The Proportional Odds Model:
Simulation Studies and Predictive Accuracy

Andrew Thomas

1 Introduction

The goal of the research that lead to this paper was chiefly to examine the proportional odds model (McCullagh, 1980)—how it fits and how it predicts. This model can be seen below in (1). It is used to model categorical response data when the response categories have a natural ordering. Let J be the number of response categories. Let Y be the random response which is a numerically coded categorical random variable taking values in the set $\{1, \ldots, J\}$. Let $\mathbf{x} = (x_1, \ldots, x_p)'$ denote the values of the p predictors. The model assumes that

$$\log \frac{P(Y \leq j|\mathbf{x})}{P(Y > j|\mathbf{x})} = \text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j - \boldsymbol{\beta}'\mathbf{x}, \quad j = 1, \ldots, J-1, \tag{1}$$

where $\alpha_1, \ldots, \alpha_{J-1}, \boldsymbol{\beta}$ are unknown parameters and logit(u)= $\log \frac{u}{1-u}$ (Agresti, 2010). This exploits the ordinality of the response because an increase in $\mathbf{x}$ leads to a decrease in probabilities for all categories less than or equal to j, not just j. We call it proportional odds because for two different predictor vectors $\mathbf{x}_1$ and $\mathbf{x}_2$

$$\log \frac{\text{odds}(Y \leq j|\mathbf{x_1})}{\text{odds}(Y \leq j|\mathbf{x_2})} = \boldsymbol{\beta}'(\mathbf{x}_1 - \mathbf{x}_2), \tag{2}$$

where the right-hand side is the same for all j. A concrete illustration of this assumption can be seen in the section 4 on predictive accuracy.

In this paper I will go over the likelihood ratio test of the proportional odds assumption and show it is valid. Second, I will look at the predictive accuracy[1] of this model compared to the multinomial logit model. Third, I will look at the utility of this model with qualitative predictors. Finally, I will look at literature where a proportional odds model was fit and see if each of the papers performed a likelihood ratio test or another test of the proportional odds assumption along with other items of interest related to fitting proportional odds models.

[1]Note: In this paper, the predictive accuracy of a model is the proportion of correct classification of response categories by said model.

2 Likelihood-Ratio Test

To begin, this project has aimed to address the specific question: is the proportional odds model and its simplifying assumption adequate in applications? Verifying this could be done via the likelihood-ratio test, fitting the model in (1) which has the proportional odds restriction, and the model where (1) is replaced by

$$\log\frac{P(Y \leq j|\mathbf{x})}{P(Y > j|\mathbf{x})} = \text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j - \boldsymbol{\beta_j}'\mathbf{x}, \quad j = 1, \ldots, J-1, \tag{3}$$

and where $\boldsymbol{\beta_1}, \ldots, \boldsymbol{\beta_{J-1}}$ are $p \times J - 1$ unknown parameters. This is called the generalized proportional odds model[2].

After the models in (1) and (2) are fit, one computes the likelihood ratio statistic $-2(L_0 - L_1)$, where L_0 is the maximized log-likelihood of the cases model implied in (1) and L_1 is the same for (2). This test-statistic approximately has the chi-squared distribution with d degrees of freedom, where d is number of free parameters in (2) minus the free parameters in (1). A significant p-value associated with this test rejects the null hypothesis that the proportional odds assumption holds.

Although the likelihood-ratio test statistics approximately follow the chi-squared distribution with d degrees of freedom, it remains to check how good this approximation is in practice. To check this approximation I ran a simulation study.

The simulation study consisted of randomly generating 1000 datasets from a proportional odds model consisting of three response categories and a single, continuous predictor and then recording the likelihood-ratio test statistics for testing the proportional odds assumption for each data-set. Two different parameter settings were used. The first setting generated response counts with relatively equal frequency (figure 1). The second setting generated response counts unequally (figure 2) so some response counts are not sufficiently large (these values can be seen in table 1). The predictor values were the vector $\mathbf{x}_1$ (with sample size 100), the selection method of which will be described in the following section. Then the p-values for the tests were recorded and plotted in figures 1 and 2. For both sets of parameters, since the null hypothesis is true, the p-values should approximately follow a standard uniform distribution.

In the case of figure 1, where the response category counts were sufficiently large for each

[2] A third form of the proportional odds model that allows for some parameters to satisfy and others to violate the proportional odds assumption. This model is called the partial proportional odds model (Peterson & Harrell, 1990) and can be seen described in the appendix.

