

บทที่ 11

การวิเคราะห์การถดถอยพหุ

1. ทฤษฎีเกี่ยวกับการวิเคราะห์การถดถอยพหุ

วัตถุประสงค์ทั่วไปของการวิเคราะห์การถดถอยพหุ คือ ต้องการศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระหรือตัวแปรทำนายหลายตัวกับตัวแปรตามหรือตัวแปรเกณฑ์ 1 ตัว ตัวอย่างเช่น การตอบสนองของสัตว์ทดลองต่อยาบางชนิดอาจขึ้นอยู่กับปริมาณยา อายุ และน้ำหนักของสัตว์ทดลองโดยกำหนดให้ตัวแปรตาม Y แทนการตอบสนองของสัตว์ทดลองและให้ตัวแปรอิสระ X_1 แทนปริมาณยา, X_2 แทนอายุ, และ X_3 แทนน้ำหนักของสัตว์ทดลอง เป็นต้น ซึ่งในบทนี้จะอธิบายเกี่ยวกับตัวแบบการถดถอย วิธีการหาสมการถดถอย การประเมินสมการถดถอยและการใช้สมการถดถอย

ข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยพหุ คือ

- (1) ตัวแปรอิสระ X_i ใด ๆ เป็นตัวแปรสุ่มแต่กำหนดให้คงที่
- (2) สำหรับแต่ละเซตของค่า X_i จะมี Y เป็นตัวแปรเชิงสุ่มที่มีการแจกแจงความน่าจะเป็นแบบปกติที่มีค่าเฉลี่ยและความแปรปรวนคือ $Y \sim N(\mu, \sigma^2)$
- (3) ค่าสังเกต Y ทุกตัวเป็นอิสระกัน
- (4) ค่าเฉลี่ยของ Y ที่แต่ละจุดของตัวแปรอิสระทุกตัว X_1, X_2, \dots, X_k เป็นฟังก์ชันเส้นตรงของ X_1, X_2, \dots, X_k นั่นคือ

$$\mu_{Y|X_1, X_2, \dots, X_k} = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

- (5) ความแปรปรวนของ Y เท่ากันหมดสำหรับทุกจุดของตัวแปรอิสระทุกตัว X_1, X_2, \dots, X_k นั่นคือ

$$\sigma_{Y|X_1, X_2, \dots, X_k}^2 = \text{Var}(Y|X_1, X_2, \dots, X_k) \equiv \sigma^2$$

(6) ตัวแปร Y มีการแจกแจงแบบปกติสำหรับทุกจุดของตัวแปรอิสระทุกตัว X_1, X_2, \dots, X_k เขียนได้ดังนี้

$$Y \sim N(\mu_{Y|X_1, X_2, \dots, X_k}, \sigma^2)$$

ตัวแบบการถดถอยพหุเชิงเส้นแบบทั่วไปที่มีตัวแปรอิสระ k ตัวคือ

$$y_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \dots + \beta_k X_{kj} + e_j$$

เมื่อ y_j คือ ค่าตัวหนึ่งจากประชากรย่อยของค่า Y ประชากรหนึ่งเป็นตัวแปรตาม

$\beta_0, \beta_1, \dots, \beta_k$ คือ สัมประสิทธิ์การถดถอยบางส่วน (partial - regression coefficient)

β_i แทนค่าการเปลี่ยนแปลงของ y สำหรับ x_i เปลี่ยนไป 1 หน่วย เมื่อตัวแปร x ตัวอื่น ๆ ทั้งหมดคงที่

$X_{1j}, X_{2j}, \dots, X_{kj}$ คือค่าเฉพาะของตัวแปรอิสระ X_1, X_2, \dots, X_k

e_j คือ ความคลาดเคลื่อนที่แสดงความแตกต่างระหว่างค่าสังเกต y แต่ละตัวกับ

$\mu_{Y|X_1, X_2, \dots, X_k}$ เป็นตัวแปรสุ่มที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนคงที่เท่ากับ σ^2 สำหรับ

แต่ละค่าของ X และความคลาดเคลื่อนทุกตัวเป็นอิสระกัน

ขั้นตอนการวิเคราะห์การถดถอยเหมือนกับการวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายที่อธิบายไว้แล้วในบทที่ 9

การประมาณสมการถดถอยพหุด้วยวิธีกำลังสองน้อยที่สุดจะได้ตัวแบบซึ่งผลบวกกำลังสองของระยะห่างระหว่างค่าสังเกตและค่าทำนายมีค่าน้อยที่สุด นั่นคือ ถ้าให้ตัวแบบที่เหมาะสมกับข้อมูลคือ

$$y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_k X_k$$

ดังนั้นผลบวกกำลังสองที่น้อยที่สุดของระยะห่างระหว่างค่าสังเกต (y) และค่าทำนาย (\hat{y}) คือ

$$\sum_{j=1}^n (y_j - \hat{y}_j)^2 = \sum_{i=1}^n (y_j - \hat{\beta}_0 - \hat{\beta}_1 X_{1j} - \dots - \hat{\beta}_k X_{kj})^2$$

ซึ่งเรียกว่าผลบวกกำลังสองของความคลาดเคลื่อนหรือ **SSE (error sum of square)**

การวิเคราะห์ความแปรปรวนสำหรับการวิเคราะห์การถดถอยพหุ โดยการคำนวณหาผลบวกกำลังสองของแต่ละเทอมในการวิเคราะห์ความแปรปรวน ซึ่งมีความสัมพันธ์กันดังนี้

$$\sum_{j=1}^n (y_j - \bar{y})^2 = \sum_{j=1}^n (\hat{y}_j - \bar{y})^2 + \sum_{j=1}^n (y_j - \hat{y}_j)^2$$

โดยที่ $\sum_{j=1}^n (y_j - \bar{y})^2$ เรียกว่า ผลบวกกำลังสองของทั้งหมด หรือ **SST (total sum of square)**

$\sum_{j=1}^n (\hat{y}_j - \bar{y})^2$ เรียกว่า ผลบวกกำลังสองของการถดถอยหรือ **SSR (regression sum of square)**

การวิเคราะห์ความแปรปรวนสำหรับการวิเคราะห์การถดถอยพหุสามารถสรุปลงในตารางวิเคราะห์ความแปรปรวนได้ดังนี้

ตารางที่ 11.1 ตารางวิเคราะห์ความแปรปรวนสำหรับการวิเคราะห์การถดถอยพหุ

Source of Variation	Degree of Freedom	Sum of Square	Mean Square	F
Regression	k	SSR	MSR	F_0
Error	n - k - 1	SSE	MSE	
Total	n - 1	SST		

จำนวนชั้นอิสระของแหล่งความแปรปรวนแหล่งต่าง ๆ ในการวิเคราะห์ความแปรปรวนคือ จำนวนชั้นอิสระของ **Regression** , **Error** , และ **Total** เท่ากับ k , n - k - 1 , และ n - 1 ตามลำดับ

การคำนวณค่าเฉลี่ยของผลบวกกำลังสอง (**Mean Square**) คำนวณจากผลบวกกำลังสองหารด้วยจำนวนชั้นอิสระของเทอมนั้น

การคำนวณค่าสถิติทดสอบ (F_0) คำนวณจาก $F_0 = MSR/MSE$ ซึ่งมีจำนวนชั้นอิสระ $df = k$ และ $n - k - 1$

2. การทดสอบสมมติฐานในการวิเคราะห์การถดถอยพหุ

การวิเคราะห์การถดถอยพหุสามารถตอบคำถามที่สำคัญ 3 ข้อ คือ

- (1) ตัวแปรอิสระทั้งหมดมีส่วนสำคัญในการใช้ทำนาย Y ใฉ่อย่างมีนัยสำคัญ
- (2) การเพิ่มตัวแปรอิสระที่สนใจเข้าไปในตัวแบบอีก 1 ตัว มีส่วนสำคัญในการใช้ทำนาย Y ใฉ่เพิ่มขึ้นอย่างมีนัยสำคัญ
- (3) การเพิ่มตัวแปรอิสระที่สนใจเข้าไปในตัวแบบอีก 1 ชุด มีส่วนสำคัญในการใช้ทำนาย Y ใฉ่เพิ่มขึ้นอย่างมีนัยสำคัญ

คำถามเหล่านี้สามารถหาคำตอบได้จากการทดสอบสมมติฐานทางสถิติคือ $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$ คู่กับ $H_1 : \text{มี } \beta_i \text{ อย่างน้อย 1 ค่าที่ไม่เท่ากับ 0 เมื่อ } i = 1, 2, \dots, k$

ผลการทดสอบสมมติฐานพิจารณาจากค่าสถิติทดสอบ F_0 ในตารางวิเคราะห์ความแปรปรวนว่ามีนัยสำคัญทางสถิติหรือไม่ ถ้าค่า Sig. มีค่าน้อยกว่าระดับนัยสำคัญที่กำหนด $\alpha = .05$ จะสรุปว่าปฏิเสธ H_0 แสดงว่ามีตัวแปรอิสระ X อย่างน้อย 1 ตัวแปรที่มีความสัมพันธ์กับตัวแปรตาม Y หมายความว่า ตัวแบบ $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$ ใช้แทนรูปแบบความสัมพันธ์ระหว่างตัวแปรตาม Y และตัวแปรอิสระ X ทั้ง k ตัวได้ หรือตัวแปรอิสระ X ทั้ง k ตัวมีส่วนสำคัญในการใช้ทำนายตัวแปรตาม Y ใฉ่อย่างมีนัยสำคัญทางสถิติ

ถ้าสรุปว่าปฏิเสธ H_0 ขึ้นต่อไปต้องทดสอบสมมติฐานเกี่ยวกับ β_i ทีละตัวเมื่อควบคุมตัวแปรอิสระตัวอื่น ๆ ใฉ่คงที่เพื่อหาว่ามีตัวแปรอิสระตัวใดบ้างที่มีความสัมพันธ์กับตัวแปรตาม Y อย่างมีนัยสำคัญ สมมติฐานที่ต้องการทดสอบคือ $H_0 : \beta_i = 0$ คู่กับ $H_1 : \beta_i \neq 0 ; i = 1, 2, \dots, k$ สถิติทดสอบคือ สถิติทดสอบ t ผลการทดสอบสมมติฐานถ้าค่าสถิติทดสอบ t มีนัยสำคัญคือ ค่า Sig. มีค่าน้อยกว่าระดับนัยสำคัญที่กำหนด $\alpha = .05$ ก็แสดงว่าตัวแปรอิสระ X_i ตัวนั้นมีความสัมพันธ์เชิงเส้นกับตัวแปรตาม Y อย่างมีนัยสำคัญทางสถิติ เรายังจะนำตัวแปรอิสระ X_i ตัวนั้นเข้าไปอยู่ในตัวแบบการถดถอยเชิงเส้น

3. การทดสอบสมมติฐานเกี่ยวกับ β_i ทีละตัว

สถิติที่ใช้ในการทดสอบสมมติฐานเกี่ยวกับค่าสัมประสิทธิ์การถดถอย β_i และ β_0 คือ สถิติทดสอบ t ผลการทดสอบสมมติฐานพิจารณาจากค่าสถิติทดสอบ t ว่ามีนัยสำคัญทางสถิติหรือไม่ ถ้าค่า Sig. มีค่าน้อยกว่าระดับนัยสำคัญที่กำหนด $\alpha = .05$ จะสรุปว่าปฏิเสธ $H_0 : \beta_i = 0$ นั่นคือ ตัวแปรอิสระ X_i สามารถอธิบายความผันแปรในตัวแปรตาม Y ได้ เมื่อกำหนดให้ตัวแปรอิสระตัวอื่น ๆ คงที่ และถ้ายอมรับ H_0 จะสรุปว่าตัวแปรอิสระ X_i ไม่สามารถอธิบายความผันแปรในตัวแปรตาม Y เมื่อกำหนดให้ตัวแปรอิสระตัวอื่น ๆ คงที่

สำหรับการทดสอบสมมติฐานเกี่ยวกับ β_0 ซึ่งเป็นค่าคงที่ในสมการถดถอย สมมติฐานที่ต้องการทดสอบคือ $H_0 : \beta_0 = 0$ คู่กับ $H_1 : \beta_0 \neq 0$ ผลการทดสอบสมมติฐานพิจารณาแบบเดียวกัน ถ้าสรุปว่ายอมรับ $H_0 : \beta_0 = 0$ นั่นคือเส้นถดถอยเชิงเส้นที่แสดงความสัมพันธ์ระหว่างตัวแปรตาม Y กับตัวแปรอิสระ X_i ต่าง ๆ ผ่านจุดกำเนิดและถ้าสรุปว่าปฏิเสธ H_0 นั่นคือเส้นถดถอยไม่ผ่านจุดกำเนิด

4. สัมประสิทธิ์การกำหนดเชิงพหุ สัมประสิทธิ์สหสัมพันธ์พหุ และสัมประสิทธิ์สหสัมพันธ์บางส่วน (Multiple Coefficient of Determination, Multiple Coefficient of Correlation and Coefficient of Partial Correlation)

