บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ธุรกิจประกันภัยเป็นธุรกิจที่รับเสี่ยงภัยจากผู้เอาประกันภัย โดยบริษัทประกันภัยเรียกเก็บเงินจำนวนหนึ่ง ซึ่งเรียกว่าเบี้ยประกันภัยจากผู้เอาประกันภัยซึ่งเบี้ยประกันภัยที่เรียกเก็บนี้จะมีความแตกต่างกัน โดยส่วนมากขึ้นกับความเสี่ยงภัยและผลประโยชน์ที่จะได้รับจากการประกันภัยนั้นๆ ในบริษัทประกันภัยจะมีนักคณิตศาสตร์ประกันภัยเป็นผู้ที่ทำหน้าที่คำนวณอัตราเบี้ยประกันภัย การคำนวณเงินสำรอง การคำนวณมูลค่ากรมธรรม์ และอื่นๆ

การคำนวณอัตราเบี้ยประกันภัย สิ่งที่ควรนำมาพิจารณา คือ

1. ความเพียงพอ (Adequacy) เบี้ยประกันภัยจะต้องมีความเพียงพอที่จะจ่ายชดเชยความเสียหายที่เกิดขึ้นตามความคุ้มครองที่ระบุไว้ในกรมธรรม์และมากเพียงพอสำหรับชดเชยค่าใช้จ่ายในการบริหารงานของบริษัทประกันภัย
2. ความเป็นธรรม (Not Unfairly Discriminate) เบี้ยประกันภัยจะต้องมีความเท่าเทียมกันระหว่างกลุ่มผู้เอาประกันภัยที่มีความเสี่ยงภัยอยู่ในระดับเดียวกัน
3. ความเหมาะสมของเบี้ยประกันภัยไม่ควรมีค่าสูงเกินไป (Not Excessive)

ในการวิเคราะห์ คำนวณอัตราหรือค่าต่างๆ ดังกล่าวข้อมูลความเสียหายที่นำมาใช้วิเคราะห์นั้นอาจมีความสัมพันธ์กับตัวแปรอื่นๆ เช่น การประกันภัยสุขภาพ จำนวนเงินที่จ่ายเนื่องจากค่ารักษาพยาบาลจะขึ้นอยู่กับ อายุ เพศ และ ระดับความเสี่ยงภัยของอาชีพของผู้เอาประกันภัย จึงทำให้อัตราเบี้ยประกันภัยที่เรียกเก็บจากผู้เอาประกันภัยแตกต่างกัน และข้อมูลความเสียหายโดยทั่วไปจะเป็นข้อมูลที่ไม่สมบูรณ์ (Incomplete Data) ตัวอย่างเช่น บริษัทประกันภัยจะจ่ายชดเชยความเสียหายให้แก่ผู้เอาประกันภัยไม่เกินความเสียหายที่เกิดขึ้นจริงและไม่เกินจำนวนเงินรับผิดชอบสูงสุดของกรมธรรม์ เช่น กรมธรรม์ประกันภัยกำหนดจำนวนเงินรับผิดชอบสูงสุดต่อกรมธรรม์ไว้เท่ากับ T_c บาท และให้ T แทนจำนวนความเสียหายจริงที่เกิดขึ้นแก่ผู้เอาประกันภัย ดังนั้นถ้า $T \leq T_c$ บริษัทจะจ่ายชดเชยเท่ากับ T บาท และบันทึกการจ่ายเงินชดเชยความเสียหายเท่ากับ T บาท แต่ถ้า $T > T_c$ บาท บริษัทจะจ่ายชดเชยเท่ากับ T_c บาท และบันทึกการจ่ายเงินชดเชยความเสียหายเท่ากับ T_c บาท

ในกรณีเช่นนี้ข้อมูลความเสียหายของบริษัทจึงไม่ใช่ข้อมูลที่แท้จริงทั้งหมดและมีค่าเฉลี่ยต่ำกว่าค่าที่แท้จริง และเนื่องจากข้อมูลที่ถูกตัดทิ้งมีความสำคัญในการนำมาวิเคราะห์เพื่อกำหนดความคุ้มครองและอัตราเบี้ยประกันภัยให้มีความเพียงพอและเป็นธรรมทั้งแก่บริษัทประกันภัยและผู้เอาประกันภัย ดังนั้นจึงควรทำการประมาณค่าข้อมูลที่ถูกตัดทิ้งโดยวิธีที่เหมาะสมเพื่อให้ได้ค่าประมาณที่ใกล้เคียงกับค่าจริงมากที่สุด และมีจำนวนข้อมูลเพียงพอที่จะนำไปศึกษาตามวัตถุประสงค์ เป็นต้นว่า เพื่อสร้างตัวแบบที่เหมาะสม หรือเพื่อศึกษาถึงการแจกแจงของค่าความเสียหาย

