



การประมาณเบี้ยประกันภัยที่แท้จริงของการประกันภัยรถยนต์โดยใช้ตัวแบบเชิงเส้นวางนัยทั่วไป

โดย

นายวัชรวิทย์ กิตติพัฒน์	เลขทะเบียน 5809610859
นายจารุวิทย์ เกิดพระ	เลขทะเบียน 5809610958
นายรัชภูมิ ขมาวิมล	เลขทะเบียน 5809680456

รายงานนี้เป็นส่วนหนึ่งของการศึกษาวิชา ส.495 โครงการพิเศษ 2

ตามหลักสูตรวิทยาศาสตรบัณฑิต

สาขาสถิติ ภาควิชาคณิตศาสตร์และสถิติ

คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

ปีการศึกษา 2561

แบบฟอร์มการส่งรายงานฉบับสมบูรณ์
วิชา ส.495 โครงงานพิเศษ 2 ปีการศึกษา 2561
กลุ่มที่ N14

แบบฟอร์ม ข4

รายชื่อสมาชิกในกลุ่ม

- 1.ชื่อ-นามสกุล นายวัชรวิทย์ กิตติพัฒน์ เลขทะเบียน 5809610859 โทรศัพท์ติดต่อ 090-0691720
- 2.ชื่อ-นามสกุล นายจารุวิทย์ เกิดพระ เลขทะเบียน 5809610958 โทรศัพท์ติดต่อ 093-5349894
- 3.ชื่อ-นามสกุล นายรัชภูมิ ษมาวิมล เลขทะเบียน 5809680456 โทรศัพท์ติดต่อ 092-2847201

อาจารย์ที่ปรึกษา ผศ.ดร.แสงหล้า ชัยมงคล

อาจารย์ที่ปรึกษาร่วม คุณวรสิทธิ์ วิจิตรนาการ

อาจารย์ที่ร่วมเป็นกรรมการสอบ

1. ผศ.สุเมศวร์ ฮาซิม
2. ผศ.ดร.แสงดาว วงศ์สาย

หัวข้อโครงงาน

(ภาษาไทย) การประมาณเบี้ยประกันภัยที่แท้จริงของการประกันภัยรถยนต์โดยใช้ตัวแบบเชิงเส้นวางนัยทั่วไป

(ภาษาอังกฤษ) Estimating Pure Premium for Automobile Insurance Using Generalized Linear Models

อาจารย์ที่ปรึกษา.....แสงหล้า ชัยมงคล
(ผศ.ดร.แสงหล้า ชัยมงคล)
นักศึกษา.....วัชรวิทย์ กิตติพัฒน์
(นายวัชรวิทย์ กิตติพัฒน์)
นักศึกษา.....จารุวิทย์ เกิดพระ
(นายจารุวิทย์ เกิดพระ)
นักศึกษา.....รัชภูมิ ษมาวิมล
(นายรัชภูมิ ษมาวิมล)
วันที่.....๑๘ ม.๓. ๒๕๖๒

หัวข้องานวิจัย	การประมาณเบี้ยประกันภัยที่แท้จริงของการประกันภัยรถยนต์โดยใช้ตัวแบบเชิงเส้นวงนัยทั่วไป
คณะผู้จัดทำ	นาย วัชรวิทย์ กิตติพัฒน์ นาย จารุวิทย์ เกิดพระ นาย รัชภูมิ ยมาวิมล
ชื่อปริญญา	วิทยาศาสตร์บัณฑิต
สาขา/คณะ/มหาวิทยาลัย	สาขาสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
อาจารย์ที่ปรึกษางานวิจัย	ผู้ช่วยศาสตราจารย์ ดร.แสงหล้า ชัยมงคล
อาจารย์ที่ปรึกษาร่วม	คุณวรสิทธิ์ ฐิตินการ
ปีการศึกษา	2561

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อประยุกต์ใช้ตัวแบบเชิงเส้นวงนัยทั่วไปในการประมาณค่าเบี้ยประกันภัยที่แท้จริงในการประกันภัยรถยนต์ โดยใช้ข้อมูลสถิติภูมิที่ได้จากฐานข้อมูลกรมธรรม์ประกันอุบัติเหตุรถยนต์ภาคสมัครใจ ชั้น 1 ประเภทรถยนต์นั่ง (รหัส 110 และ 120) ของสำนักงานอตราเบี้ยประกันวินาศภัยที่เริ่มสัญญาในปี พ.ศ.2559 และ 2560 ซึ่งการสร้างตัวแบบเบี้ยประกันที่แท้จริงจะประกอบไปด้วยตัวแบบความถี่ในการเรียกร้องค่าสินไหมทดแทน (Claim Frequency Model) และ ตัวแบบความรุนแรงในการเรียกร้องค่าสินไหมทดแทน (Claim Severity Model)

ผลการวิเคราะห์ Claim Frequency Model พบว่ามีปัจจัยเสี่ยง อายุผู้ขับขี่ (Age), อายุการใช้งานรถยนต์ (VehicleAge) และ ขนาดเครื่องยนต์ (VehicleSizeCode) ที่มีผลกระทบต่อความถี่ในการเรียกร้องค่าสินไหมทดแทน ที่ระดับนัยสำคัญ 0.05

ผลการวิเคราะห์ Claim Severity Model พบว่ามีปัจจัยเสี่ยง อายุผู้ขับขี่ (Age), อายุการใช้งานรถยนต์ (VehicleAge), ขนาดเครื่องยนต์ (VehicleSizeCode), กลุ่มรถยนต์ (VehicleGroupType) และ ทุนเอาประกัน (BandValue) ที่มีผลกระทบต่อความรุนแรงในการเรียกร้องค่าสินไหมทดแทน ที่ระดับนัยสำคัญ 0.05

ดังนั้นการประมาณเบี้ยประกันภัยที่แท้จริงควรคำนึงถึงปัจจัยเสี่ยงดังกล่าว ซึ่งผลการศึกษาจะเป็นแนวทางให้กับบริษัทประกันภัยนำไปประยุกต์ใช้กับข้อมูลที่รับความเสี่ยงเพื่อสร้างตัวแบบการประมาณเบี้ยประกันภัยที่แท้จริง

คำสำคัญ: เบี้ยประกันภัยที่แท้จริง ตัวแบบเชิงเส้นวางนัยทั่วไป ปัจจัยเสี่ยง ตัวแบบความถี่ในการเรียกร้องค่าสินไหมทดแทน ตัวแบบความรุนแรงในการเรียกร้องค่าสินไหมทดแทน

Independent Study Title	Estimating Pure Premium for Automobile Insurance Using Generalized Linear Models
Authors	Mister Watcharawit Kitipipat Mister Jaruwit Kerdphra Mister Ratchaphoom Samawimon
Degree	Bachelor of Science
Major Field/Faculty/University	Statistics Faculty of Science and Technology Thammasat University
Independent Study Advisor	Assistant Professor Saengla Chaimongkol, Ph.D.
Co-Advisor	Mister Worasit Thititanakan, ASAT.
Academic Year	2018

ABSTRACT

The objective of this study was to apply generalized linear models to estimate pure premium for automobile insurance. In this study, the data set was the 1st class voluntary insurance policies (code 110 and 120), which was provided by the Insurance Premium Rating Bureau (IPRB) that started the contract in 2016 and 2017. The model of pure premium is composed of the Claim Frequency Model and the Claim Severity Model.

For Claim Frequency Model, the result of this study show that the risk factors significantly related to Claim Frequency at the 0.05 level consisted of Age, VehicleAge and VehicleSizeCode.

For Claim Severity Model, the result of this study show that the risk factors significantly related to Claim Severity at the 0.05 level consisted of Age, VehicleAge ,VehicleSizeCode, VehicleGroup and BandValue.

Therefore, to estimate pure premium taking care of the risk factors in this study will guide the insurance company to apply to the risk-sensitive information in order to create a pure premium estimate.

Keywords: Pure Premium, Generalized Linear Models, Risk Factor, Claim Frequency Model, Claim Severity Model

กิตติกรรมประกาศ

งานวิจัยฉบับนี้สำเร็จลุล่วงได้ด้วยความกรุณาของผู้ช่วยศาสตราจารย์ ดร.แสงหล้า ชัยมงคล อาจารย์ที่ปรึกษางานวิจัย คณะผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูง ที่กรุณาสละเวลาให้คำปรึกษา คำแนะนำ ตลอดจนดูแล ตรวจสอบและแก้ไขข้อบกพร่อง เพื่อให้งานวิจัยฉบับนี้สำเร็จลุล่วงด้วยดี ทั้งนี้ด้วยความอนุเคราะห์จากสำนักงานอตราเบียบประกันวินาศภัย และ คุณวรสิทธิ์ จูติชนการ อาจารย์ที่ปรึกษาร่วม ที่อนุญาตให้นำข้อมูลการประกันภัยรถยนต์มาใช้ในการศึกษางานวิจัยในครั้งนี้

กราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ สุรเมศวร์ อาซิม และ ผู้ช่วยศาสตราจารย์ ดร.แสงดาว วงศ์สาย กรรมการสอบงานวิจัยในครั้งนี้ รวมทั้งคณาจารย์ ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ที่ได้ให้ความรู้แก่คณะผู้วิจัย และขอขอบคุณเจ้าหน้าที่ประจำภาควิชาคณิตศาสตร์และสถิติ ที่ได้ให้ความช่วยเหลือ ประสานงานและอำนวยความสะดวกให้

สุดท้ายนี้ ขอกราบขอบพระคุณ บิดา มารดา ที่คอยอบรมสั่งสอนและคอยให้กำลังใจที่ดีเสมอมา ขอขอบคุณ พี่ น้อง และเพื่อน ๆ ทุกคนที่คอยสนับสนุนและให้ความช่วยเหลือ คำปรึกษา และคำแนะนำต่าง ๆ เสมอมา

คณะผู้วิจัย

นาย วัชรวิทย์ กิตติพัฒน์

นาย จารุวิทย์ เกิดพระ

นาย รัชภูมิ ขมาวิมล

สารบัญ

หน้า

บทคัดย่อ	(1)
ABSTRACT	(3)
กิตติกรรมประกาศ	(5)
บทที่ 1 บทนำ	1
1.1 ที่มาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของการวิจัย	2
1.3 ขอบเขตของการวิจัย	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ	6
บทที่ 2 ทฤษฎี และงานวิจัยที่เกี่ยวข้อง	7
2.1 วิธีการทางสถิติ	7
2.1.1 ตัวแบบเชิงเส้นวงนัยทั่วไป (Generalized Linear Model or GLMs)	7
2.1.1.1 ส่วนประกอบของตัวแบบเชิงเส้นวงนัยทั่วไป (Component of GLMs)	8
2.1.1.2 Offset Term	10
2.1.1.3 Prior Weight	11
2.1.1.4 ระดับพื้นฐาน (Base Level)	11
2.1.1.5 การอนุมานตัวแบบเชิงเส้นวงนัยทั่วไป (Inference for Generalized Linear Models)	12
2.1.1.6 Deviance และภาวะสารูปดีของตัวแบบ GLMs (Deviance and Goodness of fit of GLMs)	12
2.1.1.7 การเลือกตัวแบบและการตรวจสอบความเหมาะสมของตัวแบบ	12
2.1.1.8 Deviance สำหรับตัวแบบปัวซอง (Deviance for Poisson models)	13
2.1.1.9 ส่วนเหลือดีวีเนียนส์ (Deviance Residual or Γ_D)	13
2.1.2 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Square Error)	14
2.1.3 ตัวแบบเบี่ยงแปรกันภัยที่แท้จริง	15

2.2 คำศัพท์ประกันภัย	16
2.2.1 Insurance	16
2.2.2 Claim Frequency	16
2.2.3 Claim Count	16
2.2.4 Earned Exposure	16
2.2.5 Claim Severity	16
2.2.6 Claim amount	17
2.2.7 Pure Premium	17
2.3 งานวิจัยที่เกี่ยวข้อง	18
2.3.1 ตัวแบบเชิงเส้นวางนัยทั่วไปกับการประกันภัย	18
2.3.2 การประมาณเบี่ยงประกันภัยที่แท้จริง	18
2.3.3 การประมาณเบี่ยงประกันภัยที่แท้จริงโดยใช้ตัวแบบเชิงเส้นวางนัยทั่วไป	18
บทที่ 3 วิธีการดำเนินงานวิจัย	20
3.1 เครื่องมือที่ใช้ในการดำเนินงานวิจัย	20
3.2 ภาพรวมของกระบวนการสร้างตัวแบบเชิงเส้นวางนัยทั่วไป	20
3.3 การเตรียมและตรวจสอบความถูกต้องของข้อมูล	21
3.3.1 การตรวจสอบความถูกต้อง	22
3.3.2 การเตรียมข้อมูล	22
3.4 การสร้างตัวแบบเชิงเส้นวางนัยทั่วไป	23
3.5 การวิเคราะห์ผลที่ได้จากตัวแบบ	24
3.6 การรวมตัวแบบเพื่อหาเบี่ยงประกันภัยแท้จริง	24
3.7 การวัดความเหมาะสมของตัวแบบ	24
3.8 การสรุปผล	25
3.8.1 ลักษณะทั่วไปของข้อมูล	25
3.8.2 ผลการวิเคราะห์ตัวแบบของกรมธรรม์ที่เริ่มสัญญาในปี 2559	25

3.8.3 ผลการวิเคราะห์ตัวแบบของกรรมธรรม์ที่เริ่มสัญญาในปี 2560	25
3.8.4 ผลการวัดความเหมาะสมของตัวแบบ	25
บทที่ 4 ผลการดำเนินงานวิจัย	26
4.1 ลักษณะทั่วไปของข้อมูล	26
4.2 ผลการวิเคราะห์ตัวแบบของกรรมธรรม์ที่เริ่มสัญญาในปี 2559	29
4.2.1 ผลการวิเคราะห์ตัวแบบความถี่ในการเรียกร้องค่าสินไหมทดแทน (Claim Frequency Model) ของปี 2559	29
4.2.2 ผลการวิเคราะห์ตัวแบบความรุนแรงในการเรียกร้องค่าสินไหมทดแทน (Claim Severity Model) ของปี 2559	32
4.2.3 ผลการรวมตัวแบบเบี่ยงแปรกันภัยที่แท้จริงของปี 2559	34
4.3 ผลการวิเคราะห์ตัวแบบของกรรมธรรม์ที่เริ่มสัญญาในปี 2560	37
4.3.1 ผลการวิเคราะห์ตัวแบบความถี่ในการเรียกร้องค่าสินไหมทดแทน (Claim Frequency Model) ในปี 2560	37
4.3.2 ผลการวิเคราะห์ตัวแบบความรุนแรงในการเรียกร้องค่าสินไหมทดแทน (Claim Severity Model) ในปี 2560	39
4.3.3 ผลการรวมตัวแบบเบี่ยงแปรกันภัยที่แท้จริงของปี 2560	42
4.4 ผลการวัดความเหมาะสมของตัวแบบ	44
บทที่ 5 สรุปผลการวิจัย อภิปราย และข้อเสนอแนะ	45
5.1 สรุปและอภิปรายผลการวิเคราะห์ปัจจัยเสี่ยงที่มีผลต่อตัวแบบ	45
5.2 ข้อเสนอแนะ	46
รายการอ้างอิง	47

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ธุรกิจประกันภัยมีส่วนช่วยในการพัฒนาเศรษฐกิจและสังคมของประเทศ รวมทั้งมีบทบาทสำคัญในการพัฒนาตลาดทุนและระดมเงินออมของประเทศให้มีศักยภาพและขีดความสามารถในการแข่งขันระดับประเทศ ดังนั้นการสร้าง ความมั่นคงเข้มแข็งทางการเงินและยกระดับคุณภาพมาตรฐานของบริษัทประกันภัย จึงถือเป็นยุทธศาสตร์หลักของสำนักงานคณะกรรมการกำกับและส่งเสริมการประกอบธุรกิจประกันภัย (คปภ.) ซึ่งพื้นฐานสำคัญในการสร้างความมั่นคงเข้มแข็งทางการเงินของบริษัทประกันภัยคือ การกำหนดเบี้ยประกันภัยที่เหมาะสม

