การสร้างตัวแบบของระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทน
สำหรับการประกันภัยโรคร้ายแรง

นางสาวภรณ์ทิวา พลทิแสง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย
The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)
are the thesis authors' files submitted through the University Graduate School.

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาการประกันภัย ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2557
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

TIME DELAY BETWEEN DIAGNOSIS AND SETTLEMENT FOR CRITICAL ILLNESS INSURANCE MODELLING

Miss Porntiva Pontisang

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Insurance
Department of Statistics
Faculty of Commerce and Accountancy
Chulalongkorn University
Academic Year 2014

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การสร้างตัวแบบของระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง
โดย	นางสาวภรณ์ทิวา พลทิแสง
สาขาวิชา	การประกันภัย
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร. สุวาณี สุรเสียงสังข์

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีคณะพาณิชยศาสตร์และการบัญชี
(รองศาสตราจารย์ ดร. พสุ เดชะรินทร์)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(รองศาสตราจารย์ วัลภา ประกอบผล)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร. สุวาณี สุรเสียงสังข์)

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

.....กรรมการภายนอกมหาวิทยาลัย
(อาจารย์ ดร. บุษยมาส พิมพ์พรรณชาติ)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภรณ์ทิวา พลทิแสง : การสร้างตัวแบบของระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง (TIME DELAY BETWEEN DIAGNOSIS AND SETTLEMENT FOR CRITICAL ILLNESS INSURANCE MODELLING) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: รศ. ดร. สุวาณี สุรเสียงสังข์, 58 หน้า.

ปัจจุบันพบว่าจำนวนผู้ป่วยโรคมะเร็งและหลอดเลือดมีแนวโน้มเพิ่มขึ้นอย่างต่อเนื่อง การทำสัญญาประกันภัยโรคร้ายแรงจึงเป็นทางเลือกที่จะช่วยให้พร้อมรับความเสี่ยงที่อาจเกิดโรคดังกล่าว แต่พบว่าปัญหาที่เกิดกับบริษัทประกันภัยคือผู้เอาประกันภัยบางรายไม่ได้แจ้งข้อมูลวันที่วินิจฉัยโรค งานวิจัยนี้จึงเสนอการหาตัวแบบระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง โดยใช้ข้อมูลผู้เอาประกันภัย 840 คน ที่ทำประกันภัยโรคร้ายแรงจากบริษัทประกันชีวิต 2 แห่งในประเทศไทย โดยข้อมูลมี 6 ตัวแปร ได้แก่ อายุ เพศ รูปแบบผลประโยชน์ จำนวนเงินเอาประกันภัย บริษัทประกันภัย และสาเหตุการเรียกร้องสินไหมทดแทน โดยรูปแบบผลประโยชน์แบ่งเป็น 2 ประเภท คือประเภทครอบคลุมการเสียชีวิต และไม่ครอบคลุมการเสียชีวิต สาเหตุการเรียกร้องสินไหมทดแทนแบ่งเป็น 8 กลุ่มโรคร้ายแรง ได้แก่ โรคมะเร็ง โรคเกี่ยวกับตับ โรคเลือด โรคระบบประสาท โรคปอดอักเสบ โรคความดันเลือดสูง โรคหัวใจ และโรคอื่นๆ ข้อมูลนี้ใช้การวิเคราะห์ทางสถิติ 2 แบบ คือตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อก จากตัวแบบนี้พบว่าอายุและบริษัทประกันภัยมีผลต่อระยะเวลาที่ล่าช้า และการวิเคราะห์แบบเบส์เซียนด้วยตัวแบบเบอร์และตัวแบบล็อกนอร์มอล พบว่าตัวแบบเบอร์มีความเหมาะสมกับข้อมูลมากกว่าตัวแบบล็อกนอร์มอล ซึ่งตัวแปรที่มีผลต่อระยะเวลาที่ล่าช้าในตัวแบบเบอร์ คือ อายุ บริษัทประกันภัย และรูปแบบผลประโยชน์

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาควิชา สถิติ ลายมือชื่อนิสิต .....

สาขาวิชา การประกันภัย ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

ปีการศึกษา 2557

5481640726 : MAJOR INSURANCE
KEYWORDS: BAYESIAN ANALYSIS, CRITICAL ILLNESS INSURANCE, DIAGNOSIS-SETTLEMENT TIME LAG, GIBBS VARIABLE SELECTION, MARKOV CHAIN MONTE CARLO

PORNTIVA PONTISANG: TIME DELAY BETWEEN DIAGNOSIS AND SETTLEMENT FOR CRITICAL ILLNESS INSURANCE MODELLING. ADVISOR: SUWANEE SURASIENGSUNK, 58 pp.

The number of people suffering from cancer and stroke continue to increase progressively. Critical illness insurance is one of the effective means for coping with the disease's effects. Although this type of insurance is very successful amongst patients who have suffered from critical illness but the insurance company still faced many threats like for example some insureds did not report their diagnosis instantly. This research exhibits the models to find time delay between diagnosis and settlement for critical illness insurance. This research use data of 840 insureds who hold the critical illness insurance from two Thai life insurance companies. Six factors considered in this research are age, gender, benefit type, benefit amount, insurance company and cause of claim. Benefit type is separated in two types. One is full accelerated and the other is stand alone. Cause of claims were separated into 8 groups as follows: cancer, liver disease, blood disorder, neurological disease, lung diseases, hypertension, heart diseases and other diseases. Two statistical models are used to analyze this data set. One is the Cox Proportional Hazard model. By using this model, age and insurance company factors are discovered affecting time delay. The other is the Bayesian approach by Burr and Log-Normal model. Burr model is more suitable than Log-Normal model. The factors affecting time delay in Burr model are age, insurance company and benefit type.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Department:	Statistics	Student's Signature	.....
Field of Study:	Insurance	Advisor's Signature	.....
Academic Year:	2014		

กิตติกรรมประกาศ

วิทยานิพนธ์นี้สำเร็จได้ด้วยความอนุเคราะห์ของบุคคลหลายท่าน ซึ่งท่านแรกที่ข้าพเจ้าจะขอบพระคุณ คือ รองศาสตราจารย์ ดร. สุวาณี สุรเสียงสังข์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ที่ได้ให้ความรู้ ชี้แนะแนวทาง พร้อมทั้งคำแนะนำ ตรวจทานและแก้ไขข้อบกพร้องต่างๆ ด้วยความเอาใจใส่ ทำให้วิทยานิพนธ์ฉบับนี้เสร็จสมบูรณ์ในที่สุด

ท่านต่อไปคือ บิดาและมารดาของข้าพเจ้าที่สนับสนุนทั้งทุนทรัพย์และกำลังใจ และขอบคุณเพื่อนๆที่ให้ความช่วยเหลือ คำปรึกษา และกำลังใจ ทำให้ข้าพเจ้ามีความพยายามที่จะทำวิทยานิพนธ์ให้สำเร็จลุล่วง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์ของการวิจัย 2

1.3 ขอบเขตของงานวิจัย 2

1.4 ข้อตกลงเบื้องต้น 3

1.5 ประโยชน์ที่คาดว่าจะได้รับ 3

1.6 ขั้นตอนการดำเนินงานวิจัยโดยย่อ 3

1.7 ลำดับขั้นตอนของรายงาน 4

บทที่ 2 5

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 5

2.1. แนวคิดและทฤษฎี 5

2.2. งานวิจัยที่เกี่ยวข้อง 18

บทที่3 วิธีดำเนินการวิจัย 20

3.1 ข้อมูลที่ใช้ในการศึกษา 20

3.2 ขั้นตอนการดำเนินงาน 21

3.3 โปรแกรมที่ใช้ในงานวิจัย 29

บทที่ 4 ผลการวิเคราะห์ข้อมูล 30

4.1 การวิเคราะห์ข้อมูลพื้นฐาน 30

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

หน้า

4.2 การวิเคราะห์แบบไม่ใช้เบส์เซียน 32
4.3 การวิเคราะห์การอยู่รอดแบบเบส์เซียน 40
4.4 การเปรียบเทียบผลการวิเคราะห์แบบไม่ใช้เบส์เซียนและการวิเคราะห์แบบเบส์เซียน 52
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ 53
5.1 สรุปผลการวิจัย 53
5.2 อภิปรายผล 54
5.3 ข้อเสนอแนะ 55
รายการอ้างอิง 56
ประวัติผู้เขียนวิทยานิพนธ์ 58

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

จากรายงานสถิติสาธารณสุข พ.ศ. 2549 – 2553 พบว่าจำนวน และอัตราผู้ป่วยโรคมะเร็งและหลอดเลือด มีแนวโน้มเพิ่มขึ้นอย่างต่อเนื่อง ซึ่งหากป่วยเป็นโรคดังกล่าวจะต้องเข้ารับการรักษาตัวในโรงพยาบาล และมีค่าใช้จ่ายที่สูงมาก ดังนั้นการทำสัญญาการประกันภัยโรคร้ายแรงซึ่งคุ้มครองกลุ่มโรคมะเร็ง และโรคหลอดเลือด จึงเป็นหนึ่งทางเลือกที่จะช่วยให้พร้อมรับความเสี่ยงที่อาจเกิดโรคดังกล่าวได้ ในทางประกันภัย บริษัทจะทำการบันทึกวันที่วินิจฉัยโรค และวันที่ชำระสินไหมทดแทน เพื่อใช้คำนวณจำนวนเงินสำรองของบริษัท ซึ่งจะเตรียมไว้เพื่อจ่ายให้กับผู้เอาประกันภัย หรือผู้รับประโยชน์ เมื่อมีการเรียกร้องเข้ามาตามสิทธิ เงินสำรองประกันภัยจัดเป็นหนี้สินหรือพันธะผูกพันธ์ที่บริษัทประกันชีวิตมีต่อผู้เอาประกันภัย และผู้รับประโยชน์ แต่มีบางกรณีที่ไม่มีข้อมูลวันที่วินิจฉัยโรค จึงอาจส่งผลกระทบต่อการคำนวณกระแสเงินสดในอนาคต และอาจทำให้เกิดการผิดพลาดในการคำนวณ ส่งผลให้บริษัทไม่สามารถดำรงเงินสำรองไว้ครบถ้วนตามประกาศสำนักงานคณะกรรมการกำกับและส่งเสริมการประกอบธุรกิจประกันภัย (คปภ.) และมีเงินไม่เพียงพอจ่ายตามกรมธรรม์ให้แก่ผู้เอาประกันภัยหรือผู้รับประโยชน์ได้

สมาคมประกันชีวิตไทยเปิดเผยว่าธุรกิจประกันชีวิตไทยในปี 2557 ยังคงมีแนวโน้มการเติบโตอย่างต่อเนื่อง ด้วยอัตราการเติบโตประมาณร้อยละ 12 คิดเป็นเบี้ยประกันภัยรับรวม 496,000 ล้านบาท ซึ่งจำแนกออกเป็น เบี้ยประกันภัยรับรายใหม่จำนวน 168,000 ล้านบาท อัตราการเติบโตร้อยละ 13 และเบี้ยประกันภัยรับปีต่อไป 327,000 บาท อัตราการเติบโตร้อยละ 11 โดยมีอัตราการคงอยู่ของกรมธรรม์ประกันชีวิตที่ร้อยละ 85 เนื่องจากมีช่องทางการจำหน่ายที่หลากหลาย และมีการเข้าถึงผู้บริโภคทุกกลุ่มโดยใช้เทคโนโลยีที่ทันสมัยเข้ามาดำเนินการ อีกทั้งมีการพัฒนาผลิตภัณฑ์ประกันชีวิตและสามารถตอบสนองความต้องการผู้บริโภคในทุกระดับ ส่งผลให้ผู้บริโภคหันมาใส่ใจและให้ความสำคัญกับเรื่องของการประกันชีวิต แม้ว่าเศรษฐกิจจะมีการชะลอตัวลงบ้างเล็กน้อย เนื่องมาจากสภาวะเศรษฐกิจโดยทั่วไปที่ส่งผลต่อกำลังซื้อของผู้บริโภค การจำกัดอัตราการปล่อยสินเชื่อของธนาคารพาณิชย์ รวมถึงปัจจัยความเสี่ยงทางการเมืองที่มีความผันผวนอยู่ในขณะนี้ แต่เชื่อว่าจะไม่ส่งผลกระทบต่อธุรกิจประกันชีวิต โดยจะเห็นได้จากการเพิ่มขึ้นของกรมธรรม์ที่เน้นด้านความคุ้มครองและแนวโน้มสินค้าประเภทนี้จะได้รับความนิยมขึ้นอีก โดยเฉพาะสินค้าที่ให้ความคุ้มครองโรคร้ายแรงและการประกันสุขภาพที่ได้รับความสนใจมากขึ้น

(สำนักงานคณะกรรมการกำกับและส่งเสริมการประกอบธุรกิจประกันภัย (คปภ.) รวบรวมโดยสำนักงานเศรษฐกิจการคลัง, ปี 2557)

วิธีการหนึ่งที่ใช้แก้ปัญหากรณีที่ขาดข้อมูลวันที่วินิจฉัยโรค คือการหาตัวแบบระยะเวลาที่ล่าช้าระหว่างเวลาที่ทำการวินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทน โดยใช้การวิเคราะห์แบบเบส์เซียน (Bayesian Analysis) ซึ่งอันดับแรกต้องวิเคราะห์ก่อนว่าการแจกแจงความน่าจะเป็นแบบใดเหมาะสมกับข้อมูลระยะเวลาที่ล่าช้านี้ และปัจจัยใดบ้างที่ส่งผลกระทบต่อระยะเวลาระหว่างสองช่วงเวลานี้ จากนั้นทำการหาตัวแบบที่เหมาะสมที่สุดที่ใช้พยากรณ์ความล่าช้าระหว่างเวลาที่ทำการวินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทน

โดยในอดีตที่ผ่านมาได้มีนักวิจัยหลายท่านเสนอวิธีการวิเคราะห์แบบเบส์เซียนมาใช้ในการวิเคราะห์ข้อมูลระยะเวลา เช่น ในปี ค.ศ. 2002 ซินยังซอง และ ซิกยำลี ได้คัดเลือกตัวแบบโดยใช้วิธีวิเคราะห์แบบเบส์เซียน เพื่อประมาณค่าเบส์แฟคเตอร์ และสังเกตการจำลองแบบจากการแจกแจงหลัง (posterior distribution) ของพารามิเตอร์ นอกจากนี้ในปี ค.ศ. 2011 ออคคอก ได้นำตัวแบบเบส์เซียนมาวิเคราะห์การแจกแจงความน่าจะเป็น และประมาณค่าพารามิเตอร์ในตัวแบบด้วยวิธีมาร์คอฟ เชน มอนติคาร์โล (Markov Chain Monte Carlo : MCMC)

ดังนั้นงานวิจัยนี้จึงสนใจหาตัวแบบระยะเวลาที่ล่าช้าระหว่างเวลาที่ทำการวินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง ของบริษัทประกันชีวิตในประเทศไทย โดยใช้การวิเคราะห์การอยู่รอดแบบไม่ใช้เบส์เซียน และการวิเคราะห์แบบเบส์เซียน

1.2 วัตถุประสงค์ของการวิจัย

1. หาปัจจัยที่มีผลต่อระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง
2. หาตัวแบบของระยะเวลาระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนของการประกันภัยโรคร้ายแรง ตามการวิเคราะห์การอยู่รอดแบบไม่ใช้เบส์เซียน (Non Bayesian) และแบบเบส์เซียน (Bayesian)

1.3 ขอบเขตของงานวิจัย

ข้อมูลที่ใช้สร้างตัวแบบเป็นข้อมูลระยะเวลาระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรงระหว่างปี พ.ศ. 2550 – 2556 ของบริษัทประกันชีวิต 2 แห่ง โดยมีผู้เอาประกันภัยทั้งหมดจำนวน 840 ราย โดยมีเงื่อนไขว่าข้อมูลที่ใช้เป็นข้อมูล

ของการประกันภัยโรคร้ายแรงที่ชำระค่าสินไหมทดแทนเพียงครั้งเดียว คือ ไม่มีการจ่ายเป็นรายงวดในระหว่างที่ผู้เอาประกันภัยทำการรักษาโรค

1.4 ข้อตกลงเบื้องต้น

1. การศึกษานี้ใช้การวิเคราะห์แบบเบส์เซียนในการหาตัวแบบที่เหมาะสมกับระยะเวลาที่ล่าช้า โดยกำหนดพารามิเตอร์ก่อน (prior parameter) เป็นแบบ Zellner' s g – prior
2. การวิเคราะห์ค่าประมาณกำหนดระดับเปอร์เซ็นไทล์ (Percentile) ที่ 2.5 ถึง 97.5

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. บริษัทประกันภัยได้ตัวแบบทางสถิติของระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทน เพื่อใช้พยากรณ์ข้อมูลเวลาที่ขาดหายไปของผู้เอาประกันภัยบางราย
2. ทราบปัจจัยที่ส่งผลต่อระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง

1.6 ขั้นตอนการดำเนินงานวิจัยโดยย่อ

1. ศึกษาการวิเคราะห์แบบไม่ใช้เบส์เซียน ที่ใช้ตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อก
2. ศึกษาการประมาณค่าพารามิเตอร์ในตัวแบบเบส์ด้วยวิธีมาร์คอฟ เชน มอนติคาร์โล
3. เก็บรวบรวมข้อมูลที่เกี่ยวข้องพร้อมปรับข้อมูลให้เหมาะสมกับการวิเคราะห์
4. วิเคราะห์ข้อมูลความล่าช้าระหว่างวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทนร่วมกับปัจจัยที่มีผลต่อความล่าช้าด้วยตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อก
5. หาตัวแบบที่เหมาะสมกับข้อมูลความล่าช้าโดยไม่รวมข้อมูลเวลาที่ขาดหายไป (missing value) ด้วยการวิเคราะห์การอยู่รอดแบบเบส์เซียน ซึ่งคัดเลือกตัวแปรด้วยวิธีการคัดเลือกตัวแปรของกิบส์ โดยกำหนดพารามิเตอร์เดิม (prior parameter) เป็น Zellner' s g – prior และหาค่าประมาณหลัง (posterior estimates) ของพารามิเตอร์
6. เปรียบเทียบค่าประมาณหลัง (posterior estimate) ของพารามิเตอร์ภายใต้ตัวแบบที่เหมาะสม ในระดับเปอร์เซ็นไทล์ที่ 2.5 ถึง 97.5 เพื่อวิเคราะห์หาปัจจัยที่ส่งผลให้เกิดความคลาดเคลื่อนของความล่าช้า
7. เขียนรายงานและทำการสรุปผลข้อมูล

1.7 ลำดับขั้นตอนของรายงาน

วิทยานิพนธ์เล่มนี้มีจุดประสงค์ที่จะนำเสนอและเปรียบเทียบตัวแบบของระยะเวลาระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนของการประกันภัยโรคร้ายแรง ด้วยการวิเคราะห์แบบไม่ใช้เบส์เซียน และแบบเบส์เซียน ซึ่งแบ่งออกเป็น 6 บท โดยบทที่ 1 กล่าวถึงความเป็นมาและความสำคัญของปัญหา วัตถุประสงค์และขอบเขตในการทำงานวิจัยนี้ ส่วนในบทที่ 2 อธิบายหลักการและทฤษฎีที่เกี่ยวข้องที่ใช้ในการหาตัวแบบของระยะเวลาระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนของการประกันภัยโรคร้ายแรง ตามการวิเคราะห์แบบไม่ใช้เบส์เซียน และแบบเบส์เซียน บทที่ 3 กล่าวถึงข้อมูลที่ใช้ วิธีการดำเนินการ และโปรแกรมที่ใช้ในการวิจัย บทที่ 4 แสดงผลการวิเคราะห์ข้อมูล และบทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1. แนวคิดและทฤษฎี

ในการศึกษาครั้งนี้จะนำการวิเคราะห์ข้อมูลของระยะเวลาถึงเหตุการณ์ที่สนใจ (Time to event) มาประยุกต์ใช้ในการวิเคราะห์ระยะเวลาระหว่างเวลาที่ทำการวินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง และศึกษาความสัมพันธ์ระหว่างระยะเวลาถึงเหตุการณ์ที่สนใจกับตัวแปรอิสระ ในที่นี้เราสนใจตัวแปรอิสระที่ส่งผลต่อความล่าช้าระหว่างเวลาที่ทำการวินิจฉัยโรค และเวลาที่เรียกร้องสินไหมทดแทน ดังนั้นกระบวนการคัดเลือกตัวแปรอิสระเข้าสู่ตัวแบบด้วยวิธีที่เหมาะสมจึงเป็นสิ่งสำคัญ ในหัวข้อนี้จะนำเสนอเกี่ยวกับรายละเอียดของการคัดเลือกตัวแปรโดยใช้การคัดเลือกตัวแปรของกิบส์ และการคัดเลือกตัวแปรโดยใช้ภาวะน่าจะเป็นส่วนริมภายใต้การประมาณค่าพารามิเตอร์ในฟังก์ชันการแจกแจงความน่าจะเป็นด้วยทฤษฎีของเบส์ โดยกำหนดพารามิเตอร์ก่อน (prior parameter) เป็น Zellner' s g – prior แต่ในหลายกรณีมีความยากมากที่จะหาตัวประมาณเบส์ หรือค่าประมาณเบส์ด้วยวิธีการเชิงวิเคราะห์ เช่น แคลคูลัส และพีชคณิตเชิงเส้น จึงจำเป็นต้องใช้การประมาณ เช่น วิธีเชิงตัวเลข และการจำลอง เป็นต้น ดังนั้นในที่นี่จะนำเสนอการหาค่าประมาณด้วยวิธีการประมาณค่าพารามิเตอร์ที่ไม่ทราบค่าด้วยวิธีมาร์คอฟ เชน มอนติคาร์โล

