บทที่ 3
วิธีดำเนินการวิจัย

งานวิจัยนี้มีลักษณะเป็นการวิจัยเชิงทดลอง เพื่อเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยในตัวแบบการถดถอยเชิงเส้นปกติเชิงเดียว (simple normal linear regression model) โดยเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอย 3 วิธี คือ วิธีกำลังสองน้อยสุด (Least Square method) วิธีเชิงเบส์เมื่อใช้การแจกแจงก่อนที่ไม่ให้ข้อมูล (Bayesian method using noninformative prior) และวิธีเชิงเบส์เมื่อใช้การแจกแจงก่อนที่ให้ข้อมูล (Bayesian method using informative prior) ด้วยการเปรียบเทียบค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของค่าประมาณสัมประสิทธิ์การถดถอย ภายใต้สถานการณ์ต่างๆ ที่กำหนด โดยการจำลองข้อมูลใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Technique) ซึ่งกระทำซ้ำ 500 ครั้งในแต่ละสถานการณ์

3.1 แผนการทดลอง

ผู้วิจัยกำหนดสถานการณ์ต่างๆ ที่ต้องการศึกษาดังนี้

3.1.1 ขนาดตัวอย่างเท่ากับ 10 30 50 และ 100

3.1.2 กลุ่มตัวอย่างของความคลาดเคลื่อนสุ่มจากประชากรที่มีการแจกแจงปกติ ซึ่งมีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.1 0.3 0.5 0.7 และ 0.9

3.1.3 ตัวแปรอิสระเป็นค่าคงที่ซึ่งสุ่มจากประชากรที่มีการแจกแจงปกติด้วยค่าเฉลี่ยเท่ากับ 1 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.1 0.15 0.2 0.25 0.3 0.5 0.7 และ 0.9 นั่นคือ เปอร์เซ็นต์ของสัมประสิทธิ์ความแปรผันของตัวแปรอิสระเท่ากับ 10% 15% 20% 25% 30% 50% 70% และ 90%

3.1.4 กำหนดค่าสัมประสิทธิ์การถดถอยในตัวแบบ เมื่อ β_0 เท่ากับ 0.5 และ β_1 เท่ากับ 0.5

3.1.5 จากข้อ 3.1.2 3.1.3 และ 3.1.4 เราจะได้ว่ากลุ่มตัวอย่างของตัวแปรตามมีการแจกแจงปกติ ซึ่งมีค่าเฉลี่ยเท่ากับ 1 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.1 0.3 0.5 0.7 และ 0.9

3.2 ขั้นตอนการวิจัย

3.2.1 การสร้างข้อมูลของตัวแปรอิสระ (X) ที่เป็นค่าคงที่ซึ่งสุ่มจากการแจกแจงปกติ โดยมีลักษณะข้อมูลแบบต่างๆ ตามค่าสัมประสิทธิ์การถดถอยที่กำหนด

3.2.2 กำหนดค่าสัมประสิทธิ์การถดถอย $(\underset{\sim}{\beta} = (\beta_0, \beta_1)')$ ที่ใช้ใน ตัวแบบการถดถอยเชิงเส้นปกติเชิงเดียว $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \ , i = 1,2,\ldots,n$

3.2.3 การสร้างข้อมูลของความคลาดเคลื่อน $(\underset{\sim}{\varepsilon})$ ให้มีการแจกแจงปกติโดยมีลักษณะข้อมูลต่างๆ ตามค่าพารามิเตอร์ที่กำหนด

3.2.4 การสร้างข้อมูลของตัวแปรตาม $(\underset{\sim}{y})$ ที่มีรูปแบบความสัมพันธ์เชิงเส้น คือ

$$\underset{\sim}{y} = X\,\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

เมื่อ $\underset{\sim}{y} = (y_1, y_2, \ldots, y_n)'$

$$X = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ x_1 & x_2 & \ldots & x_n \end{bmatrix}'$$

$$\underset{\sim}{\beta} = (\beta_0, \beta_1)'$$

และ $\underset{\sim}{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$

3.2.5 การประมาณค่าสัมประสิทธิ์การถดถอยของวิธีที่ศึกษา ซึ่งมีค่าประมาณสัมประสิทธิ์การถดถอยของวิธีกำลังสองน้อยสุด วิธีเชิงเบส์ที่ใช้การแจกแจงก่อนที่ให้ข้อมูล และวิธีเชิงเบส์ที่ใช้การแจกแจงก่อนที่ไม่ให้ข้อมูล โดยค่าประมาณเบส์จะได้จากค่าเฉลี่ยของกลุ่มตัวอย่างที่สร้างจากการแจกแจงภายหลังที่ได้จากการใช้การแจกแจงก่อนดังกล่าว ตามลำดับ

