บทที่ 3

การดำเนินการวิจัย

3.1 ขั้นตอนการดำเนินการวิจัย

ผู้วิจัยได้ใช้เทคนิคการจำลองข้อมูลแบบมอนติคาร์โล (Monte Carlo Simulation Technique) โดยใช้โปรแกรม MATLAB เพื่อสร้างข้อมูลของตัวแปรตามและตัวแปรอิสระตามขอบเขตงานวิจัย กระทำซ้ำ 1,000 ครั้ง ในแต่ละสถานการณ์ ซึ่งมีขั้นตอนการดำเนินงานดังนี้

ตัวแบบสมการถดถอยเชิงเส้นพหุ คือ $V = U\beta + \varepsilon$

1. กำหนดขนาดตัวอย่าง (n) เป็น 10, 20, 30, 50, 100

2. กำหนดสัมประสิทธิ์การถดถอยเชิงเส้นพหุ (β)

$$\beta = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 5 \\ 5 \\ 5 \end{bmatrix}, \begin{bmatrix} 5 \\ 1 \\ 1 \\ 1 \\ 5 \\ 5 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 5 \\ 1 \\ 1 \\ 5 \\ 5 \\ 5 \end{bmatrix}, \begin{bmatrix} 5 \\ 5 \\ 5 \\ 5 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 5 \\ 5 \\ 5 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 \\ 1 \\ 5 \\ 5 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

3. จำลองค่าความคลาดเคลื่อนสุ่ม (ε_i) ให้มีการแจกแจงปกติที่มีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น 0.25, 1, 9 และ 25

4. สร้างตัวแปรอิสระให้มีความสัมพันธ์เชิงเส้นต่อกัน

การวิจัยครั้งนี้ได้ศึกษาตัวแบบสมการถดถอยเชิงเส้นพหุที่มีจำนวนตัวแปรอิสระ p = 6 ตัวแปร ขนาดตัวอย่างเท่ากับ n คือ U_{ij} (i=1,2,....,n) , (j=1,2,...,6) โดยเริ่มต้นจากการสร้างตัวแปรอิสระ N_{ij} ที่มีการแจกแจงแบบปกติมาตรฐาน จากตัวแปรอิสระ N_{ij} นำมาสร้างตัวแปรอิสระ U_{ij} ให้มีสหสัมพันธ์ตามวิธีจำลองข้อมูลของ Wichern และ Churchill

$$U_{ij} = \begin{cases} (1-\rho^2)^{1/2} N_{ij} + \rho N_{i7} & i = 1,2,....,n \quad j = 1,2,3 \\ N_{ij} & i = 1,2,.....,n \quad j = 4,5,6 \end{cases}$$

โดยที่ U_{ij} เป็นค่าของตัวแปรอิสระของหน่วยตัวอย่างที่ i ของตัวแปรอิสระตัวที่ j
ซึ่งเราสามารถพิสูจน์ได้ว่า เมตริกซ์ของสหสัมพันธ์ระหว่างตัวแปรอิสระคือ

$$P = \begin{bmatrix} 1 & \rho & \rho & 0 & 0 & 0 \\ \rho & 1 & \rho & 0 & 0 & 0 \\ \rho & \rho & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$\rho(X_i, X_j) = 1 \quad ;\ i = j = 1,2,3,4,5,6$

$\rho(X_i, X_j) = \rho \quad ;\ i \neq j = 1,2,3$

$\rho(X_i, X_j) = 0 \quad ;\ i = 1,2,3 \quad j = 4,5,6$

5. สร้างตัวแปรตาม V จากตัวแบบสมการถดถอยเชิงเส้นพหุ

$$V_i = \beta_0 + \beta_1 U_{i1} + \beta_2 U_{i2} + \beta_3 U_{i3} + \beta_4 U_{i4} + \beta_5 U_{i5} + \beta_6 U_{i6} + \varepsilon_i \qquad (i=1,2,....,n)$$

6. แปลงตัวแปรอิสระ U ที่ได้จากข้อ 1

$$X_{ij} = U_{ij} - \bar{U}_j$$

7. แปลงตัวแปรตาม V

$$Y_i = V_i - \bar{V}$$

แปลง V และ U ทำให้ตัวแบบการถดถอยเชิงเส้นพหุที่ใช้ในการศึกษาครั้งนี้มีรูปแบบดังนี้

