目　錄

第一章 緒論

1.1 羅吉斯迴歸模型

本篇論文是以羅吉斯迴歸模型為主要基本架構，進而延伸出本文的相關問題。在述說羅吉斯迴歸模型前，首先說明其運用時機。當反應變數(response or outcome)為離散型，且分類只有二種或少數幾類時，羅吉斯迴歸模型即為一最適用此類資料的分析方法。當然離散型的資料其分析的方法有許多種，Cox 於 1970 年根據兩個主要的理由而選擇了羅吉斯分配(logistic distribution)：(1)依數學的觀點而言，它是一種極賦彈性且容易使用的函數。(2)在生物學上的資料，它較適合解釋其意義。

我們舉一例子說明之。在測試一治療高血壓的新藥其療效時，廠商找了 100 位患病者，在未告知所服之藥為何下，一半服用此新藥，另一半則服用維他命，在服用一段期間後，測量其血壓是否超出標準值。則應變數為一二元化的離散型資料（有高血壓或沒有高血壓)，而自變數即為廠商所提供的二種藥劑。

在羅吉斯分配下，當給定一自變數(X)時，應變數(Y)的條件期望值

為

$$E[Y \mid x] = \frac{\exp(b_0 + b_1 x)}{1 + \exp(b_0 + b_1 x)} \qquad (1.1.1)$$

此即為單變量羅吉斯迴歸模型。

我們以下述的例子作說明。假設某一肝癌病患，經由某種特殊治療後，若存活則記為「1」，若死亡則為「0」，此時應變數為

$$Y = \begin{cases} 1, & \text{若為存活者} \\ 0, & \text{若為死亡者} \end{cases}$$

令$p(x)$為「存活」的機率，$p(x) = P[Y = 1 \mid x]$。因此

$$E[Y \mid x] = p(x) = \frac{\exp(b_0 + b_1 x)}{1 + \exp(b_0 + b_1 x)} \qquad (1.1.2)$$

為一單變量的羅吉斯迴歸模型。

將上述觀念推廣至多變量的複羅吉斯迴歸模型(multiple logistic regression model)：假設有 I 個獨立的伯努利隨機變數$Y = (Y_1, \cdots, Y_I)$，每一個Y_i皆為二元應變數。令$\underline{x}_i = (x_{i0}, \cdots, x_{ik})'$為第 i 個自變數向量，含有 k 個自變數，其中$x_{i0} = 1$。則

$$p(\underline{x}_i) = \frac{\exp(\sum_{j=0}^{k} b_j x_{ij})}{1 + \exp(\sum_{j=0}^{k} b_j x_{ij})} \quad , \quad i = 1, \cdots, I \qquad (1.1.3)$$

此即為複羅吉斯迴歸模型。

羅吉斯迴歸模型在統計分析的運用上已漸普遍，在二元化的離散型

資料中，以醫學方面的使用較為廣泛。

1.2 問題緣起

目前對於二元化應變數的研究，大多僅限於當應變數為伯努利分配時，求其參數估計值以及對迴歸係數作檢定。

Agresti（1990）所著的” Categorical Data Analysis ”書中曾提及，若某一固定自變數 $\underline{x}_i = (x_{i0}, \cdots, x_{ik})'$ 的觀測點（Y_i）大於一個時，則可計算出其總觀察數（n_i）與發生「成功」事件的總數量（y_i）。顯然地，$\{Y_i : i = 1, \cdots, I\}$ 為一獨立的二項隨機變數，至於當應變數為二項分配 (Binomial distribution)時，在一維或多維以上的自變數(k≧1)的研究中也僅侷限於對迴歸係數作檢定。

舉例如下：假設某家醫院婦產科 89 年針對 n_i=251 人做子宮頸癌篩檢，計有 y_i=75 人患病，病發機率為 $p(\underline{x}_i) = P[Y = y_i \mid \underline{x}_i]$

$$Y = \begin{cases} y_i & , \text{ 病發者總數} \\ n_i - y_i, & \text{ 未病發者總數} \end{cases}$$

因此我們可針對身體各狀況的檢測值即自變數 $\underline{x}_i = (x_{i0}, \cdots, x_{ik})'$，當中有哪些因子會對子宮頸癌的病發有顯著的影響，或者病發機率是否為 p_0 做進一步的探討。

本篇論文的研究重點為在自變數為兩個時，討論如何檢定機率 $p(\underline{x}_i)$

是否可為p_0的相關問題，本篇研究的方法可推廣至多個（k≧3）自變數。

有關本篇論文的相關文獻有：

1.Agresti（1990）所著的” Categorical Data Analysis ”書中，提出利用概似方程式（likelihood equations）求其最大概似估計值。作者提出：因為$\{Y_i : i=1,\cdots,I\}$為獨立的二項隨機變數，所以當省略常數係數時，其概似函數為下式

$$L(\boldsymbol{b}_0,\boldsymbol{b}_1,\boldsymbol{b}_2 \mid y_1,\cdots y_I) \approx \prod_{i=1}^{I}\left(\exp(\sum_{j=0}^{2}\boldsymbol{b}_j x_{ij})\right)^{y_i}\left(1+\exp(\sum_{j=0}^{2}\boldsymbol{b}_j x_{ij})\right)^{-n_i}$$

