หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3

วิธีดำเนินการวิจัย

งานวิจัยครั้งนี้ต้องการศึกษาเปรียบเทียบตัวสถิติที่ใช้ตรวจสอบข้อมูลผิดปกติในการวิเคราะห์ความถดถอยเชิงเส้นอย่างง่าย ของวิธีการทดสอบ 3 วิธี คือ วิธีของทิตเจน, มัวร์ และเบคแมน(TMB) วิธีของเมอวีน จี มาราชิงห์(M) และวิธีของจีแบร์รี(GB)จะศึกษาในกรณีที่ความผิดพลาดมีการแจกแจงแบบเบ้ และแบบหางยาวกว่าการแจกแจงปกติ ซึ่งงานวิจัยครั้งนี้เป็นงานวิจัยเชิงทดลอง โดยใช้วิธีมอนติคาร์โล(Monte Carlo method)มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา ดังนั้นในส่วนแรกของบทนี้จะกล่าวถึงวิธีมอนติคาร์โลเพื่อเป็นพื้นฐานในการศึกษาวิจัย จากนั้นก็จะกล่าวถึงแผนการทดลอง ขั้นตอนการวิจัยแต่ละขั้นตอน ผังงานและโปรแกรมที่ใช้ในงานวิจัยตามลำดับ

3.1 <u>วิธีมอนติคาร์โล</u> (Monte Carlo Method)

วิธีมอนติคาร์โลเป็นวิธีการจำลองตัวแบบทางคณิตศาสตร์วิธีหนึ่งที่นิยมใช้กันอย่างแพร่หลายในปัจจุบัน ซึ่งหลักการของวิธีมอนติคาร์โลนั้นเป็นการจำลองตัวเลขสุ่ม (random number) มาช่วยในการค้นหาคำตอบของปัญหาที่ต้องการศึกษา สำหรับในงานวิจัยครั้งนี้จะใช้วิธีมอนติคาร์โลมาช่วยในการสร้างข้อมูลให้มีการแจกแจงต่างๆ ตามที่เราต้องการ โดยขั้นตอนของวิธีมอนติคาร์โล มีอยู่ 3 ขั้นตอน ดังนี้

3.1.1 <u>การสร้างตัวเลขสุ่ม</u> ตัวเลขสุ่มถือเป็นสิ่งที่สำคัญมากในการจำลองค่า โดยวิธีมอนติคาร์โล ทั้งนี้เพราะตัวเลขสุ่มที่สร้างขึ้นมาจะถูกนำไปใช้ในการสร้างประชากรเพื่อให้มีลักษณะการแจกแจงแบบต่าง ๆ ตามที่เราต้องการศึกษา ดังนั้นในขั้นตอนนี้ผู้วิจัยจึงต้องระมัดระวังในการสร้างตัวเลขสุ่ม โดยอาศัยหลักการที่ว่าจะต้องสร้างตัวเลขสุ่มให้มีลักษณะการแจกแจงแบบสม่ำเสมอในช่วง(0,1) ซึ่งวิธีการสร้างตัวเลขสุ่มนั้นมีผู้เสนอไว้หลายวิธี แต่วิธีหนึ่งที่ดีก็คือ วิธีของไวท์ และชมิดท์(White and Schemidt, 1975) เพราะให้ตัวเลขสุ่มที่มีลักษณะการแจกแจงสม่ำเสมอในช่วง(0,1) และตัวเลขสุ่มแต่ละค่าเป็นอิสระต่อกัน ดังรายละเอียดที่ได้แสดงไว้ในภาคผนวก ก.

3.1.2 <u>การประยุกต์ตัวเลขสุ่มให้สามารถใช้ได้กับปัญหาที่ต้องการศึกษา</u> ซึ่งขั้นตอนนี้ขึ้นอยู่กับลักษณะของปัญหา บางปัญหาอาจจะไม่ใช้ตัวเลขสุ่มโดยตรงแต่อาจจะมีขั้นตอนอื่น ๆ อีกหลายขั้นตอนซึ่งบางขั้นตอนต้องใช้ตัวเลขสุ่ม

3.1.3 <u>การทดลองซ้ำ ๆ</u> (replication) เมื่อสามารถประยุกต์ตัวเลขสุ่มให้ใช้กับปัญหาได้แล้ว ขั้นตอนต่อไปก็คือ การทดลองโดยใช้กระบวนการสุ่ม(random process) มากระทำในลักษณะที่ซ้ำ ๆ กันเพื่อหาคำตอบของปัญหาที่ต้องการศึกษา

3.2 <u>แผนการทดลอง</u>

งานวิจัยครั้งนี้ผู้วิจัยมีวัตถุประสงค์เพื่อเปรียบเทียบอำนาจการทดสอบ และความสามารถในการควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของตัวสถิติ 3 ตัว คือตัวสถิติของวิธีการของทิตเจน, มัวร์ และเบคแมน(TMB) ของวิธีการของเมอวีน(M) และของวิธีการของจีแบร์รี(GB) โดยมีแผนการทดลองดังนี้

3.2.1 ศึกษากรณีการวิเคราะห์ความถดถอยเชิงเส้นอย่างง่าย โดยใช้ตัวแบบเชิงเส้นดังนี้

$$y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i \qquad i=1, 2, 3, \ldots, n$$

3.2.2 การแจกแจงของความผิดพลาดที่ศึกษามี 2 ลักษณะดังแสดงในตารางที่ 3.1

<u>ตารางที่ 3.1</u> แสดงลักษณะการแจกแจงของความผิดพลาดและค่าพารามิเตอร์ที่ใช้ศึกษา

แบบหางยาวกว่าการแจกแจงปกติ	แบบเบ้
1. สเกลคอนทามิเนตนอร์มอล ศึกษากรณี $\mu = 0$, $\sigma^2 = 1$ $c = 3,5,10$ และ $p = 5\%, 10\%, 25\%$ 2. โลเคชันคอนทามิเนตนอร์มอล ศึกษากรณี $\mu = 0$, $\sigma^2 = 1$ $a = 3,5,15$ และ $p = 5\%, 10\%, 25\%$ 3. แบบที ศึกษากรณี $n = 20$ ณ ระดับความเป็นอิสระ = 18	1. ลอกนอร์มอล ศึกษากรณี $\mu = 0$, $\sigma^2 = 0.1, 0.3, 0.5, 0.7$ 2. แกมมา ศึกษากรณี $\alpha = 1,2,3$ และ 10 เมื่อ $\beta = 1$ 3. ไวบูลล์ ศึกษากรณี $\alpha = 1,2,3$ และ 10 เมื่อ $\beta = 1$

