บทที่ 5
การวิเคราะห์การถดถอยโลจิสติกด้วยโปรแกรมสำเร็จรูป SAS

5.1 กล่าวนำ

การวิเคราะห์การถดถอยโลจิสติก มีขบวนการวิเคราะห์ประมวลผลที่ค่อนข้างยุ่งยาก เป็นการลำบากถ้าจะวิเคราะห์ประมวลผลด้วยมือ จำเป็นต้องอาศัยโปรแกรมสำเร็จรูปทางสถิติช่วยในการประมวลผล โปรแกรมสำเร็จรูปทางสถิติที่ใช้ในการวิเคราะห์การถดถอยโลจิสติกมีหลายโปรแกรม เช่น GLIM3 (Generalized Linear Interactive Modelling) GENSTAT5 (General Statistical program) SAS (Statistical Analysis System) BMDP (Biomedical Package) SSPC/PC+ (Statistical Package for Social Scientists) และ EGRET (Epidermiological, Graphics, Estimation and Testing program) เป็นต้น

บทนี้จะกล่าวเฉพาะการใช้โปรแกรมสำเร็จรูป SAS (Additional SAS/STAT Procedure) Release 6.03 เนื่องจากเป็นโปรแกรมที่มีผู้ใช้กันอย่างกว้างขวาง คำสั่งที่ใช้ในการวิเคราะห์การถดถอยโลจิสติกได้แก่ คำสั่ง **proc probit** และคำสั่ง **proc catmood** ซึ่งจะให้ผลลัพธ์ที่เหมือนกัน แต่การใช้คำสั่ง **proc probit** จะใช้ง่ายและสะดวกกว่าการใช้คำสั่ง **proc catmod** ซึ่งนอกจากมีวิธีการเขียนโปรแกรมที่ซับซ้อนกว่าแล้ว การตีความผลลัพธ์ที่ได้ก็ยากกว่าด้วย ดังนั้นจะกล่าวเฉพาะการใช้คำสั่ง **proc probit** เท่านั้น

สำหรับข้อมูลที่ใช้เป็นตัวอย่างเพื่อวิเคราะห์การถดถอยโลจิสติก ด้วยโปรแกรม SAS นำมาจาก Collett (1991) มีจำนวน 2 ชุด ดังนี้

ข้อมูลชุดที่ 1 เป็นข้อมูลที่ได้จากการทดลองของ Smith (1932) ซึ่งทดลองเกี่ยวกับการใช้เซรุ่มชื่อ ' Serum number 32 ' เพื่อป้องกันโรคนิวมอเนีย (pneumonia) ที่เกิดจากเชื้อแบคทีเรีย pneumococcus การทดลองทำโดยการฉีด culture of pneumococci และเซรุ่ม anti-pneumococcus แต่ละปริมาณ จำนวน 5 ปริมาณ ให้กับหนูทดลองจำนวน 5 กลุ่ม ๆ ละ 40 ตัว เมื่อครบ 7 วัน นับจำนวนหนูที่ตายในแต่ละกลุ่มได้ดังตาราง 5.1

ตาราง 5.1 จำนวนหนูที่ตายเนื่องจากโรคนิวมอเนีย เมื่อใช้เซรุ่มในปริมาณที่แตกต่าง

ปริมาณเซรุ่มที่ฉีดให้กับหนู(ลบ.ซม.) (dose)	จำนวนหนูที่ตายจากทั้งหมด 40 (n) ตัว (y)
0.0028	35
0.0056	21
0.0112	9
0.0225	6
0.0450	1

ข้อมูลชุดที่ 2 เป็นข้อมูลจากการทดลองของ Hoblyn และ Palmer (1934) ซึ่งทดลองตัดรากของต้นพลัมพันธุ์ Common Mussel ในช่วงเวลาระหว่างเดือนตุลาคม 2477 และเดือนกุมภาพันธ์ 2478 ให้มีความยาว 2 ขนาด คือ ขนาด 12 และ 6 เซนติเมตร จำนวนขนาดละ 480 ส่วน นำครึ่งหนึ่งของแต่ละขนาดไปปลูกทันที สำหรับอีกครึ่งหนึ่งนำไปเพาะในทรายก่อนจนถึงฤดูใบไม้ผลิ แล้วจึงนำไปปลูก ทดลองจนถึงเดือนตุลาคม 2478 จึงนับจำนวนต้นพลัมที่มีชีวิตรอดอยู่ ได้ข้อมูลดังตาราง 5.2

