IPRB

คู่มือสำหรับผู้ปฏิบัติงานด้านคณิตศาสตร์ประกันภัย

การสร้างตัวแบบเชิงเส้นวางนัยทั่วไป

(Generalized Linear Models : GLMs)

คู่มือสำหรับผู้ปฏิบัติงาน

ด้านคณิตศาสตร์ประกันภัย

การสร้างตัวแบบเชิงเส้นวางนัยทั่วไป

(Generalized Linear Models)

IPRB
สำนักงานอัตราเบี้ยประกันวินาศภัย
The Insurance Premium Rating Bureau

คำนำ

ตัวแบบเชิงเส้นวางนัยทั่วไป (Generalized Linear Models) มีความสำคัญอย่างมากต่อการคำนวณสถิติที่ใช้สำหรับการดำเนินธุรกิจประกันภัย โดยเฉพาะอย่างยิ่งการคำนวณอัตราต้นทุนความเสียหาย โดยตัวแบบนี้สามารถขจัดปัญหาที่เกิดจากสหสัมพันธ์ระหว่างตัวแปรต้นที่อาจมีความสัมพันธ์ซึ่งกันและกัน และยังสามารถใช้ตรวจสอบผลกระทบจากความสัมพันธ์ซึ่งกันและกันของตัวแปรต้น และสามารถให้ผลการวิเคราะห์ที่บ่งบอกระดับความเชื่อมั่นทางสถิติหรือผลการประเมินคุณภาพของตัวแบบได้ด้วย

สำนักงานอัตราเบี้ยประกันวินาศภัยได้เล็งเห็นถึงความสำคัญของการสร้างตัวแบบดังกล่าว และเห็นว่าหนังสือตำราที่เกี่ยวข้องในปัจจุบันมีเฉพาะในส่วนของทฤษฎี จึงได้จัดทำหนังสือคู่มือฉบับนี้ขึ้น โดยผนวกวิธีการทำงานตั้งแต่การเตรียมข้อมูล ปัญหาที่มักพบ ชุดคำสั่งในการสร้างตัวแบบเชิงเส้นวางนัยทั่วไปด้วยโปรแกรม Statistical Analysis System (SAS) รวมถึงแนวทางการวิเคราะห์ โดยมีวัตถุประสงค์เพื่อให้ผู้ปฏิบัติงานด้านคณิตศาสตร์ประกันภัยสามารถใช้คู่มือฉบับนี้เป็นแนวทางในการสร้างตัวแบบเชิงเส้นวางนัยทั่วไปได้ด้วยตัวเอง สอดคล้องกับแนวทางการพัฒนางานด้านคณิตศาสตร์ประกันภัย ซึ่งเป็นหนึ่งในภารกิจสำคัญของสำนักงานฯ ที่สนับสนุนการพัฒนาอุตสาหกรรมประกันภัยอย่างยั่งยืน

สำนักงานฯ ขอขอบคุณ คุณจักรพันธ์ เหลืองนฤมิตชัย, FCAS เป็นอย่างสูง ที่ให้ความคิดเห็น พร้อมทั้งตรวจทานเนื้อหาของคู่มือให้มีความสมบูรณ์ครบถ้วน และขอขอบคุณ รศ.ดร.ฐิติวดี ชัยวัฒน์ ที่ช่วยร่วมให้ความคิดเห็น พร้อมทั้งปรับปรุงสำนวนการเขียนคู่มือปฏิบัติงานฉบับนี้ ให้มีความสละสลวยและง่ายต่อการอ่านมากยิ่งขึ้น

สำนักงานฯ หวังเป็นอย่างยิ่งว่า คู่มือนี้จะเป็นประโยชน์ต่อการทำงานของนักคณิตศาสตร์ประกันภัย และหากมีข้อบกพร่องประการใด ขอน้อมรับความคิดเห็นเพื่อนำไปปรับปรุงให้คู่มือปฏิบัติงานมีความสมบูรณ์มากยิ่งขึ้นต่อไป

สำนักงานอัตราเบี้ยประกันวินาศภัย

มิถุนายน 2559

สารบัญ

บทที่ 1 บทนำ 8

1.1. ที่มาและความสำคัญในการจัดทำคู่มือปฏิบัติงาน 8

1.2. กรอบของคู่มือปฏิบัติงาน 8

บทที่ 2 ความรู้เบื้องต้นเกี่ยวกับตัวแบบเชิงเส้นวางนัยทั่วไป (Generalized Linear Models) 10

2.1. ตัวแบบที่เกี่ยวข้อง และสมมติฐานของตัวแบบ 10

2.1.1 ตัวแบบเชิงเส้น (Linear Model) 10

2.1.2 ตัวแบบเชิงเส้นแบบดั้งเดิม (Classical Linear Model) 11

2.1.3 ตัวแบบเชิงเส้นวางนัยทั่วไป (Generalized Linear Models) 12

2.2. วิธีการทดสอบความมีนัยสำคัญของตัวแบบ 19

2.2.1 การวินิจฉัยตัวแบบ (Model Diagnostics) 19

2.2.2 Deviance Test 19

2.3. ความเหมาะสม และข้อจำกัดของการใช้ตัวแบบเชิงเส้นวางนัยทั่วไป 21

2.4. ตัวอย่างการประยุกต์ใช้ตัวแบบเชิงเส้นวางนัยทั่วไปในงานคณิตศาสตร์ประกันภัย 22

2.4.1 รูปแบบของตัวแบบอัตราการต่ออายุ 23

2.4.2 ตัวอย่างผลการวิเคราะห์ 23

2.5. ตัวอย่าง Software ที่ใช้สร้างตัวแบบเชิงเส้นวางนัยทั่วไป 24

บทที่ 3 ขั้นตอนการสร้างตัวแบบเชิงเส้นวางนัยทั่วไป 26

3.1. ภาพรวมของกระบวนการสร้างตัวแบบเชิงเส้นวางนัยทั่วไป 26

3.2. การเตรียมข้อมูล (Data Preparation) 27

3.2.1 ลักษณะของข้อมูล 27

3.2.2 การตรวจสอบความสมเหตุสมผลของข้อมูล 30

3.2.3 กรณีศึกษา ปัญหาที่มักพบในขั้นตอนการเตรียมข้อมูล และแนวทางการแก้ไขปัญหา 30

3.3. การวิเคราะห์ข้อมูลเบื้องต้น 32

3.3.1 การวิเคราะห์การแจกแจงของข้อมูล 32

3.3.2 การวิเคราะห์ทางเดียว (One-Way Analysis) 36

3.3.3 การวิเคราะห์ความมีสหสัมพันธ์ระหว่างตัวแปร (Correlation Analysis) 41

3.4. การสร้างตัวแบบเชิงเส้นวางนัยทั่วไป 46

3.4.1 การเตรียมตัวแปรก่อนการสร้างตัวแบบ 46

3.4.2 ชุดคำสั่งในการสร้างตัวแบบเชิงเส้นวางนัยทั่วไปด้วยโปรแกรม Statistical Analysis System (SAS) 47

3.4.3 การวิเคราะห์ผลและการตรวจสอบความมีนัยสำคัญของตัวแบบ (Model Validation) 50

3.4.4 การรวมตัวแบบเพื่อหาต้นทุนที่แท้จริง (Loss Cost Model) 56

บทที่ 4 การปรับปรุงตัวแบบ 59

4.1. การปรับปรุงตัวแบบเชิงเส้นวางนัยทั่วไปด้วยความเสียหายที่เกิดขึ้นแล้วแต่ยังไม่ได้มีการรายงานให้ทราบ (IBNR) และแนวโน้ม 59

4.2. การปรับปรุงตัวแบบเพื่อให้มีความสม่ำเสมอ (Smoothing) 61

4.3. การใส่ข้อจำกัดให้ตัวแบบ (Restriction) 63

บทที่ 5 ภาคผนวกและเอกสารอ้างอิง 67

5.1. รูปแบบการแจกแจงที่สำคัญ 67

5.1.1 การแจกแจงแบบปกติ (Normal Distribution) 67

5.1.2 การแจกแจงพัวซองค์ (Poisson Distribution) 67

5.1.3 การแจกแจงแบบทวินาม (Binomial Distribution) 68

5.1.4 การแจกแจงแกมมา (Gamma Distribution) 68

5.1.5 การแจกแจงแบบทวิดี (Tweedie Distribution) 69

5.2. เอกสารอ้างอิง 70

สารบัญตาราง

ตารางที่ 1 ค่าของตัวแปรต้นตามเพศของผู้ขับขี่และสถานที่ขับขี่รถยนต์.. 10
ตารางที่ 2 ตัวอย่างความแปรปรวนของการแจกแจงที่อยู่ในวงศ์ชี้กำลัง (Exponential Family)............... 13
ตารางที่ 3 รูปแบบฟังก์ชันเชื่อมโยงที่ใช้บ่อย 14
ตารางที่ 4 สรุปรูปแบบทั่วไปของตัวแบบเชิงเส้นวางนัยทั่วไป 16
ตารางที่ 5 จำนวนหน่วยเสี่ยงภัยของจำนวนประตูกับสีของรถยนต์.. 18
ตารางที่ 6 สรุปข้อดี ข้อเสียของตัวแบบเชิงเส้นวางนัยทั่วไป 21
ตารางที่ 7 การแจกแจงที่เหมาะสมกับตัวแบบแต่ละประเภท.. 33
ตารางที่ 8 คำสั่งสำหรับระบุการแจกแจงในตัวแบบ 49
ตารางที่ 9 คำสั่งสำหรับระบุฟังก์ชันเชื่อมโยงในตัวแบบ.. 49
ตารางที่ 10 ตัวอย่างการคำนวณเบี้ยประกันภัยพื้นฐานและค่าสัมประสิทธิ์ของตัวแบบต้นทุนที่แท้จริง 57
ตารางที่ 11 ตัวอย่างการพยากรณ์ต้นทุนที่แท้จริงที่ได้จากตัวแบบเชิงเส้นวางนัยทั่วไป 58
ตารางที่ 12 ตัวอย่างการคำนวณเบี้ยประกันภัยพื้นฐานที่ปรับด้วย IBNR แล้ว.. 60
ตารางที่ 13 ตัวอย่างการคำนวณเบี้ยประกันภัยพื้นฐานที่ปรับด้วย IBNR และแนวโน้มของต้นทุนที่เพิ่มขึ้นแล้ว 61

สารบัญภาพ

รูปที่ 1 อัตราการขาดต่ออายุ 23

รูปที่ 2 ผลกระทบของส่วนลด/ส่วนเพิ่มเบี้ยประกันภัยต่ออัตราการขาดต่ออายุ 24

รูปที่ 3 ตัวเลือกเครื่องมือ Distribution Analysis 33

รูปที่ 4 หน้าต่าง Data ของเครื่องมือ Distribution Analysis 34

รูปที่ 5 หน้าต่าง Distribution > Summary ของเครื่องมือ Distribution Analysis 35

รูปที่ 6 ตารางแสดงผลการทดสอบการแจกแจง 36

รูปที่ 7 การเลือกเครื่องมือ Summary Tables เพื่อการวิเคราะห์ทางเดียว 37

รูปที่ 8 หน้าต่าง Data ของ Summary Tables 38

รูปที่ 9 หน้าต่าง Summary Tables ของ Summary Tables 39

รูปที่ 10 ตัวอย่างหน้าจอแสดงผล Summary Tables 40

รูปที่ 11 แผนภาพความสัมพันธ์ของอายุรถยนต์ (ปี) ในส่วนความเสียหายตัวรถยนต์ (Own Damage) 41

รูปที่ 12 การเรียกใช้เครื่องมือ Correlations 42

รูปที่ 13 หน้าต่างของเครื่องมือ Correlations 43

รูปที่ 14 หน้าจอแสดงผลวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร 44

รูปที่ 15 ตัวอย่างชุดคำสั่ง PROC FREQ สำหรับการทดสอบสหสัมพันธ์ระหว่างตัวแปรที่ไม่ได้มีค่าในเชิงตัวเลข (non-numerical) 45

รูปที่ 16 หน้าจอแสดงผลวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรด้วยตัวสถิติ Cramer’s V 45

รูปที่ 17 ตัวอย่างการระบุให้ระดับขั้นของตัวแปรใดๆ เป็นระดับพื้นฐาน (Base Level) 47

รูปที่ 18 ตัวอย่างชุดคำสั่ง GENMOD เพื่อการสร้างตัวแบบเชิงเส้นวางนัยทั่วไป 48

รูปที่ 19 ตารางแสดงผลชุดคำสั่ง GENMOD ของโปรแกรม SAS 51

รูปที่ 20 ตัวอย่างตารางการวิเคราะห์ Type 1 (ผลหลังการจัดกลุ่ม) 53

รูปที่ 21 ตัวอย่างตารางการวิเคราะห์ Type 3 และตัวแปรที่ระดับนัยสำคัญไม่เพียงพอ (ผลก่อนการจัดกลุ่ม) 54

รูปที่ 22 ตารางการวิเคราะห์ Maximum Likelihood Parameter และระดับของตัวแปรที่ระดับนัยสำคัญไม่เพียงพอ (ผลก่อนการจัดกลุ่ม) 55

รูปที่ 23 Relativity ของความรุนแรงในการชนโดยเฉลี่ยตามอายุรถยนต์ 62

รูปที่ 24 การปรับ relativity ของความรุนแรงในการชนโดยเฉลี่ยสำหรับรถยนต์อายุ 6 ปี 63

รูปที่ 25 การปรับปรุงตัวแบบเชิงเส้นวางนัยทั่วไปด้วยการใส่ข้อจำกัดลงในปัจจัย “Factor 1”.................. 64
รูปที่ 26 ผลกระทบจากการปรับปรุงตัวแบบเชิงเส้นวางนัยทั่วไปต่อปัจจัย “Factor 2” ด้วยการใส่ข้อจำกัดลงในปัจจัย “Factor 1”.. 65
รูปที่ 27 ผลกระทบจากการปรับปรุงตัวแบบเชิงเส้นวางนัยทั่วไปต่อปัจจัย “Factor 3” ด้วยการใส่ข้อจำกัดลงในปัจจัย “Factor 1”.. 65
รูปที่ 28 ผลกระทบจากการปรับปรุงตัวแบบเชิงเส้นวางนัยทั่วไปต่อปัจจัย “Factor 4” ด้วยการใส่ข้อจำกัดลงในปัจจัย “Factor 1”.. 66
รูปที่ 29 กราฟการแจกแจงแบบ Tweedie.. 69

บทที่ 1 บทนำ

1.1. ที่มาและความสำคัญในการจัดทำคู่มือปฏิบัติงาน

รูปแบบการวิเคราะห์ข้อมูลแบบดั้งเดิมในอุตสาหกรรมการประกันภัยนั้นไม่ได้มีความซับซ้อนเชิงสถิติมากนัก โดยทั่วไปจะใช้การวิเคราะห์แบบทางเดียว (One-Way Analysis) หรือการวิเคราะห์แบบสองทาง (Two-Way Analysis) เป็นหลัก ซึ่งมีข้อดีคือ ง่ายต่อการคำนวณ แต่วิธีดังกล่าวก็มีข้อเสียคือ ผลของการวิเคราะห์นั้นอาจถูกบิดเบือน เนื่องจากวิธีดังกล่าวนั้นไม่ได้คำนึงถึงสหสัมพันธ์ (Correlation) ระหว่างตัวแปรและไม่ได้คำนึงถึงผลกระทบความสัมพันธ์ซึ่งกันและกันระหว่างตัวแปร (Interdependencies)

ตัวแบบเชิงเส้นวางนัยทั่วไป (Generalized Linear Models หรือ GLMs) ถูกนำเสนอครั้งแรกในปี ค.ศ. 1972 โดย John Nelder และ Robert Wedderburn แนวคิดของตัวแบบคือ การหาความสัมพันธ์ระหว่างตัวแปรต้น (Explanatory Variables) และตัวแปรตาม (Response Variable) โดยตัวแบบนี้สามารถขจัดปัญหาที่เกิดจากสหสัมพันธ์ระหว่างตัวแปรต้นที่อาจมีความสัมพันธ์ซึ่งกันและกัน และนอกจากนั้นยังสามารถใช้ตรวจสอบผลกระทบจากความสัมพันธ์ซึ่งกันและกันของตัวแปรต้น นอกจากนี้ตัวแบบเชิงเส้นวางนัยทั่วไปยังให้ผลการวิเคราะห์ที่บ่งบอกระดับความเชื่อมั่นทางสถิติหรือผลการประเมินคุณภาพของตัวแบบได้ด้วย

ตัวแบบเชิงเส้นวางนัยทั่วไปมีความสำคัญอย่างมากต่อการคำนวณสถิติที่ใช้สำหรับการดำเนินธุรกิจประกันภัย เช่น การคำนวณอัตราต้นทุนความเสียหายซึ่งเป็นองค์ประกอบหนึ่งของอัตราเบี้ยประกันภัย สำนักงานอัตราเบี้ยประกันวินาศภัยได้เล็งเห็นถึงความสำคัญของตัวแบบดังกล่าวจึงได้จัดทำหนังสือคู่มือฉบับนี้ขึ้นมาโดยมีวัตถุประสงค์เพื่อให้ผู้ปฏิบัติงานด้านคณิตศาสตร์ประกันภัยสามารถใช้คู่มือฉบับนี้เป็นแนวทางในการสร้างตัวแบบเชิงเส้นวางนัยทั่วไปด้วยตัวเอง

เพื่อให้ได้ประโยชน์สูงสุดในการใช้งานคู่มือปฏิบัติงานฉบับนี้ ผู้ใช้คู่มือควรมีความรู้ในวิชาแคลคูลัสและทฤษฎีทางสถิติ แต่อย่างไรก็ตาม ผู้อ่านที่ไม่มีพื้นฐานดังกล่าวสามารถทำความเข้าใจกับทฤษฎีที่เกี่ยวข้องกับตัวแบบเชิงเส้นวางนัยทั่วไปได้ตามขั้นตอนที่คู่มือนี้ได้ระบุไว้ในบทที่ 2

1.2. กรอบของคู่มือปฏิบัติงาน

คู่มือปฏิบัติงานฉบับนี้ประกอบไปด้วย 5 บท ดังต่อไปนี้

1. บทนำ

   อธิบายถึงที่มาและความสำคัญของการจัดทำและกรอบของคู่มือปฏิบัติงานฉบับนี้

2. ความรู้เบื้องต้นเกี่ยวกับตัวแบบเชิงเส้นวางนัยทั่วไป

   อธิบายถึงความรู้พื้นฐานที่เกี่ยวกับตัวแบบเชิงเส้นวางนัยทั่วไป ซึ่งรวมถึงตัวแบบที่เกี่ยวข้อง สมมติฐานของตัวแบบ วิธีการทดสอบความมีนัยสำคัญของตัวแบบ ความเหมาะสมของตัวแบบ ข้อจำกัดของการใช้ตัวแบบ ตัวอย่างการประยุกต์ใช้ตัวแบบเชิงเส้นวางนัยทั่วไปในงานคณิตศาสตร์ประกันภัย และ Software ที่นิยมนำมาใช้ในการสร้างตัวแบบ

3. ขั้นตอนการสร้างตัวแบบเชิงเส้นวางนัยทั่วไป

   อธิบายถึงขั้นตอนในการสร้างตัวแบบ ซึ่งรวมถึงข้อมูลขั้นต่ำที่ต้องใช้ในการสร้างตัวแบบ วิธีการจัดการข้อมูล วิธีการตรวจสอบความถูกต้องก่อนการนำมาสร้างตัวแบบ ปัญหาที่มักพบในการเตรียมข้อมูลและแนวทางในการแก้ไขปัญหา การวิเคราะห์ข้อมูลเบื้องต้น การสร้างตัวแบบเชิงเส้นวางนัยทั่วไปด้วยชุดคำสั่งจากโปรแกรม SAS การวิเคราะห์ผลที่ได้จากโปรแกรม SAS ปัญหาที่มักพบในการสร้างตัวแบบ และแนวทางในการแก้ปัญหา

4. การปรับปรุงตัวแบบเชิงเส้นวางนัยทั่วไป

   อธิบายถึงวิธีที่ใช้สำหรับการปรับปรุงตัวแบบเพื่อให้มีความสม่ำเสมอ และวิธีในการใส่ข้อจำกัดให้ตัวแบบ

5. ภาคผนวกและเอกสารอ้างอิง

   อธิบายถึงการแจกแจงรูปแบบต่างๆที่ควรทราบ และเอกสารอ้างอิงต่างๆ ที่ใช้ในการจัดทำคู่มือฉบับนี้

บทที่ 2 ความรู้เบื้องต้นเกี่ยวกับตัวแบบเชิงเส้นวางนัยทั่วไป (Generalized Linear Models)

2.1. ตัวแบบที่เกี่ยวข้อง และสมมติฐานของตัวแบบ

2.1.1 ตัวแบบเชิงเส้น (Linear Model)

จุดประสงค์ของการใช้ตัวแบบเชิงเส้น (Linear Model) และตัวแบบเชิงเส้นวางนัยทั่วไป (Generalized Linear Models) คือ การหาความสัมพันธ์ระหว่างตัวแปรต้น (Explanatory Variables) กับตัวแปรตาม (Response Variable) โดยมีแนวคิดว่า ตัวแปรตาม (Y) สามารถเขียนอยู่ในรูปของผลรวมของค่าเฉลี่ย (μ) กับความคลาดเคลื่อน (ε) ได้ ซึ่งสามารถเขียนให้อยู่ในรูปของสมการที่ (2.1) ดังนี้

$$Y = \mu + \varepsilon \tag{2.1}$$

สมมติฐานของการใช้ตัวแบบเชิงเส้น ในสมการที่ (2.1) คือ

- μ ค่าคาดหมายของ Y สามารถเขียนให้อยู่ในรูปของสมการเส้นตรงได้
- ε มีการแจกแจงแบบปกติ (Normal Distribution) ที่มีค่าเฉลี่ยเท่ากับ 0 และมีความแปรปรวนเท่ากับ σ^2

ตัวอย่างของการใช้ตัวแบบเชิงเส้น เช่น ให้ Y เป็นค่าเฉลี่ยของจำนวนครั้งของการเรียกร้องค่าสินไหมทดแทน (Claims Frequency) โดยมีตัวแปรต้น (X) จำนวน 2 ตัวแปร คือ 1) เพศของผู้ขับขี่ (“ผู้ชาย” และ “ผู้หญิง”) และ 2) สถานที่ขับขี่รถยนต์ (“ในเมือง” และ “นอกเมือง”)

เราสามารถเขียน Y ให้อยู่ในรูปผลรวมของตัวแปรต้น (X) ซึ่งอยู่ในรูปแบบของสมการเส้นตรงตามรูปแบบของตัวแบบเชิงเส้นได้ดังสมการที่ (2.2)

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon \tag{2.2}$$

และสามารถกำหนดค่าของตัวแปรต้น (X) ได้ดังตารางที่ 1

ตารางที่ 1 ค่าของตัวแปรต้นตามเพศของผู้ขับขี่และสถานที่ขับขี่รถยนต์

เพศของผู้ขับขี่	สถานที่ขับขี่รถยนต์	X_1	X_2	X_3
ชาย	นอกเมือง	1	0	0
ชาย	ในเมือง	1	0	1
หญิง	นอกเมือง	0	1	0
หญิง	ในเมือง	0	1	1

จากสมการ (2.2) เราจะหาค่า β_1, β_2 และ β_3 ที่ทำให้ผลรวมกำลังสองของความคลาดเคลื่อน (Sum Square Errors) มีค่าน้อยที่สุด โดยที่

β_1 คือ จำนวนครั้งเฉลี่ยของการเรียกร้องค่าสินไหมทดแทนจากการขับขี่รถยนต์นอกเมืองของผู้ชาย

β_2 คือ จำนวนครั้งเฉลี่ยของการเรียกร้องค่าสินไหมทดแทนจากการขับขี่รถยนต์นอกเมืองของผู้หญิง

β_3 คือ จำนวนครั้งเฉลี่ยของการเรียกร้องค่าสินไหมทดแทนที่เพิ่มขึ้นเนื่องจากการขับขี่รถยนต์ในเมือง

2.1.2 ตัวแบบเชิงเส้นแบบดั้งเดิม (Classical Linear Model)

สมการตัวแบบเชิงเส้นแบบดั้งเดิม (Classical Linear Model) สามารถเขียนได้ดังสมการที่ (2.3)

$$\underline{Y} = E[\underline{Y}] + \underline{\varepsilon} \quad (2.3)$$

โดยที่ $E[\underline{Y}] = \boldsymbol{X}\underline{\beta}$

ดังนั้น $$\underline{Y} = \boldsymbol{X}\underline{\beta} + \underline{\varepsilon} \quad (2.4)$$

สมมติฐานของตัวแบบเชิงเส้นแบบดั้งเดิมมีดังนี้

(1) องค์ประกอบแบบสุ่ม (Random Component)

ค่าของ $\underline{Y}$ เป็นอิสระต่อกัน และมีการแจกแจงแบบปกติ โดยที่ค่าเฉลี่ยของตัวแปร $\underline{Y}$ แต่ละตัว ไม่จำเป็นต้องมีค่าเท่ากัน แต่จะต้องมีค่าความแปรปรวนร่วมกัน (Common Variance) เท่ากับ σ^2

(2) องค์ประกอบแบบเป็นระบบ (Systematic Component)

สามารถเขียนตัวแปรต้นให้อยู่ในรูปของตัวประมาณเชิงเส้น η ได้ โดยที่ $\underline{\eta} = \boldsymbol{X}\underline{\beta}$

(3) ฟังก์ชันเชื่อมโยง (Link Function)

ความสัมพันธ์ระหว่างองค์ประกอบแบบสุ่มและองค์ประกอบแบบเป็นระบบถูกกำหนดโดยฟังก์ชันเชื่อมโยง โดยตัวแบบเชิงเส้นมีฟังก์ชันเชื่อมโยงแบบเอกลักษณ์ (Identity) โดยที่ $E[\underline{Y}] \equiv \underline{\mu} = \underline{\eta}$

ตัวแบบเชิงเส้นแบบดั้งเดิมนี้เป็นที่นิยมใช้กันอย่างแพร่หลายเนื่องจากเป็นตัวแบบที่ง่ายและไม่ซับซ้อนจนเกินไป มีเครื่องมือให้เลือกใช้หลากหลายและสามารถให้ผลข้อมูลได้รวดเร็ว แต่กลับมีข้อจำกัดอันเนื่องมาจาก ตัวแบบเชิงเส้นจะเป็นไปตามสมมติฐานที่กล่าวมาข้างต้นได้ยาก โดยทั่วไปเราจะทำการแปลงตัว

แปรเพื่อให้ตัวแบบเชิงเส้นเป็นไปตามสมมติฐาน แต่บางครั้งเราก็ไม่สามารถแปลงตัวแปรเพื่อทำให้ตัวแบบเชิงเส้นนั้นสอดคล้องกับสมมติฐานได้ จึงควรพิจารณาถึงความเหมาะสมของข้อมูลก่อนนำมาใช้สร้างตัวแบบเพื่อความถูกต้องแม่นยำของสมมติฐาน

2.1.3 ตัวแบบเชิงเส้นวางนัยทั่วไป (Generalized Linear Models)

ตัวแบบเชิงเส้นวางนัยทั่วไป (Generalized Linear Models: GLMs) เป็นชื่อที่ใช้เรียกกลุ่มของตัวแบบ ซึ่งรวมถึงตัวแบบเชิงเส้นที่กล่าวมาในข้อ 2.1.1 และ 2.1.2

สมมุติฐานของตัวแบบเชิงเส้นวางนัยทั่วไปมีดังนี้

(1) องค์ประกอบแบบสุ่ม (Random Component)

ค่าของ $\underline{\boldsymbol{Y}}$ เป็นอิสระต่อกัน และมีการแจกแจงแบบใดแบบหนึ่งในวงศ์ชี้กำลัง (Exponential Family)

(2) องค์ประกอบแบบเป็นระบบ (Systematic Component)

สามารถเขียนตัวแปรต้นให้อยู่ในรูปตัวประมาณเชิงเส้น η ได้ โดยที่ $\underline{\eta} = \boldsymbol{X}\,\underline{\beta}$

(3) ฟังก์ชันเชื่อมโยง (Link Function)

ความสัมพันธ์ระหว่างองค์ประกอบแบบสุ่มและองค์ประกอบแบบเป็นระบบถูกกำหนดด้วยฟังก์ชันเชื่อมโยง (g) โดยที่ฟังก์ชันเชื่อมโยงนั้นสามารถหาอนุพันธ์ได้ (Differentiable) และเป็นฟังก์ชันทางเดียว (Monotonic)

