บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยในครั้งนี้ต้องการศึกษาเปรียบเทียบการจำแนกกลุ่ม 2 กลุ่ม ระหว่างการวิเคราะห์การถดถอยทวิและการวิเคราะห์จำแนกประเภท โดยศึกษาในกรณีที่ประชากรของตัวแปรอิสระมีการแจกแจงแบบเบ้คือ แบบไวบูลล์ แบบแกมมาและแบบลอกนอร์มอล นอกจากนี้ยังได้นำการแจกแจงแบบปกติมาตรฐานมาศึกษาเปรียบเทียบด้วยเพราะว่าเป็นการแจกแจงที่จำเป็นสำหรับข้อสมมติเบื้องต้นสำหรับการจำแนกกลุ่ม เนื่องความชัดเจนของวิธีการประมาณค่าที่นำมาเปรียบเทียบในการเปรียบเทียบจะพิจารณาจากร้อยละเฉลี่ยที่พยากรณ์ได้ถูกต้องสำหรับข้อมูลของประชากรที่ใช้ศึกษาในครั้งนี้ กำหนดสถานการณ์ต่าง ๆ โดยใช้เทคนิคการจำลองค่าโดยวิธีมอนติคาร์โล ซึ่งเป็นวิธีหนึ่งที่นิยมใช้กันอย่างแพร่หลาย ในปัจจุบันมีผู้ทำวิทยานิพนธ์หลายท่านได้ใช้วิธีนี้ในการวิจัยเช่นกันและได้กล่าวถึงวิธีมอนติคาร์โลไว้แล้วอย่างละเอียดซึ่งสามารถหาอ่านได้ง่าย จึงไม่ขอกล่าวถึงรายละเอียดในการวิจัยครั้งนี้อีก ดังนั้นจะกล่าวถึงรายละเอียดของแผนการทดลอง ขั้นตอนการวิจัยและโปรแกรมที่ใช้ในการวิจัยตามลำดับ ซึ่งรายละเอียดต่าง ๆ เป็นดังนี้

3.1 แผนการทดลอง

การวิจัยในครั้งนี้กำหนดสถานการณ์ต่าง ๆ ที่ต้องการศึกษาโดยสุ่มตัวอย่างจากประชากรที่มีการแจกแจงแบบเดียวกัน ซึ่งลักษณะการแจกแจงแบบเบ้ที่ต้องการศึกษา คือ การแจกแจงแบบแกมมา แบบไวบูลล์และแบบลอกนอร์มอล สำหรับอีกการแจกแจงหนึ่งที่นำมาศึกษาคือการแจกแจงแบบปกติ เพื่อให้สอดคล้องกับปัญหาที่พบและสอดคล้องกับตัวแบบมากที่สุด จึงได้กำหนดให้ความผิดพลาด (error) มีการแจกแจงแบบปกติมาตรฐาน

สำหรับแต่ละการแจกแจงที่นำมาเปรียบเทียบนั้นพยายามควบคุมให้มีความแตกต่างกันน้อยที่สุด เพื่อไม่ให้มีผลต่อวิธีการพยากรณ์ การควบคุมนั้นพยายามควบคุมทั้งเทคนิคการจำลองค่าและพารามิเตอร์ที่จะนำมาใช้ โดยพิจารณาจากค่าเฉลี่ยและความแปรปรวนเป็นหลักให้ใกล้เคียงกันมากที่สุด ภายใต้ข้อจำกัดของพารามิเตอร์แต่ละการแจกแจงดังนี้

การแจกแจงแบบแกมมาและไวบูลล์ สนใจศึกษาที่ค่าพารามิเตอร์ $\beta = 1$, $\alpha = 1$ สำหรับการแจกแจงแบบลอกนอร์มอลและแบบปกติ สนใจศึกษาเมื่อ $\mu = 0$, $\sigma^2 = 1$

