CHAPTER 31

Logistic Regression Models for Categorical Outcome Variables

CHESTER L. BRITT AND DAVID WEISBURD

INTRODUCTION

The study of crime and criminal justice often results in the researcher being confronted with an outcome variable measured by two or more categories. In many applications, the variables are unordered, such as, arrested vs. not arrested, convicted vs. not convicted, or type of case disposition (e.g., acquittal, guilty plea conviction, or trial conviction). Other outcome variables of interest may be measured with ordered categories, such as type of punishment (e.g., probation sentence, jail sentence, or prison sentence), self-reported delinquent acts (e.g., none, one, or two or more), and many attitudinal items, which are commonly measured on ordinal scales that use categories ranging from Strongly Agree to Strongly Disagree. Much of the published criminological research confronted with categorical outcome variables has taken one of three approaches: (1) To collapse some number of categories on a polytomous (multicategory) outcome variable to create a binary measure, and use binary logistic regression models to analyze the data, (2) to treat an ordinal measure as if it were unordered and use multinomial logistic regression models, or (3) to treat an ordinal measure as if it were continuous, and use OLS regression to analyze the data. The first two approaches effectively throw away information about the distribution of cases on the dependent variable, while the third approach assumes that too much information is contained in the ordinal variable, consequently making all three approaches less than satisfactory.

The primary focus of this chapter is a discussion on the type of logistic regression model best suited to an analysis of categorical outcome variables. Toward that end, we discuss binary logistic regression models for situations where the dependent variable has only two categories, multinomial logistic regression for situations where the dependent variable has three or more unordered categories, and ordinal logistic regression models for situations where the dependent variable has at least three ordered categories. Interestingly, and despite the relative ease of use of ordinal logistic regression models, their use in the analysis of crime and criminal justice data is relatively rare. In an attempt to address concerns about analyzing ordinal data appropriately, we place greater emphasis on the use and interpretation of ordinal logistic regression models in this chapter.

A.R. Piquero and D. Weisburd (eds.), *Handbook of Quantitative Criminology*,
DOI 10.1007/978-0-387-77650-7_31,

The discussion is organized as follows: "Binary Logistic Regression" will present the key elements for the binary logistic regression model, since the application and interpretation of this model is foundational to the application of both multinomial and ordinal logistic regression models discussed later in this chapter. Readers who are familiar with the binary logistic regression model may skip "Binary Logistic Regression" and move on directly to "Multinomial Logistic Regression", which will present the application and interpretation of the multinomial logistic regression model. Finally, "Ordinal Logistic Regression" will then discuss two different logistic regression models for ordinal outcome variables: The proportional odds model and the partial proportional odds model. The general approach in the discussion of each model will be to highlight some of the key technical details of the model, and then to emphasize the application and the interpretation of the model. Readers who are interested in more in-depth technical discussions will be referred to the appropriate sources.

BINARY LOGISTIC REGRESSION

Preliminaries: Odds and Odds Ratios

The use of binary logistic regression models in criminology is common because of the dichotomous nature of many key outcome variables in criminology: Arrest vs. no arrest, delinquency vs. no delinquency, or incarceration sentence vs. no incarceration sentence. The following discussion of binary logistic regression models is brief and intended as a starting point (or review) for the discussion of multinomial and ordinal logistic regression models. There is an extensive literature on the use of binary logistic regression models, and readers who are less familiar with this model are encouraged to consult these sources for comprehensive and accessible treatments of binary logistic regression (Hosmer and Lemeshow 2000; Menard 2002; Pampel 2000).

We begin by assuming that we have a binary outcome variable, arbitrarily assigned the values of "1" for the category of primary theoretical or research interest and "0" for the reference category (i.e., the comparison outcome). We can define the odds of one outcome relative to the other outcome as follows:

$$\text{Odds} = \frac{P(Y=1)}{P(Y=0)} = \frac{P(Y=1)}{1-P(Y=1)}$$

where $P(Y=1)$ refers to the probability that the outcome variable (Y) takes on the value "1" (the category of primary interest), and $P(Y=0)$ refers to the probability that the outcome variable takes on the value "0" (the reference category). We can compute the odds from sample data using the proportion of cases that fall into each category. For example, if 60% of cases fall into category 1 and 40% into category 0, then the odds would be $0.60/0.40 = 1.5$, meaning that category "1" is 1.5 times more likely to be observed than category "0".

Since we are typically interested in how the odds of a particular outcome differ for different groups of cases, such as males and females or treatment and control groups, we can use the odds for each group to construct an odds ratio:

$$\text{Odds ratio} = \frac{\left(\frac{P(Y=1)}{1-P(Y=1)}\right)_{\text{group 1}}}{\left(\frac{P(Y=1)}{1-P(Y=1)}\right)_{\text{group 2}}}$$

In practice, the odds ratio provides a measure of strength of association between the outcome variable and the independent variable. For example, in the assessment of the effectiveness of a treatment program, a researcher may compute the odds of success for individuals in the control and the treatment groups. If the treatment was effective, then there would presumably be different odds of success for each group of individuals. The magnitude of this difference would then appear in the odds ratio.

Suppose, for instance, the researcher found that 70% of the individuals in the treatment group and 50% of the individuals in the control group were classified as a "success."

The odds of success for the treatment group would be:

$$\text{Odds} = \frac{P(Y = 1)}{P(Y = 0)} = \frac{0.70}{1 - 0.70} = 2.33$$

The odds of success for the control group would be:

$$\text{Odds} = \frac{P(Y = 1)}{P(Y = 0)} = \frac{0.55}{1 - 0.55} = 1.22$$

The odds ratio comparing the odds for the two groups would be:

$$\text{Odds ratio} = \frac{\left(\frac{0.70}{1 - 0.70}\right)_{\text{treatment}}}{\left(\frac{0.55}{1 - 0.55}\right)_{\text{control}}} = \frac{2.33}{1.22} = 1.91$$

The odds ratio indicates that the odds of a success are 1.91 times larger for the treatment group than the odds of a success for the control group.

Binary Logistic Regression Model

Information about the odds is used as a fundamental component of the binary logistic regression model. The "logit" is defined as the natural logarithm of the odds:

$$\text{Logit} = ln\left(\frac{P(Y = 1)}{1 - P(Y = 1)}\right)$$

The logit is then used as the dependent variable in a linear model that includes k independent variables:

$$\text{Logit} = ln\left(\frac{P(Y = 1)}{1 - P(Y = 1)}\right) = b_0 + b_1 X_1 + \cdots + b_k X_k$$

Using the logit equation, we can rewrite the equation for $P(Y = 1)$, which turns out to be the logistic distribution function:

$$P(Y = 1) = \frac{1}{1 + \exp(-Xb)}$$

The logistic distribution function is then used to estimate the model coefficients through a maximum likelihood procedure in a wide range of statistical software packages. It is important to note that other probability functions, such as the Normal or the Gompertz, may also be used to calculate $P(Y = 1)$. The focus in this chapter is on the use of the logistic regression models because of the relative ease of interpreting the coefficients from these models over those that rely on an alternative probability distribution. Readers interested in the application of the Normal Distribution, known as the probit model, are encouraged to read Pampel (2000) for a comprehensive, and accessible discussion.

Interpreting the Coefficients

The coefficients estimated from a binary logistic regression model (the b_k) can be interpreted directly in the context of a one unit change in the independent variable: For a one unit change in the independent variable, the log of the odds is expected to change by b_k, controlling for all other independent variables included in the model. Since the "log of the odds" does not have intuitive meaning for most people, we can transform the estimated coefficients by exponentiating each coefficient (i.e., $\exp(\mathrm{b}) = \mathrm{e}^b$), which then estimates the odds ratio for a single one unit change in the independent variable.

For example, suppose $b = 0.2$, we could interpret this coefficient directly to mean that for each one unit increase in the independent variable, the log of the odds increases by 0.2 units, controlling for all other variables in the statistical model. Alternatively, if we take the antilog of b, we see that the odds ratio has a value of 1.22 ($\mathrm{e}^{0.2} = 1.22$). Substantively, this tells us that for a one unit change in the independent variable, the odds of outcome $Y = 1$ increase by a factor of 1.22, controlling for all other variables in the model.

It is important to keep in mind that the odds ratio is not a linear function of the independent variables in a binary logistic regression model, which means that we cannot discuss linear changes in the odds ratio in response to changes in the independent variable. If we are interested in determining how much the odds change in response to more than a one unit change in the independent variable, we multiply that value (ΔX) by the coefficient (b) and exponentiate this quantity (i.e., $\mathrm{e}^{\Delta X b}$). If we again use $b = 0.2$, but are interested in the change in the odds for a 5-unit change in the independent variable, then we calculate $\mathrm{e}^{5*0.2} = \mathrm{e}^1 = 2.72$, which means that the odds increase by a factor of 2.72 for a 5-unit increase in the independent variable.

Model Assessment

There is no single best measure for statistically assessing the quality of a binary logistic regression model. Perhaps the most common means for assessing whether a model has a statistically significant effect on the dependent variable is to use a likelihood-ratio test that compares two models. The first model constrains all coefficients for the independent variables and estimates only the model intercept (the null model), while the second model estimates coefficients for all the independent variables (the full model). Since maximum likelihood estimation techniques are used to estimate the coefficients, one of the pieces of information produced in the process of estimation is the value of the likelihood function (essentially a probability estimate). Taking the natural logarithm of this value, and then multiplying by −2 produces a value that is useful for hypothesis testing of one or more coefficients in a binary logistic regression model. The value −2* log-likelihood is typically written as −2LL.

The hypothesis test for whether all coefficients in a model are equal to zero is given by the difference in the −2LL values for the null model and the full model.

$$\text{LR test statistic} = (-2\text{LL}_{\text{null model}}) - (-2\text{LL}_{\text{full model}})$$

The LR test statistic is distributed as a chi-square statistic with degrees of freedom equal to the number of coefficients estimated in the model. This same test can also be used to test the statistical significance of individual coefficients or subsets of coefficients from the full regression model. For a single coefficient, the degrees of freedom would be $df = 1$, while for the subset of coefficients, the degrees of freedom would be equal to the number of coefficients in the subset. Conceptually, the likelihood-ratio test for the full logistic regression model or for a subset of coefficients works in much the same way as an F-test in the typical ordinary least squares regression model.

In addition to likelihood-ratio tests, there are a number of alternative means for overall model assessment of a binary logistic regression model. Two of the more commonly reported measures are the percent of observations correctly classified by the model and the pseudo-R^2 value. Both of these tests are limited in their ability to assess model quality and will tend to be more suggestive than conclusive. However, when used in conjunction with each other and the likelihood-ratio test, a reasonably clear picture of the model's effectiveness should emerge.

The percent of observations correctly classified uses the predicted value of the dependent variable based on the hypothesized logistic regression model, and compares it to the observed distribution of the dependent variable. The predicted value of the dependent variable is usually based on the category that has the highest predicted probability of occurring. In other words, if $P(Y = 1) > 0.5$, then the predicted value of the dependent variable is estimated as category "1"; $P(Y = 1) < 0.5$, means the predicted value of the dependent variable will be estimated as "0". The percent of observations correctly classified as a "0" or a "1" is then a simple percentage:

$$\text{Percent correctly classified} = \frac{\text{Number of observations correctly placed}}{\text{Total number of observations}} \times 100\%$$

One of the primary limitations of the percent of observations correctly classified is that it is dependent on the original distribution of observations. Consequently, if the original distribution has a relatively small percentage of cases in one of the two categories (say, less than 10%), the percent correctly classified will tend to be quite large, because most (or all) of the observations will be predicted to fall in the category with the most cases. Consider the somewhat common situation in the study of crime and delinquency, where 90% of the observations fall into one category, likely no delinquency or no arrest. It is also then quite common for all sample observations to be predicted to fall into the modal category of the dependent variable, which would then result in the percent correctly classified having a value of 90%, even though the model did nothing to discriminate cases and make predictions about which cases would fall into the other category.

Pseudo-R^2 measures are based on comparing values of the −2LL values for the null and full models. Intuitively, this makes sense, because the −2LL value provides an approximation for how close the predicted data represent the observed data. Perhaps the most commonly reported pseudo-R^2 measure is Cox and Snell's R^2:

$$R^2 = 1 - e^{-[(-2\text{LL}_{\text{null model}})-(-2\text{LL}_{\text{full model}})]/N},$$

where N is the size of the sample. All pseudo-R^2 values are then interpreted as rough approximations of the proportion of variation in the dependent variable that is explained by the full model. It is important to note, however, that all pseudo-R^2 measures assume a hypothetical dependent variable that is continuously distributed and are then interpreted as an indicator of explained variance. Again, these kinds of model assessment measures should be viewed as suggestive, rather than conclusive.

Statistical Significance of Individual Coefficients

As noted above, the test of statistical significance for a single coefficient in a regression model could be conducted with a likelihood-ratio chi-square test and one degree of freedom. An alternative is to use a z-statistic or a Wald statistic (W). The z-statistic is simply the ratio of the logistic regression coefficient (b) to its standard error ($SE(b)$). The Wald statistic is the squared value of the z-statistic:

$$z = \left(\frac{b}{SE(b)}\right)$$

$$W = \left(\frac{b}{SE(b)}\right)^2 = z^2.$$

Substantively, the z-statistic, the Wald statistic, and the likelihood-ratio test will tend to give the same substantive answer in regard to the statistical significance of an individual coefficient.[1] For large samples, the results will be nearly identical. To the extent that the results differ for the likelihood-ratio compared to the z-statistic or the Wald statistic, it will be due to the values of z and W being more sensitive to small sample sizes, implying that the likelihood-ratio test will tend to give a more accurate assessment of statistical significance, regardless of the size of the sample (Long 1997).

An Example: Fayetteville Youth Project

The following analysis of data from the Fayetteville Youth Project (FYP) is intended to illustrate some of the key points in the preceding discussion, as well as the material to follow. Briefly, the FYP data were collected with a self-report survey administered in 1997 to students in the 9th, 10th, and 11th grades at Fayetteville High School (Arkansas). The Fayetteville sample was drawn from the 1,782 students enrolled in the 9th through 11th grades in the spring of 1997. The final sample consists of 1,130 student respondents, of which 489 are white males, 69 are black males, 83 are black females, and 489 are white females, generally reflecting the demographic composition of the school.

The dependent variable in the following analysis is self-reported theft of an item valued at \$2 to \$50 (Yes = 1, No = 0). The independent variables include age (in years), sex (Male = 1, Female = 0), race (White = 1, Nonwhite = 0), grade point average (rounded to

[1] Software packages differ in the default results that are reported. For example, SAS and SPSS report the Wald statistic, while Stata and LIMDEP report the z-statistic. Practically, it makes no difference which statistic (W or z) is reported, since the observed significance level of the coefficient will be the same.

nearest integer, range is 0 to 4), whether any of the respondent's friends had been picked up by the police (Yes = 1, No = 0), whether the respondent's parent(s) knew where the respondent was while away from home (Agree and Strongly Agree = 1, Undecided, Disagree, and Strongly Disagree = 0), and three items that assessed the respondent's beliefs about theft and breaking the law (all three coded as Agree and Strongly Agree = 1, Undecided, Disagree, and Strongly Disagree = 0). Descriptive statistics for these variables are presented in Table 31.1.

The logistic regression model can be represented in equation form as

$$\begin{aligned}\text{Logit (Theft)} &= b_0 + b_1\text{age} + b_2\text{sex} + b_3\text{race} + b_4\text{GPA} + b_5\text{friends picked up}\\ &\quad + b_6\text{parents know} + b_7\text{get ahead} + b_8\text{around the law}\\ &\quad + b_9\text{OK to take}\end{aligned}$$

The results of the logistic regression analysis appear in Table 31.2, which reports the coefficient value, the standard error, the z-score, and the odds ratio for each coefficient. We can see from the results that all of the independent variables are statistically significant, except for the age and the sex of the respondent. More directly, and to highlight a few results, we observe the following:

- Whites have odds of self-reported theft that are $\exp(-0.52) = 0.59$ times smaller than the odds of self-reported theft for nonwhites, controlling for all other variables in the model.

TABLE 31.1. Descriptive statistics for Fayetteville youth study ($N = 1{,}056$)

Variable	Mean	Standard deviation
Age	15.74	0.99
Male	0.49	0.50
White	0.87	0.33
GPA	3.04	0.92
Friends picked up by the police	0.59	0.49
Parents know where youth is while away	0.61	0.49
Have to do some things that are not right to get ahead	0.27	0.45
OK to get around the law if you could get away with it	0.27	0.45
OK to take things from big business	0.14	0.35

TABLE 31.2. Binary logistic regression results for Fayetteville youth study

Variable	Coefficient	SE	z-Score
Age	−0.04	0.08	−0.52
Male	0.29	0.16	1.89
White	−0.52	0.22	−2.43
GPA	−0.36	0.09	−4.21
Friends picked up by the police	1.17	0.17	6.81
Parents know where youth is while away	−0.66	0.16	−4.24
Have to do some things that are not right to get ahead	0.41	0.18	2.33
OK to get around the law if you could get away with it	0.53	0.18	2.94
OK to take things from big business	0.88	0.22	3.98
Intercept	0.52	1.32	0.40

- As GPA increases, the odds of self-reported theft decrease. For example, a one-unit increase in GPA reduces the odds of self-reported theft by a factor of $\exp(-0.36) = 0.69$, controlling for all other variables in the model.
- Respondents who have had friends picked up by the police have odds of self-reported theft that are about $\exp(1.17) = 3.21$ times greater than the odds of self-reported theft for those youth who have not had a friend picked up by the police, controlling for all other variables in the model.
- Youth who agreed that their parent(s) generally knew where they were when away from home had odds of self-reported theft that were $\exp(-0.66) = 0.52$ times smaller than the odds of self-reported theft for those youth who were undecided or disagreed that their parents generally knew where they were, controlling for all other variables in the model.
- Youth who agreed that one had to do some things that were not right to get ahead had odds of self-reported theft that were $\exp(0.41) = 1.51$ times higher than the odds of self-reported theft for the youth who were undecided or disagreed, controlling for all other variables in the model.
- Youth who agreed that it was okay to get around the law if one could get away with it had odds of self-reported theft that were $\exp(0.53) = 1.70$ times higher than the odds of self-reported theft for the youth who were undecided or disagreed, controlling for all other variables in the model.
- Youth who agreed that it was okay to take things from big business had odds of self-reported theft that were $\exp(0.88) = 2.41$ times higher than the odds of self-reported theft for the youth who were undecided or disagreed, controlling for all other variables in the model.

The model assessment statistics show the likelihood-ratio test ($\chi^2 = 277.34$, $df = 9$, $p < 0.001$), the pseudo-R^2 (0.20), and percent correctly classified (79.4%) values collectively suggest a model with a modest ability to predict self-reported theft.

MULTINOMIAL LOGISTIC REGRESSION

Multinomial Logistic Regression Model

Multinomial logistic regression is used to examine problems where there are three or more unordered categories in the dependent variable. There are many situations in the study of crime and criminal justice in which dependent variables include multiple unordered categories. For example, the type of case disposition may be measured as acquittal, guilty plea conviction, or trial conviction; type of crime committed by offenders may be measured as violent, property, or drug offenses; and, pretrial release decisions in a court may be measured as released on recognizance, released with supervision, released on bail, and denied release.

Multinomial logistic regression is conceptually a straightforward extension of the binary logistic regression model.[2] Recall that in the binary logistic regression model, we designated one of the two outcome categories as the presence of a given trait and the second as the

[2] Readers interested in more detailed and technical treatments of multinomial logistic regression should consult Agresti (2002), Long (1997), and Menard (2002).

absence of that trait. For example, in the analysis of the FYP data, we compared those who had self-reported theft ($Y = 1$) to those who did not report any theft ($Y = 0$) by estimating the logit of Y:

$$\text{Logit}(Y) = \ln\left(\frac{P(Y = 1)}{P(Y = 0)}\right) = b_0 + b_1 X_1 + \cdots + b_k X_k$$

The logit of Y in the equation requires that there be only the absence ($Y = 0$) or the presence ($Y = 1$) of a characteristic. What happens when the outcome variable has more than two categories? The problem here is that we do not have a simple change in the odds for one outcome compared to one other outcome, as we did with the self-reported theft example. In an outcome variable with three or more unordered categories, we have to take into account the changes in the odds of multiple outcomes, which then leads to multiple comparisons.

Suppose, for example, that our outcome variable has three categories (C1, C2, and C3) with the number of observations in each being represented by N_{C1}, N_{C2}, and N_{C3}. We could begin by estimating three binary logistic regression models that would allow for all possible comparisons of the outcome categories – the logits for C1 and C2, C2 and C3, and C1 and C3. The logit of Y for each regression could be written simply as

$$\ln\left(\frac{P(Y = C1)}{P(Y = C2)}\right), \quad \ln\left(\frac{P(Y = C2)}{P(Y = C3)}\right), \quad \text{and} \quad \ln\left(\frac{P(Y = C1)}{P(Y = C3)}\right)$$

for each comparison, respectively.[3]

If we were to estimate these three separate logits, the coefficients from each equation would be interpreted in the same way as we described in the discussion of the binary logistic regression model. While this approach would allow us to make comparisons of the likelihood of subjects falling in each of the three categories examined as compared to each other, it would require us to run three separate logistic regressions. Moreover, and more importantly from a statistical point of view, we would likely be working with three completely different samples in each of the three analyses: (1) $N_{C1}+N_{C2}$, (2) $N_{C2}+N_{C3}$, and (3) $N_{C1}+N_{C3}$. This is because the cases on the dependent variable are unlikely to be distributed evenly. For example, if we were studying case dispositions among criminal defendants – acquittal/dismissal, guilty plea conviction, or trial conviction – we would not expect the type of disposition for the sample of defendants to be distributed equally across each possible outcome. Consequently, each of our comparisons would be based on different samples. In comparing defendants who had only outcomes C1 and C2, our analyzable sample would be $N_{C1} + N_{C2}$. The sample size would not reflect the defendants that had outcome C3 (i.e., N_{C3}), since they would have been excluded from the comparison. But what we are really interested in is the choice among the three outcomes and how this choice is distributed in our entire sample. The statistical problem here is that the varying sample sizes would then result in incorrect standard errors for the coefficients, leading to inaccurate tests of statistical significance – the absolute values of the coefficients are not affected by the order of the comparisons. The multinomial logistic

[3] These three logits can be linked in an identity equation that illustrates how knowledge of any two logits can produce the values of the third. The identity equation can be stated as

$$\ln\left(\frac{P(Y = C1)}{P(Y = C2)}\right) + \ln\left(\frac{P(Y = C2)}{P(Y = C3)}\right) = \ln\left(\frac{P(Y = C1)}{P(Y = C3)}\right).$$

regression model simultaneously accounts for these different sample sizes, ensuring a more valid estimate of significance levels. It also has the benefit of allowing us to conduct our analysis using a single regression equation.

An important step in a multinomial regression is the definition of the "reference category." This is necessary because we need to decide which category we want to use as a baseline for all comparisons. It is an arbitrary decision about which category is designated the reference category, but to the extent that we can make a choice that has some theoretical relevance or makes the interpretation of the results simpler, that would be the preferred choice. For the case disposition example above, suppose that we choose dismissal as the reference category, which then allows us to make two comparisons between a type of conviction – guilty plea or trial – and dismissal. More directly, our multinomial logistic regression results will indicate (1) the relative likelihood of a guilty plea conviction compared to a dismissal and (2) the relative likelihood of a trial conviction compared to a dismissal. The one comparison not mentioned was the relative likelihood of a guilty plea conviction compared to a trial conviction. In the multinomial logistic regression model, this comparison is not directly estimated, but as will be illustrated shortly, the results can be obtained very simply from the results for the comparison of each conviction type to a dismissal.

The multinomial logistic regression model can be written as either a probability model or an odds ratio model. As a probability equation, the multinomial logistic equation is

$$P(Y = m) = \frac{\exp(Xb_m)}{\sum_{j=1}^{J} \exp\left(Xb_j\right)}.$$

In this equation, m refers to the outcome category of interest and has values ranging from 1 to J (the last category). The numerator to the equation tells us to exponentiate the value of *Xb* for category m. The denominator, in turn, tells us that we need to exponentiate the value of *Xb* for all categories, and then sum these values together. Since there is a redundancy built into the values of the coefficients in a multinomial logistic model, the coefficient values for the reference category ($b1$) are set at 0. The constraining of one set of coefficients to 0 results in the identification of the system of logistic regression equations, which allows for the estimation of unique coefficient estimates for the $J - 1$ logits.[4]

For the three-category case disposition variable, $m = 1, 2$, or 3. Writing out the probability equations for each outcome leads to the following formulations of the probability of each of the three outcomes in our example. For $m = 1,\ b_1 = 0$ and

$$P(Y = 1) = \frac{\exp(X0)}{\exp(X0) + \exp(Xb_2) + \exp(Xb_3)} = \frac{1}{1 + \exp(Xb_2) + \exp(Xb_3)}$$

For $m = 2$ and $m = 3$, we have

$$P(Y = 2) = \frac{\exp(Xb_2)}{1 + \exp(Xb_2) + \exp(Xb_3)}$$

$$P(Y = 3) = \frac{\exp(Xb_3)}{1 + \exp(Xb_2) + \exp(Xb_3)}$$

[4] As noted previously, the choice of the reference category is arbitrary. All possible comparisons of the outcome categories can be made based on single set of $J - 1$ logits.

The multinomial logistic regression model can also be written as an odds ratio equation comparing the probabilities for any two outcomes m and n on the dependent variable:

$$\mathrm{OR}_{m|n} = \frac{P(Y=m)}{P(Y=n)} = \frac{\frac{\exp(Xb_m)}{\sum_{j=1}^{J}\exp(Xb_j)}}{\frac{\exp(Xb_n)}{\sum_{j=1}^{J}\exp(Xb_j)}} = \frac{\exp(Xb_m)}{\exp(Xb_n)}.$$

If we are interested in computing the odds ratio for a comparison between any category (m) and the reference category ($m = 1$), where $b_1 = 0$, we obtain

$$\mathrm{OR}_{m|n} = \frac{P(Y=m)}{P(Y=1)} = \frac{\exp(Xb_m)}{\exp(X0)} = \exp(Xb_m)$$

This result also indicates how we should interpret the coefficients from the multinomial logistic regression model. Since the coefficients for the reference category have been fixed at 0, the coefficients for each of the remaining outcome categories will compare the relative likelihood of that category to the reference category.[5]

In practice what these equations tell us is that we will have $J - 1$ sets of coefficients from a multinomial logistic regression model that can be interpreted in the same way as binary logistic coefficients, where we compare each outcome (m) to the reference category ($m = 1$) for the outcome variable. In our example for case disposition, where we have designated dismissal as the reference category, one set of coefficients will give us the log of the odds or the odds ratios comparing the likelihood of a guilty plea conviction relative to a dismissal, while the second set of coefficients will give us the log of the odds or the odds ratios comparing the likelihood of a trial conviction relative to a dismissal.

An Example: Pretrial Release Status in California

The State Court Processing Statistics database includes information on random samples of individuals arrested for felony offenses in the largest court districts in the United States. To illustrate the application of the multinomial logistic regression model, the following example will use a random sample of 6,606 felony arrestees in California in the 1990s. A question of both policy and theoretical relevance is the study of the factors that affect the defendant's pretrial release status: Nonfinancial release, financial release, held on bail, or denied bail. Table 31.3 presents the coding and descriptive statistics for all the variables included in the multivariate model.

Table 31.4 presents the results from our application of the multinomial logistic regression model. Since there are four categories to the dependent variable, there are three sets of coefficients presented in the three columns of results. Each column in Table 31.4 represents a comparison of nonfinancial release (the reference category) with each of the other three financial release categories: financial release (column 1), held on bail (column 2), and denied bail (column 3).

The first column of results shows that defendants who are older, black or Hispanic (rather than white), used a public defender, and were charged with a drug or a property crime (rather than "other" crime) were less likely to have a financial release than a nonfinancial release.

[5] It is worth pointing out that the binary logistic regression model presented above is a special case of the multinomial logistic regression model, where $m = 2$.

TABLE 31.3. Variable coding and descriptive statistics for pretrial release in California ($N = 6{,}606$)

Variable	Mean	Standard deviation
Age (years)	29.94	8.87
Male (1 = male, 0 = female)	0.83	0.37
Black (1 = black, 0 = non-black)	0.25	0.44
Hispanic (1 = Hispanic, 0 = Non-Hispanic)	0.46	0.50
Under criminal justice supervision (1 = yes, 0 = no)	0.48	0.50
Represented by public defender (1 = yes, 0 = no)	0.75	0.43
Represented by private attorney (1 = yes, 0 = no)	0.15	0.35
Charged with violent crime (1 = yes, 0 = no)	0.22	0.42
Charged with property crime (1 = yes, 0 = no)	0.31	0.46
Charged with drug crime (1 = yes, 0 = no)	0.39	0.49
Number of prior felony arrests	4.06	5.88
Number of prior violent convictions	0.14	0.59

TABLE 31.4. Multinomial logistic regression results for pretrial release in California

	Coefficients for logits:		
Variable	Financial vs. non-financial release	Held on bail vs. non-financial release	Denied bail vs. non-financial release
Age	−0.01	−0.02	−0.02
Male	0.08	0.76	0.65
Black	−0.31	0.11	0.10
Hispanic	−0.09	0.79	−0.24
Under criminal justice supervision	0.23	0.95	1.94
Represented by public defender	−0.19	0.36	−0.39
Represented by private attorney	0.97	−0.22	−0.61
Charged with violent crime	0.43	0.93	0.72
Charged with property crime	−0.43	−0.03	−0.97
Charged with drug crime	−0.52	−0.45	−0.95
Number of prior felony arrests	0.03	0.10	0.12
Number of prior violent convictions	0.18	0.50	0.63
Intercept	0.10	−0.66	−2.07

Conversely, defendants who were male, under criminal justice supervision at the time of arrest, had a private defense attorney, were charged with a violent crime, had more prior felony arrests, and more prior convictions for violent crimes were more likely to have a financial release than a nonfinancial release. As in the previous section, we can also interpret each of these coefficients more directly as odds ratios. (Recall from the previous section that the exponentiation of the coefficient provides us with the odds ratio given a one-unit change in the independent variable.)

- If age is increased by 1 year, the odds of a financial release versus a nonfinancial release decrease by a factor of $\exp(-0.010) = 0.990$, controlling for all other variables in the model.
- The odds of a financial release versus a nonfinancial release are $\exp(0.077) = 1.080$ times greater for male than for female defendants, controlling for all other variables in the model.

- The odds of a financial release versus a nonfinancial release are exp(−0.313) = 0.731 times smaller for black than for white defendants, controlling for all other variables in the model.
- The odds of a financial release versus a nonfinancial release are exp(−0.091) = 0.913 times smaller for Hispanic than for white defendants, controlling for all other variables in the model.
- The odds of a financial release versus a nonfinancial release are exp(0.227) = 1.255 times greater for defendants under criminal justice supervision than for defendants not under supervision at the time of arrest, controlling for all other variables in the model.
- The odds of a financial release versus a nonfinancial release are exp(−0.190) = 0.827 times smaller for defendants with a public defender than for defendants with self or other representation, controlling for all other variables in the model.
- The odds of a financial release versus a nonfinancial release are exp(0.969) = 2.636 times greater for defendants with a private defense attorney than for defendants with self or other representation, controlling for all other variables in the model.
- The odds of a financial release versus a nonfinancial release are exp(0.426) = 1.531 times greater for defendants charged with a violent crime than for defendants with an "other" offense, controlling for all other variables in the model.
- The odds of a financial release versus a nonfinancial release are exp(−0.428) = 0.652 times smaller for defendants charged with a property crime than for defendants with an "other" offense, controlling for all other variables in the model.
- The odds of a financial release versus a nonfinancial release are exp(−0.519) = 0.595 times smaller for defendants charged with a drug crime than for defendants with an "other" offense, controlling for all other variables in the model.
- If the number of prior felony arrests is increased by one, the odds of a financial release versus a nonfinancial release increase by a factor of exp(0.028) = 1.028, controlling for all other variables in the model.
- If the number of prior convictions for violent crimes is increased by one, the odds of a financial release versus a nonfinancial release increase by a factor of exp(0.183) = 1.200, controlling for all other variables in the model.

We can similarly interpret the results in columns 2 and 3, which compare held on bail and denied bail to nonfinancial release, respectively. Note that the effects of each of the independent variables are not necessarily constant across the three different comparisons – the effect of sex or race, for example, varies by the specific comparison. These kinds of variable effects of independent variables are to be expected and reinforce the rationale for using a multinomial logistic regression model. If the dependent variable had been collapsed into two categories, say nonfinancial release versus all other pretrial statuses, then many of the interesting effects of both legal and extra-legal characteristics on pretrial release status would have been missed.

The Missing Coefficients

As noted earlier, when we estimate a multinomial logistic regression model, the coefficients for the reference category are fixed at 0, and we estimate coefficients for comparisons of all other categories with the reference category. Clearly, there are a number of other comparisons that may be of substantive interest, but are not directly estimated. For example, in our analysis of pretrial release status, we may be interested in comparing those defendants denied release to those held on bail. Based on the identity relationship of multiple logits described above, for all possible comparisons of the outcome categories, the most direct way of obtaining the missing coefficients is to simply subtract one set of coefficients from another set of coefficients. In Table 31.4, the results in column 2 represent the logit for held on bail and nonfinancial release, while those in column 3 represent the logit for denied bail and nonfinancial release. The following discussion illustrates how to compute this specific comparison, as well as highlight the general process for computing comparisons of outcome categories not directly estimated in a multinomial logistic regression model.

We start by reconsidering the logit equation for the estimated results. Since the logarithm of a fraction can be rewritten as the subtraction of the logarithm of the denominator from the logarithm of the numerator, the logits can be rewritten as

$$\ln\left(\frac{P(Y = \text{held on bail})}{P(Y = \text{nonfinancial release})}\right) = \ln\left(P(Y = \text{held on bail})\right) - \ln\left(P(Y = \text{nonfinancial release})\right)$$

and

$$\ln\left(\frac{P(Y = \text{denied bail})}{P(Y = \text{nonfinancial release})}\right) = \ln\left(P(Y = \text{denied bail})\right) - \ln\left(P(Y = \text{nonfinancial release})\right)$$

By performing simple subtractions of the logits, we can generate additional contrasts between the outcome categories. To obtain the missing coefficients for the comparison of denied bail to held on bail, we subtract the logit for held on bail and nonfinancial release from the logit for denied bail and nonfinancial release:

$$\ln\left(\frac{P(Y = \text{denied bail})}{P(Y = \text{nonfinancial release})}\right) - \ln\left(\frac{P(Y = \text{held on bail})}{P(Y = \text{nonfinancial release})}\right)$$

$$= \left[\ln\left(P(Y = \text{denied bail})\right) - \ln\left(P(Y = \text{nonfinancial release})\right)\right] - \left[\ln\left(P(Y = \text{held on bail})\right) - \ln\left(P(Y = \text{nonfinancial release})\right)\right]$$

$$= \ln\left(P(Y = \text{denied bail})\right) - \ln\left(P(Y = \text{held on bail})\right)$$

$$= \ln\left(\frac{P(Y = \text{denied bail})}{P(Y = \text{held on bail})}\right)$$

What this algebraic manipulation of logits shows us is that we can obtain the coefficients for the omitted contrast simply by subtracting one set of coefficients from another set of coefficients. If we wanted the inverse of this comparison – the logit of held on bail to denied bail – all we would need to do is alternate the order of subtraction.

When applied to the coefficients for the pretrial release status example, we obtain the results presented in Table 31.5. To highlight a few of the findings:

- The odds of being denied bail versus being held on bail are exp(0.990) $=$ 2.691 times greater for defendants under criminal justice supervision than for defendants not under criminal justice supervision at the time of arrest, controlling for all other variables in the model.

- The odds of being denied bail versus being held on bail are exp(−0.204) = 0.815 times smaller for defendants charged with a violent crime than for defendants with an "other" offense, controlling for all other variables in the model.
- If the number of prior convictions for violent crimes is increased by one, then the odds of being denied bail versus being held on bail increase by a factor of exp(0.135) = 1.145, controlling for all other variables in the model.

A second way to obtain the coefficients for the comparison of being denied bail to being held on bail, would be to redefine the statistical model so that held on bail was chosen as the reference category and re-estimate the multinomial model. Upon rerunning the multinomial logistic regression model with held on bail used as the reference category, we obtain the results presented in Table 31.6. In regard to the comparison of being denied bail to being held on bail (see column 3), note that the results are identical to those presented in Table 31.5, based

TABLE 31.5. Results for logit comparing held on bail to denied bail

Variable	Held on bail vs. denied bail
Age	0.00
Male	0.11
Black	0.01
Hispanic	1.03
Under criminal justice supervision	−0.99
Represented by public defender	0.75
Represented by private attorney	0.39
Charged with violent crime	0.21
Charged with property crime	0.94
Charged with drug crime	0.50
Number of prior felony arrests	−0.02
Number of prior violent convictions	−0.13
Intercept	1.41

TABLE 31.6. Multinomial logistic regression results for pretrial release in California (denied bail as reference category)

	Coefficients for logits:		
Variable	Non-financial release vs. denied bail	Financial release vs. denied bail	Held on bail vs. denied bail
Age	0.02	0.01	0.00
Male	−0.65	−0.57	0.11
Black	−0.10	−0.42	0.01
Hispanic	0.24	0.15	1.03
Under criminal justice supervision	−1.94	−1.71	−0.99
Represented by public defender	0.39	0.20	0.75
Represented by private attorney	0.61	1.58	0.39
Charged with violent crime	−0.72	−0.30	0.20
Charged with property crime	0.97	0.54	0.94
Charged with drug crime	0.95	0.43	0.51
Number of prior felony arrests	−0.12	−0.09	−0.02
Number of prior violent convictions	−0.63	−0.45	−0.13
Intercept	2.07	2.17	1.41

on subtracting the coefficients from the original analysis in Table 31.4. Overall, the results in Table 31.6 will lead us to the same substantive conclusions as the results in Table 31.4, with the only difference being an alternative reference category for interpretation. What this indicates is that the selection of reference categories is arbitrary and that correctly interpreted results will lead to the same substantive conclusions.

Statistical Inference

SINGLE COEFFICIENTS. The results from a multinomial logistic regression analysis complicate tests of statistical significance slightly. Since we now have multiple coefficients representing the effects of each independent variable on the dependent variable, there are questions about how to discern whether an independent variable has an effect on the dependent variable. Specifically, there are two issues of statistical inference that are important for interpreting the results from a multinomial logistic regression analysis. For each coefficient we can (1) estimate the statistical significance of each category compared to the reference category, and (2) estimate the overall significance of the independent variable in predicting the multi-category dependent variable.

To test the effect of each individual coefficient in comparison to the reference category, we would again use the Wald or the *z*-statistic described above. Table 31.7 presents the multinomial logistic coefficients from the original model, along with the standard errors (SE) of the coefficients and the *z*-scores.

We can see that criminal justice supervision at the time of offense and the number of prior felony arrests have statistically significant effects on all three pairs of outcomes. The number of prior convictions is statistically significant for the comparisons of financial release and held on bail to nonfinancial release, but not for the comparison of denied bail to nonfinancial release. Type of legal representation and type of offense also have variable effects across the three comparisons. In regard to defendants' demographic characteristics, age is the only variable that has statistically significant effects on all three comparisons. Defendants' sex has statistically significant effects on all pairs, except for that between financial and nonfinancial release, while the effects of race – measured with dummy variables for black and Hispanic – vary across all three comparisons.

MULTIPLE COEFFICIENTS. Note that in Table 31.7 there are three coefficients for each independent variable. As we noted above, the number of coefficients from a multinomial logistic regression model for each independent variable will be one less than the number of categories on the dependent variable (i.e., $J - 1$). How do we assess the overall effect of each independent variable on all of the categories of the dependent variable simultaneously? There are two key ways of doing this – one is the likelihood ratio test and the other test is an extension of the Wald test – discussed above. Regardless of the statistical software package one uses to estimate a multinomial logistic regression model, both of these methods will be available as either default or requested output.

Recall that the likelihood ratio test involves estimating two different models. One model is viewed as the "full" model (i.e., it contains all the coefficients to be estimated in that specific model) and the other model is viewed as the "reduced" model (i.e., one or more coefficients

TABLE 31.7. Multinomial logistic regression results for pretrial release in California (coefficients, standard errors, and z-scores)

	Financial vs. non-financial release			Held on bail vs. non-financial release			Denied bail vs. non-financial release		
Variable	Coefficient	SE	*z*-Score	Coefficient	SE	*z*-Score	Coefficient	SE	*z*-Score
Age	−0.01	0.005	−2.22	−0.02	0.004	−4.65	−0.02	0.007	−2.21
Male	0.08	0.098	0.78	0.76	0.087	8.75	0.65	0.177	3.66
Black	−0.31	0.108	−2.89	0.11	0.092	1.20	0.10	0.145	0.71
Hispanic	−0.09	0.094	−0.96	0.79	0.081	9.78	−0.24	0.150	−1.58
Under criminal justice supervision	0.23	0.090	2.53	0.95	0.072	13.19	1.94	0.141	13.79
Represented by public defender	−0.19	0.129	−1.48	0.36	0.108	3.35	−0.39	0.167	−2.33
Represented by private attorney	0.97	0.145	6.67	−0.22	0.139	−1.59	−0.61	0.234	−2.61
Charged with violent crime	0.43	0.181	2.35	0.93	0.157	5.91	0.72	0.223	3.23
Charged with property crime	−0.43	0.168	−2.55	−0.03	0.142	−0.19	−0.97	0.220	−4.38
Charged with drug crime	−0.52	0.160	−3.23	−0.45	0.137	−3.24	−0.95	0.210	−4.53
Number of prior felony arrests	0.03	0.011	2.43	0.10	0.009	10.75	0.12	0.011	10.68
Number of prior violent convictions	0.18	0.156	1.17	0.50	0.123	4.02	0.63	0.139	4.53
Intercept	0.10	0.261	0.38	−0.66	0.227	−2.91	−2.07	0.394	−5.25

have been excluded from the full model). The likelihood ratio test statistic is the difference in the −2LL function for each model and is distributed as a chi-square statistic with degrees of freedom equal to the number of coefficients constrained to be zero (i.e., eliminated from the full model). As a test for the statistical significance of an independent variable in a multinomial logistic regression model, the likelihood ratio test refers to estimating the full multinomial logistic regression equation with all variables, and then estimating reduced models that eliminate one independent variable ($J - 1$ coefficients) from each analysis. The difference in the −2LL function for each equation will then allow for the test of each independent variable with $df = J - 1$.

For example, in the analysis of the pretrial release status data, the value of −2LL for the full model is 13194.994. When we estimate the same model, but eliminate the variable age from the analysis, the value of the −2LL increases to 13216.834. The difference of the two log-likelihood functions is $13216.834 - 13194.994 = 21.840$. By eliminating the variable age, we have removed three coefficients from the analysis. The corresponding degrees of freedom for the test will be $df = J - 1 = 3$ to reflect the removal of the three coefficients. Table 31.8 presents the results of the LR test for all the independent variables included in the model. Note that all of the independent variables have statistically significant effects on pretrial release status well below conventional levels of significance.

An alternative test of each independent variable is to use the Wald statistic. Up to this point, we have used the Wald statistic to test the statistical significance of a single coefficient,

TABLE 31.8. Likelihood ratio and Wald statistic results for the overall effect of each independent variable

Independent variable	*df*	LR test statistic	Wald
Age	3	21.70	21.84
Male	3	90.80	89.38
Black	3	17.78	17.56
Hispanic	3	177.32	173.90
Under criminal justice supervision	3	327.31	303.86
Represented by public defender	3	40.14	41.16
Represented by private attorney	3	99.48	91.59
Charged with violent crime	3	36.38	37.39
Charged with property crime	3	28.91	30.54
Charged with drug crime	3	23.80	23.39
Number of prior felony arrests	3	198.92	163.80
Number of prior violent convictions	3	32.18	26.78
Intercept	3	36.31	35.29

but it can also be used to test the group of coefficients representing the effect of any given independent variable. Recall that the Wald test statistic for a single coefficient is computed by dividing the coefficient by its standard error and then squaring this value. The Wald statistic for a group of coefficients involves an analogous calculation, but requires using matrix algebra. Many statistical software packages (e.g., SAS and SPSS) will report the results for the Wald test as part of the standard output. In most applications, the value of the Wald statistic will be very similar to the value of the LR test, unless a small sample is being analyzed. To test the overall effect of an independent variable with the Wald statistic, we continue to use a chi-square distribution with degrees of freedom equal to the number of coefficients being tested (i.e., $df = J - 1$).

The values of the Wald test for each of the independent variables included in our analysis of pretrial release status are presented in the last column of Table 31.8. All of the independent variables have statistically significant effects on pretrial release status. As expected, the values of the Wald statistic are not identical to those of the likelihood ratio test, but the substance of these results is identical to that using the LR test.

How should we address mixed results? For example, it is not uncommon for a researcher to find that the overall effect of an independent variable is not statistically significant, but one of the individual coefficients does have a significant effect on a comparison of two outcome categories. Alternatively, the likelihood ratio or the Wald test for the overall effect of an independent variable may show it to have a statistically significant effect, but there may be individual coefficients representing the effect of that independent variable on a specific comparison that are not statistically significant.

This kind of difficulty is illustrated in the results presented in Tables 31.7 and 31.8. The overall effect of the number of prior convictions is statistically significant, but the individual coefficient for the number of prior convictions on the comparison between financial release and nonfinancial release is not statistically significant. The best approach in this type of situation is to note the significance of the overall effect of the independent variable, but to explain clearly the pattern of results for the individual coefficients. In this case, our model suggests that the number of prior convictions has an overall impact on pretrial release status, but despite this, the results do not allow us to conclude, that the number of prior convictions has a significant effect on receiving a financial release as opposed to a nonfinancial release. In multinomial logistic regression, a large number of coefficients are estimated and caution should be exercised so as not to draw selectively from the results.

OVERALL MODEL. To assess the statistical significance of the full multinomial regression model, we compute a model chi-square statistic that is identical in form to that used for the binary logistic regression model previously discussed. The model chi-square is computed as:

$$\text{Model chi-square} = (-2\text{LL}_{\text{null model}}) - (-2\text{LL}_{\text{full model}})$$

For our pretrial release status analysis, the $-2\text{LL}_{\text{null model}} = 15249.398$ and the $-2\text{LL}_{\text{full model}} = 13194.994$, resulting in a model chi-square of $15249.398 - 13194.994 = 2054.404$. The $df = 36$ for the model (3 coefficients for each of 12 independent variables), indicating that the model is statistically significant ($p < 0.001$) well beyond conventional levels of significance.

A Concluding Observation about Multinomial Logistic Regression Models

In our example using data on pretrial release in California, the dependent variable had four categories, representing a total of six different possible contrasts. Realistic applications of multinomial logistic regression models with more than four categories can quickly become unwieldy in regard to the number of contrasts that are being analyzed. For example, if we had a dependent variable with five categories, we would have four sets of coefficients to represent a total of ten different contrasts. As a way of addressing the complexity of multinomial logistic regression results, there have been some attempts to graph the coefficients (Long 1987, 1997) or plot the expected probabilities (Fox and Andersen 2006) in ways that clarify the effects of the independent variables on the dependent variable. Thus far, these methods have not been used very often in criminology and criminal justice. In turn, the complexity of results from multinomial logistic regression models has likely limited its published applications to all but the simplest results.

ORDINAL LOGISTIC REGRESSION

Thus far, the discussion has focused on logistic regression models for nominal dependent variables with two categories (binary logistic regression) and three or more categories (multinomial logistic regression). How should we analyze an ordinal dependent variable? Historically, the analysis of ordered dependent variables in criminology has tended to treat the ordinal variable either as nominal and used multinomial logistic regression or as interval and used OLS regression. Although the use of multinomial logistic regression models provides a solution to the important problem of analyzing multiple category dependent variables, it is not able to use information about the ordering of the categories, which may be theoretically and substantively important. The use of OLS regression to analyze an ordered dependent variable may not be problematic if the variable has a relatively large number of categories (Fox 2008). For an ordered dependent variable with a modest number of categories, say less than seven, the use of OLS techniques will likely be inappropriate. For example, if we examine responses to a survey question about fear of crime that was measured as a series of categories from "very fearful" to "not fearful at all," it is difficult to assume that there are equal intervals between these qualitative responses. Ordinal logistic regression models offer a means for explicitly taking into account an ordered categorical dependent variable using the logistic distribution function. A growing body of literature on ordinal logistic regression models is available to readers interested in more comprehensive treatments (e.g., Agresti 1984; Clogg and Shihadeh 1994; Long 1997; O'Connell 2006).

In order to set up the application and interpretation of the ordinal logistic regression model, we need to reconsider what a variable measured at the ordinal level represents. Ideally, an ordinal variable has ranked categories that represent an underlying continuum that cannot be directly observed and measured. For example, when respondents to a survey are presented with a statement that has response choices Strongly Agree, Agree, Disagree, and Strongly Disagree, the variable is assumed to represent an underlying continuum of agreement–disagreement with some issue. Yet, we know that however an individual responds to the question, any two individuals falling in the same category may not mean exactly the same thing. For example, if we randomly selected two individuals who had responded Strongly Disagree with a policy statement, and we were able to ask more in-depth follow-up questions, we would likely discover that there were degrees of how strongly each disagreed. The same concern would apply to comparisons within any response category.

If we assume that an ordinal variable's categories represent an underlying continuum, we can think of thresholds as those points on the continuum where an individual may move from one ordinal category to another (adjacent) category. In the example above, we could make a note of the thresholds between Strongly Agree and Agree, Agree and Disagree, and Disagree and Strongly Disagree. Figure 31.1 illustrates the link between the underlying continuum and the variable measured at the ordinal level. In Fig. 31.1, each dot represents the "true value" for an individual's attitudes about a given issue – but this true value cannot be measured directly, and we are left with the four response choices indicating degree of agreement or disagreement.[6] Each of the vertical lines marks the point between one of the possible response choices and indicates the threshold for each response category.

Proportional (Cumulative) Odds Model

The most common type of ordinal logistic regression model is known as the proportional (or cumulative) odds model. The proportional odds model represents something of a hybrid of the binary logistic and the multinomial logistic regression models. Similar to the multinomial logistic regression model's estimation of multiple model intercepts, the proportional odds model estimates multiple intercepts that represent the values of the thresholds (see again Fig. 31.1). Comparable to the binary logistic model, the proportional odds model estimates one coefficient for the effect of each independent variable on the dependent variable. In part, this is due to the added information contained in an ordinal variable, rather than a multi-category nominal variable. The interpretation of the results from the proportional odds model is also potentially much simpler than the results from the multinomial logistic model.

[6] The vertical spread of the dots is simply a convenience to illustrate the placement of cases along the Agreement–Disagreement continuum.

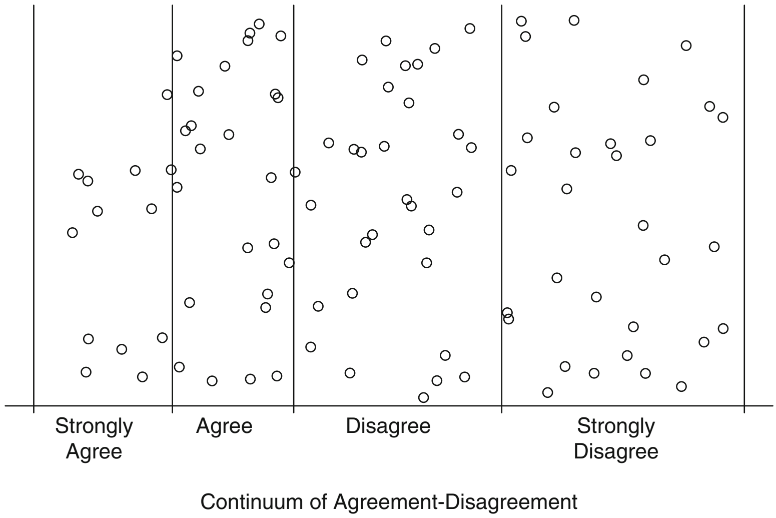

FIGURE 31.1. Hypothetical ordinal variable and underlying continuum.

One of the key differences between the proportional odds model and other logistic models is that rather than estimating the probability of a single category as in the binary and multinomial logistic regression models, this model estimates a cumulative probability (and hence the basis for its alternative name): the probability that the outcome is equal to or less than the category of interest. In equation format,

$$P(Y \leq m) = \sum_{j=1}^{m} P(Y = j),$$

where m is the category of interest and can take on values ranging from 1 to $J - 1$, while j denotes each individual category and J the total number of categories. For example, using the four response categories for agreement–disagreement above would mean that $J = 4$, and we could compute a total of $J - 1 = 4 - 1 = 3$ cumulative probabilities. If we define a Strongly Agree response as 1 and Strongly Disagree response as 4, we could then compute probabilities for $P(Y \leq 1)$, $P(Y \leq 2)$, and $P(Y \leq 3)$, representing $P(Y \leq \text{Strongly Agree})$, $P(Y \leq \text{Agree})$, and $P(Y \leq \text{Disagree})$, respectively. We would not include the final category (i.e., $P(Y \leq 4)$ or $P(Y \leq \text{Strongly Disagree})$), since it would have to be equal to 1 (or 100%) – all possible values have to fall in one of the four response categories.

Using the cumulative probabilities, we can then compute odds ratios to represent the effects of the independent variables on the dependent variable. There is an important difference in the construction and interpretation of the odds ratios, since we are now using cumulative probabilities:

$$\text{OR}_m = \frac{P(Y \leq m)}{1 - P(Y \leq m)} = \frac{P(Y \leq m)}{P(Y > m)}$$

Substantively, the odds ratio using cumulative probabilities indicates the odds of an outcome less than or equal to category m versus the odds of a category greater than m. In the context of our four response choices, the three odds ratios that we could make reference to would be the following:

- Odds of a Strongly Agree response versus the combined outcomes of Agree, Disagree, and Strongly Disagree response.
- Odds of the combined outcomes of Strongly Agree and Agree response versus the combined outcomes of Disagree and Strongly Disagree response.
- Odds of the combined outcomes of Strongly Agree, Agree, and Disagree response versus Strongly Disagree response.

Similar to the development of the binary and multinomial logistic regression models, we can use this equation for the odds ratio and rewrite it as a linear model:

$$\mathrm{OR}_m = \frac{P(Y \leq m)}{P(Y > m)} = \exp\left(\tau_m - Xb\right).$$

The general form for this equation is very similar to that of either the binary or multinomial logistic model, except that we have introduced a new term (τ_m) and now have a negative sign to the left of Xb. The (τ_m) represent the threshold parameters, which function as intercepts in the linear model and will take on values for $j = 1$ to $J - 1$.

By taking the natural logarithm of the odds ratio equation, we produce the logit equation:

$$\ln\left(\frac{P(Y \leq m)}{P(Y > m)}\right) = \ln\left[\exp\left(\tau_m - Xb\right)\right] = \tau_m - Xb,$$

which forms the basis for estimating the proportional odds model.

Interpretation of Ordinal Logistic Regression Coefficients

In our discussion of the binary logistic regression model, we illustrated how a one-unit increase in the independent variable would modify the odds of the different outcomes by a factor of $\exp(b)$. Since the equation for the proportional odds model takes a slightly different form, we cannot simply exponentiate b to obtain the effect of the independent variable.

To illustrate the modification, suppose we have two values of X: x and $x + 1$. The odds ratio for x and $x + 1$ would be

$$\frac{\mathrm{OR}_m\,(X+1)}{\mathrm{OR}_m\,(X)} = \frac{\exp\left(\tau_m - (X+1)b\right)}{\exp\left(\tau_m - Xb\right)} = \exp\left(\left[X - (X+1)\right]b\right) = \exp\left(-b\right).$$

Thus, to interpret the effect of a one-unit change in the independent variable in the proportional odds model, we will need to exponentiate the *negative* value of the estimated coefficient. We can then interpret the coefficient as indicating the odds of an outcome less than or equal to category m versus the odds of a category greater than m for a one-unit change in the independent variable.

TABLE 31.9. Proportional odds model results for Fayetteville youth study

Variable	Coefficient	Standard error	*z*-Score
Age	−0.06	0.074	−0.76
Male	0.28	0.147	1.93
White	−0.48	0.193	−2.49
GPA	−0.40	0.078	−5.12
Friends picked up by the police	1.13	0.167	6.77
Parents know where youth is while away	−0.68	0.148	−4.61
Have to do some things that are not right to get ahead	0.40	0.163	2.42
OK to get around the law if you could get away with it	0.57	0.168	3.41
OK to take things from big business	1.03	0.193	5.34
tau-1	−0.87	1.231	
tau-2	0.94	1.229	
tau-3	2.00	1.231	

An Example: Fayetteville Youth Project

Returning to the FYP data, we use the same model for self-reported theft as we used in the section on binary logistic regression, except that the dependent variable is kept in its original ordinal level of measurement: Never, Once or twice, Several times, and Many times. Table 31.9 presents the results for the ordinal logistic regression model; included in the table are the coefficient estimates, standard errors, and *z*-statistics to assess the statistical significance of the individual coefficients.

INTERPRETING THE COEFFICIENTS. While the proportional odds model accounts for the fact that the categories in the dependent variable are ranked, for example in our case from no self-reported theft to frequent self-reported thefts, the interpretation of the coefficients is similar to that used in multinomial regression. In this case, we can compare lower categories to the categories ranked above them. Since we have four ordered categories, there are three comparisons we can make: (1) no theft (i.e., Never) to one or more thefts, (2) no theft and once or twice to several and many times, and (3) no theft, once or twice, and several times to many times. In all cases, the exponent of the negative of the coefficient provides the odds ratio for change from a category indicating less delinquency to a category indicating more delinquency. Overall, the pattern of results in Table 31.9 indicates that youth who were male, non-white, younger, had lower GPAs, had friends picked up by the police, were not well supervised by their parents, and felt that it was okay to break the law or to take things were more likely to self-report higher levels of theft. The following discussion is intended to illustrate the direct interpretation of several of these coefficients.

We see that the coefficient for male is 0.283. By exponentiating the negative of 0.283 ($\exp(-0.283) = 0.754$), we see that males are likely to self-report more frequent theft than females, controlling for the other independent variables in the model. More concretely, we can state the following about the self-reported theft of male youth:

- The odds of self-reporting no theft versus at least one theft are 0.754 times smaller for males than for females, controlling for all other variables in the model.
- The odds of self-reporting theft as never and once or twice versus self-reporting theft as several times and many times are 0.754 times smaller for males than for females, controlling for all other variables in the model.

- The odds of self-reporting theft as never, once or twice, and several times versus self-reporting theft as many times are 0.754 times smaller for males than for females, controlling for all other variables in the model.

The effect of race on self-reported theft is $\exp(-(-0.482)) = 1.619$. Writing out direct interpretations of this coefficient leads to the following statements:

- The odds of self-reporting no theft versus at least one theft are 1.619 times greater for whites than for non-whites, controlling for all other variables in the model.
- The odds of self-reporting theft as never and once or twice versus self-reporting theft as several times and many times are 1.619 times greater for whites than for non-whites, controlling for all other variables in the model.
- The odds of self-reporting theft as never, once or twice, and several times versus self-reporting theft as many times are 1.619 times greater for whites than for non-whites, controlling for all other variables in the model.

The effect of having one or more friends picked up by the police is $\exp(-1.132) = 0.322$, indicating that youth who have had at least one friend picked up tend to self-report higher levels of theft. The direct interpretations are:

- The odds of self-reporting no theft versus at least one theft are 0.322 times smaller for youth who have friends picked up by the police than those who have had none picked up by the police, controlling for all other variables in the model.
- The odds of self-reporting theft as never and once or twice versus self-reporting theft as several times and many times are 0.322 times smaller for youth who have friends picked up by the police than those who have had none picked up by the police, controlling for all other variables in the model.
- The odds of self-reporting theft as never, once or twice, and several times versus self-reporting theft as many times are 0.322 times smaller for youth who have friends picked up by the police than those who have had none picked up by the police, controlling for all other variables in the model.

The coefficient for GPA is -0.402. Since GPA is measured at the interval level of measurement, we would note that for a one-unit increase in GPA, the odds ratio changes by a factor of $\exp(-(-0.402)) = 1.495$, controlling for the other variables in the model. We can write out the interpretations as follows:

- The odds of self-reporting no theft versus at least one theft increase by a factor of 1.495 for a one-unit increase in GPA, controlling for all other variables in the model.
- The odds of self-reporting theft as never and once or twice versus self-reporting theft as several times or many times increase by a factor of 1.495 for a one-unit increase in GPA, controlling for all other variables in the model.
- The odds of self-reporting theft as never, once or twice, and several times versus self-reporting theft as many times increase by a factor of 1.495 for a one-unit increase in GPA, controlling for all other variables in the model.

Note, too, that there are three threshold parameters representing the threshold points between each of the ordered categories (i.e., never, once or twice, several times, and many times).

Table 31.9 also reports the values of the z-statistic for each independent variable. Note that the only coefficient not statistically significant at a level of 5% is the effect of age. The statistical significance of the overall model is based on a model chi-square statistic that is also computed and interpreted in exactly the same way as for the binary logistic and multinomial logistic regression models. In our example, the $-2\text{LL}_{\text{null model}} = 1968.274$ and the $-2\text{LL}_{\text{full model}} = 1646.683$, resulting in a model chi-square of $1968.274 - 1646.683 = 326.591$. Since there are 9 coefficients that have been estimated (one for each independent variable), the degree of freedom value for this test is equal to 9, resulting in a model that has a high level of statistical significance ($p < 0.001$), which means that the overall model has a statistically significant effect on self-reported theft.

Parallel Slopes Tests

As we noted earlier, the proportional odds model assumes that the effects of the independent variables are constant across all categories of the dependent variable, which is analogous to our interpretation of coefficients in a multivariate linear regression model. Regardless of the level (or category) of the dependent variable, we expect the independent variable to exert a constant (i.e., proportional) effect on the dependent variable. The constant effect of each independent variable should have also been clear in the direct interpretations of the coefficients noted in the previous section. This is known more generally as the parallel slopes assumption. Most statistical packages include a score test of this assumption that informs the user regarding the appropriateness of the ordinal logistic model. Somewhat less common is the Brant Test, which tests for parallel slopes in the overall model and in each independent variable.

Score Test. Conceptually, the parallel slopes score test is based on the idea that we could estimate a series of $J - 1$ binary logistic regression models (i.e., one model less than the number of ordered categories in the dependent variable) of the form $P(Y \leq m)$ that allowed the effects of all K independent variables to vary by outcome category on the dependent variable. The test would then focus on whether a single coefficient or multiple coefficients best represented the effects of the independent variables on the dependent variable. Technically, the score test uses information about the log-likelihood for the ordinal logistic regression model, and assesses how much it would change by allowing the coefficients for all the independent variables to vary by the outcome category on the dependent variable. The degree of change in the likelihood function then indicates whether the parallel slopes assumption is met. The null hypothesis of the score test is parallel (equal) slopes. The research hypothesis is that the slopes are not parallel (equal). The value of the score test (reported in most statistical software) is distributed as a chi-square with $K(J - 2)$ degrees of freedom.

For our self-reported theft example, we have $K = 9$ (i.e., nine independent variables) and $J = 4$ (i.e., four outcome categories on the dependent variable). The corresponding degrees of freedom for our score test is equal to $9(4 - 2) = 18$. The value of the score test for our model is 16.217 ($p = 0.577$), which indicates that we should not reject our null hypothesis of parallel slopes and conclude that our model does indeed meet the parallel slopes assumption.

Brant Test. Similar to the Score Test, the Brant Test (Brant 1990) is a Wald test that assesses whether all the coefficients in a proportional odds model satisfy the parallel slopes assumption. The computation of the Brant Test is based on the values of the coefficients and their respective variances. Readers interested in a step-by-step presentation of the computation

TABLE 31.10. Brant test results for Fayetteville youth study

Variable	χ^2	*df*	*p*
Overall	14.57	18	0.691
Age	0.22	2	0.894
Male	4.87	2	0.087
White	1.59	2	0.452
GPA	0.60	2	0.742
Friends picked up by the police	1.16	2	0.561
Parents know where youth is while away	0.63	2	0.731
Have to do some things that are not right to get ahead	0.00	2	0.999
OK to get around the law if you could get away with it	1.33	2	0.514
OK to take things from big business	1.80	2	0.406

of the Brant Test should consult Long (1997).[7] In addition to providing an overall test for the parallel slopes assumption, the Brant Test can be decomposed into values for each of the independent variables in the ordinal logistic regression model.

The Brant Test for the overall model will be distributed as a chi-square with $K(J-2)$ degrees of freedom (same as in the Score Test). The test statistic of each independent variable is distributed as a chi-square with $J-2$ degrees of freedom.

The results of the Brant Test for the self-reported theft example appear in Table 31.10. The overall test again indicates that the parallel slopes assumption is satisfied for the model (chi-square $= 14.57$, $df = 18$, $p = 0.691$). Similarly, each of the independent variables in the model also satisfies the parallel slopes assumption.

Partial Proportional Odds

Although the parallel slopes assumption was satisfied in the preceding example, it is quite common for the assumption not to be met in practice. Historically, researchers confronted with results from the Score Test indicating the model failed to satisfy the parallel slopes assumption were left with a choice of fitting the proportional odds model and violating assumptions or fitting a multinomial logistic regression model by ignoring the ordinal nature of the dependent variable and then complicating the interpretation of the results through the increased number of coefficients. Recently, a class of models referred to as partial proportional odds or generalized ordinal logistic regression models has received increasing attention (Fu 1998; Lall et al. 2002; O'Connell 2006; Peterson and Harrell 1990; Williams 2005). The logic to the partial proportional odds model is to allow some or all of the coefficients of the independent variables to vary by the level of the dependent variable, much like what we see in the application of multinomial logistic regression, but to constrain other coefficients to have a single value, as in the proportional odds model.

[7] Long and Freese (2006) have written a procedure for Stata to compute the Brant Test.

We obtain the partial proportional odds model by generalizing the proportional odds equation from Sect. "Proportional (Cumulative) Odds Model" to allow the coefficients (the b_m) to vary by the level of the dependent variable (m):

$$\ln\left(\frac{P(Y \leq m)}{P(Y > m)}\right) = \ln\left[\exp\left(\tau_m - Xb_m\right)\right] = \tau_m - Xb_m.$$

Without any further constraints on the coefficients, the total number of coefficients estimated will be identical to that obtained from a multinomial logistic regression analysis. It is important to note, however, that the coefficients do not mean the same thing. Recall that the multinomial logistic regression coefficients refer to comparisons between a given category and the reference category. As noted in the equation above, the logit in the partial proportional odds model is identical to that in the proportional odds model and refers to the odds of a category less than or equal to m versus a category greater than m.

Due to the large number of coefficients in a fully generalized ordinal logit model, most researchers will want to limit the number of variables with nonconstant effects. The results from the Brant Test are useful for determining which, if any, independent variables appear to have varying effects on the different categories of the dependent variable (i.e., the effects are not parallel). If the overall Brant Test result is not statistically significant, it implies that the parallel slopes assumption is met for the full model. In this case, there is likely little to be gained by relaxing the parallel slopes assumption for a single variable. In those cases where the overall Brant Test result is statistically significant, then the Brant Test results for individual variables will point to those variables with the greatest divergence from the parallel slopes assumption and the best candidates for allowing the effects to vary across the different ordinal logits.

All other features of the partial proportional odds model – tests for statistical significance, interpretation of the coefficients, and the like – are the same as found in the proportional odds model.

AN EXAMPLE: SENTENCING DECISIONS IN CALIFORNIA. In our discussion of multinomial logistic regression, we presented the results from an analysis of data from California pretrial release decisions in the 1990s. These same data can be used to illustrate the application and interpretation of the partial proportional odds model. The following example uses the same data, but restricts it to those cases where the offender was ultimately convicted of a crime and sentenced to probation, jail, or prison ($N = 4{,}765$). The independent variables used in this example include a similar set of offender and case characteristics: age, sex, race, whether the offender had a private attorney, type of offense charged, total number of charges, whether the offender was under supervision at the time of the offense, the number of prior felony arrests, and the number of prior violent convictions. Descriptive statistics for this sample of offenders appear in Table 31.11.

Table 31.12 presents the results for the proportional odds model (column 1) and the Brant Test (column 2). We see from the values of the coefficients appearing in column 1 that offenders who were male, younger, charged with a violent crime, under criminal justice supervision at the time of the offense, and did not have a private attorney were more likely to receive more severe forms of punishment (jail and/or prison). In a similar way, as the total number of charges, prior felony arrests, and prior violent convictions increased, offenders were increasingly likely to receive jail and/or prison sentences rather than probation sentences.

TABLE 31.11. Variable coding and descriptive statistics for sentencing decisions in California ($N = 4{,}765$)

Variable	Mean	Standard deviation
Age (years)	31.15	9.05
Male (1 = male, 0 = female)	0.81	0.40
Black (1 = black, 0 = non-black)	0.44	0.50
Hispanic (1 = Hispanic, 0 = non-Hispanic)	0.04	0.21
Under criminal justice supervision (1 = yes, 0 = no)	0.51	0.50
Represented by private attorney (1 = yes, 0 = no)	0.17	0.38
Total number of charges	2.30	1.66
Charged with violent crime (1 = yes, 0 = no)	0.21	0.41
Charged with property crime (1 = yes, 0 = no)	0.34	0.47
Charged with drug crime (1 = yes, 0 = no)	0.38	0.48
Number of prior felony arrests	4.91	6.64
Number of prior violent convictions	0.17	0.67

TABLE 31.12. Proportional odds model and Brant test results for sentencing in California

Variable	Coefficient (SE)	Brant test χ^2 (p-value)[a]
Age	−0.01 (0.003)	8.38 (0.004)
Male	0.53 (0.074)	0.47 (0.491)
Black	0.07 (0.060)	0.10 (0.756)
Hispanic	−0.27 (0.140)	1.98 (0.160)
Under criminal justice supervision	0.70 (0.061)	5.73 (0.017)
Represented by private attorney	−0.18 (0.078)	0.67 (0.414)
Total number of charges	0.03 (0.018)	0.54 (0.461)
Charged with violent crime	0.39 (0.122)	0.57 (0.449)
Charged with property crime	0.14 (0.115)	0.59 (0.441)
Charged with drug crime	0.06 (0.114)	7.11 (0.008)
Number of prior felony arrests	0.07 (0.005)	16.20 (<0.001)
Number of prior violent convictions	0.63 (0.072)	6.66 (0.010)
Overall ($df = 12$)		90.74 (<0.001)

[a] Individual coefficient tests based on $df = 1$.

The results from the Brant Test in Column 2 show that the overall model fails to satisfy the parallel slopes assumption ($\chi^2 = 90.74,\ df = 12,\ p < 0.001$).[8] When the overall value is decomposed into the effects for each of the independent variables, we see that the coefficients for age, drug charge, under criminal justice supervision, the number of prior felony arrests, and the number of prior violent convictions do not satisfy the parallel slopes assumption, while those for all other independent variables do.

Table 31.13 presents the results from a partial proportional odds model where all coefficients are allowed to have different effects on the two ordinal logits:

$$\ln\left(\frac{P(Y \leq 1)}{P(Y > 1)}\right) = \ln\left(\frac{P(Y = \text{probation})}{P(Y = \text{jail or prison})}\right) = \tau_1 - Xb_1$$

[8] The Score Test similarly indicates a failure of the full model to satisfy the parallel slopes assumption ($\chi^2 = 81.39,\ df = 12,\ p < 0.001$).

TABLE 31.13. Partial proportional odds model for sentencing in California – all coefficients allowed to vary

Variable	Probation vs. jail and/or prison Coefficient (SE)	Probation and/or jail vs. prison Coefficient (SE)
Age	−0.02 (0.004)	−0.01 (0.004)
Male	0.51 (0.094)	0.60 (0.098)
Black	0.06 (0.087)	0.10 (0.070)
Hispanic	−0.08 (0.187)	−0.48 (0.194)
Under criminal justice supervision	0.57 (0.087)	0.81 (0.070)
Represented by private attorney	−0.23 (0.102)	−0.14 (0.097)
Total number of charges	0.02 (0.024)	0.04 (0.020)
Charged with violent crime	0.28 (0.167)	0.43 (0.138)
Charged with property crime	0.19 (0.158)	0.09 (0.132)
Charged with drug crime	0.30 (0.156)	−0.09 (0.132)
Number of prior felony arrests	0.04 (0.008)	0.07 (0.006)
Number of prior violent convictions	0.45 (0.099)	0.62 (0.073)
Constant (τ)	1.25 (0.232)	−2.17 (0.211)

and

$$\ln\left(\frac{P(Y \leq 2)}{P(Y > 2)}\right) = \ln\left(\frac{P(Y = \text{probation or jail})}{P(Y = \text{prison})}\right) = \tau_2 - Xb_2.$$

Since there are two different ordinal logits being estimated, there are two full sets of unique coefficients to interpret that illustrate the different effects the independent variables have on the two different ordered logits. To highlight a few of the findings in Table 31.13:

- Age:
  - The odds of probation versus jail or prison increase by a factor of $\exp(-(-0.02)) = 1.020$ for a one-unit increase in age, controlling for all other variables in the model. For a 10 year increase in age, the odds of probation versus a jail or prison sentence increase by a factor of $\exp(-10 \times -0.02) = 1.221$, controlling for all other variables in the model.
  - The odds of probation or jail versus prison increase by a factor of $\exp(-(-0.007)) = 1.007$ for a one-unit increase in age, controlling for all other variables in the model. For a 10 year increase in age, the odds of probation versus a jail or prison sentence increase by a factor of $\exp(-10 \times -0.007) = 1.073$, controlling for all other variables in the model.
- Drug charge:
  - The odds of probation versus jail or prison are $\exp(-0.305) = 0.737$ times smaller for offenders charged with a drug offense than for offenders charged with a miscellaneous offense, controlling for all other variables in the model.
  - The odds of probation or jail versus prison are $\exp(-(-0.092)) = 1.097$ times greater for offenders charged with a drug offense than for offenders charged with a miscellaneous offense, controlling for all other variables in the model.

TABLE 31.14. Partial proportional odds model for sentencing in California – selected coefficients allowed to vary

	Constrained	Probation vs. jail and/or prison	Probation and/or jail vs. prison
Variable	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)
Age		−0.02 (0.004)	−0.01 (0.004)
Male	0.55 (0.075)		
Black	0.09 (0.061)		
Hispanic	−0.27 (0.143)		
Under criminal justice supervision		0.57 (0.087)	0.80 (0.070)
Represented by private attorney	−0.18 (0.079)		
Total number of charges	0.04 (0.018)		
Charged with violent crime	0.39 (0.121)		
Charged with property crime	0.13 (0.114)		
Charged with drug crime		0.30 (0.128)	−0.09 (0.119)
Number of prior felony arrests		0.04 (0.007)	0.07 (0.006)
Number of prior violent convictions		0.43 (0.100)	0.63 (0.072)
Constant (τ)		1.19 (0.196)	−2.10 (0.191)

- CJ supervision:
  - The odds of probation versus jail or prison are $\exp(-0.575) = 0.563$ times smaller for offenders under criminal justice supervision at the time of offense than for offenders not under supervision, controlling for all other variables in the model.
  - The odds of probation or jail versus prison are $\exp(-0.807) = 0.446$ times smaller for offenders under criminal justice supervision at the time of offense than for offenders not under supervision, controlling for all other variables in the model.

Table 31.14 presents the results for the partial proportional odds model, where the only coefficients allowed to vary across the two ordinal logits are those identified through the Brant Test as not satisfying the parallel slopes assumption: age, drug charge, criminal justice supervision, prior felony arrests, and prior violent convictions. Note that the coefficients are virtually identical to those appearing in Table 31.13. Limiting the discussion to the effects of prior felony arrests and violent convictions, we see the following:

- Number of prior felony arrests:
  - The odds of probation versus jail or prison decrease by a factor of $\exp(-0.038) = 0.963$ for a one-unit increase in the number of prior felony arrests, controlling for all other variables in the model.
  - The odds of probation or jail versus prison decrease by a factor of $\exp(-0.070) = 0.932$ for a one-unit increase in the number of prior felony arrests, controlling for all other variables in the model.
- Number of prior violent convictions:
  - The odds of probation versus jail or prison decrease by a factor of $\exp(-0.432) = 0.649$ for a one-unit increase in the number of prior violent convictions, controlling for all other variables in the model.

– The odds of probation or jail versus prison decrease by a factor of $\exp(-0.631) = 0.532$ for a one-unit increase in the number of prior violent convictions, controlling for all other variables in the model.

Substantively, and consistent with much prior research, these findings indicate that as the severity of the offender's prior record increases, the likelihood of a prison sentence increases over the chances of a probation or a jail sentence.

Note on Statistical Software

Statistical software packages vary in their ability to estimate the models discussed in this chapter. All of the results presented in this chapter were estimated with multiple software packages, including LIMDEP (Greene 2008), R (R Development Core Team 2003), SAS (SAS Institute 2009), SPSS (SPSS Inc 2007), and Stata (Stata Corporation 2009). Of the commercially available software packages, Stata offers the widest range of procedures in a user-friendly environment that includes both command-line driven analysis and drop-down menus (which in turn generate the command-line syntax). The R system is an open source program. Various packages within the R system offer a wider selection of statistical procedures than is available in Stata, but the learning curve may be too steep for the preferences of some potential users. LIMDEP, SAS, and SPSS easily estimate the binary logit, multinomial logit, and proportional odds models. With additional syntax, it is possible to estimate the partial proportional odds model in SAS and SPSS (see O'Connell 2006 for examples).

SUMMARY AND DIRECTIONS FOR FUTURE APPLICATIONS

The focus of this chapter has been the application and the interpretation of logit models for variables with categorical dependent variables. For those situations where we have measured the dependent variable as a dichotomy, we can apply binary logistic regression. When the dependent variable has three or more categories, we need to consider the level of measurement: Are the categories ordered or unordered? If the categories are not ordered, then the application of a multinomial logistic regression model is most appropriate. The primary caution to the analyst is to watch the number of categories – too many categories can result in an unwieldy set of coefficients to interpret. If the categories are ordered, then an ordinal logit model is most appropriate. Key to the application and interpretation of the ordinal logit model is the satisfaction of the parallel slopes assumption. When the parallel slopes assumption is satisfied, then the proportional odds model provides the most parsimonious way to assess the effects of the independent variables on the dependent variable since the effects of the independent variables are constrained to be the same across all of the ordinal logits. When the parallel slopes assumption is not satisfied, the partial proportional odds model is a more effective choice, but the researcher will need to consider whether to allow all coefficients to vary across all the ordinal logits or to allow only individual variables that fail the parallel slopes assumption to vary.

In this chapter, there are several issues that were mentioned in passing that provide a basis for future applications of logit models in criminology and criminal justice. First, the presentation of results from multinomial logistic and ordinal logistic (proportional or partial proportional odds) models remains a challenge, due to the large number of coefficients. A description of the findings, as in this chapter, necessarily limits the amount of information that can be conveyed to the reader and tends to result in a more superficial discussion of the findings. The suggestions for more comprehensive interpretations of the results from these models have relied on the use of effects displays. Long (1997, 1987), for example, has proposed a graphical method for presenting coefficients from multinomial logit models that conveys information about statistical significance as well as information about the magnitude of the effects for all possible comparisons. Fox (2003, 2008) and Fox and Andersen (2006) have proposed a more general framework for effects displays that are applicable to all generalized linear models. Although these methods offer a way of simplifying the presentation and interpretation of numerous coefficients and of highlighting interesting findings in an analysis, they have not yet been widely used in criminology and criminal justice.

Second, research on crime and the criminal justice system is often confronted with a dichotomous outcome variable that occurs infrequently. For example, studies of self-reported delinquency will find a small number of cases of self-reported sexual assault. Similarly, a study of sentencing decisions within a single state may find few offenders who receive a sentence of death or life without parole. When the distribution of cases for an outcome variable shows less than about 5% of the observations in the category of interest, the coefficients from a binary logistic regression will be biased toward under-predicting the outcome of interest. King and Zeng (2001a, b) have proposed a "rare events" logistic regression model that is essentially a corrective procedure that adjusts the values of the coefficients and their standard errors to account for the bias in the coefficients. King and Zeng (2001a, b) also explain that the same corrective procedure can be used in the logistic regression analysis of small samples (<200 cases) with less skewed distributions on the dependent variable.[9] Aside from Piquero et al.'s (2005) application of the rare events logistic regression procedure to the analysis of homicide victimization and violent offending, the rare events procedure has not been widely used in criminology and criminal justice.

Third, conditional logit models (McFadden 1973) offer a potentially interesting approach to studying crime and the criminal justice system ((Long 1997) and Maddala (1983) also present brief discussions of this model). Conceptually, conditional logit models assume that an individual is confronted with a number of choices and picks one. Each choice has a set of attributes or unique characteristics that affect its probability of selection. Individual characteristics also affect the probability of selection, but these are invariant across the number of possible choices. The conditional logit model takes both choice attributes and individual characteristics into account.[10] With the exception of Phillips's (2003) application of the conditional logit model to the study of interpersonal violence, this model has not been used in criminology and criminal justice. In large part, the difficulty researchers will confront – especially those using secondary data sources – when trying to use this model is the lack of information on the attributes of the possible choices. For example, conditional logit models could be used to study the types of offenses individuals commit, but this would require a much better understanding of the costs and the benefits of each type of offense (i.e., the attributes of the choices) than is found in the typical analysis of offending patterns.

[9] King and Zeng (2001a, b) also note that their corrective procedure is not applicable to situations where the researcher is studying a rare event in a small sample.

[10] Through some algebraic manipulation, the conditional logit model can be shown to be equivalent to the multinomial logistic regression model (see, for example, Long 1997).

Finally, nested logit models (McFadden 1978, 1981) may offer an important avenue for understanding sequences of decisions involved in making a choice. Similar to the conditional logit model, the nested logit model assumes that an individual is confronted with a set of possible choices, each choice having a unique set of attributes that affect its probability of selection, along with the characteristics of the individual making the choice. Where nested logit models differ from conditional logit models is the assumption that the set of choices can be grouped into some smaller set of choices that essentially represent stages in a decision. For example, demographers have used nested logit models to study the resolution of a pregnancy (e.g., Plotnick 1992; South and Baumer 2001). The first stage represents whether the individual becomes pregnant and the second stage is how the pregnancy is resolved (i.e., abortion, birth, or some other form of involuntary termination).

Although nested logit models have not yet been used in the study of crime and criminal justice, there would seem to be a number of potentially relevant applications. The analysis of sentencing decisions provides an interesting possibility. A first stage may be the decision by the judge on whether or not to incarcerate an offender. Following that decision, the judge has a different set of choices that are contingent on the first choice. For those offenders being incarcerated, there may be a choice of whether to incarcerate the offender in a jail or in a prison. For those offenders not being incarcerated, there are again several choices available to many judges: probation, fine, community service, drug treatment, or some other form of therapy or intervention. Thus far, researchers who have tried to separate the various choices (probation, jail, prison) have treated all of the decisions as if they were simultaneous possibilities (e.g., Holleran and Sphon 2004) when there are clear contingencies on the final decision of the judge. Nested logit models may offer an important means for gaining a better understanding of not only the decision-making process involved in the sentencing of convicted offender, but a wide range of other criminal justice decisions as well as the choices made by offenders.

In light of the many qualitative outcome measures in the study of crime and the criminal justice system, logit models should be an important tool for criminological researchers. This chapter has outlined the primary logit models that most researchers will want to consider using in their analyses. Important to all of these applications, however, is sensitivity to the nature of the data and the research question motivating the analysis.

REFERENCES

Agresti A (1984) Analysis of ordinal categorical data. Wiley, New York

Agresti A (2002) Categorical data analysis, 2nd ed. Wiley, New York

Brant R (1990) Assessing proportionality in the proportional odds model for ordinal logistic regression. Biometrics 46:1171–1178

Clogg C, Shihadeh E (1994) Statistical models for ordinal data. Sage, Thousand Oaks, CA

Fox J, Andersen R (2006) Effect displays for multinomial and proportional-odds logit models. Sociol Methodol 36:225–255

Fox J (2003) Effect displays in R for generalised linear models. J Stat Softw 8:1–27

Fox J (2008) Applied regression analysis and generalized linear models, 2nd ed. Sage, Thousand Oaks, CA

Fu V (1998) Estimating generalized ordered logit models. Stata Technical Bulletin 8:160–164

Greene W (2008) LIMDEP 9.0 Econometric modeling guide, volumes 1 and 2. Econometric Software, Plainville, NY

Holleran D, Sphon C (2004) On the use of the total incarceration variable in sentencing research. Criminology 42:211–240

Hosmer, and Lemeshow 2000. Applied logistic regression, 2nd ed. Wiley, New York

King G, Zeng L (2001a) Explaining rare events in international relations. International Organization 55:693–715

King G, Zeng L (2001b) Logistic regression in rare events data. Polit Anal 9:137–163

Lall R, Walters SJ, Morgan K, MRC CFAS Co-operative Institute of Public Health (2002) A review of ordinal regression models applied on health-related quality of life assessments. Stat Methods Med Res 11:49–67

Long JS (1987) A graphical method for the interpretation of multinomial logit analysis. Sociological Methods and Research 15:420–446

Long JS (1997) Regression models for categorical and limited dependent variables. Sage, Thousand Oaks, CA

Long JS, Freese J (2006) Regression models for categorical dependent variables using Stata, 2nd ed. Stata, College Station, TX

Maddala GS (1983) Limited-dependent and qualitative variables in econometrics. Cambridge University Press, New York

McFadden D (1973) Conditional logit analysis of qualitative choice behavior. In: Zarembka P (ed) Frontiers of econometrics. Academic, New York, pp 105–142

McFadden D (1978) Modelling the choice of residential location. In: Karlquist A, et al. (eds) Spatial interaction theory and residential location. North-Holland, Amsterdam, pp 75–96

McFadden D (1981) Econometric models of probabilistic choice. In: Manski CF, McFaden D (eds) Structural analysis of discrete data: with econometric applications. MIT, Cambridge, MA

Menard S (2002) Applied logistic regression analysis, 2nd ed. Sage, Thousand Oaks, CA

O'Connell AA (2006) Logistic regression models for ordinal response variables. Sage, Thousand Oaks, CA

Pampel F (2000) Logistic regression: a primer. Sage, Thousand Oaks, CA

Peterson B, Harrell FE (1990) Partial proportional odds models for ordinal response variables. Appl Stat 39:205–217

Phillips S (2003) Social structure of vengeance: a test of Black's model. Criminology 41:673–708

Piquero AR, MacDonald J, Dobrin A, Daigle LH, Cullen FT (2005) Self-control, violent offending, and homicide victimization: assessing the general theory of crime. J Quant Criminol 21:55–71

Plotnick RD (1992) The effects of attitudes on teenage premarital pregnancy and its resolution. Am Sociol Rev 57:800–811

R Development Core Team (2003) R: A language and environment for statistical computing. http://www.R-project.org (accessed 16 November 2009)

SAS Institute (2009) SAS 9.2. SAS Institute, Cary, NC

South S, Baumer E (2001) Community effects on the resolution of adolescent premarital pregnancy. J Fam Issues 22:1025–1043

SPSS Inc (2007) SPSS 16.0 for Windows. SPSS, Chicago

Stata Corporation (2009) Stata 10. Stata Corporation, College Station, TX

Williams R (2005) Gologit2: a program for generalized logistic regression/partial proportional odds models for ordinal variables. http://www.nd.edu/∼rwilliam/gologit2/gologit2.pdf, Accessed April, 2009