WILD 502
Lab 3 – Introduction to CJS Models

This week, you'll estimate apparent survival from live animal recaptures or resightings using the Cormack-Jolly-Seber (CJS) model. In addition to estimating apparent survival (φ), you will also be estimating recapture probability p. Keep in mind the distinction between φ and S: $\varphi = S \cdot F = S \cdot (1 - E)$, where S is the probability that the animal is still alive, F is fidelity, and F equals $1 – E$, where E is emigration. φ estimates the probability that the animal is still alive and remains on the study area available for recapture. Thus, $\varphi \leq S$.

The lab exercise today takes advantage of excellent material written by Cooch and White on using Program MARK. You should first review Chapters 1 and 2, which provide useful introductory material and will help you review basics of MARK. To conduct this lab, you should carefully work through Chapters 3 and 4 of Cooch and White. These chapters are linked from the course web page. Chapter 3 won't take long, but you'll need to spend some quality time with Chapter 4 to truly understand what's presented, i.e., don't just run the models!

The data used in this exercise were collected on European Dippers in eastern France (presented by Lebreton et al. 1992:85-89). Birds were banded for 7 consecutive years during early summer along streams, providing 6 re-encounter occasions. For this lab, you will just work with data for the males (**ed-males.inp**, which can be downloaded from the course schedule page).

1. Examine the input file using notepad or other text editor of your choice. Notice that there is one row for each individual. This file could have been compressed into a file that reads as seen below (but without the last 2 columns). There are data for 141 animals in the file. There are 25 different encounter histories, and the 17 histories coded 0000001 could be deleted without losing any information.

EH	n		Ri	"releases"
1111110	1	;	12	6
1111000	1	;		4
1100000	4	;		8
1010000	1	;		2
1000000	5	;		5
0111100	1	;	20	4
0111000	1	;		3
0110000	7	;		14
0100000	11	;		11
0011110	1	;	25	4
0011100	4	;		12
0011000	8	;		16
0010110	1	;		3
0010000	11	;		11
0001111	6	;	22	18
0001110	3	;		9
0001100	6	;		12
0001001	1	;		1
0001000	6	;		6
0000111	10	;	22	20
0000110	3	;		6
0000100	9	;		9
0000011	12	;	23	12
0000010	11	;		11
0000001	17	;	17	
/* SUM	141 */		141	207

On occasions 1-7; 12, 20, 25, 22, 22, 23, & 17 new tags were put out.
The effective sample size is the number of animals released or re-released during the study not counting the last year, n_{ess} = 207.

2. Open MARK and work your way through Chapter 3 of Cooch & White. Much of this chapter will be old hat for you as you've already worked with MARK but … the chapter is well written and may help clear up some nagging confusion!

   a. Note the confounding of φ_6 and p_7. We will discuss this further (see Chapter 4 of C&W).

   b. Read the material on the Likelihood Ratio Tests. We won't work with significance tests but it's a nice reinforcement of material we covered earlier on the limitations of this approach to model selection.

3. Next, work your way through Chapter 4 of Cooch & White. This chapter will be a little more challenging for you as it introduces some new ideas and is, in a few places, a bit ahead of lecture. Also, some of the sidebars are quite technical and beyond our scope.

   a. Review the connections among PIMs, the Design Matrix, the link function, and the estimates of the real parameters.

      i. Because we use an identity matrix as the design matrix for every model being run this week, we can use the 'sine' link. This link function is the default choice when an identity matrix is used because when it's used MARK does a better job of counting how many parameters were estimated.

      ii. It's important to keep track of which link function was used when you go to transform the beta's into the real parameters. Fortunately, it's easy to do: *you just have to remember to do it*. One way is to look at the 2nd line of text in the beta estimates window.

      iii. For the SIN link function, an estimated beta for a real parameter is converted to an estimate of the real parameter using the equation: $real = (\sin(\hat{\beta})+1)/2$. Try this for the 1st beta estimated in the table above and you'll obtain the estimate of $\hat{\phi}_1 = 0.697$ <u>as long as you work in radians</u>. In *R*, you can just type, `(sin(0.4049138)+1)/2 # = 0.6969698`

   b. Be sure you understand how the CI's for the real parameters are obtained and why this is so. We have discussed this for known fate models, and it operates the same in CJS models. For the models used this week where only 1 beta is used to estimate each real parameter, it's easy to do: just use the CI's on the betas and the appropriate transform (sine or logit) to obtain real parameter CI's.

