บทที่ 2

วิธีที่เกี่ยวข้องกับการวิจัย

1. <u>วิธีกำลังสองน้อยที่สุด</u> (Least-Squares Method)

โดยทั่วไปเมื่อมีข้อมูลอยู่อย่างครบถ้วน วิธีกำลังสองน้อยที่สุดในการประมาณสัมประสิทธิ์ของการถดถอยโดยทำให้ผลบวกของกำลังสองของความคลาดเคลื่อน มีค่าน้อยที่สุด

1.1 <u>การถดถอยเชิงเส้นอย่างง่าย</u> (Simple Linear Regression)

แบบการถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression Model)

การหาความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรตาม (Y) และตัวแปรอิสระ (X) เมื่อมีตัวแปรอิสระเพียงตัวเดียว จะเขียนความสัมพันธ์ระหว่างตัวแปรทั้งสองในรูปแบบ (Model)

$$Y = \beta_0 + \beta_1 X + \epsilon$$

เมื่อกำหนดให้ Y เป็นตัวแปรตาม (Dependent Variable)

X เป็นตัวแปรอิสระ (Independent Variable)

β_0 เป็นระยะระหว่างจุดกำเนิดกับจุดที่เส้นถดถอยตัดกับแกน Y (Y-intercept)

β_1 เป็นค่าสัมประสิทธิ์แห่งการถดถอยของประชากร (Population Regression Coefficient) ซึ่งเป็นพารามิเตอร์ที่ยังไม่ทราบค่า

และ ϵ เป็นค่าความแตกต่างระหว่างเส้นถดถอยกับค่าจริงของ Y ซึ่งเป็นตัวแปรสุ่มที่ไม่ทราบค่า

$$E(\epsilon_i) = 0 \; ; \; i = 1, 2, \ldots\ldots\ldots n$$

$$E(\epsilon_i \epsilon_j) = \sigma^2 ; \; i = j = 1, 2, \ldots n$$

$$= 0 \; ; \; i \neq j$$

$$E(Y_i) = \beta_0 + \beta_1 X_i \; , \; i = 1, 2, \ldots n$$

$$\text{var}(Y_i) = \sigma^2 \; , \; i = 1, 2, \ldots\ldots\ldots n$$

<u>การประมาณค่าพารามิเตอร์</u>

จากตัวแบบ $Y = \beta_0 + \beta_1 X + \epsilon$

โดยวิธีกำลังสองน้อยที่สุดนี้ จะต้องทำให้ผลบวกของกำลังที่สองของความคลาดเคลื่อน (Error sum of squares)

$$\sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2 \qquad (2.1.1)$$

มีค่าน้อยที่สุด เมื่อให้ b_0^{LS} และ b_1^{LS} เป็นค่าประมาณของ β_0 และ β_1 ตามลำดับ เราอาจจะหาค่า b_0^{LS} และ b_1^{LS} โดยใช้วิธีการทางคณิตศาสตร์ จะได้สมการปกติ (Normal Equations) เป็น

$$n\, b_0^{LS} + b_1^{LS} \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i \qquad (2.1.2)$$

$$b_0^{LS} \sum_{i=1}^{n} X_i + b_1^{LS} \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i Y_i \qquad (2.1.3)$$

เมื่อแก้สมการปกติจะได้

$$b_1^{LS} = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}_n \bar{Y}_n}{\sum_{i=1}^{n} X_i^2 - n\bar{X}_n^2} \qquad (2.1.4)$$

$$\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n} \; , \quad \bar{Y}_n = \frac{\sum_{i=1}^{n} Y_i}{n}$$

และ $$b_0^{LS} = \bar{Y}_n - b_1^{LS} \bar{X}_n \qquad (2.1.5)$$

$\therefore$ สมการถดถอยที่ได้ คือ $\hat{Y}_i^{LS} = b_o^{LS} + b_i^{LS}X_i$ (2.1.6)

หรือ $\hat{Y}_i^{LS} = \bar{Y} + b_1^{LS}(X_i - \bar{X})$

โดยมี $E(b_1^{LS}) = \beta_1$, $E(b_o^{LS}) = \beta_o$,

$$\mathrm{var}(b_1^{LS}) = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$$

และ $$\mathrm{var}(b_o^{LS}) = \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]\sigma^2 = \frac{\sum_{i=1}^{n} X_i^2}{n\sum_{i=1}^{n}(X_i - \bar{X})^2}\,\sigma^2$$

จาก $\hat{Y}_i^{LS} = \bar{Y}_n + b_1^{LS}(X_i - \bar{X}_n)$ ✓

$$\therefore \quad \mathrm{var}(\hat{Y}_i^{LS}) = \mathrm{var}(\bar{Y}_n) + (X_i - \bar{X}_n)^2\mathrm{var}(b_1^{LS}) + 2(X_i - \bar{X}_n)\mathrm{cov}(\bar{Y}_n , b_1^{LS})$$

$$= \frac{\mathrm{var}(Y_i)}{n} + (X_i - \bar{X}_n)^2 \frac{\mathrm{var}(Y_i)}{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$$

$$= \text{var}\ (Y_i)\frac{1}{n} + \frac{(X_i-\bar{X}_n)^2}{\sum_{i=1}^{n}(X_i-\bar{X}_n)^2}$$

$$= \frac{\sigma^2}{n} + \frac{(X_i-\bar{X}_n)^2}{\sum_{i=1}^{n}(X_i-\bar{X}_n)^2}$$

เมื่อพิจารณาความคลาดเคลื่อน(ϵ_i) จะได้ $\epsilon_i = Y_i - (\beta_0 - \beta_1 X_i)$

$$= (Y_i - \hat{Y}_i)$$

$$\therefore\ Y_i - \hat{Y}_i = (Y_i - \bar{Y}_n) - (\hat{Y}_i - \bar{Y}_n)$$

จะเห็นได้ว่า ความคลาดเคลื่อนเกิดจากการที่ค่าจริงเบี่ยงเบนไปจากค่าเฉลี่ย และอีกส่วนหนึ่งเกิดจากค่าประมาณเบี่ยงเบนจากค่าเฉลี่ย

จาก $\epsilon_i = (Y_i - \bar{Y}_n) - (\hat{Y}_i - \bar{Y}_n)$

$$\sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} \left[(Y_i-\bar{Y}_n)-(\hat{Y}_i-\bar{Y}_n)\right]^2$$

$$\therefore \sum_{i=1}^{n} (Y_i-\hat{Y}_i)^2 = \sum_{i=1}^{n}(Y_i-\bar{Y}_n)^2 + \sum_{i=1}^{n}(\hat{Y}_i-\bar{Y}_n)^2$$

$$- 2\sum_{i=1}^{n}(Y_i-\bar{Y}_n)(\hat{Y}_i-\bar{Y}_n) \qquad (2.1.7)$$

พิจารณาเทอม $\sum_{i=1}^{n}(Y_i-\bar{Y}_n)(\hat{Y}_i-\bar{Y}_n)$

แทนค่า $\hat{Y}_i$ ด้วย $\bar{Y}_n+b_1^{LS}(X_i-\bar{X}_n)$ ซึ่งหาได้จากสมการ (2.1.8)

$$\therefore \sum_{i=1}^{n}(Y_i-\bar{Y}_n)(\hat{Y}_i-\bar{Y}_n) = \sum_{i=1}^{n}(Y_i-\bar{Y}_n) \left[\bar{Y}_n+b_1^{LS}(X_i-\bar{X}_n)-\bar{Y}_n\right]$$

$$= b_1^{LS}\sum_{i=1}^{n}(Y_i-\bar{Y}_n)(X_i-\bar{X}_n)$$

$$= (b_1^{LS})^2\sum_{i=1}^{n}(X_i-\bar{X}_n)^2$$

$$= \sum_{i=1}^{n}\left[b_1^{LS}(X_i-\bar{X}_n)\right]^2$$

$$= \sum_{i=1}^{n}(\hat{Y}_i-\bar{Y}_n)^2;$$

$$\left[\hat{Y}_i = \bar{Y}_n + b_1^{LS}(X_i-\bar{X}_n)\right]$$

แทนค่า $\sum_{i=1}^{n}(Y_i-\bar{Y}_n)(\hat{Y}_i-\bar{Y}_n)$ ในสมการ (2.1.7) จะได้

$$\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2 = \sum_{i=1}^{n}(Y_i-\bar{Y}_n)^2-\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y}_n)^2$$

$$\therefore \quad \sum_{i=1}^{n}(Y_i-\bar{Y}_n)^2 = \sum_{i=1}^{n}(\hat{Y}_i-\bar{Y}_n)^2+\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2 \qquad (2.1.8)$$

จะเห็นว่า $\sum_{i=1}^{n}(Y_i-\bar{Y}_n)^2$ ซึ่งเรียกว่าผลบวกกำลังสองของทั้งหมด (Total Sum of Squares) จะประกอบด้วย

ส่วนที่ 1 $\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y}_n)^2$ ซึ่งเรียกว่าผลบวกกำลังสองเนื่องมาจากการถดถอย (Sum of Squares due to Regression)

และ ส่วนที่ 2 $\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2$ ซึ่งเรียกว่าผลบวกกำลังสองของความคลาดเคลื่อน (Error Sum of Squares)

ถ้าให้ $R^2 = \dfrac{\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y}_n)^2}{\sum_{i=1}^{n}(Y_i-\bar{Y}_n)^2} = \dfrac{\text{ผลบวกกำลังสองเนื่องจากการถดถอย}}{\text{ผลบวกกำลังสองของทั้งหมด}}$

ค่า R^2 จะมีค่าอยู่ระหว่าง 0 ถึง 1

ถ้าค่า R^2 ใกล้ศูนย์หมายถึงจุดต่าง ๆ กระจายห่างจาก $\hat{Y}_i$ มาก ค่าของผลบวกกำลังสองเนื่องจากการถดถอยมีค่าต่ำมาก แต่ถ้าค่า $R^2 \to -1$ จนถึง 1 หมายถึงสมการถดถอยที่ได้จะเป็นตัวอธิบายความเปลี่ยนแปลงทั้งหมดของ Y เพราะผลบวกกำลังสองของทั้งหมด จะมีค่าเท่ากับผลบวกกำลังสองเนื่องจากการถดถอย

การทดสอบสมมติฐานโดยใช้การทดสอบ F (F-test)