取對數並省略常數後其對數概似函數為

$$\begin{aligned} l &= \log L(\boldsymbol{b}_0,\boldsymbol{b}_1,\boldsymbol{b}_2 \mid y_1,\cdots y_I) \\ &\approx \sum_{j=0}^{2}\boldsymbol{b}_j(\sum_{i=1}^{I} y_i x_{ij}) - \sum_{i=1}^{I} n_i \log\left\{1+\exp(\sum_{j=0}^{2}\boldsymbol{b}_j x_{ij})\right\} \end{aligned}$$

分別對每一個b_j偏微

$$\frac{\partial l}{\partial \boldsymbol{b}_j} = \sum_{i=1}^{I} y_i x_{ij} - \sum_{i=1}^{I} n_i x_{ij}\left\{\frac{\exp(\sum_{j=0}^{2}\boldsymbol{b}_j x_{ij})}{1+\exp(\sum_{j=0}^{2}\boldsymbol{b}_j x_{ij})}\right\} \qquad j=0,1,2$$

則其概似方程式為

$$\sum_{i=1}^{I} y_i x_{ij} - \sum_{i=1}^{I} n_i x_{ij} \left\{ \frac{\exp(\sum_{j=0}^{2} b_j x_{ij})}{1+\exp(\sum_{j=0}^{2} b_j x_{ij})} \right\} = 0 \qquad j = 0,1,2$$

由於此概似方程式為非線性方程組，所以作者提出”牛頓-賴福森”法以求最大概似估計式的估計值。

2.Dobson（1990）所著的” Introduction to Generalized Linear Model”中，作者提出的應變數近似分配：

（1）因為

$$Y_i \sim Bin(n_i, p_i) \qquad E[Y_i] = n_i p_i \qquad Var(Y_i) = n_i p_i (1-p_i)$$

所以作者認為可用常態分配逼近之，意即

$$Y_i \sim N(n_i p_i, n_i p_i (1-p_i))$$

（2）假設應變數

$$Y_i = \begin{cases} 1, & \text{若某事件視為成功} \\ 0, & \text{若某事件視為失敗} \end{cases}$$

考慮其成功的比例

$$p_i = \frac{y_i}{n_i}$$

與其相對應的隨機變數設為 P_i，則

$$E[P_i]=p_i \quad , \quad Var(P_i)=\frac{p_i(1-p_i)}{n_i}$$

因此作者認為可利用其漸近分配，也就是說

$$P_i \sim N(p_i,\frac{p_i(1-p_i)}{n_i})$$

3.陳慎健（1998）碩士論文”反應變數具二項分佈之羅吉斯複迴歸模型之檢定”中，針對反應變數為二項分佈且有 2 個自變數時，求其參數估計值並以三種檢定方法檢定其迴歸係數。

1.3 本研究問題之建立

依據上述的假設可知，應變數為二項隨機變數，因此

$$P(Y_i = y_i) = \binom{n_i}{y_i}[p(\underline{x}_i)]^{y_i}[1-p(\underline{x}_i)]^{n_i-y_i} \text{ ，} \quad i = 1,\cdots,I$$

其中

$$p(\underline{x}_i) = \frac{\exp(b_0 + b_1 x_{i1} + b_2 x_{i2})}{1+\exp(b_0 + b_1 x_{i1} + b_2 x_{i2})} \qquad i = 1,\cdots,I \tag{1.3.1}$$

在本篇論文為探討兩個自變數的情況下，其概似函數可表示為

$$L = L(p(\underline{x}_i) \mid y_1,\cdots,y_I) = \prod_{i=1}^{I}\binom{n_i}{y_i}[p(\underline{x}_i)]^{y_i}[1-p(\underline{x}_i)]^{n_i-y_i} \tag{1.3.2}$$

取對數後其對數概似函數為

$$\begin{aligned} l = \log(L) &= \log\left(\prod_{i=1}^{I}\binom{n_i}{y_i}\right) + \sum_{i=1}^{I} y_i \log p(\underline{x}_i) + \sum_{i=1}^{I}(n_i - y_i)\log[1-p(\underline{x}_i)] \\ &= \log\left(\prod_{i=1}^{I}\binom{n_i}{y_i}\right) + b_0\sum_{i=1}^{I} y_i + b_1\sum_{i=1}^{I} y_i x_{i1} + b_2\sum_{i=1}^{I} y_i x_{i2} \\ &\quad - \sum_{i=1}^{I} n_i \log[1+\exp(b_0 + b_1 x_{i1} + b_2 x_{i2})] \end{aligned} \tag{1.3.3}$$

吾人所關心的假設檢定為

$$H_0 : p(\underline{x}) = p_0 \text{ 對 } H_a : p(\underline{x}) \neq p_0 \tag{1.3.4}$$

其中p_0為已知常數，$0 < p_0 < 1$。對p_0做 logit，則可令

$$q_0 = \log\left(\frac{p_0}{1-p_0}\right)$$

因此（1.3.4）式的檢定可轉換成

$$H_0 : b_0 + b_1 x_{i1} + b_2 x_{i2} = q_0 \quad \text{對} \quad H_a : b_0 + b_1 x_{i1} + b_2 x_{i2} \neq q_0 \qquad \textbf{(1.3.5)}$$

