

## บทที่ 8

### การเลือกตัวแบบที่เหมาะสม

การสร้างตัวแบบนั้นนักวิจัยจะสร้างโดยการเลือกตัวแปรอิสระเพียงบางตัวจากตัวแปรอิสระจำนวนมากที่นักวิจัยคาดว่าจะมีความสำคัญต่อตัวแปรตาม การเลือกตัวแบบที่เหมาะสมกับข้อมูลนั้นคือการเลือกเฉพาะกลุ่มตัวแปรอิสระที่เหมาะสมในการพยากรณ์โดยการเปรียบเทียบตัวแบบที่มีทั้งหมดโดยใช้ทั้งเกณฑ์ที่เหมาะสมและสภาพความเป็นจริงของข้อมูลประกอบกัน Montgomery & Peck (1991, p. 265) กล่าวว่า การเลือกตัวแปรที่ต้องการนั้นมีวัตถุประสงค์ที่ขัดแย้งกัน 2 ประการ คือ (1) นักวิจัยอาจต้องการตัวแบบที่มีตัวแปรอิสระมากที่สุดเท่าที่จะมากได้เพื่อให้ใช้ประโยชน์ในการพยากรณ์ตัวแปรตามให้มากที่สุด และ (2) ต้องการให้มีตัวแปรอิสระที่น้อยที่สุด เนื่องจากความแปรปรวนของค่าพยากรณ์จะเพิ่มขึ้นเมื่อจำนวนตัวแปรอิสระเพิ่มขึ้น นอกจากนี้การเพิ่มจำนวนตัวแปรอิสระมากขึ้นในตัวแบบหมายถึงค่าใช้จ่ายในการเก็บข้อมูลในอนาคตจะมากขึ้นด้วย อย่างไรก็ตามการเลือกตัวแบบที่เหมาะสมที่จะกล่าวถึงในบทนี้เป็น การเลือกตัวแบบโดยพิจารณาจากข้อมูล (data driven) ในทางปฏิบัติแล้วการเลือกตัวแบบนั้นต้องคำนึงถึงทฤษฎีที่เกี่ยวข้องด้วย นอกจากนี้ก่อนการสร้างตัวแบบเพื่อทำการคัดเลือกนั้นจำเป็นต้องตรวจสอบและกำจัดข้อมูลที่ผิดปกติ (outlier) หรือข้อมูลที่มีอิทธิพล (influential) ออกไปเสียก่อนมิฉะนั้นตัวแบบที่ได้จะมีความคลาดเคลื่อนมาก

#### 8.1 เกณฑ์การเลือกตัวแบบที่เหมาะสม

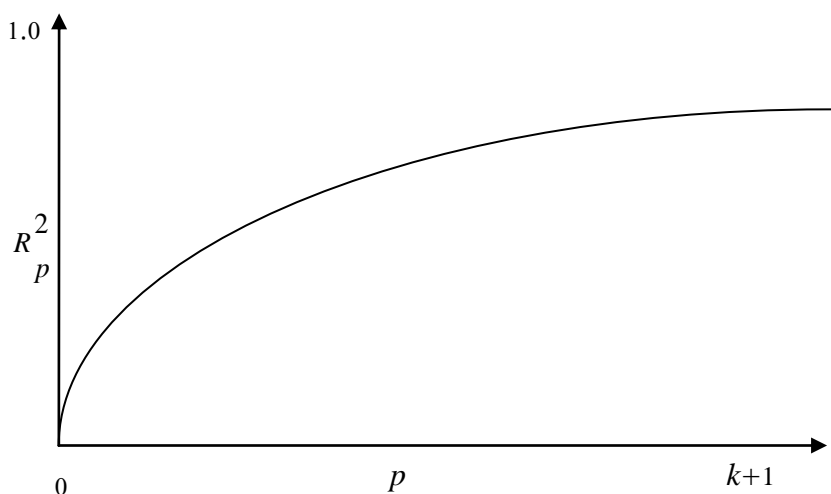
การหาตัวแบบที่เหมาะสมนั้นมี 2 ขั้นตอนหลักคือ การสร้างตัวแบบจำนวนมากเพื่อทำการคัดเลือกจากนั้นเลือกตัวแบบที่เหมาะสมโดยใช้เกณฑ์ที่จะกล่าวถึงต่อไป เกณฑ์ที่ใช้ในการเลือกตัวแบบที่เหมาะสมนั้นมีจำนวนมาก แต่ละเกณฑ์มีข้อดีและข้อเสียที่แตกต่างกันดังนี้

##### 8.1.1 ค่าสัมประสิทธิ์การตัดสินใจสำหรับตัวแปรอิสระหลายตัว

ค่า  $R^2$  เป็นการวัดความสามารถในการอธิบายความแปรผันของตัวแปรตามโดยตัวแปรอิสระในสมการถดถอย ในที่นี้ให้  $R_p^2$  เป็นค่าสัมประสิทธิ์การตัดสินใจสำหรับสมการถดถอยที่มีจำนวนพจน์  $p$  พจน์หรือมีตัวแปรอิสระหลายตัวที่มีตัวแปรอิสระจำนวน  $p - 1$  ตัวและ  $\beta_0$  อยู่ในสมการ  $R_p^2$  สามารถคำนวณได้ดังนี้

$$R_p^2 = \frac{SSR(p)}{SST} = 1 - \frac{SSE(p)}{SST} \quad (8.1)$$

โดย  $SSR(p)$  และ  $SSE(p)$  คือ ผลรวมกำลังสองถดถอยและผลรวมกำลังสองความคลาดเคลื่อนที่มีจำนวน  $p$  พจน์ตามลำดับ ตัวแบบที่มีค่า  $R_p^2$  สูงเป็นตัวแบบที่สามารถอธิบายความแปรผันของตัวแปรตามได้ดีแต่ข้อเสียของ  $R_p^2$  คือเมื่อจำนวนพจน์ในสมการเพิ่มขึ้นค่า  $R_p^2$  จะเพิ่มขึ้นด้วย ดังนั้นสมการที่มีตัวแปรอิสระทุกตัวจะเป็นสมการที่มีค่า  $R_p^2$  สูงสุดแต่การเพิ่มขึ้นของค่า  $R_p^2$  จะเพิ่มขึ้นไม่มากเมื่อถึงจุดหนึ่งดังภาพที่ 8.1



ภาพที่ 8.1 ความสัมพันธ์ระหว่างจำนวนตัวแปรอิสระกับค่า  $R_p^2$

การเลือกตัวแบบที่เหมาะสมที่สุดโดยใช้วิธีนี้จึงไม่ควรพิจารณาเฉพาะค่า  $R_p^2$  ที่สูงที่สุด แต่ควรคำนึงถึงจำนวนตัวแปรอิสระในตัวแบบด้วยโดยควรเลือกตัวแบบที่ใช้จำนวนตัวแปรอิสระที่น้อยที่สุดที่ให้ค่า  $R_p^2$  ที่สูงในระดับที่ยอมรับได้

### 8.1.2 ค่าสัมประสิทธิ์การตัดสินใจที่ปรับแล้วสำหรับตัวแปรอิสระหลายตัว

เนื่องจากข้อเสียของ  $R_p^2$  คือ เมื่อจำนวนพจน์ในตัวแบบเพิ่มขึ้นค่า  $R_p^2$  จะเพิ่มขึ้นด้วยดังนั้นนักสถิติจึงคิดค้นสัมประสิทธิ์การตัดสินใจที่ปรับแล้ว (adjusted coefficient of multiple determination) สำหรับตัวแปรอิสระหลายตัว หรือ  $R_{adj}^2(p)$  โดยการเพิ่มจำนวนพจน์ในตัวแบบอาจไม่จำเป็นที่ค่า  $R_{adj}^2(p)$  เพิ่มขึ้นตัวแบบที่มีความเหมาะสมที่สุดคือตัวแบบที่มีค่า  $R_{adj}^2(p)$  สูงสุดโดยมีสูตรการคำนวณดังนี้

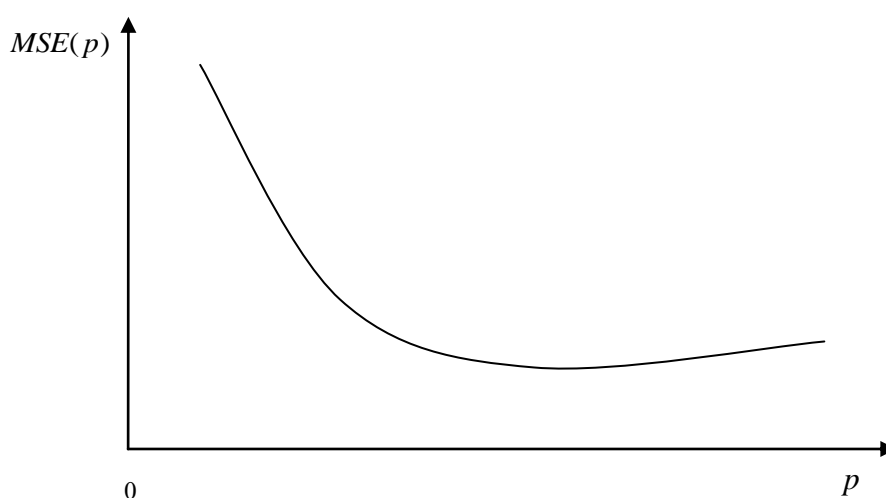
$$R_{adj}^2(p) = 1 - \left( \frac{n-1}{n-p} \right) \frac{SSE}{SST} = 1 - \left( \frac{n-1}{n-p} \right) (1 - R_p^2) \quad (8.2)$$

