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

International Council for the Exploration of the Sea

C.M. 1975/F:31
Demersal Fish (Northern) Committee
Ref D (Statistics C.)

THE PRECISION OF ESTIMATED AGE COMPOSITIONS

by

Christine Johnston, M.D. Nicholson and J.A. Pope
Marine Laboratory, Aberdeen, Scotland.

THÜNEN
Digitalization sponsored by Thünen-Institut

Abstract

This paper contains a brief description of the sampling procedure at Aberdeen Market by which estimates of numbers landed by age for a particular species are obtained. A simulation study to examine the variability underlying these estimates is described.

Introduction

In a previous contribution (Johnston et al 1974) the authors presented a description of fish market sampling in Scotland, with emphasis on sampling at Aberdeen Market. The authors also presented formulae for an estimate and its variance of the total landings of fish in any specified size range. The formula for the variance has since been amended*, and should be written as

$$\mathrm{Var}(\bar{\bar{My}}) = \frac{N}{n}\left\{\sum_{ij}^{Ns} M_{ij}\,(M_{ij}-m_{ij})\,\frac{S_{ij}^2}{m_{ij}} + \frac{N-n}{N}\sum_{i}^{N} M_i^2\,(\bar{Y}_i - \bar{\bar{Y}})^2\right\}$$

The estimates of numbers in a given size range are only intermediate objectives, a fuller objective being the production of estimates of the numbers landed of each age group. This introduces an additional stage into the estimation process, namely the employment of age-length keys. The variability of these estimates will then have two components; the first due to the sampling for length, and the second due to sampling for age.

Sampling Procedure

In this as in the previous contribution, attention will be confined to haddock sampling on Aberdeen fish market.

Sampling is carried out on a monthly basis. Landings are stratified for marketing purposes into a number of size categories termed, for example, small, medium, large, extra large. Fish are placed into identical boxes, each containing ostensibly the same weight (50.8 kg).

* The authors are grateful to Mr H B Becker (Ijmuiden) for drawing their attention to the need for this amendment.

Landings for a boat are then measured as numbers of boxes in each category.

The first stage of sampling is for vessels. The requirement that vessels should be selected at random is to some extent compromised by the need that sampling activities do not interfere with those of commerce. It is not felt that this introduces any source of bias.

Since the size categories of fish are not stable, it is necessary to select one box from each of these categories for each of the selected boats. This is the second stage of sampling. All fish in a box are then measured.

The third stage is to select a subsample from the measured fish. At each centimetre interval up to 5 fish are selected, otolithed and subsequently aged. These data are used to form the age-length key.

The length frequencies within boxes are then raised and combined to give an estimated length composition for the month. The age-length key is then applied, and estimates of numbers at age obtained.

Variance Formulae

For each age group the expectation and variance of the estimated number is the sum of the expectations and variances of the estimated numbers in each of the individual length groupings. We can write

$$E(\hat{N}_a) = \sum_{\ell} \{ E(\hat{N}_\ell) E(\hat{P}_{a\ell}) + \mathrm{Cov}(\hat{N}_\ell \hat{P}_{a\ell}) \}$$

and to a second order approximation

$$\mathrm{Var}(\hat{N}_a) = \sum_{\ell} \{ E(\hat{P}_{a\ell})^2 \mathrm{Var}(\hat{N}_\ell) + E(\hat{N}_\ell)^2 \mathrm{Var}(\hat{P}_{a\ell}) + 2 \mathrm{Cov}(\hat{N} \hat{P}_{a\ell}) \}$$

where the subscripts a and ℓ refer to age and length, and $\hat{N}_a$ and $\hat{P}_{a\ell}$ are the estimated number aged a and the estimated proportion aged a at length ℓ respectively.

There are difficulties with the variance formula. Consider the variance of $\hat{P}_{a\ell}$. Since fish selected for aging are a subsample of those measured, this can be written as the sum of two components.

ie $$\mathrm{Var}(\hat{P}_{a\ell}) \equiv \mathrm{Var}_f \{ E_s^{(f)} (\hat{P}_{a\ell}) \} + E_f \{ \mathrm{Var}_s^{(f)} (\hat{P}_{a\ell}) \}$$

where f and s refer to expectations taken over the length and age samples respectively. The expectation and variance for the age sample can be determined using the properties of simple random sampling. There are difficulties however in evaluating these quantities over the length sample. Although formulae exist for the case of cluster sampling within strata, these formulae require that numbers within clusters are weighted by the total numbers in strata. Equivalent formulae have not yet been developed for our case in which this weighting is not done. Because of this and other problems, variance formulae for the estimated numbers of fish landed by age have not yet been determined in a usable form.

To satisfy the need for some indication of the reliability of estimates obtained by market sampling, an interim solution has been sought in a computer simulation study.

Simulation study

The variances of estimated numbers in a particular length range quoted by Johnston et al (1974) referred to haddock landed at Aberdeen during July 1974 by trawlers fishing at Faroe Plateau. It was decided to direct the simulation study to a further analysis of these data.

During July 1974 a total of 4 boats had been sampled from the 54 boats which had landed. One box of each of small, medium, large and extra large fish had been subsampled for length from each boat. A further subsample for age had been drawn, and estimates of numbers at age obtained.

Empirical variances for these estimates have been obtained by generating hypothetical samples which could have been drawn from the available landings. This was done as follows.

A hypothetical population of haddock of 10 age groups was defined such that length at age followed a Normal distribution. The population parameters (mean and variances of length within age groups and the proportions by which each age group is present in the population) were estimated from the routine data which had been collected during July 1974. These are given in Table 1. The total landings by each of the 54 boats were also available, given as number of boxes in each of the size categories.

At each cycle of the simulation 4 boats (ie 4 sets of boat landings) were selected at random from the 54 available. For each boat a box of small, medium, large and extra-large fish was generated using random numbers. The fish were then measured and subsampled for age. The length frequencies were raised using the data appropriate to the 4 boats selected to give an estimated length frequency for total landings. Applying the age-length key then produced estimates of the numbers landed at each age.

By repeating this process, an empirical distribution of the estimates of numbers at age was produced from which the sampling properties of the estimates could be deduced. The number of cycles was limited to 50, which although small, required some 4 hours computer time.

Conclusions and Results

In Table 2 are given the actual estimated numbers at age as observed in July 1974, the mean estimated numbers at age as calculated from the empirical distributions, the standard deviation of the estimated numbers at age as calculated from the empirical distributions, and the coefficients of variation.

As can be seen the coefficients of variation are highest for the 1973 brood, which at this stage is only beginning to be recruited to the commercial fishery. The coefficients of variation are lowest for the 1972 brood, reflecting the general rule that the greatest accuracy is associated with greatest numbers at age. The level of accuracy of the 3+'s to 8+'s is of the order of 20-25%.

These levels of accuracy are lower than those observed for the estimated numbers in a given size range (5%; Johnston et al 1974), but sampling for age is at a much lower level than for length.

A use of these estimates is the determination of total mortality rates. The above results can provide an indication of their reliability. If N_1 and N_2 are the estimated indices of abundance of a year class in successive

years then Z, the coefficient of mortality, can be estimated by $\log N_1/N_2$. It can easily be shown that the variance of this is approximately the sum of the squared coefficients of variation. For a coefficient of variation of 20%, the standard deviation of Z will be 0.28. This is higher than would be achieved in practice, since Z would seldom be estimated from as few as 2 observations.

Although this study is not intended to obviate the need for a full theoretical solution to this problem, it is felt that this approach might be used as above as an interim solution providing some insight into the reliability of estimates obtained by fish market sampling. A full theoretical solution is still being sought however, and it is hoped to present further results in a future contribution.

Reference

Johnston, C., Nicholson, M. D. and Pope, J. A. 1974 Fish Market Sampling in Scotland. ICES C.M. 1974/F:39.

Table 1: Population parameters for Haddock in July at Faroe Plateau.

Brood	Age	Percentage Composition	Mean length (cm)	Standard Deviation
1973	1+	0.09	25.5	2.23
1972	2+	65.30	35.4	3.01
1971	3+	13.94	43.6	3.82
1970	4+	11.36	48.3	4.23
1969	5+	1.66	55.7	4.87
1968	6+	2.21	57.8	5.06
1967	7+	2.42	59.7	5.22
1966	8+	2.20	60.7	5.31
1965	9+	0.41	60.9	5.33
1964 (and older)	10+	0.41	64.2	5.62

Table 2: Results from simulation study

Brood	Age	Numbers at Age July 1974	Estimated numbers at age	Standard deviation	%age Coefficient of Variation
1973	1+	1 305	987	1 147	116.2
1972	2+	923 993	961 832	58 754	6.1
1971	3+	197 233	208 786	35 616	17.1
1970	4+	160 732	163 111	23 499	14.4
1969	5+	23 505	23 287	7 936	34.1
1968	6+	31 315	31 667	7 552	23.8
1967	7+	34 290	35 899	9 240	25.7
1966	8+	31 188	32 161	7 467	23.2
1965	9+	5 730	5 037	3 314	65.8
1964	10+	5 769	6 298	3 324	52.8