Table 1: Coefficients for Figures 1 and 2

	Fig. 1	Fig. 2
α_1	-2.065	-0.877
α_2	2.349	-0.804
β	0.925	0.451

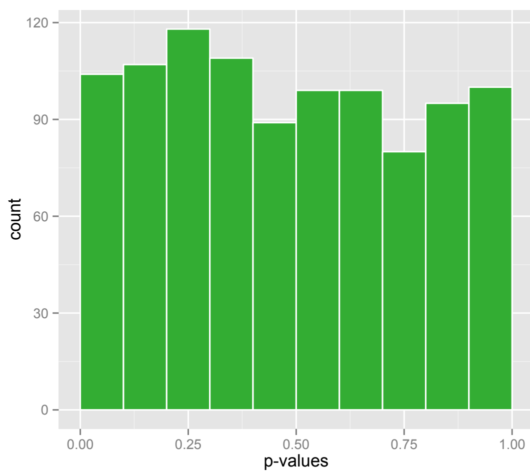

Figure 1: Well fit model p-values

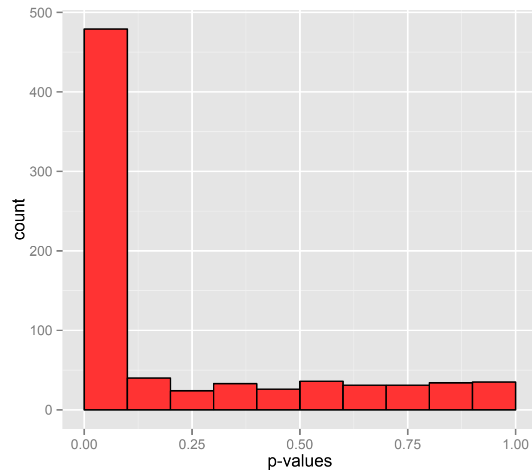

Figure 2: Poorly fit model p-values

of the 1000 trials, the histogram of the p-values in figure 1 look to approximately follow a standard uniform distribution. When the response counts are often not sufficiently large for the 1000 trials, as in figure 2, the p-values do not follow a standard uniform distribution and the p-values seen in this figure are those only for which the model could be fit. Thus, there exist data-generating models where the proportional odds model is correct, but the chi-squared approximation is poor and leads to an inflated type I error rate. For reference, the coefficients for these models corresponding to figures 1 and 2 can be seen in table 1.

3 Model Fitting

Now that the likelihood-ratio test is shown to be valid when the response categories counts are sufficiently large, I turn to model fitting. The main question I decided to pursue when it came to model-fiiting was whether or not the multinomial logit model predicts the correct response category labels at a practically higher rate than the proportional odds model when (1) the data was generated from the multinomial logit model and (2) when the data was generated from the proportional odds model. To do this, I created the R functions `polrgen` and `multgen` (note: All R functions I created for this research can be seen in a link in the appendix). These functions generate the desired trichotomous response category counts from a proportional odds model and multinomial logit model respectively. Another function I created in R was `polr2cv`, which took the predictor vector $\mathbf{x}$ and response vector $\mathbf{y}$ of lengths n as arguments and generated $2n$ values of two-fold cross-validated predictive accuracies from a given single predictor and some J-category response[3]

Before assessing the predictive accuracy of (1) and (2), I needed to randomly generate predictor values (I arbitrarily chose a sample size of 100). I did this by means of a function called `linear` that I created, and randomly generated the predictor vectors $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$, with $\mathbf{x}_1$ being least correlated with its indices and $\mathbf{x}_3$ being most correlated with its indices.

Returning to the main argument, I ran the `polrgen` and `multgen` functions to generate 6 total sets of response counts $\mathbf{y}_{ij}$ with $i = \{p, m\}$ representing the model the data was generated from[4], and $j = \{1, 2, 3\}$ representing the predictor vector used in the generation of response counts. The correlations between the proportional odds model generated response counts and correlations with the predictor vectors can be seen in table 2 and the same can be seen for the multinomial logit in table 3. The green correlations represent the correlation of a specific set of response counts and the predictor vectors from which it was fit. Unsurprisingly, if one looks at the correlation between a certain set of response counts and the other two predictor vectors, the correlations are often less owing to these not being the values from whence the response counts came. The multinomial logit "green" correlations are generally less in practice because of a deficiency in how I randomly chose the values of the coefficients $\boldsymbol{\beta}$ in the function `multgen`. This is hardly a concern because any number of models can be generated and thus the models with the desired properties can be had.

Table 2: Prop. odds fit correlations

	y-values		
$\mathbf{x}$	$\mathbf{y}_{p1}$	$\mathbf{y}_{p2}$	$\mathbf{y}_{p3}$
$\mathbf{x}_1$	**0.619**	-0.078	-0.260
$\mathbf{x}_2$	0.326	**-0.087**	-0.134
$\mathbf{x}_3$	0.177	0.042	**0.265**

Table 3: Multinomial logit fit correlations

	y-values		
$\mathbf{x}$	$\mathbf{y}_{m1}$	$\mathbf{y}_{m2}$	$\mathbf{y}_{m3}$
$\mathbf{x}_1$	**-0.413**	-0.280	-0.132
$\mathbf{x}_2$	-0.421	**0.441**	-0.061
$\mathbf{x}_3$	0.159	0.104	**0.540**

[3] All mentions of generating response counts in this paper concern only a 3-level response.