ในทางปฏิบัติ ถ้าทราบการแจกแจงของข้อมูลความเสียหาย ก็จะสามารถประมาณค่าข้อมูลความเสียหายโดยวิธีการประมาณที่ใช้พารามิเตอร์(Parametric Estimation) ได้ แต่มีบ่อยครั้งที่ไม่สามารถหาหรือไม่ทราบรูปแบบการแจกแจงของความเสียหาย ดังนั้นอาจประมาณค่าข้อมูลความเสียหายโดยวิธีการประมาณที่ไม่ใช้พารามิเตอร์(Nonparametric Estimation)

จากปัญหาดังกล่าว ผู้วิจัยจึงสนใจที่จะทำการศึกษาวิธีการประมาณค่าพารามิเตอร์แบบนอนพาราเมตริกในสมการถดถอยเชิงเส้นอย่างง่าย เมื่อตัวแปรตามมีค่าถูกตัดทิ้งทางขวา ประเภทที่ 1 โดยพิจารณาให้ตัวแปรตามเป็นความเสียหายที่เกิดขึ้นแก่ผู้เอาประกันภัยและมีการกำหนดค่าความเสียหายสูงสุดไว้ล่วงหน้า(T_c) ดังนั้นข้อมูลความเสียหายบางส่วนที่มากกว่าค่า T_c จะเป็นข้อมูลที่ถูกตัดทิ้งและถือเป็นข้อมูลที่ถูกตัดทิ้งประเภทที่ 1 ส่วนรูปแบบการแจกแจงของค่าคลาดเคลื่อนจะพิจารณา 3 รูปแบบ คือ การแจกแจงแบบปกติ(Normal Distribution) การแจกแจงแบบลอกนอร์มอล(Lognormal Distribution) และ การแจกแจงแบบไวบูลล์(Weibull Distribution) และพิจารณาให้ตัวแปรอิสระเป็นปัจจัยที่มีความสัมพันธ์กับความเสียหายที่เกิดขึ้นแก่ผู้เอาประกันภัย ศึกษารูปแบบการแจกแจง 2 แบบ คือ การแจกแจงแบบปกติ(Normal Distribution) และการแจกแจงแบบไวบูลล์ (Weibull Distribution)

การประมาณค่าพารามิเตอร์ในสมการถดถอยเชิงเส้นอย่างง่ายโดยทำการวิเคราะห์กับข้อมูลค่าสังเกตของตัวแปรตามที่บางส่วนมีค่าถูกตัดทิ้งนั้น การนำข้อมูลที่มีค่าถูกตัดทิ้งมาใช้ประมาณค่าพารามิเตอร์ อาจกระทำได้ 2 กรณี คือ

กรณีแรก จะพิจารณาว่าข้อมูลที่เป็นค่าที่ถูกตัดทิ้งเสมือนเป็นค่าที่ไม่ถูกตัดทิ้ง แล้วนำข้อมูลทั้งหมดมาวิเคราะห์

กรณีที่สอง จะไม่พิจารณาข้อมูลส่วนที่เป็นค่าที่ถูกตัดทิ้ง นั่นคือ จะทำการวิเคราะห์ข้อมูลเฉพาะกับข้อมูลที่ไม่ถูกตัดทิ้งเท่านั้น ดังนั้นจำนวนข้อมูลที่นำมาวิเคราะห์จึงมีขนาดตัวอย่างน้อยกว่ากรณีแรก

ได้มีผู้คิดวิธีการในการประมาณค่าพารามิเตอร์ในสมการถดถอยเชิงเส้น เมื่อข้อมูลของตัวแปรตามบางส่วนถูกตัดทิ้ง ดังนี้

มิลเลอร์(Miller,R.G.,Biometrica,1976) และมิลเลอร์(Rupert Miller and Jerry Halpern, Biometrica,1982) ได้ทำการประมาณค่าพารามิเตอร์แบบนอนพาราเมตริกโดยใช้วิธีกำลังสองต่ำสุดแบบถ่วงน้ำหนัก (Weighted Least Squares) โดยใช้ตัวประมาณของแคีพแลน-ไมเออร์เป็นค่าถ่วงน้ำหนัก และทำการกระทำซ้ำ(Iterative)กับค่าพารามิเตอร์ที่ประมาณได้จนกว่าค่าประมาณพารามิเตอร์จะคงตัว (Converge)

