บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยนี้เพื่อศึกษาและเปรียบเทียบ วิธีการปรับค่าประมาณความน่าจะเป็นที่คนอายุ x ปีจะเสียชีวิตภายใน 1 ปีข้างหน้า หรือค่าประมาณของค่า q_x ซึ่งกระทำภายใต้ข้อมูลที่มีลักษณะถูกตัดปลาย โดยนำเสนอวิธีการปรับค่า 3 วิธีดังนี้

1. การปรับค่าโดยใช้ค่าเฉลี่ยถ่วงน้ำหนักเคลื่อนที่
2. การปรับค่าโดยใช้รูปแบบฟังก์ชัน
3. การปรับค่าโดยใช้ส่วนโค้งพหุนามองศาสาม

การปรับค่าโดยใช้ค่าเฉลี่ยถ่วงน้ำหนักเคลื่อนที่ เป็นวิธีการปรับค่าโดยการนำค่าประมาณของค่า q_x มาเฉลี่ยถ่วงน้ำหนักด้วยค่า ๆ หนึ่ง เพื่อให้ได้ค่าประมาณที่ปรับแล้วมีความราบเรียบ การปรับค่าโดยใช้รูปแบบฟังก์ชันเป็นวิธีการปรับค่าโดยการแทนประมาณของค่า q_x ด้วยรูปแบบฟังก์ชันที่เหมาะสม ซึ่งนำเสนอ 2 รูปแบบคือ รูปแบบไวบูลล์และรูปแบบกอมเพริตซ์ แล้วใช้วิธีการกำลังสองน้อยที่สุด ในการประมาณค่าพารามิเตอร์ของฟังก์ชันดังกล่าว เพื่อหาค่าประมาณของค่า q_x ที่ปรับแล้ว การปรับค่าโดยใช้ส่วนโค้งพหุนามองศาสาม เป็นวิธีการปรับค่าโดยการแบ่งช่วงอายุเป็นช่วงย่อย ๆ และแทนแต่ละช่วงย่อยด้วยส่วนโค้งพหุนามองศาสาม ทำให้หาค่าประมาณของค่า q_x ที่ปรับแล้วได้ สำหรับการหาค่าประมาณของค่า q_x จะใช้วิธีการประมาณทางคณิตศาสตร์ประกันภัย

การวิจัยนี้ใช้เทคนิคการจำลองมอนติคาร์โล (Monte Carlo Simulation Technique) ในการสร้างสถานการณ์ต่าง ๆ เพื่อหาค่าประมาณของค่า q_x ในช่วงอายุ 0-99 ปี เมื่อขนาดตัวอย่างมี 5 ระดับ และทำการทดลอง 100 ครั้งในแต่ละสถานการณ์

<u>แผนการทดลอง</u>

การวิจัยนี้ต้องการปรับค่าประมาณความน่าจะเป็นที่คนอายุ x ปีจะเสียชีวิตภายใน 1 ปี

ข้างหน้า สำหรับข้อมูลที่มีลักษณะถูกตัดปลาย ด้วยวิธีการปรับค่า 3 วิธีภายใต้สถานการณ์ต่าง ๆ โดยศึกษาข้อมูลเกี่ยวกับระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคต ภายใต้การแจกแจง 2 รูปแบบ และระยะเวลาที่จะเกิดการถอนตัว ภายใต้การแจกแจง 2 รูปแบบเช่นเดียวกัน ซึ่งช่วงอายุที่ศึกษามี 100 อายุ และขนาดตัวอย่างข้อมูลที่นำมาศึกษามี 5 ระดับ ในการวิจัยจะทำการเปรียบเทียบค่าประมาณของค่า q_x ที่ปรับแล้วจากวิธีการทั้ง 3 วิธีดังกล่าว โดยพิจารณาเปอร์เซ็นต์ค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (MAPE) เพื่อหาวิธีการปรับค่าที่ดีที่สุด

รายละเอียดของแผนการทดลองมีดังต่อไปนี้

1. ระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคตภายใต้การแจกแจง 2 รูปแบบ คือ
   1.1 การแจกแจงแบบไวบูลล์
   1.2 การแจกแจงแบบกอมเพริตซ์
2. ระยะเวลาที่จะเกิดการถอนตัวภายใต้การแจกแจง 2 รูปแบบ คือ
   2.1 การแจกแจงแบบสม่ำเสมอ
   2.2 การแจกแจงแบบแกมมา
3. การปรับค่าประมาณของค่า q_x ศึกษาในช่วงอายุ 0-99 ปี
4. ขนาดตัวอย่างข้อมูลที่นำมาศึกษามี 5 ระดับ คือ 100, 300, 500, 700 และ 1000ตามลำดับ
5. การหาค่าประมาณของค่า q_x จะใช้วิธีการประมาณทางคณิตศาสตร์ประกันภัย

