

## The data for an ordinal approach

### Contents

1.	INTRODUCTION .....	1
2.	SURVIVAL DATA AS ORDINAL OUTCOMES .....	2
3.	EXAMPLE.....	3
3.1	SETTING UP THE ANALYSIS .....	3
3.2	DISCUSSION OF RESULTS .....	8
3.3	INTERPRETING THE OUTPUT.....	10

## 1. Introduction

The ordinal analysis illustrated in this chapter is again based on the TVSFP data. As shown in the previous section, one can also fit grouped-time survival models using dichotomous indicators of event/censoring across the study time points. To do so, requires additional data manipulation. The data set used for the ordinal approach differs from that previously discussed, and is represented by the LISREL spreadsheet file **SMKCCLC.lsf**. The first 10 records of this data set are shown below.

	School	Class	Student	SmkOnset	Event	SexM	CC	TV	CCTV
1	193.0	193101.0	193101103.0	1.0	1.0	0.0	0.0	0.0	0.0
2	193.0	193101.0	193101105.0	4.0	0.0	1.0	0.0	0.0	0.0
3	193.0	193101.0	193101108.0	4.0	0.0	0.0	0.0	0.0	0.0
4	193.0	193101.0	193101111.0	1.0	1.0	1.0	0.0	0.0	0.0
5	193.0	193101.0	193101113.0	2.0	1.0	1.0	0.0	0.0	0.0
6	193.0	193101.0	193101117.0	2.0	1.0	1.0	0.0	0.0	0.0
7	193.0	193101.0	193101121.0	4.0	0.0	1.0	0.0	0.0	0.0
8	193.0	193101.0	193101132.0	4.0	0.0	0.0	0.0	0.0	0.0
9	193.0	193101.0	193101201.0	4.0	1.0	0.0	0.0	0.0	0.0
10	193.0	193101.0	193101204.0	4.0	0.0	0.0	0.0	0.0	0.0

The variables of interest are:

- School indicates the school a student is from.
- Class identifies the classroom to which a student belongs.
- Student represents the student identification number.
- SmkOnset indicates the time at which an event occurred. It assumes a value of 1 for a WaveA measurement (*i.e.*, the event occurred at Wave A), 2 for a WaveB measurement, 3 for a WaveC measurement, and 4 for a WaveD measurement.
- Event is an indicator variable indicating whether the subject experienced the event or was censored. A value of 1 indicates that the student did experience the event (*i.e.*, onset of cigarette experimentation) at one of the time points, while a value of 0 indicates that the subject was censored and never experienced the event (*i.e.*, no onset of cigarette experimentation) at any time point that they were assessed at.
- SexM is an indicator variable for gender, with "1" indicating male respondents, and "0" female respondents.
- CC is a binary variable indicating whether a social-resistance classroom curriculum was introduced, with 0 indicating "no" and 1 "yes."
- TV is an indicator variable for the use of media (television) intervention, with a "1" indicating the use of media intervention, and "0" the absence thereof.
- CC\*TV was constructed by multiplying the variables TV and CC, and represents the CC by TV interaction.

## 2. Survival data as ordinal outcomes

Assume 4 time points with no intermittent censoring and let  $y$  denote the ordinal outcome variable. Let us first consider subjects who initiated smoking at some point in the study. For these subjects, the variable Event will be coded as 1 and the coding of the SmkOnset variable will be as follows.

SmkOnset:

- $y_{ij} = 1$ : Student first started to smoke at  $t = 1$ .
- $y_{ij} = 2$ : Student did not smoke at  $t = 1$ , but first smoked at  $t = 2$ .
- $y_{ij} = 3$ : Student did not smoke at  $t = 1$  or 2, but first smoked at  $t = 3$ .
- $y_{ij} = 4$ : Student did not smoke at  $t = 1, 2, \text{ or } 3$ , but first smoked at  $t = 4$ .

Similarly, subjects who were censored would have the variable Event coded as 0, and the following codes for the SmkOnset variable.

