This paper not to be cited without prior reference to the authors

International Council for the Exploration of the Sea

I.C.E.S. C.M. 1983/D:23
Statistics Cttee, Ref. Demersal Cttee, Pelagic Cttee, Baltic Fish Cttee

USING MEAN AGE, MEAN LENGTH AND MEDIAN LENGTH DATA TO ESTIMATE THE TOTAL MORTALITY RATE

John M. Hoenig[1]
Graduate School of Oceanography

William D. Lawing
Departments of Industrial Engineering and Computer Science and Statistics

and

Nadine A. Hoenig
Department of Computer Science and Statistics
University of Rhode Island
Kingston, RI 02881, U.S.A.

Abstract

Estimators of the total instantaneous mortality rate, Z, based on mean age or mean length data are now becoming widely established. The statistical properties of these estimators are examined and it is concluded that one estimator should no longer be used. A new estimator, based on the median length, is proposed as a potentially robust alternative. The mortality estimates are shown to be approximately normally distributed by Monte Carlo simulation methods.

Résumé

Des estimateurs du taux total de mortalité instantanée, Z, construits à partir de données d'age moyen ou de longueur moyenne sont maintenant employés fréquemment. Les propriétés statistiques de ces estimateurs sont etudiées et il est conclu qu'un de ces estimateurs ne devrait plus être employé. Un estimateur noveau, dérivé à partir de l'age median est proposé comme étant une alternative potentiellement robuste. Il est démontré par des méthodes de simulation de Monte-Carlo que les estimateurs de mortalité sont distribués approximativement suivant une courbe normale.

[1] Present address: Minnesota Department of Natural Resources, Division of Fisheries and Wildlife, Box 25 Centennial Office Bldg., St. Paul, MN 55155, U.S.A.

Simple methods of estimating mortality rates, utilizing a minimum amount of data, are required for management of new or previously uncontrolled fisheries and for management of low valued "trash" fisheries. The need is particularly acute for tropical fisheries management (Saila and Roedel, 1981).

Beverton and Holt (1956) developed estimators for the total instantaneous mortality rate, Z, based on mean age and on mean length. They assumed an exponential mortality model which implies that reproduction occurs continuously at constant level. Holt (1965) developed an estimator based on mean age for the case where the upper tail of the age distribution is truncated.

Ssentongo and Larkin (1973) erroneously corrected the bias in Beverton and Holt's (1956) estimator based on mean age and presented another estimator which assumed a geometric age distribution, i.e. that assumed annual reproduction. They also derived a new estimator based on the mean length for the exponential model. However, in practice, Ssentongo and Larkin's estimator based on mean length has been found to give substantially higher estimates than the estimator of Beverton and Holt (Pauly, 1980; Marti, 1982; Sainsbury, 1982).

In this paper, the biases of the above estimators are examined and corrected where possible. Ssentongo and Larkin's (1973) estimator based on mean length is shown to have properties which make its utility doubtful. A new estimator based on the median length is proposed as a potentially robust alternative to Beverton and Holt's (1956) estimator based on mean length. Monte Carlo simulation methods are used to establish empirically the normality of the distribution of mortality estimates.

ESTIMATORS BASED ON MEAN AGE

Continuous Reproduction

Assume that the life expectancy of an animal follows an exponential distribution so that the probability density function (pdf) is given by

$$f(t) = Ze^{-Z(t-t_c)}, \quad t > t_c$$
$$= 0 \quad , \text{ elsewhere} \tag{1}$$

where Z is the instantaneous mortality rate (hazard rate), t is the random variable for age and t_c is the age of first capture (location parameter). Let $\bar{t}$ be the mean age in a random sample of size n from such a population. The model implies that reproduction occurs continuously at constant level. The maximum likelihood estimator of Z (and the estimator obtained by the method of moments) is given by

$$\hat{Z} = \frac{1}{\bar{t} - t_c} \quad . \tag{2}$$

Beverton and Holt (1956) arrived at this estimator deterministically. Ssentongo and Larkin (1973) pointed out that the estimator is biased but incorrectly stated that multiplying the estimator in (2) by the factor n/(n+1) would correct the bias.