การกำหนดเบี้ยประกันภัยของบริษัทประกันภัยควรเป็นไปตามหลักการ 3 ประการ คือ

1. เบี้ยประกันภัยจะต้องเพียงพอ (Adequate) เพื่อให้บริษัทประกันภัยสามารถดำเนินการได้
2. เบี้ยประกันภัยจะต้องไม่สูงจนเกินไป (Not Excessive) จนผู้ที่ต้องการความคุ้มครองไม่สามารถซื้อประกันภัยได้
3. เบี้ยประกันภัยจะต้องแสดงให้เห็นถึงความแตกต่างของความเสี่ยงอย่างยุติธรรม (Fairly Discriminatory) กล่าวคือ เบี้ยประกันภัยจะต้องแสดงให้เห็นถึงค่าความเสียหายค่าใช้จ่ายที่คาดว่าจะเกิดขึ้นในอนาคตสำหรับแต่ละกลุ่มความเสี่ยง (

Robert L. Brown, 1993)

สำหรับวิธีการที่ใช้ในการคำนวณเบี้ยประกันภัยจะใช้ข้อมูลในอดีตมาประมาณเบี้ยประกันภัยที่เหมาะสมในอนาคต ซึ่งเป็นรายรับหลักของบริษัทประกันภัย (Insurer) ที่เรียกเก็บจากลูกค้าผู้เอาประกันภัย (Insured) จากสถิติที่เกิดขึ้นของผู้เอาประกันภัยแต่ละรายหรือเรียกว่าหน่วยเสี่ยงภัย (Exposure Unit) มีความแตกต่างกัน จึงต้องพิจารณาถึงประสบการณ์การเกิดภัยของหน่วยเสี่ยงภัยหรือตัวแปรปัจจัยเสี่ยงต่าง ๆ (Risk Factor) ที่มีผลกระทบต่อการประมาณเบี้ยประกันภัย โดยเบี้ยประกันภัยที่แท้จริง (Pure Premium) หมายถึง เบี้ยประกันภัยที่ครอบคลุมเฉพาะในส่วนขอความเสียหายที่คาดว่าจะเกิดขึ้นต่อหน่วยเสี่ยงภัยเท่านั้น ไม่รวมส่วนบวกเพิ่ม เช่น ค่าไรและค่าใช้จ่ายต่าง ๆ ของบริษัทประกันภัย ซึ่งตัวแบบของเบี้ยประกันภัยที่แท้จริงประกอบด้วย ความถี่ในการเรียกร้องค่าสินไหมทดแทน (Claim Frequency) และความรุนแรงของการเรียกร้องค่าสินไหมทดแทน (Claim Severity) ซึ่งมีลักษณะการแจกแจงที่ไม่เป็นการแจกแจงปกติ (Normal Distribution)

จากปัญหาและความสำคัญข้างต้น ในด้านคณิตศาสตร์ประกันภัยจึงได้มีผู้ที่ศึกษาและพัฒนาตัวแบบเชิงเส้นวางนัยทั่วไป (Generalized Linear Models: GLMs) เป็นวิธีการทางสถิติที่นำไปประยุกต์ใช้ในงานด้านการประกันภัย อาทิเช่น J. Zhou และ J. Garrido (2009) ได้ศึกษา GLMs ในการประมาณเงินสำรองค่าสินไหมทดแทน ต่อมาในปี ค.ศ. 2015 M. David ได้ใช้หลักวิธีการของ GLMs ไปประยุกต์ใช้ในการสร้างตัวแบบเบี้ยประกันภัยที่แท้จริงสำหรับการประกันภัยรถยนต์ในประเทศฝรั่งเศส ซึ่งทำให้ได้ทราบถึงปัจจัยเสี่ยงที่มีผลกระทบต่อบี้อประกันภัยแท้จริง

ด้วยเหตุนี้ ผู้วิจัยจึงสนใจที่จะนำวิธีการทางสถิติมาใช้ในการประมาณเบี้ยประกันภัยที่แท้จริงสำหรับการประกันภัยรถยนต์ในประเทศไทย โดยใช้ตัวแบบเชิงเส้นวงนัยทั่วไป

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อศึกษาวิธีการประมาณเบี้ยประกันภัยที่แท้จริงโดยใช้ตัวแบบเชิงเส้นวงนัยทั่วไป

1.2.2 เพื่อศึกษาปัจจัยที่มีผลต่อความถี่และความรุนแรงของการเรียกร้องค่าสินไหมทดแทน

1.3 ขอบเขตของการวิจัย

1.3.1 ข้อมูลที่ใช้เป็นข้อมูลทุติยภูมิที่ได้จากฐานข้อมูลกรมธรรม์ประกันอุบัติเหตุรถยนต์ภาคสมัครใจ ชั้น 1 ประเภทรยนต์นั่ง (รหัส 110 และ 120) ของสำนักงานอตราเบี้ยประกันวินาศภัยที่เริ่มสัญญาในปี 2559 และ 2560 ดังนี้

- ข้อมูลสำหรับตัวแปรตอบสนอง

- ความถี่ในการเรียกร้องค่าสินไหมทดแทน (ClaimFrequency)
- ความรุนแรงในการเรียกร้องค่าสินไหมทดแทน (ClaimSeverity)

ตาราง 1.1 ข้อมูลสำหรับตัวแปรอธิบาย

ส่วนของข้อมูลที่เป็นแบบกลุ่ม		
ปัจจัย	ตัวแปร (X)	คำอธิบาย
ลักษณะการใช้งาน (VehicleUsage)	private (base)	รถยนต์ใช้งานส่วนบุคคล (รหัส 110)
	commercial	รถยนต์ใช้งานเพื่อการพาณิชย์ (รหัส 120)
กลุ่มรถยนต์จำแนก ตามยี่ห้อและรุ่น (VehicleGroup)	Type 1	ประเภทที่ 1 (Super car, รถหรูขนาดใหญ่, SUV)
	Type 2	ประเภทที่ 2 (รถหรูระดับต้น, PPV)
	Type 3	ประเภทที่ 3 (Mid-size car, Pick-up)

ปล. จำแนกตาม Tariff 2535 ในตาราง 1.2	Type 4	ประเภทที่ 4 (Compact car)
	Type 5 (base)	ประเภทที่ 5 (Subcompact car)
ขนาดเครื่องยนต์ (VehicleSizeCode)	<=2000 CC.	รถยนต์ที่ใช้เครื่องยนต์ขนาดความจุไม่เกิน 2000 CC.
	>2000 CC. (base)	รถยนต์ที่ใช้เครื่องยนต์ขนาดความจุมากกว่า 2000 CC.
ประวัติการเรียกร้อง สินไหมทดแทนของผู้ เอาประกันภัย (BonusMalus)	Normal	ประวัติปกติ
	NCB (No Claim Bonus) (base)	ประวัติดี
	Surcharge	ประวัติไม่ดี
ส่วนของข้อมูลที่เป็นค่าตัวเลข		
ปัจจัย	ตัวแปร (X)	คำอธิบาย
อายุผู้ขับขี่	Age	หน่วย: ปี
อายุการใช้งานรถยนต์	VehicleAge	หน่วย: ปี
ทุนเอาประกันภัยของ รถยนต์	Bandvalue	หน่วย: บาท

ตาราง 1.2 ข้อมูลแบ่งกลุ่มรถยนต์สำหรับตัวแปรอธิบายกลุ่มรถยนต์ (เฉพาะรถยนต์นั่ง รหัส 110 120 เท่านั้น)

กลุ่มที่ 1		กลุ่มที่ 2		กลุ่มที่ 3		กลุ่มที่ 4		กลุ่มที่ 5	
ยี่ห้อ	รุ่น	ยี่ห้อ	รุ่น	ยี่ห้อ	รุ่น	ยี่ห้อ	รุ่น	ยี่ห้อ	รุ่น
ASTON	MARTIN	ALFA		DAIHAT	GRAN	MAZDA	323	DAIHA	MIRA
		ROMEO		SU				TSU	
BENTLE		AUDI			TERIO	MITSUBI	CHAM	HONDA	CITY
Y					S	SHI	P		
CADILLA		CHRYSLER		FIAT		TOYOTA	COROL	MAZDA	FAMILI
C							LA		A
CHEVRO		CITROEN		FORD		TOYOTA	CKD.	NISSAN	NV
LET									
DAIMLE		HOLDEN		HONDA	ACCO	NISSAN	CKD.		SUNNY
R					RD				
FERRARI		ISUZU	TROOPE		CIVIC	MITSUBI	CKD.	SUZUKI	CARRIBI
			R			SHI			AN
HAONDA	NSX	JEEP	CHERO	HYUDAI				TOYOT	SOLUNA
			KEE					A	
JAGUAR		LAND		ISUZU	CAME				
		ROVER			O				
MG		LEXUS			RODE				
					O				
NISSAN	PRESID	MAGSO			VEGA				
	ENT								
	INFINIT	MAZDA	MX5		VERTE				
	Y				X				
PORSCH		MERCEDE		KIA					
E		S-BENZ							

กลุ่มที่ 1		กลุ่มที่ 2		กลุ่มที่ 3		กลุ่มที่ 4		กลุ่มที่ 5	
ยี่ห้อ	รุ่น	ยี่ห้อ	รุ่น	ยี่ห้อ	รุ่น	ยี่ห้อ	รุ่น	ยี่ห้อ	รุ่น
ROLLS-ROYCE		MITSUBIS HI	PAJERO	MAZDA	121				
TOYOTA	SUPRA	OPEL	CALIBRA		626				
HONDA	CRV		OMAGA		ASTINA				
	LEGEND		VECTRA		CRONOS				
	ODYSSEY	PEUGEOT			LANTIS				
	PRELUDE	ROVER		MITSUBISHI	GALANT				
HONDA	CBU	SAAB		NISSAN	200SX				
		TOYOTA	CAMRY		CERFIRO				
			CELICA		PRIMER				
			CROWN		121				
			LAND CRUISER	OPEL	ASTRA				
			RAV 4		CORSA				
		VOLKSWAGEN		PROTON	SEGA				
		VOLVO		SEAT					
		BENZ		SUBARU	IMPENZA				

กลุ่มที่ 1		กลุ่มที่ 2		กลุ่มที่ 3		กลุ่มที่ 4		กลุ่มที่ 5	
ยี่ห้อ	รุ่น	ยี่ห้อ	รุ่น	ยี่ห้อ	รุ่น	ยี่ห้อ	รุ่น	ยี่ห้อ	รุ่น
		MAZDA	CBU.		LEGACY				
		TOYOTA	CBU.	SUZUKI	ESTEE M				
		NISSAN	CBU.		VITARA				
		MITSUBISHI	CBU.		CORONA				
				TOYOTA	STARLET				
				MAZDA	CKD.	หมายเหตุ CKD. หมายถึง รถที่ผลิตในประเทศ CBU. หมายถึง รถนำเข้าจากต่างประเทศ			
				HONDA	CKD.				

1.3.2 วิธีการทางสถิติที่นำมาใช้ในการวิเคราะห์ข้อมูล คือ ตัวแบบเชิงเส้นวางนัยทั่วไป

1.3.3 โปรแกรมสำเร็จรูปที่นำมาใช้ในการวิเคราะห์ข้อมูล คือ โปรแกรม RStudio

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 เข้าใจวิธีการประมาณแบบพหุคูณที่แท้จริงโดยใช้ตัวแบบเชิงเส้นวางนัยทั่วไป

1.4.2 สามารถนำวิธีการใช้ตัวแบบเชิงเส้นวางนัยทั่วไป ไปประยุกต์ใช้กับข้อมูลรูปแบบอื่น ๆ

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาวิธีการประมาณเบี่ยงแปรกันภัยที่แท้จริงโดยใช้ตัวแบบเชิงเส้นวางนัยทั่วไป ซึ่งในบทนี้จะกล่าวถึงตัวแบบเชิงเส้นวางนัยทั่วไปโดยละเอียด หลักการคำนวณเบี่ยงแปรกันภัยที่แท้จริง และงานวิจัยที่เกี่ยวข้อง เพื่อนำมาประกอบในงานวิจัย ซึ่งมีหัวข้อดังต่อไปนี้

2.1 วิธีการทางสถิติ

2.1.1 ตัวแบบเชิงเส้นวางนัยทั่วไป (Generalized Linear Model or GLMs)

ตัวแบบเชิงเส้นวางนัยทั่วไป (Generalized Linear Models) นั้นเป็นตัวแบบเชิงสถิติหนึ่งที่ถูกพัฒนาเพื่อใช้ในการวิเคราะห์ข้อมูลจำแนกประเภท ทำให้เข้าใจความเกี่ยวพันของตัวแปรตอบสนองและสามารถนำไปพยากรณ์กลุ่มต่าง ๆ ที่สนใจได้

Nelder and Wedderburn (1972) ได้เสนอตัวแบบเชิงเส้นวางนัยทั่วไปหรือ GLMs ขึ้น ซึ่งมีหลักการเพิ่มเติมโดย McCullagh and Nelder (1983,1989) GLMs เป็นตัวแบบที่พัฒนาจากตัวแบบเชิงเส้นคลาสสิก (Classical linear model) โดยขยายเพิ่มเติมจากข้อตกลงเบื้องต้นดังนี้

1. ส่วนประกอบเชิงสุ่ม (Random component)

สำหรับตัวแปรตอบสนองจากเดิมที่ใช้สำหรับการแจกแจงปกติ ขยายไปสู่ตัวแปรตอบสนองที่มีการแจกแจงในวงศ์เลขชี้กำลัง (Exponential family)

2. ส่วนประกอบเชิงระบบ (Systematic component)

จากเดิมที่เป็นผลรวมเชิงเส้นบนเทอมพารามิเตอร์ของตัวแปรอธิบายแบบต่อเนื่องเท่านั้น ขยายไปสู่ตัวแปรอธิบายแบบจำแนกประเภททั้งหมด หรือเป็นแบบผสม กล่าวคือ มีตัวแปรอธิบายทั้งแบบจำแนกประเภทและแบบต่อเนื่อง

3. ส่วนประกอบเชื่อมโยง (Link functions)

ฟังก์ชันเชื่อมโยงจากเดิมที่ใช้เฉพาะแบบ Identity link ขยายไปสู่การใช้ฟังก์ชันเชื่อมโยงแบบอื่นได้ โดยที่ฟังก์ชันเชื่อมโยงนั้นเป็นฟังก์ชันที่แปลงแล้วของค่าเฉลี่ยแบบ Monotonic differentiable function ได้หลายฟังก์ชันต่าง ๆ กัน

2.1.1.1 ส่วนประกอบของตัวแบบเชิงเส้นวางนัยทั่วไป (Component of GLMs)

ตัวแบบเชิงเส้นวางนัยทั่วไปแบบต่าง ๆ ประกอบไปด้วย 3 ส่วนประกอบ ดังนี้

1. ส่วนประกอบเชิงสุ่ม (Random component)

ส่วนประกอบเชิงสุ่ม เป็นส่วนของลักษณะการแจกแจงของตัวแปรตอบสนองเชิงสุ่ม (Y) โดยมีค่าสังเกตของ Y ขนาด N คือ $y_i, i = 1, \dots, N$ ที่ตัวแปรตอบสนอง Y เป็นอิสระต่อกันและ $Y = (y_1, \dots, y_N)$ มีการแจกแจงภายในวงศัเลขชี้กำลัง และสามารถเขียนฟังก์ชันความน่าจะเป็นได้ดังสมการที่ (2.1)

$$f_y(y; \theta, \phi) = \exp\{[y\theta - b(\theta)]/a(\phi) + c(y, \phi)\} \quad (2.1)$$

เมื่อ $a(\cdot)$, $b(\cdot)$ และ $c(\cdot)$ แทนฟังก์ชันต่าง ๆ และ (θ, ϕ) แทนพารามิเตอร์ของการแจกแจงความน่าจะเป็น

หากทราบ ϕ ที่ปรากฏในสมการที่ (2.1) แล้ว จะได้ว่าเป็นการแจกแจงหนึ่งในวงศัเลขชี้กำลังที่มีพารามิเตอร์ θ หากไม่ทราบ ϕ อาจเป็นหรือไม่เป็นการแจกแจงหนึ่งในวงศัเลขชี้กำลัง ที่มีพารามิเตอร์ 2 ตัว (θ, ϕ)

