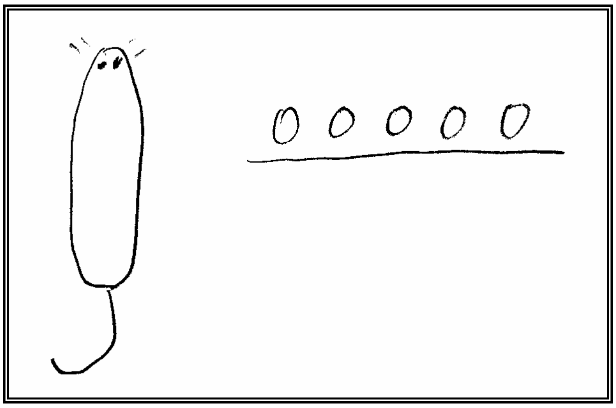
A simple mark-recapture experiment

For this experiment you'll need approximately 50 identical cards, such as business cards or index cards. On each card, draw a rough sketch of a rat (just an oval with a 'tail' will do) and add a row of five zeros and underline them, as shown in the diagram below.



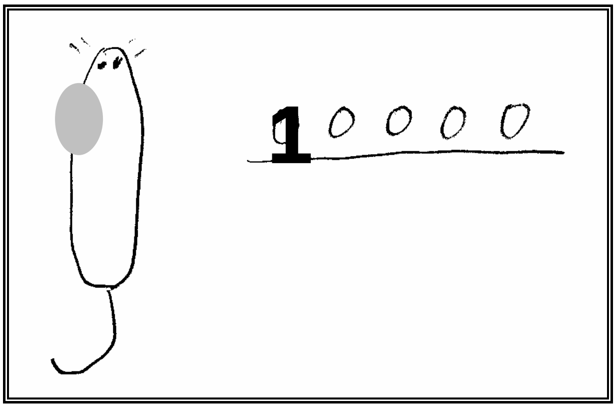
Print out the data sheet on the last page of this guide.

A. Two trapping sessions

First trapping occasion: Deal the cards into, say, 20 small stacks on the table. "Capture" the card at the top of each stack.

If these were real rats you were trapping, you could mark them by clipping the fur; in this case, shade the left shoulder of the sketched rat.

Also change the first zero to "1", as shown below.

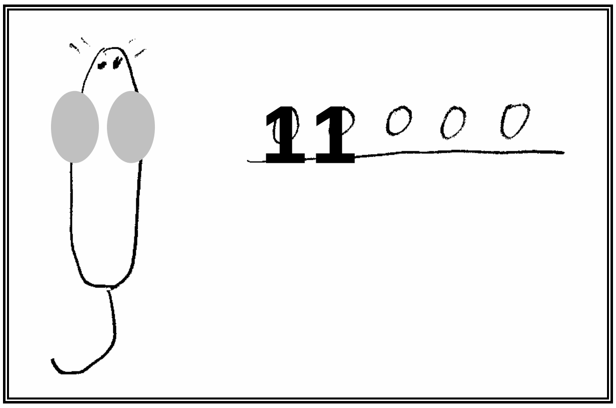


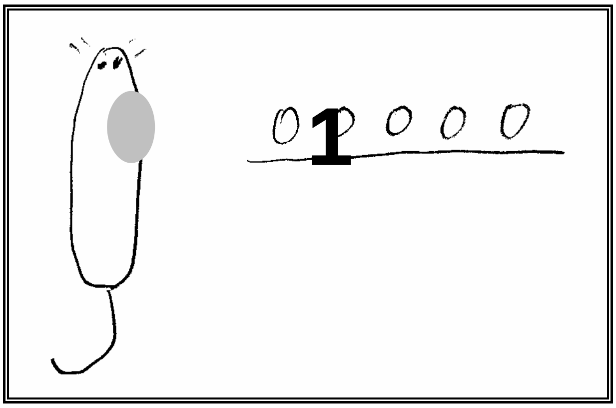
Second trapping occasion: "Release" the captured cards to the stacks where they were caught. Gather up the stacks, shuffle the cards well, and deal again into small stacks. The number of animals you capture this time may not be the same as the number you captured the first time, so now use, say, 16 stacks. "Capture" the card at the top of each stack.

