Paper Accompaniment for E-Poster 406-2017

Using New SAS 9.4 Features for Cumulative Logit Models with Partial Proportional Odds

Paul J. Hilliard, Educational Testing Service (ETS)

ABSTRACT

Multicategory logit models extend the techniques of logistic regression to response variables with three or more categories. For ordinal response variables, a cumulative logit model assumes that the effect of an explanatory variable is identical for all modeled logits (known as the assumption of proportional odds). Past research supports the finding that as the sample size and number of predictors increase, it is unlikely that proportional odds can be assumed across all predictors. An emerging method to effectively model this relationship uses a partial proportional odds model, fit with unique parameter estimates at each level of the modeled relationship only for the predictors in which proportionality cannot be assumed. Building upon features first employed in SAS/STAT® 12.1, PROC LOGISTIC in SAS® 9.4 now extends this functionality for variable selection methods in a manner in which all equal and unequal slope parameters are available for effect selection. Previously, the statistician was required to assess predictor non-proportionality a priori through likelihood tests or subjectively through graphical diagnostics. Following a review of statistical methods and limitations of other commercially available software to model data exhibiting non-proportional odds, a public-use data set is used to examine the new functionality in PROC LOGISTIC using stepwise variable selection methods. Model diagnostics and the improvement in prediction compared to a general cumulative model are noted.

INTRODUCTION

PROC LOGISTIC offers the ability to model a variety of relationships in which the response variable is categorical. In the simplest case, a model in which the response can result in one of two values is built based upon a few predictors. Similar to PROC REG for linear regression, PROC LOGISTIC also offers a variety of model-building and diagnostic tools which enable the user to compare models in terms of precision, fit, and appropriateness to the modeled scenario. It can also be used to model multinomial categorical relationships, in which three or more discrete response levels can have a natural order. When an order exists within the response levels, a cumulative logit model is appropriate to model the relationship.

This paper will explore the basic functions of PROC LOGISTIC in modeling a publicly available dataset employing a cumulative logit link. Functionality new to SAS® 9.4 which simplifies stepwise model-building and evaluation for relationships in which predictors exhibit non-proportional odds will be examined.

The paper assumes a user understanding of the syntax and fundamental capabilities of PROC LOGISTIC.

TYPES OF LOGIT MODELS – STATISTICAL THEORY

Logistic Regression

Logistic regression is a fundamental model-building approach that aims to explain a variety of relationships between predictors and a response (dependent or Y) variable which may take one of two values (such as ‘yes’ or ‘no’ to indicate an event or non-event, respectively). While predictors themselves are often categorical (such as gender or the presence of a medical symptom) quantitative predictors also may be used in the model. The adoption of statistical software which makes model-building much simpler than in the past has allowed logistic regression to become increasingly popular in a variety of industries and research organizations, particularly the pharmaceutical and medical industries.

Categorical random variables have numerous types: dichotomous versus polytomous, ordered versus unordered categories, and so forth. For example, the distribution of a binomial random variable is based

on independent and identical Bernoulli trials. For each trial, the outcome *Y* can take on values of 0 or 1, specified by probabilities *P(Y=1)* = π of success and *P(Y=0)* = 1- π of failure (where 0 ≤ $\pi \leq 1$). A Bernoulli random variable has a mean of *E(Y)* = π. For n independent and identical Bernoulli trials, the number of successes follows a binomial distribution with mean $n*\pi$. (see Agresti, 2007).

We could specify a linear probability model as follows:

$$\pi(x) = \alpha + \beta x$$

where *x* is a linear function of the predictor variable, with *β* defined as the change in the probability per one unit of change in *x.* However, this model is inappropriate for probability values that must be restricted to a range between 0 and 1. Extreme values of a parameter estimate or value of *x* could result in predictions of $\pi(x) > 1$ or $\pi(x) < 0$. Furthermore, the prediction errors for a binary response are also binary and are not normally distributed.

To address these concerns, instead of modeling the response variable itself, a *logit link* is used to model the likelihood that an event occurs. Let us consider a model with a single predictor:

$$\log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta x$$

This is called the *logistic regression model*, where $\log\left(\frac{\pi(x)}{1-\pi(x)}\right)$ is known as the *logit link*, often symbolized by "logit($\pi(x)$)". Using this link, the prediction errors for the two outcomes (event/non-event) have a binomial distribution. The likelihood that an event occurs is expressed in terms of odds instead of probability, where $\pi(x)$ is the probability of an event and 1– $\pi(x)$ is the probability of a non-event. The natural log of this value is commonly known as the *log odds.* Whereas π is restricted to the 0–1 range, the logit can be any real number, which means this model does not have the same issues encountered when using the linear probability model.

The probability of success at a given value of a predictor variable *x* is then restricted to the [0,1] range and takes the form:

$$\pi(x) = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}$$

Similarly a, model with *k* predictor variables takes the form:

$$\log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_k x_k$$

The probability of an event given a vector of values of each of the predictors can then similarly found using:

$$\pi(x) = \frac{e^{\alpha+\beta_1 x_1+\beta_2 x_2+\ldots\beta_k x_k}}{1+e^{\alpha+\beta_1 x_1+\beta_2 x_2+\ldots\beta_k x_k}}$$

Multicategory Logit Models

Multicategory logit models are an extension of logistic regression models in which the response can be classified into three or more categories. When the categories have no natural ordering (e.g. car style preference), a *baseline-category logit model* is appropriate. To simplify the explanation, let us consider a single predictor. The model takes the following form, where the response can be classified into one of *J* categories:

$\log\left(\frac{\pi_j}{\pi_J}\right) = \alpha + \beta_j x$ where *j* = 0, 1,…*J –1. (*The reference category is typically taken to be category *J*).

The logit link for this model is then $\log\left(\frac{\pi_j}{\pi_J}\right)$.

