

WILD 502

Lab 4 – CJS Models: comparing groups, QAICc, & model averaging

This lab will extend your abilities for working with data from live capture-recapture studies and give you experience with several important topics: (1) adjusting AICc values for overdispersion, (2) using model averaging to bring model-selection uncertainty to bear on inferences, & (3) working with differences between parameter estimates given that we have sampling variance-covariance for the estimates.

Today's lab builds on material in Chapters 4 & 5 of Cooch and White. The data and models we'll use are described in section 4.2.2 of Cooch & White (pg 4-19). The data come from a study of the swift (*Apus apus*) in southern France. The species nests nearly exclusively in buildings. The data analyzed were collected on birds nesting on the Museum of Natural History in Nîmes in Southern France from 1981-1989. The data come from two neighboring colonies: one colony faced west in a cloister that was closed to the public (GOOD), whereas the other colony faced south and was along a busy street and exposed to swirling winds (POOR). The birds are highly faithful to their nesting colonies. Data for the 2 colonies are entered in 2 groups with the data for Poor in group 1 and Good in group 2. Differences in colony conditions were hypothesized to affect survival rates directly by affecting the survival and indirectly by affecting the quality of individuals at each site. An analysis of these data was presented by Lebreton et al. (1992). The goodness-of-fit topics we'll use in lab are covered in pages 5-1 to 5-7, 5-27 to 5-30, and 5-37 to 5-40. The rest of chapter 5 is useful, but we won't use RELEASE or U-CARE today.



Key Topics for This Week's Lab

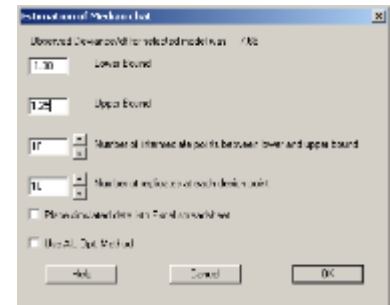
1. Model set doesn't include all possible models
 - a. model list would be thought out carefully but here we take the list as given
 - b. Only possible sources of variation considered here are *time & group*
2. Can change the PIM's if like to run various models and can do this by opening each PIM window (1 for each parameter type and group = 4 windows here). The PIM Chart offers a quick way to change the PIMs in one window. Chapter 4 of C&W shows how to use the PIM Chart (see sections 4.2.1 & 4.2.2).
3. Can also run each model using the Design Matrix. I think that this is valuable to be able to do for every model. But, often you can change the PIM to reduce the size of the Design Matrix
4. Want to check GOF for most complex model in list (Chapter 5 of C&W).
 - a. Fit problems come because
 - i. Missing covariates that are important
 - ii. Wrong variance = binomial is too simple = have extrabinomial variation
 - iii. Either problem leads to overdispersed data and underestimates of variance. This causes you to overfit models
 - b. Solutions can be obtained by a variety of methods.
 - i. For CJS, can use
 1. RELEASE & U-CARE = look for systematic fit problems
 2. Median c-hat (we'll use that)
 - c. Median c-hat procedure works with deviance/df

- i. We don't know how the dev/df is actually distributed when the model fits (e.g., when null hypothesis of good fit is correct). So, we don't know what values of dev/df to consider problematic.
- ii. Instead, we use median c -hat procedure to estimate c -hat from dev/df
 - 1. Create data with different levels of *known* overdispersion
 - 2. Analyze each simulated dataset in MARK and record true c and c -hat from dev/df
 - 3. Fit a logistic regression to data where true $c=x$ and c -hat $=y$.
 - 4. Estimate level of c as value for which 50% of simulated datasets had c -hat values above true c and 50% had values below true c . This is done by fitting the logistic regression and taking advantage of the fact that $-b_0/b_1$ yields the level of x (true c in this case) for which $\text{pr}(y=1) = 0.5$
 - 5. Once have that value, use it to adjust $\text{AICc} = \text{convert to QAICc}$ as presented in textbook and in 5.1 of C&W
 - 6. Read chapter 5 of C&W with special focus on sections 5.1, 5.7, 5.9, & 5.10.
- d. NOTE: when you apply a value of c -hat in MARK, variances are inflated by value c -hat & se 's are inflated by the square root of c -hat. This doesn't identify the source of lack of fit but does inflate variances & reduce overfitting and problems of highly precise wrong answers.
- 5. Can use QAICc to evaluate models when c -hat >1.0 . This is a way of incorporating estimated overdispersion into modeling and model selection.
- 6. Model averaging is an important tool to learn and use. Model averaging yields weighted average estimates across models: the weights used are the QAICc weights, which are based on ΔQAICc values.