4.1 สัมประสิทธิ์การกำหนดเชิงพหุ

การประเมินคุณภาพของเส้นถดถอยเชิงเส้นตรงที่ได้จากกลุ่มตัวอย่างโดยตรวจสอบว่าตัวแบบที่ได้สามารถอธิบาย Y ได้ดีอย่างไร โดยพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์พหุ (R) และค่าสัมประสิทธิ์การกำหนด (R Square) เพื่อดูว่าตัวแปรอิสระทั้งหมด (X_1, X_2, \dots, X_k) ในสมการนั้นมีความสัมพันธ์กับตัวแปรตามมากน้อยเพียงใด และสามารถอธิบายความผันแปรของตัวแปรตาม Y ได้ร้อยละเท่าไร นั่นคือ ผลบวกกำลังสองของเส้นถดถอยส่วน ผลบวกกำลังสองของทั้งหมด ค่าของสัดส่วนที่ได้นี้เรียกว่า สัมประสิทธิ์การกำหนด แทนด้วย $R_{Y.123\dots k}^2$ หรือ R^2 เขียนเป็นสูตรได้คือ

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

ความหมายของ R^2 ได้อธิบายไว้แล้วในบทที่ 10 การวิเคราะห์การถดถอยอย่างง่าย

4.2 สัมประสิทธิ์สหสัมพันธ์เชิงพหุ

เป็นการวัดความสัมพันธ์ระหว่างตัวแปรตาม Y กับตัวแปรอิสระ X_1, X_2, \dots, X_k และเป็นรูปทั่วไปของสัมประสิทธิ์สหสัมพันธ์อย่างง่ายที่เป็นความสัมพันธ์ระหว่างตัวแปรอิสระหลายตัว เขียนแทนด้วยสัญลักษณ์ $R_{y.12\dots k}$ และยังสามารถใช้แทนสัมประสิทธิ์สหสัมพันธ์อย่างง่ายระหว่าง y_i และ y คือ

$$R_{y.12\dots k} = r_{y.\hat{y}}$$

สัมประสิทธิ์สหสัมพันธ์พหุคูณ คือ รากที่สองของสัมประสิทธิ์การกำหนดเชิงพหุ

$$\begin{aligned} R_{y.12\dots k} &= \sqrt{R_{y.12\dots k}^2} \\ &= \sqrt{\frac{\sum (y - \hat{y})^2}{\sum (y_i - \bar{y})^2}} \end{aligned}$$

4.3 สัมประสิทธิ์สหสัมพันธ์บางส่วน

สัมประสิทธิ์สหสัมพันธ์บางส่วน คือ ค่าความสัมพันธ์เชิงเส้นระหว่างตัวแปร 2 ตัว เมื่อตัวแปรตัวอื่น ๆ ที่เหลือคงที่หรือควบคุมอิทธิพลของตัวแปรตัวอื่น ๆ แล้ว ตัวอย่างเช่น ค่าความสัมพันธ์ระหว่างตัวแปร Y และ X_1 เมื่อควบคุมอิทธิพลของตัวแปร X_2 หรือให้ X_2 คงที่ เขียนแทนด้วยสัญลักษณ์ $r_{y1.2}$ สูตรการคำนวณคือ

$$r_{y1.2} = \frac{(r_{y1} - r_{y2} r_{12})}{\sqrt{(1 - r_{y2}^2)(1 - r_{12}^2)}}$$

สัมประสิทธิ์สหสัมพันธ์บางส่วนระหว่าง Y และ X_2 เมื่อให้ X_1 คงที่คือ

$$r_{y2.1} = \frac{(r_{y2} - r_{y1} r_{12})}{\sqrt{(1-r_{y1}^2)(1-r_{12}^2)}}$$

และสัมประสิทธิ์สหสัมพันธ์บางส่วนระหว่าง X_1 และ X_2 เมื่อให้ Y คงที่คือ

$$r_{12.y} = \frac{(r_{12} - r_{y1} r_{y2})}{\sqrt{(1-r_{y1}^2)(1-r_{y2}^2)}}$$

เมื่อ r_{y1} คือ สัมประสิทธิ์สหสัมพันธ์อย่างง่ายระหว่าง Y และ X_1

r_{y2} คือ สัมประสิทธิ์สหสัมพันธ์อย่างง่ายระหว่าง Y และ X_2

r_{12} คือ สัมประสิทธิ์สหสัมพันธ์อย่างง่ายระหว่าง X_1 และ X_2

5. การทดสอบสมมติฐานเกี่ยวกับสัมประสิทธิ์สหสัมพันธ์พหุ

สมมติฐานที่ต้องการทดสอบคือ $H_0 : \rho_{y.12\dots k} = 0$ คู่กับ $H_1 : \rho_{y.12\dots k} \neq 0$

สถิติทดสอบคือ

$$F = \frac{R_{y.12\dots k}^2}{1 - R_{y.12\dots k}^2} \cdot \frac{n - k - 1}{k}$$

ซึ่งมีการแจกแจงแบบ F ที่มีจำนวนชั้นอิสระ k และ $n - k - 1$

เหมือนการทดสอบสมมติฐาน $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$

6. การทดสอบสมมติฐานเกี่ยวกับสัมประสิทธิ์สหสัมพันธ์บางส่วน

สมมติฐานที่ต้องการทดสอบ คือ $H_0 : \rho_{y1.2} = 0$ คู่กับ $H_1 : \rho_{y1.2} \neq 0$

สถิติทดสอบคือ

$$t = r_{y1.2} \sqrt{\frac{n-k-1}{1-r_{y1.2}^2}}$$

ซึ่งมีการแจกแจงแบบ student's t ที่มีจำนวนชั้นอิสระ $df = n - k - 1$

7. การใช้คำสั่ง **Correlate , Partial ...** คำนวณค่า และทดสอบสมมติฐานเกี่ยวกับสัมประสิทธิ์

สหสัมพันธ์บางส่วน

ตัวอย่างเช่น การศึกษาเกี่ยวกับนิโคติน หน่วยทดลองคือ ผู้ชายที่สูบบุหรี่ อายุ 24 ถึง 48 ปี ที่สูบบุหรี่ประจำ จำนวน 9 คน เก็บข้อมูล 3 อย่าง คือ puffs per cigarette (X_1) total particulate matter per cigarette (X_2) และ nicotine intake per cigarette (Y) ผู้วิจัยต้องการวิเคราะห์หาความสัมพันธ์ระหว่าง 3 ตัวแปรนี้ ข้อมูลแสดงในตารางที่ 11.2

ตารางที่ 11.2 ข้อมูลผู้สูบบุหรี่ 9 คน

X_1	X_2	Y
7.5	21.9	1.38
9.0	46.4	1.78
8.5	24.0	1.68
10.0	28.8	2.12
14.5	43.8	3.26
11.0	48.1	2.98
9.0	50.8	2.56
12.0	47.8	3.47
14.0	49.1	3.22

แหล่งที่มา : Neal L. Benowitz , Peyton Jacob III , Charles Denaro , and Roger Jenkins , “Stable ISotope Studies of Nicotine Kinetics and Bioavailability,” *Clinical Phamacology & Therapeutics* , 49 (1991) , 270-277. (อ้างถึงใน Daniel, 1995)

ข้อมูลอยู่ในแฟ้มข้อมูล **Corrpart.sav** มีขั้นตอนในการใช้คำสั่ง ดังนี้

1. ไปที่เมนูบาร์ คลิกที่ **Analyze , Correlate , Partial...** จะได้หน้าต่าง **Partial Correlations**

2. ในหน้าต่าง **Partial Correlations** ในช่องซ้ายมือ คลิกที่ตัวแปร **Y** และ **X₁** ให้ย้ายเข้าไปอยู่ในช่อง **Variables :** และคลิกที่ตัวแปร **X₂** ให้ย้ายเข้าไปอยู่ในช่อง **Controlling for :** เพื่อคำนวณหาสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร **Y** และ **X₁** เมื่อควบคุมตัวแปร **X₂**

ในกรอบของ **Test of Significance**

เลือก **Two - tailed** เพื่อทดสอบสมมติฐาน $H_0 : \rho_{y1.2} = 0$ คู่กับ $H_1 : \rho_{y1.2} \neq 0$

และเลือก **Display actual significance level** เพื่อให้แสดงค่าระดับนัยสำคัญ

คลิกที่ปุ่ม **Options ...** จะได้หน้าต่าง **Partial Correlations : Options** ดังภาพที่ 11.1

3. ในหน้าต่าง **Partial Correlations : Options**

ในกรอบของ **Statistics**

เลือก **Zero - order correlations** เพื่อคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรที่ละคู่ โดยไม่มีการควบคุมตัวแปรตัวอื่น ๆ

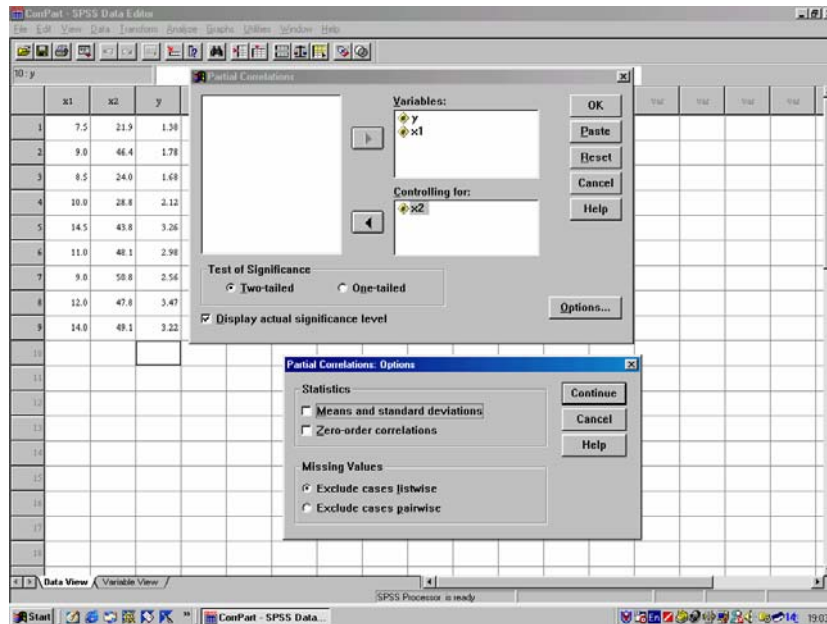
ในกรอบของ **Missing Values**

เลือก **Exclude cases listwise**

แล้วคลิกที่ปุ่ม **Continue** หน้าต่างนี้จะถูกปิดไป

4. ในหน้าต่าง **Partial Correlations**

คลิกที่ปุ่ม **OK** จะได้ผลลัพธ์ดังภาพที่ 11.2



ภาพที่ 11.1

Partial Corr

--- PARTIAL CORRELATION COEFFICIENTS ---

Zero Order Partials

	Y	X1	X2
Y	1.0000 (0) P= .	.8764 (7) P= .002	.7499 (7) P= .020
X1	.8764 (7) P= .002	1.0000 (0) P= .	.5608 (7) P= .116
X2	.7499 (7) P= .020	.5608 (7) P= .116	1.0000 (0) P= .

(Coefficient / (D.F.) / 2-tailed Significance)

". " is printed if a coefficient cannot be computed

-

--- PARTIAL CORRELATION COEFFICIENTS ---

Controlling for.. X2

	Y	X1
Y	1.0000 (0) P= .	.8323 (6) P= .010
X1	.8323 (6) P= .010	1.0000 (0) P= .

