

fishR Vignette - Precision and Accuracy in Ages

Dr. Derek Ogle, Northland College

December 16, 2013

As with most estimation techniques, errors in age estimation will occur. The first type of error that can occur is related to the validity of the method of age determination. For example, if the number of annuli found on scales reaches an asymptote that is considerably less than the maximum age of the fish then the interpretation of scales may provide invalid or inaccurate estimates of age. Another type of error is associated with the interpreters of the ageing structure, or the agers themselves. An ager may be producing biased interpretations if his or her interpretations are systematically different than what is known to be true, drift over time (away from the “truth”), or are different than another ager (supposedly interpreting the “truth”). In addition, the precision of non-systematic differences in interpretations should be measured between two agers or between an ager’s interpretations of the same structures over time. This vignette will focus on graphical and statistical measures used to identify systematic bias (Section 1) and measure precision (Section 2). More complete treatments of these items can be found in [Campana *et al.* \(1995\)](#) and [Campana \(2001\)](#).

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

```
> library(FSA)
> library(FSAdata) # for datafile
```

1 Systematic Bias

1.1 Test of Symmetry

The systematic bias of an ager can occur if the assigned ages by the ager differ from a more experienced ager, if the assigned ages differ from ages assigned to the same sample at a previous time, or if there are differences between structures. [Hoenig *et al.* \(1995\)](#) described a test of symmetry for determining if significant differences existed between structures or agers¹. The test proposed by [Hoenig *et al.* \(1995\)](#) is best described through examination of an age-agreement table (e.g., Table 1).

Table 1. The age-agreement table for two agers of Atlantic Croaker otoliths.

Reader 1	Reader 2										
	1	2	3	4	5	6	7	8	9	10	11
1	3	0	0	0	0	0	0	0	0	0	0
2	0	34	1	0	0	0	0	0	0	0	0
3	0	0	29	1	0	0	0	0	0	0	0
4	0	0	0	17	1	0	0	0	0	0	0
5	0	0	0	3	58	0	0	0	0	0	0
6	0	0	0	0	4	26	5	0	0	0	0
7	0	0	0	0	0	2	52	1	0	0	0
8	0	0	0	0	0	0	0	36	1	0	0
9	0	0	0	0	0	0	0	2	39	0	0
10	0	0	0	0	0	0	0	0	0	1	1
11	0	0	0	0	0	0	0	0	0	0	0

On an age-agreement table, the main diagonal represents the frequency of fish for which the same age was obtained by the two agers or on the two structures. Each cell off of the main diagonal represents a difference

¹[Hoenig *et al.* \(1995\)](#) first suggested this test for assessing bias in age assessments. However, this test existed previously in the statistical literature and is called a *Bowker’s Test of Symmetry*.

in assigned age between agers or structures. If there are no systematic differences between agers or structures then it would be expected that the disagreements in age would fall randomly on either side of the diagonal, producing an approximately symmetric age-agreement table. The test of symmetry proposed by [Hoenig et al. \(1995\)](#) uses a chi-square-type statistical test to determine if the age-agreement table is symmetric (the null) or not. If the age-agreement table is determined to be asymmetric then it can be concluded that there is a systematic difference in ages observed between agers or structures.

The test of symmetry described above can be computed in R with `ageBias()`. The first argument to this function is a formula of the form `col~row` where `col` is a variable name for the structure or ager that will form the columns and `row` will form the rows of the age-agreement table. Generally the structure or ager thought to provide the most accurate ages should be used as the column variable in the age-agreement table. The second argument (`data=`) is the name of the data frame that contains the `col` and `row` data given in the first argument. In addition, the columns and rows of the age agreement table can be labeled with `col.lab=` and `row.lab=`. The result of `ageBias()` should be assigned to an object. The age agreement table and the test of symmetry are extracted from that object with `summary()` using the `what="symmetry"` argument.

