ORDINARY LEAST SQUARES REGRESSION OF ORDERED CATEGORICAL DATA: INFERENTIAL IMPLICATIONS FOR PRACTICE

Beth Larrabee

H. Morgan Scott

Nora M. Bello

ORDINARY LEAST SQUARES REGRESSION OF ORDERED CATEGORICAL DATA: INFERENTIAL IMPLICATIONS FOR PRACTICE

Beth Larrabee[1], H. Morgan Scott[2] and Nora M. Bello[1]
[1]Department of Statistics, [2]Department of Diagnostic Medicine/Pathobiology,
Kansas State University, Manhattan, KS 66506

Ordered categorical responses (OCRs) are frequently encountered in many disciplines. Examples of interest in agriculture include quality assessments, such as for soil or food products, and evaluation of lesion severity, such as teat ends status in dairy cattle. OCRs are characterized by multiple categories recorded on a ranked scale that, while apprising relative order, is not informative of absolute magnitude of or proportionality between the categories. A number of statistically sound models for OCRs are available in the statistical literature, such as logistic regression and probit models, but these are commonly underutilized in practice. Instead, the ordinary least squares linear regression (OLSLR) model is often employed despite violation of basic model assumptions. In this study, the inferential implications of OLSLR-based inference on OCRs were investigated using a simulation study that evaluated realized Type I error rate and empirical statistical power. The design of the simulation study was motivated by applications reported in the subject-matter literature. A variety of plausible scenarios were considered for simulation, including various shapes of the frequency distribution and number of categories of the OCR. Using survey data on frequency of antimicrobial use in cattle feedlots, we illustrated the inferential performance of OLSLR on OCRs relative to a probit model.

Key words: ordinary least squares linear regression; ordered categorical responses; violation of assumptions; inference.

1. Introduction

Researchers in many applied disciplines are frequently faced with the challenge of analyzing ordered categorical responses (OCRs). Examples of interest in agriculture include quality assessments, such as for soil (De Groote et al., 2010) or food products (Hernandez et al., 2005), evaluation of lesion severity, such as teat ends status in dairy cows (Radostits, 2001) and semi-quantification of frequency of management practices, like antimicrobial administration in feedlot cattle (McIntosh et al., 2009; Dean et al., 2011). For instance, McIntosh (2009) worked with OCRs to describe the frequency of antimicrobial use in feedlot cattle as "never", "rarely", "sometimes", "often" and "always". In that study, categories were enumerated from 1 to 5, respectively, to reflect their natural ordering based on frequency of use. Enumeration of ranked categories is commonly observed for OCRs. However, such enumeration is only indicative of order and does not convey any information about the magnitude of the distance between consecutive categories of the response (Liu and Agresti, 2005).

The probit regression model and the proportional odds logistic regression model are examples of models developed specifically to capitalize on the information about relative order while acknowledging the categorical nature of the responses (Long, 1997). However, fitting of these models can be technically challenging in some applications due to, for instance, empty cells and data sparseness (Liu & Agresti, 2005). Additionally, interpretation of parameter

estimates from probit or logit models can be perceived as more mathematically involved and less intuitive than that of the more traditional ordinary least squares linear regression (OLSLR), which is taught widely in applied statistics courses.

The main concern of using OLSLR to fit OCRs is violation of model assumptions, which may call into question the validity of any subsequent inference. Ordinary least squares linear regression assumes that errors be mutually independent and normally distributed with mean zero and constant variance (Kutner, Nachtsheim, Neter & Li, 2005). This implies that the response variable is continuous and symmetrically distributed around a mean with the same variance parameter as the distribution of the error. In modeling OCRs with OLSLR, the underlying assumption is that the ranking of the categories nominally identified as 1,2,3... bears additional information about the relative distances between categories (Long, 1997). For instance, in the example discussed in the first paragraph, category 4 ("often") would indicate twice as frequent antimicrobial use as category 2 ("rarely"). While "often" clearly indicates more frequent use than "rarely", the postulate of doubled frequency is not tenable. Despite frequent application, the inferential implications of assuming proportionality between ranked ordered categories when fitting OLSLR to OCR are unclear. Additional features that are particular to OCRs may also impact the quality of OLSLR-based inference, such as asymmetry of frequency distributions (Abreu et al., 2008; Javaras & Ripley, 2007) and the number of categories of the OCR (Newsom, n.d.).

The objective of this study was to investigate the implications of fitting OLSLR to OCRs for statistical inference. Using a simulation study, we evaluated empirical Type I error (α) and statistical power for inference on OCRs under a variety of scenarios. These simulation scenarios are motivated by subject-matter applications encountered in the scientific literature (Abreu et al., 2008; De Groote et al., 2010; Hernandez et al., 2005; Javaras & Ripley, 2007; McIntosh et al., 2009; Dean et al., 2011; Raubertas & Shook, 1982; Russell & Bobko, 1992; Xu et al., 2007; Yi et al., 2007). We then used survey data to illustrate and compare inference on an OCR based on OLSLR and on a probit model.

