Ordinal Data Analysis: the Proportional Odds Model and the Partial Proportional Odds Model

Pei-Shan Yen

May 9 2019

Abstract

Care must be taken when dealing with ordinal data because ordinal data have both qualitative and quantitative characteristics. Before the 1980s, some researchers used the quantitative characteristics to model the ordinal data by performing the regression with assigning the scores to the categories, however this method may end in classification of the variable to the adjacent category. Some utilize the qualitative characteristic of the ordinal data and apply the generalized logit model, although using this model would possibly produce unnecessary parameter estimations, making the model too complicated for use. McCullagh (1980) proposes using the proportional odds model to solve this issue. Currently, using the proportional odds model is the most common technique for analyzing ordinal data since the model is simple and easy to interpret.

keywords: proportional odds model, ordinal regression, generalized logit model

1 Introduction

Steven (1951) classifies the scale types of variables into four classes, including nominal, ordinal, interval, and ratio scales. The ordinal variable and nominal variable belong to the scope of categorical data. The categorical variable with an ordinal scale is called an ordinal variable, but a variable measured without ordering is called a nominal variable.

Ordinal data has an explicit ordering of the levels (e.g., pain status: none, mild, discomforting, intense), and it is quantitative although the absolute distance among the levels is usually unknown. Each level on its scale represents a greater or smaller magnitude of a particular characteristic than another degree. Sometimes, the ordinal scale, which is measured by the underlying continuous variables, may collapse the categories. An ordinal variable is quite different from the qualitative variable, which is independent of magnitude of a characteristic. Because the ordinal data is inherently quantitative, the models and measures of the association share many resemblances to continuous variables.

In the early 1960s, categorical data analysis was at a primitive stage of its development. At that time, numerous well-known statistical methods for categorical data focused exclusively on the nominal data and did not consider the ordinal data. The ordinal data was typically dealt with analytical methods associated with the ratio outcome or the nominal outcome.

Between 1960-1980, researchers assigned discrete scores to the categories and used regression to model the ordinal data because they treated the ordinal outcome as a ratio outcome. Agresti (2010) suggests that this method has several limitations: (1) The cut-point for the scores is usually not clear; (2) Assigning a single score does not allow for a measurement error; (3) This approach might yield overestimation or underestimation of the predicted values (this phenomenon is also called ceiling effects and floor effects); and (4) clear cut-point sometimes produce misleading results (Anderson and Philips, 1981).

On the other hand, if the ordinal outcome was treated as the nominal outcome, the generalized logit model was applied. Researchers would apply the baseline-category logit for the ”multinomial” outcome. This model is more complicated because more parameters need to be estimated. Using this model may violate the ideal of model parsimony. Furthermore, this approach may lose some benefits of the ordinal characteristics (Agresti, 2010).

Agresti (2013) mentions that the advantages of treating the ordered categorical variable as ordinal rather than nominal data include: (1) the ordinal data can use its quantitative characteristics to calculate the numeric descriptive statistics, such as mean or correlation, (2) ordinal models are more parsimonious and easier to interpret than multinomial models, and (3) the ordinal data have greater power for detecting relative trends to the null hypothesis.

In the 1980s, numerous works on the methodology of ordinal data analysis were developed. Particular attention was given to distinguishing between nominal and ordinal data. This is because the quantitative characteristic of ordinal data provides more information when performing data analysis. Most of the statistical methods for the ordinal data were developed by utilizing their quantitative characteristic. McCullagh (1980) proposed using ordinal logistic models, including the proportional odds and the proportional hazards model, to analyze ordinal data, instead of using the traditional regression models which required assigning scores to the ordinal variable. The following sections will describe the proportional odds model, as well as its advantages and limitations.

2 Latent Variable

The latent variable is common to use when dealing with the multiclass classification. Anderson and Philips (1981) discussed using the latent variable to help analyze the outcome variable Y with 2 groups including H_1 and H_2, assuming the covariates are $x = (x_1, x_2, ...x_p)$. This classification rule maximizes the probability of correct categorization and is obtained from the posterior probability $P(H_1|x)$ and $P(H_2|x)$. The method is appropriate for distinct, qualitative outcome variables but is not appropriate for the quantitative, ordinal outcome variables. The reason is that it is possible that the classification is assigned to the adjacent categories, especially when there are more than two distinct ordered groups. However, this misclassification may decrease if we use an unobserved continuous variable Y coupled with a threshold principle as a latent variable W. Using a continuous variable as the latent variable is reasonable given the quantitative characteristic of the ordinal data.

