PROC REG

及

PROC LOGISTIC

政治大學統計系

江振東

前言

◆Significant vs. Important

◆統計顯著(Statistically Significant) vs. 實務顯著(Practically Significant)

e.g. $H_0: \mu = 170$ vs. $H_1: \mu \neq 170$

Test Statistic: $Z = \dfrac{\bar{X} - 170}{\sigma/\sqrt{n}} = \sqrt{n}\dfrac{\bar{X} - 170}{\sigma}$ (假設 $\sigma = 5$)

<u>Case 1</u>: $n = 4,\ \bar{X} = 174 \Rightarrow Z = 1.6$

<u>Case 2</u>: $n = 100,\ \bar{X} = 171 \Rightarrow Z = 2$

◆ PROC TTEST

```
PROC TTEST DATA=onesample H0=170;
   VAR x;

PROC TTEST DATA=paired;
   PAIRED pre*post;

PROC TTEST DATA=twosample;
   CLASS smoke;
   VAR bwt;
```

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
bwt	Pooled	Equal	187	2.63	0.0092
bwt	Satterthwaite	Unequal	170	2.71	0.0074

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
bwt	Folded F	114	73	1.30	0.2290

◆二獨立樣本 t 檢定

✦ $\sigma_1^2 = \sigma_2^2$

$$\frac{\overline{x}_1 - \overline{x}_2}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

，其中 $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$

✦ $\sigma_1^2 \neq \sigma_2^2$

$$\frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \sim t_{df}$$

PROC REG

資料來源

Body Fat Data (p.261, Neter et al. (1996))

x_1 : triceps skinfold thickness;

x_2 : thigh circumference

x_3 : midarm circumference;

y : body fat

```
DATA bodyfat;
   INPUT x1 x2 x3 y;
   CARDS;
19.5  43.1  29.1  11.9
24.7  49.8  28.2  22.8
30.7  51.9  37.0  18.7
29.8  54.3  31.1  20.1
19.1  42.2  30.9  12.9
25.6  53.9  23.7  21.7
31.4  58.5  27.6  27.1
27.9  52.1  30.6  25.4
22.1  49.9  23.2  21.3
25.5  53.5  24.8  19.3
31.1  56.6  30.0  25.4
30.4  56.7  28.3  27.2
18.7  46.5  23.0  11.7
19.7  44.2  28.6  17.8
14.6  42.7  21.3  12.8
29.5  54.4  30.1  23.9
27.7  55.3  25.7  22.6
30.2  58.6  24.6  25.4
22.7  48.2  27.1  14.8
25.2  51.0  27.5  21.1
;
```

```
PROC PLOT DATA=bodyfat VPERCENT=50 HPERCENT=33;
   PLOT y*(x1 x2 x3);
   PLOT x1*(x2 x3) x2*x3;

PROC CORR DATA=bodyfat NOSIMPLE;

PROC REG DATA=bodyfat;
   MODEL y=x1;
   MODEL y=x2;
   MODEL y=x3;
   MODEL y=x1 x2 x3/STB;

RUN;
```

```
    y*x1. A=1, B=2, etc.

 y |
30 +
   |
   |               AA
   |
25 +              AAA
   |               A
   |           A A
   |         A  B
20 +               A
   |            A  A
   |        A
   |
15 +          A
   |
   |    A  AA
   |       A
10 +
   |
   --+------+------+------+-
    10     20     30     40

               x1
```

```
    y*x2. A=1, B=2, etc.

 y |
30 +
   |
   |                  A A
   |
25 +             A    A A
   |                A
   |           A     A
   |           AA  A
20 +                A
   |             A A
   |     A
   |
15 +         A
   |
   |  AB
   |       A
10 +
   |
   -+----------+----------+-
   40        50         60

               x2
```

```
    y*x3. A=1, B=2, etc.

 y |
30 +
   |
   |        AA
   |
25 +     A     AA
   |           A
   |      A  A
   |    B   A
20 +            A
   |     A             A
   |         A
   |
15 +        A
   |
   | A        A A
   |   A
10 +
   |
   -+----------+----------+-
   20        30         40

               x3
```

```
   x1*x2. A=1, B=2, etc.           x1*x3. A=1, B=2, etc.           x2*x3. A=1, B=2, etc.

40 +                            40 +                            60 +
   |                               |                               |     A  A
   |                               |                               |
   |                               |                               |         A A
   |                               |                            x2 |      A
   |                  A A          |        A  A                   |    A      AA
30 +             A  B A A       30 +     A   A AA      A           |     A
   |             A   A             |      A     A                  |            A      A
   |                               |                               |        A
x1 |           AA  B            x1 |    AA  AA                  50 +    A    A
   |         A                     |        A                      |
   |           A                   |    A                          |        A
20 +   A A                      20 +         AA                    |   A
   |  A    A                       |   A        A                  |
   |                               |                               |         A
   |   A                           | A                             |          A
   |                               |                               | A          A
   |                               |                               |
10 +                            10 +                            40 +
   -+----------+----------+-       -+----------+----------+-       -+----------+----------+-
   40         50         60        20         30         40        20         30         40

               x2                              x3                              x3
```

The CORR Procedure

4 Variables: x1 x2 x3 y

Pearson Correlation Coefficients, N = 20
Prob > |r| under H0: Rho=0

	x1	x2	x3	y
x1	1.00000	0.92384 <.0001	0.45778 0.0424	0.84327 <.0001
x2	0.92384 <.0001	1.00000	0.08467 0.7227	0.87809 <.0001
x3	0.45778 0.0424	0.08467 0.7227	1.00000	0.14244 0.5491
y	0.84327 <.0001	0.87809 <.0001	0.14244 0.5491	1.00000

The REG Procedure
Model: MODEL1
Dependent Variable: y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	352.26980	352.26980	44.30	<.0001
Error	18	143.11970	7.95109		
Corrected Total	19	495.38950			

Root MSE	2.81977	R-Square	0.7111
Dependent Mean	20.19500	Adj R-Sq	0.6950
Coeff Var	13.96271		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-1.49610	3.31923	-0.45	0.6576
x1	1	0.85719	0.12878	6.66	<.0001

The REG Procedure
Model: MODEL1
Dependent Variable: y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	381.96582	381.96582	60.62	<.0001
Error	18	113.42368	6.30132		
Corrected Total	19	495.38950			

Root MSE	2.51024	R-Square	0.7710
Dependent Mean	20.19500	Adj R-Sq	0.7583
Coeff Var	12.43002		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-23.63449	5.65741	-4.18	0.0006
x2	1	0.85655	0.11002	7.79	<.0001

The REG Procedure
Model: MODEL1
Dependent Variable: y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	10.05160	10.05160	0.37	0.5491
Error	18	485.33790	26.96322		
Corrected Total	19	495.38950			

Root MSE	5.19261	R-Square	0.0203
Dependent Mean	20.19500	Adj R-Sq	-0.0341
Coeff Var	25.71236		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	14.68678	9.09593	1.61	0.1238
x3	1	0.19943	0.32663	0.61	0.5491

The REG Procedure
Model: MODEL1
Dependent Variable: y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	396.98461	132.32820	21.52	<.0001
Error	16	98.40489	6.15031		
Corrected Total	19	495.38950			

Root MSE	2.47998	R-Square	0.8014
Dependent Mean	20.19500	Adj R-Sq	0.7641
Coeff Var	12.28017		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate
Intercept	1	117.08469	99.78240	1.17	0.2578	0
x1	1	4.33409	3.01551	1.44	0.1699	4.26370
x2	1	-2.85685	2.58202	-1.11	0.2849	-2.92870
x3	1	-2.18606	1.59550	-1.37	0.1896	-1.56142

總結整理:

模型中之變數	R^2	$adj-R^2$	$\sqrt{MSE}$
x_1	0.7111	0.6950	2.81977
x_2	0.7710	0.7583	2.51024
x_3	0.0203	−0.0341	5.19261
x_1, x_2, x_3	0.8014	0.7641	2.47998

模型中之變數	b_1	b_2	b_3
x_1	0.85719 (0.12878)		
x_2		0.85655 (0.11002)	
x_3			0.19943 (0.32663)
x_1, x_2, x_3	4.33409 (3.01551)	−2.85685 (2.58202)	−2.18606 (1.59550)

待回答問題：

(1) 何以「整體模式」的檢定是顯著的，但是個別變數的檢定卻沒有一項是顯著的?