Given that the response is classified in category *j* or category *J*, this represents the log odds that the response is *j.* Let us consider a total of 4 possible response levels from 0 to 3. *For J* = 4, the model uses:

$$\log\left(\frac{\pi_0}{\pi_3}\right), \log\left(\frac{\pi_1}{\pi_3}\right), \text{ and } \log\left(\frac{\pi_2}{\pi_3}\right).$$

Odds are then calculated for each response level using the highest category as the reference category (in the example this would be category 4).

For each response level, separate parameter estimates are provided for each predictor in the model. When several predictors are used, the number of parameters in the model can grow large and become difficult to manage. This type of model will not be explored in this paper.

When there is a natural ordering to the response levels, a *cumulative logit model* is appropriate. An example of a natural ordering would be a score on a 5-point score range or the educational attainment of adults ranging from high school through post-graduate. This logit models the cumulative probability that a response (*Y*) can be classified *at or below* a given category. In its simplest form, it assumes that the same proportionality constant applies for each parameter estimate to each cumulative level (known as a *cumulative logit model with proportional odds*). As such, it assumes that, controlling for all other predictors in the model, the change in odds between a predictor value of (x+1) versus x is the same at each response level. This difference is proportional to the distance between the explanatory variables, and the difference is the same no matter which response function is used (independent of the response level considered). This is also known as the equal slopes assumption or the parallel lines assumption (Derr 2013).

This model has the following form:

$$\text{logit}[P(Y \le j)] = \alpha_j + \beta x, \text{ where j = 0 to J - 1}$$

The cumulative logit link is the following:

$$\text{logit}[P(Y \le j)] = \log\left[\frac{P(Y \le j)}{1 - P(Y \le j)}\right] = \log\left[\frac{\pi_0 + ...\pi_j}{\pi_{j+1} + ...\pi_j}\right] \text{ where j = 0 to J - 1}$$

Note that the highest category is not in the model, as the cumulative probability of the highest category is always 1.

For the scenario with 4 possible response levels, ordered from 0 to 3 (*J*=4), the three modeled logits are

$$\text{logit}[P(Y \le 0)] = \log\left(\frac{\pi_0}{\pi_1 + \pi_2 + \pi_3}\right);\ \text{logit}[P(Y \le 1)] = \log\left(\frac{\pi_0 + \pi_1}{\pi_2 + \pi_3}\right);\ \text{logit}[P(Y \le 2)] = \log\left(\frac{\pi_0 + \pi_1 + \pi_2}{\pi_3}\right)$$

The model contrasts the lower levels of response variable Y with the higher levels of Y. From left to right, as you move from the first logit function to the second and then to the third logit function, the value of the numerator increases and the denominator decreases. Overall the cumulative logits increase in value and, since the maximum response level is 3, the model forces the cumulative probability at that level to be 1.

This model assumes that proportional odds exist (that the same proportionality constant for the parameter estimates applies to each cumulative level). Although a baseline-category logit model may be fitted to ordinal data, there are several advantages to fitting a cumulative logit model. Fitting the cumulative logit model increases the power of statistical inference because there are fewer parameters than a baseline-category logit model. A baseline-category logit model is akin to simultaneous logistic regression and contains an intercept parameter estimate and separate parameter estimates for each level of each predictor. If the predictors exhibit proportionality, then this model includes unnecessary parameter estimates (i.e., is not parsimonious).

To calculate the probability that a response can be classified in a given category, one must find the cumulative probability of the response classified at or below a given ordered category level and subtract this probability from the cumulative probability of the next lowest ordered category level. For example, to calculate the probability the observation with predictor value *x* can be classified in Category 1, perform the following:

$$P(Y = 2) = P(Y \le 2) - P(Y \le 1) = \frac{e^{\alpha_2 + \beta x}}{1 + e^{\alpha_2 + \beta x}} - \frac{e^{\alpha_1 + \beta x}}{1 + e^{\alpha_1 + \beta x}}$$

In PROC LOGISTIC, specifying LINK=CLOGIT produces a model which utilizes a cumulative logit link.

Using the default settings, the procedure conducts a Score Test for the Proportional Odds Assumption (the test will be discussed further in detail in the next section). If it rejects the null hypothesis that proportional odds may be assumed, the user has two additional modeling techniques utilizing the cumulative logit link.

A *general cumulative logit model* generates separate parameter estimates (effects) for each predictor across all $J - 1$ response levels (see Agresti, 2007). This model replaces β in the proportional odds model with β_j and uses the same cumulative logit link:

$$\text{logit}[P(Y \le j)] = \alpha_j + \beta_j x, \text{ where j = 0 to J - 1}$$

The model allows for probability distribution curves that climb or fall at different rates for different response levels. This raises the possibility of these curves crossing each other at certain predictor values. This is inappropriate, because this violates the order that cumulative probabilities must have:

$$P(Y \le X) \le P(Y \le X + 1) \text{ for all values of the predictors}$$

This result could lead to negative predicted probabilities at individual response levels. Therefore, such a model can only fit adequately over a narrow range of predictor values. Using the proportional odds form of model ensures that the cumulative probabilities have the proper order for all predictor values. If a model of this form is desirable, PROC LOGISTIC will generate a general cumulative logit model if the option UNEQUALSLOPES is specified in the options for the MODEL statement.

A *partial proportional odds model* using the cumulative logit link is the second type of model that may be considered if the proportionality of the predictors at all response levels is unlikely. This model is a hybrid of the proportional odds model and the general model and uses the same cumulative logit link. Separate parameter estimates for each predictor at each response level are provided for only the predictors that fail to satisfy the "parallel lines assumption" or exhibit non-proportionality. This decision is made by chi-square tests which determine if, when the predictor is included in the model, fit is more improved when proportionality is assumed. If greater model fit occurs when proportionality cannot be assumed, separate parameter estimates for the predictor at each response level are provided.

Consider a scenario with two predictors: one predictor *x* that exhibits proportionality and one predictor *z* that exhibits non-proportionality. The model is as follows:

$$\text{logit}[P(Y \le j)] = \alpha_j + \beta x + \gamma_j z, \text{ where j = 0 to J - 1}$$

The model provides separate parameter estimates only for predictor *z*, which depend on the response level. The cumulative probabilities can be calculated similarly as for the other cumulative logit models discussed previously (see Peterson & Harrell, 1990 for additional theory and explanation).

