

## บทที่ 7

### การวิเคราะห์การถดถอยสำหรับตัวแปรอิสระเชิงคุณภาพ

ตัวแปรอิสระที่ใช้ในการวิเคราะห์การถดถอยส่วนใหญ่เป็นตัวแปรเชิงปริมาณหรือตัวแปรที่สามารถแสดงค่าในรูปของตัวเลขที่ทำการเปรียบเทียบค่าได้ เช่น อุณหภูมิ ระยะทาง คะแนนสอบ รายได้ เป็นต้น บางครั้งนักวิจัยต้องการใช้ตัวแปรเชิงคุณภาพในการพยากรณ์ตัวแปรเชิงปริมาณเช่น การพยากรณ์คะแนนสอบของนักศึกษาแต่ละเพศ หรือการพยากรณ์รายได้โดยใช้ระดับการศึกษา (ประถมศึกษา, มัธยมศึกษา, ปริญญาตรี, ปริญญาโท และปริญญาเอก) เป็นต้น การวิเคราะห์ข้อมูลโดยใช้ตัวแปรเชิงคุณภาพเหล่านี้ไม่สามารถวิเคราะห์โดยใช้รูปแบบที่ผ่านมาได้

#### 7.1 ตัวแปรอิสระเชิงคุณภาพ 1 ตัวที่มีสองกลุ่มย่อย

การวิเคราะห์ข้อมูลเชิงคุณภาพจำเป็นต้องมีการสร้างตัวแปรหุ่น (dummy variable) ขึ้นเพื่อช่วยในการสร้างสมการถดถอยโดยการแปลงข้อมูลเหล่านี้ให้เป็นตัวเลขเสียก่อน โดยตัวเลขแต่ละค่าแสดงระดับหรือกลุ่มที่แตกต่างกันของตัวแปรเชิงคุณภาพนั้นๆ เช่น ให้  $X_2$  แทนตัวแปรเพศโดยเพศชายมีค่าเป็น 0 และหญิงมีค่าเป็น 1 เป็นต้น การเลือกตัวเลขให้กับแต่ละกลุ่มนั้นจะกำหนดอย่างไรก็ได้แต่เพื่อความง่ายและสะดวกในการตีความจึงใช้ 0 กับ 1 ในกรณีนี้สมการถดถอยเชิงเส้นอย่างง่ายสามารถเขียนได้ดังนี้

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \quad (7.1)$$

หากพิจารณาเพศชาย ( $X_2 = 0$ ) แล้วสมการคือ

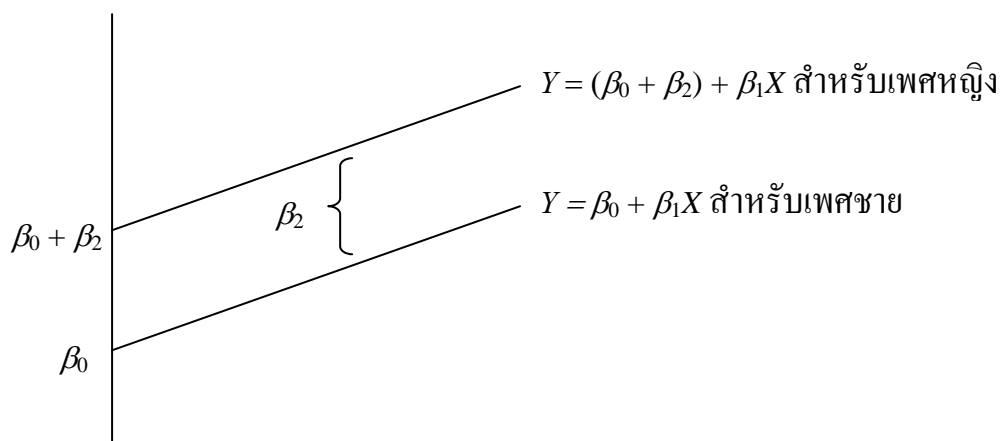
$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2(0) + \varepsilon \\ &= \beta_0 + \beta_1 X_1 + \varepsilon \end{aligned}$$

ดังนั้นความสัมพันธ์ระหว่าง  $Y$  กับ  $X_1$  จะเป็นเส้นตรงที่มีความชันเท่ากับ  $\beta_1$  และจุดตัดแกนเท่ากับ

$\beta_0$  และในกรณีเพศหญิง ( $X_2 = 1$ ) แล้วสมการคือ

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2(1) + \varepsilon \\ Y &= (\beta_0 + \beta_2) + \beta_1 X_1 + \varepsilon \end{aligned}$$

ดังนั้นความสัมพันธ์ระหว่าง  $Y$  กับ  $X_1$  จะเป็นเส้นตรงที่มีความชันเท่ากับ  $\beta_1$  และจุดตัดแกนเท่ากับ  $\beta_0 + \beta_2$  จะเห็นได้ว่าจุดตัดแกนของทั้งสองเพศนั้นแตกต่างกันเท่ากับ  $\beta_2$  ดังภาพที่ 7.1 โดย  $\beta_2$  เป็นค่าความแตกต่างระหว่างค่าของ  $Y$  ที่มีผลมาจากเพศที่แตกต่างกัน



ภาพที่ 7.1 เส้นถดถอยสำหรับเพศทั้งสอง

หากใช้จำนวนตัวแปรหุ่นเท่ากับจำนวนระดับหรือกลุ่มของตัวแปรเชิงคุณภาพแล้วคอลัมน์ในเมตริกซ์  $X'X$  จะไม่เป็นอิสระเชิงเส้นต่อกัน (linearly independent) และส่งผลให้ไม่สามารถหาเมตริกซ์ผกผันของ  $X'X$  ได้ ผลที่ตามมาคือค่าประมาณของพารามิเตอร์ที่ได้จะมีมากกว่า 1 ชุด ดังนั้นจึงจำเป็นต้องใช้จำนวนตัวแปรหุ่นน้อยกว่าจำนวนระดับหรือกลุ่มอยู่ 1 เพื่อให้ค่าประมาณของพารามิเตอร์มีชุดเดียว การสร้างสมการถดถอยจะสร้างโดยวิธีเดียวกันกับการสร้างสมการถดถอยโดยไม่มีตัวแปรหุ่น

ตัวอย่างที่ 7.1 อาจารย์ผู้สอนวิชาคณิตศาสตร์ต้องการศึกษาความสัมพันธ์ระหว่างคะแนนสอบวิชาคณิตศาสตร์ ( $Y$ ) กับเกรดสะสม ( $X$ ) และเพศ โดยมีข้อมูลดังนี้

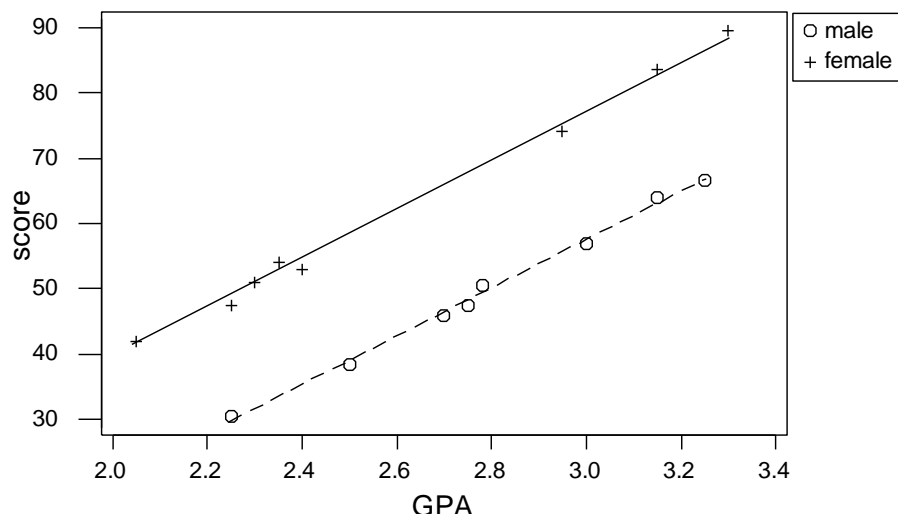
เกรด	เพศ	คะแนนสอบ
2.25	0	30.5
3.25	0	66.5
2.70	0	46.0
2.78	0	50.5

เกรด	เพศ	คะแนนสอบ
3.00	0	57.0
2.50	0	38.5
2.75	0	47.5
3.15	0	64.0
2.30	1	51.0
3.50	1	89.5
2.05	1	42.0
2.35	1	54.0
2.95	1	74.0
3.35	1	83.5
2.40	1	53.0
2.25	1	47.5

### วิธีทำ

เนื่องจากเพศเป็นตัวแปรเชิงคุณภาพที่แบ่งได้เป็น 2 กลุ่มคือ เพศชายและหญิงดังนั้นในการวิเคราะห์จึงกำหนดค่าตัวแปรเพศเป็น 0 แทนเพศชายและ 1 แทนเพศหญิง

