บทที่ 2

ระเบียบวิธีที่ใช้ในการวิจัยและผลงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงรายละเอียดของสมการถดถอยโลจิสติค วิธีประมาณค่าพารามิเตอร์ด้วยภาวะน่าจะเป็นสูงสุด ฟังก์ชันจำแนกประเภท และวิธีกำลังสองน้อยสุดถ่วงน้ำหนัก

ข้อสมมุติทั่วไปของสมการถดถอยโลจิสติค

พิจารณาข้อมูลที่อยู่ในรูป $\{(y_i, \mathbf{x}_i)\}$; $i = 1,2,\ldots,N$ ตัวแปรตาม y_i มีการแจกแจงทวินาม พารามิเตอร์ 1 และ p_i $\mathbf{x}_i$ แทน $x_{1i}, x_{2i}, \ldots, x_{mi}$ $\mathbf{x}_i$ เป็นตัวแปรอธิบายร่วมกัน m ตัว (m-covariate explanatory variable)

$$y_i = \begin{cases} 1 & \text{ด้วยความน่าจะเป็น } p_i \\ 0 & \text{ด้วยความน่าจะเป็น } 1-p_i \end{cases}$$

ให้ $\mathrm{logit}(p_i) = \ln\left(\dfrac{p_i}{1-p_i}\right)$ เป็นฟังก์ชันการแปลงโลจิสติคของความน่าจะเป็น p ซึ่งจะแปลงค่า p จากช่วง $(0,1)$ เป็นค่าของ $\mathrm{logit}(p)$ ในช่วง $(-\infty,\infty)$ กราฟของฟังก์ชัน $\mathrm{logit}(p)$ เป็นโค้งซิกมอยด์ และจะเป็นเส้นตรงเมื่อ p อยู่ระหว่าง 0.2 ถึง 0.8
ให้สมการโลจิสติคเชิงเส้น (linear logistic model) สำหรับตัวแปรตาม p_i อันเนื่องด้วยตัวแปรอธิบาย m ตัว ณ ระดับ i คือ

$$\mathrm{logit}(p_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_m x_{mi}$$

จากฟังก์ชันการแปลงและสมการโลจิสติค จะได้ว่า

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_m x_{mi}$$

$$\left(\frac{p_i}{1-p_i}\right) = e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_m x_{mi}}$$

$$p_i = \frac{e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_m x_{mi}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_m x_{mi}}}$$

p_i คือความน่าจะเป็นแบบมีเงื่อนไขของ $y = 1$ เมื่อกำหนด $\mathbf{x}$ ณ ระดับที่ i *O'hara*, *Hosmer*, *Lemeshow* และ *Hartz*(1982) กล่าวว่าเพื่อให้เหมาะสมตามทฤษฎี รูปแบบจริงของ p_i คือ $P(y = 1 \backslash \mathbf{x_i})$ ควรเขียนแทนด้วย $\pi(\mathbf{x_i})$ เพราะว่าเป็นฟังก์ชันของ $\mathbf{x_i}$

ดังนั้นฟังก์ชันโลจิสติค คือ $\pi(\mathbf{x_i}) = \dfrac{e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_m x_{mi}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_m x_{mi}}}$

ดังนั้นสมการถดถอยโลจิสติคเชิงเส้น คือ

$$\text{logit}(\pi(\mathbf{x_i})) = \ln\left(\frac{\pi(\mathbf{x_i})}{1-\pi(\mathbf{x_i})}\right)$$

$$= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_m x_{mi} \qquad \ldots(1)$$

สามารถเขียนสมการ (1) ในรูปของเมทริกซ์ ดังนี้

$$\mathbf{G(X)} = \mathbf{XB}$$

เมื่อ $\mathbf{G(X)} = \begin{bmatrix} \text{logit}(\pi(\mathbf{x}_1)) \\ \text{logit}(\pi(\mathbf{x}_2)) \\ \vdots \\ \text{logit}(\pi(\mathbf{x}_N)) \end{bmatrix}_{\text{Nx1}}$; $\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{m1} \\ 1 & x_{12} & x_{22} & \cdots & x_{m2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1i} & x_{2i} & \cdots & x_{mi} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1N} & x_{2N} & \cdots & x_{mN} \end{bmatrix}_{\text{Nx(m+1)}}$

และ $\mathbf{B} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}_{(m+1)\times 1}$

ลอการึทึมธรรมชาติของฟังก์ชันภาวะน่าจะเป็นสูงสุดในสมการถดถอยโลจิสติค (Log Likelihood F unction)

ฟังก์ชันภาวะน่าจะเป็นของ $\pi(\mathbf{x_i})$, $\mathbf{x_i}$ แทน $x_{1i}, x_{2i}, \ldots, x_{mi}$ คือ

$$l(\mathbf{B}) = \prod_{i=1}^{N} (\pi(\mathbf{x_i}))^{y_i} (1-\pi(\mathbf{x_i}))^{1-y_i}$$

ลอการึทึมธรรมชาติของฟังก์ชันภาวะน่าจะเป็น คือ

$$\begin{aligned}
L(\mathbf{B}) &= \ln(l(\mathbf{B})) \\
&= \sum_{i=1}^{N} \left\{ y_i \ln(\pi(\mathbf{x_i})) + (1-y_i)\ln(1-\pi(\mathbf{x_i})) \right\} \\
&= \sum_{i=1}^{N} \left\{ y_i \ln(\pi(\mathbf{x_i})) + \ln(1-\pi(\mathbf{x_i})) - y_i \ln(1-\pi(\mathbf{x_i})) \right\} \\
&= \sum_{i=1}^{N} \left\{ y_i \left\{ \ln(\pi(\mathbf{x_i})) - \ln(1-\pi(\mathbf{x_i})) \right\} + \ln(1-\pi(\mathbf{x_i})) \right\} \\
&= \sum_{i=1}^{N} \left\{ y_i \ln\left(\frac{\pi(\mathbf{x_i})}{1-\pi(\mathbf{x_i})} \right) + \ln(1-\pi(\mathbf{x_i})) \right\} \\
&= \sum_{i=1}^{N} \left\{ y_i \ln\left(\frac{e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_m x_{im}}}{1+e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_m x_{im}}} \frac{1}{1-\frac{e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_m x_{im}}}{1+e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_m x_{im}}}} \right) + \ln\left(1-\frac{e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_m x_{im}}}{1+e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_m x_{im}}} \right) \right\} \\
&= \sum_{i=1}^{N} \left\{ y_i \ln\left(e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_m x_{im}} \right) + \ln\left(\frac{1}{1+e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_m x_{im}}} \right) \right\} \\
&= \sum_{i=1}^{N} \left\{ y_i \ln\left(e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_m x_{im}} \right) + \ln 1 - \ln\left(1+e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_m x_{im}} \right) \right\} \\
&= \sum_{i=1}^{N} \left\{ y_i \left(\beta_0 + \beta_1 x_{1i} + \ldots + \beta_m x_{im} \right) - \ln\left(1+e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_m x_{im}} \right) \right\} \qquad \ldots(2)
\end{aligned}$$

การประมาณค่าพารามิเตอร์ด้วยภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation)

ตามหลักการของวิธีภาวะน่าจะเป็นสูงสุด จะต้องทำให้ $L(\mathbf{B})$ มีค่ามากที่สุด (maximize) โดยการหาอนุพันธ์ (differentiate) เทียบกับ $\beta_0, \beta_1, \ldots, \beta_m$ แล้วให้ผลลัพธ์เป็นศูนย์ ได้สมการที่ไม่เป็นเชิงเส้น $m+1$ สมการ ซึ่งสามารถหาค่าของ $\beta_0, \beta_1, \ldots, \beta_m$ ได้จาก วิธีของฟิชเชอร์ (*Fisher's Method*) หรือจากวิธี *Newton – Raphson* หรือจากวิธีทำซ้ำ (*Iterative Method*(*McCullagh* และ *Nelder*(1981))) ในการวิจัยนี้จะใช้วิธี *Newton – Raphson*

ตามวิธี *Newton – Raphson* เรียกอนุพันธ์ย่อย (partial derivative) เทียบกับ $\beta_0, \beta_1, \ldots, \beta_m$ ของลอการิทึมธรรมชาติของฟังก์ชันภาวะน่าจะเป็น $L(\mathbf{B})$ ว่า efficient scores แล้วนำมาเป็นสมาชิกของเวกเตอร์ $\mathbf{U(B)}$ มิติ $(m+1)\times 1$

$$\mathbf{U(B)} = \begin{bmatrix} \dfrac{\partial L(\mathbf{B})}{\partial \beta_0} \\ \dfrac{\partial L(\mathbf{B})}{\partial \beta_1} \\ \vdots \\ \dfrac{\partial L(\mathbf{B})}{\partial \beta_m} \end{bmatrix}$$

กำหนดเมทริกซ์ $\mathbf{H(B)}$ มิติ $(m+1)\times(m+1)$ มีสมาชิกเป็นอนุพันธ์ย่อยอันดับที่สอง (second partial derivative) ของ $L(\mathbf{B})$

โดยที่ สมาชิกตัวที่ (j,k) คือ $\dfrac{\partial^2 L(\mathbf{B})}{\partial \beta_j \partial \beta_k}$ $\quad j,k = 0,1,2,\ldots,m$

