c. Using conditional logistic regression, (i) conduct an exact test of the hypothesis of no effect of lymphocytic infiltration, controlling for the other variables; (ii) compute a 95% confidence interval for the effect. Interpret results.

Table 7.10

Lymphocytic Infiltration	Sex	Osteoblastic Pathology	Disease-Free Yes	Disease-Free No
High	Female	No	3	0
		Yes	2	0
	Male	No	4	0
		Yes	1	0
Low	Female	No	5	0
		Yes	3	2
	Male	No	5	4
		Yes	6	11

Source: LogXact-Turbo User Manual, Cambridge, MA: Cytel Software (1993), p. 5–23.

7.33. For the model $\text{logit}(\pi) = \alpha + \beta x$, let (x_i, y_i) denote the x and y values for subject i, $i = 1, \ldots, N$. Suppose $y_i = 0$ for all x below some point and $y_i = 1$ for all x above that point. Explain intuitively why $\hat{\beta} = \infty$. (*Note:* In technical terms, the sufficient statistics are $\sum y_i$ for α and $\sum x_i y_i$ for β. For a given $\sum y_i$, $\sum x_i y_i$ takes its maximum possible value for this assignment of $\{y_i\}$ to fixed $\{x_i\}$.)

7.34. Refer to $\{\hat{\mu}_{ijk} = n_{i+k}n_{+jk}/n_{++k}\}$ for model (XZ, YZ). Show that these fitted values have the same X-Z and Y-Z marginal totals as the observed data. For $2 \times 2 \times K$ tables, show that $\hat{\theta}_{XY(k)} = 1$. Illustrate these results for the fit of model (AM, CM) to Table 6.3.

7.35. Refer to formula (7.5.3) applied to the independence model for a 2×2 table. Show that for the constraints $\lambda_1^X = \lambda_1^Y = 0$, for which $\boldsymbol{\beta} = (\lambda, \lambda_2^X, \lambda_2^Y)$, $\mathbf{X}$ has rows $(1, 0, 0)$, $(1, 0, 1)$, $(1, 1, 0)$, $(1, 1, 1)$.

7.36. For Table 6.3, show the matrix representation of loglinear model (AC, AM, CM). Specify the parameter constraints used in your model matrix. Show the matrix representation of the corresponding logit model, when M is a response.

7.37. A GLM has form $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$, for a monotone function g. Explain what each symbol in this formula represents for fitting the ordinary linear regression model to n observations on a normally distributed response and a single predictor.

CHAPTER 8

Multicategory Logit Models

This chapter presents generalizations of logistic regression models that handle multicategory responses. Section 8.1 presents models for *nominal* response variables, and Sections 8.2 and 8.3 present models for *ordinal* response variables. As in ordinary logistic regression modeling, explanatory variables can be continuous and/or categorical.

At each combination of levels of the explanatory variables, the models assume that the response counts for the categories of Y have a *multinomial* distribution. This generalization of the binomial applies when the number of response categories exceeds two. Logistic regression models are a special case of these models for binary responses.

Like logistic regression models and unlike loglinear models, multicategory logit models treat one classification as a response and the other variables as explanatory. Nevertheless, when explanatory variables are categorical, some of these models are equivalent to loglinear models for contingency tables.

8.1 LOGIT MODELS FOR NOMINAL RESPONSES

Suppose Y is a nominal variable with J categories. The order of listing the categories is irrelevant. Let $\{\pi_1, \ldots, \pi_J\}$ denote the response probabilities, satisfying $\sum_j \pi_j = 1$. When one takes n independent observations based on these probabilities, the probability distribution for the number of outcomes that occur of each of the J types is the multinomial. It specifies the probability for each possible way of allocating the n observations to the J categories. Here, we simply mention this as a sampling model for the counts, and we will not need to calculate such probabilities. For the case $J = 2$, this is the binomial distribution (1.2.2).

Multicategory (also called *polychotomous*) logit models simultaneously refer to all pairs of categories, and describe the odds of response in one category instead of another. Once the model specifies logits for a certain $(J - 1)$ pairs of categories, the rest are redundant.

8.1.1 Baseline-Category Logits

Logit models for nominal responses pair each response category with a baseline category, the choice of which is arbitrary. When the last category (J) is the baseline, the *baseline-category logits* are

$$\log\left(\frac{\pi_j}{\pi_J}\right), \qquad j = 1, \ldots, J-1.$$

Given that the response falls in category j or category J, this is the log odds that the response is j. For $J = 3$, for instance, the logit model uses $\log(\pi_1/\pi_3)$ and $\log(\pi_2/\pi_3)$.

The logit model using the baseline-category logits with a predictor x has form

$$\log\left(\frac{\pi_j}{\pi_J}\right) = \alpha_j + \beta_j x, \qquad j = 1, \ldots, J-1. \tag{8.1.1}$$

The model consists of $J - 1$ logit equations, with separate parameters for each. That is, the effects vary according to the response category paired with the baseline. When $J = 2$, this model simplifies to a single equation for $\log(\pi_1/\pi_2) = \text{logit}(\pi_1)$, resulting in the ordinary logistic regression model for binary responses.

Parameters in the $J - 1$ equations (8.1.1) determine parameters for logits using all other pairs of response categories. For instance, for an arbitrary pair of categories a and b,

$$\begin{aligned}\log\left(\frac{\pi_a}{\pi_b}\right) &= \log\left(\frac{\pi_a/\pi_J}{\pi_b/\pi_J}\right) = \log\left(\frac{\pi_a}{\pi_J}\right) - \log\left(\frac{\pi_b}{\pi_J}\right)\\ &= (\alpha_a + \beta_a x) - (\alpha_b + \beta_b x)\\ &= (\alpha_a - \alpha_b) + (\beta_a - \beta_b)x. \end{aligned} \tag{8.1.2}$$

Thus, the logit equation for categories a and b has intercept parameter $(\alpha_a - \alpha_b)$ and slope parameter $(\beta_a - \beta_b)$.

