

Paper 446-2013

## Ordinal Response Modeling with the LOGISTIC Procedure

Bob Derr, SAS Institute Inc.

### ABSTRACT

Logistic regression is most often used for modeling simple binary response data. Two modifications extend it to ordinal responses that have more than two levels: using multiple response functions to model the ordered behavior, and considering whether covariates have common slopes across response functions.

This paper describes how you can use the LOGISTIC procedure to model ordinal responses. Before SAS/STAT<sup>®</sup> 12.1, you could use cumulative logit response functions with proportional odds. In SAS/STAT 12.1, you can fit partial proportional odds models to ordinal responses. This paper also discusses methods of determining which covariates have proportional odds. The reader is assumed to be familiar with using PROC LOGISTIC for binary logistic regression.

### INTRODUCTION

Logistic regression modeling has applications in many areas, including clinical studies, epidemiology, data mining, social sciences, marketing, and engineering. It has proved to be reliable for both prospective analyses (such as designed experiments or clinical trials) and retrospective analyses (such as found data or case-control studies).

If your response variable can take only two values (the “event” and the “non-event”), then the conditions for linear regression are not met; in particular, your errors are binary and not normally distributed. Binary logistic regression was developed to handle this case. Instead of modeling the response itself, you use logistic regression to model the probabilities of events.

Suppose your response variable  $Y_i$  has events coded as 0 and non-events coded as 1 for each observation  $i = 1, \dots, n$ , and your  $p$  covariates are  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ . Write the probability of an event as  $\pi_i = \Pr(Y_i = 1)$ . Then the logistic model is written as

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \alpha + \mathbf{x}_i \boldsymbol{\beta} = \alpha + x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{ip}\beta_p$$

where the expected behavior of the responses is modeled by a linear combination of the covariates, the *linear predictor*  $\alpha + \mathbf{X}\boldsymbol{\beta}$ . The quotient  $\frac{\pi_i}{1 - \pi_i}$  is called the *odds*, and the log of the odds is called the *logit* or the *log odds*. This model is linear in the  $\alpha$  and  $\boldsymbol{\beta}$ , and the different combinations of the  $\mathbf{X}$  covariates define subpopulations within which the response is independently binomially distributed. Under simple random sampling within the subpopulations, the data have a product binomial distribution from which you can derive a likelihood and all other results. When using continuous covariates, you might have only one observation in each combination, but this affects only certain statistical results. A general recommendation is that you should have at least 10 events and 10 non-events for each parameter (Peduzzi et al. 1996).

Although binary logistic regression is most common, logistic regression is extensible to more than two response levels. If your response variable takes values that have no inherent ordering (such as voting Democratic, Green, Independent, Republican), then your response is *nominal*. If your response takes values that have an intrinsic order (good, better, best), then your response is *ordinal*.

This paper first reviews how binary logistic regression extends to polytomous logistic regression—in particular, to a special ordinal response model, the proportional odds model combined with a cumulative logit link. Then the proportional odds model is relaxed by using the new UNEQUALSLOPES option in the LOGISTIC procedure to fit the partial proportional odds model. Methods for determining which model applies to your data are also described. The paper ends with suggestions for performing model selection while simultaneously assessing the proportional odds of the individual parameters.

## POLYTOMOUS RESPONSE MODEL

The following section introduces link functions as the response functions for polytomous responses. The next section discusses some associated linear predictors. You can use any link with any of the linear predictors, but some combinations are more standard and more feasible than others.

Binary logistic regression is a generalized linear model that uses the binomial distribution and a logit link function. When your response has more than two levels, you use a multinomial distribution and different link functions to address the nature of the responses, and different linear predictors to model the probabilities. For the following discussion, suppose your response variable  $Y$  has  $J = 4$  levels (1, 2, 3, and 4), and write  $\pi_{ij} = \Pr(Y_i = j)$ .

### Link Functions

#### Generalized Logit Link

Define a link function so that each response function contrasts a lower level with the last level:

$$\log\left(\frac{\pi_{i1}}{\pi_{i4}}\right) \quad \log\left(\frac{\pi_{i2}}{\pi_{i4}}\right) \quad \log\left(\frac{\pi_{i3}}{\pi_{i4}}\right)$$

This is the *generalized logit link*, and it ignores the order of the responses, beyond identifying the last one as the reference response. You use this link when you have nominal data.

#### Cumulative Logit Link

The most popular ordinal link function uses every probability in every function by contrasting the lower levels of  $Y$  with the higher levels of  $Y$ . Let the *cumulative probability* be denoted as  $\theta_{ij} = \Pr(Y_i \leq j)$ :

$$\text{logit}(\theta_{i1}) = \log\left(\frac{\pi_{i1}}{\pi_{i2} + \pi_{i3} + \pi_{i4}}\right) \quad \text{logit}(\theta_{i2}) = \log\left(\frac{\pi_{i1} + \pi_{i2}}{\pi_{i3} + \pi_{i4}}\right) \quad \text{logit}(\theta_{i3}) = \log\left(\frac{\pi_{i1} + \pi_{i2} + \pi_{i3}}{\pi_{i4}}\right)$$

This is the *cumulative logit link*. As you move from the first logit function to the second and from the second logit function to the third, the numerator increases and the denominator decreases, so the cumulative logits are increasing.

### Linear Predictors

#### General Model

To model the variation of the probabilities, you set each of your  $J - 1 = 3$  response functions equal to a different linear predictor,  $\alpha_j + \mathbf{X}'\boldsymbol{\beta}_j$ ,  $j = 1, 2, 3$ . Choosing this model has several consequences:

1. There are  $(J - 1)(1 + p) = 3(1 + p)$  parameters, which can be too many for a successful fit and can violate the ideal of model parsimony.
2. If you take the difference of the  $j$ th response function between any two subpopulations,  $h$  and  $i$ , you get

$$g(hj) - g(ij) = (\mathbf{x}_h - \mathbf{x}_i)' \boldsymbol{\beta}_j \quad j = 1, 2, 3$$

where  $g$  is the link function and  $g(ij)$  is the  $j$ th response function for the  $i$ th observation. The subscript on  $\boldsymbol{\beta}_j$  means that this difference depends on the specific response function, so your conclusions are drawn for each response function separately. If you use the generalized logit link function, this difference is the log odds ratio.

3. There are no constraints on the  $\alpha_j$  and  $\boldsymbol{\beta}_j$ , so there is no guarantee that the response functions are ordered.

Any of the links that are used with this model is an instance of a general (generalized) model (Agresti 2010). In PROC LOGISTIC, specifying the LINK=GLOGIT option in the MODEL statement fits this model to the generalized logit.

#### Proportional Odds Model

The general model has unequal slopes for the predictors, and you need enough data to estimate a different coefficient for each predictor in each response function. To simplify this model, you can force an

ordering on the linear predictors by using the same slope parameters for each response function and by constraining the intercepts to increase ( $\alpha_1 < \alpha_2 < \alpha_3$ ) or decrease:

$$g(j) = \alpha_j + \mathbf{X}'\boldsymbol{\beta} \quad j = 1, 2, 3$$

This model has  $J - 1 + p$  parameters and thus requires less data for an adequate fit than the general model requires. It also provides a more straightforward interpretation. Compute the difference of the  $j$ th response function between two subpopulations  $h$  and  $i$  to see the impact of this model:

$$g(hj) - g(ij) = (\mathbf{x}_h - \mathbf{x}_i)'\boldsymbol{\beta} \quad j = 1, 2, 3$$

This difference is proportional to the distance between the explanatory variables, and the difference is the same no matter which response function you consider. This is the *equal slopes assumption*, which is also called the *parallel lines assumption*. You can apply the parallel lines assumption to any of the link functions, but it is most commonly used with the cumulative logit link. When you use the cumulative logit link, the assumption is the *proportional odds assumption*, the model is the *proportional odds model*, and the difference of cumulative logits ( $g$ ) is the log cumulative odds ratio. By default, PROC LOGISTIC fits the proportional odds model combined with the cumulative logit link when you have more than two response levels.