[O’Gorman et al. \(1987\)](#) examined ages assessed from scales and otoliths for 104 Alewife (*Alosa pseudoharengus*) from Lake Huron (in `Alewifelh`). These data are read and the structure is examined with

```
> data(Alewifelh)
> str(Alewifelh)
'data.frame': 104 obs. of 2 variables:
 $ otoliths: int 0 0 1 1 1 1 1 1 1 1 ...
 $ scales : int 0 0 0 1 1 1 1 1 1 1 ...
```

Otoliths are generally thought to be more accurate than scales so the variable containing the otolith ages should appear as the columns in the age agreement table and thus should be on the left-hand-side of the formula in the first argument to `ageComp()`. The test of symmetry for detecting systematic differences between ages assigned through examination of the two structures was constructed with

```
> ab.ale <- ageBias(otoliths~scales,data=Alewifelh,col.lab="Otolith Age",
  row.lab="Scale Age")
> summary(ab.ale,what="symmetry")
Raw agreement table (square)
      Otolith Age
Scale Age 0 1 2 3 4 5 6 7 8 9 10
0      2 1 - - - - - - - -
1      - 14 - - - - - - - -
2      - 3 16 2 - - - - - -
3      - - 4 11 5 2 1 - - - -
4      - - - - 12 4 3 1 1 - -
5      - - - - - 1 4 2 5 4 - -
6      - - - - - - 2 1 - 1 2
7      - - - - - - - - - - -
8      - - - - - - - - - - -
9      - - - - - - - - - - -
10     - - - - - - - - - - -
Bowker's (Hoenig's) Test of Symmetry
df chi.sq      p
16 34.47 0.004698
```

The test of symmetry ($p = 0.0047$) indicates that there are systematic differences in assigned ages between the two structures. The age-agreement table suggests that scales underestimate age relative to otoliths after about age-5. If the otolith ages have been validated then these results would suggest that scale ages should not be used.

1.2 Age-Bias Plots

Campana *et al.* (1995) noted that several statistical methods were capable of detecting systematic ageing differences but were incapable of detecting both linear and non-linear biases in ageing. For example, the statistical methods were not sensitive to detecting if one ager systematically over-aged for younger fish but systematically under-aged for older fish. To address this problem, Campana *et al.* (1995) proposed a so-called *age-bias plot*.

In an age-bias plot, the mean (with 95% confidence interval) age from the ager, time, or structure thought to be less accurate is plotted for each distinct age from the ager, time, or structure thought to be more accurate (e.g., Figure 1). If the 95% confidence intervals do not “contain” the 1:1 agreement line then the age-bias plot suggests a systematic difference between agers, times, or structures. An age-bias plot is constructed by sending the object saved from `ageBias()` to `plot()` along with the `what="bias"` argument (which is the default).

Potential biases in the ageing of Lake Huron Alewife were further examined by producing an age-bias plot (Figure 1) of the paired otolith and scale readings with

```
> plot(ab.ale)
```

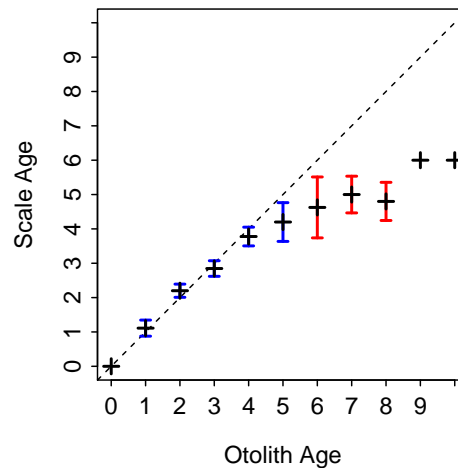


Figure 1. Age-bias plot for age-estimates from otoliths and scales of Lake Huron Alewife.

These results further suggest that the scales underestimate the age of Alewife relative to otoliths beginning at about age-5 or age=6.

