หลายในปัจจุบัน ซึ่งหลักการของวิธีมอนติคาร์โลนั้นจำลองตัวเลขสุ่ม (random number) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา

ในการวิจัยครั้งนี้จะใช้เทคนิคมอนติคาร์โลดังกล่าวในการสร้างข้อมูลที่มีสภาพการแจกแจงตามที่ต้องการ ซึ่งขั้นตอนของวิธีมอนติคาร์โลที่ใช้กันอยู่ในปัจจุบันแบ่งได้เป็น 3 ขั้นตอนดังนี้

3.1.1 การสร้างตัวเลขสุ่ม การใช้ตัวเลขสุ่มเป็นสิ่งสำคัญมากในวิธีมอนติคาร์โล ทั้งนี้เพราะว่าหลักการของวิธีมอนติคาร์โลนั้นจะใช้ตัวเลขสุ่มมาช่วยในการหาคำตอบของปัญหา ซึ่งลักษณะของตัวเลขสุ่มจะมีการแจกแจงแบบยูนิฟอร์มในช่วง (0 , 1) สำหรับวิธีการสร้างตัวเลขสุ่มนั้นมีผู้เสนอไว้หลายวิธี แต่วิธีหนึ่งที่ดีก็คือวิธีที่ไวท์และชมิดท์ (White and Schmidt 1975 : 421) เสนอไว้คือ ลักษณะของตัวเลขสุ่มที่เกิดขึ้นมีการแจกแจงแบบยูนิฟอร์มในช่วง (0 , 1) และเป็นอิสระกัน ซึ่งรายละเอียดแสดงไว้ในภาคผนวก ก

3.1.2 การประยุกต์ปัญหาที่ต้องการศึกษามาใช้กับตัวเลขสุ่ม ซึ่งขั้นตอนนี้ขึ้นอยู่กับลักษณะของปัญหาที่ต้องการศึกษา บางปัญหาอาจจะไม่ใช้ตัวเลขสุ่มโดยตรง แต่อาจมีขั้นตอนอื่นอีกหลาย ๆ ขั้นตอนซึ่งขั้นตอนเหล่านี้มีบางขั้นตอนที่ต้องใช้ตัวเลขสุ่ม

3.1.3 การทดลองกระทำ เมื่อประยุกต์ปัญหาให้ใช้กับตัวเลขสุ่มได้แล้ว ขั้นตอนต่อไปก็คือการทดลองโดยใช้กระบวนการของการสุ่ม (random process) มากระทำในลักษณะที่ซ้ำ ๆ กัน (replication) เพื่อหาคำตอบของปัญหาที่ต้องการศึกษา

3.2 แผนการทดลอง

การวิจัยในครั้งนี้กำหนดสถานการณ์ต่าง ๆ ที่ต้องการศึกษาโดย สุ่มตัวอย่างจากประชากรที่มีการแจกแจงแบบเดียวกัน ซึ่งลักษณะการแจกแจงที่สนใจศึกษาคือ โลจิสติค ดับเบิ้ล-เอ็กซ์โปเนนเชียล และปกติปลอมปน

สำหรับการแจกแจงแบบปกติปลอมปนกำหนดเปอร์เซ็นต์การปลอมปนเป็น 1% 5% 10% และ 25% ส่วนสเกลแฟคเตอร์มี 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)

นอกจากนี้ทุก ๆ การแจกแจงดังกล่าวขนาดตัวอย่างที่สนใจศึกษาคือ 30 40 50 60 70 80 90 และ 100 ส่วน จำนวนตัวแปรอิสระที่สนใจศึกษาคือ 1 2 4 6 8 และ 10 ซึ่งแสดงรายละเอียดดังตาราง 3.2

ตารางที่ 3.2 แสดงขนาดตัวอย่างและจำนวนตัวแปรอิสระทั้งหมดที่ใช้ในการวิจัยครั้งนี้

(30,1)	(30,2)	(30,4)	-	-	-
(40,1)	(40,2)	(40,4)	(40,6)	(40,8)	-
(50,1)	(50,2)	(50,4)	(50,6)	(50,8)	(50,10)
(60,1)	(60,2)	(60,4)	(60,6)	(60,8)	(60,10)
(70,1)	(70,2)	(70,4)	(70,6)	(70,8)	(70,10)
(80,1)	(80,2)	(80,4)	(80,6)	(80,8)	(80,10)
(90,1)	(90,2)	(90,4)	(90,6)	(90,8)	(90,10)
(100,1)	(100,2)	(100,4)	(100,6)	(100,8)	(100,10)