ทุก ๆ การแจกแจงจะใช้ขนาดตัวอย่าง 10, 30 และ 60 จำนวนตัวแปรอิสระเท่ากับ 1 และ 2 กำหนดสัดส่วนของตัวแปรตามระหว่าง 0 กับ 1 จำนวน 10 สัดส่วน คือ 0.50:0.50, 0.55:0.45, ..., 0.95:0.05

3.2 ขั้นตอนในการวิจัย

ขั้นตอนในการวิจัยมีดังนี้ คือ

1. การสร้างโปรแกรมย่อย (subroutine) สำหรับการแจกแจงของค่าตัวแปรอิสระที่มีการแจกแจงแบบเบ้และค่าความผิดพลาด มีการแจกแจงแบบปกติมาตรฐาน

2. การสร้างข้อมูล ที่จะใช้ในการคำนวณ คือ ตัวแปรตาม (y) ถูกสร้างจากตัวแปรอิสระ (x) และค่าความผิดพลาดที่สร้างจากขั้นตอนที่ 1 โดยให้ตัวแปรตามมีความสัมพันธ์เชิงเส้นตรงกับตัวแปรอิสระ

3. การสร้างสมการในการประมาณค่าโดยวิธีกำลังสองน้อยที่สุดแบบธรรมดา วิธีกำลังสองน้อยที่สุดแบบทั่วไปและสมการประมาณค่าโดยวิธีการวิเคราะห์จำแนกประเภทพร้อมกับคำนวณค่าพยากรณ์จากแต่ละวิธี

4. การหาค่าเฉลี่ยของร้อยละที่พยากรณ์ได้ถูกต้องโดยแบ่งการคำนวณเป็น 2 วิธี คือ

4.1 - คำนวณจากทุกหน่วยตัวอย่างที่สุ่มมา

4.2 - คำนวณจากส่วนที่เหลือเมื่อหักออกด้วยส่วนที่ถูกต้องเหมือนกัน

3.2.1 การสร้างโปรแกรมย่อยสำหรับสร้างการแจกแจงของค่าตัวแปรอิสระที่ต้องการศึกษา

การสร้างลักษณะการแจกแจงของค่าตัวแปรอิสระทุกรูปแบบนั้นใช้โปรแกรมภาษาฟอร์แทรน 77 (FORTRAN 77) โดยใช้กับเครื่องคอมพิวเตอร์ AMDAHL-5860 ในการสร้างลักษณะการแจกแจงแบบต่างๆ นั้น จะต้องใช้ตัวเลขสุ่มเป็นพื้นฐานในการสร้างนี้และสร้างโดยอาศัยตัวแปร xx ทุกการแจกแจง หลังจากนั้นจึงกำหนดค่าให้กับตัวแปรอื่น ๆที่ต้องการ

3.2.1.1 การแจกแจงแบบปกติ โปรแกรมย่อยที่ใช้ในการแจกแจงแบบปกติจะใช้วิธีของ Box และ Muller (1958) ซึ่งเป็นวิธีที่สร้างการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น μ และความแปรปรวนเป็น σ^2 แต่ในที่นี้ กำหนดค่า $\mu = 0$, $\sigma^2=1$ โดย SMEAN และ SIGMA คือค่าเฉลี่ยและความแปรปรวนที่กำหนดสำหรับรายละเอียดแสดงไว้ในภาคผนวก ก. การใช้โปรแกรมย่อยนี้ใช้คำสั่ง NORMI(SMEAN,SIGMA,XX,IX,IK)ผลลัพธ์คือ XX ที่มีค่าเฉลี่ยและความแปรปรวนที่กำหนด

3.2.1.2 การแจกแจงแบบลอกนอร์มอล คำนวณได้จากการหา Exponential ของโปรแกรมย่อย NORMI(SMEAN,SIGMA,XX,IX,IK) เมื่อ SMEAN และ SIGMA เป็นค่าเฉลี่ยและความแปรปรวนของการแจกแจงแบบปกติ สำหรับรายละเอียดแสดงไว้ในภาคผนวก ก.

