บทที่ 2

สถิติที่ใช้ในการวิจัย

ในบทนี้ กล่าวถึงวิธีการประมาณความน่าจะเป็นที่จะเสียชีวิตสำหรับข้อมูลประกันชีวิตที่ไม่สมบูรณ์ ซึ่งมี 3 วิธี ดังนี้ วิธีการประมาณแบบคลาสสิค (Classical Estimation Method) วิธีการประมาณแบบภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation Method) และวิธีการประมาณแบบเบส์ (Bayes Estimation Method) โดยจะศึกษาลักษณะการแจกแจงของระยะเวลาที่จะมีชีวิตอยู่ต่อไปในอนาคต (Future Lifetime ,T) 2 แบบคือ การแจกแจงแบบไวบูลล์ (Weibull Distribution) และการแจกแจงแบบกอมเพิรตซ์ (Gompertz Distribution) โดยมีรายละเอียดดังนี้

วิธีการประมาณแบบคลาสสิค

วิธีการประมาณแบบคลาสสิค จะกำหนดสมมติฐานว่า ผลการทดลองแต่ละผลการทดลองมีโอกาสเกิดขึ้นเท่า ๆ กัน ภายใต้ผลการทดลองจำนวนจำกัด ดังนั้นจึงทำให้สามารถกำหนดนิยามของความน่าจะเป็นของเหตุการณ์ A ใดๆ ได้ดังนี้

$$P(A) = \frac{\text{จำนวนเหตุการณ์ A จะเกิดขึ้นได้}}{\text{จำนวนวิธีทั้งหมดซึ่งการทดลองจะให้ผลการทดลองได้}}$$

ซึ่งจากแนวความคิดดังกล่าว เราสามารถประยุกต์ใช้กับการประมาณหาค่า q_x จะได้ว่า

$$\hat{q}_x = \frac{D_x}{E_x} \qquad (2.1)$$

โดยที่ D_x = จำนวนคนที่เสียชีวิต (Observed Deaths) ระหว่างอายุ x และ x+1 ปี

E_x = จำนวนผู้เสี่ยงภัย (Exposure)

$$= \sum_{i=1}^{m} (t_i - s_i)$$

เมื่อ t_i เป็นระยะเวลาของคนที่ i ออกจากช่วงที่สนใจศึกษา ซึ่ง $0 < t_i \le 1$

s_i เป็นระยะเวลาของคนที่ i เข้ามาในช่วงที่สนใจศึกษา ซึ่ง $0 < s_i < 1$

m เป็นขนาดตัวอย่าง

<u>วิธีการประมาณแบบภาวะน่าจะเป็นสูงสุด</u>

วิธีการประมาณแบบภาวะน่าจะเป็นสูงสุดนี้ เป็นการประมาณซึ่งไม่ใช้ฟังก์ชันการเสี่ยงในการคัดเลือกตัวประมาณที่เหมาะสม แต่ใช้การวิเคราะห์จากสภาพความเป็นจริง ผู้ที่ค้นพบวิธีนี้เป็นคนแรกชื่อซี เอฟ เกาส์ (C.F. Gauss .,1821) ซึ่งเป็นนักคณิตศาสตร์ชาวเยอรมัน ต่อมานักสถิติชาวอังกฤษชื่อ อาร์ เอ ฟิชเชอร์ (R. A. Fisher, 1922) ได้ปรับปรุงวิธีการและตรวจสอบคุณสมบัติต่างๆ วิธีการนี้จะใช้ได้เมื่อตัวอย่างสุ่มมีการแจกแจงแบบมีพารามิเตอร์ (Parameter Distribution)[2]

กำหนด

$$\delta_i = \begin{cases} 1 & \text{เมื่อคนที่ } i \text{ เสียชีวิตในช่วง 1 ปี (เป็นค่าสังเกตที่ไม่ถูกตัดทิ้ง)} \\ 0 & \text{เมื่อคนที่ } i \text{ มีชีวิตอยู่ในช่วง 1 ปี (เป็นค่าสังเกตที่ถูกตัดทิ้ง)} \end{cases}$$