<u>หมายเหตุ</u> ขนาดตัวอย่าง 30 และ 40 จำนวนตัวแปรอิสระจะมีได้ไม่เกิน 4 และ 8 ตามลำดับ เนื่องจากข้อจำกัดของการแบ่งข้อมูลด้วยวิธีดูเพล็กซ์ จำนวนข้อมูลต้องมีไม่น้อยกว่า 2P + 20 ถึง 2P + 30 (P เป็นจำนวนสัมประสิทธิ์ถดถอยเชิงเส้นในสมการ)

3.3 <u>ขั้นตอนในการวิจัย</u>

ขั้นตอนในการวิจัยมีดังนี้คือ

1. การสร้างโปรแกรมย่อย (subroutine) สำหรับสร้างการแจกแจงของความคลาดเคลื่อนตามที่ต้องการศึกษา

2. การสร้างข้อมูล คือตัวแปรตาม (y) จาก ตัวแปรอิสระ (x) ซึ่งเป็นค่า

คงที่ และความคลาดเคลื่อนที่มีลักษณะการแจกแจงตามที่ต้องการศึกษา โดยให้ตัวแปรตามมีความสัมพันธ์เชิงเส้นตรงกับตัวแปรอิสระ

3. การแปลงตัวแปรอิสระให้เป็นมาตรฐานและทำให้ตั้งฉากกัน

4. การแบ่งข้อมูลด้วยวิธีดูเพล็กซ์

5. การประมาณค่าสัมประสิทธิ์ถดถอยเชิงเส้นโดยวิธีกำลังสองน้อยที่สุด

6. การประมาณค่ารากที่ P ของสัดส่วนระหว่างดีเทอร์มิแนนท์ของตัวแปรอิสระที่ได้จากข้อมูลที่ใช้ในการประมาณค่ากับดีเทอร์มิแนนท์ของตัวแปรอิสระที่ได้จากข้อมูลที่ใช้ในการพยากรณ์

7. การหาค่าความน่าจะเป็นที่จะเกิดความผิดพลาดประเภทที่ 1 ของการทดสอบเชา เพื่อตรวจสอบความแกร่งของการทดสอบนี้

สำหรับรายละเอียดของแต่ละขั้นตอนเป็นดังนี้

3.3.1 การสร้างโปรแกรมย่อยสำหรับสร้างการแจกแจงของความคลาดเคลื่อนตามที่ต้องการศึกษา

การสร้างลักษณะการแจกแจงของความคลาดเคลื่อนทุกรูปแบบที่ต้องการศึกษานั้น ใช้โปรแกรมภาษาฟอร์แทรนโฟ (FORTRAN IV) โดยใช้กับเครื่อง IBM 370/3031 ซึ่งการสร้างลักษณะการแจกแจงแบบต่าง ๆ นั้นจะต้องใช้ตัวเลขสุ่มเป็นพื้นฐานในการสร้าง สำหรับรายละเอียดในการสร้างการแจกแจงแบบต่าง ๆ เป็นดังนี้

3.3.1.1 การแจกแจงแบบปกติ โปรแกรมย่อยที่ใช้ในการสร้างการแจกแจงแบบปกติจะใช้วิธีของเกาส์ (Gauss) ซึ่งเป็นวิธีที่สร้างการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น μ และความแปรปรวนเป็น σ^2 โดย SMEAN และ $(SIGMA)^2$ คือค่าเฉลี่ยและความแปรปรวนที่กำหนด สำหรับรายละเอียดแสดงไว้ในภาคผนวก ก การใช้โปรแกรมย่อยนี้ใช้คำสั่ง CALL NORMAL (SMEAN , SIGMA, X) โดยค่าเฉลี่ยและความแปรปรวนจะถูกส่งมาจากโปรแกรมหลักส่วนผลลัพธ์คือค่า X ซึ่งเป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น SMEAN และความแปรปรวนเป็น $(SIGMA)^2$

3.3.1.2 การแจกแจงแบบโลจิสติค โปรแกรมย่อยที่ใช้ในการสร้างการแจกแจงแบบโลจิสติคใช้วิธีการแปลงผกผัน (inverse transformation) รายละเอียดแสดงไว้ในภาคผนวก ก การใช้โปรแกรมย่อยนี้ใช้คำสั่ง CALL LOGIST (ALPHA, BETA, X)