[4] Indices m and p representing that the responses were randomly generated from the multinomial logit and proportional odds models respectively.

4 Predictive Accuracy

At this point, it is now prudent to check the predictive accuracy of these two models. Before I do that, I would like to make an aside. When fitting the proportional odds model on qualitative predictors, using predictive accuracy to measure the performance of the model is not always very useful—consider the data in table 4 taken from the General Social Survey 2010 (cf. Agresti, 2010). We have GRWTHELP which represents the belief that America needs economic growth to protect the environment (1=strongly agree, ..., 5=strongly disagree) versus DEGREE (0=left high school, ..., 4=graduate degree). As one can see, the table counts are concentrated in the "agree" category (*i.e.* GRWTHELP=2). Fitting the proportional odds model with GRWTHELP as response and DEGREE as predictor, the probabilities are such that in every case the "agree" category has a plurality of the probability mass and so every response category is classified as "agree". This is not useful. However, one has, if the model fits well, that the estimated cumulative odds ratio between two settings of degree is constant between settings of GRWTHELP.

Table 4: Belief that economic growth protects environment by highest education achieved

	GRWTHELP				
DEGREE	1	2	3	4	5
0	40	159	64	53	6
1	106	565	295	231	28
2	12	81	44	42	7
3	32	189	95	113	14
4	13	82	50	56	7

A p-value of 0.633 is given for the likelihood-ratio test that the proportional odds assumption holds and thus the assumption is not rejected. Note: DEGREE=0 is the baseline category. Thus, the odds ratio for GRWTHELP being less than or equal to category j between those with graduate degrees and bachelor degrees (DEGREE=4) is $\exp(0.685 - 0.548) = 1.147$. And to reiterate, this holds for $j = 1, \ldots, 5$. Thus, even though predictive accuracy is not practical, the proportional odds model with *exclusively* qualitative predictors simplifies the interpretation of the contingency table by means of the *proportional odds assumption*.

Moving back to the proportional odds model with continuous predictors[5], it will now be interesting to look at the predictive accuracy of four different models: the proportional odds model generated from the proportional odds and multinomial logit models and the

[5]Note: The results here hold for any mixture of continuous and qualitative predictors.

Table 5: Proportional odds assumption likelihood-ratio test p-values

$\mathbf{x}$	**p**-values	
	prop. odds	*mult. logit*
$\mathbf{x}_1$	0.2403	0.0044
$\mathbf{x}_2$	0.4861	0.0051
$\mathbf{x}_3$	0.1455	0.2108

Table 6: Two-fold cross-validated average predictive accuracy proportions for $\mathbf{y}_{m3}$ vs. $\mathbf{x}_3$ and $\mathbf{y}_{p1}$ vs. $\mathbf{x}_1$ ($n = 400$)

Generated from	Fit on	
	prop. odds	*mult. logit*
prop. odds	**0.648**	0.639
mult. logit	**0.763**	0.754

multinomial logit model generated from the multinomial logit and proportional odds models. This was done using the `polr2cv` function that I created. The results from this can be seen in table 5. I chose $\mathbf{y}_{m3}$ vs. $\mathbf{x}_3$ and $\mathbf{y}_{p1}$ vs. $\mathbf{x}_1$ for the comparison, because as can be seen in table 5, these both do not violate the proportional odds assumption. However, in both cases the model that the data was generated from had the greater predictive accuracy when the two-fold cross-validation was done fitting the proportional odds model[6] (this result is seen in green in table 6).

This result is not surprising. One would expect that if the proportional odds assumption holds, then the proportional odds model would fit the data better (and predict more accurately). But what of the case where the proportional odds assumption does not hold?

In the first case I only looked at values for which between the response variable and predictor the proportional odds assumption failed to be rejected. This likely has to do with the fact that correlations between $\mathbf{y}_{m3}$ vs. $\mathbf{x}_3$ and $\mathbf{y}_{p1}$ vs. $\mathbf{x}_1$ have reasonably high correlations, being 0.540 and 0.619 respectively. So for this trial, I chose $\mathbf{y}_{m1}$ vs. $\mathbf{x}_1$ and with some trial and error, generated a set of response counts from the proportional odds model which did not satisfy the proportional odds assumption. Between these values named `ypbad` and $\mathbf{x}_1$ there was a correlation -0.0583 and a p-value for the likelihood-ratio test of the proportional odds assumption of approximately 0. In this part, the results one would expect hold, namely that the model which the cross-validation was fit on predicted better

[6] A note about table 5. This table gives the p-values for the test statistics only of the response counts which are similarly indexed to the x-values. That is, $\mathbf{y}_{pj}$ fit on $\mathbf{x}_j$.

