RESTRICTED

INTERNATIONAL COMMISSION FOR THE NORTHWEST ATLANTIC FISHERIES

Serial No.2237
(D.c.2)

ICNAF Res.Doc.69/73

ANNUAL MEETING - JUNE 1969

Some Remarks on the Behaviour of Growth and Mortality Estimates Based on Age-Length Keys

by William Knight
Fisheries Research Board of Canada
Biological Station, St. Andrews, N.B.
and University of New Brunswick
Fredericton, N.B.

"The purpose of computing is insight, not numbers." Richard Hamming, Numerical Methods for Scientists and Engineers.

1. Introduction

The first statistical problem with age-length data is how to estimate mortality, growth, year-class strength, etc. This paper is not concerned with that question. Once such estimates are found, "second order" questions come up: What are the properties of these estimates, and how can the data be collected to improve these properties

We need a different kind of answer to these "second order" questions. The answer to an estimation problem is a number, the estimate, but in the second case, "the purpose ... is insight, not numbers". Whereas an estimate must be accurate and may be complicated, the description of its properties may be rough but must be simple if it is to be generally useful.

A few examples of such "second order" questions:

(i) What lengths of fish should be selected for ageing by otolith or scale? Should length groups be sampled proportionately, by equal sample sizes from each length group, or some other allocation?

(ii) How large is the statistical sampling error? How large sample sizes are needed? What sort of significance tests should be used?

(iii) Sampling programs would be less expensive if age-length keys could be pooled. When can this be done? To what extent? What is the penalty of pooling when you shouldn't?

(iv) What happens if the age-length key and the length distribution come from different populations? (This question was suggested by W.E. Ricker.)

Any study of questions like these is immediately entangled in the basic paradox of age-length keys: Although it is the age distribution (especially mortality) and the conditional distributions of

In studying the properties of growth rate and mortality estimates, as opposed to actually calculating their numerical values, it is more important that the mathematics be simple than that it be accurate. Approximations are suggested for use with age-length keys with crudely linear growth curves, and these approximations applied to a few common questions concerning age-length keys.

length at fixed ages (especially the growth rate) which are wanted, it is the complements of these which are actually collected, the length distribution and the conditional distributions of age at fixed lengths. In such a pass, any simple formulae relating the directly measured quantities to those indirectly measured but more meaningful would be a boon, even were these formulae but crude approximations, useless to the person actually performing calculations to estimate growth, mortality, etc.

Thus we finally arrive at the main purpose of this paper: To devise formulae connecting the quantities measured to the quantities actually wanted, and explore these in terms of the questions (i) - (iv).

2. Basic Approximations

It is assumed that the growth curve is linear. Although for most data it can be verified that this is not the case, there are two justifications of such a simplification. First, linearity is often, perhaps even usually, correct as a first order description. Second is the remarkable success of crude linear approximations in varied applications.

For any given age-length distribution, the relation between the growth rate, G, the slope of the least squares line of length on age, and the slope of the least squares line of age on length which is denoted B, is

$$G = r^2 B^{-1} \qquad (2.1)$$

where r is the correlation coefficient. The contributions of the age-length key and the length distribution to r can be separated,

$$r = B^2/(B^2 + F) \qquad (2.2)$$

$$F = s_{t\cdot\ell}{}^2/s_\ell{}^2$$

where s is the standard deviation of the length distribution and $s_{t\cdot\ell}{}^2$ is the mean square deviation of age from the least square line of age on length. Thus the age-length key enters the linear growth rate through B and $s_{t\cdot\ell}$ and the length distribution through s_ℓ. This would be a complete solution were it not that B is quite independent of the length distribution only if the regression of age on length is linear. (The factor $s_{t\cdot\ell}$ also depends on the length distribution but as its influence is rather small, this is not likely to be important.) As it is, (2.1) is approximate, the degree of approximation depending on the linearity of the regression.

A relation between instantaneous mortality and growth is

$$Z = GZ' \qquad (2.3)$$

where Z' is something primarily determined by the length distribution. Ricker (1958) in Chapter 2, Section G, investigates taking Z' as the decay rate of the right limb of the length distribution as Z is the decay rate of the right limb of the age distribution. His conclusion that the approximation is poor need not deter us altogether as our demands are less stringent than his.

There is also a choice of Z' which makes (2.3) exact under the assumptions of a linear regression of length on time, and that all recruitment takes place at age t_0. The second assumption is no real restriction as it can be achieved by truncating the data after collection. The first assumption again enters only through the demand that the least squares line of length on age be independent of the length distribution. This definition of Z' is

$$Z' = 1/(\bar{\ell} - \ell_0) \qquad (2.4)$$

where $\bar{\ell}$ is the mean length and ℓ_0 is the recruitment length in the sense that it corresponds to t_0 on the least squares line. The derivation requires integration and is set out in chapter 7.