For optimal efficiency, one should use software that fits the $J - 1$ logit equations (8.1.1) *simultaneously*. Estimates of the model parameters then have smaller standard errors than when binary logistic regression software fits each component equation in (8.1.1) separately. For simultaneous fitting, the same parameter estimates occur for a pair of categories no matter which category is the baseline.

This logit model is an important tool in marketing research for modeling how subjects choose one of a discrete set of options. A generalization of model (8.1.1) allows the explanatory variables to take different values for different response categories. For instance, the choice of a brand of car would likely depend on price, which would vary among the brand options. This generalized model, often called a *multinomial logit model*, is beyond the scope of this text (see Agresti (1990), Sec. 9.2). The following example deals with discrete choice in a somewhat different context: for alligators rather than humans.

Table 8.1 Alligator Size (Meters) and Primary Food Choice,[a] for 59 Florida Alligators

1.24 I	1.30 I	1.30 I	1.32 F	1.32 F	1.40 F	1.42 I	1.42 F
1.45 I	1.45 O	1.47 I	1.47 F	1.50 I	1.52 I	1.55 I	1.60 I
1.63 I	1.65 O	1.65 I	1.65 F	1.65 F	1.68 F	1.70 I	1.73 O
1.78 I	1.78 I	1.78 O	1.80 I	1.80 F	1.85 F	1.88 I	1.93 I
1.98 I	2.03 F	2.03 F	2.16 F	2.26 F	2.31 F	2.31 F	2.36 F
2.36 F	2.39 F	2.41 F	2.44 F	2.46 F	2.56 O	2.67 F	2.72 I
2.79 F	2.84 F	3.25 O	3.28 O	3.33 F	3.56 F	3.58 F	3.66 F
3.68 O	3.71 F	3.89 F					

[a] F = Fish, I = Invertebrates, O = Other.

Source: Thanks to M. F. Delany and Clint T. Moore for providing these data.

8.1.2 Alligator Food Choice Example

Table 8.1 is taken from a study by the Florida Game and Fresh Water Fish Commission of factors influencing the primary food choice of alligators. For 59 alligators sampled in Lake George, Florida, Table 8.1 shows the alligator length (in meters) and the primary food type, in volume, found in the alligator's stomach. Primary food type has three categories: Fish, Invertebrate, and Other. The invertebrates were primarily apple snails, aquatic insects, and crayfish. The "other" category includes reptiles (primarily turtles, though one stomach contained tags of 23 baby alligators that had been released in the lake during the previous year!), amphibean, mammal, plant material, and stones or other debris.

We apply logit model (8.1.1) with $J = 3$ to these data, using Y = "primary food choice" as the response and X = "length of alligator" as predictor. Table 8.2 shows ML parameter estimates and standard errors for the two logits using "other" as the baseline category. From Table 8.2, $\log(\hat{\pi}_1/\hat{\pi}_3) = 1.618 - 0.110x$, and $\log(\hat{\pi}_2/\hat{\pi}_3) = 5.697 - 2.465x$. By (8.1.2), the estimated log odds that the response is "fish" rather than "invertebrate" equals $\log(\hat{\pi}_1/\hat{\pi}_2) = (1.618 - 5.697) + [-0.110 - (-2.465)]x = -4.080 + 2.355x$.

For each logit, one interprets the estimates just as in ordinary binary logistic regression models, conditional on the event that the response outcome was one of those two categories. For instance, given that the primary food type is fish or invertebrate, the estimated probability that it is fish increases in length x according to an S-shaped curve. For alligators of length $x + 1$ meters, the estimated odds that

Table 8.2 Parameter Estimates and Standard Errors (in parentheses) for Generalized Logit Model Fitted to Table 8.1

	Food Choice Categories for Logit	
Parameter	(Fish/Other)	(Invertebrate/Other)
Intercept	1.618	5.697
Length	−0.110(.517)	−2.465(.900)

primary food type is "fish" rather than "invertebrate" equal exp(2.355) = 10.5 times the estimated odds for aligators of length x meters.

To test the hypothesis that primary food choice is independent of alligator length, we test $H_0 : \beta_j = 0$ for $j = 1, 2$ in model (8.1.1). The likelihood-ratio test takes twice the difference in maximized log likelihoods between this model and the simpler one having response independent of length. The test statistic equals 16.8 with $df = 2$, giving a P-value of .01 and strong evidence of a length effect.

8.1.3 Estimating Response Probabilities

One can alternatively express the multicategory logit model directly in terms of the response probabilities, as

$$\pi_j = \frac{\exp(\alpha_j + \beta_j x)}{\sum_h \exp(\alpha_h + \beta_h x)}, \qquad j = 1, \ldots, J-1. \qquad (8.1.3)$$

The denominator is the same for each probability, and the numerators for various j sum to the denominator, so $\sum_j \pi_j = 1$. The parameters equal zero in (8.1.3) for whichever category is the baseline in the logit expressions. For instance, the model that defines parameters with category J as the baseline for the denominator of each logit (as in (8.1.1)) has $\alpha_J = \beta_J = 0$. For category j, (α_j, β_j) then refers to the logit equation for $\log[\pi_j/\pi_J]$.

The estimates in Table 8.2 contrast "fish" and "invertebrate" to "other" as the baseline category. The estimated probabilities (8.1.3) of the outcomes (Fish, Invertebrate, Other) equal

$$\hat{\pi}_1 = \frac{\exp(1.62 - 0.11x)}{1 + \exp(1.62 - 0.11x) + \exp(5.70 - 2.47x)}$$

$$\hat{\pi}_2 = \frac{\exp(5.70 - 2.47x)}{1 + \exp(1.62 - 0.11x) + \exp(5.70 - 2.47x)}$$

$$\hat{\pi}_3 = \frac{1}{1 + \exp(1.62 - 0.11x) + \exp(5.70 - 2.47x)}$$

The 1 term in each denominator and in the numerator of $\hat{\pi}_3$ represents $\exp(\hat{\alpha}_3 + \hat{\beta}_3 x)$ using $\hat{\alpha}_3 = \hat{\beta}_3 = 0$. The three probabilities sum to 1, since the numerators sum to the common denominator. The logs of the ratios of $[\hat{\pi}_1/\hat{\pi}_3]$ and $[\hat{\pi}_2/\hat{\pi}_3]$ produce the results in Table 8.2.

