Sociology 740

John Fox

Lecture Notes

10. Logit and Probit Models For Polytomous Data

1. Goal:

- To introduce similar statistical models for polytomous response variables, including ordered categories.

2. Introduction

- I will describe three general approaches to modeling polytomous data:

1. modeling the polytomy directly as a set of unordered categories, using a generalization of the dichotomous logit model;
2. constructing a set of nested dichotomies from the polytomy, fitting an independent logit or probit model to each dichotomy; and
3. extending the unobserved-variable interpretation of the dichotomous logit and probit models to ordered polytomies.

3. The Polytomous Logit Model

- The dichotomous logit model can be extended to a polytomy by employing the multivariate-logistic distribution. This approach has the advantage of treating the categories of the polytomy in a non-arbitrary, symmetric manner.
- The response variable Y can take on any of m qualitative values, which, for convenience, we number $1, 2, ..., m$ (using the numbers only as category labels).
  - For example, in the UK, voters can vote for (1) the Conservatives, (2) Labour, or (3) the Liberal Democrats (ignoring other parties).
- Let π_{ij} denote the probability that the ith observation falls in the jth category of the response variable; that is,

$$\pi_{ij} \equiv \Pr(Y_i = j) \text{ for } j = 1, ..., m.$$

- We have k regressors, $X_1, ..., X_k$, on which the π_{ij} depend.

- More specifically, suppose that this dependence can be modeled using the *multivariate logistic distribution*:

$$\pi_{ij} = \frac{e^{\gamma_{0j}+\gamma_{1j}X_{i1}+\cdots+\gamma_{kj}X_{ik}}}{1+\sum_{l=1}^{m-1} e^{\gamma_{0l}+\gamma_{1l}X_{i1}+\cdots+\gamma_{kl}X_{ik}}}$$

for $j = 1, ..., m-1$

$$\pi_{im} = 1 - \sum_{j=1}^{m-1} \pi_{ij}$$

- There is one set of parameters, $\gamma_{0j}, \gamma_{1j}, ..., \gamma_{kj}$, for each response-variable category but the last; category m functions as a type of baseline.
- The use of a baseline category is one way of avoiding redundant parameters because of the restriction that $\sum_{j=1}^{m} \pi_{ij} = 1$.

- Some algebraic manipulation of the model produces

$$\log_e \frac{\pi_{ij}}{\pi_{im}} = \gamma_{0j} + \gamma_{1j}X_{i1} + \cdots + \gamma_{kj}X_{ik}$$

for $j = 1, ..., m-1$

- The regression coefficients affect the log-odds of membership in category j versus the baseline category.
- It is also possible to form the log-odds of membership in *any* pair of categories j and j':

$$\begin{aligned}\log_e \frac{\pi_{ij}}{\pi_{ij'}} &= \log_e \left(\frac{\pi_{ij}}{\pi_{im}} \bigg/ \frac{\pi_{ij'}}{\pi_{im}} \right) \\ &= \log_e \frac{\pi_{ij}}{\pi_{im}} - \log_e \frac{\pi_{ij'}}{\pi_{im}} \\ &= (\gamma_{0j} - \gamma_{0j'}) + (\gamma_{1j} - \gamma_{1j'})X_{i1} \\ &\quad + \cdots + (\gamma_{kj} - \gamma_{kj'})X_{ik}\end{aligned}$$

  – The regression coefficients for the logit between any pair of categories are the differences between corresponding coefficients.

▶ Now suppose that the model is specialized to a dichotomous response variable. Then, $m = 2$, and

$$\begin{aligned}\log_e \frac{\pi_{i1}}{\pi_{i2}} &= \log_e \frac{\pi_{i1}}{1 - \pi_{i1}} \\ &= \gamma_{01} + \gamma_{11} X_{i1} + \cdots + \gamma_{k1} X_{ik}\end{aligned}$$

- Applied to a dichotomy, the polytomous logit model is identical to the dichotomous logit model.

▶ The following example is adapted from work by Andersen, Heath, and Sinnott (2002) on the 2001 British election.

- The central issue addressed in the data analysis is the potential interaction between respondents' political knowledge and political attitudes in determining their vote.
- The response variable, vote, has three categories: Conservative, Labour, and Liberal Democrat.