ตาราง 5.2 อัตราการอยู่รอดของต้นพลับจากการขยายพันธุ์โดยการตัดราก

ความยาวของรากที่ตัด (length = X_j)	เวลาที่ปลูก (time = X_k)	จำนวนที่อยู่รอด (y) จากทั้งหมด 240 (n) ส่วน
6 เซนติเมตร	ปลูกทันที	107
	ปลูกในฤดูใบไม้ผลิ	31
12 เซนติเมตร	ปลูกทันที	156
	ปลูกในฤดูใบไม้ผลิ	84

5.2 การวิเคราะห์การถดถอยโลจิสติกสำหรับข้อมูลชุดที่ 1

โมเดลที่ใช้ คือ $\text{logit}(\hat{p}_i) = b_0 + b_1 \text{dose}_i$

เมื่อ $\hat{p}_i$ หมายถึงความน่าจะเป็นที่หนูซึ่งได้รับการฉีดเชรุ่มในปริมาณที่ i แล้วยังตายด้วยโรคนิวมอเนีย

dose_i หมายถึงเชรุ่มปริมาณที่ i

b_0 หมายถึงค่าคงที่

b_1 หมายถึงสัมประสิทธิ์การถดถอยเนื่องจากปริมาณเชรุ่มที่ฉีดให้กับหนู

5.2.1 โปรแกรม

```
 1        option ls=70 ps=60 nodate;
 2        data serum1;
 3        input dose y n ;
 4        cards;
 5        0.0028 35 40
 6        0.0056 21 40
 7        0.0112  9 40
 8        0.0225  6 40
 9        0.0450  1 40
10        proc probit;
11        model y/n=dose / d=logistic lackfit;
12        output out=serum2;
13        proc print;
14        run;
```


5.2.2 คำอธิบายเกี่ยวกับโปรแกรม

1. ข้อมูลที่วิเคราะห์ชื่อ **serum1** (บรรทัดที่ 2) กรณีนี้ข้อมูลที่จะวิเคราะห์อยู่ในแฟ้มโปรแกรม (program file) ถ้าเป็นกรณีที่ข้อมูลอยู่ในแฟ้มข้อมูล (data file) ต่างหากอีกแฟ้มหนึ่ง สมมติชื่อแฟ้ม serum1 สามารถเขียนโปรแกรม **infile 'serum1';** เพิ่มเติมระหว่างบรรทัดที่ 2 และ 3

2. ตัวแปรที่ใช้ได้แก่ dose แทนปริมาณเซรุ่มที่ฉีดให้กับหนู y แทนจำนวนหนูทดลองที่ตาย และ n แทนจำนวนหนูทดลองของแต่ละกลุ่ม

3. **model y/n = dose** คือการระบุโมเดลที่ใช้ว่ามี y/n เป็นตัวแปรตาม และมี dose เป็นตัวแปรอิสระ การที่ตัวแปรตามในที่นี้คือ y/n ก็เพราะเป็นข้อมูลทวิที่มาจากการแจกแจงแบบทวินาม **d = logistic** เป็นการบอกให้ทำการประมวลผลโดยใช้โมเดลโลจิสติก **lackfit** เป็นการสั่งให้ทำการทดสอบภาวะสารูปดี

5.2.3 ผลลัพธ์ที่ได้

Probit Procedure

Data Set =WORK.SERUM1

Dependent Variable=Y

Dependent Variable=N

Number of Observations= 5

Number of Events = 72 Number of Trials = 200

Log Likelihood for LOGISTIC -93.60787245

Goodness-of-Fit Tests

Statistic	Value	DF	Prob>Chi-Sq
Pearson Chi-Square	16.3416	3	0.0010
L.R. Chi-Square	15.9000	3	0.0012

Response Levels: 2 Number of Covariate Values: 5

Probit Procedure

Variable	DF	Estimate	Std Err	ChiSquare	Pr>Chi	Label/Value
INTERCPT	1	1.21789844	0.683769	3.172517	0.0749	Intercept
DOSE	1	-146.69272	61.5291	5.684026	0.0171	

