บทที่ 6
การวิเคราะห์การถดถอยที่ไม่ใช่เชิงเส้น

จากบทที่ผ่านมานั้นเป็นการหาสร้างสมการถดถอยสำหรับความสัมพันธ์ที่เป็นเส้นตรงระหว่างตัวแปรอิสระกับตัวแปรตาม หากความสัมพันธ์ระหว่างตัวแปรทั้งสองไม่เป็นเส้นตรงแล้วการใช้สมการถดถอยดังกล่าวจะทำให้การพยากรณ์คลาดเคลื่อนไปจากความเป็นจริง การวิเคราะห์การถดถอยที่ไม่ใช่เชิงเส้นใช้ประโยชน์ใน 2 กรณีหลักคือ กรณีที่นักวิจัยทราบว่าข้อมูลมีความสัมพันธ์ระหว่างกันแบบไม่เป็นเส้นตรงหรือในกรณีที่นักวิจัยไม่ทราบความสัมพันธ์ที่แน่นอนแต่ใช้ฟังก์ชันที่ไม่เป็นเชิงเส้นในการสร้างสมการความสัมพันธ์ ในบทนี้จะกล่าวถึงการวิเคราะห์การถดถอยที่ไม่ใช่เชิงเส้นรวมถึงประเด็นต่างๆ ที่สำคัญในการวิเคราะห์การถดถอยที่ไม่ใช่เชิงเส้น

6.1 รูปแบบการถดถอยที่ไม่ใช่เชิงเส้น

ตัวแบบการถดถอยที่ไม่ใช่เชิงเส้นสามารถแบ่งได้เป็น 2 รูปแบบใหญ่คือ

6.1.1 รูปแบบโพลิโนเมียลลำดับที่ p

รูปแบบโพลิโนเมียลกลุ่มนี้มีด้วยกันหลายรูปแบบขึ้นอยู่กับกำลังของตัวแปรอิสระ หากกำลังเท่ากับ 1 แล้วสมการถดถอยจะเป็นเส้นตรง หากกำลังเท่ากับสองแล้วสมการถดถอยจะเป็นเส้นโค้ง 1 โค้งและหากกำลังเท่ากับ 3 แล้วสมการถดถอยจะเป็นเส้นโค้งหลายโค้ง

6.1.2 รูปแบบที่ไม่เป็นเส้นตรงที่แปลงเป็นเส้นตรงได้

รูปแบบที่ไม่เป็นเส้นตรงกลุ่มนี้สามารถแปลงตัวแปรอิสระหรือ/และตัวแปรตามให้อยู่ในรูปสมการเส้นตรงได้เช่น รูปแบบเอ็กซ์โปเนนเชียล ($Y = \beta_0 \beta_1^X$) รูปแบบไฮเปอร์โบลา ($Y = \dfrac{X}{\beta_0 X + \beta_1}$) และรูปแบบเอ็กซ์โปเนนเชียลผกผัน ($Y = \beta_0 e^{\beta_{1/X}}$) เป็นต้น

6.2 รูปแบบโพลิโนเมียลที่มีตัวแปรอิสระ 1 ตัว

ตัวแบบสำหรับที่มีตัวแปรอิสระ 1 ตัวแปรสามารถแสดงได้ในรูปทั่วไปดังสมการคือ

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \ldots + \beta_k X^k + \varepsilon \qquad (6.1)$$

หากสมการมีการยกกำลังสองเพียง 1 ค่าเรียกว่า ตัวแบบระดับสองใน 1 ตัวแปร (second-order model) บางครั้งอาจเรียกสมการกำลังสอง (quadratic model) สามารถเขียนได้เป็น

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon \tag{6.2}$$

การวิเคราะห์ตัวแบบโพลิโนเมียลนั้น X จะอยู่ในรูปของ $(X - \bar{X})$ เพื่อป้องกันปัญหาความสัมพันธ์ระหว่าง X กับ X^2 (multicollinearity) ซึ่งจะทำให้ไม่สามารถหาค่าเมตริกซ์ผกผันของ $X'X$ ดังนั้นตัวแบบโพลิโนเมียลสามารถเขียนใหม่ได้เป็น

$$Y = \beta_0 + \beta_1 (X - \bar{X}) + \beta_2 (X - \bar{X})^2 + \varepsilon \tag{6.3}$$

ในกรณีของสมการโพลิโนเมียลในรูปทั่วไปสามารถเขียนได้เป็น

$$Y = \beta_0 + \beta_1 (X - \bar{X}) + \beta_2 (X - \bar{X})^2 + \ldots + \beta_k (X - \bar{X})^k + \varepsilon \tag{6.4}$$

การสร้างสมการโพลิโนเมียลที่มีกำลังสูงๆ นั้นควรจะสร้างด้วยความระมัดระวังอย่างมาก ทั้งนี้เนื่องจากค่าสัมประสิทธิ์จะยากแก่การอธิบายและอาจทำให้การตีความผิดพลาดได้ นอกจากนี้การสร้างตัวแบบที่มีกำลังสูงๆหรือมีรูปแบบที่ซับซ้อนอาจจะเหมาะกับข้อมูลชุดนั้นแต่ไม่เหมาะกับข้อมูลชุดอื่นโดยทั่วไปในการวิจัยเรื่องเดียวกันซึ่งทำให้อาจไม่เหมาะกับการนำตัวแบบไปใช้ในกรณีทั่วไป (generalization)

6.2.1 การหากำลังที่เหมาะสมของตัวแบบ

ในการทดสอบหากำลังที่เหมาะสมของตัวแบบว่าควรจะมีกำลังที่สูงสุดควรเป็นเท่าไรนั้นทำโดยการทดสอบกำลังในระดับสูงก่อนว่าเท่ากับ 0 หรือไม่จากนั้นทดสอบกำลังในระดับต่ำลงมาโดยในแต่ละขั้นการทดสอบเอฟบางส่วนดังที่กล่าวมาแล้วในบทที่ 5 โดยมีขั้นตอนดังนี้

(1) การทดสอบกำลังสาม การทดสอบกำลังสามทำโดยกำหนดสมมติฐานหลักเป็น

$H_0 : \beta_3 = 0$ หรือ $Y = \beta_0 + \beta_1 (X - \bar{X}) + \beta_2 (X - \bar{X})^2 + \varepsilon$ กับ

$H_1 : \beta_3 \neq 0$ หรือ $Y = \beta_0 + \beta_1 (X - \bar{X}) + \beta_2 (X - \bar{X})^2 + \beta_3 (X - \bar{X})^3 + \varepsilon$ สถิติที่ใช้คือ

$$F = \frac{SSR(\beta_3 \mid \beta_0, \beta_1, \beta_2)/1}{MSE} \tag{6.5}$$

โดย MSE เป็น MSE ที่ได้จากสมการเต็มรูปและทำการเปรียบเทียบกับ F จากตารางที่องศาเสรีเท่ากับ 1 และ $n - 4$ หากปฏิเสธสมมติหลักแสดงว่าโพลิโนเมียลกำลังสามมีความเหมาะสมกับข้อมูลนั้น แต่หากไม่ปฏิเสธสมมติฐานหลักจะต้องทำการทดสอบกำลังสอง