Table 7: Two-fold cross-validated average predictive accuracy proportions for $\mathbf{y}_{m1}$ vs. $\mathbf{x}_1$ and `ypbad` vs. $\mathbf{x}_1$ ($n = 400$)

	Fit on	
Generated from	*prop. odds*	*mult. logit*
prop. odds	**0.850**	0.845
mult. logit	0.636	**0.654**

when it was generated from the same model. This can be seen in table 7 (again, the green values).

The most important conclusion that can be drawn from this is that if the proportional odds assumption holds, even if the response counts are randomly generated from a multinomial logit model, the proportional odds model predicts better. That is, the fitted proportional odds model correctly classifies a greater proportion of response categories. What this means is that in dealing with J ordinal response categories, a proportional odds model should be fit, just in case the likelihood-ratio test of the proportional odds assumption fails to be rejected.

5 Literature Review

For this project, I intended to not only see how the proportional odds model faired under simulation, but how its assumptions were met in practice. To do this, I surveyed five different articles from various fields to get a very rudimentary picture on whether or not the proportional odds assumption was either mentioned as having been tested, or done and the p-value reported[7]. I reviewed each of the five articles in turn and summarized the results. This provided me with insight on how the model is used, in all its diversity, in the real world.

The conclusions that resulted from this review are that it is not common, at least in my biased sample, having selected the articles arbitrarily, to report any sort of test of the proportional odds assumption—as only Soon (2010) did. It is much more common to report the significance of predictors in a proportional odds model as all of the articles did (as in a p-value corresponding to a t-value). In none of the articles was the likelihood-ratio test as performed earlier in this paper, done. However, it seems like the score test as mentioned in Peterson and Harrell (1990) was likely done as two of the articles use SAS for their analysis (Lu et al., 2011; Ekholm, Strandberg-Larsen, & Grønbæk, 2011), and the method for ordinal logistic regression in SAS, PROC LOGISTIC, does a score test. Soon (2010) uses the `brant`

[7] The various tests that I looked for was the score test from Peterson and Harrell (1990), the likelihood-ratio test mentioned earlier and various others from Brant (1990).

command in STATA, conducting Wald tests of model goodness-of-fit. Overall, the articles were inexplicit in their treatment of the how the models fit or satisfied the assumptions, but in every instance significance was reported for predictors in the proportional odds models fit.

The number of models I looked at was only 5 but it provides a starting point for further investigation. It is clear that this model is very useful for survey data, as 3 of the 5 articles reviewed analyzed survey data. Also, 3 of the 5 of the articles dealt with health sciences. In conclusion, there were no instances in any of the articles of the proportional odds assumption overall being rejected, but the partial proportional odds model seemed popular, with the assumption being violated for some subsets of predictors. In none of the articles where the partial proportional odds model was used (Soon, 2010; Lu et al., 2011) was the potential structural deficiency of the model noted (Agresti, 2010).

Bullying behaviors among US youth

This was a short, but influential, paper on the frequency of bullying in schools in the United States. The population sampled from were 6^{th} to 10^{th} graders in the US and the response was ordinal with bullying frequencies as response, specifically bullying and being bullied (Nansel et al., 2001). The values of the response were I HAVEN'T, ONCE OR TWICE, SOMETIMES and ONCE OR MORE PER WEEK. There were a total of five proportional odds models fit, with the full sample for one and subsamples of the sample for the rest. It was not explicit whether or not a test of the proportional odds assumption was done. Although, an indication of tests performed is seen in the following: "The overall model for each of the outcomes was significant ($P < .001$)" (Nansel et al., 2001, p. 2097). However, there is no indication which test was conducted for these models. The results were that ordinal predictors such as smoking and fighting were associated in all models with the outcomes.

The determinants of students return intentions: A partial proportional odds model

This paper dealt with international students from two New Zealand universities and their intention to return to their home country after their studies were finished (Soon, 2010). The response had 4 levels, with DEFINITELY NOT RETURN, PROBABLY NOT RETURN, PROBABLY RETURN and DEFINITELY RETURN as the values. This model primarily dealt with the partial proportional odds model. Three of the 22 predictors were shown to significantly violate the proportional odds assumption. Furthermore, a Wald chi-squared test also from Brant (1990) was used to assess the null hypothesis that all of the predictors were zero and was soundly rejected at an approximate p-value of 0.

Overall, the tests that this paper did on the model were very encouraging and rigorous; however, there was no mention of the fact that the partial proportional odds model, like the generalized form in (2) can also be structurally deficient, as in producing negative category probabilities (Agresti, 2010). Since most of the predictors are two-level factors, this may not be an issue in this case, but is still a potential concern as there are three predictors defined on the interval $[0, \infty)$. The results of this study were that many factors have a significant effect on whether or not a student returns to his/her home country, especially if there is little infrastructure for their field in their home country upon returning.

