Ordinal Outcomes with the Continuation Ratio Model

Melissa J. McGowan, SM Statistical Consulting, Media, PA

ABSTRACT

Ordinal regression is a relatively new statistical method developed for analyzing ranked outcomes. In the past, ranked scales have often been analyzed without making full use of the ordinality of the data, or alternatively, by assigning arbitrary numerical scores to the ranks. While ordinal regression models are now available to make full use of ranked data, they are not used widely. This poster will illustrate the application of the Continuation Ratio model for an ordinal outcome representing stage of retention. The Continuation Ratio model is appropriate when the ordered categories represent a progression through stages, so those individuals must pass through each lower stage before they go on to higher stages.

INTRODUCTION

Ordinal regression is a relatively new statistical method developed for analyzing ranked outcomes. In the past, rank scales have often been analyzed without making full use of the ordinality of the data. Ordinal scales are often treated as nominal scales, with proportions calculated for each level of outcome. Ordinal outcomes may also be analyzed with binary logistic regression. Again, this method discards information, as it requires that the ordinal outcomes be forced into two levels. Polytomous logistic regression models can be used to accommodate more than two levels of outcome but do so without incorporating information on order (Thomas, 1986). In summary, methods based on either reducing ordinal scales to nominal or dichotomous ones, or assuming that ordinal scales have the properties of interval scales, have several drawbacks and may lead to erroneous statistical inferences. These approaches are still being used, even though statistical theory and software have advances sufficiently to permit exploiting the ordinal nature of the data. Statistically powerful methods, referred to as "ordinal regression models" have been designed to take full advantage of ordinal outcomes (Scott, 1997).

CONTINUATION RATIO MODEL

The Continuation Ratio model is linear and additive on the logit scale and uses maximum likelihood methods to estimate a summary odds ratio. The summary odds ratio based on the continuation ratio model represents among subjects in level j, the odds of having an outcome greater than level j relative to being in level j. The continuation ratio model is based on "conditional incremental cutpoints", with outcomes at a given level discarded after being compared to higher levels. By viewing the outcome as going from more severe to less severe, the model can be applied in reverse. However, because of the conditioning on adjacent cutpoints, the continuation ratio model is affected by the direction chosen for modeling the response variable. A test of heterogeneity over cutpoint is also available for the continuation ratio model.

COMPUTATIONAL ISSUES

Any standard statistical package that performs binary logistic regression can be used. The original data set must be restructured to account for transition through stages (Armstrong, 1989). This new data set is created by repeatedly including subset of all observations contributing to each respective cutpoint. Two new variables must also be added to this data set, one indicating the cutpoint from which the particular subset arose and a second, a binary variable indicating the dichotomous status of the outcome at that cutpoint.

If significant heterogeneity of an effect over cutpoints is found, the continuation ratio model can be easily adapted to allow effects to vary by cutpoint. Including the interaction of the effect with the cutpoint performs this.

EXAMPLE

A total of 1764 patients met basic inclusion criteria and were invited for an intake visit. Fifty-two percent (924) of the eligible patients attended an intake visit and most (870 patients) began orientation by attending another visit after the first intake. Of the 924 patients who came to the first intake visit, 54 did not return. Of the 870 patients who started orientation phase, 254 (29%) did not complete the attendance requirements and 129 patients (15%) did not complete the assessment requirements, leaving a final sample of 487 randomized patients. Are there indicators of which patients were randomized and those who were not? Potential predictors analyzed were Cocaine usage (Binge vs. Chronic) and Gender (Female vs. Male).

Our outcome consists of three stages: Stage 1 – Does the patient attend intake? Stage 2 – Does the patient attend orientation? Stage 3 – Is the patient randomized? Each stage represents an ordered category, so that each patient must pass through the lower stage before they go onto the higher stage. In this situation, the Continuation Ratio ordinal regression model exploits the ordinality and structure of the outcome measure (Allison, 1997).

The Continuation Ratio model examines the probability the patient makes it to the next stage given their current stage status. It is comparable to Cox's proportional hazards model in discrete time (Cox, 1988). A separate intercept is modeled for each stage; therefore, an odds ratio of increasing to the next stage can be acquired. Each covariate was entered separately in the model. Homogeneity of covariates over the stages is assumed. If significant heterogeneity is found, it is modeled.

The Continuation Ratio model was implemented on the restructured data set with PROC GENMOD (Stokes, 1995). The GLIMMIX macro of PROC MIXED was used to calculate proportion estimates per group from the generated risk ratios (Little, 1996).

SYNTAX FOR RESTRUCTING THE DATA

The following syntax illustrates how to restructure the data set over the various cutpoints. The initial data set PTSTAT contains the Intake date, Orientation date, and the Randomization date. All dates are filled out if patients progress that far; therefore, a missing date indicates the patient dropped from the study prior to the corresponding point. PTSTAT also contains our two measures of interest: GENDER and USAGE.

```
DATA EXAMINE;
SET PTSTAT;
POINT = 0;
IF INTAKE NE . THEN POINT = 1;
IF ORIENT NE . THEN POINT = 2;
IF RANDDATE NE . THEN POINT = 3;
RUN;

DATA CRDATA;
SET EXAMINE;
STAGE = 1;
ADVANCE = (POINT GE 1);
OUTPUT;
STAGE = 2;
ADVANCE = (POINT GE 2);
IF POINT = 0 THEN DELETE;
OUTPUT;
STAGE = 3;
ADVANCE = (POINT GE 3);
IF POINT LE 1 THEN DELETE;
OUTPUT;
RUN;
```