สำหรับพารามิเตอร์ θ เรียกว่า Natural parameter ส่วน ϕ เรียกว่า Dispersion parameter และฟังก์ชัน $a(\phi)$ ซึ่งโดยทั่วไปมีรูปแบบเป็น $a(\phi) = \phi/w_i$ โดยที่ w_i แทนน้ำหนักที่ทราบค่า เช่น \bar{y}_i แทนค่าเฉลี่ยของ N_i หน่วยที่เป็นอิสระต่อกัน จะนิยมใช้ $w_i = N_i$

ฟังก์ชันความน่าจะเป็นแบบต่อเนื่องและแบบไม่ต่อเนื่องของลักษณะการแจกแจงใด ๆ ที่สามารถเขียนได้ในรูปแบบของสมการที่ (2.1) จะแสดงลักษณะการแจกแจงที่อยู่ในวงศัเลขชี้กำลัง เช่น ฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability density function) ของการแจกแจงปกติ จะพบว่าเป็นลักษณะการแจกแจงหนึ่งในวงศัเลขชี้กำลัง เนื่องจากสามารถเขียนในรูปแบบ (2.1) ได้ดังนี้

$$\begin{aligned} f_y(y; \theta, \phi) &= \frac{1}{\sqrt{(2\pi\sigma^2)}} \exp\{-(y - \mu)^2/2\sigma^2\} \\ &= \exp\left\{(y\mu - \mu^2/2)/\sigma^2 - \frac{1}{2}(y^2/\sigma^2 + \log(2\pi\sigma^2))\right\} \end{aligned}$$

เมื่อ $\theta = \mu, b(\theta) = \theta^2/2, a(\phi) = \phi = \sigma^2$ และ $c(y, \theta) = -1/2 \{y^2/\sigma^2 + \log(2\pi\sigma^2)\}$

2. ส่วนประกอบเชิงระบบ (Systematic component)

ส่วนประกอบที่สองของ GLMs คือ ส่วนประกอบเชิงระบบของเซตของตัวแปรอธิบายที่มีระบบหรือรูปแบบเชิงเส้นในเทอมของพารามิเตอร์ หรือผลรวมเชิงเส้น $X\beta$ ซึ่งทำหน้าที่เชื่อมกับเวกเตอร์ η และเรียก $X\beta$ ว่าส่วนประกอบเชิงระบบ ดังนี้

$$\eta = X\beta$$

$$\eta = \sum_j \beta_j X_{ij} \quad \text{เมื่อ } i = 1, \dots, N \text{ และ } j = 1, \dots, p$$

โดย

η แทนตัวพยากรณ์เชิงเส้น (Linear predictor) หรือ $\eta = (\eta_1, \dots, \eta_N)$

X แทนเมทริกซ์ของตัวแปรอธิบาย อาจเรียก X ว่า Design matrix ที่มีขนาด $(N \times p)$

β แทนเวกเตอร์ของพารามิเตอร์ $(\beta_1, \dots, \beta_p)$

3. ส่วนประกอบฟังก์ชันเชื่อมโยง (Link functions)

ฟังก์ชันเชื่อมโยง คือ ฟังก์ชันของค่าเฉลี่ยของตัวแปรตอบสนองเชิงกลุ่ม (Y) ที่ใช้เป็นฟังก์ชันเชื่อมโยงระหว่างส่วนประกอบเชิงกลุ่มและส่วนประกอบเชิงระบบเข้าด้วยกัน แสดงได้ดังนี้

$$\mu_i = E(Y_i) \quad \text{เมื่อ } i = 1, \dots, N$$

โดยที่ μ_i เกี่ยวพันกับ η_i ในรูปฟังก์ชันของค่าเฉลี่ย ดังนี้

$$\eta_i = g(\mu_i) = \sum_j \beta_j X_{ij} \quad \text{เมื่อ } i = 1, \dots, N \text{ และ } j = 1, \dots, p.$$

เมื่อ g แทนฟังก์ชันแบบ Monotonic differentiable function และ p แทนจำนวนตัวแปรอธิบาย

ดังนั้น ตัวแบบที่ต้องการเชื่อมโยงระหว่างฟังก์ชันค่าเฉลี่ยของ Y กับเซตของตัวแปรอธิบาย X 's มีรูปแบบทั่วไป คือ

$$g(\mu_i) = \sum_j \beta_j X_{ij} \quad \text{เมื่อ } i = 1, \dots, N$$

ถ้า $g(\mu) = \mu$ จะได้ว่า $\eta_j = \mu_j$ นั่นคือเป็นฟังก์ชันเชื่อมโยงแบบ Identity link และอาจเรียกฟังก์ชันเชื่อมโยง อีกอย่างหนึ่งได้ว่า Canonical link โดยมีการแปลงค่าเฉลี่ยให้อยู่ในเทอมของพารามิเตอร์แสดงได้ดังนี้

$$g(\mu_i) = Q(\theta_i)$$

และ

$$Q(\theta_i) = \sum_j \beta_j X_{ij} \quad \text{เมื่อ } i = 1, 2, \dots, N \text{ และ } j = 1, 2, \dots, p$$

การใช้ตัวแบบเชิงเส้นวางนัยทั่วไป จะต้องทำการแปลงค่า Y เพื่อให้สอดคล้องกับสมมติฐานโดยการใช้ฟังก์ชันเชื่อมโยง (Link Function) เพื่อที่จะสามารถเขียน Y ให้อยู่ในรูปฟังก์ชันผลรวมของตัวแปรต้นได้ โดยที่ฟังก์ชันเชื่อมโยงจะต้องเป็นฟังก์ชันแบบทางเดียว (Monotonic) และสามารถหาอนุพันธ์ได้ (Differentiable) โดยรูปแบบฟังก์ชันเชื่อมโยงที่ใช้บ่อย แสดงได้ในตารางที่ 1 ดังนี้

ตารางที่ 1 รูปแบบฟังก์ชันเชื่อมโยงที่ใช้บ่อย

ฟังก์ชันเชื่อมโยง	$g(x)$	$g^{-1}(x)$
Identity	x	x
Log	$\ln x$	e^x
Logit	$\ln\left(\frac{x}{1-x}\right)$	$\frac{e^x}{1+e^x}$
Reciprocal	$\frac{1}{x}$	$\frac{1}{x}$

2.1.1.2 Offset Term

ในบางกรณีหากทราบผลกระทบที่จะเกิดขึ้นกับตัวแบบที่เป็นผลมาจากตัวแปรต้น จำเป็นต้องใส่ Offset Term ในตัวแบบเพื่อจัดผลกระทบที่จะเกิดขึ้นโดยกำหนดให้ตัวประมาณเชิงเส้น η เป็นไปตาม สมการที่ (2.3)

$$\eta = x\beta + \xi \quad (2.3)$$

จากสมการข้างต้น เราสามารถเขียนให้อยู่ในรูปแบบของสมการได้ดังแสดงในสมการที่ (2.4)

$$E[\underline{Y}] = \underline{\mu} = g^{-1}(X\underline{\beta} + \underline{\xi}) \quad (2.4)$$

ในด้านประกันภัย มีการนำตัวแบบเชิงเส้นวางนัยทั่วไปไปประยุกต์ใช้อย่างแพร่หลาย ดังแสดงได้ดังนี้

ตารางที่ 2 สรุปรูปแบบทั่วไปของตัวแบบเชิงเส้นวงนัยทั่วไป

	Claims Frequency	Claims Count	Average Claim Size	Probability
Link Function $g(x)$	$\ln x$	$\ln x$	$\ln x$	$\ln \frac{x}{(1-x)}$
Error (ξ)	Poisson	Poisson	Gamma	Binomial
Scale Parameter (ϕ)	1	1	Estimated	1
Variance Function $V(x)$	x	x	x^2	$\frac{x(t-x)}{t}$
Prior Weight (ω_i)	Exposure	1	Number of Claims	1
Offset Term (ξ)	0	In exposure	0	0

2.1.1.3 Prior Weight

Prior Weight คือ ข้อมูลที่แสดงถึงความน่าเชื่อถือของค่าสังเกตที่เราสามารถใส่ให้กับตัวแบบได้ โดย ค่าสังเกตที่มีจำนวนหน่วยเสี่ยงภัยมากจะมีความแปรปรวนน้อย และเป็นค่าที่มีผลต่อตัวแบบมาก

กำหนดให้

m_{ik} เป็นจำนวนครั้งของการเรียกร้องค่าสินไหมทดแทนจากหน่วยเสี่ยงภัยที่ k และจากกลุ่มที่ i

w_i เป็นจำนวนหน่วยเสี่ยงภัยของกลุ่มที่ i

Y_i เป็นความถี่ของการเรียกร้องค่าสินไหมทดแทนของกลุ่มที่ i

ดังนั้น รูปของสมการเป็นดังนี้

$$Y_i = \left(\frac{1}{w_i}\right) \sum_{k=1}^{w_i} m_{ik} \quad (2.5)$$

2.1.1.4 ระดับพื้นฐาน (Base Level)

โดยทั่วไปแล้ว การกำหนดระดับพื้นฐานสามารถกำหนดได้จากระดับที่หน่วยเสี่ยงภัยมากที่สุดในแต่ละปัจจัย

2.1.1.5 การอนุมานตัวแบบเชิงเส้นวงนัยทั่วไป (Inference for Generalized Linear Models)

การอนุมานสำหรับตัวแบบเชิงเส้นวงนัยทั่วไปเป็นการอนุมานถึงพารามิเตอร์ของตัวแบบและภาวะสารูปดีของตัวแบบ โดยทั่วไปใช้ตัวสถิติ Wald ตัวสถิติ Score และตัวสถิติ Likelihood ratio ซึ่งตัวสถิติ Likelihood ratio นิยมใช้ในรูปแบบของตัวสถิติ Deviance (D) ซึ่งจะกล่าวในหัวข้อต่อไป

2.1.1.6 Deviance และภาวะสารูปดีของตัวแบบ GLMs (Deviance and Goodness of fit of GLMs)

ตัวสถิติทดสอบภาวะสารูปดีของตัวแบบเชิงเส้นวงนัยทั่วไป นิยมใช้ตัวสถิติ ดีไวเอนส์ หรือ D (Nelder & Wedderburn (1972) ซึ่งมีการแจกแจงแบบไคสแควร์ n df ที่เกี่ยวข้องดังนี้

$$D = -2[\log L_0 - \log L_1]$$

$$D = -2[\log L(\hat{\mu}; y) - \log L(y; y)] \sim \chi^2_{N-p} \text{ เมื่อตัวอย่างมีขนาดใหญ่}$$

$$= 2 \sum_i [y_i \tilde{\theta}_i - b(\tilde{\theta}_i)]/a(\theta) - 2 \sum_i [y_i \hat{\theta}_i - b(\hat{\theta}_i)]/a(\phi)$$

โดยทั่วไป $a(\phi)$ มีรูปแบบ ϕ/ω_i ดังนั้น

$$2 \sum_i \omega_i [y_i (\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)]/a(\phi) = D(y; \hat{\mu})/\phi = D/\phi$$

เทอม D/ϕ เรียกว่า Scaled deviance และเทอม D เรียกว่า Deviance เมื่อ

$\log L_0 = \log L(\hat{\mu}, y)$ คือค่าสูงสุดของลอการิทึมของฟังก์ชันภาวะน่าจะเป็น (Log likelihood function) ของตัวแบบ GLMs ภายใต้ H_0 หรือ M_0

$\log L_1 = \log L(y, y)$ คือค่าสูงสุดของลอการิทึมของฟังก์ชันภาวะน่าจะเป็น (Log likelihood function) ของตัวแบบ GLMs ภายใต้ H_1 ซึ่งมีจำนวนพารามิเตอร์มากกว่าของตัวแบบ GLMs ภายใต้ H_0 หรืออาจใช้ตัวแบบ GLMs ภายใต้ H_1 เป็นตัวแบบเต็ม (The saturated model) หรือ M_1

2.1.1.7 การเลือกตัวแบบและการตรวจสอบความเหมาะสมของตัวแบบ

การเลือกตัวแบบนิยมใช้ตัวสถิติผลต่างของ Deviances ภายใต้ตัวแบบใน H_0 และตัวแบบใน H_1

$$\Delta D = D(y; \hat{\mu}_0) - D(y; \hat{\mu}_1) \sim \chi^2_{\Delta df} \quad \text{เมื่อตัวอย่างมีขนาดใหญ่}$$

เมื่อ Δdf แทนผลต่างขององศาอิสระ ซึ่งมีค่าเท่ากับผลต่างของจำนวนพารามิเตอร์ของ 2 ตัวแบบ

การทดสอบภาวะสารูปดีของตัวแบบภายใต้ H_0 และตัวแบบภายใต้ H_1 ซึ่งมีจำนวนพารามิเตอร์มากกว่าตัวแบบภายใต้ H_0 ทำให้ได้ในทำนองเดียวกันกับตัวสถิติ ΔD ซึ่งในหนังสือตำราต่าง ๆ ได้ใช้สัญลักษณ์ G^2 แทน D และใช้ ΔG^2 แทน ΔD ด้วย ดังนี้

$$G^2(M_0|M_1) = [G^2(M_0) - G^2(M_1)] \sim \chi^2_{\Delta df} \quad \text{เมื่อตัวแบบมีขนาดใหญ่}$$

2.1.1.8 Deviance สำหรับตัวแบบปัวซอง (Deviance for Poisson models)

เนื่องจากสำหรับ ตัวแบบปัวซอง ซึ่งเรียกตามส่วนประกอบเชิงสุ่มแบบปัวซอง จะได้ว่า

$$\hat{\theta}_i = \ln \hat{\mu}_i, b(\hat{\theta}_i) = \exp(\hat{\theta}_i) = \hat{\mu}_i$$

และในทำนองเดียวกันสำหรับตัวแบบเต็มจะได้ว่า

$$\tilde{\theta}_i = \ln y_i, b(\tilde{\theta}_i) = \exp(\tilde{\theta}_i) = y_i \text{ และ } a(\phi) = 1$$

ดังนั้นเทอม “Scale deviance” เท่ากับ “Deviance” โดยมีรูปแบบดังนี้

$$D = 2 \sum_i [y_i \log(y_i/\hat{\mu}_i) - y_i - \hat{\mu}_i]$$

เมื่อตัวแบบปัวซองรวมเทอม Intercept ด้วยสมการภาวะความน่าจะเป็นของการประมาณพารามิเตอร์จะรวมสมการ $\sum_i y_i = \sum_i \hat{\mu}_i$ ด้วย ดังนั้นเทอม Deviance จะลดรูปเป็นอย่างน้อย ดังนี้

$$D = 2 \sum_i [y_i \log(y_i/\hat{\mu}_i)]$$

สำหรับตัวแบบปัวซอง หรือตัวแบบมัลติโนเมียล ที่มีจำนวนเซลล์คงตัว (N) ตัวสถิติ D มีการแจกแจงที่ประมาณได้ด้วยการแจกแจงแบบไคสแควร์ เมื่อ $\{\mu_i\}$ มีขนาดใหญ่

2.1.1.9 ส่วนเหลือดีเวียนส์ (Deviance Residual or r_D)

ส่วนเหลือดีเวียนส์ (r_{D_i}) เป็นส่วนเหลือที่สร้างจากรากที่สองแบบคิดเครื่องหมายของผลลัพธ์ ส่วนเหลือของดีเวียนส์ในตัวแบบเชิงเส้นทั่วไป (d_i) เช่น ส่วนเหลือดีเวียนส์ที่ i ของค่าสังเกตที่ i คือ

$$r_{D_i} = \sqrt{d_i} \times \text{sign}(y_i - \hat{\mu}_i)$$

ดังนั้น

$$\sum r^2_{D_i} = \sum d_i = \text{Deviance}$$

ทางเลือกของส่วนเหลือดีเวียนส์ คือ ส่วนเหลือของเพียร์สัน (Pearson's residual)