Muir *et al.* (2008) used a slight modification of the age-bias plot where the differences in ages between the groups was plotted on the y-axis (Figure 2). Their modification can be constructed by including the `difference=TRUE` argument in `plot()`.

```
> plot(ab.ale,difference=TRUE)
```

The 95% confidence intervals in Figure 1 and Figure 2 are color coded. The color of the confidence interval corresponds to whether a one-sample t-test indicated that the mean of the ages plotted on the y-axis was equal to the corresponding age on the x-axis for the original age-bias plot or whether the difference in ages plotted on the y-axis was equal to zero on the difference age-bias plot. The significance was determined by comparing an adjusted p-value (default is the “Holm” method²) to the `sig.level=` argument in the original call to `ageBias()` (=0.05 by default). Note that the p-value adjustment makes it more difficult to reject the

²See `p.adjust.methods()` for the full list of possible adjustment methods.

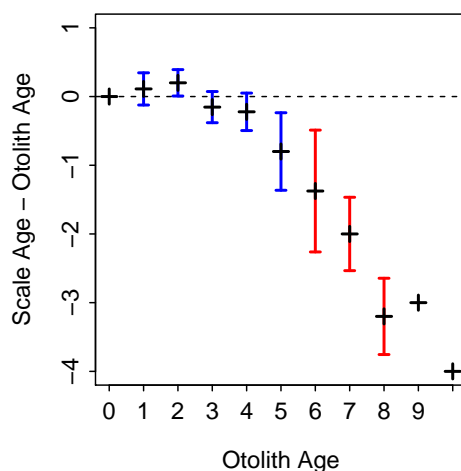


Figure 2. Alternative age-bias plot for age-estimates from otoliths and scales of Lake Huron Alewife.

null hypothesis of no difference then in a situation where no p-value adjustment occurred. The confidence intervals presented in the age-bias plot are *unadjusted* confidence intervals; thus, more differences from the 1:1 line (original age-bias plot) or the zero line (differences age-bias plot) than are warranted will appear. The significance colors presented are *adjusted* and are more conservative than the actual confidence intervals. Unadjusted p-values can be obtained by using `method="none"` in `ageBias()`. In addition, different color coding for the confidence intervals can be avoided by setting the `col.err=` and `col.err.sig=` arguments to the same color.

The fisheries scientist may want to visualize the actual observations on an age-bias plot. However, this is problematic because the discrete nature of age data results in a large amount of overplotting (i.e., individual points cannot be discerned). One possible correction is to modify the age-bias plot by plotting individual points with transparency such that a single point is very transparent but as more points are plotted on top of each other the “point” becomes darker. The number of overplotted points required to make the point appear completely dark is controlled with the `transparency=` argument. This argument can be set equal to a fraction where the denominator dictates the number of points required to remove all transparency. For example, if `transparency=1/10` then 10 overplotted points are required to have a completely dark point. These types of points are added to the age-bias plot Figure 3 by including `show.pts=TRUE` to `plot()`.

```
> plot(ab.ale, show.pts=TRUE, transparency=1/10)
```

Additionally, the age-bias plot can be further modified to include a “range bar” that shows the range of ages for the y-axis variable for each age of the x-axis variable. These types of bars are added to the age-bias plot Figure 4 by including `show.rng=TRUE` to `plot()`.

```
> plot(ab.ale, show.rng=TRUE)
```

Finally, you can see the statistics underlying an age-bias plot by adding `what="bias"` and, optionally, `difference=TRUE`, to `summary()` with the saved `ageComp()` object. For example,

```
> summary(ab.ale, what="bias")
Summary of Scale Age by Otolith Age
  otoliths  n min max mean      SE      t  adj.p  sig  LCI  UCI
    0      2  0  0  0.00    NA     NA    NA FALSE  NA  NA
    1     18  0  2  1.11  0.1111  1.00  0.33138 FALSE 0.877 1.35
    2     20  2  3  2.20  0.0918  2.18  0.16836 FALSE 2.008 2.39
    3     13  2  3  2.85  0.1041 -1.48  0.33098 FALSE 2.619 3.07
```