Docosahexaenoic acid supplementation decreases liver fat content in children with non-alcoholic fatty liver disease: Double-blind randomised controlled clinical trial

This paper dealt with a study of 60 consecutive children at an Italian hospital and the effects of the implementation of the procedures in the title of the paper (Nobili et al., 2011). The response for the proportional odds model was change in liver steatosis with 5 ordinal categories, ranging from -3 to $+1$, with steatosis originally categorized with 4 categories, ranging from less severe to more severe, at the outset of the study. Distributional assumption tests were mentioned as being made, by "inspecting probability plots" and checking "that there were no within-group differences in the changes over time in an ordinal model not making the proportional odds assumption" (Nobili et al., 2011, p. 351). The first method likely involved plotting probabilities from a proportional odds model from which the responses were generated from a fitted proportional odds model, against that of a generalized proportional odds model seen in (2) (where a linear trend indicates fit). This first procedure seems to be what is demonstrated by Kim (2003) and the second procedure is not entirely clear to me. Regardless, no specific statistical measure of goodness-of-fit or test of the proportional odds assumption was made in the paper. The results indicated that administration of docosahexaenoic acid was significant in change of liver steatosis in the model.

Influence of the recall period on a beverage-specific weekly drinking measure for alcohol intake

This paper concerned the affect of recall period on reported alcohol-usage in the Danish population (Ekholm et al., 2011). The ordinal response was daily alcohol intake with 3 categories taking the values NO-INTAKE, MODERATE and HIGH consumption of alcohol. A proportional odds model was fit for each day of the week with the predictor being the recall period (a factor with 7 levels, being the amount of days between initial interview and interview about alcohol-usage on said day—ranging from 1 to 7 days). No tests were reported or mentioned for this study. The results indicate that there was a significant association between recall period and reported alcohol-usage, but that beverage-specific questions (another predictor) were not significant.

Pain in long-term adult survivors of childhood cancers and their siblings: A report from the Childhood Cancer Survivor Study

This paper takes results from the Childhood Cancer Survivor Study taken from childhood cancer survivors in North America (Lu et al., 2011). The response for the partial proportional odds model was pain experienced by the survivors (as an adult) on an ordinal scale with 5 levels taking values NO PAIN, SMALL AMOUNT OF PAIN, MEDIUM AMOUNT OF PAIN, A LOT OF PAIN and VERY BAD EXCRUCIATING PAIN. Siblings of the survivors were used as a control group and results of the model showed that childhood cancer survivors exhibited greater pain. As far as tests of assumptions go, three factors were supposed to violate the proportional odds assumptions, but specifics were not mentioned as to how this was tested, although analysis took place in SAS.

6 Moving Forward

Thus far, I have checked the distributional assumption of the p-values of the likelihood-ratio test statistic of the proportional odds assumption and shown it to approximately follow a standard uniform distribution when the settings of the parameters are such that all of the response categories are regularly sufficiently large and the model can regularly be fit. I have also generated various realizations of response variables from the multinomial logit and proportional odds models and checked the predictive accuracies of these generated responses based on violation of the proportional odds assumption and shown if a likelihood-ratio test of the proportional odds assumption fails to be rejected for responses generated from either of these two models, then fitting a proportional odds model will provide one with a higher proportion of correctly classified response categories. I have also done a review of literature containing data analyses with the proportional odds model included and seen that oftentimes a likelihood-ratio test of the proportional odds assumption, or any equivalent test, is not commonly done or made explicitly known.

More broadly, the conclusion I have to make is that the proportional odds model should be considered as a first course of action when one is dealing with any ordinal response data[8] that one thinks is correlated with the predictor(s). Similarly, this should be done with contingency tables where one of the variables is ordered, or qualitative predictors, because the proportional odds model simplifies the interpretation of such an ordinal categorical response. In addition, a likelihood-ratio test should be done to see if the proportional odds assumption holds, because a higher predictive accuracy may be to gain from failing to reject it and the converse if it is rejected.

From here, there are a few things to look at. The first is the partial proportional odds model from Peterson and Harrell (1990). The model specification can be seen in (4) and it was used in two articles in the literature review. However, considering this for continuous predictors presents the same structural problems as the model in (2). So this not practical under normal circumstances. But looking at this for qualitative or ordered score predictors may be worthwhile.

With regards to other assumptions of the proportional odds model, it would interesting to check the distributional assumptions of the other tests that I have mentioned thus far in the paper. It would also be interesting to look at link tests[9], especially programming them into R, and implementing a procedure in R to do a sort of backwards elimination for parameters to determine which ones satisfy the proportional odds assumption.

