การเปรียบเทียบตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ด้วยวิธีการประมาณพารามิเตอร์แบบQuasi-Likelihood สำหรับข้อมูลระยะยาว

นางสาวศิรินทิพย์ เสริมสุข

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต

สาขาวิชาสถิติ ภาควิชาสถิติ

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2553

A COMPARISON BETWEEN GENERALIZED LINEAR MODEL AND GENERALIZED ESTIMATING EQUATIONS WITH QUASI-LIKELIHOOD ESTIMATION FOR LONGITUDINAL DATA

Miss. Sirintip Sermsuk

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Statistics
Department of Statistics
Faculty of Commerce and Accountancy
Chulalongkorn University
Academic Year 2010

หัวข้อวิทยานิพนธ์	การเปรียบเทียบตัวแบบ Generalized Linear Model และตัวแบบ Generalized Estimating Equations ด้วยวิธีการประมาณพารามิเตอร์แบบQuasi-Likelihood สำหรับข้อมูลระยะยาว
โดย	นางสาวศิรินทิพย์ เสริมสุข
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร. กัลยา วานิชย์บัญชา

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.... คณบดีคณะพาณิชยศาสตร์และการบัญชี
(รองศาสตราจารย์ ดร. อรรณพ ตันละมัย)

คณะกรรมการสอบวิทยานิพนธ์

.... ประธานกรรมการ
(รองศาสตราจารย์ ดร. ธีระพร วีระถาวร)

.... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร. กัลยา วานิชย์บัญชา)

.... กรรมการ
(อาจารย์ ดร. อนุภาพ สมบูรณ์สวัสดี)

.... กรรมการภายนอกมหาวิทยาลัย
(อาจารย์ ดร.อรุณี กำลัง)

ศิรินทิพย์ เสริมสุข : การเปรียบเทียบตัวแบบ Generalized Linear Model และตัวแบบ Generalized Estimating Equations ด้วยวิธีการประมาณพารามิเตอร์แบบ Quasi-Likelihoodสำหรับข้อมูลระยะยาว.(A COMPARISON BETWEEN GENERALIZED LINEAR MODEL AND GENERALIZED ESTIMATING EQUATIONS WITH QUASI-LIKELIHOOD ESTIMATION FOR LONGITUDINAL DATA) อ. ที่ปรึกษาวิทยานิพนธ์หลัก: รศ. ดร. กัลยา วานิชย์บัญชา , 57 หน้า.

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบการประมาณพารามิเตอร์ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ด้วยวิธีการประมาณพารามิเตอร์แบบ Quasi-Likelihood สำหรับข้อมูลระยะยาว เมื่อกำหนดให้ตัวแปรตามมีการแจกแจงปัวซงส์ กระทำภายใต้เงื่อนไข ตัวแปรอิสระเป็นเชิงปริมาณคือ 1 และ 3 มีความสัมพันธ์กัน คือ 0.1 ,0.5 และ 0.9 ขนาดตัวอย่าง 20 , 60 ระยะเวลาที่ใช้ในการเก็บข้อมูลซ้ำ 3 และ 6 คาบเวลา เมื่อกำหนดอัตตสัมพันธ์ของตัวแปรตามของตัวแบบ Generalized Estimating Equations และโครงสร้างความแปรปรวนร่วมของตัวแบบ Generalized Linear Model คือ Exchangeable โดยกำหนดค่าอัตตสัมพันธ์ของตัวแปรตามคือ 0.3 และ 0.8 ข้อมูลที่ใช้ในการวิจัยได้จากการจำลองข้อมูลด้วยโปรแกรม Rและกระทำการทดลองซ้ำๆกัน 1,000 ครั้งโดยใช้ค่าเฉลี่ยของผลบวกกำลังสองของความคลาดเคลื่อนของสัมประสิทธิ์ความถดถอย (AMSE) ในการเปรียบการประมาณพารามิเตอร์ทั้งสองตัวแบบ

เมื่อทำการเปรียบเทียบด้วยค่า AMSE พบว่า ตัวแบบ Generalized Estimating Equations ประมาณพารามิเตอร์ถูกต้องกว่าตัวแบบ Generalized Linear Model เมื่อมีอัตตสัมพันธ์ของตัวแปรตามสูง ขนาดตัวอย่างเล็ก และระยะเวลาการเก็บข้อมูลซ้ำมาก และตัวแบบ Generalized Linear Mode lประมาณพารามิเตอร์ถูกต้องกว่าตัวแบบ Generalized Estimating Equations เมื่ออัตตสัมพันธ์ของตัวแปรตามต่ำ ขนาดตัวอย่างใหญ่ และระยะการเก็บข้อมูลซ้ำต่ำ

ภาควิชา	สถิติ	ลายมือชื่อนิสิต	ศิรินทิพย์ เสริมสุข
สาขาวิชา	สถิติ	ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก	
ปีการศึกษา	2553		

5181921826 : MAJOR STATISTICS
KEYWORDS : Count data/Quasi-Likelihood /Exchangeable/Generalized Estimating Equation /Generalized Linear Model/longitudinal

SIRINTIP SERMSUK : A COMPARISON BETWEEN GENERALIZED LINEAR MODEL AND GENERALIZED ESTIMATING EQUATIONS WITH QUASI-LIKELIHOOD ESTIMATION FOR LONGITUDINAL DATA.THESIS ADVISOR : ASSOC. PROF. KALAYA VANITBUNCHA, Ph.D., 57 pp.

The objective of this study is to compare the estimated parameter between Generalized Linear Model (GLM) and Generalized Estimating Equations (GEE) with quasi – likelihood estimation for longitudinal data when dependent variable is poisson distribution. The study is compared under condition of independent variable of 1 and 3 with multicollinearity among independent variables are 0.1, 0.5 and 0.9. The sample sizes of 20 and 60 with the repeated periods of 3 and 6 ,the working correlation and covariance structure of data are Exchangeable. The working correlation is 0.3 and 0.8 . The data are simulated with R-program and repeat 1,000 times. Using Average Mean Square Error (AMSE) to compare Generalized Linear Model (GLM) and Generalized Estimating Equations (GEE) for each situation .

The result of Generalized Estimating Equations(GEE) is better than Generalized Linear Model (GLM) when autocorrelation of dependent variable is high(0.8) and small sample sizes and six repeated periods. Generalized Linear Model (GLM) is better than Generalized Estimating Equations(GEE) when autocorrelation of dependent variable is low(0.3) and a large sample sizes and three repeated periods.

Department : Statistics	Student's Signature Sirintip Sermsuk
Field of Study : Statistics	Advisor's Signature Kamlaya Vanidbuncha
Academic Year : 2010	

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยความช่วยเหลืออย่างดียิ่งของ รองศาสตาจารย์ ดร.กัลยา วานิชย์บัญชา อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่กรุณาให้คำแนะนำตลอดจนการช่วยเหลือแก้ไขข้อบกพร่องต่างๆด้วยดีตลอดมา จนวิทยานิพนธ์นี้เสร็จสมบูรณ์ จึงใคร่ขอขอบพระคุณเป็นอย่างสูง ณ โอกาสนี้

ผู้วิจัยขอกราบขอบพระคุณ รองศาตราจารย์ ดร.ธีรพร วีระถาวร ในฐานะประธานกรรมการ และอาจารย์ ดร.อนุภาพ สมบูรณ์สวัสดี กรรมการสอบวิทยานิพนธ์ที่กรุณาตรวจแก้ไขวิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น และขอกราบขอบพระคุณคณาจารย์ประจำภาควิชาสถิติที่ให้โอกาสทางการศึกษาและประสิทธิ์ประสาทความรู้ให้แก่ผู้เขียนจนกระทั่งสำเร็จการศึกษา

ผู้วิจัยขอกราบขอบพระคุณ อาจารย์ดร.อรุณี กำลังที่ท่านได้เสียสละเวลาอันมีค่ามาเป็นกรรมการภายนอกมหาวิทยาลัย ซึ่งทาให้วิทยานิพนธ์ฉบับนี้มีความสมบูรณ์มากยิ่งขึ้น

ท้ายนี้ ผู้วิจัยใคร่ขอขอบพระคุณกำลังใจอย่างดียิ่งจากบิดา มารดา พร้อมทั้งเพื่อนๆทั้งหลายที่มีให้จนกระทั่งจบการศึกษา

สารบัญ

บทที่ หน้า

สารบัญตาราง

ตารางที่		หน้า

ตารางที่		หน้า

สารบัญภาพ

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันนี้งานวิจัยมีบทบาทสำคัญ หลายองค์กรมีการเก็บรวบรวมข้อมูลเพื่อนำมาทำงานวิจัย โดยใช้กระบวนการทางสถิติในการวิเคราะห์ประมวลผลและนำผลของการวิจัยปรับใช้กับแนวทางขององค์กร แต่ในบางครั้งการเก็บข้อมูลต้องใช้ระยะเวลานาน ไม่ว่าจะทางการแพทย์ การเกษตร ต้องเก็บข้อมูลจากหน่วยตัวอย่างเดียวกัน ซ้ำมากกว่า 1 ครั้ง ในระยะเวลาที่ต่างกัน เรียกว่า ข้อมูลระยะยาว

วิทยานิพนธ์เรื่องนี้เป็นการศึกษาการเปรียบเทียบประสิทธิภาพของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ด้วยข้อมูลระยะยาว โดยในการศึกษานี้จะจำลองสถานการณ์ที่ตัวแปรอิสระมีความสัมพันธ์กันและกำหนดอัตตสัมพันธ์และโครงสร้างความแปรปรวนร่วม เพื่อประมาณพารามิเตอร์ทั้งสองตัวแบบด้วยวิธีเดียวกันคือ วิธี Quasi-Likelihood แล้วเปรียบเทียบดูว่าตัวแบบใดเหมาะสมกว่าสำหรับข้อมูลระยะยาวที่มีตัวแปรตามมีการแจกจงแบบปัวซงส์ เพื่อสามารถนำไปใช้ประโยชน์ได้ในอนาคตต่อไป

การแจกแจงแบบปัวซงส์นั้นเป็นการแจกแจงปัวซงส์เป็นการแจกแจงที่เกิดขึ้นในตัวแปรสุ่มของการทดลองหรือว่าเหตุการณ์ที่เกิดขึ้นในช่วงระยะเวลาใดเวลาหนึ่งที่ต่อเนื่องกัน โดยหนึ่งหน่วยนั้นอาจจะเป็น วินาที นาที ชั่วโมง วัน สัปดาห์ เดือน หรือ ปี อาจจะเป็นความยาว พื้นที่ ปริมาตร โดยข้อมูลที่ได้จากตัวแปรสุ่มนี้เป็นแบบไม่ต่อเนื่องกันชนิดหนึ่ง ตัวอย่างของตัวแปรตามที่มีการแจกแจงแบบปัวซงส์พบอยู่ทั่วไป เช่น อัตราการเกิดของทารกที่เกิดขึ้นในหนึ่งปีของกรุงเทพมหานคร จำนวนครั้งที่มีการกดเอทีเอ็มในแต่ละวัน เป็นต้น เราสามารถเขียนฟังก์ชั่นการแจกแจงในรูปนี้ $f(Y_i)=\dfrac{(t_i\mu_i)^{Y_i}\exp(-t_i\mu_i)}{Y_i!};Y_i=0,1,2,3,...$ โดยที่ t_i คือ ระยะเวลาในการเก็บข้อมูลซ้ำ ตัวแบบ Generalized Estimating Equationsเป็นตัวแบบที่ขยายจากตัวแบบ Generalized Linear Model ภายใต้สถานการณ์ที่ข้อมูลมีความสัมพันธ์กัน สามารถเก็บข้อมูลซ้ำได้หลายช่วงเวลา ซึ่งเป็นตัวแบบที่นิยมมากที่ใช้กับข้อมูลที่เป็นกลุ่มและข้อมูลการนับ โดยที่ตัวแปรตามไม่จำเป็นต้องมีการแจกแจงแบบปกติ สามารถมีการแจกแจงแบบ Binomial , Poisson , Gamma เป็นต้น สามารถเขียนในรูปแบบได้ดังนี้

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

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

และตัวแบบ Generalized Linear Model ที่นำมาใช้ในการเปรียบเทียบนี้ คือ Poisson Regression ซึ่งมีตัวแปรตามมีการแจกแจงแบบปัวซงส์เช่นกัน สามารถเขียนในรูปแบบดังนี้

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

$t = 1,2,\ldots,T$ ซึ่งรายละเอียดของตัวแปรต่างๆในสมการจะอธิบายในบทถัดไป วิธีที่ใช้ในการประมาณค่าพารามิเตอร์นี้ จะเลือกใช้วิธี Quasi – Likelihood ซึ่งเป็นการประมาณที่ดึงมาส่วนหนึ่งของการประมาณMaximum Likelihood เนื่องจากสามารถใช้ประมาณค่าพารามิเตอร์ได้ทั้งสองตัวแบบ ซึ่งมีรูปแบบที่สนใจดังนี้ $U(\beta) = 0$ ซึ่งจะได้

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

ซึ่งเรียกว่า quasi – score function โดยที่

$$D_i = \frac{\partial \mu_i}{\partial \beta_t}$$

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

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

เมื่อ V_i คือ เมทริกซ์โครงสร้างความแปรปรวนร่วมของ Y_{it}

A_i คือ diagonal matrix ความแปรปรวนของ Y_{it}

R_i คือ เมทริกซ์อัตตสัมพันธ์ของ Y_{it}

ϕ คือ overdispersion parameter

วัตถุประสงค์ของการวิจัย

เพื่อศึกษาเปรียบเทียบประสิทธิภาพ ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ด้วยวิธีการประมาณพารามิเตอร์แบบ Quasi-Likelihood สำหรับข้อมูลระยะยาว เมื่อกำหนดให้ตัวแปรตามมีการแจกแจงปัวซงส์

ขอบเขตของการวิจัย

ในการวิจัยครั้งนี้กระทำภายใต้ขอบเขตดังนี้

1. วิธีการประมาณค่าพารามิเตอร์ที่สนใจในการศึกษาในงานวิจัย คือ

   Quasi – Likelihood ซึ่งมีรูปแบบดังนี้

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

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

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

2. จำนวนตัวแปรอิสระที่นำมาศึกษาเท่ากับ 1 , 3
3. ขนาดตัวอย่างที่ศึกษาเท่ากับ 20 , 60
4. กำหนดให้อัตตสัมพันธ์และโครงสร้างความแปรปรวนร่วมของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations คือ Exchangeable
5. กำหนดรูปแบบความสัมพันธ์ระหว่างตัวแปรอิสระด้วยกัน คือ มีสัมพันธ์กันเล็กน้อย สัมพันธ์กันปานกลาง และสัมพันธ์กันมาก คือ 0.1 , 0.5 ,0.9
6. ศึกษาเมื่อระยะเวลาที่ทำการเก็บข้อมูลซ้ำ (t) เท่ากับ 3 , 6
7. กำหนดการสุ่มตัวอย่างจำลองสถานการณ์ 1000 รอบ
8. เปรียบเทียบตัวแบบด้วยค่าเฉลี่ยของผลบวกกำลังสองของความคลาดเคลื่อนของสัมประสิทธิ์ความถดถอย (AMSE)