เมื่อตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามโดยใช้แผนภาพกระจายระหว่างเกรดสะสม (GPA) กับคะแนนสอบ (score) แยกตามเพศ (gender) พบว่าเมื่อนักเรียนที่มีเกรดสะสมต่ำจะมีคะแนนสอบต่ำและนักเรียนที่มีเกรดสะสมสูงจะมีคะแนนสอบสูงด้วยเช่นกัน และยังพบอีกว่าเพศหญิงมีคะแนนสอบที่สูงกว่าเพศชายดังภาพข้างล่าง



จากข้อมูลสามารถสร้างตัวแบบอยู่ในรูปของ

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

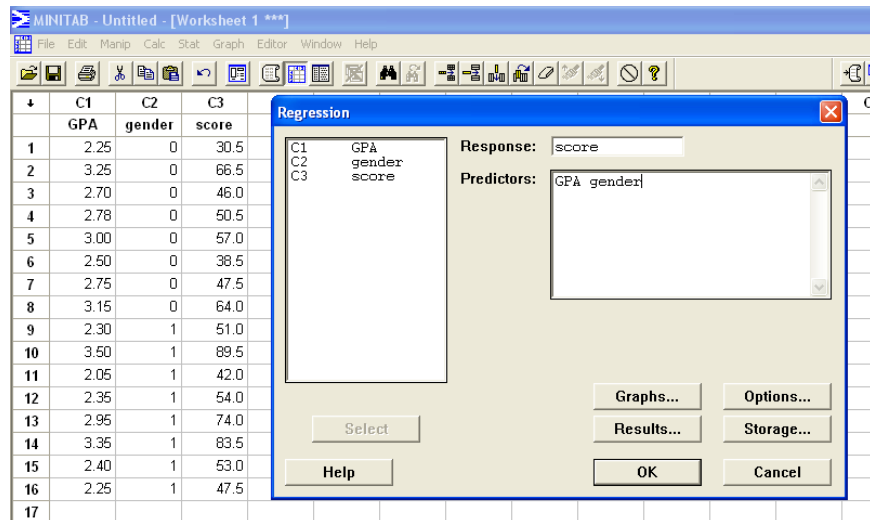
และสามารถจัดข้อมูลอยู่ในรูปเมตริกซ์ดังนี้

$$X = \begin{bmatrix} 1 & 2.25 & 0 \\ 1 & 3.25 & 0 \\ 1 & 2.70 & 0 \\ 1 & 2.78 & 0 \\ 1 & 3.00 & 0 \\ 1 & 2.50 & 0 \\ 1 & 2.75 & 0 \\ 1 & 3.15 & 0 \\ 1 & 2.30 & 1 \\ 1 & 3.50 & 1 \\ 1 & 2.05 & 1 \\ 1 & 2.35 & 1 \\ 1 & 2.95 & 1 \\ 1 & 3.35 & 1 \\ 1 & 2.40 & 1 \\ 1 & 2.25 & 1 \end{bmatrix} \quad Y = \begin{bmatrix} 30.5 \\ 66.5 \\ 46.0 \\ 50.5 \\ 57.0 \\ 38.5 \\ 47.5 \\ 64.0 \\ 51.0 \\ 89.5 \\ 42.0 \\ 54.0 \\ 74.0 \\ 83.5 \\ 53.0 \\ 47.5 \end{bmatrix}$$

การสร้างสมการถดถอยโดยใช้เมตริกซ์ใช้วิธีเดียวกับที่ผ่านมาคือ

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{Y} = \begin{bmatrix} -55.46 \\ 37.72 \\ 19.44 \end{bmatrix}$$

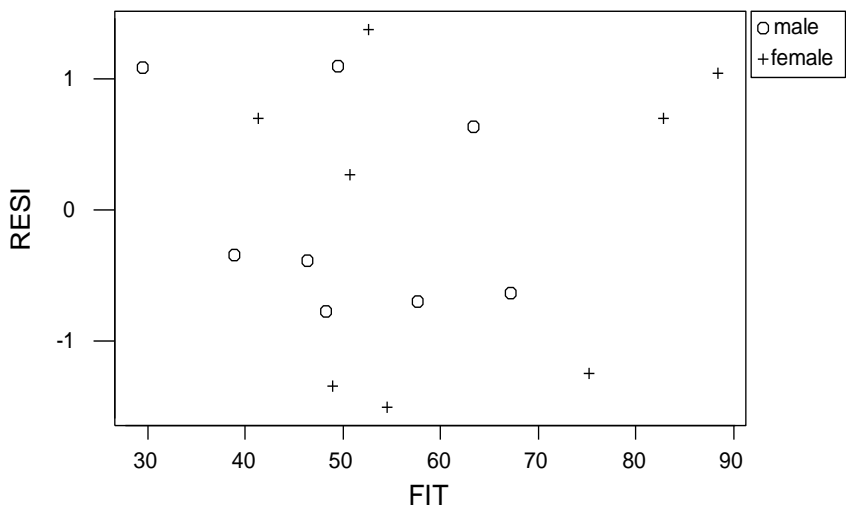
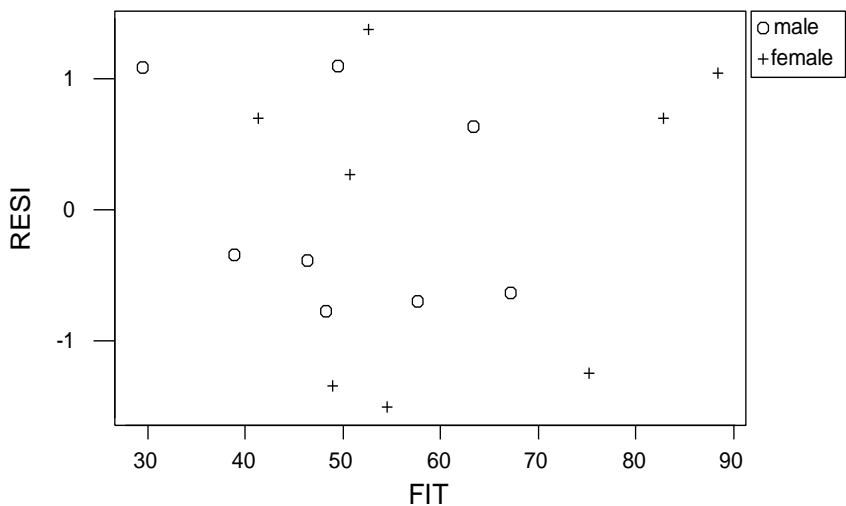
การสร้างสมการถดถอยโดยใช้โปรแกรม MINITAB สามารถทำได้โดยใช้คำสั่ง “Regression” เช่นเดียวกับการวิเคราะห์ที่ผ่านมาและระบุตัวแปรอิสระทั้งสองคือ GPA และ gender ในช่อง “Predictors:” และระบุตัวแปรตามคือ Score ในช่อง “Response:” ดังภาพข้างล่าง



จกตารางวิเคราะห์ความแปรปรวนที่ได้จากโปรแกรม MINITAB พบว่าค่าสถิติ  $F$  มีค่าเท่ากับ 1746.23 และมีค่า  $p$ -value เท่ากับ 0.000 แสดงว่าสมการถดถอยที่สร้างขึ้นสามารถใช้ในการพยากรณ์คะแนนสอบได้และมีค่าสัมประสิทธิ์การตัดสินใจ (R-Sq) เท่ากับ 99.6% แสดงว่าสมการถดถอยที่ได้สามารถพยากรณ์ได้เป็นอย่างดีและเมื่อพิจารณาความสำคัญของตัวแปรอิสระแต่ละตัวต่อตัวแปรตามพบว่าเกรดสะสม (GPA) มีค่าสถิติ  $t$  เท่ากับ 54.67 ค่า  $p$ -value เท่ากับ 0.000 และเพศ (gender) มีค่าสถิติ  $t$  เท่ากับ 35.86 ค่า  $p$ -value เท่ากับ 0.000 โดยค่า  $p$ -value ของตัวแปรทั้งสองมีค่าน้อยกว่า 0.05 ดังนั้นทั้งเกรดสะสมและเพศมีความสัมพันธ์ต่อคะแนนสอบที่ระดับนัยสำคัญ 0.05