For an alligator of length $x = 3.9$ meters, for instance, the estimated probability that primary food choice is "other" equals

$$\hat{\pi}_3 = \frac{1}{1 + \exp[1.62 - 0.10(3.9)] + \exp[5.70 - 2.47(3.9)]} = 0.23.$$

Figure 8.1 displays the three response probabilities as a function of length.

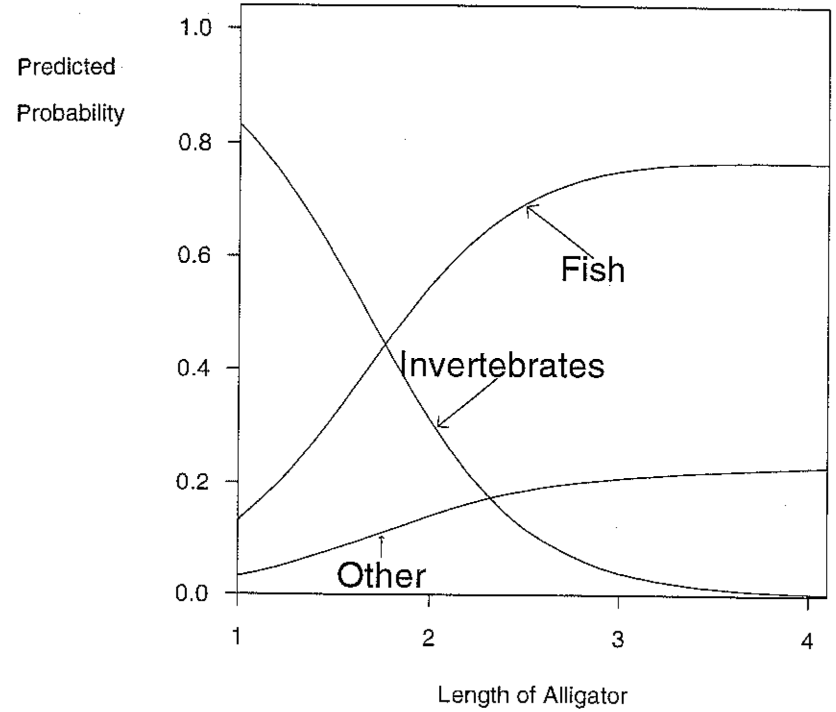

Figure 8.1 Predicted probabilities for primary food choice.

8.1.4 Belief in Afterlife Example

Logit models for nominal responses also apply when explanatory variables are categorical, or a mixture of categorical and continuous. When predictors are entirely categorical, one can display the data as a contingency table. If the data are not sparse, one can test model goodness of fit using X^2 or G^2 statistics. In addition, the models then have equivalent loglinear models.

To illustrate, Table 8.3, taken from the 1991 General Social Survey, has response variable Y = belief in life after death, with categories (Yes, Undecided, No), and explanatory variables X_1 = gender and X_2 = race. We use dummy variables for the predictors, with $x_1 = 1$ for females and 0 for males, and $x_2 = 1$ for whites and 0 for blacks. Using "no" as the baseline category for belief in life after death, we form the model

$$\log\left(\frac{\pi_j}{\pi_3}\right) = \alpha_j + \beta_j^G x_1 + \beta_j^R x_2, \qquad j = 1, 2, \qquad (8.1.4)$$

where G and R superscripts identify the gender and race parameters. The model assumes a lack of interaction between gender and race in their effects on belief in

Table 8.3 Belief in Afterlife, by Race and Gender, with Fitted Values for Generalized Logit Model

		Belief in Afterlife		
Race	Gender	Yes	Undecided	No
White	Female	371	49	74
		(372.8)	(49.2)	(72.1)
	Male	250	45	71
		(248.2)	(44.8)	(72.9)
Black	Female	64	9	15
		(62.2)	(8.8)	(16.9)
	Male	25	5	13
		(26.8)	(5.2)	(11.1)

Source: 1991 General Social Survey.

life after death. The effect parameters represent log odds ratios with the baseline category. For instance, β_1^G is the conditional log odds ratio between gender and response categories 1 and 3 (yes and no), given race; β_2^G is the conditional log odds ratio between gender and response categories 2 and 3 (undecided and no).

Table 8.3 also shows ML fitted values for this model. The goodness-of-fit statistics are $G^2 = 0.8$ and $X^2 = 0.9$. The sample has two non-redundant logits at each of four gender–race combinations, for a total of eight logits. Model (8.1.4), considered for $j = 1$ and 2, contains six parameters. Thus, the model has residual $df = 8 - 6 = 2$. The model fits well. Table 8.4 shows the parameter estimates and their standard errors. The estimated odds of a "yes" rather than a "no" response for females are $\exp(0.419) = 1.5$ times those for males, controlling for race; for whites, they are $\exp(0.342) = 1.4$ times those for blacks, controlling for gender.

To test the effect of gender, we test $H_0 : \beta_j^G = 0$ for $j = 1, 2$. The likelihood-ratio test compares $G^2 = 0.8$ ($df = 2$) to $G^2 = 8.0$ ($df = 4$) obtained by dropping gender from the model. The difference of 7.2, based on $df = 2$, has a P-value of .03 and shows evidence of a gender effect. By contrast, the effect of race is not significant, the model deleting race having $G^2 = 2.8$ ($df = 4$). This partly reflects the larger standard errors that the effects of race have, due to a much greater imbalance between sample sizes in the two race categories than occurs with gender.

