

The Proportional Odds Assumption for Ordinal Logistic Regression

Ian Dinwoodie

May 4th, 2022

1 Introduction

Generalized linear models (GLMs) provide a unified approach to developing models for arbitrarily distributed response variables with an arbitrary link function that varies linearly with its predictors. Two such models we have examined thus far are binary and multinomial logistic regression; both of which are used for modeling nominal outcomes. However, if we wish to model ordinal outcomes, then choosing a model that accounts for the natural ordering of the response—such as ordinal logistic regression—can help avoid an unnecessary loss in statistical power (Agresti, 2013). When considering the employment of an ordinal logistic regression, one oft-overlooked requirement is satisfaction of the proportional odds assumption. Fullerton & Xu (2012) claimed that “[t]he proportional odds assumption in ordered logit models is a restrictive assumption that is often violated in practice.” During the peer review cycles for my most recent publication, Dinwoodie et al. (2021), it was determined that the ordered logit model I had employed was in violation of the proportional odds assumption. Due to the time constraints for revisions in the review cycle I employed a less restrictive model without taking the time to develop a comprehensive understanding of the proportional odds assumption I had previously violated.

Therefore, I use paper to explore the proportional odds assumption required for ordinal logistic regression; namely, what the assumption is, why it is necessary, and how it can be evaluated. This paper begins with a discussion of the theory underlying ordinal logistic regression to arrive at the necessity of the proportional odds assumption from a mathematical perspective. We then put the theory to test with simulated examples of proportional and non-proportional odds and discuss further handling of the later scenario. Lastly, we conclude with a brief discussion of challenges, limitations, and topics for further consideration.

2 Background

2.1 Ordinal Data

Ordinal data is a categorical data type that has an order by which the data can be sorted, but the relative degree of difference between categories cannot be assumed equal (Stevens, 1946). Therefore, the ordinal scale only allows for less-than, equal-to, or greater-than comparisons. A ubiquitous example of ordinal data is the use of letter grades for coursework.

2.2 Ordinal Logistic Regression

Ordinal logistic regression is a statistical method used model the relationship between an ordinal response variable given one or more independent explanatory variables. Popular functions used to perform ordinal logistic regression in R (i.e., `MASS::polr`, `VGAM::vglm`, `ordinal::clm`) generate an ordered logit model; a cumulative link model with a logit link (Christensen, n.d.).

2.2.1 Model Paramterization

Let Y be an outcome with J categories such that $P(Y \leq j)$ is the cumulative probability of Y less than or equal to a specific category j . The odds of an outcome equal to or less than a particular category j can be defined as

$$\frac{P(Y \leq j)}{P(Y > j)}, \quad j = 1, \dots, J-1.$$

The logit, also known as log odds, of an outcome equal to or less than a particular category j can be defined as

$$\text{logit}(P(Y \leq j)) = \ln\left(\frac{P(Y \leq j)}{P(Y > j)}\right), \quad j = 1, \dots, J-1.$$

Due to the fact that the outcome is both categorical and ordered, we divide the probability space into J contiguous segments. Let θ_j denote the offset for each category. We use the definition of the logistic function to define the probability for an outcome in the lowest category of j (i.e., $j = 1$) when there are p predictor variables

$$P(Y = 1) = \frac{1}{1 + e^{-(\theta_1 + \beta_1 x_1 + \dots + \beta_p x_p)}}.$$

We can use this probability equation to derive the associated odds

$$\begin{aligned} \frac{P(Y = 1)}{P(Y > 1)} &= \frac{P(Y = 1)}{1 - P(Y = 1)} = \frac{\frac{1}{1 + e^{-(\theta_1 + \beta_1 x_1 + \dots + \beta_p x_p)}}}{1 - \frac{1}{1 + e^{-(\theta_1 + \beta_1 x_1 + \dots + \beta_p x_p)}}} \\ &= \frac{\frac{1}{1 + e^{-(\theta_1 + \beta_1 x_1 + \dots + \beta_p x_p)}}}{\frac{e^{-(\theta_1 + \beta_1 x_1 + \dots + \beta_p x_p)}}{1 + e^{-(\theta_1 + \beta_1 x_1 + \dots + \beta_p x_p)}}} \\ &= \frac{1}{e^{-(\theta_1 + \beta_1 x_1 + \dots + \beta_p x_p)}} \\ &= e^{\theta_1 + \beta_1 x_1 + \dots + \beta_p x_p} \end{aligned}$$