我們可進一步的令

$$q = b_0 + b_1 x_{i1} + b_2 x_{i2} - q_0$$

則（1.3.5）式又可轉換成

$$H_0 : q = 0 \quad \text{對} \quad H_a : q \neq 0 \qquad \textbf{(1.3.6)}$$

其中（1.3.4）式、（1.3.5）式及（1.3.6）式均為等價的檢定假設。

1.4 研究方法與結果

針對本研究之假設檢定問題，本文提出三種檢定方法：

(1) 概度比檢定式；

(2) 一致最強力不偏檢定；

(3) 華德檢定(Wald’s Test)。

在第二章中第一節我們介紹概度比檢定式，利用延伸自牛頓-賴福森法的反覆加權最小平方法，以求其近似的最大概似估計值。

我們在第三章第一節中敘述何謂一致最強力不偏檢定，第二節將導出本研究之假設檢定的一致最強力不偏檢定式，最後在第三節介紹由資料計算檢定式的方法。

在第四章第一節，我們將介紹如何將模型轉換成可利用加權最小平方估計參數的模式，第二節介紹如何利用加權最小平方估計參數及其檢定式。

在第五章中我們將提出一個實例，介紹如何運用上述三種檢定方法及討論其結果。

第二章　概度比檢定式

在本章中，我們針對$H_0 : p(\underline{x}) = p_0$對$H_a : p(\underline{x}) \neq p_0$的假設檢定，導出概度比檢定式。

當$H_a : p(\underline{x}) \neq p_0$時，其概似函數為（1.3.2）式及對數概似函數為（1.3.3）式，為了使對數概似函數產生最大值，我們分別對b_0，b_1，b_2微分，可得

$$\left.\begin{aligned}
\frac{\partial l}{\partial b_0} &= \sum_{i=1}^{I} y_i - \sum_{i=1}^{I} \frac{n_i \exp(b_0 + b_1 x_{i1} + b_2 x_{i2})}{1 + \exp(b_0 + b_1 x_{i1} + b_2 x_{i2})} \\
\frac{\partial l}{\partial b_1} &= \sum_{i=1}^{I} y_i x_{i1} - \sum_{i=1}^{I} \frac{n_i x_{i1} \exp(b_0 + b_1 x_{i1} + b_2 x_{i2})}{1 + \exp(b_0 + b_1 x_{i1} + b_2 x_{i2})} \\
\frac{\partial l}{\partial b_2} &= \sum_{i=1}^{I} y_i x_{i2} - \sum_{i=1}^{I} \frac{n_i x_{i2} \exp(b_0 + b_1 x_{i1} + b_2 x_{i2})}{1 + \exp(b_0 + b_1 x_{i1} + b_2 x_{i2})}
\end{aligned}\right\} \qquad (2.1.1)$$

令（2.1.1）式為零可得

$$\left.\begin{aligned}
\sum_{i=1}^{I} \frac{n_i \exp(b_0 + b_1 x_{i1} + b_2 x_{i2})}{1 + \exp(b_0 + b_1 x_{i1} + b_2 x_{i2})} &= \sum_{i=1}^{I} y_i \\
\sum_{i=1}^{I} \frac{n_i x_{i1} \exp(b_0 + b_1 x_{i1} + b_2 x_{i2})}{1 + \exp(b_0 + b_1 x_{i1} + b_2 x_{i2})} &= \sum_{i=1}^{I} y_i x_{i1} \\
\sum_{i=1}^{I} \frac{n_i x_{i2} \exp(b_0 + b_1 x_{i1} + b_2 x_{i2})}{1 + \exp(b_0 + b_1 x_{i1} + b_2 x_{i2})} &= \sum_{i=1}^{I} y_i x_{i2}
\end{aligned}\right\} \qquad (2.1.2)$$

即為概似方程式。令$\hat{b}_0$，$\hat{b}_1$，$\hat{b}_2$為（2.1.2）式的解，亦即（$\hat{b}_0, \hat{b}_1, \hat{b}_2$）為（$b_0, b_1, b_2$）的最大概似估計式。

在$H_0 : p(\underline{x}) = p_0$成立下，即$H_0 : b_0 + b_1 x_{i1} + b_2 x_{i2} = q_0$，其對數概似方

程式為

$$\log L(b_1, b_2 \mid y_1, \cdots, y_I)$$

$$= \log\left\{\prod_{i=1}^{I}\binom{n_i}{y_i}\right\} - \sum_{i=1}^{I} n_i \log\{1+\exp(q_0 + b_1(x_{i1} - x_1) + b_2(x_{i2} - x_2))\}$$

$$+q_0\sum_{i=1}^{I} y_i + b_1\sum_{i=1}^{I}(x_{i1} - x_1)y_i + b_2\sum_{i=1}^{I}(x_{i2} - x_2)y_i \qquad \textbf{(2.1.3)}$$

相同的，我們分別對 b_1， b_2微分，使對數概似函數產生最大值，並可得到下列二個等式：

$$\left.\begin{aligned} \sum_{i=1}^{I}\frac{n_i(x_{i1}-x_1)\exp(q_0+b_1(x_{i1}-x_1)+b_2(x_{i2}-x_2))}{1+\exp(q_0+b_1(x_{i1}-x_1)+b_2(x_{i2}-x_2))} &= \sum_{i=1}^{I}(x_{i1}-x_1)y_i \\ \sum_{i=1}^{I}\frac{n_i(x_{i2}-x_2)\exp(q_0+b_1(x_{i1}-x_1)+b_2(x_{i2}-x_2))}{1+\exp(q_0+b_1(x_{i1}-x_1)+b_2(x_{i2}-x_2))} &= \sum_{i=1}^{I}(x_{i2}-x_2)y_i \end{aligned}\right\} \qquad \textbf{(2.1.4)}$$

