บทที่ 1

บทนำ

1. ที่มาและความสำคัญของปัญหา

ธุรกิจประกันภัย เป็นธุรกิจที่รับเสี่ยงภัยจากผู้เอาประกันภัย และเป็นเครื่องมือกระจายความสูญเสีย ซึ่งเกิดขึ้นจากเหตุการณ์ที่ไม่อาจคาดได้ โดยบริษัทประกันภัยเรียกเก็บเงินจำนวนหนึ่ง ซึ่งเรียกว่าเบี้ยประกันภัยจากผู้เอาประกันภัยเป็นค่าตอบแทนซึ่งเป็นหน้าที่ของนักคณิตศาสตร์ประกันภัยของบริษัท ที่จะต้องคำนวณเบี้ยประกันภัยให้เพียงพอสำหรับความคุ้มครองที่บริษัทให้ความคุ้มครองแก่ผู้เอาประกันภัย และความคุ้มครองนั้นจะต้องเหมาะสมกับความสูญเสียที่จะเกิดขึ้นจริง จำนวนความสูญเสียอาจจะมีความสัมพันธ์กับตัวแปรอื่น ๆ เช่นการประกันภัยรถยนต์ จำนวนเงินความสูญเสียจะขึ้นอยู่กับ ขนาดซีซีรถ ประเภทรถ และจำนวนที่นั่ง หรือการประกันภัยสุขภาพ จำนวนเงินที่จ่ายเนื่องจากค่ารักษาพยาบาลจะขึ้นอยู่กับ อายุ เพศ และความเสี่ยงของอาชีพของผู้เอาประกันภัย เป็นต้น ซึ่งตามหลักการของการประกันภัยแล้ว จำนวนเงินที่ผู้รับประกันภัยจะจ่ายชดเชยให้แก่ผู้เอาประกันภัยจะไม่เกินความสูญเสียที่เกิดขึ้นจริงและปกติกรมธรรม์ประกันภัยจะกำหนดจำนวนเงินรับผิดชอบสูงสุดต่อกรมธรรม์ไว้ เช่นไม่เกิน Tc บาท ดังนั้นถ้า Y เป็นจำนวนความสูญเสียจริงที่มีค่ามากกว่าจำนวนเงินสูงสุดที่บริษัทรับผิดชอบตามกรมธรรม์ คือ $Y > Tc$ บริษัทจะจ่ายชดเชยจำนวน Tc บาทเท่านั้น และบันทึกการจ่ายเงินชดเชยความสูญเสียเท่ากับ Tc บาท ดังนั้นข้อมูลความสูญเสียของบริษัทประกันภัยจึงไม่ใช่ข้อมูลที่แท้จริงทั้งหมด บางส่วนถูกตัดทิ้งทำให้ข้อมูลของบริษัทประกันภัยต่ำกว่าค่าที่แท้จริง และเนื่องจากข้อมูลที่ถูกตัดทิ้งมีความสำคัญสามารถนำมาใช้วิเคราะห์เพื่อประโยชน์ในการกำหนดความคุ้มครองและเบี้ยประกันภัยของธุรกิจประกันภัยให้มีความเหมาะสม และเพื่อให้การวางแผนการจัดการทางด้านการเงินของบริษัทมีประสิทธิภาพมากยิ่งขึ้น จึงควรทำการประมาณค่าที่ถูกตัดทิ้ง โดยใช้วิธีการที่มีความเหมาะสม เพื่อให้ได้ค่าประมาณที่ใกล้เคียงกับค่าจริงและเพื่อให้มีจำนวนข้อมูลที่เพียงพอต่อการนำไปศึกษาตามวัตถุประสงค์ ดังนั้นเมื่อทำการประมาณค่าความสูญเสีย Y ที่มากกว่า Tc จะทำให้มีข้อมูลค่าความสูญเสียที่ได้จากการประมาณของค่าที่ถูกตัดทิ้ง ข้อมูลที่ได้เหล่านี้รวมกับข้อมูลที่ไม่ถูกตัดทิ้ง จะสามารถนำมาศึกษาหารูปแบบการแจกแจงของค่าความสูญเสีย (Loss Distribution)

