Chapter 6 Multicategory Logit Models

Outline

Outline

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

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

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

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 model uses $\log(\pi_1/\pi_3)$ and $\log(\pi_2/\pi_3)$.

6.1.1 Baseline-Category Logits

The baseline-category logit model with a predictor x is

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

The model has $J-1$ equations, with separate parameters for each. 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.

6.1.1 Baseline-Category Logits

These equations for these pairs of categories determine equations for all other pairs of categories. For example, for an arbitrary pair of categories a and b,

$$\begin{aligned}\log\left(\frac{\pi_a}{\pi_b}\right) &= \log\left(\frac{\pi_a/\pi_J}{\pi_b/\pi_J}\right) = \log\left(\frac{\pi_a}{\pi_J}\right) - \log\left(\frac{\pi_b}{\pi_J}\right) \\ &= (\alpha_a + \beta_a x) - (\alpha_a + \beta_b x) \\ &= (\alpha_a - \alpha_b) + (\beta_a - \beta_b)x.\end{aligned} \tag{2}$$

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 $\beta = (\beta_a - \beta_b)$.

The choice of the baseline category is arbitrary.

6.1.2 Example: Alligator Food Choice

Table 6.1 Alligator Size (Meters) and Primary Food Choice,[a] for 59 Florida Alligators

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

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

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

6.1.2 Example: Alligator Food Choice

Example

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.

6.1.2 Example: Alligator Food Choice

Example

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.

6.1.2 Example: Alligator Food Choice

Output

Let Y = primary food choice and x = alligator length. With J = 3, Table 6.2 shows some output, with other as the baseline category. The ML prediction equations are

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

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

The estimated log odds that the response is "fish" rather than "invertebrate" equals

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

6.1.2 Example: Alligator Food Choice

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

Parameter	choice	DF	Estimate	Standard Error	Wald 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

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

6.1.2 Example: Alligator Food Choice

- 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.
- The hypothesis that primary food choice is independent of alligator length is $H_0 : \beta_1 = \beta_2 = 0$ for model (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

$$\pi_j = \frac{e^{\alpha_j+\beta_j x}}{\sum_h e^{\alpha_h+\beta_h x}}, j = 1, \ldots, J. \tag{3}$$

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

6.1.3 Estimating Response Probabilities

The estimates in Table 6.3 contrast "fish" and "invertebrate" to "other" as the baseline category. The estimated probabilities (3) of the outcomes (Fish, Invertebrate, Other) equal

$$
\begin{aligned}
\hat{\pi}_1 &= \frac{e^{1.62-0.11x}}{1+e^{1.62-0.11x}+e^{5.70-2.47x}} \\
\hat{\pi}_2 &= \frac{e^{5.70-2.47x}}{1+e^{1.62-0.11x}+e^{5.70-2.47x}} \\
\hat{\pi}_3 &= \frac{1}{1+e^{1.62-0.11x}+e^{5.70-2.47x}}
\end{aligned}
$$

6.1.3 Estimating Response Probabilities

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)

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.

6.1.3 Estimating Response Probabilities

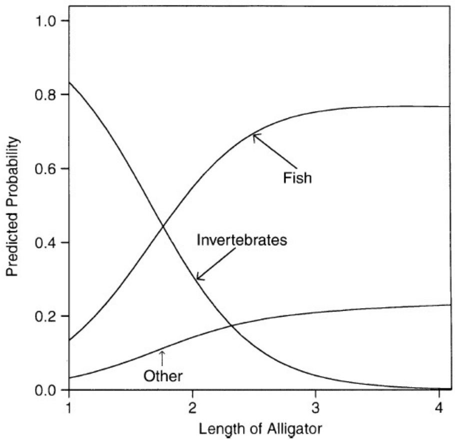

Figure 6.1. Estimated probabilities for primary food choice.

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.

6.1.4 Example: Belief in Afterlife

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

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

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

6.1.4 Example: Belief in Afterlife

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.

6.1.4 Example: Belief in Afterlife

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.

6.1.4 Example: Belief in Afterlife

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)

6.1.4 Example: Belief in Afterlife

β_1^G is the conditional log odds ratio between gender and response categories 1 and 3 (yes and no), given race. Since $\hat{\beta}_1^G = 0.419$, 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.