- There are several explanatory variables:
  - Attitude toward European integration, an 11-point scale, with high scores representing a negative attitude (so-called "Euro-scepticism").
  - Knowledge of the platforms of the three parties on the issue of European integration, with integer scores ranging from 0 through 3. (Labour and the Liberal Democrats supported European integration, the Conservatives were opposed.)
  - Other variables included in the model primarily as "controls"—age, gender, perceptions of national and household economic conditions, and ratings of the three party leaders.

- The coefficients of a polytomous logit model fit to the BEPS by ML along with their standard errors:

	Labour/Lib.Dem.	
Coefficient	*Estimate*	*Std. Error*
Constant	-0.155	0.612
Age	-0.005	0.005
Gender (male)	0.021	0.144
Perception of Economy	0.377	0.091
Perception of Household Economic Position	0.171	0.082
Evaluation of Blair (Labour leader)	0.546	0.071
Evaluation of Hague (Conservative leader)	-0.088	0.064
Evaluation of Kennedy (Liberal Democrat leader)	-0.416	0.072
Attitude Toward European Integration	-0.070	0.040
Political Knowledge	-0.502	0.155
Europe $\times$ Knowledge	0.024	0.021

	Cons./Lib.Dem.	
Coefficient	*Estimate*	*Std. Error*
Constant	0.718	0.734
Age	0.015	0.006
Gender (male)	-0.091	0.178
Perception of Economy	-0.145	0.110
Perception of Household Economic Position	-0.008	0.101
Evaluation of Blair (Labour leader)	-0.278	0.079
Evaluation of Hague (Conservative leader)	0.781	0.079
Evaluation of Kennedy (Liberal Democrat leader)	-0.656	0.086
Attitude Toward European Integration	-0.068	0.049
Political Knowledge	-1.160	0.219
Europe $\times$ Knowledge	0.183	0.028

- This model differs from those that we have seen previously in that it includes the product of two quantitative explanatory variables, representing the *linear-by-linear interaction* between these variables:
  - Focusing on the Conservative/Liberal-Democrat logit, for example, when political knowledge is 0, the slope for attitude toward European integration (“Europe”) is -0.068.
  - With each unit increase in political knowledge, the slope for Europe increases by 0.183, thus becoming increasingly positive: Those with more knowledge of the parties’ positions are more likely to vote in conformity with their own position on the issue.
  - By the same token, at low levels of Europe, the slope for political knowledge is negative, but it increases by 0.183 with each unit increase in Europe.
  - By a Wald test, this interaction coefficient is highly statistically significant, with $Z = 0.183/0.028 = 6.53$, for which $p \ll .0001$.

- An analysis-of-deviance table for the model:

Source	df	G_0^2	p
Age	2	13.87	.0009
Gender	2	0.45	.78
Perception of Economy	2	30.60	$\ll .0001$
Perception of Household Economic Position	2	5.65	.059
Evaluation of Blair	2	135.37	$\ll .0001$
Evaluation of Hague	2	166.77	$\ll .0001$
Evaluation of Kennedy	2	68.88	$\ll .0001$
Attitude Toward European Integration	2	78.03	$\ll .0001$
Political Knowledge	2	55.57	$\ll .0001$
Europe $\times$ Knowledge	2	50.80	$\ll .0001$

  - All of the terms in the model are highly statistically significant, with the exception of gender and perception of household economic position.

- There are two obstacles to interpreting the coefficient estimates:
  (i) The interaction between political knowledge and attitude toward European integration requires that we combine the estimated coefficient for the interaction with the coefficients for the "main-effect" regressors that are marginal to the interaction
  (ii) The structure of the polytomous logit model, which is for log-odds of pairs of categories (each category versus the base-line Liberal-Democrat category), makes it difficult to formulate a general understanding of the results.
- An effect plot for the interaction of attitude toward European integration with political knowledge is shown in Figure 1.

