



ใบรับรองวิทยานิพนธ์

บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์

วิทยาศาสตร์มหาบัณฑิต (สถิติ)

ปริญญา

สถิติ

สถิติ

สาขา

ภาควิชา

เรื่อง ตัวแบบเชิงเส้นวางนัยทั่วไปสำหรับการศึกษาดิตตามระยะยาวของจำนวนการเรียก
ค่าสินไหมทดแทนการประกันภัยรถยนต์ในกรุงเทพมหานคร ประเทศไทย

Generalized Linear Models for Longitudinal Study of Car Insurance Claims
in Bangkok, Thailand

นามผู้วิจัย นางสาวศิริกัญญา ชีระอนันต์ชัย

ได้พิจารณาเห็นชอบโดย

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ผู้ช่วยศาสตราจารย์ลีลี อิงศรีสว่าง, Ph.D.)

อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม

(รองศาสตราจารย์จิรัชย์ สุขะเกตุ, M.S.)

หัวหน้าภาควิชา

(อาจารย์อำไพ ทองธีรภาพ, Ph.D.)

บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์รับรองแล้ว

(รองศาสตราจารย์กัญญา ชีระกุล, D.Agr.)

คณบดีบัณฑิตวิทยาลัย

วันที่.....เดือน.....พ.ศ.....

วิทยานิพนธ์

เรื่อง

ตัวแบบเชิงเส้นวางนัยทั่วไปสำหรับการศึกษาดิตตามระยะยาวของจำนวนการเรียก
ค่าสินไหมทดแทนการประกันภัยรถยนต์ในกรุงเทพมหานคร ประเทศไทย

Generalized Linear Models for Longitudinal Study of Car Insurance Claims
in Bangkok, Thailand

โดย

นางสาวศิริกัญญา ชีระอนันต์ชัย

เสนอ

บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์
เพื่อขอความสมบูรณ์แห่งปริญญาวิทยาศาสตรมหาบัณฑิต (สถิติ)

พ.ศ. 2552

ศิริญา วีระอนันต์ชัย 2552: ตัวแบบเชิงเส้นวางนัยทั่วไปสำหรับการศึกษาดิตตามระยะยาวของ
จำนวนการเรียกค่าสินไหมทดแทนการประกันภัยรถยนต์ในกรุงเทพมหานคร ประเทศไทย ประโยชน์
วิทยาศาสตร์มหาบัณฑิต (สถิติ) สาขาสถิติ ภาควิชาสถิติ อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก:
ผู้ช่วยศาสตราจารย์ลีลี อิงศรีสว่าง, Ph.D. 116 หน้า

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างตัวแบบจำนวนการเรียกค่าสินไหมทดแทนการประกันภัยรถยนต์ใน
กรุงเทพมหานครที่มีการเก็บซ้ำในระยะเวลาติดตามตั้งแต่ปี พ.ศ. 2544-2548 โดยใช้ข้อมูลที่มีจำนวนการเรียก
ค่าสินไหมทดแทนอย่างน้อยหนึ่งครั้งในระยะเวลาติดตาม 5 ปี จำนวน 3,635 กรมธรรม์ ที่ได้จากการ
ประกันภัย ด้วยวิธี Generalized Estimating Equations (GEE) เมื่อกำหนดโครงสร้างความสัมพันธ์ของข้อมูลเป็น
แบบ First-order Autoregressive (AR(1)) และ Compound Symmetry (CS) ตามลำดับ และวิธีตัวแบบผสมเชิง
เส้นวางนัยทั่วไป (Generalized Linear Mixed Models, GLMMs) เมื่อกำหนดโครงสร้างความแปรปรวนร่วมของ
ข้อมูลเป็นแบบ AR(1) และ CS ตามลำดับ พร้อมทั้งหาตัวแบบที่เหมาะสม โดยตัวแปรตาม คือ จำนวนการเรียก
ค่าสินไหมทดแทนซึ่งมีการแจกแจงแบบปัวซอง และตัวแปรอิสระ คือ ค่าสินไหมทดแทน อายุรถยนต์ เพศ และ
กลุ่มอายุผู้ทำประกันภัย และทำการคัดเลือกตัวแปรด้วยวิธีขั้นตอนก้าวหน้า (Forward Stepwise Method)

ผลการศึกษาพบว่า ตัวแบบวิธี GEE เมื่อโครงสร้างความสัมพันธ์เป็นแบบ AR(1) มีความเหมาะสม
สำหรับข้อมูลมากกว่าการกำหนดโครงสร้างความสัมพันธ์เป็นแบบ CS ด้วยค่าสถิติ Pearson Chi-square of
residual/DF ที่มีค่าเข้าใกล้ 1 มากกว่าคือมีค่าเท่ากับ 0.64 และ 0.63 ตามลำดับ รวมทั้งค่าร้อยละของการทำนาย
ถูกต้อง (R_p^2) ที่มีค่าสูงกว่าคือมีค่าเท่ากับ 52.32 และ 52.05 ตามลำดับ ซึ่งปัจจัยที่มีผลต่อจำนวนการเรียกค่า
สินไหมทดแทนที่ระดับนัยสำคัญทางสถิติ 0.05 คือ ค่าสินไหมทดแทน อายุรถยนต์ กลุ่มอายุผู้ทำประกันภัย
อาชีพ และเพศ ส่วนตัวแบบวิธี GLMMs เมื่อกำหนดโครงสร้างความแปรปรวนร่วมเป็นแบบ AR(1) มีความ
เหมาะสมสำหรับข้อมูลมากกว่าการกำหนดโครงสร้างความแปรปรวนร่วมเป็นแบบ CS ด้วยค่าสถิติ
Generalized Chi-Square/DF เท่ากับ 0.18 และ 0.15 ตามลำดับ โดยมีค่า R_p^2 ใกล้เคียงกันคือ 53.34 และ 53.38
ตามลำดับ รวมทั้งค่า Pseudo-AIC และค่า Pseudo-BIC ของโครงสร้างความแปรปรวนร่วมแบบ AR(1) ที่มีค่าต่ำ
กว่าโครงสร้างความแปรปรวนร่วมแบบ CS เท่ากับ 36630.43 และ 36649.03 ตามลำดับ และปัจจัยที่มีผลต่อ
จำนวนการเรียกค่าสินไหมทดแทนที่ระดับนัยสำคัญทางสถิติ 0.05 คือ ค่าสินไหมทดแทน อายุรถยนต์ กลุ่มอายุ
ผู้ทำประกันภัย อาชีพ และเพศ เมื่อเปรียบเทียบตัวแบบวิธี GEE และตัวแบบวิธี GLMMs ในภาพรวมของ
ค่าเฉลี่ยประชากรจะมีค่า R_p^2 ใกล้เคียงกันประมาณร้อยละ 53 แต่เมื่อตัวแบบวิธี GLMMs ที่มีการพิจารณาเทอม
อิทธิพลค่าคงที่ผู้พบพบว่า ตัวแบบ GLMMs ที่มีโครงสร้างความแปรปรวนร่วมเป็นแบบ AR(1) ให้ประสิทธิภาพ
ในการทำนายจำนวนการเรียกค่าสินไหมทดแทนมีความถูกต้องสูงขึ้นเป็นร้อยละ 76.28

Sirinya Teeraananchai 2009: Generalized Linear Models for Longitudinal Study of Car Insurance Claims in Bangkok, Thailand. Master of Science (Statistics), Major Field: Statistics, Department of Statistics. Thesis Advisor: Assistant Professor Lily Ingsrisawang, Ph.D. 116 pages.

The goal of this study was to model a number of car insurance claims for car policyholders in Bangkok during the five-year insurance period of 2001-2005. The data used in this study came from the Department of Insurance in Thailand which consisted of 3,635 observations with at least one claim count in the 5-year period. The methodologies of GEE and GLMMs approaches were applied by taking account of correlation and covariance structures of data such as First-order Autoregressive (AR(1)) and Compound Symmetry (CS), respectively. The appropriate model is suggested. The dependent variable was the claim counts with Poisson distribution while independent variables were indemnity, car-age, age, gender, occupation, car-group and engine size. These independent variables were chosen to model the relationship between the claim counts and the potential predictors, using a forward stepwise procedure for variable selection.

The results showed that the GEE model with AR(1) correlation structure was more appropriate than the model with CS structure, as indicated by the values of Pearson Chi-square of residual/DF 0.64 and 0.63, respectively; and the percentages of correct prediction (R_p^2) with 52.32 and 52.05, respectively. The statistically significant factors at a 0.05 level consisted of indemnity, car-age, age-group, occupation and gender. On the other hand, the GLMMs model with AR(1) covariance structure was also more appropriate than the model with CS structure, as indicated by the values of Generalized Chi-Square/DF 0.18 and 0.15, respectively; and R_p^2 with 53.34 and 53.38, respectively. Moreover, the lower Pseudo-AIC and Pseudo-BIC of the GLMMs model with AR(1) structure was used to indicate the more effective covariance structure. The statistically significant factors at the 0.05 level consisted of indemnity, car-age, age-group, occupation and gender. The GLMMs and GEE for population averaged models had the same performance for estimating the claim counts with R_p^2 about 53 %, but the GLMMs model with AR(1) structure and random intercept effect showed more efficiency in predicting the number of claim counts with 76.28 %.

Student's signature

Thesis Advisor's signature

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยความกรุณาของผู้ช่วยศาสตราจารย์ ดร.ลีลี อิงศรีสว่าง อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก และรองศาสตราจารย์จิรัช สุชะเกตุ อาจารย์ที่ปรึกษา วิทยานิพนธ์ร่วม ที่ได้ให้คำแนะนำ ข้อคิดเห็น ตลอดจนช่วยเหลือในการวางแผนงานวิจัยและ ตรวจสอบแก้ไขข้อบกพร่องต่าง ๆ ผู้วิจัยขอขอบพระคุณรองศาสตราจารย์ประสิทธิ์ พัทธมพงษ์ ประธานกรรมการสอบวิทยานิพนธ์ และรองศาสตราจารย์ ดร.จุฬาลักษณ์ โกมลตรี ผู้ทรงคุณวุฒิ ภายนอก ที่กรุณาให้คำปรึกษาแนะนำและช่วยเหลือให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยดี

ผู้วิจัยขอขอบคุณพี่ ๆ ที่กรมการประกันภัย สำหรับความอนุเคราะห์ในด้านข้อมูลทั้งหมดที่ นำมาศึกษาทำวิทยานิพนธ์ให้มีความสมบูรณ์จนเป็นผลสำเร็จ

สุดท้ายผู้วิจัยขอกราบขอบพระคุณบิดา มารดา ที่ส่งเสริมและสนับสนุนการศึกษาของผู้วิจัย ตลอดจน ขอกราบขอบพระคุณครูบาอาจารย์ทุกท่านที่ได้แนะนำสั่งสอนประสิทธิ์ประสาทวิชา ความรู้ต่าง ๆ ให้ และขอขอบคุณเพื่อน ๆ พี่ ๆ และน้อง ๆ ทุกคนที่เป็นกำลังใจ ให้ความช่วยเหลือ และสนับสนุนมาโดยตลอดไว้ ณ โอกาสนี้ด้วย

ศิริัญญา ชีระอนันต์ชัย

ตุลาคม 2551

สารบัญ

	หน้า
สารบัญ	(1)
สารบัญตาราง	(2)
สารบัญภาพ	(5)
คำนำ	1
วัตถุประสงค์	4
การตรวจเอกสาร	6
อุปกรณ์และวิธีการ	40
อุปกรณ์	40
วิธีการ	40
ผลและวิจารณ์	52
ผล	52
วิจารณ์	100
สรุปและข้อเสนอแนะ	102
สรุป	102
ข้อเสนอแนะ	106
เอกสารและสิ่งอ้างอิง	108
ภาคผนวก	113
ประวัติการศึกษาและการทำงาน	116

สารบัญตาราง

ตารางที่		หน้า
1	ปริมาณรถยนต์นั่งส่วนบุคคลไม่เกิน 7 ที่นั่ง (คัน) ที่จดทะเบียนใหม่ทั้งหมดในประเทศไทยและในเขตกรุงเทพมหานครปี พ.ศ. 2544-2548	1
2	จำนวนอุบัติเหตุ (ครั้ง) จากการจราจรทางบกประเภทรถยนต์นั่งทั้งหมดในประเทศไทยและเฉพาะในเขตกรุงเทพมหานคร ปี พ.ศ. 2544-2548	2
3	โครงสร้างของข้อมูลการศึกษาระยะยาว	9
4	ลักษณะข้อมูลการประกันภัยรถยนต์ที่เฝ้าติดตามในระยะเวลา 5 ปี ตั้งแต่ปี พ.ศ. 2544-2548 จากตัวอย่างหน่วยศึกษาที่เป็นผู้ทำประกันภัยจำนวน 4 คน	42
5	ตัวแปรที่ใช้ในการศึกษาและการกำหนดรหัส	43
6	จำนวนขนาดตัวอย่างในแต่ละชุดข้อมูล	45
7	ลักษณะทั่วไปของข้อมูล	52
8	ลักษณะทั่วไปของข้อมูลเชิงปริมาณ	55
9	ค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนระหว่างตัวแปรกลุ่มแบบอันดับกับตัวแปรกลุ่มแบบอันดับ	66
10	ค่า Phi และ Cramer's V ระหว่างตัวแปรกลุ่มที่แทนลักษณะกับตัวแปรกลุ่มแบบอันดับ	67
11	ผลการวิเคราะห์ตัวแปรเดียวในตัวแบบวิธี GEE	69
12	ผลการวิเคราะห์ตัวแปรเดียวในตัวแบบวิธี GLMMs	70
13	ผลการคัดเลือกตัวแปรอิสระของข้อมูลฝึกสอน ด้วยค่าสถิติ Generalized Score Statistic (χ^2) ในตัวแบบ GEE เมื่อมีโครงสร้างความสัมพันธ์เป็นแบบ AR(1)	72
14	ผลการคัดเลือกตัวแปรอิสระของข้อมูลฝึกสอน ด้วยค่าสถิติ Generalized Score Statistic (χ^2) ในตัวแบบ GEE เมื่อมีโครงสร้างความสัมพันธ์เป็นแบบ CS	73

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
15	ผลการคัดเลือกตัวแปรอิสระในตัวแบบวิธี GLMMs จากการพิจารณาค่า -2 Residual Log Pseudo-Likelihood เมื่อกำหนดโครงสร้างความแปรปรวนร่วมเป็นแบบ AR(1)	74
16	ผลการคัดเลือกตัวแปรอิสระในตัวแบบวิธี GLMMs จากการพิจารณาค่า -2 Residual Log Pseudo-Likelihood เมื่อกำหนดโครงสร้างความแปรปรวนร่วมเป็นแบบ CS	75
17	ผลวิเคราะห์ตัวแบบวิธี GEE เมื่อกำหนดโครงสร้างความสัมพันธ์เป็นแบบ AR(1) เปรียบเทียบระหว่างชุดข้อมูลฝึกสอนและชุดข้อมูลทั้งหมด	75
18	ผลวิเคราะห์ตัวแบบวิธี GEE เมื่อกำหนดโครงสร้างความสัมพันธ์เป็นแบบ CS เปรียบเทียบระหว่างชุดข้อมูลฝึกสอนและชุดข้อมูลทั้งหมด	77
19	ผลวิเคราะห์ตัวแบบวิธี GLMMs เมื่อกำหนดโครงสร้างความแปรปรวนร่วมเป็นแบบ AR(1) เปรียบเทียบระหว่างชุดข้อมูลฝึกสอนและชุดข้อมูลทั้งหมด	80
20	ผลวิเคราะห์ตัวแบบวิธี GLMMs เมื่อกำหนดโครงสร้างความแปรปรวนร่วมเป็นแบบ CS เปรียบเทียบระหว่างชุดข้อมูลฝึกสอนและชุดข้อมูลทั้งหมด	82
21	ผลวิเคราะห์ตัวแบบวิธี GEE เมื่อเปรียบเทียบโครงสร้างความสัมพันธ์เป็นแบบ AR(1) และ CS ในชุดข้อมูลทั้งหมด	84
22	ผลวิเคราะห์ตัวแบบวิธี GLMMs เมื่อเปรียบเทียบโครงสร้างความแปรปรวนร่วมเป็นแบบ AR(1) และ CS จากข้อมูลทั้งหมด	89
23	ผลการวิเคราะห์ตัวแบบสุดท้ายที่เหมาะสมสำหรับตัวแบบวิธี GEE เมื่อกำหนดโครงสร้างความสัมพันธ์เป็นแบบ AR(1)	95
24	ผลการวิเคราะห์ตัวแบบที่เหมาะสมด้วยวิธี GLMMs ที่มีโครงสร้างความแปรปรวนร่วมเป็นแบบ AR(1)	97

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
25	ปัจจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนของตัวแบบแต่ละวิธี	104
ตารางผนวกที่		
1	การแบ่งกลุ่มรถยนต์ (เฉพาะรถยนต์นั่ง รหัส 110 120 เท่านั้น)	114
2	ลักษณะข้อมูลการจัดกลุ่มของตัวแปรค่าสินไหมทดแทนและอายุรถยนต์ตั้งแต่ปี พ.ศ. 2544-2548	115

สารบัญภาพ

ภาพที่	หน้า
1 ขั้นตอนการพัฒนาตัวแบบทางสถิติ	50
2 การแจกแจงข้อมูลของค่าสินไหมทดแทนจากข้อมูลทั้งหมด	60
3 การแจกแจงข้อมูลของค่าสินไหมทดแทนจากข้อมูลฝึกสอน	60
4 การแจกแจงข้อมูลของค่าสินไหมทดแทนจากข้อมูลทดสอบ	61
5 การแจกแจงข้อมูลค่าสินไหมทดแทนที่แปลงข้อมูลของข้อมูลทั้งหมด	62
6 การแจกแจงข้อมูลค่าสินไหมทดแทนที่แปลงข้อมูลของข้อมูลฝึกสอน	63
7 การแจกแจงข้อมูลค่าสินไหมทดแทนที่แปลงข้อมูลของข้อมูลทดสอบ	63
8 ความสัมพันธ์ระหว่างจำนวนการเรียกค่าสินไหมกับค่าสินไหมทดแทนที่แปลงข้อมูลในชุดข้อมูลทั้งหมด	64
9 ความสัมพันธ์ระหว่างจำนวนการเรียกค่าสินไหมกับค่าสินไหมทดแทนที่แปลงข้อมูลในชุดข้อมูลฝึกสอน	64
10 ความสัมพันธ์ระหว่างจำนวนการเรียกค่าสินไหมกับค่าสินไหมทดแทนที่แปลงข้อมูลในชุดข้อมูลทดสอบ	65
11 ความสัมพันธ์ของ residuals กับ $\hat{Y}$ (predicted) เมื่อมีโครงสร้างความสัมพันธ์เป็นแบบ AR(1)	88
12 ความสัมพันธ์ของ residuals กับ $\hat{Y}$ (predicted) เมื่อมีโครงสร้างความสัมพันธ์เป็นแบบ CS	88
13 ความสัมพันธ์ของ residuals กับ $\hat{Y}$ (predicted) เมื่อมีโครงสร้างความแปรปรวนร่วมเป็นแบบ AR(1)	94
14 ความสัมพันธ์ของ residuals กับ $\hat{Y}$ (predicted) เมื่อมีโครงสร้างความแปรปรวนร่วมเป็นแบบ CS	94

คำอธิบายสัญลักษณ์และคำย่อ

GLMs	=	ตัวแบบเชิงเส้นวางนัยทั่วไป (Generalized Linear Models)
GEE	=	Generalized Estimating Equations
LMMs	=	ตัวแบบผสมเชิงเส้น (Linear Mixed Models)
GLMMs	=	ตัวแบบผสมเชิงเส้นวางนัยทั่วไป (Generalized Linear Mixed Models)
AR(1)	=	First-order Autoregressive
CS	=	Compound Symmetry
DF	=	องศาแห่งความเป็นอิสระ (Degree of Freedom)
X	=	ตัวแปร
$\underline{X}$	=	เวกเตอร์ของตัวแปร
$\mathbf{X}$	=	เมทริกซ์ของตัวแปร

ตัวแบบเชิงเส้นวางนัยทั่วไปสำหรับการศึกษาติดตามระยะยาวของจำนวนการเรียก
ค่าสินไหมทดแทนการประกันภัยรถยนต์ในกรุงเทพมหานคร ประเทศไทย

Generalized Linear Models for Longitudinal Study of Car Insurance Claims
in Bangkok, Thailand

คำนำ

ปัจจุบันรถยนต์เข้ามามีบทบาทต่อการดำเนินชีวิตประจำวันของมนุษย์ และมีปริมาณเพิ่มขึ้นอย่างรวดเร็ว จึงเป็นตัวกระตุ้นให้มีการพัฒนาการขนส่งมวลชน การขนส่งสินค้า และเส้นทางคมนาคมทางบกให้มีความสะดวกมากขึ้น การใช้รถยนต์เป็นพาหนะเดินทางและใช้ในการขนส่งสินค้าทางการเกษตรและอุตสาหกรรมจึงเป็นที่นิยม สำหรับการใช้อยู่อาศัยในการเดินทางนั้น พบว่ามีการใช้รถยนต์นั่งส่วนบุคคลเป็นจำนวนมากขึ้นตามลำดับ โดยเฉพาะปริมาณรถยนต์ที่มีการจดทะเบียนใหม่มีมากขึ้นทุกปีตามรายงานสถิติที่แสดงในตารางที่ 1

ตารางที่ 1 ปริมาณรถยนต์นั่งส่วนบุคคลไม่เกิน 7 ที่นั่ง (คัน) ที่จดทะเบียนใหม่ทั้งหมดในประเทศไทยและในเขตกรุงเทพมหานครปี พ.ศ. 2544-2548

พื้นที่	ปี พ.ศ.				
	2544	2545	2546	2547	2548
ประเทศไทย	135,625	180,728	231,030	284,813	314,508
กรุงเทพมหานคร	92,064	118,564	146,945	176,933	188,936

ที่มา กรมการขนส่งทางบก กระทรวงคมนาคม (2549)

อย่างไรก็ตามปริมาณรถยนต์ที่เพิ่มขึ้นนำไปสู่ความสูญเสียที่เกิดจากอุบัติเหตุ ทำให้มีผลกระทบด้านลบต่อเศรษฐกิจและสังคม การสูญเสียทรัพยากรทั้งชีวิตและทรัพย์สิน หน่วยงานและองค์กรต่าง ๆ จึงได้พยายามหาทางป้องกัน เพื่อลดอุบัติเหตุจากการใช้รถยนต์ในทุกวิถีทางไม่ว่าจะเป็นการใช้เทคโนโลยีใหม่ๆ ในการผลิตรถยนต์ เช่น ระบบเบรกอัตโนมัติ (Automatic Brake

System :ABS) อุณหภูมิรัย และการก่อสร้างถนนที่ถูกต้องตามหลักวิศวกรรมจราจร การควบคุมผู้ใช้รถใช้ถนนให้ปฏิบัติตามกฎจราจร รวมทั้งการจัดตั้งหน่วยงานที่มีหน้าที่รับผิดชอบในการป้องกันอุบัติเหตุอันเกิดจากการใช้รถยนต์ แต่ก็ยังไม่สามารถป้องกันการเกิดอุบัติเหตุได้ ดังจะเห็นได้จากรายงานสถิติการเกิดอุบัติเหตุทางรถยนต์นั่งในตารางที่ 2

ตารางที่ 2 จำนวนอุบัติเหตุ (ครั้ง) จากการจราจรทางบกประเภทรถยนต์นั่งทั้งหมดในประเทศไทย และเฉพาะในเขตกรุงเทพมหานคร ปี พ.ศ. 2544-2548

พื้นที่	ปี พ.ศ.				
	2544	2545	2546	2547	2548
ประเทศไทย	33,907	39,279	41,949	46,658	43,538
กรุงเทพมหานคร	26,828	28,249	30,993	31,384	28,705

ที่มา สำนักงานตำรวจแห่งชาติ (2549)

การประกันภัยจึงเข้ามามีบทบาทช่วยบรรเทาปัญหาที่เกิดจากอุบัติเหตุ เช่น ความไม่ปลอดภัยสำหรับผู้ขับขี่รถยนต์ในการเดินทางและความเสียหายจากการเสี่ยงภัยของรถยนต์ การประกันภัยรถยนต์สำหรับผู้ที่ใช้รถยนต์ในการเดินทางโดยเฉพาะการประกันภัยรถยนต์ภาคสมัครใจที่คุ้มครองความเสี่ยงที่เกิดจากอุบัติเหตุทั้งผู้ขับขี่และความเสียหายของทรัพย์สิน ในระบบการประกันภัยรถยนต์ กรมการประกันภัยได้มีการเก็บรวบรวมข้อมูลการประกันภัยรถยนต์ในแต่ละปีเพื่อใช้ในการวางแผน และกำหนดนโยบายที่เกี่ยวกับการประกันภัยรถยนต์ ตลอดจนใช้เป็นแนวทางในการคำนวณอัตราเบี้ยประกันรถยนต์ (Premium Rate) สำหรับปีต่อไป โดยพิจารณาจากข้อมูลจำนวนการเรียกค่าสินไหมทดแทนในปีที่ผ่านมาของผู้ทำประกันภัย เนื่องจากในการประกันภัยรถยนต์นั้นไม่สามารถคาดการณ์จำนวนการเรียกค่าสินไหมทดแทนที่เกิดจากอุบัติเหตุหรือความเสียหายที่เกี่ยวข้องกับการประกันภัยภายในระยะเวลา 1 ปีของแต่ละบุคคลได้ ข้อมูลจำนวนการเรียกค่าสินไหมทดแทนจึงมักจะถูกใช้เป็นปัจจัยหนึ่งในการกำหนดอัตราเบี้ยประกันปีต่อไป ลักษณะการเก็บข้อมูลซ้ำของจำนวนการเรียกค่าสินไหมทดแทนอย่างต่อเนื่องกันทุกปีของผู้ทำประกันภัยคนเดิม จะเรียกเป็นข้อมูลที่มีการติดตามระยะยาว (Longitudinal Data)

ในทางสถิติได้มีการศึกษาตัวแบบสำหรับข้อมูลจำนวนการเรียกค่าสินไหมทดแทนใน

การประกันภัยรถยนต์ ณ เวลาใดเวลาหนึ่ง ซึ่งเป็นข้อมูลการนับ (Count) ที่มีลักษณะการแจกแจงแบบปัวซอง (Poisson Distribution) (Denuit, 1997) โดยได้มีการนำหลักการของตัวแบบเชิงเส้นวางนัยทั่วไป (Generalized Linear Models, GLMs) มาใช้พัฒนาตัวแบบข้อมูลจำนวนการเรียกค่าสินไหมทดแทน (Stroinski and Currie, 1989) แต่เนื่องจากวิธี GLMs มีข้อดกลงเบื้องต้นว่าค่าสังเกตของตัวแปรตามหรือตัวแปรที่สนใจจะต้องเป็นอิสระต่อกันในช่วงเวลาที่ศึกษา ทำให้สามารถอธิบายผลการวิเคราะห์ได้ในภาพรวม แต่สำหรับการวิเคราะห์ข้อมูลที่มีการติดตามในระยะยาว อย่างเช่น ข้อมูลการประกันภัยรถยนต์เป็นข้อมูลที่มีการเก็บบันทึกอย่างต่อเนื่องทุกปีของผู้ทำประกันภัยรถยนต์แต่ละคน ทำให้ค่าสังเกตข้อมูลที่มีการเก็บซ้ำในแต่ละปีของบุคคลเดียวกันมีความสัมพันธ์กันเกิดขึ้น ดังนั้นในปี ค.ศ. 1986 Liang and Zeger ได้เสนอวิธี Generalized Estimating Equations (GEE) ที่พัฒนาต่อมาจากวิธี GLMs สำหรับการสร้างตัวแบบภาพรวมของค่าเฉลี่ยประชากร (Population-Average: PA) โดยได้มีการนำโครงสร้างความสัมพันธ์ที่สมมติขึ้นสำหรับข้อมูลที่เก็บซ้ำของหน่วยศึกษาหนึ่ง ๆ กำหนดเป็น Working Correlation เข้ามาพิจารณาในขั้นตอนของการประมาณค่าพารามิเตอร์ นอกจากนี้ข้อมูลที่มีการเก็บซ้ำแบบติดตามระยะยาวสามารถวิเคราะห์ด้วยวิธีตัวแบบผสมเชิงเส้นวางนัยทั่วไป (Generalized Linear Mixed Models, GLMMs) ซึ่งเป็นตัวแบบที่อธิบายแต่ละหน่วยบุคคลหรือหน่วยศึกษา (Subject-Specific Models) โดยตัวแบบนี้พัฒนามาจากตัวแบบภาพรวมค่าเฉลี่ยประชากร (PA Model) และมีการเพิ่มเทอมอิทธิพลสุ่ม (Random Effect) เพื่อแสดงลักษณะเฉพาะของแต่ละบุคคล เช่น การยอมให้ค่าพื้นฐานหรือที่เรียกว่าค่าคงที่ (Intercept) ของแต่ละบุคคลมีค่าแตกต่างกันได้แบบสุ่ม เรียกเป็น Random Intercept เพื่อให้อธิบายความแตกต่างของแต่ละบุคคลได้ชัดเจนกว่าการอธิบายผลในภาพรวมของค่าเฉลี่ยประชากร

ดังนั้น การศึกษานี้จึงสนใจพัฒนาตัวแบบทางสถิติสำหรับข้อมูลจำนวนการเรียกค่าสินไหมทดแทนของการประกันภัยรถยนต์ที่มีการเก็บข้อมูลซ้ำแต่ละปีในระยะเวลาติดตาม 5 ปี ของผู้ทำประกันภัยคนเดียวกัน ที่ทำประกันภัยรถยนต์ภาคสมัครใจ ประเภท 1 แบบระบุชื่อผู้ขับขี่ รหัส 110 รถยนต์นั่งส่วนบุคคล เฉพาะรถยนต์ที่จดทะเบียนในกรุงเทพมหานคร เมื่อตัวแปรตามที่สนใจ คือ จำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ต่อปีที่มีการแจกแจงแบบปัวซอง (Poisson Distribution) ด้วยวิธี GEE และวิธี GLMMs พร้อมทั้งพิจารณาตัวแบบที่เหมาะสมกับข้อมูล เพื่อเป็นประโยชน์ต่อหน่วยงานที่เกี่ยวข้องสามารถนำวิธีการพัฒนาตัวแบบไปใช้กับข้อมูลในช่วงเวลาต่อไป จะได้ผลลัพธ์ข้อมูลที่ทันสมัยและใช้เป็นข้อมูลสนับสนุนในการวางแผนการ

กำหนดอัตราเบี้ยประกันภัยรถยนต์ รวมทั้งสามารถเป็นแนวทางในการนำวิธีทางสถิติมาประยุกต์ใช้กับข้อมูลการประกันภัยอื่น ๆ เมื่อตัวแปรตามที่น่าสนใจเป็นข้อมูลจำนวนนับที่มีการเก็บข้อมูลซ้ำ

วัตถุประสงค์

1. พัฒนาตัวแบบด้วยวิธี GEE และวิธี GLMMs สำหรับข้อมูลจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ที่มีการเก็บข้อมูลซ้ำในระยะเวลาติดตาม 5 ปี เมื่อสมมติโครงสร้างความสัมพันธ์ของข้อมูลในหน่วยศึกษามีรูปแบบ First-order Autoregressive และ Compound Symmetry
2. เพื่อหาตัวแบบทางสถิติที่เหมาะสมระหว่างวิธี GEE กับวิธี GLMMs สำหรับข้อมูลจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ที่มีการเก็บข้อมูลซ้ำในระยะเวลาติดตาม 5 ปี

ประโยชน์ที่คาดว่าจะได้รับ

1. หน่วยงานที่เกี่ยวข้องสามารถนำผลที่ได้จากการศึกษานี้ไปใช้ในการวางแผนการคำนวณอัตราเบี้ยประกันภัยรถยนต์ได้
2. ได้ตัวแบบทางสถิติที่เหมาะสมสำหรับการประมาณจำนวนการเรียกค่าสินไหมทดแทนการประกันภัยรถยนต์ และสามารถนำไปใช้เป็นแนวทางการศึกษาความสัมพันธ์ระหว่างปัจจัยส่วนบุคคลที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนในช่วงเวลาต่อไป
3. สามารถใช้วิธีทางสถิติจากการศึกษาไปประยุกต์กับข้อมูลการประกันภัยอื่น ๆ ที่ตัวแปรตามเป็นข้อมูลจำนวนการเรียกค่าสินไหมทดแทนที่มีการติดตามวัดซ้ำ

ขอบเขตการวิจัย

1. การศึกษานี้กำหนดขอบเขตการหาตัวแบบทางสถิติสำหรับจำนวนการเรียกค่าสินไหม

ทดแทนในการประกันภัยรถยนต์ ภาคสมัครใจ ประเภท 1 แบบระบุชื่อผู้ขับขี่ รหัส 110 รถยนต์นั่งส่วนบุคคล ที่จดทะเบียนในกรุงเทพมหานครเท่านั้น ตั้งแต่ปี พ.ศ. 2544 ถึงปี พ.ศ. 2548

2. การออกแบบการศึกษา (Study Design) เป็นการศึกษาติดตามระยะยาวของข้อมูลจำนวนการเรียกค่าสินไหมทดแทนของผู้ทำประกันภัยรถยนต์ภาคสมัครใจ ประเภท 1 แบบระบุชื่อผู้ขับขี่ รหัส 110 รถยนต์นั่งส่วนบุคคล ที่จดทะเบียนในกรุงเทพมหานคร โดยเริ่มทำประกันในปี พ.ศ. 2544 และติดตามการทำประกันภัยของผู้ทำประกันภัยรายหนึ่ง ๆ เป็นระยะเวลา 5 ปี ตั้งแต่ปี พ.ศ. 2544 ถึงปี พ.ศ. 2548

3. ประชากรที่ศึกษา คือ ผู้ทำประกันภัยรถยนต์ภาคสมัครใจ ประเภท 1 ที่จดทะเบียนในเขตกรุงเทพมหานคร

4. ตัวอย่างที่ศึกษา คือ ผู้ทำประกันภัยรถยนต์ภาคสมัครใจ ประเภท 1 ทั้งหมดที่จดทะเบียนกรุงเทพมหานครตั้งแต่ปี พ.ศ. 2544 และติดตามจนถึงปี พ.ศ. 2548

5. ข้อมูลและแหล่งข้อมูล คือ ข้อมูลการประกันภัยรถยนต์ ภาคสมัครใจ ประเภท 1 แบบระบุชื่อผู้ขับขี่ รหัส 110 รถยนต์นั่งส่วนบุคคล ที่จดทะเบียนในกรุงเทพมหานครในช่วง 5 ปี ตั้งแต่ปี พ.ศ. 2544 ถึงปี พ.ศ. 2548 โดยรวบรวมได้จากกรมประกันภัย กระทรวงพาณิชย์

6. การสร้างตัวแบบด้วยวิธี GEE และวิธี GLMMs สำหรับศึกษาความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระในที่นี้จะให้ความสำคัญกับตัวแปรอิทธิพลหลัก (Main Effects) เท่านั้น ซึ่งวิธี GEE จะกำหนดโครงสร้างความสัมพันธ์ของข้อมูลที่มีการวัดซ้ำภายในหน่วยศึกษาหนึ่ง ๆ เป็นแบบ First-order Autoregressive (AR(1)) และ Compound Symmetry (CS) ส่วนวิธี GLMMs จะกำหนดโครงสร้างความแปรปรวนร่วมของข้อมูลเป็นแบบ First-order Autoregressive และ Compound Symmetry

การตรวจเอกสาร

การตรวจเอกสารจะแบ่งออกเป็น 3 ส่วน คือ ส่วนแรกจะกล่าวถึงประเภทของการประกันภัยรถยนต์ (กรมการประกันภัย, 2550) ส่วนที่สองจะกล่าวถึงทฤษฎีสถิติที่เกี่ยวข้อง และ ส่วนที่สามจะกล่าวถึงผลงานวิจัยที่เกี่ยวข้อง

ส่วนที่ 1 ประเภทของการประกันภัยรถยนต์

กรมการประกันภัยได้แบ่งประกันภัยรถยนต์ออกเป็น 2 ประเภท คือ

1. ประกันภัยรถยนต์ภาคบังคับ (Compulsory Motor Insurance) หมายถึงการประกันภัยรถยนต์ประเภทที่กฎหมายให้เจ้าของรถซึ่งใช้หรือมีรถไว้เพื่อใช้ ต้องจัดให้มีการประกันความเสียหายสำหรับ ผู้ประสบภัยโดยประกันภัยกับบริษัทตามกฎหมายว่าด้วยการประกันวินาศภัยที่ได้รับอนุญาตให้ประกอบกิจการประเภทการประกันภัยรถ โดยรัฐบาลมีเจตจำนงเพื่อให้เกิดความคุ้มครองแก่ชีวิต ร่างกายของประชาชนที่ประสบภัยเป็นสำคัญ

2. การประกันภัยรถยนต์ภาคสมัครใจ (Voluntary Motor Insurance) เป็นการประกันภัยรถที่กฎหมายไม่ได้บังคับ ขึ้นอยู่กับความพอใจของผู้ทำประกันภัยที่เห็นถึงความเสี่ยงภัยแห่งตนและมีความคิดที่จะกระจายความเสี่ยงภัยออกไปยังบุคคลอื่นคือการประกันภัยไว้กับบริษัทผู้รับประกันภัย ซึ่งผู้ทำประกันภัยสามารถเลือกซื้อความคุ้มครองตามประเภทที่ผู้ทำประกันภัยประสงค์ โดยบริษัทประกันภัยจะออกกรมธรรม์ประกันภัยให้ไว้เป็นหลักฐาน โดยมีความคุ้มครอง เงื่อนไข และ ข้อยกเว้นตามแบบที่นายทะเบียนให้ความเห็นชอบ การประกันภัยรถยนต์ภาคสมัครใจแบ่งออกเป็น 3 ประเภทหลัก คือ

1. การประกันภัยประเภท 1 บริษัทจะจ่ายค่าสินไหมทดแทน ดังนี้

1.1 คุ้มครองความรับผิดชอบต่อชีวิตร่างกายบุคคลภายนอกรวมถึงผู้โดยสารในรถยนต์คันที่เอาประกันภัยด้วย

1.2 คุ้มครองความรับผิดชอบต่อทรัพย์สินของบุคคลภายนอก

1.3 คຸ້ມครองความเสียหายต่อตัวรถยนต์

1.4 คຸ້ມครองความสูญหายและไฟไหม้ของตัวรถยนต์

2. การประกันภัยประเภท 2 คຸ້ມครองเฉพาะความรับผิดชอบบุคคลภายนอก และความสูญหายและไฟไหม้ของตัวรถยนต์ ตามข้อ 1.1, 1.2 และ 1.4 ที่ระบุไว้ในประเภท 1

3. การประกันภัยประเภท 3 คຸ້ມครองเฉพาะความรับผิดชอบบุคคลภายนอก ตามข้อ 1 และ 2 ที่ระบุไว้ในการประกันภัยรถยนต์ประเภท 1

การประกันภัยรถยนต์ภาคสมัครใจนั้นถือว่าเป็นส่วนสำคัญของธุรกิจประกันภัยรถยนต์ สำหรับเป็นทางเลือกในการประกันภัยรถยนต์ ซึ่งการประกันภัยรถยนต์ยังแบ่งออกเป็น 2 กรณี คือ

1. การประกันภัยประเภทไม่ระบุชื่อผู้ขับขี่ การประกันภัยรถยนต์ประเภทนี้ เป็นการประกันภัยที่เข้ามาแต่ดั้งเดิมในประเทศไทย ซึ่งจะคຸ້ມครองความรับผิดชอบและความเสียหายต่อรถยนต์ที่เกิดขึ้นในระหว่างการใช้หรือการขับขี่ของบุคคลใด ๆ ก็ตามที่ใช้หรือขับขี่โดยได้รับความยินยอมจากผู้ทำประกันภัย ทั้งนี้ภายใต้หลักเกณฑ์ เงื่อนไข และข้อยกเว้น ที่ปรากฏในกรมธรรม์ประกันภัย

2. การประกันภัยประเภทระบุชื่อผู้ขับขี่ การประกันภัยรถยนต์ประเภทนี้ เป็นการประกันภัยที่มีการพัฒนาขึ้นมาใหม่ในประเทศไทย โดยหลักการที่สำคัญสำหรับการประกันภัยกรณีนี้ คือ จะคຸ້ມครองความรับผิดชอบหรือความเสียหายต่อรถยนต์ที่เกิดขึ้น ในขณะที่มีบุคคลที่ระบุชื่อในกรมธรรม์เป็นผู้ขับขี่ ในการทำประกันรถยนต์ประเภทระบุชื่อผู้ขับขี่นั้น ผู้ทำประกันภัยสามารถระบุชื่อผู้ขับขี่ที่ได้รับความคุ้มครองได้ถึง 2 คน แต่จะระบุคนเดียวก็ได้ ในกรณีที่ระบุ 2 คนนั้น การคิดคำนวณเบี้ยประกันภัย จะใช้ผู้ขับขี่ที่มีความเสี่ยงสูงเป็นฐานในการคำนวณเบี้ยประกันภัย ซึ่งปัจจัยที่ใช้เป็นตัวกำหนดความเสี่ยงสำหรับการประกันภัยระบุชื่อผู้ขับขี่ เพิ่มเติมจากการประกันภัยประเภทไม่ระบุชื่อผู้ขับขี่ก็คือ อายุของผู้ขับขี่ โดยมีการแบ่งช่วงอายุของผู้ขับขี่จากช่วงที่มีความเสี่ยงน้อยไปยังช่วงที่มีความเสี่ยงมากเป็น 4 ช่วงอายุ ดังนี้

- อายุเกิน 50 ปีขึ้นไป
- อายุ 36-50 ปี
- อายุ 25-35 ปี

- อายุ 18-24 ปี

สำหรับรถยนต์ที่ผู้ทำประกันภัยจะนำมาทำประกันภัยประเภทระบุชื่อผู้ขับขี่ได้นั้น จะต้องเป็นรถยนต์ที่ใช้เป็นส่วนบุคคลเท่านั้น รถยนต์ที่ใช้รับจ้างสาธารณะหรือใช้เพื่อการพาณิชย์ หรือใช้เพื่อการพาณิชย์พิเศษ ไม่สามารถทำประกันภัยประเภทดังกล่าวได้ ดังนั้นรถยนต์ที่จะสามารถทำประกันภัยประเภทระบุชื่อผู้ขับขี่ได้ จึงจำกัดไว้เพียง 3 ประเภทเท่านั้น คือ รถยนต์นั่งส่วนบุคคล รถยนต์โดยสารส่วนบุคคล และรถจักรยานยนต์ส่วนบุคคล

ส่วนที่ 2 ทฤษฎีสถิติที่เกี่ยวข้อง

ข้อมูลการติดตามระยะยาว เป็นข้อมูลที่มีการเก็บซ้ำ (Repeated Measures) จากหน่วยศึกษาเดิม โดยมีการเก็บข้อมูล ณ ช่วงเวลาต่าง ๆ การเก็บข้อมูลซ้ำลักษณะนี้จะพบในการศึกษาทางการแพทย์ การเกษตร และการประกันภัย เป็นต้น โดยค่าสังเกตของตัวแปรตาม (Y) จากแต่ละหน่วยศึกษา (Subject) จะมีการเก็บซ้ำตามเวลาที่กำหนด ส่วนตัวแปรอิสระ (X) อาจจะเป็นตัวแปรที่มีค่าคงที่ในทุกช่วงเวลาเช่น เพศ กลุ่มรถยนต์ และอาชีพ เป็นต้น หรือเป็นตัวแปรที่มีค่าเปลี่ยนแปลงตามเวลาเช่นเดียวกับตัวแปรตามก็ได้ เช่น อายุ เป็นต้น ลักษณะโครงสร้างของข้อมูลการศึกษาระยะยาวมีรูปแบบ ดังตารางที่ 3 (Diggle *et al.*, 2002)

เมื่อ y_{ij} คือ ค่าสังเกตของตัวแปรตาม Y จากหน่วยศึกษาที่ i ณ เวลาที่ j และตัวแปรอิสระมีจำนวน k ตัว คือ $X_1 \ X_2 \ \dots \ X_k$ โดยที่ x_{ijk} คือ ค่าสังเกตของตัวแปร X_k จากหน่วยศึกษาที่ i ณ เวลา j โดยที่ $i = 1, 2, \dots, S$ และ $j = 1, 2, \dots, t$ ส่วน S คือ จำนวนหน่วยศึกษา ตัวแบบทางสถิติสำหรับข้อมูลศึกษาระยะยาวอาจพิจารณาจาก 2 ตัวแบบ (Diggle *et al.*, 2002) ดังนี้

1. Marginal Models หรือ Population Average Models ตัวแบบนี้มีเป้าหมายของการอนุมานทางสถิติไปที่ค่าเฉลี่ยประชากรเป็นหลัก และทำการประมาณค่าพารามิเตอร์ด้วยวิธี Generalized Estimating Equation (GEE) โดยการสร้างตัวแบบความสัมพันธ์ระหว่างค่าเฉลี่ยประชากรกับตัวแปรอิสระผ่านฟังก์ชันเชื่อมโยง ส่วนความสัมพันธ์ของค่าสังเกตภายในหน่วยศึกษา จะมีการสร้างเป็นตัวแบบต่างหาก แล้วนำผลที่ได้มาใช้ในการปรับค่าประมาณพารามิเตอร์และค่าความแปรปรวนของพารามิเตอร์การถดถอยในตัวแบบความสัมพันธ์ เพื่อหาค่าประมาณของพารามิเตอร์ที่เหมาะสมของตัวแบบ