(2) การทดสอบกำลังสอง การทดสอบกำลังสองจะทำเมื่อพบว่ากำลังสามนั้นไม่เหมาะสมกับข้อมูล โดยกำหนดสมมติฐานหลักเป็น $H_0 : \beta_2 = 0$ หรือ $Y = \beta_0 + \beta_1(X - \bar{X}) + \varepsilon$ กับ $H_1 : \beta_2 \neq 0$ หรือ $Y = \beta_0 + \beta_1(X - \bar{X}) + \beta_2(X - \bar{X})^2 + \varepsilon$ สถิติที่ใช้คือ

$$F = \frac{SSR(\beta_2 \mid \beta_0, \beta_1)/1}{MSE} \tag{6.6}$$

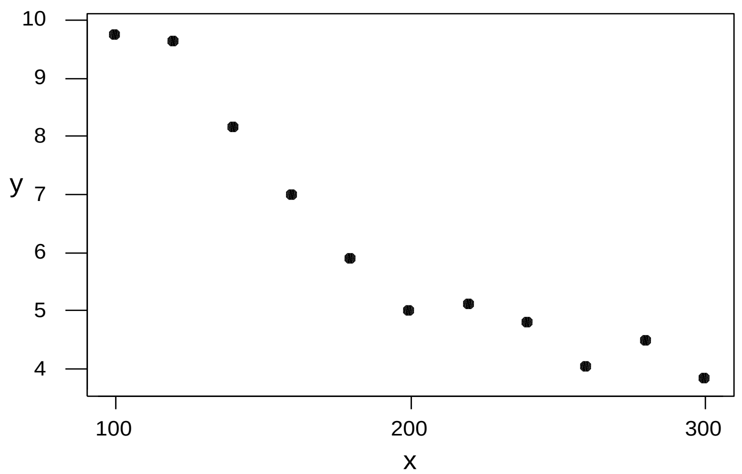
โดย MSE เป็น MSE ที่ได้จากสมการ $Y = \beta_0 + \beta_1(X - \bar{X}) + \beta_2(X - \bar{X})^2 + \varepsilon$ และทำการเปรียบเทียบกับ F จากตารางที่องศาเสรีเท่ากับ 1 และ $n - 3$ หากปฏิเสธสมมติหลักแสดงว่าโพลิโนเมียลกำลังสองมีความเหมาะสมกับข้อมูลนั้น แต่หากไม่ปฏิเสธสมมติฐานหลักจะต้องทำการทดสอบสมการถดถอยเชิงเส้นดังกล่าวมาแล้วในบทที่ 2

ตัวอย่างที่ 6.1 จากข้อมูลข้างล่างจงหาตัวแบบที่เหมาะสม

X	100	120	140	160	180	200	220	240	260	280	300
Y	9.73	9.61	8.15	6.98	5.87	4.98	5.09	4.79	4.02	4.46	3.82

วิธีทำ

การพิจารณาความสัมพันธ์ระหว่างตัวแปรทั้งสองอย่างง่ายทำได้โดยวาดกราฟระหว่างตัวแปรทั้งสอง จากกราฟข้างล่างที่เป็นเส้นโค้งเพียงโค้งเดียวดังนั้นคาดว่าตัวแปรทั้งสองมีความสัมพันธ์แบบเส้นโค้งที่มีกำลังเพียงกำลังสอง ซึ่งการวิเคราะห์โดยใช้สมการถดถอยเชิงเส้นอย่างง่ายอาจไม่เหมาะสมอีกทั้งการทดสอบกำลังสามจึงไม่จำเป็นเพื่อหาตรวจแบบที่เหมาะสมจึงทำการทดสอบเฉพาะกำลังสอง



เพื่อป้องกันการเกิดปัญหาความสัมพันธ์ระหว่างพจน์ของตัวแปรอิสระจึงนำค่าเฉลี่ยมาลบออกจากค่า X ในทุกพจน์และสมมติฐานคือ

$$H_0 : \beta_2 = 0$$

$$H_1 : \beta_2 \neq 0$$

ดังนั้นสมการเต็มรูปคือ

$$Y = \beta_0 + \beta_1(X - \bar{X}) + \beta_2(X - \bar{X})^2 + \varepsilon$$

และสมการลดรูปคือ

$$Y = \beta_0 + \beta_1(X - \bar{X}) + \varepsilon$$

จากข้อมูลพบว่า X มีค่าเฉลี่ย ($\bar{X}$) เท่ากับ 200 ดังนั้น

$$SSR(\beta_1, \beta_2 | \beta_0) = \mathbf{b'X'Y} - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}$$

$$= 459.5857 - 414.2045$$

$$= 45.3811 \quad \text{(องศาเสรีเท่ากับ 2)}$$

$$SSR(\beta_1 | \beta_0) = 41.666 \quad \text{(องศาเสรีเท่ากับ 1)}$$

ดังนั้น $SSR(\beta_2 | \beta_0, \beta_1) = 45.3811 - 41.666$

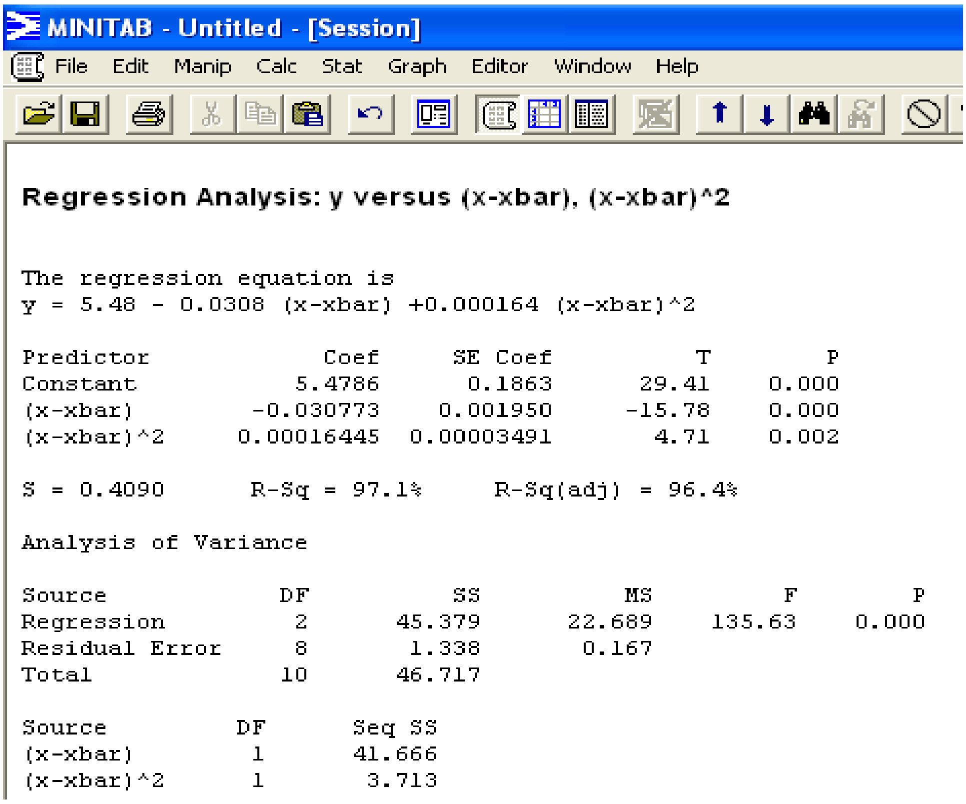
$$= 3.715 \quad \text{(องศาเสรีเท่ากับ 1)}$$

และ

$$F = \frac{SSR(\beta_2 | \beta_0, \beta_1)/1}{MSE}$$

$$= \frac{3.715/1}{0.167} = 22.246$$

โดยที่ MSE ได้จากตัวแบบที่มีตัวแปรอิสระกำลังสองอยู่หรือตัวแบบเต็มรูปแบบ เนื่องจากค่า $F = 22.246 > F_{0.05, 1, 8} = 5.32$ ดังนั้นจึงปฏิเสธสมมติฐานหลักและสรุปว่าพจน์กำลังสองมีผลต่อตัวแบบที่ระดับนัยสำคัญ 0.05 และเมื่อทำการวิเคราะห์ความแปรปรวนจะได้ผลดังตาราง ANOVA ที่ได้จากโปรแกรม MINITAB ข้างล่าง




```
Regression Analysis: y versus (x-xbar), (x-xbar)^2

The regression equation is
y = 5.48 - 0.0308 (x-xbar) +0.000164 (x-xbar)^2

Predictor            Coef      SE Coef          T       P
Constant           5.4786       0.1863      29.41   0.000
(x-xbar)        -0.030773     0.001950     -15.78   0.000
(x-xbar)^2     0.00016445   0.00003491       4.71   0.002

S = 0.4090      R-Sq = 97.1%      R-Sq(adj) = 96.4%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         2      45.379      22.689    135.63    0.000
Residual Error     8       1.338       0.167
Total             10      46.717

Source          DF      Seq SS
(x-xbar)         1      41.666
(x-xbar)^2       1       3.713
```