จากสมการถดถอยที่ได้ $\hat{Y}_i = b_o^{LS} + b_1^{LS}X_i$

เมื่อหาสมการถดถอยได้แล้ว เราก็ต้องทดสอบดูว่าสมการถดถอยที่ได้จะเหมาะสมหรือไม่ คือจะทดสอบว่ามีการถดถอยหรือไม่ โดยตั้งสมมติฐานว่า

$H_o: \beta_1 = 0$, $H_A: \beta_1 \neq 0$ โดยมีข้อสมมติว่า ϵ_i มีการแจกแจงแบบปกติ มีค่าเฉลี่ยเป็น 0 และมีค่าความแปรปรวนเป็น σ^2

ตารางวิเคราะห์ความแปรปรวน (Table of Analysis of Variance) สำหรับทดสอบสมมติฐานเรียกย่อ ๆ ว่า ตาราง ANOVA

ตารางที่ 1

ANOVA

Source of Variation	df	Sum of Squares	Mean Squares	F
Due to Regression	1	$\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y}_n)^2 =$ SSR	SSR/1 =MSR	$\frac{MSR}{MSE}$
Error	n-2	$\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2 =$ SSE	SSE/(n-2)=MSE	
Total	n-1	$\sum_{i=1}^{n}(Y_i-\bar{Y}_n)^2 =$ SST		

นำค่า $F = \frac{MSR}{MSE}$ ที่คำนวณได้ มาเปรียบเทียบกับค่า F จากตาราง ที่ชั้นแห่งความเป็นอิสระ (d.f) = (1, n-2)

ถ้า $F_{\text{คำนวณ}} > F_{\text{ตาราง}}$ เราจะไม่ยอมรับสมมติฐานที่ว่า $\beta_1 = 0$ นั่นคือ $\beta_1 \neq 0$ แสดงว่ามีการถดถอย

ถ้า $F_{\text{คำนวณ}} < F_{\text{ตาราง}}$ เราจะยอมรับสมมติฐานที่ว่า $\beta_1 = 0$ นั่นคือ $\beta_1 = 0$ แสดงว่าไม่มีการถดถอย

1.2 การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression)

ในการวิเคราะห์ปัญหาบางอย่างที่จำเป็นต้องใช้การถดถอยนั้น บางครั้งการศึกษาถึงการถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression) อาจจะไม่เป็นการเพียงพอ ทั้งนี้เพราะการประมาณค่าของ Y ให้ได้ใกล้เคียงที่สุดนั้น เรามักจะต้องพิจารณาถึงตัวแปรอิสระ (X) ที่มีอิทธิพลหรือมีความสำคัญต่อ Y มากกว่าหนึ่งตัวขึ้นไป ดังนั้นเราจะใช้การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression) ซึ่งเป็นการหาความสัมพันธ์ระหว่างข้อมูลตั้งแต่ 2 ชุดขึ้นไป เมื่อข้อมูลชุดหนึ่งเป็นตัวแปรตาม (Y) ส่วนข้อมูลชุดอื่นเป็นตัวแปรอิสระ (X) แล้วศึกษาความสัมพันธ์ของข้อมูลเหล่านี้ในรูปสมการ

$$Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \dots\dots\dots + \beta_p X_p + \epsilon \qquad (2.1.9)$$

โดยที่ Y เป็นตัวแปรตาม (Dependent Variable)

X_i เป็นตัวแปรอิสระ (Independent variable) ตัวที่ i, $i = 1,2,\dots\dots\dots p$

β_o เป็นค่าคงที่

β_i เป็นค่าสัมประสิทธิ์แห่งการถดถอยของประชากรบางส่วน (Population Partial Regression Coefficient) ของ Y ต่อ X_i เมื่อให้ $X_1, X_2, \dots X_{i-1}, X_{i+1} \dots\dots X_p$ คงที่ นั่นคือ เมื่อ X_i เปลี่ยนแปลง 1 หน่วย จะมีผลทำให้ Y เปลี่ยนแปลง β_i หน่วย เมื่อตัวแปรอิสระตัวอื่นคงที่ , $i = 1,2,\dots\dots p$

และ ϵ เป็นค่าความแตกต่างระหว่างเส้นถดถอยกับค่าจริง ซึ่งเป็นตัวแปรสุ่มที่ไม่ทราบค่า

$E(\epsilon_i) = 0, \quad i = 1, 2, \ldots\ldots n$

$E(\epsilon_i \epsilon_j) = \sigma^2; \; i = j = 1, 2, \ldots\ldots n$

$= 0; \quad i \neq j$

ค่าพารามิเตอร์ที่ไม่ทราบค่าในที่นี้ได้แก่

$\beta_i, \; i = 0, 1, 2, \ldots\ldots p$ และ σ^2

ในการประมาณค่าพารามิเตอร์ สำหรับการถดถอยพหุคูณนี้ เราจะใช้เมตริกซ์ช่วย จาก (2.1.9) เขียนตัวแบบให้อยู่ในรูปเมตริกซ์จะได้

$$Y = X\beta + \epsilon \tag{2.1.10}$$

โดยที่ Y = $n \times 1$ เวคเตอร์ที่มีคอมโปเนนเป็น $Y_1, Y_2 \ldots\ldots Y_n$

X = $n \times (p + 1)$ เมตริกซ์ที่มีสมาชิกเป็น X_{ij} ซึ่งได้แก่ค่าที่ i ของตัวแปร X_j คือ

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots\cdots & X_{1p} \\ 1 & X_{21} & X_{22} & \cdots\cdots & X_{2p} \\ 1 & X_{31} & X_{32} & \cdots\cdots & X_{3p} \\ \vdots & \vdots & \vdots & \cdots\cdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots\cdots & X_{np} \end{bmatrix}$$

β = $(p + 1)$ เวคเตอร์ที่มีคอมโปเนนเป็น $\beta_0, \beta_1, \ldots\ldots \beta_p$

$$= \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

ϵ = $n \times 1$ เวคเตอร์ของความคลาดเคลื่อน

$$= \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

ถ้าให้ $\hat{\beta}^{LS}$ เป็นค่าประมาณของ β ในการหาค่า $\hat{\beta}^{LS}$ โดยวิธีกำลังสองน้อยที่สุดนี้ จะต้องทำให้ค่าผลบวกกำลังสองของความคลาดเคลื่อนมีค่าน้อยที่สุด แล้วใช้วิธีทางคณิตศาสตร์ โดยการหาอนุพันธ์ของ $\epsilon'\epsilon$ เทียบกับ β

จากตัวแบบ $Y = X\beta + \epsilon$

$$\epsilon = (Y - X\beta)$$

$$\epsilon'\epsilon = (Y - X\beta)'(Y - X\beta)$$

$$= Y'Y - 2\beta'X'Y + \beta'X'X\beta$$

$$\frac{\partial \epsilon'\epsilon}{\partial \beta} = -2X'Y + 2X'X\beta = 0$$

แทนค่า β ด้วย $\hat{\beta}^{LS}$

$$\therefore X'X\hat{\beta}^{LS} = X'Y$$

ถ้า $X'X$ เป็นเมตริกซ์แบบนอนซิงกูลาร์ จะได้

$$\hat{\beta}^{LS} = (X'X)^{-1}X'Y \qquad (2.1.11)$$

$\therefore$ สมการถดถอยที่ได้ คือ $\hat{Y}^{LS} = X\hat{\beta}^{LS}$ (2.1.12)

โดยที่ $E(\hat{\beta}^{LS}) = \beta$ และ

$$\text{var}(\hat{\beta}^{LS}) = (X'X)^{-1}\sigma^2$$

เมื่อจะทดสอบสมมุติฐาน $H_o : \beta = 0$; $H_A : \beta \neq 0$ โดยใช้การทดสอบ F (F-test) เราใช้ตาราง ANOVA ต่อไปนี้

ตารางที่ 2

ANOVA

SOV	d.f	SS	MS	F
Regression	p+1	$\hat{\beta}^{LS'}X'\hat{Y}^{LS}$ = SSR	$\frac{SSR}{p+1}$ = MSR	$\frac{MSR}{MSE}$
Error	n-p-1	$\hat{Y}^{LS'}\hat{Y}^{LS} - \hat{\beta}^{LS'}X'\hat{Y}^{LS}$ = SSE	$\frac{SSE}{n-p-1}$ = MSE	
Total (uncorrected)	n	$\hat{Y}^{LS'}\hat{Y}^{LS}$		

2. วิธีการประมาณค่าของตัวแปรที่ขาดหายไปโดยวิธีกำลังสองต่ำสุด (The Classical Least-Squares Method)[1]

การประมาณค่าสังเกตของตัวแปรที่ขาดหายไปในการวิเคราะห์การถดถอยโดยวิธีนี้ ใช้ค่าสังเกตที่มีครบทั้งตัวแปรตาม (Y) และตัวแปรอิสระ (X) หาสมการถดถอยโดยวิธีกำลังสองน้อยที่สุด และใช้สมการถดถอยที่ได้นี้ประมาณค่าสังเกตของตัวแปรตาม (Y) ที่ขาดหายไป แล้วจึงนำข้อมูลที่ได้ใหม่ไปหาสมการถดถอย

2.1 การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression)

จากตัวแบบ $Y = \beta_o + \beta_1 X + \epsilon$

ให้ $b_o^{(1)}$ และ $b_1^{(1)}$ เป็นค่าประมาณของ β_o และ β_1 จะได้สมการถดถอยที่ต้องการ คือ

$$\hat{Y}_i^{(1)} = b_o^{(1)} + b_1^{(1)} X_i$$

เมื่อกำหนดให้ $n = (n_c+m_x+m_y)$ = จำนวนหน่วยตัวอย่างทั้งหมด

n_c = จำนวนหน่วยตัวอย่างที่มีทั้งค่าของ X และ Y

$m_x = (n-n_x)$ = จำนวนหน่วยตัวอย่างที่ค่าของ X ขาดหายไป

$m_y = (n-n_y)$ = จำนวนหน่วยตัวอย่างที่ค่าของ Y ขาดหายไป

n_x จำนวนหน่วยตัวอย่างที่มีค่าของ X ทั้งหมด

n_y จำนวนหน่วยตัวอย่างที่มีค่าของ Y ทั้งหมด

---

[1] A.A. Afifi and R.M. Elashoff, "Missing Obsevation in Multivariate Statistics I. Review of the Literature," <u>Journal of the American Statistical Association</u>, 61 (1966), pp. 595-597.

