# บทที่ 1

# บทนำ

ความเป็นมาและความสำคัญของปัญหา

ในงานวิเคราะห์ทางสถิติที่ศึกษาเกี่ยวกับเวลาที่เริ่มต้นศึกษาสิ่งที่สนใจไปจนกระทั่งเกิดเหตุการณ์ที่สนใจนั้น จะเป็นการวิเคราะห์เกี่ยวกับเวลาการอยู่รอด (Lifetime) ซึ่งพบได้ในงานวิจัยด้านต่างๆ เช่น ในงานด้านการประกันภัยสนใจศึกษาเวลาการคงอยู่ของกรมธรรม์ ในงานด้านการแพทย์สนใจศึกษาเวลาที่คนไข้เริ่มได้รับการรักษาจนกระทั่งเสียชีวิต และในงานด้านอุตสาหกรรมสนใจศึกษาเวลาที่เครื่องจักรทำงานเป็นปกติจนกระทั่งเครื่องจักรเกิดการขัดข้อง ทั้งนี้ในทางปฏิบัติการเก็บรวบรวมข้อมูลเพื่อใช้ในการวิเคราะห์นั้นบางกรณีจะต้องเสียเวลา และค่าใช้จ่ายมากเพื่อรอจนกว่าเหตุการณ์ที่สนใจจะเกิดขึ้นครบทุกหน่วยตัวอย่างที่ศึกษา ดังนั้นจึงจำเป็นต้องกำหนดเวลาที่จะสิ้นสุดการเก็บรวบรวมข้อมูลไว้ล่วงหน้า ซึ่งมีผลทำให้ข้อมูลที่เก็บรวบรวมมาได้นั้นเป็นข้อมูลที่ไม่สมบูรณ์ (Incomplete data) โดยจะพบว่าข้อมูลที่ได้เป็นข้อมูลที่ถูกตัดทิ้งทางขวา (Right-Censored data) และในการวิเคราะห์ข้อมูลเวลาการอยู่รอดนั้น สิ่งหนึ่งที่สนใจศึกษาคือฟังก์ชันการอยู่รอด (Survival function) ซึ่งสามารถประมาณได้ด้วยวิธีการทางสถิติ

สำหรับการศึกษาฟังก์ชันการอยู่รอดในงานด้านการประกันภัยนั้น อาจสนใจศึกษาการคงอยู่ของกรมธรรม์ซึ่งจะมีประโยชน์ในด้านการพิจารณาความเป็นไปได้ของกรมธรรม์ที่ออกแบบขึ้นใหม่ โดยพิจารณาการกำหนดระยะเวลาเอาประกันที่เหมาะสม หรือนำไปใช้ประกอบการคำนวณและการวิเคราะห์เงินสำรอง โดยที่การคำนวณเงินสำรองให้ได้เพียงพอและเหมาะสมตามความเป็นจริงนั้น จะทำให้บริษัทประกันสามารถวิเคราะห์ค่าใช้จ่ายในการดำเนินงานและวิเคราะห์การจ่ายเงินค่าสินไหมทดแทนได้อย่างเพียงพอ ซึ่งการคำนวณหาเงินสำรองให้เหมาะสมใกล้เคียงกับความเป็นจริงนั้นต้องอาศัยข้อมูลที่มีลักษณะสมบูรณ์ (Complete data) แต่ในทางปฏิบัติข้อมูลที่เก็บรวบรวมมาได้มักจะเป็นข้อมูลที่ไม่สมบูรณ์ เช่น เมื่อสนใจศึกษาเวลาการคงอยู่ของกรมธรรม์ในช่วงเวลา 5 ปี การเก็บรวบรวมข้อมูลจะทำโดยการบันทึกข้อมูลระยะเวลาตั้งแต่ผู้เอาประกันเริ่มทำประกันจนกระทั่งขาดส่งเบี้ยประกัน หรือมีการยกเลิกกรมธรรม์ก่อนหมดระยะเวลาเอาประกัน ซึ่งเป็นผลทำให้เกิดการขาดอายุกรมธรรม์ (Lapsation) และจะบันทึกข้อมูลระยะเวลาเป็น 5 ปี