2. Mixed Effect Models หรือ Subject-Specific Models ตัวแบบนี้จะมุ่งเน้นการอนุมานทางสถิติไปที่หน่วยศึกษาเป็นหลัก โดยการเพิ่มเทอมอิทธิพลสุ่ม (Random Effects) เข้ามาในตัวแบบ คือยอมให้เซตย่อยของสัมประสิทธิ์การถดถอยสามารถมีค่าเปลี่ยนแปลงได้อย่างสุ่มจากคนหนึ่งไปยังอีกคนหนึ่ง โดยมีการนำความสัมพันธ์ของค่าสังเกตภายในหน่วยศึกษาเข้ามาพิจารณาด้วยผ่านทางเทอมอิทธิพลสุ่ม

ตารางที่ 3 โครงสร้างของข้อมูลการศึกษาระยะยาว

หน่วยศึกษา (Subject)	ตัวแปรตาม (Y)	ตัวแปรอิสระ (X)			
		X_1	X_2	$\dots$	X_k
1	Y_{11}	X_{111}	X_{112}	$\dots$	X_{11k}
	Y_{12}	X_{121}	X_{122}	$\dots$	X_{12k}
	$\vdots$	$\vdots$			
	Y_{1t}	X_{1t1}	X_{1t2}	$\dots$	X_{1tk}
2	Y_{21}	X_{211}	X_{212}	$\dots$	X_{21k}
	Y_{22}	X_{221}	X_{222}	$\dots$	X_{22k}
	$\vdots$	$\vdots$			
	Y_{2t}	X_{2t1}	X_{2t2}	$\dots$	X_{2tk}
$\vdots$	$\vdots$				
S	Y_{s1}	X_{s11}	X_{s12}	$\dots$	X_{s1k}
	Y_{s2}	X_{s21}	X_{s22}	$\dots$	X_{s2k}
	$\vdots$	$\vdots$			
	Y_{st}	X_{st1}	X_{st2}	$\dots$	X_{stk}

การสร้างตัวแบบทางสถิติสำหรับข้อมูลการศึกษาติดตามระยะยาวจะมุ่งไปที่วัตถุประสงค์ของการศึกษาว่าสนใจที่อธิบายหรือแปลผลในภาพรวมของค่าเฉลี่ยประชากร หรือสนใจที่จะอธิบายผลในลักษณะเฉพาะของแต่ละหน่วยศึกษา อย่างไรก็ตามตัวแบบทั้งสองได้พัฒนามาจากตัวแบบเชิงเส้นวางนัยทั่วไป (GLMs) ดังนี้

1. วิธีตัวแบบเชิงเส้นวางนัยทั่วไป (Generalized Linear Models ,GLMs)

วิธีตัวแบบเชิงเส้นทั่วไป (General Linear Models) มีข้อกำหนดให้ตัวแปรตามจะต้องมีการแจกแจงปกติ แต่ในบางสถานการณ์ ตัวแปรตามอาจมีการแจกแจงในรูปแบบอื่น และบางครั้งอาจจะมีลักษณะข้อมูลเป็นข้อมูลเชิงกลุ่ม (Categorical Data) หรือข้อมูลจำนวนนับ (Count) ดังนั้นการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระจะไม่สามารถเขียนให้อยู่ในความสัมพันธ์เชิงเส้นทั่วไปได้ ในปี ค.ศ.1972 Nelder and Wedderburn ได้เสนอตัวแบบเชิงเส้นวางนัยทั่วไป (GLMs) เพื่อให้สามารถใช้ได้กับข้อมูลที่มีลักษณะไม่ต่อเนื่อง (Discrete Data) โดยตัวแบบนี้ได้พัฒนาต่อจากตัวแบบเชิงเส้นทั่วไป ที่มีข้อตกลงว่าข้อมูลต้องมีการแจกแจงปกติ (Normal Distribution) และเป็นข้อมูลต่อเนื่อง ให้เป็นตัวแทนสำหรับข้อมูลที่มีการแจกแจงจัดเป็นสมาชิกในกลุ่มเอกซ์โปเนนเชียล (Exponential Family) ซึ่งจะครอบคลุมถึง การแจกแจงทวินาม (Binomial Distribution) การแจกแจงปัวส์ซอง (Poisson Distribution) การแจกแจงปกติ การแจกแจงเรขาคณิต (Geometric Distribution) การแจกแจงทวินามลบ (Negative Binomial Distribution) การแจกแจงแกมมา (Gamma Distribution) และการแจกแจงเอกซ์โปเนนเชียล (Exponential Distribution) เป็นต้น (Myers *et al.*, 2002)

ตัวแบบ GLMs ประกอบด้วยองค์ประกอบ 3 ส่วน (Fitzmaurice *et al.*, 2004) คือ

1.1 การแจกแจงของตัวแปรสุ่ม (Y) ที่เป็นตัวแปรตาม หรือตัวแปรผลลัพธ์ที่อาจเป็นข้อมูลต่อเนื่องหรือไม่ต่อเนื่อง สามารถเขียนอยู่ในรูปแบบสมาชิกของกลุ่มเอกซ์โปเนนเชียลได้ดังนี้

$$f_Y(y ; \theta, \phi) = \exp \left\{ \frac{[y\theta - b(\theta)]}{a(\phi)} + c(y, \phi) \right\} \quad (1)$$

เมื่อ $a(\cdot)$, $b(\cdot)$ และ $c(\cdot)$ เป็นฟังก์ชันใด ๆ ที่กำหนด

โดย $b(\theta)$ เป็นฟังก์ชันในเทอมของ θ เท่านั้น

และ $c(y, \phi)$ เป็นฟังก์ชันในเทอมของ y และ ϕ

พารามิเตอร์ θ เรียกว่าพารามิเตอร์ตำแหน่ง (Location Parameter) ส่วน ϕ เรียกว่าพารามิเตอร์การกระจาย (Dispersion Parameter) สำหรับฟังก์ชัน $a(\phi)$ มักจะมีรูปแบบเป็น

$a(\phi) = \phi/w$ และฟังก์ชัน $c = c(y, \phi/w)$ โดยที่ w แทนค่าน้ำหนักที่ทราบค่าของแต่ละค่าสังเกต โดยทั่วไปจะกำหนดให้ $w = 1$ สำหรับทุก ๆ ค่าสังเกต

ค่าเฉลี่ยและค่าความแปรปรวนของตัวแปรตาม (Y) ที่เขียนในรูปแบบการแจกแจงของกลุ่มเอกซ์โปเนนเชียลสามารถหาได้จาก

$$\begin{aligned} E(Y) &= \frac{db(\theta)}{d\theta} = b'(\theta) = \mu & \text{และ} & \quad \text{Var}(Y) = \frac{d^2b(\theta)}{d\theta^2} a(\phi) \\ \therefore \theta &= b'^{-1}(\mu) & & \quad = b''(\theta) a(\phi) \\ & & & \quad = \frac{d\mu}{d\theta} a(\phi) \end{aligned}$$

โดยที่ $b'(\theta)$ คือ การหาอนุพันธ์อันดับที่ 1 ของฟังก์ชัน $b(\theta)$ เมื่อเทียบกับพารามิเตอร์ θ
 $b''(\theta)$ คือ การหาอนุพันธ์อันดับที่ 2 ของฟังก์ชัน $b(\theta)$ เมื่อเทียบกับพารามิเตอร์ θ

ความแปรปรวนของตัวแปรตาม (y) ที่เขียนในรูปแบบการแจกแจงของกลุ่มเอกซ์โปเนนเชียลสามารถเขียนเป็นผลคูณระหว่างฟังก์ชันสองฟังก์ชัน คือ $b''(\theta)$ ซึ่งเป็นฟังก์ชันที่ขึ้นกับ θ หรือค่าเฉลี่ยกับ $a(\phi)$ ซึ่งเป็นฟังก์ชันที่เป็นอิสระจาก θ ดังนั้นฟังก์ชันความแปรปรวนของ Y จะถูกพิจารณาเหมือนเป็นฟังก์ชันของค่าเฉลี่ย (μ) (Myers *et al.*, 2002) คือ

$$\begin{aligned} \text{จาก} \quad \text{Var}(Y) &= b''(\theta) a(\phi) \\ &= \frac{d\mu}{d\theta} a(\phi) \text{ หรือ } b''[b'^{-1}(\mu)] a(\phi) \\ \text{โดยที่} \quad b''(\theta) &= \frac{d\mu}{d\theta} = \text{Var}(\mu) \end{aligned}$$

ซึ่ง $\text{Var}(\mu)$ จะถูกเรียกเป็น Variance Function ที่อธิบายความสัมพันธ์ระหว่างค่าเฉลี่ยและค่าความแปรปรวน

$$\text{ดังนั้น} \quad \text{Var}(Y) = a(\phi) \text{Var}(\mu) \text{ หรือ } \text{Var}(Y) = \phi \text{Var}(\mu) \text{ ถ้า } a(\phi) = \phi$$

ในการศึกษานี้ตัวแปรตามหรือ Y เป็นจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยต่อปี โดยที่ค่าสังเกต y จะมีค่าเป็นศูนย์หรือจำนวนเต็มบวกเท่านั้น ดังนั้นฟังก์ชันการแจกแจงความน่าจะเป็นของค่าสังเกต y จะมีการแจกแจงแบบปัวซอง ดังสมการข้างล่างนี้

$$f(y; \mu) = \frac{e^{-\mu} \mu^y}{y!} \quad \text{เมื่อ } y = 0, 1, 2, \dots$$

สามารถเขียนฟังก์ชันของ y ให้อยู่ในรูปแบบของเอกซ์โปเนนเชียลได้ดังนี้

$$f(y; \mu) = \exp [y \log (\mu) - \mu - \log (y!)] \quad (2)$$

โดยที่ $\theta = \log (\mu)$

$$b(\theta) = \mu = \exp (\theta)$$

$$a(\phi) = 1 \text{ หรือ } \phi = 1$$

$$c(y, \phi) = -\log (y!)$$

จากการจัดรูปทำให้ Dispersion Parameter ของการแจกแจงปัวซองมีค่าเท่ากับ 1 คือ $\phi = 1$ ส่วนค่าเฉลี่ยและความแปรปรวนของตัวแปรตาม Y คำนวณได้จาก

$$E(Y) = b'(\theta) = \exp (\theta) = \mu \quad \text{และ} \quad \text{Var}(Y) = b''(\theta) = \exp (\theta) = \mu$$

1.2 องค์ประกอบระบบ (Systematic Component) หรือโครงสร้างหลักของ GLMs สามารถเขียนได้ในรูปแบบของตัวทำนายเชิงเส้น (Linear Predictor) ของตัวแปรอิสระ $X_1, \dots, X_k$ ที่มีความสัมพันธ์กับเวกเตอร์พารามิเตอร์ $\underline{\beta}$ ดังสมการ

$$\eta = X\underline{\beta}; \quad \eta_i = \underline{x}_i' \underline{\beta} = \beta_0 + \sum_{i=1}^k \beta_i x_i; \quad i = 1, 2, \dots, k \quad (3)$$

โดยที่ η แทนตัวทำนายเชิงเส้นเป็นเวกเตอร์ขนาด $n \times 1$

X แทนเมตริกซ์ของตัวแปรอิสระที่มีขนาด $n \times p$ ประกอบด้วย n ค่าสังเกต

$\underline{x}_i'$ แทนเวกเตอร์ของค่าสังเกตของตัวแปรอิสระ $X_1, \dots, X_k$ ในแถวข้อมูลที่ i

โดย $\underline{x}_i' = (1 \ x_1 \ x_2 \ \dots \ x_k)$

$\underline{\beta}$ แทนเวกเตอร์ของพารามิเตอร์ $(\beta_1, \dots, \beta_p)'$ ที่มีขนาด $p \times 1$; $p = k+1$

1.3 ฟังก์ชันเชื่อมโยง (Link Function) แสดงความสัมพันธ์ระหว่างค่าเฉลี่ยของตัวแปรตาม (μ) กับตัวทำนายเชิงเส้น (η) โดยเขียนแทนฟังก์ชันเชื่อมโยง ด้วยสัญลักษณ์ $g(\cdot)$ ดังสมการ

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (4)$$

สำหรับการแจกแจงแบบปัวซอง จะเชื่อมความสัมพันธ์ระหว่าง μ กับ η ด้วยฟังก์ชันของลอการิธึม เรียกว่าฟังก์ชันเชื่อมโยงของลอการิธึม (Log Link Function) คือ $g(\mu) = \log(\mu)$ สามารถเขียนแทน $g(\mu)$ ด้วย η ดังนี้

$$\log(\mu_i) = \eta_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (5)$$

หรือเรียกเป็นรูปแบบลอการิธึม (Log-linear Model) ซึ่งเขียนได้ดังสมการ

$$\log(\mu_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (6)$$

การประมาณค่าพารามิเตอร์ $\underline{\beta}$ ในตัวแบบ GLMs สามารถหาได้จากวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood; ML) กรณีตัวแปรตามมีการแจกแจงแบบปัวซอง สามารถจัดให้อยู่ในรูปแบบของกลุ่มเอกซ์โปเนนเชียลจะได้ค่าความแปรปรวนของตัวแปรตาม Y เป็น $\text{Var}(Y) = \phi \text{Var}(\mu) = \mu$ โดยที่ $\text{Var}(\mu) = \mu$ ส่วน Dispersion parameter ϕ เป็นค่าที่ทราบหรืออาจประมาณค่าขึ้นจากข้อมูลในการแจกแจงแบบปัวซองจะได้ $\phi = 1$

การประมาณพารามิเตอร์ $\underline{\beta}$ ด้วยการประมาณค่าวิธีภาวะน่าจะเป็นสูงสุด (ML) เริ่มด้วยการเขียนฟังก์ชันภาวะน่าจะเป็น (Likelihood Function) ดังนี้

$$l = \log L = \sum_{i=1}^n \left[\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right] \quad (7)$$

และทำการหาอนุพันธ์อันดับที่หนึ่งเทียบกับ $\underline{\beta}$ แล้วกำหนดให้เท่ากับศูนย์ ได้สมการประมาณค่า (Estimation Equation) เขียนในรูปเวกเตอร์ได้ ดังนี้

$$\frac{\partial l}{\partial \underline{\beta}} = \sum_{i=1}^n \left(\frac{\partial \theta_i}{\partial \underline{\beta}} \right) \left(\frac{y_i - \mu_i}{\phi} \right) = \underline{0} \quad (8)$$

เนื่องจากเป็นสมการในรูปแบบไม่เป็นเชิงเส้น (Non-Linear) ดังนั้นจึงไม่สามารถหาค่าประมาณพารามิเตอร์ $\underline{\beta}$ ของสมการนี้ได้โดยตรง จึงต้องใช้วิธีการคำนวณเชิงตัวเลขแบบวนซ้ำ (Numerical Iteration) ซึ่งมีหลายวิธี เช่น วิธีนิวตัน-ราฟสัน (Newton-Raphson Algorithm) และวิธีฟิชเชอร์ –

สกอริง (Fisher-Scoring Algorithm) เป็นต้น สามารถใช้โปรแกรมทางสถิติมาช่วยวิเคราะห์ เช่น คำสั่ง PROC GENMOD ในโปรแกรมสำเร็จรูป SAS สำหรับวิเคราะห์ด้วยแบบวิธี GLMs (Fitzmaurice *et al.*, 2004)

2. วิธี Generalized Estimating Equations (GEE)

การสร้างตัวแบบทางสถิติสำหรับข้อมูลที่มีการติดตามระยะยาว โดยค่าสังเกตของตัวแปรตามมีการเก็บข้อมูลซ้ำ ๆ ภายใต้นักศึกษาเดียวกันมักมีความสัมพันธ์กัน (Correlated Data) ในปี ค.ศ.1986 Liang and Zeger ได้เสนอวิธี Generalized Estimating Equations (GEE) ที่ขยายมาจากวิธี GLMs สำหรับใช้วิเคราะห์ข้อมูลที่มีการติดตามระยะยาว และตัวแบบที่ได้มักจะถูกเรียกว่า Marginal Model ซึ่งมีการอธิบายผลของค่าประมาณพารามิเตอร์ในภาพรวมค่าเฉลี่ยประชากร (Population Average; PA) พร้อมทั้งพิจารณาโครงสร้างความสัมพันธ์ของค่าสังเกตที่มีการวัดซ้ำบนหน่วยศึกษาหนึ่ง ๆ เรียกเป็นความสัมพันธ์ภายในหน่วยศึกษา (Within-Subject) โดยมีการสมมติรูปแบบของโครงสร้างความสัมพันธ์ในลักษณะของเมทริกซ์ความสัมพันธ์ เรียกว่า Working Correlation Matrix สำหรับวิธี GEE จะสมมติให้ y_{ij} คือ ค่าสังเกตของหน่วยศึกษาที่ $i = 1, 2, \dots, s$ จากการวัดครั้งที่ $j = 1, 2, \dots, t$ และมีความสัมพันธ์กับเวกเตอร์ค่าสังเกตของตัวแปรอิสระ $X_1, \dots, X_k$ ที่ได้จากการวัดครั้งที่ j คือ $\underline{x}'_{ij} = (1 \ x_{1j}, \dots, x_{kj})$ เมื่อ $k = 1, 2, \dots, k$ และข้อตกลงเบื้องต้นของวิธี GEE ขยายมาจากวิธี GLMs แต่ในวิธี GEE จะไม่คำนึงถึงข้อสมมติการแจกแจงของตัวแปรตาม ซึ่งไม่จำเป็นต้องจัดการแจกแจงของตัวแปรตามให้อยู่ในกลุ่มเอกซ์โปเนนเชียล แต่ค่าเฉลี่ยฟังก์ชันความแปรปรวน และฟังก์ชันเชื่อมโยงของตัวทำนายเชิงเส้นจะต้องมีคุณสมบัติเป็นไปตาม

ข้อตกลงของ GLMs ดังนั้นวิธี GEE สำหรับการวิเคราะห์ Marginal Model มีคุณลักษณะที่สำคัญ 3 ข้อ (Fitzmaurice *et al.*, 2004) คือ

2.1 ค่าเฉลี่ยโดยรวม (Marginal Expectation) ของตัวแปรตาม Y_{ij} คือ $E(Y_{ij}) = \mu_{ij}$ จะขึ้นอยู่กับตัวแปรอิสระหรือมีความสัมพันธ์ผ่านฟังก์ชันเชื่อมโยงในรูปแบบเชิงเส้น ดังนี้

$$g(\mu_{ij}) = \eta_{ij} = \underline{x}_{ij}'\underline{\beta} \quad (9)$$

2.2 ค่าความแปรปรวนโดยรวม (Marginal Variance) ของตัวแปรตาม Y_{ij} จะเป็นฟังก์ชันที่ขึ้นกับค่าเฉลี่ย คือ

$$\text{Var}(Y_{ij}) = \phi \text{Var}(\mu_{ij})$$

เมื่อ $\text{Var}(\mu_{ij})$ เป็นฟังก์ชันที่ทราบค่าของค่าเฉลี่ย (μ_{ij}) และ Dispersion Parameter ϕ ในตัวแบบ GLMs อาจเป็นค่าที่ทราบหรืออาจจะต้องประมาณค่าจากข้อมูลก็ได้

2.3 ความสัมพันธ์ของค่าสังเกตภายใต้หน่วยศึกษา หรือค่าของตัวแปรตามที่มีการวัดซ้ำของหน่วยศึกษาเดียวกันหรือบุคคลเดียวกันจะมีความสัมพันธ์กัน คือ $\text{Corr}(y_{ij}, y_{ik}) = \rho_{jk}$ เมื่อ ij แทนหน่วยศึกษาที่ i จากการวัดครั้งที่ j และ ik แทนหน่วยศึกษาที่ i จากการวัดครั้งที่ k ซึ่ง $j \neq k$ ในวิธี GEE จะพิจารณาถึงโครงสร้างความสัมพันธ์ในรูปแบบของ Working Correlation Matrix ต่าง ๆ เช่น First-order Autoregressive : AR(1) และ Compound Symmetry : CS เป็นต้น

การศึกษาครั้งนี้เป็นการสร้างตัวแบบทางสถิติสำหรับข้อมูลจำนวนการเรียกค่าสินไหมทดแทนที่มีการเก็บข้อมูลซ้ำในระยะเวลาติดตาม 5 ปี โดยที่ ตัวแปรตาม คือ จำนวนการเรียกค่าสินไหมทดแทน ซึ่งเป็นตัวแปรจำนวนนับ (Count) มีการแจกแจงแบบปัวซอง สามารถวิเคราะห์ด้วยวิธี GEE ได้ดังนี้

ตัวแปรตามจะมีความสัมพันธ์กับตัวทำนายเชิงเส้นผ่านทางฟังก์ชันเชื่อมโยงของ Log Link ดังนี้

$$g(\mu_{ij}) = \log(\mu_{ij}) = \underline{x}_{ij}'\underline{\beta} \quad (10)$$

ความแปรปรวนของตัวแปรตาม เมื่อถูกกำหนดด้วยค่าของตัวแปรอิสระ จะขึ้นกับค่าเฉลี่ยของตัวแปรตามนั้น

$$\text{จาก } \text{Var}(Y_{ij}) = \phi \text{ Var}(\mu_{ij}) \text{ เมื่อ } \text{Var}(\mu_{ij}) = \mu_{ij} \text{ และ } \phi = 1 \\ \text{จะได้ } \text{Var}(Y_{ij}) = \mu_{ij}$$

ความสัมพันธ์ระหว่างค่าสังเกตของหน่วยศึกษาเดียวกัน ที่ปรากฏใน Working Correlation Matrix จะสมมติให้อยู่ในรูปแบบของ AR(1) และ CS ดังนี้

1. AR(1) มีโครงสร้างดังนี้

$$\text{Corr}(y_{ij}, y_{i,j+t}) = \rho^t \text{ for } t = 1, 2, \dots, n_i - j \text{ เขียนในรูปเมทริกซ์ คือ } \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{pmatrix}$$

ความสัมพันธ์ของค่าสังเกตภายในหน่วยศึกษาเดียวกันจะมีค่าลดลงตามระยะห่างของช่วงเวลาที่เก็บข้อมูลซ้ำ ซึ่งค่าความสัมพันธ์จะอยู่ในรูปของ ρ^t เมื่อ t คือ จำนวนช่วงเวลาก่อนหน้า (lag time) เมื่อ $t = 1, 2, \dots, n_i - j$

2. CS หรือเรียกอีกอย่างว่า Exchangeable (Shoukri and Chaudhary, 2007) มีโครงสร้างดังนี้

$$\text{Corr}(y_{ij}, y_{ik}) = \begin{cases} 1 & ; j = k \\ \rho & ; j \neq k \end{cases} \text{ เขียนในรูปเมทริกซ์ คือ } \begin{pmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \rho & \rho & 1 & \dots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{pmatrix}$$

ความสัมพันธ์ของค่าสังเกตภายในหน่วยศึกษาเดียวกันจะมีค่าเท่ากัน เท่ากับ ρ ณ ช่วงเวลาที่ $j \neq k$

การประมาณค่าพารามิเตอร์ $\underline{\beta}$ ในการวิเคราะห์หัตถิ GEE จะขยายมาจากสมการเมื่อความสัมพันธ์ของข้อมูลเป็นอิสระกันเขียนได้ดังนี้

$$\sum_{i=1}^s \frac{\partial \mu'_i}{\partial \underline{\beta}} [V_i(\hat{\rho})]^{-1} [y_i - \mu_i] = 0 \quad ; \quad \mu'_i = [\mu_{i1}, \mu_{i2}, \dots, \mu_{it}] \quad (11)$$

เมื่อ y_i เป็นค่าสังเกตที่ i ของตัวแปรตาม Y

เนื่องจากเป็นสมการในรูปแบบไม่เป็นเชิงเส้น ในการประมาณค่าของ $\underline{\beta}$ จึงไม่สามารถแก้สมการนี้ได้โดยตรง จึงต้องใช้วิธีการคำนวณซ้ำเชิงตัวเลข (Numerical Iteration) ระหว่างวิธี Quasi-Likelihood สำหรับประมาณค่าพารามิเตอร์และวิธี Robust สำหรับประมาณค่า ρ โดยใช้คำสั่ง PROC GENMOD ในโปรแกรมสำเร็จรูป SAS

สำหรับการวิเคราะห์หัตถิ GEE และสามารถสรุปเป็นขั้นตอนได้ดังนี้

1. ประมาณค่าเริ่มต้นของพารามิเตอร์ $\underline{\beta}$ จากวิธี GLMs เมื่อสมมติให้ความสัมพันธ์ของค่าสังเกตภายในหน่วยศึกษาเดียวกันเป็นอิสระต่อกัน
2. คำนวณ Working Correlation Matrix (R) โดยประมาณค่าโครงสร้างความสัมพันธ์ภายในแต่ละหน่วยศึกษา คือ $R_i(\rho)$ จากค่าที่คำนวณได้ของ $\underline{\beta}$ ณ ปัจจุบัน แล้วคำนวณค่า Pearson Residuals (e_{ij}) ได้ดังนี้ $e_{ij} = \frac{y_{ij} - \mu_{ij}}{\sqrt{v(\mu_{ij})}}$ จากนั้นนำค่า e_{ij} ที่ได้มาคำนวณค่าประมาณของ ϕ โดยที่ $\hat{\phi} = \frac{1}{N-p} \sum_{i=1}^s \sum_{j=1}^t e_{ij}^2$; N คือจำนวนค่าสังเกตทั้งหมดของข้อมูล ($N = s \times t$) และประมาณค่าของ ρ จากฟังก์ชันของ ϕ และ e_{ij} ตามรูปแบบความสัมพันธ์ของ $R_i(\rho)$ ที่สมมติ
3. คำนวณค่าของเมทริกซ์ความแปรปรวนร่วม (Working Covariance Matrix) ของหน่วยศึกษาที่ i คือ $V_i(\rho) = \phi A_i^{\frac{1}{2}} R_i(\rho) A_i^{\frac{1}{2}}$
4. ทำการปรับค่าพารามิเตอร์ $\underline{\beta}$ ไปเรื่อย ๆ โดยการเพิ่มจำนวนการวนซ้ำดังนี้

$$\beta_{r+1} = \beta_r - \left[\sum_{i=1}^s \frac{\partial \mu'_i}{\partial \underline{\beta}} V_i^{-1} \frac{\partial \mu_i}{\partial \underline{\beta}} \right]^{-1} \left[\sum_{i=1}^s \frac{\partial \mu'_i}{\partial \underline{\beta}} V_i^{-1} (y_i - \mu_i) \right] \quad (12)$$

5. คำนวณซ้ำในข้อ 2 ถึง 4 จนกระทั่งค่าประมาณของพารามิเตอร์ $\underline{\beta}$ คือ $\hat{\underline{\beta}}$ เข้าสู่ค่าคงที่

ค่าประมาณความแปรปรวนของ $\hat{\underline{\beta}}$ ที่ได้จากการสมมติโครงสร้างความสัมพันธ์ของค่าสังเกตภายในหน่วยศึกษาเดียวกันจะถูกเรียกเป็น Empirical Variance Estimator หรือ Sandwich Estimator หรือ Robust Estimator ซึ่งคำนวณได้จาก (SAS/STAT User's Guide Version 8, 1999)

$$\hat{\text{Var}}(\hat{\underline{\beta}}) = \left(\sum_{i=1}^s \frac{\partial \mu'_i}{\partial \underline{\beta}} V_i^{-1} \frac{\partial \mu_i}{\partial \underline{\beta}} \right)^{-1} \left(\sum_{i=1}^s \frac{\partial \mu'_i}{\partial \underline{\beta}} V_i^{-1} \text{Cov}(y_i) V_i^{-1} \frac{\partial \mu_i}{\partial \underline{\beta}} \right) \left(\sum_{i=1}^s \frac{\partial \mu'_i}{\partial \underline{\beta}} V_i^{-1} \frac{\partial \mu_i}{\partial \underline{\beta}} \right)^{-1}$$

3. ตัวแบบผสมเชิงเส้น (Linear Mixed Models , LMMs)

ตัวแบบผสมเชิงเส้น (LMMs) เป็นตัวแบบทางสถิติสำหรับตัวแปรตามที่มีข้อมูลเป็นแบบต่อเนื่องใช้สำหรับศึกษาความสัมพันธ์เชิงเส้นระหว่างค่าเฉลี่ยของตัวแปรตาม (Y) กับตัวแปรอิสระหรือปัจจัย (X) ตั้งแต่หนึ่งตัวขึ้นไป ที่ประกอบด้วยปัจจัยอิทธิพลคงที่ (Fixed Effects) และปัจจัยอิทธิพลสุ่ม (Random Effects) โดยปัจจัยอิทธิพลสุ่มจะอธิบายลักษณะเฉพาะตัวของแต่ละหน่วยศึกษา

Fixed Effects เป็นเทอมที่เรียกสัมประสิทธิ์การถดถอยหรือพารามิเตอร์อิทธิพลคงที่เพื่อใช้อธิบายความสัมพันธ์ร่วมกันจากทุกหน่วยศึกษาระหว่างค่าเฉลี่ยของตัวแปรตามกับตัวแปรอิสระที่เป็นปัจจัยคงที่ (Fixed Factors) ซึ่งอาจเป็นตัวแปรสุ่มที่นักวิจัยได้กำหนดระดับที่สนใจศึกษาหรือเป็นตัวแปรต่อเนื่องที่ได้กำหนดขอบเขตที่ต้องการศึกษาไว้ก็ได้ โดยที่ระดับหรือขอบเขตของค่าที่สนใจจะกำหนดมาจากประชากรทั้งหมดที่เกี่ยวข้องเป็นหน่วยในการวิเคราะห์ (Analysis Unit) เช่น สนใจที่จะอธิบายความแตกต่างระหว่างระดับของปัจจัยคงที่ระหว่างเพศชายกับเพศหญิงที่มีต่อค่าเฉลี่ยผลลัพธ์ของตัวแปรตามและเทอมอิทธิพลคงที่ที่สามารถประมาณได้จากข้อมูล (West *et al.*, 2007)

Random Effects เป็นเทอมอิทธิพลเฉพาะของแต่ละหน่วยศึกษามีลักษณะเป็นค่าสุ่มที่สัมพันธ์กับระดับของปัจจัยสุ่ม (Random Factors) ที่เป็นตัวแทนซึ่งสุ่มมาจากระดับที่เป็นไปได้ทั้งหมดของประชากรที่ศึกษา ตัวอย่างเช่น Random Effects ที่สัมพันธ์กับระดับของปัจจัยสุ่มในตัวแบบ LMMs ที่เรียกเป็น Random Intercept จะแสดงความผันแปรสุ่มหรือความแตกต่างของค่าคงที่ในแต่ละหน่วยศึกษาจากค่าคงที่ของประชากรทั้งหมด (Overall Fixed Intercept) ส่วนใหญ่ Random Intercept มักจะหมายถึงอิทธิพลของหน่วยศึกษา (Subject Effect) นอกจากนี้ Random Effects ที่เรียกว่า Random Coefficients จะแสดงความผันแปรสุ่มของพารามิเตอร์สำหรับแต่ละหน่วยศึกษาที่แตกต่างกันจากพารามิเตอร์อิทธิพลคงที่โดยรวมของประชากร (Overall Fixed Effects) (West *et al.*, 2007)

ดังนั้นตัวแบบผสมเชิงเส้น (Linear Mixed Models, LMMs) มาจากการเพิ่มเทอมอิทธิพลสุ่ม (Random Effects) เข้ามาในตัวแบบเชิงเส้น (Linear Models) ทั่วไปเขียนได้ ดังนี้ (West *et al.*, 2007)

$$\mathbf{Y}_i = \underbrace{\mathbf{X}_i \boldsymbol{\beta}}_{\text{fixed}} + \underbrace{\mathbf{Z}_i \mathbf{u}_i}_{\text{random}} + \boldsymbol{\varepsilon}_i \quad (13)$$

เมื่อ	$\mathbf{Y}_i$	คือ เวกเตอร์ของตัวแปรตามที่เป็นค่าต่อเนื่องสำหรับหน่วยศึกษาที่ i ; $i = 1, 2, \dots, n$
	$\mathbf{X}_i$	คือ เมทริกซ์ของตัวแปรอิสระหรืออิทธิพลคงที่ที่มี k ตัวแปร โดยมีขนาดเมทริกซ์ $n \times p$; $p = k + 1$
	$\boldsymbol{\beta}$	คือ เวกเตอร์ของพารามิเตอร์อิทธิพลคงที่ที่มีขนาด $p \times 1$
	$\mathbf{Z}_i$	คือ เมทริกซ์ของอิทธิพลสุ่มที่มี q ตัวแปร ที่มีขนาด $n \times q$ โดย $q \leq p$
	$\mathbf{u}_i$	คือ เวกเตอร์ของพารามิเตอร์อิทธิพลสุ่มของหน่วยศึกษาที่ i ที่มีขนาด $q \times 1$
		เมื่อ $\mathbf{u}_i \sim N(\mathbf{0}, \mathbf{G})$
	$\boldsymbol{\varepsilon}_i$	คือ เวกเตอร์ของความคลาดเคลื่อนของหน่วยศึกษาที่ i เมื่อ $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i)$

นั่นคือ $\boldsymbol{\varepsilon}_i$ มีการแจกแจงปกติ แต่อาจจะไม่เป็นอิสระกัน หรือมีความแปรปรวนเท่ากัน จากสมการ (13) $\mathbf{Y}_i$ คือ เวกเตอร์ของตัวแปรตามสำหรับหน่วยศึกษาที่ i สามารถเขียนในรูปเวกเตอร์สำหรับค่าสังเกตแต่ละหน่วยศึกษา ดังนี้

$$\mathbf{Y}_i = [\mathbf{y}_{it}] = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{it} \end{pmatrix} ; t = 1, 2, \dots, t$$

โดยที่ y_{it} คือ ค่าสังเกตของตัวแปรตามสำหรับหน่วยศึกษาที่ i วัดครั้งที่ t ส่วน $\mathbf{X}_i$ ในสมการ (13) เป็นเมทริกซ์ของตัวแปรอิสระหรืออิทธิพลคงที่สามารถเขียนในรูปเมทริกซ์ได้ดังนี้

$$\mathbf{X}_i = \begin{pmatrix} x_{i1}^{(1)} & x_{i1}^{(2)} & \dots & x_{i1}^{(p)} \\ x_{i2}^{(1)} & x_{i2}^{(2)} & \dots & x_{i2}^{(p)} \\ \vdots & \vdots & \dots & \vdots \\ x_{it}^{(1)} & x_{it}^{(2)} & \dots & x_{it}^{(p)} \end{pmatrix}$$

โดยที่ $X^{(1)}, X^{(2)}, \dots, X^{(p)}$ เป็นตัวแปรอิสระหรืออิทธิพลคงที่มี p ตัว และ $x_{it}^{(1)}$ เป็นค่าวัดจากตัวแปร $X^{(1)}$ จากการวัดครั้งที่ t สำหรับหน่วยศึกษาที่ i

ส่วน $\mathbf{u}_i$ เป็นตัวแปรสุ่มของเทอมอิทธิพลสุ่ม ที่มีการแจกแจงแบบพหุปกติ (Multivariate Normal Distribution) มีเวกเตอร์ของค่าเฉลี่ยเท่ากับศูนย์และเมทริกซ์ความแปรปรวนร่วม (Variance-Covariance matrix) ในรูปของ G แสดงได้ดังนี้

$$G = \text{Var}(\mathbf{u}_i) = \begin{pmatrix} \text{Var}(u_{1i}) & \text{Cov}(u_{1i}, u_{2i}) & \dots & \text{Cov}(u_{1i}, u_{qi}) \\ \text{Cov}(u_{1i}, u_{2i}) & \text{Var}(u_{2i}) & \dots & \text{Cov}(u_{2i}, u_{qi}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(u_{1i}, u_{qi}) & \text{Cov}(u_{2i}, u_{qi}) & \dots & \text{Var}(u_{qi}) \end{pmatrix}$$

โดยที่สมาชิกในแนวทแยงมุม คือ ค่าความแปรปรวน (Variance) ของอิทธิพลสุ่มใน $\mathbf{u}_i$ ส่วนสมาชิกนอกเส้นทแยงมุม คือ ค่าความแปรปรวนร่วมระหว่างเทอมอิทธิพลสุ่ม โครงสร้างการแปรปรวนร่วมของเมทริกซ์ G มีรูปแบบต่าง ๆ ดังนี้

1. Variance Component (VC)

$$G = \text{Var}(u_i) = \begin{pmatrix} \sigma_{u_1}^2 & 0 \\ 0 & \sigma_{u_2}^2 \end{pmatrix} \text{ สำหรับแต่ละ Random Effect ใน } u_i$$

2. Unstructure (UN)

$$G = \text{Var}(u_i) = \begin{pmatrix} \sigma_{u_1}^2 & \sigma_{u_1, u_2} \\ \sigma_{u_1, u_2} & \sigma_{u_2}^2 \end{pmatrix}$$

ส่วน ε_i คือเวกเตอร์ของความคลาดเคลื่อนของหน่วยศึกษาที่ i มีจำนวนความคลาดเคลื่อนเท่ากับ n ซึ่งแต่ละสมาชิกจะสัมพันธ์กับค่าที่สังเกตได้ ณ เวลา t ของหน่วยศึกษาที่ i และ ε_i จะมีการแจกแจงแบบพหุปกติ มีเวกเตอร์ของค่าเฉลี่ยเท่ากับศูนย์ และมีเมทริกซ์ความแปรปรวนร่วมในรูปของ R_i แสดงได้ดังนี้

$$R_i = \text{Var}(\varepsilon_i) = \begin{pmatrix} \text{Var}(\varepsilon_{1i}) & \text{Cov}(\varepsilon_{1i}, \varepsilon_{2i}) & \dots & \text{Cov}(\varepsilon_{1i}, \varepsilon_{ti}) \\ \text{Cov}(\varepsilon_{1i}, \varepsilon_{2i}) & \text{Var}(\varepsilon_{2i}) & \dots & \text{Cov}(\varepsilon_{2i}, \varepsilon_{ti}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(\varepsilon_{1i}, \varepsilon_{ti}) & \text{Cov}(\varepsilon_{2i}, \varepsilon_{ti}) & \dots & \text{Var}(\varepsilon_{ti}) \end{pmatrix}$$

โดยมีข้อกำหนดว่าค่าความคลาดเคลื่อนที่สัมพันธ์กับหน่วยศึกษาที่ต่างกันจะเป็นอิสระต่อกัน และค่าความคลาดเคลื่อน $\varepsilon_1, \dots, \varepsilon_n$ กับอิทธิพลสุ่ม $u_1, \dots, u_n$ เป็นอิสระต่อกันด้วย

ในการกำหนดโครงสร้างของเมทริกซ์ความแปรปรวนร่วมที่ง่ายที่สุด โดยทั่วไปจะกำหนดค่าความคลาดเคลื่อนที่สัมพันธ์กับค่าสังเกตบนหน่วยศึกษาเดียวกันจะไม่มีความสัมพันธ์ หรือกำหนดว่ามีค่าความแปรปรวนเท่ากันเขียนในรูปเมทริกซ์ได้ดังนี้

1. Diagonal Structure

$$R_i = \text{Var}(\varepsilon_i) = \sigma^2 \mathbf{I} = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix}$$

2. First-order Autoregressive (AR(1))

$$R_i = \text{Var}(\varepsilon_i) = \begin{pmatrix} \sigma^2 & \sigma^2 \rho & \dots & \sigma^2 \rho^{n_i-1} \\ \sigma^2 \rho & \sigma^2 & \dots & \sigma^2 \rho^{n_i-2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2 \rho^{n_i-1} & \sigma^2 \rho^{n_i-2} & \dots & \sigma^2 \end{pmatrix}$$

3. Compound Symmetry (CS)

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

ส่วนข้อสมมติการแจกแจงของตัวแปรตาม เมื่อ $\underline{Y}_i$ เป็นเวกเตอร์ตัวแปรสุ่มเมื่อมีเงื่อนไขของเทอมอิทธิพลสุ่มเข้ามาในตัวแบบจะทำให้แต่ละหน่วยศึกษามีความแตกต่างกัน โดย $\underline{Y}_i$ จะเป็นอิสระต่อกัน และจะได้ค่าเฉลี่ยของตัวแปรตาม $\underline{Y}_i$ ที่มีเงื่อนไขขึ้นกับอิทธิพลสุ่ม $\underline{u}_i$ ดังนี้

$$E(\underline{Y}_i | \underline{u}_i) = \mathbf{X}_i \underline{\beta} + \mathbf{Z}_i \underline{u}_i \quad (14)$$

และมีค่าเฉลี่ยของตัวแปรตามที่แสดงภาพรวมของประชากร โดยรวมที่เรียกว่า Marginal Mean คือ $E(\underline{Y}_i) = \mathbf{X}_i \underline{\beta}$ ส่วนค่าความแปรปรวนของตัวแปรตาม $\underline{Y}_i$ แบบมีเงื่อนไขที่ขึ้นกับ $\underline{u}_i$ คือ

$$\text{Var}(\underline{Y}_i | \underline{u}_i) = \text{Var}(\underline{\varepsilon}_i) = R_i$$

รวมทั้งมีค่าความแปรปรวนของตัวแปรตามในภาพรวมของประชากร เรียกว่า Marginal Variance คือ

$$\begin{aligned}\text{Var}(\underline{Y}_i) &= \text{Var}(\underline{Z}_i \underline{u}_i) + \text{Var}(\underline{\varepsilon}_i) \\ &= \underline{Z}_i \text{Var}(\underline{u}_i) \underline{Z}_i' + R_i \\ &= \underline{Z}_i G \underline{Z}_i' + R_i\end{aligned}$$

ถ้า $R_i = \text{Var}(\underline{\varepsilon}_i) = \sigma^2 \mathbf{I}$ ทำให้โครงสร้างความแปรปรวนของตัวแปรตามคือ

$$\text{Var}(\underline{Y}_i) = \underline{Z}_i G \underline{Z}_i' + \sigma^2 \mathbf{I}$$

สามารถเลือกความแปรปรวนของตัวแปรตามในรูปแบบเมทริกซ์เป็นโครงสร้างความแปรปรวนร่วมแบบใดก็ได้ เนื่องจากในความเป็นจริงการวิเคราะห์ความสัมพันธ์ภายในหน่วยศึกษาเดียวกันที่มีการวัดซ้ำ สามารถกำหนดโครงสร้างความแปรปรวนร่วมของค่าความคลาดเคลื่อนหรือ R_i ที่ไม่จำเป็นต้องเป็นอิสระกัน คือ สามารถกำหนดให้เป็นรูปแบบต่าง ๆ เช่น รูปแบบ CS และ AR(1) เป็นต้น

การประมาณค่าพารามิเตอร์อิทธิพลคงที่ (Fixed Effect Parameters) $\underline{\beta}$ และพารามิเตอร์ความแปรปรวนร่วมที่ปรากฏในเมทริกซ์ G และ R_i สามารถใช้วิธี ML หรือวิธี REML (Residual Maximum Likelihood) บางครั้งเรียกเป็น Restricted ML ซึ่งวิธี ML อาศัยการหาค่า Maximize Likelihood Function ของค่าสังเกต $\underline{y}$ แต่วิธี REML จะหาค่าสูงสุดจากการ Maximize Likelihood Function ของเทอมความคลาดเคลื่อน (Residual Terms) ได้ค่าประมาณ ดังนี้

$$\left. \begin{aligned}\hat{\underline{\beta}} &= (\underline{X}' \underline{V}^{-1} \underline{X})^{-1} \underline{X}' \underline{V}^{-1} \underline{y} \\ \text{Var}(\hat{\underline{\beta}}) &= (\underline{X}' \underline{V}^{-1} \underline{X})^{-1}\end{aligned} \right\} \text{สำหรับเทอมอิทธิพลคงที่}$$

กรณีไม่ทราบค่าเมทริกซ์ความแปรปรวนที่แท้จริง $\underline{V}^{-1}$ สามารถแทนค่าประมาณคือ $\hat{\underline{V}}^{-1}$ ในสูตร ทำให้ $\hat{\underline{\beta}}$ มีคุณสมบัติเป็น Empirical Best Linear Unbiased Estimator ส่วนการประมาณค่า Random Effect หรือ Random Intercept ใช้วิธี ML โดยกำหนดให้เมทริกซ์ของความแปรปรวนเป็น

Diagonal คือ $R_i = \sigma^2 \mathbf{I}$ โดยส่วนใหญ่โปรแกรมสำเร็จรูปจะกำหนดให้ใช้วิธี REML ในการประมาณค่าพารามิเตอร์ เนื่องจากค่าพารามิเตอร์ของความแปรปรวนร่วมจะมีความไม่เอนเอียง (Unbiased) ส่วนวิธี ML ค่าพารามิเตอร์ของความแปรปรวนร่วมจะมีความเอนเอียง (biased) รายละเอียดสามารถศึกษาได้จาก West *et al.* (2007) ซึ่งการวิเคราะห์ตัวแบบ LMMs สามารถใช้โปรแกรมทางสถิติต่าง ๆ ในการวิเคราะห์ได้ เช่น SAS สำหรับ Proc Mixed, SPSS MIXED และ R: lme () เป็นต้น (West *et al.*, 2007)

4. วิธีตัวแบบผสมเชิงเส้นวางนัยทั่วไป (Generalize Linear Mixed Models, GLMMs)