SmkOnset:

- $y_{ij} = 1$ : Student did not smoke at  $t = 1$  and no data beyond  $t = 1$ .
- $y_{ij} = 2$ : Student did not smoke at  $t = 1$  or 2, and no data beyond  $t = 2$ .

- $y_{ij} = 3$ : Student did not smoke at  $t = 1, 2, \text{ or } 3$ , and no data beyond  $t = 3$  (*i.e.*, no data at  $t = 4$ ).
- $y_{ij} = 4$ : Student did not smoke at  $t = 1, 2, 3, \text{ or } 4$ .

Here, the phrase "did not smoke" is more precisely "did not answer yes to the question have you ever smoked a cigarette."

**Table 1.4: Three time points with censoring**

Outcome	Ordinal dep. Variable	Event indicator
Censor at $T_1$	1	0
Event at $T_1$	1	1
Censor at $T_2$	2	0
Event at $T_2$	2	1
Censor at $T_3$	3	0
Event at $T_3$	3	1
Censor at $T_4$	4	0
Event at $T_4$	4	1

Table 1.4 shows how values are assigned to  $y_{ij}$ , and the relationship between the  $y_{ij}$  outcomes and the event indicator.

### 3. Example

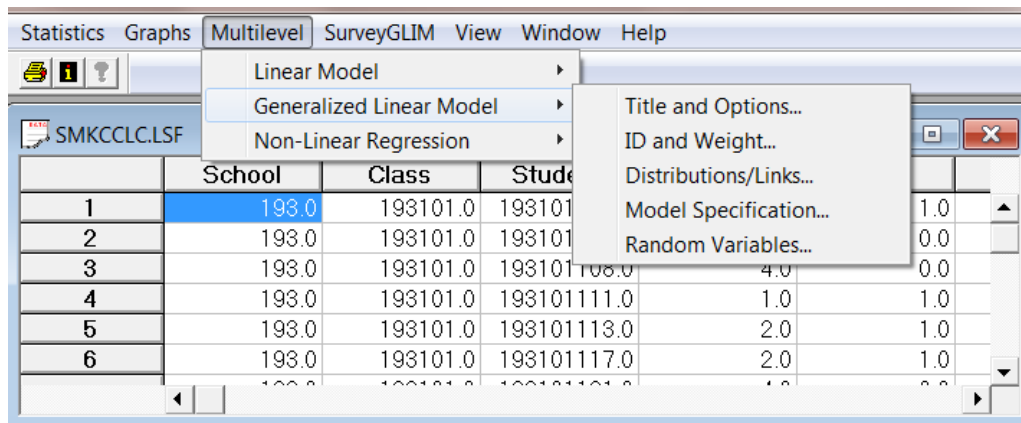
In this section, the re-formatted form of the data, as captured in **smkcclc.ss3** is imported and used to fit a model to the data with the ordinal variable SmkOnset as outcome.