(Coefficient / (D.F.) / 2-tailed Significance)

". " is printed if a coefficient cannot be computed

ภาพที่ 11.2

การแปลผล

การคำนวณค่าสัมประสิทธิ์สหสัมพันธ์อย่างง่าย r_{y1} , r_{y2} , และ r_{12} เท่ากับ 0.8764 , 0.7499 , และ 0.5608 ตามลำดับ สมมติฐานที่ต้องการทดสอบคือ $H_0 : \rho = 0$ คู่กับ $\rho \neq 0$ สถิติทดสอบ t ที่มีจำนวนชั้นอิสระ $df = n - 2 = (9 - 2) = 7$ และ P -value เท่ากับ 0.002 , 0.020 , และ 0.116 ตามลำดับจึงสรุปได้ว่า ตัวแปร Y กับ X_1 มีความสัมพันธ์กันเมื่อมีอิทธิพลของ X_2 เข้ามาเกี่ยวข้องด้วย ($P = 0.002 < 0.05$) และตัวแปร Y กับ X_2 มีความสัมพันธ์กันเมื่อมีอิทธิพลของ X_1 เข้ามาเกี่ยวข้องด้วย ($P = 0.020 < 0.05$) สำหรับตัวแปร X_1 กับ X_2 ไม่มีความสัมพันธ์กันเมื่อมีอิทธิพลของ Y เข้ามาเกี่ยวข้องด้วย ($P = 0.116 > 0.05$)

จากภาพผลลัพธ์ที่ได้ค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนระหว่างตัวแปร Y และ X_1 เมื่อควบคุมตัวแปร X_2 คือ $ry_{1.2}$ เท่ากับ 0.8323

การทดสอบสมมติฐาน $H_0 : \rho_{y1.2} = 0$ คู่กับ $H_1 : \rho_{y1.2} \neq 0$ สถิติทดสอบ t ที่มีจำนวนชั้นอิสระ $df = (n - k - 1) = (9 - 2 - 1) = 6$ และ P - value เท่ากับ 0.010 จึงตัดสินใจปฏิเสธ H_0 สรุปว่ามีความสัมพันธ์อย่างมีนัยสำคัญระหว่าง Y กับ X_1 เมื่อให้ X_2 คงที่ การทดสอบสัมประสิทธิ์สหสัมพันธ์บางส่วนตัวอื่น ๆ ก็ทำเช่นเดียวกัน

8. วิธีการเลือกตัวแปรอิสระเข้าสมการถดถอย

ในขั้นตอนการวิเคราะห์การถดถอยที่อธิบายไว้แล้วในบทที่ 11 เริ่มจากการตรวจสอบว่าตัวแบบเส้นตรงเป็นตัวแทนที่เหมาะสมและเป็นไปตามข้อตกลงเบื้องต้นเกี่ยวกับความเป็นปกติ แล้วจึงคำนวณหาสมการถดถอยเชิงเส้นตรงที่เหมาะสมที่สุดกับข้อมูล ในขั้นตอนนี้เป็นการนำตัวแปรอิสระ X_i เข้าสมการถดถอย

การสร้างตัวแทนการถดถอยที่เป็นไปได้ทั้งหมดเมื่อมีตัวแปรอิสระหรือตัวแปรทำนายทั้งหมด k ตัว สามารถสร้างตัวแทนที่เป็นไปได้ทั้งหมด $2^k - 1$ ตัวแบบ มีวิธีการหลายวิธีที่ใช้สร้างตัวแทนเหล่านี้ วิธีการที่ใช้ต่อไปนี้สามารถรวบรวมสารสนเทศที่สำคัญ ๆ ทั้งหมดในข้อมูลเพื่อจะเลือกให้ได้ตัวแทนที่ดีที่สุด ซึ่งวิธีการเลือกตัวแปรอิสระ X_i เข้าสมการมี 5 วิธี คือ (1) Enter (2) Remove (3) Forward (4) Backward (5) Stepwise

8.1 วิธี Enter

กำหนดให้สมการถดถอยประกอบด้วยตัวแปรอิสระทั้งหมด โดยปกติโปรแกรม SPSS กำหนดให้นำตัวแปรอิสระทุกตัวเข้าในสมการถดถอยอยู่แล้วโดยวิธี Enter ผู้วิเคราะห์จะต้องตัดสินใจเองว่าตัวแปรอิสระ X_i ตัวใดบ้างที่ควรอยู่ในสมการถดถอยโดยพิจารณาจากค่า Sig. ของสถิติทดสอบ F ของสมการที่ประกอบด้วยตัวแปรอิสระ X_i ทั้งหมดและค่า Sig. ของสถิติทดสอบ t ของตัวแปรอิสระ X_i แต่ละตัวว่ามีนัยสำคัญทาง

สถิติหรือไม่ (ค่า Sig. น้อยกว่าระดับนัยสำคัญที่กำหนด $\alpha = .05$) นอกจากนี้ยังต้องพิจารณาข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยเกี่ยวกับความเป็นอิสระของตัวแปรอิสระด้วย แล้วทำการเปลี่ยนแปลงตัวแบบ ดังอธิบายไว้ในหัวข้อที่ 5

8.2 วิธี Remove

เป็นการเลือกตัวแปรอิสระที่กำหนดออกจากสมการถดถอยภายในขั้นตอนเดียว วิธีนี้จะใช้คู่กับวิธี Enter โปรแกรม SPSS ไม่นอนุญาตให้เลือกใช้วิธีนี้เป็นวิธีแรกในการวิเคราะห์การถดถอย

8.3 วิธี Forward

เป็นการคัดเลือกตัวแปรอิสระ X_i เพิ่มเข้าไปอยู่ในสมการถดถอยครั้งละ 1 ตัวแปร ในที่นี้สมมติว่ามีตัวแปรอิสระทั้งหมด 3 ตัวคือ X_1, X_2, X_3 โดยมีขั้นตอนการพัฒนาตัวแบบตามวิธีนี้คือ

ขั้นที่ 1) เลือกตัวแปรอิสระ X_i ตัวแรกที่มีความสัมพันธ์กับตัวแปรตาม Y มากที่สุด กว่าตัวแปรอิสระ X_i ตัวอื่น ๆ โดยพิจารณาจากค่า r_{YX_i} แล้วคำนวณหาสมการถดถอยเชิงเส้นของตัวแปรตาม Y และตัวแปรอิสระ X_i ตัวนั้น ถ้าค่าสถิติทดสอบ F ในตารางวิเคราะห์ความแปรปรวนไม่มีนัยสำคัญเราจะหยุดและสรุปว่าไม่มีตัวแปรอิสระตัวใดเลยที่เป็นตัวทำนายที่สำคัญ แต่ถ้าค่าสถิติทดสอบ F มีนัยสำคัญ เราจะเก็บตัวแปรอิสระ X_i ตัวนั้นเข้าไปในตัวแบบ แล้วทำขั้นที่ 2 ต่อไป ในที่นี้สมมติว่าได้ X_1 อยู่ในตัวแบบ

ขั้นที่ 2) คำนวณค่าสถิติ F บางส่วนของตัวแปรอิสระ X_i ที่เหลือแต่ละตัวเมื่อในตัวแบบมีตัวแปรอิสระ X_1 ตัวแรกที่ถูกเลือกในขั้นที่ 1 ให้อยู่ในตัวแบบแล้ว ในกรณีตัวอย่างนี้คือ คำนวณค่าของ partial $F (X_2|X_1)$ และ partial $F (X_3|X_1)$

ขั้นที่ 3) คัดเลือกตัวแปรอิสระ X_i ที่มีค่าสถิติ F บางส่วนมากที่สุดกว่าตัวอื่น ๆ สมมติว่าตัวแปรอิสระ X_2 มีค่าสถิติ F บางส่วนมากที่สุดกว่าตัวอื่น ๆ

ขั้นที่ 4) ทดสอบนัยสำคัญของค่าสถิติ F บางส่วนของตัวแปรอิสระที่คัดเลือกจากขั้นที่ 3 คือ คัดเลือกได้ X_2 ให้อยู่ในตัวแบบ ถ้าผลการทดสอบพบว่าไม่มีนัยสำคัญก็ตัดสินใจเพิ่มตัวแปรอิสระ X_2 เข้าไปในตัวแบบ ถ้าผลการทดสอบพบว่าไม่มีนัยสำคัญก็หยุดและใช้ตัวแบบที่ได้จากขั้นที่ 1

ขั้นที่ 5) ถ้าในขั้นที่ 4 ได้ผลการทดสอบพบว่ามีความสำคัญก็ทำต่อโดยทำซ้ำขั้นตอนที่ 2, 3 และ 4 คือ คำนวณค่าสถิติ F บางส่วนของตัวแปรอิสระ X_i ที่เหลือแต่ละตัวเมื่อในตัวแบบมีตัวแปรอิสระที่ถูกเลือกจากขั้นตอนก่อนหน้านี้แล้ว พิจารณาตัวแปรอิสระ X_i ที่มีค่าสถิติ F บางส่วนมากที่สุดกว่าตัวอื่น ๆ และทดสอบนัยสำคัญ ถ้าพบว่ามีความสำคัญก็เพิ่มตัวแปรอิสระตัวนั้นเข้าไปในตัวแบบ แต่ถ้าเมื่อใดพบว่าค่าสถิติ F บางส่วนไม่มีความสำคัญก็จะไม่มีการเพิ่มตัวแปรอิสระตัวใด ๆ เข้าไปในตัวแบบอีกและหยุดการพัฒนาตัวแบบ ยกตัวอย่างเช่น ต้องการพิจารณาว่าควรเพิ่มตัวแปรอิสระ X_3 เข้าไปในตัวแบบหรือไม่ พิจารณาจากค่าสถิติทดสอบ partial $F (X_3|X_1, X_2)$ ถ้าพบว่าไม่มีความสำคัญก็จะไม่เพิ่มตัวแปรอิสระ X_3 เข้าไปในตัวแบบอีก และหยุดการพัฒนาตัวแบบการถดถอย

ซึ่งขั้นตอนทั้งหลายเหล่านี้ในโปรแกรม SPSS จะทำให้ถ้าเลือกวิธี Forward วิธีนี้ผู้วิเคราะห์เป็นผู้กำหนดเกณฑ์ในการเลือกตัวแปรอิสระ X_i เข้าในตัวแบบเองโดยกำหนดค่าระดับนัยสำคัญของการทดสอบสมมติฐานเพื่อคัดเลือกตัวแปรอิสระ X_i ตัวหนึ่ง ๆ เข้าไปอยู่ในสมการถดถอย คือ ถ้าค่า Sig. ของค่าสถิติทดสอบ F น้อยกว่าค่าระดับนัยสำคัญที่ผู้วิเคราะห์กำหนดให้นี้จะคัดเลือกตัวแปรอิสระ X_i ตัวนั้น ๆ เข้าสมการถดถอย ทำได้โดยการเลือกคำสั่ง options ในหน้าต่างของคำสั่ง Linear Regression แล้วเลือก Use probability of F สำหรับ Entry : แล้วกำหนดค่าระดับนัยสำคัญเอง โดยปกติโปรแกรมจะกำหนดให้เท่ากับ 0.05 หรืออาจเลือก Use F Value สำหรับ Enter : แล้วกำหนดค่า F ที่ใช้เป็นเกณฑ์ในการคัดเลือกตัวแปรอิสระ X_i ตัวหนึ่ง ๆ เข้าไปอยู่ในสมการถดถอยเอง โดยปกติโปรแกรมจะกำหนดให้เท่ากับ 3.84 ซึ่งสอดคล้องกับค่าระดับนัยสำคัญเท่ากับ 0.05

8.4 วิธี Backward

เป็นการกำหนดให้สมการถดถอยประกอบด้วยตัวแปรอิสระ X_i ทั้งหมดก่อนแล้วจึงคัดเลือกตัวแปรอิสระ X_i ที่ไม่มีความสัมพันธ์กับตัวแปรตาม Y ออกจากตัวแบบทีละตัวโดยมีขั้นตอนดังนี้คือ

- ขั้นที่ 1) กำหนดให้สมการถดถอยประกอบด้วยตัวแปรอิสระทั้งหมดสมมติว่ามี X_1, X_2, X_3
- ขั้นที่ 2) คำนวณค่าสถิติ F บางส่วนของตัวแปรอิสระ X_i ทุกตัวในตัวแบบ คือ $\text{partial } F (X_1|X_2, X_3)$, $\text{partial } F (X_2|X_1, X_3)$ และ $\text{partial } F (X_3|X_1, X_2)$
- ขั้นที่ 3) พิจารณาตัวแปรอิสระ X_i ตัวที่มีค่าสถิติ F บางส่วนต่ำกว่าตัวอื่น ๆ สมมติให้เป็น F_L
- ขั้นที่ 4) ทดสอบนัยสำคัญของค่าสถิติ F โดยการทดสอบสมมติฐาน $H_0 : \beta_i = 0$ คู่กับ $H_1 : \beta_i \neq 0$ ถ้าผลการทดสอบพบว่า F_L มีนัยสำคัญก็ใช้สมการถดถอยนี้ได้เป็นสมการที่สมบูรณ์ แต่ถ้าพบว่า F_L ไม่มีนัยสำคัญก็จะตัดตัวแปรอิสระตัวนั้นออกจากตัวแบบแล้วคำนวณสมการถดถอยสำหรับตัวแปรที่เหลือใหม่อีกครั้ง และทำซ้ำขั้นตอนที่ 2, 3, และ 4 จนกระทั่งไม่สามารถตัดตัวแปรอิสระตัวใดได้อีกแล้วจึงหยุดการพัฒนาตัวแบบ
- ซึ่งขั้นตอนทั้งหลายเหล่านี้ในโปรแกรม SPSS จะทำให้ถ้าเลือกวิธี Backward วิธีนี้ผู้วิเคราะห์เป็นผู้กำหนดเกณฑ์ในการคัดตัวแปรอิสระ X_i ออกจากตัวแบบเองโดยกำหนดค่าระดับนัยสำคัญสูงสุดของการที่ตัวแปรอิสระ X_i ตัวหนึ่ง ๆ สามารถยังคงอยู่ในสมการถดถอย ทำได้โดยการเลือกคำสั่ง options ในหน้าต่างของคำสั่ง Linear Regression แล้วเลือก Use probability of F สำหรับ Removal : แล้วกำหนดค่าระดับนัยสำคัญสูงสุดเอง โดยปกติโปรแกรมจะกำหนดให้เท่ากับ 0.10 หรืออาจเลือก Use F Value สำหรับ Removal : แล้วกำหนดค่า F ที่ใช้เป็นเกณฑ์ในการคัดตัวแปรอิสระ X_i ตัวหนึ่ง ๆ ออกจากสมการถดถอยเอง โดยปกติโปรแกรมจะกำหนดให้เท่ากับ 2.71 ซึ่งสอดคล้องกับค่าระดับนัยสำคัญเท่ากับ 0.10

