STAT475/675 TUT07

http://www.sfu.ca/~zza115/teaching.html
http://people.stat.sfu.ca/~joanh/stat475-675web.html

Zhiyang Zhou (zhiyang_zhou@sfu.ca)

2018-02-27

Recall: slice of a three-way contingency table

- Interested in pairs (X, Y), (X, Z) and (Y, Z)
  - consider a cross-sectional slice and apply techniques for two-way tables
    - two-way partial table
    - two-way marginal table
  - with $2 \times 2 \times K$ contingency table
    - conditional OR
    - marginal OR

Homogeneity of conditional ORs

- Breslow-Day Test: testing on the homogeneity of conditional ORs
  - $H_0 : \theta_{XY(1)} = \cdots = \theta_{XY(K)}$ vs $H_1 :$ otherwise
  - under H_0,

$$BD = \sum_{i,j,k} \frac{(N_{ijk} - \mathrm{E}_{H_0}(N_{ijk}))^2}{\mathrm{E}_{H_0}(N_{ijk})} \approx \chi^2(K-1)$$

- Mantel-Haenszel estimator: estimating the common conditional OR

$$\hat{\theta}_{XY,MH} = \frac{\sum_k n_{11k} n_{22k} / n_{++k}}{\sum_k n_{12k} n_{21k} / n_{++k}}$$

Testing on conditional independence of X and Y given Z

- $H_0 : \theta_{XY(1)} = \cdots = \theta_{XY(K)} = 1$ vs $H_1 :$ otherwise
- Cohran-Mantel-Haenszel Test: under H_0, with $N = n \gg 1$,

$$CMH = \frac{(\sum_k (N_{11k} - \mathrm{E}_{H_0}(N_{11k}))^2}{\sum_k \mathrm{var}(N_{11k})} \approx \chi^2(1)$$

Testing on independence of X, Y and Z

- $H_0 : \pi_{ijk} = \pi_{i++}\pi_{+j+}\pi_{++k}$ vs $H_1 :$ otherwise
- Pearson's χ^2-test: under H_0, with $n_{ijk} \geq 5$, $\hat{\mu}_{ijk} = \frac{N_{i++}N_{+j+}N_{++k}}{N^2}$ and $df = (I-1)(J-1)(K-1)$,

$$\chi^2 = \sum_{i,j,k} \frac{(N_{ijk} - \hat{\mu}_{ijk})^2}{\hat{\mu}_{ijk}} \approx \chi^2(df)$$

- LRT: under H_0, with $n_{ijk} \geq 5$, $\hat{\mu}_{ijk} = \frac{N_{i++}N_{+j+}N_{++k}}{N^2}$ and $df = (I-1)(J-1)(K-1)$,

$$G^2 = 2\sum_{i,j,k} N_{ijk} \ln \frac{N_{ijk}}{\hat{\mu}_{ijk}} \approx \chi^2(df)$$

Demo

The data included in "Arthritis" in package "vcd" is from Koch & Edwards (1988) after a double-blind clinical trial investigating a new treatment for rheumatoid arthritis.

- Check the homogeneity of conditional ORs.
- Check the conditional independence of treatment and improvement given sex.
- Check the independence of treatment, sex and improvement.

Multicategory logit models: nominal response

- $\pi_j(x_1, \ldots, x_p) = \Pr(Y = j|X_1 = x_1, \ldots, X_p = x_p) : \sum_{j=1}^{J} \pi_j(x_1, \ldots, x_p) = 1,\ J \geq 3$
- Baseline-Category Logit Model
  - Pick up one category as base level first and pair each category with the chosen baseline category
  - Odds of Category j vs Category 1 for $x_1, \ldots, x_p$ and each j,

$$\text{odds}_{j,x_1,\ldots,x_p} = \frac{\pi_j(x_1, \ldots, x_p)}{\pi_1(x_1, \ldots, x_p)}|x_1, \ldots, x_p = \exp(\alpha_j + \beta_{j1}x_1 + \ldots + \beta_{jp}x_p)$$

  - Equivalently,

$$\ln(\text{odds}_{j,x_1,\ldots,x_p}) = \alpha_j + \beta_{j1}x_1 + \ldots + \beta_{jp}x_p$$

