บทที่ 2

ระเบียบวิธีที่ใช้ในการวิจัย

การศึกษาความถดถอยเชิงเส้นอย่างง่าย เมื่อตัวแปรตามบางค่ามีค่าขาดหาย ในการศึกษาครั้งนี้ ได้ทำการศึกษาวิธีการประมาณค่า 3 วิธี คือ วิธีการกำลังสองต่ำสุด, วิธีการของมิลเลอร์, วิธีการของบัคเลย์และเจมส์ ซึ่งในบทนี้จะกล่าวถึงความรู้ ทฤษฎีพื้นฐานและรายละเอียดของวิธีการประมาณค่าเกี่ยวกับพารามิเตอร์แต่ละวิธี รายละเอียดต่าง ๆ เป็นดังนี้

2.1 <u>ความรู้และทฤษฎีพื้นฐาน</u>

2.1.1 <u>ประเภทของค่าขาดหาย</u> (Type of Censoring)

2.1.1.1 <u>ค่าขาดหายประเภทที่ 1</u> (Type I Censoring) ค่าขาดหายประเภทนี้จะกำหนดเวลาของการเกิด Censoring Time เอาไว้ล่วงหน้า ซึ่งจะเรียกว่า Fixed Censoring Time

ตัวอย่างเช่น ทำการทดลองเกี่ยวกับอายุการใช้งานของหลอดไฟ เรานำหลอดไฟมาจำนวนหนึ่ง แล้วเริ่มทำการทดลอง โดยการผ่านกระแสไฟฟ้าเข้าไป แล้วบันทึกเวลาไว้ตั้งแต่เริ่มผ่านกระแสไฟฟ้าเข้าไปจนกระทั่งหลอดขาดนั้นใช้เวลากี่ชั่วโมงหรือกี่วัน ถ้าจำนวนชั่วโมงที่หลอดใดหลอดหนึ่งขาดน้อยกว่าจำนวนชั่วโมงที่เรากำหนดไว้ จะถือว่าค่าสังเกตนี้เป็นค่าที่ไม่ขาดหาย หรือเป็น Survival Time แต่ถ้าจำนวนชั่วโมง ของหลอดไฟใดมีค่ามากกว่าจำนวนชั่วโมงที่เรากำหนดไว้ จะถือว่าค่าสังเกตนี้เป็นค่าที่ขาดหาย หรือเป็น Censoring Time ที่เราต้องกำหนดจำนวนชั่วโมงไว้ล่วงหน้า ก็เนื่องจากว่า หลอดไฟบางหลอดจะมีอายุการใช้งานที่ยาวนานมาก ซึ่งเราอาจจะไม่สามารถรอจนกระทั่งหลอดไฟนั้นเสื่อมคุณภาพได้ จำนวนชั่วโมงที่เรากำหนดไว้นี้จะเป็นตัวหยุดการทดลอง

ถ้า $T_1, \ldots, T_n$ เป็นค่าสังเกตที่ไม่ขาดหาย

t_C เป็นเวลาของการเกิดค่าขาดหาย

ค่าสังเกตใหม่ที่เราต้องการคือ $Y_1, \ldots, Y_n$ เมื่อ

$$Y_i = \begin{cases} T_i & \text{ถ้า} \quad T_i \leqslant t_C \\ t_c & \text{ถ้า} \quad t_C < T_i \end{cases}$$

2.1.1.2 <u>ค่าขาดหายประเภทที่ 2</u> (Type II Censoring) ในบางกรณีเราไม่อาจกำหนดเวลาของการเกิดค่าขาดหาย (Censoring Time) ที่เหมาะสมได้ ดังนั้น เราจะกำหนดจำนวนค่าสังเกตที่ไม่ขาดหายแทน นั่นคือเมื่อค่าสังเกตที่ไม่ขาดหายเกิดขึ้นครบตามจำนวนที่กำหนดไว้ก็จะหยุดทำการทดลอง

จากตัวอย่างที่แล้ว ถ้าจำนวนตัวอย่างของหลอดไฟที่ทดลองขาดหรือเสื่อมคุณภาพเท่ากับจำนวนตัวอย่างที่เรากำหนดไว้ก็จะหยุดทำการทดลอง