3. The Distribution of Information

Although the manner of distribution of information between the age-length key and the length distribution is easily worked out, and some aspects at least, published before (Gulland, 1956, p. 24), an explicit statement should be made. Most of the information about growth is found in the age-length key. Equation (2.3) tells us that the length distribution informs us primarily about the ratio of mortality to growth. The comparison of strengths of adjacent year-classes can be made on the basis of the age-length key alone; for separated year-classes, some kind of correction for mortality is needed.

If a von Bertalanffy curve is fitted, the distribution of information is rather interesting. In what should be (Knight, 1968) the usual situation, the asymptote, L_∞, is determined by the size of the largest fish; this must come from the length distribution. On the other hand, the linear properties of the curve come from the age-length key. A line is determined by its intercept and slope; with the von Bertalanffy curve the intercept is given by t_0, and the slope is approximately KL_∞. (The slope at t_0 is KL_∞ by a simple differentiation.) Thus t_0 and the product, KL_∞, are determined primarily by the age-length key, and L_∞ by the length distribution!

4. Some Questions Roughly Answered

We now return to the questions raised in the Introduction:

(i) In sampling otoliths or scales for age reading, should each length class be sampled proportionately, sampled equally, or some other allocation? As far as mortality and growth, but not year-class strength, are concerned, the important thing is the accuracy with which the regression of age on length is measured, this being the dominant term in (2.1) and thence by implication the important part of the age-length key appearing in (2.3). (Z' is a property of the length distribution only.) A regression line is best estimated by concentrating effort near the ends of the line. Indeed, in principle, the only reason for taking any points near the middle is to check the adequacy of the linearity assumption. On the other hand, for estimating relative year-class strength, about equal attention to all ages, and thence by the approximate linearity of the growth curve, to all lengths, is called for. These considerations pretty clearly point to a practice already in use: Take a complete sample of the largest length classes and the smallest; then take equal numbers from the rest.

(ii) What about confidence intervals and tests of significance for growth and mortality? For practical purposes it suffices to get an approximation to the variance, or what is as good, the coefficient of variation. Denoting by v(*) the coefficient of variation of *, the usual linearization technique plus some crude approximation yields

$$v(G) \doteq \sqrt{v(B)^2 + (GFv(F))^2} \qquad (4.1)$$

The derivation is relegated to a later section. For the coefficient of variation of Z the usual approximation is

$$v(Z) = \sqrt{v(Z')^2 + v(G)^2} \qquad (4.2)$$

(iii) When can age-length keys be pooled, and what is the penalty of pooling without justification? Since the age-length key's information is mostly about year-class strength and growth rate, it suffices that these be the same. Put more generally, hence vaguely, the age-length key describes some of the biological properties of the stock at some time, whereas such man-created things as gear selection and fishing intensity show up in the length distribution, thus it should be safe to mix age-length keys for the same stock over such a time interval as it remains stable.

The penalties for pooling unlike keys are these: First, the year-class strengths will be mixed. Second, any difference in the length distributions will be interpreted primarily as a difference in mortalities, even if really a difference between growth rates, for most of the information which could distinguish between different growth rates has been lost in the pooling.

(iv) What if the age-length key and the length distribution actually come from different populations? Upon considering section 3, we find that some information can be salvaged for the population from which the age-length key was drawn, but little from the other. The growth rate, and the relative strengths of nearby year-classes, are relatively insensitive to the length distribution. Except in unusual cases, no information about mortality is available.

5. Sampling Error of the Primary Quantities

Formulae (4.1) and (4.2) give rough estimates of the error of the secondary quantities on the basis of estimates of the error of the primary quantities. There remains the need of estimating the errors of the primary quantities, a considerable problem and one outside the scope of this paper. To illustrate some of the considerations involved, I remark on an attempt which failed, an endeavour to assign error bounds to the growth rate and mortality rate derived from a routine age-length key for cod, commercially caught by otter trawl in ICNAF Division 4T in 1966, extracted from the files of the St. Andrews Biological Station of the Fisheries Research Board of Canada. The reader can skip the rest of this section with no loss of continuity.

The values of G and F are readily calculated while running textbook linear regressions of length on age and age on length.

The estimation of the coefficient of variation of B cannot be done with the distribution in its final form for presentation; the numbers do not represent real fish, but elaborate weighted combinations. I first considered the possibility of using the unweighted age-length key which, though not made up, is readily calculated from worksheets available. The regression of age on length could be run on the unweighted key yielding a regression equation of no interest whatever, and an error estimate of the slope of regression whose calculation is the purpose of the exercise. Unfortunately, the textbook error term assumes that the 28 samples whose combination make up the unweighted age-length key are homogeneous, an assumption hardly tenable after the report of Dickie and Paloheimo (1965), and a similar study by the author (unpublished).

The situation for length distributions is much worse than for the age-length keys. Dickie and Paloheimo in the study noted before found far greater heterogeneity among length distributions than conditional age-at-length distributions (as measured by the likelihood ratio statistic for contingency tables). Moreover, while methods for estimating the error of a regression are commonly found in textbooks regardless of their applicability to the present situation, estimates of the error of the estimated standard deviation of the badly skewed distribution, such as the typical length distribution, and estimates of the error of the decay rate of the right limb of the length distribution are not easily found in textbooks or anywhere else.

