บทที่ 10
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 เป็นตัวแปรแบบกำหนดไม่ใช่ตัวแปรสุ่ม

(2) สำหรับแต่ละเซตของค่า X_i จะมี Y เป็นตัวแปรเชิงสุ่มที่มีการแจกแจงความน่าจะเป็นแบบปกติที่มีค่าเฉลี่ยและความแปรปรวนคือ $Y \sim N(\mu , \sigma^2)$

(3) ค่าสังเกต Y ทุกตัวเป็นอิสระกัน

(4) ค่าเฉลี่ยของ Y ที่แต่ละจุดของตัวแปรอิสระทุกตัว $X_1, X_2, ..., X_k$ เป็นฟังก์ชันเส้นตรงของ $X_1, X_2, \dots , X_k$ นั่นคือ

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

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

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

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

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

ตัวแบบการถดถอยพหุเชิงเส้นแบบทั่วไปที่มีตัวแปรอิสระ k ตัวคือ

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

เมื่อ y_j คือ ค่าตัวหนึ่งจากประชากรย่อยของค่า Y ประชากรหนึ่งเป็นตัวแปรตาม

$\beta_0, \beta_1, \ldots, \beta_k$ คือ สัมประสิทธิ์การถดถอยบางส่วน (partial – regression coefficient)

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

$x_{1j}, x_{2j}, \ldots, x_{kj}$ คือค่าเฉพาะของตัวแปรอิสระ $X_1, X_2, \ldots, X_k$

e_j คือ ความคลาดเคลื่อนที่แสดงความแตกต่างระหว่างค่าสังเกต y แต่ละตัวกับ $\mu_{Y|X_1, X_2, \ldots, X_k}$ เป็นตัวแปรสุ่มที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนคงที่เท่ากับ σ^2 สำหรับแต่ละค่าของ X และความคลาดเคลื่อนทุกตัวเป็นอิสระกัน

ขั้นตอนการวิเคราะห์การถดถอยเหมือนกับการวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายที่อธิบายไว้แล้วในบทที่ 9

การประมาณสมการถดถอยพหุด้วยวิธีกำลังสองน้อยที่สุดจะได้ตัวแบบซึ่งผลบวกกำลังสองของระยะห่างระหว่างค่าสังเกตและค่าทำนายมีค่าน้อยที่สุด นั่นคือ ถ้าให้ตัวแบบที่เหมาะสมกับข้อมูลคือ

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_k x_k$$

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

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

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 10.1 ตารางวิเคราะห์ความแปรปรวนสำหรับการวิเคราะห์การถดถอยพหุ

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

จำนวนชั้นอิสระของแหล่งความแปรปรวนแหล่งต่าง ๆ ในการวิเคราะห์ความแปรปรวน คือ จำนวนชั้นอิสระของ Regression , Error , และ Total เท่ากับ k , n – k – 1 , และ n – 1 ตามลำดับ

การคำนวณหาค่าเฉลี่ยของผลบวกกำลังสอง (Mean Square) คำนวณจากผลบวกกำลังสองหารด้วยจำนวนชั้นอิสระของเทอมนั้น

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

2. การทดสอบสมมติฐานในการวิเคราะห์การถดถอยพหุ

การวิเคราะห์การถดถอยพหุสามารถตอบคำถามที่สำคัญ 3 ข้อ คือ

(1) ตัวแปรอิสระทั้งหมดมีส่วนสำคัญในการใช้ทำนาย Y ได้อย่างมีนัยสำคัญ

(2) การเพิ่มตัวแปรอิสระที่สนใจเข้าไปในตัวแบบอีก 1 ตัว มีส่วนสำคัญในการใช้ทำนาย Y ได้เพิ่มขึ้นอย่างมีนัยสำคัญ

(3) การเพิ่มตัวแปรอิสระที่สนใจเข้าไปในตัวแบบอีก 1 ชุด มีส่วนสำคัญในการใช้ทำนาย Y ได้เพิ่มขึ้นอย่างมีนัยสำคัญ

คำถามเหล่านี้สามารถหาคำตอบได้จากการทดสอบสมมติฐานทางสถิติคือ $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$ คู่กับ H_1 : มี β_i อย่างน้อย 1 ค่าที่ไม่เท่ากับ 0 เมื่อ $i = 1, 2, \ldots, k$

ผลการทดสอบสมมติฐานพิจารณาจากค่าสถิติทดสอบ F_0 ในตารางวิเคราะห์ความแปรปรวนว่ามีนัยสำคัญทางสถิติหรือไม่ ถ้าค่า Sig. มีค่าน้อยกว่าระดับนัยสำคัญที่กำหนด $\alpha = .05$ จะสรุปว่าปฏิเสธ H_0 แสดงว่ามีตัวแปรอิสระ X อย่างน้อย 1 ตัวแปรที่มีความสัมพันธ์กับตัวแปรตาม Y หมายความว่า ตัวแบบ $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_k x_k$ ใช้แทนรูปแบบความสัมพันธ์ระหว่างตัวแปรตาม Y และตัวแปรอิสระ X ทั้ง k ตัวได้ หรือตัวแปรอิสระ X ทั้ง k ตัวมีส่วนสำคัญในการใช้ทำนายตัวแปรตาม Y ได้อย่างมีนัยสำคัญทางสถิติ

ถ้าสรุปว่าปฏิเสธ H_0 ขั้นต่อไปต้องทดสอบสมมติฐานเกี่ยวกับ β_i ทีละตัวเมื่อควบคุมตัวแปรอิสระตัวอื่น ๆ ให้คงที่เพื่อหาว่ามีตัวแปรอิสระตัวใดบ้างที่มีความสัมพันธ์กับตัวแปรตาม Y อย่างมีนัยสำคัญ สมมติฐานที่ต้องการทดสอบคือ $H_0 : \beta_i = 0$ คู่กับ $H_1 : \beta_i \neq 0$; $i = 1, 2, \ldots, k$ สถิติทดสอบคือ สถิติทดสอบ t ผลการทดสอบสมมติฐานถ้าค่าสถิติทดสอบ t มีนัยสำคัญคือ ค่า Sig. มีค่าน้อยกว่าระดับนัยสำคัญที่กำหนด $\alpha = .05$ ก็แสดงว่าตัวแปรอิสระ X_i ตัวนั้นมีความสัมพันธ์เชิงเส้นกับตัวแปรตาม Y อย่างมีนัยสำคัญทางสถิติ เราก็จะนำตัวแปรอิสระ 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. วิธีการเลือกตัวแปรอิสระเข้าสมการถดถอย

ในขั้นตอนการวิเคราะห์การถดถอยที่อธิบายไว้แล้วในบทที่ 9 เริ่มจากการตรวจสอบว่าตัวแบบเส้นตรงเป็นตัวแบบที่เหมาะสมและเป็นไปตามข้อตกลงเบื้องต้นเกี่ยวกับความเป็นปกติแล้วจึงคำนวณหาสมการถดถอยเชิงเส้นตรงที่เหมาะสมที่สุดกับข้อมูล ในขั้นตอนนี้เป็นการนำตัวแปรอิสระ X_i เข้าสมการถดถอย

