

Survival analysis for an ordinal outcome

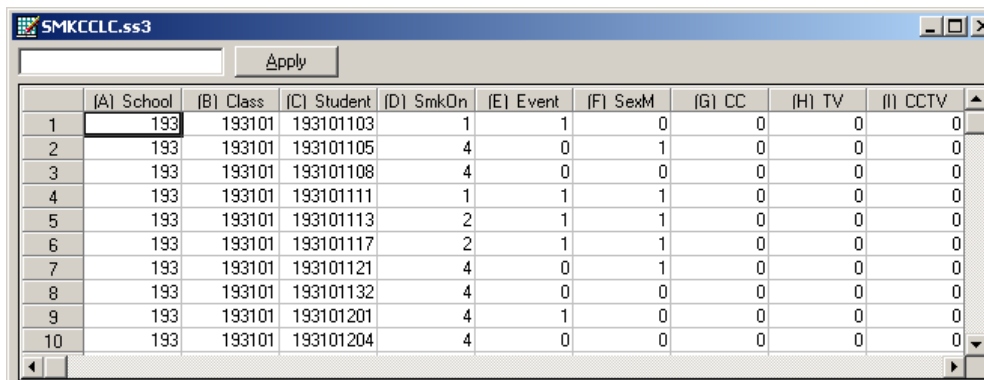
Contents

3.1	THE ORDINAL CASE	1
3.1.1	<i>The data for an ordinal approach</i>	1
3.1.2	<i>Ordinal case: 2-level model</i>	3
3.2	SURVIVAL ANALYSIS MODEL FOR AN ORDINAL OUTCOME	4
3.2.1	<i>Setting up the analysis</i>	5
3.2.2	<i>Discussion of results</i>	7
3.2.3	<i>Interpreting the output</i>	10

3.1 The ordinal case

3.1.1 The data for an ordinal approach

The ordinal analysis illustrated in this chapter is again based on the TVSF data. 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 SuperMix spreadsheet file **SMKCLC.ss3**. The first 10 records of this data set are shown below.



	(A) School	(B) Class	(C) Student	(D) SmkOn	(E) Event	(F) SexM	(G) CC	(H) TV	(I) CCTV
1	193	193101	193101103	1	1	0	0	0	0
2	193	193101	193101105	4	0	1	0	0	0
3	193	193101	193101108	4	0	0	0	0	0
4	193	193101	193101111	1	1	1	0	0	0
5	193	193101	193101113	2	1	1	0	0	0
6	193	193101	193101117	2	1	1	0	0	0
7	193	193101	193101121	4	0	1	0	0	0
8	193	193101	193101132	4	0	0	0	0	0
9	193	193101	193101201	4	1	0	0	0	0
10	193	193101	193101204	4	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.

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 8.4: 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 8.4 shows how values are assigned to y_{ij} , and the relationship between the y_{ij} outcomes and the event indicator.

3.1.2 Ordinal case: 2-level model

Let y_{ij} denote an ordinal outcome variable that takes on discrete positive values $t = 1, 2, \dots, m$. In previous examples we assumed that y_{ij} has C categories or distinct values, however here to be consistent with the survival analysis notation we will use m to represent the number of ordinal categories. The subscript (i, j) denotes subject j , $j = 1, 2, \dots, n_i$ nested within level-2 unit i , $i = 1, 2, \dots, N$. In the present context the level-1 units j indicates students and the level-2 unit i indicates schools. Note, that as another example of this type of model, one could have multiple failure times nested within individuals.

Let δ_{ij} denote the censor/event indicator, then $\delta_{ij} = 1$ if the event occurs and $\delta_{ij} = 0$ if an observation is censored. In survival analysis each ij is observed until time t_{ij} and if an event occurs $t_{ij} = t$ and $\delta_{ij} = 1$. If the observation is censored at $t_{ij} = t$ then $\delta_{ij} = 0$.