令 $\tilde{b}_1$， $\tilde{b}_2$為（2.1.4）式的解，亦即（ $\tilde{b}_1$, $\tilde{b}_2$）為（ b_1, b_2）的最大概似估計式。

在 $H_0 : b_0 + b_1x_1 + b_2x_2 = q_0$ 對 $H_a : b_0 + b_1x_1 + b_2x_2 \neq q_0$ 的假設下，概度比檢定式為

$$l = \frac{L(\tilde{p}(\underline{x}_i) \mid y_1, \ldots, y_I)}{L(\hat{p}(\underline{x}_i) \mid y_1, \ldots, y_I)} = \frac{L(\tilde{b}_1, \tilde{b}_2)}{L(\hat{b}_0, \hat{b}_1, \hat{b}_2)}$$

當給定一顯著水準 a ，若 l 小於某一臨界點時則拒絕

$H_0 : b_0 + b_1 x_1 + b_2 x_2 = q_0$的假設，即當

$$
\begin{aligned}
D &= -2\log(l) = -2[\log(\tilde{b}_1, \tilde{b}_2) - \log(\hat{b}_0, \hat{b}_1, \hat{b}_2)] \\
&= 2[\log L(\hat{b}_0, \hat{b}_1, \hat{b}_2) - \log L(\tilde{b}_1, \tilde{b}_2)] \\
&> c^2_{1;a} \qquad (2.1.5)
\end{aligned}
$$

則拒絕 H_0 的假設，其中 $c^2_{1;a}$ 表示自由度為 1 的卡方分配在右尾 100a % 臨界點。

在實際計算過程中，我們可利用統計軟體 SAS 中的 Logistic Procedure，但須注意的是 SAS 對反應變數機率的估計乃針對所輸入的資料定義為"0"的組別做估計，以簡化計算，因此針對有興趣的反應變數可先轉換成"0"的類別，接下來將反應變數Y_i轉換成$Y_i^* = (1-p_0)n_i Y_i / [Y_i + (n_i - 2Y_i)p_0]$，再把模式的反應變數在 SAS 程式中寫成$Y_i^* / n_i$，即可針對我們有興趣的自變數，以 SAS 在此分析內鍵由牛頓賴福森法演變而來的反覆加權最小平方法，求出 (2.1.2) 及 (2.1.4) 式裡的參數，再代回 (1.3.3) 及 (2.1.3) 式分別求出在 $H_0 : b_0 + b_1 x_1 + b_2 x_2 = q_0$ 與 $H_a : b_0 + b_1 x_1 + b_2 x_2 \neq q_0$ 的假設下，個別的對數概似函數值，最後得到（2.1.5）式之概度比檢定式 D 值。

第三章　一致最強力不偏檢定

3.1 一致最強力不偏檢定之定義

當我們考量單尾假設檢定時，關於（1）$H_0:\theta \le \theta_0$對$H_a:\theta > \theta_0$及（2）$H_0:\theta \le \theta_1$或$H_0:\theta \ge \theta_2$對$H_a:\theta_1 < \theta < \theta_2$的檢定問題，通常可找到其一致最強力檢定（uniformly most powerful test）。但是對於檢定問題屬於$H_0:\theta_1 \le \theta \le \theta_2$對$H_a:\theta < \theta_1$或$H_a:\theta > \theta_2$的假設，一般都無法找到一致最強力檢定。然而，它卻存在一個一致最強力不偏檢定（uniformly most powerful unbiased test）。因此，本章的主要目的在於針對本研究之假設檢定問題，探討它的一致最強力不偏檢定。我們首先定義不偏檢定（unbiased test）：

定義 3.1.1

考慮 $H_0:\theta \in \Omega_{H_0}$ 對 $H_a:\theta \in \Omega_{H_a}$ ，在 $\Omega_{H_0} \cap \Omega_{H_a} = \phi$ 的檢定問題。對顯著水準為α的檢定φ，若其檢定力函數（power function）$\beta_\varphi(\theta) = E[\varphi(X)]$滿足

$$\beta_\varphi(\theta) \le \alpha \text{，若} \theta \in \Omega_{H_0}$$

$$\beta_\varphi(\theta) \ge \alpha \text{，若} \theta \in \Omega_{H_a}$$

則稱檢定φ為不偏檢定。

根據定義 3.1.1，我們可以定義一致最強力不偏檢定為：

定義 3.1.2

若 f 為顯著水準 a 下的不偏檢定，且對任何其它同為顯著水準 a 的不偏檢定 f^* 都滿足

$$E_q[f^*(X)] \le E_q[f(q)]\text{，}\ \forall q \in W_{H_a}$$

則稱檢定 f 為一致最強力不偏檢定。

參考 Lehmann（1986）的書中提及：一個多維度參數的指數族中，若只考慮一個單維度的參數，其他參數視為干擾參數，則會存在一個一致最強力不偏檢定。針對本研究的問題： $H_0:\theta \ne \theta_0$ **對** $H_a:\theta \ne \theta_0$ **。若** $\underline{x}=(x_1,\ldots,x_I)$ **的分佈為**

$$P_{\underline{\theta}}(x)=c(\theta,\delta)\exp[\theta U(x)+\sum_{i=1}^{I-1}\delta_i T_i(x)]h(x)$$