Mark the captured rats with shading on the **right** shoulder and change the **second** zero to "1".

Now some of the rats are marked on the right shoulder only (coded "01000"), and some on both (coded "11000"): see the diagram below. There are also some in the pack which are marked on the left shoulder only (and are coded "10000").

Count the number of cards that are "recaptures". Summarize your results in the table below.





First capture	Number of rats captured and marked:	20
Second capture	Total rats captured:	16
	Number of marked rats 'recaptured':	

Now we can estimate the total number of cards. We take the proportion of recaptures in the second sample as an estimate of the proportion of marked cards in the whole pack. Since we know how many cards we marked, we can calculate how many are in the pack.

When I did this, I marked 20 cards the first time; the second time, I captured 7 marked cards, so my estimate of capture probability is 7/20 = 0.35, i.e. I've captured about 1/3 of the marked individuals. If the capture probability is the same for marked and unmarked cards, then the 16 cards in the second sample represent about 1/3 of the total number of cards, so the total is about 16 x 3 = 48. More exactly:

$$\hat{N} = 16 \times 20 / 7 = 45.7$$

Estimate the number of cards in the pack, based on your samples:

$\hat{N} =$

Count the cards and compare the true number, N, with your estimate.

This is the Petersen method, and it depends on a couple of assumptions (which obviously apply to the experiment with the cards, but might not apply when trapping real rats):

- Marked animals do not die or move out of the population between the first and second samples, and they do not lose their marks; if we mark 20 animals on the first trapping occasion, there are still 20 marked animals in the population during the second occasion.
- Capture probability is the same for marked and unmarked animals. In practice, animals which have been trapped once (and marked) often become 'trap-shy' and avoid traps in future. Occasionally – if traps are baited – animals become 'trap-happy' and more likely to be captured in future.

B. Multiple trapping sessions

You could repeat step 2 and compare the different estimates you get. But if we are going to have more than two trapping sessions, there are better ways of combining the results. First, though, we need to do a '**capture history**' for the first two sessions and begin filling in a data sheet.

We use '1' to indicate capture and '0' for non-capture, so '1 1' means captured twice, '1 0' means captured the first time but not the second, and '0 1' means captured the second time but not the first. My second trapping session resulted in recapturing 7 out of the 20 cards captured the first time, so:

1 1 = 7
1 0 = 20 - 7 = 13

There were also 9 new captures:

0 1 = 9

The first two columns of my data sheet look like this:

First Session	Second session
1 = 20	11 = 7
	10 = 20 – 7 = 13
0	01 = 9
	00

Fill in the first 2 columns of the data sheet with your results.

Third trapping session: Make sure the cards you captured in session 2 are properly marked, then “release” the captured cards, gather up the stacks and shuffle, then deal again – I used 22 small stacks this time.

Mark the captured rats with shading on the **left thigh** and change the **third** zero to “1”.

Sort the captured rats according to their capture histories: ‘1 1 1’, ‘1 0 1’, ‘0 1 1’ and ‘0 0 1’ and enter the numbers in the third column of the data sheet. Calculate the numbers of marked rats which you didn’t catch, ie. those with capture histories ‘1 0 0’, ‘1 1 0’ and ‘0 1 0’, and complete the third column.

Fourth trapping occasion: Repeat the procedure, this time marking the **right thigh** and changing the **fourth** zero to “1”. Fill in the fourth column of the data sheet.

Fifth trapping occasion: Repeat the procedure: there’s no need to mark the rats this time, but do change the **fifth** zero to “1”. Fill in the last column of the data sheet.

In this experiment, you can check the capture histories by spreading out all the cards, but that’s not possible with a real trapping study, so you will have to keep careful records of the markings on captured animals.

C. Data analysis

How do we analyze these results? We need to estimate two parameters, population size and capture probability, which are interrelated. To do this efficiently we need computer software, such as MARK or CAPTURE, which we will get to know in a moment.

First we need to put the capture histories into a computer file.