We take the log of the odds to get the logit for the outcome being in the lowest category of j

$$\text{logit}\left(\frac{P(Y = 1)}{P(Y > 1)}\right) = \ln(e^{\theta_1 + \beta_1 x_1 + \dots + \beta_p x_p}) = \theta_1 + \beta_1 x_1 + \dots + \beta_p x_p.$$

If we similarly derive the logit for the outcome being in the highest category of j we get

$$\text{logit}\left(\frac{P(Y \leq J-1)}{P(Y = J)}\right) = \theta_{J-1} + \beta_1 x_1 + \dots + \beta_p x_p.$$

We can generalize these last two derivations to arrive at the parameterization for the ordered logit model

$$\text{logit}\left(\frac{P(Y \leq j)}{P(Y > j)}\right) = \theta_j + \beta_1 x_1 + \dots + \beta_p x_p, \quad j = 1, \dots, J-1.$$

It is important to note that this is not the only parameterization of the model. Parry (2020) states that the following alternative parameterization incorporates a negative sign to allow a direct correspondence between the slope and the ranking:

$$\text{logit}\left(\frac{P(Y \leq j)}{P(Y > j)}\right) = \theta_j - (\beta_1 x_1 + \dots + \beta_p x_p), \quad j = 1, \dots, J-1.$$

According to Parry (2020) this is the parameterization model that is used by the `MASS::polr` function which we utilize later in this paper.

Regardless of the specific parameterization, it is clear from these models that the impact of a unit change in x_p on the odds of Y being in a higher ordinal category is β_p , regardless of what category of j we are looking at. Therefore, we have a single set of coefficients to explain the effect of x on Y throughout the ordinal scale (McNulty, 2021).

2.2.2 Proportional Odds Assumption

The proportional odds assumption is an immediate consequence of the single set of non-intercept coefficients that apply to all response levels of the ordered logit model. If we consider θ_j to be the intercept and $\beta_1x_1 + \dots + \beta_px_p$ to be the slope, then the ordered logit requires that the slope remain constant across $j = 1, \dots, J - 1$. If you were to plot the cumulative logits for each category, then you would expect to have a plot with $J - 1$ parallel lines; each with an offset corresponding to the respective θ_j . For this reason, the proportional odds assumption is also known as the parallel lines assumption. Violation of the proportional odds assumption for as few as one response level results in improperly fitted θ and β coefficients. We illustrate this effect in the sections that follow.

3 Evaluating the Proportional Odds Assumption

3.1 Methods

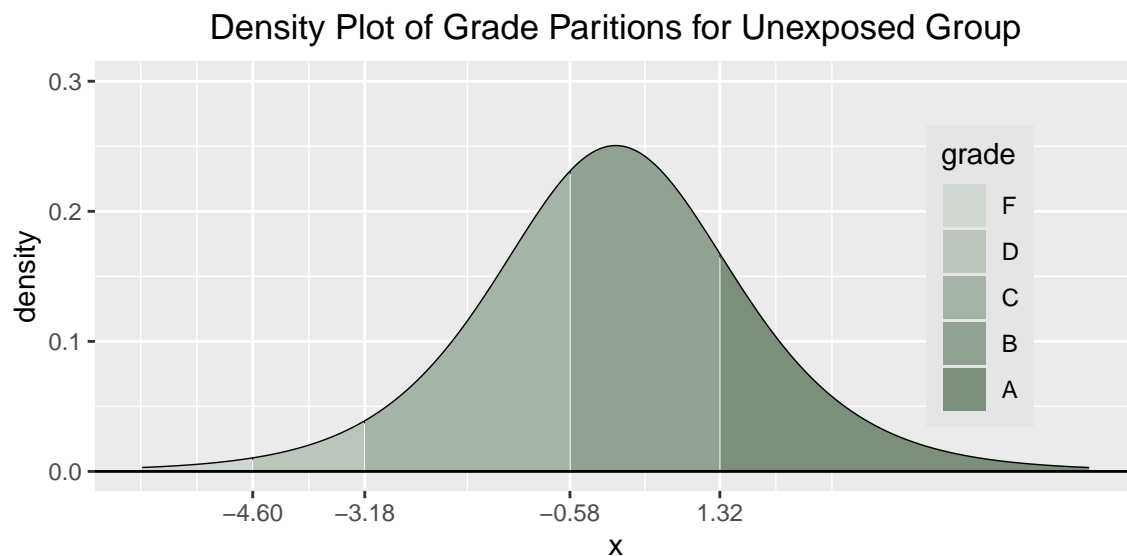
Evaluation of the proportional odds assumption in this paper was conducted using the R statistical language (R Core Team, 2022). Ordinal data sets, with and without the assumption of proportional odds, were simulated using functions and code snippets provided by the `simstudy` package (Goldfeld & Wujciak-Jens, 2020). Simulated data sets were fitted to ordinal logistic regression models using the `MASS` package (Ripley, 2022). The proportional odds assumption was verified both visually and quantitatively. Quantitative verification was performed using the `car` package (Fox & Weisberg, 2019). Plots were generated using the `ggplot2` package (Wickham, 2016). The source code for this paper has been made available at github.com/iandinwoodie/proportional-odds-assumption.

