

6. Mark-recapture analysis

Using CAPTURE and MARK to calculate multi-sample estimates of abundance with closed models

This practical uses the same dataset on humpback whales collected in 1988-1993 off West Greenland that you used for the first practical.

1. Input file for MARK

Program MARK requires data to be input in a particular format in a text file with a specific filename extension (.inp). To create the appropriate file, follow these steps:

(a) Create a text file containing the necessary data from the Excel file. To do this:

- Open the **Wghumpback** Excel file. Delete the header row and the first column with the animal identification numbers and row labels, leaving only the 1s and 0s of the capture histories.
- Add a final column containing “1;” in every row.
- Use **File > Save As** to save this file as **Text (Tab delimited)** with the filename **Wghumpback.inp**. Close Excel.

(b) Open **Wghumpback.inp** (in Notepad or similar text editor). The first few rows should look like this:

0	1	0	0	0	0	1;
0	0	1	0	1	0	1;
1	0	0	0	0	0	1;
0	0	0	1	0	0	1;

(c) The spaces between the number are the “tabs”. Use **Edit > Replace** to open the **Replace** window and delete all the “tabs”. Copy a “tab” from the data file into the **Find what** box, leave the **Replace with** box blank, and select **Replace All**. The first few rows should now look like this:

0100001;
0010101;
1000001;
0001001;

(d) Use **Edit > Replace** again to insert a space before “1;” in every row. Enter “1;” in the **Find what** box and enter “ 1;” (a space then 1;) in the **Replace with** box and select **Replace All**. The data should now be in the correct format for input to MARK:

010000 1;
001010 1;
100000 1;
000100 1;

The **1;** at the end of each line tells MARK that there is one individual with the capture history listed in each row; different whales with the same capture history are listed separately. You can also create input files where each row is a unique capture history and the number at the end is the number of animals with this capture history. Save and exit.

2. Entering data into MARK

Open program MARK.

Select **File > New** on the toolbar to show the **Enter Specifications for MARK Analysis** window.

From the list on the left side of the Window, choose **Closed Captures**.

The **Closed Captures Data Type** window appears. Select **Full Likelihood p and c** and then **OK** to close the window.

Type in a title for the analysis, e.g.: West Greenland humpback.

For **Encounter Histories File Name**, click on **Click to Select File**.

Find the **WGhumpback.inp** input file, select it and click **Open**.

Click on **View File** to check that you have the correct data file. Close the input file.

For **Encounter occasions**, change the default **5** to **6** (or however many sampling occasions you have in your data).

Ignore the part about **Attribute groups** and click **OK**.

MARK tells you that it has created a file to hold your results. Close this window.

You will now see a Window called **Capture Probability (p) Group 1 of Full Likelihood p and c**. Close this window.

NOTE: MARK uses a database program to store data and results in two files, a .DBF file and an .FPT file, each with the same name as the name of your input data file. To store your work from a particular analysis, keep these two files together. MARK saves everything you do automatically; you do not have to save your work yourself.

3. Running program CAPTURE

Select **Browse** on the toolbar to bring up the **Results Browser: Full Likelihood p and c** window. Ignore this for now; instead choose **Tests > Program CAPTURE** from the toolbar to bring up the **Program CAPTURE Models** window.

The models listed allow for different types of variation in capture probabilities. The Null model M(o) assumes a constant capture probability for all individuals across all sampling occasions. Model M(t) allows capture probability to vary by time (sampling occasion). Model M(b) allows for behavioural variability in capture probability, in particular to account for “trap happiness” or “trap shyness”. Model M(h) allows for individual variability (heterogeneity) in capture probability. Other models allow for combinations of these.

Because we don’t know which model best describes our data, check **Appropriate** to instruct the program to run tests to inform us which model is likely to fit the data best. Click **OK** to run the program.

A text file containing results will open in Notepad. Things to note include:

(a) A table summarising your data that looks like this:

Occasion	j=	1	2	3	4	5	6	
Animals caught	n(j)=	75	132	109	42	138	101	
Total caught	M(j)=	0	75	181	236	257	314	343
Newly caught	u(j)=	75	106	55	21	57	29	
Frequencies	f(j)=	176	100	52	11	3	1	

Question 1: Is the estimate of population size similar to those obtained using estimates for pairs of years in the first Practical? If not, what could be causing the difference?

If you want to get results for a different model with a smaller Criteria value, you can go back and run CAPTURE again but tick the boxes for whichever models you want to investigate rather than **Appropriate**.

Do this for models M(o), M(t) and model M(th) to see how the results differ. Write down the population estimates for these models. Remember, CV = SE divided by the Estimate.

