Analysis of Categorical Data

Multicategory Logit Models

Outline

- Logit Models for Nomial Responses
- Cumulative Logit Models for Ordinal Responses
- Paired-Category Ordinal Logits

Logit Models for Nomial Responses

- At each combination of explanatory variables (X), Y has a multinomial distribution, where $\sum_{j=1}^{J} Y_j = n$ and $\sum_{j=1}^{J} \pi_j = 1$. Allocate n observations into J categoires.
- Once we speify $J-1$ logits, the rest are redundant.
- Baseline logits (use last J^{th} category as baseline)

$$\log\left(\frac{\pi_j}{\pi_J}\right) = \alpha_j + \beta_j X, j = 1, \ldots, J-1$$

  $J-1$ logit equations, each with separate parameters.
- All other logits can be found from these $J-1$ categories.

$$\begin{aligned} \log\left(\frac{\pi_a}{\pi_b}\right) &= \log\left(\frac{\pi_a}{\pi_J}\right) - \log\left(\frac{\pi_b}{\pi_J}\right) \\ &= (\alpha_a - \alpha_b) + (\beta_a - \beta_b)X \end{aligned}$$

Example 1: Alligator Food Choice

Example 2: Belief in Afterlife

Cumulative Logit Models (Ordinal)

- Consider the jth cumulative probability:

$$\underbrace{P(Y \leq j)}_{\text{probability of Y falling into category j or below}} = \pi_1 + \pi_2 + \ldots + \pi_j, j = 1, \ldots, J$$

- Ordering has effect of:
  - simpler interpretations
  - potentially more power than multicategory logit
- The cumulative logit

$$\begin{aligned} logit[P(Y \leq j)] &= \log\left[\frac{P(Y \leq j)}{1 - P(Y \leq j)}\right] \\ &= \log \frac{\pi_1 + \pi_2 + \ldots + \pi_j}{\pi_{j+1} + \pi_{j+2} + \ldots + \pi_J} \end{aligned}$$

Proportional Odds Model

- Given explanatory variable X,

$$logit[P(Y \leq j)] = \alpha_j + \beta X, j = 1, \dots, J-1$$

- Odds Ratio

$$\frac{P(Y \leq j|X = x_2)/P(Y > j|X = x_2)}{P(Y \leq j|X = x_1)/P(Y > j|X = x_1)} = e^{\beta(x_2 - x_1)}$$

  - The same "proportionality" constant β applies to all js
  - Odds ratio does not depend on j, but rather distance $x_2 - x_1$
  - When $x_2 = x_1 + 1$, the odds ratio is e^{β}
  - If categories are reversed, then same fit but $\hat{\beta}$ has opposite sign

Proportional Odds Model (Cont'd)

- Textbook p.181 Figure 6.2
  - Separate cure for each cumulative logit
  - Each curve can be thought of as a logistic regression with outcomes $Y \leq j$ and $Y > j$
  - Common β gives curves same shape.
  - If $\beta < 0$, the curves will be descend rather than ascend.
- Textbook p.181 Figure 6.3
  - As x increases, the response on Y is more likely to fall at the low end or the ordinal scale.
  - What if $\beta < 0$?

Example 3: Political Ideology

Inference and Model Fit

- Wald and likelihhod ratio tests for βs
- Related to tests for “Conditional Independence”
- Test for proportional odds assumption
- Check model fit: G^2 and X^2
  - separate effects for the different cumulative probabilities
  - fit baseline-category logit model
  - collapse ordinal categories to make binary response (not recommended - loss of efficiency and larger SEs)

SAS Summary

- PROC LOGISTIC
  - Proportional odds
  - Model on $\log[P(Y \leq j)/P(Y > j)]$
  - $X = \left\{ \begin{array}{ll} 1 & \text{Democrat} \\ 0 & \text{Republican} \end{array} \right\} Quantitative$
  - $\hat{\beta} = 0.975$ & $e^{0.975} = 2.65 = \frac{\hat{P}(Y \leq j)/\hat{P}(Y > j)|_{Demo}}{\hat{P}(Y \leq j)/\hat{P}(Y > j)|_{Rep}}$
- PROC CATMOD
  - Proportional odds: Clogit option
  - Model on $\log[P(Y > j)/P(Y \leq j)]$
  - $X = \left\{ \begin{array}{ll} 1 & \text{Democrat} \\ 0 & \text{Republican} \end{array} \right\} Categorical$
  - $X = 0 : \hat{\beta}^{Rep} = -.4875$; $X = 1 : \hat{\beta}^{Dem} = .4875$; $2\hat{\beta} = 0.975$

Example 4: Mental Health

Adjacent-Category Logits

- Consider the logit: $\log\left(\frac{\pi_{j+1}}{\pi_j}\right) = \alpha_j + \beta_j x$, where $j = 1, \ldots, J-1$
- A simpler version: $\log\left(\frac{\pi_{j+1}}{\pi_j}\right) = \alpha_j + \beta x$
  Explanation: Effect of X on odds of higher to lower response is same for all $J-1$ logits.

$$\frac{\frac{\pi_2}{\pi_1}\Big|_{x=b}}{\frac{\pi_2}{\pi_1}\Big|_{x=a}} = e^{\beta(b-a)}$$

Example 3 (Cont'd)

SAS Summary (Cont'd)

PROC CATMOD

- Adjacent Logit model: alogit option
- Model on $\log[\pi_{j+1}/\pi_j]$
- $X = \left\{ \begin{array}{ll} 1 & \text{Democrati} \\ 0 & \text{Republican} \end{array} \right\} Categorical$
- $X = 0 : \hat{\beta}^{Rep} = .2159$;
  $X = 1 : \hat{\beta}^{Dem} = -.2159$;
  $2\hat{\beta} = 0.43$
  $e^{0.43} = \frac{\hat{\pi}_{j+1}/\hat{\pi}_j|_{Rep}}{\hat{\pi}_{j+1}/\hat{\pi}_j|_{Dem}} = 1.54$

Continuation-Ratio Logits

- Logits 1:

$$\log\left(\frac{\pi_1}{\pi_2}\right)$$

$$\log\left(\frac{\pi_1+\pi_2}{\pi_3}\right)$$

...

$$\log\left(\frac{\pi_1+\ldots+\pi_{J-1}}{\pi_J}\right)$$

- Logits 2:

$$\log\left(\frac{\pi_1}{\pi_2+\ldots+\pi_J}\right)$$

$$\log\left(\frac{\pi_2}{\pi_3+\ldots+\pi_J}\right)$$

...

$$\log\left(\frac{\pi_{J-1}}{\pi_J}\right)$$

Example 5

A Developmental Toxicity Study

Test of Conditonal Independence

Example 6: Job Satisfaction and Income

- Likelihood Ratio Test
  - cumulative logit model
  - baseline-category logit model
- Generalized Cocharn-Mantel-Haenszel Tests
- Detecting Nominal-Ordinal Conditional Association
- Detecting Nominal-Nominal Conditional Association