WILD 502
Lab 5 – Live Recaptures – Age + Individual Covariates

Today's lab works with live-recaptures data and adds consideration of age and individual covariates (*SwampSquirrels.inp*). The data are for the imaginary swamp squirrel and include 2 covariates: (1) *birth* - day of year on which the animal was born, and (2) *tail* - tail length in cm at time of initial capture). Animals were captured as either juveniles (young of the year in late summer) or as older animals (yearlings or older, which could not be distinguished from one another with the field methods used). At initial capture, each individual received a PIT tag and had their tail length measured.

The questions of interest are: (1) does apparent survival rate (φ) vary with age class (juveniles or older), birth date, and/or tail length. On the log-odds scale, please evaluate (1) a possible linear relationship between φ and length and (2) the possibility of linear and quadratic relationship φ and tail length. To keep the model list a bit simpler, all models with birthday or tail length should also include age class. Also, when combining age class with birth date and/or tail length, always allow age classes to have different intercepts but constrain them to have the same slopes. This leads to 7 model structures for φ.

Because all squirrels will be at least yearlings by the time you start estimating recapture probability (*p*) for any of them (e.g., juveniles captured in year 1 will be yearlings in year 2 when you first estimate *p* for them), you don't need to assess whether *p* might vary with age. However, please do evaluate possible temporal variation in *p* by evaluating support for models that hold *p* constant, allow *p* to be different in each week (no pattern), or constrain *p* to be a function of precipitation levels during capture sessions (values = 0.11, 1.38, 1.12, 1.58, and 1.87 for years 1 through 5). This leads to 3 model structures for *p*.

	Model Structure	
Model #	*phi*	*p*
1, 2, 3	~1	~1, ~ time, ~ precip
4, 5, 6	~age_class	~1, ~ time, ~ precip
7, 8, 9	~age_class + birth	~1, ~ time, ~ precip
10, 11, 12	~age_class + tail	~1, ~ time, ~ precip
13, 14, 15	~age_class + tail + tail^2	~1, ~ time, ~ precip
16, 17, 18	~age_class + birth + tail	~1, ~ time, ~ precip
19, 20, 21	~age_class + birt + tail + tail^2	~1, ~ time, ~ precip

Use Program MARK to evaluate the 21 models that represent all possible combinations of the 7 model structures for φ and the 3 model structures for *p*. So that you can just focus on working with the individual covariates for a scenario with 2 groups and 2 parameter types, you do <u>not</u> need to concern yourself with possible overdispersion for this lab.

You might find Chapters 7 (Age & Cohort Models) & 11 (Individual Covariates) of *Cooch & White* to be helpful to you if you get stuck, but we have covered the relevant information for running these models and interpreting the output. Notice that you will enter information on covariates in 3 different ways today: (1) you will provide information on age class at capture using 2 groups (you will have separate PIMs for each group, which allows you flexibility in working with the time-varying age covariate as needed in your modeling), (2) you will provide information on individual covariates in the rows of the input file, and (3) you'll enter time-varying environmental values that apply to all individuals in the Design Matrix.

The input file: 6 occasions, 2 groups (juveniles and older), and 2 covariates (birth and tail). The numbers in the 2 group columns indicate how many animals in each group are being represented on each row (in row 1, you are providing MARK with information for 0 juveniles and 1 older animal).

```
SwampSquirrels.inp    ×
  0           1,0           2,0
1 100000   0 1 26.5   11;
2 100000   0 1 25    24;
3 100000   1 0 22.5  13.5;
4 101100   0 1 28.5   21;
5 100000   1 0 30    24;
6 100110   0 1 23.5   24;
7 101000   1 0 20.5   15;
8 100110   0 1 27    18;
9 100000   1 0 24.5   24;
0 110100   1 0 24    12.5;
1 100000   1 0 18    17.5;
```
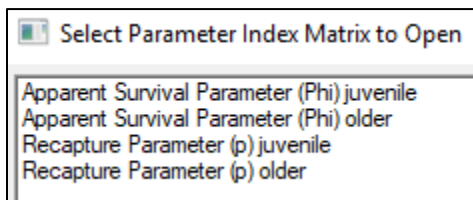
### Initiating the MARK run
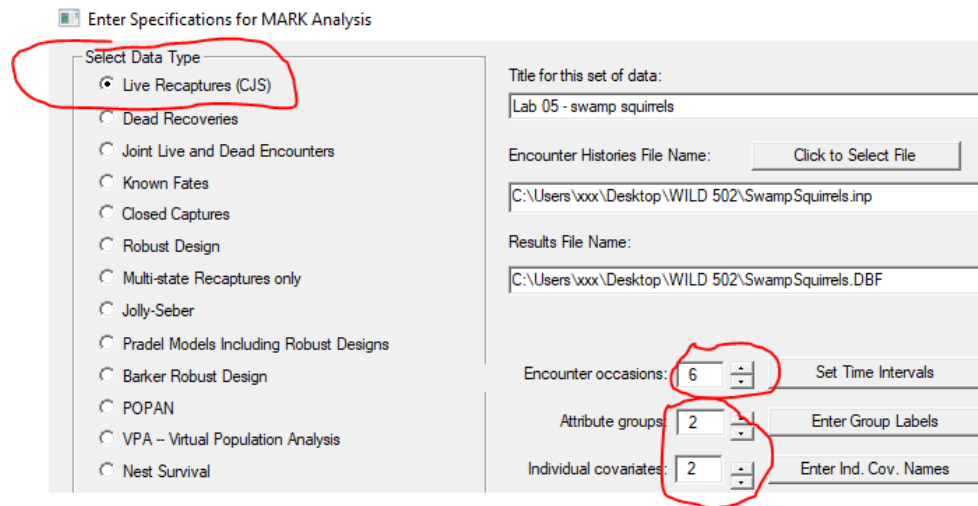
Provide names for your groups (*juvenile* and *older*) and for your 2 covariates (*birth* and *tail*).

After you've initialized the project in MARK, close the PIM window that appears and then click on the PIM menu and choose "Open Parameter Index Matrix". Notice that you have 2 parameter types for each of 2 groups.

Setting up the PIMs for your modeling

You need to think about what you want to do with the PIM values for each parameter type. You don't have any time variation in any of your models for Phi but you do have age variation in most of them. For squirrels first caught as juveniles, you'll want to number the Phi PIM with 2 age classes (1's for juveniles on the 1st diagonal and 2's everywhere else to represent that you want a different parameter for Phi for older squirrels). You can then put in all 2's in the Phi PIM for squirrels first caught as older animals.
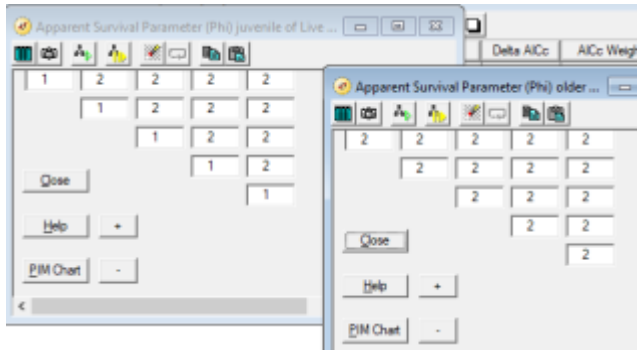

For PIMs for *p*, you have no models with age but have a lot of models with time variation. So, it's necessary to let the PIM values vary by year. You can number the PIMS for *p* the same across age groups though. This will allow you to use 1 set of PIM numbering for all of your models while not leaving you with a bunch of unnecessary rows in your Design Matrices for most models (though there will be some redundant rows in the Design Matrices for some of your simpler models).
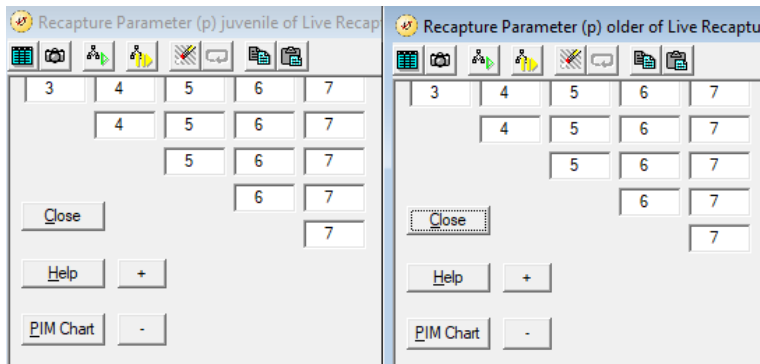
Setting up the Design Matrix for various models
Now, that you've settled on having 7 different PIM values and so 7 rows in your Design Matrix, you might find it useful to label the rows in your Design Matrix before running models.
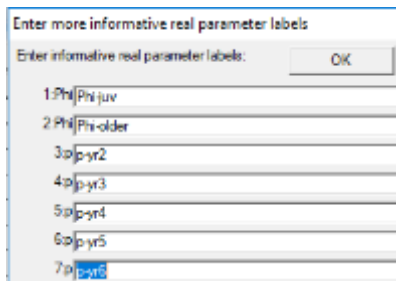
Design Matrix for "Phi(age_class), p(time)" (here are 2 ways)