To find the expected value of the estimator in (2) it is convenient to first perform a change of variable by letting $x = t - t_c$. Then Σx is distributed as a gamma variate with parameters Z and n, i.e.

$$\sum_{i=1}^{n} x_i \sim G(Z,n) \ .$$

Letting y equal the sum of the x's gives

$$E(\hat{Z}) = E(\frac{n}{y}) = nE(\frac{1}{y})$$

where the E represents the expected value operator. The expected value of 1/y is by definition

$$E(\frac{1}{y}) = \int_0^\infty \frac{1}{y} g(y)\, dy = \frac{Z}{n-1}$$

where g(y) is the pdf of the gamma distribution. Hence, the expected value of $\hat{Z}$ is

$$E(\hat{Z}) = \frac{n}{n-1} Z$$

and an unbiased estimator is

$$\hat{Z} = \frac{n-1}{n} \cdot \frac{1}{\bar{t}-t_c} \qquad (3)$$

(For large n, the bias correction can be ignored.)

The variance of (3) is given by

$$V(\hat{Z}) = \frac{Z^2}{n-2} \ . \qquad (4)$$

The variance of $\hat{Z}$ can be estimated by substituting the sample estimate of Z (from (3)) into equation (4) to give

$$\widehat{V(\hat{Z})} = \frac{(n-1)^2}{n^2(n-2)(\bar{t}-t_c)^2}$$

but this is a biased estimator. An unbiased estimator of the variance of $\hat{Z}$ is

$$\widehat{V(\hat{Z})} = \frac{n-1}{n^2} \cdot \frac{1}{(\bar{t}-t_c)^2} \qquad (5)$$

which again differs slightly from the results of Ssentongo and Larkin.

Since equation (3) is a one-to-one function of the maximum likelihood estimator given by (2), it is itself a maximum likelihood estimator and thus is asymptotically of minimum variance and normally distributed.

In some cases, animals above a certain age t_u occur with different frequency than is consistent with the exponential model. This may be due to a change in natural mortality rate with age, gear selectivity, or segregation of animals by age. Holt (1965, eqn. (10)) derived the following deterministic expression for mean age when the youngest and oldest ages are t_c and t_u, respectively[1]:

$$_c\bar{t}_u = \frac{t_c - t_u e^{-Z(t_u - t_c)}}{1 - e^{-Z(t_u - t_c)}} + \frac{1}{Z} .$$

Iterative solution of the equation for Z provides an estimate by the method of moments.

An alternative approach to estimating Z is the maximum likelihood method. The probability density function of a doubly truncated exponential distribution is given by

$$f(t) = \frac{Z e^{-Z(t_u - t_c)}}{1 - e^{-Z(t_u - t_c)}} , \; t_c < t < t_u .$$

If t_c and t_u are known, the maximum likelihood estimate of Z is the solution of

$$\frac{\bar{t} - t_c}{t_u - t_c} = \frac{1}{\hat{Z}(t_u - t_c)} - \frac{1}{e^{\hat{Z}(t_u - t_c)} - 1}$$

for $\bar{t} - t_c < 0.5(t_u - t_c)$. If $\bar{t} - t_c > 0.5(t_u - t_c)$ then $\hat{Z}$ is zero indicating that the truncated exponential model may be inappropriate.

The large sample formula for the variance of the maximum likelihood estimator is

$$Var(\hat{Z}) = (1/n)[Z^{-2} - (t_u - t_c)^2 e^{-Z(t_u-t_c)}/(1 - e^{-Z(t_u-t_c)})^2]^{-1} .$$

The reader is referred to Johnson and Kotz (1970 pp. 223-4) for details and references. The maximum likelihood estimator has desirable asymptotic qualities such as minimum variance, unbiasedness and normality.