The model fitted to the data is of the form

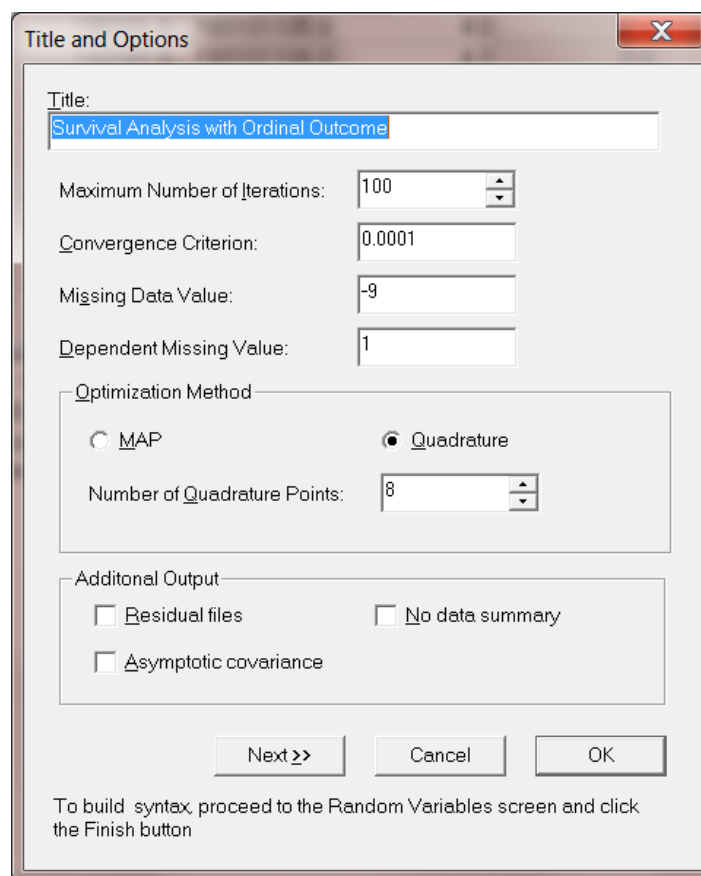
$$\log[-\log(1 - P_{ijt})] = \gamma_t + (SexM_{ij})\beta_1 + (CC_j)\beta_2 + (TV_j)\beta_3 + v_{0i}.$$

#### 3.1 Setting up the analysis

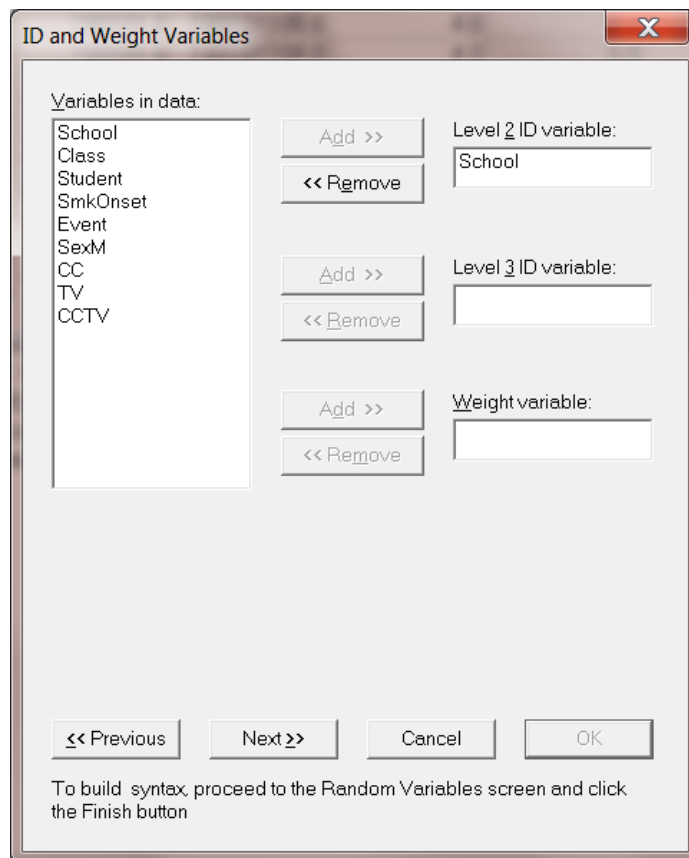
Using the data in the LISREL spreadsheet **SMKCCLC.lsf**, we start by selecting the **Multilevel, Generalized Linear Model** from the main menu bar as shown below.



Start with the **Titles and Options** dialog and enter (optional) a title in the **Title** text box. Enter a value of -9 as the global missing value and a value of 1 as the dependent variable missing value.



Specify the number of quadrature points as 8. When done, click the **Next** button to proceed to the **ID and Weight Variables** dialog shown below and select **School** as the level-2 ID variable.



Next is the **Distributions and Links** dialog shown below. Select **Multinomial** as the distribution type and **Ordinal complementary log-log** as the link function.

Choose **Add** for the model terms and then click the **Next** button to invoke the **Dependent and Independent Variables** dialog shown next.

