# 4. Mark-recapture analysis

## Using Excel to calculate 2-sample estimates of population size

### 1. The data

The practical uses a dataset on humpback whales collected in 1988-1993 off West Greenland.

Open the Excel file **WGhumpback**. The first column contains the unique number of each individually identified whale. The capture histories for each whale are in the next six columns, labelled by year. "1" indicates that the whale was seen in that year and "0" that it was not seen.

First, explore your data by generating a discovery curve and a frequency of capture histogram.

### 2. Discovery curve

To create a discovery curve, you need to generate data for the cumulative number of different individuals (whales) seen in successive sampling occasions (years). To do this, you need to determine how many individual whales were seen <u>for the first time</u> in each year. An easy way to do this is to use **Filter**, found under **Sort & Filter > Filter**.

***Excel tip:***

Highlight your data (columns A-G, including the header row and down to row 344) and turn on **Filter**; arrows appear at the top of the columns.

Click on the arrow in the **1988** column. Select "**1**". The number of rows selected should appear at the bottom left of the worksheet. Write down this number, click on the arrow again and select "**0**".

In the next column, **1989**, click on the arrow, select "**1**", and again write down the number of rows selected. Click on the arrow and select "**0**" again.

Continue this pattern until you have written down the numbers of whales seen for the first time in all sampling occasions. Turn off **Filter**.

Enter these numbers of animals seen for the first time in each year in column J (**new whales**). In column K (**cumulative**), calculate the <u>cumulative</u> number of whales identified up to and including each year. For example, for 1989, sum the number of new whales in column J for 1988 and 1989; for 1990, sum the number of new whales for 1988, 1989 and 1990; etc

Plot the cumulative number of new identifications against year - this is your discovery curve.

**<u>Question 1</u>:** Is the shape of the discovery curve what you expected to see? Have all the whales been seen or are there more to identify? Does the fact that the data were collected over several years influence your interpretation?

**Note:** Another way of showing a discovery curve is to plot the cumulative number of new identifications against the cumulative total number of identifications. This can show a smoother curve because the number of new identifications is relative to the sample size, which may vary from one sampling occasion to another. However, it loses information on timing of samples.

## 3. Frequency of capture histogram

In column H (**no. times seen**), enter the number of times each animal was seen (simply sum over the 1s and 0s of the capture history). Then use the **Histogram** tool, found under **Data > Data Analysis > Histogram**, to calculate frequencies of capture (how many animals were seen once, twice, etc) and plot the frequency histogram as a graph.

***Excel tips***:

When using the **Histogram** tool, you need to enter a range of cells for the **Bin Range** (the horizontal axis of your histogram). Highlight the numbers 1, 2, 3, 4, 5, 6 in column L (**frequency**) to copy into the **Bin Range**.

In the **Histogram** tool window, check **Chart Output** at the bottom to get a graph.

**Question 2**: What does the histogram tell you? Is it indicative of any problems?

## 4. Estimates of population size

Calculate two-sample estimates of population size for consecutive pairs of years (1988-1989, 1989-1990, etc).

To do this:

(a) In row 348, labelled $\mathbf{n_1}$, sum each column to determine the number of whales seen in each year.

(b) The next row is labelled $\mathbf{n_2}$. Set each cell in this row equal to the value for $\mathbf{n_1}$ from the next column (i.e. $\mathbf{n_2}$ for 1988 is the same as $\mathbf{n_1}$ for 1989, etc). This means you have $\mathbf{n_1}$ and $\mathbf{n_2}$ for each pair of years in the same column and allows you to do all the calculations in this single column. Note that you won't have an $\mathbf{n_2}$ for the final year 1993.

(c) Determine the number of whales seen in both years for each pair of years (1988 and 1989, 1989 and 1990, etc. The easiest way to do this is again to use **Filter**. Select your data and turn on **Filter**. Select "**1**" in each pair of columns in turn, making sure that all the other columns are set to "**Select all**" (the default). Write down these numbers from the bottom left corner and then enter them in row 350 labelled $\mathbf{m_2}$. You now have all the data needed to estimate population size.

(d) For the first pair of years (column B), use the formulae given at the end of this worksheet to calculate the Petersen and Chapman-modified estimates of population size from the cells containing $\mathbf{n_1}$, $\mathbf{n_2}$ and $\mathbf{m_2}$. Put these in the appropriate rows below. Then, for the Chapman-modified estimate, also calculate estimates of Variance, Standard Error, Coefficient of Variation, and 95% Confidence Limits using the equations given. When you enter the formulae, be sure to check your parentheses (brackets) are in the right place and matched up.

(e) When you've done this for one column, select and drag the cells containing all these calculations across to the columns C to F, to obtain the results for the other pairs of years.

Record your estimates below:

<u>1988-1989</u>

Petersen estimate = _____ ; Chapman estimate = _____ ; CV = ______

<u>1989-1990</u>

Petersen estimate = _____ ; Chapman estimate = _____ ; CV = ______

<u>1990-1991</u>

Petersen estimate = _____ ; Chapman estimate = _____ ; CV = ______

<u>1991-1992</u>

Petersen estimate = _____ ; Chapman estimate = _____ ; CV = ______

<u>1992-1993</u>

Petersen estimate = _____ ; Chapman estimate = _____ ; CV = ______

**<u>Question 3</u>**: Do your results indicate any small sample bias in the Petersen estimates?

Now plot the estimates and their confidence limits on a graph to illustrate how abundance varies across years.

**Excel tip:** In Excel, confidence limits are usually added to a graph as the <u>intervals</u> between the limits and the estimate. The upper interval is the <u>upper limit</u> minus the <u>estimate</u> and the lower interval is the <u>estimate</u> minus the <u>lower limit</u>. Calculate these in rows 363 and 364 before you create the graph.

**<u>Question 4</u>**: Are the estimates consistent across years? If not, what could be causing the variation?

| **Question 5**: How would you deal with the estimate for 1990-1991 when describing your results? |
| --- |
| |

## Formulae

Use the formulae below to calculate quantities in *Excel*.

**Note** that *Excel* format is different from the way the formulae are written on paper. Excel formats are given for most calculations you need. For each quantity in each formula, click on the cell that contains the value of that quantity. For example, $n_1$ is in cell B348 for 1988-89. Don't forget to balance parentheses correctly.

Petersen estimator

$$\hat{N} = \frac{n_1 n_2}{m_2}$$

*Excel* format: = (n1*n2)/m2

Chapman estimator

$$\hat{N} = \frac{(n_1+1)(n_2+1)}{(m_2+1)} - 1$$

*Excel* format: = (n1+1)*(n2+1)/(m2+1)-1

$$\text{var}_{\hat{N}} = \frac{(n_1+1)(n_2+1)(n_1-m_2)(n_2-m_2)}{(m_2+1)^2\,(m_2+2)}$$

*Excel* format: = (n1+1)*(n2+1)*(n1-m2)*(n2-m2)/((m2+1)^2*(m2+2))

$$SE_{\hat{N}} = \sqrt{\text{var}_{\hat{N}}}$$

*Excel* format: = sqrt(…

$$CV_{\hat{N}} = \frac{SE_{\hat{N}}}{\hat{N}}$$

$$c = e^{1.96\sqrt{\ln(1+CV_{\hat{N}}^{\,2})}}$$

*Excel* format: = exp(1.96*sqrt(ln(1+CVN^2)))

Lower 95% confidence limit = $\hat{N}/c$

Upper 95% confidence limit = $c\hat{N}$