Estimating Maturity-at-Age from Length Stratified Sampling

M. Joanne Morgan and John M. Hoenig
Department of Fisheries and Oceans, P. O. Box 5667
St. John's, Newfoundland, Canada A1C 5X1

Abstract

Much of the research sampling for age composition in commercial fish species is stratified on the basis of length, with a fixed number of samples taken from each length class. For most fish species, a given age can straddle several length classes. Further, the probability of being mature at a given age is influenced by length at that age. Often, the proportion mature at a given age is estimated from the observed fish at that age without taking into account the size frequency distribution from which the specimens are drawn. This can result in biases of unpredictable magnitude and sign in the estimation of proportion mature-at-age. This study presents a method to estimate proportion mature-at-age from a length stratified sampling scheme. The method accounts for the size distribution at age by weighting observations by the abundances of the size groups. The method is applied to cod (*Gadus morhua*) and American plaice (*Hippoglossoides platessoides*) in the Northwest Atlantic and the results compared to those obtained from the unweighted (traditional) method. In general, the differences between the weighted and unweighted estimates are small, but some comparisons show substantial differences.

Keywords: American plaice, biomass, cod, maturity, sampling

Introduction

The ability to accurately determine age-at-maturity is important to the study and management of commercially exploited fish species. Spawning stock biomass can be estimated using biomass-at-age estimates and a 'knife edge' estimate of maturity (e.g. assuming a criterion such as all age 7+ fish are mature). However, use of the estimated proportion mature at each age will produce a more accurate estimate of the biomass of spawning fish. The ability to track changes in spawning stock biomass can be crucial to monitoring the health of stocks. Also, there are often changes in the age-at-maturity with changing population size (e.g. Beacham, 1983; Pitt, 1975; Templeman *et al.*, 1978; Shelton and Armstrong, 1983). These changes will be more reliably detected if one has an unbiased and precise estimate of the proportion mature at each age.

Many sampling programs for collecting age samples from fish populations are designed on the basis of length stratification, with a fixed number of the fish being taken from each *a priori* length group. The age and maturity of each sampled fish is then determined. Maturity-at-age estimates determined from samples collected in this way are subject to errors if the number of fish in each length group is not taken into account. When an age group straddles several length classes, the probability of being mature may increase with length and, therefore, the catch-at-length and the distribution of age across length must be accounted for.

In this paper we show that maturity-at-age can be affected by length-at-age in two important commercial species in the Northwest Atlantic, the Atlantic cod (*Gadus morhua*) and the American plaice (*Hippoglossoides platessoides*). Maturity-at-age in both these species is presently sampled using a length stratified scheme. A theoretically sound method is presented here to estimate the proportion mature-at-age from this sampling regime by weighting the observations from each length group by the abundance of the length category. The results are then compared to those obtained from the historically used method which does not account for the length composition in these species.

Historically Used Method (Unweighted)

A commonly used procedure (e.g. in much of the Canadian Department of Fisheries and Oceans in Atlantic Canada) is to estimate the proportion mature at age a, M_a, using only data from an aged sample collected by taking a fixed number of samples from each length class. The estimate is obtained by dividing the observed number of mature fish at age a by the number of fish observed at that age. Thus,

$$\hat{M}_{a1} = \frac{\sum_{j=1}^{J} \hat{P}(a|j)\, \hat{P}(m|a,j)}{\sum_{j=1}^{J} \hat{P}(a|j)} \tag{1}$$

where the caret (^) indicates an estimate, and

$\hat{P}(a|j)$　　is the proportion of the sampled fish in length category j which is age a,

$\hat{P}(m|a,j)$　is the proportion of the sampled fish at age a which are mature in the length category j, and

J　　　　　is the number of length classes.

Typically, the fish are collected at a number of stations (locations) but the estimates are computed without regard to the origin of the samples with respect to station.

Weighted Method

The data from the aged sample and from a sample of length frequencies can be combined to account for the effect of the length stratification scheme. This estimator weights the observations from each length category by the abundance of the length category. Thus,

$$\hat{M}_{a2} = \frac{\sum_{j=1}^{J} \hat{P}(j)\,\hat{P}(a|j)\,\hat{P}(m|a,j)}{\sum_{j=1}^{J} \hat{P}(j)\,\hat{P}(a|j)} \qquad (2)$$

where $\hat{P}(j)$ is the estimated fraction of the catch that was length j, and the other symbols are as defined previously.

The denominator is simply the estimated number of fish observed at age a, while the numerator is the estimated number of mature fish observed at age a. The asymptotic variance of this estimator is given in Appendix I. Depending on whether part or all of the sample is measured at each sampling station, the estimation of $P(j)$ will be based on a one- or two-stage cluster sampling procedure. The length categories are then treated as post-strata and subsampled by cluster (see Cochran (1977) for a discussion of multistage sampling and post-stratification).