2.1.2 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Square Error)

$$MSE = \frac{1}{n} \sum_i (Y_i - \widehat{f(x_i)})$$

n คือ จำนวนข้อมูล

Y_i คือ ค่าของตัวแปรตอบสนองชุดที่ i

$\widehat{f(x_i)}$ คือ ค่าพยากรณ์จากตัวแปรอธิบายชุดที่ i

2.1.3 ตัวแบบเบี่ยงแปรผันที่แท้จริง

การคำนวณเบี่ยงแปรผันที่แท้จริงจะพิจารณา ความถี่ในการเรียกร่องคำสินไหมทดแทน $E(N)$ และ ความรุนแรงของการเรียกร่องคำสินไหมทดแทน $E(C)$ ซึ่งให้เป็นอิสระจากกัน ได้สมการคำนวณเบี่ยงแปรผันที่แท้จริง ดังนี้

$$E(\sum_{k=1}^{N_i} C_{ik}) = E(N_i)E(C_i) \text{๗}$$

ในงานวิจัยนี้ใช้ *GLMs* ในการวิเคราะห์ โดยผลการวิเคราะห์จะเป็นสารสนเทศเพื่อไปใช้ประกอบการตัดสินใจให้กับ ผู้ปฏิบัติงานที่เกี่ยวข้องในสายงานธุรกิจประกันวินาศภัย ซึ่งในทางปฏิบัติไม่ต้องใช้กระบวนการที่ซับซ้อนที่อาจทำให้เกิดต้นทุนที่สูงเกินความจำเป็นของบริษัท

ดังนั้น การศึกษาตัวแบบ N และ C ของงานวิจัยนี้จึงเลือกใช้การแจกแจงของทั้งสองตัวแบบให้เป็นรูปแบบทั่วไปที่ใช้ในสายงานประกันวินาศภัย นั่นคือ $E(N)$ อยู่ในตัวแบบถดถอยปัวซอง และ $E(C)$ อยู่ในตัวแบบถดถอยแบบแกมมา ซึ่งทั้งสองตัวแบบมีฟังก์ชันการเชื่อมโยงเป็น *Log-Link function* ดังนี้

$$\ln(\mu) = \beta^t x$$

$$\mu = \exp(\beta^t x)$$

$$\mu \text{ คือ } E(N) \text{ หรือ } E(C)$$

โดยที่ β คือ ค่าสัมประสิทธิ์การถดถอยของปัจจัยเสี่ยง ที่มีผลกระทบต่อ μ และ x คือ เวกเตอร์ของปัจจัยเสี่ยง

จะได้สมการในการประมาณเบี่ยงแปรผันที่แท้จริง ดังนี้

$$\begin{aligned} E\left(\sum_{k=1}^{N_i} C_{ik}\right) &= E(N_i)E(C_i) = \exp(\beta_{freq}^t x_i) \cdot \exp(\beta_{sev}^t x_i) \\ &= \exp((\beta_{freq} + \beta_{sev})^t x_i) \end{aligned}$$

โดยที่ β_{freq} คือ ค่าสัมประสิทธิ์การถดถอยของปัจจัยเสี่ยงที่มีผลกระทบต่อ N และ β_{sev} คือ ค่าสัมประสิทธิ์การถดถอยของปัจจัยเสี่ยงที่มีผลกระทบต่อ C

2.2 คำศัพท์ประกันภัย

2.2.1 Insurance การประกันภัย

- การกระจายความเสี่ยงจากผู้เอาประกันภัย (Insured) ไปยังผู้รับประกันภัย (Insurer)

2.2.2 Claim Frequency ความถี่ในการเรียกร้องค่าสินไหมทดแทน

- จำนวนครั้งของการเรียกร้องค่าสินไหมทดแทนต่อ 1 หน่วยเสี่ยงภัย

$$\text{Claim Frequency} = \frac{\text{Claim Count}}{\text{Earned Exposure}}$$

2.2.3 Claim Count จำนวนครั้งของการเรียกร้องค่าสินไหมทดแทน

2.2.4 Earned Exposure จำนวนหน่วยเสี่ยงภัยที่นับได้

- หน่วยของทรัพย์สินหรือวัตถุที่เอาประกันภัย สำหรับการประกันภัยรถยนต์ จะใช้เป็น ปีรถยนต์ (Car- Year)

$$\text{Car Year} = \text{Number of Cars} \times \text{Number of Years}$$

ตัวอย่าง รถยนต์ 1 คัน ที่ได้เอาประกันภัยเป็นระยะเวลา 12 เดือน จะนับเป็น $1 \times 1 = 1$ ปีรถยนต์

รถยนต์ 3 คัน ที่ได้เอาประกันภัยเป็นระยะเวลา 6 เดือน จะนับเป็น $3 \times 0.5 = 1.5$ ปีรถยนต์

ปด. Earned Exposure = 0 หมายถึงมีการยกเลิกกรมธรรม์

2.2.5 Claim Severity ความรุนแรงของการเรียกร้องค่าสินไหมทดแทน

- ค่าสินไหมทดแทนต่อการเรียกร้อง 1 ครั้ง

$$\text{Claim Severity} = \frac{\text{Claim Amount}}{\text{Claim Count}}$$

2.2.6 Claim amount มูลค่าสินไหมทดแทน

$$\text{Claim Amount} = \frac{\text{ทุนประกัน}}{\text{มูลค่าทรัพย์สินที่เอาประกัน}} \times \text{ความเสียหายที่เกิดขึ้น}$$

ปล. ค่าสินไหมทดแทนน้อยกว่า 0 หมายถึง มีสินไหมทดแทนรับคืน

2.2.7 Pure Premium เบี้ยประกันภัยที่แท้จริง

- เบี้ยประกันภัยที่ครอบคลุมเฉพาะในส่วนของความเสียหายที่คาดว่าจะเกิดขึ้นต่อ 1 หน่วยเสี่ยงภัย โดยยังไม่รวมส่วนบวกเพิ่ม เช่น กำไรและค่าธรรมเนียมต่าง ๆ ของบริษัท

$$\text{Pure Premium} = \text{Claim Frequency} \times \text{Claim Severity}$$

$$\begin{aligned} &= \frac{\text{Claim Count}}{\text{Earned Exposure}} \times \frac{\text{Claim Amount}}{\text{Claim Count}} \\ &= \frac{\text{Claim Amount}}{\text{Earned Exposure}} \end{aligned}$$

2.3 งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่ได้ศึกษาเกี่ยวกับการทำวิจัยครั้งนี้ แบ่งออกเป็น 3 หัวข้อใหญ่ ได้แก่ งานวิจัยที่เกี่ยวข้องกับตัวแบบเชิงเส้นวางนัยทั่วไปที่ใช้ในงานประกันภัย งานวิจัยที่เกี่ยวข้องกับการประมาณเบี้ยประกันภัยที่แท้จริง และงานวิจัยที่เกี่ยวข้องกับการประมาณเบี้ยประกันภัยที่แท้จริงโดยใช้ตัวแบบเชิงเส้นวางนัยทั่วไป สามารถสรุปได้ดังนี้

2.3.1 ตัวแบบเชิงเส้นวางนัยทั่วไปกับการประกันภัย

J. Zhou และ J. Garrido (2009) ได้ศึกษาถึงการประมาณเงินสำรองค่าสินไหมทดแทนโดยใช้ตัวแบบเชิงเส้นวางนัยทั่วไป ที่เสนอโดย Nelder and Wedderburn (1972) ซึ่งมีหลักการเพิ่มเติมโดย McCullagh and Nelder (1983,1989) ซึ่ง Nelder (1989) กล่าวว่า GLMs เป็นตัวแบบที่พัฒนาจากตัวแบบเชิงเส้นคลาสสิก (Classical linear model) จากการศึกษาพบว่าเงินสำรองค่าสินไหมทดแทนที่ทำการประมาณโดยใช้ตัวแบบเชิงเส้นวางนัยทั่วไปมีค่าใกล้เคียงกับความเป็นจริงมากกว่าวิธีการดั้งเดิม อย่าง Chain-Ladder Method และ Bornhuetter-Ferguson Method

ปวีศา สุขเรื่อย และ สรรวม จงเจริญ (2019) ได้ศึกษาตัวแบบถดถอยที่มีผลกระทบค่าศูนย์ประยุกต์ใช้กับจำนวนครั้งของการเรียกร้องค่าสินไหมทดแทนในการประกันภัยรถยนต์ภาคสมัครใจ จากการศึกษาพบว่า ตัวแบบที่ถูกเลือกคือตัวแบบถดถอยทวินามนิเสธที่มีผลกระทบค่าศูนย์ และผลการศึกษาปัจจัยที่มีผลต่อการเรียกร้องค่าสินไหมทดแทนในการประกันภัยรถยนต์ภาคสมัครใจ คือ อายุของผู้เอาประกันภัย อายุของรถยนต์ ขนาดของตัวถังรถยนต์ และรุ่นของรถยนต์

2.3.2 การประมาณเบี้ยประกันภัยที่แท้จริง

สุทธาสินี จันทร์สมบูรณ์ (2009) ได้ศึกษาการพยากรณ์เบี้ยประกันภัยที่แท้จริงของการประกันภัยรถยนต์โดยใช้ทฤษฎีความน่าเชื่อถือด้วยวิธีเบอร์แมน-สกรับ ประชากรคือ กรมธรรม์ประกันภัยรถยนต์ภาคสมัครใจประเภทชั้น 1 รถยนต์นั่งส่วนบุคคล (รหัส 110) ขนาดเครื่องยนต์ไม่เกิน 2000 ซีซี รถยนต์กลุ่มที่ 3 อายุรถยนต์ 1-5 ปี อายุผู้ขับขี่ 18-35 ปี โดยใช้ข้อมูลในปี พ.ศ. 2548-2552 เป็นรายปี จากสำนักงานอตราเบี้ยประกันวินาศภัย ผลการศึกษาพบว่า ค่าพยากรณ์เบี้ยประกันภัยที่แท้จริงรวมทุกความคุ้มครองของการประกันภัยรถยนต์ปี พ.ศ. 2552 มีค่ามากกว่าค่าสินไหมทดแทนที่เกิดขึ้นจริงในปี พ.ศ. 2552

2.3.3 การประมาณเบี้ยประกันภัยที่แท้จริงโดยใช้ตัวแบบเชิงเส้นวางนัยทั่วไป

M. David (2015) ได้ศึกษาการคำนวณเบี้ยประกันภัยรถยนต์โดยใช้ตัวแบบเชิงเส้นวางนัยทั่วไป ประชากรคือกรมธรรม์ประกันภัยรถยนต์ของบริษัทประกันภัยแห่งหนึ่งในประเทศฝรั่งเศส จำนวน 50,000 กรมธรรม์ ที่เริ่มสัญญาในปี พ.ศ. 2552 ผลการศึกษาพบว่า เบี้ยประกันภัยที่แท้จริงจะเพิ่มขึ้น เมื่อตัวแปรอธิบาย ได้แก่ อายุผู้ขับขี่ (Age), อายุสัญญา (Age of Insurance Contract), ส่วนลดส่วนเพิ่มเบี้ยประกันภัยประวัติดี (Bonus-Malus) เพิ่มขึ้น

ชนกนาถ สุขประยูร และคณะ (2017) ได้ศึกษาการใช้ตัวแบบเชิงเส้นวางนัยทั่วไปในการประมาณค่าเบี้ยประกันภัยรถยนต์ ประชากรคือ กรมธรรม์ประกันภัยรถยนต์ประเภทชั้น 1 ของบริษัทประกันภัยแห่งหนึ่งในประเทศไทย ซึ่งเป็นกรมธรรม์ประเภทรถยนต์นั่ง ที่เริ่มสัญญาในปี พ.ศ. 2550 ระยะเวลาสัญญา 1 ปี จำนวน 26,959 กรมธรรม์ ซึ่งผลการศึกษาพบว่า มีเพียงปัจจัยเสี่ยง No Claim Bonus (NCB) และ Car Code ที่มีผลกระทบต่อค่าประมาณเบี้ยประกันภัยที่แท้จริง ที่ระดับนัยสำคัญ 0.05 ดังนั้นการกำหนดค่าเบี้ยประกันภัยควรคำนึงถึงปัจจัยดังกล่าว

บทที่ 3

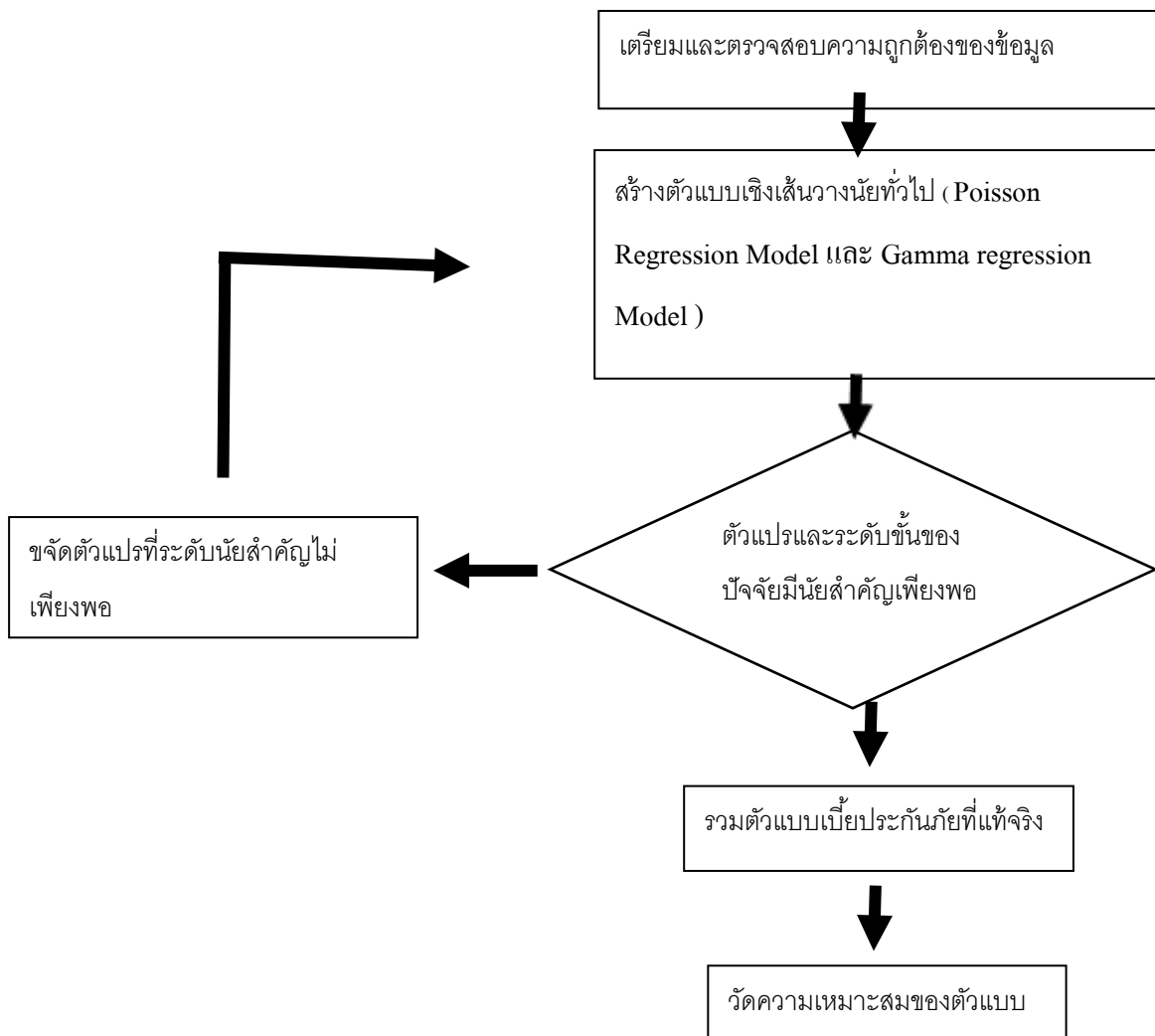
วิธีการดำเนินงานวิจัย

การประมาณเบี่ยงแปรกันภัยที่แท้จริงโดยใช้ตัวแบบเชิงเส้นวงนัยทั่วไปนั้น มีขั้นตอนการดำเนินงานวิจัย 6 ขั้นตอน ได้แก่ การเตรียมและตรวจสอบความถูกต้องของข้อมูล การสร้างตัวแบบเชิงเส้นวงนัยทั่วไป การรวมตัวแบบเพื่อหาเบี่ยงแปรกันภัยแท้จริง การวัดความเหมาะสมของตัวแบบ และการสรุปผล โดยมีรายละเอียดดังต่อไปนี้