(2) 哪一個變數是「最重要」的變數? 可否利用標準化迴歸係數來做判斷?

(3) $adj-R^2<0$ 如何解釋?

A. (a) 如何解釋 β_k？

模型：$Y \sim N(\mu_Y,\ \sigma^2)$，其中 $\mu_Y = \mu_Y(x_1,\ldots,x_p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$

$$\beta_k = \mu_Y(x_1,\ldots,x_{k-1},x_k+1,x_{k+1},\ldots,x_p) - \mu_Y(x_1,\ldots,x_{k-1},x_k,x_{k+1},\ldots,x_p)$$
$$= (\beta_0 + \beta_1 x_1 + \cdots + \beta_{k-1}x_{k-1} + \beta_k(x_k+1) + \beta_{k+1}x_{k+1} + \cdots + \beta_p x_p)$$
$$-(\beta_0 + \beta_1 x_1 + \cdots + \beta_{k-1}x_{k-1} + \beta_k x_k + \beta_{k+1}x_{k+1} + \cdots + \beta_p x_p)$$

(b) β_k 之檢定

$H_0 : \beta_k = 0$ vs. $H_1 : \beta_k \neq 0$

亦即 $H_0 : \mu_Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{k-1}x_{k-1} + \beta_{k+1}x_{k+1} + \cdots + \beta_p x_p$

vs. $H_1 : \mu_Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{k-1}x_{k-1} + \beta_k x_k + \beta_{k+1}x_{k+1} + \cdots + \beta_p x_p$ 。

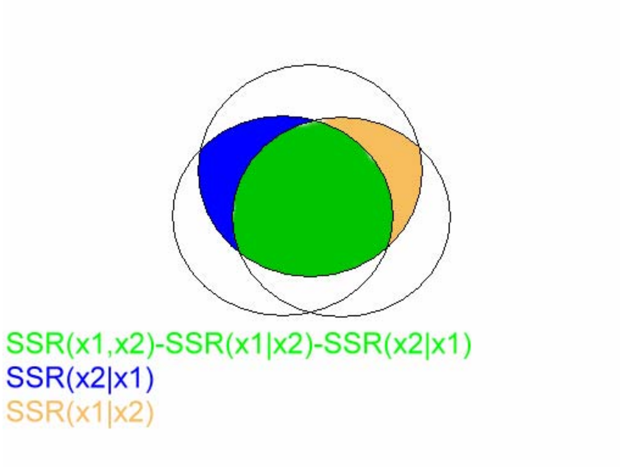
檢定統計量：$t^* = \dfrac{b_k}{s(b_k)}$ 。

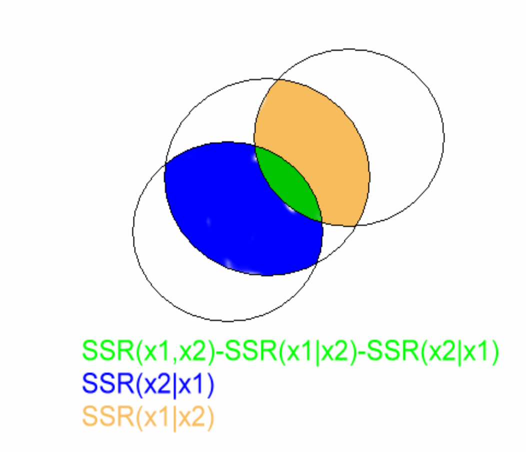
決策法則：若 $|t^*| > t(1-\alpha/2;\ n-p-1)$，則棄卻 H_0 。

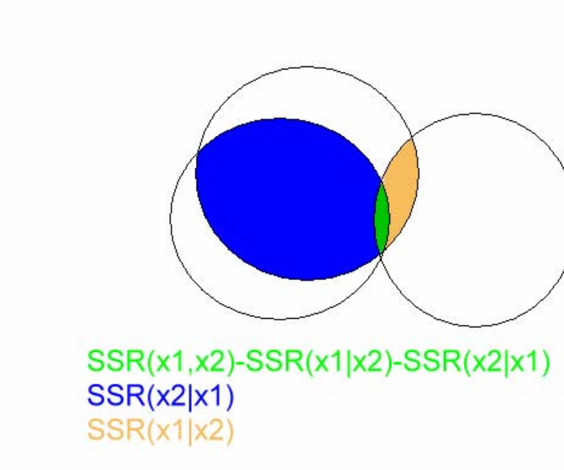
注意：

檢定顯著，並不意謂著 x_k 就是個「重要」變數；相反的，

檢定不顯著，也並不意謂著 x_k 就不會是個「重要」變數。


SSR(x1,x2)-SSR(x1|x2)-SSR(x2|x1)
SSR(x2|x1)
SSR(x1|x2)


SSR(x1,x2)-SSR(x1|x2)-SSR(x2|x1)
SSR(x2|x1)
SSR(x1|x2)


SSR(x1,x2)-SSR(x1|x2)-SSR(x2|x1)
SSR(x2|x1)
SSR(x1|x2)

(c) $H_0: \beta_1 = \cdots = \beta_p = 0$ vs. $H_1:$ not H_0

$(\Leftrightarrow\ H_0: \mu_Y = \beta_0$ vs. $H_1: \mu_Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)$

✦ 這並不是一個「適合度」檢定(goodness-of-fit test)

(d) ◇ 如果 R^2 很小，或者 p 很大時，

$$adj - R^2 = 1 - \frac{n-1}{n-p-1}\frac{SSE}{SSTO} = 1 - \frac{n-1}{n-p-1}(1-R^2) < 0$$

◇ $adj - R^2 \xleftrightarrow{1-1} MSE,\ \because adj - R^2 = 1 - \frac{SSE/(n-p-1)}{SSTO/n-1} = 1 - \frac{MSE}{SSTO/n-1}$ 。

(e) R^2 vs. MSE

◇ $R^2 = 1 - \frac{SSE}{SSTO}$ ，其中 $SSE = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - (b_0 + b_1 x_{i1} + \cdots + b_p x_{ip}))^2$

◇ $MSE = \hat{\sigma}^2$

(f) 模型 4 仍可用來預測 y，但不能用來解釋 x_1 、 x_2 、 x_3 個別對 y 的影響。

模型中之變數	MSE	$\hat{y}_h$	$s(\hat{y}_h)$
x_1	7.95	19.93	0.632
x_1, x_2	6.47	19.36	0.624
x_1, x_2, x_3	6.15	19.19	0.621

