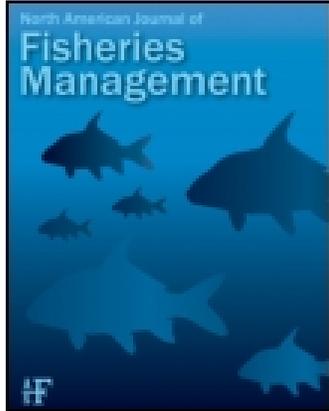


This article was downloaded by: [Tennessee Tech University]

On: 09 April 2015, At: 11:43

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



North American Journal of Fisheries Management

Publication details, including instructions for authors and subscription information:
<http://www.tandfonline.com/loi/ujfm20>

Cautionary Note about Estimating Mean Length at Age with Subsampled Data

Phillip W. Bettoli^a & Leandro E. Miranda^b

^a U.S. Geological Survey, Biological Resources Division Tennessee Cooperative Fishery Research Unit, Tennessee Technological University, Cookeville, Tennessee, 38505, USA

^b U.S. Geological Survey, Biological Resources Division, Mississippi Cooperative Fish and Wildlife Research Unit, Mississippi State University, Mississippi, Mississippi State, 39762, USA

Published online: 08 Jan 2011.

To cite this article: Phillip W. Bettoli & Leandro E. Miranda (2001) Cautionary Note about Estimating Mean Length at Age with Subsampled Data, North American Journal of Fisheries Management, 21:2, 425-428, DOI: [10.1577/1548-8675\(2001\)021<0425:CNAEML>2.0.CO;2](https://doi.org/10.1577/1548-8675(2001)021<0425:CNAEML>2.0.CO;2)

To link to this article: [http://dx.doi.org/10.1577/1548-8675\(2001\)021<0425:CNAEML>2.0.CO;2](http://dx.doi.org/10.1577/1548-8675(2001)021<0425:CNAEML>2.0.CO;2)

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Cautionary Note about Estimating Mean Length at Age with Subsampled Data

PHILLIP W. BETTOLI*

U.S. Geological Survey, Biological Resources Division,
Tennessee Cooperative Fishery Research Unit,¹
Tennessee Technological University, Cookeville, Tennessee 38505, USA

LEANDRO E. MIRANDA

U.S. Geological Survey, Biological Resources Division,
Mississippi Cooperative Fish and Wildlife Research Unit,²
Mississippi State University, Mississippi State, Mississippi 39762, USA

Abstract.—Subsampling fixed or random numbers of fish per length category to estimate mean length at age is commonplace. However, biologists often ignore the fact that those data are collected in a stratified manner and do not represent a simple random sample of the population. We demonstrate that failure to consider the stratified nature of data and use the correct formulae to calculate means and standard errors will usually result in biased estimates of mean length at age and will always inflate standard error estimates. If the distribution of lengths within a particular age is highly skewed, estimates will be severely biased if the data are not treated in a stratified manner. Subsampling in proportion to the number of fish in each length category may be superior from a statistical standpoint; however, the more commonplace sampling of a fixed number of fish per length category is superior from the standpoints of logistics and the frequent need to accurately estimate age proportions in the largest length categories.

Fishery biologists often estimate mean length at age to assess fish growth, a variable that is fundamental to understanding population dynamics and managing fish populations (Ricker 1975). Estimates about the population are often made from a sample selected at random, in which all the fish collected are assessed for age. However, biologists often resort to subsampling because a sample may contain more fish than can be processed, or because the increased reliance on otoliths requires minimizing the number of fish to be killed. Subsampling to generate estimates of ages for the en-

tire sample was first applied by Fridriksson (1934; cited in Ketchen 1949). Fridriksson proposed estimating fish ages from a subsample selected according to length-group and in proportion to the frequency distribution in the sample; the proportions of different ages in each length-group in the subsample are then used to assign ages to all fish in each length-group in the sample. Ketchen (1949) modified Fridriksson's approach by subsampling a fixed number of fish in each length-group and suggested that this method would be a more efficient means of assigning ages to the sample, because weighting all length groups equally increases representation of older age-groups, which are often a small fraction of a sample. This method of creating an age-length key has been widely used for years (e.g., Allen 1966), although several authors have commented on potential biases (e.g., Kimura 1977; Westheim and Ricker 1977).

