STA 303 | 1002

Note Title 3/22/2012

I posted (in announcement on Bb) another summer research opportunity.

<u>Reminder</u>: Chunyi has office hours tomorrow 2:00-4:00, 2W 107/109

<u>Loglinear Models</u>

What are conditions for inference to be valid?

- independent quantities being counted.
- need large enough sample sizes for MLE asymptotic tests to hold

  Rule-of-thumb: (most) $\hat{\mu}_{ijk} \geq 5$ all i, j, k

- Poisson distribution appropriate
  - variance is equal to the mean

    If not, get large deviance, "extra-Poisson variation"

$\frac{\text{Deviance}}{df}$ should be about 1

- Model fits the data – right form of explanatory variables
  – no outliers
  - Check for agreement of predicted and observed counts
    - If don't agree ⇒ large deviance

  Deviance goodness-of-fit test to detect this

__Checking Model Adequacy__

① Deviance G-O-F test

- comparison of fitted model to saturated model
Is form of model that I'm using adequate or do I need something more complicated?

Test stat: $G^2 = 2 \sum_k \sum_j \sum_i \left(y_{ijk} \log\left(\frac{y_{ijk}}{\hat{\mu}_{ijk}} \right) \right)$

Exercise: Show this is the test statistic (in Practice Problems)

2 hints: For saturated model: $\hat{\mu}_{ijk} = y_{ijk}$

$$\sum_k \sum_j \sum_i \hat{\mu}_{ijk} = n$$
$$= n \sum \sum \sum \hat{\pi}_{ijk}$$

$\underbrace{\qquad\qquad}_{=1}$

H_0: fitted model fits as well as saturated model

Under H_0, G^2 has a Chisquare distribution with df
$IJK - \#$ of parameters in fitted model

If small p-value, need a more complicated model than what was fit

Possible reasons:

- need additional explanatory variables
- outliers
- Poisson model not correct, $Var(Y_{ijk}) > \mu_{ijk}$

(2) Residuals – for detecting outliers

Raw residual: $y_{ijk} - \hat{\mu}_{ijk}$

Pearson residual: $\dfrac{y_{ijk} - \hat{\mu}_{ijk}}{\sqrt{\hat{\mu}_{ijk}}}$

Sum of squares of Pearson residuals gives Pearson chi-square test statistic (Analysis 2a)

Deviance residual: Sum of squares of deviance residuals gives test stat for deviance G-O-F test

$$\text{sign}(y_{ijk}-\hat{\mu}_{ijk})\sqrt{2y_{ijk}\log\left(\frac{y_{ijk}}{\hat{\mu}_{ijk}}\right) - y_{ijk} + \hat{\mu}_{ijk}}$$

For both Pearson & Deviance residuals, look for values >3 (>2 is small sample size)

Useful to studentize Pearson & Deviance residuals

- Pearson and Deviance are adjusted for fact that y_{ijk}'s have different variances but not for the fact that $Var(y_{ijk}-\hat{\mu}_{ijk}) \neq Var(y_{ijk})$

(<u>Idea</u>: $Var(Y_{ijk}-\hat{\mu}_{ijk}) = (1-h_{ii})\,Var(Y_{ijk})$)

(<u>Also</u> (not responsible for) there are Likelihood Residuals)

③ Extra-<u>Poisson</u> variance : Indication $\frac{\text{Deviance}}{df} > 1$

How much > 1 is important ?
Deviance G-O-F test is statistically significant.
If ruled out other problems.
- include a dispersion parameter
so $Var(Y_{ijk}) = \varphi \mu_{ijk}$

<u>or</u> - negative binomial regression

From type3 option:

LRTs for models with and without set of indicator variables for effect of interest.

- particularly useful if >2 categories in categorical explanatory variables

Example: Suppose 3 cat. variables

X - 2 categories i: 1, 2

Y - 2 categories j: 1, 2

Z - 3 categories k: 1, 2, 3

Fit Uniform association model: (XY, XZ, YZ)

Model: $\log(\mu_{ijk}) = \beta_0 + \beta_1 I_{X=1} + \beta_2 I_{Y=1}$

$$+ \beta_3 I_{Z=1} + \beta_4 I_{Z=2}$$

$$+ \beta_5 I_{X=1} * I_{Y=1}$$

$$+ \beta_6 I_{X=1} * I_{Z=1} + \beta_7 I_{X=1} * I_{Z=2}$$

$$+ \beta_8 I_{Y=1} * I_{Z=1} + \beta_9 I_{Y=1} * I_{Z=2}$$

__Test__ Do I need YZ interaction?

$\iff$ Does (XY, XZ) model fit just as well?

$H_0: \beta_8 = \beta_9 = 0$ vs H_a: at least 1 of β_8, β_9 is not 0

Type 3 option would give LRT for this