# บทที่ 2
# การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย

การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย (simple linear regression analysis) เป็นการวิเคราะห์การถดถอยของตัวแปรอิสระ 1 ตัวและตัวแปรตาม 1 ตัวโดยตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นตรงกันอาจเป็นความสัมพันธ์ตามกันหรือผกผันก็ได้ รูปแบบการวิเคราะห์นี้เป็นรูปแบบพื้นฐานที่ง่ายที่สุดของการวิเคราะห์การถดถอยโดยมีตัวแบบการถดถอยคือ

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad (2.1)$$

โดย $Y_i$ = ค่าของตัวแปรตามในลำดับที่ $i$

$\beta_0$ และ $\beta_1$ = พารามิเตอร์ที่ไม่ทราบค่า

$X_i$ = ค่าคงที่ของตัวแปรอิสระในลำดับที่ $i$

$\varepsilon_i$ = ความคลาดเคลื่อน (random error) ในลำดับที่ $i$

ความคลาดเคลื่อนมีข้อกำหนดว่าต้องเป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติโดยมีค่าเฉลี่ยหรือ $E(\varepsilon_i)$ เท่ากับ 0 และความแปรปรวนหรือ $\sigma^2(\varepsilon_i)$ เท่ากับ $\sigma^2$ และความคลาดเคลื่อนแต่ละค่ามีความเป็นอิสระต่อกัน เนื่องจาก $\varepsilon_i$ และ $\varepsilon_j$ ไม่มีความสัมพันธ์กันดังนั้นอาจกล่าวได้ว่าความแปรปรวนร่วม (covariance) มีค่าเท่ากับ 0 หรือ $\sigma(\varepsilon_i, \varepsilon_j)$ เมื่อ $i \neq j$ จากข้อตกลงของความคลาดเคลื่อนดังกล่าวส่งผลให้ตัวแปรตาม $Y$ แต่ละค่ามีความเป็นอิสระต่อกันและมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ $\beta_0 + \beta_1 X_i$ และความแปรปรวนเท่ากับ $\sigma^2$ หรือความแปรปรวนของความคลาดเคลื่อนนั่นเองหากเขียนในรูปแบบสัญลักษณ์ทางสถิติจะได้ว่า $Y_i \sim \text{NID}(\beta_0 + \beta_1 X_i, \sigma^2)$ (Montgomery & Peck, 1992, p. 10-11)

ค่าพารามิเตอร์ $\beta_0$ และ $\beta_1$ เรียกว่าสัมประสิทธิ์การถดถอย (regression coefficient) โดยค่า $\beta_1$ คือความชันของสมการถดถอยที่บอกให้ทราบถึงอัตราการเปลี่ยนแปลงของค่าเฉลี่ยของการแจกแจงของตัวแปร $Y$ เมื่อตัวแปรอิสระ $X$ มีค่าเพิ่มขึ้น 1 หน่วยในขณะที่ $\beta_0$ คือจุดตัดแกน $Y$ ของสมการถดถอยหรือเป็นค่าเฉลี่ยของการแจกแจงของตัวแปรตาม $Y$ เมื่อตัวแปรอิสระ $X$ มีค่าเท่ากับ 0 การตีความ $\beta_0$ นั้นหากข้อมูลที่นำมาศึกษามิได้ครอบคลุมค่า 0 แล้วไม่สมควรที่จะตีความค่า $\beta_0$

Hair et al (2010, p 176) กล่าวว่าขนาดตัวอย่างที่เพียงพอในการวิเคราะห์สมการถดถอยเชิงเส้นอย่างง่ายนั้นเท่ากับ 20 ข้อมูล

## 2.1 การประมาณค่าพารามิเตอร์โดยวิธีกำลังสองน้อยที่สุด

การวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายเป็นการสร้างสมการถดถอยหรือสร้างตัวแบบของประชากรโดยตัวแบบที่สร้างขึ้นแสดงได้ดังสมการ (2.1) เนื่องจากในสมการดังกล่าวนักวิจัยไม่ทราบค่าพารามิเตอร์ $\beta_0$ และ $\beta_1$ จึงต้องใช้ข้อมูลที่ได้จากตัวอย่างหรือข้อมูลที่ได้เก็บรวบรวมมาเพื่อประมาณค่าของพารามิเตอร์ทั้งสองเพื่อใช้ในพยากรณ์ค่าของตัวแปรตามหรือ $\hat{Y}$ ค่าพยากรณ์นี้เรียก fitted value โดยค่าพยากรณ์สามารถเขียนได้ดังนี้

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \tag{2.2}$$

โดยที่ $\hat{\beta}_0$ และ $\hat{\beta}_1$ คือค่าประมาณของพารามิเตอร์ $\beta_0$ และ $\beta_1$ ตามลำดับ

ในการประมาณค่าพารามิเตอร์ $\beta_0$ และ $\beta_1$ นั้นมีด้วยกันหลายวิธีแต่วิธีที่เป็นที่นิยมคือวิธีกำลังสองน้อยที่สุด (ordinary least square estimation) หลักการของวิธีนี้คือ การประมาณค่าพารามิเตอร์ให้ค่าผลรวมกำลังสองของส่วนเหลือ (residual) ที่น้อยที่สุด ส่วนเหลือ ($e_i$) คือค่าความแตกต่างระหว่างค่าจริงของตัวแปรตาม $Y$ กับค่าพยากรณ์ที่ได้จากสมการถดถอย ($\hat{Y}$) ที่ระดับเดียวกันของค่าของตัวแปรอิสระ $X$ หรือสามารถเขียนได้ดังนี้

$$e_i = Y_i - \hat{Y}_i = Y_i - (\beta_0 + \beta_1 X_i) \tag{2.3}$$

ความคลาดเคลื่อน ($\varepsilon$) เป็นค่าที่ได้จากประชากรแต่ส่วนเหลือ ($e$) เป็นความคลาดเคลื่อนที่ได้จากตัวอย่าง

### 2.1.1 ตัวประมาณค่าพารามิเตอร์ $\beta_0$ และ $\beta_1$

กำหนดให้ $Q$ เป็นค่าผลรวมกำลังสองของความคลาดเคลื่อนที่น้อยที่สุดหรือ

$$Q = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2 \tag{2.4}$$

ในการคำนวณเพื่อให้ได้ค่า $Q$ ที่น้อยที่สุดนั้นต้องหาอนุพันธ์ย่อยเทียบกับค่า $\beta_0$ และ $\beta_1$ (Abraham & Ledolter, 2006, p.28-29) โดยสามารถเขียนสมการทั้งสองได้ดังนี้

$$\frac{\partial Q}{\partial \beta_0} = -2\sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)$$

และ

$$\frac{\partial Q}{\partial \beta_1} = -2\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)X_i$$

จากนั้นกำหนดให้สมการทั้งสองมีค่าเท่ากับ 0 จะได้ค่าประมาณของ $\beta_0$ และ $\beta_1$ โดยใช้ $b_0$ และ $b_1$ แทนค่าประมาณดังกล่าวตามลำดับดังนี้

$$-2\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i) = 0$$

และ

$$-2\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)X_i = 0$$

โดยการหารทั้งสองสมการด้วย 2 จากนั้นกระจายผลบวกและย้ายข้างจะได้สมการปกติ (normal equation) ดังสมการ (2.5) และ (2.6)

$$nb_0 + b_1\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i \tag{2.5}$$

และ

$$b_0\sum_{i=1}^{n} X_i + b_1\sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i Y_i \tag{2.6}$$

จากการแก้สมการปกติทั้งสองจะได้

$$b_0 = \bar{Y} - b_1\bar{X} \tag{2.7}$$

และ

$$b_1 = \frac{\sum_{i=1}^{n} Y_i(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{S_{xy}}{S_{xx}} \tag{2.8}$$

โดย

$$S_{xx} = \sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n} \tag{2.9}$$

และ

$$S_{xy} = \sum_{i=1}^{n} Y_i(X_i - \bar{X}) = \sum_{i=1}^{n} X_i Y_i - \frac{\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{n} \tag{2.10}$$

**ตัวอย่างที่ 2.1** จากข้อมูลในตัวอย่างที่ 1.1 จงสร้างสมการถดถอยโดยวิธีกำลังสองน้อยที่สุด

**วิธีทำ**

จากข้อมูลจะสามารถคำนวณค่า $b_0$ และ $b_1$ ดังนี้

$$S_{xx} = \sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n} = 7{,}733.41 - \frac{(299.41)^2}{13} = 837.54$$

และ

$$S_{xy} = \sum_{i=1}^{n} X_i Y_i - \frac{\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{n} = 62{,}978.59 - \frac{299.41 \times 2{,}435.50}{13} = 6{,}885.28$$

ดังนั้นจากสมการ (2.6) และ (2.7) จะได้

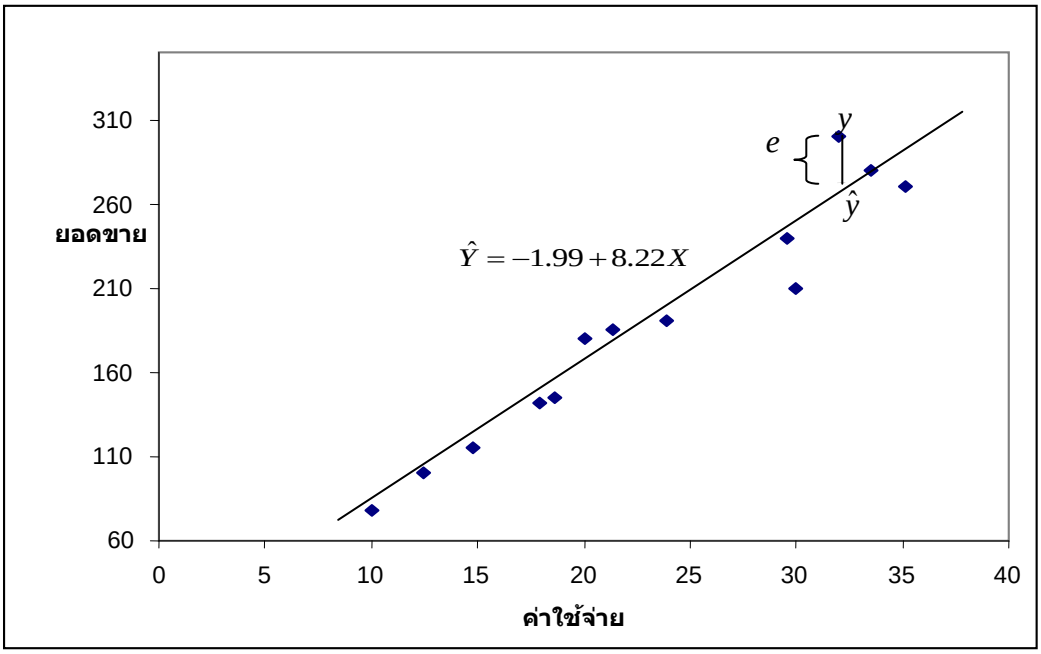
$$b_1 = \frac{S_{xy}}{S_{xx}} = \frac{6{,}885.28}{837.54} = 8.22$$

และ

$$b_0 = \bar{Y} - b_1\bar{X} = 187.35 - 8.22 \times 23.03 = -1.99$$

ดังนั้นสมการถดถอยคือ

$$\hat{Y}_i = -1.99 + 8.22X_i$$



จากสมการถดถอยที่ได้สามารถอธิบายความหมายของค่า $b_0$ และ $b_1$ ได้โดย $b_1$ เท่ากับ 8.22 หมายถึงเมื่อค่าใช้จ่ายในการโฆษณาเพิ่มขึ้น 1 ล้านบาทจะทำให้ยอดขายเฉลี่ยเพิ่มขึ้น 8.22 ล้านบาท สำหรับการอธิบายความหมายของ $b_0$ นั้นไม่สมควรจะตีความเนื่องจากค่าของข้อมูลไม่ครอบคลุมค่าใช้จ่ายในการโฆษณาเท่ากับ 0 หรือเมื่อไม่มีการโฆษณานั่นเอง ค่าพยากรณ์ของ $Y$ หรือ $\hat{Y}$ ที่ $X$ แต่ละค่าแสดงดังคอลัมน์ที่ 3 ของตารางข้างล่างและส่วนเหลือ ($e$) ได้ดังคอลัมน์ที่ 4

เมื่อมีค่าใช้จ่ายในการโฆษณา 10.01 ล้านบาทแล้วจะมียอดขายเฉลี่ย 80.30 ล้านบาทหรือ

$$\hat{y} = -1.99 + 8.22 \times 10.01 = 80.30$$

และมีค่าคลาดเคลื่อนเท่ากับ $y$ - $\hat{y}$ หรือ 77.60 – 80.30 = - 2.70

| $X_i$ | $Y_i$ | $\hat{Y}_i$ | $e_i$ |
|---|---|---|---|
| 10.01 | 77.60 | 80.30 | -2.70 |
| 14.75 | 114.90 | 119.26 | -4.36 |
| 17.94 | 141.40 | 145.49 | -4.09 |
| 23.93 | 190.80 | 194.73 | -3.93 |
| 29.61 | 239.90 | 241.43 | -1.53 |
| 35.18 | 270.00 | 287.22 | -17.22 |
| 33.50 | 280.00 | 273.41 | 6.59 |
| 12.45 | 100.50 | 100.36 | 0.14 |
| 20.04 | 180.40 | 162.75 | 17.65 |
| 18.60 | 145.00 | 150.92 | -5.92 |
| 30.00 | 210.00 | 244.63 | -34.63 |
| 32.00 | 300.00 | 261.07 | 38.93 |
| 21.40 | 185.00 | 173.93 | 11.07 |
| $\sum x_i = 299.41$ | $\sum y_i = 2{,}435.50$ | $\sum \hat{y}_i = 2{,}435.50$ | $\sum e_i = 0.00$ |