ข้อตกลงเบื้องต้น

ในการวิเคราะห์ครั้งนี้ผู้วิจัยได้กำหนดข้อตกลงเบื้องต้นดังนี้

1. ตัวแปรตามมีการแจกแจงแบบปัวซงส์
2. ตัวแปรอิสระเป็นเชิงปริมาณและมีความสัมพันธ์กัน
3. ข้อมูลที่ศึกษาไม่มีค่าติดลบ

เกณฑ์ที่ใช้ในการตัดสินใจ

ในการวิจัยนี้จะพิจารณาค่าเฉลี่ยของผลบวกกำลังสองของความคลาดเคลื่อนของสัมประสิทธิ์ความถดถอย (Average Mean Square Error : AMSE) ซึ่งมีสูตรการคำนวณดังนี้

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

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

β_{mk} คือ ค่าจริงของพารามิเตอร์ตัวที่ m ครั้งที่ k

M คือ จำนวนของพารามิเตอร์ที่มีอยู่ในสมการ

คำจำกัดความที่ใช้ในการวิจัย

ในการวิจัยครั้งนี้มีคำจำกัดความที่ใช้ดังนี้

1. ข้อมูลระยะยาว (Longitudinal Data) หมายถึง ข้อมูลที่มีการเก็บรวบรวมจากหน่วยทดลองเดิมมากกว่า 1 ครั้งขึ้นไปในช่วงเวลาที่ต่างกัน
2. พารามิเตอร์ (Parameter) หมายถึง ค่าคงที่ที่แสดงคุณลักษณะบางประการของประชากร
3. Exchangeable หมายถึง ความสัมพันธ์ของค่าสังเกตภายในหน่วยศึกษาเดียวกันจะมีค่าเท่ากัน ณ ช่วงเวลาที่ต่างกัน

ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อทราบถึงตัวแบบที่เหมาะสมในการวิเคราะห์ข้อมูล
2. เพื่อทราบถึงวิธีการประมาณค่าพารามิเตอร์ที่ตัวแปรตามมีการแจกแจงแบบปัวซงส์ด้วยวิธี Quasi - Likelihood
3. เพื่อเป็นแนวทางในการศึกษาวิจัยต่อไป

วิธีดำเนินการวิจัย

1. สร้างตัวแปรตามที่มีการแจกแจงแบบปัวซงส์
2. สร้างตัวแปรอิสระโดยกำหนดให้เป็นค่าคงที่ โดยกำหนดให้มีค่าสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient) ของตัวแปรอิสระ
3. สร้างโครงสัมพันธ์ของตัวแปรตามของตัวแบบ Generalized Estimating Equations แบบ Exchangeable
4. สร้างโครงสร้างความแปรปรวนร่วมของแบบ Generalized Linear Model แบบ Exchangeable

5. ทำการประมาณพารามิเตอร์ของตัวแบบ Generalized Linear Model ด้วยวิธี Quasi-Likelihood
6. ทำการประมาณพารามิเตอร์ของตัวแบบ Generalized Estimating Equations ด้วยวิธี Quasi-Likelihood
7. ทำการเปรียบเทียบวิธีการประมาณตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ที่ได้การวิธี Quasi-Likelihood โดยเปรียบเทียบด้วยวิธี ค่าเฉลี่ยของผลบวกกำลังสองของความคลาดเคลื่อน (Average Mean Square Error : AMSE)
8. สรุปผลการวิจัยในแต่ละสถานการณ์

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

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

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

$Y_{i(t)}$ 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 $G(Y) = X\underset{\sim}{\beta}$

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

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

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

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

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

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

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

$Y_{i(t)}$ 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

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

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

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

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

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

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

เนื่องจากในงานวิจัยนี้สนใจศึกษาอัตตสัมพันธ์แบบ Exchangeable

ϕ คือ overdispersion parameter = 1

บทที่ 3

วิธีดำเนินการวิจัย

การดำเนินการวิจัยนี้ เป็นการวิจัยเพื่อเปรียบเทียบประสิทธิภาพของการประมารพารามิเตอร์ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ด้วยวิธีการประมาณพารามิเตอร์แบบ Quasi-Likelihood สำหรับข้อมูลระยะยาว เมื่อกำหนดให้ตัวแปรตามมีการแจกแจงปัวซงส์ โดยศึกษาเปรียบเทียบค่าเฉลี่ยของผลบวกกำลังสองของความคลาดเคลื่อนของสัมประสิทธิ์ความถดถอยและสำหรับการเรียบเทียบจะกระทำภายใต้สถานการณ์ที่กำหนด 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 โดยกำหนดให้มีอัตตสัมพันธ์ของตัวแปรตามที่ 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 $\beta_0 = -1, \beta_{1t} = 0.01, \beta_{21} = 0.02,$ $\beta_{22} = 0.04,\ \beta_{23} = 0.01$ เนื่องจากกำหนดให้มีอัตตสัมพันธ์ของตัวแปรตามแบบ Exchangeable ค่าพารามิเตอร์ที่กำหนดจึงต้องเป็นค่าที่ทำให้เกิดอัตตสัมพันธ์องตัวแปรตามเป็นแบบ Exchangeable

3.1.7 กำหนดรอบประมวลผลในแต่ละสถานการณ์ 1000 รอบ

3.1.8 เปรียบเทียบตัวแบบด้วยค่าเฉลี่ยของผลบวกกำลังสองของความคลาดเคลื่อนของสัมประสิทธิ์ความถดถอย (AMSE)

3.2 ขั้นตอนในการดำเนินการวิจัย

มีขั้นตอนในการดำเนินการวิจัยดังนี้

3.2.1 การสร้างข้อมูลตัวแปรตาม (y) ที่มีการแจกแจงแบบปัวซงส์

3.2.2 การสร้างข้อมูลตัวแปรอิสระ (x) ที่มีการแจกแจงแบบปกติ

3.2.3 สร้างอัตตสัมพันธ์ของตัวแปรตามของตัวแบบ Generalized Estimating Equations แบบ Exchangeable

3.2.4 สร้างโครงสร้างความแปรปรวนร่วมตัวแปรตามของแบบ Generalized Linear Model แบบExchangeable โดยใช้ข้อมูลชุดเดียวกัน

3.2.5 ทำการประมาณพารามิเตอร์ของตัวแบบ Generalized Linear Model และ ตัวแบบGeneralized Estimating Equations ด้วยวิธี Quasi- Likelihood

3.3.6 เปรียบเทียบค่าเฉลี่ยของผลบวกกำลังสองของความคลาดเคลื่อนของสัมประสิทธิ์ความถดถอย (AMSE)ของทั้งสองตัวแบบ ว่าตัวแบบใดให้ค่านี้ต่ำกว่ากัน แล้วดูร้อยละความแตกต่างของทั้งสองตัวแบบว่าแตกต่างเท่าไหร่ และสรุปผลในรูปของตารางและกราฟในแต่ละกรณี การประมาณพารามิเตอร์ของตัวแบบใดให้ค่าเฉลี่ยของผลบวกกำลังสองของความคลาดเคลื่อนของสัมประสิทธิ์ความถดถอยต่ำกว่า แสดงว่าตัวแบบนั้นให้การประมาณพารามิเตอร์ที่ดีกว่า

3.3 การสร้างข้อมูลตัวแปรตาม (y)

การสร้างข้อมูลตัวแปรตาม (Y) ทำได้ด้วยการสุ่มตัวอย่างค่าตัวแปรตามให้มีการแจกแจงแบบปัวซงส์ ด้วยโปรแกรมRโดยกำหนดให้มีพารามิเตอร์ (λ) เท่ากับ 0.4 ซึ่งเป็นค่าพารามิเตอร์ที่ทำให้ความแปรปรวนคงที่ จะใช้คำสั่งrpois()ในการสร้างข้อมูล เมื่อมีขนาดตัวอย่าง n เท่ากับ 20 และ 60 และทำซ้ำช่วงเวลา t เท่ากับ 3 แล ะ 6 คาบเวลา จำนวนวน 1,000 รอบ โดยสร้างตัวแปรตามให้มีอัตตสัมพันธ์กันที่ 0.3 และ 0.8 ให้อยู่ในรูปแบบของ Exchangeableคือ $Corr(y_{it}, y_{iw}) = \rho \,; t \neq w$ 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

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

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

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

ในการศึกษาครั้งนี้เริ่มต้นสร้างตัวแปรตาม (Y) และตัวแปรอิสระ ($\underset{\sim}{X}$) เนื่องจากมีเงื่อนไขของอัตตสัมพันธ์ของตัวแปรตามเป็นแบบ Exchangeable ซึ่งเกิดขึ้นได้ยากเพราะเป็นความสัมพันธ์ของค่าสังเกตภายในหน่วยศึกษาเดียวกันจะมีค่าเท่ากัน ณ ช่วงเวลาต่างกัน ตัวแปรตามมีการแจกแจงแบบปัวซงส์ 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 และกำหนดค่าฟังก์ชั่นเชื่อมโยง (Link)เป็น log การสร้างพารามิเตอร์จะใช้การวนซ้ำ (Iteration) 1,000 รอบ โดยค่าพารามิเตอร์เบต้าจะหาได้จาก

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

ซึ่งเรียกว่า quasi – score function โดยที่

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

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

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

เมื่อ V_i คือ เมทริกซ์โครงสร้างความแปรปรวนร่วมของ Y_{it}

A_i คือ diagonal matrix ความแปรปรวนของ Y_{it}

R_i คือ เมทริกซ์อัตตสัมพันธ์ของ Y_{it}

ϕ คือ overdispersion parameter = 1

3.6 เปรียบเทียบค่าเฉลี่ยของผลบวกกำลังสองของความคลาดเคลื่อนของพารามิเตอร์ (AMSE)

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

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

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

β_{mk} คือ ค่าจริงของพารามิเตอร์ตัวที่ m ครั้งที่ k

M คือ จำนวนของพารามิเตอร์ที่มีอยู่ในสมการ

3.7 ร้อยละความแตกต่างของ 2 ตัวแบบ

เป็นการหาร้อยละของความแตกต่างของค่าAMSEของตัวแบบ Generalized Linear Model และตัวแบบ Generalized Estimating Equations เพื่อดูร้อยละความแตกต่างของค่า AMSE ของทั้งสองตัวแบบ ซึ่งคำนวณได้ดังนี้

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

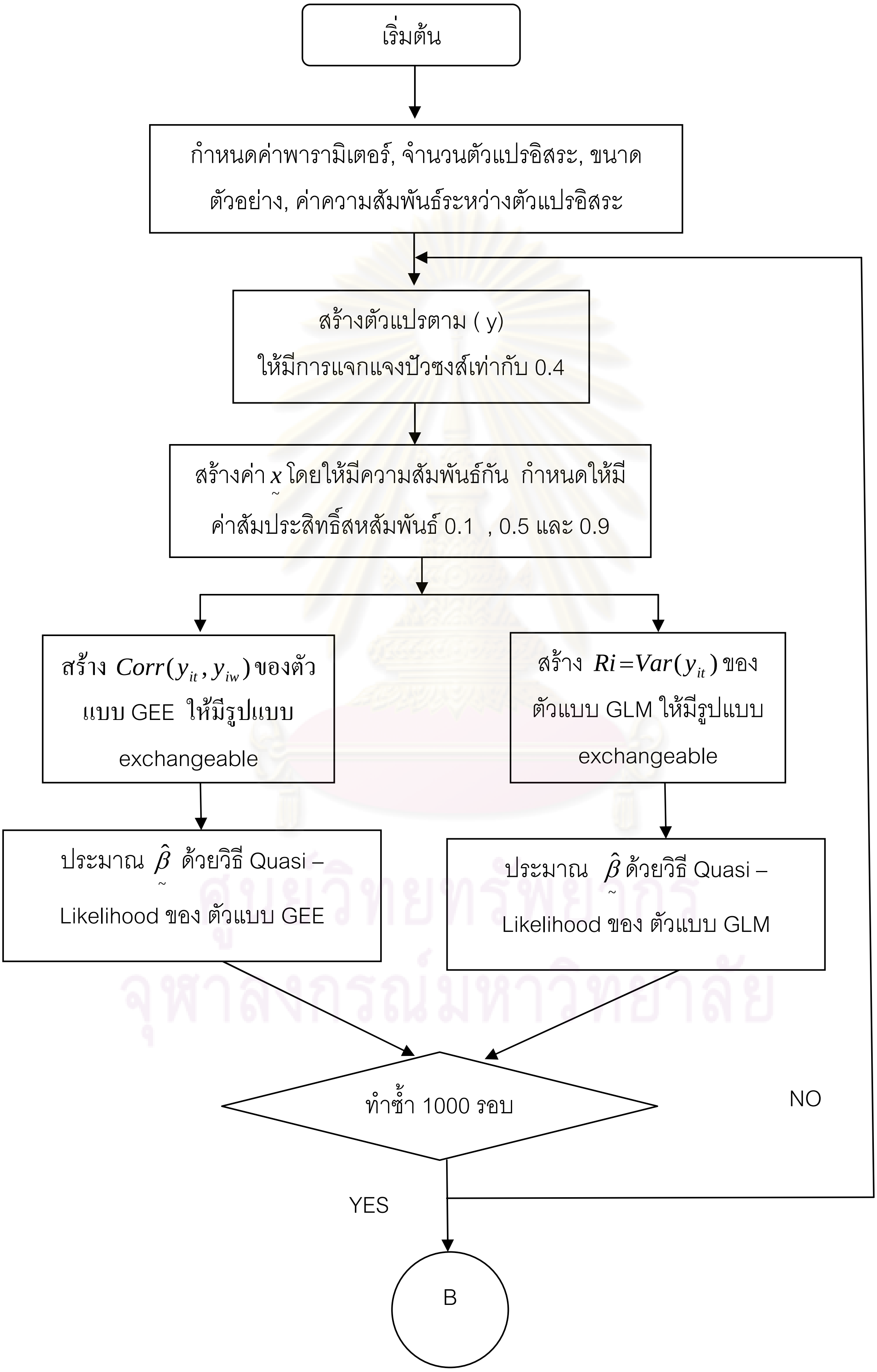
โดย $Max(AMSE)$ คือ ตัวแบบที่ให้ค่า AMSE สูง เมื่อทำการเปรียบเทียบสองตัวแบบ