และอัตราส่วนค่าความสูญเสีย (Loss Ratio) ซึ่งจะเป็นประโยชน์ในการสร้างตารางอัตราเบี้ยประกันภัย การกำหนดวงเงินความคุ้มครอง และการกำหนดค่าความรับผิดส่วนแรก (Retention) ของบริษัทประกันภัย

จากปัญหาดังกล่าวจึงทำให้ผู้วิจัยสนใจที่จะศึกษาการประมาณค่าความสูญเสีย เมื่อความสูญเสียบางส่วนถูกตัดทิ้งทางขวาประเภทที่ 1 ด้วยสมการถดถอยเชิงเส้นพหุ (Multiple Linear Regression) โดยพิจารณาให้ความสูญเสียที่เกิดขึ้นในแต่ละครั้งของผู้เอาประกันภัย เป็นตัวแปรตาม และให้ปัจจัยที่มีความสัมพันธ์กับความสูญเสียเป็นตัวแปรอิสระ ในการวิจัยครั้งนี้ผู้วิจัยพิจารณากรณีที่มีปัจจัยที่มีความสัมพันธ์กับตัวแปรตาม 2 ปัจจัย สำหรับการประมาณเพารามิเตอร์ $\beta_0, \beta_1, \beta_2$ จากสมการถดถอยเชิงเส้นพหุโดยใช้วิธีวิเคราะห์กับข้อมูลที่ค่าสังเกตของตัวแปรตามมีค่าถูกตัดทิ้งซึ่งการนำข้อมูลที่มีค่าถูกตัดทิ้งมาใช้ประมาณค่าพารามิเตอร์นั้นอาจทำได้ในลักษณะดังต่อไปนี้

กรณีแรก จะพิจารณาให้ค่าที่ถูกตัดทิ้งแทนด้วยค่าที่ไม่ถูกตัดทิ้ง แล้วจึงทำการวิเคราะห์จากข้อมูลทั้งหมด

กรณีที่สอง ไม่พิจารณาข้อมูลส่วนที่เป็นค่าถูกตัดทิ้ง นั่นคือ จะทำการวิเคราะห์ข้อมูลเฉพาะกับข้อมูลที่ไม่ถูกตัดทิ้งเท่านั้น ดังนั้นจำนวนข้อมูลที่นำมาวิเคราะห์ จึงมีขนาดตัวอย่างที่น้อยกว่ากรณีแรก

จากทั้งสองกรณีดังกล่าว เมื่อประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองต่ำสุด จะทำให้ได้ค่าประมาณที่เอนเอียง และโดยเฉลี่ยการประมาณค่าจะต่ำกว่าความเป็นจริง หรือจะทำให้ได้ช่วงความเชื่อมั่นที่แคบกว่าความเป็นจริง

ด้วยเหตุนี้จึงได้มีผู้คิดวิธีการในการประมาณค่าพารามิเตอร์เมื่อข้อมูลถูกตัดทิ้งในตัวแปรตามขึ้นคือ

มิลเลอร์ (Rupet G. Miller) ได้ทำการประมาณค่าพารามิเตอร์โดยใช้ Weight Least Squares เฉพาะกับข้อมูลที่ไม่ขาดหายและทำการวนซ้ำ (Iterative) กับค่าพารามิเตอร์ที่ประมาณจนกว่าค่าประมาณพารามิเตอร์นั้นจะคงตัว (Converge) สำหรับบัคเลย์และเจมส์ (Jonathan Buckley and Ian James) ทำการประมาณพารามิเตอร์ โดยทำการประมาณค่าข้อมูลที่ถูกตัดทิ้งขึ้นมาแล้วทำการวิเคราะห์ ข้อมูลรวมทั้งหมดซึ่งวิธีการของบัคเลย์และเจมส์จะให้ตัวประมาณที่ไม่เอนเอียงในขณะที่วิธีการของมิลเลอร์จะให้ตัวประมาณที่เกือบจะไม่เอนเอียง (nearly unbiased)

เมอร์เรย์ ไอท์เคน (Murray Aitkin) ทำการประมาณค่าพารามิเตอร์โดยวิธีภาวะน่าจะเป็นสูงสุด และทำการประมาณค่าข้อมูลที่ถูกตัดทิ้งด้วยขั้นตอนวิธีอีเอ็ม แล้ววนทำซ้ำโดยทำการวิเคราะห์จากข้อมูลรวมทั้งหมดจนกว่าพารามิเตอร์จะคงตัว

