On Length-Weight Relationships: Part II: Computing Mean Weights from Length Statistics

JAN E. BEYER
Danish Institute for Fisheries and Marine Research
Charlottenlund Castle, DK-2920 Charlottenlund
Denmark

Abstract

Assuming a known length-weight relationship, the formula for calculating the mean weight from grouped length statistics is recapitulated. Then a simple analytical method for calculating mean weight from basic length statistics is developed which applies when lengths are approximately log-normally distributed or when normally distributed with a coefficient of variation (sd/mean) less than ca. 1/3. The method can be applied to other length distributions as well provided that the coefficient of variation is small.

Introduction

In fish stock assessment, biologists frequently need to compute mean population characteristics based on measurements of individual fish. However, "average" fish do not exist and bias is often introduced in the computation of averages.

The main problem here is caused by the nonlinearity of the length-weight relationship. As an example, suppose the mean weight of a group of individual fish of average length 25 cm is required. Assuming, for simplicity's sake a length-weight relationship with a = 0.01 (gcm^{-3}), and b = 3, the weight corresponding to the mean length of the fish sampled becomes

$$0.01 \text{ g cm}^{-3} \cdot 25^3 \text{ cm}^{-3} = 156.25 \text{ g}$$

Now, the mean weight of a fish of average length is usually not a good estimate of the mean weight of the fish in the sample. In general, the greater the variability in length, the more the true mean weight will deviate from 156.25 g.

Ricker (1975, p. 211) states that the true mean weight is always greater than the weight corresponding to the mean length of the fish sampled. It will be shown below that this is not necessarily true. Also, we shall examine the empirical length distribution for any sample of fish before turning to the situation in which a quick, analytically-derived method for computing the mean weight of a group of fish is required. Box 1 summarizes the notation used in this paper.

Box 1

NOTATION:

L	body length
W	body weight
a	constant of proportionality in the L-W relationship (equals the condition factor when b=3)
b	exponent (power) in the allometric L-W relationship
N	total number of fish in sample (= $N_1 + N_2 + N_3 + ... + N_m$)
j	fish number index (1,2,3,...,N)
i	length class index (1,2,3,...,m)
h	class length (= $L_{i+1} - L_i$)
$\bar{L}_i$	class midpoint ($L_i + h/2 = L_{i+1} - h/2$)
$\bar{X}$	= E(X) expected value or mean of the variable X
s	standard deviation in the length distribution
s^2	variance in the length distribution: (see equation (8))
C	coefficient of variation in the length distribution (see equation (6))
G	coefficient of skewness in the length distribution (i.e., third central moment relative to standard deviation cubed)
α	first parameter in log-normal length distribution, i.e., mean of ln (L)
β	second parameter in log-normal length distribution, i.e. standard deviation of ln (L).

Allometric Growth and Uniform Length Distribution

The weight of an individual fish can be obtained from length using the empirical relationship:

$$W = a \cdot L^b \quad ...1)$$

The mean weight of N fish can then be computed as follows,

$$\overline{W} = N^{-1} \cdot \sum_{j=1}^{N} W_j = a \ N^{-1} \cdot \sum_{j=1}^{N} L_j^b \quad ...2)$$

provided that the L_j's, the individual lengths of all N fish are known. Usually, fish are grouped in length classes. In this case, the individual weights in equation (2) are replaced by group weights, i.e.,

$$\overline{W} = N^{-1} \cdot \sum_{i=1}^{m} N_i \overline{W}_i \quad ...3)$$

where $\overline{W}_i$ is the mean weight of the N_i uniformly distributed fish in length class i. Beyer (1987) showed that

$$\overline{W}_i = a \cdot (b+1)^{-1} \cdot h^{-1} \cdot (L_{i+1}^{b+1} - L_i^{b+1}) \quad ...4)$$

where L_i designates the lower class limit of length class i and h the class width. Then, inserting equation (4) in equation (3) gives

$$\overline{W} = a \cdot (b+1)^{-1} \cdot h^{-1} \cdot N^{-1} \cdot \sum_{i=1}^{m} N_i (L_{i+1}^{b=1} - L_i^{b+1}) \quad ...5)$$

These equations do not utilize information that may be available with respect to, say, the shape of the histogram of lengths, i.e., the empirical distribution or the type and parameters of the length distribution of the population from which the length-frequency sample is supposed to be randomly drawn.