3.2.6 การหาค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของตัวประมาณสัมประสิทธิ์การถดถอยของแต่ละวิธี

3.2.7 การเปรียบเทียบประสิทธิภาพของตัวประมาณสัมประสิทธิ์การถดถอยและสรุปผลที่ได้

3.3 การประมาณค่าสัมประสิทธิ์การถดถอยจากแต่ละวิธี

เมื่อสร้างข้อมูล (x_i, y_i) ที่มีความสัมพันธ์เชิงเส้นตรงและรูปแบบความคลาดเคลื่อนตามสถานการณ์ต่างๆที่กำหนดแล้ว ต่อมานำข้อมูล (x_i, y_i) ที่ได้ไปประมาณค่าสัมประสิทธิ์การถดถอย $\underset{\sim}{\beta}$ โดยผู้วิจัยเสนอวิธีประมาณค่าสัมประสิทธิ์การถดถอยไว้ดังนี้

3.3.1 วิธีกำลังสองน้อยสุด ซึ่งค่าประมาณกำลังสองน้อยสุด ($\underset{\sim}{b}$) สามารถคำนวณได้จาก $\underset{\sim}{b}=(X'X)^{-1}X'\underset{\sim}{y}$ เมื่อ $\underset{\sim}{b}=(\hat{\beta}_0,\hat{\beta}_1)'$

3.3.2 วิธีเชิงเบส์เมื่อใช้การแจกแจงก่อนที่ไม่ให้ข้อมูล ซึ่งค่าประมาณเบส์เมื่อใช้การแจกแจงก่อนที่ไม่ให้ข้อมูลเป็นค่าเฉลี่ยที่คำนวณได้จากกลุ่มตัวอย่างที่สร้างจากการแจกแจงภายหลังขอบ (marginal posterior distribution) ที่ใช้การแจกแจงก่อนที่ไม่ให้ข้อมูลของ β_0 และ β_1 เมื่อ $\beta_0 \sim N(\hat{\beta}_0,\sigma^2 c_{11})$ และ $\beta_1 \sim N(\hat{\beta}_1,\sigma^2 c_{22})$ โดยที่ c_{ii} คือ ค่าในตำแหน่งที่ (i,i) ของเมทริกซ์ $(X'X)^{-1}$

3.3.3 วิธีเชิงเบส์เมื่อใช้การแจกแจงก่อนที่ให้ข้อมูล ซึ่งค่าประมาณเบส์เมื่อใช้การแจกแจงก่อนที่ให้ข้อมูลเป็นค่าเฉลี่ยที่คำนวณได้จากกลุ่มตัวอย่างที่สร้างจากการแจกแจงภายหลังขอบ (marginal posterior distribution) ที่ใช้การแจกแจงก่อนที่ให้ข้อมูลของ β_0 และ β_1 เมื่อ $\beta_0 \sim N(\overline{\hat{\beta}}_0,\sigma^2 c_{11}^*)$ และ $\beta_1 \sim N(\overline{\hat{\beta}}_1,\sigma^2 c_{22}^*)$ โดยที่ c_{ii}^* คือ ค่าในตำแหน่งที่ (i,i) ของเมทริกซ์ $(A+X'X)^{-1}$ และ $\sigma^2(A+X'X)^{-1}$ คือ เมทริกซ์ความแปรปรวนร่วมภายหลัง