โดยที่
$$E[\underline{Y}] \equiv \underline{\mu} = g^{-1}(\underline{\eta})$$

การแจกแจงที่อยู่ในวงศ์ชี้กำลัง (Exponential Family) มีคุณสมบัติ 2 ประการ ดังนี้

(1) การแจกแจงสามารถเขียนได้ในรูปแบบของค่าเฉลี่ย และความแปรปรวน

(2) ความแปรปรวนเป็นฟังก์ชันของค่าเฉลี่ย

จากคุณสมบัติข้อที่ (2) เราสามารถเขียนให้อยู่ในรูปของสมการได้ดังสมการที่ (2.5)

$$Var(Y_i) = \frac{\phi V(\mu_i)}{\omega_i} \tag{2.5}$$

โดยที่ ϕ เป็นพารามิเตอร์ที่กำหนดขนาดของความแปรปรวน (Scaled Parameter) และ ω_i เป็นค่าคงที่ที่กำหนดน้ำหนักให้กับค่าสังเกต (Prior Weight) แต่ละตัว โดยการแจกแจงที่อยู่ในวงศ์ชี้กำลังและความแปรปรวนของการแจกแจงต่างๆ สามารถสรุปได้โดยย่อตามตารางที่ 2

ตารางที่ 2 ตัวอย่างความแปรปรวนของการแจกแจงที่อยู่ในวงศ์ชี้กำลัง (Exponential Family)

การแจกแจง	**ความแปรปรวน**
Normal	1
Poisson	x
Gamma	x^2
Binomial	$x(1-x)$ เมื่อจำนวนการทดลองเท่ากับ 1 ครั้ง
Inverse Gaussian	x^3

นอกจากตัวอย่างการแจกแจงที่แสดงให้เห็นในตารางด้านบนแล้ว การแจกแจงแบบ Tweedie ถือเป็นการแจกแจงที่สำคัญอีกตัวหนึ่งที่อยู่ในวงศ์ชี้กำลัง เนื่องจากการแจกแจงดังกล่าวเหมาะสำหรับนำมาใช้ในการสร้างตัวแบบที่ใช้ในการประมาณเบี้ยประกันภัยที่แท้จริง (Pure Premium) เนื่องจากการแจกแจงดังกล่าวมีมวลจำนวนหนึ่งอยู่ที่ค่า 0 ซึ่งสอดคล้องกับลักษณะของการแจกแจงของการเรียกร้องค่าสินไหมทดแทนที่มีหน่วยเสี่ยงภัยบางส่วนที่ไม่มีการเรียกร้องค่าสินไหมทดแทน

2.1.3.1 Prior Weight

Prior Weight คือ ข้อมูลที่แสดงถึงความน่าเชื่อถือของค่าสังเกตที่เราสามารถใส่ให้กับตัวแบบได้ โดยค่าสังเกตที่มีจำนวนหน่วยเสี่ยงภัยมากจะมีความแปรปรวนน้อย และเป็นค่าที่มีผลต่อตัวแบบมาก

กำหนดให้

m_{ik} เป็นจำนวนครั้งของการเรียกร้องค่าสินไหมทดแทนจากหน่วยเสี่ยงภัยที่ k และจากกลุ่มที่ i

ω_i เป็นจำนวนหน่วยเสี่ยงภัยของกลุ่มที่ i

Y_i เป็นความถี่ของการเรียกร้องค่าสินไหมทดแทนของกลุ่มที่ i

ดังนั้น รูปของสมการเป็นดังนี้

$$Y_i = \left(\frac{1}{\omega_i}\right) \sum_{k=1}^{\omega_i} m_{ik} \quad (2.6)$$

ยกตัวอย่างเช่น ถ้า m_{ik} มีการแจกแจงแบบพัวซองค์ (Poisson) ด้วยความถี่ของการเรียกร้องค่าสินไหมทดแทน f_i สำหรับทุกหน่วยเสี่ยงภัย

$$E[m_{ik}] = f_i = Var[m_{ik}] \quad (2.7)$$

ดังนั้น เมื่อหน่วยเสี่ยงภัยแต่ละหน่วยเป็นอิสระต่อกัน จะได้ว่า

$$\mu_i = E[Y_i] = \frac{1}{\omega_i} \sum_{k=1}^{\omega_i} E[m_{ik}] = \frac{1}{\omega_i} * \omega_i f_i = f_i \quad (2.8)$$

$$Var[Y_i] = \frac{1}{\omega_i^2} \sum_{k=1}^{\omega_i} Var[m_{ik}] = \frac{1}{\omega_i^2} * \omega_i f_i = \frac{1}{\omega_i} * f_i = \mu_i \left(\frac{1}{\omega_i}\right) \quad (2.9)$$

จากตัวอย่างดังกล่าว เมื่อนำไปเปรียบเทียบกับสมการที่ (2.5) จะเห็นได้ว่า $V(\mu_i) = \mu_i, \phi = 1$ และ Prior Weight ที่ใช้ในตัวอย่างที่กล่าวถึงนี้คือ จำนวนหน่วยเสี่ยงภัย (ω_i) ซึ่งในกรณีนี้ ค่าพารามิเตอร์ที่ใช้กำหนดขนาดของความแปรปรวน (Scaled Parameter) มีค่าเท่ากับ 1 และไม่มีผลต่อการวิเคราะห์

อย่างไรก็ดี ในกรณีทั่วไปเราจะประมาณค่าพารามิเตอร์ที่ใช้กำหนดขนาดของความแปรปรวนจากข้อมูลจริง แม้ว่าค่าพารามิเตอร์ที่ใช้กำหนดขนาดของความแปรปรวนจะไม่ได้ถูกใช้ในตัวแบบเชิงเส้นวางนัยทั่วไป แต่ก็ถูกใช้ในการคำนวณค่าสถิติที่สำคัญ เช่น ค่าความแปรปรวน

ในการใช้ตัวแบบเชิงเส้นวางนัยทั่วไป เราจะต้องทำการแปลงค่า Y เพื่อให้สอดคล้องกับสมมติฐานโดยการใช้ฟังก์ชันเชื่อมโยง (Link Function) เพื่อที่เราจะสามารถเขียน Y ให้อยู่ในรูปฟังก์ชันผลรวมของตัวแปรต้นได้ โดยที่ฟังก์ชันเชื่อมโยงจะต้องเป็นฟังก์ชันแบบทางเดียว (Monotonic) และสามารถหาอนุพันธ์ได้ (Differentiable) โดยรูปแบบฟังก์ชันเชื่อมโยงที่ใช้บ่อย สามารถสรุปรายละเอียดได้ตามตารางที่ 3

<u>ตารางที่ 3</u> รูปแบบฟังก์ชันเชื่อมโยงที่ใช้บ่อย

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

2.1.3.2 Offset Term

ในบางครั้งเราทราบล่วงหน้าถึงผลกระทบที่จะเกิดขึ้นกับตัวแบบที่เป็นผลมาจากตัวแปรต้น ซึ่งเราจำเป็นต้องใส่ Offset Term ในตัวแบบเพื่อขจัดผลกระทบที่จะเกิดขึ้นโดยการกำหนดให้ตัวประมาณเชิงเส้น **η** เป็นไปตามสมการที่ (2.10)

$$\underline{\eta} = \boldsymbol{X}\,\underline{\beta} + \underline{\xi} \tag{2.10}$$

จากสมการข้างต้น เราสามารถเขียนให้อยู่ในรูปแบบของสมการได้ดังแสดงในสมการที่ (2.11)

$$E[\underline{Y}] = \underline{\mu} = g^{-1}\left(\boldsymbol{X}\,\underline{\beta} + \underline{\xi}\right) \tag{2.11}$$

ในกรณีที่ต้องการประมาณจำนวนครั้งของการเรียกร้องค่าสินไหมทดแทนด้วยตัวแบบเชิงเส้นวางนัยทั่วไป เราทราบว่าค่าคาดหมายของจำนวนครั้งของการเรียกร้องค่าสินไหมทดแทน นั้นเป็นสัดส่วนกับจำนวนหน่วยเสี่ยงภัย (จำนวนหน่วยเสี่ยงภัยที่สูงขึ้นส่งผลให้เกิดจำนวนครั้งของการเรียกร้องค่าสินไหมทดแทนที่มากขึ้นด้วย) ดังนั้น จึงต้องใส่ข้อมูลดังกล่าวลงในตัวแบบเชิงเส้นวางนัยทั่วไปโดยการเพิ่ม Offset Term ลงในตัวแบบ ซึ่งสามารถเขียนได้ในรูปแบบของ Log ของจำนวนหน่วยเสี่ยงภัยของแต่ละค่าสังเกต ดังนี้

$$E[Y_i] = g^{-1}\left(\sum_j X_{ij}\beta_j + \xi_i\right) = \exp\left(\sum_j X_{ij}\beta_j + \ln(e_i)\right)$$

$$E[Y_i] = \exp\left(\sum_j X_{ij}\beta_j\right) e_i \tag{2.12}$$

ดังนั้น ตัวแบบเชิงเส้นวางนัยทั่วไปมีรายละเอียดที่สามารถสรุปได้ดังตารางที่ 4

<u>ตารางที่ 4</u> สรุปรูปแบบทั่วไปของตัวแบบเชิงเส้นวางนัยทั่วไป

	Claims Frequency	Claims Count	Average Claims Size	Probability
Link Function $g(x)$	$\ln x$	$\ln x$	$\ln x$	$\ln \frac{x}{(1-x)}$
Error (ε)	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	$\ln exposure$	0	0

2.1.3.3 Base Level (ระดับพื้นฐาน) และ Intercept Term

โดยทั่วไปแล้ว การกำหนดระดับพื้นฐานสามารถกำหนดได้จากระดับที่มีจำนวนหน่วยเสี่ยงภัยมากที่สุดในแต่ละปัจจัย ซึ่งจากตัวอย่างของการใช้ตัวแบบเชิงเส้นในข้อที่ 2.1.1 เราสามารถกำหนดรูปแบบของพารามิเตอร์ที่มีระดับพื้นฐานได้ดังตัวอย่างต่อไปนี้

ตัวอย่างเช่นกำหนดให้ผู้หญิงที่ขับขี่รถยนต์นอกเมืองเป็นระดับพื้นฐาน เราสามารถกำหนดพารามิเตอร์ได้ดังนี้

β_2 คือ จำนวนครั้งเฉลี่ยของการเรียกร้องค่าสินไหมทดแทนจากการขับขี่รถยนต์นอกเมืองของผู้หญิง

$\beta_1 - \beta_2$ คือ จำนวนครั้งเฉลี่ยของการเรียกร้องค่าสินไหมทดแทนที่เพิ่มขึ้นเนื่องจากผู้ขับขี่เป็นผู้ชาย

β_3 คือ จำนวนครั้งเฉลี่ยของการเรียกร้องค่าสินไหมทดแทนที่เพิ่มขึ้นเนื่องจากการขับขี่รถยนต์ในเมือง

ในกรณีที่เราสร้างตัวแบบโดยใช้ข้อมูลที่มีหลายตัวแปร และแต่ละตัวแปรมีหลายระดับขั้น เราสามารถลดจำนวนพารามิเตอร์ในตัวแบบได้โดยการสร้างพารามิเตอร์ที่ไม่ขึ้นอยู่กับพารามิเตอร์ตัวอื่น และสามารถใช้ได้กับข้อมูลทุกตัว ซึ่งเราเรียกพารามิเตอร์ดังกล่าวว่า Intercept Term

จากตัวอย่างข้างต้น เราสามารถกำหนดพารามิเตอร์ที่เหมาะสมจากปัจจัยในการคำนวณทั้ง 2 ปัจจัย (มีทั้งระดับพื้นฐาน และ intercept term) ได้ดังนี้

เพศของผู้ขับขี่	**พารามิเตอร์**
ชาย	$\beta_1 - \beta_2$
หญิง	

สถานที่ขับขี่	**พารามิเตอร์**
นอกเมือง	
ในเมือง	β_3

Intercept Term	β_2

ในกรณีที่เราสร้างตัวแบบที่มี Intercept Term แต่ไม่ได้มีการกำหนดระดับพื้นฐาน (Base Level) ให้กับตัวแปรแต่ละตัว จะทำให้ตัวแบบที่ได้จากการคำนวณเกิดปัญหามีความสัมพันธ์กันในรูปแบบเชิงเส้นระหว่างตัวแปรต้น (Aliasing) ซึ่งทำให้ตัวแบบที่ได้ไม่มีความเป็นเอกลักษณ์ (Unique)

2.1.3.4 Aliasing

Aliasing เกิดขึ้นเนื่องจากตัวแปรต้นนั้นมีความสัมพันธ์กันในรูปแบบเชิงเส้น ซึ่งสามารถเกิดขึ้นได้ทั้งภายในตัวแปรตัวเดียวกัน (Intrinsic Aliasing) และระหว่างตัวแปร (Extrinsic Aliasing) ดังนี้

(1) Intrinsic Aliasing

Intrinsic Aliasing เกิดขึ้นเมื่อค่าภายในตัวแปรตัวเดียวกันเองนั้นมีความสัมพันธ์กันโดยธรรมชาติ ซึ่งโดยทั่วไปจะเกิดกับตัวแปรที่มีลักษณะข้อมูลเป็นกลุ่ม (Category)

ตัวอย่างเช่น อายุของรถยนต์ถูกแบ่งออกเป็น 3 กลุ่ม ดังนี้

กลุ่มที่ 1: รถยนต์มีอายุระหว่าง 0-5 ปี (ให้ x_1มีค่าเป็น 1)
กลุ่มที่ 2: รถยนต์มีอายุระหว่าง 5-10 ปี (ให้ x_2 มีค่าเป็น 1)
กลุ่มที่ 3: รถยนต์มีอายุมากกว่า 10 ปีขึ้นไป (ให้ x_3 ค่าเป็น 1)

จากข้อมูลดังกล่าว เราสามารถเขียนความสัมพันธ์ให้อยู่ในรูปเชิงเส้นได้ดังสมการที่ (2.13)

$$x_1 + x_2 + x_3 = 1 \quad (2.13)$$

ในกรณีที่รถยนต์มีอายุน้อยกว่า 10 ปี จะทำให้ $x_3 = 0$ และในกรณีที่อายุรถยนต์มากกว่า 10 ปี จะทำให้ $x_3 = 1$ ซึ่งสามารถเขียนเป็นความสัมพันธ์ใหม่ได้ดังสมการที่ (2.14)

$$x_3 = 1 - x_1 - x_2 \quad (2.14)$$

จากสมการที่ (2.14) สามารถเขียนรูปแบบตัวประมาณเชิงเส้นได้ดังสมการต่อไปนี้

$$\eta = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \quad (2.15)$$

$$\eta = \beta_1 X_1 + \beta_2 X_2 + \beta_3 (1 - X_1 - X_2) \quad (2.16)$$

$$\eta = (\beta_1 - \beta_3) X_1 + (\beta_2 - \beta_3) X_2 + \beta_3 \quad (2.17)$$

ดังนั้น เราสามารถเขียนสมการให้อยู่ในรูปความสัมพันธ์เชิงเส้นแบบใหม่ได้ดังที่ (2.18)

$$\eta = \alpha_1 X_1 + \alpha_2 X_2 + \alpha_0 \quad (2.18)$$

โดยที่

$$\alpha_0 = \beta_3, \alpha_1 = \beta_1 - \beta_3 \text{ และ } \alpha_1 = \beta_2 - \beta_3$$

(2) Extrinsic Aliasing

Extrinsic Aliasing เกิดขึ้นในกรณีที่ค่าของตัวแปรใดตัวแปรหนึ่งมีความสัมพันธ์โดยสมบูรณ์กับค่าของตัวแปรอื่น

ตัวอย่างเช่น ในกรณีที่ไม่สามารถระบุความสัมพันธ์ของจำนวนประตูกับสีของรถยนต์ได้ในบางค่าสังเกตดังตารางที่ 5

<u>ตารางที่ 5</u> จำนวนหน่วยเสี่ยงภัยของจำนวนประตูกับสีของรถยนต์

สีรถยนต์	**จำนวนประตู**		
	2	4	ไม่ทราบ
ขาว	100	1,000	0
ดำ	50	500	0
ไม่ทราบ	0	0	20

จะเห็นได้ว่ามีหน่วยเสี่ยงภัยจำนวน 20 หน่วยเสี่ยงภัยที่ไม่ทราบสีของรถยนต์และไม่ทราบจำนวนประตู ดังนั้นสามารถตัดระดับสีของรถยนต์หรือระดับจำนวนประตูที่ไม่ทราบตัวใดตัวหนึ่งออกจากการวิเคราะห์ข้อมูลได้

ในกรณีที่ใช้ Software ในการสร้างตัวแบบเชิงเส้นวางนัยทั่วไป Software จะทำการขจัดปัญหา Alias ให้โดยอัตโนมัติ

2.2. วิธีการทดสอบความมีนัยสำคัญของตัวแบบ

2.2.1 การวินิจฉัยตัวแบบ (Model Diagnostics)

นอกจากที่เราจะสามารถใช้ตัวแบบเชิงเส้นวางนัยทั่วไปในการประมาณค่าพารามิเตอร์ได้แล้ว ตัวแบบเชิงเส้นวางนัยทั่วไปยังให้ตัวชี้วัดถึงความคงเส้นคงวาของตัวประมาณค่าพารามิเตอร์เพื่อประโยชน์ในการประเมินความคลาดเคลื่อนของตัวแบบได้อีกด้วย

เราสามารถทำการทดสอบตัวประมาณค่าพารามิเตอร์แต่ละตัวที่ได้จากตัวแบบ โดยการนำตัวประมาณค่าพารามิเตอร์ที่ได้จากตัวแบบมายกกำลังสองและหารด้วยความแปรปรวนของตัวประมาณค่าพารามิเตอร์ดังกล่าว (ซึ่งค่าที่ได้นี้จะมีการแจกแจงแบบไคสแควร์ (Chi-Square)) และนำค่าที่ได้มาทดสอบโดยใช้การทดสอบแบบไคสแควร์ เพื่อทดสอบว่าค่าของตัวแปรแต่ละค่ามีความแตกต่างกับค่าในระดับพื้นฐานอย่างมีนัยสำคัญหรือไม่

2.2.2 Deviance Test

เราสามารถใช้ Deviance Test ในการประเมินความมีนัยสำคัญสำหรับตัวแปรแต่ละตัว โดยที่ Deviance Test เป็นตัวชี้วัดว่าค่าที่ได้จากตัวแบบ (Fitted Value) มีความแตกต่างจากค่าสังเกตมากน้อยเพียงใด โดยที่เราสามารถกำหนด Deviance Function ได้ดังสมการที่ (2.19)

$$d(Y_i;\mu_i) = 2\omega_i \int_{\mu_i}^{Y_i} \left(\frac{Y_i - \zeta}{V(\zeta)}\right) d\zeta \qquad (2.19)$$

ฟังก์ชันดังกล่าวเป็นตัวชี้วัดที่แสดงถึงความแตกต่างระหว่างค่าที่ได้จากตัวแบบ และค่าสังเกต ซึ่งจะเห็นได้ว่าน้ำหนักของความแตกต่างระหว่างค่าจริง Y_i และค่าคาดหวัง μ_i จะมีค่ามากในกรณีที่ Variance Function มีค่าน้อย

เราสามารถมอง Deviance Function ให้อยู่ในรูปแบบทั่วไปของผลต่างยกกำลังสอง (Squared Error) และเมื่อเราหาผลรวมของ Deviance Function ของทุกค่าสังเกต เราจะได้ตัวชี้วัดค่า Deviance โดยรวม ซึ่งเรียกว่า Total Deviance โดยที่เราสามารถเขียนอยู่ในรูปของสมการได้ดังสมการ (2.20)

$$D = \sum_{i=1}^{n} 2\omega_i \int_{\mu_i}^{Y_i} \left(\frac{Y_i - \zeta}{V(\zeta)}\right) d\zeta \quad (2.20)$$

จากสมการข้างต้น เมื่อเรานำค่า Total Deviance มาหารด้วย Scale Parameter เราจะได้ค่า Scaled Deviance (ซึ่งเป็นรูปแบบมาตรฐานของ Sum Squared Error) ที่ปรับด้วยรูปร่างของการแจกแจงแล้วดังสมการที่ (2.21)

$$D^* = \sum_{i=1}^{n} \frac{2\omega_i}{\phi} \int_{\mu_i}^{Y_i} \left(\frac{Y_i - \zeta}{V(\zeta)}\right) d\zeta \quad (2.21)$$

และจากฟังก์ชันดังกล่าว หากการแจกแจงมีรูปแบบอยู่ภายในวงศ์ชี้กำลัง ฟังก์ชันดังกล่าวจะมีค่ากับ 2 เท่าของผลต่างระหว่างภาวะน่าจะเป็นสูงสุด (Maximum Achievable Likelihood) และภาวะน่าจะเป็นของตัวแบบ (Likelihood)

เราสามารถใช้ Deviance ในการทดสอบได้หลากหลายรูปแบบ โดยหนึ่งในวิธีที่มีประโยชน์มากในการทดสอบสามารถทำได้โดยการหาสัดส่วนของ Likelihood ของ 2 ตัวแบบที่ทับซ้อนกัน (Nested Model) ซึ่งตัวแบบที่ทับซ้อนกันนั้นหมายถึงตัวแบบที่ตัวแปรของตัวแบบที่หนึ่งเป็น Subset ของตัวแบบที่สอง ซึ่งการทดสอบดังกล่าวนั้นมีชื่อเรียกอีกชื่อหนึ่งว่า Type 3 Test

ค่าความเปลี่ยนแปลงของ Scaled Deviance ของ 2 ตัวแบบที่ทับซ้อนกันนั้นจะมีการแจกแจงแบบไคสแควร์ ที่มีองศาอิสระ (Degree of Freedom) เท่ากับจำนวนของค่าสังเกตลบกับจำนวนของพารามิเตอร์

การทดสอบดังกล่าวสามารถทดสอบถึงความมีนัยสำคัญที่เกิดจากการเปลี่ยนแปลงตัวแบบได้โดยการวัดว่าการเพิ่มขึ้นของตัวแปรนั้นส่งผลต่อตัวแบบอย่างมีนัยสำคัญหรือไม่ เมื่อกำหนดให้มีพารามิเตอร์เพิ่มขึ้นในตัวแบบ

การทดสอบไคสแควร์ (Chi-Squared Test) นั้นขึ้นอยู่กับ Scaled Deviance แต่เนื่องจากเราไม่ทราบ Scaled Deviance ในบางการแจกแจงเช่น การแจกแจงแบบพัวซองค์ (Poisson) หรือการแจกแจงแบบทวินาม (Binomial) ดังนั้นเราจึงต้องทำการประมาณค่า Scaled Deviance ก่อนการทดสอบ ซึ่งจะทำให้ความน่าเชื่อถือของการทดสอบลดลง

แต่เนื่องจากเราทราบว่าตัวประมาณค่าหลังจากการปรับองศาอิสระ และค่า True Scaled Parameter มีการแจงแจงแบบไคสแควร์ และผลหารระหว่างการแจกแจงแบบไคสแควร์ กับการแจกแจงแบบ

ไคสแควร์ นั้นมีการแจกแจงแบบเอฟ (F-Distribution) ดังนั้นสัดส่วนจากการเปลี่ยนแปลง Deviance และตัวประมาณของ Scaled Parameter นั้นจึงมีการแจกแจงแบบเอฟ (F-Distribution) เช่นกัน

$$\frac{(D_1 - D_2)}{(df_1 - df_2)D_2/df_2} \sim F_{df_1 - df_2, df_2} \qquad (2.22)$$

ซึ่งหมายความว่าเราสามารถใช้ F-test ได้ในกรณีที่เรานั้นไม่ทราบ Scaled Parameter (ยกตัวอย่างเช่นกรณีที่มีการแจกแจงแบบแกมมา (Gamma))

2.3. ความเหมาะสม และข้อจำกัดของการใช้ตัวแบบเชิงเส้นวางนัยทั่วไป

ตัวแบบเชิงเส้นวางนัยทั่วไปเป็นวิธีทางสถิติที่ใช้ในการหาผลกระทบของปัจจัยต่างๆ ที่ส่งผลต่อค่าสังเกต โดยทั่วไปในทางการประกันภัย ตัวแบบเชิงเส้นวางนัยทั่วไปถูกสร้างขึ้นเพื่อใช้หาผลกระทบที่แท้จริงของปัจจัยต่างๆ ที่ใช้ในการกำหนดเบี้ยประกันภัยต่อประสบการณ์ในการเรียกร้องค่าสินไหมทดแทน และยังถูกใช้เพื่อหาความน่าจะเป็นในการต่ออายุกรมธรรม์อีกด้วย โดยตัวแบบเชิงเส้นวางนัยทั่วไปนั้นได้คำนึงถึงการมีสหสัมพันธ์ระหว่างปัจจัยต่างๆ ในตัวแบบแล้ว

ตัวแบบเชิงเส้นวางนัยทั่วไปนั้นมีความยืดหยุ่นในการสร้างตัวแบบเป็นอย่างมาก โดยผู้สร้างตัวแบบสามารถเลือกฟังก์ชันเชื่อมโยง (Link Function) และการแจกแจงของตัวแปรตามได้ตามความเหมาะสม ซึ่งส่งผลให้ตัวแบบมีความแม่นยำมากยิ่งขึ้น และนอกจากนี้ตัวแบบเชิงเส้นวางนัยทั่วไปยังให้ตัวชี้วัดถึงความแน่นอนของตัวแบบที่ได้ ซึ่งทำให้ตัวแบบเชิงเส้นวางนัยทั่วไปมีความโปร่งใส ง่ายต่อการทำความเข้าใจ และเป็นที่ยอมรับอย่างกว้างขวาง โดยสามารถสรุปได้ดังตารางที่ 6

ตารางที่ 6 สรุปข้อดี ข้อเสียของตัวแบบเชิงเส้นวางนัยทั่วไป

ข้อดี	ข้อเสีย
1. ตัวแปรตามไม่จำเป็นต้องมีการแจกแจงแบบปกติ	1. ตัวแปรตามนั้นต้องเป็นอิสระต่อกัน
2. ฟังก์ชันเชื่อมโยงไม่ขึ้นอยู่กับองค์ประกอบแบบสุ่ม ซึ่งช่วยเพิ่มความยืดหยุ่นในการสร้างตัวแบบ	2. ฟังก์ชันเชื่อมโยงต้องเป็นฟังก์ชันทางเดียว
3. ตัวแปรตามไม่จำเป็นต้องมีความแปรปรวนร่วมกัน (common variance)	3. ต้องสามารถเขียนตัวประมาณให้อยู่ในรูปแบบเชิงเส้นได้

2.4. ตัวอย่างการประยุกต์ใช้ตัวแบบเชิงเส้นวางนัยทั่วไปในงานคณิตศาสตร์ประกันภัย

ในอดีตการสร้างตัวแบบทางคณิตศาสตร์ประกันภัยนั้นจะมุ่งเน้นไปที่การสร้างตัวแบบเพื่อใช้ในการคำนวณเบี้ยประกันภัยเป็นหลัก ซึ่งตัวแบบดังกล่าวจะส่งผลกระทบต่ออัตราการต่ออายุของผู้เอาประกันภัย เนื่องจากการใช้ตัวแบบดังกล่าวจะส่งผลให้มีการเปลี่ยนแปลงเบี้ยประกันภัย

การใช้ตัวแบบอัตราการต่ออายุจะช่วยเพิ่มประสิทธิภาพในการตัดสินใจกำหนดเบี้ยประกันภัย ช่วยในการประมาณการผลกำไร (Profitability Forecast) และยังช่วยเพิ่มประสิทธิภาพในการตัดสินใจทางการตลาดอีกด้วย

ในการสร้างตัวแบบอัตราการต่ออายุนั้น เราจำเป็นต้องใช้ข้อมูลการเสนอเบี้ยประกันภัยต่ออายุและข้อมูลการต่ออายุรายกรมธรรม์ ซึ่งเราสามารถแบ่งกลุ่มของตัวแปรต้นของข้อมูลดังกล่าวได้เป็น 3 กลุ่มดังต่อไปนี้