3.1 เครื่องมือที่ใช้ในการดำเนินงานวิจัย

เครื่องมือที่ใช้ในการดำเนินงานวิจัยจะใช้โปรแกรม RStudio ในการสร้าง Poisson Regression Model และ Gamma regression Model และ การวัดความเหมาะสมของตัวแบบ

3.2 ภาพรวมของกระบวนการสร้างตัวแบบเชิงเส้นวงนัยทั่วไป



3.3 การเตรียมและตรวจสอบความถูกต้องของข้อมูล

ข้อมูลที่ใช้ในงานวิจัยครั้งนี้ เป็นข้อมูลทุติยภูมิที่ได้จากฐานข้อมูลกรมธรรม์ประกันอุบัติเหตุรถยนต์ภาคสมัครใจ ชั้น 1 ประเภทรถยนต์นั่ง (รหัส 110 และ 120) ของสำนักงานอัตราระเบีย้ประกันวินาศภัยจำนวน 2 ชุด โดยแบ่งเป็นข้อมูลกรมธรรม์ประกันอุบัติเหตุรถยนต์ที่เริ่มสัญญาในปีพ.ศ. 2559 1 ชุดและปีพ.ศ. 2560 1 ชุด โดยข้อมูลกรมธรรม์ประกันอุบัติเหตุรถยนต์ที่เริ่มสัญญาในปี พ.ศ.2559 มีจำนวน 471,763 กรมธรรม์ และข้อมูลประกันอุบัติเหตุรถยนต์ที่เริ่มสัญญาในปี พ.ศ.2560 จำนวน 460,683 กรมธรรม์

ในการสร้างตัวแบบเชิงเส้นวางนัยทั่วไปในที่นี้ จะสร้างตัวแบบ 2 ตัวแบบ นั่นคือ

1. ตัวแบบการถดถอยแบบปัวซอง สำหรับ Claim Frequency Model มีตัวแปรตอบสนองคือ ความถี่ในการเรียกร้องค่าสินไหมทดแทน (ClaimFrequency)
2. ตัวแบบการถดถอยแบบแกมมา สำหรับ Claim Severity Model มีตัวแปรตอบสนอง คือ ความรุนแรงในการเรียกร้องค่าสินไหมทดแทน (ClaimSeverity)

โดยทั้ง 2 ตัวแบบป็นั้นจะประกอบด้วยตัวแปรอธิบายชนิดเดียวกัน 7 ตัวแปร ดังนี้

- 1.อายุผู้ขับขี่ (Age)
- 2.อายุใช้งานรถยนต์ (VehicleAge)
- 3.ขนาดเครื่องยนต์ (VehicleSizeCode)
4. กลุ่มรถยนต์ (VehicleGroup)
- 5.ประวัติการเรียกร้องสินไหมทดแทนของผู้เอาประกันภัย (BonusMalus)
6. ลักษณะการใช้งานรถยนต์ (VehicleUsage)
- 7.ทุนเอาประกันภัยของรถยนต์ (BandValue)

3.3.1 การตรวจสอบความถูกต้อง

1. สำหรับตัวแปรอายุผู้ขับขี่ (Age) จะต้องมีย่านค่ามากกว่าหรือเท่ากับ 18 ปี ดังนั้นจะไม่นำข้อมูลที่ตัวแปรอายุผู้ขับขี่มีค่าน้อยกว่า 18 มาพิจารณาในตัวแบบ
2. สำหรับตัวแปรทุนเอาประกันภัยของรถยนต์ (BandValue) จะต้องมีย่านค่ามากกว่า 0 ดังนั้นจะไม่นำข้อมูลที่ตัวแปรทุนเอาประกันภัยของรถยนต์ มีค่าน้อยกว่า 0 มาพิจารณาในตัวแบบ
3. พิจารณาปัญหา จำนวนหน่วยเสี่ยงภัย (Earned Exposure) เท่ากับ 0 ซึ่งหมายถึงมีการยกเลิกกรมธรรม์ โดยจะไม่นำข้อมูลที่ค่าจำนวนหน่วยเสี่ยงภัย เท่ากับ 0 มาพิจารณาในตัวแบบ

3.3.2 การเตรียมข้อมูล

สำหรับตัวแปรอธิบาย

1. พิจารณาปัญหาค่าสูญหาย (Missing Value) ที่เกิดขึ้นในตัวแปรที่เลือกใช้ โดยไม่นำข้อมูลที่เกิดค่าสูญหายมาพิจารณาในตัวแบบ ในที่นี้ตัวแปรที่เกิดค่าสูญหายคือ อายุผู้ขับขี่ (Age) และ ขนาดเครื่องยนต์ (VehicleSizeCode)
2. สำหรับตัวแปรประวัติการเรียกร้องสินไหมทดแทนของผู้เอาประกันภัย (BonusMalus) เป็นตัวแปรจัดประเภท (Categorical Variable) ซึ่งถูกจัดแบ่งตามอัตราส่วนลดส่วนเพิ่มเบี้ยประกันภัยที่มีหลายระดับดังนี้

Normal 0%

NCB (0%, 20%, 30%, 40%, >=50%)

Surcharge (0%, 20%, 30%, 40%, >=50%)

จึงทำการจัดตัวแปรนี้ให้เหลือเพียง 3 ระดับ คือ Normal, NCB และ Surcharge

3. เนื่องจากตัวแปรทุนเอาประกันภัยของรถยนต์ (BandValue) เป็นตัวแปรจัดประเภท (Categorical Variable) ที่มีหลายระดับมากเกินไปซึ่งถูกจัดแบ่งตามช่วงของทุนประกันภัย จึงทำการเปลี่ยนตัวแปรนี้ให้เป็นตัวแปรต่อเนื่อง (Continuous Variable) โดยใช้ค่าเฉลี่ยเลขคณิต (Mean) แทน

สำหรับตัวแปรตอบสนอง

เนื่องจากข้อมูลที่ได้รับ ไม่ได้เป็นค่าของตัวแปรตอบสนองโดยตรง จึงต้องทำการหาค่าของตัวแปรตอบสนองก่อนดังนี้

1. ทำการหาค่าของตัวแปรตอบสนอง ClaimFrequency จากสูตร

$$\text{Claim Frequency} = \frac{\text{Claim Count}}{\text{Earned Exposure}}$$

2. ทำการหาค่าของตัวแปรตอบสนอง ClaimSeverity จากสูตร

$$\text{Claim Severity} = \frac{\text{Claim Amount}}{\text{Claim Count}}$$

3. เนื่องจากค่าของตัวแปรตอบสนอง ClaimSeverity ที่คำนวณได้บางแถวมีค่าน้อยกว่าหรือเท่ากับ 0 ซึ่งไม่ตรงกับข้อตกลงเบื้องต้นของ Gamma Regression Model คือ ต้องมีค่ามากกว่า 0 ดังนั้นจึงไม่นำข้อมูลที่ค่าของตัวแปรตอบสนอง ClaimSeverity น้อยกว่าหรือเท่ากับ 0 มาพิจารณาในตัวแบบ

3.4 การสร้างตัวแบบเชิงเส้นวงนัยทั่วไป

3.4.1 แบ่งข้อมูลรายปี 2559 และ 2560 ออกเป็น 2 ส่วน

ส่วนแรก: คิดเป็น 80% ของข้อมูลทั้งหมดเป็น Training Set โดยข้อมูลส่วนนี้จะใช้ในการสร้าง Claim Frequency Model และ Claim Severity Model

ส่วนหลัง: คิดเป็น 20% ของข้อมูลทั้งหมดเป็น Test Set โดยข้อมูลส่วนนี้จะใช้ในการวัดความเหมาะสมของตัวแบบที่สร้างจาก Training Set

- 3.4.2 กำหนด Base Level สำหรับตัวแปรข้อมูลจัดประเภท จากระดับที่มีจำนวนหน่วยเสี่ยงภัยมากที่สุดในแต่ละปีจ่าย

- 3.4.3 ในการสร้าง Claim Frequency Model ที่มีตัวแปรตอบสนอง คือ ClaimFrequency และ Claim Severity Model ที่มีตัวแปรตอบสนอง คือ ClaimSeverity จะมีรูปแบบทั่วไปของตัวแบบดังตารางที่ 2 นั่นคือ

สำหรับ Claim Frequency Model มี link function เป็นแบบ Log Link, ตัวแปรตอบสนองมีการแจกแจงแบบปัวซอง, Prior Weight ใช้เป็น Earned Exposure และ Offset Term มีค่าเป็น 0

สำหรับ Claim Severity Model มี link function เป็นแบบ Log Link, ตัวแปรตอบสนองมีการแจกแจงแบบแกมมา, Prior Weight ใช้เป็น Claim Count และ Offset Term มีค่าเป็น 0

3.5 การวิเคราะห์ผลที่ได้จากตัวแบบ

ทำการสร้างตัวแบบโดยเลือกเฉพาะปัจจัยที่มีอิทธิพลต่อตัวแปรตอบสนองทั้ง Claim Frequency Model และ Claim Severity Model จนกว่าจะได้ตัวแบบที่ทุกตัวแปรที่มีอิทธิพลต่อตัวแปรตอบสนอง

3.6 การรวมตัวแบบเพื่อหาเบี่ยงแปรกันภัยแท้จริง

เนื่องจากผลจากการวิเคราะห์ที่ได้จากโปรแกรม RStudio ไม่ได้แสดงถึงสัมประสิทธิ์ของ Claim Frequency Model และ Claim Severity Model โดยตรง ดังนั้นจะต้องนำผลการวิเคราะห์ที่ได้จากโปรแกรม Rstudio (ค่า Estimate) มาแปลงให้เป็นค่าสัมประสิทธิ์ของตัวแบบตามสมการ

$$\begin{aligned} E\left(\sum_{k=1}^{N_i} C_{ik}\right) &= E(N_i)E(C_i) = \exp(\beta_{freq}^t x_i) \cdot \exp(\beta_{sev}^t x_i) \\ &= \exp((\beta_{freq} + \beta_{sev})^t x_i) \end{aligned}$$

โดยที่ β_{freq} คือ ค่าสัมประสิทธิ์การถดถอยของปัจจัยเสี่ยงที่มีผลกระทบต่อ *ClaimFrequency* และ β_{sev} คือ ค่าสัมประสิทธิ์การถดถอยของปัจจัยเสี่ยงที่มีผลกระทบต่อ *ClaimSeverity*

3.7 การวัดความเหมาะสมของตัวแบบ

หลังจากที่ได้ตัวแบบที่ทุกตัวแปรที่มีอิทธิพลต่อตัวแปรตอบสนองแล้ว จะนำข้อมูล *Test set* ที่แบ่งเอาไว้ 20% ของข้อมูลทั้งหมด มาทำการหาความคลาดเคลื่อนเฉลี่ยกำลังกำลังสอง (*MSE*) ทั้งสองตัวแบบแยกตามแต่ละปี พร้อมทั้งเปรียบเทียบค่า *Residual Deviance* ของทั้งสองตัวแบบในแต่ละปี

3.8 การสรุปผล

การประมาณเบี้ยประกันภัยที่แท้จริงสำหรับการประกันภัยรถยนต์โดยใช้ตัวแบบเชิงเส้นวางนัยทั่ว ในงานวิจัยนี้จะประกอบด้วย 2 ตัวแบบ ได้แก่ ตัวแบบการถดถอยปัวซอง (Poisson regression model) ที่มีตัวแปรตอบสนองคือ ความถี่ในการเรียกค่าสินไหมทดแทน (ClaimFrequency) และตัวแบบถดถอยแกมมา (Gamma model) ที่มีตัวแปรตอบสนองคือ ความรุนแรงในการเรียกค่าสินไหมทดแทน (ClaimSeverity) โดยทำการสรุปผลการวิเคราะห์ข้อมูลออกเป็น 4 ส่วน ดังนี้

3.8.1 ลักษณะทั่วไปของข้อมูล

ลักษณะทั่วไปของข้อมูลกรมธรรม์ประกันอุบัติเหตุรถยนต์ภาคสมัครใจ ชั้น 1 ประเภทรถยนต์นั่ง (รหัส 110 และ 120) ของสำนักงานอตราเบี้ยประกันวินาศภัยซึ่งจะแสดง ตัวแปรปัจจัยเสี่ยง คำอธิบาย ร้อยละของหน่วยเสี่ยงภัย ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน จำแนกตามปี พ.ศ. 2559 และพ.ศ. 2560

3.8.2 ผลการวิเคราะห์ตัวแบบของกรมธรรม์ที่เริ่มสัญญาในปี 2559

นำเสนอผลการวิเคราะห์โดยแสดงถึงปัจจัยเสี่ยงต่าง ๆ ที่มีผลกระทบต่อตัวแบบและการรวมตัวแบบเพื่อประมาณเบี้ยประกันภัยที่แท้จริง สำหรับปี 2559

3.8.3 ผลการวิเคราะห์ตัวแบบของกรมธรรม์ที่เริ่มสัญญาในปี 2560

นำเสนอผลการวิเคราะห์โดยแสดงถึงปัจจัยเสี่ยงต่าง ๆ ที่มีผลกระทบต่อตัวแบบและการรวมตัวแบบเพื่อประมาณเบี้ยประกันภัยที่แท้จริง สำหรับปี 2560

3.8.4 ผลการวัดความเหมาะสมของตัวแบบ

นำเสนอผลการวิเคราะห์ความคลาดเคลื่อนโดยใช้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และ ค่า Residual Deviance เปรียบเทียบ FrequencyModel2016 กับ FrequencyModel2017 และ SeverityModel2016 กับ SeverityModel2017

บทที่ 4

ผลการดำเนินงานวิจัย

ในบทนี้ได้เสนอผลการวิเคราะห์ข้อมูลของการประมาณเบี้ยประกันภัยที่แท้จริงสำหรับการประกันภัยรถยนต์โดยใช้ตัวแบบเชิงเส้นวางนัยทั่วไป โดยได้นำเสนอผลการวิเคราะห์ข้อมูลออกเป็น 3 ส่วน ดังนี้

ส่วนที่ 1 ลักษณะทั่วไปของข้อมูล

ส่วนที่ 2 ผลการวิเคราะห์ตัวแบบของกรมธรรม์ที่เริ่มสัญญาในปี 2559

ส่วนที่ 3 ผลการวิเคราะห์ตัวแบบของกรมธรรม์ที่เริ่มสัญญาในปี 2560

ส่วนที่ 4 ผลการวัดความเหมาะสมของตัวแบบ

4.1 ลักษณะทั่วไปของข้อมูล

ข้อมูลที่ใช้ในงานวิจัยครั้งนี้ เป็นข้อมูลทุติยภูมิที่ได้จากฐานข้อมูลกรมธรรม์ประกันอุบัติเหตุรถยนต์ภาคสมัครใจ ชั้น 1 ประเภทรถยนต์นั่ง (รหัส 110 และ 120) ของสำนักงานอัตรเบี้ยประกันวินาศภัย

หลังจากการเตรียมและตรวจสอบความถูกต้อง มีลักษณะข้อมูลเป็นดังนี้

สำหรับปี 2559 มีจำนวนข้อมูลทั้งสิ้น 61,710 กรมธรรม์ แบ่งเป็น Training Set 44,369 กรมธรรม์ Test Set 12,341 กรมธรรม์

สำหรับปี 2560 มีจำนวนข้อมูลทั้งสิ้น 47,997 กรมธรรม์ แบ่งเป็น Training Set 38,399 กรมธรรม์ Test Set 9,598 กรมธรรม์

ตาราง 4.1 ลักษณะข้อมูลของ Training Set ที่ใช้ในตัวแบบปี พ.ศ. 2559 แสดงจำนวนร้อยละของหน่วยเสี่ยงภัยสำหรับข้อมูลที่เป็นแบบกลุ่ม ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานสำหรับข้อมูลที่เป็นตัวเลข

