บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบตัวสถิติทดสอบที่ใช้ในการตรวจสอบข้อมูลผิดปรกติในการวิเคราะห์ความถดถอยเชิงเส้นของวิธีการตรวจสอบ 4 วิธี คือ วิธีของเมอร์วิน จี มาราชิง(MV), วิธีของฮาไดและไชมันนอฟฟี(HS), วิธีเวียนเกิดโดยลำดับ(SRM) และวิธีเวียนเกิดตัดแปร(MRM) ผู้วิจัยจะศึกษาในกรณีที่ความคลาดเคลื่อนมีการแจกแจงเบ้ และหางยาวกว่าการแจกแจงปรกติ และผู้วิจัยจะทำการแปลงข้อมูลให้คู่เข้าสู่การแจกแจงปรกติก่อนที่จะใช้วิธีกำลังสองน้อยสุด ในการประมาณค่าสัมประสิทธิ์การถดถอยเมื่อความคลาดเคลื่อนมีการแจกแจงเบ้ เพื่อนำไปคำนวณหาค่าตกค้าง (residual) ต่อไป การวิจัยครั้งนี้เป็นการวิจัยเชิงทดลองโดยใช้วิธีมอนติคาร์โล (Monte Carlo method) มาช่วยในการหาคำตอบของปัญหา ซึ่งมีรายละเอียดดังนี้

3.1 วิธีมอนติคาร์โล

ในปี พ.ศ. 2533 สมชาย รัตนเลิศนุสรณ์ กล่าวว่า "วิธีมอนติคาร์โลเป็นวิธีการจำลองตัวแบบทางคณิตศาสตร์วิธีหนึ่งที่นิยมใช้กันอย่างแพร่หลายในปัจจุบันซึ่งจะทำการจำลองเลขสุ่ม (ramdom number) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา" ในการวิจัยครั้งนี้ผู้วิจัยจะใช้วิธีมอนติคาร์โลช่วยสร้างข้อมูลให้มีการแจกแจงตามที่ต้องการตามขั้นตอนต่อไปนี้

3.1.1 <u>การสร้างตัวเลขสุ่ม</u>

สิ่งสำคัญในวิธีมอนติคาร์โล คือ การสร้างตัวเลขสุ่มเพราะต้องนำมาช่วยในการหาคำตอบของปัญหา วิธีการสร้างตัวเลขสุ่มมีหลายวิธีแต่วิธีที่ดีลักษณะของตัวเลขสุ่มที่สร้างขึ้นมาจะต้องมีการแจกแจงสม่ำเสมอ (Uniform distribution) ในช่วง (0,1) และตัวเลขสุ่มแต่ละตัวเป็นอิสระกัน รายละเอียดแสดงในภาคผนวก ก.

3.1.2 การประยุกต์ตัวเลขสุ่มให้สามารถใช้กับปัญหาที่ต้องการศึกษา

ในขั้นตอนนี้จะขึ้นอยู่กับลักษณะของปัญหา ว่าเราจะนำเลขสุ่มนั้นไปใช้โดยตรงเลยหรือไม่ เพราะบางปัญหาอาจจะไม่ใช้ตัวเลขสุ่มโดยตรงแต่อาจจะมีบางขั้นตอนที่ต้องการใช้เลขสุ่ม

3.1.3 การทดลองซ้ำ (replication)

เมื่อเราสามารถประยุกต์ตัวเลขสุ่มให้มีลักษณะที่ต้องการได้แล้ว ขั้นตอนต่อไปเราจะทำการทดลองโดยใช้กระบวนการสุ่ม (random process)ซ้ำๆ จนกว่าจะได้คำตอบที่เราต้องการ

3.2 แผนการทดลอง

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดประเภทที่ 1, ความน่าจะเป็นในการตรวจพบค่าผิดปรกติจริงทุกค่า, ความน่าจะเป็นซึ่งทำให้เกิดมาซดดิงเอฟเฟ็ค และความน่าจะเป็นซึ่งทำให้เกิดชวอมพิง ของตัวสถิติทดสอบทั้ง 4 ตัว คือ MV, HS, SRM และ MRM ในการวิเคราะห์ความถดถอยเชิงเส้นซึ่งผู้วิจัยมีแผนการทดลองดังนี้

3.2.1 ผู้วิจัยจะศึกษากรณีการวิเคราะห์ความถดถอยเชิงเส้น โดยใช้ตัวแบบเชิงเส้น

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_j x_{ij} + \varepsilon_i \quad ; \quad i = 1,2,\ldots,n \; ; j = 1, 2,\ldots, p-1$$

3.2.2 กำหนดจำนวนตัวแปรอิสระเป็น 1, 3 และ 5

3.2.3 กำหนดขนาดตัวอย่าง (sample size) เป็น 20, 50 และ 100

3.2.4 กำหนดจำนวนค่าผิดปรกติที่ศึกษาเป็น 1, 2 และ 3 ค่า

3.2.5 ผู้วิจัยจะศึกษากรณีความคลาดเคลื่อนมีการแจกแจง 2 ลักษณะ คือ

ก. การแจกแจงหางยาวกว่าการแจกแจงปรกติ มี 3 ลักษณะ คือ

1. การแจกแจงปรกติปลอมปนในตำแหน่ง(location contaminated normal distibution)