8.5 วิธี Stepwise

เป็นวิธีที่นิยมใช้กันมากที่สุด เป็นวิธีที่พัฒนามาจากวิธี Forward แต่ไม่เหมือนกัน คือ ตัวแปรอิสระ X_i ที่ถูกคัดเลือกให้เข้าไปอยู่ในสมการถดถอยแล้ว อาจถูกคัดออกได้ภายหลัง ถ้าพบว่ามีความสัมพันธ์กับตัวแปรอิสระ X_i ตัวอื่น ๆ ในปัจจุบันที่อยู่ในสมการถดถอย มีขั้นตอนดังนี้คือ

ขั้นที่ 1) เลือกตัวแปรอิสระ X_i ตัวที่มีความสัมพันธ์มากที่สุดกว่าตัวอื่น ๆ กับตัวแปรตามและค่าสถิติ F มีนัยสำคัญ

ขั้นที่ 2) เหมือนกับวิธี Forward คือเพิ่มตัวแปรอิสระ X_i เข้าไปในตัวแบบโดยพิจารณาจากค่าสถิติ F บางส่วนมีค่ามากที่สุดกว่าตัวแปรอิสระตัวอื่น ๆ และมีนัยสำคัญ

ขั้นที่ 3) พิจารณาค่าสถิติ F บางส่วนของตัวแปรอิสระ X_i ในขั้นที่ 1 สมมติว่าเป็น X_1 เมื่อเพิ่มตัวแปรอิสระในขั้นที่ 2 สมมติว่าเป็น X_2 เพื่อพิจารณาว่าจะตัดตัวแปร X_1 ออกจากสมการถดถอยหรือไม่ นั่นคือพิจารณา $\text{partial } F (X_1|X_2)$ ว่ามีนัยสำคัญหรือไม่ สมมติว่า $\text{partial } F (X_1|X_2)$ เท่ากับ 7.67 ซึ่งมากกว่าค่าวิกฤติคือ $F_{0.90; 1,9} = 3.36$ แสดงว่ามีนัยสำคัญ ดังนั้นเราจึงไม่ตัดตัวแปรอิสระ X_1 ออกจากสมการถดถอย

ขั้นที่ 4) พิจารณาว่าควรเพิ่มตัวแปรอิสระ X_i ตัวอื่น ๆ เข้าไปในสมการถดถอยอีก 1 ตัวหรือไม่ พิจารณาเช่นเดียวกับวิธี Forward ยกตัวอย่างเช่นต้องการพิจารณาว่าควรเพิ่มตัวแปรอิสระ X_3 เข้าไปในสมการหรือไม่ พิจารณาจากค่าสถิติทดสอบ $\text{partial } F (X_3|X_1, X_2)$ ถ้าพบว่าไม่มีนัยสำคัญก็จะไม่เพิ่มตัวแปรอิสระ X_3 เข้าไปในสมการถดถอยอีก และหยุดการพัฒนาตัวแบบการถดถอย

ซึ่งขั้นตอนทั้งหลายเหล่านี้ในโปรแกรม SPSS จะทำให้ถ้าเลือกวิธี Stepwise วิธีนี้ผู้วิเคราะห์จะเป็นผู้กำหนดเกณฑ์ในการเลือกตัวแปรอิสระ X_i เข้าไปอยู่ในสมการถดถอย และเกณฑ์ในการตัดตัวแปรอิสระ X_i ออกจากสมการถดถอย วิธีกำหนดเกณฑ์ทำแบบเดียวกับการกำหนดเกณฑ์ในวิธี Forward และวิธี Backward ทั้ง 2 วิธีพร้อม ๆ กัน โดยผู้วิเคราะห์ควรกำหนดให้ค่า F สำหรับ Entry : มากกว่าค่า F สำหรับ Removal : หรือกำหนดให้ค่าระดับนัยสำคัญของ F สำหรับ Entry : น้อยกว่าค่าระดับนัยสำคัญของ F สำหรับ Removal : เพื่อป้องกันไม่ให้ตัวแปรอิสระ X_i ตัวเดิมเข้าและออกจากสมการ

9. ตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยที่ว่าตัวแปรอิสระ X_i ทุกตัวเป็นอิสระกัน

ถ้าตัวแปรอิสระ X_i ทั้ง k ตัวไม่เป็นอิสระกันจะทำให้เกิดปัญหา **Multicollinearity** กรณีที่พบว่ามีความสัมพันธ์ระหว่างตัวแปรอิสระ X_i เราจะแก้ไขด้วยการตัดตัวแปรอิสระ X_i ที่มีความสัมพันธ์กันสูงออกไป โดยเลือกตัดตัวแปรอิสระ X_i ที่มีความสำคัญน้อยหรืออาจแก้ไขด้วยการสร้างตัวแปรใหม่ขึ้นมาแทนตัวแปรอิสระที่มีความสัมพันธ์กันสูง โดยอาจใช้เทคนิคการวิเคราะห์ปัจจัย (Factor Analysis) หรือ **Principal component**

การทดสอบความเป็นอิสระของตัวแปรอิสระ X_i ทุกตัว หรือการทดสอบ **Collinearity** โดยเลือกใช้คำสั่งย่อย **Collinearity Diagnostics** ในหน้าต่างของคำสั่ง **Linear regression, Statistics** จะได้ค่าสถิติ **tolerance** ค่า **VIF** (**Variance Inflation Factor**) และค่า **Condition Index**

ค่า **tolerance** ของ X_i เท่ากับ $1 - R_i^2$ มีค่าอยู่ระหว่าง 0 ถึง 1 ค่า R_i^2 คือสัมประสิทธิ์การตัดสินใจซึ่งแสดงความสัมพันธ์ระหว่างตัวแปรอิสระ X_i กับตัวแปรอิสระตัวอื่น ๆ ถ้าค่า **tolerance** มีค่าเข้าใกล้ศูนย์แสดงว่าตัวแปรอิสระ X_i มีความสัมพันธ์กับตัวแปรอิสระอื่น ๆ มาก นั่นคือ เกิด **multicollinearity** นั่นคือถ้าค่า **tolerance** ยิ่งมีค่าเข้าใกล้ 1 แสดงว่าตัวแปรอิสระ X_i ตัวนั้นมีความสัมพันธ์กับตัวแปรอื่น ๆ น้อย

ค่า **VIF** ของ X_i เท่ากับ $1/\text{tolerance}$ มีค่าอยู่ระหว่าง 1 ถึง ∞ ถ้าค่า **VIF** มีค่ามากแสดงว่าตัวแปรอิสระ X_i มีความสัมพันธ์กับตัวแปรอิสระตัวอื่น ๆ มาก

ค่า **Condition Index** เท่ากับ $\sqrt{\frac{\text{Eigenvalue max}}{\text{Eigenvalue}}}$ ถ้าค่า **Condition Index** มีค่ามากแสดงว่าตัวแปรอิสระ X_i ตัวนั้นมีความสัมพันธ์กับตัวแปรอิสระตัวอื่น ๆ มาก

10. ตรวจสอบข้อตกลงเบื้องต้นเกี่ยวกับความคลาดเคลื่อนโดยการวิเคราะห์เศษตกค้าง (Residual Analysis)

ในการวิเคราะห์การถดถอยพหุที่มีตัวแปรอิสระ k ตัว คือ X_1, X_2, \dots, X_k และตัวแปรตาม 1 ตัว คือ Y ถ้าเก็บข้อมูลเป็นค่าสังเกตจากกลุ่มตัวอย่างขนาด n จะได้ค่าสังเกตแต่ละตัวเขียนแทนได้เป็น $(y_j, x_{1j}, x_{2j}, \dots, x_{kj}) ; j = 1, 2, \dots, n$ สามารถเขียน

เป็นตัวแทนการถดถอยพหุเชิงเส้นแสดงความสัมพันธ์ระหว่างตัวแปรอิสระหลายตัวกับตัวแปรตาม 1 ตัว คือ

$$y_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \dots + \beta_k X_{kj} + e_j$$

$$; j = 1, 2, \dots, n$$

เมื่อ e_j คือ ความคลาดเคลื่อนของค่าสังเกตตัวที่ j

การประมาณสมการถดถอยพหุด้วยวิธีกำลังสองน้อยที่สุดจะได้ตัวแทนซึ่งผลบวกกำลังสองของระยะห่างระหว่างค่าสังเกตและค่าทำนายมีค่าน้อยที่สุด นั่นคือ ถ้าให้ตัวแทนที่เหมาะสมกับข้อมูลคือ

$$y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_k X_k$$

และค่าทำนายของตัวแปรตามของข้อมูลตัวที่ j คือ

$$\hat{y}_j = \hat{\beta}_0 + \hat{\beta}_1 X_{1j} + \hat{\beta}_2 X_{2j} + \dots + \hat{\beta}_k X_{kj}$$

ทำให้คำนวณค่าประมาณของ e_j ได้คือ ความแตกต่างระหว่างค่าสังเกต y_j กับค่าทำนาย \hat{y}_j เรียกว่าเศษตกค้าง (residual) เขียนแทนด้วย e_j คือ

$$e_j = y_j - \hat{y}_j$$

จะได้เซทของ e_j ซึ่งสะท้อนปริมาณความแตกต่างระหว่างค่าสังเกตกับค่าทำนาย โดยที่ e_j แต่ละตัวคือค่าประมาณของความคลาดเคลื่อน e_j ซึ่งในการวิเคราะห์ความแปรปรวนมีข้อตกลงเบื้องต้นเกี่ยวกับความคลาดเคลื่อนคือ ความคลาดเคลื่อน e_j แต่ละตัวมีความเป็นอิสระกัน และมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับศูนย์ และมีความแปรปรวนคงที่ทุกค่าของ X เท่ากับ σ^2

เนื่องจากการคำนวณค่าสัมประสิทธิ์การถดถอย b_0, b_1, \dots, b_k ในสมการถดถอยเชิงเส้นนั้นใช้วิธีการกำลังสองน้อยที่สุด (Least Square Error) คือทำให้ผลบวกกำลังสองของความคลาดเคลื่อนแต่ละตัวมีค่าน้อยที่สุด ซึ่งมีผลทำให้ผลรวมของความ

คลาดเคลื่อนแต่ละตัว (e_i) เท่ากับ 0 ดังนั้นจึงทำให้ค่าคาดหวังของความคลาดเคลื่อนหรือค่าเฉลี่ยของความคลาดเคลื่อนเท่ากับ 0 ด้วย เราจึงไม่จำเป็นต้องตรวจสอบ

สำหรับการตรวจสอบการแจกแจงของความคลาดเคลื่อน สามารถใช้คำสั่งย่อย Plot ในหน้าต่างของคำสั่ง Linear Regression ในส่วนของคำสั่ง Standardized Residual Plots เพื่อพล็อตกราฟฮิสโตแกรมของค่า Standardized Residuals และพล็อตกราฟของ Normal probability plot

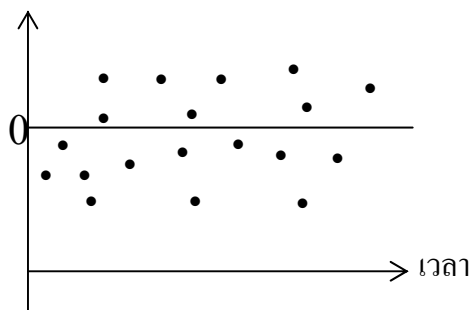
10.1 การตรวจสอบความเป็นอิสระกัน

สำหรับการตรวจสอบความเป็นอิสระกันของความคลาดเคลื่อนสามารถทำได้ 2 วิธีคือ

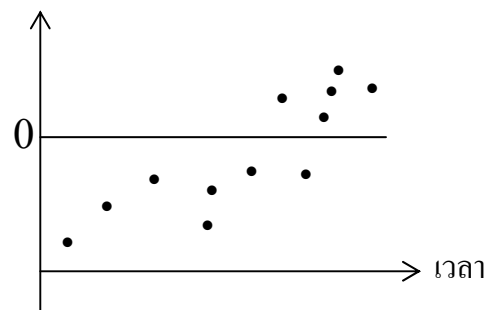
(1) ใช้วิธีการทดสอบทางสถิติ ค่าสถิติที่ใช้ทดสอบความเป็นอิสระกันของความคลาดเคลื่อนคือ ค่าสถิติ Durbin-Watson โดยใช้คำสั่ง Durbin Watson ในหน้าต่างของคำสั่ง Linear Regression, Statistics โดยพิจารณาว่าค่าสถิติ Durbin Watson มีค่าแตกต่างไปจาก 2 มากหรือไม่ ถ้าค่าเข้าใกล้ 2 แสดงว่าความคลาดเคลื่อนแต่ละค่าเป็นอิสระกัน ถ้าค่าน้อยกว่า 1.5 ระวังใกล้ 0 แสดงว่าความคลาดเคลื่อนแต่ละตัวยังมีความสัมพันธ์กันอย่างมากในเชิงบวก ถ้าค่ามากกว่า 2.5 แสดงว่าความคลาดเคลื่อนแต่ละตัวมีความสัมพันธ์ในเชิงลบ ถ้ายังเข้าใกล้ 4 แสดงว่ายังมีความสัมพันธ์กันอย่างมากในเชิงลบ

