Multinomial Data

- Recall: in binomial GLM, observe data

  $$(x_{i1}, x_{i2}, \ldots, x_{im}, \underbrace{y_i}_{y_{i1} \text{ when } J=2}),\ i = 1, 2, \ldots, I \Leftrightarrow (\mathbf{x}_i, y_i) \Leftrightarrow (\mathbf{x}, y_\mathbf{x})$$

  where $y_\mathbf{x}$ is bounded by $n_\mathbf{x}$=number of total trials

- Data: now observe a response with J categories, where $J > 2$:

  $$(x_{i1}, x_{i2}, \ldots, x_{im}, \underbrace{y_{i1}, y_{i2}, \ldots, y_{iJ}}_{\text{levels of a response factor } r}),\ i = 1, 2, \ldots, I$$

  $\Leftrightarrow (\mathbf{x}_i, Y_i) \Leftrightarrow (\mathbf{x}, Y_\mathbf{x})$, where (i) $Y_i = (y_{i1}, y_{i2}, \ldots, y_{iJ})$,

  might have same values on multiple x_i's

  (ii) y_{ij}= # of observations falling into category j at $\mathbf{x}_i$,

  (iii) $\sum_{j=1}^{J} y_{ij} = n_i$ (or denoted by $n_\mathbf{x}$), and n_i's are fixed

  - ➢ $Y_\mathbf{x} \sim$ multinomial($n_\mathbf{x}$, $p_{\mathbf{x}1}$, $p_{\mathbf{x}2}$, …, $p_{\mathbf{x}J}$), where $\sum_{j=1}^{J} p_{\mathbf{x}j} = 1$

    prob. of jth category at x

  - ➢ we may encounter grouped or ungrouped data

    distinct x_i

    - For ungrouped data, n_i=1 for i=1, 2, …, I
    - The idea of covariate classes can be similarly applied to generate grouped data from ungrouped data (**Q**: what benefit?)

      $n_x \uparrow$

      - larger n_x
      - better approximation in asymptotic theory

  - ➢ The J categories can be
    - *nominal* or *ordinal*, or
    - have a *hierarchical or nested* structure

Nominal Multinomial Responses

- **Q**: how to link $p_\mathbf{x}$=($p_{\mathbf{x}1}$, $p_{\mathbf{x}2}$, …, $p_{\mathbf{x}J}$) with $\mathbf{x}$? levels of a response factor r

  Recall 3 components in GLM

  - ➢ Recall: binomial logit model.

    $$\underbrace{(1-p_\mathbf{x}, \overbrace{p_\mathbf{x}}^{\dim=1})}_{J=2} \Rightarrow (1, p_\mathbf{x}/(1-p_\mathbf{x})) \Rightarrow p_\mathbf{x}/(1-p_\mathbf{x}) \underset{\text{exp}}{\overset{\text{log}}{\rightleftarrows}} \eta_\mathbf{x} = X\beta$$

- *Multinomial logit model*:

  $$\underbrace{(p_{\mathbf{x}1}, p_{\mathbf{x}2}, \ldots, p_{\mathbf{x}J})}_{\dim = J-1} \Rightarrow (1, \overbrace{p_{\mathbf{x}2}/p_{\mathbf{x}1}, \ \ldots\ , p_{\mathbf{x}J}/p_{\mathbf{x}1}}^{J-1})$$

  odds for multinomial

  relative proportion of two categories

  $$\begin{array}{cccc} \text{log}\updownarrow\text{exp} & \text{log}\updownarrow\text{exp} & \cdots & \text{log}\updownarrow\text{exp} \\ \eta_{\mathbf{x}1}=0 & \eta_{\mathbf{x}2}=X\beta_2 & \cdots & \eta_{\mathbf{x}J}=X\beta_J \end{array}$$

  $X\beta_1$ & $\beta_1 = 0$

  same, different

  $$\boxed{\begin{aligned} X\beta_j = 1\ &\beta_{0j} + \\ &h_1(\mathbf{x})\beta_{1j} + \cdots + \\ &h_{p-1}(\mathbf{x})\beta_{p-1,j} \end{aligned}}$$

  - ➢ The link obeys the constraints $0 \le p_{\mathbf{x}j} \le 1$, j=1,…,J, and $\Sigma_j p_{\mathbf{x}j}$=1
  - ➢ Category 1 declared as the *baseline* (note that we add $\eta_{\mathbf{x}1}$=0)

    can choose other category

  - ➢ $p_{\mathbf{x}j} = \exp(\eta_{\mathbf{x}j}) \Big/ \Big[\underbrace{1}_{\exp(\eta_{x,1})} + \sum_{j=2}^{J} \exp(\eta_{\mathbf{x}j})\Big]$ $\quad p_{xj} = \dfrac{p_{xj}/p_{x1}}{1 + p_{x2}/p_{x1} + p_{x3}/p_{x1} + \cdots + p_{xJ}/p_{x1}}$
  - ➢ Inferences
    - Estimation of the parameters: can use maximum likelihood
    - Other inferences: can use standard likelihood-based approach
    - Concepts and methods in GLM can be similarly applied (**Q**: why?) → rationale from multinomial log-linear model

➢ Interpretation of $\boldsymbol{\beta}_j$ $=(p_{xj}/p_{x1})/(p_{xj'}/p_{x1})$

- The log-odds in favor of category j over category j' is

$X^* = (1, h_1(X), \cdots, h_{p-1}(X))$

$$\log(p_{\mathbf{x}j}/p_{\mathbf{x}j'}) = \mathbf{x}^{*T}(\boldsymbol{\beta}_j - \underbrace{\boldsymbol{\beta}_{j'}}_{=0 \text{ if } j'=1}) = \eta_{xj} - \eta_{xj'}$$

Note. change in p_{xj}/p_{x1}, not change in p_{xj}

(⇒ the contrasts among the vectors $\boldsymbol{\beta}_j$ are of interest)

- The interpretation of $\boldsymbol{\beta}_j$ depends on the choice of baseline category ⇒ the k^{th} component of $\boldsymbol{\beta}_j$ represents the change in log-odds of moving from the baseline category to the j^{th} category for a unit change in the k^{th} predictor of $\mathbf{x}$ (other predictors held constant)

Let $j'=1$, $\beta_{j'}=0$

cf. Recall: nominal predictor with J categories → choose one category as reference. (1.2)

- Intercepts of $\boldsymbol{\beta}_j$ (β_{0j}) represent probabilities of J categories at $\mathbf{x}^*=\mathbf{0}$

• *Multinomial log-linear model*: $y \sim$ Poisson

SS1 →

➢ Recall: the connection between Poisson and multinomial ⇒ it is possible to fit a model for multinomial response using a Poisson

SS2 → SS3 →

a factor with I categories — response factor

$\mathbf{x}$ \ r	1	…	J	
$\mathbf{x}_1$	y_{11}	…	y_{1J}	n_1
…	…	…	…	…
$\mathbf{x}_I$	y_{I1}	…	y_{IJ}	n_I

vector — each a multinomial — fixed

➢ Procedure:

- Arrange the data in a two-way tables (LNp 5-4 ~5-13)
- Declare a nominal variable that has a level for each multinomial response category, called it *response factor*, r