

Assigning Individual Ages with an Age-Length Key

Derek H. Ogle

April 25, 2009

The assessment of ages for a large number of fish is very time-consuming. On the other hand, measuring the length of a large number of fish is relatively easy. The age structure for a large number of fish can be estimated by summarizing the relationship between age and length for a relatively small subsample of fish and then applying this summary to the entire group of fish. This summary is called an *age-length key*. The construction and application of an age-length key and the use of this key to assign ages to individual fish as described in [Isermann and Knight \(2005\)](#) is the focus of this vignette.

This vignette requires functions in the `FSA` and `NCStats` package maintained by the author. These packages are loaded into R with,

```
> library(FSA)
> library(NCStats)
```

All analyses in this document use the Lake Ontario rock bass (*Ambloplites rupestris*) data set from [Wolfert \(1980\)](#) which can be read into R with,

```
> data(RockBassL02)
> view(RockBassL02)
```

```
   age  t1
572  NA 186
600  NA 167
605  NA 193
640  NA 164
651  NA 204
1088 NA 203
```

Section 1 contains background information for age-length keys in general and the use of the methods described in [Isermann and Knight \(2005\)](#) for assigning ages to un-aged fish. **Section 2** then shows how to construct an age-length key and **Section 3** then shows how to assign ages to individual fish. **Section 4** contains additional thoughts on age-length key methods.

1 Background

1.1 Constructing Age-Length Keys

A small subsample of n fish is selected from the entire sample to be aged by randomly selecting fish *from each length interval* (rather than a simple random selection of all fish). The number of fish in each interval can either be fixed or proportional to the total number of fish in that length interval, although, [Kimura \(1977\)](#) notes that proportional selection produces statistically “better” results. More importantly, though, the range of lengths in the subsample must cover the range of lengths in the original sample. This subsample of fish is called the *age sample* as the relationship between age and length will be determined from this group of fish only.

The measured length and assessed age is recorded for all fish in the age sample. Each fish is also recorded to belong to one length interval category. For example, if 5-mm length categories are created that begin on the “0” and “5” units then a 117 mm fish will be categorized in the 115-119 mm length category. Generally all length categories are of the same width, so, for simplicity, these categorizations are usually only noted by the first length in the interval (e.g., “115” mm).

The assigned length interval category and the assessed age for fish in the age sample is summarized in a two-way contingency table with length categories as the rows and age categories as the columns. In other words, the number of fish in the age sample in each length category and age combination is determined. An example of this summary is shown below (with row and column labels to aid interpretation).

	A_1	A_2
L_1	6	2
L_2	3	3
L_3	1	4

Recall that each row corresponds to a length category and each column to an age. Thus, in this example, six fish in the age sample belong to the first length category and the first age. Similarly, one fish belongs to the last length category and the first age.

The portion of the large sample of fish that was not sampled to determine age and thus, for which only length measurements were obtained, is called the *length sample*. Age is “assigned” to fish in the length sample based on the length category of the fish and the proportion of fish of that length category of each age as determined by the fish in the age sample. For example, a fish in the length sample that belongs to the first length category has, based on the summary from the age sample, a 75% chance (i.e., $\frac{6}{8}$) of being of the first age and a 25% chance of being of the second age. Furthermore, if there were 20 fish in the length sample in the first length category then we would expect 15 (i.e., $20(0.75)$) to be of the first age and five of the second age.

Thus, assignment of ages to fish in the length sample is based on the “probability” of each age given the length category that the fish belongs as derived from the age-sample. The required conditional probabilities from the age sample are derived from the summary contingency table by dividing each cell of the table by the total for the corresponding row (thus, the contingency table becomes a “row-proportions” table). The row-proportions table derived from the example summary contingency table above is shown below.

	A_1	A_2
L_1	0.75	0.25
L_2	0.50	0.50
L_3	0.20	0.80

This table is the actual age-length key as it relates the conditional probability of an age given a particular length category.

1.2 Age Assignment to Individuals

An age-length key can be used to develop a summary frequency of ages for the length sample. In addition, methods have been developed to find the mean length-at-age for the length sample or the CPE for each age from fish in the length sample (Bettoli and Miranda 2001). However, these methods cannot be used to calculate measures of variability for these summaries (Isermann and Knight 2005). Thus, it is beneficial to use the age-length key to assign an age to each fish in the length sample and then to summarize those results (Isermann and Knight 2005). There are two methods for using the age-length key to assign ages to specific fish in the length sample: semi-random and completely random methods. The semi-random method is discussed below and the completely random method is shown in **Section 4.1**.