3.2 Simulation of Proportional Odds

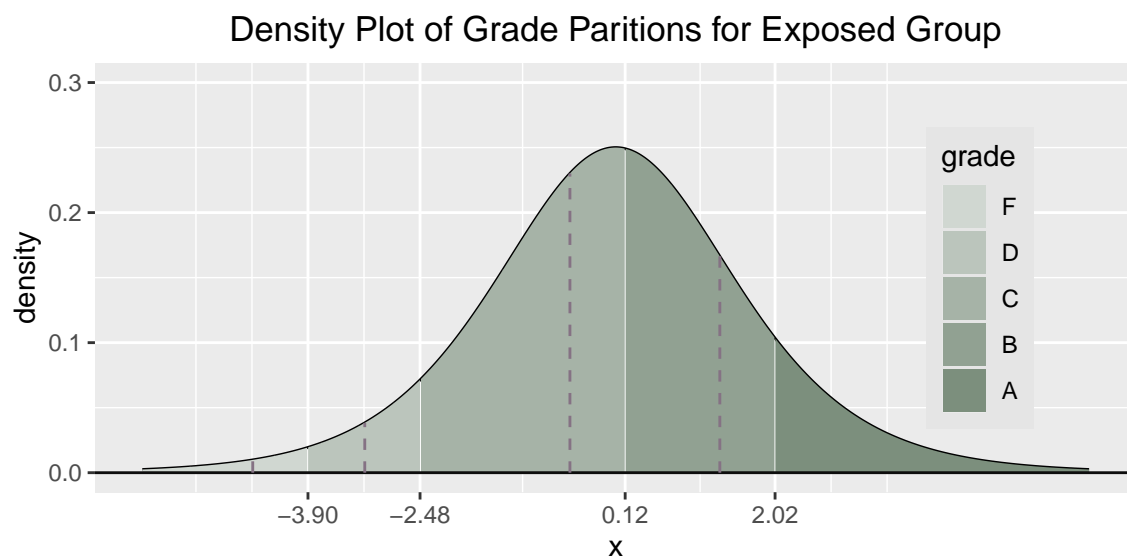
Let's start by simulating a data set with a coursework letter grade response that adheres to the following distribution.

Grade	Probability
A	0.21
B	0.43
C	0.32
D	0.03
F	0.01

We can partition a logistic distribution such that each segment corresponds to a letter grade and the area under the curve for each segment represents the probability of earning that grade. This logistic distribution is the latent process that serves as the basis for our data generation. We will assign the data to either an unexposed (i.e., control) or exposed (i.e., experimental) group for use as our predictor. The plot below illustrates this partitioning scheme for the unexposed group.



We simulate some effect of the exposed group by uniformly shifting the partitioning thresholds by a factor of $+0.7$. This adjustment represents $\beta_{exposed}$ in the model that will be produced. The plot below illustrates this partitioning scheme for the exposed group where the thresholds for the unexposed group are indicated by dashed lines.

