บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ต้องการศึกษาและเปรียบเทียบวิธีการประมาณความน่าจะเป็นที่คนอายุ x ปี จะยังคงอยู่รอดภายใน 1 ปี ข้างหน้า (p_x) สำหรับข้อมูลที่ไม่สมบูรณ์ วิธีการประมาณค่าทางสถิติที่เสนอในที่นี้มี 4 วิธี คือ

1. วิธีทางคณิตศาสตร์ ประกันภัย (The Actuarial Method)
2. ตัวประมาณ เอ (Estimator A)
3. ตัวประมาณ บี (Estimator B)
4. ตัวประมาณ ซี (Estimator C)

วิธีทางคณิตศาสตร์ประกันภัย (The Actuarial Method) เป็นวิธีการประมาณค่าอย่างง่าย โดยอาศัยหลักการเกี่ยวกับความน่าจะเป็นเบื้องต้น ตัวประมาณ เอ และ บี เป็นการประมาณค่าโดยอาศัยสมมติฐานที่ว่า การออกจากช่วงเกิดขึ้นอย่างสุ่ม ส่วนตัวประมาณ ซี นั้นต้องทราบระยะเวลาที่เสียชีวิตและระยะเวลาที่ออกจากกลุ่ม

ในแต่ละวิธีจะประมาณค่า p_x สำหรับอายุในช่วง 25-65 ปี ภายใต้สถานการณ์เมื่อขนาดตัวอย่างมี 4 ระดับ 300 , 500 , 700 และ 1,000 ตามลำดับ และ เปอร์เซ็นต์การออกจากช่วงมี 4 ระดับ คือ 10% , 20% , 30% และ 40% โดยทำซ้ำ 2,000 ครั้งในแต่ละสถานการณ์

การวิจัยครั้งนี้ใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Technique) สร้างสถานการณ์ต่างๆ ดังนั้น จะอธิบายถึงวิธีการจำลองโดยใช้เทคนิคมอนติคาร์โล แล้วจึงแสดงรายละเอียดของแผนการทดลอง ขั้นตอนการวิจัย และโปรแกรมที่ใช้ในการวิจัยตามลำดับ

วิธีการจำลองโดยใช้เทคนิคมอนติคาร์โล

เทคนิคที่ใช้แก้ปัญหาในการคำนวณทางคณิตศาสตร์นั้นมีอยู่หลายวิธี วิธีการจำลองโดยใช้เทคนิคมอนติคาร์โลเป็นวิธีหนึ่งที่นิยมนำมาใช้ในการแก้ปัญหากันมากในปัจจุบัน ซึ่งหลักการของการจำลองโดยใช้เทคนิคดังกล่าว จะใช้ตัวเลขสุ่ม (Random Number) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา

ขั้นตอนของวิธีการจำลองด้วยเทคนิคมอนติคาร์โลแบ่งได้เป็น 3 ขั้นตอนคือ

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

2 การนำตัวเลขสุ่มมาประยุกต์ใช้กับปัญหาที่ต้องการจะศึกษา ซึ่งขั้นตอนนี้ขึ้นอยู่กับลักษณะของปัญหา บางปัญหาอาจจะไม่ใช้ตัวเลขสุ่มโดยตรง แต่จะนำไปผลิตตัวเลขสุ่มที่มีการแจกแจงแบบอื่นต่อไป

3 การทดลองกระทำ เมื่อนำตัวเลขสุ่มมาประยุกต์ให้เข้ากับปัญหาที่ต้องการศึกษาได้แล้ว ขั้นตอนต่อไปคือ การทดลองโดยใช้กระบวนการของการสุ่ม (Random Process) มากระทำในลักษณะซ้ำๆ กันหลายๆ ครั้ง เพื่อหาคำตอบที่ต้องการ