3.2.1.3 การแจกแจงแบบแกมมาโปรแกรมย่อยที่ใช้คือ GAM(ALPHA, BTA,XX,IX) ฟังก์ชันนี้สร้างโดยวิธี Reproductive propertyซึ่งรายละเอียดแสดงไว้ในภาคผนวก ก. โปรแกรมย่อย GAM จะกำหนดค่า ALPHA และ BTA เพื่อให้ได้ค่าเฉลี่ยและความแปรปรวนตามที่ต้องการ ผลลัพธ์คือค่า XX จะมีการแจกแจงแบบแกมมามีค่าเฉลี่ยเป็น BTA * ALPHA และความแปรปรวนเป็น $(BTA)^2$*ALPHA

3.2.1.4 การแจกแจงแบบไวบูลล์โปรแกรมย่อยที่ใช้คือ WEIB(ALPHA, BTA,XX,IK) โปรแกรมย่อยนี้สร้างโดยวิธี inverse transformation รายละเอียดแสดงไว้ในภาคผนวก ก.โปรแกรมย่อยจะกำหนดค่า ALPHA, BTA เพื่อให้ได้ค่าเฉลี่ยและความแปรปรวนตามที่ต้องการ ผลลัพธ์คือ ค่า XX จะมีการแจกแจงแบบไวบูลล์ ซึ่งมีค่าเฉลี่ยเป็น BTA1* (1+1/ALPHA) และความแปรปรวนเป็น $(BTA)^2$*[Γ(1+2/ALPHA)-Γ^2(1+1/ALPHA)]

3.2.2 การสร้างข้อมูลให้มีความสัมพันธ์เชิงเส้นตรง

ในการวิจัยครั้งนี้จะสร้างตัวแปรอิสระ X ซึ่งมีการแจกแจงตามที่กำหนดก่อนแล้วจึงสร้างตัวแปรตาม y โดยอาศัยตัวแปรอิสระ X และค่าความผิดพลาด รูปแบบของความสัมพันธ์คือ $\underline{y} = x\underline{\beta} + \underline{\varepsilon}$ เมื่อ $\underline{\beta}$ เป็นค่าพารามิเตอร์ที่ถูกกำหนดขึ้นมา และ $\underline{\varepsilon}$ เป็นความผิดพลาดที่มีการแจกแจงเป็นแบบปกติมาตรฐาน ดังที่กล่าวมาแล้วเมื่อได้ค่า y แล้วให้แปลงค่าของตัวแปรตาม y ให้เป็นคะแนนมาตรฐาน แล้วปรับให้เป็น 1 หรือ 0 เพื่อใช้ในการกำหนดสัดส่วนของตัวแปรตาม

3.2.3 การสร้างสมการในการประมาณค่า

การประมาณค่าสัมประสิทธิ์การถดถอย การแปลงค่าพยากรณ์ทุกรูปแบบอาศัยวิธี OLS เป็นพื้นฐาน ดังนั้น โปรแกรมย่อยที่เกี่ยวข้องในการสร้างค่าประมาณของ $\underline{\beta}$ จากสมการ $\underline{\beta} = (X'X)^{-1}X'\underline{y}$ คือ

- โปรแกรมย่อย MULT สำหรับคูณเมตริกซ์ คำนวณเมตริกซ์ผกผัน (inverse Matrix)
- โปรแกรมย่อย INV2 สำหรับตัวแปรอิสระ 1 ตัวแปร
- โปรแกรมย่อย INV3 สำหรับตัวแปรอิสระ 2 ตัวแปร

