EXERCISE 5: MULTILEVEL MODELS FOR MULTIPLE PROCESSES, USING MLWIN

In this exercise, we will be using *MLwiN* to fit multilevel discrete-time event history models for multiple processes. We will consider how to allow for correlation between the unmeasured factors affecting each process by including individual-level random effects for each process which may be correlated across processes.

We will model the hazard of marital dissolution jointly with the hazard of a conception (leading to a live birth) within marriage.

1. Data

The data consist of those women from the subsample of 1000 NCDS respondents selected for Exercise 3 who have been married at some time between the ages of 16 and 42. There are 937 such women, who contribute 1065 marriages. (Note that, without exclusion criteria on the covariates, the model is likely to be poorly identified with so few women contributing more than one marriage.)

The data are already in discrete-time format, with one record per six-month time interval spent in a partnership. We will adjust for different lengths of exposure to the risk of separation within a six-month interval. The 937 women contribute a total of 28,389 six-monthly observations. The ascii dataset **`ex5.dat`** contains the following variables:

WOMAN	Woman identifier
SEP	Event indicator (1=separation, 0=still together)
CONCEPT	Conception indicator (1=conception, 0=no conception)
ATRISK	Number of months of 6-month interval for which woman exposed to risk of separation
DUR	Partnership duration, time-varying (in 6-month intervals)
PRMAR	Previously married (1=yes, 0=no)
PRCOH_ANY	Ever previously cohabited (1=yes, 0=no)
PRCOH	Previous cohabitation (1=never, 2=with current partner only, 3=with previous partner only, 4=with both current and previous partners)
AGESTART	Age at start of partnership (1=less than 20, 2=20-24, 3=25-29, 4=30+)
AGECURR	Age at start of 6-month interval (1=less than 25, 2=25-29, 3=30-34, 4=35+)
EDUC	Number of years of post-16 education, time-varying (1=none, 2=1-2 years, 3=3-5 years, 4=6+ years)
NKIDY	Presence of preschool children, time-varying (0=none, 1=1+)

NKIDO	Presence of older children, time-varying (0=none, 1=1+)

2. Data Preparation and Setting up a Simple Model

The file **`ex5_macro.txt`** is a *MLwiN* macro. This macro contains syntax to read the ascii file into *MLwiN*, stack the marital separation and conception responses into a single bivariate response variable, create indicators for the type of response (partnership or fertility), create interactions between these response indicators and partnership duration and covariates, and set up a simple multiprocess model.

> Note: As in Exercise 3, we will specify our binary logit model as a multinomial logit model. While it is possible to specify binary response models through the Equations window (see Chapter 9 of the User Guide), there is currently no single command for specifying the model using syntax.

The contents of the macro file are given at the end of this handout. We have used most of the commands before in Exercises 3 and 4. There are only two new commands required to set up data for a multiprocess model (or, more generally, a multivariate response model). The input file has each response stored in a separate column (**sep** and **concept**). Using the `vect` command (line 11), these are stacked into a single column which we call **resp**. At the same time a response indicator, **rtype**, is created. Two binary response indicator variables, **part** and **fert**, are created from **rtype**. The `repe` command (line 18) is then used to make the length of other variables the same as that of the new response variable; specifically, each value is repeated twice.

A separate equation is specified for each response by multiplying **part** and **fert** by duration and the covariates we wish to appear in each equation. Note that we do not have to have the same covariates in each equation. Here, for example, we include age at the start of the partnership as a predictor for marital separation and current age (time-varying) as a predictor for conception.

We specify a simple multiprocess model allowing for effects of partnership duration, and the presence of preschool and older children, on the hazards of both marital separation and conception. In addition, as described above, we allow for age effects on each type of response. The model is specified as a multilevel binary response model, with the response indicators and their interactions with covariates included as the explanatory variables. In this simple model we do not allow for unobserved heterogeneity, so fitting this model is equivalent to fitting two separate models, one for each response. Later, we add in random effects for each response and allow them to be correlated across responses; when this correlation is introduced, the two equations must be estimated simultaneously.