จากข้อมูลที่รวบรวมมาได้ จะหาค่า $b_1'^{LS}$ และ $b_o'^{LS}$ ได้เช่นเดียวกับ b_1^{LS} และ b_o^{LS} ซึ่งหาได้จาก (2.1.4) และ (2.1.5) ตามลำดับ จะได้

003627

$$b_1'^{LS} = \frac{\sum_{i=1}^{n_c} X_iY_i - n_c\bar{X}_c\bar{Y}_c}{\sum_{i=1}^{n_c} X_i^2 - n_c\bar{X}_c^2} \quad ; \quad \bar{X}_c = \frac{\sum_{i=1}^{n_c} X_i}{n_c}$$

$$\text{และ} \quad b_o'^{LS} = \bar{Y} - b_1'^{LS}\bar{X}_c \qquad \bar{Y}_c = \frac{\sum_{i=1}^{n_c} Y_i}{n_c}$$

สมการถดถอยที่ได้ คือ $\hat{Y}'^{LS} = b_o'^{LS} + b_1'^{LS} X_i$

ประมาณค่าสังเกต Y_i ที่ขาดหายไป โดยใช้สมการถดถอยซึ่งหาได้นี้ แล้วนำค่า $\hat{Y}_i$ ที่ประมาณได้ แทนค่าสังเกต Y_i ที่ขาดหายไป ใช้วิธีกำลังสองน้อยที่สุดเหมือนเมื่อมีข้อมูลครบถ้วน หาค่า $b_1^{(1)}$ และ $b_o^{(1)}$ ซึ่งจะหาได้เช่นเดียวกับ b_1^{LS} และ b_o^{LS} ซึ่งหาได้จาก (2.1.4) และ (2.1.5) ตามลำดับ จะได้

$$b_1^{(1)} = \frac{\sum_{i=1}^{n_x} X_iY_i - n_x\bar{Y}_{n_x}\bar{X}_{n_x}}{\sum_{i=1}^{n_x} X_i^2 - n_x\bar{X}_{n_x}^2} \quad ; \quad \bar{X}_{n_x} = \frac{\sum_{i=1}^{n_x} X_i}{n_x}$$

$$\text{และ} \quad b_o^{(1)} = \bar{Y}_{n_x} - b_1^{(1)}\bar{X}_{n_x} \qquad \bar{Y}_{n_x} = \frac{\sum_{i=1}^{n_x} Y_i}{n_x}$$

สมการถดถอยที่ได้ใหม่ คือ $\hat{Y}_i^{(1)} = b_o^{(1)} + b_1^{(1)}X_i$

เราใช้สมการถดถอยที่หาได้นี้ในการพยากรณ์ต่อไป

เมื่อจะทดสอบสมมติฐาน $H_o: \beta_1 = 0;\ H_A\ \beta_1 \neq 0$ โดยใช้การทดสอบ F (F-test) เราใช้ตาราง ANOVA ต่อไปนี้

ตารางที่ 3

ANOVA

Sov	df.	SS	MS	F
Reg	1	$\sum_{i=1}^{n_x}(\hat{Y}_i - \bar{Y}_{n_x})^2 = SSR$	$SSR/1 = MRS$	$\frac{MSR}{MSE}$
Error	$n_c - 2$	$\sum_{i=1}^{n_x}(Y_i - \hat{Y}_i)^2 = SSE$	$SSE/(n_c-2) = MSE$	
Total	$n_c - 1$	$\sum_{i=1}^{n_x}(Y - \bar{Y}_{n_x})^2 = SST$		

2.2 <u>การถดถอยเชิงเส้นพหุคูณ</u> (Multiple Linear Regression)

จากตัวแบบ $Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \cdots\cdots + \beta_p X_p + \epsilon$

หรือเขียนในรูปเมตริกซ์จะได้ $Y = X\beta + \epsilon$

ให้ $b_1^{(1)}$ และ $b_i^{(1)}$ เป็นค่าประมาณของ β_o และ β_i ตามลำดับ จะได้สมการถดถอย $Y_i^{(1)} = b_o^{(1)} + b_1^{(1)} + b_1^{(1)} X_{1i} + b_2^{(1)} X_{2i} + \cdots\cdots + b_p^{(1)} X_{pi}$

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

สมมติว่าข้อมูลที่รวบรวมมาได้มีดังนี้

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix}, \quad X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_3 \end{bmatrix}$$

โดยที่ Y_1 แทนเวคเตอร์ค่าสังเกตที่มีอยู่ ซึ่งตรงกับ X_1

Y_2 แทนเวคเตอร์ค่าสังเกตที่สูญหาย ซึ่งตรงกับ X_2

Y_3 แทนเวกเตอร์ค่าสังเกตที่มีอยู่ ซึ่งตรงกับ X_3

Y_4 แทนเวคเตอร์ค่าสังเกตที่สูญหาย ซึ่งตรงกับ X_4

X_1 แทนเมตริกซ์ที่ไม่มีค่าสังเกตตัวใดตัวหนึ่งสูญหาย

X_2 แทนเมตริกซ์ที่ไม่มีค่าสังเกตตัวใดตัวหนึ่งสูญหาย

X_3 แทนเมตริกซ์ที่มีค่าสังเกตสูญหายอย่างน้อยหนึ่งตัว

X_4 แทนเมตริกซ์ที่มีค่าสังเกตสูญหายอย่างน้อยหนึ่งตัว

โดยวิธีกำลังสองน้อยที่สุด เช่นเดียวกับสมการจาก (2.1.12) จะได้สมการถดถอย คือ

$$\hat{Y}^{LS} = X_o \hat{\beta}_o^{LS} \tag{2.2.1}$$

$$X_o = \begin{bmatrix} X_1 \\ X_2 \\ 0 \end{bmatrix} \qquad Y = \begin{bmatrix} Y_1 \\ 0 \\ Y_3 \\ 0 \end{bmatrix}$$

โดยวิธีเช่นเดียวกับสมการที่ได้จาก (2.1.15) จะได้

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

ถ้า Y_2 ที่ขาดหายไปจะแทนด้วย $\hat{Y}_2$ ซึ่งจะหาได้จาก (2.2.1) และ Corrected ด้วย $\underline{C}$[1] ซึ่งเป็น Matrix factor มีค่า

$$\underline{C} = \left[I - X_2 (X'_o X_o)^{-1} X'_2\right]^{-1}$$

---

[1] Ibid., p. 597.

จะได้ $\hat{Y}_2 = c\, X_2 \beta_o^{LS}$

นำค่า Y_2 ที่ได้ ไปแทนค่า Y_2 ที่ขาดหายไป แล้วหาค่า $\hat{\beta}_o^{LS}$ ใหม่ โดยแทน Y_3, $Y_4 = 0$ และแทน $X = X_o$ และทำเหมือนกับวิธีกำลังสองน้อยที่สุด

โดยวิธีเช่นเดียวกับสมการที่ได้จาก (2.1.11) และ (2.1.12) จะได้

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

และสมการถดถอยที่ได้ คือ $\hat{Y}^{(1)} = X_o \hat{\beta}^{(1)}$

เมื่อจะทดสอบสมมติฐาน $H_o: \beta = 0$; $H_A: \beta \neq 0$ โดยใช้การทดสอบ F (F-test) เราใช้ตาราง ANOVA ต่อไปนี้

ตารางที่ 4

ANOVA

Sov	df.	SS	MS	F
Reg	$p+1$	$\hat{\beta}^{(1)\prime} X_o' \hat{Y}^{(1)} = SSR$	$\frac{SSR}{p+1} = MSR$	$\frac{MSR}{MSE}$
Error	n_c-p-1	$\hat{Y}^{(1)\prime} \hat{Y}^{(1)} - \hat{\beta}^{(1)\prime} X_o' \hat{Y}^{(1)} = SSE$	$\frac{SSE}{n_c-p-1} = MSE$	
Total (uncorrected)	n_c	$\hat{Y}^{(1)\prime} \hat{Y}^{(1)} = SST$		

ในที่นี้ n_c คือ จำนวนหน่วยตัวอย่างที่มีทั้งค่า X และ Y อยู่ครบถ้วน

3. วิธีการประมาณค่าของตัวแปรที่ขาดหายไปโดยวิธีอันดับศูนย์ (The Zero Order Method)[1]

วิธีนี้จะประมาณค่า Y ที่ขาดหายไปด้วย $\bar{Y}_{n_y}$ และประมาณค่า X ที่ขาดหายไปด้วย $\bar{X}_{n_x}$ แล้วใช้วิธีกำลังสองน้อยที่สุด เหมือนเมื่อมีข้อมูลครบถ้วน ในการประมาณค่าพารามิเตอร์

$$\text{เมื่อ} \quad \bar{Y}_{n_y} = \frac{\sum_{i=1}^{n_y} Y_i}{n_y}$$

$$\bar{X}_{n_x} = \frac{\sum_{i=1}^{n_x} X_i}{n_x}$$

3.1 การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression)

จากตัวแบบ $Y_i = \beta_0 + \beta_1 X_i + \epsilon$

ให้ $b_0^{(2)}$ และ $b_1^{(2)}$ เป็นค่าประมาณของ β_0 และ β_1 จะได้สมการถดถอยที่ต้องการ คือ

$$\hat{Y}_i^{(2)} = b_0^{(2)} + b_1^{(2)} X_i$$

เมื่อกำหนดให้ $n = (n_c + m_x + m_y)$ = จำนวนหน่วยตัวอย่างทั้งหมด

n_c = จำนวนหน่วยตัวอย่างที่มีทั้งค่าของ X และ Y

$m_x = (n - n_x)$ = จำนวนหน่วยตัวอย่างที่ค่าของ X ขาดหายไป

$m_y = (n - n_y)$ = จำนวนหน่วยตัวอย่างที่ค่าของ Y ขาดหายไป

---

[1] A.A. Afifi and R.M. Elashoff, "Missing Observation in Multivariate Statistics II. Point Estimation in Simple Linear Regression," Journal of the American Statistical Association, 62 (1967), pp. 12-15.