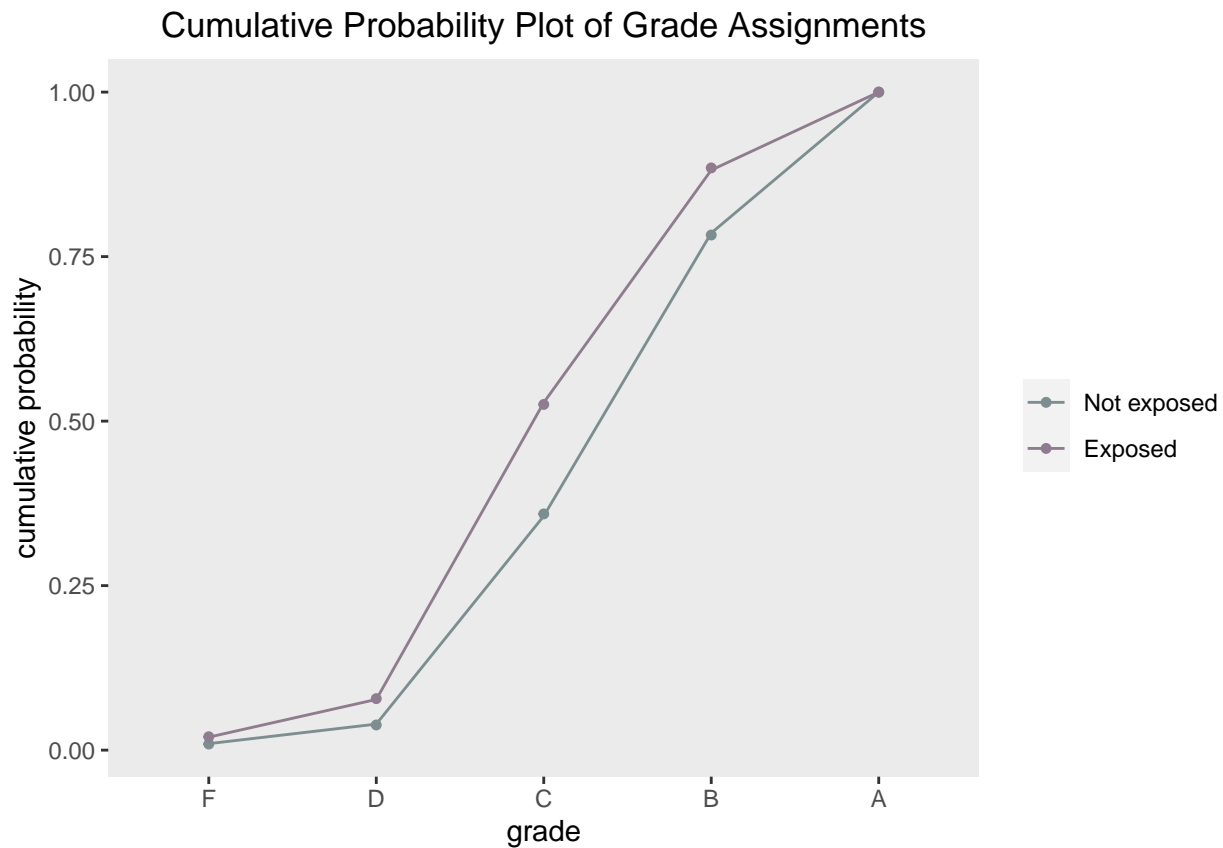


Now we fit the generated data with an ordered logit model and inspect the results. We expect to see this model estimate the parameters we used to prescribe the simulated data.

```
## Call:
## MASS::polr(formula = grade ~ exposed, data = dx, Hess = TRUE)
##
## Coefficients:
##           Value Std. Error t value
## exposed -0.7066   0.02381  -29.67
##
## Intercepts:
##      Value      Std. Error t value
## 1|2  -4.6135    0.0545  -84.6434
```

```
## 2|3    -3.1903    0.0307   -103.8934
## 3|4    -0.5938    0.0178   -33.3754
## 4|5     1.2997    0.0198    65.7407
##
## Residual Deviance: 60679.09
## AIC: 60689.09
```

We see that the coefficient closely aligns with the specified value of 0.7 and that the sign is negated due to the parameterization used by the `MASS::polr` implementation as previously discussed. Additionally, we see the intercepts closely align with the thresholds marked on the density plot for the unexposed group. We begin our evaluation of the proportional odds assumption for this model visually via a cumulative probability plot.