### 8.1.3 ค่าเฉลี่ยกำลังสองความคลาดเคลื่อน

ค่าเฉลี่ยกำลังสองความคลาดเคลื่อน (error mean square) หรือ  $MSE(p)$  ที่ได้จากตัวแบบที่มีตัวแปร  $p$  ตัวสามารถคำนวณได้ดังนี้

$$MSE(p) = \frac{SSE(p)}{n-p} \quad (8.3)$$

ข้อเสียของ  $MSE(p)$  คือ เมื่อจำนวนพจน์ในตัวแบบเพิ่มขึ้นค่า  $MSE(p)$  จะลดลงเรื่อยๆ จากนั้นจะคงที่แล้วเพิ่มขึ้นอีกครั้งดังภาพ 8.2 (Montgomery & Peck, 1991, p. 270-271)



ภาพที่ 8.2 ความสัมพันธ์ระหว่างจำนวนตัวแปรอิสระกับค่า  $MSE(p)$

การเลือกตัวแบบที่ดีที่สุดทำการเลือกตัวแบบที่มีค่า  $MSE(p)$  ที่ต่ำที่สุดโดยตัวแบบที่ได้นี้จะให้ค่า  $R_{adj}^2(p)$  ที่สูงสุดด้วยเนื่องจาก  $R_{adj}^2(p) = 1 - \frac{n-1}{S_{yy}} MSE(p)$  ดังนั้นสามารถเลือกใช้

$R_{adj}^2(p)$  สูงสุดหรือค่า  $MSE(p)$  ที่ต่ำสุดก็ได้

หากใช้โปรแกรม MINITAB ช่วยในการสร้างตัวแบบผลลัพธ์ที่ได้จะแสดงทั้ง  $MSE(p)$  และ  $S$  (โดย  $S = \sqrt{MSE(p)}$ )

### 8.1.4 $C_p$

ตัวแบบที่มีจำนวนพจน์ในตัวแบบเท่ากับ  $p$  พจน์ สามารถเขียนสมการถดถอยได้ดังนี้

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{p-1} X_{p-1} + \varepsilon$$

จากตัวแบบข้างต้นจะได้ค่า  $SSE_p$  และ Mallows'  $C_p$  หรือ  $C_p$  สามารถคำนวณได้ดังนี้

$$C_p = \frac{SSE_p}{MSE} - n + 2p \quad (8.4)$$

โดย  $MSE$  เป็นค่าเฉลี่ยกำลังสองคลาดเคลื่อนของตัวแบบเต็มรูป

$MSE$  ของตัวแบบเต็มรูปเป็นตัวประมาณค่าที่ไม่เอนเอียงของ  $\sigma^2$  แต่หากในตัวแบบเต็มรูปที่มีตัวแปรอิสระบางตัวที่ไม่มีความจำเป็นในการพยากรณ์แล้วค่า  $MSE$  จะให้ค่าประมาณที่ให้ค่าประมาณที่สูงเกินจริงส่งผลให้ค่า  $C_p$  ต่ำเกินความเป็นจริง หากค่า  $C_p$  มีค่าเท่ากับ  $p$  แล้วค่า  $C_p$  ที่ไม่เอนเอียง จากสูตรที่ (8.4) พบว่า  $C_p$  เป็นฟังก์ชันของค่า  $SSE$  เนื่องจากตัวแบบที่ดีนั้นเป็นตัวแบบที่มีค่า  $SSE$  ต่ำและตัวแบบนั้นจะให้ค่า  $C_p$  ต่ำด้วย

ตัวแบบที่ดีสามารถเลือกโดยดูจากค่า  $C_p$  ที่มีขนาดเล็กและมีค่าใกล้เคียงกับจำนวนตัวแปรอิสระทั้งหมด Stevens (2007) หน้า 238 กล่าวว่า  $C_p$  เป็นสถิติที่ดีสำหรับใช้ในการเลือกตัวแบบเนื่องจากจะป้องกันการเกิด overfitting และ underfitting ของตัวแบบ

### 8.1.5 $PRESS_p$

ค่า prediction sum of square หรือ  $PRESS_p$  เป็นค่าที่ใช้วัดว่าตัวแบบที่มีให้ค่าพยากรณ์ที่ใกล้เคียงค่าจริงมากน้อยเพียงใด โดยค่า  $PRESS_p$  แตกต่างจากค่า  $SSE$  ตรงที่  $SSE$  ได้จากค่า  $\hat{y}_i$  แต่ค่า  $PRESS_p$  เกิดจากผลรวมกำลังสองของความคลาดเคลื่อนระหว่างค่าจริง ( $y_i$ ) กับค่าพยากรณ์ที่ไม่รวมค่าจริงที่  $i$  โดยใช้ข้อมูลเพียง  $n - 1$  ตัว ( $\hat{y}_{i(i)}$ ) หรือสามารถแสดงได้ดังนี้

$$PRESS_p = \sum_{i=1}^n (y_i - \hat{y}_{i(i)})^2 = \sum_{i=1}^n \left( \frac{e_i}{1 - h_{ii}} \right)^2 \quad (8.5)$$

โดย  $h_{ii} = x_i'(X'X)^{-1}x_i$

เช่นเดียวกับ  $SSE$  ตัวแบบที่เหมาะสมเป็นตัวแบบที่มีค่า  $PRESS_p$  ต่ำที่สุดเนื่องจากค่า  $PRESS_p$  เป็นความคลาดเคลื่อนระหว่างค่าจริงกับค่าพยากรณ์หาก  $PRESS_p$  มีค่าน้อยแสดงว่าการพยากรณ์มีความถูกต้องสูง

ตัวอย่าง 8.1 จากข้อมูลข้างล่างจงเลือกตัวแบบที่เหมาะสมโดยใช้  $R_p^2$ ,  $R_{adj}^2(p)$ ,  $MSE(p)$ ,  $C_p$  และ  $PRESS_p$

$i$	$X_1$	$X_2$	$X_3$	$X_4$	$Y$
1	9.56	1	0.33	0.35	3.53
2	5.65	0	0.13	0.38	4.74
3	7.87	1	0.14	0.59	2.26
4	7.55	0	0.13	0.45	3.76
5	8.71	1	0.12	0.48	2.91
6	8.09	1	0.35	0.29	3.31
7	10.80	1	0.14	0.36	3.96
8	7.70	1	0.12	0.46	3.31
9	7.27	1	0.12	0.48	3.44
10	6.31	0	0.10	0.54	4.02
11	9.00	0	0.14	0.41	3.05
12	10.50	1	0.14	0.43	3.86
13	9.57	1	0.16	0.44	3.45
14	14.00	1	0.38	0.42	2.36
15	6.95	0	0.25	0.55	3.00
16	8.90	1	0.43	0.35	2.84
17	9.50	1	0.35	0.49	2.28
18	10.20	1	0.31	0.48	3.10
19	8.40	1	0.34	0.44	2.71
20	8.44	0	0.12	0.50	2.94
21	8.50	1	0.14	0.58	2.34
22	8.42	1	0.45	0.40	2.85
23	9.38	0	0.15	0.49	2.41
24	10.80	1	0.15	0.54	2.53
25	7.55	0	0.23	0.45	3.47

## วิธีทำ

ในที่นี้จะแสดงการคำนวณค่า  $R_{adj}^2(p)$ ,  $C_p$  และ  $PRESS_p$  เฉพาะตัวแบบที่ 5 ที่มีตัวแปรอิสระ  $X_1$  และ  $X_2$  ดังนี้

$$\begin{aligned} R_{adj}^2(p) &= 1 - \left( \frac{n-1}{n-p} \right) (1 - R_p^2) \\ &= 1 - \left( \frac{25-1}{25-3} \right) (1 - 0.184) \\ &= 0.110 \end{aligned}$$

$$\begin{aligned} C_p &= \frac{SSE_p}{MSE} - n + 2p \\ &= \frac{7.8047}{0.1671} - 25 + (2 \times 3) \\ &= 27.7 \end{aligned}$$

$$\begin{aligned} PRESS_p &= \sum_{i=1}^n \left( \frac{e_i}{1-h_{ii}} \right)^2 \\ &= \left( \frac{0.5860}{1-0.0597} \right)^2 + \dots + \left( \frac{0.0430}{1-0.1251} \right)^2 \\ &= 10.0493 \end{aligned}$$

ผลลัพธ์ทั้งหมดสามารถแสดงได้โดยตารางข้างล่าง

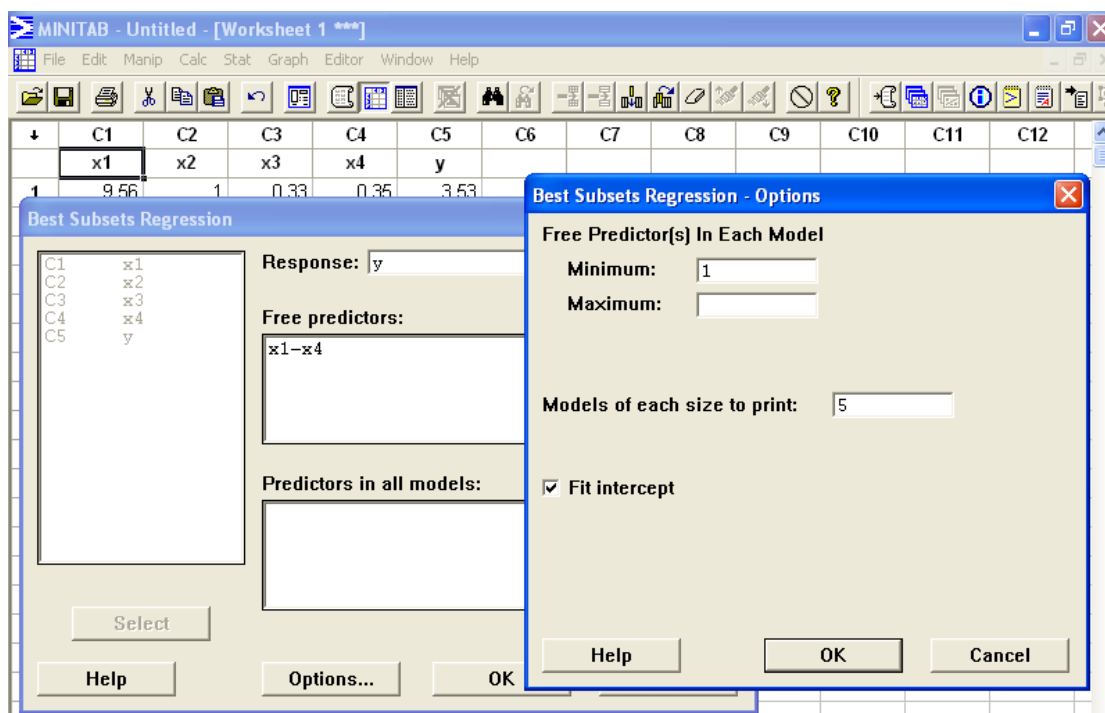
ตัวแบบ	ตัวแปรอิสระ	$R_p^2$	$R_{adj}^2(p)$	$MSE(p)$	$C_p$	$PRESS_p$
1	$x_1$	0.165	0.128	0.3473	26.8	9.6273
2	$x_2$	0.101	0.620	0.3738	30.4	10.3757
3	$x_3$	0.112	0.730	0.3693	29.8	9.7750
4	$x_4$	0.179	0.144	0.3412	26.0	9.2969
5	$x_1$ $x_2$	0.184	0.110	0.3548	27.7	10.0493
6	$x_1$ $x_3$	0.213	0.142	0.3419	26.0	9.5428
7	$x_1$ $x_4$	0.410	0.356	0.2564	14.8	7.6029

