บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

แนวคิดและทฤษฎี

การวิจัยครั้งนี้ได้เสนอวิธีการประมาณพารามิเตอร์แบบ Quasi-Likelihood ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations สำหรับข้อมูลระยะยาว เมื่อกำหนดให้ตัวแปรตามมีการแจกแจงปัวซงส์ ในบทนี้จะกล่าวถึงคุณสมบัติของตัวแบบและรายละเอียดของวิธีการประมาณค่าพารามิเตอร์เป็นลำดับต่อไป

2.1 ตัวแบบ Generalized Estimating Equationss (GEE)

ตัวแบบ Generalized Estimating Equations (GEE) เป็นตัวแบบที่ถูกพัฒนาโดย Liang และ Zeger (1986) เพื่อใช้ในการวิเคราะห์การถดถอยซึ่งคำนึงถึงสหสัมพันธ์ภายในตัวเองของตัวแปรตามในการวิจัยข้อมูลระยะยาว ซึ่งตัวแบบ Generalized Estimating Equations (GEE) เหมาะสำหรับวิเคราะห์ข้อมูลระยะยาว หรือ ข้อมูลที่มีการวัดซ้ำ โดยวิธีการนี้ได้พัฒนาจาก Generalized Linear Model ซึ่งเป็นการลดความซับซ้อนในการวิเคราะห์การถดถอยในกรณีที่ตัวแปรตามไม่มีการแจกแจงปกติ และตัวแปรตามมีความสัมพันธ์ในตัวเอง

ในการศึกษานี้ตัวแปรตามมีการแจกแจงแบบปัวซงส์ โดยจะสมมติให้ $y_{i(t)}$ คือ ค่าสังเกตของหน่วยศึกษาที่ i ระยะเวลาที่ t สามารถเขียนตัวแบบ Generalized Estimating Equations สำหรับการวิเคราะห์ Marginal Model คือ

$$\log_e(Y_{i(t)}) = \beta_0 + \beta_1 t + \beta_{2j}\sum_{j=1}^{J} x_{ijt} + corr + e_{i(t)} \quad ;$$

$$i = 1,2,\ldots,n \qquad j = 1,2,\ldots,J \qquad t = 1,2,\ldots,T$$

$Y_{i(t)}$ คือ ตัวแปรตามของหน่วยศึกษาที่ i ระยะเวลาที่ t

x_{ijt} คือ ตัวแปรอิสระที่ j ของหน่วยศึกษาที่ i ระยะเวลาที่ t ที่ขึ้นกับเวลา

T คือ ระยะเวลาที่เก็บข้อมูลซ้ำ

corr คือ อัตตสัมพันธ์ของตัวแปรตาม

n คือ ขนาดตัวอย่าง

J คือ จำนวนตัวแปรอิสระที่ขึ้นกับเวลา

$\varepsilon_{i(t)}$ คือ ความคลาดเคลื่อนของหน่วยที่ i ระยะเวลาที่ t

$\beta_0, \beta_1, \beta_{2j}$ คือ พารามิเตอร์ที่ไม่ทราบค่า (Unknown Parameter)

โดยค่าเฉลี่ยของตัวแปรตาม (μ_{it}) จะมีความสัมพันธ์กับตัวทำนายเชิงเส้นผ่านฟังก์ชันเชื่อม(Link Function) คือ $\eta_{it} = x_{it}'\beta$ ดังนั้น $g(\mu_{it}) = \log_e(\mu_{it}) = \eta_{it} = x_{ijt}'\beta$ สามารถเขียนในรูปแบบเมทริกซ์ได้ดังนี้ $G(Y) = X\underset{\sim}{\beta}$

$$X = \begin{bmatrix} 1 & x_{111} & x_{122} & \cdots & x_{1jt} \\ 1 & x_{211} & x_{222} & \cdots & x_{2jt} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n11} & \cdots & \cdots & x_{ijt} \end{bmatrix}$$

$$\underset{\sim}{\beta} = \begin{bmatrix} \beta_0 \\ \beta_{1t} \\ \beta_{21} \\ \vdots \\ \beta_{2j} \end{bmatrix}_{(j+2)x1}$$