<u>ขั้นตอนในการวิจัย</u>

ขั้นตอนในการวิจัยแบ่งเป็น 6 ขั้นตอนตามลำดับดังนี้

1. ทำการจำลองระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคตภายใต้การแจกแจงแบบไวบูลล์และการแจกแจงแบบกอมเพริตซ์

2. การจำลองระยะเวลาที่จะเกิดการถอนตัว ภายใต้การแจกแจงแบบสม่ำเสมอและการแจกแจงแบบแกมมา

3. กำหนดระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคต และระยะเวลาที่จะเกิดการถอนตัวให้มีลักษณะเป็นข้อมูลที่ถูกตัดปลาย

4. หาค่าประมาณของค่า q_x จะใช้วิธีการประมาณทางคณิตศาสตร์ประกันภัย

5. ปรับค่าประมาณของค่า q_x ด้วยวิธีการปรับค่า 3 วิธี

6. เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์ จากการปรับค่าประมาณของค่า q_x ทั้ง 3 วิธีเพื่อหาวิธีการที่ดีที่สุด

รายละเอียดแต่ละขั้นตอนมีดังต่อไปนี้

1. ทำการจำลองระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคต ภายใต้การแจกแจงแบบไวบูลล์และการแจกแจงแบบกอมเพริตซ์

1.1 การแจกแจง แบบไวบูลล์

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

เมื่อ t_i คือระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคตของคนที่ i

U คือตัวเลขสุ่มที่อยู่ระหว่าง 0 และ 1

k,n คือพารามิเตอร์เริ่มต้นสำหรับการจำลองข้อมูล โดยแสดงรายละเอียดของค่าพารามิเตอร์ไว้ในภาคผนวก

1.2 การแจกแจงแบบกอมเพริตซ์

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

เมื่อ t_i คือระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคตของคนที่ i

U คือตัวเลขสุ่มที่อยู่ระหว่าง 0 และ 1

B,c คือพารามิเตอร์เริ่มต้นสำหรับการจำลองข้อมูล โดยแสดงรายละเอียดของค่าพารามิเตอร์ไว้ในภาคผนวก

2. ทำการจำลองระยะเวลาที่จะเกิดการถอนตัว ภายใต้การแจกแจงแบบสม่ำเสมอและการแจกแจงแบบแกมมา

2.1 การแจกแจงแบบสม่ำเสมอ

$$w_i = U \qquad , i = 1, 2 ,\ldots, m$$

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

เมื่อ w_i คือระยะเวลาที่จะเกิดการถอนตัวของคนที่ i

U คือตัวเลขสุ่มที่อยู่ระหว่าง 0 และ 1

2.2 การแจกแจงแบบแกมมา

$$w_i \quad = \quad \beta . \alpha . \exp\left[1/\sqrt{2\alpha - 1}\ln(U/(1-U))\right] \quad , \quad i = 1, 2, \ldots, m$$

เมื่อ w_i คือระยะเวลาที่จะเกิดการถอนตัวของคนที่ i

U คือตัวเลขสุ่มที่อยู่ระหว่าง 0 และ 1

α , β คือพารามิเตอร์เริ่มต้นสำหรับการจำลองข้อมูล โดยที่ $\alpha = 1.5, \beta = 1.0$

3. กำหนดข้อมูลที่จำลองได้จากขั้นตอนที่ 1 และ 2 ให้มีลักษณะถูกตัดปลาย ซึ่งกำหนดช่วงเวลาที่สนใจศึกษามีระยะเวลา 1 ปี แล้วทำการเปรียบเทียบว่าเหตุการณ์ใดเกิดขึ้นก่อน หากเหตุการณ์ตายเกิดขึ้นก่อนให้นับระยะเวลาเป็น 1 ปี หากเหตุการณ์การถอนตัวเกิดขึ้นก่อนให้นับระยะเวลาที่เกิดเหตุการณ์นั้น โดยกำหนดสัดส่วนการถอนตัวแต่ละระดับตามที่แสดงไว้ในขอบเขตของการวิจัยในบทที่ 1 หากข้อมูลมีระยะเวลาเกิน 1 ปีให้ถือว่ามีชีวิตอยู่รอดเมื่อสิ้นสุดการศึกษา