ตัวแบบ	ตัวแปรอิสระ	$R_p^2$	$R_{adj}^2(p)$	$MSE(p)$	$C_p$	$PRESS_p$
8	$x_2$ $x_3$	0.155	0.780	0.3673	29.4	10.2832
9	$x_2$ $x_4$	0.332	0.272	0.2902	19.2	8.3209
10	$x_3$ $x_4$	0.534	0.492	0.2025	7.7	5.7278
11	$x_1$ $x_2$ $x_3$	0.220	0.109	0.3552	27.6	10.2467
12	$x_1$ $x_2$ $x_4$	0.445	0.366	0.2526	14.7	7.5266
<b>13</b>	<b><math>x_1</math> <math>x_3</math> <math>x_4</math></b>	<b>0.646</b>	<b>0.596</b>	<b>0.1611</b>	<b>3.2</b>	<b>4.8645</b>
14	$x_2$ $x_3$ $x_4$	0.574	0.513	0.1941	7.4	5.8049
15	$x_1$ $x_2$ $x_3$ $x_4$	0.650	0.581	0.1671	5.0	5.2672

จากตารางข้างต้นจะพบว่าตัวแบบที่ประกอบด้วยตัวแปร  $X_1$   $X_3$  และ  $X_4$  เป็นตัวแบบที่เหมาะสมที่สุดโดยการใช้  $R_p^2$ ,  $R_{adj}^2(p)$ ,  $MSE(p)$ ,  $C_p$  และ  $PRESS_p$  เป็นเกณฑ์การคัดเลือก เนื่องจากมีค่า  $MSE(p)$ ,  $C_p$  และ  $PRESS_p$  ต่ำที่สุดและมีค่า  $R_{adj}^2(p)$  สูงที่สุดและ  $C_p$  ที่มีค่าใกล้เคียงกับจำนวนตัวแปรที่สุด

#### หมายเหตุ

1. หากใช้โปรแกรม MINITAB ช่วยในการคำนวณค่า  $R_p^2$ ,  $R_{adj}^2(p)$  และ  $C_p$  สามารถทำได้ดังนี้
  - (1) เลือก “Stat” ที่เมนูบาร์
  - (2) เลือก “Regression”
  - (3) เลือก “Best Subsets...”
  - (4) ระบุตัวแปรตามใน “Response:” และตัวแปรอิสระใน “Free predictors:” หากต้องการให้ตัวแปรอิสระใดอยู่ในตัวแบบทุกตัวแบบให้ระบุชื่อตัวแบบนั้นใน “Predictors in all models:”
  - (5) หากต้องการสร้างตัวแบบเพียงบางส่วนจากตัวแปรอิสระที่มีทั้งหมดให้คลิกเข้าไปที่ “Options...” จากนั้นระบุจำนวนตัวแปรอิสระที่ต้องการในช่อง “Maximum:” จากนั้นคลิก “OK”
  - (6) ผลลัพธ์ที่ได้จะแสดงในหน้าจอ “sessions” ดังภาพที่ 8.3



ภาพที่ 8.3 หน้าจอการคำนวณค่า  $R_p^2$ ,  $R_{adj}^2(p)$  และ  $C_p$

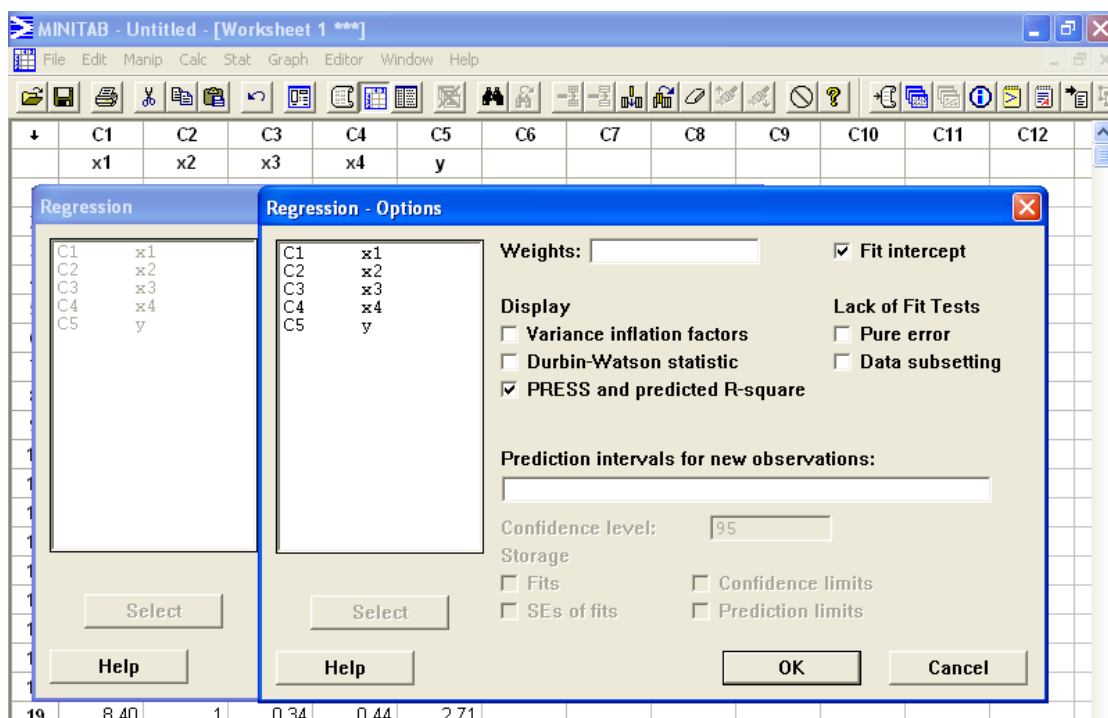
2. หากใช้โปรแกรม MINITAB ช่วยในการคำนวณค่า  $PRESS_p$  สามารถทำได้ดังนี้

- (1) เลือก “Stat” ที่เมนูบาร์
- (2) เลือก “Regression”
- (3) เลือก “Regression...”
- (4) ระบุตัวแปรตามใน “Response:” และตัวแปรอิสระใน “Predictors:”
- (5) คลิก “Options” จากนั้นคลิกเลือก “PRESS and predicted R-Square” จากนั้นคลิก

“OK”

- (6) ผลลัพธ์ที่ได้จะแสดงในหน้าจอ “sessions” ดังภาพที่ 8.4





ภาพที่ 8.4 หน้าจอการคำนวณค่า  $PRESS_p$

## 8.2 เทคนิคการเลือกตัวแบบที่เหมาะสม

เกณฑ์การเลือกตัวแบบที่เหมาะสมเป็นส่วนหนึ่งที่จะช่วยนักสถิติในการเลือกตัวแบบ แต่เทคนิคที่ใช้ในการเลือกก็มีด้วยกันหลายวิธี บางเทคนิคอาจให้ผลลัพธ์ที่แตกต่างกันออกไปและแต่ละเทคนิคมีข้อดีข้อเสียที่แตกต่างกันออกไป

### 8.2.1 การพิจารณาทุกตัวแบบ

การพิจารณาทุกตัวแบบ (All possible subsets) เป็นเทคนิคที่สร้างตัวแบบโดยใช้สร้างตัวแบบจากตัวแปรอิสระทุกตัวทีละ 1 ตัวแปรจากนั้นเพิ่มเป็น 2 ตัวแปรเพิ่มไปเรื่อยๆ จนครบจำนวนตัวแปรทั้งหมดที่มี แล้วเลือกตัวแบบที่มีค่าของสถิติที่ดีที่สุดตามเกณฑ์ในหัวข้อ 8.1 หากมีตัวแปรอิสระอยู่  $K$  ตัวจะมีตัวแบบที่สามารถเป็นไปได้ทั้งหมดจำนวน  $2^K$  ตัวแบบ เช่น หากมีจำนวนตัวแปรอิสระอยู่ 4 ตัวจะมีตัวแบบทั้งหมดอยู่  $2^4$  ตัวแบบหรือ 16 ตัวแบบและหากมีตัวแปรอิสระจำนวน 12 ตัวจะมีจำนวนตัวแบบที่เป็นไปได้ทั้งหมดอยู่  $2^{12}$  ตัวแบบหรือ 4096 ตัวแบบ ดังนั้นจะเห็นว่าวิธีนี้จะยุ่งยากและเสียเวลามากเมื่อมีจำนวนตัวแปรอิสระมาก นักวิจัยสามารถใช้โปรแกรม MINITAB ในการสร้างตัวแบบทุกตัวแบบได้โดยใช้เกณฑ์การเลือกคือ  $R_p^2$ ,  $R_{adj}^2(p)$ ,  $C_p$  และ  $S$  จากตัวแบบที่มีทั้งหมด นักวิจัยสามารถเลือกตัวแบบที่ดีที่สุดโดยใช้เกณฑ์ใดเกณฑ์หนึ่งหรือ