โดยค่า ALPHA และ BETA เป็นค่าที่กำหนดขึ้นเพื่อให้ได้ค่าเฉลี่ยและความแปรปรวนตามที่ต้องการ ค่า ALPHA และ BETA นี้จะถูกส่งมาจากโปรแกรมหลัก ส่วนผลสัพธ์คือ X ซึ่งเป็นตัวแปรสุ่มที่มีการแจกแจงแบบโลจิสติคที่มีค่าเฉลี่ยเป็น ALPHA และความแปรปรวนเป็น $\frac{1}{3}\pi^2$ (BETA)^2

3.3.1.3 การแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียล โปรแกรมย่อยที่ใช้ในการสร้างการแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียลใช้วิธีการแปลงผกผัน ซึ่งรายละเอียดแสดงไว้ในภาคผนวก ก การใช้โปรแกรมย่อยนี้ใช้คำสั่ง CALL DOUBLE (ALPHA, BETA,X) โดยค่า ALPHA และ BETA เป็นค่าที่กำหนดขึ้นเพื่อให้ได้ค่าเฉลี่ยและความแปรปรวนตามที่ต้องการ ค่า ALPHA และ BETA นี้จะถูกส่งมาจากโปรแกรมหลัก ส่วนผลสัพธ์ที่ได้คือ X ซึ่งเป็นตัวแปรสุ่มที่มีการแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียลที่มีค่าเฉลี่ยเป็น ALPHA และความแปรปรวนเป็น 2(BETA)^2

3.3.1.4 การแจกแจงแบบปกติปลอมปน โปรแกรมย่อยที่ใช้ในการสร้างการแจกแจงแบบปกติปลอมปนใช้วิธีการแปลงข้อมูลมาจากการแจกแจงแบบปกติ โดยพิจารณาการแจกแจงแบบปกติปลอมปนของตัวแปร X ซึ่งมาจาก $F = (1-P)\ N(\mu, \sigma^2) + PN(\mu, c^2\sigma^2)$ เมื่อ P เป็นเปอร์เซ็นต์การปลอมปน c เป็นสเกลแฟคเตอร์ที่จะทำให้เกิดค่าผิดปกติ ($c > 0$) หมายความว่า ตัวแปร X จะมาจากการแจกแจงแบบ $N(\mu, \sigma^2)$ ด้วยความน่าจะเป็น (1-P) และมาจากการแจกแจงแบบ $N(\mu, c^2\sigma^2)$ ด้วยความน่าจะเป็น P สำหรับรายละเอียดแสดงไว้ในภาคผนวก ก การใช้โปรแกรมย่อยนี้ใช้คำสั่ง CALL SCALE (C, P, SMEAN, SIGMA, X) C และ P เป็นค่าที่กำหนดสเกลแฟคเตอร์และเปอร์เซ็นต์การปลอมปน ส่วน SEMEAN และ (SIGMA)^2 เป็นค่าเฉลี่ยและความแปรปรวนที่กำหนด ซึ่งค่า C, P SMEAN และ SIGMA จะถูกส่งมาจากโปรแกรมหลัก ส่วนผลสัพธ์ที่ได้คือ X ซึ่งเป็นตัวแปรสุ่มที่มาจากการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น SMEAN และความแปรปรวนเป็น (SIGMA)^2 (1-P) % ส่วนที่เหลือจะมาจากการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น SMEAN และความแปรปรวนเป็น (C SIGMA)^2

3.3.2 การสร้างข้อมูลให้มีความสัมพันธ์เชิงเส้นตรง

การวิจัยในครั้งนี้จะสร้างตัวแปรอิสระ X ซึ่งเป็นค่าคงที่ก่อน แล้วจึงสร้างตัวแปร Y ที่มีความสัมพันธ์เชิงเส้นกับ X ให้มีลักษณะการแจกแจงตามความคลาดเคลื่อนที่กำหนด

ตามรูปแบบของความสัมพันธ์ดังนี้คือ $\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ เมื่อ $\underset{\sim}{\beta}$ เป็นค่าพารามิเตอร์ที่ถูกกำหนดขึ้นมา และ $\underset{\sim}{\varepsilon}$ เป็นความคลาดเคลื่อนที่มีรูปแบบการแจกแจงเป็นแบบต่าง ๆ ดังที่กล่าวมาแล้วใน 3.3.1 สำหรับการสร้าง $\underset{\sim}{Y}$ นั้นจะเริ่มจากการกำหนดขนาดตัวอย่างและจำนวนตัวแปรอิสระที่ต้องการศึกษา พารามิเตอร์ $\underset{\sim}{\beta}$ ค่าเฉลี่ย ค่าความแปรปรวนและลักษณะการแจกแจงของความคลาดเคลื่อน รวมทั้งค่าคงที่ X แล้วจึงสร้าง $\underset{\sim}{Y}$ ตามรูปแบบดังกล่าว

