

บทที่ 3

วิธีดำเนินการวิจัย

การดำเนินการวิจัยนี้ เป็นการวิจัยเพื่อเปรียบเทียบประสิทธิภาพของการประมาณพารามิเตอร์ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ด้วยวิธีการประมาณพารามิเตอร์แบบ Quasi-Likelihood สำหรับข้อมูลระยะยาว เมื่อกำหนดให้ตัวแปรตามมีการแจกแจงปัวซอง โดยศึกษาเปรียบเทียบค่าเฉลี่ยของผลบวกกำลังสองของความคลาดเคลื่อนของสัมประสิทธิ์ความถดถอยและสำหรับการเปรียบเทียบจะกระทำภายใต้สถานการณ์ที่กำหนด ด้วยโปรแกรม R ดังนี้

3.1 แผนการดำเนินงานวิจัย

ในการวิจัยครั้งนี้กำหนดสถานการณ์ต่างๆ ไว้ดังนี้

1.1.1 กำหนดจำนวนตัวแปรอิสระที่นำมาศึกษาเท่ากับ 1, 3

1.1.2 กำหนดขนาดตัวอย่างที่ศึกษาเท่ากับ 20, 60

3.1.3 กำหนดให้อัตตสัมพันธ์และโครงสร้างความแปรปรวนร่วมของตัวแบบ

Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations คือ Exchangeable โดยกำหนดให้อัตตสัมพันธ์ของตัวแปรตามที่ 0.3 และ 0.8

3.1.4 กำหนดรูปแบบความสัมพันธ์ระหว่างตัวแปรอิสระด้วยกัน คือ

3.1.4.1 กรณีมีความสัมพันธ์ระดับต่ำ ($\rho = 0.1$)

3.1.4.2 กรณีมีความสัมพันธ์ระดับปานกลาง ($\rho = 0.5$)

3.1.4.3 กรณีมีความสัมพันธ์ระดับสูง ($\rho = 0.9$)

3.1.5 กำหนดระยะเวลาที่ทำการเก็บข้อมูลซ้ำ (t) เท่ากับ 3, 6

3.1.6 กำหนดค่าพารามิเตอร์ $\beta_0 = -1, \beta_{1t} = 0.01, \beta_{21} = 0.02,$

$\beta_{22} = 0.04, \beta_{23} = 0.01$ เนื่องจากกำหนดให้อัตตสัมพันธ์ของตัวแปรตามแบบ Exchangeable ค่าพารามิเตอร์ที่กำหนดจึงต้องเป็นค่าที่ทำให้เกิดอัตตสัมพันธ์ของตัวแปรตามเป็นแบบ Exchangeable

3.1.7 กำหนดรอบประมวลผลในแต่ละสถานการณ์ 1000 รอบ

3.1.8 เปรียบเทียบตัวแบบด้วยค่าเฉลี่ยของผลบวกกำลังสองของความคลาดเคลื่อนของสัมประสิทธิ์ความถดถอย (AMSE)

3.2 ขั้นตอนในการดำเนินการวิจัย

มีขั้นตอนในการดำเนินการวิจัยดังนี้

- 3.2.1 การสร้างข้อมูลตัวแปรตาม (y) ที่มีการแจกแจงแบบปัวซองส์
- 3.2.2 การสร้างข้อมูลตัวแปรอิสระ (x) ที่มีการแจกแจงแบบปกติ
- 3.2.3 สร้างอัตตสัมพันธ์ของตัวแปรตามของตัวแบบ Generalized Estimating Equations แบบ Exchangeable
- 3.2.4 สร้างโครงสร้างความแปรปรวนร่วมตัวแปรตามของแบบ Generalized Linear Model แบบ Exchangeable โดยใช้ข้อมูลชุดเดียวกัน
- 3.2.5 ทำการประมาณพารามิเตอร์ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ด้วยวิธี Quasi- Likelihood
- 3.3.6 เปรียบเทียบค่าเฉลี่ยของผลบวกกำลังสองของความคลาดเคลื่อนของสัมประสิทธิ์ความถดถอย (AMSE) ของทั้งสองตัวแบบ ว่าตัวแบบใดให้ค่านี้ต่ำกว่ากัน แล้วดูร้อยละความแตกต่างของทั้งสองตัวแบบว่าแตกต่างเท่าไร และสรุปผลในรูปของตารางและกราฟในแต่ละกรณี การประมาณพารามิเตอร์ของตัวแบบใดให้ค่าเฉลี่ยของผลบวกกำลังสองของความคลาดเคลื่อนของสัมประสิทธิ์ความถดถอยต่ำกว่า แสดงว่าตัวแบบนั้นให้การประมาณพารามิเตอร์ที่ดีกว่า

3.3 การสร้างข้อมูลตัวแปรตาม (y)