Distributions and Links

Distribution type: Multinomial

Link function: Ordinal complementary log-log

Model terms: Add

Include intercept?

Yes  No

Dispersion parameter

Yes  Fixed value:

Estimate scale?

<< Previous Next >> Cancel OK

To build syntax, proceed to the Random Variables screen and click the Finish button

Select SmkOnset as the dependent (outcome variable) and SexM, CC, and TV are specified as the predictors (independent variables) of the fixed part of the model. Before proceeding to the **Random Variables** dialog, select Event as the **Event Variable**.

**Dependent and Independent Variables**

Variables in data:

- School
- Class
- Student
- SmkOnset
- Event
- SexM
- CC
- TV
- CCTV

Dependent variable: SmkOnset

Independent variables: SexM, CC, TV

Event Variable: Event

Buttons: Add >>, << Remove, Continuous >>, Categorical >>, << Remove, Add >>, << Remove, << Previous, Next >>, Cancel, OK

To build syntax, proceed to the Random Variables screen and click the Finish button

**Random Variables**

Variables in data:

- School
- Class
- Student
- SmkOnset
- Event
- SexM
- CC
- TV
- CCTV

Random Level 2:  Intercept

Random Level 3:  Intercept

Number of interactions: 0

Buttons: Add >>, << Remove, Add >>, << Remove, << Previous, Finish, Cancel, OK

To build syntax, click the Finish button.

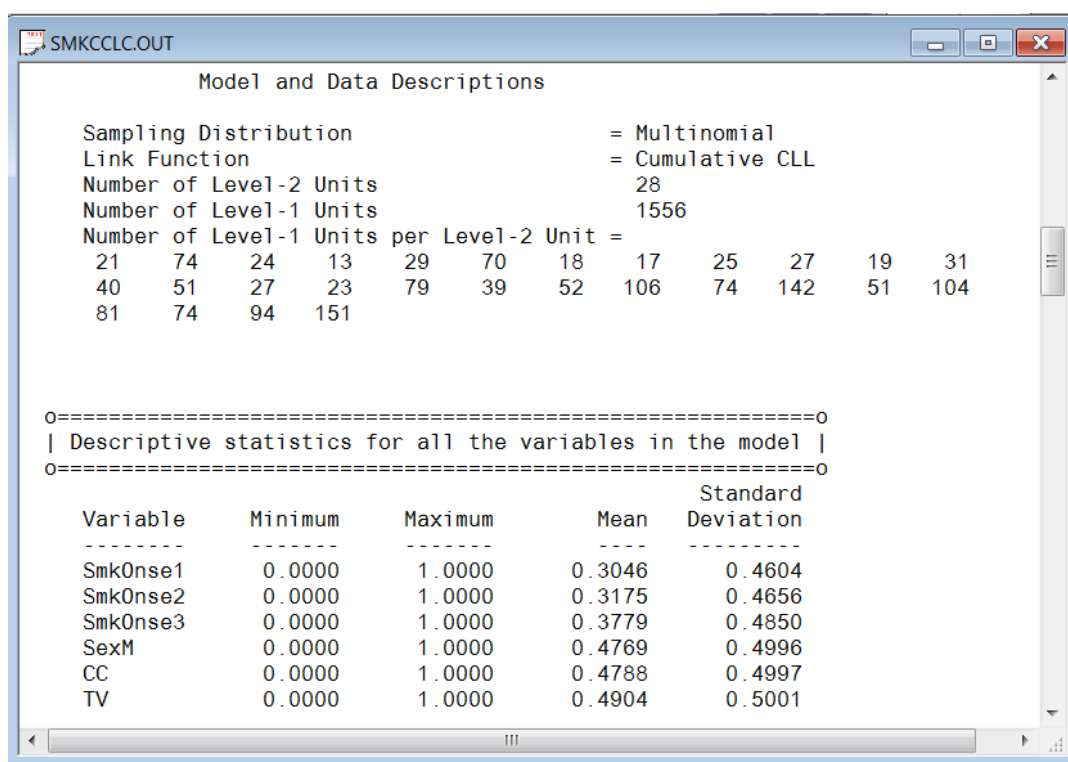
Click the **Finish** button (see above) to produce a syntax file and then click on the **Run Prelis** icon to run the analysis.