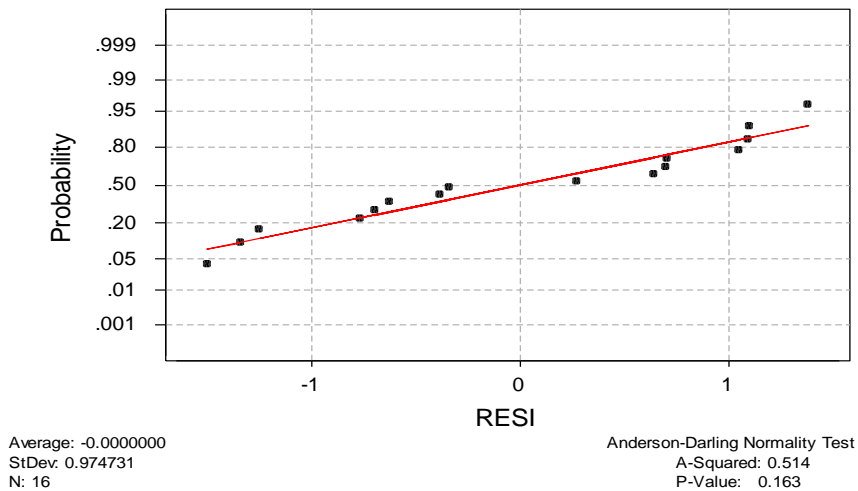
The regression equation is					
score = - 55.5 + 37.7 GPA + 19.4 gender					
Predictor	Coef	SE Coef	T	P	
Constant	-55.465	1.965	-28.22	0.000	
GPA	37.7222	0.6900	54.67	0.000	
gender	19.4359	0.5421	35.86	0.000	
S = 1.047		R-Sq = 99.6%		R-Sq(adj) = 99.6%	
Analysis of Variance					
Source	DF	SS	MS	F	P
Regression	2	3828.7	1914.3	1746.23	0.000
Residual Error	13	14.3	1.1		
Total	15	3842.9			
Source	DF	Seq SS			
GPA	1	2419.3			
gender	1	1409.4			

เมื่อพิจารณาแผนภาพกระจายระหว่างค่าพยากรณ์กับส่วนเหลือแยกตามเพศพบว่าเพศหญิงมีการกระจายตัวของค่าส่วนเหลือมากกว่าเพศชายซึ่งสอดคล้องกับแผนภาพกระจายระหว่างเกรดกับค่าส่วนเหลือ



จากแผนภาพ normal probability plot พบว่าส่วนเหลือมีการแจกแจงแบบปกติที่  $p\text{-value} = 0.163$

Normal Probability Plot



จากการวิเคราะห์โดยใช้โปรแกรม MINITAB ได้สมการถดถอยคือ

$$\hat{Y} = -55.5 + 37.7 GPA + 19.4 gender$$

โดยสามารถเขียนสมการถดถอยสำหรับเพศชายคือ

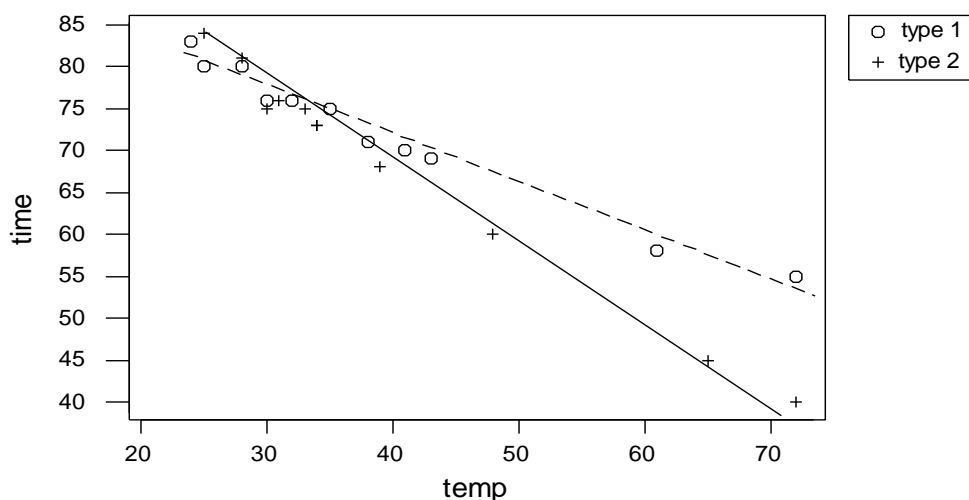
$$\begin{aligned}\hat{Y} &= -55.5 + 37.7 GPA + 19.4(0) \\ &= -55.5 + 37.7 GPA\end{aligned}$$

และสมการถดถอยสำหรับเพศหญิงคือ

$$\begin{aligned}\hat{Y} &= -55.5 + 37.7 GPA + 19.4(1) \\ &= -36.1 + 37.7 GPA\end{aligned}$$

จากสมการถดถอยพบว่านักเรียนที่มีเกรดสะสมเท่ากันแล้วเพศหญิงจะได้คะแนนสูงกว่าเพศชายอยู่ 19.4 คะแนน ทั้งนี้เมื่อพิจารณาแผนภาพกระจายระหว่างเกรดสะสมกับคะแนนสอบของทั้งสองเพศพบว่ากราฟของเพศทั้งสองมีความชันที่ไม่แตกต่างกัน

หากทั้งสองกลุ่มมีความชันที่แตกต่างกันแล้วแสดงว่าตัวแปรอิสระเชิงคุณภาพตัวนั้นกับตัวแปรอิสระเชิงปริมาณอีกตัวมีปฏิสัมพันธ์ต่อกัน (interaction) เช่น ในการทดสอบอายุการใช้งานของแบตเตอรี่ 2 ชนิดที่อุณหภูมิต่างกัน หากแผนภาพกระจายของอุณหภูมิกับอายุการใช้งานของทั้งสองชนิดมีความชันที่ไม่เท่ากันดังภาพข้างล่างแล้วแสดงว่าอุณหภูมิกับชนิดของแบตเตอรี่มีปฏิสัมพันธ์ต่อกัน โดยที่อุณหภูมิต่ำๆนั้นแบตเตอรี่ชนิดที่ 1 จะมีอายุการใช้งานที่นานกว่าชนิดที่ 2 แต่ที่อุณหภูมิสูงๆ นั้นแบตเตอรี่ชนิดที่ 2 จะมีอายุการใช้งานที่นานกว่าชนิดที่ 1 เป็นต้น



ในกรณีที่ตัวแปรทั้งสองมีปฏิสัมพันธ์กันนั้นการสร้างสมการถดถอยจำเป็นต้องเพิ่มพจน์ของปฏิสัมพันธ์ระหว่างตัวแปรทั้งสองและต้องทดสอบพารามิเตอร์ของพจน์ที่เพิ่มขึ้นนี้ โดยสามารถเขียนสมการถดถอยได้ดังนี้

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon \quad (7.2)$$

โดย  $\beta_3$  เป็นพารามิเตอร์สำหรับปฏิสัมพันธ์ระหว่างตัวแปรอิสระทั้งสอง เมื่อพิจารณาสมการถดถอยแยกตามกลุ่มของตัวแปรเชิงคุณภาพจะได้ว่าสมการถดถอยของกลุ่มที่ 1 คือ

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2(0) + \beta_3 X_1(0) + \varepsilon \\ &= \beta_0 + \beta_1 X_1 + \varepsilon \end{aligned} \quad (7.3)$$

นั่นคือสมการมีจุดตัดแกน  $Y$  ที่  $\beta_0$  และมีความชันคือ  $\beta_1$  สำหรับกลุ่มที่ 2 มีสมการถดถอยคือ

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2(1) + \beta_3 X_1(1) + \varepsilon \\ &= (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_1 + \varepsilon \end{aligned} \quad (7.4)$$

นั่นคือสมการมีจุดตัดแกน  $Y$  ที่  $\beta_0 + \beta_2$  และมีความชันคือ  $\beta_1 + \beta_3$  เนื่องจากความชันและจุดตัดแกนของทั้งสองสมการไม่เท่ากันจึงทำให้กราฟของทั้งสองกลุ่มแตกต่างกัน การทดสอบเริ่มจากการทดสอบว่าสมการถดถอยทั้งสองเหมือนกันหรือไม่โดยการตั้งสมมติฐานดังนี้

$$H_0 : \beta_2 = \beta_3 = 0$$

$$H_1 : \beta_2 \neq 0 \text{ และ/หรือ } \beta_3 \neq 0$$

ทดสอบโดยใช้สถิติ  $F$  ดังนี้



$$F = \frac{SSR(\beta_2, \beta_3 | \beta_0, \beta_1) / 2}{MSE} \quad (7.5)$$

โดยค่า  $MSE$  ได้จากสมการเต็มรูปและทำการเปรียบเทียบกับ  $F$  จากตารางที่องศาเสรีเท่ากับ 2 และ  $n - 4$  หากไม่สามารถปฏิเสธสมมติฐานหลักแสดงว่าสามารถใช้สมการถดถอยเพียงสมการเดียวในการอธิบายความสัมพันธ์ระหว่างตัวแปรอิสระเชิงปริมาณกับตัวแปรตามได้ การทดสอบว่าเส้นกราฟของทั้งสองกลุ่มมีความชันเท่ากันหรือไม่ทำได้โดยการทดสอบว่า  $\beta_3$  เท่ากับ 0 หรือไม่ โดยสามารถเขียนสมมติฐานได้ดังนี้

$$H_0 : \beta_3 = 0$$

$$H_1 : \beta_3 \neq 0$$

ทดสอบโดยใช้สถิติ  $F$  ดังนี้

$$F = \frac{SSR(\beta_3 | \beta_0, \beta_1, \beta_2) / 1}{MSE} \quad (7.6)$$

โดยค่า  $MSE$  ได้จากสมการเต็มรูปและทำการเปรียบเทียบกับ  $F$  จากตารางที่องศาเสรีเท่ากับ 1 และ  $n - 4$  หากไม่สามารถปฏิเสธสมมติฐานหลักแสดงว่าสมการถดถอยของทั้งสองกลุ่มมีความชันเท่ากันหรือตัวแปรอิสระเชิงคุณภาพไม่มีปฏิสัมพันธ์กับตัวแปรอิสระเชิงปริมาณ