(2) ใช้วิธีการพล็อตกราฟของความคลาดเคลื่อนกับเวลา ถ้าพบว่าไม่มีรูปแบบที่แน่นอนคือ การกระจายของจุดไม่เป็นระบบ ดังภาพที่ 11.3 (ก) แสดงว่าไม่มีปัญหาที่เรียกว่า Autocorrelation หรือ Serial correlation แต่ถ้าพบว่าการกระจายของจุดมีรูปแบบดังภาพที่ 11.3 (ข) ก็แสดงว่าความคลาดเคลื่อนแต่ละตัวมีความสัมพันธ์กัน

ความคลาดเคลื่อน (e)



ความคลาดเคลื่อน (e)



(ก) ความคลาดเคลื่อนแต่ละตัวเป็นอิสระกัน

(ข) ความคลาดเคลื่อนแต่ละตัวมีความสัมพันธ์กันเชิงบวก

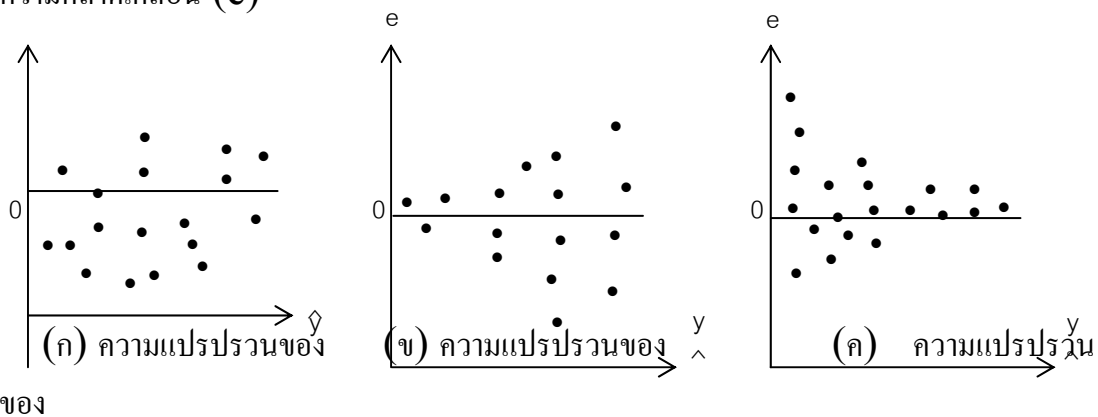
ภาพที่ 11.3

10.2 การตรวจสอบเกี่ยวกับความแปรปรวน

การตรวจสอบเกี่ยวกับความแปรปรวนของความคลาดเคลื่อนโดยวิธีการพล็อตกราฟระหว่างความคลาดเคลื่อน (e_i) กับค่าทำนายของตัวแปรตาม (y_i) โดยใช้คำสั่ง plot ในหน้าต่างของคำสั่ง Linear Regression ซึ่งในโปรแกรม SPSS จะเป็นการพล็อตระหว่าง ZRESID กับ ZPRED

ถ้าพบว่าค่า e_i กระจายอยู่รอบ ๆ ค่า 0 ไม่ว่าค่า y จะเปลี่ยนแปลงไปก็ตาม ดังภาพที่ 11.4 (ก) แสดงว่าความแปรปรวนของความคลาดเคลื่อนคงที่ แต่ถ้าพบว่าค่า e_i มีการกระจายอย่างมีรูปแบบ ดังภาพที่ 11.4 (ข) หรือภาพที่ 11.4 (ค) แสดงว่าความแปรปรวนของความคลาดเคลื่อนไม่คงที่ในทุกค่าของ X นั่นคือเกิดปัญหา Heteroscedascity การแก้ปัญหาค่าความแปรปรวนของความคลาดเคลื่อนไม่คงที่อาจใช้วิธีการแปลงข้อมูลเช่นแปลงค่าของข้อมูลให้อยู่ในรูปของ \log เป็นต้น

ความคลาดเคลื่อน (e)



ความคลาดเคลื่อนครั้งที่
ความคลาดเคลื่อน

ความคลาดเคลื่อน

ไม่คงที่

ไม่คงที่

ภาพที่ 11.4

11. การแปลงข้อมูล (Transformation)

ก่อนการวิเคราะห์การถดถอย จากข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยเกี่ยวกับตัวแปรตาม Y ที่ว่าตัวแปรตาม Y มีการแจกแจงแบบปกติสำหรับทุกจุดของตัวแปรอิสระ X_i ทุกตัว ถ้ามีการตรวจสอบการแจกแจงของตัวแปรตาม Y ก่อนการวิเคราะห์การถดถอยแล้วพบว่าตัวแปรตาม Y ที่ได้จากกลุ่มตัวอย่างไม่เป็นไปตามข้อตกลงเบื้องต้นนี้ก็สามารถแก้ไขได้โดยการแปลงข้อมูลของตัวแปรตาม Y ยกตัวอย่างเช่น

(1) ถ้าการแจกแจงของตัวแปรตาม Y เป็นแบบเบ้ขวา จึงแก้ไขโดยการแปลงข้อมูลของตัวแปรตาม Y เพื่อยังคงทำให้ความสัมพันธ์ระหว่างตัวแปรอิสระ X_i และ Y อยู่ในรูปเชิงเส้นทำได้ดังนี้คือ

$$Y' = \log(Y) \quad \text{เมื่อ } Y > 0$$

เมื่อ Y' คือ ตัวแปร Y ที่แปลงแล้ว

(2) ถ้าการแจกแจงของตัวแปรตาม Y เป็นแบบเบ้ซ้าย จึงแก้ไขโดยการแปลงข้อมูลของตัวแปรตาม Y เพื่อยังคงทำให้ความสัมพันธ์ระหว่างตัวแปรอิสระ X_i และ Y อยู่ในรูปเชิงเส้นทำได้ดังนี้คือ

$$Y' = Y^2$$

การแปลงข้อมูลของตัวแปรตาม Y อาจพิจารณาได้จากแผนภาพการกระจายของข้อมูลระหว่างตัวแปรตาม Y บนแกน Y และตัวแปรอิสระ X บนแกน X ถ้าความสัมพันธ์ไม่เป็นรูปเชิงเส้นตรง ก็อาจแปลงข้อมูลของตัวแปรตาม Y หรือตัวแปรอิสระ X หรืออาจ

แปลงข้อมูลทั้งตัวแปรตาม Y และตัวแปรอิสระ X พร้อมกันก็ได้ เพื่อทำให้ความสัมพันธ์ระหว่างตัวแปรอิสระ X_i และตัวแปรตาม Y อยู่ในรูปเชิงเส้นตรง นอกจากนี้การแปลงข้อมูลของตัวแปรตาม Y ยังสามารถพิจารณาจากค่าของความคลาดเคลื่อนและความแปรปรวนของความคลาดเคลื่อนได้ด้วย รายละเอียดเรื่องนี้สามารถดูได้จากตำราอื่น ๆ

12. ตัวอย่างการพัฒนาตัวแบบการถดถอย

ตัวอย่างเช่น การศึกษาเรื่องการผ่าตัดตัดตับ สุ่มตัวอย่างคนไข้มา 54 คน จากบันทึกคนไข้แต่ละคนมีข้อมูลการประเมินก่อนการผ่าตัดเป็นตัวแปรอิสระดังนี้

X_1 คือ blood clotting score

X_2 คือ prognostic index, which includes the age of patient

X_3 คือ enzyme function test score

X_4 คือ liver function test score

ตัวแปรตาม คือ เวลารอดชีวิต (survival time) ข้อมูลเกี่ยวกับตัวแปรอิสระและตัวแปรตามแสดงดังตาราง และเนื่องจากการแจกแจงของตัวแปรตามเป็นแบบเบ้ขวา จึงทำการแปลงข้อมูลเวลารอดชีวิต $Y' = \log_{10} Y$

ตารางที่ 11.3 ข้อมูลเกี่ยวกับตัวแปรอิสระและตัวแปรตามของตัวอย่างการผ่าตัดตัดตับ

Case Number	Blood Clotting Score X_1	Prognostic Index X_2	Enzyme Function Test X_3	Liver Function Test X_4	Survival Time Y	$Y' = \log_{10} Y$
1	6.7	62	81	2.59	200	2.3010
2	5.1	59	66	1.70	101	2.0043
3	7.4	57	83	2.16	204	2.3096
4	6.5	73	41	2.01	101	2.0043
5	7.8	65	115	4.30	509	2.7067
6	5.8	38	72	1.42	80	1.9031
7	5.7	46	63	1.91	80	1.9031
8	3.7	68	81	2.57	127	2.1038
9	6.0	67	93	2.50	202	2.3054
10	3.7	76	94	2.40	203	2.3075
11	6.3	84	83	4.13	329	2.5172
12	6.7	51	43	1.86	65	1.8129
13	5.8	96	114	3.95	830	2.9191
14	5.8	83	88	3.95	330	2.5185

Case Number	Blood Clotting Score X_1	Prognostic Index X_2	Enzyme Function Test X_3	Liver Function Test X_4	Survival Time Y	$Y' = \log_{10} Y$
15	7.7	62	67	3.40	168	2.2253
16	7.4	74	68	2.40	217	2.3365
17	6.0	85	28	2.98	87	1.9395
18	3.7	51	41	1.55	34	1.5315
19	7.3	68	74	3.56	215	2.3324
20	5.6	57	87	3.02	172	2.2355
21	5.2	52	76	2.85	109	2.0374
22	3.4	83	53	1.12	136	2.1335
23	6.7	26	68	2.10	70	1.8451
24	5.8	67	86	3.40	220	2.3424
25	6.3	59	100	2.95	276	2.4409
26	5.8	61	73	3.50	144	2.1584
27	5.2	52	86	2.45	181	2.2577
28	11.2	76	90	5.59	574	2.7589
29	5.2	54	56	2.71	72	1.8573
30	5.8	76	59	2.58	178	2.2504
31	3.2	64	65	0.74	71	1.8513
32	8.7	45	23	2.52	58	1.7634
33	5.0	59	73	3.50	116	2.0645
34	5.8	72	93	3.30	295	2.4698
35	5.4	58	70	2.64	115	2.0607
36	5.3	51	99	2.60	184	2.2648
37	2.6	74	86	2.05	118	2.0719
38	4.3	8	119	2.85	120	2.0792
39	4.8	61	76	2.45	151	2.1790
40	5.4	52	88	1.81	148	2.1703
41	5.2	49	72	1.84	95	1.9777
42	3.6	28	99	1.30	75	1.8751
43	8.8	86	88	6.40	483	2.6840
44	6.5	56	77	2.85	153	2.1847
45	3.4	77	93	1.48	191	2.2810
46	6.5	40	84	3.00	123	2.0899
47	4.5	73	106	3.05	311	2.4928
48	4.8	86	101	4.10	398	2.5999
49	5.1	67	77	2.86	158	2.1987

Case Number	Blood Clotting Score X_1	Prognostic Index X_2	Enzyme Function Test X_3	Liver Function Test X_4	Survival Time Y	$Y' = \log_{10} Y$
50	3.9	82	103	4.55	310	2.4914
51	6.6	77	46	1.95	124	2.0934
52	6.4	85	40	1.21	125	2.0969
53	6.4	59	85	2.33	198	2.2967
54	8.8	78	72	3.20	313	2.4955

แหล่งที่มา : Neter, J., Wasserman, W., Kutner, H.M., 1974.