The usual measure of relative dispersion of measurements is the coefficient of variation (C)[a] defined as the standard deviation divided by the mean, here estimated by

$$C = S/\overline{L} \quad ...6)$$

where the mean length ($\overline{L}$) is estimated by

$$\overline{L} = N^{-1} \sum_{j=1}^{N} L_j = N^{-1} \sum_{i=1}^{m} N_i \cdot \overline{L}_i \quad ...7)$$

and the variance usually is obtained as

$$S_2 = (N-1)^{-1} \sum_{j=1}^{N} (L_j - \overline{L})^2 = (N-1)^{-1} \sum_{i=1}^{m} N_i (\overline{L}_i - \overline{L})^2 \quad ...8)$$

In the grouped empirical distribution, mean and variance are usually computed by considering the lengths of all N_i fish in the i'th class to be equal to the class midlength.

Isometric Growth and Symmetrical Length Distribution

Assuming the weight of a fish is given by aL^3, the mean weight in the population becomes

$$\overline{W} = aE(L^3) \quad ...9)$$

where $E(L^3)$ denotes the third moment in the length distribution. Since we want to characterize the length distribution by its central moments (i.e., the moments about $\overline{L}=E(L)$, the mean length) we rewrite L^3 as

$$L^3 = \overline{L}^3 + 3\overline{L}^2(L-\overline{L}) + 3\overline{L}(L-\overline{L})^2 + (L-\overline{L})^3 \quad ...10)$$

Taking the expectation of both sides of this equation, and considering that $E(L - \overline{L}) = 0$ and $E((L - \overline{L})^2) = S^2$, we obtain,

$$E(L^3) = \overline{L}^3[1 + 3(S/\overline{L})^2 + (E((L-\overline{L})^3))/\overline{L}^3] \quad ...11)$$

where the last term, the third central moment relative to mean length cubed may be expressed, alternatively, as $C^3 \cdot G$, i.e., C cubed multiplied by the coefficient of skewness.

Equation (11) is valid for any type of length distribution. In the special case of a symmetrical distribution (i.e., skewness is zero) we obtain, multiplying equation (11) by a,

$$\overline{W} = a\overline{L}^3[1 + 3C^2];\ C = S/\overline{L} \quad ...12)$$

Box 2

A normally distributed length group has been identified with L = 25 cm and S = 5 cm, i.e. with a coefficient of variation of C = 5/25 or 20%. Our estimate of the mean weight for fish belonging to this (age) group becomes, using equation (12) and assuming a=0.01 in the isometric L-W relationship,

$$\overline{W} = 156.25 \cdot (1 + 3 \cdot 0.04) = 175 \text{ g}$$

In this case, the weight at mean length must be raised by 12% to estimate the correct mean weight for the population.

Allometric Growth and Log-Normal Length Distributions

If the logarithms of individual lengths are normally distributed with mean α and variance β^2 then the lengths are log-normal distributed with mean and variance:

$$\overline{L}^2 = \exp(\alpha + 1/2\ \beta^2) \quad ...13)$$

$$S^2 = \overline{L}^2\ (\exp(\beta^2) - 1) \quad ...14)$$

Given

$$\ln W = \ln a + b \ln L \quad ...15)$$

[a]This is normally abbreviated as "C.V." and expressed as percentage, i.e., $C.V. = s.d. \cdot 100/\overline{X}$.

it follows that the logarithm of weight also is normally distributed with mean ln a + bα and variance $b^2\beta^2$. Using these parameters for weight instead of α and β^2 for length as in equation (13), we obtain directly the mean weight:

$$\bar{W} = \exp(\ln a + b\alpha + 1/2\, b^2\beta^2)$$
$$= a \cdot \exp(\, b(\alpha + 1/2\, \beta^2) + 1/2\, b(b-1)\beta^2)$$

or, inserting equations (13) and (14) on the form $\exp(\beta^2) = 1 + C^2$

$$\bar{W} = a\, \bar{L}^b\, (1 + C^2)^{1/2\, b(b-1)};\ C = S/\bar{L} \qquad ...16)$$

This expression is also valid for a number of distributions apart from the log-normal if C<<1.

Box 3

If lengths are log-normally distributed, and L = 25 cm, and C = 0.2 we have from equation (16)

$$\bar{W} = a\, 25^b\, (1 + 0.04)^{1/2\, b(b-1)}$$

In the special case of isometric growth (Box 2) we obtain

$$\bar{W} = 156.25 \cdot (1 + 0.04)^3 = 175.76 \text{ g.}$$

Due to positive skewness, a log-normal length distribution will always give a slightly higher mean weight than a normal distribution with the same mean and C. In the present example, with a small C, the percentage increase is less than 1/2% (see also Box 2). This illustrates the general validity of equation (16).