การสร้างตัวแบบทางสถิติสำหรับข้อมูลที่ตัวแปรตามเป็นชนิดไม่ต่อเนื่อง หรือตัวแปรตามอาจจะมีการแจกแจงแบบอื่น ๆ เช่น การแจกแจงแบบทวินาม และการแจกแจงแบบปัวสซอง เป็นต้น ซึ่งแตกต่างจากตัวแปรตามที่เป็นข้อมูลชนิดต่อเนื่อง ซึ่งมีการแจกแจงปกติ แต่เมื่อตัวแปรตามไม่มีการแจกแจงแบบปกติ เราจะเรียกตัวแบบนี้ว่า ตัวแบบผสมเชิงเส้นวางนัยทั่วไป (GLMMs) หรือเป็นวิธีที่ใช้ในการวิเคราะห์ข้อมูลที่มีการเก็บซ้ำในระยะยาวได้ ซึ่งประกอบด้วยเทอมอิทธิพลคงที่และอิทธิพลสุ่มเช่นเดียวกัน แต่จะมีการขยายลักษณะเฉพาะ 3 ส่วนเช่นเดียวกับวิธีตัวแบบเชิงเส้นวางนัยทั่วไป (GLMs) คือ (Fitzmaurice *et al.*, 2004)

1. การแจกแจงแบบมีเงื่อนไขของแต่ละ y_{ij} ที่ถูกกำหนดด้วยเทอมอิทธิพลสุ่ม u_i สามารถเขียนให้อยู่ในรูปแบบการแจกแจงเอกซ์โปเนนเชียลแฟมิลีและสามารถหาค่าความแปรปรวนของตัวแปรตาม Y_{ij} แบบมีเงื่อนไขที่ขึ้นกับ u_i ได้จาก

$$\text{Var}(Y_{ij}|u_i) = v\{E(Y_{ij}|u_i)\} \phi$$

เมื่อ $v(\cdot)$ เป็นฟังก์ชันความแปรปรวนที่ทราบค่า ส่วน $E(Y_{ij}|u_i)$ เป็นฟังก์ชันของค่าเฉลี่ยแบบมีเงื่อนไข นอกจากนี้เมื่อกำหนดเทอมอิทธิพลสุ่ม u_i แล้วค่าสังเกต y_{ij} จะเป็นอิสระต่อกัน ซึ่งเรียกเป็นข้อตกลงของ Conditional Independence Assumption

2. ค่าเฉลี่ยของตัวแปรตามแบบมีเงื่อนไขจะขึ้นกับอิทธิพลคงที่ และอิทธิพลสุ่ม u_i ผ่านตัวทำนายเชิงเส้น ดังนี้

$$E(Y_{ij}|\underline{u}_i) = \underline{X}_{ij}\underline{\beta} + \underline{Z}_{ij}\underline{u}_i$$

จะได้

$$\eta_{ij} = \underline{X}_{ij}\underline{\beta} + \underline{Z}_{ij}\underline{u}_i \quad \text{เมื่อ } \eta_{ij} = g\{E(Y_{ij}|\underline{u}_i)\}$$

$$g\{E(Y_{ij}|\underline{u}_i)\} = \underline{X}_{ij}\underline{\beta} + \underline{Z}_{ij}\underline{u}_i \quad (15)$$

3. โดยปกติจะกำหนดให้เทอมอิทธิพลสุ่ม $\underline{u}_i$ มีการแจกแจงปกติพหุตัวแปร (Multivariate Normal Distribution) คือ $\underline{u}_i \sim N(0, G)$ มี G เป็นเมทริกซ์ความแปรปรวนร่วม และ $\underline{u}_i$ เป็นอิสระกันกับ $\underline{X}_i$ โดยสามารถเขียนโครงสร้างความแปรปรวนร่วมของตัวแปรตามได้ดังนี้

$$\begin{aligned} \text{Var}(Y_i) &= \text{Var}(Z_i \underline{u}_i) + \text{Var}(\varepsilon_i) \\ &= \underline{Z}_i G \underline{Z}_i' + R_i \end{aligned}$$

เนื่องจากการวิเคราะห์ความสัมพันธ์ภายในหน่วยศึกษาเดียวกันที่มีการวัดซ้ำจะมีรูปแบบความสัมพันธ์ของโครงสร้างความแปรปรวนร่วมของค่าความคลาดเคลื่อนคือ R_i เช่นเดียวกับตัวแบบผสมเชิงเส้น (LMMs)

ในการศึกษานี้ตัวแปรตาม คือ จำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ มีการแจกแจงแบบปัวส์ซง สามารถศึกษาวิธี GLMMs ด้วยส่วนประกอบ 3 ส่วน ดังนี้

1. เมื่อตัวแปรตามมีการแจกแจงแบบปัวส์ซง สามารถหาค่า $\text{Var}(Y_{ij}|\underline{u}_i)$ และค่า $E(Y_{ij}|\underline{u}_i)$ คือ

$$\text{Var}(Y_{ij}|\underline{u}_i) = E(Y_{ij}|\underline{u}_i) \quad \text{โดยที่ค่า } \phi = 1$$

เนื่องจากค่าเฉลี่ยและค่าความแปรปรวนของการแจกแจงแบบปัวส์ซงจะมีค่าเท่ากัน

2. ค่าเฉลี่ยของตัวแปรตามแบบมีเงื่อนไขที่ขึ้นกับ $\underline{u}_i$ จะมีความสัมพันธ์เชิงเส้นกับเทอมอิทธิพลคงที่และเทอมอิทธิพลสุ่ม ผ่านฟังก์ชันเชื่อมโยงในรูปแบบของล็อก (Log Link) ของตัวทำนายเชิงเส้น คือ

$$\begin{aligned}
\text{จาก} \quad E(Y_{ij}|\underline{u}_i) &= \underline{X}_{ij}\underline{\beta} + \underline{Z}_{ij}\underline{u}_i \\
\eta_{ij} &= \underline{X}_{ij}\underline{\beta} + \underline{Z}_{ij}\underline{u}_i & \text{เมื่อ } \eta_{ij} &= \log \{E(Y_{ij}|\underline{u}_i)\} \\
\text{ดังนั้น} \quad \log \{E(Y_{ij}|\underline{u}_i)\} &= \underline{X}_{ij}\underline{\beta} + \underline{Z}_{ij}\underline{u}_i & (16)
\end{aligned}$$

3. เมื่อกำหนด $\underline{u}_i$ มีการแจกแจงปกติพหุตัวแปร (Multivariate Normal Distribution) คือ $\underline{u}_i \sim N(0, G)$ เมื่อ G เป็นเมตริกซ์ความแปรปรวนร่วม และเป็นอิสระกับ $\underline{X}_{ij}$ โดยเขียนโครงสร้างความแปรปรวนร่วมของตัวแปรตามได้ดังนี้

$$\begin{aligned}
\text{Var}(\underline{Y}_i) &= \text{Var}(\underline{Z}_i \underline{u}_i) + \text{Var}(\underline{\varepsilon}_i) \\
&= \underline{Z}_i G \underline{Z}_i' + R_i
\end{aligned}$$

ในการศึกษานี้เป็นข้อมูลการวัดซ้ำจะสมมติให้โครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนลักษณะ AR(1) เขียนในรูปเมทริกซ์ ดังนี้

$$R_i = \text{Var}(\underline{\varepsilon}_i) = \begin{pmatrix} \sigma^2 & \sigma^2 \rho & \dots & \sigma^2 \rho^{n_i-1} \\ \sigma^2 \rho & \sigma^2 & \dots & \sigma^2 \rho^{n_i-2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2 \rho^{n_i-1} & \sigma^2 \rho^{n_i-2} & \dots & \sigma^2 \end{pmatrix}$$

โครงสร้างความแปรปรวนร่วมแบบ AR(1) จะกำหนดให้แต่ละค่าสังเกตของหน่วยศึกษาที่ i มีความแปรปรวนเท่ากัน ส่วนความแปรปรวนร่วมระหว่างค่าสังเกตจะไม่เท่ากันสำหรับบุคคลเดียวกัน แต่จะลดลงเข้าใกล้ศูนย์เมื่อช่วงห่างเวลาของการวัดซ้ำข้อมูลมากขึ้น เนื่องจากเป็นผลคูณระหว่างความแปรปรวนกับค่าความสัมพันธ์ของความล่าช้า (lag) ของเวลาในรูปยกกำลัง

นอกจากนี้ยังได้สมมติโครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนเป็นลักษณะ CS เขียนในรูปเมทริกซ์ ดังนี้

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

ค่าสังเกตจะมีค่าความแปรปรวนร่วมในแต่ละคู่เท่ากันสำหรับหน่วยศึกษาเดียวกัน ซึ่งในการวิเคราะห์ด้วยวิธี GLMMs โครงสร้างความแปรปรวนร่วมของแต่ละหน่วยศึกษาจะมีรูปแบบ CS แต่เมื่อนำทุกหน่วยศึกษามาแสดงในภาพรวมโครงสร้างความแปรปรวนร่วมจะอยู่ในรูปแบบที่เรียกว่า Variance Component (VC) ซึ่งบางครั้งจะเรียกรูปแบบ CS ว่าเป็นรูปแบบ VC (Littell *et al.*, 2000)

การประมาณค่าพารามิเตอร์ของวิธี GLMMs จะใช้วิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood) อาศัยวิธีการหาอนุพันธ์ของฟังก์ชันภาวะน่าจะเป็น $L(\beta)$ อันดับหนึ่งที่หนึ่งเทียบกับพารามิเตอร์ที่ต้องการประมาณค่าแล้วกำหนดให้เท่ากับศูนย์ ซึ่งในทางปฏิบัติจะใช้ Log-Likelihood Function หรือ $\log L(\beta)$ แทน $L(\beta)$ สามารถแสดงได้ดังนี้ (McCulloch and Searle, 2001)

$$\text{จาก } l = \log \int f_{Y|U}(y|u)f_U(u)du$$

$$l = \log \left(\prod_{i=1}^m \int_{-\infty}^{\infty} \prod_{j=1}^{n_i} \frac{\mu_{ij}^{y_{ij}} e^{-\mu_{ij}}}{y_{ij}!} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}u_i^2} du_i \right) \quad (17)$$

$$= y'X\beta - \sum_{i,j} \log y_{ij}! + \sum_i \log \int_{-\infty}^{\infty} \exp \left\{ y_i \cdot u_i - \sum_j e^{X'_{ij}\beta + u_i} \right\} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}u_i^2} du_i \quad (18)$$

เมื่อ y_{ij} คือ ค่าสังเกตที่ i จากการวัดครั้งที่ j โดย $i = 1, 2, \dots, m$ และ $j = 1, 2, \dots, n_i$

จากสมการไม่สามารถหาค่าพารามิเตอร์ได้โดยตรงเนื่องจากสมการอยู่ในรูปแบบไม่เป็นเชิงเส้นในเทอมของพารามิเตอร์ จะทำการคำนวณค่าด้วยวิธีภาวะน่าจะเป็นสูงสุดแบบวนซ้ำ (Iterative Maximum Likelihood Estimator) โดยเริ่มต้นจากวิธีภาวะน่าจะเป็นสูงสุดหาค่าเริ่มต้นของการประมาณค่าพารามิเตอร์แล้วนำไปใช้ในกระบวนการคำนวณซ้ำต่อไป กระบวนการคำนวณซ้ำจะจบลงเมื่อค่าประมาณของพารามิเตอร์ลู่เข้า (Converge) สู่ค่าประมาณภาวะน่าจะเป็นสูงสุด โดย

สามารถใช้โปรแกรมสำเร็จรูปต่าง ๆ มาช่วยในการคำนวณได้ เช่น SAS STATA MIXPREG ฯลฯ กระบวนการคำนวณแบบวนซ้ำที่นำมาใช้ในการศึกษาครั้งนี้ คือ วิธีนิวตัน-ราฟสัน

ในการศึกษาครั้งนี้ได้ใช้คำสั่ง PROC GLIMMIX จากโปรแกรมสำเร็จรูป SAS สำหรับวิเคราะห์ตัวแบบผสมเชิงเส้นวงนัยทั่วไป (GLMMs) ซึ่งในการประมาณค่าพารามิเตอร์จะใช้ข้อมูลเพียงบางส่วนมาสร้างเป็นข้อมูลเทียม (Pseudo Data) สำหรับประมาณค่าพารามิเตอร์เริ่มต้นด้วยวิธี LMMs จากนั้นประมาณค่าความแปรปรวนร่วมของพารามิเตอร์ที่เหมาะสม โดยอาศัยการคำนวณแบบวนซ้ำไปเรื่อย ๆ จนค่าประมาณพารามิเตอร์ลู่เข้า (converge) ค่าคงที่ เรียกว่าวิธีการประมาณค่านี้ว่า Restricted Pseudo-likelihood (REPL) และได้ค่าประมาณของ β

จากหลักการ Pseudo Likelihood Estimate ซึ่งอาศัยการสมมติว่าทราบค่าพารามิเตอร์ β และมีค่าเท่ากับ $\hat{\beta}_p$ ที่ประมาณได้ ณ ขณะนั้น ต่อจากนั้นประมาณค่า β ด้วยวิธี ML โดยสมมติให้ u_i มีการแจกแจงแบบปกติ จะได้ค่าคาดหวังของเทอมอิทธิพลสุ่มที่สัมพันธ์กับระดับที่ i ของปัจจัยสุ่มเมื่อทราบค่าสังเกต y_i โดยเรียกค่าคาดหวังแบบมีเงื่อนไขนี้ว่า Best Linear Unbiased Predictors (BLUPs) ของเทอมอิทธิพลสุ่ม หรือ EBLUPs (Empirical BLUPs) รายละเอียดศึกษาได้จาก Wolfinger and O'Connell (1993) และ Brown and Prescott (1999)

ในการศึกษานี้ ตัวแปรตาม คือ จำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ที่มีการเก็บข้อมูลซ้ำในระยะเวลาติดตาม 5 ปี จะกำหนดให้ค่าคงที่ (Intercept) เป็นเทอมอิทธิพลสุ่มในวิธี GLMMs คือ ยอมให้แต่ละบุคคลจะมีความแตกต่างกันที่เรียกขึ้นว่า ค่าคงที่สุ่ม (Random Intercept) และสมมติให้โครงสร้างความแปรปรวนร่วมภายในบุคคลเดียวกันของค่าสังเกตที่มีการเก็บข้อมูลซ้ำในระยะเวลาติดตาม 5 ปี หรือ R_i มีลักษณะตามรูปแบบ AR(1) และ CS

5. การตรวจสอบความเหมาะสมของตัวแบบ

การตรวจสอบความเหมาะสมของตัวแบบวิธี GEE เมื่อกำหนดโครงสร้างความสัมพันธ์ของตัวแปรตามมีรูปแบบเป็น AR(1) และ CS สามารถตรวจสอบโดยอาศัยสถิติต่อไปนี้

1. คำนวณค่า Chi-Square of Residuals (χ^2) (Shoukri and Chaudhary, 2007) จาก

$$\chi^2 = \sum \left[\frac{y_{ij} - \hat{\mu}_{ij}}{\sqrt{v(\hat{\mu}_{ij})}} \right]^2$$

ตัวแบบมีความเหมาะสมจะมีค่าเข้าใกล้กับค่าองศาความเป็นอิสระ (Degree of Freedom)

2. Chi-Square of Residuals / DF เรียกว่าค่าdispersion เป็นค่าแสดงการกระจายของความคลาดเคลื่อนของค่าสังเกต โดยค่าที่เหมาะสมจะมีค่าเข้าใกล้ 1 แต่ไม่ควรมากกว่า 1 เพราะอาจทำให้เกิดกรณีที่ค่าความแปรปรวนมากเกินไปจริง (over-dispersion) คือ ค่า ϕ มีค่ามากกว่า 1 ทำให้ไม่เป็นไปตามข้อกำหนดของการแจกแจงแบบปัวส์ซงที่ ϕ มีค่าเท่ากับ 1 แสดงได้ว่าค่าสังเกตอาจจะไม่ได้มีการแจกแจงแบบปัวส์ซง

การตรวจสอบความเหมาะสมของตัวแบบวิธี GLMMs เมื่อกำหนดโครงสร้างความแปรปรวนร่วมของตัวแปรตามมีรูปแบบเป็น AR(1) และ CS สามารถตรวจสอบโดยอาศัยสถิติต่อไปนี้

1. Generalized Chi-Square (χ_g^2) (The GLMMIX Procedure, 2006) จากการใช้ PROC GLIMMIX คำนวณมาจาก

$$\chi_g^2 = \hat{r}'V(\hat{\theta})^{-1}\hat{r} ; \hat{r} \text{ คือค่าความคลาดเคลื่อน}$$

ตัวแบบมีความเหมาะสมจะมีค่าเข้าใกล้กับค่าองศาความเป็นอิสระ (Degree of Freedom)

2. Generalized Chi-Square / DF เรียกว่าค่าdispersion เป็นค่าแสดงการกระจายของความคลาดเคลื่อนของค่าสังเกต โดยค่าที่เหมาะสมจะมีค่าเข้าใกล้ 1 แต่ไม่ควรมากกว่า 1 เพราะอาจทำให้เกิดกรณีที่ค่าความแปรปรวนมากเกินไปจริงคือ ค่า ϕ มีค่ามากกว่า 1 ทำให้ไม่เป็นไปตามข้อกำหนดของการแจกแจงแบบปัวส์ซงที่ ϕ มีค่าเท่ากับ 1 แสดงได้ว่าค่าสังเกตอาจจะไม่ได้มีการแจกแจงแบบปัวส์ซง

3. ค่า Pseudo-AIC (Akaike's Information Criterion) และค่า Pseudo-BIC (Baye's

Information Criterion) สำหรับตรวจสอบความเหมาะสมของตัวแบบวิธี GLMMs เมื่อเลือกใช้โครงสร้างความแปรปรวนร่วมต่างกัน ค่า Pseudo-AIC หรือ Pseudo-BIC ของตัวแบบที่มีค่าต่ำกว่า แสดงว่าตัวแบบนั้นเป็นตัวแบบที่เหมาะสมกับข้อมูลมากกว่า (Shoukri and Chaudhary, 2007)

$$\text{Pseudo-AIC} = -2l + 2d$$

$$\text{Pseudo-BIC} = -2l + d \log_e n$$

เมื่อ l คือ Log Likelihood, d คือ จำนวนพารามิเตอร์ และ n คือ ขนาดตัวอย่าง

สำหรับการตรวจสอบการทำนายถูกต้องของตัวแบบวิธี GEE เมื่อกำหนดโครงสร้างความสัมพันธ์ของตัวแปรตามมีรูปแบบเป็น AR(1) และ CS และตัวแบบวิธี GLMMs เมื่อกำหนดโครงสร้างความแปรปรวนร่วมของตัวแปรตามมีรูปแบบเป็น AR(1) และ CS จะทำการคำนวณค่าร้อยละของการทำนายถูกต้อง (Percentage of Correct Predictions) หรือ R_p^2 ใช้วัดความเหมาะสมของตัวแบบจากการเปรียบเทียบค่าที่ทำนายได้จากตัวแบบกับค่าจริง หากตัวแบบใดให้ค่าร้อยละของการทำนายถูกต้องสูงกว่าจะเป็นตัวแบบที่เหมาะสมมากกว่าโดยสามารถคำนวณค่า R_p^2 ได้จากสูตรดังนี้ (Studenmund, 1992)

$$R_p^2 = \frac{\text{จำนวนค่าสังเกตที่ทำนายถูก}}{\text{จำนวนค่าสังเกตทั้งหมด}} \times 100$$

6. การสร้างตัวแบบทางสถิติ

การสร้างตัวแบบทางสถิติสำหรับข้อมูลจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ภาคสมัครใจ ประเภท 1 แบบระบุผู้ขับขี่ รถยนต์นั่งส่วนบุคคลรหัส 110 มีขั้นตอนดังนี้

1. ศึกษาลักษณะข้อมูลเบื้องต้นด้วยวิธีสถิติพรรณนา เช่น จำนวน ร้อยละ ค่าเฉลี่ย มัธยฐาน และแผนภูมิ Histogram รวมทั้งตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระ ด้วยค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมน ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน และค่า Phi และ Cramer's V หากตัว

แปรอิสระคู่ใดคู่หนึ่งมีความสัมพันธ์กันสูงด้วยค่าสัมประสิทธิ์สหสัมพันธ์มากกว่า 0.80 ก็จะเลือกใช้ตัวแปรอิสระเพียงตัวใดตัวหนึ่ง และนำตัวแปรอิสระที่เหมาะสมมาใช้ในการศึกษาต่อไป

2. การพัฒนาตัวแบบทางสถิติอาศัยหลักการของ Model Validity โดยแบ่งข้อมูล (Data Splitting) ซึ่งเรียกกันว่า Cross-Validation (Neter *et al.*, 2004) ทำการแบ่งข้อมูลเป็น 2 ชุด คือ ข้อมูลฝึกสอนสำหรับนำมาใช้เป็นชุดข้อมูลสำหรับสร้างตัวแบบ และข้อมูลทดสอบจะถูกนำมาใช้เพื่อทดสอบความถูกต้องของการทำนายจากตัวแบบที่พัฒนาจากชุดข้อมูลฝึกสอน ด้วยอัตราส่วน 70:30 โดยทำการสุ่มชุดข้อมูลแต่ละชุดด้วยวิธีการสุ่มตัวอย่างแบบง่าย (Simple Random Sampling) ดังนั้นในขั้นตอนการพัฒนาตัวแบบจะเลือกใช้ข้อมูลฝึกสอนเป็นหลัก

2.1 ศึกษาลักษณะข้อมูลจากทุกตัวแปรด้วยวิธีสถิติพรรณนาในชุดข้อมูลฝึกสอน และข้อมูลทดสอบว่ามีลักษณะเดียวกันกับชุดข้อมูลทั้งหมดหรือไม่

2.2 วิเคราะห์ตัวแปรอิสระทีละตัว (Univariate Analysis) แล้วนำตัวแปรอิสระที่มีค่า p-value น้อยกว่า 0.20 จากสถิติทดสอบ Wald ของตัวแบบวิธี GEE และจากสถิติทดสอบ t ของตัวแบบวิธี GLMMs หลังจากนั้นนำตัวแปรที่คัดกรองได้มาทำการคัดเลือกด้วยวิธีขั้นตอนก้ำกั้น (Forward Stepwise Method)

2.2.1 วิธี GEE จะพิจารณาการคัดเลือกตัวแปรอิสระเข้าและออกจากตัวแบบ ด้วยค่า Probability of stepwise entry (P_E) = 0.15 และ Probability of stepwise removal (P_R) = 0.20 (Hosmer and Lemeshow, 2000) ของตัวทดสอบสถิติ Generalized Score Statistics (SAS/STAT User's Guide V.8, 1999) ซึ่งคำนวณมาจาก

$$T_{SC} = S(\tilde{\beta})' \Sigma_m L' (L \Sigma_e L')^{-1} L \Sigma_m S(\tilde{\beta})$$

เมื่อ Σ_m คือ ค่าประมาณของเมทริกซ์ความแปรปรวนร่วมจากตัวแบบ (model-based) Σ_e คือ ค่าประมาณของเมทริกซ์ความแปรปรวนร่วมจากตัวแบบที่ศึกษาจากข้อมูล (model-empirical) และ T_{SC} มีการแจกแจงเข้าสู่การแจกแจงไคสแควร์ที่องศาความเป็นอิสระเท่ากับ r เมื่อ r คือจำนวนตัวแปรอิสระที่พิจารณาเข้าตัวแบบ

2.2.2 วิธี GLMMs จะพิจารณาการคัดเลือกตัวแปรอิสระเข้าและออกจากตัวแบบ จากค่าสถิติทดสอบ G (Residual Log Pseudo-Likelihood Ratio Test Statistic) เทียบกับค่าสถิติไคสแควร์ (χ^2_{df}) เมื่อกำหนดค่า Probability of stepwise entry (P_E) = 0.15 และ Probability of stepwise removal (P_R) = 0.20 (Hosmer and Lemeshow, 2000) มีหลักการคำนวณคือ

$$G = -2[\log(L_1) - \log(L_2)] \sim \chi^2_{df}$$

เมื่อ df คือ ผลต่างของจำนวนตัวแปรอิสระ (Shoukri and Chaudhary, 2007)

3. ทำการประมาณค่าพารามิเตอร์จากตัวแบบสุดท้ายที่ได้จากข้อมูลฝึกสอนโดยที่ตัวแบบวิธี GEE จะประมาณค่าพารามิเตอร์ด้วยหลักการ Quasi-Likelihood ส่วนตัวแบบวิธี GLMMs จะประมาณค่าพารามิเตอร์ที่ด้วยวิธี ML และได้ค่าคาดหวังของเทอมอิทธิพลสุ่มซึ่งเรียกว่า BLUPs รวมทั้งทำการประมาณค่าพารามิเตอร์ของตัวแบบในชุดข้อมูลทั้งหมดโดยใช้ตัวแปรอิสระที่ได้จากการคัดเลือกในชุดข้อมูลฝึกสอน

4. ทำนายจำนวนการเรียกค่าสินไหมทดแทนในชุดข้อมูลฝึกสอน และนำตัวแบบที่ประมาณค่าพารามิเตอร์จากข้อมูลฝึกสอนมาทำนายจำนวนการเรียกค่าสินไหมทดแทนในข้อมูลทดสอบ แล้วคำนวณค่าร้อยละการทำนายถูกต้อง (R_p^2) พร้อมทั้งทำนายค่าจำนวนการเรียกค่าสินไหมทดแทนจากตัวแบบในชุดข้อมูลทั้งหมด และคำนวณค่าร้อยละการทำนายถูกต้อง R_p^2 แล้วนำมาเปรียบเทียบประสิทธิภาพในการทำนายกับข้อมูลฝึกสอนและข้อมูลทดสอบ เพื่อนำการวิเคราะห์ข้อมูลทั้งหมดไปใช้ศึกษาต่อไป

5. ตรวจสอบความเหมาะสมของตัวแบบทางสถิติจากการพิจารณาค่าสถิติ Generalized Chi-Square ค่า Generalized Chi-Square/DF ค่า Pseudo-AIC และค่า Pseudo-BIC สำหรับวิธี GLMMs ส่วนการพิจารณาค่าสถิติสำหรับวิธี GEE คือ ค่าสถิติ Chi-Square of Residuals และค่า Chi-Square of Residuals/DF นอกจากนี้ยังคำนวณค่าร้อยละการทำนายถูกต้องของการทำนายค่าที่ได้จากตัวแบบ วิธี GEE และวิธี GLMMs จากค่า R_p^2 ตามที่ได้อธิบายไว้ในหน้า 29

ส่วนที่ 3 งานวิจัยที่เกี่ยวข้อง

จากการทบทวนงานวิจัยที่เกี่ยวข้องกับการศึกษาตัวแบบสำหรับข้อมูลจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ ได้แบ่งเป็น 3 ส่วน ดังนี้ 1) ปัจจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ 2) ปัจจัยต่าง ๆ ที่เกี่ยวข้องกับการประกันภัยรถยนต์ และ 3) งานวิจัยที่เกี่ยวข้องกับวิธีทางสถิติสำหรับตัวแบบจำนวนการเรียกค่าสินไหมทดแทนของการประกันภัยรถยนต์ มีดังนี้

1. ปัจจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์

งานวิจัยเกี่ยวกับการศึกษาปัจจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ สำหรับในประเทศไทยยังไม่มีการศึกษาตัวแปรตามที่เป็นจำนวนการเรียกค่าสินไหมทดแทนโดยตรง ส่วนใหญ่จะมีความศึกษาในงานวิจัยต่างประเทศ จากงานศึกษาของ Baxter *et al.* (1989) ทำการวิเคราะห์ข้อมูลด้วยวิธี Weighted Normal-Theory Linear Model เมื่อตัวแปรตาม คือ จำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ และมีปัจจัยที่มีผลต่อตัวแปรตาม 3 ปัจจัย คือ 1) กลุ่มอายุของผู้ทำประกันภัยรถยนต์ โดยแบ่งเป็น 8 กลุ่ม ได้แก่ 17-20 ปี, 21-24 ปี, 25-29 ปี, 30-34 ปี, 35-39 ปี, 40-49 ปี, 50-59 ปี และ 60 ปีขึ้นไป 2) กลุ่มรถยนต์ ซึ่งแบ่งเป็น 4 กลุ่ม คือ กลุ่ม A, กลุ่ม B, กลุ่ม C และกลุ่ม D และ 3) กลุ่มอายุรถยนต์ แบ่งเป็น 4 กลุ่ม คือ 0-3 ปี, 4-7 ปี, 8-9 ปี และ 10 ปีขึ้นไป ซึ่งการศึกษานี้ได้นำปัจจัยเหล่านี้มาทำการศึกษาโดยข้อมูลทั้งหมดเป็นข้อมูลเชิงคุณภาพเท่านั้น ไม่มีปัจจัยหรือตัวแปรอิสระที่เป็นข้อมูลเชิงปริมาณ จึงอธิบายรายละเอียดของวิเคราะห์ได้ไม่มากนัก ส่วนงานศึกษาวิจัยอื่นๆ ที่มีการนำตัวแปรอิสระที่เป็นข้อมูลเชิงปริมาณเข้าอธิบาย ได้แก่ งานศึกษาของ Yau *et al.* (2003) ได้ทำการศึกษาดัชนีของข้อมูลที่มีการวัดซ้ำของจำนวนการเรียกค่าสินไหมทดแทนของประกันภัยรถยนต์ด้วยวิธีการวิเคราะห์การถดถอยแบบปัวซอง (Poisson Regression Analysis) มีปัจจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนของประกันภัยรถยนต์ คือ คะแนนในการขับรถยนต์ ประเภทของการใช้รถยนต์ (ใช้ส่วนตัวกับใช้เพื่อการค้า) รายได้ เขตที่อยู่อาศัย (ในเมืองกับนอกเมือง) ที่มีนัยสำคัญสถิติที่ระดับ 0.10 นอกจากนี้ยังมีงานศึกษาของ Bartoszewicz (2005) ได้ทำการสร้างตัวแบบจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ด้วยการวิเคราะห์การถดถอยปัวซองและการวิเคราะห์ถดถอยแบบทวินามนิเสธ มีปัจจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ คือ กลุ่มอายุของผู้ทำประกันภัยรถยนต์ โดยแบ่งเป็น 8 กลุ่ม ได้แก่ 17-20 ปี, 21-24 ปี, 25-29 ปี, 30-34 ปี, 35-39

ปี, 40-49 ปี, 50-59 ปี และ 60 ปีขึ้นไป กลุ่มรถยนต์ ซึ่งแบ่งเป็น 4 กลุ่ม คือ กลุ่ม A, กลุ่ม B, กลุ่ม C และกลุ่ม D และกลุ่มอายุรถยนต์ แบ่งเป็น 4 กลุ่ม คือ 0-3 ปี, 4-7 ปี และ 8 ปีขึ้นไป ซึ่งจะมีปัจจัยที่นำมาศึกษาเช่นเดียวกับ Baxter *et al.* (1989)

จากการศึกษาที่ผ่านมาพบว่าปัจจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ แบ่งเป็นปัจจัยที่เป็นข้อมูลเชิงคุณภาพ คือ 1) กลุ่มอายุของผู้ทำประกันภัยรถยนต์ 2) กลุ่มรถยนต์ 3) กลุ่มอายุรถยนต์ 4) ประเภทของการใช้รถยนต์ และ 5) เขตที่อยู่อาศัย ส่วนปัจจัยที่เป็นข้อมูลเชิงปริมาณ คือ 1) คະแนนในการขับขี และ 2) รายได้ แต่เนื่องจากข้อมูลที่มีการเก็บในประเทศไทยที่รวบรวมได้จากกรมการประกันภัยไม่มีรายละเอียดของข้อมูลตัวแปรเหล่านี้ ประเภทของการใช้รถยนต์ พื้นที่อยู่อาศัย คະแนนในการขับขี และรายได้ ดังนั้นจึงได้ทำการทบทวนเอกสารเพิ่มเติมเกี่ยวกับปัจจัยที่เกี่ยวข้องกับการประกันภัยรถยนต์ที่จะกล่าวต่อไป

2. ปัจจัยที่เกี่ยวข้องการประกันภัยรถยนต์

2.1 เพศ

จากการศึกษาของประทีป (2540) ได้ทำการวิเคราะห์ปัจจัย (Factor Analysis) พบว่า เพศเป็นปัจจัยที่มีผลต่อระดับความเสี่ยงภัยสูงสุดของการประกันภัยรถยนต์ส่วนบุคคลชนิตรายหนึ่ง อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ส่วนจารุณี (2545) ได้ศึกษาความพึงพอใจของผู้ทำประกันภัยที่มีต่อบริษัทประกันภัย ศึกษาเฉพาะกรณีการประกันภัยส่วนบุคคลภาคสมัครใจในเขตกรุงเทพมหานครโดยวิเคราะห์ด้วยสถิติทดสอบ t พบว่า ปัจจัยเพศแตกต่างกันจะมีความพึงพอใจต่อบริการของบริษัทประกันภัยแตกต่างกันที่ระดับนัยสำคัญทางสถิติ 0.01 และ 0.05 แต่ผลการศึกษาของสิริกุล (2544) เกี่ยวกับความคิดเห็นของผู้ทำประกันภัยที่มีต่อการประกันภัยรถยนต์แบบใหม่ พบว่า เพศไม่มีผลต่อความคิดเห็นเรื่องของการประกันภัยรถยนต์แบบใหม่ที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ของตัวทดสอบสถิติ t เช่นเดียวกับการศึกษาของอดิชาติ (2548) ที่ศึกษาปัจจัยที่มีผลต่อการตัดสินใจทำประกันภัยรถยนต์ภาคสมัครใจ ประเภท 1 พบว่า เพศไม่มีผลต่อความแตกต่างในการตัดสินใจที่ระดับนัยสำคัญทางสถิติ 0.05 ของตัวทดสอบสถิติ t นอกจากนี้ Yau *et al.* (2003) ได้ทำการศึกษาตัวแบบของข้อมูลที่มีการวัดซ้ำของจำนวนการเรียกค่าสินไหมทดแทนของประกันภัยรถยนต์ด้วยวิธีการวิเคราะห์การถดถอยแบบปัวส์ซง (Poisson Regression

Analysis) พบว่า เพศไม่มีผลต่อการจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ที่ระดับนัยสำคัญทางสถิติ 0.05

2.2 อายุของผู้ทำประกันภัยรถยนต์

ปัจจัยด้านอายุที่มีผลต่อการประกันภัยรถยนต์จากงานศึกษาของประทีป (2540) พบว่า อายุเป็นปัจจัยที่มีผลต่อระดับความเสี่ยงภัยสูงสุดของการประกันภัยรถยนต์ส่วนบุคคลชนิตรายณ์ต์บรรทุทอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 จากวิธีการวิเคราะห์ปัจจัย ส่วนผลการศึกษาของจากรุที (2545) ได้ศึกษาความพึงพอใจของผู้ทำประกันภัยที่มีต่อบริษัทประกันภัย ศึกษาเฉพาะกรณีการประกันภัยส่วนบุคคลภาคสมัครใจ ในเขตกรุงเทพมหานคร พบว่าอายุแตกต่างกันมีความพึงพอใจต่อบริการของบริษัทประกันภัยแตกต่างกันที่ระดับนัยสำคัญทางสถิติ 0.01 และ 0.05 ด้วยสถิติทดสอบ t แต่ในผลการศึกษาของสิริกุล (2544) ได้ศึกษาความคิดเห็นของผู้ทำประกันภัยที่มีต่อการประกันภัยรถยนต์แบบใหม่ พบว่า ปัจจัยอายุที่แตกต่างกันมีผลต่อความคิดเห็นเรื่องของการประกันภัยรถยนต์แบบใหม่ไม่แตกต่างกันที่ระดับนัยสำคัญทางสถิติ 0.05 และผลการศึกษาของอดิชาติ (2548) ได้ศึกษาปัจจัยที่มีผลต่อการตัดสินใจทำประกันภัยรถยนต์ภาคสมัครใจ ประเภท 1 พบว่า ปัจจัยอายุที่แตกต่างกันไม่มีผลต่อการตัดสินใจโดยรวมในการทำประกันภัยรถยนต์ภาคสมัครใจที่ระดับนัยสำคัญทางสถิติ 0.05 จากสถิติทดสอบ t เช่นเดียวกัน ซึ่งจะมีความแตกต่างจากการศึกษาของจากรุที (2545) นอกจากนี้ผลการศึกษาของนันทยา (2545) ได้ทำการศึกษาเปรียบเทียบต้นทุนแลผลประโยชน์ของการประกันภัยรถยนต์ภาคสมัครใจ กรณีกรมธรรม์ประเภท 1 ด้วยวิธีการวิเคราะห์การถดถอย (Regression Analysis) พบว่า ปัจจัยเสี่ยงที่มีผลต่อการคำนวณเบี้ยประกันภัยที่ระดับนัยสำคัญทางสถิติ 0.05 คือ อายุผู้ขับขี่ เช่นเดียวกับผลการศึกษาของอรณพ (2545) ได้ศึกษาปัจจัยที่สำคัญที่มีอิทธิพลต่อการทำประกันภัยภาคสมัครใจ พบว่า อายุเป็นปัจจัยสำคัญที่มีผลต่อการตัดสินใจทำประกันภัยที่ระดับนัยสำคัญทางสถิติ 0.05

2.3 อาชีพของผู้ทำประกันภัยรถยนต์

ปัจจัยด้านอาชีพที่มีผลต่อการประกันภัยรถยนต์จากการศึกษาของประทีป (2540) พบว่า อาชีพเป็นปัจจัยที่มีผลต่อระดับความเสี่ยงภัยสูงสุดของการประกันภัยรถยนต์ส่วนบุคคลชนิตรายณ์ต์โดยสารอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 จากการวิเคราะห์ปัจจัย และจากผลการศึกษาของสิริกุล (2544) ได้ศึกษาความคิดเห็นของผู้ทำประกันภัยที่มีต่อการประกันภัยรถยนต์แบบใหม่

ด้วยสถิติทดสอบ t พบว่า ปัจจัยอาชีพที่แตกต่างกันมีผลต่อความคิดเห็นเรื่องของการประกันภัยรถยนต์แบบใหม่ไม่แตกต่างกันที่ระดับนัยสำคัญทางสถิติ 0.05 เช่นเดียวกับผลการศึกษาของอดิชาติ (2548) ได้ศึกษาปัจจัยที่มีผลต่อการตัดสินใจทำประกันภัยรถยนต์ภาคสมัครใจ ประเภท 1 พบว่า ปัจจัยอาชีพที่แตกต่างกันไม่มีผลต่อการตัดสินใจโดยรวมในการทำประกันภัยรถยนต์ภาคสมัครใจที่ระดับนัยสำคัญทางสถิติ 0.05 ของสถิติทดสอบ t หรือมีการตัดสินใจไม่แตกต่างกัน

2.4 อายุรถยนต์

นันทิยา (2545) ได้ทำการศึกษาเปรียบเทียบต้นทุนและผลประโยชน์ของการประกันภัยรถยนต์ภาคสมัครใจ กรณีกรมธรรม์ประเภท 1 ด้วยวิธีการวิเคราะห์การถดถอย พบว่า ปัจจัยเสี่ยงที่มีผลต่อการคำนวณเบี้ยประกันภัยที่ระดับนัยสำคัญทางสถิติ 0.05 คือ อายุรถยนต์ เช่นเดียวกับผลการศึกษาของอรธณพ (2545) ได้ศึกษาปัจจัยสำคัญที่มีอิทธิพลต่อการทำประกันภัยภาคสมัครใจ พบว่า อายุรถยนต์เป็นปัจจัยสำคัญที่มีผลต่อการตัดสินใจทำประกันภัยที่ระดับนัยสำคัญทางสถิติ 0.05 ด้วยวิธีการวิเคราะห์การถดถอย

2.5 ค่าสินไหมทดแทน

ปัจจัยด้านค่าสินไหมทดแทนที่มีผลต่อการประกันภัยรถยนต์จากการวิเคราะห์การถดถอย ได้แก่ การศึกษาของสมภพ (2545) ที่วิเคราะห์อุปสงค์ของการประกันภัยรถยนต์ภาคสมัครใจในประเทศไทย พบว่า ค่าสินไหมทดแทนจะมีความสัมพันธ์กับอุปสงค์ของการประกันภัยรถยนต์ภาคสมัครใจที่ระดับนัยสำคัญทางสถิติ 0.05 และจากการศึกษาของปีทมา (2547) เกี่ยวกับปัจจัยที่มีผลต่ออุปสงค์และแนวโน้มของการประกันภัยรถยนต์ภาคสมัครใจ พบว่า ปัจจัยค่าสินไหมทดแทนมีผลต่อจำนวนกรมธรรม์ประกันภัยรถยนต์ภาคสมัครใจที่ระดับนัยสำคัญทางสถิติ 0.05 นอกจากนี้ กันทิมา (2545) ได้คำนวณอัตราเบี้ยประกันการประกันภัยรถยนต์ จากผลคูณระหว่างค่าสินไหมทดแทนเฉลี่ยกับจำนวนการเรียกค่าสินไหมทดแทน ทำให้สามารถอธิบายได้ว่าค่าสินไหมทดแทนมีความสัมพันธ์กับจำนวนการเรียกค่าสินไหมทดแทน

2.6 ขนาดเครื่องยนต์

ปัจจัยด้านขนาดเครื่องยนต์ที่มีผลต่อการประกันภัยรถยนต์จากการวิเคราะห์ด้วยวิธีการวิเคราะห์การถดถอย ได้แก่ ผลการศึกษาของนันทยา (2545) ที่วิเคราะห์เปรียบเทียบต้นทุนและผลประโยชน์ของการประกันภัยรถยนต์ภาคสมัครใจ กรณีกรมธรรม์ประเภท 1 พบว่าปัจจัยเสี่ยงที่มีผลต่อการคำนวณเบี้ยประกันภัยน้อยที่สุดที่ระดับนัยสำคัญทางสถิติ 0.05 คือ ขนาดเครื่องยนต์ และกันทิมมา (2545) ได้วิเคราะห์ปัจจัยที่มีผลต่อความเสียหายเพื่อกำหนดอัตราเบี้ยประกันภัยรถยนต์ในประเทศไทย พบว่าปัจจัยที่มีนัยสำคัญทางสถิติ 0.05 คือ ขนาดเครื่องยนต์

2.7 กลุ่มรถยนต์

ปัจจัยด้านกลุ่มรถยนต์ที่มีผลต่อการประกันภัยรถยนต์จากการวิเคราะห์ด้วยวิธีการวิเคราะห์การถดถอย ได้แก่ นันทยา (2545) ได้ทำการศึกษาเปรียบเทียบต้นทุนและผลประโยชน์ของการประกันภัยรถยนต์ภาคสมัครใจ กรณีกรมธรรม์ประเภท 1 พบว่า ปัจจัยเสี่ยงที่มีผลต่อการคำนวณเบี้ยประกันภัยมากที่สุดที่ระดับนัยสำคัญทางสถิติ 0.05 คือ กลุ่มรถยนต์ และกันทิมมา (2545) ได้ศึกษาการวิเคราะห์ปัจจัยที่มีผลต่อความเสียหายเพื่อกำหนดอัตราเบี้ยประกันภัยรถยนต์ในประเทศไทย พบว่าปัจจัยที่สำคัญต่อการศึกษาระดับนัยสำคัญทางสถิติ 0.05 คือ กลุ่มรถยนต์

จะเห็นว่าปัจจัยเพศ อายุของผู้ทำประกันภัยรถยนต์ อาชีพของผู้ทำประกันภัยรถยนต์ อายุรถยนต์ ค่าสินไหมทดแทน ขนาดเครื่องยนต์ และกลุ่มรถยนต์ ยังไม่มีข้อสรุปที่ชัดเจนว่ามีผลต่อจำนวนการเรียกค่าสินไหมทดแทนอย่างมีนัยสำคัญทางสถิติ นอกจากนี้โครงสร้างการประกันภัยรถยนต์ภาคสมัครใจ เดิมเป็นโครงสร้างที่ใช้มาตั้งแต่ปี 2523 และไม่มีการปรับปรุงจนกระทั่งปี 2545 โดยอัตราเบี้ยประกันภัยเดิมจะกำหนดตามปัจจัยด้านประเภทรถยนต์ ชนิดของรถยนต์ และขนาดของรถยนต์เป็นหลัก ซึ่งไม่ได้นำปัจจัยอื่น ๆ ที่มีผลต่อความเสี่ยงภัยมารวมพิจารณา ต่อมาในปี 2545 กรมการประกันภัยได้กำหนดให้การคำนวณพิกัดอัตราเบี้ยประกันภัยที่ประกอบด้วยปัจจัยต่าง ๆ คือ ประเภทของรถยนต์ ยี่ห้อรถยนต์หรือกลุ่มรถยนต์ ลักษณะการใช้รถ ขนาดของเครื่องยนต์ อายุผู้ทำประกัน และอายุกรมธรรม์เข้ามาพิจารณา เพื่อก่อให้เกิดความเป็นธรรมแก่ผู้ทำประกันภัยมากขึ้น (อรณพ, 2545) ดังนั้นการสรุปงานวิจัยที่ผ่านมาในข้อ 1 และ 2 ซึ่งตามหลักการคำนวณพิกัดอัตราเบี้ยประกันภัยของกรมประกันภัย เห็นว่า ปัจจัยที่ควรนำมาพิจารณาในการศึกษานี้ คือ ค่าสินไหมทดแทน อายุผู้ทำประกันภัย อายุรถยนต์ เพศ อาชีพ กลุ่มรถยนต์ และขนาดเครื่องยนต์