Estimated Covariance Matrix

	INTERCPT	DOSE
INTERCPT	0.467539	-33.499749
DOSE	-33.499749	3785.829606

OBS	DOSE	Y	N
1	0.0028	35	40
2	0.0056	21	40
3	0.0112	9	40
4	0.0225	6	40
5	0.0450	1	40

5.2.4 การตีความผลลัพธ์ที่ได้

1. จากค่าพารามิเตอร์ที่ประมาณได้ จะได้โมเดลเป็น

$$\text{logit}(\hat{p}_i) = b_0 + b_1 \text{dose}_i$$

$$= 1.22 - 146.69\ \text{dose}_i$$

2. Number of Events = 72 และ Number of Tirals = 200 หมายถึง จำนวนหนูทดลองทั้ง 5 กลุ่ม ที่ฉีดเชื้อรุ่มแล้วยังตายด้วยโรคนิวมอเนียจำนวน 72 ตัว จากทั้งหมด 200 ตัว

3. **Log Likelihood for Logistic** -93.60787245 คือค่าของ maximized log-likelihood ของโมเดล $\text{logit}(\hat{p}_i) = b_o + b_1 \text{dose}_i$ หรือ $\ln L_c$

4. **Goodness-of-Fit Tests** จะให้ค่า Pearson Chi-Square = 16.3416 ความน่าจะเป็นที่จะได้ไคสแควร์มีค่ามากกว่า 16.3416 เท่ากับ 0.0010 และ L.R.(Likelihood Ratio) Chi-Square = 15.9000 ซึ่งคือค่า deviance ความน่าจะเป็นที่จะได้ไคสแควร์มีค่ามากกว่า 15.9000 เท่ากับ 0.0012 มี d.f.= 3 ซึ่งสรุปได้ว่า ผลการทดสอบภาวะสารูปดีมีนัยสำคัญ ที่ระดับนัยสำคัญ 0.10 (โดยทั่วไปจะนิยมใช้ระดับนัยสำคัญ 0.10) เพราะความน่าจะเป็นมีค่าน้อยกว่าระดับนัยสำคัญ แสดงว่า โมเดลที่ประมาณได้ยังไม่เหมาะสม

5. คอลัมน์ **Estimate** และ **Std Err** แสดงค่าของพารามิเตอร์ที่ประมาณได้โดยวิธีภาวะน่าจะเป็นสูงสุด และความคลาดเคลื่อนมาตรฐานของค่าประมาณ ตามลำดับ ดังนั้น จากโมเดลข้างต้น จะได้ INTERCPT = b_o = 1.2179 ด้วยความคลาดเคลื่อนมาตรฐาน 0.6838 และ DOSE = b_1 = -146.6927 ด้วยความคลาดเคลื่อนมาตรฐาน 61.5291 คอลัมน์ **ChiSquare** แสดงค่าไคสแควร์ที่คำนวณจาก $(\text{Estimate/Std Err})^2$ และคอลัมน์ **Pr>Chi** แสดงค่าความน่าจะเป็นที่จะได้ไคสแควร์มีค่ามากกว่าค่าไคสแควร์ที่คำนวณได้ จากผลลัพธ์ที่ได้แสดงว่า การทดสอบ $H_o : \beta_o = 0$, $H_1 : \beta_o \neq 0$ ให้ค่าไคสแควร์ 3.1725 ด้วยความน่าจะเป็น 0.0749 และการทดสอบ $H_o : \beta_1 = 0$, $H_1 : \beta_1 \neq 0$ ให้ค่าไคสแควร์ 5.6840 ด้วยความน่าจะเป็น 0.0171 ซึ่งต่างก็มีนัยสำคัญ ที่ระดับนัยสำคัญ 0.10

ในกรณีที่ค่าทดสอบสถิติไคสแควร์ที่คำนวณได้มีค่ามาก ๆ ซึ่งจะให้ความน่าจะเป็นมีค่าน้อยจนเข้าใกล้ศูนย์ โปรแกรม SAS จะแสดงค่าความน่าจะเป็นไว้เท่ากับ 0.0001

