บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ ได้ทำการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบด้วยค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย สำหรับการวิเคราะห์ความถดถอยพหุนามแบบติดกลุ่ม เพื่อหาตัวแบบที่ให้ค่าพยากรณ์ถูกต้องมากที่สุด โดยเกณฑ์ที่นำมาใช้ในการคัดเลือกตัวแบบประกอบด้วย 3 เกณฑ์ ดังนี้

1. เกณฑ์การคัดเลือกตัวแบบโดยสถิติทดสอบเอฟบางส่วน(The partial F-test statistic)วิธีการกำจัดตัวแปรแบบถอยหลัง(Backward Elimination(BW))
2. เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาไคเคะ(Akaike's Information Criterion(AIC))
3. เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของเบส์(Bayesian Information Criterion(BIC))

โดยใช้เทคนิคมอนติคาร์โล(Monte Carlo Technique) หาข้อสรุปของปัญหาที่ศึกษา และใช้โปรแกรม S-plus 2000 บนเครื่องคอมพิวเตอร์ PC ในการประมวลผล ซึ่งขั้นตอนในการทำวิจัยมีรายละเอียดดังนี้

3.1 การจำลองแบบด้วยเทคนิคมอนติคาร์โล

เทคนิคมอนติคาร์โลเป็นการทดลองโดยใช้ตัวเลขสุ่มมาช่วยในการหาคำตอบของปัญหาที่ยังไม่แน่ใจในผลที่จะเกิดขึ้น เพราะเลขสุ่มมีประโยชน์หลายประการ คือ

1.ทำให้การเลือกตัวอย่างไม่มีความเอนเอียงในการสำรวจหรือทดลองในเรื่องนั้นๆ ทั้งนี้เพราะเลขสุ่มมาจากแนวคิดเกี่ยวกับการคำนวณความน่าจะเป็น

2.เลขสุ่มจะทำให้ได้มาซึ่งรูปแบบต่างๆ หรือวิธีการที่สลับซับซ้อนโดยการสร้างสถานการณ์การจำลอง(Simulation)

3.การใช้เลขสุ่มอาจทำเพื่อศึกษาคุณสมบัติทางทฤษฎีของกระบวนการทางสถิติที่มีความสำคัญสำหรับการประมาณค่า ตลอดจนนำไปสู่คำอธิบายเกี่ยวกับกำลังการทดสอบทางสถิติ (power of statistical tests)

4.เพื่อหาคำตอบในปัญหาทางคณิตศาสตร์ โดยจะพิจารณาจากการแจกแจงความน่าจะเป็นของปัญหานั้นๆ

หลักการของเทคนิคมอนติคาร์โล คือการนำเลขสุ่มมาประยุกต์ใช้ในการแก้ไขปัญหาต่างๆ ที่สนใจศึกษาถึงผลสรุปของปัญหานั้นๆ โดยมีขั้นตอนที่สำคัญ 3 ขั้นตอนดังนี้

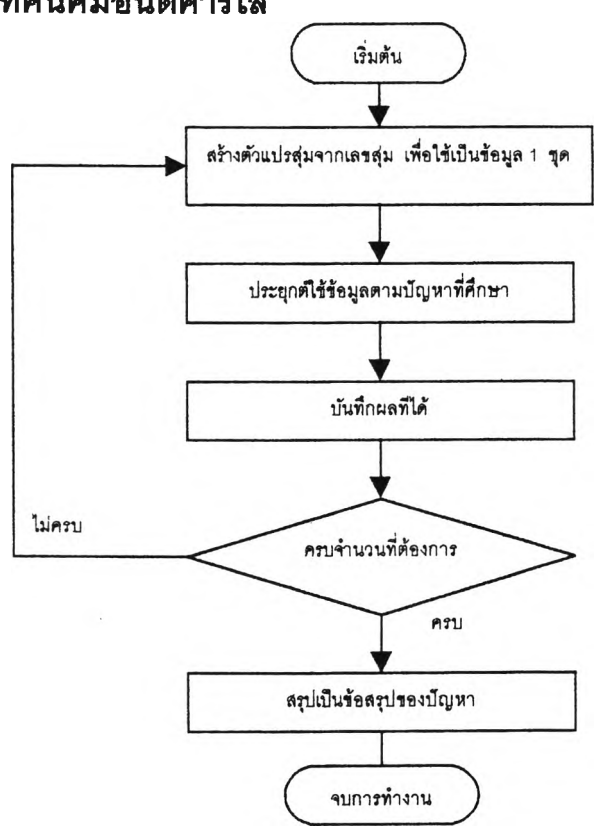
ขั้นตอนที่ 1 การสร้างเลขสุ่ม (generate random number)การสร้างเลขสุ่มจะกำหนดให้มีการแจกแจงแบบยูนิฟอร์มในช่วง[0,1] และเป็นอิสระซึ่งกันและกัน จากนั้นนำเลขสุ่มนี้ไปสร้างตัวแปร ตามลักษณะการแจกแจงที่ต้องการในปัญหาที่ศึกษา เพื่อเป็นข้อมูลของปัญหานั้นๆ