In the second edition of *Fisheries Techniques*, DeVries and Frie (1996) discussed the use of age-length keys to estimate age-frequency distributions and provided an example of how to estimate age-frequency distribution from a subsample. Currently, the most common method of subsampling is to create length-groups of 10-mm, 25-mm, or 1-in lengths and collect age data structures from a fixed number of fish per length-group. The ages of fish in the subsample are then estimated by various methodologies, and statistics such as mean length and variance are computed for each age group represented in the subsample. The practice of sampling a population of units (i.e., individual fish), classifying the units into strata (i.e., length-groups), and then selecting a sample from each stratum is formally termed double-sampling for stratification (Thompson 1992).

* Corresponding author: pbettoli@tntech.edu

¹ The Tennessee Unit is supported by the U.S. Geological Survey, Tennessee Technological University, and the Tennessee Wildlife Resources Agency.

² The Mississippi Unit is supported by the U.S. Geological Survey, Mississippi State University, and the Mississippi Department of Wildlife, Fisheries, and Parks.

Received May 31, 2000; accepted October 10, 2000

The Problem

Despite the fact that age-length keys have been in use for decades and are discussed in recent fisheries textbooks, we have noted that statistics about age-groups are often computed incorrectly when subsampling is used. The subsampling approach described by Ketchen (1949) creates a uniform length distribution if a fixed number of fish are subsampled per length-group, yet many biologists treat the subsample as if it was a randomly selected nonuniform distribution mirroring the distribution of lengths and ages in the complete sample. For instance, in the 1998 and 1999 volumes of the *North American Journal of Fisheries Management*, about half (14 of 27) of the manuscripts presenting age and growth data reported that the fish estimated for age were subsampled uniformly by length-group, yet most (9 of 14) of those studies treated the data as if they were randomly selected from the sample. Through our work with fisheries biologists in several states, we became aware that subsampling by length-group is commonplace, but the data are often treated incorrectly. Below we show that using an inappropriate design to estimate sample statistics can, under most circumstances, produce biased estimates of mean length at age and the corresponding variance. We also offer advice on how to avoid these statistical concerns. Readers, particularly marine biologists, interested in estimating maturity-at-age from length-stratified data are directed to Morgan and Hoenig (1997).

Estimation Procedures

Incorrect estimates are made when the mean length of the i th age-group (\bar{L}_i) and variance (S_i^2) are estimated directly from the subsample as:

$$\bar{L}_i = (\sum n_{ij}\bar{l}_{ij})/n_i \quad (1)$$

$$S_i^2 = [\sum (l_{ijk} - \bar{L}_i)^2]/(n_i - 1) \quad (2)$$

where n_{ij} = number of fish of the i th age-group subsampled in the j th length-group, l_{ijk} = the length of the k th fish in the j th length-group and the i th age-group, $\bar{l}_{ij} = (\sum l_{ijk})/n_{ij}$, and n_i = number of age- i fish in the subsample. Equations 1 and 2 are often used but are incorrect because they produce statistics that do not represent the sample if subsampling is not in proportion to the frequency distribution of the sample. The right way to treat these data is to extrapolate the subsample to the sample and estimate the statistics based on all the fish in the sample, as follows:

$$\bar{L}_i = (\sum N_{ij}\bar{l}_{ij})/N_i \quad (3)$$

$$S_i^2 = \{\sum N_{ij}[\sum (l_{ijk} - \bar{L}_i)^2/n_{ij}]\}/(N_i - 1) \quad (4a)$$

$$S_i^2 = [\sum N_{ij}(\bar{l}_{ij} - \bar{L}_i)^2]/(N_i - 1) \quad (4b)$$

where $N_{ij} = N_j(n_{ij}/n_j)$, N_j is the number of fish in the j th length-group, n_j is the number of fish subsampled in the j th length-group, and $N_i = \sum N_{ij}$ over all j length-groups. Equations 3 and 4b are equivalent to equations 16.1 and 16.2 in DeVries and Frie (1996) when the midpoints of each length-group are substituted for \bar{l}_{ij} . Equations 4a and 4b provide slightly different estimates (the magnitude of the difference depending on the distribution of lengths), but both are acceptable estimators of the variance.

For example, a length-frequency distribution for a sample of 200 largemouth bass *Micropterus salmoides* collected by electrofishing on Kentucky Lake, Tennessee, is depicted in Figure 1A. Otoliths were removed from as many as 10 individuals in each 25-mm length-group to determine ages and mean length at ages 2 through 4 (Figures 1B, C, and D, respectively). Using the equations presented above, we estimated \bar{L}_i and S_i^2 with both the right and the wrong procedures to demonstrate the magnitude of bias; we then calculated standard errors associated with each mean. The mean lengths generated using only the data from the subsamples do not agree with the means estimated from the entire sample. The only time these means should agree is when the length distribution for an age-group is bell-shaped and symmetrical, which is seldom the case in fish populations. For age-3 largemouth bass (Figure 1C), the two means were similar because the length-frequency distribution was nearly symmetrical. The standard error, used to generate confidence intervals around means and to conduct statistical comparisons between means (e.g., t -tests), would be overestimated when derived from equation (2), that is, $SE = (S_i^2/n_i)^{0.5}$, but unbiased when derived from an equation (4) approach ($SE = [S_i^2/N_i]^{0.5}$). An overestimated standard error facilitates type II errors, that is, accepting a null hypothesis when the alternative is true. In nonstatistical jargon, this means that when comparisons of growth estimates are made, determining significant differences is more difficult when the incorrect equations are used.

To further illustrate the biases that result from using the wrong equations with stratified data, we "sampled" randomly three fish from each 10-mm

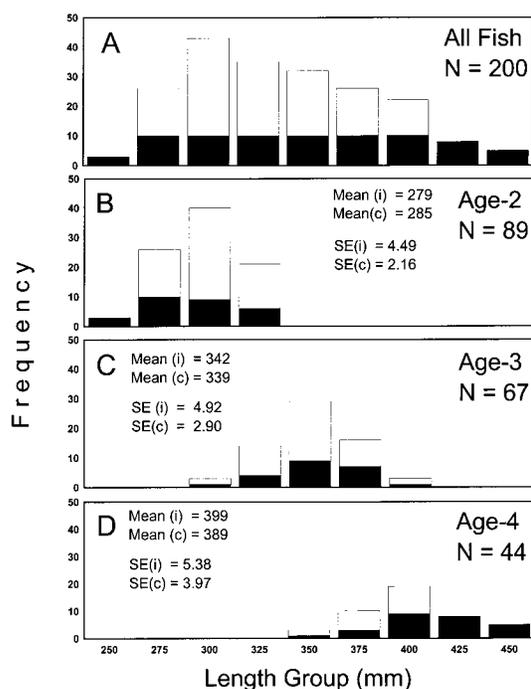


FIGURE 1.—(A) Length-frequency distribution for largemouth bass from Kentucky Lake, Tennessee, April 1994; the dark bars represent fish for which age was estimated from sagittal otoliths. The length distributions for (B) age-2, (C) age-3, and (D) age-4 fish were generated by applying an age-length key to panel A. Mean lengths and standard errors were calculated by using equations (1) and (2)—incorrect approach (i)—and equations (3) and (4a)—correct approach (c).

size-group of age-0 largemouth bass collected in a midsummer cove sample. The stratified subsample contained 27 fish from a “population” of 147 age-0 fish (all of which were measured for total length). Compared with the true mean (56 mm), the mean calculated from the subsample and using equation (1) (the incorrect approach) was severely biased because the frequency distribution was sharply skewed (Figure 2). The mean calculated from the subsample and using equation (3) (the correct approach) was nearly identical (55 mm) to the true mean. As expected, the variance calculated by using equation (2) (the wrong approach) was inflated relative to the correct sample variance calculated with equation (4a).

Computer software such as Fishcalc89 (developed by the Missouri Department of Conservation and available through the AFS Computer Users Section) can be used to estimate the mean length at age of a sample of fish for which the lengths are known, from a subsample of fish for which

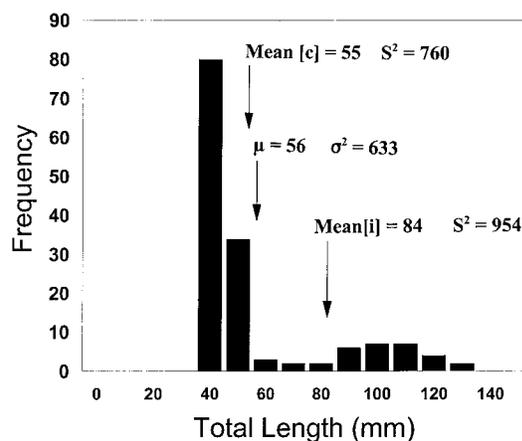


FIGURE 2.—Length-frequency distribution for 147 age-0 largemouth bass collected in Normandy Lake, Tennessee, 1994; mean length (μ) and variance (σ^2) for those fish are listed. Three fish were randomly selected from each 10-mm size-group, and the sample mean and variance were subsequently estimated by using equations (1) and (2)—incorrect approach (i)—and equations (3) and (4a)—correct approach (c).

