# Lab 07 – Models for open populations & GOF

Download the spreadsheet lab_07.xls. The spreadsheet contains worksheets entitled, Questions, Tag Recovery, and Recapture. Both include the simulated striped bass data we examined in the lecture examples. The 2nd and 3rd worksheets are divided into several sections: Data (including the m-array of observed tag recoveries or recaptures), Parameters (including initial values for the optimization), Cell probabilities, Expected values (E($m_{ij}$)), Log-likelihood (ln($\mathcal{L}(m_{ij})$)), and the Log-likelihood of the model ((ln($\mathcal{L}(\theta)$))).

## Questions

Answer the following on the spreadsheet labeled questions:

1. Describe the difference between a probability distribution and a likelihood function.

2. Compare and contrast the binomial distribution and the multinomial distribution.

3. What is over-dispersion, and how might it affect your results with respect to model selection, parameter bias, and precision of estimates?

## Tag Recovery models

Extract the file striper.inp from lab_07.zip and create a new results file in MARK by specifying recoveries only model with 5 occasions, and 1 group.

1. Cast and run the models...

   $$\{S_t\ r_t\}\ \{S.\ r_t\}\ \{S.\ r\}\ \{S.\ r_3\}$$

   using the design matrix. Create a table from the data in the Results Browser (use Output/Table of Model Results) and paste it in the top section of the MARK Results worksheet. Be sure to examine the results and eliminate or adjust parameter counts for the models where appropriate. Below the table of results interpret the model selection.

2. Cast and run the saturated model. Are the parameters in this model estimable? Is this model useful from a biological standpoint? Why or why not?

3. Construct the model $\{S_t\ r.\}$ and compare it to $\{S.\ r_t\}$. This model, $\{S_t\ r.\}$, appears to be a better model than $\{S.\ r_t\}$, yet the researchers designing this study *a priori* chose not to consider it. Can you conceive of a reason why?

## CJS models

1. Extract the file striper3.inp and create a new results file in MARK by specifying recaptures only with 6 occasions, and 1 group. **Note**: This file is based on encounter histories. MARK will not accept $m$-arrays as input for live recapture data.
2. Cast and run the models...

   $$\{\phi_t\ p_t\},\ \{\phi.\ p_t\},\ \{\phi.\ p.\},\ \text{and}\ \{\phi.\ p_3\}.$$

3. Create a table from the data in the Results Browser (use Output/Table of Model Results) and paste it in the *Recapture* worksheet.

## Goodness of Fit

Make a copy of the MARK input file dipper.inp and rename it GOF_dipper.inp. Open the input file to determine how to set up the live recaptures data type. Run models corresponding to hypotheses that survival differs among years, survival that follows a trend over time, survival differs between males and females, survival is different for each sex and year, and an additive model for sex and time trends all with recapture rates that vary by sex and year. Paste the results from these five models in the top portion of the worksheet labeled Dipper Results.

1. Examine the MARK output for the most general model (Phi(g*t) p(g*t)) and enter the values for the number of estimated parameters, model deviance, deviance degrees of freedom and deviance c-hat.
2. Run the program RELEASE GOF from within program MARK locate the Goodness of Fit Results (TEST 2 + TEST 3) by Group.
3. The Bootstrap GOF page contains 100 bootstrap simulations of the deviance for the model. What is the probability of the observed deviance c-hat? What is the recommended next step?
4. Run the median c-hat test. Enter the estimate of c-hat and its standard error in the worksheet.
5. Update the model results using your estimate of c-hat from the median test and paste the results at the bottom of the page.

Using the most general model, follow the approach outlined in the lecture notes and the MARK help file to calculate median-c-hat (Tests/Median c-hat). First run the simulations with the default bounds in the dialog box. This is used to get an initial estimate of c-hat. Then repeat the process using 10 values of c-hat $\pm$10% of the initial estimate and 100-200 trials per value. Use the resulting estimate of c-hat to calculate QAICc for the modeling results in MARK (Adjustments/c-hat). Paste these results in the middle portion of the spreadsheet.

Compare these tables, and in the lower section of the spreadsheet describe why GOF tests are important when conduction model selection based on AIC, whether lack of fit is a significant problem in this data set and how you know that, and how lack of fit influenced the model selection results in the example. In your answer be sure to consider changes $\Delta$AIC, model weights, and model uncertainty.