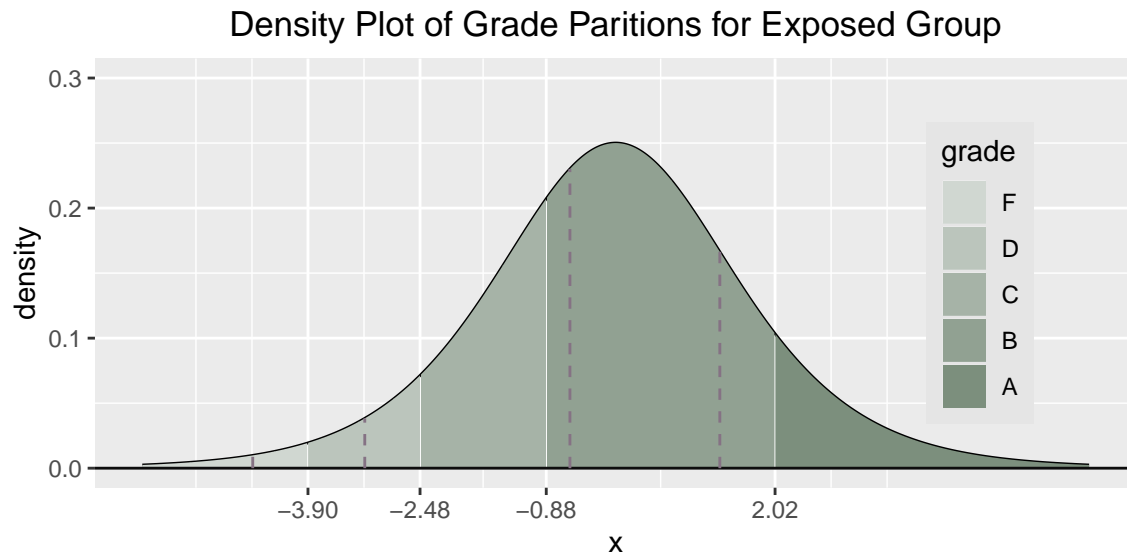
In the cumulative probability plot we see that the modeled probabilities (i.e., the lines) provide a good estimation of the probabilities we specified in the simulated data (i.e., the points). Next we perform a quantitative assessment of the proportional odds assumption using the Brant test.

```
##
## Tests for Proportional Odds
## MASS::polr(formula = grade ~ exposed, data = dx, Hess = TRUE)
##
##          b[polr]  b[>1]  b[>2]  b[>3]  b[>4] Chisquare df Pr(>Chisq)
## Overall                6.38  3    0.095 .
## exposed -0.707 -0.804 -0.758 -0.681 -0.757    6.38  3    0.095 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Our null hypothesis (H_0) is that the proportional odds assumption holds. Due to the fact that the overall and individual p-values are above a significance level of 0.05, we fail to reject the null hypothesis. We have now visually and quantitatively verified that the proportional odds assumption holds true and that an ordered logit model is an appropriate model for this data.

3.3 Simulation of Non-proportional Odds

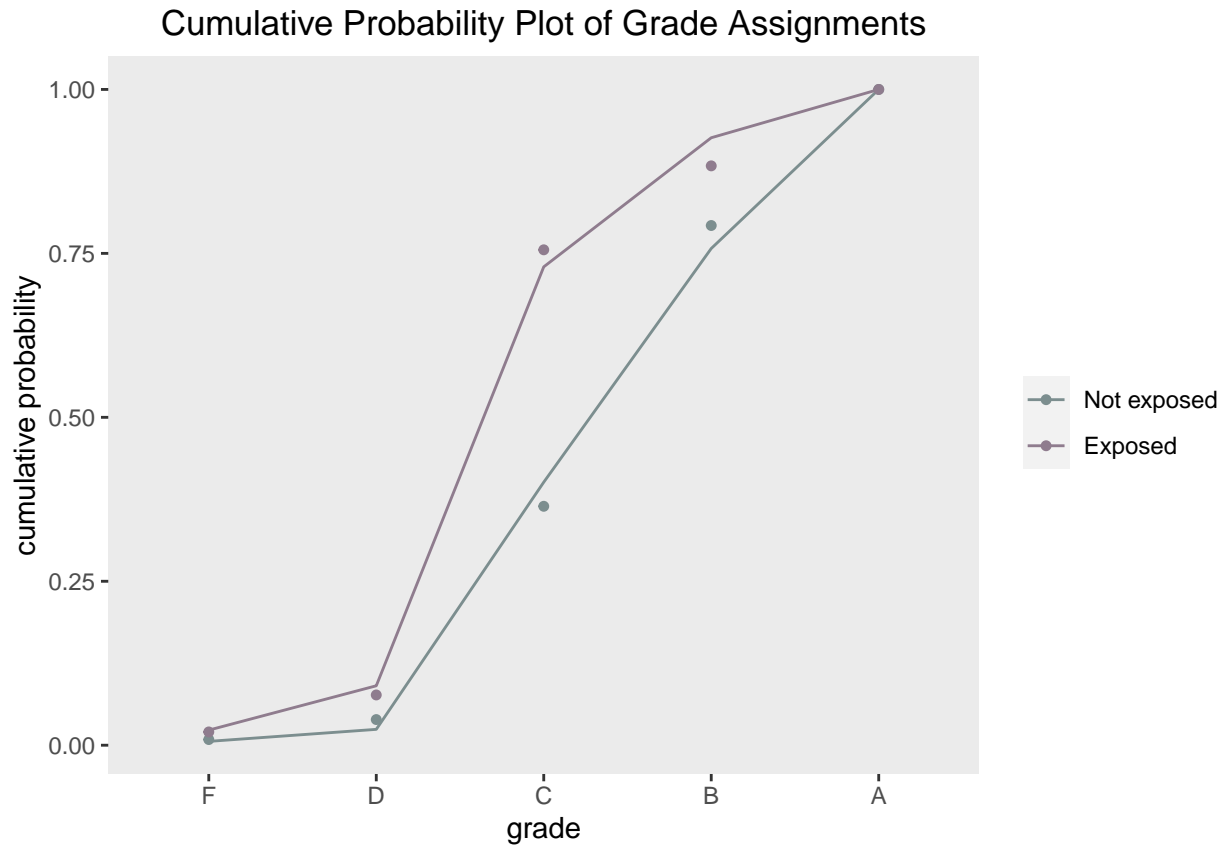
To introduce non-proportional odds we perturb the data of the exposed group by applying an additional adjustment to a single threshold. This shift to an individual threshold breaks the uniformity of the +0.7 shift that had already been applied. We shift the threshold between the third and fourth categories by an additional -1.0 and generate a density plot to observe the effect. The thresholds for the unexposed group are indicated by dashed lines.



Note that the indicator for the third threshold has shifted from 0.12 to -1.88; breaking the uniform shift in thresholds. Now we fit the generated data with an ordered logit model and inspect the results. We expect to see disparities between this model's estimates and the parameters we used to prescribe the simulated data.