From the log-normal variance, equation (14), the same procedure, substituting the length parameter β with the weight parameter bβ, yields the variance of weight:

$$S_w^2 = \bar{W}^2\, [(1 + C^2)^{b^2} - 1]\ ;\ C = S/\bar{L} \qquad ...17)$$

These expressions for the mean and variance in the weight distribution are exact for any L-W relationship and thus give the weight characteristics as a function of the log-normal length statistics.

Allometric Growth and Normal Length Distribution

In practical work with log-normal (length) distributions, it is generally accepted that (both log-length and) length is normally distributed as long as the C is less than ca. 1/3 (Hald 1952, p. 164). For such cases equations (16) and (17) can be used even for normal length distributions.

A summary of the use of the three hitherto presented equations for calculations of mean weight is presented in Fig. 1.

First Order Approximations

In cases where $C^2 << 1$ a first order approximation to equation (16) can be used as a good estimate of the mean weight. Using $(1 + x)^{power} \simeq 1 + x \cdot power$, where x is much smaller than unity (1), we have

$$\bar{W} = a\, \bar{L}^b\, (1 + 1/2\, b\, (b-1)\, C^2);\ C \leq 1/3 \qquad ...18)$$

With the same type of approximation we obtain the standard deviation of weight from equation (17),

$$S_w = b \cdot C \cdot \bar{W}\ ;\ C \leq t1/3 \qquad ...19)$$

If the length sample consists of N fish, the standard deviation of $\bar{W}$, the estimator of mean weight, or the standard error becomes

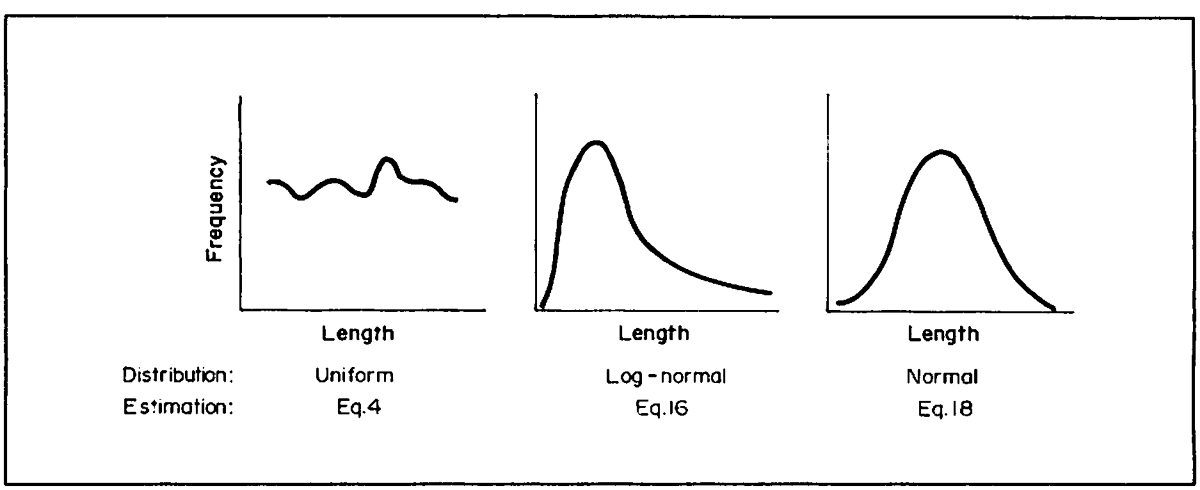

Fig. 1. Summary of three main length distributions and associated equations for calculation of mean weights. If c = s/L < 1/3, equation (16) can also be used for a number of other length distributions, including symmetrical ones. Equation (18) gives a simple approximation to equation (16), applicable for most distributions with c < 1/3.

$$S_{\overline{W}} = S_w/\sqrt{N} = b\ C\ W\ /\sqrt{N} \qquad ...20)$$

These expressions represent our final results. Equation (18) gives the mean weight as a function of the length statistics. It works for normal and log-normal length distributions if C is about equal to or less than 1/3, and for other distributions as well if C is small.