  - $(J-1) \times (p+1)$ unknown parameters in total
- Inference
  - Since $\sum_{j=1}^{J} \pi_j(x_1, \ldots, x_p) = 1$,

$$\pi_j(x_1, \ldots, x_p) = \frac{\text{odds}_{j,x_1,\ldots,x_p}}{\sum_{j=1}^{J} \text{odds}_{j,x_1,\ldots,x_p}} \Rightarrow \hat{\pi}_j = \frac{\widehat{\text{odds}}_{j,x_1,\ldots,x_p}}{\sum_{j=1}^{J} \widehat{\text{odds}}_{j,x_1,\ldots,x_p}}$$

  - Odds of Category j vs Category i for $x_1, \ldots, x_p$ regardless of baseline category:

$$\frac{\pi_j(x_1, \ldots, x_p)}{\pi_i(x_1, \ldots, x_p)} = \frac{\text{odds}_{j,x_1,\ldots,x_p}}{\text{odds}_{i,x_1,\ldots,x_p}} = \exp((\alpha_j - \alpha_i) + (\beta_{j1} - \beta_{i1})x_1 + \ldots + (\beta_{jp} - \beta_{ip})x_p)$$

    - CI: first estimate the CI for $(\alpha_j - \alpha_i) + (\beta_{j1} - \beta_{i1})x_1 + \ldots + (\beta_{jp} - \beta_{ip})x_p$ and take an exp form
  - OR of Category j vs Category 1 for $x_1 + c_1, \ldots, x_p + c_p$ and $x_1, \ldots, x_p$:

$$\frac{\text{odds}_{j,x_1+c_1,\ldots,x_p+c_p}}{\text{odds}_{j,x_1,\ldots,x_p}} = \exp(\beta_{j1}c_1 + \ldots + \beta_{jp}c_p)$$

    - CI: first estimate the CI for $\beta_{j1}c_1 + \ldots + \beta_{jp}c_p$ and take an exp form
- Implementation
  - VGAM::vglm()

Multicategory logit models: ordinal response

- $\pi_j(x_1, \ldots, x_p) = \Pr(Y = j | X_1 = x_1, \ldots, X_p = x_p) : \sum_{j=1}^{J} \pi_j(x_1, \ldots, x_p) = 1,\ J \geq 3$
  - the order of response categories contains extra information
- Cumulative Logit Model: for $j = 1, \ldots, J-1$,

$$\sum_{i=1}^{j} \pi_i(x_1, \ldots, x_p) = \Pr(Y \leq j | X_1 = x_1, \ldots, X_p = x_p) = \text{logit}^{-1}(\alpha_j + \beta_{j1}x_1 + \ldots + \beta_{jp}x_p)$$

  - special case: with proportional odds

$$\text{logit}(\Pr(Y \leq j | X_1 = x_1, \ldots, X_p = x_p)) = \alpha_j + \beta_1 x_1 + \ldots + \beta_p x_p$$

  - unable to constrain the parameters to prevent that $\Pr(Y \leq j | X_1 = x_1, \ldots, X_p = x_p) < \Pr(Y \leq j' | X_1 = x_1, \ldots, X_p = x_p)$ for some $j > j'$ (textbook pp.185)
- Inference
  - $\pi_j(x_1, \ldots, x_p) = \text{logit}^{-1}(\alpha_j + \beta_{j1}x_1 + \ldots + \beta_{jp}x_p) - \text{logit}^{-1}(\alpha_{j-1} + \beta_{j-1,1}x_1 + \ldots + \beta_{j-1,p}x_p)$
- Implementation
  - VGAM::vglm()
  - MASS::polr() for proportional case

Multicategory logit models: alternative

- Adjacent-Categories Logit Model

$$\ln \frac{\pi_{j+1}}{\pi_j} = \alpha_j + \beta_{j1}x_1 + \ldots + \beta_{jp}x_p$$

  - also allow one to take advantage of an ordinal response due to the successive comparisons
  - special case:

$$\ln \frac{\pi_{j+1}}{\pi_j} = \alpha_j + \beta_1 x_1 + \ldots + \beta_p x_p$$

- Implementation
  - VGAM::vglm()