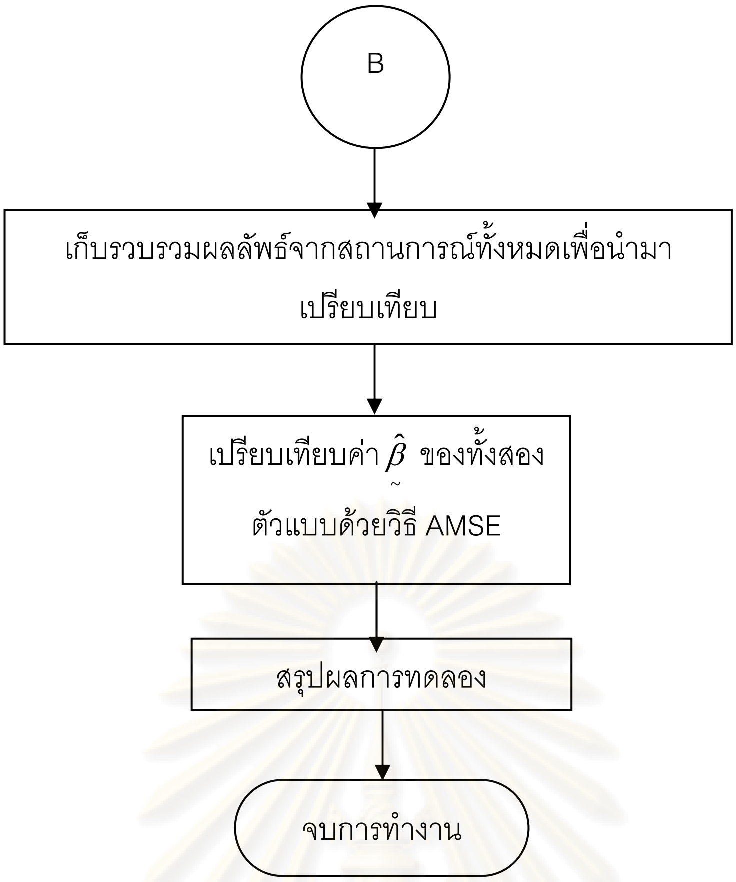
$Min(AMSE)$ คือ ตัวแบบที่ให้ค่า AMSE ต่ำ เมื่อทำการเปรียบเทียบสองตัวแบบ

3.8 สรุปผลการวิจัยในแต่ละสถานการณ์

ทำการเปรียบเทียบทุกกรณีจากการประมาณพารามิเตอร์ของทั้ง2 ตัวแบบ ว่าตัวแบบใดให้ค่าเฉลี่ยของผลบวกกำลังสองของความคลาดเคลื่อนของสัมประสิทธิ์ความถดถอยต่ำกว่าแสดงว่าตัวแบบนั้นให้การประมาณพารามิเตอร์ที่เหมาะกว่า

3.9 ขั้นตอนการทำงานของโปรแกรม





รูปที่ 3.1 แสดงขั้นตอนการทำงานของโปรแกรม

บทที่ 4

ผลการวิเคราะห์ข้อมูล

วิทยานิพนธ์เรื่องการเปรียบเทียบการประมาณพารามิเตอร์ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ด้วยวิธีการประมาณพารามิเตอร์แบบQuasi-Likelihood สำหรับข้อมูลระยะยาว มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบประสิทธิภาพการประมาณพารามิเตอร์ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ด้วยวิธีการประมาณพารามิเตอร์แบบ Quasi-Likelihood สำหรับข้อมูลระยะยาว เมื่อกำหนดให้ตัวแปรตามมีการแจกแจงปัวซงส์ และเพื่อทำการเปรียบเทียบค่าเฉลี่ยของผลบวกกำลังสองของความคลาดเคลื่อนของสัมประสิทธิ์ความถดถอย ถ้าหากค่าดังกล่าวของตัวแบบใดมีค่าน้อยกว่า แสดงว่าตัวแบบนั้นจะให้การประมาณค่าดีกว่าอีกตัวแบบหนึ่ง ซึ่งผลการวิจัยจะนำเสนอในรูปแบบตารางและกราฟ เพื่อความสะดวกในการอธิบายจึงขอใช้สัญลักษณ์ดังต่อไปนี้

Rho(x) หมายถึง สัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระ

n หมายถึง ขนาดตัวอย่าง

t หมายถึง ระยะเวลาที่เก็บข้อมูลซ้ำ

AMSE หมายถึง ค่าเฉลี่ยของผลบวกกำลังสองของความคลาดเคลื่อนของสัมประสิทธิ์ความถดถอย

* หมายถึง ค่า AMSE ที่มีค่าต่ำที่สุด

ในการศึกษาเรื่องนี้ตัวแปรตาม (y)จะมีการแจกแจงแบบปัวซงส์ ส่วนตัวแปรอิสระ (x) นั้นมีการแจกแจงปกติที่มีความสัมพันธ์กัน จะทำการทดลองภายใต้สถานการณ์ต่างๆ ดังต่อไปนี้

1. ตัวแปรตามมีการแจกแจงแบบปัวซงส์ ที่มีพารามิเตอร์เท่ากับ 0.4 และตัวแปรอิสระ 1 ตัว มีขนาดตัวอย่างเท่ากับ 20,60 ระยะการเก็บข้อมูลซ้ำเท่ากับ 3 ,6 กำหนดให้มีอัตตสัมพันธ์ เท่ากับ 0.3 และโครงสร้างความแปรปรวนแปรผันตามอัตตสัมพันธ์
2. ตัวแปรตามมีการแจกแจงแบบปัวซงส์ ที่มีพารามิเตอร์เท่ากับ 0.4 และตัวแปรอิสระ 3 ตัว โดยให้ตัวแปรอิสระมีความสัมพันธ์กันที่ 0.1 , 0.5 และ 0.9 มีขนาดตัวอย่างเท่ากับ20,60 ระยะการเก็บข้อมูลซ้ำเท่ากับ 3,6 กำหนดให้มีอัตตสัมพันธ์ เท่ากับ 0.3 และโครงสร้างความแปรปรวนแปรผันตามอัตตสัมพันธ์

3. ตัวแปรตามมีการแจกแจงแบบปัวซงส์ ที่มีพารามิเตอร์เท่ากับ 0.4 และตัวแปรอิสระ 1 ตัว มีขนาดตัวอย่างเท่ากับ 20 ,60 ระยะการเก็บข้อมูลซ้ำเท่ากับ 3,6 กำหนดให้มีอัตตสัมพันธ์ เท่ากับ 0.8 และโครงสร้างความแปรปรวนแปรผันตามอัตตสัมพันธ์
4. ตัวแปรตามมีการแจกแจงแบบปัวซงส์ ที่มีพารามิเตอร์เท่ากับ 0.4 และตัวแปรอิสระ 3 ตัว โดยให้ตัวแปรอิสระมีความสัมพันธ์กันที่ 0.1 , 0.5 และ 0.9 มีขนาดตัวอย่างเท่ากับ20,60 ระยะการเก็บข้อมูลซ้ำเท่ากับ 3,6 กำหนดให้มีอัตตสัมพันธ์ เท่ากับ 0.8 และโครงสร้างความแปรปรวนแปรผันตามอัตตสัมพันธ์

โดยในแต่ละสถานการณ์จะมีสถานการณ์ย่อยของแต่ละกรณี และสถานการณ์ที่กำหนดนี้จำทำการประมวลผล 1,000รอบ ผลลัพธ์ที่ได้จะนำมาคำนวณหาค่าเฉลี่ยของผลบวกกำลังสองของความคลาดเคลื่อนของสัมประสิทธิ์ความถดถอยของทั้งสองตัวแบบ เพื่อศึกษาว่าภายใต้สถานการณ์ที่กำหนดตัวแบบใดให้การประมาณค่าพารามิเตอร์ได้ดีกว่า ซึ่งผลการวิจัยได้นำเสนอในตารางดังกล่าวต่อไปนี้

4.1 การเปรียบเทียบค่าเฉลี่ยของผลบวกกำลังสองของความคลาดเคลื่อนของพารามิเตอร์ ของตัวแบบ แบบ Generalized Linear Modelและ ตัวแบบGeneralized Estimating Equations

4.4.1 <u>ตัวแปรตาม (Y) มีอัตตสัมพันธ์ที่ 0.3</u>

ตารางที่ 4.1: แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ที่มีขนาดตัวอย่าง (n) เท่ากับ 20 ตัวแปรอิสระ(p) 1 ตัวแปร ระยะเวลาที่เก็บข้อมูลซ้ำ (t) เท่ากับ 3 และ6 ช่วงเวลา

n	t	AMSE	
		Generalized Linear Model	Generalized Estimating Equations
20	3	0.04625*	0.04947
	6	0.02050	0.02005*

จากตารางที่ 4.1 เมื่อพิจารณาที่ระยะเวลาที่เก็บข้อมูลซ้ำที่ 3 และ 6 พบว่าทั้งตัวแบบ Generalized Linear Modelและตัวแบบ Generalized Estimating Equations ที่ ระยะเวลาเท่ากับ 6 มีค่า AMSE ต่ำกว่าที่ระยะเวลาเท่ากับ 3

เมื่อเปรียบเทียบสองตัวแบบ ณ ช่วงเวลาทั้งสอง เมื่อขนาดตัวอย่างเท่ากัน พบว่าที่ระยะเวลาที่เก็บข้อมูลซ้ำที่ 3 ตัวแบบ Generalized Linear Model ให้ค่า AMSE ต่ำกว่าตัวแบบ Generalized Estimating Equations คือ 0.04625 และ ที่ระยะเวลาที่เก็บข้อมูลซ้ำที่ 6 ตัวแบบ Generalized Estimating Equations ให้ค่า AMSE ต่ำกว่าตัวแบบ Generalized Linear Model คือ 0.02005

ตารางที่ 4.2 : แสดงค่าAMSE ของตัวแบบGeneralized Linear Models และ ตัวแบบ Generalized Estimating Equations ที่มีขนาดตัวอย่าง(n) เท่ากับ 60 ตัวแปรอิสระ(p) 1 ตัวแปร ระยะเวลาที่เก็บข้อมูลซ้ำ(t) เท่ากับ 3 และ6 ช่วงเวลา

n	t	AMSE	
		Generalized Linear Model	Generalized Estimating Equations
60	3	0.01551*	0.01650
	6	0.00970	0.00931*

จากตารางที่ 4.2 เมื่อพิจารณาที่ระยะเวลาที่เก็บข้อมูลซ้ำที่ 3 และ 6 พบว่าทั้งตัวแบบ Generalized Linear Model และตัวแบบ Generalized Estimating Equations ที่ ระยะเวลาเท่ากับ 6 มีค่า AMSE ต่ำกว่าที่ระยะเวลาเท่ากับ 3

เมื่อเปรียบเทียบสองตัวแบบ ณ ช่วงเวลาทั้งสอง เมื่อขนาดตัวอย่างเท่ากัน พบว่าที่ระยะเวลาที่เก็บข้อมูลซ้ำที่ 3 ตัวแบบ Generalized Linear Model ให้ค่า AMSE ต่ำกว่าตัวแบบ Generalized Estimating Equations คือ 0.01551และ ที่ระยะเวลาที่เก็บข้อมูลซ้ำที่ 6 ตัวแบบ Generalized Estimating Equationsให้ค่า AMSE ต่ำกว่าตัวแบบ Generalized Linear Model คือ 0.00931

ตารางที่ 4.3 : แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ที่มีขนาดตัวอย่าง (n)เท่ากับ 20 ตัวแปรอิสระ(p) 3 ตัวแปร ที่ระดับสัมประสิทธิ์สหสัมพันธ์ 3 ระดับ ระยะเวลาที่เก็บข้อมูลซ้ำ(t) 3 ช่วงเวลา

n	t	Rho(x)	AMSE	
			Generalized Linear Model	Generalized Estimating Equations
20	3	0.1	0.03093*	0.03582
		0.5	0.03148*	0.03703
		0.9	0.04056*	0.05024

จากตารางที่ 4.3 พบว่า ที่สัมประสิทธิ์สหสัมพันธ์ ของตัวแปรอิสระ ที่ 0.1 , 0.5 และ 0.9 ค่าAMSE ของตัวแบบ Generalized Linear Model และตัวแบบ Generalized Estimating Equations มีค่าเพิ่มขึ้น ตามระดับสัมประสิทธิ์สหสัมพันธ์ ของตัวแปรอิสระที่เพิ่มขึ้น

เมื่อเปรียบเทียบค่า AMSE ทั้งสองตัวแบบที่ขนาดตัวอย่างเท่ากัน พบว่าค่าAMSE ตัวแบบ Generalized Linear Model ต่ำกว่าของตัวแบบ Generalized Estimating Equations ทุกระดับสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระ ที่ 0.1 , 0.5 และ 0.9 คือ 0.03093 , 0.03148 และ 0.04056 ตามลำดับ

ตารางที่ 4.4 : แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ที่มีขนาดตัวอย่าง(n)เท่ากับ 20 ตัวแปรอิสระ(p) 3 ตัวแปร ที่ระดับสัมประสิทธิ์สหสัมพันธ์ 3 ระดับ ระยะเวลาที่เก็บข้อมูลซ้ำ(t) 6 ช่วงเวลา

n	t	Rho(x)	AMSE	
			Generalized Linear Model	Generalized Estimating Equations
20	6	0.1	0.01314	0.01300*
		0.5	0.01364*	0.01380
		0.9	0.01634*	0.01668

จากตารางที่ 4.4 พบว่า ที่สัมประสิทธิ์สหสัมพันธ์ ของตัวแปรอิสระ ที่ 0.1 , 0.5 และ 0.9 ค่าAMSEของตัวแบบ Generalized Linear Model และตัวแบบ Generalized Estimating Equations มีค่าเพิ่มขึ้น ตามระดับสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระที่เพิ่มขึ้น

เมื่อเปรียบเทียบค่า AMSE ทั้งสองตัวแบบที่ขนาดตัวอย่างเท่ากัน พบว่า ที่ระดับสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระ ที่ 0.1 ค่าAMSE ของตัวแบบ Generalized Estimating Equationsต่ำกว่าของตัวแบบ Generalized Linear Model คือ 0.01300 แต่ที่ระดับสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระ ที่ 0.5 และ 0.9 ค่าAMSE ของตัวแบบ Generalized Linear Model ต่ำกว่าของตัวแบบ Generalized Estimating Equations คือ 0.01380 และ 0.01668 ตามลำดับ

ตารางที่ 4.5: แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ที่มีขนาดตัวอย่าง(n)เท่ากับ 60 ตัวแปรอิสระ(p) 3 ตัวแปร ที่ระดับสัมประสิทธิ์สหสัมพันธ์ 3 ระดับ ระยะเวลาที่เก็บข้อมูลซ้ำ(t) 3 ช่วงเวลา

n	t	Rho(x)	AMSE	
			Generalized Linear Model	Generalized Estimating Equations
60	3	0.1	0.01010*	0.01123
		0.5	0.01026*	0.01166
		0.9	0.01240*	0.01449

จากตารางที่ 4.5 พบว่า ที่สัมประสิทธิ์สหสัมพันธ์ ของตัวแปรอิสระ ที่ 0.1 , 0.5 และ 0.9 ค่าAMSEของตัวแบบ Generalized Linear Model และตัวแบบ Generalized Estimating Equations มีค่าเพิ่มขึ้น ตามระดับสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระที่เพิ่มขึ้น