บัคเลย์และเจมส์(Jonathan Buckley and Ian James ,Biometrica,1979) ได้ทำการประมาณค่าพารามิเตอร์แบบนอนพาราเมตริกโดยทำการประมาณค่าข้อมูลที่ถูกตัดทิ้งขึ้นมาก่อนแล้วจึงทำการวิเคราะห์ข้อมูลรวมทั้งหมด โดยใช้หลักการของวิธีกำลังสองต่ำสุด

แชตเทอร์จีและแมคลีช(S.Chatterjee and D.L.McLeish,Communication Statistics,1986) ได้ทำการประมาณพารามิเตอร์ด้วยวิธีกำลังสองต่ำสุด แล้วประมาณค่าข้อมูลที่ถูกตัดทิ้งซึ่งจะมีค่ามากกว่าค่าที่ถูกตัดทิ้ง แล้วทำการกระทำซ้ำจากข้อมูลรวมทั้งหมดจนค่าพารามิเตอร์คงตัว ซึ่งวิธีนี้เป็นวิธีการแบบนอนพาราเมตริก

จงดี โรจนประศาสน์ (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย,2536) ได้ทำการศึกษาเปรียบเทียบวิธีกำลังสองต่ำสุด วิธีการของบัคเลย์และเจมส์ และวิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุดของสมการถดถอยเชิงเส้นอย่างง่าย โดยที่ค่าคลาดเคลื่อนมีการแจกแจงแบบปกติ จากการศึกษาพบว่า วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุดจะให้ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (The Square Root of Mean Square Error :RMSE) ของการประมาณค่าตัวแปรตามต่ำกว่าวิธีการอื่นๆ ที่ทำการศึกษา

จำเนียร จำนงค์รักษ์ (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย,2539) ได้ทำการศึกษาเปรียบเทียบวิธีการของบัคเลย์และเจมส์ วิธีการของแชตเทอร์จีและแมคลีช วิธีกำลังสองต่ำสุด และวิธีภาวะน่าจะเป็นสูงสุดด้วยขั้นตอนวิธีอีเอ็มของสมการถดถอยเชิงเส้นพหุ และศึกษาเมื่อค่าคลาดเคลื่อนมีการแจกแจงแบบปกติ แบบดับเบิลเอกซโพเนนเชียล และแบบลอกนอร์มอล จากการศึกษาพบว่า วิธีการประมาณด้วยวิธีภาวะน่าจะเป็นสูงสุดด้วยขั้นตอนวิธีอีเอ็ม จะให้ค่าคลาดเคลื่อน RMSE ของการประมาณค่าตัวแปรตามต่ำกว่าวิธีการอื่นๆ ที่ได้ทำการศึกษา และเมื่อการแจกแจงของค่าคลาดเคลื่อนเป็นแบบเบ้ขวา การประมาณด้วยวิธีการทั้ง 4 ที่ศึกษาจะให้ค่า RMSE เพิ่มมากขึ้นเมื่อเปอร์เซ็นต์การถูกตัดทิ้งของข้อมูลมีค่ามาก

การวิจัยครั้งนี้ศึกษาการประมาณค่าพารามิเตอร์แบบนอนพาราเมตริกในสมการถดถอยเชิงเส้นอย่างง่ายเมื่อตัวแปรตามมีค่าถูกตัดทิ้งทางขวา ประเภทที่ 1 ด้วยวิธี

1. วิธีตัวประมาณของมิลเลอร์ (Miller 's Estimator)
2. วิธีกำลังสองต่ำสุดแบบดัดแปลงแค็พแลน-ไมเออร์ (Modified Kaplan-Meier Least Squares Method)
3. วิธีการของบัคเลย์และเจมส์ (Buckley and James Method)

และศึกษาเปรียบเทียบกับวิธีกำลังสองต่ำสุด(Ordinary Least Squares Method) เนื่องจากทั้ง 3 วิธีการดังกล่าวต่างก็ใช้หลักการของวิธีกำลังสองต่ำสุด

1.2 วัตถุประสงค์ของการวิจัย

เพื่อเปรียบเทียบความถูกต้องของการประมาณค่าตัวแปรตามในสมการถดถอยเชิงเส้นอย่างง่าย เมื่อค่าสังเกตของตัวแปรตามบางค่าถูกตัดทิ้งทางขวาประเภทที่ 1 ด้วยวิธีการดังนี้