ถ้า r คือจำนวนของค่าสังเกตที่ไม่ขาดหายที่กำหนดไว้

$T_{(1)}, \ldots, T_{(r)}$ คือ ค่าสังเกตที่ไม่ขาดหาย

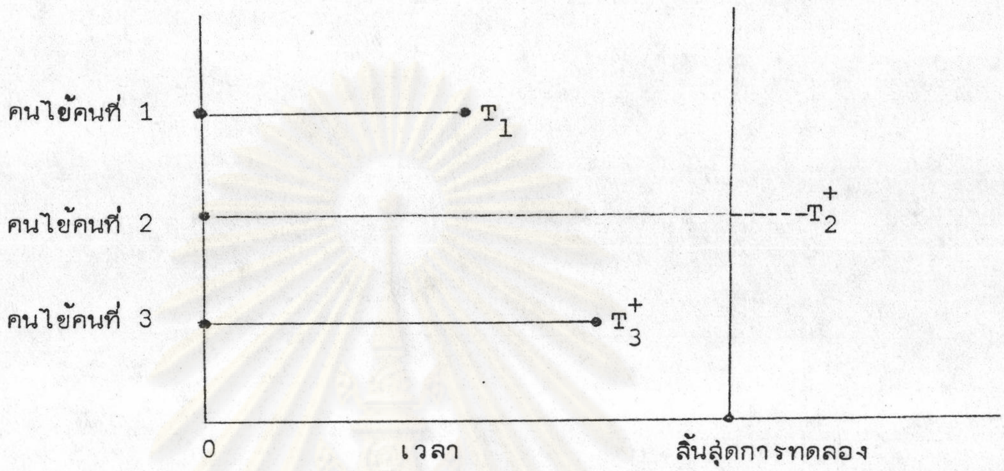
ค่าสังเกตที่ต้องการ คือ

$$\begin{aligned} Y_{(1)} &= T_{(1)} \\ Y_{(2)} &= T_{(2)} \\ &\vdots \\ Y_{(r)} &= T_{(r)} \\ Y_{(r+1)} &= T_{(r)} \\ &\vdots \\ Y_{(n)} &= T_{(r)} \end{aligned}$$

2.1.1.3 <u>ค่าขาดหายแบบสุ่ม</u> (Random Censoring) ลักษณะของค่าขาดหายประเภทนี้ส่วนใหญ่จะพบในการทดลองเกี่ยวกับการแพทย์ ซึ่งค่าขาดหายประเภทนี้จะเกิดขึ้นในการ

ทดลอง เมื่อคนไข้ถอนตัวจากการทดลอง หรือคนไข้ยังมีชีวิตอยู่เมื่อสิ้นสุดการทดลอง ซึ่งจะทำให้ไม่สามารถได้ค่าที่แน่นอนของค่าสังเกตนั้นได้

ภาพประกอบแสดงการเกิดค่าขาดหาย



คนไข้คนที่ 1 เข้าทำการทดลอง และตายที่เวลา T_1 ค่าสังเกตนี้จะเป็นค่าสังเกตที่ไม่ขาดหาย

คนไข้คนที่ 2 เข้าทำการทดลองและเมื่อสิ้นสุดการทดลองแล้วยังมีชีวิตอยู่ ค่าสังเกตนี้จะเป็นค่าสังเกตที่ขาดหาย

คนไข้คนที่ 3 เข้าทำการทดลอง และถอนตัวจากการทดลองเมื่อเวลา T_3^+ ค่าสังเกตนี้จะเป็นค่าสังเกตที่ขาดหาย

ให้ $C_1, \ldots, C_n$ เป็นค่าสังเกตที่ขาดหาย

$T_1, \ldots, T_n$ เป็นค่าสังเกตที่ไม่ขาดหาย

ค่าสังเกตใหม่ที่ต้องการคือ $(Y_1, \delta_1), \ldots, (Y_n, \delta_n)$

เมื่อ $Y_i = \min(T_i, C_i)$

$$\delta_i = \begin{cases} 1 \text{ ถ้า } T_i \leqslant C_i \text{ ; เป็นข้อมูลจากค่าที่ไม่ขาดหาย} \\ 0 \text{ ถ้า } T_i > C_i \text{ ; เป็นข้อมูลจากค่าขาดหาย} \end{cases}$$