เมื่อเปรียบเทียบค่า AMSE ทั้งสองตัวแบบที่ขนาดตัวอย่างเท่ากัน พบว่า ค่าAMSE ตัวแบบ Generalized Linear Model ต่ำกว่าของตัวแบบ Generalized Estimating Equations ทุกระดับสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระ ที่ 0.1 ,0.5 และ 0.9 คือ 0.01010 , 0.01026 และ 0.01240 ตามลำดับ

ตารางที่ 4.6: แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ที่มีขนาดตัวอย่าง(n)เท่ากับ 60 ตัวแปรอิสระ(p) 3 ตัวแปร ที่ระดับสัมประสิทธิ์สหสัมพันธ์ 3 ระดับ ระยะเวลาที่เก็บข้อมูลซ้ำ(t) 6 ช่วงเวลา

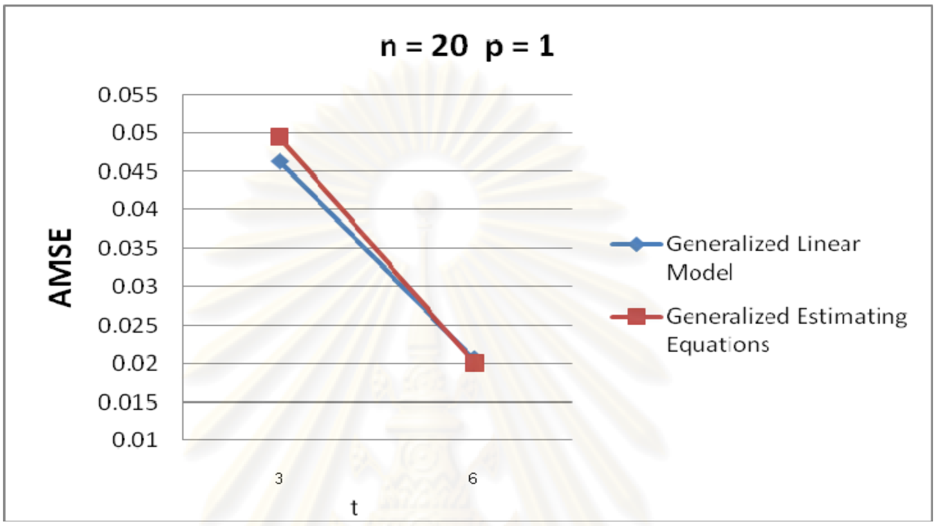
n	t	Rho(x)	AMSE	
			Generalized Linear Model	Generalized Estimating Equations
60	6	0.1	0.00624	0.00602*
		0.5	0.00631	0.00607*
		0.9	0.00720	0.00700*

จากตารางที่ 4.6 พบว่า ที่สัมประสิทธิ์สหสัมพันธ์ ของตัวแปรอิสระ ที่ 0.1 , 0.5 และ 0.9 ค่าAMSEของตัวแบบ Generalized Linear Model และตัวแบบ Generalized Estimating Equations มีค่าเพิ่มขึ้น ตามระดับสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระที่เพิ่มขึ้น

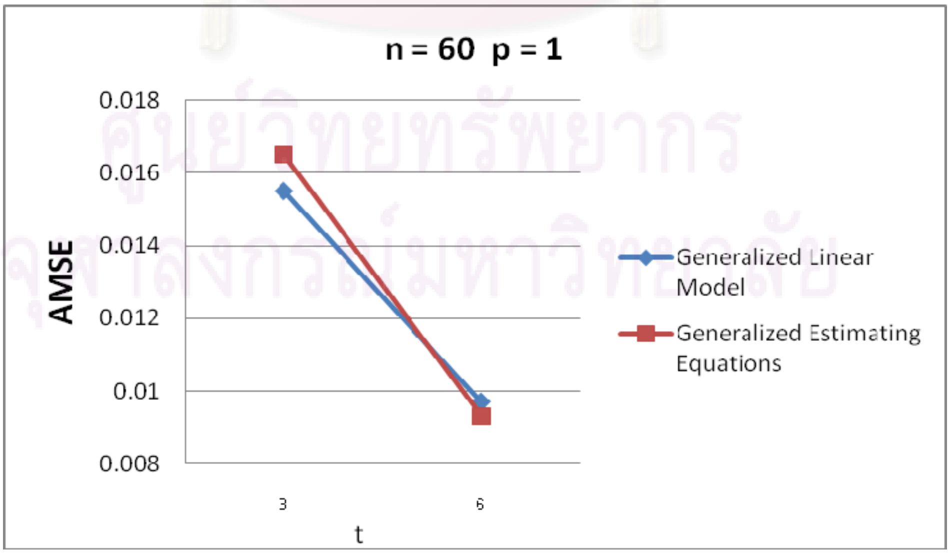
เมื่อเปรียบเทียบค่า AMSE ทั้งสองตัวแบบที่ขนาดตัวอย่างเท่ากัน พบว่า ค่าAMSE ตัวแบบ Generalized Estimating Equationsต่ำกว่าของตัวแบบ Generalized Linear Model ทุกระดับสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระ ที่ 0.1 , 0.5 และ 0.9 คือ 0.00602, 0.00607และ 0.00700ตามลำดับ

<u>ตัวแปรตาม (Y) มีอัตตสัมพันธ์ที่ 0.3</u>

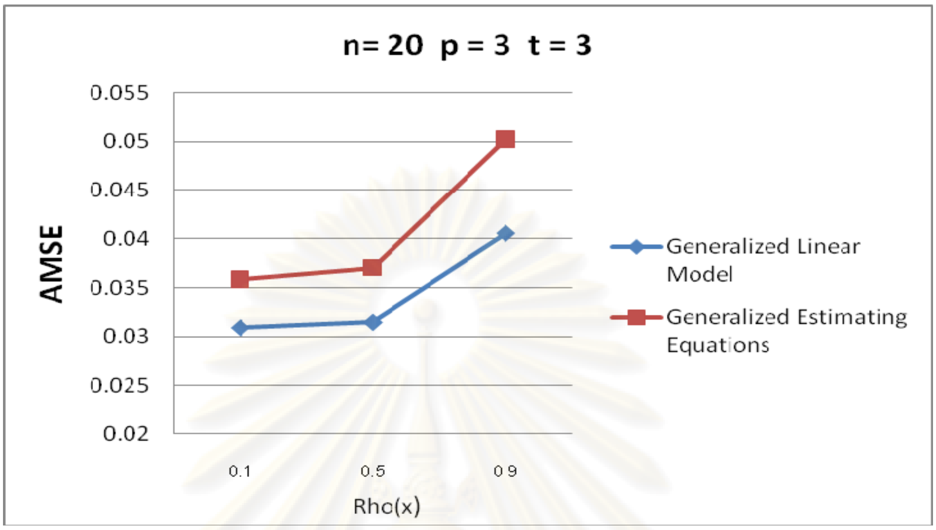
รูปที่ 4.1: แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบGeneralized Estimating Equations ที่มีขนาดตัวอย่าง (n) เท่ากับ 20 ตัวแปรอิสระ(p) 1 ตัวแปร ระยะเวลาที่เก็บข้อมูลซ้ำ (t) เท่ากับ 3 และ6 ช่วงเวลา



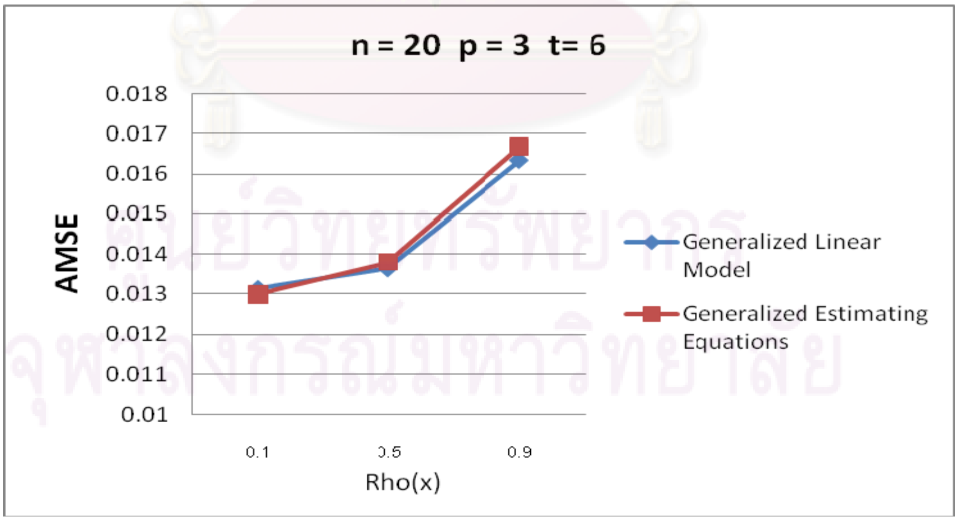
รูปที่ 4.2: แสดงค่าAMSE ของตัวแบบGeneralized Linear Model และ ตัวแบบGeneralized Estimating Equations ที่มีขนาดตัวอย่าง(n) เท่ากับ 60 ตัวแปรอิสระ(p) 1 ตัวแปร ระยะเวลาที่เก็บข้อมูลซ้ำ(t) เท่ากับ 3 และ6 ช่วงเวลา



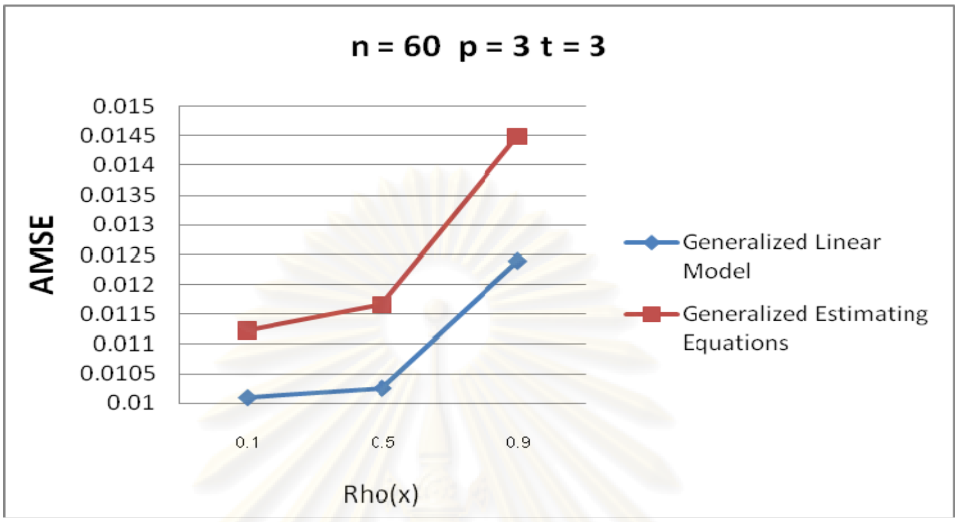
รูปที่ 4.3: แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบGeneralized Estimating Equations ที่มีขนาดตัวอย่าง (n)เท่ากับ 20 ตัวแปรอิสระ(p) 3 ตัวแปร ที่ระดับสัมประสิทธิ์สหสัมพันธ์ 3 ระดับ ระยะเวลาที่เก็บข้อมูลซ้ำ(t) 3 ช่วงเวลา



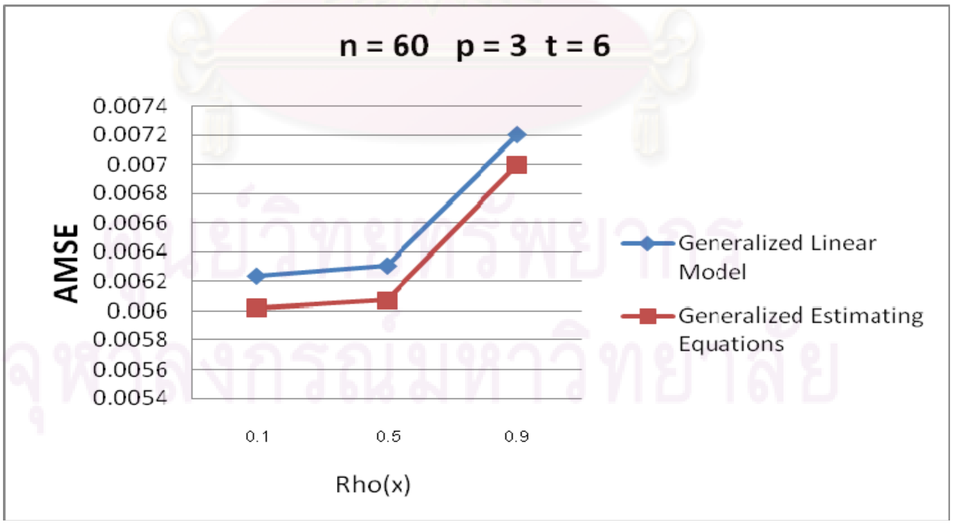
รูปที่ 4.4: แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบGeneralized Estimating Equations ที่มีขนาดตัวอย่าง(n)เท่ากับ 20 ตัวแปรอิสระ(p) 3 ตัวแปร ที่ระดับสัมประสิทธิ์สหสัมพันธ์ 3 ระดับ ระยะเวลาที่เก็บข้อมูลซ้ำ(t) 6 ช่วงเวลา



รูปที่ 4.5: แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบGeneralized Estimating Equations ที่มีขนาดตัวอย่าง(n)เท่ากับ 60 ตัวแปรอิสระ(p) 3 ตัวแปร ที่ระดับสัมประสิทธิ์สหสัมพันธ์ 3 ระดับ ระยะเวลาที่เก็บข้อมูลซ้ำ(t) 3 ช่วงเวลา



รูปที่ 4.6: แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบGeneralized Estimating Equations ที่มีขนาดตัวอย่าง(n)เท่ากับ 60 ตัวแปรอิสระ(p) 3 ตัวแปร ที่ระดับสัมประสิทธิ์สหสัมพันธ์ 3 ระดับ ระยะเวลาที่เก็บข้อมูลซ้ำ(t) 6 ช่วงเวลา



4.4.2 ตัวแปรตาม (y) มีอัตตสัมพันธ์ที่ 0.8

ตารางที่ 4.7 : แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ที่มีขนาดตัวอย่าง (n) เท่ากับ 20 ตัวแปรอิสระ(p) 1 ตัวแปร ระยะเวลาที่เก็บข้อมูลซ้ำ (t) เท่ากับ 3 และ6 ช่วงเวลา

n	t	AMSE	
		Generalized Linear Model	Generalized Estimating Equations
20	3	0.0447*	0.0493
	6	0.0943	0.0918*

จากตารางที่ 4.7 เมื่อพิจารณาที่ระยะเวลาที่เก็บข้อมูลซ้ำที่ 3 และ 6 พบว่าทั้งตัวแบบ Generalized Linear Modelและตัวแบบ Generalized Estimating Equations ที่ ระยะเวลาเท่ากับ 3 มีค่า AMSE ต่ำกว่าที่ระยะเวลาเท่ากับ 6

