Lab 2 – Linear Models, the design matrix, and modelling age structure

Last week we covered building, running, and evaluating competing models in MARK using the PIMs. This week we are going to introduce what is ultimately a much more flexible tool for model construction; the design matrix. Using the design matrix we will be constructing models of survival in a Generalized Linear Modelling (GLM) framework. GLMs are similar to ordinary linear regression, but are much more flexible in that they can accommodate multiple response variable types, including binary (1,0) response variables that would traditionally be analyzed using logistic regression. Today we will work through constructing competing GLMs of survival and detection in MARK, and we will also use the PIMs in combination with the design matrix to model age-structure.

The following are partial step-by-step instructions that correspond with the materials presented in Chapters 6 and 7 of the Gentle Introduction. So that we can move through the material in a timely fashion, I am hitting the "high points" and in some cases I'll omit topics covered in greater detail in the GI. If you will continue to work in MARK for your own research, I highly recommend working through those chapters on your own, or at least reading through them, to reinforce what we cover in lab.

In cases where we use a new procedure or concept for the first time I've spelled the steps out in great detail. In situations that are review from the previous lab I've used less detail, so refer to the handout for Lab 1 or Chapters 3 and 4 if you get stuck. I've deviated slightly from the order of topics in the Gentle Introduction in that we will cover the design matrix and age effects first, and will cover goodness-of-fit testing in the coming weeks.

Once we move into using RMark via program R, much of the apparent tedium of using the design matrix will disappear because RMark will do most of the work of constructing model structures behind the scenes based on the model statements you provide. However I think it's worth going through here in detail because it will give you a peak behind the curtain and into the black box of the analysis, and understanding how to build design matrices in MARK is a good tool to solidify your understanding of principles of regression in general.

Part I – Applying linear constraints to survival models using the design matrix.

First we will start with a brief review of the concepts of linear models. We will cover the following during lecture, but you should review the first 12 pages of Chapter 6 in the Gentle Introduction for a much more thorough treatment, and Chapter 6 of Powell and Gale for a complimentary conceptual treatment. Based on all this background, you should recall that a simple linear model takes the form:

$$y = \beta_1 + \beta_2 X_1 + \varepsilon_i$$

Where y is the response variable, B1 represents the model intercept (the value of Y at X=0.0), B2 is the slope coefficient for predictor variable X1, and Ei represents the remaining unexplained variance in Y. As with the example of swifts from the previous lab, if we wanted to test the hypothesis that the response variable Y differed as a function of two groups (colonies in

that example), we could use the above model to reflect that hypothesis. The explanatory variable X would then be a binary predictor variable where individuals from Colony A would receive a value of 1, and Colony B a value of 0. The slope coefficient B2 would then define the difference in Y with respect to colony membership. By definition B2*0 yields a value of 0, so the predicted value of Y for Colony B would be defined strictly by the intercept term, and the predicted value of Y for Colony A would be defined by both the intercept term and B2 (B2*1). If we were to look at this in matrix form, it would look something like:

$$\begin{matrix} y_A \\ y_B \end{matrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} * \begin{matrix} \beta_1 \\ \beta_2 \end{matrix}$$

Now imagine a scenario where we have more than two groups (we'll assume 3: A, B, and C). What would this model look like?

$$y = \beta_1 + \beta_2 X_A + \beta_3 X_B + \varepsilon_i$$

Here we still have the intercept term B1, and we have added two B*X terms to distinguish among the 3 groups. Why aren't there 3 B*X terms, one for each group? Again, let's define membership for group as a function of a combination of binary values, otherwise known as "dummy variables". A member of group A could have the combination (1 0 0), a member of group B (0 1 0), and group C (0 0 1). Notice though that the third series of ones and zeros is actually redundant, since group C can be defined as not belonging to either groups A or B with the series (0 0). In the above equation then, Group C can be defined simply by the intercept term. The matrix notation would be the following:

$$\begin{matrix} y_A \\ y_B \\ y_C \end{matrix} = \begin{matrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{matrix} * \begin{matrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{matrix}$$