6.1.4 Example: Belief in Afterlife

The test of the gender effect has $H_0 : \beta_1^G = \beta_2^G = 0$. 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.

6.1.4 Example: Belief in Afterlife

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

$$\frac{e^{0.883+0.419(1)+0.342(1)}}{1 + e^{0.833+0.419(1)+0.342(1)} + e^{-0.758+0.105(1)+0.271(1)}} = 0.76.$$

6.1.4 Example: Belief in Afterlife

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

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 (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.

Outline

Cumulative Probability

A *cumulative probability* for Y is the probability that Y falls at or below a particular point. For outcome category j , the cumulative probability is

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

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.

Cumulative Logits

The logits of the cumulative probabilities are

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

$$j = 1, \ldots, J-1.$$

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 $1, \ldots, j$ combine to form a single category and categories $j+1$ to J form a second category. For an explanatory variable x, the model

$$\text{logit}[P(Y \leq j)] = \alpha_j + \beta x, \ \ j = 1, \ldots, J-1 \tag{4}$$

has parameter β describing the effect of x on the log odds of response in category j or below.

6.2.1 Cumulative Logit Models with Proportional Odds Property

In this formula, β does not have a j subscript. Therefore, the model assumes that the effect of x is identical for all $J-1$ cumulative logits. When this model fits well, it requires a single parameter rather than $J-1$ parameters to describe the effect of x.

6.2.1 Cumulative Logit Models with Proportional Odds Property

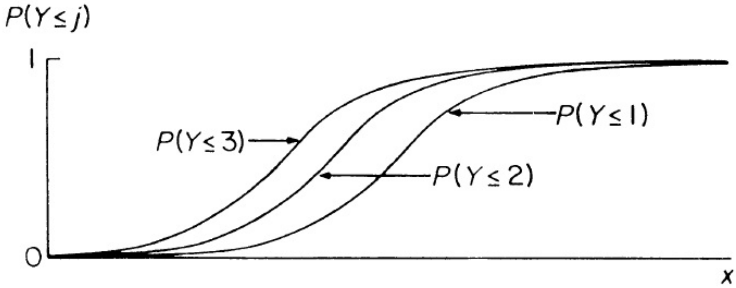

Figure 6.2. Depiction of cumulative probabilities in proportional odds model.

6.2.1 Cumulative Logit Models with Proportional Odds Property

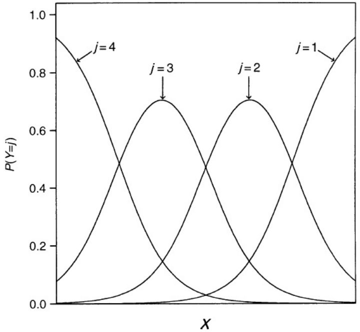

Figure 6.3. Depiction of category probabilities in proportional odds model. At any particular x value, the four probabilities sum to 1.

6.2.1 Cumulative Logit Models with Proportional Odds Property

Model interpretations can use odds ratios for the cumulative probabilities and their complements. For two values x_1 and x_2 of x, an odds ratio comparing the cumulative probabilities is

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

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.

6.2.1 Cumulative Logit Models with Proportional Odds Property

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 (4).

Explanatory variables in cumulative logit models can be quantitative, categorical (with indicator variables), or of both types. 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. 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.

6.2.2 Example: Political Ideology and Party Affiliation

Example

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.

6.2.2 Example: Political Ideology and Party Affiliation

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

Analysis of Maximum Likelihood Estimates

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

6.2.2 Example: Political Ideology and Party Affiliation

Analysis

- 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.
- The first estimated cumulative probability for Democrats $(x = 1)$ is

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

6.2.2 Example: Political Ideology and Party Affiliation

- Likewise, substituting $\hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4$ for Democrats yields $\hat{P}(Y \leq 2) = 0.38$, $\hat{P}(Y \leq 3) = 0.77$, $\hat{P}(Y \leq 4) = 0.89$.