3. งานวิจัยที่เกี่ยวข้องกับวิธีทางสถิติสำหรับตัวแบบจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์

จากการทบทวนวรรณกรรมงานวิจัยที่ผ่านมาในประเทศไทยตามที่ปรากฏในข้อ 1 และ 2 พบว่า มีการใช้วิธีวิเคราะห์ที่แตกต่างกันไปตามลักษณะของข้อมูล แต่ไม่มีการศึกษาที่ใช้ข้อมูลจำนวนการเรียกค่าสินไหมทดแทนมาวิเคราะห์โดยตรง หรือนำมาสร้างตัวแบบทางสถิติสำหรับการประกันภัยรถยนต์

ส่วนงานวิจัยที่เกี่ยวข้องในต่างประเทศมีการสร้างตัวแบบทางสถิติสำหรับข้อมูลจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ เช่น Bartoszewicz (2005) ทำการสร้างตัวแบบจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ด้วยการวิเคราะห์การถดถอยปัวส์ซง และการวิเคราะห์ถดถอยแบบทวินามนิเสธ มีปัจจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ คือ กลุ่มอายุของผู้ทำประกันภัยรถยนต์ โดยแบ่งเป็น 8 กลุ่ม ได้แก่ 17-20 ปี, 21-24 ปี, 25-29 ปี, 30-34 ปี, 35-39 ปี, 40-49 ปี, 50-59 ปี และ 60 ปีขึ้นไป กลุ่มรถยนต์ ซึ่งแบ่งเป็น 4 กลุ่ม คือ กลุ่ม A, กลุ่ม B, กลุ่ม C และกลุ่ม D และกลุ่มอายุรถยนต์ แบ่งเป็น 4 กลุ่ม คือ 0-3 ปี, 4-7 ปี และ 8 ปีขึ้นไป พบว่า ข้อมูลที่นำมาศึกษาเป็นจำนวนการนับของจำนวนการเรียกค่าสินไหมทดแทนที่มีค่าศูนย์รวมอยู่ด้วยเป็นจำนวนมาก ทำให้การสร้างตัวแบบข้อมูลจำนวนการเรียกค่าสินไหมทดแทนที่วิเคราะห์ด้วยการถดถอยเมื่อตัวแปรตามมีการแจกแจงปัวส์ซงมีความเหมาะสมน้อยเนื่องจากเกิดปัญหาการกระจายของข้อมูลที่มากเกินไป (over-dispersion) จึงแก้ปัญหาโดยการพัฒนาตัวแบบจำนวนการเรียกค่าสินไหมทดแทนด้วยตัวแบบการถดถอยแบบทวินามนิเสธ พบว่า ตัวแบบการถดถอยแบบทวินามนิเสธมีความเหมาะสมมากกว่าตัวแบบการถดถอยแบบปัวส์ซง และไม่เกิดปัญหา over-dispersion

การสร้างตัวแบบสำหรับจำนวนการเรียกค่าสินไหมทดแทนที่มีการติดตามระยะยาวของผู้ทำประกันภัยคนเดียวกัน มีงานวิจัยที่เกี่ยวข้องเช่น Yau *et al.* (2003) ทำการสร้างตัวแบบสำหรับข้อมูลการวัดซ้ำของจำนวนการเรียกค่าสินไหมทดแทนของประกันภัยรถยนต์ โดยการจำลองข้อมูลจากฐานข้อมูลจริงจำนวน 500 ตัวอย่าง แต่ละตัวอย่างมีการวัดซ้ำในระยะเวลาติดตาม 5 ปี จึงมีค่าสังเกตทั้งหมด 2,500 ค่าสังเกต เมื่อตัวแปรตาม คือ จำนวนการเรียกค่าสินไหมทดแทน มีการแจกแจงแบบปัวส์ซง และตัวแปรอิสระที่พิจารณาในตัวแบบผลการศึกษาคือ ปัจจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนที่ระดับนัยสำคัญทางสถิติ 0.10 คือ คะแนนในการขับขี้อย่างไรได้ เพศประเภทการใช้งานของรถยนต์ และเขตที่อยู่อาศัย และมีการเปรียบเทียบวิธีการวิเคราะห์ด้วยตัว

แบบเชิงเส้นวางนัยทั่วไป (GLMs) สำหรับข้อมูลติดตามระยะยาว และตัวแบบผสมเชิงเส้นวางนัยทั่วไป (GLMMs) เมื่อกำหนดให้ความสัมพันธ์ของข้อมูลวัดซ้ำเป็นแบบอิสระกัน และกำหนดค่าการกระจายของข้อมูล dispersion เป็น 3 ค่า คือ 0.5, 0.8 และ 1 พบว่า ตัวแบบผสมเชิงเส้นวางนัยทั่วไป (GLMMs) ให้ค่าความคลาดเคลื่อนของการประมาณค่าสัมประสิทธิ์การถดถอยมีค่าน้อยกว่าที่ได้จากตัวแบบเชิงเส้นวางนัยทั่วไป (GLMs) รวมทั้งค่าสถิติ Deviance ที่น้อยกว่าและค่าสถิติ Pearson Chi-square ที่มากกว่าจากการตรวจสอบความเหมาะสมของตัวแบบ แสดงว่า การวิเคราะห์ข้อมูลสำหรับการวัดซ้ำด้วยตัวแบบผสมเชิงเส้นวางนัยทั่วไป (GLMMs) มีความเหมาะสมมากกว่าตัวแบบเชิงเส้นวางนัยทั่วไป (GLMs) เมื่อข้อมูลมีค่า dispersion มีขนาดเท่ากับ 1

จะเห็นว่าวิธีการวิเคราะห์ทางสถิติในงานวิจัยที่ผ่านมาสำหรับข้อมูลจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ที่นำมาสร้างตัวแบบ สามารถสรุปได้ว่า งานวิจัยภายในประเทศไม่มีการนำจำนวนการเรียกค่าสินไหมทดแทนมาสร้างตัวแบบสำหรับข้อมูลโดยตรง ส่วนงานวิจัยต่างประเทศจะมีการนำจำนวนการเรียกค่าสินไหมทดแทนมาสร้างตัวแบบด้วยวิธีการวิเคราะห์การถดถอยเมื่อตัวแปรตามมีการแจกแจงแบบปัวส์ซง และการแจกแจงแบบทวินามนิเสธ หรือเรียกว่าการสร้างตัวแบบเชิงเส้นวางนัยทั่วไป (GLMs) และการวิเคราะห์เมื่อข้อมูลจำนวนการเรียกค่าสินไหมทดแทนที่มีการเก็บข้อมูลระยะยาวด้วยวิธีตัวแบบผสมเชิงเส้นวางนัยทั่วไป (GLMMs) ซึ่งมีประสิทธิภาพมากกว่าการวิเคราะห์ด้วยวิธีตัวแบบเชิงเส้นวางนัยทั่วไป (GLMs) และเหมาะสมกับข้อมูลมากกว่า อย่างไรก็ตามการวิเคราะห์ข้อมูลที่มีการติดตามระยะยาวสามารถใช้วิธีการวิเคราะห์ Generalized Estimating Equations (GEE) ซึ่งส่วนใหญ่มักจะปรากฏในการวิเคราะห์ข้อมูลทางการแพทย์ ซึ่งผลการวิเคราะห์จะใกล้เคียงกับวิธีตัวแบบผสมเชิงเส้นวางนัยทั่วไป (GLMMs) เมื่อพิจารณาผลในภาพรวมของค่าเฉลี่ยของประชากร ดังนั้นในการศึกษานี้จึงนำวิธีตัวแบบผสมเชิงเส้นวางนัยทั่วไป (GLMMs) และวิธี GEE มาใช้ในการสร้างตัวแบบจำนวนการเรียกค่าสินไหมทดแทนที่มีการเก็บข้อมูลซ้ำในระยะเวลาติดตาม 5 ปีของผู้ทำประกันภัยคนเดียวกันที่มีการเรียกค่าสินไหมทดแทนอย่างน้อยหนึ่งครั้ง

อุปกรณ์และวิธีการ

อุปกรณ์

1. เครื่องไมโครคอมพิวเตอร์ที่ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์
2. โปรแกรม SAS (Statistical Analysis System) version 9.13
3. ข้อมูลทุติยภูมิการประกันภัยรถยนต์ภาคสมัครใจประเภท 1 แบบระบุชื่อผู้ขับขี่ รหัส 110 รถยนต์นั่งส่วนบุคคล ที่จดทะเบียนในกรุงเทพมหานคร ตั้งแต่ปี 2544-2548 ของผู้ทำประกันภัยรถยนต์ที่เริ่มทำประกันภัยปี 2544 จากกรมการประกันภัย กระทรวงพาณิชย์

วิธีการ

การศึกษาตัวแบบสำหรับข้อมูลจำนวนการเรียกค่าสินไหมทดแทนจากผู้ทำประกันภัยรถยนต์ที่เริ่มทำประกันปี 2544 เฉพาะในกรุงเทพมหานคร อาศัยข้อมูลจากกรมการประกันภัย กระทรวงพาณิชย์ มีวิธีการดังนี้

1. รวบรวมข้อมูลที่ใช้ในการศึกษานี้ประกอบด้วย ตัวแปรตาม คือข้อมูลจำนวนการเรียกค่าสินไหมทดแทนการประกันภัยรถยนต์ที่มีการเก็บค่าในระยะเวลาติดตาม 5 ปีของผู้ทำประกันภัยแต่ละราย ส่วนตัวแปรอิสระประกอบด้วยตัวแปรต่อไปนี้

1.1 เพศ หมายถึง เพศของผู้ทำประกันภัย หรือเพศของเจ้าของกรมธรรม์ประกันภัยรถยนต์

1.2 อาชีพผู้ทำประกันภัยรถยนต์ หมายถึง อาชีพของผู้ทำประกันภัย หรืออาชีพของเจ้าของกรมธรรม์ประกันภัยรถยนต์ (ประทิบ, 2540) ซึ่งแบ่งเป็น 2 กลุ่ม คือ

- อาชีพที่มีความเสี่ยงมาก ได้แก่ นักข่าว ผู้รับเหมาก่อสร้าง วิศวกร และสถาปนิก เป็นต้น

- อาชีพที่มีความเสี่ยงน้อย ได้แก่ ครู พยาบาล แพทย์ แม่บ้าน เจ้าของกิจการ และอาจารย์ เป็นต้น

1.3 อายุของผู้ทำประกันภัยรถยนต์ (ปี) หมายถึง อายุของผู้ทำประกันภัย หรืออายุของเจ้าของกรรมสิทธิ์ประกันภัยรถยนต์ โดยมีการแบ่งช่วงอายุของผู้ขับขี่จากช่วงที่มีความเสี่ยงภัยน้อย ไปยังช่วงที่มีความเสี่ยงภัยมากเป็น 4 ช่วงอายุ(กรมการประกันภัย, 2550) ดังนี้

- อายุเกิน 50 ปีขึ้นไป
- อายุ 36-50 ปี
- อายุ 25-35 ปี
- อายุ 18-24 ปี

1.4 ขนาดของเครื่องยนต์ (cc.) หมายถึง ขนาดของเครื่องยนต์ของรถที่ทำประกันภัย ซึ่งแบ่งเป็น 2 ระดับ (ก้นทิมมา, 2545) คือ

- กลุ่มที่ 1 ขนาดเครื่องยนต์น้อยกว่าเท่ากับ 2000 cc.
- กลุ่มที่ 2 ขนาดเครื่องยนต์มากกว่า 2000 cc.

1.5 อายุรถยนต์ (ปี) หมายถึง อายุการใช้งานของรถยนต์ที่ทำการประกันภัย

1.6 ค่าสินไหมทดแทนหรือค่าเสียหาย (บาท) หมายถึง ค่าเสียหายที่ได้รับในกรณีที่เกิดความเสียหายของตัวรถยนต์เมื่อทำการซ่อม

1.7 กลุ่มรถยนต์ หมายถึง การจัดกลุ่มของรถยนต์ที่ทำประกันภัยไว้ แบ่งออกเป็น 5 กลุ่ม คือ กลุ่มที่ 1 กลุ่มที่ 2 กลุ่มที่ 3 กลุ่มที่ 4 และกลุ่มที่ 5 โดยแสดงรายละเอียดไว้ในภาคผนวก

ตารางที่ 4 ลักษณะข้อมูลการประกันภัยรถยนต์ที่เฝ้าติดตามในระยะเวลา 5 ปี ตั้งแต่

ปี พ.ศ.2544-2548 จากตัวอย่างหน่วยศึกษาที่เป็นผู้ทำประกันภัยจำนวน 4 คน

ปี พ.ศ.	บุคคล ที่	จำนวน การเรียก ค่าสินไหม ทดแทน	เพศ	อาชีพ	กลุ่ม รถยนต์	กลุ่มของ ขนาด เครื่องยนต์	อายุผู้ทำ ประกัน (ปี)	อายุ รถยนต์ (ปี)	ค่าสินไหม ทดแทน (บาท)
2544	1	0	หญิง	ครู	2	1	31	1	0.00
2545	1	1	หญิง	ครู	2	1	32	2	769.16
2546	1	1	หญิง	ครู	2	1	33	3	769.16
2547	1	1	หญิง	ครู	2	1	34	4	769.16
2548	1	1	หญิง	ครู	2	1	35	5	769.16
2544	2	0	ชาย	แพทย์	1	1	42	3	0.00
2545	2	0	ชาย	แพทย์	1	1	43	4	0.00
2546	2	1	ชาย	แพทย์	1	1	44	5	20826.00
2547	2	2	ชาย	แพทย์	1	1	45	6	27180.00
2548	2	2	ชาย	แพทย์	1	1	46	7	27180.00
2544	3	0	ชาย	นักข่าว	5	1	68	2	0.00
2545	3	3	ชาย	นักข่าว	5	1	69	3	18242.89
2546	3	3	ชาย	นักข่าว	5	1	70	4	18242.89
2547	3	4	ชาย	นักข่าว	5	1	71	5	18242.89
2548	3	7	ชาย	นักข่าว	5	1	72	6	31432.89
2544	4	0	ชาย	ครู	2	1	45	1	0.00
2545	4	0	ชาย	ครู	2	1	46	2	0.00
2546	4	0	ชาย	ครู	2	1	47	3	0.00
2547	4	0	ชาย	ครู	2	1	48	4	0.00
2548	4	0	ชาย	ครู	2	1	49	5	0.00

2. จัดเตรียมข้อมูล คือ ข้อมูลการประกันภัยรถยนต์ในช่วงปี 2544-2548 ที่มีการเรียกค่าสินไหมทดแทนอย่างน้อยหนึ่งครั้งในช่วงเวลาทำประกันภัย 5 ปี มีจำนวน 3,635 กรมธรรม์ สำหรับพัฒนาตัวแบบภาพรวมโดยเฉลี่ยของประชากรด้วยวิธี GEE และตัวแบบเฉพาะของหน่วยศึกษาด้วยวิธี GLMMs โดยมีตัวแปรและกำหนดรหัสดังแสดงในตารางที่ 5

ตารางที่ 5 ตัวแปรที่ใช้ในการศึกษาและการกำหนดรหัส

ตัวแปร	คำอธิบาย	การลงรหัส
ตัวแปรตาม		
Y	จำนวนการเรียกค่าสินไหมทดแทนในแต่ละปี	
ตัวแปรอิสระ		
	ตัวแปรอิสระที่เป็นตัวแปรเชิงปริมาณ ได้แก่	
MONEY	ค่าสินไหมทดแทน (บาท/ปี)	
TFMONEY	ค่าสินไหมทดแทนที่มีการแปลงข้อมูลด้วยวิธี Box-Cox	
CAR-AGE	อายุรถยนต์ (ปี)	
AGE	อายุผู้ทำประกันภัยรถยนต์ (ปี)	
	ตัวแปรอิสระที่เป็นตัวแปรเชิงคุณภาพ ได้แก่	
GENDER	เพศ	1 = เพศชาย 0 = เพศหญิง
OCCUP	อาชีพผู้ทำประกันภัยรถยนต์	1 = อาชีพที่มีความเสี่ยงมาก 0 = อาชีพที่มีความเสี่ยงน้อย
DA1	อายุของผู้ทำประกันภัยรถยนต์	1 = อายุ 18-24 ปี 0 = อายุเกิน 50 ขึ้นไป
DA2	อายุของผู้ทำประกันภัยรถยนต์	1 = อายุ 25-35 ปี 0 = อายุเกิน 50 ขึ้นไป
DA3	อายุของผู้ทำประกันภัยรถยนต์	1 = อายุ 36-50 ปี 0 = อายุเกิน 50 ขึ้นไป
DCC	ขนาดเครื่องยนต์	1 = ขนาดเครื่องยนต์น้อยกว่าเท่ากับ 2000 cc. 0 = ขนาดเครื่องยนต์มากกว่า 2000 cc.
DGC1	กลุ่มรถยนต์	1 = รถยนต์กลุ่มที่ 1 0 = รถยนต์กลุ่มที่ 5
DGC2	กลุ่มรถยนต์	1 = รถยนต์กลุ่มที่ 2 0 = รถยนต์กลุ่มที่ 5
DGC3	กลุ่มรถยนต์	1 = รถยนต์กลุ่มที่ 3 0 = รถยนต์กลุ่มที่ 5
DGC4	กลุ่มรถยนต์	1 = รถยนต์กลุ่มที่ 4 0 = รถยนต์กลุ่มที่ 5

3. ศึกษาลักษณะข้อมูลที่จะนำมาใช้ในการวิเคราะห์ด้วยสถิติเชิงพรรณนาคำร้อยละ ความถี่ ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน และมัธยฐาน รวมทั้งตรวจสอบความสัมพันธ์ระหว่างตัวแปรที่ใช้ในการศึกษาด้วยการหาค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนระหว่างข้อมูลกลุ่มแบบอันดับด้วยกัน และจัดระดับความสัมพันธ์ระหว่างข้อมูลกลุ่มที่แทนลักษณะกับข้อมูลกลุ่มแบบอันดับด้วยค่าสถิติ Phi และ Cramer's V (กัลยา, 2549) ตรวจสอบการแจกแจงของข้อมูลเชิงปริมาณที่มีค่าต่อเนื่อง เช่น ค่าสินไหมทดแทนด้วย Histogram และแก้ปัญหาข้อมูลเชิงปริมาณของตัวแปรอิสระกรณีที่มีการแจกแจงแบบเบ้ขวา โดยการแปลงข้อมูลแบบ Box-Cox เพื่อให้ข้อมูลมีการกระจายแบบคงที่ (Stability Variances) (Jong and Heller, 2008) โดยอาศัยหลักการแปลงข้อมูลดังนี้

$$X_j^{(\lambda)} = \begin{cases} \log_e(X_j) & ; \lambda = 0 \\ \frac{(X_j^\lambda - 1)}{\lambda} & ; \lambda \neq 0 \end{cases}$$

เมื่อ X_j คือ ตัวแปรอิสระตัวที่ j ; $j = 1, 2, \dots, k$ และ λ คือ ค่าพารามิเตอร์การแปลง ตัวแปรค่าสินไหมทดแทนมีบางค่าเป็นศูนย์ทำให้ไม่สามารถทำการแปลงข้อมูลได้หากกำหนด $\lambda = 0$ จึงได้ทดลองเปลี่ยนค่า λ ที่เหมาะสมในช่วง $[0.1, 0.4]$ และเลือกค่า λ ที่ทำให้ค่า Deviance มีค่าต่ำสุด ซึ่งในการศึกษานี้ได้ทำการทดลองจนได้ค่าการแปลงข้อมูลที่เหมาะสมคือ $\lambda = 0.1$

4. การพัฒนาตัวแบบจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์แบ่งเป็น 2 ส่วน ดังนี้

4.1 สร้างตัวแบบจำนวนการเรียกค่าสินไหมทดแทนในระยะเวลาติดตาม 5 ปี ใช้วิธี GEE เมื่อสมมติให้โครงสร้างความสัมพันธ์ของค่าสังเกตภายในหน่วยศึกษามีรูปแบบ AR(1) และ CS ตามลำดับ ส่วนการวิเคราะห์ด้วยวิธี GLMMs ได้สมมติให้โครงสร้างความแปรปรวนร่วมของตัวแปรตามมีรูปแบบ AR(1) และ CS ตามลำดับ การพัฒนาตัวแบบด้วยวิธี GEE และวิธี GLMMs ของข้อมูลจำนวนการเรียกค่าสินไหมทดแทนที่มีการเก็บข้อมูลซ้ำในระยะเวลาติดตาม 5 ปี อาศัยหลักการของ Data Splitting หรือ Cross-Validation (Neter *et al.*, 2004) แบ่งเป็น 4 ขั้นตอน คือ

4.1.1 แบ่งขนาดตัวอย่างในข้อมูลทั้งหมดเป็น 2 ชุดย่อย คือ ชุดข้อมูลฝึกสอนหรือบางครั้งเรียกเป็นชุดข้อมูลการสร้างตัวแบบและประมาณค่าพารามิเตอร์ และชุดข้อมูลทดสอบหรือเรียกเป็นชุดข้อมูลสำหรับทำนาย โดยใช้วิธีการสุ่มตัวอย่างอย่างง่าย (Simple Random Sampling)

สุ่มขนาดตัวอย่าง 70 % ของข้อมูลทั้งหมดเป็นชุดข้อมูลฝึกสอน และที่เหลือ 30 % เป็นชุดข้อมูลทดสอบ ได้ผลดังตารางที่ 6

ตารางที่ 6 จำนวนขนาดตัวอย่างในแต่ละชุดข้อมูล

ชุดข้อมูลทั้งหมด	ข้อมูลฝึกสอน (Train Set)	ข้อมูลทดสอบ (Test Set)
3,635 กรมธรรม์	2,545 กรมธรรม์	1,090 กรมธรรม์

4.1.2 นำข้อมูลฝึกสอน และข้อมูลทดสอบมาวิเคราะห์ข้อมูลเบื้องต้นด้วยสถิติเชิงพรรณนา เช่น จำนวน ร้อยละ ค่าต่ำสุด ค่าสูงสุด ค่าเฉลี่ย ค่ามัธยฐาน ค่าเบี่ยงเบนมาตรฐาน และการแจกแจงของข้อมูลในตัวแปรต่าง ๆ เช่นเกี่ยวกับการตรวจสอบลักษณะข้อมูลในชุดข้อมูลทั้งหมด เพื่อตรวจสอบว่าข้อมูลที่สุ่มมาศึกษาทั้งชุดฝึกสอนและชุดทดสอบมีลักษณะการแจกแจงเช่นเดียวกับข้อมูลทั้งหมดหรือไม่

4.1.3 พัฒนาตัวแบบทางสถิติสำหรับจำนวนการเรียกค่าสินไหมทดแทนจากชุดข้อมูลฝึกสอน เริ่มจากการวิเคราะห์ตัวแปรเดียว (Univariate Analysis) ระหว่างตัวแปรตามจำนวนการเรียกค่าสินไหมทดแทนกับตัวแปรอิสระทีละตัวด้วยตัวแบบวิธี GEE เมื่อกำหนดโครงสร้างความสัมพันธ์เป็นแบบ AR(1) และ CS ด้วยตัวทดสอบสถิติ Wald และวิเคราะห์ตัวแปรเดียวด้วยตัวแบบวิธี GLMMs เมื่อกำหนดโครงสร้างความสัมพันธ์เป็นแบบ AR(1) และ CS ด้วยตัวทดสอบสถิติ t โดยพิจารณาตัวแปรอิสระที่ให้ค่า p -value น้อยกว่า 0.20 สำหรับนำไปวิเคราะห์ในตัวแบบหลายตัวแปร (Multivariate Analysis) ต่อไป ดังนี้

ก. การพัฒนาตัวแบบด้วยวิธี GEE เมื่อกำหนดโครงสร้างความสัมพันธ์เป็นแบบ AR(1) และ CS นำตัวแปรอิสระที่ผ่านการคัดกรองด้วยการวิเคราะห์ตัวแปรเดียวมาสร้างตัวแบบสำหรับจำนวนการเรียกค่าสินไหมทดแทน ผ่านกระบวนการคัดเลือกตัวแปรอิสระด้วยวิธีขั้นตอนก้าวหน้า (Forward Stepwise Method) โดยการกำหนด Probability of stepwise entry (P_E) = 0.15 และ Probability of stepwise removal (P_R) = 0.20 (Hosmer and Lemeshow, 2000) จากค่าสถิติ Generalized Score Statistics ดังนี้

1) ขั้นแรกนำตัวแปรอิสระที่ผ่านการวิเคราะห์ที่ละตัวแปรมาพิจารณา คัดเลือก ด้วยการสร้างตัวแบบของแต่ละตัวแปรอิสระ โดยการพิจารณา Generalized Score Statistics ที่ให้ค่ามากที่สุด หรือมีค่า p-value น้อยสุดและมีค่าน้อยกว่าค่า P_E ที่กำหนดไว้ให้นำตัวแปรอิสระนั้นเข้ามาในตัวแบบเป็นอันดับแรก

2) ขั้นที่สองเมื่อนำตัวแปรอิสระที่สำคัญอันดับแรกเข้ามาในตัวแบบแล้ว ทำการสร้างตัวแบบโดยเพิ่มตัวแปรอิสระที่ยังไม่ได้เข้าตัวแบบอีกทีละตัว แล้วพิจารณาว่าตัวแปรที่จะได้รับคัดเลือกให้เข้าตัวแบบเป็นตัวถัดไปเป็นตัวใด ให้พิจารณาจากค่า Generalized Score Statistics ที่มากที่สุด หรือมีค่า p-value น้อยสุดและมีค่าน้อยกว่าค่า P_E ที่กำหนดไว้ และเมื่อได้รับคัดเลือกเข้ามาแล้วจะต้องตรวจสอบว่าตัวแปรอิสระที่ถูกเลือกเข้ามาก่อนหน้านี้ทั้งหมดนั้นยังคงมีความสำคัญหรือไม่ โดยพิจารณาจากค่า Generalized Score Statistics ที่น้อยที่สุด หรือมีค่า p-value มากที่สุด และมากกว่าค่า P_R ที่กำหนดไว้หรือไม่ ถ้ามากกว่าจะถือว่าตัวแปรอิสระตัวนั้นไม่มีความสำคัญ และควรตัดตัวแปรอิสระนั้นออกจากตัวแบบ แต่ถ้ามีค่า Generalized Score Statistics ที่น้อยสุด หรือมีค่า p-value มากที่สุด แต่มีค่าน้อยกว่าค่า P_R ที่กำหนดไว้ ตัวแปรอิสระนั้นจะยังคงอยู่ในตัวแบบต่อไป กระบวนการคัดเลือกตัวแปรอิสระเข้าตัวแบบและคัดเลือกตัวแปรอิสระออกจากตัวแบบจะถูกกระทำซ้ำในลักษณะเดิมไปเรื่อย ๆ จนกระทั่งไม่มีตัวแปรอิสระใดเลยที่สามารถถูกคัดเลือกเข้ามาในตัวแบบได้อีกแล้ว และไม่มีตัวแปรใดถูกนำออกจากตัวแบบได้อีกตามเงื่อนไขของ P_E และ P_R ที่กำหนด (พิมลรัตน์, 2547)

ข. การพัฒนาตัวแบบด้วยวิธี GLMMs เมื่อกำหนดโครงสร้างความแปรปรวนร่วมเป็นแบบ AR(1) และ CS นำตัวแปรอิสระที่ผ่านการคัดกรองด้วยการวิเคราะห์ตัวแปรเดียวมาสร้างตัวแบบสำหรับจำนวนการเรียกค่าสินไหมทดแทนผ่านการคัดเลือกตัวแปรอิสระด้วยวิธีขั้นตอนก้าวหน้าแบบอัตราส่วนภาวะน่าจะเป็น (Forward Stepwise: Likelihood Ratio) โดยการกำหนด Probability of stepwise entry (P_E) = 0.15 และ Probability of stepwise removal (P_R) = 0.20 (Hosmer and Lemeshow, 2000) จากค่าสถิติ Residual Log Pseudo-Likelihood Ratio Test Statistic (G) ดังนี้ (Hosmer and Lemeshow, 2000)

1) ขั้นแรกเริ่มจากการวิเคราะห์ตัวแบบเมื่อไม่มีตัวแปรอิสระในตัวแบบ และวิเคราะห์ตัวแปรอิสระทีละตัวที่ผ่านการคัดกรองแล้ว โดยการพิจารณาค่า G (Residual Log Pseudo-Likelihood Ratio Test Statistic) เมื่อกำหนดเกณฑ์ในการคัดเลือกตัวแปรอิสระเข้าตัวแบบที่

มีค่า p-value น้อยที่สุดและมีค่าน้อยกว่าค่า P_E ที่กำหนดไว้ หรือที่มีค่า G มากที่สุด ให้นำตัวแปรอิสระนั้นเข้ามาในตัวแบบเป็นอันดับแรก

2) ขั้นที่สองเมื่อนำตัวแปรอิสระเข้ามาในตัวแบบแล้ว ทำการสร้างตัวแบบที่มีตัวแปรอิสระตัวที่เข้ามาก่อนหน้ากับตัวแปรอิสระที่จะเข้ามาเป็นตัวต่อไป โดยการพิจารณา ค่า G ที่มากที่สุด หรือมีค่า p-value น้อยที่สุดและมีค่าน้อยกว่าค่า P_E ที่กำหนดไว้ และเมื่อได้รับคัดเลือกเข้ามาแล้วจะต้องตรวจสอบว่าตัวแปรอิสระที่ถูกเลือกเข้ามาก่อนหน้าทั้งหมดนั้นยังคงมีความสำคัญหรือไม่ โดยพิจารณาจากค่า G ที่น้อยที่สุด หรือมีค่า p-value มากที่สุดและมากกว่าค่า P_R ที่กำหนดไว้หรือไม่ กระบวนการคัดเลือกตัวแปรอิสระเข้าตัวแบบและคัดเลือกตัวแปรอิสระออกจากตัวแบบจะถูกกระทำซ้ำในลักษณะเดิมไปเรื่อย ๆ จนตัวแปรอิสระเข้ามาในตัวแบบหมดแล้ว หรือไม่มีตัวแปรอิสระใดเลยที่สามารถถูกคัดเลือกเข้ามาในตัวแบบได้ตามข้อกำหนดของ P_E และ P_R (Hosmer and Lemeshow, 2000)

4.1.4 ทำการประมาณค่าพารามิเตอร์ของตัวแบบวิธี GEE เมื่อกำหนดโครงสร้างความสัมพันธ์เป็นแบบ AR(1) และ CS และตัวแบบวิธี GLMMs เมื่อกำหนดโครงสร้างความสัมพันธ์เป็นแบบ AR(1) และ CS ในชุดข้อมูลฝึกสอน และทำการประมาณค่าพารามิเตอร์ของตัวแบบวิธี GEE เมื่อกำหนดโครงสร้างความสัมพันธ์เป็นแบบ AR(1) และ CS และตัวแบบวิธี GLMMs เมื่อกำหนดโครงสร้างความสัมพันธ์เป็นแบบ AR(1) และ CS ในชุดข้อมูลทั้งหมด โดยใช้ตัวแปรอิสระที่ได้ทำการคัดเลือกมาแล้วเช่นเดียวกันกับชุดข้อมูลฝึกสอน ซึ่งการประมาณค่าพารามิเตอร์ของตัวแบบที่ได้จากข้อมูลจะมีการทดสอบสัมประสิทธิ์การถดถอย β_j ของตัวแปรอิสระ X_j เมื่อ $j = 1, 2, \dots, p$ เป็นการทดสอบว่าตัวแปรอิสระแต่ละตัวสามารถอธิบายการเปลี่ยนแปลงของตัวแปรตาม Y ได้หรือไม่เมื่อกำหนดให้ตัวแปรอิสระตัวอื่นๆ มีค่าคงที่ และกำหนดระดับนัยสำคัญทางสถิติของการทดสอบ (α) เท่ากับ 0.05

สมมติฐานการทดสอบมีดังนี้

$$\begin{aligned} H_0: \beta_j &= 0 && ; \text{ตัวแปรอิสระ } X_j \text{ ไม่มีความสัมพันธ์กับตัวแปรตาม } Y \\ H_1: \beta_j &\neq 0 && ; \text{ตัวแปรอิสระ } X_j \text{ มีความสัมพันธ์กับตัวแปรตาม } Y \end{aligned}$$

สำหรับวิธี GEE ใช้สถิติการทดสอบที่เรียกว่า Wald-Test (Z) คือ

$$Z = \frac{b_j - \beta_j}{SE_{b_j}} \text{ สมมติฐาน } H_0 \text{ จะถูกปฏิเสธเมื่อ } |Z| > Z_{\alpha/2}$$

แสดงว่า ตัวแปรอิสระ X_j อธิบายการเปลี่ยนแปลงของตัวแปรตาม Y ได้อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05

สำหรับวิธี GLMMs ใช้สถิติทดสอบ t คือ

$$t = \frac{b_j - \beta_j}{SE_{b_j}} \text{ สมมติฐาน } H_0 \text{ จะถูกปฏิเสธเมื่อ } |t| > t_{\alpha/2}$$

แสดงว่า ตัวแปรอิสระ X_j อธิบายการเปลี่ยนแปลงของตัวแปรตาม Y ได้อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05

4.1.5 ทำนายค่าของตัวแปรตามจากตัวแบบในข้อมูลชุดฝึกสอน แล้วนำตัวแบบของข้อมูลฝึกสอนมาทำการทำนายค่าของตัวแปรตามในชุดข้อมูลทดสอบ แล้วคำนวณค่าร้อยละของการทำนายถูกต้อง (R_p^2) จากชุดข้อมูลฝึกสอน และชุดข้อมูลทดสอบ และทำนายค่าของตัวแปรตามจากตัวแบบในชุดข้อมูลทั้งหมด รวมทั้งคำนวณค่าร้อยละของการทำนายถูกต้องในชุดข้อมูลทั้งหมด แล้วนำค่า R_p^2 ของชุดข้อมูลฝึกสอน ข้อมูลทดสอบ และข้อมูลทั้งหมดมาเปรียบเทียบประสิทธิภาพความแม่นยำในการทำนาย สำหรับทำการวิเคราะห์ในข้อ 4.2 ต่อไป

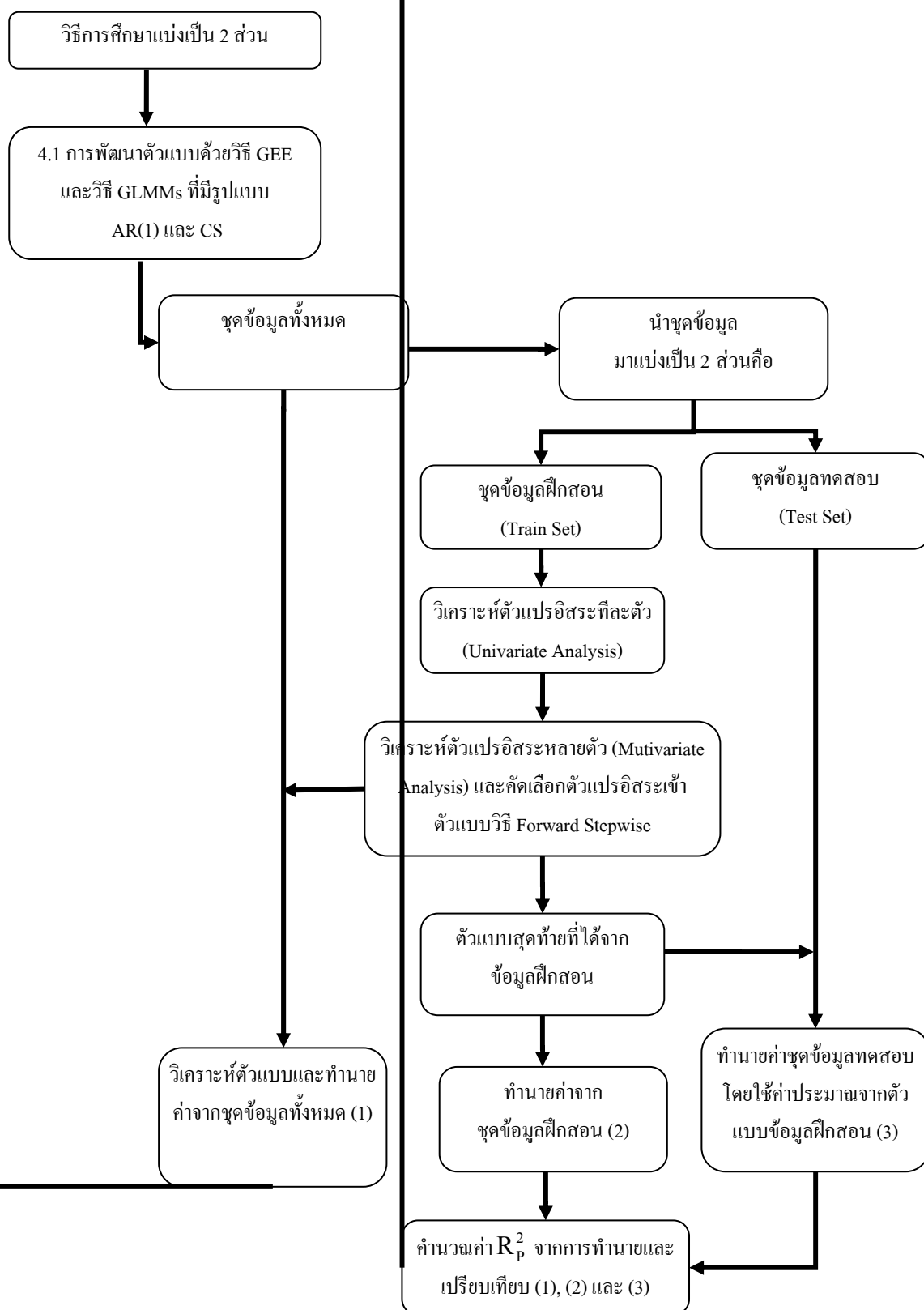
4.2 ตรวจสอบความเหมาะสมของตัวแบบวิธี GEE และวิธี GLMMs ที่ได้พัฒนาขึ้น

การตรวจสอบความเหมาะสมของตัวแบบและเปรียบเทียบผลการวิเคราะห์จากวิธี GEE เมื่อกำหนดโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) และแบบ CS สามารถพิจารณาจากค่าสถิติ Pearson Chi-Square of residual ควรมีค่าเข้าใกล้ค่าองศาแห่งความเป็นอิสระหรือค่า Pearson Chi-Square of residual / DF ที่มีค่าเข้าใกล้ 1 ส่วนวิธี GLMMs เมื่อกำหนดโครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) และแบบ CS สามารถพิจารณาได้จากค่าสถิติ Generalized Chi-Square ที่มีค่ามากที่สุด และค่า Generalized Chi-Square / DF ที่มีค่าเข้าใกล้ 1 และค่า AIC, BIC ที่มีค่าต่ำกว่าจะเป็นตัวแบบที่ดีกว่า รวมทั้งเปรียบเทียบความแม่นยำของค่าร้อยละของการทำนายถูกต้อง หรือ R_p^2 จากแต่ละตัวแบบ ตัวแบบที่ให้ค่า R_p^2 สูงกว่าจะเป็นตัวแบบที่ดีกว่า พร้อมทั้งอธิบายความหมายของสัมประสิทธิ์การถดถอยของตัวแบบซึ่งมีรายละเอียด ดังนี้

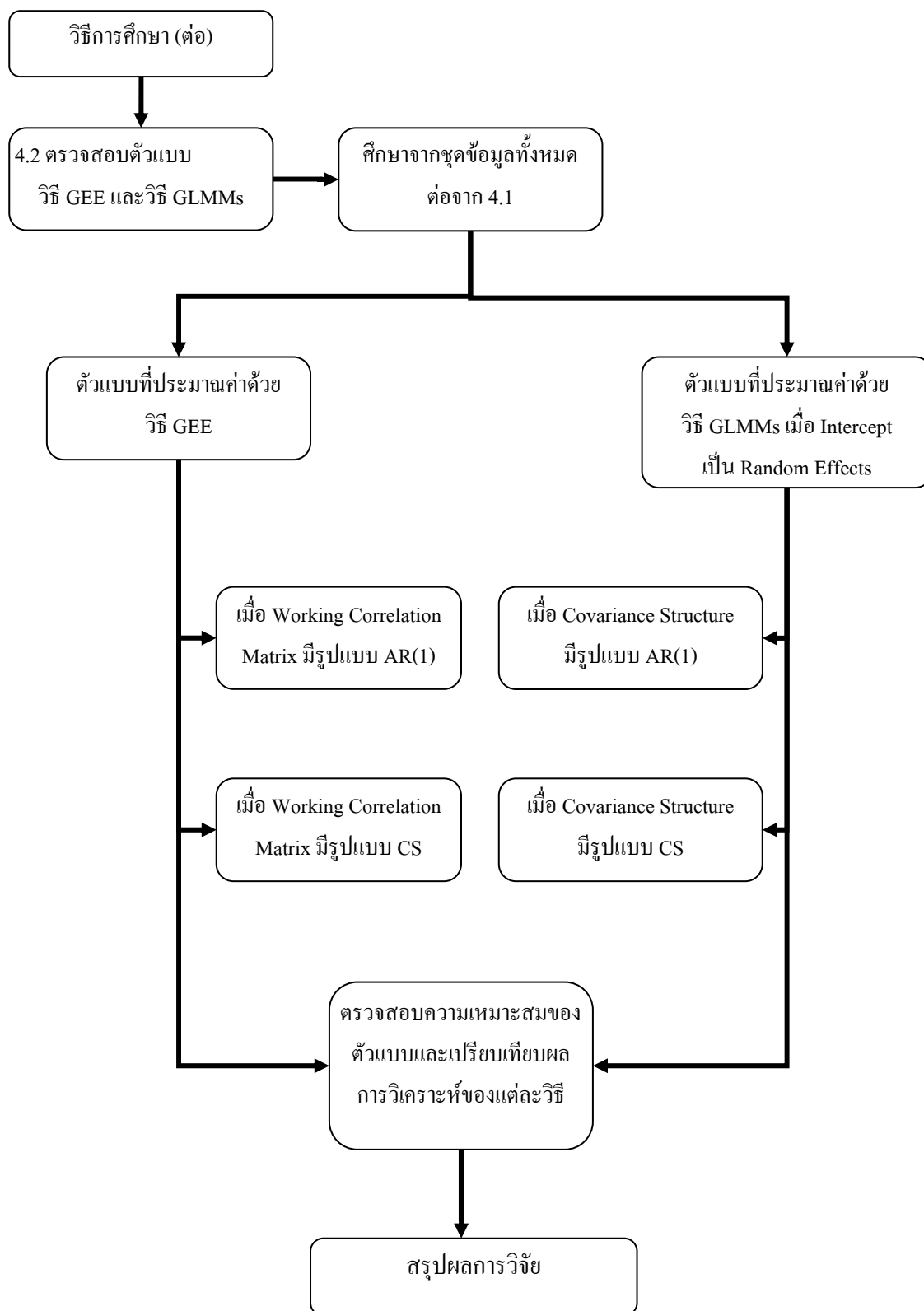
การอธิบายความหมายของตัวแปรอิสระที่เป็นข้อมูลเชิงปริมาณที่มีความสัมพันธ์กับตัวแปรตามอย่างมีนัยสำคัญทางสถิติ คือ ถ้าตัวแปรอิสระ X_j เปลี่ยนแปลงไป 1 หน่วย จะทำให้ตัวแปรตามเปลี่ยนไป $\exp(\beta_j)$ หน่วย เมื่อกำหนดให้ตัวแปรอิสระตัวอื่นๆ มีค่าคงที่

สำหรับตัวแปรอิสระที่เป็นข้อมูลเชิงคุณภาพ เช่น เพศ คือ เพศชาย = 1 และเพศหญิง = 0 ถ้าเครื่องหมายของสัมประสิทธิ์มีค่าเป็นบวก แสดงว่า เพศชายจะมีค่ามากกว่าเพศหญิงเป็น $\exp(\beta_j)$ หน่วย เมื่อกำหนดให้ตัวแปรอิสระตัวอื่นๆ มีค่าคงที่

แต่ถ้าเครื่องหมายของสัมประสิทธิ์มีค่าเป็นลบ แสดงว่า เพศชายจะมีค่าน้อยกว่าเพศหญิงเป็น $\exp(\beta_j)$ หน่วย เมื่อกำหนดให้ตัวแปรอิสระตัวอื่นๆ มีค่าคงที่ หรือ เพศหญิงจะมีค่ามากกว่าเพศชายเป็น $\exp(\beta_j)$ หน่วย เมื่อกำหนดให้ตัวแปรอิสระตัวอื่นๆ มีค่าคงที่



ภาพที่ 1 ขั้นตอนการพัฒนาตัวแบบทางสถิติ



ภาพที่ 1 (ต่อ)

ผลและวิจารณ์

ผล

การศึกษาตัวแบบทางสถิติสำหรับข้อมูลจำนวนการเรียกค่าสินไหมทดแทนที่มีการติดตามในระยะ 5 ปีของผู้ทำประกันภัยรถยนต์รายหนึ่ง ๆ โดยใช้ข้อมูลการประกันภัยรถยนต์ภาคสมัครใจประเภท 1 แบบระบุชื่อผู้ขับขี่ รหัส 110 รถยนต์นั่งส่วนบุคคล ของปี พ.ศ. 2544 ถึงปี พ.ศ. 2548 เฉพาะผู้ที่ทำการประกันภัยตั้งแต่ปี 2544 และมีจำนวนการเรียกค่าสินไหมทดแทนอย่างน้อยหนึ่งครั้งในระยะเวลาทำประกันภัย 5 ปี พบว่ามีผู้ทำประกันทั้งหมด จำนวน 3,635 กรมธรรม์ ได้ผลการวิเคราะห์ ดังนี้