n_x = จำนวนหน่วยตัวอย่างที่มีค่าของ X ทั้งหมด

n_y = จำนวนหน่วยตัวอย่างที่มีค่าของ Y ทั้งหมด

ค่า Y ที่ขาดหายไปจะประมาณด้วย $\bar{Y}_{n_y} = \dfrac{\sum_{i=1}^{n_y} Y_i}{n_y}$

ส่วนค่า X ที่ขาดหายไปจะประมาณด้วย $\bar{X}_{n_x} = \dfrac{\sum_{i=1}^{n_x} X_i}{n_x}$

เมื่อนำค่า $\bar{Y}_{n_y}$ และ $\bar{X}_{n_x}$ แทนค่าสังเกต Y_i และ X_i ที่ขาดหายไปตามลำดับแล้ว ใช้วิธีกำลังสองน้อยที่สุด เหมือนเมื่อมีข้อมูลครบถ้วน จะหาค่า $b_1^{(2)}$ และ $b_o^{(2)}$ ได้เช่นเดียวกับ b_1^{LS} และ b_o^{LS} ซึ่งหาได้จาก (2.1.4) และ (2.1.5) ตามลำดับ จะได้

$$b_1^{(2)} = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\,\bar{Y}_{n_y}}{\sum_{i=1}^{n} X_i^2 - n\bar{X}_{n_x}^2}$$

และ $$b_o^{(2)} = \bar{Y}_{n_y} - b_1^{(2)}\bar{X}_{n_x}$$

สมการถดถอยที่ได้คือ $\hat{Y}_i^{(2)} = b_o^{(2)} + b_1^{(2)} X_i$

เมื่อจะทดสอบสมมติฐาน $H_o : \beta_1 = 0$; $H_A : \beta_1 \neq 0$ โดยใช้การทดสอบ F (F-test) เราใช้ตาราง ANOVA ต่อไปนี้

ตารางที่ 5

ANOVA

Sov	df.	SS	MS	F
Reg	1	$\sum_{i=1}^{n} (\hat{Y}_i^{(2)} - \bar{Y}_{n_y})^2 = SSR$	SSR/1 =MSR	$\frac{MSR}{MSE}$
Error	n_c-2	$\sum_{i=1}^{n} (Y_i - \hat{Y}_i^{(2)})^2 = SSE$	SSE/(n_c-2) =MSE	
Total	n_c-1	$\sum_{i=1}^{n} (Y_i - \bar{Y}_{n_y})^2 = SST$		

3.2 <u>การถดถอยเชิงเส้นพหุคูณ</u> (Multiple Linear Regression)

จากตัวแบบ $Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 \ldots\ldots + \beta_p X_p + \epsilon$

หรือเขียนในรูปเมตริกซ์ จะได้ $\underline{Y} = \underline{X}\underline{\beta} + \underline{\epsilon}$

ให้ $b_i^{(2)}$ เป็นค่าประมาณของ β_i, $i = 0,1,2,\ldots\ldots,p$ ตามลำดับ จะได้สมการถดถอย เป็น

$$\hat{Y}_i^{(2)} = b_o^{(2)} + b_1^{(2)} X_{1i} + b_2^{(2)} X_{2i} + \ldots + b_p^{(2)} X_{pi}$$

หรือเขียนในรูปเมตริกซ์ จะได้ $\hat{\underline{Y}}^{(2)} = \underline{X}\hat{\underline{\beta}}^{(2)}$

สมมติว่าข้อมูลที่รวบรวมมาได้มีดังนี้

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix}, \quad \underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}, \quad \underline{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}$$

โดยที่ Y_1 แทนเวคเตอร์ค่าสังเกตที่มีอยู่ ซึ่งตรงกับ X_1

Y_2 แทนเวคเตอร์ค่าสังเกตที่สูญหาย ซึ่งตรงกับ X_2

Y_3 แทนเวคเตอร์ค่าสังเกตที่มีอยู่ ซึ่งตรงกับ X_3

Y_4 แทนเวคเตอร์ค่าสังเกตที่สูญหาย ซึ่งตรงกับ X_4

X_1 แทนเมตริกซ์ที่ไม่มีค่าสังเกตตัวใดตัวหนึ่งสูญหาย

X_2 แทนเมตริกซ์ที่ไม่มีค่าสังเกตตัวใดตัวหนึ่งสูญหาย

X_3 แทนเมตริกซ์ที่มีค่าสังเกตสูญหายอย่างน้อยหนึ่งตัว

X_4 แทนเมตริกซ์ที่มีค่าสังเกตสูญหายอย่างน้อยหนึ่งตัว

ค่า Y ที่ขาดหายไปจะประมาณด้วย $\bar{Y}_{n_y} = \dfrac{\sum_{i=1}^{n_y} Y_i}{n_y}$

ส่วนค่า x_j ที่ขาดหายไปจะประมาณด้วย $\bar{X}_j = \dfrac{\sum_{i=1}^{n_{x_j}} X_{ji}}{n_{x_j}}$; $j = 1,2,\ldots.p$

นำค่า $\bar{Y}_{n_y}$ และ $\bar{X}_j$ แทนค่าสังเกต Y_i และ X_{ji} ที่ขาดหายไปตามลำดับแล้วใช้วิธีกำลังสองน้อยที่สุดเหมือนเมื่อมีข้อมูลครบ จะหาค่า $\hat{\underline{\beta}}^{(2)}$ ได้เช่นเดียวกับ $\hat{\underline{\beta}}^{LS}$ ซึ่งหาได้จาก (2.1.11) จะได้

$$\hat{\underline{\beta}}^{(2)} = (X'X)^{-1} X'Y$$

สมการถดถอยที่ได้คือ $\hat{Y}^{(2)} = X\hat{\underline{\beta}}^{(2)}$

เมื่อจะทดสอบสมมติฐาน $H_o : \underline{\beta} = 0$, $H_A : \underline{\beta} \neq 0$ โดยใช้การทดสอบ F (F-test) เราใช้ตาราง ANOVA ต่อไปนี้

ตารางที่ 6

ANOVA

Sov	d.f	SS	MS	F
Reg	$p+1$	$\hat{\beta}^{(2)'}X'\hat{Y}^{(2)} = SSR$	$\frac{SSR}{p+1} = MSR$	$\frac{MSR}{MSE}$
Error	n_c-p-1	$\hat{Y}^{(2)'}\hat{Y}^{(2)} - \hat{\beta}^{(2)'}X'\hat{Y}^{(2)} = SSE$	$\frac{SSE}{n_c-p-1} = MSE$	
Total (uncor-rected)	n_c	$\hat{Y}^{(2)'}\hat{Y}^{(2)}$		

n_c คือจำนวนหน่วยตัวอย่างที่มีค่า X,Y อยู่ครบถ้วน

4. วิธีการประมาณค่าของตัวแปรที่ขาดหายไปโดยวิธีอันดับศูนย์ดัดแปลง

(A Modified Zero Order Method) [1]

การประมาณค่าสังเกตของตัวแปรที่ขาดหายไปในการวิเคราะห์การถดถอยโดยวิธีนี้ใช้กับการถดถอยเชิงเส้นอย่างง่ายเท่านั้น โดยจะประมาณค่าสังเกต Y ที่ขาดหายไปด้วย u และประมาณค่าสังเกต X ที่ขาดหายไปด้วย v ใช้ค่าประมาณนี้แทนค่าสังเกตที่ขาดหายไป แล้วใช้วิธีกำลังสองน้อยที่สุดในการคำนวณเหมือนเมื่อมีข้อมูลครบในการประมาณค่าพารามิเตอร์

1.4 การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression)

จากตัวแบบ $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$

---

[1] Ibid., pp 15-16.

ให้ $b_o^{(3)}$ และ $b_1^{(3)}$ เป็นค่าประมาณของ β_o และ β_1 จะได้สมการถดถอยที่ต้องการ เป็น

$$\hat{Y}_i^{(3)} = b_o^{(3)} + b_1^{(3)} X_i$$

เมื่อกำหนดให้ $n = (n_c+m_x+m_y) =$ จำนวนหน่วยตัวอย่างทั้งหมด

$n_c =$ จำนวนหน่วยตัวอย่างที่มีทั้งค่าของ X และ Y

$m_x = (n-n_x) =$ จำนวนหน่วยตัวอย่างที่ค่าของ X ขาดหายไป

$m_y = (n-n_y) =$ จำนวนหน่วยตัวอย่างที่ค่าของ Y ขาดหายไป

$n_x =$ จำนวนหน่วยตัวอย่างที่มีค่าของ X ทั้งหมด

n_y จำนวนหน่วยตัวอย่างที่มีค่า Y ทั้งหมด

จากตัวแบบ $Y = \beta_o+\beta_1 X + \epsilon$

โดยวิธีกำลังสองน้อยที่สุด จะต้องทำให้ผลบวกของกำลังที่สองของความคลาดเคลื่อน(**Error sum of Squares**)

$$\sum_{i=1}^{n}\epsilon_i^2 = \sum_{i=1}^{n}(Y_i-\beta_o-\beta_1X_i)^2 \tag{2.4.1}$$

มีค่าน้อยที่สุด เมื่อแทนค่าสังเกต Y และ X ที่ขาดหายไปด้วย **u** และ **v** ตามลำดับ จะเขียนสมการ(2.4.1) ได้ คือ

$$\sum_{i=1}^{n}\epsilon_i^2 = \sum_{i=1}^{n_c}(Y_i-\beta_o-\beta_1X_i)^2+\sum_{i=1}^{m_y}(u-\beta_o-\beta_1X_i)^2 + \sum_{i=1}^{m_x}(Y_i-\beta_o-\beta_1 v)^2 \tag{2.4.2}$$

$$\frac{\partial}{\partial\beta_o}\sum_{i=1}^{n}\epsilon_i^2 = -2\sum_{i=1}^{n_c}(Y_i-\beta_o-\beta_1X_i)-2\sum_{i=1}^{m_y}(u-\beta_o-\beta_1X_i) -2\sum_{i=1}^{m_x}(Y_i-\beta_o-\beta_1 v) = 0 \tag{2.4.3}$$

$$\frac{\partial}{\partial \beta_1}\sum_{i=1}^{n}\epsilon_i^2 = -2\sum_{i=1}^{n_c}(Y_i-\beta_o-\beta_1X_i)(X_i)$$

$$-2\sum_{i=1}^{m_y}(u-\beta_o-\beta_1X_i)(X_i)$$

$$-2\sum_{i=1}^{m_x}(Y_i-\beta_o-\beta_1v)(v)$$

$$= 0 \qquad (2.4.4)$$

$$\frac{\partial}{\partial u}\sum_{i=1}^{n}\epsilon_i^2 = 2\sum_{i=1}^{m_y}(u-\beta_o-\beta_1X_i) = 0 \qquad (2.4.5)$$

$$\frac{\partial}{\partial v}\sum_{i=1}^{n}\epsilon_i^2 = 2\sum_{i=1}^{m_x}(Y_i-\beta_o-\beta_1v)(\beta_1) = o \qquad (2.4.6)$$