3.3.3 การแปลงตัวแปรอิสระให้เป็นมาตรฐานและทำให้ตั้งฉากกัน

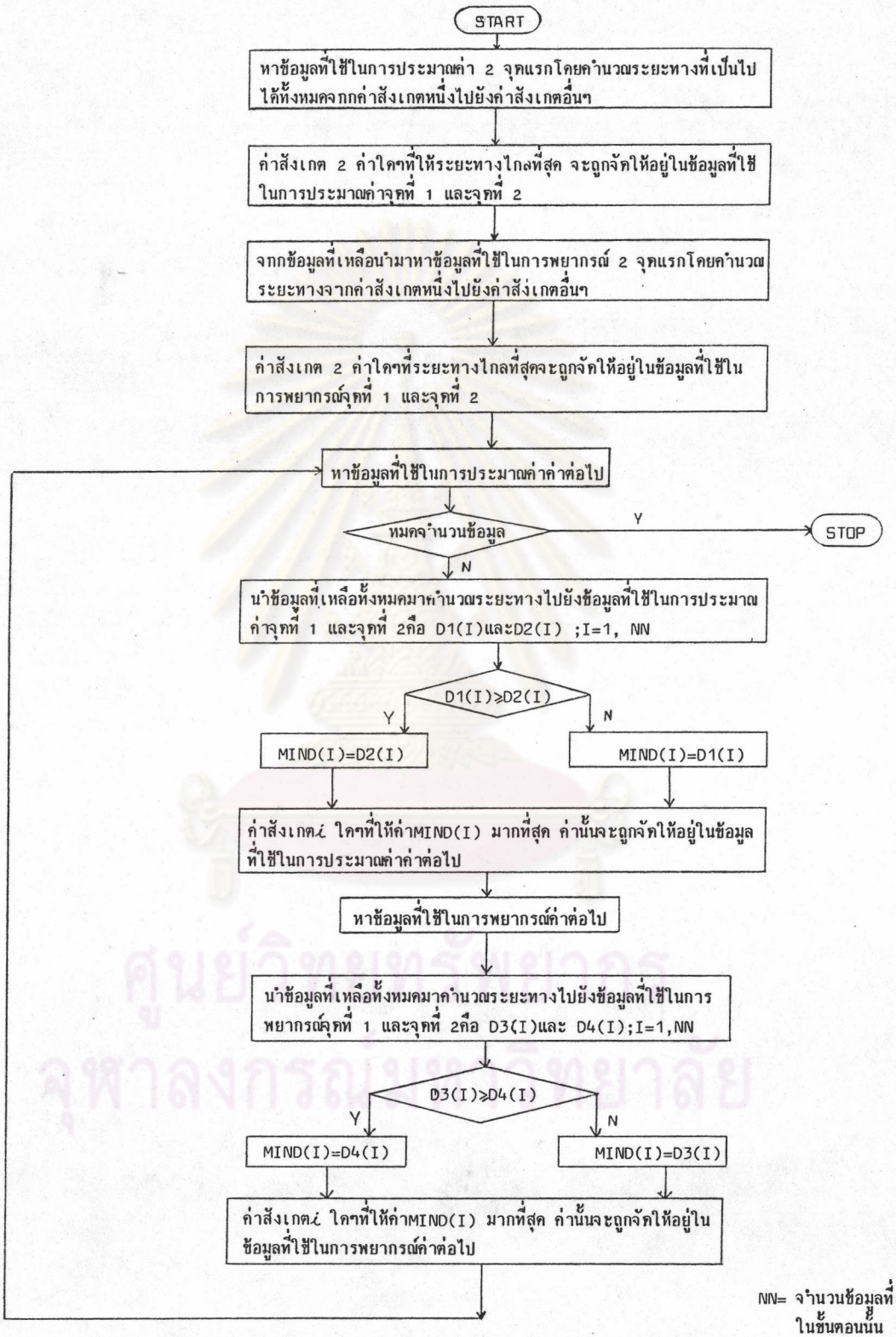
เมื่อสร้างข้อมูลที่มีความสัมพันธ์เชิงเส้นตามรูปแบบที่ต้องการได้แล้ว ในกรณีที่ตัวแปรอิสระมีมากกว่า 1 ตัวจะต้องแปลงตัวแปรอิสระให้เป็นมาตรฐานและทำให้ตั้งฉากกันก่อน ซึ่งรายละเอียดอยู่ในบทที่ 2 ส่วนในตอนนี้จะกล่าวถึงรายละเอียดในการเรียกโปรแกรมย่อยมาใช้ในการทำให้ตัวแปรอิสระตั้งฉากกัน โดยการทำให้ตัวแปรอิสระที่ถูกแปลงให้เป็นมาตรฐานแล้วจากโปรแกรมหลัก (ในที่นี้คือ Z) อยู่ในรูปใหม่คือ $W = ZT^{-1}$ ซึ่งการหาเมตริกซ์ T จะกระทำโดย ในขั้นแรกจะเรียกโปรแกรมย่อยที่ใช้ในการหาเมตริกซ์ $X\acute{X}$ การใช้โปรแกรมย่อยนี้ใช้คำสั่ง CALL PRIME (X, N, NP, XTX) เมื่อ X เป็นเมตริกซ์ของตัวแปรอิสระที่ต้องการหาเมตริกซ์ $X\acute{X}$ N และ NP เป็นจำนวนข้อมูลและจำนวนตัวแปรอิสระตามลำดับ ซึ่ง X ,N และ NP จะถูกส่งมาจากโปรแกรมหลัก