In the case of censoring it is assumed that a unit is observed at t_{ij} but not at t_{ij+1} . As described in Hedeker, Siddiqui & Hu (2000), if events occur within continuous time intervals (*i.e.*, grouped-time), for example, a student initiates smoking experimentation in the past year, use of the complementary log-log link for an ordinal outcome is equivalent to a proportional hazards model in continuous time. Therefore, the grouped-time proportional hazards mixed model can be written as:

$$\log\left[-\log\left(1 - P_{ijt}\right)\right] = \gamma_t + \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{v}_i$$

where \mathbf{x}_{ij} is a vector of explanatory variables and \mathbf{z}_{ij} a vector of random effects. Typically, the elements of \mathbf{z}_{ij} are a subset of \mathbf{x}_{ij} . For example, the elements of \mathbf{z}_{ij} might correspond to the intercept and age, whereas \mathbf{x}_{ij} would include these two terms plus any additional model covariates. It is assumed that the random effects \mathbf{v}_i are from a normal distribution with mean zero and covariance matrix $\boldsymbol{\Phi}_{(2)}$.

P_{ijt} denotes the probability that an event takes place up to and including the interval designated at time t_{ij} . Thus, P_{ijt} represents a cumulative probability of failure, whereas p_{ijt} is the interval-specific failure probability. Also, γ_t represent threshold values, and in the present context these reflect the baseline hazard (*i.e.*, the hazard when all covariates equal 0). These threshold parameters are akin to the intercept parameters α_{0t} in the dichotomous version of the model. The plus sign following γ_t means that a positive regression coefficient for a covariate indicates an increased hazard (*i.e.*, the event occurs sooner) as values of the covariate increase.

3.2 Survival analysis model for an ordinal outcome

In this section, the re-formatted form of the data, as captured in **smkcclc.ss3** is 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

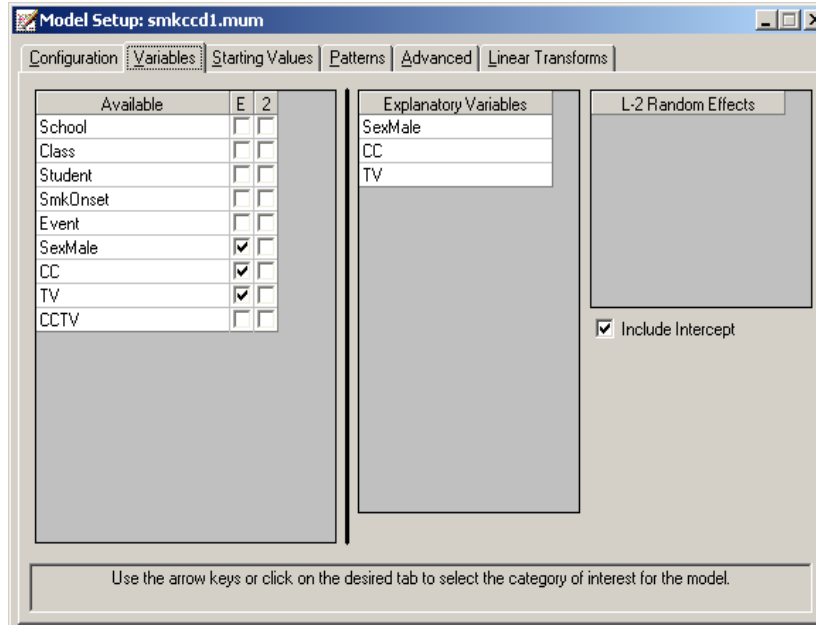
$$\log\left[-\log\left(1 - P_{ijt}\right)\right] = \gamma_t + (\text{Sex}M_{ij})\beta_1 + (\text{CC}_j)\beta_2 + (\text{TV}_j)\beta_3 + v_{0i}.$$