ผลสรุปจากการตีความพบว่า โมเดล $\text{logit}(\hat{p}_i) = b_o + b_1 \text{dose}_i$ ยังไม่เหมาะสม เมื่อมีการแปลงค่าปริมาณเชื้อรุ่มที่ฉีด (dose) ด้วย $\log_{10}$ เป็น logdose = log(dose) และให้โมเดลเป็น $\text{logit}(\hat{p}_i) = b_o + b_1 \text{logdose}_i$ สามารถเพิ่มเติมโปรแกรมเพื่อวิเคราะห์การถดถอยโลจิสติกเป็น

logdose=log(dose); แทรกระหว่างบรรทัดที่ 3 และ 4

และ model y/n=logdose / d=logistic lackfit;

output out=serum3 แทรกระหว่างบรรทัดที่ 12 และ 13 ซึ่งจะให้ผลลัพธ์เป็น

Probit Procedure

Data Set =WORK.SERUM2

Dependent Variable=Y

Dependent Variable=N

Number of Observations= 5

Number of Events = 72 Number of Trials = 200

Log Likelihood for LOGISTIC -87.06231266

Goodness-of-Fit Tests

Statistic	Value	DF	Prob>Chi-Sq
Pearson Chi-Square	2.9174	3	0.4045
L.R. Chi-Square	2.8089	3	0.4220

Response Levels: 2 Number of Covariate Values: 5

Probit Procedure

Variable	DF	Estimate	Std Err	ChiSquare	Pr>Chi	Label/Value
INTERCPT	1	-9.1893917	1.25511	53.6056	0.0001	Intercept
LOGDOSE	1	-1.8296213	0.254547	51.66384	0.0001	

Estimated Covariance Matrix

	INTERCPT	LOGDOSE
INTERCPT	1.575300	0.315849
LOGDOSE	0.315849	0.064794

OBS	DOSE	Y	N	LOGDOSE
1	0.0028	35	40	-5.87814
2	0.0056	21	40	-5.18499
3	0.0112	9	40	-4.49184
4	0.0225	6	40	-3.79424
5	0.0450	1	40	-3.10109

ผลจากการทดสอบภาวะสารูปดี ให้ค่า Pearson Chi-Square = 2.9174 ด้วยความน่าจะเป็น 0.4045 และค่า L.R.Chi-Square = 2.8089 ด้วยความน่าจะเป็น 0.4220 และจากค่าที่ประมาณได้ คือ INTERCPT = b_o = -9.19 มีค่าไคสแควร์ 53.6056 ด้วยความน่าจะเป็น 0.0001 และ LOGDOSE = b_1 = -1.83 มีค่าไคสแควร์ 51.6638 ด้วยความน่าจะเป็น 0.0001

สรุปว่า โมเดล $\text{logit}(\hat{p}_i) = b_o + b_1 \text{logdose}_i = -1.19 - 1.83\ \text{logdose}_i$ ที่ประมาณได้ เหมาะสมกับค่าสังเกตแล้ว และเมื่อต้องการให้แสดงค่า $\hat{p}_i$ ซึ่งเป็นค่าประมาณของ P_i ให้เขียนโปรแกรมเพิ่มเป็น output out=serum3 **prob=phat** ; ซึ่งจะได้ผลลัพธ์เป็น

OBS	DOSE	Y	N	LOGDOSE	PHAT
1	0.0028	35	40	-5.87814	0.82712
2	0.0056	21	40	-5.18499	0.57375
3	0.0112	9	40	-4.49184	0.27468
4	0.0225	6	40	-3.79424	0.09558
5	0.0450	1	40	-3.10109	0.02887

เมื่อ PHAT = $\hat{p}_i = [1 + \exp\{-(-1.19 - 1.83\ \text{logdose}_i)\}]^{-1}$

5.3 การวิเคราะห์การถดถอยโลจิสติกสำหรับข้อมูลชุดที่ 2

เนื่องจากมีตัวแปรอิสระมากกว่า 1 ตัวแปร ดังนั้น จะประมาณโมเดลที่เป็นไปได้ทั้งหมด แล้วนำมาเปรียบเทียบกันเพื่อหาโมเดลที่เหมาะสมที่สุด ได้เป็น