If we wanted to write out the linear equation that estimates the response y for group A, it would look like:

$$y_A = \beta_1 + (\beta_2 * 1) + (\beta_3 * 0)$$

If we wanted to write it out for group B, that would look like:

$$y_B = \beta_1 + (\beta_2 * 0) + (\beta_3 * 1)$$

And finally for group C:

$$y_C = \beta_1 + (\beta_2 * 0) + (\beta_3 * 0)$$

So you see we can define the response variable y across 3 levels, A, B, and C, as a function of one intercept term (B1) and two additional slope coefficients (B2 and B3). In this context, the parameter B2 can be thought of as the difference in the response variable y between groups A and C. Similarly B3 reflects the difference in Y between groups B and C. The distance between

the values for B2 and B3 reflect the difference between groups B and C with respect to y. The 3x3 matrix listed above is referred to as the design matrix:

$$\begin{matrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{matrix}$$

Program MARK allows us to construct our own design matrix, which gives exceptional flexibility in creating model structures that place specific constraints on how the response variable y (typically an interval-specific survival probability, detection rate, or another similar demographic rate) is estimated using maximum likelihood. Note in a standard linear model, there is theoretically no upper or lower bounds on the potential estimates of y (-infinity, +infinity). Because y in this case is a probability value that we are estimating based on encounter data (1= encountered, 0= not encountered), we have to place an additional structure on the linear model to ensure the estimates are bounded by (0,1), since be definition a probability must be ≥0.0 and ≤1.0. This is done by specifying a link function. See the sidebar on page 6-11 in the Gentle Introduction for a complete description of link functions, but for simplicity sake assume that we will typically use the logit link, which takes the form:

$$y = \frac{e^{\beta_1+\beta_2 X_A+\beta_3 X_B}}{1+e^{\beta_1+\beta_2 X_A+\beta_3 X_B}}$$

And that can be simplified to:

$$y = \frac{1}{1+e^{-(\beta_1+\beta_2 X_A+\beta_3 X_B)}}$$

By working in the logit link function, the parameter coefficients (β) will be estimated by maximum likelihood such that y (survival, recapture, or some other probability value) will always be bounded by (0,1). It turns out that when using the design matrix, the Sin function will no longer work and the logit becomes our default. If you are curious as to why this is, see the sidebar beginning on page 6-22 in the Gentle Introduction.

Now let's start a new analysis and get our first taste of working with the design matrix. First, open a new MARK file and set up a CJS model using the input file for the full European Dipper dataset (males and females), which is called "ED.inp". Take a look at the input file and set your number of occasions and group structure accordingly.

1. Even though we are going to work in the design matrix, we still need to "tell" MARK which parameters we want it to estimate, and we use the PIMs to do this. In general we'll always start with the most general model (in this case a g*t structure on both phi and p). Take a look at the PIM matrices. In this case, the default structure is full g*t variation on both phi and p. Make sure you remember why this follows logically from the PIM structure, and then run the model. Label it "phi(g*t) p(g*t) – PIM". We'll add the

PIM on the end so we can distinguish this model from one ran later using the design matrix.

2. Take a look at the results for the real parameter estimates – many are not estimating? Why is this the case?
3. Now we will open the design matrix and really start to get our hands dirty. From the menu bar at the top of the screen, select the "Design" menu, and the "Full" option. Mark will go berserk for a second or two and you'll be left with a window open that contains a spreadsheet-looking matrix of blue and white cells. The blue cells each contain a 1, and the white cells 0. This is the full design matrix that describes the complete group and time structure on both phi and p.

Let's pause here for a second to really look at this design matrix and see how it relates to the linear model that defines the structure of modelled variance in phi and p.

4. Next we will start to reduce the matrix and place additional "constraints" on our parameters. A constraint is simply a reduction in complexity of the model structure that causes the parameters to be estimated similarly in some way. The phi(.) model is a fully constrained model in that all the survivals are held at the same value. Recall that during the last lab we mentioned that using the PIMs we could define a group*time structure, but not a group + time structure. The design matrix allows us to do this.
5. In the column labeled "B8", right click in the blue cell at the top. A dropdown menu will open with a number of options you can use to alter the DM. Choose the delete multiple columns option, and in the resulting command window tell MARK to delete rows 8 to 12. Click ok and MARK will go slightly berserk again; don't panic.
6. Now notice the new structure of the design matrix for the Phi parameters. We've removed the second "stair step" from the survivals. If we look at the difference in model structure now between parameter 1 (survival of males in interval 1) and parameter 7 (survival of females in interval 1), we see that they only differ by one beta coefficient, that which describes the effect of sex on survival. This is true for each combination of parameters 1-6 and 7-12 describing interval-specific survivals. What we have done here is to define a sex + time model.