#### Partial Proportional Odds Model

The third linear predictor is a hybrid of the general model and the proportional odds model. Suppose one set of effects  $\mathbf{X}$  has  $p_1$  parameters that satisfy the parallel lines assumption (that is, they have equal slopes), but the remaining set  $\mathbf{Z}$  has  $p_2$  parameters that do not and instead require the general model (that is, they have unequal slopes). This model is written as

$$g(j) = \alpha_j + \mathbf{X}'\boldsymbol{\beta} + \mathbf{Z}'\boldsymbol{\gamma}_j \quad j = 1, 2, 3$$

It has  $(J - 1) + p_1 + (J - 1)p_2$  parameters and, when used with the cumulative logit link, is called the *partial proportional odds model* (Peterson and Harrell 1990). Interpretation of the proportional odds parameters is independent of the response function; interpretation of the general parameters depends on the response function. When you fit the partial proportional odds model, you must be especially careful to ensure that the cumulative logits remain ordered for the data being modeled.

The next sections fit proportional odds models and partial proportional odds models to several data sets.

## ASBESTOS DATA: PROPORTIONAL ODDS MODEL

This example fits a proportional odds model to the following Asbestos data set (Simonoff 2003, Section 10.2), which can be accessed from StatLib:

```
data Asbestos;
  input Task Ventilation Exposure Freq @@;
  datalines;
  0 0 1 29  0 0 2 1  0 0 3 1  0 1 1 3  0 1 2 1  0 1 3 2
  1 0 1 10  1 0 2 1  1 0 3 7  1 1 1 3  1 1 2 3  1 1 3 22
  ;
```

The **Task** variable is 0 if the worker removed asbestos tile and 1 if the worker removed asbestos insulation. The **Ventilation** variable is 0 if a negative pressure ventilation system was used and 1 if a general system was in place. The response variable **Exposure** measures the extent of the worker's exposure to asbestos; it takes the values 1 for low exposure, 2 for exposure near the legal limit, or 3 for exposure above the legal limit. The **Freq** variable counts the number of workers in each grouping.

The following code fits a proportional odds model to the Asbestos data set; the classification variables are entered as main effects. The CLASS statement is optional for these data, but you specify the PARAM=REF and REF=FIRST options in the CLASS statement to preserve the 0-1 coding of the two binary covariates (PARAM=REF is the *incremental effects parameterization*). The EFFECTPLOT statement displays a plot of the fit.

```

proc logistic data=Asbestos;
  freq Freq;
  class Task Ventilation / param=ref ref=first;
  model Exposure = Task Ventilation / aggregate scale=none;
  effectplot interaction(x=task plotby=ventilation sliceby=Exposure) / polybar;
run;

```

Because the response has more than two levels, PROC LOGISTIC by default fits cumulative logits combined with the proportional odds model, which has a different intercept for each response function but the same slopes:

$$\text{logit}(\Pr(\text{Exposure} \leq 1 | \mathbf{x}_i)) = \alpha_1 + (\text{Task}=1)\beta_1 + (\text{Ventilation}=1)\beta_2$$

$$\text{logit}(\Pr(\text{Exposure} \leq 2 | \mathbf{x}_i)) = \alpha_2 + (\text{Task}=1)\beta_1 + (\text{Ventilation}=1)\beta_2$$

This model has only four combinations of the explanatory variables, as shown in [Table 1](#).

**Table 1** Subpopulations for the Asbestos Data Set

Subpopulation	Task	Ventilation	Description
a	0	0	Workers who removed tile with negative pressure ventilation
b	1	0	Workers who removed insulation with negative pressure ventilation
c	0	1	Workers who removed tile with general ventilation
d	1	1	Workers who removed insulation with general ventilation

Each subpopulation has two cumulative logits in the model. If you write the cumulative probabilities for the subpopulations in each logit function as  $\theta_{ij} = \Pr(\text{Exposure} \leq j | \text{subpopulation} = i)$ , the model is

$$\begin{bmatrix} \text{logit}(\theta_{a1}) \\ \text{logit}(\theta_{b1}) \\ \text{logit}(\theta_{c1}) \\ \text{logit}(\theta_{d1}) \\ \text{logit}(\theta_{a2}) \\ \text{logit}(\theta_{b2}) \\ \text{logit}(\theta_{c2}) \\ \text{logit}(\theta_{d2}) \end{bmatrix} = \begin{bmatrix} \alpha_1 & & & \\ \alpha_1 & +\beta_1 & & \\ \alpha_1 & & +\beta_2 & \\ \alpha_1 & +\beta_1 & +\beta_2 & \\ \alpha_2 & & & \\ \alpha_2 & +\beta_1 & & \\ \alpha_2 & & +\beta_2 & \\ \alpha_2 & +\beta_1 & +\beta_2 & \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

The first row in the preceding equations shows you that the intercept for the first response function,  $\alpha_1$ , is the log odds of **Exposure** = 1 versus 2 or 3 for **Task** = 0 and **Ventilation** = 0. The fifth row shows you that the intercept  $\alpha_2$  is the log odds of **Exposure** = 1 or 2 versus 3 for **Task** = 0 and **Ventilation** = 0. The second (or sixth) row, compared to the first (or fifth), shows you that  $\beta_1$  is the increment in both response functions (or both types of log cumulative odds) due to **Task** = 1. The third (or seventh) row, compared to the first (or fifth), shows you that  $\beta_2$  is the increment in both types of log cumulative odds due to **Ventilation** = 1.

Exponentiating the cumulative logits produces the cumulative odds, and you can solve for the cumulative probabilities:

$$\text{logit}(\theta_j) = \alpha_j + \mathbf{X}'\boldsymbol{\beta} \rightarrow \frac{\theta_j}{(1-\theta_j)} = \exp(\alpha_j + \mathbf{X}'\boldsymbol{\beta}) \rightarrow \theta_j = \exp(\mathbf{X}'\boldsymbol{\beta}) / (1 + \exp(\mathbf{X}'\boldsymbol{\beta}))$$

[Table 2](#) displays these equations for each of the subpopulations and cumulative logits.

**Table 2** Cumulative Logits, Odds, and Probabilities for the Subpopulations

Subpopulation	Task	Ventilation	Logit	Odds	Probability
a	0	0	$\text{logit}(\theta_{a1}) = \alpha_1$	$e^{\alpha_1}$	$e^{\alpha_1}/(1 + e^{\alpha_1})$
b	1	0	$\text{logit}(\theta_{b1}) = \alpha_1 + \beta_1$	$e^{\alpha_1 + \beta_1}$	$e^{\alpha_1 + \beta_1}/(1 + e^{\alpha_1 + \beta_1})$
c	0	1	$\text{logit}(\theta_{c1}) = \alpha_1 + \beta_2$	$e^{\alpha_1 + \beta_2}$	$e^{\alpha_1 + \beta_2}/(1 + e^{\alpha_1 + \beta_2})$
d	1	1	$\text{logit}(\theta_{d1}) = \alpha_1 + \beta_1 + \beta_2$	$e^{\alpha_1 + \beta_1 + \beta_2}$	$e^{\alpha_1 + \beta_1 + \beta_2}/(1 + e^{\alpha_1 + \beta_1 + \beta_2})$
a	0	0	$\text{logit}(\theta_{a2}) = \alpha_2$	$e^{\alpha_2}$	$e^{\alpha_2}/(1 + e^{\alpha_2})$
b	1	0	$\text{logit}(\theta_{b2}) = \alpha_2 + \beta_1$	$e^{\alpha_2 + \beta_1}$	$e^{\alpha_2 + \beta_1}/(1 + e^{\alpha_2 + \beta_1})$
c	0	1	$\text{logit}(\theta_{c2}) = \alpha_2 + \beta_2$	$e^{\alpha_2 + \beta_2}$	$e^{\alpha_2 + \beta_2}/(1 + e^{\alpha_2 + \beta_2})$
d	1	1	$\text{logit}(\theta_{d2}) = \alpha_2 + \beta_1 + \beta_2$	$e^{\alpha_2 + \beta_1 + \beta_2}$	$e^{\alpha_2 + \beta_1 + \beta_2}/(1 + e^{\alpha_2 + \beta_1 + \beta_2})$

Figure 1 displays the response profiles. The note means that the first logit function compares **Exposure** = 1 to **Exposure** = 2 or 3, and the second logit function compares **Exposure** = 1 or 2 to **Exposure** = 3; that is, they compare lower exposure levels to higher exposure levels. You can compare higher values to lower values by specifying the DESCENDING response option.