In the semi-random method of assigning ages to individual fish, the exact expected number of fish with a given length interval of a given age will be assigned that age, with exceptions for fractionality discussed

below. For example, if there are 22 fish in the length interval and, from the age-length key, 50% are expected to be of age-3 and 50% are expected to be of age-4 then exactly 11 fish will be assigned age-3 and 11 fish will be assigned age-4. Which fish are assigned to specific ages is determined randomly, but the number assigned each age is set by the expected values.

One difficulty with this method is that the expected number of individuals in a given length interval assigned a certain age may contain a fraction. For example, with the same expected percentages of age-3 and age-4 fish as in the previous example, one would expect 11.5 of 23 fish in the length interval to be assigned to each age. Clearly, a half of a fish cannot be assigned to a given age category. This difficulty was called *fractionality* by [Isermann and Knight \(2005\)](#).

Fractionality may be handled by first assigning the closest integer smaller than the expected number of fish to the age-group. In the example above, 11 fish would be assigned to age-3 and 11 fish would be assigned to age-4. The remaining fish is then randomly assigned an age with a probabilistic weight equal to the expected proportion of fish in each age category. For example, the fish would be assigned an age of three with a probability of 50% and would be assigned an age of four with a probability of 50%. In other words, in this example, it is essentially a coin-flip whether the “extra” fish is assigned an age of three or an age of four. Once the number of fish within a length interval to be assigned each age is determined then the specific ages are determined randomly; but again, the number assigned each age is set.

2 Construction of Age-Length Key

In many situations one data file will contain the lengths of all fish in the sample with corresponding ages only for those fish in the age sample. The ages for the fish not in the age sample will have `NA` in their place which is how R denotes “missing data” (i.e., `NA` stands for “not available”). With this type of data file, the first step in constructing and applying the age-length key will be to separate the data file into the age sample of fish with assigned ages and the length sample of fish without assigned ages. This separation requires `Subset()` and `is.na()`. The `Subset()` function, which requires two arguments, creates a new data frame from an existing data frame based on some condition. The first argument is the original data frame from which a subset or part of the data frame should be returned. The second argument is a conditional statement on how that subset should be determined. The `is.na` function is used to identify positions in a vector where “NA”s occur. This function is used to create the condition for separating the original data frame into age and length samples. The age and length samples are constructed for the Lake Ontario rock bass data with,

```
> rb.age <- Subset(RockBassL02, !is.na(age))
> str(rb.age)

'data.frame':      135 obs. of  2 variables:
 $ age: int  6 5 7 9 9 7 8 4 6 8 ...
 $ tl : int  218 184 211 223 245 181 207 173 201 246 ...

> rb.len <- Subset(RockBassL02, is.na(age))
> str(rb.len)

'data.frame':      1153 obs. of  2 variables:
 $ age: int  NA NA NA NA NA NA NA NA NA NA ...
 $ tl : int  172 173 175 171 173 184 203 218 222 185 ...
```

The next step in constructing the age-length key is to create a variable that identifies the length interval category for each fish in the age sample. This variable is constructed, with default name `LCat`, and appended to the data frame containing the age-sample with `lencat()`. In this context, `lencat()` requires four arguments,

- `d`: the data frame containing the age-sample,

- `cl` : a number or string indicating which column of the age sample data frame contains the measured length data,
- `startcat` : a value identifying the starting length measurement category, and
- `w` : a value identifying the width of the length measurement categories.

The `lencat()` function returns a data frame that consists of the original data frame plus a variable containing the length interval categories for each fish. The default name of the new variable (`LCat`) can be changed with the `vname=` argument to `lencat()`. The `lencat()` function result must be assigned into an object, preferably named differently from the original age sample.

The starting category for 10-mm length categories for the Lake Ontario rock bass is determined by finding the minimum length in the age sample with,

```
> Summarize(rb.age$t1, numdigs = 1)
```

n	Mean	St. Dev.	Min.	1st Qu.	Median	3rd Qu.	Max.
135.0	200.7	40.3	111.0	168.0	201.0	234.5	278.0

and then starting the categories with the even-number 10-mm interval just below this value. This is completed with,

```
> rb.age1 <- lencat(rb.age, "t1", startcat = 110, w = 10)
> view(rb.age1)
```

	age	t1	LCat
3	7	211	210
37	4	143	140
67	4	208	200
102	10	236	230
125	3	137	130
126	10	275	270