เรียกเมทริกซ์ $\mathbf{H(B)}$ ว่า Hessian matrix

พิจารณาหาเวกเตอร์ $\mathbf{U}(\hat{\mathbf{B}})$ ซึ่งเป็นเวกเตอร์ของ efficient scores ของ $\mathbf{B}$ ที่ประมาณด้วยภาวะน่าจะเป็นสูงสุด ใช้ *Taylor Series* กระจาย $\mathbf{U(B)}$ รอบ $\mathbf{B}^*$ ซึ่ง $\mathbf{B}^*$ อยู่ใกล้ๆ $\hat{\mathbf{B}}$ จะได้

$$\mathbf{U}(\hat{\mathbf{B}}) \approx \mathbf{U}(\mathbf{B}^*) + \mathbf{H}(\mathbf{B}^*)(\hat{\mathbf{B}} - \mathbf{B}^*)$$

โดยนิยามของตัวประมาณภาวะน่าจะเป็นสูงสุดของ $\mathbf{B}$ จะได้ว่า

$$\left.\frac{\partial L(\mathbf{B})}{\partial \beta_j}\right|_{\hat{\mathbf{B}}} = 0 \qquad \text{สำหรับ } j = 0,1,2,\ldots,m$$

และ $\quad \mathbf{U}(\hat{\mathbf{B}}) = 0$

ดังนั้น $\quad \hat{\mathbf{B}} = \mathbf{B}^* - \mathbf{H}^{-1}(\mathbf{B}^*)\mathbf{U}(\mathbf{B}^*)$

ซึ่งชี้นำไปให้มีการประมาณ $\hat{\mathbf{B}}$ โดยการคำนวนซ้ำๆ ซึ่งค่าประมาณ $\hat{\mathbf{B}}$ ณ รอบที่ $r+1$

คือ $$\hat{\mathbf{B}}_{r+1} = \hat{\mathbf{B}}_r - \mathbf{H}^{-1}(\mathbf{B}_r)\mathbf{U}(\mathbf{B}_r) \qquad \ldots(3)$$

สำหรับ $r = 0,1,2,\ldots$ ซึ่ง $\hat{\mathbf{B}}_0$ เป็นเวกเตอร์ของตัวประมาณเริ่มต้น

1. วิธี Newton-Raphson สำหรับตัวแปรอธิบาย 1 ตัว

ข้อมูลประกอบด้วยค่าสังเกตที่เป็นอิสระ N ค่า คือ (x_i, y_i) , $y_i = 0,1$ $i = 1, 2, \ldots, \mathrm{N}$ สมการถดถอยโลจิสติค คือ

$$\begin{aligned} \operatorname{logit}(\pi(x_i)) &= \ln\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) \\ &= \beta_0 + \beta_1 x_i \end{aligned}$$

เมื่อ $\pi(x_i)$ = ความน่าจะเป็นที่ $y = 1$ กำหนดว่า $x = x_i$

$$\pi(x_i) = \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}$$

หาค่าพารามิเตอร์ β_0 และ β_1 จากการคำนวนซ้ำ ด้วยวิธี *Newton – Raphson* ด้วยสมการ

$$\hat{\mathbf{B}}_{r+1} = \hat{\mathbf{B}}_r - \mathbf{H}^{-1}(\mathbf{B}_r)\mathbf{U}(\mathbf{B}_r)$$

เมื่อ $$\mathbf{B} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} , \qquad \mathbf{U}(\mathbf{B}) = \begin{bmatrix} \dfrac{\partial L(\mathbf{B})}{\partial \beta_0} \\ \dfrac{\partial L(\mathbf{B})}{\partial \beta_1} \end{bmatrix}$$

และ $$\mathbf{H}(\mathbf{B}) = \begin{bmatrix} \dfrac{\partial^2 L(\mathbf{B})}{\partial \beta_0^2} & \dfrac{\partial^2 L(\mathbf{B})}{\partial \beta_0 \partial \beta_1} \\ \dfrac{\partial^2 L(\mathbf{B})}{\partial \beta_1 \partial \beta_0} & \dfrac{\partial^2 L(\mathbf{B})}{\partial \beta_1^2} \end{bmatrix} \qquad \ldots(4)$$

สมาชิกในเวกเตอร์ $\mathbf{U}(\mathbf{B})$ และในเมตริกซ์ $\mathbf{H}(\mathbf{B})$ หาได้จากการหาอนุพันธ์ของ $L(\mathbf{B})$ เทียบกับ β_0 และ β_1 ดังนี้

จาก (2) และ สำหรับตัวแปรอธิบาย 1 ตัว

$$\begin{aligned} L(\mathbf{B}) &= \sum_{i=1}^{N} \left\{ y_i(\beta_0 + \beta_1 x_i) - \ln(1 + e^{\beta_0+\beta_1 x_i}) \right\} \\ \frac{\partial L(\mathbf{B})}{\partial \beta_0} &= \sum_{i=1}^{N} \left\{ y_i - \frac{1}{1+e^{\beta_0+\beta_1 x_i}} (e^{\beta_0+\beta_1 x_i}) \right\} \\ &= \sum_{i=1}^{N} \left\{ y_i - \pi(x_i) \right\} \end{aligned}$$

$$\frac{\partial L(\mathbf{B})}{\partial \beta_1} = \sum_{i=1}^{N}\left\{x_i y_i - \frac{1}{1+e^{\beta_0+\beta_1 x_i}}(x_i e^{\beta_0+\beta_1 x_i})\right\}$$
$$= \sum_{i=1}^{N} x_i\left\{y_i - \pi(x_i)\right\}$$

$$\frac{\partial^2 L(\mathbf{B})}{\partial \beta_0^2} = \sum_{i=1}^{N}\left\{0-\left\{\frac{e^{\beta_0+\beta_1 x_i}(1+e^{\beta_0+\beta_1 x_i})-e^{\beta_0+\beta_1 x_i}(0+e^{\beta_0+\beta_1 x_i})}{(1+e^{\beta_0+\beta_1 x_i})^2}\right\}\right\}$$
$$= \sum_{i=1}^{N}\left\{\frac{-(e^{\beta_0+\beta_1 x_i}+e^{2(\beta_0+\beta_1 x_i)}-e^{2(\beta_0+\beta_1 x_i)})}{(1+e^{\beta_0+\beta_1 x_i})^2}\right\}$$
$$= \sum_{i=1}^{N}\left\{-\frac{e^{\beta_0+\beta_1 x_i}}{(1+e^{\beta_0+\beta_1 x_i})^2}\right\}$$
$$= \sum_{i=1}^{N}\left\{-\frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}\left(\frac{1+e^{\beta_0+\beta_1 x_i}-e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}\right)\right\}$$
$$= \sum_{i=1}^{N}\left\{-\frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}\left(1-\frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}\right)\right\}$$
$$= -\sum_{i=1}^{N}\pi(x_i)(1-\pi(x_i))$$

$$\frac{\partial^2 L(\mathbf{B})}{\partial \beta_0 \partial \beta_1} = \sum_{i=1}^{N}\left\{0-\left\{\frac{e^{\beta_0+\beta_1 x_i}x_i(1+e^{\beta_0+\beta_1 x_i})-e^{\beta_0+\beta_1 x_i}(0+e^{\beta_0+\beta_1 x_i}x_i)}{(1+e^{\beta_0+\beta_1 x_i})^2}\right\}\right\}$$
$$= \sum_{i=1}^{N}\left\{-\frac{x_i e^{\beta_0+\beta_1 x_i}+x_i e^{2(\beta_0+\beta_1 x_i)}-x_i e^{2(\beta_0+\beta_1 x_i)}}{(1+e^{\beta_0+\beta_1 x_i})^2}\right\}$$
$$= \sum_{i=1}^{N}\left\{-\frac{x_i e^{\beta_0+\beta_1 x_i}}{(1+e^{\beta_0+\beta_1 x_i})^2}\right\}$$
$$= -\sum_{i=1}^{N}x_i\pi(x_i)(1-\pi(x_i))$$

$$\frac{\partial^2 L(\mathbf{B})}{\partial \beta_1 \partial \beta_0} = \frac{\partial^2 L(\mathbf{B})}{\partial \beta_0 \partial \beta_1}$$

และ $$\frac{\partial^2 L(\mathbf{B})}{\partial \beta_1^2} = \sum_{i=1}^{N}x_i\left\{0-\left\{\frac{e^{\beta_0+\beta_1 x_i}x_i(1+e^{\beta_0+\beta_1 x_i})-e^{\beta_0+\beta_1 x_i}(0+e^{\beta_0+\beta_1 x_i}x_i)}{(1+e^{\beta_0+\beta_1 x_i})^2}\right\}\right\}$$
$$= \sum_{i=1}^{N}x_i\left\{-\left\{\frac{x_i e^{\beta_0+\beta_1 x_i}+x_i e^{2(\beta_0+\beta_1 x_i)}-x_i e^{2(\beta_0+\beta_1 x_i)}}{(1+e^{\beta_0+\beta_1 x_i})^2}\right\}\right\}$$

$$= \sum_{i=1}^{N}\left\{-\frac{x_i^2 e^{\beta_0+\beta_1 x_i}}{(1+e^{\beta_0+\beta_1 x_i})^2}\right\}$$

$$= -\sum_{i=1}^{N} x_i^2 \pi(x_i)(1-\pi(x_i))$$

นำสมาชิกของเวกเตอร์ $\mathbf{U}(\mathbf{B})$ และเมตริกซ์ $\mathbf{H}(\mathbf{B})$ ที่หาได้ ไปแทนในสมการ (4) ดังนั้นคำนวนหาค่า $\hat{\beta}_0$ และ $\hat{\beta}_1$ ได้จากสมการ

$$\hat{\mathbf{B}}_{r+1} = \hat{\mathbf{B}}_r - \mathbf{H}^{-1}(\mathbf{B}_r)\mathbf{U}(\mathbf{B}_r)$$