其中 $x_1 = 25,\ x_2 = 50,\ x_3 = 29$ ，亦即

$$\hat{y}_h = b_0 + b_1x_1 + b_2x_2 + b_3x_3 = b_0 + 25b_1 + 50b_2 + 29b_3$$

<u>說明</u>：

令 $x_2 = 2x_1$ ，則 $y = 3 + 4x_1 + x_2$

$$\begin{aligned} &= 3 + 2x_1 + 2x_2 \\ &= 3 + \qquad + 3x_2 \\ &= 3 - 2x_1 + 4x_2 \\ &= \cdots \text{。} \end{aligned}$$

我們可以發現 y 的值維持不變，但是 x_1 和 x_2 的係數可以有無限多種不同組合。

Body Fat Data

模型中之變數	R^2	$adj-R^2$	$\sqrt{MSE}$
x_1	0.7111	0.6950	2.81977
x_2	0.7710	0.7583	2.51024
x_3	0.0203	−0.0341	5.19261
x_1, x_2	0.7781	0.7519	2.54317
x_1, x_3	0.7862	0.7610	2.49628
x_2, x_3	0.7757	0.7493	2.55653
x_1, x_2, x_3	0.8014	0.7641	2.47998

模型中之變數	b_1	b_2	b_3
x_1	0.85719		
x_2		0.85655	
x_3			0.19943
x_1, x_2	0.22235	0.65942	
x_1, x_3	1.00058		−0.43144
x_2, x_3		0.85088	0.09603
x_1, x_2, x_3	4.33409	−2.85685	−2.18606

PROC LOGISTIC

⦿ 線性迴歸與邏輯斯迴歸模型

✸ 線性迴歸模型(Linear Regression Models)

$$Y = \mu_Y(x_1,\ldots,x_p) + \varepsilon$$

<u>基本假設</u>

1. Y 必須是一連續型變數(continuous variable)。

2. $\mu_Y(x_1,\ldots,x_p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$

3. $\varepsilon \sim \text{i.i.d.} N(0,\sigma^2)$

   (i.i.d.—identically and independently distributed)

(附註)

$x_1,\ldots,x_p$ 可以是連續型變數也可以是離散型變數(discrete variable)。

—變異數分析(analysis of variance)—$x_1,\ldots,x_p$ 全部都是離散型變數。

—共變異數分析(analysis of covariance)—$x_1,\ldots,x_p$ 部分為連續型，部分為離散型。

資料來源：附錄 1，Hosmer and Lemeshow(1989)。

變數名稱：ID(編號)

LOW(新生兒的體重是否過輕；0 代表體重大於等於 2500g，1 代表體重小於 2500g)

AGE(產婦的年紀)

LWT(產婦懷孕時的體重，單位為磅)

RACE(人種；1=白人，2=黑人，3=其他)

SMOKE(產婦在懷孕過程中是否抽煙；0=否，1=是)

PTL(早產紀錄；0=沒有，1=1 次，2=2 次，等)

HT(是否有高血壓的病歷；0=否，1=是)

UT(是否有尿道感染症狀；0=否，1=是)

FTV(懷孕前三個月內所作的產檢次數；0=沒有，1=1 次，2=2 次，等)

BWT(新生兒的體重，單位為公克)

