CHAPTER 6

Multicategory Log

ordinal variables (response): categories with an order

nominal variables: categories with no order

Logistic regression is used to model binary response variables. Generalizations of it model categorical responses with more than two categories. We will now study models for *nominal* response variables in Section 6.1 and for *ordinal* response variables in Sections 6.2 and 6.3. As in ordinary logistic regression, explanatory variables can be categorical and/or quantitative.

At each setting of the explanatory variables, the multicategory models assume that the counts in the categories of Y have a *multinomial* distribution. This generalization of the binomial distribution applies when the number of categories exceeds two (see Section 1.2.2).

6.1 LOGIT MODELS FOR NOMINAL RESPONSES

Let J denote the number of categories for Y. Let $\{\pi_1, \ldots, \pi_J\}$ denote the response probabilities, satisfying $\sum_j \pi_j = 1$. With n independent observations, the probability distribution for the number of outcomes of the J types is the multinomial. It specifies the probability for each possible way the n observations can fall in the J categories. Here, we will not need to calculate such probabilities.

Multicategory logit models simultaneously use all pairs of categories by specifying the odds of outcome in one category instead of another. For models of this section, the order of listing the categories is irrelevant, because the model treats the response scale as *nominal* (unordered categories).

6.1.1 Baseline-Category Logits

Logit models for nominal response variables pair each category with a baseline category. When the last category (J) is the baseline, the *baseline-category logits*

An Introduction to Categorical Data Analysis, Second Edition. By Alan Agresti

are

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

interpretation of \beta_j: change in log-odds of category j vs baseline J, from one unit increase in x

[illegible]nse falls in category j or category J, this is the log odds that the [illegible] $= 3$, for instance, the model uses $\log($[illegible]

recall in logistic regression: we're building a regression model for Y(x) = Bernoulli random variable

baseline logit model: we're building a regression model on the multinomial random variable (which takes values from {1, ..., J}

The baseline-category logit model with a predictor x i[illegible]

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

The model has $J-1$ equations, with separate parameters [illegible] The effects vary according to the 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 ordinary logistic regression for binary responses.

The equations (6.1) for these pairs of categories determine equations for all other pairs of categories. For example, for an arbitrary pair of categories a and b,

for any choice of baseline category, we will still get the same modeling equations

$$\begin{aligned}\log\left(\frac{\pi_a}{\pi_b}\right) = \log\left(\frac{\pi_a/\pi_J}{\pi_b/\pi_J}\right) &= \log\left(\frac{\pi_a}{\pi_J}\right) - \log\left(\frac{\pi_b}{\pi_J}\right) \\ &= (\alpha_a + \beta_a x) - (\alpha_b + \beta_b x) \\ &= (\alpha_a - \alpha_b) + (\beta_a - \beta_b)x \end{aligned} \tag{6.2}$$

So, the equation for categories a and b has the form $\alpha + \beta x$ with intercept parameter $\alpha = (\alpha_a - \alpha_b)$ and with slope parameter $\beta = (\beta_a - \beta_b)$.

Software for multicategory logit models fits all the equations (6.1) *simultaneously*. Estimates of the model parameters have smaller standard errors than when binary logistic regression software fits each component equation in (6.1) separately. For simultaneous fitting, the same parameter estimates occur for a pair of categories no matter which category is the baseline. The choice of the baseline category is arbitrary.

input: alligator length
output: category of food that it eats (fish, invertebrate, other) - nominal categorical variable

6.1.2 Example: Alligator Food Choice

Table 6.1 comes from a study by the Florida Game and Fresh Water Fish Commission of the foods that alligators in the wild choose to eat. For 59 alligators sampled in Lake George, Florida, Table 6.1 shows 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 included amphibian, mammal, plant material, stones or other debris, and reptiles (primarily turtles, although one stomach contained the tags of 23 baby alligators that had been released in the lake during the previous year!). The table also shows the alligator length, which varied between 1.24 and 3.89 meters.

Table 6.1. 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. M

R function: multinom( )
https://stats.idre.ucla.edu/r/dae/multinomial-logistic-regression/