และการแจกแจงของ T_i และ C_i จะเป็นอิสระต่อกัน

2.1.2 Product Limit Estimator (PL Estimator)[1]

ให้ $Y_{(1)} \leqslant Y_{(2)} \leqslant \ldots \leqslant Y_{(n)}$ เป็นค่าสังเกตที่เรียงลำดับของข้อมูลที่ขาดหาย และไม่ขาดหาย PL Estimator สามารถแสดงได้ดังนี้

$$\hat{S}(y) = \prod_{Y_{(i)} \leqslant y} \frac{n-i}{n-i+1}$$

เมื่อ n คือ จำนวนข้อมูลที่ขาดหายและไม่ขาดหาย

i คือ ลำดับที่ของข้อมูล

ตัวอย่างการหาตัวประมาณ PL Estimator

จากค่าสังเกต

9, 13, 13+, 18, 23, 28+, 31, 34, 45+, 48, 161+

เมื่อ + หมายถึง ข้อมูลที่ขาดหาย

---

[1] ดูรายละเอียดเพิ่มเติมได้จาก

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

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

จะได้ PL Estimator ดังนี้

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

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

$$\hat{S}(13) = \hat{S}(9) \times \frac{9}{10} = .82$$

$$\hat{S}(18) = \hat{S}(13) \times \frac{7}{8} = .72$$

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

$$\hat{S}(31) = \hat{S}(23) \times \frac{4}{5} = .49$$

$$\hat{S}(34) = \hat{S}(31) \times \frac{3}{4} = .37$$

$$\hat{S}(48) = \hat{S}(34) \times \frac{1}{2} = .18$$

2.2 <u>การประมาณค่าพารามิเตอร์</u>

2.2.1 <u>วิธีการกำลังสองต่ำสุด</u>

วิธีการหาตัวประมาณของพารามิเตอร์วิธีนี้ เป็นวิธีที่มีรากฐานมาจากทฤษฎีการประมาณเชิงเส้น (Theory of Linear Estimation) เป็นวิธีที่คิดขึ้นโดย คาร์ล เฟรดริก เกาส์ (Karl Freidrich Gauss 1777-1855) และอังเดร แอนดรีวิช มาร์คอฟ (Andrei Andreevich Markov 1856-1922)[2] โดยมีหลักเกณฑ์ว่าให้หาค่าประมาณของพารามิเตอร์ที่

---

[2] ประชุม สุวัตถี, ดร., <u>ทฤษฎีการอนุมานเชิงสถิติ</u>, (กรุงเทพมหานคร : 2527) หน้า 158

ทำให้ผลบวกกำลังสองของผลต่างระหว่างค่าสังเกตกับค่าประมาณมีค่าต่ำสุด ในกรณีที่ข้อมูลเป็นไปตามข้อตกลงของการวิเคราะห์ความถดถอย ตัวประมาณพารามิเตอร์โดยวิธีการกำลังสองต่ำสุดจะเป็นตัวประมาณเชิงเส้นที่ดีที่สุดและไม่เอนเอียง (Best Linear Unbiased Estimator, BLUE)

สำหรับในการศึกษาครั้งนี้ ในการประมาณค่าพารามิเตอร์โดยใช้วิธีการกำลังสองต่ำสุด จะถือว่าข้อมูลที่ขาดหายเป็นข้อมูลที่ไม่ขาดหาย นั่นคือ จะทำการวิเคราะห์ข้อมูลรวมทั้งข้อมูลที่ขาดหาย และข้อมูลไม่ขาดหาย

หลักเกณฑ์ในการหาตัวประมาณพารามิเตอร์และคุณสมบัติของตัวประมาณมีดังต่อไปนี้

2.2.1.1 <u>การหาตัวประมาณกำลังสองต่ำสุด</u>

จากสมการความสัมพันธ์จริงระหว่างตัวแปรตาม Y และตัวแปรอิสระ X คือ

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

ให้ $\hat{\underset{\sim}{\beta}} = (\hat{\alpha}, \hat{\beta}_1, \ldots, \hat{\beta}_k)'$ เป็นเวคเตอร์ของตัวประมาณค่าของพารามิเตอร์ จะได้ความสัมพันธ์คาดหมายคือ

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

เมื่อ $\underset{\sim}{e}$ คือค่าความคลาดเคลื่อนระหว่างค่าสังเกตของ $\underset{\sim}{Y}$ กับค่าประมาณ $\hat{\underset{\sim}{Y}}$ ดังนั้น

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

พิจารณา ผลบวกกำลังสองของค่าความคลาดเคลื่อน (Sum of Squared Error) จะพบว่า

$$\underset{\sim}{e}'\underset{\sim}{e} = (\underset{\sim}{Y} - X\hat{\underset{\sim}{\beta}})'(\underset{\sim}{Y} - X\hat{\underset{\sim}{\beta}})$$

$$= (\underset{\sim}{Y}' - \hat{\underset{\sim}{\beta}}'X')(\underset{\sim}{Y} - X\hat{\underset{\sim}{\beta}})$$

$$= \underset{\sim}{Y}'\underset{\sim}{Y} - 2\hat{\underset{\sim}{\beta}}'X'\underset{\sim}{Y} + \hat{\underset{\sim}{\beta}}'X'X\hat{\underset{\sim}{\beta}}$$

