Nominal and ordinal logistic regression

Patrick Breheny

April 26

Introduction

- Our goal for today is to briefly go over ways to extend the logistic regression model to the case where the outcome can have multiple categories (*i.e.*, not binary)
- We will discuss two approaches:
  - *Multinomial logistic regression*, which makes no assumptions regarding the relationship between the categories, and is most appropriate for nominal outcomes
  - The *proportional odds model*, which assumes an ordering to the categories and is most appropriate for ordinal outcomes

Notation

We will use the following notation to describe these multi-class models:

- Let Y be a random variable that can on one of K discrete value (*i.e.*, fall into one of K classes)
- Number the classes $1, \ldots, K$
- Thus, $\pi_{i2} = \Pr(Y_i = 2)$ denotes the probability that the ith individual's outcome belongs to the second class
- More generally, $\pi_{ik} = \Pr(Y_i = k)$ denotes the probability that the ith individual's outcome belongs to the kth class

The multinomial logistic regression model

- Multinomial logistic regression is equivalent to the following:
  - Let $k = 1$ denote the reference category
  - Fit separate logistic regression models for $k = 2, \ldots, K$, comparing each outcome to the baseline:

$$\log\left(\frac{\pi_{ik}}{\pi_{i1}}\right) = \mathbf{x}_i^T \boldsymbol{\beta}_k$$

- Note that this will result in $K - 1$ vectors of regression coefficients (we don't need to estimate the Kth vector because $\sum_k \pi_k = 1$)

Probabilities and odds ratios

The fitted class probabilities for an observation with explanatory variable vector $\mathbf{x}$ are therefore

$$\hat{\pi}_1 = \frac{1}{1 + \sum_k \exp(\mathbf{x}^T \widehat{\boldsymbol{\beta}}_k)}$$

$$\hat{\pi}_k = \frac{\exp(\mathbf{x}^T \widehat{\boldsymbol{\beta}}_k)}{1 + \sum_l \exp(\mathbf{x}^T \widehat{\boldsymbol{\beta}}_l)}$$

Probabilities and odds ratios

- Like logistic regression, odds ratios in the multinomial model are easily estimated as exponential functions of the regression coefficients:

$$\begin{aligned}\mathrm{OR}_{kl} = \frac{\pi_k}{\pi_l} &= \frac{\pi_k/\pi_1}{\pi_l/\pi_1} \\ &= \frac{\exp\left((\mathbf{x}_2 - \mathbf{x}_1)^T \boldsymbol{\beta}_k\right)}{\exp\left((\mathbf{x}_2 - \mathbf{x}_1)^T \boldsymbol{\beta}_l\right)} \\ &= \exp\left((\mathbf{x}_2 - \mathbf{x}_1)^T (\boldsymbol{\beta}_k - \boldsymbol{\beta}_l)\right)\end{aligned}$$

- In the simple case of changing x_j by δ_j and comparing k to the reference category,

$$\mathrm{OR}_{kl} = \exp(\delta_j \beta_{kj})$$

Flu vaccine data: Results

- This model estimates the following odds ratios, comparing vaccinated to control:

	$\widehat{\beta}$	$\widehat{\mathrm{OR}}$
Moderate	2.24	9.38
Large	2.22	9.17

- A test of the null hypothesis that the odds ratios are all 1 is significant ($p = 0.00009$)
- Note: These are the same coefficients, the same ratios (replacing OR with RR), and the same p-value for the hypothesis test as the Poisson regression approach

Proportional odds: Introduction

- Multinomial regression requires the estimation of $(K-1)p$ parameters, and assumes nothing about the relationship between the categories to assist in that estimation
- This is very flexible of course, but has the potential to lead to large variability in the estimates, especially when the number of categories is large
- A common alternative when the categories are ordered to assume that the log odds of $Y \geq k$ is linearly related to the explanatory variables
- This is called the *proportional odds* model, and requires the estimate of only one regression coefficient per explanatory variable