2. Data Simulation

Using simulation, we investigated the inferential properties of OLSLR as applied to OCRs. We evaluated OCRs with frequency distributions of uniform, belled, triangular, or exponential shape. We also considered multiple numbers of categories of the OCR, specifically, 2, 3, 4, 5, 7, and 10 categories. For each of the 24 scenarios (4 frequency distribution shapes x 6 number of categories of the OCR), 4,000 Monte Carlo replicates were produced. For each replicate, we generated 300 realizations of a normally distributed latent random variable according to the following equation:

$$y_i^* = \omega^* + \beta^* x_i + \varepsilon_i$$

where y_i^* is the ith observation (i = 1, …, 300) on a normal latent scale that is considered to underlie the OCR (see later); ω^* and β^* are the intercept and slope parameters in the latent scale; x_i is the known covariate value corresponding to the ith observation and ε_i is the error associated with the ith observation (ε_i ~ i.i.d N(0, σ^2)). Realizations of the OCR y_i were generated by discretizing the latent variable y_i^* using $J+1$ thresholds τ_0, τ_1, … τ_J between J ranked categories, such that:

$$y_i = \begin{cases} 1 & if\ \tau_0 = -\infty \leq y_i^* < \tau_1 \\ 2 & \quad if\ \tau_1 \leq y_i^* < \tau_2 \\ \vdots & \qquad \vdots \\ k & if\ \tau_{J-1} \leq y_i^* < \tau_J = \infty \end{cases}$$

These thresholds were specified by quantiles of a standard normal distribution evaluated at the cumulative probabilities listed in Tables 1 through 4. These cumulative probabilities were specified to define frequency distributions of uniform, belled, triangular or exponential shapes in the resulting OCR. For illustrative purposes, Figure 1 depicts histograms of the frequency distribution of one replicate of a simulated OCR with a total of 7 categories and a uniformly-shaped (A) or exponentially-shaped (B) frequency distribution.

We simulated each scenario using $\beta^* = 0$ and $\beta^* = 1$ in order to assess empirical Type I and Type II errors in statistical inference, respectively. For simulation purposes, the explanatory covariate x_i was allowed to assume 2 settings in each scenario. In one setting, namely $\boldsymbol{x}^{(51)}$, x_i was allowed to take 1 of 51 possible integer values ranging from -50 to 50 in intervals of 2. In the second setting, namely $\boldsymbol{x}^{(5)}$, x_i was allowed to take one of 5 possible integer values ranging from -50 to 50 in intervals of 25. The variance of the error (σ_ε^2) was set such that the statistical power for testing H_0) $\beta^* = 0$ using OLSLR on the latent scale was approximately 0.80.

3. Analysis

The following linear model was fit to each Monte Carlo replication:

$$y_i = \omega + \beta x_i + \varepsilon_i \qquad \varepsilon_i \sim i.i.d.N(0, \sigma^2) \qquad (1)$$

whereby y_i is the observed category of the OCR on the ith subject (i = 1, . . ., 300) such that categories were enumerated from 1 to J; ω and β were the OLSLR intercept and slope parameters, respectively, x_i was the ith fixed value of the predictor variable and ε_i was the error associated with the ith observation, whereby $\varepsilon_i \sim i.i.d.N(0, \sigma^2)$. We emphasize that the OLSLR model in Equation 1 assumed that the enumeration of the realized ordered response categories was indicative not only of relative ranking (as was supported by the definition of an OCR), but also of proportional magnitude between consecutive categories of the OCR. The later assumption is, by definition, incorrect (Liu and Agresti, 2005). Our interest is on the inferential robustness for such an incorrectly defined model.

In each Monte Carlo replication, we used OLSLR and probit regression to estimate β and then test the null hypothesis H_o) $\beta = 0$ based on a t test statistic with (n – 2) degrees of freedom where n is the sample size (n = 300). The test statistic was then compared with a critical value given by the (1-α) percentile of a t-distribution with (n – 2) degrees of freedom.

For each of the 24 simulated scenarios at each $\boldsymbol{x}$ setting, empirical Type I error was computed as the frequency of incorrect rejections of H_o) $\beta = 0$ at a 5% level of significance, divided by the total number of Monte Carlo replicates. Lower and upper bounds for empirical Type I errors were determined using the 2.5th and 97.5th percentiles of a binomial distribution with number of trials N = 4000, given by the number of Monte Carlo replicates, and probability of event π = 0.05, given by the desired level of significance. Similarly, empirical power was computed for each scenario and $\boldsymbol{x}$ setting as the proportion of Monte Carlo replicates for which

H_o) $\beta = 0$ had been correctly rejected at a 5% level of significance. A normal approximation to the binomial distribution was used to construct Bonferroni-adjusted confidence intervals on the differences in empirical power between probit models and their corresponding OLSLR fits, as well as between distribution shape scenarios fitted with OLSLR. All computations for data simulation and analyses were conducted using the statistical software R Version 2.9.2 (R Development Core Team, 2009).

4. Results

Empirical Type I Error. Figure 2A and 2B illustrates the results for empirical Type I error rates for all scenarios under the $\boldsymbol{x}^{(51)}$ and $\boldsymbol{x}^{(5)}$ settings, respectively. Across the frequency distribution shapes considered in this study, the empirical Type I error remained close to the nominal value of 0.05, regardless of the number of ranked categories of the OCR. For all scenarios, empirical Type I errors were within probabilistic expectation based on a binomial distribution with probability of 5% and size given by the number of Monte Carlo replicates for each scenario. The maximum and minimum empirical Type I error rates were 0.043 and 0.057, corresponding to scenarios with 4-category exponentially-shaped OCR generated under $\boldsymbol{x}^{(5)}$ and 7-category uniformly-shaped OCR generated under $\boldsymbol{x}^{(51)}$, respectively.

Empirical Statistical Power. Figures 3A, 3B, 3C and 3D illustrate empirical power for OLSLR and probit models as a function of the number of categories for OCRs with frequency distributions of uniform, bell, triangular or exponential shape, respectively. Statistical power of OLSLR fitted on the latent-scaled variable is provided as a reference line at 0.80. Overall, across scenarios of distributional shapes and $\boldsymbol{x}$ settings, statistical power for OLSLR-based inference on β was weakest when OCRs consisted of 2 categories. As the number of categories of the OCR increased, so did empirical power. However, the rate of increase in power with more OCR categories was increasingly abated as the OCRs reached 5 or more categories. For the uniform, bell, and triangularly shaped scenarios (Figures 3A, 3B and 3C), but not for the exponentially-shaped scenario (Figure 3D), empirical power for both OLSLR and probit-link models analysis approached the 0.80 reference line as the number of categories of the response increased.