(1) ข้อมูลส่วนตัวของผู้เอาประกันภัย

นอกจากปัจจัยพื้นฐาน (เช่น อายุ ที่อยู่ ประวัติการเรียกร้องค่าสินไหมทดแทน) เราควรพิจารณาถึงปัจจัยต่างๆ ให้มากที่สุดเท่าที่จะเป็นไปได้ (เช่น จำนวนปีที่ผู้เอาประกันภัยทำประกันภัยกับบริษัทประกันภัย จำนวนกรมธรรม์ที่ผู้เอาประกันภัยทำประกันภัยกับบริษัทประกันภัย วิธีการจ่ายเบี้ยประกันภัยให้บริษัทประกันภัย ช่องทางการจัดจำหน่าย) ซึ่งเราสามารถนำปัจจัยต่างๆ เหล่านี้มาวิเคราะห์เพื่ออธิบายพฤติกรรมของผู้เอาประกันภัยได้

(2) ส่วนลด/ส่วนเพิ่มเบี้ยประกันภัย

ข้อมูลส่วนลด/ส่วนเพิ่มเบี้ยประกันภัยเป็นหนึ่งในปัจจัยที่สำคัญที่สุดสำหรับตัวแบบอัตราการต่ออายุ หากมองตามทฤษฎีแล้วอัตราการต่ออายุนั้นเป็นฟังก์ชันของส่วนลดเบี้ยประกันภัย (หากบริษัทประกันภัยให้ส่วนลดเบี้ยประกันภัยมาก ผู้เอาประกันภัยก็มีแนวโน้มที่จะต่ออายุกับบริษัทประกันภัยมากตามไปด้วย) ซึ่งการใช้ตัวแบบ GLMs ในการสร้างตัวแบบอัตราการต่ออายุจะช่วยปรับความสัมพันธ์ระหว่างความยืดหยุ่นต่อราคากับปัจจัยต่างๆ ได้ดีขึ้น กล่าวคือตัวแบบจะช่วยอธิบายให้เราเข้าใจว่าเราควรกำหนดส่วนลด/ส่วนเพิ่มเบี้ยประกันภัยกับผู้เอาประกันภัยที่มีปัจจัยต่างกันอย่างไร

(3) ข้อมูลของคู่แข่ง

ในทางปฏิบัติข้อมูลของคู่แข่งเป็นข้อมูลที่หายากมากที่สุดเนื่องจากเป็นข้อมูลภายนอกที่เราไม่สามารถควบคุมได้และข้อมูลที่ได้อาจเป็นข้อมูลที่ไม่มีคุณภาพ (เช่นคำบอกกล่าวจากผู้เอาประกันภัย) เราจึงขอ

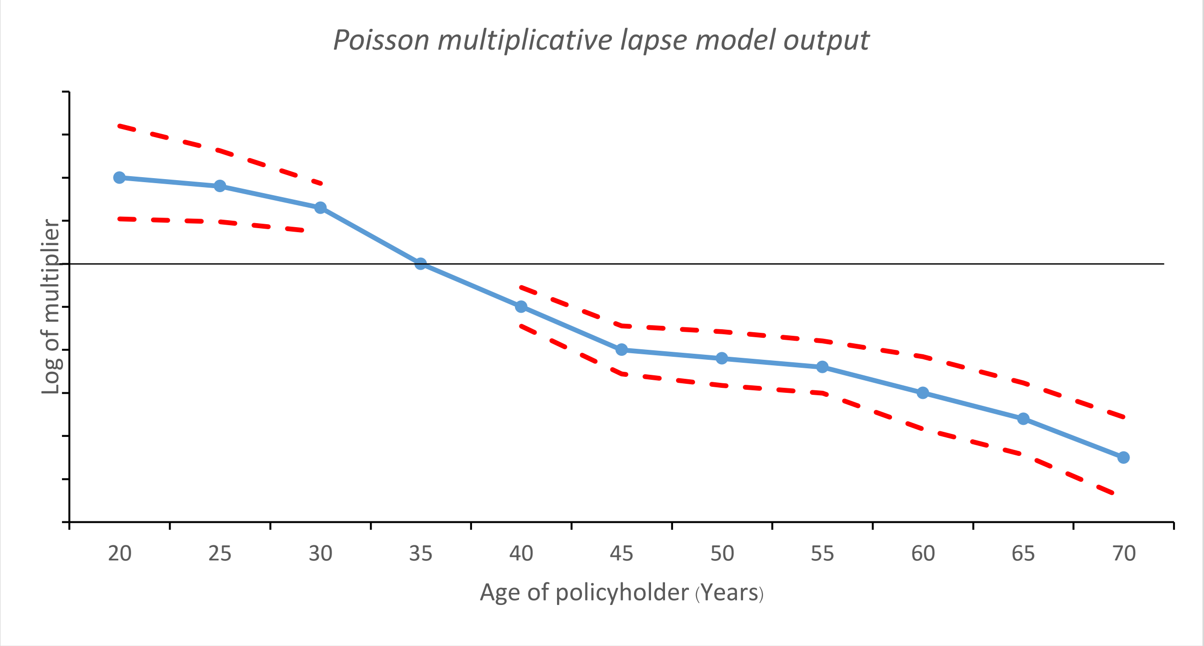
ยกตัวอย่างเพื่อเป็นแนวทางในการจัดหาข้อมูลดังกล่าว เช่น สัดส่วนเบี้ยประกันภัยเมื่อเทียบกับคู่แข่งที่บริษัทประกันภัยระบุว่าเป็นคู่แข่งทางธุรกิจในระดับเดียวกัน (อาจดูได้จากชื่อบริษัทประกันภัย สัดส่วนเบี้ยประกันภัยรับเทียบกับตลาด)

2.4.1 รูปแบบของตัวแบบอัตราการต่ออายุ

เนื่องจากลักษณะการเก็บข้อมูลการต่ออายุมีลักษณะเป็นข้อมูลแบบ 2 ค่า (0 = ไม่ต่ออายุ และ 1 = ต่ออายุ) ซึ่งสอดคล้องกับฟังก์ชันของ error term แบบ binomial และ link ฟังก์ชันแบบ logit (logit link ฟังก์ชันจะปรับค่าที่อยู่ในช่วง 0 ถึง 1 ให้มาอยู่ในช่วง -∞ ถึง ∞)

2.4.2 ตัวอย่างผลการวิเคราะห์

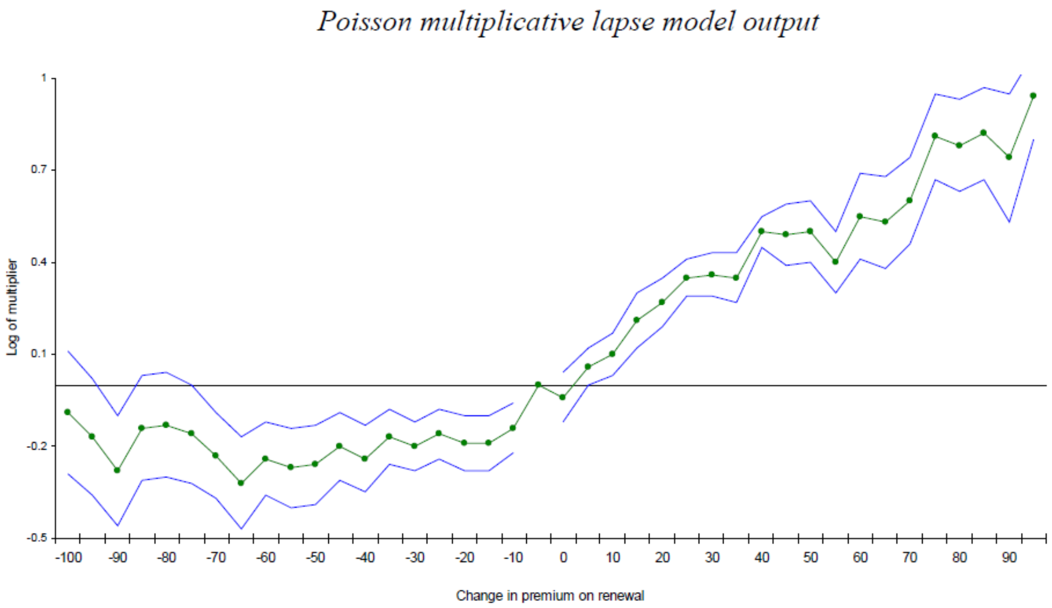
รูปที่ 1 แสดงผลที่ได้จากตัวแบบอัตราการขาดต่ออายุ (1 – อัตราการต่ออายุ) โดยกราฟเส้นทึบ แสดงถึงผลกระทบจากปัจจัยของอายุของผู้เอาประกันภัยต่ออัตราการขาดต่ออายุ และกราฟเส้นประ แสดงถึงช่วงความเชื่อมั่นที่ระดับความเชื่อมั่น 95%



รูปที่ 1 อัตราการขาดต่ออายุ

รูปที่ 2 แสดงถึงผลกระทบของส่วนลด/ส่วนเพิ่มเบี้ยประกันภัยต่ออัตราการขาดอายุกรมธรรม์ โดยรูปดังกล่าวเป็นผลจากการวิเคราะห์โดยการใช้ตัวแบบเชิงเส้นวางนัยทั่วไป จาก UK Institute of Actuaries General Insurance Research Organization (GIRO) โดยใช้ข้อมูลจากบริษัทประกันภัยที่มีส่วนแบ่งตลาด

สูงในประเทศอังกฤษในปี 1996 ซึ่งมีข้อมูลทั้งหมดประมาณ 250,000 กรมธรรม์ โดยมีสมมติฐานว่าส่วนลดของเบี้ยประกันภัยที่ลดลงส่งผลให้อัตราการขาดต่ออายุสูงขึ้น



รูปที่ 2 ผลกระทบของส่วนลด/ส่วนเพิ่มเบี้ยประกันภัยต่ออัตราการขาดต่ออายุ

จะเห็นได้ว่าเมื่อมีการเพิ่มเบี้ยประกันภัย อัตราการขาดต่ออายุกรมธรรม์จะเพิ่มขึ้น ซึ่งเป็นไปตามสมมติฐาน ในขณะที่อัตราการขาดต่ออายุไม่ได้ลดลงอย่างมีนัยสำคัญเมื่อมีการลดเบี้ยประกันภัยในจำนวนเล็กน้อย จากตัวอย่างข้อมูลข้างต้นของประเทศอังกฤษสามารถสรุปได้ว่า บริษัทประกันภัยไม่ควรเพิ่มเบี้ยประกันภัยหากไม่มีเหตุอันสมควร เช่นในกรณีที่ผู้เอาประกันภัยไม่มีการเรียกร้องค่าสินไหมทดแทน

การวิเคราะห์อัตราการต่ออายุสามารถนำมาประยุกต์ใช้ในการกำหนดกลยุทธ์ในการดำเนินงานได้ เช่นการนำผลการวิเคราะห์ที่ได้มาใช้ในการเจาะกลุ่มลูกค้าที่ดีและมีแนวโน้มที่จะต่ออายุค่อนข้างสูง ซึ่งจากกลยุทธ์ดังกล่าวสามารถนำมาพัฒนาเพื่อสร้างเป็นเงื่อนไขในการรับประกันภัยได้

2.5. ตัวอย่าง Software ที่ใช้สร้างตัวแบบเชิงเส้นวางนัยทั่วไป

เนื่องจากเทคโนโลยีสารสนเทศในปัจจุบันมีความก้าวหน้าเป็นอย่างมาก ทำให้ระบบคอมพิวเตอร์ในปัจจุบันมีความเร็วสูง สามารถรองรับการจัดการกับข้อมูลที่มีปริมาณมากได้เป็นอย่างดี และยังมี software ที่มีความสามารถในการสร้างตัวแบบ GLMs ให้เลือกอย่างหลากหลาย เช่น SAS R Emblem Pretium ซึ่ง

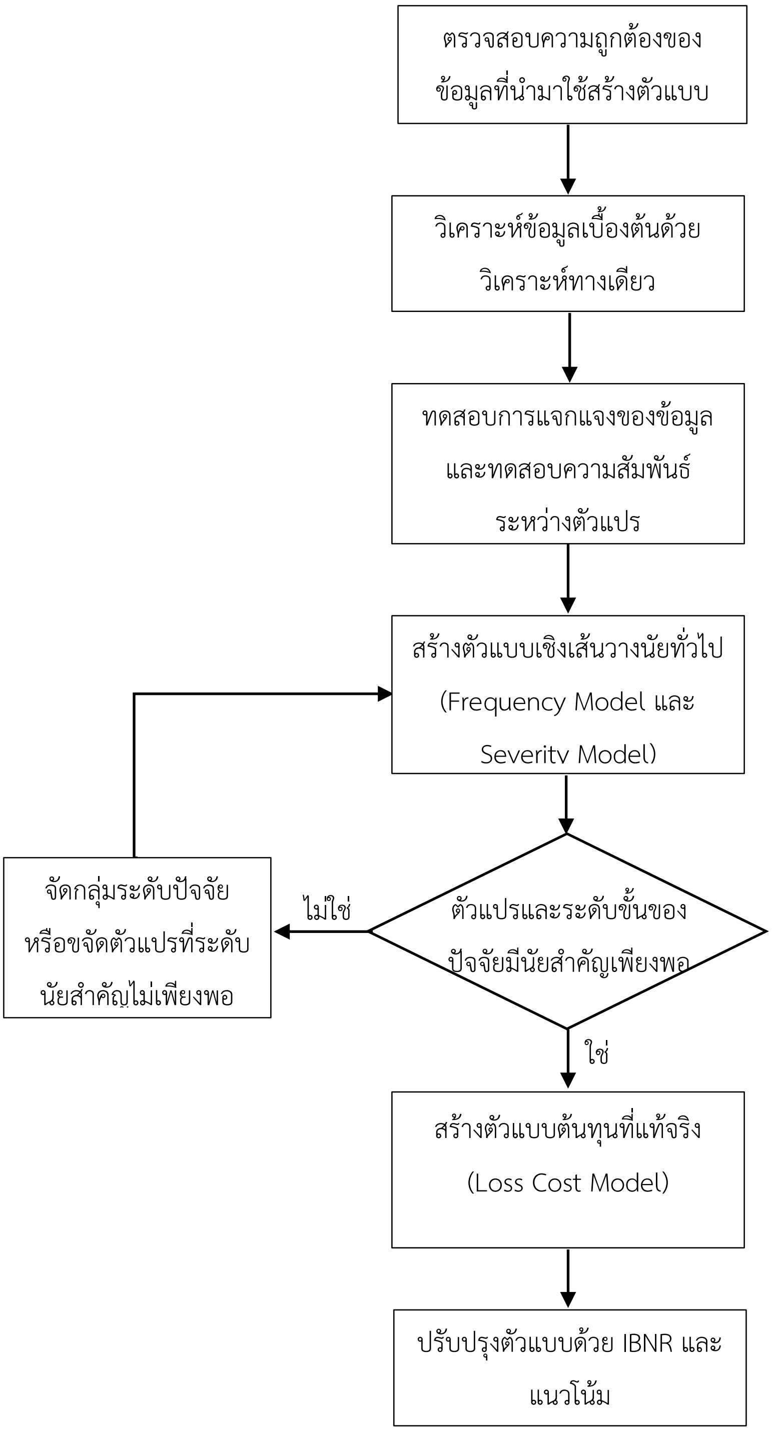
Software แต่ละตัวนั้นก็มีความแตกต่างกันทั้งในด้านความง่ายต่อการใช้งาน (ขึ้นอยู่กับ version ของ Software แต่ละตัว) ชุดคำสั่งที่ใช้ในการสร้างตัวแบบ และราคา

ในคู่มือฉบับนี้จะกล่าวถึง SAS เป็นหลัก โดยจะอธิบายถึงวิธีในการใช้ชุดคำสั่ง GENMOD (ซึ่งเป็นชุดคำสั่งที่ใช้ในการสร้างตัวแบบ GLMs) และจะอธิบายถึงวิธีการอ่านผลที่ได้จากการสร้างตัวแบบผ่านชุดคำสั่ง GENMOD

บทที่ 3 ขั้นตอนการสร้างตัวแบบเชิงเส้นวางนัยทั่วไป

ในบทนี้จะกล่าวถึงรายละเอียดของการสร้างตัวแบบเชิงเส้นวางนัยทั่วไป โดยมีภาพรวมของกระบวนการสร้างตัวแบบเชิงเส้นวางนัยทั่วไปดังนี้

3.1. ภาพรวมของกระบวนการสร้างตัวแบบเชิงเส้นวางนัยทั่วไป



3.2. การเตรียมข้อมูล (Data Preparation)

การเตรียมข้อมูลมีรายละเอียด ดังต่อไปนี้

3.2.1 ลักษณะของข้อมูล

(1) ข้อมูลที่จำเป็นสำหรับการสร้างตัวแบบเชิงเส้นวางนัยทั่วไป

ในการสร้างตัวแบบเชิงเส้นวางนัยทั่วไป จำเป็นต้องมีข้อมูลจำนวนมากในระดับหนึ่ง โดยจำนวนข้อมูลขึ้นอยู่กับความถี่ของการเรียกร้องค่าสินไหมทดแทนและจำนวนตัวแปรที่ผู้สร้างตัวแบบต้องการวิเคราะห์ ทั่วไปแล้วการวิเคราะห์ข้อมูลจะให้ผลที่น่าเชื่อถือเมื่อมีจำนวนข้อมูลประมาณ 100,000 หน่วยเสี่ยงภัยขึ้นไป นอกจากนี้ข้อมูลที่ใช้ในการสร้างตัวแบบควรมีการเก็บรวบรวมตัวอย่างอย่างน้อยสองถึงสามปี เนื่องจากการใช้ข้อมูลย้อนหลังเพียงหนึ่งปีอาจส่งผลให้ผลการวิเคราะห์มีความคลาดเคลื่อนจากปัจจัยภายนอกที่เกิดขึ้นภายในปีนั้นๆ เช่น น้ำท่วมใหญ่ในปี พ.ศ. 2554 และนโยบายรถคันแรกในปี พ.ศ. 2555

นอกจากที่กล่าวมาแล้วข้างต้น พื้นที่และปัจจัยทางภูมิศาสตร์ก็มีความสำคัญต่อการวิเคราะห์ข้อมูลเนื่องจากลักษณะทางกายภาพของแต่ละพื้นที่นั้นมีลักษณะที่แตกต่างกัน ในบางครั้งการวิเคราะห์ข้อมูลโดยพิจารณาถึงปัจจัยทางด้านลักษณะทางกายภาพของแต่ละพื้นที่อาจให้ผลการวิเคราะห์ที่เป็นประโยชน์มากกว่าการวิเคราะห์จากภาพรวม (เช่น ความเสี่ยงของการประกันภัยพืชผลทางการเกษตรนั้นขึ้นอยู่กับพื้นที่ที่เกษตรกรทำการเพาะปลูก พื้นที่เพาะปลูกที่เป็นที่ราบลุ่มเสี่ยงต่ออุทกภัยส่วนพื้นที่เพาะปลูกที่อยู่ที่ที่ราบสูงนั้นเสี่ยงต่อภัยแล้ง) นอกจากนี้ผู้สร้างตัวแบบควรพิจารณาถึงความคุ้มครองและประเภทของการประกันภัยซึ่งควรสร้างตัวแบบจากการประกันภัยที่มีลักษณะความคุ้มครองที่คล้ายคลึงกัน จะทำให้ผลการวิเคราะห์ที่ได้จากตัวแบบนั้นเข้าใจง่ายและชัดเจน

โดยทั่วไปแล้ว โครงสร้างของชุดข้อมูลที่ใช้สำหรับสร้างตัวแบบเชิงเส้นวางนัยทั่วไปจะประกอบด้วยรายละเอียดของกรมธรรม์และข้อมูลค่าสินไหมทดแทน ซึ่งข้อมูลดังกล่าวควรจัดเก็บในระดับรายความเสี่ยงภัย (Individual Risk) และในกรณีที่มีการเปลี่ยนแปลงข้อมูลในระหว่างที่กรมธรรม์ยังมีผลบังคับ ผู้สร้างตัวแบบควรแบ่งข้อมูลออกจากกันโดยให้วันที่มีผลบังคับของการเปลี่ยนแปลงข้อมูลเป็นวันสิ้นสุดของข้อมูลเดิม

ข้อมูลระดับรายความเสี่ยงภัยอาจเป็นทั้งข้อมูลรายความเสี่ยงภัย (เช่น การสร้างตัวแบบความถี่ในการเรียกร้องค่าสินไหมทดแทนโดยใช้ข้อมูลของกรมธรรม์ประกันภัยรถยนต์รายบุคคล) หรืออาจเป็นข้อมูลรายกรมธรรม์ที่ไม่เป็นข้อมูลรายความเสี่ยงภัยที่แท้จริงก็ได้ (เช่น การสร้างตัวแบบอัตราการต่ออายุของ

กรมธรรม์โดยใช้ข้อมูลรายกรมธรรม์ ซึ่งอาจเป็นกรมธรรม์แบบกลุ่ม) โดยข้อมูลดังกล่าวจะประกอบไปด้วยข้อมูลเบื้องต้นดังต่อไปนี้

- หน่วยเสี่ยงภัยที่นับได้ (Earned Exposure)
- จำนวนครั้งในการเรียกร้องค่าสินไหมทดแทนที่เกิดขึ้น (Number of Incurred Claims)
- ค่าสินไหมทดแทนที่เกิดขึ้น (Incurred Claims)
- เบี้ยประกันภัย (Premium) โดยมากจะอยู่ในรูปเบี้ยประกันภัยที่ถือเป็นรายได้ (Earned Premium) และแม้ว่าเราจะไม่ได้ใช้ข้อมูลเบี้ยประกันภัยในการสร้างตัวแบบโดยตรง แต่ผู้สร้างอาจใช้ข้อมูลดังกล่าวในการประเมินถึงผลกระทบจากการปรับเบี้ยประกันภัย และนอกจากนี้ยังสามารถใช้เป็นข้อมูลในการสร้างตารางการวิเคราะห์แบบทางเดียว (One-Way Analysis) หรือการวิเคราะห์สองทางได้ (Two-Way Analysis)
- ปัจจัยที่ใช้ในการกำหนดอัตราเบี้ยประกันภัย (Rating Factor) เช่น ข้อมูลของผู้เอาประกันภัย ลักษณะของทรัพย์สินที่เอาประกันภัย
- ปัจจัยที่ใช้ในการปรับเพื่อไม่ให้ตัวแบบมีความเอนเอียง เช่น กรณีที่ข้อมูลการเรียกร้องค่าสินไหมทดแทนมีการกระจายไปทั่วประเทศ และความเสียหายนั้นแตกต่างไปตามพื้นที่ผู้สร้างตัวแบบควรสร้างตัวแปรดัมมี่ (เช่น อำเภอ หรือจังหวัด) เพื่อใช้ในการปรับข้อมูลการเรียกร้องค่าสินไหมทดแทนเพื่อให้เป็นมาตรฐานเดียวกัน

(2) วิธีการจัดการข้อมูลให้อยู่ในรูปแบบที่สามารถใช้ในการสร้างตัวแบบเชิงเส้นวางนัยทั่วไปได้

โดยทั่วไปแล้ว ข้อมูลที่ใช้ในการสร้างตัวแบบจะถูกจัดเก็บโดยแบ่งเป็นข้อมูลการรับประกันภัย (Underwriting) และข้อมูลสินไหมทดแทน (Claim) ซึ่งเมื่อผู้สร้างตัวแบบนำข้อมูลทั้งสองชุดมารวมเป็นข้อมูลชุดเดียวกัน ผู้สร้างตัวแบบควรพิจารณาถึงข้อสังเกตที่ควรระมัดระวังดังนี้

- ชุดข้อมูลที่เกิดจากการรวมกันของข้อมูลการรับประกันภัยและข้อมูลสินไหมทดแทนต้องระบุได้ถึงความแตกต่างของหน่วยเสี่ยงภัย เช่น เมื่อเลขกรมธรรม์ต่างกัน หน่วยเสี่ยงภัยต้องเป็นคนละหน่วยกัน หรือหน่วยเสี่ยงภัยคนละหน่วยที่มีเลขกรมธรรม์เดียวกัน ก็ควรมีข้อมูลอื่นที่ใช้ระบุความแตกต่างได้ เช่น ลำดับประกันภัยหมู่ (Fleet Sequence) โดยชุดข้อมูลที่เกิดจากการรวมกันของข้อมูลการรับประกันภัยและข้อมูลสินไหมทดแทนนั้น ต้องไม่มีการทับซ้อนกันของหน่วยเสี่ยงภัย
- ในกรณีที่มีการเปลี่ยนแปลงข้อมูลกรรมธรรม์หรือมีการสลักหลังเพิ่มเติม ข้อมูลดังกล่าวจะมีข้อมูลมากกว่าหนึ่งแถวสำหรับหนึ่งเลขกรมธรรม์ ซึ่งข้อมูลเหล่านี้ควรถูกปรับปรุงให้มี

ข้อมูลเพียงหนึ่งแถวต่อหนึ่งลักษณะของหน่วยเสี่ยงภัย (เช่น ผู้เอาประกันภัยแจ้งกับบริษัทประกันภัยว่าได้ทำการเปลี่ยนสีรถยนต์คันที่ทำประกันภัย ซึ่งอาจทำให้ความเสี่ยงภัยเปลี่ยนแปลงไป ดังนั้นผู้สร้างตัวแบบควรแบ่งข้อมูลกรมธรรม์ออกเป็นสองแถวเพื่อสะท้อนถึงความเสี่ยงที่เปลี่ยนแปลงไปด้วย) โดยการถ่วงน้ำหนักหน่วยเสี่ยงภัย ปรับปรุงระยะการคุ้มครอง แบ่งส่วนเบี้ยประกันภัย เพื่อให้ข้อมูลแต่ละแถวไม่มีความทับซ้อนกันของหน่วยเสี่ยงภัย

- ค่าสินไหมทดแทนในกรณีใดกรณีหนึ่ง ต้องมีที่มาจากกรมธรรม์เดียวเท่านั้น (หนึ่งเลขสินไหมทดแทนต้องสามารถเชื่อมโยงได้กับกรมธรรม์เดียวเท่านั้น) แต่ในหนึ่งกรมธรรม์อาจมีค่าสินไหมทดแทนมากกว่าหนึ่งครั้งก็ได้ ดังนั้นผู้สร้างตัวแบบควรทำความเข้าใจถึงลักษณะของตัวแบบที่จะสร้างก่อนการพิจารณารวมข้อมูลสินไหมทดแทน (เช่น หากผู้สร้างตัวแบบต้องการสร้างตัวแบบความถี่ของค่าสินไหมทดแทน ผู้สร้างตัวแบบควรรวมข้อมูล ค่าสินไหมทดแทนทุกครั้งให้อยู่ในแถวเดียวกัน หรือหากต้องการสร้างตัวแบบความรุนแรงของค่าสินไหมทดแทน ผู้สร้างตัวแบบควรแยกค่าสินไหมทดแทนเป็นต่อครั้งแทนการรวมค่าสินไหมทดแทนทุกครั้งให้อยู่ในแถวเดียวกัน)

ผู้สร้างตัวแบบควรตรวจสอบความถูกต้องของชุดข้อมูลก่อนการนำมาใช้สร้างตัวแบบโดยการเปรียบเทียบกับข้อมูลจากแหล่งอื่นที่มีการจัดเก็บไว้ เช่น หากผู้สร้างตัวแบบต้องการนำข้อมูลเบี้ยประกันภัยที่ได้รับจากฝ่ายจัดเก็บเบี้ยประกันภัยมาใช้ในการสร้างตัวแบบ ผู้สร้างตัวแบบควรเปรียบเทียบและยืนยันความถูกต้องของข้อมูลชุดดังกล่าวกับข้อมูลเบี้ยประกันภัยที่ถูกบันทึกลงในบัญชีของฝ่ายบัญชี ซึ่งหากข้อมูลที่นำมาเปรียบเทียบกันมีความแตกต่างกันอย่างมีนัยสำคัญ (เช่น มีความแตกต่างกันเกิน 5%) ผู้สร้างตัวแบบควรทราบและอธิบายถึงที่มาของความแตกต่างนั้นได้อย่างมีเหตุผล เช่น จำนวนเบี้ยประกันภัยของฝ่ายบัญชีอาจถูกปรับปรุงโดยที่ไม่มีการบันทึกลงในระบบ

นอกจากตัวอย่างข้างต้น หากผู้สร้างตัวแบบต้องการใช้ข้อมูลจากหน่วยงานภายนอกในการสร้างตัวแบบ ผู้สร้างตัวแบบควรเลือกข้อมูลจากแหล่งที่เชื่อถือได้และหรือเป็นหน่วยงานที่มีหน้าที่กำกับดูแลข้อมูลดังกล่าวโดยตรง เช่น หากต้องการใช้ข้อมูลผลิตภัณฑ์มวลรวมภายในประเทศ (GDP) รายไตรมาส ผู้สร้างตัวแบบควรอ้างอิงข้อมูลจากธนาคารแห่งประเทศไทย หรือสำนักงานคณะกรรมการพัฒนาเศรษฐกิจและสังคมแห่งชาติ ซึ่งเป็นหน่วยงานภาครัฐที่มีหน้าที่ดูแลรับผิดชอบข้อมูลดังกล่าวโดยตรง