Pause here for a second to revisit the difference between a group*time and a group + time structure.

7. Let's go ahead and run this model and append it to our results table. For now leave the p's set in the full sex*time structure. Be sure to give the model the appropriate name for its additive structure.
8. Once the model is run, let's revisit the design matrix for that model. It should still be open in the background, but notice also if you use the "retrieve" function when right clicking on the model in the results table, it will re-open the DM for you.
9. Next let's run a model in which survival has no time structure and only varies according to sex. Take your best guess at how to alter the DM to accomplish this, run the model, and we'll compare notes.

10. Now, we want to test a hypothesis that survival differs according to annual flooding events on the streams the dippers live in. Retrieve your phi(sex) model. Since we are starting with a model that has a greater number of constraints and are trying to build a more general one, we will need to add additional beta term(s) to define the new model structure. Right click in the B3 column and select the "Insert One Column" option. Mark will do its thing and you'll be left with a blank column in the B3 spot, with the remaining parameters shifted over by 1. We know from Cooch and White that in years 2 and 3 there were major flood events in the study system for which these data were collected. If we want to test whether survival differed depending on these flood events, how can we add an additional term to the model to fit this structure? Again, take a stab at building a model of your own and we'll compare notes. Call the model Phi(sex + flood).
11. Now suppose we also have reason to believe that flooding affects males differently than females. We can test this by fitting a sex*flood interaction. Recognize that this is entirely analogous to our earlier sex*time model, but differs in how we are incorporating the temporal variance into the DM. Again, see if you can surmise this model structure on your own and run the model.

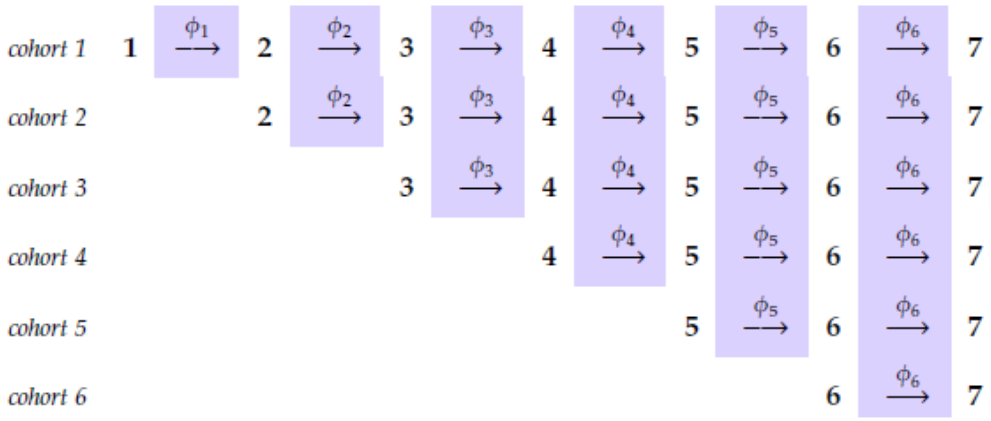
Let's take another quick pause here to talk about two things. First, to keep our heads grounded in biology rather than models and matrices, let's talk about dippers, their natural history, and why we might expect flooding to affect their survival and for that difference to depend on sex; why would we formulate these hypotheses in the first place? Second, and following Cooch and White's lead, let's consider why we wouldn't want to run a sex + flood + time model.

12. Let's go through the model results from this fairly simple set of analyses. Importantly now we will look at the Beta coefficients as well as the real parameter estimates. To access them, right click on the "best" model and click on the "Beta Estimates" option. When viewing the betas estimates, it is helpful to open the notepad window next you your open design matrix. Notice that each Beta Coefficient corresponds to a column in the design matrix.
13. Similarly, when we open the "Real Estimates", each survival or detection estimate corresponds to a row in the DM. Also notice that now the survival terms in the last interval have become estimable, whereas before they were not. This is because we have now constrained them to be the same as other, estimable parameters. Because of this constraint, those parameters are no longer defined strictly by the ϕ*p product in the last occasion. We are effectively borrowing information from other parameters to estimate survival in the last interval.