ใช้ทุกเกณฑ์ประกอบการตัดสินใจ ขั้นตอนการคำนวณโดยใช้ MINITAB นั้นทำเช่นเดียวกับการคำนวณในหมายเหตุที่ 1 ของหัวข้อ 8.1

**ตัวอย่างที่ 8.2** จากข้อมูลในตัวอย่างที่ 8.1 จงใช้เทคนิคการพิจารณาทุกตัวแบบในการเลือกตัวแบบวิธีทำ

เนื่องจากจำนวนตัวแปรอิสระทั้งหมดเท่ากับ 4 ตัว ดังนั้นจำนวนตัวแบบที่เป็นไปได้ทั้งหมดเท่ากับ  $2^4$  ตัวแบบหรือ 16 ตัวแบบผลลัพธ์ที่ได้แสดงดังรูปจาก โปรแกรม MINITAB ข้างล่างโดยมีรายละเอียดดังนี้

คอลัมน์ที่ 1 แสดงจำนวนตัวแปรอิสระในตัวแบบ

คอลัมน์ที่ 2 แสดงค่า  $R_p^2$  ในแต่ละตัวแบบ

คอลัมน์ที่ 3 แสดงค่า  $R_{adj}^2(p)$  ในแต่ละตัวแบบ

คอลัมน์ที่ 4 แสดงค่า  $C_p$  ในแต่ละตัวแบบ

คอลัมน์ที่ 5 แสดงค่า  $S$  ในแต่ละตัวแบบ

คอลัมน์ที่ 6 แสดงตัวแปรอิสระที่ใช้ในการสร้างตัวแบบ

จากผลลัพธ์ที่ได้จะเห็นว่าตัวแบบที่ประกอบด้วย  $X_1$ ,  $X_3$  และ  $X_4$  เป็นตัวแบบที่ดีที่สุด เนื่องจากมีค่า  $R_{adj}^2(p)$  สูงสุดและ  $C_p$  มีค่าใกล้เคียงกับจำนวนตัวแปรที่สุคนอกจากนี้ยังมีค่า  $S$  ที่ต่ำที่สุด

Vars	R-Sq	R-Sq(adj)	C-p	S	x x x x			
					1	2	3	4
1	17.9	14.4	26.0	0.58409				X
1	16.5	12.8	26.8	0.58936	X			
1	11.2	7.3	29.8	0.60772			X	
1	10.1	6.2	30.4	0.61137		X		
2	53.4	49.2	7.7	0.44998		X	X	
2	41.0	35.6	14.8	0.50637	X		X	
2	33.2	27.2	19.2	0.53867		X	X	
2	21.3	14.2	26.0	0.58475	X	X		
2	18.4	11.0	27.7	0.59562	X	X		
2	15.5	7.8	29.4	0.6061		X	X	
3	64.6	59.6	3.2	0.40134	X	X	X	
3	57.4	51.3	7.4	0.44053		X	X	X
3	44.5	36.6	14.7	0.50264	X	X	X	
3	22.0	10.9	27.6	0.59596	X	X	X	
4	65.0	58.1	5.0	0.40880	X	X	X	X

## 8.2.2 การเพิ่มตัวแปรอิสระ

การเพิ่มตัวแปรอิสระ (forward selection) เป็นเทคนิคที่ทำโดยการสร้างตัวแบบจากตัวแปรอิสระ 1 ตัวโดยเลือกจากตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามมากที่สุดก่อนโดยอาจพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์โดยตัวแปรอิสระตัวแรกที่ใส่เข้าไปในตัวแบบจะให้ค่า  $F$  ที่สูงที่สุด ค่า  $F$  นี้จะต้องมีค่ามากกว่าเกณฑ์ที่ตั้งไว้ ( $F$ -to-enter) โดยนักวิจัยจะเป็นผู้กำหนดค่า  $F$ -to-enter หากค่า  $F$  ที่ได้มีค่าน้อยกว่า  $F$ -to-enter แล้วแสดงว่าไม่มีตัวแปรใดที่สามารถพยากรณ์ตัวแปรตามได้ จากนั้นสร้างตัวแบบที่สองโดยการเพิ่มตัวแปรอิสระอีกตัวที่มีความสัมพันธ์กับตัวแปรตามสูงเป็นอันดับสอง จากนั้นเปรียบเทียบค่า  $F$  ที่ได้จากตัวแบบนี้กับค่า  $F$ -to-enter หากค่า  $F$  ที่ได้มีค่ามากกว่าค่า  $F$ -to-enter แสดงว่าตัวแปรอิสระตัวที่สองนี้สามารถช่วยในการพยากรณ์ตัวแปรตามได้ จากนั้นเพิ่มตัวแปรอิสระเข้าไปในตัวแบบทีละตัว หากตัวแปรใดที่เพิ่มเข้าไปแล้วให้ค่า  $F$  ที่น้อยกว่าค่า  $F$ -to-enter แสดงว่าตัวแปรอิสระนั้นไม่สามารถช่วยในการพยากรณ์ตัวแปรตามให้หยุดการค้นหาและไม่รวมตัวแปรนั้นเข้าในตัวแบบ วิธีการเพิ่มตัวแปรอิสระนี้เมื่อตัวแปรใดถูกเลือกเข้าในตัวแบบแล้วจะไม่มีมีการพิจารณาตัดออกในภายหลัง

การเลือกค่า  $F$ -to-enter นั้นขึ้นอยู่กับนักวิจัย หากตั้งค่าต่ำมากจะทำให้มีจำนวนตัวแปรอิสระในตัวแบบมากแต่หากตั้งค่าสูงมากจะทำให้มีจำนวนตัวแบบไม่มาก ค่า  $F$ -to-enter ที่เป็นที่ยอมรับของนักวิจัยอยู่ระหว่าง 2.0 - 4.0 โปรแกรม SPSS กำหนดค่า  $F$ -to-enter เท่ากับ 3.84 ให้เป็นค่าทั่วไป (default)

### หมายเหตุ

นอกจากการใช้ค่า  $F$ -to-enter แล้วนักวิจัยยังสามารถใช้ค่า  $p$ -value ที่สอดคล้องกับค่า  $F$ -to-enter เป็นตัวเปรียบเทียบได้โดยเปรียบเทียบกับค่า  $\alpha$  หากค่า  $p$ -value ที่ได้มีน้อยกว่าค่า  $\alpha$  แล้วตัวแปรอิสระตัวนั้นสามารถช่วยพยากรณ์ตัวแปรตามนั้นได้ โปรแกรม MINITAB กำหนดค่า  $\alpha$  เท่ากับ 0.25 และโปรแกรม MINITAB แสดงค่า  $t$  และค่า  $p$ -value ของแต่ละตัวแปรในแต่ละขั้นของการคำนวณ

หากใช้โปรแกรม MINITAB ช่วยในการหาตัวแบบที่เหมาะสมโดยวิธีการเพิ่มตัวแปรอิสระสามารถทำได้ดังนี้

- (1) เลือก “Stat” ที่เมนูบาร์
- (2) เลือก “Regression”
- (3) เลือก “Stepwise...”

(4) ระบุตัวแปรตามใน “Response:” และตัวแปรอิสระใน “Predictors:” หากต้องการให้ตัวแปรอิสระอยู่ในตัวแบบทุกตัวแบบให้ระบุชื่อตัวแบบนั้นใน “Predictors to include in every model:”

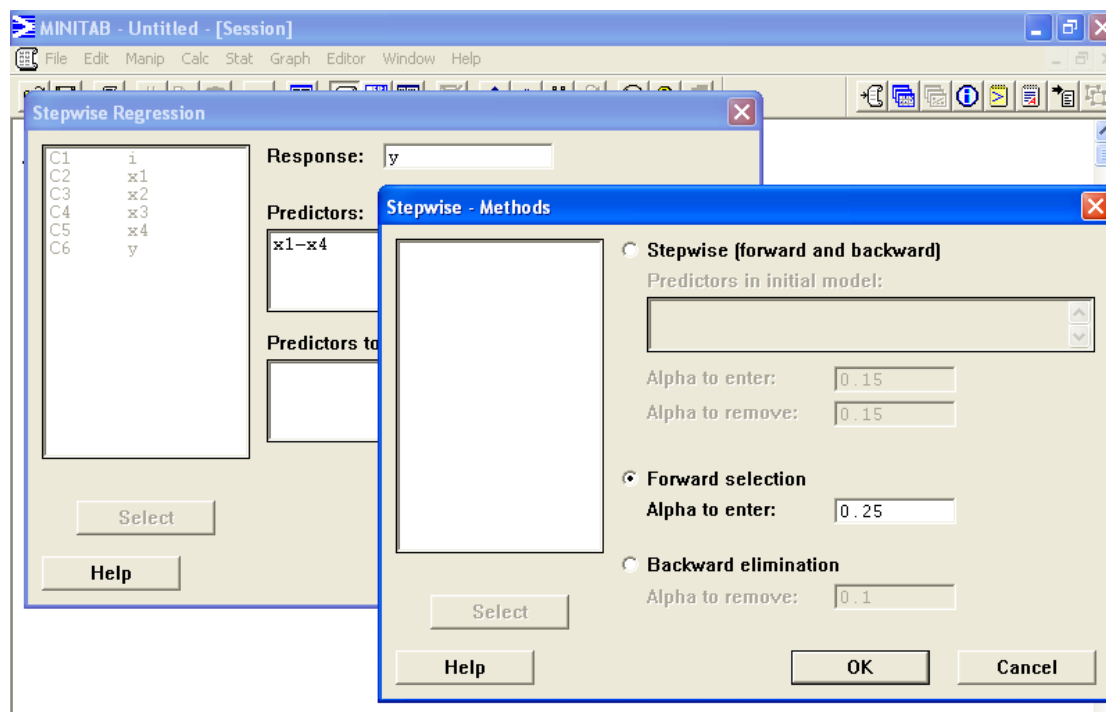
(5) คลิก “Methods...” จากนั้นคลิก “Forward selection” และระบุค่า “Alpha to enter:” หรือค่า  $\alpha$  หากไม่ระบุค่าแล้วโปรแกรมจะระบุค่าให้เท่ากับ 0.25