SYNTAX FOR MODELING HOMOGENEITY OVER CUTPOINTS

The following syntax illustrates how to model the continuation ratio model with PROC GENMOD and the GLIMMIX Macro, where the GLIMMIX macro will produce the average estimated proportions for each group. Syntax is illustrates for modeling the effect due to GENDER. Homogeneity is assumed across cutpoint.

```
PROC GENMOD DATA=CRDATA;
CLASS STAGE GENDER;
MODEL ADVANCE = STAGE GENDER/D=BINOMIAL
WALD TYPE3;
QUIT;

%INC 'C:\SAS\STAT\SAMPLE\GLIMMIX.SAS';
%GLIMMIX(DATA=CRDATA,
STMTS=%STR(
CLASS STAGE GENDER;
MODEL ADVANCE= STAGE GENDER/SOLUTION;
LSMEANS STAGE RACE/PDIFF;
), ERROR=BINOMIAL);
QUIT;
```

SYNTAX FOR MODELING HETEROGENEITY OVER CUTPOINTS

The following syntax illustrates how to model the continuation ratio model with PROC GENMOD and the GLIMMIX Macro, where the GLIMMIX macro will produce the estimated proportions for each group over the three stages: Intaked, Oriented, and Randomized.

```
PROC GENMOD DATA=CRDATA;
CLASS STAGE GENDER;
MODEL ADVANCE = STAGE|GENDER/D=BINOMIAL
WALD TYPE3;
QUIT;

%INC 'C:\SAS\STAT\SAMPLE\GLIMMIX.SAS';
%GLIMMIX(DATA=CRDATA,
STMTS=%STR(
CLASS STAGE GENDER;
MODEL ADVANCE= STAGE|GENDER/SOLUTION;
LSMEANS STAGE GENDER/PDIFF;
), ERROR=BINOMIAL);
QUIT;
```

RESULTS

Table 1 illustrates the number of patients who do not progress to the next stage and the proportion of available patients in the corresponding stage that do.

Table 1. Stages from Screening through Randomization.

Stages	Nr. Patients Per stage who Do not advance	Proportion of patients who progress
Screened And Eligible	840	52.5%
Intaked	54	94.1%
Oriented	383	56.0%
Randomized	487	N/A

Gender is a significant predictor across stage of retention (Chi-square(1)=18.52, $p < 0.0001$). The effect of Gender was found to be homogeneous across stage (Chi-square(2)=1.62, p=0.44). On average, 68.9% of the available Males were expected to progress to the next stage, whereas 75.9% of Females were expected to progress to the next stage

For Cocaine Usage, we found the effects of progression through each stage was differential across stages (Chi-square(2) = 10.79, p=0.005). For usage, a higher proportion of Binge users are intaked, whereas a higher proportion of Chronic users are oriented.

Figure 1 illustrates the homogeneity for Gender across stages and the heterogeneity for Usage across stages.

Table 2 summarizes the proportion of available patients achieving the indicated state. Proportions were acquired via the Solution option in the GLIMMIX macro.

Table 2. Estimated proportions within each stage

Effect	Intaked	Oriented	Randomized
Gender			
Female	0.474 [a]	0.919 [a]	0.471 [a]
Male	0.548 [b]	0.933 [b]	0.597 [b]
Usage			
Binge	0.574 [a]	0.924 [a]	0.594 [a]
Chronic	0.493 [b]	0.959 [b]	0.543 [a]

Note: Comparisons are done separately for each column (Intaked, Oriented, and Randomized) and predictor (Gender and Usage). Estimated proportions followed by the same letter are not statistically different ($p>0.05$).

REFERENCES

Allison, P.D. (1997). *Logit and Loglinear Analysis Using the SAS System.* Cary, NC: SAS Institute Inc.

Armstrong, B.G., and Sloan, M. (1989). Ordinal regression models for epidemiological data. *American Journal of Epidemiology*, 129: 191-204.

Cox, C. (1988). Multinomial regression models based on continuation ratios. *Statistics in Medicine*, 7, 435-441.

Littell, R.C., Milliken, G.A., Stroup, W.W., and Wolfinger, R.D. (1996). *SAS System for Mixed Models*. Cary, NC: SAS Institute Inc.

Scott, S.C., Goldberg, M., and Mayo, N.E. (1997). Statistical Assessment of Ordinal Outcome sin Comparative Studies. *Journal of Clinical Epidemiology,* 50: 45-55.

Stokes, M.E., Davis, C.S., and Koch, G.G. (1995). *Categorical Data Analysis using the SAS System*. Cary, NC: SAS Institute Inc.

Thomas, D.C., Goldberg, M., and Dewart, R. (1986). Statistical methods for relating several exposure factors to several diseases in case-heterogeneity studies. *Statistics in Medicine,* 5, 49-60.

ACKNOWLEDGEMENTS

The author would like to express her appreciation to the following people for their assistance in this paper:

Robert J. Gallop, Dept. of Biostatistics, University of Pennsylvania, Philadelphia, PA

Mark Goldberg, Epidemiology and Biostatistics Unit,
INRS – Institute of Armand-Frappier
Ville de Laval
University of Quebec, Quebec, Canada

CONTACT INFORMATION

Melissa J. McGowan
SM Statistical Consulting
P.O. Box 1613
Media, PA 19063
(215) 573-8533

website: www.smstat.com
e-mail: smstat@hotmail.com

Figure 1: Investigation of Retention over Stages