Table 8.4 Parameter Estimates and Standard Errors (in parentheses) for Generalized Logit Model Fitted to Table 8.3

	Belief Categories for Logit	
Parameter	(Yes/No)	(Undecided/No)
Intercept	0.883 (.243)	−0.758 (.361)
Gender ($F = 1$)	0.419 (.171)	0.105 (.246)
Race ($W = 1$)	0.342 (.237)	0.271 (.354)

Table 8.5 Predicted Probabilities for Belief in Afterlife

		Belief in Afterlife		
Race	Gender	Yes	Undecided	No
White	Female	0.76	0.10	0.15
	Male	0.68	0.12	0.20
Black	Female	0.71	0.10	0.19
	Male	0.62	0.12	0.26

Table 8.5 displays predicted probabilities for the three response categories. To illustrate, for white females ($x_1 = x_2 = 1$), the estimated probability of response 1 ("yes") equals

$$\frac{\exp[0.883 + 0.419(1) + 0.342(1)]}{1 + \exp[0.883 + 0.419(1) + 0.342(1)] + \exp[-0.758 + 0.105(1) + 0.271(1)]} = .76.$$

This also follows directly from the fitted values, as $372.8/(372.8 + 72.1 + 49.2) = .76$.

8.1.5 Connection with Loglinear Models

When all explanatory variables are categorical, logit models have corresponding loglinear models. The loglinear model has the most general interaction among the explanatory variables from the logit model. It has the same association and interaction structure relating the explanatory variables to the response.

To illustrate, the model fitted to Table 8.3 assumes main effects of gender (G) and race (R) on belief (B) in afterlife, with no interaction. It corresponds to the loglinear model (GR, BG, BR) of homogeneous association. The same estimated effect of G on B results from logit or loglinear parameters. For instance, model (GR, BG, BR) has estimated conditional odds ratio relating B categories 1 = "yes" and 3 = "no" to G equal to $\exp(\hat{\lambda}_{11}^{BG} + \hat{\lambda}_{32}^{BG} - \hat{\lambda}_{12}^{BG} - \hat{\lambda}_{31}^{BG}) = 1.5$, as obtained above for the logit model.

The simpler logit model that deletes the race effect on belief corresponds to loglinear model (GR, BG). For this logit model and for the one having both G and R as predictors of belief, the corresponding loglinear model contains the G-R term. This term represents the interaction among variables that are explanatory in the logit model.

8.2 CUMULATIVE LOGIT MODELS FOR ORDINAL RESPONSES

When response categories are ordered, logits can directly incorporate the ordering. This results in models having simpler interpretations and potentially greater power than ordinary multicategory logit models.

The *cumulative probabilities* are the probabilities that the response Y falls in category j or below, for each possible j. The jth cumulative probability is

$$P(Y \le j) = \pi_1 + \cdots + \pi_j, \qquad j = 1, \ldots, J.$$

The cumulative probabilities reflect the ordering, with $P(Y \leq 1) \leq P(Y \leq 2) \leq \cdots \leq P(Y \leq J) = 1$. Models for cumulative probabilities do not use the final one, $P(Y \leq J)$, since it necessarily equals 1. The logits of the first $J - 1$ cumulative probabilities are

$$\text{logit}[P(Y \leq j)] = \log\left(\frac{P(Y \leq j)}{1 - P(Y \leq j)}\right)$$
$$= \log\left(\frac{\pi_1 + \cdots + \pi_j}{\pi_{j+1} + \cdots + \pi_J}\right), \qquad j = 1, \ldots, J-1.$$

These are called *cumulative logits*.

Each cumulative logit uses all J response categories. A model for the jth cumulative logit looks like an ordinary logit model for a binary response in which categories 1 to j combine to form a single category, and categories $j+1$ to J form a second category. In other words, the response collapses into two categories. Ordinal models simultaneously provide a structure for all $J - 1$ cumulative logits. For $J = 3$, for instance, models refer both to $\log[\pi_1/(\pi_2 + \pi_3)]$ and $\log[(\pi_1 + \pi_2)/\pi_3]$.

8.2.1 Proportional Odds Model

For a predictor X, the model

$$\text{logit}[P(Y \leq j)] = \alpha_j + \beta x, \qquad j = 1, \ldots, J-1, \qquad (8.2.1)$$

has parameter β describing the effect of X on the log odds of response in category j or below. In this formula, β does not have a j subscript, so the model assumes an identical effect of X for all $J - 1$ collapsings of the response into binary outcomes. When this model fits well, it requires a single parameter rather than $J - 1$ parameters to describe the effect of X.

Interpretations for this model refer to odds ratios for the collapsed response scale, for any fixed j. For two values x_1 and x_2 of X, the odds ratio utilizes cumulative probabilities and their complements,

$$\frac{P(Y \leq j \mid X = x_2)/P(Y > j \mid X = x_2)}{P(Y \leq j \mid X = x_1)/P(Y > j \mid X = x_1)}.$$

The log of this odds ratio is the difference between the cumulative logits at those two values of x. This equals $\beta(x_2 - x_1)$, proportional to the distance between the x values. The same proportionality constant (β) applies for each possible point j for the collapsing. Because of this property, model (8.2.1) is called a *proportional odds model*. In particular, for $x_2 - x_1 = 1$, the odds of response below any given category multiply by e^{β} for each unit increase in X. When the model holds with $\beta = 0$, X and Y are statistically independent.

Explanatory variables in cumulative logit models can be continuous, categorical, or of both types. The ML fitting process uses an iterative algorithm simultaneously

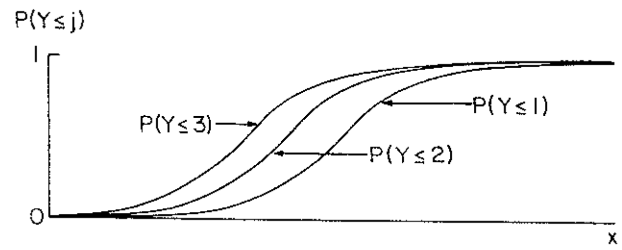

Figure 8.2 Depiction of cumulative probabilities in proportional odds model.

for all j. When the categories are reversed in order, one gets the same fit, but the sign of $\hat{\beta}$ reverses.