---

1) Note that in Holt's derivation, the t_c added to the integral was an error which did not carry over to the final result (his eqn. 10).

Annual Reproduction

When reproduction occurs at discrete intervals (e.g. annually) and age is measured to the nearest (whole) interval, the population age structure can be represented by a geometric model. Thus, the number at any age t is given by

$$N_t = N_o e^{-Z(t-t_c)} , \quad t = t_c, t_{c+1}, \ldots$$

where N_o is the number at age t_c. From this, Ssentongo and Larkin (1973) derived the following estimator using the method of moments

$$\hat{Z} = \ell n\left(\frac{1+\bar{t}-t_c}{\bar{t} - t_c}\right) . \tag{6}$$

They recognized that this estimator has a positive bias and proposed the following modification which they claimed is unbiased:

$$\hat{Z} = \ell n\left(\frac{1+\bar{t}-t_c}{\bar{t} - t_c} \cdot \frac{n}{n+1}\right) . \tag{7}$$

However, Chapman and Robson (1960, p. 367) showed that there is no unbiased estimator of Z for the geometric distribution. They proposed the following nearly unbiased estimator for the one-parameter case

$$\hat{Z} = \ell n\left(\frac{1+\bar{t} - \frac{1}{n}}{\bar{t}}\right) - \frac{(n-1)(n-2)}{n(n\bar{t}+1)(n+n\bar{t}-1)} \tag{8}$$

where the first term is the negative of the logarithm of their unbiased, minimum variance estimator of the annual survival rate (i.e., $\hat{Z} = - \ell n \hat{S}$). for the two parameter case one can simply replace $\bar{t}$ in (8) with $(\bar{t}-t_c)$. Thus,

$$\hat{Z} = \ell n\left(\frac{1+\bar{t}-t_c - \frac{1}{n}}{\bar{t} - t_c}\right) - \frac{(n-1)(n-2)}{n(n(\bar{t} - t_c)+1)(n+n(\bar{t}-t_c)-1)} . \tag{9}$$

Equation (9) is extremely close to the results of Ssentongo and Larkin given by (6) or (7), especially for large n.

An approximate (asymptotic) expression for the variance of $\hat{Z}$ can be found using the delta method (Seber, 1973 pp. 7-11; Kendall and Stuart, 1977 pp. 246-8; Effron, 1982). For any Y which is a function of the random variables X_1, X_2, ... X_k we have

$$\hat{Y} = f(X_1, X_2, \ldots X_k)$$

and

$$V(\hat{Y}) \cong \left(\frac{\delta f}{\delta X_1}\right)^2 V(X_1) + \left(\frac{\delta f}{\delta X_2}\right)^2 V(X_2) + \ldots \left(\frac{\delta f}{\delta X_k}\right)^2 V(X_k)$$
$$+ 2 \sum_{i<j}\sum \left(\frac{\delta f}{\delta X_i}\right)\left(\frac{\delta f}{\delta x_j}\right) Cov(X_i, X_j) \tag{10}$$

where the derivatives are evaluated at the parameter values (or their estimates).

The variance of $\hat{Z}$ (equation (6)) for the geometric model is thus found to be

$$V(\hat{Z}) = \frac{(1 - e^{-Z})^2}{ne^{-Z}} \qquad (11)$$

with estimates given by

$$\widehat{V(\hat{Z})} = \frac{(1 - e^{-\hat{Z}})^2}{ne^{-\hat{Z}}} \quad . \qquad (12)$$

Ssentongo and Larkin presented, without derivation, the following expression for the variance of Z

$$V(\hat{Z}) = (\frac{n}{n+1})^2 \cdot \frac{1}{n} Z^2 \quad . \qquad (13)$$

Equations (11) and (13) give extremely close results for small values of Z. When Z is equal to 1.5, Ssentongo and Larkin's estimator is approximately 17% smaller than the delta method estimator. For Z = 2.0, the former is 28% smaller than the latter. A Monte Carlo type simulation was performed to evaluate the two estimators. The true value of the variance was found to be slightly larger than the estimate given by the delta method. Hence, equation (13) is to be preferred over (11).