ขั้นตอนที่ 2 การประยุกต์ปัญหาที่ต้องการศึกษาโดยใช้เลขสุ่ม ขั้นตอนนี้ขึ้นอยู่กับปัญหาที่ศึกษา ซึ่งเป็นขั้นตอนที่นำเลขสุ่มมาใช้ในการหาค่าต่างๆ ตามปัญหาที่ต้องการตามสูตรการคำนวณในปัญหาที่ศึกษา

ขั้นตอนที่ 3 การทดลอง เมื่อประยุกต์ปัญหาที่ต้องการศึกษาโดยใช้เลขสุ่มแล้ว ขั้นต่อไปคือ การทำวิธีการนั้นซ้ำ ๆ กัน(replication) จำนวนหลายครั้ง โดยถือว่าการทำซ้ำๆ กันนั้นเป็นวิธีการเก็บรวบรวมข้อมูลให้มีจำนวนมากเพื่อลดความไม่แน่นอนของคำตอบ

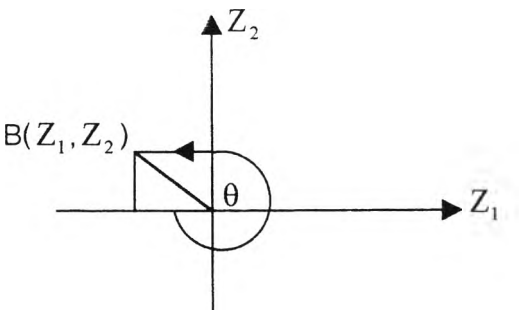
จากหลักการของเทคนิคมอนติคาร์โล จะเห็นว่าการใช้เลขสุ่มเพื่อเป็นพื้นฐานในการหาคำตอบของปัญหา เป็นวิธีการที่จะนำไปสู่แนวคิดในทางทฤษฎีที่เกี่ยวข้องกับการคำนวณโดยเฉพาะทฤษฎีความน่าจะเป็นที่จะนำไปสู่การอ้างอิงผลสรุปในสถานการณ์ของข้อมูลจริง เพราะไม่มีผลกระทบจากเรื่องอื่นๆ เข้ามาเกี่ยวข้องในการทดลองเมื่อทำซ้ำๆ กันเป็นจำนวนมากแล้ว ความคลาดเคลื่อนอย่างสุ่มที่เกิดขึ้นในการวิเคราะห์หาค่าต่าง ๆ ในแต่ละครั้งจะหมดไป (Counter balance) จากขั้นตอนเทคนิคมอนติคาร์โล สามารถเขียนแผนผังได้ดังนี้

3.1 แผนผังเทคนิคมอนติคาร์โล



3.2 การสร้างการแจกแจงแบบปกติ

การผลิตตัวเลขสุ่มที่มีการแจกแจงแบบปกติจะใช้วิธีการของ Box Muller(1958) โดยผลิตเลขสุ่มที่มีการแจกแจงปกติมาตรฐาน $N(0,1)$ พร้อมกันทั้งสองค่าและแต่ละค่าเป็นอิสระซึ่งกันและกันโดยใช้ตัวก่อกำเนิด(generator) Z_1 และ Z_2 ซึ่งพิจารณาดังรูปต่อไปนี้



พิจารณาจากรูปจะได้

$$Z_1 = B\cos(\theta) \qquad (3.1)$$

$$Z_2 = B\sin(\theta) \qquad (3.2)$$

เนื่องจาก $B^2 = Z_1^2 + Z_2^2$ มีการแจกแจงไคสแควร์ด้วยระดับขั้นความเสรี 2 และ เทียบเท่ากับการแจกแจงแบบเลขชี้กำลัง(Exponential) ค่าเฉลี่ยเท่ากับ 2 โดยใช้วิธีการแปลงผกผัน (Transformation) ซึ่งสามารถสร้างเลขสุ่มที่มีการแจกแจงแบบเลขชี้กำลัง(Exponential Distribution) ได้ดังนี้

$$B = (-2\ln r)^{\frac{1}{2}} \qquad (3.3)$$

เมื่อ r เป็นเลขสุ่มที่มีการแจกแจงสม่ำเสมอ(Uniform) ในช่วง (0,1)