เมื่อพิจารณาการทดสอบค่า β_0, β_1 และ β_2 โดยใช้สถิติ t ให้ผลสอดคล้องกับการทดสอบโดยใช้เอฟบางส่วน ดังนั้นสมการถดถอยสำหรับข้อมูลชุดนี้คือ

$$\hat{Y} = 5.48 - 0.0308(X - \bar{X}) + 0.000164(X - \bar{X})^2$$

หรือ

$$\hat{Y} = 5.48 - 0.0308(X - 200) + 0.000164(X - 200)^2$$

6.2.2 การทดสอบความเหมาะสมของสมการถดถอย

การทดสอบในหัวข้อ 6.2.1 เป็นการทดสอบค่าพารามิเตอร์ที่กำลังต่างๆ ทีละค่า ในหัวข้อนี้จะเป็นการทดสอบสมการทั้งสมการว่าเหมาะสมกับข้อมูลหรือไม่ ในการทดสอบนี้จะเปรียบเทียบตัวแบบเต็มรูปกับตัวแบบลดรูปโดยใช้สถิติ F

ในกรณีของตัวแบบระดับสองใน 1 ตัวแปรกำหนดสมมติฐานหลักและสมมติฐานทางเลือกดังนี้

$$\mathrm{H}_0 : Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$$

$$\mathrm{H}_1 : Y \neq \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$$

การทดสอบความเหมาะสมนั้นมีขั้นตอนคล้ายกับการทดสอบในกรณีการวิเคราะห์ของสมการถดถอยอย่างง่ายเชิงเส้นในบทที่ 3 การทดสอบนี้แบ่งผลรวมกำลังสองความคลาดเคลื่อน (*SSE*) ออกเป็นสองส่วนคือ ผลรวมกำลังสองความคลาดเคลื่อนแท้จริง (pure error sum of squares หรือ *SSPE*) และผลรวมกำลังสองความคลาดเคลื่อนที่ไม่เหมาะสม (lack of fit sum of squares หรือ *SSLOF*) ดังนี้

$$SSE = SSPE + SSLOF \tag{6.7}$$

โดย *SSPE* เป็นผลรวมกำลังสองของการวัดความแปรผันของค่า Y ที่ตำแหน่ง X หนึ่งเมื่อเทียบกับค่าเฉลี่ยของ Y ณ ตำแหน่ง X นั้นดังนี้

$$SSPE = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i)^2 \tag{6.8}$$

โดยองศาเสรีของ *SSPE* เท่ากับ $\sum_{i=1}^{m}(n_i - 1) = n - m$

สำหรับ *SSLOF* เป็นผลรวมกำลังสองของการวัดความแปรผันของค่าเฉลี่ยของ Y ในแต่ละระดับของ X หาก X แต่ละค่าไม่ซ้ำกันเลยอาจทำได้โดยจัดกลุ่มค่า X ที่ใกล้เคียงกันเข้าด้วยกัน เมื่อเทียบกับค่าพยากรณ์ดังนี้

$$SSLOF = \sum_{i=1}^{m} n_i(\bar{Y}_i - \hat{Y}_i)^2 \tag{6.9}$$

โดยองศาเสรีของ *SSLOF* เท่ากับ $n - p$

สำหรับสถิติที่ใช้ในการทดสอบคือ F โดยมีสูตรการคำนวณดังนี้

$$F = \frac{SSLOF/(m-p)}{SSPE/(n-m)} = \frac{MSLOF}{MSPE} \tag{6.10}$$

หาก $F \geq F_{\alpha, m-p, n-m}$ แล้วสรุปว่าตัวแบบที่ได้ที่ได้ไม่เหมาะสมกับข้อมูล

ตารางข้างล่างแสดงการวิเคราะห์ความแปรปรวนโดยแยกออกเป็นความคลาดเคลื่อนแท้จริงและความคลาดเคลื่อนที่ไม่เหมาะสม

Source of variation	***SS***	***df***	***MS***	***F***
Regression	*SSR*	p	*MSR*	
Residual	*SSE*	$n-p$	*MSE*	
Lack of fit	*SSLOF*	$c-p$	*MSLOF*	F
Pure Error	*SSPE*	$n-c$	*MSPE*	
Total	*SST*	$n-1$		

6.2.3 สัมประสิทธิ์การตัดสินใจของการถดถอยพหุ

ค่า R^2 มีวิธีการคำนวณเหมือนกับในกรณีของการวิเคราะห์การถดถอยพหุแสดงได้ดังสมการ

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{6.11}$$

ตัวอย่างที่ 6.2 จากข้อมูลในตัวอย่างที่ 6.1 จงหาค่าสัมประสิทธิ์การตัดสินใจ

วิธีทำ

เนื่องจาก SSR = 45.379 และ SST = 46.717 ดังนั้น

$$R^2 = \frac{SSR}{SST} = \frac{45.379}{46.717}$$

$$= 0.9714$$

เมื่อเพิ่มความสัมพันธ์แบบกำลังสองระหว่างตัวแปร X และ Y สามารถอธิบายความแปรปรวนในตัวแปร Y ได้ถึง 97.14%

6.3 รูปแบบโพลิโนเมียลที่มีตัวแปรอิสระมากกว่า 1 ตัว

กรณีของตัวแปรอิสระที่มากกว่า 1 ตัวนั้นจะคล้ายกับในกรณีที่มีตัวแปรอิสระ 1 ตัวโดยสามารถเขียนสมการของตัวแปรอิสระ 2 ตัวได้ดังนี้คือ

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2 + \varepsilon \tag{6.12}$$

จากสมการนี้พบว่ามีสัมประสิทธิ์ของส่วนที่เป็นสมการเชิงเส้นตรงคือ β_1 และ β_2 สัมประสิทธิ์ของส่วนที่เป็นสมการกำลังสองคือ β_{11} และ β_{22} และสัมประสิทธิ์ของส่วนที่เป็นปฏิสัมพันธ์ระหว่างตัวแปรอิสระทั้งสอง (interaction) คือ β_{12}

การสร้างตัวแบบที่ซับซ้อนต้องใช้ความระมัดระวังในการสร้างและตีความอย่างมากดังนั้นหากสามารถใช้ตัวแบบที่ไม่ยุ่งยากแต่มีค่าสัมประสิทธิ์การตัดสินใจที่ไม่แตกต่างกันกับตัวแบบที่ซับซ้อนแล้วนักวิจัยควรเลือกตัวแบบที่มีรูปแบบที่ง่ายที่สุด (parsimonious model)

การวิเคราะห์และสร้างตัวแบบโพลิโนเมียลที่มีตัวแปรมากกว่า 1 ตัวนั้นจะเหมือนกับการสร้างตัวแบบที่มีตัวแปร 1 ตัวดังนั้นจะขอกล่าวถึงเฉพาะการทดสอบสมการกำลังสองและปฏิสัมพันธ์ของสมการถดถอยโดยมีสมมติฐานคือ

$H_0 : \beta_{11} = \beta_{22} = \beta_{12} = 0$

H_1 : มีพารามิเตอร์อย่างน้อย 1 ตัวที่ไม่เท่ากับ 0

ดังนั้นสมการภายใต้สมมติฐานหลักคือ

$$Y = \beta_0 + \beta_1(X_1 - \overline{X}_1) + \beta_2(X_2 - \overline{X}_2) + \varepsilon$$