เมื่อ $\mathbf{B} = \begin{bmatrix}\beta_0 \\ \beta_1\end{bmatrix}$, $\mathbf{U}(\mathbf{B}) = \begin{bmatrix}\sum_{i=1}^{N}\left\{y_i - \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}\right\} \\ \sum_{i=1}^{N} x_i\left\{y_i - \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}\right\}\end{bmatrix}$

และ $\mathbf{H}(\mathbf{B}) = \begin{bmatrix} -\sum_{i=1}^{N}\frac{e^{\beta_0+\beta_1 x_i}}{(1+e^{\beta_0+\beta_1 x_i})^2} & -\sum_{i=1}^{N} x_i\frac{e^{\beta_0+\beta_1 x_i}}{(1+e^{\beta_0+\beta_1 x_i})^2} \\ -\sum_{i=1}^{N} x_i\frac{e^{\beta_0+\beta_1 x_i}}{(1+e^{\beta_0+\beta_1 x_i})^2} & -\sum_{i=1}^{N} x_i^2\frac{e^{\beta_0+\beta_1 x_i}}{(1+e^{\beta_0+\beta_1 x_i})^2}\end{bmatrix}$

ถ้าผลต่างระหว่าง $\hat{\mathbf{B}}$ ในรอบที่ r กับรอบที่ r+1 มีค่าน้อยมากจนถือว่าไม่แตกต่างกัน โดยกำหนดเกณฑ์ว่า $\left|\hat{\mathbf{B}}_{r+1} - \hat{\mathbf{B}}_r\right| < 0.0000001$ ค่า $\hat{\mathbf{B}}_{r+1}$ นั้นจะเป็นค่าที่ยอมรับได้

คำนวนหาค่าพยากรณ์ $\hat{\pi}(x_i)$ จาก $\hat{\pi}(x_i) = \frac{e^{\hat{\beta}_0+\hat{\beta}_1 x_i}}{1+e^{\hat{\beta}_0+\hat{\beta}_1 x_i}}$

เมื่อ $\hat{\pi}(x_i)$ คือ ความน่าจะเป็นที่ $y=1$ ณ ระดับ x_i

2. วิธี Newton-Raphson สำหรับตัวแปรอธิบาย 2 ตัว

ข้อมูลประกอบด้วย ค่าสังเกตที่เป็นอิสระ N ค่า คือ (x_{1i}, x_{2i}, y_i) , $y_i = 0,1$; $i = 1, 2, \ldots, \mathrm{N}$ สมการถดถอยโลจิสติค คือ

$$\operatorname{logit}(\pi(\mathbf{x}_i)) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$$

เมื่อ $\pi(\mathbf{x}_i)$ = ความน่าจะเป็นที่ $y=1$ กำหนดค่า $\mathbf{x}_i = x_{1i}, x_{2i}$

$$\pi(\mathbf{x}_i) = \frac{e^{\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}}}{1+e^{\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}}}$$

หาค่าพารามิเตอร์ β_0, β_1 และ β_2 จากการคำนวนซ้ำ ด้วยวิธี $Newton - Raphson$ ด้วยสมการ

$$\hat{\mathbf{B}}_{r+1} = \hat{\mathbf{B}}_r - \mathbf{H}^{-1}(\mathbf{B}_r)\mathbf{U}(\mathbf{B}_r)$$

เมื่อ $\mathbf{B} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$

$\mathbf{U}(\mathbf{B})$ เป็นเวกเตอร์มิติ 3×1 สมาชิกคือ $\dfrac{\partial L(\mathbf{B})}{\partial \beta_j}$, $j = 0,1,2$

$$\mathbf{U}(\mathbf{B}) = \begin{bmatrix} \dfrac{\partial L(\mathbf{B})}{\partial \beta_0} \\ \dfrac{\partial L(\mathbf{B})}{\partial \beta_1} \\ \dfrac{\partial L(\mathbf{B})}{\partial \beta_2} \end{bmatrix} \qquad \ldots(5)$$

และ Hessian matrix , $\mathbf{H}(\mathbf{B})$ มิติ 3×3 สมาชิกตัวที่ (i, j) คือ $\dfrac{\partial^2 L(\mathbf{B})}{\partial \beta_j \beta_k}$; $j, k = 0,1,2$

$$\mathbf{H}(\mathbf{B}) = \begin{bmatrix} \dfrac{\partial^2 L(\mathbf{B})}{\partial \beta_0^2} & \dfrac{\partial^2 L(\mathbf{B})}{\partial \beta_0 \partial \beta_1} & \dfrac{\partial^2 L(\mathbf{B})}{\partial \beta_0 \partial \beta_2} \\ \dfrac{\partial^2 L(\mathbf{B})}{\partial \beta_1 \partial \beta_0} & \dfrac{\partial^2 L(\mathbf{B})}{\partial \beta_1^2} & \dfrac{\partial^2 L(\mathbf{B})}{\partial \beta_1 \partial \beta_2} \\ \dfrac{\partial^2 L(\mathbf{B})}{\partial \beta_2 \partial \beta_0} & \dfrac{\partial^2 L(\mathbf{B})}{\partial \beta_2 \partial \beta_1} & \dfrac{\partial^2 L(\mathbf{B})}{\partial \beta_2^2} \end{bmatrix} \qquad \ldots(6)$$

สมาชิกในเวกเตอร์ $\mathbf{U}(\mathbf{B})$ และ Hessian matrix , $\mathbf{H}(\mathbf{B})$ หาได้จาก $L(\mathbf{B})$ จาก (2) และสำหรับตัวแปรอธิบาย 2 ตัว

$$L(\mathbf{B}) = \sum_{i=1}^{N} \left\{ y_i(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}) - \ln(1 + e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}}) \right\}$$

หาอนุพันธ์ของ $L(\mathbf{B})$ เทียบกับ β_0, β_1 และ β_2 จะได้

$$\frac{\partial L(\mathbf{B})}{\partial \beta_0} = \sum_{i=1}^{N} \left\{ y_i - \pi(\mathbf{x_i}) \right\}$$

$$\frac{\partial L(\mathbf{B})}{\partial \beta_1} = \sum_{i=1}^{N} x_{1i} \left\{ y_i - \pi(\mathbf{x_i}) \right\}$$

$$\frac{\partial L(\mathbf{B})}{\partial \beta_2} = \sum_{i=1}^{N} x_{2i}\left\{y_i - \pi(\mathbf{x_i})\right\}$$

โดยที่ค่า $\pi(\mathbf{x_i}) = \dfrac{e^{\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}}}{1+e^{\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}}}$

และสามารถหาอนุพันธ์ย่อยอันดับที่ 2 ได้ในรูป

$$\frac{\partial^2 L(\mathbf{B})}{\partial \beta_j^2} = -\sum_{i=1}^{N} x_{ji}^2 \pi(\mathbf{x_i})(1-\pi(\mathbf{x_i}))$$

$$\frac{\partial^2 L(\mathbf{B})}{\partial \beta_j \partial \beta_k} = -\sum_{i=1}^{N} x_{ji} x_{ki} \pi(\mathbf{x_i})(1-\pi(\mathbf{x_i})) \qquad ;\ x_{0i} = 1$$

ซึ่งทำให้ได้

$$\frac{\partial^2 L(\mathbf{B})}{\partial \beta_0^2} = -\sum_{i=1}^{N} \pi(\mathbf{x_i})(1-\pi(\mathbf{x_i}))$$

$$\frac{\partial^2 L(\mathbf{B})}{\partial \beta_1^2} = -\sum_{i=1}^{N} x_{1i}^2 \pi(\mathbf{x_i})(1-\pi(\mathbf{x_i}))$$

$$\frac{\partial^2 L(\mathbf{B})}{\partial \beta_2^2} = -\sum_{i=1}^{N} x_{2i}^2 \pi(\mathbf{x_i})(1-\pi(\mathbf{x_i}))$$

$$\frac{\partial^2 L(\mathbf{B})}{\partial \beta_0 \partial \beta_1} = \frac{\partial^2 L(\mathbf{B})}{\partial \beta_1 \partial \beta_0} = -\sum_{i=1}^{N} x_{1i} \pi(\mathbf{x_i})(1-\pi(\mathbf{x_i}))$$

$$\frac{\partial^2 L(\mathbf{B})}{\partial \beta_0 \partial \beta_2} = \frac{\partial^2 L(\mathbf{B})}{\partial \beta_2 \partial \beta_0} = -\sum_{i=1}^{N} x_{2i} \pi(\mathbf{x_i})(1-\pi(\mathbf{x_i}))$$

และ $$\frac{\partial^2 L(\mathbf{B})}{\partial\beta_1\partial\beta_2} = \frac{\partial^2 L(\mathbf{B})}{\partial\beta_2\partial\beta_1}$$
$$= -\sum_{i=1}^{N} x_{1i}x_{2i}\pi(\mathbf{x_i})(1-\pi(\mathbf{x_i}))$$

โดยที่ค่า $\pi(\mathbf{x_i})(1-\pi(\mathbf{x_i})) = \dfrac{e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}}}{(1+e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}})^2}$

นำสมาชิกของเวกเตอร์ $\mathbf{U(B)}$ และเมทริกซ์ $\mathbf{H(B)}$ ที่หาได้ ไปแทนในสมการ (5) และ (6) ดังนั้นคำนวนหาค่า $\hat{\beta}_0$, $\hat{\beta}_1$ และ $\hat{\beta}_2$ ได้จากสมการ

$$\hat{\mathbf{B}}_{r+1} = \hat{\mathbf{B}}_r - \mathbf{H}^{-1}(\mathbf{B}_r)\mathbf{U}(\mathbf{B}_r)$$