หากใช้โปรแกรม MINITABช่วยในการคำนวณสามารถทำได้ดังนี้

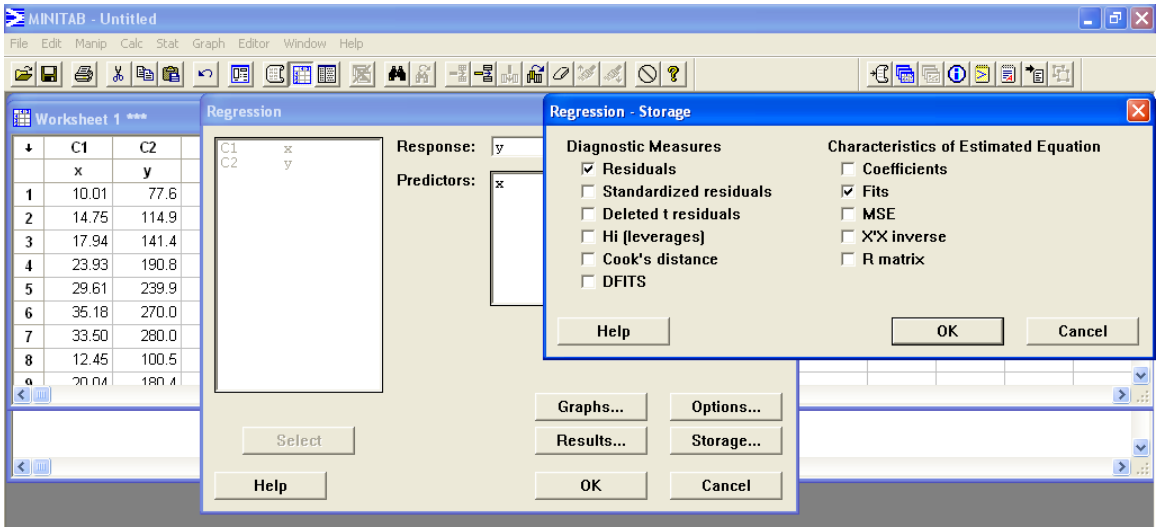
1. เลือก “Stat” ที่เมนูบาร์
2. เลือก “Regression”
3. เลือก “Regression…”
4. ระบุตัวแปรตามใน “Response:” และตัวแปรอิสระใน “Predictors:”

5. หากต้องการให้ทดสอบข้อตกลงให้เลือกเข้าไประบุที่ “Options”

6. หากต้องการให้เก็บค่าคลาดเคลื่อนและชนิดต่างๆ และค่าพยากรณ์ให้เข้าไประบุที่ “Storage”

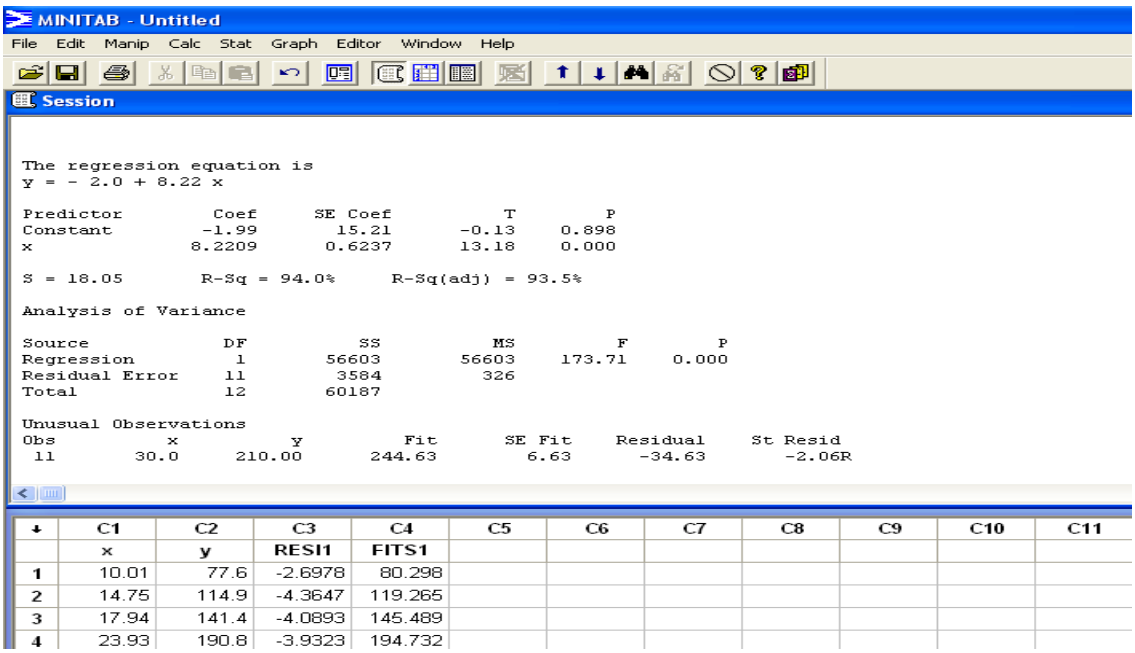
7. หากต้องการให้แสดงผลลัพธ์ต่างๆ เช่น ตาราง ANOVA และค่าสัมประสิทธิ์การตัดสินใจ ($R^2$) เป็นต้นให้เข้าไประบุที่ “Results”

8. หากต้องการให้แสดงกราฟชนิดต่างๆ ให้เข้าไประบุที่ “Graphs” จากนั้นคลิก “OK” ดังภาพที่ 2.1




ภาพที่ 2.1 หน้าจอการวิเคราะห์การถดถอย

จากตัวอย่างที่ 2.1 หากใช้โปรแกรม MINITAB ช่วยในการคำนวณแล้วจะได้ผลลัพธ์ดังภาพที่ 2.2 โดยในหน้าต่าง “Session” แสดงสมการถดถอยในบรรทัดแรกและส่วนถัดมาแสดงค่า $b_0$ และ $b_1$ พร้อมทั้งการทดสอบและค่า $p$-value (จะอธิบายในลำดับต่อๆ ไป) ส่วนถัดมาเป็นตาราง ANOVA (จะอธิบายในลำดับต่อๆ ไป) อีกทั้งแสดงค่าที่โปรแกรมสงสัยว่าจะเป็นค่าที่ผิดปกติ (outlier) สำหรับในหน้าต่าง “Worksheet” แสดงค่าส่วนเหลือ (Res1) และค่าพยากรณ์ (FITS1)




ภาพที่ 2.2 หน้าจอผลลัพธ์ของการวิเคราะห์การถดถอย

### 2.1.2 คุณสมบัติของตัวประมาณค่า $b_0$ และ $b_1$

โดยทฤษฎีของเกาส์-มาร์คอฟ (Guass-Markov theorem) จะได้ว่าตัวประมาณค่า $b_0$ และ $b_1$ ที่ได้จากวิธีกำลังสองน้อยที่สุดเป็นตัวประมาณค่าที่ไม่เอนเอียง (unbiased) หรือ $E(b_0) = \beta_0$ และ $E(b_1) = \beta_1$ และมีความแปรปรวนน้อยที่สุดในบรรดาตัวประมาณค่าเชิงเส้นที่ไม่เอนเอียง (unbiased linear estimator) หรืออาจเรียกตัวประมาณค่าทั้งสองตัวนี้ว่า best linear unbiased estimator (BLUE) โดยที่ best หมายถึงการที่มีความแปรปรวนที่น้อยที่สุด โดยค่าความแปรปรวนของ $b_0$ และ $b_1$ มีค่าดังนี้

$$V(b_0) = \sigma^2\left(\frac{1}{n} + \frac{\overline{X}^2}{S_{xx}}\right) \tag{2.11}$$

และ

$$V(b_1) = \frac{\sigma^2}{S_{xx}} \tag{2.12}$$

โดย $\sigma^2$ = ความแปรปรวนของความคลาดเคลื่อน ($\varepsilon$)

### 2.1.3 การประมาณค่าความแปรปรวน

เนื่องจากค่าความแปรปรวน ($\sigma^2$) เป็นค่าที่ไม่ทราบค่าจึงต้องทำการประมาณค่า $\sigma^2$ โดยใช้ค่าเฉลี่ยกำลังสองความคลาดเคลื่อน (mean square error หรือ *MSE*) การคำนวณหาค่า *MSE* เริ่มจากการหาผลรวมกำลังสองความคลาดเคลื่อน (error sum of squares หรือ *SSE*) ทำได้ดังนี้

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \tag{2.13}$$

หรือสามารถเขียน *SSE* ในรูปของ $S_{yy}$ และ $S_{xy}$ ดังนี้

$$SSE = S_{yy} - b_1 S_{xy} \tag{2.14}$$

โดยที่ $S_{yy} = \sum_{i=1}^{n} Y_i^2 - \dfrac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}$

และ $S_{xy} = \sum_{i=1}^{n} X_i Y_i - \dfrac{\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{n}$

ในการคำนวณค่าความแปรปรวนของตัวอย่างทำโดยการหารด้วยองศาเสรี โดยองศาเสรีของ *SSE* เท่ากับ $n$ - 2 เนื่องจากการสูญเสียองศาเสรีไป 2 ค่าในการประมาณค่าพารามิเตอร์ $\beta_0$ และ $\beta_1$ ด้วย $b_0$ และ $b_1$ ดังนั้น

$$MSE = \hat{\sigma}^2 = \frac{SSE}{n-2} \tag{2.15}$$

นอกจากนี้พบว่า *MSE* เป็นตัวประมาณค่าที่ไม่เอนเอียงของ $\sigma^2$ สำหรับตัวประมาณค่าของส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อน ($\sigma$) แล้วสามารถใช้ $\sqrt{MSE}$ เป็นตัวประมาณค่า หนังสือบางเล่มอาจเรียกค่า $\sqrt{MSE}$ ว่าค่าความคลาดเคลื่อนมาตรฐานของการถดถอย (standard error of regression) ซึ่งจะง่ายในการตีความมากกว่าความแปรปรวนเนื่องจากมีหน่วยเดียวกันกับค่าของตัวแปรตาม *Y*

**ตัวอย่างที่ 2.2** จากตัวอย่างที่ 2.1 จงประมาณค่าของ $\sigma^2$

**วิธีทำ**

เนื่องจากตัวประมาณค่าของ $\sigma^2$ คือ *MSE* สามารถคำนวณได้ดังนี้

$$S_{yy} = \sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}$$

$$= 516{,}468.79 - \frac{(2{,}435.50)^2}{13} = 60{,}187.23$$

$S_{xy} = 6{,}885.28$ (จากตัวอย่าง 2.1)

ดังนั้น

$$SSE = S_{yy} - b_1 S_{xy}$$

$$= 60{,}187.23 - 8.22 \times 6{,}885.28 = 3584.26$$

เนื่องจาก *SSE* มีองศาเสรีเท่ากับ $13 - 2 = 11$ ดังนั้น

$$MSE = \frac{SSE}{n-2}$$

$$= \frac{3584.26}{13-2} = 325.84$$

และค่าคลาดเคลื่อนมาตรฐานของการถดถอย ( $\sqrt{MSE}$ ) เท่ากับ $\sqrt{325.84} = 18.05$ ล้านบาท สามารถอธิบายความหมายของ $\sqrt{MSE}$ ได้โดยพิจารณาที่ค่าใช้จ่าย ($X$) เท่ากับ 10.01 ล้านบาทแล้วพบว่าค่าเฉลี่ยของการแจกแจงของยอดขายมีค่าประมาณ 80.30 ล้านบาท เมื่อพิจารณาความแปรผันของยอดขายที่ค่าใช้จ่ายเท่ากับ 10.01 ล้านบาทแล้วพบว่าความแปรผันของยอดขายนี้ค่อนข้างมาก (18.05 ล้านบาท) เมื่อเทียบกับค่าเฉลี่ยของยอดขายคือ 80.30 ล้านบาท

หากพิจารณาที่หน้า “Session” ของ output ในตัวอย่างที่ 2.1 จะเห็นว่าค่าองศาเสรี *SSE* และ *MSE* จะอยู่ในบรรทัดที่ 2 (Residual error) ของตาราง ANOVA และค่า $\sqrt{MSE}$ คือค่า $S$ ที่อยู่ระหว่างส่วนของค่าประมาณ $b_0$ และ $b_1$ กับตาราง ANOVA

### 2.1.4 คุณสมบัติของค่าพยากรณ์และส่วนเหลือ

ค่าพยากรณ์ ( $\hat{Y}$ ) และส่วนเหลือ ($e$) ที่ได้จากวิธีกำลังสองน้อยที่สุดนั้นมีคุณสมบัติดังนี้

(1) ผลรวมของส่วนเหลือเท่ากับ 0

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i) = \sum_{i=1}^{n} e_i = 0$$