Start a text editor such as Notepad, and enter the final capture histories for your card experiment; this is the right-hand column on your data sheet. On each line put one capture history with no spaces between the ‘0’s and ‘1’s, then a space, then the number of cards with that capture history. My results looked like this:

```
/* My card experiment */
11101 3;
11100 2;
11001 1;
11000 1;
10111 1;
10101 3;
10100 4;
10011 1;
10001 2;
10000 2;
01100 2;
01010 1;
01001 3;
01000 5;
00111 2;
00110 1;
00101 1;
```

```
00100 3;
00011 1;
00010 1;
00001 2;
```

Save the file, giving it a name ending with '.inp' such as 'cards.inp'.

D. Getting started with MARK

The best place to download the latest version of MARK is www.phidot.org/software/mark/download/. As with any .exe file, it's wise to download it, run a virus check with an up-to-date virus scanner, and create a Windows Restore Point (go to 'Start > Programs > Accessories > System Tools > System Restore') before running the setup program.

If you use the results from MARK in any report or paper, please cite it as: White, G.C. and K. P. Burnham. 1999. Program MARK: Survival estimation from populations of marked animals. Bird Study 46 Supplement, 120-138.

E. Setting up the project and entering data in MARK

Start MARK and select 'File | New' from the pull-down menus. In the "Enter Specifications for MARK Analysis" window, under "Select Data Type", click on "Closed Captures". A new "Model Selection..." window appears.

We use a 'closed captures' analysis because we know that the same cards stayed in the pack throughout the experiment. The Closed Captures option has 12 possible sub-options, but only 3 are relevant to us.

The simplest possible situation is when capture is equally probable for all animals on all trapping occasions: this is often called the "M0" model. We can design more complicated models:

Mb : the animal's **behavior** changes after being trapped once, so that capture and recapture have different probabilities.

Mt : capture probability is not the same on each occasion, it varies with **time**.

Mh : **heterogeneity** - capture probabilities are not the same for all animals.

And it's possible to combine two or three effects, but then it gets complicated! To keep it simple, MARK offers three possibilities:

Closed captures (the "plain vanilla" option) – for M0, Mb, Mt and Mbt (no heterogeneity)

Closed captures with heterogeneity – for Mh (no behavior or time effect)

Full closed captures with heterogeneity – for all models, including Mtbh

There are corresponding "Huggins" versions, which use a rather different way of calculating the total population, and also six "misidentification" versions, developed for use when 'recapture' means collecting DNA, which might not identify to individual animals with certainty. It's possible to switch between models during the analysis, so we'll start with the basic one and change if we need to.

Highlight Closed captures and click on OK. In the 'Enter Specifications...' window, type a title for the set of data (eg "cards experiment"). Then click on 'Click to select file', browse to 'cards.inp' and click 'Open'.

MARK enters the name of the Results file as 'cards.dbf', but you can change that if you wish (e.g., if you are doing different sets of analyses with the same input file).