เมื่อเปรียบเทียบสองตัวแบบ ณ ช่วงเวลาทั้งสอง เมื่อขนาดตัวอย่างเท่ากัน พบว่าที่ระยะเวลาที่เก็บข้อมูลซ้ำที่ 3 ตัวแบบ Generalized Linear Model ให้ค่า AMSE ต่ำกว่าตัวแบบ Generalized Estimating Equations คือ 0.0447 และ ที่ระยะเวลาที่เก็บข้อมูลซ้ำที่ 6 ตัวแบบ Generalized Estimating Equationsให้ค่า AMSE ต่ำกว่าตัวแบบ Generalized Linear Model คือ 0.0918

ตารางที่ 4.8 : แสดงค่าAMSE ของตัวแบบGeneralized Linear Model และ ตัวแบบ Generalized Estimating Equations ที่มีขนาดตัวอย่าง(n) เท่ากับ 60 ตัวแปรอิสระ(p) 1 ตัวแปร ระยะเวลาที่เก็บข้อมูลซ้ำ(t) เท่ากับ 3 และ6 ช่วงเวลา

n	t	AMSE	
		Generalized Linear Model	Generalized Estimating Equations
60	3	0.0186*	0.0193
	6	0.0569	0.0564*

จากตารางที่ 4.8 เมื่อพิจารณาที่ระยะเวลาที่เก็บข้อมูลซ้ำที่ 3 และ 6 พบว่าทั้งตัวแบบ Generalized Linear Modelและตัวแบบ Generalized Estimating Equations ที่ ระยะเวลาเท่ากับ 3 มีค่า AMSE ต่ำกว่าที่ระยะเวลาเท่ากับ 6

เมื่อเปรียบเทียบสองตัวแบบ ณ ช่วงเวลาทั้งสอง เท่ากัน พบว่าที่ระยะเวลาที่เก็บข้อมูลซ้ำที่ 3 ตัวแบบ Generalized Linear Mode ให้ค่า AMSE ต่ำกว่าตัวแบบ Generalized Estimating Equations คือ 0.0186 และ ที่ระยะเวลาที่เก็บข้อมูลซ้ำที่ 6 ตัวแบบ Generalized Estimating Equationsให้ค่า AMSE ต่ำกว่าตัวแบบ Generalized Linear Modelคือ 0.0564

ตารางที่ 4.9 : แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ที่มีขนาดตัวอย่าง (n)เท่ากับ 20 ตัวแปรอิสระ(p) 3 ตัวแปร ที่ระดับสัมประสิทธิ์สหสัมพันธ์ 3 ระดับ ระยะเวลาที่เก็บข้อมูลซ้ำ(t) 3 ช่วงเวลา

n	t	Rho(x)	AMSE	
			Generalized Linear Model	Generalized Estimating Equations
20	3	0.1	0.0313	0.0305*
		0.5	0.0307	0.0297*
		0.9	0.0400	0.0367*

จากตารางที่ 4.9 พบว่า ที่ระดับสัมประสิทธิ์สหสัมพันธ์ ของตัวแปรอิสระ ที่ 0.1 และ 0.9 ค่า AMSE ของตัวแบบ Generalized Linear Model และตัวแบบ Generalized Estimating Equations มีค่าเพิ่มขึ้น ตามระดับสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระที่เพิ่มขึ้น ขณะที่ที่ระดับสัมประสิทธิ์สหสัมพันธ์ ของตัวแปรอิสระ ที่ 0.5 ค่าAMSEของตัวแบบ Generalized Linear Modelและตัวแบบ Generalized Estimating Equations มีค่าลดลง

เมื่อเปรียบเทียบค่า AMSE ทั้งสองตัวแบบพบว่า ค่า AMSE ตัวแบบ Generalized Estimating Equations จะให้ค่า AMSE ต่ำกว่าตัวแบบ Generalized Linear Model ทุกระดับสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระ ที่ 0.1 ,0.5 และ 0.9 คือ 0.0305 ,0.0297 และ 0.0367 ตามลำดับ

ตารางที่ 4.10 : แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ที่มีขนาดตัวอย่าง(n)เท่ากับ 20 ตัวแปรอิสระ(p) 3 ตัวแปร ที่ระดับสัมประสิทธิ์สหสัมพันธ์ 3 ระดับ ระยะเวลาที่เก็บข้อมูลซ้ำ(t) 6 ช่วงเวลา

n	t	Rho(x)	AMSE	
			Generalized Linear Model	Generalized Estimating Equations
20	6	0.1	0.05529	0.0539*
		0.5	0.05534	0.0540*
		0.9	0.0576	0.0565*

จากตารางที่ 4.10 พบว่า ที่ระดับสัมประสิทธิ์สหสัมพันธ์ ของตัวแปรอิสระ ที่ 0.1 , 0.5 และ 0.9 ค่าAMSEของตัวแบบ Generalized Linear Model และตัวแบบ Generalized Estimating Equations มีค่าเพิ่มขึ้น ตามระดับสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระที่เพิ่มขึ้น

เมื่อเปรียบเทียบค่า AMSE ทั้งสองตัวแบบพบว่า ค่าAMSE ตัวแบบ Generalized Estimating Equationsต่ำกว่าของตัวแบบ Generalized Linear Model ทุกระดับสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระ ที่ 0.1 , 0.5 และ 0.9 คือ 0.0539, 0.0540 และ 0.0565ตามลำดับ

ตารางที่ 4.11: แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ที่มีขนาดตัวอย่าง(n)เท่ากับ 60 ตัวแปรอิสระ(p) 3 ตัวแปร ที่ระดับสัมประสิทธิ์สหสัมพันธ์ 3 ระดับ ระยะเวลาที่เก็บข้อมูลซ้ำ(t) 3 ช่วงเวลา

n	t	Rho(x)	AMSE	
			Generalized Linear Model	Generalized Estimating Equations
60	3	0.1	0.0114*	0.0129
		0.5	0.0117*	0.0125
		0.9	0.0132*	0.0141

จากตารางที่ 4.11 พบว่า ที่ระดับสัมประสิทธิ์สหสัมพันธ์ ของตัวแปรอิสระ ที่ 0.1 , 0.5 และ 0.9 ค่าAMSEของตัวแบบ Generalized Linear Model และตัวแบบ Generalized Estimating Equations มีค่าเพิ่มขึ้น ตามระดับสัมประสิทธิ์สหสัมพันธ์ ของตัวแปรอิสระที่เพิ่มขึ้น

เมื่อเปรียบเทียบค่า AMSE ทั้งสองตัวแบบพบว่า ค่าAMSE ตัวแบบ Generalized Linear Model ต่ำกว่าของตัวแบบ Generalized Estimating Equations ทุกระดับสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระ ที่ 0.1 , 0.5 และ 0.9 คือ 0.0114, 0.0117และ 0.0132 ตามลำดับ

ตารางที่ 4.12: แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ที่มีขนาดตัวอย่าง(n)เท่ากับ 60 ตัวแปรอิสระ(p) 3 ตัวแปร ที่ระดับสัมประสิทธิ์สหสัมพันธ์ 3 ระดับ ระยะเวลาที่เก็บข้อมูลซ้ำ(t) 6 ช่วงเวลา

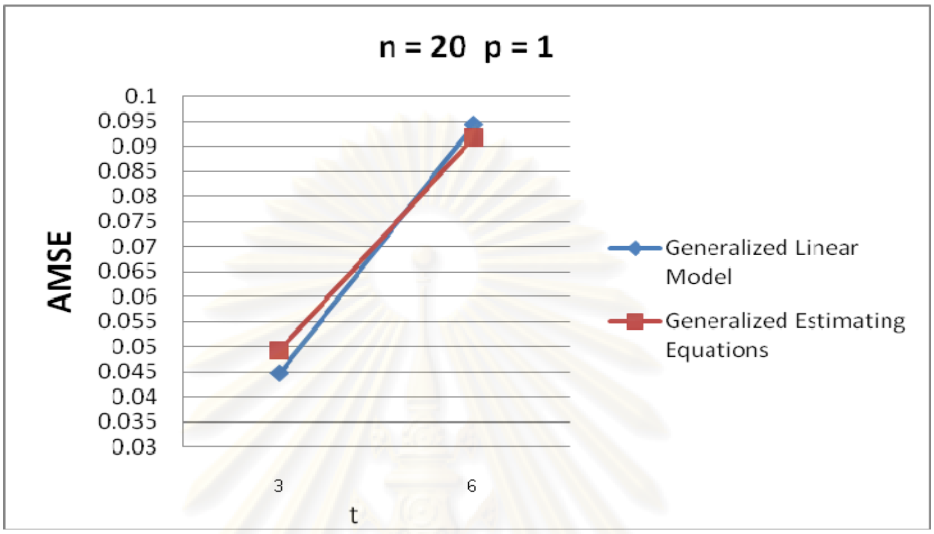
n	t	Rho(x)	AMSE	
			Generalized Linear Model	Generalized Estimating Equations
60	6	0.1	0.0338*	0.0341
		0.5	0.0342*	0.0346
		0.9	0.0347*	0.0352

จากตารางที่ 4.12 พบว่า ที่ระดับสัมประสิทธิ์สหสัมพันธ์ ของตัวแปรอิสระ ที่ 0.1 , 0.5 และ 0.9 ค่าAMSEของตัวแบบ Generalized Linear Model และตัวแบบ Generalized Estimating Equations มีค่าเพิ่มขึ้น ตามระดับสัมประสิทธิ์สหสัมพันธ์ ของตัวแปรอิสระที่เพิ่มขึ้น

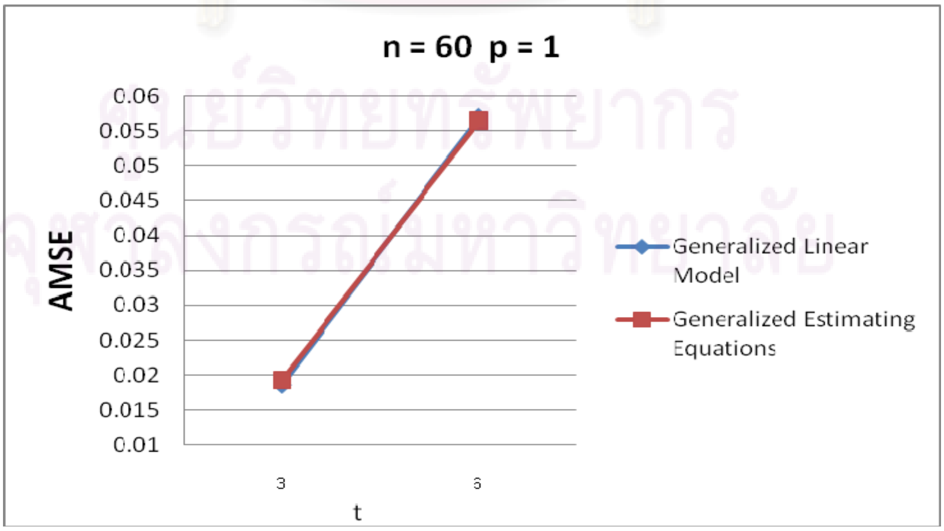
เมื่อเปรียบเทียบค่า AMSE ทั้งสองตัวแบบพบว่า ค่า AMSEของ ตัว Generalized Linear Model ต่ำกว่าของตัวแบบ Generalized Estimating Equations ทุกระดับสัมประสิทธิ์สหสัมพันธ์ ของตัวแปรอิสระ ที่ 0.1 , 0.5 และ 0.9 คือ 0.0338, 0.0342และ 0.0347 ตามลำดับ

<u>ตัวแปรตาม (y) มีอัตตสัมพันธ์ที่</u> <u>0.8</u>

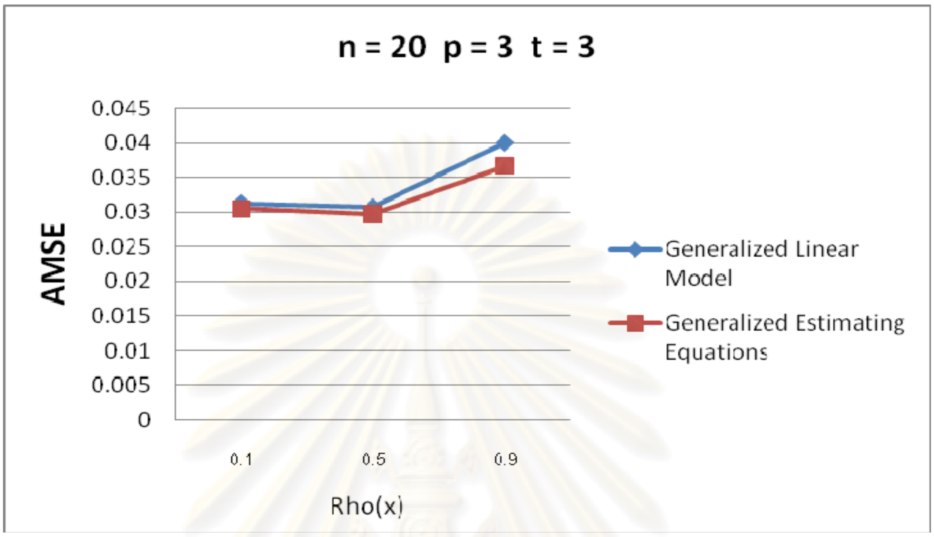
รูปที่ 4.7: แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบGeneralized Estimating Equations ที่มีขนาดตัวอย่าง (n) เท่ากับ 20 ตัวแปรอิสระ(p) 1 ตัวแปร ระยะเวลาที่เก็บข้อมูลซ้ำ (t) เท่ากับ 3 และ6 ช่วงเวลา



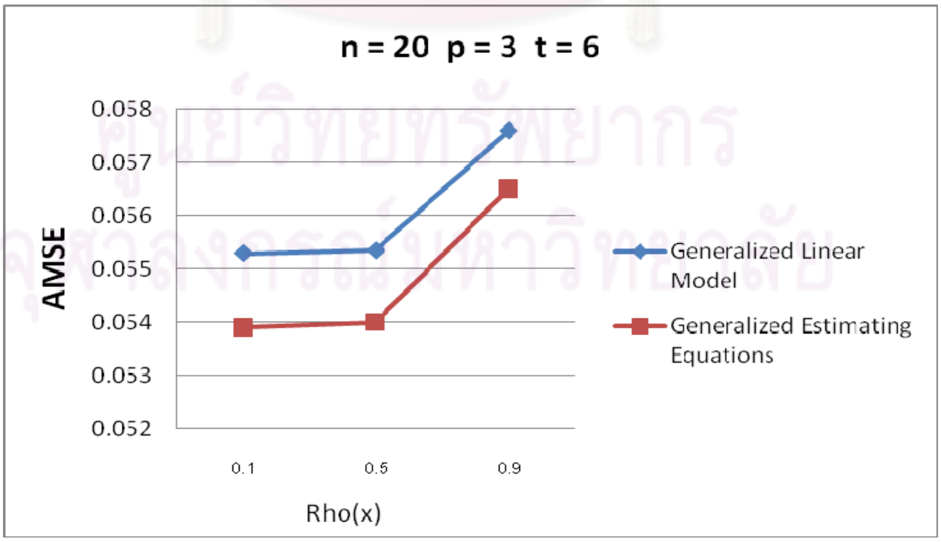
รูปที่ 4.8: แสดงค่าAMSE ของตัวแบบGeneralized Linear Model และ ตัวแบบGeneralized Estimating Equations ที่มีขนาดตัวอย่าง(n) เท่ากับ 60 ตัวแปรอิสระ(p) 1 ตัวแปร ระยะเวลาที่เก็บข้อมูลซ้ำ(t) เท่ากับ 3 และ6 ช่วงเวลา