3.4 การหาค่า Z ที่เหมาะสม

เนื่องจากค่าเฉลี่ยก่อน (prior mean) ของการแจกแจงก่อนที่ให้ข้อมูลมีผลต่อตัวประมาณเบส์เมื่อใช้การแจกแจงก่อนดังกล่าว ดังนั้นผู้วิจัยจึงกำหนดค่าในเมทริกซ์ความแปรปรวนร่วมก่อน (prior covariance) ให้เป็นค่าคงที่ใดๆ และกำหนดค่าของเวกเตอร์ค่าเฉลี่ยก่อนให้เป็นฟังก์ชันของค่าปกติมาตรฐาน Z แล้วทำการหาค่า Z ที่เหมาะสมที่ทำให้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของตัวประมาณสัมประสิทธิ์การถดถอยด้วยวิธีการเชิงเบส์ที่ใช้การแจกแจงก่อนที่ให้ข้อมูลมีค่าสูงสุด แต่มีค่าต่ำกว่าค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของวิธีกำลังสองน้อยสุด โดย Z ที่มีค่ายิ่งสูงจะทำให้ค่าเฉลี่ยก่อนมีค่ายิ่งห่างจากค่าสัมประสิทธิ์การถดถอย ซึ่งจะส่งผลให้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของตัวประมาณมีค่าสูงขึ้น

ผู้วิจัยใช้โปรแกรมคอมพิวเตอร์คำนวณหาค่า Z ที่เหมาะสมตามขั้นตอนต่อไปนี้

1. กำหนดค่าของเวกเตอร์ค่าเฉลี่ยก่อน (prior mean)

$$\underset{\sim}{\overline{\beta}} = \underset{\sim}{\beta} \pm z\sqrt{\sigma^2}\,\underset{\sim}{a}$$

เมื่อ $\underset{\sim}{\overline{\beta}}$ คือ เวกเตอร์ค่าเฉลี่ยก่อน

$\underset{\sim}{\beta}$ คือ เวกเตอร์สัมประสิทธิ์การถดถอย

$\underset{\sim}{a}$ คือ เวกเตอร์ของค่ารากที่สองของค่าในแนวเส้นทแยงมุมของเมทริกซ์ A^{-1}

และ $\sigma^2 A^{-1}$ คือ เมทริกซ์ความแปรปรวนร่วมก่อน

และกำหนดค่าในเมทริกซ์ความแปรปรวนร่วมก่อน (prior covariance)

2. คำนวณหาค่า Z ที่เหมาะสมโดยกำหนดค่า Z = 0 0.1 0.2 ... จนกระทั่งได้ค่า Z ที่เหมาะสม

3.5 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองและการเปรียบเทียบประสิทธิภาพของตัวประมาณสัมประสิทธิ์การถดถอย

3.5.1 การคำนวณค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ของค่าประมาณสัมประสิทธิ์การถดถอยของแต่ละวิธี โดยกระทำซ้ำ 500 ครั้งในแต่ละสถานการณ์ มีดังนี้

$$MSE(LS)_i = \left\{(\hat{\beta}_{0Li} - \beta_0)^2 + (\hat{\beta}_{1Li} - \beta_1)^2\right\}/2$$
$$MSE(UNI)_i = \left\{(\hat{\beta}_{0Ui} - \beta_0)^2 + (\hat{\beta}_{1Ui} - \beta_1)^2\right\}/2$$
$$MSE(NOR)_i = \left\{(\hat{\beta}_{0Ni} - \beta_0)^2 + (\hat{\beta}_{1Ni} - \beta_1)^2\right\}/2$$

เมื่อ β_0 , β_1 คือ ค่าพารามิเตอร์สัมประสิทธิ์การถดถอย

$\hat{\beta}_{0Li}, \hat{\beta}_{1Li}$ คือ ค่าประมาณ สัมประสิทธิ์การถดถอยวิธีกำลังสองน้อยสุด

$\hat{\beta}_{0Ui}, \hat{\beta}_{1Ui}$ คือ ค่าประมาณสัมประสิทธิ์การถดถอยวิธีเชิงเบส์ที่ใช้การแจกแจงก่อนที่ไม่ให้ข้อมูล และ $\hat{\beta}_{0Ni}, \hat{\beta}_{1Ni}$ คือ ค่าประมาณสัมประสิทธิ์การถดถอยวิธีเชิงเบส์ที่ใช้การแจกแจงก่อนที่ให้ข้อมูล

3.5.2 การคำนวณค่าเฉลี่ยของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (AMSE) มีดังนี้

$$AMSE(LS) = \sum_{i=1}^{500} MSE(LS)_i \Big/ 500$$
$$AMSE(UNI) = \sum_{i=1}^{500} MSE(UNI)_i \Big/ 500$$
$$AMSE(NOR) = \sum_{i=1}^{500} MSE(NOR)_i \Big/ 500$$