```
## Call:
## MASS::polr(formula = grade ~ exposed, data = dx, Hess = TRUE)
##
## Coefficients:
##           Value Std. Error t value
## exposed -1.393   0.02584   -53.9
##
## Intercepts:
##      Value   Std. Error t value
## 1|2  -5.1410    0.0563   -91.2626
## 2|3  -3.6985    0.0330  -112.1193
## 3|4  -0.4009    0.0174   -23.0016
## 4|5   1.1386    0.0192    59.1802
##
## Residual Deviance: 56596.83
## AIC: 56606.83
```

As expected, the coefficient for `exposed` does not align with any of the prescribed shifts and the estimation of the category thresholds are notable worse than our prior model. We begin our evaluation of the proportional odds assumption for this model visually via a cumulative probability plot.



In the cumulative probability plot we see that the modeled probabilities (i.e., the lines) provide a poor estimation of the probabilities we specified in the simulated data (i.e., the points). Next we perform a quantitative assessment of the proportional odds assumption using the Brant test.

```
##
## Tests for Proportional Odds
## MASS::polr(formula = grade ~ exposed, data = dx, Hess = TRUE)
##
##          b[polr]    b[>1]    b[>2]    b[>3]    b[>4] Chisquare df Pr(>Chisq)
## Overall                1238   3    <2e-16
## exposed   -1.393   -0.863   -0.712   -1.685   -0.685    1238   3    <2e-16
##
## Overall ***
## exposed ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Our null hypothesis (H_0) is that the proportional odds assumption holds. As expected, the test results in p-values that exceed a significance level of 0.05 leading us to reject the null hypothesis. We have now visually and quantitatively verified that the proportional odds assumption is violated and that an ordered logit model is not an appropriate model for this data.

