Sampling Statistics in the Atlantic Menhaden Fishery

SEDAR27-RD-07

9

NOAA Technical Report NMFS 9

Sampling Statistics in the Atlantic Menhaden Fishery

Alexander J. Chester

August 1984

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
National Marine Fisheries Service

NOAA TECHNICAL REPORTS NMFS

The major responsibilities of the National Marine Fisheries Service (NMFS) are to monitor and assess the abundance and geographic distribution of fishery resources, to understand and predict fluctuations in the quantity and distribution of these resources, and to establish levels for optimum use of the resources. NMFS is also charged with the development and implementation of policies for managing national fishing grounds, development and enforcement of domestic fisheries regulations, surveillance of foreign fishing off United States coastal waters, and the development and enforcement of international fishery agreements and policies. NMFS also assists the fishing industry through marketing service and economic analysis programs, and mortgage insurance and vessel construction subsidies. It collects, analyzes, and publishes statistics on various phases of the industry.

The NOAA Technical Report NMFS series was established in 1983 to replace two subcategories of the Technical Reports series: "Special Scientific Report—Fisheries" and "Circular." The series contains the following types of reports: Scientific investigations that document long-term continuing programs of NMFS, intensive scientific reports on studies of restricted scope, papers on applied fishery problems, technical reports of general interest intended to aid conservation and management, reports that review in considerable detail and at a high technical level certain broad areas of research, and technical papers originating in economics studies and from management investigations.

Copies of NOAA Technical Report NMFS are available free in limited numbers to governmental agencies, both Federal and State. They are also available in exchange for other scientific and technical publications in the marine sciences. Individual copies may be obtained from: Publications Services Branch (E/AI 13), National Environmental Satellite, Data, and Information Service, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, 3300 Whitehaven St., Washington, DC 20235.

NOAA Technical Report NMFS 9

Sampling Statistics in the Atlantic Menhaden Fishery

Alexander J. Chester

August 1984

U.S. DEPARTMENT OF COMMERCE
Malcolm Baldrige, Secretary
National Oceanic and Atmospheric Administration
John V. Byrne, Administrator
National Marine Fisheries Service
William G. Gordon, Assistant Administrator for Fisheries

CONTENTS

Figures

Tables

Sampling Statistics in the Atlantic Menhaden Fishery

ALEXANDER J. CHESTER[1]

ABSTRACT

Atlantic menhaden, *Brevoortia tyrannus,* the object of a major purse-seine fishery along the U.S. east coast, are landed at plants from northern Florida to central Maine. The National Marine Fisheries Service has sampled these landings since 1955 for length, weight, and age. Together with records of landings at each plant, the samples are used to estimate numbers of fish landed at each age. This report analyzes the sampling design in terms of probability sampling theory. The design is classified as two-stage cluster sampling, the first stage consisting of purse-seine sets randomly selected from the population of all sets landed, and the second stage consisting of fish randomly selected from each sampled set. Implicit assumptions of this design are discussed with special attention to current sampling procedures. Methods are developed for estimating mean fish weight, numbers of fish landed, and age composition of the catch, with approximate 95% confidence intervals. Based on specific results from three ports (Port Monmouth, N.J., Reedville, Va., and Beaufort, N.C.) for the 1979 fishing season, recommendations are made for improving sampling procedures to comply more exactly with assumptions of the sampling design. These recommendations include adopting more formal methods for randomizing set and fish selection, increasing the number of sets sampled, considering the bias introduced by unequal set sizes, and developing methods to optimize the use of funds and personnel.

INTRODUCTION

Since 1955 the National Marine Fisheries Service Laboratory at Beaufort, N.C., has sampled catches of Atlantic menhaden, *Brevoortia tyrannus,* for length, weight, and age at reduction plants located from Florida to Massachusetts. This information, along with records of landings, numbers of purse-seine sets, and estimates by vessel captains of catch per set, has been used by fishery scientists to investigate overall population structure and migratory patterns of menhaden (Nicholson 1971a, 1972; June 1972; Nelson et al. 1977), to evaluate the effects of fishing on menhaden stocks (Nicholson 1971b; Schaaf and Huntsman 1972; Schaaf 1979), and to derive management principles for the fishery (Schaaf 1975; Schaaf et al. 1975). In 1981 the Atlantic Menhaden Management Board recommended in its Fishery Management Plan for Atlantic Menhaden that at least 10% of all fish landed be age 3 or older. Although estimates of the number of fish of each age landed annually are basic to this management proposal and to the conclusions of the other cited studies, the overall reliability of the estimates and the factors affecting sampling efficiency have not been satisfactorily addressed.

The purposes of this report are to 1) define the basic menhaden sampling procedure in terms of probability sampling theory, 2) examine the assumptions of the statistical model and the consequences of their violation, 3) demonstrate methods for calculating approximate confidence intervals about some commonly estimated parameters, 4) explore implications of the sampling design for the optimal allocation of resources to this survey, and 5) provide suggestions for improving the sampling methodology so that it complies more exactly with underlying assumptions. I will also recommend areas requiring further research to document the validity of assumptions and to develop and present additional theory.

[1]Southeast Fisheries Center Beaufort Laboratory, National Marine Fisheries Service, NOAA, Beaufort, NC 28516-9722.

BRIEF HISTORY OF THE SAMPLING PROGRAM

Preliminary sampling of commercial menhaden landings began at the Lewes, Del., reduction factory in 1952 (June and Reintjes 1959). This pilot survey was intended to provide information on length, weight, and age distributions that could be used to develop an efficient design for sampling the entire fishery. At that time, processing plants were operating at 14 ports along the Atlantic coast (Fig. 1). Sampling was expanded during the 1953 and 1954 seasons, but was limited to ports from Chesapeake Bay northward. Sampling was expanded to include ports south of Chesapeake Bay in 1955, the first year of relatively complete coverage of the fishery. Although intensive data collection has continued since then, the number of ports decreased after the decline of the fishery in the 1960's. Today only six ports remain active. These are routinely sampled by on-site personnel hired on a seasonal basis.

June and Reintjes (1959) gave an explanation of the early sampling strategy. They verified that fish of similar size and age tend to school together and concluded from the pilot survey that a sample size of 20 fish was adequate to estimate the mean length of fish in a purse-seine set to within $\pm$ 2% ($P = 0.05$). Requirements for estimating age composition were not considered. Samples from purse-seine catches were collected directly from the holds of vessels by shoveling fish from the top of each load into buckets. Thus, each sample was assumed to represent fish from the last set made and did not necessarily reflect the composition of the entire boatload of fish. One hundred fish were sequentially withdrawn from buckets. A 20-fish sample was then constituted by selecting every fifth fish. These were measured (mm FL), weighed (g), sexed, and scales were removed for later age determination. Port samplers were instructed to gather 10 to 15 samples per week at each port sampled, but this number varied, depending on vessel activity and areas fished.

