M3: Binary regression (with the logit model)

week 3: 25.09.2017

Assumptions

1. $Y_j \sim \text{bin}(n_j, \pi_j)$
2. $\eta_j = x_j^T \beta$
3. $\eta_j = \ln\left(\frac{\pi_j}{1-\pi_j}\right)$ ← logit link

Aim: estimate parameters β

maximize to find maximum likelihood estimat

expected Fisher info matrix

$$\begin{tikzcd}
 & & & \mathrm{Cov}(s(\beta)) \arrow[r, phantom, "="] & F(\beta) \arrow[d, phantom, "\parallel"] \\
 & & & & E(H(\beta)) \\
L(\beta) \arrow[r, "\ln"] & l(\beta) \arrow[r, "\frac{\partial l}{\partial \beta}"] & s(\beta) \arrow[uur] \arrow[r, "-\frac{\partial}{\partial \beta^T}"] & H(\beta) \arrow[ur] &
\end{tikzcd}$$

$L(\beta)$: likelihood, joint distribution

$l(\beta)$: log-likelihood

$s(\beta)$: $p \times 1$

$H(\beta)$: $p \times p$ — -Hessian — observed Fisher info. matrix

$\hat{\beta}$: found from $s(\hat{\beta}) = 0$, by $\beta^{(t+1)} = \beta^{(t)} + F(\beta^{(t)})^{-1} s(\beta^{(t)})$

Fisher scoring / Newton Raphson

$$\hat\beta \approx N_p(\beta, \underbrace{F^{-1}(\hat\beta)}_{(X^T W X)^{-1}})$$

$$W = diag(n_j \pi_j (1-\pi_j))$$

$$\hat\beta_k \approx N(\beta_k, a_{kk}(\hat\beta))$$

$A = F^{-1}$, a_{kk} is element kk of A

Q: why $a_{kk}(\hat\beta)$ and not $a_{kk}(\beta_k)$?

Because $F(\hat\beta)$ is a function of the full $\hat\beta$ vector, and we take the inverse to get $\widehat{Cov}(\hat\beta_k)$.

Wald test

$$H_0: \underbrace{C}_{r\times p}\beta = d \quad vs. \quad H_1: C\beta \neq d$$

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

$$C\hat\beta \approx N_r(\underbrace{C\beta}_{d}, C F^{-1}(\hat\beta) C^T)$$

$$W = (C\hat\beta - d)^T [C F^{-1}(\hat\beta) C^T]^{-1} (C\hat\beta - d) \approx \underbrace{\chi^2_r}_{\text{when } H_0 \text{ is true}}$$

P-value: $P(\chi^2_r > W)$

Likelihood ratio test

A: larger model H_1 true $\Rightarrow \hat{\beta}_A$ with $L(\hat{\beta}_A)$

B: smaller model H_0 true $\Rightarrow \hat{\beta}_B$ $L(\hat{\beta}_B)$

Q: Can $L(\hat{\beta}_B) > L(\hat{\beta}_A)$? No, same as for SSE! MLR

$$LRT = -2\left(\ln L(\hat{\beta}_B) - \ln L(\hat{\beta}_A)\right)$$

$$\approx \chi^2_{\#par\,A - \#par\,B}$$

asymptotic when H_0 true $\leftarrow$ Stat. Inf TMA4295 (ch 10)

Deviance

G groups with n_j obs in group j

Saturated model: a perfect fit to each group

"imaginary model"

$$\tilde{\pi}_j = \frac{y_j}{n_j}$$

$$\frac{\exp(x_j^T \hat{\beta})}{1+\exp(x_j^T \hat{\beta})}$$

Candidate model: our model $\hat{\pi}_j = h(\eta(\hat{\beta}))$

$\hat{\beta}$ is our MLE

Deviance = LRT with A = saturated
B = Candidate

$$D = -2\left(\underset{\substack{\uparrow \\ \text{cand}}}{l(\hat{\pi})} - l(\tilde{\pi}) \right)$$

$$\approx \chi^2_{G-p}$$

Used to evaluate if a model is "good".