เริ่มต้นจะอธิบายแนวคิด และลักษณะความคุ้มครองของการประกันภัยโรคร้ายแรง จากนั้นจะเสนอปัจจัยที่คาดว่ามีผลต่อระยะเวลาที่ล่าช้าระหว่างวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทน และอธิบายการวิเคราะห์ระยะเวลาถึงเหตุการณ์ที่สนใจ ซึ่งแยกประเภทของข้อมูลที่ไม่สมบูรณ์แบ่งออกเป็น 2 ประเภท ได้แก่ข้อมูลที่มีการเซนเซอร์ (Censored data) และข้อมูลที่มีการตัดเป็นท่อน (Truncation) จากนั้นจะอธิบายการวิเคราะห์การอยู่รอดแบบไม่ใช้เบส์เซียน ซึ่งใช้ตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อก ส่วนการวิเคราะห์การอยู่รอดแบบเบส์เซียน ซึ่งจะใช้ตัวแบบเบอร์ และตัวแบบล็อกนอร์มอล ในการวิเคราะห์การอยู่รอดแบบเบส์เซียนจะอธิบายขั้นตอนการหาความน่าจะเป็นหลัง (posterior probability) จากนั้นจึงอธิบายการคัดเลือกตัวแปรด้วยวิธีการคัดเลือกตัวแปรของกิบส์ (Gibbs variable Selection) โดยกำหนดพารามิเตอร์ก่อน (prior parameter) เป็น Zellner' s g – prior ซึ่งจะอธิบายลักษณะของการแจกแจงของเวกเตอร์สัมประสิทธิ์ และการหาค่าของตัวประมาณภายหลัง (posterior estimate)

2.1.1. **การประกันภัยโรคร้ายแรง** (Critical Illness Insurance)

แนวคิดของการทำประกันภัยโรคร้ายแรง คือเพื่อแบ่งเบาหรือบรรเทาค่ารักษาพยาบาล และค่าใช้จ่ายจำนวนมากซึ่งมีโอกาสเกิดขึ้นโดยที่ไม่คาดหวังหรือตั้งใจ เหมือนเป็นการลงทุนซื้อเครื่องมือป้องกันความเสี่ยงเกี่ยวกับความสูญเสียทางด้านการเงินที่เราต้องเผชิญ สำหรับผู้ที่ไม่มีประกันภัยเพื่อรองรับโรคร้ายแรง อาจถึงกับต้องสูญเสียทรัพย์สมบัติเพื่อนำมาเป็นค่ารักษาพยาบาลหากเกิดการเจ็บป่วยขึ้น

การประกันภัยโรคร้ายแรงเป็นผลิตภัณฑ์ของทั้งบริษัทประกันชีวิต และบริษัทประกันวินาศภัย

การประกันภัยโรคร้ายแรงในบริษัทประกันชีวิต

การประกันภัยโรคร้ายแรงในบริษัทประกันชีวิตจัดอยู่ในสัญญาเพิ่มเติม

สัญญาเพิ่มเติมการประกันภัยโรคร้ายแรง เป็นสัญญาที่ให้ผลประโยชน์และความคุ้มครองในกรณีเสียชีวิต ทุพพลภาพและเป็นโรคร้ายแรงที่นอกเหนือจากผลประโยชน์และความคุ้มครองจากแบบประกันชีวิตหลักซึ่งบริษัทจะจ่ายผลประโยชน์เท่ากับทุนประกันภัยที่ได้ระบุไว้ของสัญญาเพิ่มเติมนี้ เมื่อเกิดเหตุการณ์ใดเหตุการณ์หนึ่งต่อไปนี้กับผู้เอาประกันภัย

- ผู้เอาประกันภัยเสียชีวิต หรือ
- ผู้เอาประกันภัยตกอยู่ในสภาวะทุพพลภาพสิ้นเชิงถาวร หรือ
- ผู้เอาประกันภัยได้รับการวินิจฉัยว่าเป็นโรคร้ายแรงตามคำนิยามที่ได้กำหนดไว้ในสัญญาเพิ่มเติมโดยการนิยามโรคร้ายแรงในแต่ละกรมธรรม์นั้นแตกต่างกัน คือบางกรมธรรม์ระบุ 7 โรค ได้แก่ โรคมะเร็ง กล้ามเนื้อหัวใจตายจากการขาดเลือด โรคหลอดเลือดสมอง การศัลยกรรมหลอดเลือดเลี้ยงหัวใจ ไตวายการผ่าตัดเปลี่ยนอวัยวะ และอัมพาต บางกรมธรรม์ระบุ 5 โรค ได้แก่ โรคมะเร็ง โรคกล้ามเนื้อหัวใจตายเฉียบพลันจากการขาดเลือด โรคหลอดเลือดสมองแตกหรืออุดตัน โรคสมองเสื่อม และทุพพลภาพสิ้นเชิงถาวร

การประกันภัยโรคร้ายแรงในบริษัทประกันวินาศภัย

การประกันภัยโรคร้ายแรงในบริษัทประกันวินาศภัยจัดอยู่ในรูปแบบประกันภัยที่ให้ความคุ้มครองผู้ที่เป็นโรคร้ายแรงโดยตรง และในรูปแบบสัญญาเพิ่มเติมของการประกันภัย

- **แบบประกันภัยที่ให้ความคุ้มครองผู้ที่เป็นโรคร้ายแรงโดยตรง**

แบบประกันภัยที่ให้ความคุ้มครองผู้ที่เป็นโรคร้ายแรงโดยตรงเป็นการประกันภัยที่ให้ผลประโยชน์และความคุ้มครองในกรณีที่ผู้เอาประกันถูกวินิจฉัยว่าเป็นโรคตามที่ระบุไว้ในเงื่อนไขของกรมธรรม์ โดยในแต่ละกรมธรรม์จะมีรูปแบบความคุ้มครองที่ต่างกัน เช่นบางกรมธรรม์ให้ความ

คุ้มครองโรคมะเร็งเพียงโรคเดียว บางกรมธรรม์ให้ความคุ้มครอง 10 โรคร้ายแรง ได้แก่ โรคมะเร็งระยะลุกลาม โรคกล้ามเนื้อหัวใจตายครั้งแรกจากการขาดเลือด โรคหลอดเลือดสมอง โรคไตวายเรื้อรังที่ต้องได้รับการล้างไตอย่างสม่ำเสมอ โรคตับวาย โรคปอดระยะสุดท้าย การผ่าตัดปลูกถ่ายอวัยวะสำคัญ ภาวะหมดสติ การทุพพลภาพถาวรสิ้นเชิง โดยส่วนมากจะจ่ายเงินก้อนเต็มวงเงินผลประโยชน์ให้ผู้เอาประกันทันทีเพียงแสดงหลักฐานทางการแพทย์ที่ได้รับการวินิจฉัยว่าเป็นโรค

- **สัญญาเพิ่มเติมของการประกันภัยโรคร้ายแรง**

สัญญาเพิ่มเติมการประกันภัยโรคร้ายแรงที่มีจำหน่ายในบริษัทประกันวินาศภัย ส่วนมากจะเป็นสัญญาเพิ่มเติมของการประกันสุขภาพโดยให้ผลประโยชน์และความคุ้มครองในกรณีเป็นโรคร้ายแรงตามเงื่อนไขที่กรมธรรม์กำหนด ซึ่งบริษัทประกันภัยจะจ่ายผลประโยชน์สำหรับค่าใช้จ่ายที่จำเป็นและสมควร ที่เกิดจากการรักษาพยาบาลตามความจำเป็นทางการแพทย์ให้ตรงตามผลประโยชน์ที่ระบุไว้ในสัญญาเพิ่มเติม

จากข้อมูลของสำนักงานปลัดกระทรวงสาธารณสุขย้อนหลังตั้งแต่ปี พ.ศ.2549 ถึง 2556 เกี่ยวกับสาเหตุการตายที่สำคัญพบว่าโรคมะเร็งเป็นสาเหตุการตายอันดับ 1 ตามด้วยอุบัติเหตุ ส่วนโรคหัวใจเป็นสาเหตุการตายอันดับ 3 ทั้งนี้ โรคมะเร็ง และโรคหัวใจ เป็นโรคที่ต้องการการรักษาอย่างทันท่วงที ในหลายๆกรณีต้องได้รับการดูแลรักษาในระยะยาว หากตรวจพบอาการของโรคได้เร็วจะส่งผลให้อัตราการรอดชีวิตเพิ่มสูงขึ้นด้วย จากสถานการณ์ดังกล่าวได้ก่อให้เกิดภาระค่าใช้จ่ายในการดูแลรักษาสูงมาก และประเด็นที่น่าสนใจคือ ค่าใช้จ่ายในส่วนนี้อาจกลายเป็นวิกฤตทางการเงินได้หากการเจ็บป่วยนั้น ทำให้ต้องออกจากงาน หรือขาดรายได้ประจำ และหากมองในแง่รายได้เฉลี่ยต่อหัวของประชากรไทยที่ 150,118 บาทต่อปี (จีดีพี/ยอดรวมประชากรไทยทั้งหมด) รวบรวมโดยสำนักสถิติพยากรณ์ สำนักงานสถิติแห่งชาติ ซึ่งผลสำรวจนี้บ่งชี้สัญญาณอันตรายทางการเงินที่คนไทยส่วนใหญ่ต้องเผชิญหากไม่ได้จัดการกับความเสี่ยงนี้

2.1.2. ปัจจัยที่คาดว่าจะเกี่ยวข้องกับระยะเวลาที่ล่าช้าระหว่างวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทน

จากการศึกษาของออคคอค และคณะ ในปี ค.ศ. 2012 พบว่า อายุ เพศ จำนวนเงินเอาประกันภัย ระยะเวลาที่กรมธรรม์ให้ความคุ้มครอง และสาเหตุของการเรียกร้องสินไหมทดแทนมีผลต่อระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทน (ออคคอค และคณะ ,2012) ในการศึกษาครั้งนี้จึงนำปัจจัยดังกล่าวมาพิจารณาด้วยดังนี้

- อายุ

ช่วงอายุที่แตกต่างกันจะมีโอกาสในการเกิดโรคร้ายแรงที่ต่างกันด้วย เช่น ผู้หญิงที่มีอายุมากกว่า 40 ปีขึ้นไป จะมีความเสี่ยงต่อการเป็นมะเร็งเต้านม โดยพบบ่อยในหญิงที่มีอายุ 50 ปีขึ้นไป

และมะเร็งลำไส้ใหญ่ที่อาจเกิดขึ้นได้ในทุกอายุ แต่อัตราเสี่ยงของการเกิดมะเร็งลำไส้ใหญ่ จะสูงขึ้นในผู้สูงอายุ ผู้ป่วยส่วนใหญ่มักจะถูกวินิจฉัยในช่วงอายุระหว่าง 50-70 ปี(เอกสารแจกจากโรงพยาบาลสมิติเวช อ้างใน www.moseng169.in.th)

- เพศ

โรคร้ายแรงบางชนิดมีโอกาสเกิดขึ้นในแต่ละเพศแตกต่างกัน ข้อมูลจากทะเบียนมะเร็งของสถาบันมะเร็งแห่งชาติในปี พ.ศ. 2553 พบว่า โรคมะเร็งในเพศชายที่พบบ่อยเรียงตามลำดับคือ มะเร็งปอด มะเร็งลำไส้ใหญ่ มะเร็งตับ มะเร็งท่อน้ำดี และมะเร็งหลอดอาหาร ส่วนโรคมะเร็งในเพศหญิงที่พบบ่อยเรียงตามลำดับคือ มะเร็งเต้านม มะเร็งลำไส้ใหญ่ มะเร็งปอด และมะเร็งมดลูก

- จำนวนเงินเอาประกันภัย

กรมธรรม์ที่มีจำนวนเงินเอาประกันภัยมากอาจส่งผลให้ระยะเวลาที่ล่าช้าระหว่างวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทนนั้นมากกว่ากรมธรรม์ที่มีจำนวนเงินเอาประกันภัยน้อย เนื่องจากกรมธรรม์ที่มีจำนวนเงินเอาประกันภัยมากอาจให้ความคุ้มครองมากกว่ากรมธรรม์ที่มีจำนวนเงินเอาประกันภัยน้อย

- ระยะเวลาที่กรมธรรม์ให้ความคุ้มครอง

กรมธรรม์ที่ระยะเวลาที่ให้ความคุ้มครองยาวนานอาจส่งผลให้ระยะเวลาที่ล่าช้าระหว่างวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทนนั้นมากกว่ากรมธรรม์ที่ให้ความคุ้มครองสั้นกว่า

- สาเหตุของการเรียกร้องสินไหมทดแทน

สาเหตุของการเรียกร้องสินไหมทดแทน คือ โรคร้ายแรงที่เกิดขึ้น ซึ่งโรคร้ายแรงบางชนิดมีระดับความรุนแรงมาก เช่น โรคมะเร็ง เมื่อพบว่าเป็นโรคแล้ว อาจมีระยะเวลาในการอยู่รอดไม่มาก จึงส่งผลให้ระยะเวลาที่ล่าช้าระหว่างวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทนนั้นน้อยตามไปด้วย ต่างจากโรคร้ายแรงบางชนิดที่เมื่อเป็นแล้วได้รับการรักษาที่ถูกต้องก็ส่งผลให้ดำเนินชีวิตได้ตามปกติ จึงส่งผลให้ระยะเวลาที่ล่าช้าระหว่างวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทนยาวนานมากขึ้น

2.1.3. การวิเคราะห์ข้อมูลของระยะเวลาถึงเหตุการณ์ที่สนใจ (Time to event)

การวิเคราะห์ข้อมูลของระยะเวลาถึงเหตุการณ์ที่สนใจ (Time to event) คือการพิจารณาหารูปแบบของระยะเวลาดังกล่าว โดยการวิเคราะห์ทำได้ 2 รูปแบบ คือแบบพาราเมตริกซ์ (parametric) และแบบไม่ใช่พาราเมตริกซ์(nonparametric) โดยแบบพาราเมตริกซ์จะเป็นการพิจารณาหาการแจกแจงความน่าจะเป็นของระยะเวลาดังกล่าว ส่วนแบบไม่ใช่พาราเมตริกซ์จะไม่หาการแจกแจงความน่าจะเป็นของระยะเวลานั้น

การวิเคราะห์ข้อมูลของระยะเวลาถึงเหตุการณ์ที่สนใจเป็นเทคนิคที่นำไปใช้ในการศึกษาได้หลายสาขาวิชา เช่น สถิติการประกันภัย แพทยศาสตร์ สังคมศาสตร์ ประชากรศาสตร์ เป็นต้น

ตัวอย่างข้อมูลของระยะเวลาถึงเหตุการณ์ที่สนใจ

- ในงานวิจัยที่มีการติดตามผู้ป่วยมะเร็งปอด หากสนใจผู้ป่วยมะเร็งปอดที่หายจากโรคเมื่อได้รับการรักษาด้วยยาชนิดหนึ่ง เหตุการณ์ (event) คือ การหายจากโรคมะเร็งปอดหลังจากได้รับการรักษาด้วยยาชนิดหนึ่ง ระยะเวลา (time) คือ ระยะเวลาที่นับจากรับการรักษาด้วยยาจนถึงการหายจากโรคโดยระยะเวลานี้เรียกว่า เวลาการอยู่รอด (Survival time)
- ในงานวิจัยที่มีการติดตามนิสิตที่จบการศึกษาแล้ว หากสนใจนิสิตที่จบการศึกษาแล้วจะได้งานทำครั้งแรก เหตุการณ์ (event) ที่สนใจ คือ การได้งานครั้งแรกหลังเรียนจบ ระยะเวลา (time) คือ ระยะเวลาที่นับจากสำเร็จการศึกษาจนถึงการได้งานทำครั้งแรก

กำหนดให้ T แทนระยะเวลาถึงเหตุการณ์ที่สนใจ เมื่อพิจารณาให้ T เป็นตัวแปรสุ่มจะได้ฟังก์ชันการแจกแจงสะสมคือ $F(t)=\Pr(T\le t)$ และฟังก์ชันความหนาแน่นน่าจะเป็นคือ $f(t)=\dfrac{\partial F(t)}{\partial t}$ สำหรับการเก็บข้อมูลการอยู่รอด สามารถแบ่งลักษณะข้อมูลเป็น 2 กรณีคือ กรณีที่1 คือกรณีที่เหตุการณ์ที่สนใจไม่เกิดขึ้นภายในช่วงเวลาที่กำหนดหรือสูญหายขณะทดลอง สามารถเขียนความน่าจะเป็นของเหตุการณ์ดังกล่าวด้วยฟังก์ชันการอยู่รอด (Survival function) ซึ่งเขียนได้ในรูป

$$S(t)=\Pr(T>t)=1-F(t)=1-\Pr(T\le t)$$

และกรณีที่ 2 คือกรณีที่เหตุการณ์ที่สนใจเกิดขึ้นภายในช่วงเวลาที่กำหนด ณ เวลา t เราสามารถเขียนความน่าจะเป็นของเหตุการณ์ที่บ่งบอกความเสี่ยงดังกล่าวได้จากฟังก์ชันความเสี่ยง (Hazard function) ที่เขียนได้ในรูป

$$h(t)=\lim_{\Delta t\to 0}\frac{\Pr[t\le T<t+\Delta t\,|\,T\ge t]}{\Delta t}$$

เมื่อ t คือจุดเริ่มต้นของเวลาที่สนใจทำการศึกษา และ T คือจุดสิ้นสุดของเวลาที่ทำการศึกษา

ถ้า $f(t)$ คือฟังก์ชันความหนาแน่นของความน่าจะเป็น เราสามารถเขียนความสัมพันธ์ของฟังก์ชันการอยู่รอดและฟังก์ชันความเสี่ยงในรูป

$$h(t)=\frac{f(t)}{S(t)}=-\frac{d\ln[S(t)]}{dt}$$

อินทิเกรตฟังก์ชันความเสี่ยงจะเท่ากับฟังก์ชันความเสี่ยงสะสม โดย

$$H(t)=\int_0^t h(t)dt=-\ln[S(t)]$$

และสามารถหาฟังก์ชันการอยู่รอดได้จาก

$$S(t) = \exp[-H(t)]$$

ลักษณะของข้อมูลของระยะเวลาถึงเหตุการณ์ที่สนใจ

การวิเคราะห์ข้อมูลของระยะเวลาถึงเหตุการณ์ที่สนใจส่วนใหญ่มักจะพบปัญหาเกี่ยวกับข้อมูลที่มีลักษณะเป็นข้อมูลไม่สมบูรณ์ คือการมีข้อมูลที่ไม่ทราบค่าสังเกตที่แท้จริงว่าเกิดเหตุการณ์ใดขึ้นก่อนสิ้นสุดการติดตาม (Klein and Moeschberger, 2003) โดยสาเหตุที่ค่าสังเกตมีการขาดหายไป อาจเกิดจากการหยุดเข้าร่วมการศึกษากลางคันของผู้ป่วยการที่ผู้ป่วยไม่มารับการรักษาการที่ผู้ป่วยเสียชีวิตจากโรคอื่นที่ไม่ใช่โรคที่ศึกษาหรือการสิ้นสุดการศึกษาก่อนที่กำหนดไว้

ข้อมูลไม่สมบูรณ์แบ่งออกเป็น 2ประเภท ได้แก่

1) **ข้อมูลที่มีการเซนเซอร์** (Censored data) แบ่งเป็น 3 แบบ ดังนี้

- ข้อมูลที่มีการเซนเซอร์ทางด้านขวา(Right censoring) เกิดเมื่อสิ้นสุดการศึกษาแล้วยังไม่เกิดเหตุการณ์ที่สนใจในช่วงเวลาที่ศึกษา หรือบุคคลขอออกจากการศึกษาก่อนจะเกิดเหตุการณ์ที่สนใจ
- ข้อมูลที่มีการเซนเซอร์ทางด้านซ้าย (Left censoring) เกิดเมื่อเหตุการณ์ที่สนใจเกิดขึ้นก่อนเริ่มการศึกษา
- ข้อมูลที่ไม่สมบูรณ์ในช่วง (Interval censoring) เกิดเมื่อเหตุการณ์ที่สนใจเกิดขึ้นในเวลาที่ศึกษา แต่ไม่ทราบเวลาที่แน่นอนว่าเกิด ณ จุดใด

2) **ข้อมูลที่มีการตัดทิ้ง**(Truncation)แบ่งเป็น 2 แบบ ดังนี้

- ข้อมูลที่มีการตัดทิ้งทางด้านซ้าย (Left Truncation) คือการพิจารณาข้อมูล โดยข้อมูลก่อนช่วงเวลาที่สนใจจะไม่ถูกนำมาใช้ในการวิเคราะห์
- ข้อมูลที่มีการตัดเป็นทิ้งทางด้านขวา (Right Truncation)คือการพิจารณาข้อมูล โดยข้อมูลที่เกิดเหตุการณ์ที่สนใจ ณ ช่วงเวลานั้นจะถูกนำมาใช้ในการวิเคราะห์
- ข้อมูลที่มีการตัดทิ้งทางด้านซ้ายและขวา (Left and Right Truncation)คือการพิจารณาข้อมูลโดยข้อมูลจากก่อนช่วงเวลาที่สนใจจะไม่ถูกนำมาใช้ในการวิเคราะห์แต่จะวิเคราะห์ข้อมูลที่เกิดเหตุการณ์ที่สนใจ ณ ช่วงเวลานั้น

โดยทั่วไปการวิเคราะห์ข้อมูลการอยู่รอดแบ่งเป็น 2 ประเภท คือ การวิเคราะห์ข้อมูลด้วยวิธีตามรุ่น (Cohort Analysis) และการวิเคราะห์ข้อมูลด้วยวิธีตามช่วงเวลา (Period Analysis)ซึ่งการวิเคราะห์ข้อมูลด้วยวิธีตามรุ่นสามารถประมาณข้อมูลการอยู่รอดย้อนหลังไปในหลายๆ ปีได้