ส่วนผลลัพธ์ที่ได้คือ XTX ซึ่งเป็นเมตริกซ์ของ $X\acute{X}$ ที่เป็นเมตริกซ์จัตุรัส (square matrix) และมีขนาด (NPxNP) จากนั้นจะหาเมตริกซ์ T โดยการใช้โปรแกรมย่อยที่ใช้คำสั่ง CALL ORTHO (XTX , NP, T) โดย XTX เป็นเมตริกซ์ของ $X\acute{X}$ และ NP คือขนาดของเมตริกซ์ $X\acute{X}$ ซึ่ง XTX และ NP จะถูกส่งมาจากโปรแกรมหลัก ส่วนผลลัพธ์คือเมตริกซ์ T ที่มีขนาด (NPxNP) เมื่อได้เมตริกซ์ T แล้วขั้นต่อไปจะหาเมตริกซ์ T^{-1} โดยเรียกโปรแกรมย่อยที่ใช้คำสั่ง CALL MINV (AAA, N,DDD, IJK,MK) โดย AAA เป็นเมตริกซ์ที่ต้องการหาอินเวอร์ส และ N เป็นจำนวนข้อมูล ส่วนผลลัพธ์ที่ได้คือ เมตริกซ์ AAA ที่เป็นอินเวอร์สของเมตริกซ์ที่ถูกส่งเข้ามา DDD คือ ดีเทอร์มิแนนท์ของเมตริกซ์ที่ถูกส่งเข้ามาหาอินเวอร์ส IJK และ MK คือค่าที่ใช้คำนวณภายในโปรแกรมย่อย เมื่อได้เมตริกซ์ T^{-1} แล้ว หลังจากนั้นจึงนำมาใช้ในการแปลงให้ตัวแปรอิสระอยู่ในรูปใหม่คือ $W = ZT^{-1}$