Let Y = primary food choice and x = alligator length. For model (6.1) with $J = 3$, Table 6.2 shows some output (from PROC LOGISTIC in SAS), with "other" as the baseline category. The ML prediction equations are

$$\log(\hat{\pi}_1/\hat{\pi}_3) = 1.618 - 0.110x$$

and

$$\log(\hat{\pi}_2/\hat{\pi}_3) = 5.697 - 2.465x$$

Table 6.2. Computer Output for Baseline-Category Logit Model with Alligator Data

```
          Testing Global Null Hypothesis: BETA = 0
     Test                 Chi-Square   DF   Pf > ChiSq
     Likelihood Ratio       16.8006     2      0.0002
     Score                  12.5702     2      0.0019
     Wald                    8.9360     2      0.0115

         Analysis of Maximum Likelihood Estimates
                                   Standard      Wald
Parameter  choice  DF  Estimate     Error    Chi-Square  Pr > ChiSq
Intercept     F     1    1.6177    1.3073       1.5314      0.2159
Intercept     I     1    5.6974    1.7938      10.0881      0.0015
length        F     1   -0.1101    0.5171       0.0453      0.8314
length        I     1   -2.4654    0.8997       7.5101      0.0061

                    Odds Ratio Estimates
                            Point           95% Wald
     Effect   choice   Estimate    Confidence Limits
     length     F         0.896      0.325       2.468
     length     I         0.085      0.015       0.496
```

By equation (6.2), the estimated log odds that the response is "fish" rather than "invertebrate" equals

$$\log(\hat{\pi}_1/\hat{\pi}_2) = (1.618 - 5.697) + [-0.110 - (-2.465)]x = -4.08 + 2.355x$$

Larger alligators seem relatively more likely to select fish rather than invertebrates.

The estimates for a particular equation are interpreted as in binary logistic regression, conditional on the event that the 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 primary food type is "fish" rather than "invertebrate" equal $\exp(2.355) = 10.5$ times the estimated odds at length x meters.

The hypothesis that primary food choice is independent of alligator length is H_0: $\beta_1 = \beta_2 = 0$ for model (6.1). The likelihood-ratio test takes twice the difference in log likelihoods between this model and the simpler one without length as a predictor. As Table 6.2 shows, the test statistic equals 16.8, with $df = 2$. The P-value of 0.0002 provides strong evidence of a length effect.

6.1.3 Estimating Response Probabilities

The multicategory logit model has an alternative expression in terms of the response probabilities. This is

$$\pi_j = \frac{e^{\alpha_j + \beta_j x}}{\sum_h e^{\alpha_h + \beta_h x}}, \quad j = 1, \ldots, J \tag{6.3}$$

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 equation (6.3) for whichever category is the baseline in the logit expressions.

The estimates in Table 6.3 contrast "fish" and "invertebra line category. The estimated probabilities (6.3) of the outc Other) equal

provides a direct equation for the modeled probabilities P(Y(x) = j)

$$\hat{\pi}_1 = \frac{e^{1.62 - 0.11x}}{1 + e^{1.62 - 0.11x} + e^{5.70 - 2.47x}}$$

$$\hat{\pi}_2 = \frac{e^{5.70 - 2.47x}}{1 + e^{1.62 - 0.11x} + e^{5.70 - 2.47x}}$$

$$\hat{\pi}_3 = \frac{1}{1 + e^{1.62 - 0.11x} + e^{5.70 - 2.47x}}$$

The "1" term in each denominator and in the numerator of $\hat{\pi}_3$ represents $e^{\hat{\alpha}_3 + \hat{\beta}_3 x}$ for $\hat{\alpha}_3 = \hat{\beta}_3 = 0$ with the baseline category.

Table 6.3. Parameter Estimates and Standard Errors (in parentheses) for Baseline-category Logit Model Fitted to Table 6.1

	Food Choice Categories for Logit	
Parameter	(Fish/Other)	(Invertebrate/Other)
Intercept	1.618	5.697
Length	−0.110 (0.517)	−2.465 (0.900)

For example, for an alligator of the maximum observed length of $x = 3.89$ meters, the estimated probability that primary food choice is "other" equals

$$\hat{\pi}_3 = 1/\{1 + e^{1.62-0.11(3.89)} + e^{5.70-2.47(3.89)}\} = 0.23.$$

Likewise, you can check that $\hat{\pi}_1 = 0.76$ and $\hat{\pi}_2 = 0.005$. Very large alligators apparently prefer to eat fish. Figure 6.1 shows the three estimated response probabilities as a function of alligator length.

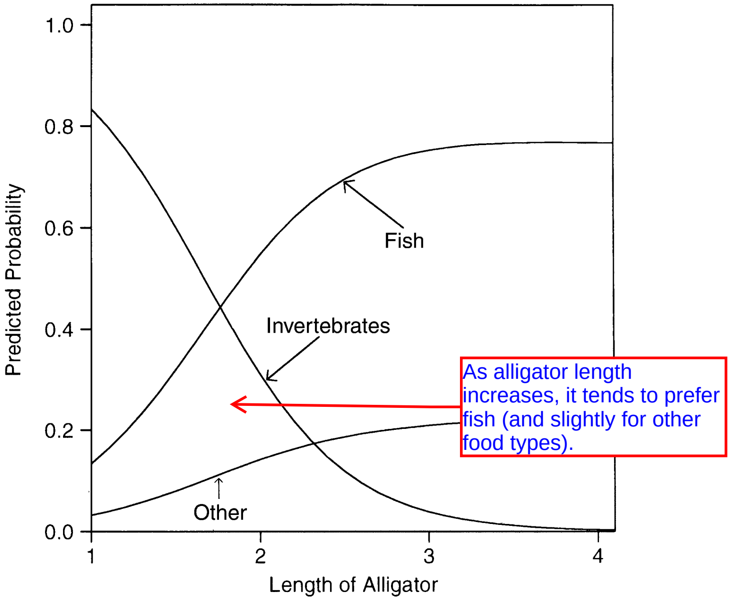


Figure 6.1. Estimated probabilities for primary food choice.