แทนค่า β_o และ β_1 ด้วย b'_o และ b'_1 ตามลำดับ

จาก(2.4.5) $\therefore \sum_{i=1}^{m_y} u = \sum_{i=1}^{m_y}b'_o + \sum_{i=1}^{m_y}b'_1X_i$

$$m_y \cdot u = m_yb'_o+b'_1m_y\bar{X}_{m_y}, \quad \bar{X}_{m_y} = \frac{\sum_{i=1}^{m_y}X_i}{m_y}$$

$$\therefore u = b'_o+ b'_1\bar{X}_{m_y} \qquad (2.4.7)$$

จาก (2.4.6) หารตลอดด้วย $2b_1$ จะได้

$$\sum_{i=1}^{m_x}(Y_i-b'_o-b'_1v) = 0$$

$$m_x\bar{Y}_{m_x} - m_xb'_o-m_xb'_1v = 0; \quad \bar{Y}_{m_x} = \frac{\sum_{i=1}^{m_x}Y_i}{m_x}$$

$$\therefore\ v = \frac{\bar{Y}_{m_x} - b'_o}{b'_1} \qquad (2.4.8)$$

แทนค่า u และ v จากสมการ (2.4.7) และ (2.4.8) ใน (2.4.3) จะได้

$$-2\sum_{i=1}^{n_c}(Y_i-b'_o-b'_1X_i)-2\sum_{i=1}^{m_y}\left[(b'_o+b'_1\bar{X}_{m_y})-b_o-b_1X_i\right]-2\sum_{i=1}^{m_x} Y_i-b'_o-b'_1\left(\frac{\bar{Y}_{m_x}-b_o}{b_1}\right) = 0$$

$$\therefore\ \sum_{i=1}^{n_c}(Y_i-b'_o-b'_1X_i)-b'\sum_{1i=1}^{m_y}(X_i-\bar{X}_{m_y})-\sum_{i=1}^{m_x}(Y_i-\bar{Y}_{m_x}) = 0$$

$$\sum_{i=1}^{n_c}(Y_i-b'_o-b'_1X_i)-b_1\left(\sum_{i=1}^{m_y}X_i-m_y\bar{X}_{m_y}\right)-\left(\sum_{i=1}^{m_x}Y_i-m_x\bar{Y}_{m_x}\right) = 0$$

$$\therefore\ n_c b'_o = \sum_{i=1}^{n_c}Y_i-b'_1\sum_{i=1}^{n_c}X_i$$

$$\therefore\ b'_o = \bar{Y}_c - b_1\bar{X}_c \qquad (2.4.9)$$

แทนค่า u และ v ใน (2.4.4) จะได้

$$-2\sum_{i=1}^{n_c}(Y_i-b'_o-b'_1X_i)(X_i)-2\sum_{i=1}^{m_y}\left|(b'_o-b_1\bar{X}_{m_y})-b'_o-b'_1X_i\right|(X_i)$$

$$-2\sum_{i=1}^{m_x}\left[Y_i-b'_o-b'_1\frac{(\bar{Y}_{m_x}-b'_o)}{b'_1}\right]\frac{(\bar{Y}_{m_x}-b'_o)}{b'_1} = 0$$

$$\therefore \sum_{i=1}^{n_c}(Y_i-b'_o-b'_1X_i)(X_i)-b'_1\sum_{i=1}^{m_y}(X_i-\bar{X}_{m_y})(X_i)+\left(\frac{\bar{Y}_{m_x}-b'_o}{b_1}\right)\sum_{i=1}^{m_x}(Y-\bar{Y}_{m_x}) = 0$$

$$\therefore \sum_{i=1}^{n_c}(Y_i-b'_o-b'_1X_i)(X_i)-b'_1\sum_{i=1}^{m_y}(X_i-\bar{X}_{m_y})(X_i) = 0$$

แทนค่า b'_o จาก (2.4.9)

$$\therefore \sum_{i=1}^{n_c}(Y_i-\bar{Y}_c+b'_1\bar{X}_c-b'_1X_i)(X_i)-b'_1\sum_{i=1}^{m_y}(X_i-\bar{X}_{m_y})(X_i) = 0$$

$$\sum_{i=1}^{n_c}X_iY_i-\bar{Y}_c\sum_{i=1}^{n_c}X_i-b'_1\sum_{i=1}^{n_c}(X_i-\bar{X}_c)(X_i)-b'_1\sum_{i=1}^{m_y}(X_i-\bar{X}_{m_y})(X_i) = 0$$

$$\left(\sum_{i=1}^{n_c}X_iY_i-n_c\bar{Y}_c\bar{X}_c\right)-b'_1\left[\left(\sum_{i=1}^{n_c}X_i^2-n_c\bar{X}_c^2\right)+\left(\sum_{i=1}^{m_y}X_i^2-m_y\bar{X}_{m_y}\right)\right] = 0$$

$$\sum_{i=1}^{n_c}(X_i-\bar{X}_c)(Y_i-\bar{Y}_c)-b'_1\left[\sum_{i=1}^{n_c}(X_i-\bar{X}_c)^2+\sum_{i=1}^{m_y}(X_i-\bar{X}_{m_y})^2\right] = 0$$

$$\therefore b'_1 = \frac{\sum_{i=1}^{n_c}(X_i-\bar{X}_c)(Y_i-\bar{Y}_c)}{\sum_{i=1}^{n_c}(X_i-\bar{X}_c)^2+\sum_{i=1}^{m_y}(X_i-\bar{X}_{m_y})^2} \qquad (2.4.10)$$

แทนค่า b'_o ที่ได้จาก (2.4.9) ใน (2.4.7) และ (2.4.8) จะได้

จาก (2.4.7) $\therefore u = \bar{Y}_c-b'_1\bar{X}_c+b'_1\bar{X}_{m_y}$

$$u = \bar{Y}_c+b'_1(\bar{X}_{m_y}-\bar{X}_c) \qquad (2.4.11)$$

จาก (2.4.8) $\therefore v = \dfrac{\bar{Y}_{m_x} - (\bar{Y}_c - b'_1\bar{X}_c)}{b_1}$

$$v = \bar{X}_c + \frac{\bar{Y}_{m_x} - \bar{Y}_c}{b'_1} \qquad (2.4.12)$$

นำค่า u และ v ที่คำนวณได้จาก (2.4.11) และ (2.4.12) แทนค่า Y_i และ X_i ที่ขาดหายไปตามลำดับ แล้วใช้วิธีกำลังสองน้อยที่สุด เหมือนเมื่อมีข้อมูลครบถ้วน จะหาค่า $b_1^{(3)}$ และ $b_o^{(3)}$ ได้เช่นเดียวกับ b_1^{LS} และ b_o^{LS} ซึ่งหาได้จาก (2.1.4) และ (2.1.5) ตามลำดับจะได้

$$b_1^{(3)} = \frac{\sum_{i=1}^{n} X_iY_i - n\bar{X}_n\bar{Y}_n}{\sum_{i=1}^{n} X_i^2 - n\bar{X}_n^2}$$

และ $$b_o^{(3)} = \bar{Y}_n - b_1^{(3)}\bar{X}_n$$

สมการถดถอยที่ได้คือ $$\hat{Y}_i^{(3)} = b_o^{(3)} + b_1^{(3)}X_i$$

เมื่อจะทดสอบสมมุติฐาน $H_o: \beta_1 = 0$; $H_A: \beta_1 \neq 0$ โดยใช้การทดสอบ F (F-test) เราใช้ตาราง ANOVA ต่อไปนี้

ตารางที่ 7

ANOVA

Sov	df.	SS	MS	F
Reg	1	$\sum_{i=1}^{n}(\hat{Y}_i^{(3)} - \bar{Y}_n)^2 = SSR$	SSR/1 =MSR	$\frac{MSR}{MSE}$
Error	n_c-2	$\sum_{i=1}^{n}(Y_i-\hat{Y}_i^{(3)})^2 = SSE$	SSE/(n_c-2)=MSE	
Total	n_c-1	$\sum_{i=1}^{n}(Y_i-\bar{Y}_n)^2 = SST$		

5. <u>วิธีการประมาณค่าของตัวแปรที่ขาดหายไป โดยวิธีถดถอยอันดับหนึ่ง</u>
(A First Order Regression Method) [1]

การประมาณค่าสังเกตของตัวแปรที่ขาดหายไปในการวิเคราะห์การถดถอย โดยวิธีนี้จะกล่าวถึงวิธีการถดถอยเชิงเส้นอย่างง่ายเท่านั้น ซึ่งมีวิธีการที่สลับซับซ้อนพอสมควร โดยมีวิธีทำเป็นขั้น ๆ ดังนี้

---

[1] A.A. Afifi and R.M. Elashoff. "Missing Observation in Multivariate Statistics III. Large Sample Analysis of Simple Linear Regression," <u>Journal of the American Statistical Association</u>, 64 (Number 325) pp. 346-350.