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

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 ($P < 0.0001$). *How to calculate?*
- 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. *What's that?*
- Similar strong evidence results from the Wald test, using $z^2 = (\hat{\beta}/SE)^2 = (0.975/0.129)^2 = 57.1$.
- A 95% CI for β is $0.975 \pm 1.96 \times 0.129$. *What's the CI for the odds ratio of cumulative probabilities?*

6.2.4 Checking Model Fit

- 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. 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.*
- Comparing model(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.

6.2.4 Checking Model Fit

- 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 (4) with β_j. Problem?
- When the model fit is inadequate, another alternative is to fit baseline-category logit models [recall equation (1)] and use the ordinality in an informal way in interpreting the associations. A disadvantage this approach shares with the previous one mentioned is the increase in the number of parameters.
- 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.

6.2.5 Example: Modeling Mental Health

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

6.2.5 Example: Modeling Mental Health

Example

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).

6.2.5 Example: Modeling Mental Health

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. 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

6.2.5 Example: Modeling Mental Health

Analysis

- For checking fit, the Pearson X^2 and deviance G^2 statistics are valid only for nonsparse contingency tables. They are inappropriate here. Permitting interaction yields a model with ML fit

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

- An alternative test of fit, presented inTable 6.10, is the score test of the proportional odds assumption. 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 $\text{logit}[P(Y \leq 1)]$, $\text{logit}[P(Y \leq 2)]$, and $\text{logit}[P(Y \leq 3)]$. 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.

6.2.7 Latent Variable Motivation

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.

Latent Variable

An unobserved variable assumed to underlie what we actually observe is called a *latent variable*.

6.2.7 Latent Variable Motivation

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

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

In other words, we observe Y in category j when the latent variable falls in the j th interval of values.

6.2.7 Latent Variable Motivation

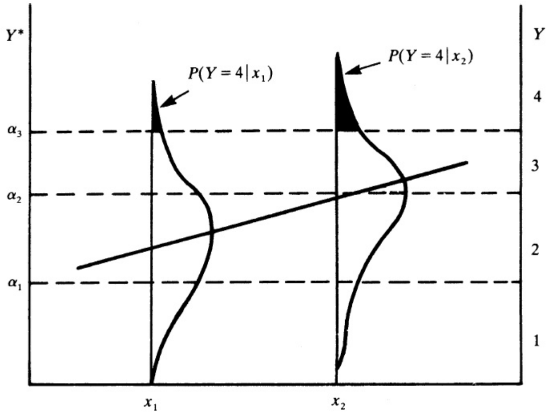

Figure 6.4. Ordinal measurement, and underlying regression model for a latent variable.

6.2.7 Latent Variable Motivation

Suppose the latent variable Y^* satisfies an ordinary regression model relating its mean to the predictor values. Then, one can show 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^*.

6.2.7 Latent Variable Motivation

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. 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

- The effect parameters are invariant to the choice of categories for Y.
- 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.

Outline

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

6.3.1 Adjacent-Categories Logits

Adjacent-Categories Logits

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

$$\log\left(\frac{\pi_{j+1}}{\pi_j}\right) = \alpha_j + \beta_j x, j = 1, \ldots, J-1. \tag{5}$$

$$\log\left(\frac{\pi_{j+1}}{\pi_j}\right) = \alpha_j + \beta x, j = 1, \ldots, J-1. \tag{6}$$

6.3.1 Adjacent-Categories Logits

The adjacent-categories logits, like the baseline-category logits, determine the logits for all pairs of response categories. For the simpler model (6), the coefficient of x for the logit, $\log(\pi_a/\pi_b)$, equals $\beta(a-b)$. The effect depends on the distance between categories, so this model recognizes the ordering of the response scale.

6.3.2 Example: Political Ideology Revisited

Example

Let's return to Table 6.7 and model political ideology using the adjacent-categories logit model (6) of proportional odds form. Let $x = 0$ for Democrats and $x = 1$ for Republicans.

Result

$\hat{\beta} = 0.435$

$G^2 = 5.5$

6.3.3 Continuation-Ratio Logits

Continuation-Ratio Logits

The models apply simultaneously to

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

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.

6.3.3 Continuation-Ratio Logits

Continuation-Ratio Logits

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.