จากการแจกแจงปกติที่สมมาตร จะได้ว่ามุม θ มีการแจกแจงสม่ำเสมอระหว่าง 0 ถึง 2π และรัศมี B ทำมุมกับ θ จาก (3.1) (3.2) และ (3.3) เราสามารถสร้างเลขสุ่มที่มีการแจกแจงปกติมาตรฐาน จากเลขสุ่ม 2 ชุด r_1 และ r_2 กล่าวคือ

$$Z_1 = (-2\ln r_1)^{\frac{1}{2}}\cos(2\pi r_1)$$

$$Z_2 = (-2\ln r_2)^{\frac{1}{2}}\sin(2\pi r_2)$$

ซึ่ง r_1 และ r_2 เป็นตัวเลขสุ่มที่สร้างจากฟังก์ชัน Random.seed(.....) เมื่อได้เลขสุ่มที่มีการแจกแจงปกติมาตรฐานแล้ว ทำการแปลงเลขสุ่มดังกล่าวโดยใช้ฟังก์ชัน

$$rn_1 = \mu + \sigma Z_1$$

$$rn_2 = \mu + \sigma Z_2$$

ซึ่งจะได้ว่า rn_1 และ rn_2 มีการแจกแจงปกติด้วยค่าเฉลี่ยเท่ากับ μ และมีความแปรปรวนเท่ากับ σ^2 สำหรับโปรแกรม S-plus 2000 จะใช้ฟังก์ชัน morm(n, μ, d) ในการสร้างการแจกแจงแบบปกติ โดย n แทน ขนาดตัวอย่าง μ แทนค่าเฉลี่ย และ d แทน ค่าเบี่ยงเบนมาตรฐาน

3.3 ขั้นตอนในการดำเนินงานวิจัย

1.กำหนดความคลาดเคลื่อนอย่างสุ่มจากประชากรที่มีการแจกแจงปกติ มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5 10 15 และ 25

1.1 กำหนดขนาดตัวอย่างที่ศึกษาเท่ากับ 35 50 75 และ 100

1.2 กำหนดตัวแปรอิสระเริ่มต้นเป็นค่าคงที่

1.3 กำหนดระดับนัยสำคัญ (α)คือ 0.05 และ 0.01

1.4 กำหนดตัวแปรอิสระเริ่มต้นที่ใช้สำหรับการสร้างตัวแปรตามที่เหมาะสมคือ2 ตัวแปร

1.5 กำหนดกำลังสูงสุดของตัวแปรอิสระใช้สำหรับการสร้างตัวแปรตามที่เหมาะสม คือ กำลังหก

2.สร้างข้อมูลของตัวแปรตาม y จากตัวแปรอิสระตามระดับความสัมพันธ์ต่างๆ ที่กำหนดไว้ในขอบเขตของการวิจัย

3.ทำการแปลงข้อมูลให้เป็นไปตามคุณสมบัติของพหุนามเชิงตั้งฉาก(Orthogonal Polynomial) และ การแปลงค่าเข้าสู่ศูนย์กลาง(Certering)

4.กำหนดตัวแบบเริ่มต้นในการคัดเลือกตัวแบบที่เหมาะสม

5.ทำการสร้างตัวแบบจากเกณฑ์การคัดเลือกตัวแบบทั้ง 3 เกณฑ์คือ

5.1 เกณฑ์การคัดเลือกตัวแบบโดยสถิติทดสอบเอฟบางส่วน(The partial F-test statistics)วิธีการกำจัดตัวแปรแบบถอยหลัง(Backward Elimination(BW))

5.2 เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาไคเคะ (Akaike's Information Criterion(AIC))

