Note on the result (2.36)

$$W = \cancel{2}\left(-\cancel{\frac{n}{2}}\hat{\sigma}^2 - \frac{D(\hat{\beta})}{\cancel{2}\hat{\sigma}^2} + \cancel{\frac{n}{2}}\hat{\sigma}^2 + \frac{D(\hat{\beta})}{\cancel{2}\hat{\sigma}^2} \right)$$

$$= \frac{-D(\hat{\beta}) + D(\hat{\beta}_0)}{\hat{\sigma}^2}$$

$$= \frac{-\|y - \hat{y}\|^2 + \|y - \hat{y}_0\|}{\hat{\sigma}^2}$$

$$= \frac{-\cancel{\|y\|} + \|\hat{y}\| + \cancel{\|y\|} - \|\hat{y}_0\|}{\hat{\sigma}^2} = \frac{\|\hat{y} - \hat{y}_0\|}{\hat{\sigma}^2}$$

$$F = \frac{\left(D(\hat{\beta}_0) - D(\hat{\beta})\right)/q}{D(\hat{\beta})/(n-p)} \sim F_{q,n-p}$$

Likelihood quantities in the case of binomial distr.

$Y_i \in \{0,1\}$ $\quad Y_i \sim Bi(1,\pi)$ Bernulli distrib.

$Y = \sum_{i=1}^{n} Y_i \sim Bi(n,\pi)$ $\quad \pi$ = probability of success

$\hookrightarrow$ # successes in n trials

$$P_Y(y;\pi) = \binom{n}{y}\pi^y(1-\pi)^{n-y}$$

$$\ell(\pi) = y\log\pi + (n-y)\log(1-\pi)$$

$$\ell_\pi(\pi) = \frac{y}{\pi} + \frac{(n-y)}{1-\pi}(-1)$$

$$\frac{y}{\pi} - \frac{n-y}{1-\pi} = 0 \qquad \frac{y - \cancel{y\pi} - n\pi + \cancel{y\pi}}{\pi(1-\pi)} = 0$$

$$\Leftrightarrow \hat{\pi} = \frac{y}{n}$$

$$\ell_{\pi\pi}(\pi) = -\frac{y}{\pi^2} - \frac{n-y}{(1-\pi)^2}$$

$$j(\pi) = \frac{y}{\pi^2} + \frac{n-y}{(1-\pi)^2}$$

$$i(\pi) = E\left[\frac{Y}{\pi^2} + \frac{n-Y}{(1-\pi)^2}\right] = \frac{n\cancel{\pi}}{\pi^{\cancel{2}}} + \frac{n-n\pi}{(1-\pi)^2}$$

$$= \frac{n}{\pi} + \frac{n\cancel{(1-\pi)}}{(1-\pi)^{\cancel{2}}}$$

$$= \frac{n - \cancel{n\pi} + \cancel{n\pi}}{\pi(1-\pi)} = \frac{n}{\pi(1-\pi)}$$

$$se(\hat{\pi}) = \sqrt{i^{-1}(\pi)}\,\Big|_{\pi=\hat{\pi}} = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$

Comparison of two groups

Brasilian bank example

X = age

	young (1)	old (2)	tot
low	84	34	118
high	225 y_1	157 y_2	382
tot	309 n_1	191 n_2	500

Y = satisfaction

dbinom(y_1, n_1, $\hat{\pi}_1$, log=TRUE)

$$\ell(\hat{\pi}_1, \hat{\pi}_2) = c + y_1 \log \hat{\pi}_1 + (n_1 - y_1)\log(1-\hat{\pi}_1) + \underbrace{y_2 \log \hat{\pi}_2 + (n_2 - y_2)\log(1-\hat{\pi}_2)}$$

dbinom(y_2, n_2, $\hat{\pi}_2$, log=TRUE)

$$\hat{\pi}_1 = \frac{y_1}{n_1} = \frac{225}{309} \approx 0.728$$

$$\hat{\pi}_2 = \frac{y_2}{n_2} = \frac{157}{191} \approx 0.822$$

$$H_0: \pi_1 = \pi_2 \iff \pi_1 - \pi_2 = 0$$

$$H_1: \pi_1 - \pi_2 \neq 0$$

$$\pi_1 = \pi_2 = \pi$$

$$w = 2\left(\ell(\hat{\pi}_1, \hat{\pi}_2) - \ell(\hat{\pi}, \hat{\pi})\right)$$

$$\hat{\pi} = \frac{y_1 + y_2}{n_1 + n_2}$$

$$D_a = -2\,\ell(\hat{\pi}_a)$$

$$D = D_0 - D_1$$

Logistic regression

Y is dichotomous (two possible values)

we cannot use Gaussian regression → extension

↘ probability of success depends on the covariates

$$\hat{\pi}(x) = \frac{e^{\eta(x)}}{1 + e^{\eta(x)}}$$

$\hat{\pi}_{young}$

where $\eta(x) = \beta_0 + \beta_1 \overset{1 \text{ for young}}{x_1} + \dots + \beta_p x_{p-1}$

$$\frac{e^{\eta(x)}}{1 + e^{\eta(x)}} = \text{logistic function}$$

its inverse $\hat{\pi} = \dfrac{e^{\eta}}{1 + e^{\eta}}$ $\qquad \pi = \dfrac{\text{odds}}{1 + \text{odds}}$

$$\hat{\pi} + \hat{\pi} e^{\eta} = e^{\eta}$$

$$e^{\eta}(1 - \hat{\pi}) = \hat{\pi}$$

$$e^{\eta} = \frac{\hat{\pi}}{1 - \hat{\pi}} \quad \text{odds}$$

$$\eta(x) = \log\left(\frac{\hat{\pi}}{1 - \hat{\pi}}\right) = \text{logit}(\hat{\pi})$$

In broad generality,

$$g(E[Y|x]) = \eta(x) = X\beta \qquad \text{GLM}$$

g ↖ link function

logistic regression $g = \text{logit}$

$p(Y|x)$ belongs to the exponential family

- Gaussian
- gamma } continuous Y
- binomial } binary Y
- Poisson
- inverse binomial } counts Y
- inverse Gaussian