เมื่อ j+2 คือ จำนวนของพารามิเตอร์ , i คือ ขนาดตัวอย่าง , j คือ จำนวนของตัวแปรอิสระ และ t คือ ระยะเวลาที่เก็บข้อมูลซ้ำ

เมื่อ $Var(Y_{it}) = \phi Var(\mu_{it})$ เมื่อ $Var(\mu_{it}) = \mu_{it}$, $\phi = 1$ จะได้ว่า $Var(Y_{it}) = Var(\mu_{it})$

2.1.1 **อัตตสัมพันธ์ของตัวแปรตาม** (Working Correlation)

อัตตสัมพันธ์ของค่าสังเกตของหน่วยศึกษา (y) จะสมมติให้อยู่ในรูปแบบของ Exchangeable

Exchangeable เป็นความสัมพันธ์ของค่าสังเกตภายในหน่วยศึกษาเดียวกันจะมีค่าเท่ากัน ณ ช่วงเวลาที่ต่างกัน ซึ่งความสัมพันธ์จะอยู่ในรูปของ $Corr(y_{it}, y_{im}) = \rho\,; t \neq m$ สามารถเขียนใน

รูปเมทริกซ์ คือ $\begin{pmatrix} 1 & \rho & \rho \ldots & \rho \\ \rho & 1 & \rho \ldots & \rho \\ \rho & \rho & 1 \ldots & \rho \\ \vdots & \vdots & \vdots \ddots & \vdots \\ \rho & \rho & \rho \ldots & 1 \end{pmatrix}$

2.2 ตัวแบบ Generalized Linear Model (GLM)

ในการศึกษาตัวแบบ Generalized Linear Model นี้ จะใช้ ตัวแบบ Poisson Regression ซึ่งตัวแปรตามมีการแจกแจงแบบปัวซงส์ โดยการแจกแจงแบบปัวซงส์นั้นเป็นการแจกแจงปัวซงส์เป็นการแจกแจงที่เกิดขึ้นในตัวแปรสุ่มของการทดลองหรือว่าเหตุการณ์ที่เกิดขึ้นในช่วงระยะเวลาใดเวลาหนึ่งที่ต่อเนื่องกัน โดยหนึ่งหน่อยนั้นอาจจะเป็น วินาที นาที ชั่วโมง วัน สัปดาห์ เดือน หรือ ปี อาจจะเป็นความยาว พื้นที่ ปริมาตร โดยข้อมูลที่ได้จากตัวแปรสุ่มนี้เป็นแบบไม่ต่อเนื่องกันชนิดหนึ่ง สามารถเขียนตัวแบบ ได้ดังนี้

$$\log_e(Y_{i(t)}) = \beta_0 + \beta_1 t + \sum_{j=1}^{J} \beta_{2j} x_{ijt} + \varepsilon_{i(t)} \quad ;$$

$$i = 1,2,\ldots,n \qquad j = 1,2,\ldots,J$$

$$t = 1,2,\ldots,T$$

$Y_{i(t)}$ คือ ตัวแปรตามของหน่วยศึกษาที่ i ระยะเวลาที่ t

x_{ijt} คือ ตัวแปรอิสระที่ j ของหน่วยศึกษาที่ i ระยะเวลาที่ t ที่ขึ้นกับเวลา

T คือ ระยะเวลาที่เก็บข้อมูลซ้ำ

n คือ ขนาดตัวอย่าง

J คือ จำนวนตัวแปรอิสระที่ขึ้นกับเวลา

$\varepsilon_{i(t)}$ คือ ความคลาดเคลื่อนของหน่วยที่ i ระยะเวลาที่ t

$\beta_0, \beta_1, \beta_{2j}$ คือ พารามิเตอร์ที่ไม่ทราบค่า (Unknown Parameter)

โดยมีค่าเฉลี่ยของตัวแปรตาม คือ $E(Y_{it}) = \exp(x'_{ijt}\beta)$ ดังนั้น $\log E(Y_{it}) = x'_{ijt}\beta$ และความแปรปรวนของตัวแปรตาม คือ $Var(Y_{it}) = \mu_{it}$ จะได้ว่า $g(\mu_{it}) = x'_{ijt}\beta$

2.2.1 โครงสร้างความแปรปรวนร่วมของตัวแปรตาม

ในการศึกษานี้เป็นข้อมูลวัดซ้ำจะสมมติให้โครงสร้างความแปรปรวนร่วมให้อยู่ในรูปแบบของ Exchangeable

