6 Multicategory Logit Models

- Logistic regression is a technique for relating a binary response variable Y to one or more explanatory variables. The explanatory variables may be categorical, continuous, or both. Here we extend the methods of logistic regression to include responses that can be any of several categories. Such models are called *multicategory (or polychotomous) logit models*. We will first study models with nominal categorical responses and then models with ordinal categorical responses.

- At each combination of levels of the explanatory variables, the model assumes that the counts for the categories of Y have a multinomial distribution. The models are also known as *multinomial logit models*.

  A generalization of this model is referred to as the *discrete choice model* in the business and econometrics literature.

Example: A study was undertaken to assess factors associated with women's knowledge, attitude, and behavior toward mammography. The variables in the study are in the table on the next page. The response variable was mammography experience.

Variable	Description	Codes/Values	Name
1	Identification Code	1-412	OBS
2	Mammograph Experience	0 = Never	ME
		1 = Within One Year	
		2 = $>$ One Year Ago	
3	”You do not need a mammogram	1 = Strongly Agree	SYMPT
	unless you develop symptoms”	2 = Agree	
		3 = Disagree	
		4 = Stongly Disagree	
4	Perceived benefit of mammography	5 - 20	PB
5	Mother or Sister with a history	0 = No, 1 = Yes	HIST
	of breast cancer		
6	”Has anyone taught you how to	0 = No, 1 = Yes	BSE
	examine your own breasts:		
	that is BSE”		
7	”How likely is it that a mamogram	1= Not likely	DETC
	could find a new case of	2 = Somewhat likely	
	breast cancer”	3 = Very likely	

*The variable PB is the sum of five scaled responses, each on a four point scale. A low value is indicative of a woman with strong agreement with the benefits of mammography.

Example: Agresti presents an example where 59 alligators were sampled in Florida. The response is primary food type: Fish, Invertebrate, and Other. The explanatory variable is length of the alligator in meters.

6.1 Logit Models for Nominal Responses

- We suppose that the response Y is a nominal variable with J categories. The ordering of the categories is irrelevant.

- Let $\{\pi_1, \ldots, \pi_J\}$ denote the response probabilities. Then $\sum_j \pi_j = 1$.

- If we have n independent observations based on these probabilities, the counts in the categories have a *multinomial* distribution.

6.1.1 Baseline Category Logits

- We need to generalize from ordinary logits to *generalized logits* to handle the J categories.

- Once the model has formed logits for certain $J-1$ pairs of categories, any other logits are redundant.

- For dichotomous data, we model the logit:

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$$

- Let's try to generalize to the three category case. Consider the three following binary logit models:

$$\begin{aligned}
\log\left(\tfrac{\pi_1}{1-\pi_1}\right) &= \alpha_1 + \beta_1 x \\
\log\left(\tfrac{\pi_2}{1-\pi_2}\right) &= \alpha_2 + \beta_2 x \\
\log\left(\tfrac{\pi_3}{1-\pi_3}\right) &= \alpha_3 + \beta_3 x
\end{aligned}$$

This approach is not workable because the probabilities must satisfy

$$\pi_1 + \pi_2 + \pi_3 = 1.$$

- Instead we formulate models for each pair of categories:

$$
\begin{aligned}
\log\left(\frac{\pi_1}{\pi_3}\right) &= \alpha_1 + \beta_1 x \\
\log\left(\frac{\pi_2}{\pi_3}\right) &= \alpha_2 + \beta_2 x \\
\log\left(\frac{\pi_1}{\pi_2}\right) &= \alpha_3 + \beta_3 x
\end{aligned}
$$

These equations are mutually consistent, and one is redundant. We can obtain the third equation from the first two:

$$
\begin{aligned}
\log\left(\frac{\pi_1}{\pi_2}\right) &= \log\left(\frac{\pi_1}{\pi_3}\right) - \log\left(\frac{\pi_2}{\pi_3}\right) \\
&= (\alpha_1 + \beta_1 x) - (\alpha_2 + \beta_2 x) \\
&= (\alpha_1 - \alpha_2) + (\beta_1 - \beta_2) x
\end{aligned}
$$

This implies that $\alpha_3 = \alpha_1 - \alpha_2$ and $\beta_3 = \beta_1 - \beta_2$.

- For J categories, there are $\binom{J}{2} = \frac{J(J-1)}{2}$ pairs of categories. We only need to specify certain $J-1$ pairs, and the other pairs will be redundant.