c. Be sure you understand how the likelihood function for this multinomial process looks for each model. That is, consider what each model is saying about time variation in φ and p. Then consider again why we're doing maximum likelihood estimation and what the -2lnL scores are used for.

d. Consider how the -2lnL value for each model and model structure (# parameters) is used to obtain AIC and AICc values.

e. Work through how ΔAICc values and AICc weights are obtained and how they are used. The text in chapter 4 of *C&W* is very helpful on these topics.

f. The -2 log likelihood for the saturated model is needed to compute the deviance. For CJS models, this is computed using each of the observed encounter histories. For each encounter history, the quantity $n_j \cdot \ln\left(\frac{n_j}{R_j}\right)$ is computed, and then these values are summed across all encounter histories (and groups, if multiple groups are used). Thus, the $\ln L_{sat}$ = -138.34
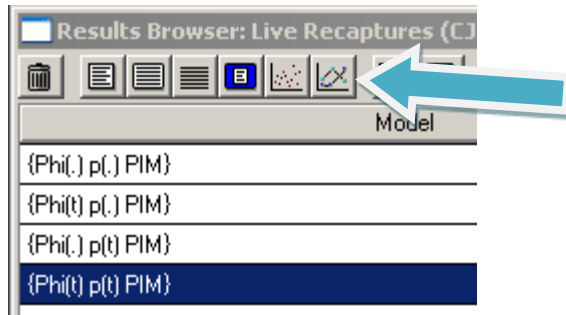
$$1 \cdot \ln\left(\frac{1}{12}\right) + 1 \cdot \ln\left(\frac{1}{12}\right) + 4 \cdot \ln\left(\frac{4}{12}\right) + 1 \cdot \ln\left(\frac{1}{12}\right) + 5 \cdot \ln\left(\frac{5}{12}\right) +$$
$$1 \cdot \ln\left(\frac{1}{20}\right) + 1 \cdot \ln\left(\frac{1}{20}\right) + 7 \cdot \ln\left(\frac{7}{20}\right) + 11 \cdot \ln\left(\frac{11}{20}\right) +$$
$$...$$
$$12 \cdot \ln\left(\frac{12}{23}\right) + 11 \cdot \ln\left(\frac{11}{23}\right) +$$
$$17 \cdot \ln\left(\frac{17}{17}\right) =$$
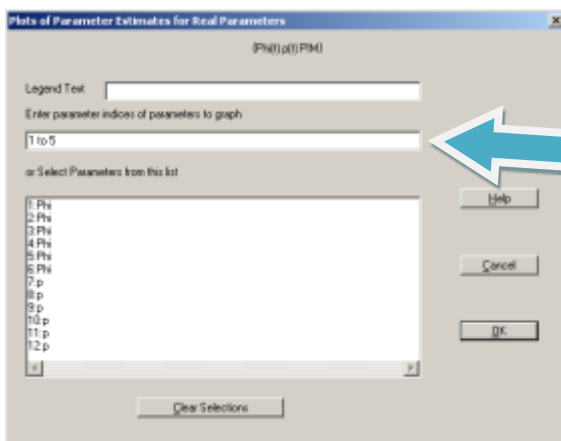$$-138.34$$

-2lnL$_{sat}$ = 276.68

    i. Note that: $17 \cdot \ln\left(\frac{17}{17}\right) = 0$ and so contributes nothing to the likelihood, i.e., the encounter history 000001 didn't help us estimate anything.

    ii. 276.68 is the −2lnL for a model that fits perfectly, i.e., it's analogous to a line fit to 2 data points, and can be used as a baseline for evaluating the fit of other models, i.e., by measuring how much worse their -2lnL score is than 276.68.

Models that fit better have higher Likelihood values. Likelihood values are usually small positive numbers (e.g., 0<L<1). Thus, lnL values are typically negative numbers, and better models have higher lnL values. E.g., L_1=.8, lnL_1=-0.22; L_2=.4, lnL_2=-0.92. When converted to -2lnL values, things reverse! Models that fit better have smaller -2lnL values: $-2lnL_1$=0.446 and $-2lnL_2$=1.833.

g. Here's how to obtain the graphics described in chapter 4.