เมื่อ $$\mathbf{B} = \begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\end{bmatrix}, \quad \mathbf{U(B)} = \begin{bmatrix} \sum_{i=1}^{N}\left\{y_i - \dfrac{e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}}}{1+e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}}}\right\} \\ \sum_{i=1}^{N}x_{1i}\left\{y_i - \dfrac{e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}}}{1+e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}}}\right\} \\ \sum_{i=1}^{N}x_{2i}\left\{y_i - \dfrac{e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}}}{1+e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}}}\right\} \end{bmatrix}$$

$$\mathbf{H(B)} = \begin{bmatrix} -\sum_{i=1}^{N}\dfrac{e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}}}{(1+e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}})^2} & -\sum_{i=1}^{N}x_{1i}\dfrac{e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}}}{(1+e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}})^2} & -\sum_{i=1}^{N}x_{2i}\dfrac{e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}}}{(1+e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}})^2} \\ -\sum_{i=1}^{N}x_{1i}\dfrac{e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}}}{(1+e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}})^2} & -\sum_{i=1}^{N}x_{1i}^2\dfrac{e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}}}{(1+e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}})^2} & -\sum_{i=1}^{N}x_{1i}x_{2i}\dfrac{e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}}}{(1+e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}})^2} \\ -\sum_{i=1}^{N}x_{2i}\dfrac{e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}}}{(1+e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}})^2} & -\sum_{i=1}^{N}x_{1i}x_{2i}\dfrac{e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}}}{(1+e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}})^2} & -\sum_{i=1}^{N}x_{2i}^2\dfrac{e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}}}{(1+e^{\beta_0+\beta_1x_{1i}+\beta_2x_{2i}})^2} \end{bmatrix}$$

ถ้าผลต่างระหว่าง $\hat{\mathbf{B}}$ ในรอบที่ r กับรอบที่ r+1 มีค่าน้อยมากจนถือว่าไม่แตกต่างกัน โดยกำหนดเกณฑ์ว่า $\left|\hat{\mathbf{B}}_{r+1} - \hat{\mathbf{B}}_r\right| < 0.0000001$ ค่า $\hat{\mathbf{B}}_{r+1}$ นั้นจะเป็นค่าที่ยอมรับได้

คำนวนหาค่าพยากรณ์ $\hat{\pi}(\mathbf{x_i})$ จาก $\hat{\pi}(\mathbf{x_i}) = \dfrac{e^{\hat{\beta}_0+\hat{\beta}_1x_{1i}+\hat{\beta}_2x_{2i}}}{1+e^{\hat{\beta}_0+\hat{\beta}_1x_{1i}+\hat{\beta}_2x_{2i}}}$

เมื่อ $\hat{\pi}(\mathbf{x_i})$ คือ ความน่าจะเป็นที่ $y=1$ ณ ระดับ x_{1i} และ x_{2i}

ฟังก์ชันจำแนกประเภทเชิงเส้นของฟิชเชอร์ (Fisher's Linear Discriminant Function)

สมมติเวกเตอร์สุ่ม $\mathbf{X}$ เป็นประชากรของตัวแปรสุ่มร่วม m ตัว ที่มีการแจกแจงปกติ (m-dimension normal population) ประชากรถูกแบ่งเป็นกลุ่มย่อย 2 กลุ่ม มีค่าเฉลี่ยต่างกัน แต่มีความแปรปรวนเท่ากัน กล่าวคือ

$$\mathbf{X} \sim N(\mathbf{U}_1, \Sigma) \quad \text{ด้วยความน่าจะเป็น} \quad \pi_1$$

$$\mathbf{X} \sim N(\mathbf{U}_0, \Sigma) \quad \text{ด้วยความน่าจะเป็น} \quad \pi_0$$

ซึ่ง $\pi_1 + \pi_0 = 1$

ถ้าพารามิเตอร์ π_1 , π_0 , $\mathbf{U}_1$, $\mathbf{U}_0$ และ Σ ทราบค่า จะได้ว่า เวกเตอร์สุ่ม $\mathbf{X}$ สามารถถูกกำหนดเข้าอยู่ในกลุ่มย่อยได้ด้วยฟังก์ชันที่ชื่อว่า ฟังก์ชันจำแนกประเภทเชิงเส้นของฟิชเชอร์ (Fisher's Linear Discriminant Function) ฟังก์ชันนี้ คือ

$$\lambda(\mathbf{X}) = \beta_0 + \mathbf{B}'\mathbf{X} \qquad \ldots(7)$$

เมื่อ

$$\beta_0 = \ln\left(\frac{\pi_1}{\pi_0}\right) - \tfrac{1}{2}\left(\mathbf{U}_1 - \mathbf{U}_0\right)' \Sigma^{-1}\left(\mathbf{U}_1 + \mathbf{U}_0\right) \qquad \ldots(8)$$

$$\mathbf{B}' = \left(\mathbf{U}_1 - \mathbf{U}_0\right)' \Sigma^{-1} \qquad \ldots(9)$$

$\mathbf{X}$ อยู่ในประชากรกลุ่มย่อย 1 ถ้า $\lambda(\mathbf{X}) > 0$ และ $\mathbf{X}$ อยู่ในประชากรกลุ่มย่อย 0 ถ้า $\lambda(\mathbf{X}) < 0$

ในทางปฏิบัติ พารามิเตอร์ π_1 , π_0 , $\mathbf{U}_1$, $\mathbf{U}_0$ และ Σ เป็นพารามิเตอร์ที่ไม่ทราบค่า แต่กลุ่มตัวอย่าง $(y_1, \mathbf{X}_1)$, $(y_2, \mathbf{X}_2), \ldots, (y_N, \mathbf{X}_N)$ ทำให้หาตัวประมาณค่าของพารามิเตอร์เหล่านี้ได้ โดย y_j จะชี้ว่า $\mathbf{X}_j$ มาจากประชากรกลุ่มย่อยกลุ่มใด

$$y_j = \begin{cases} 1 & \text{ด้วยความน่าจะเป็น} \quad \pi_1 \\ 0 & \text{ด้วยความน่าจะเป็น} \quad \pi_0 \end{cases}$$

และ $\mathbf{X}_j \setminus y_j \sim N(\mathbf{U}_{y_j}, \Sigma)$

และสมมติให้ ค่าสังเกต $(y_i, \mathbf{X}_i)$ เป็นอิสระซึ่งกันและกัน ซึ่งทำให้ตัวประมาณค่าภาวะน่าจะเป็นสูงสุดของพารามิเตอร์เหล่านี้ หาได้จาก

$$\hat{\pi}_1 = \frac{n_1}{\mathrm{N}}$$

$$\hat{\pi}_0 = \frac{n_0}{\mathrm{N}}$$

$$\hat{\mathbf{U}}_1 = \overline{\mathbf{X}}_1 = \frac{1}{n_1}\sum_{y_j=1}\mathbf{X}_j$$

$$\hat{\mathbf{U}}_0 = \overline{\mathbf{X}}_0 = \frac{1}{n_0}\sum_{y_j=0}\mathbf{X}_j$$

และ $$\Sigma = \frac{1}{N}\left\{\sum_{y_j=1}(\mathbf{X}_j - \overline{\mathbf{X}}_1)(\mathbf{X}_j - \overline{\mathbf{X}}_1)' + \sum_{y_j=1}(\mathbf{X}_j - \overline{\mathbf{X}}_0)(\mathbf{X}_j - \overline{\mathbf{X}}_0)'\right\}$$

เมื่อ n_1 และ n_0 เป็นจำนวนค่าสังเกตที่ $y = 1$ และ $y = 0$ ตามลำดับ

เมื่อแทนค่าพารามิเตอร์ใน (7) (8) และ (9) ด้วยค่าประมาณเหล่านี้ จะได้ฟังก์ชันค่าประมาณของฟังก์ชันจำแนกประเภทเชิงเส้น

$$\hat{\lambda}(\mathbf{X}) = \hat{\beta}_0 + \hat{\mathbf{B}}'\mathbf{X}$$

ฟังก์ชันนี้จะเป็น ตัวกำหนดค่าเวคเตอร์สุ่ม $\mathbf{X}$ ตัวใหม่ ว่าจะอยู่ในประชากร 1 หรือประชากร 0 ตามที่ $\hat{\lambda}(\mathbf{X})$ มากกว่า 0 หรือ น้อยกว่า 0 เรียกวิธีการจำแนกกลุ่ม 2 กลุ่มนี้ว่า Normal Discriminant Procedue

ทฤษฎีบทของเบส์ (Bayes's theorem) อธิบายได้ว่า ฟังก์ชันจำแนกประเภท $\lambda(\mathbf{X})$ นี้ แท้จริงคือ log odds ratio สำหรับประชากร 1 ต่อประชากร 0 ณ ค่าสังเกต $\mathbf{X}_i$ กล่าวคือ

$$\lambda(\mathbf{X}) = \ln\left(\frac{\pi_1(\mathbf{X}_i)}{\pi_0(\mathbf{X}_i)}\right)$$

เมื่อ $\pi_1(\mathbf{X}_i)$ คือ ความน่าจะเป็นที่ $y = 1$ ณ ระดับที่ $\mathbf{X}_i$

$\pi_0(\mathbf{X}_i)$ คือ ความน่าจะเป็นที่ $y = 0$ ณ ระดับที่ $\mathbf{X}_i$

สำหรับค่าสังเกต $(y_1, \mathbf{X}_1)$, $(y_2, \mathbf{X}_2), \ldots, (y_N, \mathbf{X}_N)$ เมื่อ y_i เป็นตัวแปรตามชนิด 2 ค่า

$$\text{Prob}\{y_i = 1 \backslash \mathbf{X}_i\} = \pi_1(\mathbf{X}_i) = \frac{e^{\beta_0 + \mathbf{B}'\mathbf{X}_i}}{1 + e^{\beta_0 + \mathbf{B}'\mathbf{X}_i}}$$