เมื่อความคลาดเคลื่อนมีการแจกแจงปรกติปลอมปนในตำแหน่งผู้วิจัยจะศึกษากรณีที่ $\mu = 0, \sigma^2 = 1$ มีพารามิเตอร์แสดงตำแหน่ง a = 3, 5 และ 10 ตามลำดับ และมีร้อยละการปลอมปนตามขนาดตัวอย่างและจำนวนค่าผิดปรกติที่ศึกษาดังตารางที่ 3.1

2. <u>การแจกแจงแบบปรกติปลอมปนในสเกล</u>(scale contaminated normal distribution)

เมื่อความคลาดเคลื่อนมีการแจกแจงปรกติปลอมปนในสเกลผู้วิจัยจะศึกษากรณีที่ $\mu = 0$, $\sigma^2 = 1$ มีพารามิเตอร์แสดงสเกล $c = 3, 5$ และ 10 ตามลำดับ และมีร้อยละการปลอมปนตามขนาดตัวอย่างและจำนวนค่าผิดปรกติที่ศึกษาดังตารางที่ 3.1

3. <u>การแจกแจงที</u> (t-distribution)

เมื่อความคลาดเคลื่อนมีการแจกแจงที ผู้วิจัยจะศึกษาเฉพาะกรณีที่ขนาดตัวอย่างเท่ากับ 20 ที่ระดับขั้นความเสรี (d.f.) เท่ากับ 18, 16 และ 14 ตามลำดับ โดยที่จำนวนตัวแปรอิสระเท่ากับ 1, 3 และ 5 ตามลำดับ เนื่องจากเมื่อขนาดตัวอย่างมากขึ้นการแจกแจงทีจะลู่เข้าสู่การแจกแจงปกติมากขึ้นทำให้ความคลาดเคลื่อนต่ำกว่าความเป็นจริง ซึ่งจะเป็นสาเหตุที่ทำให้ผลการทดลองไม่เป็นจริงได้

ตารางที่ 3.1 แสดงร้อยละการปลอมปนของการแจกแจงปรกติปลอมปนในตำแหน่งและการแจกแจงปรกติปลอมปนในสเกล

จำนวนค่าผิดปรกติ	ขนาดตัวอย่าง		
	20	50	100
1	5%	2%	1%
2	10%	4%	2%
3	15%	6%	3%

ข. การแจกแจงเบ้ ผู้วิจัยจะศึกษากรณีที่ความคลาดเคลื่อนมีการแจกแจงลอกนอร์มอล, แกมมาและไวบูลล์ โดยที่

1. <u>การแจกแจงลอกนอร์มอล</u> ผู้วิจัยจะศึกษากรณีที่ $\mu = 0$, $\sigma^2 = 0.1, 0.2$ และ 0.7 ตามลำดับ

2. <u>การแจกแจงแกมมาและไวบูลล์</u> ผู้วิจัยจะศึกษากรณีที่ $\beta = 1$, $\alpha = 1, 2, 3$ และ 10 ตามลำดับ

3.3 ขั้นตอนการวิจัย

ผู้วิจัยได้แบ่งขั้นตอนการวิจัยออกเป็น 4 ขั้นตอนหลักๆดังนี้

1. การสร้างโปรแกรมย่อยสำหรับการแจกแจงของความคลาดเคลื่อน ตามที่เรากำหนดในแผนการทดลอง
2. การสร้างข้อมูล (x,y) ให้มีความสัมพันธ์เชิงเส้นตรง
3. การสร้างค่าผิดปรกติ
4. การคำนวณค่าสถิติของทั้ง 4 วิธี คือ วิธี HS, MV, MRM และ SRM
5. การหาค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1, ความน่าจะเป็นที่จะพบค่าผิดปรกติจริงที่กำหนดทุกค่า(p_1), ความน่าจะเป็นซึ่งทำให้เกิดมาชดคิงเอฟเฟ็ค(p_2) และ ความน่าจะเป็นซึ่งทำให้เกิดชวอมพิงเอฟเฟ็ค(p_3)

แต่ละขั้นตอนมีรายละเอียดดังนี้

3.3.1 <u>การสร้างโปรแกรมย่อยสำหรับการแจกแจงของความคลาดเคลื่อน</u>

ก. <u>การแจกแจงแบบปรกติปลอมปนในตำแหน่ง</u> เราสร้างจากการแจกแจงปรกติมาตรฐานปลอมปนด้วยการแจกแจงปรกติที่มีค่าเฉลี่ยไม่เท่ากับ 0 โดยมีขั้นตอนดังนี้

<u>ขั้นที่ 1</u> เราสร้างการแจกแจงปรกติ โดยใช้วิธีของบ็อกซ์และมูลเลอร์(1958) ซึ่งเป็นวิธีสร้างการแจกแจงปรกติที่มีค่าเฉลี่ยเท่ากับ μ และความแปรปรวนเท่ากับ σ^2 เราสร้างไว้ในฟังก์ชันชื่อ NORMAL(DMEAN,SIGMA) เมื่อ DMEAN และ $SIGMA^2$ คือ ค่าเฉลี่ยและความแปรปรวนที่เรากำหนด รายละเอียดแสดงในภาคผนวก ก.