ตัวประมาณกำลังสองต่ำสุด คือตัวประมาณที่ได้จากการทำให้ผลบวกกำลังสองของค่าความคลาดเคลื่อน หรือ $\underset{\sim}{e}'\underset{\sim}{e}$ มีค่าต่ำสุด

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

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

2.2.1.2 <u>คุณสมบัติของตัวประมาณกำลังสองต่ำสุด</u>

2.2.1.2.1 <u>ความไม่เอนเอียง</u> (Unbiasedness)

$\hat{\underset{\sim}{\beta}} = (\hat{\alpha}, \hat{\beta}_1, \ldots, \hat{\beta}_k)$ เป็นตัวประมาณที่ไม่เอนเอียงของพารามิเตอร์ $\underset{\sim}{\beta} = (\alpha, \beta_1, \ldots, \beta_k)$ นั่นคือ $E(\hat{\underset{\sim}{\beta}}) = \underset{\sim}{\beta}$

การพิสูจน์ $E(\hat{\underset{\sim}{\beta}}) = \underset{\sim}{\beta}$ แสดงได้ดังนี้

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

$$= (X'X)^{-1}X'(X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon})$$

$$= (X'X)^{-1}(X'X)\underset{\sim}{\beta} + (X'X)^{-1}X'\underset{\sim}{\varepsilon}$$

$$= \underset{\sim}{\beta} + (X'X)^{-1}X'\underset{\sim}{\varepsilon}$$

$$E(\hat{\underset{\sim}{\beta}}) = \underset{\sim}{\beta} + (X'X)^{-1}X'E(\underset{\sim}{\varepsilon})$$

แต่ $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$ ตามข้อตกลงของการวิเคราะห์ความถดถอย

ดังนั้น $E(\hat{\underset{\sim}{\beta}}) = \underset{\sim}{\beta}$

2.2.1.2.2 <u>ความแปรปรวนของ $\hat{\boldsymbol{\beta}}$</u> $\hat{\boldsymbol{\beta}} = (\hat{\alpha}, \hat{\beta}_1, \ldots, \hat{\beta}_k)$ เป็นตัวประมาณที่มีความแปรปรวนต่ำสุด (Minimum Variance) หรือ $\hat{\boldsymbol{\beta}}$ เป็นตัวประมาณค่าที่ดีที่สุด (Best Estimator)

ความแปรปรวนของ $\hat{\boldsymbol{\beta}}$ คือ $V(\hat{\boldsymbol{\beta}}) = \sigma^2 (X\ X)^{-1}$

การพิสูจน์ $V(\hat{\boldsymbol{\beta}}) = \sigma^2 (X'X)^{-1}$ แสดงได้ดังนี้

$$
\begin{aligned}
V(\hat{\boldsymbol{\beta}}) &= E\{\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})\}\{\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})\}' \\
&= E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \\
&= E\{\boldsymbol{\beta} + (X'X)^{-1} X'\boldsymbol{\varepsilon} - \boldsymbol{\beta}\}\{\boldsymbol{\beta} + (X'X)^{-1} X'\boldsymbol{\varepsilon} - \boldsymbol{\beta}\}' \\
&= E\{(X'X)^{-1} X'\boldsymbol{\varepsilon}\}\{(X'X)^{-1} X'\boldsymbol{\varepsilon}\}' \\
&= (X'X)^{-1} X'E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') X (X'X)^{-1} \\
&= (X'X)^{-1} X'\sigma^2 I\, X(X'X)^{-1} \\
&= \sigma^2 (X'X)^{-1} (X'X)(X'X)^{-1} \\
&= \sigma^2 (X'X)^{-1}
\end{aligned}
$$

2.2.1.2.3 <u>การแจกแจงของ $\hat{\boldsymbol{\beta}}$</u> $\hat{\boldsymbol{\beta}}$ มีการแจกแจงดังนี้คือ $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2 (X'X)^{-1})$

พิสูจน์ เนื่องจาก $\boldsymbol{Y} \sim N(X\boldsymbol{\beta}, \sigma^2 I_n)$

และ $\hat{\boldsymbol{\beta}}$ เป็นฟังก์ชั่นเชิงเส้นของตัวแปรสุ่ม $\boldsymbol{Y}$

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

โดยอาศัยวิธีโมเมนต์เทคนิคจะได้

$$M_{\hat{\underset{\sim}{\beta}}}(t) = \exp\{t'(A\mu) + \frac{t'(A\sigma^2 I_n A')t}{2}\}$$