โมเดล (1) ; $\text{logit}(\hat{p}_{jk}) = b_o$

โมเดล (2) ; $\text{logit}(\hat{p}_{jk}) = b_o + b_1X_j$

โมเดล (3) ; $\text{logit}(\hat{p}_{jk}) = b_o + b_2X_k$

โมเดล (4) ; $\text{logit}(\hat{p}_{jk}) = b_o + b_1X_j + b_2X_k$

โมเดล (5) ; $\text{logit}(\hat{p}_{jk}) = b_o + b_1X_j + b_2X_k + b_3X_jX_k$

เมื่อ $\hat{p}_{jk}$ คือ ความน่าจะเป็นที่ต้นพลัมจากรากจะมีชีวิตอยู่รอดเมื่อรากที่ปลูกมีขนาด j และปลูกในเวลา k ; $j = 1, 2$

$k = 1, 2$

b_o คือ ค่าคงที่

b_1 คือ สัมประสิทธิ์การถดถอยเนื่องจากความยาวราก

b_2 คือ สัมประสิทธิ์การถดถอยเนื่องจากเวลาปลูก

b_3 คือ สัมประสิทธิ์การถดถอยเนื่องจากอิทธิพลร่วมระหว่างความยาวรากและเวลาปลูก

X_j คือ ความยาวรากขนาดที่ j

$X_1 = 1$ = รากยาว 6 ซม.

$X_2 = 0$ = รากยาว 12 ซม.

X_k คือ เวลาปลูกที่ k

$X_1 = 1$ = ปลูกทันที

$X_2 = 0$ = ปลูกในฤดูใบไม้ผลิ

5.3.1 โปรแกรม

```
1       option ls=70 ps=60 nodate;
2       data plums1;
3       input length time y n lt;
4       cards;
5       1 1 107 240 1
6       1 2  31 240 0
7       2 1 156 240 0
8       2 2  84 240 0
9       proc probit;
10      class length time lt;
11      model y/n= /d=logistic;
12      model y/n=length/d=logistic lackfit;
13      model y/n=time/d=logistic lackfit;
14      model y/n= length time / d=logistic lackfit;
15      model y/n=length time lt/ d=logistic lackfit;
16      output out= plums2;
17      proc print;
18      run;
```

5.3.2 คำอธิบายเกี่ยวกับโปรแกรม

1. ตัวแปรที่ใช้ได้แก่ length แทนความยาวราก time แทนเวลาที่ปลูก lt แทนอิทธิพลร่วมระหว่างความยาวรากและเวลาปลูก y แทนจำนวนต้นพลับที่มีชีวิตรอด และ n แทนจำนวนรากทั้งหมดของแต่ละกลุ่ม

2. เมื่อตัวแปรที่ใช้ในการวิเคราะห์การถดถอยเป็นตัวแปรแบบกลุ่ม จะต้องใช้คำสั่ง **class** ตามด้วยตัวแปรกลุ่ม ซึ่งในที่นี้ได้แก่ตัวแปร length time lt

3. เมื่อตัวแปรอิสระเป็นตัวแปรแบบกลุ่ม โปรแกรม SAS จะสร้างตัวแปรหุ่นให้จำนวน m-1 ตัว สำหรับแต่ละตัวแปรกลุ่มที่มีจำนวน m ระดับ และจะให้ค่าของตัวแปรหุ่นตัวที่ i มีค่าเป็น 1 เมื่อตัวแปรกลุ่มนั้น ๆ อยู่ในระดับที่ i สำหรับตัวแปรหุ่นที่เหลือจะให้ค่าเป็น 0 เช่น ตัวแปรกลุ่ม X มีจำนวน 3 ระดับ คือ 1, 2 และ 3 โปรแกรม SAS จะสร้างตัวแปรหุ่นให้จำนวน 3-1 = 2 ตัว สมมติว่าคือ D_1 และ D_2

เมื่อ X = 1 จะได้ $D_1 = 1$, $D_2 = 0$

เมื่อ X = 2 จะได้ $D_1 = 0$, $D_2 = 1$

เมื่อ X = 3 จะได้ $D_1 = 0$, $D_2 = 0$

ดังนั้น การสร้างค่าของตัวแปรกลุ่มที่เป็นอิทธิพลร่วมระหว่างตัวแปรกลุ่มด้วยกัน ต้องคำนึงถึงหลักการนี้ ค่าที่ได้จะอยู่ในรูปของแมทริกซ์ ซึ่งค่อนข้างจะยุ่งยากเมื่อมีตัวแปรกลุ่มหลายตัว และแต่ละตัวแปรกลุ่มมีหลายระดับ ในที่นี้จะได้ค่าของตัวแปร length time และ lt เป็น

length (ความยาวราก)	time (เวลาที่ปลูก)	lt (อิทธิพลร่วม)
6 ซม. = 1	ปลูกทันที = 1	1
6 ซม. = 1	ปลูกในฤดูใบไม้ผลิ = 0	0
12 ซม. = 0	ปลูกทันที = 1	0
6 ซม. = 0	ปลูกในฤดูใบไม้ผลิ = 0	0