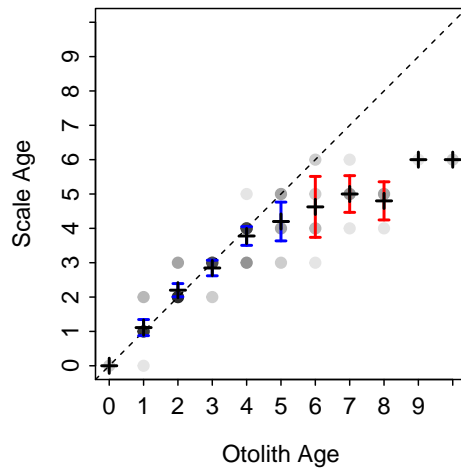


Figure 3. Age-bias plot for age-estimates from otoliths and scales of Lake Huron Alewife. Darker points on the plot represent more individuals.

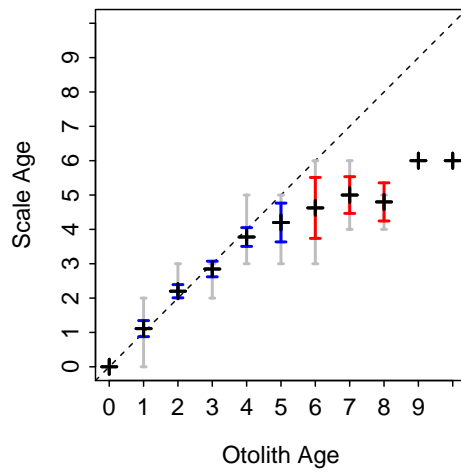


Figure 4. Age-bias plot for age-estimates from otoliths and scales of Lake Huron Alewife. The gray intervals represent the range of observed ages.

```

 4 18  3  5 3.78 0.1292 -1.72 0.31112 FALSE 3.505 4.05
 5 10  3  5 4.20 0.2494 -3.21 0.05354 FALSE 3.636 4.76
 6  8  3  6 4.62 0.3750 -3.67 0.04801  TRUE 3.738 5.51
 7  7  4  6 5.00 0.2182 -9.16 0.00071  TRUE 4.466 5.53
 8  5  4  5 4.80 0.2000 -16.00 0.00071  TRUE 4.245 5.36
 9  1  6  6 6.00      NA      NA      NA FALSE      NA  NA
10  2  6  6 6.00      NA      NA      NA FALSE      NA  NA

```

```
> summary(ab.ale,what="bias",difference=TRUE)
```

```
Summary of Scale Age by Otolith Age
```

otoliths	n	min	max	mean	SE	t	adj.p	sig	LCI	UCI
0	2	0	0	0.00	NA	NA	NA	FALSE	NA	NA
1	18	0	2	1.11	0.1111	1.00	0.33138	FALSE	0.877	1.35
2	20	2	3	2.20	0.0918	2.18	0.16836	FALSE	2.008	2.39
3	13	2	3	2.85	0.1041	-1.48	0.33098	FALSE	2.619	3.07
4	18	3	5	3.78	0.1292	-1.72	0.31112	FALSE	3.505	4.05
5	10	3	5	4.20	0.2494	-3.21	0.05354	FALSE	3.636	4.76
6	8	3	6	4.62	0.3750	-3.67	0.04801	TRUE	3.738	5.51
7	7	4	6	5.00	0.2182	-9.16	0.00071	TRUE	4.466	5.53
8	5	4	5	4.80	0.2000	-16.00	0.00071	TRUE	4.245	5.36
9	1	6	6	6.00	NA	NA	NA	FALSE	NA	NA
10	2	6	6	6.00	NA	NA	NA	FALSE	NA	NA

2 Precision

In age assessment, precision is the reproducibility of estimated ages either between- or within-agers regardless of whether the estimated ages are accurate or not (Campana 2001). Measures of precision can be used as a relative measure of the ease with which ages are assigned, reproducibility of an individual’s age assessments, or for comparing multiple ager’s age assignments to the same structures. It must be made clear, though, that highly precise estimates should not imply highly accurate age assessments; it is possible, and common, for age assessments to be highly reproducible but wrong!