資料總數： 189 筆新生兒資料

(1)簡單線性機率模型(Simple Linear Probability Model)：$P(Y=1|x)=\beta_0+\beta_1 x$

令$Y=LOW$ ， $X=LWT$

$\Rightarrow \hat{P}(LOW=1)=0.6467-0.0026LWT$

—“可能”解釋：大體上而言，產婦懷孕時的體重每增加一磅，新生兒的體重過輕

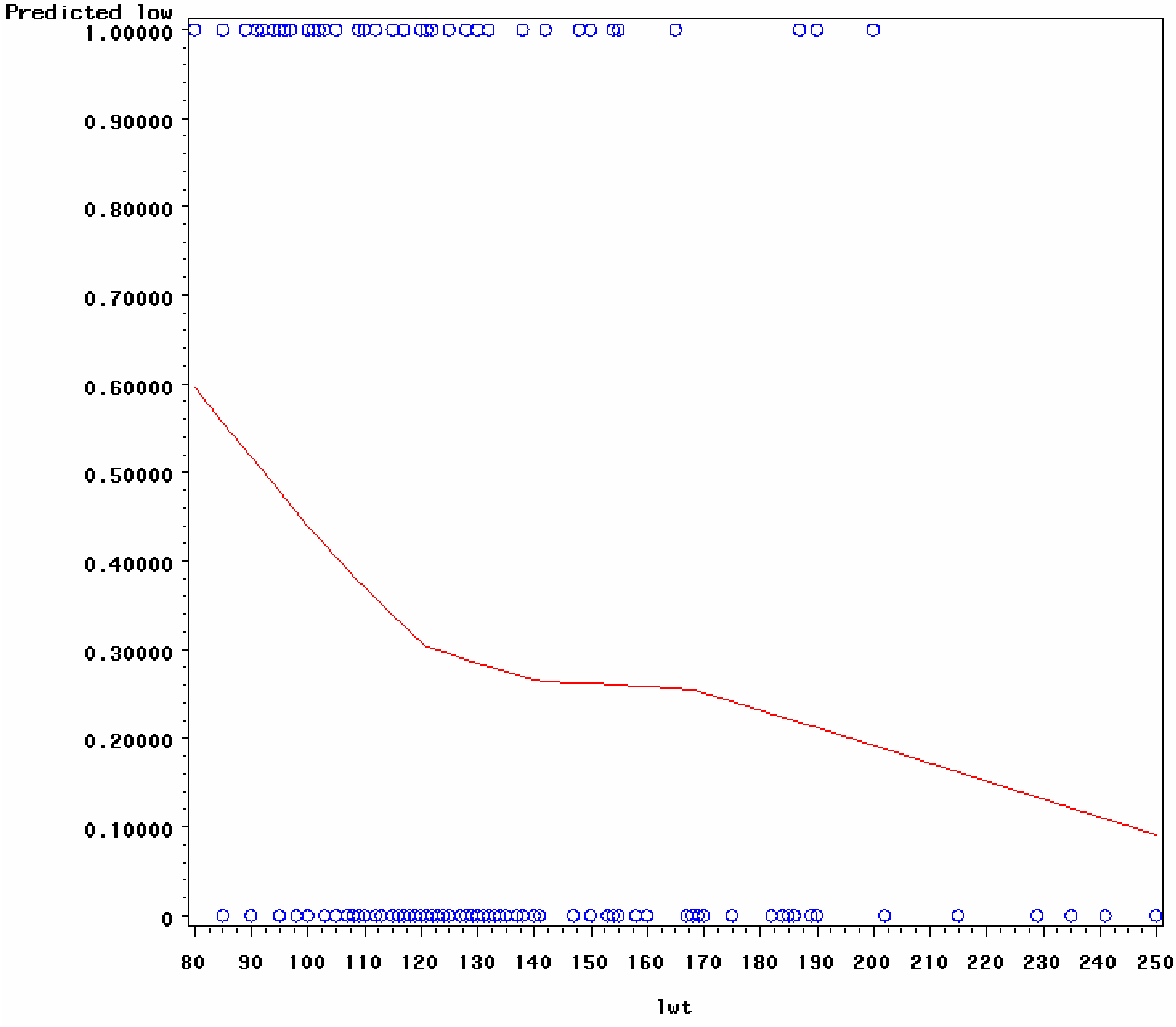
的機率將減少 0.26 個百分點。

—模型是否適用？

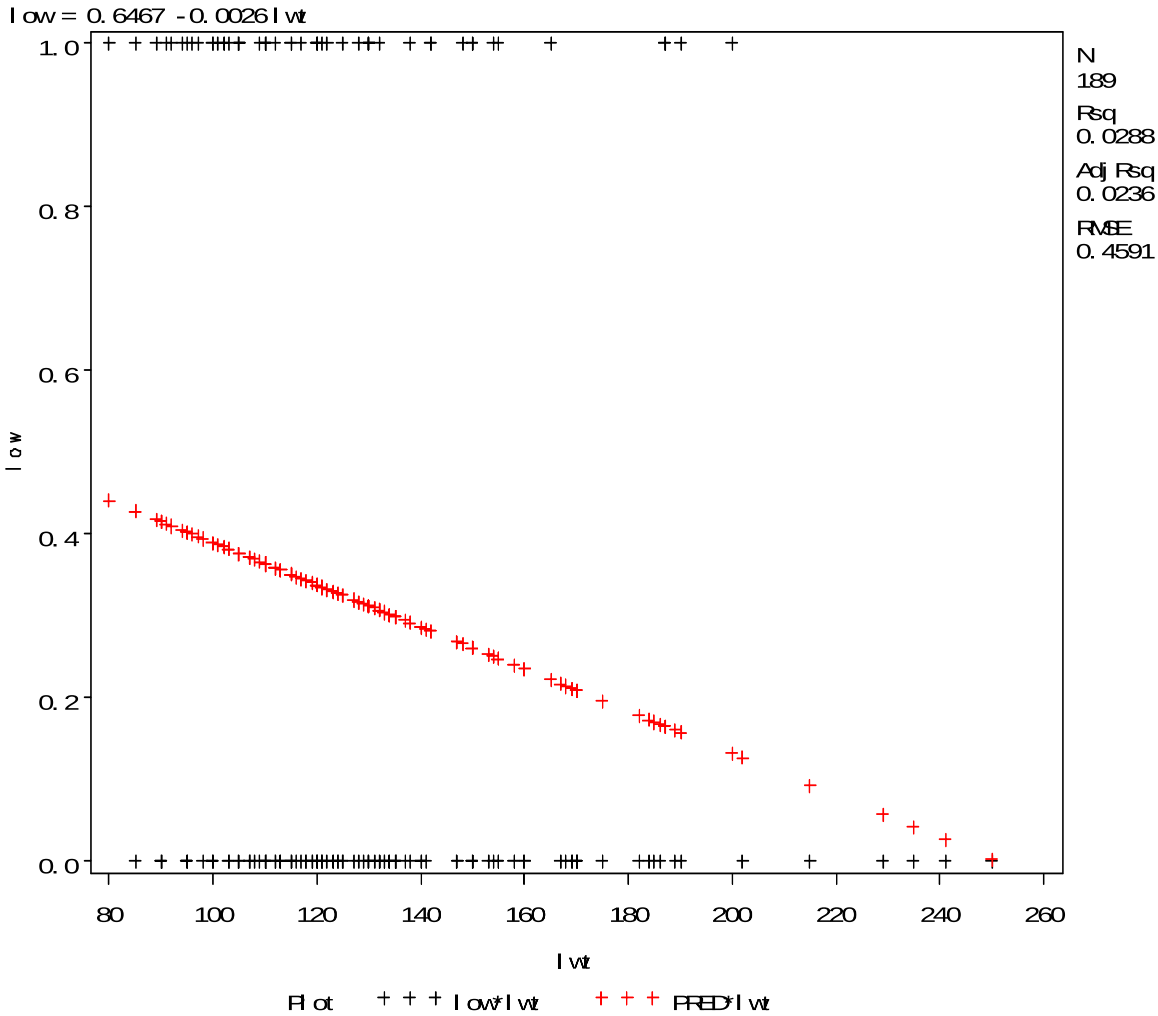
①殘差顯然是LWT 的函數

②殘差並非常態分配

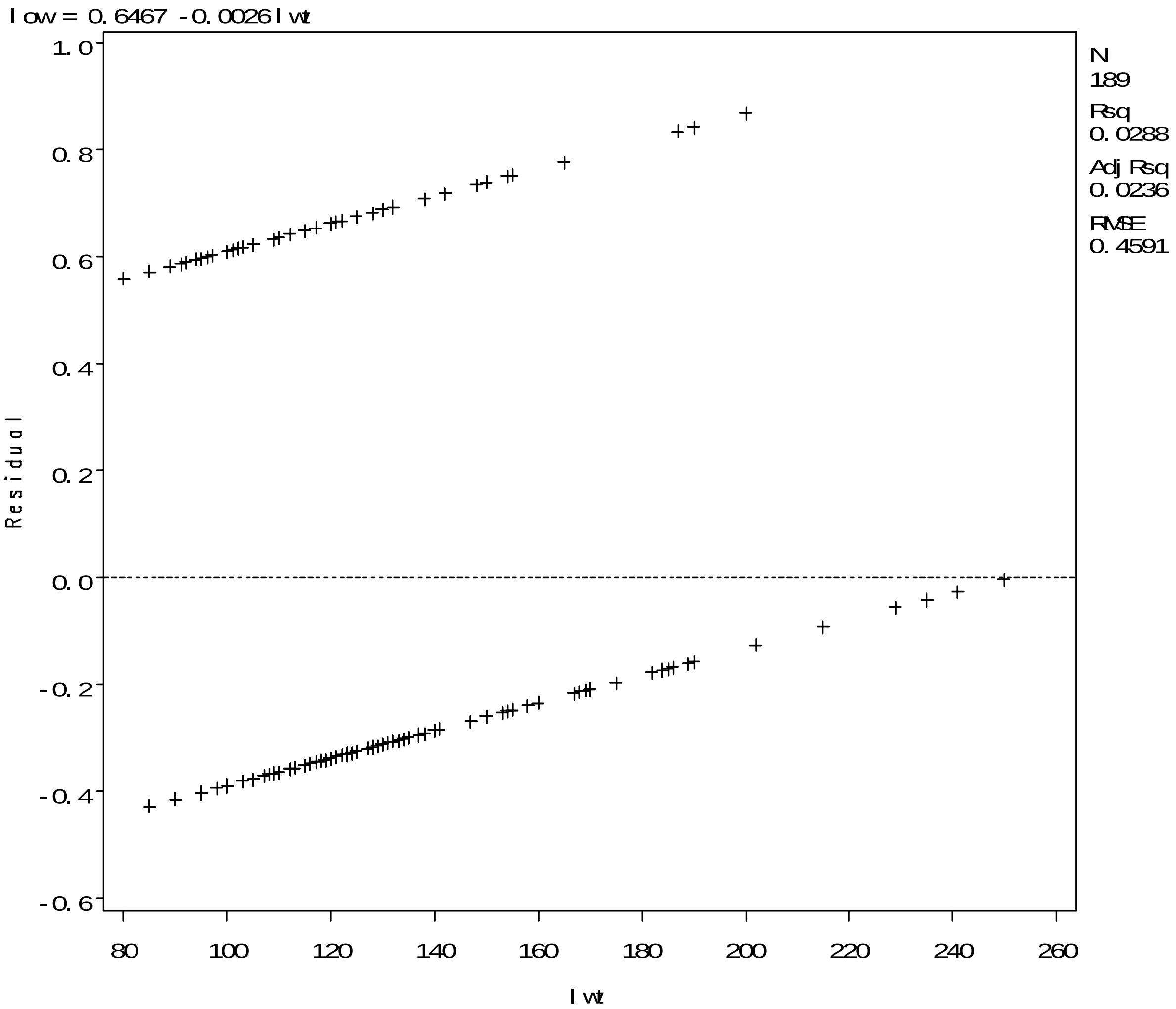
$\Rightarrow$相關議題的推論可能都不正確



Predicted low
1.00000
0.90000
0.80000
0.70000
0.60000
0.50000
0.40000
0.30000
0.20000
0.10000
0
80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250
lwt