(6) หากต้องการให้แสดงตัวแบบที่ดีที่สุดมากกว่า 1 ตัวแบบให้คลิก “Options” จากนั้นระบุจำนวนตัวแบบที่ดีที่สุดลงในช่อง “Number of alternative predictors to show:” เช่น 2 หมายถึงต้องการให้แสดงตัวแบบที่ดีที่สุด 2 ตัวแบบแรก เป็นต้น

(7) หากต้องการให้ทำการคำนวณมากกว่า 1 ครั้งในแต่ละครั้งของการเลือกตัวแปรอิสระให้คลิก “Options” จากนั้นระบุจำนวนครั้งของการคำนวณที่ “Number of steps between pauses:”

(8) หากต้องการให้แสดงค่า PRESS และ predicted R-Square ของแต่ละ “PRESS and predicted R-Square” ให้คลิก “Options” จากนั้นคลิก “Display PRESS and predicted R-Square” จากนั้นคลิก “OK”

(9) ผลลัพธ์ที่ได้จะแสดงในหน้าจอ “sessions” ดังภาพที่ 8.5



ภาพที่ 8.5 หน้าจอการคำนวณโดยวิธีการเพิ่มตัวแปรอิสระ

ตัวอย่างที่ 8.3 จากตัวอย่างที่ 8.1 จงสร้างตัวแบบที่เหมาะสมกับข้อมูลชุดนี้ด้วยโปรแกรม MINITAB โดยใช้วิธีการเพิ่มตัวแปรอิสระที่ใช้ค่า  $\alpha$  เท่ากับ 0.10

### วิธีทำ

เนื่องจากกำหนดค่า  $\alpha$  เท่ากับ 0.10 ดังนั้นให้ระบุค่า Alpha to enter เท่ากับ 0.10 ในขั้นที่ 1 ตัวแปรอิสระที่ถูกเลือกเข้าสู่ตัวแบบคือ  $X_4$  ในขั้นที่ 2 ตัวแปรอิสระที่ถูกเลือกเข้าสู่ตัวแบบคือ  $X_3$  และขั้นสุดท้ายตัวแปรอิสระที่ถูกเลือกเข้าสู่ตัวแบบคือ  $X_1$  และตัวแบบที่ดีที่สุดที่ได้จากโปรแกรมประกอบด้วยตัวแปร  $X_1$   $X_3$  และ  $X_4$  โดยมีค่า  $S$  เท่ากับ 0.401 ค่า  $R_p^2$  เท่ากับ 64.63 ค่า  $R_{adj}^2(p)$  เท่ากับ 59.58 และค่า  $C_p$  เท่ากับ 3.2 ซึ่งสอดคล้องกับผลลัพธ์ที่ได้จากตัวอย่างที่ 8.2 ผลลัพธ์ที่ได้แสดงดังภาพข้างล่าง ตัวแบบที่ได้มีสมการถดถอยคือ

$$\hat{y} = 7.796 - 0.130x_1 - 3.16x_3 - 6.2x_4$$

MINITAB - Untitled - [Session]

File Edit Manip Calc Stat Graph Editor Window Help

Stepwise Regression: y versus x1, x2, x3, x4

Forward selection. Alpha-to-Enter: 0.1

Response is y on 4 predictors, with N = 25

Step	1	2	3
Constant	4.756	6.744	7.796
x4	-3.6	-6.2	-6.2
T-Value	-2.24	-4.47	-5.07
P-Value	0.035	0.000	0.000
x3		-3.74	-3.16
T-Value		-4.09	-3.74
P-Value		0.000	0.001
x1			-0.130
T-Value			-2.58
P-Value			0.017
S	0.584	0.450	0.401
R-Sq	17.95	53.42	64.63
R-Sq(adj)	14.38	49.18	59.58
C-p	26.0	7.7	3.2

### 8.2.3 การลดตัวแปรอิสระ

การลดตัวแปรอิสระ (backward elimination) เป็นเทคนิคที่ทำตรงข้ามกับวิธีการเพิ่มตัวแปรอิสระโดยการสร้างตัวแบบจากตัวแปรอิสระทุกตัวเช่น มีตัวแปรอิสระทั้งหมด 4 ตัวคือ  $X_1$   $X_2$   $X_3$  และ  $X_4$  จากนั้นทำการตัดตัวแปรอิสระออก 1 ตัวจากตัวแบบนี้ดังนั้นตัวแบบใหม่จะมีทั้งหมด 4 ตัวแบบคือ (1)  $X_1$   $X_2$  และ  $X_3$  (2)  $X_1$   $X_2$  และ  $X_4$  (3)  $X_1$   $X_3$  และ  $X_4$  (4)  $X_2$   $X_3$  และ

$X_4$  แล้วเลือกตัวแบบที่ให้ค่าให้  $F$  ที่มีค่าต่ำกว่าค่า  $F$ -to-remove ที่น้อยที่สุดสมมติตัวแบบที่ (1) ให้ค่า  $F$  ที่มีค่าต่ำกว่าค่า  $F$ -to-remove ที่น้อยที่สุดจึงเลือกตัวแบบที่ (1) นั่นคือตัดตัวแปรอิสระ  $X_4$  ออกจากตัวแบบ จากนั้นตัดตัวแปรอิสระอีก 1 ตัวออกจากตัวแบบที่ (1) แล้วคำนวณค่า  $F$  ใหม่จากตัวแบบใหม่ที่ได้โดยการทำเช่นนี้เรื่อยไปจนไม่มีตัวแปรใดที่มีค่า  $F$  ที่มีค่าต่ำกว่าค่า  $F$ -to-remove จึงหยุด

การเลือกค่า  $F$ -to-remove นั้นขึ้นอยู่กับนักวิจัยเช่นเดียวกับ  $F$ -to-enter โปรแกรม SPSS กำหนดค่า  $F$ -to-remove เท่ากับ 2.71 ให้เป็นค่าทั่วไป (default)

### หมายเหตุ

นอกจากการใช้ค่า  $F$ -to-remove แล้วนักวิจัยยังสามารถใช้ค่า  $p$ -value ที่สอดคล้องกับค่า  $F$ -to-remove เป็นตัวเปรียบเทียบได้โดยเปรียบเทียบกับค่า  $\alpha$  หากค่า  $p$ -value ที่ได้มีน้อยกว่าค่า  $\alpha$  แล้วตัวแปรอิสระตัวนั้นสามารถช่วยพยากรณ์ตัวแปรตามนั้นได้ โปรแกรม MINITAB กำหนดค่า  $\alpha$  เท่ากับ 0.10 และโปรแกรม MINITAB แสดงค่า  $t$  และค่า  $p$ -value ของแต่ละตัวแปรในแต่ละขั้นของการคำนวณ

หากใช้โปรแกรม MINITAB ช่วยในการหาตัวแบบที่เหมาะสมโดยวิธีการลดตัวแปรอิสระสามารถทำได้ดังนี้

- (1) เลือก “Stat” ที่เมนูบาร์
- (2) เลือก “Regression”
- (3) เลือก “Stepwise...”

(4) ระบุตัวแปรตามใน “Response:” และตัวแปรอิสระใน “Predictors:” หากต้องการให้ตัวแปรอิสระใดอยู่ในตัวแบบทุกตัวแบบให้ระบุชื่อตัวแบบนั้นใน “Predictors to include in every model:”

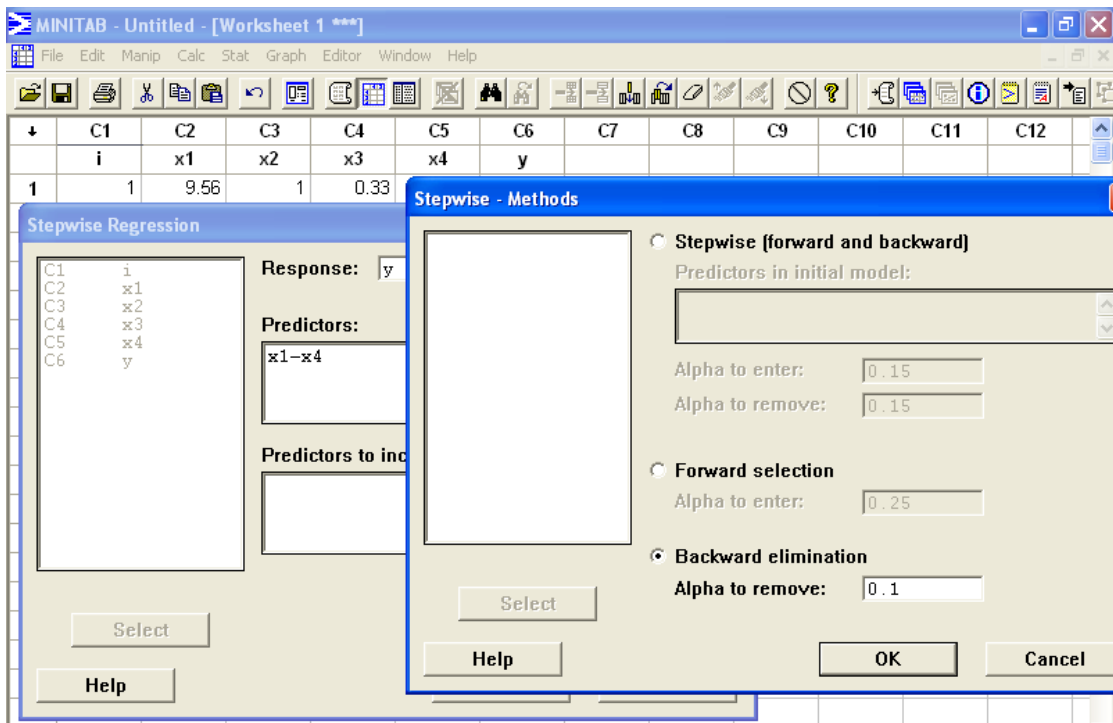
(5) คลิก “Methods...” จากนั้นคลิก “Backward selection” และระบุค่า “Alpha to remove:” หรือค่า  $\alpha$  หากไม่ระบุค่าแล้วโปรแกรมจะระบุค่าให้เท่ากับ 0.10