A common measure of precision is to compute the percent of all paired age assessments that are in agreement. However, the percent-agreement measure should not be used solely as it varies widely among species and among ages within a species. The classic example of the inadequacy of the percent-agreement measure was shown by Beamish and Fournier (1981). They noted that 95% agreement to within one year between two agers is very poor agreement for a species that consists of a few year-classes, whereas the same level of agreement is very good agreement for a species that has several year-classes spread over a long-period of time.

Two other statistical measures have been forwarded as “better” measures of precision. Beamish and Fournier (1981) recommended the average percent error (APE),

$$APE = \frac{\sum_{j=1}^n APE_j}{n} \quad \text{where} \quad APE_j = 100 * \frac{\sum_{i=1}^R \frac{|x_{ij} - \bar{x}_j|}{\bar{x}_j}}{R} \quad (1)$$

where APE_j is the average percent error for the j th fish, x_{ij} is the i th age estimate on the j th fish, \bar{x}_j is the mean age estimate for the j th fish, R is the number of times that each fish was aged (assumed to be the same for all fish), and n is the number of aged fish in the sample. The APE formula (1) tacitly assumes that the standard deviation of the age estimates are proportional to the mean of the age estimates (Chang 1982). To avoid this assumption, Chang (1982) suggested that precision should be measured by the coefficient of variation (CV),

$$CV = \frac{\sum_{j=1}^n CV_j}{n} \quad \text{where} \quad CV_j = 100 * \frac{s_j}{\bar{x}_j} = 100 * \frac{\sqrt{\frac{\sum_{i=1}^R (x_{ij} - \bar{x}_j)^2}{R-1}}}{\bar{x}_j} \quad (2)$$

where CV_j is the coefficient of variation for the j th fish and s_j is the standard deviation of the age estimates for the j th fish.

The CV measure is the preferred measure because the coefficient of variation has known statistical properties. However, [Chang \(1982\)](#) shows that the CV will be a constant value greater than APE and [Campana \(2001\)](#) showed that the CV was approximately 40% greater than the APE in practice. Thus, for relative comparisons, it does not matter whether APE or CV is used.

A “cut-off” criterion for what is an acceptable value of precision cannot be made as it will depend on the species being examined, the experience of the agers, and the structure being examined. However, [Campana \(2001\)](#) found in an examination of a larger number of studies that the median CV was 7.6% and suggested that a CV of 5% could serve as a general reference point. Thus, a CV of less than 5% suggests that the assigned ages are acceptably precise and values greater than 5% suggest that the ages are relatively imprecise.

All measures of precision discussed in this section can be computed with `agePrecision()`. The arguments to `agePrecision()` are a formula of the form `~var1+var2+...`, where `var1`, `var2`, etc. are the variables containing the age assessments³ and the corresponding data frame in `data=`. The results of `agePrecision()` should be saved to an object such that specific results can be extracted with `summary()`. The summary measures of precision discussed above are extracted with `summary()` using `what="precision"` (the default). A table of percentage agreement between the structures within various amounts (e.g., ages agree within ± 1 year) is extracted with `what="agreement"` in `summary()`.