To open and run the macro in *MLwiN*:

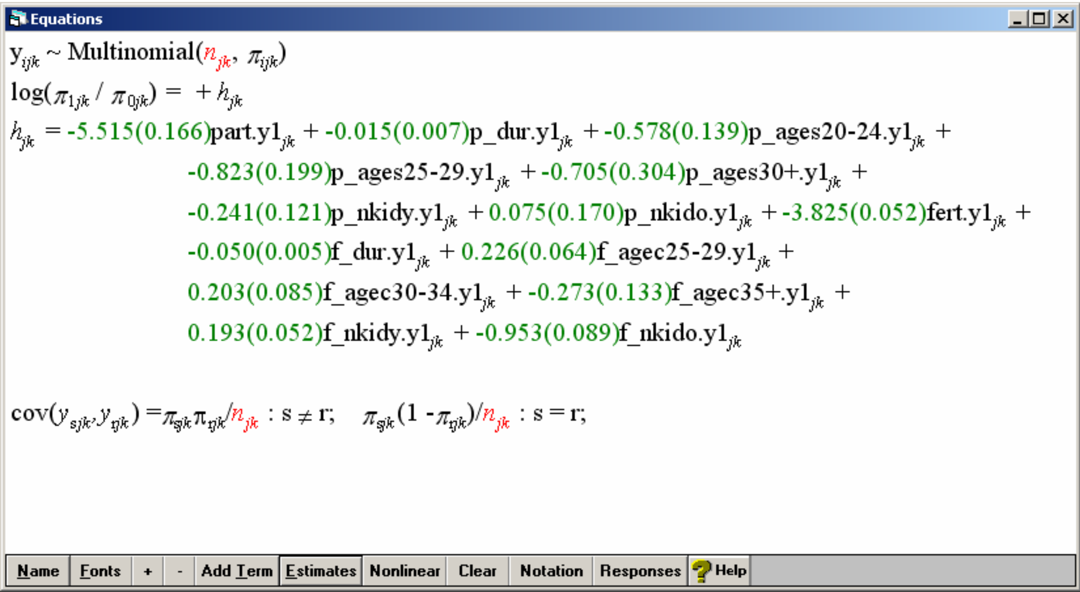
> From the **File** menu, select **Open Macro**
>
> Locate the file **`ex5_macro.txt`**
>
> Click on the **Execute** tab to run the macro

3. Fitting a Simple Multiprocess Model

To run the single-level model which ignores unobserved heterogeneity:

Open the **Equations** window (**Model** → **Equations**)
Click **Nonlinear** and select **Use Defaults**
Click **Start**
Click on **Estimates** to see the estimates

You should get the following results:



The coefficient of **part.y1** is the intercept in the martial separation equation, while **fert.y1** is the intercept in the conception equation. The coefficients of variables with a **p** prefix are effects on the log-odds of a marital separation, and those with prefix **f** are effects on the log-odds of a conception.