A common sampling design is to divide the survey area into strata. Each stratum is sampled independent of the other strata with the number of sampling stations in each stratum specified *a priori*. In this case, the results from the different strata must be weighted appropriately to account for the different sizes of the strata. The true proportion mature at age a, M_a, is

$$M_a = \frac{\sum_{k=1}^{K} S_k D_k \sum_{j=1}^{J} P(m|a,j,k)\,P(a|j,k)\,P(j|k)}{\sum_{k=1}^{K} S_k D_k \sum_{j=1}^{J} P(a|j,k)\,P(j|k)} \qquad (3)$$

where　S_k　　　is the size (area) of the kth stratum,

D_k　　　is the density (number per unit area) of fish in the kth stratum,

$P(m|a,j,k)$　is the proportion of fish of age a and length j in stratum k which is mature,

$P(a|j,k)$　is the proportion of the fish of length j in stratum k which is age a,

$P(j|k)$　　is the proportion of fish in stratum k which is length j,

K　　　　is the number of strata, and

J　　　　is the number of length classes.

An estimate, $\hat{M}_{aS}$, of the proportion mature at age a can be obtained by substituting sample estimates for each of the above parameters into equation (3). Sample estimates are:

$$\hat{P}(m|a,j,k) = \frac{C_{majk}}{C_{ajk}}$$

$$\hat{P}(a|j,k) = \frac{C_{ajk}}{C_{jk}}$$

$$\hat{P}(j,k) = \frac{C_{jk}}{C_k}$$

$$\hat{D}_k = \overline{C}_k$$

where　C_x is the total number of fish observed in class x, and

$\overline{C}_k$ is the mean catch of fish per unit effort in stratum k.

Here, the stratum sizes S_k are assumed to be known and the expected catch-per-unit-effort is assumed proportional to the mean density of fish.

An alternative way to write the estimator is:

$$\hat{M}_{aS} = \frac{\sum_{k=1}^{K} S_k \sum_{j=1}^{J} \overline{C}_{jk}\hat{P}(a|j,k)\,\hat{P}(m|a,j,k)}{\sum_{k=1}^{K} S_k \sum_{j=1}^{J} \overline{C}_{jk}\hat{P}(a|j,k)} \qquad (4)$$

where $\overline{C}_{jk}$ is the mean catch-per-unit-effort of fish of length j in stratum k and the other symbols are as defined previously.

A special case is where proportional allocation is used to assign the number of sampling stations to each stratum, and a constant sampling fraction is taken from the catch at each stratum. That is, the

number of stations in each stratum is proportional to the area of the stratum. Then, equation (4) reduces to:

$$\hat{M}_{a3} = \frac{\sum_{j=1}^{J} \overline{C}_j \hat{P}(a|j) \hat{P}(m|a,j)}{\sum_{j=1}^{J} \overline{C}_j \hat{P}(a|j)}$$

$$= \frac{\sum_{j=1}^{J} \hat{P}(j) \hat{P}(a|j) \hat{P}(m|a,j)}{\sum_{j=1}^{J} \hat{P}(j) \hat{P}(a|j)} \quad (5)$$

where:

$$\hat{P}(j) = \frac{\sum_{k=1}^{K} \overline{C}_{jk}}{\sum_{k=1}^{K} \overline{C}_k}$$

$\overline{C}_j$ is the mean catch-per-unit-effort of fish of length j (over all strata combined)

$$\hat{P}(a|j) = \frac{\sum_{k=1}^{K} C_{ajk}}{\sum_{k=1}^{K} C_{jk}}$$

$$\hat{P}(m|a,j) = \frac{\sum_{k=1}^{K} C_{majk}}{\sum_{k=1}^{K} C_{ajk}}$$

and the other symbols are as described previously.

In other words, under proportional allocation sampling, the data from all strata can be combined as if there were no stratification (i.e. equation 5 = equation 2). $\hat{P}(j)$ is simply the fraction of fish encountered in the survey (all areal strata combined) that is length j. $\hat{P}(a|j)$ is the fraction of the fish of length j encountered in the entire survey area which is of age a, and likewise for $\hat{P}(m|a,j)$.

Examples

The unweighted estimator, $\hat{M}_{a1}$ (equation 1) and the weighted estimator $\hat{M}_{a2}$ (equation 2) were compared for two species, Atlantic cod and American plaice, using data collected in NAFO Div. 3L on the northern Grand Bank off Newfoundland. Samples were collected during spring multispecies groundfish surveys conducted by the Department of Fisheries and Oceans, Canada, using an otter trawl with a fine mesh liner in the cod end. In the surveys, a depth stratified design was used where sets in the depth strata were chosen at random, with the number of sets being approximately proportional to the stratum area (Doubleday, 1981). We treated the data as coming from a proportional allocation sampling scheme. For American plaice, data from the spring surveys from 1971–92 were used, and for cod the data were from the spring surveys in the years 1978–92. For both species, otoliths were collected for ageing using a length stratified scheme. For cod, 25 samples were collected from each 3 cm length class. For American plaice, a fixed number of fish were sampled for each 2 cm length class. The fixed number varied with the size of fish, with fewer fish sampled at the smallest and largest lengths. In all surveys an attempt was made to spread the sampling throughout the entire Div. 3L.