During the late 1960's, fishery biologists in the menhaden program concluded from theoretical and field studies that age and size variability within vessels was smaller than variability among vessels. They decided that more efficient estimates would be ob-

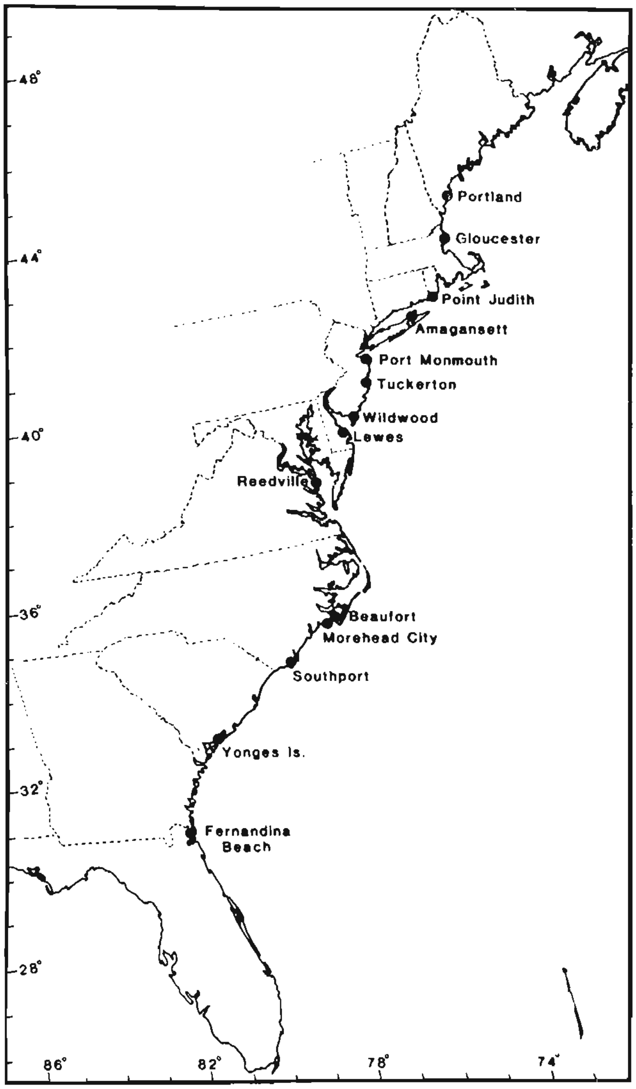

Figure 1.—Map showing the locations of menhaden reduction plants in operation along the U.S. east coast during 1955.

tained if the number of vessels sampled was increased and the number of fish per sample was decreased. Therefore, in 1971, sample size was decreased to 10 fish and the number of vessels sampled per week was increased to 20-25. In practice the total number of samples per week was often less than recommended, partially because of a decline in the number of landings (sampling opportunities).

STRUCTURE OF THE SAMPLING DESIGN

For any given port-week combination, the sampling design outlined above may be classified as two-stage cluster sampling (also called subsampling) (Cochran 1977), where purse-seine sets are the primary sampling units of unequal population size and individual fish are the secondary sampling units. Both primary and secondary sampling units are selected with equal probabilities and without replacement. The application of probability theory to this problem, while not completely straightforward, requires random selection at all stages of sampling and minimal selection and estimation biases. The present sampling procedure will be examined in light of these assumptions.

Because it is assumed that only the last purse-seine set of the trip is actually sampled, the boatload cannot be considered the primary sampling unit. I assume that a set is the fundamental unit and that sampling the last set of the trip yields a random sample from the population of all sets landed at a port during a given week. The selected sets are a true random sample if 1) individual schools are located and set upon at random, so that, in aggregate, last sets per trip are a random sample of all sets taken (there must not be a geographic stratification whereby, for instance, the last set of the day is always made close to home port); 2) sets are sampled at random from the total population of sets landed at a port.

Violation of the first assumption is especially critical if the age and size distribution of fish is regulated in any way, such as by proximity to shore. The tendency toward nonrandom sampling by set location was examined for pooled data from the Reedville, Va., port. Logbooks from each vessel were used to determine the 1979 total numbers of 1) sets taken in Chesapeake Bay for trips when all sets were made in the bay, 2) sets taken outside Chesapeake Bay for trips when all sets were made outside, 3) sets taken inside the bay for trips when fish were caught both inside and outside, and 4) sets taken outside the bay for trips when fish were caught both inside and outside. A log likelihood ratio test (*G* statistic) (Sokal and Rohlf 1981) was used to test the hypothesis of random sampling of sets for various combinations of inside and outside sets (Table 1). Under the null hypothesis, the ratio of inside to outside sets sampled would be expected to equal the ratio of inside to outside sets taken. When all inside and outside sets were included,

Table 1.—Tests of the hypothesis of random sampling of sets taken inside and outside Chesapeake Bay (Reedville, Va., port) during 1979.

A) All inside and outside sets taken:

		Number of sets sampled	
		Observed (f_i)	Expected ($\hat{f}_i$)
Inside sets	7,223	198	162.50
Outside sets	3,445	42	77.50

$G = 2\,\Sigma f_i \ln\left(\frac{f_i}{\hat{f}_i}\right) = 26.79$

$\chi_{.05,1} = 3.87$

∴ Reject the hypothesis of random sampling.

B) Inside and outside sets from "pure" trips on which either all sets were taken inside or all sets were taken outside the bay:

		Number of sets sampled	
		Observed (f_i)	Expected ($\hat{f}_i$)
Inside sets	5,320	159	151.11
Outside sets	1,334	30	37.89

$G = 2.18$

∴ Do not reject the hypothesis of random sampling.

C) Inside and outside sets from "mixed" trips on which both inside and outside sets were landed:

		Number of sets sampled	
		Observed (f_i)	Expected ($\hat{f}_i$)
Inside sets	1,903	39	24.18
Outside sets	2,111	12	26.82

$G = 17.99$

∴ Reject the hypothesis of random sampling.

the hypothesis of random sampling was rejected ($P < 0.05$); sets taken inside Chesapeake Bay were more likely to be sampled. However, when sets were divided into those from "pure" trips (trips during which all sets were made either inside or outside the bay) and those from "mixed" trips (trips during which sets were made both inside and outside), a possible source for the nonrandom behavior was discovered. When only pure trips were considered, the random-selection hypothesis could not be rejected; when only mixed trips were considered the sampling distribution was strikingly nonrandom. Apparently on trips catching fish both inside and outside the bay, vessels were more likely to make the last set of the cruise inside, and that is the set available for sampling. This possibility was tested by tabulating the actual number of last sets made inside and outside the bay and comparing these with the number of last inside and outside sets expected from the ratio of total inside to outside sets made on all mixed trips (Table 2). The G-test indicated ($P < 0.05$) that the last sets were not randomly distributed; i.e., vessels tended to make the last set inside Chesapeake Bay on their way to home port. This is the likely reason that inside sets are favored in the sampling process. The degree of bias introduced to our landing estimates depends on the degree of size and age difference between inside and outside fish. For the Chesapeake Bay example, I attempted to estimate the magnitude and direction of the bias by comparing sampled sets taken inside and outside the bay during mixed trips. The average weight of fish in sampled sets was significantly greater outside the bay (169.92 g) than inside (142.16 g) (t-test, $P < 0.05$, df = 49). I calculated that the observed oversampling of inside sets during 1979 may have introduced a 5.4% underestimate of average fish weight and an overestimate of numbers of fish landed for mixed trips. This bias would be somewhat ameliorated because only about 35% of all sets came from mixed trips.