และมีการแบ่งส่วนของผลรวมกำลังสองเป็น

$$SSR(\beta_{11}, \beta_{22}, \beta_{12} | \beta_1, \beta_2, \beta_0) = SSR(\beta_1, \beta_2, \beta_{11}, \beta_{22}, \beta_{12} | \beta_0) - SSR(\beta_1, \beta_2 | \beta_0) \qquad (6.13)$$

และการทดสอบทำโดยใช้สถิติ F ดังสมการ

$$F = \frac{SSR(\beta_{11}, \beta_{22}, \beta_{12} | \beta_0, \beta_1, \beta_2)/3}{MSE} \qquad (6.14)$$

โดย MSE เป็น MSE ที่ได้จากสมการ $Y = \beta_0 + \beta_1(X_1 - \overline{X}_1) + \beta_2(X_2 - \overline{X}_2) + \varepsilon$ และเปรียบเทียบค่า F ที่คำนวณได้กับ F จากตารางที่องศาเสรีเท่ากับ 3 และ $n - 3$ หากปฏิเสธสมมติหลักแสดงว่าพารามิเตอร์โพลิโนเมียลกำลังสองและปฏิสัมพันธ์มีความเหมาะสมกับข้อมูลนั้น

ตัวอย่างที่ 6.3 การศึกษาความสัมพันธ์ระหว่างผลผลิตของสารเคมีที่ได้ (Y) กับเวลา (X_1) และอุณหภูมิในการทำปฏิกิริยา (X_2) โดยมีข้อมูลดังนี้

X_1	76.0	80.5	78.0	89.0	93.0	92.1	77.8	84.0	87.3	75.0
X_2	170	165	182	185	180	172	170	180	165	172
Y	50.95	47.35	50.99	44.96	41.89	41.44	51.79	50.78	41.48	49.80

X_1	85.0	90.0	85.0	79.2	83.0	82.0	94.0	91.4	95.0	81.1
X_2	185	176	178	174	168	179	181	184	173	169
Y	48.74	46.20	50.49	52.78	49.71	52.75	39.41	43.63	38.19	50.92

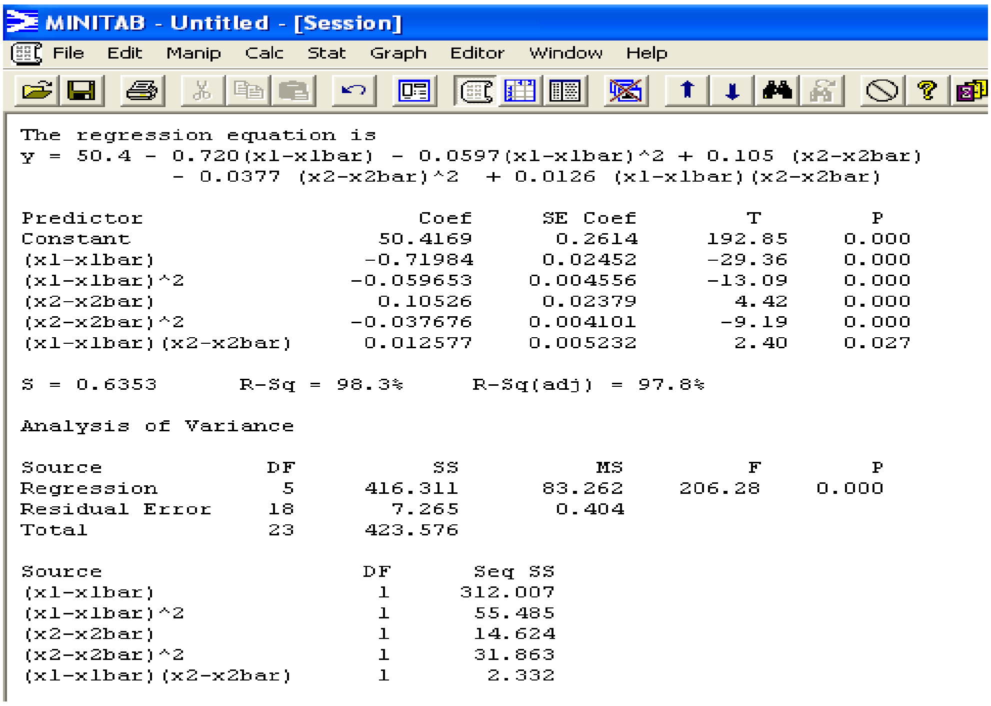
X_1	88.8	91.0	87.0	86.0
X_2	183	178	175	175
Y	46.55	44.28	48.72	49.13

วิธีทำ

จากข้อมูลพบว่าค่าเฉลี่ยของ X_1 และ X_2 เท่ากับ 85.467 และ 175.792 ตามลำดับ หากสร้างตัวแบบที่มีกำลังสองและปฏิสัมพันธ์จะมีสมการดังนี้

$$Y = \beta_0 + \beta_1(X_1 - \bar{X}_1) + \beta_2(X_2 - \bar{X}_2) + \beta_{11}(X_1 - \bar{X}_1)^2 + \beta_{22}(X_2 - \bar{X}_2)^2 + \beta_{12}(X_1 - \bar{X}_1)(X_2 - \bar{X}_2) + \varepsilon$$

เมื่อทำการวิเคราะห์ความแปรปรวนของตัวแบบนี้พบว่ามีอย่างน้อย 1 ตัวแปรอิสระที่สามารถอธิบายตัวแปรตามได้เนื่องจากค่า p-value ที่ได้จากตารางวิเคราะห์ความแปรปรวนน้อยกว่า 0.05 ดังภาพข้างล่างที่ได้จากโปรแกรม MINITAB

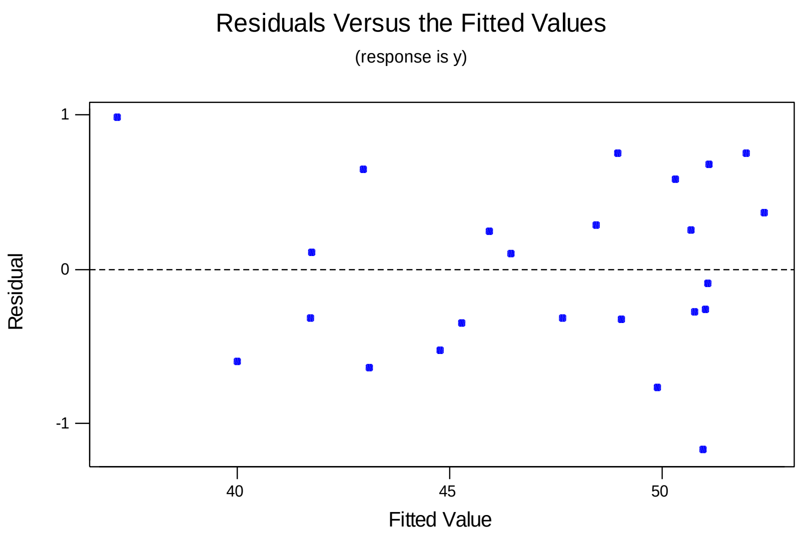



```
The regression equation is
y = 50.4 - 0.720(x1-x1bar) - 0.0597(x1-x1bar)^2 + 0.105 (x2-x2bar)
         - 0.0377 (x2-x2bar)^2  + 0.0126 (x1-x1bar)(x2-x2bar)

Predictor                  Coef     SE Coef        T       P
Constant                50.4169      0.2614   192.85   0.000
(x1-x1bar)             -0.71984     0.02452   -29.36   0.000
(x1-x1bar)^2          -0.059653    0.004556   -13.09   0.000
(x2-x2bar)              0.10526     0.02379     4.42   0.000
(x2-x2bar)^2          -0.037676    0.004101    -9.19   0.000
(x1-x1bar)(x2-x2bar)   0.012577    0.005232     2.40   0.027

S = 0.6353       R-Sq = 98.3%      R-Sq(adj) = 97.8%

Analysis of Variance

Source            DF        SS         MS        F       P
Regression         5   416.311     83.262   206.28   0.000
Residual Error    18     7.265      0.404
Total             23   423.576

Source                 DF     Seq SS
(x1-x1bar)              1    312.007
(x1-x1bar)^2            1     55.485
(x2-x2bar)              1     14.624
(x2-x2bar)^2            1     31.863
(x1-x1bar)(x2-x2bar)    1      2.332
```