ข้อจำกัดที่สำคัญของวิธีนี้คือ ไม่สามารถนำข้อมูลบางส่วนมาใช้ในการคำนวณได้ เพราะระยะเวลาที่เริ่มรับการรักษาถึงระยะเวลาที่ทำการศึกษายังต่ำกว่าช่วงเวลาที่ต้องการคำนวณอัตราการอยู่รอด จึงอาจส่งผลต่อความแม่นยำในการวิเคราะห์ดังนั้นการศึกษาครั้งนี้จะใช้การวิเคราะห์ข้อมูลของระยะเวลาถึงเหตุการณ์ที่สนใจตามช่วงเวลาเพราะการวิเคราะห์ด้วยวิธีนี้จะนำเฉพาะข้อมูลที่มีความเป็นปัจจุบันในช่วงเวลาที่สนใจมาใช้ในการคำนวณ เพื่อประโยชน์ในการประมาณค่าข้อมูลระยะเวลาดังกล่าวในระยะยาว

2.1.4.การวิเคราะห์ข้อมูลของระยะเวลาถึงเหตุการณ์ที่สนใจแบบไม่ใช้เบส์เซียน

ในการวิเคราะห์ข้อมูลของระยะเวลาถึงเหตุการณ์ที่สนใจแบบไม่ใช้เบส์เซียน มีวิธีการวิเคราะห์ 2 แบบ คือ การประมาณกรณีตัวแปรเดียว (Univariate Estimation) โดยใช้ตัวประมาณแคปลานและไมย์เออร์ และการประมาณกรณีมีตัวแปรร่วมอื่น ที่มีผลต่อระยะเวลาการอยู่รอด (Multivariate Estimation) โดยใช้ตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อก และตัวแบบการเร่งเวลาการเสียหาย โดยมีรายละเอียดดังนี้

ตัวประมาณแคปลานและไมย์เออร์ (Kaplan-Meier estimator)

ตัวประมาณแคปลานและไมย์เออร์ เป็นตัวประมาณมาตรฐานสำหรับฟังก์ชันการอยู่รอดที่เสนอโดยแคปลานและไมย์เออร์ ในปี ค.ศ. 1958 จึงถูกเรียกว่าตัวประมาณแคปลานและไมย์เออร์ ซึ่งกำหนดโดย

$$\hat{S}(t)=\begin{cases}1 & ,t<t_1\\ \prod_{t_i\le t}[1-\frac{d_i}{y_i}] & ,t\ge t_1\end{cases}$$

เมื่อ t_1 แทนเวลาที่เหตุการณ์แรกเกิดขึ้น และ $t_{\max}$ แทนเวลาสุดท้ายที่สังเกต

จากการนิยามตัวประมาณแคปลานและไมย์เออร์ จะพบว่าฟังก์ชันดังกล่าวเป็นฟังก์ชันขั้นบันได ซึ่งมีการเปลี่ยนค่า (jump) เกิดขึ้นที่ t_i ขนาดของค่าที่เปลี่ยนจะขึ้นอยู่กับจำนวนเหตุการณ์ที่เกิดขึ้น คือ d_i และหาค่าประมาณความแปรปรวนได้ด้วย 2 สูตร ได้แก่

สูตรของกรีนวูด (Greenwood's formula)

$$\hat{V}[\hat{S}(t)]=\hat{S}(t)^2\sum_{t_i\le t}\frac{d_i}{y_i(y_i-d_i)}$$

และต่อมาเอเลนและโจฮานเซนได้เสนอว่า

$$\hat{V}[\hat{S}(t)]=\hat{S}(t)^2\sum_{t_i\le t}\frac{d_i}{y_i^{\,2}}$$

โดยตัวประมาณค่าของความแปรปรวนนี้จะให้ค่าที่ต่ำกว่าความเป็นจริงเมื่อจำนวนหน่วยตัวอย่างมีขนาดเล็กถึงขนาดกลาง แต่โดยเฉลี่ยแล้วค่าประมาณความแปรปรวนที่ใช้สูตรของกรีนวูดจะให้ค่าที่ใกล้เคียงกับค่าจริง และให้ค่าน้อย

ตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อก (Cox Proportional Hazard Model)

ตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อก เป็นตัวแบบทางสถิติเพื่อใช้ในการศึกษาความสัมพันธ์ระหว่างตัวแปรตามที่สนใจ กับตัวแปรอิสระต่างๆ ตั้งแต่หนึ่งตัวขึ้นไป โดยเป็นข้อมูลที่มีช่วงเวลาเข้ามาเกี่ยวข้อง ซึ่งถูกพัฒนาขึ้นมาโดยค็อก (Cox, 1972) โดยที่ตัวแปรตามเป็นฟังก์ชันความเสี่ยงต่อการเกิดเหตุการณ์ที่สนใจ ณ เวลา t และตัวแปรอิสระสามารถเป็นได้ทั้งตัวแปรเชิงปริมาณและตัวแปรเชิงกลุ่มที่ไม่ขึ้นกับเวลา โดยมีข้อตกลงเบื้องต้นคืออิทธิพลของตัวแปรอิสระแต่ละตัวในสมการจะเท่ากันตลอดระยะเวลาของการศึกษา นั่นคืออัตส่วนของอิทธิพลของตัวแปรอิสระระหว่างกลุ่มหรือทุกหน่วยการศึกษาจะมีค่าคงที่ และตัวแปรอิสระแต่ละตัวไม่ขึ้นกับเวลา

กำหนดให้ $Z_j(t) = Z_j = (z_{j1},...,z_{jp})'$ เป็นตัวแปรร่วมที่เป็นอิสระไม่ขึ้นกับเวลาและ p แทนจำนวนตัวแปรร่วมที่สนใจ โดยสามารถเขียนตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อก ได้ดังนี้

$$h(t \mid z) = h_0(t)c(\beta' z)$$

เมื่อ $h_0(t)$ คือความเสี่ยงฐาน (Baseline hazard) ซึ่งหมายถึงความเสี่ยงของแต่ละบุคคลเมื่อตัวแปรอิสระมีค่าเท่ากับศูนย์ $c(\beta' z)$ คือฟังก์ชันที่ทราบและสร้างได้ด้วยวิธีพาราเมตริกซ์ โดยฟังก์ชันที่นิยมใช้คือ

$$c(\beta' z) = e^{\beta' z} = \exp\left(\sum_{k=1}^{p} \beta_k z_k\right)$$

ทำการประมาณค่า β ด้วยวิธีของนิวตันราฟสัน (Newton Raphson algorithm) โดยกำหนดค่าเริ่มต้น $b_0 = 0$ และประมาณค่าถึงเวลาที่ m โดยที่ $b_m = b_{m-1} + \dfrac{U(b_{m-1})}{I(b_{m-1})}$ จากนั้นทำกระบวนการนี้ซ้ำจนเกิดการลู่เข้าเมื่อ $\dfrac{LL(b_m) - LL(b_{m-1})}{|LL(b_{m-1})|}$ มีค่าน้อยกว่า 0.0001

เมื่อ

$b = (b_1, b_2, ..., b_p)^t$ เป็นค่าประมาณน่าจะเป็นสูงสุด (maximum likelihood estimate) ของ β

$I(\beta)$ เป็นเมทริกซ์ขนาด $p \times p$ ของ β

$U(\beta) = (U_1(\beta),...,U_p(\beta))^t$

$$LL(\beta) = \beta d_1 - \sum_{i=1}^{n} \ln\left[Y_{0i} + Y_{1i}e^{\beta}\right]$$

$$U(\beta) = d_1 - \sum_{i=1}^{n} \frac{Y_{1i}e^{\beta}}{Y_{0i} + Y_{1i}e^{\beta}}$$

และ $I(\beta)=\sum_{i=1}^{n}\left[\frac{Y_{1i}e^{\beta}}{\left(Y_{0i}+Y_{1i}e^{\beta}\right)}-\frac{{Y^{2}}_{1i}e^{2\beta}}{\left(Y_{0i}+Y_{1i}e^{\beta}\right)^{2}}\right]$

ตัวแบบการเร่งเวลาการเสียหาย (Accelerated Failure Time Model)

กำหนดให้ X เป็นระยะเวลาตั้งแต่ได้รับการวินิจฉัยว่าเป็นโรคถึงเวลาที่ชำระสินไหมทดแทน และ Z เป็นเวกเตอร์ของตัวแปรร่วมที่ไม่ขึ้นกับเวลา ดังนั้นตัวแบบนี้จะระบุว่าฟังก์ชันของระยะเวลาถึงเหตุการณ์ที่สนใจ หน่วยตัวอย่างที่มีตัวแปรร่วม Z ณ เวลา X จะมีค่าเท่ากับฟังก์ชันของระยะเวลาถึงเหตุการณ์ที่สนใจ ของหน่วยตัวอย่างที่เป็นฐาน (baseline) ณ เวลา $x\exp(\theta' z)$ โดยที่ $\theta'=(\theta_1,...,\theta_p)$ เป็นเวกเตอร์ของสัมประสิทธิ์การถดถอย จึงเขียนตัวแบบการเร่งเวลาการเสียหายได้ในรูปความสัมพันธ์ดังนี้

$$S(x\mid z)=S_0[\exp(\theta' z)x] \text{ สำหรับทุกค่าของ } x$$

เมื่อ $\exp(\theta' z)$ คือปัจจัยการเร่ง (acceleration factor) ซึ่งเป็นตัวที่บ่งบอกว่าค่าตัวแปรร่วมที่เปลี่ยนไปทำให้ขนาดเวลา (time scale) เปลี่ยนไปจากขนาดเวลาที่เป็นฐาน (baseline time scale) มากน้อยเช่นไร จากตัวแบบดังกล่าวสามารถหาฟังก์ชันความเสี่ยงอันตราย (hazard function) ได้ดังนี้

$$h(x\mid z)=\exp(\theta' z)h_0[\exp(\theta' z)x] \text{ สำหรับทุกค่าของ } x$$

นอกจากนั้นสามารถใช้ตัวแบบเชิงเส้นสำหรับลอการิทึมของเวลาการอยู่รอดได้ด้วยสมการ

$$Y=\ln X=\mu+\gamma^{t}z+\sigma W$$

โดยที่ $\gamma^{t}=(\gamma_1,...,\gamma_p)$ เป็นสัมประสิทธิ์การถดถอย และ W เป็นการแจกแจงของความคลาดเคลื่อน (Error distribution) จากสมการข้างต้นจะได้

$$X\exp[-\gamma^{t}z]=\exp[\mu+\sigma W] \qquad \text{...(2.1)}$$

ทำให้ $\exp[\mu+\sigma W]$ เป็นตัวแปรสุ่มใน $S_0(x)$ นั่นคือ

$$\exp[\mu+\sigma W]=X\exp[\theta^{t}z] \qquad \text{...(2.2)}$$

ดังนั้นจากสมการที่ (2.1) และ (2.2) จะได้ $\theta=-\gamma$

2.1.5. การวิเคราะห์แบบเบส์เซียน (Bayesian analysis)

การศึกษาครั้งนี้ใช้ตัวแบบเบอร์ และตัวแบบล็อกนอร์มอล มาทำการวิเคราะห์แบบเบส์เซียน โดยคัดเลือกตัวแปรด้วยวิธีการคัดเลือกตัวแปรของกิบส์ โดยกำหนดพารามิเตอร์ก่อน (prior parameter)เป็น Zellner' s g – prior มีรายละเอียดดังนี้

ตัวแบบเบอร์

ให้ D_i แทนระยะเวลาที่ล่าช้าของการเรียกร้องค่าสินไหมทดแทนของกรมธรรม์ที่ i และตัวแบบเบอร์มีฟังก์ชันความหนาแน่นน่าจะเป็น คือ

$$f(d_i;\alpha,\tau,s)=\frac{\alpha\tau(d_i/s)^{\tau}}{d_i[1+(d_i/s)^{\tau}]^{\alpha+1}},\ i=1,...,n$$

เมื่อ $d_i,\alpha,\tau,s>0$ โดยที่ α และ τ เป็นพารามิเตอร์บ่งรูปร่าง (shape parameter) s เป็นพารามิเตอร์บ่งขนาด (scale parameter) และ n เป็นจำนวนกรมธรรม์ที่เกิดการล่าช้าในการเรียกร้องสินไหมทดแทน และมีฟังก์ชันการแจกแจงเป็น

$$F_D(d)=1-\left(\frac{s^{\tau}}{s^{\tau}+d^{\tau}}\right)^{\alpha}$$

และมี

$$E(D^k)=s^k\Gamma(\alpha-\frac{k}{\tau})\Gamma(1+\frac{k}{\tau})/\Gamma(\alpha)\ \text{เมื่อ}\ \alpha\tau>k$$

ในตัวแบบเชิงเส้นโดยนัยทั่วไป (generalised linear type model) จะใช้การเชื่อมโยงแบบล็อก (logarithmic link) ทำให้ได้ตัวแบบเบอร์เชิงเส้นโดยทั่วไป (Burr generalised linear type model) ที่มี D_i มีการแจกแจงแบบเบอร์ด้วยพารามิเตอร์ α τ และ s_i โดยที่

$$E(D_i)=\exp(\eta_i)=s_i\frac{\Gamma(\alpha-1/\tau)\Gamma(1+1/\tau)}{\Gamma(\alpha)} \qquad (2.3)$$

เมื่อ $\eta=X\beta$ โดยที่ η เป็นเวกเตอร์เมทริกซ์มิติ $n\times1$, X เป็นเมทริกซ์มิติ $n\times(p+1)$ และ β เป็นสัมประสิทธิ์ของเวกเตอร์ในรูปเมทริกซ์มิติ $(p+1)\times1$ โดย p เป็นจำนวนตัวแปรร่วม จากสมการ (2.3) จะได้

$$s_i=\frac{\Gamma(\alpha)}{\Gamma(\alpha-\frac{1}{\tau})\Gamma(1+\frac{1}{\tau})}\exp(\eta_i)$$

โดยที่

$$\eta_i=\beta_0+\sum_{j=1}^{p}\beta_j z_{ij}$$

เมื่อ $z_{ij}=(x_{ij}-\bar{x}_j)/sd(x_j)$ เป็นค่าสังเกตมาตรฐาน (standardised observation)

x_{ij} เป็นค่าสังเกตที่แท้จริง (original observation)

β_0 เป็น standardised intercept term และ

β_j สัมประสิทธิ์ โดยที่ $\beta_j\neq0$

Chulalongkorn University

กำหนดการแจกแจง non informative prior ของพารามิเตอร์ในตัวแบบ ดังนี้

α มีการแจกแจงแบบแกมมาด้วยพารามิเตอร์ 0.001 และ 0.001

τ มีการแจกแจงแบบแกมมาด้วยพารามิเตอร์ 0.001 และ 0.001

β_j มีการแจกแจงแบบปกติด้วยพารามิเตอร์ 0 และ 0.001

จะได้ภาวะน่าจะเป็นของตัวแบบ คือ

$$f(D\mid\alpha,\tau,\beta)=\alpha^n\tau^n\prod_i^n((G(\alpha,\tau))^{-1}\exp(\eta_i))^{\alpha\tau}d_i^{\tau-1}\times[((G(\alpha,\tau))^{-1}\exp(\eta_i))^{\tau}+d_i^{\tau}]^{-(\alpha+1)}$$

เมื่อ $G(\alpha,\tau)=\dfrac{\Gamma(\alpha-\frac{1}{\tau})\Gamma(1-\frac{1}{\tau})}{\Gamma(\alpha)}$

โดยที่ $\Gamma(t)=\int_0^\infty x^{t-1}e^{-x}dx$

ตัวแบบล็อกนอร์มอล

ให้ D_i แทนระยะเวลาที่ล่าช้าของการเรียกร้องค่าสินไหมทดแทนของกรมธรรม์ที่ i D_i มีการแจกแจงแบบล็อกนอร์มอลด้วยพารามิเตอร์ μ_i และ σ^2 โดยที่

$$\mu_i=\eta_i=\beta_0+\sum_{j=1}^{p}\beta_j z_{ij}$$

เมื่อ $i=1,...,n$

การแจกแจง non informative prior ของ β ในตัวแบบ มีการแจกแจกแบบทั่วไปด้วยพารามิเตอร์ 0 และ 0.001 และ σ^2 มีการแจกแจงแบบอินเวอร์สแกมมาด้วยพารามิเตอร์ 0.001 และ 0.001

จะได้ภาวะน่าจะเป็นของตัวแบบ คือ

$$f(D\mid\sigma^2,\beta)=\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n\prod_i^n\frac{1}{d_i}\exp\left(-\frac{\sum(\log(d_i)-\eta_i)^2}{2\sigma^2}\right)$$

การวิเคราะห์แบบเบส์เซียน

ให้ D เป็นตัวแปรสุ่มประเภทไม่ต่อเนื่อง โดยมี $D_1,D_2,...,D_n$ เป็นตัวอย่างสุ่ม โดยที่ θ เป็นพารามิเตอร์ในฟังก์ชันการแจกแจงของ D และ θ เป็นค่าที่เป็นไปได้ของพารามิเตอร์สุ่ม Θ ดังนั้นการแจกแจงความน่าจะเป็นของ D เป็นการแจกแจงแบบมีเงื่อนไข

ให้ Θ มีฟังก์ชันความหนาแน่น $f(\theta)$ และให้ตัวแบบที่เป็นไปได้ทั้งหมดแทนด้วย m_t เมื่อ $t=1,...,k$ และ $f(D\mid m_t)$ เป็นภาวะน่าจะเป็นส่วนริม(marginal likelihood)ภายใต้ m_t

โดยทฤษฎีบทเบส์(Bayes theorem) จะได้ความน่าจะเป็นหลัง(posterior probability)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

$$f(m_t \mid D) = \frac{f(D \mid m_t) f(m_t)}{\int f(D \mid m_t) f(m_t) dm_t}$$

เมื่อ $f(m_t)$ คือ ฟังก์ชันความน่าจะเป็นของตัวแบบก่อน (prior model probability)

จากนั้นเปรียบเทียบ 2 ตัวแบบ เช่น m_i กับ m_j เพื่อหาตัวแบบที่เหมาะสมที่สุดด้วย posterior odd ratio

$$\frac{f(m_i \mid D)}{f(m_j \mid D)} = \frac{f(D \mid m_i) f(m_i)}{f(D \mid m_j) f(m_j)}$$

เมื่อ $i, j \in [1,...,k]$ ซึ่ง $B_{ij} = \frac{f(D \mid m_i)}{f(D \mid m_j)}$ เป็น Bayes factor และ $\frac{f(m_i)}{f(m_j)}$ เป็น prior odds ratio

โดย (Kassและ Raftery, 1995) นำการตรวจสอบสมมติฐานแบบเบส์เซียนไปใช้โดยมีเกณฑ์ ดังนี้

$\log B_{ij}$	<0	$0-1$	$1-3$	>3
	สนับสนุน m_i	ไม่มีข้อสรุป	สนับสนุน m_j	สนับสนุน m_j อย่างมาก

โดยทั่วไปภาวะน่าจะเป็นส่วนริม(marginal likelihood) แทนด้วย

$$f(D \mid m_t) = \int f(D \mid \theta_t, m_t) f(\theta_t \mid m_t) d\theta_t$$

เมื่อ θ_t เป็นพารามิเตอร์เวกเตอร์ใน m_t

$f(D \mid \theta_t, m_t)$ เป็นฟังก์ชันภาวะน่าจะเป็น (likelihood function) ภายใต้ m_t

$f(\theta_t \mid m_t)$ เป็นฟังก์ชันความหนาแน่นก่อน (prior density function) ของพารามิเตอร์เวกเตอร์ θ_t

มีวิธีการหลายวิธีการที่ใช้ในการประมาณค่า ในที่นี่จะใช้วิธีการคัดเลือกตัวแปรของกิบส์ โดยพิจารณาพารามิเตอร์เดิม (prior parameter) สำหรับการแจกแจงของ β เป็นแบบ Zellner 's g – prior ซึ่งมีรายละเอียดดังต่อไปนี้

- **การคัดเลือกตัวแปรของกิบส์** (Gibbs variable selection :GVS)

ให้ γ เป็น $p \times 1$ indicator vector โดยที่ $\gamma \in \{0,1\}^p$ และ p แทนจำนวนตัวแปรร่วมทั้งหมดในตัวแบบ จากนั้นเพิ่ม γ ลงไปในการพยากรณ์เชิงเส้นสำหรับตัวแบบ จะได้

$$\eta_i = \beta_0 + \sum_{j=1}^{p} \gamma_j \beta_j z_{ij}$$

โดยที่ γ_j มีการแจกแจงแบบเบอร์นูลลีด้วยพารามิเตอร์ 0.5 เพื่อให้พารามิเตอร์แต่ละตัวมีค่าน้ำหนักที่เท่ากัน โดยพารามิเตอร์ก่อน (prior parameters) สำหรับการแจกแจงของ β เป็นแบบ Zellner' s g – prior

Zellner' s g – prior (Zellner, 1986) จะใช้สำหรับตัวแบบปกติซึ่ง Normal – Inverse Gamma conjugate prior หาได้จากงานวิจัยของ Ntzoufras ในปี ค.ศ. 2009 และงานวิจัยของ ออคคอค ในปี ค.ศ. 2010 เพื่อที่จะใช้กับการแจกแจงของข้อมูลที่มีจะใช้การประมาณค่าเมทริกซ์ความแปรปรวน และความแปรปรวนร่วม (variance – covariance matrix) โดยที่ β มีการแจกแจงแบบ $N(\mu, S)$ ในที่นี่ β เป็นเวกเตอร์สัมประสิทธิ์ (coefficient vector $\mu = \mu_0, ..., \mu_p$ เป็นเวกเตอร์ค่าเฉลี่ยก่อน (prior mean vector) และ S เป็นเมทริกซ์ความแปรปรวนร่วมก่อน (prior covariance matrix) ซึ่ง

$$\mu_j = (1-\gamma_j)\bar{\mu}_{\beta_j}, \ j = 0, ..., p$$

$$S_{j,k} = (\gamma_j \gamma_k) n \sum_{jk} + (1-\gamma_j \gamma_k) I\{j=k\} \bar{\sigma}^2_{\beta_j}, \ j,k = 0, ..., p$$

เมื่อ $\sum = (-D^2 l(\hat{\beta}))^{-1}$ ถูกประมาณด้วยเมทริกซ์ความแปรปรวน และความแปรปรวนร่วม (variance – covariance matrix) ซึ่ง $D^2 l(\hat{\beta})$ คือ Hessian matrix ของ loglikelihood ที่ประมาณด้วยภาวะน่าจะเป็น-สูงสุด(maximum likelihood) (Kassและ Wasserman, 1995 ;Ntzoufrasและคณะ, 2003)