6.3.4 Example: A Developmental Toxicity Study

Table 6.11. Outcomes for Pregnant Mice in Developmental Toxicity Study[a]

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

[a] 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.

6.3.4 Example: A Developmental Toxicity Study

Example

Table 6.11 comes from 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 10 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.

6.3.4 Example: A Developmental Toxicity Study

- We apply 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.
- When models for different continuation-ratio logits have separate parameters, separate fitting of ordinary binary logistic regression models for different logits gives the same results as simultaneous fitting.
- The sum of the separate deviance statistics is an overall goodness-of-fit statistic pertaining to the simultaneous fitting. For Table 6.11, the deviance G^2 values are 5.8 for the first logit and 6.1 for the second.

6.3.5 Overdispersion in Clustered Data

- The above 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 from the same litter.
- The model also treats fetuses from different litters at a given concentration level as having the same response probabilities.

Overdispersion

Either statistical dependence or heterogeneous probabilities violates the binomial assumption. They typically cause *overdispersion*-greater variation than the binomial model implies

Outline

It is often useful to check whether one variable has an effect on another after we control for a third variable. This section shows ways to test the hypothesis of conditional independence in three-way tables when the response variable is multicategory.

6.4.1 Example: Job Satisfaction and Income

Table 6.12. Job Satisfaction and Income, Controlling for Gender

		Job Satisfaction			
Gender	Income	Very Dissatisfied	A Little Satisfied	Moderately Satisfied	Very Satisfied
Female	<5000	1	3	11	2
	5000–15,000	2	3	17	3
	15,000–25,000	0	1	8	5
	>25,000	0	2	4	2
Male	<5000	1	1	2	1
	5000–15,000	0	3	5	1
	15,000–25,000	0	0	7	3
	>25,000	0	1	9	6

Source: General Social Survey, 1991.

6.4.1 Example: Job Satisfaction and Income

Example

Table 6.12, from the 1991 General Social Survey, refers to the relationship between Y = job satisfaction and income, stratified by gender, for black Americans. Let us test the hypothesis of conditional independence using a cumulative logit model. Let x_1 = gender and x_2 = income.

Model

We will treat income as quantitative, by assigning scores to its categories.

$$\text{logit}[P(Y \leq j)] = \alpha_j + \beta_1 x_1 + \beta_2 x_2, j = 1, 2, 3,$$

$$\text{logit}[P(Y \leq j)] = \alpha_j + \beta_1 x_1, j = 1, 2, 3.$$

6.4.1 Example: Job Satisfaction and Income

Analysis

For income, we use scores $\{3, 10, 20, 35\}$, which use midpoints of the middle two categories, in thousands of dollars. The model with an income effect has deviance $13.9(df = 20)$, and the model with no income effect has deviance $19.6(df = 19)$. The difference between the deviances is 5.7, based on $df = 20 - 19 = 1$. This gives $P = 0.017$ and provides evidence of an association.

Analysis

$$\log\left(\frac{\pi_j}{\pi_4}\right) = \alpha_j + \beta_{j1}x_1 + \beta_{j2}x_2 + \beta_{j3}x_3 + \beta_{j4}x_4.$$

6.4.1 Example: Job Satisfaction and Income

Analysis

$x_2 = 1$ if income is in the first category, $x_3 = 1$ if income is in the second category, and $x_4 = 1$ if income is in the third category, in each case 0 otherwise. For this model, conditional independence of job satisfaction and income is

$$H_0 : \beta_{j2} = \beta_{j3} = \beta_{j4} = 0, j = 1, 2, 3$$

equating nine parameters equal to 0.

6.4.1 Example: Job Satisfaction and Income

The difference of deviances equals 12.3, $df = 9$, P-value is 0.20. This test has the advantage of not assuming as much about the model structure. A disadvantage is that it often has low power, because the null hypothesis has so many (nine) parameters.

6.4.2 Generalized Cochran-Mantel-Haenszel Tests

Alternative tests of conditional independence generalize the Cochran-Mantel-Haenszel(CMH) statistic to $I \times J \times K$ tables. Like the CMH statistic and the model-based statistics without interaction terms, these statistics perform well when the conditional association is similar in each partial table. There are three versions, according to whether both, one, or neither of Y and the predictor are treated as ordinal.