Of particular interest was the comparison of statistical power between OLSLR and probit models across simulated scenarios (Figure 3). Overall, there was no evidence for differences in empirical power between OLSLR and the probit models for any combination of frequency distributions of bell, uniform or triangular shape, $\boldsymbol{x}$ setting and number of categories of the OCR considered in this study. No evidence for differences between OLSLR and probit model inference was apparent for scenarios with exponentially shaped frequency distributions where the OCR was characterized by 7 or fewer categories. However, OLSLR-based empirical statistical power for 10-category OCRs with an exponentially shaped frequency distribution was significantly decreased compared to that of a probit model analysis.

Figure 4 illustrates OLSLR-based empirical statistical power for OCRs with uniform, belled, triangular and exponentially-shaped frequency distributions. In most cases, comprising any given number of OCR categories and $\boldsymbol{x}$ settings, we found no evidence for differences in OLSLR-based empirical power between uniform, triangular and belled-shaped frequency

distributions of the OCR. As an exception we note the scenario with an OCR with 4 categories, whereby OLSLR-based empirical power was greater with a bell-shaped frequency distribution in comparison to a triangular-shaped frequency distribution. When the OCR had an exponentially-shaped frequency distribution and more than 4 categories, OLSLR-based empirical power was significantly decreased relative to any other frequency distribution shapes. For both OLSLR and probit models, empirical power for the exponentially-shaped scenario increased until the OCRs had 5 categories, then peaked, and finally decreased beyond this point.

Interpretation of the Slope Parameter Estimate. Following from OLSLR-based inference on the association between an OCR and an explanatory variable, one may inquire about the interpretation of the estimated slope parameter ($\hat{\beta}$) in the context of the research problem. To address this question, we collected the OLSLR point estimates $\hat{\beta}$ for all 4000 Monte Carlo replications in a given scenario and used them to describe the empirical sampling distributions of $\hat{\beta}$. Figure 5 illustrates the empirical sampling distribution of $\hat{\beta}$ for 2 representative scenarios, as described next. When the OCRs were generated based on $\beta^* = 0$, the empirical sampling distributions of $\hat{\beta}$ were observed to be bell-shaped and centered about zero for all simulated scenarios (Figure 5A). In fact, in all simulated scenarios, the 25th and 75th percentiles of the sampling distributions of $\hat{\beta}$ had negative and positive signs, respectively, indicating a consistent coverage of the null value 0. In contrast, when the OCR was generated based on $\beta^* = 1$ in the latent scale, the sampling distributions of $\hat{\beta}$ never overlapped with the true parameter value of $\beta^* = 1$ (Figure 5B), regardless of frequency distribution shapes and number of categories of the OCRs. In fact, the sampling distributions of $\hat{\beta}$ under $\beta^* = 1$ had minimum values that frequently fell below 0 and maximum values that failed to exceed 0.04 in all simulated scenarios. This indicates that the magnitude of the OLSLR estimate $\hat{\beta}$ is not directly comparable to the slope parameter in the latent scale β^*, which was used to generate the data. This may be partially explained by the different scales, namely the latent continuous scale of the data generation process versus the ordered categorical scale used for inference. In addition, we further emphasize that the assumption of proportionality across categories of the OCR that is inherent to OLSLR is not necessarily supported by the data generation process. Such incorrect model specification further impairs meaningful interpretation of the OLSLR estimate of the slope parameter $\hat{\beta}$ as a rate of change on the OCR per unit increase in an explanatory variable X.

5. Case Study Based on Real Survey Data

To illustrate OLSLR-based inference on an OCR in a real data problem, we adapt a data set obtained from a survey study conducted by McIntosh et al. (2009). The dataset used for our case study included all 269 feedlot operators and veterinarians that returned the survey, instead of just the 103 veterinarians originally reported by McIntosh et al. The original research question of interest was the frequency of use of antimicrobials in feedlot cattle as it relates to Icek Ajzen's theory of planned behavior (Ajzen, 2011). This theory states that behavioral beliefs, normative beliefs, and belief about perceived behavioral control jointly define an intention that may then result in an action (Ajzen, 2011). In other words, the researchers were interested in investigating the association between frequency of antimicrobial use in feedlot cattle and behavioral, normative and control beliefs. The researchers used these beliefs to motivate questions for a 20

page survey that was distributed to veterinarians and feedlot operators in the Midwestern US. The individuals surveyed were asked to answer questions about their motivating beliefs and to rate their frequency of antimicrobial use as denoted by the categories "always", "often", "sometimes", "rarely", "often" and "never." Reported frequencies of antimicrobial use were labeled 1 to 5, respectively, to indicate relative ranking and used as the OCR for data analysis.

First, the research question was addressed by fitting a probit regression model to the OCR. A selected aspect of behavioral belief from the theory of planned behavior, in particular a belief in regard to the efficacy and profitability of antimicrobial use in feedlot cattle, was used as an explanatory covariate in the model. The regression coefficient under the probit model was found to be significantly greater than zero ($P<0.001$), with a point estimate of 0.6076 and a standard error of 0.0739. This result supports a significant association between frequency of antimicrobial use and the combined effect of one's beliefs that antimicrobial use is effective, that its use will bring profit and that efficacy and profitability are important to the respondent. The estimate of the regression coefficient in the probit link scale may be interpreted as the average change in z-score per unit increase of the explanatory covariate. Alternatively, in the data scale, a function of this estimate may be interpreted as the estimated change in the probability of a given category of frequency of antimicrobial use relative to the baseline category of "never" and to a mean value of the explanatory variable X, for each unit increase of the explanatory variable of behavioral belief.