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \delta_i$$

8. ประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุด้วยวิธีการต่าง ๆ ดังนี้

8.1 วิธีการถดถอยองค์ประกอบหลัก ตามขั้นตอนดังนี้

- คำนวณค่า eigenvalue ของ $X'X$ คือ $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6$
- คำนวณค่า eigenvector ของ $X'X$
- พิจารณาจำนวนองค์ประกอบหลัก โดยวิธีที่จะนำมาใช้ในการพิจารณาจำนวนองค์ประกอบหลักมี 4 วิธี คือ

1.วิธี PCA: ใช้จำนวนองค์ประกอบหลักที่มีค่าไอเกนมากกว่าหรือเท่ากับค่าเฉลี่ย

$$\bar{\lambda} = \frac{1}{p}\sum_{j=1}^{p}\lambda_j$$

2.วิธี PC3: ใช้จำนวนองค์ประกอบหลัก 3 องค์ประกอบที่มีค่าไอเกนสูงสุด 3 อันดับแรก

3.วิธี PC4: ใช้จำนวนองค์ประกอบหลัก 4 องค์ประกอบที่มีค่าไอเกนสูงสุด 4 อันดับแรก

4.วิธี PC5: ใช้จำนวนองค์ประกอบหลัก 5 องค์ประกอบที่มีค่าไอเกนสูงสุด 5 อันดับแรก

- คำนวณค่า $\hat{\beta}_{PC}$ ตามสมการ $\hat{\beta}_{PC} = T_{p-s}(\Lambda^{-1}_{p-s} T'_{p-s} X'Y)$

เมื่อ s คือจำนวนองค์ประกอบที่ตัดออกโดยทำทั้งกรณี PCA PC3 PC4 และ PC5

8.2. วิธีประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุด้วยวิธีริดจ์ รีเกรสชันที่มีค่าเบื้องต้น$\left(\hat{\beta}_{RJ}\right)$ ตามขั้นตอนดังนี้

- คำนวณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุด้วยวิธีกำลังสองน้อยที่สุด

$$\hat{\beta}_{LS} = (X'X)^{-1}X'Y \text{ และความแปรปรวน } \hat{\sigma}^2 = \frac{Y'Y - \hat{\beta}_{LS}'X'Y}{n-p}$$

- ให้ $J = \left[\sum_{j=1}^{p} \frac{\hat{\beta}_{LS_j}}{p} \right] 1_{p \times 1}$ โดยที่ $1_{p \times 1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{p \times 1}$

$$I = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & \cdots & 1 \end{bmatrix}_{p \times p}$$

- คำนวณค่าประมาณของ

$$k = \begin{cases} \dfrac{p\hat{\sigma}^2}{(\hat{\beta}_{LS} - J)'(\hat{\beta}_{LS} - J) - \hat{\sigma}^2 tr(X'X)^{-1}} & \text{if} \quad (\hat{\beta}_{LS} - J)'(\hat{\beta}_{LS} - J) - \hat{\sigma}^2 tr(X'X)^{-1} > 0 \\ \dfrac{p\hat{\sigma}^2}{(\hat{\beta}_{LS} - J)'(\hat{\beta}_{LS} - J)} & \text{other} \end{cases}$$

- คำนวณค่า $\hat{\beta}_{RJ}$ ตามสมการ $\hat{\beta}_{RJ} = (X'X + \hat{k}I)^{-1}(X'Y + \hat{k}J)$

8.3 ประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุด้วยวิธีมูนิซและไคเบรียริดจ์รีเกรสชัน ตามวิธีการดังนี้

- คำนวณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุด้วยวิธีกำลังสองน้อยที่สุด

$\hat{\alpha} = (Z'Z)^{-1}Z'Y$ และความแปรปรวน $\hat{\sigma}^2 = \dfrac{Y'Y - \hat{\alpha}'Z'Y}{n-p}$

- คำนวณค่า t_{max}

- คำนวณค่า $k_M = \underset{1\le j\le 6}{\text{median}}\left(\dfrac{1}{\sqrt{\dfrac{t_{max}\hat{\sigma}^2}{(n-p)\hat{\sigma}^2 + t_{max}\hat{\alpha}_j^2}}}\right)$