คุณสมบัตินี้มาจากสมการปกติ (2.5) ดังในคอลัมน์ที่ 4 ของตารางในตัวอย่าง 2.1 แต่บางครั้งจะพบว่าผลรวมอาจไม่เท่ากับ 0.00 ทั้งนี้อาจเนื่องมาจากการปัดเศษก็ได้

(2) ผลรวมของค่าจริง $Y$ เท่ากับผลรวมของค่าพยากรณ์ $\hat{Y}$

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i$$

จากคอลัมน์ที่ 2 และ 3 ของตารางในตัวอย่างที่ 2.1 จะเห็นว่าผลรวมของทั้งสองเท่ากัน

(3) ผลรวมของส่วนเหลือกำลังสอง ( $\sum_{i=1}^{n} e_i^2$ ) โดยวิธีนี้มีค่าน้อยที่สุดเมื่อเปรียบเทียบกับค่าที่ได้จากวิธีอื่นๆ

(4) เส้นถดถอยจะลากผ่านจุดที่เป็นค่ากลาง $(\overline{X},\overline{Y})$เสมอ

(5) ผลรวมของส่วนเหลือที่ถ่วงน้ำหนักด้วยค่า $X$ จะมีค่าเท่ากับ 0

$$\sum_{i=1}^{n} X_i e_i = 0$$

(6) ผลรวมของส่วนเหลือที่ถ่วงน้ำหนักด้วยค่าพยากรณ์จะมีค่าเท่ากับ 0

$$\sum_{i=1}^{n} \hat{Y}_i e_i = 0$$

### 2.1.5 ตัวแบบถดถอยที่ลบออกด้วยค่ากลาง

ตัวแบบถดถอยนี้สามารถทำได้ง่ายๆ โดยการหักลบค่าตัวแปรอิสระออกด้วยค่าเฉลี่ยของตัวแปรอิสระหรือ $(X_i - \overline{X}_i)$ ค่า $X$ ชนิดนี้เรียก centered $X$ สมการถดถอยแสดงได้โดยการเพิ่มค่า $\beta_1\overline{X}$ และ - $\beta_1\overline{X}$ เข้าไปในสมการถดถอยดังนี้

$$\begin{aligned} Y_i &= \beta_0 + \beta_1(X_1 - \overline{X}) + \beta_1\overline{X} + \varepsilon_i \\ &= (\beta_0 + \beta_1\overline{X}) + \beta_1(X_1 - \overline{X}) + \varepsilon_i \\ &= \beta_0' + \beta_1(X_1 - \overline{X}) + \varepsilon_i \end{aligned} \tag{2.16}$$

โดย $\beta_0' = \beta_0 + \beta_1\overline{X}$ หลังจากประมาณค่าพารามิเตอร์พบว่าจุดเริ่มต้นของข้อมูลในตัวแบบนี้จะอยู่ที่ค่าเฉลี่ย $(\overline{X},\overline{Y})$ สามารถแสดงได้โดยพิจารณาจากสมการ (2.7) ดังนี้

$$b_0' = b_0 + b_1\overline{X} = (\overline{Y} - b_1\overline{X}) + b_1\overline{X} = \overline{Y} \tag{2.17}$$

$$b_1 = \frac{\sum_{i=1}^{n} Y_i(X_i - \overline{X})}{\sum_{i=1}^{n} (X_i - \overline{X})^2} = \frac{S_{xy}}{S_{xx}} \tag{2.18}$$

จากสมการ (2.17) จะเห็นว่าจุดตัดแกน $Y$ จะอยู่ที่ค่าเฉลี่ยของ $Y$ และยังพบอีกว่าค่าความชัน ($b_1$) ไม่เปลี่ยนแปลงดังนั้นค่าพยากรณ์สามารถเขียนได้ดังนี้

$$\hat{Y}_i = \overline{Y} + b_1(X_i - \overline{X}) \tag{2.19}$$

ค่าพยากรณ์ที่ได้โดยวิธีนี้จะเท่ากับค่าพยากรณ์โดยวิธีที่กล่าวมาแล้วแต่อาจแตกต่างกันได้เล็กน้อยเนื่องจากการปัดเศษ ข้อดีของวิธีนี้คือสมการปกติคำนวณได้ง่ายกว่าสมการ (2.7)

เพราะค่า $b_0$ คือค่าเฉลี่ยของตัวแปรตามนั่นเอง นอกจากนี้ค่า $b_0$ และ $b_1$ เป็นอิสระต่อกันหรือ COV( $b_0$, $b_1$) = 0 ทำให้การคำนวณช่วงความเชื่อมั่นของค่าคาดหวังของ $Y$ ง่ายขึ้น

**ตัวอย่าง 2.3** จากข้อมูลในตัวอย่าง 2.1 จงสร้างสมการถดถอยโดยการลบออกด้วยค่ากลาง

**วิธีทำ**

จากตัวอย่างที่ 2.1 พบว่า $\bar{x}$ = 23.03, $\bar{y}$ = 187.35 และ $b_1$ = 8.22 ดังนั้นสมการถดถอยคือ

$$\hat{Y}_i = \bar{Y} + b_1(X_i - \bar{X})$$
$$= 187.35 + 8.22 \times (X - 23.03)$$

พิจารณาค่าพยากรณ์ของข้อมูลแรก, x = 10.01 จะได้

$$\hat{y} = 187.35 + 8.22 \times (10.01 - 23.03)$$
$$= 80.33$$

พบว่ามีค่าใกล้เคียงกับค่าที่ได้ในตัวอย่าง 2.1 (80.30) ค่าที่ได้แตกต่างกันเล็กน้อยเนื่องจากการปัดเศษ

## 2.2 การอนุมานของตัวประมาณค่าของสมการถดถอย

ในหัวข้อ 2.1 ได้กล่าวถึงการประมาณค่าพารามิเตอร์ $\beta_0$ และ $\beta_1$ แบบจุดโดยใช้สมการ (2.7) และ (2.8) บางครั้งนักวิจัยอาจสนใจในการอนุมานค่าหรือต้องการประมาณค่าแบบช่วงและทดสอบสมมติฐานของค่าพารามิเตอร์ทั้งสอง

### 2.2.1 การประมาณค่าแบบช่วงของตัวประมาณค่าของสมการถดถอย

ในหัวข้อนี้จะกล่าวถึงการสร้างช่วงความเชื่อมั่นหรือการประมาณค่าแบบช่วงของค่าพารามิเตอร์ $\beta_0$ และ $\beta_1$

**2.2.1.1 การแจกแจงของตัวประมาณค่า $b_0$ และ $b_1$** การสร้างช่วงความเชื่อมั่นของตัวประมาณค่านั้นจำเป็นต้องทราบถึงการแจกแจงของตัวประมาณค่าทั้งสองก่อนเพื่อให้เข้าใจการอนุมานได้ดีขึ้น การใช้วิธีกำลังสองน้อยที่สุดในการสร้างสมการถดถอยมีข้อตกลงว่าค่าคลาดเคลื่อนต้องมีการแจกแจงแบบปกติ จากข้อตกลงนี้ตัวแปรตาม $Y$ และค่าประมาณ $b_0$ และ $b_1$ มีการแจกแจงแบบปกติ

$b_0$ มีค่าเฉลี่ยเท่ากับ $\beta_0$ และมีความแปรปรวนเท่ากับ $\sigma^2\left[\frac{1}{n} + \frac{\bar{X}^2}{S_{xx}}\right]$ หรือสามารถเขียนได้ในรูป $b_0 \sim N\left(\beta_0, \sigma^2\left[\frac{1}{n} + \frac{\bar{X}^2}{S_{xx}}\right]\right)$ จะเห็นว่า $b_0$ เป็นตัวประมาณค่าที่ไม่เอนเอียง (unbiased

estimator) เนื่องจากค่าเฉลี่ยมีค่าเท่ากับค่าพารามิเตอร์ นอกจากนี้หากพิจารณาค่าความแปรปรวนจะเห็นว่าค่าความแปรปรวนของค่าประมาณนี้ขึ้นกับขนาดของข้อมูล ($n$) ที่นำมาสร้างสมการถดถอยและขอบเขตของตัวแปรอิสระ $X$ โดยความแปรปรวนของค่าประมาณ $b_0$ จะลดลงเมื่อขนาดตัวอย่างมากขึ้นเมื่อกำหนดให้ขอบเขตของ $X$ คงที่และความแปรปรวนของค่าประมาณ $b_0$ จะลดลงเมื่อขอบเขตของ $X$ มากขึ้นเมื่อกำหนดให้ขนาดตัวอย่างคงที่

$b_1$ มีค่าเฉลี่ยเท่ากับ $\beta_1$ และมีความแปรปรวนเท่ากับ $\frac{\sigma^2}{S_{xx}}$ หรือสามารถเขียนได้ในรูป $b_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$ จะเห็นว่า $b_1$ เป็นตัวประมาณค่าที่ไม่เอนเอียงเช่นเดียวกับ $b_0$ แต่ค่าความแปรปรวนของ $b_1$ ไม่ขึ้นกับขนาดของตัวอย่างแต่ขึ้นอยู่กับขอบเขตของตัวแปรอิสระ $X$

ในการคำนวณค่าความแปรปรวนของค่าประมาณ $b_0$ และ $b_1$ นั้นต้องมีการประมาณค่า $\sigma^2$ โดยใช้ค่า $MSE$ ดังนั้นความแปรปรวนของค่าประมาณ $b_0$ และ $b_1$ คือ

$$S^2(b_0) = MSE\left[\frac{1}{n} + \frac{\overline{X}^2}{S_{xx}}\right] \tag{2.20}$$

และ

$$S^2(b_1) = \frac{MSE}{S_{xx}} \tag{2.21}$$

หากนำค่าความแปรปรวนของตัวประมาณค่าทั้งสองมาถอดรากจะได้ค่าความคลาดเคลื่อนมาตรฐาน (standard error) ของ $b_0$ และ $b_1$ หรือ $se(b_0)$ และ $se(b_1)$ ค่าความแปรปรวนหรือค่าคลาดเคลื่อนมาตรฐานจะบอกถึงความถูกต้องแม่นยำในการประมาณค่าทั้งสอง

**2.2.1.2 ช่วงความเชื่อมั่นของตัวประมาณค่า $b_0$ และ $b_1$** ช่วงความเชื่อมั่นของตัวประมาณค่าจะบอกถึงคุณภาพของตัวประมาณค่า ตัวประมาณค่าที่ดีควรมีช่วงความเชื่อมั่นที่แคบ เนื่องจาก $b_0$ และ $b_1$ มีการแจกแจงแบบปกติดังนั้นหากนำค่าประมาณแต่ละค่ามาลบด้วยค่าเฉลี่ยและหารด้วยส่วนเบี่ยงเบนมาตรฐานแล้วจะได้การแจกแจงทีที่มีองศาเสรีเท่ากับ $n-2$ ดังนี้

$$\frac{b_0 - \beta_0}{\sqrt{MSE\left[\frac{1}{n} + \frac{\overline{X}^2}{S_{xx}}\right]}} \sim t_{n-2}$$

และ

$$\frac{b_1 - \beta_1}{\sqrt{\frac{MSE}{S_{xx}}}} \sim t_{n-2}$$

ดังนั้นช่วงความเชื่อมั่น 100(1- $\alpha$)% ของ $\beta_0$ คือ

$$b_0 - t_{\alpha/2,n-2}\sqrt{MSE\left(\frac{1}{n}+\frac{\overline{X}^2}{S_{xx}}\right)} \le \beta_0 \le b_0 + t_{\alpha/2,n-2}\sqrt{MSE\left(\frac{1}{n}+\frac{\overline{X}^2}{S_{xx}}\right)} \qquad (2.22)$$

และช่วงความเชื่อมั่น 100(1- $\alpha$)% ของ $\beta_1$ คือ

$$b_1 - t_{\alpha/2,n-2}\sqrt{\frac{MSE}{S_{xx}}} \le \beta_1 \le b_1 - t_{\alpha/2,n-2}\sqrt{\frac{MSE}{S_{xx}}} \qquad (2.23)$$

การอธิบายความหมายของช่วงความเชื่อมั่นทั้งสองสามารถอธิบายได้ดังนี้คือ หากสุ่มตัวอย่าง 100 ครั้งโดยให้มีขนาดเท่าเดิมทุกครั้งและที่ค่า $X$ ชุดเดียวกันแล้วนำมาสร้างช่วงความเชื่อมั่น 95% ของค่า $\beta_0$ แล้วพบว่า 95 ช่วงใน 100 ช่วงที่ได้ครอบคลุมค่าที่แท้จริงของ $\beta_0$ ในกรณีของ $\beta_1$ สามารถอธิบายได้เช่นเดียวกัน