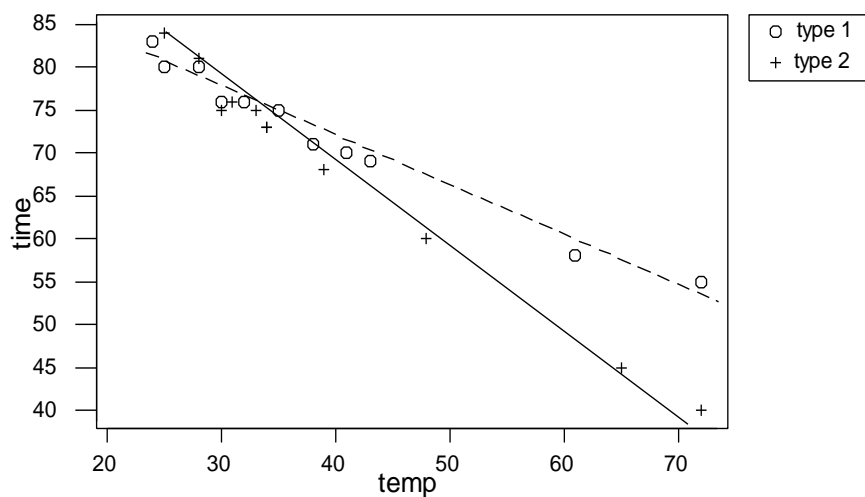
**ตัวอย่างที่ 7.2** ในการวิเคราะห์หาความสัมพันธ์ระหว่างอุณหภูมิใช้งานของแบตเตอรี่ 2 ชนิดกับอายุการใช้งานมีข้อมูลดังตารางข้างล่าง จงสร้างสมการถดถอยที่เหมาะสมกับข้อมูล

อุณหภูมิ (temp)	ชนิด (type)	เวลาที่ใช้ (time)
28	0	80
25	0	80
61	0	58
24	0	83
32	0	76
43	0	69
38	0	71
30	0	76
41	0	70

อุณหภูมิ (temp)	ชนิด (type)	เวลาที่ใช้ (time)
72	0	55
35	0	75
48	1	60
65	1	45
33	1	75
34	1	73
39	1	68
31	1	76
34	1	73
72	1	40
25	1	84
30	1	75
28	1	81

### วิธีทำ

จากข้อมูลข้างต้นนำมาวาดแผนภาพกระจายระหว่างอุณหภูมิกับอายุการใช้แยกตามประเภทพบว่าเส้นกราฟของแบตเตอรี่ทั้งสองประเภทมีความชันที่แตกต่างกัน



จากการวิเคราะห์สมการถดถอยโดยให้ประเภทของแบตเตอรี่เป็นตัวแปรหุ่นโดยใช้โปรแกรม MINITAB พบว่าสมการถดถอยสามารถพยากรณ์อายุการใช้งานของแบตเตอรี่ด้วยค่า  $F$

เท่ากับ 154.27 และ  $p$ -value เท่ากับ 0.000 นอกจากนี้เมื่อพิจารณาตัวแปรอุณหภูมิและประเภทของแบตเตอรี่นั้นพบว่าตัวแปรทั้งสองมีความสำคัญต่ออายุการใช้งานด้วย  $p$ -value เท่ากับ 0.000 และ 0.020 ตามลำดับและค่าสัมประสิทธิ์การตัดลิไนใจ (R-Sq) เท่ากับ 94.2% แสดงว่าสมการถดถอยที่ได้ นี้มีความถูกต้องในการพยากรณ์อยู่ในระดับที่สูงดังภาพข้างล่าง

```

The regression equation is
time = 101 - 0.751 temp - 3.23 type

Predictor      Coef      SE Coef      T      P
Constant      101.396   1.917        52.90  0.000
temp          -0.75142  0.04346     -17.29  0.000
type          -3.226    1.266        -2.55  0.020

S = 2.969      R-Sq = 94.2%      R-Sq(adj) = 93.6%

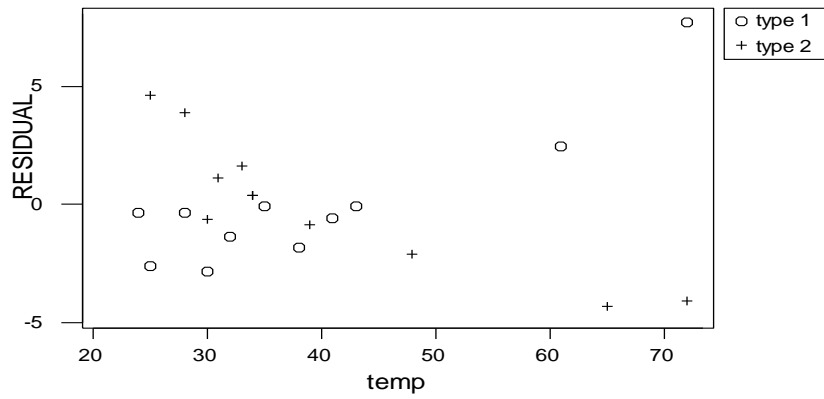
Analysis of Variance

Source      DF      SS      MS      F      P
Regression  2      2719.1  1359.6  154.27  0.000
Residual Error  19    167.4   8.8
Total      21    2886.6

Source      DF      Seq SS
temp        1      2662.0
type        1      57.2

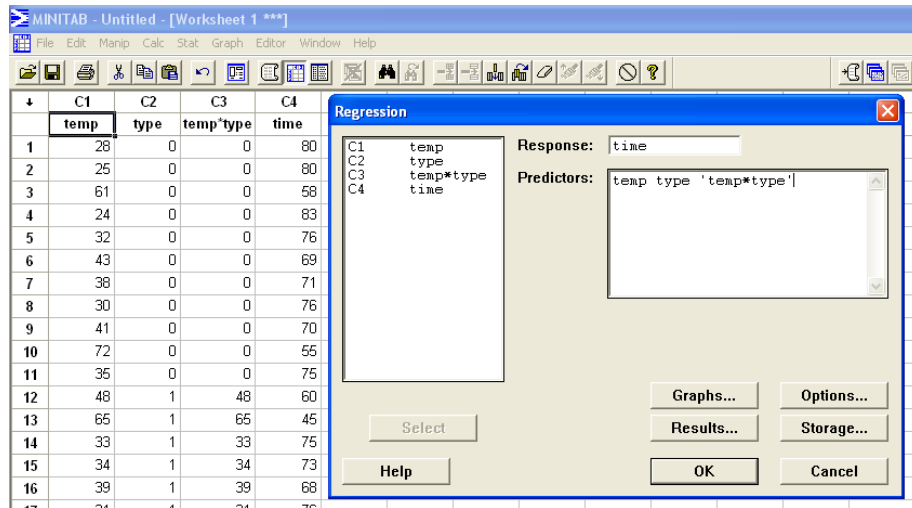
```

เมื่อพิจารณาแผนภาพกระจายระหว่างตัวแปรอุณหภูมิกับค่าส่วนเหลือแยกตามประเภทของแบตเตอรี่แล้วพบว่าที่อุณหภูมิต่างๆ แบตเตอรี่ประเภทที่ 1 จะมีความคลาดเคลื่อนในการพยากรณ์ต่ำแต่ความคลาดเคลื่อนในการพยากรณ์เพิ่มขึ้นเมื่ออุณหภูมิเพิ่มขึ้นในทางกลับกันที่อุณหภูมิต่างๆ แบตเตอรี่ประเภทที่ 2 จะมีความคลาดเคลื่อนในการพยากรณ์สูงและลดลงเมื่ออุณหภูมิสูงขึ้น จากรูปแบบที่เกิดขึ้นในแผนภาพนี้อาจมาจากการละตัวแปรบางตัวแปรที่จำเป็นในสมการถดถอย นอกจากนี้เมื่อพิจารณาแผนภาพกระจายของอุณหภูมิกับอายุการใช้งานแยกตามประเภทของแบตเตอรี่นั้นจะเห็นตัวแปรทั้งสองมีปฏิสัมพันธ์กันจึงควรเพิ่มพจน์ของปฏิสัมพันธ์ระหว่างตัวแปรทั้งสอง การเพิ่มพจน์ดังกล่าวสามารถทำได้โดยการคูณตัวแปรทั้งสองเข้าด้วยกัน ดังตารางข้างล่างจากนั้นสร้างสมการถดถอยโดยใช้ตัวแปรอิสระทั้ง 3 ตัวคือ temp type และ temp\*type



temp	type	temp*type	time
28	0	0	80
25	0	0	80
61	0	0	58
24	0	0	83
32	0	0	76
43	0	0	69
38	0	0	71
30	0	0	76
41	0	0	70
72	0	0	55
35	0	0	75
48	1	48	60
65	1	65	45
33	1	33	75
34	1	34	73
39	1	39	68
31	1	31	76
34	1	34	73
72	1	72	40
25	1	25	84
30	1	30	75
28	1	28	81