สำหรับกรณีข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 1 (Type 1 Censored Data)

$$L = \prod_{i=1}^{m} [f(t_i)]^{\delta_i} \left[{}_{t_i}p_x\right]^{1-\delta_i}$$

เมื่อ $f(t_i)$ คือ ฟังก์ชันความหนาแน่นของระยะเวลาสำหรับคนที่ i จะมีชีวิตอยู่ต่อไปในอนาคต

${}_{t_i}p_x$ คือ ความน่าจะเป็นที่คนอายุ x จะอยู่รอดต่อไปอีก t_i ปี

L คือ ฟังก์ชันภาวะน่าจะเป็น (Likelihood Function)

1 T มีการแจกแจงแบบไวบูลล์

$$L = \prod_{i=1}^{m} \left\{ kt_i^n \exp\left[-\frac{k}{n+1} t_i^{n+1}\right]\right\}^{\delta_i} \left\{\exp\left[-\frac{k}{n+1} t_i^{n+1}\right]\right\}^{1-\delta_i}$$

$$= \prod_{i=1}^{m} \left(kt_i^n\right)^{\delta_i} \exp\left[-\frac{k}{n+1} t_i^{n+1}\right]$$

---

[2]ธีระพร วีระถาวร, ดร. <u>การอนุมานเชิงสถิติขั้นกลาง : โครงสร้างและความหมาย</u> (พิฑักษ์การพิมพ์ : 2531) หน้า 99

$$\ln L = \ln k \sum_{i=1}^{m} \delta_i + n \sum_{i=1}^{m} \delta_i \ln t_i - \frac{k}{n+1} \sum_{i=1}^{m} t_i^{n+1}$$

$$g_1(k,n) = \frac{\partial \ln L}{\partial k} = \frac{1}{k}\sum_{i=1}^{m}\delta_i - \frac{1}{n+1}\sum_{i=1}^{m} t_i^{n+1} \qquad (2.2)$$

$$g_{11}(k,n) = \frac{\partial^2 \ln L}{\partial k^2} = -\frac{1}{k^2}\sum_{i=1}^{m}\delta_i \qquad (2.3)$$

$$g_{12}(k,n) = \frac{\partial^2 \ln L}{\partial n \partial k} = \frac{1}{(n+1)^2}\sum_{i=1}^{m} t_i^{n+1} - \frac{1}{n+1}\sum_{i=1}^{m}(\ln t_i) t_i^{n+1} \qquad (2.4)$$

$$g_2(k,n) = \frac{\partial \ln L}{\partial n} = \sum_{i=1}^{m}\delta_i \ln t_i + \frac{k}{(n+1)^2}\sum_{i=1}^{m} t_i^{n+1} - \frac{k}{n+1}\sum_{i=1}^{m}(\ln t_i) t_i^{n+1} \qquad (2.5)$$

$$g_{21}(k,n) = \frac{\partial^2 \ln L}{\partial k \partial n} = \frac{1}{(n+1)^2}\sum_{i=1}^{m} t_i^{n+1} - \frac{1}{n+1}\sum_{i=1}^{m}(\ln t_i) t_i^{n+1} \qquad (2.6)$$

$$g_{22}(k,n) = \frac{\partial^2 \ln L}{\partial n^2} = -\frac{2k}{(n+1)^3}\sum_{i=1}^{m} t_i^{n+1} + \frac{2k}{(n+1)^2}\sum_{i=1}^{m}(\ln t_i) t_i^{n+1} - \frac{k}{n+1}\sum_{i=1}^{m}(\ln t_i)^2 t_i^{n+1}$$
$$(2.7)$$

