# Ordinal Logistic Regression models and Statistical Software: What You Need to Know

Stephen Parry

## 1 Overview

Ordinal logistic regression is a statistical analysis method that can be used to model the relationship between an ordinal response variable and one or more explanatory variables. An ordinal variable is a categorical variable for which there is a clear ordering of the category levels. The explanatory variables may be either continuous or categorical. Estimating ordinal logistic regression models with statistical software is not difficult, but the interpretation of the model output can be cumbersome.

Ordinal logistic regression is an extension of logistic regression  where the logit (i.e. the log odds) of a binary response is linearly related to the independent variables. If instead the response variable has $k$ levels, then there are $k$-1 logits. A major assumption of ordinal logistic regression is the assumption of proportional odds: the effect of an independent variable is constant for each increase in the level of the response. Hence the output of an ordinal logistic regression will contain an intercept for each level of the response except one, and a single slope for each explanatory variable.

There are several ways in which an ordinal regression model can be parameterized and different statistical software packages use different parameterizations. Thus, great care should be taken when interpreting the output from ordinal regression models. We will consider an example to illustrate the different model parameterizations and corresponding interpretation for several commonly used statistical software packages.

## 2 Example dataset

Suppose that customers at a bedding store are asked to rate how comfortable they find a newly engineered mattress on a scale from 1 to 3; 1 for uncomfortable, 2 for comfortable, 3 for very comfortable. The categorical explanatory variable of interest is the gender of the respondent; 0 for female, 1 for male. The simulated dataset consists of 400 total observations. Table 2.1 displays the number and proportion of participants within each gender responding with each of the rating categories.

*Table 2.1: Number and proportion of females and males who responded in each rating category.*

| | Female (0) | Male (1) |
|---|---|---|
| Uncomfortable (1) | 28 (0.136) | 30 (0.155) |
| Comfortable (2) | 63 (0.306) | 64 (0.33) |

| | Female (0) | Male (1) |
|---|---|---|
| Very Comfortable (3) | 115 (0.558) | 100 (0.515) |

# 3 Parameterizations of ordinal logistic regression

A cumulative logit parameterization is used in ordinal logistic regression models. However, there are several ways in which this can be done. Table 3.1 shows the common parameterizations for the cumulative logit model, where *J* represents the number of levels in the categorical response variable, and *p* represents the number of explanatory variables. The most common parameterizations are models 1 and 2 where the outcome of interest is observing "Y less than or equal to *j*" where *j* is one of the ordered categories the response variable. For model 3, the cumulative logit parameterization specifies that the outcome of interest is observing "Y greater than *j*". Regardless of the parameterization, the model will have *J*-1 cutoffs (also referred to as intercepts or threshold values), denoted by $\alpha_j$ in the parameterizations below, and one parameter for each explanatory variable. This allows for the intercept to vary for each cumulative logit. However, the model assumes that each explanatory variable exerts the same effect on each cumulative logit. This is why the ordinal logistic regression model is also known as a proportional-odds model.

*Table 3.1: Three parameterizations of the ordinal logistic regression model.*

| | Parameterization |
|---|---|
| Model 1 | $\log\left(\frac{P(Y \le j)}{1-P(Y \le j)}\right) = \alpha_j - \left(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p\right),\ j = 1, \ldots, J-1$ |
| Model 2 | $\log\left(\frac{P(Y \le j)}{1-P(Y \le j)}\right) = \alpha_j + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p,\ j = 1, \ldots, J-1$ |
| Model 3 | $\log\left(\frac{P(Y > j)}{1-P(Y > j)}\right) = \alpha_j + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p,\ j = 2, \ldots, J-1, J$ |

Model 1 incorporates a negative sign so that there is a direct correspondence between the slope and the ranking. Thus a positive coefficient indicates that as the value of the explanatory variable increases, the likelihood of a higher ranking increases. This is also the case for the parameterization of model 3, but notice that the intercepts will differ between model 1 and model 3.

# 4 Software packages for fitting ordinal logistic regression

Ordinal logistic regression models can be estimated in most statistical software packages. Some possible implementations include:

- SAS: proc logistic or proc genmod
- R: clm in the "ordinal" package, vglm in the "VGAM" package, polr in the "MASS" package, and lrm in the "rms" package
- Stata: ologit command
- JMP: fit model menu with the response variable classified as ordinal
- SPSS: generalized linear model menu or the ordinal regression menu

Besides knowing the parameterization of the cumulative logit implemented by a software package, a researcher must also be aware of the coding scheme and choice of reference level for categorical explanatory variables. R, Stata, SPSS, and SAS (using proc genmod) use dummy coding, while JMP and SAS (using proc logistic) use effect coding . Both R and Stata use the first level alphanumerically as the reference level, whereas SAS, JMP, and SPSS use the last level as the reference level. However, it is possible to customize the reference level in each of these programs.

