Supplemental Instructions for Ordinal Regression

I use the VGAM library in R for the ordinal analyses. The program defines a model as being "forward" or "reverse" in ways that I sometimes find counter-intuitive but I do not deviate from the conventions of the VGAM package. In this supplement, I describe the different ordinal models and the forward and reverse variations of them. When describing the models, I use the symbol π to refer to a population probability and the notation π_j to refer to the probability associated with category "j" on the response scale of the outcome. I will use a four category outcome and two predictors to characterize the models.

The Adjacent Category Model

For the adjacent category model, the "pairs" comprising each binary regression are defined by the adjacent categories of the outcome, i.e., category 2 versus category 1; category 3 versus category 2; and category 4 versus category 3. The focus of the binary modeling is comparing the probability of being in category 2 versus the probability of being in category 1 (i.e., π_2 / π_1), the probability of being in category 3 versus the probability of being in category 2 (i.e., π_3 / π_2), and the probability of being in category 4 versus the probability of being in category 3 (π_4 / π_3). Note that when comparing probabilities, we do so using ratios rather than differences, so when the ratio is 1.00, the probabilities in the two categories are equal. For the comparison π_2 versus π_1, if the ratio is 2.0, then the probability of being in category 2 is twice that of being in category 1. If the ratio is 0.50, then the probability of being in category 2 is half that of being in category 1. And so on. For statistical reasons, we actually model the log of these ratios and in this case, there are three equations:

Category 2 vs. Category 1: $\text{Log}\,(\pi_2 / \pi_1) = \alpha_1 + \beta_1 X_1 + \beta_2 X_2$ [1]

Category 3 vs. Category 2: $\text{Log}\,(\pi_3 / \pi_2) = \alpha_2 + \beta_3 X_1 + \beta_4 X_2$ [2]

Category 4 vs. Category 3: $\text{Log}\,(\pi_4 / \pi_3) = \alpha_3 + \beta_5 X_1 + \beta_6 X_2$ [3]

where X_1 and X_2 are the predictor variables.

Adjacent category models are often said to model odds. Technically, an odds is a probability divided by one minus that probability, so odds are conceptually distinct from probabilities. A probability of 0.67 translates into an odds of 0.67/(1-.67) = 2.0. The adjacent category method "localizes" the analysis to individuals in the two target categories by focusing only on the individuals in those two categories. For the "2 versus 1" equation, my focus is on $\pi_2 / (1 - \pi_2) = \pi_2 / \pi_1$. Because the definitions of π in the adjacent category model are localized to just individuals in the two target categories, the adjacent category model is said to focus on ***local odds***.

For the **forward model**, the higher of the two categories is assigned a 1 in a given binary regression equation and the lower of the two categories is assigned a value of 0. For the **reverse model**, the opposite is true, i.e., the lower of the two categories is assigned a 1 in a given binary regression equation and the higher of the two categories is assigned a value of 0.

The Proportional Odds Model

For a proportional odds model, the "pairs" within the series of binary regressions are defined differently than in the adjacent category model. A pair of categories in the proportional odds model is defined by collapsing categories on the outcome metric. A pair is a given target category and all the categories below it versus all categories above the target category. For an outcome with 4 categories, one "pair" is category 1 versus categories 2, 3 and 4 combined. Another pair is categories 1 and 2 combined versus categories 3 and 4 combined. The final pair is categories 1, 2, and 3 combined versus category 4.

The **forward model** is defined by the following equations:

1 versus 2, 3, 4: $\text{Log}\,[\pi_1 / (\pi_2 + \pi_3 + \pi_4)] = \alpha_1 + \beta_1 X_1 + \beta_2 X_2$

1, 2 versus 3, 4: $\text{Log}\,[(\pi_1 + \pi_2) / (\pi_3 + \pi_4)] = \alpha_2 + \beta_3 X_1 + \beta_4 X_2$

1, 2, 3 versus 4: $\text{Log}\,[(\pi_1 + \pi_2 + \pi_3) / (\pi_4)] = \alpha_3 + \beta_5 X_1 + \beta_6 X_2$

Where the categories to the left of "versus" are assigned 1 and the categories to the rig hog "versus are assigned 0 for a given binary regression. Note the proportional odds model does ***not*** work with local odds because all individuals are included or involved in each binary regression. We are not segregating out only individuals in two categories in any of the equations. Despite this, the pair of categories for any given equation is still defined in a way that the ratio of the "pair" of probabilities represents an odds, i.e., a person is either in one member of the pair or the other member of the pair. In this sense, the model focuses on odds.

