International Commission for the Northwest Atlantic Fisheries

Serial No. 3181
(D.c.9)

ICNAF Res.Doc.74/34

ANNUAL MEETING - JUNE 1974

Costs of Surveying Recruits to the Georges Bank Herring Fishery[1]

by

J. E. Reeves

National Marine Fisheries Service
Northeast Fisheries Center
Woods Hole, Massachusetts 02543

The purpose of this paper is to provide an indication of the cost of determining the abundance of the recruiting year class(es) to a fishery. Such determinations are necessary for accurate adjustment of catch quotas. For this purpose, an estimate is made of the annual dollar cost of surveying (with net hauls) the abundance of herring recruiting to the Georges Bank fishery. To accomplish this, certain information is needed about the total number of net hauls required to obtain the abundance estimate within given limits of error (e.g. ± 50%), the trawl hauls per day possible, and the cost per day for operating a research vessel. These items can then be combined to obtain a figure for the overall cost of a survey.

An estimate of the abundance of recruits can be obtained if we have information on the average density of fish per sample unit and on the total number of units inhabited by the population we want to estimate. We will assume that the estimate of density provides the only source of error to the estimate of abundance. We will also assume that the survey technique is unbiased. Then, the total number of trawl hauls needed can be related to the desired level of precision depending only on the variance of the density index of the recruits.

Information on the variance of density indices of Georges Bank herring recruits was obtained from data in Dornheim (1973). The extent of the population was defined as shown in Figure 1. Only those peripheral rectangles extending halfway or more into the area outlined were included in the analysis. Numbers per 30-minute haul in hauls that caught herring with mean lengths corresponding to three-year-olds (i.e. the recruiting 1970 year class) were used for calculation of the density and variance (see Table 1). Since these values predominated, it was assumed that zero catch hauls and hauls where no measurements were taken were representative of the distribution of the 1970 year class. Since the data are highly skewed (i.e. many hauls with low numbers/haul, and few hauls with high numbers/haul), the variance figure obtained from the data may not be very reliable for determining the number of hauls needed. To hopefully get a more representative estimate of the variance the data were transformed to their natural log values. The original data from Dornheim, arranged to show the skewness in distribution, and the resulting variance estimate is shown in Table 1. Using the estimate of the variance of ln numbers per tow, the number of trawl hauls needed to meet various requirements of precision is estimated according to a formula derived from Aitchison and Brown (1969). They show that an interval estimate of the true mean (density in this case) is

$$\alpha = a_1 \pm \frac{t a_1}{\sqrt{n}} \sqrt{S_y^2 + \frac{S_y^4}{2}} \quad \text{where}$$

[1] Revision of Res.Doc.74/34 presented to the Special Commission Meeting, FAO, Rome, January 1974.

a_1 is their estimate of the mean numbers per tow reconverted from the natural log scale and S^2_y is an estimate of the variance of ln numbers per tow (y = ln x, where x is the original density) and t is the Student's t statistic for a desired level of probability. Setting the second term on the right hand side equal to a d_1, where d is a proportion of the mean a_1, and solving for the sample size n gives

$$n = \frac{t^2 \, a_1^2 \left[S_y^2 + \frac{S_y^4}{2} \right]}{(a_1 d)^2}$$

For each level of precision sample sizes were calculated for two levels of significance and two powers of test. The power of test may be taken into account for calculating sample sizes by using multipliers supplied by Snedecor (1967, p. 113-114). The multipliers used are shown in Table 2 for the two levels of significance and power. These values have been substituted for t in the formula above to give

$$n = \frac{m \left(S_y^2 + \frac{S_y^4}{2} \right)}{d^2}$$

The "d" values required by the formula are specified percentages which correspond to the various error intervals of $\pm$ 50%, $\pm$ 25% and $\pm$ 10%. Thus, from Table 3 we see that to estimate the mean density at any given time for the area of Georges Bank and surrounding grounds (see Dornheim, 1973, Figure 1) 95% of the time (with a power of 80%) to within $\pm$ 50%, we need 1,592 trawl hauls during the survey; to within $\pm$ 25%, 6,363 hauls; and to within $\pm$10%, 39,772 hauls, etc. If we want to increase the power of the test to 90%, more sampling is required.

On the other hand, if we are willing to accept 90% confidence in our estimates (i.e., a higher level of significance), fewer samples are needed for a given power. The accuracy of these estimates of sample size depends to some degree upon how well the data fit the normal distribution. This assumption of normality is probably valid even for skewed data such as is being used here (Table 1) since it has been shown (Cochran 1963) that sample means (i.e., what we are trying to estimate) from skewed data tend to be normally distributed. However, to insure that normality is approached, a rule of thumb given by Cochran (p. 41) has been applied. This rule, for populations with positive skewness, is that the sample size (n) should be n $25G_1^2$ to approach normality. G_1 is a measure of skewness of the data. The calculations to obtain G_1 are shown in the appendix and indicate that a minimum of about 130 net hauls is required. It is obvious that the sampling requirements given in Table 3 are much in excess of this value.

Values of cost per day for estimating the total cost of a survey are taken from information on the U.S. R/V Albatross IV, a 187-ft, 1,000-ton diesel-powered craft used extensively in the Northwest Atlantic for fisheries surveys. Current data indicate a cost per day of about $3,000 to operate this ship. This figure includes costs of fuel, food, crew wages, and maintenance and repairs. The original cost of the vessel and depreciation and opportunity costs are not included in this figure. Thus, the figure of $3,000/day is on the conservative side.

The number of hauls that can be made in a day is primarily dependent upon survey logistics. The average number of hauls per survey day can be affected by weather, type of bottom, size of area to be surveyed (distance between stations), and size of catches among other things. Based upon experience of the Albatross IV a range of from 5 to 15 trawl hauls per day seems reasonable. Thus in Table 4 cost figures are presented for 5, 10 and 15 hauls per day for ± 50% and ± 25% precision levels. Sample levels required to achieve ± 10% precision are unrealistic and are not considered further (attaining ± 25% may also be unrealistic). We see that to obtain ± 50% precision in an abundance estimate, a survey could cost between $250,000 and $1,300,000 depending on the value specified for level of significance, power, and sampling rate. If we require more precision, say ± 25%, a considerable increase in expenditures is required.

As might be expected, the estimated variance plays a large part in determining the value of sample sizes, and in practical application is probably the largest source of error to such calculations. For example, if we are satisfied with the variance estimate obtained from the untransformed data (Table 1), we may use the well-known

$$n = \frac{t^2 \, S_x^2}{d^2}$$

(substituting m for t^2) to obtain sample sizes. In this case d is the estimated mean multiplied by a percentage corresponding to a given level of precision. Tables 5 and 6 give the results, which are markedly different from the results obtained using a log transform on the data.

In deciding which are the "right" numbers, it is useful to consider the purpose to which the data are to be put. For abundance estimates we are normally interested in the arithmetic mean density, rather than the geometric mean or some other transformed value. Thus, use of transformations for variance stabilization or meeting the assumption of normality (usually not critical for means) necessitates reconverting to the arithmetic mean by some indirect method. A better approach may be to deal directly with untransformed data and accept the initial variance estimates as approximate, with the hope of obtaining more reliable variances with continued sampling. In the case of the sample size estimates of Table 5, it should be pointed out the values for ± 50% precision are marginal in relation to the minimum sample size required for normality, as estimated previously by the Cochran "rule of thumb."

The cost per day figure of $3,000 is conservative. Thus, actual costs of a survey could be higher. In addition, costs of data processing and analysis after collection have not been considered here. Also, any error in the estimate of the total number of sample units occupied by the population being estimated will increase the number of tows required. On the other hand, a survey could be designed to estimate abundance of more than one species, thus reducing the survey costs per species. Also, it is likely that, after some experience with sampling the population, some stratified sampling scheme could be devised. Grosslein (1971) indicates that stratified sampling using net hauls can lead to considerable reduction in sampling error, which leads to lowered costs of surveying.

REFERENCES

Aitchison, J., and J. A. C. Brown, 1969. The Lognormal Distribution, with special reference to its uses in economics. Cambridge Univ. Press.

Cochran, W. G. 1963. Sampling Techniques, Second Edition. John Wiley & Sons, Inc., New York, 413 p.

Dornheim, H., 1973. A preliminary report on the German Young-Herring-Survey carried out by R/V Walther Herwig in ICNAF Subarea 5 and Statistical Area 6 in February-March 1973. ICNAF Res. Doc. 73/84, 4 p.

Grosslein, M. D., 1971. Some observations on accuracy of abundance indices derived from research vessel surveys. ICNAF Redbook 1971, Part III, 249-266.

Snedecor, W. S., and W. G. Cochran, 1967. Statistical methods. Sixth Edition. Iowa State University Press, Ames, Iowa, 593 p.

Table 1. Values of density of herring recruits (from Dornheim, 1973) and an estimate of the variance* from them, using the natural log transformation for variance stabilization.

Numbers of three-year-old herring per 30-minute haul

0	56	170	2850
0	62	202	3550
0	62	264	4950
0	64	363	5850
0	70	725	9450
3	92	1000	11750
6	98	1150	14000
12	149	1700	16250
28	155	2450	

	Untransformed scale	Converted to natural log scale(ln(x+1))
Mean density	2,215.2	5.09
Variance (σ^2)	17,859,703.68	9.084

*We will assume that the coefficient of variation (CV= standard deviation/mean) for numbers per haul is approximately the same as the CV for numbers per volume sampled by a 30-minute net haul (assuming the sample units are measured in volume), and also the same as the CV for final estimates of abundance.

Table 2. Multipliers used for sample size calculations which take the power of test (1- β , where β = type II error) into account, for two levels of significance and powers of test.

Level of significance	Power of test	m
.05	.80	7.9
.05	.90	10.5
.10	.80	6.2
.10	.90	8.6

Table 3. Number of 30-minute trawl hauls, based upon ln-transformed data, needed to estimate the average density of herring recruiting to the Georges Bank fishery at any given point in time.

Estimate of average density precise to:	d	$\frac{S_y^2 + S_{y/2}^4}{d^2}$	Level of significance	Power of test	Required number of hauls
± 50	.50	201.4	.05	.80	1,592
				.90	2,115
			.10	.80	1,249
				.90	1,732
± 25	.25	805.5	.05	.80	6,363
				.90	8,458
			.10	.80	4,994
				.90	6,927
± 10	.10	5034.4	.05	.80	39,772
				.90	52,861
			.10	.80	31,213
				.90	43,296

Table 4. Estimated costs of surveys, based upon ln-transformed data, to determine the abundance of herring recruiting to the Georges Bank fishery, using a cost per day of $3,000.

Estimate precise to:	Level of significance	Power of test	Total hauls required	Hauls per day	Survey cost ($)
± 50%	.05	.80	1592	5	955,200
				10	477,600
				15	318,400
		.90	2115	5	1,269,000
				10	634,500
				15	423,000
	.10	.80	1249	5	749,400
				10	374,700
				15	249,800
		.90	1732	5	1,039,200
				10	519,600
				15	346,400
± 25%	.05	.80	6363	5	3,817,800
				10	1,908,900
				15	1,272,600
		.90	8458	5	5,074,800
				10	2,537,400
				15	1,691,600
	.10	.80	4994	5	2,996,400
				10	1,498,200
				15	998,800
		.90	6927	5	4,156,200
				10	2,078,100
				15	1,385,400

B 6

Table 5. Number of 30-minute trawl hauls, based upon untransformed data, needed to estimate the average density of herring recruiting to the Georges Bank fishery at any given point in time.

Estimate of average density precise to:	Corresponding deviation from average density from table 1 (d)	$\frac{S_x^2}{d^2}$	Level of Significance	Power of test	Required number of hauls
± 50	1107.6	14.56	.05	.80	115
				.90	153
			.10	.80	90
				.90	125
± 25	553.8	58.23	.05	.80	460
				.90	611
			.10	.80	361
				.90	501
± 10	221.5	364.02	.05	.80	2876
				.90	3822
			.10	.80	2257
				.90	3131

Table 6. Estimated costs of surveys, based upon untransformed data, to determine the abundance of herring recruiting to the Georges Bank fishery, using a cost per day of $3,000.

Estimate precise to:	Level of significance	Power of test	Total Hauls required	Hauls per day	Survey cost ($)
± 50	.05	.80	115	5	69,000
				10	34,500
				15	23,000
		.90	153	5	91,800
				10	45,900
				15	30,600
	.10	.80	90	5	54,000
				10	27,000
				15	18,000
		.90	125	5	75,000
				10	37,500
				15	25,000
± 25	.05	.80	460	5	276,000
				10	138,000
				15	92,000
		.90	611	5	366,600
				10	183,300
				15	122,200
	.10	.80	361	5	216,600
				10	108,300
				15	72,200
		.90	501	5	300,600
				10	150,300
				15	100,200
± 10	.05	.80	2876	5	1,725,600
				10	862,800
				15	575,200
		.90	3822	5	2,293,200
				10	1,146,600
				15	764,400
	.10	.80	2257	5	1,354,200
				10	677,100
				15	451,400
		.90	3131	5	1,878,600
				10	939,300
				15	626,200

Appendix. Calculation of **minimum sample size** needed to approach normality with the data given in table 1.

Interval	Coded scale yi	Frequency f_i	$f_i y_i$	$f_i y_i^2$	$f_i y_i^3$
0 - 1000	1	24	24	24	24
1001 - 2000	2	2	4	8	16
2001 - 3000	3	2	6	18	54
3001 - 4000	4	1	4	16	64
4001 - 5000	5	1	5	25	125
5001 - 6000	6	1	6	36	216
6001 - 7000	7	0	0	0	0
7001 - 8000	8	0	0	0	0
8001 - 9000	9	0	0	0	0
9001 - 10000	10	1	10	100	1000
10001 - 11000	11	0	0	0	0
11001 - 12000	12	1	12	144	1728
12001 - 13000	13	0	0	0	0
13001 - 14000	14	1	14	196	2744
14001 - 15000	15	0	0	0	0
15001 - 16000	16	0	0	0	0
16001 - 17000	17	1	17	289	4913
Totals		35	102	856	10884

$$\bar{Y} = \frac{102}{35} = 2.91429 \qquad \kappa_3 = E(y_i - \bar{Y})^3 = E(y_i^3) - 3E(y_i^2)\bar{Y} + 2\bar{Y}^3$$

$$E(y_i^2) = 856/35 = 24.45714 \qquad = 310.97143 - 213.82560 + 49.50263$$

$$E(y_i^3) = 10884/35 = 310.97143 \qquad = 146.64846$$

$$\sigma^2 = E(y_i^2) - \bar{Y}^2 = 15.96405 \qquad G_1 = \frac{\kappa_3}{\sigma^3} = \frac{146.64846}{63.78442} = 2.3$$

$$n > 132$$

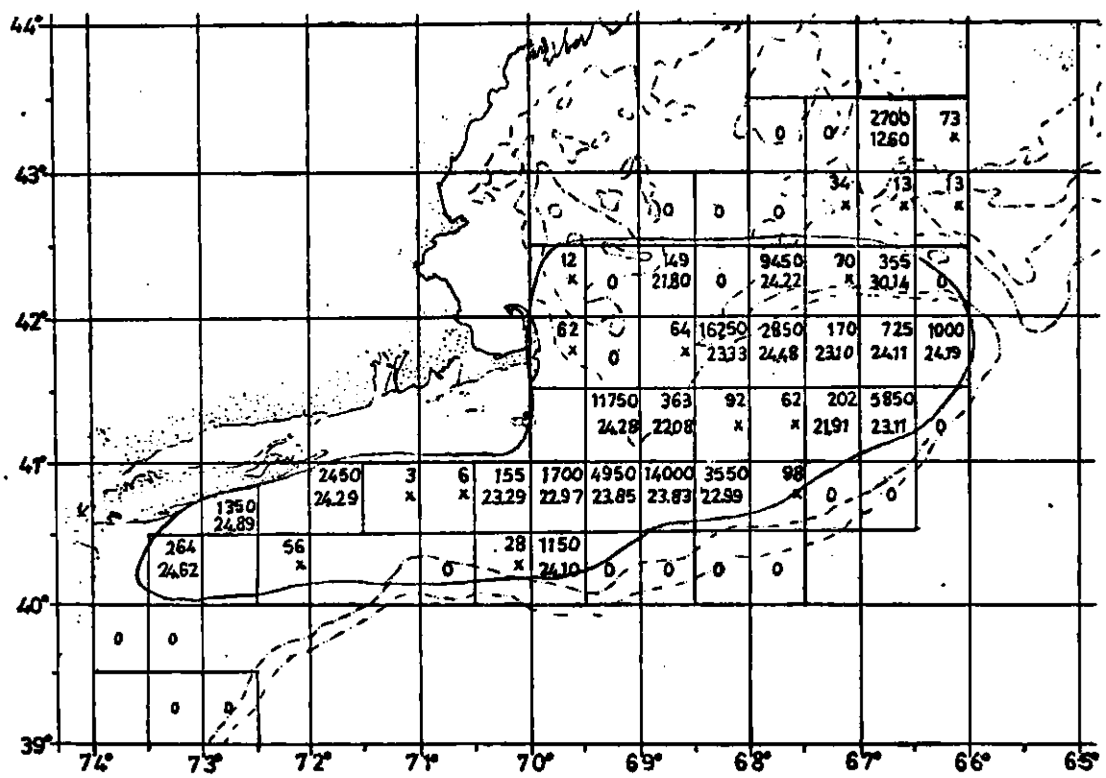

Fig. 1. Distribution of herring during the Young Herring Survey of *R/V Walther Herwigin* February - March 1973.
Upper figure: numbers of herring caught per 30 min.
Lower figure: mean length (cm).
X: no mean length calculated because less than 100 herring.

(From Dornheim, 1973)