其中 $\delta=(\delta_1,\ldots,\delta_{I-1})$ **，** $T=(T_1,\ldots,T_{I-1})$ **，很顯然的，（U ，T）為（q，d）之充分統計量，其聯合機率密度函數為**

$$P_{\underline{q}}(u,t)=c(q,d)\exp[qu+\sum_{i=1}^{I-1}d_i t_i]h(u,t)$$

因此，在我們給定 T=t 時，U 為僅剩的變數，則其機率密度函數為

$$P_{\underline{\theta}}(u\mid t)=c(\theta)\exp(\theta u)h(u)$$

關於我們的假設檢定： $H_0:q=q_0$ **對** $H_a:q\ne q_0$ **，可求其一致最強力不偏**

檢定：

$$f(u)=\begin{cases}1 & \text{若}\, u<c_1(t)\,\text{或}\, u>c_2(t)\\ g_i & \text{若}\, u=c_i(t),\quad i=1,2\\ 0 & \text{若}\, c_1(t)<u<c_2(t)\end{cases}$$

其中 $c_1(t)$**、** $c_2(t)$ **和** g_1**、** g_2 **滿足**

$$E_{q_0}[f(U,T)\,|\,t]=a$$

$$E_{q_0}[Uf(U,T)\,|\,t]=aE_{q_0}[U\,|\,t]$$

3.2 一致最強力不偏檢定式

在本研究的問題中，當我們給定任一獨立變數為(x_1 **,** x_2 **)的水準下，其反應變數的值出現 1 之機率為** p_0 **時，令**

$$\log it(p_0)=q_0$$

$$q=b_0+b_1x_1+b_2x_2-q_0$$

則 $b_0=q_0+q-b_1x_1-b_2x_2$ **，反應變數(** Y_1 **, . . . ,** Y_I **)的聯合機率密度函數可寫成**

$$P(Y_1=y_1,\ldots,Y_I=y_I\,|\,\underline{n},\underline{x}_1,\underline{x}_2;q,b_1,b_2)$$

$$=\prod_{i=1}^{I}\binom{n_i}{y_i}\prod_{i=1}^{I}\left\{1+\exp\left((q_0+q)+b_1(x_{i1}-x_1)+b_2(x_{i2}-x_2)\right)\right\}^{-n_i}$$

$$*\exp\left\{q\sum_{i=1}^{I}y_i+q_0\sum_{i=1}^{I}y_i+b_1\sum_{i=1}^{I}(x_{i1}-x_1)y_i+b_2\sum_{i=1}^{I}(x_{i2}-x_2)y_i\right\} \qquad (3.2.1)$$

根據 Neyman-Fisher 分解定理，從聯合機率密度函數（3.2.1）式可得知：

$$T_0=\sum_{i=1}^{I}y_i\ ，\ T_1=\sum_{i=1}^{I}y_i(x_{i1}-x_1)\ ，\ T_2=\sum_{i=1}^{I}y_i(x_{i2}-x_2)$$

為(q**,**b_1**,**b_2**)的聯合充分統計量。因此，依據前一節的結果，關於本研究的假設檢定問題，我們討論如下：**

當假設檢定：$H_0:q=0$**對**$H_a:q\neq 0$**時，給定**$T_0=t_0$**，**$T_1=t_1$**及**$T_2=t_2$**，在顯著水準為**a**下，則一致最強力不偏檢定式為：**

$$y(t_0)=\begin{cases}1 & \text{若}\ t_0<c_1(t_1,t_2)\ \text{或}\ t_0>c_2(t_1,t_2)\\ g_i & \text{若}\ t_0=c_i(t_1,t_2),\quad i=1,2\\ 0 & \text{若}\ c_1(t_1,t_2)<t_0<c_2(t_1,t_2)\end{cases} \qquad (3.2.2)$$

其中$c_1(t_1,t_2)$**、**$c_2(t_1,t_2)$**及**g_1**、**g_2**滿足以下二個條件：**

(a) $P(T_0<c_1(t_1,t_2)\mid T_1=t_1,T_2=t_2;H_0)$

$+g_1P(T_0=c_1(t_1,t_2)\mid T_1=t_1,T_2=t_2;H_0)$

$+g_2P(T_0=c_2(t_1,t_2)\mid T_1=t_1,T_2=t_2;H_0)$

$+P(T_0>c_2(t_1,t_2)\mid T_1=t_1,T_2=t_2;H_0)=a$ **(3.2.3)**

(b) $\sum_{c_1(t_1,t_2)}^{c_2(t_1,t_2)}t_0P(T_0=t_0\mid T_1=t_1,T_2=t_2;H_0)$

$$= E[T_0 \mid T_1 = t_1, T_2 = t_2; H_0] * (1-\alpha) \qquad (3.2.4)$$

3.3 計算方法

計算（3.2.3）式子中$c_1(t_1,t_2)$和$c_2(t_1,t_2)$時，我們知道當給定充分統計量$T_1 = t_1$及$T_2 = t_2$且在$H_0: q = 0$成立之下，T_0的條件機率密度函數與參數（b_1，b_2）獨立，因此，（3.2.1）式其聯合機率密度函數可寫為

$$P(Y_1 = y_1,\ldots,Y_I = y_I \mid \underline{n}, \underline{x}_1, \underline{x}_2; q = b_1 = b_2 = 0)$$

$$= \prod_{i=1}^{I} \binom{n_i}{y_i} \left(1+e^{q_0}\right)^{-n_i} \exp(q_0 \sum_{i=1}^{I} y_i) \qquad (3.3.1)$$