We also fitted an OLSLR model to the OCR. The response was the 1-to-5 rank associated with the category of frequency of antimicrobial use and the explanatory covariate was the same behavioral belief predictor fitted with the probit model. The OLSLR coefficient was estimated to be -0.4474 with a standard error of 0.0487 and found to be significantly different from zero ($P < 0.001$). Some practitioners may support a pragmatic interpretation of this estimate as a decrease of approximately half category per unit increase in the predictor, or alternatively, as a decrease of almost one category per 2-unit increase in the predictor. We emphasize the limitations of this interpretation based on the assumption of proportionality in the relative magnitude of the ranked categories of the OCR, as well as on unreasonable inference outside of the range of the categories defined for frequency of antimicrobial use in this application.

In comparing inference between the probit model and OLSLR fitted on the OCR of frequency of antimicrobial use, we note the similarity of the overall conclusion, that is, that the null hypothesis was rejected. However, the point estimates for the corresponding regression coefficients are not comparable in magnitude nor in interpretation. This is consistent with the results of our simulation study.

6. Discussion

In this study, we evaluated the inferential implications of analyzing OCRs with OLSLR methodology. Based on a simulation study, we explored empirical Type I error and empirical statistical power for OLSLR-based inference on an OCR under a variety of scenarios, including various frequency distribution shapes and increasing number of categories of the OCR, as well as 2 settings of the explanatory covariate, namely $\boldsymbol{x}^{(5I)}$ and $\boldsymbol{x}^{(5)}$. Overall, our results indicated that OLSLR-based inference on OCR was generally robust to violation of model assumptions in terms of Type I error. These results should provide some level of reassurance to subject-matter

scientists in support of research conclusions on OCR that were obtained based on OLSLR methodology. Furthermore, statistical power of OLSLR was comparable to that of probit models, except for rather extreme scenarios of exponential shape of the frequency distribution of an OCR with more than 7 categories.

Inferential robustness of analysis of variance (ANOVA), in terms of Type I error, has also been reported for inference on OCRs. Hsu and Feldt (1969) showed that as the number of categories of OCR increased, so did the probability of correctly failing to reject the null hypothesis. We also found OLSLR of OCR to be robust to violations of assumptions, although we did not find any evidence for an increasing robustness of OLSLR as a function of the number of categories of the OCR. The similarity of robustness results is not necessarily surprising given that both ANOVA and OLSLR are based on ordinary least squares estimation and can be shown to be equivalent models under simple circumstances. For instance, a 1-way factorial ANOVA with 2 levels can be shown to be equivalent to OLSLR when the explanatory covariate takes only 2 values (Kutner et al., 2005).

Our design of the simulation study included various frequency distribution shapes of the OCR, and thus, provides some opportunity to assess the impact of skewness and kurtosis on the quality of OLSLR based inference. For instance, as skewness and kurtosis increase jointly, we can expect a deleterious effect on the power of OLSLR. It can be expected that as the number of categories of the OCR decreases, the impact of skewness and kurtosis on the shape of the frequency distribution would be attenuated. This is likely associated with a more rectangular shape of the frequency distribution when categories are few. In contrast, as the number of categories of the OCR increase, the frequency distribution shape can be expected to reflect skew and kurtosis in more pronounced, non-rectangular shapes. It is noted that with OCRs that had an exponentially shaped frequency distribution (high skew and high kurtosis), we did not see departures of the empirical Type I error from its nominal value regardless of the numbers of categories of the response. This may be interpreted as considerable robustness of OLSLR-based inference on OCRs to skewness and kurtosis, at least in the context of the scenarios considered in this study. Glass, Peckham, and Sanders (1972) have claimed that leptokurtosis may enhance statistical power, but that platykurtosis may have an opposite, detrimental impact on it when considering ordinary least squares techniques as applied to continuous variables. These findings did not appear to extend to our particular use of OLSLR-based analysis of OCR. Instead, we found that the less skewed and less kurtotic the frequency distribution was, the greater the power. For instance, the bell-shaped distribution (no skew, no excess kurtosis) tended to perform the best and yielded the greatest power relative to the other scenarios considered when the Bonferroni adjustment used was relaxed to only account for the actual comparisons done rather than all possible comparisons that might have been of interest. In turn, statistical power for scenarios with a triangular (moderate skew) or uniform (moderate platykurtosis) shapes were not significantly different from the bell scenario (no skew, no excess kurtosis) with the Bonferroni adjustment in use for most of this paper. The exponential scenario (high skew and high kurtosis) tended to have the worst power and truly distinguished itself as the poorest after 5 categories on the OCR. As was mentioned previously, the frequency distribution shapes were more similar when the responses had few categories. Differences in shape only become apparent with increasing number of categories; this may provide a partial explanation for the increasingly poor performance of the exponential scenario as the number of categories of the response increased.

In to modeling responses with various distribution shapes, it is also common to model a response as a function of more than one covariate at a time. For an OCR, it is not clear how OLSLR-based inference would perform under more than 1 explanatory variable, in particular under the combinations of quantitative and categorical predictors, as well as potential interaction terms. Russell and Boboko (1992) have mentioned the need for investigating interaction models for OCRs analyzed with OLSLR because situations where an interaction effect would be expected to be present are frequently encountered in their discipline of psychology. We conducted a data analysis with real data to demonstrate OLSLR in a practical application. The OLSLR analysis of a 5-category OCR agreed with a theoretically sound probit model analysis in terms of the significance of the continuous covariate under inspection. Solid investigation on quality of statistical inference on an OCR as a function of interactions between explanatory variables remains to be done. Another consideration in relation to $\boldsymbol{x}$ is that our models only considered a first order (linear) explanatory term. We did not investigate the inferential implications of fitting OLSLR for OCRs when the underlying latent variable was generated as polynomial function of the explanatory variable. These realms of inquiry remain open to investigation.