5.3.3 ผลลัพธ์ที่ได้

สำหรับโมเดล (1)

Probit Procedure

Data Set =WORK.PLUMS1

Dependent Variable=Y

Dependent Variable=N

Number of Observations= 4

Number of Events = 378 Number of Trials = 960

Log Likelihood for LOGISTIC -643.5801466

Probit Procedure

Variable	DF	Estimate	Std Err	ChiSquare	Pr>Chi	Label/Value
INTERCPT	1	-0.4315763	0.066058	42.68336	0.0001	Intercept

Estimated Covariance Matrix

	INTERCPT
INTERCPT	0.004364

สำหรับโมเดล (2)

Probit Procedure

Class Level Information

Class	Levels	Values
LENGTH	2	1 2

Number of observations used = 4

Probit Procedure

Data Set =WORK.PLUMS1

Dependent Variable=Y

Dependent Variable=N

Number of Observations= 4

Number of Events = 378 Number of Trials = 960

Log Likelihood for LOGISTIC -620.6616959

Goodness-of-Fit Tests

Statistic	Value	DF	Prob>Chi-Sq
Pearson Chi-Square	101.9440	2	0.0000
L.R. Chi-Square	105.1824	2	0.0000

Response Levels: 2 Number of Covariate Values: 4

Probit Procedure

Variable	DF	Estimate	Std Err	ChiSquare	Pr>Chi	Label/Value
INTERCPT	1	8.237E-17	0.651741	1.6E-32	1.0000	Intercept
LENGTH	1			0.873289	0.3500	
	1	-0.907557	0.971169	0.873289	0.3500	1
	0	0	0	.	.	2

Estimated Covariance Matrix

	INTERCPT	LENGTH.1
INTERCPT	0.424766	-0.424766
LENGTH.1	-0.424766	0.943170

สำหรับโมเดล (3)

Probit Procedure

Class Level Information

Class	Levels	Values
TIME	2	1 2

Number of observations used = 4

Probit Procedure

Data Set =WORK.PLUMS1

Dependent Variable=Y

Dependent Variable=N

Number of Observations= 4

Number of Events = 378 Number of Trials = 960

Log Likelihood for LOGISTIC -594.7906947

Goodness-of-Fit Tests

Statistic	Value	DF	Prob>Chi-Sq
Pearson Chi-Square	52.3158	2	0.0000
L.R. Chi-Square	53.4404	2	0.0000

Response Levels: 2 Number of Covariate Values: 4

Probit Procedure

Variable	DF	Estimate	Std Err	ChiSquare	Pr>Chi	Label/Value	
INTERCPT	1	-1.1549652	0.546923	4.459484	0.0347	Intercept	
TIME	1			3.496259	0.0615		
	1	1.3472219	0.720506	3.496259	0.0615		1
	0	0	0	.	.		2

Estimated Covariance Matrix

	INTERCPT	TIME.1
INTERCPT	0.299125	-0.299125
TIME.1	-0.299125	0.519128

สำหรับโมเดล (4)

Probit Procedure

Class Level Information

Class	Levels	Values
LENGTH	2	1 2
TIME	2	1 2

Number of observations used = 4

Probit Procedure

Data Set =WORK.PLUMS1

Dependent Variable=Y

Dependent Variable=N

Number of Observations= 4

Number of Events = 378 Number of Trials = 960

Log Likelihood for LOGISTIC -569.2174083

Goodness-of-Fit Tests

Statistic	Value	DF	Prob>Chi-Sq
Pearson Chi-Square	2.2705	1	0.1319
L.R. Chi-Square	2.2938	1	0.1299

Response Levels: 2 Number of Covariate Values: 4

Probit Procedure

Variable	DF	Estimate	Std Err	ChiSquare	Pr>Chi	Label/Value	
INTERCPT	1	-0.7137712	0.121669	34.41576	0.0001	Intercept	
LENGTH	1			48.93569	0.0001		
	1	-1.0176915	0.14548	48.93569	0.0001		1
	0	0	0	.	.		2
TIME	1			95.00046	0.0001		
	1	1.42754237	0.146462	95.00046	0.0001		1
	0	0	0	.	.		2