low = 0.6467 -0.0026 lwt
1.0
0.8
0.6
0.4
0.2
0.0
low
80
100
120
140
160
180
200
220
240
260
lwt
N
189
Rsq
0.0288
Adj Rsq
0.0236
RMSE
0.4591
Plot
low*lwt
PRED*lwt


low = 0.6467 -0.0026 lwt
N
189
Rsq
0.0288
Adj Rsq
0.0236
RMSE
0.4591
Residual
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
80
100
120
140
160
180
200
220
240
260
lwt

—可行方案，嘗試進行變數轉換後，再作分析。

—如何作轉換？

①繪製 $P(Y=1|x)$ 相對於 x 的圖

②就模型 $P(Y=1|x)=\beta_0+\beta_1 x$ 而言，$0\le P(Y=1|x)\le 1$，然而 $\beta_0+\beta_1 x\in R$。

如何改進這個缺失：

$$0\le P(Y=1|x)\le 1$$

$$\Rightarrow 0\le \frac{P(Y=1|x)}{1-P(Y=1|x)}$$

$$\Rightarrow \log\frac{P(Y=1|x)}{1-P(Y=1|x)}\in R$$

(附註)

1.odds(成敗比、勝算、優勢)及 logit

$$odds(Y=1\,|\,x)=\frac{P(Y=1\,|\,x)}{1-P(Y=1\,|\,x)}=\frac{P(Y=1\,|\,x)}{P(Y=0\,|\,x)}$$

$$\begin{aligned}\mathrm{logit}(Y\,|\,x)&=\log(odds(Y=1\,|\,x))\\&=\log\frac{P(Y=1\,|\,x)}{1-P(Y=1\,|\,x)}\\&=\log\frac{P(Y=1\,|\,x)}{P(Y=0\,|\,x)}\end{aligned}$$

例：{40 女生，60 男生}

$$\Rightarrow P(\text{男生})=0.6\text{，}odds(\text{男生})=\frac{P(\text{男生})}{1-P(\text{男生})}=\frac{P(\text{男生})}{P(\text{女生})}=\frac{0.6}{0.4}=\frac{3}{2}=1.5$$

說明：男生所佔的比例是 60%；男女的比例是 3 比 2，男生是女生的 1.5 倍，或男生的 $odds$ (成敗比)是 1.5。

2.(a) $$odds(Y=1)=\frac{P(Y=1)}{1-P(Y=1)}=\frac{P(Y=1)}{P(Y=0)}$$

(b) $$P(Y=1)=\frac{odds(Y=1)}{1+odds(Y=1)}\left(=\frac{e^{\mathrm{logit}(Y)}}{1+e^{\mathrm{logit}(Y)}}\right)$$

(2)簡單邏輯斯迴歸模型(Simple Logistic Regression Model)：

$$\text{logit}(Y\,|\,x) = \log\frac{P(Y=1\,|\,x)}{1-P(Y=1\,|\,x)} = \beta_0 + \beta_1 x$$

$$(\Leftrightarrow P(Y=1\,|\,x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}})$$

$Y = LOW$ ， $X = LWT$ $\Rightarrow$ $\widehat{\text{logit}}(LOW) = 0.998 - 0.014LWT$

✸邏輯斯迴歸模型(Logistic Regression Models)

$$\text{logit}(Y \mid x_1,\ldots,x_p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p \quad \left(\Leftrightarrow P(Y=1 \mid x_1,\ldots,x_p) = \frac{e^{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p}}{1+e^{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p}}\right)$$

其中$Y \in \{0,1\}$，$x_1,\ldots,x_p$可以是連續型變數，也可以是離散型變數。

(附註)

1.如果自變數x_i為一名目變數(nominal variable)(比如，性別、人種、婚姻狀況等等)，則如同線性迴歸的處理方式一般，我們需要考慮所謂的假變數(dummy variables)。

例：(1) x_1(性別)$\in$ {男、女}

如果為男生，則可令$x_{11}=0$；要不然令$x_{11}=1$。

(2) x_2(人種)$\in$ {白人、黑人、其他}

假變數

人種	x_{21}	x_{22}
白人	0	0
黑人	1	0
其他	0	1

(3) x_3(學歷)$\in${小學、初中、高中、大學以上}

假變數

學歷	x_{31}	x_{32}	x_{33}
小學	1	0	0
初中	0	1	0
高中	0	0	1
大學以上	0	0	0

2.依此類推，就一個擁有c個可能選項的名目變數而言，我們需要定義$c-1$個假變數。

⊙ 評估模型配適好壞的幾個統計量

(1)線性迴歸

—$R^2 = \dfrac{SSR}{SSTO} = 1 - \dfrac{SSE}{SSTO}$， 其中 $SSE = \sum (Y_i - \hat{Y}_i)^2$ ，$SSTO = \sum (Y_i - \overline{Y})^2$ 。

—$F = \dfrac{SSR/p}{SSE/N-(p+1)}$

(附註)

1. R^2 ：判定係數(coefficient of determination)

    —變異數可以被解釋的比例

    —誤差減少的比例

2. F 統計量可以用來進行 $H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$ (或 $H_0 : R^2 = 0$)的檢定，藉以了解利用 $\hat{Y}$ 來做配適是否真的比 $\overline{Y}$ 來的好。

3. $F = \dfrac{R^2/p}{(1-R^2)/(N-p-1)}$ ，$R^2 = \dfrac{pF}{pF+N-p-1}$

4. 理想的情況是 F 和 R^2 都很大。但是 F 值很大、而 R^2 很小，或者 F 很小、而 R^2 很大的情形，也可能發生。

(2)邏輯斯迴歸

假定 L 為概似函數(likelihood function)，則 $-2\log L \xrightarrow{D} \chi^2_{\text{df}}$ 。

$-2\log L$ 的值越大，通常代表模型配適的情形越差。

<u>定義</u>：

$$D_0 = -2\log L\,|_{\text{logit}(Y)=\beta_0}$$

(Intercept Only) (PROC LOGISTIC)

$$D_M = -2\log L\,|_{\text{logit}(Y)=\beta_0+\beta_1 x_1+\cdots+\beta_p x_p}$$

(Intercept and Covariate) (PROC LOGISTIC)

$$G_M = D_0 - D_M$$

(Chi-Square for Covariates) (PROC LOGISTIC)

(附註)

1. $D_0 \leftrightarrow SSTO$

   $D_M \leftrightarrow SSE$

   $G_M \leftrightarrow SSR$

2. $D_0 \geq D_M \Rightarrow G_M \geq 0$

3. G_M 是一個可以用來檢定 $H_0:\beta_1=\beta_2=\cdots=\beta_p=0$ 的統計量。

4. D_M 也可以作為一個模型配適好壞的指標。此外，Hosmer and Lemeshow Goodness-of-Fit Test 也有相同的目的。

5. 理想的情況是 G_M 很大，而 D_M 值很小。不過通常 G_M 是首要考量。

◉ 邏輯斯迴歸模型的解釋及推論

(1)簡單模型(自變數為連續型變數)

$$Y \sim \text{Bernoulli}(\pi(x))$$

其中 $\pi(x) = P(Y=1 \mid x) = \dfrac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$

$$\Rightarrow \text{logit}(Y \mid x) = \log\frac{\pi(x)}{1-\pi(x)} = \beta_0 + \beta_1 x$$

