Presentation of Ordinal Regression Analysis on the Original Scale

Murray Hannah
Agriculture Victoria, Dairy Research Institute, 
RMB 2460, Ellinbank, Victoria 3820, Australia

and

Paul Quigley
Pastoral and Veterinary Institute, 
PB 105, Hamilton, Victoria 3300, Australia

SUMMARY

Frequently, ordinal measurement scales are constructed either by coarse measurement of interval or ratio scales, or by assigning numeric scores to ordinal categories. When data on such scales are analyzed using the ordinal regression techniques of McCullagh (1980, Journal of the Royal Statistics Society, Series B 42, 109–142), inference is usually performed, and results presented, on the scale of the linear predictor. This can be unsatisfactory from the point of view of the applied scientist familiar with the original scale definition. It is demonstrated how ordinal regression results can be presented on the original ordinal score scale.

1. Introduction

Ordinal regression models of the type introduced by McCullagh (1980) are useful for the analysis of data on an ordinal scale. They are used widely in medical work, psychology, sociology, and agriculture, and constitute the major parametric tool available for the analysis of ordinal data. Implicit in these models is a change of scale from the original ordinal scale to a continuous, linear predictor scale upon which biologically meaningful parameters are usually presented. However, the consultant statistician may encounter a lack of understanding, or even a lack of acceptance, of ordinal regression analysis among clients which can be difficult to overcome and which appears to be due primarily to the implicit change of scale. The relationship between model parameters which appear in the linear predictor and the original ordinal scale is obscure.

Ordinary linear regression (OLR) of the score data, while conceivably more attractive to some clients since both analysis and presentation could occur on the original scale, involves model assumptions (in particular, continuity and normality of distribution, and homoscedasticity) which would be difficult to justify. Results from OLR analysis depend upon the arbitrary score definition, and their validity would be at least questionable.

It may be argued that a change of scale is entirely appropriate and ought not be a barrier to the adoption of ordinal regression. By definition, the ordinal scale is arbitrary, apart from order, and hence inference on the scale of the linear-predictor, which is independent of the assigned ordinal scores, captures all the relevant information on location or scale treatment effects on a hypothetical underlying liability scale.

Frequently, applied statisticians must deal with discrete data derived from an interval scale having poor resolution. For example, visual keys have been constructed for the assessment of some plant diseases. “Percent leaf area affected” may be assessed as 0, 10, 25, 50, 75, 90, or 100 by visual matching of samples against a carefully calibrated key. Ordinal regression techniques can be useful in dealing with the discrete, at-least-ordinal, nature of such data.

Key words: Ordinal; Ordinal regression; Presentation; Scale; Scores; Uncertainty intervals.
Even where no underlying quantitative scale is apparent, subjective numeric scores are sometimes attached to the ordinal categories (e.g., Table 1 based on Corbin, Brockwell, and Gault (1977)). The scores become a short-hand way of referring to the physical states to which they correspond. Workers in a given field of endeavour readily interpret score values in terms of their scale definition.

Such quantitative information, or labelling, is lost when results are presented on the scale of the linear predictor. Scientists who have labored to develop, what is to them, a meaningful “quantitative” scale may feel discontent with a form of analysis which reports results on a unitless linear-predictor scale having no clear physical interpretation.

The purpose of this note is to describe a simple means of presentation of ordinal regression analysis results on the original scale.

2. Method

Suppose there are $k$ ordered categories and we observe a multinomial outcome

$$Y \sim mn(n, k, \pi),$$

where $\pi$ is a vector of cell probabilities, $(\pi_1, \ldots, \pi_k)^T$. Define the cumulative cell probabilities,

$$\gamma_j = \sum_{a=1}^{j} \pi_a, \quad j = 1, \ldots, k.$$

For simplicity of exposition, consider $t$ treatment groups and a location ordinal regression model specified by

$$g(\gamma_{ij}) = \theta_j - \alpha_i, \quad i = 1, \ldots, t; \quad j = 1, \ldots, k - 1; \quad \sum \alpha_i = 0,$$

where $g(\cdot)$ is a suitable link function, $\alpha_i$ is a location effect for treatment $i$, and $\theta_j$ is the $j$th cut-point parameter.

Now suppose that each ordinal class has an associated score, $s_j, j = 1, \ldots, k$, such that $s_j < s_{j'}$ iff $j < j'$. The mean score for treatment $i, \mu_i$, may be defined as the inner product,

$$\mu_i = s^T \pi_i = s^T \nabla \gamma_i = s^T \nabla g^{-1}(\theta - \alpha_i),$$

where $\nabla$ is the first difference operator, and $\pi_i$ and $\gamma_i$ are the vectors of cell and cumulative cell probabilities, respectively, for treatment group $i$. Thus $\mu_i$ is a function of $\theta$ and $\alpha_i$. Maximum likelihood estimates for $\theta$ and $\alpha_i$ lead to the MLE for $\mu_i$.