เมื่อ $\gamma_0 = 1$ ที่ $\bar{\mu}_{\beta_j}$ และ $\bar{\sigma}^2_{\beta_j}$ เป็นตัวประมาณหลัง(posterior estimate)ของค่าเฉลี่ย และความแปรปรวนของ β_j

2.1.6. วิธีมาร์คอฟ เชน มอนติคาร์โล(Markov Chain Monte Carlo Method :MCMC)

วิธีมาร์คอฟ เชน มอนติคาร์โล(Markov Chain Monte Carlo Method :MCMC) เป็นวิธีที่ใช้มอนติ คาร์โลอินทิเกรชั่น(Monte Carlo Integration) ด้วยมาร์คอฟ เชน (Markov Chains)

เนื่องจากการประมาณค่าพารามิเตอร์หลายวิธีมีความยากมากที่จะหาตัวประมาณเชิงวิเคราะห์ได้ ดังนั้น มอนติ คาร์โลอินทิเกรชั่น จะเป็นวิธีแก้ปัญหาการคำนวณค่าคาดหวังจากอินทิกรัลที่ซับซ้อน และยุ่งยากดังกล่าว โดยการหาค่าเฉลี่ยตัวอย่างจากฟังก์ชันความหนาแน่นที่ต้องการแทน โดยกำหนดให้ X เป็นตัวแปรสุ่มที่มีการแจกแจง π และ h เป็นฟังก์ชันที่ต้องการประเมิน

$$E[h(x)] = \int h(x)\pi(x)dx$$

และมีแนวคิดมาจากการสุ่มตัวอย่าง ซึ่งทำการจำลองตัวอย่างจากฟังก์ชันความหนาแน่นที่ต้องการ $X_1, X_2, ..., X_n \sim \pi$ เพื่อนำมาหาค่าเฉลี่ยตัวอย่าง

$$\bar{h} = \frac{1}{n}\sum_{i=1}^{n} h(x_i)$$

โดยตัวอย่างสุ่มของมอนติคาร์โลอินทิเกรชั่น จะต้องเป็นตัวอย่างสุ่มที่เป็นอิสระกัน แต่เนื่องจากการแจกแจงของฟังก์ชันความหนาแน่นที่ต้องการ (π) ไม่ได้เป็นการแจกแจงมาตรฐาน ดังนั้นจึงต้องอาศัยหลักการของ มาร์คอฟ เชน ในการจำลองตัวอย่างสุ่ม จากฟังก์ชันความหนาแน่นที่ต้องการ $\pi(.)$ ซึ่งเป็นการแจกแจงเป้าหมาย (Target distribution) โดยหลักการของมาร์คอฟ เชน มีจุดประสงค์เพื่อจำลองลำดับตัวอย่างสุ่ม (X_t) เมื่อ $\mathrm{t}=1,2,3...$ จากความหนาแน่นแบบเคอร์เนล (Kernel Density) $P(x_{1+t},x_t)$ เมื่อ $X_t=x_t$ แล้ว $X_{t+1}\sim P(x_{t+1},x_t)$

กำหนดให้ x_t เป็นค่าของตัวแปรสุ่ม X ณ เวลา t และกำหนดให้ปริภูมิสถานะ (State Space) เป็นเซตของค่า x_t ที่สามารถเป็นได้ และ $x_t\in\Omega$ โดยที่กระบวนการมาร์คอฟจะเกิดขึ้นได้ก็ต่อเมื่อความน่าจะเป็นในการเคลื่อนที่ระหว่างค่าที่เป็นไปได้ในปริภูมิสถานะ ขึ้นอยู่กับค่าในสถานะปัจจุบันเท่านั้น

$$P(x_{t+1}\in A\,|\,x_0,x_1,...,x_t)=P(x_{t+1}\in A\,|\,x_t)$$

สำหรับเซต $A\subset\Omega$

โดยที่ $P(x_{t+1}\in A\,|\,x_t)$ อาจเขียนเป็น $P(x_{t+1}:x_t)$ โดยที่ $\int P(x_{t+1}:x_t)dx_t=1$ ดังนั้นจะพบว่าความน่าจะเป็นของค่าที่เป็นไปได้ของ X_{t+1} เมื่อทราบ $(X_t,X_{t-1},...,X_0)$ จะมีค่าเท่ากับความน่าจะเป็นค่าที่เป็นไปได้ของ X_{t+1} เมื่อทราบ X_t

เมื่อกำหนดให้กระบวนการข้างต้นมีค่าความน่าจะเป็นคงที่ทุกช่วงเวลา แล้ว $P(*:*)$ จะไม่ขึ้นกับ t ใดๆ ดังนั้นการแจกแจงของ X_t เมื่อกำหนด X_0 จะมีค่าเท่ากับ $P^{(t)}(x_t:x_0)$ โดยที่ X_t จากสมการของแชปแมน-โคโมโกรอฟ จะได้ว่า

$$P^{(t+1)}(x_{t+1}:x_0)=\int P^{(t)}(x_t:x_0)P(x_{t+1}:x_t)dx_t$$

และกระบวนการมาร์คอฟ เชน จะเป็นการแจกแจงที่เสถียร (Stationary distribution : $\pi(\cdot)$)คือ

$$\pi(x_{t+1})=\int\pi(x_t)P(x_{t+1}:x_t)dx_t$$

2.2. งานวิจัยที่เกี่ยวข้อง

งานวิจัยในต่างประเทศ

Beirlantและคณะ(1998) ได้ขยายการแจกแจงแบบเบอร์ (Burr distribution) ไปสู่ตัวแบบการถดถอยโดยพิจารณาพารามิเตอร์บ่งรูปร่าง(shape parameter) หรือพารามิเตอร์บ่งขนาด (scale parameter)ของตัวแบบ และทำการประมาณพารามิเตอร์ โดยใช้ภาวะน่าจะเป็นสูงสุด (maximum likelihood) จากนั้น Beirlant และGuillou (2001) ได้พิจารณากระบวนการที่ข้อมูลมีค่าผิดปกติไปมาก(extreme value method) ของการแจกแจงพาเรโตภายใต้การเซนเซอร์ และ

ประมาณค่าดัชนีค่าผิดปกติสุดขีด(extreme valueindex) ด้วยภาวะน่าจะเป็นสูงสุด(maximum likelihood) ของการแจกแจงเบอร์ซึ่งถูกนิยามในรูปของการใช้พารามิเตอร์บ่งรูปร่าง(shape parameter) หรือการใช้พารามิเตอร์บ่งขนาด(scale parameter) ต่อมา Beirlant และ Goegebeur (2003) ได้ทำการอนุมานตัวแบบถดถอยบนดัชนีนี้กับตัวแปรร่วม และประมาณค่าด้วยภาวะน่าจะเป็นสูงสุด(maximum likelihood)

Xin – Yuan Song และ Sik – Yum Lee (2002) ได้คัดเลือกตัวแบบด้วยวิธีวิเคราะห์แบบเบส์เซียน โดยใช้เบส์แฟคเตอร์ (Bayes factor) โดยมีแนวคิดหลักเพื่อจะสร้างส่วนที่เชื่อมโยงกับตัวแบบแล้วประมาณค่าเบส์แฟคเตอร์ และสังเกตการจำลองแบบจากการแจกแจงหลัง (posterior distribution)ของพารามิเตอร์

Ozkok(2011) ได้สร้างตัวแบบการแจกแจงของเวลาที่ล่าช้าระหว่างเวลาที่ทำการวินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนของการประกันภัยโรคร้ายแรง จากข้อมูลในประเทศอังกฤษ ปีคริสต์ศักราช1999-2005 ข้อมูลดังกล่าวพบว่าในจำนวน 19,127 ของการเรียกร้องค่าสินไหมทดแทนมีผู้เอาประกันภัยจำนวน 15,860 คนที่มีการบันทึกทั้งวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทน จึงทำการจำแนกกลุ่มการเรียกร้องค่าสินไหมทดแทนในระหว่างวันที่กรมธรรม์มีผลบังคับใช้ จนถึงวันที่ทำการวินิจฉัยว่าเป็นโรค (the insured event) โดยประยุกต์ใช้ตัวแบบเบส์เซียนวิเคราะห์หาการแจกแจงความน่าจะเป็น และแก้ไขข้อมูลที่ขาดหายไปโดยการเพิ่มพารามิเตอร์ที่ถูกประมาณในการวิเคราะห์ด้วยมาร์คอฟ เชน มอนติคาร์โล จากข้อมูลเวลาที่ล่าช้าดังกล่าวพบว่าการแจกแจงแบบเบอร์ในรูปแบบเชิงเส้นทั่วไปมีความเหมาะสมที่สุด ซึ่งการแจกแจงนี้มีการนำไปใช้อย่างมากในทางประกันภัย

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่3
วิธีดำเนินการวิจัย

3.1 ข้อมูลที่ใช้ในการศึกษา

การวิจัยครั้งนี้ต้องการวิเคราะห์ระยะเวลาระหว่างเวลาที่ทำการวินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง และศึกษาความสัมพันธ์ระหว่างระยะเวลาถึงเหตุการณ์ที่สนใจกับตัวแปรร่วม โดยใช้การวิเคราะห์การอยู่รอด 2 แบบ คือ การวิเคราะห์การอยู่รอดแบบไม่ใช้เบส์เซียน ซึ่งใช้ตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อก และการวิเคราะห์การอยู่รอดแบบเบส์เซียน ซึ่งจะใช้ตัวแบบเบอร์ และตัวแบบล็อกนอร์มอล โดยข้อมูลที่ใช้เป็นข้อมูลระยะเวลาระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง ระหว่างวันที่ 18 กรกฎาคม พ.ศ. 2546 ถึงวันที่ 7 ธันวาคม 2556 ของบริษัทประกันชีวิต 2 แห่ง มีผู้เอาประกันภัยทั้งหมดจำนวน 840 ราย

ลักษณะของตัวแปรร่วม แบ่งเป็น 2 แบบคือ

- ตัวแปรร่วมที่เป็นข้อมูลเชิงปริมาณ ได้แก่
    1. อายุของผู้เอาประกันภัย ณ วันที่เรียกร้องสินไหมทดแทน มีหน่วยเป็นปี
    2. จำนวนเงินเอาประกันภัย
- ตัวแปรร่วมที่เป็นข้อมูลเชิงคุณภาพ ได้แก่
    1. เพศของผู้เอาประกันภัย โดยกำหนดให้เพศชายแทนด้วย 0 เพศหญิงแทนด้วย 1
    2. รูปแบบผลประโยชน์ โดยแบ่งเป็นผลประโยชน์ที่ครอบคลุมการเสียชีวิต (full accelerated) แทนด้วย 0 และ ผลประโยชน์ที่ไม่ครอบคลุมการเสียชีวิต (stand-alone) แทนด้วย 1
    3. สาเหตุของการเรียกร้องค่าสินไหมทดแทน ในงานวิจัยนี้ คือโรคที่เกิดขึ้นกับผู้เอาประกันภัย โดยแบ่งเป็นกลุ่มโรคร้ายแรง 8 กลุ่ม ดังนี้
    - กลุ่มโรคมะเร็ง
    - กลุ่มโรคเกี่ยวกับตับ หรือตับอ่อน
    - กลุ่มโรคเลือด อวัยวะสร้างเลือด และความผิดปกติเกี่ยวกับกลไกของภูมิคุ้มกัน
    - กลุ่มโรคเกี่ยวกับระบบประสาท
    - กลุ่มโรคปอดอักเสบ และโรคอื่นๆของปอด
    - กลุ่มโรคความดันเลือดสูง และโรคเส้นเลือดในสมอง

- กลุ่มโรคหัวใจ
- กลุ่มโรคอื่นๆ

ตัวแปรสาเหตุของการเรียกร้องค่าสินไหมทดแทนนี้ แบ่งเป็น 8 กลุ่มจึงสร้าง dummy variable โดยให้ตัวแปรกลุ่มโรคมะเร็งเป็น baseline ได้ 7 ตัว ดังนี้

ตาราง 3.1 dummy variable ของตัวแปรสาเหตุของการเรียกร้องค่าสินไหมทดแทน

	1	2	3	4	5	6	7
กลุ่มโรคมะเร็ง	0	0	0	0	0	0	0
กลุ่มโรคเกี่ยวกับตับ	1	0	0	0	0	0	0
กลุ่มโรคเลือด	0	1	0	0	0	0	0
กลุ่มโรคระบบประสาท	0	0	1	0	0	0	0
กลุ่มโรคปอด	0	0	0	1	0	0	0
กลุ่มโรคความดัน	0	0	0	0	1	0	0
กลุ่มโรคหัวใจ	0	0	0	0	0	1	0
กลุ่มโรคอื่นๆ	0	0	0	0	0	0	1

4. บริษัทประกันภัยที่ทำการศึกษา ซึ่งในงานวิจัยนี้ทำการศึกษา 2 บริษัท โดยบริษัทที่ 1 แทนด้วย 0 และบริษัทที่ 2 แทนด้วย 1

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3.2 ขั้นตอนการดำเนินงาน

1. การวิเคราะห์การอยู่รอดแบบไม่ใช้เบส์เซียน

1.1 การประมาณกรณีตัวแปรเดียว (Univariate Estimation) โดยใช้ตัวประมาณแคปลานและไมย์เออร์

การประมาณกรณีตัวแปรเดียวโดยใช้ตัวประมาณแคปลานและไมย์เออร์ ในงานวิจัยนี้ทดสอบด้วยวิธีเบรสโลว์ ซึ่งเป็นการเปรียบเทียบความแตกต่างของฟังก์ชันการอยู่รอด โดยวิธีนี้จะใช้กับการอยุ่รอดที่มีระยะเวลาสั้น โดยมีสมมติฐานหลักว่าผู้เอาประกันภัยทั้งหมดมีฟังก์ชันการแจกแจงการอยู่รอดเท่ากัน โดยที่

$$L(\beta) = \prod_{i=1}^{D} \frac{\exp(\beta^t S_i)}{\sum_{j \in R_i} \exp(\beta^t Z_j)^{d_i}}$$

เมื่อ D แทนจำนวนครั้งการเกิดเหตุการณ์การเรียกร้องสินไหมทดแทนเมื่อเวลา $t_1 < t_2 < ... < t_p$

d_i แทนจำนวนการเกิดเหตุการณ์การเรียกร้องสินไหมทดแทนที่เวลา t_i

D_i แทนเซตของผู้เอาประกันภัยที่เรียกร้องสินไหมทดแทนที่เวลา t_i

Z_j แทนตัวแปรร่วมที่เป็นอิสระกับเวลา และ $S_i = \sum_{j \in D_i} Z_j$

การแบ่งตัวแปรอายุ และจำนวนเงินเอาประกันภัยเป็นช่วงๆ และสร้าง dummy variable ให้กับตัวแปรดังกล่าว ตามตาราง 3.2 และ ตาราง 3.3 เพื่อใช้ในการวิเคราะห์ข้อมูลในหัวข้อ 4.2.1

ตาราง 3.2 dummy variable ของตัวแปรอายุผู้เอาประกันภัย

อายุ	1	2	3	4
20-29	1	0	0	0
30-39	0	1	0	0
40-49	0	0	1	0
50-59	0	0	0	1
60-69	0	0	0	0

ตาราง 3.3 dummy variable ของตัวแปรจำนวนเงินเอาประกันภัย

จำนวนเงินเอาประกันภัย	1	2	3	4
1-200000	1	0	0	0
200001-400000	0	1	0	0
400001-600000	0	0	1	0
600001-800000	0	0	0	1
800001-1000000	0	0	0	0

การทดสอบความแตกต่างของฟังก์ชันการอยู่รอดด้วยวิธีเบรสโลว์ (Breslow's method) โดยที่ค่าสถิติคือ

$$\chi_{k-1}^{\ 2} = U' V_w^{-1} U$$

เมื่อ $U_j = \sum_{i=1}^{k} w_j (d_{ij} - e_{ij})$

$$(V_j)_{ii} = \frac{n_{ij}(n_j - n_{ij}) d_j (n_j - d_j)}{n_j^{\ 2}(n_j - 1)} \ , \ i = 1, 2, ..., k-1$$

$$e_{ij} = \frac{n_{ij} d_j}{n_j}$$

e_{ij} คือค่าคาดหวังการตายของกลุ่มที่ i คนที่ k ณ เวลาการสำรวจที่ต่างกัน j

w_j คือค่าน้ำหนักของการทดสอบของเบรสโลว์

d_j คือจำนวนเหตุการณ์การเรียกร้องสินไหมทดแทนที่เกิดขึ้น

n_{ij} คือจำนวนความเสี่ยงในกลุ่มที่ i ก่อนเวลาการสำรวจที่ต่างกัน j

1.2 การประมาณกรณีมีตัวแปรร่วมอื่น (Multivariate Estimation) โดยใช้ตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อก

ตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อก $h(t \mid Z) = h_0(t)\exp(\sum_{k=1}^{p}\beta_k Z_k)$ เมื่อ t คือ ระยะเวลาที่ล่าช้าของแต่ละกรรมธรรม์ Z คือเวกเตอร์ของตัวแปรร่วม β คือเวกเตอร์ของสัมประสิทธิ์ถดถอย $h_0(t)$ คือความเสี่ยงฐาน (Baseline hazard) ซึ่งหมายถึงความเสี่ยงของแต่ละบุคคลเมื่อตัวแปรอิสระมีค่าเท่ากับศูนย์ p แทนจำนวนตัวแปรร่วมที่สนใจ

ค่าสัมประสิทธิ์การถดถอย (β) ประมาณได้จากวิธี Partial maximum likelihood ซึ่งเป็นการเปรียบเทียบภาวะน่าจะเป็นของตัวแบบที่กำหนดให้ตัวแปรร่วมทุกตัวไม่มีผลต่อตัวแบบ เขียนแทนด้วย $L(0)$ กับภาวะน่าจะเป็นของตัวแบบที่ตัวแปรอิสระมีผลต่อตัวแบบ เขียนแทนด้วย $L(\beta)$ แล้วคำนวณค่า $L(\beta)$ ที่มากที่สุดเท่าที่จะเป็นไปได้ ที่ทำให้ได้ค่า $L(0)/L(\beta)$ หรืออัตราส่วนภาวะน่าจะเป็น (Likelihood ratio) มีค่ามากที่สุด ดังนั้น Partial likelihood สามารถเขียนได้ดังนี้

$$L(\beta) = \prod_{i=1}^{D} \frac{\exp\left[\sum_{k=1}^{p}\beta_k Z_{(i)k}\right]}{\sum_{j\in R(t_i)}\exp\left[\sum_{k=1}^{p}\beta_k Z_{jk}\right]} \quad \text{เมื่อ } j = 1,\ldots,n \tag{3.1}$$

เมื่อ $Z_{(i)k}$ คือตัวแปรร่วมตัวที่ k ของตัวอย่างที่มีระยะเวลาที่ล่าช้า t_i

$R(t_i)$ คือกลุ่มของตัวอย่างที่อยู่ในการศึกษาก่อนถึงเวลา t_i

p คือจำนวนตัวแปรร่วม

n คือจำนวนกลุ่มตัวอย่างที่ทำการศึกษา

จากสมการ (3.1) จะได้

$$\ln L(\beta) = \sum_{i=1}^{D}\left\{\sum_{k=1}^{p}\beta_k Z_{(i)k} - \ln\sum_{j\in R(t_i)}\exp\left[\sum_{k=1}^{p}\beta_k Z_{jk}\right]\right\} \tag{3.2}$$

แทน $LL(\beta) = \ln L(\beta)$ ในสมการ (3.2) และหาอนุพันธ์เพื่อหาค่าประมาณภาวะน่าจะเป็นสูงสุดบางส่วน (maximum partial likelihood estimator) หรือค่า β ได้จาก

$$\frac{\partial}{\partial \beta_h} LL(\beta) = \sum_{i=1}^{D} Z_{(i)h} - \sum_{i=1}^{D} \frac{\sum_{j \in R(t_i)} Z_{jh} \exp\left[\sum_{k=1}^{p} \beta_k Z_{jk}\right]}{\sum_{j \in R(t_i)} \exp\left[\sum_{k=1}^{p} \beta_k Z_{jk}\right]}$$

ให้ $U_h(\beta) = \frac{\partial}{\partial \beta_h} LL(\beta)$ จะประมาณค่า β ได้จากการหาผลเฉลยของ $U_h(\beta) = 0$

ในการวิเคราะห์ตัวแบบค็อก ส่วนใหญ่เรียกว่าตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อก เพราะทำการวิเคราะห์ความเสี่ยงสัมพัทธ์ (relative risk หรือ hazard ratio) ของตัวอย่างที่มีตัวแปรร่วม Z เทียบกับตัวอย่างที่มีตัวแปรร่วม Z^* คือ

$$\frac{h(t \mid Z)}{h(t \mid Z^*)} = \frac{h_0(t) \exp\left[\sum_{k=1}^{p} \beta_k Z_k\right]}{h_0(t) \exp\left[\sum_{k=1}^{p} \beta_k Z^*{}_k\right]} = \exp\left[\sum_{k=1}^{p} \beta_k (Z_k - Z^*{}_k)\right]$$

ขั้นตอนการสร้างตัวแบบตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อก

ในงานวิจัยนี้ผู้วิจัยได้ใช้โปรแกรม SPSS ในการสร้างตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อก โดยมีขั้นตอน ดังนี้

1.1 วิเคราะห์กราฟฟังก์ชันการแจกแจงการอยู่รอดของข้อมูลระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทน สำหรับการประกันภัยโรคร้ายแรง ภายใต้ตัวแปรร่วมอายุ เพศของผู้เอาประกันภัย รูปแบบผลประโยชน์ จำนวนเงินเอาประกันภัย สาเหตุของการเรียกร้องค่าสินไหมทดแทน และบริษัทประกันชีวิต 2 บริษัท ณ วันที่เรียกร้องสินไหมทดแทนที่แตกต่างกัน

