[Binary case] $Y \sim \text{Binomial}(n, \pi)$

$$
\begin{array}{ll}
1 & (n_1, y_1, x_1) \\
2 & \vdots \\
\vdots & \\
N & (n_N, y_N, x_N)
\end{array}
$$

Def) $g(\pi) = \log \frac{\pi}{1-\pi}$ (logit)

Model : $\pi_i = \dfrac{e^{\beta^T x_i}}{1 + e^{\beta^T x_i}}$

$\Rightarrow$ We can calculate $l, \frac{dl}{d\beta}, \mathcal{J}$

$\Rightarrow$ Get MLE from $X^T W X \beta = X^T W z$

Example 5.6.1

$$D = 2\left[\ell(\hat{\beta}_{max}) - \ell(\hat{\beta}_{model})\right]$$

$$\ell(\beta) = \log \prod \binom{n_i}{y_i} \pi_i^{y_i} (1-\pi_i)^{n_i - y_i}$$

$$= \sum y_i \log \pi_i + (n_i - y_i)\log(1-\pi_i) + \log\binom{n_i}{y_i}$$

$$= 2\left[\sum y_i \log \frac{y_i}{\hat{y}_i} + (n_i - y_i) \log \frac{n_i - y_i}{n_i - \hat{y}_i}\right]$$

$$\hat{\beta}_{max} = y_i$$

$$\hat{\beta}_{model} = \hat{\pi}_i \Rightarrow \hat{y}_i = n_i \hat{\pi}_i$$

Observation: D has log terms, so it may lead to over/under flow problem.

Suggestion: $X^2 = \sum \frac{(y_i - n_i\hat{\pi}_i)^2}{n_i\hat{\pi}_i(1-\hat{\pi}_i)}$

$\overset{D}{\sim} X^2(N-p)$

<Outline>

$$D = 2\sum y_i \log \frac{y_i}{n_i\hat{\pi}_i} + (n_i - y_i)\log \frac{n_i - y_i}{n_i - n_i\hat{\pi}_i}$$

$\hat{y}_i = n_i\hat{\pi}_i$

Note that $f(s) = s\log\frac{s}{t}$

$$\approx f(t) + f'(t)(s-t) + \frac{f''(t)}{2}(s-t)^2$$

$$= (s-t) + \frac{1}{2t}(s-t)^2$$

$$f'(s) = \log\frac{s}{t} + s\frac{\frac{1}{t}}{\frac{s}{t}} = \log\frac{s}{t} + 1$$

$$f''(s) = \frac{\frac{1}{t}}{\frac{s}{t}} = \frac{1}{s}$$

$$= \cancel{2}\sum (\cancel{y_i} - \cancel{n_i\hat{\pi}_i}) + \frac{(y_i - n_i\hat{\pi}_i)^2}{\cancel{2}\, n_i\hat{\pi}_i}$$

$$+ (\cancel{n_i} - \cancel{y_i} - \cancel{n_i} + \cancel{n_i\hat{\pi}_i}) + \frac{(\cancel{n_i} - y_i - \cancel{n_i} + n_i\hat{\pi}_i)^2}{\cancel{2}(n_i - n_i\hat{\pi}_i)}$$

$$= \sum \frac{1}{n_i \hat{\pi}_i (1-\hat{\pi}_i)}$$

$$\left[(1-\hat{\pi}_i)(y_i - n_i\hat{\pi}_i)^2 + \hat{\pi}_i (y_i - n_i \hat{\pi}_i)^2 \right]$$

$$= \sum \frac{(y_i - n_i\hat{\pi}_i)^2}{n_i \hat{\pi}_i (1-\hat{\pi}_i)} = X^2$$

Because $D \overset{D}{\sim} \chi^2(N-p)$,

$$\Rightarrow X^2 \overset{D}{\sim} \chi^2(N-p) \quad ///$$

Recall) $S \overset{\text{def}}{=} e^T e = (y - X\beta)^T (y - X\beta)$

$$\Rightarrow S_{model} \underset{\uparrow}{=} (y - X\hat{\beta})^T (y - X\hat{\beta})$$

$\hat{\beta}$ is OLS

$$= y^T y - y^T X\hat{\beta} - \hat{\beta}^T X^T y + \underbrace{\hat{\beta}^T X^T X \hat{\beta}}_{\hat{\beta}^T X^T y}$$

$$\underset{\uparrow}{=} y^T y - \hat{\beta}^T X^T y$$

$\hat{\beta} = (X^T X)^{-1} X^T y$

$$S_{null} \underset{\uparrow}{=} (y - 1\bar{y})^T (y - 1\bar{y})$$

$\hat{\beta} = \bar{y}$

$$= \sum (y_i - \bar{y})^2 = \sum y_i^2 - N\bar{y}^2$$

$$R^2 = \frac{S_{null} - S_{model}}{S_{null}}$$

$$\text{Pseudo } R^2 = \frac{l(\hat{\pi}_{null}) - l(\hat{\pi}_{model})}{l(\hat{\pi}_{null})}$$

where $\hat{\pi}_{null} = \frac{\sum y_i}{\sum n_i}$

Defs) $AIC = -2l(\hat{\pi}) + 2p$

$BIC = -2l(\hat{\pi}) + p\ln(\#\text{data})$

(낮을수록 좋은 모델이다.)

Defs [Residuals]

Recall) $\hat{e}_i = y_i - \hat{y}_i$

$$\Rightarrow \quad r_i = \frac{\hat{e}_i}{\hat{\sigma}(1-h_i)^{1/2}}$$

(where $h_i = \text{diag}\, X(X^T X)^{-1} X)_{ii}$)

$$X_i = \frac{y_i - n_i \hat{\pi}_i}{\sqrt{n_i \hat{\pi}_i (1-\hat{\pi}_i)}} \quad (\Rightarrow X^2 = \sum X_i^2)$$

$$r_{pi} = \frac{X_i}{(1-h_i)^{1/2}} \quad \text{(standardization)}$$

or

$$d_i = \text{sign}(y_i - n_i \hat{\pi}_i) \left[2\, y_i \log \frac{y_i}{\hat{y}_i} + (n_i - y_i) \log \frac{n_i - y_i}{n_i - \hat{y}_i} \right]^{1/2}$$

$$r_{Di} = \frac{d_i}{(1-h_i)^{1/2}} \quad (\text{Std})$$