1. ข้อมูลทั่วไปของผู้ประกันภัยรถยนต์จากการวิเคราะห์ด้วยสถิติเชิงพรรณนา

1.1 ลักษณะข้อมูลผู้ทำประกันภัยรถยนต์ทั้งหมด

ตารางที่ 7 ลักษณะทั่วไปของข้อมูล

ตัวแปร	ข้อมูลทั้งหมด (n=3,635) จำนวน (ร้อยละ)	ข้อมูลฝึกสอน (n=2,545) จำนวน (ร้อยละ)	ข้อมูลทดสอบ (n=1,090) จำนวน (ร้อยละ)
เพศ			
ชาย	2,057 (56.7)	1,445 (56.8)	612 (56.1)
หญิง	1,578 (43.3)	1,100 (43.2)	478 (43.9)
อาชีพ			
มีความเสี่ยงมาก	95 (2.6)	71 (2.8)	24 (2.2)
มีความเสี่ยงน้อย	3,540 (97.4)	2,474 (97.2)	1,066 (97.8)
อายุผู้ทำประกันภัยรถยนต์			
อายุ 18-24 ปี	42 (1.2)	25 (1.0)	17 (1.6)
อายุ 25-35 ปี	992 (27.2)	686 (27.0)	306 (28.1)

ตารางที่ 7 (ต่อ)

ตัวแปร	ข้อมูลทั้งหมด (n=3,635) จำนวน (ร้อยละ)	ข้อมูลฝึกสอน (n=2,545) จำนวน (ร้อยละ)	ข้อมูลทดสอบ (n=1,090) จำนวน (ร้อยละ)
อายุเกิน 50 ปี	464 (12.8)	327 (12.8)	137 (12.6)
อายุเฉลี่ย ($\bar{X} \pm SD$)	40.79 $\pm$ 8.73	40.85 $\pm$ 6.87	40.66 $\pm$ 8.88
กลุ่มรถยนต์			
กลุ่มที่ 1	65 (1.8)	50 (2.0)	15 (1.4)
กลุ่มที่ 2	663 (18.2)	473 (18.6)	190 (17.4)
กลุ่มที่ 3	1,259 (34.6)	901 (35.4)	358 (32.8)
กลุ่มที่ 4	888 (24.4)	605 (23.8)	283 (26.0)
กลุ่มที่ 5	760 (21.0)	516 (20.3)	244 (22.4)
ขนาดเครื่องยนต์			
กลุ่มที่ 1 ขนาดเครื่องยนต์ น้อยกว่าเท่ากับ 2000 cc.	2,430 (66.8)	1692 (66.5)	738 (67.7)
กลุ่มที่ 2 ขนาดเครื่องยนต์ มากกว่า 2000 cc.	1,205 (33.2)	853 (33.5)	352 (32.3)

ข้อมูลที่น่ามาศึกษาจากตารางที่ 7 พบว่า ข้อมูลทั้งหมดของผู้ทำประกันภัยรถยนต์ส่วนใหญ่เป็นเพศชายมีถึงร้อยละ 56.7 นอกนั้นเป็นเพศหญิงมีร้อยละ 43.3 โดยที่ร้อยละ 97.4 ของผู้ทำประกันภัยรถยนต์ประกอบอาชีพอยู่ในกลุ่มที่มีความเสี่ยงน้อย เช่น ครู พยาบาล แพทย์ และแม่บ้าน เป็นต้น และมีเพียงร้อยละ 2.6 ที่ทำงานอยู่ในกลุ่มอาชีพที่มีความเสี่ยงมาก เช่น นักข่าว ผู้รับเหมาก่อสร้าง วิศวกร และสถาปนิก เป็นต้น ส่วนอายุผู้ทำประกันภัยรถยนต์โดยเฉลี่ยมีอายุ 40.79 ปี และมีค่าเบี่ยงเบนมาตรฐานเท่ากับ 8.73 ปี ผู้ทำประกันภัยรถยนต์ส่วนใหญ่อยู่ในกลุ่มอายุ 36-50 ปี มากที่สุดคิดเป็นร้อยละ 58.8 รองลงมาคือ กลุ่มอายุ 25-35 ปี กลุ่มอายุเกิน 50 ปีขึ้นไป และกลุ่มอายุ 18-24 ปี คิดเป็นร้อยละ 27.2, 12.8 และ 1.2 ตามลำดับ นอกจากนี้ยังพบว่ากลุ่มรถยนต์ที่มีการทำประกันภัยรถยนต์สูงสุดถึงร้อยละ 34.6 รองลงมาคือ กลุ่มที่ 4 กลุ่มที่ 5 กลุ่มที่ 2 และกลุ่มที่ 1 คิดเป็นร้อยละ 24.4 , 21.0, 18.2 และ 1.8 ตามลำดับ และพบว่ารถยนต์ที่มีขนาดเครื่องยนต์ในกลุ่มที่ 1 ที่มีขนาดเครื่องยนต์น้อยกว่าเท่ากับ 2000 cc. มีการทำประกันภัยสูงสุดถึงร้อยละ 66.8 รองลงมาเป็นรถยนต์กลุ่มที่ 2 ขนาดเครื่องยนต์มากกว่าเท่ากับ 2000 cc. คิดเป็นร้อยละ 33.2

และเมื่อทำการแบ่งข้อมูลออกเป็น 2 ส่วน คือ ข้อมูลฝึกสอน และข้อมูลทดสอบ พบว่า ข้อมูลฝึกสอนมีลักษณะการแจกแจงของข้อมูลเช่นเดียวกับข้อมูลทั้งหมด คือ ผู้ทำประกันภัยรถยนต์ส่วนใหญ่เป็นเพศชายมีถึงร้อยละ 56.8 นอกนั้นเป็นเพศหญิงมีร้อยละ 43.2 โดยที่ร้อยละ 97.2 ประกอบอาชีพอยู่ในกลุ่มอาชีพที่มีความเสี่ยงน้อย เช่น ครู พยาบาล แพทย์ และแม่บ้าน เป็นต้น และมีเพียงร้อยละ 2.8 ที่ทำงานอยู่ในกลุ่มอาชีพที่มีความเสี่ยงมาก เช่น นักข่าว ผู้รับเหมาก่อสร้าง วิศวกร และสถาปนิก เป็นต้น ส่วนอายุผู้ทำประกันภัยรถยนต์โดยเฉลี่ยมีอายุ 40.85 ปี และมีค่าเบี่ยงเบนมาตรฐานเท่ากับ 6.87 ปี ผู้ทำประกันภัยรถยนต์ส่วนใหญ่อยู่ในกลุ่มอายุ 36-50 ปี มากที่สุดคิดเป็นร้อยละ 59.2 รองลงมาคือ กลุ่มอายุ 25-35 ปี กลุ่มอายุเกิน 50 ปีขึ้นไป และกลุ่มอายุ 18-24 ปี คิดเป็นร้อยละ 27.0, 12.8 และ 1.0 ตามลำดับ นอกจากนี้ยังพบว่ากลุ่มรถยนต์กลุ่มที่ 3 ซึ่งเป็นกลุ่มรถยนต์ที่มีการทำประกันภัยรถยนต์สูงสุดถึงร้อยละ 35.4 รองลงมาคือ กลุ่มที่ 4 กลุ่มที่ 5 กลุ่มที่ 2 และกลุ่มที่ 1 คิดเป็นร้อยละ 23.8, 20.3, 18.6 และ 2.0 ตามลำดับ และปรากฏว่ารถยนต์ที่มีขนาดเครื่องยนต์ในกลุ่มที่ 1 ที่มีขนาดเครื่องยนต์น้อยกว่าเท่ากับ 2000 cc. มีการทำประกันภัยสูงสุดถึงร้อยละ 66.5 รองลงมาเป็นรถยนต์กลุ่มที่ 2 ขนาดเครื่องยนต์มากกว่า 2000 cc. คิดเป็นร้อยละ 33.5

ส่วนข้อมูลทดสอบ พบว่า ผู้ทำประกันภัยรถยนต์ส่วนใหญ่เป็นเพศชายมีถึงร้อยละ 56.1 นอกนั้นเป็นเพศหญิงมีร้อยละ 43.9 โดยที่ร้อยละ 97.8 ประกอบอาชีพอยู่ในกลุ่มอาชีพที่มีความเสี่ยงน้อย เช่น ครู พยาบาล แพทย์ และแม่บ้าน เป็นต้น และมีเพียงร้อยละ 2.2 ที่ทำงานอยู่ในกลุ่มอาชีพที่มีความเสี่ยงมาก เช่น นักข่าว ผู้รับเหมาก่อสร้าง วิศวกร และสถาปนิก เป็นต้น ส่วนอายุผู้ทำประกันภัยรถยนต์โดยเฉลี่ยมีอายุ 40.66 ปี และมีค่าเบี่ยงเบนมาตรฐานเท่ากับ 8.88 ปี ผู้ทำประกันภัยรถยนต์ส่วนใหญ่อยู่ในกลุ่มอายุ 36-50 ปี มากที่สุดคิดเป็นร้อยละ 57.8 รองลงมาคือ กลุ่มอายุ 25-35 ปี กลุ่มอายุเกิน 50 ปีขึ้นไป และกลุ่มอายุ 18-24 ปี คิดเป็นร้อยละ 28.1, 12.6 และ 1.6 ตามลำดับ และพบว่ากลุ่มรถยนต์กลุ่มที่ 3 เป็นกลุ่มรถยนต์ที่มีการทำประกันภัยรถยนต์สูงสุดถึงร้อยละ 32.8 รองลงมาคือ กลุ่มที่ 4 กลุ่มที่ 5 กลุ่มที่ 2 และกลุ่มที่ 1 คิดเป็นร้อยละ 26.0, 22.4, 17.4 และ 1.4 ตามลำดับ นอกจากนี้รถยนต์ที่มีขนาดเครื่องยนต์ในกลุ่มที่ 1 ที่มีขนาดเครื่องยนต์น้อยกว่าเท่ากับ 2000 cc. มีการทำประกันภัยสูงสุดถึงร้อยละ 67.7 รองลงมาเป็นรถยนต์กลุ่มที่ 2 ขนาดเครื่องยนต์มากกว่า 2000 cc. คิดเป็นร้อยละ 32.3

ตารางที่ 8 ลักษณะทั่วไปของข้อมูลเชิงปริมาณ

ตัวแปร		ข้อมูลทั้งหมด (n=3,635)				
		ต่ำสุด	สูงสุด	มัธยฐาน	ค่าเฉลี่ย	ค่าเบี่ยงเบน
		(Minimum)	(Maximum)	(Median)	($\bar{X}$)	มาตรฐาน (SD.)
อายุรถยนต์		0	25	4	4.20	2.74
จำนวนการเรียก	ปี 2544	0	10	0	0.82	1.17
ค่าสินไหม	ปี 2545	0	15	1	0.99	1.36
ทดแทน	ปี 2546	0	19	2	1.92	1.92
(ครั้งต่อปี)	ปี 2547	0	19	2	2.74	2.37
	ปี 2548	0	19	3	3.51	2.72
ค่าสินไหม	ปี 2544	0.00	486,804.71	0.00	7,501.45	20,928.87
ทดแทน	ปี 2545	0.00	1,662,959.49	282.00	9,152.57	33,654.87
(บาท)	ปี 2546	0.00	1,662,959.49	8,500.00	17,041.94	39,013.86
	ปี 2547	0.00	1,662,959.49	14,711.51	23,758.08	47,454.17
	ปี 2548	0.00	1,662,959.49	20,154.20	30,034.13	51,592.59

ตัวแปร		ข้อมูลฝึกสอน (n=2,545)				
		ต่ำสุด	สูงสุด	มัธยฐาน	ค่าเฉลี่ย	ค่าเบี่ยงเบน
		(Minimum)	(Maximum)	(Median)	($\bar{X}$)	มาตรฐาน (SD.)
อายุรถยนต์		0	25	4	4.22	2.79
จำนวนการเรียก	ปี 2544	0	10	0	0.84	1.19
ค่าสินไหม	ปี 2545	0	15	1	1.00	1.35
ทดแทน	ปี 2546	0	19	2	1.92	1.92
(ครั้งต่อปี)	ปี 2547	0	19	2	2.74	2.36
	ปี 2548	0	19	3	3.51	2.72
ค่าสินไหม	ปี 2544	0.00	461,929.18	0.00	7,832.09	21,440.16
ทดแทน	ปี 2545	0.00	479,913.20	299.07	9,084.23	21,420.83
(บาท)	ปี 2546	0.00	486,247.80	8,548.00	16,977.59	28,731.61
	ปี 2547	0.00	1,061,205.15	14,598.13	23,864.24	41,180.50
	ปี 2548	0.00	1,082,962.16	20,193.39	29,930.37	45,048.13

ตารางที่ 8 (ต่อ)

ตัวแปร		ข้อมูลทดสอบ (n=1,090)				
		ต่ำสุด	สูงสุด	มัธยฐาน	ค่าเฉลี่ย	ค่าเบี่ยงเบน
		(Minimum)	(Maximum)	(Median)	($\bar{X}$)	มาตรฐาน (SD.)
อายุรณต์		0	19	4	4.17	2.62
จำนวนการเรียก ค่าสินไหม ทดแทน (ครั้งต่อปี)	ปี 2544	0	7	0	0.76	1.12
	ปี 2545	0	13	1	0.99	1.37
	ปี 2546	0	15	2	1.92	1.92
	ปี 2547	0	18	2	2.74	2.38
	ปี 2548	0	18	3	3.51	2.71
ค่าสินไหม	ปี 2544	0.00	486,804.71	0.00	6,729.45	19,671.35
ทดแทน	ปี 2545	0.00	1,662,959.49	227.12	9,312.13	52,035.62
(บาท)	ปี 2546	0.00	1,662,959.49	8,413.55	17,192.18	56,131.18
	ปี 2547	0.00	1,662,959.49	14,894.94	23,510.21	59,606.16
	ปี 2548	0.00	1,662,959.49	20,066.27	30,276.39	64,355.58

จากสถิติพื้นฐานในตารางที่ 8 สำหรับข้อมูลที่ใช้ในการศึกษาทั้งหมด พบว่า อายุรณต์ที่มี การทำประกันภัยจะมีอายุการใช้งานเฉลี่ยประมาณ 4 ปี คือมีค่าเฉลี่ยเท่ากับ 4.20 ปี ค่ามัธยฐาน เท่ากับ 4 ปี และมีค่าเบี่ยงเบนมาตรฐานเท่ากับ 2.74 ปี ส่วนรณต์ที่มีอายุการใช้งานนานที่สุดคือ 25 ปี และที่มีอายุการใช้งานน้อยสุดเป็นรณต์ที่เพิ่งทำการจดทะเบียนใหม่ไม่ถึงปีคือ 0 ปี

จำนวนการเรียกค่าสินไหมทดแทนของผู้ทำประกันภัยรณต์ที่เริ่มทำประกันภัยในปี 2544 และถูกติดตามจนถึงปี 2548 พบว่ามีค่าเฉลี่ย $\pm$ ค่าเบี่ยงเบนมาตรฐานของจำนวนการเรียกค่าสินไหมเฉลี่ยเท่ากับ 0.82 ± 1.17 , 0.99 ± 1.36 , 1.92 ± 1.92 , 2.74 ± 2.37 และ 3.51 ± 2.72 ครั้ง ตามลำดับ แต่เมื่อพิจารณาจากค่ามัธยฐานพบว่าจำนวนการเรียกค่าสินไหมทดแทนในปี 2544 จะไม่มีการเรียกค่าสินไหมทดแทนคือ 0 ครั้ง สำหรับปี 2545 และปี 2548 มีค่ามัธยฐานเท่ากับ 1 และ 3 ครั้ง ตามลำดับ ส่วนปี 2546 และปี 2547 มีค่ามัธยฐานเท่ากัน คือ 2 ครั้ง อย่างไรก็ตามในปี 2544 และปี 2545 มีผู้ทำประกันภัยรณต์ที่มีการเรียกค่าสินไหมทดแทนสูงสุดถึง 10 ครั้ง และ 15 ครั้ง ตามลำดับ ส่วนปี 2546 ถึงปี 2548 มีผู้ทำประกันภัยรณต์ที่มีการเรียกค่าสินไหมทดแทนสูงสุด เท่ากันถึง 19 ครั้ง และมีค่าต่ำสุดคือ 0 ครั้งสำหรับทุกปี คือไม่มีการเรียกค่าสินไหมทดแทนเลย

ค่าสินไหมทดแทนระหว่างปี 2544 ถึงปี 2548 มีค่าเฉลี่ยเท่ากับ 7,501.45, 9,152.57, 17,041.94, 23,758.08 และ 30,034.13 บาท ตามลำดับ และมีค่าเบี่ยงเบนมาตรฐานเท่ากับ 20,928.87, 33,654.87, 39,013.86, 47,454.17 และ 51,592.59 บาท ตามลำดับ เมื่อพิจารณาค่ามัธยฐานของจำนวนการเรียกค่าสินไหมทดแทนในปี 2544 ไม่มีการเรียกค่าสินไหมทดแทนจึงไม่มีค่าเสียหายเกิดขึ้น คือ 0.00 บาท ซึ่งผู้ทำประกันภัยรถยนต์มีการเรียกค่าสินไหมทดแทนสูงสุดเท่ากับ 20,154.20 บาทในปี 2548 รองลงมาในปี 2547 ปี 2546 และปี 2545 มีการเรียกค่าสินไหมทดแทนเท่ากับ 14,711.51, 8,500.00 และ 282.00 บาท ตามลำดับ และจะ ไม่มีการเรียกค่าสินไหมทดแทนเลยจากค่ามัธยฐานในปี 2544 โดยค่าสินไหมทดแทนที่ผู้ทำประกันภัยรถยนต์ได้รับมีตั้งแต่ไม่มีการเรียกค่าสินไหมทดแทนจนถึงได้รับค่าสินไหมทดแทนสูงสุดเท่ากับ 486,804.71 บาทในปี 2544 และ 1,662,959.49 บาท ในปี 2545 ถึงปี 2548

1.2 ลักษณะข้อมูลผู้ทำประกันภัยรถยนต์จากชุดข้อมูลฝึกสอน

จากการวิเคราะห์ข้อมูลฝึกสอนด้วยสถิติเชิงพรรณนา พบว่า อายุรถยนต์ที่มีการทำประกันภัยมีอายุการใช้งานเฉลี่ย 4.22 ปี และค่าเบี่ยงเบนมาตรฐานเท่ากับ 2.79 ปี โดยมีค่ามัธยฐานของอายุการใช้งานรถยนต์เท่ากับ 4 ปี ส่วนรถยนต์ที่มีอายุการใช้งานนานที่สุดคือ 25 ปี และรถยนต์ที่มีการใช้งานน้อยสุดเป็นรถยนต์ที่เพิ่งทำการจดทะเบียนใหม่ไม่ถึงปีคือ 0 ปี ซึ่งสถิติเหล่านี้ที่ได้จากข้อมูลฝึกสอนมีค่าใกล้เคียงกับค่าสถิติที่คำนวณจากข้อมูลทั้งหมด

จำนวนการเรียกค่าสินไหมทดแทนของผู้ทำประกันภัยรถยนต์ที่เริ่มทำประกันภัยในปี 2544 และถูกติดตามจนถึงปี 2548 พบว่ามีค่าเฉลี่ย $\pm$ ค่าเบี่ยงเบนมาตรฐานของจำนวนการเรียกค่าสินไหมเฉลี่ยเท่ากับ 0.84 ± 1.19 , 1.00 ± 1.35 , 1.92 ± 1.92 , 2.74 ± 2.3 และ 3.51 ± 2.72 ครั้งตามลำดับ ซึ่งค่าเฉลี่ยในปี 2546 ถึงปี 2548 มีค่าเท่ากับข้อมูลทั้งหมด แต่เมื่อพิจารณาจากค่ามัธยฐานพบว่าจำนวนการเรียกค่าสินไหมทดแทนในปี 2544 จะไม่มีการเรียกค่าสินไหมทดแทนคือ 0 ครั้ง สำหรับปี 2545 และปี 2548 มีค่ามัธยฐานเท่ากับ 1 และ 3 ครั้ง ตามลำดับ ส่วนปี 2546 และปี 2547 มีค่ามัธยฐานเท่ากัน คือ 2 ครั้ง จะเห็นได้ว่าการวิเคราะห์ในข้อมูลฝึกสอนมีค่าเท่ากับข้อมูลทั้งหมด และมีค่าสูงสุดและค่าต่ำสุดของจำนวนการเรียกค่าสินไหมทดแทนของข้อมูลฝึกสอนเท่ากับข้อมูลทั้งหมดคือ ในปี 2544 และปี 2545 ผู้ทำประกันภัยรถยนต์มีการเรียกค่าสินไหมทดแทนสูงสุดถึง 10 ครั้ง และ 15 ครั้ง ตามลำดับ ส่วนปี 2546 ถึงปี 2548 ผู้ทำประกันภัยรถยนต์มีการเรียก

ค่าสินไหมทดแทนสูงสุดเท่ากันถึง 19 ครั้ง และมีค่าต่ำสุดคือ 0 ครั้งสำหรับทุกปี คือ ไม่มีการเรียกค่าสินไหมทดแทนเลย

ค่าสินไหมทดแทนระหว่างปี 2544 ถึงปี 2548 มีค่าเฉลี่ยเท่ากับ 7,832.09, 9,084.23, 16,977.59, 23,864.24 และ 29,930.37 บาท ตามลำดับ มีค่าเบี่ยงเบนมาตรฐานเท่ากับ 21,440.16, 21,420.83, 28,731.61, 41,180.50 และ 45,048.13 บาท ตามลำดับ เมื่อพิจารณาค่ามัธยฐานของจำนวนการเรียกค่าสินไหมทดแทนในปี 2544 ไม่มีการเรียกค่าสินไหมทดแทนจึงไม่มีค่าเสียหายเกิดขึ้น คือ 0.00 บาท ซึ่งปี 2548 ผู้ทำประกันภัยรถยนต์จะได้รับค่าสินไหมทดแทนสูงสุดเท่ากับ 20,193.39 บาท รองลงมาคือปี 2547, ปี 2546 และปี 2545 เท่ากับ 14,598.13, 8,548.00 และ 299.07 บาท ตามลำดับ โดยค่าสินไหมทดแทนในปี 2544 ถึงปี 2548 จะมีค่าต่ำสุดเท่ากันคือ 0.00 บาท หมายความว่าไม่มีการเสียค่าสินไหมทดแทนเลย และค่าสินไหมทดแทนในปี 2548 มีค่าสูงสุดคือ 1,082,962.16 บาท รองลงมาคือ ปี 2547, ปี 2546, ปี 2545 และปี 2544 มีค่าสูงสุดคือ 1,061,205.15, 486,247.80, 479,913.20 และ 461,929.18 บาท ตามลำดับ

1.3 ลักษณะข้อมูลผู้ทำประกันภัยรถยนต์จากชุดข้อมูลทดสอบ

ผลการวิเคราะห์ข้อมูลทดสอบด้วยสถิติเชิงพรรณนาพบว่า อายุรถยนต์ที่มีการทำประกันภัยมีอายุการใช้งานเฉลี่ย 4.17 ปี และค่าเบี่ยงเบนมาตรฐานเท่ากับ 2.62 ปีโดยมีค่ามัธยฐานของอายุการใช้งานรถยนต์เท่ากับ 4 ปี ส่วนรถยนต์ที่มีอายุการใช้งานนานที่สุดคือ 19 ปี และที่มีอายุการใช้งานน้อยสุดเป็นรถยนต์ที่เพิ่งทำการจดทะเบียนใหม่ไม่ถึงปีคือ 0 ปี

จำนวนการเรียกค่าสินไหมทดแทนของผู้ทำประกันภัยรถยนต์ที่เริ่มทำประกันภัยในปี 2544 และถูกติดตามจนถึงปี 2548 พบว่ามีค่าเฉลี่ย $\pm$ ค่าเบี่ยงเบนมาตรฐานของจำนวนการเรียกค่าสินไหมเฉลี่ยเท่ากับ 0.76 ± 1.12 , 0.99 ± 1.37 , 1.92 ± 1.92 , 2.74 ± 2.38 และ 3.51 ± 2.71 ครั้งตามลำดับ ซึ่งค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานมีค่าใกล้เคียงกับข้อมูลทั้งหมด แต่เมื่อพิจารณาจากค่ามัธยฐานพบว่าจำนวนการเรียกค่าสินไหมทดแทนในปี 2544 จะไม่มีการเรียกค่าสินไหมทดแทนคือ 0 ครั้ง สำหรับปี 2545 และปี 2548 มีค่ามัธยฐานเท่ากับ 1 และ 3 ครั้ง ตามลำดับ ส่วนปี 2546 และปี 2547 มีค่ามัธยฐานเท่ากัน คือ 2 ครั้ง ซึ่งมีค่าเท่ากับข้อมูลทั้งหมดและข้อมูลฝึกสอน ส่วนค่าต่ำสุดของจำนวนการเรียกค่าสินไหมทดแทนของข้อมูลทดสอบจะเท่ากับข้อมูลทั้งหมดและข้อมูลฝึกสอนคือ 0 ครั้งสำหรับทุกปี หรือไม่มีการเรียกค่าสินไหมทดแทนเลย และในปี 2544, ปี

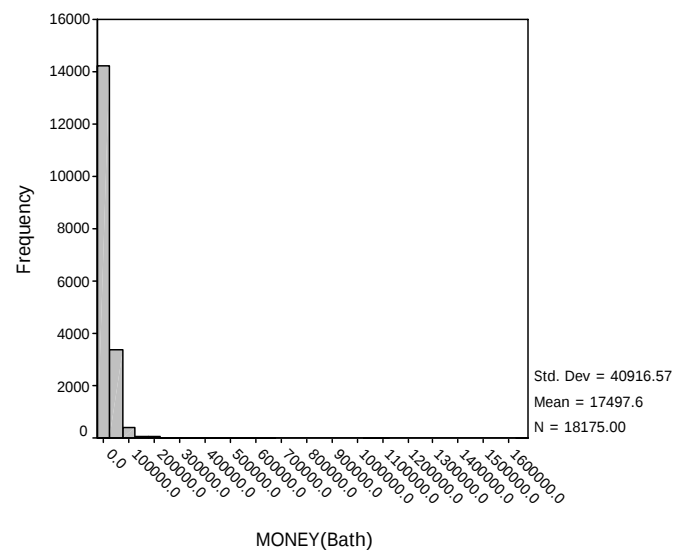
2545 และปี 2546 ผู้ทำประกันภัยรถยนต์มีการเรียกค่าสินไหมทดแทนสูงสุดเท่ากับ 7, 13 และ 15 ครั้ง ตามลำดับ ส่วนปี 2547 และปี 2548 ผู้ทำประกันภัยรถยนต์มีการเรียกค่าสินไหมทดแทนสูงสุดเท่ากันถึง 18 ครั้ง

ค่าสินไหมทดแทนระหว่างปี 2544 ถึงปี 2548 มีค่าเฉลี่ยเท่ากับ 6,729.45, 9,312.13, 17,192.18, 23,510.21 และ 30,276.39 บาท ตามลำดับ มีค่าเบี่ยงเบนมาตรฐานเท่ากับ 19,671.35, 52,035.62, 56,131.18, 59,606.16 และ 64,355.58 บาท ตามลำดับ ส่วนค่ามัธยฐานของจำนวนการเรียกค่าสินไหมทดแทนในปี 2544 ไม่มีการเรียกค่าสินไหมทดแทนจึงไม่มีค่าเสียหายเกิดขึ้น คือ 0.00 บาท ซึ่งปี 2548 ผู้ทำประกันภัยรถยนต์จะได้รับค่าสินไหมทดแทนสูงสุดเท่ากับ 20,066.27 บาท รองลงมาคือปี 2547, ปี 2546 และปี 2545 เท่ากับ 14,894.94, 8,413.55 และ 227.12 บาท ตามลำดับ โดยค่าสินไหมทดแทนในปี 2544 ถึงปี 2548 จะมีค่าต่ำสุดเท่ากันคือ 0.00 บาท หมายความว่าไม่มีการเรียกค่าสินไหมทดแทนจึงไม่มีค่าเสียหายเกิดขึ้น และค่าสินไหมทดแทนในปี 2545 ถึงปี 2548 มีค่าสูงสุดคือ 1,662,959.49 บาท รองลงมาคือ ปี 2544 มีค่าสูงสุดคือ 486,804.71 บาท ตามลำดับ

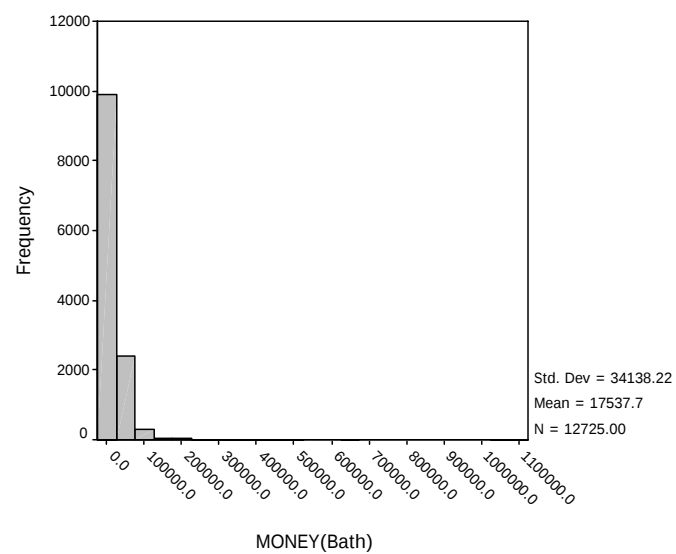
2. ผลการตรวจสอบการแจกแจงและความสัมพันธ์ของข้อมูล

2.1 ผลการตรวจสอบการแจกแจงข้อมูลค่าสินไหมทดแทน

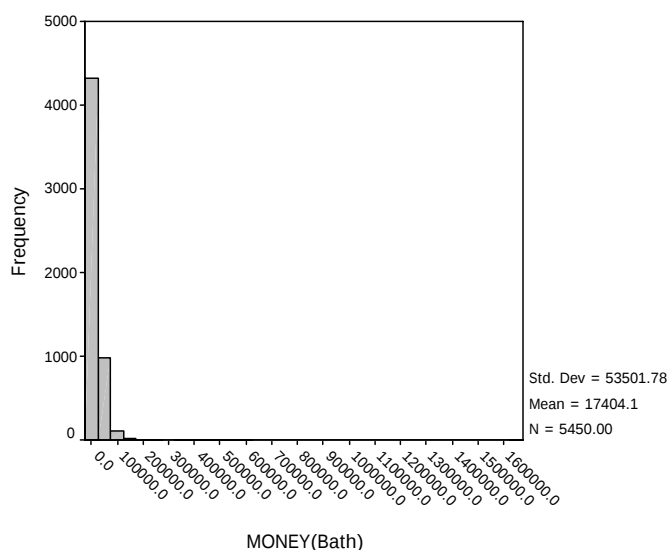
จากการศึกษาลักษณะการแจกแจงของข้อมูลเชิงปริมาณที่เป็นข้อมูลการเงินเกี่ยวกับค่าสินไหมทดแทน โดยการสร้าง Histogram จากชุดข้อมูลทั้งหมด ชุดข้อมูลฝึกสอน และชุดข้อมูลทดสอบ ดังภาพที่ 2, 3 และ 4 ตามลำดับ



ภาพที่ 2 การแจกแจงข้อมูลของค่าสินไหมทดแทนจากข้อมูลทั้งหมด



ภาพที่ 3 การแจกแจงข้อมูลของค่าสินไหมทดแทนจากข้อมูลฝึกสอน

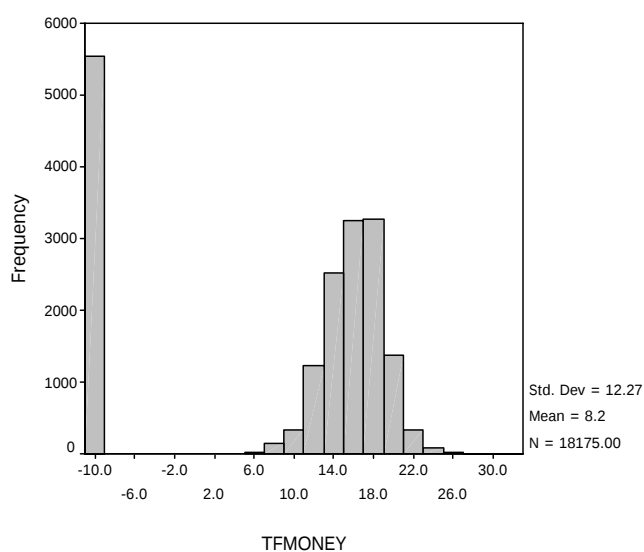


ภาพที่ 4 การแจกแจงข้อมูลของค่าสินไหมทดแทนจากข้อมูลทดสอบ

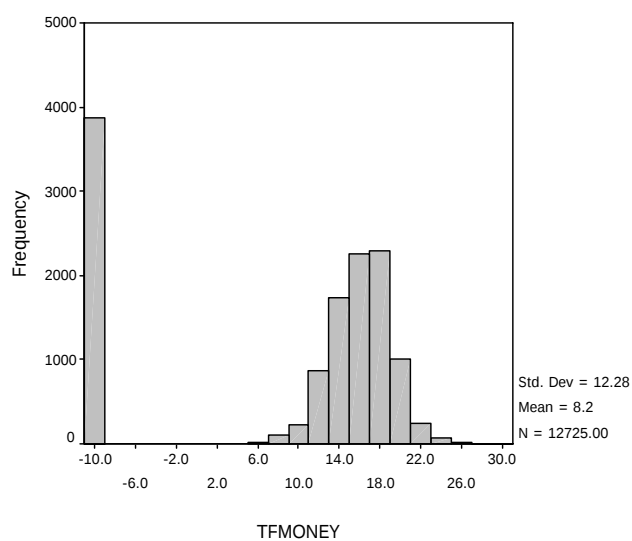
สังเกตได้ว่า ข้อมูลค่าสินไหมทดแทนมีการแจกแจงของข้อมูลเป็นลักษณะเบ้ขวาจากข้อมูลทั้ง 3 ชุด เนื่องจากเป็นข้อมูลการเงิน และมีการเก็บข้อมูลแตกต่างกันตามกลุ่มรถยนต์ทำให้เกิดความแปรปรวนของข้อมูลมาก เช่น ค่าสินไหมทดแทน ที่อยู่ในกลุ่มรถยนต์ที่ 5 เช่น รถยนต์ยี่ห้อ HONDA รุ่น City รถยนต์ยี่ห้อ Daihatsu รุ่น Mira เป็นต้น จะมีค่าเสียหายในการเรียกค่าสินไหมทดแทนต่ำ เนื่องจากเป็นกลุ่มรถยนต์ที่มีเครื่องยนต์ขนาดเล็ก ผลิตอยู่ในภูมิภาคเอเชีย ส่วนกลุ่มรถยนต์ที่ 1 เช่น รถยนต์ยี่ห้อ JAGUAR รถยนต์ยี่ห้อ FERRARI เป็นต้น จะมีค่าเสียหายในการเรียกค่าสินไหมทดแทนสูงกว่า เนื่องจากเป็นกลุ่มรถยนต์ที่มีเครื่องยนต์ขนาดใหญ่และมีการผลิตอยู่ในยุโรป ดังนั้นทำให้ข้อมูลมีความแตกต่างกันมาก และมีความแปรปรวนไม่คงที่ เมื่อนำมาทำการประมาณค่าพารามิเตอร์ในตัวแบบจะทำให้เกิดปัญหาค่าความคลาดเคลื่อนของค่าพารามิเตอร์มีค่าต่ำเกินไปจากการทดลองวิเคราะห์ข้อมูลเบื้องต้น (Jong and Heller, 2008)

และเมื่อพิจารณาลักษณะการกระจายของข้อมูลฝึกสอน และข้อมูลทดสอบที่ทำการแบ่งด้วยวิธี Cross-Validation พบว่า ตัวแปรค่าสินไหมทดแทน ที่นำมาสร้าง Histogram มีลักษณะการกระจายแบบเบ้ขวาเช่นเดียวกับลักษณะการกระจายของข้อมูลทั้งหมด ดังนั้นจึงทำการแปลงค่าตัวแปรอิสระคือ ค่าสินไหมทดแทน ตามรูปแบบการแปลงของ Box-Cox โดยได้ค่าประมาณสำหรับพารามิเตอร์ (λ) ที่ใช้ในการแปลงสำหรับตัวแปรค่าสินไหมทดแทนที่ทำให้ค่า Deviance จากตัวแบบ GEE มีค่าต่ำสุดในที่นี้ได้ $\lambda = 0.1$ เป็นค่าที่เหมาะสมสำหรับการแปลงข้อมูล เมื่อทำการแปลง

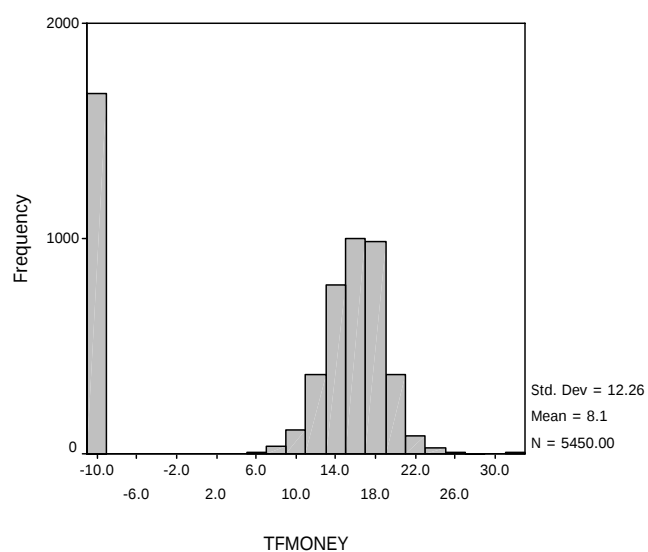
ค่าของข้อมูลของตัวแปรแล้วนำมาสร้าง Histogram ทั้งในชุดข้อมูลทั้งหมด ข้อมูลฝึกสอน และ ข้อมูลทดสอบ สังเกตได้ว่าจะมีลักษณะการกระจาย แบ่งเป็น 2 ส่วนคือ ส่วนที่ไม่มีค่าสินไหมทดแทนจะอยู่ในกลุ่มค่าศูนย์ และส่วนที่มีค่าสินไหมทดแทน โดยลักษณะการกระจายของข้อมูลจากการสร้าง Histogram จะมีลักษณะเช่นเดียวกับ Jong and Heller (2008) ที่แก้ปัญหาลักษณะข้อมูลที่มีการกระจายแบบเบ้ขวาด้วยวิธี Box-Cox แสดงได้ดังภาพที่ 5 ถึงภาพที่ 7 ดังนี้



ภาพที่ 5 การแจกแจงข้อมูลค่าสินไหมทดแทนที่แปลงข้อมูลของข้อมูลทั้งหมด



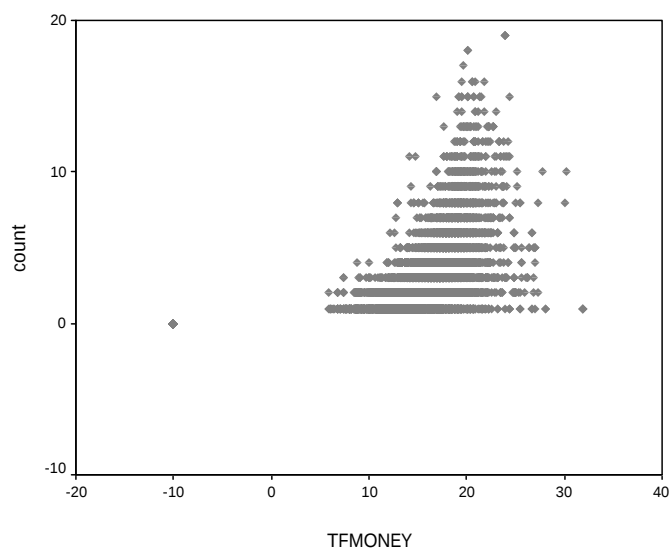
ภาพที่ 6 การแจกแจงข้อมูลค่าสินไหมทดแทนที่แปลงข้อมูลของข้อมูลฝึกสอน



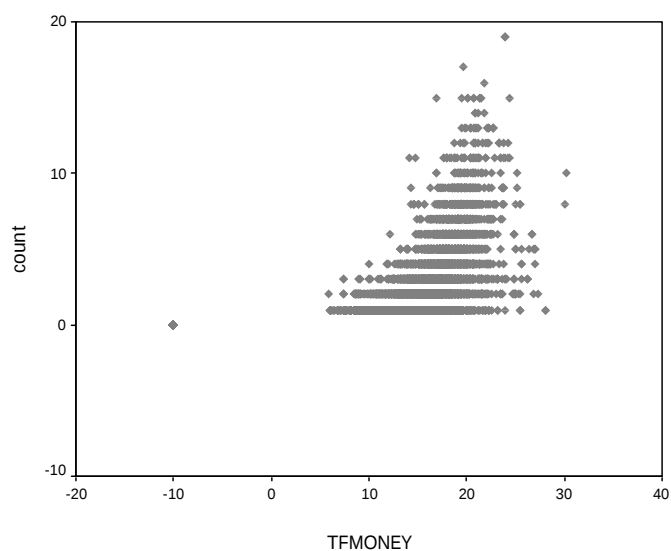
ภาพที่ 7 การแจกแจงข้อมูลค่าสินไหมทดแทนที่แปลงข้อมูลของข้อมูลทดสอบ

จากการแปลงข้อมูลค่าสินไหมทดแทนในชุดข้อมูลทั้งหมด ข้อมูลฝึกสอน และข้อมูลทดสอบจะมีลักษณะการกระจายของ Histogram แบบเดียวกัน เมื่อนำข้อมูลค่าสินไหมทดแทนที่แปลงข้อมูลมาตรวจสอบลักษณะความสัมพันธ์กับตัวแปรตามคือจำนวนค่าสินไหมทดแทนพบว่ามี ความสัมพันธ์กันในรูปเอ็กโพเนนเชียลเนื่องจากตัวแปรตามมีการแจกแจงแบบปัวส์ซองจากการสร้าง

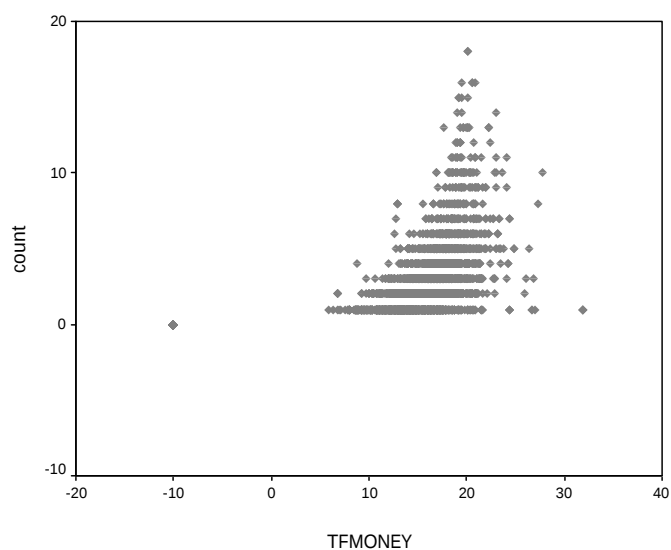
Scatter Plots ในชุดข้อมูลทั้งหมด ข้อมูลฝึกสอน และข้อมูลทดสอบ แสดงได้ดังภาพที่ 8 ถึงภาพที่ 10 ดังนี้



ภาพที่ 8 ความสัมพันธ์ระหว่างจำนวนการเรียกค่าสินไหมกับค่าสินไหมทดแทนที่แปลงข้อมูลในชุดข้อมูลทั้งหมด



ภาพที่ 9 ความสัมพันธ์ระหว่างจำนวนการเรียกค่าสินไหมกับค่าสินไหมทดแทนที่แปลงข้อมูลในชุดข้อมูลฝึกสอน



ภาพที่ 10 ความสัมพันธ์ระหว่างจำนวนการเรียกค่าสินไหมกับค่าสินไหมทดแทนที่แปลงข้อมูลในชุดข้อมูลทดสอบ

2.2 ผลการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระ

จากการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระ โดยแยกเป็นความสัมพันธ์ระหว่าง

- 1) ตัวแปรกลุ่มแบบอันดับกับตัวแปรกลุ่มแบบอันดับด้วยค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมน
- และ 2) ตัวแปรกลุ่มที่แทนลักษณะกับตัวแปรกลุ่มแบบอันดับด้วยค่า Phi และ Cramer's V แสดงผลดังตารางที่ 9 ถึงตารางที่ 10

ตารางที่ 9 ค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนระหว่างตัวแปรกลุ่มแบบอันดับกับตัวแปรกลุ่มแบบอันดับ