ดังนั้นสมการถดถอยที่ได้คือ

$$Y = 50.4169 - 0.7198(X_1 - 85.467) + 0.1053(X_2 - 175.792) - 0.0597(X_1 - 85.467)^2 - 0.0377(X_2 - 175.792)^2 + 0.0126(X_1 - 85.467)(X_2 - 175.792)$$

เมื่อพิจารณากราฟความสัมพันธ์ระหว่างส่วนเหลือกับค่าพยากรณ์ไม่พบว่ามีรูปแบบที่ปกติแต่อย่างใดดังภาพข้างล่าง



หากทำการทดสอบว่าสมการกำลังสองและปฏิสัมพันธ์มีความสำคัญต่อข้อมูลชุดนี้หรือไม่ โดยมีสมมติฐานคือ

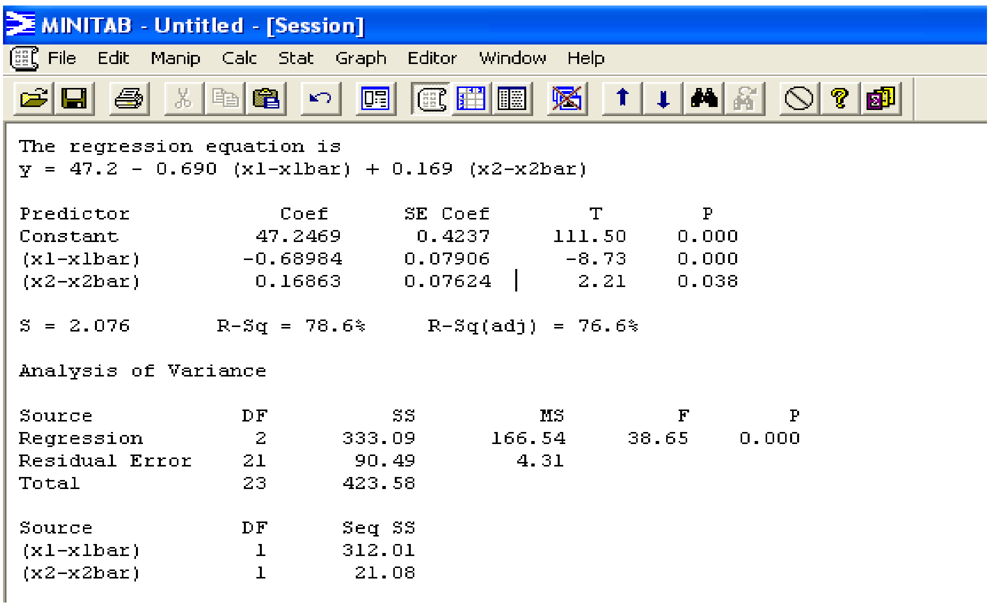
$$H_0 : \beta_{11} = \beta_{22} = \beta_{12} = 0$$

H_1 : มีพารามิเตอร์อย่างน้อย 1 ตัวที่ไม่เท่ากับ 0

ดังนั้นสมการถดถอยลดรูปคือ

$$Y = \beta_0 + \beta_1(X_1 - \bar{X}_1) + \beta_2(X_2 - \bar{X}_2) + \varepsilon$$

เมื่อทำการวิเคราะห์ความแปรปรวนโดยใช้สมการลดรูปจะได้ผลลัพธ์ดังภาพข้างล่าง




```
The regression equation is
y = 47.2 - 0.690 (x1-x1bar) + 0.169 (x2-x2bar)

Predictor          Coef     SE Coef        T       P
Constant        47.2469      0.4237   111.50   0.000
(x1-x1bar)     -0.68984     0.07906    -8.73   0.000
(x2-x2bar)      0.16863     0.07624     2.21   0.038

S = 2.076       R-Sq = 78.6%     R-Sq(adj) = 76.6%

Analysis of Variance

Source            DF        SS        MS        F        P
Regression         2    333.09    166.54    38.65    0.000
Residual Error    21     90.49      4.31
Total             23    423.58

Source            DF    Seq SS
(x1-x1bar)         1    312.01
(x2-x2bar)         1     21.08
```

จากตารางวิเคราะห์ความแปรปรวนของสมการลดรูปได้

$$SSR(\beta_{11}, \beta_{22}, \beta_{12} | \beta_1, \beta_2, \beta_0) = SSR(\beta_1, \beta_2, \beta_{11}, \beta_{22}, \beta_{12} | \beta_0) - SSR(\beta_1, \beta_2 | \beta_0)$$
$$= 416.311 - 333.09$$
$$= 83.221$$

และสถิติ F คือ

$$F = \frac{SSR(\beta_{11}, \beta_{22}, \beta_{12} | \beta_0, \beta_1, \beta_2)/3}{MSE}$$
$$= \frac{83.221/3}{0.404} = 68.66$$

เมื่อเปรียบเทียบค่าสถิติเอฟ $F = 68.66 > F_{0.05,\ 3,\ 18} = 5.09$ ดังนั้นจึงปฏิเสธสมมติฐานหลักและสรุปว่ามีอย่างน้อยหนึ่งพารามิเตอร์ที่มีผลต่อตัวแบบที่ระดับนัยสำคัญ 0.05 และเมื่อพิจารณาการทดสอบ t ของพารามิเตอร์แต่ละตัวของตัวแบบเต็มรูปพบว่าพารามิเตอร์ทุกตัวมีค่า p-value ที่น้อยกว่า 0.05 ดังนั้นพารามิเตอร์ทุกตัวมีความสำคัญต่อการพยากรณ์ข้อมูลชุดนี้

6.4 รูปแบบที่ไม่เป็นเส้นตรงที่แปลงเป็นเส้นตรงได้

หากพิจารณาหาความสัมพันธ์ของข้อมูลโดยการวาดกราฟระหว่างตัวแปรแล้วพบว่าข้อมูลมีความสัมพันธ์ที่ไม่เป็นเส้นตรงแต่สามารถแปลงให้อยู่ในรูปที่เป็นเส้นตรงได้ เช่น ความสัมพันธ์ในรูปแบบเอ็กซ์โปเนนเชียล รูปแบบไฮเปอร์โบลาหรือรูปแบบเอ็กซ์โปเนนเชียลผกผัน เป็นต้น ในกรณีเช่นนี้วิธีที่ง่ายที่สุดในการสร้างตัวแบบคือการแปลงให้อยู่ในรูปของสมการเส้นตรงแล้วจึงทำการวิเคราะห์เหมือนการวิเคราะห์สมการถดถอยเชิงเส้น การแปลงสามารถทำได้ดังนี้

6.4.1 รูปแบบเอ็กซ์โปเนนเชียล

ข้อมูลที่มีความสัมพันธ์แบบเอ็กซ์โปเนนเชียลสามารถเขียนอยู่ในรูปดังนี้

$$Y = \beta_0 \beta_1^X \tag{6.15}$$

ในการวิเคราะห์ข้อมูลสามารถแปลงข้อมูลได้โดยการใช้ลอการิทึมธรรมชาติดังนี้

$$\ln Y = \ln \beta_0 + (\ln \beta_1) X \tag{6.16}$$

หรือ

$$Y' = \beta_0' + \beta_1' X \tag{6.17}$$

โดย $\ln Y = Y'$

$\ln \beta_0 = \beta_0'$

$\ln \beta_1 = \beta_1'$

เมื่อทำการแปลงข้อมูลแล้วนักวิจัยสามารถวิเคราะห์ข้อมูลโดยการวิเคราะห์การถดถอยเชิงเส้น หากต้องการคำนวณค่าพยากรณ์นั้นนักวิจัยจำเป็นต้องแปลงตัวแบบให้กลับมาอยู่ในรูปเดิมเพื่อให้ค่าพยากรณ์มีหน่วยเหมือนกับข้อมูลเดิม