6.1.4 Example: Belief in Afterlife

When explanatory variables are entirely categorical, a contingency table can summarize the data. If the data are not sparse, one can test model goodness of fit using the X^2 or G^2 statistics of Section 5.2.2.

To illustrate, Table 6.4, from a General Social Survey, has Y = belief in life after death, with categories (Yes, Undecided, No), and explanatory variables x_1 = gender and x_2 = race. Let $x_1 = 1$ for females and 0 for males, and $x_2 = 1$ for whites and 0 for blacks. With "no" as the baseline category for Y, the model is

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

where G and R superscripts identify the gender and race parameters.

Table 6.4. Belief in Afterlife by Gender and Race

		Belief in Afterlife		
Race	Gender	Yes	Undecided	No
White	Female	371	49	74
	Male	250	45	71
Black	Female	64	9	15
	Male	25	5	13

Source: General Social Survey.

For these data, the goodness-of-fit statistics are $X^2 = 0.9$ and $G^2 = 0.8$ (the "deviance"). The sample has two logits at each of four gender–race combinations, for a total of eight logits. The model, considered for $j = 1$ and 2, contains six parameters. Thus, the residual $df = 8 - 6 = 2$. The model fits well.

The model assumes a lack of interaction between gender and race in their effects on belief in life after death. Table 6.5 shows the parameter estimates. The effect parameters represent log odds ratios with the baseline category. For instance, β_1^{G} is

Table 6.5. Parameter Estimates and Standard Errors (in parentheses) for Baseline-category Logit Model Fitted to Table 6.4

	Belief Categories for logit	
Parameter	(Yes/No)	(Undecided/No)
Intercept	0.883 (0.243)	−0.758 (0.361)
Gender ($F = 1$)	0.419 (0.171)	0.105 (0.246)
Race ($W = 1$)	0.342 (0.237)	0.271 (0.354)

Table 6.6. Estimated 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

the conditional log odds ratio between gender and response categories 1 and 3 (yes and no), given race. Since $\hat{\beta}_1^{\mathrm{G}} = 0.419$, for females the estimated odds of response "yes" rather than "no" on life after death are $\exp(0.419) = 1.5$ times those for males, controlling for race. For whites, the estimated odds of response "yes" rather than "no" on life after death are $\exp(0.342) = 1.4$ times those for blacks, controlling for gender.

The test of the gender effect has H_0: $\beta_1^{\mathrm{G}} = \beta_2^{\mathrm{G}} = 0$. 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 deviances of $8.0 - 0.8 = 7.2$ has $df = 4 - 2 = 2$. The P-value of 0.03 shows evidence of a gender effect. By contrast, the effect of race is not significant: The model deleting race has $G^2 = 2.8$ ($df = 4$), which is an increase in G^2 of 2.0 on $df = 2$. This partly reflects the larger standard errors that the effects of race have, due to a much greater imbalance between sample sizes in the race categories than in the gender categories.

Table 6.6 displays estimated probabilities for the three response categories. To illustrate, for white females ($x_1 = x_2 = 1$), the estimated probability of response 1 ("yes") on life after death equals

$$\frac{e^{0.883+0.419(1)+0.342(1)}}{1 + e^{0.883+0.419(1)+0.342(1)} + e^{-0.758+0.105(1)+0.271(1)}} = 0.76$$

6.1.5 Discrete Choice Models

The multicategory logit model is an important tool in marketing research for analyzing how subjects choose among a discrete set of options. For example, for subjects who recently bought an automobile, we could model how their choice of brand depends on the subject's annual income, size of family, level of education, and whether he or she lives in a rural or urban environment.

A generalization of model (6.1) allows the explanatory variables to take different values for different Y categories. For example, the choice of brand of auto would likely depend on price, which varies among the brand options. The generalized model is called a *discrete choice model*.[1]

[1] See Agresti (2002, Section 7.6) and Hensher et al. (2005) for details.

6.2 CUMULATIVE LOGIT MODELS FOR ORDINAL RESPONSES

response variable is an ordinal variable

When response categories are ordered, the logits can utiliz
results in models that have simpler interpretations and potentially greater power than baseline-category logit models.

A *cumulative probability* for Y is the probability tha
point. For outcome category j, the cumulative probab

e.g., very left, left, moderate, right, very right (political spectrum)

we want to build a regression model on predicting the probabilities of being within these ordered categories (e.g., probability of a person having moderate or left-wing ideoologies)

$$P(Y \leq j) = \pi_1 + \cdots + \pi_j, \quad j$$

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 cumulative probabilities are

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

These are called *cumulative logits*. For $J = 3$, for example, models use both $\text{logit}[P(Y \leq 1)] = \log[\pi_1/(\pi_2 + \pi_3)]$ and $\text{logit}[P(Y \leq 2)] = \log[(\pi_1 + \pi_2)/\pi_3]$. Each cumulative logit uses all the response categories.

6.2.1 Cumulative Logit Models with Proportional Odds Property

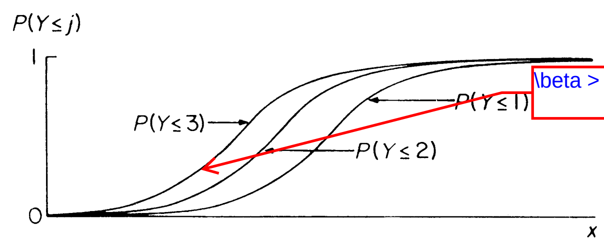
A model for cumulative logit j looks like a binary logistic regression model in which categories 1–j combine to form a single category and categories $j + 1$ to J form a second category. For an explanatory variable x, the model

$$\text{logit}[P(Y \leq j)] = \alpha_j + \beta x, \quad j = 1, \ldots, J - 1 \tag{6.4}$$

- \beta: the increase in log-odds of falling into or below any category given one-unit increase in x
- suppose \beta > 0:
    - as x increases, we will have greater probability of falling into LOWER-valued categories
    - P(Y <= j) will yield increasing curves in x
- suppose \beta < 0:
    - as x increases, we will have greater probability of falling into HIGHER-valued categories
    - P (Y <= j) will yield decreasing curves in x

has parameter β describing the effect of x on t
j or below. In this formula, β does not have
assumes that the effect of x is identical for all
model fits well, it requires a single parameter rat
the effect of x.

Figure 6.2 depicts this model for a four categ
cumulative probability has it own curve, descr
The curve for $P(Y \leq j)$ looks like a logistic re
with pair of outcomes $(Y \leq j)$ and $(Y > j)$. The
that the three curves have the same shape. Any one curve is identical to any of the



$\beta > 0$

Figure 6.2. Depiction of cumulative probabilities in proportional odds model.

others 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 6.2 has $\beta > 0$. Figure 6.3 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 curves in Figure 6.2 descend rather than ascend, and the labels in Figure 6.3 reverse order.

$\beta > 0$

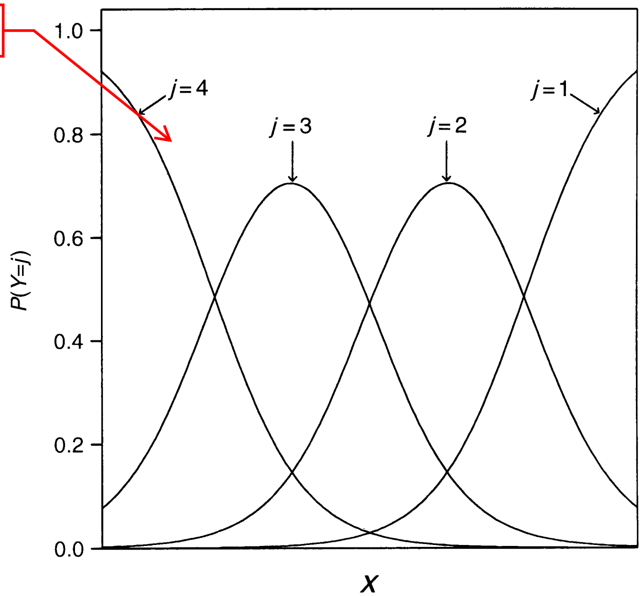


Figure 6.3. Depiction of category probabilities in proportional odds model. At any particular x value, the four probabilities sum to 1.

Then, as x increases, Y is more likely to fall at the high end of the scale.[2] When the model holds with $\beta = 0$, the graph has a horizontal line for each cumulative

for ANY category, the ratio of the cumulative odds at different predictors x1 and x2 remains the same over any category = exp(\beta (x2-x1))

- hence: proportional odds

are statistically independent.

n use odds ratios for the cumulative probabilities and their

es x_1 and x_2 of x, an odds ratio comparing the cumulative

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

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. 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.

For this log odds ratio $\beta(x_2 - x_1)$, the same proportionality constant (β) applies for each cumulative probability. This property is called the *proportional odds* assumption of model (6.4).

more flexible model (without proportional odds):
- separate slope parameters \beta_j for each of j = 1, ... , J-1

Explanatory variables in cumulative lo (with indicator variables), or of both typ algorithm simultaneously for all j. When the categories are reversed in order, the same fit results but the sign of $\hat{\beta}$ reverses.

6.2.2 Example: Political Ideology and Party Affiliation

Table 6.7, from a General Social Survey, relates political ideology to political party affiliation. Political ideology has a five-point ordinal scale, ranging from very liberal to very conservative. Let x be an indicator variable for political party, with $x = 1$ for Democrats and $x = 0$ for Republicans.

Table 6.7. Political Ideology by Gender and Political Party

		Political Ideology				
Gender	Political Party	Very Liberal	Slightly Liberal	Moderate	Slightly Conservative	Very Conservative
Female	Democratic	44	47	118	23	32
	Republican	18	28	86	39	48
Male	Democratic	36	34	53	18	23
	Republican	12	18	62	45	51

Source: General Social Survey.

R function polr ( )
https://data.princeton.edu/wws509/r/c6s5

Table 6.8 shows output (from P) for the ML fit of model (6.4). With $J = 5$ response categories, the model has four $\{\alpha_j\}$ intercepts. Usually

[2]The model is sometimes written instead as $\text{logit}[P(Y \leq j)] = \alpha_j - \beta x$, so that $\beta > 0$ corresponds to Y being more likely to fall at the high end of the scale as x increases.

Table 6.8. Computer Output (SAS) for Cumulative Logit
Ideology Data

\beta > 0:
- a larger value of x results in greater probability of falling on or below a category j
- here: democrats are more likely to have more liberal views than republicans - this makes sense

Analysis of Maximum Likelihood

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept 1	1	−2.4690	0.1318	350.8122	<.0001
Intercept 2	1	−1.4745	0.1091	182.7151	<.0001
Intercept 3	1	0.2371	0.0948	6.2497	.0124
Intercept 4	1	1.0695	0.1046	104.6082	<.0001
party	1	0.9745	0.1291	57.0182	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
party	2.650	2.058	3.412

Testing Global Null Hypothesis: BETA = 0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	58.6451	1	<.0001
Score	57.2448	1	<.0001
Wald	57.0182	1	<.0001

Deviance and Pearson Goodness-of-Fit Statistics

Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	3.6877	3	1.2292	0.2972
Pearson	3.6629	3	1.2210	0.3002

these are not of interest except for estimating response probabilities. The estimated effect of political party is $\hat{\beta} = 0.975$ ($SE = 0.129$). For any fixed j, the estimated odds that a Democrat's response is in the liberal direction rather than the conservative direction (i.e., $Y \leq j$ rather than $Y > j$) equal $\exp(0.975) = 2.65$ times the estimated odds for Republicans. A fairly substantial association exists, with Democrats tending to be more liberal than Republicans.

The model expression for the cumulative probabilities themselves is

$$P(Y \leq j) = \exp(\alpha_j + \beta x)/[1 + \exp(\alpha_j + \beta x)]$$

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

$$\hat{P}(Y \leq 1) = \frac{\exp[-2.469 + 0.975(1)]}{1 + \exp[-2.469 + 0.975(1)]} = 0.18$$

Likewise, substituting $\hat{\alpha}_2$, $\hat{\alpha}_3$, and $\hat{\alpha}_4$ for Democrats yields $\hat{P}(Y \leq 2) = 0.38$, $\hat{P}(Y \leq 3) = 0.77$, and $\hat{P}(Y \leq 4) = 0.89$. Category probabilities are differences of cumulative probabilities. For example, the estimated probability that a Democrat is moderate (category 3) is

$$\hat{\pi}_3 = \hat{P}(Y = 3) = \hat{P}(Y \leq 3) - \hat{P}(Y \leq 2) = 0.39$$

6.2.3 Inference about Model Parameters

For testing independence (H_0: $\beta = 0$), Table 6.8 reports that the likelihood-ratio statistic is 58.6 with $df = 1$. This gives extremely strong evidence of an association ($P < 0.0001$). The test statistic equals the difference between the deviance value for the independence model (which is 62.3, with $df = 4$) and the model allowing a party effect on ideology (which is 3.7, with $df = 3$).

Since it is based on an ordinal model, this test of independence uses the ordering of the response categories. When the model fits well, it is more powerful than the tests of independence presented in Section 2.4 based on $df = (I - 1)(J - 1)$, because it focuses on a restricted alternative and has only a single degree of freedom. The df value is 1 because the hypothesis of independence (H_0: $\beta = 0$) has a single parameter. Capturing an effect with a smaller df value yields a test with greater power (Sections 2.5.3 and 4.4.3). Similar strong evidence results from the Wald test, using $z^2 = (\hat{\beta}/SE)^2 = (0.975/0.129)^2 = 57.1$.

A 95% confidence interval for β is $0.975 \pm 1.96 \times 0.129$, or $(0.72, 1.23)$. The confidence interval for the odds ratio of cumulative probabilities equals $[\exp(0.72), \exp(1.23)]$, or $(2.1, 3.4)$. The odds of being at the liberal end of the political ideology scale is at least twice as high for Democrats as for Republicans. The effect is practically significant as well as statistically significant.

6.2.4 Checking Model Fit

As usual, one way to check a model compares it with models that contain additional effects. For example, the likelihood-ratio test compares the working model to models containing additional predictors or interaction terms.

For a global test of fit, the Pearson X^2 and deviance G^2 statistics compare ML fitted cell counts that satisfy the model to the observed cell counts. When there are at most a few explanatory variables that are all categorical and nearly all the cell counts are at least about 5, these test statistics have approximate chi-squared distributions. For the political ideology data, Table 6.8 shows that $X^2 = 3.7$ and $G^2 = 3.7$, based on $df = 3$. The model fits adequately.

Some software also presents a score test of the proportional odds assumption that the effects are the same for each cumulative probability. This compares model (6.4), which has the same β for each j, to the more complex model having a separate β_j for each j. For these data, this statistic equals 3.9 with $df = 3$, again not showing evidence of lack of fit.

The model with proportional odds form implies that the distribution of Y at one predictor value tends to be higher, or tends to be lower, or tends to be similar, than the distribution of Y at another predictor value. Here, for example, Republicans tend to be higher than Democrats in degree of conservative political ideology. When x refers to two groups, as in Table 6.7, the model does *not* fit well when the response distributions differ in their variability, so such a tendency does not occur. 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 variability than the Democrats'. The two ideology distributions would be quite different, but the model would not detect this.

When the model does not fit well, one could use the more general model with separate effects for the different cumulative probabilities. This model replaces β in equation (6.4) with β_j. It implies that curves for different cumulative probabilities climb or fall at different rates, but then those curves cross at certain predictor values. This is inappropriate, because this violates the order that cumulative probabilities must have [such as $P(Y \leq 2) \leq P(Y \leq 3)$ for all x]. Therefore, such a model can fit adequately only over a narrow range of predictor values. Using the proportional odds form of model ensures that the cumulative probabilities have the proper order for all predictor values.

When the model fit is inadequate, another alternative is to fit baseline-category logit models [recall equation (6.1)] and use the ordinality in an informal way in interpreting the associations. A disadvantage this approach shares with the one just mentioned is the increase in the number of parameters. Even though the model itself may have less bias, estimates of measures of interest such as odds ratios or category probabilities may be poorer because of the lack of model parsimony. We do not recommend this approach unless the lack of fit of the ordinal model is severe in a practical sense.

Some researchers collapse ordinal responses to binary so they can use ordinary logistic regression. However, a loss of efficiency occurs in collapsing ordinal scales, in the sense that larger standard errors result. In practice, when observations are spread fairly evenly among the categories, the efficiency loss is minor when you collapse a large number of categories to about four categories. However, it can be severe when you collapse to a binary response. It is usually inadvisable to do this.

6.2.5 Example: Modeling Mental Health

Table 6.9 comes from a study of mental health for a random sample of adult residents of Alachua County, Florida. Mental impairment is ordinal, with categories (well, mild symptom formation, moderate symptom formation, impaired). The study related $Y =$ mental impairment to two explanatory variables. The life events index x_1 is a composite measure of the number and severity of important life events such as birth of child, new job, divorce, or death in family that occurred to the subject within the past three years. In this sample it has a mean of 4.3 and standard deviation of 2.7. Socioeconomic status ($x_2 = SES$) is measured here as binary ($1 =$ high, $0 =$ low).

Table 6.9. Mental Impairment by SES and Life Events

Subject	Mental Impairment	SES	Life Events	Subject	Mental Impairment	SES	Life Events
1	Well	1	1	21	Mild	1	9
2	Well	1	9	22	Mild	0	3
3	Well	1	4	23	Mild	1	3
4	Well	1	3	24	Mild	1	1
5	Well	0	2	25	Moderate	0	0
6	Well	1	0	26	Moderate	1	4
7	Well	0	1	27	Moderate	0	3
8	Well	1	3	28	Moderate	0	9
9	Well	1	3	29	Moderate	1	6
10	Well	1	7	30	Moderate	0	4
11	Well	0	1	31	Moderate	0	3
12	Well	0	2	32	Impaired	1	8
13	Mild	1	5	33	Impaired	1	2
14	Mild	0	6	34	Impaired	1	7
15	Mild	1	3	35	Impaired	0	5
16	Mild	0	1	36	Impaired	0	4
17	Mild	1	8	37	Impaired	0	4
18	Mild	1	2	38	Impaired	1	8
19	Mild	0	5	39	Impaired	0	8
20	Mild	1	5	40	Impaired	0	9

The main effects model of proportional odds form is

$$\text{logit}[P(Y \leq j)] = \alpha_j + \beta_1 x_1 + \beta_2 x_2$$

Table 6.10 shows SAS output. The estimates $\hat{\beta}_1 = -0.319$ and $\hat{\beta}_2 = 1.111$ suggest that the cumulative probability starting at the "well" end of the scale decreases as life

Table 6.10. Output for Fitting Cumulative Logit Model to Table 6.9

Score Test for the Proportional Odds Assumption

Chi-Square	DF	Pr > ChiSq
2.3255	4	0.6761

Parameter	Estimate	Std Error	Like Ratio 95% Conf Limits		Chi-Square	Pr > ChiSq
Intercept1	−0.2819	0.6423	−1.5615	0.9839	0.19	0.6607
Intercept2	1.2128	0.6607	−0.0507	2.5656	3.37	0.0664
Intercept3	2.2094	0.7210	0.8590	3.7123	9.39	0.0022
life	−0.3189	0.1210	−0.5718	−0.0920	6.95	0.0084
ses	1.1112	0.6109	−0.0641	2.3471	3.31	0.0689

events increases and increases at the higher level of SES. Given the life events score, at the high SES level the estimated odds of mental impairment below any fixed level are $e^{1.111} = 3.0$ times the estimated odds at the low SES level.

For checking fit, the Pearson X^2 and deviance G^2 statistics are valid only for non-sparse contingency tables. They are inappropriate here. Instead, we can check the fit by comparing the model to more complex models. Permitting interaction yields a model with ML fit

$$\text{logit}[\hat{P}(Y \leq j)] = \hat{\alpha}_j - 0.420x_1 + 0.371x_2 + 0.181x_1x_2$$

The coefficient 0.181 of x_1x_2 has $SE = 0.238$. The estimated effect of life events is -0.420 for the low SES group ($x_2 = 0$) and $(-0.420 + 0.181) = -0.239$ for the high SES group ($x_2 = 1$). The impact of life events seems more severe for the low SES group, but the difference in effects is not significant.

An alternative test of fit, presented in Table 6.10, is the score test of the proportional odds assumption. This tests the hypothesis that the effects are the same for each cumulative logit. It compares the model with one parameter for x_1 and one for x_2 to the more complex model with three parameters for each, allowing different effects for logit$[P(Y \leq 1)]$, logit$[P(Y \leq 2)]$, and logit$[P(Y \leq 3)]$. Here, the score statistic equals 2.33. It has $df = 4$, because the more complex model has four additional parameters. The more complex model does not fit significantly better ($P = 0.68$).

6.2.6 Interpretations Comparing Cumulative Probabilities

Section 6.2.1 presented an odds ratio interpretation for the model. An alternative way of summarizing effects uses the cumulative probabilities for Y directly. To describe effects of quantitative variables, we compare cumulative probabilities at their quartiles. To describe effects of categorical variables, we compare cumulative probabilities for different categories. We control for quantitative variables by setting them at their mean. We control for qualitative variables by fixing the category, unless there are several in which case we can set them at the means of their indicator variables. In the binary case, Section 4.5.1 used these interpretations for ordinary logistic regression.

We illustrate with $P(Y \leq 1) = P(Y = 1)$, the *well* outcome, for the mental health data. First, consider the SES effect. At the mean life events of 4.3, $\hat{P}(Y = 1) = 0.37$ at high SES (i.e., $x_2 = 1$) and $\hat{P}(Y = 1) = 0.16$ at low SES ($x_2 = 0$). Next, consider the life events effect. The lower and upper quartiles for life events are 2.0 and 6.5. For high SES, $\hat{P}(Y = 1)$ changes from 0.55 to 0.22 between these quartiles; for low SES, it changes from 0.28 to 0.09. (Comparing 0.55 with 0.28 at the lower quartile and 0.22 with 0.09 at the upper quartile provides further information about the SES effect.) The sample effect is substantial for each predictor.

6.2.7 Latent Variable Motivation*

With the proportional odds form of cumulative logit model, a predictor's effect is the same in the equations for the different cumulative logits. Because each predictor has only a single parameter, it is simpler to summarize and interpret effects than in the baseline-category logit model (6.1).

One motivation for the proportional odds structure relates to a model for an assumed underlying continuous variable. With many ordinal variables, the category labels relate to a subjective assessment. It is often realistic to conceive that the observed response is a crude measurement of an underlying continuous variable. The example in Section 6.2.2 measured political ideology with five categories (very liberal, slightly liberal, moderate, slightly conservative, very conservative). In practice, there are differences in political ideology among people who classify themselves in the same category. With a precise enough way to measure political ideology, it is possible to imagine a continuous measurement. For example, if the underlying political ideology scale has a normal distribution, then a person whose score is 1.96 standard deviations above the mean is more conservative than 97.5% of the population. In statistics, an unobserved variable assumed to underlie what we actually observe is called a *latent variable*.

Let Y^* denote a latent variable. Suppose $-\infty = \alpha_0 < \alpha_1 < \cdots < \alpha_J = \infty$ are *cutpoints* of the continuous scale for Y^* such that the observed response Y satisfies

$$Y = j \quad \text{if } \alpha_{j-1} < Y^* \leq \alpha_j$$

In other words, we observe Y in category j when the latent variable falls in the jth interval of values. Figure 6.4 depicts this. Now, suppose the latent variable Y^* satisfies an ordinary regression model relating its mean to the predictor values. Then,

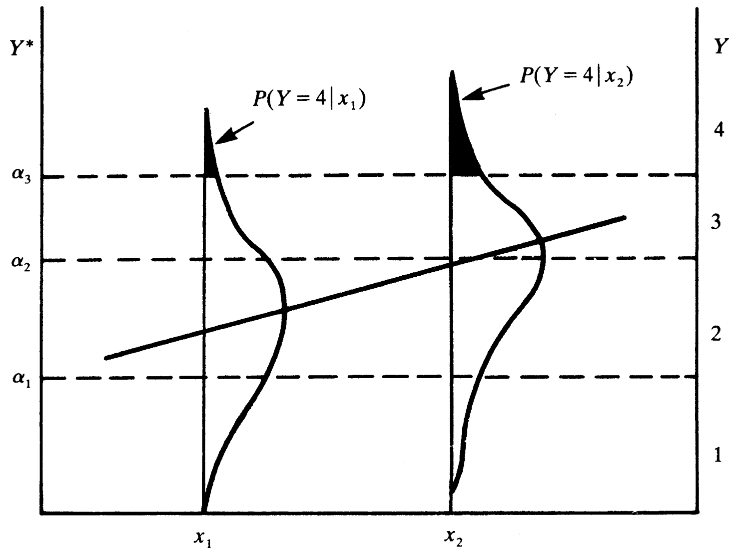


Figure 6.4. Ordinal measurement, and underlying regression model for a latent variable.

one can show[3] that the categorical variable we actually observe satisfies a model with the same linear predictor. Also, the predictor effects are the same for each cumulative probability. Moreover, the shape of the curve for each of the $J - 1$ cumulative probabilities is the same as the shape of the *cdf* of the distribution of Y^*.

At given values of the predictors, suppose Y^* has a normal distribution, with constant variance. Then a probit model holds for the cumulative probabilities. If the distribution of Y^* is the *logistic distribution*, which is bell-shaped and symmetric and nearly identical to the normal, then the cumulative logit model holds with the proportional odds form.

Here is the practical implication of this latent variable connection: If it is plausible to imagine that an ordinary regression model with the chosen predictors describes well the effects for an underlying latent variable, then it is sensible to fit the cumulative logit model with the proportional odds form.

6.2.8 Invariance to Choice of Response Categories

In the connection just mentioned between the model for Y and a model for a latent variable Y^*, the same parameters occur for the effects regardless of how the cutpoints $\{\alpha_j\}$ discretize the real line to form the scale for Y. The effect parameters are *invariant* to the choice of categories for Y.

For example, if a continuous variable measuring political ideology has a linear regression with some predictor variables, then the same effect parameters apply to a discrete version of political ideology with the categories (liberal, moderate, conservative) or (very liberal, slightly liberal, moderate, slightly conservative, very conservative). An implication is this: Two researchers who use different response categories in studying a predictor's effect should reach similar conclusions. If one models political ideology using (very liberal, slightly liberal, moderate, slightly conservative, very conservative) and the other uses (liberal, moderate, conservative), the parameters for the effect of a predictor are roughly the same. Their estimates should be similar, apart from sampling error. This nice feature of the model makes it possible to compare estimates from studies using different response scales.

To illustrate, we collapse Table 6.7 to a three-category response, combining the two liberal categories and combining the two conservative categories. Then, the estimated party affiliation effect changes only from 0.975 ($SE = 0.129$) to 1.006 ($SE = 0.132$). Interpretations are unchanged.

6.3 PAIRED-CATEGORY ORDINAL LOGITS

Cumulative logit models for ordinal responses use the entire response scale in forming each logit. Alternative logits for ordered categories use *pairs* of categories.

[3]For details, see Agresti 2002, pp. 277–279.