12.1 ขั้นตอนการพัฒนาตัวแบบการถดถอย

(1) ตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์ความแปรปรวนเกี่ยวกับการแจกแจงของตัวแปรตาม Y จากการพล็อตกราฟของข้อมูล y_i ทั้งหมดในกลุ่มตัวอย่าง โดยใช้คำสั่ง **Graphs, Histogram** แล้วเลือกตัวแปรตาม Y เข้าไปในช่อง **Variable** แล้วเลือกคำสั่ง **Display normal curve** เพื่อดูว่าข้อมูล y มีการแจกแจงเป็นปกติหรือไม่

(2) ตรวจสอบว่าตัวแปรอิสระ X_i มีความสัมพันธ์เชิงเส้นกับตัวแปรตาม Y หรือไม่ โดยใช้คำสั่ง **Graphs, Scatter ...** จะได้หน้าต่าง **Scatterplot** เลือก **Matrix** แล้ว **Define Matrix Variables**

(3) ตรวจสอบความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระ X_i เพื่อดูค่าสัมประสิทธิ์สหสัมพันธ์ โดยใช้คำสั่ง **Analyze, Correlate, Bivariate** แล้วเลือกตัวแปรทั้งหมดเข้าไปอยู่ในช่อง **Variables** : ในกรอบคำสั่ง **Correlation Coefficients** เลือก **Pearson** และ **Test of Significance** แบบ **Two-tailed**

(4) หาสมการถดถอยที่เหมาะสม ในที่นี้จะอธิบายเฉพาะวิธี **Stepwise** โดยใช้คำสั่ง **Analyze, Regression, Linear ...** จะได้หน้าต่าง **Linear Regression** เลือกตัวแปรตาม Y เข้าไปในช่อง **Dependent** เลือกตัวแปรอิสระ X_i ทั้งหมดเข้าไปในช่อง **Independents** ที่คำสั่ง **Method** : เลือก **Stepwise** ในหน้าต่างของคำสั่ง **Linear Regression** : **Statistics** ในส่วนของคำสั่ง **Regression Coefficients**

- เลือกคำสั่งย่อย **Estimates** และ **Confidence intervals** เพื่อประมาณค่าสัมประสิทธิ์การถดถอยและช่วงความเชื่อมั่นของสัมประสิทธิ์การถดถอย แล้วทดสอบ **t-test** ในการทดสอบสมมติฐานเกี่ยวกับ $H_0 : \beta_i = 0$ คู่กับ $H_1 : \beta_i \neq 0$

- เลือกคำสั่งย่อย **Model fit** และ **R squared change** เพื่อวิเคราะห์ความแปรปรวนแบบทางเดียว (**Oneway ANOVA**) ในการทดสอบสมมติฐานเกี่ยวกับ $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$ คู่กับ $H_1 : \beta_i \neq 0$ อย่างน้อย 1 ค่า ใช้ทดสอบความเหมาะสมของตัวแบบการถดถอย

- เลือกคำสั่งย่อย **Part and partial correlations** เพื่อแสดงระดับความสัมพันธ์ระหว่างตัวแปรอิสระ X_i แต่ละตัวกับตัวแปรตาม Y เมื่อไม่ได้ควบคุมอิทธิพลของตัวแปรอิสระตัวอื่น ๆ เรียกว่า **Zero-order Correlations** ส่วนค่า **Partial Correlations** เป็นค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนของตัวแปรอิสระ X_i ตัวหนึ่งกับตัวแปรตาม Y เมื่อควบคุมอิทธิพลของตัวแปรอิสระตัวอื่น ๆ ให้คงที่

ในหน้าต่างของคำสั่ง **Linear Regression : Options** ในส่วนของคำสั่ง **Stepping Method Criteria**

- เลือกคำสั่งย่อย **Use probability of F** สำหรับ **Entry** : ใส่ระดับนัยสำคัญเท่ากับ **.05** และสำหรับ **Removal** : ใส่ระดับนัยสำคัญเท่ากับ **.10** ซึ่งโดยปกติโปรแกรมจะกำหนดค่านี้ให้อยู่แล้ว เพื่อกำหนดเกณฑ์ในการเลือกตัวแปรอิสระ X_i เข้าไปอยู่ในสมการถดถอย และเกณฑ์ในการคัดตัวแปรอิสระ X_i ออกจากสมการถดถอย

(5) ตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยเกี่ยวกับความเป็นอิสระของตัวแปรอิสระ X_i ทั้งหมด หลังจากได้สมการถดถอยมาแล้วในขั้นตอนก่อน ควรทำการตรวจสอบสมการถดถอยที่ได้นี้ว่าเป็นไปตามข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยจึงสามารถใช้สมการถดถอยที่ได้จากกลุ่มตัวอย่าง ในการทำนายค่าของตัวแปรตามมีความน่าเชื่อถือจากกรอบคำสั่ง **Linear Regression : Statistics** เลือกคำสั่ง **Collinearity diagnostics** เพื่อดูค่าสถิติ **tolerance**, **VIF**, และ **Condition Index**

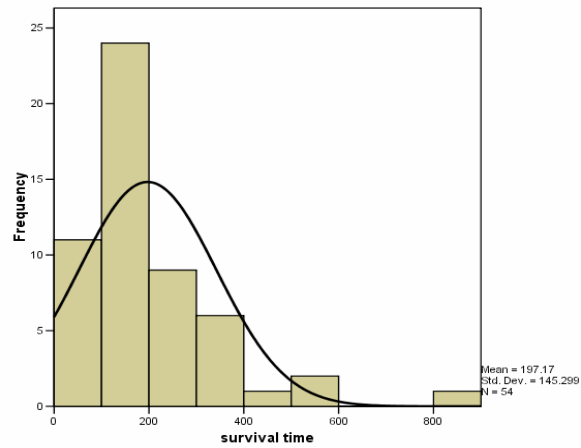
(6) ตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยเกี่ยวกับความคลาดเคลื่อน คือ ความคลาดเคลื่อน (e_j) แต่ละตัวมีความเป็นอิสระกันและมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับศูนย์ และมีความแปรปรวนคงที่ทุกค่าของ X เท่ากับ σ^2

- ตรวจสอบว่าค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติหรือไม่ จากกรอบคำสั่ง **Linear Regression : Plots** ในส่วนของคำสั่ง **Standardized Residuals Plots** โดยพล็อตกราฟ **Histogram** และ **Normal Probability Plot** ของคะแนนมาตรฐานของความคลาดเคลื่อน (**Standardized Residuals**)
- ตรวจสอบว่าค่าความคลาดเคลื่อนแต่ละค่าเป็นอิสระกันหรือไม่ จากกรอบคำสั่ง **Linear Regression : Statistics** ในส่วนของคำสั่ง **Residuals** เลือก **Durbin-Watson** เพื่อดูค่าสถิติ **Durbin-Watson** ซึ่งเรามักใช้กับข้อมูลที่เป็นอนุกรมเวลา
- ตรวจสอบว่าความแปรปรวนของค่าความคลาดเคลื่อนคงที่หรือไม่ จากกรอบคำสั่ง **Linear Regression : Plots** เพื่อพล็อตกราฟระหว่างค่าคะแนนมาตรฐานของค่าพยากรณ์ (**Standardized Predicted Value** หรือ **ZPRED**) บนแกน **X** กับคะแนนมาตรฐานของค่าเศษตกค้าง (**Standardized Residuals** หรือ **ZRESID**) บนแกน **Y**

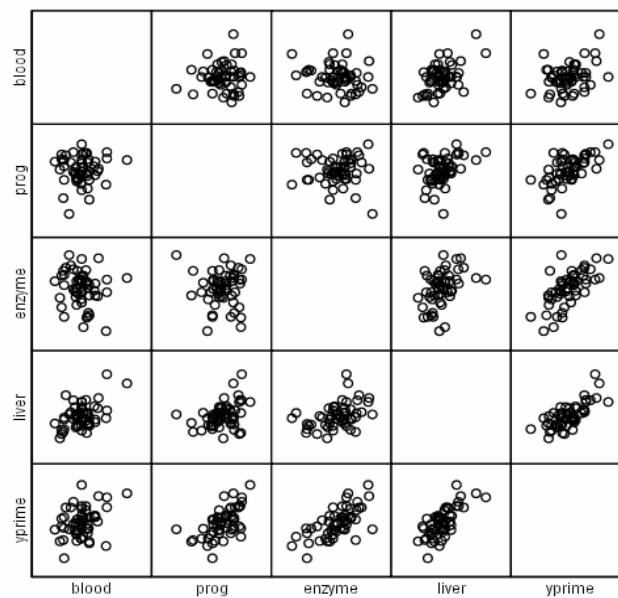
12.2 ผลลัพธ์การพัฒนาตัวแบบการถดถอย

จากตัวอย่างการศึกษาเรื่องการผ่าตัดตัดตับ มีวัตถุประสงค์เพื่อหารูปแบบความสัมพันธ์ของตัวแปรอิสระ **blood, prog, enzyme, liver** กับตัวแปรตามเวลารอดชีวิต (**Survival**) เพื่อนำไปพยากรณ์เวลารอดชีวิตของคนไข้ต่อไป ดำเนินการพัฒนาตัวแบบการถดถอยตามขั้นตอนข้างต้นได้ผลลัพธ์ดังนี้

(1) กราฟฮีสโตแกรมของตัวแปรตาม **survival** แสดงการแจกแจงของตัวแปรตาม **survival** พบว่าตัวแปรตาม **survival** มีการแจกแจงไม่เป็นปกติคือเป็นแบบเบ้ขวา ดังภาพที่ 11.5 ดังนั้นจึงทำการแปลงข้อมูล **survival** ให้มีรูปแบบใหม่คือ $y_{\text{prime}} = \log_{10} \text{survival}$



(2) กราฟแสดงความสัมพันธ์ระหว่างตัวแปรต่าง ๆ ระหว่างตัวแปรใด ๆ 2 ตัวแปร ดังภาพที่ 11.6 เพื่อดูว่าความสัมพันธ์ระหว่างตัวแปรตาม (yprime) กับตัวแปรอิสระ blood, prog, enzyme, และ liver ระหว่างตัวแปรใด ๆ 2 ตัวแปร มีลักษณะเป็นเชิงเส้นหรือไม่



ภาพที่ 11.6

(3) ตารางแสดงค่าความสัมพันธ์และผลการทดสอบความสัมพันธ์ระหว่างตัวแปรใด ๆ 2 ตัวแปร โดยไม่มีการควบคุมอิทธิพลของตัวแปรตัวอื่น ๆ ดูจากตาราง **Correlations** ดังภาพที่ 11.7

Correlations

Correlations

		blood	prog	enzyme	liver	yprime
blood	Pearson Correlation	1	.090	-.150	.502**	.346*
	Sig. (2-tailed)	.	.517	.280	.000	.010
	N	54	54	54	54	54
prog	Pearson Correlation	.090	1	-.024	.369**	.593**
	Sig. (2-tailed)	.517	.	.865	.006	.000
	N	54	54	54	54	54
enzyme	Pearson Correlation	-.150	-.024	1	.416**	.665**
	Sig. (2-tailed)	.280	.865	.	.002	.000
	N	54	54	54	54	54
liver	Pearson Correlation	.502**	.369**	.416**	1	.726**
	Sig. (2-tailed)	.000	.006	.002	.	.000
	N	54	54	54	54	54
yprime	Pearson Correlation	.346*	.593**	.665**	.726**	1
	Sig. (2-tailed)	.010	.000	.000	.000	.
	N	54	54	54	54	54

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

ภาพที่ 11.7

(4) จากขั้นตอนที่ 4, 5 และขั้นตอนที่ 6 เฉพาะการตรวจสอบความเป็นอิสระของความคลาดเคลื่อน จะได้ผลลัพธ์เป็นตารางต่าง ๆ ดังต่อไปนี้ ดังภาพที่ 11.8

- ตาราง **Variables Entered/Removed**
- ตาราง **Model Summary** ที่ให้ค่า **R Square, R Square Change**, และค่าสถิติ **Durbin-Watson**

- ตาราง ANOVA แสดงการวิเคราะห์ความแปรปรวนแบบทางเดียวของตัวแบบต่าง ๆ
- ตาราง Coefficients ที่แสดงค่าสัมประสิทธิ์การถดถอยสำหรับตัวแบบต่าง ๆ และช่วงความเชื่อมั่นของสัมประสิทธิ์การถดถอย ตามด้วยคอลัมน์ของ Correlations ที่แสดงค่า Zero-order Correlations และ Partial Correlations และต่อด้วยคอลัมน์ Collinearity Statistics ที่แสดงค่า Tolerance และ VIF
- ตาราง Collinearity Diagnostics ของตัวแบบต่าง ๆ ที่แสดงค่า Eigenvalue และค่า Condition Index
- ตาราง Residuals Statistics

Regression

Variables Entered/Removed ^a

Model	Variables Entered	Variables Removed	Method
1	liver		Stepwise (Criteria: Probability- of- F-to-enter <= .050, Probability- of- F-to-remov e >= .100).
2	enzyme		Stepwise (Criteria: Probability- of- F-to-enter <= .050, Probability- of- F-to-remov e >= .100).
3	prog		Stepwise (Criteria: Probability- of- F-to-enter <= .050, Probability- of- F-to-remov e >= .100).
4	blood		Stepwise (Criteria: Probability- of- F-to-enter <= .050, Probability- of- F-to-remov e >= .100).
5		liver	Stepwise (Criteria: Probability- of- F-to-enter <= .050, Probability- of- F-to-remov e >= .100).