แผนการทดลอง

การวิจัยครั้งนี้ต้องการประมาณความน่าจะเป็นที่คนอายุ x จะมีชีวิตอยู่รอดต่อไปอีก 1 ปีข้างหน้าสำหรับข้อมูลที่ไม่สมบูรณ์ หรือเป็นข้อมูลที่ถูกตัดปลายทางขวา (Censored Data) โดยอาศัยวิธีการทางสถิติ 4 วิธี ในการคำนวณหาค่าประมาณ p_x ภายใต้ข้อมูลเกี่ยวกับระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคตมีการแจกแจง 2 แบบ และ ระยะเวลาที่จะออกจากกลุ่มมีการแจกแจง 2 แบบ และได้กำหนดเปอร์เซ็นต์การออกจากกลุ่มมี 4 ระดับ และช่วงอายุที่ศึกษามี 41 อายุ และขนาดตัวอย่างที่นำมาศึกษามี 4 ระดับ ซึ่งรวมทั้งสิ้น 2,624 สถานการณ์ ผู้วิจัยจะเปรียบเทียบค่าประมาณ p_x ที่ได้จาก 4 วิธี โดยพิจารณาค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Percentage Error , MAPE) เพื่อหาวิธีที่ดีที่สุดในแต่ละสถานการณ์ต่อไป รายละเอียดของแผนการทดลองมีดังนี้

1 ข้อมูลเกี่ยวกับระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคต มีการแจกแจง 2 แบบ คือ

1.1 การแจกแจงแบบไวบูลล์

1.2 การแจกแจงแบบกอมเพิรตซ์

2 ข้อมูลเกี่ยวกับระยะเวลาที่จะออกจากช่วง มีการแจกแจง 2 แบบ คือ

2.1 การแจกแจงแบบสม่ำเสมอ

2.2 การแจกแจงแบบเบต้า

3 ขนาดตัวอย่างที่นำมาศึกษามี 4 ระดับคือ 300 , 500 , 700 และ 1,000 ตามลำดับ

4 ผู้วิจัยสนใจศึกษาค่า p_x สำหรับอายุ X ในช่วง 25-65 ปี

5 เปอร์เซ็นต์การออกจากช่วงที่ศึกษามี 4 ระดับคือ 10% , 20% , 30% และ 40% ตามลำดับ

<u>ขั้นตอนในการวิจัย</u>

แบ่งเป็น 5 ขั้นตอนหลักดังนี้

1 จำลองระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคต จากการแจกแจงแบบไวบูลล์ และกอมเพิรตซ์

2 จำลองระยะเวลาที่จะออกจากช่วงจากการแจกแจงแบบสม่ำเสมอ และเบต้า

3 กำหนดระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคต ให้มีลักษณะของข้อมูลเป็นแบบถูกตัดปลายทางขวา

4 หาค่าประมาณ p_x ด้วยวิธีการทางสถิติ 4 วิธี

5 หาค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ จากค่าประมาณ p_x ที่ได้ และทำการเปรียบเทียบเพื่อที่จะหาวิธีการประมาณที่ดีที่สุด

รายละเอียดของแต่ละขั้นตอนมีดังนี้

1 การจำลองระยะเวลาที่จะมีชีวิตอยู่ต่อไปภายในอนาคตจากการแจกแจง 2 แบบ คือ การแจกแจงแบบไวบูลล์ และ การแจกแจงแบบกอมเพิรตซ์

1.1 เมื่อ T มีการแจกแจงแบบไวบูลล์

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

โดย t_i เป็นระยะเวลาของคนที่ i จะมีชีวิตอยู่ต่อไปในอนาคต

U เป็นตัวเลขสุ่มซึ่งอยู่ระหว่าง 0 และ 1

k,n เป็นค่าพารามิเตอร์เริ่มต้นสำหรับการจำลองข้อมูล โดยรายละเอียดของค่าพารามิเตอร์ แสดงอยู่ในภาคผนวก

1.2 เมื่อ T มีการแจกแจงแบบกอมเพิรตซ์

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

โดย t_i เป็นระยะเวลาของคนที่ i จะมีชีวิตอยู่ต่อไปในอนาคต

U เป็นตัวเลขสุ่มซึ่งอยู่ระหว่าง 0 และ 1

B,c เป็นค่าพารามิเตอร์เริ่มต้นสำหรับการจำลองข้อมูล โดยรายละเอียดของค่าพารามิเตอร์ แสดงอยู่ในภาคผนวก