6.4.2 รูปแบบไฮเปอร์โบลา

ข้อมูลที่มีความสัมพันธ์แบบไฮเปอร์โบลาสามารถเขียนอยู่ในรูปดังนี้

$$Y = \frac{X}{\beta_0 X + \beta_1} \tag{6.18}$$

ในการวิเคราะห์ข้อมูลสามารถแปลงข้อมูลได้โดยการคำนวณส่วนกลับของตัวแปรทั้งสองดังนี้

$$Y' = \beta_0' + \beta_1' X \tag{6.19}$$

โดย $\frac{1}{Y} = Y'$

และ $\frac{1}{X} = X'$

เช่นเดียวกับรูปแบบเอ็กซ์โปเนนเชียลหลังจากการแปลงข้อมูลให้อยู่ในรูป (6.19) แล้วจึงทำการวิเคราะห์โดยการใช้สมการถดถอยเชิงเส้นและการคำนวณค่าพยากรณ์ต้องแปลงกลับให้อยู่ในรูปเดิมเช่นกัน

6.4.3 รูปแบบเอ็กซ์โปเนนเชียลผกผัน

ข้อมูลที่มีความสัมพันธ์แบบเอ็กซ์โปเนนเชียลผกผันสามารถเขียนอยู่ในรูปดังนี้

$$Y = \beta_0 e^{\beta_1 / X} \tag{6.20}$$

ในการวิเคราะห์ข้อมูลสามารถแปลงข้อมูลได้โดยการคำนวณส่วนกลับของตัวแปรอิสระและลอการิทึมธรรมชาติของตัวแปรตามดังนี้

$$Y' = \beta_0' + \beta_1' X \tag{6.21}$$

โดย $\ln Y = Y'$

และ $\frac{1}{X} = X'$

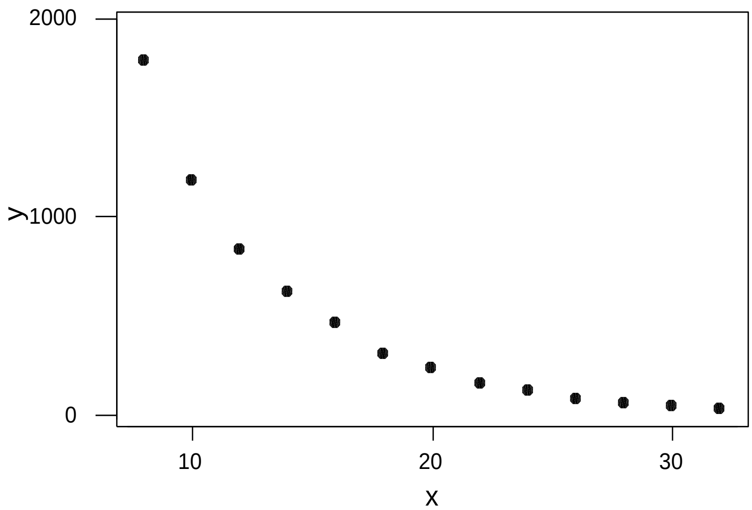
เช่นเดียวกับรูปแบบเอ็กซ์โปเนนเชียลหลังจากการแปลงข้อมูลให้อยู่ในรูป (6.21) แล้วจึงทำการวิเคราะห์โดยการใช้สมการถดถอยเชิงเส้นแต่การคำนวณค่าพยากรณ์ต้องแปลงกลับให้อยู่ในรูปเดิมของข้อมูล

ตัวอย่างที่ 6.4 จากข้อมูลข้างล่างจงสร้างสมการถดถอยที่เหมาะสมกับข้อมูลชุดนี้

X	*Y*
8	1788
10	1179
12	834
14	617
16	459
18	309
20	234
22	157
24	121
26	81
28	53
30	44
32	31

วิธีทำ

จากการพิจารณาแผนภาพกระจายระหว่างตัวแปรทั้งสองพบว่าข้อมูลมีความสัมพันธ์แบบเอ็กซ์โปเนนเชียลดังภาพข้างล่าง



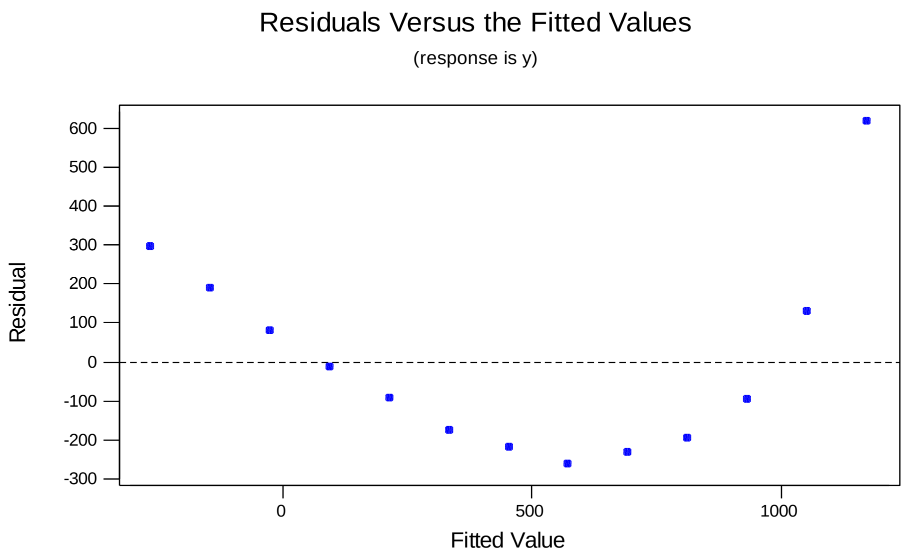
เมื่อนำข้อมูลมาสร้างสมการถดถอยเชิงเส้นอย่างง่ายพบว่าตัวแบบที่ได้ไม่เหมาะสมกับข้อมูลดังผลลัพธ์ที่ได้จากโปรแกรม MINITAB ข้างล่างโดยพิจารณาค่า *p*-value ของ “Lack of fit test” ที่มีค่าน้อยกว่า 0.05 ซึ่งสอดคล้องกับแผนภาพระหว่างส่วนเหลือกับค่าพยากรณ์เนื่องจากแผนภาพที่ได้มีลักษณะเป็นเส้นโค้งแสดงให้เห็นว่าตัวแบบที่ได้นี้ไม่เหมาะสมกับข้อมูล

MINITAB - CHAP6.MPJ - [Session]

File Edit Manip Calc Stat Graph Editor Window Help

Regression Analysis: y versus x

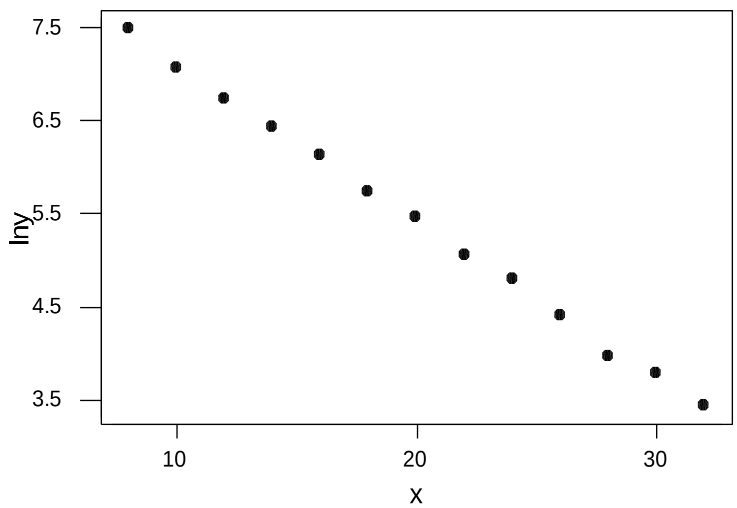
```
The regression equation is
y = 1651 - 59.8 x

Predictor        Coef    SE Coef         T       P
Constant       1650.9      211.6      7.80   0.000
x             -59.827      9.907     -6.04   0.000

S = 267.3        R-Sq = 76.8%     R-Sq(adj) = 74.7%

Analysis of Variance

Source            DF         SS         MS        F       P
Regression         1    2605702    2605702    36.47   0.000
Residual Error    11     785953      71450
Total             12    3391655


Lack of fit test
Possible curvature in variable x (P-Value = 0.003)
Possible lack of fit at outer X-values          (P-Value = 0.000)
Overall lack of fit test is significant at P = 0.000
```