แชตเทอร์จีและแมคลีช (S. Chatterjee and D.L. McLeish) ได้ทำการประมาณเพารามิเตอร์ด้วยวิธีกำลังสองต่ำสุด แล้วประมาณค่าข้อมูลที่ถูกตัดทิ้งซึ่งจะมีค่ามากกว่าค่าที่ถูกตัดทิ้ง

แล้ววนทำซ้ำจากข้อมูลรวมทั้งหมดจนพารามิเตอร์คงตัว ซึ่งวิธีการของแชตเทอร์จีและแมคลีชเป็นวิธีนอนพาราเมตริกซ์

จงดี โรจนประศาสน์ ได้ทำการศึกษาเปรียบเทียบวิธีกำลังสองต่ำสุด วิธีการของบัคเลย์และเจมส์ และวิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุดของสมการถดถอยเชิงเส้นอย่างง่าย จากการศึกษาพบว่า วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุดจะให้ค่าความคลาดเคลื่อน RMSE ของการประมาณค่าตัวแปรตามต่ำกว่าวิธีบัคเลย์และเจมส์ และวิธีกำลังสองต่ำสุด

ดังนั้นเพื่อให้ค่าประมาณใกล้เคียงกับค่าจริง การวิจัยครั้งนี้จึงศึกษาภายใต้การประมาณค่าพารามิเตอร์ $\beta_0, \beta_1, \beta_2$ ด้วยวิธี

1. วิธีการของบัคเลย์และเจมส์ (Buckley and James Method)
2. วิธีการของแชตเทอร์จีและแมคลีช (S. Chatterjee and D.L. McLeish)
3. วิธีกำลังสองต่ำสุดสามัญ (Odinary Least Squares Method) ซึ่งจากนี้จะขอใช้คำสั้นๆ ว่า วิธีกำลังสองต่ำสุด
4. วิธีการภาวะน่าจะเป็นสูงสุดด้วยขั้นตอนวิธีอีเอ็ม (Maximum Likelihood Method via EM Algorithm)

2. วัตถุประสงค์ของการวิจัย

เพื่อทำการเปรียบเทียบการประมาณค่าตัวแปรตาม ในสมการถดถอยเชิงเส้นพหุ เมื่อตัวแปรตามบางค่าถูกตัดทิ้งทางขวาประเภทที่ 1 โดยการประมาณพารามิเตอร์ในสมการถดถอยด้วยวิธีการดังนี้

1. วิธีการของบัคเลย์และเจมส์
2. วิธีการของแชตเทอร์จีและแมคลีช
3. วิธีกำลังสองต่ำสุด
4. วิธีการภาวะน่าจะเป็นสูงสุดด้วยขั้นตอนวิธีอีเอ็ม

3. สมมติฐานการวิจัย

3.1 จากการศึกษาวิจัยครั้งนี้ วิธีการประมาณพารามิเตอร์ด้วยภาวะน่าจะเป็นสูงสุดด้วยขั้นตอนวิธีอีเอ็ม เป็นวิธีที่ประมาณค่าตัวแปรตามได้ใกล้เคียงกับค่าจริงมากกว่าทุกวิธีที่ทำการศึกษาในครั้งนี้

3.2 ภายใต้ลักษณะการแจกแจงของความคลาดเคลื่อนที่มีรูปแบบต่างกันจะมีผลทำให้ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (The Square Root of Mean Square Error : RMSE) แตกต่างกัน

4. ข้อตกลงเบื้องต้น

4.1 การศึกษาภายใต้รูปแบบสมการถดถอยเชิงเส้นพหุ มีรูปแบบดังนี้

$T_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \quad ; \ i = 1,2,3,...,N$

T_i เป็นตัวแปรตาม

x_{ij} เป็นตัวแปรอิสระ $j=1,2$

ε_i เป็นค่าคลาดเคลื่อนสุ่ม

N เป็นขนาดตัวอย่างทั้งหมด

4.2 การแจกแจงของค่าที่ถูกตัดทิ้ง และไม่ถูกตัดทิ้ง เป็นอิสระต่อกัน

4.3 ตัวแปรตาม (Dependent Variable) เท่านั้นเป็นค่าที่ถูกตัดทิ้ง และเป็นการตัดทิ้งทางขวาประเภทที่ 1