Once the length category variable has been added to the age sample data frame, the `table()` function can be used to construct the summary contingency table of numbers of fish in each combined length and age category. The row variable (length category) is the first and the column variable (age) is the second argument to this function. The results of `table()` should be assigned to an object and then submitted as the first argument to `prop.table()` along with `margin=1` as a second argument¹ to construct a row-proportions table. The resulting row-proportions table is the actual age-length key determined from the age sample and is ready to be applied to the length sample. The summary contingency table and the row-proportion table, i.e., the age-length key, are constructed with (note: the results were rounded for display purposes only),

```
> rb.raw <- table(rb.age1$LCat, rb.age1$age)
> rb.key <- prop.table(rb.raw, margin = 1)
> round(rb.key, 3)
```

	3	4	5	6	7	8	9	10	11
110	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
120	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
130	0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
140	0.10	0.90	0.00	0.00	0.00	0.00	0.00	0.00	0.00
150	0.00	0.80	0.20	0.00	0.00	0.00	0.00	0.00	0.00
160	0.00	0.70	0.20	0.10	0.00	0.00	0.00	0.00	0.00

¹`margin=1` indicates "rows", whereas as `margin=2` would indicate "columns".

```

170 0.00 0.50 0.30 0.20 0.00 0.00 0.00 0.00 0.00
180 0.00 0.40 0.40 0.10 0.10 0.00 0.00 0.00 0.00
190 0.00 0.20 0.30 0.20 0.30 0.00 0.00 0.00 0.00
200 0.00 0.10 0.20 0.20 0.40 0.10 0.00 0.00 0.00
210 0.00 0.00 0.10 0.30 0.50 0.00 0.10 0.00 0.00
220 0.00 0.00 0.00 0.30 0.30 0.20 0.20 0.00 0.00
230 0.00 0.00 0.00 0.00 0.40 0.10 0.20 0.20 0.10
240 0.00 0.00 0.00 0.00 0.10 0.50 0.20 0.20 0.00
250 0.00 0.00 0.00 0.00 0.30 0.20 0.20 0.20 0.10
260 0.00 0.00 0.00 0.00 0.00 0.25 0.50 0.25 0.00
270 0.00 0.00 0.00 0.00 0.00 0.20 0.00 0.80 0.00
280

```

3 Assigning Ages to Individuals with the Age-Length Key

3.1 Age Assignment

The semi-random age assignment method described in **Section 1.2** is implemented with `age.key()`. This function requires the following arguments,

- `key` : A numeric matrix containing the age-length key (as constructed with `prop.table()`).
- `d1` : A data frame containing the length-sample of fish.
- `c1` : A number or character string indicating which column of `d1` contains the length measurements.
- `ca` : A number or character string indicating which column of `d1` should receive the age assignments. If the column does not exist in the current data frame then one will be appended with the name given in `ca`.
- `type` : A character string indicating whether to use the semi-random (`type="SR"`, *default*) or completely random (`type="CR"`) technique for assigning ages to individual fish.

The `age.key()` function will determine the length categories to construct based on the age-length key sent in the `key=` argument. The results of `age.key()` should be assigned to an object, preferably with a name different from the original length sample. Random ages were assigned to the un-aged fish in the length sample with,

```

> rb.len1 <- age.key(rb.key, rb.len, c1 = "t1", ca = "age")
> view(rb.len1)

```

```

      age t1
253    5 185
418    4 150
459    4 203
956    5 166
968   10 234
1142   6 197

```

The original (not modified) age sample data frame and the modified length sample data frame (i.e., now containing the ages assigned via the age-length key) should then be row-bound together to construct a data frame that consists of lengths and ages for all fish in the original sample. These two data frames are combined with,

```
> rb.comb <- rbind(rb.age, rb.len1)
> view(rb.comb)
```

```
   age t1
321  7 196
359  5 162
403  6 213
783  4 191
861  5 203
987  6 224
```

3.2 Summary Computations

The assigned ages in the `rb.comb` data frame can then be used to, for example, compute an age-frequency,

```
> table(rb.comb$age)
```

```
 3  4  5  6  7  8  9 10 11
5 318 264 217 310  74  63 29  8
```

or calculate summary statistics of size-at age for ALL individuals in the sample,

```
> Summarize(rb.comb$t1 ~ rb.comb$age, numdigs = 2)
```

	n	Mean	St. Dev.	Min.	1st Qu.	Median	3rd Qu.	Max.
3	5	129.40	12.28	111	125.0	131.0	137.0	143
4	318	175.68	15.71	130	164.0	175.0	186.8	208
5	264	187.42	14.07	153	179.5	187.0	196.0	217
6	217	199.77	17.21	162	188.0	202.0	213.0	228
7	310	209.61	15.57	180	198.0	208.0	217.0	258
8	74	226.80	17.89	200	207.2	226.0	242.0	270
9	63	229.05	13.90	210	220.5	225.0	237.5	265
10	29	244.83	14.80	230	234.0	240.0	250.0	278
11	8	241.50	11.41	230	233.5	235.5	253.2	258