1. วิธีกำลังสองต่ำสุด(Ordinary Least Squares Method)
2. วิธีตัวประมาณของมิลเลอร์(Miller 's Estimator)
3. วิธีกำลังสองต่ำสุดแบบดัดแปลงแค็พแลน-ไมเออร์(Modified Kaplan-Meier Least Squares Method)
4. วิธีการของบัคเลย์และเจมส์(Buckley and James Method)

1.3 สมมติฐานการวิจัย

1.3 . 1 ภายใต้ลักษณะการแจกแจงของความคลาดเคลื่อนแบบปกติ วิธีการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองต่ำสุด เป็นวิธีการประมาณค่าตัวแปรตามได้ใกล้เคียงกับค่าจริงมากกว่าวิธีตัวประมาณของมิลเลอร์ วิธีกำลังสองต่ำสุดแบบดัดแปลงแค็พแลน-ไมเออร์ และวิธีการของบัคเลย์และเจมส์ โดยมีค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (The Square Root of Mean Square Error :RMSE) ต่ำสุด

1.3.2 ภายใต้สถานการณ์ที่มีการแจกแจงของค่าคลาดเคลื่อน, เปอร์เซ็นต์การถูกตัดทิ้งของข้อมูล, ขนาดตัวอย่าง, ค่าสูงสุดของข้อมูลที่ถูกตัดทิ้ง และวิธีการประมาณค่าตัวแปรตาม เดียวกัน ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการประมาณที่ได้จะมีค่าใกล้เคียงกัน โดยไม่คำนึงถึงว่าการแจกแจงของตัวแปรอิสระจะเป็นแบบใด

หอสมุดกลาง
จุฬาลงกรณ์มหาวิทยาลัย

1.4 ข้อตกลงเบื้องต้น

1.4.1 การวิจัยจะศึกษาภายใต้รูปแบบสมการถดถอยเชิงเส้นอย่างง่าย

$$T_i = \alpha + \beta x_i + \varepsilon_i \quad ; i = 1,2,3,...,N$$

เมื่อ T_i เป็นตัวแปรตาม

X_i เป็นตัวแปรอิสระ

ε_i เป็นค่าคลาดเคลื่อนสุ่ม

α, β เป็นพารามิเตอร์ที่ไม่ทราบค่า

N เป็นขนาดตัวอย่างทั้งหมด

1.4.2 การแจกแจงของค่าที่ถูกตัดทิ้งและค่าที่ไม่ถูกตัดทิ้ง เป็นอิสระต่อกัน

1.4.3 ตัวแปรตาม(Response Variable or Dependent Variable) เท่านั้นที่เป็นค่าที่ถูกตัดทิ้ง และเป็นการตัดทิ้งทางขวา ประเภทที่ 1

1.4.4 ค่าสังเกตที่สนใจศึกษา คือ

$$Y_i = \begin{cases} T_i \; ; & T_i \le T_c \quad i = 1,2,...,n \\ T_c \; ; & T_i > T_c \quad i = n+1, n+2, ..., N \end{cases}$$

เมื่อ T_i เป็นค่าสังเกตที่ไม่ถูกตัดทิ้ง

T_c เป็นค่าสูงสุดที่ถูกตัดทิ้ง

n เป็นขนาดตัวอย่างที่ไม่ถูกตัดทิ้ง

N เป็นขนาดตัวอย่างทั้งหมด ดังนั้น N-n เป็นขนาดตัวอย่างที่ถูกตัดทิ้ง

1.5 ขอบเขตของการวิจัย

1.5.1 ในการวิจัยครั้งนี้จะทำการเปรียบเทียบการประมาณค่าตัวแปรตามในสมการถดถอยเชิงเส้นอย่างง่าย เมื่อค่าสังเกตของตัวแปรตามบางส่วนถูกตัดทิ้งทางขวา ประเภทที่ 1 โดยการประมาณค่าพารามิเตอร์ด้วยวิธี