For the second assumption (p. 2), random set selection has been defined operationally in terms of random vessel selection. It is important to point out, however, that random vessel selection does not, in and of itself, guarantee random set selection. In situations where certain vessels routinely land more sets than average and fish in these sets are significantly older or younger than in sets landed by other vessels, an unknown bias may be introduced. Such a situation may be imagined where the vessels of one company opt to land many small sets of older fish, and the vessels of another company opt to land a few large sets of younger fish. At any rate, port samplers have been instructed to sample vessels "randomly," i.e., not to follow patterns or show favoritism in the selection of vessels. This involves an element of judgment on the part of port samplers and as such may be a possible source of bias in that the actual method of random selection is not exactly specified (Williams 1978). As a first approximation, the assumption of random vessel selection was investigated for the port of Reedville, Va., for the 1979 fishing season (Table 3). Based on the number of landings made by each vessel, the expected number of sampling events per vessel was calculated under the assumption of random selection. The goodness-of-fit test (G statistic) applied to the data pooled from plants A and B led to a rejection of the hypothesis of random sampling ($P < 0.05$). The degree and direction of selection bias may be judged for each vessel (Table 3). In general, plant A vessels were underrepresented with respect to plant B vessels. To attain a random selection of vessels a more rigorous procedure should be specified. For example, a port sampler might obtain a list of all vessels expected to land that day and choose the one(s) to be sampled by tossing a six-sided die, picking names from a hat, or other appropriate randomization device. In the development of statistical methodology for this report, random vessel selection is assumed.

Table 2.—Test of the hypothesis that the last set of the day is randomly distributed between inside and outside sets when both kinds of sets are made on a single trip. Data are from Reedville, Va., for the 1979 fishing season.

		Number of final sets made	
		Observed	Expected
Inside sets	1,903	305	189.64
Outside sets	2,111	95	210.36

$G = 138.82$

$\chi_{.05,1} = 3.87$

∴ Reject the hypothesis that the final set of the day was randomly distributed between inside and outside sets.

For each set selected (at random), the port sampler presently obtains a subsample of 10 fish, which is itself considered to be randomly chosen from the population of fish constituting the set. The method of subsample selection has varied over the years. Initially, every fifth fish from a bucketful was selected to make up the subsample. However, less formalized methods have been employed since the decision to reduce subsample size from 20 to 10 individual fish. Presently, port samplers try to collect a bucketful of fish from the top of the hold before offloading begins and pick out the 10-fish sample "randomly" from the bucket. For practical reasons, such as time constraints or vessel configuration, this is not always possible, and a sample may be obtained from the

Table 3.—Test of the hypothesis of random vessel selection for the port of Reedville, Va., during the 1979 fishing season. Pluses and minuses indicate vessels where the observed number of sampling events differs from the expected by more than 3.

Plant	Vessel	No. of landings	No. of times sampled	No. of sampling times expected
A	1	64	6	8.60
	2	65	8	8.74
	3	102	16	13.71
	4	98	6	13.18 −
	5	96	2	12.91 −
	6	94	8	12.64 −
	7	94	11	12.64
	8	95	11	12.77
	9	95	16	12.77 +
	10	41	5	5.51
	11	23	2	3.09
	12	7	2	0.94
	13	4	0	0.54
B	14	71	16	9.55 +
	15	76	14	10.22 +
	16	63	3	8.47 −
	17	60	11	8.07
	18	61	12	8.20 +
	19	54	2	7.26 −
	20	62	13	8.34 +
	21	71	16	9.55 +
	22	62	9	8.34
	23	71	6	9.55 −
	24	74	14	9.95 +
	25	78	17	10.49 +

$G = 2 \Sigma f_i \ln\left(\frac{f_i}{\hat{f}_i}\right) = 55.89$

$\chi_{.05,24} = 36.42$

∴ Reject the hypothesis of random vessel selection.

"drag" or at other points along the offloading procedure. In some cases a deep hold might require a dip-net collection, and it is even possible that an occasional sample may result from simply picking 10 fish from the hold. The general instructions have been to choose fish as randomly as possible within a set, consistent with safety and practicality, so as to minimize possible bias (Gulland 1966). Here again judgment of the port samplers is involved, and a more rigorous plan might be specified. A good critique of the substantial problems inherent in the random selection of fish from a vessel is given by Tomlinson (1971), who concluded that the task is operationally impossible. In any case, we assume that any added bias is small.

BASIC CALCULATIONS IN THEORY AND PRACTICE

In this section, the basic catch statistics and their derivatives are explored with respect to problems of estimation and the development of variance formulae. Three main topics are considered. I begin by discussing how mean weight per fish is currently calculated for any port-week combination, suggest some potential difficulties with the method, and estimate the sample variance and resulting confidence intervals. I then follow the current computational method through to arrive at the total number of fish landed at a plant during a given week and for the entire fishing season, and derive confidence intervals for the estimates. Finally, I examine the problem of estimating age composition in the catch and attempt to evaluate the precision with which numbers of specific age fish landed are now estimated.

In the following discussion, the 1979 menhaden catch records are examined for three ports. These ports were selected because they reflect broad geographic differences in age composition, and because they demonstrate how difficulties encountered in estimating catch characteristics differ among locations. Port Monmouth, N.J., is representative of northern waters. Landings are made between late May and early October and are composed of older (age 2 and older) fish. At Reedville, Va., where the fishing season extends from mid-May to late October, younger (age 1 and 2) fish are landed. At Beaufort, N.C., there are two fishing seasons. During the first season (early May to late October) the catch is composed mainly of age 1 and age 2 fish. During the second season (November to mid-January) the North Carolina "fall fishery" takes advantage of impressive aggregations of southerly migrating fish to land fish of all ages. These include older, spawning fish from northern waters and small, young-of-the-year fish that move into the ocean from bays and sounds.

In the development of sample computations and variance estimates, I will make use of the following notation patterned closely after Cochran (1977):

N = number of sets made per week per port

n = number of sets sampled per week per port

M_i = number of fish in the ith sampled set

m_i = number of fish sampled from the ith set

y_{ij} = value obtained from the jth fish sampled from ith set

$f_1 = \frac{n}{N}$ = sampling fraction in the first stage (sets)

$f_2 = \frac{m}{M}$ = sampling fraction in the second stage (fish).

Further notation will be supplied as needed.

Estimating the Mean Weight Per Fish Landed

Present method and biases.—In practice the mean weight of an individual fish landed at a given port during a given week has been estimated by the following method:

If y_{ij} = the weight of the jth fish from the ith sampled set, then

$$\bar{y}_i = \sum_{j=1}^{m} \frac{y_{ij}}{m} = \text{average weight of a fish in the } i\text{th set,} \quad (1)$$

$$\text{and, } \bar{\bar{y}} = \sum_{i=1}^{n} \frac{\bar{y}_i}{n} = \text{average weight of a fish in } n \text{ sampled sets.} \quad (2)$$

That is, the overall sample mean is simply the average of the mean weights per sampled set. The estimate is inherently biased because the sets are composed of unequal numbers of fish (Pope 1956). This bias arises because the inequality of set size (M_i) affects the probabilities of including a given fish in a subsample. This bias may not be appreciable, however, if the set sizes do not vary considerably, if there is no correlation between set size and true mean weight per fish, or if the number of sets sampled (n) is large (Sukhatme and Sukhatme 1970).