**Figure 1** Response Information

Response Profile		
Ordered Value	Exposure	Total Frequency
1	1	45
2	2	6
3	3	32

Probabilities modeled are cumulated over the lower Ordered Values.

Figure 2 displays a score test for the proportional odds assumption; the test does not reject the null hypothesis that the proportional odds assumption holds. The degrees of freedom are the difference between the number of parameters in a general model and the number of parameters in a proportional odds model; in this case,  $6 - 4 = 2$ . This score test actually tends to reject the null hypothesis more often than it should; Stokes, Davis, and Koch (2012) say that this statistic needs approximately five observations (or frequencies) for each outcome at each level of each main effect, because small samples might make the statistic artificially large. This score test is a good confirmatory test if it does not reject the null; however, if it rejects the null, then you need other means to justify the proportional odds assumption.

**Figure 2** Test for Equal Slopes Is Not Rejected

Score Test for the Proportional Odds Assumption		
Chi-Square	DF	Pr > ChiSq
1.6130	2	0.4464

The AGGREGATE SCALE=NONE options produce the goodness-of-fit statistics that are displayed in Figure 3. These tests compare the actual number of observations in each subpopulation to the number that is predicted by the model, and both tests suggest an adequate fit. The degrees of freedom are the number of response functions times the number of profiles minus the number of parameters,  $4 \times 2 - 4 = 4$ . The note in Figure 3, "Number of unique profiles: 4," refers to the four subpopulations that are identified earlier. Stokes, Davis, and Koch (2012) suggest that you need at least five observations per subpopulation and response in order for

the test to be valid. If you have a continuous variable in your model, then you should not use these tests; you should instead compare your model to a larger model and show that your model is just as good.

**Figure 3** The Model Fits the Data Adequately

Deviance and Pearson Goodness-of-Fit Statistics				
Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	2.0001	4	0.5000	0.7357
Pearson	1.8628	4	0.4657	0.7610
Number of unique profiles: 4				

The fit statistics that are shown in [Figure 4](#) are often used to compare nested models. The difference of the  $-2 \text{ Log L}$  statistics forms the likelihood ratio statistic that is shown in [Figure 5](#).

**Figure 4** Fit Statistics

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	151.620	107.914
SC	156.457	117.590
$-2 \text{ Log L}$	147.620	99.914

The three global tests that are displayed in [Figure 5](#) evaluate the significance of all the predictors combined. They tell you only whether the model has some significance; they don't say anything about the effect of individual predictors. There are 4 parameters – 2 intercepts = 2 degrees of freedom for these tests. All tests in [Figure 5](#) reject the null hypothesis that the covariates are unrelated to **Exposure**; that is, the model explains a significant amount of variation in the data.

**Figure 5** Tests for Parameters Being Jointly Zero Are Rejected

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	47.7055	2	<.0001
Score	41.0749	2	<.0001
Wald	29.3468	2	<.0001

The tests that are displayed in [Figure 6](#) are Wald tests that the parameters in a given effect are jointly 0. Because the classification covariates are binary and are coded in the design matrix as a single column, the tests all have one degree of freedom.

**Figure 6** Tests That the Parameters for a CLASS Effect Are All Zero Are Rejected

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Task	1	13.6857	0.0002
Ventilation	1	14.5921	0.0001

The parameter estimates are displayed in [Figure 7](#). There are two intercepts for the two logit functions modeled, but only one slope is reported. The parameter labeled “Intercept 1” is the  $\alpha_1$  parameter for **Exposure** = 1 versus **Exposure** = 2 or 3, and “Intercept 2” is the  $\alpha_2$  parameter for **Exposure** = 1 or 2 versus **Exposure** = 3.

**Figure 7** Parameter Estimates

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	
Intercept 1	1	2.4751	0.5633	19.3075	<.0001	
Intercept 2	1	3.0208	0.6034	25.0599	<.0001	
Task	1	-2.2870	0.6182	13.6857	0.0002	
Ventilation	1	-2.1594	0.5653	14.5921	0.0001	

The fitted model is

$$\text{logit}(\text{Pr}(\text{Exposure} = 1)) = 2.4751 - 2.2870 \times (\text{Task} = 1) - 2.1594 \times (\text{Ventilation} = 1)$$

$$\text{logit}(\text{Pr}(\text{Exposure} \leq 2)) = 3.0208 - 2.2870 \times (\text{Task} = 1) - 2.1594 \times (\text{Ventilation} = 1)$$

Both cumulative logits (or log odds) decrease as the explanatory effects increase. [Table 3](#) shows the fitted cumulative logits for the subpopulations. Recall the interpretation of the parameter estimates in terms of logits; this table shows you that  $\alpha_1=2.4751$  is the log odds of **Exposure** = 1 versus 2 or 3 for a worker with **Task** = 0 and **Ventilation** = 0,  $\alpha_2$  is the log odds of **Exposure** = 1 or 2 versus 3 for a worker with **Task** = 0 and **Ventilation** = 0,  $\beta_1$  is the increment in both log odds for **Task** = 1, and  $\beta_2$  is the increment in both log odds for **Ventilation** = 1.

**Table 3** Fitted Cumulative Logits for the Subpopulations

Subpopulation	Task	Ventilation	Logit		
a	0	0	$\text{logit}(\theta_{a1}) = \alpha_1$	=	2.4751
b	1	0	$\text{logit}(\theta_{b1}) = \alpha_1 + \beta_1$	=	$2.4751 - 2.2870$ = 0.1881
c	0	1	$\text{logit}(\theta_{c1}) = \alpha_1 + \beta_2$	=	$2.4751 - 2.1594$ = 0.3157
d	1	1	$\text{logit}(\theta_{d1}) = \alpha_1 + \beta_1 + \beta_2$	=	$2.4751 - 2.2870 - 2.1594$ = -1.9713
a	0	0	$\text{logit}(\theta_{a2}) = \alpha_2$	=	3.0208
b	1	0	$\text{logit}(\theta_{b2}) = \alpha_2 + \beta_1$	=	$3.0208 - 2.2870$ = 0.7338
c	0	1	$\text{logit}(\theta_{c2}) = \alpha_2 + \beta_2$	=	$3.0208 - 2.1594$ = 0.8614
d	1	1	$\text{logit}(\theta_{d2}) = \alpha_2 + \beta_1 + \beta_2$	=	$3.0208 - 2.2870 - 2.1594$ = -1.4256

[Figure 8](#) displays the odds ratios.

Figure 8 Odds Ratios

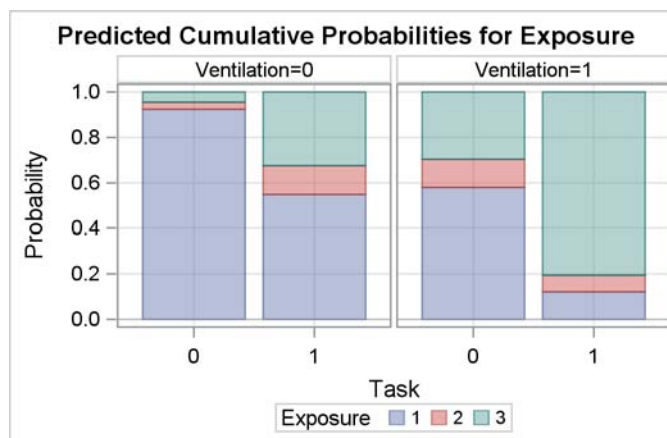
Odds Ratio Estimates				
Effect		Point Estimate	95% Wald Confidence Limits	
Task	1 vs 0	0.102	0.030	0.341
Ventilation	1 vs 0	0.115	0.038	0.349

Note that the interpretation of the parameter estimates in Figure 7 depends on the PARAM= option, but the odds ratios in Figure 8 are independent of the PARAM= option.

The interpretation of these odds ratios is that a person who performs **Task** = 1 has 0.102 times the odds of a low exposure versus a higher exposure than a person who performs **Task** = 0 both for **Exposure** = 1 versus 2 or 3 and for **Exposure** = 1 or 2 versus 3. That is, a person who removes insulation is more likely to have a higher exposure to asbestos than a person who removes tile. Similarly, a person who works in an environment that has general ventilation is more likely to have a higher exposure to asbestos than a person who works in an environment that has negative pressure ventilation. This summary holds for both response functions, because of the proportional odds assumption.