<u>ขั้นที่ 2</u> เราสร้างการแจกแจงปรกติปลอมปนในตำแหน่งโดยส่วนหนึ่งเรียกมาจากฟังก์ชัน NORMAL(DMEAN, SIGMA) ซึ่งแจกแจงปรกติมาตรฐานด้วยความน่าจะเป็น 1 - P และอีกส่วนหนึ่งมาจาก NORMAL(DMEANC, SIGMA) ด้วยความน่าจะเป็น P โดยที่ P เป็นร้อยละของการปลอมปน รายละเอียดแสดงในภาคผนวก ก.

ข. <u>การแจกแจงปรกติปลอมปนในสเกล</u> เราสร้างจากการแจกแจงปรกติมาตรฐาน โดยเรียกจากฟังก์ชัน NORMAL(DMEAN, SIGMA) ด้วยความน่าจะเป็น (1 - P) และเรียกจากฟังก์ชัน NORMAL(DMEAN, SIGMC) ด้วยความน่าจะเป็น P รายละเอียดในภาคผนวก ก.

ค. <u>การแจกแจงที</u> เราสร้างไว้ในฟังก์ชันชื่อ TDIST(NDF,DMEAN,SIGMA) โดยสร้างมาจากสัดส่วนของการแจกแจงปรกติและการแจกแจงไคกำลังสอง ด้วยระดับขั้นความเสรี NDF ค่าเฉลี่ยและความแปรปรวนของการแจกแจงปรกติเป็น DMEAN และ $SIGMA^2$ ตามลำดับ รายละเอียดในภาคผนวก ก.

ง. <u>การแจกแจงลอกนอร์มอล</u> เราสร้างจากการหา Exponential ของตัวเลขสุ่มที่มาจาก NORMAL(DMEAN,SIGMA) เมื่อ DMEAN และ $SIGMA^2$ เป็นค่าเฉลี่ยและความแปรปรวนของการแจกแจงปรกติ รายละเอียดในภาคผนวก ก.

จ. <u>การแจกแจงแกมมา</u> เราสร้างไว้ในฟังก์ชันชื่อ GAMA(ALPHA1,BETA1) ด้วยวิธี Reproductive property โดยที่ ALPHA1 เป็นพารามิเตอร์แสดงรูปร่าง BETA1 เป็นพารามิเตอร์แสดงสเกล รายละเอียดแสดงในภาคผนวก ก.

ฉ. <u>การแจกแจงไวบูลล์</u> เราสร้างไว้ในฟังก์ชันชื่อ WEI(ALPHA1,BETA1) โดยที่ ALPHA1 และ BETA1 เป็นพารามิเตอร์แสดงรูปร่างและสเกลตามลำดับและใช้วิธีการแปลงผกผันในการสร้างฟังก์ชัน รายละเอียดแสดงในภาคผนวก ก.

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

ในการสร้างข้อมูล (x,y) ให้มีความสัมพันธ์เชิงเส้นตรง เราจะแบ่งออกเป็น 2 กรณี คือ

<u>กรณีที่ 1</u> เมื่อความคลาดเคลื่อนมีการแจกแจงหางยาวกว่าการแจกแจงปรกติเราจะสร้างตัวแปรอิสระ x ซึ่งเป็นค่าคงที่ก่อนแล้วจึงสร้างตัวแปรตาม y ให้มีลักษณะการแจกแจงของความคลาดเคลื่อนที่เรากำหนด และมีความสัมพันธ์เชิงเส้นกับตัวแปรอิสระตามรูปแบบของความ

สัมพันธ์ คือ $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ เมื่อ $\underset{\sim}{\beta} = (10, 1, 1, \ldots, 1)'$ เป็นพารามิเตอร์ที่กำหนดขึ้นมา และ $\underset{\sim}{\varepsilon}$ เป็นความคลาดเคลื่อนที่มีการแจกแจงเป็นแบบหางยาวกว่าการแจกแจงปรกติที่ต้องการศึกษา

<u>กรณีที่ 2</u> เมื่อความคลาดเคลื่อนมีการแจกแจงเบ้ เราจะสร้างตัวแปรอิสระ X ซึ่งเป็นค่าคงที่ก่อน จากนั้นเราจะสร้าง y ให้มีการแจกแจงเบ้ แล้วแปลง y ให้มีการแจกแจงปรกติและมีความสัมพันธ์เชิงเส้นกับ X โดยจะทำการแปลงดังนี้

ก. การแจกแจงลอกนอร์มอล เราจะแปลงข้อมูลโดย $y' = \ln(y)$ เมื่อ y มีการแจกแจงลอกนอร์มอล ซึ่งจะทำให้ได้ y' มีการแจกแจงปรกติ

ข. การแจกแจงแกมมาและไวบูลล์ เราจะแปลงข้อมูลด้วยวิธีของฟาเบียน เยอร์นันเดซและริชาร์ด เอ จอห์นสัน(1980) ซึ่งจะทำให้ข้อมูลมีการแจกแจงปรกติโดยประมาณ และเราจะทำการตรวจสอบการแจกแจงปรกติด้วยตัวสถิติ W ของชาปิโรและไวลก์ (Shapiro and Wilk, 1965) โดยใช้ขั้นตอนวิธีของ เจ พี รอยส์ตัน(J. P. Royston, 1982) ซึ่งได้พัฒนาให้ตัวสถิติ W สามารถใช้ได้กับข้อมูลขนาดไม่เกิน 2000 (รายละเอียดแสดงในภาคผนวก ค.)