We have considered the case where we have a single outcome variable free from any complicated multivariate structure. It would be useful for application to assess robustness of OLSLR fitted simultaneously to more than one response variable, as it may be desired to infer upon the joint behavior of multiple survey responses. For example, in our case study, we focused on the frequency of antimicrobial use of feedlot operators and veterinarians as it pertains to *en masse* treatment (Dean et al., 2011). These individuals were also questioned about their frequency of antimicrobial use for acutely ill cattle, for chronically ill cattle, and for subtherapeutic use in at-risk cattle. A joint analysis of these OCRs would permit a more comprehensive assessment of antimicrobial use, simultaneously accounting for potential compensatory or complementary behaviors across scenarios.

Extension of OLSLR-based inference on OCRs in the presence of random and fixed effects can provide additional insight regarding performance of the inference with a hierarchical data structure. For example, we may wish to see what the inferential implications are for fitting linear mixed models when our dependent variable is an OCR measured over time instead of a continuous variable (Liu and Agresti, 2005). This situation may occur in agriculture when measuring score-type responses on a given animal at multiple times throughout the course of a study.

Finally, while OLSLR appears robust to violations of model assumptions, it is advised that methods with sound statistical foundations be employed for analyzing OCRs, as there are additional advantages for using theoretically appropriate techniques, such as relevant interpretation of parameter estimates in the data scale. For instance, the disciplinary scientist may implement probit regression or logistic regression to infer upon and estimate the relative probability of being in a certain response category for a given level of the explanatory covariate. Furthermore, it is worth noticing that the robustness of OLSLR outside of the scenarios considered in the simulation study presented herein is not known. Given the broad availability of theoretically sound techniques combined with the uncertainty that remains in regard to OLSLR

of OCRs, it is advised that OLSLR only be considered an alternative for inference on OCRs in problematic situations. Specifically, OLSLR-based analysis of OCR may be considered as screening tools in cases where statistically sound techniques become impossible to implement due to technical complications such as empty cells or data sparseness.

7. Summary

This research investigated the inferential implications of fitting OLSLR to OCRs. It appears that OLSLR may be a reasonable choice for investigating the hypothesis of $\beta = 0$, even when the response variables are of ordered categorical nature and violate basic OLSLR assumptions. Simulation studies did not suggest major problems with control of Type I error or empirical power for most scenarios considered. However, the estimate for the slope parameter itself should not be regarded as informative of magnitude of effect in the subject-matter context due to the fact that the rankings of the ordered categorical variables do not themselves grant any information about their relative distances.

8. References

Abreu, M. N. S., Siqueira, A. L., Cardoso, C. S., & Caiaffa, W. T. (2008). Ordered categorical logistic regression models: Application in quality of life studies. *Cadernos De Saúde Pública, 24*, s581-s591.

Ajzen, Icek. *Theory of planned behavior.* Retrieved 3/21/2011 from http://www.people.umass.edu/aizen/tpb.html

Dean, W.R., McIntosh, W.A., Scott, H.M., & Barling, K.S. (2011). The role of trust and moral obligation in beef cattle feed-lot veterinarians' contingent adoption of antibiotic metaphylaxis recommendations. *International Journal of Sociology of Agriculture and Food, 18*(2), 104-120.

De Groote, H., Rutto, E., Odhiambo, G., Kanampiu, F., Khan, Z., Coe, R., & Vanlauwe, B. (2010). Participatory evaluation of integrated pest and soil fertility management options using ordered categorical data analysis. *Agricultural Systems, 103*(5), 233-244. doi:10.1016/j.agsy.2009.12.005

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research, 42*(3), 237.

Hernandez, P., Guerrero, L., Ramirez, J., Mekkawy, W., Pla, M., Arino, B., & Blasco, A. (2005). A bayesian approach to the effect of selection for growth rate on sensory meat quality of rabbit. *Meat Science, 69*(1), 123-127. doi:10.1016/j.meatsci.2004.06.013

Hsu, T., & Feldt, L. S. (1969). The effect of limitations on the number of criterion score values on the significance level of the F-test. *American Educational Research Journal, 6*(4), pp. 515-527.

Javaras, K. N., & Ripley, B. D. (2007). An "Unfolding" latent variable model for likert attitude data. *Journal of the American Statistical Association, 102*(478), 454-463. doi:10.1198/016214506000000960

Kutner, M., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models.* Boston: McGraw-Hill Irwin.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks: Sage Publications.

Liu, I., & Agresti, A. (2005). The analysis of ordered categorical data: An overview and a survey of recent developments. *Test, 14*(1), 1-73.

McIntosh, W. M. A., Schulz, S., Dean, W., Scott, M. H., Barling, K. S., & Takei, I. (2009). Feedlot veterinarians' moral and instrumental beliefs regarding antimicrobial use in feedlot cattle. *Journal of Community & Applied Social Psychology, 19*(1), 51-67. doi:10.1002/casp.976

Newsom, J. T. (n.d.). *Lecture 1: Types of scale & levels of measurement.* Retrieved 5/20/2011 from http://www.upa.pdx.edu/IOA/newsom/pa551/lecture1.htm.

R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Radostits, O.M. "Herd Health: Food Animal Production Medicine" W.B. Saunders Company. Pensylvannia, 2001. 3rd edition.

Raubertas, R., & Shook, G. (1982). Relationship Between Lactation Measures of Somatic-Cell Concentration and Milk-Yield. *Journal of Dairy Science, 65*(3), 419.