4. หาค่าประมาณของค่า q_x ด้วยวิธีการประมาณทางคณิตศาสตร์ประกันภัย ซึ่งได้แสดงขั้นตอนการคำนวณไว้ในบทที่ 2 โดยในการจำลองข้อมูลระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคตและระยะเวลาการถอนตัว จะทำการทดลอง 100 ครั้งในแต่ละสถานการณ์ ดังนั้นสามารถหาค่าประมาณของค่า q_x ได้ดังนี้

$$q'_x \quad = \quad \frac{1}{100}\sum_{i=1}^{100} q'^{(i)}_x$$

เมื่อ $q'^{(i)}_x$ คือค่าประมาณของค่า q_x สำหรับการทดลองครั้งที่ i

i คือครั้งที่ทำการทดลอง

5. ปรับค่าประมาณของค่า q_x ด้วยวิธีการปรับค่า 3 วิธี ดังต่อไปนี้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

5.1 การปรับค่าโดยใช้ค่าเฉลี่ยถ่วงน้ำหนักเคลื่อนที่

จากรูปแบบค่า R_z น้อยที่สุด $\left(\text{Minimum-}R_z \text{ Formula}\right)$ จะทำการหาค่าสัมประสิทธิ์ a_r โดยการหาอนุพันธ์ของฟังก์ชันR เทียบกับค่า a_r ดังสมการ (2.7) แล้วให้สมการอนุพันธ์มีค่าเท่ากับศูนย์ จะได้สมการคือ

$$a_r = \left[(n+1)^2 - r^2\right]\left[(n+2)^2 - r^2\right]..\left[(n+z)^2 - r^2\right]\left[h + k\left(n^2 - r^2\right)\right]$$

ซึ่งค่า h และ k สามารถหาได้จากผลเฉลยของระบบสมการพีชคณิตเชิงเส้น ด้วยกฎของคราเมอร์ดังสมการ (2.11)

สำหรับค่า z และ n ในการวิจัยนี้ จะทำการศึกษาที่ $z = 0, 1, 2, 3$ และ 4 ตามลำดับ โดยที่ $n = 1, 2, \ldots, 10$ และเลือกค่า z และ n ที่ให้เปอร์เซ็นต์ค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ต่ำสุด หลังจากนั้นสามารถหาค่าประมาณของค่าq_x ที่ปรับแล้วได้จากสมการ(2.4)

5.2 การปรับค่าโดยใช้รูปแบบฟังก์ชัน

สำหรับการวิจัยนี้ได้นำเสนอรูปแบบฟังก์ 2 รูปแบบคือ รูปแบบไวบูลล์และรูปแบบกอมเพริตซ์ เพื่อให้สอดคล้องกับการแจกแจงของระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคต โดยแทนค่าประมาณของค่า q_x ที่ได้ด้วยรูปแบบฟังก์ชันทั้งสอง แล้วจึงเลือกรูปแบบฟังก์ชันที่เหมาะสม ซึ่งพิจารณาจากค่า MAPE ที่มีค่าต่ำสุด

5.2.1 รูปแบบไวบูลล์

ภายใต้กำลังของมรณะของรูปแบบไวบูลล์จะได้ว่า

$$\mu_x = kx^n \quad , k > 0 , n > 0$$

สามารถหาค่าพารามิเตอร์ k และ n ได้จากการแปลงรูป และใช้วิธีกำลังสองน้อยที่สุด โดยการหาอนุพันธ์ของฟังก์ชัน SS เทียบกับค่าพารามิเตอร์แต่ละค่า แล้วให้สมการอนุพันธ์มีค่าเท่ากับศูนย์ ดังแสดงในสมการ (2.13) ถึง (2.17) จากนั้นสามารถหาค่าประมาณของค่า q_x ที่ปรับแล้วได้จากสมการ (2.18)

5.2.2 รูปแบบกอมเพริตซ์

ภายใต้กำลังของมรณะของรูปแบบกอมเพริตซ์จะได้ว่า

$$\mu_x \quad = \quad Bc^x \qquad , B > 0 \, , c > 1$$

สามารถหาค่าพารามิเตอร์ B และ c ได้จากการแปลงรูป และใช้วิธีกำลังสองน้อยที่สุด โดยการหาอนุพันธ์ของฟังก์ชัน SS เทียบกับค่าพารามิเตอร์แต่ละค่า แล้วให้สมการอนุพันธ์มีค่าเท่ากับศูนย์ ดังแสดงในสมการ (2.20) ถึง (2.24) จากนั้นสามารถหาค่าประมาณของค่า q_x ที่ปรับแล้วได้จากสมการ (2.25)