^a. Dependent Variable: yprime

Model Summary^f

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.726 ^a	.527	.518	.1900229	.527	58.021	1	52	.000	
2	.829 ^b	.687	.674	.1562702	.159	25.889	1	51	.000	
3	.940 ^c	.883	.876	.0964678	.196	83.831	1	50	.000	
4	.986 ^d	.972	.970	.0473368	.089	158.652	1	49	.000	
5	.986 ^e	.972	.971	.0468797	.000	.039	1	49	.844	2.141

- a. Predictors: (Constant), liver
- b. Predictors: (Constant), liver, enzyme
- c. Predictors: (Constant), liver, enzyme, prog
- d. Predictors: (Constant), liver, enzyme, prog, blood
- e. Predictors: (Constant), enzyme, prog, blood
- f. Dependent Variable: yprime

ANOVA^f

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2.095	1	2.095	58.021	.000 ^a
	Residual	1.878	52	.036		
	Total	3.973	53			
2	Regression	2.727	2	1.364	55.840	.000 ^b
	Residual	1.245	51	.024		
	Total	3.973	53			
3	Regression	3.507	3	1.169	125.633	.000 ^c
	Residual	.465	50	.009		
	Total	3.973	53			
4	Regression	3.863	4	.966	430.981	.000 ^d
	Residual	.110	49	.002		
	Total	3.973	53			
5	Regression	3.863	3	1.288	585.889	.000 ^e
	Residual	.110	50	.002		
	Total	3.973	53			

- a. Predictors: (Constant), liver
- b. Predictors: (Constant), liver, enzyme
- c. Predictors: (Constant), liver, enzyme, prog
- d. Predictors: (Constant), liver, enzyme, prog, blood
- e. Predictors: (Constant), enzyme, prog, blood
- f. Dependent Variable: yprime

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
		1	(Constant)	1.70			.072		23.645	.000	1.552	1.840	
	liver	.186	.024	.726	7.617	.000	.137	.235	.726	.726	.726	1.000	1.000
2	(Constant)	1.39	.084		16.441	.000	1.219	1.558					
	liver	.139	.022	.543	6.302	.000	.095	.183	.726	.662	.494	.827	1.210
	enzyme	.006	.001	.439	5.088	.000	.003	.008	.665	.580	.399	.827	1.210
3	(Constant)	.942	.071		13.198	.000	.799	1.086					
	liver	.082	.015	.320	5.464	.000	.052	.112	.726	.611	.264	.683	1.464
	enzyme	.007	.001	.543	9.980	.000	.006	.008	.665	.816	.483	.790	1.265
	prog	.008	.001	.488	9.156	.000	.006	.010	.593	.791	.443	.826	1.211
4	(Constant)	.489	.050		9.728	.000	.388	.590					
	liver	.002	.010	.007	.198	.844	-.018	.021	.726	.028	.005	.391	2.555
	enzyme	.009	.000	.735	23.907	.000	.009	.010	.665	.960	.568	.596	1.678
	prog	.009	.000	.571	21.187	.000	.008	.010	.593	.950	.503	.776	1.289
	blood	.069	.005	.401	12.596	.000	.058	.079	.346	.874	.299	.556	1.799
5	(Constant)	.484	.043		11.344	.000	.398	.569					
	enzyme	.010	.000	.739	31.077	.000	.009	.010	.665	.975	.731	.978	1.023
	prog	.009	.000	.574	24.296	.000	.009	.010	.593	.960	.571	.992	1.008
	blood	.069	.004	.405	16.974	.000	.061	.077	.346	.923	.399	.970	1.031

a. Dependent Variable: yprime

Excluded Variables^e

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
1	blood	-.025 ^a	-.221	.826	-.031	.748	1.338	.748
	prog	.376 ^a	4.217	.000	.509	.864	1.158	.864
	enzyme	.439 ^a	5.088	.000	.580	.827	1.210	.827
2	blood	.235 ^b	2.412	.020	.323	.592	1.690	.500
	prog	.488 ^b	9.156	.000	.791	.826	1.211	.683
3	blood	.401 ^c	12.596	.000	.874	.556	1.799	.391
5	liver	.007 ^d	.198	.844	.028	.391	2.555	.391

a. Predictors in the Model: (Constant), liver

b. Predictors in the Model: (Constant), liver, enzyme

c. Predictors in the Model: (Constant), liver, enzyme, prog

d. Predictors in the Model: (Constant), enzyme, prog, blood

e. Dependent Variable: yprime

Collinearity Diagnostics^a

Model	Dimension	Eigen value	Condition Index	Variance Proportions				
				(Constant)	liver	enzyme	prog	blood
1	1	1.933	1.000	.03	.03			
	2	.067	5.362	.97	.97			
2	1	2.891	1.000	.01	.01	.01		
	2	.074	6.261	.19	.97	.09		
	3	.035	9.097	.80	.02	.91		
3	1	3.835	1.000	.00	.01	.00	.00	
	2	.077	7.076	.08	.73	.00	.13	
	3	.068	7.510	.01	.12	.44	.28	
	4	.020	13.785	.91	.15	.56	.59	
4	1	4.776	1.000	.00	.00	.00	.00	.00
	2	.082	7.616	.01	.07	.25	.00	.16
	3	.076	7.923	.03	.32	.06	.17	.01
	4	.056	9.234	.02	.11	.03	.43	.20
	5	.009	22.571	.94	.50	.66	.40	.63
5	1	3.843	1.000	.00		.00	.00	.00
	2	.081	6.882	.00		.49	.02	.32
	3	.060	8.028	.00		.07	.71	.33
	4	.017	15.201	1.00		.44	.26	.35

a. Dependent Variable: yprime

Residuals Statistics^a

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	1.604248	2.863085	2.206143	.2699701	54
Residual	-.1020467	.1383532	.0000000	.0455336	54
Std. Predicted Value	-2.229	2.433	.000	1.000	54
Std. Residual	-2.177	2.951	.000	.971	54

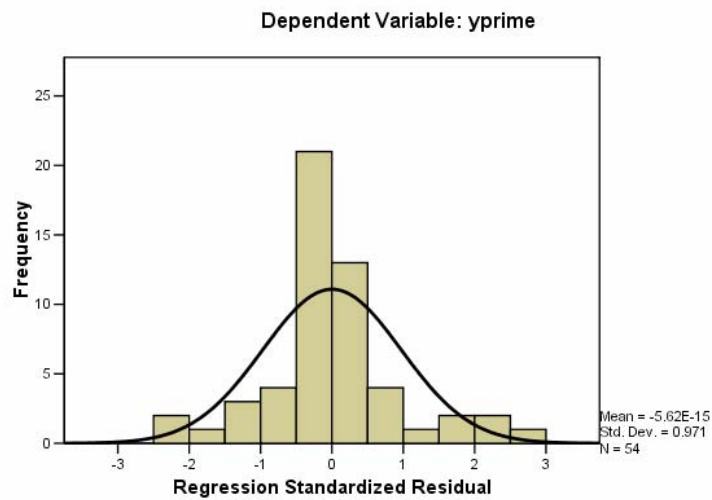
a. Dependent Variable: yprime

ภาพที่ 11.8

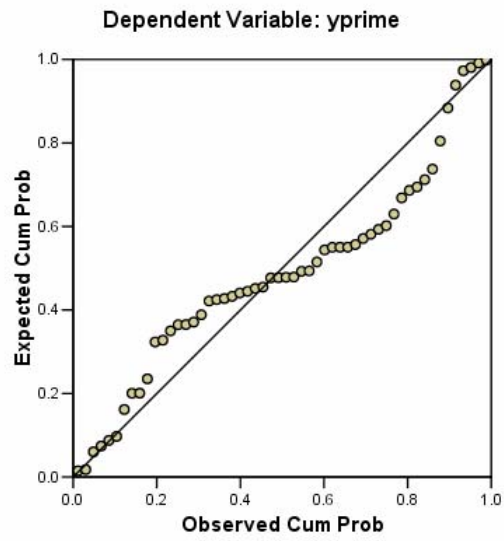
(5) กราฟแสดงการแจกแจงของคะแนนมาตรฐานของความคลาดเคลื่อนแบบสี่โต
แกรมและ Normal P-P Plot และแผนภาพการกระจายระหว่าง Standardized
Residual กับ Standardized Predicted Value ดังภาพที่ 11.9

Charts

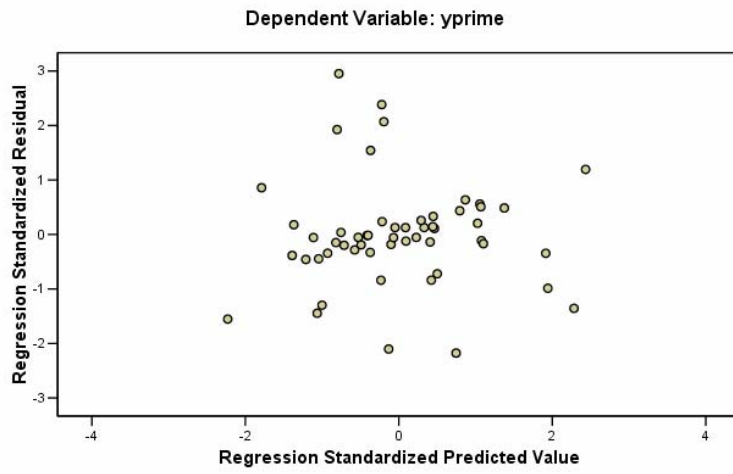
Histogram



Normal P-P Plot of Regression Standardized Residual



Scatterplot



ภาพที่ 11.9

12.3 การสรุปผลการพัฒนาตัวแบบการถดถอย

ผลการพัฒนาตัวแบบการถดถอยแบบพหุที่มีตัวแปรอิสระหรือตัวแปรทำนาย 4 ตัวแปร คือ blood, prog, enzyme, และ liver เพื่อทำนายตัวแปรตามคือ เวลารอดชีวิต (survival) ซึ่งถูกแปลงข้อมูลเป็น $\log_{10}(\text{survival})$ แล้วตั้งชื่อตัวแปรใหม่เป็น yprime พัฒนาตัวแบบการถดถอยด้วยวิธี Stepwise สรุปผลได้ดังนี้คือ

(1) คุณสมบัติของค่าความสัมพันธ์ระหว่างตัวแปรทั้งหมดเป็นรายคู่โดยไม่มีการควบคุมอิทธิพลของตัวแปรตัวอื่น ๆ ได้จากตาราง Correlations เขียนใหม่เป็นตารางเมตริกความสัมพันธ์ระหว่างตัวแปรทั้งหมดได้ดังนี้

ตารางที่ 11.4 ตารางเมตริกความสัมพันธ์ระหว่างตัวแปรทั้งหมด

	yprime	blood	prog	enzyme	liver
yprime	1.000	.346*	.593**	.665**	.726**
blood	.346*	1.000	.090	-.150	.502**
prog	.593**	.090	1.000	-.024	.369**
enzyme	.665**	-.150	-.024	1.000	.416**
liver	.726**	.502**	.369**	.416**	1.000

* ค่าความสัมพันธ์มีนัยสำคัญที่ .05 (ทดสอบแบบ 2 ทาง)

** ค่าความสัมพันธ์มีนัยสำคัญที่ .01 (ทดสอบแบบ 2 ทาง)

(2) ผลการทดสอบสมมติฐานของตัวแบบการถดถอย $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ คู่กับ $H_1 : \text{มี } \beta_i \text{ อย่างน้อย 1 ค่าที่ไม่เท่ากับ } 0$ ได้จากตาราง ANOVA ในตัวแบบที่ 5 ซึ่งมีค่าสถิติ F เท่ากับ 585.889 ค่า Sig. เท่ากับ .000 สรุปผลการทดสอบได้ว่าปฏิเสธ H_0 นั่นคือ ตัวแบบที่ 5 เป็นตัวแบบการถดถอยที่เหมาะสมซึ่งประกอบด้วยตัวแปรทำนาย 3 ตัวแปรคือ enzyme, prog, และ blood

(3) ค่าสัมประสิทธิ์ต่าง ๆ ที่ใช้อธิบายตัวแบบการถดถอย ได้แก่ ค่า R Square, R Square Change ได้จากตาราง Model Summary บนบรรทัดของตัวแบบที่ 5 ได้ค่า R Square เท่ากับ .972, ค่า R Square Change เท่ากับ .000, ค่า F Change เท่ากับ .039 และค่า R เท่ากับ .986

(4) ค่าสัมประสิทธิ์การถดถอยของตัวแบบการถดถอย ได้จากตาราง
Coefficients บนบรรทัดของตัวแบบที่ 5 ในคอลัมน์ Unstandardized
Coefficients สำหรับ B ได้สัมประสิทธิ์การถดถอย 4 ตัว คือ

- (Constant) เท่ากับ 0.484 95% CI คือ (.398 , .569)
- enzyme เท่ากับ 0.010 95% CI คือ (.009 , .010)
- prog เท่ากับ 0.009 95% CI คือ (.009 , .010)
- blood เท่ากับ 0.069 95% CI คือ (.061 , .077)

เขียนเป็นตัวแบบการถดถอยได้คือ

$$\hat{y} = 0.484 + 0.10 \text{ enzyme} + 0.009 \text{ prog} + 0.069 \text{ blood}$$

(5) ค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนของตัวแปรทำนายแต่ละตัวได้จากตาราง
Coefficients บนบรรทัดของตัวแบบที่ 5 ในคอลัมน์ Correlations สำหรับ
Partial ได้ค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนของตัวแปรตาม (yprime) กับตัวแปร
ทำนายตัวใดตัวหนึ่งเมื่อควบคุมตัวแปรทำนายตัวอื่น ๆ ได้ผลคือ

- enzyme มีค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนเท่ากับ .975
- prog มีค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนเท่ากับ .960
- blood มีค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนเท่ากับ .923

(6) ผลการตรวจสอบความเป็นอิสระของตัวแปรทำนายทั้ง 3 ตัว โดยพิจารณาจาก
ตาราง Coefficients ในตัวแบบที่ 5 ในคอลัมน์ Collinearity Statistics ซึ่งมี
ค่า Tolerance และ VIF คือ

ตัวแปร	Tolerance	VIF
- enzyme	.978	1.023
- prog	.992	1.008
- blood	.970	1.031

และในตาราง Collinearity Diagnostics ได้ค่า Condition Index ของตัวแปรทำนายทั้ง 3 ตัว คือ

ตัวแปร	Condition Index
- enzyme	6.882
- prog	8.028
- blood	15.201

(7) ผลการตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยเกี่ยวกับความเป็นอิสระกันของความคลาดเคลื่อน ดูได้จากตาราง Model Summary บนบรรทัดของตัวแบบที่ 5 ในคอลัมน์ Durbin-Watson ได้ค่าสถิติ Durbin-Watson เท่ากับ 2.141