✦ β_1 的解釋

$$\begin{aligned}
\beta_1 &= \text{logit}(Y \mid x+1) - \text{logit}(Y \mid x) \\
&= \log\frac{\pi(x+1)}{1-\pi(x+1)} - \log\frac{\pi(x)}{1-\pi(x)} \\
&= \log\frac{\pi(x+1)/(1-\pi(x+1))}{\pi(x)/(1-\pi(x))} \\
&= \log\frac{odds(Y=1 \mid x+1)}{odds(Y=1 \mid x)}
\end{aligned}$$

(附註)

1. β_1 代表 x 每增加一個單位，logit 的變化量。

$\beta_1 > 0 \Rightarrow e^{\beta_1} > 1 \ \Rightarrow$ 觀測到 $Y=1$ 的機會會隨著 x 的增加而增加。

2. $e^{\beta_1} = \dfrac{odds(Y=1 \mid x+1)}{odds(Y=1 \mid x)}$ ： odds − ratio (相對成敗比、勝算比、優勢比)

亦即 x 每增加一個單位，$Y=1$ 的成敗比變成為原來的 e^{β_1} 倍。

3. x 的變化對 $\pi(x)$ 的影響：

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} \ \Rightarrow \frac{d}{dx}\pi(x) = \beta\pi(x)\left[1-\pi(x)\right]$$

因此 $P(Y=1)$ 的變動率為 $\beta\pi(x)\left[1-\pi(x)\right]$。

✦ $H_0:\beta_1=0$ 的檢定

(亦即 $H_0:\text{logit}(Y\mid x)=\beta_0$ vs. $H_1:\text{logit}(Y\mid x)=\beta_0+\beta_1 x_1$)

在 H_0 成立的前提下，$Z=\dfrac{\hat{\beta}_1}{\text{s.e.}(\hat{\beta}_1)}\xrightarrow{D}N(0,1)$ (單尾或雙尾檢定)

$$W=\left(\frac{\hat{\beta}_1}{\text{s.e.}(\hat{\beta}_1)}\right)^2\xrightarrow{D}\chi_1^2 \quad \text{(雙尾檢定)}$$

$$G_M=D_0-D_M\xrightarrow{D}\chi_1^2 \quad \text{(雙尾檢定)}$$

(附註)

1. W 稱為 Wald 統計量。

2. W 和 G_M 這兩種檢定方式在樣本數足夠大的情形下，可以視為是相同的檢定。不過在實際應用上，G_M 通常較為可靠。

(2)簡單模型(自變數為名目變數)

假設x(人種) $\in\{$白人、黑人、其他$\}$

設立假變數如下：

人種	x_{11}	x_{12}
白人	0	0
黑人	1	0
其他	0	1

考慮模式$\text{logit}(Y \mid x) = \beta_0 + \beta_{11}x_{11} + \beta_{12}x_{12}$

亦即 $\text{logit}(Y \mid \text{白人}) = \text{logit}(Y \mid x_{11} = 0, x_{12} = 0) = \beta_0$

$\text{logit}(Y \mid \text{黑人}) = \text{logit}(Y \mid x_{11} = 1, x_{12} = 0) = \beta_0 + \beta_{11}$

$\text{logit}(Y \mid \text{其他}) = \text{logit}(Y \mid x_{11} = 0, x_{12} = 1) = \beta_0 + \beta_{12}$

✦ β_{11},β_{12} 的解釋

情況 1

$$\text{logit}(Y\,|\,\text{黑人}) - \text{logit}(Y\,|\,\text{白人}) = (\beta_0 + \beta_{11}) - \beta_0 = \beta_{11}$$

$$\Rightarrow \frac{odds(\text{黑人})}{odds(\text{白人})} = e^{\beta_{11}}$$

亦即黑人產婦生出體重過輕嬰兒的成敗比是白人產婦的 $e^{\beta_{11}}$ 倍。

情況 2

$$\text{logit}(Y\,|\,\text{其他}) - \text{logit}(Y\,|\,\text{白人}) = (\beta_0 + \beta_{12}) - \beta_0 = \beta_{12}$$

$$\Rightarrow \frac{odds(\text{其他})}{odds(\text{白人})} = e^{\beta_{12}}$$

亦即其他人種的產婦生出體重過輕嬰兒的成敗比是白人產婦的 $e^{\beta_{12}}$ 倍。

情況 3

$$\text{logit}(Y \mid \text{其他}) - \text{logit}(Y \mid \text{黑人}) = (\beta_0 + \beta_{12}) - (\beta_0 + \beta_{11}) = \beta_{12} - \beta_{11}$$

$$\Rightarrow \frac{odds(\text{其他})}{odds(\text{黑人})} = e^{\beta_{12} - \beta_{11}}$$

亦即其他人種的產婦生出體重過輕嬰兒的成敗比是黑人產婦的 $e^{\beta_{12}-\beta_{11}}$ 倍。

(附註)

$$\frac{odds(\text{其他})}{odds(\text{黑人})} = \frac{odds(\text{其他})/odds(\text{白人})}{odds(\text{黑人})/odds(\text{白人})} = \frac{e^{\beta_{12}}}{e^{\beta_{11}}} = e^{\beta_{12} - \beta_{11}}$$

✦ $H_0: \beta_{11} = \beta_{12} = 0$ 的檢定

(亦即 $H_0: \text{logit}(Y \mid x_{11}, x_{12}) = \beta_0$ vs.

$H_1: \text{logit}(Y \mid x_{11}, x_{12}) = \beta_0 + \beta_{11}x_{11} + \beta_{12}x_{12}$)

檢定統計量：(在 H_0 成立的前提下) $G_M = D_0 - D_M \xrightarrow{D} \chi_2^2$ 或 $W \xrightarrow{D} \chi_2^2$

(3)複迴歸模型(沒有交互作用項)

假設 x_1=產婦懷孕前的體重，x_2=人種(令 x_{21}, x_{22} 為對應的假變數)

考慮模型如下：

$$\text{logit}(Y \mid x_1, x_{21}, x_{22}) = \beta_0 + \beta_1 x_1 + \beta_{21} x_{21} + \beta_{22} x_{22}$$

$$\Rightarrow \text{logit}(Y \mid x_1, \text{白人}) = \text{logit}(Y \mid x_1, x_{21} = 0, x_{22} = 0) = \beta_0 + \beta_1 x_1$$

$$\text{logit}(Y \mid x_1, \text{黑人}) = \text{logit}(Y \mid x_1, x_{21} = 1, x_{22} = 0) = (\beta_0 + \beta_{21}) + \beta_1 x_1$$

$$\text{logit}(Y \mid x_1, \text{其他}) = \text{logit}(Y \mid x_1, x_{21} = 0, x_{22} = 1) = (\beta_0 + \beta_{22}) + \beta_1 x_1$$

✦ $\beta_1, \beta_{21}, \beta_{22}$ 的解釋

在 x_1 固定不變的情形下(亦即產婦懷孕時的體重相同的情形下)

$$\text{logit}(Y \mid x_1, \text{黑人}) - \text{logit}(Y \mid x_1, \text{白人}) = \beta_{21}$$

$$\text{logit}(Y \mid x_1, \text{其他}) - \text{logit}(Y \mid x_1, \text{白人}) = \beta_{22}$$

$$\text{logit}(Y \mid x_1, \text{其他}) - \text{logit}(Y \mid x_1, \text{黑人}) = \beta_{22} - \beta_{21}$$

在人種固定不變的情形下(亦即就相同人種的考量下)

$$\text{logit}(Y \mid x_1 + 1, x_{11}, x_{12}) - \text{logit}(Y \mid x_1, x_{11}, x_{12}) = \beta_1$$

✦ 假設檢定

① $H_0:\beta_1=0$ vs. $H_1:\beta_1\neq 0$

(亦即 $H_0:\text{logit}(Y\mid x_1,x_{21},x_{22})=\beta_0+\beta_{21}x_{21}+\beta_{22}x_{22}$ vs.