5.3 การปรับค่าโดยใช้ส่วนโค้งพหุนามองศาสาม

การปรับค่าประมาณด้วยวิธีนี้ จะแบ่งช่วงอายุที่สนใจศึกษาออกเป็นช่วงย่อย ๆ สำ หรับการเลือกจุดแบ่งช่วงนั้น จะพิจารณาจากลักษณะการเคลื่อนไหวของค่า q'_x ตลอดช่วงอายุ (0-99 ปี) โดยทำการแบ่ง ณ ช่วงที่ค่า q'_x เริ่มมีการเปลี่ยนรูป กล่าวคือช่วงที่ค่า q'_x เริ่มมีค่าราบเรียบขึ้นหรือมีค่าขึ้นลงไม่มากนัก ซึ่งจากการวิจัยนี้พบว่าในช่วงอายุประมาณ 70 ขึ้นไป ค่า q'_x จะมีลักษณะค่อนข้างราบเรียบแล้ว จึงไม่มีการเลือกจุดแบ่งใดที่ช่วงอายุนี้

สำหรับการวิจัยนี้การเลือกจุดใดเป็นจุดแบ่งช่วงของอายุ จะแยกพิจารณาออกเป็น 2 กรณีดังนี้ กรณีแรกเมื่อ T มีการแจกแจงแบบไวบูลล์และแบบกอมเพริตซ์ ส่วน W มีการแจกแจงแบบสม่ำเสมอ กรณีที่สองเมื่อ T มีการแจกแจงแบบไวบูลล์และแบบกอมเพริตซ์ ส่วน W มีการแจกแจงแบบแกมมา พบว่าที่ขนาดตัวอย่างทุกระดับลักษณะการเคลื่อนไหวของค่า q'_x ในแต่ละกรณีจะมีรูปแบบไม่แตกต่างกันมากนัก จึงตั้งข้อสังเกตว่าในการเลือกจุดใดเป็นจุดแบ่งช่วงควรจะมีค่าใกล้เคียงกัน และจากการทดลองแบ่งช่วงอายุของค่า q'_x ที่ทุกระดับขนาดตัวอย่าง ซึ่งทำการเลือกจุดแบ่งที่อายุ 20, 40, 50 และ 75 ปีในกรณีแรก และที่อายุ 25, 30, 40, 50 และ 70 ปีในกรณีที่สอง พบว่าค่า MAPE ที่ได้มีค่าไม่แตกต่างกันมากนัก โดยประสิทธิภาพในการปรับค่ายังคงเป็นไปตามสมมติฐานที่ตั้งไว้ ดังนั้นจึงเลือกจุดแบ่งช่วงอายุตามที่ได้กล่าวข้างต้นในการปรับค่าประมาณนี้

จากช่วงอายุที่สนใจศึกษาจะแบ่งออกเป็นช่วงย่อย $(n+1)$ ช่วง ที่อายุ $x = k_1 , k_2 , \ldots , k_n$ แต่ละช่วงย่อยนี้จะแทนด้วยพหุนามองศาสาม ซึ่งมีสมการคือ

$$p_i(x) = p_0(x)+c_5(x-k_1)^3+...+c_{i+4}(x-k_i)^3 \quad , i = 1, 2 ,..., k_n$$

โดยที่ $p_0(x) = c_1 + c_2 x + c_3 x^2 + c_4 x^3$

สามารถหาค่าพารามิเตอร์ $c_1 , c_2 , ... , c_{i+4}$ ได้จากวิธีการกำลังสองน้อยที่สุด โดยการหาอนุพันธ์ของฟังก์ชัน SS เทียบกับค่าพารามิเตอร์แต่ละค่า แล้วให้สมการอนุพันธ์มีค่าเท่ากับศูนย์ ดังแสดงในสมการ (2.27) ถึง (2.28) จากนั้นสามารถหาค่าประมาณของค่า q_x ที่ปรับแล้วได้จากสมการ (2.26)

6. เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์ระหว่างค่า q_x กับค่าประมาณของค่า q_x ที่ปรับแล้วทั้ง 3 วิธีที่แต่ละอายุ พร้อมทั้งเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ของแต่ละวิธี เพื่อหาวิธีการปรับค่าที่ดีที่สุด

การหาค่าความคลาดเคลื่อนสัมบูรณ์ และค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ จะแสดงในรูปของเปอร์เซ็นต์ซึ่งมีค่าดังนี้

$$\text{APE} = \frac{|q_x - q_x''|}{q_x} * 100\ \%$$

$$\text{MAPE} = \frac{1}{v}\sum_x \text{APE}$$

เมื่อ APE คือเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (Absolute Percent Error)

MAPE คือค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Percent Error)