แต่ $\mu = E(\underset{\sim}{Y}) = X\underset{\sim}{\beta}$

ดังนั้น $$M_{\hat{\underset{\sim}{\beta}}}(t) = \exp\{t'(AX\underset{\sim}{\beta}) + \frac{t'(A\sigma^2 I_n A')t}{2}\}$$

พิจารณา $$AX\underset{\sim}{\beta} = \{(X'X)^{-1}X'\}X\underset{\sim}{\beta} = \underset{\sim}{\beta}$$

$$A\sigma^2 I_n A' = \sigma^2\{(X'X)^{-1}X'\}\{(X'X)^{-1}X'\}' = \sigma^2(X'X)^{-1}X'X(X'X)^{-1}$$

$$= \sigma^2(X'X)^{-1}$$

นั่นคือ $$M_{\hat{\underset{\sim}{\beta}}}(t) = \exp\{t'\underset{\sim}{\beta} + \frac{t'\sigma^2(X'X)^{-1}t}{2}\}$$

แสดงว่า $$\hat{\underset{\sim}{\beta}} \sim N(\underset{\sim}{\beta}, \sigma^2(X'X)^{-1})$$

2.2.2 <u>วิธีการของมิลเลอร์</u> (Miller Method)

ในกรณีไม่มีค่าขาดหาย (No Censoring) สามารถหาค่าประมาณ $\hat{\alpha}$ และ $\hat{\beta}$ จากการทำให้ค่า $\sum_{i=1}^{n}(Y_i - \alpha - \beta X_i)^2$ มีค่าต่ำสุด หรือ ก็คือการทำให้ $\sum_{i=1}^{n}(Y_i - \alpha - \beta X_i)^2 = n\int_{-\infty}^{\infty}\varepsilon^2 d\hat{F}_n(\varepsilon)$, เมื่อ $\varepsilon = Y_i - \alpha - \beta X_i$ มีค่าต่ำสุด

ในกรณีมีค่าขาดหาย (With Censoring) มิลเลอร์ได้ทำการ หาค่าประมาณ $\hat{\alpha}$ และ $\hat{\beta}$ จากการทำให้ $n\int_{-\infty}^{\infty}\varepsilon^2 d\hat{F}(\varepsilon) = \sum_{i=1}^{n}\hat{W}_i(\beta)(Y_i - \alpha - \beta X_i)^2$ เมื่อ

$\varepsilon = Y_i - \alpha - \beta X_i$ และ $\hat{F}(\varepsilon) = 1 - \hat{S}(\varepsilon)$ เมื่อ $\hat{S}(\varepsilon)$ คือ PL Estimator มีค่าต่ำสุด

ขั้นตอนการหาตัวประมาณ $\hat{\alpha}$ และ $\hat{\beta}$ สำหรับวิธีการของมิลเลอร์ มีดังนี้

1. ให้หาค่าประมาณ $\hat{\alpha}_o$ และ $\hat{\beta}_o$ เริ่มต้น โดยการใช้วิธีการกำลังสองต่ำสุด เฉพาะกับข้อมูลที่ไม่ขาดหาย

$$\hat{\beta}_o = \frac{\Sigma_{uc} Y_i (X_i - \bar{X}^{uc})}{\Sigma_{uc} (X_i - \bar{X}^{uc})^2}$$

$$\hat{\alpha}_o = \bar{Y}^{uc} - \hat{\beta}_o \bar{X}^{uc}$$

เมื่อ Σ_{uc} เป็นค่าผลบวกเฉพาะกับข้อมูลที่ไม่ขาดหาย

$\bar{X}^{uc}$ เป็นค่าเฉลี่ยของ X_i เฉพาะกับข้อมูลที่ไม่ขาดหาย

$\bar{Y}^{uc}$ เป็นค่าเฉลี่ยของ Y_i เฉพาะกับข้อมูลที่ไม่ขาดหาย

2. จากข้อมูลทั้งหมด คือ รวมทั้งข้อมูลที่ขาดหาย และข้อมูลไม่ขาดหาย ให้หาค่าความคลาดเคลื่อน e จาก

$$e_i = Y_i - \hat{\beta}_o X_i \quad ; \; i = 1, \ldots, n$$

3. จากขั้นตอนที่ 2 ให้เรียงลำดับที่ (Rank) ของค่าคลาดเคลื่อน e จะได้ $e_{(1)}, e_{(2)}, \ldots, e_{(n)}$ ในกรณีที่ลำดับที่ของค่าคลาดเคลื่อนของค่าขาดหาย และค่าที่ไม่ขาดหายมีค่าเท่ากัน ให้ลำดับที่ของค่าที่ไม่ขาดหาย นำหน้า ค่าขาดหาย