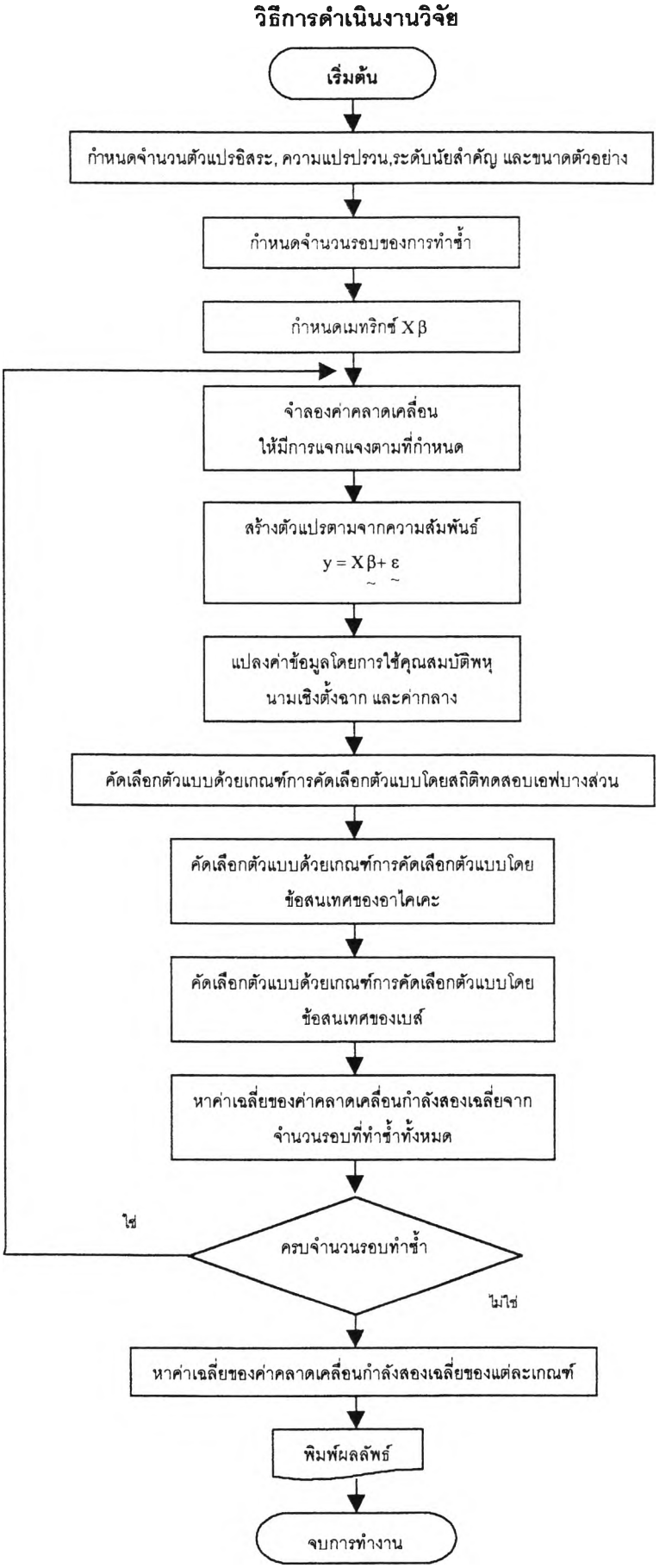
5.3 เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของเบส์(Bayesian Information Criterion(BIC))

6.คำนวณหาค่าคลาดเคลื่อนกำลังสองเฉลี่ยจากการทำซ้ำจำนวน 500 รอบ ของตัวแบบที่ได้จากเกณฑ์การคัดเลือกตัวแบบทั้ง 3 เกณฑ์

7.ทำการเปรียบเทียบค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบที่ได้จากเกณฑ์การคัดเลือกตัวแบบทั้ง 3 เกณฑ์ โดยสรุปผลในรูปของตารางและรูปภาพ(หน้า 42 – 117)

I 19962861

3.2 แผนผังแสดงขั้นตอนการดำเนินการวิจัย



สำหรับรายละเอียดของแต่ละขั้นตอนมีดังนี้

ขั้นตอนที่1 กำหนดขนาดตัวอย่าง จำนวนตัวแปรอิสระในตัวแบบ จำนวนรอบของการทำซ้ำ ระดับนัยสำคัญตามที่กำหนด และค่าคลาดเคลื่อนที่มีการแจกแจงแบบปกติโดยใช้ฟังก์ชัน morm (n,u,d) สำหรับโปรแกรม S-plus 2000 จะใช้คำสั่ง polynomial.sim<-(x,n,d,ind,inde,alpha) สำหรับกำหนดค่าข้างต้น

ขั้นตอนที่ 2 ทำการแปลงข้อมูลโดยใช้คุณสมบัติพหุนามเชิงตั้งฉาก (Orthogonal Polynomial) และการแปลงค่าเข้าสู่ศูนย์กลาง (Centering) จะกำหนดระดับความสัมพันธ์ โดยแบ่งออกเป็น 3 ระดับ นั้นคือ

1. x_1 และ x_2 สัมพันธ์กันในระดับสูง(สหสัมพันธ์=1)

   $x_1 = x \qquad x_2 = 2x_0 \quad , x_0 = 1,2,3,.....,n \, , x = 1,2,3,.....,n$

2. x_1 และ x_2 สัมพันธ์กันในระดับปานกลาง(สหสัมพันธ์=0.5)

   $$x_1 = x \qquad x_2 = \frac{1}{(1+\exp(1-(0.5 * x_0))} \quad , x_0, x = 1,2,3,.....,n$$