ก. คำนวณการถดถอยของ Y ในเทอม X จากข้อมูลซึ่งมีทั้งค่าของ Y และ X โดยวิธีกำลังสองน้อยที่สุด และประมาณค่า Y's ที่ขาดหายไปโดยใช้สมการถดถอยที่ได้นี้

ข. คำนวณการถดถอยของ X ในเทอม Y จากข้อมูลซึ่งมีทั้งค่า Y และ X โดยวิธีกำลังสองน้อยที่สุด และประมาณค่า X's ที่ขาดหายไป โดยใช้สมการถดถอยที่ได้นี้

ค. นำค่า Y_i ที่ประมาณได้จากข้อ ก. และค่า X_i ที่ประมาณได้จากข้อ ข. แทนค่า Y_i และ X_i ที่ขาดหายไปตามลำดับ

ง. คำนวณการถดถอยของ Y ในเทอม X ใหม่ โดยใช้วิธีกำลังสองน้อยที่สุดในการคำนวณเหมือนเมื่อมีข้อมูลครบถ้วน

จากตัวแบบ $$Y = \beta_o + \beta_1 X + \epsilon$$

ให้ $b_o^{(4)}$ และ $b_1^{(4)}$ เป็นค่าประมาณของ β_o และ β_1 จะได้สมการถดถอยที่ต้องการ คือ

$$\hat{Y}_i^{(4)} = b_o^{(4)} + b_1^{(4)} X_i$$

เมื่อกำหนดให้ $n = (n_c + m_x + m_y) =$ จำนวนหน่วยตัวอย่างทั้งหมด

$n_c =$ จำนวนหน่วยตัวอย่างที่มีทั้งค่าของ X และ Y

$m_x = (n-n_x) =$ จำนวนหน่วยตัวอย่างที่ค่าของ X ขาดหายไป

$m_y = (n-n_y) =$ จำนวนหน่วยตัวอย่างที่ค่าของ Y ขาดหายไป

n_x จำนวนหน่วยตัวอย่างที่มีค่าของ X ทั้งหมด

n_y จำนวนหน่วยตัวอย่างที่มีค่าของ Y ทั้งหมด

ให้นำข้อมูลที่รวบรวมมาได้ ซึ่งมีทั้งค่า X และ Y (n_c) คำนวณหาการถดถอยของ Y on X โดยหาค่า $b_1'^{LS}$ และ $b_o'^{LS}$ ได้เช่นเดียวกับ b_1^{LS} และ b_o^{LS} ซึ่งหาได้จาก (2.1.4) และ (2.1.5) ตามลำดับ จะได้

$$b_1'^{LS} = \frac{\sum_{i=1}^{n_c} X_i Y_i - n_c \bar{Y}_c \bar{X}_c}{\sum_{i=1}^{n_c} X_i^2 - n_c \bar{X}_c^2} \quad ; \quad \bar{X}_c = \frac{\sum_{i=1}^{n_c} X_i}{n_c}$$

$$\bar{Y}_c = \frac{\sum_{i=1}^{n_c} Y_i}{n_c}$$

และ $b_o'^{LS} = \bar{Y}_c - b_1'^{LS} \bar{X}_c$

ดังนั้นสมการถดถอยที่ได้คือ $\hat{Y}_i'^{LS} = b_o'^{LS} + b_1'^{LS} X_i$ (2.5.1)

จากข้อมูลเดียวกันนี้ จะคำนวณการถดถอยของ X on Y ได้เช่นเดียวกัน โดยหาค่า d_1^{LS} และ d_o^{LS} ได้ เช่นเดียวกับ b_1^{LS} และ b_o^{LS} ซึ่งหาได้จาก (2.1.4) และ (2.1.5) ตามลำดับ จะได้

$$d_1^{LS} = \frac{\sum_{i=1}^{n_c} X_i Y_i - n_c \bar{Y}_c \bar{X}_c}{\sum_{i=1}^{n_c} Y_i^2 - n_c \bar{Y}_c^2}$$

$$d_o^{LS} = \bar{X}_c - d_1^{LS} \bar{Y}_c$$

สมการถดถอยที่ได้ คือ $\hat{X}_i^{LS} = d_o^{LS} + d_1^{LS} Y_i$ (2.5.2)

ประมาณค่าสังเกต Y's และ X's ที่ขาดหายไป โดยใช้สมการถดถอยซึ่งหาได้จาก (2.5.1) และ (2.5.2) ตามลำดับ แล้วให้นำค่า Y_i และ X_i ที่ประมาณได้ไป

แทนค่า สังเกต Y_i และ X_i ที่ขาดหายไปตามลำดับ แล้วใช้วิธีกำลังสองน้อยที่สุด เหมือนเมื่อมีข้อมูลครบถ้วน หาค่า $b_1^{(4)}$ และ $b_o^{(4)}$ ซึ่งจะหาได้เช่นเดียวกับ b_1^{LS} และ b_o^{LS} ซึ่งหาได้จาก (2.1.4) และ (2.1.5) ตามลำดับ จะได้

$$b_1^{(4)} = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}_n \bar{Y}_n}{\sum_{i=1}^{n} X_i^2 - n\bar{X}_n^2}$$

และ $$b_o^{(4)} = \bar{Y}_n - b_1^{(4)}\bar{X}_n$$

สมการถดถอยที่ได้คือ $\hat{Y}_i^{(4)} = b_o^{(4)} + b_1^{(4)} X_i$

โดยที่ $$\bar{Y}_n = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{1}{n}\left\{ \sum_{i=1}^{n_c} Y_i + \sum_{i=1}^{m_x} Y_i + \sum_{i=1}^{m_y} Y_i \right\}$$

$$= \frac{1}{n}\left\{ \sum_{i=1}^{n_c} Y_i + \sum_{i=1}^{m_x} Y_i + \sum_{i=1}^{m_y}\left[\bar{Y}_c + b_1^{LS}(X_i - \bar{X}_c) \right] \right\}$$

$$= \frac{1}{n}\left\{ n_c\bar{Y}_c + m_x \bar{Y}_{m_x} + m_y \bar{Y}_c + m_y b_1^{LS}(\bar{X}_{m_y} - \bar{X}_c) \right\}$$

$$= \frac{(n_c + m_y)}{n} \cdot \bar{Y}_c + \frac{m_x}{n}\bar{Y}_{m_x} + \frac{m_y}{n} \cdot b_1^{LS}(\bar{X}_{m_y} - \bar{X}_c)$$

$$= \frac{n_x}{n} \cdot \bar{Y}_c + \frac{m_x}{n}\bar{Y}_{m_x} + \frac{m_y}{n} \cdot b^{LS}(\bar{X}_{m_y} - \bar{X}_c)$$

$$\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{1}{n}\left\{ \sum_{i=1}^{n_c} X_i + \sum_{i=1}^{m_x} X_i + \sum_{i=1}^{m_y} X_i \right\}$$

$$= \frac{1}{n}\left\{\sum_{i=1}^{n_c} X_i + \sum_{i=1}^{m_x}\left[\bar{X}_c + d_1^{LS}(Y_i - \bar{Y}_c)\right] + \sum_{i=1}^{m_y} X_i\right\}$$

$$= \frac{1}{n}\left\{n_c\bar{X}_c + m_x\bar{X}_c + m_x d_1^{LS}(\bar{Y}_{m_x} - \bar{Y}_c) + m_y\bar{X}_{m_y}\right\}$$

$$= \frac{(n_c+m_x)}{n}.\bar{X}_c + \frac{m_y}{n}\bar{X}_{m_y} + \frac{m_x}{n} d_1^{LS}(\bar{Y}_{m_x} - \bar{Y}_c)$$

$$= \frac{n_y}{n}.\bar{X}_c + \frac{m_y}{n}.\bar{X}_m + \frac{m_x}{n}.d_1^{LS}(\bar{Y}_{m_x} - \bar{Y}_c)$$

เมื่อจะทดสอบสมมุติฐาน $H_o : \beta_1 = 0$; $H_A : \beta_1 \neq 0$ โดยใช้การทดสอบ F (F-test) เราใช้ตาราง ANOVA ต่อไปนี้

ตารางที่ 8

ANOVA

SOV	df.	SS	MS	F
Reg	1	$\sum_{i=1}^{n}(\hat{Y}_i^{(4)} - \bar{Y}_n)^2 = SSR$	$SSR/1 = MSR$	$\frac{MSR}{MSE}$
Error	n_c-2	$\sum_{i=1}^{n}(Y_i - \hat{Y}_i^{(4)})^2 = SSE$	$SSE/(n_c-2) = MSE$	
Total	n_c-1	$\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2 = SST$		

6. <u>วิธีการประมาณค่าของตัวแปรที่ขาดหายไป โดยวิธีถดถอยสองขั้น</u>

(Another First Order Regression Method or Two-Stage Regression Method) [1]

การประมาณค่าสังเกตของตัวแปรที่ขาดหายไปในการวิเคราะห์การถดถอยโดยวิธีนี้ จะกล่าวถึงการถดถอยเชิงเส้นอย่างง่ายเท่านั้น มีวิธีการทำเป็นขั้น ๆ ดังนี้

ก. คำนวณการถดถอยของ X ในเทอม Y จากข้อมูลซึ่งมีค่าทั้ง Y และ X โดยวิธีกำลังสองน้อยที่สุดและประมาณค่า X's ที่ขาดหายไปโดยใช้สมการถดถอยที่ได้นี้

ข. นำค่า X_i ที่ประมาณได้จากข้อ ก. แทนค่า X_i ที่ขาดหายไป

ค. คำนวณการถดถอยของ Y ในเทอม X ใหม่ จากข้อมูลที่มีทั้ง Y และ X $(=n_y)$ โดยใช้วิธีกำลังสองน้อยที่สุด

จากตัวแบบ $$Y = \beta_0 + \beta_1 X + \epsilon$$

ให้ $b_0^{(5)}$ และ $b_1^{(5)}$ เป็นค่าประมาณของ β_0 และ β_1 จะได้สมการถดถอยที่ต้องการ คือ

$$\hat{Y}^{(5)} = b_0^{(5)} + b_1^{(5)} X_i$$

เมื่อกำหนดให้ $n = (n_c+m_x+m_y)$ = จำนวนหน่วยตัวอย่างทั้งหมด

n_c = จำนวนหน่วยตัวอย่างที่มีทั้งค่าของ X และ Y

$m_x = (n-n_x)$ = จำนวนหน่วยตัวอย่างที่ค่าของ X ขาดหายไป

---

[1] Ibid., pp. 350-351.