การสร้างสมการถดถอยโดยใช้โปรแกรม MINITAB ทำโดยการเพิ่มตัวแปรปฏิสัมพันธ์ “temp\*type” ในช่อง “Predictors:” ดังภาพข้างล่าง



เมื่อพิจารณาผลลัพธ์ที่ได้จากการโปรแกรม MINITAB ในรูปข้างล่างพบว่าพจน์ของปฏิสัมพันธ์ระหว่างตัวแปรอิสระทั้งสองหรือ temp\*type มีความสำคัญต่ออายุการใช้งานด้วย  $p$ -value เท่ากับ 0.000 และค่าสัมประสิทธิ์การตัดสินใจ (R-Sq(adj)) เท่ากับ 98.5% เพิ่มขึ้นจากเดิมนั้นคือสมการถดถอยมีความถูกต้องในการพยากรณ์มากขึ้นแต่อย่างไรก็ตามสมการถดถอยที่ไม่ได้พิจารณาปฏิสัมพันธ์เข้าในการพิจารณานั้นให้ค่าสัมประสิทธิ์การตัดสินใจที่สูงมากอยู่แล้ว

The regression equation is  

$$\text{time} = 94.7 - 0.581 \text{ temp} + 9.96 \text{ type} - 0.334 \text{ temp*type}$$

Predictor	Coef	SE Coef	T	P
Constant	94.736	1.251	75.73	0.000
temp	-0.58063	0.03009	-19.30	0.000
type	9.957	1.770	5.63	0.000
temp*type	-0.33421	0.04209	-7.94	0.000

S = 1.437      R-Sq = 98.7%      R-Sq(adj) = 98.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	2849.40	949.80	459.74	0.000
Residual Error	18	37.19	2.07		
Total	21	2886.59			

Source	DF	Seq SS
temp	1	2661.96
type	1	57.18
temp*type	1	130.26

เมื่อทดสอบความสำคัญของปฏิสัมพันธ์โดยใช้สถิติ  $F$  โดยใช้สมการที่ (7.6) หรือ

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

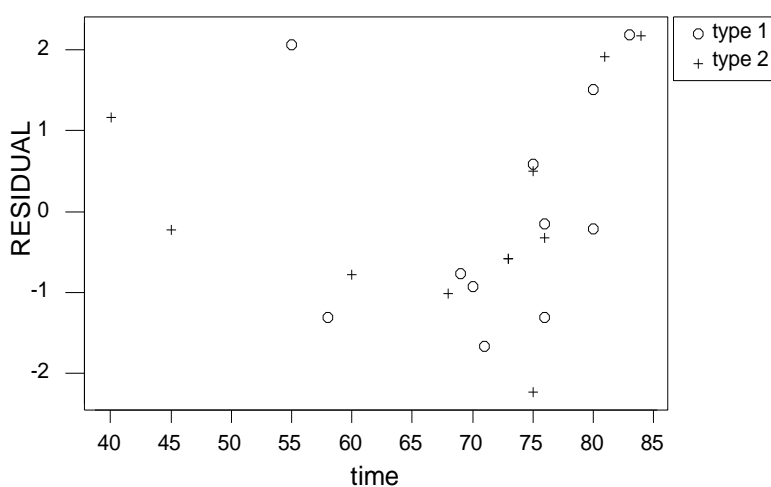
โดย

$$SSR(\beta_3 | \beta_0, \beta_1, \beta_2) = 130.26$$

ดังนั้น

$$F = \frac{SSR(\beta_3 | \beta_0, \beta_1, \beta_2)/1}{MSE} = \frac{130.26/1}{2.07} = 62.93$$

เนื่องจากค่าสถิติ  $F > F_{0.05, 1, 18} = 4.41$  ดังนั้นสามารถสรุปว่าปฏิสัมพันธ์ระหว่างตัวแปรทั้งสองมีความสำคัญกับสมการถดถอยซึ่งสอดคล้องกับค่า  $t$  ที่ได้จากโปรแกรม MINITAB นอกจากนี้เมื่อพิจารณาแผนภาพกระจายระหว่างตัวแปรอุณหภูมิกับค่าส่วนเหลือแยกตามประเภทของแบตเตอรี่พบว่าแผนภาพมีการกระจายตัวที่ดีขึ้นจากเดิมหรือแผนภาพไม่มีรูปแบบที่ชัดเจนเหมือนกับที่ได้จากตัวแบบแรก



ดังนั้นสมการถดถอยที่ได้คือ

$$\text{time} = 94.70 - 0.581 \text{ temp} + 9.96 \text{ type} - 0.334 \text{ temp} * \text{type}$$

และสมการสำหรับแบตเตอรี่ทั้งสองประเภทคือ

ประเภทที่ 1:  $\text{time} = 94.70 - 0.581 \text{ temp}$

ประเภทที่ 2:  $\text{time} = 104.66 - 0.915 \text{ temp}$

จากสมการทั้งสองพบว่าสมการทั้งสองมีความชันและจุดตัดแกน  $Y$  ที่แตกต่างกันซึ่งสอดคล้องกับแผนภาพกระจาย

## 7.2 ตัวแปรอิสระเชิงคุณภาพ 1 ตัวที่มีมากกว่าสองกลุ่มย่อย

ในกรณีที่ตัวแปรอิสระเชิงคุณภาพนั้นมีจำนวนกลุ่มมากกว่า 2 กลุ่มนักวิจัยต้องสร้างตัวแปรหุ่นจำนวนเท่ากับระดับหรือกลุ่มของตัวแปรนั้นลบออก 1 เช่น หากระดับการศึกษาแบ่งออกเป็น 5 ระดับ (ประถมศึกษา, มัธยมศึกษา, ปริญญาตรี, ปริญญาโท และปริญญาเอก) แล้วตัวแปรหุ่นสำหรับระดับการศึกษาจะมี 4 ตัวแปรดังนี้

$X_2$	$X_3$	$X_4$	$X_5$	
0	0	0	0	หากผู้นั้นจบระดับประถมศึกษา
1	0	0	0	หากผู้นั้นจบระดับมัธยมศึกษา
0	1	0	0	หากผู้นั้นจบระดับปริญญาตรี
0	0	1	0	หากผู้นั้นจบระดับปริญญาโท
0	0	0	1	หากผู้นั้นจบระดับปริญญาเอก

และมีสมการถดถอยเชิงเส้นคือ

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon \quad (7.7)$$

หากผู้นั้นจบการศึกษาระดับประถมศึกษาแล้วสมการ (7.1) จะกลายเป็น

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

หากผู้นั้นจบการศึกษาระดับมัธยมศึกษาแล้วสมการถดถอยคือ

$$Y = (\beta_0 + \beta_2) + \beta_1 X_1 + \varepsilon$$

หากผู้นั้นจบการศึกษาระดับปริญญาตรีแล้วสมการถดถอยคือ

$$Y = (\beta_0 + \beta_3) + \beta_1 X_1 + \varepsilon$$

หากผู้นั้นจบการศึกษาระดับปริญญาโทแล้วสมการถดถอยคือ

$$Y = (\beta_0 + \beta_4) + \beta_1 X_1 + \varepsilon$$

หากผู้นั้นจบการศึกษาระดับปริญญาเอกแล้วสมการถดถอยคือ

$$Y = (\beta_0 + \beta_5) + \beta_1 X_1 + \varepsilon$$

จากสมการข้างต้นพบว่า  $\beta_2$   $\beta_3$   $\beta_4$  และ  $\beta_5$  ทำให้จุดตัดแกน  $Y$  มีค่าเปลี่ยนแปลงไปหรืออาจกล่าวได้ว่า  $\beta_2$   $\beta_3$   $\beta_4$  และ  $\beta_5$  ใช้วัดผลกระทบของการศึกษาระดับมัธยมศึกษา ปริญญาตรี ปริญญาโทและปริญญาเอกเทียบกับคนที่มีการศึกษาระดับประถมศึกษาตามลำดับ หากสัมประสิทธิ์ใดมีค่า