Finally, you can generate plots of the fitted model. The EFFECTPLOT statement produces the plot in Figure 9, where you can easily see that the predicted probability of a low exposure is the highest for workers who remove tile in negative pressure ventilation, and the probability of exposure above the legal limit is highest for workers who remove insulation in general ventilation.

Figure 9 Fitted Probabilities



## COAL DATA: PROPORTIONAL ODDS MODEL

The preceding analysis uses two classification explanatory variables; the following data have a continuous explanatory variable. The following data, from McCullagh and Nelder (1989, p. 179), contain a measure of the severity of pneumoconiosis (black lung disease) in coal miners and the number of years of exposure:

```
data Coal;
  input Severity $ @@;
  do i=1 to 8; input Exposure freq @@; lnExposure=log(Exposure); output; end;
  datalines;
Normal 5.8 98 15 51 21.5 34 27.5 35 33.5 32 39.5 23 46 12 51.5 4
Moderate 5.8 0 15 2 21.5 6 27.5 5 33.5 10 39.5 7 46 6 51.5 2
Severe 5.8 0 15 1 21.5 3 27.5 8 33.5 9 39.5 8 46 10 51.5 5
;
```



The following program fits a proportional odds model to these data and displays plots of the effect of the **lnExposure** variable. By default, PROC LOGISTIC orders the response levels alphanumerically; the ORDER=DATA and DESCENDING options are specified to override the default by fitting cumulative logits of higher versus lower severity.

You should not use the standard goodness-of-fit tests, because the **lnExposure** variable is continuous. A good alternative is to compare your main-effects model to a larger one, and test whether your smaller model is sufficient. A reasonable extension of this model is to square the continuous effect. To get this test in PROC LOGISTIC, you specify a forward selection and include your model. If the fit is good, no other effects are added to your model, and the score test is reported in the "Residual Chi-Square Test" table.

The three EFFECTPLOT statements display the relationship of **lnExposure** to **Severity** on three different scales: cumulative probabilities over a range of exposure levels extended by 50%, logit of cumulative probabilities (because the link function here is the cumulative logit link), and individual probabilities.

```
proc logistic data=Coal;
  model Severity(order=data descending)=lnExposure|lnExposure
    / selection=forward include=1 stop=2;
  freq freq;
  effectplot / noobs extend=1.5;
  effectplot / noobs link;
  effectplot / noobs individual;
run;
```

The proportional odds assumption is not rejected ( $p = 0.7103$ ), so it is reasonable to believe that the response functions have a common **lnExposure** parameter. The global tests are all significant. The residual chi-square test for goodness of fit does not reject your model ( $p = 0.356$ ), where the degrees of freedom are the difference in the number of parameters in each of these models,  $df = 2 - 1 = 1$ .

The parameters are significant, as shown in [Figure 10](#). Because you are fitting two response functions, you have two intercept parameters: the log odds of severe pneumoconiosis versus moderate or normal pneumoconiosis for coal miners who have no log(exposure), and the log odds of severe or moderate pneumoconiosis versus normal pneumoconiosis for coal miners who have **Exposure** = 1. Because of the proportional odds assumption, there is only one **lnExposure** parameter  $\beta_1$ , which is the increment for both types of log odds due to a one-unit increment in log(exposure).

**Figure 10** Parameter Estimates

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	Severe	1	-10.5817	1.3454	61.8569	<.0001
Intercept	Moderate	1	-9.6761	1.3241	53.4042	<.0001
lnExposure		1	2.5968	0.3811	46.4299	<.0001

[Figure 11](#) displays the odds ratios.

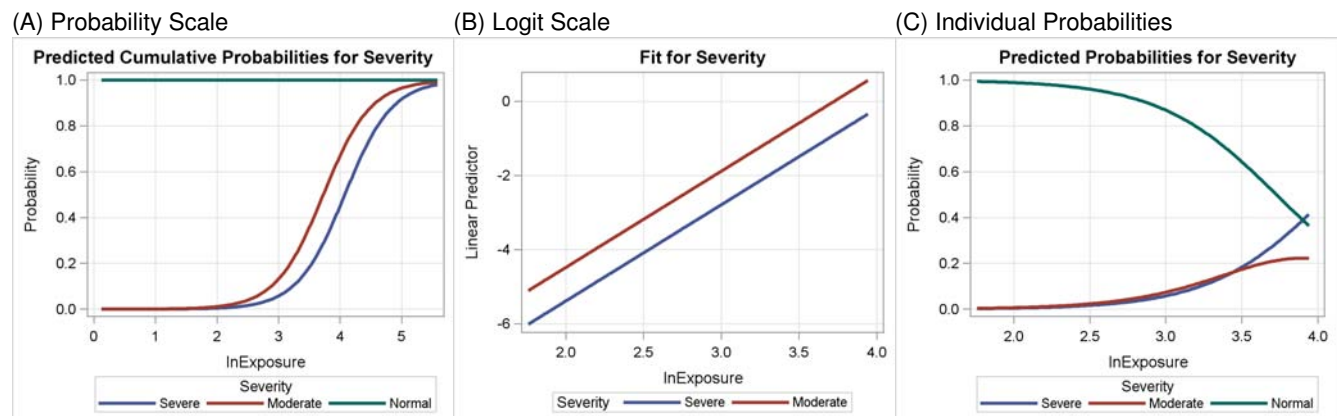
**Figure 11** Odds Ratio for lnExposure

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
lnExposure	13.421	6.359	28.325

The interpretation is that incrementing  $\log(\text{exposure})$  by 1 increases a miner's odds of having more severe pneumoconiosis by a factor of 13.

The EFFECTPLOT statements create several different views of the fitted model. First, Figure 12(A) shows the three estimated cumulative probability functions. The plot is extended beyond the data to display the S-shaped curve of the logistic function. Because the **InExposure** values are fit toward the left half of the logit functions, the predicted probabilities of higher severity are generally quite low. Second, Figure 12(B) shows the fit on the logit scale, which has the parallel lines that you expect when making the proportional odds assumption. Third, Figure 12(C) shows the model-predicted probabilities of a miner having each level of **Severity** across the range of the **InExposure** values. A miner who has little exposure has the highest probability of being normal. As exposure increases, the higher severity levels become more likely and the probability of a normal level decreases. When exposure exceeds 33 years, the probability of a moderate severity levels off. Presumably miners are moving from normal to moderate severity and from moderate to severe, so after this amount of exposure they mostly move into the severe range.

**Figure 12** Displays of the Fitted Model



## BACKACHE DATA: PROPORTIONAL ODDS MODEL

The previous examples illustrate the use of the proportional odds model when the proportional odds assumption is satisfied. The "Backache in Pregnancy" data set from Chatfield (1995, Exercise D.2), accessed from StatLib, challenges those assumptions. The following data were gathered from all women who gave birth in the London Hospital over a four-month period in 1973:

```
data Backache;
  input Severity Month Age @@;
  lnAge=log(Age);
  if Severity=0 then Severity=1;
  Trimester=(Month>6)+(Month>3)+1;
  datalines;

  ... more lines ...

  1 0 23 3 0 36 1 0 21 1 0 30 1 0 42 1 0 34 3 3 26 1 7 18 3 6 39 1 0 25
;
```

The variables that are extracted from this data set are **Severity**, which takes the values 1 = none or very little pain, 2 = troublesome pain, and 3 = severe pain; **Trimester**, which identifies in which trimester the pain starts; and **lnAge**, which is the log of the mother's age in years.

The following program fits a proportional odds model to the data. The third trimester is the reference level for the **Trimester** parameterization. Specifying the DESCENDING option means that the probability of higher severity is modeled.

```
proc logistic data=Backache;
  class Trimester / param=ref;
  model Severity(descending)=lnAge Trimester;
run;
```

The analysis (not shown) indicates that **lnAge** and **Trimester** do indeed significantly affect **Severity**, but the score test for the proportional odds assumption in [Figure 13](#) is strongly rejected.

**Figure 13** Proportional Odds Rejected!

Score Test for the Proportional Odds Assumption		
Chi-Square	DF	Pr > ChiSq
22.2794	3	<.0001

As mentioned earlier, rejection tends to occur more often than it should. Now what should you do?