3.3.3 <u>การสร้างค่าผิดปรกติ</u>

การสร้างค่าผิดปรกติเราจะแบ่งเป็น 2 ลักษณะตามรูปแบบการแจกแจง คือ

1. <u>การแจกแจงแบบลอกนอร์มอล, แกมมา, ไวบูลล์ และ ที</u> ค่าผิดปรกติของการแจกแจงเหล่านี้เราจะสร้างไว้ในลำดับที่ n เมื่อกำหนดจำนวนค่าผิดปรกติ 1 ค่า ลำดับที่ n-1 และ n เมื่อกำหนดจำนวนค่าผิดปรกติ 2 ค่า และลำดับที่ n-2, n-1 และ n เมื่อกำหนดจำนวนค่าผิดปรกติ 3 ค่า โดยที่ค่าผิดปรกติจะสร้างจาก

$$y_i = \text{mean} + (a \times sd)$$

เมื่อ mean และ sd^2 คือค่าเฉลี่ยและความแปรปรวนของการแจกแจงลอกนอร์มอล, แกมมา, วบูลล์ และที (ดูรายละเอียดในบทที่ 1) ค่า a คือค่าที่ทำให้พื้นที่สะสมของการแจกแจงจาก 0 ถึง a เท่ากับ 99.99% ของพื้นที่ทั้งหมด

2. <u>การแจกแจงปรกติปลอมปนในตำแหน่ง</u> เราใช้ฟังก์ชันการสร้างค่าผิดปรกติ

$$F(x) = (n-k)\ N(0,1) + k\ N(a,1)$$

เมื่อ n คือ ขนาดตัวอย่าง

k คือ จำนวนข้อมูลผิดปรกติที่กำหนด

และ a คือ พารามิเตอร์แสดงตำแหน่งที่ทำให้เกิดค่าผิดปรกติ

3. <u>การแจกแจงปรกติปลอมปนในสเกล</u> เราใช้ฟังก์ชันการสร้างค่าผิดปรกติ

$$F(x) = (n-k)\ N(0,1) + k\ N(0,c)$$

เมื่อ n คือ ขนาดตัวอย่าง

k คือ จำนวนข้อมูลผิดปรกติที่กำหนด

และ c คือ พารามิเตอร์แสดงสเกลที่ทำให้เกิดค่าผิดปรกติ

3.3.3 <u>การคำนวณค่าสถิติทั้ง 4 วิธี</u>

เมื่อได้ตัวแปรอิสระ X และตัวแปรตาม y มาแล้วจากขั้นตอนที่ 3.3.2 เราจะคำนวณหาค่าสถิติของทั้ง 4 วิธี โดยมีโปรแกรมย่อยที่ต้องใช้ และขั้นตอนการหาดังนี้

ก. การสร้างโปรแกรมย่อยเพื่อประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีกำลังสองน้อยสุด โดยสร้างไว้ในโปรแกรมย่อยชื่อ SBHAT(NN, LL, SX, SY, SB, AA) ซึ่งตัวแปรที่ใช้ในโปรแกรมย่อย หมายถึง จำนวนค่าสังเกต, จำนวนตัวแปรอิสระ, ตัวแปรอิสระ, ตัวแปรตาม, เวกเตอร์ของสัมประสิทธิ์การถดถอย และเมทริกซ์ผกผัน ตามลำดับ โดยมีลำดับการทำงานคือ

1. คำนวณค่า $X'y$ และ $X'X$

2. คำนวณค่า $(X'X)^{-1}$ โดยเรียกโปรแกรมย่อย INVS(LL, AA) ซึ่งเป็นโปรแกรมย่อยที่ใช้สำหรับหาเมทริกซ์ผกผัน (Inverse matrax) โดยที่ AA คือเมทริกซ์ผกผันมีขนาด LL

3. คำนวณ b โดยที่ $b = (X'X)^{-1}X'y = SB$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

I17082671

ข. สร้างโปรแกรมย่อย SBHAT(NN, LL, SX, SY, SB, BHAT) เพื่อประมาณค่าความคลาดเคลื่อนหรือคำนวณหาค่าตกค้าง(residual) โดยที่ EHAT คือค่าตกค้างที่เราต้องการ

ค. สร้างโปรแกรมย่อย SDH(NN, LL, SX, AA, DH) เพื่อหาค่า h_{ii} ซึ่งเป็นสมาชิกในเส้นทแยงมุมของเมทริกซ์ฉายา $H = X'(X'X)^{-1}X$ โดยที่ DH คือค่า h_{ii} ที่เราต้องการ