The body of this paper concerns the new features in SAS 9.4® for PROC LOGISTIC that greatly simplify the techniques needed to generate a partial proportional odds model for data that exhibits non-proportionality amongst predictors. A justification for its use and its comparison to a proportional odds model in terms of predictive accuracy will be noted.

THE NEED FOR A PARTIAL PROPORTIONAL ODDS MODEL

Apart from the simplest model with just a few categorical predictors, modeled relationships are often complex and based upon a variety of both categorical and continuous predictors. When using a proportional odds model, the test of the proportional odds assumption frequently rejects the null hypothesis that proportionality can be assumed.

The test can best be described as anti-conservative and nearly always results in rejection of the proportional odds assumption, particularly when the number of explanatory variables is large, the sample size is large, or if there are continuous predictors in the model (Allison, 1999; Brant, 1990; O'Connell 2006). For small samples, Peterson and Harrell (1990) noted that the *p*-value for the test could be artificially small, which would result in unnecessarily rejecting the proportional odds assumption.

Since the test is imperfect and rejection is frequent, it is often desirable to interpret the degree of predictor proportionality graphically. This can be accomplished by computing "empirical logits" based upon the cross-tabulations of the true response with each predictor in the full model. If all cross-tabulated cell counts are larger than about five, sample sizes should be adequate for this graphical analysis (Stokes, Davis & Koch, 2012). If graphical evidence of non-proportionality is lacking, a user could then decide that the assumption is adequately met.

Upon rejection of the test for the proportional odds assumption and a review of any graphical displays of the degree of parallelism amongst response levels for the predictors, a user is left with the option of fitting either a general cumulative logit model or a partial proportional odds model. For reasons noted in the previous section, a general cumulative logit model is limited by its design. The partial proportional odds model offers the more powerful and more appropriate model. It often results in more predictive estimates.

PROC LOGISTIC - FOR PARTIAL PROPORTIONAL ODDS MODELS

Beginning with SAS/STAT 12.1, PROC LOGISTIC offered the ability to generate a partial proportional odds model. Using the UNEQUALSLOPES option for the MODEL statement, a user could list any predictors that do not satisfy the proportional odds assumption. This option produces a different parameter estimate for every response level (the predictor has unequal slopes across the response functions). All other predictors have a single parameter estimate common to all response functions.

If a user has many predictors that lack proportionality, multiple models can be created by selecting combinations of predictors with and without equal slopes. This is done by varying the list of predictors in the UNEQUALSLOPES list. Statistical methods (such as log-likelihood tests and comparisons of the Akaike Information Criterion (AIC)) or graphical methods such as assessment plots could be used to determine model fit (Agresti, 2007). Derr (2013) offers several dataset examples in which the SAS/STAT 12.1 functionality for partial proportional models is used.

While this additional ability is of obvious importance, the user is often left with the task of assessing predictor non-proportionality *a priori* through likelihood tests or subjective graphical diagnostics. When relationships include several predictors, an exhaustive investigation of model fit and proportionality could be time-consuming.

Beginning in SAS 9.4 with the TS1M2 revision, a user can include the SELECTION= option to have PROC LOGISTIC simultaneously determine which predictors a) are significant in the overall model and b) exhibit non-proportional odds. Using a given selection method, SAS then chooses a final model based on user specifications (such as alpha-level significance). If both the EQUALSLOPES and UNEQUALSLOPES options are specified, the selection method will select a final model from a set of candidate effects that includes all equal and unequal slope parameters. The predictor with unequal slope parameters will then be included in the model when both non-proportionality exists and the predictor is significant to the modeled relationship.

Another option is a restriction that requires every model during selection to include the equal slope parameters. The INCLUDE=EQUALSLOPES option then tests the unequal slope parameters for a significant effect and retains them in the model if significant at a user-specified alpha level. This option allows the model to contain, for a given predictor, both the single equal and multiple unequal slope parameters. The multiple unequal slope parameters then serve as individual adjustments to account for the non-proportionality. While this method may be useful in the interpretation of predictor effects, it unnecessarily adds complexity (and parameters) to the model optimized using the SELECTION= method. The effect of adding unnecessary parameters can be judged by comparing the AIC for the model with the AIC for a simpler model.

Other statistical programming tools currently do not offer the power and ease of use for generating partial proportional odds models using model selection methods. The Ordinal package in R (Christensen, 2015) allows for partial proportional odds models and the ability to compare models which are subsets of one another. These model-building methods are related, in a sense, to stepwise regression methods, but are much less user-friendly and require the generation of multiple models which must be pre-specified by the user. The *GoLogit2* program in Stata (Williams, 2006) has an autofit option that offers a kind of backward selection technique for partial proportional odds models. However, stepwise or forward selection methods are unavailable. Finally, a personal review of SPSS® suggests that it does not offer the ability to generate partial proportional odds models nor does it offer the forward, backward, or stepwise model building techniques for ordinal regression (only for binary logistic regression).

DATA EXAMPLE – USING STEPWISE REGRESSION TO BUILD A CUMULATIVE LOGIT MODEL WITH PARTIAL PROPORTIONAL ODDS

To illustrate the newest capabilities in PROC LOGISTIC to build a cumulative logit model that employs partial proportional odds with stepwise selection, a publicly available dataset, titled "Adolescent Placement Study" was used. The data and in-depth variable descriptions are available at http://www.umass.edu/statdata/statdata/stat-logistic.html. The dataset contains 11 predictors that are related to the placement location of 508 adolescents in the study. Several of the predictors (such as gender, history of violence, emotional disturbance, and elopement risk) are dichotomous. Patient age and length of hospitalization are quantitative predictors while variables such as neuropsychiatric disturbance contain four ordered categories. The model would be designed to predict recommended placement status for adolescents with psychiatric problems. The dependent variable (PLACE) can take one of four coded values: Outpatient (coded as 0), Day Treatment (1), Intermediate Residential (2) and Residential (3). Because the values represent increasingly restrictive placements, it is logical to utilize the coded order. A higher predicted placement would signify that a given patient should be placed in a more restrictive environment for treatment.