- For the case $J = 3$, we choose the first two pairs of categories and get expression (6.1) from the text:

$$\log\left(\frac{\pi_j}{\pi_J}\right) = \alpha_j + \beta_j x$$

where $j = 1, 2$ and $J = 3$.

- The probabilities for the three categories corresponding to the logit equations are

$$\pi_1 = \frac{e^{\alpha_1+\beta_1 x}}{1 + e^{\alpha_1+\beta_1 x} + e^{\alpha_2+\beta_2 x}}$$

$$\pi_2 = \frac{e^{\alpha_2+\beta_2 x}}{1 + e^{\alpha_1+\beta_1 x} + e^{\alpha_2+\beta_2 x}}$$

$$\pi_3 = \frac{1}{1 + e^{\alpha_1+\beta_1 x} + e^{\alpha_2+\beta_2 x}}$$

Note that

$$\pi_1 + \pi_2 + \pi_3 = 1.$$

- For the logits defined by formula (6.1), the last (J^{th}) category is called the **baseline (or reference) category**.

- The analysis does not require that the response be ordered in any particular way. The choice of baseline category is arbitrary, but it is usually chosen in a way to facilitate interpretation of the data.

- Statistical software should estimate all $J-1$ logit equations simultaneously. We can use either the **CATMOD** procedure or the **LOGISTIC** procedure in SAS to analyze generalized logits. For the **LOGISTIC** procedure, use `link=GLOGIT` option in the model statement.

- For simultaneous fitting, the same estimates are obtained for comparing any pair of categories no matter which category is the baseline.

- For $J = 3$, the degrees of freedom for testing each effect for modeling two generalized logits are twice what you obtain for modeling one logit. This occurs because you are simultaneously modeling two response functions to estimate parameters for each logit.

- For each effect, the first row is for the first logit, the second row for the second logit, and so on.

- **CATMOD**, like **LOGISTIC**, use “Chi-square” to refer to a Wald chi-squared statistic. Also, “likelihood ratio” refers to the deviance statistic. This can be used to test goodness of fit if there are relatively few covariate patterns.

- Since you are modeling more than one response function per subpopulation, the sample size needs to be large enough to support the number of functions that you are estimating. If there are not enough data, you may encounter problems with parameter estimation and receive warning about infinite parameter estimates. Sometimes reducing the response structure to a meaningful dichotomy can help.

- The multinomial ($J > 2$) logit coefficients must be interpreted as effects on contrasts between pairs of categories, never on the probability of being in a particular category.

- In binary logit analysis, if a covariate x has a positive coefficient, it indicates that an increase in x results in an increase in the probability of the designated outcome. This is not always true in the multinomial model.

- Notice in the figure on slide 10, $\pi_I(x) > \pi_F(x) > \pi_O(x)$ for small xs, then $\pi_F(x) > \pi_I(x) > \pi_O(x)$ for moderate xs, and $\pi_F(x) > \pi_O(x) > \pi_I(x)$ for larger xs.

- For each logit, one interprets the estimates just as in ordinary logistic regression models, conditional on the event that the response outcome was one of the two categories in the logit.

6.1.2 Alligator Food Choice Example

Agresti describes a data set from a study by the Florida Game and Fresh Water Fish Commission of factors influencing the food choice of alligators.

- The response was $Y =$ "primary food choice" taking values in $\{F, I, O\}$.
- The predictor is $x =$ "length of alligator" in meters.
- For an alligator with length x, define