Let W be a finite threshold value which defines two groups H_1 and H_2 in population H. H_1 is observed if $Y \geq w$ and H_2 is observed if $Y \leq W$. Tallis (1975) assumes the joint distribution of Y and x is a multivariate normal distribution and Y follows a standard normal distribution. Therefore, the conditional distribution of Y given x is a normal distribution. Hence, the posterior probability is $P(H_2|x) = P(Y > W|x) = \Phi(\alpha_0+\alpha^t x)$ where the $\Phi(.)$ is the cumulative distribution of normal distribution. An individual will be allocated to group H_2 if $\alpha_0 + \alpha^t x \geq 0$, and to group H_1 otherwise. The allocation method is called the probit discriminant function as opposed to the logistic discriminant function. The posterior probability can be calculated as $P(H_2|x) = \frac{exp(\alpha_0+\alpha^t x)}{1+exp(\alpha_0+\alpha^t x)}$ when the cumulative distribution is a logistic function (Anderson 1972).

We may extend the concepts of the latent variable for the outcome variable Y with k ordered categories (Winship and Mare, 1984), as shown in Fig. 2. The k ordered categories are considered as k non-overlapping and exhaustive intervals on the continuous scale. Next, assume that Z is a continuous latent variable. Z has unknown cut points $\theta_1, \theta_2, ..., \theta_k - 1$ where $\theta_1 \leq \theta_2 \leq ... \leq \theta_{k-1}$ and $\theta_0 = -\infty$ and $\theta_k = \infty$. Using this mapping concept of the continuous interval can avoid the imposition of an arbitrary discrete scoring system for the categories. Furthermore, applying the reference scales $\{\theta_j\}$ can help to interpret the model in a straightforward and incisive manner (Plackett, 1981). Therefore, the ordinal variable is also referred to as a grouped continuous variable (Greenwood and Farewell, 1988) .

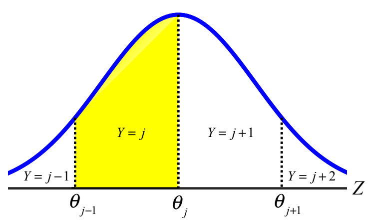

Figure 1: Latent Variable

As a result, constructing a regression model on Y given the covariate x is equivalent to the construction of the regression model for Z. The probability of the event Y_j given x is

$$P(Y = j|x) = P(\theta_{j-1} \leq Z \leq \theta_j|x) \quad (1 \leq j < k). \tag{1}$$

Now, the threshold Z model is constructed (Agresti, 2015). First, we define $Z = \alpha + \beta^t x + \epsilon$, where α is an intercept parameter which can be treated as a nuisance parameter. $\beta^t = (\beta_1, \beta_2, ...\beta_p)$ represents the slope parameters which are representative of the regression coefficients, and the error term ϵ has the cumulative distribution $\Psi(.)$ which is assumed to be a logistic function.

$P(Y \leq j|x)$ can then be expressed in the logistic form through the concepts of the latent variable. The model given by Eq. (2) is called an Ordinal Regression Model. Here it is important to note that we can use $k-1$ dummy binary variables to describe the model. This approach is, however, more complicated because it has a considerable number of parameters.

$$P(Y \leq j|x) = P(Z \leq \theta_j|x) = P(\epsilon \leq \theta_j - \alpha - \beta^t x) = \Psi_\epsilon(\theta_j - \beta^t x) = \frac{exp(\theta_j - \beta^t x)}{1 + exp(\theta_j - \beta^t x)} \tag{2}$$

3 Proportional Odds Model

Using the concept of latent variables, Eq. (2) yields the binary logistic model with unknown parameters θ_j and β (Cox, 1970):

$$logit[P(Y \leq j|x)] = \log \left[\frac{P(Y \leq j|x)}{P(Y > j|x)} \right] = \theta_j - \beta^t x \qquad (1 \leq j < k) \tag{3}$$

3.1 Proportional Odds Model (PO)

McCullagh (1980) later proposes a variant of this logistic model to analyze the ordinal data. This model can apply the ordinal nature of the ordinal data, eliminating the needs for assigning scores or assuming cardinality that might cause an overestimation or underestimation. McCullagh collapse the ordinal outcome into two categories performs a series of binary logistic models regression simultaneously and compares the difference of the cumulative logits (Williams, 2016). He would like to make the cumulative logits model as a simpler model, comparing to the generalized logit model. As a result, the objective of this model is to assume that all of the covariates have the same regression coefficients β.