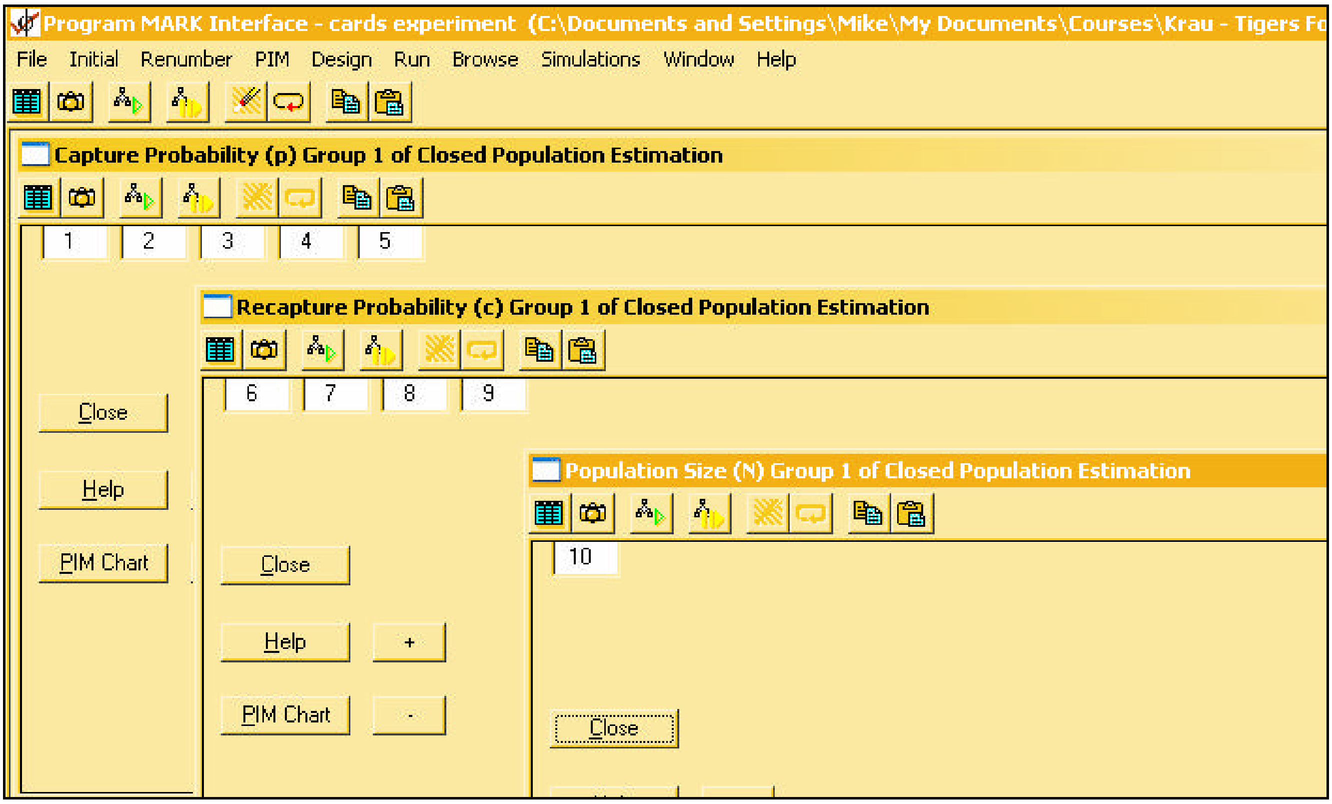
Check that 'Encounter occasions' is 5, and leave 'Attribute groups' as 1. Click on OK.

A dialogue box pops up to inform you that the output file 'CARDS.DBF' was created. Click OK.

F. Running a first analysis in MARK

When you've entered all the specifications, a PIM (Parameter Index Matrix) window opens. There are three of these, so let's open all of them before discussing the details.

From the pull-down menus, select 'PIM | Open Parameter Index Matrix'. In the box that appears click 'Select All' then 'OK'. Drag the three windows so that you can see the main items in all three (see screen shot).



The three windows show the parameter indices. The default uses the maximum number of parameters, so there are:

- 5 parameters for Capture Probability (p), numbered 1-5, one for each trapping occasion;
- 4 parameters for Recapture Probability (c), numbered 6-9 (only 4, because you can't recapture animals on the first trapping occasion);
- 1 parameter for Population Size (N).

In this case the model with so many parameters produces errors (try it if you like!), so let's begin with the minimum number of parameters, the M0 model, with all the capture and recapture probabilities equal.

Change all the Capture Probabilities to "1": you could type "1" in each box, but a quicker way is to click on any of the boxes and select 'Initial | Constant' from the pull-down menus. Then change all the Recapture Probabilities to "1" as well: click anywhere in the Recapture window and select 'Initial | Make c=p' from the menus.

This tells MARK to use the same Beta coefficient (β_1 in this case) for all the capture and recapture probabilities. We are not saying that the capture probability =1. It would make no difference to the result if you told MARK to use β_1 for the population size and β_2 for the capture and recapture probabilities. In the Parameter Index Matrix (PIM) we are only dealing with the indices (the little numbers identifying the β's), not with the values of the parameters.

The number in the Population Size box is still "10", but that doesn't matter, provided it is different from the other numbers.

Select 'Run | Current Model' from the pull-down menus.

In the 'Setup Numerical Estimation Run' window, give the model the name "M0", leave all the other default values as they are and click 'OK to Run'.

Click 'Yes' when MARK asks if it should use an identity design matrix, and 'Yes' again when it asks if it should add the output to the database.