$$\pi_1 = P(Y = F), \qquad \pi_2 = P(Y = I), \qquad \pi_3 = P(Y = O).$$

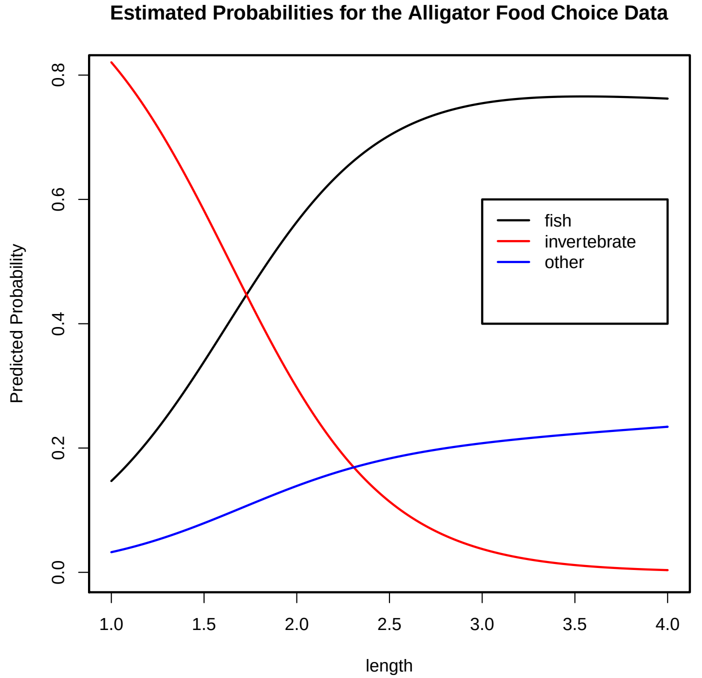
- The baseline logit model was fit to the data and the following estimated logits were obtained.

$$\begin{aligned} \log\left(\frac{\hat{\pi}_1}{\hat{\pi}_3}\right) &= 1.618 - 0.110x \\ \log\left(\frac{\hat{\pi}_2}{\hat{\pi}_3}\right) &= 5.697 - 2.465x \end{aligned}$$

- We can compute the other estimated generalized logit:

$$\begin{aligned} \log\left(\frac{\hat{\pi}_1}{\hat{\pi}_2}\right) &= \log\left(\frac{\hat{\pi}_1}{\hat{\pi}_3}\right) - \log\left(\frac{\hat{\pi}_2}{\hat{\pi}_3}\right) \\ &= (1.618 - 0.110x) - (5.697 - 2.465x) \\ &= -4.080 + 2.355x \end{aligned}$$

- The predicted probabilities for primary food choice as a function of length appear in the following plot:



- An alternative method of fitting multicategory logit models is to fit the ordinary logistic regression model for each pair of categories. This is equivalent to conditional logistic regression, given that the response is in one of the two categories.

  The following table compares the estimates from simultaneous fitting with those obtained by individual fitting:

Generalized Logits (Simultaneous Fitting)

	F vs O	I vs O	F vs I
Intercept	$1.618\ (1.307)$	$5.697\ (1.794)$	$-4.080\ (1.469)$
Length	$-0.110\ (0.517)$	$-2.465\ (0.900)$	$2.355\ (0.803)$

Ordinary Logits (Individual Fitting)

	F vs O	I vs O	F vs I
Intercept	$1.614\ (1.299)$	$5.133\ (1.875)$	$-4.320\ (1.546)$
Length	$-0.109\ (0.514)$	$-2.179\ (0.955)$	$2.478\ (0.839)$

- The parameter estimates are similar, but not identical
- The estimates in both sets are asymptotically unbiased. However, the estimates in the second set are less efficient. They have somewhat larger standard errors.
- The estimates of F vs. I in the second set cannot be found by taking differences of F vs. O and I vs. O as we did for the first set. Also, the estimated probabilities do not sum to 1.
- When fitting logits separately, we cannot test whether all the coefficients of a given covariate (say length) are equal ($p = .003$) or equal zero ($p = .012$).

6.1.3 Belief in Afterlife Example

Agresti examines a data set taken from the 1991 General Social Survey concerning the belief in life after death.

- The response variable is $Y =$ belief in life after death with categories (yes, undecided, no).
- The explanatory variables are $x_1 =$ gender ($= 1$ for females and $= 0$ for males) and $x_2 =$ race ($= 1$ for whites and $= 0$ for blacks).
- Using “no” as the baseline category, the logistic regression model is

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

- For each effect, the first row is for the first logit, etc. We get the following estimates for the two models, default (effects) and reference category (dummy variable):

model	function	MLE		
		intercept	race effect	sex effect
effects	1st function (Yes/No)	1.2632	0.1709	0.2093
ref. category	1st function (Yes/No)	0.8831	0.3418	0.4186
effects	2nd function (Undecided/No)	-0.5700	0.1355	0.0525
ref. category	2nd function (Undecided/No)	-0.7580	0.2710	0.1051

Note: $1.2632 - 0.1709 - 0.2093 = 0.8831$ and $-0.5700 - 0.1355 - 0.0525 = -0.7580$.

6.1.4 Attitude toward Mammography Example

Hosmer and Lemeshow presented a data set that investigated factors that influence a woman's attitude toward mammography. See slide 2 for the variables that were recorded in the study. We will consider various multinomial logit models with the response ME which takes on the values 0,1, and 2. The predictor variables are SYMPT, PB, HIST, BSE, and DETC. SYMPT, HIST, BSE, and DETC are treated as nominal variables. PB is treated as a numerical variable.

After an examination of the fit of the baseline odds model, H&L recommend using dichotomized versions of SYMPT and DETC. This results in a simpler model with adequate fit.

See H&L for a detailed analysis of these data including assessment of the fit of the model.

6.1.5 Using Word Counts to Determine Authorship

Authors have distinctive literary styles that serve as a way to identify them from their writing. One approach to identification uses frequencies of usage of words with little contextual meaning. We will analyze a set of data for four authors, Jane Austen, Jack London, John Milton, and William Shakespeare, where word counts were made from blocks of text by these authors, each containing 1700 words. The original data had 69 words. The authors found 11 words that loaded highly on the first 3 principal components. Baseline logistic models were fit with Shakespeare as the baseline category using frequencies of the 11 words as predictors.

6.2 Ordinal Response Models

We will consider three approaches to modeling multicategory logistic models with ordinal responses.

1. Cumulative logit models
2. Adjacent-categories logit models
3. Continuation-ratio logit models

- When the response has only two categories, these three ordinal models and the baseline-category logit model all reduce to the usual binary logit model.
- In some cases it is necessary to collapse categories.
  - This usually results in some loss of information.
  - This sometimes obscures what you are trying to study.
- It would not be incorrect to ignore the ordering and model generalized logits. However, the models that incorporate ordering
  - have simpler interpretations.
  - often have greater power.

- Models incorporating ordering impose restrictions on data that may be inappropriate. When you use an ordinal model, it is important to check whether its restrictions are valid.

- Of the three models for ordinal responses:
  - The cumulative logit model is the most widely applicable and the one that is the easiest to use in SAS.
  - The adjacent-categories logit model is an attractive general approach, but SAS can only estimate the model when the data are grouped. The covariates must all be categorical or discretized.
  - The continuation-ratio logit model is more specialized. It is designed for situations where the ordered categories represent a progression through stages.

6.2.1 Cumulative Logit Models

- The *cumulative probabilities* are the probabilities that the response Y falls in category j or below, for $j = 1, \ldots, J$. The j^{th} cumulative probability is

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

- The cumulative probabilities reflect the ordering

$$P(Y \le 1) \le P(Y \le 2) \le \cdots \le P(Y \le J) = 1.$$

- The logits of the first $J - 1$ cumulative probabilities are

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

  – Each cumulative logit uses all J categories.

  – A model for the j^{th} cumulative logit looks like an ordinary logit model for a binary response where categories $\{1, \ldots, j\}$ combine to form a single category, and categories $\{j + 1, \ldots, J\}$ form a second category. Thus, Y is collapsed into two categories for each $j < J$.

- For example, for $J = 3$ and data $\{(x, Y)\}$, we could separately fit logistic models:

$$\begin{aligned}\log\left(\frac{\pi_1}{\pi_2+\pi_3}\right) &= \alpha_1 + \beta_1 x \\ \log\left(\frac{\pi_1+\pi_2}{\pi_3}\right) &= \alpha_2 + \beta_2 x\end{aligned}$$

The probabilities based on this model are

$$\pi_1 = \frac{e^{\alpha_1+\beta_1 x}}{1+e^{\alpha_1+\beta_1 x}},\ \pi_2 = \frac{e^{\alpha_2+\beta_2 x} - e^{\alpha_1+\beta_1 x}}{(1+e^{\alpha_1+\beta_1 x})(1+e^{\alpha_2+\beta_2 x})},\ \pi_3 = \frac{1}{1+e^{\alpha_2+\beta_2 x}}$$

and we have $\pi_1 + \pi_2 + \pi_3 = 1$.

Note: These probabilities are not guaranteed to be positive.

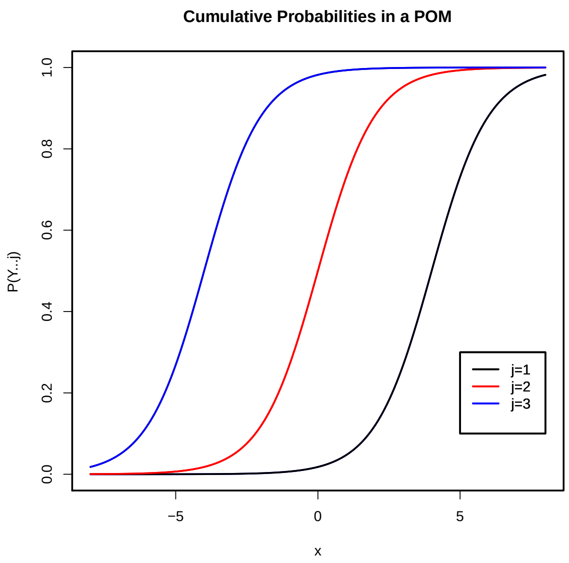
- The cumulative logit model assumes that the effect of x is identical for all $J-1$ cumulative logits. In the above, $\beta_1 = \beta_2$. In general this leads to the expression

$$\log\left(\frac{\pi_1 + \cdots + \pi_j}{\pi_{j+1} + \cdots + \pi_J}\right) = \alpha_j + \beta x, \qquad j = 1, \ldots, J-1,$$

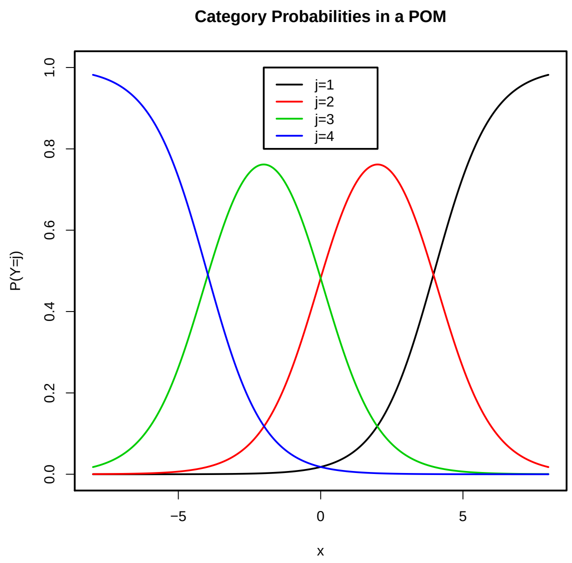
where the ordering of the cumulative probabilities implies that $\alpha_1 < \alpha_2 < \cdots < \alpha_{J-1}$. For all $j < (J-1)$, the model implies that the odds of being in the j^{th} category or lower are multiplied by a factor of e^{β} for each unit increase in x. (The log odds change by a factor of β.) We call this model a **proportional odds model (POM)**.

- We show graphs of the cumulative probabilities $P(Y \le j)$ and the category probabilities $P(Y = j)$ when $(\alpha_1, \alpha_2, \alpha_3, \beta) = (-4, 0, 4, 1)$.
  - The curve for $P(Y \le j)$ looks like a logistic regresssion curve
  - The size of $|\beta|$ determines how quickly the probabilities climb.
  - Because of the common value of β, the probability curves have the same shape.
  - At any fixed x value, the curves have the same ordering as the cumulative probabilities: $P(Y \le 1) \le P(Y \le 2) \le \cdots \le P(Y \le J) = 1$.

Cumulative Probabilities in a POM



Category Probabilities in a POM



- A nice feature of the proportional odds model (POM) is its *invariance* to choice of response categories. This means that if a POM holds for a given response scale, say $\{1, 2, 3, 4, 5\}$, it holds with the same effects for any collapsing of the response categories, say $\{< 3, = 3, > 3\}$.
- Whenever proc LOGISTIC encounters more than 2 categories for the response variable Y, it automatically fits a cumulative logit model and provides MLEs for the parameters in the POM. By default, Y is in an ascending scale, but can be changed using the DESCEND or ORDER options.

6.2.2 Latent Variable Motivation for Proportional Odds Model

Suppose that the ordered categories are based upon an underlying continuous variable, Y^*, which is related to the predictor x through a linear model where

$$E(Y^*) = \alpha + \beta x.$$

We use the cutpoints $-\infty = \alpha_0 < \alpha_1 < \cdots < \alpha_J = \infty$ to determine the ordinal response using

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

One can than show that if the distribution of Y^* is logistic, then the categorical variable that we observe follows a model with the same linear predictor and that the proportional odds model with the coefficient β of x results. The same parameter β appears in the model no matter what cutpoints $\{\alpha_j\}$ we use. If the continuous variable measuring the response follows a linear regression with some predictor variables, the same coefficients apply to a discrete version with the categories defined by cutpoints.

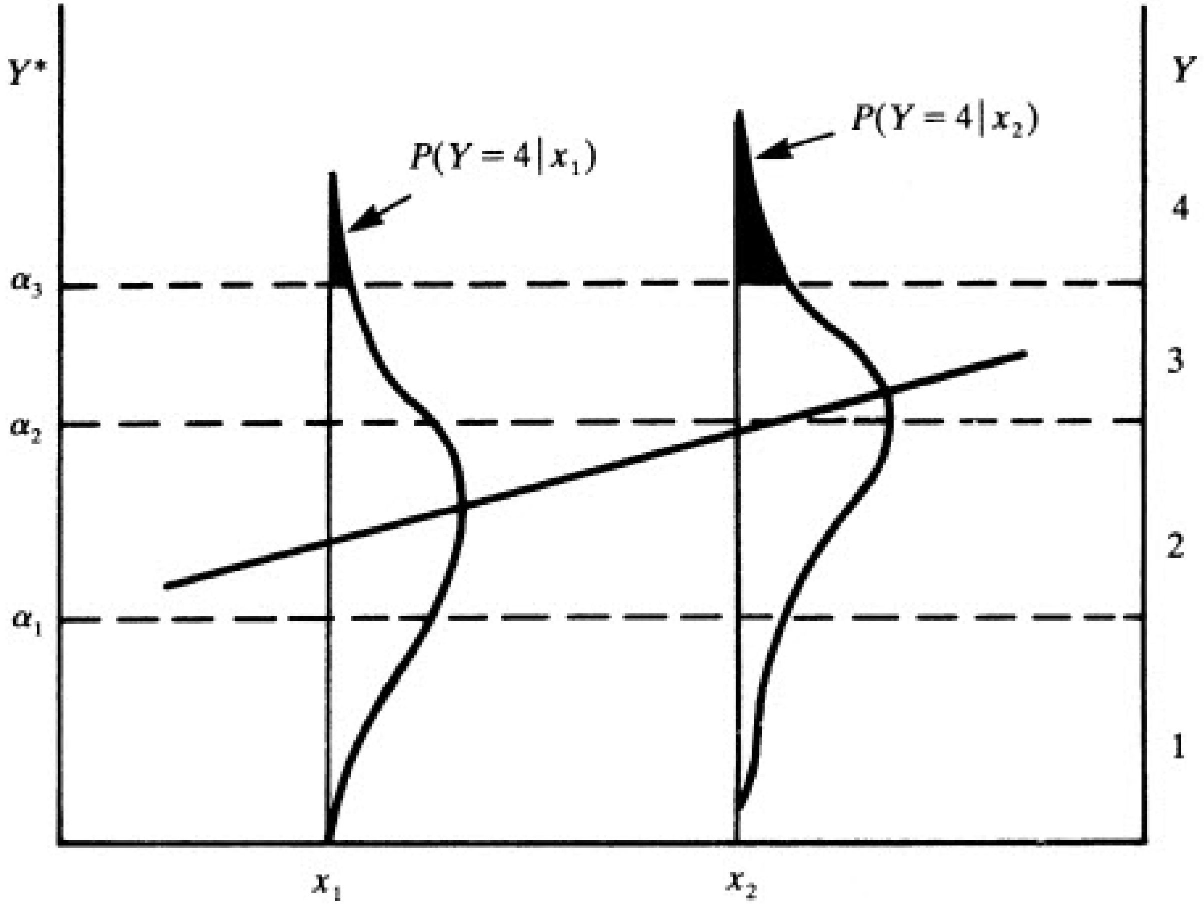


Figure 6.4. Ordinal measurement, and underlying regression model for a latent variable.

Example: **Political Idealogy Example**

- The response $Y =$ political ideology with $\{1 =$ very liberal, $2 =$ liberal, $3 =$ moderate, $4 =$ conservative, $5 =$ very conservative$\}$.
- $x =$ party affiliation with $\{0 =$ Republican, $1 =$ Democrat$\}$.
- Note that the response scale 1–5 is arbitrary. The parameter estimates are the same if you use any scale with the same ordering.
- The signs of the estimates will be reversed if you reverse the order of the categories. It is important to interpret the ORs accordingly
- We can see the effect of collapsing categories from 5 to 3 for the POM.
  - The mle $\hat{\beta}$ for the party effect changes from 0.9755 (SE$=0.1291$) for $J = 5$ to 1.0059 (SE$=0.1322$) for $J = 3$.
  - There is some loss in efficiency when collapsing ordinal scales resulting in larger SEs.
- We can compare separate fitting of logits versus the POM model for $J = 3$.

log-odds of	fitting	intercept	party	party effect	Odds ratio	95% Wald
L vs M+C	separate	-1.47 (0.13)	0.97 (0.16)	based one	Estimate	Conf. Limits
	POM	-1.50 (0.11)	1.01 (0.13)	L vs M+C	2.63	(1.91,3.61)
L+M vs C	separate	0.20 (0.10)	1.04 (0.15)	L+M vs C	2.83	(2.09,3.81)
	POM	0.21 (0.10)	1.01 (0.13)	POM	2.73	(2.11,3.54)

- The POM (6.4) constrains the covariate effect for the two binary logit models to be the same. The intercepts are allowed to differ.
- For this data set, the MLEs for the intercept are similar for both models.
- For the party effect, the MLE for the POM is a kind of “weighted average” of the MLEs from separate fitting.
- When the POM holds, the estimates are more efficient.
- Since the separate fittings agree with the POM, there is no reason to doubt the adequacy of the POM.

• In SAS we can use either CATMOD or LOGISTIC to fit the POM. However, they use different methods of estimating the parameters.
  - CATMOD uses weighted least squares (WLS) while LOGISTIC uses ML.
  - In large samples with categorical predictors, the two fits will be nearly identical. Here $\hat{\beta} = 0.9745\ (0.1291)$ for ML and $\hat{\beta} = 0.9709\ (0.1291)$ for WLS.
  - The orderings for the response functions and the parameterizations are different for the two procedures.

Estimate	LOGISTIC	CATMOD
$\hat{\beta}$	0.9745	-0.9709
$\hat{\alpha}_1$	-2.4690	$0.6591 + 1.8044 = 2.4635$

- Tests for the $x =$ party effect

  - In Chapter 2, we tested $H_0 : x$ and Y are independent using the Pearson χ^2 test, the LR χ^2 test, the Mantel-Haenszel linear trend test, or the Cochran-Armitage trend test.

  - The independence hypothesis is equivalent to $H_0 : \beta = 0$ in the POM.

  - The baseline category model is the saturated model for this example. The independence hypothesis is equivalent to $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ in this model.

  - All these tests for the party effect have the same H_0, but different alternatives. The tests with $df = 4$ are asymptotically equivalent to each other.

  - **Interpretation:** Based on the POM for any fixed $j < 5$, the estimated odds that a Democrat's response is in the liberal direction rather than the conservative direction is 2.65 (95% ci: 2.06 to 3.41) times the estimated odds for a Republican. A strong association exists with Democrats tending to be more liberal than Republicans.

- We obtain the following table for $-2\log\ell$:

Model	$-2\log\ell$
Independence	2533.6
POM	2475.0
Saturated (6.1)	2471.3

  Differences in these yield the LR tests on the output.

- **Tests for the Proportional Odds Assumption:**
  - For the ideology data, the three asymptotically equivalent test statistic values were Score$= 3.91$, Wald$= 3.84$, and LR$= 3.69$ with $df = 3$.

    The high $P-$values support the goodness of fit of the POM.
  - The hypothesis tested is that there is a common parameter vector $(\beta_1, \ldots, \beta_m)$ for the m predictors across the $J-1$ logit models instead of distinct $(\beta_{j1}, \ldots, \beta_{jm}),\ j = 1, \ldots, J-1$. Thus, we are testing

$$H_0 : \beta_{j,1} = \beta_1, \ldots, \beta_{j,m} = \beta_m,\ j = 1, \ldots, J-1$$

    Since the test is comparing m parameters across $J-1$ models, it has $m \times (J-2)$ degrees of freedom. Here, $m = 1,\ J = 5$, and $df = 3$ in the ideology example.

- The sample size requirements are fairly demanding. We need at least 5 observations at each response category at each level of the main effect, or roughly the same sample size as if you were fitting a generalized logit model.

- SAS/STAT User's Guide warns that the test may tend to reject the POM assumption more often than is warranted. Small samples or a large number of response categories or explanatory variables may make the test statistic large, resulting in rejecting the POM.