Figure 8.2 depicts the proportional odds model for a four category response and a single continuous x predictor. A separate curve applies to each cumulative probability, describing its change as a function of x. The curve for $P(Y \leq j)$ looks like a logistic regression curve for a binary response with pair of outcomes $(Y \leq j)$ and $(Y > j)$. The common effect β for each j implies that the three response curves have the same shape. Any one curve is identical to any of the others simply shifted to the right or shifted to the left. As in logistic regression, the size of $|\beta|$ determines how quickly the curves climb or drop. At any fixed x value, the curves have the same ordering as the cumulative probabilities, the one for $P(Y \leq 1)$ being lowest.

Figure 8.2 has $\beta > 0$. Figure 8.3 shows corresponding curves for the category probabilities, $P(Y = j) = P(Y \leq j) - P(Y \leq j - 1)$. As x increases, the response on Y is more likely to fall at the low end of the ordinal scale. When $\beta < 0$, the

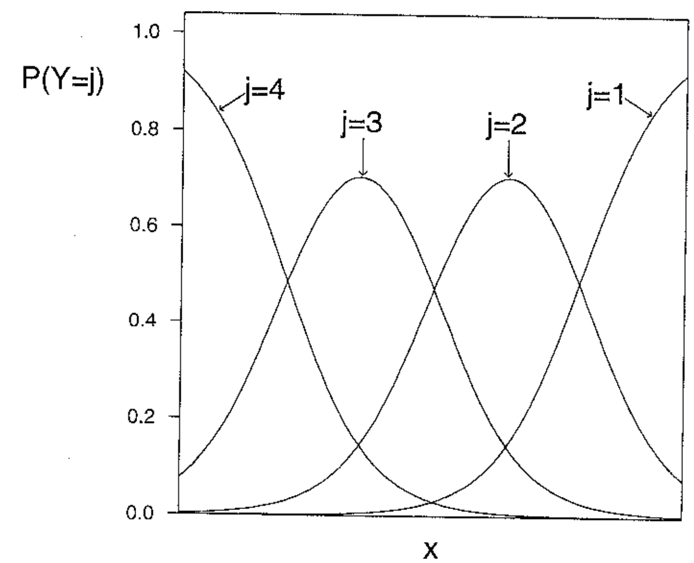

Figure 8.3 Depiction of category probabilities in proportional odds model.

curves in Figure 8.2 descend rather than ascend, and the labels in Figure 8.3 reverse order. Then, as x increases, Y is more likely to fall at the high end of the scale. (The model is sometimes written instead as $\text{logit}[P(Y \le j)] = \alpha_j - \beta x$, so that $\beta > 0$ corresponds to Y being more likely to fall at the high end of the scale as x increases.)

8.2.2 Political Ideology Example

Table 8.6, from the 1991 General Social Survey, relates political ideology to political party affiliation. Political ideology uses a five-point ordinal scale, ranging from very liberal to very conservative. Let x be a dummy variable for political party, with $x = 1$ for Democrats and $x = 0$ for Republicans. The ML fit of the proportional odds model (8.2.1) has estimated effect $\hat{\beta} = 0.975$ ($ASE = 0.129$). For any fixed j, the estimated odds that a Democrat's response is in the liberal direction rather than the conservative direction (i.e., $Y \le j$ rather than $Y > j$) equal $\exp(0.975) = 2.65$ times the estimated odds for Republicans. A 95% confidence interval for this odds ratio equals $\exp(0.975 \pm 1.96 \times 0.129)$, or $(2.1, 3.4)$. A fairly substantial association exists, Democrats tending to be more liberal than Republicans.

Table 8.6 also displays the fitted values for the model. These have an odds ratio of $\exp(\hat{\beta}) = 2.65$ for each of the four collapsings of the data to a 2×2 table. For instance, for the estimated odds of a very liberal response,

$$\frac{(78.4)(44.0 + 151.7 + 75.5 + 104.0)}{(83.2 + 168.2 + 49.1 + 49.1)(31.8)} = 2.65.$$

One can use the fitted values or parameter estimates to calculate estimated probabilities. The cumulative probabilities equal

$$P(Y \le j) = \frac{\exp(\alpha_j + \beta x)}{1 + \exp(\alpha_j + \beta x)}.$$

For instance, $\hat{\alpha}_1 = -2.469$, so the first estimated cumulative probability for Democrats ($x = 1$) equals

$$\frac{\exp[-2.469 + .975(1)]}{1 + \exp[-2.469 + .975(1)]} = .18.$$

From the fitted values for the 428 Democrats, this also equals 78.4/428.

Table 8.6 Political Ideology by Party Affiliation, with Fitted Values for Cumulative Logit Model

	Political Ideology					
Party Affiliation	Very Liberal	Slightly Liberal	Moderate	Slightly Conservative	Very Conservative	Total
Democratic	80	81	171	41	55	428
	(78.4)	(83.2)	(168.2)	(49.1)	(49.1)	
Republican	30	46	148	84	99	407
	(31.8)	(44.0)	(151.7)	(75.5)	(104.0)	

Source: 1991 General Social Survey.

The model fits Table 8.6 well, with $G^2 = 3.7$ and $X^2 = 3.7$, based on $df = 3$. The model with $\beta = 0$ specifies independence between ideology and party affiliation, and is equivalent to the loglinear model of independence. The G^2 test of fit of that special case is simply the G^2 test of independence (2.4.3) for two-way contingency tables, which equals $G^2 = 62.3$ based on $df = 4$. The model permitting an effect fits considerably better than the independence model.

The likelihood-ratio statistic for an ordinal test of independence ($H_0 : \beta = 0$) is the difference between G^2 values for the independence model and the proportional odds model. The difference in G^2 values of $62.3 - 3.7 = 58.6$, based on $df = 4 - 3 = 1$, gives extremely strong evidence of an association ($P < .0001$). Like the test for the ordinal loglinear model presented in Section 7.2.3, this test uses the ordering of the response categories. When the model fits well, it is more powerful than the tests of Section 2.4 based on $df = (I - 1)(J - 1)$, since it focuses on a restricted alternative and has only a single degree of freedom. Similar strong evidence results from the Wald test, using $z^2 = (\hat{\beta}/ASE)^2 = (0.975/0.129)^2 = 57.1$.

8.2.3 Invariance to Choice of Response Categories

When the proportional odds model holds for a given response scale, it also holds with the same effects for any collapsing of the response categories. For instance, if a model for categories (Very liberal, Slightly liberal, Moderate, Slightly conservative, Very conservative) fits well, approximately the same estimated effects result from collapsing the response scale to (Liberal, Moderate, Conservative). This *invariance* to the choice of response categories is a nice feature of the model. Two researchers who use different response categories in studying an association should reach similar conclusions.

To illustrate, we collapse Table 8.6 to a three-category response, combining the two liberal categories and combining the two conservative categories. Then, the ML estimated effect of party affiliation changes only from 0.975 ($ASE = .129$) to 1.006 ($ASE = 0.132$). Interpretations are unchanged. Some loss of efficiency occurs in collapsing ordinal scales, resulting in larger standard errors. In practice, when observations are spread fairly evenly among the categories, the efficiency loss is minor unless one collapses to a binary response. It is usually inadvisable to collapse ordinal data to binary.

The proportional odds model implies trends upward or downward among distributions of Y at different values of explanatory variables. When X refers to two groups, as in Table 8.6, the model fits well when subjects in one group tend to make higher responses on the ordinal scale than subjects in the other group. The model does not fit well when the response distributions differ in their dispersion rather than their average. If Democrats tended to be primarily moderate in ideology, while Republicans tended to be both very conservative and very liberal, then the Republicans' responses would show greater dispersion than the Democrats'. The two ideology distributions would be quite different, but the proportional odds model would not detect this if the average responses were similar.

When a proportional odds model does not fit well, an improved fit may result from a more complex model having quadratic or interaction terms, or from a different link that allows $P(Y \le j)$ to approach 1 at a different rate than it approaches 0 and permits a nonsymmetric appearance for the category probability curves of Figure 8.3 (Agresti (1990), Section 9.5). When simplified ordinal models do not fit well, one can fit ordinary multicategory logit models of form (8.1.1) and use the ordinality in an informal way in interpreting the associations.

8.3 PAIRED-CATEGORY LOGITS FOR ORDINAL RESPONSES*

Cumulative logit models for ordinal responses use the entire response scale in forming each logit. This section presents alternative logits for ordered categories that, like baseline-category logits for nominal responses, use pairs of categories.

8.3.1 Adjacent-Categories Logits

One alternative forms $J - 1$ logits using all pairs of adjacent categories. The *adjacent-categories logits* are

$$\log\left(\frac{\pi_{j+1}}{\pi_j}\right), \qquad j = 1, \ldots, J-1.$$

For $J = 3$, for instance, the logits are $\log(\pi_2/\pi_1)$ and $\log(\pi_3/\pi_2)$. A model using these logits with a predictor x has form

$$\log\left(\frac{\pi_{j+1}}{\pi_j}\right) = \alpha_j + \beta_j x, \qquad j = 1, \ldots, J-1. \tag{8.3.1}$$

These logits, like the baseline-category logits, determine logits for all pairs of response categories. Model (8.3.1) is equivalent to logit model (8.1.1), with β_j in (8.3.1) being equivalent to $\beta_{j+1} - \beta_j$ in (8.1.1).

A simpler version of model (8.3.1),

$$\log\left(\frac{\pi_{j+1}}{\pi_j}\right) = \alpha_j + \beta x, \qquad j = 1, \ldots, J-1. \tag{8.3.2}$$

has identical effects $\{\beta_j = \beta\}$ for each pair of adjacent categories. For this model, the effect of X on the odds of making the higher instead of the lower response is the same for all pairs of adjacent categories. This model and proportional odds model (8.2.1) use a single parameter, rather than $J - 1$ parameters, for the effect of X. When the model holds, independence is equivalent to $\beta = 0$.

The simpler adjacent-categories logit model (8.3.2) implies that the coefficient of x for the logit based on arbitrary response categories a and b equals $\beta(a - b)$. The effect depends on the distance between categories, so this model recognizes the ordering of the response scale. One can check whether this simplification is justified by comparing the model fit to that of (8.3.1), or equivalently to logit model (8.1.1).

8.3.2 Political Ideology Example Revisited

To illustrate the adjacent-categories logit model, we return to Table 8.6 and model political ideology using the simple model (8.3.2) having a common effect for each logit. Let $x = 0$ for Democrats and $x = 1$ for Republicans.

The ML estimate of the party affiliation effect is $\hat{\beta} = 0.435$. The estimated odds that a Republican's ideology classification is in category $j + 1$ instead of j are $\exp(\hat{\beta}) = 1.54$ times the estimated odds for Democrats. This is the estimated odds ratio for each of the four 2×2 tables consisting of a pair of adjacent columns of Table 8.6. For instance, the estimated odds of "slightly conservative" instead of "moderate" ideology are 54% higher for Republicans than for Democrats. The estimated odds ratio for an arbitrary pair of columns $a > b$ equals $\exp[\hat{\beta}(a - b)]$. The estimated odds that a Republican's ideology is "very conservative" (category 5) instead of "very liberal" (category 1) are $\exp[0.435(5 - 1)] = (1.54)^4 = 5.7$ times those for Democrats. Republicans tend to be much more conservative than Democrats.

The model fit has $G^2 = 5.5$ with $df = 3$, a reasonably good fit. The special case of the model with $\beta = 0$ specifies independence of ideology and party affiliation and is equivalent to the loglinear model of independence. The G^2 fit of that model is simply the G^2 statistic (2.4.3) for testing independence, which equals $G^2 = 62.3$ with $df = 4$. The model permitting a party affiliation effect fits much better than the independence model.

The likelihood-ratio test statistic for the hypothesis that party affiliation has no effect on ideology ($H_0: \beta = 0$) is based on the difference between the G^2 values for the two models. It equals $62.3 - 5.5 = 56.8$ with $df = 4 - 3 = 1$. There is very strong evidence of an association ($P < .0001$). Results are similar to those for the cumulative-logit analysis in Section 8.2.2.

For two-way contingency tables, the more general model (8.3.1) and the equivalent logit model (8.1.1) are saturated. An advantage of using the single effect parameter is that the simpler model is unsaturated.

8.3.3 Connection with Loglinear Models

Many loglinear models for ordinal variables have simple representations as adjacent-category logit models. For instance, Section 7.2.1 introduced the linear-by-linear association model,

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \beta u_i v_j, \tag{8.3.3}$$

for which the association term uses monotone scores for the rows and columns. For this model with column scores $\{v_j = j\}$, the adjacent-category logits within row i are

$$\begin{aligned}\log\left(\frac{\pi_{j+1}}{\pi_j}\right) &= \log\left(\frac{\mu_{i,j+1}}{\mu_{ij}}\right)\\ &= \log(\mu_{i,j+1}) - \log(\mu_{ij})\\ &= (\lambda + \lambda_i^X + \lambda_{j+1}^Y + \beta u_i v_{j+1}) - (\lambda + \lambda_i^X + \lambda_j^Y + \beta u_i v_j)\\ &= (\lambda_{j+1}^Y - \lambda_j^Y) + \beta u_i.\end{aligned}$$

This has form (8.3.2), identifying α_j with $(\lambda_{j+1}^Y - \lambda_j^Y)$ and the row scores $\{u_i\}$ with the levels of x. In fact, the two models are equivalent. The logit representation (8.3.2) provides an interpretion for model (8.3.3). The conclusions presented above for fitting logit model (8.3.2) to Table 8.6 also result from fitting the loglinear model (8.3.3) with equally-spaced column scores.

Analogous remarks apply to multi-way tables. For instance, consider Table 7.5 on job satisfaction (S), income (I), and gender (G), analyzed in Section 7.3. Unlike the loglinear models described there, logit models treat job satisfaction as a response variable and income and gender as explanatory variables. Let g denote a dummy variable for gender, and let $\{x_i\}$ denote scores assigned to the levels of income. Consider the model for job satisfaction response probabilities,

$$\log\left(\frac{\pi_{j+1}}{\pi_j}\right) = \alpha_j + \beta_1 x_i + \beta_2 g, \qquad j = 1, 2, 3. \tag{8.3.4}$$

The model assumes identical effects of explanatory variables for each adjacent pair of response categories.

This model is equivalent to a special case of loglinear model (GI, GS, IS) in which the G-S and I-S associations each have linear-by-linear ($L \times L$) form, with equally-spaced scores for S and the scores $\{x_i\}$ for I and g for G. That special case is also a special case of the homogeneous $L \times L$ model (7.3.1) (which assumed $L \times L$ structure for the I-S association) that also provides linear-by-linear structure for the G-S association. Using income categories $\{1, 2, 3, 4\}$ as scores, model (8.3.4) fits well, with $G^2 = 12.6$ having $df = 19$. It provides estimates $\hat{\beta}_1 = 0.389$ ($ASE = 0.155$) and $\hat{\beta}_2 = 0.045$ ($ASE = 0.314$). Satisfaction is positively associated with income, given gender, but shows no evidence of association with gender, given income.

8.3.4 Continuation-Ratio Logits

Another approach forms logits for ordered response categories in a sequential manner. One constructs models simultaneously for

$$\log\left(\frac{\pi_1}{\pi_2}\right), \log\left(\frac{\pi_1 + \pi_2}{\pi_3}\right), \ldots, \log\left(\frac{\pi_1 + \cdots + \pi_{J-1}}{\pi_J}\right).$$

These are called *continuation-ratio logits*. They refer to a binary response that contrasts each category with a grouping of categories from lower levels of the response scale. A second type of continuation-ratio logit contrasts each category with a grouping of categories from higher levels of the response scale; that is,

$$\log\left(\frac{\pi_1}{\pi_2 + \cdots + \pi_J}\right), \log\left(\frac{\pi_2}{\pi_3 + \cdots + \pi_J}\right), \ldots, \log\left(\frac{\pi_{J-1}}{\pi_J}\right).$$

Models using these logits have different parameter estimates and goodness-of-fit statistics than models using the other continuation-ratio logits.

Table 8.7 Outcomes for Pregnant Mice in Developmental Toxicity Study

Concentration	Response		
(mg/kg per day)	Dead	Malformation	Normal
0 (controls)	15	1	281
62.5	17	0	225
125	22	7	283
250	38	59	202
500	144	132	9

Source: Based on results in C. J. Price et al., *Fund. Appl. Toxicol., 8:* 115–126 (1987).
I thank Dr. Louise Ryan for showing me these data.

We illustrate continuation-ratio logits using Table 8.7, which refers to a developmental toxicity study. Rodent studies are commonly used to test and regulate substances posing potential danger to developing fetuses. This study administered diethylene glycol dimethyl ether, an industrial solvent used in the manufacture of protective coatings, to pregnant mice. Each mouse was exposed to one of five concentration levels for ten days early in the pregnancy. Two days later, the uterine contents of the pregnant mice were examined for defects. Each fetus had the three possible outcomes (Dead, Malformation, Normal). The outcomes are ordered, normal being the most desirable result.