การหาค่าประมาณพารามิเตอร์ k และ n จะกระทำได้โดยใช้เทคนิคการวิเคราะห์เชิงตัวเลข (Numerical Analysis) ด้วยวิธีนิวตัน-ราพซ์สัน (Newton-Raphson Method) ซึ่งพิจารณาสองสมการนี้

$$g_1(k,n) = 0$$
$$g_2(k,n) = 0$$

โดยใช้วิธีการประมาณเชิงเส้น และกำหนดค่าเริ่มต้นเป็น k_0 และ n_0 โดยวิธีเปอร์เซ็นต์ไทล์ แมทซ์ชิ่ง (Percentile Matching) จะได้สมการดังข้างล่างนี้

$$g_1(k_0,n_0) + g_{11}(k_0,n_0)(k-k_0) + g_{12}(k_0,n_0)(n-n_0) = 0$$
$$g_2(k_0,n_0) + g_{21}(k_0,n_0)(k-k_0) + g_{22}(k_0,n_0)(n-n_0) = 0$$

หรือสามารถเขียนให้อยู่ในรูปเมตริกซ์ได้ คือ

$$\begin{bmatrix} g_{11}(k_0,n_0) & g_{12}(k_0,n_0) \\ g_{21}(k_0,n_0) & g_{22}(k_0,n_0) \end{bmatrix} \begin{bmatrix} k-k_0 \\ n-n_0 \end{bmatrix} = \begin{bmatrix} -g_1(k_0,n_0) \\ -g_2(k_0,n_0) \end{bmatrix}$$

และใช้กระบวนการการทำซ้ำ(Iterative Procedure)จนกระทั่งได้ค่า $\hat{k}$ และ $\hat{n}$ ที่ต้องการ หลังจากหาค่า $\hat{k}$ และ $\hat{n}$ ได้แล้ว เราสามารถหาค่า $\hat{q}_x$ ได้ดังนี้

$$\begin{aligned} \hat{q}_x &= P(T \le 1) \\ &= 1-\exp\left[\frac{\hat{k}}{\hat{n}+1}\right] \end{aligned} \qquad (2.8)$$

<u>การกำหนดค่าเริ่มต้น k_0 และ n_0</u>

ในการวิจัยครั้งนี้ ผู้วิจัยได้ใช้วิธีเปอร์เซ็นต์ไทล์ แมทช์ชิ่ง (Percentile Matching) ซึ่งเป็นวิธีการประมาณค่าพารามิเตอร์อย่างง่าย และไม่ค่อยมีความซับซ้อนมากนักในการคำนวณ คือ การแก้สมการ p สมการ เพื่อหาค่าประมาณพารามิเตอร์ p ค่า โดยสมการจะอยู่ในรูป

$$F_T(t_i;\theta) = F_m(t_i) \quad , \quad i = 1,2,\ldots,p$$

โดย $F_T(t_i;\theta)$ เป็นฟังก์ชันการแจกแจง (Distribution Function) ของ T

$F_m(t_i)$ เป็นการแจกแจงเชิงการทดลอง (Empirical Distribution)

θ เป็นเวคเตอร์ p มิติ สำหรับพารามิเตอร์ของการแจกแจง

$t_1,t_2,\ldots,t_p$ เป็นค่าใดๆ ที่เลือกมาจากข้อมูลที่มีอยู่

สำหรับการแจกแจงแบบไวบูลล์ ซึ่งมีพารามิเตอร์ p = 2 ค่า โดยสามารถเขียนสมการได้ดังนี้

$$1-\exp\left[-\frac{k_0}{n_0+1}t_1^{n_0+1}\right] = F_m(t_1)$$

$$1-\exp\left[-\frac{k_0}{n_0+1}t_2^{n_0+1}\right] = F_m(t_2)$$

จากสองสมการนี้ แก้สมการหาค่า n_0 และ k_0 ได้ดังนี้

$$n_0 = \frac{\ln\{\ln[1-F_m(t_1)]\} - \ln\{\ln[1-F_m(t_2)]\}}{\ln t_1 - \ln t_2} - 1 \qquad (2.9)$$

คำนวณค่า k_0 โดยแทนค่า n_0 ที่คำนวณได้จากข้างต้นลงในสูตรข้างล่างนี้

$$k_0 = -\frac{(n_0+1)\ln[1-F_m(t_2)]}{t_2^{n_0+1}} \qquad (2.10)$$

2 T มีการแจกแจงแบบกอมเพิรตซ์

$$\begin{aligned} L &= \prod_{i=1}^{m}\left\{Bc^{t_i}\exp\left[-\frac{B}{\ln c}(c^{t_i}-1)\right]\right\}^{\delta_i}\left\{\exp\left[-\frac{B}{\ln c}(c^{t_i}-1)\right]\right\}^{1-\delta_i} \\ &= \prod_{i=1}^{m}\left(Bc^{t_i}\right)^{\delta_i}\exp\left[-\frac{B}{\ln c}(c^{t_i}-1)\right] \end{aligned}$$

$$\ln L = \ln B\sum_{i=1}^{m}\delta_i + \ln c\sum_{i=1}^{m} t_i\delta_i - \frac{B}{\ln c}\sum_{i=1}^{m}(c^{t_i}-1)$$

$$g_1(B,c) = \frac{\partial \ln L}{\partial B} = \frac{1}{B}\sum_{i=1}^{m}\delta_i - \frac{1}{\ln c}\sum_{i=1}^{m}(c^{t_i}-1) \qquad (2.11)$$

$$g_{11}(B,c) = \frac{\partial^2 \ln L}{\partial B^2} = -\frac{1}{B^2}\sum_{i=1}^{m}\delta_i \qquad (2.12)$$

$$g_{12}(B,c) = \frac{\partial^2 \ln L}{\partial c\,\partial B} = \frac{1}{c(\ln c)^2}\sum_{i=1}^{m}(c^{t_i}-1) - \frac{1}{\ln c}\sum_{i=1}^{m} t_i c^{t_i-1} \qquad (2.13)$$

$$g_2(B,c) = \frac{\partial \ln L}{\partial c} = \frac{1}{c}\sum_{i=1}^{m}\delta_i t_i + \frac{B}{c(\ln c)^2}\sum_{i=1}^{m}(c^{t_i}-1) - \frac{B}{\ln c}\sum_{i=1}^{m} t_i c^{t_i-1}$$

(2.14)

$$g_{21}(B,c) = \frac{\partial^2 \ln L}{\partial B \partial c} = \frac{1}{c(\ln c)^2}\sum_{i=1}^{m} (c^{t_i} - 1) - \frac{1}{\ln c}\sum_{i=1}^{m} t_i c^{t_i - 1} \quad (2.15)$$

$$g_{22}(B,c) = \frac{\partial^2 \ln L}{\partial c^2} = -\frac{1}{c^2}\sum_{i=1}^{m} \delta_i t_i - \left[\frac{B}{(c\ln c)^2} + \frac{2B}{c^2(\ln c)^3}\right]\sum_{i=1}^{m} (c^{t_i} - 1)$$

$$+ \frac{2B}{c(\ln c)^2}\sum_{i=1}^{m} t_i c^{t_i - 1} - \frac{B}{\ln c}\sum_{i=1}^{m} t_i (t_i - 1) c^{t_i - 2}$$

(2.16)

การหาค่าประมาณพารามิเตอร์ B และ c จะกระทำได้โดยอาศัยเทคนิคการวิเคราะห์เชิงตัวเลขด้วยวิธีนิวตัน-ราพซ์สัน ซึ่งพิจารณาสองสมการนี้

$$g_1(B,c) = 0$$
$$g_2(B,c) = 0$$

โดยใช้วิธีการประมาณเชิงเส้นและกำหนดค่าเริ่มต้นเป็น B_0 และ c_0 โดยวิธีเปอร์เซ็นต์ไทล์ แมทช์ชิ่ง จะได้สมการดังข้างล่างนี้

$$g_1(B_0,c_0) + g_{11}(B_0,c_0)(B - B_0) + g_{12}(B_0,c_0)(c - c_0) = 0$$
$$g_2(B_0,c_0) + g_{21}(B_0,c_0)(B - B_0) + g_{22}(B_0,c_0)(c - c_0) = 0$$

หรือสามารถเขียนให้อยู่ในรูปเมตริกซ์ได้ คือ

$$\begin{bmatrix} g_{11}(B_0,c_0) & g_{12}(B_0,c_0) \\ g_{21}(B_0,c_0) & g_{22}(B_0,c_0) \end{bmatrix} \begin{bmatrix} B - B_0 \\ c - c_0 \end{bmatrix} = \begin{bmatrix} -g_1(B_0,c_0) \\ -g_2(B_0,c_0) \end{bmatrix}$$

และใช้กระบวนการการทำซ้ำ จนกระทั่งได้ค่า $\hat{B}$ และ $\hat{c}$ ที่ต้องการหลังจากหาค่า $\hat{B}$ และ $\hat{c}$ ได้แล้ว เราสามารถหาค่า $\hat{q}_x$ ได้ดังนี้

$$\hat{q}_x = P(T \le 1)$$

$$= 1-\exp\left[-\frac{\hat{B}}{\ln \hat{c}}(\hat{c}-1)\right] \quad (2.17)$$

<u>การกำหนดค่าเริ่มต้น B_0 และ c_0</u>

ผู้วิจัยได้ใช้วิธีเปอร์เซ็นต์ไทล์ แมทช์ชิ่ง เช่นเดียวกับการแจกแจงแบบไวบูลล์ ซึ่งสามารถเขียนสมการให้อยู่ในรูป

$$1-\exp\left[-\frac{B_0}{\ln c_0}\left(c_0^{t_1}-1\right)\right] = F_m(t_1)$$

$$1-\exp\left[-\frac{B_0}{\ln c_0}\left(c_0^{t_2}-1\right)\right] = F_m(t_2)$$

เนื่องจากการแก้สมการสองสมการดังกล่าวข้างต้นนี้กระทำได้โดยยาก ดังนั้น จึงต้องอาศัยเทคนิคการวิเคราะห์เชิงตัวเลข คือ วิธีแบ่งครึ่ง (Bisection Method) โดยพยายามจัดสมการให้อยู่ในรูปใหม่ดังนี้

$$\frac{c_0^{t_1}-1}{c_0^{t_2}-1} = \frac{\ln\left[1-F_m(t_1)\right]}{\ln\left[1-F_m(t_2)\right]} = K$$

$c_0^{t_1} - 1 = Kc_0^{t_2} - K$ หรือ $Kc_0^{t_2} - c_0^{t_1} + (1-K) = 0$

กำหนดให้ $x = c_0^{t_1}$, $r = \dfrac{t_2}{t_1}$

$$f(x) = Kx^r - x + (1-K) = 0 \quad (2.18)$$

พิจารณากรณี $0<r<1$ $\left(t_2<t_1\right)$ และหาอนุพันธ์สมการ (2.18) พบว่า

$$f'(x) = Krx^{r-1} - 1 = 0$$

ซึ่งสามารถแก้สมการหาคำตอบได้ดังนี้

$$x_0 = (Kr)^{1/(1-r)}$$

หลังจากนี้จะใช้วิธีแบ่งครึ่งในการหาคำตอบซึ่งอยู่ในช่วง $(0, x_0)$ เมื่อคำนวณหาค่า x ได้แล้ว จะกระทำการคำนวณหาค่า c_0 ได้ดังนี้

$$c_0 = \exp\left(\frac{\ln x}{t_1}\right) \qquad (2.19)$$

และคำนวณหาค่า B_0 โดยแทนค่า c_0 ที่คำนวณได้จากข้างต้นลงสูตรข้างล่างนี้

$$B_0 = -(\ln c_0)\frac{\ln[1-F_m(t_2)]}{c_0^{t_2}-1} \qquad (2.20)$$

วิธีการประมาณแบบเบส์

การประมาณด้วยวิธีแบบคลาสสิคนั้น จะอาศัยข้อมูลจากกลุ่มตัวอย่างที่นำมาศึกษาเพียงอย่างเดียว แต่สำหรับวิธีการประมาณแบบเบส์นี้ เป็นวิธีทางสถิติซึ่งไม่เพียงแต่อาศัยข้อมูลจากกลุ่มตัวอย่างที่นำมาศึกษาเท่านั้น แต่ยังอาศัยข้อมูลหรือประสบการณ์ในอดีตเข้าช่วยในการประมาณค่าอีกด้วย โดยจากข้อมูลหรือประสบการณ์ในอดีตนี้สามารถนำมาหาการแจกแจงก่อนการทดลอง(Prior Distribution) และเมื่อทราบการแจกแจงก่อนการทดลองรวมทั้งข้อมูลจากกลุ่มตัวอย่างที่นำมาศึกษา ทำให้สามารถคำนวณการแจกแจงหลังการทดลอง(Posterior Distribution) ได้ โดยการแจกแจงดังกล่าวนี้ จะนำไปใช้ในการประมาณค่าที่ต้องการ ดังนั้น จึงสรุปได้ว่าวิธีการประมาณแบบเบส์นี้ จะให้ค่าประมาณที่ดีหรือไม่นั้นขึ้นอยู่กับข้อมูลจากกลุ่มตัวอย่างที่นำมาศึกษาและข้อมูลหรือประสบการณ์ในอดีตที่ผ่านมา

ภายใต้กำลังของมรณะ (Force of Mortality , $\mu_{x+0.5}$) เป็นค่าคงที่

กำหนดให้ $\mu_{x+0.5} = \mu$ ซึ่งมีการแจกแจงแบบเอกซโพเนนเชียล โดยมีฟังก์ชันความหนาแน่นก่อนการทดลอง (Prior Density Function) ดังนี้

$$\pi(\mu) = \beta e^{-\beta\mu} \quad , \mu \geq 0$$

เนื่องจากภายใต้กำลังของมรณะเป็นค่าคงที่ ฟังก์ชันความหนาแน่นของ T อยู่ในรูป

$$f(t_i) \;=\; \mu^{\delta_i}\exp[-\mu t_i] \quad , t_i \geq 0$$

ฟังก์ชันภาวะน่าจะเป็น หรือ ฟังก์ชันความหนาแน่นร่วมระหว่าง T และ μ อยู่ในรูป

$$p(\underset{\sim}{t};\mu) \;=\; (\mu)^{\sum_{i=1}^{m}\delta_i}\exp\left[-\mu\sum_{i=1}^{m}t_i\right]\pi(\mu)$$

$$= \;\beta(\mu)^{\sum_{i=1}^{m}\delta_i}\exp\left[-\left(\beta+\sum_{i=1}^{m}t_i\right)\mu\right] \qquad (2.21)$$

ฟังก์ชันความหนาแน่นหลังการทดลอง (Posterior Density Function) คือ

$$\pi\left(\mu\middle|\underset{\sim}{t}\right) \;=\; \frac{\beta(\mu)^{\sum_{i=1}^{m}\delta_i}\exp\left[-\left(\beta+\sum_{i=1}^{m}t_i\right)\mu\right]}{\int_0^\infty \beta(s)^{\sum_{i=1}^{m}\delta_i}\exp\left[-\left(\beta+\sum_{i=1}^{m}t_i\right)s\right]ds}$$

$$= \;\frac{\left(\beta+\sum_{i=1}^{m}t_i\right)^{(1+\sum_{i=1}^{m}\delta_i)}}{\Gamma\left(1+\sum_{i=1}^{m}\delta_i\right)}(\mu)^{\sum_{i=1}^{m}\delta_i}\exp\left[-\left(\beta+\sum_{i=1}^{m}t_i\right)\mu\right] \qquad (2.22)$$

นั่นคือ $\mu\big|\underset{\sim}{T}$ มีการแจกแจงแบบแกมมา 2 พารามิเตอร์ คือ

$\alpha' = 1+\sum_{i=1}^{m}\delta_i$ และ $\beta' = \beta+\sum_{i=1}^{m}t_i$ ตามลำดับ

ดังนั้น $\hat{q}_x \;=\; \hat{E}\left(1-e^{-\mu}\middle|T\right)$

$$= \;1-\left(\frac{\hat{\beta}'}{1+\hat{\beta}'}\right)^{\hat{\alpha}'} \qquad (2.23)$$