To illustrate the possible effect of unequal set size on the estimate of mean weight, consider a hypothetical population of 81 sets (Fig. 2), where the number of fish in a set is an inverse linear function of the average weight per fish. The example assumes no subsample variability, i.e., all fish in a set are the same weight, and any subsample of 10 fish will estimate the true mean per set. The true population mean in the example is 235.46 g/fish. From this population we now randomly sample 10 sets with equal probability and calculate two means for each sample. The first mean is the simple average of the 10 sets sampled, analogous to the $\bar{\bar{y}}$ computed above. The second mean is a ratio-to-size estimate (Cochran 1977: 303), which is essentially an average weighted by set size:

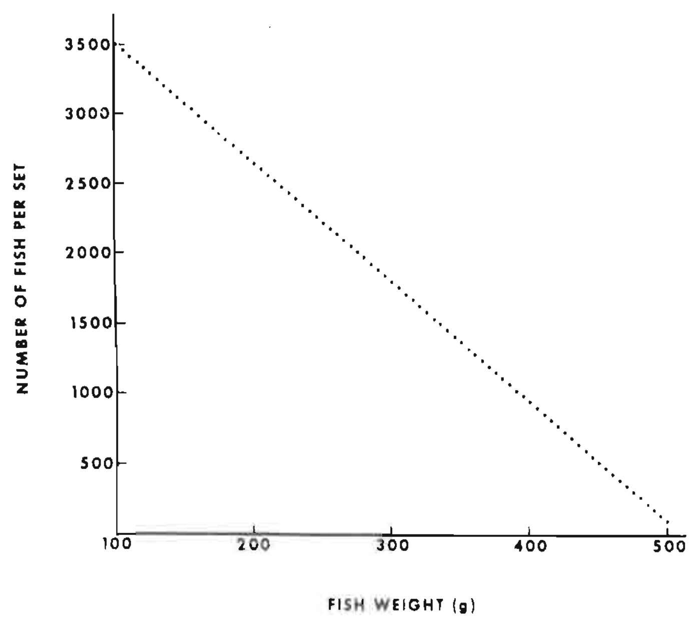

Figure 2.—Hypothetical population of 81 sets in which the number of fish per set is inversely related to the average weight per fish.

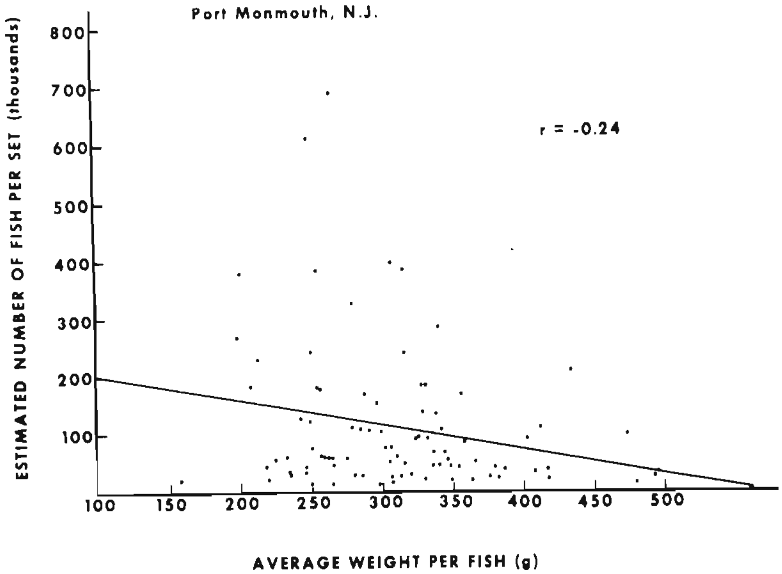

Figure 3.—Relationship between average weight per fish (estimated from dockside samples) and estimated set size (from captains' logbook records) for Port Monmouth, N.J., during 1979.

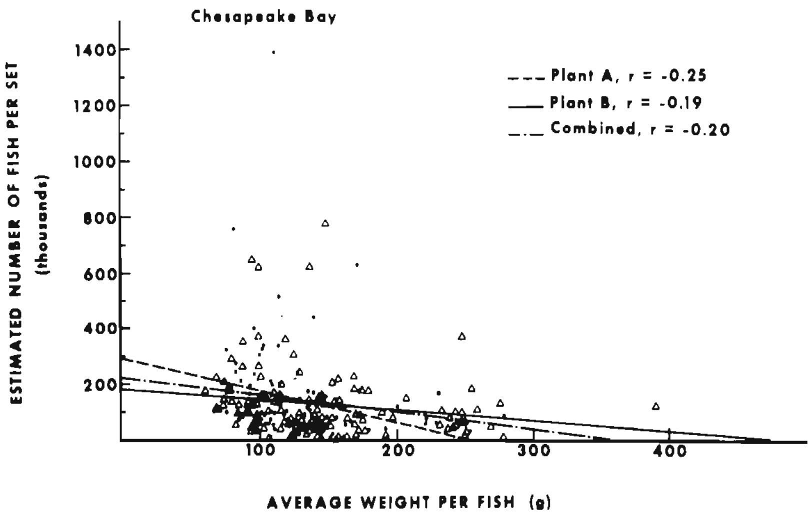

Figure 4.—Relationship between average weight per fish (estimated from dockside samples) and estimated set size (from captains' logbook records) for Reedville, Va., during 1979.

$$\bar{\bar{y}}_R = \sum_{i=1}^{n} M_i \bar{y}_i \Big/ \sum_{i=1}^{n} M_i \quad . \tag{3}$$

This sampling procedure was repeated 25 times and the average $\bar{\bar{y}}$ and $\bar{\bar{y}}_R$ computed. In the first instance, $\bar{\bar{y}}$ = 295.18 g, while the weighted statistic ($\bar{\bar{y}}_R$) equalled 227.39 g, much closer to the true population value.

This example demonstrates the potential dangers inherent in the present method of calculating average weight. Later we will see that such a bias could, if it exists, result in profound errors in the estimation of total numbers of fish caught. It is, therefore, critical to determine the extent to which set size is correlated with average weight per fish. June (1972) observed a significant negative correlation ($r = -0.60$) between mean fork length and numbers of fish per set for 275 summer purse-seine sets made during 1955-62. Whether his data are valid for our purposes is open to question, since the sets are representative of a wide geographic area and long time interval. For sets taken within any single port-week combination the correlation may not be so striking. I have examined similar relationships in the three ports over the 1979 fishing season (Figs. 3-5). The number of fish per set, computed by dividing the captains' estimated catch per set by the average weight per fish from dockside samples, was itself plotted

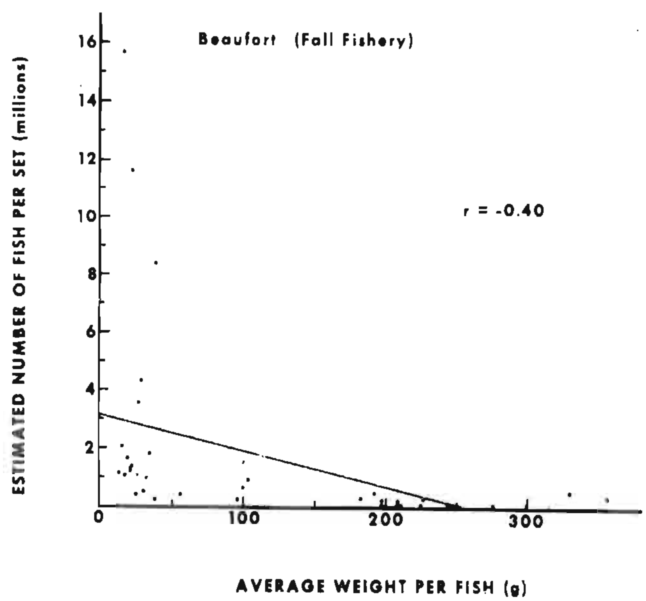

Figure 5.—Relationship between average weight per fish (estimated from dockside samples) and estimated set size (from captains' logbook records) for Beaufort, N.C., during 1979.

against average weight. Relationships of this form (i.e., a ratio vs. its denominator) commonly yield spuriously high self-correlations, but this effect is reduced here, since the variance in average fish weight is small compared with the variance in estimated catch (Kenney 1982). A correlation between set size and average weight is not strongly evidenced for either Port Monmouth or Reedville ($r = -0.24$ and -0.20, respectively; neither significantly differs from zero at $P < 0.01$). For the Beaufort fall fishery, however, the correlation coefficient ($r = -0.40$) is significantly negative and the relationship appears nonlinear. It may be tentatively concluded that, for most of the fishery, the present method of calculating average weight is not materially biased by the implicit assumption of equal set size. For the North Carolina fall fishery, however, this bias may be larger, and, depending on the age distribution of the catch, the estimate of average weight per fish may be overstated. Hence, total numbers landed would be underestimated.

A further verification that the present method of calculating mean weight is reasonable is a comparison between the simple average, $\bar{\bar{y}}$, and the approximate weighted mean, $\bar{\bar{y}}_R$ (Table 4). The computation of $\bar{\bar{y}}_R$ is limited to Port Monmouth and Reedville because it requires the captains' daily fishing report estimates of set size for every sample taken. This knowledge is incomplete for the port of Beaufort. In almost every instance the two means are in close agreement; $\bar{\bar{y}}_R$ is always included within the 95% confidence intervals about $\bar{\bar{y}}$.

Approximate confidence intervals.—Turning now to the problem of calculating confidence intervals about the mean weight estimate, we will continue to assume equal set sizes in the development of variance functions. If the n primary units (sampled sets) and the m subunits (fish within sets) from each chosen unit are selected by simple random sampling, Cochran (1977) gives the sample variance as:

$$v(\bar{\bar{y}}) = \frac{1-f_1}{n} s_1^2 + \frac{f_1(1-f_2)}{mn} s_2^2, \tag{4}$$

where $s_1^2 = \dfrac{\sum_{i=1}^{n} (\bar{y}_i - \bar{\bar{y}})^2}{n-1}$,

and $s_2^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} \dfrac{(y_{ij} - \bar{y}_i)^2}{n(m-1)}$.

There are, therefore, two parts to the variance expression. In analysis of variance terms, s_1^2 is equivalent to Among Set Mean

Table 4.—Number of sets sampled, mean weight per fish ($\bar{\bar{y}}$), standard error (SE) of $\bar{\bar{y}}$, and the ratio-to-size estimate of mean weight ($\bar{\bar{y}}_R$) for Port Monmouth, N.J., Reedville, Va., and Beaufort, N.C., for weeks during the 1979 fishing season.

	Port Monmouth, N.J.				Reedville, Va.				Beaufort, N.C.			
Week ending	n	$\bar{\bar{y}}$	SE	$\bar{\bar{y}}_R$	n	$\bar{\bar{y}}$	SE	$\bar{\bar{y}}_R$	n	$\bar{\bar{y}}$	SE	$\bar{\bar{y}}_R$
5/12									3	123.7	17.8	
5/19									2	96.0	0.8	
5/26					8	142.6	5.2	141.6	3	98.7	5.0	
6/ 2					10	125.2	4.8	117.9	5	94.1	4.3	
6/ 9					10	137.7	5.0	144.1	4	118.4	7.4	
6/16	9	252.2	10.8	254.8								
6/23	5	317.6	28.8	308.8	10	130.7	5.8	134.9	4	112.4	3.3	
6/30	4	305.9	17.3	306.6	10	126.6	8.3	107.8	5	116.2	2.8	
7/ 7	6	266.4	28.8	237.0	10	154.2	21.5	143.4	2	95.6	25.6	
7/14					10	146.1	17.6	130.8	2	100.2	3.0	
7/21	10	274.4	16.0	265.3	10	148.5	7.8	152.0	4	106.6	3.3	
7/28	8	286.6	26.3	259.9	10	168.5	14.6	171.1	3	108.0	3.8	
8/ 4	7	322.0	35.0	379.4	8	142.6	9.1	143.5	2	106.2	22.2	
8/11	6	318.5	22.8	—	8	162.1	24.8	129.9	2	111.0	5.8	
8/18	6	330.9	6.8	—	10	154.1	15.8	143.8				
8/25					10	155.5	11.8	154.0				
9/ 1	4	259.1	6.6	261.6	8	187.6	23.4	187.9	2	78.9	69.8	
9/ 8	4	309.3	17.1	314.9	6	112.3	3.0	109.3				
9/15	4	382.9	33.4	411.5	8	105.0	6.6	97.8				
9/22	6	399.1	41.0	342.6	10	159.3	21.0	160.1				
9/29	4	319.6	9.9	317.8	6	96.6	5.7	98.3				
10/ 6	4	359.5	22.5	351.2	10	153.7	17.4	135.6	3	96.0	9.7	
10/13					8	107.3	11.5	104.3				
10/20					10	96.7	11.1	83.9	4	139.9	10.9	
10/27	4	343.5	7.0	—	8	123.9	28.0	106.0	5	104.3	41.8	
11/ 3					10	133.1	16.4	112.4	6	54.5	20.0	
11/10					8	97.0	22.5	79.3	7	92.0	32.9	
11/17					2	137.0	14.3	—				
11/24					2	382.1	6.7	376.7	11	206.8	9.9	
12/ 1									10	193.4	13.9	
12/ 8									20	110.3	17.4	
12/15									16	114.4	30.8	
12/22									14	27.8	2.8	
12/29												
1/ 5									9	18.6	1.2	
1/12									4	22.4	2.0	

Square/m, a reflection of the variance among the sampled set means. In contrast, s_2^2 is identical to the error mean square (MS), and as such it estimates the amount of variance within a set. The f_1 and f_2 terms are sampling fractions for the sets and fish, respectively, and are included as finite population corrections which express the degree to which the whole population is sampled. As f_1 and f_2 increase, the overall variance decreases because we gain progressively more information about the entire population. In the menhaden sampling survey, f_2 is always infinitely small (and generally unknown) because each set contains large numbers of fish, and f_1 is generally < 0.10. For these reasons the variance formula can be simplified to:

$$v(\bar{\bar{y}}) \cong \frac{s_1^2}{n} = \frac{\sum_{i=1}^{n} (\bar{y}_i - \bar{\bar{y}})^2}{n(n-1)}. \qquad (5)$$

The estimated variance is, therefore, dependent only on the unit (set) means. If f_1 is not small, this simplification gives a conservatively high estimate of the variance. The standard error is, of course, $\sqrt{v(\bar{\bar{y}})}$, and confidence regions may be estimated as $\bar{\bar{y}} \pm 2(\text{SE})$, if the sample size is large.

To illustrate the calculation of $\bar{\bar{y}}$ and its variance, consider the sample data for Port Monmouth for the week ending 6/16/79 (Table 5). For this case, $\bar{\bar{y}}$ was estimated at 252.19 g/fish landed. Assuming no finite population correction, the variance can be estimated from the sample means by the equation given above, or equivalently from the analysis of variance:

$$v(\bar{\bar{y}}) = \frac{\text{Among Set Mean Square}}{n\ m}. \qquad (6)$$

In either case, $v(\bar{\bar{y}}) = 116.49$, SE = 10.79, and the confidence intervals are estimated as $252.19 \pm 2(10.79)$ g. (Standard errors are reported for the three selected ports for weeks during 1979 in Table 4.)

Optimal allocation of sampling resources.—The data in Table 2 may be used to examine the efficient allocation of sampling effort within a two-stage sampling design. To demonstrate the effect of varying the number of sets sampled (n) and the number of fish sampled per set (m), we first consider the theoretical variance of the population mean:

$$v(\bar{\bar{y}}) = \frac{\sigma_a^2}{n} + \frac{\sigma_w^2}{nm}. \qquad (7)$$

Here, the true population statistics, σ_a^2 and σ_w^2, are best estimated by the among set component of variance (s_a^2) and the within set component of variance (s_w^2) calculated from the analysis of vari-

Table 5.—Survey data collected during the week ending 6/16/79 at Port Monmouth, N.J. Values reported are weight (g) for each fish in the sample. Values in parentheses are fish ages estimated from scales. The analysis of variance table, estimate of mean weight per fish ($\bar{\bar{y}}$), variance and standard error of $\bar{\bar{y}}$, and approximate 95% confidence intervals are also given.

		Sampling dates								
Fish no.		6/12	6/12	6/12	6/13	6/13	6/14	6/14	6/15	6/15
1		329(2)	274(2)	247(2)	278(2)	160(2)	237(2)	184(2)	227(2)	215(2)
2		254(3)	244(3)	216(2)	305(3)	260(3)	257(2)	223(3)	271(2)	258(3)
3		274(2)	407(3)	178(2)	176(2)	208(2)	204(2)	401(3)	240(2)	209(2)
4		314(3)	401(3)	318(3)	185(2)	175(2)	246(3)	250(2)	252(3)	268(2)
5		312(3)	338(3)	263(3)	161(2)	269(3)	219(2)	157(2)	371(3)	305(3)
6		288(2)	332(3)	276(3)	355(2)	272(3)	207(2)	147(2)	230(3)	231(2)
7		339(3)	385(3)	261(2)	309(3)	181(2)	174(2)	197(2)	219(2)	154(2)
8		259(2)	194(2)	200(3)	218(2)	278(2)	263(2)	205(2)	213(3)	213(2)
9		176(2)	293(3)	271(3)	266(2)	237(2)	263(3)	294(2)	218(2)	241(2)
10		296(2)	350(3)	327(3)	221(2)	203(2)	278(3)	147(2)	246(2)	230(2)
	$\bar{y}_i$	284.1	321.8	255.7	247.4	224.3	234.8	220.5	248.7	232.4
	s	47.36	70.07	47.71	64.97	44.41	32.91	78.75	46.51	40.34

ANOVA table

Source	df	SS	MS
Total	89	324,155.7889	
Among sets	8	83,874.0889	10,484.2611
Within sets	81	240,281.7000	2,966.4407

$$\bar{\bar{y}} = \sum_{i=1}^{n} \frac{y_i}{n} = 252.19$$

$$v(\bar{\bar{y}}) \cong \frac{\sum_{i=1}^{n} (\bar{y}_i - \bar{\bar{y}})^2}{n(n-1)} \cong \frac{\text{Among Set MS}}{nm} \cong 116.49$$

$$\text{SE}_{\bar{\bar{y}}} = 10.79$$

95% Confidence Intervals about $\bar{\bar{y}} \cong 252.19 \pm 2(10.79) \cong 230.60 \rightarrow 273.78$

ance of the existing sample (Snedecor and Cochran 1967):

$$s_a^2 = s_1^2 - \frac{s_2^2}{m} = \frac{\text{Among Set MS} - \text{Error MS}}{m} = \frac{10{,}484 - 2{,}966}{10} = 751.8 \tag{8}$$

$$s_w^2 = s_2^2 = \text{Error MS} = 2{,}966.44. \tag{9}$$

If we were to conduct a new survey of the catch for that particular port-week, the predicted variance of the mean weight would be:

$$v(\bar{y}) = \frac{s_a^2}{n} + \frac{s_w^2}{nm} = \frac{751.78}{9} + \frac{2{,}966.44}{(9)\,(10)} = 116.49. \tag{10}$$

These data can be used to help address two related matters. First, we can now predict the variance we would obtain with different values of n and m. Second, given s_a^2 and s_w^2 we can determine what combinations of n and m will yield a required variance. For example, by sampling 5 fish per set from 30 sets we would expect the variance to be lowered to 44.84. These conclusions are valid only for the given port-week. More general conclusions rely on similar analyses for several port-week combinations. Clearly, however, sampling requirements in the menhaden fishery vary according to location and time of year.

For a given level of variance, several different values of n and m are possible. The "best" combination depends on the relative cost of sampling primary and secondary units. For the menhaden survey it has been estimated (Huntsman[2]) that a port sampler needs 45 min to collect a subsample and 7.5 min to age, weigh, and measure the length of each fish. The relative cost of subsample selection versus individual fish analysis is approximately 6:1. If we ignore fixed costs, the expense of the sample can be estimated by:

$$\text{cost} = c_1 n + c_2 nm, \tag{11}$$

where c_1 is the cost of sampling a set and c_2 is the cost of processing a fish. To estimate the optimal number of fish per set, the product of the cost and variance functions are minimized (Snedecor and Cochran 1967):

$$m_{\text{opt}} = \sqrt{\frac{c_1 s_w^2}{c_2 s_a^2}} = \sqrt{\frac{(6)\,(2{,}966.44)}{(1)\,(751.78)}} = 5 \text{ fish}. \tag{12}$$

Assuming that c_1/c_2 remains constant, the optimal subsample size (m) is directly related to $\sqrt{s_w^2/s_a^2}$. As the variance among sets increases relative to the variance within sets, this fraction becomes smaller, and optimal allocation requires greater numbers of sets to be sampled with fewer fish per set. Quick calculations for Port Monmouth, Reedville, and Beaufort reveal optimal subsampling sizes of about 4 fish with an upper 95% confidence bound of about 10 fish.

Knowing the optimal value for m, one can use the cost and variance equations [Equations (10) and (11)] to either 1) minimize the sampling cost subject to a variance constraint, or 2) minimize the variance subject to a cost constraint. In the Table 5 example, with n = 9, m = 10, and variance = 116.49, the cost of sampling is (45) (9) + (7.5) (9) (10) = 1,080 min. To predict the minimum cost required to halve the variance, first replace m with m_{opt} to solve for n in the variance equation and then solve the cost equation:

$$\frac{751.78}{n} + \frac{2{,}966.44}{5n} = 58.25,$$

n = 23 sets, and

45 (23) + 7.5 (23) (5) = 1,898 min.