You can first try to determine whether the proportional odds assumption is actually valid. There are other ways to assess the proportional odds structure besides relying on this test. Two graphical methods are discussed in the next section.

## GRAPHICAL ASSESSMENT OF PROPORTIONAL ODDS

One way to judge whether the proportional odds assumption holds is to look at plots of the empirical cumulative logit function. Recall that if you fit a proportional odds model to data, then plots of the resulting fitted cumulative logits are parallel. (See [Figure 12](#).) Conversely, if the empirical cumulative logits look approximately parallel, then this provides evidence that a proportional odds model is appropriate. The %EmpiricalLogitPlot macro uses the code that is provided in SAS Note 37944 (2012) to produce empirical cumulative logit plots. (See [support.sas.com/statpapercode](http://support.sas.com/statpapercode) for this code and other code used in this paper.) For a classification covariate **A**, compute the empirical cumulative logits as

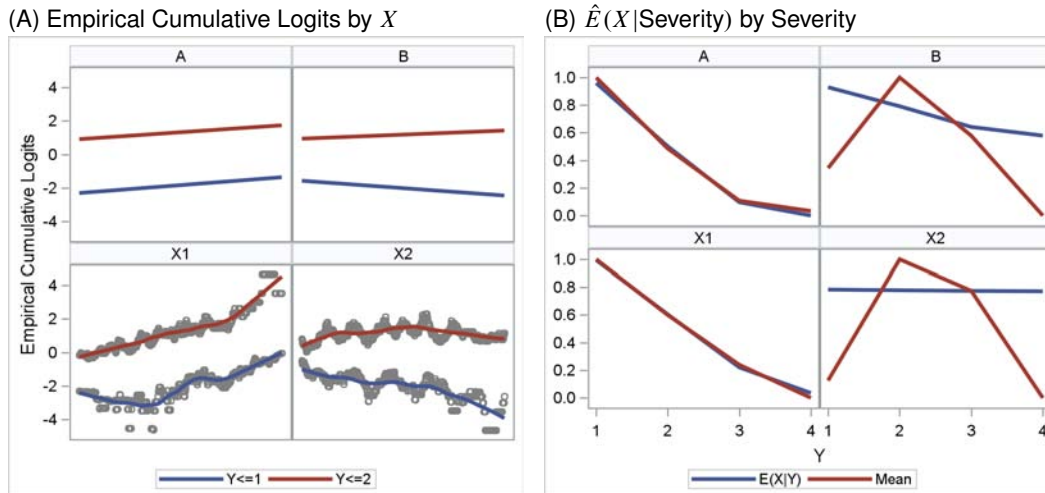
$$\log \frac{\#\{Y \leq j \mid A = a\}}{\#\{Y > j \mid A = a\}} \quad \text{for } j = 1, 2 \text{ and } a = 0, 1$$

To compute empirical cumulative logits for the continuous covariates, accumulate counts  $\#\{Y \leq j\}$ ,  $j = 1, 2$  in a neighborhood of each value, use these values to compute an empirical cumulative logit, and smooth the resulting plot. The empirical cumulative logits are marginal; that is, they do not adjust for any other variables in the model. An example of these curves for continuous variables **X1** and **X2** and classification variables **A** and **B**, where both the **A** and **X1** variables have equal slopes and the **B** and **X2** variables have unequal slopes, is displayed in [Figure 14\(A\)](#). In order for the variables to satisfy the proportional odds assumption, the empirical cumulative logit curves should move in a similar fashion while holding an approximately constant distance between them.

Harrell (2001) proposes an alternative graphical technique for assessing proportionality. He compares the mean value of the predictor  $X$  within each level of the response  $Y$  to the model-expected value of  $X \mid Y = j$ , given that the proportional odds assumption holds for the single-effect model. Again, the effects are not adjusted for the other parameters in the model, so this provides an unadjusted check of the assumptions. The %EXPlot macro computes these statistics. An example of this plot is shown in [Figure 14\(B\)](#), which uses the same variables as (A). These plots are interpreted as follows:

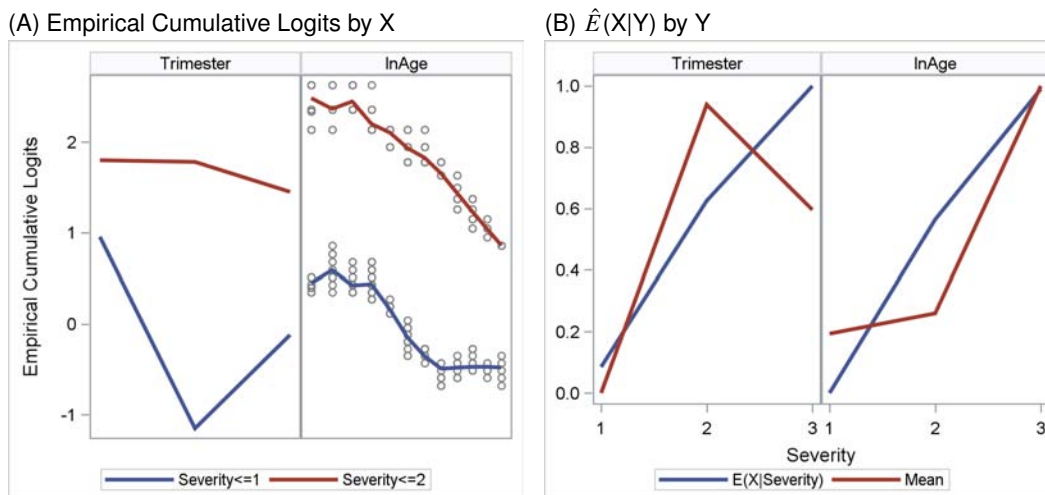
- To confirm the ordinality of the response for a predictor, the means should be strictly increasing or decreasing with respect to the  $Y$  variable.
- To assess the proportional odds assumption for a predictor, the model-expected value curve should closely follow the mean curve.

The macro requires CLASS variables to be numeric; otherwise, you can use the coding from the design matrix data set that is provided by the OUTDESIGN= option in the PROC LOGISTIC statement.

**Figure 14** Assessment Plots (A and X1 have proportional odds)

The following code applies these macros to the Backache data set and produces [Figure 15](#):

```
%EmpiricalLogitPlot (class=Trimester, cont=lnAge, data=Backache, y=Severity) ;
%EXPPlot (Trimester lnAge, data=Backache, y=Severity) ;
```

**Figure 15** Assessment Plots for Backache Data

The plots for **lnAge** show no profound departure from proportionality (the mean curve in (B) is increasing), but the plots for **Trimester** certainly do show a departure. But how do you generate a model in which the proportional odds structure holds for one covariate but not for another? The answer is the *partial proportional odds model*, a hybrid method in which the **lnAge** variable has a proportional odds structure and the **Treatment** variable has a general structure. The partial proportional odds model for the Backache data is

$$\text{logit}(\Pr(\text{Severity} = 3)) = \alpha_1 + (\text{Trimester} = 1)\beta_{11} + (\text{Trimester} = 2)\beta_{21} + \text{lnAge}\beta_3$$

$$\text{logit}(\Pr(\text{Severity} \geq 2)) = \alpha_2 + (\text{Trimester} = 1)\beta_{12} + (\text{Trimester} = 2)\beta_{22} + \text{lnAge}\beta_3$$

where the response functions share the **lnAge** parameter but each function has its own version of the intercept, the **Trimester** = 1 parameter, and the **Trimester** = 2 parameter. To interpret these parameters and their odds ratios, the **lnAge** parameter is treated as before—your interpretation applies to both response functions. However, when you discuss the **Trimester** = 1 parameter, you have to distinguish between the

**Severity** = 3 versus 2 or 1 logit function and the **Severity** = 3 or 2 versus 1 logit function. The partial proportional odds model is fit in the next section.