當$H_0: q = 0$時，（T_0，T_1，T_2）及（T_1，T_2）的條件機率分別為

$$P(T_0 = t_0, T_1 = t_1, T_2 = t_2 \mid H_0) = e^{q_0 t_0}\left(1+e^{q_0}\right)^{-\sum n_i} * \sum_{A_1}\left\{\left[\prod_{i=1}^{I}\binom{n_i}{y_i}\right]\right\}$$

$$P(T_1 = t_1, T_2 = t_2 \mid H_0) = \left(1+e^{q_0}\right)^{-\sum n_i} * \sum_{A_2}\left\{\left[\prod_{i=1}^{I}\binom{n_i}{y_i}\right]e^{q_0 \sum y_i}\right\}$$

其中

$$A_1 = \left\{ \underline{y} : \sum_{i=1}^{I} y_i = t_0, \sum_{i=1}^{I} (x_{i1} - x_1) y_i = t_1, \sum_{i=1}^{I} (x_{i2} - x_2) y_i = t_2 \right\} \quad (3.3.2)$$

$$A_2 = \left\{ \underline{y} : \sum_{i=1}^{I} (x_{i1} - x_1) y_i = t_1, \sum_{i=1}^{I} (x_{i2} - x_2) y_i = t_2 \right\} \quad (3.3.3)$$

所以當給定 $T_1 = t_1$ 及 $T_2 = t_2$ 且在 $H_0 : q = 0$ 成立之下， T_0 的條件機率密度函數為

$$P(T_0 = t_0 \mid T_1 = t_1, T_2 = t_2; H_0) = \frac{e^{q_0 t_0} \sum_{A_1} \left\{ \prod_{i=1}^{I} \binom{n_i}{y_i} \right\}}{\sum_{A_2} \left\{ \left[\prod_{i=1}^{I} \binom{n_i}{y_i} \right] e^{q_0 \sum y_i} \right\}} \quad (3.3.4)$$

而在實際的計算過程中，可以分為以下四個步驟：

（1）由觀察值計算出 $T_0 = t_0 = \sum_{i=1}^{I} y_i$ **，** $T_1 = t_1 = \sum_{i=1}^{I} y_i (x_{i1} - x_1)$ **，** $T_2 = t_2 = \sum_{i=1}^{I} y_i (x_{i2} - x_2)$ **。**

（2）計算出滿足 A_2 集合中所有 $\underline{y}$ 的可能組合，求得（3.3.4）式中的分母。

（3）由 A_2 的集合計算出所有滿足 A_1 集合中 $\underline{y}$ 的可能組合，則可求得（3.3.4）式中的分子。

（4）給定一顯著水準 a ，則 $c_1(t_1, t_2)$ 和 $c_2(t_1, t_2)$ 的選取分別由 T_0 的最小

值及最大值開始，計算其條件機率與累積條件機率，再漸增減 T_0 值。當 $c_1(t_1,t_2)=k_1$ 和 $c_2(t_1,t_2)=k_2$ 時，若累積條件機率等於 a，則 $c_1(t_1,t_2)$ 和 $c_2(t_1,t_2)$ 的值即為 k_1、k_2，此時 $g_1=g_2=1$。若當 $T_0=k_1$ 或 k_2 時，其累積條件機率大於 a 且當 $T_0=k_1-1$ 或 k_2+1 時，其累積條件機率小於 a，則 $c_1(t_1,t_2)$ 和 $c_2(t_1,t_2)$ 的值即為（k_1，k_2+1）、（k_1-1，k_2）或（k_1-1，k_2+1）。待數組累積條件機率小於 a 的解求出後，最後選擇累積條件機率最接近 a 且符合（3.2.4）式的解。至於 g_1 及 g_2 的值可由滿足（3.2.3）及（3.2.4）的等式解出。其中（3.2.4）式為

$$E[T_0 \mid T_1=t_1, T_2=t_2; H_0] = \frac{\sum_{A_1}\left\{\sum_{i=1}^{I} y_i * e^{q_0 t_0}\prod_{i=1}^{I}\binom{n_i}{y_i}\right\}}{\sum_{A_2}\left\{\left[\prod_{i=1}^{I}\binom{n_i}{y_i}\right] * e^{q_0\sum y_i}\right\}}$$

其中 A_1，A_2 分別為滿足（3.3.2）及（3.3.3）式之集合。

第 四 章　華德檢定

4.1 模型轉換

本節主要在於對羅吉斯模式進行模型轉換。首先，令

$$p_i = \frac{y_i}{n_i}$$

為第 i 組樣本「成功」的比例且令 $q_i = 1 - p_i$，及定義 P_i 和 Q_i 分別為 p_i 與 q_i 相對應之隨機變數。由（1.3.1）式

$$\log it(\pi_i) = \ln \frac{\pi_i}{1-\pi_i} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} \quad , \quad i = 1,2,\ldots,I$$