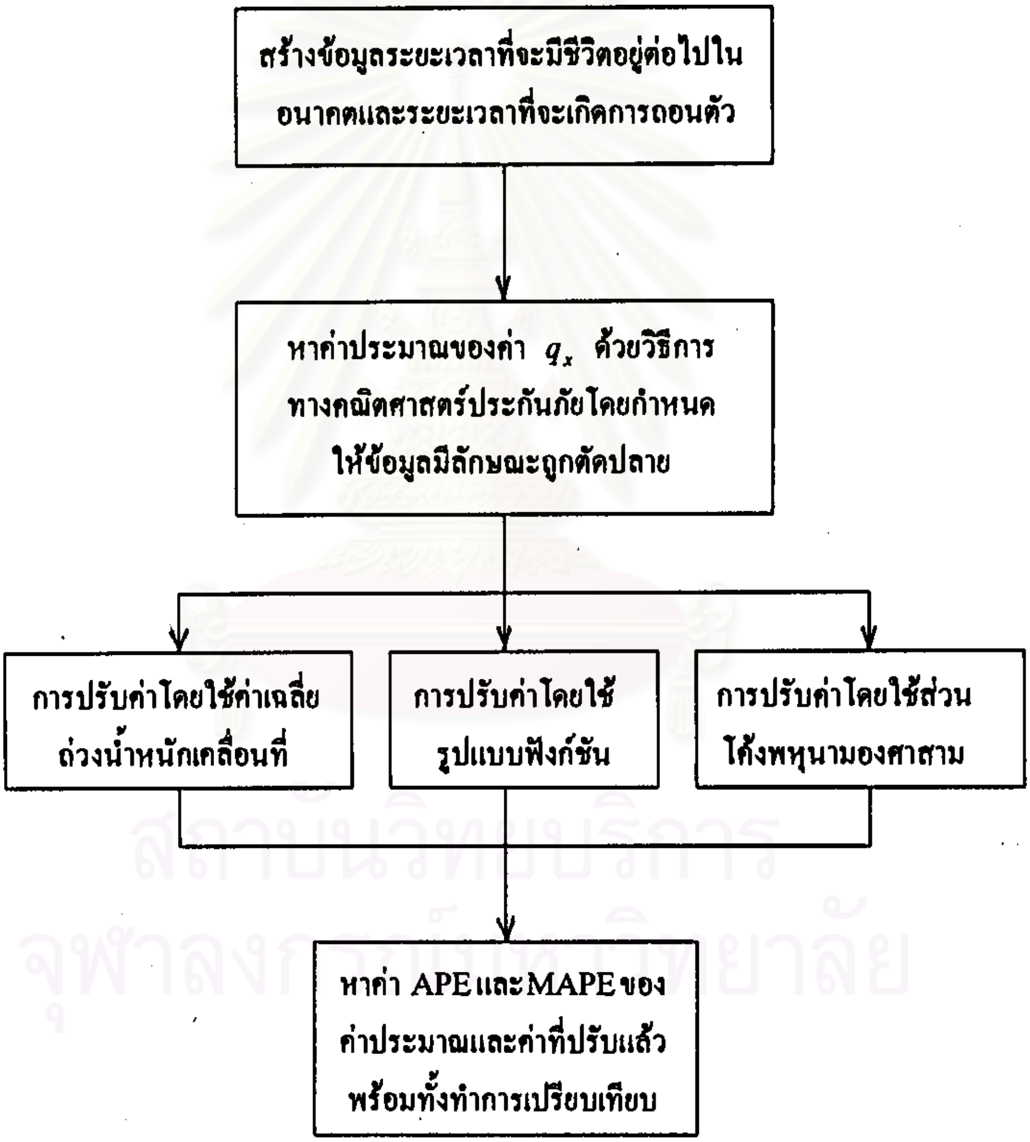
v คือจำนวนค่าของอายุในช่วงที่สนใจศึกษา

<u>โปรแกรมที่ใช้ในการวิจัย</u>

โปรแกรมที่ใช้ในการวิจัยนี้ เขียนด้วยภาษาฟอร์แทรน (FORTRAN) โดยใช้กับเครื่อง AMDAHL 5860 ซึ่งเป็นโปรแกรมการหาค่าประมาณของค่า q_x และการปรับค่าทั้ง 3 วิธี สำหรับรายละเอียดของโปรแกรมได้แสดงไว้ในภาคผนวก

<u>ขั้นตอนการทำงานของโปรแกรม</u>

การหาค่าประมาณของค่า q_x และการปรับค่าทั้ง 3 วิธี แสดงได้ดังแผนผังต่อไปนี้



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย