Lab 3 – Modelling continuous effects on survival – time trends and covariate effects.

Last week we covered modelling survival using a generalized linear modelling framework that allowed us to place constraints on survival associated with variation among groups of individuals and through time. We used the design matrix in MARK as a very flexible method to apply these constraints in our analysis. So far we've worked strictly with variation that could be classified by categorical effects (dummy variables) such as group membership or discrete time intervals. However it will often be the case that we are interested in modelling survival or other parameters as a function of continuous effects: those that produce a range of responses in survival depending on the relative level of effect. Logical examples might include the influence of weather variables such as precipitation and temperature, effects of individual characteristics such as size or body condition, or habitat characteristics like tree basal area, water temperature, grass height, etc. To do this requires that we consider the effects of continuous "covariates" on our modelled parameters. Today we will cover two distinct types of covariates – those associated with groups, and those associated with individuals. We will also explore the situation where survival changes directionally through time by explicitly incorporating temporal trends into the DM.

As with last week, I'm going to give progressively less details in these handouts as we move forward in terms of step-by-step instructions on "button-pushing" in the software. Hopefully this will help you to become more comfortable with the mechanics of the program as you learn by doing. If you are working through this handout again on your own, refer to the weeks 1 and 2 handouts, or the Gentle Introduction Ch. 3-7, if any of the steps are not clear.

Keep in mind that we are still fitting linear models to our data. For a single variable model, the form would be:

$$logit(\phi) = \beta_1 + \beta_2 X_1 \qquad \text{eq. 1}$$

Previously with discrete grouping variables, in matrix form the above equation would look like the following

$$\begin{matrix}\phi_A \\ \phi_B\end{matrix} = \begin{bmatrix}1 & 1 \\ 1 & 0\end{bmatrix} * \begin{matrix}\beta_1 \\ \beta_2\end{matrix} \qquad \text{Matrix 1}$$

Where ΦA and ΦB are survival probabilities for two discrete groups of individuals, and the "data" that populates the matrix comes in the form of "dummy" variables that are applied across groups. If we want to include data that describe variation among individuals, rather than groups, we have to work with individual covariates and we have to also think in individual terms within our matrix:

$$\phi = [1 \quad X_1] * \begin{matrix}\beta_1 \\ \beta_2\end{matrix} \qquad \text{Matrix 2}$$

Where now the effect of B2 is applied equally across groups as a function of variable X1, which is defined by data that are associated with each individual included in the analysis as part of their encounter history. For example, in the following 4-occasion history with only 1 group:

```
0110  1  3.5  ;
1011  1  1.1  ;
1101  1  0.8  ;
```

The first columns of 1s and 0s reflects the history of encounters, the second column of 1s indicates there is 1 individual for each encounter, and the third column of values are individual data for covariate X. If we were to also include group membership, the linear equation would be:

$$logit(\phi) = \beta_1 + \beta_2 X_1 + \beta_3 X_2 \qquad \text{eq. 2}$$

And the matrix equation would look like this:

$$\begin{matrix}\phi_A \\ \phi_B\end{matrix} = \begin{bmatrix}1 & 1 & X_2 \\ 1 & 0 & X_2\end{bmatrix} * \begin{matrix}\beta_1 \\ \beta_2 \\ \beta_3\end{matrix} \qquad \text{Matrix 3}$$

Where B1 is the intercept, B2 is the group-level effect, and B3 is the covariate effect. Notice that the B3 covariate effect is applied equally across groups, because values for X2 are contained in all cells in the third column. However each individual in the analysis has an independent value for X, and the "data" that populate the third column is drawn from each individual. For a thorough explanation of how these individual data are incorporated into the model likelihood, see the first 3 pages of CH 11.

In Matrix 3, we could express the group effect, B2, similarly, as:

$$\begin{matrix}\phi_A \\ \phi_B\end{matrix} = \begin{bmatrix}1 & X_1 & X_2 \\ 1 & X_1 & X_2\end{bmatrix} * \begin{matrix}\beta_1 \\ \beta_2 \\ \beta_3\end{matrix} \qquad \text{Matrix 4}$$

However, remember that as a dummy variable, X1 is defined by a value of 1 for group A and 0 for group B. So the 1 and 0 values in the second column of matrix 3 reflect the fact that all individuals in each group share common values for X1. In that scenario it is far simpler to populate the matrix with data directly. We can extend this principle to a scenario where a continuous value is expressed across all individuals within groups. As an example let's first consider a situation where we monitor two groups (males and females) across four years. The fully time-varying linear model would look like this:

$$logit(\phi) = \beta_1 + \beta_2 S + \beta_3 Y_1 + \beta_4 Y_2 + \beta_5 Y_3 \qquad \text{eq. 3}$$