Model M(o): Estimate = ____; SE = ____; CV = ____; 95% conf interval = ____

Model M(t) [Darroch]: Estimate = ____; SE = ____; CV = ____; 95% conf interval = ____

Model M(th) [Chao]: Estimate = ____; SE = ____; CV = ____; 95% conf interval = ____

Question 2: Is there much difference in the estimates from the different models? Why might you expect the estimate from model M(th) to be higher than the others?

Close the CAPTURE results file(s).

4. Running closed population models in MARK

On the toolbar, select **PIM > Open Parameter Index Matrix**. In the **Select Parameter Index Matrix to Open** window, click **Select All** and then **OK**.

Three windows appear containing the Parameter Index Matrices (PIMs) for:

- **Capture Probability (p) Group 1 of Full Likelihood p and c**
- **Recapture Probability (c)** etc
- **Never encountered (f0)** etc


These windows define the parameters that are in the currently defined model.

Select the **Capture Probability (p)** window. There are six small boxes labelled 1, 2, 3, 4, 5, 6. These represent the six capture probability parameters (one for each sampling occasion) in the model to be estimated from the data. The number is simply the number given to the parameter in the model.


Now select the **Recapture Probability (c)** window, where there are five small boxes labelled 7, 8, 9, 10, 11. These represent the five recapture probability parameters for each of the last five sampling occasions (obviously there can be no recapture probability on the first occasion).

Having separate parameters for recapture probabilities allows the model to account for any behavioural response in which recapture probability (after first capture) may be different from initial capture. This model with separate capture and recapture probabilities for each sampling occasion is equivalent to Model M(tb) in CAPTURE.

However, for photo-identification studies, we can usually assume that recapture probabilities are the same as capture probabilities, and we can modify the model to take account of this and make it simpler.

Still in the **Recapture Probability (c)** window, click on the  button to decrement the numbers in the small boxes from 7, 8, ... 11 down to 2, 3, ... 6, respectively. This gives the recapture probability parameters the same number as the capture probability parameters for the last five sampling occasions; that is, there is a single parameter representing capture (**p**) and recapture (**c**) probability in each sampling occasion.

This is known as the **c=p** option in MARK. This model that has capture and recapture probabilities the same for each sampling occasion, but varying by sampling occasion is equivalent to Model M(t) in CAPTURE.

Now select the **Never encountered (f0)** window, which contains one small box labelled 12. This represents a single model parameter for the number of animals never captured. Click on the  button again to decrement the number in the small window from 12 down to 7 (because there are now only 6 other parameters in the model).

Close the three **PIM** windows.

On the Toolbar select **PIM > Parameter Index Chart** to bring up a window showing the parameters pictorially. The type of parameter is shown vertically and the parameter number in the model is shown horizontally.

You should see a blue bar for **p** parameter number 1, bars representing **c** and **p** for parameter numbers 2 to 6, and one **f0** parameter (number 7). It is a good idea always to look at this chart to check that the model you are about to run is indeed what you think it is. This is now your “current model”.

Close the **Parameter Index Chart** window.

Now run your model using **Run > Current Model** from the toolbar to show the **Setup Numerical Estimation Run** window. Give the model a name (e.g. Closed captures $c(t)=p(t)$). The (t) is the convention for a parameter that varies over time.

Under **Link function**, select **Logit**. Accept the default for **Variance Estimation**. Click **OK to Run**.

Select **Yes** to accept the identity design matrix in the **Info** window. Select **Yes** in the **{model name} Results** window to append the model output to the database, i.e. to save your results.

Your model should now appear in the **Results Browser: Full Likelihood p and c** window. Note the AICc (a measure of model fit – the lower the better), the number of estimated parameters, and the residual deviance (a measure of how much variability in the data remains unexplained by the model). None of this is informative except when comparing different models, so ignore it for now.

Right click on the model name and select **Real Estimates** to open a table of results in Notepad. An alternative way to do this is to click on the fourth button from the left in the toolbar in **Results Browser**. The Table shows the estimates of capture/recapture probabilities for the six sampling occasions (1:p, 2:p, ... 6:p), and animals never captured (7:f0), together with standard errors and 95% confidence intervals.

Record these values:

1:p = _____; SE = _____

2:p = _____; SE = _____

3:p = _____; SE = _____

4:p = _____; SE = _____

5:p = _____; SE = _____

6:p = _____; SE = _____

7:f0 = _____; SE = _____

Question 3: Is there much variability among the six years in the estimates of capture probability? What could cause such variability?

Right click on the model name again and select **Derived Estimates** to see the estimate of population size (**N-hat**), which is calculated from the sum of the total number of captured animals (from the input data) and the estimate of the number of animals never captured (**f0**). Record the results.

