บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ต้องการศึกษาและเปรียบเทียบวิธีการประมาณความน่าจะเป็นที่คนอายุ x จะเสียชีวิตภายใน 1 ปีข้างหน้า (q_x) สำหรับข้อมูลประกันชีวิตที่ไม่สมบูรณ์ วิธีการประมาณค่าทางสถิติที่เสนอในที่นี้มี 3 วิธี คือ

1. วิธีการประมาณแบบคลาสสิค
2. วิธีการประมาณแบบภาวะน่าจะเป็นสูงสุด
3. วิธีการประมาณแบบเบส์

วิธีการประมาณแบบคลาสสิค เป็นวิธีการประมาณค่า q_x อย่างง่าย โดยหลักการเกี่ยวกับความน่าจะเป็นเบื้องต้นคือจำนวนผู้ที่เสียชีวิตหารด้วยจำนวนผู้เสี่ยงภัยทั้งหมด วิธีการประมาณแบบภาวะน่าจะเป็นสูงสุด เป็นวิธีการประมาณค่า q_x โดยอาศัยการแจกแจงของระยะเวลาที่จะมีชีวิตต่อไปในอนาคต(Future Lifetime, T)ซึ่งในการวิจัยครั้งนี้จะใช้การแจกแจงแบบไวบูลล์และการแจกแจงแบบกอมเพิรตซ์ ดังได้กล่าวไปแล้วในบทที่หนึ่ง และบทที่สอง โดยจะทำการประมาณค่าพารามิเตอร์ของการแจกแจงดังกล่าวและอาศัยเทคนิคการวิเคราะห์เชิงตัวเลข (Numerical Analysis) เข้ามาช่วยในการประมาณค่าพารามิเตอร์และหลังจากได้ประมาณค่าพารามิเตอร์แล้วก็จะสามารถประมาณค่า q_x ได้ ส่วนวิธีการประมาณแบบเบส์นั้นเป็นวิธีการประมาณค่าq_x โดยการกำหนดการแจกแจงก่อนการทดลองเป็นเอกซโพเนนเชียล ซึ่งจะได้การแจกแจงหลังการทดลองเป็นแบบแกมมา และจากการแจกแจงหลังการทดลองนี้ จะทำให้สามารถประมาณค่า q_x ได้

ในแต่ละวิธีจะประมาณค่า q_x สำหรับอายุ x ในช่วง 25 - 65 ปี ภายใต้สถานการณ์เมื่อขนาดตัวอย่างมี 8 ระดับ คือ 30, 50, 70, 100, 300, 500, 700 และ 1,000 ตามลำดับ โดยทำซ้ำ 500 ครั้งในแต่ละสถานการณ์

การวิจัยครั้งนี้ใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Technique)

สร้างสถานการณ์ต่างๆ ดังนั้น จะอธิบายถึงวิธีการจำลองโดยใช้เทคนิคมอนติคาร์โลแล้วจึงแสดง

รายละเอียดของแผนการทดลอง ขั้นตอนการวิจัย และ โปรแกรมที่ใช้ในการวิจัยตามลำดับ

วิธีการจำลองโดยใช้เทคนิคมอนติคาร์โล

เทคนิคที่ใช้แก้ปัญหาในการคำนวณทางคณิตศาสตร์นั้นมีอยู่หลายวิธี วิธีการจำลองโดยใช้เทคนิคมอนติคาร์โลเป็นวิธีหนึ่งที่นิยมนำมาใช้แก้ปัญหากันมากในปัจจุบัน ซึ่งหลักการของการจำลองโดยใช้เทคนิคดังกล่าว จะใช้ตัวเลขสุ่ม (Random Numbers) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา

ขั้นตอนของวิธีการจำลองด้วยเทคนิคมอนติคาร์โลที่ใช้กันในปัจจุบันแบ่งได้เป็น 3 ขั้นตอนคือ

1 การสร้างตัวเลขสุ่ม การใช้ตัวเลขสุ่มเป็นสิ่งที่สำคัญมากในเทคนิคนี้ เพราะว่าหลักการของการจำลองแบบมอนติคาร์โลนั้น จะใช้ตัวเลขสุ่มเข้ามาช่วยในการหาคำตอบของปัญหาโดยลักษณะของตัวเลขสุ่มที่นำมาใช้ จะมีการแจกแจงแบบสม่ำเสมอ (Uniform Distribution) ในช่วง (0, 1)สำหรับวิธีการสร้างตัวเลขสุ่มนั้นมีผู้เสนอไว้หลายวิธี แต่วิธีที่ดีนั้นลักษณะของตัวเลขสุ่มที่ถูกสร้างขึ้น จะต้องมีการแจกแจงแบบสม่ำเสมอในช่วง (0, 1) ตัวเลขสุ่มแต่ละตัวเป็นอิสระต่อกันและมีช่วงยาวก่อนจะเกิดเลขสุ่มซ้ำ