**ตัวอย่างที่ 2.4** จากข้อมูลในตัวอย่าง 2.1 จงสร้างช่วงความเชื่อมั่น 95% ของค่า $\beta_0$ และ $\beta_1$

**วิธีทำ**

ความคลาดเคลื่อนมาตรฐานของ $\beta_0$ เท่ากับ

$$se(b_0) = \sqrt{MSE\left[\frac{1}{n}+\frac{\overline{X}^2}{S_{xz}}\right]}$$

$$= \sqrt{325.84\left[\frac{1}{13}+\frac{23.03^2}{837.54}\right]} = 15.21$$

และความคลาดเคลื่อนมาตรฐานของ $\beta_1$ เท่ากับ

$$se(b_1) = \sqrt{\frac{MSE}{S_{xx}}}$$

$$= \sqrt{\frac{325.84}{837.54}} = 0.624$$

ค่า $t$ ได้จากตารางที่ 1 ในภาคผนวกโดย $t_{\alpha/2,\,n-2} = t_{0.025,\,13} = 2.201$ ดังนั้นช่วงความเชื่อมั่น 95% ของ $\beta_0$ เท่ากับ

$$b_0 - t_{\alpha/2,n-2}\sqrt{MSE\left(\frac{1}{n}+\frac{\overline{X}^2}{S_{xx}}\right)} \le \beta_0 \le b_0 + t_{\alpha/2,n-2}\sqrt{MSE\left(\frac{1}{n}+\frac{\overline{X}^2}{S_{xx}}\right)}$$

$$-1.99 - 2.201\times 15.21 \le \beta_0 \le -1.99 + 2.201\times 15.21$$

$$-35.472 \le \beta_0 \le 31.492$$

และช่วงความเชื่อมั่น 95% ของ $\beta_1$ เท่ากับ

$$b_1 - t_{\alpha/2,n-2}\sqrt{\frac{MSE}{S_{xx}}} \le \beta_1 \le b_1 - t_{\alpha/2,n-2}\sqrt{\frac{MSE}{S_{xx}}}$$

$$8.22 - 2.201 \times 0.624 \le \beta_1 \le 8.22 + 2.201 \times 0.624$$

$$6.847 \le \beta_1 \le 9.593$$

ดังนั้นเชื่อมั่นได้ 95% ว่ายอดขายเฉลี่ยมีค่าเพิ่มขึ้นอยู่ระหว่าง 6.847 และ 9.593 ล้านบาทเมื่อเพิ่มค่าใช้จ่ายในการโฆษณาขึ้น 1 ล้านบาท แต่เนื่องจากในโจทย์นี้อาจไม่เหมาะสมที่จะตีความหมายของช่วงความเชื่อมั่นของ $\beta_0$ เนื่องจากข้อมูลมิได้ครอบคลุมค่า x ที่เท่ากับ 0 หรือมิได้ครอบคลุมเมื่อไม่มีการโฆษณาเกิดขึ้นแต่แสดงให้ดูในที่นี้เพื่อเป็นตัวอย่างในการคำนวณเท่านั้น

### 2.2.2 การทดสอบสมมติฐานของค่าพารามิเตอร์ $\beta_0$ และ $\beta_1$

นอกเหนือจากการสร้างช่วงความเชื่อมั่นของค่าพารามิเตอร์แล้วการทดสอบสมมติฐานของค่าพารามิเตอร์ $\beta_0$ และ $\beta_1$ จากหัวข้อ 2.2.1.2 กล่าวว่า $(b_0 - \beta_0)/\ se(b_0)$ และ $(b_1 - \beta_1)/\ se(b_1)$ มีการแจกแจงแบบ $t$ มีองศาเสรีเท่ากับ $n - 2$

การทดสอบสมมติฐานของค่า $\beta_0$ และ $\beta_1$ กับค่าใดๆ ($\beta_{00}$ และ $\beta_{10}$) สามารถทดสอบได้ทั้งหางเดียวและสองหาง การทดสอบทำขึ้นเพื่อให้ทราบว่าค่าพารามิเตอร์ทั้งสองมีค่าแตกต่างจากค่าที่สนใจหรือไม่ โดยในกรณีของ $\beta_0$ จะทดสอบว่าที่ค่า $x = 0$ แล้วค่า y มีค่าเท่ากับค่าที่สนใจ ($\beta_{00}$) หรือไม่และในกรณีของ $\beta_1$ เป็นการทดสอบว่าค่าความชันเท่ากับค่าที่สนใจ ($\beta_{10}$) หรือไม่ โดยมีขั้นตอนดังนี้

(1) ตั้งสมมติฐาน

กรณีของ $\beta_0$

$H_0$: $\beta_0 = \beta_{00}$

$H_1$: $\beta_0 \neq \beta_{00}$ (กรณีทดสอบสองหาง)

หรือ $H_1$: $\beta_0 < \beta_{00}$ (กรณีทดสอบหางซ้าย)

หรือ $H_1$: $\beta_0 > \beta_{00}$ (กรณีทดสอบหางขวา)

กรณีของ $\beta_1$

$H_0$: $\beta_1 = \beta_{10}$

$H_1$: $\beta_1 \neq \beta_{10}$ (กรณีทดสอบสองหาง)

หรือ $H_1$: $\beta_1 < \beta_{10}$ (กรณีทดสอบหางซ้าย)

หรือ $H_1$: $\beta_1 > \beta_{10}$ (กรณีทดสอบหางขวา)

(2) กำหนดระดับนัยสำคัญ ($\alpha$)

การกำหนดระดับนัยสำคัญเป็นการกำหนดโอกาสในการตัดสินใจผิดพลาดแบบที่ 1

(3) ค่าวิกฤต (critical value)

ค่าวิกฤตของทั้ง $\beta_0$ และ $\beta_1$ จะเหมือนกัน ในกรณีของการทดสอบสองหางแล้วค่าวิกฤตของการทดสอบทั้งสองคือ $t_{\alpha/2,\,n-2}$ สำหรับกรณีของการทดสอบหางซ้ายค่าวิกฤตคือ $-t_{\alpha,\,n-2}$ และกรณีของการทดสอบหางขวาค่าวิกฤตคือ $t_{\alpha,\,n-2}$ โดยที่ค่าวิกฤตได้จากการเปิดตารางสถิติ $t$ ในภาคผนวก

(4) ค่าสถิติ

กรณีของ $\beta_0$

$$t = \frac{(b_0 - \beta_0)}{\sqrt{MSE\left(\frac{1}{n} + \frac{\bar{X}^2}{S_{xx}}\right)}} \tag{2.24}$$

กรณีของ $\beta_1$

$$t = \frac{(b_1 - \beta_1)}{\sqrt{\frac{MSE}{S_{xx}}}} \tag{2.25}$$

(5) เปรียบเทียบค่าสถิติกับค่าวิกฤตและทำการตัดสินใจ

กรณีการทดสอบสองหาง

หาก $|t| \le t_{\alpha/2,\,n-2}$ แล้วไม่ปฏิเสธสมมติฐานหลัก ($H_0$)

หาก $|t| > t_{\alpha/2,\,n-2}$ แล้วปฏิเสธสมมติฐานหลัก ($H_0$)

กรณีการทดสอบหางซ้าย

หาก $t \ge -t_{\alpha,\,n-2}$ แล้วไม่ปฏิเสธสมมติฐานหลัก ($H_0$)

หาก $t < -t_{\alpha,\,n-2}$ แล้วปฏิเสธสมมติฐานหลัก ($H_0$)

กรณีการทดสอบหางขวา

หาก $t \le t_{\alpha,\,n-2}$ แล้วไม่ปฏิเสธสมมติฐานหลัก ($H_0$)

หาก $t > t_{\alpha,\,n-2}$ แล้วปฏิเสธสมมติฐานหลัก ($H_0$)

**หมายเหตุ**

1. ส่วนใหญ่มักจะไม่ทดสอบค่า $\beta_0$ เนื่องจากส่วนใหญ่นักวิจัยไม่ได้เก็บข้อมูลที่ครอบคลุมค่า 0

2. หากทำการทดสอบ $\beta_1$ เทียบกับค่า 0 แล้วเป็นการทดสอบว่าตัวแปร $X$ และ $Y$ มีความสัมพันธ์เชิงเส้นตรงกันหรือไม่ หากปฏิเสธสมมติฐานหลักแสดงว่าตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นตรงกัน หรือสามารถกล่าวได้อีกแบบหนึ่งว่าตัวแปร $X$ สามารถอธิบายความแปรผันในตัวแปร $Y$ ได้

3. หากนักวิจัยทราบค่าความแปรปรวนของประชากร ($\sigma^2$) แล้วสมการ (2.24) และ (2.25) จะมีการแจกแจงแบบปกติมาตรฐาน ($Z$) ไม่ใช่การแจกแจงแบบ $t$ และแทนค่า $MSE$ ในสมการทั้งสองด้วย $\sigma^2$ และเปิดตาราง $Z$ แทนตาราง $t$ นอกจากนี้หากขนาดตัวอย่างใหญ่แล้วสามารถใช้การแจกแจงแบบ $Z$ ได้

4. นอกจากการตัดสินใจโดยการเปรียบเทียบจากค่าสถิติแล้วยังสามารถใช้ค่า $p$-value ในการตัดสินใจได้อีกด้วย โปรแกรมทางสถิติส่วนใหญ่จะแสดงค่า $p$-value บางครั้งเรียกค่า $p$-value ว่า observed level of significance

5. ค่า $p$-value คือ ค่าความน่าจะเป็นที่ได้ค่าสถิติภายใต้สมมติฐานหลัก ในการตัดสินใจสามารถทำได้โดยการเปรียบเทียบค่า $p$-value กับระดับนัยสำคัญ ($\alpha$) หาก $p$-value < $\alpha$ แล้วจะปฏิเสธสมมติฐานหลักและหาก $p$-value > $\alpha$ แล้วจะไม่ปฏิเสธสมมติฐานหลัก เช่น หากค่า $p$-value มีค่า 0.40 เมื่อค่า $t$ = 2.561 แสดงว่าโอกาสที่จะได้ค่าสถิติเท่ากับ 2.561 นั้นมีสูงถึง 40% หากสมมติฐานหลักเป็นจริง ดังนั้นหากกำหนด $\alpha$ = 0.05 แล้วจะไม่ปฏิเสธสมมติฐานหลัก เป็นต้น

6. การเปรียบเทียบค่า $p$-value กับค่า $\alpha$ จะเหมือนกันทั้งการทดสอบสองหางและหางเดียว

7. หากค่า $p$-value มีค่าต่ำกว่า 0.000 แล้วโปรแกรมสถิติส่วนใหญ่จะแสดงค่า $p$-value เป็น 0.000

**ตัวอย่าง 2.5** จากข้อมูลในตัวอย่าง 2.1 จงทดสอบว่า (1) $\beta_1$ = 9 หรือไม่และ (2) ตัวแปรทั้งสองมีความสัมพันธ์กันเชิงเส้นตรงหรือไม่โดยทดสอบที่ระดับนัยสำคัญ 0.05

**วิธีทำ**

(1) เนื่องจาก $\beta_{10} = 9$ และเป็นการทดสอบสองหางดังนั้นสมมติฐานคือ

$$H_0 : \beta_1 = 9$$

$$H_1 : \beta_1 \neq 9$$

ที่ระดับนัยสำคัญ ($\alpha$) 0.05 และ $n$ = 13 ดังนั้นค่าวิกฤตคือ $t_{\alpha/2,\, n-2} = t_{0.025,\ 11} = 2.201$ ดังนั้นค่าสถิติ $t$ คือ

$$t = \frac{(b_1 - \beta_1)}{\sqrt{\frac{MSE}{S_{xx}}}}$$

$$= \frac{(8.22 - 9)}{\sqrt{\frac{325.84}{837.54}}} = -1.251$$

เมื่อเปรียบเทียบค่าสถิติกับค่าวิกฤตพบว่า | -1.251 | < 2.201 จึงไม่ปฏิเสธสมมติฐานหลัก ($H_0$) นั่นคือค่าความชันของสมการถดถอยนี้เท่ากับ 9 ที่ระดับนัยสำคัญ 0.05

(2) เนื่องจากต้องการทดสอบว่าตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นตรงกันหรือไม่คือการทดสอบว่า $\beta_1 = 0$ หรือไม่ ดังนั้นสมมติฐานคือ

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

ค่าวิกฤตจะเท่ากับค่าวิกฤตในข้อ (1) คือ 2.201

ดังนั้นค่าสถิติ *t* คือ

$$t = \frac{b_1}{\sqrt{\frac{MSE}{S_{xx}}}}$$

$$= \frac{8.22}{\sqrt{\frac{325.84}{837.54}}} = 13.179$$

เมื่อเปรียบเทียบค่าสถิติกับค่าวิกฤตพบว่า | 13.179 | > 2.201 จึงปฏิเสธสมมติฐานหลัก ($H_0$) นั่นคือตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นตรงกันที่ระดับนัยสำคัญ 0.05

**หมายเหตุ**

หากเปรียบเทียบค่าสถิติที่ได้จาก (2) กับผลลัพธ์จาก MINITAB ในตัวอย่าง 2.1 โดยดูจากบรรทัดที่ 5 คอลัมน์ที่ 4 นอกจากนี้ MINITAB ยังแสดงค่า *p*-value ในคอลัมน์สุดท้ายอีกด้วย โดย *p*-value = 0.000 < $\alpha$ (0.05)

```
The regression equation is
y = - 2.0 + 8.22 x

Predictor        Coef     SE Coef          T        P
Constant        -1.99       15.21      -0.13    0.898
x              8.2209       0.6237      13.18    0.000
```