3. x_1 และ x_2 สัมพันธ์กันในระดับต่ำ(สหสัมพันธ์=0)

   $$x_1 = x \qquad x_2 = \frac{1}{8}x^4 - x^2 \quad , x = 1,2,3,.....,n \text{ และ } x_0 = x - \frac{1}{2}(n+1)$$

สำหรับรายละเอียดการแปลงข้อมูลสามารถดูได้จาก Bernd Droge(1997)

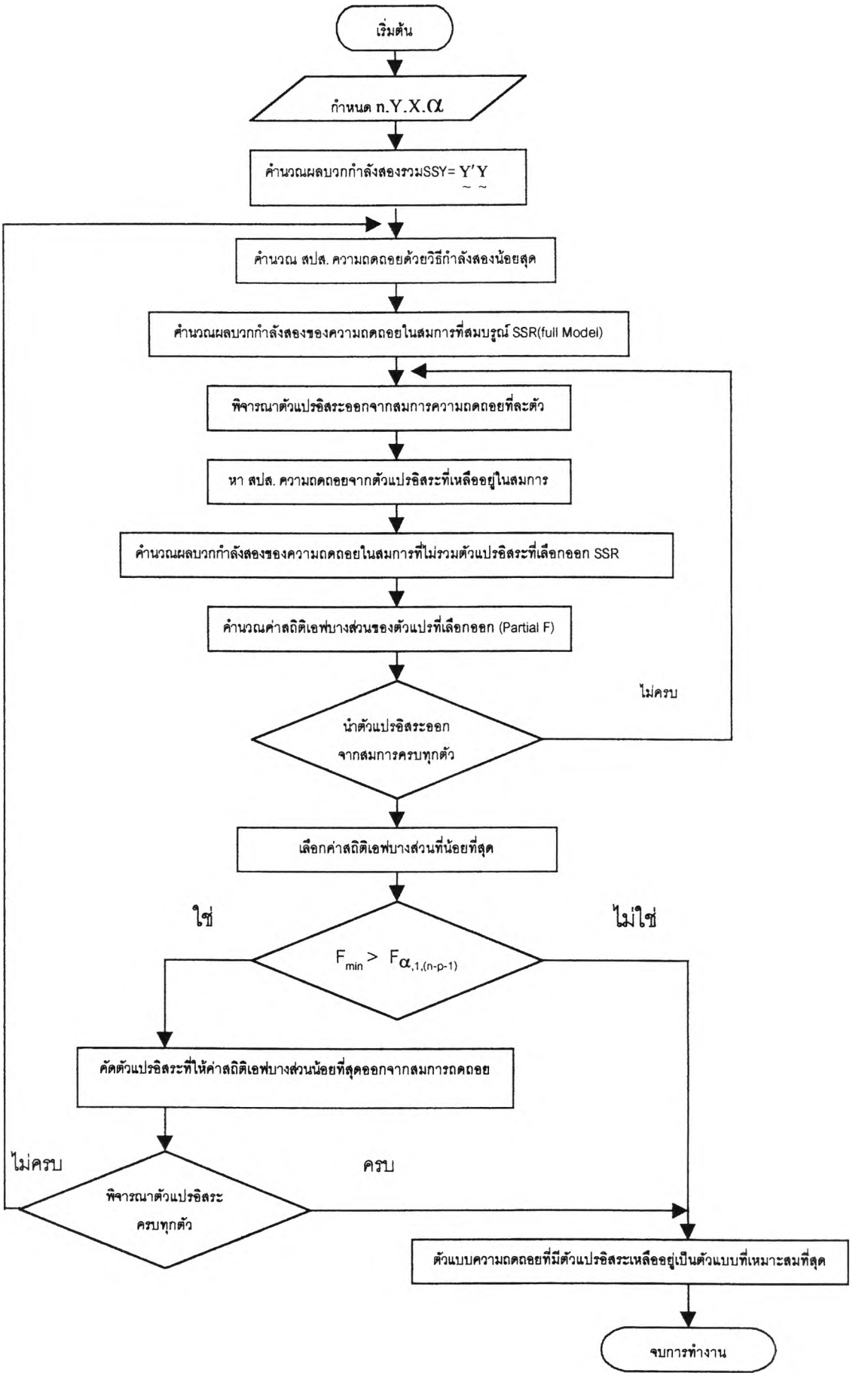
ขั้นตอนที่ 3 กำหนดตัวแบบที่สมบูรณ์(Complete model)ในการคัดเลือกตัวแบบที่เหมาะสม