2 การนำตัวเลขสุ่มมาประยุกต์ใช้กับปัญหาที่ต้องการจะศึกษา ซึ่งขั้นตอนนี้ขึ้นอยู่กับลักษณะของปัญหา บางปัญหาอาจจะไม่ใช้ตัวเลขสุ่มโดยตรง แต่จะนำไปผลิตตัวเลขสุ่มที่มีการแจกแจงแบบอื่นต่อไป

3 การทดลองกระทำ เมื่อนำตัวเลขสุ่มมาประยุกต์ให้เข้ากับปัญหาที่ต้องการศึกษาได้แล้วขั้นต่อไปคือ การทดลองโดยใช้กระบวนการของการสุ่ม (Random Process) มากระทำในลักษณะซ้ำๆ กันหลายๆ ครั้ง เพื่อหาคำตอบที่ต้องการ

แผนการทดลอง

การวิจัยครั้งนี้ต้องการประมาณความน่าจะเป็นที่คนอายุ x จะเสียชีวิตภายใน 1 ปีข้างหน้าสำหรับข้อมูลประกันชีวิตที่ไม่สมบูรณ์ หรือเป็นข้อมูลที่ถูกตัดปลายทางขวา (Censored Data) โดยอาศัยวิธีการทางสถิติ 3 วิธีในการคำนวณหาค่าประมาณ q_x ภายใต้ข้อมูลเกี่ยวกับระยะเวลาที่

จะมีชีวิตอยู่ต่อไปในอนาคตมีการแจกแจง 2 แบบ ช่วงอายุที่ศึกษามี 41 อายุ และขนาดตัวอย่างที่

นำมาศึกษามี 8 ระดับ ซึ่งรวมทั้งสิ้น 656 สถานการณ์ ผู้วิจัยจะเปรียบเทียบค่าประมาณ q_x ที่ได้จาก 3 วิธี โดยพิจารณาเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (Absolute Percentage Error , APE) เพื่อหาวิธีที่ดีที่สุดในแต่ละสถานการณ์ต่อไป รายละเอียดของแผนการทดลองมีดังต่อไปนี้

1 ข้อมูลเกี่ยวกับระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคต มีการแจกแจง 2 แบบ คือ

1.1 การแจกแจงแบบไวบูลล์

1.2 การแจกแจงแบบกอมเพิรตซ์

2 ผู้วิจัยสนใจศึกษาค่า q_x สำหรับอายุ x ในช่วง 25 - 65 ปี

3 ขนาดตัวอย่างที่นำมาศึกษามี 8 ระดับคือ 30, 50, 70, 100, 300, 500, 700 และ 1000 ตามลำดับ

4 สำหรับวิธีการประมาณแบบเบส์ ผู้วิจัยได้ทำการกำหนดการแจกแจงก่อนการทดลองเป็นแบบเอกซโพเนนเชียล

<u>ขั้นตอนในการวิจัย</u>

แบ่งเป็น 4 ขั้นตอนหลัก ดังนี้

1 จำลองระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคตจากการแจกแจงแบบไวบูลล์ และการแจกแจงแบบกอมเพิรตซ์

2 กำหนดระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคต ให้มีลักษณะของข้อมูลเป็นแบบถูกตัดปลายทางขวา

3 หาค่าประมาณ q_x ด้วยวิธีการทางสถิติ 3 วิธี โดยอาศัยระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคต

4 หาค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์จากการประมาณค่า q_x และทำการเปรียบเทียบเพื่อที่จะหาวิธีการประมาณที่ดีที่สุด

รายละเอียดของแต่ละขั้นตอนมีดังนี้

1 การจำลองระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคตจากการแจกแจง 2 แบบ คือ การแจกแจงแบบไวบูลล์ และ การแจกแจงแบบกอมเพิรตซ์

1.1 เมื่อ T มีการแจกแจงแบบไวบูลล์

$$t_i \quad = \quad \left[-\frac{n+1}{k}\ln(1-U)\right]^{1/_{n+1}} \qquad ,i=1,2,\ldots,m$$

โดย t_i เป็นระยะเวลาของคนที่ i จะมีชีวิตอยู่ต่อไปในอนาคต

U เป็นตัวเลขสุ่มซึ่งอยู่ระหว่าง 0 และ 1

k,n เป็นค่าพารามิเตอร์เริ่มต้นสำหรับการจำลองข้อมูล โดยรายละเอียดของค่าพารามิเตอร์นี้ แสดงอยู่ในภาคผนวก ก