Estimated Covariance Matrix

	INTERCPT	LENGTH.1	TIME.1
INTERCPT	0.014803	-0.007761	-0.010726
LENGTH.1	-0.007761	0.021164	-0.003359
TIME.1	-0.010726	-0.003359	0.021451

สำหรับโมเดล (5)

Probit Procedure

Class Level Information

Class	Levels	Values
LENGTH	2	1 2
TIME	2	1 2
LT	2	0 1

Number of observations used = 4

Probit Procedure

Data Set =WORK.PLUMS1

Dependent Variable=Y

Dependent Variable=N

Number of Observations= 4

Number of Events = 378 Number of Trials = 960

Log Likelihood for LOGISTIC -568.0704886

Probit Procedure

Variable	DF	Estimate	Std Err	ChiSquare	Pr>Chi	Label/Value	
INTERCPT	1	-0.1662909	0.381422	0.190076	0.6629	Intercept	
LENGTH	1			30.02857	0.0001		
	1	-1.2893078	0.235282	30.02857	0.0001		1
	0	0	0	.	.		2
TIME	1			41.84648	0.0001		
	1	1.23807842	0.19139	41.84648	0.0001		1
	0	0	0	.	.		2
LT	1			2.264049	0.1324		
	1	-0.4527483	0.300894	2.264049	0.1324		0
	0	0	0	.	.		1

Estimated Covariance Matrix

	INTERCPT	LENGTH.1	TIME.1	LT.1
INTERCPT	0.145482	-0.073673	-0.054945	-0.108852
LENGTH.1	-0.073673	0.055358	0.018315	0.055358
TIME.1	-0.054945	0.018315	0.036630	0.036630
LT.1	-0.108852	0.055358	0.036630	0.090537

OBS	LENGTH	TIME	Y	N	LT
1	1	1	107	240	1
2	1	2	31	240	0
3	2	1	156	240	0
4	2	2	84	240	0

5.3.4 การตีความผลลัพธ์ที่ได้

1. ค่าพารามิเตอร์ที่ประมาณได้ของแต่ละโมเดล เป็นดังนี้

โมเดล (1) ; $\text{logit}(\hat{p}_{jk}) = b_o$

$= - 0.43$

โมเดล (2) ; $\text{logit}(\hat{p}_{jk}) = b_o + b_1 X_j$

$= (8.24 \times 10^{-17}) - 0.91 X_j$

โมเดล (3) ; $\text{logit}(\hat{p}_{jk}) = b_o + b_2 X_k$

$= - 1.15 + 1.35 X_k$

โมเดล (4) ; $\text{logit}(\hat{p}_{jk}) = b_o + b_1 X_j + b_2 X_k$

$= - 0.71 - 1.02 X_j + 1.43 X_k$

โมเดล (5) ; $\text{logit}(\hat{p}_{jk}) = b_o + b_1 X_j + b_2 X_k + b_3 X_j X_k$

$= - 0.17 - 1.29 X_j + 1.24 X_k - 0.45 X_j X_k$

2. การพิจารณาว่าโมเดลใดจะเหมาะสม อาจพิจารณาได้ง่าย ๆ จากการทดสอบภาวะสารูปดี โดยดูว่าค่า Pearson Chi-Square หรือค่า L.R.Chi-Square ของโมเดลใดไม่มีนัยสำคัญ หรือพิจารณาจากการวิเคราะห์ deviance ดังนี้

โมเดล	Pearson Chi-Square(Prob>Chi-Sq)	L.R.Chi-Square(Prob>Chi-Sq)
(1)	-	-
(2)	101.9440^{***} (0.0000)	105.1834^{***} (0.0000)
(3)	52.3158^{***} (0.0000)	53.4404^{***} (0.0000)
(4)	2.2705^{NS} (0.1319)	2.2938^{NS} (0.1299)
(5)	-	-

สำหรับของโมเดล (1) และโมเดล (5) โปรแกรม SAS จะไม่คำนวณค่า Pearson Chi-Square และค่า L.R.Chi-Square ให้