- คำนวณค่า $\hat{\alpha}_{RM} = (Z'Z + k_M I)^{-1} Z'Y$

- คำนวณค่า $\hat{\beta}_{RM} = T\hat{\alpha}_{RM}$

9. จากสมการ $Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6$ ทำการแปลงกลับให้อยู่ในรูป $V = \beta_0 + \beta_1 U_1 + \beta_2 U_2 + \beta_3 U_3 + \beta_4 U_4 + \beta_5 U_5 + \beta_6 U_6$

10. คำนวณค่า $AMSE(\hat{\beta})$ ของทั้ง 6 วิธี คือ PCA PC3 PC4 PC5 RJ RM และเปรียบเทียบค่า $AMSE(\hat{\beta})$ ของทั้ง 6 วิธี โดยเลือกวิธีที่ให้ $AMSE(\hat{\beta})$ ต่ำสุดในแต่ละกรณี

$$MSE(\hat{\beta}_j) = \frac{1}{1000}\sum_{a=1}^{1000}(\hat{\beta}_{aj} - \beta_j)^2$$

$$AMSE(\hat{\beta}) = \frac{1}{7}\sum_{j=0}^{6} MSE(\hat{\beta}_j)$$

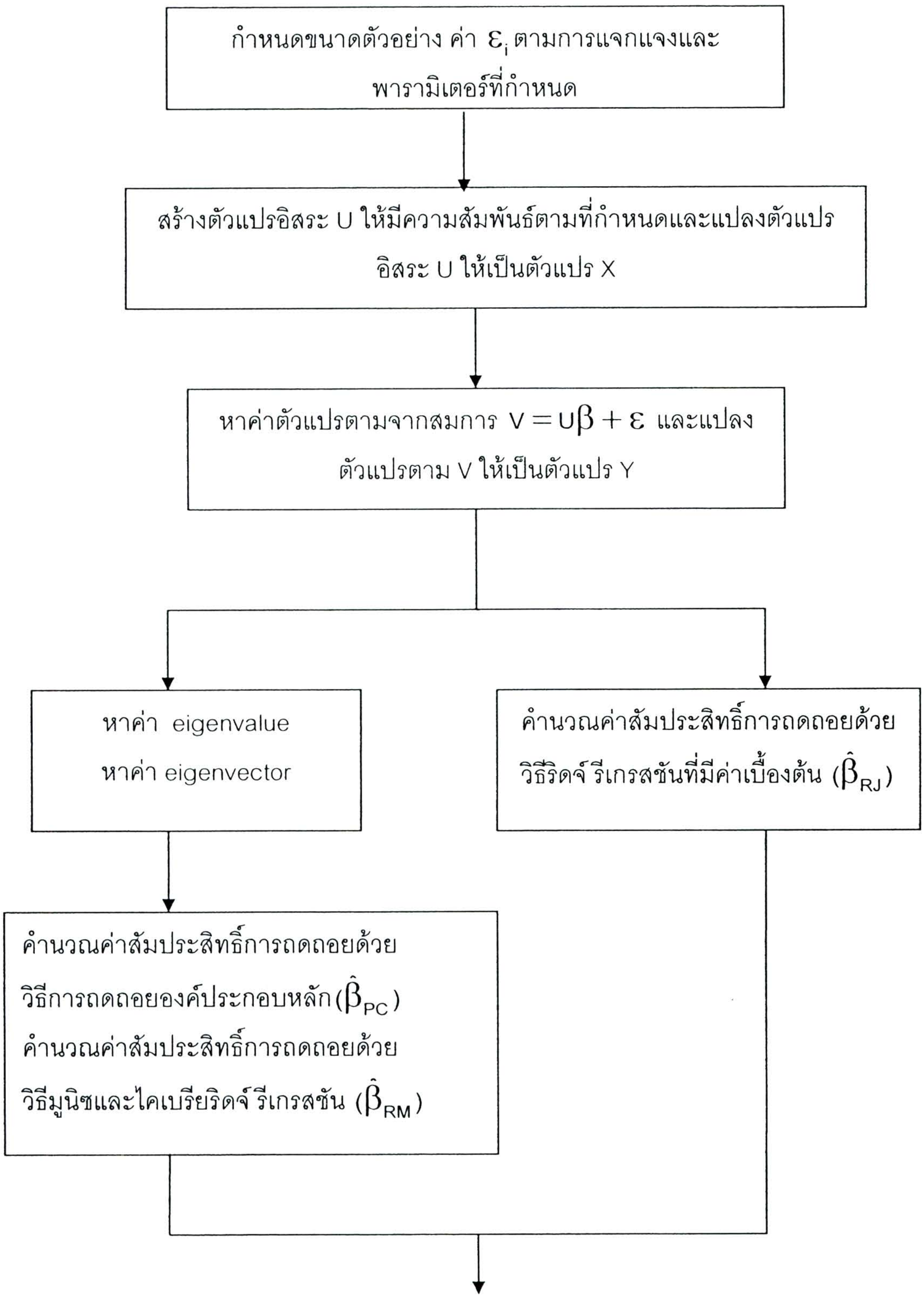
11.คำนวณค่า $ABIAS(\hat{\beta})$ ของทั้ง 6 วิธี โดยเลือกวิธีที่ให้ $ABIAS(\hat{\beta})$ ต่ำสุดในแต่ละกรณี