และ $$\text{Prob}\{y_i = 0 \backslash \mathbf{X}_i\} = \pi_0(\mathbf{X}_i) = \frac{1}{1 + e^{\beta_0 + \mathbf{B}'\mathbf{X}_i}}$$

การประมาณค่า β_0 และ $\mathbf{B}$ ทำได้โดยทำให้ ฟังก์ชันภาวะน่าจะเป็น คือ

$$f_{\beta_0,\mathbf{B}}(y_1, y_2, \ldots, y_N \backslash \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N) = \prod_{i=1}^{N}\left(\pi_1(\mathbf{X}_i)\right)^{y_i}\left(1 - \pi_1(\mathbf{X}_i)\right)^{1-y_i}$$

มีค่ามากที่สุด (maximize) ด้วยการหาอนุพันธ์เทียบกับ β_0 และ $\mathbf{B}$ ค่ามากที่สุดที่ได้ ทำให้ได้ฟังก์ชันประมาณค่าของฟังก์ชันจำแนกประเภทเชิงเส้น $\hat{\lambda}(\mathbf{X}) = \hat{\beta}_0 + \hat{\mathbf{B}}'\mathbf{X}$ ซึ่งใช้เป็นค่าตัดสินใจว่า $\mathbf{X}$ จะอยู่ในประชากรกลุ่มใด เรียกวิธีจำแนกกลุ่มวิธีนี้ว่า Logistic Regression Procedue

ที่กล่าวมานี้ได้แสดงให้เห็นความสัมพันธ์ระหว่าง ฟังก์ชันโลจิสติคของสมการถดถอยโลจิสติค กับ ฟังก์ชันจำแนกประเภท กล่าวคือ

$$\ln\left(\frac{\pi_1(\mathbf{X}_i)}{1-\pi_1(\mathbf{X}_i)}\right) = \lambda(\mathbf{X})$$
$$= \beta_0 + \mathbf{B}'\mathbf{X}_i$$

การประมาณค่าพารามิเตอร์ด้วยฟังก์ชันจำแนกประเภท (Discriminant Function)

1. กรณีตัวแปรอธิบาย 1 ตัว

Cornfield(1962) นำฟังก์ชันจำแนกประเภทมาใช้ประมาณค่าพารามิเตอร์ในสมการถดถอยโลจิสติค โดยชี้ให้เห็นว่า ฟังก์ชันความน่าจะเป็นที่หาได้ภายหลัง (posterior probability) ในฟังก์ชันจำแนกประเภท เป็นฟังก์ชันโลจิสติค ที่กำหนดด้วย $\pi(x) = \dfrac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$ กล่าวได้ว่า สำหรับตัวแปรอธิบาย x ที่ถูกกำหนดด้วยค่าของ y ให้อยู่แยกกันเป็นสองกลุ่ม ถ้าตัวแปรอธิบาย x ในแต่ละกลุ่มมีการแจกแจงปกติ มีค่าเฉลี่ยต่างกัน แต่มีความแปรปรวนเดียวกัน จะได้ว่า การแจกแจงอย่างมีเงื่อนไขของ y กำหนด x เป็นฟังก์ชันโลจิสติค

นั่นคือ ถ้า $x \setminus y = j$ มีการแจกแจง $N(\mu_j, \sigma^2)$, $j = 0,1$ จะได้ว่า

$P(y=1 \setminus x) = \pi(x)$

ภายใต้ข้อตกลงเบื้องต้นนี้ *Bruch*(1975) ประมาณค่าพารามิเตอร์ β_0 และ β_1 จาก

$$\beta_0 = \ln\left(\frac{\theta_1}{\theta_0}\right) - 0.5\left(\frac{\mu_1^2 - \mu_0^2}{\sigma^2}\right)$$

และ $$\beta_1 = \frac{(\mu_1 - \mu_0)}{\sigma^2}$$

เมื่อ $$\theta_j = P(y=j) \quad ; j = 0,1$$

ประมาณค่า β_0 และ β_1 ได้โดยแทนค่าด้วยตัวประมาณของ μ_j, θ_j ; $j = 0,1$ และ σ^2 ในสมการทั้งสอง

ตัวประมาณที่ใช้ คือ

$\hat{\mu}_j = \bar{x}_j$ ซึ่งเป็นค่าเฉลี่ยของ x ในกลุ่มย่อยที่ $y = j$

$\theta_1 = \dfrac{n_1}{N}$ เป็นค่าเฉลี่ยของ y

$\theta_0 = 1 - \theta_1$

และ $$\hat{\sigma}^2 = \frac{(n_0 - 1)S_0^2 + (n_1 - 1)S_1^2}{n_0 + n_1 - 2}$$

เมื่อ S_j^2 เป็นตัวประมาณที่ไม่เอนเอียงของ σ^2 คำนวนจากกลุ่มย่อยที่ $y = j \ ; j = 0,1$

2. <u>กรณีตัวแปรอธิบายหลายตัว</u>

Truett,*Cornfield* และ *Kannel*(1967) กล่าวถึงข้อตกลงเบื้องต้นที่จำเป็นในการนำฟังก์ชันจำแนกประเภทมาประมาณค่าพารามิเตอร์ว่า ถ้าการแจกแจงอย่างมีเงื่อนไขของ $\mathbf{x}$ (ซึ่งเป็นเวคเตอร์ของ covariate random variable m ตัว) กำหนดด้วยตัวแปรตาม y มีการแจกแจงปกติพหุนาม มีเวกเตอร์ค่าเฉลี่ยที่ขึ้นกับค่า y แต่เมทริกซ์ความแปรปรวนร่วมคงที่

นั่นคือ ถ้า $\mathbf{x} \setminus y = j$ มีการแจกแจง $N(\mathbf{U}_j, \Sigma) \ ; j = 0,1$

เมื่อ $\mathbf{U}_j$ เป็นเวกเตอร์ค่าเฉลี่ยของตัวแปรอธิบาย m ตัว ในประชากรย่อยที่กำหนดด้วย $y = j$

Σ เป็นเมทริกซ์ความแปรปรวนร่วม มิติ $m \times m$ ภายใต้ข้อตกลงเบื้องต้นดังกล่าว

จะได้ว่า $P(y = 1 \setminus \mathbf{x}) = \pi(\mathbf{x})$

ค่าพารามิเตอร์ในฟังก์ชันโลจิสติคหาได้จากสมการต่อไปนี้

$$\beta_0 = \ln\left(\frac{\theta_1}{\theta_0}\right) - (0.5)\left(\mathbf{U}_1 - \mathbf{U}_0\right)' \Sigma^{-1}\left(\mathbf{U}_1 + \mathbf{U}_0\right)$$

และ

$$\mathbf{B}' = \left(\mathbf{U}_1 - \mathbf{U}_0\right)' \Sigma^{-1}$$

เมื่อ $\theta_1 = P(y = 1 \setminus \mathbf{x})$ และ $\theta_0 = 1 - \theta_1$ เป็นสัดส่วนของประชากรที่มีค่า $y = 1$ และ $y = 0$ ตามลำดับ

ประมาณค่า β_0 และ $\mathbf{B}$ ได้โดยแทนค่าด้วยตัวประมาณสำหรับ $\mathbf{U}_j \ ; \ j = 0,1$, Σ , θ_1 และ θ_0

ตัวประมาณของ $\mathbf{U}_j$ คือ $\hat{\mathbf{U}}_j = \bar{\mathbf{x}}_\mathbf{i}$ ซึ่งเป็นเวกเตอร์ค่าเฉลี่ยในกลุ่มย่อยที่มี $y = j \ ; j = 0,1$

ตัวประมาณของเมทริกซ์ความแปรปรวนร่วม Σ คือ เมตริกซ์ $\mathbf{S}$

$$\mathbf{S} = \frac{(n_0 - 1)\mathbf{S}_0 + (n_1 - 1)\mathbf{S}_1}{n_0 + n_1 - 2}$$

เมื่อ $\mathbf{S}_j$ คือ เมทริกซ์ $m \times m$ ของตัวประมาณที่ไม่เอนเอียงของความแปรปรวนและความแปรปรวนร่วมภายในกลุ่มย่อยที่มี $y = j \ ; j = 0,1$

การเตรียมข้อมูลเพื่อการคำนวนค่าพารามิเตอร์โดยฟังก์ชันจำแนกประเภท

1. กรณีตัวแปรอธิบาย 1 ตัว

ตารางการเตรียมข้อมูล

$x \setminus y = 1$	$x \setminus y = 0$
x_1 x_2 $\vdots$ x_{n_1}	x_1 x_2 $\vdots$ x_{n_0}
$\hat{\mu}_1 = \bar{x}_1 = \dfrac{1}{n_1}\sum_{i=1}^{n_1} x_i$ $S_1^2 = \dfrac{1}{n_1 - 1}\sum_{i=1}^{n_i}(x_i - \bar{x}_1)^2$	$\hat{\mu}_0 = \bar{x}_0 = \dfrac{1}{n_0}\sum_{i=1}^{n_0} x_i$ $S_0^2 = \dfrac{1}{n_0 - 1}\sum_{i=1}^{n_0}(x_i - \bar{x}_0)^2$