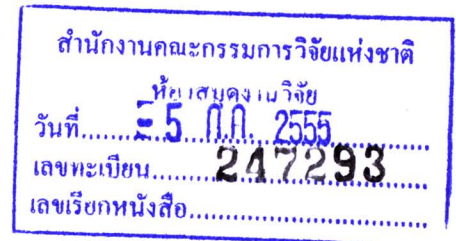
การสร้างข้อมูลตัวแปรตาม (Y) ทำได้ด้วยการสุ่มตัวอย่างค่าตัวแปรตามให้มีการแจกแจงแบบปัวซองส์ ด้วยโปรแกรม R โดยกำหนดให้มีพารามิเตอร์ (λ) เท่ากับ 0.4 ซึ่งเป็นค่าพารามิเตอร์ที่ทำให้ความแปรปรวนคงที่ จะใช้คำสั่ง `rpois()` ในการสร้างข้อมูล เมื่อมีขนาดตัวอย่าง n เท่ากับ 20 และ 60 และทำซ้ำช่วงเวลา t เท่ากับ 3 และ 6 คาบเวลา จำนวนวน 1,000 รอบ โดยสร้างตัวแปรตามให้อัตตสัมพันธ์กันที่ 0.3 และ 0.8 ให้อยู่ในรูปแบบของ Exchangeable คือ $Corr(y_{it}, y_{iw}) = \rho; t \neq w$ สามารถเขียนใน

รูปเมทริกซ์ คือ
$$\begin{pmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \rho & \rho & 1 & \dots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{pmatrix}$$
 โดยที่ t และ w คือช่วงเวลาที่ต่างกัน และ



โครงสร้างความแปรปรวนร่วมของตัวแปรตามให้อยู่ในรูปแบบของ Exchangeable คือ

$$R_i = \text{Var}(y_{it}) = \begin{pmatrix} \sigma^2 & \sigma_1 & \sigma_1 & \dots & \sigma_1 \\ \sigma_1 & \sigma^2 & \sigma_1 & \dots & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 & \dots & \sigma_1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_1 & \sigma_1 & \sigma_1 & \dots & \sigma^2 \end{pmatrix}$$



ซึ่งความแปรปรวนร่วมของตัวแปรตาม จะแปรผันกับอัตตสัมพันธ์ของตัวแปรตาม

3.4 การสร้างข้อมูลตัวแปรอิสระ (X)

การสร้างข้อมูลตัวแปรอิสระ (X) จะถูกกำหนดให้มีการแจกแจงปกติ คือ Normal(mean, var) โดยที่กำหนดให้ mean = 0 และ var = 4 ด้วยโปรแกรม R โดยใช้คำสั่ง rnorm() โดยจำนวนตัวแปรอิสระ (X) จะที่นำมาศึกษา คือ 1 และ 3 ให้ตัวแปรอิสระมีระดับสัมพันธ์กันคือ มีความสัมพันธ์กันน้อย มีความสัมพันธ์กันปานกลาง และมีความสัมพันธ์กันมาก จะสร้างความสัมพันธ์ของตัวแปรอิสระให้อยู่ในรูป

$$\begin{pmatrix} 1 & \rho & \rho \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \rho & \rho & 1 & \dots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{pmatrix}$$

ในการศึกษาครั้งนี้เริ่มต้นสร้างตัวแปรตาม (Y) และตัวแปรอิสระ (X) เนื่องจากมีเงื่อนไขของอัตตสัมพันธ์ของตัวแปรตามเป็นแบบ Exchangeable ซึ่งเกิดขึ้นได้ยากเพราะเป็นความสัมพันธ์ของค่าสังเกตภายในหน่วยศึกษาเดียวกันจะมีค่าเท่ากัน ณ ช่วงเวลาต่างกัน ตัวแปรตามมีการแจกแจงแบบปัวซองส์ และตัวแปรอิสระมีความสัมพันธ์กัน

3.5 การประมาณพารามิเตอร์ด้วยวิธี Quasi-Likelihood

ในการประมาณค่าพารามิเตอร์วิธี Quasi-likelihood ในตัวแบบ Generalized Estimating Equations จะใช้ package geepack ในโปรแกรม R เรียกใช้คำสั่ง `geeglm()` และ ในตัวแบบ Generalized Linear Model เรียกใช้คำสั่ง `glmFamily() = quasipoisson` โดยกำหนด เมทริกซ์ X เมทริกซ์ Y และ ค่าพารามิเตอร์ ตามเงื่อนไขในหัวข้อ 3.1 และกำหนดค่าฟังก์ชันเชื่อมโยง (Link) เป็น log การสร้างพารามิเตอร์จะทำการวนซ้ำ (Iteration) 1,000 รอบ โดยค่าพารามิเตอร์เบต้าจะหาได้จาก

$$U(\beta) = \sum D^T V^{-1} (Y - \mu) = 0$$

ซึ่งเรียกว่า quasi - score function โดยที่

$$D_i = \frac{\partial \mu_{ii}}{\partial \beta_i}$$

$$V_i = (A_i^{1/2} R_i A_i^{1/2}) \phi$$

$$\phi = \frac{1}{n-p} \sum_i \sum_i \frac{y_{ii} - \mu_{ii}}{\sqrt{\text{var}(\mu_{ii})}}$$