A sample of 408 observations were used for model building, leaving 100 aside to serve as an unbiased sample that will be used to evaluate the prediction power of the models. PROC SURVEYSELECT with SEED=535113001 and Simple Random Sampling (SRS) was used to create the evaluation set.

Option 1: A Proportional Odds Model

For ordinal logistic regression, the default method of model building uses a proportional odds model. Generally, a user would assume proportionality of the predictors unless statistical or graphical evidence suggests otherwise. As stated in the statistical theory section of this paper, the proportional odds model contains fewer parameters than more generalized models and is simpler and easier to interpret. It also restricts the response functions so that they cannot intersect one another, avoiding the possibility of negative individual predicted probabilities at certain response levels.

The following code uses Stepwise selection with a minimal alpha level ($\alpha = 0.10$ as specified in the SLE= and SLS= options) to enter or exit the model. The cumulative logit model is specified (LINK=CLOGIT). The dataset “MB” is comprised of 408 of the 508 observations in the dataset. Dataset “XV” contains 100 observations and will be used for cross-validation purposes for the model.

```
proc logistic data= MB outest=SASGlobal_est_4PROPODDS;
      model PLACE = AGE RACE GENDER NEURO EMOT DANGER ELOPE LOS BEHAV CUSTD VIOL /
            LINK=CLOGIT SELECTION=STEPWISE SLE=0.10 SLS=0.10;
      score data= XV out=SasGlob.Pred_XV_4PROPODDS;
      ods output association=SasGlob.SASGlobal_assoc_4PROPODDS;
run;
```

A “stepwise” selection method begins with an intercept-only model. The method then evaluates each available predictor to determine which predictor most improves model fit (defined by the greatest change in log-likelihood or similarly the predictor with the highest Score Test Chi-Square test statistic). After each step, the method re-examines the significance of all predictors in the model and removes any predictor which, after controlling for other predictors in the model, is no longer significant at the alpha level that is user-specified (a Wald Chi-Square test statistic is employed). The iterative process continues until no more predictors meet the significance level for entry. For a more in-depth explanation of stepwise regression, see Chapter 5 of Agresti (2007).

The options specified in PROC LOGISTIC include the “OUTEST” dataset which provides parameter estimates and a covariance matrix. The ODS OUTPUT dataset “association” provides Association Statistics such as percent concordant / discordant and the *c* statistic for model classification accuracy. A very useful option is specified by the SCORE statement. A user specifies the dataset which can be “scored” by the resulting model. The “out” option provides a separate dataset with these scores right-appended to the original dataset. The scores include the estimated probabilities that the response would be found in a given category (response level). The SCORE statement also assigns a response level score to the observation with the most likely predicted category. This value can then be used to evaluate the classification accuracy of the model. The SCORE dataset (e.g., XV) must include observations which have the same list of identically-named predictors as the dataset used to build the model (e.g., MB).

For this dataset, stepwise regression resulted in seven of the eleven predictors entering the model as significant predictors of psychological placement:

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	0	1	3.8320	0.9915	14.9364	0.0001
Intercept	1	1	5.3762	1.0107	28.2949	<.0001
Intercept	2	1	7.3393	1.0444	49.3814	<.0001
AGE		1	-0.1142	0.0604	3.5777	0.0586
RACE		1	-0.6136	0.2025	9.1809	0.0024
NEURO		1	-0.3507	0.1028	11.6421	0.0006
DANGER		1	0.2543	0.1215	4.3798	0.0364
LOS		1	-0.0532	0.00686	60.0603	<.0001
BEHAV		1	-0.3195	0.0627	25.9971	<.0001
CUSTD		1	-2.3526	0.2342	100.9056	<.0001

With the exception of DANGER, all parameter estimates are negative. This indicates that, controlling for all other predictors in the model, as the value of a given predictor <u>increases</u>, the probability that the observation is more likely to be classified in a lower category level decreases (a less restrictive treatment setting). For the CUSTD predictor, if CUSTD=1 then the adolescent was in state custody. When predicting placement using the model, the value obtained when CUSTD=1 (in this case -2.3526, the product of the parameter estimate and 1) then indicates an increased probability that the adolescent would be predicted to be placed in a more restrictive setting.

These parameter estimates can also be used to compute odds ratios. The odds ratio for NEURO (Neuropsychiatric Disturbance) is $e^{-0.3507.}$ = 0.7042. Controlling for all other predictors in the model, at any fixed placement level, a one unit increase in the value of NEURO results in the odds of being categorized at a lower placement level as 0.7042 times the odds of being categorized at a higher level. The odds that a response is categorized at a higher placement level are greater.

For a cumulative logit link, the default estimation method assumes proportionality. PROC LOGISTIC automatically generates a Score Test to assess the assumption. For the modeled relationship, the following results were obtained after the final iteration of stepwise selection:

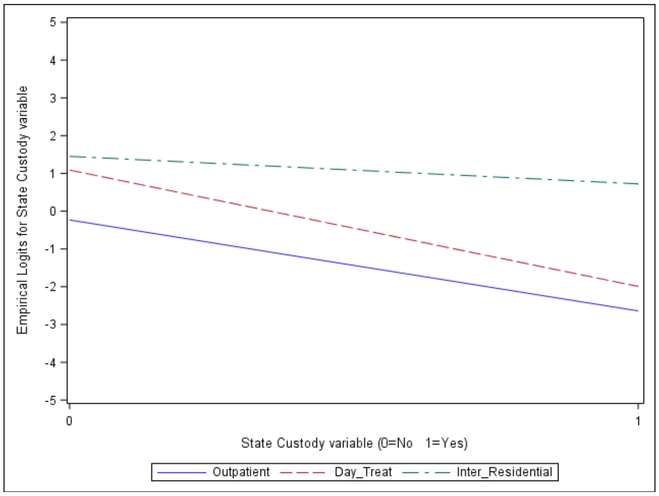
Score Test for the Proportional Odds Assumption		
Chi-Square	DF	Pr > ChiSq
136.0528	14	<.0001

The Test Statistic is statistically significant. This suggests that the proportionality of the predictors in the modeled relationship cannot be assumed. As stated previously, the test is anti-conservative and the null hypothesis is frequently rejected for large datasets with several predictors. It is often useful to assess the degree of non-proportionality graphically. This assumption can be assessed visually for a given predictor by plotting the predictor versus the "empirical logits" (see SAS Note 37944). These logits are based upon the observed counts of the dependent variable and can be calculated similarly to the cumulative logits.