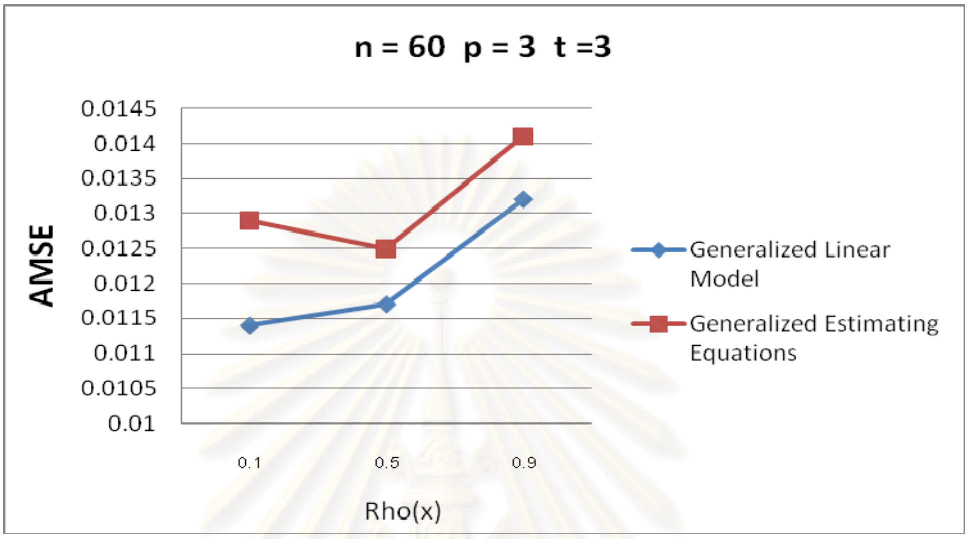
รูปที่ 4.9: แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบGeneralized Estimating Equations ที่มีขนาดตัวอย่าง (n)เท่ากับ 20 ตัวแปรอิสระ(p) 3 ตัวแปร ที่ระดับสัมประสิทธิ์สหสัมพันธ์ 3 ระดับ ระยะเวลาที่เก็บข้อมูลซ้ำ(t) 3 ช่วงเวลา



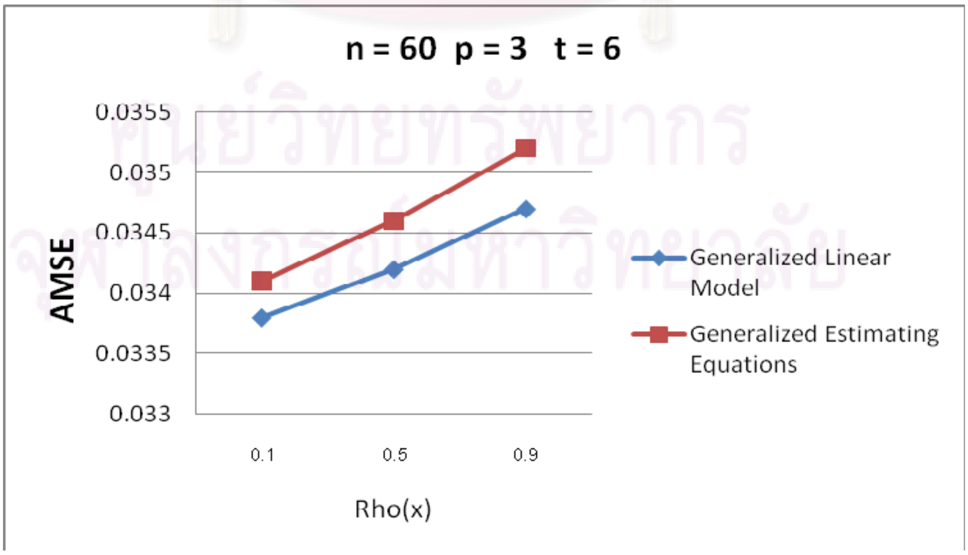
รูปที่ 4.10: แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบGeneralized Estimating Equations ที่มีขนาดตัวอย่าง(n)เท่ากับ 20 ตัวแปรอิสระ(p) 3 ตัวแปร ที่ระดับสัมประสิทธิ์สหสัมพันธ์ 3 ระดับ ระยะเวลาที่เก็บข้อมูลซ้ำ(t) 6 ช่วงเวลา



รูปที่ 4.11: แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบGeneralized Estimating Equations ที่มีขนาดตัวอย่าง(n)เท่ากับ 60 ตัวแปรอิสระ(p) 3 ตัวแปร ที่ระดับสัมประสิทธิ์สหสัมพันธ์ 3 ระดับ ระยะเวลาที่เก็บข้อมูลซ้ำ(t) 3 ช่วงเวลา



รูปที่ 4.12: แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบGeneralized Estimating Equations ที่มีขนาดตัวอย่าง(n)เท่ากับ 60 ตัวแปรอิสระ(p) 3 ตัวแปร ที่ระดับสัมประสิทธิ์สหสัมพันธ์ 3 ระดับ ระยะเวลาที่เก็บข้อมูลซ้ำ(t) 6 ช่วงเวลา



4.2 สรุปการเปรียบเทียบค่าเฉลี่ยของผลบวกกำลังสองของความคลาดเคลื่อนของพารามิเตอร์ ของตัวแบบ Generalized Linear Modelและ ตัวแบบGeneralized Estimating Equations โดยที่ตัวแปรตาม (y)มีอัตตสัมพันธ์ที่ 0.3 และ0.8

ตารางที่ 4.13: แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ที่มีขนาดตัวอย่าง (n) เท่ากับ 20และ 60 ตัวแปรอิสระ(p) 1 ตัวแปร ระยะเวลาที่เก็บข้อมูลซ้ำ (t) เท่ากับ 3 ช่วงเวลา โดยที่ตัวแปรตาม (y)มีอัตตสัมพันธ์ที่ 0.3 และ 0.8

อัตตสัมพันธ์ของตัวแปรตาม	t	n	AMSE		ร้อยละความ** แตกต่างค่า AMSEของ2ตัวแบบ
			Generalized Linear Model	Generalized Estimating Equations	
0.3	3	20	0.04625*	0.04947	6.962162
		60	0.01551*	0.01650	6.382979
0.8		20	0.0447*	0.0493	10.29083
		60	0.0186*	0.0193	3.763441

จากตารางที่ 4.13 เมื่อพิจารณาที่อัตตสัมพันธ์ของตัวแปรตามที่ 0.3 ที่ขนาดตัวอย่างเพิ่มขึ้น 20 เป็น60 แต่ระยะเวลาการเก็บข้อมูลซ้ำ เท่ากับ 3 พบว่าค่า AMSE ของตัวแบบ Generalized Linear Model ต่ำกว่าของตัวแบบ Generalized Estimating Equations โดยมีร้อยละความแตกต่างค่า AMSE ของ 2ตัวแบบ คือ 6.962162 และ 6.382979 ตามลำดับ

เมื่อพิจารณาที่อัตตสัมพันธ์ของตัวแปรตามที่ 0.8 ที่ขนาดตัวอย่างเพิ่มขึ้น 20 เป็น60 แต่ระยะเวลาการเก็บข้อมูลซ้ำ เท่ากับ 3 จะพบว่าค่า AMSE ของตัวแบบ Generalized Linear Model ต่ำกว่าตัวแบบ Generalized Estimating Equations โดยมีร้อยละความแตกต่างค่า AMSE ของ 2ตัวแบบ คือ 10.29083 และ 3.763441

ทำการเปรียบเทียบทั้งสองอัตตสัมพันธ์ของตัวแปรตามพบว่า เมื่ออัตตสัมพันธ์เพิ่มขึ้นโดยส่วนใหญ่จะส่งผลให้ค่า AMSE ของตัวแบบ Generalized Linear Modelต่ำกว่าของตัวแบบ Generalized Estimating Equations เมื่อระยะเวลาการเก็บข้อมูลซ้ำ เท่ากับ 3

หมายเหตุ ** คือ เป็นการหาร้อยละของความแตกต่างของค่า AMSE ของ2ตัวแบบ 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

ตารางที่ 4.14 : แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ที่มีขนาดตัวอย่าง (n) เท่ากับ 20 และ 60 ตัวแปรอิสระ(p) 1 ตัวแปร ระยะเวลาที่เก็บข้อมูลซ้ำ (t) เท่ากับ 6 ช่วงเวลา โดยที่ตัวแปรตาม (y)มีอัตตสัมพันธ์ที่ 0.3 และ 0.8

อัตตสัมพันธ์ของตัวแปรตาม	t	n	AMSE		ร้อยละความ** แตกต่างค่า AMSEของ2ตัวแบบ
			Generalized Linear Model	Generalized Estimating Equations	
0.3	6	20	0.02050	0.02005*	2.244389
		60	0.00970	0.00931*	4.189044
0.8		20	0.0943	0.0918*	2.723312
		60	0.0569	0.0564*	0.886525

จากตารางที่4.14 เมื่อพิจารณาที่อัตตสัมพันธ์ของตัวแปรตามที่ 0.3 ที่ขนาดตัวอย่างเพิ่มขึ้น 20 เป็น60 แต่ระยะเวลาการเก็บข้อมูลซ้ำ เท่ากับ 6 พบว่าค่า AMSE ของตัวแบบ Generalized Estimating Equations ต่ำกว่าของตัวแบบ Generalized Linear Model โดยมีร้อยละความแตกต่างค่า AMSE ของ 2ตัวแบบ คือ 2.244389 และ 4.189044ตามลำดับ

เมื่อพิจารณาที่อัตตสัมพันธ์ของตัวแปรตามที่ 0.8 ที่ขนาดตัวอย่างเพิ่มขึ้น 20 เป็น60 แต่ระยะเวลาการเก็บข้อมูลซ้ำ เท่ากับ 6 พบว่าค่า AMSE ของตัวแบบ Generalized Estimating Equations ต่ำกว่าของตัวแบบ Generalized Linear Model โดยมีร้อยละความแตกต่างค่าAMSE ของ 2ตัวแบบ คือ 2.723312 และ 0.886525ตามลำดับ

ทำการเปรียบเทียบทั้งสองอัตตสัมพันธ์ของตัวแปรตามพบว่าเมื่ออัตตสัมพันธ์เพิ่มขึ้นโดยส่วนใหญ่จะส่งผลให้ค่า AMSE ของตัวแบบ Generalized Estimating Equations ต่ำกว่าของตัวแบบ Generalized Linear Model เมื่อระยะเวลาการเก็บข้อมูลซ้ำ เท่ากับ 6

ตารางที่ 4.15 : แสดงค่า AMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ที่มีขนาดตัวอย่าง(n)เท่ากับ 20 ตัวแปรอิสระ (x) 3 ตัวแปร ที่ระดับสัมประสิทธิ์สหสัมพันธ์(Rho (x)) 3 ระดับ ระยะเวลาที่เก็บข้อมูลซ้ำ (t) 3 ช่วงเวลา โดยที่ตัวแปรตาม (y)มีอัตตสัมพันธ์ที่ 0.3 และ 0.8

อัตตสัมพันธ์ของตัวแปรตาม	t	n	Rho (x)	AMSE		ร้อยละความ** แตกต่างค่า AMSEของ2ตัวแบบ
				Generalized Linear Model	Generalized Estimating Equations	
0.3	3	20	0.1	0.03093*	0.30582	15.8099
			0.5	0.03148*	0.03703	17.6302
			0.9	0.04056*	0.05024	23.8659
0.8			0.1	0.0313	0.0305*	2.622951
			0.5	0.0307	0.0297*	3.367003
			0.9	0.0400	0.0367*	8.991826

จากตารางที่ 4.15 เมื่อพิจารณาที่อัตตสัมพันธ์ของตัวแปรตามที่ 0.3 ที่ขนาดตัวอย่างเท่ากับ 20 ที่สัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระเท่ากับ 0.1 ,0.5 และ 0.9 ระยะเวลาการเก็บข้อมูลซ้ำ เท่ากับ 3 พบว่าค่า AMSE ของตัวแบบ Generalized Linear Model มีค่าต่ำกว่าของตัวแบบ Generalized Estimating Equations โดยมีร้อยละความแตกต่างค่า AMSE ของ 2 ตัวแบบ คือ 15.8099 , 17.6302 และ 23.8659 ตามลำดับ

เมื่อพิจารณาที่อัตตสัมพันธ์ของตัวแปรตามที่ 0.8 ที่ขนาดตัวอย่างเท่ากับ 20 ที่สัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระเท่ากับ0.1 ,0.5 และ 0.9 ระยะเวลาการเก็บข้อมูลซ้ำ เท่ากับ 3 พบว่าค่า AMSE ของตัวแบบ Generalized Estimating Equationsมีค่าต่ำกว่าของตัวแบบ Generalized Linear Model โดยมีร้อยละความแตกต่างค่า AMSE ของ 2 ตัวแบบ คือ 2.622951 , 3.367003 และ 8.991826 ตามลำดับ

ทำการเปรียบเทียบทั้งสองอัตตสัมพันธ์ของตัวแปรตามพบว่า เมื่ออัตตสัมพันธ์เพิ่มขึ้นจะส่งผลให้ค่าAMSEของตัวแบบ Generalized Estimating Equationsต่ำกว่าของตัวแบบ Generalized Linear Model เมื่อระยะเวลาการเก็บข้อมูลซ้ำ เท่ากับ 3

ตารางที่ 4.16 : แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ที่มีขนาดตัวอย่าง(n)เท่ากับ 20 ตัวแปรอิสระ (x) 3 ตัวแปร ที่ระดับสัมประสิทธิ์สหสัมพันธ์(Rho (x)) 3 ระดับ ระยะเวลาที่เก็บข้อมูลซ้ำ (t) 6 ช่วงเวลา โดยที่ตัวแปรตาม (y)มีความสัมพันธ์ที่ 0.3 และ 0.8

อัตตสัมพันธ์ของตัวแปรตาม	t	n	Rho (x)	AMSE		ร้อยละความ** แตกต่างค่า AMSEของ2ตัวแบบ
				Generalized Linear Model	Generalized Estimating Equations	
0.3	6	20	0.1	0.01314	0.01300*	1.06545
			0.5	0.01364*	0.01380	1.17302
			0.9	0.01634*	0.01668	2.08078
0.8			0.1	0.05529	0.0539*	2.57885
			0.5	0.05534	0.0540*	2.48148
			0.9	0.0576	0.0565*	1.94690