### 2.2.3 การประมาณค่าแบบช่วงของตัวพยากรณ์

นอกจากตัวประมาณค่าของพารามิเตอร์ทั้งสองแล้วบางครั้งนักวิจัยอาจสนใจตัวประมาณค่าอื่นๆ เช่น ค่าคาดหวังของตัวแปรตามหรือค่าพยากรณ์ เป็นต้น ตัวประมาณค่าที่จะกล่าวถึงในที่นี้คือ (1) ตัวประมาณค่าพยากรณ์ของค่าคาดหวังของ $Y$ หรือค่าเฉลี่ยของ $Y$ (mean response) หรือ $E(y|x_0)$ เมื่อ $x = x_0$ หรือเมื่อ x มีค่าใดๆ ที่อยู่ในช่วงของข้อมูลที่สนใจ (2) ตัวประมาณค่าพยากรณ์ของ $Y$ จำนวน 1 ค่าเมื่อ $x = x_0$ โดยที่ $x_0$ คือค่าที่ไม่ได้อยู่ข้อมูลและ (3) ตัวประมาณค่าเฉลี่ยของค่าพยากรณ์ของ $Y$ จำนวน $m$ ค่า ตัวประมาณค่าพยากรณ์ทั้งสามจะเขียนได้ในรูปของ $\hat{y}_0 = b_0 + b_1 x_0$ ความแตกต่างระหว่างค่าพยากรณ์ของค่าคาดหวังกับค่าพยากรณ์ของ $Y$ คือค่าพยากรณ์ของค่าคาดหวังเป็นการพยากรณ์ค่าเฉลี่ยของการแจกแจงของ $Y$ ที่ $x_0$ เช่น ค่าพยากรณ์ของยอดขายโดยเฉลี่ยที่ค่าใช้จ่ายเท่ากับ 8 ล้านบาท เป็นต้น แต่ค่าพยากรณ์ของค่า $Y$ เป็นการพยากรณ์ค่าของข้อมูลแต่ละตัว เช่น ค่าพยากรณ์ของยอดขายที่ค่าใช้จ่ายเท่ากับ 8 ล้านบาท เป็นต้น หากทำการพยากรณ์ ณ ค่า x ใดที่มากกว่า 1 ครั้งแล้วจะเป็นการพยากรณ์จำนวน $m$ ครั้งเมื่อ $m$ มีค่ามากกว่า 1 ซึ่งค่าที่ได้เป็นค่าเฉลี่ยของค่าพยากรณ์และโดยทฤษฎีแล้วเมื่อ $m$ มากขึ้นเรื่อยแล้วการพยากรณ์ค่าเฉลี่ยนี้จะเข้าสู่การพยากรณ์ค่าคาดหวัง

**2.2.3.1 การแจกแจงของตัวประมาณค่าพยากรณ์** ตัวประมาณค่าพยากรณ์ก็เช่นเดียวกับตัวประมาณค่าพารามิเตอร์ที่มีการแจกแจงแบบปกติแต่มีค่าเฉลี่ยและความแปรปรวนที่แตกต่างกันดังนี้

ตัวประมาณค่าพยากรณ์ของค่าคาดหวังของ $Y$ หรือ $E(y|x_0)$ มีค่าเฉลี่ยเท่ากับ $\beta_0 + \beta_1 X$ และความแปรปรวนเท่ากับ $\sigma^2\left[\frac{1}{n} + \frac{(x_0 - \bar{X})^2}{S_{xx}}\right]$ หรือสามารถเขียนได้ในรูป

$$\hat{y}_0 \sim N\left(\beta_0 + \beta_1 X, \sigma^2\left[\frac{1}{n} + \frac{(x_0 - \bar{X})^2}{S_{xx}}\right]\right)$$

ตัวประมาณค่าพยากรณ์ของ $Y$ จำนวน 1 ค่ามีค่าเฉลี่ยเท่ากับ $\beta_0 + \beta_1 X$ และความแปรปรวนเท่ากับ $\sigma^2\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{X})^2}{S_{xx}}\right]$ หรือสามารถเขียนได้ในรูป

$\hat{y}_0 \sim N\left(\beta_0 + \beta_1 X, \sigma^2\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{X})^2}{S_{xx}}\right]\right)$ เมื่อพิจารณาความแปรปรวนพบว่าความแปรปรวนของตัวประมาณค่าพยากรณ์ของ $Y$ นั้นมีค่ามากกว่าตัวประมาณค่าพยากรณ์ของค่าคาดหวังของ $Y$

ตัวประมาณค่าเฉลี่ยของค่าพยากรณ์ของ $Y$ จำนวน $m$ ค่ามีค่าเฉลี่ยเท่ากับ $\beta_0 + \beta_1 X$ และความแปรปรวนเท่ากับ $\sigma^2\left[\frac{1}{m}+\frac{1}{n}+\frac{(x_0-\bar{X})^2}{S_{xx}}\right]$ หรือสามารถเขียนได้ในรูป $\hat{y}_0 \sim N\left(\beta_0+\beta_1 X, \sigma^2\left[\frac{1}{m}+\frac{1}{n}+\frac{(x_0-\bar{X})^2}{S_{xx}}\right]\right)$ จะเห็นว่าความแปรปรวนนี้จะลดลงเมื่อจำนวนค่าพยากรณ์มากขึ้นที่ค่าที่ $X$ หนึ่งๆ นอกจากนี้หากจำนวนที่พยากรณ์มากๆ ($m$ เข้าใกล้ค่าอนันต์) แล้วความแปรปรวนนี้จะเข้าใกล้ความแปรปรวนของตัวประมาณค่าคาดหวังทั้งนี้ค่าคาดหวังคือค่าเฉลี่ยนั่นเอง

จะเห็นว่าตัวประมาณค่าพยากรณ์ทั้งสามเป็นตัวประมาณค่าที่ไม่เอนเอียง เนื่องจากมีค่าเฉลี่ยเท่ากับ $\beta_0 + \beta_1 X$ นอกจากนี้เมื่อค่า $X$ ที่ต้องการพยากรณ์เข้าใกล้ค่าเฉลี่ย ($\bar{X}$) ค่าความแปรปรวนของตัวประมาณค่าพยากรณ์จะลดลงและจะมีค่าน้อยสุดเมื่อทำการพยากรณ์ที่ค่าเฉลี่ยเนื่องจากค่า $(\bar{x}-\bar{x})^2 = 0$ นอกจากนี้ยังพบว่าหากทำการพยากรณ์ที่ $x_0 = 0$ แล้วค่าความแปรปรวนของตัวประมาณค่าพยากรณ์ของค่าคาดหวังจะกลายเป็นค่าความแปรปรวนของ $b_0$ นั่นเองดังในสมการ (2.10) ทั้งนี้เนื่องจากเมื่อพิจารณาสมการ $\hat{y}_0 = b_0 + b_1 x_0$ ที่ $x_0 = 0$ แล้วจะได้ $\hat{y}_0 = b_0$

**2.2.3.2 ช่วงความเชื่อมั่นของตัวประมาณค่าพยากรณ์** เนื่องจากตัวประมาณค่าทั้งสามมีการแจกแจงแบบปกติดังนั้นการสร้างช่วงความเชื่อมั่นจะคล้ายกับการสร้างช่วงความเชื่อมั่นของค่าพารามิเตอร์โดยใช้การแจกแจงแบบ $t$ ช่วยในการสร้างช่วงและใช้ $MSE$ ในการประมาณค่าของ $\sigma^2$ ดังนี้

ช่วงความเชื่อมั่น 100(1- $\alpha$)% ของค่าคาดหวังของ $Y$ หรือ $Y_0$ คือ

$$\hat{y}_0 - t_{\alpha/2,n-2}\sqrt{MSE\left(\frac{1}{n}+\frac{(x_0-\bar{X})^2}{S_{xx}}\right)} \le y_0 \le \hat{y}_0 + t_{\alpha/2,n-2}\sqrt{MSE\left(\frac{1}{n}+\frac{(x_0-\bar{X})^2}{S_{xx}}\right)} \qquad (2.26)$$

ช่วงความเชื่อมั่น 100(1- $\alpha$)% ของค่า $Y$ ใหม่ ณ $x_0$ ที่ไม่ได้อยู่ในข้อมูลจำนวน 1 ค่าหรือ $Y_0$ คือ

$$\hat{y}_0 - t_{\alpha/2,n-2}\sqrt{MSE\left(1+\frac{1}{n}+\frac{(x_0-\bar{X})^2}{S_{xx}}\right)} \le y_0 \le \hat{y}_0 + t_{\alpha/2,n-2}\sqrt{MSE\left(1+\frac{1}{n}+\frac{(x_0-\bar{X})^2}{S_{xx}}\right)} \qquad (2.27)$$

ช่วงความเชื่อมั่น 100(1- $\alpha$)% ของค่าเฉลี่ยของค่า $Y$ ใหม่ ณ $x_0$ ที่ไม่ได้อยู่ในข้อมูลจำนวน $m$ ค่าหรือ $Y_0$ คือ

$$\hat{y}_0 - t_{\alpha/2,n-2}\sqrt{MSE\left(\frac{1}{m}+\frac{1}{n}+\frac{(x_0-\bar{X})^2}{S_{xx}}\right)} \le y_0 \le \hat{y}_0 + t_{\alpha/2,n-2}\sqrt{MSE\left(\frac{1}{m}+\frac{1}{n}+\frac{(x_0-\bar{X})^2}{S_{xx}}\right)} \quad (2.28)$$

**ตัวอย่าง 2.6** จากข้อมูลในตัวอย่าง 2.1 จงสร้างช่วงความเชื่อมั่นของยอดขายที่ระดับความเชื่อมั่น 95% ของ (1) ค่าคาดหวังที่ค่าใช้จ่ายเท่ากับ 12 ล้านบาท (2) ค่าใช้จ่ายเท่ากับ 12 ล้านบาทโดยพยากรณ์เพียงค่าเดียว และ (3) ค่าใช้จ่ายเท่ากับ 12 ล้านบาทโดยพยากรณ์จำนวน 10 ค่า

**วิธีทำ**

(1) เมื่อ $x_0 = 12$ แล้ว $\hat{y}_0 = b_0 + b_1 x_0 = -1.99 + 8.22 \times 12 = 96.65$

ที่ระดับความเชื่อมั่น 95% จะมีค่า $t_{\alpha/2,\,n-2} = t_{0.025,\ 11} = 2.201$ ($t$ จากตาราง) และช่วงความเชื่อมั่นของค่าคาดหวังของยอดขายที่ค่าใช้จ่าย 12 ล้านบาทคือ

$$\hat{y}_0 - t_{\alpha/2,n-2}\sqrt{MSE\left(\frac{1}{n}+\frac{(x_0-\bar{X})^2}{S_{xx}}\right)} \le y_0 \le \hat{y}_0 + t_{\alpha/2,n-2}\sqrt{MSE\left(\frac{1}{n}+\frac{(x_0-\bar{X})^2}{S_{xx}}\right)}$$

$$96.65 - 2.201\sqrt{325.84\left(\frac{1}{13}+\frac{(12-23.03)^2}{837.54}\right)} \le y_0 \le 96.65 + 2.201\sqrt{325.84\left(\frac{1}{13}+\frac{(12-23.03)^2}{837.54}\right)}$$

$$77.923 \le y_0 \le 115.377$$

ดังนั้นเชื่อมั่นได้ 95% ว่ายอดขายเฉลี่ยที่ได้จะอยู่ระหว่าง 77.923 กับ 115.377 ล้านบาทหากใช้ค่าใช้จ่ายในการโฆษณาเท่ากับ 12 ล้านบาท

(2) จาก (1) ที่ $x_0 = 12$ แล้ว $\hat{y}_0 = 96.65$ และ $t_{0.025,\ 11} = 2.201$ และช่วงความเชื่อมั่นของยอดขายที่ค่าใช้จ่าย 12 ล้านบาทคือ

$$\hat{y}_0 - t_{\alpha/2,n-2}\sqrt{MSE\left(1+\frac{1}{n}+\frac{(x_0-\bar{X})^2}{S_{xx}}\right)} \le y_0 \le \hat{y}_0 + t_{\alpha/2,n-2}\sqrt{MSE\left(1+\frac{1}{n}+\frac{(x_0-\bar{X})^2}{S_{xx}}\right)}$$

$$96.65 - 2.201\sqrt{325.84\left(1+\frac{1}{13}+\frac{(12-23.03)^2}{837.54}\right)} \le y_0 \le 96.65 + 2.201\sqrt{325.84\left(1+\frac{1}{13}+\frac{(12-23.03)^2}{837.54}\right)}$$

$$52.727 \le y_0 \le 140.573$$

ดังนั้นเชื่อมั่นได้ 95%ว่ายอดขายที่ได้จะอยู่ระหว่าง 52.727 กับ 140.573 ล้านบาทหากใช้ค่าใช้จ่ายในการโฆษณาเท่ากับ 12 ล้านบาท

(3) จาก (1) ที่ $x_0 = 12$ แล้ว $\hat{y}_0 = 96.65$ และ $t_{0.025,\ 11} = 2.201$ และช่วงความเชื่อมั่นของยอดขายเฉลี่ยที่ค่าใช้จ่าย 12 ล้านบาทโดยใช้ค่าพยากรณ์ 10 ค่าคือ

$$\hat{y}_0 - t_{\alpha/2,n-2}\sqrt{MSE\left(\frac{1}{m}+\frac{1}{n}+\frac{(x_0-\bar{X})^2}{S_{xx}}\right)} \le y_0 \le \hat{y}_0 + t_{\alpha/2,n-2}\sqrt{MSE\left(\frac{1}{m}+\frac{1}{n}+\frac{(x_0-\bar{X})^2}{S_{xx}}\right)}$$

$$96.65 - 2.201\sqrt{325.84\left(\frac{1}{10}+\frac{1}{13}+\frac{(12-23.03)^2}{837.54}\right)} \le y_0 \le 96.65 + 2.201\sqrt{325.84\left(\frac{1}{10}+\frac{1}{13}+\frac{(12-23.03)^2}{837.54}\right)}$$

$$74.099 \le y_0 \le 119.201$$

ดังนั้นเชื่อมั่นได้ 95%ว่ายอดขายเฉลี่ยที่ได้จะอยู่ระหว่าง 74.099 กับ 119.201 ล้านบาทหากใช้ค่าใช้จ่ายในการโฆษณาเท่ากับ 12 ล้านบาทโดยทำการพยากรณ์ 10 ค่า

