บทที่ 1

บทนำ

ที่มาและความสำคัญของปัญหา

ในงานวิจัยด้านการแพทย์ ด้านวิศวกรรมและอุตสาหกรรม ส่วนใหญ่จะเป็นการศึกษาเกี่ยวกับเวลาที่เริ่มต้นศึกษาสิ่งที่สนใจไปจนกระทั่งเกิดเหตุการณ์ที่ต้องการศึกษาขึ้นกับสิ่งนั้นเรียกว่าเวลาการอยู่รอด (Survival time) เช่นในงานด้านอุตสาหกรรมสนใจศึกษาเวลาที่เริ่มใช้หลอดไฟใหม่ไปจนกระทั่งหลอดไฟเสีย ในงานด้านการแพทย์สนใจศึกษาเวลาที่คนไข้โรคมะเร็งเม็ดเลือดขาวเริ่มได้รับการรักษาจากโรงพยาบาลไปจนกระทั่งเสียชีวิต การศึกษาวิเคราะห์ข้อมูลเวลาการอยู่รอดสามารถใช้วิธีการทางสถิติเพื่อประมาณฟังก์ชันการอยู่รอด (Survival Funetion) ทั้งที่เป็นตัวประมาณแบบจุดและแบบช่วง

ในทางปฏิบัติแล้วพบว่าข้อมูลส่วนใหญ่เป็นข้อมูลที่มีค่าที่ถูกตัดทิ้ง (Censored Data) คือไม่สามารถทราบถึงเวลาการอยู่รอดที่แน่นอนได้ ซึ่งอาจเนื่องมาจากสาเหตุต่าง ๆ สำหรับในงานด้านอุตสาหกรรมข้อมูลส่วนใหญ่จะเป็นข้อมูลที่มีค่าที่ถูกตัดทิ้งทางขวา เช่นสนใจศึกษาเวลาที่เริ่มใช้หลอดไฟใหม่จนกระทั่งหลอดไฟเสีย ซึ่งถ้าต้องการเก็บข้อมูลทั้งหมดจะใช้เวลานาน ดังนั้นจะมีการกำหนดเวลาของการสิ้นสุดการเก็บข้อมูลเอาไว้ล่วงหน้า ส่วนในงานด้านการแพทย์ข้อมูลส่วนใหญ่จะเป็นข้อมูลที่มีค่าที่ถูกตัดทิ้งแบบสุ่ม เช่นต้องการศึกษาถึงระยะเวลาที่คนไข้โรคมะเร็งเม็ดเลือดขาวเข้ารับการรักษากับทางโรงพยาบาลจนกระทั่งเสียชีวิต โดยจะทำการเก็บข้อมูลตั้งแต่คนไข้เข้ารับการรักษาจนกระทั่งคนไข้เสียชีวิต ซึ่งจะใช้เวลาในการเก็บข้อมูลเป็นเวลานาน ฉะนั้นจะมีการกำหนดเวลาของการสิ้นสุดการเก็บข้อมูลเอาไว้ล่วงหน้า เมื่อเก็บข้อมูลถึงเวลาที่กำหนดก็จะหยุดเก็บข้อมูล นอกจากนี้ในระหว่างช่วงเวลาที่ทำการเก็บข้อมูลอาจมีคนไข้บางคนที่ขาดการติดต่อกับทางโรงพยาบาลทำให้การเก็บข้อมูลของคนไข้ไม่สามารถทำได้ หรือคนไข้เสียชีวิตเนื่องจากสาเหตุอื่นนอกจากการเสียชีวิตด้วยโรคมะเร็งเม็ดเลือดขาว ข้อมูลที่เก็บมาสามารถนำมาวิเคราะห์หาความน่าจะเป็นที่คนไข้จะมีชีวิตอยู่รอดมากกว่า t โดยการประมาณฟังก์ชันการอยู่รอด $S(t) = Prob(T>t)$ เมื่อ T คือ ตัวแปรสุ่มระยะเวลาที่คนไข้จะมีชีวิตอยู่รอด