ส่วนของข้อมูลที่เป็นแบบกลุ่ม			
ปัจจัย	ตัวแปร X)	คำอธิบาย	ร้อยละของหน่วยเสี่ยงภัย
ลักษณะการใช้งาน (VehicleUsage)	private (base)	รถยนต์ใช้งานส่วนบุคคล (รหัส 110)	99.9959%

	commercial	รถยนต์ใช้งานเพื่อการพาณิชย์ (รหัส 120)	0.0041%
กลุ่มรถยนต์ จำแนกตามยี่ห้อ และรุ่น (VehicleGroup)	Type 1	ประเภทที่ 1	0.2269%
	Type 2	ประเภทที่ 2	6.1111%
	Type 3	ประเภทที่ 3	36.7579%
	Type 4	ประเภทที่ 4	14.9669%
	Type 5 (base)	ประเภทที่ 5	41.9372%
ขนาดเครื่องยนต์ (VehicleSizeCode)	<=2000 CC.	รถยนต์ที่ใช้เครื่องยนต์ ขนาดความจุไม่เกิน 2000 CC.	19.5143%
	>2000 CC. (base)	รถยนต์ที่ใช้เครื่องยนต์ ขนาดความจุมากกว่า 2000 CC.	80.4857%
ประวัติการ เรียกรถยนต์ใหม่ ทดแทนของผู้เอา ประกันภัย (BonusMalus)	Normal	ประวัติปกติ	4.8370%
	NCB (No Claim Bonus) (base)	ประวัติดี	94.9847%
	Surcharge	ประวัติไม่ดี	0.1782%
ส่วนของข้อมูลที่เป็นค่าตัวเลข			
ปัจจัย	ตัวแปร (X)	คำอธิบาย	ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน)
อายุผู้ขับขี่	Age	หน่วย: ปี	39.3601 (9.1495)
อายุการใช้งาน รถยนต์	VehicleAge	หน่วย: ปี	5.7671 (2.6498)
ทุนเอาประกันภัย ของรถยนต์	Bandvalue	หน่วย: บาท	2726359 (3871637)
ปัจจัย	ตัวแปร (Y)	คำอธิบาย	ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน)

ความถี่ในการ เรียกร้องค่า สินไหมทดแทน	ClaimFrequency	หน่วย: ครั้งต่อ 1 หน่วย เสี่ยงภัย	1.4956 (1.2359)
ความรุนแรงใน การเรียกร้องค่า สินไหมทดแทน	ClaimSeverity	หน่วย: บาทต่อครั้ง	17,058.43 (46,566.114)

ตาราง 4.2 ลักษณะข้อมูลของ Training Set ที่ใช้ในตัวแบบปี พ.ศ. 2560 แสดงจำนวนร้อยละของหน่วยเสี่ยงภัยสำหรับข้อมูลที่เป็นแบบกลุ่ม ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานสำหรับข้อมูลที่เป็นตัวเลข

ส่วนของข้อมูลที่เป็นแบบกลุ่ม			
ปัจจัย	ตัวแปร (X)	คำอธิบาย	ร้อยละของหน่วยเสี่ยงภัย
ลักษณะการใช้งาน (VehicleUsage)	private (base)	รถยนต์ใช้งานส่วนบุคคล (รหัส 110)	99.9922%
	commercial	รถยนต์ใช้งานเพื่อการ พาณิชย์ (รหัส 120)	0.0078%
กลุ่มรถยนต์ จำแนกตามยี่ห้อ และรุ่น (VehicleGroup)	Type 1	ประเภทที่ 1	0.3490%
	Type 2	ประเภทที่ 2	6.8465%
	Type 3	ประเภทที่ 3	37.0843%
	Type 4	ประเภทที่ 4	15.1801%
	Type 5 (base)	ประเภทที่ 5	40.5401%
ขนาดเครื่องยนต์ (VehicleSizeCode)	<=2000 CC.	รถยนต์ที่ใช้เครื่องยนต์ ขนาดความจุไม่เกิน 2000 CC.	19.2401%
	>2000 CC. (base)	รถยนต์ที่ใช้เครื่องยนต์ ขนาดความจุมากกว่า 2000 CC.	80.7599%
	Normal	ประวัติปกติ	3.7110%

ประวัติการ เรียกร้องสินไหม ทดแทนของผู้เอา ประกันภัย (BonusMalus)			
	NCB (No Claim Bonus) (base)	ประวัติดี	96.1405%
	Surcharge	ประวัติไม่ดี	0.1484%
ส่วนของข้อมูลที่เป็นค่าตัวเลข			
ปัจจัย	ตัวแปร (X)	คำอธิบาย	ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน)
อายุผู้ขับขี่	Age	หน่วย: ปี	40.0115 (9.2764)
อายุการใช้งาน รถยนต์	VehicleAge	หน่วย: ปี	6.2860 (2.6677)
ทุนเอาประกันภัย ของรถยนต์	Bandvalue	หน่วย: บาท	493,483.8 (423213.6)
ปัจจัย	ตัวแปร (Y)	คำอธิบาย	ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน)
ความถี่ในการ เรียกร้องค่า สินไหมทดแทน	ClaimFrequency	หน่วย: ครั้งต่อ 1 หน่วย เสี่ยงภัย	1.6076 (1.3998)
ความรุนแรงใน การเรียกร้องค่า สินไหมทดแทน	ClaimSeverity	หน่วย: บาทต่อครั้ง	17,138.86 (40,750.3)

4.2 ผลการวิเคราะห์ตัวแบบของกรมธรรม์ที่เริ่มสัญญาในปี 2559

ผลการวิเคราะห์ข้อมูลของการประมาณเบี้ยประกันภัยที่แท้จริงสำหรับการประกันภัยรถยนต์โดยใช้ตัวแบบเชิงเส้น
วางนัยทั่วไปสำหรับปี 2559 เป็นดังต่อไปนี้

4.2.1 ผลการวิเคราะห์ตัวแบบความถี่ในการเรียกร้องค่าสินไหมทดแทน (Claim Frequency Model) ของปี 2559

ตารางที่ 4.1 ค่าประมาณสัมประสิทธิ์การถดถอยของ Claim Frequency Model ของปี 2559

ปัจจัยเสี่ยง	Estimate	p-value
(Intercept)	4.248e-01	<2e-16 *
Age	-2.511e-03	8.22e-09 *
VehicleAge	6.194e-03	5.02e-05 *
VehicleGroupType3	1.306e-02	0.184586
VehicleGroupType4	1.038e-03	0.929754
VehicleGroupType2	5.827e-02	0.000518 *
VehicleGroupType1	7.414e-02	0.341164
VehicleSizeCode<=2000 CC.	-4.111e-02	0.000141 *
VehicleUsageCommercial	-4.068e-01	0.565182
BonusMalusNormal	2.001e-02	0.249027
BonusMalusSurcharge	1.105e-01	0.187012
Bandvalue	1.777e-09	0.099473

* หมายถึง ตัวแปรปัจจัยเสี่ยงที่มีผลต่อตัวแปรตอบสนอง ClaimFrequency ที่ระดับนัยสำคัญ 0.05

จากตารางที่ 4.1 พบว่ามีปัจจัยเสี่ยง Age, VehicleAge, VehicleGroupType2, VehicleSizeCode ที่มีผลกระทบต่อความถี่ของการเรียกร้องค่าสินไหมทดแทนในปี 2559

หลังจากพิจารณาดำเนินการแล้ว จะได้ Reduced Claim Frequency Model ของปี 2559 ดังตารางต่อไปนี้

ตารางที่ 4.2 ค่าประมาณสัมประสิทธิ์การถดถอยของ Reduced Claim Frequency Model ของปี 2559

ปัจจัยเสี่ยง	Estimate(SE.)	p-value
(Intercept)	0.429492(0.017228)	<2e-16 *
Age	-0.002228(0.000427)	1.81e-07 *
VehicleAge	0.5527(0.001457)	0.000148 *
VehicleSizeCode<=2000 CC.	-0.02792(0.009728)	0.005028 *

* หมายถึง ตัวแปรปัจจัยเสี่ยงที่มีผลต่อตัวแปรตอบสนอง ClaimFrequency ที่ระดับนัยสำคัญ 0.05

จากผลลัพธ์ พบว่าปัจจัยเสี่ยง Age, VehicleAge, VehicleSizeCode ทั้ง 3 ปัจจัยมีผลกระทบต่อความถี่ในการเรียกร้องค่าสินไหมทดแทนปี 2559 อย่างมีนัยสำคัญที่ระดับ 0.05

เมื่อให้ค่าตัวแปรปัจจัยเสี่ยงตัวอื่น ๆ คงที่ในตัวแบบ

- เมื่อค่าปัจจัยเสี่ยง Age เปลี่ยนแปลงเพิ่มขึ้น 1 ปี จะส่งผลกระทบต่อความถี่ในการเรียกร้องค่าสินไหมทดแทนลดลงมีค่าเป็น $e^{(-0.002228)(1)} = 0.998$ ซึ่งหมายความว่า จะมีค่าความถี่ในการเรียกร้องค่าสินไหมทดแทนลดลง 0.2%
- เมื่อค่าปัจจัยเสี่ยง VehicleAge เปลี่ยนแปลงเพิ่มขึ้น 1 ปี จะส่งผลกระทบต่อความถี่ในการเรียกร้องค่าสินไหมทดแทนลดลงมีค่าเป็น $e^{(0.005527)(1)} = 1.006$ ซึ่งหมายความว่า จะมีค่าความถี่ในการเรียกร้องค่าสินไหมทดแทนเพิ่มขึ้น 0.6%
- ปัจจัยเสี่ยง VehicleSizeCode <= 2000 CC จะส่งผลกระทบต่อความถี่ในการเรียกร้องค่าสินไหมทดแทนมีค่าเป็น $e^{(-0.02792)(1)} = 0.973$ เปรียบเทียบกับกลุ่ม Base Level คือ VehicleSizeCode > 2000 CC ซึ่งหมายความว่า กรมธรรม์ในกลุ่ม VehicleSizeCode <= 2000 CC จะมีความถี่ในการเรียกร้องค่าสินไหมทดแทน น้อยกว่ากรมธรรม์ในกลุ่ม VehicleSizeCode > 2000 CC เท่ากับ 2.7%

4.2.2 ผลการวิเคราะห์ตัวแบบความรุนแรงในการเรียกร้องค่าสินไหมทดแทน (Claim Severity Model) ของปี 2559

ตารางที่ 4.3 ค่าประมาณสัมประสิทธิ์การถดถอยของ Claim Severity Model ของปี 2559

ปัจจัยเสี่ยง	Estimate (SE.)	p-value
(Intercept)	9.790e+00 (4.391e-02)	<2e-16 *
Age	-4.203e-03 (1.089e-03)	0.000114 *
VehicleAge	-2.902e-02 (3.848e-03)	4.74e-14 *
VehicleGroupType3	1.682e-01 (2.484e-02)	1.29e-11 *
VehicleGroupType4	1.386e-01 (2.963e-02)	2.88e-06 *
VehicleGroupType2	1.025e+00 (4.239e-02)	<2e-16 *
VehicleGroupType1	6.469e-01 (1.966e-01)	0.000998 *
VehicleSizeCode<=2000 CC.	7.992e-02 (2.732e-02)	0.003448 *
VehicleUsageCommercial	5.075e-01 (1.784e+00)	0.776013
BonusMalusNormal	2.611e-02 (4.377e-02)	0.550828
BonusMalusSurcharge	3.325e-01 (2.112e-01)	0.115450
Bandvalue	7.62e-09 (2.724e-09)	0.005122 *

* หมายถึง ตัวแปรปัจจัยเสี่ยงที่มีผลต่อตัวแปรตอบสนอง ClaimSeverity ที่ระดับนัยสำคัญ 0.05

จากผลลัพธ์ พบว่ามีปัจจัยเสี่ยง Age, VehicleAge, VehicleGroupType, VehicleSizeCode และ BandValue ที่มีผลกระทบต่อความรุนแรงของการเรียกร้องค่าสินไหมทดแทนในปี 2559

หลังจากพิจารณานำตัวแปรออก จะได้ Reduced Claim Severity Model ของปี 2559 ดังตารางต่อไปนี้

ตารางที่ 4.4 ค่าประมาณสัมประสิทธิ์การถดถอยของ Reduced Claim Severity Model ของปี 2559

ปัจจัยเสี่ยง	ค่าสัมประสิทธิ์ของตัวแปร (SE.)	p-value
(Intercept)	9.791e+00 (4.390e-02)	<2e-16 *
Age	-4.180e-03 (1.090e-03)	0.000126 *
VehicleAge	-2.884e-02 (0.851e-03)	7.01e-14 *
VehicleGroupType3	1.673e-01 (2.486e-02)	1.74e-11 *
VehicleGroupType4	1.380e-01 (2.965e-02)	3.29e-06 *
VehicleGroupType2	1.026e+00 (4.241e-02)	<2e-16 *
VehicleGroupType1	6.478e-01 (1.968e-01)	0.000994 *
VehicleSizeCode<=2000 CC.	8.001e-02 (2.735e-02)	0.003441 *
Bandvalue	7.624e-09 (2.726e-09)	0.005162 *

* หมายถึง ตัวแปรปัจจัยเสี่ยงที่มีผลต่อตัวแปรตอบสนอง ClaimFrequency ที่ระดับนัยสำคัญ 0.05

จากตารางที่ 4.4 พบว่าปัจจัยเสี่ยง Age, VehicleAge, VehicleGroupType, VehicleSizeCode และ BandValue ทุกปัจจัยมีผลกระทบต่อความรุนแรงในการเรียกร้องค่าสินไหมทดแทนอย่างมีนัยสำคัญที่ระดับ 0.05 โดยเมื่อให้ค่าตัวแปรปัจจัยเสี่ยงตัวอื่น ๆ คงที่ในตัวแบบ โดยแปลความได้ดังนี้

- เมื่อค่าปัจจัยเสี่ยง Age เปลี่ยนแปลงเพิ่มขึ้น 1 ปี จะส่งผลกระทบต่อความรุนแรงในการเรียกร้องค่าสินไหมทดแทนลดลง มีค่าเป็น $e^{(-0.00418)(1)} = 0.996$ ซึ่งหมายความว่า จะมีค่าความรุนแรงในการเรียกร้องค่าสินไหมทดแทนลดลง 0.4%
- เมื่อค่าปัจจัยเสี่ยง VehicleAge เปลี่ยนแปลงเพิ่มขึ้น 1 ปี จะส่งผลกระทบต่อความรุนแรงในการเรียกร้องค่าสินไหมทดแทนลดลง มีค่าเป็น $e^{(-0.02884)(1)} = 0.972$ ซึ่งหมายความว่า จะมีค่าความรุนแรงในการเรียกร้องค่าสินไหมทดแทนลดลง 2.8%
- ปัจจัยเสี่ยง VehicleGroupType 1 จะส่งผลกระทบต่อความรุนแรงในการเรียกร้องค่าสินไหมทดแทนมีค่าเป็น $e^{(0.6478)(1)} = 1.911$ เปรียบเทียบกับกลุ่ม Base Level คือ VehicleGroupType 5 ซึ่งหมายความว่า กรมธรรม์ในกลุ่ม VehicleGroupType 1 จะมีความรุนแรงในการเรียกร้องค่าสินไหมทดแทน มากกว่ากรมธรรม์ในกลุ่ม VehicleGroupType 5 เท่ากับ 91.1%