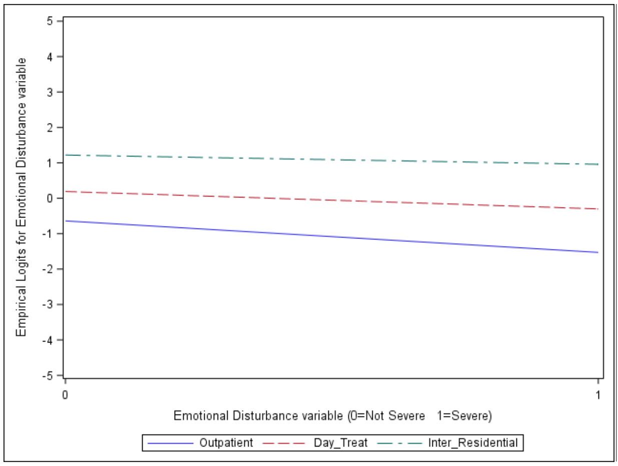
For response level *i* when there are *J* levels from 0 to J, the empirical logit can be calculated by:

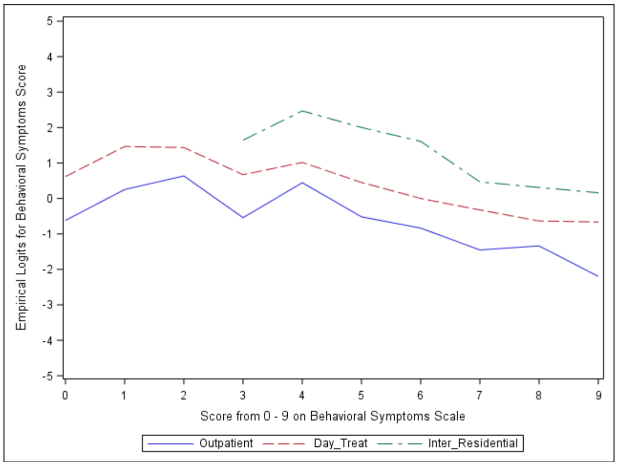
$$\text{EmpLogit}_i = \log\left(\frac{y_0 + y_1 + \ldots + y_i}{y_{i+1} + y_{i+2} + \ldots + y_J}\right)$$

For the Adolescent Placement Study, the following plot (using PROC SGPLOT) displays the empirical logits for the State Custody variable (CUSTD). While the empirical logits for Placement levels 0 and 2 (Outpatient and Inter-Residential) are fairly parallel to one another, the empirical logit for Placement level 1 (Day Treatment) is not parallel to the others. Taken in sum, the plot indicates a lack of proportionality.



Empirical logits were also plotted for the EMOT and BEHAV variables. A lack of proportionality does not appear to exist for EMOT and perhaps a moderate degree of non-proportionality exists for BEHAV.





When several of these predictors are present in the model, with some lacking proportionality, the resulting model fails to satisfy the proportional odds assumption. See Derr (2013) for macros that generate empirical logits for all predictors in a given model and other visual methods for assessing the Proportional Odds Assumption.

Option 2: A General Cumulative Logit Model

When proportionality cannot be assumed, the user has the option of creating a "general" cumulative logit model, which generates separate parameter estimates (effects) for each predictor across all *J* – 1 response levels. PROC LOGISTIC will generate a general cumulative logit model if the option UNEQUALSLOPES is specified in the options for the MODEL statement.

The following code generates this type of model.

```
proc logistic data=SasGlob.SASGlobal_MB outest=SASGlobal_est_4GEN;
      model PLACE = AGE RACE GENDER NEURO EMOT DANGER ELOPE LOS BEHAV CUSTD VIOL /
            LINK=CLOGIT SELECTION=STEPWISE SLE=0.10 SLS=0.10
      unequalslopes maxiter=100;
      score data=SasGlob.SASGlobal_XV out=SasGlob.Pred_XV_4GEN;
      ods output association=SasGlob.SASGlobal_assoc_4GEN;
run;
quit;
```

A model with 7 predictors and 21 parameter estimates is generated, with 3 parameter estimates provided for the *J* – 1 response levels (*J* = 4 levels in the data). However, when UNEQUALSLOPES is specified, the following warning occurs in the LOG, which is a frequent occurrence when fitting a general cumulative logit model.

```
WARNING: Negative individual predicted probabilities were identified in the final model fit
         during Step 6.  You may want to modify your UNEQUALSLOPES specification.
WARNING: The LOGISTIC procedure continues in spite of the above warning. Results shown are based
         on the last maximum likelihood iteration. Validity of the model fit is questionable.
```

If for a given observation, the probability that the predicted response category is less than 2 is lower than the probability that the predicted category is less than 1, then the probability that the category= 2 would be negative. When the SCORE statement estimates predicted response levels for observations, missing values are assigned to any observations where negative individual predicted probabilities occur.

As such, a general cumulative logit model is inappropriate to model the relationship, yet proportional odds cannot be safely assumed.