**หมายเหตุ**

หากเปรียบเทียบช่วงความเชื่อมั่นทั้งสามแบบจะเห็นว่าช่วงความเชื่อมั่นของค่าคาดหวังจะแคบสุดและช่วงความเชื่อมั่นของการพยากรณ์เพียงค่าเดียวจะกว้างสุดทั้งนี้เนื่องจากการพยากรณ์แบบค่าคาดหวังเป็นการพยากรณ์จำนวนมากแล้วคำนวณค่าเฉลี่ยดังนั้นความถูกต้องจึงมากกว่าทำให้ช่วงแคบกว่าทั้งสองแบบ

## 2.3 การวิเคราะห์ความแปรปรวน

ในการทดสอบสมมติฐานของการวิเคราะห์การถดถอยนั้นสามารถใช้การวิเคราะห์ความแปรปรวน(ANOVA)เข้ามาช่วยได้และมีประโยชน์อย่างมากในการวิเคราะห์การถดถอยพหุ (multiple regression analysis) การวิเคราะห์ความแปรปรวนในที่นี้เป็นการทดสอบสมมติฐานว่าค่าความชัน ($\beta_1$) มีค่าเท่ากับ 0 หรือไม่หรืออีกนัยหนึ่งคือตัวแปรอิสระสามารถใช้พยากรณ์ตัวแปรตามหรือไม่

### 2.3.1 ผลรวมกำลังสอง

การวิเคราะห์ความแปรปรวนมาจากแบ่งความแปรผันของ $Y$ ออกเป็นสองส่วนคือความแปรผันที่อธิบายได้โดย $X$ กับความแปรผันที่ไม่ทราบสาเหตุ หากตัวแปร $X$ กับ $Y$ มีความสัมพันธ์กันแล้วความแปรผันของ $Y$ ควรจะประกอบด้วยความแปรผันที่ $X$ สามารถอธิบายได้เป็นส่วนใหญ่ ความแปรผันของ $Y$ เป็นความแตกต่างระหว่างค่าสังเกตแต่ละค่ากับค่าเฉลี่ย $(Y_i - \bar{Y})$ ซึ่งสามารถแบ่งได้เป็นความแตกต่างระหว่างค่าพยากรณ์แต่ละค่ากับค่าเฉลี่ย $(\hat{Y}_i - \bar{Y})$ หรือเป็นความแปรผันที่อธิบายโดย $X$ กับความแตกต่างระหว่างค่าสังเกตกับค่าพยากรณ์ $(Y_i - \hat{Y}_i)$ หรือความแปรผันที่ไม่ทราบสาเหตุหรือไม่สามารถอธิบายได้โดย $X$ สามารถเขียนได้ดังนี้

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

หากยกกำลังสองแต่ละพจน์และหาผลรวมแล้วจะได้

$$\sum_{i=1}^{n}(Y_i-\bar{Y}) = \sum_{i=1}^{n}(\hat{Y}_i-\bar{Y}) + \sum_{i=1}^{n}(Y_i-\hat{Y}_i) + 2\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})(Y_i-\hat{Y}_i) \quad (2.29)$$

พิจารณาพจน์สุดท้ายของสมการจะได้

$$\begin{aligned} 2\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})(Y_i-\hat{Y}_i) &= 2\sum_{i=1}^{n}\hat{Y}_i(Y_i-\hat{Y}_i) - 2\bar{Y}\sum_{i=1}^{n}(Y_i-\hat{Y}_i) \\ &= 2\sum_{i=1}^{n}\hat{Y}_i e_i - 2\bar{Y}\sum_{i=1}^{n}e_i \\ &= 0 \end{aligned}$$

จากคุณสมบัติข้อ 1 และข้อ 6 ของค่าคลาดเคลื่อนและค่าพยากรณ์จะได้ว่าผลรวมของค่าคลาดเคลื่อนเท่ากับ 0 และผลรวมของค่าคลาดเคลื่อนที่ถ่วงน้ำหนักด้วยค่าพยากรณ์เท่ากับ 0 ดังนั้น

$$\sum_{i=1}^{n}(Y_i-\bar{Y}) = \sum_{i=1}^{n}(\hat{Y}_i-\bar{Y}) + \sum_{i=1}^{n}(Y_i-\hat{Y}_i) \quad (2.30)$$

หรือ $SST = SSR + SSE$

โดยที่ *SST* คือผลรวมกำลังสองทั้งหมด (total sum of squares) หรือค่า $S_{yy}$ เป็นการวัดความแปรผันในค่าสังเกตทั้งหมด หากค่าสังเกตทุกค่าเท่ากันหมดแล้ว $SST = 0$ ดังนั้นเมื่อค่าสังเกตมีค่าแตกต่างกันมากค่า *SST* จะใหญ่ขึ้น

*SSR* คือผลรวมกำลังสองถดถอย (regression sum of squares) เป็นการวัดความแปรผันในค่าสังเกตที่สามารถอธิบายได้โดยสมการถดถอย หากค่า *SSR* ใหญ่ขึ้นเมื่อเทียบกับ *SST* แสดงว่าสมการถดถอยสามารถอธิบายความแปรผันในค่าสังเกตได้ดี

*SSE* คือผลรวมกำลังสองความคลาดเคลื่อน (error sum of squares) เป็นการวัดความแปรผันที่ไม่สามารถอธิบายได้โดยสมการถดถอยหรือตัวแปร $X$ หากสมการถดถอยสามารถอธิบายค่าสังเกตได้ดี ค่าพยากรณ์ทุกค่าจะมีค่าเท่ากับค่าสังเกตทำให้ $SSE = 0$

จากสมการ (2.30) และ (2.14) จะสามารถคำนวณหา *SSR* ได้ดังนี้

$$SSR = b_1 S_{xy} \quad (2.31)$$

### 2.3.2 องศาเสรี

ในทำนองเดียวกันกับผลรวมกำลังสองพบว่าองศาเสรีทั้งหมดสามารถแบ่งได้เป็นสองส่วนคือ องศาเสรีของสมการถดถอยและองศาเสรีของความคลาดเคลื่อน องศาเสรีทั้งหมดเป็นองศาเสรีของ *SST* โดยมีค่าเท่ากับ $n - 1$ การสูญเสียความเป็นอิสระ 1 ค่าเนื่องจากข้อกำหนดที่ว่าผลรวมของค่าความแตกต่างระหว่างค่าสังเกตกับค่าเฉลี่ยหรือ $\sum_{i=1}^{n}(Y_i-\bar{Y})$ เท่ากับ 0 องศาเสรีของ *SSE* เท่ากับ $n - 2$ การสูญเสียความเป็นอิสระไป 2 ค่าเนื่องจากการประมาณค่า $\beta_0$ และ $\beta_1$ สำหรับองศาเสรีของ *SSR* เท่ากับ 1 เนื่องจากองศาเสรีรวมของ *SSR* กับ *SSE* ต้องเท่ากับ *SST* ดังนี้

$$n - 1 = 1 + (n - 2)$$

### 2.3.3 ค่าเฉลี่ยกำลังสอง

ค่าเฉลี่ยกำลังสอง (mean square) เกิดจากการนำเอาผลรวมกำลังสองมาหารด้วยองศาเสรี ในการวิเคราะห์ความแปรปรวนจะใช้ค่าเฉลี่ยกำลังสองเฉพาะค่าเฉลี่ยกำลังสองถดถอยและความคลาดเคลื่อน

ค่าเฉลี่ยกำลังสองถดถอย (regression mean square หรือ *MSR*) คำนวณได้จากการหารผลรวมกำลังสองถดถอยด้วยองศาเสรีดังนี้

$$MSR = \frac{SSR}{1} = SSR \tag{2.32}$$

ค่าเฉลี่ยกำลังสองความคลาดเคลื่อน (error mean square หรือ *MSE*) คำนวณได้จากการหารผลรวมกำลังสองความคลาดเคลื่อนด้วยองศาเสรีดังนี้

$$MSE = \frac{SSE}{n-2} \tag{2.33}$$

เนื่องจาก *MSR* และ *MSE* ต่างเป็นตัวแปรสุ่มดังนั้นตัวแปรสุ่มทั้งสองมีค่าคาดหวังดังนี้

$$E(MSR) = \sigma^2 + \beta_1^2 S_{xx}$$

และ

$$E(MSE) = \sigma^2$$

### 2.3.4 ค่าสถิติ *F*

การวิเคราะห์ความแปรปรวนในที่นี้เป็นการทดสอบสมมติฐานว่าค่าความชัน ($\beta_1$) มีค่าเท่ากับ 0 หรือไม่หรือ $H_0$: $\beta_1 = 0$ และ $H_1$: $\beta_1 \neq 0$ จะเห็นว่าเป็นการทดสอบเช่นเดียวกับการทดสอบโดยใช้การทดสอบแบบ *t* ในหัวข้อ 2.2.2 แต่สถิติที่ใช้คือ *F* โดยมีวิธีคำนวณดังนี้

$$F = \frac{MSR}{MSE} \tag{2.34}$$

การทดสอบสมมติฐานทำโดยเปรียบเทียบกับค่า *F* จากตารางที่ 2 ในภาคผนวกที่องศาเสรีเท่ากับ 1 และ $n - 2$ หากค่า *F* ที่คำนวณได้มีค่าน้อยกว่าหรือเท่ากับค่า *F* จากตารางแล้วจะไม่ปฏิเสธสมมติฐานหลัก ความแปรผันที่อธิบายได้โดย *X* จะมีค่าไม่แตกต่างจากความแปรผันที่อธิบายไม่ได้แต่หากค่า *F* ที่คำนวณได้มีค่ามากกว่าค่า *F* จากตารางแล้วจะปฏิเสธสมมติฐานหลักแสดงว่า *X* สามารถอธิบายตัวแปร *Y* ได้ดีดังนั้นความแปรผันที่อธิบายได้โดย *X* จะมีค่าสูงกว่าความแปรผันที่อธิบายไม่ได้มาก ผลที่ได้จากการทดสอบ *F* จะสอดคล้องกับผลที่ได้จากการทดสอบ *t*

และเมื่อยกกำลังสองค่าสถิติ $t$ จะมีค่าเท่ากับสถิติ $F$ แต่สถิติ $t$ นั้นสามารถทดสอบสองหางหรือหางเดียวก็ได้ในขณะที่ $F$ ใช้ทดสอบได้เฉพาะสองหางเท่านั้น

การเพิ่มอำนาจการทดสอบของสถิติ $F$ นั้นทำได้โดยการเพิ่มขนาดตัวอย่างหรือขยายพิสัยของค่าของตัวแปรอิสระให้กว้างขึ้นหรือการลดความแปรปรวนของส่วนเหลือลง (Weisberg, 2005, p. 31)

การนำเสนอค่าสถิติต่างๆ ของการวิเคราะห์ความแปรปรวนที่ได้กล่าวมาแล้วส่วนใหญ่จะนำเสนอในรูปของของตารางเพื่อให้ง่ายแก่การอ่านค่าดังนี้

| **Source of variation** | ***SS*** | ***df*** | ***MS*** | ***F*** |
|---|---|---|---|---|
| **Regression** | $SSR = b_1 S_{xy}$ | 1 | $MSR = SSR$ | $F = \dfrac{MSR}{MSE}$ |
| **Error** | $SSE = SST - SSR$ | $n-2$ | $MSE = \dfrac{SSE}{n-2}$ | |
| **Total** | $SST = S_{yy}$ | $n-1$ | | |

**ตัวอย่าง 2.7** จากข้อมูลในตัวอย่าง 2.1 จงทดสอบค่าความชันโดยใช้สถิติ $F$ ที่ระดับความเชื่อมั่น 95% และสร้างตารางวิเคราะห์ความแปรปรวน

**วิธีทำ**

เนื่องจากการทดสอบสมมติฐานโดยใช้ $F$ นั้นเป็นการทดสอบสองหางดังนั้นสมมติฐานคือ

$H_0$: $\beta_1 = 0$

$H_1$: $\beta_1 \neq 0$

จากตัวอย่างที่ 2.1 จะได้ $SST$ = 60,187.23 ($S_{yy}$ จากตัวอย่าง 2.1)

$$SSR = b_1 S_{xy}$$
$$= 8.22 \times 6{,}885.28 = 56{,}597.00$$

ดังนั้น
$$SSE = SST - SSR$$
$$= 60{,}187.23 - 56{,}597.00$$
$$= 3{,}590.23$$
$$MSR = SSR / 1$$
$$= 56{,}597.00$$
$$MSE = \frac{SSE}{n-2}$$

$$= \frac{3{,}590.23}{11} = 326.38$$

$$F = \frac{MSR}{MSE}$$

$$= \frac{56{,}597.00}{326.38} = 173.41$$

เนื่องจากค่าวิกฤต $F_{0.05,(1,11)} = 4.54$ ซึ่งน้อยกว่าค่า $F$ ที่คำนวณได้ (173.41) ดังนั้นจึงปฏิเสธสมมติฐานหลักและสรุปว่าความชันไม่เท่ากับ 0 หรือตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นตรงกันที่ระดับนัยสำคัญ 0.05