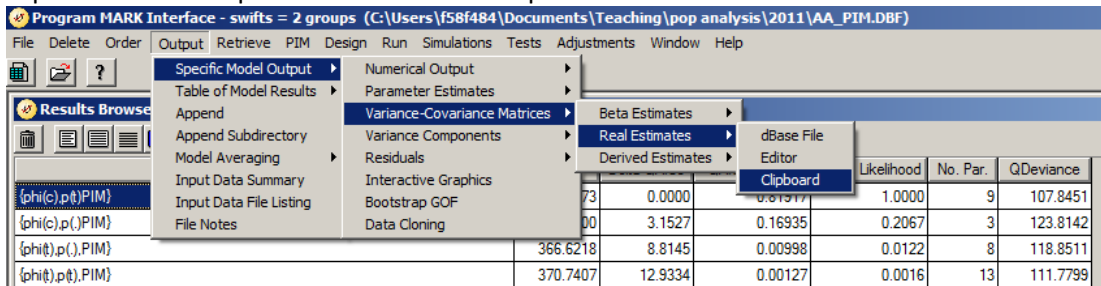
Lab Assignment

1. Run the 6 candidate models described in section 4.2.2 of Cooch & White (see the 2nd list of 6 models on page 4-19) by two methods:
 - a. Adjusting the PIMs as described. Run all 6 models and save a copy of the Results File as Lab04_PIM.dbf. You can easily do this with the *File menu* in the top left of the MARK window. Report the AICc table for the models using Lab04_PIM.dbf.
 - b. Using the Design Matrix to create the models when the PIMs are numbered 1-7 for φ_{Poor} , 8-14 for φ_{Good} , 15-21 for p_{Poor} , and 22-28 for p_{Good} . For example, the Design Matrix for the model $\varphi_c * p_t$ is on the left below, and the DM for the model $\varphi_i p_i$ is on the right. Save a copy of the Results file for the 6 models built using the Design Matrix as Lab04_DM.dbf. Report the AICc table for the models using Lab04_DM.dbf.

- c. Do the results differ or are they the same when you run models by changing PIMs versus using the Design Matrix? Do you see advantages or disadvantages to either approach?
 - d. When you run the model $\varphi_c + t\rho_t$, how many actual parameters can be estimated and what are they?
2. Highlight the most complex model in your dataset and run the median c-hat procedure on this model after carefully setting up your window like the one below: the key being to set the upper bound for c-hat to 1.25 (for this case where there's not much overdispersion, if you set it to something high like 5.0, you'll get lots of failures and the simulations will take a very long time to complete. Setting it to something low will provide us with more simulations in the proper range. **Warning:** if you make an error and try to stop MARK from completing the simulations, MARK will abort, which isn't a big deal but does mean you have to re-open the file, choose the model of interest, and try again.
- a. Once the c-hat procedure is complete, examine the graph that is put on your screen and report the estimated c-hat that appears on the graph title.
 - b. Also, report the estimates for the intercept and slope that are estimated from the regression that's done and report the value of $\frac{Intercept(Truth)}{Slope(Truth)}$, e.g., $10.090504/9.7453796=1.035414$. Each set of simulations will, of course, come up with somewhat different values for the intercept, slope, and c-hat.
 - c. Use the 'Adjustments' window to change the c-hat value from 1.0 to the value you just estimated. Report the QAICc table that results and comment on how the AICc table differs from the QAICc table. Why is there so little difference?
 - d. Provide one convincing example from your results that shows that variances were truly inflated by c-hat when you switched from AICc to QAICc, e.g., examine the estimates for one parameter (especially the estimated SE) with c-hat set to 1 and to the value estimated from the median c-hat procedure.
3. Conduct model-averaging on your QAICc table to provide your best set of estimates of apparent survival rate for the good sites and poor sites for occasions 1 through 6 (we'll ignore the estimates for the last occasion because some of the models can't estimate that one).