Option 3: A Partial Proportional Odds Model

A Partial Proportional Odds model using the cumulative logit link can be specified in a manner similar to a general model. The only difference is the inclusion of the "EQUALSLOPES" option along with the "UNEQUALSLOPES" option in the MODEL statement. The following code was used to build the first partial proportional odds model for the Adolescent Placement Study. It once again uses the Stepwise method although a user may also opt to use Forward or Backward selection methods with these same options:

```
proc logistic data=SasGlob.SASGlobal_MB outest=SASGlobal_est_4PPODDS;
      model PLACE = AGE RACE GENDER NEURO EMOT DANGER ELOPE LOS BEHAV CUSTD VIOL /
            LINK=CLOGIT SELECTION=STEPWISE SLE=0.10 SLS=0.10
      equalslopes unequalslopes details maxiter=10000;
      score data=SasGlob.SASGlobal_XV out=SasGlob.Pred_XV_4PPODDS;
      ods output association=SasGlob.SASGlobal_assoc_4PPODDS;
run;
quit;
```

This code considers all predictors in both their equal and non-equal slope forms across the response levels. These predictors are then simultaneously evaluated stepwise to build the model. For example, for CUSTD, the proportional odds parameter estimate (with one degree of freedom) is evaluated along with the general model parameter estimate (with *J -1* degrees of freedom for *J* levels) and all other predictors. The predictor with the most significant Score Test Chi-Square test statistic (of all available predictors) is added to the model.

If the general model unequal slope predictor enters the model, the equal slope predictor is no longer eligible for entry. Conversely, if the equal slope predictor enters the model, the general model unequal slope predictor retains eligibility for entry during a later step (though with one less degree of freedom because the predictor is present in the model). This is due to the stepwise nature of the model building process. As the model becomes more complex, it is possible that when controlling for other predictors in

the model, either the predictor is no longer significant and exits the model at that iteration, or the unequal slope predictor gains significance during a later iteration.

The summary of the Stepwise selection for the model appears below. During the first step, the CUSTD predictor with unequal slopes entered the model (denoted as U_CUSTD). The Score Test Chi-Square test statistic is highly significant. From step 2 through step 6, the next five predictors that entered the model had common slopes across response levels. This suggests that the proportionality assumption may likely be assumed. During step 7, the U_EMOT predictor entered the model. Finally, the quantitative predictor AGE (with equal or common slope) entered the model. The Stepwise selection re-evaluates all predictors in the model following each selection. No predictors were removed in this example.

Summary of Stepwise Selection

	Effect						
Step	*Entered*	*Removed*	*DF*	*Number In*	*Score Chi-Square*	*Wald Chi-Square*	*Pr > ChiSq*
1	U_CUSTD		3	1	183.1529		<.0001
2	LOS		1	2	121.9317		<.0001
3	BEHAV		1	3	20.9883		<.0001
4	NEURO		1	4	10.9480		0.0009
5	RACE		1	5	7.7161		0.0055
6	DANGER		1	6	5.4120		0.0200
7	U_EMOT		3	7	8.0716		0.0446
8	AGE		1	8	3.1047		0.0781

Another option is a restriction that requires every model during selection to include the equal slope parameters. The INCLUDE=EQUALSLOPES option then tests the unequal slope parameters for a significant effect and retains them in the model if significant at a user-specified alpha level. This option allows the model to contain, for a given predictor, both the single equal and multiple unequal slope parameters. The multiple unequal slope parameters then serve as individual adjustments to account for the non-proportionality. The following code forces all eleven predictors with equal slopes into the model prior to stepwise selection of the unequal slopes:

```
proc logistic data=SasGlob.SASGlobal_MB outest=SASGlobal_est_4PPODDS_INC;
      model PLACE = AGE RACE GENDER NEURO EMOT DANGER ELOPE LOS BEHAV CUSTD VIOL /
            LINK=CLOGIT SELECTION=STEPWISE SLE=0.10 SLS=0.10
      equalslopes unequalslopes include=equalslopes details maxiter=10000;
      score data=SasGlob.SASGlobal_XV out=SasGlob.Pred_XV_4PPODDS_INC;
      ods output association=SasGlob.SASGlobal_assoc_4PPODDS_INC;
run;
```

The summary appears on the following page. Three additional unequal slopes predictors entered the model. As noted in the previous model, since the equal slope parameter estimates for these predictors are already in the model, they contain one fewer degree of freedom than the *J - 1* response levels. Contrasting this model to the previous model, unequal slope parameter estimates for ELOPE were added to this model while both models contained the U_CUSTD and U_EMOT parameter estimates.

Summary of Stepwise Selection							
	Effect						
Step	Entered	Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	U_CUSTD		2	12	108.2995		<.0001
2	U_EMOT		2	13	6.0004		0.0498
3	U_ELOPE		2	14	4.9389		0.0846

While this method may be useful in determining which of all the predictors exhibit non-proportionality, there are at least two potential drawbacks. Firstly, forcing all equal slope predictors into the model may unnecessarily add complexity (and parameters) to the model whereas a single optimized parameter may be found using the stepwise selection method that starts with the intercept only. The effect of adding unnecessary parameters can be judged by comparing the AIC for the model with this value for a simpler model. Secondly, the resulting model may have non-significant predictors. Of the eleven predictors in this example, four did not meet the $\alpha=0.10$ level specified for entry into the model through stepwise selection. Forcing these non-significant parameters into our model has increased standard errors and decreased the predictive ability of the model.

A COMPARISON OF MODELS

Often a user will develop many models that require an analytical comparison to determine which is most optimal. We have explored three models in this paper which provide stable prediction categories for all observations. These were:

- Proportional Odds Model (labeled **PropOdds**),
- Partial Proportional Odds Model (**PPODDS≠**) and
- Partial Proportional Odds Model with all equal slope predictors forced into the model (**PPODDSwith=**).

Since the general cumulative logit model failed to assign predicted response categories to some observations, it cannot be considered as a viable model for the relationship.

One method of comparing two models is a log-likelihood test. When one model contains a subset of predictors from a more complex model, the difference in log-likelihood forms a chi-square test statistic that can be tested against a chi-square distribution with degrees of freedom equal to the difference in the number of parameters between the full model and reduced model. While this method is often sufficient, in some instances in iterative model building one model is not necessarily a subset of another.