การสร้างตัวแบบการถดถอยที่เป็นไปได้ทั้งหมดเมื่อมีตัวแปรอิสระหรือตัวแปรทำนายทั้งหมด k ตัว สามารถสร้างตัวแบบที่เป็นไปได้ทั้งหมด 2^k ตัวแบบ มีวิธีการหลายวิธีที่ใช้สร้างตัวแบบเหล่านี้ วิธีการที่ใช้ต่อไปนี้สามารถรวบรวมสารสนเทศที่สำคัญ ๆ ทั้งหมดในข้อมูลเพื่อจะเลือกให้ได้ตัวแบบที่ดีที่สุด ซึ่งวิธีการเลือกตัวแปรอิสระ X_i เข้าสมการมี 5 วิธี คือ (1) Enter (2) Remove (3) Forward (4) Backward (5) Stepwise

4.1 วิธี Enter

กำหนดให้สมการถดถอยประกอบด้วยตัวแปรอิสระทั้งหมด โดยปกติโปรแกรม SPSS กำหนดให้นำตัวแปรอิสระทุกตัวเข้าในสมการถดถอยอยู่แล้วโดยวิธี Enter ผู้วิเคราะห์จะต้องตัดสินใจเองว่าตัวแปรอิสระ X_i ตัวใดบ้างที่ควรอยู่ในสมการถดถอยโดยพิจารณาจากค่า Sig. ของสถิติทดสอบ F ของสมการที่ประกอบด้วยตัวแปรอิสระ X_i ทั้งหมดและค่า Sig. ของสถิติทดสอบ t ของตัวแปรอิสระ X_i แต่ละตัวว่ามีนัยสำคัญทางสถิติหรือไม่ (ค่า Sig. น้อยกว่าระดับนัยสำคัญที่กำหนด $\alpha = .05$) นอกจากนี้ยังต้องพิจารณาข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยเกี่ยวกับความเป็นอิสระของตัวแปรอิสระด้วย แล้วทำการเปลี่ยนแปลงตัวแบบ ดังอธิบายไว้ในหัวข้อที่ 5

4.2 วิธี Remove

เป็นการเลือกตัวแปรอิสระที่กำหนดออกจากสมการถดถอยภายในขั้นตอนเดียว วิธีนี้จะใช้คู่กับวิธี Enter โปรแกรม SPSS ไม่อนุญาตให้เลือกใช้วิธีนี้เป็นวิธีแรกในการวิเคราะห์การถดถอย

4.3 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 $\sum_{j=1}^{n}(y_j - \bar{y}.)^2$ แล้วคำนวณหาสมการถดถอยเชิงเส้นของตัวแปรตาม Y และตัวแปรอิสระ X_i ตัวนั้น ถ้าค่าสถิติทดสอบ F ในตารางวิเคราะห์ความแปรปรวนไม่มีนัยสำคัญเราจะหยุดและสรุปว่าไม่มีตัวแปรอิสระตัวใดเลยที่เป็นตัวทำนายที่สำคัญแต่ถ้าค่าสถิติทดสอบ F มีนัยสำคัญ เราจะเก็บตัวแปรอิสระ X_i ตัวนั้นเข้าไว้ในตัวแบบแล้วทำขั้นที่ 2 ต่อไป ในที่นี้สมมติว่าได้ X_1 อยู่ในตัวแบบ

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

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

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

ขั้นที่ 5) คำนวณค่าสถิติ F บางส่วนของตัวแปรอิสระ X_i ที่เหลือแต่ละตัวเมื่อในตัวแบบมีตัวแปรอิสระที่ถูกเลือกจากขั้นตอนก่อนหน้านี้แล้ว พิจารณาตัวแปรอิสระ X_i ที่มีค่าสถิติ F บางส่วนมากที่สุดกว่าตัวอื่น ๆ และทดสอบนัยสำคัญ ถ้าพบว่ามีนัยสำคัญก็เพิ่มตัวแปรอิสระตัวนั้นเข้าไปในตัวแบบ แต่ถ้าเมื่อใดพบว่าค่าสถิติ F บางส่วนไม่มีนัยสำคัญก็จะไม่มีการเพิ่มตัวแปรอิสระตัวใด ๆ เข้าไปในตัวแบบอีกและหยุดการพัฒนาตัวแบบ ยกตัวอย่างเช่น ต้องการพิจารณาว่าควรเพิ่มตัวแปรอิสระ X_3 เข้าไปในตัวแบบหรือไม่ พิจารณาจากค่าสถิติทดสอบ partial F $(X_3|X_1, X_2)$ ถ้าพบว่าไม่มีนัยสำคัญก็จะไม่เพิ่มตัวแปรอิสระ X_3 เข้าไปในตัวแบบอีก และหยุดการพัฒนาตัวแบบการถดถอย

ซึ่งขั้นตอนทั้งหลายเหล่านี้ในโปรแกรม SPSS จะทำให้ถ้าเลือกวิธี Forward วิธีนี้ผู้วิเคราะห์เป็นผู้กำหนดเกณฑ์ในการเลือกตัวแปรอิสระ X_i เข้าในตัวแบบเองโดยกำหนดค่าระดับนัยสำคัญของการทดสอบสมมติฐานเพื่อคัดเลือกตัวแปรอิสระ X_i ตัวหนึ่ง ๆ เข้าไปอยู่ในสมการ

ถดถอย 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

4.4 วิธี Backward

เป็นการกำหนดให้สมการถดถอยประกอบด้วยตัวแปรอิสระ X_i ทั้งหมดก่อนแล้วจึงคัดเลือกตัวแปรอิสระ X_i ที่ไม่มีความสัมพันธ์กับตัวแปรตาม Y ออกจากตัวแบบทีละตัว โดยมีขั้นตอนดังนี้คือ

ขั้นที่ 1) กำหนดให้สมการถดถอยประกอบด้วยตัวแปรอิสระทั้งหมดสมมติว่ามี X_1 , X_2 , X_3

ขั้นที่ 2) คำนวณหาค่าสถิติ F บางส่วนของตัวแปรอิสระ X_i ทุกตัวในตัวแบบคือ partial F ($X_1|X_2 , X_3$) , partial F ($X_2|X_1 , X_3$) และ 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 ที่ใช้เป็นเกณฑ์ใน

การคัดตัวแปรอิสระ X_i ตัวหนึ่ง ๆ ออกจากสมการถดถอยเอง โดยปกติโปรแกรมจะกำหนดให้เท่ากับ 2.71 ซึ่งสอดคล้องกับค่าระดับนัยสำคัญเท่ากับ 0.10