ง. สร้างโปรแกรมย่อย RANK(NN, RR, SY, SX, OO) เพื่อใช้ในการจัดเรียงค่าความคลาดเคลื่อนต่างๆ โดยที่ตัวแปรตาม, ตัวแปรอิสระ และลำดับของค่าสังเกต (OO) ก็จะถูกจัดเรียงไปพร้อมๆกันตามค่าตกค้าง(RR) ทั้งนี้เพื่อความสะดวกในการพิจารณาจัดเซตของค่าสังเกตที่ทำให้เกิดค่าตกค้าง

จ. การคำนวณค่าสถิติวิธีเวียนเกิดโดยลำดับและวิธีเวียนเกิดดัดแปร

1. ในขั้นตอนแรกเราจะหาค่าตกค้างที่ปรับแล้ว(adjusted residual) $A_i = \frac{e_i}{\sqrt{1-h_{ii}}}$; $i = 1,2,\ldots,n$ โดยเรียกใช้โปรแกรมย่อย SBHAT, SEHAT และ SDH ตามลำดับ

2. จัดเรียงค่าตกค้างที่ปรับแล้วโดยเรียกโปรแกรม RANK

3. ให้ p ค่าสังเกตแรกในข้อ 2. เป็นเซตย่อยพื้นฐาน และเรียกโปรแกรมย่อย SBHAT(L1, L1, X, Y, BHAT, AA) เพื่อนำไปคำนวณหาค่าตกค้างเวียนเกิด(recursive residual (W_i))

4. เรียกโปรแกรมย่อย WI (N, L1, X, Y, SSSE, RAB, BHAT, AA, WS) เพื่อคำนวณค่า W_i เมื่อ RAB, SSSE และ WS คือ ค่าตกค้าง, ผลบวกค่าตกค้างกำลังสอง และ ค่าสถิติ W_i ตามลำดับ ซึ่งโปรแกรมย่อย WI มีรายละเอียดดังนี้

4.1 เราจะคำนวณหาผลบวกค่าตกค้างกำลังสองเมื่อตัดค่าสังเกตที่ i ออก โดยที่ $s_{(-i)} = \sum_{i=1}^{n} e_i^2 - A_i^2$

4.2 คำนวณค่า W_i ลำดับที่ $i = p + 1$

4.3 คำนวณค่า BHAT และ AA ใหม่จาก BHAT และ AA เดิม เพื่อนำไปคำนวณหา ค่า W_i ลำดับที่ $i = p+2$ ทำเช่นนี้ไปเรื่อยๆจนกระทั่งถึงลำดับที่ n

ฉ. คำนวณค่าสถิติของวิธีเมอร์วิน จี มาราชิง ในขั้นแรกเราจะเรียกโปรแกรมย่อย SBHAT, SEHAT และ SDH ตามลำดับเพื่อสร้างค่าตกค้างที่ปรับแล้ว (A_i) จากนั้นเรียกโปรแกรมย่อย RANK เพื่อจัดเรียงค่า $|A_i|$ และเก็บค่า $\max|A_i|$ ไว้ในเซต K แล้วหาผลบวกค่าตกค้างกำลังสองของรอบแรก รอบต่อไปก็จะทำในลักษณะเดียวกันกับรอบแรก แต่ขนาดตัวอย่างจะลดลงทีละ 1 จากการตัดค่าสังเกตที่มีค่า $\max|A_i|$ ในรอบแรกออกไป จนกระทั่งได้เซตย่อย K มีขนาด k^* จึงนำมาคำนวณค่าสถิติ F_k^* ต่อไป

ช. คำนวณค่าสถิติของวิธีฮาไดและไชมันนอฟฟ์ ในขั้นแรกเราทำการคำนวณหาเซตของข้อมูลสะอาด (M) ที่มีขนาด $h = (n + p - 1)/2$ โดยคำนวณหาค่าตกค้างที่ปรับแล้ว (A_i) จัดเรียงค่าสังเกตตาม $|A_i|$ จากน้อยไปมาก เช่นเดียวกับวิธีในข้อ จ. และ ฉ. แล้วแบ่งข้อมูลออกเป็น 2 เซตย่อย เซตแรกมีขนาด $s = p + 1$ เซตที่สองขนาด $n - p - 1$ นำเซตย่อยแรกไปคำนวณหาค่าตกค้าง $|B_i|$ แล้วจัดเรียงค่า $|B_i|$ เพื่อจัดเซตย่อยทั้งสองใหม่โดยที่เซตแรกก็จะมีขนาดเพิ่มขึ้นทีละ 1 และเซตที่ 2 จะมีขนาดลดลงทีละ 1 และเราจะทำซ้ำในลักษณะนี้อีกจนกระทั่งเซตย่อยแรกมีขนาดเท่ากับ h จากนั้นนำเซตย่อยแรกที่ได้ใหม่ไปคำนวณหาค่า $|M_i|$ ก่อนที่จะนำไปจัดเรียงเพื่อเปรียบเทียบค่ากับค่าวิกฤติต่อไป ดูรายละเอียดในบทที่ 2