คำนวนหาค่า $\hat{\beta}_0$ จาก $\hat{\beta}_0 = \ln\left(\dfrac{\theta_1}{\theta_0}\right) - 0.5\left(\dfrac{\mu_1^2 - \mu_0^2}{\sigma^2}\right)$

และ คำนวนหาค่า $\hat{\beta}_1$ จาก $\hat{\beta}_1 = \dfrac{(\mu_1 - \mu_0)}{\sigma^2}$

เมื่อ $\theta_1 = \dfrac{n_1}{N}$

$\theta_0 = 1 - \theta_1$

และ $\hat{\sigma}^2 = \dfrac{(n_0 - 1)S_0^2 + (n_1 - 1)S_1^2}{n_0 + n_1 - 2}$

คำนวนหาค่าพยากรณ์ $\hat{\pi}(x_i)$ จาก $\hat{\pi}(x_i) = \dfrac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}}$

เมื่อ $\hat{\pi}(x_i)$ คือ ความน่าจะเป็นที่ $y = 1$ ณ ระดับ x_i

2. กรณีตัวแปรอธิบาย 2 ตัว

ตารางการเตรียมข้อมูล

$\mathbf{X} \setminus y = 1$	$\mathbf{X} \setminus y = 0$
x_{11}, x_{21} x_{12}, x_{22} $\vdots$ x_{1n_1}, x_{2n_1}	x_{11}, x_{21} x_{12}, x_{22} $\vdots$ x_{1n_0}, x_{2n_0}
$\bar{x}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} x_{1i}$ $\bar{x}_2 = \frac{1}{n_1}\sum_{i=1}^{n_1} x_{2i}$	$\bar{x}_1 = \frac{1}{n_0}\sum_{i=1}^{n_0} x_{1i}$ $\bar{x}_2 = \frac{1}{n_0}\sum_{i=1}^{n_0} x_{1i}$
$\hat{\sigma}^2{}_{x_1} = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(x_{1i} - \bar{x}_1)^2$ $\hat{\sigma}^2{}_{x_2} = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(x_{2i} - \bar{x}_2)^2$ $\sigma_{x_1x_2} = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(x_{1i} - \bar{x}_1)(x_{21} - \bar{x}_2)$	$\hat{\sigma}^2{}_{x_1} = \frac{1}{n_0 - 1}\sum_{i=1}^{n_0}(x_{1i} - \bar{x}_1)^2$ $\hat{\sigma}^2{}_{x_2} = \frac{1}{n_0 - 1}\sum_{i=1}^{n_0}(x_{1i} - \bar{x}_1)^2$ $\sigma_{x_1x_2} = \frac{1}{n_0 - 1}\sum_{i=1}^{n_0}(x_{1i} - \bar{x}_1)(x_{21} - \bar{x}_2)$
$\mathbf{U}_1 = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix}$ $\mathbf{S}_1 = \begin{bmatrix} \hat{\sigma}^2_{x_1} & \hat{\sigma}_{x_1x_2} \\ \hat{\sigma}_{x_2x_1} & \hat{\sigma}^2_{x_2} \end{bmatrix}$	$\mathbf{U}_0 = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix}$ $\mathbf{S}_0 = \begin{bmatrix} \hat{\sigma}^2_{x_1} & \hat{\sigma}_{x_1x_2} \\ \hat{\sigma}_{x_2x_1} & \hat{\sigma}^2_{x_2} \end{bmatrix}$

คำนวนหาค่า $\hat{\beta}_0$ ได้จาก $\hat{\beta}_0 = \ln\left(\frac{\theta_1}{\theta_0}\right) - (0.5)\left(\mathbf{U}_1 - \mathbf{U}_0\right)' \Sigma^{-1}\left(\mathbf{U}_1 + \mathbf{U}_0\right)$

และ คำนวนหาค่า $\hat{\beta}_1$, $\hat{\beta}_2$ จาก $\mathbf{B}' = \left(\mathbf{U}_1 - \mathbf{U}_0\right)' \Sigma^{-1}$

เมื่อ $\theta_1 = \frac{n_1}{N}$

$\theta_0 = 1 - \theta_1$

และ $\Sigma \approx \mathbf{S} = \frac{(n_0 - 1)\mathbf{S}_0 + (n_1 - 1)\mathbf{S}_1}{n_0 + n_1 - 2}$

คำนวนหาค่าพยากรณ์ $\hat{\pi}(\mathbf{x_i})$ จาก $\hat{\pi}(\mathbf{x_i}) = \dfrac{e^{\hat{\beta}_0+\hat{\beta}_1 x_{1i}+\hat{\beta}_2 x_{2i}}}{1+e^{\hat{\beta}_0+\hat{\beta}_1 x_{1i}+\hat{\beta}_2 x_{2i}}}$

เมื่อ $\hat{\pi}(\mathbf{x_i})$ คือ ความน่าจะเป็นที่ $y=1$ ณ ระดับ x_1 และ x_2

การประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยสุดถ่วงน้ำหนัก (Weighted Least Squares)

ในการสร้างสมการถดถอยโลจิสติค $Cox(1970)$ ให้ฟังก์ชันการแปลง คือ ฟังก์ชันโลจิต

$$\text{logit}\,\pi(\mathbf{x_i}) = \ln\left(\frac{\pi(\mathbf{x_i})}{1-\pi(\mathbf{x_i})}\right)$$

$$= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_m x_{mi}$$

ซึ่งทำให้ได้ฟังก์ชันโลจิสติค

$$\pi(\mathbf{x_i}) = \frac{e^{\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+\ldots+\beta_m x_{mi}}}{1+e^{\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+\ldots+\beta_m x_{mi}}} \qquad \ldots(10)$$

เมื่อ $\pi(\mathbf{x_i})$ = ความน่าจะเป็นที่ $y=1$ กำหนดค่า $\mathbf{x}$ ระดับที่ i

ในการประมาณค่าพารามิเตอร์ $\beta_0, \beta_1, \ldots, \beta_m$ ด้วยวิธีกำลังสองน้อยสุดถ่วงน้ำหนัก

จะพิจารณาให้ $$p^* = \ln\left(\frac{E(y \setminus \mathbf{x})}{1-E(y \setminus \mathbf{x})}\right)$$

$$= \ln\left(\frac{p}{1-p}\right)$$

$$= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_m x_{mi} \qquad \ldots(11)$$

และจะต้องให้มีการสังเกต n_i ครั้ง ณ ระดับ $\mathbf{x_i}$ ให้ y_i เป็นจำนวนครั้งที่ $y=1$, $i = 1,2,\ldots,\mathrm{N}$

การประมาณค่าพารามิเตอร์ $\beta_0, \beta_1, \ldots, \beta_m$ ให้ทำผลรวมของความคลาดเคลื่อนกำลังสองให้มีค่าน้อยที่สุด (minimize Sum of Squares Error (SSE))

$$\text{SSE} = \sum_{i=1}^{N}(\tilde{p}_i - \beta_0 + \beta_1 x_{1i} + \ldots + \beta_m x_{mi})^2$$

$$= \sum_{i=1}^{N}\left(\frac{y_i}{n_i} - \beta_0 + \beta_1 x_{1i} + \ldots + \beta_m x_{mi}\right)^2$$

โดยที่ y_i มีการแจกแจงทวินาม พารามิเตอร์ n_i, p_i

$$Var(y_i) = n_i p_i (1-p_i)$$

ดังนั้น $$Var\left(\frac{y_i}{n_i}\right) = \frac{p_i(1-p_i)}{n_i}$$

โดยที่ ความแปรปรวนของสัดส่วนของ $y=1$ ไม่คงที่ ความแปรปรวนขึ้นอยู่กับ p_i คือค่าความน่าจะเป็น (ที่ $y=1$) จริง ซึ่งไม่ทราบค่า

การประมาณค่าพารามิเตอร์จึงต้องใช้วิธีทำผลรวมของความคลาดเคลื่อนกำลังสองถ่วงน้ำหนัก ให้มีค่าน้อยที่สุด

ผลรวมของความคลาดเคลื่อนกำลังสองถ่วงน้ำหนัก คือ $\sum_{i=1}^{N} w_i(\tilde{p}_i - \beta_0 + \beta_1 x_{1i} + \ldots + \beta_m x_{mi})^2$

เมื่อ $$w_i = \frac{1}{\frac{p_i(1-p_i)}{n_i}}$$

โดยที่ p_i เป็นพารามิเตอร์ที่ไม่ทราบค่า จึงให้ใช้สัดส่วนค่าสังเกต $\tilde{p}_i$ คำนวนน้ำหนัก w_i และโดยที่วิธีนี้ จะต้องไม่มี $p_i = 0$ หรือ 1 จึงแทน p_i ด้วย $\frac{0.5}{\mathrm{N}}$ ถ้า $y=0$ และแทน p_i ด้วย $\frac{\mathrm{N}-0.5}{\mathrm{N}}$ ถ้า $y=1$

ในที่สุด $$\mathrm{SSE} = \sum_{i=1}^{N} w_i(\tilde{p}_i - \beta_0 + \beta_1 x_{1i} + \ldots + \beta_m x_{mi})^2$$

หา $\beta_0, \beta_1, \ldots, \beta_m$ ที่ทำให้ SSE มีค่าน้อยที่สุด โดยการหาอนุพันธ์ของ SSE เทียบกับ β_j , $j = 1,2,\ldots,m$ แล้วให้เท่ากับ 0 จะได้สมการปกติ $m+1$ สมการ แก้สมการหาค่า $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_m$

จาก (11) จะได้ $$\hat{p}^* = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \ldots + \hat{\beta}_m x_{mi}$$

