Multinomial logit models

NOV 22

Logistic Regression with more than 2 outcomes.

Logit is log of probability ratio

$$\log \frac{\pi}{1-\pi} = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

For an outcome with c categories, have $c-1$ logits, and a regression equation for each one.

For c categories

$$\log\left(\frac{\pi_1}{\pi_c}\right) = \beta_{01} + \beta_{11} x_1 + \cdots + \beta_{k1} x_k$$

$$\log\left(\frac{\pi_2}{\pi_c}\right) = \beta_{02} + \beta_{12} x_1 + \cdots + \beta_{k2} x_k$$

$$\vdots$$

$$\log\left(\frac{\pi_{c-1}}{\pi_c}\right) = \beta_{0,c-1} + \beta_{1,c-1} x_1 + \cdots + \beta_{k,c-1} x_k$$

If $c = 2$, have regular logistic regression

Example! 3 outcomes, 2 x variables

First-year university students. Next Fall term, are they

- Full-time π_3
- Part-time π_2
- Gone π_1

Explanatory variables

$$x_1 = \text{HS GPA}$$

$$x_2 = \begin{cases} 1 & \text{lives on campus} \\ 0 & \text{lives off campus} \end{cases}$$

The category without a regression equation is the reference category

$$\log \frac{\pi_1 \leftarrow P(\text{Gone})}{\pi_3} = \beta_{01} + \beta_{11} x_1 + \beta_{21} x_2$$

$$\log \frac{\pi_2 \leftarrow P(\text{Part time})}{\pi_3} = \beta_{02} + \beta_{12} x_1 + \beta_{22} x_2$$

Call π_1/π_3 & π_2/π_3 "Relative probabilities" Relative to P(Full time)

$$\log \frac{\pi_1}{\pi_3} = \beta_{01} + \beta_{11} x_1 + \beta_{21} x_2$$

$$\log \frac{\pi_2}{\pi_3} = \beta_{02} + \beta_{12} x_1 + \beta_{22} x_2$$

Controlling for HSGPA the relative probability of being Gone next year is ______ times as great as for those living in residence.

$$\frac{e^{\beta_{01} + \beta_{11} x_1 + \beta_{21}}}{e^{\beta_{01} + \beta_{11} x_1}} = e^{\beta_{21}}$$

Like an odds ratio

Controlling for HSGPA, the relative probability of being part time next year is ______ times as great for a student living in residence.

$$\frac{e^{\beta_{02} + \beta_{12} x_1 + \beta_{22}}}{e^{\beta_{02} + \beta_{12} x_1 + 0}} = e^{\beta_{22}}$$

$$\log \frac{\pi_1}{\pi_3} = \beta_{01} + \beta_{11} x_1 + \beta_{21} x_2$$

$$\log \frac{\pi_2}{\pi_3} = \beta_{02} + \beta_{12} x_1 + \beta_{22} x_2$$

Taking living location into account, an increase of 5 HSGPA points multiplies the relative probability of being gone by ________.

$$\frac{e^{\cancel{\beta_{01}} + \beta_{11}(x_1+5) + \cancel{\beta_{21} x_2}}}{e^{\cancel{\beta_{01}} + \beta_{1} x_1 + \cancel{\beta_{21} x_2}}} = e^{5\beta_{11}}$$

Solve for the probabilities

$$\log \frac{\pi_1}{\pi_3} = \beta_{01} + \beta_{11} x_1 + \cdots + \beta_{k1} x_k = L_1$$

$$\log \frac{\pi_2}{\pi_3} = \beta_{02} + \beta_{12} x_1 \cdots + \beta_{k2} x_k = L_2$$

Have

$$\log \frac{\pi_1}{\pi_3} = L_1 \Rightarrow \frac{\pi_1}{\pi_3} = e^{L_1} \Rightarrow \pi_1 = \pi_3 e^{L_1}$$

$$\log \frac{\pi_2}{\pi_3} = L_2 \Rightarrow \frac{\pi_2}{\pi_3} = e^{L_2} \Rightarrow \pi_2 = \pi_3 e^{L_2}$$

So need to solve

$$\pi_1 = \pi_3 e^{L_1}$$

$$\pi_2 = \pi_3 e^{L_2}$$

$$\pi_1 + \pi_2 + \pi_3 = 1$$

$$\pi_3 e^{L_1} + \pi e^{L_2} + \pi_3 = 1 = \pi_3 (1 + e^{L_1} + e^{L_2})$$

$$\Rightarrow \pi_3 = \frac{1}{1 + e^{L_1} + e^{L_2}}$$

$$\pi_1 = \frac{e^{L_1}}{1 + e^{L_1} + e^{L_2}}, \quad \pi_2 = \frac{e^{L_2}}{1 + e^{L_1} + e^{L_2}}$$

If more than 3 categories,

$$\log\frac{\pi_1}{\pi_6} = L_1 \Rightarrow \frac{\pi_1}{\pi_6} = e^{L_1} \Rightarrow \pi_1 = \pi_6 e^{L_1}$$

$$\log\frac{\pi_2}{\pi_6} = L_2 \Rightarrow \frac{\pi_2}{\pi_6} = e^{L_2} \Rightarrow \pi_2 = \pi_6 e^{L_2}$$

$$\vdots \qquad\qquad \vdots$$

$$\log\frac{\pi_5}{\pi_6} = L_5 \Rightarrow \frac{\pi_5}{\pi_6} = e^{L_5} \Rightarrow \pi_5 = \pi_6 e^{L_5}$$

$$\pi_1 + \cdots + \pi_6 = 1, \quad \text{so}$$

$$= \pi_6 e^{L_1} + \pi_6 e^{L_2} + \cdots + \pi_6 e^{L_5} + \pi_6$$

$$\Rightarrow \pi_6 = \frac{1}{1 + \sum_{j=1}^{c-1} e^{L_j}}$$

$$\pi_1 = \frac{e^{L_1}}{1 + \sum_{j=1}^{5} e^{L_j}}$$

$$\pi_2 = \frac{e^{L_2}}{1 + \sum_{j=1}^{5} e^{L_j}}$$

$$\vdots$$

These probabilities $\pi_1, \ldots \pi_c$ are actually

$$\pi_1(\underset{\sim}{x}_i), \pi_2(\underset{\sim}{x}_i) \cdots \pi_c(\underset{\sim}{x}_i)$$

Put in multinomial likelihood

$$\underset{\sim}{y}_i \overset{ind}{\sim} M(1, \underset{\sim}{\pi}(\underset{\sim}{x}_i))$$

What does $\underset{\sim}{y}_i$ look like?

$\underset{\sim}{y}_i = (y_{i1}, y_{i2}, \ldots, y_{ic})$,

All zeros, except for a single 1 at the position of the outcome for case i.

$$\ell(\underset{\sim}{\beta}) = \prod_{i=1}^{n} \pi_1(\underset{\sim}{x}_i)^{y_{i1}} \pi_2(\underset{\sim}{x}_i)^{y_{i2}} \cdots \pi_c(\underset{\sim}{x}_i)^{y_{ic}}$$

$$= \prod_{i=1}^{n} \left(\frac{e^{L_{i1}}}{1+\Sigma e^{L_{ir}}}\right)^{y_{i1}} \left(\frac{e^{L_{i2}}}{1+\Sigma e^{L_{ir}}}\right)^{y_{i2}} \cdots \left(\frac{e^{L_{i,c-1}}}{1+\Sigma e^{L}}\right)^{y_{i,c-1}} \left(\frac{1}{1+\Sigma e^{L}}\right)^{y_{ic}}$$

R's mlogit package

8

Need to install from within R