ขั้นตอนที่ 4 นำข้อมูลที่กำหนดมาทำการคัดเลือกตัวแบบที่เหมาะสมจากทั้ง 3 เกณฑ์ดังต่อไปนี้

4.1 การคัดเลือกตัวแบบด้วยวิธีการกำจัดตัวแปรแบบถอยหลัง

ในวิธีการกำจัดตัวแปรแบบถอยหลัง จะเริ่มต้นจากการกำหนดสมการความถดถอยที่ประกอบไปด้วยตัวแปรอิสระทั้งหมด และจะทำการคัดเลือกตัวแปรอิสระโดยการคำนวณค่าสถิติเอฟบางส่วนทุกๆ ตัวแปรในตัวแบบ จากนั้นนำตัวแปรที่ให้ค่าสถิติเอฟบางส่วนน้อยที่สุดมาเทียบกับค่าระดับนัยสำคัญ (α) ตามที่กำหนด ถ้าไม่มีระดับนัยสำคัญก็จะทำการคัดตัวแปรนั้นออกจากตัวแบบ หลังจากนั้นก็นำตัวแปรที่เหลือมาคำนวณค่าสถิติเอฟบางส่วนอีกครั้ง ทำเช่นนี้ไปจนกว่าจะไม่มีตัวแปรใดถูกคัดออกจากตัวแบบ แต่ถ้าตัวแปรนั้นมีระดับนัยสำคัญ (α) ตามที่กำหนดก็จะเก็บตัวแปรนั้นไว้ และหยุดทำการพิจารณา