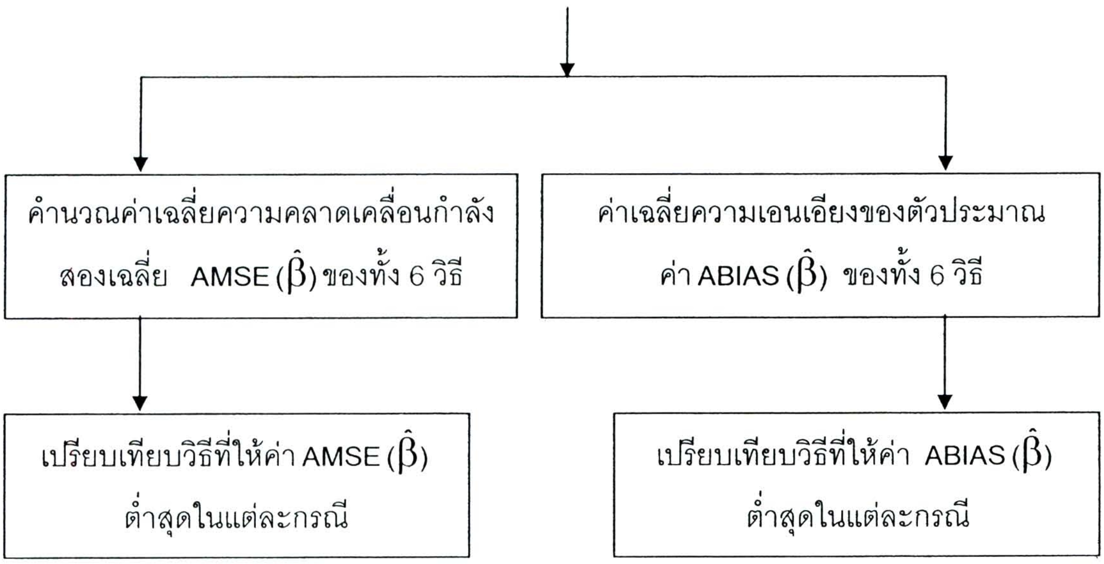
$$BIAS(\hat{\beta}_j) = \frac{1}{1000}\sum_{a=1}^{1000}(\hat{\beta}_{aj} - \beta_j)$$

$$ABIAS(\hat{\beta}) = \frac{1}{7}\sum_{j=0}^{6}\left|BIAS(\hat{\beta}_j)\right|$$

12. สร้างแฟ้มข้อมูลเพื่อเตรียมการประมวลผลเปรียบเทียบระหว่างวิธีการด้วยโปรแกรม SPSS

แผนภาพที่ 3.1 **ขั้นตอนการดำเนินการวิจัย**




คำนวณค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ย AMSE ($\hat{\beta}$) ของทั้ง 6 วิธี
ค่าเฉลี่ยความเอนเอียงของตัวประมาณค่า ABIAS ($\hat{\beta}$) ของทั้ง 6 วิธี
เปรียบเทียบวิธีที่ให้ค่า AMSE ($\hat{\beta}$) ต่ำสุดในแต่ละกรณี
เปรียบเทียบวิธีที่ให้ค่า ABIAS ($\hat{\beta}$) ต่ำสุดในแต่ละกรณี