TMA4315 H3w2
BINARY REGRESSION

Parameter estimation

$$S(\hat{\beta}) = 0 \qquad S(\beta) = \sum_{i=1}^{n} \underbrace{x_i}_{p\times 1} (y_i - \pi_i)$$

$$\pi_i = \frac{\exp(x_i^T\beta)}{1+\exp(x_i^T\beta)}$$

$$S(\hat{\beta}) \approx S(\beta^{(0)}) + \underbrace{\frac{\partial S(\beta)}{\partial \beta^T}\Bigg|_{\beta=\beta^{(0)}}}_{-H(\beta^{(0)})} (\hat{\beta} - \beta^{(0)})$$

1st order Taylor expansion

$-H(\beta^{(0)})$ observed Fisher information

$S(\hat{\beta}) = 0$ solve for $\hat{\beta}$

$$S(\beta^{(0)}) - H(\beta^{(0)})(\hat{\beta} - \beta^{(0)}) = 0$$

$$H(\beta^{(0)})\hat{\beta} = S(\beta^{(0)}) + H(\beta^{(0)}) \cdot \beta^{(0)}$$

$$\hat{\beta} = \beta^{(0)} + H^{-1}(\beta^{(0)}) S(\beta^{(0)})$$

$$\underbrace{\beta^{(t+1)}}_{p\times 1} = \underbrace{\beta^{(t)}}_{p\times 1} + \underbrace{[H(\beta^{(t)})]^{-1}}_{p\times p} \underbrace{S(\beta^{(t)})}_{p\times 1}$$ Newton Raphson method

$$H(\beta) = \cdots = \sum_{i=1}^{n} x_i x_i^T \pi_i (1-\pi_i)$$

expected Fisher info

Fisher scoring: replace $H(\beta)$ with $F(\beta) = E(H(\beta))$

AND: $F(\beta) = E(H(\beta)) = \boxed{E\left(-\frac{\partial^2 l}{\partial\beta\partial\beta^T}\right) = E\left(\frac{\partial l}{\partial \beta}\cdot\left(\frac{\partial l}{\partial\beta}\right)^T\right)}$

useful

$= Cov(S(\beta))$

Properties of $S(\beta)$

$S(\beta)$ is a function of Y_i's and therefore a random vector

$$S(\beta) = \sum_{i=1}^{n} (Y_i - \pi_i)\, x_i$$

$$E(S(\beta)) = E\left(\sum_{i=1}^{n} (Y_i - \pi_i) x_i\right) = \sum_{i=1}^{n} \overbrace{\left(\underbrace{E(Y_i)}_{\pi_i} - \pi_i\right) x_i}^{E(s_i(\beta))} = 0$$

$$E(s_i(\beta)) = 0$$

$$\text{Cov}(S(\beta)) = \text{Cov}\left(\sum_{i=1}^{n} s_i(\beta)\right) = \sum_{i=1}^{n} \underbrace{\text{Cov}(s_i(\beta))}$$

Independent pairs (x_i, Y_i)

$$\text{Cov}(Z) = E\left[(Z-\mu_Z)(Z-\mu_Z)^T\right]$$

$$= \sum_{i=1}^{n} E\left(\left(s_i(\beta) - \underbrace{E(s_i(\beta))}_{0}\right)\left(s_i(\beta) - \underbrace{E(s_i(\beta))}_{0}\right)^T\right)$$

$$= \sum_{i=1}^{n} \underbrace{E\left(s_i(\beta)\, s_i(\beta)^T\right)}_{F_i(\beta)} = \sum_{i=1}^{n} F_i(\beta)$$

$$F_i(\beta) = E\left(s_i(\beta) \cdot s_i(\beta)^T\right) = E\left((Y_i - \pi_i) x_i \, (Y_i - \pi_i) x_i^T\right)$$

$$= \underbrace{E\left((Y_i - \pi_i)^2\right)}_{\text{Var}(Y_i)} x_i x_i^T = \pi_i (1-\pi_i) \cdot x_i x_i^T$$

$$F(\beta) = \sum_{i=1}^{n} \pi_i (1-\pi_i)\, x_i x_i^T$$

Canonical link

$Y_i \sim$ univ. exp. family θ_i, ϕ

param of interest

$$\theta_i \xrightarrow{b'(\theta_i)} \mu_i \xrightarrow{g} \eta_i$$

canonical link

$\theta_i = \eta_i$

$$\theta_i = \eta_i$$
$$\|$$
$$\theta_i = g(\mu_i)$$

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

logit link: $\underbrace{g(\mu_i)}_{= \eta_i} = g(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \theta_i$

For GLM's with canonical link → nice things happen!

$F(\beta) = H(\beta)$

Asymptotic properties of $\hat{\beta}$

$$\hat{\beta} \approx N_p\left(\beta, F^{-1}(\hat{\beta})\right)$$

$$F(\beta) = \sum_{i=1}^{n} \pi_i (1-\pi_i)\, x_i x_i^T = X^T \underset{\substack{\uparrow \\ \text{diag}(\pi_i(1-\pi_i))}}{W} X$$

$$\pi_i = \frac{e^{\eta_i}}{1+e^{\eta_i}}, \qquad \eta_i = x_i^T \beta$$

so $\quad Cov(\hat{\beta}) = F^{-1}(\hat{\beta}) = (X^T W X)^{-1}$

and we use $\quad \widehat{Cov}(\hat{\beta}) = (X^T \hat{W} X)^{-1}$

$$\hat{W} = \text{diag}\left(\hat{\pi}_i (1-\hat{\pi}_i)\right)$$

Confidence intervals and hypothesis tests: about β_j

$$\hat{\beta}_j \approx N\left(\beta_j, \hat{a}_{jj}(\hat{\beta})\right)$$

$\hat{a}_{jj}$: diagonal element (j,j) of $(X^T \hat{W} X)^{-1}$

For hypothesis testing the summary.glm gives the result of a Wald test.
As for MLR we may also do LRT, and a third possibility is a score test (more in M4).

Next: model assessment and model choice

Devience

unique coverete pattern

only use if grouped data: G groups, n_j obs in group j

$j = 1, \dots, G$ $\quad$ y_j successes

Saturated model: $\tilde{\pi}_j = \frac{y_j}{n_j}$

Candidate model: $\hat{\pi}_j = e^{x_j^T \hat{\beta}} / (1 + e^{x_j^T \hat{\beta}})$

$$D = -2\left(l(\hat{\pi}) - l(\tilde{\pi}) \right) = \dots =$$

$$2 \sum_{j=1}^{G} \left(y_j \cdot \ln\left(\frac{y_j}{n_j \hat{\pi}_j}\right) + (n_j - y_j) \cdot \ln\left(\frac{n_j - y_j}{n_j - n_j \hat{\pi}_j}\right)\right)$$

Asymptotically $D \sim \chi^2_{G-p}$

is used for model assessment

Model choice: use AIC

Overdispersion: Verience in groups $> n_j \hat{\pi}_j (1 - \hat{\pi}_j)$