7 Appendix

You can find the R code for all the functions in the document here and a .csv file with all the randomly generated data in the document here. The partial proportional odds model from Peterson and Harrell (1990) can be seen in the model

$$\text{logit}[P(Y \leq j)] = \alpha_j - \boldsymbol{\beta}'\mathbf{x} - \boldsymbol{\gamma_j}'\mathbf{u}, \quad j = 1, \ldots, J-1, \tag{4}$$

where the predictors are partitioned into sets $\mathbf{x}$ of the predictors with the proportional odds assumption and $\mathbf{u}$ those without the proportional odds assumption (Peterson & Harrell, 1990; Agresti, 2010). In the case that $\forall j,\ \boldsymbol{\gamma_j} = 0$, the model is the same as in (1).

[8] Or, if one does not care to use the canonical logit link function as I have during the course of this investigation, one could alternatively use the probit or log-log link functions that also carry with them a sort of proportionality assumption (Agresti, 2010).

[9] As mentioned earlier and in the footnote above, there are alternative link functions and there are tests mentioned by Agresti (2010) that determine the best one from a continuum

Acknowledgments

I would like to thank professor Adam J. Rothman in the School of Statistics at the University of Minnesota for being my faculty mentor for this research project.

References

Agresti, A. (2010). *Analysis of ordinal categorical data* (2nd ed.). Hoboken, NJ: Wiley.

Brant, R. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*, *46*(4), 1171–1178.

Ekholm, O., Strandberg-Larsen, K., & Grønbæk, M. (2011). Influence of the recall period on a beverage-specific weekly drinking measure for alcohol intake. *European Journal of Clinical Nutrition*, *65*, 520–525.

Kim, J.-H. (2003). Assessing practical significance of the proportional odds assumption. *Statistics & Probability Letters*, *65*(3), 233–239.

Lu, Q., Krull, K. R., Leisenring, W., Owend, J. E., Kawashima, T., Tsao, J. C., ... Zeltzer, L. K. (2011). Pain in long-term adult survivors of childhood cancers and their siblings: A report from the childhood cancer survivor study. *Pain*, *152*, 2616–2624.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society. Series B (Methodological)*, *42*(2), 109–142.

Nansel, T. R., Overpeck, M., Pilla, R. S., Ruan, W. J., Simons-Morton, B., & Scheidt, P. (2001). Bully behaviors among us youth: Prevalence and association with psychological adjustment. *Journal of the American Medical Association*, *285*(16), 2094–2100.

Nobili, V., Bedogni, G., Alisi, A., Pietrobattista, A., Risé, P., Galli, C., & Agostoni, C. (2011). Docosahexaenoic acid supplementation decreases liver fat content in children with non-alcoholic fatty liver disease: double-blind randomised controlled clinical trial. *Archives of Disease in Childhood*, *96*, 350–353.

Peterson, B., & Harrell, F. E. (1990). Partial proportional odds models for ordinal response variables. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *39*(2), 205–217.

Soon, J.-J. (2010). The determinants of students return intentions: A partial proportional odds model. *Journal of Choice Modelling*, *3*(2), 89–112.