ส่วนในงานด้านการประกันชีวิต การศึกษาฟังก์ชันการอยู่รอดของกรมธรรม์ประกันภัยจะมีประโยชน์ในด้านการนำไปใช้คำนวณอัตราเบี้ยประกันภัย การคำนวณเงินสำรองและอื่น ๆ ซึ่งถือว่าเป็นหน้าที่ที่สำคัญของนักคณิตศาสตร์ประกันภัย การคำนวณเบี้ยประกันภัยที่เหมาะสมจะทำให้บริษัทมีเงินเพียงพอที่จะจ่ายเงินผลประโยชน์ตามกรมธรรม์และเพียงพอต่อค่าใช้จ่ายในการดำเนินงาน ถ้าอัตราเบี้ยประกันภัยที่กำหนดมีค่าต่ำเกินไปจะมีผลต่อความมั่นคงของบริษัท แต่ถ้าอัตราเบี้ยประกันภัยที่กำหนดมีค่าสูงเกินไปจะมีผลทำให้บริษัทอาจสูญเสียตลาดให้แก่คู่แข่งได้ การที่จะได้มาซึ่งอัตราเบี้ยประกันภัยที่เหมาะสมใกล้เคียงความเป็นจริงนั้นต้องใช้ข้อมูลที่มีลักษณะสมบูรณ์ (Complete Data) แต่ในทางปฏิบัติพบว่าข้อมูลส่วนใหญ่เป็นข้อมูลที่ไม่สมบูรณ์ (Incomplete Data) ทำให้อัตราเบี้ยประกันภัยที่กำหนดมีค่าสูงหรือต่ำเกินไป ดังนั้นจึงควรที่จะศึกษารูปแบบการแจกแจงและการประมาณฟังก์ชันการอยู่รอดให้ใกล้เคียงกับความเป็นจริงมากที่สุดเพื่อให้การคำนวณเบี้ยประกันภัยได้มีความถูกต้องยิ่งขึ้น ลักษณะข้อมูลในงานด้านการประกันภัย เช่นศึกษาถึงระยะเวลาของการอยู่รอดของกรมธรรม์ประกันภัยในช่วงเวลา 3 ปี หรือ 5 ปี ว่ามีกรมธรรม์ใดบ้างที่ผู้เอาประกันภัยเสียชีวิตในช่วงเวลาที่ศึกษา หรือเมื่อครบกำหนดช่วงเวลาที่ศึกษาจะมีกรมธรรม์ที่อยู่รอดเท่าใด ซึ่งในการเก็บข้อมูลอาจมีผู้เอาประกันบางคนที่ขาดการส่งเบี้ยประกันมีผลทำให้กรมธรรม์ขาดอายุ (Lapse) ดังนั้นข้อมูลที่เก็บมาจะเป็นข้อมูลที่ไม่สมบูรณ์

การประมาณฟังก์ชันการอยู่รอดสามารถทำได้ทั้งที่เป็นตัวประมาณแบบจุดและแบบช่วงสำหรับในงานวิจัยนี้สนใจการประมาณฟังก์ชันการอยู่รอดแบบช่วง ภายใต้เงื่อนไขต่าง ๆ ผู้วิจัยจะศึกษาหาวิธีการประมาณแบบช่วงที่เหมาะสมสำหรับฟังก์ชันการอยู่รอด (Survival Function) โดยจะทำการเปรียบเทียบวิธีการประมาณฟังก์ชันการอยู่รอดแบบช่วง 3 วิธี คือ

1. วิธีการประมาณช่วงความเชื่อมั่นความเที่ยงเท่ากัน (Equal Precision (EP))
2. วิธีการประมาณช่วงความเชื่อมั่นฮอล-เวลเนอร์ (Hall-Wellner (HW))
3. วิธีการประมาณช่วงความเชื่อมั่นเรนยี (Renyi (R))

สำหรับวิธีการประมาณช่วงความเชื่อมั่นความเที่ยงเท่ากันได้ถูกนำเสนอโดย V.N. Nair (ค.ศ. 1981) ซึ่งทำการประมาณช่วงความเชื่อมั่นเมื่อข้อมูลมีค่าที่ถูกตัดทิ้งแบบสุ่มและกลุ่มตัวอย่างมีขนาดใหญ่ ส่วนวิธีการประมาณช่วงความเชื่อมั่นฮอล-เวลเนอร์ได้ถูกนำเสนอโดย W.J. Hall and Jon A. Wellner (ค.ศ. 1980) ซึ่งเป็นวิธีที่พัฒนาขึ้นมาจากช่วงความเชื่อมั่นของ Kolmogorov ศึกษาในกรณีที่ข้อมูลมีค่าที่ถูกตัดทิ้งแบบสุ่ม และวิธีการประมาณช่วงความเชื่อมั่นเรนยีได้ถูกนำเสนอโดย Alfred Renyi (ค.ศ. 1953)

วัตถุประสงค์ของการวิจัย

เพื่อศึกษาเปรียบเทียบช่วงความเชื่อมั่นสำหรับฟังก์ชันการอยู่รอดเมื่อข้อมูลมีค่าที่ถูกตัดทิ้ง 3 วิธี คือ

