บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันการวิจัยทางด้านการแพทย์ ด้านวิศวกรรม และอุตสาหกรรม จำเป็นต้องอาศัยวิธีการทางสถิติ ซึ่งส่วนใหญ่จะเป็นการศึกษาเกี่ยวกับเวลาที่เริ่มต้นศึกษาสิ่งที่สนใจไปจนกระทั่งเกิดเหตุการณ์ที่ต้องการศึกษาขึ้นกับสิ่งนั้น เรียกว่า เวลาการอยู่รอด (Survival time) เช่น เวลาที่เริ่มใช้หลอดไฟฟ้าใหม่จนกระทั่งหลอดเสีย เวลาที่คนไข้โรคมะเร็ง เริ่มได้รับการรักษาจากโรงพยาบาลไปจนกระทั่งเสียชีวิต ฯลฯ จากตัวอย่างคนไข้เป็นตัวอย่างการวิจัยทางด้านการแพทย์ เพื่อจะเปรียบเทียบผลของวิธีในการรักษาแบบใหม่กับแบบเดิม ดูว่าวิธีการรักษาแบบใหม่ซึ่งคิดว่าดีกว่าจะให้เวลาการอยู่รอดนานกว่าแบบเดิมจริงหรือไม่ สิ่งที่สนใจคือ คนไข้โรคมะเร็ง เวลาที่เริ่มต้นศึกษาคือ เวลาที่คนไข้เริ่มได้รับการรักษาจากโรงพยาบาล และเหตุการณ์ที่ต้องการศึกษาคือ การเสียชีวิตของคนไข้หลังจากได้รับการรักษา โดยปกติแล้วจะให้เหตุการณ์ที่ต้องการศึกษาเป็น เวลาการอยู่รอด (Surviyal time) และการศึกษาวิเคราะห์ข้อมูลเวลาการอยู่รอดเพื่อตอบปัญหาต่าง ๆ ที่ต้องการทราบ ดังตัวอย่างข้างต้น เรียกว่า การวิเคราะห์การอยู่รอด (Survival analysis)

ซึ่งปัญหาต่าง ๆ เหล่านี้ สามารถตอบได้โดยใช้ฟังก์ชันของเวลาการอยู่รอด (Function of survival time) ได้แก่ ฟังก์ชั่นความหนาแน่น (density function) ฟังก์ชันการอยู่รอด (survival function) และฟังก์ชันความสูญเสีย(failure or hazard function) เช่น ถ้าทราบฟังก์ชันความหนาแน่นของเวลาการอยู่รอด ก็สามารถบอกได้ว่าเวลาการอยู่รอดมีการแจกแจงแบบใด จากตัวอย่างข้างต้น สมมติฟังก์ชันความหนาแน่นของเวลาการอยู่รอดของคนไข้เป็น $f(t) = \lambda \exp(-\lambda t)$ แสดงว่า เวลาการอยู่รอดของคนไข้มีการแจกแจงแบบเอกซ์โปเนนเชียลด้วยพารามิเตอร์ λ

ในทางปฏิบัติแล้วพบว่ามีข้อมูลจำนวนมากที่มีค่าสังเกตบางค่าเป็นค่าสังเกตไม่สมบูรณ์ (incomplete or censored observation) คือ ไม่สามารถทราบถึงเวลาการอยู่รอดที่

แน่นอนได้ ซึ่งอาจเนื่องมาจากสาเหตุต่าง ๆ เช่น หน่วยสังเกตไม่แสดงถึงเหตุการณ์ที่ต้องการศึกษาเมื่อสิ้นสุดการศึกษา หรือคนไข้ขาดการติดต่อกับโรงพยาบาลหลังจากได้รับการรักษาแล้ว นอกจากนั้นข้อมูลส่วนใหญ่มักจะไม่ทราบถึงการแจกแจง หรือข้อมูลที่ใช้ในการทดลองศึกษามีจำนวนน้อยมาก ทำให้ไม่สามารถกำหนดการแจกแจงได้ ดังนั้นในการเลือกใช้ตัวสถิติทดสอบที่เหมาะสมจึงนิยมใช้วิธีนอนพาราเมตริก

ในกรณีที่ต้องการศึกษาถึงการทดสอบการเปรียบเทียบการแจกแจงการอยู่รอด (survival distribution) ของ 2 กลุ่มตัวอย่าง เมื่อมีค่าสังเกตไม่สมบูรณ์นั้น ได้มีนักสถิติหลายท่านเสนอตัวสถิติที่ใช้ในการทดสอบเหตุการณ์ดังกล่าวขึ้น ซึ่งในที่นี้จะกล่าวถึงตัวสถิติทดสอบดังกล่าว 4 ตัว ได้แก่

1. วิธี Cox - Mantel test
2. วิธี Peto and Peto's Generalized Wilcoxon test
3. วิธี Logrank test
4. วิธี Modified Logrank test