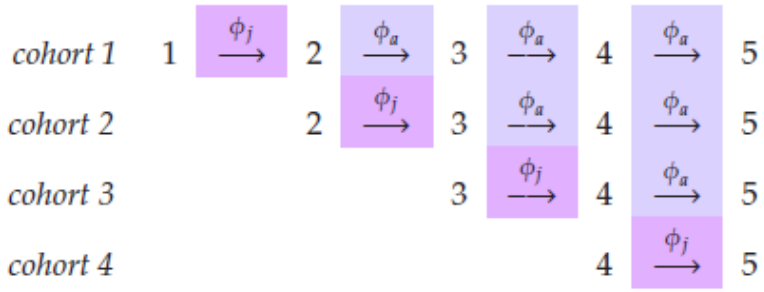
We'll close here with a discussion of interpreting the Beta estimates and their variance, as well as how we can use the betas to estimate the survival rates on our own. Next week, we will move past thinking about constraints simply in terms of "dummy variables" and will consider cases where we wish to model continuous effects on survival and recapture probabilities.

Part II – Incorporating Age Structure using the PIMs

So far we've been working in fairly simple group and time structures; group membership is discrete and does not change within individuals through time. Often we will be interested in evaluating the effects of age on survival; do individuals in older age classes have an inherently different survival probability than those in younger age classes? Life history theory predicts this to be the case, so the need to incorporate age into our analyses of survival should be intuitive. Furthermore, the survival models we've been using so far implicitly assume that age-related variation in survival does not exist. All individuals in a study population are assumed to have equal probability of survival and recapture. Age structure is one way in which this assumption may likely be violated, so how can we incorporate the fact that we are sampling individuals of different ages? You might think that we can include this as a group-level effect, and you would be correct, but it's also not quite that simple. First consider a PIM structure for survival from one of our previous analyses:



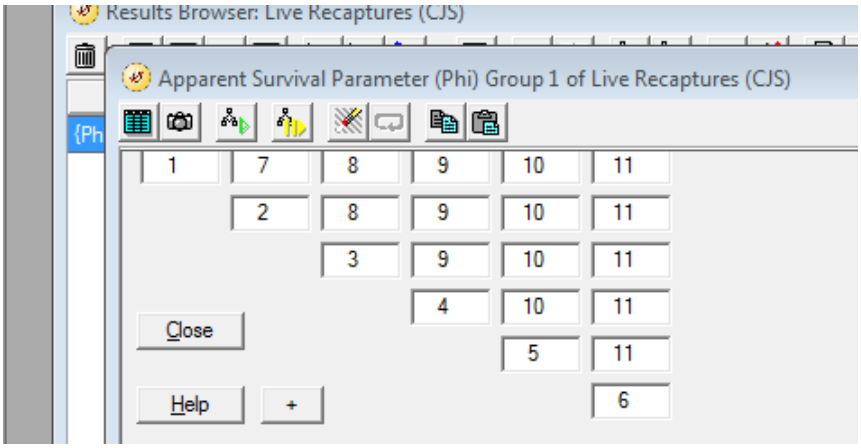
If we were to add a second group, with one constituting young and the second adults, using our basic group structure there is no way for young individuals to transition into adulthood as they grow older. So we need to think about setting the PIMs slightly differently in order to "graduate" individuals among age classes, so that our model structure reflects the following:



In the above conceptual PIM, ϕj reflects survival of juveniles and ϕa reflects survival of adults. We will start with a very simple example that would be similar to a case where we initially capture all individuals as juveniles, but re-encounter them later in life. Examples might include fish that are tagged as smolts and recaptured and released through adulthood, or birds that are captured and color-banded as nestlings and the bands read later as adults.

Example 1 -

1. Start a new project using the file "AGE.inp". Based on what you've learned up to this point to set the model specifications and open the results file, given the information contained in the input file.
2. There is only 1 group in this dataset. Per Cooch and White, here we are going to test for an age effect in these data by manipulating the PIM structure. Open the PIMs and look at the default structure. This should be familiar to you as the general Phi(t) p(t) model. Go ahead and run this model. Notice that because we haven't opened the design matrix, the Sin link function is still available. For consistency from here on out, let's use the logit link function for all models unless otherwise warranted.
3. Now let's adjust the PIMs for Phi to reflect the conceptual diagram above, where the age of individuals is allowed to differ among juvenile and adult cohorts, but we still retain time variation as well. The diagonal cells in the PIM represent the survivals of individuals during the first interval after initial capture, whereas the internal cells of the matrix reflect survival during intervals at least 1 year following initial capture. Using this fact, we see that we can adjust the internal cells to differ from the diagonal and to reflect the structure depicted in the figure above. There are a number of ways we can accomplish this. Since the diagonal already counts up from 1-6, we can start with the second column of row one and change that cell to a value of "7", the one to the right of it gets an "8", the next a "9", etc. For the second row, the second column would then be given an "8", then a "9", etc. The idea here is for the diagonal cells (the phi terms for juveniles) to advance through time, and for the inner cells (the phi terms for adults) to also advance through time independently from the diagonal but consistent with other adults in the same interval. The resulting PIM structure should look like this:



4. Carefully compare this PIM structure with the conceptual diagram on the previous page and make sure you see the connections – this is the foundational basis for modelling age structure in MARK.
5. Now before we run the model there is some parameter accounting we need to take care of. Here we have indexed 11 parameters in the survival PIM. Open the PIM for recaptures and take a look at it. Now we've got substantial overlap between the Phi's and p's, so adjust the p's to eliminate this overlap (Hint – the p's should start with parameter 12). For the time being let's assume no age structure on p.
6. Run the model. Call it Phi(t*age) p(t) to reflect the fact that we've just built a fully general age*time interaction on phi.
7. Take a look at the model selection results as well as the real parameter estimates. We should find model selection support for this more complex age structure, as well as difference among the real parameter estimates. Why are there only 11 rather than 12 (6 intervals and two ages) survival estimates?
8. Now, let's try a model where instead of only two age classes, there are three (juveniles, subadults, adults). This would reflect a more complex age structure. After the preceding example, the way we do this should be apparent with a bit of though – see if you can figure it out and I'll advise as needed.
9. Run the model – notice that Cooch and White provide a fairly formal convention to naming models to reflect differing age structures. I prefer a slightly more intuitive convention here and will call this more complex model Phi(t*Age-3 Class) p(t).
10. Examine the model selection results and parameter estimates - Is there support for a 3 age class structure in these data?

At this point it's probably appropriate to quickly review how these model structures are represented in the model likelihoods by considering the probability statement for each potential history. Cooch and White discuss this in detail beginning on page 7-13. Consider the first (and best-supported) example of the 2 age class model and an abbreviated set of potential capture histories where all individuals are detected each time following initial capture.

capture history	*probability*
1111111	$\phi_1 p_2 \phi_7 p_3 \phi_8 p_4 \phi_9 p_5 \phi_{10} p_6 \phi_{11} p_7$
0111111	$\phi_2 p_3 \phi_8 p_4 \phi_9 p_5 \phi_{10} p_6 \phi_{11} p_7$
0011111	$\phi_3 p_4 \phi_9 p_5 \phi_{10} p_6 \phi_{11} p_7$
0001111	$\phi_4 p_5 \phi_{10} p_6 \phi_{11} p_7$
0000111	$\phi_5 p_6 \phi_{11} p_7$
0000011	$\phi_6 p_7$

Notice that the first Phi in each probability statement reflects the indexed parameter value from the corresponding position along the diagonal of the PIM, whereas the subsequent Phis are indexed to correspond with the internal PIM cells. Cooch and White also make the point that

because p7 is not estimable, both the phi11 and phi6 terms are not identifiable (do to the inherent confounding between the final p and phi terms in the CJS). The real parameter estimates reflect this.

If we were to continue this analysis, we'd also want to run phi(.) and phi(age) models in the interest of compiling a complete model set. But for now, we can just skip ahead in the interest of time.

Example 2 -

Now let's go through a slightly more complex example where we capture, mark, and recapture both young and adults simultaneously. In this scenario, let's assume we have two groups that reflect discrete age classes, individuals captured as juveniles (<1 year of age) and those captured as adults (> 1 year of age). We could construct two PIMs, one for each group, to reflect group-level differences related to age. But remember that under a simple g*t PIM structure we apply the same year effect (columns) across all cohorts (rows). What this means is that during interval 2 the survival rate for juveniles captured in year 1 (which are now technically adults) would be estimated the same as for individuals captured as juveniles during year 2. In reality the interval 2 survival for juveniles captured in year 1 should be identical to that for the adult group in interval 2, because at this point they are all adults and as adults presumably experience the same mortality risk.

1. Start a new project using the "age_ya.inp" file. Take a look at the Input file. Notice that now instead of a series of 1's and 0's representing group membership, we've got a summary of the number of individuals represented by each encounter history – a slight variation on the previous input file structures we've used, but still valid. The first group represents juveniles and the second reflects adults (this is where adding a descriptive comment to the head of the input file would be useful).
2. Create the file and open the PIM windows for the survival terms for both groups. We know from the previous example that we need to change the diagonal term for the juveniles to reflect a "graduation" into the adult age class. Let's do that for starters.

We'll take a brief side track here to explore the different shortcuts for tweaking the PIM windows using the menu options accessed by right clicking.

3. So at this point we could just do some quick parameter accounting and change the adult PIM to start counting up at 12, but would that be appropriate? Remember in this analysis there are only two age classes, and juveniles "graduate" to adults after their first year. Given that, does it make sense to model survival for adults originally marked as juveniles differently than adults marked as adults? Maybe in some very specific cases, but for the most part, no. So, how can we adjust the PIMs to reflect a common time-varying structure for survival among adults in each group? I'll give you a few minutes to mull it over, and then we'll compare notes. Remember in contrast to the previous exercise, we now have adults that are marked in occasion #1.

Once we've talked about the general principles for #3, let's sidetrack just a bit and again emphasize the importance of "accounting" when working with parameters, as there are a number of different ways you can index the parameters in this scenario and also some easy ways to screw up.

4. Set the Ps to reflect a time trend but no group (age) structure. Go ahead and run the model and label it phi(age*t) p(t).
5. As with the general group and time models, working in the PIMs is a less-flexible way to construct models than using the design matrix. However, we will always have to "set" the PIMs at least once in an analysis to the most general structure we plan to model. In the case of this analysis, you've just done this, and we can move to the design matrix from here.
6. From the menu bar, click on the Design menu and chose the "Reduced" option. Notice that if you try to select the "Full" option, MARK will tell you that it's not possible. This is because we've already altered the PIMs so that the most general model is no longer possible, and MARK only "knows" how to build a full, square matrix. This is no big deal as we need to know how to build a design matrix from scratch anyhow.
7. After choosing the reduced option, MARK will ask you how many columns you want to build. This should correspond to the number of parameters that will ultimately be needed for the linear model(s) you are trying to construct. We will start with the model we just built using the PIMs, the phi(age*t) p(t) model. First, recognize that we are really constructing two distinct linear models here; one for survival and one for detection. The detection model is the more straight-forward of the two, and has the following form:

$$logit(p) = \beta_1 + \beta_2 T_1 + \beta_3 T_2 + \beta_4 T_3 + \beta_5 T_4 + \beta_6 T_5$$

