NOT TO BE CITED WITHOUT
PRIOR REFERENCE TO AUTHOR(S)

International Commission for the Northwest Atlantic Fisheries

Serial No. 4040
(D.c.3)

ICNAF Res.Doc. 76/XII/144

NINTH SPECIAL COMMISSION MEETING-DECEMBER 1976

Using the USA Research Vessel Spring Bottom Trawl Survey as an Index of Atlantic Mackerel Abundance

by

M. P. Sissenwine
National Marine Fisheries Service
Northeast Fisheries Center
Woods Hole Laboratory
Woods Hole, Massachusetts 02543 USA

Publ. in ICNAF Sel. Papers No. 3: 49-56

Abstract

The catch frequency distributions for mackerel from the USA research vessel spring bottom trawl survey were examined and found to be highly skewed. A linear relationship (on a log-log scale), with a slope of approximately two, between the means and variances of a series of samples from the survey indicate that a log transform is appropriate in order to normalize the distribution of survey catches. The log transformation was tested by simulating sampling from hypothetical cumulative distribution functions which were based on observed survey frequency distributions. These simulations indicate that the mean of log transformed catches has a symmetric distribution with a strong central tendency.

Introduction

The use of the USA research vessel spring bottom trawl survey as an index of Atlantic mackerel (*Scomber scombrus*) abundance was most recently proposed by Anderson (1976). Because mackerel are a pelagic schooling species, catches of mackerel in the survey have been more sporadic than catches of most common demersal species. Therefore, special consideration of the frequency distribution of catches from the survey and of the distributional characteristics of indices based on this data is warranted. In this paper, catch frequency distributions from the survey are presented, an appropriate transformation is derived and the distributional properties of several indices are investigated by Monte Carlo simulation.

Method

USA research vessel spring bottom trawl surveys have been conducted between Cape Hatteras and Nova Scotia each year since 1968. The survey has a stratified random sampling design where strata are selected on the bases of depth and area. The strata are shown in Figure 1. A No. 36 Yankee otter trawl was used prior to 1973 and is still used in autumn surveys. Since 1973, spring surveys have used a modified No. 41 Yankee high-opening bottom trawl. More details of the surveys are given by Grosslein (1969).

Mackerel abundance is best reflected by tows from strata 1-25 and 61-76. Spring bottom trawl survey catches of mackerel in the Gulf of Maine and further north are rare. About 200 tows are made in these strata during each spring survey. The mean ($\overline{X}$) and variance (S^2) of the population from which these samples are drawn are estimated by:

$$\overline{X} = \frac{1}{A} \sum_h A_h \overline{X}_h \qquad (1)$$

$$S^2 = \frac{1}{A}\left[\sum_h A_h (\overline{X}_h)^2 - A(\overline{X})^2\right] + \frac{1}{A} \sum_h \left[(A_h - 1) + \left(\frac{A_h - A}{A}\right)\left(\frac{A_h - n_h}{n_h}\right) S_h^{\,2}\right] \qquad (2)$$

where A is the area of all strata considered, A_h is the area of strata h, $\overline{X}_h$ and S_h^2 are strata means and variances and n_h is the strata sample size. Equation (2) was derived by J. A. Brennan (Northeast Fisheries Center, personal communication).

Taylor (1961) empirically derived the "Power Law." Taylor's power law states that the variance of a population is proportional to a fractional power of its mean:

$$S^2 = a\ \overline{X}^b \tag{3}$$

$$\log S^2 = \log a + b \log \overline{X} \tag{4}$$

The parameter a depends chiefly upon the size of the sampling unit. Parameter b is an index of dispersion and varies from 0 for a regular distribution to 1 for a random distribution with values greater than 1 indicating a contiguous distribution. The slope of the linear relationship between the log S^2 and log $\overline{X}$ for a series of samples is an estimate b. Once b is estimated, a common transformation can be applied to the catch from each tow thus stabilizing the sample variance and allowing the application of normal statistics. The appropriate transformation is to replace the catch from each tow (X) by X^p where $p = 1- {}^b/2$. When p = 0, a log transformation should be used (Elliot, 1971). The slope of the relationship between the $\log_e S^2$ and $\log_e \overline{X}$ for the series of spring surveys between 1968-1976 was used to estimate b. Sample strata means and variances for strata in which at least four tows were made in a specific spring survey were also used to estimate b. Since the variability of a population within strata is likely to be lower than the variability over all strata combined, the latter set of data may underestimate b.

Once an appropriate transformation was derived, the distributional properties of the sample mean of transformed data were investigated using a Monte Carlo simulation. Monte Carlo simulation was recently used by Barrett and Goldsmith (1976) in a similar manner to investigate the distribution of sample means drawn from nonnormal distributions. For a brief but general discussion of the method of Monte Carlo simulation, see Gordon (1969). The specific application used in this work is discussed below.

The cumulative distribution function (F (X)) of a population defines the probability of an observed value from that population being less than X. Therefore, F(X) is nondecreasing and ranges from 0 to 1 for all populations.

$$F\ (X) = \int_{-\infty}^{X} f\ (X)dx \tag{5}$$

where F(X) is a probability density function. Then if u is a uniformly distributed random variable between 0 and 1 and the equation u = F (X) is solved for X, this value of X will have F (X) as its probability density function. Therefore, random numbers with a desired distribution can be generated from uniformly distributed random numbers based on a cumulative distribution function. For an empirically determined cumulative distribution function, the transformation is most easily accomplished graphically. If F(X) is plotted on the vertical axis and X on the horizontal axis, then for a random value of u, the corresponding value of X is determined by projecting the point on the curve with level u on the vertical axis to the horizontal axis. When a series of points from a cumulative distribution function is known, then the transformation is easily accomplished by linearly interpolating between points. Uniform random numbers between 0 and 1 can be generated by a subroutine written by IBM (1970).

A program was written to generate random numbers distributed according to a specified cumulative distribution. Several cumulative distributions were considered some of which corresponded to observed catch frequencies of mackerel from the spring bottom trawl surveys. Samples of 25, 100 and 200 values were generated and 100 samples of each sample size were considered for each distribution. Frequency diagrams of sample means of transformed and untransformed data were prepared in order to compare their distribution characteristics under various circumstances.