Russell, C. J., & Bobko, P. (1992). Moderated regression analysis and likert scales: Too coarse for comfort. *Journal of Applied Psychology, 77*(3), 336.

Xu W. H., Dai, Q., Xiang, Y. B., Zhao, G. M., Ruan, Z. X., & Cheng, J.R. (2007). Nutritional factors in relation to endometrial cancer: a report from a population-based case-control study in Shanghai, China. *International Journal of Cancer, 120*, 1776-1781.

Yi, N., Samprit, B., Pomp, D., & Yandell, B. S. (2007). Bayesian Mapping of Genomewide Interacting Trait Loci for Ordered Categorical Traits. *Genetics, 176*(3), 1855-1864.

Table 1. Probability that an ordered categorical response (OCR) takes on a certain value, given the total number of categories of the response. These probabilities were used to define thresholds in the latent scale that separate sequentially-ranked categories of an OCR with a uniformly-shaped frequency distribution.

	Number of categories of the ordered categorical response.					
Realized value of an ordered categorical response.	**2**	**3**	**4**	**5**	**7**	**10**
1	1/2	1/3	1/4	1/5	1/7	1/10
2	1/2	1/3	1/4	1/5	1/7	1/10
3	.	1/3	1/4	1/5	1/7	1/10
4	.	.	1/4	1/5	1/7	1/10
5	.	.	.	1/5	1/7	1/10
6	.	.	.	.	1/7	1/10
7	.	.	.	.	1/7	1/10
8	.	.	.	.	.	1/10
9	.	.	.	.	.	1/10
10	.	.	.	.	.	1/10

Table 2. Probability that an ordered categorical response (OCR) takes on a certain value, given the total number of categories of the response. These probabilities were used to define thresholds in the latent scale that separate sequentially-ranked categories of an OCR with a bell-shaped frequency distribution.

	Number of categories of the ordered categorical response.					
Realized value of an ordered categorical response.	**2**	**3**	**4**	**5**	**7**	**10**
1	1/2	1/4	1/6	1/9	1/16	1/30
2	1/2	2/4	2/6	2/9	2/16	2/30
3	.	1/4	2/6	3/9	3/16	3/30
4	.	.	1/6	2/9	4/16	4/30
5	.	.	.	1/9	3/16	5/30
6	.	.	.	.	2/16	5/30
7	.	.	.	.	1/16	4/30
8	.	.	.	.	.	3/30
9	.	.	.	.	.	2/30
10	.	.	.	.	.	1/30

Table 3. Probability that an ordered categorical response (OCR) takes on a certain value, given the total number of categories of the response. These probabilities were used to define thresholds in the latent scale that separate sequentially-ranked categories of an OCR with a triangularly-shaped frequency distribution.

	Number of categories of the ordered categorical response.					
Realized value of an ordered categorical response.	**2**	**3**	**4**	**5**	**7**	**10**
1	1/3	1/6	1/10	1/15	1/28	1/55
2	2/3	2/6	2/10	2/15	2/28	2/55
3	.	3/6	3/10	3/15	3/28	3/55
4	.	.	4/10	4/15	4/28	4/55
5	.	.	.	5/15	5/28	5/55
6	.	.	.	.	6/28	6/55
7	.	.	.	.	7/28	7/55
8	.	.	.	.	.	8/55
9	.	.	.	.	.	9/55
10	.	.	.	.	.	10/55

Table 4. Probability that an ordered categorical response (OCR) takes on a certain value, given the total number of categories of the response. These probabilities were used to define thresholds in the latent scale that separate sequentially-ranked categories of an OCR with an exponentially-shaped frequency distribution.

	Number of categories of the ordered categorical response.					
Realized value of an ordered categorical response	**2**	**3**	**4**	**5**	**7**	**10**
1	1/3	1/7	1/15	1/31	1/127	1/1023
2	2/3	2/7	2/15	2/31	2/127	2/1023
3	.	4/7	4/15	4/31	4/127	4/1023
4	.	.	8/15	8/31	8/127	8/1023
5	.	.	.	16/31	16/127	16/1023
6	.	.	.	.	32/127	32/1023
7	.	.	.	.	64/127	64/1023
8	.	.	.	.	.	128/1023
9	.	.	.	.	.	256/1023
10	.	.	.	.	.	512/1023

Figure 1. Frequency distributions of an ordered categorical response with 7 sequentially-ranked categories and A) a uniform shape or B) an exponential shape. Histograms are based on a single simulated Monte Carlo replication with 300 observations.