or plot (**Figure 1**) length-at-age for ALL individuals in the sample,

4 Further Thoughts

4.1 Completely Random Age Assignment with Age-Length Key

In the completely random method of assigning ages to individual fish, the age assigned to a fish is randomly assigned with probabilistic weights proportional to the proportion of fish in each age of the same length interval. This is essentially the same rule used to assign age to the fish “left-over” because of fractionality in the semi-random method. For example, suppose that 50% of the fish in a length interval in the age-sample were age-1, 25% were age-2, and 25% were age-3. A single fish in the same length interval in the length sample would then have a 50% chance of being assigned an age of 1, a 25% chance of being assigned an age of 2, and a 25% chance of being assigned an age of 3. The specific age it is assigned depends completely on randomness. Thus, in this example, we would expect 11 of 22 fish in this length interval in the length-sample to be assigned an age of 1 but in actuality it could be more or it could be less.

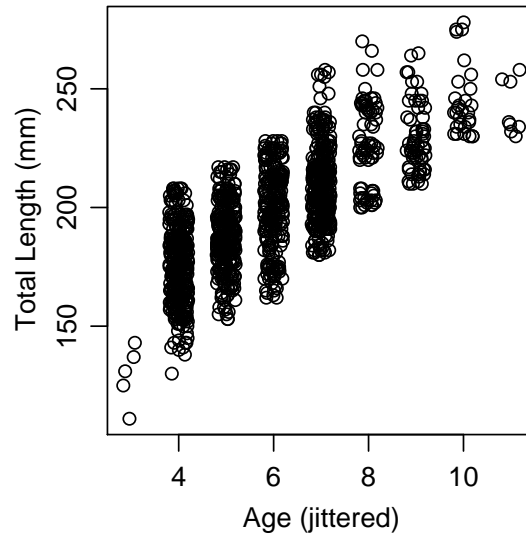


Figure 1. Length-at-age for Lake Ontario rock bass.

The completely random method of assigning ages to individual fish is also implemented with `age.key()`. The arguments to the function are exactly the same as they were for the semi-random method except that a `type="CR"` argument must be included. Use of the completely random method of assigning ages and summary results are computed with,

```
> rb.len2 <- age.key(rb.key, rb.len, type = "cr", cl = "tl", ca = "age")
> rb.comb2 <- rbind(rb.age, rb.len2)
> table(rb.comb2$age)
```

```
 3  4  5  6  7  8  9 10 11
5 298 288 209 311 78 63 33  3
```

```
> Summarize(rb.comb2$tl ~ rb.comb2$age, numdigs = 2)
```

	n	Mean	St. Dev.	Min.	1st Qu.	Median	3rd Qu.	Max.
3	5	129.40	12.28	111	125.0	131.0	137	143
4	298	175.97	16.04	130	164.0	175.5	187	208
5	288	187.00	14.86	150	178.0	187.0	197	218
6	209	198.77	17.23	160	185.0	198.0	213	228
7	311	209.46	15.37	181	200.0	207.0	217	257
8	78	225.86	17.49	200	208.0	226.5	240	270
9	63	229.59	13.27	210	221.0	226.0	239	265
10	33	246.42	13.70	230	236.0	242.0	253	278
11	3	242.33	13.65	233	234.5	236.0	247	258

The semi-random method of assigning ages to individual fish is the preferred method, especially when the age-length key is used to extrapolate the ages of fish in a single length sample. The completely random method is used primarily when studying the theoretical sources of variability inherent in an age-length key analysis. In other words, the semi-random method is the method most often used by the practicing biologist.

References

- Bettoli, P. W. and L. E. Miranda. 2001. A cautionary note about estimating mean length at age with subsampled data. *North American Journal of Fisheries Management* 21:425–428. [2](#)
- Isermann, D. A. and C. T. Knight. 2005. A computer program for age–length keys incorporating age assignment to individual fish. *North American Journal of Fisheries Management* 25:1153–1160. [1](#), [2](#), [3](#)
- Kimura, D. A. 1977. Statistical assessment of the age-length key. *Journal of the Fisheries Research Board of Canada* 34:317–324. [1](#)
- Wolfert, D. R. 1980. Age and growth of rock bass in eastern Lake Ontario. *New York Fish and Game Journal* 27:88–90. [1](#)