6.4.2 Generalized Cochran-Mantel-Haenszel Tests

When both variables are ordinal, the test statistic generalizes the correlation statistic (2.10) for two-way tables. It is designed to detect a linear trend in the association that has the same direction in each partial table. The generalized correlation statistic has approximately a chi-squared distribution with $df = 1$. Its formula is complex and we omit computational details. It is available in standard software (e.g., PROC FREQ in SAS).

6.4.2 Generalized Cochran-Mantel-Haenszel Tests

For Table 6.12 with the scores $\{3, 10, 20, 35\}$ for income and $\{1, 3, 4, 5\}$ for satisfaction, the sample correlation between income and job satisfaction equals 0.16 for females and 0.37 for males. The generalized correlation statistic equals 6.2 with $df = 1(P = 0.01)$. This gives the same conclusion as the ordinal-model-based likelihood-ratio test of the previous subsection.

6.4.3 Detecting Nominal-Ordinal Conditional Association

When the predictor is nominal and Y is ordinal, scores are relevant only for levels of Y. We summarize the responses of subjects within a given row by the mean of their scores on Y, and then average this row-wise mean information across the K strata. The test of conditional independence compares the I rows using a statistic based on the variation among the I averaged row mean responses. This statistic is designed to detect differences among their true mean values. It has a large-sample chi-squared distribution with $df = (I - 1)$.

6.4.3 Detecting Nominal- Ordinal Conditional Association

For Table 6.12, this test treats job satisfaction as ordinal and income as nominal. The test searches for differences among the four income levels in their mean job satisfaction. Using scores $\{1, 2, 3, 4\}$, the mean job satisfaction at the four levels of income equal $(2.82, 2.84, 3.29, 3.00)$ for females and $(2.60, 2.78, 3.30, 3.31)$ for males. For instance, the mean for the 17 females with income < 5000 equals $[1(1) + 2(3) + 3(11) + 4(2)]/17 = 2.82$. The pattern of means is similar for each gender, roughly increasing as income increases. The generalized CMH statistic for testing whether the true row mean scores differ equals 9.2 with $df = 3 (P = 0.03)$. The evidence is not quite as strong as with the fully ordinal analyses above based on $df = 1$.

6.4.4 Detecting Nominal - Nominal Conditional Association

Another CMH-type statistic, based on $df = (I-1)(J-1)$, provides a "general association" test. It is designed to detect any type of association that is similar in each partial table. It treats the variables as nominal, so it does not require category scores.

For Table 6.12, the general association statistic equals 10.2, with $df = 9 (P = 0.34)$. We pay a price for ignoring the ordinality of job satisfaction and income. For ordinal variables, the general association test is usually not as powerful as narrower tests with smaller df values that use the ordinality.

6.4.4 Detecting Nominal - Nominal Conditional Association

Table 6.13 summarizes results of the three generalized CMH tests applied to Table 6.12. The format is similar to that used by SAS with the CMH option in PROC FREQ. Normally, we would prefer a model-based approach to a CHM-type test. A model, besides providing significance tests, provides estimates of sizes of the effects.

6.4.4 Detecting Nominal - Nominal Conditional Association

Table 6.13. Summary of Generalized Cochran–Mantel–Haenszel Tests of Conditional Independence for Table 6.12

Alternative Hypothesis	Statistic	*df*	*P*-value
General association	10.2	9	0.34
Row mean scores differ	9.2	3	0.03
Nonzero correlation	6.6	1	0.01

Outline

Homework 6

1. Analysis the GDS5037 data. Suppose the samples are randomly chosen. Let Y denote patient's status, $Y = M$ for mild asthma (MMA), $Y = S$ for severe asthma (SA), and $Y = C$ for control.
   (1) Build baseline-category logit model use $Y = C$ as baseline with any 4 identifiers without interaction. Explain your model.
   (2) Patients' status can be viewed as ordinal variable. Let $S > M > C$, build cumulative logit models with the 4 identifiers in (1) without interaction. Which model is better, with proportional odds property or not? Why?
2. Problems in textbook 6.3, 6.4, 6.6, 6.7, 6.12