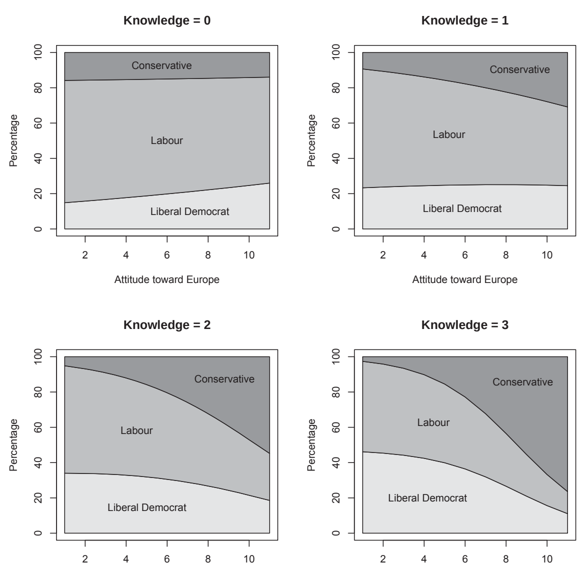

Figure 1. Effect display for the interaction between attitude toward European integration and political knowledge.

4. Nested Dichotomies

▶ Perhaps the simplest approach to polytomous data is to fit separate models to each of a set of dichotomies derived from the polytomy.
- These dichotomies are *nested,* making the models statistically independent.
- Logit models fit to a set of nested dichotomies constitute a model for the polytomy, but are not equivalent to the polytomous logit model previously described.

▶ A nested set of $m-1$ dichotomies is produced from an m-category polytomy by successive binary partitions of the categories of the polytomy.
- Two examples for a four-category variable are shown in Figure 2.
  – In part (a), the dichotomies are {12, 34}, {1, 2}, and {3, 4}.
  – In part (b), the nested dichotomies are {1, 234}, {2, 34}, and {3, 4}.

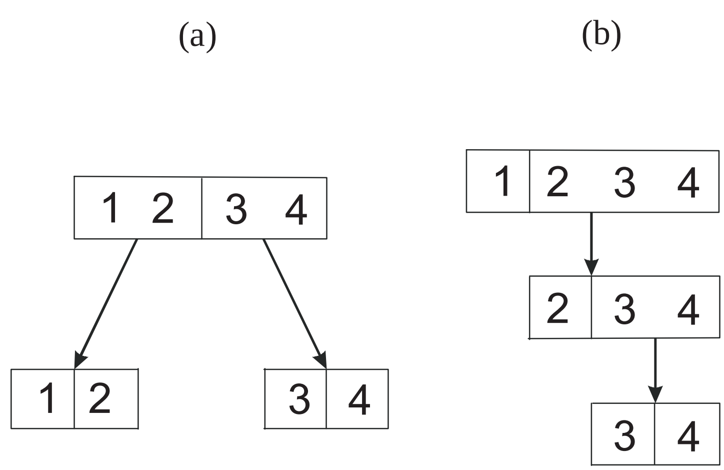

Figure 2. Alternative sets of nested dichotomies for a four-category response.

- Because the results of the analysis and their interpretation depend upon the set of nested dichotomies that is selected, this approach to polytomous data is reasonable only when a particular choice of dichotomies is substantively compelling.
- Nested dichotomies are attractive when the categories of the polytomy represent ordered progress through the stages of a process (called *continuation dichotomies*).
  - Imagine that the categories in (b) represent adults' attained level of education: (1) less than high school; (2) high-school graduate; (3) some post-secondary; (4) post-secondary degree.
  - Since individuals normally progress through these categories in sequence, the dichotomy {1, 234) represents the completion of high school; {2, 34} the continuation to post-secondary education, conditional on high-school graduation; and {3, 4} the completion of a degree conditional on undertaking a post-secondary education.

5. Ordered Logit and Probit Models

- Imagine that there is a latent variable ξ that is a linear function of the X's plus a random error:
$$\xi_i = \alpha + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i$$
  - Suppose that instead of dividing the range of ξ into two regions to produce a dichotomous response, the range of ξ is dissected by $m-1$ boundaries or *thesholds* into m regions.
  - Denoting the thresholds by $\alpha_1 < \alpha_2 < \cdots < \alpha_{m-1}$, and the resulting

response by Y, we observe

$$Y_i = \begin{cases} 1 & \text{if } \xi_i \le \alpha_1 \\ 2 & \text{if } \alpha_1 < \xi_i \le \alpha_2 \\ \cdot & \\ \cdot & \\ \cdot & \\ m-1 & \text{if } \alpha_{m-2} < \xi_i \le \alpha_{m-1} \\ m & \text{if } \alpha_{m-1} < \xi_i \end{cases}$$

- The thresholds, regions, and corresponding values of ξ and Y are represented graphically in Figure 3.