### 3.2 Discussion of results

Selected portions of the output file **smkccd1.out** are shown below.

#### Data summary and descriptive statistics

The portion of the output file shown below indicates that there are 28 schools, with 1556 students nested within these. This is followed by descriptive statistics for all the variables. Note that all three predictor variables are dichotomous in nature.



#### Fixed effects estimates

This is followed by the results for the model specified, but without any random effects. In this format, none of the included predictors are significant. It will be interesting to compare these results with those obtained once the hierarchical structure of the data has been taken into account.



```

SMKCLC.OUT
=====0
| Optimization Method: Adaptive Quadrature |
=====0

Number of quadrature points =          8
Number of free parameters =          7
Number of iterations used =          5

-2lnL (deviance statistic) =        3187.38817
Akaike Information Criterion      3201.38817
Schwarz Criterion                 3238.83729

                                Estimated regression weights

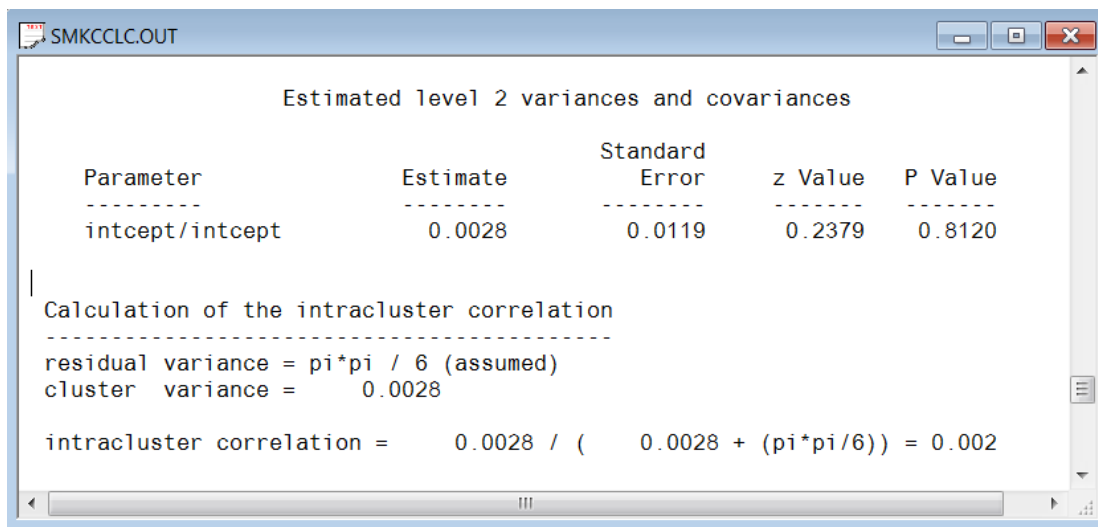
Parameter      Estimate      Standard      z Value      P Value
-----
Thresh1        -1.6564        0.0950       -17.4269     0.0000
Thresh2        -0.9431        0.0869       -10.8519     0.0000
Thresh3        -0.4313        0.0848        -5.0846     0.0000
SexM           0.0574        0.0798         0.7190     0.4722
CC             0.0449        0.0842         0.5333     0.5938
TV             0.0213        0.0833         0.2555     0.7984

```

Parameter estimates are given in the next part of the output. Taking the hierarchical structure into account and allowing for the intercept to vary randomly over the schools had little effect on the significance level of the 3 covariates: all are still non-significant. We note that the three thresholds, which represent the cumulative baseline hazard, are estimated as  $-1.6564$ ,  $-0.9431$ , and  $-0.4313$  respectively. An alternative parameterization is also given. Here, the first threshold has been set to zero and as a result, the intercept and second and third threshold estimates are calculated as  $-1.6564$ ,  $0.7133$ , and  $1.2251$  respectively.

### Random effects estimates and intraclass correlation (ICC)

This part of the output shows the estimates of the random effects and an estimate of the intraclass correlation. There is no evidence of significant random variation in the intercept over the schools ( $p = 0.8120$ ). The intraclass correlation coefficient shown is based on the use of the complementary log-log link function for these data, which results in a residual variance of  $\pi^2 / 6$  (see Agresti, 2002).



#### Population Average Estimates

Parameter	Estimate	Standard Error	z Value	P Value
Thresh1	-1.6553	0.0942	-17.5716	0.0000
Thresh2	-0.9422	0.0863	-10.9226	0.0000
Thresh3	-0.4308	0.0845	-5.0987	0.0000
SexM	0.0574	0.0798	0.7190	0.4722
CC	0.0449	0.0842	0.5334	0.5938
TV	0.0213	0.0833	0.2554	0.7984

### 3.3 Interpreting the output

#### Comparing binary and ordinal models

When the number of measurement occasions is not too large, the binary outcome model utilizing dummy variables to represent the measurement occasions can be useful in fitting survival analysis models. Additionally, the binary model easily allows relaxation of the proportional hazards assumption for model covariates through inclusion of interaction terms with the time point indicators. Finally, though not illustrated here, the binary model can also handle time-dependent covariates in the same manner as the covariate by time interactions. When the number of occasions is very large, however, the number of time point indicators that must be created for the binary model, and the resulting size of the data set, can get very large and unwieldy. In this case, the ordinal outcome model such as the model discussed in this section is perhaps the better analysis option (though covariates must follow the proportional hazards assumptions and time-dependent covariates are not allowed). If the complementary log-log link function is selected (*i.e.*, the model is specified as a proportional hazards model), the binary and ordinal outcome models yield identical estimates for parameters that do not depend on time (Lara & Matthews, 1985). This is shown in Table 1.12. The regression coefficients are exactly the same for Male, CC, and TV. This is also true of their standard errors and so the *p*-values for both sets are identical. However, the intercept and threshold parameters, which do represent time-related information, are not the same with the exception of the first intercept. The reason for this is that the intercepts in the binary model represent the *interval-specific* baseline hazard, whereas their corresponding threshold parameters in the ordinal model represent the

*cumulative* baseline hazard across the time intervals. These are only equivalent only for the first time interval and thereafter diverge in value and meaning. Finally, it should be mentioned that if one uses the logit link, in place of the complementary log-log link, the estimates (of the parameters not involving time) from the binary and ordinal models are not equivalent, though similar.

Notice also that the likelihood values for the two representations are identical, as are the AIC values. The Schwarz values are not the same because the numbers of observations in the two representations are different. That is, because the binary-case data set consists of multiple person-time indicators for each outcome, the numbers of observations in the binary-case data set is inflated, relative to the ordinal case.

**Table 1.12: Comparison of results of binary and ordinal outcome models**

Term	Binary outcome (EVENT)	Ordinal outcome (SmkOnset)
Wave B baseline hazard binary $\alpha_{01}$ or ordinal $\gamma_1$	-1.6564	-1.6564
Wave C baseline hazard binary $\alpha_{01} + \alpha_{02}$ or ordinal $\gamma_2$	$-1.65654 + 0.0399 = -1.6165$	-0.9431
Wave D baseline hazard binary $\alpha_{01} + \alpha_{03}$ or ordinal $\gamma_3$	$-1.6564 + 0.3103 = -1.3461$	-0.4313
Male $\beta_1$	0.0574	0.0574
CC $\beta_2$	0.0449	0.0449
TV $\beta_3$	0.0213	0.0213
$-2\ln L$	3187.38817	3187.38817
AIC	3201.38817	3201.38817
Schwarz	3243.94116	3238.83729
No. of parameters	7	7