3.2.3 กำหนดขนาดตัวอย่าง(sample size)ให้มีขนาด 3 ระดับ คือ 20 , 50 และ 100

3.2.4 จำนวนค่าผิดปกติที่ศึกษา k = 1, 2, 3 ทุกขนาดตัวอย่างที่ศึกษา

3.3 ขั้นตอนการวิจัย

สามารถแบ่งออกเป็น 4 ขั้นตอนหลัก ๆ ดังนี้

3.3.1 การสร้างโปรแกรมย่อยสำหรับการแจกแจงของความผิดพลาด (ε) ตามที่กำหนดในแผนการทดลอง

3.3.2 การสร้างข้อมูล(x,y) ให้มีความสัมพันธ์เชิงเส้นตรง

3.3.3 การคำนวณค่าสถิติทั้ง 3 วิธี คือวิธีของทิตเจน, มัวร์ และเบคแมน วิธีของเมอวีน จี มาราชิงห์ และวิธีของจีแบร์รี

3.3.4 การหาค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบสำหรับทดสอบสมมติฐานเกี่ยวกับค่าผิดปกติ

รายละเอียดของแต่ละขั้นตอน มีดังต่อไปนี้

3.3.1 การสร้างโปรแกรมย่อยสำหรับการแจกแจงของความผิดพลาด

ในการสร้างลักษณะการแจกแจงแบบต่าง ๆ นั้น จะต้องอาศัยตัวเลขสุ่มเป็นพื้นฐานในการสร้าง และภาษาคอมพิวเตอร์ที่ใช้เขียนโปรแกรมก็คือ ภาษาฟอร์แทรน โฟ(FORTRAN IV) บนเครื่อง IBM 370/3031 ณ สถาบันบริการคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย สำหรับรายละเอียดในการสร้างการแจกแจงแบบต่าง ๆ เป็นดังนี้

ก) **การแจกแจงแบบสเกลคอนทามิเนตนอร์มอล** สร้างโดยใช้วิธีการแปลงข้อมูลที่มาจากการแจกแจงปกติ โดยมีขั้นตอนการสร้าง 2 ขั้นตอน คือ

ขั้นแรก สร้างการแจกแจงแบบปกติ ซึ่งโดยทั่วไปแล้วจะใช้วิธีของBox และ Muller(1958) ซึ่งเป็นวิธีที่สร้างการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น μ และความแปรปรวนเป็น σ^2 โดยที่ DMEAN และ $(SIGMA)^2$ คือ ค่าเฉลี่ย และความแปรปรวนที่กำหนด (รายละเอียดของโปรแกรมได้แสดงไว้ในภาคผนวก ก. สร้างไว้โดยใช้ชื่อฟังก์ชัน NORMAL(DMEAN,SIGMA))

ขั้นที่สอง สร้างการแจกแจงแบบสเกลคอนทามิเนตนอร์มอล โดยส่วนหนึ่งมาจาก NORMAL(DMEAN, SIGMA)ด้วยความน่าจะเป็น 1-p อีกส่วนหนึ่งได้จาก

NORMAL(DMEAN, SIGM2) ด้วยความน่าจะเป็น p โดยที่ DMEAN, SIGMA และ p เป็น ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และเปอร์เซนต์การปลอมปน ตามลำดับ ส่วน SIGM2 เป็นผลคูณของค่า scale factor(c)กับ SIGMA สำหรับรายละเอียดของโปรแกรมได้แสดงไว้ในภาคผนวก ก.

ข) **การแจกแจงแบบโลเคชันคอนทามิเนตนอร์มอล** ใช้หลักการสร้างทำนองเดียวกับข้อ ก) โดยส่วนหนึ่งสร้างจาก NORMAL(DMEAN, SIGMA)ด้วยความน่าจะเป็น 1-p อีกส่วนหนึ่งสร้างจาก NORMAL(DMEAN2, SIGMA)ด้วยความน่าจะเป็น p เมื่อ DMEAN2 เป็นผลบวกของค่า location factor(a)กับ DMEAN สำหรับรายละเอียดของโปรแกรมได้แสดงไว้ในภาคผนวก ก.

ค) **การแจกแจงแบบที** สร้างเป็นฟังก์ชันชื่อว่า TDIST(NDF, DMEAN, SIGMA) เมื่อต้องการเรียกใช้ฟังก์ชันนี้ทำได้โดยการเรียกชื่อฟังก์ชันเท่านั้นจะได้ผลลัพธ์คือค่าตัวแปร TDIST ที่มีระดับความเป็นอิสระ(degree of freedom)เป็น NDF ซึ่งสร้างจากสัดส่วนการแจกแจงแบบปกติ และการแจกแจงแบบไคสแควร์ที่มีระดับความเป็นอิสระเป็น NDF สำหรับรายละเอียดของโปรแกรมได้แสดงไว้ในภาคผนวก ก.

ง) **การแจกแจงแบบลอกนอร์มอล** สร้างโดยการหาค่า exponential ของฟังก์ชัน NORMAL(DMEAN, SIGM3) เมื่อ DMEAN และ SIGM3 เป็นค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของการแจกแจงปกติ ตามลำดับ สำหรับรายละเอียดได้แสดงไว้ในภาคผนวก ก.

จ) **การแจกแจงแบบแกมมา** สร้างเป็นฟังก์ชันชื่อว่า GAMMA(ALPHA1, BETA1)เมื่อต้องการเรียกใช้ฟังก์ชันนี้ทำได้โดยการเรียกชื่อฟังก์ชันเท่านั้น จะได้ผลลัพธ์คือค่าตัวแปร GAMMA ที่มีการแจกแจงแบบแกมมา โดยสามารถกำหนดค่า ALPHA1 และ BETA1 ตามที่ต้องการศึกษาเพื่อให้ได้ค่าเฉลี่ยเป็น ALPHA1*BETA1 และความแปรปรวนเป็น ALPHA1*(BETA1)^2 ตามลำดับ สำหรับรายละเอียดของโปรแกรมได้แสดงไว้ในภาคผนวก ก.

ฉ) **การแจกแจงแบบไวบูลล์** สร้างเป็นฟังก์ชันชื่อว่า WEIBUL(ALPHA1, BETA1)เมื่อต้องการเรียกใช้ฟังก์ชันนี้ทำได้โดยการเรียกชื่อฟังก์ชันเท่านั้น จะได้ผลลัพธ์คือค่าตัวแปร WEIBUL ที่มีการแจกแจงแบบไวบูลล์ โดยสามารถกำหนดค่า ALPHA1 และ BETA1 ตามที่ต้องการศึกษาเพื่อให้ได้ค่าเฉลี่ยเป็น BETA1*Γ(1+1/ALPHA1) และความแปรปรวนเป็น (BETA1)^2*[Γ(1+2/ALPHA1)- Γ^2(1+1/ALPHA1)] ตามลำดับ สำหรับรายละเอียดของโปรแกรมได้แสดงไว้ในภาคผนวก ก.