Formal inference can be performed as usual, using changes in the scaled deviance and contrasts between parameters in the linear predictor (McCullagh, 1980). In addition, (1) may be used to back-transform treatment effect estimates from the linear-predictor scale to scores on the original ordinal scale. These may be accompanied by uncertainty intervals (Snee, 1981) which likewise can be back-transformed to the original scale.

Initially, uncertainty interval end points are defined on the scale of the linear predictor as

$$\hat{\alpha}_i \pm \hat{\delta}_i,$$

where $\hat{\delta}_i$ is chosen according to a desired level of uncertainty. In particular, we may specify $\delta_i$ such that (2) defines least significant intervals, or LSIs (Snee, 1981). For this we require

$$\delta_i + \delta_{i'} \geq \text{LSD}_{ii'},$$

where $\text{LSD}_{ii'}$ is the least significant difference for $\alpha_i$ and $\alpha_{i'}$, and $\zeta_q$ is the $q$-quantile of the standard normal distribution. In general, the (conservative) inequality in (3) is required to preserve the LSI interpretation, namely, that disjoint intervals indicate significant differences at the given level. Equation (3) is achieved by defining

$$\delta_i = \max_{i'} \left\{ \frac{\text{LSD}_{ii'}}{\text{se}_i + \text{se}_{i'}} \right\} \text{se}_i$$
where

\[ se^2_i = \text{var}(\hat{\alpha}_i). \]

3. Example
Rhizobia are bacteria which inhabit the roots of leguminous plants. They facilitate the fixation of atmospheric nitrogen via root nodulation, resulting in increased soil nitrogen levels. Hence rhizobia hold an important place in agricultural systems.

In an experiment to ascertain root nodulation effectiveness of endemic rhizobia strains relative to commercially available rhizobia strains, rhizobia were collected from 18 sites in southwest Victoria, Australia. Each of these and each of three commercial rhizobia strains was used to inoculate the roots of 20 individually potted subterranean clover seedlings. Seedlings were grown in the glasshouse according to a completely randomized design and subsequently scored for nodulation using the ordinal scale defined in Table 1.

### Table 1

Ordinal score definition for root nodulation in subterranean clover based on Corbin et al. (1977)

<table>
<thead>
<tr>
<th>Crown</th>
<th>Elsewhere</th>
<th>Nodulation score definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1-4</td>
<td>0.5</td>
</tr>
<tr>
<td>0</td>
<td>&gt;10</td>
<td>1</td>
</tr>
<tr>
<td>≤10</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>≤10</td>
<td>≤10</td>
<td>2.5</td>
</tr>
<tr>
<td>&gt;10</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>&gt;10</td>
<td>≤10</td>
<td>4</td>
</tr>
<tr>
<td>&gt;10</td>
<td>&gt;10</td>
<td>5</td>
</tr>
</tbody>
</table>

Although the score definition in Table 1 is derived from a bivariate nodule count, the far superior nitrogen fixation associated with crown nodulation in practice guarantees the ordinality of the scale which has been used by plant bacteriologists for almost two decades (e.g., Diatloff and Brockwell, 1976). The tap root crown is considerably thicker and has a ready supply of nutrients from foliar photosynthesis and soil, compared with the much narrower, deeper, lateral roots. Accordingly, nodules tend to be larger on the tap root compared with those on lateral roots.

No nodulation score of 1.5 was observed, eliminating this category from the analysis.

The data were analyzed using McCullagh’s ordinal regression FORTAN program, PLUM, with the parameterization

\[
\text{logit}(\gamma_{ij}) = \theta_j - (\beta_i - \bar{\beta}),
\]

\[
\beta_1 \equiv 0.
\]

The scaled deviance difference for this model was 212.4 on 20 degrees of freedom (df) indicating significant effects of rhizobia strain. The residual scaled deviance was 82.3 on 120 df, consistent with the model assumption of multinomial sampling and the adequacy of a location model. No significant change in scaled deviance resulted from the inclusion of scale parameters for rhizobia strains. Parameter estimates for the fitted model (5) are presented in Table 2.

Reparameterization of (5) according to

\[ \alpha_i = \beta_i - \bar{\beta}. \]

allows calculation of LSIs defined by (2) and (4). The maximization in (4) was performed by complete enumeration of standard errors and pair-wise LSDs based on the inverse observed information matrix.

Estimates and uncertainty interval end-points were then back-transformed to the score scale via (1) for graphical presentation (Figure 1). Fifteen of the 18 sites had mean nodulation score estimates in the region of 4. One of these, Site 1, was significantly (at the 5% level, with no adjustment for multiple comparisons) higher than one of the commercial strains.
Table 2
Maximum likelihood estimates of ordinal regression model parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cut Points</td>
<td></td>
<td></td>
</tr>
<tr>
<td>θ₁</td>
<td>-6.36</td>
<td>0.49</td>
</tr>
<tr>
<td>θ₂</td>
<td>-6.08</td>
<td>0.46</td>
</tr>
<tr>
<td>θ₃</td>
<td>-5.90</td>
<td>0.45</td>
</tr>
<tr>
<td>θ₄</td>
<td>-3.42</td>
<td>0.25</td>
</tr>
<tr>
<td>θ₅</td>
<td>-3.07</td>
<td>0.22</td>
</tr>
<tr>
<td>θ₆</td>
<td>-0.84</td>
<td>0.12</td>
</tr>
<tr>
<td>θ₇</td>
<td>0.73</td>
<td>0.12</td>
</tr>
<tr>
<td>C(WU95)</td>
<td>β₁</td>
<td>0</td>
</tr>
<tr>
<td>B(TA1)</td>
<td>β₂</td>
<td>0.33</td>
</tr>
<tr>
<td>P(cc2483g)</td>
<td>β₃</td>
<td>-0.18</td>
</tr>
<tr>
<td>Site 1</td>
<td>β₄</td>
<td>1.15</td>
</tr>
<tr>
<td>Site 2</td>
<td>β₅</td>
<td>0.53</td>
</tr>
<tr>
<td>Site 3</td>
<td>β₆</td>
<td>-0.29</td>
</tr>
<tr>
<td>Site 4</td>
<td>β₇</td>
<td>0.33</td>
</tr>
<tr>
<td>Site 5</td>
<td>β₈</td>
<td>0.46</td>
</tr>
<tr>
<td>Site 6</td>
<td>β₉</td>
<td>0.20</td>
</tr>
<tr>
<td>Site 7</td>
<td>β₁₀</td>
<td>0.51</td>
</tr>
<tr>
<td>Site 8</td>
<td>β₁₁</td>
<td>0.36</td>
</tr>
<tr>
<td>Site 9</td>
<td>β₁₂</td>
<td>0.38</td>
</tr>
<tr>
<td>Site 10</td>
<td>β₁₃</td>
<td>0.33</td>
</tr>
<tr>
<td>Site 11</td>
<td>β₁₄</td>
<td>0.09</td>
</tr>
<tr>
<td>Site 12</td>
<td>β₁₅</td>
<td>-0.02</td>
</tr>
<tr>
<td>Site 13</td>
<td>β₁₆</td>
<td>-0.11</td>
</tr>
<tr>
<td>Site 14</td>
<td>β₁₇</td>
<td>-0.90</td>
</tr>
<tr>
<td>Site 15</td>
<td>β₁₈</td>
<td>-2.35</td>
</tr>
<tr>
<td>Site 16</td>
<td>β₁₉</td>
<td>-0.68</td>
</tr>
<tr>
<td>Site 17</td>
<td>β₂₀</td>
<td>-5.42</td>
</tr>
<tr>
<td>Site 18</td>
<td>β₂₁</td>
<td>-5.17</td>
</tr>
</tbody>
</table>

Figure 1. Nodulation score estimates and least significant intervals (5% level) for three commercial rhizobia strains and eighteen rhizobia populations collected from sites in southwest Victoria, Australia.
4. Discussion
For the present model and data, the percentage error in LSI, defined as

\[
\frac{100(\delta_i + \delta_{i'} - \text{LSD}_{i,i'})}{\text{LSD}_{i,i'}}
\]

was computed for each \(i, i'\) pair. Two hundred and seven (out of the 210) of these percent-error values fell within [0\%, 5\%]. Two of the remaining three were close to 7\%, and the other, pertaining to the contrast between Sites 17 and 18, was 36\%. Standard errors of difference were calculated for all pair-wise treatment contrasts based on the inverse observed information matrix. These were all of a similar magnitude, as could be expected for equally replicated treatments, except for contrasts associated with sites 17 and 18, which were a little larger (cf. Table 2).

The degree of conservatism of the LSIs depends upon the covariance structure of the parameter estimates, through (3). The parameter estimates, \(\hat{\alpha}_i\), for Sites 17 and 18 were more highly correlated (\(\hat{\rho} = 0.64\)) than for all other treatment pairs (\(\hat{\rho} < 0.51\)). Apart from this one contrast for which the LSI was excessively conservative, \(\{\delta_i + \delta_{i'}\}\) provided close approximations (exact or slightly conservative) to the pairwise LSDs. The excessive conservatism is inconsequential in this example as these two sites had very similar nodulation scores and were very different from the other rhizobia strains.

The proposed method offers an effective graphical device for the presentation of results from ordinal regression analyses. However, the method should not be employed uncritically. Our example provides an indication that interval widths can be unduly conservative in the presence of unevenly correlated parameter estimates. With Snee (1981), we emphasise that such intervals are intended to complement more precise contrast inferences, not to replace them.

Acknowledgements

The authors are grateful to John Reynolds, Graham Hepworth, an unknown referee, and an associate editor for their helpful comments and suggestions. The work was supported by Agriculture Victoria.

Résumé


References


Received November 1994; revised April 1995; accepted June 1995.