4.5 วิธี Stepwise

เป็นวิธีที่นิยมใช้กันมากที่สุด เป็นวิธีที่พัฒนามาจากวิธี Forward แต่ไม่เหมือนกันคือ 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 partial F $(X_1|X_2)$ ว่ามีนัยสำคัญหรือไม่ สมมติว่า 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 ยกตัวอย่างเช่นต้องการ พิจารณาว่าควรเพิ่มตัวแปรอิสระ X_3 เข้าไปในสมการหรือไม่ พิจารณาจากค่าสถิติทดสอบ partial F $(X_3|X_1,X_2)$ ถ้าพบว่าไม่มีนัยสำคัญก็จะไม่เพิ่มตัวแปรอิสระ X_3 เข้าไปในสมการถดถอยอีก และหยุดการพัฒนาตัวแบบการถดถอย

ซึ่งขั้นตอนทั้งหลายเหล่านี้ในโปรแกรม SPSS จะทำให้ถ้าเลือกวิธี Stepwise วิธีนี้ผู้วิเคราะห์จะเป็นผู้กำหนดเกณฑ์ในการเลือกตัวแปรอิสระ X_i เข้าไปอยู่ในสมการถดถอย และเกณฑ์ในการคัดตัวแปรอิสระ 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

5. 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)

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 นั่นคือถ้าค่า tolerance ยิ่งมีค่าเข้าใกล้ 1 แสดงว่าตัวแปรอิสระ X_i ตัวนั้นมีความสัมพันธ์กับตัวแปรอื่น ๆ น้อย แต่ถ้าค่า tolerance มีค่าเข้าใกล้ศูนย์แสดงว่าตัวแปรอิสระ X_i มีความสัมพันธ์กับตัวแปรอิสระ อื่น ๆ มาก นั่นคือ เกิด multicollinearity ถ้าค่าต่ำกว่า 0.1 แสดงว่ามีปัญหาเกี่ยวกับ multicollinearity ขั้นรุนแรง และถ้าค่าต่ำกว่า 0.2 แสดงว่ามีปัญหาอยู่ในระดับมาก (Menard, 1995. อ้างถึงใน Field, 2000)

ค่า VIF ของ X_i เท่ากับ 1/tolerance มีค่าอยู่ระหว่าง 1 ถึง ∞ ถ้าค่า VIF มีค่ามาก หมายความตัวแปรอิสระ X_i มีความสัมพันธ์กับตัวแปรอิสระตัวอื่น ๆ มาก ถ้าค่า VIF มากกว่า 10 หมายความว่า มีหลักฐานเกี่ยวกับความสัมพันธ์ที่มากพอ (Myers, 1990; Bowerman and O'Connell, 1990. อ้างถึงใน Field, 2000)

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

6. ตรวจสอบข้อตกลงเบื้องต้นเกี่ยวกับความคลาดเคลื่อนโดยการวิเคราะห์เศษตกค้าง (Residual Analysis)

ข้อตกลงเบื้องต้นเกี่ยวกับเศษตกค้าง (residual) โดยที่ความคลาดเคลื่อน e เป็นค่าประมาณของเศษตกค้าง คือ ความคลาดเคลื่อน e_i แต่ละตัวมีความเป็นอิสระกันและมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับศูนย์และมีความแปรปรวนคงที่ทุกค่าของ X เท่ากับ σ^2 โดยที่ความคลาดเคลื่อน (e_i) เท่ากับผลต่างระหว่างค่าสังเกตของตัวแปรตามตัวที่ i กับค่าทำนายของตัวแปรตามตัวที่ i ($\hat{y}_i - y_i$)

เนื่องจากในการคำนวณค่าสัมประสิทธิ์การถดถอย $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$ ในสมการถดถอยเชิงเส้นนั้นใช้วิธีการกำลังสองน้อยที่สุด (Least Square Error) คือทำให้ผลบวกกำลังสองของความคลาดเคลื่อนแต่ละตัวมีค่าน้อยที่สุด ซึ่งมีผลทำให้ผลรวมของความคลาดเคลื่อนแต่ละตัว (e_j) เท่ากับ 0 ดังนั้นจึงทำให้ค่าคาดหวังของความคลาดเคลื่อนหรือค่าเฉลี่ยของความคลาดเคลื่อนเท่ากับ 0 ด้วย เราจึงไม่จำเป็นต้องตรวจสอบ

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

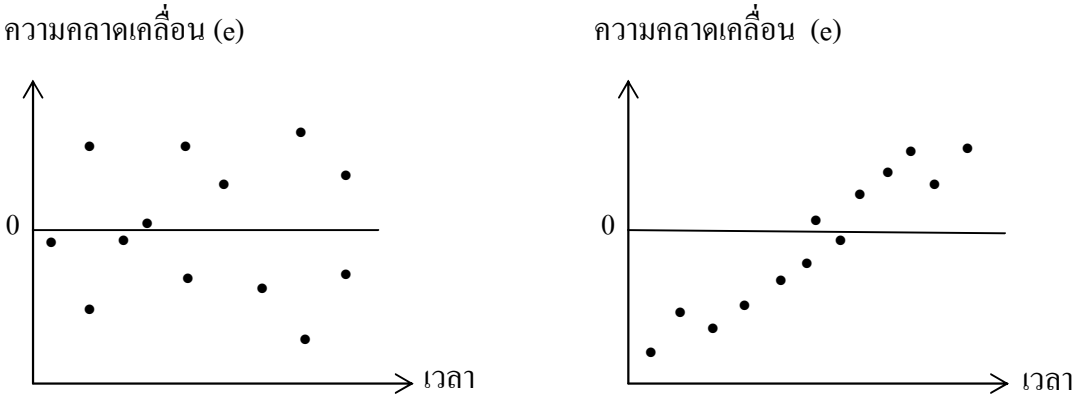
6.1 การตรวจสอบความเป็นอิสระกัน

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 วิ่งเข้าใกล้ 0 แสดงว่าความคลาดเคลื่อนแต่ละตัวยิ่งมีความสัมพันธ์กันอย่างมากในเชิงบวก ถ้าค่ามากกว่า 2.5 แสดงว่าความคลาดเคลื่อนแต่ละตัวมีความสัมพันธ์ในเชิงลบ ถ้ายิ่งเข้าใกล้ 4 แสดงว่ายิ่งมีความสัมพันธ์กันอย่างมากในเชิงลบ

(2) ใช้วิธีการพล็อตกราฟของความคลาดเคลื่อนกับเวลา ถ้าพบว่าไม่มีรูปแบบที่แน่นอน คือ การกระจายของจุดไม่เป็นระบบ แสดงว่าไม่มีปัญหาที่เรียกว่า Autocorrelation หรือ Serial

correlation แต่ถ้าพบว่าการกระจายของจุดมีรูปแบบก็แสดงว่าความคลาดเคลื่อนแต่ละตัวมีความสัมพันธ์กันดังภาพ