For purposes of these models, the comparison of the AIC values is useful, with lower values indicating a preferable model. The AIC uses the log-likelihood values from each model, but adds a “penalty” for including unnecessary predictors to the model that add little to the overall relationship. A summary appears below:

Model	# Parameters Estimated	AIC	-2 LL (Log-Likelihood)
PropOdds	10 (7 + 3 intercept)	853	833
PPODDS≠	**15 (12 + 3 intercept)**	**780**	750
PPODDSwith=	20 (17 + 3 intercept)	785	**745**

The AIC value is lowest for the PPODDS≠ model. Though there is no statistical test to determine the significance of the difference in AIC values, the AIC value is much lower than for the PropOdds model (as is the log-likelihood). The model contains five additional parameter estimates, but their inclusion

improved model fit. The PPODDSwith= model included five additional parameter estimates compared to the PPODDS≠ model, yet the -2 LL decreased very slightly and the AIC was higher. This suggests this model contains unnecessary parameters. However, both models employing the partial proportional odds option represent an improvement over the proportional odds model.

Since the Score Test for the proportional odds assumption was statistically significant, the PropOdds model is not particularly attractive. While SAS provides parameter estimates and fits a model, the estimated probabilities across response categories (response functions) are not proportional to the values of some of the predictors.

Finally, we may evaluate the ability of the models to predict placement level for additional observations (adolescents) in the study by examining placement accuracy. Recall that of the 508 observations in the dataset, 100 responses were randomly withheld to serve as an evaluation set, XV, which was used as the scoring dataset in each of the sets of example code.

The following table displays the 2 x 2 classification of the true placement level with the predicted placement level for the **PropOdds** model (obtained from the predicted placements written out to the SCORE dataset):

Frequency — *Table of F_PLACE by I_PLACE*

F_PLACE(TRUE PLACEMENT)	*I_PLACE(PREDICTED PLACEMENT)* 0	1	2	3	Total
0	21	6	3	1	31
1	8	3	5	1	17
2	3	3	16	9	31
3	4	2	4	11	21
Total	36	14	28	22	100

FINDINGS:

- The correct placement was predicted 51% of the time (exact agreement)
- The placement levels were adjacent to one another (+- 1 level) an additional 35% of the time
- 14% of the placements differed by at least 2 levels (poor prediction)
- Simple kappa statistic: 0.3302
- Weighted kappa: 0.5668

Both the simple kappa and weighted kappa statistics were calculated using PROC FREQ. While kappa is a fairly standard metric, weighted kappa, a variation of a metric designed by Cohen (see Agresti, 2007), is an adjustment with respect to expected agreements that is based on observed marginal frequencies, as some chance agreement always exists between two independent measures. Quadratic weights are assigned to each contingency table cell by squaring the difference in the actual response level and predicted response level. Values typically range between 0 and 1, with higher values associated with stronger association (agreement). Values of 0.70 or higher are generally characteristic of highly predictive models while values between 0.60 and 0.69 are considered marginally acceptable and characteristic of models with some predictive value. A value of 0 would indicate that agreement between the two measures is no better than agreement by chance alone.

The weighted kappa value of 0.5668 for this model indicates it is of limited predictive value. The model correctly placed the adolescent just over half the time and was very poor at predicting a placement level of 1 (Day Treatment), only predicting that 3 of the 17 observations with a true placement level of 1 would be classified as such. The predicted placements were fairly symmetric in nature to the true placements. The model did not tend to be biased upward or downward in predicting placement level.

Next, the following table displays the 2 x 2 classification of the true placement level with the predicted placement level for the **PPODDS≠** model:

Frequency	Table of F_PLACE by I_PLACE				
	I_PLACE(PREDICTED PLACEMENT)				
F_PLACE(TRUE PLACEMENT)	0	1	2	3	Total
0	23	5	3	0	31
1	6	7	4	0	17
2	3	2	23	3	31
3	4	3	3	11	21
Total	36	17	33	14	100

FINDINGS:

- The correct placement was predicted 64% of the time
- The placement levels were adjacent to one another (+- 1 level) an additional 23% of the time
- 13% of the placements differed by at least 2 levels (poor prediction)
- Kappa statistic: 0.5054
- Weighted Kappa: 0.6200

The weighted kappa value of 0.6200 for this model indicates it has some predictive value. It correctly predicts placement level in nearly 2 out of 3 observations (13 more adolescents out of 100 total are now correctly placed using the model). Prediction remains somewhat poor at Placement Level 1 while there continues to be a number of observations that are poorly predicted by both models. These could be adolescents whose behavioral characteristics are not well-captured by the available predictors in the modeled relationship. Nonetheless, there is a meaningful increase in prediction accuracy compared to the PropOdds model.

Finally, the following table displays the 2 x 2 classification of the true placement level with the predicted placement level for the **PPODDSwith=** model:

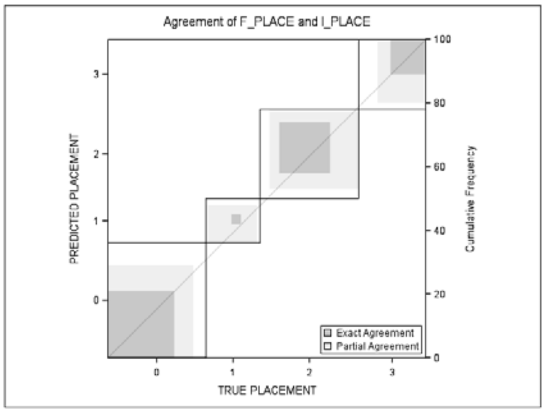
Frequency	Table of F_PLACE by I_PLACE				
	I_PLACE(PREDICTED PLACEMENT)				
F_PLACE(TRUE PLACEMENT)	0	1	2	3	Total
0	24	3	3	1	31
1	7	6	3	1	17
2	4	1	21	5	31
3	4	3	4	10	21
Total	39	13	31	17	100

