STA 303 / 1002

Note Title 2/16/2012

TEST: Monday Feb. 27 EX 310 / 320

All the details are on Blackboard

See Blackboard for office hours next week (announcements)

COME EARLY !!

Assignment 1 marks will be on Blackboard sometime tomorrow.

Last: Binomial logistic regression model

Deviance GOF test compares fitted model to saturated model

Test Statistic: $\hat{y}_i$ is binomial count est'd from fitted model

$$G^2 = 2\sum_{i=1}^{n}\left\{y_i \log(y_i) - y_i \log(\hat{y}_i) + (m_i - y_i)\log(m_i - y_i) - (m_i - y_i)\log(m_i - \hat{y}_i)\right\}$$

Why is there no Deviance GOF test in the Binary logistic regression case?

In Binary case, $m_i = 1$ (all i) and $y_i = \begin{cases} 0 \\ 1 \end{cases}$

G^2 becomes $2\sum_{i=1}^{n}\{-y_i \log(\hat{y}_i) - (n_i - y_i)\log(n_i - \hat{y}_i)\}$

The terms that came from $\log L_S$ are gone so G^2 no longer includes both L_M & L_S

For model assessment in Binomial Logistic Regression:

1. Plot observed logits (logits of response proportions) versus quantitative explanatory variable to see if linear relationship is appropriate.
2. Deviance GOF test

③ Examine residuals

"Response residuals"

$$\frac{y_i}{m_i} - \hat{\pi}_{M,i}$$

$\frac{y_i}{m_i}$: observed proportion

$\hat{\pi}_{M,i}$: fitted proportion

- useful to look for outliers
- need to standardize them

2 choices:

① Pearson residual

$$P_{res,i} = \frac{y_i - n_i \hat{\pi}_{M,i}}{\sqrt{n_i \hat{\pi}_{M,i}(1 - \hat{\pi}_{M,i})}}$$

Σ est s.d. of Binomial dist

SAS: RESCHI

② Deviance Residual - defined so that the sum of the squares of the deviance residuals is the deviance (G^2)

$$D_{res,i} = \text{sign}\left(y_i - m_i \hat{\pi}_{M,i}\right) * \sqrt{2\left\{y_i \log\left(\frac{y_i}{m_i \hat{\pi}_{M,1}}\right) + (m_i - y_i)\log\left(\frac{m_i - y_i}{m_i - m_i \hat{\pi}_{M,i}}\right)\right\}}$$

SAS: RESDEV

Which standardized residual should you use?

- I usually look at both
- Both are asymptotically Normally distributed (standard)

- if $|Pres_i|$ or $|Dres_i| > 2$, possibly an outlier, particularly if a gap between the residual and the next larger residual
- if $|Pres_i|$ or $|Dres_i| > 3$, classify as outlier

- For small sample sizes, the deviance residuals are closer to normally distributed than the Pearson residuals
- Pearson residuals are unstable when $\hat{\pi}$ close to 0 or 1

Looking at both Pearson and Deviance residuals for Krunnit example, all are < 2 in absolute value, so no outliers