4.4 ค่าสังเกตที่สนใจศึกษา คือ

$$Y_i = \begin{cases} T_i & \text{เมื่อ } T_i \le Tc \quad ; i = 1,2,...,n \\ Tc & \text{เมื่อ } T_i > Tc \quad ; i = n+1,...,n+m \end{cases}$$

เมื่อ T_i เป็นค่าสังเกตที่ไม่ถูกตัดทิ้ง

Tc เป็นค่าสูงสุด

n เป็นจำนวนตัวอย่างที่ไม่ถูกตัดทิ้ง

m เป็นจำนวนตัวอย่างที่ถูกตัดทิ้ง

5. ขอบเขตของการวิจัย

5.1 ในการวิจัยครั้งนี้ ทำการศึกษาเปรียบเทียบการประมาณค่าตัวแปรตามในสมการถดถอยเชิงเส้นพหุ เมื่อค่าสังเกตของตัวแปรตามบางส่วนถูกตัดทิ้งทางขวาประเภทที่ 1 โดยการประมาณพารามิเตอร์ด้วยวิธี

5.1.1 วิธีการของบัคเลย์และเจมส์ (Buckley and James Method)

5.1.2 วิธีการของแชตเทอร์จีและแมคลีช (S.Chatterjee and D.L. McLeish)

5.1.3 วิธีกำลังสองต่ำสุด (Ordinary Least Squares Method)

5.1.4 วิธีการภาวะน่าจะเป็นสูงสุดด้วยขั้นตอนวิธีอีเอ็ม (Maximum Likelihood Method via EM Algorithm)

5.2 ในการศึกษาครั้งนี้ศึกษาภายใต้ตัวแปรอิสระจำนวน 2 ตัวแปรซึ่งจำลองตัวแปรอิสระ X_{i1} และ X_{i2} ดังนี้

5.2.1 จำลอง X_{i1} ด้วยการแจกแจงไวบูลล์ (Weibull Distribution) ซึ่งมีฟังก์ชันความหนาแน่นดังนี้

$$f(x) = \begin{cases} \alpha\beta^{-\alpha}x^{\alpha-1}\exp\left(-(x/\beta)^{\alpha}\right) & ; x>0, \alpha>0, \beta>0 \\ 0 & ; \text{อื่นๆ} \end{cases}$$

ในการวิจัยครั้งนี้จะศึกษาที่ $\alpha = 2$ และ $\beta = 2$

5.2.1 จำลอง X_{i2} ด้วยการแจกแจงปกติ (Normal Distribution) ซึ่งมีฟังก์ชันความหนาแน่นดังนี้

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-(x-\mu)^2/2\sigma^2\right) \quad ; \quad -\infty < x < \infty$$

ในการวิจัยครั้งนี้จะศึกษาที่ $\mu = 10$ และ $\sigma^2 = 12$

5.3 ศึกษาภายใต้ค่าคลาดเคลื่อน ε แจกแจง 3 แบบ และภายใต้การแจกแจงของค่าคลาดเคลื่อนแต่ละแบบกำหนดค่าสูงสุดของข้อมูลที่ถูกตัดทิ้ง 3 ระดับ โดยกำหนดจุดตัดระดับที่ 1 ที่ $\mu_T + \sigma_T$ ระดับที่ 2 ที่ $\mu_T + 1.5\sigma_T$ ระดับที่ 3 ที่ $\mu_T + 2\sigma_T$ เมื่อ μ_T เป็นค่าเฉลี่ยของ T และ σ_T เป็นส่วนเบี่ยงเบนมาตรฐานของ T (Standard Deviation) ซึ่งภายใต้การวิจัยครั้งนี้กำหนดการแจกแจงของค่าคลาดเคลื่อนและค่าสูงสุดที่ถูกตัดทิ้งเป็นดังนี้

5.3.1 การแจกแจงปกติ ซึ่งมีฟังก์ชันความหนาแน่นดังนี้

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-(x-\mu)^2/2\sigma^2\right] \quad ; \quad -\infty < x < \infty$$