เมื่อ $MSE(LS)_i$ คือ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณกำลังสองน้อยสุด

$MSE(UNI)_i$ คือ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณเบส์ที่ใช้การแจกแจงก่อนที่ไม่ให้ข้อมูล

และ $MSE(NOR)_i$ คือ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณเบส์ที่ใช้การแจกแจงก่อนที่ให้ข้อมูล

3.5.3 การคำนวณเปอร์เซ็นต์ของอัตราส่วนผลต่างค่าเฉลี่ยความคลาดเคลื่อนกำลังสองซึ่งเป็นเกณฑ์ที่ใช้ในการเปรียบเทียบวิธีการแต่ละวิธี โดยการเปรียบเทียบกับวิธีที่ให้ AMSE ต่ำสุด มีดังนี้

$$DIFF = \frac{AMSE_i - AMSE_{\min}}{AMSE_{\min}} \times 100, \quad i = 1,2,3$$

เมื่อ $DIFF$ คือ เปอร์เซ็นต์ของอัตราส่วนผลต่างค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง

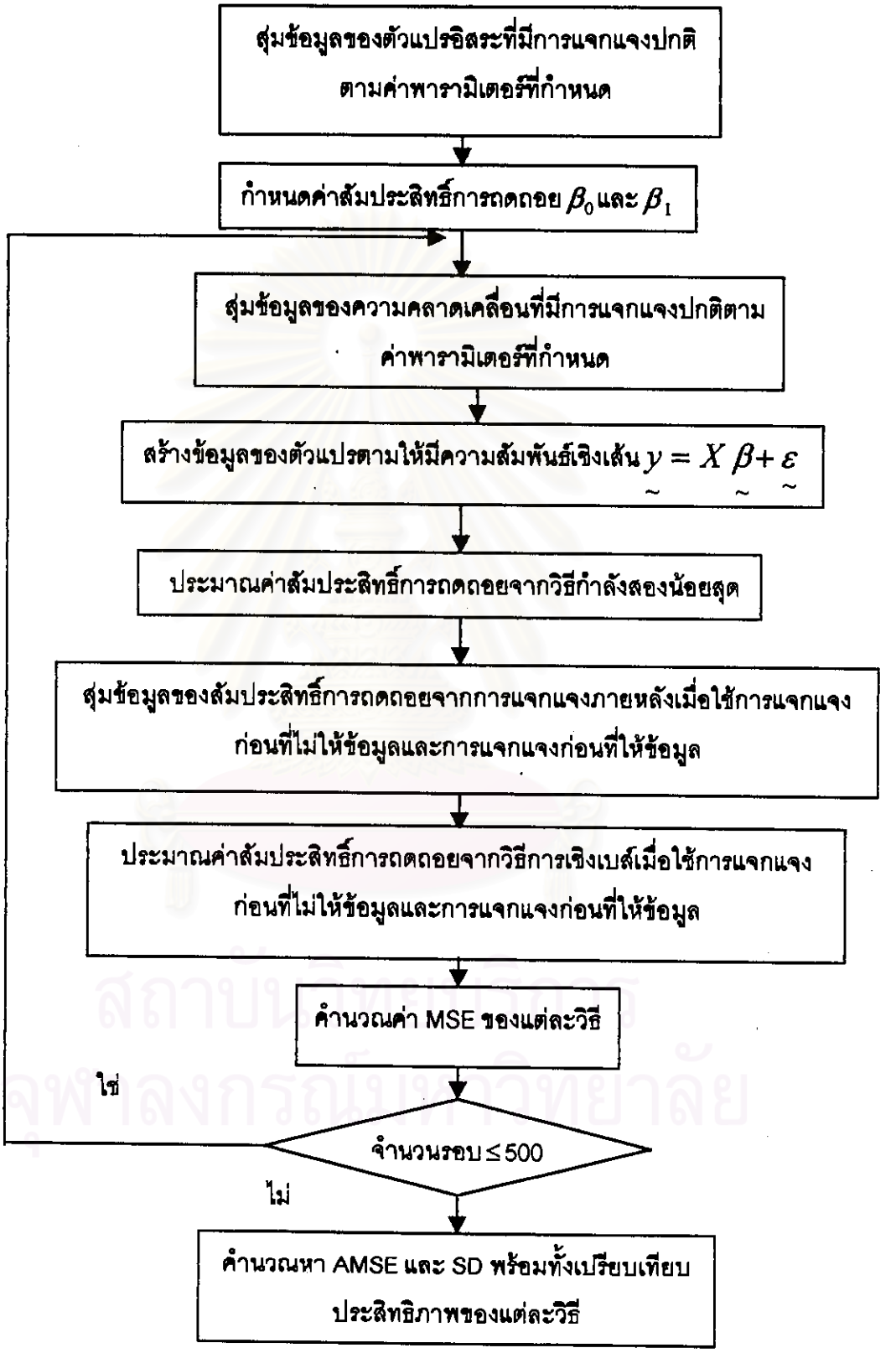
$AMSE_i$ คือ $AMSE$ ของแต่ละวิธี

และ $AMSE_{\min}$ คือ $AMSE$ ของวิธีที่ให้ค่า $AMSE$ ต่ำสุด

ผู้วิจัยแสดงแผนผังการเขียนโปรแกรมในการหาค่าประมาณสัมประสิทธิ์การถดถอยและค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง และแสดงตารางรายละเอียดของโปรแกรมที่ใช้ในงานวิจัยนี้ ดังนี้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

<u>แผนผังที่ 1</u> การเขียนโปรแกรมในการหาค่าประมาณสัมประสิทธิ์การถดถอยและค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของแต่ละวิธี



รายละเอียดของโปรแกรมที่ใช้ในงานวิจัยนี้

ชื่อโปรแกรม	คุณสมบัติของโปรแกรม	โปรแกรมที่เรียกใช้
<u>โปรแกรมหลัก</u>		