จากตารางที่ 4.16 เมื่อพิจารณาที่อัตตสัมพันธ์ของตัวแปรตามที่ 0.3 ที่ขนาดตัวอย่างเท่ากับ 20 ที่สัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระเท่ากับ0.1 ,0.5 และ 0.9 ระยะเวลาการเก็บข้อมูลซ้ำ เท่ากับ 6 พบว่าค่า AMSE ของตัวแบบ Generalized Linear Model มีค่าต่ำกว่าของตัวแบบ Generalized Estimating Equations โดยมีร้อยละความแตกต่างค่า AMSE ของ 2ตัวแบบคือ1.06545% , 1.17302% และ 2.08078 ตามลำดับ

เมื่อพิจารณาที่อัตตสัมพันธ์ของตัวแปรตามที่ 0.8 ที่ขนาดตัวอย่างเท่ากับ 20 ที่สัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระเท่ากับ0.1 ,0.5 และ 0.9 ระยะเวลาการเก็บข้อมูลซ้ำ เท่ากับ 6 พบว่าค่า AMSE ของตัวแบบ Generalized Estimating Equations มีค่าต่ำกว่าของตัวแบบ Generalized Linear Model โดยมีร้อยละความแตกต่างค่าAMSEของ 2 ตัวแบบ คือ 2.57885% , 2.48148%และ 1.94690 %ตามลำดับ

ทำการเปรียบเทียบทั้งสองอัตตสัมพันธ์ของตัวแปรตาม พบว่า เมื่อความสัมพันธ์เพิ่มขึ้นจะส่งผลให้ค่า AMSE ของตัวแบบ Generalized Estimating Equations ต่ำกว่าของตัวแบบ Generalized Linear Model เมื่อระยะเวลาการเก็บข้อมูลซ้ำ เท่ากับ 6

จากตารางที่ 4.15 และ ตารางที่ 4.16 พบว่า เมื่อขนาดตัวอย่างเท่ากับ 20 อัตตสัมพันธ์เท่ากับ 0.3 ระยะเวลาการเก็บข้อมูลซ้ำ เท่ากับ 3 ตัวแบบ Generalized Linear Model มีค่าต่ำกว่า แต่เมื่อเพิ่มระยะเวลาการเก็บข้อมูลซ้ำเท่ากับ 6 ตัวแบบ Generalized Estimating Equationsต่ำกว่า ในขณะที่อัตตสัมพันธ์เท่ากับ 0.8 ระยะเวลาการเก็บข้อมูลซ้ำ เท่ากับ 3 และ6 ตัวแบบ Generalized Estimating Equationsต่ำกว่า

ตารางที่ 4.17: แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ที่มีขนาดตัวอย่าง(n)เท่ากับ 60 ตัวแปรอิสระ (x) 3 ตัวแปร ที่ระดับสัมประสิทธิ์สหสัมพันธ์(Rho (x)) 3 ระดับ ระยะเวลาที่เก็บข้อมูลซ้ำ (t) 3 ช่วงเวลา โดยที่ตัวแปรตาม (y)มีอัตตสัมพันธ์ที่ 0.3 และ 0.8

อัตตสัมพันธ์ของตัวแปรตาม	t	n	Rho (x)	AMSE		ร้อยละความ** แตกต่างค่า AMSEของ2ตัวแบบ
				Generalized Linear Model	Generalized Estimating Equations	
0.3	3	60	0.1	0.01010*	0.01123	11.18812
			0.5	0.01026*	0.01166	13.64522
			0.9	0.01240*	0.01449	16.85484
0.8			0.1	0.0114*	0.0129	13.15789
			0.5	0.0117*	0.0125	6.837607
			0.9	0.0132*	0.0141	6.818182

จากตารางที่ 4.17 เมื่อพิจารณาที่อัตตสัมพันธ์ของตัวแปรตามที่ 0.3 ที่ขนาดตัวอย่างเท่ากับ 60 ที่สัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระเท่ากับ0.1 ,0.5 และ 0.9 ระยะเวลาการเก็บข้อมูลซ้ำ เท่ากับ 3 พบว่าค่า AMSE ของตัวแบบ Generalized Linear Model มีค่าต่ำกว่าของตัวแบบ Generalized Estimating Equations โดยมีร้อยละความแตกต่างค่า AMSE ของ 2 ตัวแบบ คือ11.18812 , 13.64522 และ16.85484

เมื่อพิจารณาที่อัตตสัมพันธ์ของตัวแปรตามที่ 0.8 ที่ขนาดตัวอย่างเท่ากับ 60 ที่สัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระเท่ากับ 0.1 ,0.5และ 0.9 ระยะเวลาการเก็บข้อมูลซ้ำ เท่ากับ 3 พบว่าค่า AMSE ของตัวแบบ Generalized Linear Model มีค่าต่ำกว่าของตัวแบบ

Generalized Estimating Equations โดยมีร้อยละความแตกต่างค่า AMSE ของ 2 ตัวแบบ คือ 013.15789, 6.837607และ 6.818182ตามลำดับ

ทำการเปรียบเทียบทั้งสองอัตตสัมพันธ์พบว่า เมื่ออัตตสัมพันธ์เพิ่มขึ้นโดยส่วนใหญ่จะส่งผลให้ค่าAMSEของตัวแบบ Generalized Linear Modelต่ำกว่าของตัวแบบGeneralized Estimating Equations

ตารางที่ 4.18 : แสดงค่าAMSE ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ที่มีขนาดตัวอย่าง(n)เท่ากับ 60 ตัวแปรอิสระ (x)) 3 ตัวแปร ที่ระดับสัมประสิทธิ์สหสัมพันธ์(Rho (x)) 3 ระดับ ระยะเวลาที่เก็บข้อมูลซ้ำ (t) 6 ช่วงเวลา โดยที่ตัวแปรตาม (y)มีความสัมพันธ์ที่ 0.3 และ 0.8

อัตตสัมพันธ์ของตัวแปรตาม	t	n	Rho (x)	AMSE Generalized Linear Model	AMSE Generalized Estimating Equations	ร้อยละความ** แตกต่างค่าAMSE ของ2ตัวแบบ
0.3	6	60	0.1	0.00624	0.00602*	3.654485
			0.5	0.00631	0.00607*	3.953871
			0.9	0.00720	0.00700*	2.857143
0.8			0.1	0.0338*	0.0341	0.887574
			0.5	0.0342*	0.0346	1.169591
			0.9	0.0347*	0.0352	1.440922

จากตารางที่ 4.18 เมื่อพิจารณาที่อัตตสัมพันธ์ของตัวแปรตามที่ 0.3 ที่ขนาดตัวอย่างเท่ากับ 60 ที่สัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระเท่ากับ 0.1 ,0.5และ 0.9 ระยะเวลาการเก็บข้อมูลซ้ำ เท่ากับ 6 พบว่าค่า AMSE ของตัวแบบ Generalized Estimating Equationsมีค่าต่ำกว่าของตัวแบบ Generalized Linear Model โดยมีร้อยละความแตกต่างค่า AMSE ของ 2 ตัวแบบ คือ 3.654485 , 3.953871และ 2.857143 ตามลำดับ

เมื่อพิจารณาที่อัตตสัมพันธ์ของตัวแปรตามที่ 0.8 ที่ขนาดตัวอย่างเท่ากับ 60 ที่สัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระเท่ากับ0.1 ,0.5 และ 0.9 ระยะเวลาการเก็บข้อมูลซ้ำเท่ากับ 6 พบว่าค่า AMSE ของตัวแบบ Generalized Linear Model มีค่าต่ำกว่าของตัวแบบ Generalized Estimating Equations โดยมีร้อยละความแตกต่างความแตกต่างค่า AMSE ของ 2 ตัวแบบคือ 0.887574,1.169591และ1.440922 ตามลำดับ

ทำการเปรียบเทียบทั้งสองอัตตสัมพันธ์พบว่า เมื่อความสัมพันธ์เพิ่มขึ้นจะส่งผลให้ค่า AMSE ของตัวแบบGeneralized Linear Modelต่ำกว่าของตัวแบบ Generalized Estimating Equations

จากตารางที่ 4.17 และ ตารางที่ 4.18 พบว่า เมื่อขนาดตัวอย่างเท่ากับ 60 อัตตสัมพันธ์เท่ากับ 0.3 ระยะเวลาการเก็บข้อมูลซ้ำ เท่ากับ 3 ตัวแบบ Generalized Linear Modelมีค่าต่ำกว่า แต่เมื่อเพิ่มระยะเวลาการเก็บข้อมูลซ้ำเท่ากับ 6 ตัวแบบ Generalized Estimating Equationsต่ำกว่า ในขณะที่อัตตสัมพันธ์เท่ากับ 0.8 ระยะเวลาการเก็บข้อมูลซ้ำ เท่ากับ 3และ6 โดยส่วนใหญ่ตัวแบบ Generalized Linear Model

บทที่ 5
สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

วิทยานิพนธ์เรื่องนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบประสิทธิภาพการประมาณพารามิเตอร์ ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ด้วยวิธีการประมาณพารามิเตอร์แบบ Quasi-Likelihood สำหรับข้อมูลระยะยาว เมื่อกำหนดให้ตัวแปรตามมีการแจกแจงปัวซงส์ และเพื่อทำการเปรียบเทียบค่าเฉลี่ยของผลบวกกำลังสองของความคลาดเคลื่อนของสัมประสิทธิ์ความถดถอย โดยผลการวิจัยมีข้อสรุปดังนี้

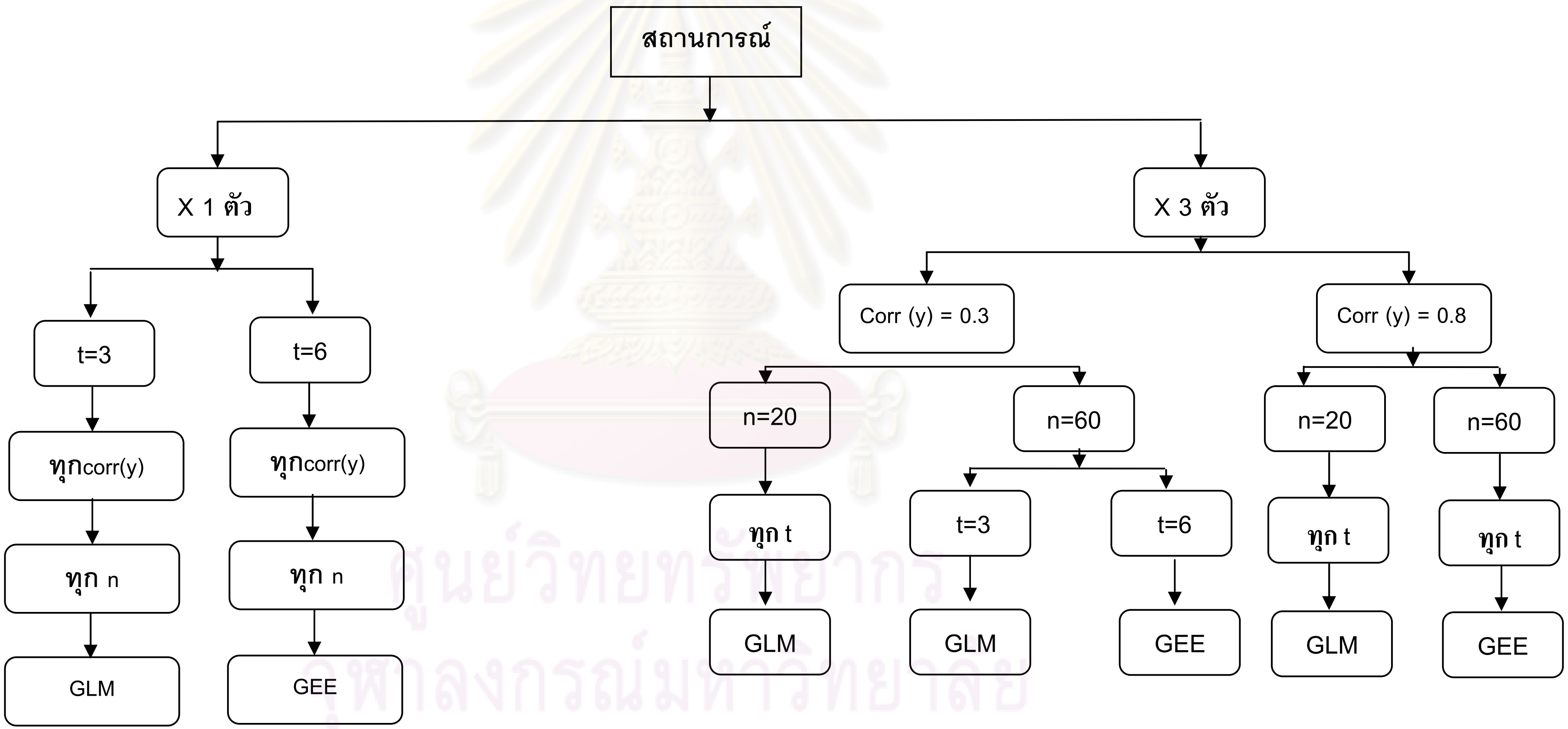
5.1 สรุปผลการวิจัย

จากการเปรียบเทียบค่า AMSE ของตัวแบบ Generalized Linear Model และตัวแบบ Generalized Estimating Equations สรุปได้ว่า

1. กรณีที่ตัวแปรอิสระ 1 ตัวแปร เมื่อระยะเวลาการเก็บข้อมูลซ้ำ 3 อัตตสัมพันธ์ของตัวแปรตามเพิ่มขึ้นจาก 0.3 จะได้ว่า ตัวแบบ Generalized Linear Model ให้การประมาณพารามิเตอร์ได้ดีกว่า
2. กรณีที่ตัวแปรอิสระ 1 ตัวแปร เมื่อระยะเวลาการเก็บข้อมูลซ้ำ 6 อัตตสัมพันธ์ของตัวแปรตามเพิ่มขึ้นจาก 0.8 จะได้ว่า ตัวแบบ Generalized Estimating Equations ให้การประมาณพารามิเตอร์ได้ดีกว่า
3. กรณีที่ตัวแปรอิสระ 3 ตัวแปร เมื่ออัตตสัมพันธ์ของตัวแปรตาม 0.3 ขนาดตัวอย่างเท่ากับ 20 ระยะเวลาการเก็บข้อมูลซ้ำเท่ากับ 3 และ 6 จะได้ว่า ตัวแบบ Generalized Linear Model ให้การประมาณพารามิเตอร์ได้ดีกว่า
4. กรณีที่ตัวแปรอิสระ 3 ตัวแปร เมื่ออัตตสัมพันธ์ของตัวแปรตาม 0.8 ขนาดตัวอย่างเท่ากับ 20 ระยะเวลาการเก็บข้อมูลซ้ำเท่ากับ 3 และ 6 จะได้ว่า ตัวแบบ Generalized Estimating Equations ให้การประมาณพารามิเตอร์ได้ดีกว่า
5. กรณีที่ตัวแปรอิสระ 3 ตัวแปร เมื่ออัตตสัมพันธ์ของตัวแปรตาม เท่ากับ 0.3 ขนาดตัวอย่างเท่ากับ 60 ระยะเวลาการเก็บข้อมูลซ้ำ 3 เป็น 6 ตัวแบบ Generalized Estimating Equationsได้ดีกว่า เมื่ออัตตสัมพันธ์ของตัวแปรตาม เท่ากับ 0.8 ระยะเวลาการเก็บข้อมูลซ้ำ เท่ากับ 3และ6โดยส่วนใหญ่ตัวแบบ Generalized Linear Model ได้ดีว่า