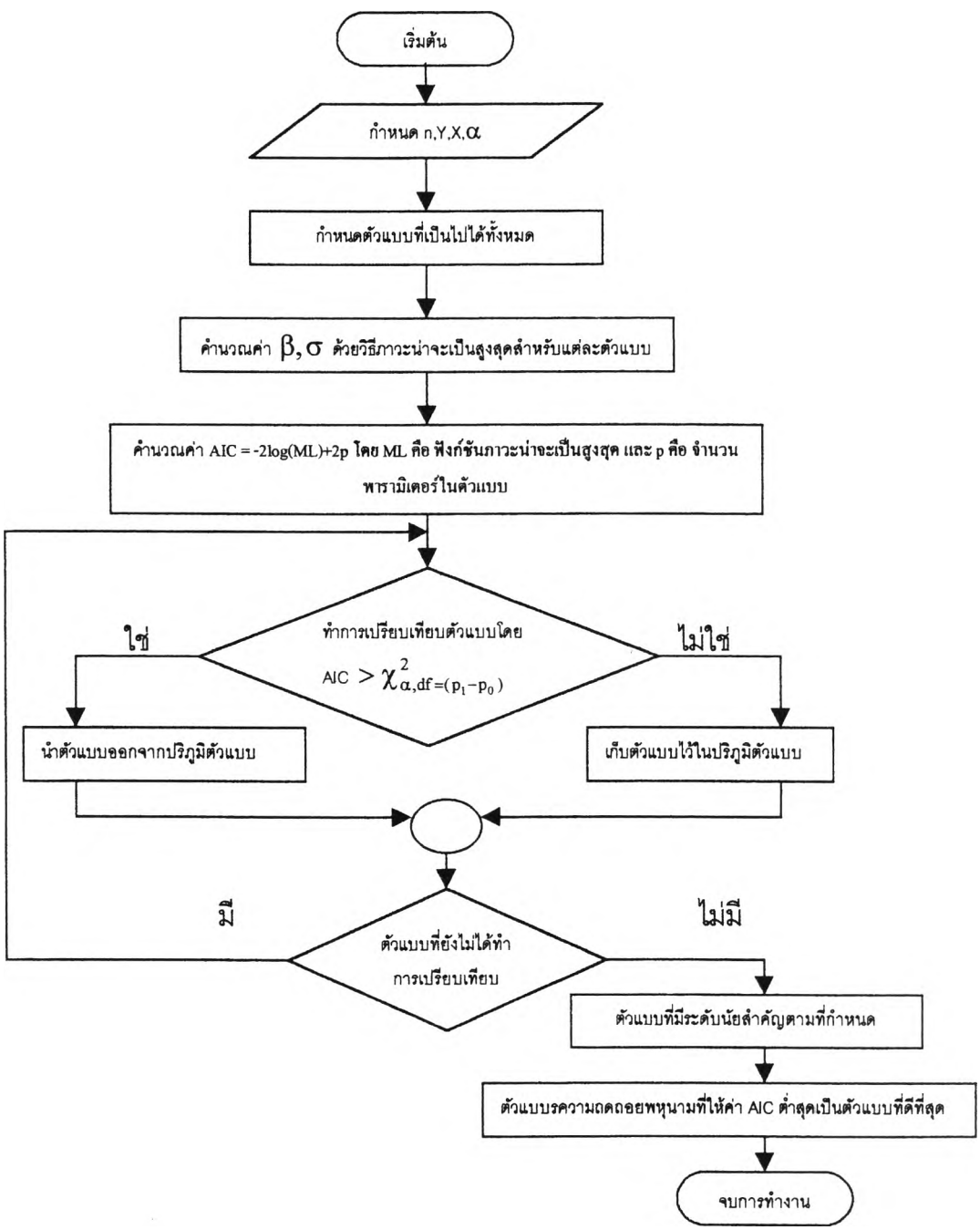
3.3 แผนผังวิธีการกำจัดตัวแปรแบบถอยหลัง



4.2 **เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาไคเคะ** (Akaike's Information Criterion (AIC))

เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาไคเคะ จะเริ่มจากการสร้างตัวแบบที่เป็นไปได้ทั้งหมดจากตัวแบบที่กำหนด แล้วทำการเปรียบเทียบตัวแบบด้วยการทดสอบสมมติฐานแบบติดกลุ่ม และเก็บตัวแบบที่มีระดับนัยสำคัญไว้ จากนั้นนำตัวแบบที่มีระดับนัยสำคัญมาพิจารณาค่า AIC โดยตัวแบบที่ให้ค่า AIC ต่ำสุดเป็นตัวแบบที่ให้ค่าพยากรณ์ถูกต้องและแม่นยำมากที่สุด

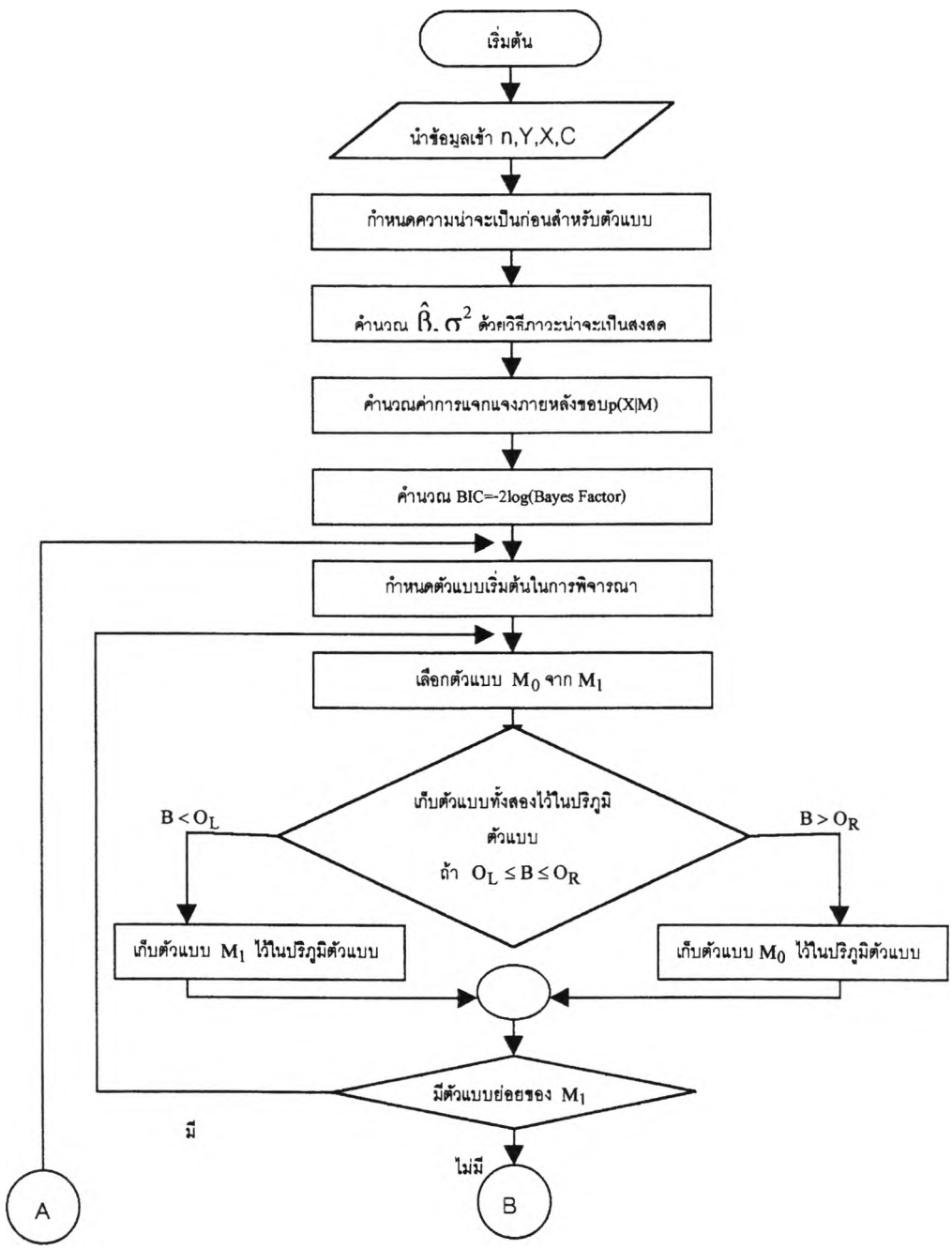
3.4 **แผนผังเกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาไคเคะ** (Akaike's Information Criterion(AIC))

