บทที่ 2

ตัวสถิติที่ใช้ในงานวิจัย

การวิเคราะห์ข้อมูลที่เกี่ยวกับอายุการใช้งาน (life time) หรือช่วงเวลาก่อนล้มเหลว (failure time) มักจะเกี่ยวข้องกับงานด้านวิศวกรรมศาสตร์ วิทยาศาสตร์ และงานด้านการแพทย์ โดยข้อมูลมักจะมีลักษณะเกิดการตัดทิ้งอันเนื่องมาจากข้อกำหนดของเวลาและข้อจำกัดอื่น ๆ ในบทนี้จะกล่าวถึงรายละเอียดเกี่ยวกับชนิดของข้อมูลที่มีค่าถูกตัดทิ้ง (types of data censoring) ฟังก์ชันการอยู่รอด และวิธีการประมาณช่วงความเชื่อมั่นของฟังก์ชันการอยู่รอดซึ่งมีรายละเอียดต่าง ๆ ดังนี้

ชนิดของข้อมูลที่มีค่าถูกตัดทิ้ง (Types of data censoring)

1. ข้อมูลถูกตัดทิ้งประเภทที่ 1 (Type I censoring) เป็นลักษณะของข้อมูลที่มีการกำหนดเวลาของการเกิดค่าถูกตัดทิ้งเอาไว้ล่วงหน้าตัวอย่างของข้อมูลที่ถูกตัดทิ้งประเภทที่ 1 เช่นการทำประกันชีวิตแบบสะสมทรัพย์ (endowment insurance) เป็นการทำประกันชีวิตที่มีการกำหนดระยะเวลาที่แน่นอนของสัญญาประกันชีวิต โดยอาจกำหนดเวลาของการคุ้มครองไว้ 15 ปี ถ้าผู้เอาประกันเสียชีวิตภายใน 15 ปีก็จะได้รับความคุ้มครองการเสียชีวิตซึ่งข้อมูลในลักษณะนี้เป็นข้อมูลที่ไม่ถูกตัดทิ้ง (Uncensored Data) เพราะทราบเวลาที่ผู้เอาประกันเสียชีวิตที่แน่นอน และเมื่อครบกำหนดสัญญากรมธรรม์จะมีผู้เอาประกันบางคนที่ยังมีชีวิตอยู่รอดซึ่งก็จะได้รับเงินคืนเท่ากับจำนวนเงินเอาประกันข้อมูลในลักษณะนี้จะเป็นข้อมูลที่ถูกตัดทิ้ง (Censored data) เพราะไม่สามารถทราบเวลาที่ผู้เอาประกันเสียชีวิตที่แน่นอนได้

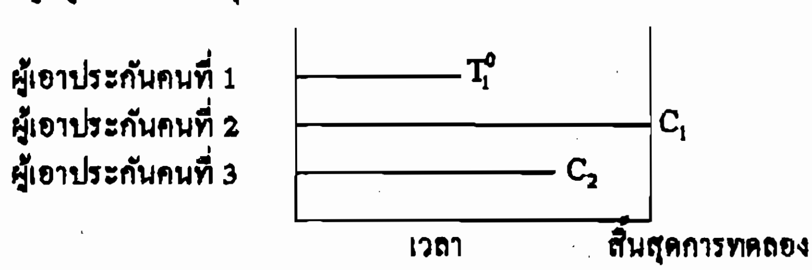
ให้ t_c เป็นเวลาที่กำหนดไว้ และให้ $T_1^0, T_2^0, \ldots\ldots, T_n^0$ เป็นตัวแปรสุ่มเวลาของการอยู่รอดมีการแจกแจงที่เหมือนกันและเป็นอิสระกันจะได้ค่าสังเกตสุ่ม $T_1, T_2, \ldots\ldots, T_n$ ซึ่ง

$$T_i = \begin{cases} T_i^0 & ; \quad T_i^0 \le t_c \quad \text{ข้อมูลไม่ถูกตัดทิ้ง} \\ t_c & ; \quad T_i^0 > t_c \quad \text{ข้อมูลถูกตัดทิ้ง} \end{cases}$$

2. ข้อมูลถูกตัดทิ้งประเภทที่ 2 (Type II censoring) เป็นลักษณะของข้อมูลที่จะต้องกำหนดจำนวนข้อมูลที่ไม่ถูกตัดทิ้งไว้ล่วงหน้า นั่นคือเมื่อจำนวนของค่าสังเกตที่ไม่ถูกตัดทิ้งเกิดขึ้นครบตามจำนวนที่กำหนดไว้จะหยุดทำการทดลอง ตัวอย่างของข้อมูลที่ถูกตัดทิ้งประเภทที่ 2 เช่นในการทำประกันของผู้เอาประกันชีวิตจำนวน n คน ต้องการศึกษาถึงระยะเวลาที่ผู้เอาประกันจะเสียชีวิตโดยจะทำการเก็บข้อมูลคือระยะเวลาที่ผู้เอาประกันชีวิตจะเสียชีวิตจำนวน r ($r \le n$) คน เมื่อเก็บข้อมูลครบ r คนแล้วจึงหยุดการเก็บข้อมูล จะได้ค่าสังเกตที่ไม่ถูกตัดทิ้งเป็น $T_1^0 \le T_2^0 \le \ldots\ldots \le T_r^0$ และ $T_{r+1}^0 \le T_{r+2}^0 \le \ldots\ldots \le T_n^0$ เป็นลำดับของตัวแปรสุ่มเวลาของการอยู่รอดที่เป็นค่าสังเกตที่ถูกตัดทิ้งซึ่ง $T_i^0 \ge T_r^0$, $i = r+1, r+2, \ldots\ldots, n$ ดังนั้นค่าสังเกต T_i , $i = 1,2,\ldots\ldots,n$ ได้ว่า

$$
\begin{aligned}
T_1 &= T_1^0 \\
T_2 &= T_2^0 \\
&\vdots \\
T_r &= T_r^0 \\
T_{r+1} &= T_r^0 \\
&\vdots \\
T_n &= T_r^0
\end{aligned}
$$

3. ข้อมูลถูกตัดทิ้งแบบสุ่ม (Random censoring) เป็นลักษณะของข้อมูลที่มีการตัดข้อมูลในระหว่างการทดลองและหลังสิ้นสุดการทดลอง เช่นในงานด้านการประกันชีวิตต้องการศึกษาถึงระยะเวลาของการอยู่รอดของกรมธรรม์ประกันชีวิตในช่วงระยะเวลา 15 ปี ผู้เอาประกันชีวิตบางคนอาจเสียชีวิตก่อนระยะเวลาที่กำหนดข้อมูลในลักษณะนี้จะเป็นข้อมูลที่ไม่ถูกตัดทิ้ง (Uncensored Data) แต่จะมีผู้เอาประกันชีวิตบางคนที่ขาดการส่งเบี้ยประกันมีผลทำให้กรมธรรม์ประกันชีวิตขาดอายุ (Lapse) หรือบางคนที่มีชีวิตอยู่รอดเมื่อครบกำหนดอายุกรมธรรม์ (Ender) ข้อมูลในลักษณะนี้จะเป็นข้อมูลที่ถูกตัดทิ้ง (Censored Data) รูปที่ 2.1 แสดงลักษณะความเป็นไปได้ของข้อมูลถูกตัดทิ้งแบบสุ่ม



รูปที่ 2.1

ผู้เอาประกันคนที่ 1 เริ่มทำประกันชีวิต ณ เวลา $t = 0$ และเสียชีวิตที่เวลา T_1^0 ค่าสังเกตนี้เป็นค่าสังเกตที่ไม่ถูกตัดทิ้ง

ผู้เอาประกันคนที่ 2 เริ่มทำประกันชีวิต ณ เวลา $t = 0$ และยังมีชีวิตอยู่รอดเมื่อครบกำหนดอายุกรมธรรม์ค่าสังเกตนี้เป็นค่าสังเกตที่ถูกตัดทิ้ง

ผู้เอาประกันคนที่ 3 เริ่มทำประกันชีวิต ณ เวลา $t = 0$ และหยุดการส่งเบี้ยประกันเมื่อเวลา C_2 ค่าสังเกตนี้เป็นค่าสังเกตที่ถูกตัดทิ้ง

ถ้า $T_1^0, T_2^0, \ldots\ldots, T_n^0$ เป็นตัวแปรสุ่มเวลาของการอยู่รอดมีการแจกแจงเหมือนกันและเป็นอิสระกันมีฟังก์ชันการแจกแจงเป็น F และ $C_1, C_2, \ldots\ldots, C_n$ เป็นตัวแปรสุ่มของค่าสังเกตที่ถูกตัดทิ้งมีการแจกแจงที่เหมือนกันและเป็นอิสระกันมีฟังก์ชันการแจกแจงเป็น G

ดังนั้นถ้า T_i^0 และ C_i , $i = 1,2,\ldots\ldots,n$ เป็นอิสระกันให้

$T_i = \min(T_i^0, C_i)$ จะได้ค่าสังเกตสุ่ม $T_1, T_2, \ldots\ldots, T_n$ ดังนี้

$$T_i = \begin{cases} T_i^0 & ; \quad T_i^0 \le C_i \quad \text{ข้อมูลไม่ถูกตัดทิ้ง} \\ C_i & ; \quad T_i^0 > C_i \quad \text{ข้อมูลถูกตัดทิ้ง} \end{cases}$$

ให้ δ_i เป็นครรชนีของ T_i แสดงว่า T_i เป็นข้อมูลที่ถูกตัดทิ้งหรือเป็นข้อมูลที่ไม่ถูกตัดทิ้ง โดยค่าของ δ_i กำหนดดังนี้

$$\delta_i = \begin{cases} 1 & ; \quad T_i^0 \le C_i \\ 0 & ; \quad T_i^0 > C_i \end{cases}$$

ในงานวิจัยครั้งนี้ผู้ทำวิจัยจะทำการศึกษาในกรณีที่ข้อมูลมีค่าถูกตัดทิ้งแบบสุ่มเท่านั้น

การแจกแจงการอยู่รอด

ให้ตัวแปรสุ่ม T แทนเวลาการอยู่รอด

f(t) แทน ฟังก์ชันความหนาแน่น (Density function)

F(t) แทน ฟังก์ชันการแจกแจง (Distribution function)

S(t) แทน ฟังก์ชันการอยู่รอด (Survival function) มีค่าเท่ากับความน่าจะเป็นที่หน่วยตัวอย่างมีการอยู่รอดนานกว่าเวลา t นั่นคือ

$$
\begin{aligned}
S(t) &= \text{Prob(หน่วยตัวอย่างมีการอยู่รอดนานกว่าเวลา } t) \\
&= \text{Prob}(T > t) \\
&= 1 - F(t)
\end{aligned}
$$

คุณสมบัติของ S(t) มีดังนี้

1. S(t) เป็นฟังก์ชันไม่เพิ่ม (nonincreasing function) นั่นคือ $S(t) \geq S(t+1)$
2. S(t) เป็นฟังก์ชันต่อเนื่องของ t
3. $S(t) = 1$ เมื่อ $t = 0$
4. $S(t) = 0$ เมื่อ $t = \infty$

h(t) แทนฟังก์ชันการสูญเสีย (Hazard function) มีค่าเท่ากับลิมิตของความน่าจะเป็นที่หน่วยตัวอย่างสูญเสียในช่วงเวลาสั้น ๆ จาก t ถึง $t+\Delta t$ ต่อหน่วยเวลา Δt เมื่อแต่ละหน่วยตัวอย่างมีการอยู่รอดถึงเวลา t นั้นคือ

$$h(t) = \lim_{\Delta t \to 0} \text{Prob\{หน่วยตัวอย่างที่มีอายุ } t \text{ สูญเสียในช่วง}(t, t+\Delta t)\}/\Delta t$$

ฟังก์ชันการสูญเสียมีคุณสมบัติดังนี้

1. $h(t) \geq 0 \quad ; \quad -\infty < t < \infty$
2. $\lim\limits_{t \to -\infty} \int_{-\infty}^{t} h(t)\,dt = 0$ และ

$$\lim_{t \to \infty} \int_{-\infty}^{t} h(t)\,dt = \infty$$

ความสัมพันธ์ของฟังก์ชันการอยู่รอด ฟังก์ชันความหนาแน่น ฟังก์ชันการแจกแจง และฟังก์ชันการสูญเสีย

1. $f(t) = F'(t) = -S'(t)$
2. $h(t) = \dfrac{f(t)}{S(t)} = \dfrac{-S'(t)}{S(t)} = \dfrac{-d\ln(S(t))}{dt}$
3. $\int_0^t h(u)\,du = -\int_0^t \dfrac{S'(u)}{S(u)}\,du = -\ln(S(t))$

$$S(t) = \exp\left(-\int_0^t h(u)\,du\right)$$

และ $f(t) = h(t)\exp\left(-\int_0^t h(u)\,du\right)$

รูปแบบการแจกแจงการอยู่รอดที่ใช้ในงานวิจัยนี้มี 3 แบบคือ

1. การแจกแจงไวบูลล์ (Weibull Distribution)

การแจกแจงแบบไวบูลล์ให้ค่าฟังก์ชันการสูญเสียหลายค่า ซึ่งขึ้นกับค่าพารามิเตอร์ λ และ γ

ถ้า $\gamma < 1$ และ $\lambda = 1$ แล้วฟังก์ชันการสูญเสียจะลดลงเมื่อค่า t เพิ่มขึ้น

ถ้า $\gamma > 1$ และ $\lambda = 1$ แล้วฟังก์ชันการสูญเสียจะเพิ่มขึ้นเมื่อค่า t เพิ่มขึ้น

ถ้า $\gamma = 1$ และ $\lambda = 1$ แล้วฟังก์ชันการสูญเสียจะคงที่เมื่อค่า t เพิ่มขึ้น

โดยที่รูปแบบของฟังก์ชันต่าง ๆ เป็นดังนี้

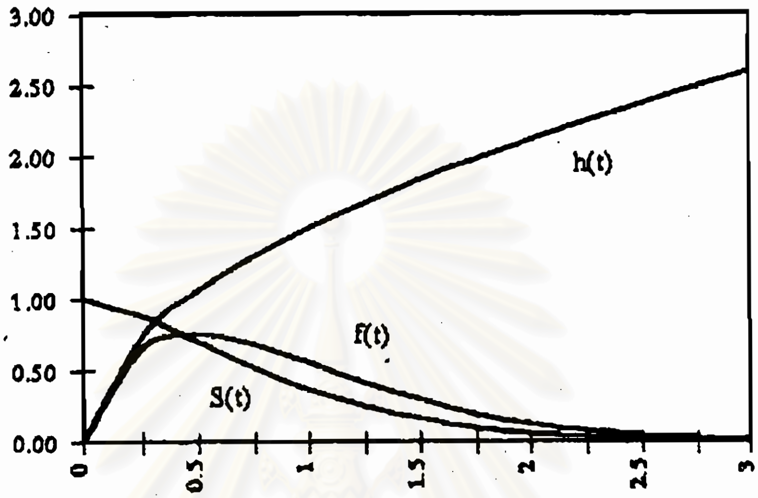
$$f(t) = \begin{cases} \lambda\gamma(\lambda t)^{\gamma-1}\exp(-(\lambda t)^{\gamma}) & ;\quad 0 < t < \infty,\ \lambda > 0,\ \gamma > 0 \\ 0 & ;\quad \text{อื่น ๆ} \end{cases}$$

$$F(t) = 1 - \exp(-(\lambda t)^{\gamma})$$

$$S(t) = \exp(-(\lambda t)^{\gamma})$$

$$h(t) = \lambda\gamma(\lambda t)^{\gamma-1}$$

สำหรับรูปแบบของฟังก์ชัน f(t) ,S(t) และ h(t) เมื่อการแจกแจงการอยู่รอดมีรูปแบบเป็นไวบูลล์ที่พารามิเตอร์ $\lambda = 1.00$ และ $\gamma = 1.50$ สามารถแสดงได้ดังรูปที่ 2.2



รูปที่ 2.2

2. การแจกแจงลอกนอร์มอล (Lognormal Distribution)

การแจกแจงลอกนอร์มอลพบว่าเมื่อ t เพิ่มมากขึ้น ฟังก์ชันการสูญเสียจะมีค่าเพิ่มขึ้นเรื่อยๆ เมื่อเพิ่มจนถึงค่าสูงสุดแล้วจะลดลงเข้าสู่ 0 เมื่อเวลาเข้าใกล้อนันต์ (Watson&Wells,1961) ดังนั้นการแจกแจงนี้จึงเหมาะสม สำหรับรูปแบบของเวลาการอยู่รอดที่มีฟังก์ชันการสูญเสียเพิ่มขึ้นในช่วงระยะแรกและลดลงในเวลาต่อมา รูปแบบของฟังก์ชันต่าง ๆ เป็นดังต่อไปนี้

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

$$S(t) = \frac{1}{\sigma\sqrt{2\pi}}\int_t^{\infty}\frac{1}{u}\exp\left(-\frac{(\ln(u)-\mu)^2}{2\sigma^2}\right)du$$

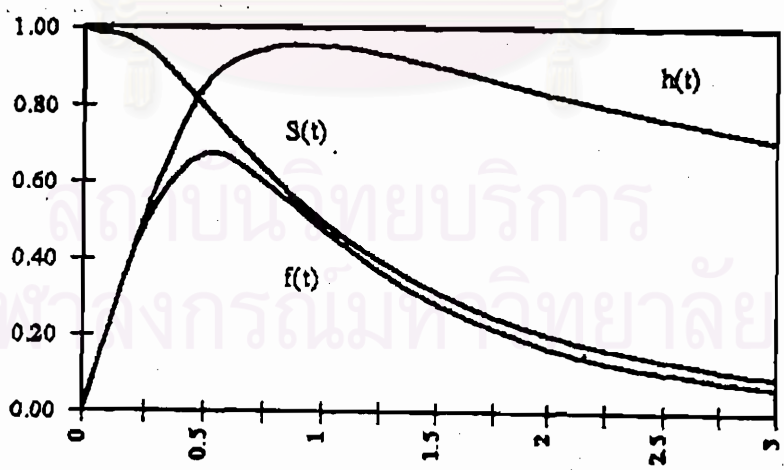
$$= 1-\Phi\left(\frac{\ln(t)-\mu}{\sigma}\right)$$

$\Phi(x)$ คือฟังก์ชันการแจกแจงปกติมาตรฐาน

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x \exp(-\frac{u^2}{2})\,du$$

$$h(t) = \frac{\frac{1}{t\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln(t)-\mu)^2}{2\sigma^2}\right)}{1-\Phi\left(\frac{\ln(t)-\mu}{\sigma}\right)}$$

สำหรับรูปแบบของฟังก์ชัน f(t) , S(t) และ h(t) เมื่อการแจกแจงการอยู่รอดมีรูปแบบเป็นลอกนอร์มอลที่พารามิเตอร์ $\mu = 0.0$ และ $\sigma^2 = 0.7$ สามารถแสดงได้ดังรูปที่ 2.3



รูปที่ 2.3

3. การแจกแจงกอมเพริตซ์ (Gompertz Distribution)

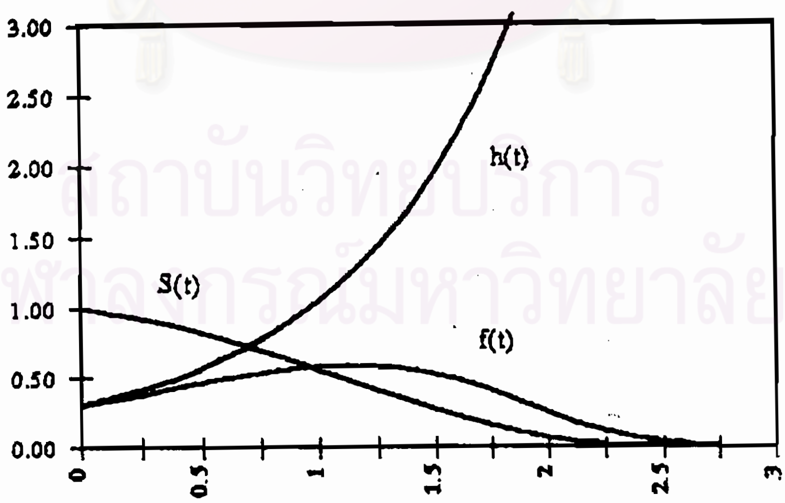
การแจกแจงกอมเพริตซ์พบว่าเมื่อ t เพิ่มมากขึ้น ฟังก์ชันการสูญเสียจะมีค่าเพิ่มขึ้นเรื่อย ๆ ดังนั้นการแจกแจงนี้จึงเหมาะสมสำหรับรูปแบบของเวลาการอยู่รอดที่มีฟังก์ชันการสูญเสียเพิ่มขึ้นเรื่อย ๆ เมื่อเวลาเพิ่มขึ้น รูปแบบของฟังก์ชันต่าง ๆ เป็นดังต่อไปนี้

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

$$S(t) = \exp\left(\frac{B}{\ln(C)}(1-C^t)\right)$$

$$h(t) = BC^t$$

สำหรับรูปแบบของฟังก์ชัน f(t) , S(t) และ h(t) เมื่อการแจกแจงการอยู่รอดมีรูปแบบเป็นกอมเพริตซ์ที่พารามิเตอร์ B = 0.30 และ C = 3.50 สามารถแสดงได้ดังรูปที่ 2.4



รูปที่ 2.4

สถิติที่ใช้ในการวิจัย

ตัวประมาณพีแอล (Product Limit Estimator)[1]

ในหัวข้อนี้กล่าวถึงวิธีการประมาณฟังก์ชันการอยู่รอด $S(t)$ โดยใช้ตัวประมาณพีแอล ซึ่งตัวประมาณพีแอลถูกพัฒนาขึ้นโดยแค็พแลนและไมเออร์ (Kaplan and Meier 1958)

แนวคิดของฟังก์ชันการอยู่รอดคือ ตัวอย่างที่มีชีวิตอยู่รอด 2 ปี หมายถึง ตัวอย่างที่มีชีวิตอยู่รอดในปีแรกและมีชีวิตอยู่รอดมากกว่า 1 ปี ดังนั้นความน่าจะเป็นของการมีชีวิตอยู่รอด 2 ปี คือความน่าจะเป็นของการมีชีวิตอยู่รอดในปีแรกและมีชีวิตอยู่รอดมากกว่า 1 ปี

$S(2)$ = Prob(มีชีวิตอยู่รอดในปีแรกและมีชีวิตอยู่รอดมากกว่า 1 ปี)

= Prob(มีชีวิตอยู่รอด 2 ปีเมื่อมีชีวิตอยู่รอดมาแล้ว 1 ปี) * Prob(มีชีวิตอยู่รอด 1 ปี)

$= \text{Prob}(T>2/T>1) * \text{Prob}(T>1)$

สำหรับในกรณีทั่วไป สามารถเขียนเป็นสูตรได้ดังนี้เมื่อ $k \geq 2$

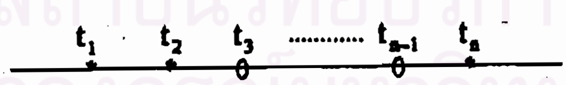
$$S(k) = p_1 * p_2 * \ldots\ldots * p_k$$

$$= \prod_{i \leq k} p_i$$

p_i คือค่าสัดส่วนของตัวอย่างที่อยู่รอดถึงเวลา t_i หลังจากอยู่รอดมาแล้ว t_{i-1}

$$= \text{Prob}(T > t_i \,/\, T > t_{i-1})$$

ในทางปฏิบัติจะสามารถคำนวนฟังก์ชันการอยู่รอดได้โดยสมมติว่าค่าสังเกตของอายุที่อยู่รอด (Survival Time) จำนวน n มีค่าสังเกตเป็น $t_1, t_2, \ldots\ldots, t_n$ นำค่าสังเกตมาเรียงลำดับจากค่าน้อยไปมากจะได้ $t_1 < t_2 < \ldots\ldots < t_n$



โดยที่เครื่องหมาย * หมายถึง ค่าสังเกตที่ไม่ถูกตัดทิ้ง

0 หมายถึง ค่าสังเกตที่ถูกตัดทิ้ง

[1] Kaplan E.L, and Meier P. (1958) "Nonparametric Estimation From Incomplete Ob Observations," Journal of the American Statistical Association, 53 , 457-481

Rupert G. Miller **Survival Analysis** (New York John Wiley 1981) pp. 46-50

ฟังก์ชันการอยู่รอดหาได้จากสูตร

$$\hat{S}(t) = \prod_{\{i: t_i \le t\}} \left(\frac{n-i}{n-i+1}\right)^{\delta_i}$$

n เป็นจำนวนข้อมูลทั้งหมดที่ไม่ถูกตัดทิ้งและถูกตัดทิ้ง

i เป็นลำดับที่ของข้อมูลในการเรียงลำดับค่าจากน้อยไปมาก

การประมาณโดยตัวประมาณพีแอล (PL Estimator) รายละเอียดดูได้จากภาคผนวก ก.

i เป็นลำดับที่ของ t_i ในการเรียงลำดับค่าจากน้อยไปมาก

$$\hat{V}ar(\hat{S}(t)) = (\hat{S}(t))^2 \sum_{\{i: t_i \le t\}} \frac{\delta_i}{(n-i)(n-i+1)}$$

คุณสมบัติของตัวประมาณพีแอล[2]

ตัวประมาณพีแอลเป็นตัวประมาณที่ไม่เอนเอียงและคงที่ (Unbiased and Consistent)

ตัวประมาณแบบช่วง

ในกรณีที่ตัวอย่างมีขนาดใหญ่จะใช้การประมาณปรกติ (Normal Approximate) ในการประมาณช่วงความเชื่อมั่นของฟังก์ชันการอยู่รอด ทั้งนี้เพราะตัวประมาณพีแอลจะมีการแจกแจงปกติ[3]

จะได้ช่วงความเชื่อมั่นสำหรับฟังก์ชันการอยู่รอดเป็น

$$\hat{S}(t) \pm Z_{\frac{\alpha}{2}} SE(\hat{S}(t))$$

---

[2] Diok London, FSA **Survival Models and Their Estimation** (ACTEX Publications 1988) pp. 167

[3] Mahesh K.B. Parmar and David Machin **Survival Analysis A Practical Approach** (New York John Wiley&sons 1995) pp. 36

สำหรับในการวิจัยนี้เนื่องจากข้อมูลมีค่าถูกตัดทิ้ง ดังนั้นตัวสถิติที่ใช้ในการประมาณค่าแบบช่วงของฟังก์ชันการอยู่รอดจะสามารถคำนวณได้ดังนี้

ภายใต้เงื่อนไขการตัดทิ้งแบบสุ่ม (Random censoring model)

$$T_i = \min(T_i^0, C_i)$$

โดยที่ T_i^0 เป็นตัวแปรสุ่มของค่าสังเกตุที่ไม่ถูกตัดทิ้งมีฟังก์ชันการแจกแจงเป็น F และฟังก์ชันการอยู่รอดเป็น S

C_i เป็นตัวแปรสุ่มของค่าสังเกตุที่ถูกตัดทิ้งมีฟังก์ชันการแจกแจงเป็น G และฟังก์ชันการอยู่รอดเป็น H

ถ้า F และ G เป็นการแจกแจงแบบต่อเนื่อง และ $T < T_F$ โดย $F(t) < 1$

ดังนั้น $n \to \infty$; $n^{\frac{1}{2}}\left(\hat{S}_n(t) - S(t)\right) \to Z^*(t)$

โดยที่ $Z^*(t)$ คือ Gaussian process ด้วย mean = 0 และ covariance function

$$\mathrm{Cov}\{Z^*(s), Z^*(t)\} = \sigma^2(s)S(s)S(t) \quad ; \quad s \le t$$

$$\sigma^2(s) = \int_0^s (1-F)^{-2}(1-H)^{-1}\,dF \quad ; s < T_F$$

ในทางปฏิบัติสามารถประมาณ $\sigma^2(s)$ ได้โดย

$$\hat{\sigma}^2(t) = n \sum_{\{i: t_i \le s\}} \frac{\delta_i}{(n-i)(n-i+1)}$$

$$\hat{K}(t) = \frac{\hat{\sigma}^2(t)}{1+\hat{\sigma}^2(t)} \quad ; \quad 0 \le t \le T_F$$

$$\overline{K}(t) = 1 - \hat{K}(t) = \left(1+\hat{\sigma}^2(t)\right)^{-1}$$

และเมื่อ $n \to \infty$; $n^{\frac{1}{2}}\left(\frac{\hat{S}_n(t)-S(t)}{S(t)}\right)\overline{K}(t) \to W^0(K(t))$; $0 \le t \le T$

โดยที่ $W^0(u)$ คือกระบวนการวีเนอร์ (Wiener process on (0,1))[4]

คุณสมบัติของ Wiener process

สมมติให้ $\{X(t)\}$ เป็น Wiener process

1. $\{X(t), t \ge 0\}$ มีส่วนเพิ่มที่เป็นอิสระและคงที่ (stationary independent increments)
2. สำหรับทุก ๆ $t > 0$, $X(t)$ มีการแจกแจงปรกติ
3. สำหรับทุกค่าของ $t > 0$, $E[X(t)] = 0$
4. $X(0) = 0$
5. $V[X(t)-X(s)] = \sigma^2|t-s|$

ให้ $\psi(u)$ เป็นฟังก์ชันที่เป็นบวกในช่วง (0,1) จะได้ตัวสถิติ

$$\sup_{0<t<T} \frac{n^{\frac{1}{2}}|\hat{S}(t)-S(t)|}{\hat{S}(t)}\overline{K}(t)\psi(\hat{K}(t)) \to \sup_{0\le t\le T}|W^0(K(t))|\psi(K(t))$$

1. ช่วงความเชื่อมั่นความเที่ยงเท่ากัน (Equal Precision (EP))

วิธีการประมาณนี้พัฒนาขึ้นโดยเนอร์ (Nair) โดยกำหนดให้

$$\psi(\hat{K}(t)) = \left(\hat{K}(t)(1-\hat{K}(t))\right)^{-\frac{1}{2}}$$

จะได้

$$\sup_{0<t<T} \frac{n^{\frac{1}{2}}|\hat{S}(t)-S(t)|}{\hat{S}(t)\hat{\sigma}(t)} \to \sup_{0\le t\le T}|W^0(K(t))|\left(\hat{K}(t)(1-\hat{K}(t))\right)^{-\frac{1}{2}}$$

ดังนั้นช่วงความเชื่อมั่นของวิธีประมาณความเที่ยงเท่ากัน (Equal Precision) คือ

$$\mathrm{Prob}\left(\hat{S}(t)-e_\alpha n^{-\frac{1}{2}}\hat{S}(t)\hat{\sigma}(t) < S(t) < \hat{S}(t)+e_\alpha n^{-\frac{1}{2}}\hat{S}(t)\hat{\sigma}(t)\right) = 1-\alpha$$

$$\hat{S}(t) \pm e_\alpha n^{-\frac{1}{2}}\hat{S}(t)\hat{\sigma}(t)$$

---

[4] D.R.Cox&H.D.Miller **The Theory of Stochastic processes** (London Chapman and Hall 1965) pp. 205-213

U.Narayam Bhat **Elements of Applied Stochastic Processes** (U.S.A. John Wiley&sons 1984) pp. 19,321

ค่าวิกฤต e_α หาได้จาก

$$\text{Prob}\left\{\sup_{0\le t\le T}\frac{|W^0(K(t))|}{\left(\hat{K}(t)(1-\hat{K}(t))\right)^{\frac{1}{2}}} \le e_\alpha\right\} = 1-\alpha$$

ซึ่งสามารถประมาณค่า e_α ได้โดยการกำหนดให้ $A(e_\alpha) = \frac{\alpha}{2}$

โดยที่ $$A(x) = \frac{x}{\sqrt{8\pi}}\exp(-\frac{x^2}{2})\ln\left(\frac{(1-a)b}{a(1-b)}\right)$$

กำหนดค่า $a = 0.05$ หรือ 0.01

และ $b = \min\left(0.95,\hat{K}(t_n^0)\right)$ หรือ $\min\left(0.99,\hat{K}(t_n^0)\right)$

t_n^0 เป็นค่าสังเกตที่ใหญ่ที่สุดของข้อมูลที่ไม่ถูกตัดทิ้ง

2. ช่วงความเชื่อมั่นฮอล-เวลเนอร์ (Hall-Wellner (HW))

วิธีการประมาณนี้พัฒนาขึ้นโดยฮอลและเวลเนอร์ (Hall and Wellner)

โดยกำหนดให้ $$\psi\left(\hat{K}(t)\right) = 1$$

จะได้

$$\sup_{0<t<T}\frac{n^{\frac{1}{2}}|\hat{S}(t)-S(t)|}{\hat{S}(t)}\overline{K}(t) \to \sup_{0\le t\le T}|W^0(K(t))|$$

ดังนั้นช่วงความเชื่อมั่นของวิธีประมาณฮอล-เวลเนอร์ (Hall-Wellner) คือ

$$\text{Prob}\left(\hat{S}(t)-h_\alpha n^{-\frac{1}{2}}\frac{\hat{S}(t)}{\overline{K}(t)} < S(t) < \hat{S}(t)+h_\alpha n^{-\frac{1}{2}}\frac{\hat{S}(t)}{\overline{K}(t)}\right) = 1-\alpha$$

$$\hat{S}(t) \pm h_\alpha n^{-\frac{1}{2}}\frac{\hat{S}(t)}{\overline{K}(t)}$$

ค่าวิกฤต h_α หาได้จาก

$$\text{Prob}\left\{\sup_{0\le t\le T}|W^0(K(t))| \le h_\alpha\right\} = 1-\alpha$$

ซึ่งสามารถประมาณค่า h_α ได้โดยการกำหนดให้ $B(h_\alpha) = \alpha$

โดยที่ $$B(x) = 2\sum_{k=1}^{\infty}(-1)^{k+1}\exp(-2k^2x^2)$$

$$\approx 2\exp(-2x^2)$$

3. ช่วงความเชื่อมั่นเรนยี (Renyi (R))

วิธีการประมาณนี้พัฒนาขึ้นโดย เรนยี (Renyi) โดยกำหนดให้

$$\psi\left(\hat{K}(t)\right) = \left(1-\hat{K}(t)\right)^{-1}$$

จะได้

$$\sup_{0<t<T} \frac{n^{\frac{1}{2}}\left|\hat{S}(t)-S(t)\right|}{\hat{S}(t)} \rightarrow \sup_{0\le t\le T}\left|W^{0}\left(K(t)\right)\right|\left(1-\hat{K}(t)\right)^{-1}$$

ดังนั้นช่วงความเชื่อมั่นของวิธีประมาณเรนยี (Renyi) คือ

$$\text{Prob}\left(\hat{S}(t)-r_{\alpha}n^{-\frac{1}{2}}\hat{S}(t) < S(t) < \hat{S}(t)+r_{\alpha}n^{-\frac{1}{2}}\hat{S}(t)\right) = 1-\alpha$$

$$\hat{S}(t) \pm r_{\alpha}n^{-\frac{1}{2}}\hat{S}(t)$$

ค่าวิกฤต r_{α} หาได้จาก

$$\text{Prob}\left\{\sup_{0\le t\le T}\frac{\left|W^{0}\left(K(t)\right)\right|}{1-\hat{K}(t)} \le r_{\alpha}\right\} = 1-\alpha$$

ซึ่งสามารถประมาณค่า r_{α} ได้โดยการกำหนดให้ $r_{\alpha} = \left(\frac{b}{1-b}\right)^{\frac{1}{2}}W_{\alpha}$

กำหนดค่า $b = \min\left(0.8,\hat{K}(t_{n}^{0})\right)$ หรือ $\min\left(0.9,\hat{K}(t_{n}^{0})\right)$

t_{n}^{0} เป็นค่าสังเกตที่ใหญ่ที่สุดของข้อมูลที่ไม่ถูกตัดทิ้ง

สำหรับค่า W_{α} สามารถหาได้จากการกำหนดให้ $B(W_{\alpha}) = 1-\alpha$

โดยที่ $B(x) = \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^{k}}{2k+1}\exp\left(\frac{-\pi^{2}}{8x^{2}}(2k+1)^{2}\right)$

$$\approx 4\Phi(x)-2$$

$\Phi(x)$ คือ ฟังก์ชันการแจกแจงปกติมาตรฐาน

ตัวอย่างต่อไปนี้จะแสดงถึงลักษณะของข้อมูลที่มีการแจกแจงแบบไวบูลล์โดยข้อมูลมีค่าสังเกตที่ถูกตัดทิ้งแบบสุ่ม จะแสดงวิธีการคำนวณวิธีการประมาณแบบช่วงสำหรับฟังก์ชันการอยู่รอดทั้ง 3 วิธี ข้อมูลที่นำเสนอมีทั้งสิ้น 25 ค่า โดยมีข้อมูลที่มีค่าที่ถูกตัดทิ้ง 5 ค่า และไม่ทราบค่าเนื่องจากมีค่ามากกว่าเวลาที่กำหนดอีก 1 ค่า ลักษณะของข้อมูลเป็นดังนี้

0.030,0.170,0.210,0.216,0.281,0.287,0.350*,0.366,0.458*,0.479,0.482,0.503,0.591,0.599,

0.621,0.734,0.796,0.806,0.884*,0.961,1.499,1.815,1.857*,1.887*,2.000*

นำค่าสังเกตมาเรียงลำดับค่าจากน้อยไปมากจะได้ดังรูปที่ 2.5

t_i	i	$(\frac{n-i}{n-i+1})^{\delta_i}$	$\frac{\delta_i}{(n-i)(n-i+1)}$
0.030	1	24/25	1/(24*25)
0.170	2	23/24	1/(23*24)
0.210	3	22/23	1/(22*23)
0.216	4	21/22	1/(21*22)
0.281	5	20/21	1/(20*21)
0.287	6	19/20	1/(19*20)
0.350*	7	1	0
0.366	8	17/18	1/(17*18)
0.458*	9	1	0
0.479	10	15/16	1/(15*16)
0.482	11	14/15	1/(14*15)
0.503	12	13/14	1/(13*14)
0.591	13	12/13	1/(12*13)
0.599	14	11/12	1/(11*12)
0.621	15	10/11	1/(10*11)
0.734	16	9/10	1/(9*10)
0.796	17	8/9	1/(8*9)
0.806	18	7/8	1/(7*8)
0.884*	19	1	0
0.961	20	5/6	1/(5*6)
1.499	21	4/5	1/(4*5)
1.815	22	3/4	1/(3*4)
1.857*	23	1	0
1.887*	24	1	0
2.000*	25	1	0

รูปที่ 2.5

* คือข้อมูลที่ถูกตัดทิ้ง

จากตัวประมาณพีแแอล (PL Estimator) จะได้

$$\hat{S}(t) = \prod_{(i:t_i \le t)} \left(\frac{n-i}{n-i+1}\right)^{\delta_i} \quad ; \quad \hat{\sigma}^2(t) = n \sum_{(i:t_i \le t)} \frac{\delta_i}{(n-i)(n-i+1)}$$

$$\hat{K}(t) = \frac{\hat{\sigma}^2(t)}{1+\hat{\sigma}^2(t)} \quad ; \quad \overline{K}(t) = 1-\hat{K}(t)$$

แสดงค่า $\hat{S}(t)$, $\hat{\sigma}^2(t)$, $\hat{K}(t)$, $\overline{K}(t)$ ได้ดังรูปที่ 2.6

t	$\hat{S}(t)$	$\hat{\sigma}^2(t)$	$\hat{K}(t)$	$\overline{K}(t)$
0.25	(24/25)*(23/24)*(22/23)*(21/22) = 0.84	25*(1/(24*25)+1/(23*24)+1/(22*23) +1/(21*22)) = 0.1905	0.16	0.84
0.50	0.84*(20/21)*(19/20)*1*(17/18)*1* (15/16)*(14/15) = 0.6281	0.1905+25*(1/(20*21)+1/(19*20)+0+ 1/(17*18)+0+1/(15*16)+1/(14*15)) = 0.6207	0.3830	0.617
0.75	0.6281*(13/14)*(12/13)*(11/12)* (10/11)*(9/10) = 0.4038	0.6207+25*(1/(13*14)+1/(12*13)+ 1/(11*12)+1/(10*11)+1/(9*10)) = 1.6128	0.6173	0.3827
1.00	0.4038*(8/9)*(7/8)*1*(5/6) = 0.2617	1.6128+25*(1/(8*9)+1/(7*8)+0+ 1/(5*6)) = 3.2398	0.7641	0.2359
1.25	0.2617	3.2398	0.7641	0.2359
1.50	0.2617*(4/5) = 0.2094	3.2398+25*(1/(4*5)) = 4.4898	0.8178	0.1822
1.75	0.2094	4.4898	0.8178	0.1822
2.00	0.2094*(3/4)*1*1*1 = 0.1570	4.4898+25*(1/(3*4)+0+0+0) = 6.5731	0.8680	0.1320

รูปที่ 2.6

นำค่า $\hat{S}(t)$, $\hat{\sigma}^2(t)$, $\hat{K}(t)$, $\overline{K}(t)$ มาคำนวณหาช่วงความเชื่อมั่นของฟังก์ชันการอยู่รอดโดยวิธีการประมาณทั้ง 3 วิธี ที่ค่าสัมประสิทธิ์ความเชื่อมั่นเป็น 95% ตามสมการที่กล่าวมาแล้วข้างต้น สำหรับข้อมูลนี้มาจากประชากรที่มีการแจกแจงแบบไวบูลล์
$\lambda = 1.00$, $\gamma = 1.50$ ดังนั้น $S(t) = \exp(-(\lambda t)^{\gamma})$ แสดงค่าต่าง ๆ ได้ดังรูปที่ 2.7

t	LEP	UEP	LHW	UHW	LR	UR	S(t)
0.25	0.6089	1.0000	0.5684	1.0000	0.0869	1.0000	0.8825
0.50	0.3162	0.9400	0.3516	0.9045	0.0650	1.0000	0.7022
0.75	0.0806	0.7270	0.1172	0.6903	0.0418	0.7657	0.5223
1.00.	0.0000	0.5586	0.0000	0.5631	0.0271	0.4963	0.3679
1.25	0.0000	0.5586	0.0000	0.5631	0.0271	0.4963	0.2472
1.50	0.0000	0.4890	0.0000	0.5215	0.0217	0.3971	0.1593
1.75	0.0000	0.4890	0.0000	0.5215	0.0217	0.3971	0.0988
2.0	0.0000	0.4108	0.0000	0.4800	0.0162	0.2978	0.0591

รูปที่ 2.7

โดยที่ LEP คือ ขีดจำกัดล่างของวิธีการประมาณ EP

UEP คือ ขีดจำกัดบนของวิธีการประมาณ EP

LHW คือ ขีดจำกัดล่างของวิธีการประมาณ HW

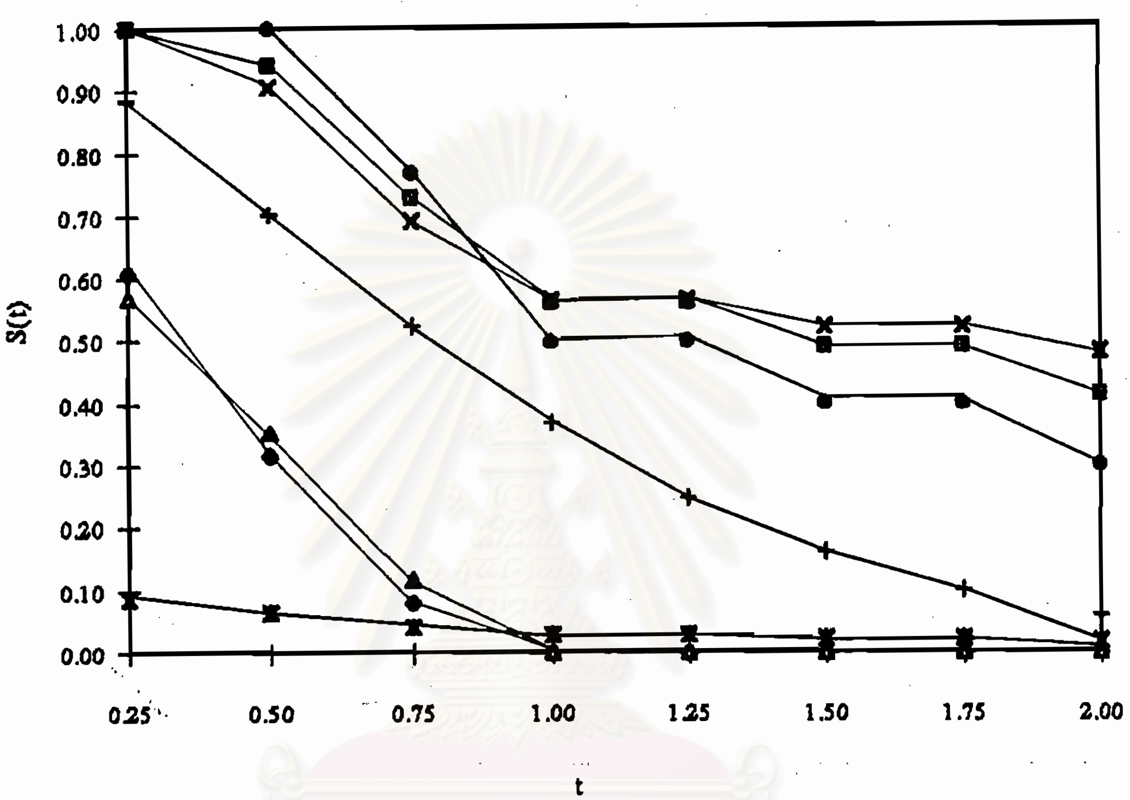
UHW คือ ขีดจำกัดบนของวิธีการประมาณ HW

LR คือ ขีดจำกัดล่างของวิธีการประมาณ R

UR คือ ขีดจำกัดบนของวิธีการประมาณ R

S(t) คือ ฟังก์ชันการอยู่รอดของประชากรที่มีการแจกแจงแบบไวบูลล์ที่พารามิเตอร์ $\lambda = 1.00$, $\gamma = 1.50$

นำค่าของช่วงความเชื่อมั่นความเที่ยงเท่ากัน (LBP,UBP) ,ช่วงความเชื่อมั่นฮอล-เวลเนอร์ (LHW,UHW) ,ช่วงความเชื่อมั่นเรนยี (LR,UR) และค่าฟังก์ชันการอยู่รอดที่แท้จริง (S(t)) มาเขียนกราฟได้ดังรูปที่ 2.8



รูปที่ 2.8

———— คือกราฟแสดงช่วงความเชื่อมั่นความเที่ยงเท่ากัน

———— คือกราฟแสดงช่วงความเชื่อมั่นฮอล-เวลเนอร์

———— คือกราฟแสดงช่วงความเชื่อมั่นเรนยี

———— คือกราฟแสดงค่าฟังก์ชันการอยู่รอดที่แท้จริง

จากกราฟจะพบว่าช่วงความเชื่อมั่นความเที่ยงเท่ากันจะมีลักษณะกว้างในช่วงกลาง และช่วงต้นกับท้ายจะแคบคือความยาวของช่วงจะสั้นเมื่อเวลามีค่าน้อยหรือเมื่อเวลามีค่ามากและความยาวของช่วงจะกว้างเมื่อเวลามีค่าปานกลาง ส่วนช่วงความเชื่อมั่นของฮอล-เวลเนอร์จะมีลักษณะแคบในช่วงต้นและค่อย ๆ กว้างออกเมื่อเวลามีค่าเพิ่มขึ้น และช่วงความเชื่อมั่นเรนยีจะมีลักษณะกว้างในช่วงต้นและค่อย ๆ แคบเข้าเมื่อเวลามีค่าเพิ่มขึ้น

เกณฑ์ที่ใช้ในการเปรียบเทียบวิธีการประมาณค่าแบบช่วง

การเปรียบเทียบวิธีการประมาณค่าแบบช่วงสำหรับฟังก์ชันการอยู่รอด เมื่อข้อมูลมีค่าที่ถูกตัดทิ้งทั้ง 3 วิธีนี้ จะทำการเปรียบเทียบค่าระดับความเชื่อมั่นและค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่คำนวณได้จากแต่ละวิธีที่สัมประสิทธิ์ความเชื่อมั่น 3 ระดับคือ 90% ,95% และ 99% ในแต่ละสถานการณ์ของการทดลอง ทำการทดลองซ้ำ 2,000 ครั้ง การเปรียบเทียบค่าระดับความเชื่อมั่นของช่วงความเชื่อมั่นที่คำนวณได้จากแต่ละวิธีการประมาณ กับค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด ในการพิจารณาว่าวิธีการใดจะให้ค่าระดับความเชื่อมั่นจากการทดลองไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด ผู้วิจัยอาศัยการทดสอบสมมติฐานโดยใช้ตัวสถิติ Z ดังนี้

$$-Z_{1-\frac{\alpha}{2}} < \frac{\hat{p}-p}{\left(\frac{p(1-p)}{n}\right)^{\frac{1}{2}}} < Z_{1-\frac{\alpha}{2}}$$

$$p-Z_{1-\frac{\alpha}{2}}\left(\frac{p(1-p)}{n}\right)^{\frac{1}{2}} < \hat{p} < p+Z_{1-\frac{\alpha}{2}}\left(\frac{p(1-p)}{n}\right)^{\frac{1}{2}}$$

จะได้ $$\left(p-Z_{1-\frac{\alpha}{2}}\left(\frac{p(1-p)}{n}\right)^{\frac{1}{2}}, p+Z_{1-\frac{\alpha}{2}}\left(\frac{p(1-p)}{n}\right)^{\frac{1}{2}}\right)$$

นั่นคือที่ค่าสัมประสิทธิ์ความเชื่อมั่น 90% ,95% และ 99% ถ้าวิธีการประมาณใดให้ค่าระดับความเชื่อมั่นจากการทดลองไม่ต่ำกว่า 0.8890 ,0.9405 และ 0.9843 ตามลำดับจะถือว่าวิธีการประมาณนั้น ให้ค่าระดับความเชื่อมั่นจากการทดลองไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดในสถานการณ์นั้น ๆ จากนั้นจึงมาพิจารณาเปรียบเทียบค่าความยาวเฉลี่ยของช่วงความเชื่อมั่น ถ้าวิธีการประมาณใดให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นต่ำที่สุดจะถือว่าเป็นวิธีการประมาณที่เหมาะสมที่สุด ทั้งนี้ในการเปรียบเทียบค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นจะเปรียบเทียบเฉพาะ ในกรณีที่วิธีการประมาณให้ค่าระดับความเชื่อมั่นจากการทดลองไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดเท่านั้น