R Code

```
#LIKELIHOOD RATIO TEST
linear=function(n=100, xp=0.5)
{
x=sort(rnorm(n))
y=x; x=sample(x)
predictor=xp*x+(1-xp)*y
return(predictor)
}
x1=linear(xp=0)
x=x1
ybad=polrgen(x, b=0.4508689, a1=-0.8772445, a2=-0.804251,
random=FALSE, check=TRUE, n=1000)
poor=ybad$p[is.na(ybad$p)==FALSE]
ggplot()+geom_histogram(aes(poor), colour="black",
fill="#FF3333", binwidth=0.1)+xlab("p-values")+xlim(0,1)
ygood=polrgen(x, b=0.9253194, a1=-2.065255, a2=2.348769,
random=FALSE, check=TRUE, n=1000)
ggplot()+geom_histogram(aes(ygood$p), colour="white",
fill="#33AD33", binwidth=0.1)+xlab("p-values")+xlim(0,1)

#MODEL FITTING
x2=linear(xp=0.5)
x3=linear(xp=1)
yp1=polrgen(x1)
yp2=polrgen(x2)
yp3=polrgen(x3)
ym1=multgen(x1)
ym2=multgen(x2)
ym3=multgen(x3)
yms=data.frame(cbind(ym1, ym2, ym3))
yps=data.frame(cbind(yp1, yp2, yp3))
xs=data.frame(cbind(x1, x2, x3))
cor(xs, yps); cor(xs, yms)
```

```
cor(yp1, x1); cor(yp2, x2); cor(yp3, x3)
cor(ym1, x1); cor(ym2, x2); cor(ym3, x3)


#PREDICTIVE ACCURACY
yp1=ordered(yp1); yp2=ordered(yp2); yp3=ordered(yp3)
ym1=ordered(ym1); ym2=ordered(ym2); ym3=ordered(ym3)
po1=checkpo(x1, yp1)
po2=checkpo(x2, yp2)
po3=checkpo(x3, yp3)
pm1=checkpo(x1, ym1)
pm2=checkpo(x2, ym2)
pm3=checkpo(x3, ym3)
matrix(c(po1, po2, po3, pm1, pm2, pm3), nrow=3)
pomu=polr2cv(x1, yp1, method="multi")
popo=polr2cv(x1, yp1)
mupo=polr2cv(x3, ym3, method="multi")
mumu=polr2cv(x3, ym3)
mean(pomu$acc); mean(popo$acc);
mean(mupo$acc); mean(mumu$acc)
ypbad=polrgen(x1)
pomubad=polr2cv(x1, ypbad, method="multi")
popobad=polr2cv(x1, ypbad)
mupobad=polr2cv(x1, ym1, method="multi")
mumubad=polr2cv(x1, ym1)
mean(pomubad$acc); mean(popobad$acc)
mean(mupobad$acc); mean(mumubad$acc)

#FUNCTIONS
polr2cv=function(x, y, n=200,
method=c("logistic", "probit", "cloglog", "multi"), flevels=levels(y))
{
if (missing(method)==TRUE)
  {
```

```
method="logistic"
}
# THIS FUNCTION DOES TWO-WAY CROSS-VALIDATION n TIMES
# TO ASSESS PREDICTIVE ACCURACY BETWEEN TWO SUBSETS OF EQUAL SIZE
# OF A PROPORTIONAL ODDS MODEL OR A MULTINOMIAL LOGIT MODEL.
# THIS DOES THIS FOR A FACTORED RESPONSE y
# AND A SINGLE PREDICTOR x. THE LEVELS OF THE FACTOR y
# CAN OPTIONALLY BE SPECIFIED,
# IF IT DOES NOT CORRESPOND WITH DEFAULT.
require(MASS); require(nnet)
# SIMPLE FUNCTION THAT COMPUTES
# THE LENGTH OF THE LEVELS OF THE ARGUMENT IT TAKES
ll=function(x) return(length(levels(factor(x))))
predictive_acc=NULL; counts=NULL
#CHECKS TO SEE IF LENGTH OF X AND Y ARE EQUAL
if (length(x)!=length(y))
{
  stop("lengths of arguments are not equal")
}
# FACTORS AN UNFACTORED ARGUMENT
if (is.factor(y)==FALSE)
{
  flevels=levels(factor(y))
}
lex=length(x)
# DOES TWO-WAY CROSS-VALIDATION ON THE RESPECTIVE DATA SUBSETS
for (i in 1:n)
{
  k=(2*i-1); j=2*i
  checkdex=NULL
  checknot=NULL
  #THIS SETS TO ZERO THE LENGTH OF THE LEVELS
  # OF YDEX (LYD) AND YNOT (LYN)
  ylevels=ll(y); lyd=0; lyn=0; count=0
```

```
while (lyd<ylevels | lyn<ylevels)
{
  dex=sample(lex, lex/2)
  xdex=x[dex]; ydex=y[dex]
  xnot=x[-dex]; ynot=y[-dex]
  lyd=ll(ydex); lyn=ll(ynot)
  count=count+1
  if (count > 10)
  {
    break
  }
}
counts[i]=count
if (method=="multi")
{
  # FACTORS Y-VALUES
  ydex=factor(ydex, ordered=FALSE)
  ynot=factor(ynot, ordered=FALSE)
  # FITS MULTINOMIAL LOGIT MODELS
  mod_dex=multinom(ydex~xdex)
  mod_not=multinom(ynot~xnot)
  # PREDICTS CATEGORIES FOR OTHER HALF OF
  # XVALUES NOT USED IN RESPECTIVE MODELS
  checkdex=predict(mod_dex, newdata=data.frame(xdex=xnot))
  checknot=predict(mod_not, newdata=data.frame(xnot=xdex))
  # FACTORS PREDICTED Y VALUES
  checkdex=factor(checkdex, levels=flevels)
  checknot=factor(checknot, levels=flevels)