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

3.2.4 การคำนวณค่าเฉลี่ยของร้อยละ ที่พยากรณ์ได้ถูกต้อง

การเปรียบเทียบประสิทธิภาพสำหรับวิธีการวิเคราะห์การถดถอยและการวิเคราะห์จำแนกกลุ่มเพื่อการแยกกลุ่ม 2 กลุ่ม ในสัดส่วนของตัวแปรตามแต่ละแบบ โดยจำแนกตามการแจกแจง ขนาดตัวอย่างและวิธีการพยากรณ์ต่าง ๆ จะพิจารณาจากความสามารถในการพยากรณ์หรือการจำแนกกลุ่มได้ถูกต้อง หรืออาจจะพิจารณาจาก misclassification error แต่การวิจัยครั้งนี้จะพิจารณาจากร้อยละที่พยากรณ์ได้ถูกต้อง โดยพิจารณาแยกเป็น 2 กรณี

3.2.4.1 พิจารณาว่าพยากรณ์ได้ถูกต้องร้อยละเท่าใดจากทั้งหมดที่ได้พยากรณ์ พิจารณาจาก confusion Matrix ต่อไปนี้

Predicted membership

		¥1	¥0	
Actual membership	¥1	n_{1c}	$n_{1m} = n_1 - n_{1c}$	n_1
	¥0	$n_{0m} = n_0 - n_{0c}$	n_{0c}	n_0

n_{1c} = จำนวนวัตถุหรือตัวอย่างที่เดิมเคยอยู่ใน ¥1 และจัดกลุ่ม ¥1 ได้ถูกต้อง

n_{0c} = จำนวนวัตถุหรือตัวอย่างที่เดิมเคยอยู่ใน ¥0 และจัดกลุ่ม ¥0 ได้ถูกต้อง

n_{1m} = จำนวนวัตถุหรือตัวอย่างที่เดิมเคยอยู่ใน ¥1 แต่จัดเข้ากลุ่มที่ ¥0 (จัดกลุ่มผิด)

n_{0m} = จำนวนวัตถุหรือตัวอย่างที่เดิมเคยอยู่ใน ¥0 แต่จัดเข้ากลุ่มที่ ¥1 (จัดกลุ่มผิด)

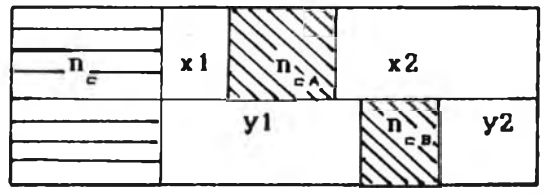
c = correctly classified

m = misclassified

ดังนั้น ร้อยละที่จัดเข้ากลุ่มถูกต้อง $= \dfrac{(n_{1c} + n_{0c})}{(n_1 + n_0)} \times 100$

3.2.4.2 พิจารณาร้อยละที่จำแนกได้ถูกต้องเมื่อหักออกด้วยจำนวนที่จำแนกได้ถูกต้องตรงกัน การพิจารณาในกรณีนี้จะทำให้เห็นชัดเจนกว่ากรณี 3.2.4.1 เพราะสามารถบอกได้ว่าการจำแนกกลุ่มพิจารณาจากส่วนที่เหลือซึ่งมีจำนวนน้อยทำให้เห็นชัดเจนมากยิ่งขึ้นและยังเปรียบเทียบให้เห็นว่าประสิทธิภาพในการจำแนกกลุ่มของวิธีการใดมีลักษณะเด่น