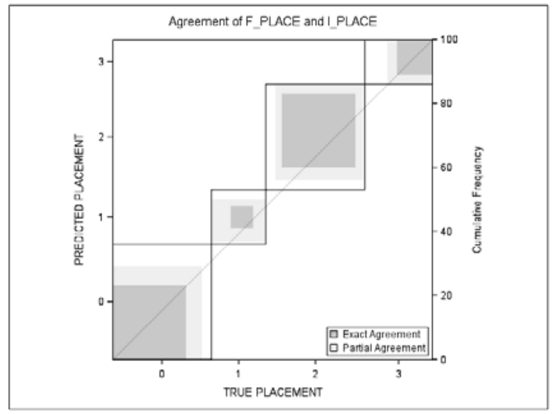
FINDINGS:

- The correct placement was predicted 61% of the time
- The placement levels were adjacent to one another (+- 1 level) an additional 23% of the time
- 16% of the placements differed by at least 2 levels (poor prediction)
- Kappa statistic: 0.4622
- Weighted Kappa: 0.5728

Compared to the previous model employing partial proportional odds, the kappa and weighted kappas and the correct placement rates are lower. The rate of poor predictions (where predicted placement differed by 2+ ordered levels) was actually the highest of all three models. The decrease in predictive value of this model could be due to its inclusion of additional non-significant predictors. If several of these predictors are related to one another, the model could suffer from multicollinearity, which has the effect of increasing the standard errors of the estimates. Removing any redundant predictors could potentially reduce standard errors and increase predictive ability (Agresti, 2007). However, as specified, of the two models using partial proportional odds, the **PPODDS≠** model that does not force all equal slope predictors into the model has superior predictive ability.

Finally, when WT=FC is specified in the options for PROC FREQ, a visual interpretation of the 2 x 2 classification tables is produced. The following tables compare the true and predicted placement levels for the **PropOdds** and the **PPODDS≠** models. The darker grey within each box indicates exact agreement while the lighter shade of grey indicates partial or adjacent agreement. At all score levels, the **PPODDS≠** model has an equal or higher rate of exact agreement. This improvement is particularly notable for Placement Level=2 (Intermediate Residential).





The increase in correct classification (agreement) is of crucial importance if a model is ever used to make or support high-stakes decisions. While it may be inappropriate in the context of psychological placement, it could assist in other situations. For example, along with a single human rater, models or “machine” scorers often serve as second raters of writing content. When the model score agrees with the human score within a given threshold (often 1 point), the two scores are combined to arrive at an assigned score. When the model disagrees with the human rater, a second human rater is tasked with “adjudicating” the score. An accurate scoring model would greatly reduce the expense and need to employ additional human raters.

For the context of the data used in this paper, it may also be true that if two psychiatrists were independently placing the same adolescents, they may agree on placement level no more often than the model agrees with one of the psychiatrists. For complex relationships where different psychiatrists weigh different adolescent characteristics in a different manner, the placement decisions made by these individuals may be no more reliable than the placements predicted by the model.

CONCLUSION

This paper has illustrated the additional model building options deployed in SAS® 9.4 to create partial proportional odds models for relationships in which the assumptions necessary for the proportional odds model are not satisfied. For relationships beyond those that are simple with few predictors and observations, the proportional odds assumption is often violated. The general cumulative logit model is often complex and only fits adequately over a narrow range of predictor values. The partial proportional odds model therefore offers users an effective solution that addresses the proportionality assumption and offers an improvement in model fit over general models.

SAS offers a model building approach unique to the industry. Rather than assessing predictor non-proportionality *a priori* through likelihood tests or graphical diagnostics, PROC LOGISTIC now is able to assess all predictors simultaneously in both their equal and non-equal slope forms for model inclusion. Only the predictors which are significant to the relationship are included in the model, and only in the form that maximizes model fit. Users also have many of the same options in PROC LOGISTIC that are available for other logit models in order to customize the approach best suited to their data. In all, SAS offers a robust, straightforward, and user-friendly solution that can be used to model these complex relationships.

REFERENCES

Agresti, A. (2007). *An Introduction to Categorical Data Analysis*. 2nd ed., Hoboken, NJ: John Wiley & Sons.

Allison, P. D. (1999). “Logistic Regression Using the SAS System: Theory and Application,” Cary, NC.: SAS Institute.

Brant, R. (1990). Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression,” Biometrics, 46, 1171-1178.

Christensen, H.R.B. (2015). “Regression Models for Ordinal Data,” https://cran.r-project.org/web/packages/ordinal/ordinal.pdf”

Derr, B. (2013). “Ordinal Response Modeling with the LOGISTIC Procedure,” http://support.sas.com/resources/papers/proceedings13/446-2013.pdf.

Harrell, F. E. (2001), *Regression Modeling Strategies*, New York: Springer-Verlag.

Koch, G. G., Amara, I. A., and Singer, J. M. (1985). “A Two-Stage Procedure for the Analysis of Ordinal Categorical Data,” in P. K. Sen, ed., Biostatistics: Statistics in Biomedical, Public Health, and Environmental Sciences, Amsterdam: Elsevier Science.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed., London: Chapman & Hall.

O'Connell, A. (2006). *Logistic Regression Models for Ordinal Response Variables*. Thousand Oaks: Sage Publications.

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., and Feinstein, A. R. (1996). “A Simulation Study of the Number of Events per Variable in Logistic Regression Analysis,” Journal of Clinical Epidemiology, 49, 1373–1379.

Peterson, B. L. and Harrell, F. E. (1990). “Partial Proportional Odds Models for Ordinal Response Variables,” Journal of the Royal Statistical Society, Series B, 39, 205–217.

SAS Institute Inc. (2012). “SAS Note 37944: Plots to Assess the Proportional Odds Assumption in an Ordinal Logistic Model,” http://support.sas.com/kb/37944.

Stokes, M. E., Davis, C. S., and Koch, G. G. (2012). *Categorical Data Analysis Using SAS*, 3rd ed., Cary, NC: SAS Institute Inc.

Williams, R. A. (2006). “Generalized Ordered Logit/Partial Proportional Odds Models for Ordinal Dependent Variables,” The Stata Journal, 1, 58-82.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Paul J. Hilliard
Educational Testing Service
660 Rosedale Road (MS 20-T)
Princeton, NJ 08541

phillard@ets.org