Nominal multinomial regression

data $(\underset{\sim}{x}_i, \underbrace{y_{i1}, \ldots, y_{iJ}}_{\text{sum to } n_i})$ $\quad i = 1, \ldots, n$

on n_i independent trials,
y_{ij} had the j-th of J
(unordered) outcomes

so

$$Pr\{Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ}\} =$$

$$\left(\frac{n_i!}{y_{i1}! \ldots y_{iJ}!}\right) p_{i1}^{y_{i1}} \cdots p_{iJ}^{y_{iJ}}$$

Ch. 5 example

Survey of $n = 944$ voters

Y_i = party identification
"REP.", "DEM.", "IND."

$\underset{\sim}{x}_i$ = income, age, education

numeric

categorical

So $n_i = 1$, $i = 1, \dots, 944$

$J = 3$

Exp. family? GLM?

grey area, since response var. is J-1 dimensional

variance function?

Can still think link structure though

$$g\begin{pmatrix} p_2 \\ \vdots \\ p_J \end{pmatrix} = \begin{pmatrix} \log\left\{\frac{p_2}{1-(p_2+\dots+p_J)}\right\} \\ \vdots \\ \log\left\{\frac{p_J}{1-(p_2+\dots+p_J)}\right\} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{\beta}_2^T \\ \vdots \\ \underset{\sim}{\beta}_J^T \end{pmatrix} \underset{\sim}{X}$$

Interpretation

$$\frac{p_j}{p_1} = e^{\underset{\sim}{\beta}_j^T \underset{\sim}{x}}$$

conditional odds?

$$\frac{p_j}{p_k} = e^{(\underset{\sim}{\beta}_j - \underset{\sim}{\beta}_k)^T \underset{\sim}{x}}$$

$$(p_1, \ldots, p_J) = \frac{1}{\left(1 + \sum_{j=2}^{J} e^{\underset{\sim}{\beta}_j^T \underset{\sim}{x}}\right)} \left(1, e^{\underset{\sim}{\beta}_2^T \underset{\sim}{x}}, \ldots, e^{\underset{\sim}{\beta}_J^T \underset{\sim}{x}}\right)$$

Note: typically want

$\underset{\sim}{x} = (1, \text{predictors})$

first col. of $\left(\quad B \quad \right)$

are intercept params.

Ex.

$$\hat{\beta} = \begin{array}{c c} & \begin{array}{cc} \text{Intcpt} & \text{Income (\$1000)} \end{array} \\ \begin{array}{c} \text{"IND"} \\ \text{"REP"} \end{array} & \begin{pmatrix} -1.18 & 0.0161 \\ -0.95 & 0.0176 \end{pmatrix} \end{array}$$

$$\widehat{p \mid INC=0} \propto \left(e^{0}, e^{-1.18}, e^{-0.95}\right)$$

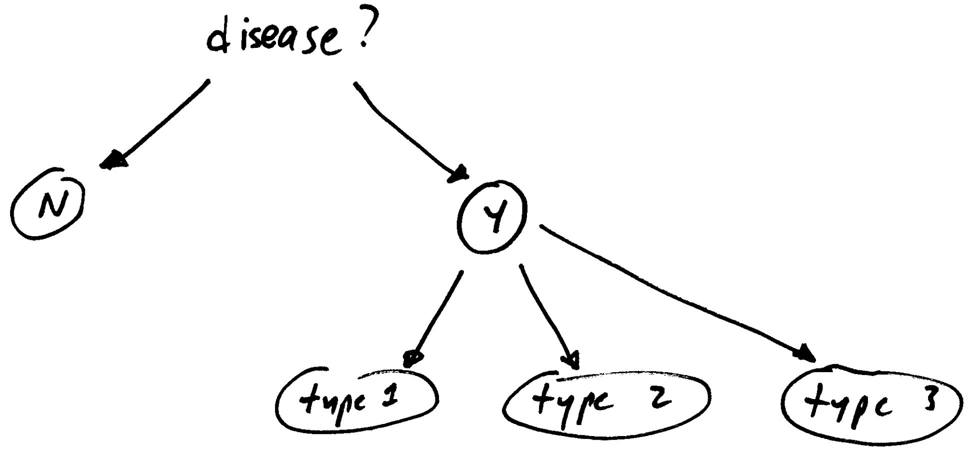
$$= \left(\underset{\text{"D"}}{0.59}, \underset{\text{"I"}}{0.18}, \underset{\text{"R"}}{0.23}\right)$$

Cond. OR associated with \$1000 increase in income?

$$e^{.0161} \approx 1.0162$$

$$e^{.017} \approx 1.0178$$

Hierarchical / Nested



Strategy - separately fit

Binomial model for DISEASE

Multinomial model for TYPE | DISEASE=YES

Statistical justification?

Start with multimonial model

$$\text{lik.} \propto \prod_{i=1}^{n} p_1^{y_{i1}} p_2^{y_{i2}} p_3^{y_{i3}} p_4^{y_{i4}}$$

reparameterize $q = \Pr\{\text{Dis.} = Y\} = p_2 + p_3 + p_4$

$$r_i = \Pr\{\text{Type} = i \mid \text{Dis.} = Y\} = \frac{p_{i+1}}{p_2 + p_3 + p_4}$$

So $$\text{lik.} \propto \prod_{i=1}^{n} (1-q)^{y_{i1}} (q r_1)^{y_{i2}} (q r_2)^{y_{i3}} (q r_3)^{y_{i4}}$$

$$\propto \left\{ \prod_{i=1}^{n} (1-q)^{y_{i1}} q^{y_{i2}+y_{i3}+y_{i4}} \right\} \left\{ \prod_{i=1}^{n} r_1^{y_{i2}} r_2^{y_{i3}} r_3^{y_{i4}} \right\}$$

Fundamentally different modelling assumptions by choosing different parameterization for "linking"

e.g. nested model says

$$\log\left(\frac{p_2 + p_3 + p_4}{p_1}\right) \text{ linear in } \underset{\sim}{x}$$

regular multinomial model says

$$\frac{p_2 + p_3 + p_4}{p_1} = e^{\underset{\sim}{\beta}_2^T \underset{\sim}{x}} + e^{\underset{\sim}{\beta}_3^T \underset{\sim}{x}} + e^{\underset{\sim}{\beta}_4^T \underset{\sim}{x}}$$