เมื่อ V_i คือ เมทริกซ์โครงสร้างความแปรปรวนร่วมของ Y_{ii}

A_i คือ diagonal matrix ความแปรปรวนของ Y_{ii}

R_i คือ เมทริกซ์อัตโนมัติสัมพันธ์ของ Y_{ii}

ϕ คือ overdispersion parameter = 1

3.6 เปรียบเทียบค่าเฉลี่ยของผลบวกกำลังสองของความคลาดเคลื่อนของพารามิเตอร์ (AMSE)

คำนวณหาค่าเฉลี่ยของผลบวกกำลังสองของความคลาดเคลื่อนของสัมประสิทธิ์

ความถดถอย (Average Mean Square Error : AMSE) จากการประมวลผล 1,000 รอบของพารามิเตอร์ที่ได้จากการประมาณของทั้ง 2 ตัวแบบ

$$AMSE = \frac{\sum_{k=1}^{1000} \sum_{m=0}^M (\hat{\beta}_{mk} - \beta_{mk})^2}{1000M} \quad \text{โดยที่ } m=1,2,\dots,M \quad k=1,2,\dots,1000$$

$\hat{\beta}_{mk}$ คือ ค่าประมาณของพารามิเตอร์ตัวที่ m ครั้งที่ k

β_{mk} คือ ค่าจริงของพารามิเตอร์ตัวที่ m ครั้งที่ k

M คือ จำนวนของพารามิเตอร์ที่มีอยู่ในสมการ

3.7 ร้อยละความแตกต่างของ 2 ตัวแบบ

เป็นการหาร้อยละของความแตกต่างของค่า AMSE ของตัวแบบ Generalized Linear Model และตัวแบบ Generalized Estimating Equations เพื่อดูร้อยละความแตกต่างของค่า AMSE ของทั้งสองตัวแบบ ซึ่งคำนวณได้ดังนี้

$$\frac{Max(AMSE) - Min(AMSE)}{Min(AMSE)} \times 100$$

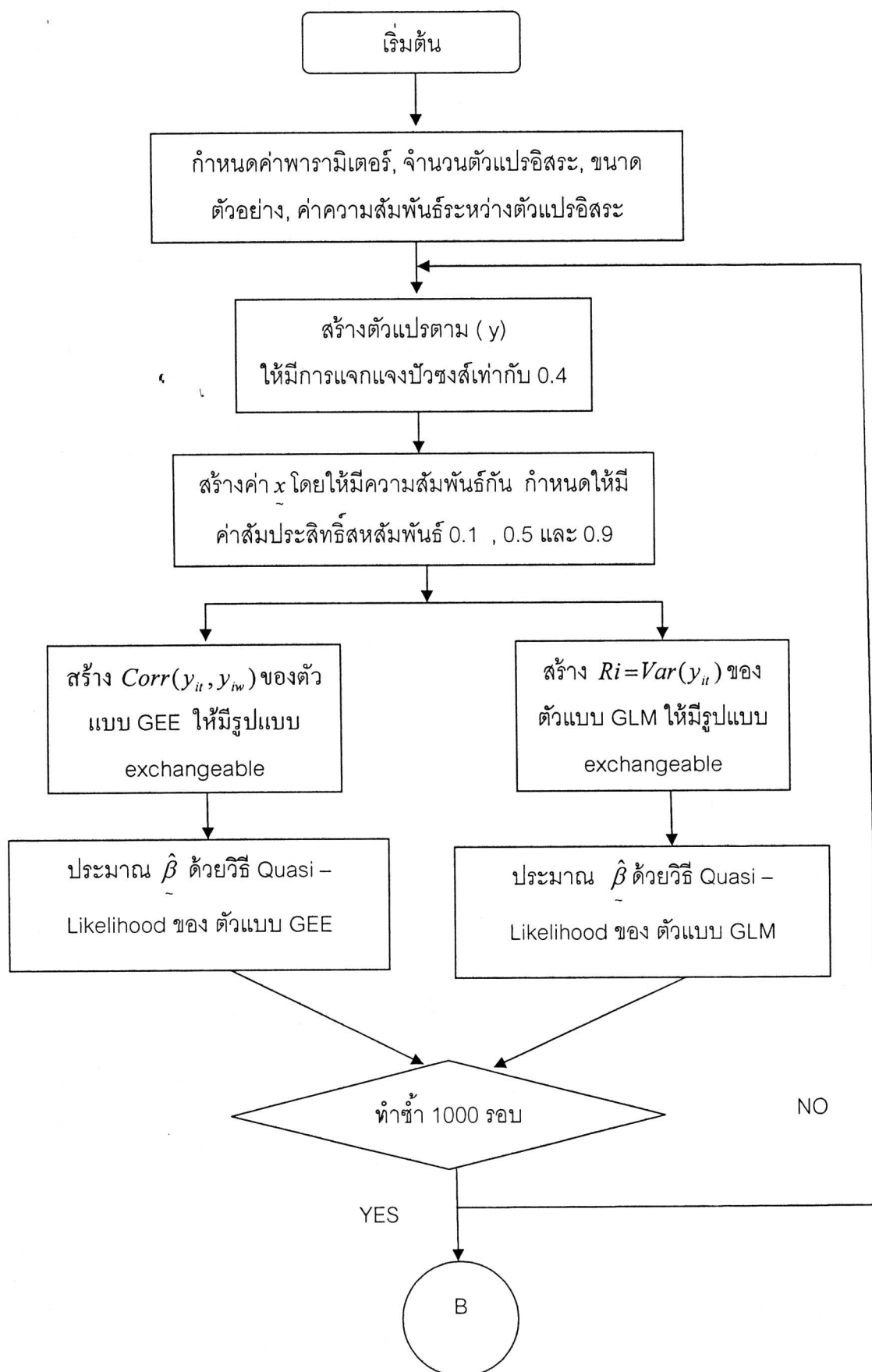
โดย $Max(AMSE)$ คือ ตัวแบบที่ให้ค่า AMSE สูง เมื่อทำการเปรียบเทียบสองตัวแบบ