ก. ความคลาดเคลื่อนแต่ละตัวเป็นอิสระกัน

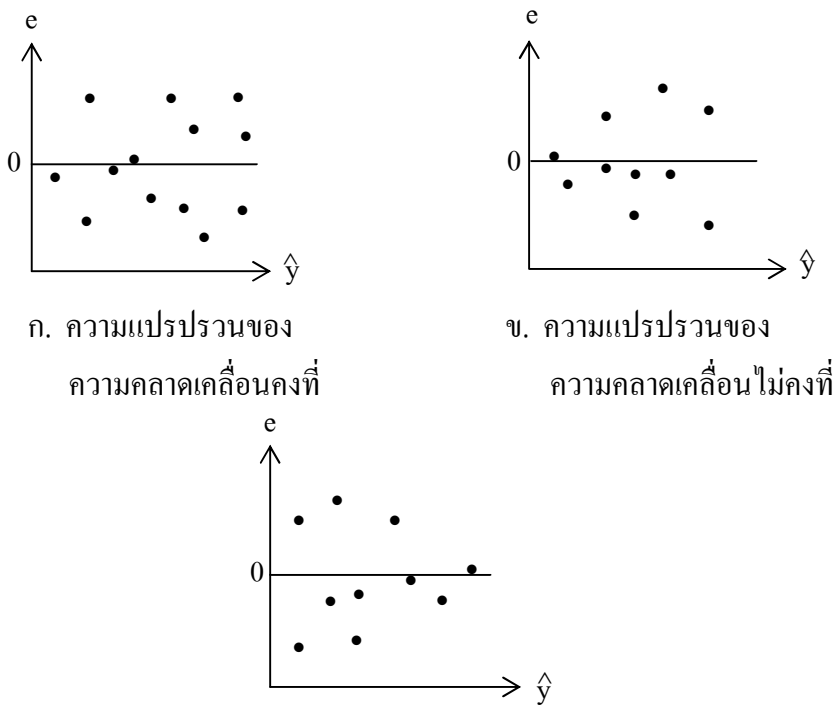
ข. ความคลาดเคลื่อนแต่ละตัวมีความสัมพันธ์กันเชิงบวก

ภาพที่ 10.1

6.2 การตรวจสอบเกี่ยวกับความแปรปรวน

การตรวจสอบเกี่ยวกับความแปรปรวนของความคลาดเคลื่อนโดยวิธีการพล็อตกราฟระหว่างความคลาดเคลื่อน (e_i) กับค่าทำนายของตัวแปรตาม ($\hat{y}_i$) โดยใช้คำสั่ง plot ในหน้าต่างของคำสั่ง Linear Regression ซึ่งในโปรแกรม SPSS จะเป็นการพล็อตระหว่าง ZRESID กับ ZPRED

ถ้าพบว่าค่า e_i กระจายอยู่รอบ ๆ ค่า O ไม่ว่าค่า $\hat{y}$ จะเปลี่ยนแปลงไปก็ตามแสดงว่าความแปรปรวนของความคลาดเคลื่อนคงที่ แต่ถ้าพบว่าค่า e_i มีการกระจายอย่างมีรูปแบบ แสดงว่าความแปรปรวนของความคลาดเคลื่อนไม่คงที่ในทุกค่าของ X นั่นคือเกิดปัญหา heterogeneity การแก้ปัญหาความแปรปรวนของความคลาดเคลื่อนไม่คงที่อาจใช้วิธีการแปลงข้อมูลเช่นแปลงค่าของข้อมูลให้อยู่ในรูปของ log เป็นต้น



ก. ความแปรปรวนของความคลาดเคลื่อนคงที่

ข. ความแปรปรวนของความคลาดเคลื่อนไม่คงที่

ค. ความแปรปรวนของความคลาดเคลื่อนไม่คงที่

ภาพที่ 10.2

7. การแปลงข้อมูล (transformation)

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

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

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

เมื่อ Y' คือ ตัวแปร Y ที่เปลี่ยนรูปแบบแล้ว

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

$$Y' = Y^2$$

การแปลงข้อมูลของตัวแปรตาม Y อาจพิจารณาได้จากแผนภาพการกระจายของข้อมูลระหว่างตัวแปรตาม Y บนแกน Y และตัวแปรอิสระ X บนแกน X ถ้าความสัมพันธ์ไม่เป็นรูปเชิงเส้นตรง ก็อาจเปลี่ยนรูปแบบของตัวแปรตาม Y หรือตัวแปรอิสระ X หรืออาจเปลี่ยนทั้งตัวแปรตาม Y และตัวแปรอิสระ X พร้อมกันก็ได้ เพื่อทำให้ความสัมพันธ์ระหว่างตัวแปรอิสระ X_i และตัวแปรตาม Y อยู่ในรูปเชิงเส้นตรง นอกจากนี้การเปลี่ยนรูปแบบของตัวแปรตาม Y ยังสามารถพิจารณาจากค่าของความคลาดเคลื่อนและความแปรปรวนของความคลาดเคลื่อนได้ด้วย รายละเอียดเรื่องนี้สามารถดูได้จากตำราอื่น ๆ

8. ตัวอย่างการพัฒนาตัวแบบการถดถอย

ตัวอย่างเช่น การศึกษาเรื่องการผ่าตัดตับ 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) ข้อมูลเกี่ยวกับตัวแปรอิสระและตัวแปรตามแสดงดังตารางที่ 10.2 และเนื่องจากการแจกแจงของตัวแปรตามเป็นแบบเบ้ขวา จึงทำการแปลงข้อมูลเวลารอดชีวิต $Y' = \log_{10} Y$ ข้อมูลอยู่ในแฟ้มข้อมูล Regress2.sav

ตารางที่ 10.2 ข้อมูลเกี่ยวกับตัวแปรอิสระและตัวแปรตามของตัวอย่างการผ่าตัดตับ

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
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

(ต่อ)

Case Number	Blood Clotting Score X_1	Prognostic Index X_2	Enzyme Function Test X_3	Liver Function Test X_4	Survival Time Y	$Y' = \log10\ Y$
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
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.