1.2 เมื่อ T มีการแจกแจงแบบกอมเพิรตซ์

$$t_i \quad = \quad \frac{\ln\left[1-\frac{\ln c}{B}\ln(1-U)\right]}{\ln c} \qquad ,i=1,2,\ldots,m$$

โดย t_i เป็นระยะเวลาของคนที่ i จะมีชีวิตอยู่ต่อไปในอนาคต

U เป็นตัวเลขสุ่มซึ่งอยู่ระหว่าง 0 และ 1

B,c เป็นค่าพารามิเตอร์เริ่มต้นสำหรับการจำลองข้อมูล โดยรายละเอียดของค่าพารามิเตอร์นี้ แสดงอยู่ในภาคผนวก ก

2 กำหนดระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคต ซึ่งจำลองได้จากหัวข้อ 1 ให้มีลักษณะของข้อมูลเป็นแบบถูกตัดปลายทางขวาโดยกำหนดให้ t_i = 1 เมื่อ t_i > 1 นั่นคือ การวิจัยครั้งนี้ได้กำหนดช่วงเวลาที่สนใจจะศึกษา(Observation Period) มีระยะเวลาเป็น 1 ปี ดังนั้นถ้าผู้ใดมีระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคตเกิน 1 ปี ให้ถือว่าผู้นั้นเป็นผู้อยู่รอดเมื่อสิ้นสุดการศึกษา

3 หาค่าประมาณ q_x ด้วยวิธีการทางสถิติ 3 วิธีโดยอาศัยระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคต ซึ่งได้จากหัวข้อ 3.3.1 และ หัวข้อ 2 มีขั้นตอนดังนี้

3.1 <u>การประมาณค่า q_x ด้วยวิธีแบบคลาสสิค</u>

1) หาจำนวนผู้ที่เสียชีวิตในช่วงเวลาที่สนใจศึกษาจากจำนวนผู้เสี่ยงภัยทั้งหมด ซึ่งก็คือผู้ที่มีชีวิตอยู่ต่อไปในอนาคตน้อยกว่าหนึ่งปี (t_i < 1)

2) หาค่าประมาณ q_x จากสมการที่ (2.1)

3.2 การประมาณค่า q_x ด้วยวิธีแบบภาวะน่าจะเป็นสูงสุด

1) เมื่อ T มีการแจกแจงแบบไวบูลล์

$$L = \prod_{i=1}^{m}\left(kt_i^n\right)^{\delta_i} \exp\left[-\frac{k}{n+1}t_i^{n+1}\right]$$

ประมาณค่าพารามิเตอร์ k และ n โดยหาอนุพันธ์บางส่วน (Partial Derivative) ของล็อกของฟังก์ชันภาวะน่าจะเป็นเทียบกับ k และ n และให้สมการอนุพันธ์บางส่วนเท่ากับศูนย์ ดังได้แสดงในสมการที่ (2.2) ถึงสมการที่ (2.7) ซึ่งจะต้องอาศัยเทคนิคการวิเคราะห์เชิงตัวเลขโดยใช้วิธีนิวตัน - ราพซ์สันในการแก้สมการต่อไปนี้

$$g_1(k_0,n_0) + g_{11}(k_0,n_0)(k-k_0) + g_{12}(k_0,n_0)(n-n_0) = 0$$
$$g_2(k_0,n_0) + g_{21}(k_0,n_0)(k-k_0) + g_{22}(k_0,n_0)(n-n_0) = 0$$

จะกำหนดค่าเริ่มต้น n_0 และ k_0 จากสมการที่ (2.9) และ (2.10) ตามลำดับ และจะต้องใช้กระบวนการการทำซ้ำจนกระทั่งได้ค่าพารามิเตอร์ $\hat{k}$ และ $\hat{n}$ ที่ต้องการ หลังจากนั้นสามารถนำไปหาค่า q_x ได้โดยอาศัยสมการที่ (2.8)

2) เมื่อ T มีการแจกแจงแบบกอมเพิรตซ์

$$L = \prod_{i=1}^{m}\left(Bc^{t_i}\right)^{\delta_i} \exp\left[-\frac{B}{\ln c}\left(c^{t_i}-1\right)\right]$$