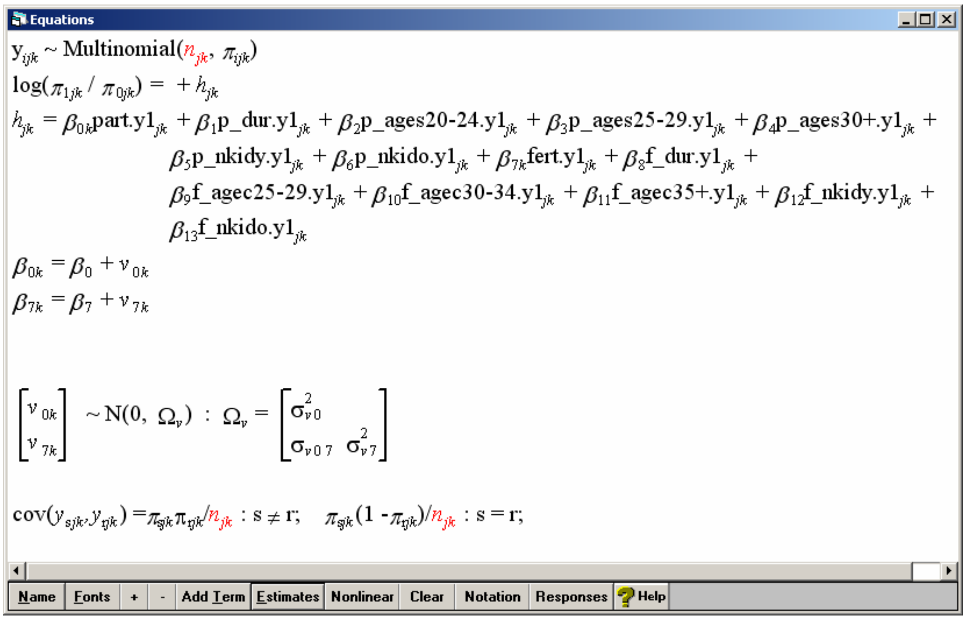
What are the effects of the presence of children on the odds of marital separation and the odds of another conception?

4. Allowing for Unobserved Heterogeneity

To add random effects for each state, we need to allow the coefficients of the variables **part.y1** and **fert.y1** to vary randomly across women.

In the **Equations** window click on **part.y1** (or its coefficient) and check **k(woman_long)**, then **Done**
Repeat for **fert.y1**

The **Equations** window should look like this (you will need to click **Estimates** first):