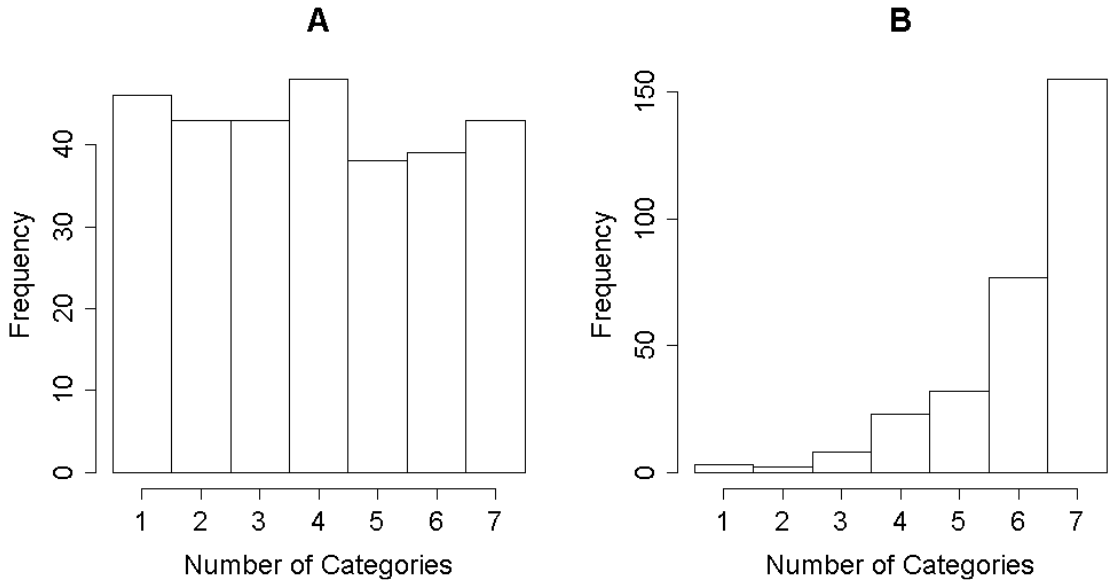

Figure 2. Empirical Type I error for inference on H_0) $\beta = 0$ based on ordinary least squares regression fitted to an ordered categorical response with increasing number of categories and Uniform, Bell, Triangular, or Exponentially shaped frequency distributions. Each scenario is represented by 4000 Monte Carlo replications. The figures represent settings where the explanatory covariate consisted of A) 51 or B) 5 levels ranging from -50 to50. Bounds on the point estimate for empirical power correspond to 2.5th and 97.5th percentiles of a binomial distribution with probability of success π =0.05 and number of trials N= 4000.

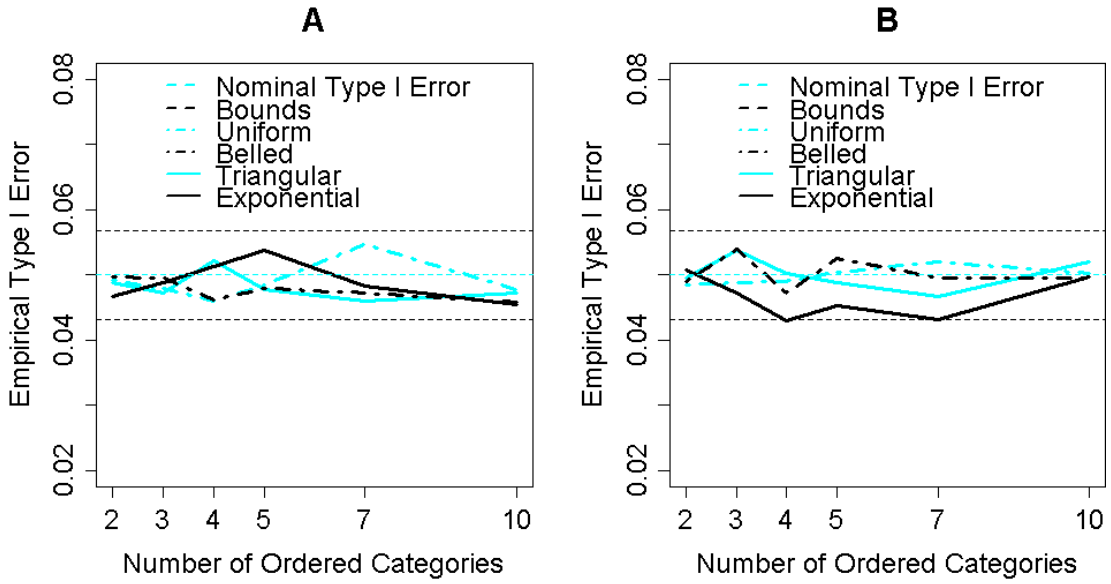

Figure 3. Empirical statistical power for correctly rejecting H_0) $\beta = 0$ based on a probit or ordinary least squares linear regression fitted to an ordered categorical response characterized by increasing number of categories with an A) Uniformly- , B) Bell-, C) Triangular- or D) Exponentially-Shaped frequency distribution. The figure represents settings where the explanatory covariate consisted of either 51 levels (i.e. $\boldsymbol{x}^{(51)}$) ranging from -50 to 50 in intervals of 2 or 5 levels (i.e. $\boldsymbol{x}^{(5)}$) ranging from -50 to 50 in intervals of 25.