Where B2 is the effect of group (Sex; S), and B3, B4, and B5 reflect the relative difference in survival for years 1, 2, and 3, and year 4 is defined by the model intercept. The matrix equation would look like this:

$$\begin{matrix}\phi_{M1}\\ \phi_{M2}\\ \phi_{M3}\\ \phi_{M4}\\ \phi_{F1}\\ \phi_{F2}\\ \phi_{F3}\\ \phi_{F4}\end{matrix} = \begin{bmatrix}1 & 1 & 1 & 0 & 0\\ 1 & 1 & 0 & 1 & 0\\ 1 & 1 & 0 & 0 & 1\\ 1 & 1 & 0 & 0 & 0\\ 1 & 0 & 1 & 0 & 0\\ 1 & 0 & 0 & 1 & 0\\ 1 & 0 & 0 & 0 & 1\\ 1 & 0 & 0 & 0 & 0\end{bmatrix} * \begin{matrix}\beta_1\\ \beta_2\\ \beta_3\\ \beta_4\\ \beta_5\end{matrix}$$

Matrix 5

Before you panic, the above design matrix should look remarkably similar to the Group + Time models we built in MARK last week, because that's exactly what this model reflects. Survival varies among sexes and among time intervals, but the influence of time is the same for both males and females. Now, suppose we want to test the hypothesis that annual variation in survival is a function of annual rainfall patterns. Because all individuals occupy the same study system, they experience the same level of rainfall and we can apply the effect across individuals. The linear equation would look like this:

$$logit(\phi) = \beta_1 + \beta_2 S + \beta_3 X_1 \qquad \text{eq. 4}$$

Where B3 is the effect of rainfall and X1 is the measured value in each year. If we measure (cm) rainfall across 4 years as (25, 43, 18, 36), we can actually populate the matrix directly with these data.

$$\begin{matrix}\phi_{M1}\\ \phi_{M2}\\ \phi_{M3}\\ \phi_{M4}\\ \phi_{F1}\\ \phi_{F2}\\ \phi_{F3}\\ \phi_{F4}\end{matrix} = \begin{bmatrix}1 & 1 & 25\\ 1 & 1 & 43\\ 1 & 1 & 18\\ 1 & 1 & 36\\ 1 & 0 & 25\\ 1 & 0 & 43\\ 1 & 0 & 18\\ 1 & 0 & 36\end{bmatrix} * \begin{matrix}\beta_1\\ \beta_2\\ \beta_3\end{matrix}$$

Matrix 6

And the reason this works is because the covariate values are applied equally across all individuals: they are group-level covariates. In this case, we are applying the effect evenly among groups, but we can also test for variable group-level effects using a group * covariate interaction. We will see how this works in a later exercise.

Finally, let's consider a scenario where we wanted to test for a systematic change (either increase or decline) in survival through time. To do so requires that we apply a specific constraint to survival that causes it to trend at a systematic rate through time. For a linear trend, this only requires one beta coefficient that defines the slope of the trend. The linear model is:

$$logit(\phi) = \beta_1 + \beta_2 S + \beta_3 T \qquad \text{eq. 5}$$

Where we'll use an upper case "*T*" to distinguish a temporal trend from general time variation (lower case "t"). The matrix equation is given as:

$$\begin{matrix}\phi_{M1}\\ \phi_{M2}\\ \phi_{M3}\\ \phi_{M4}\\ \phi_{F1}\\ \phi_{F2}\\ \phi_{F3}\\ \phi_{F4}\end{matrix} = \begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 2\\ 1 & 1 & 3\\ 1 & 1 & 4\\ 1 & 0 & 1\\ 1 & 0 & 2\\ 1 & 0 & 3\\ 1 & 0 & 4\end{bmatrix} * \begin{matrix}\beta_1\\ \beta_2\\ \beta_3\end{matrix}$$

Matrix 7

Because the value for T increases by a fixed amount (1) at each successive time step, B3 gives a linear slope that matches some fixed level of increase or decline. Again, the data are populated directly into the design matrix, and the effect is considered equally across all groups.

Part 1 – Group-level Covariates and Time Trends.