3.2.2 การตรวจสอบความสมเหตุสมผลของข้อมูล

ก่อนที่ผู้สร้างตัวแบบจะนำข้อมูลมาทำการวิเคราะห์ตามขั้นตอนของการคำนวณต้นทุนความเสียหายอ้างอิงนั้น ผู้สร้างตัวแบบควรตรวจสอบความถูกต้องของชุดข้อมูลที่ใช้สำหรับสร้างตัวแบบ และหากพบว่าข้อมูลมีความผิดปกติหรือพบค่าที่อยู่นอกเหนือจากค่าที่เป็นไปได้ (Possible Value) ผู้สร้างตัวแบบควรจัดการกับข้อมูลดังกล่าวก่อนที่จะนำมาเข้าสู่กระบวนการวิเคราะห์ เพื่อให้ผลที่ได้จากการวิเคราะห์มีความถูกต้องและน่าเชื่อถือมากยิ่งขึ้น

ผู้สร้างตัวแบบควรทำการตรวจสอบทุกตัวแปรที่ต้องการนำมาใช้ในการสร้างตัวแบบ โดยอาจแยกเป็นข้อมูลที่ใช้สำหรับการกำหนดอัตราเบี้ยประกันภัย (Rating Factor) และข้อมูลอื่นๆ โดยอาจกำหนดวิธีการตรวจสอบข้อมูลดังตัวอย่างต่อไปนี้

- จำนวนหน่วยเสี่ยงภัย (Earned Exposure Unit) ต้องมีค่ามากกว่าหรือเท่ากับ 0
- จำนวนเงินเอาประกันภัยรวม (Sum Insured) ต้องมีค่ามากกว่า 0
- จำนวนเงินเอาประกันภัยรวมมีค่ามากกว่าเบี้ยประกันภัย
- ค่าสินไหมทดแทนจ่าย (Paid) มีค่ามากกว่าหรือเท่ากับ 0
- วันสิ้นสุดความคุ้มครอง ต้องไม่เกิดขึ้นก่อนวันเริ่มความคุ้มครอง

หากผู้สร้างตัวแบบสามารถระบุถึงความผิดพลาดและสามารถแก้ไขข้อมูลให้เป็นไปตามความเป็นจริงได้ ควรทำการแก้ไขข้อมูลให้ถูกต้องเพื่อความแม่นยำในการพยากรณ์ของตัวแบบ หรือหากไม่สามารถระบุความผิดพลาดได้ ก็อาจพิจารณาลบข้อมูลที่ผิดพลาดออกไป

3.2.3 กรณีศึกษา ปัญหาที่มักพบในขั้นตอนการเตรียมข้อมูล และแนวทางการแก้ไขปัญหา

<u>กรณีศึกษาที่ 1</u> สินไหมทดแทนที่มีการปิดเรื่องแล้วแต่ไม่มีการจ่ายเงิน (Claim closed without payments)

ในกรณีที่ชุดข้อมูลสินไหมทดแทนที่นำมาใช้ในการสร้างตัวแบบมีค่าสินไหมทดแทนที่ไม่มีการจ่ายเงิน (มีค่าเท่ากับ 0 บาท) ผู้สร้างตัวแบบควรพิจารณาตัดข้อมูลค่าสินไหมทดแทนดังกล่าวออกจากชุดข้อมูลก่อนการสร้างตัวแบบ เนื่องจากข้อมูลค่าสินไหมทดแทนดังกล่าวจะส่งผลกระทบต่อค่าสินไหมทดแทนเฉลี่ย (ทำให้ค่าสินไหมทดแทนเฉลี่ยมีค่าน้อยลง) และยังทำให้ไม่สามารถสร้างตัวแบบได้ในบางกรณี (กรณีที่ตัวแบบเชิงเส้นวางนัยทั่วไปมี error ฟังก์ชันอยู่ในรูปแกมมา (Gamma) ซึ่งไม่รองรับค่าสังเกตที่มีค่าเท่ากับ 0)

ตัวอย่าง นาย ก ตรวจสอบพบว่าข้อมูลค่าสินไหมทดแทนของกรมธรรม์ประกันภัยรถยนต์ประเภท 1 บางรายการมีค่าเป็น 0 เนื่องจากผู้เอาประกันภัยมีความประสงค์ที่จะไม่เรียกร้องค่าสินไหมทดแทนและยินยอมให้ปิดเรื่องแล้ว นาย ก จึงทำการตัดข้อมูลดังกล่าวออกจากชุดข้อมูลที่นำไปใช้ในการสร้างตัวแบบ

นอกจากวิธีการตัดข้อมูลแล้ว ยังมีแนวทางอื่นๆ ในการจัดการกับปัญหาดังกล่าว ซึ่งสามารถสรุปได้ดังนี้

- แก้ไขจำนวนครั้งของค่าสินไหมทดแทนที่ไม่มีการจ่ายเงินให้เป็น 0 (ส่งผลให้ไม่มีผลกระทบต่อค่าสินไหมทดแทนเฉลี่ย)
- แยกประเภทของค่าสินไหมทดแทนที่ไม่มีการจ่ายเงินออกจากค่าสินไหมทดแทนที่มีการจ่ายเงิน โดยการสร้างข้อมูลค่าสินไหมทดแทนประเภทใหม่ (เพื่อประโยชน์ในการสร้างตัวแบบสำหรับค่าใช้จ่าย)

กรณีศึกษาที่ 2 Incurred But Not Report (IBNR) กับข้อมูลที่ใช้ในการสร้างตัวแบบเชิงเส้นวางนัยทั่วไปที่ยังไม่สมบูรณ์ (Un Ultimate Data)

โดยทั่วไปแล้ว ข้อมูลค่าสินไหมทดแทนที่นำมาใช้ในการสร้างตัวแบบจะอยู่ในรูปปีอุบัติเหตุ (Accident Year) ซึ่งข้อมูลค่าสินไหมทดแทนที่อยู่ในรูปแบบดังกล่าวนั้นยังสามารถพัฒนาต่อไปได้อีกในอนาคตภายหลังจากที่นำข้อมูลมาใช้ในการสร้างตัวแบบ ดังนั้นผู้สร้างตัวแบบควรรอให้เวลาผ่านไปสักระยะหนึ่งก่อนการนำข้อมูลมาสร้างตัวแบบเพื่อลดความผิดพลาดจากความไม่แม่นยำของการประมาณการณ์ค่าสินไหมทดแทนค้างจ่าย

ตัวอย่าง นาย ก นำข้อมูลค่าสินไหมทดแทนของกรมธรรม์ประกันภัยรถยนต์ประเภท 1 ที่เกิดขึ้นในปีอุบัติเหตุ 2556 มาใช้สร้างตัวแบบโดยดึงข้อมูล ณ สิ้นปี พ.ศ. 2557 ซึ่งจะเห็นได้ว่า นาย ก รอให้ค่าสินไหมทดแทนพัฒนาเข้าสู่จุดสมบูรณ์โดยดึงข้อมูลที่เกิดอุบัติเหตุในปี พ.ศ. 2556 มาใช้เมื่อข้อมูลผ่านไปแล้ว 1 ปี

ระยะเวลาที่เหมาะสมนั้นไม่จำเป็นต้องเป็น 1 ปีเสมอไป ผู้สร้างตัวแบบควรพิจารณาจากลักษณะของความคุ้มครองของกรมธรรม์ (ประกันอัคคีภัย ประกันภัยรถยนต์ มีความเหมาะสมของระยะเวลาในการพัฒนาที่ไม่เหมือนกัน) หรือแบบแผนในการเรียกร้องค่าสินไหมทดแทน (IBNR Pattern)

โดยทั่วไปแล้ว ผู้สร้างตัวแบบจะทำการปรับปรุงข้อมูลเพื่อให้สะท้อนถึงผลกระทบจาก IBNR หลังจากที่ได้ตัวแบบสุดท้าย และนอกจากนี้ผู้สร้างตัวแบบควรพิจารณาถึงผลกระทบที่เกิดขึ้นจากเวลา ซึ่งผลกระทบดังกล่าวส่งผลให้ ค่าสัมประสิทธิ์ (relativity) ของตัวแบบเกิดความเอนเอียงและไม่เหมาะสมที่จะนำไปใช้ในทันที

ในการจัดการกับปัญหาดังกล่าวนั้นอาจทำได้หลายวิธี เช่น

- ไม่สนใจการพัฒนาการของค่าสินไหมทดแทน โดยใช้สมมติฐานว่าพัฒนาการของค่าสินไหมทดแทนไม่ส่งผลกระทบต่อพารามิเตอร์ของตัวแบบ
- เพิ่มตัวแปร dummy (accident year) ลงในตัวแบบเพื่อขจัดผลกระทบที่เกิดจากเวลา โดยที่ไม่นำตัวแปรดังกล่าวมาใช้ในตัวแบบและทำการปรับปรุงตัวแบบโดยการปรับระดับพื้นฐาน (Base Level) แทนการใช้ตัวแปรดังกล่าว
- สร้างตัวแบบด้วยชุดข้อมูลที่มีการพัฒนาของค่าสินไหมทดแทนที่หลากหลายและนำผลลัพธ์ที่ได้มาเปรียบเทียบกันเพื่อหาสัดส่วนของสัมประสิทธิ์เพื่อนำไปใช้ปรับปรุงตัวแบบ (เช่น การนำสัมประสิทธิ์ของตัวแบบที่ได้จากข้อมูลค่าสินไหมทดแทนปีอุบัติเหตุ 2556 ณ สิ้นปี พ.ศ. 2556 มาเปรียบเทียบกับสัมประสิทธิ์ของตัวแบบที่ได้จากข้อมูลค่าสินไหมทดแทนปีอุบัติเหตุ 2556 ณ สิ้นปี 2557 ซึ่งจะเห็นได้ว่าข้อมูลมีพัฒนาของค่าสินไหมทดแทนต่างกัน 1 ปี) และนำสัดส่วนดังกล่าวไปคูณกับสัมประสิทธิ์ที่ได้จากการสร้างตัวแบบที่ใช้ชุดข้อมูลล่าสุด (เช่น นำสัดส่วนดังกล่าวไปปรับปรุงสัมประสิทธิ์ ของตัวแบบที่ได้จากชุดข้อมูลปีอุบัติเหตุ 2557 ณ สิ้นปี พ.ศ. 2557) แต่กรณีที่ใช้วิธีดังกล่าวผู้สร้างตัวแบบจะต้องมั่นใจว่าความเสี่ยงภัยของบริษัทไม่มีการเปลี่ยนแปลงอย่างมีนัยสำคัญ
- ปรับปรุงตัวแบบด้วยการพัฒนาการของค่าสินไหมทดแทนจากข้อมูลในอดีต โดยผู้สร้างตัวแบบสามารถศึกษาวิธีดังกล่าวในบทที่ 4

3.3. การวิเคราะห์ข้อมูลเบื้องต้น

ผู้สร้างตัวแบบควรวิเคราะห์ข้อมูลเบื้องต้น ซึ่งมีรายละเอียด ดังต่อไปนี้

3.3.1 การวิเคราะห์การแจกแจงของข้อมูล

ในทางทฤษฎี ผู้สร้างตัวแบบควรวิเคราะห์ถึงลักษณะการแจกแจงของข้อมูลก่อนการสร้างตัวแบบ เนื่องจากตัวแบบเชิงเส้นวางนัยทั่วไปมีสมมติฐานที่เกี่ยวข้องกับตัวแปรตาม (สมมติฐานที่ (1)) ซึ่งกล่าวเอาไว้ว่าค่าของตัวแปรตามมีความเป็นอิสระต่อกันและมีการแจกแจงแบบใดแบบหนึ่งในวงศ์ชี้กำลัง (Exponential Family) ซึ่งการทราบถึงลักษณะการแจงที่ถูกต้องของข้อมูล ส่งผลให้ตัวแบบที่ได้มีความแม่นยำในการพยากรณ์มากยิ่งขึ้นด้วย

ในกรณีที่ผู้สร้างตัวแบบไม่สามารถทำการวิเคราะห์ถึงลักษณะการแจกแจงของข้อมูลก่อนการสร้างตัวแบบ ผู้สร้างตัวแบบสามารถพิจารณากำหนดรูปแบบการแจกแจงของตัวแปรตามได้โดยใช้รูปแบบการแจกแจงที่เหมาะสมกับประเภทของตัวแบบได้ตามตารางที่ 7

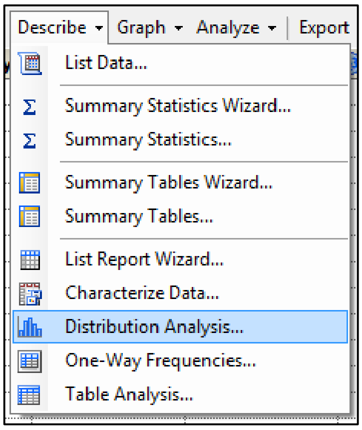
<u>ตารางที่ 7</u> การแจกแจงที่เหมาะสมกับตัวแบบแต่ละประเภท

ตัวแบบ	**การแจกแจงที่เหมาะสม**
จำนวนครั้งของค่าสินไหมทดแทน (Claims Count)	Poisson
ความถี่ของค่าสินไหมทดแทน (Claims Frequency)	Poisson
ความรุนแรงของค่าสินไหมทดแทน (Claims Severity)	Gamma
ต้นทุนของค่าสินไหมทดแทน (Burning Cost)	Tweedie
ความน่าจะเป็นในการต่ออายุ (Retention Rate)	Logit

ในกรณีที่ผู้สร้างตัวแบบต้องการทำการวิเคราะห์ถึงลักษณะการแจกแจงของข้อมูลก่อนการสร้างตัวแบบ ผู้สร้างตัวแบบสามารถทำการวิเคราะห์ถึงลักษณะการแจกแจงของข้อมูลได้โดยด้วยโปรแกรม SAS โดยผ่านการใช้ชุดเครื่องมือ “Distribution Analysis” ตามขั้นตอนดังต่อไปนี้

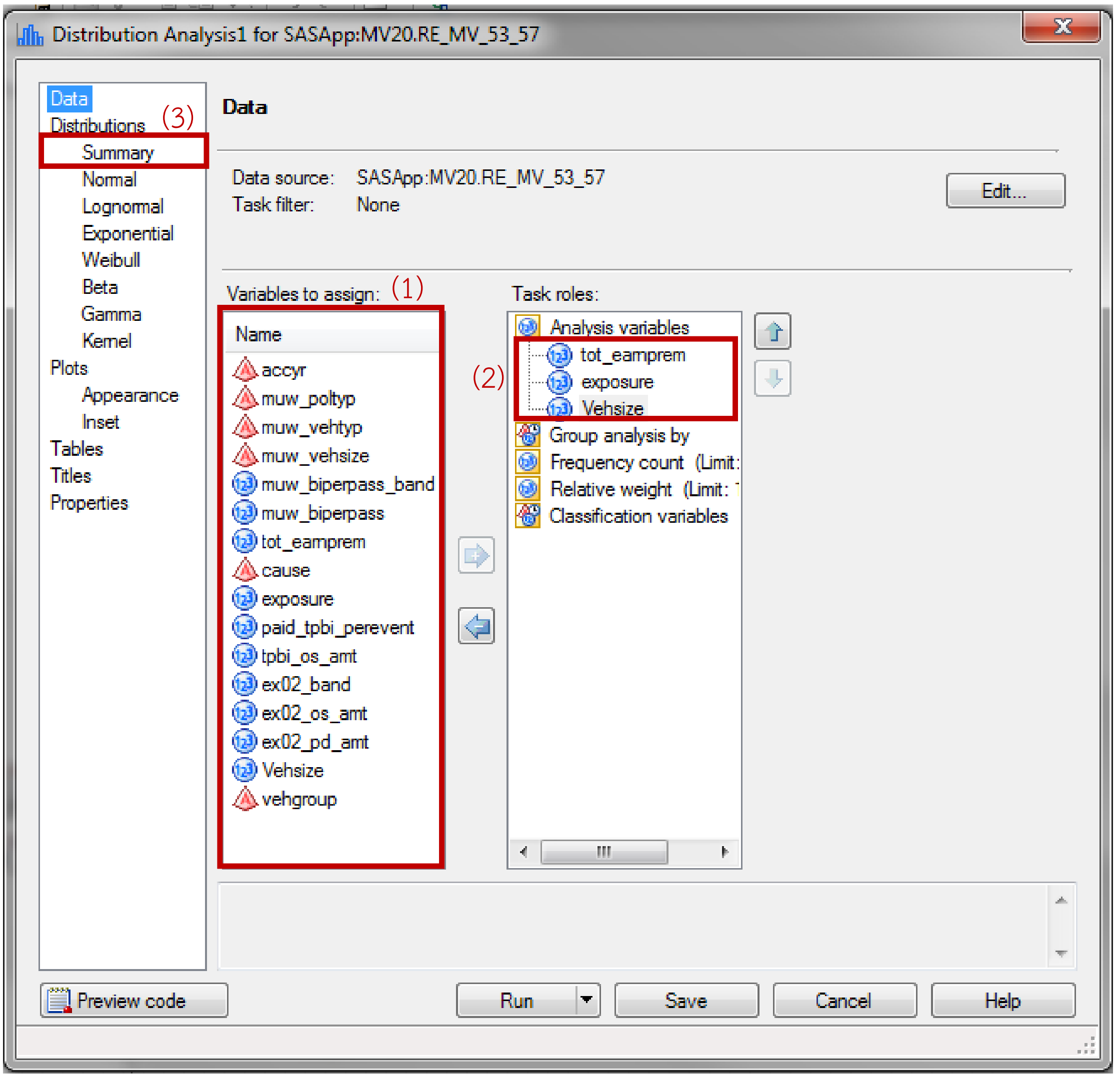
(1) เข้าไปยังชุดข้อมูลที่ต้องการวิเคราะห์ความแจกแจง

(2) เลือก Describe -> Distribution Analysis ดังรูปที่ 3



รูปที่ 3 ตัวเลือกเครื่องมือ Distribution Analysis

(3) จะปรากฏเป็นหน้าต่างตั้งค่ากำหนดการวิเคราะห์ดังรูปที่ 4

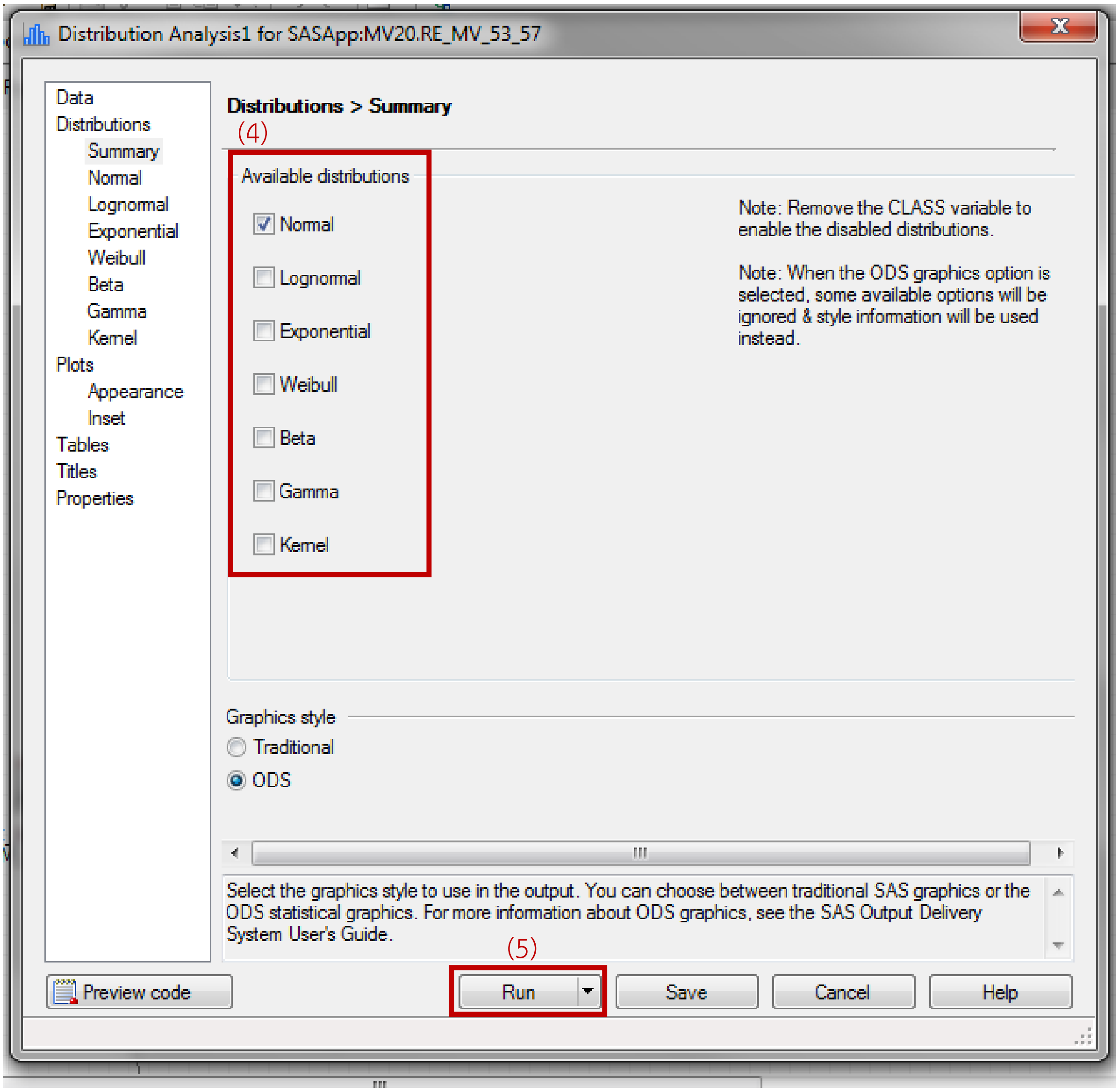


รูปที่ 4 หน้าต่าง Data ของเครื่องมือ Distribution Analysis

(4) เลือกตัวแปรที่ต้องการวิเคราะห์จากช่องหมายเลข (1) ดังรูปที่ 4

(5) ลากตัวแปรที่ต้องการวิเคราะห์จากช่องหมายเลข (1) มาวางไว้ในช่องหมายเลข (2) ดังรูปที่ 4 และตัวแปรดังกล่าวจะปรากฏขึ้นในช่องหมายเลข (2) โดยตัวแปรที่นำมาวิเคราะห์จะต้องเป็นตัวแปรที่เป็นค่าในเชิงตัวเลข (Numeric) เท่านั้น ซึ่งผู้สร้างตัวแบบสามารถวิเคราะห์ได้มากกว่าหนึ่งตัวแปรในแต่ละครั้ง

(6) เลือก Distributions > Summary จากช่องหมายเลข (3) เพื่อเข้าสู่หน้าต่างรูปแบบการแจกแจงที่ต้องการทดสอบ ซึ่งจะปรากฏในหน้าต่างดังรูปที่ 5



รูปที่ 5 หน้าต่าง Distribution > Summary ของเครื่องมือ Distribution Analysis

(7) เลือกการแจกแจงที่ต้องการทดสอบในช่องที่ (4) ดังรูปที่ 5

(8) กดปุ่ม Run ในช่องที่ (5) เพื่อเริ่มต้นการวิเคราะห์ และเมื่อการวิเคราะห์เสร็จสิ้นจะปรากฏผลลัพธ์ในรูปแบบของตารางดังรูปที่ 6 (ตัวอย่างเพียงบางส่วนจากผลการวิเคราะห์)

Page Break

Distribution analysis of: tot_earnprem

(6)

The UNIVARIATE Procedure

Fitted Normal Distribution for tot_earnprem (tot_earnprem)

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	10125.81
Std Dev	Sigma	10542.64

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.17	Pr > D	<0.010
Cramer-von Mises	W-Sq	134819.48	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	1022379.05	Pr > A-Sq	<0.005

(7)

Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	0.0	-14400.04
5.0	0.0	-7215.29
10.0	0.0	-3385.13
25.0	1999.0	3014.91
50.0	8889.0	10125.81
75.0	15778.0	17236.71
90.0	20116.0	23636.74
95.0	24434.0	27466.90
99.0	49100.0	34651.65

รูปที่ 6 ตารางแสดงผลการทดสอบการแจกแจง

จากผลการวิเคราะห์ตามรูปที่ 6 สามารถสรุปผลจากการอ่านค่าสถิติได้ดังนี้

<u>สมมติฐาน</u>

H_0 : ลักษณะการแจกแจงของข้อมูล มีการแจกแจงแบบปกติ (เป็นไปตามสมมติฐาน)

H_1 : ลักษณะการแจกแจงของข้อมูล ไม่ได้มีแจกแจงแบบปกติ (ไม่เป็นไปตามสมมติฐาน)

เนื่องจากค่า p-value ของตัวสถิติ Anderson-Darling ในช่องที่ (7) มีค่าน้อยกว่า 0.005 (ที่ระดับนัยสำคัญที่ 99.5%) จึงปฏิเสธสมมติฐานดังกล่าวข้างต้น ดังนั้นสามารถสรุปได้ว่า ลักษณะการแจกแจงของข้อมูลไม่ได้มีการแจกแจงแบบปกติ

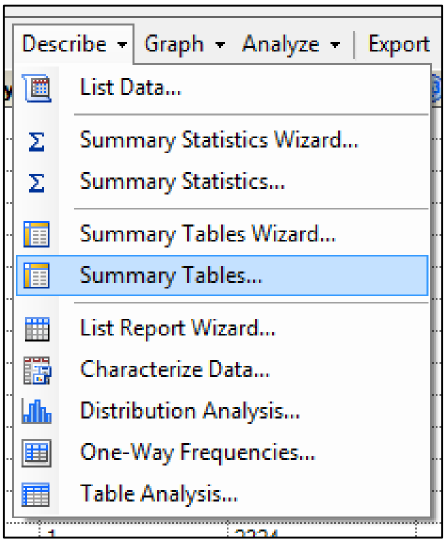
3.3.2 การวิเคราะห์ทางเดียว (One-Way Analysis)

ถึงแม้ว่าตัวแบบเชิงเส้นวางนัยทั่วไปจะเป็นการวิเคราะห์ข้อมูลแบบพหุปัจจัยที่สะท้อนถึงผลกระทบจากปัจจัยต่างๆ โดยคำนึงถึงสหสัมพันธ์ระหว่างตัวแปรต้นด้วยกันแล้ว แต่ผู้สร้างตัวแบบก็ควรใช้การวิเคราะห์

ทางเดียวเพื่อวิเคราะห์ข้อมูลเบื้องต้น และเพื่อพิจารณาปรับปรุงชุดข้อมูลให้มีความถูกต้องและเหมาะสมมากยิ่งขึ้น เช่น การวิเคราะห์ถึงความเหมาะสมในการตัดตัวแปรออกจากการวิเคราะห์ด้วยจำนวนหน่วยเสี่ยงภัย (Earned Exposure Unit) ซึ่งหากหน่วยเสี่ยงภัยเกือบทั้งหมดของตัวแปรดังกล่าวนั้นอยู่ในระดับขั้นเพียงขั้นเดียว (เช่น ค่าของของมูล 99% มีค่าอยู่ในระดับ “N/A”) ผู้สร้างตัวแบบอาจพิจารณาตัดตัวแปรดังกล่าวออกจากการสร้างตัวแบบ เนื่องจากตัวแบบไม่สามารถใช้ตัวแปรดังกล่าวในการแยกระดับความเสี่ยงภัยได้