3.3.4 <u>การคำนวณค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1, ความน่าจะเป็นที่จะตรวจพบค่าผิดปรกติที่กำหนดทุกค่า (p_1), ความน่าจะเป็นซึ่งทำให้เกิดมาซคิงเอฟเฟ็ค (p_2) และความน่าจะเป็นซึ่งทำให้เกิดซาอมพิงเอฟเฟ็ค (p_3)</u>

ความน่าจะเป็นของความผิดพลาดประเภทที่ 1, p_1, p_2 และ p_3 ผู้วิจัยคำนวณดังนี้

ก. <u>การคำนวณค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1</u>

เมื่อคำนวณได้ค่าสถิติทั้ง 4 วิธีแล้วเราจะทำการเปรียบเทียบค่าสถิติที่ได้กับค่าวิกฤติ ถ้าปฏิเสธสมมติฐานว่างเราจะนับจำนวนครั้งของการปฏิเสธจนครบจำนวนรอบ (500 รอบ) แล้วนำไปคำนวณค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1 โดยคำนวณจากสัดส่วนของจำนวนครั้งของการปฏิเสธสมมติฐานว่างกับจำนวนรอบทั้งหมด

ข. <u>การคำนวณค่าความน่าจะเป็น p_1, p_2 และ p_3</u>

เมื่อทำการเปรียบเทียบค่าสถิติของตัวสถิติทดสอบทั้ง 4 ตัวกับค่าวิกฤติแล้ว ปฏิเสธสมมติฐานว่าง เราจะตรวจเช็คค่าที่ถูกปฏิเสธว่า เป็นค่าผิดปรกติจริงที่เรากำหนด หรือเป็นค่าปรกติที่เมื่อใช้ตัวสถิตินั้นๆแล้วทำให้ตรวจพบว่าเป็นค่าผิดปรกติไป โดยตรวจเช็คจากลำดับของค่าสังเกตที่เรากำหนดไว้ในขั้นแรกของแต่ละรอบ กล่าวคือ ลำดับที่ 1 ถึง nk คือค่าปรกติ และลำดับที่ nk + 1 ถึง n คือค่าผิดปรกติที่เรากำหนด เมื่อ nk คือ ผลต่างของขนาดตัวอย่างและจำนวนค่าผิดปรกติที่ต้องการศึกษา (n - k) โดยเราจะทำการพิจารณาค่า p_1, p_2 และ p_3 ดังนี้

1. ถ้าพบว่าค่าที่ถูกปฏิเสธเป็นค่าผิดปรกติจริงที่เรากำหนดทุกค่า เราจะนับจำนวนครั้งของการปฏิเสธสมมติฐานว่าง เพื่อนำไปคำนวณค่าความน่าจะเป็นที่จะตรวจพบค่าผิดปรกติจริงที่กำหนดทุกค่า p_1

2. ถ้าค่าที่ถูกปฏิเสธเป็นค่าผิดปรกติจริงอย่างน้อย 1 ค่า เราจะนับจำนวนครั้งของการปฏิเสธสมมติฐานว่าง เพื่อนำไปคำนวณค่าความน่าจะเป็นซึ่งทำให้เกิดมาซดดิงเอฟเฟ็ค (p_2) ซึ่งจะได้เท่ากับ 1 - p_0 เมื่อ p_0 คือความน่าจะเป็นที่จะตรวจพบค่าผิดปรกติจริงอย่างน้อย 1 ค่า

3. ถ้าค่าที่ถูกปฏิเสธเป็นค่าปรกติที่ถูกตรวจพบว่าเป็นค่าผิดปรกติ เราจะนับจำนวนครั้งของการปฏิเสธสมมติฐานว่าง เพื่อนำไปคำนวณความน่าจะเป็นซึ่งทำให้เกิดซวอมพิงเอฟเฟ็ค (p_3)

สรุปหลักการนับจำนวนครั้งของการปฏิเสธสมมติฐานสำหรับการหาค่า p_1, p_2 และ p_3 แสดงในตารางที่ 3.3.4.1

ตารางที่ 3.3.4.1 แสดงสรุปหลักการหาค่า p_1, p_2 และ p_3

เงื่อนไขการนับ	ความน่าจะเป็นที่ต้องการหา		
	p_1	p_0	p_3
IR1 = 0 และ IR3 > 0			√
IR1 = k และ IR3 = 0	√	√	
IR1 = k และ IR3 > 0		√	√
0 < IR1 < k และ IR3 = 0		√	
0 < IR1 < k และ IR3 > 0		√	√