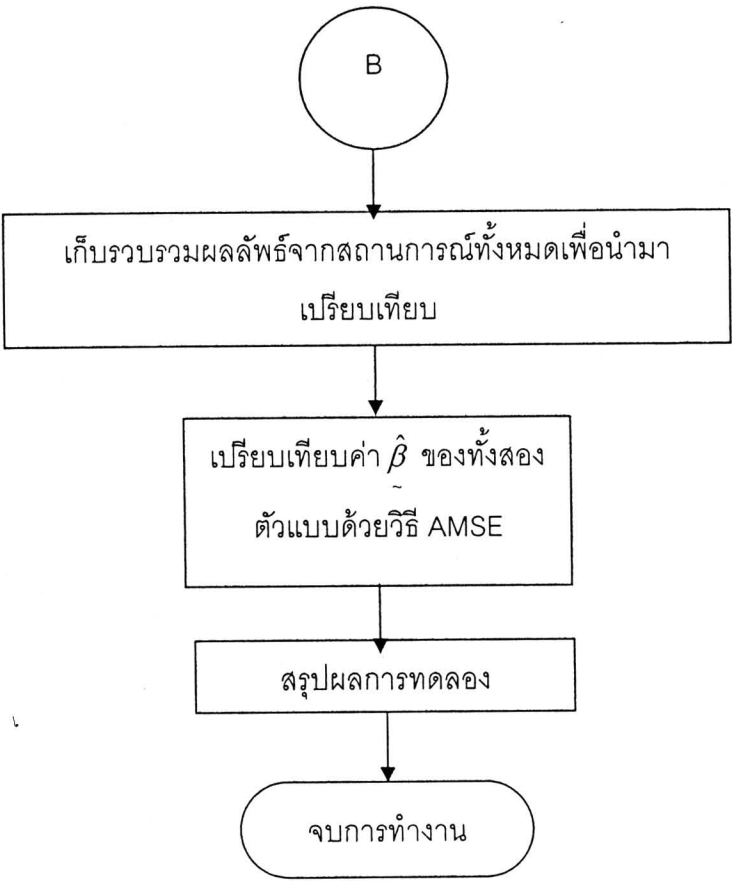
$Min(AMSE)$ คือ ตัวแบบที่ให้ค่า AMSE ต่ำ เมื่อทำการเปรียบเทียบสองตัวแบบ

3.8 สรุปผลการวิจัยในแต่ละสถานการณ์

ทำการเปรียบเทียบทุกกรณีจากการประมาณพารามิเตอร์ของทั้ง 2 ตัวแบบ ว่าตัวแบบใดให้ค่าเฉลี่ยของผลบวกกำลังสองของความคลาดเคลื่อนของสัมประสิทธิ์ความถดถอยต่ำกว่าแสดงว่าตัวแบบนั้นให้การประมาณพารามิเตอร์ที่เหมาะสมกว่า

3.9 ขั้นตอนการทำงานของโปรแกรม





รูปที่ 3.1 แสดงขั้นตอนการทำงานของโปรแกรม