The **reverse model** uses the following ratios instead:

2, 3, 4 versus 1: $\text{Log}\,[(\pi_2 + \pi_3 + \pi_4) / (\pi_1)] = \alpha_1 + \beta_1 X_1 + \beta_2 X_2$

3, 4 versus 1, 2: $\text{Log}\,[(\pi_3 + \pi_4) / (\pi_1 + \pi_2)] = \alpha_2 + \beta_3 X_1 + \beta_4 X_2$

4 versus 1, 2, 3: $\text{Log}\,[(\pi_4) / (\pi_1 + \pi_2 + \pi_3)] = \alpha_3 + \beta_5 X_1 + \beta_6 X_2$

The reverse model simply reverses the numerator and denominator of the forward model.

The Forward-Stopping Model

Each equation in this model focuses on a specific category, j. A pair is defined by the probability associated with category j, π_j, versus the combined π for all categories greater than j. This yields the following pairings for a four category outcome, expressed as ratios:

1 versus 2, 3, and 4: $\pi_1 / (\pi_2 + \pi_3 + \pi_4)$

2 versus 3 and 4: $\pi_2 / (\pi_3 + \pi_4)$

3 versus 4: π_3 / π_4

Traditionally, one models the log of each ratio as a function of the predictors, X, i.e., we use a logit function.

If the categories of your outcome represent a forward progression on some dimension of interest, then this model essentially examines the odds of stopping or "stagnating" in that forward sequence at a certain "stage" or category level. For example, suppose the outcome reflects different levels of mastery of mathematics, with higher categories/numbers indicating greater levels of mastery. We want to model the odds that people will stagnate at category 1 (versus move forward), the odds they will stagnate at category 2 (versus move forward from there), and the odds they will stagnate at category 3 versus move forward from there. The above equations accomplish this.

The category that is the target of the stop is scored 1 and the other categories are scored 0. So, the intercept in the first equation is the log odds of stopping or stagnating in category 1 when all predictors equal 0; the intercept in the second equation is the log odds of stopping or stagnating in category 2 when all predictors equal 0; and the intercept in the third equation is the log odds of stopping or stagnating in category 3 when all predictors equal 0.

The coefficient for a predictor indicates how much the log odds of stagnating at a given category increases given a one unit increase in the predictor.

Reverse-Stopping Model

This model focuses on category, j+1. A given pair is defined by π for category j+1 versus the combined π for all categories less than j+1. This yields the following unique pairings for a four category outcome:

2 versus 1: $\pi_2 / (\pi_1)$

3 versus 2 and 1: $\pi_3 / (\pi_1 + \pi_2)$

4 versus 3, 2, and 1: $\pi_4 / (\pi_1 + \pi_2 + \pi_3)$

Again, we typically model the log of each ratio as a function of the predictors, i.e., we use a logit function.

This model applies to a sequence that moves from the highest category down to the lowest category and the direction of stopping or "stagnating" in a category is from high to low rather than low to high. For example, the outcome might be an index of symptom severity, where 1 = no symptoms, 2 = mild symptoms, 3 = moderate symptoms, and 4 = severe symptoms. When a health professional treats people who have the illness in question, we might be interested in the likelihood people will "stagnate" in category 4 (severe symptoms), the likelihood they will "stagnate" in category 3 (moderate symptoms), and the likelihood they will "stagnate" in category 2 (mild symptoms). We might model these processes as a function of being in a treatment versus a control group, which would then be a dummy coded predictor in each equation.

The coefficient for a predictor indicates how much the log odds of stagnating at a given category increases given a one unit increase in the predictor.

Forward-Continuation Model

This model focuses on category, j. A given pair is defined by π for the combined categories greater than j versus the π for j. This yields the following unique pairings, beginning with category j=1 and increasing j by one until all pairings have been specified:

2, 3, 4 versus 1: $(\pi_2 + \pi_3 + \pi_4) / (\pi_1)$

3, 4 versus 2: $(\pi_3 + \pi_4) / (\pi_2)$

4 versus 3 $(\pi_4) / (\pi_3)$

One models the log of each ratio as a function of the predictors, X, using the logit function. Note this model is the same as the forward stopping model, but with the numerator and denominators of the ratios for each equation switched.

Reverse-Continuation Model

This model focuses on a specific category, j+1. A given pair is defined by π for the combined categories less than j+1 versus the π for category j+1. This yields the following unique ratio pairings, beginning with category j=1 and increasing j by one until all pairings have been specified:

1 versus 2: $(\pi_1) / (\pi_2)$

1, 2 versus 3: $(\pi_1 + \pi_2) / (\pi_3)$

1, 2, 3 versus 4: $(\pi_1 + \pi_2 + \pi_3) / (\pi_4)$

One models the log of each ratio as a function of the predictors, X, using a logit function. Note this model is the same as the reverse stopping model, but with the numerator and denominators of the ratios for each equation switched.