เมื่อ IR1 คือ จำนวนค่าผิดปรกติจริงที่พบใน 1 รอบ

IR3 คือ จำนวนค่าปรกติที่ถูกตรวจพบว่าเป็นค่าผิดปรกติที่พบใน 1 รอบ

k คือ จำนวนค่าผิดปรกติจริงที่เรากำหนด

p_0 คือ ความน่าจะเป็นที่จะตรวจพบค่าผิดปรกติจริงอย่างน้อย 1 ค่า

และ $p_2 = 1 - p_0$

3.4 ผังงานและลักษณะการทำงานของโปรแกรม

ในการวิจัยครั้งนี้ผู้วิจัยจะใช้โปรแกรมและโปรแกรมย่อยต่างๆ ซึ่งผู้วิจัยจะแสดงชื่อโปรแกรม ลักษณะการทำงาน และโปรแกรมย่อยที่เรียกใช้ในตารางที่ 3.4.1 ส่วนผังงานซึ่งใช้อธิบายขั้นตอนการทำงานของโปรแกรมด้วยแผนผัง แสดงในรูปที่ 3.4.1 และมีรายละเอียดการทำงานของโปรแกรมตามลำดับดังนี้

1. เริ่มต้นการทำงาน
2. กำหนดค่าพารามิเตอร์ต่างๆที่ใช้ในการวิจัย เช่น ค่าเฉลี่ย, ความแปรปรวน, พารามิเตอร์แสดงรูปร่าง, พารามิเตอร์แสดงสเกล และอื่นๆ
3. กำหนดจำนวนตัวแปรอิสระเป็น 1, 3 และ 5 ตามลำดับ
4. กำหนดขนาดตัวอย่างที่ต้องการศึกษาเป็น 20, 50 และ 100 ของทุกระดับจำนวนตัวแปรอิสระ
5. สร้างตัวแปรอิสระ X

6. กำหนดการแจกแจงของความคลาดเคลื่อนที่ต้องการศึกษาของทุกขนาดตัวอย่าง และทุกระดับจำนวนตัวแปรอิสระ

7. กำหนดจำนวนค่าผิดปรกติเป็น 0, 1, 2 และ 3 ค่า เมื่อครบทั้ง 4 ค่าโปรแกรมจึงจะเปลี่ยนการแจกแจง ขนาดตัวอย่าง และจำนวนตัวแปรอิสระ ตามลำดับ

8. ตรวจสอบจำนวนรอบของการทดลองว่าครบ 500 รอบในแต่ละสถานการณ์หรือไม่ ถ้ายังไม่ครบโปรแกรมก็จะดำเนินการทดลองในขั้นต่อไปคือในข้อที่ 9. แต่ถ้าครบก็จะตรวจเช็คว่า จำนวนค่าผิดปรกติที่กำหนดเท่ากับ 0 หรือไม่ ($k = 0$)

ก. ถ้าจำนวนค่าผิดปรกติที่เรากำหนดเท่ากับ 0 โปรแกรมจะทำการคำนวณความน่าจะเป็นของความผิดพลาดประเภทที่ 1

ข. ถ้าจำนวนค่าผิดปรกติที่เรากำหนดไม่เท่ากับ 0 โปรแกรมจะทำการคำนวณค่า p_1 , p_2 และ p_3

ค. พิมพ์ผลลัพธ์ของการทดลอง เสร็จแล้วกลับไปที่ ข้อ 7. เพื่อตรวจเช็คว่าครบทุกจำนวนค่าผิดปรกติหรือไม่ ถ้าครบแล้วก็โปรแกรมจะตรวจเช็คว่าครบทุกการแจกแจงของความคลาดเคลื่อน, ขนาดตัวอย่าง และจำนวนตัวแปรอิสระหรือไม่ ในข้อที่ 6. ข้อที่ 4. และข้อที่ 3. ตามลำดับจนจบการทำงาน