Results and Discussion

The catch frequencies in weight of mackerel in spring surveys from 1968-1976 are given in Table 1. The weight of mackerel caught is considered instead of the number of mackerel since Anderson's (1976) index is based on weight. Weight intervals in which there were no occurrences are omitted from

Table 1. Catch frequency diagrams for three typical years are given in Figures 2-4. The skewness of sample distributions is obvious. From 63 to 92% of the tows in each year had no mackerel. The highest catch was greater than 5,000 kg (occurring in 1973), more than 20 times the catch in any other tow. This catch approached a physical limit (on catch in any single tow) thus indicating that the catch distribution is finite with a finite variance.

The $\log_e(\bar{X})$ and $\log_e(S^2)$ for each year 1968-1976 are plotted in Figure 5. These variables are also plotted for each strata in which four or more tows were made in a single year in Figure 6. The points in Figure 6 from surveys prior to 1973 were adjusted to values equivalent to expected catches using a No. 41 net as described by Anderson (1976). The points marked by x's in Figure 5 were adjusted upward to correspond to the No. 41 net, but these were not considered in fitting the regression line. It is obvious that the line fits the points based on actual catches and adjusted catches equally well. The slope of both regression lines is approximately two (1.96 and 1.91 for Figures 5 and 6, respectively) and therefore p approaches 0 indicating a log transformation. If the slopes are interpreted exactly, the appropriate transforms would be $X^{0.02}$ or $X^{0.04}$. Both the log transform ($\log_e X + 1.0$) and an exponential transform ($X^{0.02}$) were considered in the simulations described below.

Two hypothetical cumulative distribution functions are shown in Figure 7. These distributions are indicated by 1973 and 1975 survey samples. The 1973 and 1975 samples are representative of high and low levels of mackerel abundance in the survey, respectively. The means of these distributions are 26.2 and 0.65. Simulated samples of 25, 100, and 200 were drawn from each population. The means of the untransformed and transformed samples were calculated and the procedure was repeated 100 times. Frequency distributions of sample means are plotted in Figures 8-11. The means of exponentially transformed samples are not shown, but their frequency distributions are similar in shape to log transformed means. Therefore, there is no benefit in using the less common exponential transformation.

According to the central limit theory, the probability density function of the sample mean converges to a normal distribution as sample size increases when the population from which the sample is drawn has a finite variance. Usually, a sample size of 30-60 is assumed adequate for a sample mean to approach a normal distribution, but this may not be true for samples from a highly skewed population. The symmetry and central tendency of the normal distribution are highly desirable characteristics for an abundance index based on random sampling.

For both empirical distributions, the arithmetic means are highly variable and asymmetric for a sample size of 25 and 100. For the low abundance (1975) distribution and a sample size of 200, the arithmetic means appear more symmetric and exhibit some central tendency. This does not occur for the high abundance (1973) distribution even with a sample size of 200.

Log transformed means are less variable and more symmetric than untransformed means for all three sample sizes and both distributions. The log transformation appears adequate to allow the application of normal statistics to survey means for sample sizes of 100-200.

Some words of caution on the use of means of log transformed catches from the bottom trawl survey are necessary. Firstly, while the work reported in this paper does indicate that the means of log transformed catches have desirable characteristics, the use of this statistic as an index of abundance depends on the assumption that the survey randomly samples the population. For mobile populations with oriented movements, sampling should be random in both time and space. In fact, it is virtually impossible to randomize the order in which randomly selected stations are sampled.

Secondly, it is important to note that catch frequency distributions with vastly different means might have the same log transformed means. Therefore, it may not always be possible to distinguish between some populations with different means when using a log transformed index even with an extremely large sample size.

In conclusion, the log transformation does tend to normalize the sample means of mackerel catches from spring research vessel bottom trawl surveys. The significance of nonrandom temporal sampling and of the relationship between transformed and untransformed means of a population should be considered further.

References

Anderson, E. D. 1976. Measures of abundance of Atlantic mackerel off the Northeast Coast of the United States. ICNAF Research Bulletin No. 12: 5-22.

Barrett, J. P., and L. Goldsmith. 1976. "When is n sufficiently large?" American Statistician 30(2): 67-70.

Elliott, J. M. 1971. Some methods for the statistical analysis of samples of benthic invertebrates. Freshwater Biological Association, Special Publication No. 25, 144 pp.

Gordon, G. 1969. System Simulations. Prentice-Hall, Inc., New Jersey, 303 pp.

Grosslein, M. D. 1969. Groundfish survey program at BCF Woods Hole. Commercial Fisheries Review 31(8-9): 22-25.

IBM. 1970. System 360 scientific subroutine package version III programmer's manual. No. GH20-0205-4.

Taylor, L. R. 1961. Aggregation, variance and the mean. Nature, Lond. 189: 732-735.

Table 1. Catch (in kilograms) frequency of mackerel in spring bottom trawl surveys in strata 1-25 and 61-76 for 1968 - 1976.

Year	Weight (kg)	No. of Tows	% of Tows	Cumulative %
1968	0	142	78.9	78.9
	0-0.5	10	5.6	84.5
	0.5-1	4	2.2	86.7
	1-2	2	1.1	87.8
	2-3	5	2.8	90.6
	3-4	2	1.1	91.7
	4-5	2	1.1	92.8
	5-10	5	2.8	95.6
	10-100	6	3.2	98.8
	175-176	1	.6	99.4
	228-229	1	.6	100.0
1969	0	168	91.5	91.5
	0-0.5	10	5.5	92.0
	0.5-1	2	1.0	98.0
	1-2	1	.5	98.5
	2-3	2	1.0	99.5
	14-15	1	.5	100.0
1970	0	135	69.6	69.6
	0-0.5	12	6.2	75.8
	0.5-1	18	9.3	85.1
	1-2	6	3.1	88.2
	2-3	4	2.1	90.3
	3-10	6	3.1	93.4
	10-30	8	4.2	97.6
	30-100	4	2.1	99.7
	120-121	1	.3	100.0
1971	0	146	76.8	76.8
	0-0.5	9	4.7	81.5
	0.5-1	13	6.8	88.3
	1-2	6	3.2	91.5
	2-5	7	3.7	95.2
	5-10	2	1.1	96.3
	10-20	2	1.1	97.4
	40-70	4	2.1	99.5
	214-215	1	.5	100.0
1972	0	145	74.0	74.0
	0-0.5	17	8.7	82.7
	0.5-1	14	7.1	89.8
	1-5	9	4.6	94.4
	5-10	3	1.6	96.0
	10-20	4	2.0	98.0
	20-100	4	2.0	100.0