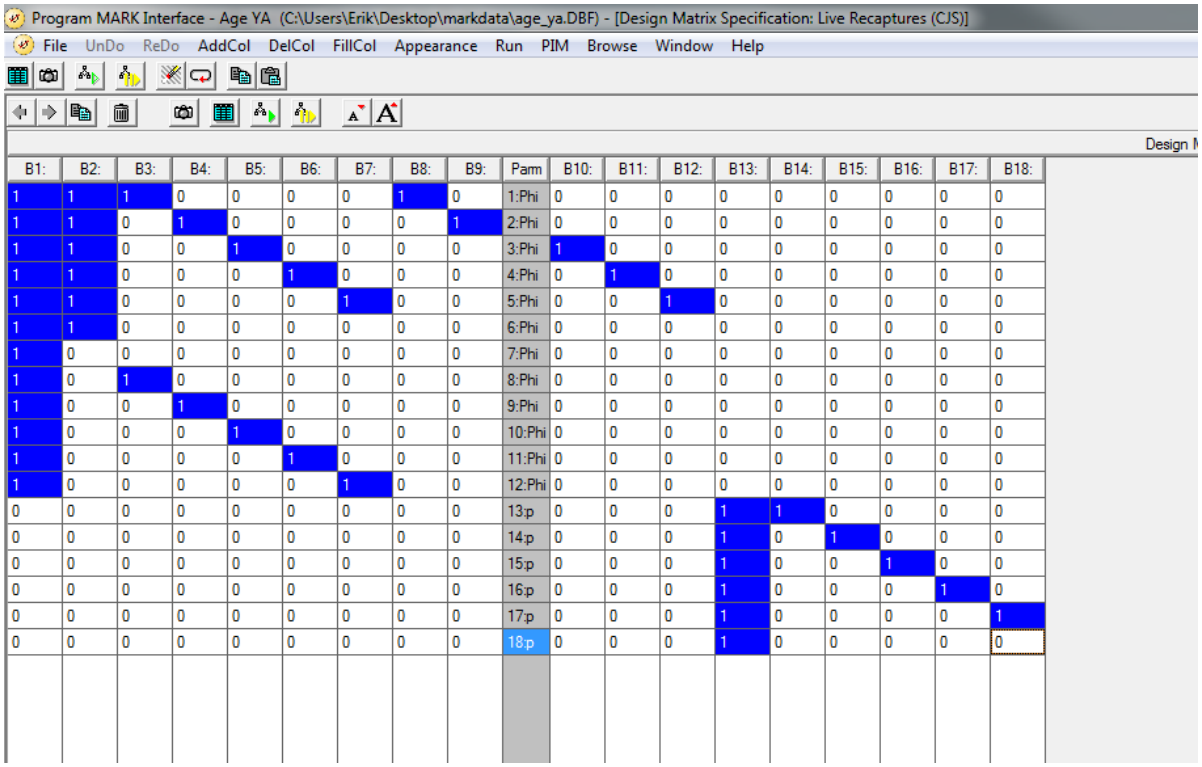
We see that there are 6 parameters associated with detection we will need to estimate. Also notice that the logit() notation tells us we are estimating the function on the logit scale. For just a second, let's revisit why there is no B7T6 term in this model?

So now how about the survival model. It will take the following, more complex form:

$$logit(\phi) = \beta_1 + \beta_2 A + \beta_3 T_1 + \beta_4 T_2 + \beta_5 T_3 + \beta_6 T_4 + \beta_7 T_5 + \beta_8 T_1 A + \beta_9 T_2 A + \beta_{10} T_3 A + \beta_{11} T_4 A + \beta_{12} T_5 A$$

Where now we've added additional terms describing the interaction between individual age and time (the BTA terms) for a total of 12 parameters. This means the combined number of parameters for the overall model should be 18 (12 survival and 6 recapture), which is the default number MARK gives us. Notice that as with earlier in lab today, we can add and delete columns from the DM as needed, so getting this number 100% correct is not crucial (you can build as you go). However, understanding your model well enough to know how many parameters you need to build is good practice in general.

8. At this point, rather than writing out step-by-step instructions we'll go through building the DM together in lab. If you are working back through this on your own, or if you missed class, see the text beginning on 7-23 for a more complex (and thorough) example from The Gentle Introduction. Ultimately, you will be left with something that like this:



B1:	B2:	B3:	B4:	B5:	B6:	B7:	B8:	B9:	Pam	B10:	B11:	B12:	B13:	B14:	B15:	B16:	B17:	B18:
1	1	1	0	0	0	0	1	0	1:Phi	0	0	0	0	0	0	0	0	0
1	1	0	1	0	0	0	0	1	2:Phi	0	0	0	0	0	0	0	0	0
1	1	0	0	1	0	0	0	0	3:Phi	1	0	0	0	0	0	0	0	0
1	1	0	0	0	1	0	0	0	4:Phi	0	1	0	0	0	0	0	0	0
1	1	0	0	0	0	1	0	0	5:Phi	0	0	1	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	6:Phi	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	7:Phi	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	8:Phi	0	0	0	0	0	0	0	0	0
1	0	0	1	0	0	0	0	0	9:Phi	0	0	0	0	0	0	0	0	0
1	0	0	0	1	0	0	0	0	10:Phi	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	11:Phi	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	1	0	0	12:Phi	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	13:p	0	0	0	1	1	0	0	0	0
0	0	0	0	0	0	0	0	0	14:p	0	0	0	1	0	1	0	0	0
0	0	0	0	0	0	0	0	0	15:p	0	0	0	1	0	0	1	0	0
0	0	0	0	0	0	0	0	0	16:p	0	0	0	1	0	0	0	1	0
0	0	0	0	0	0	0	0	0	17:p	0	0	0	1	0	0	0	0	1
0	0	0	0	0	0	0	0	0	18:p	0	0	0	1	0	0	0	0	0