ตัวสถิติที่กล่าวมาทั้ง 4 วิธีข้างต้นนี้ เป็นวิธีนอนพาราเมตริก สำหรับข้อมูลที่มีค่าสังเกตไม่สมบูรณ์โดยเฉพาะ โดยที่จะพิจารณาถึงค่าสังเกตไม่สมบูรณ์ 2 กรณีดังนี้

ก. แบบ Right Censored data

ค่าขาดหายประเภทนี้ เกิดจากการที่กำหนดให้มีค่าสังเกตไม่สมบูรณ์ เท่ากับจำนวนที่กำหนดให้ โดยที่ค่าสังเกตไม่สมบูรณ์จะมีค่ามากที่สุด หรืออาจมีค่าเท่ากับค่าสังเกตสมบูรณ์ที่มีค่ามากที่สุด

ข. แบบ Random Censored data

ค่าขาดหายประเภทนี้ส่วนใหญ่จะพบในการทดลองเกี่ยวกับการแพทย์ ซึ่งค่าขาดหายประเภทนี้จะกำหนดให้ข้อมูลมีค่าสังเกตไม่สมบูรณ์เท่ากับจำนวนที่กำหนดขึ้น และค่าสังเกตไม่สมบูรณ์แต่ละค่าอาจมีค่าเท่ากับค่าสังเกตสมบูรณ์ที่มีค่ามากที่สุด ซึ่งเกิดจากหน่วยสังเกตยังไม่เกิดการสูญเสีย

เมื่อสิ้นสุดการศึกษา (ค่าสังเกตไม่สมบูรณ์ประเภทนี้เรียกว่า Singly censored observation)[1] หรือค่าสังเกตไม่สมบูรณ์อาจจมีค่าไม่เท่ากับค่าสังเกตใด ๆ เลยก็ได้ ซึ่งกรณีนี้อาจเกิดจากการที่หน่วยสังเกตขาดการติดต่อหรือหน่วยสังเกตถอนตัวจากการทดสอบ (ค่าสังเกตไม่สมบูรณ์ประเภทนี้เรียกว่า Multiply time cersored observation)[2] ซึ่งจะทำให้ไม่สามารถได้ค่าที่แน่นอนของค่าสังเกตได้

1.2 วัตถุประสงค์ของการวิจัย

เพื่อศึกษาเปรียบเทียบอำนาจการทดสอบของตัวสถิตินอนพาราเมตริก ในการทดสอบการแจกแจงการอยู่รอดใน 2 กลุ่มตัวอย่างย่อย เมื่อมีจำนวนค่าสังเกตไม่สมบูรณ์ในตัวอย่างย่อยต่าง ๆ กัน โดยวิธี

- Cox - Mantel Test
- Peto and Peto's Generalized Wilcoxon Test
- Logrank Test
- Modified Logrank Test

1.3 สมมติฐานในการวิจัย

ขนาดตัวอย่างและเปอร์เซ็นต์ของค่าสังเกตไม่สมบูรณ์ จะมีผลทำให้อำนาจการทดสอบของตัวสถิติทดสอบ Modified Logrank test สูงสุด

1.4 ข้อตกลงเบื้องต้น

ในการศึกษาครั้งนี้ จะถือว่าอำนาจการทดสอบ (Power of test) และความสามารถในการควบคุมความคลาดเคลื่อนแบบที่ 1 เป็นเกณฑ์ในการเลือกตัวสถิติทดสอบ

[1,2] W. Nelson, Applied Life Data Analysis, (John Wiley & Sons : New York, 1982), pp. 7-9.

1.5 ขอบเขตของการวิจัย

1. สำหรับการวิเคราะห์ข้อมูลการอยู่รอด ฟังก์ชั่นการอยู่รอด (Survival Function, S(t)) เป็นสิ่งสำคัญและนำไปใช้ประโยชน์ได้มากที่สุด เพราะฟังก์ชั่นการอยู่รอดสามารถบอกถึงรูปแบบของการแจกแจงการอยู่รอดได้ และเนื่องจากการแจกแจงการอยู่รอดโดยทั่วไปแล้ว พบว่ามักจะเป็นแบบเบ้ (Skewed) ดังนั้นในการศึกษาครั้งนี้จะศึกษาในรูปแบบของการแจกแจงการอยู่รอดที่มักพบอยู่เสมอ ได้แก่

ก. การแจกแจงแบบเอกซ์โปเนนเชียล (Exponential distribution) ฟังก์ชั่นความหนาแน่นอยู่ในรูป

$$f(t) = \begin{cases} \beta e^{-\beta t} & ; \ t \geq 0, \ \beta > 0 \\ 0 & ; \ t < 0 \end{cases}$$

$$E(t) = 1/\beta$$

$$V(t) = 1/\beta^2$$

ในการวิจัยครั้งนี้ จะศึกษาที่ $\beta = 0.5, 1, 2$

ข. การแจกแจงแบบไวบูลล์ (Weibull distribution) ฟังก์ชั่นความหนาแน่นอยู่ในรูป

$$f(t) = \begin{cases} \dfrac{\alpha t^{\alpha} \exp(t/\beta)^{\alpha}}{\beta^{\alpha}} & ; \ \alpha, \beta > 0, \ t > 1 \\ 0 & ; \text{อื่น ๆ} \end{cases}$$