Proportional odds model

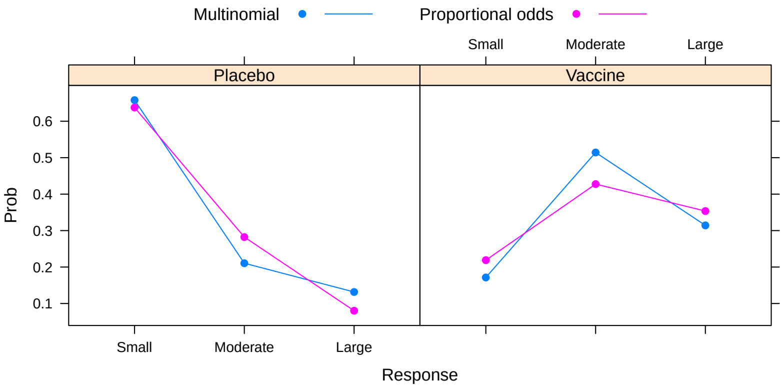
- Specifically, the proportional odds model assumes

$$\log\left(\frac{\pi_k + \cdots + \pi_K}{1 + \cdots + \pi_{k-1}}\right) = \beta_{0k} + \mathbf{x}^T\boldsymbol{\beta}$$

- Thus, we still have to estimate $K-1$ intercepts, but only p linear effects, where p is the number of explanatory variables (note that $K+p-1 < (K-1)(p+1)$ if $K>2$)
- Note: Writing down the proportional odds model requires us to modify the notation we've used all semester – so in the above, $\mathbf{x}$ and $\boldsymbol{\beta}$ do not include a term for the intercept

Proportional odds: Results

The proportional odds model estimates that the odds ratio for $Y \in \{\text{Moderate}, \text{Large}\}$ given vaccination is $\exp(\hat{\beta}_1) = 6.3$; furthermore, by assumption of the model, this is also the odds ratio for a large response relative to $\{\text{Small}, \text{Moderate}\}$ given vaccination

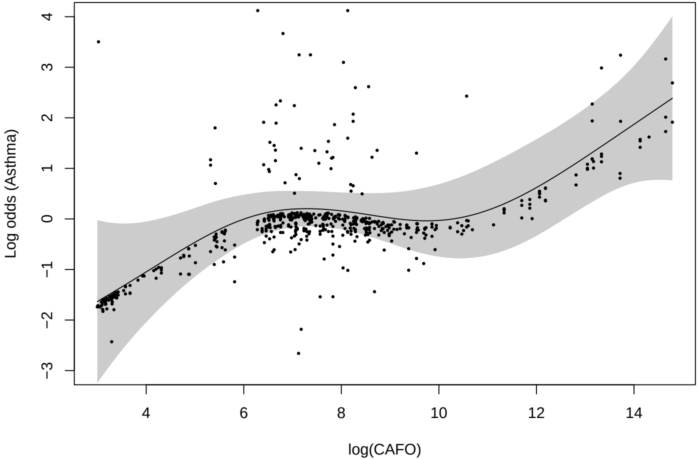


Non-linear models

- Linear and generalized linear models are certainly the most important class of models, but they are not the only kind of model
- For example, we have already alluded to the idea that we sometimes wish to allow the effect of an explanatory variable to be a smooth curve rather than a line:

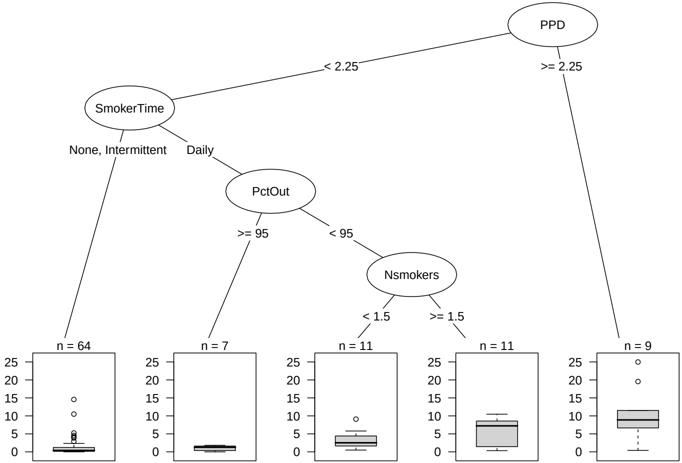
$$g(\mu_i) = f(x_i)$$

Non-linear models: Example



Tree-based models

There are also tree-based models:



And much more. . .

- There are many additional extensions/modifications/alternatives that have been proposed as well:
  - Robust regression
  - Distribution-free methods for inference
  - Discriminant analysis
  - Principal component analysis
  - Methods for dealing with highly correlated explanatory variables
  - Methods for variable selection that avoid the problems of subset selection
- These topics form the basis of BST 764: Applied Statistical Modeling, which I will be teaching next fall