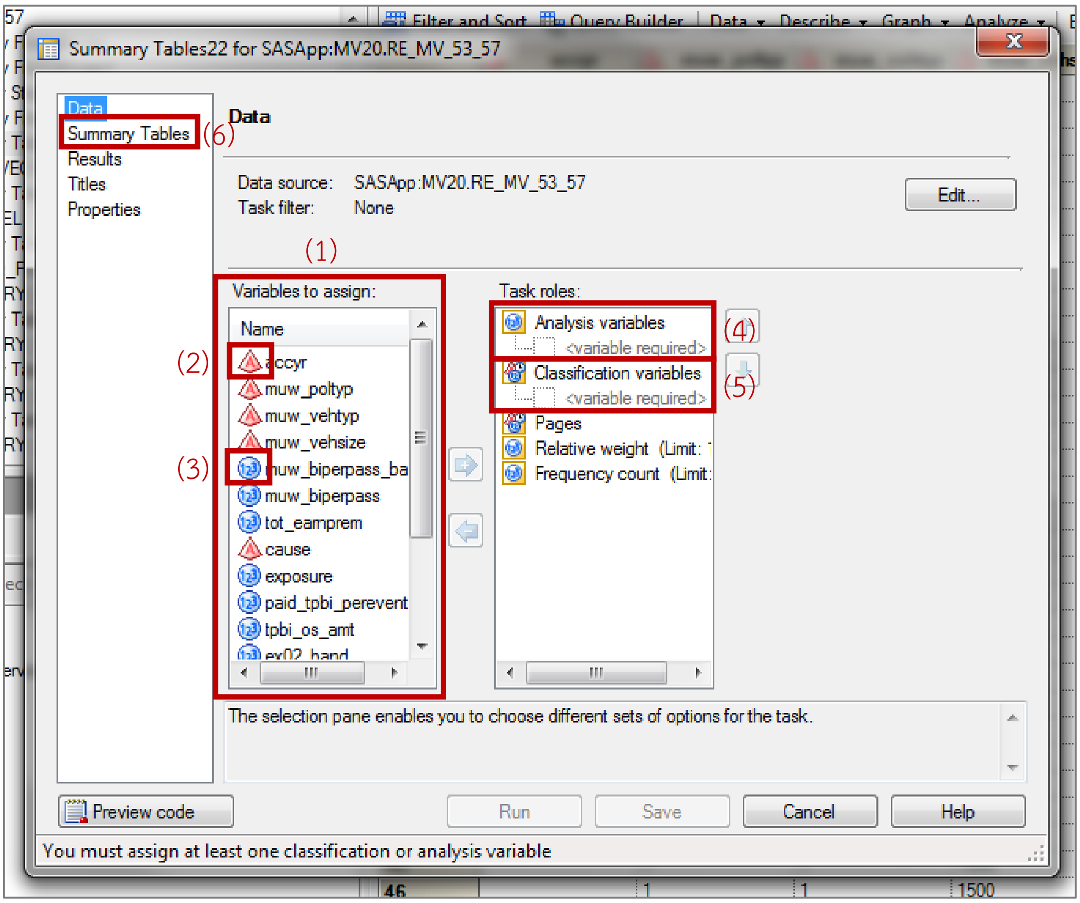
ในการวิเคราะห์ทางเดียวด้วยโปรแกรม SAS ผู้สร้างตัวแบบสามารถทำการวิเคราะห์ผ่านเครื่องมือ Summary Tables เพื่อวิเคราะห์ความสัมพันธ์ทางเดียวระหว่างตัวแปรต้นและตัวแปรตาม โดยผู้สร้างตัวแบบสามารถทำการวิเคราะห์ได้ตามขั้นตอนดังต่อไปนี้

(1) เลือก Describe -> Summary Tables... ดังรูปที่ 7



รูปที่ 7 การเลือกเครื่องมือ Summary Tables เพื่อการวิเคราะห์ทางเดียว

(2) จะปรากฏหน้าต่างสำหรับระบุข้อมูลเพื่อให้กำหนดตัวแปรสำหรับจำแนกข้อมูล ตัวแปรในการวิเคราะห์และตัวเลือกอื่นๆ ดังรูปที่ 8



รูปที่ 8 หน้าต่าง Data ของ Summary Tables

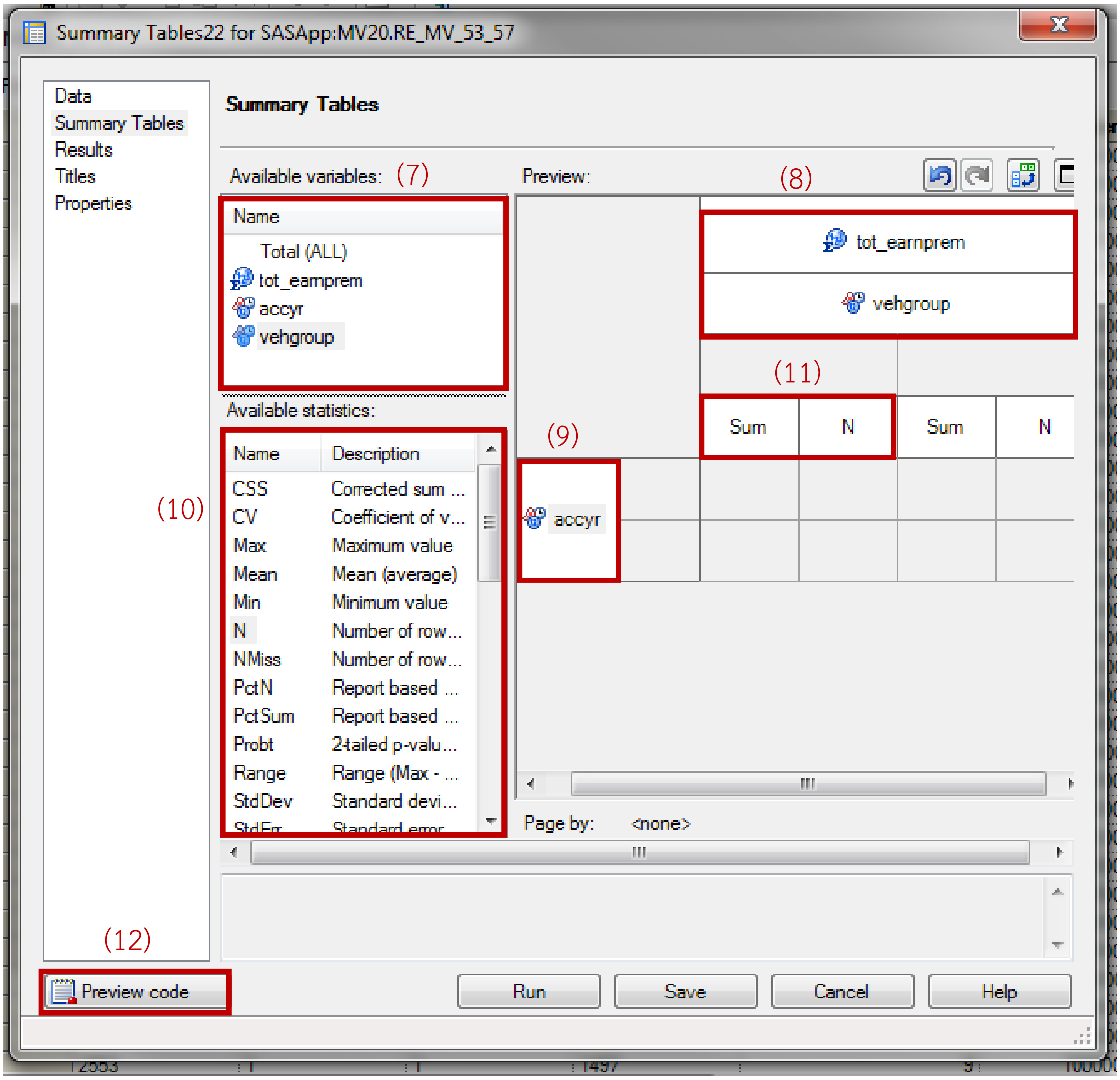
(3) เลือกตัวแปรที่ต้องการวิเคราะห์จากช่องหมายเลข (1) ดังรูปที่ 8 ซึ่งสามารถแบ่งชนิดของตัวแปรได้ตามสัญลักษณ์ดังช่องหมายเลข (2) และ (3) ซึ่งเป็น แบบตัวอักษร (Character) และแบบตัวเลข (Numeric) ตามลำดับ

(4) ลากตัวแปรที่ต้องการวิเคราะห์จากช่องหมายเลข (1) มาวางไว้ในช่องหมายเลข (4) ดังรูปที่ 8 โดยตัวแปรที่สามารถวางในช่องหมายเลข (4) ได้ จะต้องเป็นตัวแปรที่เป็นค่าในเชิงตัวเลข (Numeric) เท่านั้น ซึ่งผู้สร้างตัวแบบสามารถวิเคราะห์ได้มากกว่าหนึ่งตัวแปรในแต่ละครั้ง และตัวแปรในช่องหมายเลข (4) จะถูกนำมาใช้วิเคราะห์ตามเงื่อนไขที่ผู้สร้างตัวแบบต้องการ เช่น จำนวนครั้งของค่าสินไหมทดแทน (Claims Count) และจำนวนค่าสินไหมทดแทนที่เกิดขึ้น (Claims Incurred)

(5) ลากตัวแปรที่ต้องการวิเคราะห์จากช่องหมายเลข (1) มาวางไว้ในช่องหมายเลข (5) ดังรูปที่ 8 โดยตัวแปรที่สามารถวางในช่องหมายเลข (5) ได้นั้นสามารถเป็นตัวแปรชนิดใดก็ได้ ซึ่งผู้สร้างตัวแบบสามารถวิเคราะห์ได้มากกว่าหนึ่งตัวแปรในแต่ละครั้ง และตัวแปรในช่องหมายเลข (5) จะถูกนำมาแบ่งตามจำนวนระดับของตัวแปรดังกล่าวเพื่อแสดงให้เห็นถึงความแตกต่างตาม

เงื่อนไขที่ผู้สร้างตัวแปรกำหนดไว้ในข้อที่ (4) ของแต่ละระดับขั้นนั้นๆ เช่น จำนวนครั้งของค่าสินไหมทดแทน (Claims Count) ของผู้ชายและผู้หญิงจะถูกแบ่งออกเป็น 2 ตาราง เป็นต้น

(6) ให้เลือกไปที่หน้าต่าง Summary Tables ดังช่องหมายเลข (6) ดังรูปที่ 8 จะปรากฏหน้าต่างใหม่ดังรูปที่ 9



รูปที่ 9 หน้าต่าง Summary Tables ของ Summary Tables

(7) ผู้สร้างสามารถใช้ตัวแปรในช่องหมายเลข (7) เพื่อสร้างตารางการวิเคราะห์ทางเดียว หรือตารางการวิเคราะห์สองทางได้โดยการนำตัวแปรในช่องหมายเลข (7) ไปวางไว้ในช่องหมายเลข (8) และช่องหมายเลข (9) โดยที่ตัวแปรที่ถูกวางในช่องหมายเลข (8) จะแสดงในรูปแบบคอลัมน์ และตัวแปรที่ถูกวางในช่องหมายเลข (9) จะแสดงในรูปแบบแถว

(8) ผู้สร้างสามารถเลือกแสดงค่าทางสถิติที่ต้องการโดยการลากตัวเลือกในช่องหมายเลข (10) มาวางในช่องหมายเลข (11) ซึ่งสามารถแสดงผลได้ทั้งในรูปแบบแถวและคอลัมน์ แต่ผู้สร้างจะต้องเลือกให้แสดงค่าสถิติในแนวใดแนวหนึ่งเท่านั้น (ไม่สามารถแสดงผลในทั้งสองแนวพร้อมกันได้)

(9) ผู้สร้างสามารถดูชุดคำสั่งที่ใช้ในการสั่งเพื่อให้แสดงผลตามตารางที่ต้องการได้โดยการกด ปุ่ม Preview Code ในช่องหมายเลข (12)

(10) กดปุ่ม Run หลังจากเลือกตัวแปรและกำหนดรูปแบบตารางตามที่ต้องการ เพื่อทำการวิเคราะห์ข้อมูล

ผลจากการวิเคราะห์นั้นสามารถแสดงได้ถึงความสัมพันธ์ระหว่างตัวแปรที่ผู้สร้างตัวแบบนำมาวิเคราะห์กับค่าสถิติที่ต้องการวิเคราะห์ ซึ่งตัวอย่างของผลลัพธ์ที่ได้จากการวิเคราะห์นั้นเป็นไปดังรูปที่ 10

Summary Tables

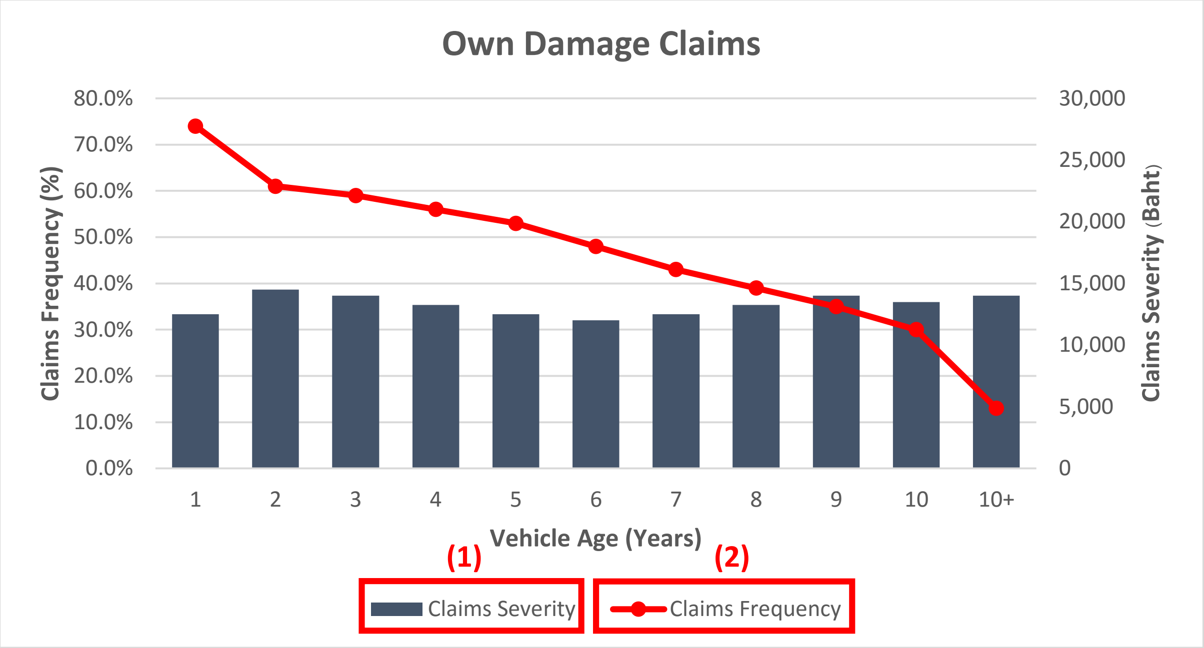
	tot_earnprem					
	accyr					
		2553	2554	2555	2556	2557
	Sum	Sum	Sum	Sum	Sum	Sum
vehgroup						
1Bike	745436480.50	6852114.28	14867848.37	16810615.62	25839716.14	28125719.26
2Car	119638531325	5870146104.1	14818730416	17123206049	21134279053	17983469605
3P-Car	35541452006	1791603593.1	4484363752.5	4569560708.1	5042833785.4	3982209296.7
4T1	23125918188	934175872.82	2195769397.8	2448925892.1	2676360410.4	2229280922.3
5T2	5631292097.9	208157835.13	468845294.94	593134705.97	693805819.68	514855905.33
6T3	6233298474.8	253135001.55	583835046.51	607195324.15	656070829.85	537267705.09
7T4	10923597043	405422902.64	978961143.19	1172725887.6	1478044780.0	1410763441.9
8Header	5742705626.7	255554605.67	652034845.54	847263546.59	1066553962.5	904514856.31
9Trailer	3172430788.5	55536729.73	138117211.82	160584455.86	199250657.89	178795093.51
Other	2511388383.5	152669318.72	314254062.69	308522839.90	338190154.90	239247964.25

Generated by the SAS System ('SASApp', X64_SRV08) on December 16, 2015 at 3:20:15 PM

Page Break

รูปที่ 10 ตัวอย่างหน้าจอแสดงผล Summary Tables

เพื่อความสะดวกในการวิเคราะห์และแปลผล คู่มือนี้ได้นำข้อมูลผลการวิเคราะห์ทางเดียวที่ได้จากโปรแกรม SAS มาสร้างเป็นแผนภาพโดยใช้โปรแกรม Microsoft Excel ดังตัวอย่างตามรูปที่ 11



รูปที่ 11 แผนภาพความสัมพันธ์ของอายุรถยนต์ (ปี) ในส่วนความเสียหายตัวรถยนต์ (Own Damage)

(1) แท่งกราฟในช่องหมายเลข (1) แสดงถึงความรุนแรงของค่าสินไหมทดแทนของตัวรถยนต์คันที่เอาประกันภัยตามอายุรถยนต์ (ปี) ซึ่งจากกราฟจะเห็นได้ว่า ความรุนแรงของค่าสินไหมทดแทนของตัวรถยนต์ (Own Damage) คันที่เอาประกันภัยนั้นไม่มีแนวโน้ม ซึ่งแสดงให้เห็นว่าความรุนแรงของค่าสินไหมทดแทนของตัวรถยนต์คันที่เอาประกันภัยไม่มีความสัมพันธ์กับอายุรถยนต์

(2) เส้นกราฟในช่องหมายเลข (2) แสดงถึงความถี่ของค่าสินไหมทดแทนของตัวรถยนต์คันที่เอาประกันภัยตามอายุรถยนต์ (ปี) ซึ่งจากกราฟจะเห็นได้ว่า ความถี่ของค่าสินไหมทดแทนของตัวรถยนต์คันที่เอาประกันภัยมีแนวโน้มลดลงตามอายุการใช้งานที่เพิ่มขึ้น ซึ่งแสดงให้เห็นว่าความถี่ของค่าสินไหมทดแทนของตัวรถยนต์คันที่เอาประกันภัยมีความสัมพันธ์กับอายุรถยนต์

3.3.3 การวิเคราะห์ความมีสหสัมพันธ์ระหว่างตัวแปร (Correlation Analysis)

เมื่อผู้สร้างตัวแบบสามารถระบุถึงตัวแปรที่ต้องการใช้ในการสร้างตัวแบบได้แล้ว ผู้สร้างตัวแบบสามารถทำการทดสอบถึงความมีสหสัมพันธ์ระหว่างตัวแปรเพื่อช่วยเสริมสร้างความเข้าใจในข้อมูล และเพื่อใช้เป็นข้อมูลเบื้องต้นในการตัดสินใจคัดเลือกตัวแปรที่จะใช้ในการสร้างตัวแบบได้อีกด้วย

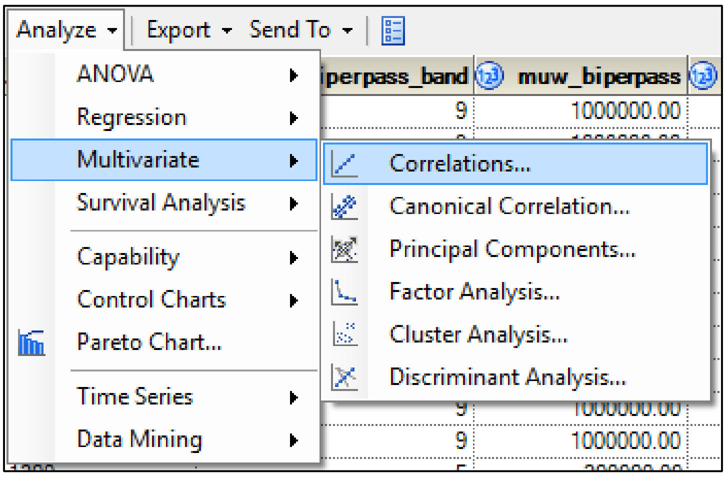
ในกรณีที่ตัวแปรมีสหสัมพันธ์ระหว่างกันสูง (ไม่ว่าเป็นไปในทิศทางเดียวกันหรือเป็นไปในทิศทางตรงกันข้าม) ผู้สร้างตัวแบบสามารถพิจารณาคัดเลือกตัวแปรตัวใดตัวหนึ่งออกจากตัวแบบได้ เพื่อลดความซับซ้อนของตัวแบบ (และยังช่วยลดปัญหา Aliasing)

หากผู้สร้างตัวแบบไม่แน่ใจถึงผลกระทบ ผู้สร้างตัวแบบไม่จำเป็นต้องคัดเลือกตัวแปรที่มีสหสัมพันธ์ระหว่างกันสูงออกจากตัวแบบด้วยตัวเอง เนื่องจากตัวแบบเชิงเส้นวางนัยทั่วไปมีความสามารถในการขจัดปัญหาความมีสหสัมพันธ์ระหว่างตัวแปรสูงได้ด้วยตัวเอง

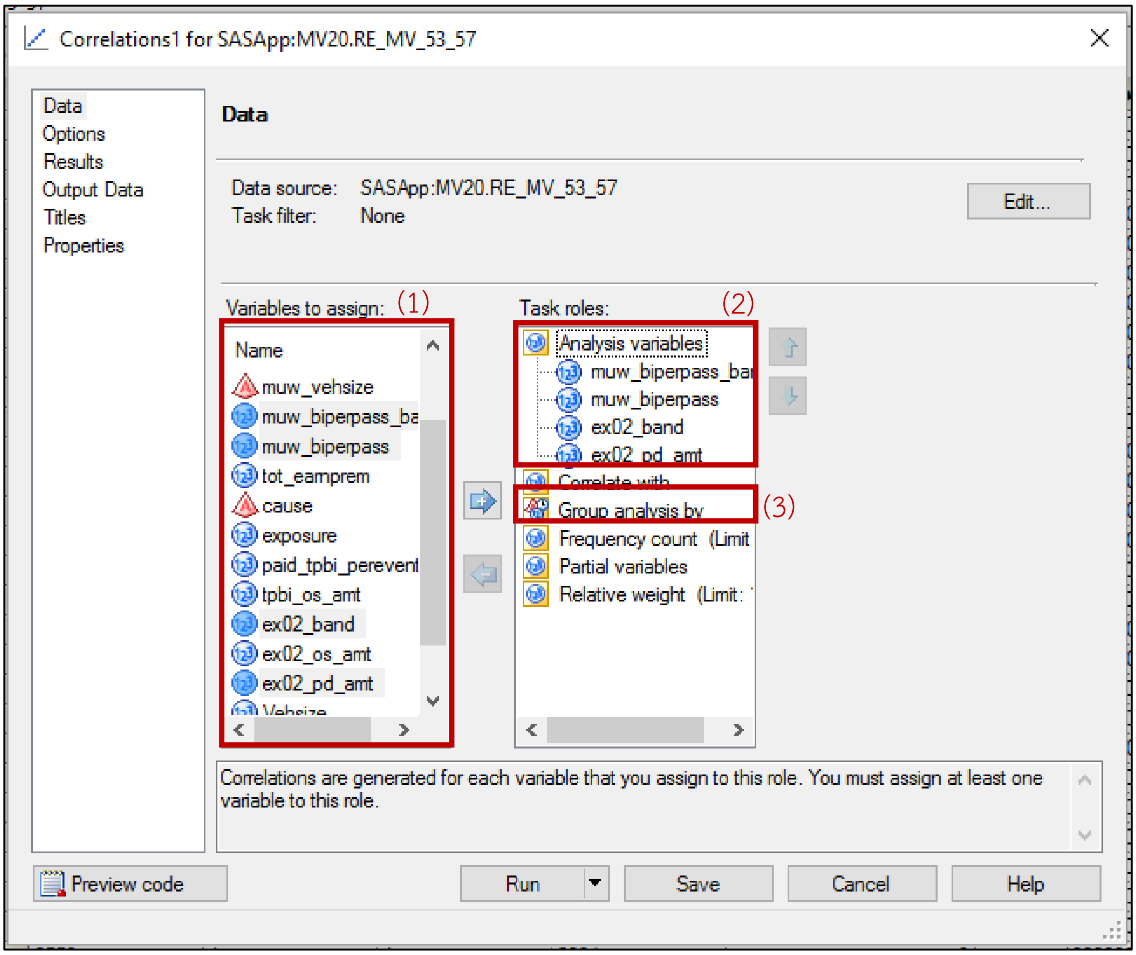
อย่างไรก็ตาม หากผู้สร้างตัวแบบต้องการวิเคราะห์ถึงความสัมพันธ์ระหว่างตัวแปร ผู้สร้างตัวแบบสามารถใช้โปรแกรม SAS ในการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรได้โดยมีขั้นตอนการใช้เครื่องมือสำหรับการวิเคราะห์ดังนี้

(1) เลือกชุดข้อมูลที่ต้องการทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร

(2) เลือกชุดเครื่องมือ Analyze -> Multivariate -> Correlations... ดังรูปที่ 12



รูปที่ 12 การเรียกใช้เครื่องมือ Correlations



รูปที่ 13 หน้าต่างของเครื่องมือ Correlations

(3) เมื่อเปิดเครื่องมือ Correlations จะพบหน้าต่างดังรูปที่ 13

(4) ลากตัวแปรที่ต้องการวิเคราะห์จากช่องหมายเลข (1) มาวางไว้ในช่องหมายเลข (2) ดังรูปที่ 13 โดยตัวแปรที่สามารถวางในช่องหมายเลข (2) ได้ จะต้องเป็นตัวแปรที่เป็นค่าในเชิงตัวเลข (Numeric) เท่านั้น ซึ่งผู้สร้างตัวแบบสามารถวิเคราะห์ได้มากกว่าหนึ่งตัวแปรในแต่ละครั้ง และตัวแปรในช่องหมายเลข (2) จะถูกนำมาใช้หาค่าความสัมพันธ์ระหว่างตัวแปรแต่ละคู่

(5) ลากตัวแปรจากช่องหมายเลข (1) มาวางในช่องหมายเลข (3) ดังรูปที่ 13 เพื่อให้โปรแกรม SAS แบ่งการวิเคราะห์ออกตามระดับขั้นของตัวแปรดังกล่าว ซึ่งจะแสดงให้เห็นถึงความสัมพันธ์ระหว่างตัวแปรแต่ละคู่ตามแต่ละระดับขั้น เช่น ความสัมพันธ์ของจำนวนครั้งของค่าสินไหมทดแทนกับเพศของผู้ขับขี่ ตามปีกรมธรรม์

Correlation Analysis

The CORR Procedure

4 Variables:	muw_biperpass_band	muw_biperpass	ex02_band	ex02_pd_amt

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
muw_biperpass_band	33282953	7.26571	2.20575	241824195	0	13.00000	muw_biperpass_band
muw_biperpass	33282953	1259344	7329701	4.19147E13	0	99999999	muw_biperpass
ex02_band	11852743	1.72424	0.85892	20436931	0	6.00000	ex02_band
ex02_pd_amt	11852743	38.96084	2291	461792832	0	1568126	ex02_pd_amt

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

		muw_biperpass_band	muw_biperpass	ex02_band	ex02_pd_amt
muw_biperpass_band		1.00000	0.25048	0.59240	0.00012
			<.0001	<.0001	0.6914
muw_biperpass_band		33282953	33282953	11852743	11852743
muw_biperpass	(4)	0.25048	1.00000	0.09189	-0.00074
	(5)	<.0001		<.0001	0.0112
muw_biperpass	(6)	33282953	33282953	11852743	11852743
ex02_band		0.59240	0.09189	1.00000	0.00514
		<.0001	<.0001		<.0001
ex02_band		11852743	11852743	11852743	11852743
ex02_pd_amt		0.00012	-0.00074	0.00514	1.00000
		0.6914	0.0112	<.0001	
ex02_pd_amt		11852743	11852743	11852743	11852743

รูปที่ 14 หน้าจอแสดงผลวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร

(6) เมื่อกดปุ่ม Run ดังรูปที่ 13 โปรแกรม SAS จะทำการวิเคราะห์ข้อมูลและแสดงผลจากการวิเคราะห์ดังรูปที่ 14

(7) ช่องหมายเลข (4) แสดงถึงค่าความสัมพันธ์ระหว่างตัวแปรแต่ละคู่ โดยที่ค่าความสัมพันธ์จะมีค่าตั้งแต่ -1 ถึง 1 (ค่าความสัมพันธ์ที่มีค่าเป็นลบหมายความว่าตัวแปรคู่ดังกล่าวมีความสัมพันธ์กันในทิศทางตรงกันข้าม และหากค่าความสัมพันธ์สัมบูรณ์ของตัวแปรคู่ดังกล่าวมีค่าความเข้าใกล้ 1 หมายความว่าตัวแปรคู่ดังกล่าวมีความสัมพันธ์ในทิศทางเดียวกันเกือบสมบูรณ์)

(8) ช่องหมายเลข (5) แสดงถึงค่าสถิติ P-Value ซึ่งแสดงถึงความมีนัยสำคัญของค่าความสัมพันธ์ของตัวแปรแต่ละคู่ในช่องหมายเลข (4) ซึ่งเมื่อค่าสถิติ P-Value มีค่าน้อยกว่าระดับความคลาดเคลื่อนที่สามารถรับได้ เช่น 0.05 หมายความว่าตัวแปรคู่ดังกล่าวมีความสัมพันธ์กันอย่างมีนัยสำคัญ