The Monte Carlo Simulation also showed that the estimates of Z given by (6) through (9) are approximately normally distributed. Hence, the usual procedures for confidence intervals and tests of hypotheses (based on the normal distribution) are appropriate.

No estimator for the mortality rate, Z, has been derived for the geometric model when the upper tail is truncated. However, Chapman and Robson (1960, eq. (18)) gave the maximum likelihood estimator for the annual survival rate, S.

A natural estimator of Z would be given by

$$\hat{Z} = -\ell n \ \hat{S} \quad .$$

The approximate variance of this estimator can again be found using the delta method (10). Thus,

$$V(\hat{Z}) \cong \frac{1}{S^2} V(\hat{S})$$

where $V(\hat{S})$ is the variance of the maximum likelihood estimator of S given by Chapman and Robson (1960, eq. (20)).

ESTIMATORS BASED ON MEAN LENGTH

Method of Moments

Beverton and Holt (1956, p. 81) derived the following deterministic expression for mean length by assuming an exponential mortality model and von Bertalanffy growth:

$$\bar{\ell} = L_\infty\left(1 - \frac{Z}{Z+K} \cdot \frac{L_\infty - \ell_c}{L_\infty}\right),$$

where K and L_∞ are paramaters of the von Bertalanffy growth equation and ℓ_c is the length when the fish become vulnerable to the gear. Solving this for Z they obtained an estimator by the method of moments. Thus,

$$\hat{Z} = \frac{K(L_\infty - \bar{\ell})}{\bar{\ell} - \ell_c} \quad . \qquad (14)$$

The delta method (10) can be used to derive an expression for the variance of the estimate given by (14). It is assumed that the mean length in the population is estimated independently of the von Bertalanffy growth parameters. Then

$$V(\hat{Z}) \cong \left(\frac{L_\infty - \bar{\ell}}{\bar{\ell} - \ell_c}\right)^2 V(\hat{K}) + \left(\frac{K}{\bar{\ell} - \ell_c}\right)^2 V(\hat{L}_\infty)$$

$$+ K^2 \frac{(\ell_c - L_\infty)^2}{(\bar{\ell} - \ell_c)^4} V(\hat{\bar{\ell}}) + 2\left(\frac{L_\infty - \bar{\ell}}{\bar{\ell} - \ell_c}\right)\left(\frac{K}{\bar{\ell} - \ell_c}\right) \mathrm{Cov}(\hat{K},\hat{L}_\infty) \qquad (15)$$

where ℓ_c is assumed to be known perfectly. Again, substituting sample estimates into (15) gives an estimator of the variance.

Method of Maximum Likelihood

Ssentongo and Larkin (1973) attempted to derive a maximum likelihood estimator for Z by working with the probability density function of a transformation of length. They defined y and y_c to be the functions given by

$$y = -\ell n\left(1 - \frac{\ell}{L_\infty}\right), \qquad (16)$$

$$y_c = -\ell n\left(1 - \frac{\ell_c}{L_\infty}\right)$$

and showed that y follows a two parameter exponential distribution with probability density function

$$f(y) = \frac{Z}{K} \exp((-Z/K)(y-y_c))$$

for $y > y_c$. The maximum likelihood estimator is thus

$$\widehat{Z/K} = \frac{1}{\bar{y} - y_c} \qquad (17)$$

where

$$\bar{y} = \sum_{i=1}^{n} y_i/n \quad . \tag{18}$$

That is, $\bar{y}$ is the mean of the y values calculated for each individual length in the sample. Ssentongo and Larkin erroneously evaluated the function for y given by (16) at the mean length in the sample and used this in place of $\bar{y}$. The result is a serious positive bias.