Exchangeable ความแปรปรวนร่วมระหว่างค่าสังเกต (σ_1)จะเท่ากันสำหรับบุคคลเดียวกัน สามารถเขียนในรูปเมทริกซ์ คือ

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

2.3. Quasi – Likelihood

วิธี Quasi – Likelihood ถูกพัฒนาจาก Weddderburn (1974) เป็นวิธีที่มีคุณสมบัติเหมือนกับวิธี Maximum Likelihood แต่ Full Likelihood ไม่ต้องการเจาะจง เป็นวิธีทการประมาณพารามิเตอร์ที่ใช้สำหรับการแจกแจงที่มีสมมติฐานอ่อน โดยเจาะจงเพียงรูปแบบของ Linear Parameter (β) และ Variance Function ($V(\mu)$) โดยการประมาณนี้จะใช้ได้กับทั้งสองตัวแบบ

2.3.1 **โครงสร้างของ** Quasi - Likelihood

พิจารณาส่วนประกอบเดียวของตัวแปรตอบสนอง Y โดยไม่มีsubscript ภายใต้เงื่อนไขดังนี้

$$U = u(\mu|;Y) = \frac{Y-\mu}{\sigma^2 V(\mu)}$$

ตามมาด้วยคุณสมบัติพื้นฐานที่ได้มาจาก log-likelihood

$$E(U) = 0,$$
$$\mathrm{var}(U) = 1/\{\sigma^2 V(\mu)\},$$
$$-E(\frac{\partial U}{\partial \mu}) = 1/\{\sigma^2 V(\mu)\}$$

โดยที่ Quasi – Likelihood Function หาได้จาก

$Q(\mu;y) = \int_y^\mu u(y|y)dt = \int_y^\mu \frac{y-t}{\sigma^2 V(t)}dt$ ซึ่งมีลักษณะเหมือนฟังก์ชัน log-likelihood

Quasi – likelihood สำหรับข้อมูลที่สมบูรณ์เป็นผลรวมของแต่ละ $Q_i(\mu_i;y_i)$ คือ $Q(\mu;y) = \sum Q_i(\mu_i;y_i)$ ให้ฟังก์ชัน quasi – deviance สำหรับค่าสังเกตเดียว คือ $D(\mu;y) = 2\sigma^2 Q(\mu;y) = 2\int_y^\mu \frac{y-t}{V(t)}dt$ ที่เป็นอย่างถูกต้องที่เชื่อถือได้ ยกเว้น $y=\mu$ ผลรวมของ deviance $D(\mu;y)$ หามาได้จากผลรวมของแต่ละส่วนประกอบ คือ คำนวณจากฟังก์ชันที่ขึ้นอยู่กับ y และ μ แต่ไม่ขึ้นกับ σ^2

2.3.2 การประมาณพารามิเตอร์

Quasi - Likelihood Estimating Equation สำหรับพารามิเตอร์ (β) สามารถหาได้จาก $Q(\mu;y)$ อาจจะเขียนอยู่ในรูปของ $U(\beta)=0$ ซึ่งจะได้

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

ซึ่งเรียกว่า quasi – score function โดยที่

$$D_i \doteq \frac{\partial \mu_{it}}{\partial \beta_t}$$

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

$$\phi = \frac{1}{n-p}\sum_i \sum_t \frac{y_{it}-\mu_{it}}{\sqrt{\mathrm{var}(\mu_{it})}}$$

เมื่อ V_i คือ เมทริกซ์โครงสร้างความแปรปรวนร่วมของ Y_{it}

A_i คือ diagonal matrix ความแปรปรวนของ Y_{it} โดยเขียนให้อยู่ในรูปเมทริกซ์

ได้ดังนี้ $\begin{bmatrix} \mu_{1t} & 0 & \cdots & 0 \\ 0 & \mu_{2t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_{nt} \end{bmatrix}$

R_i คือ เมทริกซ์อัตตสัมพันธ์ของ Y_{it} โดยเขียนให้อยู่ในรูปเมทริกซ์

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

เนื่องจากในงานวิจัยนี้สนใจศึกษาอัตตสัมพันธ์แบบ Exchangeable

ϕ คือ overdispersion parameter = 1