2 การจำลองระยะเวลาที่จะออกจากช่วงก่อนสิ้นสุดการศึกษาจากการแจกแจง 2 แบบ คือ การแจกแจงแบบสม่ำเสมอ และการแจกแจงแบบเบต้า โดยให้จำนวนคนที่ออกจากช่วงเป็น 10% , 20% , 30% และ 40% ตามลำดับ

2.1 เมื่อ W มีการแจกแจงแบบสม่ำเสมอ

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

โดย w_i เป็นระยะเวลาของคนที่ i จะออกจากช่วงก่อนสิ้นสุดการศึกษา

U เป็นตัวเลขสุ่มซึ่งอยู่ระหว่าง 0 และ 1

2.2 เมื่อ W มีการแจกแจงแบบเบต้า

กำหนดให้ $\alpha_1 = 5$ และ $\alpha_2 = 1.5$

$$x_1 = \exp\left(1/\sqrt{(2\alpha_1-1)}\ln(U/1-U)\right)\alpha_1$$

$$x_2 = \exp\left(1/\sqrt{(2\alpha_2-1)}\ln(U/1-U)\right)\alpha_2$$

$$w_i = \frac{x_1}{x_1 + x_2}$$

โดย w_i เป็นระยะเวลาของคนที่ i จะออกจากช่วงก่อนสิ้นสุดการศึกษา

U เป็นตัวเลขสุ่มซึ่งอยู่ระหว่าง 0 และ 1

α_1, α_2 เป็นค่าพารามิเตอร์เริ่มต้นสำหรับการจำลองข้อมูล

3 การวิจัยครั้งนี้ได้กำหนดช่วงเวลาที่สนใจจะศึกษา มีระยะเวลาเป็น 1 ปี กำหนดระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคต ซึ่งจำลองได้จากหัวข้อ 1 ให้มีลักษณะของข้อมูลเป็นแบบถูกตัดปลายทางขวา โดยกำหนดให้ $t_i = 1$ เมื่อ $t_i > 1$ และระยะเวลาการออกจากกลุ่มก่อนสิ้นช่วงการศึกษา (w_i) ซึ่งจำลองได้จากหัวข้อ 2 แล้วแต่เหตุการณ์ใดจะเกิดขึ้นก่อนกัน ถ้าผู้ใดมีระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคตเกิน 1 ปี ให้ถือว่าผู้นั้นเป็นผู้อยู่รอดเมื่อสิ้นสุดการศึกษา เช่น $w_i = t_i = 1$ ถ้า $t_i < w_i$ ให้ถือว่าผู้นั้นเป็นผู้ที่เสียชีวิต ถ้า $w_i < t_i$ ให้ถือว่าผู้นั้นเป็นผู้ที่ออกจากกลุ่มก่อนสิ้นช่วงศึกษา ถ้า $w_i < t_i$ และ $t_i < 1$ ให้ถือว่าผู้นั้นเป็นผู้ที่เสียชีวิตของกลุ่มคนที่ออกจากช่วงก่อนสิ้นช่วงศึกษาจนเมื่อได้จำนวนคนที่ออกจากช่วงก่อนสิ้นช่วงศึกษาจนครบตามแต่ละเปอร์เซ็นต์การออกจากช่วง ซึ่ง มี 4 ระดับ คือ 10% , 20% , 30% และ 40% หลังจากนั้นก็จะจำลองข้อมูลเฉพาะระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคต นั่นคือ ถ้า $t_i = 1$ ให้ถือว่าผู้นั้นเป็นผู้อยู่รอดเมื่อสิ้นสุดการศึกษา และถ้า $t_i < 1$ ให้ถือว่าผู้นั้นได้เสียชีวิตก่อนสิ้นช่วงศึกษา