*Table 4.1: Output for models 1, 2, and 3 in different software packages.*

| | Stata, R (polr or clm) | R (vglm) | R (lrm) | SPSS | JMP or SAS (proc logistic) | SAS (proc genmod) |
|---|---|---|---|---|---|---|
| Model: | 1 | 2 | 3 | 1 | 2 | 2 |
| Coding: | Dummy | Dummy | Dummy | Dummy | Effect | Dummy |
| Threshold 1, $\alpha_1$: | $-1.858$ | $-1.858$ | 1.858 | $-1.690$ | $-1.774$ | $-1.691$ |
| Threshold 2, $\alpha_2$: | $-0.232$ | $-0.232$ | 0.232 | $-0.064$ | $-0.148$ | $-0.064$ |
| coefficient for Gender=1 indicator | $-0.168$ | 0.168 | $-0.168$ | *na* | *na* | *na* |
| coefficient for Gender=0 indicator | *na* | *na* | *na* | 0.168 | $-0.084$ | $-0.168$ |

# 5 Model interpretation

As an example, using the Stata output we can write the functional form of the ordinal regression as follows:

$$\log\left(\frac{P(Y \leq 1)}{1 - P(Y \leq 1)}\right) = -1.858 + 0.168 * \text{Gender}$$

One way to interpret the coefficients is via a proportional odds ratio. The model parameterization dictates the interpretation of the odds ratio. Using Stata's estimates, the odds ratio for gender is $\exp(-\beta_1) = \exp(0.168) = 1.18$. Thus the odds of rating a lower score is 1.18 times higher for man than it is for women.

In R (vglm), the same interpretation holds but the odds ratio is computed by exponentiating the parameter estimate without adding the negative sign: $\exp(\beta_1) = \exp(0.168) = 1.18$.

However, for SAS proc genmod we would say that the odds of women rating a mattress with a higher score is 0.84 times as large as it is for men: $\exp(\beta_1) = \exp(-0.168) = 0.84$. Note this is the same interpretation as above because we are dividing the odds for women by the odds for men, and $0.84 = 1/1.18$.

# 6 Predicted probabilities and proportional odds assumption

As in binary logistic regression, we can compute predicted probabilities in an ordinal logistic regression. For example, using the Model 2 parameterization,

$$\log\left(\frac{P(Y \leq j)}{1 - P(Y \leq j)}\right) = \alpha_j + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p,$$

the predicted probabilities are

$$P(Y \leq j) = \frac{e^{\alpha_j + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p}}{1 + e^{\alpha_j + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p}}.$$

When the assumption of proportional odds is satisfied, the predicted probabilities from the model will be similar to the observed proportions. Table 6.1 shows the predicted probabilities from the ordinal logistic regression model as well as the observed proportions (in parentheses) of each ratings within each gender. Note that although the model outputs in Table 4.1 are different due to the parameterizations used by each software package, they all agree in interpretation and estimate the same predicted probabilities.

*Table 6.1: Predicted probability of each rating for males and females along with observed proportions (in parentheses).*

| | Female (0) | Male (1) |
|---|---|---|
| Uncomfortable (1) | 0.135 (0.136) | 0.156 (0.155) |
| Comfortable (2) | 0.307 (0.306) | 0.328 (0.33) |
| Very Comfortable (3) | 0.558 (0.558) | 0.516 (0.515) |

Tests are available to assess the assumption of proportional odds. In Stata, the brant command applied after an ordinal logistic model provides one method for testing the assumption of proportional odds. In R, the `nominal_test()` function in the ordinal package can be used to test this assumption. SAS includes the test for the proportional odds assumption automatically in the output, as does SPSS's ordinal regression menu. JMP does not offer a test of proportional odds. In the absence of a test, one can fit both an ordinal logistic regression and a multinomial logistic regression to compare the AIC values. If the proportional odds assumption is not met, one can use a multinomial logistic regression model, an adjacent-categories logistic model, or a partial proportional odds model.

# 7 References

Agresti, Alan. "Categorical Data Analysis." New York: Wiley, 2002.

Le, Chap T. "Applied Categorical Data Analysis." New York: Wiley, 1998.

---

Created June 2016. Last updated September 2024.