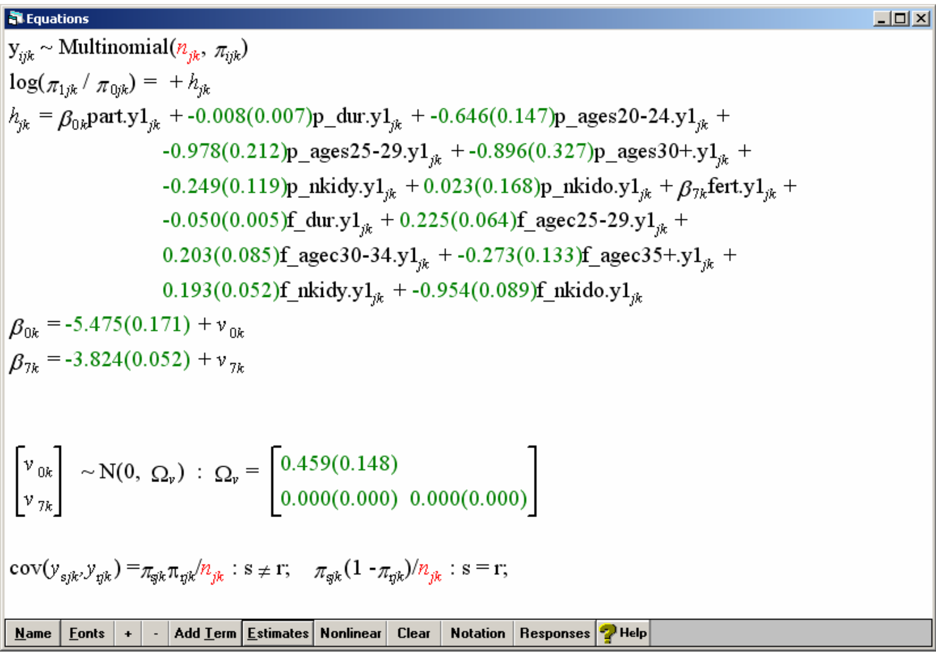

$$y_{ijk} \sim \text{Multinomial}(n_{jk}, \pi_{ijk})$$
$$\log(\pi_{1jk} / \pi_{0jk}) = +h_{jk}$$
$$\begin{aligned} h_{jk} = {} & \beta_{0k}\text{part.y1}_{jk} + \beta_1\text{p\_dur.y1}_{jk} + \beta_2\text{p\_ages20-24.y1}_{jk} + \beta_3\text{p\_ages25-29.y1}_{jk} + \beta_4\text{p\_ages30+.y1}_{jk} + \\ & \beta_5\text{p\_nkidy.y1}_{jk} + \beta_6\text{p\_nkido.y1}_{jk} + \beta_{7k}\text{fert.y1}_{jk} + \beta_8\text{f\_dur.y1}_{jk} + \\ & \beta_9\text{f\_agec25-29.y1}_{jk} + \beta_{10}\text{f\_agec30-34.y1}_{jk} + \beta_{11}\text{f\_agec35+.y1}_{jk} + \beta_{12}\text{f\_nkidy.y1}_{jk} + \\ & \beta_{13}\text{f\_nkido.y1}_{jk} \end{aligned}$$
$$\beta_{0k} = \beta_0 + v_{0k}$$
$$\beta_{7k} = \beta_7 + v_{7k}$$
$$\begin{bmatrix} v_{0k} \\ v_{7k} \end{bmatrix} \sim \text{N}(0,\ \Omega_v)\ :\ \Omega_v = \begin{bmatrix} \sigma^2_{v0} & \\ \sigma_{v07} & \sigma^2_{v7} \end{bmatrix}$$
$$\text{cov}(y_{sjk}, y_{rjk}) = \pi_{sjk}\pi_{rjk}/n_{jk}\ :\ s \neq r; \quad \pi_{sjk}(1 - \pi_{sjk})/n_{jk}\ :\ s = r;$$

σ^2_{v0} is the between-woman variance in the log-odds of separation, and σ^2_{v7} is the between-woman variance in the log-odds of a conception. The cross-process covariance between the random effects is σ_{v07}. If women with a high propensity to separate have a low propensity to conceive during marriage, we would expect a negative covariance estimate

We will begin by estimating the model using quasi-likelihood procedures, but as these are unreliable for small clusters we will switch to MCMC methods after convergence.

Click **More** to fit the random effects model

You will get the following estimates:

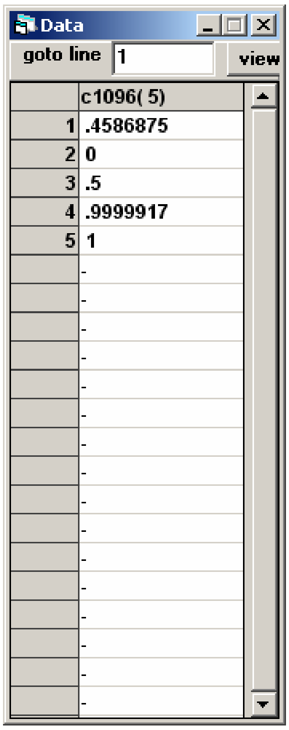

$$y_{ijk} \sim \text{Multinomial}(n_{jk}, \pi_{ijk})$$
$$\log(\pi_{1jk} / \pi_{0jk}) = +h_{jk}$$
$$\begin{aligned} h_{jk} = {} & \beta_{0k}\text{part.y1}_{jk} + \text{-0.008(0.007)p\_dur.y1}_{jk} + \text{-0.646(0.147)p\_ages20-24.y1}_{jk} + \\ & \text{-0.978(0.212)p\_ages25-29.y1}_{jk} + \text{-0.896(0.327)p\_ages30+.y1}_{jk} + \\ & \text{-0.249(0.119)p\_nkidy.y1}_{jk} + \text{0.023(0.168)p\_nkido.y1}_{jk} + \beta_{7k}\text{fert.y1}_{jk} + \\ & \text{-0.050(0.005)f\_dur.y1}_{jk} + \text{0.225(0.064)f\_agec25-29.y1}_{jk} + \\ & \text{0.203(0.085)f\_agec30-34.y1}_{jk} + \text{-0.273(0.133)f\_agec35+.y1}_{jk} + \\ & \text{0.193(0.052)f\_nkidy.y1}_{jk} + \text{-0.954(0.089)f\_nkido.y1}_{jk} \end{aligned}$$
$$\beta_{0k} = \text{-5.475(0.171)} + v_{0k}$$
$$\beta_{7k} = \text{-3.824(0.052)} + v_{7k}$$
$$\begin{bmatrix} v_{0k} \\ v_{7k} \end{bmatrix} \sim \text{N}(0,\ \Omega_v)\ :\ \Omega_v = \begin{bmatrix} \text{0.459(0.148)} & \\ \text{0.000(0.000)} & \text{0.000(0.000)} \end{bmatrix}$$
$$\text{cov}(y_{sjk}, y_{rjk}) = \pi_{sjk}\pi_{rjk}/n_{jk}\ :\ s \neq r; \quad \pi_{sjk}(1 - \pi_{sjk})/n_{jk}\ :\ s = r;$$