4 หาค่าประมาณ p_x ด้วยวิธีการทางสถิติ 4 วิธี มีขั้นตอนดังนี้

4.1 <u>การประมาณค่า p_x ด้วยวิธีทางคณิตศาสตร์ประกันภัย</u>

1) หาจำนวนผู้ที่เสียชีวิต (D_x) ในช่วงเวลาที่สนใจศึกษา จากจำนวนผู้เสี่ยงภัยทั้งหมด และหาจำนวนผู้ที่ออกจากกลุ่มก่อนสิ้นสุดช่วงการศึกษา (w_x)

2) หาค่าประมาณ p_x จากสมการที่ 2.1

4.2 <u>การประมาณค่า p_x ด้วยวิธีตัวประมาณ เอ</u>

1) หาจำนวนผู้ที่เสียชีวิต (d_x และ d'_x) ในช่วงเวลาที่สนใจศึกษาจากจำนวนผู้เสี่ยงภัยทั้งหมด หาจำนวนผู้ที่ออกจากกลุ่มก่อนสิ้นสุดช่วงการศึกษา (w_x) และหาจำนวนผู้ที่อยู่รอดจนสิ้นช่วงการศึกษา (s_x)

2) หาค่าประมาณ p_x จากสมการที่ 2.6

และอาศัยเทคนิคการวิเคราะห์เชิงตัวเลขโดยใช้วิธีนิวตัน-ราฟซ์สันในการแก้สมการ

4.3 การประมาณค่า p_x ด้วยวิธีตัวประมาณ บี

1) หาจำนวนผู้ที่เสียชีวิตในช่วงเวลาที่สนใจศึกษา จากจำนวนผู้เสี่ยงภัยที่ออกจากช่วงก่อนสิ้นช่วงการศึกษา (d'_x) หาจำนวนผู้ที่ออกจากกลุ่มก่อนสิ้นสุดช่วงการศึกษา (w_x) และหาจำนวนผู้ที่อยู่รอดจนสิ้นช่วงการศึกษา (s_x)

2) หาค่าประมาณ p_x จากสมการที่ 2.10

4.4 การประมาณค่า p_x ด้วยวิธีตัวประมาณ ซี

1) หาจำนวนผู้ที่อยู่รอดจนสิ้นช่วงการศึกษา (s_x) หาระยะเวลาที่เสียชีวิตและหาระยะเวลาที่ออกจากช่วงก่อนสิ้นช่วงการศึกษา

2) หาค่าประมาณ p_x จากสมการที่ 2.16

5 ในการทดลองได้จำลองข้อมูลซ้ำกัน 2,000 ครั้ง ในแต่ละสถานการณ์ กำหนดให้ i แทนรอบที่ทำซ้ำ i = 1 , 2 , ... , 2,000 ดังนั้น