▶ Using the model for the latent variable, along with category thresholds, we can determine the cumulative probability distribution of Y:

$$\begin{aligned} \Pr(Y_i \le j) &= \Pr(\xi_i \le \alpha_j) \\ &= \Pr(\alpha + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i \le \alpha_j) \\ &= \Pr(\varepsilon_i \le \alpha_j - \alpha - \beta_1 X_{i1} - \cdots - \beta_k X_{ik}) \end{aligned}$$

- If the errors ε_i are independently distributed according to the standard

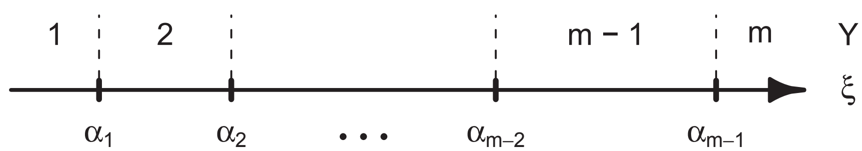

Figure 3. The thresholds $\alpha_1 < \alpha_2 < \cdots < \alpha_{m-1}$ divide the latent continuum ξ into m regions, corresponding to the values of the observable variable Y.

normal distribution, then we obtain the ordered probit model.

- If the errors follow the similar logistic distribution, then we get the ordered logit model:

$$\text{logit}[\Pr(Y_i \leq j)] = \log_e \frac{\Pr(Y_i \leq j)}{\Pr(Y_i > j)} = \alpha_j - \alpha - \beta_1 X_{i1} - \cdots - \beta_k X_{ik}$$

- Equivalently,

$$\text{logit}[\Pr(Y_i > j)] = \log_e \frac{\Pr(Y_i > j)}{\Pr(Y_i \leq j)} = (\alpha - \alpha_j) + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}$$

for $j = 1, 2, ..., m-1$.

- The logits in this model are for cumulative categories — at each point contrasting categories above category j with category j and below.
- The slopes for each of these regression equations are identical; the equations differ only in their intercepts.

  - The logistic regression surfaces are therefore horizontally parallel to each other, as illustrated in Figure 4 for $m = 4$ response categories and a single X.
- For a fixed set of X's, any two different cumulative log-odds — say, at categories j and j' — differ only by the constant $(\alpha_j - \alpha_{j'})$.

– The odds, therefore, are proportional to one-another, and for this reason, the ordered logit model is called the *proportional-odds model*.

▶ There are $(k+1)+(m-1) = k+m$ parameters to estimate in the proportional-odds model, including the regression coefficients $\alpha, \beta_1, ..., \beta_k$ and the category thresholds $\alpha_1, ..., \alpha_{m-1}$.

- There is an extra parameter in the regression equations, since each equation has its own constant, $-\alpha_j$, along with the common constant α.
- A simple solution is to set $\alpha = 0$ (and to absorb the negative sign in α_j), producing

$$\text{logit}[\Pr(Y_i > j)] = \alpha_j + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}$$

▶ Figure 5 illustrates the proportional-odds model for $m = 4$ response categories and a single X.

- The conditional distribution of the latent response variable ξ is shown for two representative values of the explanatory variable, x_1 and x_2.

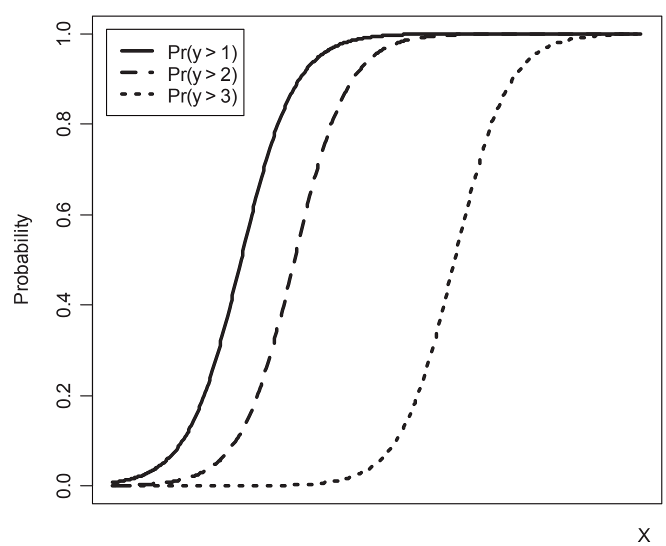

Figure 4. The proportional-odds model for four response categories and a single explanatory variable X.