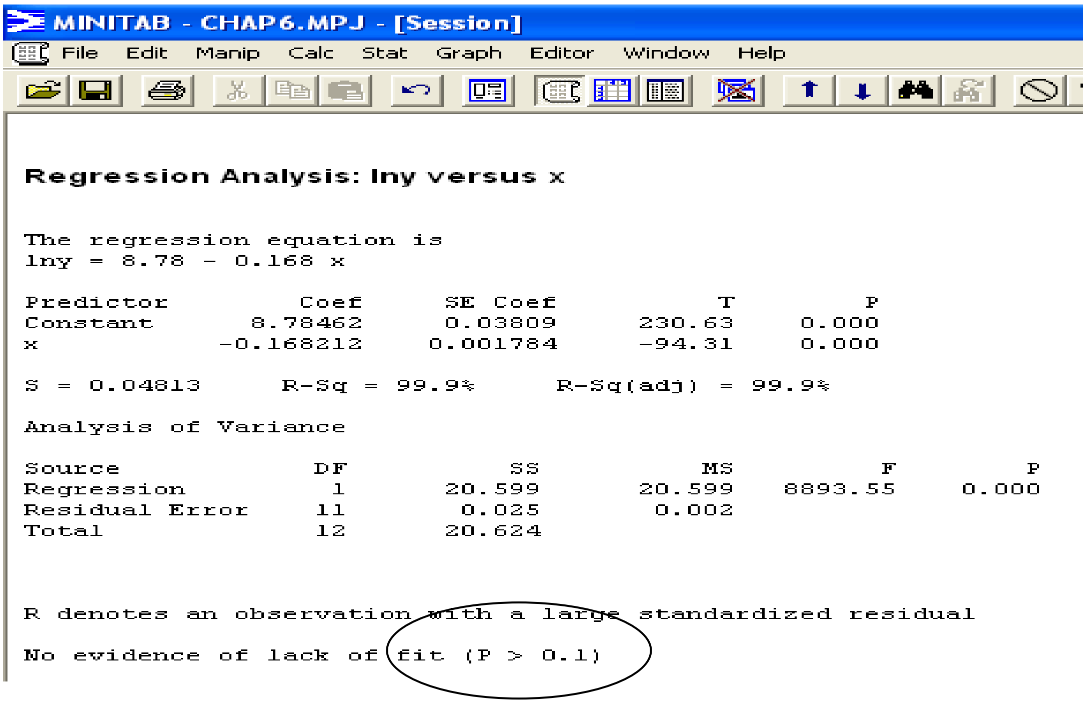
ดังนั้นจึงควรแปลงข้อมูลเพื่อให้สามารถวิเคราะห์ข้อมูลโดยใช้ตัวแบบการถดถอยเชิงเส้นได้ เนื่องจากความสัมพันธ์ระหว่างตัวแปรทั้งสองเป็นเอ็กซ์โปเนนเชียลจึงแปลงข้อมูลโดยการใช้ลอการิทึมธรรมชาติกับตัวแปรตามดังสมการ (6.16) และมีค่า Y' ดังนี้

Y	$Y' = \ln Y$
1788	7.48885
1179	7.07242
834	6.72623
617	6.42487
459	6.12905
309	5.73334
234	5.45532
157	5.05625
121	4.79579
81	4.39445
53	3.97029
44	3.78419
31	3.43399

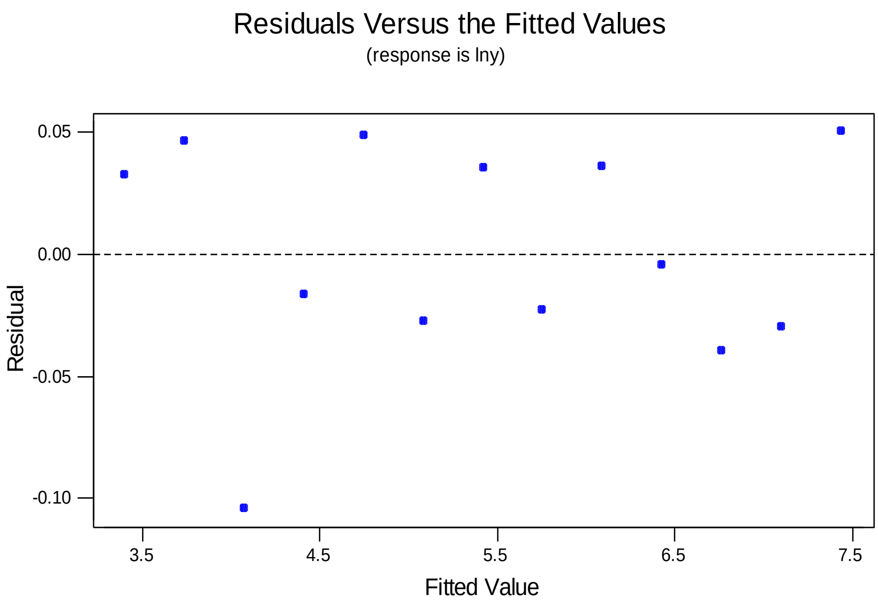
จากแผนภาพกระจายระหว่างตัวแปรอิสระและตัวแปรตามที่แปลงแล้วพบว่าความสัมพันธ์ที่ได้เป็นเส้นตรงดังแผนภาพข้างล่าง



เมื่อพิจารณาสมการถดถอยที่ได้จากตัวแปรอิสระและตัวแปรตามที่แปลงแล้วพบว่าสมการมีความเหมาะสมโดยพิจารณาจาก “Lack of fit test” ที่มีค่า *p*-value ที่มากกว่า 0.10 ดังภาพที่ได้จากโปรแกรม MINITAB ซึ่งสอดคล้องกับผลที่ได้จากแผนภาพกระจายระหว่างส่วนเหลือและค่าพยากรณ์ที่ไม่มีรูปแบบ



```
MINITAB - CHAP6.MPJ - [Session]
File  Edit  Manip  Calc  Stat  Graph  Editor  Window  Help

Regression Analysis: lny versus x

The regression equation is
lny = 8.78 - 0.168 x

Predictor        Coef      SE Coef        T        P
Constant      8.78462      0.03809   230.63    0.000
x           -0.168212     0.001784   -94.31    0.000

S = 0.04813     R-Sq = 99.9%     R-Sq(adj) = 99.9%

Analysis of Variance

Source            DF        SS        MS        F        P
Regression         1    20.599    20.599  8893.55    0.000
Residual Error    11     0.025     0.002
Total             12    20.624

R denotes an observation with a large standardized residual

No evidence of lack of fit (P > 0.1)
```