3.2.1 Setting up the analysis

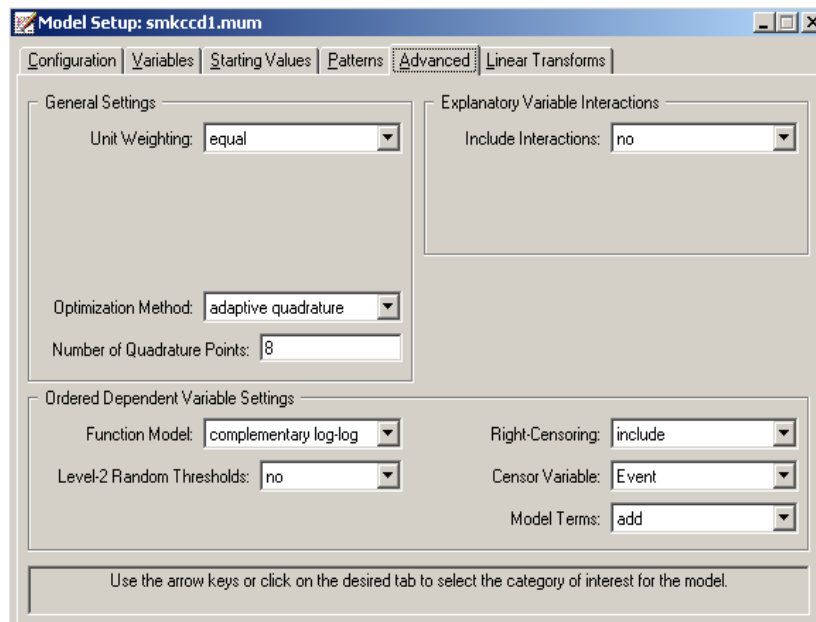
Using the data in the SuperMix spreadsheet **SMKCCLC.ss3**, we start by selecting the **New Model Setup** option on the **File** menu to open the **Model Setup** window. Enter (optional) titles in the **Title 1** and **Title 2** text boxes. Select the ordinal outcome variable **SmkOnset** from the **Dependent Variable** drop-down list box. Note that when the variable is selected, the **Categories** field is populated with values 1 through 4. In these data, the value "1" represents missing data because this value indicates failure or censoring at Wave A (*i.e.*, the pre-intervention time point). As previously noted, the intent was to focus on the post-intervention time points only (*i.e.*, Waves B, C, and D). Indicate this by setting the **Missing Values Present** field to true, and entering the value "1" in the **Missing value for the Dependent Var** field. The **Categories** field now shows the remaining three categories only. The variable **School**, which defines the units within which students are nested, is selected as the Level-2 ID from the **Level-2 IDs** drop-down list box. The completed dialog box is shown below.

Categories	Value
1	2
2	3
3	4

Click on the **Variables** tab of the **Model Setup** window. **SexM**, **CC**, and **TV** are specified as the predictors (explanatory variables) of the fixed part of the model by checking the corresponding boxes in the **E** column of the **Available** grid on the **Variables** screen. By default, it is assumed that the intercept is allowed to vary randomly over the level-2 units, as indicated by the checked box in the **Include Intercept** field.



To specify the number of quadrature points, link function (**Function Model**), and right censoring, we proceed to the **Advanced** screen by clicking on the **Advanced** tab. Change **Model Terms** from subtract to add (so that the model terms are added to the thresholds as specified in the ordinal version of the survival analysis model) and select **complementary log-log** as the **Function Model** (to yield a proportional hazards model). Note that the default **Number of Quadrature Points** of **10** is replaced by **8**. Only 8 quadrature points were used here since the values of the estimated parameters and $-2\ln L$ statistic remain unchanged, up to 5 decimal places, for this or a larger number of quadrature points. Finally, we indicate that **Right Censoring** is to be included and that the variable for this is Event (which is coded 0 = censor and 1 = event).



Use the **File, Save** option to save the model setup to a file named **smkccd1.mum**. Next, use the **Analysis, Run** option on the main menu bar to run the analysis.