m_y = $(n-n_y)$ = จำนวนหน่วยตัวอย่างที่ค่าของ Y ขาดหายไป

n_x = จำนวนหน่วยตัวอย่างที่มีค่าของ X ทั้งหมด

n_y = จำนวนหน่วยตัวอย่างที่มีค่าของ Y ทั้งหมด

ให้นำข้อมูลที่รวบรวมมาได้ ซึ่งมีทั้งค่า X และ Y (n_c) คำนวณหาการถดถอยของ X on Y โดยหาค่า d_1^{LS} และ d_o^{LS} ได้เช่นเดียวกับ b_1^{LS} และ b_o^{LS} ซึ่งหาได้จาก (2.1.4) และ (2.1.5) ตามลำดับจะได้

$$d_1^{LS} = \frac{\sum_{i=1}^{n_c} X_i Y_i - n_c \bar{Y}_c \bar{X}_c}{\sum_{i=1}^{n_c} Y_i^2 - n_c \bar{Y}_c^2} \quad ; \quad \bar{X}_c = \frac{\sum_{i=1}^{n_c} X_i}{n_c}$$

$$\bar{Y}_c = \frac{\sum_{i=1}^{n_c} Y_i}{n_c}$$

และ $d_1^{LS} = \bar{X}_c - d_1^{LS} \bar{Y}_c$

สมการถดถอยที่ได้ คือ $\hat{X}_i^{LS} = d_o^{LS} + d_1^{LS} Y_i$

ประมาณค่าสังเกต X's ที่ขาดหายไป โดยใช้สมการถดถอย ซึ่งหาได้นี้แล้วนำค่า X_i ที่ประมาณได้ แทนค่าสังเกต X_i ที่ขาดหายไปตามลำดับ แล้วใช้วิธีกำลังสองน้อยที่สุด เหมือนเมื่อมีข้อมูลครบถ้วน หาค่า $b_1^{(5)}$ และ $b_o^{(5)}$ ซึ่งจะหาได้เช่นเดียวกับ b_1^{LS} และ b_o^{LS} ซึ่งหาได้จาก (2.1.4) และ (2.1.5) ตามลำดับ จะได้

$$b_1^{(5)} = \frac{\sum_{i=1}^{n_y} X_i Y_i - n_y \bar{Y}_{n_y} \bar{X}_{n_y}}{\sum_{i=1}^{n_y} X_i^2 - n_y \bar{X}_{n_y}^2} \quad ;$$

$$b_o^{(5)} = \bar{Y}_{n_y} - b_1^{(5)} \bar{X}_{n_y}$$

โดยที่ $\bar{Y}_{n_y} = \dfrac{\sum_{i=1}^{n_y} Y_i}{n_y}$

$$\bar{X}_{n_y} = \frac{\sum_{i=1}^{n_y} X_i}{n_y}$$

$$= \frac{1}{n_y}\left\{\sum_{i=1}^{n_c} X_i + \sum_{i=1}^{m_x} X_i\right\}$$

$$= \frac{1}{n_y}\left\{n_c\bar{X}_c + \sum_{i=1}^{m_x}\left[X_c + d_1^{LS}(Y_i - \bar{Y}_c)\right]\right\}$$

$$= \frac{1}{n_y}\left\{n_c \cdot \bar{X}_c + m_x \cdot \bar{X}_c + m_x d_1^{LS}(\bar{Y}_{m_x} - \bar{Y}_c)\right\}$$

$$= \frac{n_c + m_x}{n_y} \cdot \bar{X}_c + \frac{m_x}{n_y} \cdot d_1^{LS}(\bar{Y}_{m_x} - \bar{Y}_c)$$

$$= \bar{X}_c + \frac{m_x}{n_y}\, d_1^{LS}(\bar{Y}_{m_x} - \bar{Y}_c)$$

สมการถดถอยที่ได้ คือ $\hat{Y}_i^{(5)} = b_o^{(5)} + b_1^{(5)} X_i$

เมื่อจะทดสอบสมมุติฐาน $H_o : \beta_1 = 0 \; ; \; H_A : \beta_1 \neq 0$ โดยใช้การทดสอบ F (F-test) เราใช้ตาราง ANOVA ต่อไปนี้

ตารางที่ 9

ANOVA

SOV	df.	SS	MS	F
Reg	1	$\sum_{i=1}^{n_y} (\hat{Y}_i^{(5)} - \bar{Y}_{n_y})^2 = SSR$	SSR/1 =MSR	$\frac{MSR}{MSE}$
Error	n_c-2	$\sum_{i=1}^{n_y} (Y_i - \hat{Y}_i^{(5)})^2 = SSE$	SSE/(n_c-2)=MSE	
Total	n_c-1	$\sum_{i=1}^{n_y} (Y_i - \bar{Y}_{n_y})^2 = SST$		

7. <u>วิธีการประมาณค่าของตัวแปรที่ขาดหายไปโดยวิธีผสม</u> (Mixed Method)

การประมาณค่าสังเกตของตัวแปรที่ขาดหายไปในการวิเคราะห์การถดถอยโดยวิธีนี้ เป็นวิธีที่ผู้ทำวิจัยได้นำเสนอมาโดยอาศัยแนวความคิดจากวิธีอันดับศูนย์และวิธีถดถอยอันดับหนึ่ง มีวิธีการดำเนินการเป็นขั้น ๆ ดังนี้

ก. หาค่าเฉลี่ยของค่าสังเกต X และ Y เท่าที่ข้อมูล X และ Y มีอยู่ จะได้ค่า $\bar{X}_{n_x}$ และ $\bar{Y}_{n_y}$

ข. แทนค่าสังเกต X และ Y ที่ขาดหายไปด้วย $\bar{X}_{n_x}$ และ $\bar{Y}_{n_y}$ ตามลำดับ แล้วนำข้อมูลที่ได้ทั้งหมด คำนวณการถดถอยของ Y ในเทอม X และของ X ในเทอม Y โดยวิธีกำลังสองน้อยที่สุด ซึ่งจะได้สมการถดถอย คือ

$$\hat{Y}_i^o = b_o^o + b_1^o X_i$$

$$\hat{X}_i^o = d_o^o + d_1^o Y_i$$

ค. หาค่าสังเกต Y_i และ X_i ที่ขาดหายไปโดยใช้สมการถดถอยที่ได้จากข้อ ข.

ง. หาค่าเฉลี่ยของ $\frac{\bar{X}_{n_x}+\hat{X}_i^o}{2}$ และ $\frac{\bar{Y}_{n_y}+\hat{Y}_i^o}{2}$ แล้วนำค่าเฉลี่ยที่ได้นี้แทนค่า สังเกต X_i และ Y_i ที่ขาดหายไปตามลำดับ

จ. คำนวนการถดถอยของ Y ในเทอม X ใหม่ โดยใช้วิธีกำลังสองน้อยที่สุด ทำเช่นเดียวกับกรณีที่มีข้อมูลครบถ้วน จึงจะได้สมการถดถอย คือ

$$\hat{Y}_i^{(6)} = b_o^{(6)} + b_1^{(6)} X_i$$

ฉ. ใช้สมการถดถอยที่ได้จากข้อ จ. ในการวิเคราะห์การถดถอย

7.1 <u>การถดถอยเชิงเส้นอย่างง่าย</u> (Simple Linear Regression)

จากตัวแบบ $Y = \beta_o + \beta_1 X + \epsilon$

ให้ $b_o^{(6)}$ และ $b_1^{(6)}$ เป็นค่าประมาณของ β_o และ β_1 ตามลำดับ จะได้สมการถดถอยที่ต้องการ เป็น

$$\hat{Y}_i^{(6)} = b_o^{(6)} + b_1^{(6)} X_i$$

เมื่อกำหนดให้ $n = (n_c+m_x+m_y)$ จำนวนหน่วยตัวอย่างทั้งหมด

n_c = จำนวนหน่วยตัวอย่างที่มีทั้งค่า ของ X และ Y

$m_x = (n-n_x)$ จำนวนหน่วยตัวอย่างที่ค่าของ X ขาดหายไป

$m_y = (n-n_y)$ จำนวนหน่วยตัวอย่างที่ค่าของ Y ขาดหายไป

n_x = จำนวนหน่วยตัวอย่างที่มีค่าของ X ทั้งหมด

n_y = จำนวนหน่วยตัวอย่างที่มีค่าของ Y ทั้งหมด

การประมาณค่าที่ขาดหายไปในการวิเคราะห์การถดถอยโดยวิธีผสมมีขั้นตอนดังนี้

(1) ให้นำข้อมูลที่รวบรวมมาได้มาหาค่า $\bar{X}_{n_x}$ และ $\bar{Y}_{n_y}$ โดยที่

$$\bar{X}_{n_x} = \frac{\sum_{i=1}^{n_x} X_i}{n_x}$$

$$\bar{Y}_{n_y} = \frac{\sum_{i=1}^{n_y} Y_i}{n_y}$$

(2) นำค่า $\bar{X}_{n_x}$ และ $\bar{Y}_{n_y}$ แทนค่าสังเกต X_i และ Y_i ที่ขาดหายไปตามลำดับ แล้วนำข้อมูลที่ได้ทั้งหมดไปหาสมการถดถอยของ Y ในเทอม X และของ X ในเทอม Y โดยวิธีกำลังสองน้อยที่สุด ซึ่งจะได้

$$\hat{Y}_i^o = b_o^o + b_1^o X_i \qquad (2.7.1)$$

และ

$$\hat{X}_i^o = d_o^o + d_1^o Y_i \qquad (2.7.2)$$

โดยที่ b_1^o, b_o^o, d_1^o และ d_o^o จะหาได้เช่นเดียวกับ b_1^{LS} และ b_o^{LS} ซึ่งหาได้จาก (2.1.4) และ (2.1.5) ตามลำดับ จะได้

$$b_1^o = \frac{\sum_{i=1}^{n_c} X_i Y_i - n_c \bar{Y}_{n_y} \bar{X}_{n_x}}{\sum_{i=1}^{n_x} X_i^2 - n_x \bar{X}_{n_x}^2}$$

$$b_o^o = \bar{Y}_{n_y} - b_1^o \bar{X}_{n_x}$$

$$d_1^o = \frac{\sum_{i=1}^{n_c} X_i Y_i - n_c \bar{Y}_{n_y} \bar{X}_{n_x}}{\sum_{i=1}^{n_y} Y_i^2 - n_y \bar{Y}_{n_y}^2}$$

$$d_o^o = \bar{X}_{n_x} - d_1^o \bar{Y}_{n_y}$$

(3) หาค่าสังเกต Y_i และ X_i ที่ขาดหายไปโดยใช้สมการถดถอยที่ได้จาก (2.7.1) และ (2.7.2) ตามลำดับ