017433

3.3.2 การสร้างข้อมูล(x,y)ให้มีความสัมพันธ์เชิงเส้นตรง

สามารถแบ่งออกเป็น 2 กรณี คือ

กรณีที่ 1 เมื่อการแจกแจงของความผิดพลาดเป็นแบบ หางยาวกว่าการแจกแจงแบบปกติ

โดยขั้นแรกจะสร้างตัวแปรอิสระ X ก่อน จากนั้นจึงสร้างการแจกแจงของความผิดพลาดตามที่ต้องการศึกษา คือ แบบสเกลคอนทามิเนตนอร์มอล โลเคชันคอนทามิเนตนอร์มอล และแบบที่โดยสร้างทีละการแจกแจง จากนั้นจึงกำหนดค่าพารามิเตอร์ β ที่มีขนาด 2X1 คือ β_0 กับ β_1 แล้วแทนค่าต่าง ๆ ที่ได้ตามรูปแบบความสัมพันธ์เชิงเส้นตรงอย่างง่ายคือ

$$y = X\beta + \varepsilon$$

เมื่อ $\beta = (\beta_0 , \beta_1)'$ ในงานวิจัยครั้งนี้ใช้ค่า $\beta_0 = 10$ และ $\beta_1 = 1$ ก็จะได้ตัวแปร y มีการแจกแจงตามรูปแบบดังกล่าว

กรณีที่ 2 เมื่อการแจกแจงของความผิดพลาดเป็นแบบเบ้

ขั้นแรกจะสร้างตัวแปรอิสระ X ก่อน จากนั้นจึงสร้างตัวแปรตาม (y)ให้มีการแจกแจงแบบเบ้ตามที่ต้องการศึกษา คือ แบบลอกนอร์มอล แกมมา และไวบูลล์ แล้วอาศัยการแปลงที่อยู่ในรูปยกกำลังของ Box และ Cox เพื่อสร้างตัวแปรที่มีความสัมพันธ์เชิงเส้นกับ X ตามรูปแบบของความสัมพันธ์ คือ

$$y^{\lambda} = X\beta + \varepsilon$$

เมื่อ β เป็นค่าพารามิเตอร์ที่ถูกกำหนดขึ้นมา โดย $\beta_0 = 10$ และ $\beta_1 = 1$

λ เป็นค่าพารามิเตอร์ของการแปลงที่อยู่ในรูปยกกำลังของ Box และ Cox ซึ่งจะนำไปใช้ในการประมาณค่าสัมประสิทธิ์การถดถอย(β)(ดูรายละเอียดในภาคผนวก ค) และ ε เป็นความผิดพลาดที่มีการแจกแจงเป็นแบบสมมาตร และเป็นแบบปกติโดยประมาณ ซึ่งเราสามารถตรวจสอบการแจกแจงแบบปกติโดยใช้ตัวสถิติของ shapiro และ wilk

สำหรับการสร้างค่าผิดปกติของแต่ละการแจกแจงของความผิดพลาดที่ต้องการศึกษา มีดังนี้

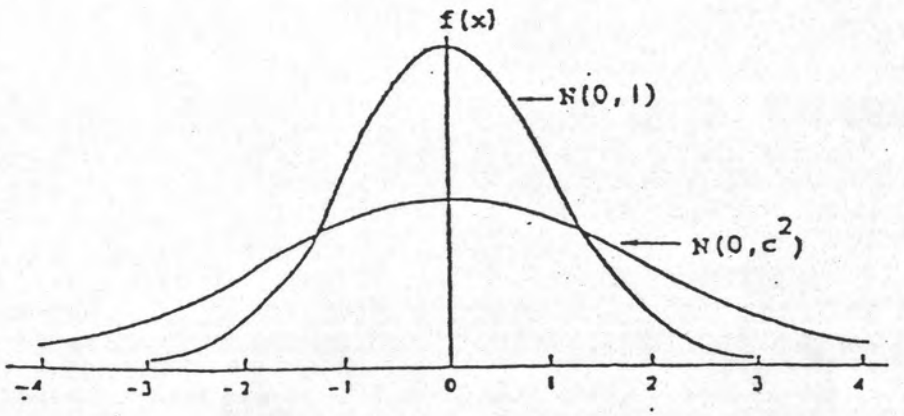
1. การแจกแจงแบบสเกลคอนทามิเนตนอร์มอล ใช้ฟังก์ชันการสร้างค่าผิดปกติ ดังนี้

$$f(x) = (N-N_1)(1-p)N(0,1) + N_1pN(0,c^2)$$

เมื่อ N เป็นจำนวนข้อมูลทั้งหมด และ N_1 เป็นจำนวนข้อมูลที่ผิดปกติ

c เป็น scale factor ที่ทำให้เกิดค่าผิดปกติ

p เป็น เปอร์เซนต์ของการปลอมปน ในงานวิจัยนี้ใช้ p = 5%, 10% และ 25%



รูปที่ 3.1 แสดงการแจกแจงแบบสเกลคอนทามิเนตนอร์มอล เมื่อ varaince $=c^2$ โดยที่ c = 3, 5, 10 ตามลำดับ

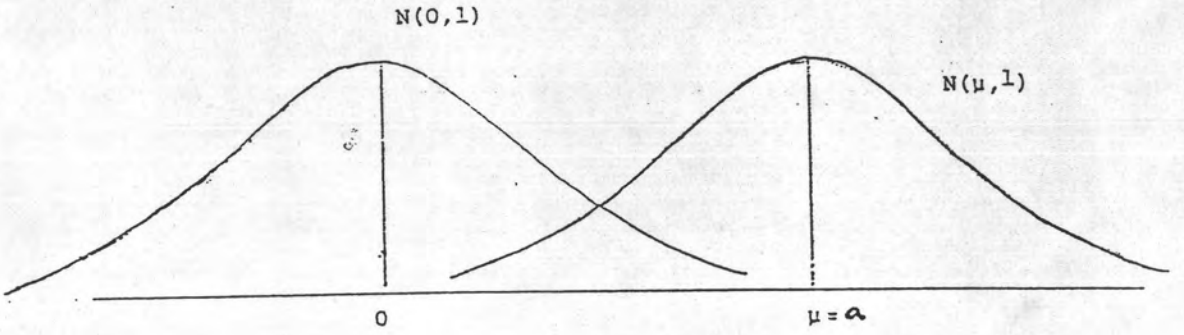
2. <u>การแจกแจงแบบโลเคชันคอนทามิเนตนอร์มอล</u> ใช้ฟังก์ชันการสร้างค่าผิดปกติ ดังนี้