Maturity stages were determined at sea according to the criteria of Templeman et al. (1978). In this scheme there are 9 main maturity stages for females and 7 for males. For each sex the first stage is immature and all other stages show some evidence of maturing to spawn in the present year or of having spawned in the past and are classed as mature or adult fish. During these surveys, length frequency data were collected from every successful fishing set to provide an estimate of the length distribution of the populations. All fish were sampled for length from the vast majority of sets, however, when catches were very large the entire catch was weighed and then a subsample taken for length. The sampling design does not strictly meet the criterion of a constant sampling fraction from the catch in each stratum but is considered to vary only slightly from this design.

To determine if there can be an effect of length on the proportion mature-at-age, the proportion of mature fish at one age at each length class was calculated for each sex using fish from the aged sample. For both cod and American plaice there can be an effect of length on the proportion mature-at-age, with the probability of a fish being mature at a given age increasing with increasing length. As examples, results for age 9 American plaice are presented in Fig. 1 and for age 6 cod in Fig. 2. The effect of length on proportion mature was statistically significant (for age 9 American plaice males: $\chi^2 = 43.14$, df $= 15$, p<0.0001 and for age 9 American plaice females: $\chi^2 = 84.08$, df $= 13$, p<0.0001. For age 6 cod males: $\chi^2 = 33.67$, df $= 15$, p<0.005 and for age 6 cod females: $\chi^2 = 45.93$, df $= 12$, p<0.0001).

The difference between the estimates obtained from weighting by the abundance of each length category (weighted estimates, equation 2) and the estimates of percentage mature-at-age from the unweighted method (equation 1) was calculated by
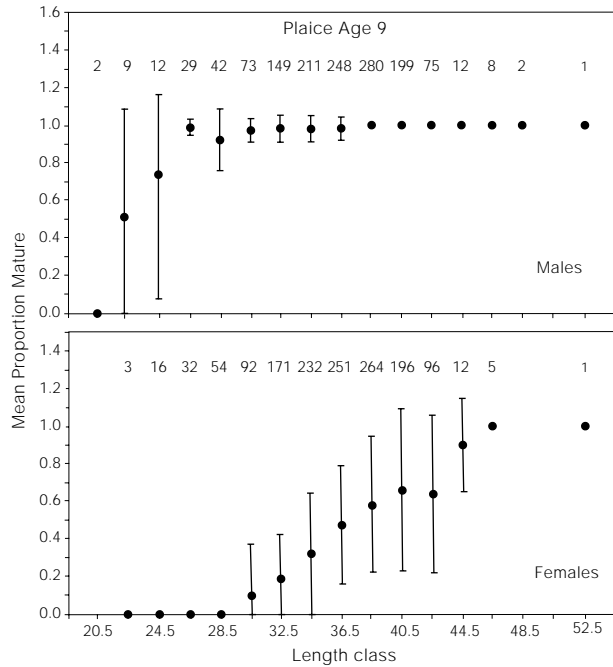
Fig. 1.   Mean (±1 Std. dev.) proportion mature for age 9 American plaice males and females at each length class. The means and standard deviations are over the years 1971–92. The numbers above the points are the number of fish examined over the years.
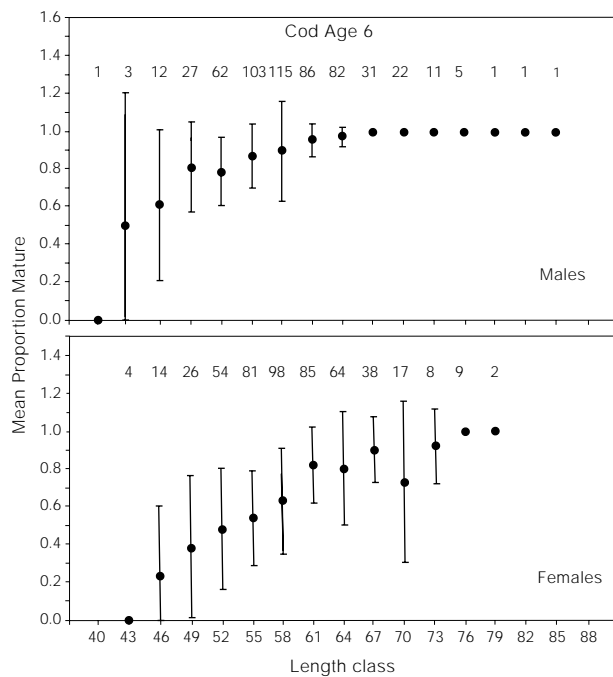

Fig. 2.   Mean (±1 Std. dev.) proportion mature for age 6 cod males and females at each length class. The means are over the years 1978–92. The numbers above the points are the number of fish examined.

subtracting the weighted from the unweighted estimate for each age in each year. The two estimates were compared in this manner for each species and sex for every age in every year, for example for male cod, all sampled ages were compared in each of 13 years. Age at 50% maturity was estimated for each species, for each sex and year, using probit analysis assuming a normal distribution. This was done for both the weighted and unweighted estimates.

An estimate of spawning stock biomass in Div. 3L was also calculated for each method. For American plaice, the proportion mature-at-age for each sex was multiplied by the numbers-at-age for each sex from the research surveys and the commercial weights-at-age for all of Div. 3LNO, sexes combined (Brodie *et al.*, MS 1993; Brodie, MS 1985). For cod, the proportion mature-at-age for each sex was multiplied by the numbers-at-age for each sex from the surveys and the research vessel weights-at-age for Div. 3L in spring, sexes combined (Bishop *et al.*, MS 1993).