B0:	B1:	B2:	Parm	B3:	B4:	B5:	B6:
1	0	0	1:Phi-juv	0	0	0	0
0	1	0	2:Phi-older	0	0	0	0
0	0	1	3:p-yr2	0	0	0	0
0	0	0	4:p-yr3	1	0	0	0
0	0	0	5:p-yr4	0	1	0	0
0	0	0	6:p-yr5	0	0	1	0
0	0	0	7:p-yr6	0	0	0	1

or

B0:	B1:	B2:	Parm	B3:	B4:	B5:	B6:
1	0	0	1:Phi-juv	0	0	0	0
1	1	0	2:Phi-older	0	0	0	0
0	0	1	3:p-yr2	0	0	0	0
0	0	0	4:p-yr3	1	0	0	0
0	0	0	5:p-yr4	0	1	0	0
0	0	0	6:p-yr5	0	0	1	0
0	0	0	7:p-yr6	0	0	0	1

Design Matrix for "Phi(age_class + birth + tail + tail^2), p(time)" - notice that here, I used the "Appearance" menu that is available when the Design Matrix window is open and labeled the columns. You don't have to do this but it does make the output easier to follow.

B0 int-phi	B1 int_adj-older	B2 birth	Parm	B3 tail	B4 tail-sq	B5 int-p	B6 precip
1	0	birth	1:Phi-juv	tail	product(tail,tail)	0	0
1	1	birth	2:Phi-older	tail	product(tail,tail)	0	0
0	0	0	3:p-yr2	0	0	1	0.11
0	0	0	4:p-yr3	0	0	1	1.38
0	0	0	5:p-yr4	0	0	1	1.12
0	0	0	6:p-yr5	0	0	1	1.58
0	0	0	7:p-yr6	0	0	1	1.87

You should be able to work out how to run all of your models and do so quite efficiently by retrieving a model and tweaking the Design Matrix as necessary. I think you might be wise to label both rows and columns to help you keep track of things (I won't be checking on that, so do as you see fit). I also recommend that you run all of your models using the logit link, so you don't have to switch back and forth. That said, do pay attention to your parameter counts and make sure all is as it should be. Here are results for estimated beta's for the model directly above.

```
         LOGIT Link Function Parameters of {Phi(age+birth+tail+tail.sq), p(precip)}
                                                   95% Confidence Interval
Parameter                    Beta         Standard Error      Lower            Upper
-------------------------   -------------  --------------   --------------   --------------
    1:int-phi               -5.0243288     2.0296061        -9.0023568       -1.0463008
    2:int_adj-older          1.6786643     0.3385935         1.0150211        2.3423075
    3:birth                 -0.0582655     0.0297969        -0.1166674        0.1363510E-003
    4:tail                   0.8179373     0.2474928         0.3328514        1.3030231
    5:tail-sq               -0.0225186     0.0070058        -0.0362500       -0.0087872
    6:int-p                 -1.4161917     0.2060092        -1.8199699       -1.0124136
    7:precip                 0.3596395     0.1412747         0.0827411        0.6365379
```

```
     Real Function Parameters of {Phi(age+birth+tail+tail.sq), p(precip)}
   Following estimates based on unstandardized individual covariate values:
       Variable    Value
       ---------   -------------
       BIRTH       22.416667
       TAIL        17.479333

                                                      95% Confidence Interval
    Parameter                Estimate     Standard Error     Lower        Upper
    --------------------     ----------   --------------   ----------   ----------
       1:Phi-juv             0.7477242    0.0546093        0.6269349    0.8394218
       2:Phi-older           0.9407647    0.0207238        0.8845541    0.9705185
       3:p-yr2               0.2015506    0.0309226        0.1476379    0.2689386
       4:p-yr3               0.2849805    0.0184503        0.2502349    0.3224754
       5:p-yr4               0.2663159    0.0183688        0.2318925    0.3038279
       6:p-yr5               0.2998607    0.0205091        0.2612599    0.3415280
       7:p-yr6               0.3222030    0.0261415        0.2732329    0.3754161
```

I think you can see that by labeling the beta columns and the parameter names that the output is easier to keep track of. Of course, it comes at the cost of doing some extra typing, and … if you make errors in your labeling (!), you might never notice and make some incorrect statements (e.g., imagine mistakes you might well make if you mis-label your beta columns). So, be careful when you label.