3.3.4 การแบ่งข้อมูลด้วยวิธีดูเพล็กซ์

สำหรับรายละเอียดของการแบ่งข้อมูลด้วยวิธีดูเพล็กซ์นี้ได้กล่าวมาแล้วในบทที่ 2 ดังนั้นในส่วนนี้จะแสดงผังงาน (flowchart) ของการแบ่งข้อมูลด้วยวิธีดูเพล็กซ์ ซึ่งสรุปได้ดังรูป 3.1

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

รูป 3.1 แสดงผังงานในการแบ่งข้อมูลด้วยวิธีคูเพล็กซ์



NN= จำนวนข้อมูลที่เหลือในขั้นตอนนั้น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

3.3.5 การประมาณค่าสัมประสิทธิ์ถดถอยเชิงเส้นโดยวิธีกำลังสองน้อยที่สุด

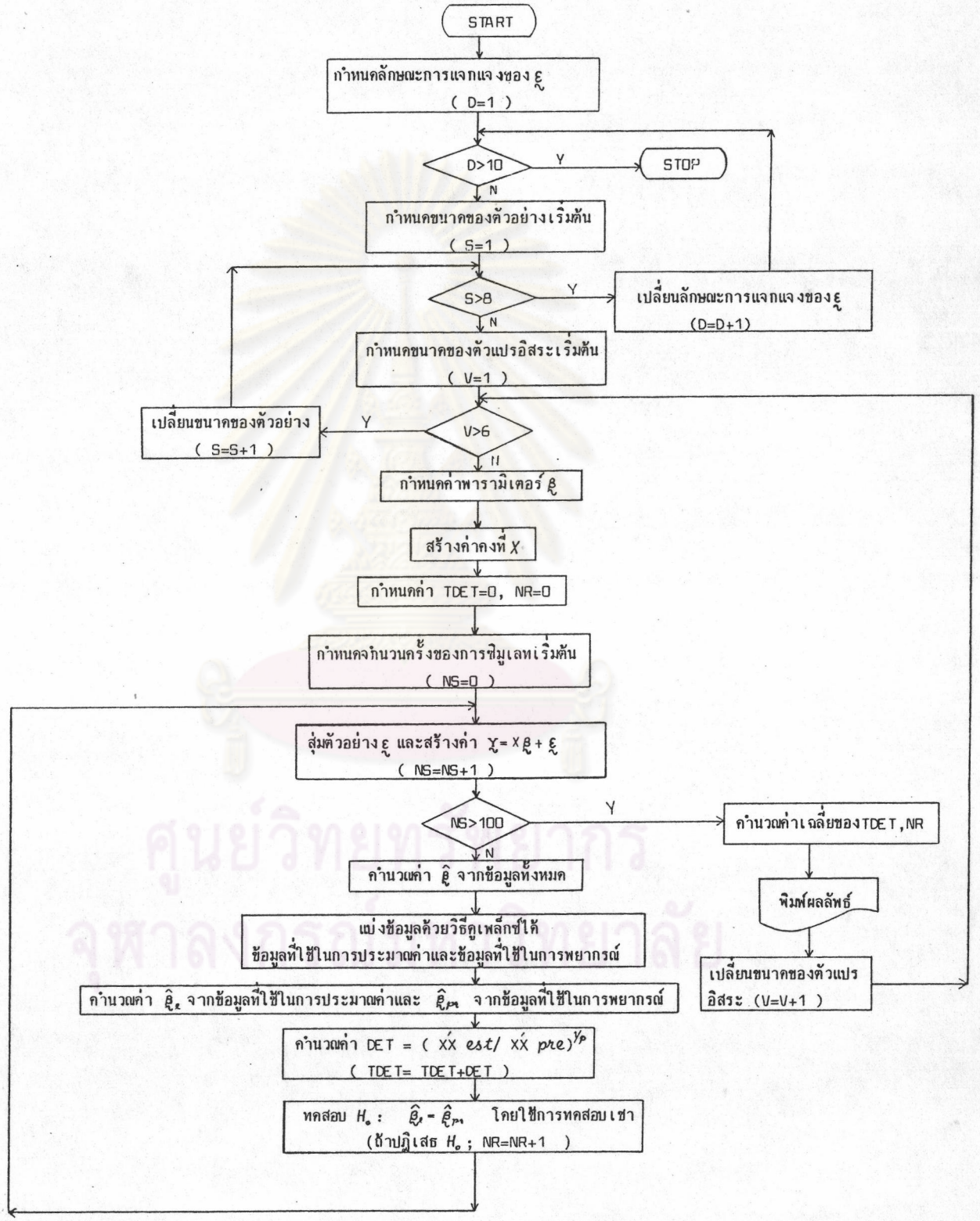
โปรแกรมย่อยที่ใช้ในการประมาณค่าสัมประสิทธิ์ถดถอยเชิงเส้นจากข้อมูล จะใช้โปรแกรมย่อยที่ใช้คำสั่ง CALL OLS (B, X, N, NP1, NP2) โดย X คือเมตริกซ์ของตัวแปรอิสระและตัวแปรตาม N คือจำนวนข้อมูล NP1 คือจำนวนตัวแปรอิสระบวก 1 และ NP2 คือจำนวนตัวแปรอิสระบวก 2 ส่วนผลลัพธ์ที่ได้คือ B ซึ่งเป็นเวคเตอร์ของสัมประสิทธิ์ถดถอยเชิงเส้นที่มีขนาด (NP1x1)