- a. What are the estimates by occasion and by group?
 - b. Briefly describe in words how these model-averaged estimates are achieved and which models are most heavily contributing to the estimates.
 - c. How do the estimated SE's from the model-averaged estimates compare to those from the top model? Why is this so?
4. Open "Lab04_PIM.dbf", adjust c-hat to 1.035414, & use MARK's output menu to view the estimated variance-covariance matrix for $\hat{\varphi}_{Poor}$ (the 1st group) and $\hat{\varphi}_{Good}$ (the 2nd group) based on results from the $\varphi_c p_t$ model as illustrated below. I find it easiest to output the variance-covariance matrix to the clipboard and then paste it into R or into a spreadsheet for further work with the estimates.



If you use the "Clipboard" option, you only get variance-covariance values, e.g.:

0.00616	0.00015	-0.00013	-0.00051	-0.00041	-0.00041	-0.00030	-0.00051	-0.00093
0.00015	0.00157	-0.00002	-0.00017	-0.00018	-0.00020	-0.00036	-0.00116	-0.00111
-0.00013	-0.00002	0.00758	0.00013	0.00004	0.00002	0.00001	0.00002	0.00003
-0.00051	-0.00017	0.00013	0.01090	0.00073	0.00025	0.00010	0.00017	0.00019
-0.00041	-0.00018	0.00004	0.00073	0.01339	0.00057	0.00017	0.00021	0.00019
-0.00041	-0.00020	0.00002	0.00025	0.00057	0.01115	0.00051	0.00035	0.00024
-0.00030	-0.00036	0.00001	0.00010	0.00017	0.00051	0.00798	0.00129	0.00054
-0.00051	-0.00116	0.00002	0.00017	0.00021	0.00035	0.00129	0.01142	0.00216
-0.00093	-0.00111	0.00003	0.00019	0.00019	0.00024	0.00054	0.00216	0.00932

And, the only values that we're interested in here are those highlighted in yellow, which we want to as many decimal places as are available for accuracy (I've rounded them above for ease of display).

If you choose "Editor" instead of "Clipboard," you'll see variance values on the diagonal of the output matrix, correlations above the diagonal, and covariances below the diagonal. This is all very useful but takes some diligence to understand, especially for matrices with rows that are so long that they wrap onto multiple lines on your screen.

You should obtain the following estimated variance-covariance matrix for φ_{Good} and φ_{Poor} and understand what those values mean.

	φ_{Poor}	φ_{Good}
φ_{Poor}	0.0061633030	0.0001537251
φ_{Good}	0.0001537251	0.0015690689

- a. Calculate the estimate of the difference between φ_{Good} and φ_{Poor} using the following formulas (show your work):

- i. Estimated difference between $\hat{\varphi}_{Good}$ and $\hat{\varphi}_{Poor}$ is simply $\hat{\varphi}_{Good} - \hat{\varphi}_{Poor}$
- ii. $Var(\hat{\theta}_i - \hat{\theta}_j) = Var(\hat{\theta}_i) + Var(\hat{\theta}_j) - 2Cov(\hat{\theta}_i, \hat{\theta}_j)$ where $\hat{\theta}_i = \hat{\varphi}_{Good}$ & $\hat{\theta}_j = \hat{\varphi}_{Poor}$.

NOTE: this is using the delta method to estimate the variance of the difference between two estimated quantities & incorporates variances and covariances.

- iii. $SE(\hat{\theta}_i - \hat{\theta}_j) = \sqrt{Var(\hat{\theta}_i - \hat{\theta}_j)}$
- iv. 95% CI on $(\hat{\theta}_i - \hat{\theta}_j) = (\hat{\theta}_i - \hat{\theta}_j) \pm 1.96 \cdot SE(\hat{\theta}_i - \hat{\theta}_j)$

Show your work and report the difference and 95% CI for the difference.
You might find pages 6-66 through 6-70 of C&W useful as you work on this.

- b. Comment on the estimate you obtain and the possible biological significance of the result.