สำหรับกรมธรรม์ที่ยังคงส่งเบี้ยประกันจนถึงเวลาที่ทำการศึกษา นั่นคือ ข้อมูลที่เก็บรวบรวมได้จะมีลักษณะเป็นข้อมูลที่มีค่าถูกตัดทิ้งประเภทที่ 1 (Type I Censoring) ดังนั้นการวิเคราะห์ความน่าจะเป็นที่กรมธรรม์จะคงอยู่มากกว่าเวลา t ก็คือ การประมาณค่าฟังก์ชันการอยู่รอด ($S(t)=\Pr(T>t)$) เมื่อ $T$ เป็นตัวแปรสุ่มของเวลาจากเริ่มทำประกันจนกระทั่งเกิดการขาดอายุกรมธรรม์ ส่วนในงานด้านการแพทย์ต้องการทราบฟังก์ชันการอยู่รอดของกลุ่มผู้ป่วยที่ได้รับการรักษาด้วยยาชนิดใหม่ หรือในงานด้านอุตสาหกรรมการผลิตเครื่องใช้ไฟฟ้า เพื่อการควบคุมคุณภาพสินค้า ผู้ผลิตต้องการทดสอบระยะเวลาที่เครื่องใช้ไฟฟ้าที่ผลิตขึ้นมาจะสามารถทำงานได้เป็นปกติไม่บกพร่องในช่วงเวลาที่รับประกันคุณภาพสินค้า ดังนั้นการประมาณค่าฟังก์ชันการอยู่รอดให้ได้ใกล้เคียงกับความเป็นจริงมากที่สุดจะเป็นประโยชน์มากในงานวิจัยด้านต่างๆ ข้างต้น

ในทางปฏิบัติ ถ้าทราบการแจกแจงของระยะเวลาตั้งแต่เริ่มต้นศึกษาจนกระทั่งเกิดเหตุการณ์ที่ต้องการศึกษา ก็จะสามารถประมาณค่าฟังก์ชันการอยู่รอดโดยวิธีการประมาณที่ใช้พารามิเตอร์ (Parametric Estimation) ได้ แต่ในการวิจัยสำหรับกรมธรรม์หรือสินค้าใหม่นั้นมักไม่ทราบการแจกแจงของข้อมูลระยะเวลาที่จะเกิดเหตุการณ์ที่สนใจศึกษา ดังนั้นอาจประมาณค่าฟังก์ชันการอยู่รอดโดยวิธีการประมาณที่ไม่ใช้พารามิเตอร์ (Nonparametric Estimation)

จากปัญหาดังกล่าว ผู้วิจัยสนใจที่จะทำการศึกษาการประมาณค่าฟังก์ชันการอยู่รอดโดยวิธีการประมาณที่ไม่ใช้พารามิเตอร์ สำหรับข้อมูลที่ไม่สมบูรณ์ซึ่งเป็นข้อมูลที่ถูกตัดทิ้งทางขวาโดยมีการกำหนดเวลาสิ้นสุดการเก็บข้อมูลไว้ล่วงหน้า คือเป็นข้อมูลที่มีค่าถูกตัดทิ้งประเภทที่ 1 และทำการศึกษาเปรียบเทียบวิธีการประมาณค่าฟังก์ชันการอยู่รอด 3 วิธี ดังต่อไปนี้

1. วิธีลิมิตผลคูณ หรือวิธีพีแอล (Product Limit (PL) Method)

2. วิธีฟังก์ชันภาวะภัย (Hazard Function Method)

3. วิธีเบส์ (Bayes Method) สำหรับวิธีนี้ผู้วิจัยพิจารณาการแจกแจงก่อน (Prior Distribution) จากเอกสาร The Annals of Statistics 7 (1979) ซึ่งได้เสนอการแจกแจงก่อนไว้ 3 แบบ คือ กระบวนการโฮโมจีเนียสอย่างง่าย (A simple homogeneous process) กระบวนการแกมมา (Gamma process) และกระบวนการดีริชเลต์ (Dirichlet process) ทั้งนี้ เนื่องจากกระบวนการโฮโมจีเนียสอย่างง่าย และกระบวนการแกมมาให้ผลไม่แตกต่างกัน ดังนั้นในที่นี้ผู้วิจัยได้พิจารณาเปรียบเทียบการแจกแจงก่อน เป็น 2 แบบเท่านั้น คือ

3.1) การแจกแจงก่อนเป็นกระบวนการแกมมา (Gamma Process)

3.2) การแจกแจงก่อนเป็นกระบวนการดีริชเลต์ (Dirichlet Process)

## วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาและเปรียบเทียบวิธีการประมาณค่าฟังก์ชันการอยู่รอด สำหรับข้อมูลที่มีค่าถูกตัดทิ้งโดยการกำหนดเวลาสิ้นสุดการเก็บรวบรวมข้อมูลไว้ล่วงหน้า ด้วยวิธีการประมาณที่ไม่ใช้พารามิเตอร์ 3 วิธี คือ วิธีพีแอล วิธีฟังก์ชันภาวะภัย และวิธีเบส์

2. เพื่อศึกษาผลการทดสอบการเปรียบเทียบวิธีการประมาณค่าฟังก์ชันการอยู่รอด เมื่อจำนวนค่าสังเกตไม่สมบูรณ์ในตัวอย่างเพิ่มขึ้น

3. เพื่อศึกษาผลการทดสอบการเปรียบเทียบวิธีการประมาณค่าฟังก์ชันการอยู่รอด เมื่อกำหนดเวลาสิ้นสุดการเก็บรวบรวมข้อมูลไว้ล่วงหน้าแตกต่างกัน

## สมมติฐานการวิจัย

1. ในกรณีที่เวลาที่กำหนดไว้ล่วงหน้ามีค่าน้อยกว่าค่าเฉลี่ยของการแจกแจงของระยะเวลาอยู่รอด วิธีฟังก์ชันภาวะภัยจะเป็นวิธีการประมาณที่ให้ค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Percentage Error:MAPE) ต่ำที่สุดใน 3 วิธี

2. ในกรณีที่เวลาที่กำหนดไว้ล่วงหน้ามีค่าเท่ากับ และมากกว่าค่าเฉลี่ยของการแจกแจงของระยะเวลาอยู่รอด วิธีเบส์จะเป็นวิธีการประมาณที่ให้ค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Percentage Error:MAPE) ต่ำที่สุดใน 3 วิธี

## ข้อตกลงเบื้องต้น

1. สำหรับการวิจัยนี้ศึกษาข้อมูลในการทดลองแบบปิด คือไม่มีหน่วยตัวอย่างเข้าใหม่และไม่มีหน่วยตัวอย่างภายใต้การทดลองออกจากกลุ่ม

2. ศึกษาในกรณีที่การแจกแจงเป็นแบบเบ้ขวา เนื่องจากการแจกแจงการอยู่รอดมักพบเป็นแบบเบ้ขวา

3. ในการวิจัยครั้งนี้กำหนดให้เริ่มต้นเก็บข้อมูลที่เวลาเดียวกัน คือ $t = 0$ และศึกษาลักษณะของข้อมูลที่ไม่สมบูรณ์ที่มีค่าถูกตัดทิ้งประเภทที่ 1 (Type I Censoring) คือมีการกำหนดเวลาการสิ้นสุดการเก็บข้อมูลไว้ล่วงหน้า (Predetemined fixed time period : Tc)

4. กำหนดเวลาที่สิ้นสุดการเก็บข้อมูลที่ถูกกำหนดไว้ล่วงหน้า (Tc) ให้มีค่าแตกต่างกัน ดังนี้คือ มีค่าน้อยกว่า เท่ากับและมากกว่าค่าเฉลี่ยของการแจกแจงของระยะเวลาอยู่รอด โดยให้น้อยกว่า และมากกว่าค่าเฉลี่ยของการแจกแจงของระยะเวลาอยู่รอดเป็นระยะห่าง 25%, 50% และ 75 %

5. ศึกษากรณีการประมาณค่าฟังก์ชันการอยู่รอด $S(t)$ ณ เวลาการอยู่รอด t โดยมีค่า t ตั้งแต่ 0.25 ถึงเวลาที่กำหนดไว้ล่วงหน้า (เพิ่มขึ้นทีละ 0.25)

<u>ขอบเขตของการวิจัย</u>

1. ในการวิจัยครั้งนี้ทำการศึกษาเปรียบเทียบการประมาณค่าฟังก์ชันการอยู่รอด ด้วยวิธีพีแอล วิธีฟังก์ชันภาวะภัย และวิธีเบส์

2. ศึกษากรณีที่ค่าเฉลี่ยของการแจกแจงระยะเวลาการอยู่รอด (E(T)) เป็น 4.0

3. ศึกษากรณีที่ข้อมูล $T_1, T_2, T_3, \ldots, T_n$ เป็นตัวอย่างสุ่มของเวลาการอยู่รอดที่เป็นอิสระกัน และมีการแจกแจงเหมือนกันเป็น

3.1) การแจกแจงแบบไวบูลล์ (Weibull Distribution) ซึ่งมีฟังก์ชันความหนาแน่นอยู่ในรูป

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

ในการวิจัยครั้งนี้จะศึกษาที่ E(T) = 4.0 และ $\gamma$ = 1.5 [1]

3.2) การแจกแจงแบบลอกนอร์มอล (Lognormal Distribution) ซึ่งมีฟังก์ชันความหนาแน่นอยู่ในรูป

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

ในการวิจัยครั้งนี้ศึกษาที่ E(T) = 4.0 และ $\sigma$ = 0.9 [1]

---

[1] ผู้วิจัยได้ทดลองด้วยค่าเฉลี่ยของการแจกแจงหลายค่า และค่าพารามิเตอร์ต่างๆ ที่ทำให้การแจกแจงมีลักษณะเบ้ขวาแล้ว พบว่าให้ผลสรุปไม่แตกต่างกัน

4. ศึกษาในกรณีที่เปอร์เซ็นต์ของข้อมูลที่ถูกตัดทิ้งเป็น 10%, 20%, 30% และ 40% ของขนาดตัวอย่าง

5. ศึกษาในกรณีที่ขนาดตัวอย่างมี 5 ระดับ คือ 10, 20, 30, 40 และ 50

6. ศึกษากรณีที่เวลาที่สิ้นสุดการเก็บข้อมูลที่ถูกกำหนดไว้ล่วงหน้า Tc มีค่าน้อยกว่า เท่ากับ และมากกว่าค่าเฉลี่ยของการแจกแจงของระยะเวลาอยู่รอด โดยให้น้อยกว่าและมากกว่าค่าเฉลี่ยของการแจกแจงของระยะเวลาอยู่รอดเป็นระยะห่าง 25%, 50% และ 75% ดังนั้นเวลาที่ถูกกำหนดไว้ล่วงหน้า คือ 1.0, 2.0, 3.0, 4.0, 5.0, 6.0 และ 7.0 ตามลำดับ

7. ศึกษาในกรณีที่เวลาการอยู่รอดมีค่าตั้งแต่ 0.25 ถึงเวลา Tc โดยกำหนดให้มีค่าเพิ่มขึ้นทีละ 0.25

8 ในการวิจัยครั้งนี้สร้างแบบจำลองข้อมูลโดยใช้เทคนิคมอนติคาร์โล (Monte Carlo Simulation Method) ทำการทดลองซ้ำ 1,000 ครั้ง ในแต่ละสถานการณ์

<u>เกณฑ์การตัดสินใจ</u>

ในการพิจารณาว่าวิธีการประมาณค่าฟังก์ชันการอยู่รอดวิธีใดจะให้ค่าประมาณมีความคลาดเคลื่อนโดยเฉลี่ยต่ำที่สุด จะพิจารณาเปรียบเทียบค่าความคลาดเคลื่อนระหว่างค่าประมาณจากวิธีต่างๆ ($\hat{S}(t)$) กับค่าฟังก์ชันการอยู่รอดของการแจกแจง ($S(t)$) ตามสมมติฐาน ในรูปของค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Percentage Error : MAPE) ซึ่งวิธีการประมาณวิธีใดให้ค่า MAPE ต่ำกว่าจะเป็นวิธีการประมาณที่ดีกว่า

<u>ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย</u>

1. เพื่อเป็นแนวทางให้นักวิจัยมีผลสรุปในการนำวิธีการประมาณที่ไม่ใช้พารามิเตอร์ที่เหมาะสมในการประมาณค่าฟังก์ชันการอยู่รอดไปใช้กับข้อมูลจริงที่มีค่าสังเกตถูกตัดทิ้งซึ่งมีการกำหนดเวลาการเก็บข้อมูลไว้ล่วงหน้า

2. เพื่อเป็นแนวทางในการศึกษา และเปรียบเทียบวิธีการประมาณที่ไม่ใช้พารามิเตอร์ในการประมาณค่าฟังก์ชันการอยู่รอด เมื่อข้อมูลมีค่าถูกตัดทิ้งแบบอื่นๆ ต่อไป

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

<u>คำจำกัดความ</u>

**ข้อมูลที่มีค่าถูกตัดทิ้งประเภทที่ 1** (Type I Censoring) **เป็นข้อมูลชนิดหนึ่งซึ่งเมื่อข้อมูลที่มีค่ามากกว่าค่าที่กำหนดไว้ล่วงหน้าแล้วจะถูกบันทึกค่าให้มีค่าเท่ากับค่าที่กำหนดไว้ล่วงหน้าเท่านั้น นั่นคือ**

ถ้าให้ $T_1, T_2, T_3, \ldots, T_n$ เป็นตัวแปรสุ่มเวลาการอยู่รอดที่มีการแจกแจงเหมือนกัน และเป็นอิสระกัน และ Tc เป็นเวลาที่ถูกกำหนดไว้ล่วงหน้าแล้ว จะได้ค่าสังเกตสุ่ม $Y_1, Y_2, Y_3, \ldots, Y_n$ โดยที่

$Y_i = T_i$ ; ถ้า $T_i \leq Tc$ นั่นคือ ข้อมูลไม่ถูกตัดทิ้ง

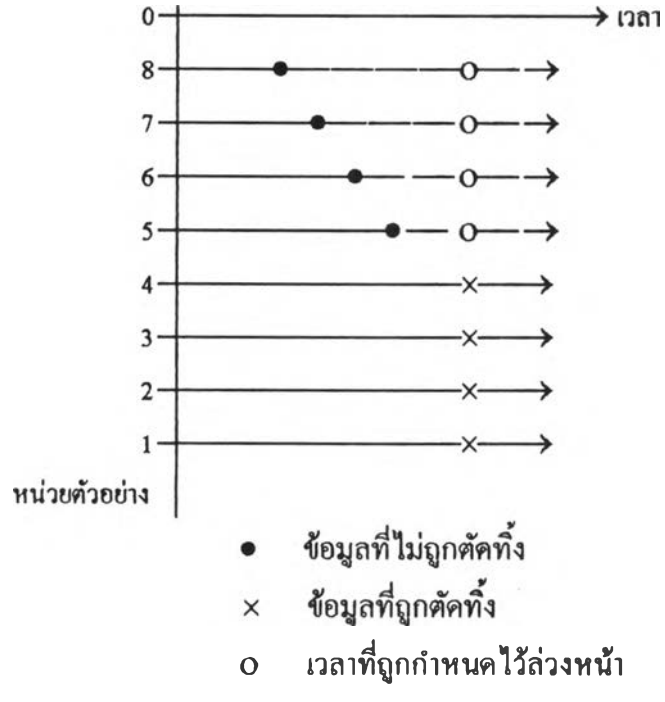
$= Tc$ ; ถ้า $T_i > Tc$ นั่นคือ ข้อมูลถูกตัดทิ้ง

แสดงลักษณะของค่าสังเกตสุ่มได้ดังรูปที่ 1.1

ถ้าให้ $\delta_i$ เป็นครรชนีของ $Y_i$ ที่แสดงว่า $Y_i$ เป็นข้อมูลที่มีค่าถูกตัดทิ้งหรือไม่ โดยกำหนดค่าของ $\delta_i$ ดังนี้

$\delta_i = 1$ ; ถ้า $T_i \leq Tc$ นั่นคือ ข้อมูลไม่ถูกตัดทิ้ง

$= 0$ ; ถ้า $T_i > Tc$ นั่นคือ ข้อมูลถูกตัดทิ้ง




**รูปที่ 1.1**