(8) ผลการตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยเกี่ยวกับความคลาดเคลื่อน ดูได้จาก Charts ของ Histogram และ Normal P-P Plot of Regression Standardized Residual พบว่ามีการแจกแจงเข้าใกล้แบบปกติและดูจาก Scatterplot ซึ่งพล็อตกราฟระหว่างค่าของ Regression Standardized Residual บนแกน Y กับค่าของ Regression Standardized Predicted Value บนแกน X พบว่าจุดมีการกระจายอย่างไม่เป็นระบบ แสดงว่าความแปรปรวนของค่าความคลาดเคลื่อนค่อนข้างคงที่

12.4 การแปลความหมายการพัฒนาตัวแบบการถดถอย

จากการสรุปผลการพัฒนาตัวแบบการถดถอยด้วยวิธี Stepwise ของตัวอย่างการศึกษาเรื่องการผ่าตัดตัดตับ สามารถแปลความหมายได้คือ

(1) จากผลของค่าความสัมพันธ์ระหว่างตัวแปรทั้งหมดเป็นรายคู่โดยไม่มีการควบคุมอิทธิพลของตัวแปรตัวอื่น ๆ ทำการทดสอบสมมติฐานเกี่ยวกับ $H_0 : \rho = 0$ คู่กับ $H_1 : \rho \neq 0$ พบว่าตัวแปรอิสระทุกตัวมีความสัมพันธ์กับตัวแปรตาม (yprime) โดยเฉพาะ liver มีค่ามากที่สุด และ blood มีค่าน้อยที่สุด และยังแสดงความสัมพันธ์ระหว่างตัวแปร

อิสระด้วยตัวเองด้วย โดยเฉพาะความสัมพันธ์ระหว่าง liver กับ blood และ prog และ enzyme มีค่าพอสมควร

(2) ผลการพัฒนาตัวแบบการถดถอย ได้ตัวแบบการถดถอยคือ

$$\hat{yprime} = 0.484 + 0.10 \text{ enzyme} + 0.009 \text{ prog} + 0.069 \text{ blood}$$

- ค่า 0.10 ของตัวแปร enzyme หมายถึง ค่า yprime ที่เปลี่ยนแปลงไปเมื่อ enzyme เปลี่ยนแปลงไป 1 หน่วย โดย prog และ blood อยู่ในตัวแบบ แต่กำหนดให้คงที่

- ค่า 0.009 ของตัวแปร prog หมายถึง ค่า yprime ที่เปลี่ยนแปลงไปเมื่อ prog เปลี่ยนแปลงไป 1 หน่วย โดย enzyme และ blood อยู่ในตัวแบบ แต่กำหนดให้คงที่

- ค่า 0.069 ของตัวแปร blood หมายถึง ค่า yprime ที่เปลี่ยนแปลงไปเมื่อ blood เปลี่ยนแปลงไป 1 หน่วย โดย enzyme และ prog อยู่ในตัวแบบ แต่กำหนดให้คงที่

การเปรียบเทียบอิทธิพลหรือความสำคัญของตัวแปรทำนายทั้ง 3 ตัวนี้ที่มีต่อตัวแปรตาม yprime สามารถดูได้จากค่า Standardized Coefficients ของ Beta ซึ่งคำนวณได้จาก

$$Beta_j = b_j \frac{S_x}{S_y}$$

เมื่อ b_j คือ สัมประสิทธิ์ความถดถอยของตัวแปรทำนายที่ j

S_x คือ ค่าเบี่ยงเบนมาตรฐานของตัวแปรทำนายที่ j

S_y คือ ค่าเบี่ยงเบนมาตรฐานของตัวแปรตาม Y

และสามารถดูค่า Standardized Coefficients ของ Beta ได้จากตาราง Coefficients ดังนี้

ตัวแปร	Standardized Coefficients ของ Beta
--------	------------------------------------

enzyme	.739
prog	.574
blood	.405

แสดงว่าตัวแปรอิสระ enzyme มีความสำคัญต่อตัวแปรตาม yprime มากที่สุด รองลงมาคือตัวแปรอิสระ prog และ blood ตามลำดับ

- ค่าประมาณแบบช่วงของสัมประสิทธิ์การถดถอย β_i ที่ระดับความเชื่อมั่น 95% (95% Confidence Interval for B) ของตัวแปรอิสระ enzyme , prog , blood และ constant ไม่คลุมค่าศูนย์ ซึ่งเหมือนกับผลการทดสอบโดยใช้สถิติทดสอบ t ที่ทำการทดสอบสมมติฐานเกี่ยวกับค่าคงที่ (β_0) คือ $H_0 : \beta_0 = 0$ คู่กับ $H_1 : \beta_0 \neq 0$ ได้ค่าสถิติทดสอบ $t = 11.344$ ค่า Sig. = .000 สรุปผลการทดสอบได้ว่า ปฏิเสธ H_0

ทดสอบสมมติฐานเกี่ยวกับสัมประสิทธิ์การถดถอย (β_i) คือ

$H_0 : \beta_{enzyme} = 0$ คู่กับ $H_1 : \beta_{enzyme} \neq 0$

ได้ค่าสถิติทดสอบ $t = 31.077$ ค่า Sig. = .000 สรุปผลการทดสอบได้ว่า ปฏิเสธ H_0

$H_0 : \beta_{prog} = 0$ คู่กับ $H_1 : \beta_{prog} \neq 0$

ได้ค่าสถิติทดสอบ $t = 24.296$ ค่า Sig. = .000 สรุปผลการทดสอบได้ว่า ปฏิเสธ H_0

$H_0 : \beta_{blood} = 0$ คู่กับ $H_1 : \beta_{blood} \neq 0$

ได้ค่าสถิติทดสอบ $t = 16.974$ ค่า Sig. = .000 สรุปผลการทดสอบได้ว่า ปฏิเสธ H_0

นั่นคือ สรุปได้ว่า

ตัวแปรอิสระ enzyme มีความสัมพันธ์กับตัวแปรตาม yprime เมื่อกำหนดให้ตัวแปรอิสระ prog และ blood คงที่

ตัวแปรอิสระ prog มีความสัมพันธ์กับตัวแปรตาม yprime เมื่อกำหนดให้ตัวแปรอิสระ enzyme และ blood คงที่

ตัวแปรอิสระ blood มีความสัมพันธ์กับตัวแปรตาม yprime เมื่อกำหนดให้ตัวแปรอิสระ enzyme และ prog คงที่

(3) ค่าสัมประสิทธิ์ต่าง ๆ ที่ใช้อธิบายตัวแบบการถดถอย

- ค่า R Square คือ สัมประสิทธิ์การกำหนดเชิงพหุ (Multiple Coefficient of Determination : R^2 หรือ r^2) มีความหมายคือ เป็นสัดส่วนหรือเปอร์เซ็นต์ที่ตัวแปรทำนายสามารถอธิบายการเปลี่ยนแปลงของตัวแปรตาม หรือคือสัดส่วนหรือเปอร์เซ็นต์ของความผันแปรของตัวแปรตามที่มีสาเหตุเนื่องจากความผันแปรของตัวแปรทำนายเขียนเป็นสูตรได้ดังนี้

$$R^2 = \frac{\text{ความผันแปรของ } Y \text{ เนื่องจากอิทธิพลของ } X_1, X_2, \dots, X_k}{\text{ความผันแปรทั้งหมด}}$$

$$= \frac{SSR}{SST}$$

ค่า R Square ของตัวแบบการถดถอยเท่ากับ .972 หมายความว่าตัวแปรอิสระ enzyme , prog , และ blood ในตัวแบบการถดถอยสามารถอธิบายการเปลี่ยนแปลงของเวลารอดชีวิตได้ 97.2%

- ค่า R Square Change จะแสดงค่า R^2 ที่เปลี่ยนไปเมื่อสมการเปลี่ยนหรือเมื่อมีตัวแปรอิสระเพิ่มขึ้นหรือลดลงในสมการถดถอยใช้พิจารณาความสำคัญของตัวแปรอิสระที่มีต่อตัวแปรตาม ถ้าค่า R Square Change ของ X_i มีค่ามากแสดงว่า X_i มีความสัมพันธ์กับ Y มาก จึงควรนำ X_i เข้าสมการถดถอย แต่ถ้ามีค่าน้อยแสดงว่า X_i ที่เพิ่มเข้ามาในสมการถดถอยอาจมีความสัมพันธ์กับ Y น้อยมากหรืออาจไม่มีความสัมพันธ์กับ Y เลยก็ได้

ค่า R Square Charge ของตัวแบบการถดถอย เท่ากับ .000 หมายความว่า การตัดตัวแปร liver ออกจากสมการถดถอยไม่ทำให้ค่า R^2 เปลี่ยน

- ค่า R คือ สัมประสิทธิ์สหสัมพันธ์เชิงพหุ (Multiple Coefficient of Correlation) ได้จากการถอดรากที่สองของสัมประสิทธิ์การตัดสินใจเชิงพหุ มีค่าอยู่ระหว่าง 0 และ 1 หมายถึงค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง Y กับ Y ถ้าค่า R เข้าใกล้ 0 แสดงว่า Y มีความสัมพันธ์กับ X_1, X_2, \dots, X_k น้อยมาก ถ้า R เท่ากับ 0 แสดงว่า Y ไม่มีความสัมพันธ์กับ X_1, X_2, \dots, X_k ถ้าค่า R เข้าใกล้ 1 แสดงว่า Y มีความสัมพันธ์มากกับตัวแปรอิสระทั้ง k ตัว

ค่า R ของตัวแบบการถดถอยเท่ากับ .986 หมายความว่าตัวแปรตาม เวลารอดชีวิต มีความสัมพันธ์มากกับตัวแปรอิสระ enzyme , prog , และ blood

- ค่า F Change จะแสดงค่าสถิติทดสอบ F ที่เปลี่ยนไป พร้อมทั้งค่า $Sig.$ ของ F Change เมื่อจำนวนตัวแปรอิสระเปลี่ยนไปคือ เพิ่มขึ้นหรือลดลง ใช้พิจารณาความสำคัญหรือความสัมพันธ์ของตัวแปรอิสระที่เพิ่มเข้าหรือตัดออกจากสมการถดถอย ถ้าค่าสถิติทดสอบ F เพิ่มขึ้นอย่างมากแสดงว่าตัวแปร X ที่เพิ่มเข้ามาในสมการถดถอยมีความสัมพันธ์กับ Y มาก

ค่า F Change ของตัวแบบการถดถอยเท่ากับ .039 ที่ df เท่ากับ 1 และ 51 และค่า $Sig.$ เท่ากับ .844 หมายความว่าเมื่อตัดตัวแปรอิสระ liver ออกจากสมการถดถอย แล้วทดสอบสมมติฐาน $H_0 : \beta_{\text{liver}} = 0$ คู่กับ $H_1 : \beta_{\text{liver}} \neq 0$ จะมีค่าสถิติทดสอบ $F_{1,51} = .039$ และมีค่า $Sig. F$ Change = .844 ซึ่งมากกว่าระดับนัยสำคัญที่กำหนดในการเลือกตัวแปรออก ($\alpha = .10$) จึงสรุปผลการทดสอบว่ายอมรับ H_0 นั่นคือตัดตัวแปรอิสระ liver ออกจากสมการถดถอย

- ค่า $Std. Error$ of the Estimate คือ ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า y_{prime} เท่ากับ .047

(4) ความเป็นอิสระของตัวแปรทำนายทั้ง 3 ตัวแปร เมื่อพิจารณาค่า Tolerance ของตัวแปรทำนายแต่ละตัวจะพบว่าทั้ง 3 ตัวแปร คือ enzyme , prog , และ blood มีค่า Tolerance มากเข้าใกล้ 1 แสดงว่าตัวแปรแต่ละตัวมีความสัมพันธ์กับตัวแปรตัวอื่น ๆ น้อยมาก

ส่วนค่า VIF ของทั้ง 3 ตัวแปรนี้ มีค่าน้อยมากเข้าใกล้ 1 แสดงว่าตัวแปรแต่ละตัวมีความสัมพันธ์กับตัวแปรตัวอื่น ๆ น้อยมาก

ส่วนค่า condition Index ของทั้ง 3 ตัวแปรนี้มีค่าน้อยแสดงว่าตัวแปรแต่ละตัวมีความสัมพันธ์กับตัวแปรตัวอื่น ๆ น้อย

ซึ่งค่า Tolerance , ค่า VIF , และค่า Condition Index ให้ผลสอดคล้องกันทั้ง 3 ค่า

(5) ความเป็นอิสระกันของความคลาดเคลื่อน โดยใช้สถิติทดสอบ Durbin-Watson ทดสอบสมมติฐานเกี่ยวกับ $H_0 :$ ค่าความคลาดเคลื่อน (e_i) เป็นอิสระกัน คู่กับ $H_1 :$ ค่าความคลาดเคลื่อนไม่เป็นอิสระกัน ได้ค่าสถิติ Durbin-Watson =

2.141 ซึ่งมีค่า ≈ 2 จึงสรุปผลการทดสอบได้ว่ายอมรับ H_0 นั่นคือ ค่าความคลาดเคลื่อนเป็นอิสระกัน