$H_1:\text{logit}(Y\mid x_1,x_{21},x_{22})=\beta_0+\beta_1x_1+\beta_{21}x_{21}+\beta_{22}x_{22}$)

檢定統計量：$Z=\dfrac{\hat{\beta}_1}{\text{s.e.}(\hat{\beta}_1)}\xrightarrow{D}N(0,1)$(在 H_0 成立的情形下)

$$W=\left(\frac{\hat{\beta}_1}{\text{s.e.}(\hat{\beta}_1)}\right)^2\xrightarrow{D}\chi_1^2$$

$$\Delta G=G_{M_1}-G_{M_2}\xrightarrow{D}\chi_1^2$$

其中 $G_{M_1}=D_0-D_{M_1}$，$G_{M_2}=D_0-D_{M_2}$

M_1 代表模型 $\text{logit}(Y\mid x_1,x_{21},x_{22})=\beta_0+\beta_1x_1+\beta_{21}x_{21}+\beta_{22}x_{22}$

M_2 代表模型 $\text{logit}(Y\mid x_1,x_{21},x_{22})=\beta_0+\beta_{21}x_{21}+\beta_{22}x_{22}$

$(\Rightarrow \Delta G=D_{M_2}-D_{M_1})$

② $H_0 : \beta_{21} = \beta_{22} = 0$

(亦即 $\text{logit}(Y \mid x_1, x_{21}, x_{22}) = \beta_0 + \beta_1 x_1$ vs.

$\text{logit}(Y \mid x_1, x_{21}, x_{22}) = \beta_0 + \beta_1 x_1 + \beta_{21} x_{21} + \beta_{22} x_{22}$)

檢定統計量：(在 H_0 成立的情形下) $\Delta G = G_{M_1} - G_{M_2} \xrightarrow{D} \chi_2^2$ 或 $W \xrightarrow{D} \chi_2^2$

其中 M_1 代表模型 $\text{logit}(Y \mid x_1, x_{21}, x_{22}) = \beta_0 + \beta_1 x_1 + \beta_{21} x_{21} + \beta_{22} x_{22}$

M_2 代表模型 $\text{logit}(Y \mid x_1, x_{21}, x_{22}) = \beta_0 + \beta_1 x_1$

③ $H_0 : \beta_1 = \beta_{21} = \beta_{22} = 0$

(亦即 $\text{logit}(Y \mid x_1, x_{21}, x_{22}) = \beta_0$ vs.

$\text{logit}(Y \mid x_1, x_{21}, x_{22}) = \beta_0 + \beta_1 x_1 + \beta_{21} x_{21} + \beta_{22} x_{22}$)

檢定統計量：(在 H_0 成立的情形下) $G_M = D_0 - D_M \xrightarrow{D} \chi_3^2$ 或 $W \xrightarrow{D} \chi_3^2$

其中 M 指的是模型 $\text{logit}(Y \mid x_1, x_{21}, x_{22}) = \beta_0 + \beta_1 x_1 + \beta_{21} x_{21} + \beta_{22} x_{22}$

(4)複迴歸模型(存在交互作用項)

假設 x_1=產婦懷孕時的體重，x_2=人種(令 x_{21}, x_{22} 為對應的假變數)

考慮模型如下：

$$\text{logit}(Y \mid x_1, x_{21}, x_{22}) = \beta_0 + \beta_1 x_1 + \beta_{21} x_{21} + \beta_{22} x_{22} + \beta_{31} x_1 x_{21} + \beta_{32} x_1 x_{22}$$

$$\Rightarrow \text{logit}(Y \mid x_1, \text{白人}) = \text{logit}(Y \mid x_1, x_{21} = 0, x_{22} = 0) = \beta_0 + \beta_1 x_1$$

$$\text{logit}(Y \mid x_1, \text{黑人}) = \text{logit}(Y \mid x_1, x_{21} = 1, x_{22} = 0) = \beta_0 + \beta_1 x_1 + \beta_{21} + \beta_{31} x_1$$

$$= (\beta_0 + \beta_{21}) + (\beta_1 + \beta_{31}) x_1$$

$$\text{logit}(Y \mid x_1, \text{其他}) = \text{logit}(Y \mid x_1, x_{21} = 0, x_{22} = 1) = \beta_0 + \beta_1 x_1 + \beta_{22} + \beta_{32} x_1$$

$$= (\beta_0 + \beta_{22}) + (\beta_1 + \beta_{32}) x_1$$

✦ 自變數的變化對 $\text{logit}(Y)$ 的影響

在 x_1 固定不變的情況下，

$$\text{logit}(Y \mid x_1, \text{黑人}) - \text{logit}(Y \mid x_1, \text{白人}) = \beta_{21} + \beta_{31} x_1$$

$$\text{logit}(Y \mid x_1, \text{其他}) - \text{logit}(Y \mid x_1, \text{白人}) = \beta_{22} + \beta_{32} x_1$$

$$\text{logit}(Y \mid x_1, \text{其他}) - \text{logit}(Y \mid x_1, \text{黑人}) = (\beta_{22} - \beta_{21}) + (\beta_{32} - \beta_{31}) x_1$$

(前述三個 logit 的變化量會隨著 x_1 的不同而改變)

在人種固定不變的情況下

$$\text{logit}(Y \mid x_1+1,\text{白人}) - \text{logit}(Y \mid x_1,\text{白人}) = \beta_1$$

$$\text{logit}(Y \mid x_1+1,\text{黑人}) - \text{logit}(Y \mid x_1,\text{黑人}) = \beta_1 + \beta_{31}$$

$$\text{logit}(Y \mid x_1+1,\text{其他}) - \text{logit}(Y \mid x_1,\text{其他}) = \beta_1 + \beta_{32}$$

(這三個logit的變化量也會隨著人種的不同而不同)

✦ $H_0 : \beta_{31} = \beta_{32} = 0$

(亦即 $H_0 : \text{logit}(Y \mid x_1, x_{21}, x_{22}) = \beta_0 + \beta_1 x_1 + \beta_{21} x_{21} + \beta_{22} x_{22}$ vs.