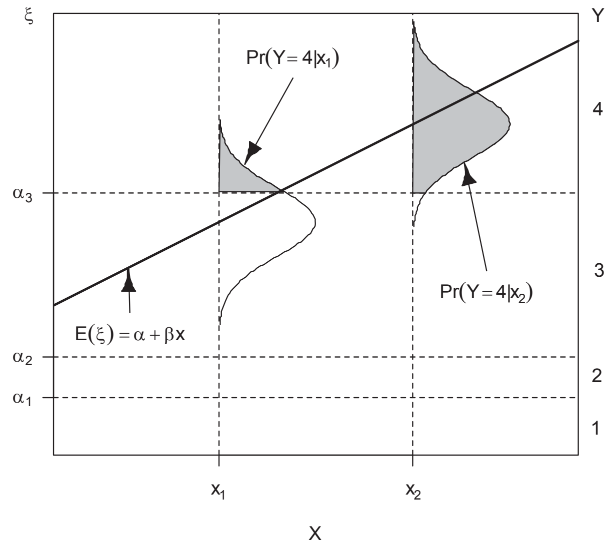

Figure 5. The proportional-odds model for four response categories and a single explanatory variable X.

- To illustrate the use of the proportional-odds model, I draw on data from the World Values Survey (WVS) of 1995–97
  - Although the WVS collects data in many countries, to provide a manageable example, I will restrict attention to only four: Australia, Sweden, Norway, and the United States. The combined sample size for these four countries is 5381.
  - The response variable in the analysis is the answer to the question, "Do you think that what the government is doing for people in poverty is about the right amount, too much, or too little." There are, therefore, three ordered categories: *too little*, *about right*, *too much*.

- There are several explanatory variables:
  - gender (represented by a dummy variable coded 1 for *men* and 0 for *women*);
  - whether or not the respondent belonged to a religion (coded 1 for *yes*, 0 for *no*);
  - whether or not the respondent had a university degree (coded 1 for *yes* and 0 for *no*);
  - age (in years, ranging from 18 to 87); preliminary analysis of the data suggested a roughly linear age effect;
  - country (entered into the model as a set of three dummy regressors, with *Australia* as the base-line category).

- Analysis of deviance for an initial model fit to the data incorporating interactions between country and each of the other explanatory variables:

Source	df	G_0^2	p
Country	3	250.881	$\ll .0001$
Gender	1	10.749	.0010
Religion	1	4.132	.042
Education	1	4.284	.038
Age	1	49.950	$\ll .0001$
Country×Gender	3	3.049	.38
Country×Religion	3	21.143	$< .0001$
Country×Education	3	12.861	.0049
Country×Age	3	17.529	.0005

– As usual, the likelihood-ratio tests in the table are computed by contrasting the deviances for alternative models, with and without the terms in question.
– These tests were formulated in conformity with the principal of marginality (i.e., “Type-II” tests).

- Estimated coefficients and their standard errors for a final model, removing the non-significant interaction between country and gender:

Coefficient	*Estimate*	*Standard Error*
Gender (Men)	0.1744	0.0532
Country (Norway)	0.1516	0.3355
Country (Sweden)	-1.2237	0.5821
Country (United States)	1.2225	0.3068
Religion (Yes)	0.0255	0.1120
Education (Degree)	-0.1282	0.1676
Age	0.0153	0.0026
Country (Norway)×Religion	-0.2456	0.2153
Country (Sweden)×Religion	-0.9031	0.5125
Country (United States)×Religion	0.5706	0.1733
Country (Norway)×Education	0.0524	0.2080
Country (Sweden)×Education	0.6359	0.2141
Country (United States)×Education	0.3103	0.2063

Coefficient	*Estimate*	*Standard Error*
Country (United States)×Education	0.3103	0.2063
Country (Norway)×Age	-0.0156	0.0044
Country (Sweden)×Age	-0.0090	0.0047
Country (United States)×Age	0.0008	0.0040
Thresholds		
$-\widehat{\alpha}_1$ (Too Little $\mid$ About Right)	0.7699	0.1491
$-\widehat{\alpha}_2$ (About Right $\mid$ Too Much)	2.5882	0.1537

- Interpretation of the estimated coefficients for the proportional-odds model is complicated by the interactions in the model and by the multiple-category response.
  – For example, consider the interaction between age and country: The age slope is positive in the base-line country of Australia, this slope is nearly zero in Norway, smaller in Sweden than in Australia, and very slightly larger in the United States than in Australia, but a more detailed understanding of the age-by-country interaction is hard to discern from the coefficients alone.
- Figures 6 and 7 show alternative effect plots for the age-by-country interaction.