$$f(x) = (N-N_1)(1-p)N(0,1) + N_1pN(a,1)$$

เมื่อ N เป็นจำนวนข้อมูลทั้งหมด และ N_1 เป็นจำนวนข้อมูลที่ผิดปกติ

a เป็น location factor ที่ทำให้เกิดค่าผิดปกติ

p เป็น เปอร์เซนต์ของการปลอมปน ในงานวิจัยนี้ใช้ p = 5%, 10% และ 25%

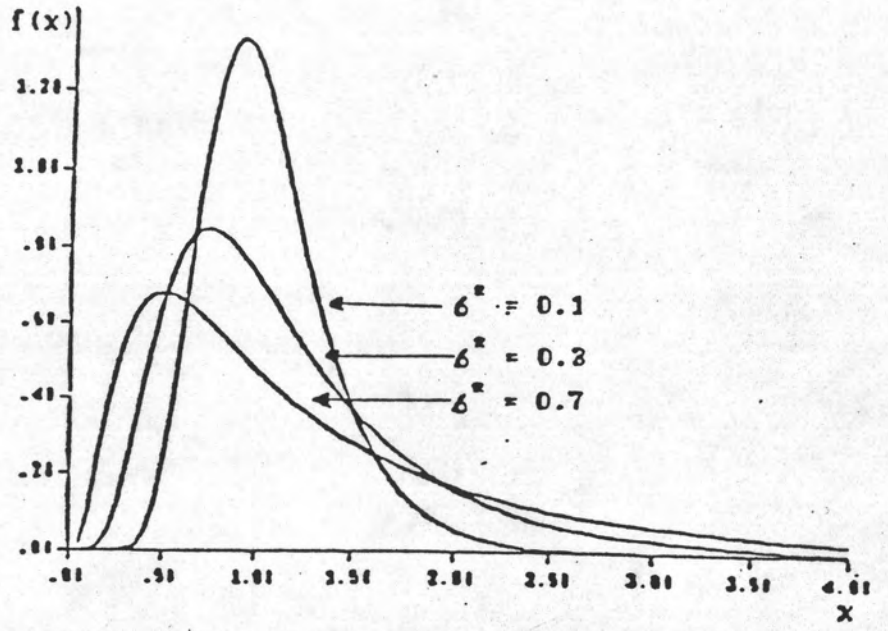


รูปที่ 3.2 แสดงการแจกแจงแบบโลเคชันคอนทามิเนตนอร์มอล เมื่อ mean(μ)=a โดยที่ a = 3, 5, 15 ตามลำดับ

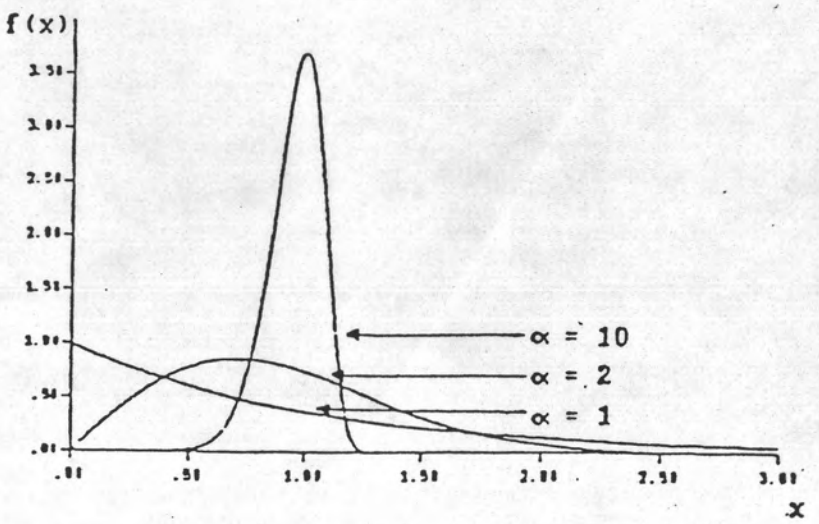
3. <u>การแจกแจงแบบที ลอกนอร์มอล แกมมา และไวบูลล์</u> ใช้เกณฑ์การสร้างค่าผิดปกติ โดยทำให้เป็นค่าสูงมากผิดปกติ(มนตรี, 2529) ดังนี้

$$O_t = \text{MEAN} + 3.5 \times \text{STD}$$

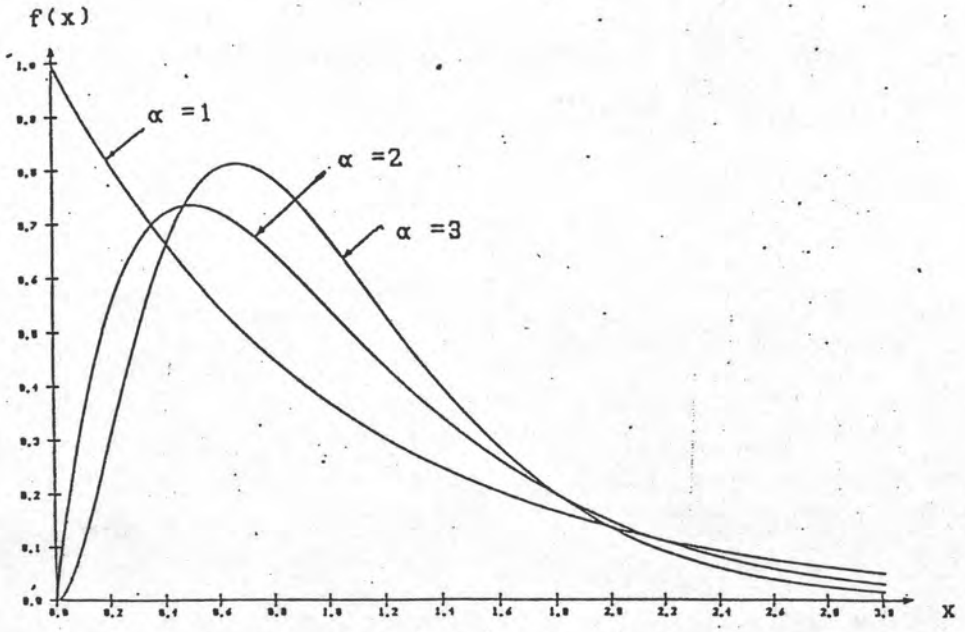
เมื่อ O_t คือค่าผิดปกติ ส่วน MEAN และ STD เป็นค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของประชากร ตามลำดับ