得到想法，我們可令

$$w_i = \ln \frac{p_i}{q_i} \text{，} \quad i = 1,2,\ldots,I$$

且 W_i 為 w_i 相對應之隨機變數。則在大樣本下，我們可預期 W_i 將滿足以下二項結果：

（1） $E[W_i] \approx \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$

（2） $Var(W_i) \approx \dfrac{1}{n_i \pi_i (1-\pi_i)}$

經由上述（1）、（2）之模型轉換後，我們即可利用加權最小平方法來估

計參數 b_0、 b_1和 b_2，並針對本研究之假設檢定討論之。

4.2 加權最小平方法

本節將介紹如何利用加權最小平方法去估計參數。根據加權最小平方法之理論，合適的加權數為W_i的變異數之倒數，亦即$n_i p_i(1-p_i)$，但p_i為未知，因而我們考慮利用樣本比例 p_i去估計p_i，得到W_i的近似變異數

$$S^2\{W_i\} = \frac{1}{n_i p_i(1-p_i)}$$

因此，在加權最小平方法計算過程中所用的權數為估計值$n_i p_i(1-p_i)$。令

$$U = \sum_{i=1}^{I} n_i p_i(1-p_i)\{w_i - b_0 - b_1 x_{i1} - b_2 x_{i2}\}^2$$

為使U 產生最小值，我們分別對 b_0、 b_1和 b_2作偏微分

$$\frac{\partial U}{\partial b_j} = \frac{\partial}{\partial b_j}\sum_{i=1}^{I} n_i p_i(1-p_i)\{w_i - b_0 - b_1 x_{i1} - b_2 x_{i2}\}^2 \text{，} \quad j = 0,1,2 \qquad \textbf{(4.2.1)}$$

令（4.2.1）式為零並移項而得到下列正規方程式：

$$\left.\begin{aligned}
&\boldsymbol{b}_0\sum_{i=1}^{I} n_i p_i q_i + \boldsymbol{b}_1\sum_{i=1}^{I} x_{i1} n_i p_i q_i + \boldsymbol{b}_2\sum_{i=1}^{I} x_{i2} n_i p_i q_i = \sum_{i=1}^{I} w_i n_i p_i q_i \\
&\boldsymbol{b}_0\sum_{i=1}^{I} x_{i1} n_i p_i q_i + \boldsymbol{b}_1\sum_{i=1}^{I} x_{i1}{}^2 n_i p_i q_i + \boldsymbol{b}_2\sum_{i=1}^{I} x_{i1} x_{i2} n_i p_i q_i = \sum_{i=1}^{I} w_i x_{i1} n_i p_i q_i \\
&\boldsymbol{b}_0\sum_{i=1}^{I} x_{i2} n_i p_i q_i + \boldsymbol{b}_1\sum_{i=1}^{I} x_{i1} x_{i2} n_i p_i q_i + \boldsymbol{b}_2\sum_{i=1}^{I} x_{i2}{}^2 n_i p_i q_i = \sum_{i=1}^{I} w_i x_{i2} n_i p_i q_i
\end{aligned}\right\} \qquad (4.2.2)$$

為解（4.2.2）式，首先令 $\hat{b}_0$、$\hat{b}_1$ 和 $\hat{b}_2$ 為（4.2.2）式的解，並將（4.2.2）式改寫成

$$A\underline{\hat{b}} = \underline{Z} \qquad (4.2.3)$$

其中

$$A=\begin{bmatrix}
\sum_{i=1}^{I} n_i p_i q_i & \sum_{i=1}^{I} x_{i1} n_i p_i q_i & \sum_{i=1}^{I} x_{i2} n_i p_i q_i \\
\sum_{i=1}^{I} x_{i1} n_i p_i q_i & \sum_{i=1}^{I} x_{i1}{}^2 n_i p_i q_i & \sum_{i=1}^{I} x_{i1} x_{i2} n_i p_i q_i \\
\sum_{i=1}^{I} x_{i2} n_i p_i q_i & \sum_{i=1}^{I} x_{i1} x_{i2} n_i p_i q_i & \sum_{i=1}^{I} x_{i2}{}^2 n_i p_i q_i
\end{bmatrix}$$

$$\underline{\hat{b}} = \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} \quad , \qquad \underline{Z} = \begin{bmatrix} \sum_{i=1}^{I} w_i n_i p_i q_i \\ \sum_{i=1}^{I} w_i x_{i1} n_i p_i q_i \\ \sum_{i=1}^{I} w_i x_{i2} n_i p_i q_i \end{bmatrix}$$

由（4.2.3）式可得 $\underline{b}$ 的加權最小平方估計式

$$\underline{\hat{b}} = A^{-1}\underline{Z}$$

且在大樣本下

$$\underline{\hat{b}} \sim N\left(\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} \quad , \quad \sum(b)\right)$$