1.2 ทำ global test สำหรับแต่ละปัจจัยที่คาดว่าจะมีผลต่อระยะเวลาที่ล่าช้าระหว่างเวลาที่ทำการวินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทน โดยกำหนดสมมติฐานหลัก $\beta_1 = \beta_2 = ... = \beta_p = 0$ จะปฏิเสธสมมติฐานหลักเมื่อมี β อย่างน้อยหนึ่งตัวไม่เท่ากับศูนย์ โดยที่ช่วงความเชื่อมั่น 95% (95% confidence interval) ของ $\exp(\beta_i)$ หาได้จาก $\exp(b_i \pm z_{1-\frac{0.05}{2}} \mathrm{se}(b_i))$ เมื่อ b_i แทนค่าประมาณสัมประสิทธิ์การถดถอย

1.3 เลือกปัจจัยที่มีนัยสำคัญทางสถิติสูงสุดเข้าตัวแบบ

1.4 ทำ local test สำหรับแต่ละปัจจัยที่เหลือที่คาดว่าจะมีผลต่อระยะเวลาที่ล่าช้าระหว่างเวลาที่ทำการวินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทน แล้วเลือกปัจจัยที่มีนัยสำคัญทางสถิติสูงสุดเข้ามาและหยุดเมื่อพบว่าไม่มีปัจจัยใดมีนัยสำคัญทางสถิติ

2 การวิเคราะห์การอยู่รอดแบบเบส์เซียน

การศึกษาครั้งนี้ใช้ตัวแบบเบอร์ (Burr model) และตัวแบบล็อกนอร์มอล (Log Normal) มาทำการวิเคราะห์แบบเบส์เซียน โดยมีขั้นตอนดังนี้

2.1 การวิเคราะห์ตัวแบบโดยไม่นำตัวแปรเข้ามาเกี่ยวข้อง (Null model)

2.1.1 ตัวแบบเบอร์

ให้ D_i แทนระยะเวลาที่ล่าช้าของการเรียกร้องค่าสินไหมทดแทนของกรมธรรม์ที่ i และตัวแบบเบอร์มีฟังก์ชันความหนาแน่นน่าจะเป็น คือ

$$f(D\,|\,\alpha,\tau,\beta)=\alpha^n\tau^n\prod_i^n((G(\alpha,\tau))^{-1}\exp(\beta_0))^{\alpha\tau}d_i^{\tau-1}[((G(\alpha,\tau))^{-1}\exp(\beta_0))^{\tau}+d_i^{\tau}]^{-(\alpha+1)} \tag{3.3}$$

เมื่อ $d_i,\alpha,\tau,s>0$ โดยที่ α และ τ เป็นพารามิเตอร์บ่งรูปร่าง (shape parameter) s เป็นพารามิเตอร์บ่งขนาด (scale parameter) และ n เป็นจำนวนกรมธรรม์ที่เกิดการล่าช้าในการเรียกร้องสินไหมทดแทน

และมี

$$E(D_i)=\exp(\eta_i)=\exp(\beta_0)=s_iG(\alpha,\tau)$$

ให้ $G(\alpha,\tau)=\dfrac{\Gamma(\alpha-\frac{1}{\tau})\Gamma(1-\frac{1}{\tau})}{\Gamma(\alpha)}$

เมื่อ $\beta_0=\log(s)+\log G(\alpha,\tau)$

และ $s=\left(G(\alpha,\tau)\right)^{-1}\exp(\beta_0)$

กำหนดการแจกแจงก่อน (non informative prior) ของพารามิเตอร์ในตัวแบบ ดังนี้

α มีการแจกแจงแบบแกมมาด้วยพารามิเตอร์ 0.001 และ 0.001

τ มีการแจกแจงแบบแกมมาด้วยพารามิเตอร์ 0.001 และ 0.001

s มีการแจกแจงแบบแกมมาด้วยพารามิเตอร์ 0.001 และ 0.001

2.1.2 ตัวแบบล็อกนอร์มอล

ให้ D_i แทนระยะเวลาที่ล่าช้าของการเรียกร้องค่าสินไหมทดแทนของกรมธรรม์ที่ i และตัวแบบล็อกนอร์มอลมีฟังก์ชันความหนาแน่นน่าจะเป็น คือ

$$f(D \mid \sigma^2, \beta) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \prod_i^n \frac{1}{d_i} \exp\left(- \frac{\sum (\log(d_i) - \eta_i)^2}{2\sigma^2} \right)$$

(3.4)

ให้ D_i มีการแจกแจงแบบล็อกนอร์มอลด้วยพารามิเตอร์ μ_i และ σ^2 โดยที่

$$\mu_i = \eta_i = \beta_0$$

เมื่อ $i = 1, ..., n$

การแจกแจงก่อน (non informative prior) ของพารามิเตอร์

σ^2 มีการแจกแจงแบบอินเวอร์สแกมมาด้วยพารามิเตอร์ 0.001และ 0.001

2.2 การวิเคราะห์ตัวแบบที่มีตัวแปรร่วม (Analysis with covariate)

งานวิจัยนี้มีตัวแปรทั้งหมด 6 ตัว ได้แก่ อายุ เพศ รูปแบบผลประโยชน์ จำนวนเงินผลประโยชน์ สาเหตุของการเรียกร้องค่าสินไหมทดแทน แบ่งเป็น 8 กลุ่มโรคร้ายแรง และบริษัทประกันภัยที่ทำการศึกษา โดยในงานวิจัยนี้ทำการศึกษา 2 บริษัท

2.2.1 ตัวแบบเบอร์

การวิเคราะห์ตัวแบบเบอร์ร่วมกับตัวแปรมีฟังก์ชันความหนาแน่นน่าจะเป็นเช่นเดียวกับสมการ (3.3)

โดยที่

$$\eta_i = \beta_0 + \sum_{j=1}^{4} \beta_j Z_{ij} + \beta_{5,Cause_i} + \beta_{6,Office_i} \qquad (3.5)$$

เมื่อ $i = 1, ..., 840 \quad Office_i = 1, 2 \quad Cause_i = 1, ..., 8$

แทน $z_{ij} = (x_{ij} - \bar{x}_j) / sd(x_j)$ เป็นค่าสังเกตมาตรฐาน (standardized observation)

x_{ij} เป็นค่าสังเกตที่แท้จริง (original observation)

β_0 เป็น standardized intercept term

$\beta_1, ..., \beta_4 \quad \beta_{5.1}, ..., \beta_{5.8}$ และ $\beta_{6.1}, \beta_{6.2}$ เป็นสัมประสิทธิ์ของตัวแปรร่วม โดยที่ $\beta_j \neq 0$

กำหนดการแจกแจงก่อน (non informative prior) ของพารามิเตอร์ในตัวแบบ ดังนี้

α มีการแจกแจงแบบแกมมาด้วยพารามิเตอร์ 0.001 และ 0.001

τ มีการแจกแจงแบบแกมมาด้วยพารามิเตอร์ 0.001 และ 0.001

β_j มีการแจกแจงแบบปกติด้วยพารามิเตอร์ 0 และ 0.001 เมื่อ $j = 1, ..., 4$

$\beta_{5,k}$ มีการแจกแจงแบบปกติด้วยพารามิเตอร์ 0 และ 0.001 เมื่อ $k = 1, ..., 8$

$\beta_{6,l}$ มีการแจกแจงแบบปกติด้วยพารามิเตอร์ 0 และ 0.001 เมื่อ $l = 1, 2$

2.2.2 ตัวแบบล็อกนอร์มอล

ให้ D_i แทนระยะเวลาที่ล่าช้าของการเรียกร้องค่าสินไหมทดแทนสำหรับกรมธรรม์ที่ i การวิเคราะห์ตัวแบบล็อกนอร์มอลร่วมกับตัวแปรมีฟังก์ชันความหนาแน่นน่าจะเป็นเช่นเดียวกับสมการ (3.4) โดยที่

D_i มีการแจกแจงแบบล็อกนอร์มอลด้วยพารามิเตอร์ μ_i และ σ^2

และ

$$\eta_i = \beta_0 + \sum_{j=1}^{4} \beta_j Z_{ij} + \beta_{5,Cause_i} + \beta_{6,Office_i}$$

เมื่อ $i = 1,...,840$ $Office_i = 1,2$ $Cause_i = 1,...,8$

กำหนดการแจกแจงก่อน (non informative prior) ของพารามิเตอร์ในตัวแบบ ดังนี้

σ^2 มีการแจกแจงแบบอินเวอร์สแกมมาด้วยพารามิเตอร์ 0.001และ 0.001

β_j มีการแจกแจงแบบแบบปกติด้วยพารามิเตอร์ 0 และ 0.001 เมื่อ $j = 1,...,4$

$\beta_{5,k}$ มีการแจกแจงแบบแบบปกติด้วยพารามิเตอร์ 0 และ 0.001 เมื่อ $k = 1,...,8$

$\beta_{6,l}$ มีการแจกแจงแบบแบบปกติด้วยพารามิเตอร์ 0 และ 0.001 เมื่อ $l = 1,2$

2.3 คัดเลือกตัวแบบที่เหมาะสม

ในการสร้างตัวแบบของระยะเวลาในงานวิจัยนี้ ใช้ตัวแปรร่วมทั้งหมด 12 ตัวแปร ทำให้เกิดตัวแบบที่เป็นไปได้เท่ากับจำนวนสับเซตทั้งหมดของจำนวนตัวแปรร่วม นั่นคือ 2^{12} ตัวแบบ เขียนแทนด้วย m_t เมื่อ $t = 1,...,2^{12}$ และ D เป็นตัวแปรสุ่มประเภทไม่ต่อเนื่องที่แทนระยะเวลาระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทน โดยมี $D_1, D_2, ..., D_n$ เป็นตัวอย่างสุ่ม และ $f(D \mid m_t)$ เป็นภาวะน่าจะเป็นส่วนริม (marginal likelihood) ภายใต้ m_t

โดยทฤษฎีบทเบส์(Bayes theorem) จะได้ความน่าจะเป็นหลัง(posterior probability)

$$f(m_t \mid D) = \frac{f(D \mid m_t) f(m_t)}{\sum_{t=1}^{k} f(D \mid m_t) f(m_t)}$$

เมื่อ $f(m_t)$ คือ ฟังก์ชันความน่าจะเป็นของตัวแบบก่อน (prior model probability)

จากนั้นเปรียบเทียบ 2 ตัวแบบ เช่น m_i กับ m_j เพื่อหาตัวแบบที่เหมาะสมที่สุดด้วย posterior odd ratio

$$\frac{f(m_i \mid D)}{f(m_j \mid D)} = \frac{f(D \mid m_i) f(m_i)}{f(D \mid m_j) f(m_j)}$$

เมื่อ $i, j \in [1,...,64]$ ซึ่ง $B_{ij} = \frac{f(D|m_i)}{f(D|m_j)}$ เป็น Bayes factor และ $\frac{f(m_i)}{f(m_j)}$ เป็น prior odds ratio

ในงานวิจัยนี้จะคัดเลือกตัวแปรด้วยวิธีการคัดเลือกตัวแปรของกิบส์ โดยพิจารณาพารามิเตอร์เดิม (prior parameter) สำหรับการแจกแจงของ β เป็นแบบ Zellner 's g – prior ซึ่งมีรายละเอียดดังต่อไปนี้

2.2.3 การคัดเลือกตัวแปรของกิบส์ (Gibbs Variable Selection)

ให้ γ เป็นเวกเตอร์ชี้วัด (indicator vector) ขนาด $p \times 1$ โดยที่ $\gamma \in \{0,1\}^p$ และ $p = 6$ แทนจำนวนตัวแปรร่วมทั้งหมดในตัวแบบ จากนั้นเพิ่ม γ ลงไปในการพยากรณ์เชิงเส้นสำหรับตัวแบบ จะได้

$$\eta_i = \beta_0 + \sum_{j=1}^{4} \gamma_j \beta_j Z_{ij} + \gamma_5 \beta_{5,Cause_i} + \gamma_6 \beta_{6,Office_i}$$

โดยที่ γ_j มีการแจกแจงแบบเบอร์นูลลีด้วยพารามิเตอร์ 0.5 เพื่อให้พารามิเตอร์แต่ละตัวมีค่าน้ำหนักที่เท่ากัน โดยพารามิเตอร์ก่อน (prior parameters) สำหรับการแจกแจงของ β เป็น Zellner' s g – prior

Zellner' s g – prior (Zellner, 1986) จะใช้สำหรับตัวแบบปกติซึ่ง Normal – Inverse Gamma conjugate prior หาได้จากงานวิจัยของ Ntzoufras (2009) และงานวิจัยของ Ozkok (2010) เพื่อที่จะใช้กับการแจกแจงของข้อมูลที่มีจะใช้การประมาณค่าเมทริกซ์ความแปรปรวนและความแปรปรวนร่วม (variance – covariance matrix) โดยที่

β มีการแจกแจงแบบปกติด้วยพารามิเตอร์ μ และ S

ในที่นี้ β เป็นเวกเตอร์สัมประสิทธิ์ (coefficient vector)

$\mu = \mu_0,..., \mu_p$ เป็นเวกเตอร์ค่าเฉลี่ยก่อน (prior mean vector) และ S เป็นเมทริกซ์ความแปรปรวนร่วมก่อน (prior covariance matrix) ซึ่ง

$$\mu_j = (1-\gamma_j)\bar{\mu}_{\beta_j}, \ j = 0,..., p$$

$$S_{j,k} = (\gamma_j \gamma_k) n \sum_{jk} + (1-\gamma_j \gamma_k) I\{j = k\} \bar{\sigma}^2_{\beta_j}, \ j,k = 0,..., p$$

เมื่อ $\sum = (-D^2 l(\hat{\beta}))^{-1}$ ถูกประมาณด้วยเมทริกซ์ความแปรปรวน และความแปรปรวนร่วม (variance – covariance matrix) ซึ่ง $D^2 l(\hat{\beta})$ คือ Hessian matrix ของ loglikelihood ที่ประมาณด้วยภาวะน่าจะเป็น-สูงสุด (maximum likelihood) (Kass และ Wasserman, 1995 ;Ntzoufras และคณะ, 2003) เมื่อ $\gamma_0 = 1$ ที่ $\bar{\mu}_{\beta_j}$ และ $\bar{\sigma}^2_{\beta_j}$ เป็นตัวประมาณหลัง (posterior estimate) ของค่าเฉลี่ย และความแปรปรวนของ β_j

3 เกณฑ์การเลือกตัวแบบในการวิเคราะห์การอยู่รอดแบบเบส์เซียน

เกณฑ์ที่ใช้ในการตัดสินว่าตัวแบบเบอร์ หรือตัวแบบล็อกนอร์มอล เป็นตัวแบบที่เหมาะสมกับข้อมูลระยะเวลาที่ล่าช้าระหว่างเวลาที่ทำการวินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง สำหรับการวิเคราะห์การอยู่รอดแบบเบส์เซียน ทั้งการวิเคราะห์ตัวแบบโดยไม่นำตัวแปรร่วมเข้ามาเกี่ยวข้อง และการวิเคราะห์ตัวแบบที่มีตัวแปรร่วม คือค่า DIC (deviance information criterion) โดยที่

$$DIC = \bar{D} + p_D$$

เมื่อ $\bar{D} = -2\log L(\theta)$

p_D แทนความคลาดเคลื่อนของพารามิเตอร์ (the effective number of parameter)
ตัวแบบที่มีค่า DIC ต่ำจะเหมาะสมกับข้อมูลมากกว่าตัวแบบที่มีค่า DIC สูงกว่า

3.3 โปรแกรมที่ใช้ในงานวิจัย

ในงานวิจัยนี้ผู้วิจัยได้ใช้โปรแกรม SPSS ซึ่งเป็นโปรแกรมสำหรับการวิเคราะห์ทางสถิติ มาวิเคราะห์การอยู่รอดแบบไม่ใช้เบส์เซียนด้วยตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อก และใช้โปรแกรม WinBUGS ซึ่งเป็นโปรแกรมทางสถิติสำหรับการประมาณค่าประมาณเบย์ โดยใช้วิธี MCMC ในการประมาณค่าพารามิเตอร์ สามารถดาวน์โหลดโปรแกรมโดยไม่เสียค่าใช้จ่ายได้ที่ http://www.mrc-bsu.cam.ac.uk/bugs/

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 4
ผลการวิเคราะห์ข้อมูล

งานวิจัยนี้มีวัตถุประสงค์เพื่อหาตัวแบบของระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง ตามการวิเคราะห์แบบไม่ใช้เบส์เซียน และแบบเบส์เซียน โดยที่การวิเคราะห์การอยู่รอดแบบไม่ใช้เบส์เซียน ใช้ตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อก และการวิเคราะห์การอยู่รอดแบบเบส์เซียน จะใช้ตัวแบบเบอร์ และตัวแบบล็อกนอร์มอล

สัญลักษณ์ต่างๆที่ใช้ในงานวิจัย

SE คือ ส่วนเบี่ยงเบนมาตรฐาน

Wald คือ ค่าทดสอบทางสถิติของวอล์ด (Wald test)

df คือ degree of freedom

p – value คือ ค่า p - value

Lower คือ ขอบล่างของช่วงความเชื่อมั่นที่ร้อยละ 95

Upper คือ ขอบบนของช่วงความเชื่อมั่นที่ร้อยละ 95

Burr คือ การแจกแจงแบบเบอร์

LN คือ การแจกแจงแบบล็อกนอร์มอล

BA คือ จำนวนเงินเอาประกันภัย

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4.1 การวิเคราะห์ข้อมูลพื้นฐาน

ข้อมูลที่ใช้ในงานวิจัยนี้เป็นข้อมูลระยะเวลาระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง ของบริษัทประกันชีวิต 2 บริษัท มีหน่วยเป็นวัน โดยประกอบด้วยตัวแปรร่วม 6 ตัว ได้แก่ อายุของผู้เอาประกันภัย ณ วันที่เรียกร้องสินไหมทดแทน มีหน่วยเป็นปี เพศของผู้เอาประกันภัย รูปแบบผลประโยชน์ โดยแบ่งเป็นผลประโยชน์ที่ครอบคลุมการเสียชีวิต (full accelerated) และผลประโยชน์ที่ไม่ครอบคลุมการเสียชีวิต (stand-alone) จำนวนเงินเอาประกันภัย สาเหตุของการเรียกร้องค่าสินไหมทดแทน ในงานวิจัยนี้ คือโรคที่เกิดขึ้นกับผู้เอาประกันภัย แบ่งเป็นกลุ่มโรคร้ายแรง 8 กลุ่ม และบริษัทประกันชีวิต 2 บริษัท พบว่าข้อมูลของทั้ง 2 บริษัทมีค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าสูงสุด และค่าต่ำสุดของระยะเวลาที่

ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรงดังตารางที่ 4.1

ตาราง 4.1 ค่าเฉลี่ยเลขคณิต มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ค่าสูงสุด และค่าต่ำสุดของระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรงของบริษัทประกันภัย 2 แห่งในประเทศไทย (หน่วยเป็นวัน)

บริษัทที่	จำนวนข้อมูล	มัธยฐาน	ส่วนเบี่ยงเบนมาตรฐาน	ค่าสูงสุด	ค่าต่ำสุด
1	511	85	147.55	1427	5
2	329	80	141.85	1407	7

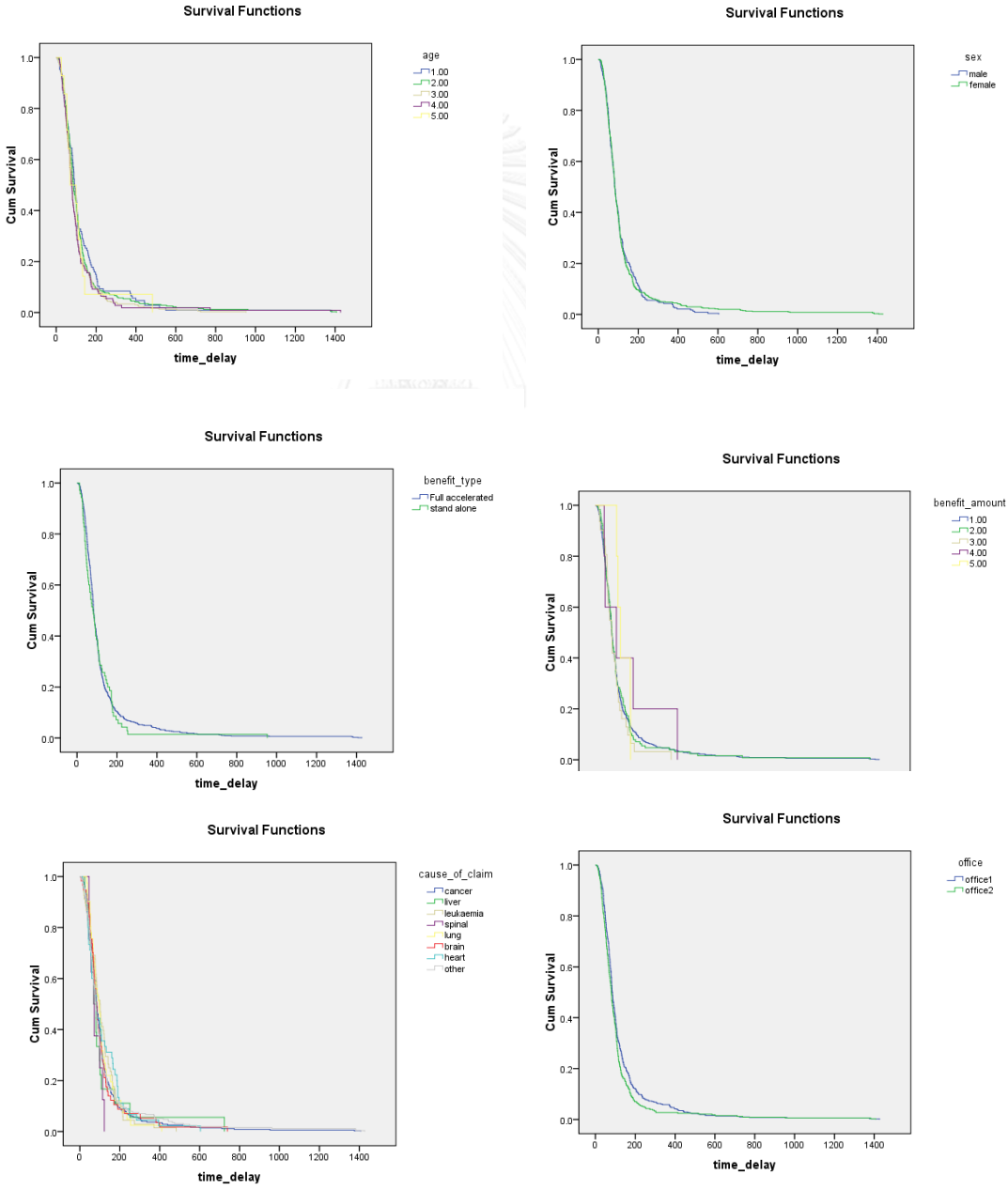
พิจารณาตัวแปรร่วมที่คาดว่าจะมีผลต่อระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรงของทั้งสองบริษัทพบว่า บริษัทที่ 1 มีผู้เอาประกันภัยทั้งหมด 511คน ส่วนมากมีอายุ 36 ปี ณ วันที่เรียกร้องสินไหมทดแทน ในจำนวนนี้มีเพศชายร้อยละ 32 เพศหญิงร้อยละ 68 ทำประกันภัยรูปแบบผลประโยชน์ที่ครอบคลุมการเสียชีวิตร้อยละ 97 จำนวนเงินเอาประกันภัยส่วนใหญ่ 100,000 บาท คิดเป็นร้อยละ 52 และสาเหตุการเรียกร้องสินไหมทดแทนส่วนมากเกิดจากโรคมะเร็งคิดเป็นร้อยละ 39 ของจำนวนผู้เอาประกันภัยทั้งหมด ส่วนบริษัทที่ 2 มีผู้เอาประกันภัยทั้งหมด 329 คน ส่วนมากมีอายุ 41 ปี ณ วันที่เรียกร้องสินไหมทดแทน ในจำนวนนี้มีเพศชายร้อยละ 22 เพศหญิงร้อยละ 78 ทำประกันภัยรูปแบบผลประโยชน์ที่ครอบคลุมการเสียชีวิตร้อยละ 83 จำนวนเงินเอาประกันภัยส่วนใหญ่ 100,000 บาท คิดเป็นร้อยละ 51 และสาเหตุการเรียกร้องสินไหมทดแทนส่วนมากเกิดจากโรคมะเร็งคิดเป็นร้อยละ 45 ของจำนวนผู้เอาประกันภัยทั้งหมด