เป็นบวกแสดงว่าระดับการศึกษานั้นให้ค่าตัวแปรตามที่สูงกว่าระดับการศึกษาระดับประถมศึกษา เช่น หาก  $\beta_2$  มีค่าเป็นบวกแสดงว่าผู้จบมัธยมศึกษา มีค่าของตัวแปรตามสูงกว่าผู้จบประถมศึกษา เป็นต้น ในทางกลับกันหากสัมประสิทธิ์ใดมีค่าเป็นลบแสดงว่าระดับการศึกษานั้นให้ค่าตัวแปรตามต่ำกว่าระดับการศึกษาระดับประถมศึกษา เช่น หาก  $\beta_3$  มีค่าเป็นลบแสดงว่าผู้จบปริญญาตรี มีค่าของตัวแปรตามต่ำกว่าผู้จบประถมศึกษา เป็นต้น

การวิเคราะห์ข้อมูลประเภทนี้จะไม่แตกต่างจากการวิเคราะห์ข้อมูลเชิงคุณภาพที่มีสองกลุ่ม ในหัวข้อที่แล้วและข้อมูลในแต่ละกลุ่มอาจมีปฏิสัมพันธ์ระหว่างกันซึ่งสามารถตรวจสอบได้โดยวิธีเดียวกับหัวข้อที่แล้ว

**ตัวอย่างที่ 7.3** การศึกษาความสัมพันธ์ระหว่างจำนวนครัวเรือน (หมื่นครัวเรือน) ในแต่ละจังหวัดของ 3 ภาคกับยอดขายของห้างสรรพสินค้า (ล้านบาท) โดยใช้สมการถดถอยดังนี้

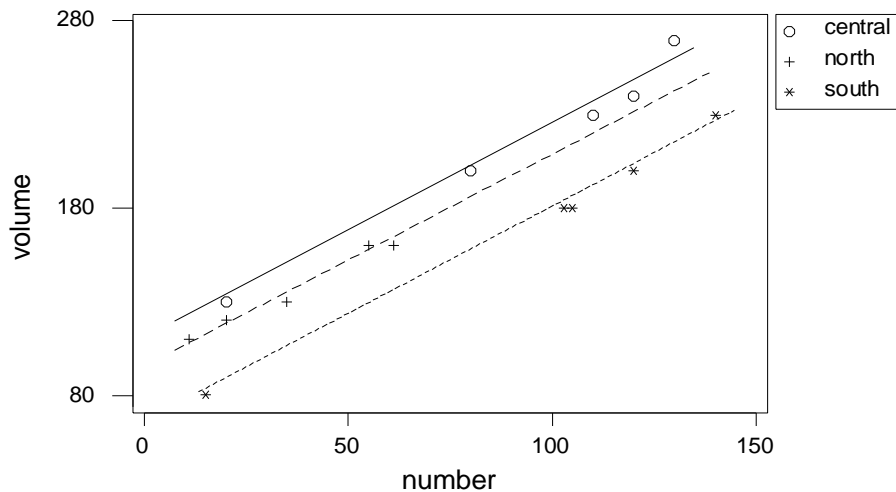
จำนวนครัวเรือน (number)	ภาค (region)	ยอดขาย (volume)
61	เหนือ	160
11	เหนือ	110
35	เหนือ	130
20	เหนือ	120
55	เหนือ	160
120	กลาง	240
80	กลาง	200
130	กลาง	270
110	กลาง	230
20	กลาง	130
120	ใต้	200
103	ใต้	180
140	ใต้	230
15	ใต้	80
105	ใต้	180

#### วิธีทำ

เมื่อพิจารณาแผนภาพกระจายระหว่างจำนวนครัวเรือนกับยอดขายแยกตามภาคแล้วพบว่าภาคกลางแต่ละภาคมีจำนวนยอดขายที่แตกต่างกันในขณะที่จำนวนครัวเรือนไม่แตกต่างกันโดยภาค



กลาง (central) มีจำนวนยอดขายมากกว่าภาคอื่นที่จำนวนครัวเรือนเท่ากัน นอกจากนี้ยังพบอีกว่า เส้นกราฟทั้งสามขนานกันหรือมีความชันที่ไม่แตกต่างกันแสดงว่าจำนวนครัวเรือนไม่มีปฏิสัมพันธ์กับภาคจึงไม่จำเป็นต้องเพิ่มพจน์ปฏิสัมพันธ์



เนื่องจากภาค (region) เป็นตัวแปรอิสระเชิงคุณภาพที่มี 3 กลุ่มดังนั้นในการวิเคราะห์ข้อมูล จึงจำเป็นต้องสร้างตัวแปรหุ่น 2 ตัวดังนี้

region1	region2	
0	0	หากห้างสรรพสินค้าอยู่ในภาคเหนือ
1	0	หากห้างสรรพสินค้าอยู่ในภาคกลาง
0	1	หากห้างสรรพสินค้าอยู่ในภาคใต้

ดังนั้นตารางข้อมูลสามารถเขียนใหม่ได้ดังนี้

number	region1	region2	volume
61	0	0	160
11	0	0	110
35	0	0	130
20	0	0	120
55	0	0	160
120	1	0	240
80	1	0	200

number	region1	region2	volume
130	1	0	270
110	1	0	230
20	1	0	130
120	0	1	200
103	0	1	180
140	0	1	230
15	0	1	80
105	0	1	180

จากการวิเคราะห์สมการถดถอยโดยใช้โปรแกรม MINITAB พบว่าสมการถดถอยสามารถพยากรณ์ยอดขาย (volume) ด้วยค่า  $F$  เท่ากับ 465.24 และ  $p$ -value เท่ากับ 0.000 นอกจากนี้พบว่าตัวแปรอิสระทุกตัวมีความสำคัญต่อการพยากรณ์ยอดขายด้วย  $p$ -value เท่ากับ 0.000, 0.008 และ 0.000 ตามลำดับและค่าสัมประสิทธิ์การตัดสินใจ (R-Sq) เท่ากับ 99.2% แสดงว่าสมการถดถอยที่ได้ี้มีความถูกต้องในการพยากรณ์อยู่ในระดับที่สูงดังภาพข้างล่าง

The regression equation is					
volume = 93.5 + 1.17 number + 13.1 region1 - 32.3 region2					
Predictor	Coef	SE Coef	T	P	
Constant	93.496	2.818	33.18	0.000	
number	1.16770	0.03940	29.64	0.000	
region1	13.076	4.070	3.21	0.008	
region2	-32.295	4.170	-7.74	0.000	
S = 5.423		R-Sq = 99.2%		R-Sq(adj) = 99.0%	
Analysis of Variance					
Source	DF	SS	MS	F	P
Regression	3	41050	13683	465.24	0.000
Residual Error	11	324	29		
Total	14	41373			
Source	DF	Seq SS			
number	1	35783			
region1	1	3503			
region2	1	1764			

สมการถดถอยที่ได้สามารถเขียนได้ดังนี้

$$\text{volume} = 93.5 + 1.17 \text{ number} + 13.1 \text{ region1} - 32.3 \text{ region2}$$

ค่าสัมประสิทธิ์ของตัวแปร number เท่ากับ 1.17 หมายความว่าเมื่อพิจารณาภาคใดภาคหนึ่งเพียงภาคเดียว ถ้าจำนวนครัวเรือนเพิ่มขึ้นหนึ่งหมื่นครัวเรือนจะทำให้ยอดขายเพิ่มขึ้น 1.17 ล้านบาท

ค่าสัมประสิทธิ์ของตัวแปร region1 เท่ากับ 13.1 หมายความว่าเมื่อจำนวนครัวเรือนเท่ากันแล้วภาคกลางมียอดขายมากกว่าภาคเหนืออยู่ 13.1 ล้านบาท

ค่าสัมประสิทธิ์ของตัวแปร region2 เท่ากับ -32.1 หมายความว่าเมื่อจำนวนครัวเรือนเท่ากันแล้วภาคใต้มียอดขายน้อยกว่าภาคเหนืออยู่ 32.1 ล้านบาท

นอกจากนี้สามารถแยกสมการถดถอยสำหรับแต่ละภาคได้ดังนี้คือ

ภาคเหนือ:  $\text{volume} = 93.5 + 1.17 \text{ number}$

ภาคกลาง:  $\text{volume} = 106.6 + 1.17 \text{ number}$

ภาคใต้:  $\text{volume} = 61.2 + 1.17 \text{ number}$

### 7.3 ตัวแปรอิสระเชิงคุณภาพสองตัว

หากข้อมูลมีตัวแปรอิสระเชิงคุณภาพสองตัวแล้วการวิเคราะห์ข้อมูลชุดนั้นต้องสร้างตัวแปรหุ่น 2 ชุด จากนั้นจึงทำการวิเคราะห์เช่นเดียวกับการวิเคราะห์ในหัวข้อที่ผ่านมา นอกจากนี้หากตัวแปร มีปฏิสัมพันธ์ระหว่างกันแล้วการสร้างสมการถดถอยของปฏิสัมพันธ์นั้นให้ทำเช่นเดียวกับกรณีข้อมูลที่มีตัวแปรเชิงคุณภาพเพียงตัวเดียวกล่าวคือสร้างตัวแปรใหม่โดยนำตัวแปรทั้งสองมาคูณกันแล้วเพิ่มเข้าไปในสมการถดถอย