8.1 ขั้นตอนการพัฒนาตัวแบบการถดถอย

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

(2) ตรวจสอบว่าตัวแปรอิสระ X_i มีความสัมพันธ์เชิงเส้นกับตัวแปรตาม Y Prime หรือไม่ โดยใช้คำสั่ง Graphs, Scatter … จะได้หน้าต่าง Scatterplot เลือก Matrix แล้ว Define

Matrix Variables โดยคลิกเลือกตัวแปรอิสระ X_i ทุกตัว และตัวแปรตาม Y Prime ให้เข้าไปอยู่ในช่อง Matrix Variables:

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

(4) หาสมการถดถอยที่เหมาะสม ในที่นี้จะอธิบายเฉพาะวิธี Stepwise โดยใช้คำสั่ง Analyze, Regression, Linear … จะได้หน้าต่าง Linear Regression เลือกตัวแปรตาม Y Prime 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 (One-way ANOVA) ในการทดสอบสมมติฐานเกี่ยวกับ $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$ คู่กับ $H_1 : \beta_i \neq 0$ อย่างน้อย 1 ค่า ใช้ทดสอบความเหมาะสมของตัวแบบการถดถอย
- เลือกคำสั่งย่อย Part and partial correlations เพื่อแสดงระดับความสัมพันธ์ระหว่างตัวแปรอิสระ X_i แต่ละตัวกับตัวแปรตาม Y Prime เมื่อไม่ได้ควบคุมอิทธิพลของตัวแปรอิสระตัวอื่น ๆ เรียกว่า Zero-order Correlations ส่วนค่า Partial Correlations เป็นค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนของตัวแปรอิสระ X_i ตัวหนึ่งกับตัวแปรตาม Y Prime 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 ในการทำนายค่าของตัวแปรตามมีความน่าเชื่อถือ จากกรอบคำสั่ง Linear Regression : Statistics เลือกคำสั่ง Collinearity diagnostics เพื่อดูค่าสถิติ tolerance, VIF, และ Condition Index

(6) ตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยเกี่ยวกับความคลาดเคลื่อนคือความคลาดเคลื่อน (e_j) แต่ละตัวมีความเป็นอิสระกันและมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับศูนย์ และมีความแปรปรวนคงที่ทุกค่าของ X เท่ากับ σ^2

- ตรวจสอบว่าค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติหรือไม่ จากกรอบคำสั่ง Linear Regression : Plots ในส่วนของคำสั่ง Standardized Residuals Plots โดยพล็อตกราฟ Histogram และ Normal Probability Plot ของคะแนนมาตรฐานของเศษตกค้าง (Standardized Residuals)

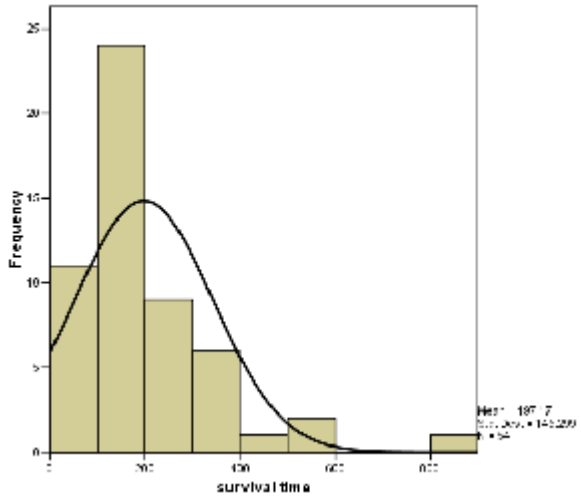
- ตรวจสอบว่าค่าความคลาดเคลื่อนแต่ละค่าเป็นอิสระกันหรือไม่ จากกรอบคำสั่ง Linear Regression : Statistics ในส่วนของคำสั่ง Residuals เลือก Durbin-Watson เพื่อดูค่าสถิติ Durbin-Watson ซึ่งเรามักใช้กับข้อมูลที่เป็นอนุกรมเวลา

- ตรวจสอบว่าความแปรปรวนของค่าความคลาดเคลื่อนคงที่หรือไม่ จากกรอบคำสั่ง Linear Regression : Plots เพื่อพล็อตกราฟระหว่างค่าคะแนนมาตรฐานของค่าพยากรณ์ (Standardized Predicted Value หรือ ZPRED) บนแกน X กับคะแนนมาตรฐานของค่าเศษตกค้าง (Standardized Residuals หรือ ZRESID) บนแกน Y

8.2 ผลลัพธ์การพัฒนาตัวแบบการถดถอย

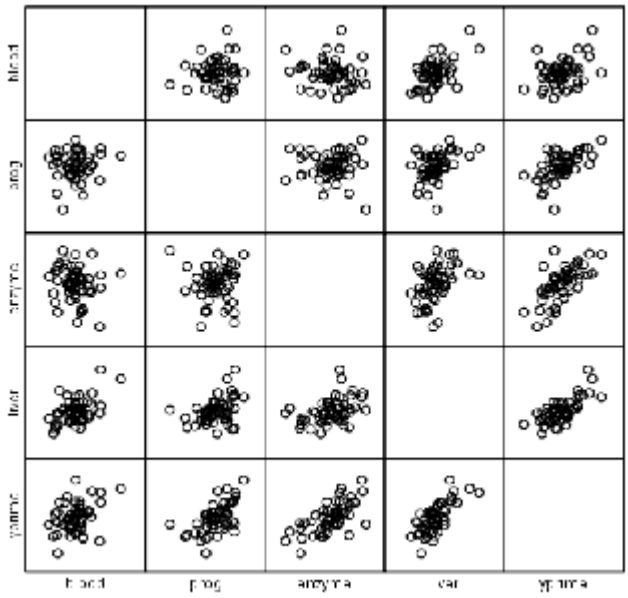
จากตัวอย่างการศึกษาเรื่องการผ่าตัดตับ มีวัตถุประสงค์เพื่อหารูปแบบความสัมพันธ์ของตัวแปรอิสระ blood, prog, enzyme, liver กับตัวแปรตามเวลารอดชีวิต (Survival) เพื่อนำไปพยากรณ์เวลารอดชีวิตของคนไข้ต่อไป ดำเนินการพัฒนาตัวแบบการถดถอยตามขั้นตอนข้างต้นได้ผลลัพธ์ดังนี้

(1) กราฟแสดงการแจกแจงของตัวแปรตาม survival ดังภาพที่ 10.3 พบว่าตัวแปรตาม survival มีการแจกแจงไม่เป็นปกติคือเป็นแบบแบ้ขวา ดังนั้นจึงทำการเปลี่ยนรูปแบบของข้อมูล survival ให้มีรูปแบบใหม่คือ yprime = $\log_{10}$ survival



ภาพที่ 10.3 กราฟฮีสโตแกรมของตัวแปรตาม (survival)

(2) กราฟแสดงความสัมพันธ์ระหว่างตัวแปรต่าง ๆ ดังภาพที่ 10.4 เพื่อดูว่าความสัมพันธ์ระหว่างตัวแปรตาม (yprime) กับตัวแปรอิสระ blood, prog, enzyme, และ liver มีลักษณะเป็นเชิงเส้นหรือไม่



ภาพที่ 10.4 กราฟแสดงความสัมพันธ์ระหว่างตัวแปรทั้งหมดเป็นรายคู่

(3) ตารางแสดงค่าความสัมพันธ์และผลการทดสอบความสัมพันธ์ระหว่างตัวแปรทั้งหมดเป็นรายคู่ โดยไม่มีการควบคุมอิทธิพลของตัวแปรตัวอื่น ๆ ดูจากตาราง Correlations ดังภาพที่ 10.5