4. ให้เปลี่ยนค่าลำดับที่ของค่าขาดหายเป็น 0 ส่วนค่าลำดับที่ของค่าที่ไม่ขาดหายให้คงไว้

5. หาค่า PL Estimator $\hat{S}$ จาก

$$\hat{S}(\varepsilon) = \prod_{e_{(i)} \leq e_{(i+1)}}^{n} \frac{(n - i)}{(n - i + 1)}$$

เมื่อ i คือ ค่าลำดับที่ของค่าคลาดเคลื่อน

n คือ จำนวนตัวอย่างทั้งหมด

ซึ่งค่า PL Estimator $\hat{S}$ นี้ก็คือ ค่าของ Survival Time

6. หาค่าฟังก์ชั่น $\hat{F}(\varepsilon)$ จาก

$$\hat{F}(\varepsilon) = 1 - \prod_{e_{(i)} \leq e_{(i+1)}}^{n} \frac{n-i}{n-i+1}$$

$$= 1 - \hat{S}(\varepsilon)$$

7. หาค่าตัวถ่วงน้ำหนัก $W(\hat{\beta}_o)$ จาก

$$W_1(\hat{\beta}_o) = \hat{F}_1$$

$$W_2(\hat{\beta}_o) = \hat{F}_2 - \hat{F}_1$$

$$\vdots$$

$$W_n(\hat{\beta}_o) = \hat{F}_n - \hat{F}_{n-1}$$

ในกรณีที่ลำดับที่สูงสุดของค่าความคลาดเคลื่อนเป็นค่าของข้อมูลที่ขาดหาย ให้ปรับตัวถ่วงน้ำหนัก $W(\hat{\beta}_o)$ ใหม่ ดังนี้

$$W_i^*(\hat{\beta}_o) = \frac{W_i(\hat{\beta}_o)}{\sum_{uc} W_j(\hat{\beta}_o)}$$

เมื่อ Σ_{uc} เป็นค่าผลบวกเฉพาะกับข้อมูลที่ไม่ขาดหาย

8. ประมาณค่าพารามิเตอร์ β จาก

$$\hat{\beta} = \frac{\Sigma_{uc} W_i(\hat{\beta}_o) Y_i (X_i - \bar{X}^*)}{\Sigma_{uc} W_i(\hat{\beta}_o)(X_i - \bar{X}^*)^2}$$

เมื่อ $\bar{X}^* = \Sigma_{uc} W_i(\hat{\beta}_o) X_i$

9. แทนค่า $\hat{\beta}$ ที่ได้จากขั้นตอนที่ 8 ลงในขั้นตอนที่ 2 แล้วทำการวนซ้ำ จากขั้นตอนที่ 2 จนถึงขั้นตอนที่ 8 ทำไปเรื่อย ๆ จนกระทั่งค่าของ $\hat{\beta}$ ที่ได้เท่ากับค่าของ $\hat{\beta}$ ในรอบที่แล้วก็จะหยุด ก็จะได้ค่าประมาณของ $\hat{\beta}$

ในบางครั้งค่าของ $\hat{\beta}$ จะแกว่งอยู่ระหว่าง 2 ค่า ในกรณีนี้จะใช้ค่าเฉลี่ยระหว่าง 2 ค่านั้นเป็นค่าประมาณของ $\hat{\beta}$

10. ประมาณค่า $\hat{\alpha}$ จาก

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

เมื่อ $\bar{Y} = \Sigma_{uc} W_i(\hat{\beta}_o) Y_i$

$$\bar{X} = \Sigma_{uc} W_i(\hat{\beta}_o) X_i$$

ค่าประมาณความแปรปรวนของพารามิเตอร์ $\hat{\beta}$ สำหรับวิธีการของมิลเลอร์ คือ

$$V(\hat{\beta}) = \frac{\Sigma_{uc} \{W_i(\hat{\beta})\}^2 (Y_i - \hat{\alpha} - \hat{\beta}X_i)^2}{\Sigma_{uc} W_i(\hat{\beta})(X_i - \bar{X}^*)^2}$$

เมื่อ $\bar{X}^* = \Sigma W_i(\hat{\beta}) X_i$

2.2.3 <u>วิธีการของบัคเลย์และเจมส์</u> (Buckley & James Method)