4 Handling Non-proportional Odds

The first choice that must be made when handling the non-proportional odds case is whether or not a proportional odds model will continue to be pursued. If one opts not to pursue a proportional odds model, then they could attempt to employ another GLM such as multinomial regression. The downsides of multinomial regression include loss in statistical power and increased interpretation complexity due to having separate sets of coefficients for each response level. If it is possible to condense the problem space to a binomial response by collapsing response categories and ease of interpretation is a high priority, then binomial logistic regression can be employed. In addition to the models mentioned, there are other advanced models available that aim to address these situations (e.g., adjacent-category logistic model and continuation-ratio logistic model) that are outside the scope of this paper.

If one chooses to continue pursuing a proportional odds model, then a commonly suggested route is to reduce the number of categories by adjusting the number and width of category thresholds (i.e., “squashing” the categories). Goldfeld & Wujciak-Jens (2020) provides evidence in his vignettes that there are scenarios where this solution does not work. Christensen (2019) provides guidance in the **ordinal** package vignette for relaxing the proportional odds assumption through the use of nominal and scale effects.

5 Challenges, Limitations, and Topics for Further Discussion

The main challenge I encountered during this project was attempting to generate simulated ordinal response data by hand. Thankfully, after some searching on CRAN I came across the very helpful **simstudy** package and vignettes produced by Goldfeld & Wujciak-Jens (2020) which was designed with ordinal response simulation in mind. An additional challenge I had was choosing a package to use to perform the ordinal regression model fitting. Ultimately I decided to use the **MASS::polr** function due to the wide support for the models it generates.

In order to keep this paper manageable I did not dive into a discussion on the theory behind the Brant test which was used to quantitatively verify the proportional odds assumption. In the same vein, I did not delve into the topics of category “squashing” or relaxing the proportional odds assumption; both of which are topics that could easily lend themselves to their own paper of a similar size.

For further exploration and discussion I would like to investigate the limitations around category “squashing” and relaxing the proportional odds assumption. I think it would also be beneficial to compare such augmented models with binomial or multinomial logistic regression models to compare performance. I believe it would be of great use to produce a chart of the many ways to handle the non-proportional odds case with a clear list of pros and cons for each method.

References

- Agresti, A. (2013). *Categorical data analysis* (3rd ed). Wiley.
- Christensen, R. H. B. (n.d.). *Cumulative Link Models for Ordinal Regression with the R*. https://cran.r-project.org/web/packages/ordinal/vignettes/clm_article.pdf
- Christensen, R. H. B. (2019). *Ordinal: Regression models for ordinal data*. <https://github.com/runehaubo/ordinal>
- Dinwoodie, I. R., Zottola, V., & Dodman, N. H. (2021). An investigation into the effectiveness of various professionals and behavior modification programs, with or without medication, for the treatment of canine aggression. *Journal of Veterinary Behavior*, 43, 46–53. <https://doi.org/10.1016/j.jveb.2021.02.002>
- Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third). Sage. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>
- Fullerton, A. S., & Xu, J. (2012). The proportional odds with partial proportionality constraints model for ordinal response variables. *Social Science Research*, 41(1), 182–198. <https://doi.org/10.1016/j.ssresearch.2011.09.003>
- Goldfeld, K., & Wujciak-Jens, J. (2020). Simstudy: Illuminating research methods through data generation. *Journal of Open Source Software*, 5(54), 2763. <https://doi.org/10.21105/joss.02763>
- McNulty, K. (2021). *Peopleanalyticsdata: Data sets for keith McNulty’s handbook of regression modeling in people analytics*. <http://peopleanalytics-regression-book.org>
- Parry, S. (2020). Ordinal Logistic Regression models and Statistical Software: What You Need to Know. In *Cornell Statistical Consulting Unit*. https://cscu.cornell.edu/wp-content/uploads/91_ordlogistic.pdf
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ripley, B. (2022). *MASS: Support functions and datasets for venables and ripley’s MASS*. <http://www.stats.ox.ac.uk/pub/MASS4/>
- Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science*, 103(2684), 677–680. <https://doi.org/10.1126/science.103.2684.677>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>