The Results Browser appears, which will look familiar if you have used PRESENCE or DISTANCE before. Nothing interesting in the Browser yet, as we have only run one model, but you might note that No. Parameters is 2, as we intended; these are

- capture/recapture probability (*p* = *c*) and
- population size (N).

Let's look at the detailed output.

Right-click in the Results Browser and select 'Output in Notepad' (note that the results for the highlighted model appear, not the one you clicked on!) or click on the second button on the toolbar.

The first part summarizes the input, including the model name and setup details. Then there's some technical information and, near the bottom, the *real* stuff. Just like PRESENCE, MARK deals in probabilities which vary from 0 to 1, and these cannot be modeled directly by a linear function. Like PRESENCE, MARK uses a 'linear estimator' which is linked to the real values of *p* and *N*; in this case the default is the SIN link function (PRESENCE uses the logit link by default).

These are the results for my experiment:

Real Function Parameters of {M0}

Parameter	Estimate	Standard Error	95% Confidence Interval Lower	Upper
1:p	0.3871155	0.0386994	0.3145026	0.4651181
2:N	45.464471	2.5022502	42.973006	54.335552

The estimated capture (and recapture) probability is quite high at 0.4, and the estimated number of cards is 45.5 for my experiment.

Now let's see what happens with more complex (and maybe more realistic) models.

G. Comparing models

Open the three PIM windows again. If you can't see them, check the 'Windows' pull-down menu, and if they have been closed, use 'PIM | Open Parameter Index Matrices' to get them back. Arrange them so that you can see all the values.

Notice that MARK did some tidying up before running the previous model and renumbered the Population Size parameter to "2".

Mb : Let's run a model where animals behave differently after being trapped the first time, so recapture probability is different from capture probability. But neither changes with time.

In the Recapture Probability window, click on the "+" button: all the values will increase to "2". Do the same in the Population Size window, changing the "2" to "3". Run the analysis as before, naming the model "Mb".

You could set the index for Recapture Probability to "3" and leave Population Size as "2", but the output is easier to interpret if you always have Population Size values last. In any case, do make sure that the Population size index is not the same as the Recapture index (to avoid that mistake and the nonsense results that follow, it's best to open *all* the PIM windows).

Look at the output for this model. Compare the results for this model. In particular check to see if *c* and *p* are similar. Would you expect them to be different for this experiment?

Mt : Now try the model where capture and recapture probabilities are equal, but they vary with time.

Open the three PIM windows again. Click in the Capture Probability window and select 'Initial | Time' from the pull-down menu, then 'Initial | Make c=p'.

The parameters 1 to 5 will be used to model 5 different capture probabilities for the 5 trapping occasions, and the parameters 2 to 5 will also be used for the recapture probabilities – remember that there are no *re*captures for the first trapping occasion.

Change the parameter index for Population Size to 6 (or any number above 5). Run the model as before, naming it "Mt".

Check if the population estimate and its confidence limits look reasonable. The output contains 5 separate estimates for *p*: are these what you would expect?

Mh: So far we've assumed all animals have the same probability of being captured. MARK can't handle separate capture probabilities for each animal, but it can work with 2 or sometimes 3 groups with different probabilities and mix the results. We'll try mixing just two, but to do that we have to change the data type.

Select 'PIM | Change Data Type' from the pull-down menus. In the list of options that appears, highlight 'Closed Captures with Heterogeneity' and click on 'OK'. Leave the Number of Mixtures as "2" in the next box and click 'OK'.

This data type allows for a basic heterogeneity model equivalent to mixing two M0 models, so there is only one PIM window for "*p*", with two indices. The parameter *pi* (π) determines how much of each model goes into the mix. If we can divide the population into two groups of animals with different *p* values, *pi* indicates the proportion of the population in each group. There isn't much we can change in this model, so we'll use the defaults.

Run the model as before, naming it "Mh"

Look carefully at the output, as MARK may not have been able to separate the cards into two groups (which would not be surprising!) Indicators are:

- the estimate for *pi* is near 0 or 1; if so, the animals are all in the same group.
- the two estimates of capture probability are almost the same; in this case, MARK has identified two groups, but there is no difference between them.

Now look at the summary of the models in the Results Browser. The screen shot is for my results: yours will be different.