$H_1 : \text{logit}(Y \mid x_1, x_{21}, x_{22}) = \beta_0 + \beta_1 x_1 + \beta_{21} x_{21} + \beta_{22} x_{22} + \beta_{31} x_1 x_{21} + \beta_{32} x_1 x_{22}$)

檢定統計量：(在 H_0 成立的前提下)

$$\Delta G = G_{H_1} - G_{H_0} = D_{H_0} - D_{H_1} \xrightarrow{D} \chi_2^2 \quad \text{或}$$

$$W \xrightarrow{D} \chi_2^2$$

```
DATA lowbwt;
   INFILE 'c:\logistic\hosmer\data\appndix1.dat';
   INPUT id 1-3 low 4 age 5-6 lwt 7-9 race 10 smoke 11 ptl 12
         ht 13 ui 14 ftv 15 bwt 16-19;


PROC LOGISTIC DATA=lowbwt;       /* Model 1 */
   MODEL low(EVENT='1')=lwt/LACKFIT;

PROC LOGISTIC DATA=lowbwt;       /* Model 2 */
   CLASS race(REF='1')/PARAM=REF;
   MODEL low(EVENT='1')=race/LACKFIT;


PROC LOGISTIC DATA=lowbwt;       /* Model 3 */
   CLASS race(REF='1')/PARAM=REF;
   MODEL low(EVENT='1')=lwt race/LACKFIT;

PROC LOGISTIC DATA=lowbwt;       /* Model 4 */
   CLASS race(REF='1')/PARAM=REF;
   MODEL low(EVENT='1')=lwt|race/LACKFIT;


DATA lowbwt2;
   SET lowbwt;
   IF race=2 THEN r1=1;
      ELSE r1=0;
   IF race=3 THEN r2=1;
      ELSE r2=0;
   lwtr1=lwt*r1;
   lwtr2=lwt*r2;
   KEEP low lwt race r1 r2 lwtr1 lwtr2;

PROC LOGISTIC DATA=lowbwt2;      /* Model 2 */
   MODEL low(EVENT='1')=r1 r2/LACKFIT;
PROC LOGISTIC DATA=lowbwt2;      /* Model 3 */
   MODEL low(EVENT='1')=lwt r1 r2/LACKFIT;
PROC LOGISTIC DATA=lowbwt2;      /* Model 4 */
   MODEL low(EVENT='1')=lwt r1 r2 lwtr1 lwtr2/LACKFIT;


RUN;
```

Model Information

Data Set	WORK.LOWBWT
Response Variable	low
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	189
Number of Observations Used	189

Response Profile

Ordered Value	low	Total Frequency
1	0	130
2	1	59

Probability modeled is low=1.

Class Level Information

Class	Value	Design Variables	
race	1	0	0
	2	1	0
	3	0	1

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	236.672	231.259
SC	239.914	244.226
-2 Log L	234.672	223.259

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	11.4129	3	0.0097
Score	10.7572	3	0.0131
Wald	10.1316	3	0.0175

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
lwt	1	5.5886	0.0181
race	2	5.4024	0.0671

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	0.8057	0.8452	0.9088	0.3404
lwt		1	-0.0152	0.00644	5.5886	0.0181
race	2	1	1.0811	0.4881	4.9065	0.0268
race	3	1	0.4806	0.3567	1.8156	0.1778

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
lwt	0.985	0.973	0.997
race 2 vs 1	2.948	1.133	7.672
race 3 vs 1	1.617	0.804	3.253

Association of Predicted Probabilities and Observed Responses

Percent Concordant	64.1	Somers' D	0.293
Percent Discordant	34.8	Gamma	0.296
Percent Tied	1.1	Tau-a	0.127
Pairs	7670	c	0.647

The LOGISTIC Procedure

Partition for the Hosmer and Lemeshow Test

		low = 1		low = 0	
Group	Total	Observed	Expected	Observed	Expected
1	19	2	2.37	17	16.63
2	21	4	4.25	17	16.75
3	20	5	4.80	15	15.20
4	19	6	5.07	13	13.93
5	19	6	5.50	13	13.50
6	19	6	6.22	13	12.78
7	20	6	7.21	14	12.79
8	20	6	7.95	14	12.05
9	20	12	9.21	8	10.79
10	12	6	6.43	6	5.57

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
3.1459	8	0.9249

Model Information

Data Set	WORK.LOWBWT
Response Variable	low
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	189
Number of Observations Used	189

Response Profile

Ordered Value	low	Total Frequency
1	0	130
2	1	59

Probability modeled is low=1.

Class Level Information

Class	Value	Design Variables	
race	1	0	0
	2	1	0
	3	0	1

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	236.672	233.882
SC	239.914	253.332
-2 Log L	234.672	221.882

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	12.7905	5	0.0254
Score	11.7189	5	0.0388
Wald	11.0939	5	0.0495

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
lwt	1	2.3845	0.1225
race	2	1.0123	0.6028
lwt*race	2	1.3324	0.5137

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	0.7923	1.2548	0.3986	0.5278
lwt		1	-0.0151	0.00979	2.3845	0.1225
race	2	1	-0.0981	2.0259	0.0023	0.9614
race	3	1	1.9209	2.0981	0.8382	0.3599
lwt*race	2	1	0.00823	0.0145	0.3240	0.5692
lwt*race	3	1	-0.0124	0.0174	0.5063	0.4767

Association of Predicted Probabilities and Observed Responses

Percent Concordant	64.4	Somers' D	0.297
Percent Discordant	34.7	Gamma	0.300
Percent Tied	1.0	Tau-a	0.128
Pairs	7670	c	0.649

The LOGISTIC Procedure

Partition for the Hosmer and Lemeshow Test

		low = 1		low = 0	
Group	Total	Observed	Expected	Observed	Expected
1	19	2	2.29	17	16.71
2	20	6	3.83	14	16.17
3	19	5	4.35	14	14.65
4	20	3	5.27	17	14.73
5	20	7	5.78	13	14.22
6	19	6	6.14	13	12.86
7	21	6	7.54	15	13.46
8	19	6	7.89	13	11.11
9	20	11	9.41	9	10.59
10	12	7	6.51	5	5.49

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
5.2301	8	0.7327

The LOGISTIC Procedure

Model Information

Data Set	WORK.LOWBWT2
Response Variable	low
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	189
Number of Observations Used	189

Response Profile

Ordered Value	low	Total Frequency
1	0	130
2	1	59

Probability modeled is low=1.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	236.672	231.259
SC	239.914	244.226
-2 Log L	234.672	223.259

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	11.4129	3	0.0097
Score	10.7572	3	0.0131
Wald	10.1316	3	0.0175

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.8057	0.8452	0.9088	0.3404
lwt	1	-0.0152	0.00644	5.5886	0.0181
r1	1	1.0811	0.4881	4.9065	0.0268
r2	1	0.4806	0.3567	1.8156	0.1778

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
lwt	0.985	0.973	0.997
r1	2.948	1.133	7.672
r2	1.617	0.804	3.253

Association of Predicted Probabilities and Observed Responses

Percent Concordant	64.1	Somers' D	0.293
Percent Discordant	34.8	Gamma	0.296
Percent Tied	1.1	Tau-a	0.127
Pairs	7670	c	0.647

Partition for the Hosmer and Lemeshow Test

Group	Total	low = 1 Observed	low = 1 Expected	low = 0 Observed	low = 0 Expected
1	19	2	2.37	17	16.63
2	21	4	4.25	17	16.75
3	20	5	4.80	15	15.20
4	19	6	5.07	13	13.93
5	19	6	5.50	13	13.50
6	19	6	6.22	13	12.78
7	20	6	7.21	14	12.79
8	20	6	7.95	14	12.05
9	20	12	9.21	8	10.79
10	12	6	6.43	6	5.57

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
3.1459	8	0.9249