เมื่อวิเคราะห์ข้อมูลพื้นฐานแล้วจึงทำการวิเคราะห์ข้อมูลแบบไม่ใช้เบส์เซียน โดยทำการประมาณกรณีตัวแปรเดียว (Univariate Estimation) ด้วยตัวประมาณแคปลานและไมย์เออร์ (Kaplan-Meier) และการประมาณกรณีมีตัวแปรร่วมอื่น (Multivariate Estimation) ด้วยตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อก แล้วจึงวิเคราะห์แบบเบส์เซียนด้วยตัวแบบเบอร์และตัวแบบล็อกนอร์มอลต่อไป

4.2 การวิเคราะห์แบบไม่ใช้เบส์เซียน

4.2.1 การประมาณกรณีตัวแปรเดียว (Univariate Estimation) โดยใช้ตัวประมาณแคปลานและไมย์เออร์

จากการวิเคราะห์ด้วยตัวประมาณแคปลานและไมย์เออร์โดยใช้วิธีเบรสโลว์ในการทดสอบความแตกต่างของฟังก์ชันการอยู่รอดของข้อมูลระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง ในตัวแปรร่วมแต่ละตัวได้ผลดังภาพที่ 4.1



ภาพที่ 4.1 แสดงฟังก์ชันการแจกแจงการอยู่รอด (survival distribution function) ของข้อมูลระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทน สำหรับการประกันภัยโรคร้ายแรง ภายใต้ตัวแปรร่วมอายุ เพศของผู้เอาประกันภัย รูปแบบผลประโยชน์ จำนวนเงินเอาประกันภัย สาเหตุของการเรียกร้องค่าสินไหมทดแทน และบริษัทประกันชีวิต 2 บริษัท ณ วันที่เรียกร้องสินไหมทดแทนที่แตกต่างกัน โดยใช้ตัวประมาณแคปลานและไมย์เออร์

หมายเหตุ

ตัวแปรร่วมอายุ (age)

1.00 แทนผู้เอาประกันภัยอายุ 20-29 ปี 2.00 แทนผู้เอาประกันภัยอายุ 30-39 ปี 3.00 แทนผู้เอาประกันภัยอายุ 40-49 ปี 4.00 แทนผู้เอาประกันภัยอายุ 50-59 ปี และ 5.00 แทนผู้เอาประกันภัยอายุ 60-69 ปี

ตัวแปรร่วมจำนวนเงินเอาประกันภัย (benefit_amount)

1.00 แทนจำนวนเงินเอาประกันภัย 1-200000 บาท 2.00 แทนจำนวนเงินเอาประกันภัย 200001-400000 บาท 3.00 แทนจำนวนเงินเอาประกันภัย 400001-600000 บาท 4.00 แทนจำนวนเงินเอาประกันภัย 600001-800000 บาท และ 5.00 แทนจำนวนเงินเอาประกันภัย 800001-1000000 บาท

จากภาพที่ 4.1 เมื่อพิจารณาด้วยค่ามัธยฐานของฟังก์ชันการอยู่รอด (median survival function) ของระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง และทำการเปรียบเทียบพบว่า

- ตัวแปรร่วมอายุของผู้เอาประกันภัยที่แตกต่างกัน มีค่ามัธยฐานของฟังก์ชันการอยู่รอดเป็น 92 วัน 87 วัน 79 วัน 78 วัน และ 70 วัน สำหรับผู้เอาประกันภัยอายุ 20-29 ปี 30-39 ปี 40-49 ปี 50-59 ปี และ 60-69 ปี ตามลำดับ นั่นคือ ผู้เอาประกันภัยที่อายุน้อยจะมีระยะเวลาระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรงยาวนานกว่าผู้เอาประกันภัยที่มีอายุมากกว่า
- ตัวแปรร่วมเพศหญิงมีค่ามัธยฐานของฟังก์ชันการอยู่รอดเท่ากับเพศชาย คือ 83 วัน
- ตัวแปรร่วมรูปแบบผลประโยชน์ พบว่ารูปแบบผลประโยชน์แบบครอบคลุมการเสียชีวิตมีค่ามัธยฐานของฟังก์ชันการอยู่รอด 83 วัน และรูปแบบผลประโยชน์แบบไม่ครอบคลุมการเสียชีวิตมีค่ามัธยฐานของฟังก์ชันการอยู่รอด 60 วัน นั่นคือผู้เอาประกันภัยที่มีรูปแบบผลประโยชน์แบบครอบคลุมการเสียชีวิตจะมีระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรค และ

เวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรงยาวนานกว่าผู้เอาประกันภัยที่มีรูปแบบผลประโยชน์แบบไม่ครอบคลุมการเสียชีวิต

- ตัวแปรร่วมจำนวนเงินเอาประกันภัยที่แตกต่างกัน มีค่ามัธยฐานของฟังก์ชันการอยู่รอดเป็น 82 วัน 84 วัน 82 วัน 106 วัน และ 126 วัน สำหรับจำนวนเงินเอาประกันภัย 1-200,000 บาท 200,001-400,000 บาท 400,001-600,000 บาท 600,001-800,000 บาท และ 800,001-1,000,000 บาทตามลำดับ นั่นคือผู้เอาประกันภัยที่มีจำนวนเงินเอาประกันภัยน้อยจะมีระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนน้อยกว่าผู้เอาประกันภัยที่มีจำนวนเงินเอาประกันภัยมากกว่า
- ตัวแปรร่วมสาเหตุการเรียกร้องสินไหมทดแทน พบว่าได้ค่ามัธยฐานของฟังก์ชันการอยู่รอดดังตาราง 4.2

ตาราง 4.2 ค่ามัธยฐานของฟังก์ชันการอยู่รอด (median survival function) ของระยะเวลาระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนของการประกันภัยโรคร้ายแรง สำหรับการเปรียบเทียบภายใต้ภายใต้ตัวแปรร่วมสาเหตุการเรียกร้องสินไหมทดแทน (หน่วยเป็นวัน)

สาเหตุ	โรคมะเร็ง	โรคเกี่ยวกับตับ	โรคเลือด	โรคประสาท	โรคปอด	โรคสมอง	โรคหัวใจ	โรคอื่นๆ
มัธยฐาน	85	70	101	70	97	85	82	78

จากตาราง 4.2 พบว่าตัวแปรร่วมโรคเลือดมีค่ามัธยฐานของฟังก์ชันการอยู่รอดมากกว่าโรคอื่นๆ คือ 101 วัน นั่นคือผู้เอาประกันภัยที่เรียกร้องค่าสินไหมทดแทนเนื่องจากป่วยเป็นโรคเลือดจะมีระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนมากกว่าผู้เอาประกันภัยที่เรียกร้องค่าสินไหมทดแทนเนื่องจากป่วยเป็นโรคร้ายแรงอื่นๆ

- ตัวแปรร่วมบริษัทพบว่า บริษัทที่ 1 มีค่ามัธยฐานของฟังก์ชันการอยู่รอด 85 วัน และบริษัทที่ 2 มีค่ามัธยฐานของฟังก์ชันการอยู่รอด 80 วัน นั่นคือบริษัทที่ 1 มีระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนมากกว่าบริษัทที่ 2

จากนั้นทำการวิเคราะห์ด้วยตัวประมาณแคปลานและไมย์เออร์โดยใช้วิธีเบรสโลว์ในการทดสอบความแตกต่างของฟังก์ชันการอยู่รอดได้ผลดังตารางที่ 4.3

ตาราง 4.3 เปรียบเทียบฟังก์ชันการแจกแจงการอยู่รอด (survival distribution function) ของข้อมูล ระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรงโดยใช้ตัวประมาณ Kaplan-Meier

ตัวแปรร่วม	chi-square	df	p-value
อายุ	6.455	4	0.168
เพศ	0.048	1	0.826
รูปแบบผลประโยชน์	0.874	1	0.350
จำนวนเงินเอาประกันภัย	3.822	4	0.431
สาเหตุการเรียกร้องสินไหมทดแทน	3.840	7	0.798
บริษัท	6.054	1	0.014

จากตารางที่ 4.3 พบว่าตัวแปรร่วมที่มีนัยสำคัญทางสถิติ $(\alpha = 0.05)$ คือ ตัวแปรร่วมบริษัท แต่เมื่อพิจารณาจากภาพที่ 4.1 จะเห็นว่ามีตัวแปรร่วมอื่นที่น่าจะส่งผลต่อตัวแบบนอกจากตัวแปรร่วมบริษัท คือตัวแปรร่วมเพศ รูปแบบผลประโยชน์ และจำนวนเงินเอาประกันภัย

4.2.2 การประมาณกรณีมีตัวแปรร่วมอื่น (Multivariate Estimation) โดยใช้ตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อก

การประมาณกรณีมีตัวแปรร่วมอื่น โดยใช้ตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อกนั่น ขั้นแรกจะนำข้อมูลของผู้เอาประกันภัยทั้งหมดจำนวน 840 ราย มาทำการทดสอบสมมติฐานหลัก $\beta_1 = \beta_2 = ... = \beta_p = 0$ โดยใช้วิธี Forward และทดสอบด้วยการทดสอบทางสถิติของวอล์ด (Wald test) ซึ่งแสดงค่าในตาราง 4.4

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตาราง 4.4 ค่า chi-square df และ p-value ของตัวแปรร่วมทุกตัว เมื่อวิเคราะห์ด้วยตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อก

ตัวแปรร่วม	Score	df	p-value
อายุ	4.774	1	0.029
เพศ	0.026	1	0.873
รูปแบบผลประโยชน์	0.445	1	0.505
จำนวนเงินเอาประกันภัย	0.729	1	0.393
โรคเกี่ยวกับตับ	0.084	1	0.771
โรคเลือด	0.393	1	0.531
โรคระบบประสาท	1.342	1	0.247
โรคปอด	0.084	1	0.772
โรคความดันเลือดสูง	0.090	1	0.764
โรคหัวใจ	0.320	1	0.572
โรคอื่นๆ	0.004	1	0.950
บริษัท	5.384	1	0.020

จากตาราง 4.4 จะเห็นว่าตัวแปรร่วมอายุ และบริษัทมีนัยสำคัญทางสถิติ จึงทำการทดสอบต่อไปดังตาราง 4.5

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตาราง 4.5 ค่า chi-square df และ p-value ของขั้นตอนการคัดเลือกตัวแปร โดยใช้ตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อก

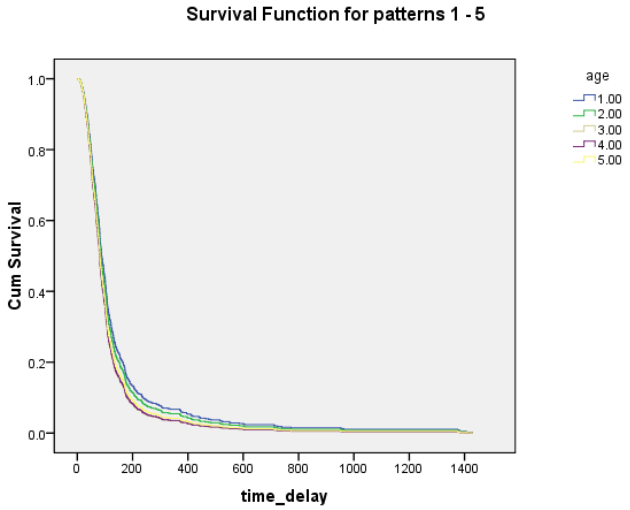
ตัวแปรร่วม		Score	df	p-value
ขั้นที่ 1	อายุ	4.327	1	0.038
	เพศ	0.102	1	0.749
	รูปแบบผลประโยชน์	0.001	1	0.988
	จำนวนเงินเอาประกันภัย	1.038	1	0.308
	โรคเกี่ยวกับตับ	0.067	1	0.796
	โรคเลือด	0.306	1	0.580
	โรคระบบประสาท	1.759	1	0.185
	โรคปอด	0.052	1	0.820
	โรคความดันเลือดสูง	0.036	1	0.850
	โรคหัวใจ	0.233	1	0.629
	โรคอื่นๆ	0.020	1	0.889
ขั้นที่ 2	เพศ	0.095	1	0.758
	รูปแบบผลประโยชน์	0.069	1	0.793
	จำนวนเงินเอาประกันภัย	1.218	1	0.270
	โรคเกี่ยวกับตับ	0.001	1	0.996
	โรคเลือด	0.465	1	0.495
	โรคระบบประสาท	1.755	1	0.185
	โรคปอด	0.083	1	0.773
	โรคความดันเลือดสูง	0.248	1	0.618
	โรคหัวใจ	0.377	1	0.539
	โรคอื่นๆ	0.118	1	0.731

จากตาราง 4.5 ขั้นตอนที่ 1 นำตัวแปรร่วมที่มีค่า p-value น้อยที่สุด คือตัวแปรร่วมบริษัทออกมาก่อน จากนั้นพิจารณาค่า p-value ของตัวแปรร่วมที่เหลือ จะพบว่าตัวแปรร่วมอายุมีค่า p-value น้อยกว่า 0.05 จึงทำขั้นตอนที่ 2 โดยนำตัวแปรร่วมอายุออกมา และพบว่าไม่มีตัวแปรร่วมใดมีนัยสำคัญทางสถิติอีก ทำให้ได้ค่าสัมประสิทธิ์การถดถอยของตัวแปรร่วมอายุ และบริษัทดังตารางที่ 4.6

ตาราง 4.6 แสดงค่าสัมประสิทธิ์การถดถอยของตัวแปรร่วมอายุ และบริษัท ที่ได้รับการคัดเลือกเข้าสู่ตัวแบบ

	ตัวแปรร่วม	β	SE	Wald	df	p-value	Exp(B)	95% CI for $\exp(\beta)$	
								Lower	Upper
ขั้นที่ 1	บริษัท	0.164	0.071	5.372	1	0.020	1.179	1.026	1.354
ขั้นที่ 2	อายุ	0.008	0.004	4.326	1	0.038	1.008	1.000	1.015
	บริษัท	0.157	0.071	4.920	1	0.027	1.171	1.018	1.345

จากตารางที่ 4.6 พบว่าตัวแปรร่วมที่มีนัยสำคัญทางสถิติ คือ ตัวแปรร่วมอายุ และบริษัท โดยมีกราฟฟังก์ชันการแจกแจงการอยู่รอด ของทั้ง 2 ตัวแปรร่วมดังภาพที่ 4.2 และภาพที่ 4.3

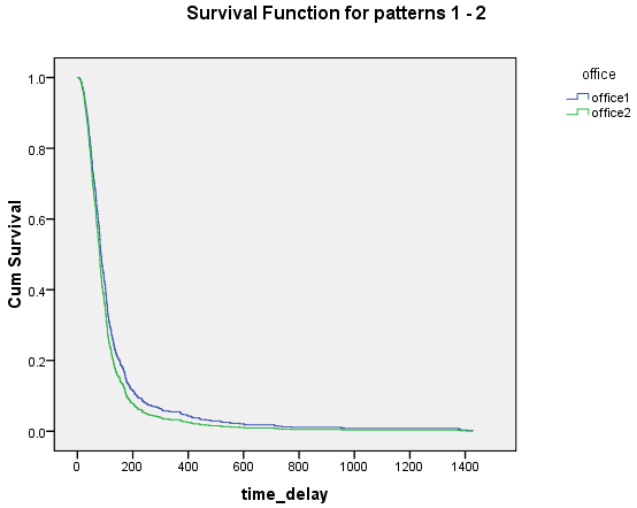


ภาพที่ 4.2

CHULALONGKORN UNIVERSITY

แสดงฟังก์ชันการแจกแจงการอยู่รอด (survival distribution function) สำหรับข้อมูลระยะเวลาระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง ภายใต้อายุผู้เอาประกันภัย ณ วันที่เรียกร้องสินไหมทดแทนที่แตกต่างกัน สำหรับผู้เอาประกันภัยอายุ 20-29 ปี 30-39 ปี 40-49 ปี 50-59 ปี และ 60-69 ปี โดยใช้ตัวแบบ การเสี่ยงภัยแบบอัตราส่วนค็อก

หมายเหตุ: 1.00 แทนผู้เอาประกันภัยอายุ 20-29 ปี 2.00 แทนผู้เอาประกันภัยอายุ 30-39 ปี 3.00 แทนผู้เอาประกันภัยอายุ 40-49 ปี 4.00 แทนผู้เอาประกันภัยอายุ 50-59 ปี และ 5.00 แทนผู้เอาประกันภัยอายุ 60-69 ปี



ภาพที่ 4.3 **แสดงฟังก์ชันการแจกแจงการอยู่รอด** (survival distribution function) **สำหรับข้อมูลระยะเวลาระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนของการประกันภัยโรคร้ายแรง ภายใต้บริษัทประกันภัยที่ทำการศึกษาทั้ง 2 บริษัท โดยใช้ตัวแบบ การเสี่ยงภัยแบบอัตราส่วนค็อก**

จากภาพที่ 4.2 พบว่าเมื่ออายุเพิ่มขึ้น ความน่าจะเป็นของการเกิดเหตุการณ์การเรียกร้องสินไหมทดแทนจะลดลงแตกต่างอย่างมีนัยสำคัญทางสถิติ นั่นคือ ผู้เอาประกันภัยที่อายุน้อยจะมีระยะเวลาระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนของการประกันภัยโรคร้ายแรงยาวนานกว่าผู้เอาประกันภัยที่มีอายุมากกว่า

จากภาพ 4.3 พบว่าบริษัทที่ 1 มีเส้นกราฟอยู่เหนือเส้นกราฟของบริษัทที่ 2 นั่นคือบริษัทที่ 1 มีระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนมากกว่าบริษัทที่ 2

ตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อกมีฟังก์ชันการเสี่ยงภัย ดังนี้

$$h(t \mid z) = h_0(t)\exp(0.008Z_1 + 0.157Z_{12}) \qquad (4.1)$$

เมื่อ $h_0(t)$ คือความเสี่ยงฐาน (Baseline hazard)

Z_1 คือตัวแปรแสดงอายุของผู้เอาประกันภัย

Z_{12} คือตัวแปรแสดงบริษัทประกันภัย โดย 0 คือบริษัทที่ 1 และ 1 คือบริษัทที่ 2

หากต้องการเปรียบเทียบค่าอัตราการเสี่ยงภัย (Hazard rate) ของผู้เอาประกันภัย 2 คน ที่มีลักษณะดังนี้คือ นาย ก อายุ 30 ปี ทำประกันภัยโรคร้ายแรงกับบริษัทที่ 1 ส่วนนาย ข อายุ 40 ปี ทำประกันภัยโรคร้ายแรงกับบริษัทที่ 1 เช่นเดียวกัน จากตัวแบบในสมการที่ 4.1 จะได้ค่าอัตราการเสี่ยงภัยของนาย ก เทียบกับนาย ข ดังนี้

$$\frac{h_{ก}(t)}{h_{ข}(t)} = \exp\big(0.008(30-40)\big) = 0.9231$$

นั่นคือ ความเสี่ยงที่จะเกิดระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนของนาย ก น้อยกว่านาย ข เนื่องจากอัตราการเสี่ยงภัยของนาย ก เทียบกับนาย ข มีค่าน้อยกว่า 1

4.3 การวิเคราะห์การอยู่รอดแบบเบส์เซียน

4.2.1 วิเคราะห์ตัวแบบโดยไม่นำตัวแปรเข้ามาเกี่ยวข้อง (Null model)