<u>รูปที่ 3.3</u> แสดงการแจกแจงแบบลอกนอร์มอล เมื่อ μ=0 σ^2= 0.1, 0.3, 0.7



<u>รูปที่ 3.4</u> แสดงการแจกแจงแบบไวบูลล์ เมื่อ β = 1 α = 1, 2, 10



รูปที่ 3.5 แสดงการแจกแจงแบบแกมมา เมื่อ $\beta = 1$ $\alpha = 1, 2, 3$

3.3.3 การคำนวณค่าตัวสถิติทั้ง 3 วิธี คือวิธีของทิตเจน, มัวร์ และเบคแมน(TMB) วิธีของเมอวิน จี มาราชิงห์(M) และวิธีของจีแบร์รี(GB)

หลังจากสร้างข้อมูล(x,y)ให้มีความสัมพันธ์เชิงเส้นตามรูปแบบที่ต้องการได้แล้ว ขั้นตอนต่อไปก็คือการคำนวณค่าตัวสถิติทั้ง 3 วิธี โดยเริ่มการคำนวณค่าตัวสถิติของวิธี TMB และวิธี GB ก่อน จากนั้นจึงค่อยคำนวณค่าตัวสถิติของวิธี M โดยมีขั้นตอนการคำนวณค่าตัวสถิติทั้ง 3 วิธี ดังนี้

1. การคำนวณค่าสถิติของวิธี TMB และวิธี GB ขั้นแรกทำการประมาณค่าสัมประสิทธิ์การถดถอย(β) และหาความคลาดเคลื่อน(residual) จากนั้นก็นำค่าความคลาดเคลื่อนที่ได้ไปหาค่าความคลาดเคลื่อนมาตรฐาน(studentized residual หรือบางครั้งเรียก standardized residual) ขั้นตอนต่อมาทำการคำนวณตัวสถิติของวิธี TMB โดยนำค่าความคลาดเคลื่อนมาตรฐานที่คำนวณได้ทุกค่ามาใส่ค่าสัมบูรณ์(absolute) แล้วเลือกค่าความคลาดเคลื่อนมาตรฐานที่มีค่าสูงสุด(maximum absolute studentized residual)จะได้ตัวสถิติของวิธี TMB ขั้นตอนสุดท้ายคำนวณตัวสถิติของวิธี GB ได้โดยการนำตัวสถิติของวิธี TMB มายกกำลังสอง แล้วหารด้วยผลต่างของขนาดตัวอย่าง กับจำนวนตัวแปรอิสระ (ดูรายละเอียดของตัวสถิติวิธี TMB และวิธี GB ได้ในบทที่ 2)

2. การคำนวณค่าสถิติของวิธี M ขั้นแรกทำการกำหนดจำนวนรอบสูงสุดที่จะทำการวิเคราะห์ความถดถอย (k^*) ซึ่งมีค่าตั้งแต่ 2 ถึง 5 รอบ จากนั้นก็ทำการคำนวณค่าสัมประสิทธิ์การถดถอย ค่าความคลาดเคลื่อน(residual) ค่าความคลาด

เคลื่อนที่ปรับแล้วที่มีค่าสูงสุด(maximum adjusted residual, t_i) และคำนวณค่าผลรวมกำลังสองของความคลาดเคลื่อน ในแต่ละรอบ จากนั้นก็คำนวณค่า

$$Q_{k*} = \sum_{i=1}^{k^*} (t_i)^2$$

และค่าตัวสถิติทดสอบ

$$F_{k*} = (S-Q_{k*})/ S$$

เมื่อ $S = (n-p)s^2$ โดยที่ $s^2 = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 / (n-p)$

ดังนั้น S ก็คือ ผลรวมกำลังสองของความคลาดเคลื่อน (ดูรายละเอียดของวิธีการของเมอวีน(M) และขั้นตอนการคำนวณได้ในบทที่ 2)

3.3.4 การคำนวณค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบ

เมื่อคำนวณค่าตัวสถิติทั้ง 3 วิธีเสร็จ ก็นำค่าตัวสถิติที่คำนวณได้ไปเปรียบเทียบกับค่าวิกฤตตามระดับนัยสำคัญ ซึ่งใช้ $\alpha = 0.05$ และ 0.01 เพื่อหาข้อสรุปเกี่ยวกับสมมติฐาน ดังนั้นในขั้นตอนนี้จึงต้องกล่าวถึง วิธีการนับจำนวนครั้งที่ปฏิเสธสมมติฐานว่างของการวิจัยนี้

การนับจำนวนครั้งที่ปฏิเสธสมมติฐานว่าง ในงานวิจัยนี้ทดลองเมื่อค่าผิดปกติมีจำนวน 1, 2 และ 3 ค่า ตามลำดับ ในกรณีที่มีค่าผิดปกติ 1 ค่า การนับจำนวนครั้งที่ปฏิเสธสมมติฐานว่างมีหลักการดังนี้ จะทำการทดลอง 2 ครั้งเพื่อตัดสินใจว่าจะยอมรับ หรือปฏิเสธสมมติฐานว่าง โดยที่การทดลองครั้งแรกจะต้องปฏิเสธสมมติฐานว่าง ครั้งที่สองจะต้องยอมรับสมมติฐานว่าง จึงจะสรุปได้ว่ามีค่าผิดปกติ 1 ค่า ส่วนกรณีอื่น ๆ จะยอมรับสมมติฐานว่าง กรณีที่มีค่าผิดปกติ 2 ค่า ก็มีหลักการนับจำนวนครั้งที่ปฏิเสธ H_0 ดังนี้คือ จะทำการทดลอง 3 ครั้งเพื่อตัดสินใจว่าจะยอมรับหรือปฏิเสธสมมติฐานว่าง โดยที่ครั้งที่หนึ่งและสองจะต้องปฏิเสธสมมติฐานว่าง ส่วนครั้งที่สามจะต้องยอมรับสมมติฐานว่าง จึงจะสรุปได้ว่ามีค่าผิดปกติ 2 ค่า ส่วนกรณีอื่นๆ จะยอมรับสมมติฐานว่าง กรณีค่าผิดปกติ 3 ค่า ก็ใช้หลักการนับจำนวนครั้งที่ปฏิเสธสมมติฐานว่างทำนองเดียวกันคือ ทำการทดลอง 4 ครั้งโดยที่ ครั้งแรก ครั้งที่สอง และครั้งที่สาม

จะต้องปฏิเสธสมมติฐานว่าง ส่วนครั้งที่สี่จะต้องยอมรับสมมติฐานว่าง จึงจะสรุปได้ว่ามีค่าผิดปกติ 3 ค่า ส่วนกรณีอื่น ๆ จะยอมรับสมมติฐานว่าง