ตารางวิเคราะห์ความแปรปรวนคือ

| **Source of variation** | ***SS*** | ***df*** | ***MS*** | ***F*** |
|---|---|---|---|---|
| **Regression** | 56,597.00 | 1 | 56,597.00 | 173.41 |
| **Error** | 3,590.23 | 11 | 326.38 | |
| **Total** | **60,187.23** | **12** | | |

**หมายเหตุ**

หากเปรียบเทียบตารางวิเคราะห์ความแปรปรวนที่ได้จากการคำนวณด้วยมือกับผลลัพธ์จาก MINITAB ในตัวอย่าง 2.1 พบว่าค่าต่างๆ แตกต่างกันเล็กน้อยทั้งนี้อาจเนื่องมาจากการปัดเศษในการคำนวณนอกจากนี้ MINITAB ยังแสดงค่า $p$-value ในคอลัมน์สุดท้ายอีกด้วย โดย $p$-value = 0.000 < $\alpha$ (0.05)

Analysis of Variance

| Source | DF | SS | MS | F | P |
|---|---|---|---|---|---|
| Regression | 1 | 56603 | 56603 | 173.71 | 0.000 |
| Residual Error | 11 | 3584 | 326 | | |
| Total | 12 | 60187 | | | |

## 2.4 สัมประสิทธิ์การตัดสินใจ

*SST* เป็นการวัดความแปรผันในค่าสังเกต $Y$ โดยไม่คำนึงถึงตัวแปร $X$ และ *SSE* เป็นการวัดความแปรผันที่เหลือใน $Y$ จากการนำสมการถดถอยมาใช้ซึ่งได้มาจากการนำความแปรผันทั้งหมดใน $Y$ มาหักลบออกด้วยความแปรผันที่อธิบายด้วยสมการถดถอย บางครั้งนักวิจัยอาจต้องการทราบว่าสัดส่วนของความแปรผันที่อธิบายได้โดยสมการถดถอยต่อความแปรผันทั้งหมดเป็นเท่าใด

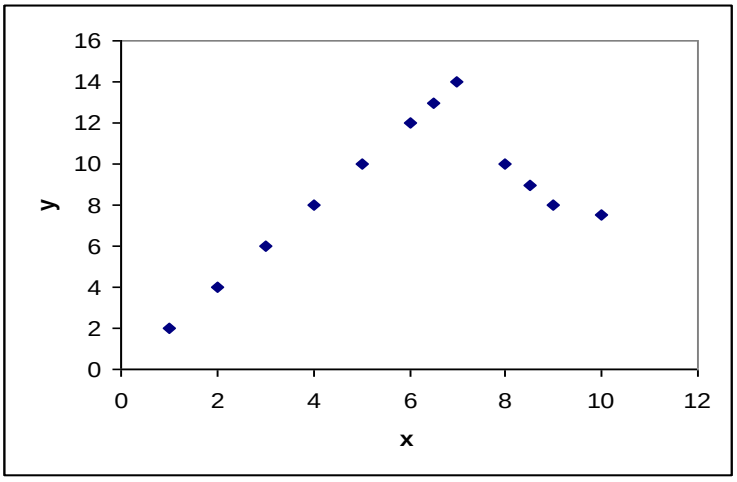
นักวิจัยสามารถใช้สัมประสิทธิ์การตัดสินใจ (coefficient of determination) หรือใช้ตัวย่อว่า $R^2$ ในการวัดสัดส่วนของความแปรผัน สูตรที่ใช้ในการคำนวณสัมประสิทธิ์การตัดสินใจคือ

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{2.35}$$

เนื่องจากค่า $R^2$ เป็นค่าสัดส่วนดังนั้นจึงมีค่าได้ตั้งแต่ 0 ถึง 1 เท่านั้น หากค่า $R^2$ ที่ได้มีค่าต่ำหรือเข้าใกล้ 0 แสดงว่าสมการถดถอยหรือตัวแปร $X$ มีความสามารถต่ำในการอธิบายความแปรผันใน $Y$ หรือความถูกต้องในการพยากรณ์ที่จะต่ำ เมื่อค่า $R^2$ เข้าใกล้ 1 แสดงว่าความแปรผันใน $Y$ เกือบทั้งหมดสามารถอธิบายได้โดยตัวแปร $X$ หรือสมการถดถอยมีความถูกต้องในการพยากรณ์ตัวแปร $Y$ ได้ดี ผู้อ่านอาจมีคำถามว่าค่า $R^2$ ควรจะมีค่าเท่าไรจึงจะถือว่าดี ? คำตอบของคำถามนี้คือขึ้นอยู่กับธรรมชาติของสิ่งที่ศึกษา หากเป็นงานวิจัยทางสังคมศาสตร์แล้วบางครั้งการได้ค่า $R^2$ ประมาณ 0.60 หรือ 0.70 อาจถือว่าสูงมากแต่ทางวิทยาศาสตร์อาจถือว่ายังค่อนข้างต่ำทั้งนี้ในงานวิจัยทางสังคมศาสตร์หรืองานวิจัยที่เกี่ยวข้องกับพฤติกรรมของมนุษย์นั้นมีความแปรผันและมีปัจจัยที่เกี่ยวข้องค่อนข้างมากจึงยากแก่การพยากรณ์ ตัวอย่างของการอธิบายความหมายของ $R^2$ เช่น $R^2 = 0.80$ หมายถึง สามารถใช้ตัวแปร $X$ ในการอธิบายความแปรผันในตัวแปร $Y$ ได้ 80% เป็นต้น

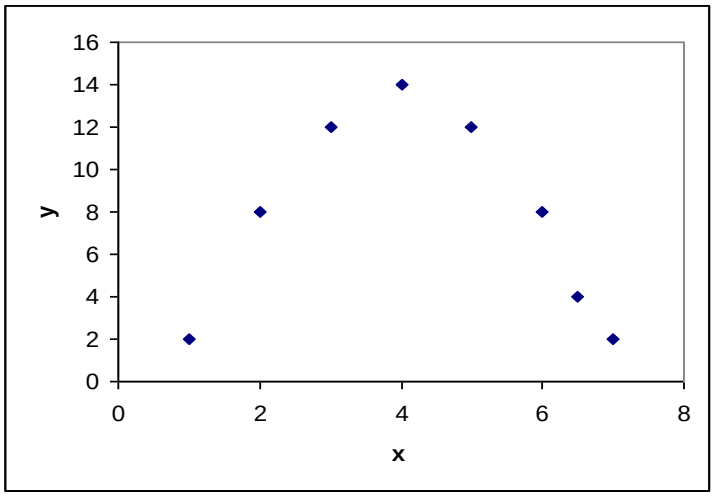
ข้อควรระวังในการใช้ $R^2$ มีดังนี้

(1) ค่า $R^2$ ที่สูงไม่ได้หมายความว่าสมการถดถอยจะพยากรณ์ตัวแปร $Y$ ได้ดีเสมอไปทั้งนี้ขึ้นอยู่กับขอบเขตของการพยากรณ์และขอบเขตของข้อมูลที่นำมาสร้างสมการถดถอยบางครั้งในขอบเขตที่ทำการศึกษานั้นตัวแปรทั้งสองมีความสัมพันธ์กันเชิงเส้นตรงแต่นอกขอบเขตที่ศึกษาความสัมพันธ์อาจไม่เป็นเส้นตรงก็ได้ดังภาพที่ 2.3 ขอบเขตในการศึกษาอยู่ระหว่าง 1 ถึง 7 และความสัมพันธ์ในช่วงดังกล่าวเป็นเส้นตรงแต่กลับลดลงหลังจากค่า x เท่ากับ 7 ดังนั้นหากทำการพยากรณ์นอกเหนือขอบเขตที่ศึกษาหรือทำการพยากรณ์ในช่วงตั้งแต่ 8 ขึ้นไปแล้วค่าที่พยากรณ์ได้จะไม่ตรงกับสภาพความเป็นจริงและนักวิจัยไม่มีโอกาสทราบได้เนื่องจากไม่ได้ทำการศึกษาในช่วง x ที่มากกว่า 7 ขึ้นไป




ภาพที่ 2.3 ความสัมพันธ์ที่เป็นเส้นตรงในช่วงแรก

(2) บางครั้งสมการที่มีค่า $R^2$ เข้าใกล้ 0 ไม่ได้หมายความว่าตัวแปรทั้งสองไม่มีความสัมพันธ์กันแต่อาจมีความสัมพันธ์เชิงเส้นโค้งก็ได้ หากใช้ตัวแบบที่ไม่เป็นเชิงเส้นมาพยากรณ์จะมีความเหมาะสมมากกว่าสมการถดถอยเชิงเส้นดังภาพที่ 2.4




ภาพที่ 2.4 ความสัมพันธ์แบบเส้นโค้ง

(3) เนื่องจากค่าคาดหวังของค่า $R^2$ เท่ากับ $\frac{b_1^2 S_{xx}}{b_1^2 S_{xx} + \sigma^2}$ ดังนั้นเมื่อขอบเขตของค่า $X$ กว้างขึ้นหรือ $S_{xx}$ ใหญ่ขึ้นจะทำให้ค่า $R^2$ สูงขึ้นด้วย แต่หากขอบเขตของ $X$ กว้างมากจนเกินความจำเป็นจะทำให้ผลที่ได้ใช้นำไปใช้ประโยชน์ได้ไม่เท่าที่ควร

(4) ค่า $R^2$ เป็นค่าที่ได้จากการใช้ข้อมูลชุดที่ทำการศึกษา หากข้อมูลที่เก็บได้แต่ละครั้งแตกต่างกันค่า $R^2$ จะแตกต่างกันไป นักวิจัยบางท่านอาจพยายามที่จะปรับเปลี่ยนสมการถดถอยเพื่อให้มีค่า $R^2$ สูงขึ้นแต่สมการถดถอยที่ได้อาจเหมาะสมกับข้อมูลชุดใดชุดหนึ่งเท่านั้น บางครั้งจะพบว่านักวิจัยแบ่งข้อมูลบางส่วนเพื่อใช้ทดสอบว่าสมการถดถอยที่ได้ยังคงเหมาะสมกับข้อมูลอีกชุดหรือไม่

(5) ค่า $R^2$ ไม่ได้แสดงถึงความสัมพันธ์เชิงเส้นตรงเสมอไป ในกรณีของสมการถดถอยที่ไม่เป็นเชิงเส้นแล้วค่า $R^2$ ที่ได้ไม่ได้แสดงถึงความสัมพันธ์ที่เป็นเชิงเส้นแต่แสดงถึงสัดส่วนของความแปรผันที่ตัวแปร $X$ สามารถอธิบายได้

(6) ค่า $R^2$ ขึ้นกับขนาดตัวอย่าง กล่าวคือ เมื่อขนาดตัวอย่างยิ่งน้อยจะยิ่งทำให้ค่า $R^2$ สูงขึ้น (Cohen et al, 2003, p. 83)

## 2.5 สัมประสิทธิ์สหสัมพันธ์

ค่าสัมประสิทธิ์สหสัมพันธ์ (coefficient of correlation) หรือย่อว่า $r$ เป็นค่าที่แสดงถึงความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรสองตัว ความสัมพันธ์ระหว่างค่าสัมประสิทธิ์สหสัมพันธ์และสัมประสิทธิ์การตัดสินใจคือ

$$r = \pm\sqrt{R^2} \tag{2.36}$$

ค่า $r$ มีขอบเขตตั้งแต่ -1 ถึง 1 หาก $r$ มีค่าเข้าใกล้ -1 แสดงว่าตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นตรงแบบผกผันกันกล่าวคือเมื่อตัวแปรหนึ่งมีค่ามากขึ้นอีกตัวแปรหนึ่งจะมีค่าลดลง แต่หาก $r$ มีค่าเข้าใกล้ 1 แสดงว่าตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นตรงแบบตามกันกล่าวคือเมื่อตัวแปรหนึ่งมีค่ามากขึ้นอีกตัวแปรหนึ่งจะมีค่าเพิ่มขึ้นด้วยแต่หาก $r$ มีค่าเท่ากับ 0 แล้วแสดงว่าตัวแปรทั้งสองไม่มีความสัมพันธ์เชิงเส้นตรงกัน หากคำนวณค่า $r$ จาก $R^2$ แล้วจะต้องพิจารณาเครื่องหมายของ $b_1$ เป็นหลัก หาก $b_1$ มีเครื่องหมายเป็นบวกค่า $r$ จะมีเครื่องหมายเป็นบวกเช่นกัน หาก $b_1$ มีเครื่องหมายเป็นลบค่า $r$ จะมีเครื่องหมายเป็นลบ โดยค่าทั้งสองมีความสัมพันธ์กันดังนี้

$$b_1 = \left(\frac{S_{yy}}{S_{xx}}\right)^2 r \tag{2.37}$$

นอกจากนี้สามารถคำนวณค่า $r$ ได้โดยตรงจากสูตร

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \qquad (2.38)$$

**ตัวอย่าง 2.8** จากข้อมูลในตัวอย่าง 2.1 จงคำนวณค่าสัมประสิทธิ์การตัดสินใจและสัมประสิทธิ์สหสัมพันธ์

**วิธีทำ**

สัมประสิทธิ์การตัดสินใจเท่ากับ $R^2 = \frac{SSR}{SST}$

$$= \frac{56{,}597.00}{60{,}187.23} = 0.94$$

ดังนั้นค่าใช้จ่ายในการโฆษณาสามารถอธิบายความแปรผันของยอดขายได้ถึง 94% หรือความแปรผันของยอดขายที่ไม่ได้มาจากการโฆษณามีเพียง 6% เท่านั้น