ตาราง 4.7 ค่าประมาณหลังของพารามิเตอร์สำหรับตัวแบบที่ไม่มีตัวแปรมาเกี่ยวข้อง

	Burr		LN	
	Mean	SD	Mean	SD
α	0.7303	0.04221	μ 4.426	0.027
τ	2.745	0.0653	τ 1.681	0.077
Mean	197.34		193.73	
Median	109.12		114.54	
SD	∞		221.54	
$\bar{D}$	8428		12280	
$\hat{D}$	7908		12270	
p_D	519.8		0.9687	
DIC	8947		12280	

จากตาราง 4.7 ที่แสดงค่าพารามิเตอร์สำหรับตัวแบบเบอร์ และตัวแบบล็อกนอร์มอลที่ไม่มีตัวแปรร่วม พบว่าตัวแบบเบอร์ให้ค่า DIC เท่ากับ 8947 ซึ่งน้อยกว่าค่า DIC ของตัวแบบล็อกนอร์มอล ดังนั้นตัวแบบเบอร์จึงมีความเหมาะสมกับตัวแบบของระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคถึงระยะเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรงมากกว่าตัวแบบล็อกนอร์มอล

4.2.2 วิเคราะห์ตัวแบบที่มีตัวแปรร่วม (Analysis with covariate)

วิเคราะห์ตัวแบบทีมีตัวแปรร่วมทั้ง 6 ตัวแปร ได้แก่ อายุ เพศ รูปแบบผลประโยชน์ จำนวนเงินเอาประกันภัย สาเหตุของการเรียกร้องสินไหมทดแทน และบริษัทประกันชีวิตทั้ง 2 บริษัท ทำให้ได้ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และค่าที่ช่วงความเชื่อมั่นร้อยละ 2.5 ถึง 97.5 ของพารามิเตอร์

พร้อมทั้งแสดงกราฟค่าประมาณหลังของสัมประสิทธิ์การถดถอยสำหรับตัวแบบเบอร์และตัวแบบล็อกนอร์มอล โดยแสดงค่าในตาราง 4.8 และตาราง 4.9

ตาราง 4.8 ค่าประมาณหลังของพารามิเตอร์สำหรับตัวแบบเบอร์ที่มีตัวแปรร่วม

Parameter	Mean	Std dev	2.50%	50%	97.50%
Intercept	4.357	0.026	4.311	4.354	4.405
อายุ	-0.062	0.035	-0.140	-0.049	-0.003
เพศ	-0.104	0.041	-0.150	-0.123	-0.012
รูปแบบผลประโยชน์	-0.042	0.034	-0.095	-0.035	0.014
จำนวนเงินเอาประกันภัย	0.001	0.036	-0.065	-0.003	0.059
บริษัท	-0.068	0.071	-0.199	-0.051	0.029
โรคเกี่ยวกับตับ	0.032	0.044	-0.051	0.033	0.103
โรคเลือด	0.017	0.029	-0.046	0.022	0.063
โรคระบบประสาท	0.076	0.027	0.031	0.073	0.125
โรคเกี่ยวกับปอด	0.030	0.030	-0.030	0.042	0.071
โรคความดันเลือด	-0.042	0.034	-0.118	-0.039	0.025
โรคหัวใจ	0.078	0.045	0.001	0.086	0.154
โรคอื่นๆ	0.155	0.039	0.088	0.155	0.224
α	0.896	0.053	0.803	0.892	1.002
τ	2.507	0.107	2.292	2.526	2.655

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตาราง 4.9 ค่าประมาณหลังของพารามิเตอร์สำหรับตัวแบบล็อกนอร์มอลที่มีตัวแปรร่วม

Parameter	Mean	Std dev	2.50%	50%	97.50%
Intercept	4.427	0.026	4.380	4.427	4.477
อายุ	-0.012	0.052	-0.114	-0.012	0.085
เพศ	0.002	0.045	-0.087	0.000	0.092
รูปแบบผลประโยชน์	-0.074	0.048	-0.171	-0.072	0.024
จำนวนเงินเอาประกันภัย	0.067	0.043	-0.021	0.069	0.146
บริษัท	-0.017	0.047	-0.114	-0.015	0.073
โรคเกี่ยวกับตับ	0.030	0.047	-0.062	0.030	0.125
โรคเลือด	0.013	0.034	-0.054	0.014	0.082
โรคระบบประสาท	-0.050	0.056	-0.157	-0.051	0.067
โรคเกี่ยวกับปอด	0.045	0.047	-0.044	0.043	0.137
โรคความดันเลือด	0.028	0.042	-0.055	0.027	0.112
โรคหัวใจ	0.034	0.040	-0.039	0.033	0.116
โรคอื่นๆ	0.019	0.052	-0.081	0.016	0.119
σ^2	0.597	0.029	0.541	0.597	0.661

ใช้ค่า DIC (deviance information criterion) เป็นเกณฑ์ในการตัดสินว่าตัวแบบเบอร์หรือตัวแบบล็อกนอร์มอลเป็นตัวแบบที่เหมาะสมกับข้อมูลระยะเวลาที่ล่าช้าระหว่างเวลาที่ทำการวินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง สำหรับการวิเคราะห์การอยู่รอดแบบเบส์เซียน โดยแสดงค่าในตารางต่อไปนี้

ตาราง 4.10 ค่า DIC ของตัวแบบเบอร์ และตัวแบบล็อกนอร์มอลที่มีตัวแปรร่วม

Model	$\bar{D}$	$\hat{D}$	p_D	DIC
Burr	8553	8109	443.6	8997
LN	9387	9374	13.28	9400

จากตาราง 4.10 พบว่าตัวแบบเบอร์ให้ค่า DIC น้อยกว่าตัวแบบล็อกนอร์มอล ดังนั้นการวิเคราะห์ตัวแบบร่วมกับตัวแปรด้วยตัวแบบเบอร์จึงมีความเหมาะสมกับข้อมูลระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคถึงเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรงมากกว่าตัวแบบล็อกนอร์มอล

จากตาราง 4.7 ซึ่งแสดงค่าประมาณหลังของพารามิเตอร์สำหรับตัวแบบเบอร์จะได้ว่า

i) เนื่องจาก β_1 = -0.062 จะได้ว่า ผู้เอาประกันภัยที่มีอายุมาก จะมีระยะเวลาระหว่างเวลาที่วินิจฉัยโรคถึงเวลาที่ชำระสินไหมทดแทนน้อยกว่าผู้เอาประกันภัยที่มีอายุน้อยกว่า

ii) เนื่องจาก β_2 = -0.104 จะได้ว่า ผู้เอาประกันภัยเพศหญิง (x = 1) จะมีระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคถึงเวลาที่ชำระสินไหมทดแทนน้อยกว่าผู้เอาประกันภัยเพศชาย (x = 0)

iii) เนื่องจาก β_3 = -0.042 จะได้ว่า กรมธรรม์ที่มีจำนวนเงินเอาประกันภัยน้อยจะมีระยะเวลาที่ล้าช้าระหว่างเวลาที่วินิจฉัยโรคถึงเวลาที่ชำระสินไหมทดแทนน้อยกว่ากรมธรรม์ที่มีจำนวนเงินเอาประกันภัยจำนวนมากกว่า

iv) เนื่องจาก β_4 = 0.001 จะได้ว่า รูปแบบผลประโยชน์ประเภทไม่คุ้มครองการเสียชีวิต(x = 1) จะมีระยะเวลาระหว่างเวลาที่วินิจฉัยโรคถึงเวลาที่ชำระสินไหมทดแทนน้อยกว่ารูปแบบผลประโยชน์ประเภทคุ้มครองการเสียชีวิต (x =0)

v) เนื่องจาก β_5 = -0.068 จะได้ว่าผู้เอาประกันภัยที่ทำประกันภัยกับบริษัทที่ 2 (x =1) จะมีระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทน น้อยกว่าผู้เอาประกันภัยที่ทำประกันภัยกับบริษัทที่ (x =0)

สร้างสถานการณ์ (Scenarios) โดยแบ่งเป็น 11 สถานการณ์ เพื่อพยากรณ์ระยะเวลาระหว่างวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง โดยที่สถานการณ์ที่ 1 คือ ผู้เอาประกันภัยเป็นเพศชาย มีรูปแบบผลประโยชน์ประเภทคุ้มครองการเสียชีวิต อายุ 36 ปี จำนวนเงินเอาประกันภัย 100,000 บาท เรียกร้องค่าสินไหมทดแทนเนื่องจากเป็นโรคมะเร็ง และทำประกันชีวิตกับบริษัทที่ 1 จากนั้นทำการเปรียบเทียบระยะเวลาระหว่างวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรงที่พยากรณ์ได้ กับสถานการณ์อื่นๆ โดยทำการเปลี่ยนลักษณะของตัวแปรร่วม จนครบ 11 แบบ ดังแสดงในตาราง 4.11

ตาราง 4.11 สถานการณ์ (Scenarios) เพื่อพยากรณ์การแจกแจงของระยะเวลาระหว่างวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง

สถานการณ์	1	2	3	4	5	6
เพศ	M	F	M	M	M	M
แบบผลประโยชน์	FA	FA	SA	FA	FA	FA
อายุ	36	36	36	29	49	36
เงินผลประโยชน์	100,000	100,000	100,000	100,000	100,000	200,000
สาเหตุ	Blood	Blood	Blood	Blood	Blood	Blood
บริษัท	1	1	1	1	1	1
สถานการณ์	7	8	9	10	11	
เพศ	M	M	M	M	M	
แบบผลประโยชน์	FA	FA	FA	FA	FA	
อายุ	36	36	36	36	40	
เงินผลประโยชน์	300,000	100,000	100,000	100,000	100,000	
สาเหตุ	Blood	Heart	Liver	Lung	Blood	
บริษัท	1	1	1	1	2	

ทำการหาค่าประมาณค่าเฉลี่ยหลัง (posterior mean estimates) ของระยะเวลาระหว่างวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรงในสถานการณ์ต่างๆ โดยที่ ค่าประมาณค่าเฉลี่ยหลัง หาได้จาก

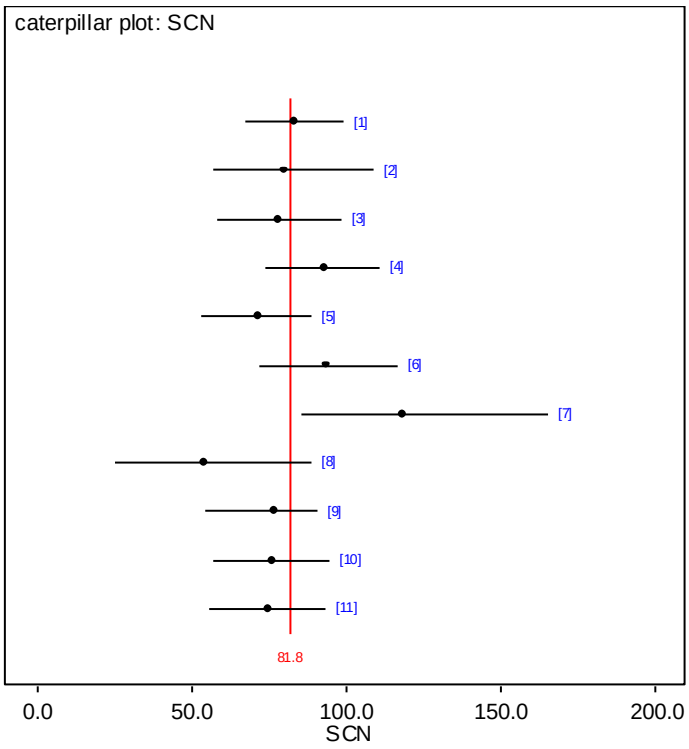
$$E(D_i) = \exp(\eta_i)$$

โดยที่ η_i คือค่าในสมการที่ (3.5)

ตาราง 4.12 ค่าประมาณค่าเฉลี่ยหลัง (posterior mean estimates) ของระยะเวลาระหว่างวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรงในสถานการณ์ในตาราง 4.11 สำหรับตัวแบบเบอร์

Scenario	Burr Model				
	Mean	SD	2.50%	50%	97.50%
1	83.25	8.19	67.60	83.42	99.56
2	80.13	12.90	57.21	79.50	109.20
3	78.00	9.90	58.71	77.25	98.67
4	93.06	9.66	73.86	93.70	111.20
5	71.43	8.91	53.45	70.81	88.92
6	93.66	11.90	72.17	93.08	117.10
7	118.30	20.96	85.49	115.10	165.70
8	53.90	17.06	25.51	54.65	89.30
9	76.77	8.69	54.57	78.24	91.23
10	76.23	9.30	57.21	76.48	94.84
11	74.71	8.38	56.19	74.63	93.43

ค่าประมาณค่าเฉลี่ยหลัง (posterior mean estimates) ของระยะเวลาระหว่างวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรงในสถานการณ์ต่างๆ สำหรับตัวแบบเบอร์ แสดงดังภาพที่ 4.4



ภาพที่ 4.4 ค่าประมาณค่าเฉลี่ยหลัง (posterior mean estimates) ของระยะเวลาระหว่างวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทน สำหรับการประกันภัยโรคร้ายแรงในสถานการณ์ต่างๆ สำหรับตัวแบบเบอร์ โดยเส้นตรงแสดงค่าที่ระดับเปอร์เซ็นไทล์ที่ 2.5 ถึง 97.5 จุดแสดงค่าเฉลี่ยหลัง (posterior means) และเส้นสีแดงแสดงค่าเฉลี่ยของค่าประมาณค่าเฉลี่ยหลังของระยะเวลาระหว่างวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทนใน 11สถานการณ์

จากภาพที่ 4.4 จะเห็นว่าสถานการณ์ที่ 1 และ 2 มีค่าประมาณค่าเฉลี่ยหลังของระยะเวลาที่ล่าช้าระหว่างเวลาที่ทำการวินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนใกล้เคียงกัน นั่นคือตัวแปรร่วมเพศไม่มีผลต่อระยะเวลาที่ล่าช้าระหว่างเวลาที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทน

สถานการณ์ที่ 7 มีค่าประมาณค่าเฉลี่ยหลังของระยะเวลาที่ล่าช้าระหว่างเวลาที่ทำการวินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนมากกว่าสถานการณ์อื่นๆ

4.2.3 การคัดเลือกตัวแบบที่เหมาะสม

ในการคัดเลือกตัวแบบที่เหมาะสมโดยคัดเลือกตัวแปรด้วยวิธีการคัดเลือกตัวแปรของกิบส์ พิจารณาพารามิเตอร์เดิม (prior parameter) สำหรับการแจกแจงของ β เป็นแบบ Zellner 's g – prior โดยใช้ γ เป็นเวกเตอร์ชี้วัด (indicator vector) นั้น เมื่อพารามิเตอร์ γ มีค่าเฉลี่ยน้อยกว่า 0.5 จะนำตัวแปรร่วมดังกล่าวออกจากตัวแบบ และจะนำตัวแปรร่วมที่มีค่าเฉลี่ยของพารามิเตอร์ γ

เท่ากับ 1 เข้าสู่ตัวแบบ ทำให้เกิดตัวแบบที่เหมาะสมกับข้อมูลระยะเวลาระหว่างเวลาที่วินิจฉัยว่าเป็นโรคและเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง โดยค่าเฉลี่ยของพารามิเตอร์ γ ที่ใช้การแจกแจงก่อนเป็นแบบ Zellner' s g สำหรับตัวแบบเบอร์แสดงดังตาราง 4.13

ตาราง 4.13 ค่าเฉลี่ยของพารามิเตอร์ γ ที่ใช้การแจกแจงก่อนเป็นแบบ Zellner' s g สำหรับตัวแบบเบอร์

พารามิเตอร์	Zellner's g-prior Mean	SD
อายุ	1	0
เพศ	0.446	0.498
จำนวนเงินเอาประกันภัย	0.082	0.277
รูปแบบผลประโยชน์	1	0
บริษัท	1	0
สาเหตุ	0.441	0.486

จากตาราง 4.13 พบว่าตัวแปรร่วมที่มีค่าเฉลี่ยของพารามิเตอร์ γ เข้าใกล้ 1 คือตัวแปรร่วม อายุ รูปแบบผลประโยชน์ และบริษัท จึงนำตัวแปรร่วมดังกล่าวเข้าสู่ตัวแบบ

สร้างสถานการณ์ (Scenarios) โดยแบ่งเป็น 4 สถานการณ์ เพื่อพยากรณ์ระยะเวลาระหว่างวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง โดยที่สถานการณ์ที่ 1 คือ ผู้เอาประกันภัยอายุ 36 ปี มีรูปแบบผลประโยชน์ประเภทครอบคลุมการเสียชีวิตและทำประกันภัยกับบริษัทที่1จากนั้นทำการเปรียบเทียบระยะเวลาระหว่างวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรงที่พยากรณ์ได้ กับสถานการณ์อื่นๆ โดยทำการเปลี่ยนลักษณะของตัวแปรร่วม จนครบ 4 แบบ ดังแสดงในตาราง 4.14

ตาราง 4.14 สถานการณ์ (Scenarios) เพื่อพยากรณ์การแจกแจงของระยะเวลาระหว่างวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง สำหรับตัวแบบเบอร์ที่มีตัวแปรร่วมคือ อายุ รูปแบบผลประโยชน์ และบริษัท

Scenarios	1	2	3	4
อายุ	36	49	36	36
แบบผลประโยชน์	Full	Full	SA	Full
บริษัท	1	1	1	2

ทำการหาค่าประมาณค่าเฉลี่ยหลัง (posterior mean estimates) ของระยะเวลาระหว่างวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรงในสถานการณ์ต่างๆ โดยที่ ค่าประมาณค่าเฉลี่ยหลัง หาได้จาก

$$E(D_i) = \exp(\eta_i)$$

โดยที่ η_i คือค่าในสมการที่ (3.5)

ตาราง 4.15 การประมาณหลัง (posterior estimates) ของค่าเฉลี่ยระยะเวลาระหว่างวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรงในสถานการณ์ต่างๆ สำหรับตัวแบบเบอร์

Scenario	BURR Model				
	Mean	SD	2.50%	50%	97.50%
1	80.71	6.33	68.23	79.06	95.06
2	61.94	4.26	55.64	61.56	71.39
3	49.44	4.37	42.70	48.92	58.64
4	74.43	4.59	68.10	74.08	85.52

จะเห็นว่าสถานการณ์ที่ 3 มีค่าประมาณค่าเฉลี่ยหลังของระยะเวลาที่ล่าช้าระหว่างเวลาที่ทำการวินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนน้อยกว่าสถานการณ์อื่นๆ คือ เมื่อพิจารณาผู้เอาประกันภัยที่ทำประกันภัยกับบริษัทที่1 อายุ 36 ปีเท่ากัน พบว่าผู้เอาประกันภัยที่ทำประกันภัยรูปแบบไม่ครอบคลุมการเสียชีวิตจะมีระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนน้อยกว่าผู้เอาประกันภัยที่ทำประกันภัยรูปแบบครอบคลุมการเสียชีวิต

สถานการณ์ที่ 1 มีค่าประมาณค่าเฉลี่ยหลังของระยะเวลาที่ล่าช้าระหว่างเวลาที่ทำการวินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนมากกว่าสถานการณ์ที่ 4 คือ เมื่อพิจารณาผู้เอาประกันภัยอายุ 36 ปี และรูปแบบผลประโยชน์เป็นแบบครอบคลุมการเสียชีวิต จะพบว่าผู้เอาประกันภัยที่ทำประกันภัยกับบริษัทที่ 1 จะมีระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนมากกว่าผู้เอาประกันภัยที่ทำประกันภัยกับบริษัทที่ 2

จากนั้นทำการคัดเลือกตัวแบบที่เหมาะสมโดยใช้ตัวแบบล็อกนอร์มอล ด้วยการคัดเลือกตัวแปรตามวิธีการคัดเลือกตัวแปรของกิบส์ เช่นเดียวกับตัวแบบเบอร์ โดยค่าเฉลี่ยของพารามิเตอร์ γ ที่ใช้การแจกแจงก่อนเป็นแบบ Zellner' s g แสดงในตารางต่อไปนี้

ตาราง 4.16 ค่าเฉลี่ยของพารามิเตอร์ γ ที่ใช้การแจกแจงก่อนเป็นแบบ Zellner' s g สำหรับตัวแบบล็อกนอร์มอล

พารามิเตอร์	Zellner's g-prior Mean	SD
อายุ	1	0
เพศ	0.411	0.492
จำนวนเงินเอาประกันภัย	0.003	0.212
รูปแบบผลประโยชน์	1	0
บริษัท	1	0
สาเหตุ	0.441	0.495

จากตาราง 4.16 พบว่าแต่ละตัวแปรร่วมมีค่าเฉลี่ยของพารามิเตอร์ γ ที่ใช้การแจกแจงก่อนเป็นแบบ Zellner' s g สำหรับตัวแบบล็อกนอร์มอล มีค่าใกล้เคียงกับการแจกแจงแบบเบอร์ ดังนั้นจึงได้ผลสรุปในทำนองเดียวกัน คือ อายุ รูปแบบผลประโยชน์ และบริษัท เป็นตัวแปรร่วมที่ทำให้ได้ตัวแบบที่เหมาะสมกับข้อมูลระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยว่าเป็นโรคและเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง

สร้างสถานการณ์ (Scenarios) โดยแบ่งเป็น 4 สถานการณ์ เพื่อพยากรณ์ระยะเวลาระหว่างวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง เช่นเดียวกับตาราง 4.13

ทำการหาค่าประมาณค่าเฉลี่ยหลัง (posterior mean estimates) ของระยะเวลาระหว่างวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรงในสถานการณ์ต่างๆ โดยที่ ค่าประมาณค่าเฉลี่ยหลัง หาได้จาก

$$E(D_i) = \exp(\eta_i)$$

โดยที่ η_i คือค่าในสมการที่ (3.5)

ตาราง 4.17 การประมาณหลัง (posterior estimates) ของค่าเฉลี่ยระยะเวลาระหว่างวันที่ทำการวินิจฉัยโรคและวันที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรงในสถานการณ์ต่างๆ สำหรับตัวแบบล็อกนอร์มอล