บัคเลย์และเจมส์ ได้ทำการ ประมาณค่าของข้อมูลที่ขาดหายดังนี้

$$Y_i^* = Y_i\delta_i + E(Y_i/Y_i > t_i)(1-\delta_i) \;;\; i = 1,\ldots, n$$

เมื่อ $\delta_i = \begin{cases} 1 \;; & \text{ข้อมูลที่ไม่ขาดหาย} \\ 0 \;; & \text{ข้อมูลที่ขาดหาย} \end{cases}$

แต่เนื่องจากค่าของ $E(Y_i/Y_i > t_i)$ ไม่อาจทราบค่าได้ ดังนั้นจะประมาณค่าด้วย PL Estimator ดังนั้นข้อมูลที่ขาดหายจะถูกแทนด้วย

$$\bar{Y}_i(\hat{\beta}_o) = \frac{\hat{\beta}_o X_i + \Sigma_{uc} W(\hat{\beta})(Y_k - \hat{\beta}_o X_k)}{1-F(C-\hat{\beta}_o X)}$$

ขั้นตอนการหาตัวประมาณ $\hat{\alpha}$ และ $\hat{\beta}$ สำหรับวิธีการของบัคเลย์และเจมส์มีดังนี้

1. ให้หาค่าประมาณ $\hat{\alpha}_o$ และ $\hat{\beta}_o$ เริ่มต้น โดยการใช้วิธีการกำลังสองต่ำสุด เฉพาะกับข้อมูลที่ไม่ขาดหาย

$$\hat{\beta}_o = \frac{\Sigma_{uc} Y_i(X_i - \bar{X}^{uc})}{\Sigma_{uc}(X_i - \bar{X}^{uc})^2}$$

$$\hat{\alpha}_o = \bar{Y}^{uc} - \hat{\beta}_o \bar{X}^{uc}$$

เมื่อ Σ_{uc} เป็นค่าผลบวกเฉพาะข้อมูลที่ไม่ขาดหาย

$\bar{X}^{uc}$ เป็นค่าเฉลี่ยของ X_i เฉพาะกับข้อมูลที่ไม่ขาดหาย

$\bar{Y}^{uc}$ เป็นค่าเฉลี่ยของ Y_i เฉพาะกับข้อมูลที่ไม่ขาดหาย

2. จากข้อมูลทั้งหมด คือ รวมทั้งข้อมูลที่ขาดหาย และข้อมูลที่ไม่ขาดหาย ให้หาค่าความคลาดเคลื่อน e จาก

$$e_i = Y_i - \hat{\beta}_o X_i \qquad ; \; i = 1,\ldots, n$$

3. จากขั้นตอนที่ 2 ให้เรียงลำดับที่ (Rank) ของค่าคลาดเคลื่อน e จะได้ $e_{(1)}, e_{(2)}, \ldots, e_{(n)}$ ในกรณีที่ลำดับที่ของค่าความคลาดเคลื่อนของค่าขาดหายและค่าที่ไม่ขาดหายมีค่าเท่ากัน ให้ลำดับที่ของค่าที่ไม่ขาดหายนำหน้าค่าขาดหาย

4. ให้เปลี่ยนค่าลำดับที่ของค่าขาดหายเป็น 0 ส่วนค่าลำดับที่ของค่าที่ไม่ขาดหาย ให้คงไว้

5. หาค่า PL Estimator $\hat{S}$ จาก

$$\hat{S}(\varepsilon) = \prod_{e_{(i)} \leqslant e_{(i+1)}}^{n} \frac{(n-i)}{(n-i+1)}$$

เมื่อ i คือ ค่าลำดับที่ของค่าคลาดเคลื่อน

n คือ จำนวนตัวอย่างทั้งหมด

ซึ่งค่า PL Estimator $\hat{S}$ นี้ คือ ค่าของ Survival Time

6. หาค่าฟังก์ชั่น $\hat{F}(\varepsilon)$ จาก

$$\hat{F}(\varepsilon) = 1 - \prod_{e_{(i)} \leqslant e_{(i+1)}}^{n} \frac{n-i}{n-i+1}$$

7. หาค่าตัวถ่วงน้ำหนัก $W(\hat{\beta}_o)$ จาก

$$W_1(\hat{\beta}_o) = \hat{F}_1$$

$$W_2(\hat{\beta}_o) = \hat{F}_2 - \hat{F}_1$$

$$\vdots$$

$$W_n(\hat{\beta}_o) = \hat{F}_n - \hat{F}_{n-1}$$

ในกรณีที่ลำดับที่สูงสุดของความคลาดเคลื่อนเป็นค่าขาดหาย ให้ปรับตัวถ่วงน้ำหนัก $W(\hat{\beta}_o)$ ใหม่ดังนี้

$$W_i^*(\hat{\beta}_o) = \frac{W_i(\hat{\beta}_o)}{\Sigma_{uc} W_j(\hat{\beta}_o)}$$