เมื่อ β เป็น Scale paramcter

α เป็น Shape parameter

$$E(t) = \beta\alpha$$

$$V(t) = \beta^2 \alpha$$

ในการวิจัยครั้งนี้ จะศึกษาที่ $\beta = 1$ และ $\alpha = 1, 2, 4$

ค. การแจกแจงแบบลอกนอร์มอล (Lognormal distribution) ฟังก์ชันความหนาแน่นอยู่ในรูป

$$f(t) = \begin{cases} \dfrac{1.e^{-\frac{1}{2}(/og_e t - \mu)/\sigma^2}}{t\sigma\sqrt{2\pi}} ; & t > 0,\ \sigma > 0 \\ 0 ; & \text{อื่น ๆ} \end{cases}$$

$$E(t) = \exp(\mu + \sigma^2/2)$$

$$V(t) = \exp(2\mu + \sigma^2) \cdot (\exp \sigma^2 - 1)$$

เมื่อ μ และ σ^2 เป็นค่าเฉลี่ยและความแปรปรวนของ Y โดยที่ $Y = \log_e t$ และ Y มีการแจกแจงแบบปกติ

ในการวิจัยครั้งนี้จะศึกษาที่ $\mu = 0$ และ $\sigma^2 = 0.1, 0.3, 0.5$

2. ขนาดตัวอย่างที่ใช้ในการศึกษา คือ 20, 50 โดยแต่ละกลุ่มตัวอย่างมีขนาดเท่ากันและมีจำนวนค่าสังเกตไม่สมบูรณ์เท่ากับ 10, 20, 30 เปอร์เซนต์ของขนาดตัวอย่าง

3. กำหนดให้รูปแบบของค่าสังเกตไม่สมบูรณ์เป็นแบบ Right Censored data และแบบ Random Censored data โดยกำหนดให้เปอร์เซนต์ของค่าสังเกตไม่สมบูรณ์ของทั้ง 2 กลุ่มตัวอย่าง ดังนี้

- เปอร์เซนต์ค่าสังเกตไม่สมบูรณ์ในกลุ่มที่ 1 เป็น 10% และในกลุ่มที่ 2 เป็น 10% ของขนาดตัวอย่าง
- เปอร์เซนต์ค่าสังเกตไม่สมบูรณ์ในกลุ่มที่ 1 เป็น 20% และในกลุ่มที่ 2 เป็น 10% ของขนาดตัวอย่าง
- เปอร์เซนต์ค่าสังเกตไม่สมบูรณ์ในกลุ่มที่ 1 เป็น 20% และในกลุ่มที่ 2 เป็น 20% ของขนาดตัวอย่าง
- เปอร์เซนต์ค่าสังเกตไม่สมบูรณ์ในกลุ่มที่ 1 เป็น 30% และในกลุ่มที่ 2 เป็น 20% ของขนาดตัวอย่าง

4. ศึกษาที่ระดับนัยสำคัญ 0.1 และ 0.05

5. การศึกษาครั้งนี้ สร้างแบบจำลองข้อมูลให้มีสถานการณ์ตามต้องการ โดยใช้เทคนิคมอนติคาร์โลซิมูเลชั่น จากเครื่องคอมพิวเตอร์ AMDAHL 3860 เขียนโปรแกรมด้วยภาษาฟอร์แทรน 77 โดยจะทำการทดลองซ้ำประมาณ 500 ครั้ง ในแต่ละสถานการณ์

1.6 คำจำกัดความ

ความคลาดเคลื่อนประเภทที่ 1 (Type I error) เป็นความผิดพลาดที่เกิดจากการปฏิเสธสมมติฐาน H_o เมื่อสมมติฐาน H_o ถูกต้อง

ความคลาดเคลื่อนประเภทที่ 2 (Type II error) เป็นความผิดพลาดที่เกิดจากการยอมรับสมมติฐาน H_o เมื่อสมมติฐาน H_o นั้นผิด

อำนาจการทดสอบ (Power of the test) คือ ความน่าจะเป็นที่จะปฏิเสธสมมติฐาน H_o เมื่อสมมติฐาน H_o ผิด ซึ่งจะมีค่าเท่ากับ $1-\beta$ เมื่อ β คือ ความคลาดเคลื่อนประเภทที่ 2

1.7 ประโยชน์ที่คาดว่าจะได้รับ

ทำให้ทราบถึงวิธีนอนพาราเมตริกที่เหมาะสม สำหรับการทดสอบการเปรียบเทียบการแจกแจงการอยู่รอดของ 2 กลุ่มตัวอย่าง เมื่อข้อมูลมีค่าสังเกตไม่สมบูรณ์