3.2.2 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.

SuperMix - [smkccd1.out]

```

o=====o
| Survival Analysis Treat first wave as missing data |
| Comp_Log-log link function, Level-3 Model         |
o=====o

          Model and Data Descriptions

Sampling Distribution           = Multinomial
Link Function                   = Cumulative CLL
Number of Level-2 Units        = 28
Number of Level-1 Units        = 1556
Number of Level-1 Units per Level-2 Unit =
21  74  24  13  29  70  18  17  25  27  19  31
40  51  27  23  79  39  52  106  74  142  51  104
81  74  94  151
  
```

Save As... Close

SuperMix - [smkccd1.out]

```

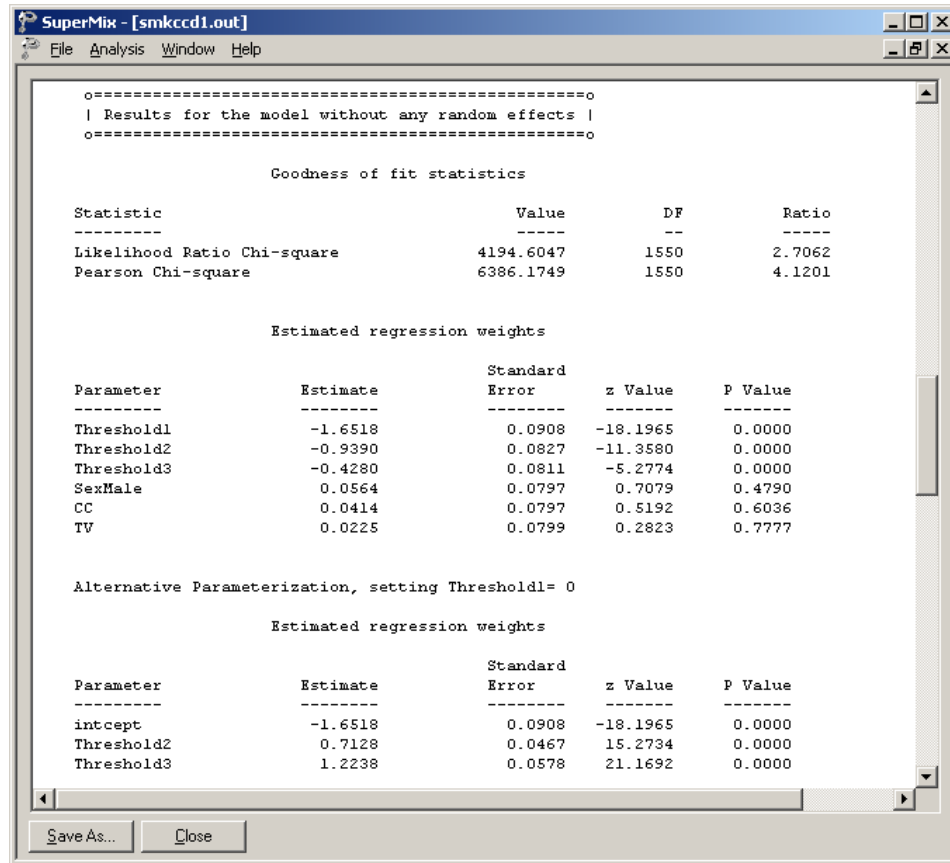
o=====o
| Descriptive statistics for all the variables in the model |
o=====o

Variable           Minimum      Maximum      Mean      Standard
-----           -
SmkOnset1          0.0000      1.0000      0.3046     0.4604
SmkOnset2          0.0000      1.0000      0.3175     0.4656
SmkOnset3          0.0000      1.0000      0.3779     0.4850
SexMale            0.0000      1.0000      0.4769     0.4996
CC                 0.0000      1.0000      0.4788     0.4997
TV                 0.0000      1.0000      0.4904     0.5001
  