In theory, one could calculate the proper sample value of $\bar{y}$ (from (18)) for insertion into equation (17). However, a cursory examination of equation (16) shows that whenever ℓ is greater than or equal to L_∞ the value of y will be undefined. Values of ℓ close to L_∞ will essentially be interpreted as representing unreasonably old ages.

We therefore conclude that the maximum likelihood approach of Ssentongo and Larkin should not be used in either form. Pauly (1980) converted Ssentongo and Larkin's estimator to a form that utilizes weight data. His estimator suffers from the same deficiencies as the estimator based on length and is likewise not recommended.

ESTIMATORS BASED ON MEDIAN LENGTH

The median length in a sample can often be found rather easily by arranging fish in ascending order by size. To derive an estimator based on the median length, we follow Ssentongo and Larkin (1973) in defining y and y_c as in (16). Thus, y is distributed as a two-parameter exponential variate with parameters Z/K and y_c. The median, y_m is given by

$$y_m = \frac{K \ \ell n \ 2}{Z} + y_c \quad . \tag{19}$$

Solving for Z, we obtain

$$\tilde{Z} = \frac{K \ \ell n \ 2}{\tilde{y}_m - y_c} \tag{20}$$

Where $\tilde{y}_m$ is the sample estimate of the median.

Assume for the moment that K and L_∞ are perfectly known. Equation (19) states that the median is a function of 1/Z. The estimator (20) consists of equating the mortality rate with a nonlinear function of the median, i.e. $1/\tilde{y}_m$. Hence, $\tilde{Z}$ is not necessarily unbiased. In practice, the bias in using (20) is sufficiently small to be ignored. (See Hoenig (1983) or Hoenig and Lawing (1982) for information on the expected value and variance of the reciprocals of order statistics (including the median) from an exponential distribution).

To find the approximate variance of $\tilde{Z}$, let $\tilde{x} = \tilde{y}_m - y_c$.

Then,

$$\tilde{Z} = \frac{K \ln 2}{x}$$

and

$$V(\tilde{Z}) = (K \ln 2)^2 V(1/\tilde{x}) . \quad (21)$$

The approximate variance of 1/x is derived in Hoenig (1983) as

$$V(1/\tilde{x}) = \frac{V(\tilde{x})}{[E(\tilde{x})]^4} - \frac{[V(\tilde{x})]^2}{[E(\tilde{x})]^6}$$

where $E(\tilde{x})$ is the expected value of the median from a sample size n (n odd) given by

$$E(\tilde{x}) = \frac{1}{Z} \sum_{i=1}^{(n+1)/2} (n - i + 1)^{-1}$$

$$V(\tilde{x}) = \frac{1}{Z^2} \sum_{i=1}^{(n+1)/2} (n - i + 1)^{-2}$$

The expected value and variance of $\tilde{Z}$ have not been worked out for the case where K and L_∞ are not perfectly known.

DISCUSSION

The methods evaluated here are generally easier to perform or require less data than traditional catch curve analysis. They provide explicit formulae for the variance of the estimated mortality rate in terms of the input parameters. thus a preliminary estimate of the mortality rate or other relevant parameters will allow one to estimate necessary sample sizes for further study.

The mortality estimator of Ssentongo and Larkin (1973), based on mean length, and that of Pauly (1980), based on mean weight, have been found to have undesirable statistical properties. Their use is not recommended. Unfortunately, Csirke and Caddy (1983) listed these methods as possible sources of mortality data for their stock production model.

A new estimator, based on the median length in a sample, is presented here as a potentially robust alternative to Beverton and Holt's estimator based on mean length. The median length may be rather insensitive to variability in year class strength. Also, for an exponential distribution, the median (y_m) is smaller than the mean by a factor of $\ln 2$. Though the variable y is distributed as an exponential variate, the y's represent a monotonic transformation of length. The relationship between age and length is strongest for small sizes. Thus the estimator based on the transformation of median length makes use of length data which is strongly related to age. This results in robustness with respect to variability in growth rate.