หรือมีประสิทธิภาพพอ ๆ กัน การคำนวณจะใช้วิธีการง่าย ๆ โดยพิจารณาจากรูปต่อไปนี้



= ถูกต้องตรงกัน n_c = จำนวนพยากรณ์ถูกต้องตรงกัน

= ถูกต้องแตกต่างกัน

n_{cA} = จำนวนที่วิธี A พยากรณ์ได้ถูกต้อง แต่วิธี B พยากรณ์ไม่ถูกต้อง

n_{cB} = จำนวนที่วิธี B พยากรณ์ได้ถูกต้อง แต่วิธี A พยากรณ์ไม่ถูกต้อง

x1,x2,y1 และ y2 คือส่วนที่ทั้ง 2 วิธีการพยากรณ์ผิดพลาด

ดังนั้น ร้อยละที่พยากรณ์ได้ถูกต้องบางส่วน สำหรับแต่ละวิธีการพยากรณ์ คือ

$$\text{วิธี A} = \frac{n_{cA} \times 100}{n - n_c}$$

$$\text{วิธี B} = \frac{n_{cB} \times 100}{n - n_c}$$

3.3 ขั้นตอนในการทำงานของโปรแกรมที่ใช้ในการวิจัย

แบ่งการทำงานออกเป็น 3 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 : กำหนดค่าพารามิเตอร์และสร้างข้อมูลเพื่อใช้ในการวิเคราะห์

MAIN PROGRAM 1

1.1 สร้างข้อมูลของตัวแปรอิสระที่มีการแจกแจงแบบเบ้ สำหรับการแจกแจง
-แบบไวบูลล์ ใช้โปรแกรมย่อย ชื่อ WEIB
-แบบแกมมา ใช้โปรแกรมย่อย ชื่อ GAM
-แบบลอกนอร์มอล ใช้โปรแกรมย่อย ชื่อ NORM1

1.2 สร้างข้อมูลของตัวแปรอิสระที่มีการแจกแจงแบบปกติ ใช้โปรแกรมย่อยชื่อ NORM1

1.3 สร้างข้อมูลค่าผิดพลาดที่มีการแจกแจงแบบปกติ ใช้โปรแกรมย่อย ชื่อ NORM1

1.4 สร้างข้อมูลของตัวแปรตามจากข้อมูล 1.1 - 1.3

ขั้นตอนที่ 2 : ส่วนของการวิเคราะห์ข้อมูล แบ่งเป็น 3 โปรแกรมหลัก คือ

MAIN PROGRAM 2

2.1 คำนวณค่าประมาณของพารามิเตอร์ β ซึ่งเป็นค่าเริ่มต้น โดยวิธี OLS ใช้โปรแกรมย่อยชื่อ MULT และ INV2 หรือ INV3

2.2 คำนวณค่าพยากรณ์โดยวิธี OLS กำหนดให้เป็นค่ากลาง (Intermediate value)

2.3 ปรับค่าพยากรณ์ให้อยู่ในช่วง [0,1] โดยใช้เส้นโค้งการแจกแจงในรูปแบบต่อไปนี้

-Normit model ปรับโดยใช้โปรแกรมย่อยชื่อ NORMIT

-Logit model

-Gompit model

2.4 คำนวณค่าประมาณของ V และ P

2.5 พยากรณ์ค่าตัวแปรตามสำหรับแต่ละวิธีการและแบ่งกลุ่มค่าพยากรณ์ออกเป็น 1 หรือ 0

2.6 เปรียบเทียบค่าพยากรณ์ที่ได้จาก 2.2 และ 2.5 กับค่า y ที่สร้างจาก 1.3

MAIN PROGRAM 3 : โปรแกรมสำหรับคำนวณการวิเคราะห์จำแนกประเภททั้ง 2 วิธี คือ DSC และ EMC

MAIN PROGRAM 4 : คำนวณร้อยละที่พยากรณ์ถูกต้องจากทุกวิธี

4.1 ร้อยละที่พยากรณ์ถูกต้อง ที่คำนวณจากทุกหน่วยตัวอย่าง

4.2 ร้อยละที่พยากรณ์ถูกต้อง ที่คำนวณจากส่วนที่เหลือ เมื่อหักออกด้วยส่วนที่พยากรณ์ถูกต้องตรงกัน ใช้โปรแกรมย่อยชื่อ SCORE และ TRUE

ขั้นตอนที่ 3 : สรุปผล

MAIN PROGRAM 5 : สรุปผลการคำนวณจากการทำซ้ำจำนวน 500 รอบ

5.1 คำนวณค่าเฉลี่ยของร้อยละที่พยากรณ์ถูกต้อง

5.2 พิมพ์ผลลัพธ์จากการวิเคราะห์