สรุปหลักการนับจำนวนครั้งที่ปฏิเสธสมมติฐานว่างได้ดังแสดงไว้ในตารางที่ 3.2

ตารางที่ 3.2 แสดงสรุปหลักการนับจำนวนครั้งที่ปฏิเสธ H_0

เมื่อ H_0 : ไม่มีข้อมูลผิดปกติ

*จำนวนครั้งที่ทดลองเพื่อตัดสินใจ	ผลการทดลองครั้งที่				สรุปผลการทดลอง	นับจำนวน Reject.H_0
	1	2	3	4		
2	Reject H_0	Accept H_0	----	----	มีค่าผิดปกติ 1 ค่า	1 ครั้ง
3	Reject H_0	Reject H_0	Accept H_0	----	มีค่าผิดปกติ 2 ค่า	1 ครั้ง
4	Reject H_0	Reject H_0	Reject H_0	Accept H_0	มีค่าผิดปกติ 3 ค่า	1 ครั้ง

เมื่อนับจำนวนครั้งที่ปฏิเสธ H_0 ได้แล้ว ก็สามารถนำมาคำนวณหาค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบได้

---

* การทดลองในครั้งแรกจะต้องปฏิเสธสมมติฐานว่าง(H_0) จากนั้นจึงทำการทดลองครั้งต่อไปได้ ในกรณีที่การทดลองครั้งแรกยอมสมมติฐานว่างจะไม่ทำการทดลองในครั้งต่อไปเนื่องจากตัวสถิติทดสอบไม่สามารถตรวจสอบค่าผิดปกติได้ ดังนั้นจึงหยุดการทดสอบ