There are several other methods for estimating mortality rates from size distributions. These techniques are generally based on deterministic models and entail estimation by the least squares approach. The reader is referred to Silliman (1945), Ricker (1958 pp. 202-4), Powell (1976), Van Sickle (1977), Jones (1981), and Sainsbury (1982). Mortality estimators based on mean and median length remain to be derived for the geometric model.

ACKNOWLEDGEMENTS

Drs. Saul Saila and Edward Carney encouraged us to pursue this study.

REFERENCES

BEVERTON, R.J.H., and S.J. HOLT. 1956. A review of methods for estimating mortality rates in exploited fish populations, with special reference to sources of bias in catch sampling. Rapp. Pr.-Verb. Reuns. Cons. Perm. Int. Explor. Mer 140:67-83.

CHAPMAN, D., and D. ROBSON. 1960. The analysis of a catch curve. Biometrics 16:354-368.

CSIRKE, J., and J. CADDY. 1983. Production modeling using mortality estimates. Can. J. Fish. Aquat. Sci. 40:43-51.

EFFRON, B. 1982. The jackknife, the bootstrap and other resampling plans. Soc. Inuds. Appl. Math., Philadelphia. 92 pp.

HOENIG, J.M. 1983. Estimating total mortality rate from longevity data. Ph.D. diss., Univ. Rhode Island, Kingston. 92 pp.

HOENIG, J.M., and W.D. LAWING. 1982. Estimating the total mortality rate using the largest order statistic for age. Int. Counc. Explor. Sea C.M. 1982/D:7. 13 pp.

HOLT, S.J. 1965. A note on the relation between the mortality rate and the duration of life in an exploited fish population. ICNAF Res. Bull. 2:73-75.

JOHNSON, N.L., and S. KOTZ. 1970. Continuous univariate distributions - 1. John Wiley & Sons, New York. 300 pp.

JONES, R. 1981. The use of length composition data in fish stock assessment (with notes on VPA and cohort analysis). FAO Fish. Circ. 734, FAO, Rome. 55 pp.

KENDALL, M., and A. STUART. 1977. The advanced theory of statistics, Vol. 1, 4th ed. Macmillan, New York. 472 pp.

MARTI, K. 1982. Population age structure and mortality determination for the sea scallop (Placopecten magellanicus). M.S. Thesis, Univ. Rhode Island, Kingston. 85 pp.

PAULY, D. 1980. A selection of simple methods for the assessment of tropical fish stocks. FAO Fish. Circ. No. 729. FIRM/C701, 54 pp.

POWELL, D.G. 1976. Estimation of mortality and growth parameters from the length frequency of a catch. Int. Council. Explor. Sea Spec. Meeting on Pop. assess. of shellfish stocks.

RICKER, W.E. 1958. Handbook of computations for biological statistics of fish populations. Bull. Fish. Res. Bd. Can. 119, 300 pp.

SAILA, S.B., and P.M. ROEDEL (eds.). 1981. Proceedings of the international workshop on tropical small-scale fishery stock assessment, Univ. Rhode Island, Sept. 1979. 181 pp.

SAINSBURY, K.J. 1982. Population dynamics and fishery management of the paua, *Haliotus iris* I. Population structure, growth, reproduction, and mortality. N.Z.J. Mar. Freshw. Res. 16:147-161.

SEBER, G.A.F. 1973. The estimation of animal abundance. Hafner Press, New York. 506 pp.

SILLIMAN, R.P. 1945. Determination of mortality rates from length frequencies of the pilchart or sardine, *Sardinops caerulea*. Copeia 1945(4):191-196.

SSENTONGO, G.M., and P.A. LARKIN. 1973. Some simple methods of estimating mortality rates of exploited fish population. J. Fish. Res. Bd. Can. 30:695-698.

VANSICKLE, J. 1977. Mortality rates from size distributions - the application of a conservation law. Oecologia (Berl.) 27:311-318.