(9) ช่องหมายเลข (6) แสดงถึงจำนวนค่าสังเกตที่โปรแกรม SAS ใช้ในการวิเคราะห์

ในกรณีที่ผู้สร้างตัวแบบต้องการทดสอบความมีสหสัมพันธ์ของตัวแปรคู่ใดคู่หนึ่งที่ไม่ได้มีค่าในเชิงตัวเลข (non-numerical) ผู้สร้างตัวแบบสามารถใช้โปรแกรม SAS ในการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรได้ โดยใช้ตัวสถิติ Cramer's V ซึ่งสามารถทำการวิเคราะห์ผ่านชุดคำสั่ง PROC FREQ ได้ดังนี้

```
proc freq data =  work.test;   (1)
(2) tables sex*sports / chisq;
(3) weight exposure;
run;
```

รูปที่ 15 ตัวอย่างชุดคำสั่ง PROC FREQ สำหรับการทดสอบสหสัมพันธ์ระหว่างตัวแปรที่ไม่ได้มีค่าในเชิงตัวเลข (non-numerical)

(1) ช่องหมายเลข (1) ในรูปที่ 15 คำสั่งที่ใช้เพื่อระบุชุดข้อมูลที่นำมาใช้ในการทดสอบสหสัมพันธ์
(2) ช่องหมายเลข (2) คำสั่งที่ใช้เพื่อระบุตัวแปรที่นำมาใช้ในการทดสอบสหสัมพันธ์ (sex และ sports)
(3) ช่องหมายเลข (3) คำสั่งที่ใช้เพื่อระบุตัวแปรที่นำมาใช้เป็นน้ำหนัก (exposure)

Statistics for Table of Sex by Sports

Statistic	DF	Value	Prob
Chi-Square	1	26666.6667	<.0001
Likelihood Ratio Chi-Square	1	29110.3166	<.0001
Continuity Adj. Chi-Square	1	26664.0001	<.0001
Mantel-Haenszel Chi-Square	1	26666.2222	<.0001
Phi Coefficient		0.6667	
Contingency Coefficient		0.5547	
(4) Cramer's V		0.6667	

รูปที่ 16 หน้าจอแสดงผลวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรด้วยตัวสถิติ Cramer's V

(4) ตัวสถิติ Cramer's V ดังในรูปที่ 16 ช่องหมายเลข (4) แสดงถึงค่าความสัมพันธ์ระหว่างตัวแปรแต่ละคู่ โดยที่ค่าความสัมพันธ์จะมีค่าตั้งแต่ -1 ถึง 1 (ค่าความสัมพันธ์ที่มีค่าเป็นลบหมายความว่าตัวแปรคู่ดังกล่าวมีความสัมพันธ์กันในทิศทางตรงกันข้าม และหากค่า

ความสัมพันธ์สัมบูรณ์ของตัวแปรคู่ดังกล่าวมีค่าความเข้าใกล้ 1 หมายความว่าตัวแปรคู่ดังกล่าวมีความสัมพันธ์ในทิศทางเดียวกันเกือบสมบูรณ์)

ในการวิเคราะห์ความมีสหสัมพันธ์ระหว่างตัวแปร ผู้สร้างตัวแบบควรพิจารณาถึงค่าสังเกตที่ผิดปกติ (outlier) ซึ่งอาจส่งผลกระทบต่อผลการวิเคราะห์สหสัมพันธ์ เช่น ในกรณีที่มีค่าสังเกตที่สูงหรือต่ำผิดปกติอยู่ในชุดข้อมูล ค่าสหสัมพันธ์ที่คำนวณได้จากชุดข้อมูลดังกล่าวจะมีค่าน้อยกว่าค่าสหสัมพันธ์ที่คำนวณได้จากชุดข้อมูลที่มีการตัดค่าที่สูงหรือต่ำผิดปกติดังกล่าวออกไป ซึ่งหากค่าสหสัมพันธ์ดังกล่าวมีความแตกต่างกันมาก ผู้สร้างตัวแบบควรพิจารณาตัดค่าสังเกตที่ผิดปกตินั้นออกจากชุดข้อมูลที่ใช้ในการสร้างตัวแบบ

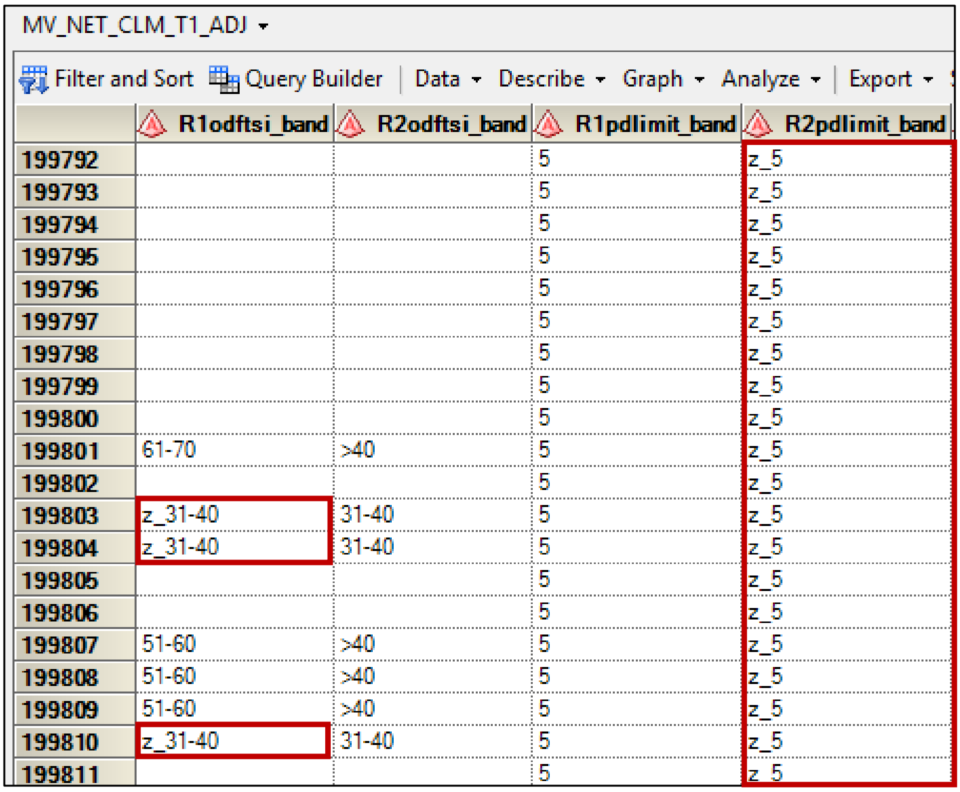
3.4. การสร้างตัวแบบเชิงเส้นวางนัยทั่วไป

การสร้างตัวแบบเชิงเส้นวางนัยทั่วไป มีรายละเอียดดังต่อไปนี้

3.4.1 การเตรียมตัวแปรก่อนการสร้างตัวแบบ

ระดับพื้นฐานของตัวแปร คือระดับขั้นที่ผู้สร้างตัวแบบกำหนดให้เป็นระดับความเสี่ยงภัยพื้นฐานของตัวแปร โดยที่ระดับขั้นอื่นในตัวแปรจะเป็นปัจจัยที่ทำให้ความเสี่ยงภัยเพิ่มขึ้นหรือลดลงไปจากระดับพื้นฐานดังกล่าว โดยทั่วไปแล้วนิยมใช้ระดับขั้นที่มีจำนวนหน่วยเสี่ยงภัยสูงที่สุดในการกำหนดให้เป็นระดับพื้นฐาน

เนื่องจากโปรแกรม SAS จะกำหนดให้ระดับของตัวแปรที่มีลำดับตัวอักษรอยู่ในลำดับสุดท้ายเป็นระดับพื้นฐาน ดังนั้นผู้สร้างตัวแบบสามารถกำหนดระดับพื้นฐานด้วยโปรแกรม SAS โดยการเพิ่มค่า “z_” ไว้ที่หน้าค่าของระดับขั้นที่ต้องการกำหนดให้เป็นระดับพื้นฐานดังรูปที่ 17 (ผู้สร้างตัวแบบใช้ควรระมัดระวังในการกำหนดระดับพื้นฐานในกรณีที่มีระดับขั้นอื่นที่ขึ้นต้นด้วยอักษร z)



	R1odftsi_band	R2odftsi_band	R1pdlimit_band	R2pdlimit_band
199792			5	z_5
199793			5	z_5
199794			5	z_5
199795			5	z_5
199796			5	z_5
199797			5	z_5
199798			5	z_5
199799			5	z_5
199800			5	z_5
199801	61-70	>40	5	z_5
199802			5	z_5
199803	z_31-40	31-40	5	z_5
199804	z_31-40	31-40	5	z_5
199805			5	z_5
199806			5	z_5
199807	51-60	>40	5	z_5
199808	51-60	>40	5	z_5
199809	51-60	>40	5	z_5
199810	z_31-40	31-40	5	z_5
199811			5	z_5

รูปที่ 17 ตัวอย่างการระบุให้ระดับขั้นของตัวแปรใดๆ เป็นระดับพื้นฐาน (Base Level)

3.4.2 ชุดคำสั่งในการสร้างตัวแบบเชิงเส้นวางนัยทั่วไปด้วยโปรแกรม Statistical Analysis System (SAS)

ในการสร้างตัวแบบเชิงเส้นวางนัยทั่วไปด้วยโปรแกรม SAS ผู้สร้างตัวแบบสามารถสร้างตัวแบบได้โดยการใช้ชุดคำสั่ง GENMOD ซึ่งเป็นชุดคำสั่งที่ใช้สำหรับการสร้างตัวแบบเชิงเส้นวางนัยทั่วไปโดยเฉพาะ โดยชุดคำสั่งดังกล่าวจะมีรูปแบบการรับค่าพารามิเตอร์ที่เกี่ยวข้องกับตัวแบบเชิงเส้นวางนัยทั่วไป ซึ่งเป็นไปตามรูปแบบคำสั่งดังรูปที่ 18

```
                  (1)
PROC GENMOD DATA = library.file name ;
(2) CLASS             Var1 Var2 Var3 ;
(3) MODEL    Resp = Var1 Var2 Var3 / DIST     = poisson   (4)
                                     LINK     = log       (5)
                                     OFFSET   = Var4      (6)
                                     TYPE1
                                     TYPE3;  (7)
(8) WEIGHT  Var5 ;
(9) OUTPUT OUT    = deslibrary.outputname
                              (10) RESDEV     = ResName
                              (11) PREDICTED = PredName ;
RUN;
```

รูปที่ 18 ตัวอย่างชุดคำสั่ง GENMOD เพื่อการสร้างตัวแบบเชิงเส้นวางนัยทั่วไป

*** ตัวอย่างชุดคำสั่ง ใช้การระบุค่าที่สื่อถึงรูปแบบคำสั่ง อาจไม่สอดคล้องกับตัวอย่างการแสดงผล ***

(1) คำสั่งที่ใช้เพื่อระบุชุดข้อมูลที่นำมาใช้ในการสร้างตัวแบบ

(2) คำสั่งที่ใช้เพื่อระบุให้ตัวแปรในชุดข้อมูลเป็นตัวแปรต้นที่ใช้ในการสร้างตัวแบบ (Var1 Var2 Var3)

(3) คำสั่งที่ใช้เพื่อระบุลักษณะความสัมพันธ์ของตัวแปรตามและตัวแปรต้น โดยที่ด้านซ้ายมือของสมการคือตัวแปรตาม (Resp) และด้านขวามือของสมการคือตัวแปรต้น (Var1 Var2 Var3)

(4) คำสั่งที่ใช้เพื่อระบุรูปแบบการแจกแจงของตัวแปรตาม ผู้สร้างตัวแบบสามารถระบุการแจกแจงของตัวแปรตามได้ตามความเหมาะสม (สามารถตรวจสอบการแจกแจงที่เหมาะสมกับตัวแบบแต่ละประเภทได้จากตารางที่ 7 ที่ได้กล่าวไว้ก่อนหน้านี้) โดยสามารถเลือกรูปแบบการแจกแจงที่ต้องการได้ดังตารางที่ 8

ตารางที่ 8 คำสั่งสำหรับระบุการแจกแจงในตัวแบบ

DIST =	Distribution	Default Link Function
BINOMIAL | BIN | B	Binomial	Logit
GAMMA | GAM | G	Gamma	Inverse (power (-1))
IGUASSIAN | IG	Inverse Gaussian	Inverse squared (power (-2))
MULTINOMIAL | MULT	Multinomial	Cumulative logit
NEGBIN | NB	Negative Binomial	Log
NORMAL | NOR | N	Normal	Identity
POISSON | POI | P	Poisson	Log

(5) คำสั่งที่ใช้เพื่อระบุฟังก์ชันเชื่อมโยง ผู้สร้างตัวแบบสามารถระบุฟังก์ชันเชื่อมโยงได้ตามความเหมาะสม โดยสามารถเลือกรูปแบบของฟังก์ชันเชื่อมโยงที่ต้องการได้ดังตารางที่ 9 ซึ่งหากผู้สร้างตัวแบบไม่ได้ระบุรูปแบบของฟังก์ชันเชื่อมโยงที่ต้องการ โปรแกรม SAS จะทำการระบุฟังก์ชันเชื่อมโยงให้เป็นไปตามค่าพื้นฐานตามการแจกแจงของตัวแปรตามที่ได้กำหนดไว้โดยอัตโนมัติ (รูปแบบของฟังก์ชันเชื่อมโยงพื้นฐานเป็นไปดังตารางที่ 8)

ตารางที่ 9 คำสั่งสำหรับระบุฟังก์ชันเชื่อมโยงในตัวแบบ

LINK =	Link Function
CUMCLL | CCLL	Cumulative Complementary Log-Log
CUMLOGIT | CLOGIT	Cumulative Logit
CUMPROBIT | CPROBIT	Cumulative Probit
CLOGLOG | CLL	Complementary Log-Log
IDENTITY | ID	Identity
LOG	Log
LOGIT	Logit
PROBIT	Probit
POWER(*number*) | POW(*number*)	Power with λ = (*number*)

(6) คำสั่งที่ใช้เพื่อระบุตัวแปรที่จะนำมาใช้เป็นค่า Offset ให้กับตัวแบบ (Var4) โดยตัวแปรที่นำมาใช้จะต้องไม่ถูกระบุให้เป็นตัวแปรต้นหรือตัวแปรตามของตัวแบบ (ต้องไม่ใช่ตัวแปรที่ถูก

ระบุอยู่ในคำสั่งที่ (3)) ผู้สร้างตัวแบบสามารถระบุ Offset ให้กับตัวแบบได้ตามความเหมาะสม (สามารถตรวจสอบ Offset ที่เหมาะสมกับตัวแบบแต่ละประเภทได้จากตารางที่ 4)

(7) คำสั่งที่ใช้เพื่อระบุให้โปรแกรม SAS สร้างตารางการแสดงผลการทดสอบแบบ Type 1 และการทดสอบแบบ Type 3 ซึ่งผู้สร้างตัวแบบสามารถระบุให้แสดงผลเพียงการทดสอบเดียวก็ได้ (จะกล่าวถึงในหัวข้อที่ 3.4.3.1)

(8) คำสั่งที่ใช้เพื่อระบุตัวแปรที่จะนำมาใช้เป็นค่า Prior Weight ให้กับตัวแบบ (Var5) โดยตัวแปรที่นำมาใช้จะต้องไม่ถูกระบุให้เป็นตัวแปรต้นหรือตัวแปรตามของตัวแบบ (ต้องไม่ใช่ตัวแปรที่ถูกระบุอยู่ในคำสั่งที่ (3)) ผู้สร้างตัวแบบสามารถระบุ Prior Weight ให้กับตัวแบบได้ตามความเหมาะสม (สามารถตรวจสอบ Prior Weight ที่เหมาะสมกับตัวแบบแต่ละประเภทได้จากตารางที่ 4)

(9) คำสั่งที่ใช้เพื่อระบุให้โปรแกรม SAS แสดงผลการวิเคราะห์ในตำแหน่งและชื่อที่ต้องการ

(10)คำสั่งที่ใช้เพื่อระบุให้โปรแกรม SAS แสดงค่า Deviance Residual โดยระบุให้อยู่ในรูปตัวแปรตามชื่อที่ได้ระบุเอาไว้ (ResName)

(11)คำสั่งที่ใช้เพื่อระบุให้โปรแกรม SAS แสดงค่าพยากรณ์ของตัวแปรตาม โดยระบุให้อยู่ในรูปตัวแปรตามชื่อที่ได้ระบุเอาไว้ (PredName)

3.4.3 การวิเคราะห์ผลและการตรวจสอบความมีนัยสำคัญของตัวแบบ (Model Validation)

เมื่อผู้สร้างตัวแบบทำการ Run ชุดคำสั่ง GENMOD ตามเงื่อนไขที่ได้กล่าวไปแล้วข้างต้น โปรแกรม SAS จะทำการวิเคราะห์และแสดงผลในรูปแบบตารางรายงานผลดังรูปที่ 19

The GENMOD Procedure

(1)

Model Information		
Data Set	MOTOR_MV.MV_NET_CLM_T1_ADJ_3	
Distribution	Gamma	
Link Function	Log	
Dependent Variable	paid_ft	paid_ft
Scale Weight Variable	C_FT	C_FT

Number of Observations Read	17741194
Number of Observations Used	12788
Sum of Weights	12788
Missing Values	17728406

Class Level Information		
Class	Levels	Values
SFT_odftsi_band	7	0 1-10 11-2 121- 21-3 >130 z_31
SFT_MD	2	0.40-0.99 z_>=1.00
SFT_vehsize_code	2	L z_S
SFT_drv1age	3	18- >35 z_0
SFT_fleet	2	0.90 z_1.

(2)

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	13E3	41781.6602	3.2701
Scaled Deviance	13E3	16813.6467	1.3159
Pearson Chi-Square	13E3	22768.8450	1.7820
Scaled Pearson X2	13E3	9162.5683	0.7171
Log Likelihood		-163054.9037	
Full Log Likelihood		-163054.9037	
AIC (smaller is better)		326133.8073	
AICC (smaller is better)		326133.8318	
BIC (smaller is better)		326223.2825	

(3)

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	12.3752	0.0262	12.3239	12.4265	223563	<.0001
SFT_odftsi_band	1-10	1	-1.2524	0.1677	-1.5811	-0.9237	55.78	<.0001
SFT_odftsi_band	11-2	1	-0.6410	0.0854	-0.8083	-0.4737	56.39	<.0001
SFT_odftsi_band	121-	1	1.2042	0.1653	0.8802	1.5283	53.05	<.0001
SFT_odftsi_band	21-3	1	-0.3530	0.0544	-0.4597	-0.2464	42.11	<.0001
SFT_odftsi_band	>130	1	1.3992	0.2717	0.8667	1.9317	26.52	<.0001
SFT_odftsi_band	z_31	0	0.0000	0.0000	0.0000	0.0000	.	.
SFT_MD	0.40-0.99	1	-0.1185	0.0291	-0.1754	-0.0616	16.64	<.0001
SFT_MD	z_>=1.00	0	0.0000	0.0000	0.0000	0.0000	.	.
SFT_vehsize_code	L	1	-0.2228	0.0286	-0.2789	-0.1668	60.66	<.0001
SFT_vehsize_code	z_S	0	0.0000	0.0000	0.0000	0.0000	.	.
SFT_drv1age	18-	1	-0.3725	0.0743	-0.5181	-0.2269	25.15	<.0001
SFT_drv1age	>35	1	-0.2990	0.0517	-0.4003	-0.1977	33.47	<.0001
SFT_drv1age	z_0	0	0.0000	0.0000	0.0000	0.0000	.	.
SFT_fleet	0.90	1	0.7763	0.0546	0.6693	0.8834	202.12	<.0001
SFT_fleet	z_1.	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		1	0.4024	0.0041	0.3945	0.4105		

The scale parameter was estimated by maximum likelihood.

(4)

LR Statistics For Type 1 Analysis				
Source	2*LogLikelihood	DF	Chi-Square	Pr > ChiSq
Intercept	-326792.34			
SFT_odftsi_band	-326507.29	5	285.05	<.0001
SFT_MD	-326464.88	1	42.41	<.0001
SFT_vehsize_code	-326419.63	1	45.25	<.0001
SFT_drv1age	-326360.85	2	58.77	<.0001
SFT_fleet	-326109.81	1	251.05	<.0001

(5)

LR Statistics For Type 3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
SFT_odftsi_band	5	247.90	<.0001
SFT_MD	1	16.68	<.0001
SFT_vehsize_code	1	60.71	<.0001
SFT_drv1age	2	50.26	<.0001
SFT_fleet	1	251.05	<.0001

รูปที่ 19 ตารางแสดงผลชุดคำสั่ง GENMOD ของโปรแกรม SAS

(1) ช่องหมายเลข (1) แสดงถึงข้อมูลทั่วไปของชุดข้อมูลที่ใช้ในการสร้างตัวแบบ ซึ่งผู้สร้างตัวแบบสามารถใช้ข้อมูลจากช่องหมายเลข (1) เพื่อตรวจสอบความถูกต้องเบื้องต้นได้ (เช่น จำนวนแถวของชุดข้อมูลที่เข้าสู่โปรแกรม SAS ตรงกับจำนวนแถวของชุดข้อมูลเริ่มต้น) โดยช่องหมายเลข (1) จะประกอบไปด้วยตารางดังต่อไปนี้
   - ตารางด้านบนแสดงถึง ชุดข้อมูลที่โปรแกรม SAS ใช้ในการวิเคราะห์ การแจกแจงของตัวแปรตาม ฟังก์ชันเชื่อมโยง ตัวแปรตาม และตัวแปร Prior Weight ที่ผู้สร้างตัวแบบกำหนด
   - ตารางที่สองแสดงถึง จำนวนแถวของชุดข้อมูลที่เข้าสู่โปรแกรม SAS จำนวนแถวของชุดข้อมูลที่โปรแกรม SAS ใช้ในการวิเคราะห์ และจำนวนแถวของชุดข้อมูลที่ถูกตัดออกจากการวิเคราะห์
   - ตารางด้านล่างแสดงถึง จำนวนระดับขั้นของตัวแปร และค่าของแต่ละระดับขั้นของตัวแปร
(2) ช่องหมายเลข (2) แสดงถึงค่าสถิติที่ได้จากการวิเคราะห์ข้อมูล (ผู้สร้างตัวแบบสามารถแปลผลได้โดยใช้ค่าสถิติจากช่องหมายเลข (3) – (5))
(3) ช่องหมายเลข (3) แสดงถึงความมีนัยสำคัญของระดับขั้นในแต่ละตัวแปร
(4) ช่องหมายเลข (4) แสดงถึงผลลัพธ์จากการทดสอบแบบ Type 1 (จะนำมาอธิบายในหัวข้อที่ 3.4.3.1)
(5) ช่องหมายเลข (5) แสดงถึงผลลัพธ์จากการทดสอบแบบ Type 3 (จะนำมาอธิบายในหัวข้อที่ 3.4.3.1)

3.4.3.1 วิธีการตรวจสอบระดับการมีนัยสำคัญของตัวแปร

ผู้สร้างตัวแบบควรทำการตรวจสอบถึงความมีนัยสำคัญของตัวแปรแต่ละตัว และระดับขั้นของตัวแปรทุกระดับขั้น เพื่อสร้างความมั่นใจว่าตัวแบบที่ได้จากการวิเคราะห์เป็นตัวแบบที่สามารถพยากรณ์ได้อย่างมีประสิทธิภาพ โดยผู้สร้างตัวแบบสามารถตรวจสอบความมีนัยสำคัญได้โดยการแปลผลจากค่า P-Value ของค่าสถิติ Chi-Square (Pr > ChiSq) ซึ่งหากค่า P-Value มีค่าน้อยกว่าระดับความคลาดเคลื่อนที่ยอมรับได้ เช่น 0.01 หมายความว่าตัวแปรหรือระดับขั้นของตัวแปรดังกล่าวมีนัยสำคัญต่อตัวแบบ

จากผลการทดสอบแบบ Type 1 ดังรูปที่ 20 ผู้สร้างตัวแบบสามารถสรุปได้ว่าการเพิ่มตัวแปรเข้าสู่ตัวแบบจะส่งผลให้ตัวแบบมีประสิทธิภาพเพิ่มขึ้นอย่างมีนัยสำคัญเมื่อค่า P-Value มีค่าน้อยกว่าระดับความคลาดเคลื่อนที่ยอมรับได้ (โดยทั่วไปอยู่ที่ 0.05) โดยมีเงื่อนไขดังต่อไปนี้

LR Statistics For Type 1 Analysis					
Source	2*LogLikelihood	DF	Chi-Square	Pr > ChiSq	
Intercept	-326792.34				
SFT_odftsi_band	-326507.29	5	285.05	<.0001	(1)
SFT_MD	-326464.88	1	42.41	<.0001	(2)
SFT_vehsize_code	-326419.63	1	45.25	<.0001	(3)
SFT_drv1age	-326360.85	2	58.77	<.0001	
SFT_fleet	-326109.81	1	251.05	<.0001	(4)

รูปที่ 20 ตัวอย่างตารางการวิเคราะห์ Type 1
(ผลหลังการจัดกลุ่ม)

(1) จากค่า P-Value ในช่องหมายเลข (1) ผู้สร้างตัวแบบสามารถสรุปได้ว่าการเพิ่มตัวแปร SFT_odftsi_band เข้าสู่ตัวแบบส่งผลให้ตัวแบบมีประสิทธิภาพเพิ่มขึ้นที่ระดับนัยสำคัญ 99.99%

(2) จากค่า P-Value ในช่องหมายเลข (2) ผู้สร้างตัวแบบสามารถสรุปได้ว่าการเพิ่มตัวแปร SFT_MD เข้าสู่ตัวแบบ (เมื่อมีตัวแปร SFT_odftsi_band อยู่ในตัวแบบ) ส่งผลให้ตัวแบบมีประสิทธิภาพเพิ่มขึ้นที่ระดับนัยสำคัญ 99.99%

(3) จากค่า P-Value ในช่องหมายเลข (3) ผู้สร้างตัวแบบสามารถสรุปได้ว่าการเพิ่มตัวแปร SFT_vehsize_code เข้าสู่ตัวแบบ (เมื่อมีตัวแปร SFT_odftsi_band และ SFT_MD อยู่ในตัวแบบ) ส่งผลให้ตัวแบบมีประสิทธิภาพเพิ่มขึ้นที่ระดับนัยสำคัญ 99.99%

(4) จากค่า P-Value ในช่องหมายเลข (4) ผู้สร้างตัวแบบสามารถสรุปได้ว่าการเพิ่มตัวแปร SFT_fleet เข้าสู่ตัวแบบ (เมื่อมีตัวแปร SFT_odftsi_band, SFT_MD, SFT_vehsize_code และ SFT_drv1age อยู่ในตัวแบบ) ส่งผลให้ตัวแบบมีประสิทธิภาพเพิ่มขึ้นที่ระดับนัยสำคัญ 99.99%

จากผลการทดสอบแบบ Type 3 ดังรูปที่ 21 ผู้สร้างตัวแบบจะสามารถสรุปได้ว่าตัวแปรมีนัยสำคัญต่อตัวแบบเมื่อค่า P-Value มีค่าน้อยกว่าระดับความคลาดเคลื่อนที่ยอมรับได้ (โดยทั่วไปอยู่ที่ 0.05) โดยมีเงื่อนไขดังต่อไปนี้

LR Statistics For Type 3 Analysis				
Source	DF	Chi-Square	Pr > ChiSq	
SFT_odftsi_band	143	514.96	<.0001	(1)
SFT_MD	56	258.08	<.0001	
SFT_vehage	10	46.38	<.0001	
SFT_vehsize_code	1	42.84	<.0001	
SFT_drv1age	54	124.43	<.0001	
SFT_vehusage	1	3.49	0.0618	(2)
FT_fleet	1	236.41	<.0001	

รูปที่ 21 ตัวอย่างตารางการวิเคราะห์ Type 3 และตัวแปรที่ระดับนัยสำคัญไม่เพียงพอ (ผลก่อนการจัดกลุ่ม)

(1) จากค่า P-Value ในช่องหมายเลข (1) ผู้สร้างตัวแบบสามารถสรุปได้ว่าตัวแปร SFT_odftsi_band มีนัยสำคัญต่อตัวแบบที่ระดับนัยสำคัญ 99.99% เมื่อมีตัวแปร SFT_MD SFT_vehage, SFT_vehsize_code, SFT_drv1age, SFT_vehusage และ SFT_fleet อยู่ในตัวแบบ