- ปัจจัยเสี่ยง VehicleGroupType 2 จะส่งผลกระทบให้ความรุนแรงในการเรียกร้องค่าสินไหมทดแทนมีค่าเป็น $e^{(1.026)}(1) = 2.79$ เปรียบเทียบกับกลุ่ม Base Level คือ VehicleGroupType 5 ซึ่งหมายความว่า กรมธรรม์ในกลุ่ม VehicleGroupType 2 จะมีความรุนแรงในการเรียกร้องค่าสินไหมทดแทน มากกว่ากรมธรรม์ในกลุ่ม VehicleGroupType 5 เท่ากับ 179% หรือ 1.79 เท่า
- ปัจจัยเสี่ยง VehicleGroupType 3 จะส่งผลกระทบให้ความรุนแรงในการเรียกร้องค่าสินไหมทดแทนมีค่าเป็น $e^{(0.1683)}(1) = 1.183$ เปรียบเทียบกับกลุ่ม Base Level คือ VehicleGroupType 5 ซึ่งหมายความว่า กรมธรรม์ในกลุ่ม VehicleGroupType 3 จะมีความรุนแรงในการเรียกร้องค่าสินไหมทดแทน มากกว่ากรมธรรม์ในกลุ่ม VehicleGroupType 5 เท่ากับ 18.3%
- ปัจจัยเสี่ยง VehicleGroupType 4 จะส่งผลกระทบให้ความรุนแรงในการเรียกร้องค่าสินไหมทดแทนมีค่าเป็น $e^{(0.138)}(1) = 1.148$ เปรียบเทียบกับกลุ่ม Base Level คือ VehicleGroupType 5 ซึ่งหมายความว่า กรมธรรม์ในกลุ่ม VehicleGroupType 4 จะมีความรุนแรงในการเรียกร้องค่าสินไหมทดแทน มากกว่ากรมธรรม์ในกลุ่ม VehicleGroupType 5 เท่ากับ 14.8%
- ปัจจัยเสี่ยง VehicleSizeCode ≤ 2000 CC จะส่งผลกระทบให้ความรุนแรงในการเรียกร้องค่าสินไหมทดแทนมีค่าเป็น $e^{(0.08001)}(1) = 1.083$ เปรียบเทียบกับกลุ่ม Base Level คือ VehicleSizeCode > 2000 CC ซึ่งหมายความว่า กรมธรรม์ในกลุ่ม VehicleSizeCode ≤ 2000 CC จะมีความรุนแรงในการเรียกร้องค่าสินไหมทดแทน มากกว่ากรมธรรม์ในกลุ่ม VehicleSizeCode > 2000 CC เท่ากับ 8.3%
- เมื่อค่าปัจจัยเสี่ยง BandValue เปลี่ยนแปลงเพิ่มขึ้น 1 บาท จะส่งผลกระทบให้ความรุนแรงในการเรียกร้องค่าสินไหมทดแทนลดลงมีค่าเป็น $e^{(0.000000008)}(1) \sim 1$ ซึ่งหมายความว่า จะมีค่าความรุนแรงในการเรียกร้องค่าสินไหมทดแทน เพิ่มขึ้นจากเดิมไม่มาก

4.2.3 ผลการรวมตัวแบบเบี่ยงแปรผันที่แท้จริงของปี 2559

ตารางที่ 4.5 แสดงผลการวิเคราะห์ GLMs กับการประมาณค่าเบี่ยงแปรผันที่แท้จริงที่เกิดจากการรวมตัวแบบของ Claim

Frequency Model และ Claim Severity Model ของปี 2559

ปัจจัยเสี่ยง	ค่าสัมประสิทธิ์ของตัวแปร		
	ตัวแบบถดถอยปัวซอง	ตัวแบบถดถอยแกมมา	ตัวแบบเบย์ประกันภัยแท้จริง
Intercept	$\widehat{\beta}_{\text{freq}} = 0.4295$ (0.0172)	$\widehat{\beta}_{\text{sev}} = 9.791$ (0.0439)	$\widehat{\beta}_{\text{freq}} + \widehat{\beta}_{\text{sev}} = 10.2205$
Age	$\widehat{\beta}_{\text{freq}} = -0.0022$ (0.0004)	$\widehat{\beta}_{\text{sev}} = -0.0042$ (0.0011)	$\widehat{\beta}_{\text{freq}} + \widehat{\beta}_{\text{sev}} = -0.0064$
VehicleAge	$\widehat{\beta}_{\text{freq}} = 0.0055$ (0.0015)	$\widehat{\beta}_{\text{sev}} = -0.0288$ (0.0039)	$\widehat{\beta}_{\text{freq}} + \widehat{\beta}_{\text{sev}} = -0.0233$
VehicleSizeCode <=2000 CC.	$\widehat{\beta}_{\text{freq}} = -0.0273$ (0.0097)	$\widehat{\beta}_{\text{sev}} = 0.080$ (0.0274)	$\widehat{\beta}_{\text{freq}} + \widehat{\beta}_{\text{sev}} = 0.0527$
Bandvalue	-	$\widehat{\beta}_{\text{sev}} = 7.624 \times 10^{-9}$ (2.726×10^{-9})	$\widehat{\beta}_{\text{freq}} + \widehat{\beta}_{\text{sev}} = 7.624 \times 10^{-9}$
VehicleGroup Type 3	-	$\widehat{\beta}_{\text{sev}} = 0.1673$ (0.0249)	$\widehat{\beta}_{\text{freq}} + \widehat{\beta}_{\text{sev}} = 0.1673$
VehicleGroup Type 4	-	$\widehat{\beta}_{\text{sev}} = 0.1380$ (0.0297)	$\widehat{\beta}_{\text{freq}} + \widehat{\beta}_{\text{sev}} = 0.1380$
VehicleGroup Type 2	-	$\widehat{\beta}_{\text{sev}} = 1.026$ (0.0424)	$\widehat{\beta}_{\text{freq}} + \widehat{\beta}_{\text{sev}} = 1.026$
VehicleGroup Type 1	-	$\widehat{\beta}_{\text{sev}} = 0.6478$ (0.1968)	$\widehat{\beta}_{\text{freq}} + \widehat{\beta}_{\text{sev}} = 0.6478$
() คือ ค่า Standard Error (SE)			

จากตารางที่ 4.5 จะได้ตัวแบบประมาณเบี่ยงประกันภัยที่แท้จริงดังนี้

$$E\left(\sum_{k=1}^{N_i} C_{ik}\right) \\ = \text{Exp}[10.2205 - 0.0064x_{Age} - 0.0233x_{VehAge} + 0.0527x_{\leq 2000 \text{ CC}} \\ + (7.624 * 10^{-9})x_{Band} + 0.6478x_{Type1} + 1.026x_{Type2} \\ + 0.1673x_{Type3} + 0.1380x_{Type4}]$$

จากตัวแบบข้างต้น สามารถประมาณเบี่ยงประกันภัยที่แท้จริง ได้ดังตัวอย่างต่อไปนี้

ข้อมูลกรรมกรรม Age = 25 , VehicleAge = 1 , VehicleSizeCode = <2000 CC , BandValue = 4,900,000.5 , VehicleGroup = Type

1

$$E\left(\sum_{k=1}^{N_i} C_{ik}\right) \\ = \text{Exp}[10.2205 - 0.0064(25) - 0.0233(1) + 0.0527(1) \\ + (7.624 * 10^{-9})(4,900,000.5) + 0.6478(1) + 1.026(0) + 0.1673(0) \\ + 0.1380(0)]$$

= 47,813 23 บาท

4.3 ผลการวิเคราะห์ตัวแบบของกรมธรรม์ที่เริ่มสัญญาในปี 2560

ผลการวิเคราะห์ข้อมูลของการประมาณเบี้ยประกันภัยที่แท้จริงสำหรับการประกันภัยรถยนต์โดยใช้ตัวแบบเชิงเส้นวางนัยทั่วไปสำหรับปี 2560 เป็นดังต่อไปนี้

4.3.1 ผลการวิเคราะห์ตัวแบบความถี่ในการเรียกร้องค่าสินไหมทดแทน (Claim Frequency Model) ในปี 2560

ตารางที่ 4.6 ค่าประมาณสัมประสิทธิ์การถดถอยของ Claim Frequency Model ในปี 2560

ปัจจัยเสี่ยง	ค่าสัมประสิทธิ์ของตัวแปร (SE.)	p-value
(Intercept)	3.986e-01 (2.232e-02)	<2e-16 *
Age	-1.441e-03 (4.986e-04)	0.00384 *
VehicleAge	9.674e-03 (1.907e-02)	3.91e-07 *
VehicleGroupType3	-4.923e-03 (1.131e-02)	0.66336
VehicleGroupType4	7.663e-03 (1.358e-02)	0.57248
VehicleGroupType2	4.003e-02 (2.512e-02)	0.11098
VehicleGroupType1	-8.319e-03 (7.907e-02)	0.91621
VehicleSizeCode<=2000 CC.	-1.760e-02 (1.255e-02)	0.16069
VehicleUsageCommercial	6.383e-01 (3.536e-01)	0.07106
BonusMalusNormal	1.945e-02 (2.290e-02)	0.39559
BonusMalusSurcharge	8.337e-02 (1.074e-01)	0.43745
Bandvalue	3.010e-08 (1.536e-08)	0.05004

* หมายถึง ตัวแปรปัจจัยเสี่ยงที่มีผลต่อตัวแปรตอบสนอง ClaimFrequency ที่ระดับนัยสำคัญ 0.05

จากตารางที่ 4.6 พบว่ามีปัจจัยเสี่ยง Age และ VehicleAge ที่มีผลกระทบต่อความถี่ของการเรียกร้องค่าสินไหมทดแทนในปี 2560

หลังจากพิจารณานำตัวแปรออก จะได้ Reduced Claim Frequency Model ของปี 2560 ดังตารางต่อไปนี้

ตารางที่ 4.7 ค่าประมาณสัมประสิทธิ์การถดถอยของ Reduced Claim Frequency Model ในปี 2560

ปัจจัยเสี่ยง	ค่าสัมประสิทธิ์ของตัวแปร (SE.)	p-value
(Intercept)	0.4159553 (0.0203272)	<2e-16 *
Age	-0.0011839 (0.0004851)	0.0147 *
VehicleAge	0.0075747 (0.0016732)	5.98e-06 *

* หมายถึง ตัวแปรปัจจัยเสี่ยงที่มีผลต่อตัวแปรตอบสนอง ClaimFrequency ที่ระดับนัยสำคัญ 0.05

จากตารางที่ 4.7 พบว่าปัจจัยเสี่ยง Age และ VehicleAge ทั้ง 2 ปัจจัยมีผลกระทบต่อความถี่ในการเรียกร้องค่าสินไหมทดแทนในปี 2560 อย่างมีนัยสำคัญที่ระดับ 0.05

เมื่อให้ค่าตัวแปรปัจจัยเสี่ยงตัวอื่น ๆ คงที่ในตัวแบบ

- เมื่อค่าปัจจัยเสี่ยง Age เปลี่ยนแปลงเพิ่มขึ้น 1 ปี จะส่งผลกระทบต่อความถี่ในการเรียกร้องค่าสินไหมทดแทนลดลงมีค่าเป็น $e^{(-0.0011839)(1)} = 0.999$ ซึ่งหมายความว่า จะมีค่าความถี่ในการเรียกร้องค่าสินไหมทดแทนลดลง 0.1%
- เมื่อค่าปัจจัยเสี่ยง VehicleAge เปลี่ยนแปลงเพิ่มขึ้น 1 ปี จะส่งผลกระทบต่อความถี่ในการเรียกร้องค่าสินไหมทดแทนลดลงมีค่าเป็น $e^{(0.0075747)(1)} = 1.008$ ซึ่งหมายความว่า จะมีค่าความถี่ในการเรียกร้องค่าสินไหมทดแทนเพิ่มขึ้น 0.8%

4.3.2 ผลการวิเคราะห์ตัวแบบความรุนแรงในการเรียกร้องค่าสินไหมทดแทนในปี 2560

ตารางที่ 4.8 ค่าประมาณสัมประสิทธิ์การถดถอยของ Claim Severity Model ในปี 2560

ปัจจัยเสี่ยง	ค่าสัมประสิทธิ์ของตัวแปร (SE.)	p-value
(Intercept)	9.613e+00 (5.134e-02)	<2e-16 *
Age	-3.436e-03 (1.128e-03)	0.002322 *
VehicleAge	-1.086e-02 (4.429e-03)	0.014163 *
VehicleGroupType3	8.608e-02 (2.585e-02)	0.000868 *
VehicleGroupType4	8.546e-02 (3.087e-02)	0.005635 *
VehicleGroupType2	4.706e-01 (5.998e-02)	4.42e-15 *
VehicleGroupType1	7.223e-02 (1.783e-01)	0.685412
VehicleSizeCode<=2000 CC.	-7.546e-03 (2.862e-02)	0.792053
VehicleUsageCommercial	-3.701e-01 (8.031e-01)	0.644955
BonusMalusNormal	-8.200e-02 (5.198e-02)	0.114710
BonusMalusSurcharge	2.060e-01 (2.439e-01)	0.398361
Bandvalue	3.708e-07 (3.849e-08)	<2e-16 *

* หมายถึง ตัวแปรปัจจัยเสี่ยงที่มีผลต่อตัวแปรตอบสนอง ClaimSeverity ที่ระดับนัยสำคัญ 0.05

จากตาราง 4.8 พบว่ามีปัจจัยเสี่ยง Age, VehicleAge, VehicleGroupType และ Bandvalue ที่มีผลกระทบต่อความรุนแรงของการเรียกร้องค่าสินไหมทดแทนในปี 2560

ตารางที่ 4.9 ค่าประมาณสัมประสิทธิ์การถดถอยของ Reduced Claim Severity Model ในปี 2560

ปัจจัยเสี่ยง	ค่าสัมประสิทธิ์ของตัวแปร (SE.)	p-value
(Intercept)	9.613e+00 (5.068e-02)	<2e-16 *
Age	-3.438e-03 (1.127e-03)	0.002293 *
VehicleAge	-1.117e-02 (4.374e-03)	0.010643 *
VehicleGroupType3	8.520e-02 (2.456e-02)	0.000522 *
VehicleGroupType4	8.466e-02 (3.066e-02)	0.005764 *
VehicleGroupType2	4.706e-01 (5.995e-02)	4.28e-15 *
VehicleGroupType1	7.894e-02 (1.784e-01)	0.658093
Bandvalue	3.665e-07 (3.784e-08)	<2e-16 *

* หมายถึง ตัวแปรปัจจัยเสี่ยงที่มีผลต่อตัวแปรตอบสนอง ClaimSeverity ที่ระดับนัยสำคัญ 0.05

จากตารางที่ 4.9 พบว่าปัจจัยเสี่ยง Age, VehicleAge, VehicleGroupType และ Bandvalue ทุกปัจจัยมีผลกระทบต่อความรุนแรงในการเรียกร้องค่าสินไหมทดแทนในปี 2560 อย่างมีนัยสำคัญที่ระดับ 0.05

เมื่อให้ค่าตัวแปรปัจจัยเสี่ยงตัวอื่น ๆ คงที่ในตัวแบบ

- เมื่อค่าปัจจัยเสี่ยง Age เปลี่ยนแปลงเพิ่มขึ้น 1 ปี จะส่งผลกระทบต่อความรุนแรงในการเรียกร้องค่าสินไหมทดแทนลดลง มีค่าเป็น $e^{(-0.003438)(1)} = 0.997$ ซึ่งหมายความว่า จะมีค่าความรุนแรงในการเรียกร้องค่าสินไหมทดแทนลดลง 0.3%
- เมื่อค่าปัจจัยเสี่ยง VehicleAge เปลี่ยนแปลงเพิ่มขึ้น 1 ปี จะส่งผลกระทบต่อความรุนแรงในการเรียกร้องค่าสินไหมทดแทนลดลง มีค่าเป็น $e^{(-0.01117)(1)} = 0.989$ ซึ่งหมายความว่า จะมีค่าความรุนแรงในการเรียกร้องค่าสินไหมทดแทนลดลง 1.1%
- ปัจจัยเสี่ยง VehicleGroupType 2 จะส่งผลกระทบต่อความรุนแรงในการเรียกร้องค่าสินไหมทดแทนมีค่าเป็น $e^{(0.4706)(1)} = 1.601$ เปรียบเทียบกับกลุ่ม Base Level คือ VehicleGroupType 5 ซึ่งหมายความว่า กรมธรรม์ในกลุ่ม VehicleGroupType 2 จะมีความรุนแรงในการเรียกร้องค่าสินไหมทดแทน มากกว่ากรมธรรม์ในกลุ่ม VehicleGroupType 5 เท่ากับ 60.1%