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 10.5 ตารางแสดงค่าความสัมพันธ์และผลการทดสอบความสัมพันธ์ระหว่างตัวแปรทั้งหมดเป็นรายคู่

(4) จากขั้นตอนที่ 4, 5 และขั้นตอนที่ 6 เฉพาะการตรวจสอบความเป็นอิสระของความคลาดเคลื่อน จะได้ผลลัพธ์เป็นตารางต่าง ๆ ดังต่อไปนี้ ดังภาพที่ 10.6

- 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-remove >= .100).
2	enzyme	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
3	prog	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
4	blood	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
5	.	liver	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .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]

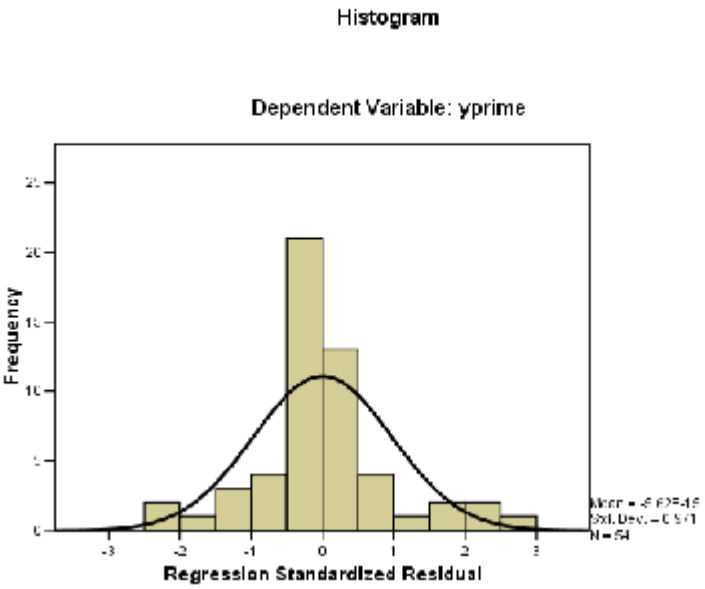
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	1.604248	2.863085	2.206143	.2699701	54
Residual	-.1020467	.1383532	.0000000	.0455336	54
Std. Predicted Value	-2.229	2.433	.000	1.000	54
Std. Residual	-2.177	2.951	.000	.971	54

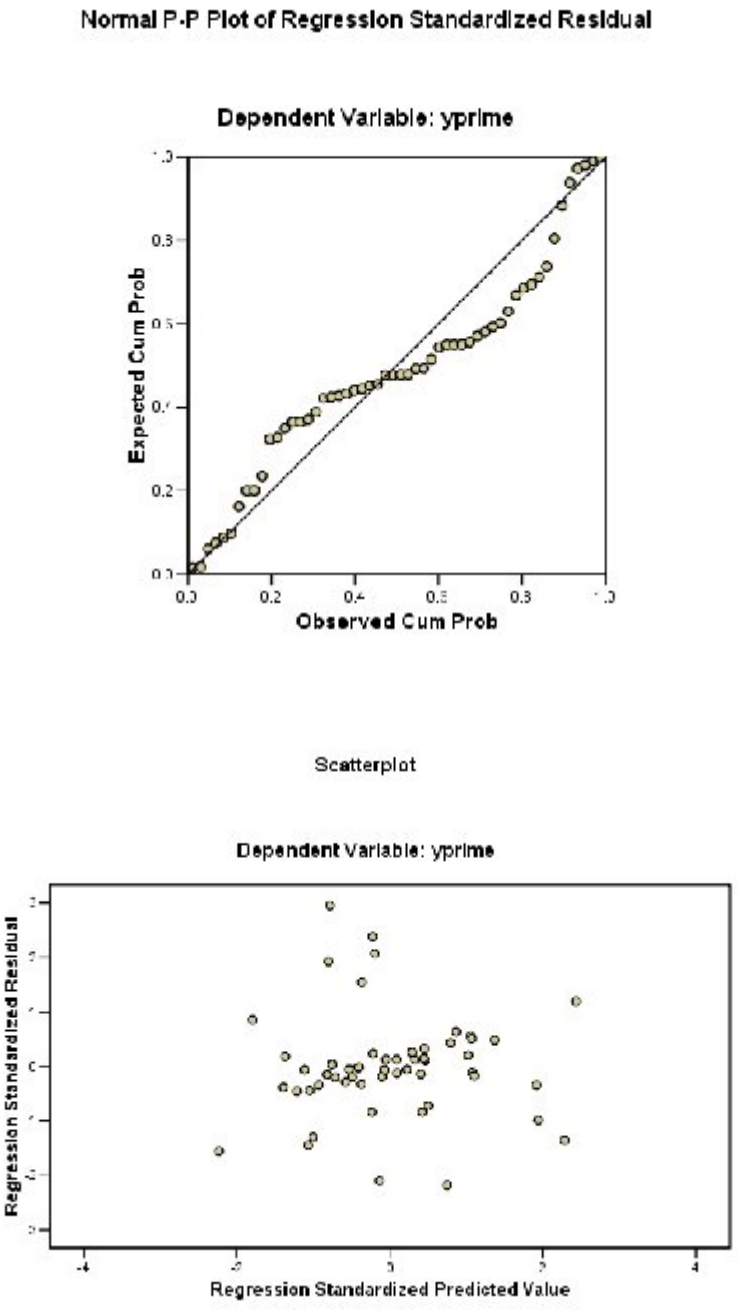
a. Dependent Variable: yprime

ภาพที่ 10.6 ผลลัพธ์จากขั้นตอนการพัฒนาตัวแบบการถดถอยในขั้นตอนที่ 4, 5 และขั้นตอนที่ 6 เฉพาะการตรวจสอบความเป็นอิสระของความคลาดเคลื่อน

(5) กราฟแสดงการแจกแจงของคะแนนมาตรฐานของความคลาดเคลื่อนแบบฮีสโตแกรมและ Normal P-P Plot และแผนภาพการกระจายระหว่าง Standardized Residual กับ Standardized Predicted Value ดังภาพที่ 10.7

Charts





ภาพที่ 10.7 ฮีสโตแกรม, Normal P-P Plot ของคะแนนมาตรฐานของความคลาดเคลื่อนและ Scatterplot ระหว่างคะแนนมาตรฐานของความคลาดเคลื่อนกับคะแนนมาตรฐานของค่าทำนาย Y

8.3 การสรุปผลการพัฒนาตัวแบบการถดถอย

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

(1) ดูผลของค่าความสัมพันธ์ระหว่างตัวแปรทั้งหมดเป็นรายคู่โดยไม่มีการควบคุมอิทธิพลของตัวแปรตัวอื่น ๆ ได้จากตาราง Correlations เขียนใหม่เป็นตารางเมตริกความสัมพันธ์ระหว่างตัวแปรทั้งหมดได้ดังนี้