We use continuation-ratio logits to model the probability of a dead fetus, using $\log[\pi_1/(\pi_2 + \pi_3)]$, and the conditional probability of a malformed fetus, given that the fetus was live, using $\log(\pi_2/\pi_3)$. We used scores $\{0, 62.5, 125, 250, 500\}$ for concentration level. The two models are ordinary logistic regression models in which the responses are column 1 and columns 2–3 combined for one fit and column 2 and column 3 for the second fit. The estimated effect of concentration level is 0.0064 ($ASE = 0.0004$) for the first logit, and 0.0174 ($ASE = 0.0012$) for the second logit. In each case, the less desirable outcome is more likely as concentration level increases. For instance, given that a fetus was live, the estimated odds that it was malformed rather than normal changes by a multiplicative factor of $\exp(100 \times 0.0174) = 5.7$ for every 100-unit increase in concentration level.

When models for different continuation-ratio logits have separate parameters, as in this example, separate fitting of models for different logits gives the same results as simultaneous fitting. The sum of the separate G^2 statistics is an overall goodness-of-fit statistic pertaining to the simultaneous fitting of the models. For Table 8.7, the G^2 values are 5.8 for the first logit and 6.1 for the second, each based on $df = 3$. We summarize the fit by their sum, $G^2 = 11.8$, based on $df = 6$.

This analysis treats pregnancy outcomes for different fetuses as independent observations. In fact, each pregnant mouse had a litter of fetuses, and statistical dependence may exist among different fetuses in the same litter. The model also treats different fetuses at a given concentration level as having the same response probabilities. Heterogeneity of various types among the litters (for instance, due to different physical conditions of different pregnant mice) would usually cause these probabilities to vary somewhat among litters. Either statistical dependence or heterogeneous probabilities violates the binomial assumption and typically causes *overdispersion*—greater variation than the binomial model predicts. For example, at a fixed concentration level, the number of mice that die in a litter may vary among pregnant mice to a greater degree than if the counts were independent and identical binomial variates.

The total G^2 for testing the continuation-ratio model shows evidence of lack of fit. The structural form chosen for the model may be incorrect. The lack of fit may partly or entirely, however, reflect overdispersion caused by dependence within litters or heterogeneity among litters. Both factors are common in developmental toxicity studies. See Collett ((1992), Ch. 6) and Morgan ((1992), Ch. 6) for ways of handling overdispersion.

PROBLEMS

8.1. Refer to the example in Section 8.1.2.

- **a.** Using the model fit, calculate an odds ratio that describes the estimated effect of length on primary food choice being either "invertebrate" or "other."
- **b.** Estimate the probability that food choice is "invertebrate," for an alligator of length 3.9 meters.
- **c.** Find the length at which the outcomes "invertebrate" and "other" are equally likely.

8.2. Table 8.8 displays primary food choice for a sample of alligators, classified by length ($\leq$ 2.3 meters, $>$ 2.3 meters) and by the lake in Florida in which they were caught.

Table 8.8

Lake	Size	Primary Food Choice: Fish	Invertebrate	Reptile	Bird	Other
Hancock	≤ 2.3	23	4	2	2	8
	> 2.3	7	0	1	3	5
Oklawaha	≤ 2.3	5	11	1	0	3
	> 2.3	13	8	6	1	0
Trafford	≤ 2.3	5	11	2	1	5
	> 2.3	8	7	6	3	5
George	≤ 2.3	16	19	1	2	3
	> 2.3	17	1	0	1	3

Source: Wildlife Research Laboratory, Florida Game and Fresh Water Fish Commission.

- **a.** Use a logit model to describe effects of length and lake on primary food choice.
- **b.** Using your model, estimate the probability that the primary food choice is "fish," for the various length and lake combinations. Interpret.
- **c.** Which loglinear model is equivalent to this logit model?

8.3. Refer to the example in Section 8.1.4.

- **a.** Using the fitted model, estimate the probability of response "no" for black females, (i) using parameter estimates, (ii) using fitted values.
- **b.** Fit the logit model for which race has no effect on belief in an afterlife, controlling for gender. Test the goodness of fit, and interpret. Which loglinear model is equivalent to this model?
- **c.** Describe the gender effect for the model in **(b)** by reporting the estimated odds ratio for each pair of response categories. Interpret, and explain how these relate to the marginal gender–belief odds ratios.
- **d.** Does the model in **(b)** imply that race and belief in an afterlife are marginally independent, collapsing over gender? (*Hint:* Consider collapsibility for the corresponding loglinear model.)

8.4. Refer to Problem 6.7. Treating vote for President (P) as the response variable and B and D as explanatory variables, find a logit model that fits these data well. Interpret parameter estimates for the effects of B on P. Which loglinear model is equivalent to this model? Test the significance of the B-P conditional association, and interpret.

8.5. Refer to Problem 2.21 with Table 2.15. Using modeling methods, describe and make inferences about the effect of mammography experience on women's attitudes.

8.6. Table 8.9 shows results of a survey of women, relating frequency of breast self-examination and age. Fit a proportional odds model. Analyze goodness of fit, and conduct descriptive and inferential analyses about the association.

Table 8.9

Age	Frequency of Breast Self-Examination: Monthly	Occasionally	Never
< 45	91	90	51
45–59	150	200	155
60+	109	198	172

Source: R. T. Senie et al., *Am. J. Public Health, 71*: 583–590 (1981). Reprinted with permission of the American Public Health Association.

8.7. For a sample of 40 subjects, Table 8.10 relates Y = mental impairment (1 = None , 2 = Mild, 3 = Moderate, 4 = Impaired) to socioeconomic status (SES = 1, high; SES = 0, low) and a life events index, which is a composite measure of both the number and severity of important life events (such as loss of job, divorce, death in family) that the subject experienced within the past three years.

- **a.** Fit a proportional odds model, and interpret effects.
- **b.** Add an interaction term between the predictors. Interpret this model by describing the impact of life events on mental impairment separately at each level of SES. Check whether this model provides an improved fit.