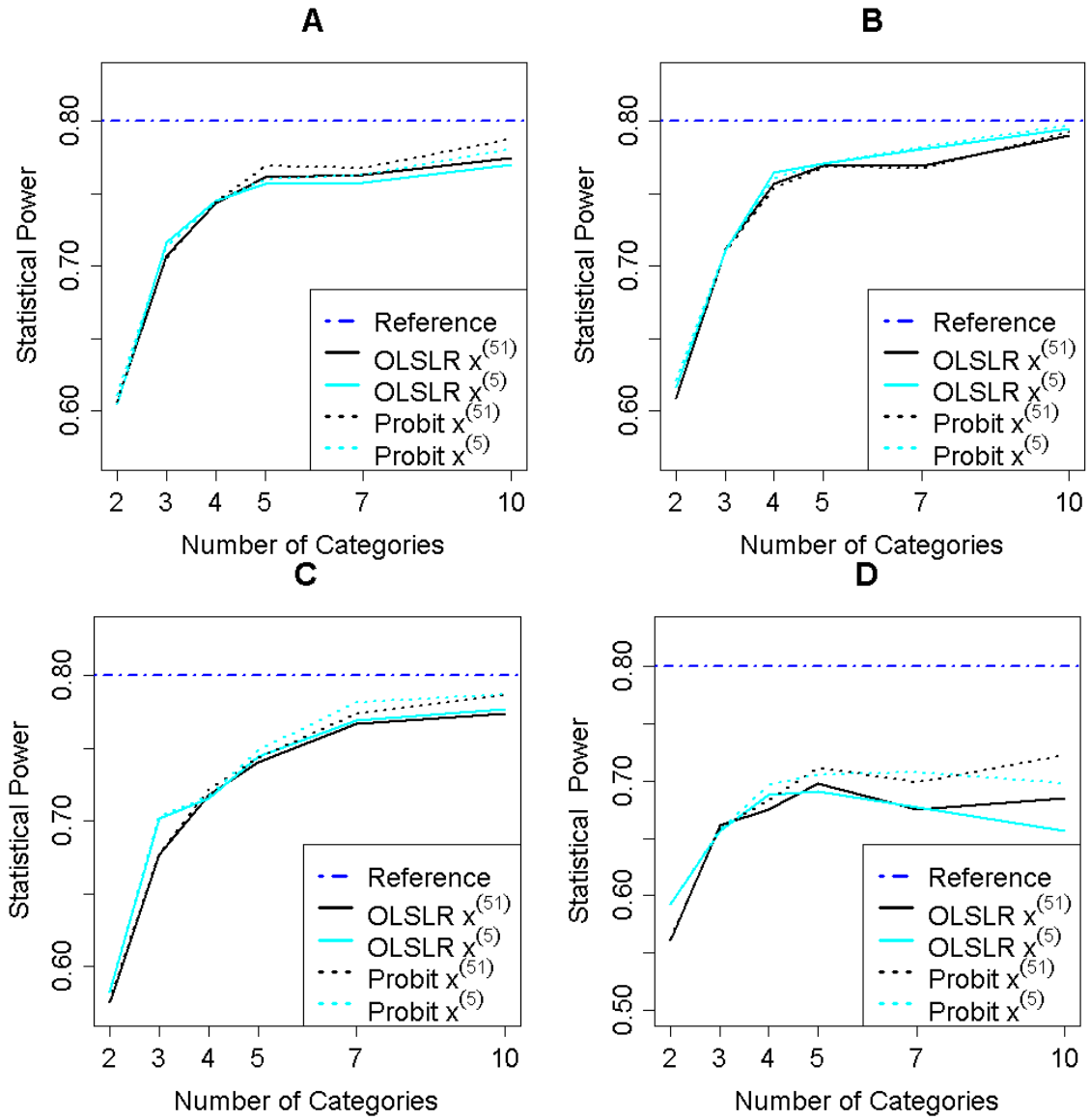

Figure 4. Empirical statistical power for inference on H_0) $\beta = 0$ based on ordinary least squares linear regression fitted to an ordered categorical response characterized by increasing number of categories and Uniform, Bell, Triangle, or Exponentially shaped frequency distributions. Each scenario is represented by 4000 Monte Carlo replications. The figure represents settings where the explanatory covariate consisted of A) 51 levels ($x_i^{(51)}$) ranging from -50 to 50 in intervals of 2 or B) 5 levels ($x_i^{(5)}$) ranging from -50 to 50 in intervals of 25.

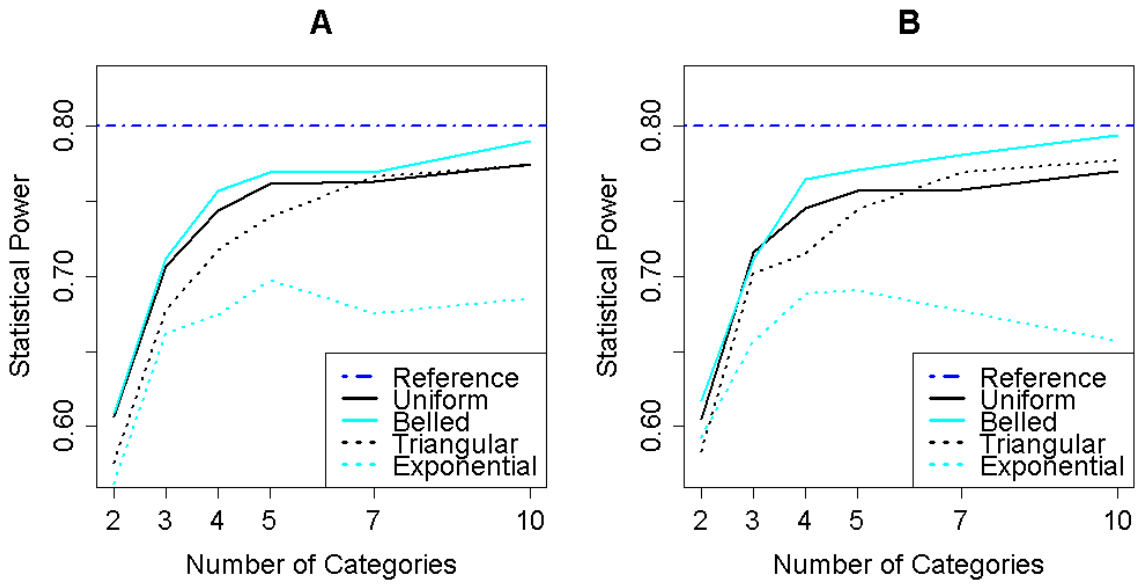

Figure 5. Empirical sampling distribution for the estimated slope parameter $\widehat{\boldsymbol{\beta}}$ obtained from fitting ordinary least squares linear regression to an ordered categorical response with 2 categories and a Uniformly-Shaped frequency distribution. Figures are for illustrative purposes only. The histogram is based on 4000 Monte Carlo replications generated under the condition that A) $\beta^* = 0$ or B) $\beta^* = 1$. The figure represents the setting where the explanatory covariate consisted of 51 Levels ($x^{(51)}$) ranging from -50 to 50 in increments of 2.