Discussion

Ricker's (1975) statement that the true mean weight is always greater than the weight at mean length is not necessarily correct. For example, in the case of isometric growth, the two terms in equation (11) for the second and third central moments may cancel out. This happens

$$\text{IF } G \cdot C = -3 \text{ THEN } E(L^3) = \bar{L}^3$$

That is, the product of the coefficients of skewness and variation for length equals -3 in which case $\overline{W} = q\bar{L}^3$ is exact. If $G{\cdot}C < -3$ then $W < qL^3$. Such situations where the length distribution is considerably skewed to the left (i.e., negative skewness) are sometimes seen in pooled samples from surveys, e.g., of larval fish.

For most practical assessments, however, Ricker's statement is correct. Most distributions can be described adequately by normal distributions (i.e., G=0), or by log-normal distributions (i.e., distributions with positive skewness).

For any L-W relationship, we have a simple analytical expression in equation (18) for the mean weight in an approximative normal length distribution provided C is less than about 1/3,

$$\overline{W} = a\ \bar{L}^b\ (1 + \Delta)\ ;\ \Delta = 1/2\ b(b-1)\ C^2;\ C \le 1/3$$

For a given value of the power, b, and a certain mean length, $\bar{L}$, the relative error or bias, Δ, by using the weight at mean length is directly proportional to the variance in the length distribution. Thus, a doubling, say, in standard deviation increases the relative error by a factor of 4.

The mean weight in equation (5) is obtained by integrating the length histogram. The basic element is equation (4) for the mean weight of fish uniformly distributed over a length class (Beyer 1987). However, based on equation (4) it is difficult to distinguish between the effects of changes in the power (b) and in the class length (h) on the mean weight. It may therefore be noted that equation (18) actually appears through a Taylor expansion of equation (4) to the third order,

$$(1+x)^{b+1} = 1 + (b+1)x + 1/2b(b+1)x^2 + 1/6(b-1)b(b+1)x^3$$

where $x = \pm h/(2L_i)$ is small. Note that the variance in the uniform distribution is $h^2/12$. As a consequence, equation (18) may in practice be used instead of equation (4). This is demonstrated in Box 4. Likewise, equation (16) may be used instead of equation (9) due to its general validity.

Box 4

Equation (18) may also be used for computation of the mean weight, for example, in a uniformly distributed length class. We may use Example 2 in Beyer (1987) as an illustration. The class considered is (10,11), i.e. $\bar{L}$=10.5 and h=1 and, the L-W relationship is specified by a=0.009 and b=3.193. Using equation (4) gives $\overline{W}$ = 16.445525 as the exact result. The variance of a uniform distribution is $S^2 = h^2/12$. From equation (18)

$$\overline{W} = 0.009 \cdot 10.5^{3.193} \cdot (1+1/2 \cdot 3.193 \cdot 2.193/(12 \cdot 10.5^2))$$

or 16.445524 reflecting a 7-digit precision. Such a good approximation is not achieved because the length distribution is uniform but mainly because C is very small (C=0.0275, i.e., less than 3%).

If length is normally distributed then weight will not be so. The distribution of weight will have positive skewness and show resemblance to a log-normal distribution (because continuing to multiply identical distributions produces a log-normal distribution). If length is log-normally distributed and growth is allometric, weight will also be log-normally distributed.

Returning to the characteristics of a log-normal distribution we note that equation (16), divided by a, gives a simple and exact expression for any moment of a log-normal distribution. Thus, if a (stochastic) variable X belongs to Log-N (α,β^2), denoting the mean by $E(X) = \mu$ and the variance by $Var(X) = \sigma^2$, we have,

$$E(X^b) = \exp(b(\alpha + 1/2\ b\beta^2)) = \mu^b\ [1+(\sigma/\mu)^2]^{1/2 b(b-1)};$$

here μ and σ^2 are the true (but usually unknown) mean and variance which are estimated by X and S_x^2, respectively, (as done throughout the text in order to simplify the notation). When raw data are grouped, the grouped variance (equation (8) right) is (on the average) equal to the ungrouped variance (equation (8) left) plus $h^2/12$ where h is the class length. The smaller the class length the smaller the loss of information. With respect to the reliability of statistical tests based on grouped data, the safe maximum class length is determined by the variance. A useful rule is that the class length should be smaller than the half standard deviation (i.e., $h^2 < S^2/4$).

Conclusions

If the allometric L-W relationship, $W = aL^b$, is known, then the mean weight of the fish in a sample can be calculated quickly by a simple formula using length statistics only. Due to the nonlinearity of the L-W relationship, the mean weight depends on the relative variability (i.e., dispersion) of lengths. The key parameter is therefore the coefficient of length variation, C.