Let's get started working with these continuous effects on survival. We will use the full European Dipper dataset (males and females) based on the "ED.inp" input file. If you have access to the results file you worked with last week, go ahead and open it. If not, I've added a zipped folder that contains the analysis from last week on the Blackboard content page under "Dipper Analysis".

1. Open the file and retrieve the Group + Flood model. Remember that groups in this case are males and females, and we were modelling an effect of flood versus non flood years on survival.
2. In the column currently occupied by the flood term, beginning with the top cell and working downward, enter the values 1,2,3,4,5,6. Then repeat this sequence for the female segment of the population. Why don't we continue counting upward (7,8,…12)?
3. Run the model. We will label it phi(Group + T) using the upper case T to distinguish a time trend.
4. Let's take a second to look at the model results, including the AIC results, beta coefficients, and a graph of the survival estimates.
5. What if we expected a non-linear trend on survival through time? We could fit this using a quadratic term, where the linear model would take the form:

$$logit(\phi) = \beta_1 + \beta_2 S + \beta_3 T + \beta_4 T^2 \qquad \text{eq. 6}$$

   How would we fit this model using the design matrix? Take your best guess, and we'll compare notes.
6. Run that model (call it phi(Group + T^2)) and we'll take a look at the results, including a graph of the estimates.
7. Now, let's finish with a scenario where we want to fit a continuous value that describes the rate of flooding each year, rather than the discrete flood vs. non-flood effect we tested for last week. In the real world, we might use average flow rates from a river gauging station to estimate the relative degree of flooding in each year. Let's pretend

(these are haphazardly generated data by me) that the average annual flow for this stream is the following:

Y1	5
Y2	25
Y3	36
Y4	14
Y5	7
Y6	13

It ought to be fairly intuitive at this point how we can incorporate these values directly into the design matrix, because it is directly analogous to modelling a trend effect. Run this model and label it phi(Group + Flow).

8. What if we want to model an interaction between flow rate and group? Can that be done in a group covariate context? Run this model as (Group* Flow) and we'll interpret the results.

Part 2 – Individual Covariates

The following is drawn from materials contained in Ch. 11 of the gentle introduction. We are going to be evaluating the effects of body mass on survival of a hypothetical animal – these data are simulated.

1. Start a new analysis using the data file "indcov1.inp". Take a look at the inp file. Notice there are 8 occasions, 1 group, and two columns for continuous individual covariates. The first covariate value is individual body mass (measured at capture) and the second is simply the body mass covariate squared (BM^2).
2. Under the Individual Covariate option, tell MARK that we've got 2 individual covariates. Using the "Enter Ind. Cov. Names" button, give the covariates names. This is a fairly important step, as we'll use these names to specify where the covariates should be included in the DM. You should pick something intuitive and short – In this case lets label them BM and BM2. Case sensitivity is not important. Note that if you fail to complete this step, the Var1, Var2 labels will be used as defaults, which doesn't seem like a bad thing until you have dozens of covariates and need to remember which is which!
3. First, take a look at the PIMS. This is a fairly simple model with no group structure. In the book, they suggest setting the PIMs to be constant due to the artificial nature of the data and the "known" lack of time structure. However, under a real analysis you'll never be in a position where the survival or detection functions are known before hand, so let's proceed with the full time structures on both phi and p. Run this model and label it appropriately.
4. Now let's open the design matrix. We could open a full matrix in this case, since we haven't altered the PIMs at all, but why don't we use a reduced and build it ourselves for the practice. Think about how many betas would be required to fit this model (it may

help to write out the linear equation) and build your DM accordingly. Run the Phi(time) p(time) model again from the DM, just to satisfy yourself that you've built it correctly.

What's that – the results are different? Let's side-track for a second to discuss why and how we can remedy the situation.