Notice the zero estimate for the between-woman variance associated with fertility. Before switching to MCMC estimation, we will need to change this to a non-zero value as MCMC requires a positive definite covariance matrix for starting values. We will change this value to 0.5 as follows:

From the **Data Manipulation** menu, select **View or edit data**

Click **View** and select column **C1096**, where the current random parameter estimates are stored. The estimates are stored in lower triangular form, i.e. $(\sigma^2_{v0}, \sigma_{v07}, \sigma^2_{v7})$.

Edit the **third row** to **0.5**, as shown below



The value for this parameter should be updated in the **Equations** window.

Now click **Estimation Control** then **MCMC**

We will run MCMC with a burn-in of 500 (the default) and a chain length of **1000**

Click **Done** to activate the MCMC settings, then **Start**

MCMC will take a few minutes to run. In the meantime you may wish to start looking at the exercises below.

You should get the following results:

Equations

$$
\begin{aligned}
&\mathrm{y}_{ijk} \sim \mathrm{Multinomial}(n_{jk}, \pi_{ijk}) \\
&\log(\pi_{1jk} / \pi_{0jk}) = +h_{jk} \\
&h_{jk} = \beta_{0k}\mathrm{part.y1}_{jk} + 0.006(0.008)\mathrm{p\_dur.y1}_{jk} + {-0.669}(0.166)\mathrm{p\_ages20\text{-}24.y1}_{jk} + \\
&\qquad -1.213(0.233)\mathrm{p\_ages25\text{-}29.y1}_{jk} + {-1.115}(0.331)\mathrm{p\_ages30+.y1}_{jk} + \\
&\qquad -0.197(0.121)\mathrm{p\_nkidy.y1}_{jk} + {-0.065}(0.172)\mathrm{p\_nkido.y1}_{jk} + \beta_{7k}\mathrm{fert.y1}_{jk} + \\
&\qquad -0.052(0.006)\mathrm{f\_dur.y1}_{jk} + 0.246(0.067)\mathrm{f\_agec25\text{-}29.y1}_{jk} + \\
&\qquad 0.239(0.096)\mathrm{f\_agec30\text{-}34.y1}_{jk} + {-0.225}(0.137)\mathrm{f\_agec35+.y1}_{jk} + \\
&\qquad 0.166(0.053)\mathrm{f\_nkidy.y1}_{jk} + {-0.978}(0.090)\mathrm{f\_nkido.y1}_{jk} \\
&\beta_{0k} = -6.098(0.175) + v_{0k} \\
&\beta_{7k} = -3.838(0.046) + v_{7k} \\
&\begin{bmatrix} v_{0k} \\ v_{7k} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_v) \ :\ \Omega_v = \begin{bmatrix} 1.146(0.250) & \\ -0.058(0.046) & 0.048(0.008) \end{bmatrix} \\
&\mathrm{cov}(y_{sjk}, y_{rjk}) = \pi_{sjk}\pi_{rjk}/n_{jk} \ :\ s \neq r;\quad \pi_{sjk}(1-\pi_{rjk})/n_{jk} \ :\ s = r; \\
&\mathit{Deviance(MCMC)} = 21266.290(56778 \text{ of } 56778 \text{ cases in use})
\end{aligned}
$$

Name | Fonts | + | - | Add Term | Estimates | Nonlinear | Clear | Notation | Responses | Help

Again we need to check whether convergence has been achieved before interpreting the results. If we look at the trajectories, we see that mixing is extremely poor. Below are the results from a burn-in of 5000 and a chain length of 50,000.