Scenario	Lognormal Model				
	Mean	SD	2.50%	50%	97.50%
1	94.59	4.724	85.87	94.43	104.5
2	89.44	3.087	83.73	89.3	95.76
3	78.86	2.587	73.93	78.78	84.45
4	87.02	2.711	82.03	86.94	92.44

ใช้ค่า DIC เป็นเกณฑ์ในการตัดสินว่าตัวแบบเบอร์ หรือตัวแบบล็อกนอร์มอลเป็นตัวแบบที่เหมาะสมกับข้อมูลระยะเวลาที่ล่าช้าระหว่างเวลาที่ทำการวินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง สำหรับการวิเคราะห์การอยู่รอดแบบเบส์เซียน โดยแสดงค่าในตารางต่อไปนี้

ตาราง 4.18 ค่า DIC ของตัวแบบเบอร์ และตัวแบบล็อกนอร์มอลที่มีตัวแปรร่วมคือ อายุรูปแบบผลประโยชน์ และบริษัท

Model	$\bar{D}$	$\hat{D}$	p_D	DIC
Burr	8474	7981	492.7	8967
LN	9378	9372	5.436	9383

จากตาราง 4.18 พบว่าตัวแบบเบอร์ให้ค่า DIC น้อยกว่าตัวแบบล็อกนอร์มอล ดังนั้นการวิเคราะห์ตัวแบบที่มีตัวแปรร่วม ด้วยตัวแบบเบอร์จึงมีความเหมาะสมกับข้อมูลระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคถึงเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรงมากกว่าตัวแบบล็อกนอร์มอล และได้ค่าประมาณหลังของพารามิเตอร์สำหรับตัวแบบเบอร์ที่มีตัวแปรอายุรูปแบบผลประโยชน์ และบริษัท ดังตาราง 4.19

ตาราง 4.19 ค่าประมาณหลังของพารามิเตอร์สำหรับตัวแบบเบอร์ที่มีตัวแปรอายุ รูปแบบผลประโยชน์ และบริษัท

Parameter	Mean	Std dev	2.50%	50%	97.50%
Intercept	4.322	0.048	4.241	4.317	4.391
อายุ	-0.188	0.032	-0.482	-0.170	0.010
รูปแบบผลประโยชน์	-0.490	0.016	-0.618	-0.510	-0.289
บริษัท	-0.081	0.012	-0.251	-0.074	0.022
α	0.849	0.082	0.729	0.843	1.019
τ	2.594	0.143	2.332	2.659	2.792

จากตาราง 4.19 จะได้ตัวแบบระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง คือ

$$\exp(\eta_i) = \exp(4.322 - 0.188Z_{i1} - 0.490X_{i2} - 0.081X_{i3})$$

เมื่อ Z_{i1} คือค่ามาตรฐานของผู้เอาประกันภัยคนที่ i ภายใต้ตัวแปรอายุ โดยที่ $Z_{i1} = \frac{X_{i1} - \bar{X}}{S.D.}$ เมื่อ X_{i1} คืออายุของผู้เอาประกันภัย $\bar{X}$ คือค่าเฉลี่ยเลขคณิตของอายุผู้เอาประกันภัย มีค่าเท่ากับ 39.378 และ $S.D.$ คือส่วนเบี่ยงเบนมาตรฐานของอายุผู้เอาประกันภัย มีค่าเท่ากับ 9.232 X_{i2} คือรูปแบบผลประโยชน์ผู้เอาประกันภัยคนที่ i โดยที่ $X_{i2} = 0$ เมื่อผู้เอาประกันภัยทำประกันภัยประเภทครอบคลุมการเสียชีวิต และ $X_{i2} = 1$ เมื่อผู้เอาประกันภัยทำประกันภัยประเภทไม่ครอบคลุมการเสียชีวิต X_{i3} คือบริษัทประกันภัยของผู้เอาประกันภัยคนที่ i โดยที่ $X_{i3} = 0$ เมื่อผู้เอาประกันภัยทำประกันภัยกับบริษัทที่ 1 และ $X_{i3} = 1$ เมื่อผู้เอาประกันภัยทำประกันภัยกับบริษัทที่ 2

ตัวอย่างเช่น ผู้เอาประกันภัยอายุ 30 ปี มีรูปแบบผลประโยชน์ประเภทครอบคลุมการเสียชีวิต และทำประกันภัยกับบริษัทที่ 1 สามารถหาระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนได้ด้วย

$$\exp(\eta_i) = \exp\left(4.322 - 0.188\left(\frac{30 - 39.378}{9.232}\right) - 0.490(0) - 0.081(0)\right) = 91.195 \text{ วัน}$$

4.4 การเปรียบเทียบผลการวิเคราะห์แบบไม่ใช้เบส์เซียนและการวิเคราะห์แบบเบส์เซียน

สำหรับการวิเคราะห์แบบไม่ใช้เบส์เซียน โดยใช้ตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อก พบว่าตัวแปรร่วมที่มีนัยสำคัญทางสถิติคือ ตัวแปรอายุ และบริษัท โดยผู้เอาประกันภัยที่มีอายุมากจะมีระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนน้อยกว่าผู้เอาประกันภัยที่มีอายุน้อยกว่า และผู้เอาประกันภัยที่ทำประกันภัยกับบริษัทที่ 1 จะมีระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรงมากกว่าผู้เอาประกันภัยที่ทำประกันภัยกับบริษัทที่ 2 แต่เมื่อพิจารณาจากกราฟที่แสดงค่าความแตกต่างของการอยู่รอดของตัวแปรร่วมแต่ละตัว ดังภาพที่ 4.1 พบว่ามีตัวแปรอื่นที่อาจมีผลต่อตัวแบบ ได้แก่ ตัวแปรรูปแบบผลประโยชน์ พบว่าผู้เอาประกันภัยที่มีรูปแบบผลประโยชน์แบบครอบคลุมการเสียชีวิตจะมีระยะเวลาที่ล่าช้ายาวนานกว่าผู้เอาประกันภัยที่มีรูปแบบผลประโยชน์แบบไม่ครอบคลุมการเสียชีวิต และตัวแปรร่วมจำนวนเงินเอาประกันภัยที่แตกต่างกัน พบว่าผู้เอาประกันภัยที่มีจำนวนเงินเอาประกันภัยน้อยจะมีระยะเวลาที่ล่าช้าน้อยกว่าผู้เอาประกันภัยที่มีจำนวนเงินเอาประกันภัยมากกว่า

ส่วนการวิเคราะห์แบบเบส์เซียนพบว่าตัวแบบเบอร์มีความเหมาะสมกับข้อมูลมากกว่าตัวแบบล็อกนอร์มอล โดยตัวแบบที่ได้ประกอบด้วยตัวแปรอายุ รูปแบบผลประโยชน์ และบริษัท โดยตัวแปรอายุ และบริษัทได้ผลเช่นเดียวกับการวิเคราะห์แบบไม่ใช้เบส์เซียน และการวิเคราะห์ด้วยวิธีนี้สามารถพยากรณ์ระยะเวลาที่ล่าช้าของผู้เอาประกันภัยแต่ละคนได้ โดยพิจารณาตามสถานการณ์ต่างๆ เช่น ผู้เอาประกันภัยเพศชาย อายุ 36 ปี ทำประกันภัยรูปแบบครอบคลุมการเสียชีวิตกับบริษัทประกันภัยที่ 1 เมื่อพิจารณาที่ระดับเปอร์เซ็นไทล์ที่ 2.5 ถึง 97.5 พบว่าจะมีระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนอยู่ในช่วง 68 ถึง 95 วัน ดังนั้น การวิเคราะห์แบบเบส์เซียนด้วยตัวแบบเบอร์จึงเหมาะสมกับข้อมูลที่ใช้ในงานวิจัยนี้เนื่องจากสามารถหาปัจจัยที่มีผลต่อระยะเวลาที่ล่าช้า และพยากรณ์ระยะเวลาที่ล่าช้าเมื่อพิจารณาสถานการณ์ของผู้เอาประกันภัยแต่ละคนได้

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

การวิจัยครั้งนี้ต้องการหาปัจจัยที่มีผลต่อระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง และหาตัวแบบของระยะเวลา ตามการวิเคราะห์แบบไม่ใช้เบส์เซียน และแบบเบส์เซียน โดยการวิเคราะห์แบบไม่ใช้เบส์เซียนจะใช้ตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อก ส่วนการวิเคราะห์แบบเบส์เซียนจะใช้ตัวแบบเบอร์และตัวแบบล็อกนอร์มอล โดยการคัดเลือกตัวแปรใช้วิธีการคัดเลือกตัวแปรของกิบส์ ซึ่งกำหนดพารามิเตอร์ก่อนเป็น Zellner' s g – prior ข้อมูลที่ใช้ในการศึกษาคือข้อมูลระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรงของบริษัทประกันภัย 2 แห่ง และตัวแปรร่วมที่ใช้ในการศึกษาคือ อายุของผู้เอาประกันภัย ณ วันที่เรียกร้องสินไหมทดแทน มีหน่วยเป็นปี เพศของผู้เอาประกันภัย รูปแบบผลประโยชน์ โดยแบ่งเป็นผลประโยชน์ที่ครอบคลุมการเสียชีวิต และผลประโยชน์ที่ไม่ครอบคลุมการเสียชีวิต จำนวนเงินเอาประกันภัย บริษัทประกันภัย 2 บริษัท สาเหตุของการเรียกร้องค่าสินไหมทดแทน ในงานวิจัยนี้ คือโรคที่เกิดขึ้นกับผู้เอาประกันภัย แบ่งเป็นกลุ่มโรคร้ายแรง 8 กลุ่ม คือ กลุ่มโรคมะเร็ง กลุ่มโรคเกี่ยวกับตับ กลุ่มโรคเลือด กลุ่มโรคเกี่ยวกับระบบประสาท กลุ่มโรคปอดอักเสบ กลุ่มโรคความดันเลือดสูง และโรคเส้นเลือดในสมอง กลุ่มโรคหัวใจ และกลุ่มโรคอื่นๆ

ผลการศึกษาสรุปได้ดังนี้

สำหรับการวิเคราะห์แบบไม่ใช้เบส์เซียนด้วยตัวแบบการเสี่ยงภัยแบบอัตราส่วนค็อก พบว่าตัวแปรร่วมที่มีนัยสำคัญทางสถิติคือ ตัวแปรอายุ และบริษัทประกันภัย โดยพบว่าผู้เอาประกันภัยที่มีอายุมากจะมีระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนน้อยกว่าผู้เอาประกันภัยที่มีอายุน้อยกว่า และผู้เอาประกันภัยที่ทำประกันภัยกับบริษัทที่ 1 จะมีระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนมากกว่าผู้เอาประกันภัยที่ทำประกันภัยกับบริษัทที่ 2 แต่เมื่อพิจารณาจากกราฟที่แสดงค่าความแตกต่างของการอยู่รอดของตัวแปรร่วมแต่ละตัว พบว่ามีตัวแปรอื่นที่อาจมีผลต่อตัวแบบ ได้แก่ ตัวแปรร่วมรูปแบบผลประโยชน์ พบว่าผู้เอาประกันภัยที่มีรูปแบบผลประโยชน์แบบครอบคลุมการเสียชีวิตจะมีระยะเวลาที่ล่าช้ายาวนานกว่าผู้เอาประกันภัยที่มีรูปแบบผลประโยชน์แบบไม่ครอบคลุมการเสียชีวิต และตัวแปรจำนวนเงินเอาประกันภัยที่แตกต่าง

กัน พบว่าผู้เอาประกันภัยที่มีจำนวนเงินเอาประกันภัยน้อยจะมีระยะเวลาที่ล่าช้าน้อยกว่าผู้เอาประกันภัยที่มีจำนวนเงินเอาประกันภัยมากกว่า

การวิเคราะห์แบบเบส์เซียนที่ใช้ตัวแบบเบอร์ และตัวแบบล็อกนอร์มอล พบว่าตัวแบบเบอร์มีความเหมาะสมกับข้อมูลมากกว่าตัวแบบล็อกนอร์มอล โดยตัวแบบที่ได้ประกอบด้วยตัวแปร อายุ รูปแบบผลประโยชน์ และบริษัท โดยได้ผลเช่นเดียวกับการวิเคราะห์แบบไม่ใช้เบส์เซียน และการวิเคราะห์ด้วยวิธีนี้สามารถพยากรณ์ระยะเวลาที่ล่าช้าของผู้เอาประกันภัยแต่ละคนได้ โดยพิจารณาตามสถานการณ์ต่างๆ

5.2 อภิปรายผล

จากงานวิจัยของ Ozkok (2011) ที่ใช้ข้อมูลระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรงของบริษัทประกันภัยในประเทศอังกฤษระหว่างปี ค.ศ. 1999 ถึง 2005 พบว่าตัวแบบเบอร์มีความเหมาะสมกับข้อมูลมากกว่าตัวแบบล็อกนอร์มอล และตัวแปรร่วมที่มีผลต่อระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทน คือรูปแบบผลประโยชน์ จำนวนเงินเอาประกันภัย สาเหตุการเรียกร้องค่าสินไหมทดแทน และบริษัทประกันภัย โดยตัวแปรที่ให้ผลเช่นเดียวกับงานวิจัยนี้ คือตัวแปรรูปแบบผลประโยชน์ โดยผู้เอาประกันภัยที่มีรูปแบบผลประโยชน์ประเภทครอบคลุมการเสียชีวิต จะมีระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนมากกว่าผู้เอาประกันภัยที่มีรูปแบบผลประโยชน์ประเภทไม่ครอบคลุมการเสียชีวิต และบริษัทประกันภัยที่ต่างกันก็มีระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนแตกต่างกันด้วย

รูปแบบผลประโยชน์ประเภทครอบคลุมการเสียชีวิตมีระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนมากกว่ารูปแบบผลประโยชน์ประเภทไม่ครอบคลุมการเสียชีวิต เนื่องจากทางบริษัทประกันภัยอาจต้องใช้เวลาในการพิสูจน์ว่าผู้เอาประกันภัยที่ทำประกันภัยในรูปแบบผลประโยชน์ประเภทครอบคลุมการเสียชีวิตนั้น เสียชีวิตเนื่องจากสาเหตุใด คือได้เสียชีวิตเนื่องจากโรคร้ายแรงตามที่กรมธรรม์ระบุไว้หรือไม่ และต้องทำการพิสูจน์ว่าเสียชีวิตด้วยการเจ็บป่วยเนื่องจากโรคที่ผู้เอาประกันภัยได้เป็นมาก่อนวันที่บริษัทประกันภัยตกลงรับทำสัญญาหรือวันต่ออายุเนื่องจากสัญญาขาดผลบังคับครั้งสุดท้ายหรือไม่ หรือเป็นในช่วงเวลาที่บริษัทกำหนดไว้ในเงื่อนไขกรรมธรรม์ว่าจะชำระค่าสินไหมทดแทนให้หรือไม่ เช่นบางบริษัทประกันภัยกำหนดว่าจะชำระค่าสินไหมทดแทนให้กับผู้เอาประกันภัยก็ต่อเมื่อผู้เอาประกันภัยได้เจ็บป่วยหรือเสียชีวิตเนื่องจากโรคร้ายแรง ภายหลังจาก 90 วัน นับตั้งแต่วันที่บริษัทประกันภัยตกลงรับทำสัญญา หรือนับตั้งแต่วันที่อนุมัติให้ต่ออายุสัญญาครั้งสุดท้าย

บริษัทประกันภัยที่ต่างกันจะมีระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนต่างกัน เนื่องจากรูปแบบสัญญาการประกันภัยที่ต่างกัน มีข้อกำหนดในการจ่ายค่าสินไหมทดแทนที่ต่างกัน บางบริษัทประกันภัยอาจมีรูปแบบสัญญาการประกันภัยที่มีข้อกำหนดมาก จึงต้องใช้ระยะเวลานานในการพิสูจน์ว่าสมควรจ่ายค่าสินไหมทดแทนให้แก่ผู้เอาประกันภัยหรือไม่

จากการวิเคราะห์แบบเบส์เซียนที่ใช้ตัวแบบเบอร์ พบว่าตัวแปรร่วมที่มีผลต่อระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรคและเวลาที่ชำระสินไหมทดแทนสำหรับการประกันภัยโรคร้ายแรง คือ ตัวแปรร่วมอายุ รูปแบบผลประโยชน์ และบริษัทประกันภัย จึงทำให้บริษัทประกันภัยได้ตัวแบบทางสถิติเพื่อใช้พยากรณ์ข้อมูลเวลาที่ขาดหายไปของผู้เอาประกันภัยบางราย โดยสามารถพยากรณ์ระยะเวลาที่ล่าช้าระหว่างเวลาที่วินิจฉัยโรค และเวลาที่ชำระสินไหมทดแทนที่ระดับเปอร์เซ็นไทล์ที่ 2.5 ถึง 97.5 เพื่อใช้เป็นข้อมูลพื้นฐานให้บริษัทประกันภัยได้เตรียมเงินสำรองให้พร้อมภายในช่วงเวลาดังกล่าว และทำให้คงเงินสำรองสำหรับการประกันภัยโรคร้ายแรงไว้ได้

5.3 ข้อเสนอแนะ

1. การวิเคราะห์แบบเบส์เซียนที่ใช้ในงานวิจัยครั้งนี้ได้กำหนดค่าเริ่มต้นให้กับพารามิเตอร์ที่ไม่ทราบค่าให้กับตัวแบบเบอร์ และตัวแบบล็อกนอร์มอล และกำหนดค่าความน่าจะเป็นก่อนดังนี้

   $\beta_1,...,\beta_p$ มีการแจกแจงแบบปกติ ด้วยพารามิเตอร์ 0 และ 0.001

   สำหรับผู้ที่สนใจนำไปประยุกต์ควรกำหนดค่าพารามิเตอร์และค่าความน่าจะเป็นก่อนที่แตกต่าง หรือหาวิธีการกำหนดรูปแบบของการกำหนดค่าพารามิเตอร์ของการแจกแจงความน่าจะเป็นก่อนที่เหมาะสมกับข้อมูล
2. ควรมีการศึกษาการวิเคราะห์การอยู่รอดแบบเบส์เซียนเพิ่มเติม โดยเปรียบเทียบตัวแบบของระยะเวลาระหว่างวันที่วินิจฉัยโรคและวันที่ชำระสินไหมทดแทน เมื่อใช้ข้อมูลระยะเวลาที่ไม่รวมข้อมูลเวลาที่ขาดหายไป และข้อมูลระยะเวลาที่รวมข้อมูลระยะเวลาที่ขาดหายไป จากนั้นหาค่าการแจกแจงหลัง (posterior distribution) ของพารามิเตอร์ θ จากค่าสังเกตที่มีอยู่ โดยให้ตัวแปร $D=(D_{obs},D_{mis})$ เมื่อ D_{obs} แทนค่าสังเกตที่มีอยู่ และ D_{mis} แทนค่าสังเกตที่หายไป ซึ่งฟังก์ชันความหนาแน่นน่าจะเป็น ของค่าสังเกตที่หายไปเป็น $P(D_{mis}\mid D_{obs})=\int P(D_{mis},\theta\mid D_{obs})d\theta$ โดยความหนาแน่นของค่าที่หายไปสามารถประมาณค่าได้ด้วยวิธีมาร์คอฟเชนมอนติคาร์โล

ควรมีการนำตัวแบบระยะเวลาที่ล้าช้าไปใช้ในทางประกันวินาศภัย โดยสนใจเวลาที่ล่าช้าระหว่างเวลาที่เกิดอุบัติเหตุและเวลาที่ชำระสินไหมทดแทน

รายการอ้างอิง

Antonio K., & Beirlant J. (2008). Issue in claim reserving and credibility: a semiparametric approach with mixed models. *The Journal of Risk and Insurance, 75*, 643-676.

Beirlant J., & Goegebeur Y. (2003). Regression with response distribution of Pareto type Computational Statistic and Data Analysis 595-619.

Beirlant J., Goegebeur Y., & Verlaak R. (1998). Burr regression and portfolio segmentation .Insurance:Mathematics and Economics. *23*, 231-250.

Beirlant J., & Guillou A. (2001). Pareto index estimation under moderate right censoring. Scandinavian Actuarial *2*, 111-125.

Dellaportas P., Forster J.J., & Ntzoufras I. (2002). On Bayesian model and variable selection using MCMC. *Statistics and Computing, 12*, 27-36.

Gelman A., Carlin J.B., Stern H.S., & Rubin D.B. (2000). *Bayesian Data Analysis*.

Gilks W.R., Richardson S., & Spiegelhalter D.J. (1996). Markov Chain Monte Carlo in Practice. *Chapman & Hall*.

Kaplan EL., & Meier P. (1958). J Am Stat Assoc. 457-481.

Kass R.E., & Wasserman L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association, 90*, 928-934.

Klein., & Moeschberger. (2003). Survival Analysis Techniques for Censored and Truncated Data.

Ntzoufras I. (2002). Gibbs variable selection using BUGS. *Journal of Statistical Software, 7*.

Ntzoufras I. (2009). Bayesian Modeling Using WinBUGS. *John Wiley & Sons Inc. USA ISBN: 978-0-470-14114-4*.

Ntzoufras I., Dellaportas P., & Forster J.J. . (2003). Bayesian variable and link determination for generalized linear models. *Journal of Statistical Planning and Inference, 111*, 165-180.

Ozkok. (2011). A Stochastic model for Critical Illness Insurance.

Xin-Yuan Song, & Sik-Yum Lee. (2002). A Bayesian model selection method with applications. *Department of Statistics, Chinese University of Hong Kong*.

Zellner A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior Distribution. *North – Holland, Amsterdam*, 233-243.

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวภรณ์ทิวา พลทิแสง เกิดวันที่ 3 กุมภาพันธ์ พ.ศ. 2532 สำเร็จการศึกษาปริญญาวิทยาศาสตร์บัณฑิต (วท.บ.) สาขาวิชาคณิตศาสตร์ ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2553 และเข้าศึกษาต่อในหลักสูตรวิทยาศาสตร์มหาบัณฑิต (วท.ม.) สาขาการประกันภัย ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชีจุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2554

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University