Old Dominion University in conjunction with the Virginia Marine Resources Commission (VMRC) conducts routine ageing of Virginia’s marine catch to provide the VMRC with the fundamental demographic data necessary for management. One of the species studied in their continuing project is the Atlantic Croaker (*Micropogonias undulatus*) which is one of the five most targeted species by anglers. For part of their quality control analysis of the ageing process they had two different agers examine 317 otolith thin sections. Their data (in the file **Croaker1**) is analyzed here to estimate precision among readers⁴.

```
> data(Croaker1)
> str(Croaker1)
'data.frame': 317 obs. of 2 variables:
 $ reader1: int  1 1 1 2 2 2 2 2 2 2 ...
 $ reader2: int  1 1 1 2 2 2 2 2 2 2 ...

> ap.croak <- agePrecision(~reader1+reader2,data=Croaker1)
> summary(ap.croak,what="precision")

Precision summary statistics
  n R   CV   APE PercAgree
317 2 0.8834 0.6247    93.06

> summary(ap.croak,what="agreement")

Percentage of fish by differences in ages between pairs of assignments
  0 1
93.06 6.94
```

These results indicate very high agreement between the two agers (i.e., APE and CV <1%, 93% exact agreement, 100% agreement within one year).

³If only two variables are being considered this formula can also be of the same form used in `ageBias()`.

⁴One should check for a lack of bias between readers before estimating precision.

References

- Beamish, R. J. and D. A. Fournier. 1981. A method for comparing the precision of a set of age determinations. *Canadian Journal of Fisheries and Aquatic Sciences* 38:982–983. [6](#)
- Campana, S. E. 2001. Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. *Journal of Fish Biology* 59:197–242. [1](#), [6](#), [7](#)
- Campana, S. E., M. C. Annand, and J. I. McMillan. 1995. Graphical and statistical methods for determining the consistency of age determinations. *Transactions of the American Fisheries Society* 124:131–138. [1](#), [3](#)
- Chang, W. Y. B. 1982. A statistical method for evaluating the reproducibility of age determination. *Canadian Journal of Fisheries and Aquatic Sciences* 39:1208–1210. [6](#), [7](#)
- Hoening, J. M., M. J. Morgan, and C. A. Brown. 1995. Analysing differences between two age determination methods by tests of symmetry. *Canadian Journal of Fisheries And Aquatic Systems* 52:364–368. [1](#), [2](#)
- Muir, A. M., M. P. Ebener, J. X. He, and J. E. Johnson. 2008. A comparison of the scale and otolith methods of age estimation for lake whitefish in Lake Huron. *North American Journal of Fisheries Management* 28:625–635. [3](#)
- O’Gorman, R., D. H. Barwick, and C. A. Bowen. 1987. Age and Growth of Fish, chapter Discrepancies between ages determined from scales and otoliths for alewives from the Great Lakes, pp. 203–210. The Iowa State University Press, Ames, IA. [2](#)

Reproducibility Information

Version Information

- **Compiled Date:** Mon Dec 16 2013
- **Compiled Time:** 8:59:32 PM
- **Code Execution Time:** 1.85 s

R Information

- **R Version:** R version 3.0.2 (2013-09-25)
- **System:** Windows, i386-w64-mingw32/i386 (32-bit)
- **Base Packages:** base, datasets, graphics, grDevices, methods, stats, utils
- **Other Packages:** FSA_0.4.3, FSAdat_0.1.4, gdata_2.13.2, knitr_1.5.15
- **Loaded-Only Packages:** bitops_1.0-6, car_2.0-19, caTools_1.16, cluster_1.14.4, evaluate_0.5.1, formatR_0.10, Formula_1.1-1, gplots_2.12.1, grid_3.0.2, gtools_3.1.1, highr_0.3, Hmisc_3.13-0, KernSmooth_2.23-10, lattice_0.20-24, MASS_7.3-29, multcomp_1.3-1, mvtnorm_0.9-9996, nlme_3.1-113, nnet_7.3-7, plotrix_3.5-2, quantreg_5.05, sandwich_2.3-0, sciplot_1.1-0, SparseM_1.03, splines_3.0.2, stringr_0.6.2, survival_2.37-4, tools_3.0.2, zoo_1.7-10
- **Required Packages:** FSA, FSAdat and their dependencies (car, gdata, gplots, Hmisc, knitr, multcomp, nlme, plotrix, quantreg, sciplot)