- ปัจจัยเสี่ยง VehicleGroupType 3 จะส่งผลกระทบต่อให้ความรุนแรงในการเรียกร้อยค่าสินไหมทดแทนมีค่าเป็น $e^{(0.08520)(1)} = 1.089$ เปรียบเทียบกับกลุ่ม Base Level คือ VehicleGroupType 5 ซึ่งหมายความว่า กรรมธรรม์ในกลุ่ม VehicleGroupType 3 จะมีความรุนแรงในการเรียกร้อยค่าสินไหมทดแทน มากกว่ากรรมธรรม์ในกลุ่ม VehicleGroupType 5 เท่ากับ 8.9%
- ปัจจัยเสี่ยง VehicleGroupType 4 จะส่งผลกระทบต่อให้ความรุนแรงในการเรียกร้อยค่าสินไหมทดแทนมีค่าเป็น $e^{(0.08466)(1)} = 1.088$ เปรียบเทียบกับกลุ่ม Base Level คือ VehicleGroupType 5 ซึ่งหมายความว่า กรรมธรรม์ในกลุ่ม VehicleGroupType 4 จะมีความรุนแรงในการเรียกร้อยค่าสินไหมทดแทน มากกว่ากรรมธรรม์ในกลุ่ม VehicleGroupType 5 เท่ากับ 8.8%
- เมื่อค่าปัจจัยเสี่ยง BandValue เปลี่ยนแปลงเพิ่มขึ้น 1 บาท จะส่งผลกระทบต่อให้ความรุนแรงในการเรียกร้อยค่าสินไหมทดแทนลดลงมีค่าเป็น $e^{(3.665 \times 10^{-7})(1)} \sim 1$ ซึ่งหมายความว่า จะมีค่าความรุนแรงในการเรียกร้อยค่าสินไหมทดแทน เพิ่มขึ้นจากเดิมไม่มาก

4.3.3 ผลการรวมตัวแบบเบี่ยงแปรกันภัยที่แท้จริงของปี 2560

ตารางที่ 4.10 แสดงผลการวิเคราะห์ GLMs กับการประมาณค่าเบี่ยงแปรกันภัยที่แท้จริงที่เกิดจากการรวมตัวแบบของ Claim Frequency Model และ Claim Severity Model ของปี 2560

ปัจจัยเสี่ยง	ค่าสัมประสิทธิ์ของตัวแปร		
	ตัวแบบถดถอยปัวซอง	ตัวแบบถดถอยแกมมา	ตัวแบบเบี่ยงแปรกันภัยที่แท้จริง
Intercept	$\widehat{\beta}_{freq} = 0.416$ (0.0203)	$\widehat{\beta}_{sev} = 9.613$ (0.0507)	$\widehat{\beta}_{freq} + \widehat{\beta}_{sev} = 10.029$
Age	$\widehat{\beta}_{freq} = -0.0012$ (0.0005)	$\widehat{\beta}_{sev} = -0.0034$ (0.0011)	$\widehat{\beta}_{freq} + \widehat{\beta}_{sev} = -0.0046$
VehicleAge	$\widehat{\beta}_{freq} = 0.0076$ (0.0017)	$\widehat{\beta}_{sev} = -0.0112$ (0.0044)	$\widehat{\beta}_{freq} + \widehat{\beta}_{sev} = -0.0036$
Bandvalue	-	$\widehat{\beta}_{sev} = 3.665 \times 10^{-7}$ (3.784×10^{-8})	$\widehat{\beta}_{freq} + \widehat{\beta}_{sev} = 3.665 \times 10^{-7}$
VehicleGroup Type 3	-	$\widehat{\beta}_{sev} = 0.0852$ (0.0246)	$\widehat{\beta}_{freq} + \widehat{\beta}_{sev} = 0.0852$
VehicleGroup Type 4	-	$\widehat{\beta}_{sev} = 0.0847$ (0.0307)	$\widehat{\beta}_{freq} + \widehat{\beta}_{sev} = 0.0847$
VehicleGroup Type 2	-	$\widehat{\beta}_{sev} = 0.4706$ (0.06)	$\widehat{\beta}_{freq} + \widehat{\beta}_{sev} = 0.4706$
() คือ ค่า Standard Error (SE)			

จากตาราง 4.10 จะได้ตัวแบบประมาณเบี่ยงแปรกันภัยที่แท้จริงดังนี้

$$E \left(\sum_{k=1}^{N_i} C_{ik} \right) \\ = \text{Exp}[10.029 - 0.0046x_{Age} - 0.0036x_{VehAge} + (3.665 * 10^{-7})x_{Band} \\ + 0.4706x_{Type2} + 0.0852x_{Type3} + 0.0847x_{Type4}]$$

จากตัวแบบข้างต้น สามารถประมาณเบี่ยงแปรกันภัยที่แท้จริง ได้ดังตัวอย่างต่อไปนี้

ข้อมูลกรมธรรม์ Age = 25 , VehicleAge = 1 , VehicleSizeCode = <2000 CC , BandValue = 4,900,000.5 , VehicleGroup = Type

1

$$E \left(\sum_{k=1}^{N_i} C_{ik} \right) \\ = \text{Exp}[10.029 - 0.0046(25) - 0.0036(1) + (3.665 * 10^{-7})(4,900,000.5) \\ + 0.4706(0) + 0.0852(0) + 0.0847(0)]$$

= 121,327.67 บาท

4.4 ผลการวัดความเหมาะสมของตัวแบบ

ตารางที่ 4.11 แสดงผลลัพธ์ที่ได้จากการหาค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ที่ทดสอบกับ Test Set ของ Claim Frequency Model และ Claim Severity Model ในแต่ละปี

ตัวแบบ	ค่า MSE
FrequencyModel2016	12.23725
FrequencyModel2017	1.186171
SeverityModel2016	2,605,417,228
SeverityModel2017	16,967,445,967

จากตารางที่ 4.11 จะแปลผลได้ดังนี้

1. FrequencyModel2017 มีความคลาดเคลื่อนที่น้อยกว่า FrequencyModel2016
2. SeverityModel2016 มีความคลาดเคลื่อนที่น้อยกว่า SeverityModel2016

ตารางที่ 4.12 แสดงผลลัพธ์ Residual Deviance ที่ได้จาก Claim Frequency Model และ Claim Severity Model ในแต่ละปี

ตัวแบบ	ค่า Residual Deviance
FrequencyModel2016	17,642
FrequencyModel2017	12,844
SeverityModel2016	81,971
SeverityModel2017	54,684

จากตารางที่ 4.12 จะเห็นว่าค่า Residual Deviance ของโมเดลทั้งสองในปี 2017 คือ FrequencyModel2017 และ SeverityModel2017 มีค่าน้อยกว่าโมเดลในปี 2016 จึงแปลผลได้ว่าโมเดลทั้งสองในปี 2017 มีความเหมาะสมมากกว่า โมเดลในปี 2016

บทที่ 5

สรุปผลการวิจัย อภิปราย และข้อเสนอแนะ

การศึกษาวิจัย เรื่อง การประมาณเบี้ยประกันภัยที่แท้จริงโดยใช้ตัวแบบเชิงเส้นวางนัยทั่วไป มีวัตถุประสงค์เพื่อศึกษาวิธีการประมาณเบี้ยประกันภัยที่แท้จริงโดยใช้ตัวแบบเชิงเส้นวางนัยทั่วไป และ เพื่อศึกษาปัจจัยที่มีผลต่อความถี่และความรุนแรงของการเรียกร้องค่าสินไหมทดแทน โดยใช้ข้อมูลอุบัติเหตุที่ได้จากฐานข้อมูลกรมธรรม์ประกันอุบัติเหตุรถยนต์ภาคสมัครใจ ชั้น 1 ประเภทรถยนต์นั่ง (รหัส 110 และ 120) ของสำนักงานอัตรเบี้ยประกันวินาศภัยที่เริ่มสัญญาในปี พ.ศ.2559 และ 2560 จำนวน 61,710 และ 47,997 กรมธรรม์ ตามลำดับ โดยขั้นตอนการสร้างตัวแบบเบี้ยประกันที่แท้จริงของแต่ละปีจะประกอบไปด้วยตัวแบบความถี่ในการเรียกร้องค่าสินไหมทดแทน (Claim Frequency Model) และ ตัวแบบความรุนแรงในการเรียกร้องค่าสินไหมทดแทน (Claim Severity Model) ซึ่งผลการวิจัย อภิปรายผล และข้อเสนอแนะ สามารถสรุปได้ดังนี้

5.1 สรุปและอภิปรายผลการวิเคราะห์ปัจจัยเสี่ยงที่มีผลต่อตัวแบบ

สำหรับปี 2559

ตัวแบบความถี่ในการเรียกร้องค่าสินไหมทดแทน (Claim Frequency Model) มีเพียงปัจจัยเสี่ยง Age, VehicleAge และ VehicleSizeCode ที่มีผลกระทบต่อความถี่ในการเรียกร้องค่าสินไหมทดแทน ที่ระดับนัยสำคัญ 0.05

ตัวแบบความรุนแรงในการเรียกร้องค่าสินไหมทดแทน (Claim Severity Model) ปัจจัยเสี่ยง Age, VehicleAge, VehicleSizeCode, VehicleGroupType และ BandValue ที่มีผลกระทบต่อความรุนแรงในการเรียกร้องค่าสินไหมทดแทน ที่ระดับนัยสำคัญ 0.05

สำหรับปี 2560

ตัวแบบความถี่ในการเรียกร้องค่าสินไหมทดแทน (Claim Frequency Model) มีเพียงปัจจัยเสี่ยง Age และ VehicleAge ที่มีผลกระทบต่อความถี่ในการเรียกร้องค่าสินไหมทดแทน ที่ระดับนัยสำคัญ 0.05

ตัวแบบความรุนแรงในการเรียกร้องค่าสินไหมทดแทน (Claim Severity Model) ปัจจัยเสี่ยง Age, VehicleAge, VehicleGroupType และ BandValue ที่มีผลกระทบต่อความรุนแรงในการเรียกร้องค่าสินไหมทดแทน ที่ระดับนัยสำคัญ 0.05

จะเห็นว่าตัวแบบของ Claim Frequency Model และ Claim Severity Model ทั้งสองปี มีปัจจัยเสี่ยงที่มีอิทธิพลเหมือนกัน นอกจากนี้จะเห็นว่า ลักษณะของข้อมูลที่ใช้ทำตัวแบบของทั้ง Claim Frequency Model และ Claim Severity Model เป็นไปในทิศทางเดียวกัน นั่นคือ สัดส่วนหน่วยเสี่ยงภัย (Earned exposure) ของแต่ละตัวแปรมีค่าใกล้เคียงกัน

ดังนั้นการประมาณเบี้ยประกันภัยที่แท้จริงควรคำนึงถึงปัจจัยเสี่ยงดังกล่าว ซึ่งผลการศึกษาจะเป็นแนวทางให้กับบริษัทประกันภัยนำไปประยุกต์ใช้กับข้อมูลที่รับความเสี่ยงเพื่อสร้างตัวแบบการประมาณเบี้ยประกันภัยที่แท้จริง ซึ่งผลการวิเคราะห์ข้อมูลของแต่ละบริษัทอาจให้ค่าปัจจัยเสี่ยงที่มีผลกระทบต่อค่าประมาณเบี้ยประกันภัยที่แท้จริงที่แตกต่างไปจากผลการศึกษาเนื่องจากลักษณะข้อมูลของการรับประกันความเสี่ยงมีความแตกต่างกัน

5.2 ข้อเสนอแนะ

ในการวิจัยครั้งนี้ ผู้วิจัยได้ทำการศึกษาและวิเคราะห์การประมาณเบี้ยประกันภัยที่แท้จริงสำหรับการประกันภัยรถยนต์โดยใช้ตัวแบบเชิงเส้นวงนัยทั่วไป ซึ่งในการที่จะสร้างตัวแบบเบี้ยประกันภัยที่แท้จริง (Pure Premium Model) ได้นั้น จะต้องสร้างตัวแบบ 2 ตัวแบบ ขึ้นมาก่อน คือ

1. ตัวแบบความถี่ในการเรียกร้องค่าสินไหมทดแทน (Claim Frequency Model) ซึ่งผู้วิจัยเลือกใช้เป็น Poisson Regression Model
2. ตัวแบบความรุนแรงในการเรียกร้องค่าสินไหมทดแทน (Claim Severity Model) ผู้วิจัยเลือกใช้เป็น Gamma Regression Model

จากข้อดกลงเบื้องต้นของ Gamma Regression Model ที่ตัวแปรตอบสนองต้องมีค่ามากกว่า 0 ทำให้ผู้วิจัยจำเป็นต้องตัด record ที่ Claim Severity น้อยกว่าหรือเท่ากับ 0 ทั้งใน Claim Severity Model และ Claim Frequency Model ทิ้งไป เพื่อให้มีจำนวน record ที่เท่ากัน ทั้งนี้หากมีโอกาสในการวิจัยครั้งต่อไป ควรที่จะมีการศึกษาการนำข้อมูล ที่ Claim Severity น้อยกว่าหรือเท่ากับ 0 มาปรับตัวแบบในภายหลัง เพื่อให้เบี้ยประกันภัยที่แท้จริงจากตัวแบบสะท้อนถึงการได้รับคืนค่าสินไหมทดแทนของบริษัท ($\text{Claim Severity} < 0$) และกรณีธรรมที่ ไม่มีการเรียกร้องค่าสินไหมจริง ๆ ($\text{Claim Severity} = 0$)

อีกกรณีคือ ข้อมูลตัวแปรตอบสนองของ Claim Frequency Model มีลักษณะ Over disperse นั่นคือ ข้อมูลมีความแปรปรวนมากกว่าค่าเฉลี่ย ซึ่งไม่สอดคล้องกับข้อดกลงเบื้องต้นของ Poisson Regression Model ที่ความแปรปรวนต้องเท่ากับค่าเฉลี่ย ทำให้การ fitting ตัวแบบมีความคลาดเคลื่อน แต่อย่างไรก็ตาม ยังมีอีกหลายตัวแบบที่เหมาะสมจะใช้กับ Claim Frequency Model ยกตัวอย่างเช่น Negative Binomial Regression Model ซึ่งยังไม่ได้ศึกษาหรือนำมาวิเคราะห์ในงานวิจัยครั้งนี้ และควรที่จะมีการศึกษาในการวิจัยครั้งต่อไป

รายการอ้างอิง

David M. (2015) . Auto Insurance Premium Calculation Using Generalized Models. Retrieve November 1 ,2018.

<https://www.sciencedirect.com/science/article/pii/S2212567115000593>

Goldburd, M. and Tevet, D. (2016). Generalized Linear Model for Insurance rating. Retrieve March 4,2019.

<https://www.casact.org/pubs/monographs/papers/05-Goldburd-Khare-Tevet.pdf>

Hastie, T. and Tibshirani, R. (2013) . Model Comparison and Out of Sample Prediction. Retrieve April 18,2019.

<https://www2.stat.duke.edu/courses/Spring17/sta521/knitr/Lec-9-Selection/Resampling.pdf>

Zhou, J. and Garrido, J. (2009). A Loss Reserving Method Based on Generalized Linear Models. Retrieve September 11,2018.

<https://www.researchgate.net/publication/320559204/download>

ชนกนาค สุขประยูร และคณะ (2560). การใช้ตัวแบบเชิงเส้นวางนัยทั่วไปในการประมาณค่าเบี้ยประกันภัย

รถยนต์, 5 พฤศจิกายน 2561.

<http://journal.maththai.org/ejournal/index.php/math/article/viewFile/33/24>

วีรณันท์ พงศาภักดิ์. (2555). การวิเคราะห์ข้อมูลจำแนกประเภท : ทฤษฎีและการประยุกต์ด้วยGLIM,SPSS,SAS และ MTB. นครปฐม: โรงพิมพ์มหาวิทยาลัยศิลปากร.

ปวีรศา สุขเรื่อย และ สำรวม จงเจริญ. (2561). ตัวแบบถดถอยที่มีผลกระทบค่าศูนย์ประยุกต์ใช้กับจำนวนครั้งของการเรียกร้องค่าสินไหมทดแทนในการประกันภัยรถยนต์ภาคสมัครใจ, 15 เมษายน 2562.

<https://www.tci-thaijo.org/index.php/swujournal/article/view/170333/122433>

สุทธาสินี จันทร์สมบูรณ์. (2552). การพยากรณ์เบี้ยประกันภัยแท้จริงของการประกันภัยรถยนต์, 11 กันยายน2561.

<http://proceedings.bu.ac.th/index.php/com-phocadownload-controlpanel/grc?download=199>

สำนักงานอตราเบี้ยประกันวินาศภัย. (2559). คู่มือสำหรับผู้ปฏิบัติงานด้านคณิตศาสตร์ประกันภัย การสร้างตัว แบบเชิงเส้นวางนัยทั่วไป, 11 กันยายน 2561.

https://www.tgia.org/upload/file_group/3/download_861.pdf