1) วิธีกำลังสองต่ำสุด

2) วิธีตัวประมาณของมิลเลอร์

3) วิธีกำลังสองต่ำสุดแบบดัดแปลงแค็พแลน-ไมเออร์

4) วิธีการของบัคเลย์และเจมส์

1.5.2 ลักษณะข้อมูลที่ใช้ศึกษาในครั้งนี้ จะมีการกำหนดการแจกแจง ค่าเฉลี่ย และค่าความแปรปรวน ของตัวแปรอิสระและตัวแปรตาม โดยนำข้อมูลจากการจ่ายค่าสินไหมทดแทนของสัญญาเพิ่มเติมการประกันภัยสุขภาพประจำปี 2538 มาทำการวิเคราะห์ด้วยโปรแกรมสำเร็จรูปทางสถิติ ได้ผลดังนี้

1) ตัวแปรอิสระจะกำหนดให้เป็นอายุของผู้เอาประกันภัยขณะที่ซื้อสัญญาเพิ่มเติมการประกันภัยสุขภาพ มีหน่วยเป็นปี มีค่าเฉลี่ยเท่ากับ 34 ปี และมีการแจกแจงแบบปกติหรือแบบไวบูลล์

2) ตัวแปรตามจะกำหนดให้เป็นค่าใช้จ่ายในการรักษาพยาบาลของผู้เอาประกันภัย มีหน่วยเป็นพันบาท กรณีที่ค่าคลาดเคลื่อนมีการแจกแจงแบบปกติและแบบลอกนอร์มอลจะกำหนดให้ค่าเฉลี่ยของตัวแปรตามเท่ากับ 5.5 พันบาทและความแปรปรวนเท่ากับ 36 โดยที่ค่าคลาดเคลื่อนมีค่าเฉลี่ยเท่ากับศูนย์ และเพื่อความสะดวกในการกำหนดค่าพารามิเตอร์ของการแจกแจงแบบไวบูลล์ ดังนั้นในกรณีที่ค่าคลาดเคลื่อนมีการแจกแจงแบบไวบูลล์จะกำหนดให้ค่าเฉลี่ยของตัวแปรตามเท่ากับ 7.5 พันบาทและความแปรปรวนเท่ากับ 20 โดยที่ค่าคลาดเคลื่อนมีค่าเฉลี่ยเท่ากับ 2 โดยปกติค่าความรับผิดชอบสูงสุดของบริษัทประกันภัยจะมีค่ามากกว่าค่าเฉลี่ยความเสียหาย(μ_T) ในการวิจัยครั้งนี้จึงศึกษาที่ 3 ระดับ คือ $\mu_T + 0.5\sigma_T$, $\mu_T + \sigma_T$ และ $\mu_T + 1.5\sigma_T$

1.5.3 ในการวิจัยครั้งนี้ศึกษาภายใต้ลักษณะการแจกแจงของตัวแปรอิสระ 1 ตัวแปร ที่มีการแจกแจงแบบต่อเนื่อง 2 รูปแบบ คือ

1) การแจกแจงแบบปกติ(Normal Distribution) มีฟังก์ชันความหนาแน่น คือ

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

ในการวิจัยครั้งนี้จะศึกษาที่ $\mu = 34$ และ $\sigma^2 = 144$

2) การแจกแจงแบบไวบูลล์(Weibull Distribution) มีฟังก์ชันความหนาแน่น คือ

$$f(x) = \alpha \beta^{-\alpha} x^{\alpha-1} \exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right] \quad ; \quad x > 0$$

ในการวิจัยครั้งนี้จะศึกษาที่ $\alpha = 3.3$ และ $\beta = 38$

1.5.4 ในการวิจัยครั้งนี้ศึกษาการแจกแจงของค่าคลาดเคลื่อน 3 แบบ และภายใต้การแจกแจงของค่าคลาดเคลื่อนแต่ละแบบกำหนดค่าสูงสุดของข้อมูลที่ถูกตัดทิ้ง 3 ระดับ ดังนี้

1) การแจกแจงแบบปกติ ศึกษาที่ $\mu = 0$ และ $\sigma^2 = 36$ และกำหนดค่าสูงสุดของข้อมูลที่ถูกตัดทิ้ง ที่ $\mu_T+0.5\sigma_T$, $\mu_T+\sigma_T$ และ $\mu_T+1.5\sigma_T$ ซึ่งเท่ากับ 8.5 , 11.5 และ 14.5 ตามลำดับ เมื่อ μ_T เป็นค่าเฉลี่ยของ T มีค่าเท่ากับ 5.5 และ σ_T เป็นส่วนเบี่ยงเบนมาตรฐาน(Standard Deviation) ของ T มีค่าเท่ากับ 6