1. วิธีการประมาณช่วงความเชื่อมั่นความเที่ยงเท่ากัน (Equal Precision (EP))
2. วิธีการประมาณช่วงความเชื่อมั่นฮอล-เวลเนอร์ (Hall-Wellner (HW))
3. วิธีการประมาณช่วงความเชื่อมั่นเรนยี (Renyi (R))

การเปรียบเทียบจะทำการเปรียบเทียบด้วยค่าระดับความเชื่อมั่นของช่วงความเชื่อมั่นและค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่คำนวณจากแต่ละวิธีการประมาณที่ระดับค่าสัมประสิทธิ์ความเชื่อมั่น 3 ระดับคือ 90%,95% และ 99%

สมมติฐานของการวิจัย

1. ในกรณีที่ฟังก์ชันการอยู่รอดมีค่าน้อย ๆ (เข้าใกล้ 0) หรือมีค่ามาก (เข้าใกล้ 1) ช่วงความเชื่อมั่นที่ประมาณด้วยวิธีความเที่ยงเท่ากัน (Equal Precision (EP)) จะให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นต่ำกว่าวิธีอื่น ๆ

2. ส่วนในกรณีที่ฟังก์ชันการอยู่รอดมีค่าปานกลางช่วงความเชื่อมั่นที่ประมาณด้วยวิธีฮอล-เวลเนอร์ (Hall-Wellner (HW)) จะให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นต่ำกว่าวิธีอื่น ๆ

ข้อตกลงเบื้องต้น

1. ศึกษาในกรณีที่การแจกแจงเป็นแบบเบ้ขวา เนื่องจากการแจกแจงการอยู่รอดโดยทั่วไปแล้วพบว่ามักจะเป็นแบบเบ้

2. การแจกแจงของค่าที่ถูกตัดทิ้ง และค่าที่ไม่ถูกตัดทิ้งเป็นอิสระต่อกัน

3. ศึกษากรณีของข้อมูลเป็นประเภทมีค่าที่ถูกตัดทิ้งแบบสุ่ม

4. ศึกษากรณีประมาณฟังก์ชันการอยู่รอด (S(t)) เป็นแบบช่วง ณ เวลาของการอยู่รอด (t) มีค่าตั้งแต่ 0.25 ถึง 3.00 โดยเพิ่มขึ้นทีละ 0.25

ขอบเขตของการวิจัย

1. ในการวิจัยครั้งนี้ทำการศึกษาเปรียบเทียบการประมาณค่าช่วงความเชื่อมั่นของฟังก์ชันการอยู่รอดเมื่อข้อมูลมีค่าที่ถูกตัดทิ้งแบบสุ่มด้วยวิธี

1. วิธีการประมาณช่วงความเชื่อมั่นความเที่ยงเท่ากัน (Equal Precision (EP))
2. วิธีการประมาณช่วงความเชื่อมั่นฮอล-เวลเนอร์ (Hall-Wellner (HW))
3. วิธีการประมาณช่วงความเชื่อมั่นเรนยี (Renyi (R))

2. ศึกษาในกรณีที่ข้อมูลไม่ถูกตัดทิ้ง $T_1^0, T_2^0, \ldots, T_n^0$ เป็นอิสระกันและมีการแจกแจงเหมือนกันมีรูปแบบเป็น

2.1. การแจกแจงไวบูลล์ (Weibull Distribution)

ฟังก์ชันความหนาแน่นอยู่ในรูป

$$f(t) = \begin{cases} \lambda\gamma(\lambda t)^{\gamma-1}\exp\left(-(\lambda t)^{\gamma}\right) & ; \ 0<t<\infty,\ \lambda>0,\ \gamma>0 \\ 0 & ; \ \text{อื่น ๆ} \end{cases}$$

ในการวิจัยครั้งนี้จะศึกษาที่ $\lambda = 1.00$ และ $\gamma = 1.50$[1]

2.2. การแจกแจงลอกนอร์มอล (Lognormal Distribution)

ฟังก์ชันความหนาแน่นอยู่ในรูป

$$f(t) = \begin{cases} \dfrac{1}{t\sigma\sqrt{2\pi}}\exp\left(-\dfrac{(\ln(t)-\mu)^2}{2\sigma^2}\right) & ; \ t>0,\ -\infty<\mu<\infty,\ \sigma>0 \\ 0 & ; \ \text{อื่น ๆ} \end{cases}$$