McCullagh's work is contingent upon several assumptions, including (1) Y have the probabilities $\pi_1(x), \pi_2(x), ..., \pi_k(x)$,(2) The cumulative probability of the event $Y \leq j$ given x is $P(Y \leq j|x) = \gamma_j(x) = \pi_1(x)+\pi_2(x)+...+\pi_j(x)$, and (3) The cumulative odds is $\kappa_j(x) = \frac{\gamma_j(x)}{1-\gamma_j(x)}$. Given that a natural logarithm is a monotonic increasing function, $\kappa_j(x)$ also increases with j.

Next, McCullagh shows that the logistic model can, given the $k-1$ cumulative logits $\gamma_j(x)$, be recast as

$$logit[P(Y \leq j|x)] = logit[r_j(x)] = \log \left[\frac{\gamma_j(x)}{1-\gamma_j(x)} \right] = \log \left[\kappa_j(x) \right] = \theta_j - \beta^t x \tag{4}$$

McCullagh further derives one property of Eq. (5) by taking the exponential function of Eq. (4), which is

$$\kappa_j(x) = \frac{\gamma_j(x)}{1-\gamma_j(x)} = exp(\theta_j - \beta^t x) = exp(\theta_j)exp(-\beta^t x) = \kappa_j exp(-\beta^t x). \tag{5}$$

where $\theta_j = log(\kappa_j)$ increases with respect to j. The parameters $\{\theta_j\}$ can be treated as the nuisance parameters but are usually serve as the cut points of the latent variable Z. The regression coefficient β describes how the logit is related to the covariates x.

The cumulative odds ratio will satisfy

$$\frac{\kappa_j(x_1)}{\kappa_j(x_2)} = exp\{\beta^t(x_2 - x_1)\} \tag{6}$$

Finally, taking the natural logarithm function of Eq. (6) leads to Eq. (7)

$$log[\kappa_j(x_1)/\kappa_j(x_2)] = logit[\gamma_j(x_1)] - logit[\gamma_j(x_2)] = \beta^t(x_2 - x_1) \tag{7}$$

3.2 The Characteristics of Proportional Odds Model

Eq. (7) implies that the difference between the cumulative logits, namely $logit[\gamma_j(x_1)] - logit[\gamma_j(x_2)]$, is independent of the category j involved. The model assumes that the relationship between covariates x and Y does not depend on j. That is, the cumulative odds ratio is proportional to the difference between the covariate values, or $x_2 - x_1$. As a result, the model given by Eq. (7) is also called the Proportional Odds Model (PO). In some fields, this model is also called Ordered Logit Model or Proportional Odds Version of the Cumulative Logit Model (Agresti, 2010).

In other words, it is assumed that the effects β of the covariates are the same across the cumulative logits. Thus, the model provides a single estimate of β of the cumulative log odds ratio over the cut-off points (Lall et al., 2002). The simple model only need to estimate $(k-1)+p$ parameters while the generalized logit model need to estimate $(k-1)(p+1)$ parameters. Due to the ease of interpretation and the model parsimony, currently, the proportional odds model is the most popular model to deal with ordinal data.

3.3 Link Choice

The proportional odds model (Eq. (7)) is constructed using the logit link. McCullagh also demonstrates that any other monotonic increasing function which can map the interval $(0, 1)$ onto $(-\infty, \infty)$ can be regarded as a link function. Examples include the probit link or the complementary log-log link. Besides, if we have evidence to support the underlying distribution is asymmetric, we can use loglog or complementary log link. If we do not have information about the underlying distribution, the determination of a suitable link function should pertain to the ease of the model interpretation. Consequently, the logit link is usually preferred because it is not only straightforward to describe the relationship between the outcome variables and the covariates but also it is computationally more attractive (Hastie et al., 1989).

3.4 Test of Proportional Odds Assumption

The test for the proportional odds assumption is equivalent to testing whether the effects of the covariates are the same for each cumulative logit. The test is also called the parallelism test because the fitted surfaces for the cumulative logits are all parallel. In other words, the slopes of all of the covariates are equal if the proportional odds assumption holds (Stokes et al., 2012). Violating the proportional odds assumption indicates at least one of the covariates does not have proportional odds and those specific covariates need to further estimate their parameters. The null hypothesis and the alternative hypothesis of the test are

$$logit[P(Y \leq j|x)] = \theta_j - \beta^t x \quad v.s. \quad logit[P(Y \leq j|x)] = \theta_j - \beta_j^t x \quad (1 \leq j < k) \tag{8}$$

and the approximate chi-squared distribution for this test has a degree of freedom $df = p(k-2)$.

Agresti (2010) states that the Likelihood Ratio Test and the Wald Test are not preferred methods to verify the assumption of proportional odds. The two approaches are computationally costly because they require maximization of the likelihood function under the assumption of a multinomial setting. Also, it is possible to have diverged estimation when the phenomenon of separation occurs, or the sample size is too small in the dataset. Instead, the Score Test is more applicable to test the homogeneity of the proportional odds ratio because the implementation of the score statistics is only evaluating the rate of change of the likelihood. In other words, the Score Test has more power to detect the small deviations for the null hypothesis of proportionality.

The Score Test has several limitations. Allison (1999) mentions that if there is a continuous covariate, this may produce a small P-value. Brant (1990) indicates that the larger number of covariates will result in the rejection of the assumption. Besides, having an adequate sample size is the key for this test. Clogg and Shihadeh (1994) suggests that a large sample size may lead to a statistically significant P-value. Peterson and Harrell (1990) discusses that the test is too liberal when the sample size is small, which means that the P-value might be artificially too small. Stokes et al. (2012) suggests that if all cell counts in the cross-tabulations between the outcome and each covariate reach 5 counts, the sample size should be adequate.

Beyond the above chi-square tests, there are alternative methods that help to evaluate the proportional odds assumption. Harrell (2001) recommends that the graphical technique is useful to access the parallelism assumption subjectively. McCullagh (1980) suggests that the standardized residual test may be used to test the assumption of proportional odds. Brant (1990) develops a specific test of the proportional odds assumption by examining the separate fits to the underlying binary logistic models. This test evaluates whether the deviations between the observed values and the predicted values are by chance or not.

Agresti (2010) states that the proportional odds model is preferred even if it slightly violates the proportional odds assumption. If β_j for different logits are not substantially different, a proportional odds model should be selected because the model is parsimonious and easy to interpret.

3.5 Parameter Estimation

The link function is strictly stochastic ordering. Hence, if we take two subgroups with covariates x_1 and x_2, we will have

$$link[\gamma_j(x_1)] - link[\gamma_j(x_2)] = \beta^t(x_2 - x_1) = \Delta \tag{9}$$

where the difference of the links is denoted by Δ.

Therefore, all of the categories j will have the following properties:

$$\gamma_j(x_1) > \gamma_j(x_2) \quad if \quad \Delta > 0 \quad or \quad \gamma_j(x_1) < \gamma_j(x_2) \quad if \quad \Delta < 0 \tag{10}$$

Eq. (10) implies that the ordinal data are not permutation invariant. This suggests that if we arbitrarily permute the categories, either the parameter estimation or the fit will be affected. However, reversing the ordered categories is allowable when using a logit or probit link. The estimation parameter β merely changes sign and the $\{\theta_j\}$ reverse sign and order.

If the proportional odds does not have a strict stochastic ordering among the outcome, a general form can be expressed as (McCullagh, 1980):

$$link[r_{ij}(x)] = \frac{\theta_j - \beta^t x_i}{\tau_i} \qquad (1 \leq j < k) \tag{11}$$

The quantity $\beta^t x_i$ are the location parameters for the i^{th} row and τ_i are the scale parameters for the i^{th} row. That is, we have scale parameters associated with the covariates x. In order to estimate the unknown scale parameters, it is convenient to apply a constraint, for instance, $\tau_1 = 1$ or $\Sigma \log \tau_i = 0$.