```

Save As... Close

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.



o=====o
| Results for the model without any random effects |
o=====o

Goodness of fit statistics

Statistic	Value	DF	Ratio
Likelihood Ratio Chi-square	4194.6047	1550	2.7062
Pearson Chi-square	6386.1749	1550	4.1201

Estimated regression weights

Parameter	Estimate	Standard Error	z Value	P Value
Threshold1	-1.6518	0.0908	-18.1965	0.0000
Threshold2	-0.9390	0.0827	-11.3580	0.0000
Threshold3	-0.4280	0.0811	-5.2774	0.0000
SexMale	0.0564	0.0797	0.7079	0.4790
CC	0.0414	0.0797	0.5192	0.6036
TV	0.0225	0.0799	0.2823	0.7777

Alternative Parameterization, setting Threshold1= 0

Estimated regression weights

Parameter	Estimate	Standard Error	z Value	P Value
intcept	-1.6518	0.0908	-18.1965	0.0000
Threshold2	0.7128	0.0467	15.2734	0.0000
Threshold3	1.2238	0.0578	21.1692	0.0000

Save As... Close

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.


```

SuperMix - [smkccd1.out]
File Analysis Window Help

o=====o
| Optimization Method: Adaptive Quadrature |
o=====o

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

-2lnL (deviance statistic) =      3187.38945
Akaike Information Criterion =    3201.38945
Schwarz Criterion           =    3238.83857

Estimated regression weights

Parameter          Estimate          Standard
-----          -
Threshold1         -1.6568          0.0951
Threshold2         -0.9435          0.0870
Threshold3         -0.4316          0.0850
SexMale            0.0575          0.0798
CC                 0.0452          0.0844
TV                 0.0212          0.0837
z Value           P Value
-----          -
-17.4289         0.0000
-10.8481         0.0000
-5.0802          0.0000
0.7201           0.4714
0.5358           0.5921
0.2532           0.8001

Alternative Parameterization, setting Threshold1= 0

Estimated regression weights

Parameter          Estimate          Standard
-----          -
intcept           -1.6568          0.0951
Threshold2         0.7134          0.0467
Threshold3         1.2252          0.0581
z Value           P Value
-----          -
-17.4289         0.0000
15.2638          0.0000
21.1000          0.0000

```

Random effects estimates and intraclass correlation (ICC)

```

SuperMix - [smkccd1.out]
File Analysis Window Help

Estimated level 2 variances and covariances

Parameter          Estimate          Standard
-----          -
intercept/intercept 0.0031          0.0113
z Value           P Value
-----          -
0.2790           0.7803

```

```

SuperMix - [smkccd1.out]
File Analysis Window Help

Calculation of the intraclass correlation
-----
residual variance = pi*pi / 6 (assumed)
cluster variance = 0.0031

intraclass correlation = 0.0031 / ( 0.0031 + (pi*pi/6) ) = 0.002

```

The last part of the output shows the estimates of the random effects and an estimate of the intracluster correlation. There is no evidence of significant random variation in the intercept over the schools ($p = 0.8120$). The intracluster 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).

3.2.3 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 (Laara & Matthews, 1985). This is shown in Table 8.12. 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 8.12: Comparison of results of binary and ordinal outcome models

Term	Binary outcome (EVENT)	Ordinal outcome (SmkOnset)
Wave B baseline hazard binary α_{01} or ordinal γ_1	-1.6564	-1.6564
Wave C baseline hazard binary $\alpha_{01} + \alpha_{02}$ or ordinal γ_2	$-1.65654 + 0.0399 = -1.6165$	-0.9431
Wave D baseline hazard binary $\alpha_{01} + \alpha_{03}$ or ordinal γ_3	$-1.6564 + 0.3103 = -1.3461$	-0.4313
Male β_1	0.0574	0.0574
CC β_2	0.0449	0.0449
TV β_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