When the weighted estimate of percentage mature-at-age calculated using equation 2 was compared to the unweighted percent mature-at-age (equation 1), the differences were sometimes negative and sometimes positive (Fig. 3 and 4). For a given age they varied from year to year both in magnitude and direction of the difference. However, for both sexes of both species approximately one-third of the differences were less than 1%.

For male American plaice (Fig. 3), 26.2% of the comparisons had a difference of less than 1%. There was a difference of greater than 5% in 28.9% of the comparisons, and the largest difference was 49.3%. For American plaice females (Fig. 3), 32.7% of the comparisons differed by less than 1%, 15.7% of the comparisons differed by more than 5% and the largest difference between unweighted and weighted estimates was 25.7%.

For male cod (Fig. 4), 27.6% of the comparisons differed by less than 1%, 16.2% differed by more than 5% and the largest difference was 16.2%. For female cod (Fig. 4), 29.5% of the comparisons differed by less than 1% with 22.8% differing by more than 5% and the largest difference was 15.0%.

For both species, there were only small differences (less than one year) in the estimates of age at 50% maturity from the two methods (Tables 1 and 2). The magnitude of the differences between the two estimates was not consistent over years nor was the direction of the difference consistent.

The estimates of spawning stock biomass using the two methods differed by 60 (0.1%) to 13 000
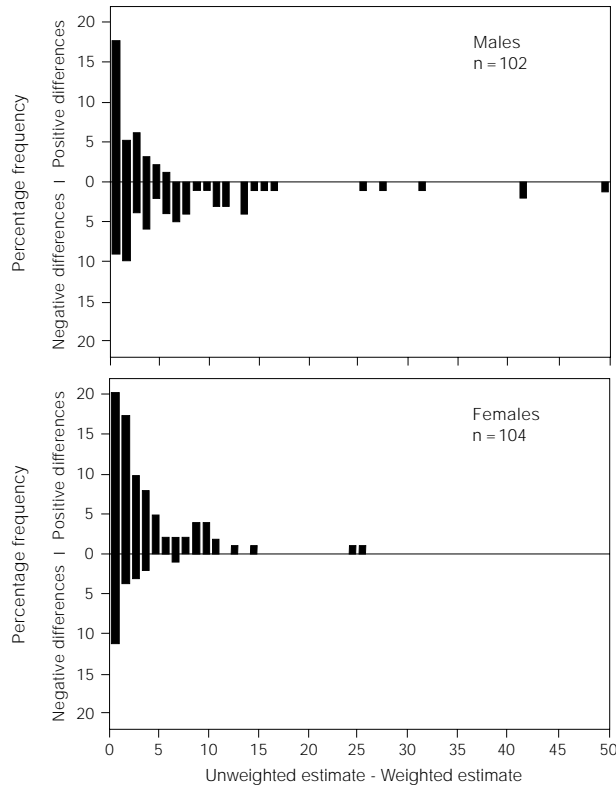
Fig. 3. Percent frequency of occurrence of differences between unweighted and weighted percentage mature at each age in each year for male and female American plaice. (*n* = total number of age× year combinations compared.)
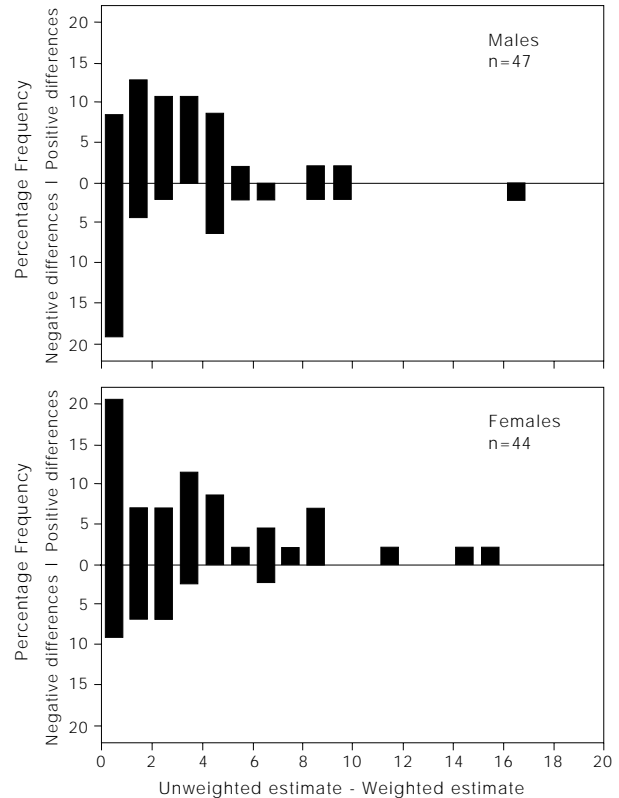

Fig. 4. Percent frequency of occurrence of differences between unweighted and weighted percentage mature at each age in each year for male and female cod. (*n* = total number of age× year combinations compared.)

(8.9%) tons for American plaice and by 230 (0.1%) to 4 400 (6.0%) tons for cod (Tables 3 and 4). Again, neither the magnitude nor direction of the differences between the estimates was consistent over years. For cod all but one of the differences was positive, while for American plaice all differences up until 1982 were positive and from 1984 to 1992 all differences but one were negative.

Discussion

For both cod and American plaice the probability of being mature at a given age increases with size. This means that estimates of maturity-at-age, based on a length stratified sampling scheme, will be biased if the distribution of size-at-age is not accounted for properly. This potential bias can be easily avoided using the estimators presented here. Estimators (2), (4) and (5) are consistent and asymptotically unbiased under simple random sampling (Cochran, 1977).

The differences between the weighted and unweighted estimates of the percentage mature-at-

age were generally small. However, some large differences did occur, up to 49% for male American plaice. The differences were not consistent in magnitude or direction across ages, or for a given age across years. Therefore, a constant empirical correction factor cannot be applied. There were no apparent differences in the data or sampling between instances when the two estimates were very similar and when they were very different. There was no tendency to have many large differences in the same year nor was there any indication that the sample size differed substantially from the average for comparisons that showed large differences between the estimates.

There was also little difference in the estimate of age at 50% maturity using estimates of percentage mature-at-age from the two methods. Again the magnitude and direction of the differences were not consistent, precluding the application of a constant empirical correction factor. Estimates of spawning stock biomass using the two methods differed by as much as 13 000 tons (8.9%) although most of the differences were less than 1 000 tons (0.1%).

TABLE 1.   Estimates of age at 50% maturity in American plaice using weighted (equation 2) and unweighted (equation 1) estimates of proportion mature-at-age.

Year	Males		Females	
	Weighted	Unweighted	Weighted	Unweighted
1971	6.15	6.41	10.61	10.52
1972	5.90	6.14	10.93	10.83
1973	5.53	5.47	9.81	9.78
1974	5.67	6.29	11.09	10.82
1975	6.88	7.23	11.61	10.79
1976	6.46	6.78	10.99	10.68
1977	6.35	6.69	11.17	10.93
1978	5.51	5.72	10.15	9.79
1979	6.10	6.25	10.00	9.68
1980	6.54	6.78	10.73	10.56
1981	6.25	6.62	9.85	9.68
1982	6.70	7.34	10.42	10.29
1984	5.06	5.38	8.50	8.59
1985	4.86	5.44	8.40	8.47
1986	5.01	5.63	8.43	8.53
1987	5.62	5.95	7.99	8.03
1988	3.76	4.52	7.78	7.89
1989	3.98	4.96	7.92	7.90
1990	4.47	4.78	7.74	7.75
1991	4.73	5.46	8.71	8.64
1992	4.03	4.73	7.95	7.92

TABLE 2.   Estimates of age at 50% maturity for cod using weighted (equation 2) and unweighted (equation 1) estimates of proportion mature-at-age.

Year	Males		Females	
	Weighted	Unweighted	Weighted	Unweighted
1978	5.10	5.04	6.07	6.05
1979	4.82	4.85	5.91	5.59
1980	5.00	4.96	5.63	5.58
1981	5.02	5.08	5.64	5.61
1982	4.74	4.63	5.42	5.35
1985	5.18	5.10	6.21	6.11
1986	5.04	5.01	6.04	5.88
1987	5.22	5.29	5.98	6.00
1988	4.78	4.96	5.93	5.93
1989	5.74	5.70	6.00	6.00
1990	5.03	4.98	5.27	5.25
1991	5.48	5.35	6.07	5.89
1992	4.06	4.24	5.08	4.93

Although differences in the weighted and unweighted estimates were generally small, the method presented is simple to apply, and results in a more accurate estimate of proportion mature-at-age.

Acknowledgements

TABLE 3.   Estimates of spawning stock biomass (in tons) of American plaice in Div. 3L using unweighted (equation 1) and weighted (equation 2) estimates of proportion mature-at-age.

Year	Unweighted	Weighted	Difference
1971	226 068	225 468	600
1972	139 297	139 232	65
1973	50 930	50 868	62
1974	131 663	130 075	1 588
1975	150 648	137 178	13 470
1976	235 429	229 912	5 517
1977	265 772	256 254	9 518
1978	218 159	208 656	9 503
1979	264 157	255 928	8 229
1980	193 376	188 548	4 828
1981	183 720	182 474	973
1982	222 783	218 942	3 841
1984	92 142	92 994	-852
1985	180 770	181 629	-859
1986	144 486	144 994	-508
1987	167 697	167 351	346
1988	178 309	178 790	-481
1989	126 549	126 640	-91
1990	102 220	102 372	-152
1991	45 578	46 255	-677
1992	16 888	16 997	-109

TABLE 4.   Estimates of spawning stock biomass (in tons) of cod in Div. 3L using unweighted (equation 1) and weighted (equation 2) estimates of proportion mature-at-age.

Year	Unweighted	Weighted	Difference
1978	41 108	40 356	752
1979	74 310	69 829	4 481
1980	93 400	92 327	1 073
1981	171 199	170 300	899
1982	107 481	106 562	919
1985	165 730	161 896	3 834
1986	156 571	154 225	2 346
1987	164 641	163 533	1 108
1988	188 903	189 133	-230
1989	130 955	129 875	1 080
1990	165 938	164 046	1 892
1991	50 230	48 178	2 052
1992	20 023	19 474	549

References

BEACHAM, T. D. 1983. Variability in median size and age at sexual maturity of Atlantic cod, *Gadus morhua*, on the Scotian Shelf in the Northwest Atlantic Ocean. *Fish. Bull. U.S.*, **81**: 303–321.

BISHOP, C. A., E. F. MURPHY, M. B. DAVIS, J. W. BAIRD, and G. A. ROSE. MS 1993. An assessment of the cod stock in NAFO Divisions 2J+3KL. *NAFO SCR Doc.*, No. 86, Serial No. N2271, 51 p.

BRODIE, W. B. MS 1985. An assessment update of the American plaice stock in NAFO Divisions 3LNO. *NAFO SCR Doc.*, No. 51, Serial No. N1000, 28 p. + addendum.

BRODIE, W. B., D. POWER, and M. J. MORGAN. MS 1993. An assessment of the American plaice stock in NAFO Divisions 3LNO. *NAFO SCR Doc.*, No. 91, Serial No.

N2277, 60 p.

COCHRAN, W.G. 1977. Sampling techniques. 3rd edition. John Wiley and Sons, New York.

DOUBLEDAY, W. G. 1981. Manual on groundfish surveys in the Northwest Atlantic. *NAFO Sci. Coun. Studies*, **2**: 55 p.

PITT, T. K. 1975. Changes in abundance and certain biological characteristics of Grand Bank American plaice, *Hippoglossoides platessoides. J. Fish. Res. Board Canada*, **32**: 1383–1398.

SHELTON, P. A., and M. J. ARMSTRONG. 1983. Variations in parent stock and recruitment of pilchard and anchovy populations in the southern Benguela system. *In:*. Proceedings of the Expert Consultation to Examine Change in Abundance and Species Composition of Neritic Fish Resources, San Jose, Costa Rica, 18–29 April 1983. G. D. Sharp and J. Csirke (eds.). *FAO Fish. Rep.,* **291**(3); 1113–1132.

TEMPLEMAN, W., V. M. HODDER, and R. WELLS. 1978. Sexual maturity and spawning in haddock *Melanogrammus aeglefinus* of the southern Grand Bank. *ICNAF Res. Bull.*, **13**: 53–65.

APPENDIX 1: Variance of the Estimated Proportion Mature at age *a*

The estimated proportion mature at age *a*, $\hat{M}_{a2}$, is given by equation (2) (see Introduction above) as:

$$\hat{M}_{a2} = \frac{\sum_{j=1}^{J} \hat{P}(j)\, \hat{P}(a\,|\,j)\, \hat{P}(m\,|\,a,j)}{\sum_{j=1}^{J} \hat{P}(j)\hat{P}(a\,|\,j)} \tag{A1}$$

where the symbols are as defined in the main text. To reduce the amount of writing we define:

$$X_j = \hat{P}(j)$$
$$Y_j = \hat{P}(a\,|\,j)$$
$$Z_j = \hat{P}(m\,|\,a,j)$$
$$N = \sum_{j=1}^{J} \hat{P}(j)\, \hat{P}(a\,|\,j)\, \hat{P}(m\,|\,a,j)$$
$$D = \sum_{j=1}^{J} \hat{P}(j)\, \hat{P}(a|j)$$

Then equation A1 can be written as:

$$\hat{M}_{a2} = \frac{\sum_{j=1}^{J} X_j Y_j Z_j}{\sum_{j=1}^{J} X_j Y_j} \quad .$$

$\hat{M}_{a2}$ is thus a function of $3J$ random variables.

The variance of a function of random variables can be approximated by the delta or Taylor's series method. The variance of equation A1 is thus found to be:

$$V(\hat{M}_{a2}) = \sum_{i=1}^{J}\sum_{j=1}^{J} \frac{\partial \hat{M}_{a2}}{\partial X_i}\frac{\partial \hat{M}_{a2}}{\partial X_j}Cov(X_i, X_j) + \sum_{i=1}^{J}\sum_{j=1}^{J} \frac{\partial \hat{M}_{a2}}{\partial Y_i}\frac{\partial \hat{M}_{a2}}{\partial Y_j} Cov(Y_i, Y_j)$$

$$+ \sum_{i=1}^{J}\sum_{j=1}^{J} \frac{\partial \hat{M}_{a2}}{\partial Z_i}\frac{\partial \hat{M}_{a2}}{\partial Z_j}Cov(Z_i, Z_j) + \sum_{i=1}^{J}\sum_{j=1}^{J} \frac{\partial \hat{M}_{a2}}{\partial X_i}\frac{\partial \hat{M}_{a2}}{\partial Y_j} Cov(X_i, Y_j)$$

$$+ \sum_{i=1}^{J}\sum_{j=1}^{J} \frac{\partial \hat{M}_{a2}}{\partial X_i}\frac{\partial \hat{M}_{a2}}{\partial Z_j}Cov(X_i, Z_j) + \sum_{i=1}^{J}\sum_{j=1}^{J} \frac{\partial \hat{M}_{a2}}{\partial Y_i}\frac{\partial \hat{M}_{a2}}{\partial Z_j} Cov(Y_i, Z_j)$$

The required derivatives are:

$$\frac{\partial \hat{M}_{a2}}{\partial X_i} = (D\, Y_i\, Z_i - N\, Y_i)\, D^{-2}$$

$$\frac{\partial \hat{M}_{a2}}{\partial Y_i} = (D\, X_i\, Z_i - N\, X_i)\, D^{-2} \tag{A2}$$

$$\frac{\partial \hat{M}_{a2}}{\partial Z_i} = \frac{X_i\, Y_i}{D}$$

Then the variances and covariances remain to be determined. These will depend on the particular sampling design employed, so that it is not possible to derive generally applicable relationships. Because of the lack of a precisely defined sampling design, variances can not be calculated for the data used here. Instead an example is provided.

Equation A2 also holds for stratified random sampling with proportional allocation used for the areal strata. The required variances and covariances will depend on the sampling design used within strata.

Example

Suppose n = 3 stations are randomly selected for trawling and at each station a fixed fraction f_1 of the fish are sexed and measured for length. Of the females, a fixed fraction f_2 = 10% are aged and the maturity is ascertained. For illustrative purposes we assume the data in Table A.1 are obtained; we estimate the proportion mature at age 2 and the variance. Note that for illustrative purposes we have kept the data set (particularly the number of stations) very small. Thus, we would expect that the precision would be poor and also that the validity of the asymptotic approximations would be questionable.

The estimated proportion in each length class j, X_j, the estimated proportion that is age 2 in each length class, Y, and the proportion of age 2 fish that are mature in each length class, Z_j, are given in the first three lines ^{j}of Table A.2. Thus, N = 0.0604, D = 0.1813, and the estimated proportion mature is $M_{a=2}$ = 0.3333. The partial derivatives (equation A2) evaluated at the parameter estimates are given in lines 4 through 6 of Table A.2.

The X_j, are estimates of a multinomial proportion obtained from cluster sampling. The covariances are estimated by:

$$C\hat{o}v\,(X_j,\,X_k) = \frac{\sum\limits_{i=1}^{n} (C_{ij} - m_i\,X_j)\,(C_{ik} - m_i\,X_k)}{n\,(n-1)\overline{m}^2}$$

where C_{ij} is the number of measured fish of length j at station (cluster) i, m_i is the total number of fish measured at station i, and $\overline{m}$ is the mean of the m_i.

TABLE A.1.   Hypothetical data on length composition, age and maturity of female fish at 3 sampling stations (clusters). The m_i are the sums of the catches at each station ($m_i = \Sigma_j\,C_{ij}$).

Station, i	Variable	Length Class j				
		1	2	3	4	5
1	catch, C_{1j}	90	180	240	160	30
	number aged, a_{1j}	9	18	24	16	3
	number age 2, A_{1j}	3	5	2	0	0
	number age 2 & mature, S_{1j}	1	1	1	0	0
2	catch, C_{2j}	120	80	60	70	10
	number aged, a_{2j}	12	8	6	7	1
	number age 2, A_{2j}	3	3	3	0	0
	number age 2 & mature, S_{2j}	1	1	1	0	0
3	catch, C_{3j}	150	190	240	170	30
	number aged, a_{3j}	15	19	24	17	3
	number age 2, A_{3j}	6	5	2	1	0
	number age 2 & mature, S_{3j}	1	2	1	1	0

TABLE A.2. Estimated proportions, derivatives and covariances derived from the data in Table A.1. No fish of age 2 in a length category is denoted by ---.

Estimated proportion	symbol	Length class j				
		1	2	3	4	5
$P(j)$	X_j	0.1978	0.2473	0.2967	0.2198	0.0385
$P(a = 2\,|\,j)$	Y_j	0.3333	0.2889	0.1296	0.0250	0.0000
$P(m\,|\,a = 2, j)$	Z_j	0.2500	0.3077	0.4286	1.0000	---

derivatives evaluated at the parameter estimates

Variable	Length class j				
	1	2	3	4	5
X_j	-0.15319865	-0.04085297	0.06808829	0.09191919	---
Y_j	-0.09090909	-0.03496503	0.15584416	0.80808081	---
Z_j	0.36363636	0.39393939	0.21212121	0.03030303	0.00000000

covariances of estimated proportions at length ($\times 10^5$)

	$[X_1]$	$[X_2]$	$[X_3]$	$[X_4]$	$[X_5]$
$[X_1]$	233.18	-24.35	-170.21	-24.51	-14.10
$[X_2]$	-24.35	3.29	16.55	2.98	1.53
$[X_3]$	-170.21	16.55	126.27	17.20	10.20
$[X_4]$	-24.51	2.98	17.20	2.82	1.52
$[X_5]$	-14.10	1.53	10.20	1.52	0.86

covariances of estimated proportions age 2 by length ($\times 10^5$)

	$[Y_1]$	$[Y_2]$	$[Y_3]$	$[Y_4]$	$[Y_5]$
$[Y_1]$	231.48	-109.05	-257.20	78.12	0
$[Y_2]$	-109.05	55.82	141.75	-26.81	0
$[Y_3]$	-257.20	141.75	381.04	-40.51	0
$[Y_4]$	78.12	-26.81	-40.51	48.87	0
$[Y_5]$	0.00	0.00	0.00	0.00	0

covariances of estimated proportions of 2 year olds that are mature by length ($\times 10^5$)

	$[Z_1]$	$[Z_2]$	$[Z_3]$	$[Z_4]$	$[Z_5]$
$[Z_1]$	390.62	-332.84	-191.33	0	---
$[Z_2]$	-332.84	451.66	-54.34	0	---
$[Z_3]$	-191.33	-54.34	374.84	0	---
$[Z_4]$	0.00	0.00	0.00	0	---
$[Z_5]$	---	---	---	---	---

TABLE A.2.   (Continued). Estimated proportions, derivatives and covariances derived from the data in Table A.1. No fish of age 2 in a length category is denoted by ---.

covariances of estimated proportions length j with estimated proportions that are age 2 by length $(\times 10^5)$

	$[Y_1]$	$[Y_2]$	$[Y_3]$	$[Y_4]$	$[Y_5]$
$[X_1]$	-130.57	88.14	216.76	15.84	0
$[X_2]$	2.77	-5.11	-16.71	-7.62	0
$[X_3]$	113.21	-71.09	-167.99	-1.73	0
$[X_4]$	7.55	-6.94	-19.42	-5.06	0
$[X_5]$	7.04	-5.01	-12.64	-1.43	0

covariances of estimated proportions length j with estimated proportions of 2 year olds that are mature $(\times 10^5)$

	$[Z_1]$	$[Z_2]$	$[Z_3]$	$[Z_4]$	$[Z_5]$
$[X_1]$	22.08	178.62	-266.16	0	---
$[X_2]$	14.72	-33.98	20.52	0	---
$[X_3]$	-44.15	-105.15	206.27	0	---
$[X_4]$	7.36	-27.49	23.84	0	---
$[X_5]$	0.00	-12.00	15.53	0	---

estimated covariances of proportions age 2 with proportions of 2 year olds that are mature $(\times 10^5)$

	$[Z_1]$	$[Z_2]$	$[Z_3]$	$[Z_4]$	$[Z_5]$
$[Y_1]$	-12500.00	6410.26	7142.86	0	---
$[Y_2]$	6111.11	-1082.62	-4920.63	0	---
$[Y_3]$	13888.89	4273.50	-15873.02	0	---
$[Y_4]$	-7187.50	7788.46	1250.00	0	---
$[Y_5]$	0.00	0.00	0.00	0	---

The Y_j are estimates of the fractions of the fish in length category j that are age 2. The covariances of the Y's are computed by:

$$C\hat{o}v\,(Y_j,\,Y_k) = \frac{\sum_{i=1}^{n}(A_{ij} - a_{ij}\,Y_j)\,(A_{ik} - a_{ik}\,Y_k)}{n\,(n-1)\bar{a}_j\,\bar{a}_k}$$

where A_{ij} is the number of fish of age 2 in length j from station i, a_{ij} is the number of fish of length j aged from station i, and $\bar{a}_j$ is the mean of the a_{ij} calculated over all stations i.

Similarly, if we let S_{ij} represent the number of fish of age 2 and length j from station i that are mature, then the covariances of the Z_j's can be estimated by:

$$C\hat{o}v\,(Z_j,\,Z_k) = \frac{\displaystyle\sum_{i=1}^{n} (S_{ij} - A_{ij}\,Z_j)\,(S_{ik} - A_{ik}\,Z_k)}{n\,(n-1)\bar{A}_j\,\bar{A}_k}$$

where $\bar{A}_j$ is the mean of the A_{ij} calculated over all stations i.

The remaining covariances can be estimated by:

$$C\hat{o}v\,(X_j,\,Y_k) = \frac{\displaystyle\sum_{i=1}^{n} (C_{ij} - m_i\,X_j)\,(A_{ik} - a_{ik}\,Y_k)}{n\,(n-1)\bar{m}\bar{a}_k}$$

$$C\hat{o}v\,(X_j,\,Z_k) = \frac{\displaystyle\sum_{i=1}^{n} (C_{ij} - m_i\,X_j)\,(S_{ik} - A_{ik}\,Z_k)}{n\,(n-1)\bar{m}\bar{A}_k}$$

$$C\hat{o}v\,(Y_j,\,Z_k) = \frac{\displaystyle\sum_{i=1}^{n} (A_{ij} - a_{ji}\,Y_j)\,(S_{ik} - A_{ik}\,Z_k)}{n\,(n-1)\bar{a}_j\bar{A}_k}$$

The estimated covariance matrices are shown in Table A.2.

The estimated variance of $\hat{M}_{a=2}$ is 0.0115 (standard error = 0.11) thus indicating that the estimated proportion mature is not very precise. This is not surprising given the small number of stations visited and the small numbers examined for age and maturity.