$$\hat{p}_x = \frac{1}{2,000}\sum_{i=1}^{2,000}\hat{p}_x^{(i)}$$

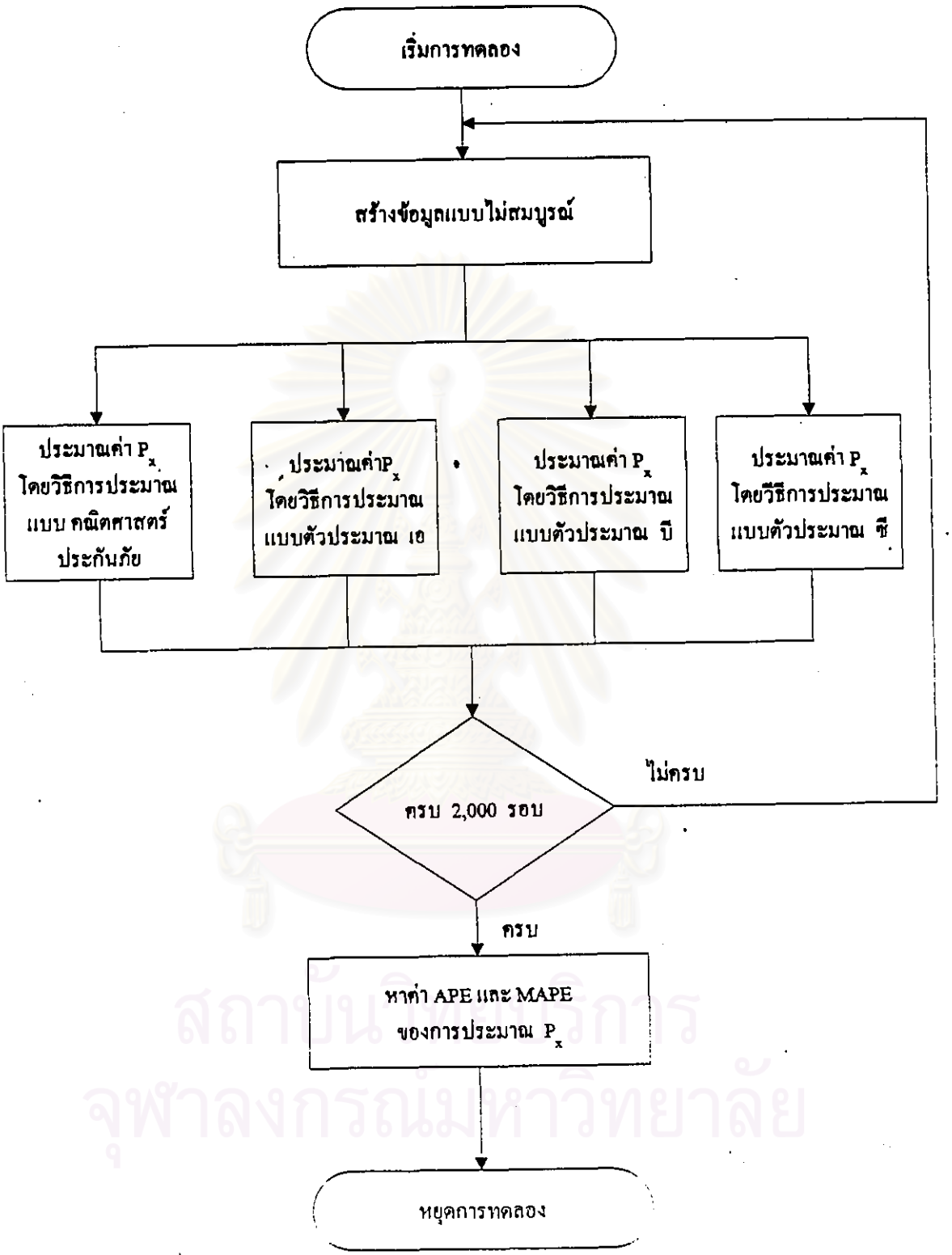
6 คำนวณค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ดังสมการต่อไปนี้

$$\text{MAPE} = \frac{1}{41}\sum_{i=25}^{65}\left|\frac{p_x - \hat{p}_x}{p_x}\right| \times 100\%$$

จากนั้นจะนำค่า MAPE ของการประมาณค่า p_x ของวิธีการทั้ง 4 มาเปรียบเทียบเพื่อหาว่าวิธีการใดให้ค่า MAPE ต่ำที่สุด วิธีการนั้นจะเป็นวิธีที่ประมาณค่า p_x ที่ดีที่สุดของแต่ละสถานการณ์

โปรแกรมที่ใช้ในการวิจัย

โปรแกรมที่ใช้ในการวิจัยทั้งหมดนี้ เขียนด้วยภาษาฟอร์แทรน (FORTRAN) ซึ่งในแต่ละสถานการณ์ของการทดลอง ลักษณะการทำงานของโปรแกรมจะเหมือนกัน สำหรับรายละเอียดของโปรแกรมจะแสดงไว้ในภาคผนวก



รูปที่ 3.1 แสดงผังงานสำหรับหาค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) จากการประมาณด้วย วิธีการทั้ง 4 วิธี

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย