บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยนี้ทำการศึกษา และเปรียบเทียบวิธีการประมาณค่าฟังก์ชันการอยู่รอด ($S(t)$) สำหรับข้อมูลไม่สมบูรณ์ที่มีค่าถูกตัดทิ้งประเภทที่ 1 ซึ่งได้เสนอไว้ 3 วิธี คือ วิธีพีแอล วิธีฟังก์ชันภาวะภัย และวิธีเบส์ สำหรับวิธีเบส์นั้นได้ทำการกำหนดการแจกแจงก่อนเป็น 2 แบบ คือ แบบกระบวนการแกมมา และแบบกระบวนการดีริชเลต์ ซึ่งจะพิจารณาค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ และค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ที่ได้จากแต่ละวิธีนำมาเปรียบเทียบกัน โดยทำการศึกษาสำหรับข้อมูลที่มีการแจกแจง 2 แบบ คือแบบไวบูลล์ และแบบลอกนอร์มอลที่มีขนาดตัวอย่างเป็น 10, 20, 30, 40 และ 50 เปอร์เซ็นต์ของข้อมูลที่มีค่าถูกตัดทิ้งเป็น 10%, 20%, 30% และ 40% เวลาสิ้นสุดการเก็บรวบรวมข้อมูลที่ถูกกำหนดไว้ล่วงหน้า (Tc) มีค่าเท่ากับ น้อยกว่าและมากกว่าค่าเฉลี่ยของการแจกแจงของของระยะเวลาอยู่รอด โดยให้น้อยกว่า และมากกว่าค่าเฉลี่ยเป็นระยะห่าง 25%, 50% และ 75 % ตามลำดับ

สำหรับการวิจัยครั้งนี้ใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Technique) สร้างสถานการณ์ต่างๆ ในการทดลองโดยใช้โปรแกรมภาษาฟอร์แทรน 77 (Fortran 77) ทำการทดลองซ้ำ 1000 ครั้ง ซึ่งเทคนิคมอนติคาร์โล เป็นเทคนิคในการสร้างข้อมูลโดยการใช้ตัวเลขสุ่ม (Random Number) จากตารางเลขสุ่ม (Random Numbers Table) หรือจากโปรแกรมคอมพิวเตอร์ และนำความน่าจะเป็นสะสมมาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา โดยสามารถสร้างค่าตัวแปรแบบสุ่มที่มีการกระจายของความน่าจะเป็นที่มีลักษณะตามแบบที่ต้องการได้จากตัวเลขสุ่มที่มีลักษณะการกระจายแบบสม่ำเสมอ (Uniform Distribution) ในช่วง (0,1) แล้วจึงนำค่าตัวแปรแบบสุ่มที่ถูกสร้างขึ้นมานั้นไปประยุกต์ใช้กับปัญหาที่ต้องการศึกษา โดยทดลองกระทำในลักษณะซ้ำๆ กันหลายๆ ครั้งเพื่อหาคำตอบที่ต้องการ

<u>แผนการทดลอง</u>

การวิจัยครั้งนี้ ต้องการศึกษาและเปรียบเทียบวิธีการประมาณค่าฟังก์ชันการอยู่รอดเมื่อข้อมูลมีค่าถูกตัดทิ้งประเภทที่ 1 จากวิธีการประมาณ 3 วิธี โดยผู้วิจัยจะเปรียบเทียบค่าประมาณที่

ได้ โดยพิจารณาค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Percentage Error, MAPE) เพื่อหาวิธีที่ดีที่สุดในแต่ละสถานการณ์ต่อไป รายละเอียดของแผนการทดลองมีดังต่อไปนี้

1. กำหนดให้ข้อมูลระยะเวลามีการแจกแจง 2 แบบ คือ

1.1) การแจกแจงแบบไวบูลล์ ศึกษาที่ค่าเฉลี่ยของการแจกแจง ($E(T)$) เท่ากับ 4.0 และที่พารามิเตอร์ $\gamma = 1.5$

1.2) การแจกแจงแบบลอกนอร์มอล ศึกษาที่ค่าเฉลี่ยของการแจกแจง ($E(T)$) เท่ากับ 4.0 และที่พารามิเตอร์ $\sigma = 0.9$

2. เวลาที่ถูกกำหนดไว้ล่วงหน้า (Tc) มีค่าเท่ากับ น้อยกว่า และมากกว่าค่าเฉลี่ยของการแจกแจงโดยศึกษาระยะห่างจากค่าเฉลี่ยของการแจกแจงเป็น 3 ระดับ คือ 25%, 50% และ 75% นั่นคือ ศึกษาเวลาที่ถูกกำหนดไว้ล่วงหน้าเป็น 1.0, 2.0, 3.0, 4.0, 5.0, 6.0 และ 7.0

3. กำหนดขนาดตัวอย่างที่นำมาศึกษาเป็น 5 ระดับ คือ 10, 20, 30, 40 และ 50

4. กำหนดเปอร์เซ็นต์ของข้อมูลที่ถูกตัดทิ้ง 4 ระดับ คือ 10%, 20%, 30% และ 40%

5.กำหนดให้มีการคำนวณค่าฟังก์ชันการอยู่รอด ณ เวลาที่มีค่าตั้งแต่ 0.25 ถึงเวลาที่ถูกกำหนดไว้ล่วงหน้า (ตามข้อ 2) โดยเพิ่มขึ้นทีละ 0.25

6. สำหรับวิธีเบส์ ผู้วิจัยได้ทำการกำหนดการแจกแจงก่อนเป็น 2 แบบ คือ

6.1) การแจกแจงก่อนเป็นแบบกระบวนการแกมมา

6.2) การแจกแจงก่อนเป็นแบบกระบวนการดีริชเลต์

<u>ขั้นตอนการวิจัย</u>

แบ่งออกเป็น 5 ขั้นตอน ดังนี้

1.จำลองข้อมูลระยะเวลาจากการแจกแจงแบบไวบูลล์ และการแจกแจงแบบลอกนอร์มอล

2. กำหนดข้อมูลให้มีลักษณะเป็นข้อมูลที่ถูกตัดทิ้งประเภทที่ 1

3. เรียงลำดับข้อมูลจากค่าน้อยไปมาก สำหรับข้อมูลที่ไม่มีค่าสังเกตซ้ำหาลำดับที่ของข้อมูล ส่วนกรณีที่มีค่าสังเกตซ้ำจะนับจำนวนตัวอย่างที่มีค่าอยู่ในช่วงที่ศึกษา

4. คำนวณหาค่าฟังก์ชันการอยู่รอด ด้วยวิธีการประมาณทั้ง 3 วิธี

5. หาค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ และค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์จากการประมาณค่าฟังก์ชันการอยู่รอด และนำมาเปรียบเทียบเพื่อที่จะหาวิธีการประมาณที่ดีที่สุด

แต่ละขั้นตอนข้างต้นมีรายละเอียดดังนี้

1. การจำลองข้อมูลจากการแจกแจง 2 แบบ คือ การแจกแจงแบบไวบูลล์ และการแจกแจงแบบลอกนอร์มอล โดยที่การสร้างค่าตัวแปรแบบสุ่มให้มีลักษณะการแจกแจงตามที่ต้องการได้นั้น จะ ซึ่งมีวิธีการดังนี้

1.1) เมื่อ T มีการแจกแจงแบบไวบูลล์

การสร้างตัวเลขสุ่มที่มีการแจกแจงแบบไวบูลล์จะใช้เทคนิคการแปลงผกผัน (Inverse Transformation) ซึ่งมีวิธีการดังนี้

ให้ R ซึ่งเป็นตัวเลขสุ่ม (Random Number) ในช่วง (0,1) มีค่าเท่ากับ ฟังก์ชันการแจกแจงสะสมของการแจกแจงที่ต้องการจำลองข้อมูล

$$R = F(t)$$

และจะได้ค่าตัวแปรสุ่ม $$t = F^{-1}(R)$$

$$t = \frac{\left(-\ln(R)\right)^{\frac{1}{\gamma}}}{\lambda}$$

โดยที่ t เป็นค่าตัวแปรแบบสุ่มซึ่งถูกสร้างขึ้น ซึ่งในที่นี้คือเวลาที่ยังคงอยู่

R เป็นตัวเลขสุ่มซึ่งมีค่าอยู่ระหว่าง 0 และ 1

γ, λ เป็นพารามิเตอร์ของการแจกแจงแบบไวบูลล์

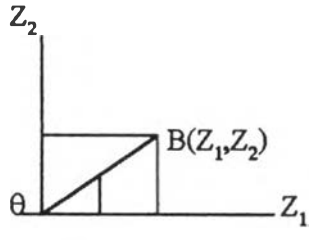
1.2 เมื่อ T มีการแจกแจงแบบลอกนอร์มอล

สำหรับการสร้างค่าตัวแปรแบบสุ่มที่มีการแจกแจงแบบลอกนอร์มอลนั้น ทำได้โดยการหาค่าของ Exponential ของตัวแปรแบบสุ่มที่มีการแจกแจงแบบปกติ โดยที่การสร้างตัวแปรแบบสุ่มที่มีการแจกแจงแบบปกติมีวิธีการดังนี้

การสร้างตัวแปรแบบสุ่มที่มีการแจกแจงแบบปกติจะใช้เทคนิคการแปลงโดยตรงจาก

$$\phi(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \quad ,-\infty < x < \infty$$

โดย Box และ Muller(ค.ศ.1958) สร้างเลขสุ่มที่มีการแจกแจงแบบปกติมาตรฐานที่มีค่าเฉลี่ย 0 และค่าความแปรปรวน 1 พร้อมๆ กัน 2 ค่าดังนี้



เมื่อ $Z_1 = B\cos\theta$

$Z_2 = B\sin\theta$

ดังนั้น $B = Z_1^2 + Z_2^2$ จะมีการแจกแจงแบบไคสแควร์(Chi-square Distribution) ระดับองศาความเป็นอิสระ 2 ซึ่งก็คือการแจกแจงแบบเอกซ์โพเนนเชียล(Exponential Distribution) ที่มีค่าเฉลี่ยเท่ากับ 2 แล้วจะได้ค่ารัศมี B ดังนี้

$$B = (-2\ln R)^{1/2}$$

เมื่อ R เป็นเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอในช่วง(0,1)

พิจารณาการสมมาตรของการแจกแจงแบบปกติ จะได้ θ มีการแจกแจงแบบสม่ำเสมอระหว่าง 0 กับ 2π เรเดียน ซึ่งค่า B และ θ เป็นอิสระกัน(Mutually Independent)

$$Z_1 = (-2\ln R_1)^{1/2}\cos(2\pi R_2)$$

$$Z_2 = (-2\ln R_1)^{1/2}\sin(2\pi R_2)$$

เมื่อได้เลขสุ่มที่มีการแจกแจงปกติมาตรฐาน(Standard-Normal Distribution) แล้วทำการแปลงค่าเลขสุ่มดังกล่าวให้เป็นเลขสุ่มที่มีการแจกแจงปกติมีค่าเฉลี่ย μ และค่าเบี่ยงเบนมาตรฐาน σ ดังนี้

$$X_1 = Z_1\sigma + \mu$$

$$X_2 = Z_2\sigma + \mu$$

โปรแกรมย่อยที่ใช้ในการสร้างเลขสุ่มที่มีการแจกแจงปกติ มีค่าเฉลี่ย μ และค่าเบี่ยงเบนมาตรฐาน σ คือ ฟังก์ชัน Normal(Dmean,Sigma) เมื่อ Dmean และ Sigma เป็นค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของการแจกแจงแบบปกติ ดังแสดงในภาคผนวก

2. กำหนดข้อมูลระยะเวลาที่ยังคงอยู่ ซึ่งจำลองได้จากหัวข้อที่ 1 ให้มีลักษณะของข้อมูลเป็นแบบถูกตัดทิ้งทางขวาประเภทที่ 1 โดยกำหนดให้ $T_i = \mathrm{Tc}$ เมื่อ $T_i > \mathrm{Tc}$ นั่นคือ การวิจัยครั้งนี้จะกำหนดเวลาสิ้นสุดการเก็บข้อมูลไว้ล่วงหน้า ดังนั้น ถ้าข้อมูลที่ได้จากการจำลองข้อมูลมีค่าเกินกว่าเวลาที่ถูกกำหนดไว้ล่วงหน้าให้ถือว่าข้อมูลนั้นมีค่าถูกตัดทิ้ง

3. นำข้อมูลที่ได้จากการดำเนินการตามข้อที่ 1 และ 2 มาเรียงลำดับค่าจากน้อยไปมาก นั่นคือ จะนำข้อมูลทั้งที่ถูกตัดทิ้งและไม่ถูกตัดทิ้งมาเรียงลำดับเพื่อใช้ในการคำนวณขั้นตอนต่อไป

4. การประมาณค่าฟังก์ชันการอยู่รอด ด้วยวิธีการประมาณที่ไม่ใช้พารามิเตอร์ทั้ง 3 วิธี โดยแต่ละวิธีมีขั้นตอนดังนี้

4.1) วิธีพีแอล

1) นำข้อมูลมาเรียงลำดับค่าจากน้อยไปมาก เพื่อหาลำดับที่ของข้อมูลสำหรับข้อมูลที่ไม่มีค่าสังเกตซ้ำ หรือนับจำนวนตัวอย่างที่มีค่าอยู่ในแต่ละช่วงที่ต้องการประมาณค่าเมื่อข้อมูลมีค่าสังเกตซ้ำ

2) หาค่าประมาณฟังก์ชันการอยู่รอด $\hat{S}_{PL}(t)$ จากสมการ (2.1) สำหรับกรณีที่ข้อมูลมีค่าสังเกตซ้ำ และเมื่อข้อมูลไม่มีค่าสังเกตซ้ำจะประมาณค่าฟังก์ชันการอยู่รอดจากสมการ (2.2)

4.2) วิธีฟังก์ชันภาวะภัย

1) นำข้อมูลมาเรียงลำดับค่าจากน้อยไปมาก เพื่อหาลำดับที่ของข้อมูล เพื่อหาลำดับที่ของข้อมูลสำหรับข้อมูลที่ไม่มีค่าสังเกตซ้ำ หรือนับจำนวนตัวอย่างที่มีค่าอยู่ในแต่ละช่วงที่ต้องการประมาณค่าเมื่อข้อมูลมีค่าสังเกตซ้ำ

2) หาค่าประมาณฟังก์ชันการอยู่รอด $\hat{S}_{HA}(t)$ จากสมการ (2.8) สำหรับกรณีที่ข้อมูลมีค่าสังเกตซ้ำ และเมื่อข้อมูลไม่มีค่าสังเกตซ้ำจะประมาณค่าฟังก์ชันการอยู่รอดจากสมการ (2.10)

4.3) วิธีเบส์

1) เมื่อกำหนดการแจกแจงก่อนเป็นกระบวนการแกมมา

โดยที่มีพารามิเตอร์คือ $\gamma(t)$ และ τ ซึ่งในการคำนวณตัวประมาณเบส์ที่ไม่ใช้พารามิเตอร์จะต้องมีการกำหนดการแจกแจงคาดเดาก่อนการทดลอง (prior guess) ของ $S(t)$ เพื่อใช้ในการประมาณค่าพารามิเตอร์ สำหรับงานวิจัยนี้กำหนดการแจงแจงคาดเดาก่อนการทดลอง ($S_0(t)$) เป็นการแจกแจงแบบเอกซโพเนนเชียล (Exponential Distribution) ดังนี้

$$S_0(t) = \exp(-t\theta)$$

เมื่อ θ เป็นพารามิเตอร์ของการแจกแจงแบบเอกซโพเนนเชียล ซึ่งในที่นี้ประมาณ θ ด้วยวิธีการประมาณภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation) โดยสามารถหาค่าได้จาก

$$\hat{\theta} = \frac{r}{\sum_{i=1}^{n} t_i}$$

เมื่อ r เป็นจำนวนค่าสังเกตที่ไม่ถูกตัดทิ้ง และเพื่อจะหาประมาณ $\gamma(t)$ จะต้องกำหนด τ โดยในที่นี้กำหนดให้ $\tau = 1$ ดังนั้นจากสมการที่ (2.31) จะได้ว่า

$$\gamma(t) = \log(S_0(t))/\log(0.5)$$

นั่นคือ สามารถประมาณค่าฟังก์ชันการอยู่รอดด้วยวิธีเบส์ เมื่อกำหนดการแจกแจงก่อนเป็นกระบวนการแกมมา $\hat{S}_G(t)$ ได้จากสมการที่ (2.30)

2) เมื่อกำหนดการแจกแจงก่อนเป็นกระบวนการดีริชเลต์

โดยการกำหนดการแจกแจงคาดเดาก่อนการทดลอง (prior guess) ของ $S(t)$ เป็นการแจกแจงแบบเอกซโพเนนเชียล (Exponential Distribution) และกำหนด $\alpha(R) = 1$ ดังนั้นจากสมการที่ (2.38) จะได้

$$\alpha(t) = S_0(t) = \exp(-t\hat{\theta})$$

นั่นคือ สามารถประมาณค่าฟังก์ชันการอยู่รอดด้วยวิธีเบส์ เมื่อกำหนดการแจกแจงก่อนเป็นกระบวนการดีริชเลต์ $\hat{S}_D(t)$ ได้จากสมการที่ (2.37)

5. ในการทดลองได้ทำการจำลองข้อมูลซ้ำกันจำนวน 1000 ครั้ง ในแต่ละสถานการณ์ที่กำหนด เมื่อให้ i แทนรอบที่ทำซ้ำ โดยที่ $i = 1,2,3,\ldots,1000$ ดังนั้น

$$\hat{S}(t) = \frac{1}{1000}\sum_{i=1}^{1000}\hat{S}(t)^{(i)}$$

โดยที่ $\hat{S}(t)^{(i)}$ เป็นค่าประมาณ $S(t)$ สำหรับการทำซ้ำรอบที่ i
และคำนวณค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (APE) และค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของแต่ละวิธีตามสมการต่อไปนี้

$$\text{APE}_t = \left|\frac{S(t)-\hat{S}(t)}{S(t)}\right| \times 100\ \%$$

เมื่อ $S(t)$ เป็นค่าฟังก์ชันการอยู่รอด และ $\hat{S}(t)$ เป็นค่าประมาณของฟังก์ชันการอยู่รอดด้วยวิธีการประมาณต่างๆ โดยที่เมื่อจะประมาณค่าฟังก์ชันการอยู่รอดด้วยวิธีพีแอลสามารถประมาณ $S(t)$ ด้วย $\hat{S}_{PL}(t)$ จากสมการที่ (2.1) วิธีฟังก์ชันภาวะภัยสามารถประมาณด้วย $\hat{S}_{HA}(t)$ จากสมการที่ (2.10) วิธีเบส์เมื่อการแจกแจงก่อนเป็นกระบวนการแกมมาและกระบวนการดีริชเลต์ประมาณด้วย $\hat{S}_G(t), \hat{S}_D(t)$ จากสมการที่ (2.30) และ (2.37) ตามลำดับ

$$\text{MAPE} = \frac{\sum_{t=0.25}^{Tc}\text{APE}_t}{\text{NT}}$$

เมื่อ NT เป็นจำนวนจุดเวลาตั้งแต่ 0.25 ถึงเวลาที่ถูกกำหนดไว้ล่วงหน้า โดยเพิ่มขึ้นทีละ 0.25 แล้วนำค่า MAPE ของการประมาณค่าในแต่ละวิธีมาเปรียบเทียบกัน โดยวิธีที่ให้ค่า MAPE ต่ำที่สุดจะเป็นวิธีการประมาณที่ดีที่สุดของแต่ละสถานการณ์

โปรแกรมที่ใช้ในการวิจัย

โปรแกรมที่ใช้ในการวิจัยนี้ เขียนด้วยภาษาฟอร์แทรน (FORTRAN) โดยใช้กับเครื่อง AMDAHL 5860 ซึ่งในแต่ละสถานการณ์ของการทดลองลักษณะการทำงานของโปรแกรมจะเหมือนกัน สำหรับรายละเอียดของผังงานแสดงการทำงานของโปรแกรม และโปรแกรมที่ใช้ในการประมาณค่า $S(t)$ ด้วยวิธีต่างๆ แสดงไว้ในภาคผนวก ข