(6) หากต้องการให้แสดงตัวแบบที่ดีที่สุดมากกว่า 1 ตัวแบบให้คลิก “Options” จากนั้นระบุจำนวนตัวแบบที่ดีที่สุดลงในช่อง “Number of alternative predictors to show:” เช่น 2 หมายถึงต้องการให้แสดงตัวแบบที่ดีที่สุด 2 ตัวแบบแรก เป็นต้น

(7) หากต้องการให้ทำการคำนวณมากกว่า 1 ครั้งในแต่ละครั้งของการเลือกตัวแปรอิสระให้คลิก “Options” จากนั้นระบุจำนวนครั้งของการคำนวณที่ “Number of steps between pauses:”

(8) หากต้องการให้แสดงค่า PRESS และ predicted R-Square ของแต่ละ “PRESS and predicted R-Square” ให้คลิก “Options” จากนั้นคลิก “Display PRESS and predicted R-Square” จากนั้นคลิก “OK”

(9) ผลลัพธ์ที่ได้จะแสดงในหน้าจอ “sessions” ดังภาพที่ 8.6



ภาพที่ 8.6 หน้าจอการคำนวณโดยวิธีการลดตัวแปรอิสระ

ตัวอย่างที่ 8.4 จากตัวอย่างที่ 8.1 จงสร้างตัวแบบที่เหมาะสมกับข้อมูลชุดนี้ด้วยโปรแกรม MINITAB โดยใช้วิธีการลดตัวแปรอิสระที่ใช้ค่า  $\alpha$  เท่ากับ 0.10

วิธีทำ

เนื่องจากกำหนดค่า  $\alpha$  เท่ากับ 0.10 ดังนั้นให้ระบุค่า Alpha to remove เท่ากับ 0.10 ในขั้นที่ 1 สร้างตัวแบบจากตัวแปรอิสระทั้งหมด ในขั้นที่ 2 ซึ่งเป็นขั้นสุดท้ายตัวแปรอิสระที่ถูกตัดทิ้งออกจากตัวแบบคือ  $X_2$  ดังนั้นตัวแบบที่ดีที่สุดที่ได้จากโปรแกรมประกอบด้วยตัวแปร  $X_1$ ,  $X_3$  และ  $X_4$  โดยมีค่า  $S$  เท่ากับ 0.401 ค่า  $R_p^2$  เท่ากับ 64.63 ค่า  $R_{adj}^2(p)$  เท่ากับ 59.58 และค่า  $C_p$  เท่ากับ 3.2 ซึ่งสอดคล้องกับผลลัพธ์ที่ได้จากตัวอย่างที่ 8.2 และ 8.3 ผลลัพธ์ที่ได้แสดงดังภาพข้างล่าง ตัวแบบที่ได้มีสมการถดถอยคือ

$$\hat{y} = 7.796 - 0.130x_1 - 3.16x_3 - 6.2x_4$$

Backward elimination. Alpha-to-Remove: 0.1

Response is y on 4 predictors, with N = 25

Step	1	2
Constant	7.732	7.796
x1	-0.118	-0.130
T-Value	-2.09	-2.58
P-Value	0.049	0.017
x2	-0.10	
T-Value	-0.49	
P-Value	0.630	
x3	-3.05	-3.16
T-Value	-3.43	-3.74
P-Value	0.003	0.001
x4	-6.2	-6.2
T-Value	-4.96	-5.07
P-Value	0.000	0.000
S	0.409	0.401
R-Sq	65.05	64.63
R-Sq(adj)	58.06	59.58

## 8.2.4 การเพิ่มตัวแปรอิสระแบบขั้นตอน

การเพิ่มตัวแปรอิสระแบบขั้นตอน (stepwise regression) เป็นเทคนิคที่คัดแปลงมาจากวิธีการเพิ่มตัวแปรอิสระโดยมีการพิจารณาตัดตัวแปรอิสระที่มีอยู่ในตัวแบบออกหลังจากเพิ่มตัวแปรอิสระตัวใหม่เข้าไปในตัวแบบทั้งนี้เนื่องจากตัวแปรอิสระเดิมที่มีอยู่ในตัวแบบอาจมีความสัมพันธ์กับตัวแปรอิสระตัวใหม่ที่เข้าไปทำให้ไม่มีความจำเป็นในการเก็บตัวแปรเดิมไว้ในตัวแบบ หากตัวแปรอิสระตัวใดมีค่า  $F$  ที่น้อยกว่าค่า  $F$ -to-remove และให้ค่าน้อยที่สุดตัวแปรนั้นจะถูกตัดออกจากตัวแบบและตัวแปรอิสระตัวใดที่มีค่า  $F$  ที่มากกว่าค่า  $F$ -to-enter และให้ค่ามากที่สุดตัวแปรนั้นจะถูกรวมเข้าสู่ตัวแบบ ทำเช่นนี้เรื่อยไปจนไม่สามารถเพิ่มตัวแปรใหม่เข้ามาในตัวแบบได้หรือไม่สามารถตัดตัวแปรที่มีอยู่ในตัวแบบได้จึงหยุด ดังนั้นเทคนิคนี้จะมีการใช้ทั้งค่า  $F$ -to-enter และ  $F$ -to-remove ส่วนใหญ่มักกำหนดค่า  $F$ -to-enter มากกว่า  $F$ -to-remove เพื่อให้การเพิ่มตัวแปรอิสระในตัวแบบทำได้ยากกว่าการลดจำนวนตัวแปรอิสระในตัวแบบ แต่นักวิจัยบางท่านกำหนดให้ค่าทั้งสองมีค่าเท่ากัน นักวิจัยยังสามารถใช้ค่า  $p$ -value ที่สอดคล้องกับค่า  $F$  ทั้งสองเป็นตัวเปรียบเทียบได้โดยเปรียบเทียบกับค่า  $\alpha$  นักวิจัยส่วนใหญ่จะใช้ค่า  $\alpha$  เท่ากันทั้งการพิจารณาการเพิ่มตัวแปรเข้าหรือการตัดตัวแปรออกโดยให้เท่ากับ 0.05 หรือ 0.10 โปรแกรม MINITAB กำหนดค่า  $\alpha$  ทั้งสองเท่ากับ 0.15



การรวมตัวแปรที่ไม่สำคัญเข้าไปในตัวแบบจัดเป็นการทำความผิดพลาดประเภทที่ 1 (Type I error) ซึ่งอาจเกิดผลเสียน้อยกว่าการที่ไม่รวมตัวแปรที่สำคัญเข้ามาในตัวแบบหรือการทำความผิดพลาดประเภทที่ 2 (Type II error)

หากใช้โปรแกรม MINITAB ช่วยในการหาตัวแบบที่เหมาะสมโดยวิธีการเพิ่มตัวแปรอิสระแบบขั้นตอนสามารถทำได้ดังนี้

(1) เลือก “Stat” ที่เมนูบาร์

(2) เลือก “Regression”

(3) เลือก “Stepwise...”

(4) ระบุตัวแปรตามใน “Response:” และตัวแปรอิสระใน “Predictors:” หากต้องการให้ตัวแปรอิสระใดอยู่ในตัวแบบทุกตัวแบบให้ระบุชื่อตัวแบบนั้นใน “Predictors to include in every model:”

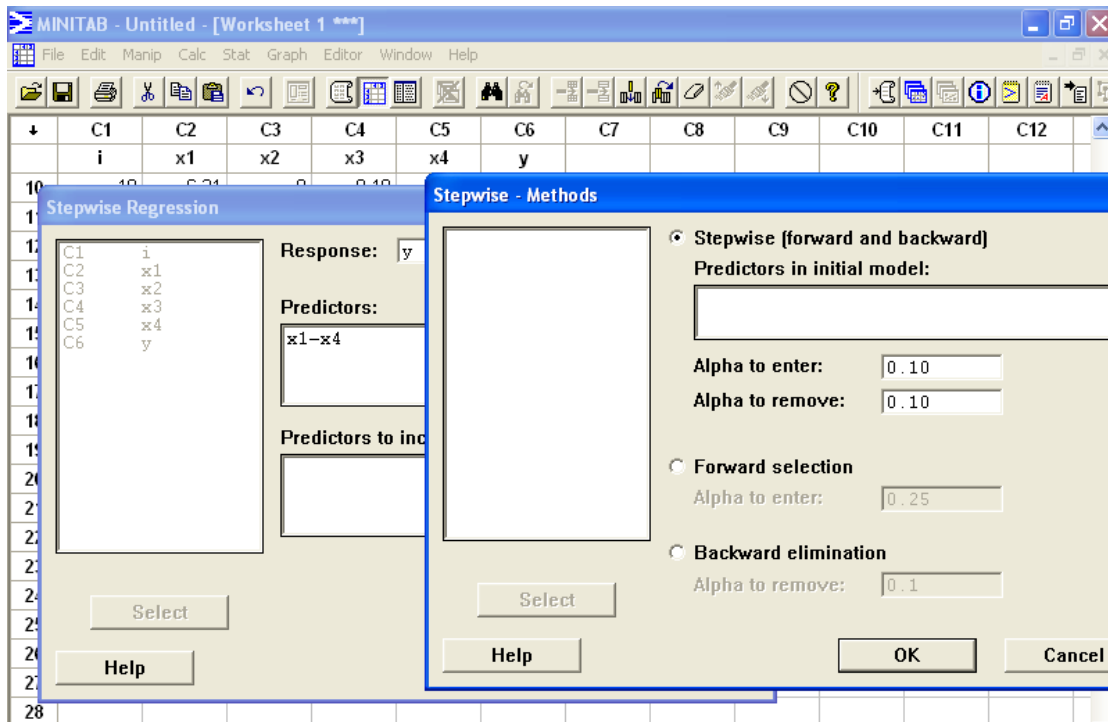
(5) คลิก “Methods...” จากนั้นคลิก “Stepwise (forward and backward)” และระบุค่า “Alpha to enter:” และ “Alpha to remove:” หากไม่ระบุค่าแล้วโปรแกรมจะระบุค่าให้เท่ากับ 0.15