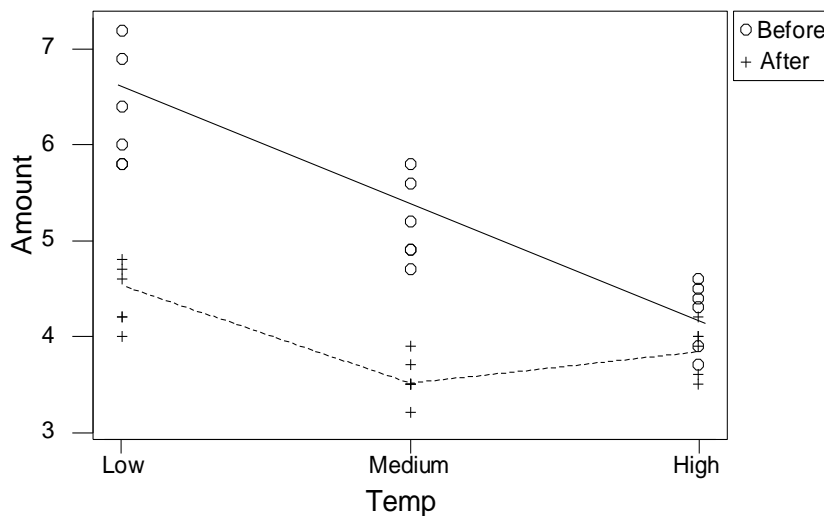
ตัวอย่างที่ 7.4 การวิเคราะห์หาความสัมพันธ์ระหว่างการใช้ฉนวนกันความเย็นภายในบ้านกับปริมาณแก๊สธรรมชาติที่ใช้ในแต่ละวันที่ระดับอุณหภูมิที่แตกต่างกัน 3 ระดับ โดยมีข้อมูลดังนี้

ฉนวน (Insulate)	ช่วงอุณหภูมิ (Temp)	ปริมาณ (Amount)
ก่อนใช้	ต่ำ	7.2
ก่อนใช้	ต่ำ	6.9
ก่อนใช้	ต่ำ	6.4
ก่อนใช้	ต่ำ	6.0
ก่อนใช้	ต่ำ	5.8
ก่อนใช้	ต่ำ	5.8

นวนวน (Insulate)	ช่วงอุณหภูมิ (Temp)	ปริมาณ (Amount)
ก่อนใช้	กลาง	5.6
ก่อนใช้	กลาง	4.7
ก่อนใช้	กลาง	5.8
ก่อนใช้	กลาง	5.2
ก่อนใช้	กลาง	4.9
ก่อนใช้	กลาง	4.9
ก่อนใช้	สูง	4.3
ก่อนใช้	สูง	4.4
ก่อนใช้	สูง	4.5
ก่อนใช้	สูง	4.6
ก่อนใช้	สูง	3.7
ก่อนใช้	สูง	3.9
หลังใช้	ต่ำ	4.8
หลังใช้	ต่ำ	4.6
หลังใช้	ต่ำ	4.7
หลังใช้	ต่ำ	4.0
หลังใช้	ต่ำ	4.2
หลังใช้	ต่ำ	4.2
หลังใช้	กลาง	3.2
หลังใช้	กลาง	3.9
หลังใช้	กลาง	3.5
หลังใช้	กลาง	3.7
หลังใช้	กลาง	3.5
หลังใช้	กลาง	3.5
หลังใช้	สูง	4.2
หลังใช้	สูง	4.0
หลังใช้	สูง	3.9
หลังใช้	สูง	3.5
หลังใช้	สูง	4.0
หลังใช้	สูง	3.6

**วิธีทำ**

เมื่อพิจารณาแผนภาพกระจายระหว่งการใช้ฉนวนกับปริมาณแก๊สที่ใช้ที่ระดับอุณหภูมิ ทั้ง 3 พบว่าก่อนการใช้ฉนวนมีการใช้แก๊สที่มากกว่าหลังการใช้ฉนวนแต่พบว่าเส้นกราฟทั้งสอง ไม่ได้ขนานกันดังนั้นแสดงว่าการใช้ฉนวนกับระดับอุณหภูมิมิ่ปฏิสัมพันธ์กันดังนั้นจึงต้องเพิ่มพจน์ปฏิสัมพันธ์เข้าไปในการสร้างสมการถดถอย



ข้อมูลชุดนี้มีตัวแปรอิสระเชิงคุณภาพ 2 ชุดคือ การใช้ฉนวนกับระดับอุณหภูมิ ดังนั้นการวิเคราะห์ข้อมูลจึงจำเป็นต้องสร้างตัวแปรหุ่น 3 ตัวคือ Insulate Temp1 และ Temp2 ดังนี้

$$\text{Insulate} = \begin{cases} 0 & \text{สำหรับก่อนการใช้ฉนวน} \\ 1 & \text{สำหรับหลังการใช้ฉนวน} \end{cases}$$

Temp1	Temp2	
0	0	สำหรับอุณหภูมิระดับต่ำ
1	0	สำหรับอุณหภูมิระดับกลาง
0	1	สำหรับอุณหภูมิระดับสูง

นอกจากนี้ยังต้องเพิ่มตัวแปรสำหรับปฏิสัมพันธ์ระหว่างตัวแปรหุ่น Insulate Temp1 และ Temp2 ดังนี้คือ Temp1\*Insulate และ Temp2\*Insulate ดังนั้นตารางข้อมูลสามารถเขียนใหม่ได้ดังนี้

<b>Insulate</b>	<b>Temp1</b>	<b>Temp2</b>	<b>Temp1*Insulate</b>	<b>Temp2*Insulate</b>	<b>Amount</b>
0	0	0	0	0	7.2
0	0	0	0	0	6.9
0	0	0	0	0	6.4
0	0	0	0	0	6.0
0	0	0	0	0	5.8
0	0	0	0	0	5.8
0	1	0	0	0	5.6
0	1	0	0	0	4.7
0	1	0	0	0	5.8
0	1	0	0	0	5.2
0	1	0	0	0	4.9
0	1	0	0	0	4.9
0	0	1	0	0	4.3
0	0	1	0	0	4.4
0	0	1	0	0	4.5
0	0	1	0	0	4.6
0	0	1	0	0	3.7
0	0	1	0	0	3.9
1	0	0	0	0	4.8
1	0	0	0	0	4.6
1	0	0	0	0	4.7
1	0	0	0	0	4.0
1	0	0	0	0	4.2
1	0	0	0	0	4.2
1	1	0	1	0	3.2
1	1	0	1	0	3.9
1	1	0	1	0	3.5
1	1	0	1	0	3.7
1	1	0	1	0	3.5

Insulate	Temp1	Temp2	Temp1*Insulate	Temp2*Insulate	Amount
1	1	0	1	0	3.5
1	0	1	1	0	4.2
1	0	1	1	0	4.0
1	0	1	1	0	3.9
1	0	1	1	0	3.5
1	0	1	1	0	4.0
1	0	1	1	0	3.6

จากการวิเคราะห์สมการถดถอยโดยใช้โปรแกรม MINITAB พบว่าสมการถดถอยสามารถพยากรณ์ปริมาณการใช้แก๊สด้วยค่า  $F$  เท่ากับ 41.75 และ  $p$ -value เท่ากับ 0.000 นอกจากนี้พบว่าตัวแปรอิสระทุกตัวยกเว้นพจน์ปฏิสัมพันธ์ระหว่าง Temp1\*Insulate มีความสำคัญต่อการพยากรณ์การใช้แก๊สด้วย  $p$ -value เท่ากับ 0.000, 0.000, 0.000 และ 0.000 ตามลำดับแต่ในขณะที่พจน์ปฏิสัมพันธ์ระหว่าง Temp1\*Insulate ไม่มีความสำคัญต่อการพยากรณ์ปริมาณการใช้แก๊สเนื่องจากค่า  $p$ -value เท่ากับ 0.350 และค่าสัมประสิทธิ์การตัดสินใจ (R-Sq) เท่ากับ 87.4% แสดงว่าสมการถดถอยที่ได้นี้มีความถูกต้องในการพยากรณ์อยู่ในระดับที่สูงคงภาพข้างล่าง

The regression equation is					
Amount = 6.35 - 1.17 temp1 - 1.93 insulate - 2.12 temp2					
+ 0.30 temp1*insulate + 1.57 temp2*insulate					
Predictor	Coef	SE Coef	T	P	
Constant	6.3500	0.1580	40.19	0.000	
temp1	-1.1667	0.2234	-5.22	0.000	
insulate	-1.9333	0.2234	-8.65	0.000	
temp2	-2.1167	0.2234	-9.47	0.000	
temp1*insulate	0.3000	0.3160	0.95	0.350	
temp2*insulate	1.5667	0.3160	4.96	0.000	
S = 0.3870		R-Sq = 87.4%		R-Sq(adj) = 85.3%	
Analysis of Variance					
Source	DF	SS	MS	F	P
Regression	5	31.2667	6.2533	41.75	0.000
Residual Error	30	4.4933	0.1498		
Total	35	35.7600			
Source	DF	Seq SS			
temp1	1	0.9800			
insulate	1	15.4711			
temp2	1	10.6667			
temp1*insulate	1	0.4672			
temp2*insulate	1	3.6817			