Table 1. (continued)

Year	Weight (kg)	No. of Tows	% of Tows	Cumulative %
1973	0	149	69.0	69.0
	0-0.5	25	11.6	80.6
	0.5-1	20	9.3	89.9
	1-2	8	3.6	93.5
	2-5	5	2.3	95.8
	5-10	2	.9	96.7
	10-50	4	1.8	98.5
	72-73	1	.5	99.0
	194-195	1	.5	99.5
	5182-5183	1	.5	1000
1974	0	101	63.5	63.5
	0-0.5	26	16.4	79.9
	0.5-1	15	9.4	89.3
	1-2	5	3.1	92.4
	2-3	3	1.9	94.3
	3-5	2	1.3	95.6
	5-6	3	1.9	97.5
	10-1	1	.6	98.1
	50-100	3	1.9	100.0
1975	0	158	79.3	79.3
	0-0.5	15	7.6	86.9
	0.5-1	13	6.6	93.5
	1-2	5	2.5	96.0
	2-10	3	1.5	97.5
	15-16	1	.5	98.0
	25-35	4	2.0	100.0
1976	0	152	79.7	79.7
	0-0.5	2	1.0	80.7
	0.5-1	23	12.0	92.7
	1-2	7	3.7	96.4
	2-10	3	1.6	98.0
	15-16	1	.5	98.5
	29-30	1	.5	99.0
	69-70	1	.5	99.5
	78-79	1	.5	100.0

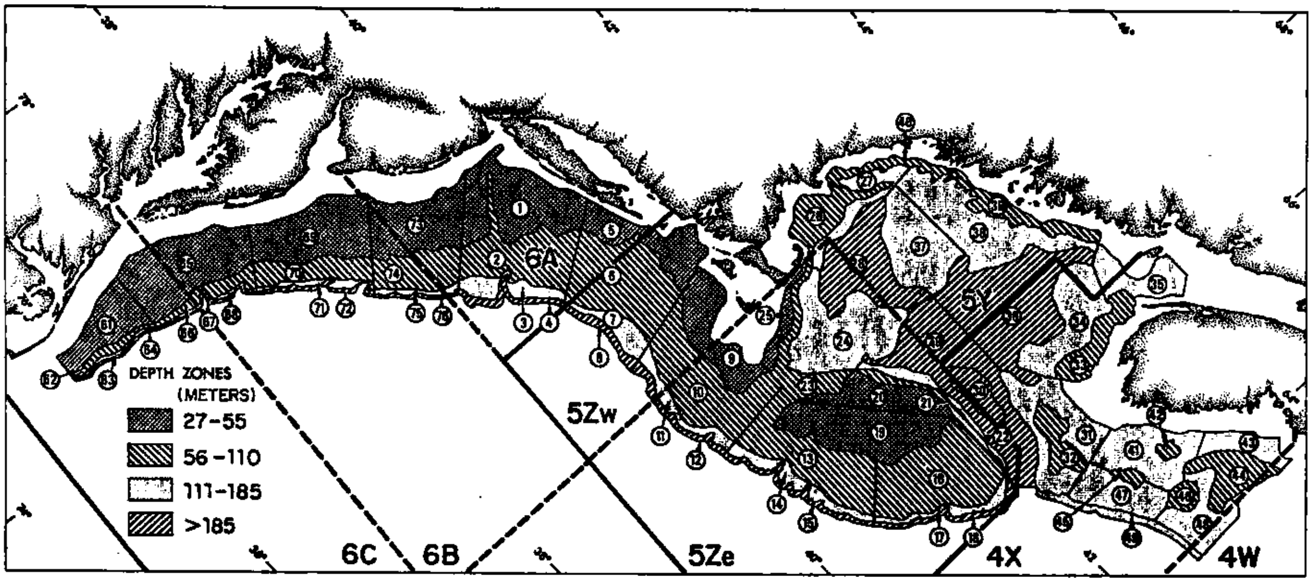

Fig. 1. US research vessel bottom trawl survey strata.

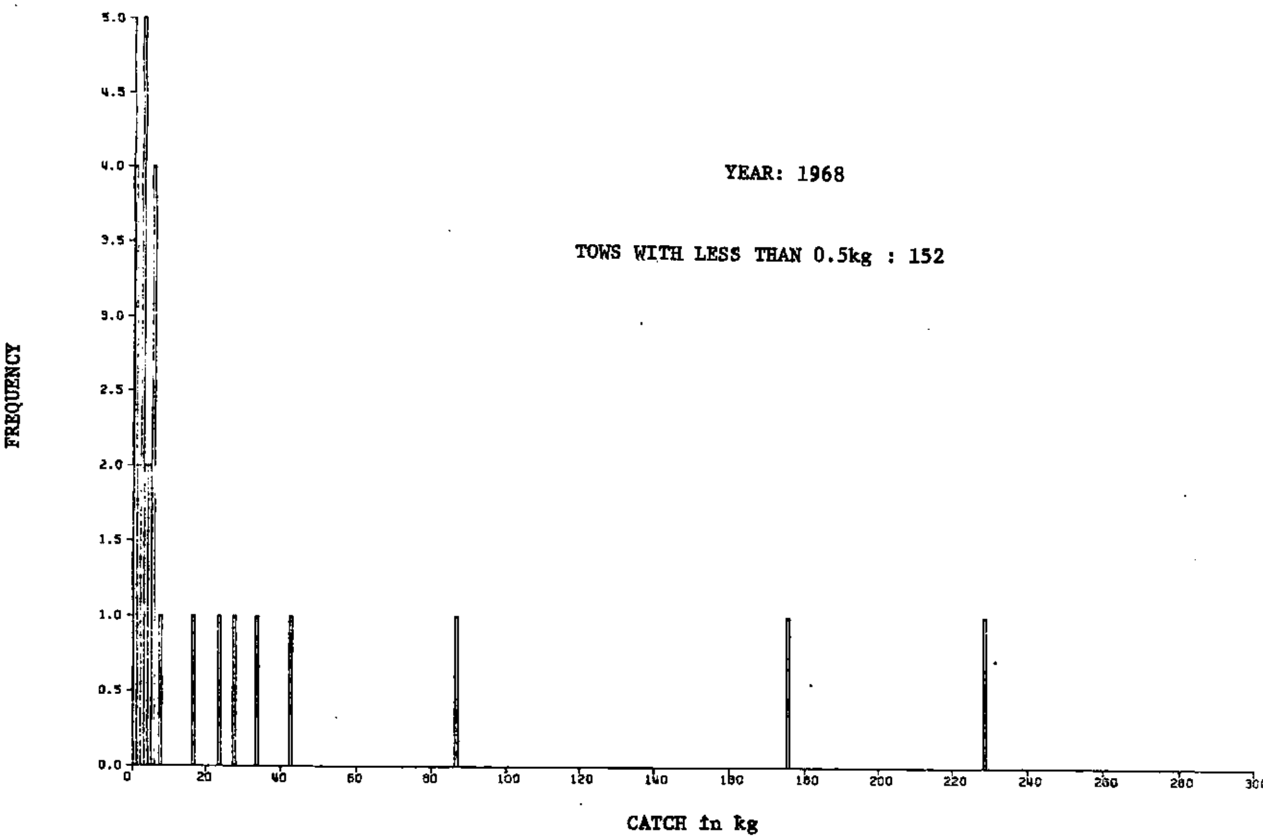

Fig. 2. Catch frequency from 1968 spring bottom trawl survey.

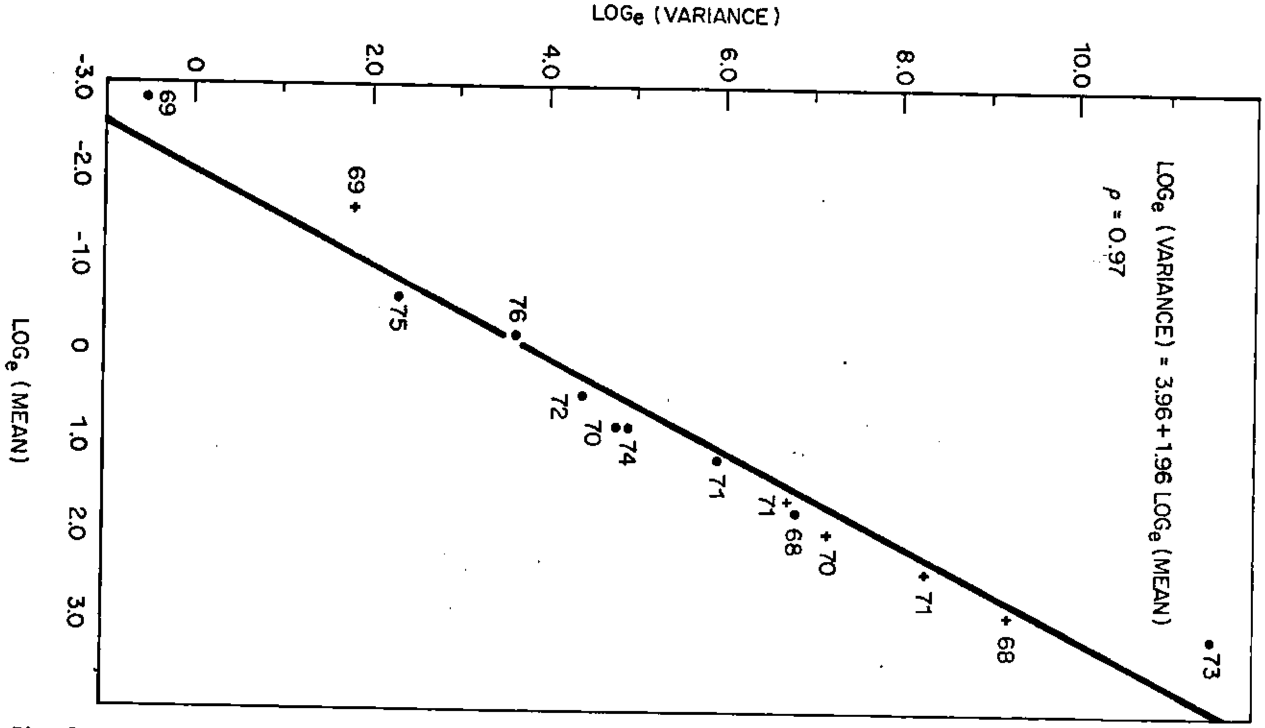

Fig. 5. The relationship between annual survey sample means and variances. Dots are actual sample statistics while x's are adjusted to expected values of catch using the No. 41 net.

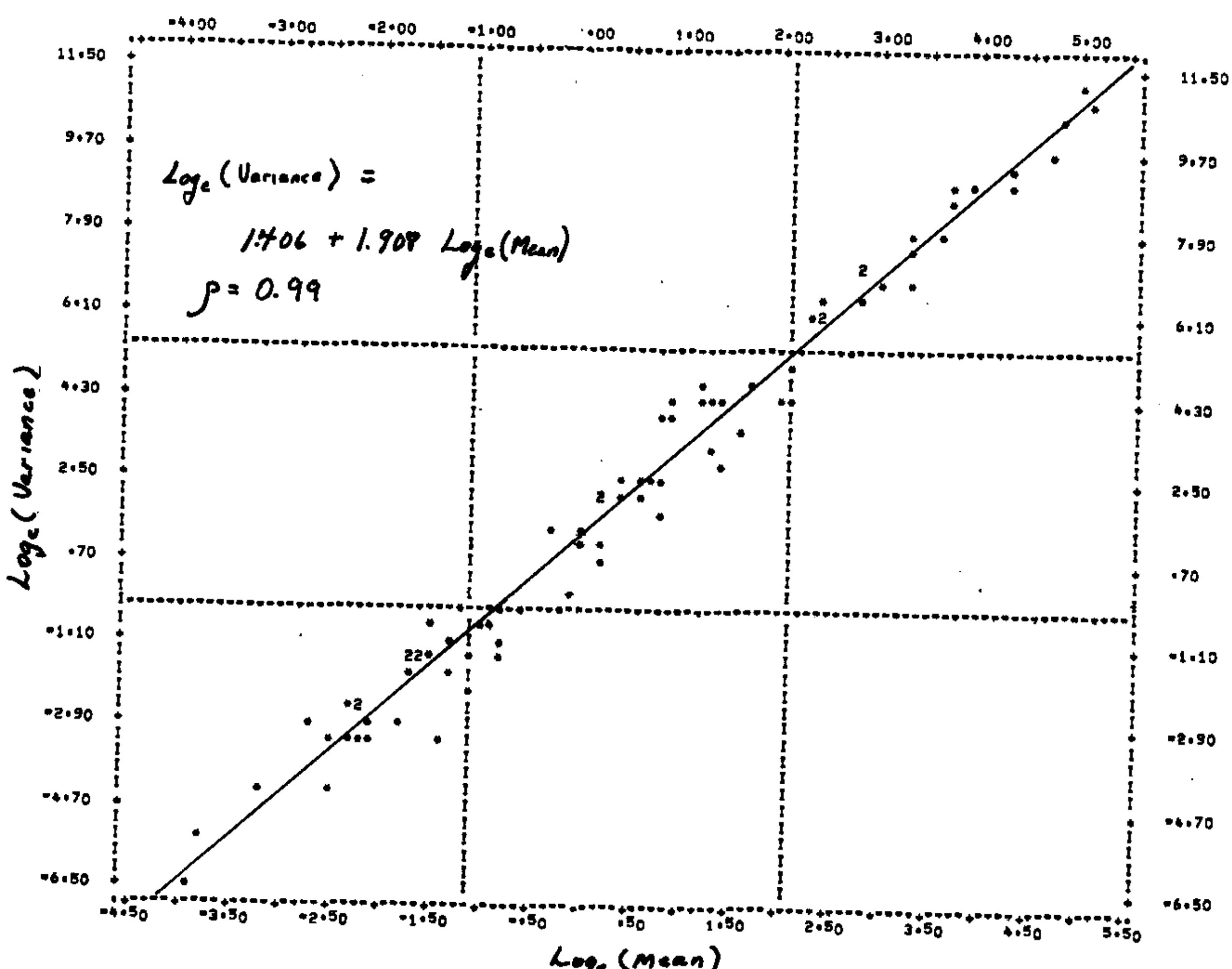

Fig. 6. The relationship between strata means and variances.

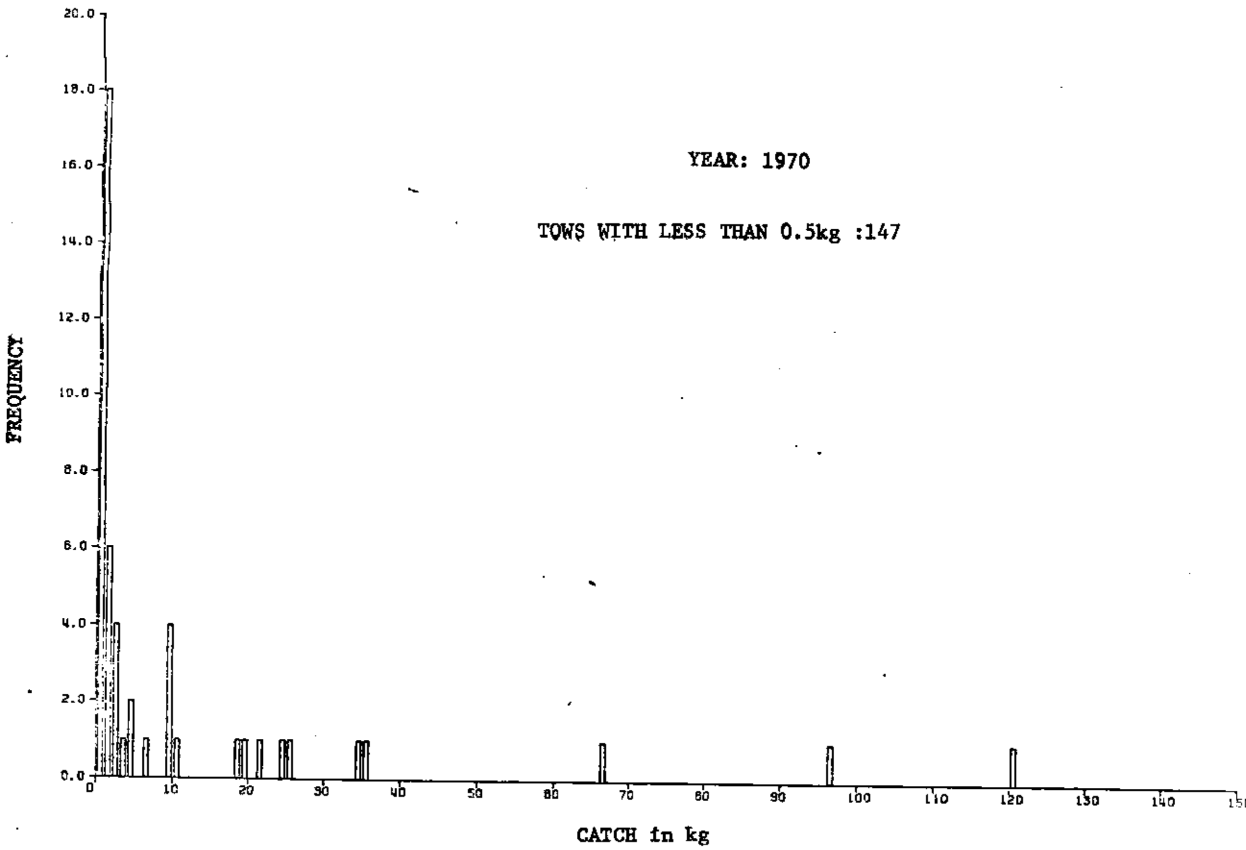

Fig. 3. Catch frequency from 1970 spring bottom trawl survey.

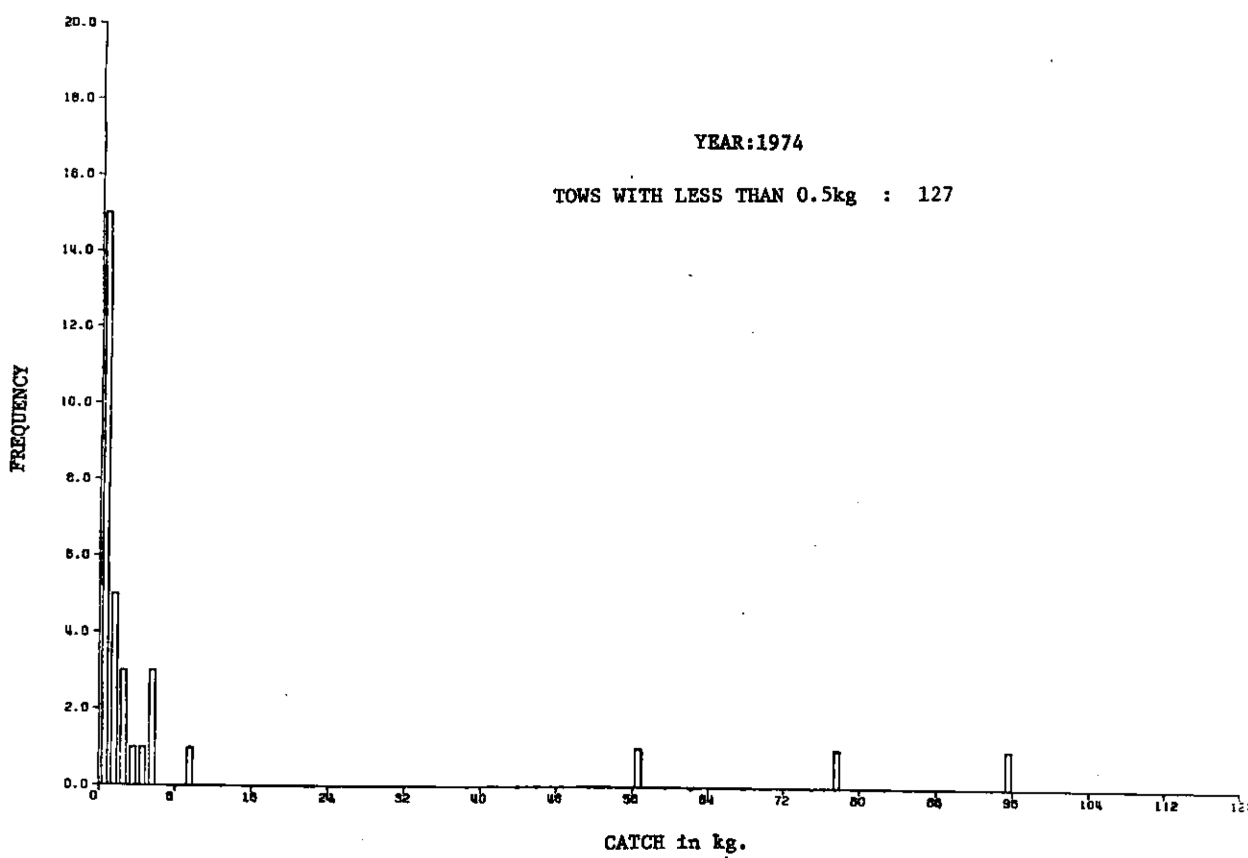

Fig. 4. Catch frequency from 1974 spring bottom trawl survey.

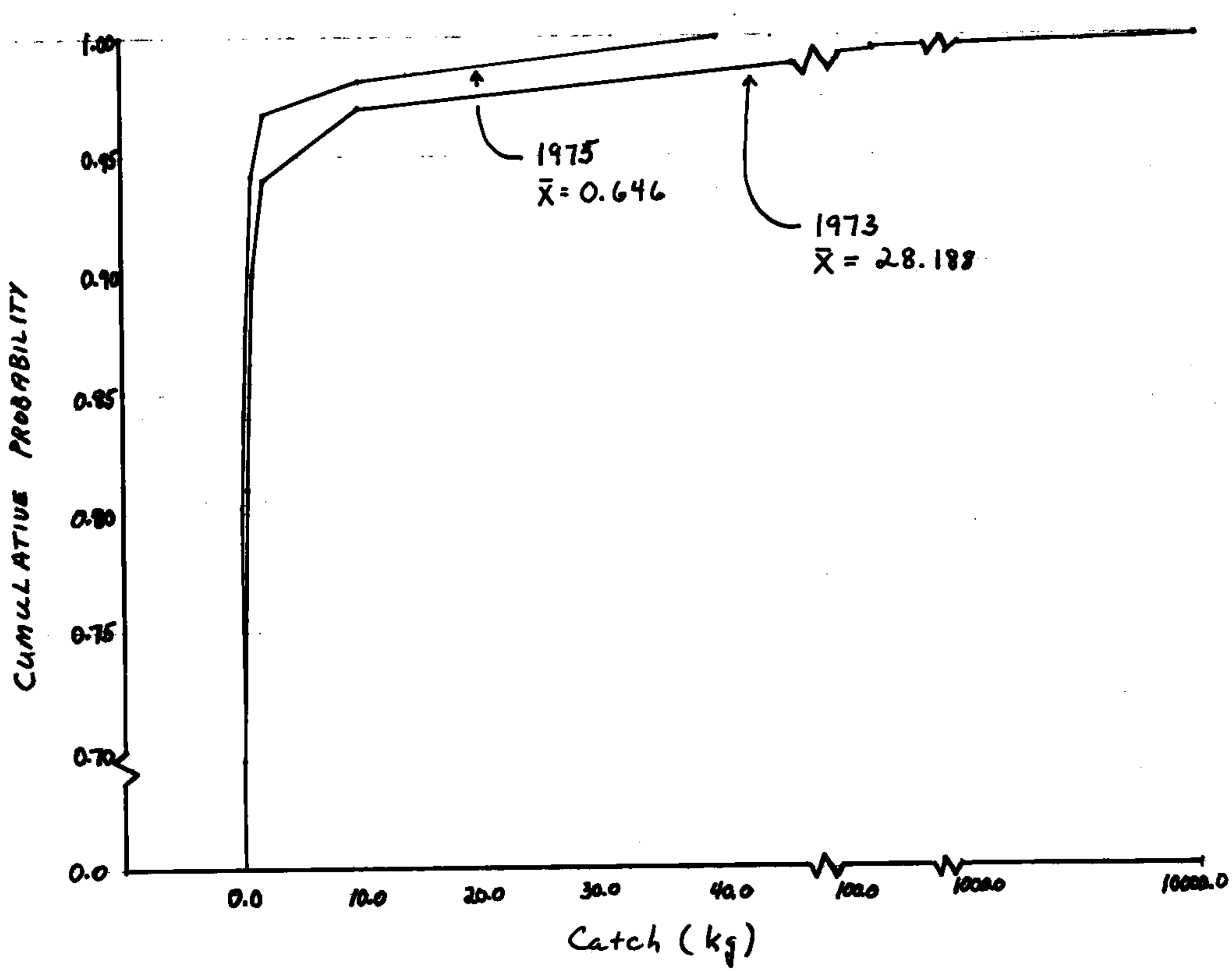

Fig. 7. Hypothetical cumulative distribution functions based on 1973 and 1975 catches of mackerel in US spring bottom trawl surveys.

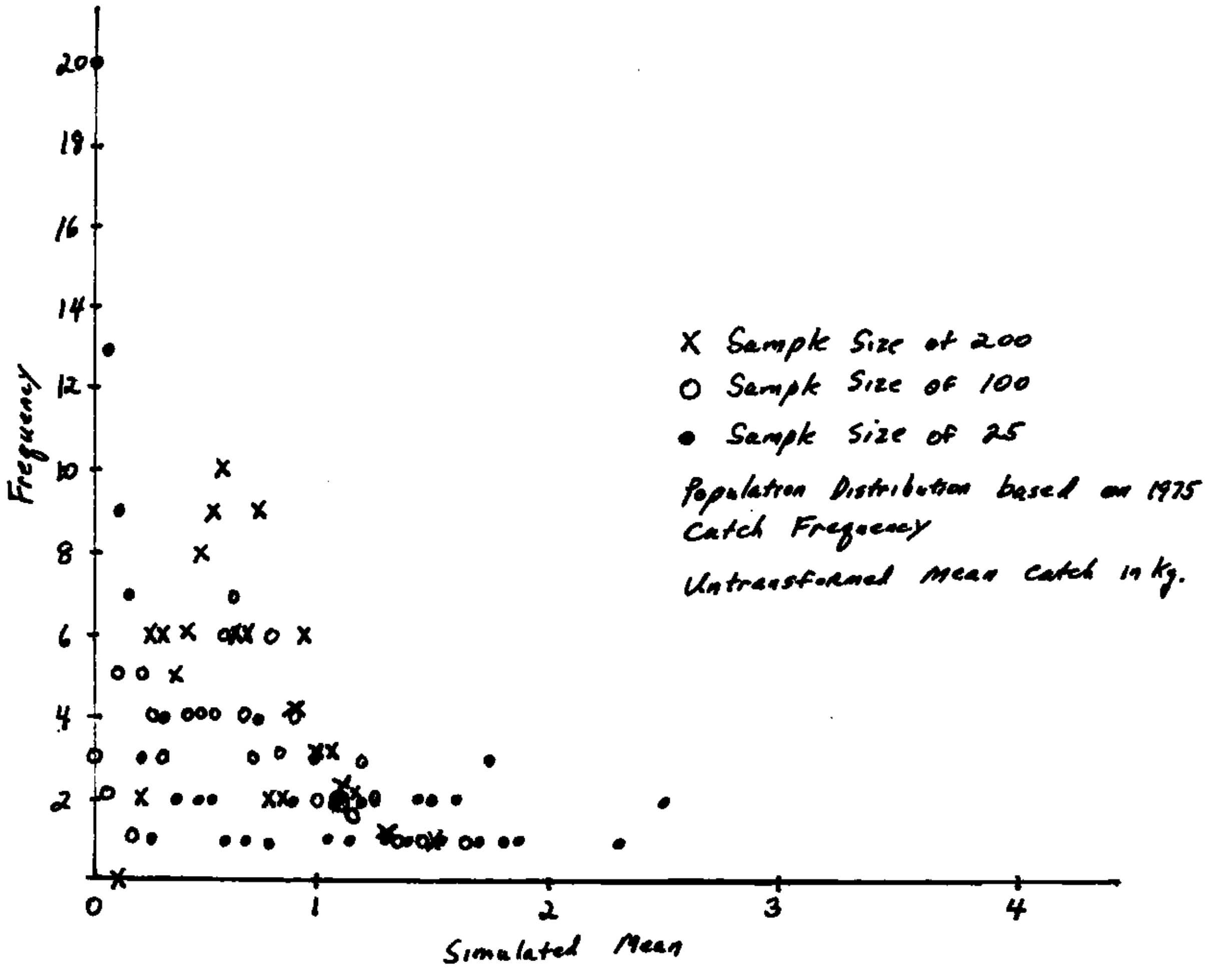

Fig. 8. Frequency distribution of simulated untransformed means.

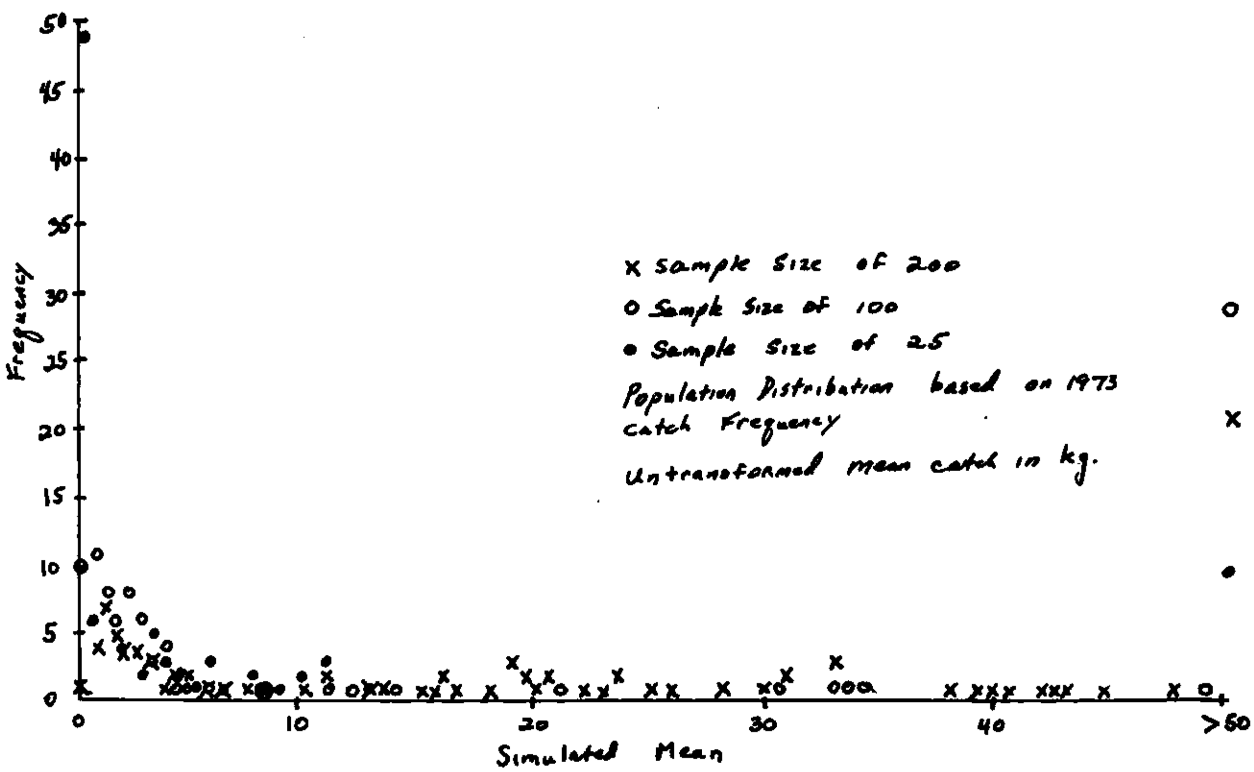

Fig. 9. Frequency distribution of simulated untransformed means.

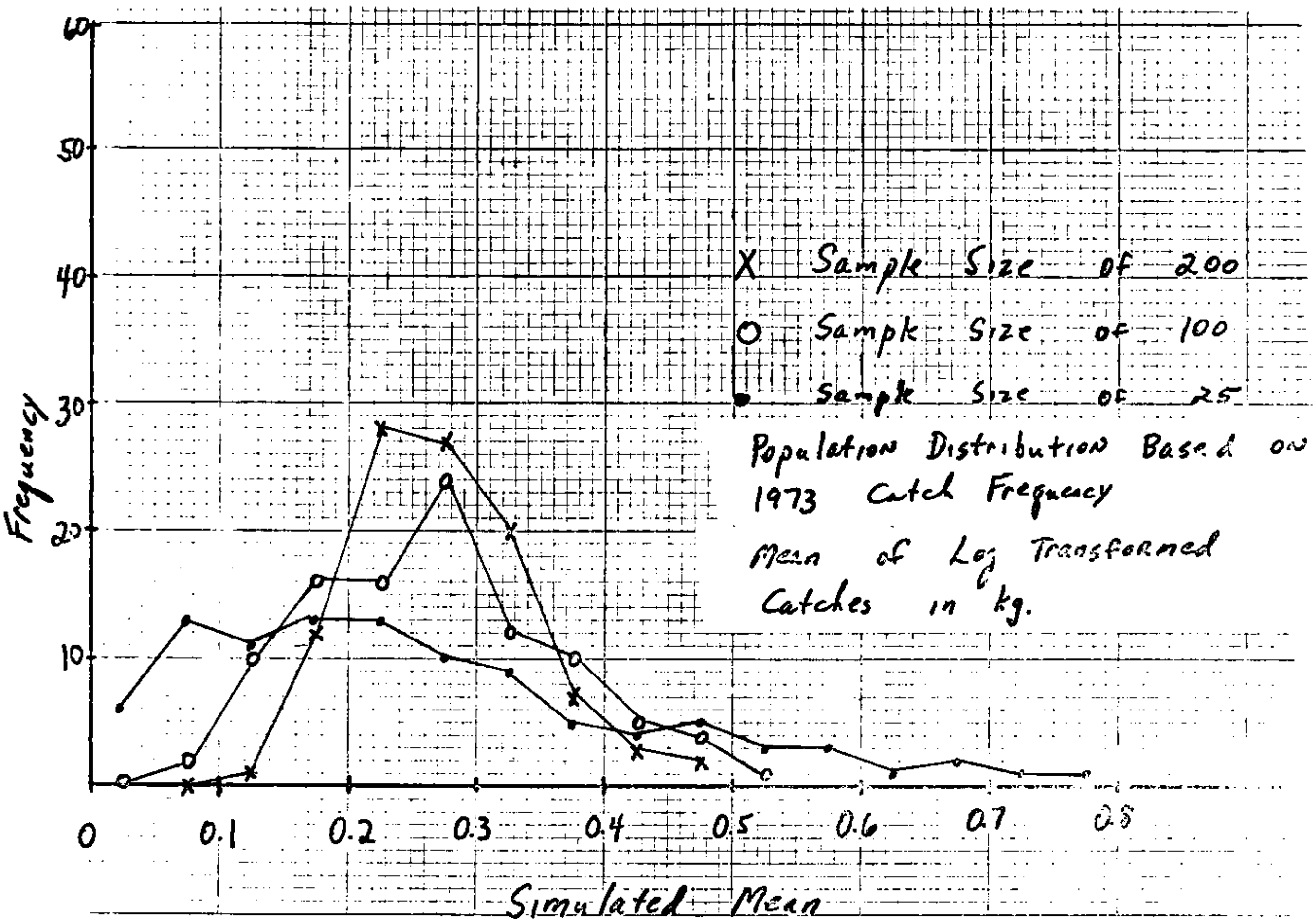

Fig. 10. Frequency distribution of simulated log transformed means.