其中

$$\sum(b) = B(\underline{b})^{-1}\Lambda(\underline{b})B(\underline{b})^{-1}$$

$$B(\underline{b}) = \begin{bmatrix} \sum p_i(1-p_i) & \sum x_{i1}p_i(1-p_i) & \sum x_{i2}p_i(1-p_i) \\ \sum x_{i1}p_i(1-p_i) & \sum x_{i1}{}^2 p_i(1-p_i) & \sum x_{i1}x_{i2}p_i(1-p_i) \\ \sum x_{i2}p_i(1-p_i) & \sum x_{i2}p_i(1-p_i) & \sum x_{i2}{}^2 p_i(1-p_i) \end{bmatrix}$$

$$L(\underline{b}) = \begin{bmatrix} \sum \frac{p_i(1-p_i)}{n_i} & \sum \frac{x_{i1}p_i(1-p_i)}{n_i} & \sum \frac{x_{i2}p_i(1-p_i)}{n_i} \\ \sum \frac{x_{i1}p_i(1-p_i)}{n_i} & \sum \frac{x_{i1}{}^2 p_i(1-p_i)}{n_i} & \sum \frac{x_{i1}x_{i2}p_i(1-p_i)}{n_i} \\ \sum \frac{x_{i2}p_i(1-p_i)}{n_i} & \sum \frac{x_{i1}x_{i2}p_i(1-p_i)}{n_i} & \sum \frac{x_{i2}{}^2 p_i(1-p_i)}{n_i} \end{bmatrix}$$

令

$$\hat{\Sigma}(\underline{b}) = \hat{B}^{-1}(b)\hat{L}(b)\hat{B}^{-1}(b)$$

其中

$$\hat{B}(\underline{b}) = B(b)\big|(\underline{p} = \underline{p})$$

$$\hat{\Lambda}(\underline{b}) = \Lambda(b)\big|(\underline{p} = \underline{p})$$

$$\hat{\Sigma}(\underline{b}) = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix}$$

則對本研究的假設$H_0 : p(\underline{x}) = p_0$**與**$H_a : p(\underline{x}) \neq p_0$**做檢定時，我們可利用**

華德檢定的概念，把欲檢定的參數函數除以其標準差得到下列統計量來檢定之

$$Q=\frac{\left|\hat{b}_0+\hat{b}_1x_1+\hat{b}_2x_2-q_0\right|}{\sqrt{v_{11}+x_1^{\ 2}v_{22}+x_2^{\ 2}v_{33}+2x_1v_{12}+2x_2v_{13}+2x_1x_2v_{23}}}$$

其中 $q_0=\ln\frac{p_0}{1-p_0}$ **，由於** $Q\sim N(0,1)$ **，則當給定一顯著水準** a **時，若** $Q>z_{a/2}$ **即拒絕** $H_0:p(\underline{x})=p_0$ **之假設。其中** $z_{a/2}$ **為標準常態分配下，右尾機率為** $a/2$ **的臨界點。**

第五章　實例

本實例取自 Hosmer 和 Lemeshow（1989）”Applied Logistic Regression”一書。本實例之資料為 1989 年美國某家醫院的內科部所做的調查，其主要目的為針對體重過輕（小於 2500 公克）的新生嬰兒，研究其造成的因素。調查對象為 189 名婦女($\sum n_i = 189$)，其中 59 名婦女所生的嬰兒屬於體位過輕($\sum y_i = 59$)，130 名婦女所生的嬰兒體重正常；由於資料型態的關係，所以本研究僅討論三種可能的自變數變因，針對二項自變數下，我們的假設是：婦女所生的嬰兒屬於體位過輕的機率為，：婦女所生的嬰兒屬於體位過輕的機率不為”：

（1）Race（Race：1=White　2=Black　3=Other）；

（2）Smoking Status During Period（Smoke：1=Yes　0=No）；

（3）Presence of Uterine Irritability（Ui：1=Yes　0=No）。

我們將原始資料整理後列於附錄中表 A1，因為 Race 有三個類別，所以我們可用二個虛擬變數（Race1,Race2）分別以（0,0）、（1,0）、（0,1）的代號來劃分該三類變項。對於概度比檢定，利用反覆加權最小平方法求出的參數，並非本研究探討的目的，因此只列出概度比檢定式的 D 值及 P 值於表 5.1.1~3。對於一致最強力不偏檢定，我們將檢定法則的臨界值及□值列於表 5.2.1~3。對於大樣本檢定，我們將檢定統計量 Q 值和 P 值列於表 5.3.1~3。由表 5.1.1 至表 5.3.3，我們可清楚的看到：針對本研究的假設檢定所用的三種檢定方法，其結果完全一致。

表 5.1.1 對 之概度比檢定結果( )

race1	race2	smoke			D 值	P-value
0	0	0	0.137	0.3 0.2 0.1 0.05	9.3311 1.8174 1.1206 9.9453	0.0022＊ 0.1776 0.2898 0.0016＊
0	0	1	0.326	0.5 0.4 0.3 0.2	7.5638 1.5236 0.3950 5.5991	0.0060＊ 0.2171 0.5297 0.0179＊
1	0	0	0.319	0.5 0.4 0.3 0.2	3.0917 0.7248 0.2717 2.3318	0.0787 0.3946 0.6022 0.1268
1	0	1	0.589	0.8 0.7 0.6 0.5	4.9335 1.1995 0.0427 0.5969	0.0263＊ 0.2734 0.8363 0.4398
0	1	0	0.325	0.5 0.4 0.3 0.2	7.8666 1.6071 0.3883 5.7306	0.0050＊ 0.2049 0.5332 0.0167＊
0	1	1	0.595	0.7 0.6 0.5 0.45	1.3588 0.0122 0.9617 2.2471	0.2438 0.9122 0.3268 0.1339

註：「＊」表示小於顯著水準 0.05。

表5.1.2 對 之概度比檢定結果( )

race1 race2 ui