(6) หากต้องการให้แสดงตัวแบบที่ดีที่สุดมากกว่า 1 ตัวแบบให้คลิก “Options” จากนั้นระบุจำนวนตัวแบบที่ดีที่สุดลงในช่อง “Number of alternative predictors to show:” เช่น 2 หมายถึงต้องการให้แสดงตัวแบบที่ดีที่สุด 2 ตัวแบบแรก เป็นต้น

(7) หากต้องการให้ทำการคำนวณมากกว่า 1 ครั้งในแต่ละครั้งของการเลือกตัวแปรอิสระให้คลิก “Options” จากนั้นระบุจำนวนครั้งของการคำนวณที่ “Number of steps between pauses:”

(8) หากต้องการให้แสดงค่า PRESS และ predicted R-Square ของแต่ละ “PRESS and predicted R-Square” ให้คลิก “Options” จากนั้นคลิก “Display PRESS and predicted R-Square” จากนั้นคลิก “OK”

(9) ผลลัพธ์ที่ได้จะแสดงในหน้าจอ “sessions” ดังภาพที่ 8.7



ภาพที่ 8.7 หน้าจอกำหนดค่าโดยวิธีการเพิ่มตัวแปรอิสระแบบขั้นตอน

ตัวอย่างที่ 8.5 จากตัวอย่างที่ 8.1 จงสร้างตัวแบบที่เหมาะสมกับข้อมูลชุดนี้ด้วยโปรแกรม MINITAB โดยใช้วิธีการเพิ่มตัวแปรอิสระแบบขั้นตอนที่ใช้ค่า  $\alpha$  to enter และ  $\alpha$  to remove เท่ากับ 0.10

#### วิธีทำ

เนื่องจากกำหนดค่า  $\alpha$  ทั้งสองเท่ากับ 0.10 ดังนั้นให้ระบุค่า Alpha to enter และ Alpha to remove เท่ากับ 0.10 ขั้นที่ 1 เลือกตัวแบบที่มาจากตัวแปรอิสระที่ให้ค่า  $F$ -to-enter สูงที่สุดหรือให้ค่า  $p$ -value ที่ต่ำที่สุดพบว่า  $X_4$  เป็นตัวแปรที่ให้ค่า  $p$ -value ที่ต่ำที่สุดเท่ากับ 0.035 ขั้นที่ 2 เพิ่มตัวแปรอิสระอีก 1 ตัวโดยเลือกตัวแปรที่ให้ค่า  $F$ -to-enter สูงที่สุดหรือให้ค่า  $p$ -value ที่ต่ำสุด โดยทั้งนี้ต้องต่ำกว่า 0.10 พบว่า  $X_3$  เป็นตัวแปรที่ให้ค่า  $p$ -value ที่ต่ำที่สุดเท่ากับ 0.000 ขั้นที่ 3 พิจารณาว่าควรเอาตัวแปรใดที่มีอยู่เดิม (ในที่นี้คือ  $x_4$ ) ออกจากตัวแบบหรือไม่โดยดูจากตัวแปรที่มีค่า  $F$ -to-remove ต่ำสุดหรือให้ค่า  $p$ -value ที่สูงสุดโดยทั้งนี้ต้องมากกว่า 0.10 ในที่นี้โปรแกรม MINITAB ไม่ได้แสดงผลลัพธ์ให้เห็น ขั้นที่ 4 ให้ทำตามขั้นที่ 2 และ 3 ทำเช่นนี้จนไม่สามารถเพิ่มหรือลดตัวแปรจากตัวแบบอีก ตัวแบบที่ดีที่สุดที่ได้จากโปรแกรมประกอบด้วยตัวแปร  $X_1$ ,  $X_3$  และ  $X_4$  โดยมีค่า  $S$  เท่ากับ 0.401 ค่า  $R_p^2$  เท่ากับ 64.63 ค่า  $R_{adj}^2(p)$  เท่ากับ 59.58 และค่า  $C_p$  เท่ากับ 3.2

ซึ่งสอดคล้องกับผลลัพธ์ที่ได้จากตัวอย่างที่ 8.2 - 8.4 ผลลัพธ์ที่ได้แสดงดังภาพข้างล่าง ตัวแบบที่ได้มีสมการถดถอยคือ

$$\hat{y} = 7.796 - 0.130x_1 - 3.16x_3 - 6.2x_4$$

Stepwise Regression: y versus x1, x2, x3, x4

Alpha-to-Enter: 0.1 Alpha-to-Remove: 0.1

Response is y on 4 predictors, with N = 25

Step	1	2	3
Constant	4.756	6.744	7.796
x4	-3.6	-6.2	-6.2
T-Value	-2.24	-4.47	-5.07
P-Value	0.035	0.000	0.000
x3		-3.74	-3.16
T-Value		-4.09	-3.74
P-Value		0.000	0.001
x1			-0.130
T-Value			-2.58
P-Value			0.017
S	0.584	0.450	0.401
R-Sq	17.95	53.42	64.63
R-Sq(adj)	14.38	49.18	59.58
C-p	26.0	7.7	3.2

ในกรณีที่มีข้อมูลมีตัวแปรอิสระจำนวนมากนั้นวิธีการเพิ่มตัวแปรอิสระแบบขั้นตอนเป็นวิธีที่เหมาะสมกว่าวิธีการพิจารณาทุกตัวแบบ ทั้งนี้เนื่องจากเมื่อจำนวนตัวแปรเพิ่มขึ้นการพิจารณาทุกตัวแบบจะใช้เวลามาก

ตัวอย่างที่ 8.6 จากข้อมูลในตัวอย่างที่ 5.10 จงใช้วิธีการเพิ่มตัวแปรอิสระแบบขั้นตอนเพื่อหาตัวแบบที่เหมาะสมโดยใช้โปรแกรม MINITAB ที่ใช้ค่า  $\alpha$  to enter และ  $\alpha$  to remove เท่ากับ 0.10

วิธีทำ โดยการ ใช้โปรแกรม MINITAB ได้ผลดังรูปข้างล่างซึ่งสอดคล้องกับผลที่ได้ในตัวอย่างที่ 5.10 คือ ตัวแปรความคาดหวังถึงการยอมรับจากสังคมที่คาดว่าจะได้รับ ( $X_1$ ) ระดับความมุ่งมั่นของผู้บริหาร ( $X_5$ ) การนำระบบการจัดการความปลอดภัยของอาหารมาใช้ของกิจการอื่น ( $X_6$ ) และระดับการมีปฏิสัมพันธ์กับองค์การภายนอกที่เกี่ยวข้องกับการจัดการความปลอดภัยของอาหาร ( $X_7$ ) สามารถใช้ในการพยากรณ์ระดับการนำระบบการจัดการความปลอดภัยของอาหารมาใช้ ( $Y$ )

**Stepwise Regression: y versus x1, x2, x3, x4, x5, x6, x7**

Alpha-to-Enter: 0.1 Alpha-to-Remove: 0.1

Response is y on 7 predictors, with N = 195

Step	1	2	3	4
Constant	0.2931	-0.2224	-0.4214	-0.6307
x5	0.867	0.737	0.629	0.611
T-Value	12.01	9.75	7.98	7.79
P-Value	0.000	0.000	0.000	0.000
x1		0.254	0.240	0.204
T-Value		4.25	4.14	3.43
P-Value		0.000	0.000	0.001
x7			0.216	0.200
T-Value			3.70	3.44
P-Value			0.000	0.001
x6				0.126
T-Value				2.25
P-Value				0.026
S	0.539	0.516	0.500	0.495
R-Sq	42.76	47.69	51.19	52.46
R-Sq(adj)	42.46	47.14	50.43	51.46
Mallows Cp	36.3	18.7	6.8	3.8

**สรุป**

ตัวแบบที่เหมาะสมคือ ตัวแบบที่มีความถูกต้องในการพยากรณ์มากที่สุด ในบรรดาตัวแบบที่มีทั้งหมด การเลือกตัวแบบที่เหมาะสมคือ การเลือกตัวแปรอิสระที่มีความสำคัญต่อตัวแปรตาม จากตัวแปรอิสระที่มีทั้งหมด เกณฑ์ที่ใช้ในการเลือกตัวแบบมีด้วยกันหลายแบบซึ่งส่วนมากแล้ว เกณฑ์เหล่านี้จะให้ข้อสรุปที่คล้ายกัน บางเกณฑ์มีข้อจำกัดในการใช้ที่แตกต่างกัน เช่น การลดตัวแปรอิสระไม่เหมาะสมกับงานวิจัยที่มีตัวแปรอิสระจำนวนมาก เป็นต้น