9. คำนวนค่าสถิติทั้ง 4 วิธี คือ SRM, MRM, MV และ HS ตามลำดับ

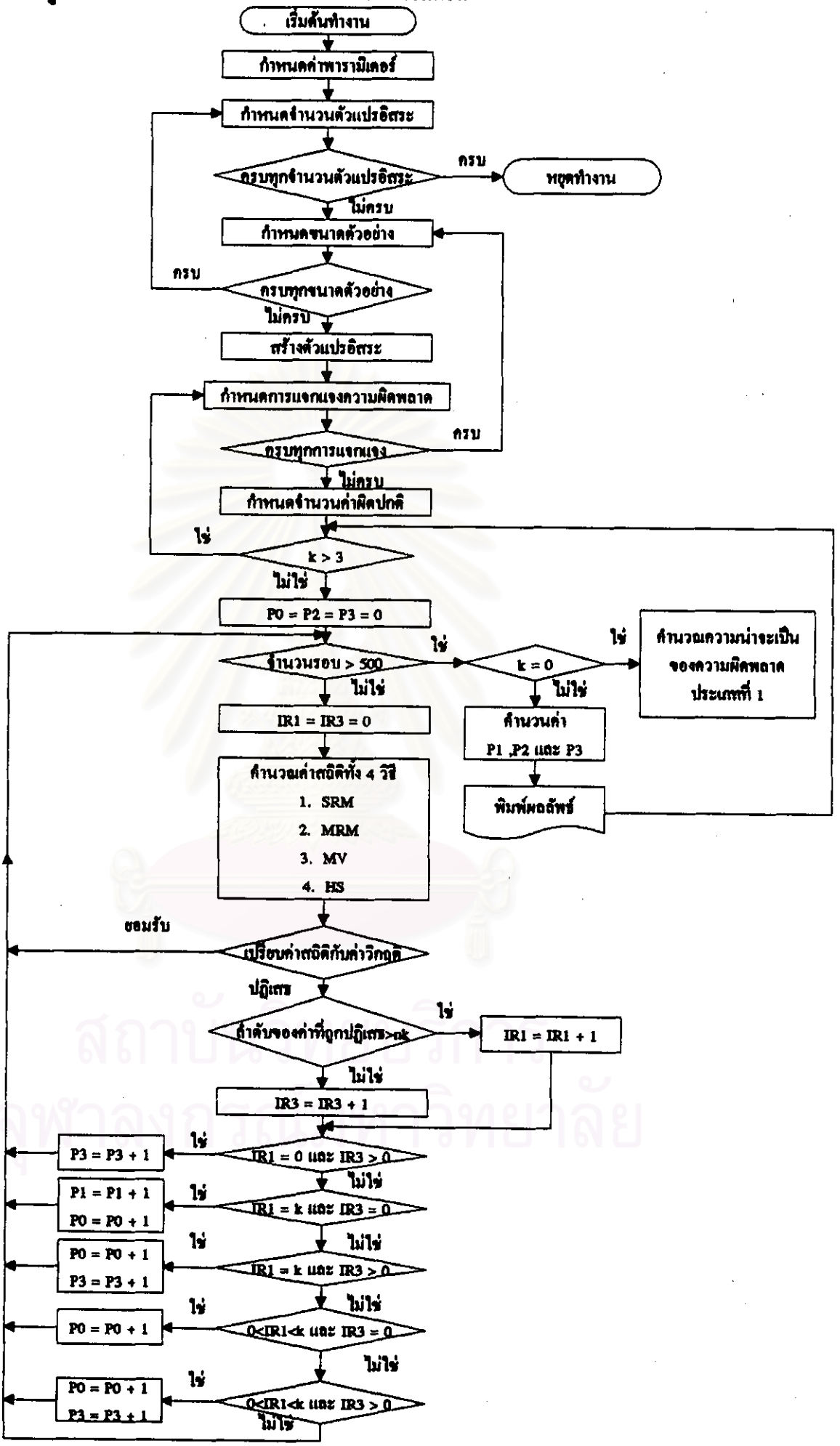
10. เปรียบเทียบค่าสถิติของแต่ละวิธีที่คำนวณได้กับค่าวิกฤติเพื่อทำการทดสอบสมมติฐานว่าง H : ไม่มีค่าผิดปรกติ ซึ่งจะได้ผล 2 ลักษณะคือ

ก. ถ้ายอมรับสมมติฐานว่าง โปรแกรมจะกลับไปทำที่ข้อ 8.

ข. ถ้าปฏิเสธสมมติฐานว่าง โปรแกรมจะทำการตรวจเช็คค่าที่ถูกปฏิเสธว่าเป็นค่าผิดปรกติจริงที่เรากำหนดไว้หรือไม่ แล้วนับจำนวนไว้เพื่อนำไปหาค่า p_1, p_2 และ p_3 ต่อไป

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รูปที่ 8.4.1 แสดงแผนผังการทำงานของโปรแกรม



ตารางที่ 3.4.1 แสดงลักษณะการทำงานของโปรแกรม

ชื่อโปรแกรม	ลักษณะการทำงาน	โปรแกรมย่อยและฟังก์ชันที่เรียกใช้
RAND*	สร้างเลขสุ่มที่มีการแจกแจงสม่ำเสมอในช่วง(0,1)	—
NORMAL*	สร้างเลขสุ่มที่มีการแจกแจงปรกติ $N(\mu,\sigma^2)$	RAND
TDIST*	สร้างเลขสุ่มที่มีการแจกแจงที	NORMAL
GAMA*	สร้างเลขสุ่มที่มีการแจกแจงแกมมา	RAND
WEI	สร้างเลขสุ่มที่มีการแจกแจงไวบูลล์	RAND
RANK	เรียงลำดับค่าของข้อมูลจากน้อยไปมากพร้อมกันหลายตัวแปร	—
INVS*	หาค่าเมทริกซ์ผกผัน	—
SBHAT	ประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นด้วยวิธีกำลังสองน้อยสุด	—
SDH	คำนวณค่าสมาชิกในเส้นทแยงมุมของเมทริกซ์ฉายา	—
SEHAT	คำนวณค่าตกค้าง	—
WI	คำนวนค่าสถิติของวิธีเวียนเกิดโดยลำดับและวิธีเวียนเกิดคัดแปร	—
PROB	คำนวณค่า p_1, p_2 และ p_3	—
MAIN	ใช้ในการคำนวณค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1, p_1, p_2 และ p_3 ของทุกวิธี และ ทุกการแจกแจง	RAND, NORMAL, SBHAT, SEHAT, TDIST, RANK, WI, SDH, WEI, PROB, GAMA

* โปรแกรมที่นำมาจาก สมชาย รัตนเลิศนุสรณ์(2533)