และจาก (10) จะได้ $$\hat{\pi}(\mathbf{x_i}) = \frac{e^{\hat{p}^*}}{1+e^{\hat{p}^*}}$$

<u>1. กรณีตัวแปรอธิบาย 1 ตัว</u>

$$\hat{\pi}(x_i) = \frac{e^{\hat{p}^*}}{1+e^{\hat{p}^*}}$$

เมื่อ $\hat{p}^* = \hat{\beta}_0 + \hat{\beta}_1 x_i$

หาค่า $\hat{\beta}_0$ และ $\hat{\beta}_1$ จากการทำให้ SSE มีค่าน้อยที่สุด โดยการหาอนุพันธ์เทียบกับ $\hat{\beta}_0$ และ $\hat{\beta}_1$ แล้วให้เท่ากับ 0 ได้สมการปกติ 2 สมการ คือ

$$\beta_0 \sum_{i=1}^{N} w_i + \beta_1 \sum_{i=1}^{N} w_i x_i \quad = \quad \sum_{i=1}^{N} w_i \tilde{p}_i$$
$$\beta_0 \sum_{i=1}^{N} w_i x_i + \beta_1 \sum_{i=1}^{N} w_i x_i^2 \quad = \quad \sum_{i=1}^{N} w_i x_i \tilde{p}_i$$

ซึ่งเขียนในรูปเมทริกซ์

$$(\mathbf{W}'\mathbf{W})\mathbf{B} \quad = \quad \mathbf{W}'\mathbf{P}$$

เมื่อ $\mathbf{W} = \begin{bmatrix} \sqrt{w_1} & \sqrt{w_1}x_1 \\ \sqrt{w_2} & \sqrt{w_2}x_2 \\ \vdots & \vdots \\ \sqrt{w_i} & \sqrt{w_i}x_i \\ \vdots & \vdots \\ \sqrt{w_N} & \sqrt{w_N}x_N \end{bmatrix}_{\mathrm{Nx2}}$, $\mathbf{P} = \begin{bmatrix} \sqrt{w_1}\tilde{p}_1 \\ \sqrt{w_2}\tilde{p}_2 \\ \vdots \\ \sqrt{w_i}\tilde{p}_i \\ \vdots \\ \sqrt{w_N}\tilde{p}_N \end{bmatrix}_{\mathrm{Nx1}}$ และ $\mathbf{B} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$

ดังนั้นหา $\hat{\mathbf{B}}$ ได้จาก $\hat{\mathbf{B}} \quad = \quad (\mathbf{W}'\mathbf{W})^{-1}(\mathbf{W}'\mathbf{P})$

คำนวนหาค่าพยากรณ์ $\hat{\pi}(x_i)$ จาก $\hat{\pi}(x_i) \quad = \quad \dfrac{e^{\hat{\beta}_0+\hat{\beta}_1 x_i}}{1+e^{\hat{\beta}_0+\hat{\beta}_1 x_i}}$

เมื่อ $\hat{\pi}(x_i)$ คือ ความน่าจะเป็นที่ $y=1$ ณ ระดับ x_i

<u>2. กรณีตัวแปรอธิบาย 2 ตัว</u>

$$\hat{\pi}(\mathbf{x_i}) \quad = \quad \frac{e^{\hat{p}^*}}{1+e^{\hat{p}^*}}$$

เมื่อ $\hat{p}^* \ = \ \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i}$

หาค่า $\hat{\beta}_0$ $\hat{\beta}_1$ และ $\hat{\beta}_2$ จากการทำให้ SSE มีค่าน้อยที่สุด โดยการหาอนุพันธ์เทียบกับ $\hat{\beta}_0$ $\hat{\beta}_1$ และ $\hat{\beta}_2$ แล้วให้เท่ากับ 0 ได้สมการปกติ 3 สมการ คือ

$$\beta_0 \sum_{i=1}^{N} w_i + \beta_1 \sum_{i=1}^{N} w_i x_{1i} + \beta_2 \sum_{i=1}^{N} w_i x_{2i} \quad = \quad \sum_{i=1}^{N} w_i \tilde{p}_i$$
$$\beta_0 \sum_{i=1}^{N} w_i x_{1i} + \beta_1 \sum_{i=1}^{N} w_i x_{1i}^2 + \beta_2 \sum_{i=1}^{N} w_i x_{1i} x_{2i} \quad = \quad \sum_{i=1}^{N} w_i x_{1i} \tilde{p}_i$$

$$\beta_0 \sum_{i=1}^{N} w_i x_{2i} + \beta_1 \sum_{i=1}^{N} w_i x_{2i} x_{1i} + \beta_2 \sum_{i=1}^{N} w_i x_{2i}^2 \quad = \quad \sum_{i=1}^{N} w_i x_{2i} \tilde{p}_i$$

ซึ่งเขียนในรูปเมทริกซ์

$$(\mathbf{W}'\mathbf{W})\mathbf{B} \quad = \quad \mathbf{W}'\mathbf{P}$$

เมื่อ $\mathbf{W} = \begin{bmatrix} \sqrt{w_1} & \sqrt{w_1}x_{11} & \sqrt{w_1}x_{21} \\ \sqrt{w_2} & \sqrt{w_2}x_{12} & \sqrt{w_2}x_{22} \\ \vdots & \vdots & \vdots \\ \sqrt{w_i} & \sqrt{w_i}x_{1i} & \sqrt{w_i}x_{2i} \\ \vdots & \vdots & \vdots \\ \sqrt{w_N} & \sqrt{w_N}x_{1N} & \sqrt{w_N}x_{2N} \end{bmatrix}_{\text{Nx3}}$, $\mathbf{P} = \begin{bmatrix} \sqrt{w_1}\tilde{p}_1 \\ \sqrt{w_2}\tilde{p}_2 \\ \vdots \\ \sqrt{w_i}\tilde{p}_i \\ \vdots \\ \sqrt{w_N}\tilde{p}_N \end{bmatrix}_{\text{Nx1}}$ และ $\mathbf{B} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$

ดังนั้นหา $\hat{\mathbf{B}}$ ได้จาก $\hat{\mathbf{B}} \quad = \quad (\mathbf{W}'\mathbf{W})^{-1}(\mathbf{W}'\mathbf{P})$

คำนวนหาค่าพยากรณ์ $\hat{\pi}(\mathbf{x}_\mathbf{i})$ จาก $\hat{\pi}(\mathbf{x}_\mathbf{i}) \quad = \quad \dfrac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i}}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i}}}$

เมื่อ $\hat{\pi}(\mathbf{x}_\mathbf{i})$ คือ ความน่าจะเป็นที่ $y = 1$ ณ ระดับ x_1 และ x_2

ตัวสถิติ DEVIANCE

Nelder และ *Wedderburn* (1972) แนะนำตัวสถิติ Deviance สำหรับใช้กับสมการถดถอยโลจิสติค Deviance มีบทบาทเหมือนผลรวมของความคลาดเคลื่อนกำลังสอง (Sum of Squares Error (SSE)) และ เกี่ยวข้องกับ ลอการิทึมธรรมชาติของฟังก์ชันภาวะน่าจะเป็น

$$L(\mathbf{B}) \quad = \quad \ln(l(\mathbf{B})) \quad = \quad \ln\left(\prod_{i=1}^{N} \left(\pi_1(\mathbf{X}_i)\right)^{y_i} \left(1 - \pi_1(\mathbf{X}_i)\right)^{1-y_i}\right)$$

$$= \quad \sum_{i=1}^{N} \left\{ y_i \ln(\pi(\mathbf{x}_\mathbf{i})) + (1 - y_i)\ln(1 - \pi(\mathbf{x}_\mathbf{i})) \right\}$$

ซึ่งเป็นฟังก์ชันที่รวบรวมข้อสนเทศของข้อมูลเกี่ยวกับตัวพารามิเตอร์ที่ไม่ทราบค่าไว้ เรียกค่าฟังก์ชันภาวะน่าจะเป็นที่แทนค่าพารามิเตอร์ด้วยตัวประมาณ $\hat{\mathbf{B}}$ ว่า ภาวะน่าจะเป็นสูงสุดภายใต้ตัวแบบปัจจุบัน เขียนแทนด้วย $\hat{L}_c$ ค่านี้ไม่เป็นอิสระจากจำนวนค่าสังเกตของข้อมูล จำเป็นต้องนำมาเปรียบเทียบกับ ภาวะน่าจะเป็นสูงสุดของตัวแบบของค่าสังเกตทั้งหมด ซึ่งเรียกว่า

ตัวแบบเต็มรูป เขียนแทนด้วย $\hat{L}_f$

ในการเปรียบเทียบ $\hat{L}_c$ กับ $\hat{L}_f$ เพื่อความเหมาะสม และเป็นไปได้ทางคณิตศาสตร์ ให้นิยามความแตกต่างเป็น Deviance, D

$$D = -2\ln\left(\frac{\hat{L}_c}{\hat{L}_f}\right)$$

$$= -2\left\{\ln \hat{L}_c - \ln \hat{L}_f\right\}$$

D จะมีค่าใหญ่ ถ้า $\hat{L}_c$ มีค่าน้อยกว่า $\hat{L}_f$ นี่แสดงว่าตัวแบบปัจจุบันไม่เหมาะสม ในทางกลับกัน D จะมีค่าเล็ก ถ้า $\hat{L}_c$ ไม่ต่างจาก $\hat{L}_f$ ซึ่งเป็นการบอกว่าตัวแบบปัจจุบันเป็นตัวแบบที่ดี

ภายใต้ตัวแบบปัจจุบัน

$$\hat{L}_c = \sum_{i=1}^{N}\left\{y_i \ln(\pi(\mathbf{x_i})) + (1-y_i)\ln(1-\hat{\pi}(\mathbf{x_i}))\right\}$$