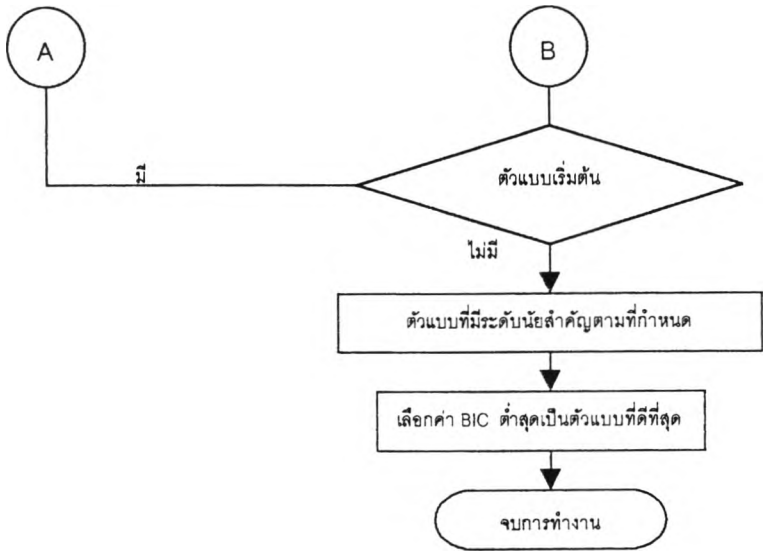


6.3 **เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของเบส์**(Bayesian Information Criterion (BIC))

เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของเบส์ จะเริ่มจากการสร้างตัวแบบที่เป็นไปได้ทั้งหมดจากตัวแบบที่กำหนด แล้วทำการเปรียบเทียบตัวแบบด้วยวิธีการ Occam's Window สำหรับทดสอบสมมติฐานแบบติดกลุ่ม และเก็บตัวแบบที่มีระดับนัยสำคัญตามที่กำหนดไว้ จากนั้นนำตัวแบบที่มีระดับนัยสำคัญมาพิจารณาค่า BIC โดยตัวแบบที่ให้ค่า BIC ต่ำสุดเป็นตัวแบบที่ให้ค่าพยากรณ์ถูกต้องและแม่นยำมากที่สุด

3.5 **แผนผังเกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของเบส์** (Bayesian Information Criterion(BIC))





ในการพิจารณาตัวแบบขั้นตอนแรกจะใช้กระบวนการ "Up" ในการเปรียบเทียบตัวแบบและทำการกำหนดตัวแบบเริ่มต้นจากกระบวนการ "Down" โดยการกระบวนการ "Up" เป็นการค้นหาตัวแบบโดยการเพิ่มตัวแปร และ "Down" เป็นกระบวนการในการค้นหาตัวแบบเริ่มต้นโดยการลดตัวแปร รายละเอียดสามารถอ่านได้จากบทที่ 2 ในหัวข้อ 2.5 (หน้า 27-29)

ขั้นตอนที่ 6 คำนวณหาค่าคลาดเคลื่อนกำลังสองเฉลี่ยจากการทำซ้ำจำนวน 500 รอบ ของตัวแบบที่ได้จากเกณฑ์การคัดเลือกตัวแบบทั้ง 3 เกณฑ์

ขั้นตอนที่ 7 ทำการเปรียบเทียบค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย สำหรับตัวแบบที่ได้จากเกณฑ์การคัดเลือกตัวแบบทั้ง 3 เกณฑ์ โดยสรุปผลในรูปของตารางและรูปภาพ(หน้า 42 – 117)

สำหรับรายละเอียดของโปรแกรมดูได้จากภาคผนวก