ในการวิจัยครั้งนี้กำหนด $\mu = 0$, $\sigma^2 = 100$ และกำหนดค่าสูงสุดของข้อมูลถูกตัดทิ้งที่ $\mu_T + \sigma_T$, $\mu_T + 1.5\sigma_T$ และ $\mu_T + 2\sigma_T$ เท่ากับ 53, 58 และ 63 ตามลำดับ

5.3.2 การแจกแจงแบบดับเบิลเอกซโพเนนเชียล (Double Exponential Distribution) ซึ่งมีฟังก์ชันความหนาแน่นดังนี้

$$f(x) = \frac{1}{2\sigma}\exp\left(-|x-\mu|/\sigma\right) \quad ; \quad -\infty < x < \infty$$

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ในการวิจัยครั้งนี้กำหนด $\mu = 0$, $\sigma = \sqrt{50}$ และกำหนดค่าสูงสุดของข้อมูลถูกตัดทิ้งที่ $\mu_T + \sigma_T$, $\mu_T + 1.5\sigma_T$ และ $\mu_T + 2\sigma_T$ เท่ากับ 53, 58 และ 63 ตามลำดับ

5.3.3 การแจกแจงแบบล็อกนอร์มอล (Lognormal Distribution) ซึ่งมีฟังก์ชันความหนาแน่นดังนี้

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right] ; x > 0, -\infty < \mu < \infty, \sigma > 0$$

ในการวิจัยครั้งนี้กำหนด $\mu = 1$, $\sigma = 1.1025$ และกำหนดค่าสูงสุดของข้อมูลถูกตัดทิ้งที่ $\mu_T + \sigma_T$, $\mu_T + 1.5\sigma_T$ และ $\mu_T + 2\sigma_T$ เท่ากับ 55, 58 และ 61 ตามลำดับ

5.5 ศึกษาเมื่อเปอร์เซ็นต์การถูกตัดทิ้งของข้อมูลเป็น 10 %, 20%, 30% และ 40% ของขนาดตัวอย่าง

5.6 ศึกษาเมื่อขนาดตัวอย่างเป็น 20, 30, 40, 50, 60 และ 70

5.7 การวิจัยครั้งนี้ จำลองสถานการณ์ตามที่ต้องการศึกษา โดยใช้เทคนิคมอนติคาร์โล (Monte Carlo Simulation Technique) จากเครื่องคอมพิวเตอร์ AMDAHL 5860 เขียนโปรแกรมภาษาฟอร์แทรน (FORTRAN 77) ทำการจำลองข้อมูลซ้ำ ๆ กัน 1000 ครั้ง ในแต่ละสถานการณ์

6. เกณฑ์การตัดสินใจ

เกณฑ์การพิจารณาว่า วิธีการประมาณค่าพารามิเตอร์ในสมการถดถอยเชิงเส้นพหุด้วยวิธีใดจะให้ค่าการประมาณตัวแปรตามที่ดีกว่า จะพิจารณาโดยการเปรียบเทียบค่าประมาณของตัวแปรตามกับค่าจริงก่อนถูกตัดทิ้งของตัวแปรตาม ด้วยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (The Square Root of Mean Square Error : RMSE) วิธีใดให้ RMSE ต่ำกว่าเป็นวิธีที่ดีกว่า

7. ประโยชน์ที่คาดว่าจะได้รับ

ผลจากการวิจัยจะเป็นแนวทางให้นักคณิตศาสตร์ประกันภัย และผู้ที่สนใจ สามารถนำไปประยุกต์ใช้ เพื่อเลือกวิธีการประมาณค่าความถูกเสีย โดยอาศัยปัจจัยต่าง ๆ ที่มีผลต่อความถูกเสียได้อย่างมีประสิทธิภาพ และเหมาะสมกับข้อมูลของการประกันภัย ซึ่งจะทำให้การกำหนดความคุ้มครองมีความเหมาะสมยิ่งขึ้น และเป็นแนวทางสำหรับผู้ที่สนใจจะทำการศึกษาวิจัยในหัวข้ออื่น ๆ ที่เกี่ยวข้องต่อไป