ตัวแปรอิสระ	ค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมน (p-value)					
	ค่าสินไหมทดแทน	อายุรถยนต์	อาชีพ	กลุ่มอายุ	กลุ่มรถยนต์	ขนาดเครื่องยนต์
ค่าสินไหมทดแทน	1	-0.018 (0.017*)	-0.008 (0.302)	-0.031 (0.000*)	-0.011 (0.126)	0.025 (0.001*)
อายุรถยนต์	-0.018 (0.017*)	1	-0.024 (0.001*)	0.094 (0.000*)	-0.138 (0.000*)	-0.032 (0.000*)
อาชีพ	-0.008 (0.302)	-0.024 (0.001*)	1	-0.016 (0.027*)	0.023 (0.002*)	-0.024 (0.001*)
กลุ่มอายุ	-0.031 (0.000*)	0.094 (0.000*)	-0.016 (0.027*)	1	0.142 (0.001*)	0.139 (0.000*)
กลุ่มรถยนต์	-0.011 (0.126)	-0.138 (0.000*)	0.023 (0.002*)	-0.142 (0.001*)	1	-0.554 (0.000*)
ขนาดเครื่องยนต์	0.025 (0.001*)	-0.032 (0.000*)	-0.024 (0.001*)	0.139 (0.000*)	-0.554 (0.000*)	1

หมายเหตุ * มีระดับนัยสำคัญทางสถิติที่ระดับ 0.05

จากตารางที่ 9 การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรกลุ่มแบบอันดับกับตัวแปรกลุ่มแบบอันดับ พบว่า ระหว่างกลุ่มค่าสินไหมทดแทนกับกลุ่มอายุรถยนต์ (แสดงรายละเอียดการจัดกลุ่มในตารางภาคผนวกที่ 2) อาชีพ กลุ่มอายุผู้ทำประกันภัย กลุ่มรถยนต์ และขนาดเครื่องยนต์ มีค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนเท่ากับ -0.018 (p-value = 0.017), -0.003 (p-value = 0.728), -0.029 (p-value = 0.000), -0.039 (p-value = 0.000) และ 0.051 (p-value = 0.000) ตามลำดับ แสดงว่ากลุ่มค่าสินไหมทดแทนมีความสัมพันธ์กับกลุ่มอายุรถยนต์ กลุ่มอายุผู้ทำประกันภัย กลุ่มรถยนต์ และขนาดเครื่องยนต์ที่ระดับต่ำอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ยกเว้นอาชีพที่ไม่มี ความสัมพันธ์กัน ทางด้านกลุ่มอายุรถยนต์กับอาชีพ กลุ่มอายุผู้ทำประกันภัย กลุ่มรถยนต์ และขนาดเครื่องยนต์มีค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนเท่ากับ -0.024 (p-value = 0.001), 0.094 (p-value = 0.000), -0.138 (p-value = 0.000) และ -0.032 (p-value = 0.000) แสดงว่ากลุ่มอายุรถยนต์มีความสัมพันธ์กับอาชีพ กลุ่มอายุผู้ทำประกันภัย กลุ่มรถยนต์ และขนาดเครื่องยนต์ที่ระดับต่ำอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ด้านอาชีพกับกลุ่มอายุผู้ทำประกันภัย กลุ่มรถยนต์ และขนาดเครื่องยนต์ มีค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนเท่ากับ -0.016 (p-value = 0.027), 0.023 (p-value = 0.002) และ -0.024 (p-value = 0.001) ตามลำดับ แสดงว่าอาชีพมีความสัมพันธ์กับกลุ่มอายุผู้ทำประกันภัย กลุ่มรถยนต์ และขนาดเครื่องยนต์ที่ระดับต่ำอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ด้านอายุกับกลุ่มรถยนต์ และขนาดเครื่องยนต์ มีค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนเท่ากับ 0.142 (p-value = 0.001) และ 0.139 (p-value = 0.000) แสดงว่า อายุมีความสัมพันธ์กับกลุ่มรถยนต์ และขนาดเครื่องยนต์ที่ระดับต่ำอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ส่วน กลุ่มรถยนต์กับขนาดเครื่องยนต์มีค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนเท่ากับ -0.554 (p-value = 0.000) แสดงว่า กลุ่มรถยนต์ มีความสัมพันธ์กับขนาดเครื่องยนต์ที่ระดับปานกลางอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05

ตารางที่ 10 ค่า Phi และ Cramer's V ระหว่างตัวแปรกลุ่มที่แทนลักษณะกับตัวแปรกลุ่มแบบอันดับ

ตัวแปรอิสระ	เพศ			
	Phi	p-value	Cramer's V	p-value
กลุ่มค่าสินไหมทดแทน	0.014	0.190	0.014	0.190
กลุ่มอายุรถยนต์	0.021	0.016*	0.021	0.016*
อาชีพ	0.022	0.003*	0.022	0.003*
กลุ่มอายุ	0.122	0.000*	0.122	0.000*
กลุ่มรถยนต์	0.140	0.000*	0.140	0.000*

ตารางที่ 10 (ต่อ)

ตัวแปรอิสระ	เพศ			
	Phi	p-value	Cramer's V	p-value
ขนาดเครื่องยนต์	0.129	0.000*	0.129	0.000*

หมายเหตุ * มีระดับนัยสำคัญทางสถิติที่ระดับ 0.05

จากตารางที่ 10 การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรกลุ่มที่แทนลักษณะกับตัวแปรกลุ่มแบบอันดับ พบว่า ระหว่างเพศกับกลุ่มอายุรถยนต์ อาชีพ กลุ่มอายุผู้ทำประกันภัย กลุ่มรถยนต์ และขนาดเครื่องยนต์ มีค่าสถิติ Phi และ Cramer's เท่ากับ 0.014 (p-value = 0.016), 0.022 (p-value = 0.003), 0.122 (p-value = 0.000), 0.140 (p-value = 0.000) และ 0.129 (p-value = 0.000) แสดงว่า เพศมีความสัมพันธ์กับกลุ่มอายุรถยนต์ อาชีพ กลุ่มอายุผู้ทำประกันภัย กลุ่มรถยนต์ และขนาดเครื่องยนต์ในระดับต่ำอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ส่วนเพศกับค่าสินไหมทดแทนมีค่าสถิติ Phi และ Cramer's เท่ากับ 0.014 (p-value = 0.190) แสดงว่าเพศไม่มีความสัมพันธ์กับค่าสินไหมทดแทนที่ระดับนัยสำคัญทางสถิติที่ระดับ 0.05

ในการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระด้วยกันเอง พบว่า มีเพียงกลุ่มรถยนต์กับขนาดเครื่องยนต์มีค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนจะอยู่ในระดับปานกลาง แต่จากงานศึกษาที่ผ่านมากลุ่มรถยนต์ และขนาดเครื่องยนต์ ถูกจัดอยู่ในกลุ่มปัจจัยเดียวกันที่มีผลในการกำหนดอัตราเบี้ยประกันในการประกันภัยรถยนต์ (กันทิมา, 2545) นั่นคือตัวแปรอิสระเหล่านี้มีความสัมพันธ์กันตามหลักการวิเคราะห์ปัจจัย จึงนำตัวแปรขนาดเครื่องยนต์เท่านั้นมาวิเคราะห์ในตัวแบบ เพื่อป้องกันการเกิดปัญหาตัวแปรอิสระมีความสัมพันธ์กัน (Multicollinearity) ดังนั้นตัวแปรอิสระที่จะนำมาพิจารณาในการศึกษานี้ คือ ค่าสินไหมทดแทน กลุ่มอายุรถยนต์ อาชีพ เพศ กลุ่มอายุผู้ทำประกันภัย และขนาดเครื่องยนต์

3. ผลการวิเคราะห์ตัวแปรเดียวด้วยวิธี GEE และวิธี GLMMs

ผลจากข้อมูลฝึกสอนที่ทำการวิเคราะห์ตัวแปรอิสระทีละตัว (Univariate Analysis) ด้วยวิธี GEE เมื่อกำหนดโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) และ CS และวิเคราะห์ด้วยวิธี GLMMs เมื่อกำหนดโครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) และ CS แสดงในตารางที่ 11 และตารางที่ 12

ตารางที่ 11 ผลการวิเคราะห์ตัวแปรเดียวในตัวแบบวิธี GEE

ตัวแปรอิสระ	ตัวแบบวิธี GEE			
	AR(1)		CS	
	Estimate (S.E.)	Z [p-value]	Estimate (S.E.)	Z [p-value]
ตัวแปรเชิงปริมาณ				
ค่าสินไหมทดแทนที่แปลงค่า (TFMONEY)	0.1514 (0.0033)	46.57 [<.0001]*	0.1680 (0.0042)	40.47 [<.0001]*
อายุรถยนต์ (CAR-AGE)	0.1272 (0.0065)	19.70 [<.0001]*	0.1670 (0.0084)	19.92 [<.0001]*
ตัวแปรเชิงคุณภาพ				
อาชีพ (OCCUP)	0.3029 (0.0776)	3.90 [<.0001]*	0.3064 (0.0826)	3.71 [0.0002]*
เพศ (GENDER)	-0.0275 (0.0293)	-0.94 [0.3479]	-0.0223 (0.0314)	-0.71 [0.4770]
อายุผู้ทำประกันภัย 18-24 ปี (DA1)	0.4371 (0.1528)	2.86 [0.0042]*	0.4309 (0.1682)	2.56 [0.0104]*
อายุผู้ทำประกันภัย 25-35 ปี (DA2)	0.1275 (0.0494)	2.58 [0.0098]*	0.1362 (0.0530)	2.57 [0.0101]*
อายุผู้ทำประกันภัย 36-50 ปี (DA3)	0.0405 (0.0452)	0.90 [0.3702]	0.0364 (0.0489)	0.75 [0.4562]
ขนาดเครื่องยนต์น้อยกว่าเท่ากับ 2000 cc. (DCC)	-0.0161 (0.0312)	-0.52 [0.6054]	-0.0251 (0.0336)	-0.75 [0.4541]

หมายเหตุ * มีระดับนัยสำคัญทางสถิติที่ระดับ 0.05

ตารางที่ 12 ผลการวิเคราะห์ตัวแปรเดียวในตัวแบบวิธี GLMMs

ตัวแปรอิสระ	ตัวแบบวิธี GLMMs			
	AR(1)		CS	
	Estimate (S.E.)	t [p-value]	Estimate (S.E.)	t [p-value]
ตัวแปรเชิงปริมาณ				
ค่าสินไหมทดแทนที่แปลงค่า (TFMONEY)	0.1893 (0.0031)	60.50 [<.0001]*	0.1680 (0.0042)	40.47 [<.0001]*
อายุรถยนต์ (CAR-AGE)	0.2202 (0.0052)	42.31 [<.0001]*	0.1670 (0.0084)	19.92 [<.0001]*
ตัวแปรเชิงคุณภาพ				
อาชีพ (OCCUP)	0.3022 (0.0772)	3.92 [<.0001]*	0.3064 (0.0826)	3.71 [0.0002]*
เพศ (GENDER)	-0.0284 (0.0290)	-0.98 [0.3272]	-0.0223 (0.0314)	-0.71 [0.4770]
อายุผู้ทำประกันภัย 18-24 ปี (DA1)	0.4383 (0.1506)	2.91 [0.0036]*	0.4309 (0.1682)	2.56 [0.0105]*
อายุผู้ทำประกันภัย 25-35 ปี (DA2)	0.1259 (0.0490)	2.57 [0.0101]*	0.1362 (0.0530)	2.57 [0.0102]*
อายุผู้ทำประกันภัย 36-50 ปี (DA3)	0.0413 (0.0448)	0.92 [0.3563]	0.0364 (0.0489)	0.75 [0.4563]
ขนาดเครื่องยนต์น้อยกว่าเท่ากับ 2000 cc. (DCC)	-0.0157 (0.03117)	-0.50 [0.6138]	-0.0249 (0.0327)	-0.73 [0.4550]

หมายเหตุ * มีระดับนัยสำคัญทางสถิติที่ระดับ 0.05

จากตารางที่ 11 และตารางที่ 12 เมื่อนำตัวแปรอิสระมาตรวจสอบโดยการวิเคราะห์ทีละตัวแปร พบว่า ในตัวแบบวิธี GEE เมื่อกำหนดโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) และ CS และตัวแบบวิธี GLMMs เมื่อกำหนดโครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) และ CS มีตัวแปร 4 ตัวที่ได้จากการทดสอบจากตัวสถิติทดสอบ Wald และสถิติ t ตามลำดับ เช่นเดียวกัน คือ ค่าสินไหมทดแทนที่แปลงค่า อายุรถยนต์ อาชีพ และกลุ่มอายุผู้ทำประกันภัยรถยนต์ ให้ค่า p-value น้อยกว่าระดับนัยสำคัญทางสถิติที่ระดับ 0.05 และตัวแปรเหล่านี้ได้ถูกคัดเลือกเป็นตัวแปรนำเข้าไปในตัวแบบการวิเคราะห์พหุตัวแปร (Multivariate Analysis) ตามเงื่อนไขการคัดกรองตัวแปรที่ระดับ $P_R = 0.20$ ต่อไป ยกเว้นตัวแปรเพศ และขนาดเครื่องยนต์ที่ไม่ผ่านการคัดกรองตามเงื่อนไข ในการศึกษานี้จะนำตัวแปรเพศเข้ามาคัดเลือกในตัวแบบด้วยเนื่องจากข้อมูลที่ใช้ในการศึกษาเป็นข้อมูลการประกันภัยรถยนต์ภาคสมัครใจ ประเภท1 แบบระบุชื่อผู้ขับขี่ จึงควรมนำตัวแปรเพศเข้ามาพิจารณา

เช่นเดียวกับการศึกษาของ Yau *et al.* (2003) ที่ปัจจัยเพศไม่ถูกคัดเลือกเข้ามาในตัวแบบ แต่ในการวิเคราะห์ภาพรวมทั้งหมดจะนำตัวแปรเพศนี้เข้ามาพิจารณาด้วย ดังนั้นตัวแปรอิสระที่นำมาพิจารณาในการคัดเลือกเข้าตัวแบบวิธี GEE เมื่อกำหนดโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) และ CS และตัวแบบวิธี GLMMs เมื่อกำหนดโครงสร้างความสัมพันธ์ร่วมของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) และ CS คือ ค่าสินไหมทดแทน อายุรถยนต์ อาชีพ และกลุ่มอายุผู้ทำประกันภัยรถยนต์ และเพศ

4. ผลการพัฒนาตัวแบบทางสถิติสำหรับจำนวนการเรียกค่าสินไหมทดแทนจากชุดข้อมูลฝึกสอน

จากตัวแปรอิสระที่ผ่านการคัดกรองในส่วนที่ 3 จะถูกนำมาพิจารณาเป็นตัวแปรนำเข้าในตัวแบบที่วิเคราะห์หลายตัวแปรโดยใช้กระบวนการคัดเลือกตัวแปรด้วยวิธีขั้นตอนก้าวหน้า

4.1 ตัวแบบวิธี GEE

การใช้ข้อมูลฝึกสอนการคัดเลือกตัวแปรอิสระเมื่อกำหนดโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) และ CS ด้วยการพิจารณาค่าสถิติ Generalized Score Statistic (χ^2) โดยกำหนดเกณฑ์ในการคัดเลือกตัวแปรอิสระเข้าตัวแบบที่ระดับนัยสำคัญทางสถิติ $P_E = 0.15$ และเกณฑ์ในการคัดเลือกตัวแปรอิสระออกจากตัวแบบที่ระดับนัยสำคัญทางสถิติ $P_R = 0.20$ พบว่าตัวแปรอิสระที่ได้รับการพิจารณาให้อยู่ในตัวแบบประกอบด้วยตัวแปรอิสระต่อไปนี้ดังแสดงในตารางที่ 13 และตารางที่ 14

1. ค่าสินไหมทดแทนที่แปลงค่า (TFMONEY)
2. อายุรถยนต์ (CAR-AGE)
3. กลุ่มอายุผู้ทำประกันภัยรถยนต์ (DA1, DA2 และ DA3)
4. อาชีพของผู้ทำประกันภัยรถยนต์ (OCCUP)
5. เพศ (GENDER)

ตารางที่ 13 ผลการคัดเลือกตัวแปรอิสระของข้อมูลฝึกสอน ด้วยค่าสถิติ Generalized Score Statistic (χ^2) ในตัวแบบ GEE เมื่อมีโครงสร้างความสัมพันธ์เป็นแบบ AR(1)

ขั้นตอนที่ 1		ขั้นตอนที่ 2		ขั้นตอนที่ 3		ขั้นตอนที่ 4		ขั้นตอนที่ 5	
ตัวแปรอิสระ	χ^2 (p-value)	ตัวแปรอิสระ	χ^2 (p-value)	ตัวแปรอิสระ	χ^2 (p-value)	ตัวแปรอิสระ	χ^2 (p-value)	ตัวแปรอิสระ	χ^2 (p-value)
TFMONEY	1672.54 (<.0001)	TFMONEY	1504.91 (<.0001)	TFMONEY	1502.19 (<.0001)	TFMONEY	1502.38 (<.0001)	TFMONEY	1504.75 (<.0001)
		CAR-AGE	198.48 (<.0001)	CAR-AGE	214.69 (<.0001)	CAR-AGE	215 (<.0001)	CAR-AGE	216.84 (<.0001)
				DA1	7.22 (0.0072)	DA1	6.92 (0.0085)	DA1	6.41 (0.0114)
				DA2	25.27 (<.0001)	DA2	26.50 (<.0001)	DA2	22.86 (<.0001)
				DA3	7.17 (0.0074)	DA3	6.93 (0.0085)	DA3	5.69 (0.0171)
						OCCUP	19.75 (<.0001)	OCCUP	20.3 (<.0001)
								GENDER	7.94 (0.0048)

ตารางที่ 14 ผลการคัดเลือกตัวแปรอิสระของข้อมูลฝึกสอน ด้วยค่าสถิติ Generalized Score Statistic (χ^2) ในตัวแบบ GEE เมื่อมีโครงสร้างความสัมพันธ์เป็นแบบ CS

ขั้นตอนที่ 1		ขั้นตอนที่ 2		ขั้นตอนที่ 3		ขั้นตอนที่ 4		ขั้นตอนที่ 5	
ตัวแปรอิสระ	χ^2 (p-value)	ตัวแปรอิสระ	χ^2 (p-value)	ตัวแปรอิสระ	χ^2 (p-value)	ตัวแปรอิสระ	χ^2 (p-value)	ตัวแปรอิสระ	χ^2 (p-value)
TFMONEY	1649.33 (<.0001)	TFMONEY	1065.29 (<.0001)	TFMONEY	1040.42 (<.0001)	TFMONEY	1046.53 (<.0001)	TFMONEY	1064.62 (<.0001)
		CAR-AGE	224.50 (<.0001)	CAR-AGE	236.41 (<.0001)	CAR-AGE	236.02 (<.0001)	CAR-AGE	237.52 (<.0001)
				DA1	6.97 (0.0083)	DA1	6.54 (0.0105)	DA1	6.09 (0.0136)
				DA2	22.19 (<.0001)	DA2	23.37 (<.0001)	DA2	19.99 (<.0001)
				DA3	6.42 (0.0113)	DA3	6.19 (0.0129)	DA3	5.00 (0.0253)
						OCCUP	18.66 (<.0001)	OCCUP	19.36 (<.0001)
								GENDER	9.45 (0.0021)

4.2 ตัวแบบวิธี GLMMs

สำหรับผลการคัดเลือกตัวแปรอิสระด้วยวิธีขั้นตอนก้าวหน้าของข้อมูลฝึกสอน โดยวิธี GLMMs เมื่อกำหนดโครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) และ CS โดยพิจารณาจากค่า G (Residual Log Pseudo-Likelihood Ratio Test Statistic) เทียบกับค่าสถิติไคสแควร์ (χ^2_{df}) เมื่อกำหนดเกณฑ์ในการคัดเลือกตัวแปรอิสระเข้าตัวแบบที่ระดับนัยสำคัญทางสถิติ $P_E = 0.15$ และเกณฑ์ในการคัดเลือกตัวแปรอิสระออกจากตัวแบบที่ระดับนัยสำคัญทางสถิติ $P_R = 0.20$ ที่แสดงในตารางที่ 15 และตารางที่ 16 คือ

1. ค่าสินไหมทดแทนที่แปลงค่า (TFMONEY)
2. อายุรถยนต์ (CAR-AGE)
3. กลุ่มอายุผู้ทำประกันภัยรถยนต์ (DA1, DA2 และ DA3)
4. อาชีพของผู้ทำประกันภัยรถยนต์ (OCCUP)
5. เพศ (GENDER)

ตารางที่ 15 ผลการคัดเลือกตัวแปรอิสระในตัวแบบวิธี GLMMs จากการพิจารณาค่า -2 Residual Log Pseudo-Likelihood เมื่อกำหนดโครงสร้างความแปรปรวนร่วมเป็นแบบ AR(1)

ขั้นตอน	ตัวแปรอิสระที่เข้า	-2 Residual Log Pseudo-Likelihood	G	df	p-value
0		29508.58			
1	TFMONEY	27184.39	2324.19	1	0.000*
2	CARAGE	25812.51	1371.88	1	0.000*
3	AGE	25779.08	33.43	2	0.000*
4	OCCUP	25751.83	27.25	1	0.000*
5	GENDER	25748.31	3.52	1	0.051*

หมายเหตุ G คือ Residual Log Pseudo-Likelihood Ratio Test Statistic

* คือ ถูกคัดเลือกเข้าตัวแบบตามเกณฑ์ที่กำหนด

ตารางที่ 16 ผลการคัดเลือกตัวแปรอิสระในตัวแบบวิธี GLMMs จากการพิจารณาค่า -2 Residual Log Pseudo-Likelihood เมื่อกำหนดโครงสร้างความแปรปรวนร่วมเป็นแบบ CS

ขั้นตอน	ตัวแปรอิสระที่เข้า	-2 Residual Log Pseudo-Likelihood	G	df	p-value
0		36633.87			
1	TFMONEY	31477.98	5155.89	1	0.000*
2	CARAGE	30290.83	1187.15	1	0.000*
3	OCCUP	30207.74	83.09	1	0.000*
4	AGE	30185.88	21.86	2	0.010*
5	GENDER	30170.70	15.18	1	0.010*

หมายเหตุ G คือ Residual Log Pseudo-Likelihood Ratio Test Statistic

* คือ ถูกคัดเลือกเข้าตัวแบบตามเกณฑ์ที่กำหนด

5. ผลการพัฒนาตัวแบบทางสถิติสำหรับจำนวนการเรียกค่าสินไหมทดแทนจากชุดข้อมูลทั้งหมด

5.1 ตัวแบบวิธี GEE

ผลการประมาณค่าพารามิเตอร์ของตัวแบบสำหรับข้อมูลจำนวนการเรียกค่าสินไหมทดแทนที่มีการเก็บข้อมูลซ้ำในระยะเวลาติดตาม 5 ปี จากตัวแบบวิธี GEE เมื่อกำหนดโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) และ CS รายละเอียดดังตารางที่ 17 และตารางที่ 18

ตารางที่ 17 ผลวิเคราะห์ตัวแบบวิธี GEE เมื่อกำหนดโครงสร้างความสัมพันธ์เป็นแบบ AR(1) เปรียบเทียบระหว่างชุดข้อมูลฝึกสอนและชุดข้อมูลทั้งหมด

ตัวแปรอิสระ	ชุดข้อมูลฝึกสอน		ชุดข้อมูลทั้งหมด	
	Estimate (S.E.)	Z [p-value]	Estimate (S.E.)	Z [p-value]
ค่าคงที่ (INTERCEPT)	-1.8413 (0.0661)	-27.87 [<.0001]*	-1.8074 (0.0530)	-34.08 [<.0001]*
ค่าสินไหมทดแทนที่แปลงค่า (TFMONEY)	0.1498 (0.0032)	47.51 [<.0001]*	0.1464 (0.0029)	50.23 [<.0001]*

ตารางที่ 17 (ต่อ)

ตัวแปรอิสระ	ชุดข้อมูลฝึกสอน		ชุดข้อมูลทั้งหมด	
	Estimate (S.E.)	Z [p-value]	Estimate (S.E.)	Z [p-value]
อายุรถยนต์ (CAR-AGE)	0.0421 (0.0028)	15.12 [<.0001]*	0.0452 (0.0025)	18.15 [<.0001]*
อายุผู้ทำประกันภัย 18-24 ปี (DA1)	0.2872 (0.0910)	3.16 [0.0016]*	0.1113 (0.0835)	1.33 [0.1828]
อายุผู้ทำประกันภัย 25-35 ปี (DA2)	0.1714 (0.0355)	4.82 [<.0001]*	0.1656 (0.0350)	4.73 [<.0001]*
อายุผู้ทำประกันภัย 36-50 ปี (DA3)	0.0779 (0.0327)	2.38 [0.0172]*	0.0787 (0.0329)	2.39 [0.0168]*
อาชีพ (OCCUP)	0.3522 (0.0559)	6.30 [<.0001]*	0.3488 (0.0495)	7.04 [<.0001]*
เพศ (GENDER)	-0.0613 (0.0217)	-2.83 [0.0047]*	-0.0366 (0.0190)	-1.92 [0.0546]
Model fit Criteria	ชุดข้อมูลฝึกสอน		ชุดข้อมูลทั้งหมด	
Pearson Chi-Square of residual	7841.364		11682.680	
Pearson Chi-Square of residual / DF	0.62		0.64	
R_p^2 (%)	ชุดข้อมูลฝึกสอน		ชุดข้อมูลทั้งหมด	
$\hat{Y}$	52.36		52.32	
$\hat{Y}_T$ ของชุดข้อมูลทดสอบ	53.58		-	

หมายเหตุ เมื่อ *	มีระดับนัยสำคัญทางสถิติที่ระดับ 0.05
S.E.	คือ ค่าความคลาดเคลื่อนมาตรฐานของค่าสัมประสิทธิ์การถดถอย
R_p^2	คือ ค่าร้อยละของการทำนายถูกต้อง
$\hat{Y}$	คือ จำนวนการเรียกค่าสินไหมทดแทนที่ทำนายจากตัวแบบ
$\hat{Y}_T$	คือ จำนวนการเรียกค่าสินไหมทดแทนที่ทำนายได้ในชุดข้อมูลทดสอบโดยใช้ตัวแบบทำนายจากชุดข้อมูลฝึกสอน
-	ไม่มีผลการวิเคราะห์

จากตารางที่ 17 พบว่า ผลการพัฒนาตัวแบบด้วยวิธี GEE สำหรับประมาณค่าพารามิเตอร์เมื่อโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) ในชุดข้อมูลฝึกสอน มีปัจจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนที่ระดับนัยสำคัญทางสถิติ 0.05 คือ ค่า

สินไหมทดแทนที่แปลงค่า (p-value < .0001) อายุรถยนต์ (p-value < .0001) กลุ่มอายุผู้ทำประกันภัยรถยนต์ (p-value = 0.0016, < .0001 และ 0.0172 ตามลำดับ) อาชีพ (p-value < .0001) และเพศ (p-value = 0.0047) เมื่อพิจารณาความเหมาะสมของตัวแบบจากค่า Pearson Chi-Square of residual /DF เท่ากับ 0.62 ซึ่งไม่เกิน 1 แสดงว่าไม่เกิดปัญหาค่าความแปรปรวนมากเกินไป (over-dispersion) สำหรับค่าร้อยละของการทำนายถูกต้องของ $\hat{Y}$ จากการประมาณค่าพารามิเตอร์ของตัวแบบเท่ากับ 52.36 และเมื่อนำการประมาณค่าพารามิเตอร์ของตัวแบบจากชุดข้อมูลฝึกสอนมาทำนายค่าของตัวแปรตามหรือจำนวนการเรียกค่าสินไหมทดแทนในชุดข้อมูลทดสอบจะมีค่าร้อยละของการทำนายถูกต้องของ $\hat{Y}$ เท่ากับ 53.58 ซึ่งมีค่าใกล้เคียงกับค่าร้อยละของการทำนายถูกต้องของ $\hat{Y}$ ในชุดข้อมูลฝึกสอน และเมื่อนำชุดข้อมูลทั้งหมด มาประมาณค่าพารามิเตอร์ของตัวแบบจะมีปัจจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนที่ระดับนัยสำคัญทางสถิติ 0.05 คือ ค่าสินไหมทดแทนที่แปลงค่า (p-value < .0001) อายุรถยนต์ (p-value < .0001) กลุ่มอายุผู้ทำประกันภัยรถยนต์ (p-value = 0.1828, < .0001 และ 0.0168 ตามลำดับ) และอาชีพ (p-value < .0001) และเมื่อพิจารณาความเหมาะสมของตัวแบบจาก Pearson Chi-Square of residual /DF เท่ากับ 0.64 ซึ่งไม่เกิน 1 แสดงว่าไม่เกิดปัญหาค่าความแปรปรวนมากเกินไป สำหรับค่าร้อยละของการทำนายถูกต้องของ $\hat{Y}$ เท่ากับ 52.32 ซึ่งมีค่าใกล้เคียงกับชุดข้อมูลฝึกสอน และค่า Pearson Chi-Square of residual /DF ของข้อมูลชุดฝึกสอนและชุดข้อมูลทั้งหมดมีค่าใกล้เคียงกัน ดังนั้นสามารถนำผลวิเคราะห์ตัวแบบจากชุดข้อมูลฝึกสอนมาใช้ในการประมาณค่าพารามิเตอร์ของตัวแบบด้วยวิธี GEE ของข้อมูลทั้งหมดเมื่อกำหนดโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) ได้

ตารางที่ 18 ผลวิเคราะห์ตัวแบบวิธี GEE เมื่อกำหนดโครงสร้างความสัมพันธ์เป็นแบบ CS
เปรียบเทียบระหว่างชุดข้อมูลฝึกสอนและชุดข้อมูลทั้งหมด

ตัวแปรอิสระ	ชุดข้อมูลฝึกสอน		ชุดข้อมูลทั้งหมด	
	Estimate (S.E.)	Z [p-value]	Estimate (S.E.)	Z [p-value]
ค่าคงที่ (INTERCEPT)	-2.0445 (0.0780)	-26.20 [<.0001]*	-2.0178 (0.0612)	-32.97 [<.0001]*
ค่าสินไหมทดแทนที่แปลงค่า (TFMONEY)	0.1611 (0.0040)	40.70 [<.0001]*	0.1574 (0.0037)	42.97 [<.0001]*
อายุรถยนต์ (CAR-AGE)	0.0485 (0.0029)	16.52 [<.0001]*	0.0527 (0.0027)	19.20 [<.0001]*
อายุผู้ทำประกันภัย 18-24 ปี (DA1)	0.3052 (0.0980)	3.12 [0.0018]*	0.1435 (0.0909)	1.58 [0.1143]

ตารางที่ 18 (ต่อ)

ตัวแปรอิสระ	ชุดข้อมูลฝึกสอน		ชุดข้อมูลทั้งหมด	
	Estimate (S.E.)	Z [p-value]	Estimate (S.E.)	Z [p-value]
อายุผู้ทำประกันภัย 25-35 ปี (DA2)	0.1755 (0.0391)	4.49 [<.0001]*	0.1737 (0.0397)	4.38 [<.0001]*
อายุผู้ทำประกันภัย 36-50 ปี (DA3)	0.0811 (0.0363)	2.23 [0.0256]*	0.0824 (0.0376)	2.19 [0.0282]*
อาชีพ (OCCUP)	0.3607 (0.0585)	6.16 [<.0001]*	0.3536 (0.0538)	6.58 [<.0001]*
เพศ (GENDER)	-0.0722 (0.0234)	-3.08 [0.0020]*	-0.0428 (0.0209)	-2.05 [0.0404]*
Model fit Criteria	ชุดข้อมูลฝึกสอน		ชุดข้อมูลทั้งหมด	
Pearson Chi-Square of residual	7648.054		11414.060	
Pearson Chi-Square of residual /DF	0.60		0.63	
R_p^2 (%)	ชุดข้อมูลฝึกสอน		ชุดข้อมูลทั้งหมด	
$\hat{Y}$	52.06		52.05	
$\hat{Y}_T$ ของชุดข้อมูลทดสอบ	53.30		-	

หมายเหตุ	เมื่อ *	มีระดับนัยสำคัญทางสถิติที่ระดับ 0.05
	S.E.	คือ ค่าความคลาดเคลื่อนมาตรฐานของค่าสัมประสิทธิ์การถดถอย
	R_p^2	คือ ค่าร้อยละของการทำนายถูกต้อง
	$\hat{Y}$	คือ จำนวนการเรียกค่าสินไหมทดแทนที่ทำนายจากตัวแบบ
	$\hat{Y}_T$	คือ จำนวนการเรียกค่าสินไหมทดแทนที่ทำนายได้ในชุดข้อมูลทดสอบโดยใช้ตัวแบบทำนายจากชุดข้อมูลฝึกสอน
	-	ไม่มีผลการวิเคราะห์

จากตารางที่ 18 พบว่า ผลการพัฒนาตัวแบบด้วยวิธี GEE สำหรับประมาณค่าพารามิเตอร์เมื่อกำหนดโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ CS ในชุดข้อมูลฝึกสอน มีปัจจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนที่ระดับนัยสำคัญทางสถิติ 0.05 เช่นเดียวกับการประมาณค่าพารามิเตอร์เมื่อกำหนดรูปแบบความสัมพันธ์ของจำนวนการเรียกค่า

สินไหมทดแทนเป็นแบบ AR(1) คือ ค่าสินไหมทดแทนที่แปลงค่า ($p\text{-value} < .0001$) อายุรถยนต์ ($p\text{-value} < .0001$) กลุ่มอายุผู้ทำประกันภัยรถยนต์ ($p\text{-value} = 0.0018, < .0001$ และ 0.0256 ตามลำดับ) อาชีพ ($p\text{-value} < .0001$) และเพศ ($p\text{-value} = 0.0020$) สำหรับการพิจารณาความเหมาะสมของตัวแบบมีค่าสถิติ Pearson Chi-Square of residual /DF เท่ากับ 0.60 ซึ่งไม่เกิน 1 แสดงว่าไม่เกิดปัญหาที่มีความแปรปรวนมากเกินไปจริง สำหรับค่าร้อยละของการทำนายถูกต้องของ $\hat{Y}$ จากการประมาณค่าพารามิเตอร์ของตัวแบบ เท่ากับ 52.06 และเมื่อนำการประมาณค่าพารามิเตอร์ของตัวแบบจากชุดข้อมูลฝึกสอนมาทำนายค่าของตัวแปรตามหรือจำนวนการเรียกค่าสินไหมทดแทนในชุดข้อมูลทดสอบ จะมีค่าร้อยละของการทำนายถูกต้องของ $\hat{Y}$ เท่ากับ 53.30 ซึ่งมีค่าใกล้เคียงกับค่าร้อยละของการทำนายถูกต้องของ $\hat{Y}$ ในชุดข้อมูลฝึกสอน และเมื่อนำชุดข้อมูลทั้งหมดมาประมาณค่าพารามิเตอร์ของตัวแบบจะมีปัจจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนที่ระดับนัยสำคัญทางสถิติ 0.05 เช่นเดียวกับการประมาณค่าพารามิเตอร์เมื่อกำหนดโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) คือ ค่าสินไหมทดแทนที่แปลงค่า ($p\text{-value} < .0001$) อายุรถยนต์ ($p\text{-value} < .0001$) กลุ่มอายุผู้ทำประกันภัยรถยนต์ ($p\text{-value} = 0.1143, < .0001$ และ 0.0282 ตามลำดับ) อาชีพ ($p\text{-value} < .0001$) และเพศ ($p\text{-value} = 0.0404$) ส่วนการพิจารณาความเหมาะสมของตัวแบบจากค่าสถิติ Pearson Chi-Square of residual /DF เท่ากับ 0.63 ซึ่งไม่เกิน 1 เช่นเดียวกัน สำหรับค่าร้อยละของการทำนายถูกต้องของ $\hat{Y}$ เท่ากับ 52.05 ซึ่งมีค่าใกล้เคียงกับชุดข้อมูลฝึกสอน และค่า Pearson Chi-Square of residual /DF ของข้อมูลชุดฝึกสอนและชุดข้อมูลทั้งหมดมีค่าใกล้เคียงกัน ดังนั้นตัวแบบที่พัฒนามาจากชุดข้อมูลฝึกสอนสามารถนำมาใช้ในการประมาณค่าพารามิเตอร์สำหรับตัวแบบด้วยวิธี GEE ของข้อมูลทั้งหมดเมื่อกำหนดโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ CS

5.2 ตัวแบบวิธี GLMMs

ผลการประมาณค่าพารามิเตอร์ของตัวแบบสำหรับข้อมูลจำนวนการเรียกค่าสินไหมทดแทนที่มีการเก็บข้อมูลซ้ำในระยะเวลาติดตาม 5 ปี จากตัวแบบวิธี วิธี GLMMs เมื่อโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) และ CS รายละเอียดดังตารางที่ 19 และตารางที่ 20

ตารางที่ 19 ผลวิเคราะห์ด้วยแบบวิธี GLMMs เมื่อกำหนดโครงสร้างความแปรปรวนร่วม
เป็นแบบ AR(1) เปรียบเทียบระหว่างชุดข้อมูลฝึกสอนและชุดข้อมูลทั้งหมด

ตัวแปรอิสระ	ชุดข้อมูลฝึกสอน		ชุดข้อมูลทั้งหมด	
	Estimate (S.E.)	t [p-value]	Estimate (S.E.)	t [p-value]
ค่าคงที่ (INTERCEPT)	-2.7366 (0.0549)	-49.87 [<.0001]*	-2.7130 (0.0467)	-58.07 [<.0001]*
ค่าสินไหมทดแทนที่แปลงค่า (TFMONEY)	0.1821 (0.0029)	63.38 [<.0001]*	0.1805 (0.0024)	75.19 [<.0001]*
อายุรถยนต์ (CAR-AGE)	0.0838 (0.0031)	27.11 [<.0001]*	0.0861 (0.0026)	32.82 [<.0001]*
อายุผู้ทำประกันภัย 18-24 ปี (DA1)	0.3316 (0.1014)	3.27 [0.0011]*	0.1390 (0.0862)	1.61 [0.1068]
อายุผู้ทำประกันภัย 25-35 ปี (DA2)	0.1858 (0.0360)	5.16 [<.0001]*	0.1703 (0.0310)	5.49 [<.0001]*
อายุผู้ทำประกันภัย 36-50 ปี (DA3)	0.0935 (0.0335)	2.79 [0.0052]*	0.0744 (0.0288)	2.58 [0.0098]*
อาชีพ (OCCUP)	0.3190 (0.0578)	5.52 [<.0001]*	0.3105 (0.0514)	6.04 [<.0001]*
เพศ (GENDER)	-0.0650 (0.0211)	-3.08 [0.0020]*	-0.0487 (0.0179)	-2.72 [0.0065]*
Model fit Criteria	ชุดข้อมูลฝึกสอน		ชุดข้อมูลทั้งหมด	
Generalized Chi-Square	2324.84		3345.96	
Generalized Chi-Square / DF	0.18		0.18	
R_p² (%)	ชุดข้อมูลฝึกสอน		ชุดข้อมูลทั้งหมด	
$\hat{Y}_u$	75.94		76.28	
$\hat{Y}$	53.03		53.34	
$\hat{Y}_T$ ของชุดข้อมูลทดสอบ	54.75		-	

หมายเหตุ เมื่อ * มีระดับนัยสำคัญทางสถิติที่ระดับ 0.05

S.E. คือ ค่าความคลาดเคลื่อนมาตรฐานของค่าสัมประสิทธิ์การถดถอย

R_p² คือ ค่าร้อยละของการทำนายถูกต้อง

$\hat{Y}$ คือ จำนวนการเรียกค่าสินไหมทดแทนที่ทำนายจากตัวแบบ

ตารางที่ 19 (ต่อ)

$\hat{Y}_T$	คือ จำนวนการเรียกค่าสินไหมทดแทนที่ทำนายได้ในชุดข้อมูลทดสอบ โดยใช้ตัวแบบทำนายจากชุดข้อมูลฝึกสอน
$\hat{Y}_u$	คือ จำนวนการเรียกค่าสินไหมทดแทนที่ทำนายจากตัวแบบ เมื่อมีเทอมอิทธิพลค่าคงที่สุ่ม
-	ไม่มีผลการวิเคราะห์

จากตารางที่ 19 พบว่า ผลการพัฒนาตัวแบบด้วยวิธี GLMMs สำหรับประมาณค่าพารามิเตอร์ เมื่อกำหนดโครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) ในชุดข้อมูลฝึกสอน พบว่า ปัจจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนที่ระดับนัยสำคัญทางสถิติ 0.05 คือ ค่าสินไหมทดแทนที่แปลงค่า (p-value < .0001) อายุรถยนต์ (p-value < .0001) กลุ่มอายุผู้ทำประกันภัยรถยนต์ (p-value = 0.0011, < .0001 และ 0.0052 ตามลำดับ) อาชีพ (p-value < .0001) และเพศ (p-value = 0.0020) เมื่อพิจารณาความเหมาะสมของตัวแบบจากค่าสถิติ Generalized Chi-Square /DF เท่ากับ 0.18 ซึ่งไม่เกิน 1 แสดงว่าไม่มีปัญหาค่าความแปรปรวนมากเกินไปจริง สำหรับค่าร้อยละของการทำนายถูกต้องของ $\hat{Y}$ จากการประมาณค่าพารามิเตอร์ของตัวแบบเท่ากับ 53.03 และเมื่อนำการประมาณค่าพารามิเตอร์ของตัวแบบจากชุดข้อมูลฝึกสอนมาทำนายค่าของตัวแปรตามหรือจำนวนการเรียกค่าสินไหมทดแทนในชุดข้อมูลทดสอบจะมีค่าร้อยละของการทำนายถูกต้องของ $\hat{Y}$ เท่ากับ 54.75 ซึ่งมีค่าใกล้เคียงกับค่าร้อยละของการทำนายถูกต้องของ $\hat{Y}$ ในชุดข้อมูลฝึกสอน และมีค่าร้อยละของการทำนายถูกต้องของตัวแบบเมื่อมีเทอมค่าคงที่สุ่ม (Random Intercept) ของแต่ละบุคคล $R_p^2(\hat{Y}_u)$ เท่ากับ 75.94 ส่วนการนำชุดข้อมูลทั้งหมดมาประมาณค่าพารามิเตอร์ของตัวแบบพบว่า ปัจจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนที่ระดับนัยสำคัญทางสถิติ 0.05 คือ ค่าสินไหมทดแทนที่แปลงค่า (p-value < .0001) อายุรถยนต์ (p-value < .0001) กลุ่มอายุผู้ทำประกันภัยรถยนต์ (p-value = 0.1068, < .0001 และ 0.0098 ตามลำดับ) อาชีพ (p-value < .0001) และเพศ (p-value = 0.0065) เมื่อพิจารณาความเหมาะสมของตัวแบบจากค่าสถิติ Generalized Chi-Square /DF เท่ากับ 0.18 ซึ่งไม่เกิน 1 สำหรับค่าอัตราส่วนร้อยละของการทำนายถูกต้องของ $\hat{Y}$ เท่ากับ 53.34 และมีค่าอัตราส่วนร้อยละของการทำนายถูกต้องของตัวแบบเมื่อมีเทอมค่าคงที่สุ่มของแต่ละบุคคล $R_p^2(\hat{Y}_u)$ เท่ากับ 76.28 ซึ่งมีค่าใกล้เคียงกับชุดข้อมูลฝึกสอน และค่า Generalized Chi-Square /DF ของข้อมูลชุดฝึกสอนและชุดข้อมูลทั้งหมดมีค่าเท่ากัน ดังนั้นผลวิเคราะห์ที่พัฒนามาจากชุดข้อมูลฝึกสอนสามารถนำมาใช้ใน

การประมาณค่าพารามิเตอร์สำหรับตัวแบบวิธี GLMMs ของข้อมูลทั้งหมดได้ เมื่อกำหนดโครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1)

ตารางที่ 20 ผลวิเคราะห์ตัวแบบวิธี GLMMs เมื่อกำหนดโครงสร้างความแปรปรวนร่วมเป็นแบบ CS เปรียบเทียบระหว่างชุดข้อมูลฝึกสอนและชุดข้อมูลทั้งหมด

ตัวแปรอิสระ	ชุดข้อมูลฝึกสอน		ชุดข้อมูลทั้งหมด	
	Estimate (S.E.)	t [p-value]	Estimate (S.E.)	t [p-value]
ค่าคงที่ (INTERCEPT)	-2.9705 (0.0538)	-55.23 [<.0001]*	-2.9525 (0.0458)	-64.54 [<.0001]*
ค่าสินไหมทดแทนที่แปลงค่า (TFMONEY)	0.1917 (0.0030)	64.08 [<.0001]*	0.1904 (0.0025)	76.69 [<.0001]*
อายุรถยนต์ (CAR-AGE)	0.09128 (0.0035)	26.12 [<.0001]*	0.0938 (0.0030)	31.55 [<.0001]*
อายุผู้ทำประกันภัย 18-24 ปี (DA1)	0.3169 (0.0600)	5.28 [<.0001]*	0.1445 (0.0886)	1.63 [0.1028]
อายุผู้ทำประกันภัย 25-35 ปี (DA2)	0.3319 (0.1066)	3.11 [0.0019]*	0.1755 (0.0323)	5.43 [<.0001]*
อายุผู้ทำประกันภัย 36-50 ปี (DA3)	0.1902 (0.0375)	5.07 [<.0001]*	0.0751 (0.0301)	2.50 [0.0126]*
อาชีพ (OCCUP)	0.0940 (0.0349)	2.70 [0.0070]*	0.3054 (0.0537)	5.69 [<.0001]*
เพศ (GENDER)	-0.0669 (0.0219)	-3.06 [0.0022]*	-0.0517 (0.0186)	-2.79 [0.0053]*
Model fit Criteria	ชุดข้อมูลฝึกสอน		ชุดข้อมูลทั้งหมด	
Generalized Chi-Square	1882.99		2677.77	
Generalized Chi-Square / DF	0.15		0.15	
R_p² (%)	ชุดข้อมูลฝึกสอน		ชุดข้อมูลทั้งหมด	
$\hat{Y}_u$	78.81		79.36	
$\hat{Y}$	53.18		53.38	
$\hat{Y}_T$ ของชุดข้อมูลทดสอบ	54.62		-	