If we use the logit link, then Eq. (11) can be rewritten as:

$$logit[r_{ij}(x)] = \frac{\theta_j - \beta^t x_i}{\tau_i} \tag{12}$$

We further derives one property of Eq. (13) by taking the exponential function of Eq. (12), which is

$$\kappa_{ij}(x) = \frac{\gamma_j(x)}{1-\gamma_j(x)} = exp\Big[\frac{\theta_j - \beta^t x}{\tau_i}\Big] = exp\Big(\frac{\theta_j}{\tau_i}\Big)exp\Big(\frac{\beta^t xi}{\tau_i}\Big) \tag{13}$$

The cumulative odds ratio will satisfy

$$\frac{\kappa_{1j}(x_1)}{\kappa_{2j}(x_2)} = \frac{exp\Big(\frac{\theta_j}{\tau_1}\Big)exp\Big(\frac{-\beta^t x_1}{\tau_1}\Big)}{exp\Big(\frac{\theta_j}{\tau_2}\Big)exp\Big(\frac{-\beta^t x_2}{\tau_2}\Big)} = exp\Big[\theta_j\big(\frac{1}{\tau_1} - \frac{1}{\tau_2}\big) + \beta^t\big(\frac{x_2}{\tau_2} - \frac{x_1}{\tau_1}\big)\Big] \tag{14}$$

Finally, taking the natural logarithm function of Eq. (14) leads to Eq. (15)

$$log[\kappa_{1j}(x_1)/\kappa_{2j}(x_2)] = logit[\gamma_{1j}(x_1)] - logit[\gamma_{2j}(x_2)] = \theta_j\big(\frac{1}{\tau_1} - \frac{1}{\tau_2}\big) + \beta^t\big(\frac{x_2}{\tau_2} - \frac{x_1}{\tau_1}\big) \tag{15}$$

The difference between the cumulative logits, i.e., $logit[\gamma_{1j}(x_1)] - logit[\gamma_{2j}(x_2)]$, is a function of the category j and the scale parameters τ_i. In other words, the model non-proportion odds model. In practice, a linear trend will be detected in the log-odds ratios (Peterson and Harrell, 1990). Testing the equality of the τ_i is equivalent to testing the assumption of proportional odds.

4 Partial Proportional Odds Models (PPO)

In practice, the proportional odds assumption are frequently violated (Long and Frees 2014). To deal with this issue, Peterson and Harrell (1990) propose the Partial Proportional Odds model (PPO) when the cumulative odds ratio are not identical in a subset of covariates. In general, there are two types of partial proportional odds model: the Unconstrained Partial Proportional Odds Model and the Constrained Partial Proportional Odds Model.

4.1 The Unconstrained Partial Proportional Odds Model

The Unconstrained Partial Proportional Odds Model indicates that some of the covariates have proportional odds while some have non-proportional odds. The model can be expressed as the form:

$$logit[P(Y \leq j|x)] = \theta_j - \beta^t x - \eta_{js}^t c \qquad (1 \leq j < k) \tag{16}$$

where (1) $x = (x_1, x_2, ..., x_p)$ is the covariates, (2) β is the regression coefficients associated with x, (3) c contains the q covariates $(q \leq p)$ which the proportional odds assumption either is not assumed or is to be tested, and (4) η_{js} $(s = 1, 2, ..., q)$ is the regression coefficients associated with c. The coefficients η_{js} represent deviations from proportionality.

Therefore, the model in Eq. (16) reduces to the proportional odds model if $\eta_{js} = 0$. This indicates that a simultaneous test of the proportional odds assumption for the q covariates in c is a test of the null hypothesis that $\eta_{js} = 0$. Since $\eta_{js} = 0$, the reduced model only uses $\theta_j - \beta^t x$ to estimate the cumulative odds ratio.