3.3.6 การประมาณค่ารากที่ P ของสัดส่วนระหว่างดีเทอร์มิแนนท์ของตัวแปรอิสระที่ได้จากข้อมูลที่ใช้ในการประมาณค่ากับดีเทอร์มิแนนท์ของตัวแปรอิสระที่ได้จากข้อมูลที่ใช้ในการพยากรณ์

เมื่อทำการแบ่งข้อมูลด้วยวิธีดูเพล็กซ์จนได้ข้อมูลเป็น 2 ชุดคือ ข้อมูลที่ใช้ในการประมาณค่าและข้อมูลที่ใช้ในการพยากรณ์ ขั้นตอนต่อไปก็คือ จะต้องหาค่าดีเทอร์มิแนนท์ของเมตริกซ์ $X\acute{X}$ จากข้อมูลทั้ง 2 ชุด โดยใช้โปรแกรมย่อยที่ใช้คำสั่ง CALL PRIME (X,N,MP,XTX) ซึ่งจะได้เมตริกซ์ $X\acute{X}$ ของข้อมูลทั้ง 2 ชุด จากนั้นก็ใช้โปรแกรมย่อยที่ใช้คำสั่ง CALL MINV (AAA, N, DDD, IJK, MK) จะได้ค่าดีเทอร์มิแนนท์ของข้อมูลทั้ง 2 ชุด แล้วจึงนำมาหาค่ารากที่ P ของสัดส่วนระหว่างดีเทอร์มิแนนท์ที่ได้จากข้อมูลทั้ง 2 ชุด เมื่อ P คือจำนวนสัมประสิทธิ์ถดถอยเชิงเส้นในสมการ ในการวิจัยครั้งนี้จะสร้างข้อมูล 100 ครั้ง แต่ละครั้งจะคำนวณค่ารากที่ P ของสัดส่วนระหว่างดีเทอร์มิแนนท์ที่ได้จากข้อมูลที่ใช้ในการประมาณค่ากับดีเทอร์มิแนนท์ที่ได้จากข้อมูลที่ใช้ในการพยากรณ์บวกสะสมไว้ แล้วย้อนกลับไปสุ่มตัวอย่างชุดใหม่ ซึ่งยังคงใช้ค่าคงที่ X ชุดเดิม จนกระทั่งครบ 100 ครั้ง แล้วคำนวณค่าเฉลี่ยของค่ารากที่ P ของสัดส่วนระหว่างดีเทอร์มิแนนท์ที่ได้จากข้อมูลทั้ง 2 ชุดใน 100 ครั้ง จากนั้นจะเปลี่ยนขนาดตัวอย่างและจำนวนตัวแปรจนครบทุกรูปแบบที่ต้องการศึกษา โดยในแต่ละรูปแบบจะกระทำซ้ำกัน 100 ครั้ง เมื่อได้ขนาดตัวอย่างและจำนวนตัวแปรครบทุกรูปแบบที่ต้องการศึกษาแล้ว ขั้นต่อไปก็จะเปลี่ยนลักษณะการแจกแจงของความคลาดเคลื่อนจนครบทุกการแจกแจง โดยแต่ละการแจกแจงของความคลาดเคลื่อนจะใช้ขนาดตัวอย่างและจำนวนตัวแปรจนครบทุกรูปแบบที่ต้องการศึกษา ซึ่งสรุปเป็นผังงานได้ดังรูป 3.2

I10292214

รูป 3.2 แสดงผังงานสำหรับการหาค่ารากที่ p ของสัดส่วนระหว่างดีเทอร์มิแนนท์ของตัวแปรอิสระที่ได้จากข้อมูลที่ใช้ในการประมาณค่ากับดีเทอร์มิแนนท์ของตัวแปรอิสระที่ได้จากข้อมูลที่ใช้ในการพยากรณ์ และการหาค่าความน่าจะเป็นที่จะเกิดความผิดพลาดประเภทที่ 1 ของการทดสอบเชา