Conversely, given a time resource limit of X minutes, it is a simple matter to choose n and m such that variance is minimized. For example, given $m_{\text{opt}} = 5$, what is the minimum possible variance for our original cost of 1,080 min? Substitute m_{opt} into Equation (11) and solve for n (answer: n = 13 sets; minimum variance = 103.47).

The relationship between cost and variance may be represented graphically (Fig. 6). As n and m increase, the variance decreases and the cost of sampling increases. To solve for minimum cost at a stated variance, the cost curve is moved, and the point of tangency with the variance isoline defines optimal n and m. To solve for minimum variance, the cost curve is fixed and the variance curve is moved to a point of tangency.

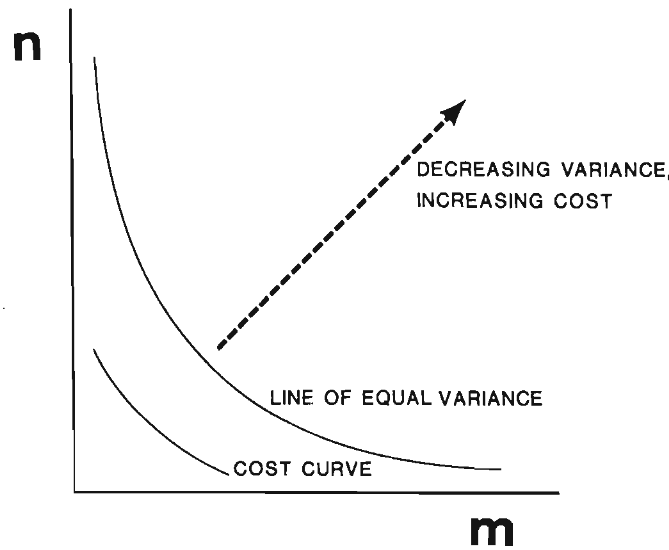

Figure 6.—Theoretical relationship between n (number of sets sampled) and m (number of fish sampled per set) and the variance in average weight and cost of the sampling program.

Estimating the Total Number of Fish Landed

Number of fish landed by port-week.—The total number of fish landed per port-week (C_{pw}) is presently estimated by dividing the total weight of the catch for each plant (W_t) by the mean weight per fish landed ($\bar{\bar{y}}$):

$$C_{pw} = W_t \,/\, \bar{\bar{y}}. \tag{13}$$

I assume that W_t is measured without error and that $\bar{\bar{y}}$ is a random variable subject to the uncertainty described previously. For the sample data shown in Table 5, the weekly catch total was 2,356.87 t. Dividing this figure by 252.19 g/fish, an estimated

[2]G. R. Huntsman, Fishery Biologist, Southeast Fisheries Center Beaufort Laboratory, National Marine Fisheries Service, NOAA, Beaufort, NC, 28516-9722, pers. commun. April 1981.

9.346×10^6 fish were harvested at Port Monmouth during the week ending 6/16/79.

The variance of the estimate of total fish caught is derived by considering the basic calculation to be of the form $a\bar{x}$, where a is the constant W_t, and $\bar{x}$ is the inverse of the mean weight per fish, $1/\bar{\bar{y}}$. The variance of this product may be expressed as:

$v(a\bar{x}) = a^2\, v(\bar{x})$, and by analogy,

$$v(W_t / \bar{\bar{y}}) = W_t^2\, v(\bar{\bar{y}}^{-1}). \tag{14}$$

The variance of $\bar{\bar{y}}^{-1}$ is estimated as for $\bar{\bar{y}}$ in the previous section, but an inverse transformation is applied to the individual fish weights initially. The ANOVA table derived from the data in Table 5, and the calculation of variance and confidence intervals are given in Table 6. For that example, the 95% confidence intervals extended to ± 8.6% of the total catch estimate.

Table 6.—Estimate of the total number of fish landed at Port Monmouth, N.J., during the week ending 6/16/19 with approximate 95% confidence intervals.

ANOVA table			
Source	df	SS	MS
Total	89	9.178×10^{-5}	
Among sets	8	2.085×10^{-5}	2.6061×10^{-6}
Within sets	81	7.093×10^{-5}	8.7569×10^{-7}

$$v(\bar{\bar{y}}^{-1}) = \frac{\text{Among Set MS}}{nm} = 2.900\times10^{-8}$$

$$W_t = 2.35687\times10^9 \text{ g}$$

$$C_{pw} = \frac{W_t}{\bar{\bar{y}}} = \frac{2.35687\times10^9}{252.19} = 9{,}345{,}612 \text{ fish landed}$$

$$v(W_t / \bar{\bar{y}}) = W_t^2\, v(\bar{\bar{y}}^{-1}) = (2.35687\times10^9)^2\, 2.900\times10^{-8} = 1.611\times10^{11}$$

SE = 401,361 fish

95% Confidence Interval $\cong 9.346\times10^6 \pm 2(401{,}361)$

$\cong 8{,}708{,}694 \rightarrow 9{,}982{,}550$

$\cong$ estimate ± 8.6%

Number of fish landed over ports and weeks.—Similar calculations can be made for any port-week combination and we can build on these individual catch estimates to obtain composite measures of the total catch over any period of time or combinations of ports. For example, to get an estimate for the total number of fish caught in the Beaufort fall fishery, simply add up the weekly totals. (I assume that samples are available for every week included in the estimate.) If k weeks are to be summed:

$$C_T = C_{pw_1} + C_{pw_2} + \ldots + C_{pwk}. \tag{15}$$

Assuming that weekly totals are independently estimated, the variance of the grand total (C_T) may be expressed as:

$$v(C_T) = v(C_{pw_1}) + v(C_{pw_2}) + \ldots + v(C_{pwk}). \tag{16}$$

This equation ignores any possible covariance terms among weekly catch totals. These terms may exist, for example between C_{pw_1} and C_{pw_2}, if seasonal trends recur. Such an effect would inflate the variance and, hence, decrease the confidence about seasonal totals.

Tabulated in Table 7 are the weekly and season totals for Port Monmouth and Reedville; only the fall fishery totals are given for Beaufort. Confidence intervals, expressed as a percentage of the individual estimates, lead to the conclusion that the northern summer fishery is less uncertain with respect to catch estimates than is the North Carolina fall fishery. As indicated earlier, it is the greater variance associated with estimates of average fish weight that supplies the explanation.

Table 7.—Estimate of the total number of fish landed by port-week and for the entire 1979 season. Approximate 95% confidence intervals are included and are expressed as a percentage of the estimate.