**คำถามท้ายบท**

8.1 ผลลัพธ์ข้างล่างได้จากโปรแกรม MINITAB จงเลือกตัวแบบที่ดีที่สุดโดยใช้  $R_p^2$   $R_{adj}^2(p)$   $C_p$  และ  $S$  พร้อมทั้งเปรียบเทียบผลลัพธ์ที่ได้ว่าเหมือนหรือต่างกันอย่างไร

Best Subsets Regression: y versus x1, x2, x3, x4, x5, x6											
Response is y											
Vars	R-Sq	R-Sq(adj)	C-p	S	x x x x x x						
					1	2	3	4	5	6	
1	25.3	22.0	1.2	4.2193				X			
1	3.6	0.0	7.6	4.7909	X						
1	2.9	0.0	7.8	4.8090						X	
2	30.5	24.2	1.6	4.1587	X	X					
2	29.2	22.8	2.0	4.1977		X	X				
2	27.1	20.4	2.6	4.2621		X		X			
3	34.6	25.3	2.4	4.1296	X	X		X			
3	33.9	24.5	2.6	4.1518	X	X	X				
3	33.2	23.6	2.8	4.1757	X	X		X			
4	38.3	26.0	3.3	4.1105	X	X	X	X			
4	38.1	25.8	3.4	4.1166	X	X	X	X			
4	35.2	22.2	4.2	4.2135	X	X	X		X		
5	38.9	22.9	5.1	4.1958	X	X	X	X	X		
5	38.7	22.5	5.2	4.2051	X	X	X	X	X		
5	38.6	22.5	5.2	4.2060	X	X	X	X	X		
6	39.3	19.1	7.0	4.2971	X	X	X	X	X	X	

8.2 จงอธิบายความแตกต่างระหว่างค่า  $R_p^2$  และ  $R_{adj}^2(p)$

8.32 ตัวแบบที่ดีควรมีค่า  $C_p$  เป็นอย่างไร

8.4 จากข้อ 5.10 จงหาตัวแบบที่เหมาะสมโดยใช้  $R_p^2$   $R_{adj}^2(p)$   $C_p$   $MSE(p)$   $PRESS_p$  และ  $S$  พร้อมทั้งเปรียบเทียบผลลัพธ์ที่ได้ว่าเหมือนหรือต่างกันอย่างไรกับข้อ 5.10

8.5 จากข้อมูลข้างล่างจงหาตัวแบบที่เหมาะสมโดยใช้  $R_p^2$   $R_{adj}^2(p)$   $C_p$   $MSE(p)$

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$Y$
57.5	18	5.7	26	30	2.0	79
58.5	20	7.5	37	17	3.5	82
57.5	20	6.6	32	27	2.5	78
56.5	12	5.1	24	27	2.5	78
55.0	17	6.6	34	23	2.5	77
56.5	13	5.7	28	12	3.1	75
59.0	19	5.4	24	20	3.0	77
58.5	18	6.9	32	23	2.5	85
55.5	22	6.3	28	10	4.0	81
61.0	18	6.3	27	22	2.5	81
55.5	19	6.6	32	18	3.4	77
56.0	20	6.9	33	20	2.6	75
59.5	20	6.6	35	8	4.0	85
60.0	22	5.1	21	25	2.5	77
57.0	17	5.4	27	15	3.8	79
55.5	18	6.0	29	18	2.1	79
55.5	21	6.9	32	18	3.6	80
57.5	18	6.6	32	20	2.0	77
63.0	12	6.6	27	17	1.4	80
57.5	21	6.0	20	10	2.3	75
57.0	15	7.2	20	30	2.8	83
57.5	20	6.3	20	25	3.5	81
57.5	15	6.9	22	22	2.7	84

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$Y$
56.0	23	6.9	20	14	2.6	80
57.5	18	6.9	20	18	3.5	80

- 8.6 วิธีการลดตัวแปรและวิธีการเพิ่มตัวแปรมีข้อดี ข้อเสียต่างกันอย่างไร
- 8.7 จากข้อ 8.5 จงหาตัวแบบที่เหมาะสมโดยใช้วิธีการลดตัวแปรที่  $\alpha$  to remove เท่ากับ 0.10
- 8.8 จากข้อ 8.5 จงหาตัวแบบที่เหมาะสมโดยใช้วิธีการเพิ่มตัวแปรที่  $\alpha$  to enter เท่ากับ 0.10
- 8.9 จากข้อ 8.5 จงหาตัวแบบที่เหมาะสมโดยใช้วิธีการเพิ่มตัวแปรอิสระแบบขั้นตอนที่ใช้ค่า  $\alpha$  to enter และ  $\alpha$  to remove เท่ากับ 0.10 แล้วเปรียบเทียบผลที่ได้จากข้อ 8.5 - 8.9 ว่าเหมือนหรือต่างกันอย่างไร
- 8.10 การใช้  $\alpha$  ที่แตกต่างกันจะให้ผลแตกต่างกันอย่างไรในการหาตัวแบบที่เหมาะสมและเมื่อใดจึงควรใช้ค่า  $\alpha$  ที่สูง
- 8.11 นักอสังหาริมทรัพย์ต้องการศึกษาปัจจัยที่มีผลต่อยอดขายของบ้าน โดยปัจจัยมีดังนี้คือ ราคาบ้าน พื้นที่บ้าน อายุของผู้ซื้อและจำนวนอาชญากรรมเฉลี่ยในแต่ละเดือนในเขตนั้น

ยอดขาย	ราคา	พื้นที่	อายุผู้ซื้อ	จำนวนอาชญากรรมเฉลี่ย
12.3	154.25	95	27.4	17.1
6.1	173.25	95	28.9	18.2
25.3	154.00	100	25.2	16.6
10.4	184.75	101	29.4	18.2
28.7	184.25	100	27.7	17.7
20.9	210.25	110	30.6	18.8
19.2	181.00	100	27.8	17.7
12.4	176.00	98	29.0	18.8
4.1	191.00	100	31.1	18.2
11.7	198.25	105	30.0	19.2
7.1	186.25	98	29.4	18.5
7.8	216.00	107	30.2	19.0
20.8	180.50	105	28.6	17.7
21.2	205.25	110	31.6	18.8

ยอดขาย	ราคา	พื้นที่	อายุผู้ซื้อ	จำนวนอาชญากรรมเฉลี่ย
22.1	187.75	100	30.5	18.2
20.9	162.75	100	26.4	16.9
29.0	195.75	105	30.8	17.3
22.9	209.25	107	31.6	19.3
16.0	183.75	102	30.5	18.5
16.5	211.75	109	30.1	18.2

- จากข้อมูลเบื้องต้นจงหาตัวแบบที่ดีที่สุดที่มีตัวแปรอิสระ 2 ตัวและ 3 ตัวโดยพิจารณาค่า  $C_p$
- 8.12 จากข้อ 8.11 จงหาตัวแบบที่ดีที่สุดโดยพิจารณาค่า  $C_p$  แล้วเปรียบเทียบผลที่ได้จาก ข้อ 8.7
- 8.13 จากข้อ 8.11 จงหาตัวแบบที่เหมาะสมโดยใช้วิธีการลดตัวแปรที่  $\alpha$  to remove เท่ากับ 0.05 และ 0.10 แล้วเปรียบเทียบผลลัพธ์ที่ได้
- 8.14 นักวิจัยต้องการศึกษาว่าปัจจัยใดมีผลต่อ IQ ปัจจัยมีดังต่อไปนี้ คือ พื้นที่ผิวของ Corpus Collasum (Corpus) เส้นผ่าศูนย์กลางของศีรษะ (head cir) เพศ (sex) พื้นที่ผิวของสมอง (brain surface) ปริมาตรของสมอง (brain volume) และน้ำหนักตัว (body weight)

Corpus	brain			body		IQ
	head cir	sex	surface	brain volume	weight	
6.08	54.7	ญ	1913.88	1005	67.0	96
5.73	54.2	ญ	1684.89	963	58.9	89
6.22	53.0	ญ	1902.36	1035	64.2	87
5.80	52.9	ญ	1860.24	1027	58.5	87
7.99	56.8	ญ	2264.25	1281	63.9	101
8.42	56.9	ญ	2216.40	1272	61.7	103
8.84	56.6	ญ	1866.99	1051	90.0	103
6.44	55.3	ญ	1850.64	1079	107.5	96
6.48	56.4	ญ	1743.04	1034	62.1	110
7.83	59.2	ญ	1709.30	1070	83.0	120
7.99	57.2	ช	1689.60	1173	85.0	120
8.76	57.2	ช	1806.31	1079	61.3	96

Corpus	head cir	sex	brain		body	IQ
			surface	brain volume	weight	
6.32	55.2	ช	2136.37	1067	97.0	93
6.32	57.2	ช	2018.92	1104	93.0	103
8.90	55.8	ช	1966.81	1347	97.5	94
6.12	53.5	ช	2154.67	1439	127.0	85
6.03	57.2	ช	1767.56	1029	91.0	97
6.59	58.7	ช	1827.92	1100	88.5	114
7.52	59.2	ช	1773.83	1204	80.0	113
7.67	58.5	ช	1971.63	1160	72.5	124

จงหาตัวแบบที่เหมาะสมโดยใช้  $R_p^2$ ,  $R_{adj}^2(p)$ ,  $C_p$ ,  $MSE(p)$ ,  $PRESS_p$  และ  $S$

8.15 จากข้อ 8.14 ท่านคิดว่าวิธีการเลือกตัวแปรแบบใดให้ผลลัพธ์ที่ดีที่สุดและจำนวนทั้ง 3 วิธีเปรียบเทียบกันโดยใช้  $\alpha$  to enter และ  $\alpha$  to remove เท่ากับ 0.05