ประมาณค่าพารามิเตอร์ B และ c โดยหาอนุพันธ์บางส่วน (Partial Derivative) ของล็อกของฟังก์ชันภาวะน่าจะเป็นเทียบกับ B และ c และให้สมการอนุพันธ์บางส่วนเท่ากับศูนย์ ดังได้แสดงในสมการที่ (2.11) ถึงสมการที่ (2.16) ซึ่งจะต้องอาศัยเทคนิคการวิเคราะห์เชิงตัวเลขโดยใช้วิธีนิวตัน - ราพซ์สันในการแก้สมการดังต่อไปนี้

$$g_1(B_0,c_0) + g_{11}(B_0,c_0)(B-B_0) + g_{12}(B_0,c_0)(c-c_0) = 0$$
$$g_2(B_0,c_0) + g_{21}(B_0,c_0)(B-B_0) + g_{22}(B_0,c_0)(c-c_0) = 0$$

กำหนดค่าเริ่มต้น B_0 และ c_0 จากสมการที่ (2.19) และ (2.20) ตามลำดับและจะต้องใช้กระบวนการการทำซ้ำจนกระทั่งได้ค่าพารามิเตอร์ $\hat{B}$ และ $\hat{c}$ ที่ต้องการ หลังจากนั้นสามารถนำไปหาค่า q_x ได้โดยอาศัยสมการที่ (2.17)

3.3 <u>การประมาณค่า q_x</u> <u>ด้วยวิธีแบบเบส์</u>

กำหนดการแจกแจงก่อนการทดลองเป็นแบบเอกซโพเนนเชียล ซึ่งมีพารามิเตอร์ คือ β ซึ่งมีหลักในการกำหนดค่า β ดังนี้

$$\hat{\mu} = E(\mu) = \frac{1}{\beta}$$

ภายใต้กำลังของมรณะเป็นค่าคงที่ สามารถคำนวณหาค่าประมาณ μ จากสมการดังต่อไปนี้

$$\hat{q}_x = 1-e^{-\hat{\mu}}$$

$$\hat{\mu} = -\ln(1-\hat{q}_x)$$ (โดยค่า $\hat{q}_x$ คำนวณจากสมการ (2.1))

ดังนั้น

$$\hat{\beta} = -\frac{1}{\ln(1-\hat{q}_x)}$$

หลังจากที่กำหนดการแจกแจงก่อนการทดลองและทราบค่าประมาณ $\hat{\beta}$ ของพารามิเตอร์ β แล้วคำนวณหาการแจกแจงหลังการทดลอง และจากสมการที่ (2.22) พบว่า การแจกแจงหลังการทดลองเป็นการแจกแจงแบบแกมมาซึ่งมีพารามิเตอร์คือ $\alpha' = 1+\sum_{i=1}^{m}\delta_i$ และ $\beta' = \beta+\sum_{i=1}^{m}t_i$ ตามลำดับ ดังนั้น จะสามารถประมาณค่า q_x ได้จากสมการที่ (2.23)

4 ในการทดลองได้จำลองข้อมูลซ้ำกันจำนวน 500 ครั้ง ในแต่ละสถานการณ์กำหนดให้ i แทนรอบที่ทำซ้ำ i = 1, 2, ... , 500 ดังนั้น

$$\hat{q}_x = \frac{1}{500}\sum_{i=1}^{500}\hat{q}_x^{(i)}$$

โดย $\hat{q}_x^{(i)}$ เป็นค่าประมาณ q_x สำหรับการทำซ้ำรอบที่ i

หลังจากนี้จะคำนวณค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (APE) ดังสมการต่อไปนี้