Once you've run your models and identified which models are well-supported by the data, you'll want to examine the beta-hats and produce estimates of φ and p. You can certainly get useful plots from Program MARK or Excel using the "individual covariates plot" tool. If you'd like to make plots in R, you can use MARK's "individual covariates plot" tool to send predicted values to Excel. For example, for the model above, you can set up a plot using the information below.
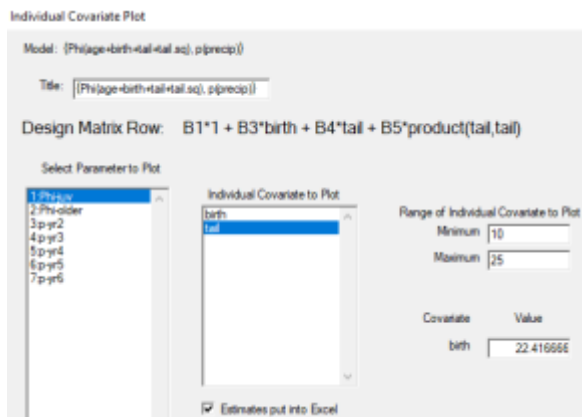
This will send key data to Excel. If you then do the same for "Phi-older" with the exact same set up for covariate conditions, you'll end up with 2 Excel spreadsheets. You can then copy and paste the rows of data for the older squirrels in rows just below the rows for juveniles. Then, you'll need to (a) rename the 2nd column as "phi", (b) add a column 6 that's "age_class" (fill in rows with "juvenile" or "older"), and (c) a column 7 that contains the value you used for "birth" (e.g., 22.42 for average). You can then use the code below in R to obtain a nice plot using things you've learned about copying data to the clipboard and then to R. You can adapt this for other covariate conditions of interest and for other models with just a bit of work. Wherever you make your plots, set your axes to be the same on all plots as it's harder to compare plots if the y-axis limits vary among plots. When copying data from Excel to R via the clipboard, use header = FALSE in the R code if you only copy the data and not the row of column names.

```
 phi_hats <- read.table("clipboard", header = FALSE,
            col.names = c("tail", "phi", "se", "lcl",
                          "ucl", "age_class", "birth"))
library(ggplot2)
ggplot(phi_hats, aes(x = tail, y = phi)) +
  geom_ribbon(aes(ymin = lcl, ymax = ucl), fill = "grey80") +
  geom_line(aes(color = age_class)) + ylim(0, 1) +
  facet_wrap( ~ age_class) + theme(legend.position="none") +
  annotate("text", label = "for avg birth date (22.4)",
           x = 17, y = 0.1, size = 4)
```

Lab Assignment

1. Briefly describe the data set (key features of mark-recapture study & covariate conditions).

2. Provide a table of model-selection results and text describing what you learned about possible sources of variation in φ and p from the set of competing models and data you worked with in this lab. Be sure to discuss model-selection uncertainty, what well-supported models had in common, and how well-supported models differed.

3. Expand on your presentation of model-selection results by providing key information on (a) the regression coefficients (and associated measures of uncertainty) in well-supported models and (b) estimates of φ and p for relevant covariate conditions given the model structures that were well-supported. This portion of your results should include tables and appropriate graphics.

   If your well-supported models contain both birth date and tail length, consider how best to present the results of how φ changes as a function of age class and both continuous covariates. One way to do it is to present a separate graphic for each age class with one continuous covariate on the x-axis while holding the 2ⁿᵈ continuous covariate at a particular value or values of interest (e.g., it's mean, 5% quantile, and 95% quantile). If you predict survival for each age class across the range of tail lengths for squirrels with (a) mean birth date, (b) 5% quantile for birth date, and (c) 95% quantile for birth date, you would have 6 plots (2 age classes x 3 birth date values). You can obtain such results by working with Program MARK's individual covariate tools and Excel and/or R using skills you've learned in previous weeks. If you think that this should be done for multiple models, please do so.

4. Based on your model-selection results and what you learn from each well-supported model, what inferences can you draw about sources of variation in φ and values for φ?

5. Based on your model-selection results and what you learn from each well-supported model, what inferences can you draw about sources of variation in *p* and values for *p*?

6. Please provide model-averaged predictions of φ for a juvenile squirrel and for an older squirrel with average covariate values and provide comments on (a) how much rates varied for a given parameter across the models, (b) how much uncertainty there was among the estimates form well-supported models, and (c) how much the model-averaged estimate and SE for the model-averaged value differed from the values from the best-supported model.

   The default when you run a model is for MARK to estimate the real parameter values using the average covariate values. As long as you didn't change that default on any of your model runs, the model-averaging here is easy to do. You simply choose the 1ˢᵗ and 2ⁿᵈ cells in the 1ˢᵗ row on the chart that comes up under the tab for apparent survival of juveniles (this provides you with a juvenile rate and a rate for older squirrels [2ⁿᵈ parameter]). Of course, if you wanted model-averaged values for lots of other covariate conditions, it would take some work and this is where RMark can be very useful. But, doing the model averaging here, let's you see the output for each model and get some experience with the output and ideas.

7. Please briefly list out any questions that you have about this week's lab that you'd like to have discussed at the start of next week's lab.