Multinomial and ordinal logistic regression using PROC LOGISTIC
Peter L. Flom National Development and Research Institutes, Inc

ABSTRACT
Logistic regression may be useful when we are trying to model a categorical dependent variable (DV) as a function of one or more independent variables. This paper reviews the case when the DV has more than two levels, either ordered or not, gives and explains SAS® code for these methods, and illustrates them with examples.

Keywords: Ordinal Multinomial Logistic.

INTRODUCTION
In logistic regression, the goal is the same as in ordinary least squares (OLS) regression: we wish to model a dependent variable (DV) in terms of one or more independent variables (IVs). However, OLS regression is for continuous (or nearly continuous) DVs; logistic regression is for DVs that are categorical. The DV may have two categories (e.g., alive/dead; male/female; Republican/Democrat) or more than two categories. If it has more than two categories they may be ordered (e.g. none/some/a lot) or unordered (e.g. married/single/divorced/widowed/other). This paper deals with modeling multiple category DVs (ordered or not) with SAS PROC LOGISTIC.

WHY LOGISTIC REGRESSION IS NEEDED
One might try to use OLS regression with categorical DVs. There are several reasons why this is a bad idea:

1. The residuals cannot be normally distributed (as the OLS model assumes), since they can only take on one of several values for each combination of level of the IVs
2. The OLS model makes nonsensical predictions, since the DV is not continuous - e.g., it may predict that someone does something more than 'all the time'.
3. For nominal DVs, the coding is completely arbitrary, and for ordinal DVs it is (at least supposedly) arbitrary up to a monotonic transformation. Yet recoding the DV will give very different results.

A VERY QUICK INTRODUCTION TO LOGISTIC REGRESSION
Logistic regression deals with these issues by transforming the DV. Rather than using the categorical responses, it uses the log of the odds ratio of being in a particular category for each combination of values of the IVs. The odds is the same as in gambling, e.g., 3-1 indicates that the event is three times more likely to occur than not. We take the ratio of the odds in order to allow us to consider the effect of the IVs. We then take the log of the ratio so that the final number goes from $-\infty$ to $\infty$, so that 0 indicates no effect, and so that the result is symmetric around 0, rather than 1. For more details on logistic regression, see Hosmer and Lemeshow (2000), Agresti (2002), or Long (1997).

MODEL SELECTION
Methods such as forward, backward, and stepwise selection are available, but, in logistic as in other regression methods, are not to be recommended. They give incorrect estimates of the standard errors and p-values, can delete variables that are critical to include, and, perhaps most important, allow the researcher not to think (Harrell, 2001). It is much better to compare models based on their results, reasonableness, and fit (as measured, e.g. by the Akaike Information Criterion (AIC) — note that a lower AIC indicates better fit). A good text on this is Burnham and Anderson (2002).

ORDINAL LOGISTIC REGRESSION
THE MODEL
As noted, ordinal logistic regression refers to the case where the DV has an order; the multinomial case is covered below. The most common ordinal logistic model is the proportional odds model. If we pretend that the DV is really continuous, but is recorded ordinally (as might, for instance, happen if income were asked about in terms of ranges, rather than precise numbers), but that it has been divided into $J$ categories then if the ‘real’ DV is $Y^*$, the model is

$$y_j^* = x \beta + \epsilon_i$$
However, since the DV is categorized, we must instead use
\[
c_k(x) = \ln \frac{P(Y \leq j | x)}{P(Y > j | x)} 
= \ln \frac{\phi_0(x) + \phi_1(x) + \ldots + \phi_j(x)}{\phi_{j+1}(x) + \phi_{j+2}(x) + \ldots + \phi_J(x)} 
= \tau_j - x^t \beta
\]
where \( \tau_j \) are the cutpoints between the categories, and \( \phi_i(x) \) is the probability of being in class \( i \) given covariates \( x \).

**THE PROPORTIONAL ODDS ASSUMPTION**

The proportional odds assumption is then that the \( \beta \)s are independent of \( j \) (note that the \( \beta \)s have no subscripts). In other words, it assumes that if we looked at (binary) logistic regressions of category 1 vs. 2, category 2 vs. 3, and so on, then the intercepts in the equations might vary, but the parameters would be identical for each model. SAS uses the score test to test the proportional odds assumption, but this test is anticonservative (that is, it rejects the assumption too often); for details on this test see [SAS Institute, Inc., 2004]. Graphical methods are better [Harrell 2001], but are not yet available in SAS. Another method is to compare the ordinal model with the binomial models, and determine whether the slopes are meaningfully different.

If the proportional odds assumption is not met, there are several options:

1. Collapse two or more levels, particularly if some of the levels have small N.
2. Do bivariate ordinal logistic analyses, to see if there is one particular IV that is operating differently at different levels of the DV.
3. Use the partial proportional odds model (available in SAS through PROC GENMOD).
4. Use multinomial logistic regression (see below).

**CHECKING MODEL FIT, RESIDUALS AND INFLUENTIAL POINTS**

Assessment of fit, residuals, and influential points can be done by the usual methods for binomial logistic regression, performed on each of \( j - 1 \) regressions. SAS has extensive facilities for this, including the excellent ODS graphics (new to version 9), but a discussion of these is beyond the scope of the current paper.

**EXAMPLE**

In a sample of youth from Bushwick, NY, I looked at the relationship between drug use (none, marijuana, hard drugs) and sex, age, and a factor representing peer norms about drug use [Flom et al. 2001]. The norms factor was based on responses to questions such as ‘How many of your friends encourage you to use ____’ and ‘How many of your friends would object if you used ____ ’ where the blanks were filled in with different drugs.

I compared several models which seemed substantively reasonable:

1. A null model (with no covariates)
2. A main effects model
3. A model with only the norm factor and
4. A model with 2-way interactions among the three IVs

The AICs for these models are shown in Table 1:

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null model</td>
<td>1142</td>
</tr>
<tr>
<td>Main effects model</td>
<td>920</td>
</tr>
<tr>
<td>Norm only model</td>
<td>977</td>
</tr>
<tr>
<td>Interactions model</td>
<td>921</td>
</tr>
</tbody>
</table>

Table 1: Comparison of ordinal logistic regression models on AIC criterion

The AIC suggests that either the main effects model or the interactions model are reasonable; given this I opted for the simpler model, for ease of interpretation and parsimony. The score test indicated no problem with the proportional odds assumption.
INTERPRETATION OF RESULTS

There are several ways to interpret the results:

1. In terms of odds ratios
2. In terms of parameters
3. In terms of probabilities
   (a) For each individual category
   (b) For cumulative categories

The odds ratios and the confidence limits are in table 2. The interpretation of the ORs is that the odds of women doing hard drugs (as opposed to marijuana or no drugs) are 0.25 those for men doing so, holding all other variables constant. Similarly, the odds of women doing hard drugs or marijuana, as opposed to no drugs, are 0.25 those for men doing so, holding all other variables constant. Similarly, for each year of age, the odds of doing hard drugs (as opposed to none or marijuana) increase by a multiple of 1.09, as do the odds of doing hard drugs or marijuana (as opposed to no drugs). Finally, for each one-unit increase in the norm factor, each of these odds decrease by a multiple of 0.27.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Point estimate</th>
<th>95% confidence limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor 1</td>
<td>0.269</td>
<td>0.209 0.345</td>
</tr>
<tr>
<td>Age</td>
<td>1.093</td>
<td>1.004 1.191</td>
</tr>
<tr>
<td>Female</td>
<td>0.248</td>
<td>0.172 0.359</td>
</tr>
</tbody>
</table>

Table 2: Odds ratios and confidence limits: Ordinal model

The parameter estimates for this model are given in table 3. These parameter estimates are more useful when the dependent variable can be viewed as a continuous variable that has been categorized (which is hard to see, here). Briefly, the parameter estimates are estimates of the $\beta$s in the ordinal logistic regression equation. For more on interpreting these estimates, see the references, especially Long (1997).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard error</th>
<th>Wald Chi-square</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept 2: Hard 1</td>
<td>1</td>
<td>-0.3685</td>
<td>0.9375</td>
<td>0.1545</td>
<td>0.6943</td>
</tr>
<tr>
<td>Intercept 1: MJ</td>
<td>1</td>
<td>1.1565</td>
<td>0.9395</td>
<td>1.5155</td>
<td>0.2183</td>
</tr>
<tr>
<td>Factor 1</td>
<td>1</td>
<td>-1.3145</td>
<td>0.1284</td>
<td>104.85</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age</td>
<td>1</td>
<td>0.0892</td>
<td>0.0437</td>
<td>4.1685</td>
<td>0.0412</td>
</tr>
<tr>
<td>Female</td>
<td>1</td>
<td>-1.3931</td>
<td>0.1876</td>
<td>55.1246</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Table 3: Parameter estimates and standard errors: Ordinal models

Another way to interpret these results is in terms of predicted probabilities of different levels of drug use for people with different levels of the IVs. In table 4, I present the predicted probabilities of using no drugs, marijuana, or hard drugs for people at various levels of the different independent variables.

<table>
<thead>
<tr>
<th>Sex</th>
<th>Age</th>
<th>Factor</th>
<th>P(no drugs)</th>
<th>P(MJ)</th>
<th>P(Hard drugs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>18</td>
<td>-2.17</td>
<td>0.01</td>
<td>0.03</td>
<td>0.96</td>
</tr>
<tr>
<td>M</td>
<td>18</td>
<td>0.95</td>
<td>0.45</td>
<td>0.34</td>
<td>0.21</td>
</tr>
<tr>
<td>M</td>
<td>24</td>
<td>-2.48</td>
<td>0.01</td>
<td>0.02</td>
<td>0.97</td>
</tr>
<tr>
<td>M</td>
<td>24</td>
<td>0.95</td>
<td>0.34</td>
<td>0.36</td>
<td>0.30</td>
</tr>
<tr>
<td>F</td>
<td>18</td>
<td>-2.17</td>
<td>0.05</td>
<td>0.16</td>
<td>0.79</td>
</tr>
<tr>
<td>F</td>
<td>18</td>
<td>0.95</td>
<td>0.71</td>
<td>0.21</td>
<td>0.08</td>
</tr>
<tr>
<td>F</td>
<td>24</td>
<td>-2.48</td>
<td>0.02</td>
<td>0.07</td>
<td>0.90</td>
</tr>
<tr>
<td>F</td>
<td>24</td>
<td>0.95</td>
<td>0.78</td>
<td>0.23</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Table 4: Predicted probabilities of different levels of drug use: Ordinal model

The cumulative probabilities (not shown) are the probabilities of being in a given category or a lower one, here there would be three possibilities:
1. Using no drugs
2. Using marijuana or no drugs
3. Using hard drugs or marijuana or no drugs (by definition, this will always equal one).

It is also useful to know how well the predicted values match the actual values. Although SAS prints out the ‘percent concordant’ and ‘percent discordant’, it is also useful to know how the discordant results are discordant. I compared the predicted drug use (that, the drug use with the highest predicted probability) with actual drug use (see table 5); it is evident that the model predicts reasonably well for nonusers and hard drug users, but not that well for marijuana.

<table>
<thead>
<tr>
<th>Actual level</th>
<th>Predicted level</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>None</td>
<td>144</td>
</tr>
<tr>
<td>Marijuana</td>
<td>Marijuana</td>
<td>50</td>
</tr>
<tr>
<td>Hard drugs</td>
<td>Hard drugs</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>222</td>
</tr>
</tbody>
</table>

Table 5: Predicted drug use and actual drug use: Ordinal model

SAS CODE

```sas
title 'Main effects model';
proc logistic data = today desc;
    /* 'desc' is often needed to correctly order the DV */
    model drugcat = normfactor age sex;
    /* Same model syntax as dichotomous logistic, or glm */
run;
```

Predicted probabilities (either for individual levels or cumulatively) can be added easily

```sas
title 'Main effects model'; proc logistic data = today desc;
    model drugcat = normfactor age sex;
    output pred = predicted predprobs = i c;
    /* i option shows probability of individual levels (none, MJ, hard drug. c option shows cumulative probabilities (none, none or MJ, none or MJ or hard drug */
run;
```

MULTINOMIAL LOGISTIC REGRESSION

THE MODEL

In the ordinal logistic model with the proportional odds assumption, the model included j-1 different intercept estimates (where j is the number of levels of the DV) but only one estimate of the parameters associated with the IVs. If the DV is not ordered, however, this assumption makes no sense (i.e., because we could reorder the levels of the DV arbitrarily). The multinomial model generates j-1 sets of parameter estimates, comparing different levels of the DV to a base level. This makes the model considerably more complex, but also much more flexible.

The model can be written as

\[
pr(y_i = 1|\mathbf{x}_i) = \frac{1}{1 + \sum_{j=2}^{J} \exp(\mathbf{x}_i \beta_j)} \quad \text{for } m = 1
\]

\[
pr(y_i = m|\mathbf{x}_i) = \frac{\exp(\mathbf{x}_i \beta_m)}{1 + \sum_{j=2}^{J} \exp(\mathbf{x}_i \beta_j)} \quad \text{for } m > 1
\] (2)
CHECKING MODEL FIT, RESIDUALS, AND INFLUENTIAL POINTS

For the multinomial model, one way to check model fit is to use check each of the binomial models separately. An observation with a residual that is far from 0 (in either direction) is poorly fit by the model. A point with high leverage has a large influence on the parameter estimates. Several measures have been proposed for analyzing residuals, influential points, and high leverage points, but they are beyond the scope of this paper, for details, see Hosmer and Lemeshow (2000) and SAS Institute, Inc. (2004). Be sure to check ODS graphics, which are new (and experimental) in SAS 9, and, in my opinion, a great feature.

EXAMPLE

We can analyze the same data set as above; although the ordinal model is simpler, easier to interpret, and has more power since it includes the ‘ordinality’ of the DV, it is useful to compare the model with the multinomial model, both to check the assumptions and to see if interesting things happen.

Here, the AIC gives slight preference to the model with two way interactions. However, I fit the main effects model since it can be compared directly to the ordinal model above, and since the difference in AIC was very small (927.01 for the interaction model, 928.78 for the main effects model).

Output from the main effects model is similar to that from the ordinal model in that it includes ORs, parameter estimates, and predicted probabilities. However, it is different in that there are now more parameter estimates and ORs. Odds ratios from the main effects model are in table 6, parameter estimates are shown in table 7. Here I compare marijuana to no drugs, and hard drugs to no drugs. SAS allows you to specify the reference group.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard error</th>
<th>Wald Chi-square</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept MJ</td>
<td>1</td>
<td>2.3067</td>
<td>1.2364</td>
<td>3.4805</td>
<td>0.06221</td>
</tr>
<tr>
<td>Intercept Hard drugs</td>
<td>1</td>
<td>0.1027</td>
<td>1.3053</td>
<td>0.062</td>
<td>0.9373</td>
</tr>
<tr>
<td>Norm: MJ</td>
<td>1</td>
<td>-0.9586</td>
<td>0.2172</td>
<td>19.4890</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Norm: Hard drugs</td>
<td>1</td>
<td>-1.7810</td>
<td>0.2149</td>
<td>68.6934</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age: MJ</td>
<td>1</td>
<td>-0.0345</td>
<td>0.0578</td>
<td>0.3555</td>
<td>0.5510</td>
</tr>
<tr>
<td>Age: Hard drugs</td>
<td>1</td>
<td>0.1361</td>
<td>0.0610</td>
<td>4.9849</td>
<td>0.0256</td>
</tr>
<tr>
<td>Female: MJ</td>
<td>1</td>
<td>-0.9922</td>
<td>0.2397</td>
<td>17.1265</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Female: Hard drugs</td>
<td>1</td>
<td>-1.7980</td>
<td>0.2647</td>
<td>46.1403</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Table 7: Parameter estimates and standard errors: nominal models

INTERPRETATION OF RESULTS

The essential thing to remember here is that there are really two equations (one fewer than the number of categories). One formula compares people who use marijuana to those who use no drugs, the other compares those who use hard drugs to those who use no drugs. So, the odds ratios can be interpreted as saying, e.g., that the odds of a woman using hard drugs compared to no drugs are 0.166 times those of a man using hard drugs compared to no drugs. The parameter estimates are for use in the formula for the model (see equation 2).

CHOOSING BETWEEN ORDINAL AND NOMINAL MODELS

Key questions regarding the choice between the ordinal and the multinomial are whether the more complex model offers either:

1. Greater insight into the substantive area
2. Better fit or
3. Substantially different fitted values.

One substantive difference between the two models is that, in the nominal model, we see that age has a negligible effect on the ORs comparing marijuana use to non-use of drugs (OR = 0.966 per year), but a large and statistically significant effect on the ORs comparing hard drug use to non-use (OR = 1.146 per year). It should be remembered that these ORs are per year. The ages of the subjects ranged from 18 to 24; the model predicts that 24 year olds will have odds of using hard drugs (vs. no drugs) that are 1.146/(24 – 18) = 2.2652 times those of 18 year olds, but their odds of using marijuana (vs. no drugs) will be 0.8126 times those of 18 year olds.

One way to compare the fit of the two models is to compare the predicted drug use with the actual drug use for each model. The fit for the ordinal model was shown in table 5, those for the nominal model are in table 8. The nominal model actually does slightly worse than the ordinal model.

<table>
<thead>
<tr>
<th>Actual level</th>
<th>Predicted level</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>144</td>
<td>16</td>
</tr>
<tr>
<td>Marijuana</td>
<td>51</td>
<td>55</td>
</tr>
<tr>
<td>Hard drugs</td>
<td>29</td>
<td>15</td>
</tr>
<tr>
<td>Total</td>
<td>224</td>
<td>60</td>
</tr>
</tbody>
</table>

Table 8: Predicted drug use and actual drug use: Nominal model

The predicted probabilities for representative people under the multinomial model are given in table 9. They are not substantially different than those in the ordinal model (see table 4).

<table>
<thead>
<tr>
<th>Sex</th>
<th>Age</th>
<th>Factor</th>
<th>P(no drugs)</th>
<th>P(MJ)</th>
<th>P(Hard drugs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>18</td>
<td>-2.17</td>
<td>0.01</td>
<td>0.14</td>
<td>0.86</td>
</tr>
<tr>
<td>M</td>
<td>18</td>
<td>0.95</td>
<td>0.45</td>
<td>0.37</td>
<td>0.18</td>
</tr>
<tr>
<td>M</td>
<td>24</td>
<td>-2.48</td>
<td>0.00</td>
<td>0.04</td>
<td>0.96</td>
</tr>
<tr>
<td>M</td>
<td>24</td>
<td>0.95</td>
<td>0.39</td>
<td>0.26</td>
<td>0.35</td>
</tr>
<tr>
<td>F</td>
<td>18</td>
<td>-2.17</td>
<td>0.04</td>
<td>0.25</td>
<td>0.71</td>
</tr>
<tr>
<td>F</td>
<td>18</td>
<td>0.95</td>
<td>0.73</td>
<td>0.22</td>
<td>0.05</td>
</tr>
<tr>
<td>F</td>
<td>24</td>
<td>-2.48</td>
<td>0.01</td>
<td>0.09</td>
<td>0.90</td>
</tr>
<tr>
<td>F</td>
<td>24</td>
<td>0.95</td>
<td>0.72</td>
<td>0.17</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Table 9: Predicted probabilities of different levels of drug use: Multinomial model

In summary, there is no reason to prefer the more complicated model:

1. The proportional odds assumption is not violated.
2. The ordinal model fits the data slightly better than the nominal model.
3. The predicted drug use of representative people is quite similar in the two models.

SAS CODE

The main effects model can be coded with:

```sas
proc logistic data = today;
  /* ’desc’ is not needed, there is no order to the DV*/;
  model drugcat(ref = ’0: None’) = normfactor age sex/link = glogit;
  /* model statement same as for ordinal model */
  /* ’link = glogit’ fits the multinomial model - new in v9 */
  /* ’ref’ defines the reference category */
run;
```

Code for the interactions model should be clear, and is left as an exercise.
USEFUL OPTIONS IN PROC LOGISTIC

Most of these options are not specific to ordinal or multinomial logistic regression, but they can be very helpful, and may be underutilized.

- The **UNITS** statement. Occasionally, the unit of one or more of the IVs is not ideal. For example, if income were recorded in dollars per year, then the effect of a single unit change would obviously be minimal, and would lead to uninterpretable ORs. To adjust this, PROC LOGISTIC offers the UNITS statement, which allows you to adjust the units either by specifying a number (positive or negative), SD or -SD, or a number*SD. It is important to note that this affects only the estimation of the OR and confidence intervals, not the parameters.

- The **PARAM =** keyword option on the **CLASS** statement. This option specifies the parameterization for the classification variable or variables. Available choices include (but are not limited to):
  - **EFFECT** for effect coding
  - **POLY** for polynomial coding.
  - **REF** for reference cell coding.

  these also have orthogonal variants, see [SAS Institute, Inc. (2004)](http://support.sas.com/rnd/app/usage/using/ordmult.pdf) for more information.

- The **REF =** option on the **CLASS** statement specifies the reference level for **PARAM = EFFECT**, and **PARAM = REF** and their orthogonalizations. You can specify a level or use keywords **first** or **last**. A similar option on the model statement specifies the reference category for ordinal or dichotomous logistic regression.

TROUBLESHOOTING

Here I list a few things that can cause strange results, and suggest solutions.

- Improperly ordered DV or IVs. Always check the ordering of your DV when doing ordinal logistic regression (it is printed near the beginning of the output), and check the ordering of any ordinal IVs, as well. You can change the default ordering of the DV with the **DESCEND** and **ORDER =** options on the **MODEL** statement, and of the IVs with the same options on the **CLASS** statement.

- Sensible units. Improper choice of the unit on any of the IVs can lead to ORs that are hard to interpret. This can be adjusted with the **UNITS** statement discussed above.

- Huge confidence intervals. If you see that one or more IVs has a very large confidence interval, then it is a sign that something is wrong. The next step is to look at the distribution of that IV and the DV, using PROC FREQ (for categorical IVs) or PROC MEANS with a BY statement (for continuous IVs).

- Collinearity. SAS offers very good collinearity diagnostics in PROC REG. These are not available in PROC LOGISTIC, but, since collinearity is a problem among the IVs, you can use PROC REG even when the DV is not suited to. For more on collinearity see [Belsley (1991)](http://www.amazon.com/Regression-Diagnostics-Linear-Model-Statistics/dp/0471098817).

- Complete or quasicomplete separation. Complete separation occurs when an IV or set of IVs perfectly relates to the DV ([Hosmer and Lemeshow](2000)). Suppose, in our example, that all 18 year old females did no drugs. When this happens, the maximum likelihood estimates do not exist. Quasicomplete separation indicates that there is very little overlap, e.g., if nearly all 18 year old females did no drugs. In both cases, SAS issues a warning but produces output, often with very large or small ORs with huge confidence intervals. The ideal solution to this problem is to gather more data; if this is not possible, one or more IVs may need to be dropped, or levels of one or more of the IVs may need to be combined.

FURTHER READING

[Hosmer and Lemeshow](2000) is a good general book on logistic regression at a moderate mathematical level. Chapter 8 deals with the multinomial and ordinal logistic regression models. In general, they cover logistic regression in more depth than [Long](1997). Particular strengths include the section on assessing fit and using diagnostics.

[Long](1997) is a great resource for categorical and limited dependent variables. It is at a similar mathematical level to [Hosmer and Lemeshow](2000), but has less depth and more breadth; however, his coverage of the multinomial and ordinal logistic regression models is quite extensive. Chapter 5 covers ordinal logistic, and chapter 6 the multinomial case. Particular strengths include clarity and the integration of the material for various regression models.
Agresti (2002) is a classic on categorical data analysis. It is at a slightly higher mathematical level than Long (1997) or Hosmer and Lemeshow (2000), but is very clearly written considering the mathematical rigor. Particular strengths include thoroughness and inclusion of mathematical details. Readers who want a less mathematical introduction may want to consider Agresti (1996), although I have not read this book.

For details on logistic regression using the SAS system, in addition to the SAS/STAT manuals, there is Stokes et al. (2000), although it is slightly dated.

Two excellent books on regression modeling generally are Harrell (2001) and Burnham and Anderson (2002), although they don’t use SAS. In particular, I think the first chapter of Burnham and Anderson (2002) will be eye-opening for some people, and Harrell (2001) offers very good advice on general strategies for model fitting.

SUMMARY

Ordinal and multinomial logistic regression offer ways to model two important types of dependent variable, using regression methods that are likely to be familiar to many readers (and data analysts). Although there are subtleties to interpretation of the parameter estimates, the essential ideas are similar to binomial logistic regression, and, to a lesser extent, to ordinary least squares regression.

SAS offers PROC LOGISTIC to fit both these types of models; the ability to model multinomial logistic models in PROC LOGISTIC rather than GENMOD is new, and makes using this model considerably more ‘user-friendly’. ODS graphics make a powerful addition to PROC LOGISTIC, although they are not yet fully implemented for ordinal and multinomial models.

* 

REFERENCES


ACKNOWLEDGMENTS
This work was supported by NIDA grant P30DA11041, I would also like to thank Ron Fehd for providing help with \LaTeX.

CONTACT INFORMATION
Peter L. Flom
National Development and Research Institutes, Inc.
71 W. 23rd St.
8th floor
New York, NY 10010
Phone: (212) 845-4485
Fax: (917) 438-0894
flom@ndri.org
Personal webpage: www.peterflom.com
Work webpage: http://cduhr.ndri.org

SAS® and all other SAS Institute Inc., product or service names are registered trademarks or trademarks of SAS Institute Inc., in the USA and other countries. © indicates USA registration. Other brand names and product names are registered trademarks or trademarks of their respective companies.