1 Length data processing

1.1 Introduction

1.2 Input files

This subsection details the input data needed and the format of the files. There are three types of data required.

- Sample data
- Proportion at length data
- Catch data

1.2.1 Sample data

This data contains information on the numbers of fish sampled for each year, fishery, and sex. There are three files; males, females and combined. Each file contains a table of the form:

	Year	RV_4VWX	HS	LL_NAFO3_Obs	LL_NAFO4_Obs	OT_NAFO3_Obs	OT_NAFO4_Obs
1	1970	24	0	0	0	0	0
2	1971	37	0	0	0	0	0
3	1972	25	0	0	0	0	0
4	1973	36	0	0	0	0	0
5	1974	46	0	0	0	0	0
6	1975	62	0	0	0	0	0

Table 1: Example Sample data table

The files should be formatted as ".txt" files with a space seperating each column of the table, and a new line seperating each row of the table. An example is in the master folder under: "NumbersCombinedMeasured.txt".

The columns can be in any order but the following names must be used consistently:

- "Year"
- "RV_4VWX"
- "HS"
- "LL_NAFO3_Obs"
- "LL_NAFO4_Obs"
- "OT_NAFO3_Obs"

- "OT_NAFO4_Obs"

The name of the input file can also be changed, but the new name must be updated in the file "Filenames_control.r".

1.2.2 Proportion at length data

This data contains information of the proportion of fish sampled at each length class. More specifically, for each fishery it shows the proportion of fish sampled at each length class on each year. The total for any given year is 1. Once again, there are three files; males, females and combined. Each file contains a number of tables of the form:

	1971	1972	1973	..	2013
1	0.02058	0	0.0643	..	0
2	0	0.0461	0	..	0.01675
3	0	0.29965	0.03347	..	0.02519
4	0.05797	0.11525	0.1876	..	0.07879
5	0.03739	0.02305	0.06385	..	0.08773
6	0.24151	0.11406	0.17949	..	0.07983

Table 2: Example Sample data table

There is one such table for each fishery, and each table is labelled with a line at the top with a "#", a space, and the fishery name. (OneF of the following: RV_4VWX, HS, LL_NAFO3_Obs, LL_NAFO4_Obs, OT_NAFO3_Obs, OT_NAFO4_Obs).

Once again, columns in the table are seperated by spaces, and lines are seperated by new lines. An example of one such file may be:

```
# RV_4VWX
0.02058 0 0.0643 0 0 ..
0 0.0461 0 0 0 ..
0 0.29965 0.03347 0.08376 0.02001 ..
0.05797 0.11525 0.1876 0.03861 0.11247 ..
0.03739 0.02305 0.06385 0.09824 0.10712 ..
.. .. .. .. .. ..
# HS
-1 -1 -1 -1 -1 ..
-1 -1 -1 -1 -1 ..
-1 -1 -1 -1 -1 ..
-1 -1 -1 -1 -1 ..
-1 -1 -1 -1 -1 ..
-1 -1 -1 -1 -1 ..
.. .. .. .. .. ..
```

Once again, the order does not matter, as long as the fishery labels are consistent. The files can be given any name, as long as it is set in the file "Filenames_control.r".

1.3 Step 1: Weigh catch at length data by catch and sample size

We aggregate the proportion at length data for the commercial fisheries into 1, 2 or 4 fisheries, depending our needs.

- Case 1 - 4 fisheries
- Case 2 - 2 fisheries
- Case 3 - 1 fishery

Case 1 is quite simple, since there are already 4 fisheries. There is nothing to aggregate, so there is no need to calculate weights, and we can leave the proportion at length files as is.

Case 2 is a little more complicated. We want to aggregate the length proportion data from each of 4 commercial fisheries into 1. We also want to give a higher weight to fisheries with higher catch and fisheries with higher sampling effort. First some definitions. Let $p_{g,a,s,y,l}$ be:

$$p_{s,l,y,g,a} = \begin{array}{l}\text{The proportion of sex } s \text{ fish of length } l \text{ sampled for year } y, \\ \text{sampled for gear type } g \text{ and area } a.\end{array} \tag{1}$$

Then we sum over the gear type and area to obtain a new proportion at length $\tilde{p}_{s,l,y}$ representing all 4 fisheries, in the following manner:

$$\tilde{p}_{s,l,y} = \sum_{g,a} w_{s,y,g,a} p_{s,l,y,g,a} \tag{2}$$

where the weights $w_{s,y,g,a}$ are calculated as:

$$w_{s,y,g,a} = \frac{C_{y,g,a} S_{s,y,g,a}}{\sum_{g,a} C_{y,g,a} S_{s,y,g,a}} \tag{3}$$

where:

$$C_{y,g,a} = \text{Total catch (tonnes) in year } y, \text{ gear type } g \text{ and area } a \tag{4}$$

$$S_{s,y,g,a} = \text{Number of fish of sex } s \text{ sampled in year } y, \text{ gear type } g \text{ and area } a \tag{5}$$

Note that the weights sum to 1, i.e.

$$\sum_{g,a} w_{s,y,g,a} = 1 \tag{6}$$

Case 3 is even more complicated. This time we want to aggregate the length proportion data from each of 4 commercial fisheries into 2 fisheries, based on gear type. Again we also want to give a higher weight to fisheries with higher catch and fisheries with higher sampling effort. We calculate $\bar{p}_{s,y,g}$ as:

$$\bar{p}_{s,l,y,g} = \sum_{a} w_{s,y,a} p_{s,l,y,g,a} \tag{7}$$

where the weights $w_{s,y,a}$ are calculated as:

$$w_{s,y,a} = \frac{C_{y,g,a} S_{s,y,g,a}}{\sum_a C_{y,g,a} S_{s,y,g,a}} \tag{8}$$