Equations

$$
\begin{aligned}
&\mathrm{y}_{ijk} \sim \mathrm{Multinomial}(n_{jk}, \pi_{ijk}) \\
&\log(\pi_{1jk} / \pi_{0jk}) = +h_{jk} \\
&h_{jk} = \beta_{0k}\mathrm{part.y1}_{jk} + 0.003(0.009)\mathrm{p\_dur.y1}_{jk} + {-0.726}(0.178)\mathrm{p\_ages20\text{-}24.y1}_{jk} + \\
&\qquad -1.243(0.267)\mathrm{p\_ages25\text{-}29.y1}_{jk} + {-1.194}(0.366)\mathrm{p\_ages30+.y1}_{jk} + \\
&\qquad -0.233(0.126)\mathrm{p\_nkidy.y1}_{jk} + {-0.018}(0.180)\mathrm{p\_nkido.y1}_{jk} + \beta_{7k}\mathrm{fert.y1}_{jk} + \\
&\qquad -0.051(0.005)\mathrm{f\_dur.y1}_{jk} + 0.237(0.065)\mathrm{f\_agec25\text{-}29.y1}_{jk} + 0.226(0.088)\mathrm{f\_agec30\text{-}34.y1}_{jk} + \\
&\qquad -0.246(0.136)\mathrm{f\_agec35+.y1}_{jk} + 0.171(0.052)\mathrm{f\_nkidy.y1}_{jk} + {-0.970}(0.089)\mathrm{f\_nkido.y1}_{jk} \\
&\beta_{0k} = -5.995(0.239) + v_{0k} \\
&\beta_{7k} = -3.834(0.053) + v_{7k} \\
&\begin{bmatrix} v_{0k} \\ v_{7k} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_v) \ :\ \Omega_v = \begin{bmatrix} 1.179(0.417) & \\ -0.053(0.042) & 0.041(0.009) \end{bmatrix} \\
&\mathrm{cov}(y_{sjk}, y_{rjk}) = \pi_{sjk}\pi_{rjk}/n_{jk} \ :\ s \neq r;\quad \pi_{sjk}(1-\pi_{rjk})/n_{jk} \ :\ s = r; \\
&\mathit{Deviance(MCMC)} = 21268.170(56778 \text{ of } 56778 \text{ cases in use})
\end{aligned}
$$

Name | Fonts | + | - | Add Term | Estimates | Nonlinear | Clear | Notation | Responses | Help

Note that the estimated covariance between the random effects is close to zero with a large standard error. We therefore conclude that there is little evidence of correlation between the unobserved woman-level characteristics affecting separation and those affecting the hazard of a conception. Put another way, we may conclude that the presence of children (at least of those *conceived during marriage*) is exogenous to marital separation. If we refitted the model excluding the covariance, which is equivalent to fitting a separate random effects model to each response, the effects of the presence of preschool and older children on the risk of marital separation would be very similar to those obtained from the multiprocess model above.

5. Exercises

Because of the length of time required to estimate models using MCMC, we will ignore unobserved heterogeneity for now. In practice, it is advisable to carry out preliminary analysis using single-level models, fitting only selected models with random effects. Alternatively, the estimates obtained using quasi-likelihood methods are useful as an approximation.