1. Highlight the model of interest and then click on this icon

2. Fill in the numbers for the parameters that you want plotted

h. Text in the Addendum to Chapter 4 of C&W (starts on page 4-70) is very useful regarding how the number of estimated parameters is determined (though some of it does get quite technical and might take some solid math background). The text in the Addendum regarding non-identifiable parameters is also very well written and worth working through carefully.

i. Text starting on page 4-33 of C&W provides a nice overview on working with ΔAICc to compare models and make inferences in the face of model-selection uncertainty.   We won't get into BIC, LRT much but at least you can read of them and gain some exposure to the ideas that other model-selection methods exist.

Some rules of thumb for interpreting ΔAICc (from text starting on C&W 4-39)
When using AIC values, how much 'support' is there for selecting one model over the other? When the difference in AIC between 2 models (ΔAIC) is < 2, then we are reasonably safe is saying that both models have approximately equal weight in the data. If 2 < ΔAIC < 7, then there is considerable support for a real difference between the models, and if ΔAIC > 7, then there is strong evidence to support the conclusion of differences between the models. The ΔAIC values can be used to calculate relative model weights, too.

Your homework assignment

1. For a study with 7 occasions, use words, graphics, or both to explain
   a. How many $\hat{\phi}_i$ and $\hat{p}_i$ exist and
   b. How many of those can be estimated if data are present for all occasions.
2. Run the pre-defined $\varphi(t),p(t)$ model using the *sine link*.
   a. How many parameters are listed as having been estimated in the *Results Browser*?
   b. Run the model again but this time use the *logit link*. Now how many parameters are listed as having been estimated in the *Results Browser*? Take a look at the parameter estimates and their estimated standard errors. Can you find a parameter estimate that's on a boundary? Which is it, and what is its value and SE?
   c. For the model run using the *logit link*, adjust the number of parameters to the appropriate number. Once this is done, report on how the AICc values compare for the model run with the *sine link* vs. the *logit link*.
   d. What is the MLE for $\hat{p}_i$ for occasion 3? From looking at the input data and the model structure, explain why this value makes sense.
3. Clean up the list of models in the Results Browser so that only the following 4 models are listed: $\varphi(t),p(t)$; $\varphi(t),p(.)$; $\varphi(.),p(t)$;   and $\varphi(.),p(.)$
   a. Provide a modified version of MARK's AIC table that also includes 3 other columns: the -2lnL, lnL, and L value for each model.
   b. Based on ΔAIC values and model weights, what evidence is there for temporal variation in φ and p, and how strong is that evidence?
   c. What can you learn about φ_6 from each model in the set of 4 models?
   d. What inferences can you draw about φ and p from this study and how certain are you of those inferences (express the uncertainty as you see fit).
4. Given the MLE's for the best model in the list, what is the probability of observing each of the following encounter histories (show your calculations):
   a. 1111111
   b. 0011001
   c. 0001101
5. What does the M_{ij} array look like for this study? MARK will provide it to you if you click on 'Output', then on 'Input Data Summary' on the top menu.
6. What do the numbers in the M_{ij} array represent and how do they relate to the encounter histories and effective sample size?
7. Use page 421 of your textbook, the input data, and estimates for model $\varphi(.),p(.)$.
   a. What is the formula for calculating the expected number of recaptures in the M_{ij} array for animals released on the occasion 1 and first recaptured on occasion 3?
   b. What is the expected number of recaptures in the M_{ij} array for animals released on the occasion 1 and first recaptured on occasion 3? Show your work.
8. Using 2 examples for species that you are interested in, and familiar with, describe scenarios where CJS modeling of φ and p would be:
   a. a useful way to learn about survival in some species, and
   b. not a useful way to learn to learn about survival in some species.
9. Provide a reference to 1 paper in the ecological literature from the past 5 years that used a CJS modeling approach to investigate survival rates. Provide a very brief summary of the question that they addressed, the length of the study, the number of animals studied, and the main conclusion of the work.