สมการถดถอยที่ได้สามารถเขียนได้ดังนี้

$$\text{Amount} = 6.35 - 1.17 \text{ temp1} - 1.93 \text{ insulate} - 2.12 \text{ temp2} + 0.30 \text{ temp1} * \text{insulate} \\ + 1.57 \text{ temp2} * \text{insulate}$$

## สรุป

การวิเคราะห์ข้อมูลที่มีตัวแปรอิสระเป็นตัวแปรเชิงคุณภาพเช่น เพศ เชื้อชาติหรือศาสนา เป็นต้นนั้นจะมีการวิเคราะห์ที่แตกต่างจากการวิเคราะห์ที่ผ่านมาโดยต้องสร้างตัวแปรหุ่นให้มีจำนวนน้อยกว่าจำนวนกลุ่มของตัวแปรเชิงคุณภาพอยู่หนึ่งตัว ตัวแปรหุ่นแต่ละตัวจะแทนกลุ่มของตัวแปรเชิงคุณภาพเทียบกับกลุ่มที่ไม่ได้ถูกสร้างเป็นตัวแปรหุ่น การสร้างสมการถดถอยจะสร้างโดยวิธีเดียวกันกับการสร้างสมการถดถอยโดยไม่มีตัวแปรหุ่น

## คำถามท้ายบท

- 7.1 จงอธิบายว่าทำไมจึงต้องสร้างตัวแปรหุ่นให้กับตัวแปรอิสระเชิงคุณภาพและหากไม่สร้างตัวแปรหุ่นแล้วจะมีผลต่อการวิเคราะห์ข้อมูลอย่างไร
- 7.2 จงอธิบายว่าทำไมจึงไม่ควรสร้างตัวแปรหุ่นที่มีจำนวนเท่ากับจำนวนกลุ่มของตัวแปรอิสระเชิงคุณภาพนั้น
- 7.3 การวางแผนภาพกระจายระหว่างตัวแปรตามกับตัวแปรอิสระเชิงปริมาณ โดยแยกตามตัวแปรอิสระเชิงคุณภาพช่วยบอกถึงความสำคัญของตัวแปรอิสระเชิงคุณภาพต่อตัวแปรตามได้อย่างไร
- 7.4 เกษตรกรสังเกตเห็นว่าสภาพอากาศที่แตกต่างกันเช่นร้อนและเย็นอาจทำให้เป็ดที่เลี้ยงโตช้าลง เกษตรกรจึงทำการทดลองแบ่งเป็ดออกเป็น 2 กลุ่มโดยกลุ่มแรกเลี้ยงในที่โล่งแจ้งจำนวน 20 ตัว และกลุ่มที่ 2 เลี้ยงในที่ร่มจำนวน 30 ตัวแล้วชั่งน้ำหนักของเป็ด จงระบุตัวแปรอิสระและตัวแปรตาม หากต้องสร้างสมการถดถอยท่านจะสร้างตัวแปรหุ่นอย่างไรและตีความ  $\beta$  อย่างไร
- 7.5 ในการศึกษาความสัมพันธ์ระหว่างอายุและเพศที่มีต่อความดันโลหิตมีข้อมูลดังนี้

อายุ ( $X_1$ )	63	65	45	53	60	42	58	62	47	56	48	55
เพศ ( $X_2$ )	ชาย	หญิง	ชาย	ชาย	หญิง	หญิง	ชาย	ชาย	หญิง	ชาย	หญิง	หญิง
ความดัน ( $Y$ )	161	208	130	148	200	145	152	140	135	141	158	173

จงวางแผนภาพกระจายระหว่างอายุกับความดันโลหิตแยกตามเพศ พร้อมทั้งอธิบายแผนภาพที่ได้



- 7.6 จากข้อ 7.5 จงสร้างสมการถดถอยระหว่างอายุกับความดันโลหิต พร้อมทั้งวิเคราะห์ส่วนเหลือว่าสมการถดถอยที่ได้เหมาะสมหรือไม่
- 7.7 จากข้อ 7.5 จงสร้างสมการถดถอยระหว่างอายุกับความดันโลหิตโดยใช้เพศเป็นตัวแปรหุ่นอธิบายสมการถดถอยของแต่ละเพศ พร้อมทั้งวิเคราะห์ส่วนเหลือว่าสมการถดถอยที่ได้เหมาะสมหรือไม่
- 7.8 จากข้อ 5.10 หากน้ำผลไม้มี 3 กลิ่นคือ ส้ม, มะนาวและองุ่น จงเขียนสมการเชิงเส้นที่รวมตัวแปรกลิ่นเข้าในสมการ
- 7.9 จากข้อ 7.8 หากกลิ่นส้มคือข้อมูล 5 ค่าแรก กลิ่นมะนาว 5 ค่าต่อมาและกลิ่นองุ่น 4 ค่าที่เหลือ จงวาดแผนภาพกระจายระหว่าง pH กับความพึงพอใจแยกตามกลิ่น พร้อมทั้งอธิบายแผนภาพที่ได้
- 7.10 จากข้อ 7.9 จงสร้างสมการถดถอยและพิจารณาว่ากลิ่นมีผลต่อความพึงพอใจหรือไม่ที่ระดับนัยสำคัญ 0.05
- 7.11 หากต้องการศึกษาผลของความดันและชนิดของแผ่นกรองที่มีต่ออัตราการกรอง โดยมีข้อมูลดังนี้

ความดัน ( $X_1$ )	2	2	2	4	4	4	6	6	6	8
ชนิดของแผ่นกรอง ( $X_2$ )	1	2	3	1	2	3	1	2	3	1
อัตราการกรอง ( $Y$ )	26	35	35	52	66	87	52	93	91	60
ความดัน	8	8	10	10						
ชนิดของแผ่นกรอง	2	3	1	2						
อัตราการกรอง	101	138	65	115						

- จงวาดแผนภาพกระจายระหว่างความดันกับอัตราการกรองแยกตามชนิดของแผ่นกรอง แผ่นกรองทั้ง 3 ชนิดมีความชันที่เหมือนกันหรือไม่
- 7.12 จากข้อ 7.11 จงสร้างตัวแปรหุ่นของชนิดของแผ่นกรอง พร้อมทั้งสร้างสมการถดถอย ค่า  $\beta$  ที่ได้จะอธิบายได้อย่างไร ตัวแปรใดบ้างที่มีผลต่ออัตราการกรอง
- 7.13 จากข้อ 7.11 จงเพิ่มพจน์ปฏิสัมพันธ์ระหว่างความดันและชนิดของแผ่นกรองแล้วอธิบายผลลัพธ์ที่ได้
- 7.14 นักวิจัยต้องการศึกษาความสัมพันธ์ระหว่างระยะทางที่จับ (พันกิโลเมตร) และยี่ห้อของรถยนต์ที่มีต่อค่าใช้จ่ายในการบำรุงรักษา (พันบาท) โดยมีข้อมูลดังนี้

ยี่ห้อ ( $X_1$ )	ระยะทาง ( $X_2$ )	ค่าใช้จ่าย ( $Y$ )
ฮอนด้า	112	45.6
ฮอนด้า	173	82.8
ฮอนด้า	111	50.0
ฮอนด้า	110	48.9
ฮอนด้า	67	38.7
โตโยต้า	124	53.1
โตโยต้า	153	65.0
โตโยต้า	113	47.5
โตโยต้า	82	47.4
โตโยต้า	123	53.3
นิสสัน	77	39.6
นิสสัน	143	61.8
นิสสัน	88	47.4
นิสสัน	136	63.9
นิสสัน	71	45.7
ฟอร์ด	87	46.0
ฟอร์ด	65	43.3
ฟอร์ด	131	62.1
ฟอร์ด	99	46.0
ฟอร์ด	137	55.3

จงวาดแผนภาพกระจายระหว่างระยะทางที่จับกับค่าใช้จ่ายแยกตามยี่ห้อของรถยนต์ จงอธิบายแผนภาพที่ได้

7.15 จากข้อ 7.14 จงสร้างตัวแปรหุ่นของยี่ห้อของรถยนต์พร้อมทั้งสร้างสมการถดถอย ระยะทางที่จับและยี่ห้อของรถยนต์มีผลต่อค่าใช้จ่ายในการบำรุงรักษาหรือไม่ที่ระดับนัยสำคัญ 0.10

7.16 จากสมการถดถอยที่ได้ในข้อ 7.15 จงวิเคราะห์ส่วนเหลือว่าตัวแบบที่ได้เหมาะสมหรือไม่