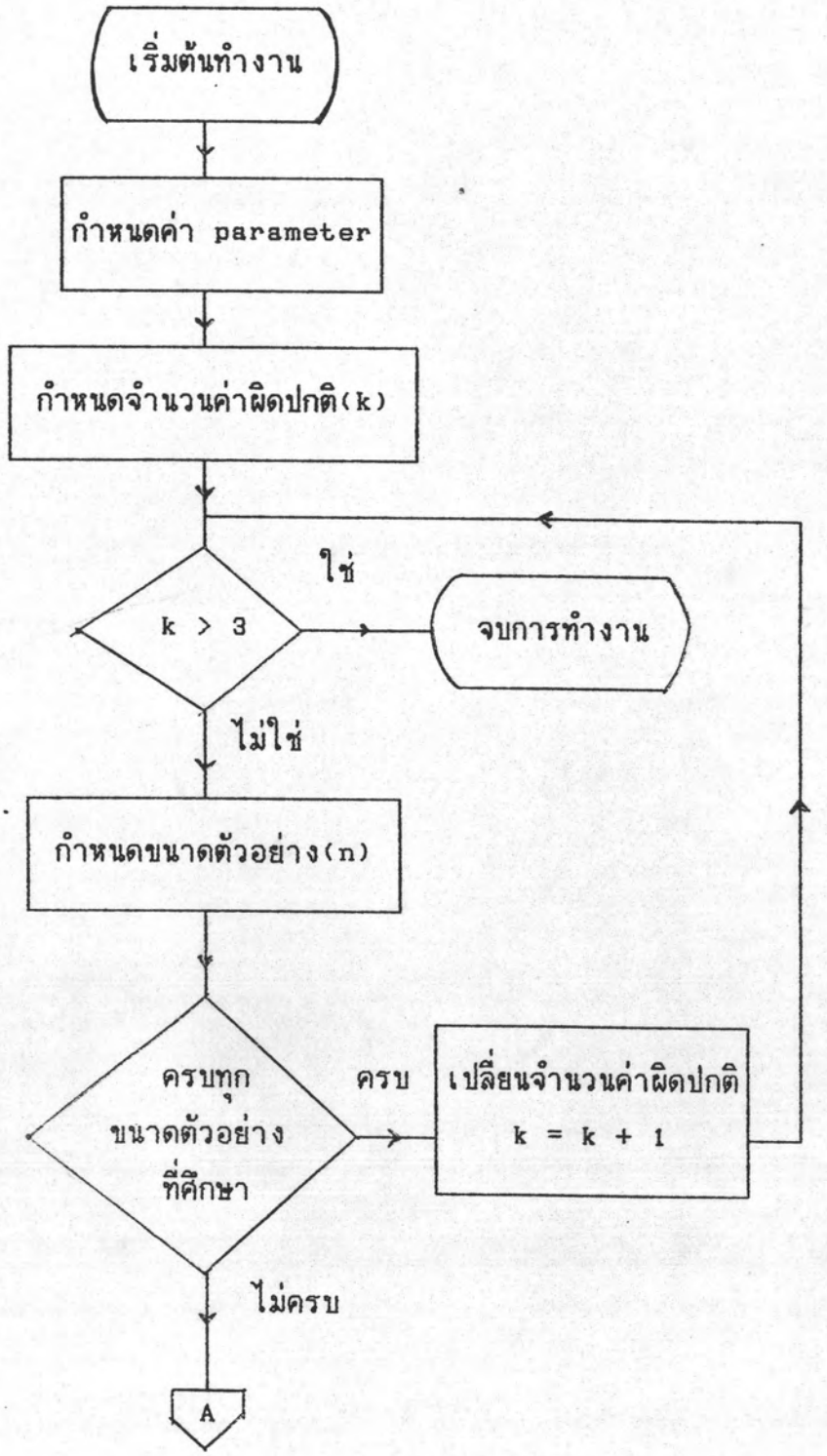
3.4 ผังงานและโปรแกรมที่ใช้ในงานวิจัย

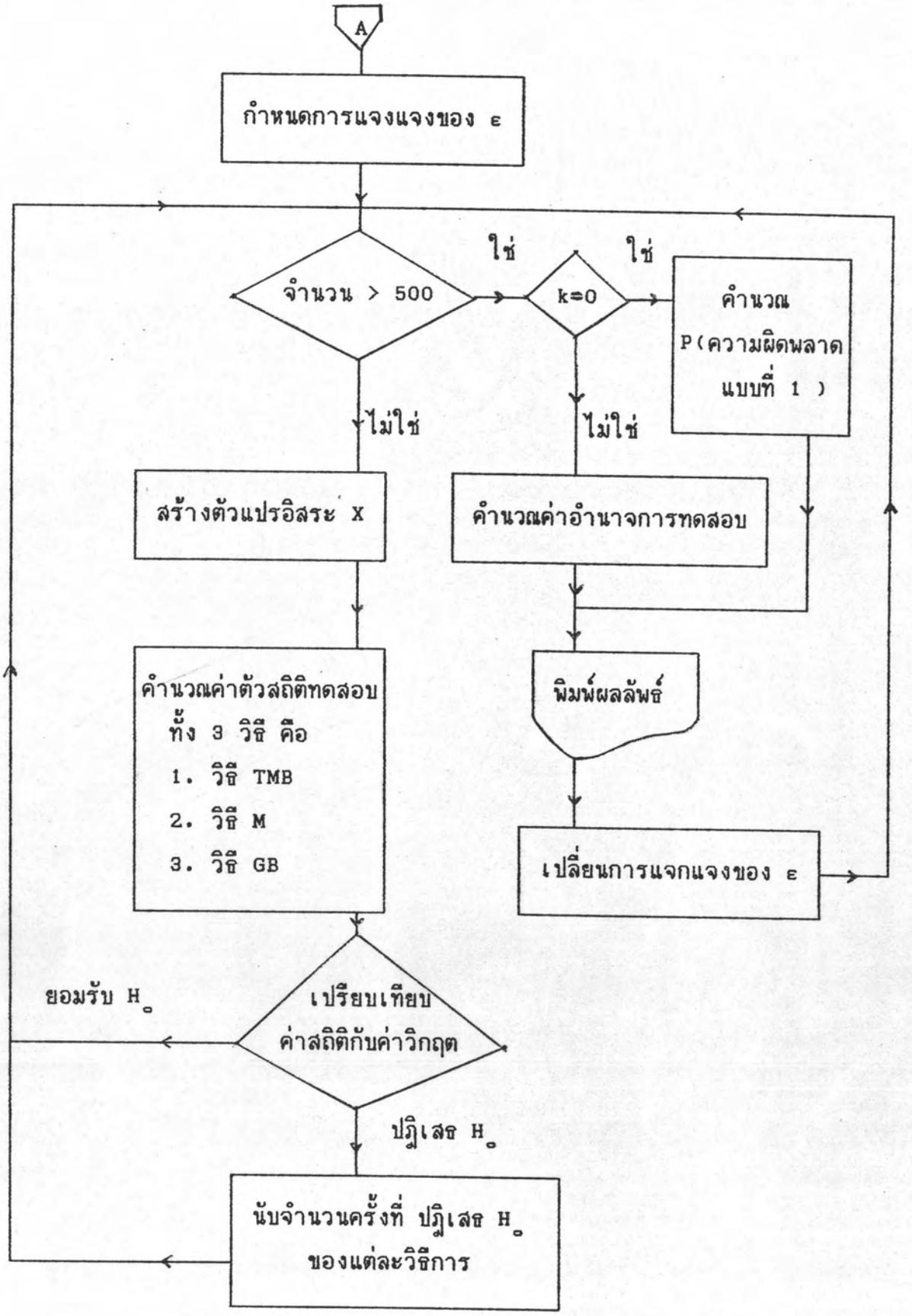
ในการเขียนโปรแกรมเพื่อแก้ปัญหาใดปัญหาหนึ่งโดยใช้เครื่องคอมพิวเตอร์ จะมีวิธีการตามขั้นตอน ดังต่อไปนี้

1. การวิเคราะห์งาน(job analization)
2. การเขียนผังงาน(flowcharting)
3. การเขียนโปรแกรม(programming)
4. การทดสอบและแก้ไขโปรแกรม(testing and debugging program)
5. การทำเอกสารประกอบการใช้และการบำรุงรักษาโปรแกรม (documentation and maintenance program)

วัตถุประสงค์ของงานวิจัยครั้งนี้ก็เพื่อคำนวณค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการทดสอบของสถานการณ์แต่ละสถานการณ์ โดยใช้การทดสอบ 3 วิธีคือ วิธี TMB วิธี M และวิธี GB ซึ่งสามารถวิเคราะห์งานและเขียนเป็นผังงาน(flowchart) เพื่อลำดับการทำงานให้เครื่องคอมพิวเตอร์ทำงานได้อย่างมีประสิทธิภาพ โดยได้แสดงผังงานดังกล่าวไว้ ดังนี้

ผังงานแสดงการหาค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของการทดสอบสมมติฐานเกี่ยวกับค่าผิดปกติ(H_0 : ไม่มีค่าผิดปกติ)




A
กำหนดการแจงแจงของ ε
จำนวน > 500
ใช่
k=0
ใช่
คำนวณ P(ความผิดพลาดแบบที่ 1)
ไม่ใช่
ไม่ใช่
สร้างตัวแปรอิสระ X
คำนวณค่าอำนาจการทดสอบ
คำนวณค่าตัวสถิติทดสอบ ทั้ง 3 วิธี คือ
1. วิธี TMB
2. วิธี M
3. วิธี GB
พิมพ์ผลลัพธ์
เปลี่ยนการแจกแจงของ ε
ยอมรับ H0
เปรียบเทียบค่าสถิติกับค่าวิกฤต
ปฏิเสธ H0
นับจำนวนครั้งที่ ปฏิเสธ H0 ของแต่ละวิธีการ

การทำงานของเครื่องคอมพิวเตอร์ จะทำตามลำดับดังต่อไปนี้

1. เริ่มต้นการทำงาน

2. กำหนดค่า parameter ต่างๆ ที่ใช้ในการทดลอง เช่น ค่าเฉลี่ย ความแปรปรวน สเกลแฟคเตอร์ โลเคชันแฟคเตอร์ เปอร์เซนต์การปลอมปน และอื่น ๆ ที่จำเป็น

3. กำหนดจำนวนค่าผิดปกติ ในงานวิจัยครั้งนี้ต้องการศึกษาค่าผิดปกติเท่ากับ 1, 2 และ 3 ค่าเท่านั้น ถ้าค่าผิดปกติมากกว่า 3 ค่า โปรแกรมจะหยุดการทำงานทันที

4. กรณีที่จำนวนค่าผิดปกติ เป็น 0, 1, 2 และ 3 ค่า จะทำการกำหนดขนาดตัวอย่างที่ต้องการศึกษา(n) ซึ่งในงานวิจัยนี้ต้องการศึกษาขนาดตัวอย่าง 3 ระดับ คือ 20, 50 และ 100 โดยในการทดลองนี้จะต้องทำการทดลองจนครบทุกขนาดตัวอย่างที่ต้องการศึกษาก่อน จึงค่อยเปลี่ยนจำนวนค่าผิดปกติที่ศึกษาขึ้นทีละ 1 ค่า

5. กำหนดการแจกแจงของความผิดพลาดซึ่งในงานวิจัยนี้ต้องการศึกษา 2 กรณี คือ กรณีการแจกแจงแบบหางยาวกว่าการแจกแจงปกติ และการแจกแจงแบบเบ้ ในการทดลองนี้จะทำการทดลองทีละการแจกแจงจนครบทุกขนาดตัวอย่างที่ต้องการศึกษาก่อน แล้วจึงค่อยเปลี่ยนไปศึกษาการแจกแจงอื่นต่อไป