## BACKACHE DATA: PARTIAL PROPORTIONAL ODDS MODEL

With SAS/STAT 12.1, you can use the LOGISTIC procedure to fit the partial proportional odds model. The following code fits this model to the Backache data; the model is specified in the same way as you would specify a proportional odds model, except that you add the UNEQUALSLOPES option.

```
proc logistic data=Backache;
  class Trimester / param=ref;
  model Severity(descending)=lnAge Trimester / unequalslopes=Trimester;
  oddsratio lnAge;
  oddsratio Trimester;
run;
```

Specifying the UNEQUALSLOPES option modifies models that use the cumulative link functions. Instead of applying the proportional odds assumption to every effect, you can apply the assumption to specific effects. If you do not specify any effects in the UNEQUALSLOPES option, then the general model is fit to the cumulative links. The general model has the same number of parameters that you get by specifying LINK=GLOGIT (although the link is different). If you have a predictor that does not satisfy the proportional odds assumption, then the UNEQUALSLOPES option enables you to fit a different parameter for every response function; that is, the predictor has *unequal slopes* across the response functions. If you have a predictor that satisfies the proportional odds assumption, then it has *equal slopes* across the response functions.

In this case, specifying the UNEQUALSLOPES=**Trimester** option creates a different parameter for each level of the **Trimester** effect in each response function, and the proportional odds structure still holds for the **lnAge** variable. The ODDSRATIO statements are also specified to compute odds ratios that compare all levels of the **Trimester** effect.

The global tests in [Figure 16](#) conclude that the parameters are significantly nonzero, and the Type 3 tests show that the group of four parameters that make up the **Trimester** effect (**Trimester** = 1 and 2 for each response function) are not all 0.

**Figure 16** Tests of Effect Parameters

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	44.0609	5	<.0001
Score	41.5082	5	<.0001
Wald	35.9914	5	<.0001
Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
lnAge	1	5.0543	0.0246
Trimester	4	33.7063	<.0001

As in the analysis of the Coal data set, you should not use the standard goodness-of-fit tests because the **lnAge** variable is continuous. Instead compare your model to a reasonable extension: in this case, take the interaction and square the continuous effect. You can produce the Rao score test in [Figure 17](#) from a forward selection step as shown in the following code:

```
proc logistic data=Backache;
  class Trimester / param=ref;
  model Severity(descending)=lnAge Trimester lnAge*lnAge lnAge*Trimester
    / unequalslopes=(Trimester lnAge*Trimester)
    selection=forward include=2 stop=3;
  ods select ResidualChiSq;
run;
```

Figure 17 Tests of Effect Parameters

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
8.8720	5	0.1143

The score test in Figure 17 suggests the main-effects model fits the data adequately.

Alternatively, you can use a likelihood ratio test to compare the two models, just as the likelihood ratio global test compares your model to an intercept-only model. The following %LRTest macro call uses a likelihood ratio test to compare these two models and does not reject the main-effects model ( $p = 0.357$ ):

```
%LRTest (data=Backache, class=Trimester / param=ref,
  model1=Severity(descending) = lnAge Trimester / unequalslopes=Trimester,
  model2=Severity(descending) = lnAge Trimester lnAge*lnAge lnAge*Trimester
    / unequalslopes=Trimester);
```

Figure 18 shows that the **Trimester** effects for the first response function are not significantly nonzero.

Figure 18 Parameter Estimates

Analysis of Maximum Likelihood Estimates						
Parameter	Severity	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	3	1	-7.2266	2.5883	7.7958	0.0052
Intercept	2	1	-5.5761	2.5526	4.7719	0.0289
lnAge		1	1.7670	0.7860	5.0543	0.0246
Trimester 1	3	1	-0.3949	0.4961	0.6334	0.4261
Trimester 1	2	1	-1.1475	0.3873	8.7778	0.0030
Trimester 2	3	1	-0.3262	0.5533	0.3476	0.5555
Trimester 2	2	1	1.0891	0.4427	6.0524	0.0139

The fitted model is

$$\begin{aligned} \text{logit}(\text{Pr}(\text{Severity} = 3)) &= -7.2266 + 1.767 \times \text{lnAge} - 0.3949 \times (\text{Trimester} = 1) - 0.3262 \times (\text{Trimester} = 2) \\ \text{logit}(\text{Pr}(\text{Severity} \geq 2)) &= -5.5761 + 1.767 \times \text{lnAge} - 1.1475 \times (\text{Trimester} = 1) + 1.0891 \times (\text{Trimester} = 2) \end{aligned}$$

Even though the **Trimester** effects for the first logit function are not significant, PROC LOGISTIC does not set the estimates to 0. Peterson and Harrell (1990) introduce a *constrained* partial proportional odds model, which enables you to multiply the **Trimester** parameters by a different specified constant for each response function—multiplying the parameter in the first response function by 0 and in the second response function by 1 removes the effect from the first response function. You can look for the constrained model in a future release of PROC LOGISTIC.

You can interpret the intercepts and the **InAge** parameter as before. The first intercept ( $-7.2266$ ) is the log odds of **Severity** = 3 versus 2 or 1 for **InAge** = 0 and **Trimester** = 3. The second intercept ( $-5.5761$ ) is the log odds of **Severity** = 3 or 2 versus 1 for **InAge** = 0 and **Trimester** = 3. The **InAge** parameter ( $1.767$ ) is the increment in log odds of a higher **Severity** for a one-unit increase in **InAge**.

The interpretation of the **Trimester** parameters is complicated because they depend on the response function. In the first response function, the **Trimester** = 1 parameter ( $-0.3949$ ) is the increment in log odds for **Severity** = 3 versus 2 or 1 for pain onset in **Trimester** = 1, and the **Trimester** = 2 parameter ( $-0.3262$ ) is the increment in log odds for **Severity** = 3 versus 2 or 1 for pain onset in **Trimester** = 2. Neither of these is significant. In the second response function, the **Trimester** = 1 parameter ( $-1.1475$ ) is the increment in log odds for **Severity** = 3 or 2 versus 1 for pain onset in **Trimester** = 1, and the **Trimester** = 2 parameter ( $1.0891$ ) is the increment in log odds for **Severity** = 3 or 2 versus 1 for pain onset in **Trimester** = 2.

Similarly, the **Trimester** odds ratios that are displayed in [Figure 19](#) also depend on the response functions. For the response function that compares **Severity** = 3 versus 2 or 1, every confidence interval contains 1, so mothers who experience pain onset in any trimester are equally likely to have higher-severity pain; that is, the **Trimester** effect does not describe any variation in the first response function. However, in the second response function, which compares **Severity** = 3 or 2 versus 1, mothers who experience pain onset in the first trimester are less likely to have higher-severity pain than mothers who experience pain onset in the later trimesters, and mothers with pain onset in the second trimester are more likely to have higher-severity pain than mothers with pain onset in the third trimester. A mother with a one-unit increase in **InAge** has 5.853 times the odds of having (is more likely to have) a higher severity, for both response functions.

**Figure 19** Odds Ratios

Odds Ratio Estimates and Wald Confidence Intervals			
Label	Estimate	95% Confidence Limits	
<b>InAge</b>	5.853	1.254	27.318
<b>Severity 3: Trimester 1 vs 2</b>	0.934	0.334	2.611
<b>Severity 2: Trimester 1 vs 2</b>	0.107	0.047	0.244
<b>Severity 3: Trimester 1 vs 3</b>	0.674	0.255	1.782
<b>Severity 2: Trimester 1 vs 3</b>	0.317	0.149	0.678
<b>Severity 3: Trimester 2 vs 3</b>	0.722	0.244	2.135
<b>Severity 2: Trimester 2 vs 3</b>	2.971	1.248	7.076

## ALTERNATIVE METHODS TO ASSESS THE PROPORTIONAL ODDS ASSUMPTION

Although the previously discussed plots can suggest model appropriateness, performing tests gives you a more formal foundation for your model fitting. All tests of the proportionality assumption for an effect rely on comparing a model in which that effect has a proportional odds structure to a model in which that effect does not. The differences in the tests are how you make adjustments for other variables in the model and whether you implement these tests as likelihood ratio, score, or Wald tests.

Three different ways of handling other effects in the model are as follows:

- Up Start small with a proportional odds model and compare it to unequal slopes on each effect in turn (Peterson and Harrell 1990).
- Down Start big with a general model and compare it to equal slopes on each effect in turn (Koch, Amara, and Singer 1985, p. 372).
- OneUp Study proportionality for a series of one-effect models, for which Up and Down are equivalent (based on Harrell 2001, p. 332).

None of these types of test are perfect: Up and OneUp are apt to signal nonproportional odds too often; Down, which starts big, can require a lot of data and might be less able to detect proportional odds.