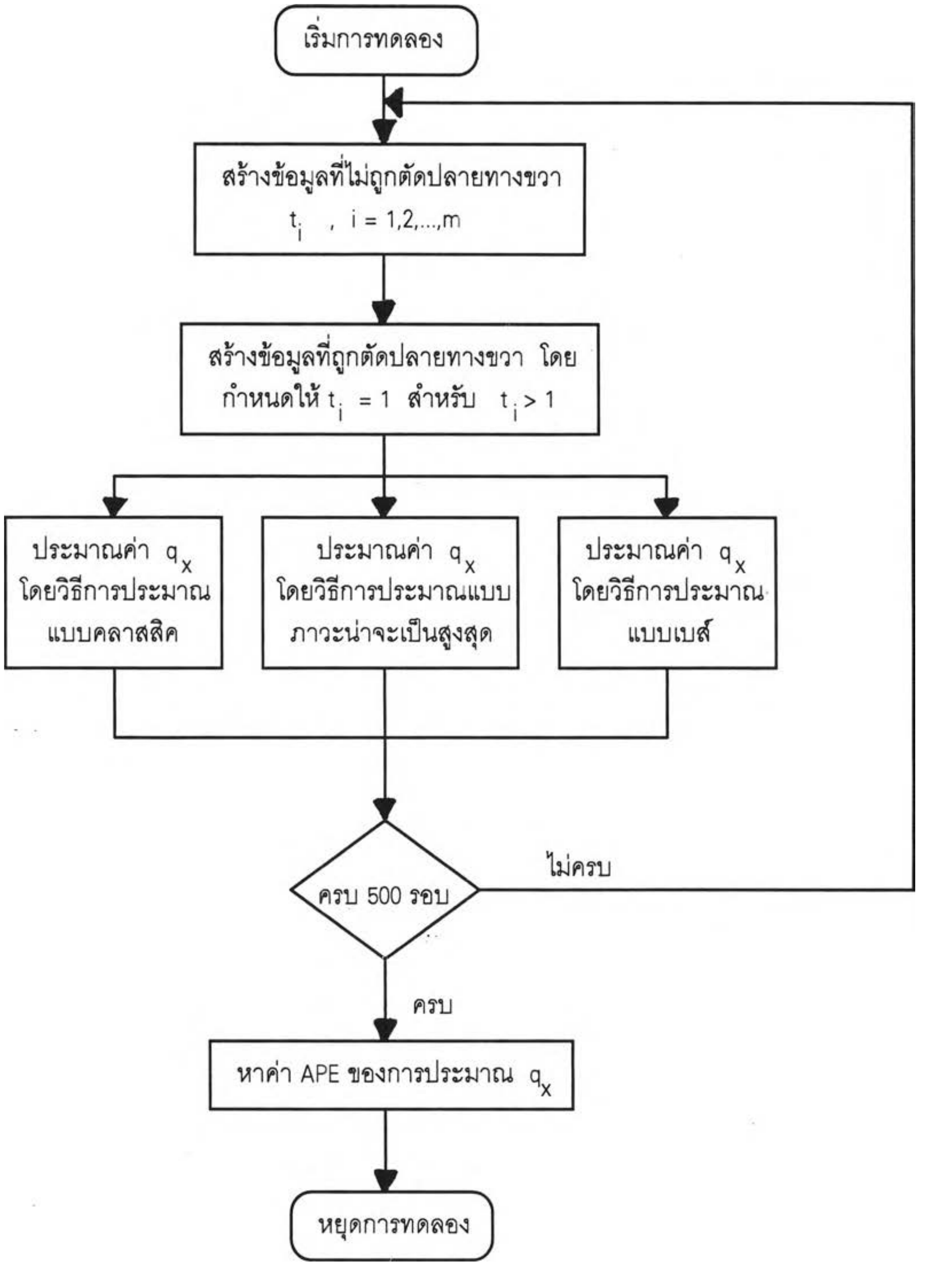
$$\text{APE} = \left| \frac{q_x - \hat{q}_x}{q_x} \right| \times 100\,\%$$

จากนั้นจะนำค่า APE ของการประมาณค่า q_x ของวิธีการทั้ง 3 วิธีมาเปรียบเทียบเพื่อหาว่าวิธีการใดให้ค่า APE ต่ำที่สุดวิธีการนั้นจะเป็นวิธีที่ประมาณค่า q_x ที่ดีที่สุดของแต่ละสถานการณ์

การคำนวณค่า APE ของการประมาณค่า q_x ของทั้ง 3 วิธี ในการทดลองที่อายุหนึ่งๆจะเปลี่ยนขนาดตัวอย่างเป็น 8 ระดับ และจะเปลี่ยนการแจกแจงของระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคต 2 แบบ โดยแต่ละสถานการณ์จะทดลองซ้ำๆ จำนวน 500 ครั้งจนครบทุกสถานการณ์ ซึ่งขั้นตอนของการทดลองดังกล่าว แสดงได้ดังรูปที่ 3.1

<u>โปรแกรมที่ใช้ในการวิจัย</u>

โปรแกรมที่ใช้ในการวิจัยทั้งหมดนี้ เขียนด้วยภาษาฟอร์แทรน (FORTRAN) โดยใช้กับเครื่องAMDAHL 5860 ซึ่งในแต่ละสถานการณ์ของการทดลอง ลักษณะการทำงานของโปรแกรมจะเหมือนกัน สำหรับรายละเอียดของโปรแกรมจะแสดงไว้ในภาคผนวก ข ซึ่งจะเป็นโปรแกรมการทำงานของแต่ละวิธีการ คือ วิธีการประมาณแบบคลาสสิค วิธีการประมาณแบบภาวะน่าจะเป็นสูงสุด และวิธีการประมาณแบบเบส์



รูปที่ 3.1 แสดงผังงานสำหรับหาเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (APE) จากการประมาณค่า q_x ด้วยวิธีการทั้ง 3 วิธี