6. Derivations of Equation (2.2)

Equation (2.2) is a descriptive identity and holds whether or not growth is linear although it will not be relevant for extreme non-linearity. It is derived from the following well known formulae:

$$GB = r^2 \qquad (6.1)$$

$$B\, s_\ell^2 = \text{mean of products about mean} = G\, s_t^2 \qquad (6.2)$$

$$r^2 = 1 - s_{t\cdot\ell}^2/s_t^2 \qquad (6.3)$$

where s_t is the age standard deviation; (6.1) and (6.3) can be found in Steel and Torrie (1960, p. 188); (6.2) is immediate from the definitions of B and G. Solving (6.2) for s_t^2,

$$s_t^2 = B\, s_\ell^2/G \qquad (6.4)$$

and substituting (6.1) and (6.4) into (6.3) to eliminate r and s_t^2 respectively,

$$G = 1 - G\, s_{t\cdot\ell}^2/B\, s_t^2 \qquad (6.5)$$

which is solved for G yielding (2.1).

7. Derivation of (2.4)

Equation (2.4) rests on the fact that the reciprocal of average age is an estimate of mortality. That this is true for the case of constant mortality is a simple exercise with the exponential distribution. For any distribution the following holds:

Let Z(t) be the mortality at time t, that is

$$(df(t)/dt)/f(t) \qquad (7.1)$$

where f(t) denotes the age distribution density function. It is usually the mortality for large t in which we are interested; we express this by taking as our overall figure for mortality the weighted average, where the weight is age over the recruitment age, thus

$$Z = \int_{t_o}^{\infty} (t-t_o)Z(t)f(t)dt \Big/ \int_{t_o}^{\infty} (t-t_o)f(t)dt \qquad (7.2)$$

The numerator is

$$\int_{t_o}^{\infty} (t-t_o)f'(t)dt = \int_{t_o}^{\infty} f(t)dt \qquad (7.3)$$

by integration by parts. We then have weighted average

$$Z(t) = \frac{1}{\bar{t} - t_o} = \frac{G}{\bar{\ell} - \ell_o} \qquad (7.4)$$

where $\bar{t}$ and $\bar{\ell}$ are respectively the average age and length of fish of age t_o or more. The right equality is merely a restatement of the equation for the least squares line, we have (2.4). Note however that G here is the slope of the line fitted to fish of age t_o or more, not the entire population, hence the necessity of the assumption that G is stable.

8. Derivation of (4.1)

Equation (4.1) follows from the usual approximation with partial derivatives, plus a crude approximation. The partial of G with respect to B:

$$\frac{\partial G}{\partial B} = \frac{\partial}{\partial B}\left[\frac{B}{B^2+F}\right]$$

$$= G\left(\frac{1}{B} - \frac{2B}{B^2+F}\right)$$

$$= \frac{G}{B}(1-2GB)$$

$$= \frac{G}{B}(1-2r^2) \qquad (8.1)$$

In an approximate formula, subtractions can be downright dangerous, but noting that $|1-2r^2| \leq 1$, the approximation below is conservative.

$$\frac{\partial G}{\partial B} \leq \frac{G}{B} \qquad (8.2)$$

The partial of G with respect to F

$$\frac{\partial G}{\partial F} = -\frac{G}{B^2+F} = -\frac{G^2}{B} \qquad (8.3)$$

Taking V(*) to mean the variance of *, the usual linear approximation,

$$V(G) = \left(\frac{\partial G}{\partial B}\right)^2 V(B) + \left(\frac{\partial G}{\partial F}\right)^2 V(F) \qquad (8.4)$$

leads to,

$$\frac{V(G)}{G^2} = \frac{V(B)}{B^2} + (GF)^2 \frac{V(F)}{F^2} \qquad (8.5)$$

References

Dickie, L.M., and J.E. Paloheimo. 1965. Heterogeneity among samples of the length and age compositions of commercial groundfish landings. ICNAF Research Bulletin No. 2: 48-52.

Gulland, J.A. 1956. The study of fish populations by the analysis of commercial catches. Rapports et Procès-Verbaux des Réunions, Conseil Permanent International pour l'Exploration de la Mer 140(1): 21-27.

Knight, William. 1968. Asymptotic growth: an example of nonsense disguised as mathematics. J. Fish. Res. Bd. Canada 25(6): 1303-1307.

Knight, William. 1962. Groundfish sampling studies. Unpublished manuscript. Orig. MS 940 of the Biological Station of the Fish. Res. Bd. Canada, St. Andrews, New Brunswick.

Ricker, W.E. 1958. Handbook of computations for biological statistics of fish populations. Queen's Printer, Ottawa.

Steel, Robert G.D., and James H. Torrie. 1960. Principles and procedures of statistics. McGraw Hill.

Watson, G.S., and M.R. Leadbetter. 1964. Hazard analysis. I. *Biometrika* 51: 175-184.