In most length-based assessment work C < 1/3, in which case the mean weight may be obtained from equation (16) for most distributions.

The more the length distribution resembles a symmetrical distribution (such as a normal or a uniform distribution) the less critical the assumption of C < 1/3 becomes.

It is the power, b, which expresses the relationship between variability in length and weight. For approximative symmetrical distributions, we have (cf equation (9))

$$\text{coefficient of W-variation} \approx b \cdot \left[\text{coefficient of L-variation}\right]$$

This relation is also useful in expressing the precision obtained for the mean weight based on a certain number of length measurements (and assuming an exact L-W relationship)

$$\text{coefficient of W-variation} \approx b \cdot n^{1/2} \cdot \left[\text{coefficient of L-variation}\right]$$

where n is the number of length measurements. Thus, if $n = b^2$ or ca 10 then the relative error on the mean weight is about equal to the coefficient of variation in length. Box 5 deals with such a case in which $C_w = C = 0.10$ or 10% implying that the true mean weight is within ±10% of the estimated mean weight with probability 68%.

The problem discussed here is not specific to tropical fish stock assessment but I hope that this note illustrates the usefulness of obtaining simple expressions on a closed analytical form by thinking in terms of probability distributions and applying various type of approximations.

Box 5

Pienaar and Ricker (1968) present a theoretical example with a=0.5 and b=3.3 for a sample of 16 fish drawn from a N(100,100), normal length distribution. The weight at mean length 100 becomes $0.5 \cdot 100^{3.3}$ or 1990536 which is an underestimate (-5%) of the mean weight estimated from the sample (2090820). Noting that S^2 = 133.33 (or S=11.547 based on the 16 length observations), the estimated C (11.547/100) squared is equal to 0.013333. As this is less than 1/3, equation (18) yields

$$\overline{W} = 1990536 \cdot (1+1/2\ \ 3.3 \cdot 2.3 \cdot 0.013333) = 2091254$$

or only 0.45% more than 2081800, the mean weight estimated from the sample. From equation (20) with N=16 we obtain the standard error

$$S_{\overline{W}} = 3.3 \cdot 0.1155 \cdot 2091000\ /\ 4 \approx 200000$$

This standard deviation measures dispersion about the true mean weight in the following (general) way: The interval ($W-S_w$, $W+S_w$) or (1891000, 2291000) contains the true mean weight with a probability of 68%. In this simulation example we know the true mean weight because the true variance of length is 100, i.e., C=0.01. Hence, from equation (18), the true mean weight in the population becomes $\overline{W}_{true}$ = 2066000. This means that the interval ($\overline{W}_{true} - S_{\overline{W}}$, $\overline{W}_{true} + S_{\overline{W}}$) or (1866000, 2266000) contains 68% of the $\overline{W}$ observations. The sample in Pienaar and Ricker produces only one observation of $\overline{W}$ which is 2081000 as noted above. If 100 simulations were carried out (sampling 16 fish from the N(100,100) length distribution in each case) we would get 100 observations of $\overline{W}$ and would expect 68 of these to be contained in the interval (1866000, 2266000). Of course, with 100 times as many observations as in this example we would be able to reduce the standard error and, hence, the confidence intervals by a factor of 10. If the intervals are constructed with $2S_{\overline{W}}$ instead of $S_{\overline{W}}$, the interpretation is the same, but with 95% probability instead of 68%.

Acknowledgements

This work has mainly been carried out because I was foolish enough to call my 1987 contribution to *Fishbyte* "Part I". Daniel Pauly has been chasing me and "Part II" since that time (using an impressive amount of argumentation). I must apologize to the readers of *Fishbyte* for my late call response. I'm not sure it is worthwhile to thank Daniel for indirectly generating the present note but I do think this story indicates a much more important acknowledgement: the tremendous amount of work, persistency and continuity Daniel Pauly has put into making *Fishbyte* a success.

References

Beyer, J.E. 1987. On length-weight relationships. Part I: Computing the mean weight of the fish in a given length class. Fishbyte 5(1): 11-13.

Hald, A. 1952. Statistical theory with engineering applications. Wiley and Sons, Inc., New York. 783 p.

Pienaar, L.V. and W.E. Ricker. 1968. Estimating mean weight from length statistics. J. Fish. Res. Board Can. 25: 2743-2747.

Ricker, W.E.. 1975. Computation and interpretation of biological statistics of fish populations. Bull. Fish. Res. Board Can. 25, No. 191.