จากผลลัพธ์ที่ได้จากโปรแกรม MINITAB พบว่าสมการถดถอยคือ

$$\hat{Y}' = 8.78 - 0.168X$$

แต่เนื่องจากค่า $\hat{Y}'$ เป็นค่าลอการิทึมดังนั้นจึงต้องแปลงกลับมาให้อยู่ในรูปข้อมูลเดิมโดย $b_0 = e^{8.78} = 6502.877$ และ $b_1 = e^{-0.168} = 0.845$ ดังนั้นสมการถดถอยคือ

$$\hat{Y} = 6502877(0.845)^X$$

สรุป

การนำตัวแบบที่เป็นเชิงเส้นมาใช้กับข้อมูลที่มีความสัมพันธ์รูปแบบอื่นๆ ที่ไม่เป็นเส้นตรงนั้นถือว่าไม่เหมาะสมจะทำให้ความถูกต้องในการพยากรณ์ต่ำ หากข้อมูลใดสามารถแปลงให้อยู่ในรูปของเส้นตรงได้ควรแปลงข้อมูลก่อนการใช้ตัวแบบเชิงเส้น สำหรับข้อมูลโพลิโนเมียลทำได้โดยการเพิ่มพจน์ที่มีกำลังสูงเข้าไปในตัวแบบ ทั้งนี้ควรตรวจสอบความเหมาะสมของรูปแบบที่ได้

คำถามท้ายบท

6.1 วิธีตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามว่าเป็นเส้นตรงหรือไม่อย่างง่ายทำได้อย่างไร

6.2 จากผลลัพธ์ที่ได้จากโปรแกรม MINITAB ข้างล่างจงคำนวณค่า $SSR(\beta_1 | \beta_0)$ $SSR(\beta_1, \beta_2 | \beta_0)$ และ $SSR(\beta_2 | \beta_0, \beta_1)$

Regression Analysis: y versus x1, x2

```
The regression equation is
y = - 20 + 13.4 x1 + 244 x2

Predictor        Coef     SE Coef          T        P
Constant        -20.4       652.7      -0.03    0.976
X1             13.350       7.672       1.74    0.107
X2             243.71       63.51       3.84    0.002

S = 547.7       R-Sq = 58.2%     R-Sq(adj) = 51.3%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         2     5018232     2509116      8.36    0.005
Residual Error    12     3600196      300016
Total             14     8618428


Source       DF      Seq SS
X1            1      600498
X2            1     4417734
```

6.3 จากข้อมูลข้างล่าง

Y	7.2	7.3	7.5	10	11	13	17	28	39	58
X	2	5	8	10	12	15	18	20	22	25

(1) จงวาดแผนภาพกระจายระหว่างตัวแปรทั้งสอง

(2) จงสร้างสมการถดถอยอย่างง่ายพร้อมทั้งคำนวณค่าสัมประสิทธิ์การตัดสินใจ

(3) จงวาดแผนภาพกระจายระหว่าง X กับค่าคลาดเคลื่อนพร้อมทั้งอธิบาย

(4) จงอธิบายว่าตัวแบบที่ได้ในข้อ (2) เหมาะสมหรือไม่

6.4 จากข้อมูลใน 6.3 จงสร้างสมการถดถอยโดยเพิ่มตัวแปรอิสระกำลังสองและทดสอบ $H_0 : \beta_2 = 0$ ว่ากำลังสองมีความจำเป็นหรือไม่ที่ระดับนัยสำคัญ 0.05 พร้อมทั้งวาดแผนภาพกระจายระหว่าง X กับค่าคลาดเคลื่อน

6.5 ท่านสามารถใช้แผนภาพกระจายของค่าคลาดเคลื่อนช่วยในการตรวจสอบความเหมาะสมของตัวแบบได้อย่างไร

6.6 จากข้อมูลข้างล่าง

Y	20	20	20	25	25	30	33	33	35
X	123	124	120	195	198	285	335	337	390

(1) จงสร้างสมการถดถอยโดยเพิ่มตัวแปรอิสระกำลังสองและทดสอบ $H_0 : \beta_2 = 0$ ว่ากำลังสองมีความจำเป็นหรือไม่ที่ระดับนัยสำคัญ 0.05

(2) จงสร้างสมการถดถอยเชิงเส้นอย่างง่ายพร้อมทั้งทดสอบตัวแบบว่าเหมาะสมหรือไม่

(3) เปรียบเทียบแผนภาพกระจายของตัวแปร X กับค่าคลาดเคลื่อนระหว่างตัวแบบทั้งสอง

6.7 จากข้อมูลในข้อ 5.6 จงสร้าง

(1) สมการถดถอยพหุและทดสอบความเหมาะสมของตัวแบบที่ระดับนัยสำคัญ 0.05 พร้อมทั้งวาดแผนภาพกระจายของค่าคลาดเคลื่อนกับค่าพยากรณ์

(2) สมการถดถอยพหุโดยเพิ่มตัวแปรโพลิโนเมียลและปฏิสัมพันธ์ระหว่างตัวแปรอิสระทั้งสอง จากนั้นทดสอบความเหมาะสมของตัวแบบที่ระดับนัยสำคัญ 0.05

6.8 จากข้อมูลในข้อ 5.6 จงทดสอบว่าสามารถละตัวแปรความสูงจากสมการได้หรือไม่ที่ระดับนัยสำคัญ 0.05 โดยกำหนดให้สมการมีตัวแปรเส้นผ่าศูนย์กลางอยู่แล้วโดยใช้ค่าเอฟ พร้อมทั้งคำนวณค่า *SSR*(เส้นผ่าศูนย์กลาง) *SSR*(ความสูง) *SSR*(ความสูง| เส้นผ่าศูนย์กลาง) และ *SSR*(เส้นผ่าศูนย์กลาง| ความสูง)

6.9 จากข้อมูลในข้อ 5.9 จง

(1) ทดสอบว่าสามารถละตัวแปรความดันจากสมการได้หรือไม่ที่ระดับนัยสำคัญ 0.05 โดยกำหนดให้สมการมีตัวแปรอุณหภูมิอยู่แล้วโดยใช้ค่า F

(2) ทดสอบว่า SSR(อุณหภูมิ) + SSR(ความดัน| อุณหภูมิ) = SSR(ความดัน) + SSR(อุณหภูมิ| ความดัน) หรือไม่

(3) หาค่าสัมประสิทธิ์การตัดสินใจสำหรับตัวแปรอิสระหลายตัว

6.10 จากข้อมูลในข้อ 5.10 จง

(1) สร้างสมการถดถอยพหุโดยเพิ่มตัวแปรโพลิโนเมียลและปฏิสัมพันธ์ระหว่างตัวแปรอิสระทั้งสอง จากนั้นว่าตัวแปรกำลังสองแต่ละตัวและปฏิสัมพันธ์มีความจำเป็นหรือที่ระดับนัยสำคัญ 0.05 โดยใช้ F

(2) หาค่า SSR(pH) เท่ากับ SSR(pH | ปริมาณกรด) หรือไม่ เพราะเหตุใด

(3) จงวาดแผนภาพกระจายของค่าคลาดเคลื่อนที่ได้จาก (1) กับตัวแปรอิสระแต่ละตัว พร้อมทั้งอธิบายค่าที่ได้

(4) หากต้องทำการแปลงข้อมูลควรทำอย่างไร

6.11 จากข้อมูลข้างล่างจง

Y	1	1	2	2	3	3	4	4	5	5	6	6
X	44	40	56	62	74	76	82	78	68	74	54	60

(1) วาดแผนภาพกระจายระหว่างตัวแปรทั้งสอง

(2) สร้างสมการถดถอยเชิงเส้นอย่างง่ายพร้อมทั้งทดสอบความเหมาะสมของตัวแบบที่ระดับนัยสำคัญ 0.05 พร้อมทั้งอธิบาย

(3) วาดแผนภาพกระจายระหว่างค่าคลาดเคลื่อนกับตัวแปรอิสระพร้อมทั้งอธิบาย

(4) สร้างสมการถดถอยโดยเพิ่มตัวแปรอิสระกำลังสองและทดสอบ $H_0 : \beta_2 = 0$ ว่ากำลังสองมีความจำเป็นหรือไม่ที่ระดับนัยสำคัญ 0.05

(5) คำนวณค่าสัมประสิทธิ์การตัดสินใจพร้อมทั้งอธิบาย

(6) หากต้องแปลงข้อมูลท่านจะแปลงข้อมูลอย่างไร