(4) ประมาณหาค่า Y_i และ X_i ที่ขาดหายไปด้วย $\hat{Y}_i'$ และ $\hat{X}_i'$ ตามลำดับ แล้วนำค่าที่ได้นี้แทนค่า Y_i และ X_i ที่ขาดหายไป ตามลำดับ โดยที่

$$\hat{Y}_i' = \frac{\bar{Y}_{n_y} + \hat{Y}_i^o}{2}$$

และ $$\hat{X}_i' = \frac{\bar{X}_{n_x} + \hat{X}_i^o}{2}$$

(5) นำข้อมูลที่ได้ใหม่นี้มาคำนวณการถดถอยของ Y on X โดยวิธีกำลังสองน้อยที่สุดเหมือนเมื่อมีข้อมูลครบถ้วน ซึ่งจะหาค่า $b_1^{(6)}$ และ $b_o^{(6)}$ ได้เช่นเดียวกับ b_1^{LS} และ b_o^{LS} ซึ่งหาได้จาก (2.1.4) และ (2.1.5) ตามลำดับ จะได้

$$b_1^{(6)} = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{Y}_n \bar{X}_n}{\sum_{i=1}^{n} X_i^2 - n\bar{X}_n^2}$$

และ $$b_o^{(6)} = \bar{Y}_n - b_1^{(6)} \bar{X}_n$$

โดยที่ $$\bar{Y}_n = \frac{\sum_{i=1}^{n} Y_i}{n}$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n_y} Y_i + \sum_{i=1}^{m_y} Y_i\right]$$

$$= \frac{1}{n}\left[n_y \cdot \bar{Y}_{n_y} + \sum_{i=1}^{m_y} \frac{\bar{Y}_{n_y} + \hat{Y}_i^o}{2}\right]$$

$$= \frac{1}{n}\left\{n_y.\bar{Y}_{n_y} + \frac{1}{2}\sum_{i=1}^{m_y}\left[\bar{Y}_{n_y} + \bar{Y}_{n_y} + b_1^o(X_i - \bar{X}_{n_x})\right]\right\}$$

$$= \frac{1}{n}\left[(n_y + m_y).\bar{Y}_{n_y} + \frac{1}{2}b_1^o\sum_{i=1}^{m_y}(X_i - \bar{X}_{n_x})\right]$$

$$= \bar{Y}_{n_y} + \frac{1}{2n} b_1^o\sum_{i=1}^{m_y}(X_i - \bar{X}_{n_x})$$

$$\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n_x} X_i + \sum_{i=1}^{m_x} X_i\right]$$

$$= \frac{1}{n}\left[n_x.\bar{X}_{n_x} + \sum_{i=1}^{m_x}\frac{\bar{X}_{n_x} + \hat{X}_i^o}{2}\right]$$

$$= \frac{1}{n}\left\{n_x.\bar{X}_{n_x} + \frac{1}{2}\sum_{i=1}^{m_x}\left[\bar{X}_{n_x} + \bar{X}_{n_x} + d_1^o(Y_i - \bar{Y}_{n_y})\right]\right\}$$

$$= \frac{1}{n}\left[(n_x + m_x)\bar{X}_{n_x} + \frac{1}{2} d_1^o\sum_{i=1}^{m_x}(Y_i - \bar{Y}_{n_y})\right]$$

$$= \bar{X}_{n_x} + \frac{1}{2n} \cdot d_1^o\sum_{i=1}^{m_x}(Y_i - \bar{Y}_{n_y})$$

สมการถดถอยที่ได้คือ $\hat{Y}_i^{(6)} = b_o^{(6)} + b_1^{(6)}X_i$

เมื่อจะทดสอบสมมุติฐาน $H_o : \beta_1 = 0$; $H_A : \beta_1 \neq 0$ โดยใช้การทดสอบ F (F-test) เราใช้ตาราง ANOVA ต่อไปนี้

ตารางที่ 10

ANOVA

SOV	df.	SS	MS	F
Reg	1	$\sum_{i=1}^{n}(\hat{Y}_i^{(6)} - \bar{Y}_n)^2 = SSR$	$SSR/1 = MSR$	$\frac{MSR}{MSE}$
Error	n_c-2	$\sum_{i=1}^{n}(Y_i - \hat{Y}_i^{(6)})^2 = SSE$	$SSE/(n_c-2) = MSE$	
Total	n_c-1	$\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2 = SST$		

7.2 <u>การถดถอยเชิงเส้นพหุคูณ</u> (Multiple Linear Regression)

จากตัวแบบ $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \beta_p X_p + \epsilon$

หรือเขียนในรูปเมตริกซ์จะได้ $\underline{Y} = \underline{X}\,\underline{\beta} + \underline{\epsilon}$

ให้ $b_i^{(6)}$ เป็นค่าประมาณของ β_i, $i = 0, 1, 2, \ldots p$ ตามลำดับ โดยวิธีผสม จะได้สมการถดถอย เป็น

$$\hat{Y}_i^{(6)} = b_o^{(6)} + b_1^{(6)}X_{1i} + b_2^{(6)}X_{2i} + \ldots\ldots + b_p^{(6)}X_{p_i}$$

หรือเขียนในรูปเมตริกซ์จะได้ $\hat{\mathrm{Y}}^{(6)} = \mathrm{X}\beta^{(6)}$

สมมติว่าข้อมูลที่รวบรวมมาได้ มีดังนี้

$$\mathrm{Y} = \begin{bmatrix} \mathrm{Y}_1 \\ \mathrm{Y}_2 \\ \mathrm{Y}_3 \\ \mathrm{Y}_4 \end{bmatrix}, \quad \mathrm{X} = \begin{bmatrix} \mathrm{X}_1 \\ \mathrm{X}_2 \\ \mathrm{X}_3 \\ \mathrm{X}_4 \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}$$

โดยที่ Y_1 แทนเวคเตอร์ค่าสังเกตที่มีอยู่ ซึ่งตรงกับ X_1

Y_2 แทนเวคเตอร์ค่าสังเกตที่สูญหาย ซึ่งตรงกับ X_2

Y_3 แทนเวคเตอร์ค่าสังเกตที่มีอยู่ ซึ่งตรงกับ X_3

Y_4 แทนเวคเตอร์ค่าสังเกตที่สูญหาย ซึ่งตรงกับ X_4

X_1 แทนเมตริกซ์ที่ไม่มีค่าสังเกตตัวใดตัวหนึ่งสูญหาย

X_2 แทนเมตริกซ์ที่ไม่มีค่าสังเกตตัวใดตัวหนึ่งสูญหาย

X_3 แทนเมตริกซ์ที่มีค่าสังเกตสูญหายอย่างน้อยหนึ่งตัว

X_4 แทนเมตริกซ์ที่มีค่าสังเกตสูญหายอย่างน้อยหนึ่งตัว

เราประมาณค่าที่ขาดหายไปโดยใช้ขั้นตอนดังนี้

(1) ให้นำข้อมูลที่รวบรวมมาได้มาหาค่า $\bar{Y}_{n_y}$ และ $\bar{X}_{n_{x_j}}$ โดยที่

$$\bar{Y}_{n_y} = \frac{\sum_{i=1}^{n_y} Y_i}{n_y} \; ; \quad n_y \text{ : จำนวนหน่วยตัวอย่างที่มีค่า Y ทั้งหมด}$$

$$\bar{X}_j = \frac{\sum_{i=1}^{n_{x_j}} X_i}{n_{x_j}} ; \quad n_{x_j} = \text{ จำนวนหน่วยตัวอย่างที่มีค่า } X_j \text{ ทั้งหมด}$$

$$j = 1, 2 \ldots\ldots p$$

(2) นำค่า $\bar{Y}_{n_y}$ และ $\bar{X}_j$ แทนค่าสังเกต Y และ X_{ij} ที่ขาดหายไปตามลำดับ แล้วนำข้อมูลที่ได้ทั้งหมดไปหาสมการถดถอยของ Y on X และของ X_j on Y โดยวิธีกำลังสองน้อยที่สุด ซึ่งจะได้

$$\hat{Y} = X \hat{\beta}' \tag{2.7.3}$$

และ $$\hat{Y}_j'' = X_o \hat{\beta}'' \tag{2.7.4}$$

เมื่อ Y_j คือค่า X_j โดยการแทนค่า Y กับ X_j

$$\beta' = [X'X]^{-1} X'Y$$

และ $$\beta'' = [X'_o X_o]^{-1} X'_o Y$$

(3) หาค่าสังเกต Y_i และ X_{ij} ที่ขาดหายไปโดยใช้สมการถดถอยที่ได้จาก (2.7.3) และ (2.7.4) ตามลำดับ

(4) ประมาณค่า Y_i และ X_{i_j} ที่ขาดหายไปด้วย $\hat{Y}_i'$ และ $\hat{X}_{ij}'$ ตามลำดับ แล้วนำค่าที่ได้แทนค่า Y_i และ X_i ที่ขาดหายไปตามลำดับ โดยที่

$$\hat{Y}_i' = \frac{\bar{Y}_{n_y} + \hat{Y}_i}{2}$$

และ $$\hat{X}_{i_j}' = \frac{\bar{X}_{n_y} + \hat{X}_{ij}}{2}$$

(5) นำข้อมูลที่ได้ใหม่นี้มาคำนวณการถดถอยของ Y on X โดยวิธีกำลังสองน้อยที่สุด เหมือนเมื่อมีข้อมูลครบถ้วน จะหาค่า $\hat{\beta}^{(6)}$ ได้เช่นเดียวกับ $\hat{\beta}^{LS}$ ซึ่งหาได้จาก (2.1.11) จะได้

$$\hat{\beta}^{(6)} = (X'X)^{-1}X'Y$$

สมการถดถอยที่ได้คือ $\hat{Y}^{(6)} = X\hat{\beta}^{(6)}$

เมื่อจะทดสอบสมมติฐาน $H_o : \beta = 0$; $H_A : \beta \neq 0$ โดยใช้การทดสอบ F (F.test) ตามตาราง ANOVA ดังนี้

ตารางที่ 11

ANOVA

SOV	df	SS	MS	F
Reg	$p+1$	$\hat{\beta}^{(6)}X'\hat{Y}^{(6)}$	$\frac{SSR}{p+1} = MSR$	$\frac{MSR}{MSE}$
Error	$n_c-(p+1)$	$\hat{Y}'^{(6)}\hat{Y}^{(6)} - \hat{\beta}^{(6)}X'\hat{Y}^{(6)}$	$\frac{SSE}{n_c-(p+1)} = MSE$	
Total	n_c	$\hat{Y}'^{(6)}\hat{Y}^{(6)}$		