- If you are concerned about low $P-$values for this test, you might find it useful to fit separate $J-1$ binary-logit models and examine the estimates for similarities or differences.

- If there appears to be no proportionality, the best approach may be to treat the data as nominal and fit the baseline-category logit model.

- The POM implies trends upward or downward in the distributions of the response levels at different values of the explanatory variables. The model does not fit well when the response distributions differ in their dispersion rather than their average.

6.2.3 Adjacent-Categories Logit Models

- The adjacent category logits are

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

- A model with a predictor x has the form

$$\log\left(\frac{\pi_{j+1}}{\pi_j}\right) = \alpha_j^* + \beta_j^* x, \quad j = 1, \ldots, J-1$$

  - For $j = 3$ the logits are $\frac{\pi_2}{\pi_1}$ and $\frac{\pi_3}{\pi_2}$.
  - The baseline logits model has the logits $\frac{\pi_1}{\pi_3}$ and $\frac{\pi_2}{\pi_3}$.
  - These logits, like the baseline logits, determine the logits for all pairs of response categories:

$$\begin{aligned}\log\left(\tfrac{\pi_{j+1}}{\pi_j}\right) &= \log\left(\tfrac{\pi_{j+1}}{\pi_J}\right) - \log\left(\tfrac{\pi_j}{\pi_J}\right) \\ &= (\alpha_{j+1} + \beta_{j+1}x) - (\alpha_j + \beta_j x) = (\alpha_{j+1} - \alpha_j) + (\beta_{j+1} - \beta_j)x,\end{aligned}$$

  Thus, $\alpha_j^* = \alpha_{j+1} - \alpha_j$ and $\beta_j^* = \beta_{j+1} - \beta_j$.

– A simpler model that reflects the ordering of the response categories has identical effects for adjacent categories:

$$\log\left(\frac{\pi_{j+1}}{\pi_j}\right) = \tilde{\alpha}_j + \tilde{\beta}x, \quad j = 1, \ldots, J-1$$

– This model implies that the effect of x depends on the distance between the categories. The coefficient of x for comparing adjacent categories is $\tilde{\beta}$ whereas the coefficient for comparing categories 1 and 3 is $2\tilde{\beta}$.

6.2.4 Continuation-Ratio Logit Models

- Continuation-ratio logits can be modeled by fitting the $J-1$ logits:

$$\log\left(\frac{\pi_{i,1}}{\pi_{i,2}+\cdots+\pi_{i,J}}\right), \log\left(\frac{\pi_{i,2}}{\pi_{i,3}+\cdots+\pi_{i,J}}\right), \cdots, \log\left(\frac{\pi_{i,J-1}}{\pi_{i,J}}\right)$$

- This model compares the probability of a category to that of all the higher categories.

- Continuation-ratio logits can also be modeled by fitting the $J-1$ logits:

$$\log\left(\frac{\pi_{i,1}}{\pi_{i,2}}\right), \log\left(\frac{\pi_{i,1}+\pi_{i,2}}{\pi_{i,3}}\right), \cdots, \log\left(\frac{\pi_{i,1}+\cdots+\pi_{i,J-1}}{\pi_{i,J}}\right)$$

- This model contrasts each category probability with the sum of the probabilities of all the lower categories.

- Similar to the other models for ordinal data, we can have different intercepts and different coefficients of x_i for the $J-1$ logits. Alternatively, we can fit a model where the covariate effects (β_j) are all the same for the $J-1$ logit models.

- The cumulative logit and adjacent categories logit models are reasonable candidates for almost any ordered categorical response variable.

- The continuation-logit model is most useful for situations where a sequential mechanism, such as survival through various age periods or a progression of stages, defines the ordered response categories.

- Suppose $Y \in \{1, 2, \ldots, J\}$. The j^{th} logit for the cumulative logit model is the log-odds of $P(Y \leq j)$.

- For the continuation-ratio logit model, the j^{th} logit models the log-odds of the conditional probability $P(Y \leq j | Y > j-1)$. That is,

$$
\begin{aligned}
\text{logit}\left[P(Y \leq j | Y > j-1)\right] &= \text{logit}\left(\frac{P(Y = j)}{P(Y > j-1)}\right) \\
&= \log\left(\frac{P(Y = j)}{P(Y > j)}\right) \\
&= \log\left(\frac{\pi_j}{\pi_{j+1} + \cdots + \pi_J}\right)
\end{aligned}
$$