เมื่อ Σ_{uc} เป็นค่าผลบวกเฉพาะกับข้อมูลที่ไม่ขาดหาย

8. ให้เปลี่ยนค่าลำดับที่ของค่าที่ไม่ขาดหายเป็น 0 ล่วนค่าลำดับที่ของค่าที่ขาดหายให้คงไว้

9. หาค่า PL Estimator $\hat{S}$

$$\hat{S}(\varepsilon) = \prod_{e_{(i)} \leqslant e_{(i+1)}}^{n} \frac{(n-i)}{(n-i+1)}$$

ซึ่งค่า PL Estimator $\hat{S}$ นี้ คือค่าของ Censored Time

10. หาค่าฟังก์ชั่น $\hat{F}(\varepsilon)$ จาก

$$\hat{F}(\varepsilon) = 1 - \prod_{e_{(i)} \leqslant e_{(i+1)}}^{n} \frac{n-i}{n-i+1}$$

$$= 1 - \hat{S}(\varepsilon)$$

11. หาค่าประมาณสำหรับค่าที่ขาดหายจาก

$$\bar{Y}(\hat{\beta}_o) = \hat{\beta}_o X_i + \frac{\Sigma_{uc} W(\hat{\beta}).(Y_k - \hat{\beta}_o X_k)}{1-F(C-\hat{\beta}_o X)}$$

12. คำนวณค่า $\hat{\beta}$ จาก

$$\hat{\beta} = \frac{\left\{ \Sigma_{uc} Y_i (X_i - \bar{X}) + \Sigma_c \bar{Y}(\hat{\beta}_o)(X_i - \bar{X}) \right\}}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

เมื่อ Σ_{uc} เป็นค่าผลบวกเฉพาะกับค่าที่ไม่ขาดหาย

Σ_c เป็นค่าผลบวกเฉพาะกับค่าที่หาดขาย

$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$; n คือ จำนวนตัวอย่างทั้งหมด

13. แทนค่า $\hat{\beta}$ ที่ได้จากขั้นตอนที่ 11 ลงในขั้นตอนที่ 2 แล้วทำการวนซ้ำจากขั้นตอนที่ 2 จนถึงขั้นตอนที่ 11 ทำไปเรื่อย ๆ จนกระทั่งค่าของ $\hat{\beta}$ ที่ได้เท่ากับค่าของ $\hat{\beta}$ ในรอบที่แล้วก็จะหยุด ก็จะได้ค่าประมาณของ $\hat{\beta}$

ในบางครั้งค่าของ $\hat{\beta}$ จะแกว่งอยู่ระหว่างค่า 2 ค่า ในกรณีนี้จะใช้ค่าเฉลี่ยระหว่าง 2 ค่านั้น เป็นค่าประมาณของ $\hat{\beta}$

14. คำนวณค่า $\hat{\alpha}$ จาก

$$\hat{\alpha} = \frac{1}{n}\left\{\Sigma_{uc} Y_i + \Sigma_c \bar{Y}\ (\hat{\beta})\right\} - \hat{\beta}\bar{X}$$

ค่าประมาณความแปรปรวนของพารามิเตอร์ β สำหรับวิธีการของบัคเลย์และเจมส์คือ

$$V(\hat{\beta}) = \frac{\sigma_u^2}{\Sigma_{uc}\ (X_i - \bar{X}_u)^2}$$

เมื่อ $$\sigma_u^2 = \frac{1}{n_u-2}\Sigma_{uc}\ (Y_i - \bar{Y}_{uc} - \hat{\beta}(X_i - \bar{X}_{uc}))^2$$

เมื่อ $\bar{X}_{uc}$ เป็นค่าเฉลี่ยของ X เฉพาะกับข้อมูลที่ไม่ขาดหาย

$\bar{Y}_{uc}$ เป็นค่าเฉลี่ยของ Y เฉพาะกับข้อมูลที่ไม่ขาดหาย

n_u จำนวนข้อมูลที่ไม่ขาดหาย

สำหรับในการศึกษาครั้งนี้ เพื่อให้เครื่องคอมพิวเตอร์ลดเวลาในการคำนวณ ได้ทำการวนซ้ำ 15 รอบ ทั้งวิธีการของมิลเลอร์ และวิธีการของบัคเลย์และเจมส์