You should recognize the familiar group*time structure (two stair-steps) on the phis, and the time-only structure on the p's from Part I of this lab. Go ahead and run this model, label it "phi(age*t) p(t) – DM" to contrast with the earlier model we ran using the PIMs. You should get identical results.

9. Now, let's tweak the DM to do a formal test for age structure with these data. Using what you learned earlier, and keeping the general time structure on p, run the following models:

Phi(age + t) p(t) Phi(t) p(t) Phi(age) p(t) Phi(.) p(t)

10. Lets look through the results, including the model selection table, real parameters, and betas.
11. We will finish with a brief discussion and example of model-averaging for age-structured analyses, and will touch on cohort variation that is not related to age. (time permitting)

Take home message – the biggest thing to take away from these exercises is that we can realistically incorporate an age structure into our models. This allows us to both account for an important potential source of heterogeneity among individuals in our datasets, and to test for age-related effects on demographic rates. The PIMs allow us to index the parameters to

appropriately graduate individuals among age classes, and the Design Matrix gives us a flexible way of modelling age-associated variation.

If you read through Chapters 6 and 7 of the Gentle Introduction you will see there are some potentially useful tools we did not cover this week. Specifically I did not include consideration of temporal trends or group-level continuous covariates, which the design matrix allows us to incorporate and are presented in Ch 6. We will cover these along with individual covariates next week when we explore various ways to model survival as a function of continuous effects (until now all effects we've tested have been discrete). Also at the end of Ch 7 you will find a detailed description of accounting for transience in encounter data in a manner that is conceptually very similar to modelling age structure. Transience is the case where you expect that some non-trivial number of individuals may permanently emigrate from your study area immediately after capture, never to be seen again (for example, if you are capturing resident and migratory animals in the same system). If you feel this is relevant to the type of work you will be doing, or your semester project, I highly recommend reading through that portion of the Chapter (begins page 7-29).

Lab 2 Assignment

This assignment will require you to apply principles we covered today related to both setting age structure in the PIMs and using the Design Matrix to run models. On Blackboard under the Assignments Page you will find an input file labeled Lab2HW.inp. This contains a simulated capture history for two age classes of hypothetical organisms. Assume these are waterfowl (ducks) that are banded in a large wetland complex each year at the end of the summer during their molt when they are flightless. At capture, birds are aged as either Hatch Year (HY) or After Hatch Year (AHY), which naturally suggests that HY birds become AHY birds in their second year. You are going to evaluate whether the survival from August to August each year differs between HY and AHY groups. For simplicity we will assume that recapture success is constant across years and age classes, so you can use a p(.) structure for all models. I would also like you to test the hypothesis that survival varies through time for only the HY birds, whereas survival is constant for the AHY birds. This hypothesis is rooted in life history theory, which predicts that for certain life history strategies the survival of mature animals should be less sensitive in general to environmental variation compared to younger organisms. The suite of models you should run are:

1. Phi (Age * Year) p (.)
2. Phi (Age + Year) p (.)
3. Phi (Age) p (.)
4. Phi (Year) p (.)
5. Phi (.) p (.)
6. Phi (Age + Year[HY Only]) p (.)

I suspect the last model might be a little challenging for you to figure out, but if you can make the connection between about how the beta estimates (columns in the Design Matrix) correspond with the linear model you are attempting to construct, and how that in turn defines your model structure, you've got a good working understanding of what we've covered up to this

point. From your completed MARK analysis, export your AIC table and your real parameter estimates, copy and paste them into a Word document, and upload this as your completed assignment. Also please attach the .dbf and .fpt files associated with your MARK analysis. Don't spend a lot of time worrying about formatting for you results tables, but please make sure what you submit is readable. The assignment is due prior to the start of lab next week.

When imprinting studies go awry