ภายใต้ตัวแบบเต็มรูป

$$\hat{L}_f = \sum_{i=1}^{N}\left\{y_i \ln(\tilde{p}_i) + (1-y_i)\ln(1-\tilde{p}_i)\right\}$$

ซึ่ง $\tilde{p}_i = y_i$ ทำให้ได้ $y_i \ln y_i$ และ $(1-y_i)\ln(1-y_i)$ มีค่าเป็น 0 ทั้งคู่ สำหรับค่าที่เป็นไปได้ของ $y_i = 0$ และ 1 ดังนั้น $\hat{L}_f = 0$

ดังนั้น $$D = -2\sum_{i=1}^{N}\left\{y_i \ln(\hat{\pi}(\mathbf{x_i})) + (1-y_i)\ln(1-\hat{\pi}(\mathbf{x_i}))\right\}$$

$$= -2\sum_{i=1}^{N}\left\{y_i \ln(\hat{\pi}(\mathbf{x_i})) + \ln(1-\hat{\pi}(\mathbf{x_i})) - y_i\ln(1-\hat{\pi}(\mathbf{x_i}))\right\}$$

$$= -2\sum_{i=1}^{N}\left\{y_i \ln\left(\frac{\hat{\pi}(\mathbf{x_i})}{1-\hat{\pi}(\mathbf{x_i})}\right) + \ln(1-\hat{\pi}(\mathbf{x_i}))\right\}$$

$$= -2\sum_{i=1}^{N}\left\{y_i \,\mathrm{logit}(\hat{\pi}(\mathbf{x_i})) + \ln(1-\hat{\pi}(\mathbf{x_i}))\right\}$$

สามารถพิสูจน์ได้ว่า $\sum_{i=1}^{N} y_i \,\mathrm{logit}(\hat{\pi}(\mathbf{x_i})) = \sum_{i=1}^{N} \hat{\pi}(\mathbf{x_i})\log(\hat{\pi}(\mathbf{x_i}))$

ดังนั้น $$D = -2\sum_{i=1}^{N}\left\{\hat{\pi}(\mathbf{x_i})\,\mathrm{logit}(\hat{\pi}(\mathbf{x_i})) + \ln(1-\hat{\pi}(\mathbf{x_i}))\right\}$$

ต่อไปนี้คือ การพิสูจน์ย่อย

จาก (2)

$$L(\mathbf{B}) \;=\; \sum_{i=1}^{N}\left\{y_i\left(\beta_0+\beta_1 x_{1i}+\ldots+\beta_m x_{im}\right)-\ln\left(1+e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_m x_{im}}\right)\right\}$$

หาอนุพันธ์ของ $L(\mathbf{B})$ เทียบกับพารามิเตอร์ β_j ; $j = 0,1,2,\ldots,m$

$$\begin{aligned}\frac{\partial L(\mathbf{B})}{\partial \beta_j} &= \sum_{i=1}^{N}\left\{y_i(x_{ji})-\frac{1}{1+e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_m x_{im}}}(x_{ji}e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_m x_{im}})\right\}\\ &= \sum_{i=1}^{N}\left\{\left\{y_i-\frac{e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_m x_{im}}}{1+e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_m x_{im}}}\right\}(x_{ji})\right\}\\ &= \sum_{i=1}^{N}\left\{\left\{y_i-\pi(\mathbf{x_i})\right\}x_{ji}\right\}\end{aligned}$$

ดังนั้น

$$\beta_j\frac{\partial L(\mathbf{B})}{\partial \beta_j} = \sum_{i=1}^{N}\left\{\left\{y_i-\pi(\mathbf{x_i})\right\}\beta_j x_{ji}\right\}$$

$$\sum_{j=0}^{m}\beta_j\frac{\partial L(\mathbf{B})}{\partial \beta_j} = \sum_{i=1}^{N}\left\{\left\{y_i-\pi(\mathbf{x_i})\right\}\sum_{j=0}^{m}\beta_j x_{ji}\right\} \quad , \quad x_{01} = 1$$

$$= \sum_{i=1}^{N}\left\{\left\{y_i-\pi(\mathbf{x_i})\right\}\operatorname{logit}(\pi(\mathbf{x_i}))\right\}$$

เพราะว่า $\hat{\mathbf{B}}$ เป็นค่าประมาณภาวะน่าจะเป็นสูงสุดของ $\mathbf{B}$ อนุพันธ์ทางซ้ายมือของสมการนี้ ณ $\hat{\mathbf{B}}$ มีค่าเท่ากับ 0 ดังนั้นความน่าจะเป็น $\hat{\pi}(\mathbf{X}_i)$ ต้องสอดคล้องกับ

$$\sum_{i=1}^{N}\left\{\left\{y_i-\hat{\pi}(\mathbf{x_i})\right\}\operatorname{logit}(\hat{\pi}(\mathbf{x_i}))\right\} = 0$$

นั้นคือ
$$\sum_{i=1}^{N} y_i \operatorname{logit}(\hat{\pi}(\mathbf{x_i})) = \sum_{i=1}^{N}\hat{\pi}(\mathbf{x_i})\operatorname{logit}(\hat{\pi}(\mathbf{x_i}))$$

McCullagh และ *Nelder* (1983) กล่าวว่า Deviance เป็นตัวสถิติไลค์ลิฮูดเรโช ผลการทดสอบไลค์ลิฮูดเรโช Deviance มีการแจกแจงใกล้การแจกแจงไคสแคว์ (Asymtotic Chi-Square) ชั้นความอิสระ (N - m)

ผลงานวิจัยที่เกี่ยวข้อง

Efron (1975) ทำวิจัยเปรียบเทียบประสิทธิภาพของการถดถอยโลจิสติค โดยวิธีภาวะน่าจะเป็นสูงสุด(ML) กับ การวิเคราะห์จำแนกประเภท(DA) สำหรับตัวแปรอธิบายที่มีการแจกแจงปกติ พบว่า ML ให้ค่าพารามิเตอร์เช่นเดียวกับ DA งานวิจัยคำนวนค่า asymtotic relative efficiency ของทั้งสองวิธี ค่าสถิติของพารามิเตอร์ ML มีประสิทธิภาพ 1 ใน 2 และ 2 ใน 3 ของ DA

O' Hara, *Hosmer*, *Lemeshow* และ *Hartz* (1980) ทำการวิจัยโดยใช้ข้อมูลจริงเกี่ยวกับการเป็นโรคระบาด โดยจัดข้อมูลชนิดแยกประเภท(categorical data) ในรูปตารางการจร(contingency table) 2×3 และ 2×3^3 ข้อมูลมีทั้งโรคทั่วไป และโรคหายาก ซึ่งเป็นสถานการณ์ความสัมพันธ์อย่างอ่อนและอย่างแข็งระหว่างตัวทำนาย x และตัวพยากรณ์ y หาค่าพารามิเตอร์โดยวิธีฟังก์ชันจำแนกประเภท(DF) และ ML และเปรียบเทียบผลด้วยค่าสถิติ ค่าความเอนเอียงสัมพัทธ์ (Relative Bias) และ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ (Relative Mean Squares Error) ผลการวิจัยมีหลักฐานพอที่จะสนับสนุนว่าวิธี ML ให้ผลดีกว่า

Albert และ *Anderson* (1983) กล่าวถึง การหาตัวประมาณ ML บางครั้งหาไม่ได้ ซึ่ง *Mccullagh*(1986) ยืนยันเรื่องนี้ ท่านทั้งสองเขียนทฤษฎีเกี่ยวกับ The Existence of Maximum Likelihood Estimates in Logistic Regression Model ทฤษฎีหนึ่งกล่าวว่า ถ้าข้อมูลมีจุดข้อมูลของกลุ่มหนึ่งเกินไปอยู่ในอีกกลุ่มหนึ่ง (overlap) จะได้ว่าตัวประมาณ ML หาได้และมีค่าเดียว

Brenn และ *Arnesen* (1985) ศึกษาเปรียบเทียบคุณค่าของวิธี DA, วิธี ML และวิธี Cox's regression model (CR) โดยทำวิจัยเรื่องการเลือกปัจจัยเสี่ยงของการเสียชีวิตด้วยโรคหัวใจของชาย 6595 คน อายุ 20-49 ปี ติดตามผล 9 ปี มีตัวแปรปัจจัยให้เลือก 10 ตัวแปร คำนวนด้วย ซอฟแวร์ BMDP บนเครื่อง CYBER 171 ด้วยหน่วยความจำ 1.96 Mbyte พบว่าเวลาที่ใช้ในการคำนวนของโปรแกรม ML และ CR ยาวมากเป็น 2 เท่า หรือมากกว่าของ DA ส่วนหนึ่งของรายงานชิ้นนี้กล่าวว่า วิธี ML และ CR เลือกได้กลุ่มปัจจัยเดียวกัน แต่มีบางกลุ่มแตกต่างกันบ้าง หนึ่งหรือสองปัจจัย, วิธี DA แยกตัวแปรปัจจัยเสี่ยงได้ดี ทั้งๆ ที่ข้อมูลไม่มีคุณสมบัติการแจกแจงปกติ แต่อย่างไรก็ตามก็มีค่าพารามิเตอร์บางค่าผิดพลาด เกือบทุกกรณี วิธี DA ให้ค่าพารามิเตอร์ใกล้เคียงกับของ วิธี ML และ CR มีบางค่าเล็กกว่า บางค่าสูงกว่าพารามิเตอร์ของ วิธี ML และ CR ไม่มีความแตกต่างทางสถิติ