หมายเหตุ เมื่อ * มีระดับนัยสำคัญทางสถิติที่ระดับ 0.05
S.E. คือ ค่าความคลาดเคลื่อนมาตรฐานของค่าสัมประสิทธิ์การถดถอย

ตารางที่ 20 (ต่อ)

R_p^2	คือ ค่าร้อยละของการทำนายถูกต้อง
$\hat{Y}$	คือ จำนวนการเรียกค่าสินไหมทดแทนที่ทำนายจากตัวแบบ
$\hat{Y}_T$	คือ จำนวนการเรียกค่าสินไหมทดแทนที่ทำนายได้ในชุดข้อมูลทดสอบโดยใช้ตัวแบบทำนายจากชุดข้อมูลฝึกสอน
$\hat{Y}_u$	คือ จำนวนการเรียกค่าสินไหมทดแทนที่ทำนายจากตัวแบบเมื่อมีเทอมอิทธิพลค่าคงที่สุ่ม
-	ไม่มีผลการวิเคราะห์

จากตารางที่ 20 พบว่า ผลการพัฒนาตัวแบบด้วยวิธี GLMMs สำหรับประมาณค่าพารามิเตอร์ เมื่อกำหนดโครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ CS ในชุดข้อมูลฝึกสอน พบว่า ปัจจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนที่ระดับนัยสำคัญทางสถิติ 0.05 เช่นเดียวกับรูปแบบ AR(1) คือ ค่าสินไหมทดแทนที่แปลงค่า (p-value < .0001) อายุรถยนต์ (p-value < .0001) กลุ่มอายุผู้ทำประกันภัยรถยนต์ (p-value < .0001, 0.0019 และ < .0001 ตามลำดับ) อาชีพ (p-value = 0.0070) และเพศ (p-value = 0.0022) เมื่อพิจารณาความเหมาะสมของตัวแบบจากค่าสถิติ Generalized Chi-Square /DF เท่ากับ 0.15 ซึ่งไม่เกิน 1 แสดงว่าไม่เกิดปัญหามีค่าความแปรปรวนมากเกินไป สำหรับค่าร้อยละของการทำนายถูกต้องของ $\hat{Y}$ จากการประมาณค่าพารามิเตอร์ของตัวแบบเท่ากับ 53.18 และเมื่อนำการประมาณค่าพารามิเตอร์ของตัวแบบจากชุดข้อมูลฝึกสอนมาทำนายค่าของตัวแปรตามหรือจำนวนการเรียกค่าสินไหมทดแทนในชุดข้อมูลทดสอบจะมีค่าร้อยละของการทำนายถูกต้องของ $\hat{Y}$ เท่ากับ 54.62 ซึ่งมีค่าใกล้เคียงกับค่าร้อยละของการทำนายถูกต้องของ $\hat{Y}$ ในชุดข้อมูลฝึกสอน และมีค่าร้อยละของการทำนายถูกต้องของตัวแบบเมื่อมีเทอมค่าคงที่สุ่มของแต่ละบุคคลได้ $R_p^2(\hat{Y}_u)$ เท่ากับ 78.81 ส่วนการนำชุดข้อมูลทั้งหมดมาประมาณค่าพารามิเตอร์ของตัวแบบพบว่า ปัจจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนที่ระดับนัยสำคัญทางสถิติ 0.05 เช่นเดียวกับรูปแบบ AR(1) คือ ค่าสินไหมทดแทนที่แปลงค่า (p-value < .0001) อายุรถยนต์ (p-value < .0001) กลุ่มอายุผู้ทำประกันภัยรถยนต์ (p-value = 0.1028, < .0001 และ 0.0126 ตามลำดับ) อาชีพ (p-value < .0001) และเพศ (p-value = 0.0053) และเมื่อพิจารณาความเหมาะสมของตัวแบบจากค่าสถิติ Generalized Chi-Square /DF เท่ากับ 0.15 ซึ่งไม่เกิน 1 สำหรับค่าร้อยละของการทำนายถูกต้องของ $\hat{Y}$ เท่ากับ 53.38 และมีค่าร้อยละของการทำนายถูกต้องของตัวแบบเมื่อมีเทอมค่าคงที่สุ่มของแต่ละบุคคลได้ $R_p^2(\hat{Y}_u)$ เท่ากับ 79.36 ซึ่งมีค่า

ใกล้เคียงกับชุดข้อมูลฝึกสอน และค่า Generalized Chi-Square /DF ของข้อมูลชุดฝึกสอนและชุดข้อมูลทั้งหมดมีค่าเท่ากัน ดังนั้นผลวิเคราะห์ตัวแบบที่พัฒนามาจากชุดข้อมูลฝึกสอนสามารถนำมาใช้ในการประมาณค่าพารามิเตอร์ของตัวแบบวิธี GLMMs ของข้อมูลทั้งหมดได้ เมื่อกำหนดโครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ CS

6. ผลการวิเคราะห์ตัวแบบจำนวนการเรียกค่าสินไหมทดแทนด้วยวิธี GEE และวิธี GLMMs

การเปรียบเทียบผลการวิเคราะห์ตัวแบบสำหรับข้อมูลของจำนวนการเรียกค่าสินไหมทดแทนที่มีการเก็บข้อมูลซ้ำในระยะเวลาติดตาม 5 ปี ด้วยวิธี GEE และวิธี GLMMs เมื่อโครงสร้างเป็นแบบ AR(1) และ CS จากข้อมูลทั้งหมด แสดงผลวิเคราะห์ได้ดังนี้

6.1 ตัวแบบวิธี GEE

ตารางที่ 21 ผลวิเคราะห์ตัวแบบวิธี GEE เมื่อเปรียบเทียบโครงสร้างความสัมพันธ์เป็นแบบ AR(1) และ CS ในชุดข้อมูลทั้งหมด

ตัวแปรอิสระ	ตัวแบบวิธี GEE			
	AR(1)		CS	
Fixed-effect Parameter	Estimate (S.E.)	Z [p-value]	Estimate (S.E.)	Z [p-value]
ค่าคงที่ (INTERCEPT)	-1.8074 (0.0530)	-34.08 [<.0001]*	-2.0178 (0.0612)	-32.97 [<.0001]*
ค่าสินไหมทดแทนที่แปลงค่า (TFMONEY)	0.1464 (0.0029)	50.23 [<.0001]*	0.1574 (0.0037)	42.97 [<.0001]*
อายุรถยนต์ (CAR-AGE)	0.0452 (0.0025)	18.15 [<.0001]*	0.0527 (0.0027)	19.20 [<.0001]*
อายุผู้ทำประกันภัย 18-24 ปี (DA1)	0.1113 (0.0835)	1.33 [0.1828]	0.1435 (0.0909)	1.58 [0.1143]
อายุผู้ทำประกันภัย 25-35 ปี (DA2)	0.1656 (0.0350)	4.73 [<.0001]*	0.1737 (0.0397)	4.38 [<.0001]*
อายุผู้ทำประกันภัย 36-50 ปี (DA3)	0.0787 (0.0329)	2.39 [0.0168]*	0.0824 (0.0376)	2.19 [0.0282]*
อาชีพ (OCCUP)	0.3488 (0.0495)	7.04 [<.0001]*	0.3536 (0.0538)	6.58 [<.0001]*
เพศ (GENDER)	-0.0366 [0.0190]	-1.92 [0.0546]	-0.0428 (0.0209)	-2.05 [0.0404]*

ตารางที่ 21 (ต่อ)

Model fit Criteria	ตัวแบบวิธี GEE	
	AR(1)	CS
Pearson Chi-Square of residual	11682.680	11414.060
Pearson Chi-Square of residual / DF	0.64	0.63
R_p^2 (%) ของ $\hat{Y}$	52.32	52.05

หมายเหตุ	เมื่อ *	มีระดับนัยสำคัญทางสถิติที่ระดับ 0.05
	S.E.	คือ ค่าความคลาดเคลื่อนมาตรฐานของค่าสัมประสิทธิ์การถดถอย
	R_p^2	คือ ค่าร้อยละของการทำนายถูกต้อง
	$\hat{Y}$	คือ จำนวนการเรียกค่าสินไหมทดแทนที่ทำนายจากตัวแบบ
	-	ไม่มีผลการวิเคราะห์

จากตารางที่ 21 พบว่า การวิเคราะห์ตัวแบบด้วยวิธี GEE สำหรับข้อมูลทั้งหมดเมื่อโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) โดยการทดสอบสมมติฐานของสัมประสิทธิ์การถดถอยของตัวแปรอิสระแต่ละตัวจากสถิติทดสอบ Wald หรือ $|Z|$ พบว่าตัวแปรที่มีระดับนัยสำคัญทางสถิติ 0.05 คือ ค่าสินไหมทดแทนที่แปลงค่า อายุรถยนต์ อายุผู้ทำประกันภัย 25-35 ปี อายุผู้ทำประกันภัย 36-50 ปี และอาชีพ มีค่าเท่ากับ 50.23 (p-value < .0001), 18.15 (p-value < .0001), 4.73 (p-value < .0001), 2.39 (p-value < .0001) และ 7.04 (p-value < .0001) แสดงว่าตัวแปรอิสระเหล่านี้สามารถอธิบายการเปลี่ยนแปลงจำนวนการเรียกค่าสินไหมทดแทนได้ ดังนั้นตัวแบบของการทำนายจำนวนการเรียกค่าสินไหมทดแทนสำหรับโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) เขียนได้ดังนี้

$$\log_e y = -1.8074 + 0.1464 (\text{ค่าสินไหมทดแทนที่แปลงค่า}) + 0.0452 (\text{อายุรถยนต์}) + 0.1113 (\text{อายุผู้ทำประกันภัย 18-24 ปี}) + 0.1656 (\text{อายุผู้ทำประกันภัย 25-35 ปี}) + 0.0787 (\text{อายุผู้ทำประกันภัย 36-50 ปี}) + 0.3488 (\text{อาชีพ}) + (-0.0366) (\text{เพศ})$$

(* มีระดับนัยสำคัญทางสถิติที่ระดับ 0.05)

ส่วนการวิเคราะห์ตัวแบบวิธี GEE สำหรับข้อมูลทั้งหมดเมื่อโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ CS โดยการทดสอบสมมติฐานของสัมประสิทธิ์การถดถอยของตัวแปรอิสระแต่ละตัวจากค่าสถิติทดสอบ Wald หรือ $|Z|$ พบว่าตัวแปรที่มีนัยสำคัญทางสถิติที่ระดับ 0.05 คือ ค่าสินไหมทดแทนที่แปลงค่า อายุรถยนต์ อายุผู้ทำประกันภัย 25-35 ปี อายุผู้ทำประกันภัย 36-50 ปี อาชีพ และเพศ มีค่าเท่ากับ 42.97 (p-value < .0001), 19.20 (p-value < .0001), 4.38 (p-value < .0001), 2.19 (p-value < .0001), 6.58 (p-value < .001) และ -2.05 (p-value = 0.0404) แสดงว่าตัวแปรอิสระเหล่านี้สามารถอธิบายการเปลี่ยนแปลงจำนวนการเรียกค่าสินไหมทดแทน ดังนั้นตัวแบบของการทำนายจำนวนการเรียกค่าสินไหมทดแทนสำหรับโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ CS เขียนได้ดังนี้

$$\begin{aligned} \log_e y = & -2.0178 + 0.1574 (\text{ค่าสินไหมทดแทนที่แปลงค่า})^* + 0.0527 (\text{อายุรถยนต์})^* \\ & + 0.1435 (\text{อายุผู้ทำประกันภัย 18-24 ปี}) + 0.1737 (\text{อายุผู้ทำประกันภัย 25-35 ปี})^* \\ & + 0.0824 (\text{อายุผู้ทำประกันภัย 36-50 ปี})^* + 0.3536 (\text{อาชีพ})^* + (-0.0428) (\text{เพศ})^* \end{aligned}$$

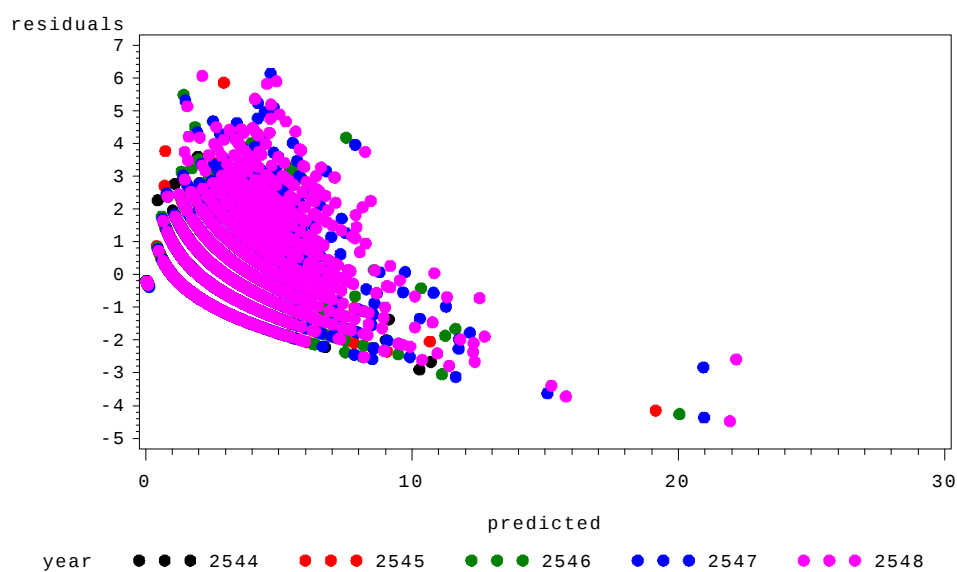
(* มีระดับนัยสำคัญทางสถิติที่ระดับ 0.05)

เมื่อพิจารณาจากค่าความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์การถดถอยแต่ละปัจจัยของตัวแบบด้วยวิธี GEE เมื่อกำหนดโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) และ CS พบว่า ค่าความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์การถดถอยของตัวแบบเมื่อโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) มีค่าต่ำกว่ารูปแบบ CS แต่มีค่าไม่ต่างกันจึงเห็นว่าควรใช้ตัวแบบที่มีโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) ไปใช้ในการทำนายจำนวนการเรียกค่าสินไหมทดแทนต่อไป

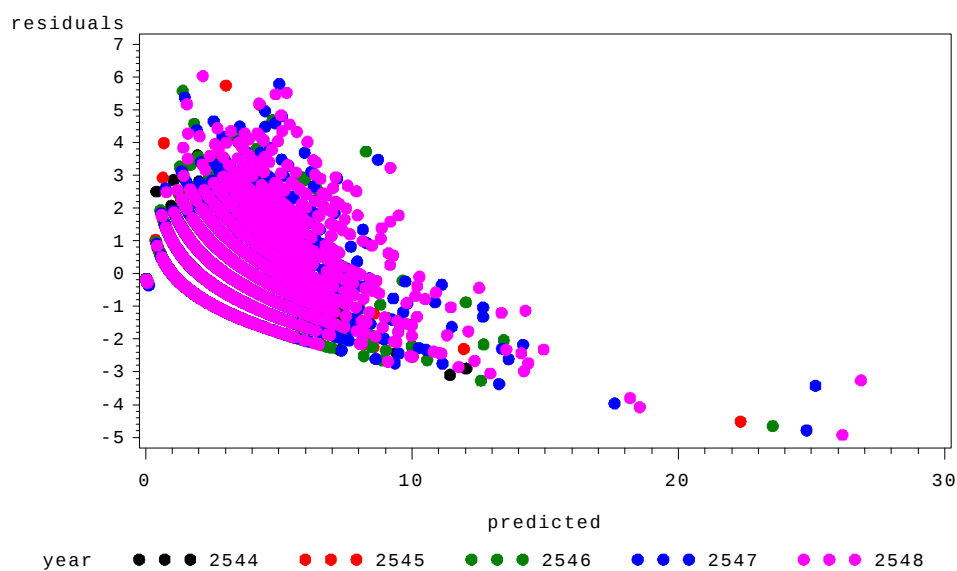
สำหรับการตรวจสอบความเหมาะสมของตัวแบบพบว่า ค่าสถิติของการวิเคราะห์ตัวแบบด้วยวิธี GEE เมื่อโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) พบว่าค่าสถิติ Pearson Chi-Square of residual /DF เท่ากับ 0.64 ส่วนรูปแบบ CS มีค่าสถิติ Pearson Chi-Square of residual /DF เท่ากับ 0.63 จะเห็นว่า ทั้งสองรูปแบบให้ค่าใกล้เคียงกัน แต่มีค่าไม่เกิน

1 แสดงว่าไม่เกิดปัญหามีค่าความแปรปรวนมากเกินไปจริงของข้อมูลทั้งหมด เนื่องจากตัวแปรตามเป็นข้อมูลที่มีจำนวนการเรียกค่าสินไหมทดแทนอย่างน้อยหนึ่งครั้งในระยะเวลา 5 ปี และพิจารณาค่าร้อยละของความถูกต้องที่ได้จากการทำนายของวิธี GEE เมื่อโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) และ CS มีค่าความถูกต้องเท่ากับร้อยละ 52.32 และ 52.05 ตามลำดับ แต่เมื่อพิจารณาค่าสถิติทดสอบเป็นหลักตัวแบบวิธี GEE ที่มีโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) จะมีค่า Pearson Chi-Square of residual เข้าใกล้ค่า Degree of freedom มากกว่ารูปแบบ CS จึงมีความเหมาะสมสำหรับใช้เป็นโครงสร้างความสัมพันธ์มากกว่าในตัวแบบวิธี GEE

และเมื่อนำตัวแบบที่ได้จากการประมาณค่าพารามิเตอร์ในวิธี GEE เมื่อกำหนดโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) และ CS มาทำการพล็อตค่า Residuals คู่กับค่าทำนายบนเส้นถดถอย (Predicted) หรือ $\hat{Y}$ ในระยะเวลาดำเนินการ 5 ปี ของการทำประกันภัยรถยนต์พบว่า การกระจายของข้อมูลมีความสัมพันธ์กันลักษณะแบบเส้นโค้ง หรืออยู่ในรูปแบบเอกซ์โพเนนเชียล ทั้งในวิธี GEE ที่มีรูปแบบ AR(1) และ CS ซึ่งตรงตามกับลักษณะของตัวแบบที่ตัวแปรอยู่ในรูปของลอการิทึม และตัวแปรตามหรือจำนวนการเรียกค่าสินไหมทดแทนมีการแจกแจงแบบปัวซอง แสดงว่า ตัวแบบของข้อมูลทั้งหมดมีความเหมาะสม ดังแสดงในภาพที่ 11 และภาพที่ 12



ภาพที่ 11 ความสัมพันธ์ของ residuals กับ $\hat{Y}$ (predicted) เมื่อมีโครงสร้างความสัมพันธ์เป็นแบบ AR(1)



ภาพที่ 12 ความสัมพันธ์ของ residuals กับ $\hat{Y}$ (predicted) เมื่อมีโครงสร้างความสัมพันธ์เป็นแบบ CS

6.2 ตัวแบบวิธี GLMMs

ตารางที่ 22 ผลวิเคราะห์ตัวแบบวิธี GLMMs เมื่อเปรียบเทียบโครงสร้างความแปรปรวนร่วม
เป็นแบบ AR(1) และ CS จากข้อมูลทั้งหมด

ตัวแปรอิสระ	วิธี GLMMs			
	AR(1)		CS	
Fixed-effect Parameter	Estimate (S.E.)	t [p-value]	Estimate (S.E.)	t [p-value]
ค่าคงที่ (INTERCEPT)	-2.7130 (0.0467)	-58.07 [<.0001]*	-2.9525 (0.0458)	-64.54 [<.0001]*
ค่าสินไหมทดแทนที่แปลง ค่า (TFMONEY)	0.1805 (0.0024)	75.19 [<.0001]*	0.1904 (0.0025)	76.69 [<.0001]*
อายุรถยนต์ (CAR-AGE)	0.0861 (0.0026)	32.82 [<.0001]*	0.0938 (0.0030)	31.55 [<.0001]*
อายุผู้ทำประกันภัย 18-24 ปี (DA1)	0.1390 (0.0862)	1.61 [0.1068]	0.1445 (0.0886)	1.63 [0.1028]
อายุผู้ทำประกันภัย 25-35 ปี (DA2)	0.1703 (0.0310)	5.49 [<.0001]*	0.1755 (0.0323)	5.43 [<.0001]*
อายุผู้ทำประกันภัย 36-50 ปี (DA3)	0.0744 (0.0288)	2.58 [0.0098]*	0.0751 (0.0301)	2.50 [0.0126]*
อาชีพ (OCCUP)	0.3105 (0.0514)	6.04 [<.0001]*	0.3054 (0.0537)	5.69 [<.0001]*
เพศ (GENDER)	-0.0487 (0.0179)	-2.72 [0.0065]*	-0.0517 (0.0186)	-2.79 [0.0053]*
Covariance Parameter	Estimate (S.E.)		Estimate (S.E.)	
Intercept	0.2416 (0.0071)		0.2791 (0.0075)	
Type	0.4464 (0.0110)		0.1474 (0.0018)	
Residual	0.1842 (0.0034)		0.1474 (0.0018)	
Random Subject Effect	Estimate (S.E.)		Estimate (S.E.)	
ID 1	0.4135 (0.2593)		0.5648 (0.1876)	
ID 2	-0.4155 (0.2105)		-0.4252 (0.1606)	
ID 3	0.4216 (0.1413)		0.4393 (0.1099)	
ID 4	0.3017 (0.2782)		0.4210 (0.2169)	

ตารางที่ 22 (ต่อ)

Random Subject Effect	วิธี GLMMs	
	AR(1)	CS
	Estimate (S.E.)	Estimate (S.E.)
ID 5	0.1005 (0.2003)	0.1284 (0.1517)
Model fit Criteria	AR(1)	CS
Generalized Chi-Square	3345.96	2677.77
Generalized Chi-Square / DF	0.18	0.15
Pseudo-AIC	36630.43	40278.53
Pseudo-BIC	36649.03	40290.93
R_p^2 (%) of $\hat{Y}_u$	76.28	79.36
R_p^2 (%) of $\hat{Y}$	53.34	53.38

หมายเหตุ เมื่อ *	มีระดับนัยสำคัญทางสถิติที่ระดับ 0.05
S.E.	คือ ค่าความคลาดเคลื่อนมาตรฐานของค่าสัมประสิทธิ์การถดถอย
$\hat{Y}$	คือ จำนวนการเรียกค่าสินไหมทดแทนที่ทำนายจากตัวแบบ
$\hat{Y}_u$	คือ จำนวนการเรียกค่าสินไหมทดแทนที่ทำนายจากตัวแบบ เมื่อมีเทอมอิทธิพลค่าคงที่สุ่ม
R_p^2	คือ ค่าร้อยละของการทำนายถูกต้อง

จากตารางที่ 22 พบว่า การวิเคราะห์ตัวแบบด้วยวิธี GLMMs สำหรับข้อมูลทั้งหมด เมื่อโครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) โดยการทดสอบสมมติฐานของสัมประสิทธิ์การถดถอยของตัวแปรอิสระแต่ละตัวจากค่าสถิติ t ที่ระดับนัยสำคัญทางสถิติ 0.05 คือ ค่าสินไหมทดแทนที่แปลงค่า อายุรถยนต์ อายุผู้ทำประกันภัย 25-35 ปี อายุผู้ทำประกันภัย 36-50 ปี อาชีพ และเพศ มีค่าเท่ากับ 75.19 (p-value < .0001), 32.82 (p-value < .0001), 5.49 (p-value < .0001), 2.58 (p-value = 0.0098), 6.04 (p-value < .0001) และ -2.72 (p-value = 0.0065) แสดงว่าตัวแปรอิสระเหล่านี้สามารถอธิบายการเปลี่ยนแปลงจำนวนการเรียกค่าสินไหมทดแทน ดังนั้นตัวแบบวิธี GLMMs ของการทำนายจำนวนการเรียกค่าสินไหมทดแทน

สำหรับโครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) เขียนได้ดังนี้

$$\begin{aligned}\log_e y = & (-2.7130 + \text{Random Subject Effect}) + 0.1805 (\text{ค่าสินไหมทดแทนที่แปลงค่า}) * \\ & + 0.0861 (\text{อายุรถยนต์}) * + 0.1390 (\text{อายุผู้ทำประกันภัย 18-24 ปี}) \\ & + 0.1703 (\text{อายุผู้ทำประกันภัย 25-35 ปี}) * + 0.0744 (\text{อายุผู้ทำประกันภัย 36-50 ปี}) * \\ & + 0.3105 (\text{อาชีพ}) * + (-0.0487) (\text{เพศ}) *\end{aligned}$$

(* มีระดับนัยสำคัญทางสถิติที่ระดับ 0.05)

ในการวิเคราะห์ด้วยวิธี GLMMs ที่มีรูปแบบ AR(1) จะมีการสร้างเทอมอิทธิพลค่าคงที่สุ่ม (Random Intercept) คือ หน่วยศึกษาจะมีค่าคงที่ (Intercept) แตกต่างกัน โดยการเพิ่มเทอมอิทธิพลสุ่มของหน่วยศึกษา (Random Subject Effect) ทำให้ตัวแบบของแต่ละหน่วยศึกษามีความแตกต่างกันแสดงได้ดังตัวอย่าง

หน่วยศึกษาที่ 1 มีค่าคงที่ = 0.4135 ดังนั้นตัวแบบสำหรับหน่วยศึกษาที่ 1 เขียนได้เป็น

$$\begin{aligned}\log_e y = & -2.2995 + 0.1805 (\text{ค่าสินไหมทดแทนที่แปลงค่า}) * + 0.0861 (\text{อายุรถยนต์}) * \\ & + 0.1390 (\text{อายุผู้ทำประกันภัย 18-24 ปี}) + 0.1703 (\text{อายุผู้ทำประกันภัย 25-35 ปี}) * \\ & + 0.0744 (\text{อายุผู้ทำประกันภัย 36-50 ปี}) * + 0.3105 (\text{อาชีพ}) * + (-0.0487) (\text{เพศ}) *\end{aligned}$$

โดยที่ -2.2995 มาจาก $-2.7130 + 0.4135$

หน่วยศึกษาที่ 2 มีค่าคงที่ = -0.4155 ดังนั้นตัวแบบสำหรับหน่วยศึกษาที่ 2 เขียนได้เป็น

$$\begin{aligned}\log_e y = & -3.1285 + 0.1805 (\text{ค่าสินไหมทดแทนที่แปลงค่า}) * + 0.0861 (\text{อายุรถยนต์}) * \\ & + 0.1390 (\text{อายุผู้ทำประกันภัย 18-24 ปี}) + 0.1703 (\text{อายุผู้ทำประกันภัย 25-35 ปี}) * \\ & + 0.0744 (\text{อายุผู้ทำประกันภัย 36-50 ปี}) * + 0.3105 (\text{อาชีพ}) * + (-0.0487) (\text{เพศ}) *\end{aligned}$$

โดยที่ -3.1285 มาจาก $-2.7130 + (-0.4155)$

ส่วนการวิเคราะห์ตัวแบบด้วยวิธี GLMMs สำหรับข้อมูลทั้งหมด เมื่อโครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ CS โดยการทดสอบสมมติฐานของสัมประสิทธิ์การถดถอยของตัวแปรอิสระแต่ละตัวจากค่าสถิติ t ที่ระดับนัยสำคัญทางสถิติ 0.05 คือค่าสินไหมทดแทนที่แปลงค่า อายุรถยนต์ อายุผู้ทำประกันภัย 25-35 ปี อายุผู้ทำประกันภัย 36-50 ปี อาชีพ และเพศ มีค่าเท่ากับ 76.69 (p-value < .0001), 31.55 (p-value < .0001), 5.43 (p-value < .0001), 2.50 (p-value = 0.0126), 5.69 (p-value < .0001) และ -2.79 (p-value = 0.0053) แสดงว่าตัวแปรอิสระเหล่านี้สามารถอธิบายการเปลี่ยนแปลงจำนวนการเรียกค่าสินไหมทดแทน ดังนั้นตัวแบบวิธี GLMMs ของการทำนายจำนวนการเรียกค่าสินไหมทดแทนสำหรับโครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ CS เขียนได้ดังนี้

$$\begin{aligned} \log_e y = & (-2.9525 + \text{Random Subject Effect}) + 0.1904 (\text{ค่าสินไหมทดแทนที่แปลงค่า}) * \\ & + 0.0938 (\text{อายุรถยนต์}) * + 0.1445 (\text{อายุผู้ทำประกันภัย 18-24 ปี}) \\ & + 0.1755 (\text{อายุผู้ทำประกันภัย 25-35 ปี}) * + 0.0751 (\text{อายุผู้ทำประกันภัย 36-50 ปี}) * \\ & + 0.3054 (\text{อาชีพ}) * + (-0.0517) (\text{เพศ}) * \end{aligned}$$

(* มีระดับนัยสำคัญทางสถิติที่ระดับ 0.05)

และการวิเคราะห์ตัวแบบด้วยวิธี GLMMs ที่มีรูปแบบ CS จะมีการสร้างเทอมอิทธิพลค่าคงที่สุ่ม เช่นเดียวกับรูปแบบ AR(1) คือ หน่วยศึกษาจะมีค่าคงที่ที่แตกต่างกัน โดยการเพิ่มเทอมอิทธิพลสุ่มของหน่วยศึกษา (Random Subject Effect) ทำให้ตัวแบบของแต่ละหน่วยศึกษามีความแตกต่างกันแสดงได้ดังตัวอย่าง

หน่วยศึกษาที่ 1 มีค่าคงที่ = 0.5648 ดังนั้นตัวแบบสำหรับหน่วยศึกษาที่ 1 เขียนได้เป็น

$$\begin{aligned} \log_e y = & -2.3877 + 0.1904 (\text{ค่าสินไหมทดแทนที่แปลงค่า}) * + 0.0938 (\text{อายุรถยนต์}) * \\ & + 0.1445 (\text{อายุผู้ทำประกันภัย 18-24 ปี}) + 0.1755 (\text{อายุผู้ทำประกันภัย 25-35 ปี}) * \\ & + 0.0751 (\text{อายุผู้ทำประกันภัย 36-50 ปี}) * + 0.3054 (\text{อาชีพ}) * + (-0.0517) (\text{เพศ}) * \end{aligned}$$

โดยที่ -2.3877 มาจาก -2.9525 + 0.5648

หน่วยศึกษาที่ 2 มีค่าคงที่ = -0.4252 ดังนั้นตัวแบบสำหรับหน่วยศึกษาที่ 2 เขียนได้เป็น

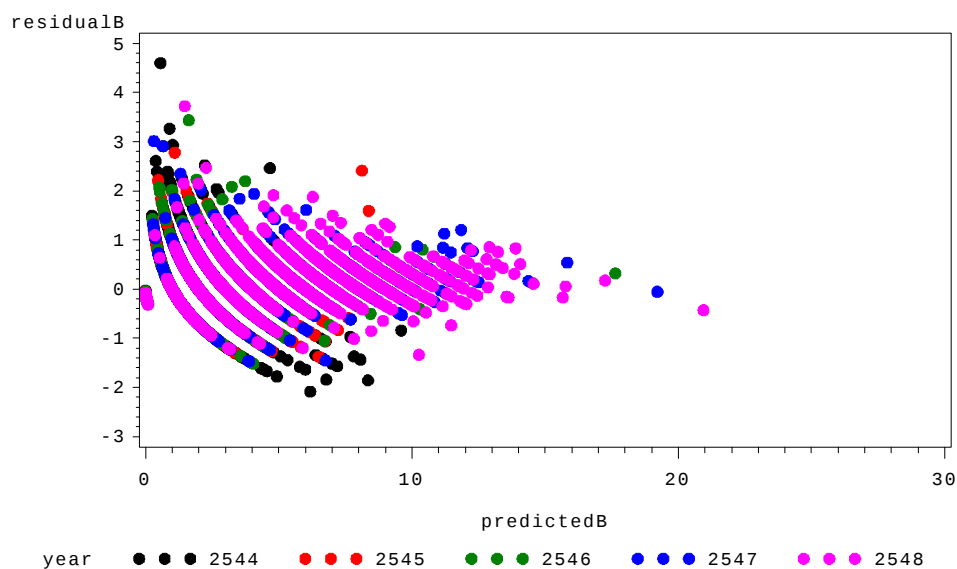
$$\begin{aligned} \log_e y = & -3.3777 + 0.1904 (\text{ค่าสินไหมทดแทนที่แปลงค่า})^* + 0.0938 (\text{อายุรถยนต์})^* \\ & + 0.1445 (\text{อายุผู้ทำประกันภัย 18-24 ปี}) + 0.1755 (\text{อายุผู้ทำประกันภัย 25-35 ปี})^* \\ & + 0.0751 (\text{อายุผู้ทำประกันภัย 36-50 ปี})^* + 0.3054 (\text{อาชีพ})^* + (-0.0517) (\text{เพศ})^* \end{aligned}$$

โดยที่ -3.3777 มาจาก $-2.9525 + (-0.4252)$

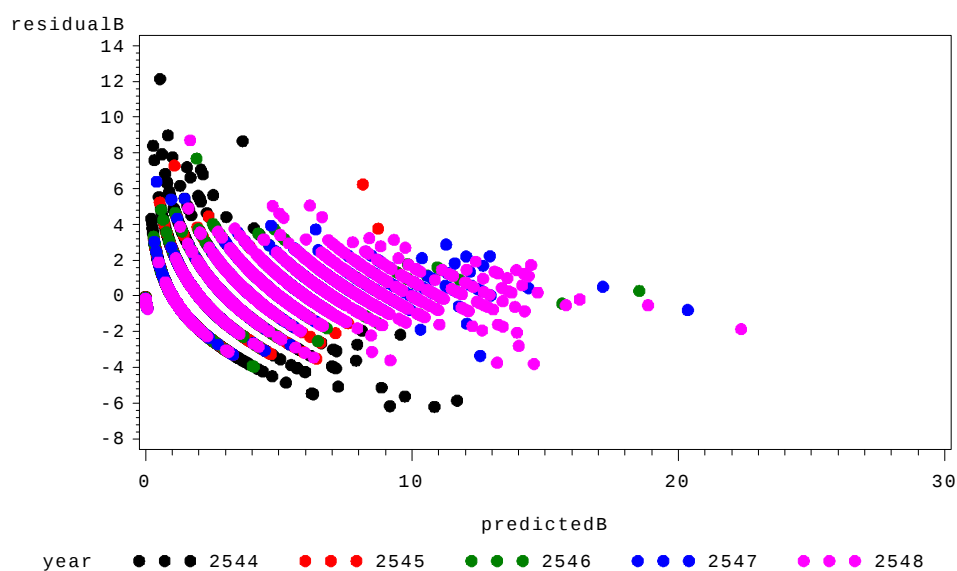
สำหรับการตรวจสอบความเหมาะสมของตัวแบบพบว่า ค่าสถิติของการวิเคราะห์ตัวแบบด้วยวิธี GLMMs เมื่อโครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) และ CS มีค่าสถิติ Generalized Chi-Square / DF เท่ากับ 0.18 และ 0.15 ตามลำดับ ซึ่งมีค่าน้อยกว่า 1 แสดงว่าไม่เกิดปัญหามีค่าความแปรปรวนมากเกินไปจริง และเนื่องจากในข้อมูลทั้งหมดตัวแปรตามเป็นข้อมูลที่มีจำนวนการเรียกค่าสินไหมทดแทนอย่างน้อยหนึ่งครั้งในระยะเวลา 5 ปี และพิจารณาค่าร้อยละของความถูกต้องที่ได้จากการทำนายของตัวแบบในภาพรวมของค่าเฉลี่ยประชากรในรูปแบบ AR(1) และ CS มีค่าความถูกต้องเท่ากับร้อยละ 53.34 และ 53.38 ตามลำดับ ซึ่งมีค่าใกล้เคียงกัน ส่วนค่าร้อยละของความถูกต้องที่ได้จากการทำนายของตัวแบบเมื่อมีเทอมอิทธิพลค่าคงที่ของแต่ละบุคคลมีค่าเท่ากับ 76.28 และ 79.36 ตามลำดับ แสดงว่าการเพิ่มเทอมค่าคงที่สำหรับแต่ละบุคคลจะทำให้ตัวแบบของข้อมูลทั้งหมด จะมีความสามารถในการทำนายค่ามากขึ้น นอกจากนี้การพิจารณาความเหมาะสมของโครงสร้างความแปรปรวนร่วมของตัวแบบวิธี GLMMs ด้วยค่าสถิติ Pseudo-AIC จากรูปแบบ AR(1) และ CS มีค่าเท่ากับ 36630.43 และ 40278.53 ตามลำดับ และ Pseudo-BIC รูปแบบ AR(1) และ CS คือ 36649.03 และ 40290.93 ตามลำดับ แสดงว่ารูปแบบ AR(1) มีค่าสถิติ Pseudo-AIC และ Pseudo-BIC มีค่าต่ำกว่ารูปแบบ CS ดังนั้นการใช้โครงสร้างความแปรปรวนร่วมรูปแบบ AR(1) จะมีความเหมาะสมมากกว่ารูปแบบ CS

และเมื่อนำตัวแบบที่ได้จากการประมาณค่าพารามิเตอร์วิธี GLMMs เมื่อกำหนดโครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) และ CS มาทำการพล็อตค่า Residuals คู่กับค่าทำนายบนเส้นถดถอย (Predicted) หรือ $\hat{Y}_u$ ในระยะเวลาติดตาม 5 ปีของการทำประกันภัยรถยนต์พบว่า การกระจายของข้อมูลมีความสัมพันธ์กันลักษณะแบบเส้นโค้ง หรืออยู่ในรูปแบบเอกซ์โพเนนเชียล ทั้งในวิธี GLMMs ที่มีรูปแบบ AR(1) และ CS ซึ่งตรงตามกับลักษณะของตัวแบบที่ตัวแปรอยู่ในรูปของลอการิทึม และตัวแปรตามหรือจำนวนการเรียกค่าสินไหมทดแทนมี

การแจกแจงแบบปัวซองแสดงว่า การวิเคราะห์ด้วยวิธี GLMMs สามารถนำไปใช้เป็นตัวแทน
สำหรับจำนวนการเรียกค่าสินไหมทดแทนได้ ดังแสดงในภาพที่ 13 และภาพที่ 14



ภาพที่ 13 ความสัมพันธ์ของ residuals กับ $\hat{Y}_u$ (predicted) เมื่อมีโครงสร้างความแปรปรวนร่วม
เป็นแบบ AR(1)



ภาพที่ 14 ความสัมพันธ์ของ residuals กับ $\hat{Y}_u$ (predicted) เมื่อมีโครงสร้างความแปรปรวนร่วม
เป็นแบบ CS

7. ผลการวิเคราะห์การหาตัวแบบที่เหมาะสมสำหรับข้อมูลจำนวนการเรียกค่าสินไหมทดแทน

7.1 ตัวแบบวิธี GEE

ตารางที่ 23 ผลการวิเคราะห์ตัวแบบสุดท้ายที่เหมาะสมสำหรับตัวแบบวิธี GEE เมื่อกำหนดโครงสร้างความสัมพันธ์เป็นแบบ AR(1)

ตัวแปรอิสระ	ตัวแบบวิธี GEE				
	Estimate (S.E.)	Z [p-value]	RR exp (b)	95% Confidence Interval of RR	
ค่าคงที่ (INTERCEPT)	-1.8074 (0.0530)	-34.08 [<.0001]*	0.1641	0.1479	0.1821
ค่าสินไหมทดแทนที่แปลงค่า (TFMONEY)	0.1464 (0.0029)	50.23 [<.0001]*	1.1577	1.1511	1.1644
อายุรถยนต์ (CAR-AGE)	0.0452 (0.0025)	18.15 [<.0001]*	1.0462	1.0411	1.0514
อายุผู้ทำประกันภัย 18-24 ปี (DA1)	0.1113 (0.0835)	1.33 [0.1828]	1.1177	0.9489	1.3165
อายุผู้ทำประกันภัย 25-35 ปี (DA2)	0.1656 (0.0350)	4.73 [<.0001]*	1.1801	1.1018	1.2640
อายุผู้ทำประกันภัย 36-50 ปี (DA3)	0.0787 (0.0329)	2.39 [0.0168]*	1.0819	1.0143	1.1540
อาชีพ (OCCUP)	0.3488 (0.0495)	7.04 [<.0001]*	1.4174	1.2862	1.5617
เพศ (GENDER)	-0.0366 (0.0190)	-1.92 [0.0546]	0.9641	0.9288	1.0007
Model fit Criteria					
Pearson Chi-Square of residual	11682.680				
Pearson Chi-Square of residual / DF	0.64				
R_p^2 (%) ของ $\hat{Y}$	52.32				

หมายเหตุ เมื่อ * มีระดับนัยสำคัญทางสถิติที่ระดับ 0.05
S.E. คือ ค่าความคลาดเคลื่อนมาตรฐานของค่าสัมประสิทธิ์การถดถอย

ตารางที่ 23 (ต่อ)

$\hat{Y}$	คือ จำนวนการเรียกค่าสินไหมทดแทนที่ทำนายจากตัวแบบ
R_p^2	คือ ค่าร้อยละของการทำนายถูกต้อง
RR	คือ Relative Rate

จากตารางที่ 23 การวิเคราะห์ด้วยวิธี GEE เมื่อกำหนดโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ เป็นแบบ AR(1) จะเป็นรูปแบบที่เหมาะสมสำหรับข้อมูลจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ ซึ่งใช้ในการวิเคราะห์ภาพรวมของค่าเฉลี่ยประชากร ปีจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนคือ ค่าสินไหมทดแทนที่แปลงค่า อายุรถยนต์ อายุผู้ทำประกันภัย 25-35 ปี อายุผู้ทำประกันภัย 36-50 ปี และอาชีพ ซึ่งแต่ละปีจัยมีจำนวนการเรียกค่าสินไหมทดแทนเฉลี่ยอยู่ในช่วงความเชื่อมั่น 95 % ของ Relative Rate (RR) คือ (1.1511, 1.1644), (1.0411, 1.0514), (1.1018, 1.2640), (1.0143, 1.1540) และ (1.2862, 1.5617) ครั้งต่อปี ตามลำดับโดยตัวแบบวิธี GEE ที่มีโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) สามารถทำนายความถูกต้องของจำนวนการเรียกค่าสินไหมทดแทนในภาพรวมได้ร้อยละ 52.32 มีค่าสถิติ Pearson Chi-Square of residual/DF เท่ากับ 0.64 ดังนั้นในการวิเคราะห์ภาพรวมของค่าเฉลี่ยประชากรของข้อมูลจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ของการศึกษานี้จึงควรใช้ตัวแบบวิธี GEE ที่มีโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) สำหรับข้อมูลทั้งหมดโดยเขียนตัวแบบได้ดังนี้

$$\begin{aligned} \log_e y = & -1.8074 + 0.1464(\text{ค่าสินไหมทดแทนที่แปลงค่า}) * + 0.0452(\text{อายุรถยนต์}) * \\ & + 0.1113(\text{กลุ่มอายุ 18-24 ปี}) + 0.1656(\text{กลุ่มอายุ 25-35 ปี}) * \\ & + 0.0787(\text{กลุ่มอายุ 36-50 ปี}) * + 0.3488 (\text{อาชีพ}) * + (-0.0366) (\text{เพศ}) \end{aligned}$$

(* มีนัยสำคัญทางสถิติที่ระดับ 0.05)

การอธิบายความหมาย ของวิธี GEE สำหรับข้อมูลจำนวนการเรียกค่าสินไหมทดแทน ที่มี การวัดซ้ำในระยะเวลาติดตาม 5 ปี เมื่อโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) ที่ได้จากข้อมูลทั้งหมดนี้คือ ตัวแปรอธิบายแต่ละตัวจะมีการควบคุมซึ่งกัน

และกัน ดังนั้นเมื่อต้องการเปรียบเทียบจำนวนการเรียกค่าสินไหมทดแทนจากตัวแปรอธิบายใด ๆ จำเป็นต้องควบคุมตัวแปรอื่นให้มีค่าคงที่

สำหรับตัวแปรอิสระเป็นข้อมูลเชิงปริมาณ

ถ้าปัจจัยค่าสินไหมทดแทนเพิ่มขึ้น 1 หน่วยข้อมูลที่แปลงค่า ทำให้จำนวนการเรียกค่าสินไหมทดแทน เพิ่มขึ้นโดยเฉลี่ยเป็น 1.16 ครั้ง/คน/ปี เมื่อปัจจัยอื่น ๆ คงที่ [$\exp(0.1464) = 1.16$]

ถ้าปัจจัยอายุรถยนต์เพิ่มขึ้น 1 ปี ทำให้จำนวนการเรียกค่าสินไหมทดแทน เพิ่มขึ้นโดยเฉลี่ยเป็น 1.05 ครั้ง/คน/ปี เมื่อปัจจัยอื่น ๆ คงที่ [$\exp(0.0452) = 1.05$]

สำหรับตัวแปรอิสระเป็นข้อมูลเชิงคุณภาพ

กลุ่มอายุ 18-24 ปีจะมีจำนวนการเรียกค่าสินไหมทดแทนโดยเฉลี่ยเป็น 1.12 ครั้ง/คน/ปี เมื่อเทียบกับกลุ่มอายุเกิน 50 ปี โดยที่ปัจจัยอื่น ๆ คงที่ [$\exp(0.1113) = 1.12$]