N-hat = _____; SE = _____; CV = _____; 95% confidence interval = _____ to _____

Question 4: Note that the Standard Error for **N-hat** is the same as for **f0**. Why is this?

Question 5: How does this estimate of population size compare to those you obtained from CAPTURE?

5. Investigating model improvements

The model you have just run is the equivalent of model $M(t)$ in CAPTURE. Is it the best model? We can investigate other models to find out.

5.1 The simplest possible model

It is often easier to set up or modify a model using the **Parameter Index Chart** (blue bars).

Open the **Parameter Index Chart** again. Right click somewhere on the bottom bar (**p**) and choose **Constant**. Do the same for the second bar up (**c**). Drag the second (**c**) bar to the left so it sits on top of the first (**p**) bar. You now have a single parameter (number 1) for capture/recapture probability which applies to all sampling occasions; that is, it doesn't vary by time. Drag the top (**f0**) bar to the left to close up the gap but keep it as a separate parameter (number 2).

These are some of the quick options for modifying your model but sometimes you may still need to use the **Parameter Index Matrix**.

Close the **Parameter Index Chart** and run the current model. Give it a name, e.g. Closed captures $c(.)=p(.)$. The $(.)$ is the convention for a parameter that is constant. Use the same options for running the model as before. The results will appear as a new row in the **Results Browser** window.

The model with the lowest AICc will appear at the top. The Delta AICc for any model is the difference between the AICc for that model and the AICc for the model with the lowest AICc. The AICc Weight is a reflection of how much support a model has from the data (these weights always sum to 1 over all models shown).

The convention is that if the Delta AICc for a model is greater than 2 units, then the model with the lower AICc has more support from the data and should usually be considered the best model. If the Delta AICc is less than 2 units, then both models have more or less equivalent support from the data.

Question 6: Is model $c(.)=p(.)$ better than the previous model $c(t)=p(t)$? Why?

Look at your results for the $c(.)=p(.)$ model and record them.

1:p = _____; SE = _____

2:f0 = _____; SE = _____

N-hat = _____; SE = _____; CV = _____; 95% confidence interval = _____ to _____

Question 7: How does this estimate of population size compare to those you obtained from CAPTURE?

5.2 Investigating heterogeneity

To investigate heterogeneity in capture probabilities, we need to change model type.

Select **PIM > Change Data Type** from the toolbar and in the **Select new data type** window select **Full likelihood heterogeneity pi, p and c**. In the **Number of mixtures** window choose **2**.

This sets up a model in which it is assumed that the population comprises a mixture of two types of animals. Each type of animal has a constant (but different) capture probability. This is known as a “mixture model” and is a different way of modelling heterogeneity from that used in CAPTURE.

Open the **Parameter Index Chart**. You will see a parameter **pi**, which is the probability of an animal being of the first type. There are now double the number of **p** and **c** parameters, one set for each of the two animal types. To set $c=p$, as before, from the **Parameter Index Chart** select **Initial > Make c=p**.

You'll now see the same pattern for the **p** and **c** parameters as for model $c(t)=p(t)$ but with parameter numbers 2-7 and 8-13 representing the two types, respectively. Close up the gap between the **f0** parameter and these parameters.

Close the **Parameter Index Chart**. Run the model giving it a suitable name, e.g. $c(t)=p(t)$ 2 types. Look at the model diagnostics in the **Results Browser**.

Question 8: Is this model an improvement on the previous models?

Now look at and record the results:

1:pi = _____; SE = _____

2:p = _____; SE = _____ 8:p = _____; SE = _____

3:p = _____; SE = _____ 9:p = _____; SE = _____

4:p = _____; SE = _____ 10:p = _____; SE = _____

5:p = _____; SE = _____ 11:p = _____; SE = _____

6:p = _____; SE = _____ 12:p = _____; SE = _____

7:p = _____; SE = _____ 13:p = _____; SE = _____

14:f0 = _____; SE = _____

N-hat = _____; SE = _____; CV = _____; 95% confidence interval = _____ to _____

The value of the \mathbf{pi} parameter can be interpreted as the estimated proportion of the population in the first type; $(1-\mathbf{pi})$ is therefore the estimated proportion in the second type. The first six \mathbf{p} parameters (2-7) are for the six sampling occasions for the first type and the second six (8-13) are for the second type.

Question 9: What do you notice about the capture probabilities in the first type compared to the second type? What can you infer from this pattern and the estimate of \mathbf{pi} ?

Question 10: How does the estimate of population size compare with that from model $\mathbf{c(t)=p(t)}$ without heterogeneity? Is this what you expected?

Question 11: How does the estimate of population size compare with that from model $\mathbf{M(th)}$ from CAPTURE? All things considered, which would you choose as the best model?