MAIN	1. อ่านค่าพารามิเตอร์ที่กำหนด 2. คำนวณค่าประมาณสัมประสิทธิ์การถดถอยของแต่ละวิธี 3. คำนวณค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของแต่ละวิธี 4. คำนวณและเปรียบเทียบค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของแต่ละวิธี	XNOR,CALXTX,INVRS,BET, ERR,BUILDY,UNI,NOR, SSE
<u>โปรแกรมย่อย</u>		
VARBIAS	คำนวณค่าความแปรปรวนและความเอนเอียงของตัวประมาณเบส์เมื่อใช้การแจกแจงก่อนที่ให้ข้อมูล	
BET	สร้างเวกเตอร์ของสัมประสิทธิ์การถดถอย	
ERR	สร้างเวกเตอร์ของความคลาดเคลื่อน	NORMAL
XNOR	สร้างเมทริกซ์ของตัวแปรอิสระ	NORMAL
BUILDY	คำนวณค่าตัวแปรตาม	
OLS	คำนวณค่าประมาณสัมประสิทธิ์การถดถอยด้วยวิธีกำลังสองน้อยสุด	INVRS
UNI	คำนวณค่าประมาณสัมปรสิทธิ์การถดถอยด้วยวิธีเชิงเบส์เมื่อใช้การแจกแจงก่อนที่ไม่ให้ข้อมูล	NORMAL,TDIST
NOR	คำนวณค่าประมาณสัมปรสิทธิ์การถดถอยด้วยวิธีเชิงเบส์เมื่อใช้การแจกแจงก่อนที่ให้ข้อมูล	INVRS,NORMAL,TDIST
CALXTX	คำนวณค่าของเมทริกซ์ XX	
INVRS	คำนวณค่าของเมทริกซ์ผกผัน	

รายละเอียดของโปรแกรมที่ใช้ในงานวิจัยนี้ (ต่อ)

ชื่อโปรแกรม	คุณสมบัติของโปรแกรม	โปรแกรมที่เรียกใช้
<u>โปรแกรมย่อย</u>		
SSE	คำนวณค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของแต่ละวิธี	
<u>ฟังก์ชัน</u>		
RAND	สร้างเลขสุ่มที่มีการแจกแจงสม่ำเสมอ	
NORMAL	สร้างเลขสุ่มที่มีการแจกแจงปกติ	RAND
TDIST	สร้างเลขสุ่มที่มีการแจกแจงที	CHISQ
CHISQ	สร้างเลขสุ่มที่มีการแจกแจงไคกำลังสอง	RAND
SD	คำนวณค่าส่วนเบี่ยงเบนมาตรฐาน	

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย