Conseil International pour l'Exploration de la Mer

C.M. 1953
Contribution No. 17
(North Sea Sub-Committee)

The Sampling Variance in Measurements of Commercial Landings

By

J. A. Gulland

A knowledge of the length composition of commercial catches of fish is of great importance in studying any fish population and the influenceoof fishing upon it. The accuracy with which this can be estimated is influenced greatly by the methods used as well as the total number of fish measured. In particular a suitable choice of the number of fish in each separate sample is important.

If a large number of very small samples are used, then an obviously undue proportion of the time is spent in preparing to take the sample, rather than in actually measuring fish, but a very large sample can be equally wasteful. This is because the length distributions of different parts of the same ship's catch are generally much more alike, than those from different ships. Therefore, beyond a certain stage, measuring more fish from one ship will only increase the accuracy of the knowledge of the length-composition of that ship, and not of the landings as a whole. These can be more exactly shown in mathematical terms.

I will take as the quantity to be estimated the total number of fish landed of a certain length group, or, since the number of ships landing fish is known, the average number landed by each ship. The variance of their estimate can be calculated, and the effect of a change in sampling design shown. For the sake of simplicity I assume here that all ships land the same weight of fish.

Let n = number of samples taken.
m = number of fish in each sample.
Lx = true number of fish of required length in the xth sampled boat $x = 1, 2, 3 \ldots n$.
L = average number of fish of required length in each ship.
lx = estimate of Lx.
l = estimate of L.

$$\therefore l = \frac{1}{n} Sl_x \qquad \text{where } Sl_x \text{ denotes the sum of all the } l_x\text{'s.}$$

$$\therefore l = \frac{1}{n} S(lx - Lx) + \frac{1}{n} S(Lx - L) + L. \quad \ldots\ldots (1)$$

In the right hand expression of equation 1, the first term contains the errors due to sampling variations, and the second those due to differences between the ships. All these terms are independent, so that the variances of l can be expressed as the sum of the variances of the terms on the right of equation 1. If the length distributions in various ships do not differ greatly, the sampling variance in each will be the same, and will be inversely proportional to the size of sample,

i.e. variance $(lx - Lx) = \frac{1}{m} S_1{}^2$, and also, for the variance between ships
variance $(Lx - L) = S_2{}^2$

$\therefore$ from equation (1).

$$\text{variance } l = \frac{1}{n} \times \frac{1}{m} S_1{}^2 + \frac{1}{n} S_2{}^2 \quad \ldots\ldots (2)$$

This equation gives directly the accuracy with which the numbers of fish of a given length is known. In it S_1^2 and S_2^2 are constants determined by the composition of the catches, but m and n can be varied by the degree and design of the sampling. If the first term is much the larger, then the variance, and the accuracy, will be determined mostly by m x n, i.e. the total number of fish measured, but if the second is larger, then only increasing n (the number of samples) can appreciably increase the accuracy, and increasing m, the size of each sample gives no benefit. For a given number of fish measured the accuracy is greatest for samples consisting of only one fish, but so long as the first term in equation (2) is the larger it is more convenient, and involves little loss in efficiency, to increase the size of sample.

The relative sizes of the two terms in equation (2) can be calculated by taking two or more independent samples from a number of different ships and computing the variances within and between ships. This has been done at Lowestoft, taking two samples of 50 medium plaice from each of twelve ships. The results are shown in the form of analyses of variance in Table I.

Table 1 Analysis of Variance of Length Measurements

Length Group		30 - 34 cm.		35 - 39 cm.		40 - 44 cm.	
	D.F.	Sum of sq.	Mean sq.	Sum of sq.	Mean sq.	Sum of sq.	Mean sq.
Within ships	12	141	11.75	145	12.08	62.5	5.21
Between ships	11	965.5	87.77	622.8	56.62	88.5	8.04
Total	23	1106.5		767.83		151	
Average number		18.75		28.08		3.04	
Expected variance		11.71		12.31		2.86	

Besides the normal analysis of variance, Table 1 gives the average number of each length group. If there were no particular tendency for fish of the same size to occur together, and the sampling were truly random, then the numbers at each length would be distributed multinomialy. The variance for this distribution can be calculated at once, and this "expected variance" is given in the last row of Table 1.

For the two smaller length groups the within-ship variance differs little from the expected variance, so that it is reasonable to suppose that this variance is due mainly to errors of random sampling but the between-ship variance is much greater, the difference being significant for both length groups. This shows that there is a consistent difference in size between the catches of different ships. The values of S_1^2 and S_2^2 can be calculated at once, and are for the 30-34 cm., 35-39 cm. and 40-44 cm. length groups, 587.5 and 76.06; 604 and 44.54; 260.5 and 2.83 respectively; and the variances of the means of one sample from each ship are from equation (2).

$$\frac{1}{12} \cdot \frac{1}{50} \cdot 587.5 + \frac{1}{12} \cdot 76.06 = 0.98 + 6.34 = 7.32$$

$$\frac{1}{12} \cdot \frac{1}{50} \cdot 604 + \frac{1}{12} \cdot 44.54 = 1.01 + 3.71 = 4.72$$

$$\frac{1}{12} \cdot \frac{1}{50} \cdot 260.5 + \frac{1}{12} \cdot 2.83 = 0.43 + 0.24 = 0.67$$

for the three length-groups. The means with their standard deviations are therefore

$$18.75 \pm 2.7, \quad 28.08 \pm 2.2 \quad \text{and} \quad 3.04 \pm 0.82$$

In both the smaller length-groups, which comprise the bulk of the fish measured, the second term in the variance is much the bigger, showing that increased accuracy can only be obtained by increasing the number of samples. This is shown in Table 2 below for four different sampling designs. The table gives the variance of each length group, the total number of fish measured, and the time taken for each design. This last was calculated on the basis of taking seven minutes to measure 100 medium plaice, and that the time taken in moving to the extra samples has been added to all except the first design. On the Lowestoft market it takes about three minutes from finishing one sample to the start of the next.

Table 2 Variances of different sampling designs

Design	A	B	C	D
Nomber of samples	30	20	10	5
Number per sample	25	50	150	400
Total measured	750	1,000	1,500	2,000
Time taken (min.)	117.5	115	120	140
Variance of 30-34 group	3.32	4.39	8.00	15.40
Variance of 35-39 group	2.29	2.83	4.86	9.28
Variance of 40-44 group	0.44	0.40	0.46	0.70

Table II shows at once that the first two designs, though taking rather less time than the others are more accurate - very much so for the smaller sizes. The relative efficiency of two designs, for any length-group may be calculated as the ratio of the variances, for equal sampling time. Thus, for example, for the 30-34 cm. group design D is only 23% efficient relative to design B. That is, design B can be four times as accurate, for the same sampling effort, or can obtain the same accuracy with only a quarter of the effort.

Similar analyses have been made for other market categories of plaice, and for hake and cod, and these have also shown that the limiting size of sample, beyond which additional measuring gives little gain in accuracy, is small, as little as about 15-20 for some sizes and generally less than 50.