ตารางที่ 10.3 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) ดูผลการทดสอบสมมติฐานของตัวแบบการถดถอย $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ คู่กับ H_1 : มี β_i 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) ดูค่าสัมประสิทธิ์ต่าง 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) ดูค่าสัมประสิทธิ์การถดถอยของตัวแบบการถดถอย 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) ดูค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนของตัวแปรทำนายแต่ละตัวได้จากตาราง Coefficients บนบรรทัดของตัวแบบที่ 5 ในคอลัมน์ Correlations สำหรับ Partial ได้ค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนของตัวแปรตาม (yprime) กับตัวแปรทำนายตัวใดตัวหนึ่งเมื่อควบคุมตัวแปรทำนายตัวอื่น ๆ ได้ผลคือ

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

(6) ผลการตรวจสอบความเป็นอิสระของตัวแปรทำนายทั้ง 3 ตัว โดยพิจารณาจากตาราง Coefficients ในตัวแบบที่ 5 ในคอลัมน์ Collinearity Statistics ซึ่งมีค่า 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

และในตาราง Excluded Variables ในคอลัมน์ Collinearity Statistics ที่มีค่า Minimum Tolerance ของตัวแปร liver เท่ากับ 0.391 ซึ่งไม่อยู่ในตัวแบบการถดถอย บนบรรทัดของตัวแบบที่ 5 หมายความว่า ตัวแปร liver มีความสัมพันธ์กับตัวแปรอิสระตัวอื่น ๆ มาก

(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 พบว่าจุดมีการกระจัดกระจายอย่างไม่เป็นรูปแบบ แสดงว่าความแปรปรวนของค่าความคลาดเคลื่อนค่อนข้างคงที่

8.4 การแปลความหมายการพัฒนาตัวแบบการถดถอย

จากการสรุปผลการพัฒนาตัวแบบการถดถอยด้วยวิธี Stepwise ของตัวอย่างการศึกษาเรื่องการผ่าตัดตับ สามารถแปลความหมายได้คือ

(1) จากผลของค่าความสัมพันธ์ระหว่างตัวแปรทั้งหมดเป็นรายคู่โดยไม่มีการควบคุมอิทธิพลของตัวแปรตัวอื่น ๆ ทำการทดสอบสมมติฐานเกี่ยวกับ $H_0 : \rho = 0$ คู่กับ $H_1 : \rho \neq 0$ พบว่าตัวแปรอิสระทุกตัวมีความสัมพันธ์กับตัวแปรตาม (yprime) โดยเฉพาะ liver มีค่ามากที่สุด และ blood มีค่าน้อยที่สุด และยังแสดงความสัมพันธ์ระหว่างตัวแปรอิสระด้วยกันเองด้วย โดยเฉพาะความสัมพันธ์ระหว่าง liver กับ blood และ prog และ enzyme มีค่าพอสมควร

(2) ผลการพัฒนาตัวแบบการถดถอย ได้ตัวแบบการถดถอยคือ

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

- ค่า 0.10 ของตัวแปร enzyme หมายถึง ค่า yprime ที่เพิ่มขึ้นเมื่อ enzyme เพิ่มขึ้น 1 หน่วย โดย prog และ blood คงที่อยู่ในตัวแบบ
- ค่า 0.009 ของตัวแปร prog หมายถึง ค่า yprime ที่เพิ่มขึ้นเมื่อ prog เพิ่มขึ้น 1 หน่วย โดย enzyme และ blood คงที่อยู่ในตัวแบบ
- ค่า 0.069 ของตัวแปร blood หมายถึง ค่า yprime ที่เพิ่มขึ้นเมื่อ blood เพิ่มขึ้น 1 หน่วย โดย enzyme และ prog คงที่อยู่ในตัวแบบ

การเปรียบเทียบอิทธิพลหรือความสำคัญของตัวแปรทำนายทั้ง 3 ตัวนี้ที่มีต่อตัวแปรตาม yprime สามารถดูได้จากค่า Standardized Coefficients Beta ซึ่งคำนวณได้จาก

$$Beta_i = b_i \frac{S_i}{S_y}$$

เมื่อ b_i คือ สัมประสิทธิ์การถดถอยบางส่วน (partial – regression coefficient) ของตัวแปรทำนายที่ i

S_i คือ ค่าเบี่ยงเบนมาตรฐานของตัวแปรทำนายที่ i

S_y คือ ค่าเบี่ยงเบนมาตรฐานของตัวแปรตาม Y

และสามารถดูค่า Standardized Coefficients Beta ได้จากตาราง Coefficients ดังนี้

ตัวแปร	Standardized Coefficients 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 สรุปผลการทดสอบได้ว่าปฏิเสธ 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 สรุปผลการทดสอบได้ว่าปฏิเสธ H_0

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

ได้ค่าสถิติทดสอบ t = 24.296 ค่า Sig. = .000 สรุปผลการทดสอบได้ว่าปฏิเสธ H_0

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

ได้ค่าสถิติทดสอบ t = 16.974 ค่า Sig. = .000 สรุปผลการทดสอบได้ว่าปฏิเสธ 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

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

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

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 Change ของตัวแบบการถดถอย เท่ากับ .000 หมายความว่าการตัดตัวแปร liver ออกจากสมการถดถอยไม่ทำให้ค่า R^2 เปลี่ยน

- ค่า R คือ สัมประสิทธิ์สหสัมพันธ์เชิงพหุ (Multiple Coefficient of Correlation) ได้จากการถอดรากที่สองของสัมประสิทธิ์การตัดสินใจเชิงพหุ มีค่าอยู่ระหว่าง 0 และ 1 หมายถึงค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรตาม Y กับตัวแปรอิสระ $X_1 , X_2 , \dots , X_k$ 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 เวลารอดชีวิตมีความสัมพันธ์มากกับตัวแปรอิสระ 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 $H_0 : \beta_{liver} = 0$ คู่กับ $H_1 : \beta_{liver} \neq 0$ จะมีค่าสถิติทดสอบ $F_{1,51} = .039$ และมีค่า Sig. F Change = .844 ซึ่งมากกว่าระดับนัยสำคัญที่กำหนดในการเลือกตัวแปรออก (α = .10) จึงสรุปผลการทดสอบว่ายอมรับ H_0 นั่นคือตัดตัวแปรอิสระ liver ออกจากสมการถดถอย

- ค่า Std. Error of the Estimate คือ ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า yprime 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

แบบฝึกหัดบทที่ 10

1. ในการศึกษาเกี่ยวกับปัจจัยที่มีอิทธิพลต่อรายได้ของผู้ขายเร่บนท้องถนน (World Development, Feb. 1998 อ้างถึงใน Mendenhall, w. and Sincich, T. 2003) ทำการศึกษาโดยการสัมภาษณ์ผู้ค้าเร่บนท้องถนนในเมืองหนึ่งของแม็กซิโก โดยเก็บข้อมูลเกี่ยวกับอายุ ชั่วโมงทำงานต่อวัน รายได้ต่อปี และระดับการศึกษา ตัดข้อมูลมาเพียงบางส่วนดังตาราง

ตาราง ข้อมูลรายได้ต่อปี อายุ และชั่วโมงทำงานต่อวันของผู้ขายเร่บนท้องถนน

จำนวนผู้ขายเร่	รายได้ต่อปี (ดอลล์ล่า)	อายุ	จำนวนชั่วโมงทั้งหมดต่อวัน
21	2841	29	12
53	1876	21	8
60	2934	62	10
184	1552	18	10
263	3065	40	11
281	3670	50	11
354	2005	65	5
401	3215	44	8
515	1930	17	8
633	2010	70	6
677	3111	20	9
710	2882	29	9
800	1683	15	5
914	1817	14	7
997	4066	33	12

แหล่งที่มา : Adapted from Smith, P.A., and Metzger, M.R. “The return to education : Street vendors in Mexico.” World Development, Vol. 26, No.2, Feb. 1998, pp. 289-296. อ้างถึงใน Mendenhall, W. and Sincich, T. 2003.

ก. จงเขียนตัวแบบการถดถอยของรายได้ต่อปี (y) ที่ถดถอยบนอายุ (x_1) และจำนวนชั่วโมงทำงานต่อวัน (x_2)

ข. จงตรวจสอบว่าตัวแปร x_1 และ x_2 มีความสัมพันธ์เชิงเส้นกับ y หรือไม่

ค. จงใช้ข้อมูลในตารางเพื่อหาสมการถดถอยของกลุ่มตัวอย่างผู้ขายเร่ตามตัวแบบในข้อ ก. และแปลความหมาย

ง. จงทดสอบความเหมาะสมของเส้นถดถอยเชิงเส้นตรงที่ได้จากข้อ ค.

จ. จงตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยเกี่ยวกับความเป็นอิสระของตัวแปร x_1 และ x_2

ฉ. จงตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยเกี่ยวกับความคลาดเคลื่อน

2. นักพฤกษศาสตร์ของมหาวิทยาลัยโตรอนโตได้ทำการศึกษาเกี่ยวกับนิสัยการกินอาหารของลูกห่าน (Journal of Applied Ecology, Vol. 32, 1995 อ้างถึงใน Mendenhall, W. and Sincich, T. 2003) ทำการทดลองโดยการให้ลูกห่าน 42 ตัว อดอาหารจนไม่มีอาหารเหลืออยู่ในท้องเลย หลังจากนั้นจึงให้กินอาหารเป็นเวลา 6 ชั่วโมง โดยให้อาหารเป็นพืช หรือข้าวเปือก หลังจากนั้น 2.5 ชั่วโมง ทำการบันทึกข้อมูลเกี่ยวกับน้ำหนักของลูกห่านที่เปลี่ยนไปจากตอนเริ่มการทดลอง (คิดเป็นเปอร์เซ็นต์) และประสิทธิภาพการย่อยอาหาร (คิดเป็นเปอร์เซ็นต์) และปริมาณของ acid-detergent fibre (คิดเป็นเปอร์เซ็นต์) ได้ข้อมูลดังตาราง

ตาราง ข้อมูลอาหาร น้ำหนักที่เปลี่ยน ประสิทธิภาพการย่อย และปริมาณของ acid-detergent fibre

ลูกห่าน	อาหาร	น้ำหนักที่เปลี่ยน (%)	ประสิทธิภาพการย่อย (%)	acid-detergent fibre (%)
1	พืช	-6	0	28.5
2	พืช	-5	2.5	27.5
3	พืช	-4.5	5	27.5
4	พืช	0	0	32.5
5	พืช	2	0	32
6	พืช	3.5	1	30
7	พืช	-2	2.5	34
8	พืช	-2.5	10	36.5
9	พืช	-3.5	20	28.5
10	พืช	-2.5	12.5	29

(ต่อ)

ลูกห่าน	อาหาร	น้ำหนักที่เปลี่ยน (%)	ประสิทธิภาพการย่อย (%)	acid-detergent fibre(%)
11	พืช	-3	28	28
12	พืช	-8.5	30	28
13	พืช	-3.5	18	30
14	พืช	-3	15	31
15	พืช	-2.5	17.5	30
16	พืช	-.5	18	22
17	พืช	0	23	22.5
18	พืช	1	20	24
19	พืช	2	15	23
20	พืช	6	31	21
21	พืช	2	15	24
22	พืช	2	21	23
23	พืช	2.5	30	22.5
24	พืช	2.5	33	23
25	พืช	0	27.5	30.5
26	พืช	.5	29	31
27	พืช	-1	32.5	30
28	พืช	-3	42	24
29	พืช	-2.5	39	25
30	พืช	-2	35.5	25
31	พืช	.5	39	20
32	พืช	5.5	39	18.5
33	พืช	7.5	50	15
34	ข้าวเปือก	0	62.5	8
35	ข้าวเปือก	0	63	8
36	ข้าวเปือก	2	69	7
37	ข้าวเปือก	8	42.5	7.5
38	ข้าวเปือก	9	59	8.5
39	ข้าวเปือก	12	52.5	8
40	ข้าวเปือก	8.5	75	6
41	ข้าวเปือก	10.5	72.5	6.5
42	ข้าวเปือก	14	69	7

แหล่งที่มา : Gadallah, F.L., and Jefferies, R.L. “Forage quality in brood rearing areas of the lesser snow goose and the growth of captive goslings.” *Journal of Applied Biology*, Vol. 32, No.2, 1995, pp.281-282 (adapted from Figures 2 and 3). อ้างถึงใน Mendenhall, W. and Sincich, T. 2003.

ก. จงเขียนตัวแบบการถดถอยของน้ำหนักของลูกห่านที่เปลี่ยนแปลง (y) ถดถอยบนประสิทธิภาพการย่อย (x_1) และปริมาณ acid-detergent fibre (x_2)

ข. จงตรวจสอบว่าตัวแปร x_1 และ x_2 มีความสัมพันธ์เชิงเส้นกับ y หรือไม่

ค. จงใช้ข้อมูลในตารางเพื่อหาสมการถดถอยของกลุ่มตัวอย่างลูกห่าน และแปลความหมาย

ง. จงทดสอบความเหมาะสมของเส้นถดถอยเชิงเส้นตรงที่ได้จากข้อ ค.

จ. จงตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยเกี่ยวกับความเป็นอิสระของตัวแปร x_1 และ x_2

ฉ. จงตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยเกี่ยวกับความคลาดเคลื่อน