(2) จากค่า P-Value ในช่องหมายเลข (2) ที่มีค่ามากกว่าระดับความคลาดเคลื่อนที่ยอมรับได้ เช่น 0.05 นั่นคือ ผู้สร้างตัวแบบสามารถสรุปได้ว่าตัวแปร SFT_vehusage ไม่มีนัยสำคัญต่อตัวแบบที่ระดับนัยสำคัญ 95% เมื่อมีตัวแปร SFT_odftsi_band, SFT_MD SFT_vehage, SFT_vehsize_code, SFT_drv1age และ SFT_fleet อยู่ในตัวแบบ

(3) ดังนั้นสามารถสรุปได้ว่า จากข้อมูลข้างต้นตัวแปร SFT_vehusage มีระดับนัยสำคัญไม่เพียงพอ ควรได้รับการปรับปรุงโดยการจัดกลุ่มหรือขจัดออก

นอกจากการตรวจสอบความมีนัยสำคัญของตัวแปรแล้ว ผู้สร้างตัวแบบควรตรวจสอบถึงความมีนัยสำคัญของระดับขั้นของตัวแปรในทุกระดับขั้นด้วย โดยพิจารณาถึงความมีนัยสำคัญจากค่า P-Value ด้วยหลักการเดียวกับการพิจารณาถึงความมีนัยสำคัญในระดับตัวแปร ซึ่งผู้สร้างตัวแบบสามารถตรวจสอบได้จากตารางแสดงผลภาวะน่าจะเป็นสูงที่สุด ดังรูปที่ 22

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	12.3826	0.0805	12.2248	12.5404	23641.9	<.0001
SFT_odftsi_band	2	1	-4.1009	1.6527	-7.3401	-0.8618	6.16	0.0131
SFT_odftsi_band	4	1	-2.2207	1.5804	-5.3183	0.8769	1.97	0.1600
SFT_odftsi_band	5	1	-2.3590	0.9074	-4.1375	-0.5805	6.76	0.0093
SFT_odftsi_band	6	1	-1.5313	0.3566	-2.2302	-0.8324	18.44	<.0001
SFT_odftsi_band	7	1	-2.4375	0.6025	-3.6183	-1.2567	16.37	<.0001
SFT_odftsi_band	8	1	-0.9506	0.3983	-1.7313	-0.1699	5.69	0.0170
SFT_odftsi_band	9	1	-1.1012	0.3948	-1.8749	-0.3274	7.78	0.0053
SFT_odftsi_band	10	1	-1.2322	0.3536	-1.9253	-0.5392	12.14	0.0005
SFT_odftsi_band	11	1	-1.2473	0.3342	-1.9023	-0.5924	13.93	0.0002
SFT_odftsi_band	12	1	-0.7958	0.2909	-1.3659	-0.2257	7.49	0.0062
SFT_odftsi_band	13	1	-0.7920	0.3444	-1.4670	-0.1170	5.29	0.0215
SFT_odftsi_band	14	1	-0.6564	0.2772	-1.1996	-0.1131	5.61	0.0179
SFT_odftsi_band	15	1	-0.7738	0.4180	-1.5931	0.0455	3.43	0.0641
SFT_odftsi_band	16	1	-0.5446	0.2272	-0.9899	-0.0993	5.75	0.0165
SFT_odftsi_band	17	1	-0.6970	0.3126	-1.3096	-0.0844	4.97	0.0258

รูปที่ 22 ตารางการวิเคราะห์ Maximum Likelihood Parameter
และระดับของตัวแปรที่ระดับนัยสำคัญไม่เพียงพอ (ผลก่อนการจัดกลุ่ม)

ในกรณีที่มีบางระดับขั้นของตัวแปรที่ไม่มีนัยสำคัญต่อตัวแบบดังรูปที่ 22 ผู้สร้างตัวแบบควรจัดกลุ่มระดับขั้นของตัวแปรใหม่เพื่อให้ทุกระดับขั้นของตัวแปรมีนัยสำคัญต่อตัวแบบตามที่ได้กล่าวไปแล้วข้างต้น โดยผู้สร้างตัวแบบสามารถทำการจัดกลุ่มใหม่ได้ตามหลักการดังต่อไปนี้

(1) ผู้สร้างตัวแบบควรเริ่มต้นด้วยการรวมกลุ่มของระดับขั้นของตัวแปรทีละคู่ (โดยดูจากระดับขั้นของตัวแปรที่มีค่าสัมประสิทธิ์ใกล้เคียงกัน) และทำการวิเคราะห์ซ้ำด้วยชุดข้อมูลหลังการจัดกลุ่มใหม่ ซึ่งหากผลจากการวิเคราะห์แสดงให้เห็นว่ายังมีระดับขั้นของตัวแปรที่ไม่มีนัยสำคัญต่อตัวแบบ ผู้สร้างตัวแบบควรทำการรวมกลุ่มใหม่จนกว่าทุกระดับขั้นของตัวแปรจะมีนัยสำคัญต่อตัวแบบ

(2) ผู้สร้างตัวแบบอาจพิจารณารวมระดับขั้นของตัวแปรโดยการนำระดับขั้นของตัวแปรที่ไม่มีนัยสำคัญมารวมกับระดับขั้นพื้นฐาน และทำการวิเคราะห์ซ้ำด้วยชุดข้อมูลหลังการจัดกลุ่มใหม่ ซึ่งหากผลจากการวิเคราะห์แสดงให้เห็นว่ายังมีระดับขั้นของตัวแปรที่ไม่มีนัยสำคัญต่อตัวแบบ ผู้สร้างตัวแบบควรทำการรวมกลุ่มใหม่จนกว่าทุกระดับขั้นของตัวแปรจะมีนัยสำคัญต่อตัวแบบ

(3) ในกรณีที่ผู้สร้างตัวแบบไม่สามารถรวมระดับขั้นของตัวแปรเพื่อให้ทุกระดับขั้นของตัวแปรมีนัยสำคัญกับตัวแบบ ผู้สร้างตัวแบบสามารถพิจารณาตัดตัวแปรดังกล่าวออกจากตัวแบบได้

เพื่อความสมเหตุสมผลและง่ายต่อการทำความเข้าใจของผู้ที่เกี่ยวข้องกับการใช้ตัวแบบ ผู้สร้างตัวแบบควรคำนึงถึงความสมเหตุสมผลในการพิจารณาจัดกลุ่มระดับขั้นของตัวแปรด้วย เช่น ในกรณีที่ค่าของแต่ละระดับขั้นของตัวแปรมีลักษณะเป็นข้อมูลแบบต่อเนื่อง (เช่น จำนวนเงินเอาประกันภัย) ผู้สร้างตัวแบบควรทำการรวมกลุ่มของระดับขั้นของตัวแปรโดยให้ระดับขั้นของตัวแปรที่อยู่ในกลุ่มเดียวกันมีค่าต่อเนื่องกัน (ระดับขั้นที่ 1 จำนวนเงินเอาประกันภัยตั้งแต่ 0 ถึง 1 ล้านบาท, ระดับขั้นที่ 2 จำนวนเงินเอาประกันภัยตั้งแต่ 1 ล้านถึง 2 ล้านบาท เป็นต้น) หรือในกรณีที่ผู้สร้างตัวแบบทราบถึงลักษณะทางกายภาพของพื้นที่ ผู้สร้างตัวแบบอาจพิจารณารวมกลุ่มระดับขั้นของตัวแปรที่มีลักษณะทางกายภาพของพื้นที่ที่ใกล้เคียงกัน เป็นต้น

เพื่อความแม่นยำในการพยากรณ์ของตัวแบบ ผู้สร้างตัวแบบควรทำการแบ่งชุดข้อมูลที่ใช้ในการสร้างตัวแบบเอาไว้บางส่วนเพื่อใช้ในการทดสอบประสิทธิภาพของตัวแบบ โดยทั่วไปแล้วผู้สร้างตัวแบบอาจแบ่งข้อมูลไว้ประมาณ 10% หรือ 20% ของข้อมูลก็ได้

เมื่อได้ตัวแบบที่ต้องการแล้ว ผู้สร้างตัวแบบควรทำการทดสอบความแม่นยำจากการพยากรณ์ โดยนำข้อมูลที่แบ่งเอาไว้ 10% หรือ 20% ดังข้างต้น มาเปรียบเทียบกับค่าที่พยากรณ์ได้จากตัวแบบเพื่อหาค่าความคลาดเคลื่อนที่เกิดขึ้นจากการพยากรณ์ ซึ่งค่าความคลาดเคลื่อนดังกล่าวไม่ควรมีค่าเกินกว่าระดับที่ผู้สร้างตัวแบบสามารถยอมรับได้ (โดยทั่วไปแล้ว ค่าความคลาดเคลื่อนไม่ควรเกิน 5%)

3.4.4 การรวมตัวแบบเพื่อหาต้นทุนที่แท้จริง (Loss Cost Model)

ผู้สร้างตัวแบบสามารถคำนวณหาต้นทุนที่แท้จริง (หรือต้นทุนโดยเฉลี่ยต่อหน่วยเสี่ยงภัย) ได้โดยการนำของความถี่ของค่าสินไหมทดแทน (Frequency) มาคูณกับความรุนแรงเฉลี่ยของค่าสินไหมทดแทน (Average Severity) ตามสมการที่ (3.1)

$$Risk\ Premium = Frequency * Average\ Severity \qquad (3.1)$$

เนื่องจากผลจากการวิเคราะห์ที่ได้จากโปรแกรม SAS ไม่ได้แสดงถึงสัมประสิทธิ์ของตัวแบบโดยตรง ดังนั้นผู้สร้างตัวแบบจะต้องนำผลการวิเคราะห์ที่ได้จากโปรแกรม SAS (ค่า Estimate จากช่องหมายเลข (3) ดังรูปที่ 19) มาแปลงให้เป็นค่าสัมประสิทธิ์ของตัวแบบตามสมการที่ (3.2)

$$\beta = Relativity = e^{(Estimate)} \qquad (3.2)$$

ผู้สร้างตัวแบบสามารถสร้างตัวแบบต้นทุนที่แท้จริง (Loss Cost Model) ได้โดยการนำ relativity ของตัวแปรตัวเดียวกันที่อยู่ในระดับขั้นเดียวกันของตัวแบบความถี่ของค่าสินไหมทดแทนและตัวแบบความ

รุนแรงเฉลี่ยของค่าสินไหมทดแทนมาคูณกัน (หากตัวแบบไม่มีตัวแปรหรือระดับขั้นที่ตรงกับตัวแบบอื่น ให้ผู้สร้างตัวแบบแทนค่า 1 ลงในระดับขั้นหรือตัวแปรดังกล่าว) ดังตารางที่ 10

<u>ตารางที่ 10</u> ตัวอย่างการคำนวณเบี้ยประกันภัยพื้นฐานและค่าสัมประสิทธิ์ของตัวแบบต้นทุนที่แท้จริง

Parameter	Level	Frequency Model Relativity	Severity Model Relativity	Loss Cost Model Relativity
		(1)	(2)	(3) = (1) * (2)
Gender	Male	1.000	1.000	1.000
	Female	1.100	0.900	0.990
Region	Urban	1.000	1.000	1.000
	Rural	0.800	1.050	0.840
Vehicle Age	1	1.000	1.000	1.000
	2	0.900	1.000	0.900
	3	0.850	1.000	0.850
	3+	0.800	1.000	0.800

Intercept (เบี้ยประกันภัยพื้นฐาน)	80%	10,000	8,000

ผู้สร้างตัวแบบสามารถพยากรณ์ต้นทุนที่แท้จริงได้จากตัวแบบโดยการนำเบี้ยประกันภัยพื้นฐานคูณกับค่าสัมประสิทธิ์ของปัจจัยในแต่ละระดับขั้น ดังตัวอย่างดังตารางที่ 11

ตารางที่ 11 ตัวอย่างการพยากรณ์ต้นทุนที่แท้จริงที่ได้จากตัวแบบเชิงเส้นวางนัยทั่วไป

No	Information	Base Premium	Gender	Region	Vehicle Age	Loss Cost
		(1)	(2)	(3)	(4)	(5)=(1)*(2)*(3)*(4)
1	Gender=Male Region=Urban Vehicle Age=1	8,000	1.000	1.000	1.000	8,000
2	Gender=Male Region=Rural Vehicle Age=2	8,000	1.000	0.840	0.900	6,048
3	Gender=Female Region=Rural Vehicle Age=2	8,000	0.990	0.840	0.900	5,988
4	Gender=Female Region=Rural Vehicle Age=3+	8,000	0.990	0.840	0.800	5,322

บทที่ 4 การปรับปรุงตัวแบบ

4.1. การปรับปรุงตัวแบบเชิงเส้นวางนัยทั่วไปด้วยความเสียหายที่เกิดขึ้นแล้วแต่ยังไม่ได้มีการรายงานให้ทราบ (IBNR) และแนวโน้ม

เนื่องจากเบี้ยประกันภัยพื้นฐานที่ได้จากตัวแบบเชิงเส้นวางนัยทั่วไปนั้นคำนวณมาจากข้อมูลในอดีตที่ยังไม่สมบูรณ์ (โดยเฉพาะข้อมูลปีปัจจุบันซึ่งข้อมูลค่าสินไหมทดแทนอาจมีการเปลี่ยนแปลงได้ในอนาคต) ผู้สร้างตัวแบบควรพิจารณาปรับปรุงเบี้ยประกันภัยพื้นฐานโดยคำนึงถึงความสมบูรณ์ของค่าสินไหมทดแทนด้วยวิธีการดังต่อไปนี้

(1) คำนวณค่าสินไหมทดแทนตามจำนวนเดือน, ไตรมาส หรือปี (Incurred Claims) โดยแบ่งตามการพัฒนาการของค่าสินไหมทดแทนที่ผู้สร้างตัวแบบมีข้อมูล

(2) คำนวณการพัฒนาการค่าปัจจัยสมบูรณ์ (Ultimate Factor) ของค่าสินไหมทดแทน ด้วยวิธีทางคณิตศาสตร์ประกันภัย

(3) คำนวณค่าสินไหมทดแทนที่สมบูรณ์ (Ultimate Claims) โดยนำค่าสินไหมทดแทนจากข้อ (1) มาคูณกับการพัฒนาการของค่าสินไหมทดแทนจากข้อ (2)

(4) คำนวณค่าปัจจัยสมบูรณ์ (Ultimate Factor) เพื่อเป็นตัวคูณเพื่อใช้ในการปรับเบี้ยประกันภัยพื้นฐาน (Adjustment Factor) โดยนำค่าสินไหมทดแทนที่สมบูรณ์แล้วจากข้อ (3) รวมทุกปีมาหารด้วยค่าสินไหมทดแทนที่ยังไม่สมบูรณ์จากข้อ (1) รวมทุกปี

(5) คำนวณเบี้ยประกันภัยพื้นฐานจากตัวแบบเชิงเส้นวางนัยทั่วไป

(6) คำนวณเบี้ยประกันภัยพื้นฐานที่ปรับแล้ว โดยการนำเบี้ยประกันภัยพื้นฐานที่ได้จากตัวแบบเชิงเส้นวางนัยทั่วไปมาคูณกับค่าปัจจัยสมบูรณ์เพื่อใช้ในการปรับเบี้ยประกันภัยพื้นฐานจากข้อ (4)

<u>ตารางที่ 12</u> ตัวอย่างการคำนวณเบี้ยประกันภัยพื้นฐานที่ปรับด้วย IBNR แล้ว

ปี ค.ศ.	ค่าสินไหมทดแทน (ที่ยังไม่สมบูรณ์)	ค่าปัจจัยสมบูรณ์ (Ultimate Factor)	ค่าสินไหมทดแทนสมบูรณ์
	(1)	(2)	(3) = (1) * (2)
2011	10,000,000	1.0000	10,000,000
2012	10,300,000	1.0000	10,300,000
2013	10,609,000	1.0000	10,609,000
2014	10,817,997	1.0101	10,927,270
2015	10,129,579	1.1111	11,255,088
รวม	**51,856,577**	**1.0238**	**53,091,358**

Ultimate Factor (4) = 53,091,358/51,856,577	1.0238
เบี้ยประกันภัยพื้นฐาน (5)	8,000
เบี้ยประกันภัยพื้นฐานที่ปรับแล้ว (6) = (4) * (5)	8,190

นอกจากที่ผู้สร้างตัวแบบจะต้องพิจารณาถึงความสมบูรณ์ของค่าสินไหมทดแทนแล้ว ค่าสินไหมทดแทนที่เพิ่มขึ้นในทุกๆ ปีก็เป็นต้นทุนที่บริษัทประกันภัยต้องแบกรับ ซึ่งต้นทุนดังกล่าวอาจเกิดขึ้นจากค่าแรงการซ่อมบำรุงของการประกันภัยรถยนต์ที่เพิ่มขึ้นตามภาวะเงินเฟ้อ หรือค่าอะไหล่ทดแทนที่สูงขึ้นตามต้นทุนของผู้ผลิตอะไหล่ทดแทน ดังนั้นผู้สร้างตัวแบบควรพิจารณาถึงแนวโน้มต้นทุนที่เพิ่มขึ้นด้วยวิธีการดังต่อไปนี้

(1) คำนวณอัตราการเพิ่มขึ้นของต้นทุนรายปี

(2) คำนวณจุดกึ่งกลางของข้อมูลที่ใช้ในการสร้างตัวแบบ เช่น ผู้สร้างตัวแบบใช้ข้อมูลเริ่มตั้งแต่วันที่ 1 มกราคม ค.ศ. 2012 ถึงวันที่ 31 ธันวาคม ค.ศ. 2015 โดยใช้ข้อมูล ณ วันที่ 31 ธันวาคม ค.ศ. 2015 ดังนั้นจุดกึ่งกลางของข้อมูลที่ใช้ในการสร้างตัวแบบคือวันที่ 31 ธันวาคม ค.ศ. 2013

(3) คำนวณปีที่คาดว่าเบี้ยประกันภัยจากตัวแบบจะมีผลบังคับใช้ เช่น ผู้สร้างตัวแบบคาดว่าเบี้ยประกันภัยที่ได้จากตัวแบบจะเริ่มมีผลบังคับใช้ในวันที่ 1 มกราคม ค.ศ.2017 จนถึงวันที่ 31 ธันวาคม ค.ศ. 2019 ดังนั้นเบี้ยประกันภัยจะมีผลบังคับใช้ถึงวันที่ 31 ธันวาคม ค.ศ. 2020 (เนื่องจากกรมธรรม์ที่เริ่มในวันที่ 31 ธันวาคม ค.ศ. 2019 จะมีผลบังคับถึงวันที่ 31 ธันวาคม ค.ศ. 2020)

(4) คำนวณจุดกึ่งกลางของช่วงเวลาที่เบี้ยประกันภัยจากตัวแบบจะมีผลบังคับใช้จากข้อ (3) เช่น ผู้สร้างตัวแบบคาดว่าเบี้ยประกันภัยที่ได้จากตัวแบบจะเริ่มมีผลบังคับใช้ในวันที่ 1 มกราคม ค.ศ. 2017 จนถึงวันที่ 31 ธันวาคม ค.ศ. 2020 ดังนั้นจุดกึ่งกลางของช่วงเวลาที่เบี้ยประกันภัยจากตัวแบบจะมีผลบังคับใช้คือวันที่ 31 ธันวาคม ค.ศ. 2018

(5) คำนวณเบี้ยประกันภัยที่ปรับแล้ว โดยนำเบี้ยประกันภัยพื้นฐานมาปรับด้วยแนวโน้มการเพิ่มขึ้นของต้นทุนต่อปีด้วยจำนวนปีจากจุดกึ่งกลางของข้อมูลที่ใช้ในการสร้างตัวแบบจากข้อ (2) ถึงจุดกึ่งกลางของช่วงเวลาที่เบี้ยประกันภัยจากตัวแบบจะมีผลบังคับใช้จากข้อ (4)

<u>ตารางที่ 13</u> ตัวอย่างการคำนวณเบี้ยประกันภัยพื้นฐานที่ปรับด้วย IBNR และแนวโน้มของต้นทุนที่เพิ่มขึ้นแล้ว

ปี	**จำนวนครั้งของค่าสินไหมทดแทน**	**ค่าสินไหมทดแทนสมบูรณ์**	**ความรุนแรงของค่าสินไหมทดแทนเฉลี่ย**	**แนวโน้มของต้นทุนที่เพิ่มขึ้น**
	(1)	**(2)**	**(3)**	**(4)**
2011	10,000	10,000,000	1,000	
2012	10,000	10,300,000	1,030	3.0%
2013	10,000	10,609,000	1,061	3.0%
2014	10,000	10,927,270	1,093	3.0%
2015	10,000	11,255,088	1,126	3.0%
รวม	**50,000**	**53,091,358**	**1,062**	**3.0%**

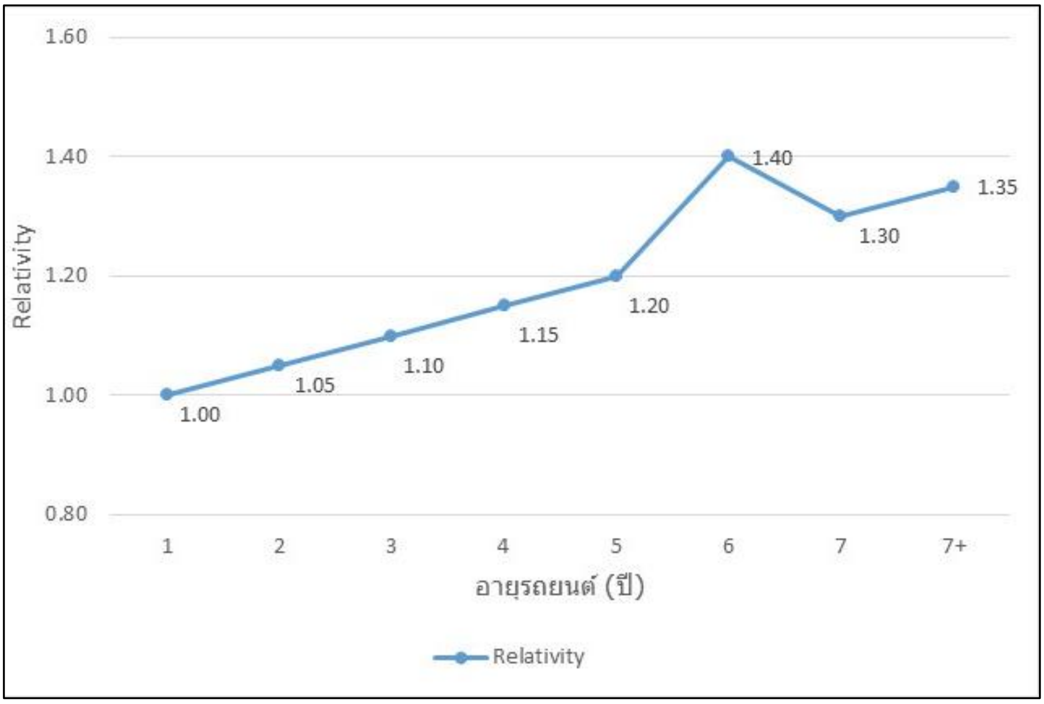
จำนวนปีจากวันที่ 31/12/2013 ถึง 31/12/2018 (5)	5
เบี้ยประกันภัยพื้นฐานที่ปรับความสมบูรณ์ของสินไหมแล้ว (6)	8,190
เบี้ยประกันภัยพื้นฐานที่ปรับความสมบูรณ์ของค่าสินไหมทดแทนและแนวโน้มแล้ว (7) = (6) * ((1 + (แนวโน้ม = 3%)) ^ (5))	9,494

4.2. การปรับปรุงตัวแบบเพื่อให้มีความสม่ำเสมอ (Smoothing)

การปรับปรุงตัวแบบเพื่อให้มีความสม่ำเสมอใช้ในกรณีที่ผู้สร้างตัวแบบต้องการปรับปรุงตัวแบบให้มีความเหมาะสมยิ่งขึ้น โดยผู้สร้างตัวแบบจะทำการรวมองค์ความรู้ของผู้สร้างตัวแบบเข้ากับตัวแบบเพื่อเพิ่ม

ประสิทธิภาพในการพยากรณ์ โดยผู้สร้างตัวแบบสามารถปรับปรุงตัวแบบด้วยวิธีที่ใช้หลักการทางคณิตศาสตร์ (เช่นการสร้างตัวแบบใหม่โดยใช้สมมติฐานแบบ polynomial) หรือไม่ก็ได้ (ปรับปรุงตัวเลขเพื่อให้ดูมีความสอดคล้องยิ่งขึ้น)

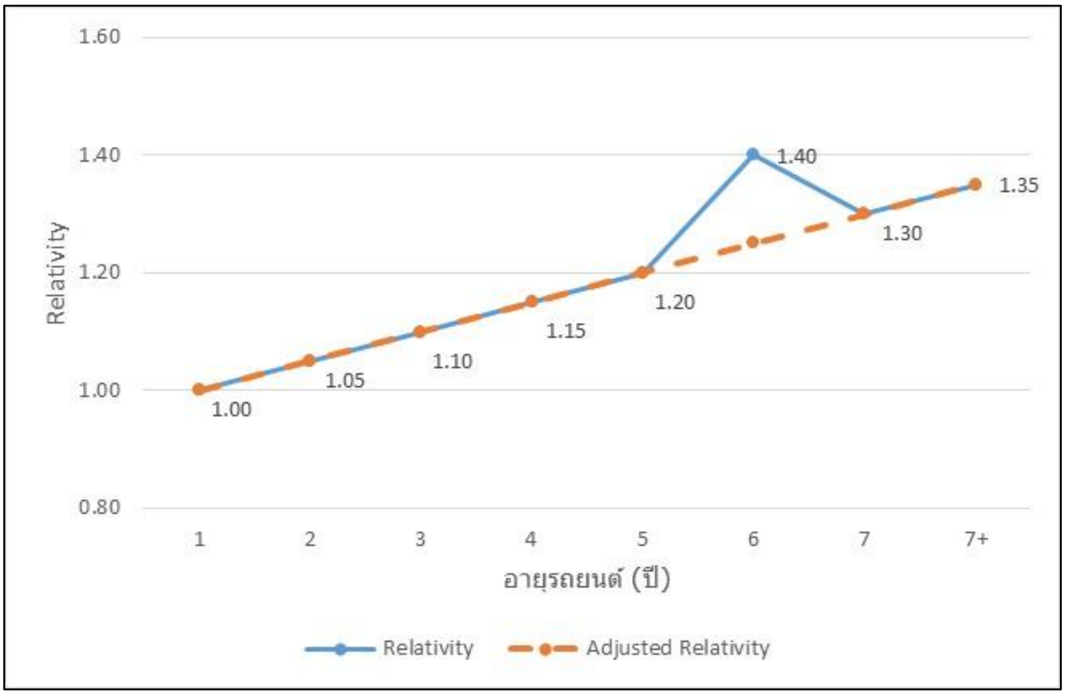
<u>ตัวอย่าง</u> Relativity ของความรุนแรงในการชนโดยเฉลี่ยแบ่งตามอายุรถยนต์จากตัวแบบเป็นไปตามกราฟในรูปที่ 23



รูปที่ 23 Relativity ของความรุนแรงในการชนโดยเฉลี่ยตามอายุรถยนต์

จากกราฟจะเห็นได้ว่ารถยนต์ที่มีอายุ 6 ปีมี relativity ของความรุนแรงในการชนโดยเฉลี่ยเท่ากับ 1.40 ในขณะที่รถยนต์ที่มีอายุ 5 ปีและ 7 ปี มี relativity ของความรุนแรงในการชนโดยเฉลี่ยเท่ากับ 1.20 และ 1.30 ตามลำดับ ซึ่งจากการตรวจสอบพบว่าสาเหตุที่ทำให้ relativity ของความรุนแรงในการชนโดยเฉลี่ยมีค่าสูงผิดปกติเกิดจากเรียกร้องค่าสินไหมทดแทนที่มีมูลค่าสูงจนผิดปกติ หรืออาจกล่าวได้ว่าลักษณะของการเรียกร้องค่าสินไหมทดแทนดังกล่าวไม่ได้เกิดขึ้นอย่างเป็นระบบ (not systematic in the experience)

จากกรณีดังกล่าวสามารถปรับปรุงตัวแบบให้มีความสม่ำเสมอโดยการปรับ relativity ของความรุนแรงในการชนโดยเฉลี่ยสำหรับรถยนต์อายุ 6 ปีให้เท่ากับ 1.25 ตามกราฟในรูปที่ 24