You can implement these tests in three ways:

- Wald tests can be computed directly by using the TEST statement in PROC LOGISTIC, but they are usually not as good as either likelihood ratio tests or score tests.
- Likelihood ratio tests require only a little SAS programming to extract the values of the fitted log likelihood from the appropriate PROC LOGISTIC table and compare the difference of their log likelihoods to a chi-square distribution. The degrees of freedom are equal to the difference of the degrees of freedom of the two models. The %LRTest macro facilitates this procedure.
- Score tests (Peterson and Harrell 1990) are traditionally computed in PROC LOGISTIC by using the model selection machinery to produce them from the residual chi-square test for adding the effect in question last.

You can run the likelihood ratio test version of the Up, Down, and OneUp tests for every variable in the model by calling the %LRTestCycle macro as in the following code. The macro uses the UNEQUALSLOPES option to generate the models in the Up, Down, and OneUp tests that involve nonproportional odds.

```
%LRTestCycle(Up, data=Backache, class=Trimester/param=ref, model=Severity = lnAge Trimester);
%LRTestCycle(Down, data=Backache, class=Trimester/param=ref, model=Severity = lnAge Trimester);
%LRTestCycle(OneUp, data=Backache, class=Trimester/param=ref, model=Severity = lnAge Trimester);
```

Results from performing these likelihood ratio tests on the Backache data set, and from performing the corresponding score tests, are summarized in [Table 4](#).

The following program computes the Up version of a score test for evaluating the **lnAge** variable from the Backache data set. It copies the **lnAge** variable and specifies the copy, **lnAgeB**, as the last variable in the MODEL statement and in the UNEQUALSLOPES option. [Table 4](#) summarizes the results from these tests, in addition to score tests for the Down and OneUp tests. Small *p*-values reject the smaller model.

```
data Backache; set Backache; lnAgeB=lnAge; TrimesterB=Trimester;
proc logistic data=Backache;
  class Trimester TrimesterB / param=ref;
  model Severity(descending)=lnAge Trimester lnAgeB / unequalslopes=lnAgeB
    selection=forward include=2 stop=3;
  ods select ResidualChiSq;
run;
```

Note that, in the preceding code, if you specify every copied variable in the MODEL statement and in the UNEQUALSLOPES option, then the “Residual Chi-Square” test is the score test for the proportional odds assumption that PROC LOGISTIC reports by default.

**Table 4** *p*-Values from Tests of Proportionality of Effects

Variable	LRTUp	ScoreUp	LRTDown	ScoreDown	LRTOneUp	ScoreOneUp
<b>lnAge</b>	0.038	0.041	0.258	0.275	0.060	0.064
<b>Trimester</b>	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

[Table 4](#) reports that the Up tests reject proportional odds for both variables, but the OneUp and Down tests disagree for **lnAge**. To break the tie, compare the two competing models. Both models have **Trimester** with unequal slopes, one model has **lnAge** with proportional odds, and the other model has **lnAge** with unequal slopes. This is just the Down test that you already ran, so you can conclude that the partial proportional odds model fit in the previous section is appropriate.

## DOCVISIT DATA: VARIABLE SELECTION

The preceding sections discuss the proportional odds model and provide several methods of determining which effects should have equal slopes and which effects should have unequal slopes. If you are performing a variable selection, you can work the proportional odds determination into the selection process. There are two ways to proceed. The first way is to determine which effects should have unequal slopes and then invoke a standard selection algorithm. The second way is to build the proportional odds determination into the steps



of the selection algorithm. The following sections discuss these two approaches by using the Docvisit data set (Cameron and Trivedi 1998, p. 68) from Example 54.17 of the PROC LOGISTIC documentation in the SAS/STAT 12.1 User's Guide.

```
data Docvisit;
  input Sex Age Agesq Income Levyplus Freepoor Freerepa
        Illness Actdays Hscore Chcond1 Chcond2 Dvisits;
  if ( Dvisits > 2) then Dvisits = 2;
  Dvisitsp1= Dvisits+1;
datalines;
1 0.19 0.0361 0.55 1 0 0 1 4 1 0 0 1
... more lines ...
0 0.72 0.5184 0.25 0 0 1 0 0 0 0 0 0
;
```

The dependent variable, **Dvisits**, contains the number of doctor visits in the past two weeks (0, 1, or 2, where 2 represents two or more visits). Details about the other variables in the data set can be found in the PROC LOGISTIC documentation.

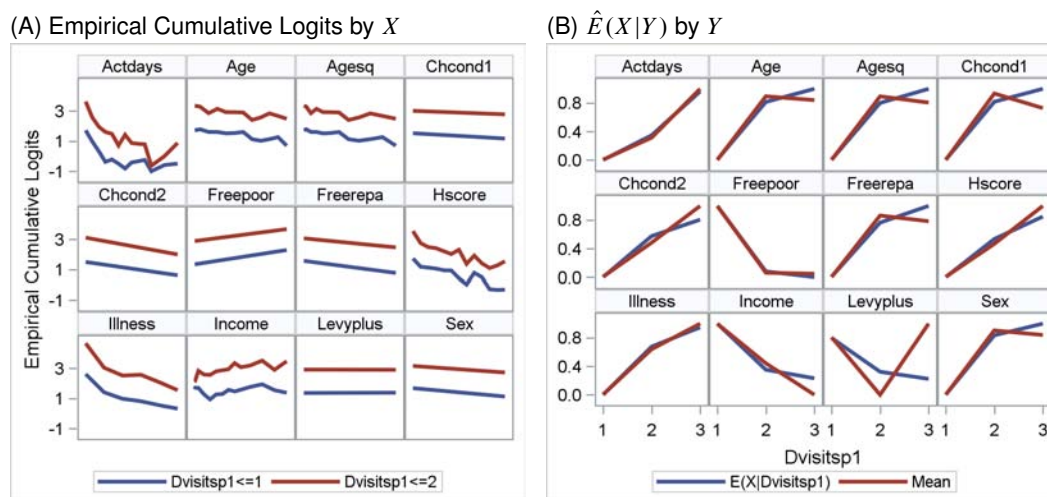
### Proportional Odds Determination Followed by Stepwise Selection

In this section, you determine which effects should have unequal slopes before invoking the usual stepwise selection method. For the determination of unequal slopes, you can use the plots and the likelihood ratio tests that are described earlier.

First, create the two graphical assessment plots discussed earlier and see what they suggest. The macros require the response variable to take the values 1, 2, 3,..., so the **Dvisitsp1** variable is used. Because all the continuous variables take only a small number of distinct values, every variable is listed in the CLASS= option of the %EmpiricalLogitPlot macro. The following code creates the plots in Figure 20:

```
%EmpiricalLogitPlot (data=Docvisit,y=Dvisitsp1,class=Sex Age Agesq Income
  Levyplus Freepoor Freerepa Illness Actdays Hscore Chcond1 Chcond2);
%EXPlot (Sex Age Agesq Income Levyplus Freepoor Freerepa Illness
  Actdays Hscore Chcond1 Chcond2,data=Docvisit,y=Dvisitsp1);
```

**Figure 20** Assessment Plots for Docvisit Data



The empirical cumulative logits in Figure 20(A) suggest that **Actdays** and **Income** should have unequal slopes, whereas **Age**, **Agesq**, and **Hscore** might have unequal slopes. The  $\hat{E}(X|Y)$  plots in Figure 20(B) show that the non-ordinality of the **Levyplus** and **Chcond1** mean curves gives them unequal slopes. The slight non-ordinality of the **Age**, **Agesq**, **Freerepa**, and **Sex** mean curves and their slight divergence from the expected curves suggest they might have unequal slopes. Finally, choose effects that either plot is sure

of or that both sets of plots agree on: let **Actdays**, **Income**, **Levyplus**, **Chcond1**, **Age**, and **Agesq** have unequal slopes. Table 5 contains these conclusions in the Graphical row.

Second, the results from the following Down likelihood ratio tests are displayed in Table 5, and tests with  $p$ -values less than 0.1 are chosen to have unequal slopes.

```
%LRTestCycle(Down,data=Docvisit,class=Sex Levyplus Freepoor Freerepa
Chcond1 Chcond2 / param=ref ref=first,model= Dvisits = Sex Age Agesq
Income Levyplus Freepoor Freerepa Illness Actdays Hscore Chcond1 Chcond2);
```

Third, you can modify the proportional odds detection by iterating the tests. One possibility is to start from a general model with the initial Down test results, remove the effect with the largest  $p$ -value from the UNEQUALSLOPES list, and continue (using the %LRTest macro) until all  $p$ -values are less than 0.1. The conclusions are displayed as the Iterative row in Table 5.