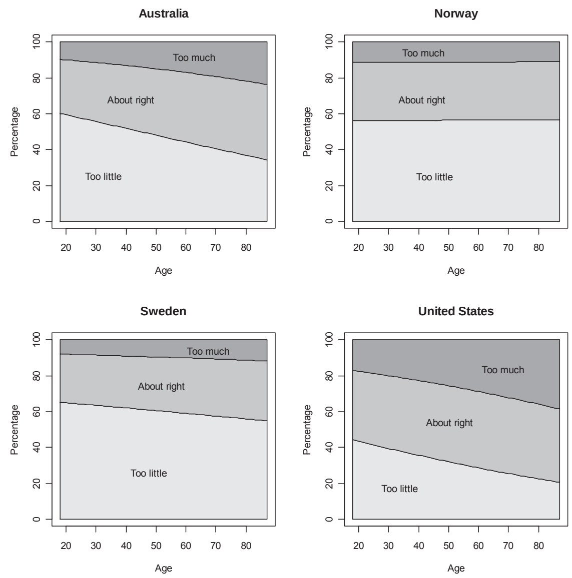

Figure 6. Effect display for the interaction of age with country.

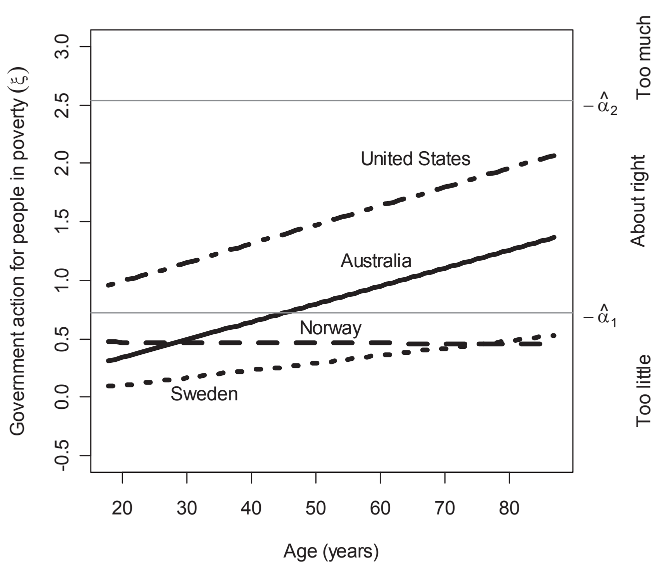

Figure 7. Alternative effect display for the proportional-odds model fit to the World Value Survey data, showing fitted values of the latent response.

6. Comparison of the Three Approaches

- The three approaches to modeling polytomous data — the polytomous logit model, logit models for nested dichotomies, and the proportional-odds model — address different sets of log-odds, corresponding to different dichotomies constructed from the polytomy.
- Consider, for example, the ordered polytomy {1, 2, 3, 4}:
  - Treating category 1 as the baseline, the coefficients of the polytomous logit model apply directly to the dichotomies {1, 2}, {1, 3}, and {1, 4}, and indirectly to any pair of categories.
  - Forming continuation dichotomies (one of several possibilities), the nested-dichotomies approach models {1, 234}, {2, 34}, and {3, 4}.
  - The proportional-odds model applies to the dichotomies {1, 234}, {12, 34}, and {123, 4}, imposing the restriction that only the intercepts of the three regression equations differ.

- Which of these models is most appropriate depends partly on the structure of the data and partly upon our interest in them.

7. Summary

- Several approaches can be taken to modeling polytomous data, including:
  (a) modeling the polytomy directly using a logit model based on the multivariate logistic distribution;
  (b) constructing a set of $m-1$ nested dichotomies to represent the m categories of the polytomy; and
  (c) fitting the proportional-odds model to a polytomous response variable with ordered categories.