รูปที่ 24 การปรับ relativity ของความรุนแรงในการชนโดยเฉลี่ยสำหรับรถยนต์อายุ 6 ปี

หลังจากการปรับปรุงตัวแบบดังกล่าวแล้วควรพิจารณาว่าการปรับปรุงตัวแบบนั้นส่งผลกระทบต่อตัวแบบมากน้อยเพียงใด ซึ่งจากการปรับ relativity ของความรุนแรงในการชนโดยเฉลี่ยสำหรับรถยนต์อายุ 6 ปี จาก 1.40 ให้เป็น 1.25 ทางผู้สร้างตัวแบบเห็นว่าการปรับปรุงดังกล่าวนั้นไม่ส่งผลกระทบต่อตัวแบบอย่างรุนแรง ดังนั้นผู้สร้างตัวแบบไม่จำเป็นต้องทำการคำนวณตัวแบบ GLMs เพื่อหา relativity ใหม่

ในทางกลับกัน หากการปรับ relativity ของความรุนแรงในการชนโดยเฉลี่ยสำหรับรถยนต์อายุ 6 ปี นั้นเปลี่ยนจาก 1.80 ให้เป็น 1.25 ทางผู้สร้างตัวแบบเห็นว่าการปรับปรุงดังกล่าวเป็นการปรับปรุงค่าที่ค่อนข้างมากและควรทำการคำนวณตัวแบบ GLMs เพื่อหา relativity ใหม่ ซึ่งเรียกวิธีดังกล่าวว่าการใส่ข้อจำกัดให้ตัวแบบ (Restriction)

จะเห็นได้ว่าการปรับปรุงตัวแบบให้มีความสม่ำเสมอนั้นไม่ได้มีการระบุเอาไว้เป็นกฎตายตัว และการปรับปรุงตัวแบบนั้นขึ้นอยู่กับวิจารณญาณของผู้สร้างตัวแบบโดยตรง

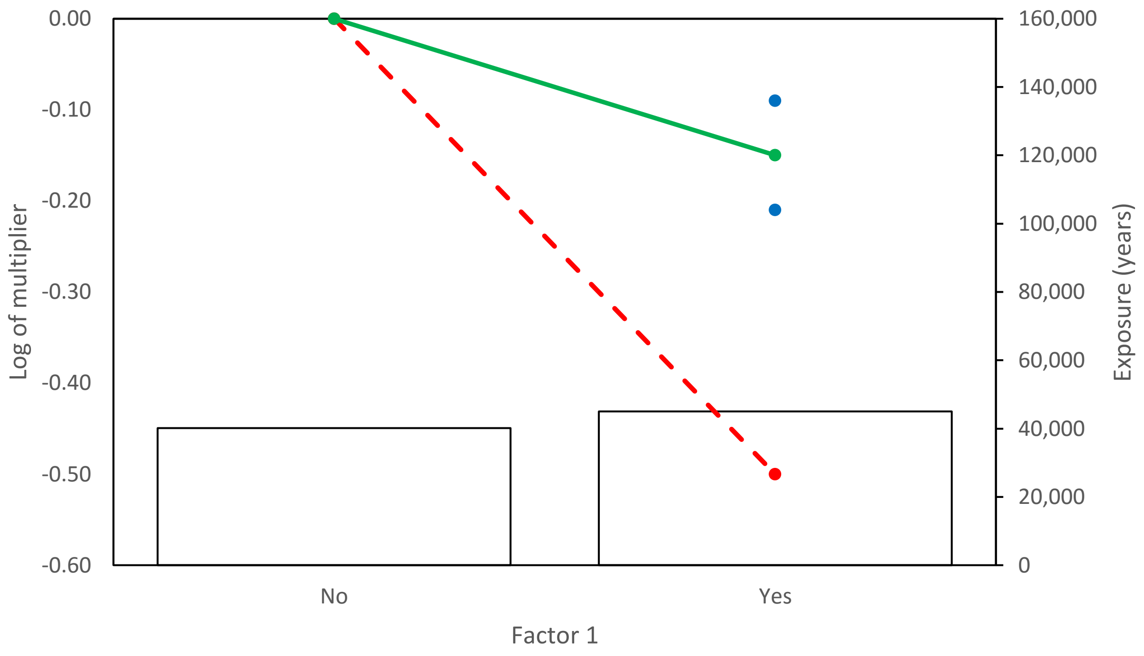
4.3. การใส่ข้อจำกัดให้ตัวแบบ (Restriction)

บ่อยครั้งที่เบี้ยประกันภัยที่แท้จริงตามทฤษฎีที่ได้จากตัวแบบ GLMs นั้นแตกต่างจากเบี้ยประกันภัยที่นำมาใช้จริง เนื่องจากในทางปฏิบัติจำเป็นต้องคำนึงถึงการแข่งขันในตลาด ข้อจำกัดในทางกฎหมาย และ

ข้อจำกัดอื่นๆ (เช่นการให้ส่วนลดประวัติดี) ซึ่งทำให้ต้องปรับปรุงตัวแบบเพื่อทดแทนส่วนที่ขาดหายไปเมื่อคำนึงถึงข้อจำกัดดังกล่าว

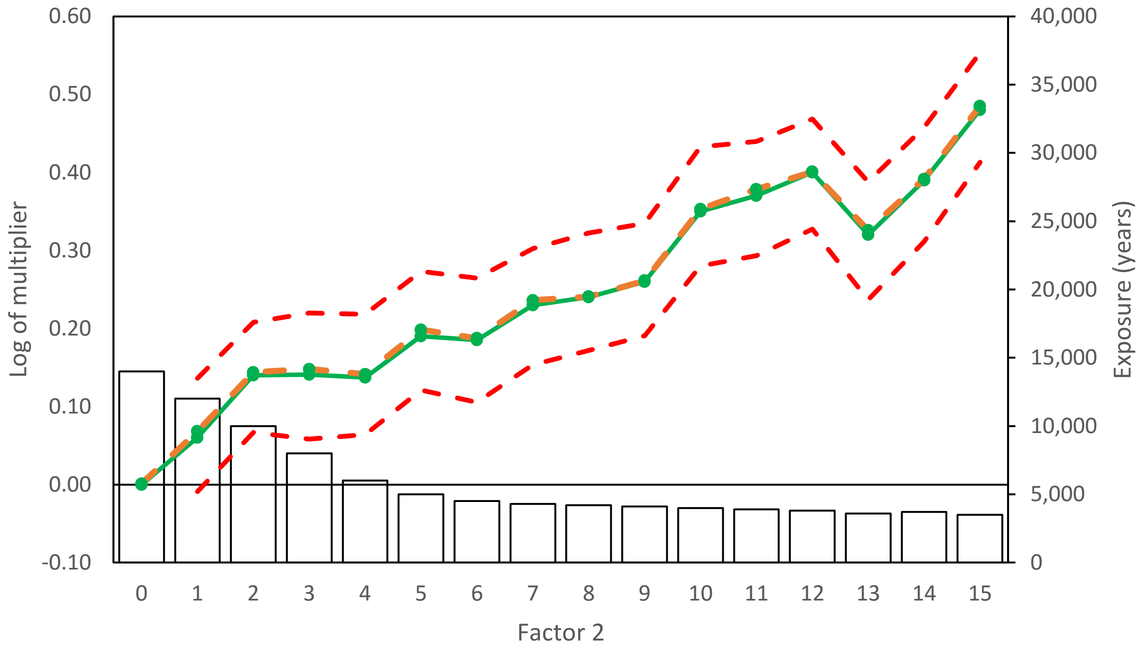
ผู้สร้างตัวแบบสามารถใส่ข้อจำกัดดังกล่าวในตัวแบบ GLMs ได้โดยการกำหนดค่าของตัวแปรที่ต้องการลงใน offset term โดยค่าที่กำหนดดังกล่าวจะถูกนำไปใช้ในการปรับข้อมูลแต่ละตัว โดยที่ตัวแปรที่ใช้ในการระบุข้อจำกัดดังกล่าวนั้นจะต้องไม่อยู่ในตัวแบบ GLMs (เปรียบเสมือนการระบุค่าพารามิเตอร์ β ที่เป็นไปตามข้อบังคับลงไปในตัวแบบ GLMs โดยตรง)

<u>ตัวอย่าง</u> กราฟเปรียบเทียบผลกระทบต่อพารามิเตอร์ของปัจจัยต่างๆ จากการปรับปรุงตัวแบบเชิงเส้นวางนัยทั่วไปด้วยข้อจำกัด

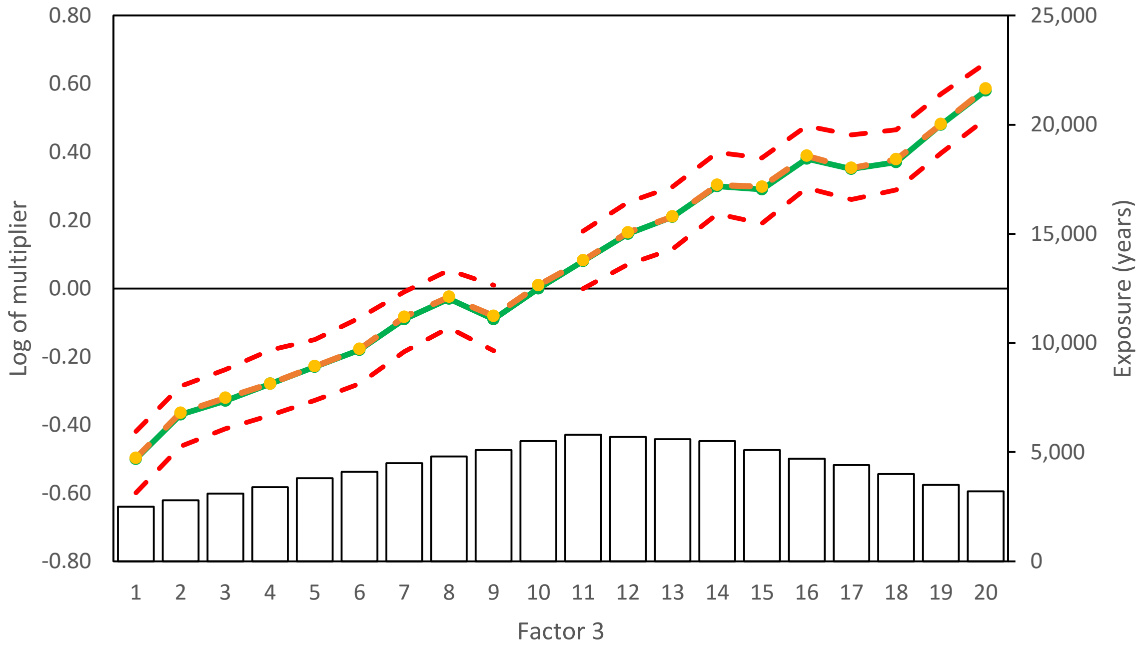


รูปที่ 25 การปรับปรุงตัวแบบเชิงเส้นวางนัยทั่วไปด้วยการใส่ข้อจำกัดลงในปัจจัย “Factor 1”

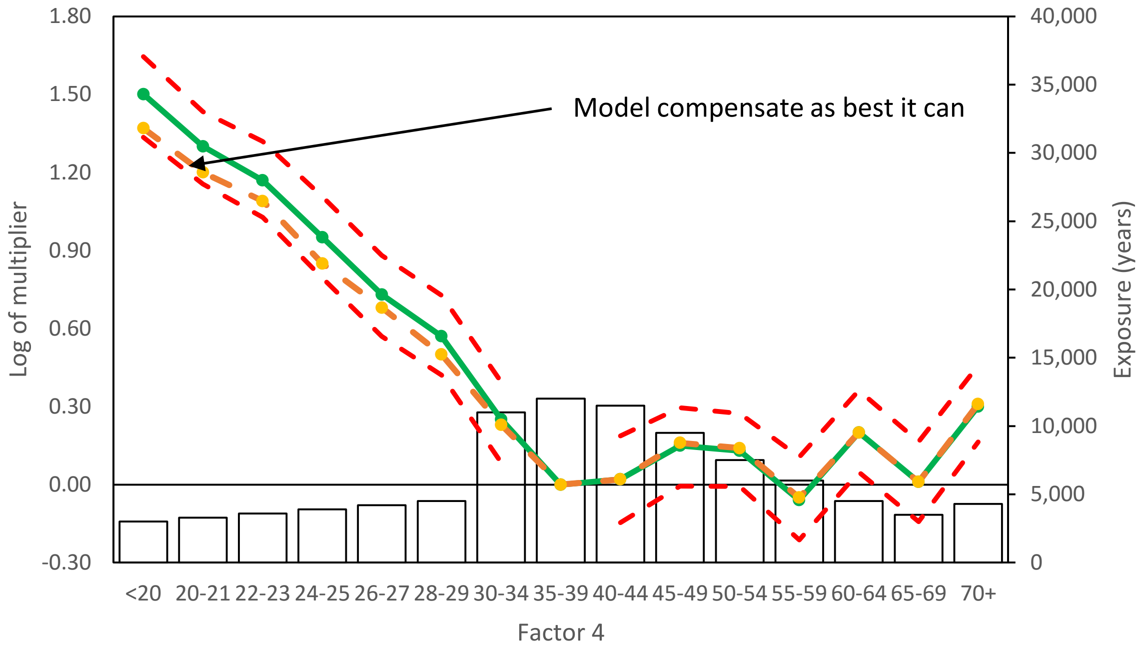
จากรูปที่ 25 เส้นทึบสีเขียวแสดงถึงพารามิเตอร์ตามทฤษฎีที่ได้จากตัวแบบเชิงเส้นวางนัยทั่วไป (จุดสีฟ้าแสดงถึงช่วงความเชื่อมั่นที่ 95% ของค่าพารามิเตอร์ดังกล่าว) และเส้นประสีแดงแสดงถึงพารามิเตอร์ที่สร้างตัวแบบปรับปรุงให้เป็นไปตามข้อจำกัด ซึ่งจากการปรับปรุงดังกล่าวส่งผลให้เกิดการชดเชยผลกระทบจากการปรับปรุงต่อปัจจัยอื่นๆ ดังรูปที่ 26 ถึง 28



รูปที่ 26 ผลกระทบจากการปรับปรุงตัวแบบเชิงเส้นวางนัยทั่วไปต่อปัจจัย “Factor 2” ด้วยการใส่ข้อจำกัดลงในปัจจัย “Factor 1”



รูปที่ 27 ผลกระทบจากการปรับปรุงตัวแบบเชิงเส้นวางนัยทั่วไปต่อปัจจัย “Factor 3” ด้วยการใส่ข้อจำกัดลงในปัจจัย “Factor 1”



รูปที่ 28 ผลกระทบจากการปรับปรุงตัวแบบเชิงเส้นวางนัยทั่วไปต่อปัจจัย “Factor 4” ด้วยการใส่ข้อจำกัดลงในปัจจัย “Factor 1”

จากรูปที่ 26 ถึง 28 เส้นทึบสีเขียวแสดงถึงพารามิเตอร์ตามทฤษฎีที่ได้จากตัวแบบเชิงเส้นวางนัยทั่วไป (เส้นประสีแดงแสดงถึงช่วงความเชื่อมั่นที่ 95% ของค่าพารามิเตอร์ดังกล่าว) และเส้นประที่มีจุดสีส้มแสดงถึงผลกระทบจากการใส่ข้อจำกัดให้กับปัจจัย “Factor 1”

ถึงแม้ว่าการใส่ข้อจำกัดนั้นสามารถทำได้ทั้งในตัวแบบ frequency และตัวแบบ severity แต่การใส่ข้อจำกัดที่เหมาะสมนั้นควรทำที่ขั้นตอนการคำนวณหาต้นทุนที่แท้จริง โดยสามารถทำได้โดยการคำนวณหาต้นทุนที่แท้จริงที่ไม่มีข้อจำกัดในขั้นต้นและทำการใส่ข้อจำกัดลงในตัวแบบสุดท้ายเพื่อหาตัวแบบที่ได้มีการปรับปรุงจากข้อจำกัดแล้ว ซึ่งจะทำให้ตัวแบบที่ได้มีความสมดุลมากที่สุดเนื่องจากตัวแบบที่ได้จะมีการปรับความสมดุลจากทุกปัจจัยในตัวแบบ

แม้ว่าผู้สร้างตัวแบบจะรู้อยู่แล้วว่าจะต้องมีการใส่ข้อจำกัดลงในตัวแบบ แต่ผู้สร้างตัวแบบก็ควรที่จะประเมินผลกระทบด้วยการสร้างตัวแบบที่ไม่มีข้อจำกัดและทำการเปรียบเทียบกับตัวแบบที่มีข้อจำกัดเพื่อเปรียบเทียบผลกระทบที่เกิดจากการใส่ข้อจำกัดลงในตัวแบบ

บทที่ 5 ภาคผนวกและเอกสารอ้างอิง

5.1. รูปแบบการแจกแจงที่สำคัญ

5.1.1 การแจกแจงแบบปกติ (Normal Distribution)

กำหนดให้ Y เป็นตัวแปนสุ่มที่การแจกแจงแบบปกติ โดยมีพารามิเตอร์ μ และ σ^2 ฟังก์ชันความหนาแน่นของ Y เขียนแทนด้วยสมการที่ (5.1)

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma^2}\exp[-\frac{1}{2\sigma^2}(y-\mu)^2] \tag{5.1}$$

จากฟังก์ชันความหนาแน่น $f(y)$ จะต้องจัดรูปแบบให้อยู่ในรูปแบบทั่วไปตามสมการด้านบน ซึ่งสามารถจัดรูปแบบได้ดังนี้

จาก $e^{\ln(y)} = y$ จะได้ดังสมการที่ (5.2) ถึงสมการที่ (5.4)

$$f(y) = \exp\left[\ln(1) - \ln(\sqrt{2\pi}\sigma^2)\right] + \exp\left[-\frac{1}{2\sigma^2}(y^2 - 2y\mu + \mu^2)\right] \tag{5.2}$$

$$f(y) = \exp\left[-\ln\left(\sqrt{2\pi}\sigma^2\right) - \frac{1}{2}\left(\frac{y^2-2y\mu+\mu^2}{\sigma^2}\right)\right] \tag{5.3}$$

$$f(y) = \exp\left[\left(\frac{y\mu-\frac{\mu^2}{2}}{\sigma^2}\right) - \frac{1}{2}\left(\frac{y^2}{2\sigma^2} + \ln(2\pi\sigma^2)\right)\right] \tag{5.4}$$

จากรูปแบบการแจกแจงในสมการที่ (5.1) ทำให้ได้พจน์ต่างๆ ดังนี้

$$y_i\theta_i = y\mu$$

$$a(\emptyset) = \sigma^2$$

$$b(\theta) = \mu^2/2$$

$$c(y_i, \emptyset) = -\frac{1}{2}\left(\frac{y^2}{2\sigma^2} + \ln(2\pi\sigma^2)\right)$$

ดังนั้น เมื่อ $Y \sim Normal(\mu, \sigma^2)$ จะได้ $b(\theta) = \mu^2/2$

5.1.2 การแจกแจงพัวซองค์ (Poisson Distribution)

ให้ Y เป็นตัวแปรสุ่มที่มีการแจกแจงแบบพัวซองค์ โดยมีพารามิเตอร์ λ ฟังก์ชันความหนาแน่นของ X เขียนแทนดังสมการที่ (5.5)

$$f(y) = \frac{e^{-\lambda}\lambda^{y}}{y!} \qquad (5.5)$$

จาก $e^{\ln(y)} = y$ จะได้

$$f(y) = \exp[-\lambda + \ln\lambda^{y} - \ln y!] \qquad (5.6)$$

$$f(y) = \exp[-\lambda + y\ln\lambda - \ln y!] \qquad (5.7)$$

ดังนั้น เมื่อ $Y \sim Poi(\lambda)$ จะได้ $b(\theta) = \lambda$

5.1.3 การแจกแจงแบบทวินาม (Binomial Distribution)

ให้ Y เป็นตัวแปรสุ่มที่มีการแจกแจงแบบทวินาม โดยมีพารามิเตอร์ np ฟังก์ชันความหนาแน่นของ Y เขียนแทนดังสมการที่ (5.8)

$$f(y) = \binom{n}{y} p^{y} q^{n-y} \quad ; \quad q = 1-p \qquad (5.8)$$

จาก $e^{\ln y} = y$ จะได้

$$f(y) = \exp\left[\ln\binom{n}{y} + \ln p^{y} + \ln(1-p)^{n-y}\right] \qquad (5.9)$$

$$f(y) = \exp\left[\ln\binom{n}{y} + \frac{\ln p^{y}}{\ln(1-p)^{y}} + n\ln(1-p)\right] \qquad (5.10)$$

ดังนั้น เมื่อ $Y \sim Binomial(np)$ จะได้ $b(\theta) = -n\ln(1-p)$

5.1.4 การแจกแจงแกมมา (Gamma Distribution)

ให้ Y เป็นตัวแปรสุ่มที่มีการแจกแจงแบบแกมมา โดยมีพารามิเตอร์ (μ, δ) ฟังก์ชันความหนาแน่นของ Y เขียนแทนได้ดังสมการที่ (5.11) ถึงสมการที่ (5.14)

$$f(y) = \left(\frac{\delta}{\mu}\right)^{\delta} \frac{1}{\Gamma(\delta)} y^{\delta-1} \exp\left[-\frac{\delta y}{\mu}\right] \qquad (5.11)$$

$$f(y) = \exp\left[\ln\left(\frac{\delta}{\mu}\right)^{\delta} + \ln\frac{1}{\Gamma(\delta)} + \ln y^{\delta-1}\right] \cdot \exp\left[-\frac{\delta y}{\mu}\right] \qquad (5.12)$$

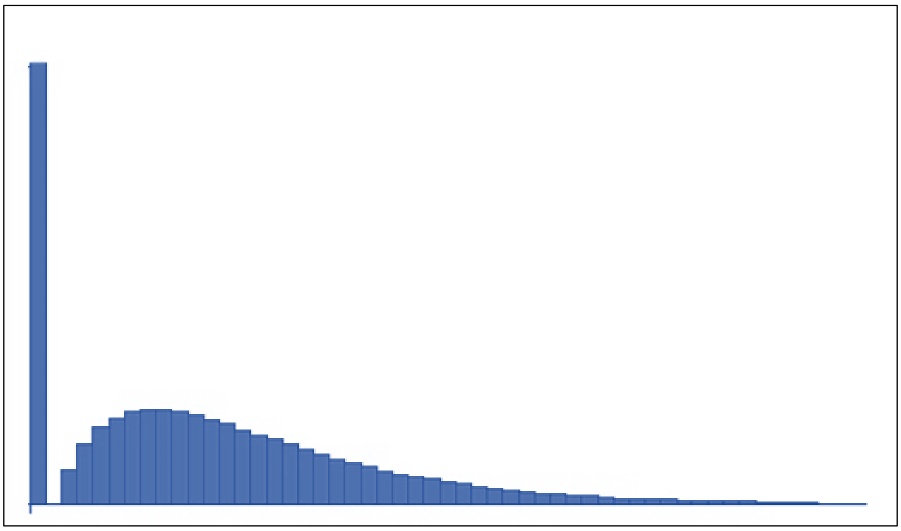
$$f(y) = \exp\left[\delta \ln\left(\frac{\delta}{\mu}\right) - \ln\Gamma(\delta) + (\delta - 1)\ln y - \frac{\delta_y}{\mu}\right] \quad (5.13)$$

$$f(y) = \exp\left[\frac{\left(-\frac{1}{\mu}y - \ln\mu\right)}{\frac{1}{\delta}} + \delta\ln\delta + (\delta - 1)\ln y - \ln(\Gamma(\delta))\right] \quad (5.14)$$

ดังนั้น เมื่อ $Y \sim Gamma\ (\mu, \delta)$ จะได้ $b(\theta) = -\ln(-\theta)$

5.1.5 การแจกแจงแบบทวิดี (Tweedie Distribution)

ในการแจกแจงรูปแบบนี้ จะมีลักษณะเฉพาะที่เป็นเอกลักษณ์ของการแจกแจงนี้ คือมีค่ากระจุกตัวอยู่ที่ “0” และมีการกระจายตัวในค่าหลังจากนั้น ซึ่งลักษณะแบบนี้เองที่มีความสอดคล้องกับข้อมูลค่าสินไหมทดแทนที่ส่วนใหญ่จะไม่เกิดการเรียกร้องค่าสินไหมทดแทนจากผู้เอาประกันภัย และระดับความเสียหายก็จะมีการกระจายตัวในค่าอื่นๆ หลังจากนั้น มีลักษณะดังกราฟด้านล่างในรูปที่ 29



รูปที่ 29 กราฟการแจกแจงแบบ Tweedie

โดยการแจกแจงแบบทวิดีจะมีค่า p เป็นพารามิเตอร์พิเศษที่ใช้บอกลักษณะของการแจกแจง และในกรณีที่ $1 < p < 2$ การแจกแจงจะมีค่า point mass ที่ 0 และมีลักษณะของการแจกแจงแบบพัวซองค์ผสมกับแกมมา โดยที่จะมีลักษณะของการแจกแจงจะเหมือนพัวซองค์ เมื่อ $p \to 1$ และเหมือนแกมมา เมื่อ $p \to 2$

โดยรวมแล้วการแจกแจงนี้มีพารามิเตอร์ที่สำคัญอยู่สามตัวคือ ค่าเฉลี่ย ค่าการกระจายตัว และพารามิเตอร์บอกรูปร่าง “p” โดยเมื่อ $1 < p < 2$ จะเขียนอยู่ในรูปของ α ได้ดังนี้

$$p = \frac{\alpha - 2}{\alpha - 1}$$

และในกรณีที่ $1 < p < 2$

$$f(y; \theta, \lambda, \alpha) = \sum_{n=1}^{\infty} \frac{\left[(\lambda\omega)^{1-\alpha}\kappa_\alpha\left(-\frac{1}{y}\right)\right]^n}{\Gamma(-n\alpha)n!y} . \exp\{\lambda\omega[\theta_0 y - \kappa_\alpha(\theta_0)]\} \quad (5.15)$$

โดยที่ $y > 0$ และ $p(Y = 0) = \exp\{-\lambda\omega\kappa_\alpha(\theta_0)\}$

เมื่อ $\kappa_\alpha(\theta) = \frac{\alpha-1}{\alpha} . \left(\frac{\theta}{\alpha-1}\right)^\alpha$

$\theta_0 = \theta\lambda^{1/(1-\alpha)}$

โดยที่ ω คือค่า Prior weight ของหน่วยเสี่ยงภัยในชุดที่ทำการสำรวจข้อมูล และฟังก์ชันความแปรปรวนของการแจกแจงแบบทวิดี สามารถเขียนได้ดังสมการที่ (5.16)

$$V(\mu) = \frac{1}{\lambda}\mu^p \quad (5.16)$$

ในทางปฏิบัติแล้ว พารามิเตอร์บอกรูปร่าง “p” สามารถสมมติให้มีค่าค่าหนึ่ง หรือประมาณเป็นส่วนหนึ่งของ Maximum Likelihood Process ก็ได้ โดยทั่วไปแล้วถ้าค่าของ p น้อยกว่า 1.5 จะเป็นการประมาณสำหรับข้อมูลค่าสินไหมทดแทนของรถยนต์

5.2. เอกสารอ้างอิง

“A Practitioner's Guide to Generalized Linear Models”. [ออนไลน์]. เข้าถึงได้จาก: https://www.casact.org/pubs/dpp/dpp04/04dpp1.pdf 2007

ส่งความคิดเห็นถึงผู้เขียน

หากผู้อ่านมีคำถาม คำแนะนำ หรือ คำติชม สามารถติดต่อมาได้ที่ feedback.iprb@tgia.org

วิสัยทัศน์และพันธกิจของสำนักงานอัตราเบี้ยประกันวินาศภัย

วิสัยทัศน์:

สร้างรากฐานธุรกิจประกันวินาศภัยให้เติบโตอย่างมั่นคงและยั่งยืน ด้วยบริการด้านข้อมูลและการคำนวณต้นทุนความเสียหายอ้างอิงที่เหมาะสมและเป็นธรรม

พันธกิจ:

จัดทำต้นทุนความเสียหายอ้างอิงของผลิตภัณฑ์ประกันวินาศภัย และให้บริการข้อมูลสารสนเทศและงานวิจัย เพื่อเป็นพื้นฐานในการดำเนินงาน การกำกับดูแล และสนับสนุนการพัฒนาของธุรกิจประกันวินาศภัย