และ *** หมายความว่าค่าทดสอบสถิติมีนัยสำคัญ ที่ระดับนัยสำคัญ 0.01

^{NS} หมายความว่าค่าทดสอบสถิติไม่มีนัยสำคัญ ที่ระดับนัยสำคัญ 0.10

เมื่อพิจารณาจากการทดสอบภาวะสารูปดีจะสรุปได้ว่า โมเดล (4) เป็นโมเดลที่เหมาะสมเนื่องจากค่า Pearson Chi-Square และค่า L.R.Chi-Square ไม่มีนัยสำคัญ

เนื่องจากโปรแกรม SAS จะไม่คำนวณค่า Pearson Chi-Square และค่า L.R.Chi-Square ของโมเดล (1) ให้ อาจเกิดข้อสงสัยว่า มีความจำเป็นหรือไม่ที่จะต้องเพิ่มตัวแปรให้กับโมเดล ในกรณีนี้สามารถทำการวิเคราะห์ deviance โดยใช้ค่า Log Likelihood ของแต่ละโมเดล ซึ่งแสดงดังตาราง

โมเดล	Log Likelihood
(1)	-643.58
(2)	-620.66
(3)	-594.79
(4)	-569.22
(5)	-568.07

ทำการวิเคราะห์ค่า deviance ได้ดังนี้

deviance ของโมเดล (2) ลดลงจากของโมเดล (1) $= -2[L_{c1} - L_{c2}]$

$= -2[-643.58 - (-620.66)]$

$= 45.84$

deviance ของโมเดล (3) ลดลงจากของโมเดล (1) = $-2[L_{c1} - L_{c3}]$

$= -2[-643.58 - (-594.79)]$

$= 97.58$

ค่า deviance ที่ลดลง 45.84 และ 97.58 ต่างก็มีค่ามากกว่าค่าไคสแควร์ 2.706 จากตารางที่ d.f. 1 [d.f.ของโมเดล (1) - d.f.ของโมเดล (2) = 3 - 2 = 1] ระดับนัยสำคัญ 0.10 ซึ่งหมายความว่า ตัวแปร length และ time ควรจะมีอยู่ในโมเดลทั้งสองตัวแปร โดยควรนำตัวแปร time เข้าในโมเดลก่อน เพราะค่า deviance ที่ลดลงมีค่ามากกว่า และเมื่อนำตัวแปร time เข้าในโมเดลได้เป็นโมเดล (3) แล้วลองนำตัวแปร length เข้าในโมเดล ได้เป็นโมเดล (4) จะได้ deviance มีค่าลดลงจากของโมเดล (3) เท่ากับ -2 [-594.79 - (-569.22)] = 51.14 มีค่ามากกว่าค่ามากกว่า 2.706 จากตารางที่ d.f. 1 [d.f.ของโมเดล (3) - d.f.ของโมเดล (4) = 2 - 1 = 1] ระดับนัยสำคัญ 0.10

ในทำนองเดียวกัน สามารถพิจารณาว่าตัวแปร It ควรจะถูกนำเข้าในโมเดลหรือไม่ พบว่าค่า deviance ของโมเดล (5) ลดลงจากของโมเดล (4) เท่ากับ -2 [-569.22 - (-568.07)] = 2.30 ซึ่งมีค่าน้อยกว่า 2.706 จากตารางที่ d.f. 1 [d.f.ของโมเดล (4) - d.f.ของโมเดล (5) = 1 - 0 = 1] ระดับนัยสำคัญ 0.10 นั่นคือ โมเดล (4) = $\text{logit}(\hat{p}_{jk}) = -0.71 - 1.02X_j + 1.43X_k$ เป็นโมเดลที่เหมาะสม และสามารถประมาณค่าความน่าจะเป็น $\hat{p}_{jk}$ ของโมเดล (4) จากโปรแกรม SAS ได้เป็น

OBS	LENGTH	TIME	Y	N	PHAT
1	1	1	107	240	0.42460
2	1	2	31	240	0.15040
3	2	1	156	240	0.67123
4	2	2	84	240	0.32877

เมื่อ $PHAT = \hat{p}_{jk} = [1 + \exp\{-(-0.71 - 1.02X_j + 1.43X_k)\}]^{-1}$

$j, k = 1, 2$

$X_1 = 0, \quad X_2 = 1$