}
else
{
  # FACTORS Y-VALUES
  ydex=ordered(ydex)
  ynot=ordered(ynot)
```

```
      # FITS CUMULATIVE LOGIT (PROPORTIONAL ODDS) MODELS
      mod_dex=polr(ydex~xdex, method=method)
      mod_not=polr(ynot~xnot, method=method)
      # PREDICTS CATEGORIES FOR OTHER HALF OF
      # XVALUES NOT USED IN RESPECTIVE MODELS
      checkdex=predict(mod_dex, newdata=data.frame(xdex=xnot))
      checknot=predict(mod_not, newdata=data.frame(xnot=xdex))
      # FACTORS PREDICTED Y VALUES
      checkdex=ordered(checkdex, levels=flevels)
      checknot=ordered(checknot, levels=flevels)
    }
    # RETURNS THE PREDICTIVE ACCURACY
    # (PERCENT OF CATEGORIES CORRECT)
    # AND RETURNS A VECTOR OF THESE 2n VALUES
    predictive_acc[k]=mean(checkdex==ynot)
    predictive_acc[j]=mean(checknot==ydex)
  }
  list("acc"=predictive_acc, "counts"=counts)
}

#CHECKS PROPORTIONAL ODDS ASSUMPTION
checkpo=function(x, y)
{
  pol=NULL
  y=ordered(y)
  agresti=clm(y~x)
  nogrest=clm(y~1, nominal=~x)
  pol=anova(agresti,nogrest)$Pr[2]
  return(pol)
}

# THIS FUNCTION GENERATES A 3-LEVEL CATEGORICAL VARIABLE
# FROM THE PROPORTIONAL ODDS MODEL ACCORDING TO SOME VECTOR OF X
# VALUES, EITHER USING SPECIFIED PARAMETER VALUES
```

```
# OR RANDOMLY GENERATING FROM AN (ARBITRARILY CHOSEN) STANDARD LOGISTIC
# DISTRIBUTION AND ORDERING THEM.
# NOTE ONLY THIS IS ONLY FOR A MODEL WITH A SINGLE PARAMETER.
polrgen=function(x, b, a1, a2, random=TRUE, ...)
{
  if (is.element("check", names(list(...))))
  {
    check=list(...)$check
    if (is.element("n", names(list(...))))
    {
      n=list(...)$n
    }
    else
    {
      n=200
    }
  }
  p.val=NULL
  require(MASS)
  if (random==TRUE)
  {
    pine=rlogis(3)
    cone=sample(pine, 2)
    a2=max(cone); a1=min(cone)
    b=setdiff(pine, cone)
  }
  proresp=NULL
  ypro=NULL
  generate=function(proresp, x, b, a1, a2)
  {
    for (i in 1:length(x))
    {
      ypro[i]=rlogis(1, location=x[i]*b)
      if (ypro[i] <= a1)
```

```
      {
        proresp[i]=1
      }
      else if (ypro[i] > a1 && ypro[i] <= a2)
      {
        proresp[i]=2
      }
      else
      {
        proresp[i]=3
      }
    }
    return(proresp)
  }
  if (exists("check")==TRUE && check==TRUE)
  {
    for (j in 1:n)
    {
      y=generate(proresp, x, b, a1, a2)
      p.val[j]=checkpo(x, y)
    }
    coef=list("A1: "=a1, "A2: "=a2, "B: "=b)
    list("coef"=coef, "p"=p.val)
  }
  else
  {
    y=generate(proresp, x, b, a1, a2)
    cat("A1: ", a1, "\n", "A2: ", a2, "\n", "B: ", b, "\n")
    return(y)
  }
}

# THIS FUNCTION GENERATES A 3-LEVEL CATEGORICAL RESPONSE VARIABLE
# FROM THE MULTINOMIAL LOGIT MODEL ACCORDING TO SOME VECTOR
```

```
# X VALUES, EITHER USING SPECIFIED PARAMETER VALUES
# OR RANDOMLY GENERATING FROM AN (ARBITRARILY CHOSEN) STANDARD NORMAL
# DISTRIBUTION AND ORDERING THEM.
# NOTE ONLY THIS IS ONLY FOR A MODEL WITH A SINGLE PARAMETER.
multgen=function(x, a2, a3, b2, b3, random=TRUE)
{
  require(nnet)
  if (random==TRUE)
  {
    pine=rnorm(4)
    b2=pine[1]; b3=pine[2]; a2=pine[3]; a3=pine[4]
  }
  denom_i=NULL
  mulresp=NULL
  for (i in 1:length(x))
  {
    denom_i[i]=1+exp(a2+x[i]*b2)+exp(a3+x[i]*b3)
    prob=c(1/(denom_i[i]), exp(a2+x[i]*b2), exp(a3+x[i]*b3))
    jhuh=rmultinom(1, 1, prob=prob)
    for (j in 1:length(jhuh))
    {
      # SINCE THE RANDOM MULTINOMIAL ONLY
      # GENERATES A SINGLE VALUE IN ANY OF THE THREE CATEGORIES
      # THE BELOW CODE DETERMINES WHICH ONE OF THE J IT IS IN.
      if (jhuh[j]>0)
      {
        mulresp[i]=j
      }
    }
  }
  cat("A2: ", a2, "\n", "A3: ", a3, "\n", "B2: ",
  b2, "\n", "B3: ", b3, "\n")
  return(mulresp)
}
```