ในการวิจัยครั้งนี้จะศึกษาที่ $\mu = 0.0$ และ $\sigma^2 = 0.7$[1]

2.3. การแจกแจงกอมเพริตซ์ (Gompertz Distribution)

ฟังก์ชันความหนาแน่นอยู่ในรูป

$$f(t) = \begin{cases} BC^t\exp\left(\dfrac{B}{\ln(C)}(1-C^t)\right) & ; \ t>0,\ B>0,\ C>1 \\ 0 & ; \ \text{อื่น ๆ} \end{cases}$$

---

[1] ผู้วิจัยได้ทดลองด้วยค่าพารามิเตอร์หลาย ๆ ค่าที่ทำให้การแจกแจงมีลักษณะเบ้ขวาแล้วพบว่าผลสรุปไม่แตกต่างกัน

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ในการวิจัยครั้งนี้จะศึกษาที่ B = 0.30 และ C = 3.50[2]

3. ศึกษาในกรณีที่ข้อมูลถูกตัดทิ้ง $C_1, C_2, \ldots, C_n$ เป็นอิสระกันและมีการแจกแจงเหมือนกันมีรูปแบบเป็น

3.1 การแจกแจงแบบสม่ำเสมอ (Uniform Distribution)

ฟังก์ชันความหนาแน่นอยู่ในรูป

$$f(c) = \begin{cases} \frac{1}{b-a} & ; \ a < c < b \\ 0 & ; \ \text{อื่น ๆ} \end{cases}$$

ในการวิจัยครั้งนี้จะศึกษาที่ a = 0 และ b = 3

3.2 การแจกแจงเอกซ์โปเนนเชียล (Exponential Distribution)

ฟังก์ชันความหนาแน่นอยู่ในรูป

$$f(c) = \begin{cases} \lambda \exp(-\lambda c) & ; \ c > 0 \\ 0 & ; \ \text{อื่น ๆ} \end{cases}$$

ในงานวิจัยครั้งนี้จะศึกษาที่ $\lambda = 0.4$

4. ศึกษาในกรณีที่ค่าสัดส่วนของข้อมูลที่ถูกตัดทิ้งเป็น 10% ,20% และ 30% ของขนาดตัวอย่าง

5. ศึกษาในกรณีที่ขนาดตัวอย่างมี 4 ระดับคือ 25 ,50 ,100 และ 200

6. ศึกษาในกรณีที่ค่าสัมประสิทธิ์ความเชื่อมั่นมี 3 ระดับคือ 90% ,95% และ 99%

7. ศึกษาในกรณีที่เวลาของการอยู่รอดมีค่าตั้งแต่ 0.25 ถึง 3.00 โดยเพิ่มขึ้นทีละ 0.25

8. ในการวิจัยครั้งนี้สร้างแบบจำลองข้อมูลโดยใช้เทคนิคมอนติคาร์โล (Monte Carlo Simulation Method) ทำการทดลองซ้ำ 2,000 ครั้งในแต่ละสถานการณ์

คำจำกัดความ

1. สัมประสิทธิ์ความเชื่อมั่น (Confidence coefficient) หมายถึง ความน่าจะเป็นที่ช่วงสุ่มจะครอบคลุมค่าจริงของฟังก์ชันการอยู่รอด

---

[2] ผู้วิจัยได้ทดลองด้วยค่าพารามิเตอร์หลาย ๆ ค่าที่ทำให้การแจกแจงมีลักษณะเบ้ขวาแล้วพบว่าผลสรุปไม่แตกต่างกัน

2. ช่วงความเชื่อมั่น (Confidence interval) หมายถึง ช่วงตัวอย่างที่คำนวณจากข้อมูลตัวอย่างหนึ่งชุดใด ๆ ซึ่งใช้ในการประมาณค่าฟังก์ชันการอยู่รอดแบบช่วง

ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

1. เพื่อเป็นแนวทางให้นักวิจัยมีผลสรุปในการเลือกวิธีการประมาณช่วงความเชื่อมั่นสำหรับฟังก์ชันการอยู่รอดเมื่อข้อมูลมีค่าที่ถูกตัดทิ้งแบบสุ่มที่เหมาะสม

2. เพื่อเป็นแนวทางในการศึกษาและเปรียบเทียบวิธีการประมาณช่วงความเชื่อมั่นสำหรับฟังก์ชันการอยู่รอดเมื่อข้อมูลมีค่าที่ถูกตัดทิ้งในแบบอื่น ๆ ต่อไปเช่นเมื่อข้อมูลมีค่าที่ถูกตัดทิ้งประเภทที่ 1 (Type I Censoring)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย