บทที่ 2

ระเบียบวิธีที่ใช้ในการวิจัย

การศึกษาความฉลาดเชิงเส้นพหุเมื่อตัวแปรตามบางค่าถูกตัดทิ้งทางขวาประเภทที่ 1 โดยในการศึกษาครั้งนี้ได้ทำการศึกษาเปรียบเทียบด้วยวิธีการประมาณเพารามิเตอร์ 4 วิธี คือ วิธีกำลังสองต่ำสุด วิธีการของแฮตเทอร์จีและแมคลีช วิธีการของบัคเลย์และเจมส์ และวิธีภาวะน่าจะเป็นสูงสุดด้วยขั้นตอนวิธีอีเอ็ม ซึ่งในบทนี้กล่าวถึงความรู้ ทฤษฎีพื้นฐาน และรายละเอียดของวิธีการประมาณเพารามิเตอร์แต่ละวิธี รายละเอียดต่างๆ เป็นดังนี้

2.1 ทฤษฎีพื้นฐาน

2.1.1 ประเภทของการถูกตัด (Type of Censoring)

ก. การถูกตัดประเภทที่ 1 (Type I Censoring)

ข้อมูลที่ถูกตัดประเภทที่ 1 เกิดขึ้นเนื่องจากมีการกำหนดค่าสูงสุดของข้อมูลล่วงหน้าด้วยค่าคงที่ Tc ซึ่งจะเรียกว่า Fixed Censoring Time ตัวอย่างเช่นการทดลองยากับหนูทดลองที่กำหนดระยะเวลาของการทดลองไว้เท่ากับ 3 เดือนแล้วเริ่มทำการทดลอง โดยการฉีดยาที่ต้องการทดลองให้กับหนู แล้วเริ่มบันทึกเวลาตั้งแต่เริ่มทำการทดลองจนกระทั่งหนูทดลองตาย ถ้าหนูตัวใดตายภายใน 3 เดือนจะถือว่าเป็นข้อมูลที่ไม่ถูกตัด และจำนวนวันตั้งแต่เริ่มทดลองจนหนูตายเรียกว่าเป็น Survival Time แต่เมื่อครบระยะเวลาทดลอง 3 เดือนแล้วหนูทดลองตัวใดที่ไม่ตาย จะถือว่าค่าสังเกตของหนูตัวนั้นถูกตัดที่ Fixed Censoring Time เท่ากับ 3 เดือน สำหรับข้อมูลทางด้านประกันภัยที่กรมธรรม์จะกำหนดความเสียหายสูงสุดที่บริษัทรับผิดชอบไว้ เช่นการรับประกันสุขภาพที่บริษัทจะจ่ายค่ารักษาพยาบาลให้ตามที่ผู้เอาประกันภัยจ่ายไปจริงแต่ไม่เกิน 10,000 บาท ถ้าผู้เอาประกันภัยรายใดต้องเสียค่ารักษาจริงมากกว่า 10,000 บาท บริษัทจะจ่ายชดเชยให้เพียง 10,000 บาทเท่านั้น และถือว่าข้อมูลนี้ถูกตัดที่ Tc เท่ากับ 10,000 บาท

ถ้า $T_1, T_2, ..., T_N$ เป็นค่าสังเกตุที่ไม่ถูกตัดที่มีการแจกแจงเหมือนกันและเป็นอิสระกัน Tc เป็นเวลาหรือค่าสูงสุดที่กำหนดไว้ล่วงหน้า จะได้ตัวแปรสุ่มของค่าสังเกต $Y_1, Y_2, ..., Y_N$ ซึ่ง

$$Y_i = \begin{cases} T_i & ; \ T_i \le Tc \\ Tc & ; \ T_i > Tc \end{cases}$$

โดยมีฟังก์ชันภาวะน่าจะเป็น (Likelihood Function) ดังนี้

$$L(y_i) = \begin{cases} f(y_i) & \text{ถ้าค่าสังเกตไม่ถูกตัด} \\ P(T_i > Tc) = S(Tc) & \text{ถ้าค่าสังเกตถูกตัด} \end{cases}$$

และมีฟังก์ชันภาวะน่าจะเป็นรวมดังนี้

$$L = \prod_{i \in u} f(y_i) \prod_{i \in c} S(Tc)$$

$i \in u$ หมายถึงค่าสังเกตที่ไม่ถูกตัดทิ้ง

$i \in c$ หมายถึงค่าสังเกตที่ถูกตัดทิ้ง

ข. การลดตัดประเภทที่ 2 (Type II Censoring)

ในบางกรณีที่ไม่อาจกำหนดเวลา หรือค่าสูงสุดของการตัดข้อมูลที่เหมาะสมได้ ดังนั้นจะกำหนดจำนวนค่าสังเกตที่ไม่ถูกตัดแทน นั่นคือเมื่อค่าสังเกตที่ไม่ถูกตัดเกิดขึ้นครบตามจำนวนที่กำหนดแล้วก็จะหยุดการทดลอง เช่นการทดสอบอายุการใช้งานของหลอดไฟ จะกำหนดจำนวนหลอดไปที่เสื่อมสภาพไว้ล่วงหน้า เริ่มทดลองโดยเปิดให้หลอดไฟทำงานทั้งหมด เริ่มบันทึกเวลาและนับจำนวนหลอดไฟที่เสื่อมสภาพ เมื่อได้จำนวนหลอดที่เสื่อมสภาพครบแล้วก็จะหยุดทำการทดสอบ

ถ้า N คือจำนวนข้อมูลทั้งหมดและกำหนด n คือจำนวนค่าสังเกตที่ไม่ถูกตัด $n \le N$ ให้ $T_1 \le T_2 \le, ..., T_n$ เป็นค่าสังเกตที่ไม่ถูกตัด และ $T_{n+1} \le T_{n+2} \le .... \le T_N$ เป็นค่าสังเกตที่ถูกตัด ซึ่ง $T_i \ge T_n \quad ; i = n+1, n+2, ..., N$ ไม่ทราบค่าที่แท้จริงของค่าสังเกต ดังนั้น Y_i เป็นตัวแปรสุ่มของค่าสังเกตซึ่ง

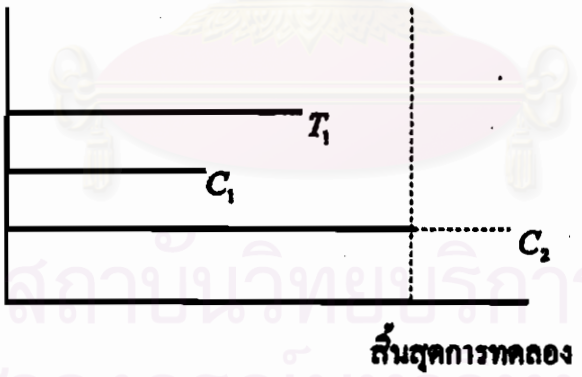
$$Y_i = \begin{cases} T_i & \text{เมื่อ } i \le n \\ T_n & \text{เมื่อ } i = n+1, n+2, \ldots, N \end{cases}$$

ฟังก์ชันความหนาแน่นร่วมของค่าสังเกตคือ

$$\frac{N!}{(N-n)!} f(y_1) f(y_2) \ldots f(y_n) [S(y_n)]^{N-n}$$

ค. การตัดแบบสุ่ม (Random Censoring)

ข้อมูลลักษณะนี้เกิดขึ้นคล้ายกับการตัดประเภทที่ 1 คือมีการกำหนดค่าสูงสุดหรือกำหนดเวลาไว้ล่วงหน้า แต่การตัดข้อมูลอาจจะเกิดขึ้นก่อนนั้นได้ เช่นการทดลองทางการแพทย์ที่คนไข้ถอนตัวออกจากการทดลองก่อนสิ้นสุดการทดลอง หรือคนไข้ยังมีชีวิตอยู่รอดเมื่อสิ้นสุดการทดลอง หรือคนไข้เสียชีวิตเนื่องจากสาเหตุอื่นที่ไม่เกี่ยวกับสิ่งที่กำลังศึกษาทดลอง เป็นต้น จึงไม่สามารถทราบค่าที่แน่นอนของค่าสังเกตนั้นได้ รูปที่ 2.1 แสดงลักษณะความเป็นไปได้ของข้อมูลถูกตัดแบบสุ่ม



รูปที่ 2.1 แผนภาพแสดงการเกิดค่าถูกตัดแบบสุ่ม

คนไข้คนที่ 1 เข้าทำการทดลองตั้งแต่เริ่มการทดลอง และเสียชีวิตที่เวลา T_1 จะถือว่าเป็นค่าสังเกตที่ไม่ถูกตัดทิ้ง

คนไข้คนที่ 2 เข้าทำการทดลองตั้งแต่เริ่มการทดลอง และถอนตัวจากการทดลองที่เวลา C_1 จะถือว่าเป็นค่าสังเกตที่ถูกตัดทิ้ง

คนไข้คนที่ 3 เข้าทำการทดลองตั้งแต่เริ่มการทดลอง และมีชีวิตอยู่รอดจนสิ้นสุดการทดลอง จะถือว่าเป็นค่าสังเกตที่ถูกตัดทิ้ง C_2

ถ้า $T_1, T_2, \ldots, T_N$ เป็นตัวแปรสุ่มของค่าสังเกตที่ไม่ถูกตัดที่มีการแจกแจงเหมือนกันและเป็นอิสระกันมีฟังก์ชันการอยู่รอดและฟังก์ชันความหนาแน่น S และ f ตามลำดับ และ $C_1, C_2, \ldots, C_N$ เป็นตัวแปรสุ่มของค่าสังเกตที่ถูกตัดทิ้งมีการแจกแจงเหมือนกันและเป็นอิสระกันมีฟังก์ชันการอยู่รอดและฟังก์ชันความหนาแน่น G และ g ตามลำดับ

ดังนั้น T_i และ C_i , i=1,2,...,N เป็นอิสระกัน จากนิยามการตัดทิ้งแบบสุ่ม นิยามให้ $Y_i = \min(T_i, C_i)$ จะได้ค่าสังเกตสุ่ม $Y_1, Y_2, \ldots, Y_N$ ดังนี้

$$Y_i = \begin{cases} T_i & \text{ถ้า} \quad T_i \le C_i \quad (\text{ไม่ถูกตัด}) \\ C_i & \text{ถ้า} \quad T_i > C_i \quad (\text{ถูกตัด}) \end{cases}$$

$$\delta_i = \begin{cases} 1 & \text{ถ้า} \quad T_i \le C_i \quad (\text{ไม่ถูกตัด}) \\ 0 & \text{ถ้า} \quad T_i > C_i \quad (\text{ถูกตัด}) \end{cases}$$

จะมีฟังก์ชันภาวะน่าจะเป็นดังนี้

$$L(y_i, \delta_i) = \begin{cases} f(y_i)G(y_i) & \text{ถ้า} \quad \delta_i = 1 \\ g(y_i)S(y_i) & \text{ถ้า} \quad \delta_i = 0 \end{cases}$$

และมีฟังก์ชันภาวะน่าจะเป็นรวมดังนี้

$$L = \left[\prod_{i \in u} f(y_i) \prod_{i \in u} G(y_i)\right]\left[\prod_{i \in c} g(y_i) \prod_{i \in c} S(y_i)\right]$$

เนื่องจาก $G(y_i)$ และ $g(y_i)$ ไม่เกี่ยวข้องกับพารามิเตอร์ที่สนใจ จึงละไว้โดยใช้ฟังก์ชันภาวะน่าจะเป็นดังนี้

$$L = \left[\prod_{i \in u} f(y_i) \prod_{i \in c} S(y_i)\right]$$

$i \in u$ หมายถึงค่าสังเกตที่ไม่ถูกตัดทิ้ง

$i \in c$ หมายถึงค่าสังเกตที่ถูกตัดทิ้ง

๔. การตัดแบบอื่นๆ (Other Type of Censoring)

นอกจากการตัดระเภทที่ 1 การตัดประเภทที่ 2 และการตัดแบบสุ่ม ซึ่งที่กล่าวมาเป็นการตัดทางขวา ยังมีการตัดทางซ้าย (Left Censoring) และการตัดทางซ้ายและทางขวา (Left and Right Censoring) ข้อมูลที่ถูกตัดทางซ้าย เช่นกรมธรรม์ประกันภัยรถยนต์กำหนดความเสียหายส่วนแรก (Deductible) ที่ผู้เอาประกันภัยจะต้องรับผิดชอบเอง ดังนั้นบริษัทประกันภัยจึงมีเฉพาะข้อมูลการจ่ายความสูญเสียที่มากกว่าความเสียหายส่วนแรกที่กำหนดไว้เท่านั้น และถ้าบริษัทรับประกันภัยนี้ทำประกันภัยต่อกับบริษัทรับประกันภัยต่อด้วย ซึ่งบริษัทรับประกันภัยจะกำหนดจำนวนความรับผิดชอบสูงสุดของบริษัทเอาไว้ ความสูญเสียส่วนที่เกินจากที่กำหนดนี้บริษัทรับประกันภัยต่อจะเป็นผู้รับผิดชอบ ดังนั้นข้อมูลที่บริษัทรับประกันภัยมีอยู่จึงเป็นความสูญเสียที่มากกว่าความเสียหายส่วนแรกและไม่เกินจำนวนเงินสูงสุดที่รับผิดชอบ

2.1.2 <u>ฟังก์ชันการอยู่รอดและฟังก์ชันการสูญเสีย</u> (Survival Function and Hazard Function)

ให้ T เป็นตัวแปรสุ่มต่อเนื่อง

$f(t)$ เป็นฟังก์ชันความหนาแน่นของ T (Probability density function)

$F(t)$ เป็นฟังก์ชันการแจกแจงสะสมของ T (Distribution Function)

$S(t)$ เป็นฟังก์ชันการอยู่รอดของ T (Survival Function)

$h(t)$ เป็นฟังก์ชันความสูญเสียหรืออัตราการสูญเสีย (Hazard Function or Hazard Rate)

นิยามฟังก์ชัน $S(t)$ คือความน่าจะเป็นที่ตัวแปรสุ่ม T จะมีค่ามากกว่าหรือเท่ากับ t

$$\begin{aligned} S(t) &= P(T > t) \\ &= \int_t f(t)dt \end{aligned}$$

นิยามฟังก์ชัน $h(t)$ แทนฟังก์ชันการสูญเสียมีค่าเท่ากับลิมิตของความน่าจะเป็นที่ตัวแปรสุ่ม T จะมีค่าอยู่ในช่วงเวลาสั้นๆ $(t, t+\Delta t)$ ต่อหน่วยเวลา Δt เมื่อกำหนดค่า $T > t$ และ $h(t)$ กำหนดได้ดังนี้

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{P(t \le T < t+\Delta t / T \ge t)}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{[F(t+\Delta t) - F(t)]}{(1-F(t))\Delta t} \end{aligned}$$

$$= \frac{d}{dt}F(t)\cdot\frac{1}{1-F(t)}$$

$$= \frac{f(t)}{1-F(t)} = \frac{f(t)}{S(t)}, \quad t>0$$

กรณีเมื่อ T เป็นตัวแปรสุ่มที่ไม่ต่อเนื่อง และมีค่าเป็น $t_1, t_2, t_3, \ldots$ โดยที่ $0 \le t_1 < t_2 < t_3 < \ldots$ ดังนั้นฟังก์ชันความน่าจะเป็น $p(t_j)$ (Probability Function) กำหนดได้ดังนี้

$$p(t_j) = P(T = t_j) \qquad j = 1,2,3,\ldots$$

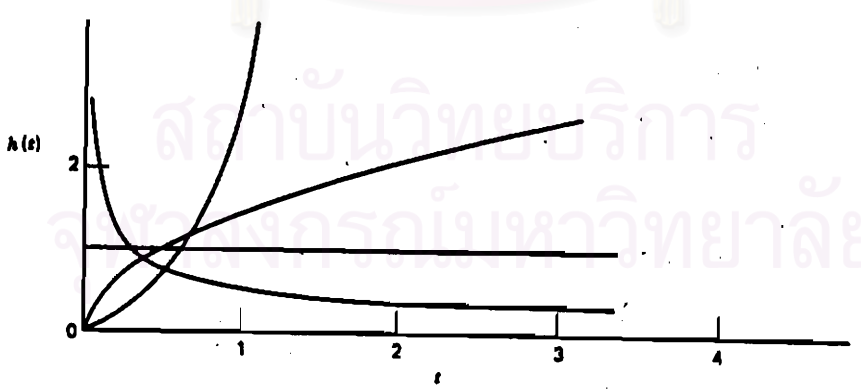
และฟังก์ชันการอยู่รอด $S(t)$ คือ

$$S(t) = P(T \ge t_j) = \sum_{j:j\ge t} p(t_j).$$

ดังนั้นฟังก์ชันความสูญเสีย $h(t)$ แสดงได้ดังนี้

$$h(t) = P(T = t_j / T \ge t_j)$$

$$= \frac{P(t_j)}{S(t_j)}$$

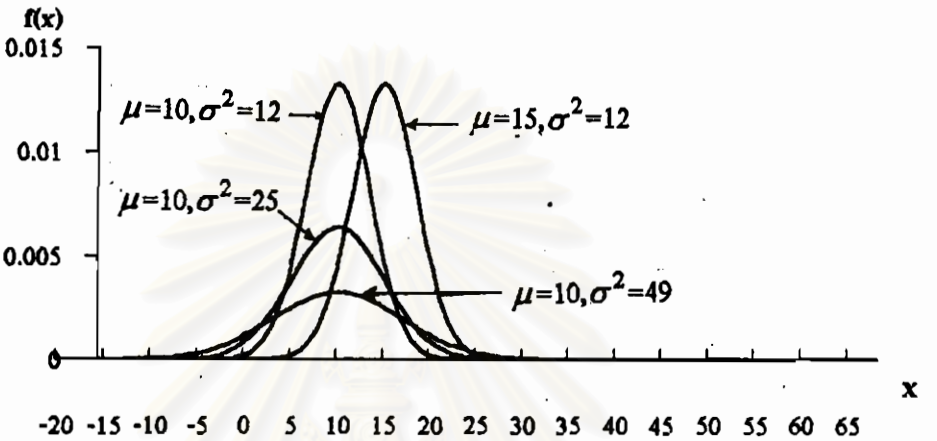


รูปที่ 2.2 แสดงตัวอย่างลักษณะต่างๆ ของฟังก์ชันการสูญเสีย $h(t)$ ในรูปแบบต่าง ๆ

2.1.3 <u>ฟังก์ชันความหนาแน่น ค่าคาดหวัง และความแปรปรวน</u> ของการแจกแจงภายใต้การศึกษาครั้งนี้

2.1.3.1 การแจกแจงปกติ (Normal Distribution)

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-(x-\mu)^2/2\sigma^2\right] \quad ; \ -\infty < x < \infty$$



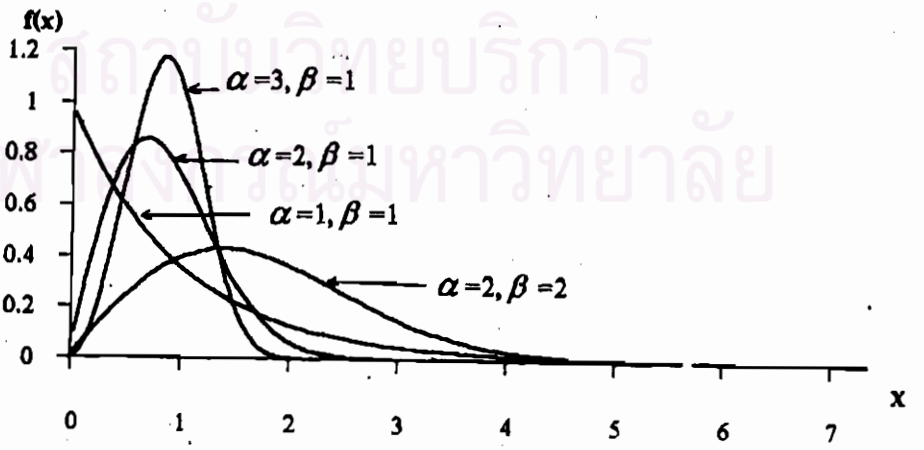
$E(x) = \mu$

$V(x) = \sigma^2$

รูปที่ 2.3 แสดงการแจกแจงปกติ

2.1.3.2 การแจกแจงแบบไวบูลล์ (Weibull Distribution)

$$f(x) = \alpha\beta^{-\alpha}x^{\alpha-1}\exp\left(-(x/\beta)^{\alpha}\right) \quad ; \ x > 0$$



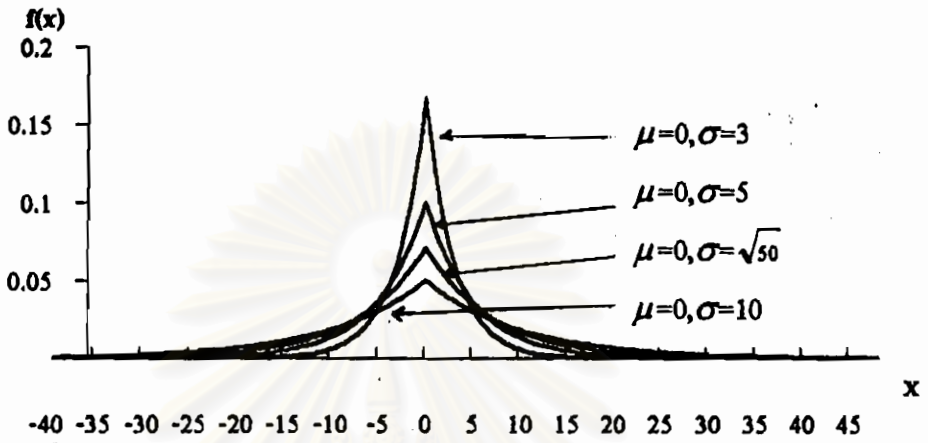
$$E(x) = \frac{\beta}{\alpha}\Gamma(1/\alpha)$$

$$V(x) = \frac{\beta^2}{\alpha}\left\{2\Gamma\left(\frac{2}{\alpha}\right) - \frac{1}{\alpha}\left(\Gamma\left(\frac{1}{\alpha}\right)\right)^2\right\}$$

รูปที่ 2.4 แสดงการแจกแจงไวบูลล์

2.1.3.3 การแจกแจงแบบดับเบิลเอกซโพเนนเชียล (Double Exponential Distribution)

$$f(x) = \frac{1}{2\sigma}\exp\left(-|x-\mu|/\sigma\right) \quad ; \quad -\infty < x < \infty$$



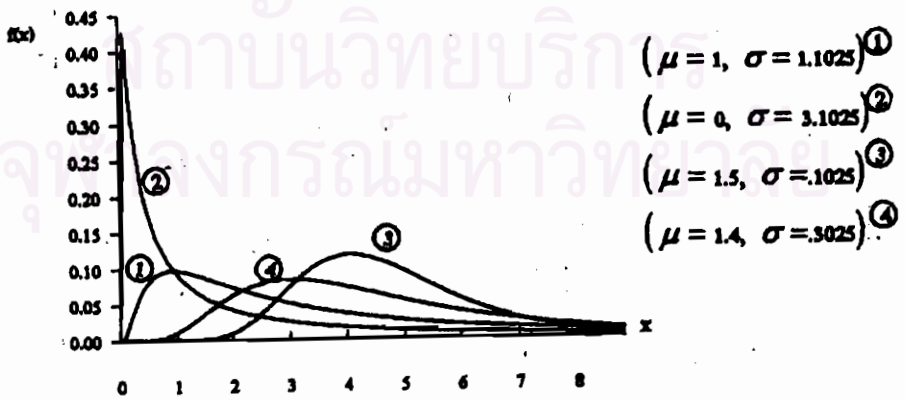
$$E(x) = \mu$$

$$V(x) = 2\sigma^2$$

รูปที่ 2.5 แสดงการแจกแจงแบบดับเบิลเอกซโพเนนเชียล

2.1.3.4 การแจกแจงแบบล็อกนอร์มอล (Lognormal Distribution)

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}}\exp\left[-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right] \quad ; x > 0, -\infty < \mu < \infty, \sigma > 0$$



$$E(x) = e^{\mu+\sigma^2/2}$$

$$V(x) = e^{2\mu+\sigma^2/2}\left(e^{\sigma^2}-1\right)$$

รูปที่ 2.6 แสดงการแจกแจงแบบล็อกนอร์มอล

2.1.4 <u>ตัวประมาณพีแอล</u> (Product Limit Estimator : PL Estimator)[1]

ตัวประมาณพีแอลพัฒนาขึ้นโดย แค็พแลน และไมเออร์ เป็นวิธีการประมาณฟังก์ชันการอยู่รอดที่เป็นนอนพาราเมตริกซ์ (Nonparametric) เมื่อข้อมูลบางส่วนถูกตัดทิ้ง แนวคิดของฟังก์ชันการอยู่รอด t_k ปี $S(t_k)$ คือความน่าจะเป็นที่ระยะเวลาของอายุที่รอด (Survival Time) มากกว่า t_k ปี ดังนั้นจะได้

$$\begin{aligned} S(t_k) &= P(T > t_k) \\ &= P(T > t_1)P(T > t_2 / T > t_1)....P(T > t_k / Tt_{k-1}) \\ &= p_1 p_2 p_3 .... p_k \end{aligned}$$

เมื่อ p_i คือความน่าจะเป็นที่จะมีชีวิตรอด t_i ปี หลังจากมีชีวิตอยู่รอดมาแล้ว t_{i-1} ปี

$$p_i = P(T > t_i / T > t_{i-1})$$

ฟังก์ชันการอยู่รอดมีค่าที่ $S(t_0) = 1$ และ $S(t_N) = 0$ และเป็นฟังก์ชันขั้นบันได

สมมติค่าสังเกตของอายุที่รอด (Survival Time) จำนวน N คนมีค่าเป็น $y_1 y_2 y_3 ... y_N$ นำอายุที่รอดมาเรียงตามลำดับเป็น $y(1)<y(2)<.....y(N)$
ดังนั้นหาค่าประมาณ $\hat{p}_i$ โดย

ให้ $$\hat{q}_i = \frac{1}{n_i}$$

---

[1] Kaplan E.L. and Meier P., "Nonparametric Estimation from Incomplete Observations" <u>Journal of the American Statistical Association</u> 53 (June 1958) : 475-81.

Rupert G. Miller, <u>Survival Analysis</u>. (New York : John Wiley and Sons Inc., 1981), PP.46-50

$$\hat{p}_i = 1-\hat{q}_i = \begin{cases} 1-\frac{1}{n_i} & \text{เมื่อ } \delta_{(i)} = 1 \\ 1 & \text{เมื่อ } \delta_{(i)} = 0 \end{cases}$$

$$\delta_i = \begin{cases} 1 & \text{เมื่อข้อมูลลำดับที่ } i \text{ ไม่ถูกตัดทิ้ง} \\ 0 & \text{เมื่อข้อมูลลำดับที่ } i \text{ ถูกตัดทิ้ง} \end{cases}$$

ตัวประมาณพีแอลสามารถแสดงได้ดังนี้

$$\hat{S}(t) = \prod_{y(i)\le t}\left(\frac{N-i}{N-i+1}\right)^{\delta_{(i)}}$$

และตัวประมาณพีแอลเมื่อมีจำนวนค่าสังเกตที่ถูกตัด ณ $y(i)$ เท่ากับ d_i ค่า สามารถแสดงได้ดังนี้

$$\hat{S}(t) = \prod_{y(i)\le t}\left(1-\frac{d_i}{n_i}\right)^{\delta_{(i)}}$$

เมื่อ $y(i)$ เป็นค่าสังเกตที่ไม่ถูกตัดทิ้ง ลำดับที่ i

i เป็นลำดับที่ของข้อมูล

N เป็นจำนวนข้อมูลทั้งหมดที่ไม่ถูกตัด และที่ถูกตัด

n_i คือจำนวนคนที่มีชีวิตรอด ณ เวลา $y(i)$.

d_i คือจำนวนคนที่เสียชีวิตรอด ณ เวลา $y(i)$

ตัวอย่างการหาตัวประมาณพีแอล (PL Estimator) จากค่าสังเกตต่อไปนี้ 9 , 12 , 12^+, 18, 23 , 28^+, 30, 34 , 45 , 47 , 160^+ เมื่อ + หมายถึงข้อมูลถูกตัดทิ้ง
จะได้ PL Estimator ดังนี้

$$\hat{S}(0) = 1$$

$$\hat{S}(9) = \hat{S}(0) \times \frac{10}{11} = .91$$

$$\hat{S}(12) = \hat{S}(9) \times \frac{9}{10} = .82$$

$$\hat{S}(18) = \hat{S}(12) \times \frac{7}{8} = .72$$

$$\hat{S}(23) = \hat{S}(18) \times \frac{6}{7} = .61$$

$$\hat{S}(30) = \hat{S}(23) \times \frac{4}{5} = .49$$

$$\hat{S}(34) = \hat{S}(30) \times \frac{3}{4} = .37$$

$$\hat{S}(47) = \hat{S}(34) \times \frac{1}{2} = .18$$

2.1.3 อีเอ็ม อัลกอริทึม (EM Algorithm)[2]

เนื่องจากสถิติที่เพียงพอ (Sufficient Statistics) เป็นสถิติที่เพียงพอสำหรับพารามิเตอร์ของข้อมูลที่สมบูรณ์เท่านั้น (Complete Data) สำหรับข้อมูลที่ถูกตัดบางส่วนถือว่าเป็นข้อมูลที่ไม่สมบูรณ์ (Incomplete Data) อีเอ็ม อัลกอริทึม เป็นการประมาณค่าที่ประกอบไปด้วย 2 ขั้นตอนคือ คือขั้นการประมาณค่าสังเกตที่ถูกตัดด้วยค่าคาดหวังอย่างมีเงื่อนไขซึ่งเรียกขั้นนี้ว่า ขั้นการหาค่าคาดหวัง (Expectation Step : E Step) ซึ่งเมื่อรวมค่าประมาณของค่าสังเกตที่ถูกตัดและค่าสังเกตที่ไม่ถูกตัดแล้วจะทำให้ได้ข้อมูลที่สมบูรณ์ขึ้น จากขั้นนี้จะทำให้ได้สถิติที่เพียงพอสำหรับพารามิเตอร์ ขั้นต่อมาจะนำข้อมูลทั้งหมดประมาณค่าสูงสุดของพารามิเตอร์ ซึ่งเรียกขั้นนี้ว่า ขั้นการประมาณค่าสูงสุด (Maximization Step : M Step) ซึ่งในการศึกษาครั้งนี้วิธีภาวะน่าจะเป็นสูงสุดใช้ขั้นตอนวิธีอีเอ็มในการประมาณพารามิเตอร์

2.2 การประมาณค่าพารามิเตอร์

2.2.1 วิธีกำลังสองต่ำสุด

การหาค่าประมาณของพารามิเตอร์วิธีนี้ มีรากฐานมาจากทฤษฎีการประมาณเชิงเส้น (Theory of Linear Estimation) ซึ่งคิดขึ้นโดย คาร์ล เฟรดริก เกาส์

---

[2] A.P. Dempster, N.M. Laird and D.B. Rubin , "Maximum Likelihood from Incomplete Data via the EM Algorithm " Journal of the Royal Statistical Society, B,39 (1977) : 1-38.

และอังเดร แอนดรีวิช มาร์คอฟ [3] โดยมีหลักเกณฑ์ว่าค่าประมาณพารามิเตอร์ที่ทำให้ผลบวกกำลังสองของผลต่างระหว่างค่าสังเกตกับค่าประมาณมีค่าต่ำสุด จะเป็นตัวประมาณเชิงเส้นที่ดีที่สุดและไม่เอนเอียง (Best Linear Unbiased Estimation , BLUE) เมื่อข้อมูลเป็นไปตามข้อตกลงเบื้องต้นของการวิเคราะห์ความถดถอย คือ

1. ค่าความคลาดเคลื่อนต้องมีการแจกแจงปกติ ที่มีค่าเฉลี่ย 0 และความแปรปรวนเป็น σ^2

2. ค่าความคลาดเคลื่อนจะต้องเป็นอิสระต่อกัน คือ ε_i และ ε_j ไม่มีความสัมพันธ์ต่อกันเมื่อ $i \neq j$, $i = 1,2,3,...,N$, $j = 1,2,3,...,N$ เมื่อ N คือขนาดตัวอย่าง

3. ค่าความคลาดเคลื่อน ε_i จะต้องเป็นอิสระกับตัวแปรอิสระ X หรือ $Cov(\varepsilon_i, X_i)$ เท่ากับ 0 เมื่อ $i = 1,2,3,...,N$ เมื่อ N คือขนาดตัวอย่าง

แต่เนื่องจากข้อมูลที่ถูกตัดเสมือนข้อมูลที่ไม่ได้ถูกตัดทิ้ง ดังนั้น วิธีกำลังสองต่ำสุดจะทำให้ได้ตัวประมาณที่เอนเอียง โดยเฉลี่ยจะน้อยกว่าความเป็นจริง

หลักเกณฑ์ในการหาตัวประมาณกำลังสองต่ำสุด จากความสัมพันธ์ระหว่างตัวแปรตาม Y และตัวแปรอิสระ X คือ

$$\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\beta} + \varepsilon_i$$

ให้ $\underset{\sim}{\hat{\beta}} = (\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k)$ เป็นเวกเตอร์ของตัวประมาณพารามิเตอร์จะได้ความสัมพันธ์ดาดหมายคือ

$$\underset{\sim}{\hat{Y}} = \underset{\sim}{X}\underset{\sim}{\hat{\beta}}$$

เมื่อ $\underset{\sim}{\varepsilon}$ คือความคลาดเคลื่อนระหว่างค่าสังเกตของ $\underset{\sim}{Y}$ กับค่าประมาณ $\underset{\sim}{\hat{Y}}$ ดังนั้น

$$\underset{\sim}{\varepsilon} = \underset{\sim}{Y} - \underset{\sim}{X}\underset{\sim}{\hat{\beta}}$$

พิจารณาผลบวกกำลังสองของความคลาดเคลื่อน (Sum of Squared Error)

$$\begin{aligned}
\underset{\sim}{\varepsilon}'\underset{\sim}{\varepsilon} &= (\underset{\sim}{Y} - \underset{\sim}{X}\underset{\sim}{\hat{\beta}})'(\underset{\sim}{Y} - \underset{\sim}{X}\underset{\sim}{\hat{\beta}}) \\
&= (\underset{\sim}{Y}' - \underset{\sim}{\hat{\beta}}'\underset{\sim}{X}')(\underset{\sim}{Y} - \underset{\sim}{X}\underset{\sim}{\hat{\beta}}) \\
&= \underset{\sim}{Y}'\underset{\sim}{Y} - 2\underset{\sim}{\hat{\beta}}'\underset{\sim}{X}'\underset{\sim}{Y} + \underset{\sim}{\hat{\beta}}\underset{\sim}{X}'\underset{\sim}{X}\underset{\sim}{\hat{\beta}}
\end{aligned}$$

---

[3] ประชุม สุวัตถี, ดร., ทฤษฎีการอนุมานเชิงสถิติ (กรุงเทพมหานคร : 2527), หน้า 158

ตัวประมาณกำลังสองต่ำสุด คือตัวประมาณที่ได้จากการทำให้ $\underset{\sim}{\varepsilon}'\underset{\sim}{\varepsilon}$ มีค่าต่ำสุด หาได้โดยการหาอนุพันธ์ (Differentiate) เทียบกับ $\underset{\sim}{\beta}$ แล้วทำให้เท่ากับ $\underset{\sim}{0}$

$$\frac{\partial}{\partial \underset{\sim}{\beta}}(\underset{\sim}{Y}'\underset{\sim}{Y}-2\hat{\underset{\sim}{\beta}}'X'\underset{\sim}{Y}+\hat{\underset{\sim}{\beta}}XX\hat{\underset{\sim}{\beta}}) = 0$$

จะได้ $$\hat{\underset{\sim}{\beta}} = (XX)^{-1}X'\underset{\sim}{Y}$$

2.2.2 <u>วิธีการของแชตเทอร์จีและแมคลีช</u>[4]

เป็นวิธีการประมาณพารามิเตอร์ซึ่งเสนอโดยแชตเทอร์จีและแมคลีช (S.Chatterjee and D.L. McLeich) ได้พัฒนาขึ้นจากวิธีการประมาณแชงเส้นเมื่อค่าสังเกตบางค่าสูญหาย ซึ่งคิดขึ้นโดย เยทส์ (Yates) และ ซีเบอร์ (Seber) โดย เยทส์และซีเบอร์ประมาณค่าที่สูญหายด้วยค่าคาดหวังที่ทำให้ผลรวมกำลังสองของความคลาดเคลื่อนมีค่าต่ำสุด วิธีของแชตเทอร์จีและแมคลีชเป็นวิธีการนอนพาราเมตริกซ์ (Nonparametric Method) โดยมีข้อสมมติว่าความคลาดเคลื่อน ε_i มีการแจกแจงที่เหมือนและเป็นอิสระกันด้วยค่าเฉลี่ย เท่ากับ 0 และความแปรปรวน σ^2 วิธีการนี้ใช้การวนทำซ้ำซึ่งในแต่ละรอบจะประมาณค่าคาดหวังของค่าสังเกตที่ถูกตัด $\hat{Y}_i$ โดยมีหลักว่าค่าคาดหวังที่ได้จะมีค่ามากกว่าค่าที่ถูกตัด Tc ดังนี้

$$\hat{Y}_i \begin{cases} Tc & \text{เมื่อ } Tc-\sum_j \hat{\beta}_j X^c_{ij} \geq 0 \\ \sum_j \hat{\beta}_j X^c_{ij} & \text{เมื่อ } Tc-\sum_j \hat{\beta}_j X^c_{ij} < 0 \end{cases}$$

และข้อมูลที่นำมาวิเคราะห์จะถือว่าค่าสังเกตที่ถูกตัดแทนเป็นค่าสังเกตที่ไม่ถูกตัด แล้วประมาณพารามิเตอร์ $\hat{\underset{\sim}{\beta}}$ ด้วยวิธีกำลังสองต่ำสุด

---

[4] S. Chatterjee and D.L. McLeish. Fitting Linear Regression to Censored Data by Least Squares and Maximum Likelihood Methods. <u>Communications Statistics.</u> (1986), A 15, 11, pp. 3227-3243.

ขั้นตอนการหาตัวประมาณ $\hat{\beta}$ สำหรับวิธีของแมคเทอร์ลีและแมคลีช

ขั้นที่ 1 หาค่าประมาณ $\hat{\beta}$ เริ่มต้น โดยการใช้วิธีการกำลังสองต่ำสุดกับเฉพาะค่าสังเกตที่ไม่ถูกตัดทิ้ง

$$\hat{\beta} = \left(X^{u'}X^{u}\right)^{-1}X^{u'}Y^{u}$$

เมื่อ X^{u} เป็นเมตริกซ์ของตัวแปรอิสระของข้อมูลที่ไม่ถูกตัด

Y^{u} เป็นเวกเตอร์ของตัวแปรตามของข้อมูลที่ไม่ถูกตัด

ขั้นที่ 2 สำหรับข้อมูลที่ถูกตัด คำนวณค่าประมาณ $\hat{Y}_i$ ดังนี้

$$\hat{Y}_i = \sum_{j=0}^{2} \hat{\beta}_j X_{ij} \quad \text{เมื่อ} \quad X_{i0} = 1$$

เปรียบเทียบค่าประมาณ $\hat{Y}_i$ ที่ได้กับค่าที่ถูกตัด Tc ถ้าค่า $\hat{Y}_i$ มากกว่าค่า Tc ให้แทนค่าที่ถูกตัดด้วย $\hat{Y}_i$ ถ้าค่า $\hat{Y}_i$ น้อยกว่าค่า Tc ให้คงค่าที่ถูกตัด Tc ไว้อย่างเดิม

ขั้นที่ 3 นำค่าประมาณที่คำนวณได้จากขั้นที่ 2 และข้อมูลที่ไม่ได้ถูกตัดคำนวณหาตัวประมาณ $\hat{\beta}$ ด้วยวิธีกำลังสองต่ำสุด

ขั้นที่ 4 แทนค่า $\hat{\beta}$ ที่ได้จากขั้นที่ 3 ลงในขั้นที่ 2 แล้วทำการวนซ้ำขั้นที่ 2 ถึงขั้นที่ 3 ทำไปจนกระทั่งค่าของ $\hat{\beta}$ ที่ได้เท่ากับค่าของ $\hat{\beta}$ ในรอบที่ผ่านมา ก็จะหยุดการวนทำซ้ำจะได้ $\hat{\beta}$ เป็นค่าประมาณของพารามิเตอร์

ในบางครั้งซึ่งค่า $\hat{\beta}$ จะแกว่งอยู่ระหว่าง 2 ค่า ในกรณีนี้จะใช้ค่าเฉลี่ยระหว่างสองค่าเป็นค่าประมาณของ $\hat{\beta}$

2.2.3 วิธีการประมาณของบัคเลย์และเจมส์ [5]

วิธีการของบัคเลย์และเจมส์เสนอโดยโจนาธานบัคเลย์และเอียนเจมส์ (Jonathan Buckley and Ian James) เป็นวิธีการประมาณพารามิเตอร์ในสมการถดถอยเชิงเส้นที่เป็นนอนพาราเมตริก (Nonparametric Method) ซึ่งวิธีการของบัคเลย์และเจมส์มีข้อสมมติว่า ค่าคลาดเคลื่อน ε_i เป็นอิสระ และมีฟังก์ชันการแจกแจง F ที่ไม่มีรูปแบบเฉพาะ มีค่าเฉลี่ยเป็น 0 ความแปรปรวน σ^2 และมีฟังก์ชันการอยู่รอดเป็น $S=1-F$

บัคเลย์และเจมส์ได้ทำการประมาณค่าของข้อมูลที่ถูกตัดด้วยค่าคาดหวังอย่างมีเงื่อนไข $E\left(Y_i/Y_i\rangle Tc, X_i \underset{\sim}{\beta}\right)$ เมื่อ Tc เป็นค่าที่ถูกตัด แต่เนื่องจากไม่ทราบฟังก์ชันการแจกแจง F จึงไม่สามารถหาค่า $E\left(Y_i/Y_i\rangle Tc, X_i \underset{\sim}{\beta}\right)$ จึงประมาณค่า F ด้วยตัวประมาณพีแอล

ดังนั้นข้อมูลที่ถูกตัดจะถูกแทนด้วย

$$\bar{Y}_i\left(\underset{\sim}{\hat{\beta}}\right) = \left(\hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2}\right) + \frac{\sum_{k} W_k\left(\hat{\beta}\right)\left(Y_k - \left(\hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2}\right)\right)}{1 - \hat{F}\left(Tc - \left(\hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2}\right)\right)}$$

เมื่อ $\delta_{(i)} = \begin{cases} 1 & \text{เมื่อข้อมูลลำดับที่ } i \text{ ไม่ถูกตัด} \\ 0 & \text{เมื่อข้อมูลลำดับที่ } i \text{ ถูกตัด} \end{cases}$

ขั้นตอนการหาตัวประมาณ $\underset{\sim}{\hat{\beta}}$ สำหรับวิธีของบัคเลย์และเจมส์

ขั้นที่ 1 หาค่าประมาณ $\underset{\sim}{\hat{\beta}}$ เริ่มต้น โดยการใช้วิธีการกำลังสองต่ำสุดกับเฉพาะค่าสังเกตที่ไม่ถูกตัดทิ้ง

---

[5] Jonathan Buckley and Ian James. Linear Regression with Censored Data. Biometrika. (1979),66,3, pp. 429-436

Rupert G. Miller. Survival Analysis. (New York : John Willey,1981), pp. 150-154

$$\underset{\sim}{\hat{\beta}} = \left(X^{u'}X^{u}\right)^{-1}X^{u'}Y^{u}$$

เมื่อ X^{u} เป็นเมตริกซ์ของตัวแปรอิสระของข้อมูลที่ไม่ถูกตัด

Y^{u} เป็นเวกเตอร์ของตัวแปรตามของข้อมูลที่ไม่ถูกตัด

ขั้นที่ 2 จากข้อมูลทั้งหมดคือรวมทั้งข้อมูลที่ถูกตัดและข้อมูลที่ไม่ถูกตัด ให้หาความคลาดเคลื่อน e จาก

$$e_i = Y_i - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} \quad ; \; i = 1,2,\ldots,N$$

ขั้นที่ 3 ให้เรียงลำดับที่ (Rank) ความคลาดเคลื่อน จากน้อยไปมาก จะได้ $e_{(1)}, e_{(2)}, \ldots, e_{(N)}$ ในกรณีที่ลำดับที่ของค่าที่ถูกตัดและค่าที่ไม่ถูกตัดมีค่าเท่ากันให้ลำดับที่ของค่าที่ไม่ถูกตัดมาหน้าค่าที่ถูกตัด

ขั้นที่ 4 ให้เปลี่ยนค่าลำดับที่ของค่าที่ถูกตัดเป็น 0 ส่วนค่าลำดับที่ของค่าที่ไม่ถูกตัดให้คงไว้

ขั้นที่ 5 หาค่า PL Estimator $\hat{S}$ จาก

$$\hat{S}(e_i, \underset{\sim}{\hat{\beta}}) = \prod_{i:e_{(i)} \le e_i}^{N} \left[\frac{N-i}{N-i+1}\right]^{\delta(i)}$$

เมื่อ i คือลำดับที่ของความคลาดเคลื่อน

N คือจำนวนตัวอย่างทั้งหมด

PL Estimator $\hat{S}$ นี้คือ ค่า Survival Time

ขั้นที่ 6 คำนวณหาฟังก์ชัน $\hat{F}(e_i, \underset{\sim}{\hat{\beta}})$ จาก

$$\hat{F}(e_i, \underset{\sim}{\hat{\beta}}) = 1 - \prod_{i:e_{(i)} \le e_i}^{N} \left[\frac{N-i}{N-i+1}\right]^{\delta(i)}$$

ขั้นที่ 7 คำนวณหาค่าถ่วงน้ำหนัก $W_i(\hat{\beta})$ จาก

$$W_1(\hat{\beta}) = \hat{F}_1$$
$$W_2(\hat{\beta}) = \hat{F}_2 - \hat{F}_1$$
$$\vdots \qquad \vdots$$
$$W_N(\hat{\beta}) = \hat{F}_N - \hat{F}_{N-1}$$

ในกรณีลำดับที่สูงสุดของความคลาดเคลื่อนเป็นค่าที่ถูกตัดให้ปรับตัวถ่วงน้ำหนัก $W_i(\hat{\beta})$ ใหม่ดังนี้

$$W^*_i(\hat{\beta}) = \frac{W_i(\hat{\beta})}{\sum_{u} W_j(\hat{\beta})}$$

เมื่อ $\sum_{u}$ เป็นค่าผลบวกเฉพาะกับข้อมูลที่ไม่ถูกตัด

ขั้นที่ 8 ให้เปลี่ยนค่าลำดับที่ของค่าที่ไม่ถูกตัดเป็น 0 ส่วนค่าลำดับที่ของค่าที่ถูกตัดให้คงไว้

ขั้นที่ 9 หาค่า PL Estimator $\hat{S}$

$$\hat{S}(e_i, \underset{\sim}{\hat{\beta}}) = \prod_{i:e_{(i)} \le e_i}^{N} \left[\frac{N-i}{N-i+1}\right]^{\delta(i)}$$

ขั้นที่ 10 หาฟังก์ชัน $\hat{F}(e_i, \underset{\sim}{\hat{\beta}})$ จาก

$$\hat{F}(e_i, \underset{\sim}{\hat{\beta}}) = 1 - \hat{S}(e_i, \underset{\sim}{\hat{\beta}})$$
$$\hat{F}(e_i, \underset{\sim}{\hat{\beta}}) = 1 - \prod_{i:e_{(i)} \le e_i}^{N} \left[\frac{N-i}{N-i+1}\right]^{\delta(i)}$$

ขั้นที่ 11 ประมาณค่าที่ถูกตัดด้วย

$$\hat{Y}_i\left(\underset{\sim}{\hat{\beta}}\right) = (\hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2}) + \frac{\sum_{u} W_k(\hat{\beta})\left(Y_k - (\hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2})\right)}{1 - \hat{F}\left(Tc - (\hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2})\right)}$$

ขั้นที่ 12 จากข้อมูลที่ไม่ถูกและค่าประมาณของข้อมูลที่ถูกตัดคำนวณหา $\hat{\beta}$ จาก

$$\hat{\beta} = (X'X)^{-1}X'Y$$

ขั้นที่ 13 แทนค่า $\hat{\beta}$ ที่ได้จากขั้นที่ 12 ลงในขั้นที่ 2 แล้วคำนวณจากขั้นที่ 2 จนถึงขั้นที่ 12 คำนวณซ้ำดังนี้จนกระทั่งค่า $\hat{\beta}$ ที่ได้เท่ากับค่า $\hat{\beta}$ ในรอบที่ผ่านมาจะหยุด ก็จะได้ค่าประมาณของ $\hat{\beta}$ ในบางครั้งการประมาณค่า $\hat{\beta}$ จะแกว่งระหว่าง 2 ค่า ในกรณีนี้จะใช้ค่าเฉลี่ยระหว่าง 2 ค่านั้น เป็นค่าประมาณ $\hat{\beta}$

ขั้นที่ 14 คำนวณ $\hat{\alpha}$ จาก

$$\hat{\alpha} = \frac{1}{N}\left\{\sum_{u} Y_i + \sum_{c} \overline{Y}_i(\hat{\beta})\right\} - \left(\hat{\beta}_1\overline{X}_1 + \hat{\beta}_2\overline{X}_2\right)$$

เมื่อ $\sum_{u}$ เป็นค่าผลบวกเฉพาะข้อมูลที่ไม่ถูกตัด

$\sum_{c}$ เป็นค่าผลบวกเฉพาะข้อมูลที่ถูกตัด

$\overline{X}_1$ เป็นค่าเฉลี่ยของตัวแปรอิสระที่ 1

$\overline{X}_2$ เป็นค่าเฉลี่ยของตัวแปรอิสระที่ 2

2.2.4 <u>วิธีการภาวะน่าจะเป็นสูงสุดด้วยขั้นตอนวิธีอีเอ็ม</u> [6]

การประมาณพารามิเตอร์ในสมการถดถอยเชิงเส้นเมื่อตัวแปรตามบางค่าถูกตัดทางขวาด้วยภาวะน่าจะเป็นสูงสุด ซึ่งเสนอโดยเมอร์เรย์ ไอท์เคน (Murrey Aitkin) เป็นวิธีการที่อาศัย อีเอ็ม อัลกอริทึม (EM Algorithm) ซึ่งเสนอโดยเด็มเสตอร์ ลาเอ็ด และ รูบิน

---

[6] Murry Aitkin., A Note on The Regression Analysis of Censored Data. <u>Technometrics</u>. (1981),23,2, pp. 161-163.

(Dempster Laird and Rubin) เป็นวิธีที่วนทำซ้ำ และในแต่ละรอบของการทำซ้ำจะประกอบไปด้วย 2 ขั้นตอนคือขั้นตอนของการประมาณค่าคาดหวังอย่างมีเงื่อนไขของข้อมูลที่ถูกตัด และขั้นตอนของการประมาณค่าสูงสุดของพารามิเตอร์โดยอาศัยทั้งข้อมูลที่ไม่ถูกตัดและถูกตัด

จากรูปแบบความสัมพันธ์

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_{i2} X_{i2} + \varepsilon_i$$

เมื่อ $\varepsilon_i \sim N(0,\sigma^2)$

กำหนดให้ $\mu_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}$

$$z_i = (y_i - \mu_i)/\sigma$$

$$f(z_i) = \frac{1}{\sqrt{2\pi}}\exp(-z^2/2)$$

$$S(z_i) = 1 - F(z_i) = \int_{z_i}^{\infty} f(t)dt$$

$$h(z_i) = \frac{f(z_i)}{1-F(z_i)}$$

$$\phi(y_i) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{\sqrt{2}\sigma^2}(y_i-\mu_i)^2\right)$$

$$= \frac{1}{\sigma} f(z_i)$$

ดังนั้นฟังก์ชันภาวะน่าจะเป็นของค่าที่ไม่ถูกตัดและค่าที่ถูกตัดคือ

$$L(\beta,\sigma) = \prod_{i=1}^{n} f(y_i) \prod_{n+1}^{n+m} S(y_i)$$

$$= \frac{1}{\sigma^n} \prod_{i=1}^{n} f(z_i) \prod_{n+1}^{n+m} S(z_i)$$

เมื่อ n คือจำนวนตัวอย่างที่ไม่ถูกตัด

m คือจำนวนตัวอย่างที่ถูกตัด

และฟังก์ชันลือกภาวะน่าจะเป็นคือ

$$\log L(\beta,\sigma) = -n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu_i)^2 + \sum_{i=n+1}^{n+m}\log S\left(\frac{y_i-\mu_i}{\sigma}\right)$$

$$= -n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i^2 - 2y_i\sum_{j=1}^{2}\beta_j x_{ij} + \left(\sum\beta_j x_{ij}\right)^2\right)$$

$$+ \sum_{i=n+1}^{n+m}\log S\left(\frac{y_i-\mu_i}{\sigma}\right)$$

หาอนุพันธ์บางส่วน (Partial Derivatives) ของล็อกภาวะน่าจะเป็นเทียบกับ β_j และฟังก์ชันล็อกภาวะน่าจะเป็นคือ

$$\frac{\partial \log L}{\partial \beta_j} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\mu_i)x_{ij} + \frac{1}{\sigma}\sum_{i=n+1}^{n+m}\frac{f(z_i)}{1-F(z_i)}x_{ij}$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\mu_i)x_{ij} + \frac{1}{\sigma}\sum_{i=n+1}^{n+m}h(z_i)x_{ij}$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n+m}(W_i-\mu_i)x_{ij}$$

โดยที่ $$W_i = \begin{cases} y_i & \text{เมื่อ } i = 1,2,....n \\ \mu_i + \sigma h(z_i) & \text{เมื่อ } i = n+1,....,n+m \end{cases}$$

หาอนุพันธ์บางส่วนของล็อกภาวะน่าจะเป็นเทียบกับ σ และให้เท่ากับ 0

$$\frac{\partial \log L}{\partial \sigma} = \frac{-n}{\sigma} + \frac{1}{\sigma^3}\sum_{1}^{n}(y_i-\mu_i)^2 + \frac{1}{\sigma^2}\sum_{n+1}^{n+m}h(z_i)(y_i-\mu_i) = 0$$

$$-n + \sum_{1}^{n}h(z_i)z_i + \frac{1}{\sigma^2}\sum_{n+1}^{n+m}(y_i-\mu_i)^2 = 0$$

$$\sigma^2 = \left(\sum_{i=1}^{n}(y_i-\mu_i)^2\right)\Big/\left(n - \sum_{n+1}^{n+m}h(z_i)z_i\right)$$

สถิติที่เพียงพอสำหรับพารามิเตอร์ของข้อมูลที่สมบูรณ์คือ $\sum_{i=1}^{n+m}x_{ij}y_i$ และ $\sum_{i=1}^{n+m}y_i^2$ และค่าคาดหวังอย่างมีเงื่อนไขคือ

$$E\left(\sum_{1}^{n+m}x_{ij}y_i\right) = \sum_{1}^{n}x_{ij} + \sum_{n+1}^{n+m}x_{ij}E\left(y_i / y_i > Tc, \beta, \sigma^2\right)$$

$$E\left(\sum_{1}^{n+m}y_i^2\right) = \sum_{1}^{n}y_i^2 + \sum_{n+1}^{n+m}E\left(y_i^2 / y_i > Tc, \beta, \sigma^2\right)$$

เมื่อ $$E\left(y_i / y_i > Tc, \beta, \sigma^2\right) = \mu_i + \sigma h(z_i)$$

$$E(y_t^2 / y_t > Tc, \beta, \sigma^2) = \mu_t + \sigma^2 + \sigma(y_t + \mu_t)h(z_t)$$

ดังนั้นสำหรับข้อมูลที่ถูกตัด หาค่าประมาณ $\hat{y}_t$ และ $\hat{y}_t^2$ ได้ดังนี้

$$\hat{y}_t = \hat{\mu}_t + \hat{\sigma}h(z_t) \quad (1)$$

$$\hat{y}_t^2 = \hat{\mu}^2{}_t + \hat{\sigma}^2 + \hat{\sigma}(y_t + \hat{\mu}_t)h(z_t) \quad (2)$$

การประมาณค่าสูงสุดของพารามิเตอร์ในขั้นเอ็ม (Maximization Step : M Step) ในรอบที่ k+1 จะประมาณเพารามิเตอร์ $\underset{\sim}{\hat{\beta}}$ ด้วยการแทนค่าที่ถูกตัดด้วย $\hat{\mu}_t + \hat{\sigma}h(z_t)$ จากรอบที่ k แล้ว คำนวณหา $\underset{\sim}{\hat{\beta}}$ ด้วยวิธีกำลังสองต่ำสุด และ $\hat{\sigma}^2$ คำนวณหาโดยใน $\sum_{t=1}^{n+m}(y_t - \hat{\mu}_t^k)^2$ จะแทน y_t ด้วย $\hat{\mu}_t + \hat{\sigma}h(z_t)$ และแทน y_t^2 ด้วย $\hat{\mu}^2{}_t + \hat{\sigma}^2 + \hat{\sigma}(y_t + \hat{\mu}_t)h(z_t)$ ดังนั้น

$$\hat{\sigma}^2{}_{(k+1)} = \frac{1}{(n+m)}\left[\sum_{1}^{n}(y_t - \hat{\mu}_t^{(k)})^2 + \hat{\sigma}^2{}_{(k)}\sum_{n+1}^{n+m}\left(1 + z_t^{(k)}h(z_t^{(k)})\right)\right]$$

ขั้นตอนการประมาณพารามิเตอร์ $\underset{\sim}{\hat{\beta}}$ และ $\hat{\sigma}^2$ โดยวิธีอารสาวะน่าจะเป็น มีขั้นตอนดังนี้

ขั้นที่ 1 ประมาณค่าพารามิเตอร์ $\underset{\sim}{\hat{\beta}}$, $\hat{\sigma}^2$ เริ่มต้นด้วยวิธีกำลังสองต่ำสุด โดยข้อมูลที่นำมาวิเคราะห์จะถือว่าข้อมูลที่ถูกตัดทิ้งเสมือนข้อมูลที่ไม่ถูกตัดทิ้งค่าประมาณเพารามิเตอร์ที่ได้คือ $\underset{\sim}{\hat{\beta}}$, $\hat{\sigma}^2$

ขั้นที่ 2 เฉพาะข้อมูลที่ถูกตัดทิ้ง และใช้พารามิเตอร์จากขั้น 1 คำนวณ

$$\hat{\mu}_i = \hat{\beta}_0 + \sum_{j=1}^{2}\hat{\beta}_j x_{ij} \quad ; i = n+1, n+2, \ldots, n+m$$

$$z_i = (Tc - \hat{\mu}_i)/\hat{\sigma}$$

$$f(z_i) = \frac{1}{\sqrt{2\pi}}\exp(-z^2/2)$$

$$S(z_i) = 1 - F(z_i) = \int_{z_i}^{\infty} f(t)dt$$

$$h(z_i) = \frac{f(z_i)}{1 - F(z_i)}$$

ขั้นที่ 3 ประมาณข้อมูลที่ถูกตัดทิ้งด้วยค่าคาดหวังอย่างมีเงื่อนไข

$$E(y_i / y_i > Tc, \beta, \sigma^2) = W_i$$

$$W_i = \hat{\mu}_i + \hat{\sigma}h(z_i) \quad ; i = n+1, n+2, \ldots, n+m$$

แล้วคำนวณค่า $y_i^*(\beta)$

$$y_i^*(\beta) = y_i\delta_i + E(y_i / y_i > Tc, \beta, \sigma^2)(1-\delta_i)$$
$$y_i^*(\beta) = y_i\delta_i + W_i(1-\delta_i)$$

เมื่อ $\delta_i = \begin{cases} 1 & \text{เมื่อข้อมูลไม่ถูกตัดทิ้ง} \\ 0 & \text{เมื่อข้อมูลถูกตัดทิ้ง} \end{cases}$

ขั้นที่ 4 นำค่า $y_i^*(\beta)$ จากขั้นที่ 3 เพื่อประมาณค่าพารามิเตอร์ $\hat{\beta}$ ด้วยวิธีกำลังสองต่ำสุดและ $\hat{\sigma}^2$ จาก

$$\hat{\sigma}^2 = \left[\sum_{1}^{n}(y_i - \hat{\mu}_i) + \hat{\sigma}^2 \sum_{n+1}^{n+m}\left(1 + z_i h(z_i)\right)\right] \Big/ (n+m)$$

ขั้นที่ 5 เปรียบเทียบค่าประมาณพารามิเตอร์ $\hat{\beta}$, $\hat{\sigma}^2$ จากขั้นที่ 4 กับขั้นที่ 1 ถ้ายังไม่เท่ากันให้คำนวณรอบต่อไป

ขั้นที่ 6 เฉพาะข้อมูลที่ถูกตัด และพารามิเตอร์ $\hat{\beta}, \hat{\sigma}^2$ จากขั้นที่ 4 คำนวณ

$$\hat{\mu}_i = \hat{\beta}_0 + \sum \hat{\beta}_j x_{ij}$$
$$z_i = (Tc - \hat{\mu}_i) / \hat{\sigma}$$
$$f(z_i) = \frac{1}{\sqrt{2\pi}} \exp(-z_i^2 / 2)$$
$$S(z_i) = 1 - F(z_i) = \int_{z_i}^{\infty} f(t)dt$$
$$S(z_i) = \frac{f(z_i)}{1 - F(z_i)}$$

ขั้นที่ 7 ประมาณค่าที่ถูกตัดทิ้งด้วยค่าคาดหวังอย่างมีเงื่อนไข

$$E(y_i / y_i > Tc, \beta, \sigma^2) = W_i$$
$$W_i = \hat{\mu}_i + \hat{\sigma} h(z_i)$$

แล้วคำนวณหา $y_i^*(\beta) = y_i\delta_i + W_i(1-\delta_i)$

เมื่อ $\delta_i = \begin{cases} 1 & \text{เมื่อข้อมูลไม่ถูกตัดทิ้ง} \\ 0 & \text{เมื่อข้อมูลถูกตัดทิ้ง} \end{cases}$

ขั้นที่ 8 นำค่าประมาณพารามิเตอร์ $\hat{\beta}, \hat{\sigma}^2$ จากรอบที่ผ่านมา (k) และ $y_i^*(\beta)$ จากขั้นที่ 7 เพื่อประมาณค่าพารามิเตอร์รอบปัจจุบัน (k+1) คือ $\hat{\beta}^{(k+1)}$ ด้วยวิธีกำลังสองต่ำสุดและ $\hat{\sigma}^{2}_{(k+1)}$ จาก

$$\hat{\sigma}^{2}_{(k+1)} = \left[\sum_{1}^{n}\left(y_i - \hat{\mu}_i^{(k)}\right) + \hat{\sigma}^{2}_{(k)} \sum_{n+1}^{n+m}\left(1 + \hat{z}_i^{(k)} h\left(\hat{z}_i^{(k)}\right)\right)\right] \Big/ (n+m)$$

ขั้นที่ 9 เปรียบเทียบค่าประมาณพารามิเตอร์จากขั้นที่ 4 กับขั้นที่ 8 คือเปรียบเทียบ $\hat{\beta}$ ขั้นที่ 8 กับ $\hat{\beta}$ ในขั้นที่ 4 และ $\hat{\sigma}^2$ ในขั้นที่ 8 กับ $\hat{\sigma}^2$ ในขั้นที่ 4 ถ้าค่าประมาณของพารามิเตอร์ไม่เท่ากันก็จะนำค่าประมาณพารามิเตอร์ในขั้นที่ 8 แทนค่าลงในขั้นที่ 6 และวนทำซ้ำรอบใหม่จากขั้นที่ 6 ถึงขั้นที่ 9 ถ้าค่าประมาณพารามิเตอร์มีค่าอยู่ระหว่างสองค่าแล้วค่าประมาณพารามิเตอร์คือค่าเฉลี่ยของสองค่านั้น