To remove the random effects:

Click **Estimation Control** then **IGLS/RIGLS**, then **Done**

Click on **part.y1** and deselect **k(woman_long)**

Repeat for **fert.y1**. The woman-level random effect covariance matrix should disappear

Click **Start** to refit the single-level model

- Modify `ex5_macro.txt` to add in other covariates to the model. Remember to declare variables with more than two categories as categorical.
- Carry out significance tests for these covariates. Which factors affect the hazard of separation, and which the hazard of a(nother) conception?

```
NOTE: Exercise 5 - macro to set up multiprocess model in MLwiN
NOTE: Marital dissolution and fertility (conception leading to live birth)

dinp c1-c13
ex5.dat

name c1 'woman' c2 'sep' c3 'concept' c4 'atrisk' c5 'dur' c6 'prmar' c7 'prcoh_any'
name c8 'prcoh' c9 'agestart' c10 'agecurr' c11 'educ' c12 'nkidy' c13 'nkido'

NOTE: Calculate duration-squared variable
calc c14='dur'*'dur'
name c14 'dursq'

NOTE: Stack partnership and conception responses
vect 2 'sep' 'concept' c15 c16
name c15 'resp' c16 'rtype'

NOTE: Create response type indicators
calc c17=('rtype'==1)
calc c18=1-c17
name c17 'part' c18 'fert'

NOTE: Repeat values of other variables twice
repe 2 c1-c14 c1-c14

NOTE: Create level 1 ID (coded 1,2, . . ., 56778)
```

```
code 56778 1 1 c19
name c19 'lev1id'

NOTE: Calculate interactions between partnership response indicator
NOTE: and state indicators with duration and age
calc c20='part'*'dur'
calc c21='part'*'dursq'
calc c22='part'*'agestart'
calc c23='part'*'nkidy'
calc c24='part'*'nkido'

name c20 'p_dur'
name c21 'p_dursq'
name c22 'p_ages'
name c23 'p_nkidy'
name c24 'p_nkido'

NOTE: Interactions between fertility response indicator and duration with duration and age
NOTE: Assume effects of partnership duration and age on hazard of conception is same for
NOTE: marriage and cohabitation
calc c25='fert'*'dur'
calc c26='fert'*'dursq'
calc c27='fert'*'agecurr'
calc c28='fert'*'nkidy'
calc c29='fert'*'nkido'

name c25 'f_dur'
name c26 'f_dursq'
name c27 'f_agec'
name c28 'f_nkidy'
name c29 'f_nkido'
NOTE: Declare age variables as categorical and name categories
catn 1 'p_ages' 1 'p_ages<20' 2 'p_ages20-24' 3 'p_ages25-29' 4 'p_ages30+'
catn 1 'f_agec' 1 'f_agec<25' 2 'f_agec25-29' 3 'f_agec30-34' 4 'f_agec35+'

NOTE: Declare response as multinomial
NOTE: The response is placed in C24 (equal to SEP in the binary case)
NOTE: A response indicator is placed C25 (equal to 1 for binary)
NOTE: Further details of the multinomial command given in Exercise 4 on competing risks
catn 1 'resp' 0 'none' 1 'y1'
mnom 0 'resp' c30 c31 0

NOTE: Declare identifiers for different levels (1 and 2 are the same for a binary response)
NOTE: In the multilevel model we only specify random effects at the woman level
iden 1 c31 2 'lev1id' 3 'woman'

NOTE: Divide response by ATRISK to obtain probability of each type of event within 6-month
NOTE: interval, then declare as response
NOTE: Create binomial response (rate of separation per month)
calc c36=c30/'atrisk'
name c36 'y'
resp 'y'

NOTE: Declare ATRISK as the denominator
calc c37='atrisk'
```

```
name c37 'denom'

NOTE: Add explanatory variables
addt 'part'
addt 'p_dur'
addt 'p_ages'
addt 'p_nkidy'
addt 'p_nkido'
addt 'fert'
addt 'f_dur'
addt 'f_agec'
addt 'f_nkidy'
addt 'f_nkido'
```