สัมประสิทธิ์สหสัมพันธ์เท่ากับ $r = +\sqrt{R^2}$

$$= +\sqrt{0.94} = 0.97$$

เนื่องจากค่า $b_1$ ในตัวอย่างที่ 2.1 มีค่าเป็นบวกดังนั้นค่า $r$ จึงมีค่าเป็นบวกด้วยและค่าใช้จ่ายในการโฆษณากับยอดขายมีความสัมพันธ์เชิงเส้นตรงกันอย่างมาก

**หมายเหตุ**

หากเปรียบเทียบค่าสัมประสิทธิ์การตัดสินใจที่ได้จากการคำนวณด้วยมือกับผลลัพธ์จาก MINITAB ในตัวอย่าง 2.1 ในบรรทัดที่ 6 ในรูปของ R-Sq พบว่าค่าที่ได้จะเท่ากัน นอกจากนี้ผลลัพธ์ที่ได้จาก MINITAB ยังแสดงค่า R-Sq(adj) ซึ่งจะอธิบายในบทที่ 5

```
The regression equation is
y = - 2.0 + 8.22 x

Predictor        Coef     SE Coef          T        P
Constant        -1.99       15.21      -0.13    0.898
x              8.2209       0.6237      13.18    0.000

S = 18.05       R-Sq = 94.0%     R-Sq(adj) = 93.5%
```

## 2.6 สมการถดถอยผ่านจุดกำเนิด

ที่ผ่านมานั้นเป็นการกล่าวถึงสมการถดถอยที่มีค่าตัวแปรทั้งสองไม่ผ่านจุดกำเนิดหรือจุดที่ $(x,y) = (0,0)$ แต่ในบางกรณีธรรมชาติของข้อมูลจำเป็นต้องเริ่มจากจุดกำเนิดซึ่งจะพบได้มากใน

กรณีของการทดลองทางเคมี เช่น การเกิดปฏิกิริยาของสารบางชนิดจะไม่เกิดขึ้นหากไม่มีตัวเร่งปฏิกิริยาเมื่อเพิ่มปริมาณของตัวเร่งปฏิกิริยาก็จะทำให้มีอัตราการเกิดปฏิกิริยาเพิ่มขึ้นเป็นต้น เนื่องจากข้อมูลมีการผ่านจุดกำเนิดดังนั้นตัวแบบจะไม่มีจุดตัดแกน $Y$ หรือ $\beta_0$ โดยมีตัวแบบคือ

$$Y_i = \beta_1 X_i + \varepsilon \tag{2.39}$$

ดังนั้นสมการปกติจึงมีเพียงสมการเดียวคือสมการของ $\beta_1$ ดังนี้

$$b_1 \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} Y_i X_i \tag{2.40}$$

และตัวประมาณค่าของ $\beta_1$ คือ

$$b_1 = \frac{\sum_{i=1}^{n} Y_i X_i}{\sum_{i=1}^{n} X_i^2} \tag{2.41}$$

ตัวประมาณค่า $b_1$ นี้เช่นเดียวกับในตัวแบบที่ผ่านมาคือเป็นตัวประมาณค่าที่ไม่เอนเอียงคือค่าเฉลี่ยเท่ากับ $\beta_1$ แต่มีค่าความแปรปรวนเท่ากับ $\frac{\sigma^2}{\sum_{i=1}^{n} X_i^2}$

สมการถดถอยที่ได้คือ

$$\hat{Y}_i = b_1 X_i \tag{2.42}$$

และมีตัวประมาณค่าของ $\sigma^2$ คือ

$$\hat{\sigma}^2 = MSE = \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n-1} = \frac{\sum_{i=1}^{n} Y_i^2 - b_1 \sum_{i=1}^{n} Y_i X_i}{n-1} \tag{2.43}$$

โดยมีองศาเสรีเท่ากับ $n$ - 1 เนื่องจากมีการประมาณค่าพารามิเตอร์เพียง $\beta_1$ ค่าเดียวเท่านั้น

การทดสอบค่าพารามิเตอร์มีเพียงค่าเดียวคือ $\beta_1$ และใช้สถิติ $t$ ที่มีองศาเสรีเท่ากับ $n - 1$ โดยในกรณีของการทดสอบ $H_1$: $\beta_1 \neq 0$ สามารถคำนวณได้ดังนี้

$$t = \frac{b_1}{\sqrt{\frac{MSE}{\sum_{i=1}^{n} X_i^2}}} \sim t_{\alpha/2, n-1} \tag{2.44}$$

การคำนวณค่าสัมประสิทธิ์การตัดสินใจใช้สูตรดังนี้

$$R^2 = \frac{\sum_{i=1}^{n} \hat{Y}_i^2}{\sum_{i=1}^{n} Y_i^2} \tag{2.45}$$

**ตัวอย่าง 2.9** ในการศึกษาเวลาที่ใช้ในการผลิต ($X$) กับผลผลิตที่ได้ ($Y$) มีข้อมูลดังนี้

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| เวลา (นาที) | 3.8 | 1.5 | 12.0 | 9.0 | 6.4 | 1.0 | 10.6 | 15.5 |
| ปริมาณผลผลิต (ตัน) | 8.0 | 4.3 | 29.1 | 21.5 | 15.9 | 3.4 | 25.0 | 34.7 |

จงสร้างสมการถดถอยและคำนวณค่าสัมประสิทธิ์การตัดสินใจ

**วิธีทำ**

เนื่องจากข้อมูลชนิดนี้เป็นข้อมูลที่มีจุดเริ่มต้นที่จุดกำเนิด หากไม่มีการผลิตจะไม่มีผลผลิต ดังนั้นสมการถดถอยจึงเป็นสมการถดถอยผ่านจุดกำเนิด จากข้อมูลจะได้

$$b_1 = \frac{\sum_{i=1}^{n} Y_i X_i}{\sum_{i=1}^{n} X_i^2}$$

$$= \frac{1{,}487.56}{636.26} = 2.34$$

ดังนั้นสมการถดถอยคือ

$$\hat{Y}_i = 2.34X_i$$

จากสมการถดถอยจะได้ค่า

$$R^2 = \frac{\sum_{i=1}^{n} \hat{Y}_i^2}{\sum_{i=1}^{n} Y_i^2}$$

$$= \frac{3{,}477.88}{3{,}485.01} = 0.998$$

**หมายเหตุ**

หากสร้างสมการถดถอยแบบปกติจะได้สมการถดถอยดังนี้

$$\hat{Y}_i = 0.878 + 2.26X_i$$

แต่เมื่อทดสอบสมมติฐาน $H_0$: $\beta_0 = 0$ จะพบว่าค่า $p$-value = 0.200 แสดงว่าข้อมูลชนิดนี้ไม่จำเป็นต้องมีจุดตัดแกนและสมการถดถอยผ่านจุดกำเนิดนี้เหมาะกับข้อมูลชนิดนี้มากกว่าสมการถดถอยแบบปกติ

## สรุป

การวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายใช้ในการวิเคราะห์ข้อมูลที่มีตัวแปรอิสระ 1 ตัว และตัวแปรตาม 1 ตัวโดยตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นตรงกัน การสร้างสมการถดถอยใช้

วิธีกำลังสองน้อยที่สุดในการประมาณค่าพารามิเตอร์ $\beta_0$ และ $\beta_1$ ในการทดสอบค่าพารามิเตอร์ทั้งสองสามารถทำได้โดยการใช้การทดสอบ $t$ และในการทดสอบว่าพารามิเตอร์ทั้งสองหรือสมการถดถอยสามารถพยากรณ์ตัวแปรตามหรือไม่ทำได้โดยการวิเคราะห์ความแปรปรวน สัมประสิทธิ์การตัดสินใจเป็นสถิติอีกตัวหนึ่งที่ใช้ช่วยในการตรวจสอบความสามารถในการพยากรณ์ตัวแปรตามของสมการถดถอยที่ได้

## คำถามท้ายบท

2.1 จากข้อมูลข้างล่างจงใช้วิธีกำลังสองน้อยที่สุดในการสร้างสมการถดถอย

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $X$: | 21 | 13 | 20 | 25 | 19 | 24 | 16 | 13 |
| $Y$: | 13 | 6 | 12 | 7 | 19 | 10 | 24 | 19 |

2.2 จากข้อมูลในข้อ 2.1 จงทดสอบค่าพารามิเตอร์โดยใช้สถิติ $t$ และ $F$ ที่ระดับนัยสำคัญ 0.10

2.3 จากข้อมูลในข้อ 2.1 จงคำนวณสัมประสิทธิ์การตัดสินใจ

2.4 ร้านขายคอมพิวเตอร์ต้องการทราบว่าจำนวนตัวแทนขายมีผลต่อยอดขายคอมพิวเตอร์หรือไม่ โดยร้านค้าเก็บข้อมูลในแต่ละเดือนดังนี้

| เดือนที่ | จำนวนเครื่อง | จำนวนตัวแทน |
|---|---|---|
| 1 | 21 | 6 |
| 2 | 17 | 6 |
| 3 | 10 | 4 |
| 4 | 5 | 2 |
| 5 | 11 | 3 |
| 6 | 12 | 4 |

จากข้อมูลข้างต้นจงสร้างสมการถดถอยโดยใช้วิธีกำลังสองน้อยที่สุด

2.5 จากข้อมูลในข้อ 2.4 จงวาดแผนภาพกระจายระหว่างตัวแปรทั้งสองและวาดเส้นถดถอยที่ได้จากสมการในข้อ 2.4 พร้อมทั้งอธิบายความสัมพันธ์ระหว่างตัวแปรทั้งสองจากแผนภาพกระจายและพิจารณาว่าเส้นถดถอยที่ได้สอดคล้องกับข้อมูลส่วนใหญ่หรือไม่

2.6 จากข้อมูลในข้อ 2.4 จงคำนวณค่าพยากรณ์และส่วนเหลือที่จำนวนตัวแทนแต่ละค่า

2.7 จากข้อมูลในข้อ 2.4 จงหาสัมประสิทธิ์การตัดสินใจพร้อมทั้งอธิบายค่าที่ได้

2.8 จงใช้โปรแกรม MINITAB สร้างสมการถดถอยของข้อมูลในข้อ 2.4

2.9 ข้อมูลข้างล่างคือเกรดเฉลี่ยสะสมกับคะแนนสอบวิชาคณิตศาสตร์ของนักศึกษากลุ่มหนึ่ง

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| เกรดเฉลี่ยสะสม | 56 | 54 | 52 | 58 | 52 | 62 | 66 | 63 |
| คะแนน | 3.18 | 3.45 | 3.68 | 2.55 | 2.07 | 3.41 | 3.18 | 2.79 |

จงสร้างสมการถดถอยเพื่อพยากรณ์คะแนนสอบ

2.10 จากข้อมูลในข้อ 2.9 จงหาสัมประสิทธิ์การตัดสินใจพร้อมทั้งอธิบายค่าที่ได้

2.11 จากข้อมูลในข้อ 2.9 จงทดสอบว่าสมการถดถอยที่ได้เหมาะสมกับข้อมูลชุดนี้หรือไม่ที่ระดับนัยสำคัญ 0.05 โดยใช้สถิติ $F$

2.12 จากผลลัพธ์ที่ได้ข้างล่างจงทดสอบว่าตัวแปรอิสระ $X$ สามารถพยากรณ์ตัวแปรตาม $Y$ ได้หรือไม่ที่ระดับนัยสำคัญ 0.05

```
The regression equation is
y = - 92.8 + 0.273 x

Predictor        Coef     SE Coef          T        P
Constant       -92.81       40.11      -2.31    0.043
x             0.27275     0.06977       3.91    0.003

S = 10.83       R-Sq = 60.4%     R-Sq(adj) = 56.5%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1      1791.0      1791.0     15.28    0.003
Residual Error    10      1172.0       117.2
Total             11      2963.0
```

2.13 จากข้อมูลความดันเลือดและอายุ

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| อายุ | 63 | 61 | 45 | 53 | 60 | 42 | 58 | 62 | 47 | 56 |
| ความดัน | 161 | 217 | 130 | 148 | 162 | 144 | 152 | 140 | 135 | 122 |

จงวาดแผนภาพกระจายระหว่างอายุกับความดันพร้อมทั้งอธิบายความสัมพันธ์ที่ได้

2.14 จากข้อมูลในข้อ 2.13 จงสร้างสมการถดถอยและคำนวณค่าสัมประสิทธิ์การตัดสินใจพร้อมทั้งเปรียบเทียบค่าสัมประสิทธิ์การตัดสินใจที่ได้กับแผนภาพกระจายในข้อ 2.13