6. ตรวจสอบจำนวนรอบของการทดลอง(round) ว่าครบ 500 รอบในแต่ละสถานการณ์ หรือไม่

6.1 กรณีทดลองครบจำนวนรอบที่ต้องการแล้วก็ให้ไปทำการตรวจสอบว่าค่า k=0 หรือไม่

ก) กรณี k=0 ก็ทำการคำนวณค่า type I error ของตัวสถิติทดสอบแต่ละวิธี จากนั้นก็ไปทำงานต่อข้อ ค)

ข) กรณี k≠0 (k=1, 2, 3) ก็ทำการคำนวณค่าอำนาจการทดสอบ จากนั้นทำข้อต่อมา (ข้อ ค))

ค) พิมพ์ผลการทดลองของตัวสถิติแต่ละวิธีเสร็จแล้วทำงานในข้อต่อมา (ข้อ ง))

ง) เปลี่ยนลักษณะการแจกแจงของ ε

6.2 กรณีทดลองยังไม่ครบจำนวนรอบที่ต้องการ จะไปทำงานที่ข้อ 7

7. สร้างตัวแปรอิสระ x

8. สร้างค่า y และค่า ε จากตัวแบบเชิงเส้นอย่างง่าย

$$y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i \qquad i=1, 2, 3, \ldots, n$$

9. คำนวณค่าตัวสถิติทดสอบจาก 3 วิธีการ คือ วิธีของทิตเจน, ม้วร์ และเบคแมน วิธีของจีแบร์รี และวิธีของเมอวีน ตามลำดับ

10. เปรียบเทียบค่าตัวสถิติที่คำนวณได้กับค่าวิกฤต เพื่อทำการทดสอบสมมติฐานว่าง(H_0 : ไม่มีค่าผิดปกติ) แล้วนับจำนวนปฏิเสธสมมติฐานว่างไว้ จากนั้นก็ทำการทดลองในรอบต่อไป

<u>โปรแกรมที่ใช้ในงานวิจัย</u> จากผังงานที่เขียนขึ้นมาสามารถนำไปเขียนเป็นโปรแกรม โดยใช้ภาษาฟอร์แทรนโฟ(FORTRAN IV)บนเครื่อง IBM 370/3010 ณ สถาบันบริการคอมพิวเตอร์จุฬาลงกรณ์ โดยแบ่งแยกโปรแกรมที่ใช้งานออกเป็น 2 โปรแกรมหลัก(MAIN PROGRAM) คือ MAIN PROGRAM ที่ 1 และ MAIN PROGRAM ที่ 2 ซึ่งรายละเอียดของแต่ละ MAIN PROGRAM ได้แสดงไว้ในภาคผนวก ข. ซึ่งในส่วนนี้จะกล่าวถึงลักษณะการทำงานของโปรแกรมทั้งหมดที่ใช้ในงานวิจัย โดยสามารถสรุปได้ดังตารางที่ 3.3

ตารางที่ 3.3 แสดงลักษณะการทำงานของโปรแกรมทั้งหมดที่ใช้ในงานวิจัย

ลำดับที่	ชื่อโปรแกรม	ลักษณะการทำงานของโปรแกรม	โปรแกรมย่อย(subroutine) หรือฟังก์ชัน(function) ที่เรียกใช้
1	MAIN PROGRAM1	ใช้คำนวณค่าความน่าจะเป็นของค่าความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของการทดสอบ 3 วิธี เมื่อการแจกแจงของความผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงปกติ ใช้ได้ 3 การแจกแจง คือ 1. แบบสเกลคอนทามิเนตนอร์มอล 2. แบบโลเคชันคอนทามิเนตนอร์มอล 3. แบบที	RAND NORMAL OLS TDIST MAXI
2	MAIN PROGRAM2	ใช้คำนวณค่าความน่าจะเป็นของค่าความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของการทดสอบ 3 วิธี เมื่อการแจกแจงของความผิดพลาดมีการแจกแจงแบบเบ้ ใช้ได้ 3 การแจกแจง คือ 1. แบบลอกนอร์มอล 2. แบบแกมมา 3. แบบไวบูลล์	RAND NORMAL YTRAN OLSBOX SHAPWK,OLS2 MAXI,GAMMA WEIBULL

ตารางที่ 3.3(ต่อ)

ลำดับที่	ชื่อโปรแกรม	ลักษณะการทำงานของโปรแกรม	โปรแกรมย่อย(subroutine) หรือฟังก์ชัน(function) ที่เรียกใช้
1	SUBROUTINE RAND	สร้างเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอในช่วง(0,1)	-----
2	FUNCTION NORMAL	สร้างเลขสุ่มที่มีการแจกแจงแบบปกติ โดยวิธีการของ Box และ Muller(1958)	RAND
3	FUNCTION TDIST	สร้างเลขสุ่มที่มีการแจกแจงแบบที	NORMAL
4	FUNCTION GAMMA	สร้างเลขสุ่มที่มีการแจกแจงแบบแกมมา	RAND
5	FUNCTION WEIBUL	สร้างเลขสุ่มที่มีการแจกแจงแบบไวบูลล์	RAND
6	SUBROUTINE OLS	คำนวณค่าสัมประสิทธิ์การถดถอย (β) โดยวิธีการกำลังสองน้อยที่สุด คำนวณค่าความคลาดเคลื่อนและค่า diagonal ของ hat matrix (เมตริกฉายา)	MULTI
7	SUBROUTINE OLSBOX	คำนวณค่าพารามิเตอร์ λ ของวิธี Box และ Cox โดยใช้วิธีกำลังสองน้อยที่สุด	-----

ตารางที่ 3.3(ต่อ)

ลำดับที่	ชื่อโปรแกรม	ลักษณะการทำงานของโปรแกรม	โปรแกรมย่อย(subroutine) หรือฟังก์ชัน(function) ที่เรียกใช้
8	SUBROUTINE OLS2	คำนวณเหมือน SUB.OLS ต่างกันที่จะใช้ OLS2 เมื่อทดสอบแล้วว่าข้อมูลมีการแจกแจงเป็นแบบปกติ	MULTI
9	SUBROUTINE SHAPWK	ใช้ทดสอบว่าข้อมูลมีการแจกแจงเป็นแบบปกติ โดยใช้ตัวสถิติ Shapiro-Wilk	RANK
10	SUBROUTINE YTRAN	ใช้คำนวณค่า y ในแต่ละค่าของ ramda(λ)	-----
11	SUBROUTINE RANK	ใช้เรียงอันดับค่าของข้อมูลจากน้อยไปมาก	-----
12	SUBROUTINE MULTI	ใช้หาผลคูณของเมตริกซ์	-----
13	SUBROUTINE MAXI	ใช้หาค่าสูงสุดของค่าคลาดเคลื่อนมาตรฐาน(maximum studentized residual)	-----