6. เมื่อมีการเพิ่มระดับสัมประสิทธิ์สหสัมพันธ์ ของตัวแปรอิสระมากขึ้นยิ่งทำให้เกิดความผิดพลาดในการประมาณค่าพารามิเตอร์เพิ่มขึ้นตามไปด้วย

รูปที่ 5.1 แสดงการเลือกใช้วิธีการประมารพารามิเตอร์ของตัวแบบ Generalized Linear Model และ ตัวแบบ Generalized Estimating Equations ด้วยวิธีการประมาณพารามิเตอร์แบบ Quasi-Likelihood สำหรับข้อมูลระยะยาว



47

หมายเหตุ จากรูปที่ 5.1

GLM หมายถึง ตัวแบบ Generalized Linear Model

GEE หมายถึง ตัวแบบ Generalized Estimating Equations

n หมายถึง ขนาดตัวอย่าง

Corr(y) หมายถึง อัตตสัมพันธ์ของตัวแปรตาม

t หมายถึง ระยะเวลาการเก็บข้อมลซ้ำ

5.2 ข้อเสนอแนะผลการวิจัยในครั้งนี้มีข้อเสนอแนะ 2 ด้าน คือ

5.2.1 ด้านการนำไปใช้ประโยชน์

1. เมื่อต้องการประมาณค่าพารามิเตอร์ที่เหมาะสมสำหรับตัวแบบตัวแบบ Generalized Linear Model และ ตัวแบบตัวแบบ Generalized Estimating Equations สามารถนำการประมาณนี้ ไปใช้ได้ตามแต่ละสถานการณ์
1. การวิจัยครั้งนี้ทำให้รู้ได้ว่าข้อมูลลักษณะไหนเหมาะกับตัวแบบใด
2. เมื่อเปรียบเทียบร้อยละความแตกต่างของทั้งสองตัวแบบ พบว่ามีบางสถานการณ์ที่มีร้อยละความแตกต่างของทั้งสองตัวแบบต่ำ ดังนั้น ในทางปฏิบัติควรเลือกตัวแบบที่สามารถประมาณพารามิเตอร์ได้ง่าย

5.2.2 ด้านการศึกษาวิจัย

1. ในการวิจัยครั้งนี้ได้ทำการศึกษาการเปรียบเทียบการประมาณพารามิเตอร์ของทั้งสองตัวแบบ โดยข้อมูลตัวแปรตามมีการแจกแจงแบบปัวซงส์ ในการวิจัยครั้งต่อไปอาจจะทำการเปรียบเทียบลักษณะเดียวกันเพียงแต่ข้อมูลของตัวแปรตามอาจจะมีการแจกแจงแบบ ต่างๆ เช่น ทวินาม หรือ แกมมา เป็นต้น
2. ในการวิจัยครั้งนี้ได้ทำการศึกษาการเปรียบเทียบการประมาณพารามิเตอร์ของทั้งสองตัวแบบ โดยข้อมูลตัวแปรตามมีอัตตสัมพันธ์แบบ Exchangeable การวิจัยครั้งต่อไปอาจจะให้ข้อมูลตัวแปรตามมีอัตตสัมพันธ์แบบอื่นๆ เช่น Autoregressive , Unstructure เป็นต้น

รายการอ้างอิง

ภาษาไทย

กัลยา วานิชย์บัญชา. การวิเคราะห์ข้อมูลหลายตัวแปร. พิมพ์ครั้งที่ 2. กรุงเทพมหานคร: สำนักพิมพ์ ธรรมสาร 2550

กานธนิกา ชุณหะวัต. การประมาณพารามิเตอร์สำหรับตัวแบบถดถอยเชิงเส้นปัวซงส์. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์ และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย, 2545

ศิริญญา ธีระอนันต์ชัย และ ลี่ลี อิงศรีสว่าง. ตัวแบบเชิงเส้นวางนัยทั่วไปสำหรับการศึกษาติดตามระยะยาวของจำนวนการเรียกค่าสินไหมทดแทนการประกันภัยรถยนต์ในกรุงเทพมหานคร. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์, 2552

ภาษาอังกฤษ

Kung-Yee Liang and Scott L. Zeger .Longitudinal Data Analysis Generalized Linear Models. Biometrika 1986:13-22

Gary A. Ballinger. Using Generalized Estimating Equations For Longitudinal Data Analysis. Organizational Research Method 2004: 127-150

P. David Wilson. Longitudinal Poisson Regression With Disturbed Random Intercept. Commun. Statist.-Theory Meth.27,9 (1998):2275-2292

ภาคผนวก

ตัวอย่างการใช้โปรแกรม R ในการดำเนินงานวิจัย

Main Program

```
# Correlation Matrix
cor.matrix <- function(p, corv=0){
   corMat <- matrix(corv, p, p)
   diag(corMat) <- 1
   output <- corMat
   return(output)   }
# Y generator with correlation matrix and loops
genY <- function(n, lambda=0.4, Rep=3, corv, error=0.1, nloop=1000){
   corMat  <- cor.matrix(Rep, corv)
   cholMat <- chol(corMat)
   num <- new("list")
   for(i in 1:nloop){
      stop.loop <- 0
      while(stop.loop<1){
         num[[i]] <- matrix(rpois(n*Rep, lambda), n)
         num[[i]] <- num[[i]]%*%cholMat
         num[[i]] <- round(num[[i]])
         if(sum(var(num[[i]])==0)<1){
            if(max(abs(cor(num[[i]])-corMat))<=error){
               stop.loop <- stop.loop+1   }
         }
         stop.loop }
   }
   output <- num
   return(output)
}
```

```
# X generator with correlation matrix and loops
genX <- function(n, Mean=0, SD=4, P=3, corv=0, error=0.1, nloop=1000){
   corMat  <- cor.matrix(P, corv)
   cholMat <- chol(corMat)
   num <- new("list")
   for(i in 1:nloop){
      stop.loop <- 0
      while(stop.loop<1){
         num[[i]] <- matrix(rnorm(n*P, Mean, SD), n)
         num[[i]] <- num[[i]]%*%cholMat
         if(sum(var(num[[i]])==0)<1){
            if(max(abs(cor(num[[i]])-corMat))<=error){
               stop.loop <- stop.loop+1  }
         }
         stop.loop  }
   }
   output <- num
   return(output)
}
# Merge data
mergeData <- function(x, y, Names=NULL){
   Length   <- length(x)
   dataTemp <- new("list")
   for(i in 1:Length){
      dataTemp[[i]] <- as.data.frame(cbind(c(1:length(y[[i]])), x[[i]], y[[i]]))
      if(!is.null(Names)){
         names(dataTemp[[i]]) <- Names
      }
   }
   output <- dataTemp
   return(output)   }
```

```
# Merge data for GEE
mergeDataGEE <- function(x, y, n, Rep, Names=NULL){
   Length   <- length(x)
   dataTemp <- new("list")
   for(i in 1:Length){
      dataTemp[[i]] <- as.data.frame(cbind(c(1:length(y[[i]])), x[[i]], y[[i]], rep(1:Rep,
each=n)))
      if(!is.null(Names)){
         names(dataTemp[[i]]) <- Names    }
   }
   output <- dataTemp
   return(output)
}
initLamb  <- 0.4    #define the lambda value
#iniBeat=[b0, b1(t), b2(x1), b3(x2), b4(x3)]
nLoop     <- 1000
seedNmber <- 1
glmFamily <- poisson
glmLink   <- "log"
geeFamily <- poisson
geeLink   <- "log"
corStruc  <- "exchangeable" #GEE exit with 'independence', 'exchangeable',
                    #'ar1', 'unstructured' and 'userdefined'
#Formula
P1F <- formula(y ~ Time + x1)
P3F <- formula(y ~ Time + x1 + x2 + x3)
```

```
# Generalized Linear Model + Generalized Estimating Equations

#GLM Case1: Y conditions: n=20 and t=3; X conditions: p=1, n=20*3, cor=0
set.seed(seedNmber)
gen60Y  <- genY(n=20, lambda=initLamb, Rep=3, corv=0.8, error=0.1, nLoop)
gen60Yt <- lapply(gen60Y, as.vector)
set.seed(seedNmber)
gen60Y00X <- genX(n=60, Mean=0, SD=4, P=1, corv=0, error=0.1, nLoop)
DTA60Y00X <- mergeData(gen60Y00X, gen60Yt, Names=c("Time", "x1", "y"))
fitGLM60Y00X  <- lapply(DTA60Y00X, glm, formula=P1F,
family=glmFamily(link=glmLink))
coefGLM60Y00X <- lapply(fitGLM60Y00X, coefficients)
betaGLMC01 <- matrix(unlist(coefGLM60Y00X),length(coefGLM60Y00X), 3, byrow=T)

library(geepack)  #Call library 'geepack'
#GEE Case1: Y conditions: n=20 and t=3; X conditions: p=1, n=20*3, cor=0
GEE60Y00X <- mergeDataGEE(gen60Y00X, gen60Yt, n=20, Rep=3,
   Names=c("Time", "x1", "y", "IDT"))
fitGEE60Y00X <- new("list")
for(i in 1:length(GEE60Y00X)){
   fitGEE60Y00X[[i]]  <- geeglm(P1F, data=GEE60Y00X[[i]], id=IDT,
      family=geeFamily(geeLink), corstr=corStruc)
}
coefGEE60Y00X <- lapply(fitGEE60Y00X, coefficients)
betaGEEC01 <- matrix(unlist(coefGEE60Y00X),length(coefGEE60Y00X), 3, byrow=T)
#GLM Case2: Y conditions: n=20 and t=3; X conditions: p=3, n=20*3, cor=0.1
set.seed(seedNmber)
gen60Y01X <- genX(n=60, Mean=0, SD=4, P=3, corv=0.1, error=0.1, nLoop)
DTA60Y01X <- mergeData(gen60Y01X, gen60Yt, Names=c("Time", "x1", "x2", "x3", "y"))
fitGLM60Y01X  <- lapply(DTA60Y01X, glm, formula=P3F,
family=glmFamily(link=glmLink))
```

```
coefGLM60Y01X <- lapply(fitGLM60Y01X, coefficients)
betaGLMC02 <- matrix(unlist(coefGLM60Y01X),length(coefGLM60Y01X), 5, byrow=T)

#GEE Case2: Y conditions: n=20 and t=3; X conditions: p=3, n=20*3, cor=0.1
GEE60Y01X <- mergeDataGEE(gen60Y01X, gen60Yt, n=20, Rep=3,
   Names=c("Time", "x1", "x2", "x3", "y", "IDT"))
fitGEE60Y01X <- new("list")
for(i in 1:length(GEE60Y01X)){
   fitGEE60Y01X[[i]]  <- geeglm(P3F, data=GEE60Y01X[[i]], id=IDT,
      family=geeFamily(geeLink), corstr=corStruc)
}
coefGEE60Y01X <- lapply(fitGEE60Y01X, coefficients)
betaGEEC02 <- matrix(unlist(coefGEE60Y01X),length(coefGEE60Y01X), 5, byrow=T)

#GLM Case3: Y conditions: n=20 and t=3; X conditions: p=3, n=20*3, cor=0.5
set.seed(seedNmber)
gen60Y05X <- genX(n=60, Mean=0, SD=4, P=3, corv=0.5, error=0.1, nLoop)
DTA60Y05X <- mergeData(gen60Y05X, gen60Yt, Names=c("Time", "x1", "x2", "x3", "y"))
fitGLM60Y05X  <- lapply(DTA60Y05X, glm, formula=P3F,
family=glmFamily(link=glmLink))
coefGLM60Y05X <- lapply(fitGLM60Y05X, coefficients)
betaGLMC03 <- matrix(unlist(coefGLM60Y05X),length(coefGLM60Y05X), 5, byrow=T)


#GEE Case3: Y conditions: n=20 and t=3; X conditions: p=3, n=20*3, cor=0.5
GEE60Y05X <- mergeDataGEE(gen60Y05X, gen60Yt, n=20, Rep=3,
   Names=c("Time", "x1", "x2", "x3", "y", "IDT"))
fitGEE60Y05X <- new("list")
for(i in 1:length(GEE60Y05X)){
   fitGEE60Y05X[[i]]  <- geeglm(P3F, data=GEE60Y05X[[i]], id=IDT,
      family=geeFamily(geeLink), corstr=corStruc)  }
```

```
coefGEE60Y05X <- lapply(fitGEE60Y05X, coefficients)
betaGEEC03 <- matrix(unlist(coefGEE60Y05X),length(coefGEE60Y05X), 5, byrow=T)
#GLM Case4: Y conditions: n=20 and t=3; X conditions: p=3, n=20*3, cor=0.9
set.seed(seedNmber)
gen60Y09X <- genX(n=60, Mean=0, SD=4, P=3, corv=0.9, error=0.1, nLoop)
DTA60Y09X <- mergeData(gen60Y09X, gen60Yt, Names=c("Time", "x1", "x2", "x3", "y"))
fitGLM60Y09X  <- lapply(DTA60Y09X, glm, formula=P3F,
family=glmFamily(link=glmLink))
coefGLM60Y09X <- lapply(fitGLM60Y09X, coefficients)
betaGLMC04 <- matrix(unlist(coefGLM60Y09X),length(coefGLM60Y09X), 5, byrow=T)

#GEE Case4: Y conditions: n=20 and t=3; X conditions: p=3, n=20*3, cor=0.9
GEE60Y09X <- mergeDataGEE(gen60Y09X, gen60Yt, n=20, Rep=3,
   Names=c("Time", "x1", "x2", "x3", "y", "IDT"))
fitGEE60Y09X <- new("list")
for(i in 1:length(GEE60Y09X)){
   fitGEE60Y09X[[i]]  <- geeglm(P3F, data=GEE60Y09X[[i]], id=IDT,
      family=geeFamily(geeLink), corstr=corStruc)
}
coefGEE60Y09X <- lapply(fitGEE60Y09X, coefficients)
betaGEEC04 <- matrix(unlist(coefGEE60Y09X),length(coefGEE60Y09X), 5, byrow=T)
```

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวศิรินทิพย์ เสริมสุข เกิดเมื่อวันศุกร์ ที่ 10 มกราคม พ.ศ. 2529 สำเร็จการศึกษาระดับปริญญาวิทยาศาสตรบัณฑิต (วท.บ.) สาขาสถิติประยุกต์ ภาควิชาสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ในปีการศึกษา 2550 และเข้าศึกษาต่อในหลักสูตรสถิติศาสตรมหาบัณฑิต (สต.ม.) สาขาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปี พ.ศ. 2551