2) การแจกแจงแบบลอกนอร์มอล (Lognormal Distribution) ซึ่งมีฟังก์ชันความหนาแน่น คือ

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

ในการวิจัยครั้งนี้จะศึกษาที่ $\mu = -5.5$ และ $\sigma^2 = 7.3$ และกำหนดค่าสูงสุดของข้อมูลที่ถูกตัดทิ้งที่ $\mu_T+0.5\sigma_T$, $\mu_T+\sigma_T$ และ $\mu_T+1.5\sigma_T$ ซึ่งเท่ากับ 8.5 , 11.5 และ 14.5 ตามลำดับ เมื่อค่าเฉลี่ยของ T (μ_T) เท่ากับ 5.5 และส่วนเบี่ยงเบนมาตรฐานของ T (σ_T) เท่ากับ 6

3) การแจกแจงแบบไวบูลล์ ศึกษาที่ $\alpha = 0.5$ และ $\beta = 1$ และกำหนดค่าสูงสุดของข้อมูลที่ถูกตัดทิ้งที่ $\mu_T+0.5\sigma_T$, $\mu_T+\sigma_T$ และ $\mu_T+1.5\sigma_T$ ซึ่งเท่ากับ 10 , 12 และ 14 ตามลำดับ เมื่อค่าเฉลี่ยของ T (μ_T) เท่ากับ 7.5 และส่วนเบี่ยงเบนมาตรฐานของ T (σ_T) เท่ากับ $\sqrt{20}$

1.5.5 ศึกษาเมื่อเปอร์เซ็นต์การถูกตัดทิ้งของข้อมูลเท่ากับ 10%, 20% , 30% และ 40%

1.5.6 ศึกษาเมื่อขนาดตัวอย่างเท่ากับ 20, 30, 40, 50 และ 60

1.5.7 ในการวิจัยครั้งนี้จำลองสถานการณ์ตามที่ต้องการศึกษา โดยใช้เทคนิคมอนติคาร์โล (Monte Carlo Simulation technique) เขียนโปรแกรมด้วยภาษาฟอร์แทรน (FORTRAN77) ทำการจำลองข้อมูลซ้ำๆ กัน 1,000 ครั้งในแต่ละสถานการณ์

1.6 <u>เกณฑ์การตัดสินใจ</u>

เกณฑ์การตัดสินใจว่าวิธีการประมาณค่าพารามิเตอร์ในสมการถดถอยเชิงเส้นอย่างง่ายด้วยวิธีใดจะให้ค่าประมาณของตัวแปรตามใกล้เคียงกับค่าจริงได้ดีกว่า จะพิจารณาโดยการเปรียบเทียบค่าประมาณของตัวแปรตามกับค่าจริงของตัวแปรตามด้วยค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (The Square Root of Mean Square Error :RMSE) วิธีใดให้ค่า RMSE ต่ำกว่าจะเป็นวิธีการประมาณที่ดีกว่า คำนวณจากสูตร

$$MSE = \sum_{i=1}^{N} \frac{\left[T_i - \hat{Y}_i\right]^2}{N}$$

$$RMSE = \sqrt{MSE}$$

เมื่อ N เป็นขนาดตัวอย่างทั้งหมด

T_i เป็นค่าจริงของตัวแปรตาม

$\hat{Y}_i$ เป็นค่าประมาณของตัวแปรตาม

1.7 ประโยชน์ที่คาดว่าจะได้รับ

1.7.1 ผลจากการวิจัยจะเป็นแนวทางในการเลือกวิธีการประมาณค่าความเสียหาย เมื่อมีปัจจัยที่มีผลต่อค่าความเสียหายได้อย่างมีประสิทธิภาพ และเหมาะสมกับลักษณะของข้อมูลเกี่ยวกับการประกันภัย และช่วยให้นักคณิตศาสตร์ประกันภัยกำหนดความคุ้มครองและคำนวณอัตราเบี้ยประกันภัย หรือค่าอื่นๆ ที่เกี่ยวข้องได้อย่างถูกต้องยิ่งขึ้น และส่งผลให้การวางแผนการจัดการด้านการเงินของบริษัทประกันภัยมีประสิทธิภาพยิ่งขึ้น

1.7.2 เป็นแนวทางสำหรับผู้ที่สนใจจะทำการศึกษาวิจัยในหัวข้ออื่น ๆ ที่เกี่ยวข้องต่อไป เช่น การศึกษาและเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์แบบนอนพาราเมตริก ในสมการถดถอยเชิงเส้นพหุ เมื่อตัวแปรตามมีค่าถูกตัดทิ้งทางขวา เป็นต้น