5. Now that we're back on track, return to the DM and remove the group and temporal structure on p. Run the resulting model and label it appropriately.
6. You should have found that the p(.) structure was superior to the p(time) structure – this is not surprising since Cooch and White have confessed that no temporal structure was used when generating these data. Now, let's do the same for the phi terms – remove the time-varying structure, label correctly, and run the model.
7. Again, we find that the phi(.) model is better supported than the phi(time) model, and again this is not surprising for the same reasons we outlined in #6.
8. So we've determined there is no evidence to support temporal variance in survival from this hypothetical dataset. But remember our objective for this analysis was to test for variation among individuals with respect to mortality risk (1-survival probability) associated with body condition. To do this, we'll include our individual covariates. First, go back to DM for the phi(.) p(.) model and insert another column. In the first cell for the survival parameters, type in BM to indicate the covariate value. When you click out of the cell, it should highlight red.
9. In accordance with what we saw for Matrix 3, we need the covariate value to be applied to the survival parameter for each interval, that is, we need BM to appear in each cell that corresponds with a survival rate. You can either type "BM" repeatedly in each cell, or alternatively (and more conveniently) can right click in the cell you just filled in, chose the "copy value down" option, chose 7 rows, and hit o.k. We now have the phi (body mass) model.
10. Run the model. You may notice a slightly longer run time. This is because with the addition of an individual covariate, the model likelihood has become exceptionally more complex. See the first pages of Ch 11 if you are curious as to why.
11. Let's take a look at the results of this model, which should be familiar to you at this point. In fact, the interpretation of results from an individual covariate model are really no different than any other results we've seen so far using linear models in the DM.
12. Although we saw a good amount of support for a linear effect of body mass, recall we have a second individual covariate value to work with that describes body mass squared. This will be useful for fitting a quadratic effect of body mass on survival, and there is a very good reason to want to do so.

Pause briefly to discuss the evolutionary biology that makes a (Body Mass)^2 hypothesis a good one to test.

13. First, let's fit a quadratic effect of body mass using the second supplied covariate. Remember (because I showed you in eq. 6) that to fit a quadratic term you must also include the simple linear effect. How would we now change the DM to model a quadratic effect of body mass? Take a guess, and then we'll compare notes and run the models.

14. There is actually a few alternatives to including squared values in your input file every time you want to fit a quadratic effect. MARK allows you to specify a number of simple mathematical operators in the design matrix that can be used to modify covariate values (see sidebar beginning on 11-20). One of the most useful is the "product" function. Open your DM for the phi (body mass^2) model, and in place of the top BM2 value, type in the term:

    product(bm,bm)

    and copy it down to the cells for the other 6 survival parameters. Run this as "phi(body mass^2) p(.) PROD", or something similar. Now do the same, but use the alternative "power" function as:

    power(bm,2)

    Run this as "phi(body mass^2) p(.) POW" or similar. The results from all three models should be identical, so you see there are multiple ways to arrive at the same endpoint, which in part demonstrates the versatility of fitting models using the DM and covariates in MARK. Also note that you can use the POW function to fit >2nd order polynomials on your covariate effects, but you should do so with caution and a keen eye towards the biological interpretation.

15. We'll finish with a discussion of estimating survival probabilities for individuals with specific covariate values, and also for using the parameter coefficients and the reconstructed logit to plot effect curves and for covariate values.

At this point we've covered all of the "basics" of running survival analyses in program MARK using the basic CJS model. Most of the more advanced applications you are likely to run into are simply logical extensions of these first principles. Next up, we'll to cover some advanced techniques associated with more complex analyses and different data types.

Assignment – *Note – this assignment will only require information presented up to and including Part 1 of this week's lab handout. I will incorporate individual covariates (Part 2) into next week's assignment.*

For this assignment, let's pretend we've used mark-recapture to study the annual survival of an imaginary pool-breeding amphibian that breeds in vernal pools but overwinters in upland habitat by burying itself in leaf litter, under downed woody debris, etc. This would be the type of data we might collect, for example, if we were using pitfall traps to tag individuals coming to and from vernal pools during the spring/summer breeding season. The encounter history (Lab3HW.inp) will consist of two groups; Group 1 = Females, Group 2 = Males.

There are two concerns related to the survival of this population. First, we want to evaluate whether there is evidence to suggest that survival is declining for this species. Second, we are interested in whether wintertime temperatures are correlated with survival probabilities, and whether the effects of winter temp on survival differ among males and females. Using what you've learned so far, I would like you to answer these two (really three) questions. Start this

exercise by determining the most appropriate structure for the recapture probability component of the model. I recommend you leave the survival component of the model in the full group*time structure while working through the structures for p. Once you've found an appropriate model for p, use that structure and then run the range of survival models that you feel are required to test the two questions I've outlined above. The average winter temperature during the study (degrees C) is listed below:

Year	Avg. Temp
1	-2
2	0
3	-3.5
4	2.5
5	3.5

You should use the design matrix to construct all models, and at least try to build your design matrix from scratch rather than allowing MARK to do it for you. Please turn in your MARK file, AIC results table, and real parameter estimates.