4.2 The Constrained Partial Proportional Odds Model

The Unconstrained Partial Proportional Odds Model is complex because it requires to estimate $(k-1) + (p-q) + (k-1)q$ parameters which are regression coefficients associated with x and c. A special case of the model called The Constrained Partial Proportional Odds Model is proposed if all of the covariate x_c have nonproportional odds. This model only requires to estimate $k+p$ parameters. Suppose the cumulative logits changes by certain patterns, such as linear trend, in the category j. This simpler model is preferred because it requires to estimate only one additional parameter. The constrained model can be rewritten as:

$$logit[P(Y \leq j|x)] = \theta_j - \beta^t x - E_j \eta_s^t c \tag{17}$$

where E_j are fixed prespecified scalars. The new parameter η_s is not dependent on the category j. The E_j constants can be used to constrain the logits to have a specified relationship among them. For example, if we specific $E_j = j - 1$, then the cumulative logits for category j is $\beta + (j-1)\eta_s$ which presents a linear trend. It makes senses to apply the constrained model only if the categories more than 2. The ideal choice of constraints should be determined using pilot data or one can choose some values which are based on prior knowledge about the possible behavior of the odds ratios. If not, the constraints can be decided by the examination of the odds ratios obtained from the unconstrained model (Lall et al., 2002).

The above model that allows the constraints E_j to vary is

$$logit[P(Y \leq j|x)] = \theta_j - \beta^t x - (diag E_j)\eta_s^t c \tag{18}$$

The most general model would permit some covariates to have proportional odds, some covariates have unconstrained non-proportional odds, and the remaining covariates have constrained non-proportional odds.

5 Conclusion and Discussion

The proportional odds model is the priority choice when dealing with ordinal data. The traditional regression with assigning scores to the categories or the generalized logit model would not be considered for this purpose because the two models possibly produce misclassification or an overly complicated model. If the data patterns hold or slightly violate the assumption of proportional odds, the proportional odds model is recommended because of model parsimony.

Several methods help to detect the assumption of proportional odds. The Score Test is better than both the Wald Test and the Likelihood Ratio Test because it has more power to detect the deviation of the odds ratio across each logit. However, the Score Test has some limitations based on the sample size and the type of covariates. As a result, Graphical Assessment, the Brant Test, and the Standardized Residuals Test could be supplementary tools to help verify this assumption. If a subset of covariates does have a significant non-proportional odds pattern, we may use the partial proportional odds model to fit the ordinal data. The reduced form for this model may be found through the trend of the odds ratio. To conclude, when selecting a model to deal with ordinal data, the proportional odds model or the partial proportional odds model are the most appropriate models.

References

Agresti, A. (2010). *Analysis of Ordinal Categorical Data*. Wiley.

Agresti, A. (2013). *Categorical Data Analysis*. Wiley John + Sons.

Agresti, A. (2015). *Foundations of Linear and Generalized Linear Models*. John Wiley & Sons Inc.

Allison, P. D. (1999). Comparing logit and probit coefficients across groups. *Sociological Methods & Research*, 28(2):186–208.

Anderson, J. and Philips, P. (1981). Regression, discrimination and measurement models for ordered categorical variables. *Applied Statistics*, 30(1):22.

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

Clogg, C. C. and Shihadeh, E. S. (1994). *Statistical Models for Ordinal Variables*. SAGE Publications, Inc.

Cox, D. R. (1970). *Analysis of Binary Data*. Chapman and Hall/CRC.

Greenwood, C. and Farewell, V. (1988). A comparison of regression models for ordinal data in an analysis of transplanted-kidney function. *Canadian Journal of Statistics*, 16(4):325–335.

Harrell, F. E. (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer.

Hastie, T. J., Botha, J. L., and Schnitzler, C. M. (1989). Regression with an ordered categorical response. *Statistics in Medicine*, 8(7):785–794.

Lall, R., Campbell, M. J., Walters, S. J., and and, K. M. (2002). A review of ordinal regression models applied on health-related quality of life assessments. *Statistical Methods in Medical Research*, 11(1):49–67.

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

Peterson, B. and Harrell, F. E. (1990). Partial proportional odds models for ordinal response variables. *Applied Statistics*, 39(2):205.

Plackett, R. L. (1981). *The analysis of categorical data (Monograph)*. Griffin.

Steven (1951). *Handbook of Experimental Psychology*. John Wiley & Sons Inc.

Stokes, M. E., Davis, C. S., and Koch, G. G. (2012). *Categorical Data Analysis Using SAS, Third Edition*. SAS Institute.

Williams, R. (2016). Understanding and interpreting generalized ordered logit models. *The Journal of Mathematical Sociology*, 40(1):7–20.

Winship, C. and Mare, R. D. (1984). Regression models with ordinal variables. *American Sociological Review*, 49(4):512.