lengths and ages are known. Fishcalc89 is a fairly easy program to master and, in our example, provided the same estimates of mean lengths at age and their variances as equations (3) and (4b) when we shortened our length-groups to 10-mm and used midpoints of each length-group in the calculations. Alternatively, equations (3) and (4) given above may be programmed into commercially available software to facilitate calculations.

The question of whether it is preferable to subsample a fixed number of fish from each length-group (i.e., fixed-age subsampling), or subsample in proportion to the number of fish in each length-group (i.e., random-age sampling), has been debated for decades. Kimura (1977) assessed the statistical validity of data generated using both approaches and concluded that random-age subsampling was superior to fixed-age subsampling. The drawback of that approach, as noted by Ketchen (1949), is that age proportions in the largest length groups, which could contain many age-groups, could be poorly estimated. When constructing catch curves to assess variable recruitment (e.g., Maceina 1997; Maceina and Bettoli 1998), good estimates of the abundance of older age-classes are critical. Therefore, it is important to oversample the largest length groups, each of which might contain several age-classes, through fixed-age subsampling. Ensuring that older, rarer fish are adequately represented in the subsample is also im-

portant when estimating maturity at length (Morgan and Hoenig 1997). Another concern with random-age subsampling is logistics; it is much easier for field crews to collect a fixed number of fish per length-group for age determination, especially when fish that are not killed can be released alive after routine measurements are recorded. Given that the examples of age subsampling found in the last two issues of the *North American Journal of Fisheries Management* all represented fixed-age subsampling, we assume that most biologists (at least, those who publish their findings) prefer that method. For those wishing to avoid any uncertainties involved with extrapolating ages from fish in a subsample, the simplest method of guaranteeing unbiased results is to estimate the age of all the fish collected in a large random sample.

Acknowledgments

Collection of the largemouth bass data used in the example was made possible by a grant from the Tennessee Valley Authority to the senior author. S. Hale, H. Schramm, S. Sammons, D. Isermann, and M. Maceina contributed thoughtful insights on this topic and facilitated the preparation of this manuscript. This manuscript benefited from the constructive comments offered by S. Cadrin and two anonymous reviewers.

References

- Allen, K. R. 1966. Determination of age distribution from age-length keys and length distributions. *Transactions of the American Fisheries Society* 95: 230–231.
- DeVries, D. R., and R. V. Frie. 1996. Determination of age and growth. Pages 483–512 in B. R. Murphy and D. W. Willis, editors. *Fisheries techniques*, 2nd edition. American Fisheries Society, Bethesda, Maryland.
- Ketchen, K. S. 1949. Stratified subsampling for determining age distributions. *Transactions of the American Fisheries Society* 79:205–212.
- Kimura, D. K. 1977. Statistical assessment of the age-length key. *Journal of the Fisheries Research Board of Canada* 34:317–324.
- Maceina, M. J. 1997. Simple application of using residuals from catch-curve regressions to assess year-class strength in fish. *Fisheries Research* 32:115–121.
- Maceina, M. J., and P. W. Bettoli. 1998. Variation in largemouth bass recruitment in four mainstem impoundments of the Tennessee River. *North American Journal of Fisheries Management* 18:998–1003.
- Morgan, M. J., and J. M. Hoenig. 1997. Estimating maturity-at-age from length-stratified sampling. *Journal of Northwest Atlantic Fisheries Science* 21:51–63.
- Ricker, W. E. 1975. Computation and interpretation of biological statistics of fish populations. *Bulletin* 191, Fisheries Research Board of Canada, Ottawa.
- Thompson, S. K. 1992. *Sampling*. Wiley, New York.
- Westrheim, S. J., and W. E. Ricker. 1977. Bias in using an age-length key to estimate age-frequency distributions. *Journal of the Fisheries Research Board of Canada* 35:184–189.