**Table 5** Test Results for Proportional Odds of Effects

Method	Sex	Age	Agesq	Income	Levyplus	Freepoor	Freerepa	Illness	Actdays	Hscore	Chcond1	Chcond2
Graphical	equal	neq	neq	neq	neq	equal	equal	equal	neq	equal	neq	equal
LRTDown	0.298	0.113	0.053	0.003	0.899	0.598	0.158	0.354	0.011	0.827	0.405	0.804
Iterative	equal	neq	neq	neq	equal	equal	neq	equal	neq	equal	equal	equal

Besides **Agesq**, **Income**, and **Actdays**, Table 5 shows that the methods have little agreement on which variables should have nonproportional odds. However, you might not have enough data to make a determination for nonsignificant effects, so the disagreement might ultimately be unimportant.

Perform a stepwise selection for the three competing models. The following program runs the model that is given in the Graphical row of Table 5, and the other models differ only in the UNEQUALSLOPES specification. The results are summarized in the section “Selection Results.”

```
proc logistic data=Docvisit;
class Sex Levyplus Freepoor Freerepa Chcond1 Chcond2 / param=ref;
model Dvisits = Sex Age Agesq Income Levyplus Freepoor Freerepa
Illness Actdays Hscore Chcond1 Chcond2
/ unequalslopes=(Age Agesq Income Levyplus Actdays Chcond1)
selection=stepwise;
run;
```

### Proportional Odds Determination Built into Stepwise Selection

You can insert the nonproportional odds check into stepwise selection as follows: (1) Begin with the intercept-only model. (2) Select an effect to enter the model under nonproportional odds by using forward selection with the INCLUDE= and STOP= options. (3) Determine whether that variable should have proportional odds with a likelihood ratio test. (4) Use backward selection to check whether any variables currently in the model should be dropped. (5) Repeat steps (2) through (4) until no more variables enter the model. (6) Use a likelihood ratio test to recheck the proportional odds assumptions for every selected effect. The steps are not shown here, but the resulting model is identical to that of the method in the next paragraph.

It is easier to copy every variable in the model, make one copy have equal slopes and the other unequal slopes, and then run a stepwise selection, which uses score tests to compare models on the forward steps and uses Wald tests to compare models on the backward steps. If an effect and its copy (for example, **Age** and **AgeB**, respectively) end up in the model, then **Age** acts like a trend and can be dropped with no loss of information. The results of the following program are summarized in the next section:

```
data two; set Docvisit;
SexB=Sex; AgeB=Age; AgesqB=Agesq; IncomeB=Income; LevyplusB=Levyplus;
FreepoorB=Freepoor; FreerepaB=Freerepa; IllnessB=Illness;
ActdaysB=Actdays; HscoreB=Hscore; Chcond1B=Chcond1; Chcond2B=Chcond2;
run;
```

```

proc logistic data=two;
  class Sex Levyplus Freepoor Freerepa Chcond1 Chcond2 SexB LevyplusB
    FreepoorB FreerepaB Chcond1B Chcond2B / param=ref ref=first;
  model Dvisits = Sex Age Agesq Income Levyplus Freepoor Freerepa Illness
    Actdays Hscore Chcond1 Chcond2 SexB AgeB AgesqB IncomeB LevyplusB
    FreepoorB FreerepaB IllnessB ActdaysB HscoreB Chcond1B Chcond2B /
    unequalslopes=(SexB AgeB AgesqB IncomeB LevyplusB FreepoorB FreerepaB
      IllnessB ActdaysB HscoreB Chcond1B Chcond2B)
  selection=stepwise details;
run;

```

## Selection Results

The selection methods in the preceding sections all choose **Sex**, **Freepoor**, **Illness**, and **Hscore** as proportional odds effects and **Agesq** and **Actdays** as general effects. Of these methods, copying the variables and using the selection methods is the easiest to implement using PROC LOGISTIC. Finish the Docvisit analysis by fitting the selected model as follows:

```

proc logistic data=Docvisit;
  class Sex Freepoor / param=ref ref=first;
  model Dvisits = Sex Agesq Freepoor Illness Actdays Hscore /
    unequalslopes=(Agesq Actdays);
run;

```

Looking at the odds ratios for the proportional odds parameters in [Figure 21](#), you see that mothers with **Sex** = 1 have 0.740 times the odds of fewer doctor visits than mothers with **Sex** = 0, for both types of cumulative logits. Mothers with **Freepoor** = 1 have 2.27 the odds of fewer doctor visits than mothers with **Freepoor** = 0. Mothers with one more day of illness have 0.748 times the odds of fewer doctor visits. Mothers who score one point higher on their **Hscore** have 0.938 times the odds of fewer doctor visits. So mothers with **Sex** = 0, **Freepoor** = 1, less illness, and lower **Hscores** are less likely to have more doctor visits.

Now look at the odds ratios for the general parameters. Mothers whose **Agesq** is 1 unit higher have 0.335 times the odds of 0 doctor visits than 1 or 2 visits. The **Agesq** parameter for the second response function is not significantly different from 0, so these mothers whose **Agesq** is 1 unit higher are equally as likely to have 0 or 1 doctor visits as they are to have 2 visits. Both parameters for **Actdays** are around 0.8, so mothers with lower **Actdays** are less likely to have more doctor visits, for both response functions.

**Figure 21** Odds Ratios

Odds Ratio Estimates				
Effect	Dvisits	Point Estimate	95% Wald Confidence Limits	
<b>Sex</b> 1 vs 0		0.740	0.635	0.862
<b>Agesq</b>	0	0.335	0.224	0.500
<b>Agesq</b>	1	0.793	0.399	1.575
<b>Freepoor</b> 1 vs 0		2.270	1.392	3.701
<b>Illness</b>		0.748	0.710	0.788
<b>Actdays</b>	0	0.857	0.838	0.876
<b>Actdays</b>	1	0.828	0.807	0.849
<b>Hscore</b>		0.938	0.909	0.969

## CONCLUSION

This paper shows you how to use the LOGISTIC procedure to fit proportional odds models and how to interpret the results. The UNEQUALSLOPES option, available in SAS/STAT 12.1, is introduced to fit partial proportional odds models and general models to cumulative logits. Two graphical methods for assessing

proportionality are discussed. The UNEQUALSLOPES option also enables you to determine which variables satisfy the proportional odds assumptions and which do not.

Expanding the scope of the LOGISTIC procedure is an active area of SAS/STAT development. You can look for the following features in a future release of PROC LOGISTIC: enabling effects to have both general and proportional odds structures, fitting the unconstrained partial proportional odds model, fitting generalized logit links with proportional odds and partial proportional odds models, and providing alternative ordinal links such as the adjacent-category logit.

## REFERENCES

- Agresti, A. (2010), *Analysis of Ordinal Categorical Data*, 2nd Edition, New York: John Wiley & Sons.
- Cameron, A. C. and Trivedi, P. K. (1998), *Regression Analysis of Count Data*, Cambridge: Cambridge University Press.
- Chatfield, C. (1995), *Problem Solving: A Statistician's Guide*, 2nd Edition, Boca Raton, FL: Chapman & Hall/CRC.
- 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 Edition, London: Chapman & Hall.
- 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>.
- Simonoff, J. S. (2003), *Analyzing Categorical Data*, New York: Springer-Verlag.
- Stokes, M. E., Davis, C. S., and Koch, G. G. (2012), *Categorical Data Analysis Using SAS*, 3rd Edition, Cary, NC: SAS Institute Inc.

## ACKNOWLEDGMENTS

The author is grateful to Maura Stokes, Randy Tobias, David Schlotzhauer, Ed Huddleston, Anne Baxter, and Tim Arnold of the Advanced Analytics Division at SAS Institute Inc. for their valuable assistance in the preparation of this paper.

## CONTACT INFORMATION

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

Bob Derr  
SAS Institute Inc.  
SAS Campus Drive  
Cary, NC 27513  
bob.derr@sas.com

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 and product names are trademarks of their respective companies.