3.3.7 การหาค่าความน่าจะเป็นที่จะเกิดความผิดพลาดประเภทที่ 1 ของการทดสอบเชา

ภายหลังจากการประมาณค่าสัมประสิทธิ์ถดถอยเชิงเส้นจากข้อมูลทั้งหมด ($\hat{\beta}$) จากข้อมูลที่ใช้ในการประมาณค่า ($\hat{\beta}_{e}$) และจากข้อมูลที่ใช้ในการพยากรณ์ ($\hat{\beta}_{pr}$) เรียบร้อยแล้ว ขั้นต่อไปก็คือ การหาค่าความน่าจะเป็นที่จะเกิดความผิดพลาดประเภทที่ 1 ของการทดสอบเชา โดยต้องคำนวณค่าสถิติเอฟของเชาก่อน จากนั้นนำค่าสถิติมาเปรียบเทียบกับค่าวิกฤต เพื่อจะได้ตัดสินใจว่าจะปฏิเสธหรือยอมรับสมมติฐาน $H_o : \beta_e = \beta_{pr}$ ในกรณีที่ปฏิเสธสมมติฐาน H_o ให้นับจำนวนครั้งที่ปฏิเสธบวกสะสมไว้ จากนั้นย้อนกลับไปสุ่มตัวอย่างชุดใหม่ ซึ่งยังคงใช้ค่าคงที่ X ชุดเดิม จนกระทั่งครบ 100 ครั้ง แล้วคำนวณค่าเฉลี่ยของจำนวนครั้งที่ปฏิเสธสมมติฐาน H_o จากนั้นจึงเปลี่ยนขนาดตัวอย่างและจำนวนตัวแปรอิสระจนครบทุกรูปแบบที่ต้องการศึกษา โดยในแต่ละรูปแบบจะกระทำซ้ำกัน 100 ครั้ง เมื่อได้ขนาดตัวอย่างและจำนวนตัวแปรอิสระครบทุกรูปแบบที่ต้องการศึกษาแล้ว ขั้นต่อไปจะเปลี่ยนลักษณะการแจกแจงของความคลาดเคลื่อนจนครบทุกการแจกแจง โดยแต่ละการแจกแจงของความคลาดเคลื่อนจะใช้ขนาดตัวอย่างและจำนวนตัวแปรอิสระครบทุกรูปแบบที่ต้องการศึกษา ซึ่งสรุปเป็นผังงานได้ดังรูป 3.2

3.4 โปรแกรมที่ใช้ในการวิจัย

โปรแกรมที่ใช้ในการวิจัยทั้งหมดเขียนด้วยภาษาฟอร์แทรนโฟ ซึ่งใช้กับเครื่อง IBM 370/3031 โดยลักษณะการทำงานของโปรแกรมแสดงในตาราง 3.3

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

ตาราง 3.3 แสดงลักษณะการทำงานของโปรแกรมทั้งหมดที่ใช้ในการวิจัย

ลำดับที่ของโปรแกรม	ลักษณะการทำงาน	โปรแกรมย่อยที่เรียกใช้
1	คำนวนค่าสัมประสิทธิ์ถดถอยเชิงเส้นของข้อมูลทั้งหมดเมื่อกำหนดขนาดของข้อมูลมีค่า 30 40 50 60 70 80 90 และ 100 จำนวนตัวแปรอิสระมีค่า 1 2 4 6 8 และ 10 โดยลักษณะการแจกแจงของความคลาดเคลื่อนคือ โลจิสติค ดับเบิ้ลเอ็กซ์โปเนนเชียล และปกติปสอมปน	โปรแกรมสร้างตัวเลขสุ่ม โปรแกรมสร้างลักษณะการแจกแจงของความคลาดเคลื่อน โปรแกรมหาค่าสัมประสิทธิ์ถดถอยเชิงเส้น
2	แบ่งข้อมูลด้วยวิธีดูเพล็กซ์เป็น 2 ชุดคือ ข้อมูลที่ใช้ในการประมาณค่า และข้อมูลที่ใช้ในการพยากรณ์	โปรแกรมหาเมตริกซ์ XX́ โปรแกรมที่ทำให้ตัวแปรอิสระตั้งฉากกัน โปรแกรมหาเมตริกซ์อินเวอร์ส
3	คำนวณค่าสัมประสิทธิ์ถดถอยเชิงเส้นจากข้อมูลทั้ง 2 ชุด	โปรแกรมหาค่าสัมประสิทธิ์ถดถอยเชิงเส้น
4	คำนวณค่ารากที่ P ของสัดส่วนระหว่างดีดีเทอร์มิแนนท์ที่ได้จากข้อมูลที่ใช้ในการประมาณค่ากับดีเทอร์มิแนนท์ที่ได้จากข้อมูลที่ใช้ในการพยากรณ์	โปรแกรมหาเมตริกซ์ XX́ โปรแกรมหาค่าดีเทอร์มิแนนท์
5	คำนวณค่าความน่าจะเป็นที่จะเกิดความผิดพลาดประเภทที่ 1ของการทดสอบเช่า	เหมือนโปรแกรม 1 2 และ 3