Results Browser: Closed Captures with Heterogeneity

Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance
{Mt}	46.4549	0.0000	0.75483	1.0000	6	25.8075
{Mb}	50.2942	3.8393	0.11070	0.1467	3	35.9442
{M0}	50.5162	4.0613	0.09907	0.1312	2	38.2247
{Mh}	52.5748	6.1199	0.03539	0.0469	3	38.2247

MARK uses AICc: the usual Akaike Information Criterion (AIC) is negatively biased (i.e., too small) when the sample size is small compared with the number of parameters, and AICc includes a small-sample correction term.

As with DISTANCE and PRESENCE, the models are ranked according to AICc, with the lowest AICc indicating the 'best' model in the sense that it is a good compromise between complexity (= number of parameters) and fit (= likelihood of getting this set of data if this model is correct).

H. Summarizing the results

1. Since you know how the experiment works, you should know which is in fact the correct model. Is this the best model based on the data (according to MARK)?

2 . Summarize the population estimates for the four models:

Model	Estimated population	Confidence interval		Capture probability	delta AICc
		lower	upper		
M0					
Mb				p = c =	
Mt				(5 values)	
Mh				p_1 = p_2 =	

3. Count the number of cards in the pack. Did the experiment give the right answers? Does the true answer fall within the confidence interval?

4. Compare the results with other people in the group. How often did the software come up with the right model? Were the estimates generally close to the right answer? How often was the right answer within the confidence interval?

I. Getting finished

MARK automatically saves all the results of analyses when you run them in files with the .dbf and .fpt extensions. You do not need to save results manually. (If you copy these files to a new location, be sure to copy *both* of them.)

To exit MARK, select 'File | Exit' or press Alt-F4 or click on the ☒ button at the top right of the window.

To re-open the project, use 'File | Open' and select the appropriate .dbf file. If you do this when a project is already open, MARK will automatically close it. You cannot have two projects open at the same time.

Further resources

The best resource if you want to learn more about MARK, the range of analyses it can do, and the math behind the software is Cooch and White (2006), which is available for free download. The recent book edited by Amstrup et al (2005) is a mine of information on current aspects of mark-recapture studies. Mark-recapture is a well-worn technique and there are several chapters on it in Krebs (1999) and Williams et al (2002).

Amstrup, S. C., T. L. McDonald, and B. F. J. Manly (eds). 2005. Handbook of capture-recapture analysis. Princeton University Press, Princeton, NJ.

Carothers, A. D. 1973. Capture-recapture methods applied to a population with known parameters. J Animal Ecology 42:125-146.

Cooch, E., and G. White 2005. Program MARK: a gentle introduction. Available in .pdf format for free download at http://www.phidot.org/software/mark/docs/book/

Karanth, K. U., and J. D. Nichols. 1998. Estimation of tiger densities in India using photographic captures and recaptures. Ecology 79:2852-2862.

Karanth, K. U., J. D. Nichols, N. S. Kumar, W. A. Link, and J. E. Hines. 2004. Tigers and their prey: Predicting carnivore densities from prey abundance. Proceedings of the National Academy of

Sciences 101:4854-4858 (Freely accessible at http://www.pnas.org/cgi/content/full/101/14/4854)

Krebs, C. J. 1999. Ecological Methodology. Addison Wesley Longman, Menlo Park CA

Seber, G. A. F. 1982. The estimation of animal abundance and related parameters. MacMillan, New York

Williams, B. K., J. D. Nichols, and M. J. Conroy 2002. Analysis and management of animal populations. Academic Press, San Diego CA

Capture history for card experiment				
First session	Second session	Third sessions	Fourth session	Fifth session
1	11	111	1111	11111
				11110
			1110	11101
				11100
		110	1101	11011
				11010
			1100	11001
				11000
	10	101	1011	10111
				10110
			1010	10101
				10100
		100	1001	10011
				10010
			1000	10001
				10000
0	01	011	0111	01111
				01110
			0110	01101
				01100
		010	0101	01011
				01010
			0100	01001
				01000
	00	001	0011	00111
				00110
			0010	00101
				00100
		000	0001	00011
				00010
			0000	00001
				00000

Do not put entries in the grey cells!