STAT475/675 TUT10

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

Zhiyang Zhou (zhiyang_zhou@sfu.ca)

2018-03-18

Model checking

- Data $\{(y_i, x_{i1}, \dots, x_{ip}) : i = 1, \dots, I\}$
  - logit model: use the data in aggregated form, i.e., y_i is the realization of $Y_i \sim \text{Binom}(n_i, \pi_i(x_{i1}, \dots, x_{ip}))$, aka the number of successes with n_i trials and treatment $(x_{i1}, \dots, x_{ip})$
  - loglinear model: y_i is the count associated with $(x_{i1}, \dots, x_{ip})$, aka the realization of $Y_i \sim \text{Pois}(\mu_i(x_{i1}, \dots, x_{ip}))$
  - rule of thumb: regroup the data to make sure that
    * logit model: $n_i \geq 5$ and $n = \sum_i n_i \gg 1$
    * loglinear model: $\mu_i(x_{i1}, \dots, x_{ip})$ is as large as possible
    * different grouping leads to different conclussions

- Inferential method
  - $H_0 : M$ is correct vs H_1 : otherwise
    * special case: checking independence for contingency tables
  - r is the number of non-redundant parameters in M
  - Pearson's χ^2-test: under H_0 with $\text{df}_M = I - r$,

$$\mathcal{K}^2 = \sum_{i=1}^{I} \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i} \approx \chi^2(\text{df}_M)$$

  - LRT: under H_0 with $\text{df}_M = I - r$,

$$\mathcal{G}^2 = 2 \sum_{i=1}^{I} y_i \ln \frac{y_i}{\hat{y}_i} \approx \chi^2(\text{df}_M),$$

- Graphical method: residual plots
  - Pearson's residual:

$$e_i = \frac{y_i - \hat{y}_i}{\sqrt{\widehat{\text{var}}(Y_i)}}$$

    * logit model: $\hat{y}_i = n_i \hat{\pi}_i$ and $\widehat{\text{var}}(Y_i) = n_i \hat{\pi}_i (1 - \hat{\pi}_i)$
    * loglinear model: $\hat{y}_i = \hat{\mu}_i = \widehat{\text{var}}(Y_i) = \frac{y_i - \hat{y}_i}{\sqrt{\widehat{\text{var}}(Y_i)}}$
  - standardized (adjusted) Pearson's residual (approximately normal-distributed):

$$e_i^* = \frac{e_i}{\sqrt{1 - h_{ii}}} = \frac{y_i - \hat{y}_i}{\sqrt{\widehat{\text{var}}(Y_i - \hat{Y}_i)}}$$

    where h_{ii} is the i-th observation's leverage: the i-th diagonal element of $H = V^{\frac{1}{2}} X (X^{\mathrm{T}} V X)^{-1} X^{\mathrm{T}} V^{\frac{1}{2}}$ with $V = \text{diag}(\widehat{\text{var}}(Y_1), \dots, \widehat{\text{var}}(Y_I))$
  - extreme residuals: implies extra variability not well-explained by the model:
    * size of residuals

        - $|e_i| \geq 2$ (or $|e_i^*| \geq 2$): 5% if the model is correct
        - $|e_i| \geq 3$ (or $|e_i^*| \geq 3$): extremely rare (0.1%) if the model is correct
        - $|e_i| \geq 4$ (or $|e_i^*| \geq 4$): unexpected at all if the model is correct
    - graph of residual vs explanatory variable
        - check the appropriateness of the form of explanatory variables
    - graph of residual vs $\hat{y}$ or $g(\hat{y})$
        - check the appropriateness of link function $g(\cdot)$

Model comparison and variable selection

- LRT: to compare a "smaller" model to a "larger" model, i.e., with $M_0 \subset M_1$,

$$H_0 : M_0 \text{ vs } H_1 : M_1$$

  - under H_0, $\mathcal{G}^2(M_0|M_1) = \mathcal{G}^2(M_0|M_s) - \mathcal{G}^2(M_1|M_s) = -2\ln\frac{\max L_{M_0}}{\max L_{M_1}} \approx \chi^2(\text{df}_{M_0} - \text{df}_{M_1})$
  - M_s is the saturated model
  - $\text{df}_{M_0} - \text{df}_{M_1}$ = the difference on numbers of non-redundant parameters
  - M_0 ought to be nested into M_1

- Information criteria: to achieve the min AIC, or corrected AIC or BIC
  - general form

$$\text{IC}(k) = -2\ln(L(\hat{\beta}|\text{data})) + kr$$

    with r non-redundant parameters
  - Akaike's Information Criterion (AIC):

$$\text{AIC} = \text{IC}(2) = -2\ln(L(\hat{\beta}|\text{data})) + 2r$$

  - corrected AIC (AIC_c):

$$\text{AIC}_\text{c} = \text{IC}\left(\frac{2n}{I-r-1}\right) = -2\ln(L(\hat{\beta}|\text{data})) + \frac{2Ir}{I-r-1}$$

  - Bayesian information criterion (BIC; Schwarz criterion):

$$\text{BIC} = \text{IC}(\ln I) = -2\ln(L(\hat{\beta}|\text{data})) + r\ln I$$

  - R functions
    - computation: AIC() and BIC()
    - model auto-selection: step() with options
      - "scope": the range of models examined in the search
      - "direction": "both", "backward", or "forward". If "scope" is missing, "direction" is always "backward".
      - "k": the k for IC(k)

Demo I

Data "UCBAdmissions" (included in R default Package "datasets") is on applicants to graduate school at Berkeley for the six largest departments in 1973 classified by admission and sex.

- Admit: Admitted, Rejected
- Gender: Male, Female
- Dept: A, B, C, D, E, F

Demo II

250 groups went to a park for fishing. Each group was questioned about

- count (integer): number of fishes they caught;
- persons (integer): number of people were in the group;
- camper (categorical): whether or not they brought a camper;
- livebait (categorical): whether or not they used live bait;
- child (categorical): number of children were in the group.

See https://stats.idre.ucla.edu/r/dae/zip/ for more details.