กลุ่มอายุ 25-35 ปีจะมีจำนวนการเรียกค่าสินไหมทดแทนโดยเฉลี่ยเป็น 1.18 ครั้ง/คน/ปี เมื่อเทียบกับกลุ่มอายุเกิน 50 ปี โดยที่ปัจจัยอื่น ๆ คงที่ [$\exp(0.1656) = 1.18$]

กลุ่มอายุ 36-50 ปีจะมีจำนวนการเรียกค่าสินไหมทดแทนโดยเฉลี่ยเป็น 1.08 ครั้ง/คน/ปี เมื่อเทียบกับกลุ่มอายุเกิน 50 ปี โดยที่ปัจจัยอื่น ๆ คงที่ [$\exp(0.0787) = 1.08$]

ผู้ทำประกันภัยรถยนต์ที่มีอาชีพเสี่ยงมากจะมีจำนวนการเรียกค่าสินไหมทดแทนโดยเฉลี่ยเป็น 1.42 ครั้ง/คน/ปี เมื่อเทียบกับผู้ที่มีอาชีพเสี่ยงน้อย โดยที่ปัจจัยอื่น ๆ คงที่ [$\exp(0.3488) = 1.42$]

7.2 ตัวแบบวิธี GLMMs

ตารางที่ 24 ผลการวิเคราะห์ตัวแบบที่เหมาะสมด้วยวิธี GLMMs ที่มีโครงสร้างความแปรปรวนร่วมเป็นแบบ AR(1)

ตัวแปรอิสระ	ตัวแบบวิธี GLMMs				
	Estimate (S.E.)	t [p-value]	RR exp (b)	95% Confidence Interval of RR	
ค่าคงที่ (INTERCEPT)	-2.7130 (0.0467)	-58.07 [<.0001]*	0.0663	0.0605	0.0727
ค่าสินไหมทดแทนที่แปลงค่า (TFMONEY)	0.1805 (0.0024)	75.19 [<.0001]*	1.1978	1.1922	1.2035

ตารางที่ 24 (ต่อ)

ตัวแปรอิสระ	ตัวแบบวิธี GLMMs				
	Estimate (S.E.)	t [p-value]	RR exp (b)	95% Confidence Interval of RR	
อายุรถยนต์ (CAR-AGE)	0.0861 (0.0026)	32.82 [<.0001]*	1.0899	1.0843	1.0955
อายุผู้ทำประกันภัย 18-24 ปี (DA1)	0.1390 (0.0862)	1.61 [0.1068]	1.1491	0.9705	1.3607
อายุผู้ทำประกันภัย 25-35 ปี (DA2)	0.1703 (0.0310)	5.49 [<.0001]*	1.1857	1.1157	1.2600
อายุผู้ทำประกันภัย 36-50 ปี (DA3)	0.0744 (0.0288)	2.58 [0.0098]*	1.0773	1.0181	1.1400
อาชีพ (OCCUP)	0.3105 (0.0514)	6.04 [<.0001]*	1.3641	1.2334	1.5086
เพศ (GENDER)	-0.0487 (0.0179)	-2.72 [0.0065]*	0.9524	0.9196	0.9865
Model fit Criteria					
Generalized Chi-Square			3345.96		
Generalized Chi-Square/DF			0.18		
R ² _p (%)					
Ŷ			53.34		
Ŷ _u			76.28		

หมายเหตุ	เมื่อ *	มีระดับนัยสำคัญทางสถิติที่ระดับ 0.05
	S.E.	คือ ค่าความคลาดเคลื่อนมาตรฐานของค่าสัมประสิทธิ์การถดถอย
	$\hat{Y}$	คือ จำนวนการเรียกค่าสินไหมทดแทนที่ทำนายจากตัวแบบ
	$\hat{Y}_u$	คือ จำนวนการเรียกค่าสินไหมทดแทนที่ทำนายจากตัวแบบ เมื่อมีเทอมอิทธิพลค่าคงที่สุ่ม
	R_p^2	คือ ค่าร้อยละของการทำนายถูกต้อง
	RR	คือ Relative Rate

จากตารางที่ 24 การวิเคราะห์ด้วยวิธี GLMMs เมื่อกำหนดโครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) จะเป็นรูปแบบที่เหมาะสมสำหรับข้อมูลจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ ซึ่งใช้ในการวิเคราะห์เน้นเฉพาะในแต่ละหน่วยศึกษา คือแต่ละหน่วยศึกษามีค่าคงที่ที่แตกต่างกันหรือค่าคงที่สุ่ม ซึ่งแตกต่างกับวิธี GEE โดยปัจจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนคือ ค่าสินไหมทดแทนที่แปลงค่า อายุรถยนต์ อายุผู้ทำประกันภัย 25-35 ปี อายุผู้ทำประกันภัย 36-50 ปี และอาชีพ ซึ่งแต่ละปัจจัยมีจำนวนการเรียกค่าสินไหมทดแทนเฉลี่ยอยู่ในช่วงความเชื่อมั่น 95 % ของ Relative Rate (RR) คือ (1.1922,1.2035), (1.0843,1.0955), (1.1157,1.2600), (1.0181,1.1400), (1.2334,1.5086) และ (0.9196,0.9865) รั้งต่อปี ตามลำดับ ตัวแบบวิธี GLMMs ที่มีโครงสร้างความแปรปรวนของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) สามารถทำนายความถูกต้องของจำนวนการเรียกค่าสินไหมทดแทนในภาพรวมได้ร้อยละ 52.34 และทำนายความถูกต้องของจำนวนการเรียกค่าสินไหมทดแทนเมื่อแต่ละหน่วยศึกษามีค่าคงที่ที่แตกต่างกันได้ร้อยละ 76.28 รวมทั้งมีค่าสถิติ Generalized Chi-Square/DF เท่ากับ 0.18 ดังนั้นในการวิเคราะห์เน้นเฉพาะในแต่ละหน่วยศึกษาของข้อมูลจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ของการศึกษานี้จึงควรใช้ตัวแบบวิธี GLMMs ที่มีโครงสร้างความแปรปรวนของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) สำหรับข้อมูลทั้งหมด โดยเขียนตัวแบบได้ดังนี้

$$\log_e y = (-2.7130 + \text{Random Subject Effect}) + 0.1805(\text{ค่าสินไหมทดแทนที่แปลงค่า}) * \\ + 0.0861(\text{อายุรถยนต์}) * + 0.1390(\text{กลุ่มอายุ 18-24 ปี}) + 0.1703(\text{กลุ่มอายุ 25-35 ปี}) * \\ + 0.0744(\text{กลุ่มอายุ 36-50 ปี}) * + 0.3105(\text{อาชีพ}) * + (-0.0487)(\text{เพศ}) *$$

(* มีนัยสำคัญทางสถิติที่ระดับ 0.05)

การอธิบายความหมาย ของวิธี GLMMs เมื่อกำหนดโครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) สำหรับข้อมูลทั้งหมด มีลักษณะเช่นเดียวกับวิธี GEE คือ

สำหรับตัวแปรอิสระเป็นข้อมูลเชิงปริมาณ

ถ้าปัจจัยค่าสินไหมทดแทนเพิ่มขึ้น 1 หน่วยที่แปลงค่า ทำให้จำนวนการเรียกค่าสินไหมทดแทนเพิ่มขึ้นโดยเฉลี่ยเป็น 1.20 ครั้ง/คน/ปี เมื่อปัจจัยอื่น ๆ คงที่ [$\exp(0.1805) = 1.20$]

ถ้าปัจจัยอายุรถยนต์เพิ่มขึ้น 1 ปี ทำให้จำนวนการเรียกค่าสินไหมทดแทนเพิ่มขึ้น โดยเฉลี่ยเป็น 1.09 ครั้ง/คน/ปี เมื่อปัจจัยอื่น ๆ คงที่ [$\exp(0.0861) = 1.09$]

สำหรับตัวแปรอิสระเป็นข้อมูลเชิงคุณภาพ

กลุ่มอายุ 18-24 ปีจะมีจำนวนการเรียกค่าสินไหมทดแทนโดยเฉลี่ยเป็น 1.15 ครั้ง/คน/ปี เมื่อเทียบกับกลุ่มอายุเกิน 50 ปี โดยที่ปัจจัยอื่น ๆ คงที่ [$\exp(0.1390) = 1.15$]

กลุ่มอายุ 25-35 ปีจะมีจำนวนการเรียกค่าสินไหมทดแทนโดยเฉลี่ยเป็น 1.19 ครั้ง/คน/ปี เมื่อเทียบกับกลุ่มอายุเกิน 50 ปี โดยที่ปัจจัยอื่น ๆ คงที่ [$\exp(0.1703) = 1.19$]

กลุ่มอายุ 36-50 ปีจะมีจำนวนการเรียกค่าสินไหมทดแทนโดยเฉลี่ยเป็น 1.08 ครั้ง/คน/ปี เมื่อเทียบกับกลุ่มอายุเกิน 50 ปี โดยที่ปัจจัยอื่น ๆ คงที่ [$\exp(0.0744) = 1.08$]

ผู้ทำประกันภัยรถยนต์ที่มีอาชีพเสี่ยงมากจะมีจำนวนการเรียกค่าสินไหมทดแทนโดยเฉลี่ยเป็น 1.36 ครั้ง/คน/ปี เมื่อเทียบกับผู้ที่มีอาชีพเสี่ยงน้อย โดยที่ปัจจัยอื่น ๆ คงที่ [$\exp(0.3105) = 1.36$]

ผู้ทำประกันภัยรถยนต์ที่เป็นเพศชายจะมีจำนวนการเรียกค่าสินไหมทดแทนโดยเฉลี่ยเป็น 1.36 ครั้ง/คน/ปี เมื่อเทียบกับผู้ทำประกันภัยรถยนต์ที่เป็นเพศหญิง โดยที่ปัจจัยอื่น ๆ คงที่ [$\exp(0.3105) = 1.36$]

วิจารณ์

1. ข้อมูลที่ใช้ในการศึกษานี้เป็นข้อมูลการประกันภัยรถยนต์เพียงบางส่วนจากข้อมูลการประกันภัยรถยนต์ทั้งหมดที่จดทะเบียนในกรุงเทพมหานคร ตั้งแต่ปี 2544 – 2548 เป็นการประกันภัยภาคสมัครใจ ประเภท 1 แบบระบุผู้ขับขี่ รถยนต์นั่งส่วนบุคคล รหัส 110 ซึ่งได้รับการอนุเคราะห์จากกรมการประกันภัยแต่ยังขาดข้อมูลที่สำคัญของปัจจัยเกี่ยวกับรายได้ของผู้ทำประกันภัย คะแนนในการขับขี่ และระยะทางในการขับขี่ของรถยนต์ เป็นต้น และเนื่องจากการเป็นประกันภัย แบบระบุชื่อผู้ขับขี่ ซึ่งสามารถระบุชื่อผู้ขับขี่ได้ 2 คน จึงเป็นไปได้ที่ผู้ขับขี่อาจไม่ใช่ผู้ทำประกันภัยรถยนต์ ซึ่งพฤติกรรม การขับขี่ก็อาจต่างกัน ดังนั้นข้อมูลที่เกี่ยวข้องกับผู้ทำประกันภัยรถยนต์ จึงอธิบายร้อยละการทํานายถูกต้องของจำนวนการเรียกค่าสินไหมทดแทนได้ในระดับที่ยังไม่ดีเท่าที่ควร

2. การคัดกรองตัวแปรอิสระเข้ามาในตัวแบบทีละตัวสำหรับสร้างตัวแบบทางสถิติ ด้วยวิธี GEE และวิธี GLMMs พบว่า ปัจจัยเพศ จะไม่ถูกคัดกรองเข้ามาอธิบายในตัวแบบทั้งสองวิธี ซึ่งให้ผลสอดคล้องกับการศึกษาของ Yau *et al.* (2003) แต่เมื่อวิเคราะห์ในภาพรวมจะนำตัวแปร

เพศมาศึกษาด้วย ดังนั้นการศึกษานี้จึงนำมาตัวแปรเพศเข้ามาพิจารณาด้วย เนื่องจากที่นำมาศึกษาเป็นข้อมูล แบบระบุชื่อผู้ขับขี่ และพบว่าปัจจัยเพศถูกคัดเลือกเข้ามาในตัวแบบทางสถิติวิธี GEE และวิธี GLMMs ซึ่งจริง ๆ แล้วปัจจัยเพศอาจไม่มีความจำเป็นมากนักเนื่องจากการทำประกันแบบระบุชื่อผู้ขับขี่ สามารถระบุชื่อผู้ขับขี่ได้ 2 คน ทำให้ปัจจัยเพศอธิบายการเปลี่ยนแปลงจำนวนครั้งในการเรียกค่าสินไหมทดแทนได้ไม่มากนัก

3. การแบ่งข้อมูลเป็น 2 ชุดจากการสุ่ม คือ ข้อมูลฝึกสอน และข้อมูลทดสอบ เพื่อตรวจสอบความถูกต้องในการทำนายของตัวแบบ เมื่อข้อมูลฝึกสอน ข้อมูลทดสอบ และข้อมูลทั้งหมดมีลักษณะการแจกแจงที่ปรากฏในข้อ 1 และ 2 จากหน้า 51 เป็นไปในรูปแบบเดียวกันจะนำข้อมูลฝึกสอนมาสร้างตัวแบบและใช้ข้อมูลทดสอบในการทำนายค่าจากตัวแบบที่ได้จากข้อมูลฝึกสอน หากผลการทำนายที่ได้มีค่าใกล้เคียงกันทั้งในชุดข้อมูลฝึกสอนและข้อมูลทดสอบ และผลการทำนายที่ได้จากตัวแบบของข้อมูลทั้งหมด จะแสดงว่าตัวแบบถูกต้องและมีความเหมาะสมกับข้อมูล แต่ถ้าวการแจกแจงของชุดข้อมูลทั้งสามไม่เป็นไปในลักษณะเดียวกันอาจมีผลต่อค่าร้อยละในการทำนายความถูกต้องของตัวแบบ

4. ผลการวิเคราะห์ของตัวแบบทางสถิติด้วยวิธี GEE และวิธี GLMMs ที่มีรูปแบบ AR(1) และ CS เมื่อพิจารณาจากค่าร้อยละการทำนายถูกต้องของทั้งสองวิธีในการวิเคราะห์ภาพรวมของประชากรจะมีค่าน้อยคือประมาณร้อยละ 52 เนื่องจากตัวแปรอิสระที่นำมาสร้างตัวแบบจำนวนการเรียกค่าสินไหมทดแทนไม่ใช่ตัวแปรที่สำคัญ ดังนั้นควรพิจารณาปัจจัยด้านอื่น ๆ เช่น พฤติกรรมของผู้ขับขี่ ระยะทางของรถยนต์ และคะแนนของผู้ขับขี่ เป็นต้น ส่วนการวิเคราะห์ด้วยวิธี GLMMs เมื่อมีความแตกต่างของเทอมค่าคงที่สุ่มในแต่ละหน่วยศึกษาหรือบุคคลจะมีค่าร้อยละของการทำนายถูกต้องสูงขึ้นคือประมาณร้อยละ 78 แสดงได้ว่าพื้นฐานของแต่ละบุคคลมีความแตกต่างกัน และให้ค่าร้อยละการทำนายในภาพรวมใกล้เคียงกับวิธี GEE คือประมาณร้อยละ 53 ดังนั้นหากมีการนำปัจจัยที่เหมาะสมมาอธิบายจำนวนการเรียกค่าสินไหมทดแทนได้มากขึ้นจะทำให้ตัวแบบสามารถทำนายจำนวนการเรียกค่าสินไหมทดแทนได้ถูกต้องมากขึ้นด้วย และเป็นตัวแบบที่มีประสิทธิภาพ

สรุปและข้อเสนอแนะ

สรุป

การศึกษานี้เป็นการสร้างตัวแบบสำหรับจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ที่เก็บข้อมูลซ้ำในระยะเวลาติดตาม 5 ปีของผู้ทำประกันภัยคนเดียวกัน ตั้งแต่ปี 2544-2548 โดยใช้ข้อมูลการประกันภัยรถยนต์ประกันภัยรถยนต์ภาคสมัครใจ ประเภท 1 แบบระบุชื่อผู้ขับขี่ รหัส 110 รถยนต์นั่งส่วนบุคคล ที่จดทะเบียนในกรุงเทพมหานคร ที่เริ่มทำประกันภัยในปี 2544 จากกรมการประกันภัย กระทรวงพาณิชย์ โดยเป็นข้อมูลจำนวนการเรียกค่าสินไหมทดแทนอย่างน้อยหนึ่งครั้งในระยะเวลาทำประกันภัย 5 ปี จำนวน 3,635 กรมธรรม์ ด้วยตัวแบบวิธี GEE เมื่อสมมติโครงสร้างความสัมพันธ์ของข้อมูลในหน่วยศึกษามีรูปแบบ AR(1) และ CS และตัวแบบวิธี GLMMs เมื่อกำหนดโครงสร้างความแปรปรวนร่วมของข้อมูลในหน่วยศึกษามีรูปแบบ AR(1) และ CS รวมทั้งหาตัวแบบทางสถิติที่เหมาะสมสำหรับวิธี GEE กับวิธี GLMMs พร้อมทั้งหาปัจจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์

ในการศึกษาตัวแบบทางสถิติสำหรับจำนวนครั้งการเรียกค่าสินไหมทดแทน มีวิธีการศึกษาดังนี้

1. ศึกษาลักษณะข้อมูลด้วยสถิติเชิงพรรณนา และการแจกแจงของข้อมูล รวมทั้งแก้ปัญหาความผิดพลาดของข้อมูลในชุดข้อมูลทั้งหมด และทำการแบ่งข้อมูลเป็น 2 ชุด ตามหลัก Cross-Validation ด้วยวิธีการสุ่มอย่างง่ายคือ ทำการสุ่มขนาดตัวอย่าง 70 % ของข้อมูลทั้งหมดเป็นชุดข้อมูลฝึกสอน และที่เหลือ 30 % เป็นชุดข้อมูลทดสอบ รวมทั้งตรวจสอบข้อมูลและการกระจายของข้อมูลทั้งสองชุดว่าสอดคล้องในลักษณะเดียวกันกับข้อมูลทั้งหมดหรือไม่

2. นำข้อมูลฝึกสอนมาพัฒนาตัวแบบทางสถิติ เริ่มจากการคัดกรองตัวแปรอิสระด้วยการวิเคราะห์ตัวแปรเดียวระหว่างตัวแปรตามจำนวนการเรียกค่าสินไหมทดแทนกับตัวแปรอิสระทีละตัวด้วยตัวแบบวิธี GEE เมื่อกำหนดโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) และ CS ด้วยตัวทดสอบสถิติ Wald และวิเคราะห์ตัวแปรเดียวด้วยตัวแบบวิธี GLMMs เมื่อกำหนดโครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) และ CS ด้วยตัวทดสอบสถิติ t โดยพิจารณาตัวแปรอิสระที่ให้ค่า p-value น้อยกว่า 0.20

แล้วนำไปวิเคราะห์หลายตัวแปรเพื่อคัดเลือกตัวแปรอิสระด้วยวิธี Forward Stepwise ในตัวแบบวิธี GEE โดยพิจารณาจากค่าสถิติ Generalized Score Statistic และในตัวแบบวิธี GLMMs โดยพิจารณาจากค่าสถิติ G (Residual Log Pseudo-Likelihood Ratio Test Statistic) ซึ่งการคัดเลือกตัวแปรของทั้ง 2 ตัวแบบกำหนดเกณฑ์การคัดเลือกตัวแปรอิสระเข้าตัวแบบที่ระดับนัยสำคัญทางสถิติ 0.15 ($P_E = 0.15$) และเกณฑ์ในการคัดเลือกตัวแปรอิสระออกจากตัวแบบที่ระดับนัยสำคัญทางสถิติ 0.20 ($P_R = 0.20$)

3. นำตัวแปรอิสระที่ผ่านการคัดเลือกทั้งหมดจากชุดข้อมูลฝึกสอนมาสร้างตัวแบบทางสถิติและทำการทำนายค่าจำนวนการเรียกค่าสินไหมทดแทน รวมทั้งนำตัวแบบที่ได้จากชุดข้อมูลฝึกสอนมาทำนายจำนวนการเรียกค่าสินไหมทดแทนในชุดข้อมูลทดสอบ นอกจากนี้ให้นำตัวแปรอิสระที่ผ่านการคัดเลือกจากข้อมูลฝึกสอนมาทำการสร้างตัวแบบในชุดข้อมูลทั้งหมดและทำนายจำนวนการเรียกค่าสินไหมทดแทนเช่นกัน ค่าพารามิเตอร์ของการทำนายถูกต้องในชุดข้อมูลฝึกสอน ข้อมูลทดสอบ และข้อมูลทั้งหมดว่ามีค่าใกล้เคียงกันหรือไม่ ถ้ามีค่าใกล้เคียงกันแสดงว่าสามารถนำตัวแบบนี้ไปเป็นตัวแทนในการวิเคราะห์ข้อมูลทั้งหมดได้

4. ตรวจสอบความเหมาะสมของตัวแบบที่ได้พัฒนาจากวิธี GEE และวิธี GLMMs เพื่อหาตัวแบบสุดท้ายที่เหมาะสมสำหรับจำนวนการเรียกค่าสินไหมทดแทน

โดยมีรายละเอียดการสรุปผลการวิจัยเพื่อตอบวัตถุประสงค์ในการศึกษาดังนี้

1. ปัจจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนของตัวแบบวิธี GEE ที่มีโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) และตัวแบบวิธี GLMMs ที่มีโครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) ซึ่งเป็นตัวแบบที่เหมาะสมสำหรับข้อมูลทั้งหมด สามารถสรุปได้ดังตารางที่ 25

ตารางที่ 25 ปัจจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนของตัวแบบแต่ละวิธี

ตัวแปรอิสระ	ข้อมูลทั้งหมด	
	วิธี GEE	วิธี GLMMs
	AR(1)	AR(1)
ค่าสินไหมทดแทน (TFMONEY)	✓	✓
อายุรถยนต์ (CAR-AGE)	✓	✓
กลุ่มอายุ 18-24 ปี (DA1)	-	-
กลุ่มอายุ 25-35 ปี (DA2)	✓	✓
กลุ่มอายุ 36-50 ปี (DA3)	✓	✓
อาชีพ (OCCUP)	✓	✓
เพศ (GENDER)	-	✓

หมายเหตุ ✓ คือ มีนัยสำคัญทางสถิติที่ระดับ 0.05

2. การพัฒนาตัวแบบด้วยวิธี GEE เมื่อกำหนดโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) พบว่ามีความเหมาะสมกับข้อมูล โดยมีปัจจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนที่ระดับนัยสำคัญทางสถิติ 0.05 คือ ค่าสินไหมทดแทนที่แปลงค่า อายุรถยนต์ อายุผู้ทำประกันภัย 25-35 ปี อายุผู้ทำประกันภัย 36-50 ปี และอาชีพ สรุปเป็นสมการของตัวแบบได้ดังนี้

$$\begin{aligned} \log_e y = & -1.8074 + 0.1464(\text{ค่าสินไหมทดแทนที่แปลงค่า}) * + 0.0452(\text{อายุรถยนต์}) * \\ & + 0.1113(\text{กลุ่มอายุ 18-24 ปี}) + 0.1656(\text{กลุ่มอายุ 25-35 ปี}) * \\ & + 0.0787(\text{กลุ่มอายุ 36-50 ปี}) * + 0.3488 (\text{อาชีพ}) * + (-0.0366) (\text{เพศ}) \end{aligned}$$

(* มีนัยสำคัญทางสถิติที่ระดับ 0.05)

ซึ่งความเหมาะสมของตัวแบบนี้พิจารณาได้จากค่าสถิติ Pearson Chi-Square of residual /DF เท่ากับ 0.64 มีค่าไม่เกิน 1 และมีค่าร้อยละของการทำนายถูกต้องของจำนวนการเรียกค่าสินไหม

ทดแทนเท่ากับ 52.32 สำหรับตัวแบบในภาพรวมของจำนวนการเรียกค่าสินไหมทดแทนโดยเฉลี่ยของประชากร รวมทั้งการพล็อตค่า Residuals กับค่าทำนายบนเส้นถดถอย (Predicted) หรือ $\hat{Y}$ ในระยะเวลาติดตาม 5 ปี ของการทำประกันภัยรถยนต์พบว่า การกระจายของข้อมูลมีความสัมพันธ์เป็นแบบเส้นโค้ง หรืออยู่ในรูปแบบเอกซ์โพเนนเชียล ซึ่งตรงตามกับลักษณะของตัวแบบที่ตัวแปรอยู่ในรูปของลอการิทึม ดังนั้นจึงสรุปได้ว่า ตัวแบบวิธี GEE เมื่อกำหนดโครงสร้างความสัมพันธ์ของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) มีความเหมาะสมกับข้อมูลที่ศึกษานี้

3. การพัฒนาตัวแบบด้วยวิธี GLMMs เมื่อกำหนดโครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) พบว่ามีความเหมาะสมกับข้อมูล โดยมีปัจจัยที่มีผลต่อจำนวนการเรียกค่าสินไหมทดแทนที่ระดับนัยสำคัญทางสถิติ 0.05 คือ ค่าสินไหมทดแทนที่แปลงค่า อายุรถยนต์ อายุผู้ทำประกันภัย 25-35 ปี อายุผู้ทำประกันภัย 36-50 ปี อาชีพ และเพศ สรุปสมการของตัวแบบได้ดังนี้

3.1 ตัวแบบสำหรับภาพรวมของค่าเฉลี่ยประชากร

$$\begin{aligned} \log_e y = & -2.7130 + 0.1805(\text{ค่าสินไหมทดแทนที่แปลงค่า}) * + 0.0861(\text{อายุรถยนต์}) * \\ & + 0.1390(\text{กลุ่มอายุ 18-24 ปี}) + 0.1703(\text{กลุ่มอายุ 25-35 ปี}) * \\ & + 0.0744(\text{กลุ่มอายุ 36-50 ปี}) * + 0.3105 (\text{อาชีพ}) * + (-0.0487) (\text{เพศ}) * \end{aligned}$$

(* มีนัยสำคัญทางสถิติที่ระดับ 0.05)

3.2 ตัวแบบสำหรับแต่ละหน่วยศึกษาเมื่อมีเทอมอิทธิพลค่าคงที่สุ่ม

$$\begin{aligned} \log_e y = & (-2.7130 + \text{Random Subject Effect}) + 0.1805(\text{ค่าสินไหมทดแทนที่แปลงค่า}) * \\ & + 0.0861(\text{อายุรถยนต์}) * + 0.1390(\text{กลุ่มอายุ 18-24 ปี}) + 0.1703(\text{กลุ่มอายุ 25-35 ปี}) * \\ & + 0.0744(\text{กลุ่มอายุ 36-50 ปี}) * + 0.3105 (\text{อาชีพ}) * + (-0.0487) (\text{เพศ}) * \end{aligned}$$

(* มีนัยสำคัญทางสถิติที่ระดับ 0.05)

ซึ่งความเหมาะสมของตัวแบบนี้พิจารณาได้จากค่าสถิติ Generalized Chi-Square/DF เท่ากับ 0.18 มีค่าไม่เกิน 1 และค่า Pseudo-AIC และ Pseudo-BIC มีค่าต่ำกว่ากรณีที่มีการเลือกใช้โครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ CS แสดงว่าการใช้โครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนรูปแบบ AR(1) มีความเหมาะสมกับข้อมูลมากกว่า รวมทั้งการคำนวณค่าร้อยละของการทำนายถูกต้องของจำนวนการเรียกค่าสินไหมทดแทนในกรณีภาพรวมของค่าเฉลี่ยประชากรมีค่าเท่ากับ 53.34 และค่าร้อยละของการทำนายถูกต้องของจำนวนการเรียกค่าสินไหมทดแทนกรณีที่มีค่าอิทธิพลคงที่สุ่ม (Random Intercept) ในตัวแบบ มีค่าเท่ากับ 76.28 นอกจากนี้เมื่อทำการพล็อตกราฟแสดงความสัมพันธ์ระหว่าง Residuals กับ $\hat{Y}$ (Predicted) แสดงให้เห็นว่าข้อมูลมีการกระจายในรูปแบบเอกซ์โพเนนเชียล ซึ่งตรงกับสมการของตัวแบบที่ประมาณค่าได้ เนื่องจากตัวแปรตาม คือ จำนวนการเรียกค่าสินไหมทดแทนมีการแจกแจงแบบปัวซอง

สรุป ตัวแบบวิธี GLMMs เมื่อกำหนดโครงสร้างความแปรปรวนร่วมของจำนวนการเรียกค่าสินไหมทดแทนเป็นแบบ AR(1) มีความเหมาะสมสำหรับใช้เป็นตัวแบบของข้อมูลที่ศึกษานี้ โดยสามารถอธิบายการเปลี่ยนแปลงของจำนวนการเรียกค่าสินไหมทดแทนได้ทั้งภาพรวมค่าเฉลี่ยประชากรและแต่ละหน่วยศึกษาของผู้ทำประกันภัยรถยนต์

ข้อเสนอแนะ

1. ให้ตัดตัวแปรอิสระค่าสินไหมทดแทนออกไป เพื่อหลีกเลี่ยงปัญหาข้อมูลมีการกระจายแบบเบ้ขวา (Jong and Heller, 2005) หรือทำการแปลงข้อมูลค่าสินไหมทดแทนด้วยวิธีอื่น ๆ เช่น การแปลงข้อมูลค่าสินไหมทดแทนให้มีการแจกแจงแบบแกมมา หรือเพิ่มตัวแปรอิสระเกี่ยวกับพฤติกรรมการขับขี่ การใช้รถ โดยอาจจะต้องทำแบบสอบถามเพิ่มเติม หรือขอข้อมูลจากบริษัทประกันภัยซึ่งจะมีส่วนช่วยให้การสร้างตัวแบบทำนายจำนวนการเรียกค่าสินไหมทดแทนมีความถูกต้องมากขึ้น และเหมาะสมกับความเป็นจริง

2. การพัฒนาตัวแบบสำหรับข้อมูลที่มีการแบ่งข้อมูลตามหลักวิธี Cross-Validation ควรทำการแบ่งข้อมูลฝึกสอนและข้อมูลทดสอบในสัดส่วนที่แตกต่างกันออกไป เช่น ข้อมูลฝึกสอนร้อยละ 80 และข้อมูลทดสอบร้อยละ 20 , ข้อมูลฝึกสอนร้อยละ 60 และข้อมูลทดสอบร้อยละ 40 และ ข้อมูลฝึกสอนร้อยละ 50 และข้อมูลทดสอบร้อยละ 50 เพื่อสามารถเปรียบเทียบผลที่ได้จากการแบ่ง

ในสัดส่วนที่แตกต่างกันว่าให้ผลการศึกษาแตกต่างกันหรือไม่ ซึ่งในการศึกษานี้ได้ทำการแบ่งข้อมูลเพียงสัดส่วนเดียวคือ ข้อมูลฝึกสอนร้อยละ 70 และข้อมูลทดสอบร้อยละ 30 จึงไม่สามารถตรวจสอบความแตกต่างได้

3. ตัวแบบทางสถิติวิธี GEE และวิธี GLMMs ที่นำมาใช้กับข้อมูลการประกันภัยรถยนต์ ให้ค่าทำนายความถูกต้องในภาพรวมของจำนวนค่าสินไหมทดแทนได้ใกล้เคียงกันประมาณร้อยละ 52 หากมีข้อมูลเกี่ยวกับพฤติกรรมการใช้รถที่แท้จริงจะเป็นประโยชน์ต่อการสร้างตัวแบบที่ให้ค่าทำนายถูกต้องมากขึ้น นอกจากนี้ผลการศึกษาที่ได้สามารถนำไปประยุกต์ใช้กับข้อมูลการประกันภัยและธุรกิจประกันภัยรถยนต์ได้ในกรณีที่ต้องการทำนายจำนวนการเรียกค่าสินไหมทดแทนในภาพรวมของค่าเฉลี่ยประชากรจากตัวแบบวิธี GEE และหากสนใจวิเคราะห์ในรายบุคคลสามารถใช้ตัวแบบวิธี GLMMs โดยเพิ่มตัวแปรอิสระที่สามารถอธิบายจำนวนการเรียกค่าสินไหมทดแทนในการประกันภัยรถยนต์ได้ รวมทั้งควรเพิ่มการพิจารณาตัวแปรอิสระในรูปแบบเทอมอิทธิพลร่วม (Interaction Effect) ในการวิเคราะห์ตัวแบบ

4. ในการกำหนดเทอมอิทธิพลร่วมเพื่อศึกษาความแตกต่างของแต่ละบุคคลในหน่วยศึกษานั้นเราสามารถกำหนดให้ปัจจัยอื่น ๆ เป็นเทอมอิทธิพลร่วมได้เช่นเดียวกันไม่จำเป็นต้องกำหนดให้มีเฉพาะค่าอิทธิพลคงที่ร่วมเท่านั้น

เอกสารและสิ่งอ้างอิง

- กันทิมา ศิริพาณิชย์. 2545. การวิเคราะห์ปัจจัยที่มีผลต่อความเสียหายเพื่อการกำหนดอัตราดอกเบี้ยประกันภัยรถยนต์ในประเทศไทย. วิทยานิพนธ์ปริญญาโท, จุฬาลงกรณ์มหาวิทยาลัย.
- กานธนิกา ชูณหะวัด. 2545. การประมาณค่าพารามิเตอร์สำหรับตัวแบบถดถอยเชิงเส้นปัวส์ซง. วิทยานิพนธ์ปริญญาโท, จุฬาลงกรณ์มหาวิทยาลัย.
- กัลยา วานิชย์บัญชา. 2549. การใช้ SPSS for Windows ในการวิเคราะห์ข้อมูล. ครั้งที่ 9. บริษัทธรรมสาร จำกัด, กรุงเทพฯ.
- กรมการประกันภัย. 2550. ประเภทของการประกันภัยรถยนต์. แหล่งที่มา: <http://www.doi.go.th>, พฤษภาคม 2550.
- จารุณี ชินะพงศ์ไพศาล. 2545. ความพึงพอใจของผู้เอาประกันที่มีต่อบริษัทประกันภัย: ศึกษาเฉพาะกรณีการประกันภัยรถยนต์ส่วนบุคคลภาคสมัครใจในเขตกรุงเทพมหานคร. วิทยานิพนธ์ปริญญาโท, จุฬาลงกรณ์มหาวิทยาลัย.
- นันทิยา วุฒิมงคลชัย. 2545. การศึกษาเปรียบเทียบต้นทุนและผลประโยชน์ของการประกันภัยรถยนต์ภาคสมัครใจ. วิทยานิพนธ์ปริญญาโท, มหาวิทยาลัยเกษตรศาสตร์.
- ปัทมา อินทรชิต. 2547. ปัจจัยที่มีผลต่ออุปสงค์และแนวโน้มของการประกันภัยรถยนต์ภาคสมัครใจ. วิทยานิพนธ์ปริญญาโท, มหาวิทยาลัยเกษตรศาสตร์.
- ประทีป ประดิษฐ์วรคุณ. 2540. การวิเคราะห์เชิงสถิติเพื่อหาอัตราเบี้ยประกันภัยรถยนต์ที่เหมาะสม. วิทยานิพนธ์ปริญญาโท, จุฬาลงกรณ์มหาวิทยาลัย.
- วิรัช พานิชวงค์. 2549. การวิเคราะห์การถดถอย. ครั้งที่ 4. ศูนย์ผลิตตำราเรียน สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, กรุงเทพฯ.

- วีรนนท์ พงศาภักดี. 2537. การวิเคราะห์ข้อมูลเชิงกลุ่ม: ทฤษฎีและการประยุกต์. ครั้งที่ 2. ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร, กรุงเทพฯ.
- สมภพ วงศ์วิชิต. 2545. การวิเคราะห์อุปสงค์ของการประกันภัยรถยนต์ภาคสมัครใจในประเทศไทย. วิทยานิพนธ์ปริญญาโท, มหาวิทยาลัยธุรกิจบัณฑิตย์.
- สิริกุล สันต์วานิช. 2544. ความคิดเห็นของผู้เอาประกันภัยที่มีต่อการประกันภัยรถยนต์แบบใหม่. การศึกษาค้นคว้าด้วยตัวเอง, มหาวิทยาลัยเกษตรศาสตร์.
- อดิชาติ โพธิเกษม. 2548. ปัจจัยที่มีผลต่อการตัดสินใจทำประกันภัยรถยนต์ภาคสมัครใจ. วิทยานิพนธ์ปริญญาโท, มหาวิทยาลัยเกษตรศาสตร์.
- อรรณพ ไสระเวช, สมศักดิ์ พงศ์พันธุ์, กิรติกา แพงลาด, ธนิสรา สมรรถพันธุ์ และ ไตรรักษ์ คุรุเวช. 2545. ผลงานวิจัยเรื่องปัจจัยสำคัญที่มีอิทธิพลต่อการตัดสินใจทำประกันภัยรถยนต์. วิทยานิพนธ์ปริญญาโท, สถาบันบัณฑิตพัฒนบริหารศาสตร์.
- Breslow, N.E. and D.G. Clayton. 1993. Approximate Inference in Generalized Linear Mixed Model. **Journal of the American Statistical Association** 88: 9-25.
- Brown H, and R. Prescott. 1999. **Applied Mixed Models in Medicine**. John Wiley & Sons, Ltd., New York: 261-294.
- Cook, R.D. and S. Weisberg. 1999. **Applied Regression Including Computing and Graphics**. Hoboken NJ: John Wiley,
- Denuit, M. 1997. A new distribution of Poisson type for the number of claims. **ASTN Bulletin** 27: 229-242.

- Diggle, P., P. Heagerty, K.Y. Liang and S.L. Zeger. 2002. **Analysis of Longitudinal Data**. 2nd edition. Oxford: Oxford University Press.
- Fitzmaurice, G.M., N.M. Laird and J.H. Ware. 2004. **Applied Longitudinal Analysis**. John Wiley & Sons, Inc., New Jersey.
- Gordon, J. and S. Maura. **Applications of GEE Methodology Using the SAS System**. Cary, NC. SAS Institute Inc.
- Guerrero, V.M. and R.A. Johnson. 1982. Use of the Box-Cox transformation with binary response models. **Biometrika** 62 : 309-14.
- Hosmer, D.W. and S. Lemeshow. 2000. **Applied Logistic Regression**. Wiley Chichester, UK.
- P.de Jong and G.Z.Heller. 2008. **Generalized Linear Models for Insurance Data**. 1 ed. Cambridge University Press, United Kingdom.
- Kutner, M.H., C.J. Nachtsheim and J. Neter. 2004. **Applied Linear Regression Models**. 4th edition. The McGraw-Hill Companies, Inc., New York.
- Liang, K.Y. and S.L. Zeger. 1986. Longitudinal data analysis using generalized linear Models. **Biometrika** 73: 13-22.
- Lindstrom, M.J. and Bates, D.M. 1990. Nonlinear Mixed Effects Models for Repeated Measures Data. **Biometrics** 46: 673-687.
- Littell, R.C., J. Pendergast and R. Natarajan. 2000. Modelling covariance structure in the analysis of repeated measures data. **Statistics in Medicine** 19: 1793-1819.

- McCullagh, P. and J. A. Nelder. 1989. **Generalized Linear Models** (Second Edition). Chapman and Hall, London.
- McCulloch, C.E. and S.R. Searle. 2001. **Generalized, Linear and Mixed Models**. John Wiley & Sons, Inc., Canada.
- McGilchrist, C.A. 1994. Estimation in Generalized Mixed Models. **Journal of the Royal Statistical Society** 1: 61-69.
- Myers, R.H., D.C. Montgomery and G.G. Vining. 2002. **Generalized Linear Models With Applications in Engineering and the Sciences**. John Wiley & Sons, Inc., New York.
- Nelder, J.A. and R.W.M. Wedderburn. 1972. Generalized Linear Model. **J. Roy. Statist Soc A** 135: 370-384.
- Oliver S. 2006. **Introducing the GLIMMIX Procedure for Generalized Linear Mixed Models**. Cary, NC. SAS Institute Inc.
- SAS Institute Inc. 1999. **SAS/STAT User's Guide, Version 8**. Cary, NC: SAS Institute Inc.
- SAS Institute Inc. 2006. **The GLIMMIX Procedure**. Cary, NC: SAS Institute Inc.
- Shoukri, M.M. and M.A. Chaudhary. 2007. **Analysis of Correlated Data with SAS and R**. 3 rd. ed. Taylor & Francis Group, LLC., The United States of America.
- Stroinski, K.J. and I.D. Currie. 1989. Selection of variables for automobile insurance rating. **Mathematics and Economics** 8: 35-46.

West, B.T., K.B. Welch and A.T. Galecki. 2007. **Linear Mixed Models: A Practical Guide Using Statistical Software**. Taylor & Francis Group, LLC., The United States of America.

Wolfinger, R. and M. O'Connell. 1993. Generalized linear mixed models: A pseudo-likelihood approach. **Journal of Statistical Computation and Simulation** 48: 233-243.

Yau, K.K.W., K.C.H. Yip and H.K. Yuen. 2003. Modelling repeated insurance claim frequency data using the generalized linear mixed model. **Journal of Applied Statistic** 30: 857-865.

ภาคผนวก

ตารางผนวกที่ 1 การแบ่งกลุ่มรถยนต์ (เฉพาะรถยนต์นั่ง รหัส 110 120 เท่านั้น)

กลุ่มที่ 1		กลุ่มที่ 2		กลุ่มที่ 3		กลุ่มที่ 4		กลุ่มที่ 5	
ชื่อรถ	รุ่นรถ	ชื่อรถ	รุ่นรถ	ชื่อรถ	รุ่นรถ	ชื่อรถ	รุ่นรถ	ชื่อรถ	รุ่นรถ
ASTON	MARTIN	ALFA	ROMEO	CHEVROLET	ZAFIRA	MAZDA	323	DAIHATSU	MIRA
BENTLEY		AUDI		DAIHATSU	GRAN	MITSUBISHI	CHAMP	HONDA	CITY
CADILLAC		BMW			TERIOS	TOYOTA	COROLLA	MAZDA	FAMILIA
CHEVROLET		CHRYSLER		FIAT			SPORT	NISSAN	NV
							RIDER		
DAIMLER		CITROEN		FORD					SUNNY
FERRARI		HOLDEN		HONDA	ACCORD	TOYOTA	CKD.	SUZUKI	CARRIBIAN
HONDA	LEGEND	ISUZU	TROOPER		CIVIC	NISSAN	CKD.	TOYOTA	SOLUNA
	NSX	JEEP			CRV	MITSUBISHI	CKD.		
		CHEROKEE							
	ODYSSEY	LAND ROVER		HYUNDAI					
	PRELUDE	LEXUS		ISUZU	CAMEO				
JAGUAR		MAGSO			RODEO				
MASERATI		MAZDA	MX5		VEGA				
MG		MERCEDES-BENZ			VERTEX				
NISSAN	PRESIDENT	MITSUBISHI	PAJERO	KIA					
	INFINITY	OPEL	CALIBRA	MAZDA	121				
PORSCHE			OMAGA		626				
ROLLS-ROYCE			VECTRA		ASTINA				
TOYOTA	SUPRA	PEUGEOT			CRONOS				
		ROVER			LANTIS				
HONDA	CBU.	SAAB		MITSUBISHI	GALANT				
		TOYOTA	CELICA	NISSAN	CEFIRO				
			CROWN		PRIMERA				
			LAND CRUISER		200SX				
			RAV 4		121				
		VOLKSWAGEN		OPEL	ASTRA				
		VOLVO			CORSA				
		MAZDA	CBU.	PROTON	SEGA				
		TOYOTA	CBU.	SEAT					
		NISSAN	CBU.	SUBARU	IMPREZA				
		MITSUBISHI	CBU.		LEGACY				
				SUZUKI	ESTEEM				
					VITARA				
				TOYOTA	CAMRY				
					CORONA				
					STARLET				
				MAZDA	CKD.				
				HONDA	CKD.				

ตารางผนวกที่ 2 ลักษณะข้อมูลการจัดกลุ่มของตัวแปรค่าสินไหมทดแทนและอายุรณย์ตั้งแต่
ปี พ.ศ. 2544-2548

ตัวแปร	จำนวน (n=18175)*	ร้อยละ
กลุ่มค่าสินไหมทดแทน		
ไม่มีค่าสินไหมทดแทน	5,540	30.5
ตั้งแต่ 1-7792.52 บาท	3,549	19.5
ตั้งแต่ 7792.53 บาทขึ้นไป	9,086	50.0
กลุ่มอายุรณย์		
อายุรณย์ 0 ปี	225	1.2
อายุรณย์ 1-4 ปี	10,395	57.2
อายุรณย์ตั้งแต่ 5 ปีขึ้นไป	7,555	41.6

หมายเหตุ * จำนวนผู้ทำประกันภัยรณย์ทั้งหมด 3,635 คน โดยรวมระยะเวลาทั้งหมด 5 ปี

ประวัติการศึกษา และการทำงาน

ชื่อ –นามสกุล

นางสาวศิริกัญญา ชีระอนันต์ชัย

วัน เดือน ปี ที่เกิด

วันที่ 14 กันยายน 2526

สถานที่เกิด

อำเภอบ้านโป่ง จังหวัดราชบุรี

ประวัติการศึกษา

วท.บ.(สถิติ) มหาวิทยาลัยธรรมศาสตร์ (พ.ศ.2547)