Week	Port Monmouth, N.J.	Reedville, Va.	Beaufort, N.C.
5/12			
5/19			
5/26		3.678×10^7 ± 7.5%	
6/ 2		5.528×10^7 ± 8.1%	
6/ 9		5.534×10^7 ± 7.2%	
6/16	9.346×10^6 ± 8.6%	6.898×10^7 ±16.0%	
6/23	6.837×10^6 ±20.4%	5.963×10^7 ± 8.6%	
6/30	6.833×10^6 ±13.1%	5.933×10^7 ±16.8%	
7/ 7	1.043×10^7 ±23.7%	6.271×10^7 ±30.0%	
7/14		5.984×10^7 ±25.2%	
7/21	1.368×10^7 ± 9.6%	4.226×10^7 ±11.8%	
7/28	6.236×10^6 ±17.1%	7.168×10^7 ±16.0%	
8/ 4	9.652×10^6 ±33.3%	4.740×10^7 ±12.9%	
8/11	8.316×10^6 ±15.8%	3.917×10^7 ±27.6%	
8/18	8.011×10^6 ± 4.9%	6.718×10^7 ±22.7%	
8/25		3.017×10^7 ±15.2%	
9/ 1	2.347×10^6 ± 4.4%	4.474×10^7 ±30.9%	
9/ 8	3.951×10^6 ±21.9%	4.871×10^7 ± 5.7%	
9/15	4.144×10^6 ±16.6%	5.120×10^7 ±10.7%	
9/22	5.843×10^6 ±22.0%	5.799×10^7 ±31.2%	
9/29	2.044×10^6 ± 6.5%	4.801×10^7 ±16.3%	
10/ 6	2.165×10^6 ±13.5%	4.812×10^7 ±23.0%	
10/13		5.906×10^7 ±24.4%	
10/20		1.478×10^8 ±15.3%	
10/27	2.235×10^6 ± 3.3%	5.140×10^7 ±36.4%	
11/ 3		4.514×10^7 ±29.4%	4.743×10^6 ±164.2%
11/10		3.245×10^7 ±45.4%	3.858×10^6 ±395.0%
11/17			
11/24		8.157×10^6 ± 5.7%	1.428×10^7 ± 15.8%
12/ 1			6.825×10^6 ± 33.9%
12/ 8			6.544×10^7 ± 73.0%
12/15			9.171×10^7 ±119.8%
12/22			4.088×10^8 ± 18.3%
12/29			
1/ 5			5.955×10^8 ± 11.3%
1/12			1.750×10^8 ± 18.4%
Season total	1.021×10^8 ± 5.1%	1.398×10^9 ± 4.3%	1.367×10^9 ± 11.8%

Estimating the Age Composition of the Catch

Basic calculations and variance formulations.—The age composition of the catch for any particular port-week is presently estimated by averaging the relative proportions of each age class over the n sampled sets. The age of each sampled fish, determined from scale readings, is assumed to be known without error. Let p_{ij} be the proportion of fish falling in the jth age category for the subsample from the ith sampled set. The proportion of all other age categories within the subsample may then be expressed as $q_{ij} = 1-p_{ij}$. The mean proportion of the jth age category in the population is then estimated to be:

$$\bar{p}_j = \sum_{i=1}^{n} p_{ij}/n. \tag{17}$$

This estimate is subject to the same sort of biases as encountered for the mean weight estimate. That is, we must assume that age composition is relatively independent of overall set size (M_i).

Assuming the p_{ij}'s are normally distributed within the population of N sets, an estimate of the sample variance may be derived. The two parts to the complete expression, the among set (s_1^2) and the within set (s_2^2) variances, are given by Cochran (1977) as:

$$s_1^2 = \frac{\sum_{i=1}^{n} (p_{ij} - \bar{p}_j)^2}{n-1} \tag{18}$$

$$s_2^2 = \frac{m}{n(m-1)} \sum_{i=1}^{n} p_{ij} q_{ij}. \tag{19}$$

The derivation of these formulae and their relationship to the analysis of variance are not as readily apparent as in the case of mean weight per fish. These variance components are derived in the Appendix.

Given s_1^2 and s_2^2, the variance of $\bar{p}_j$ may now be estimated in a fashion similar to that of the variance of mean weight:

$$v(\bar{p}_j) = \frac{1-f_1}{n(n-1)} \sum_{i=1}^{n} (p_{ij}-\bar{p}_j)^2 + \frac{f_1(1-f_2)}{n^2(m-1)} \sum_{i=1}^{n} p_{ij} q_{ij} \tag{20}$$

$$= \frac{1-f_1}{n} s_1^2 + \frac{f_1(1-f_2)}{mn} s_2^2 . \tag{21}$$

Once more, assuming that f_1 and f_2 are small, the estimate simplifies to:

$$v(\bar{p}_j) = \frac{s_1^2}{n} = \frac{\sum_{i=1}^{n} (p_{ij}-\bar{p}_j)^2}{n(n-1)}. \tag{22}$$

As an example, consider the Port Monmouth data from Table 5. The proportions of age 2 fish for those nine sampled sets are 0.6, 0.2, 0.4, 0.8, 0.7, 0.7, 0.8, 0.6, 0.8, respectively. The following calculations may be performed:

$$\bar{p}_2 = \frac{\sum_{i=1}^{n} p_{i2}}{n} = \frac{5.60}{9} = 0.622222$$

$$v(\bar{p}_2) = \frac{0.33556}{(9)\ (8)} = 0.00466$$

$$\text{SE} = \sqrt{v(\bar{p}_2)} = 0.06827$$

$$\text{95\% Confidence interval} \cong 0.6222 \pm 2(0.06827)$$
$$\cong 0.4857 - 0.7588.$$

Estimates of mean age compositions and variances are presented for each of the three ports by week for the 1979 fishing season (Tables 8-10).

It is an extremely practical matter for the management of the fishery to be able to state an average age composition with approximate confidence intervals over time intervals and geographic areas greater than 1 wk and one port. For example, the 1981 Menhaden Fishery Management Plan set a goal that at least 10% of all fish landed be age 3 or older. Specific administrative actions are keyed to this criterion. Obviously then, a knowledge of the precision with which overall age proportions are estimated is critical to the credibility of this part of the management plan. We approached the problem by conceptualizing a stratified two-stage sampling design whereby strata are individual weeks whose relative population sizes are indexed by the estimates of total catch. The underlying assumption remains that all sets within a given port-week have the same number of fish, but now set size may vary from week to week. For a given port over the entire season:

$$\bar{P}_{jkST} = \frac{\Sigma N_{wk} \bar{P}_{jwk}}{N_{tk}} = \Sigma W_{wk} \bar{P}_{jwk}, \tag{23}$$

where $\bar{P}_{jkST}$ is the overall average proportion of the jth age category over all strata from the kth port, N_{wk} is the estimated total catch for the wth week at the kth port, N_{tk} is the total catch over all weeks at the kth port, and $\bar{P}_{jwk}$ is the average proportion of age j fish in the wth week and the kth port. W_{wk} is, therefore, a week specific weighting factor. The variance about this overall proportion is:

$$v(P_{jkST}) = \sum_{w=1}^{52} W_{wk}^2\, v(P_{jwk}). \tag{24}$$

Similarly, average age composition over the entire fishery may be estimated by considering each port as a stratum within the total fishery. Overall estimates of the average age composition (for ages 0, 1, 2, 3+) and approximate 95% confidence intervals by port and over all three ports for the 1979 fishing season are given in Tables 8-11.

Optimal allocation of resources in age composition sampling.—The solution of the optimal allocation problem for proportional age sampling is by no means simple. Where only two age categories are present, the binomial distribution provides an appropriate theoretical framework within which to proceed. For more than two age classes, however, the multinomial distribution is preferable, but the theoretical properties with respect to two-stage sampling have not been adequately addressed. Cochran (1977) advised that sample size be determined by considering each age category separately versus the remaining categories as a binomial distribution. In the case of simple random sampling, one would presumably use the largest estimated sample size for the survey. For a two-stage design the optimal number of primary and secondary units is more difficult to determine. It is clear, however, that what is really wanted is the simultaneous confidence interval approach provided by the multinomial distribution. Tortora (1978) has devised such a technique for simple random sampling in a multinomial population, and has concluded that the binomial approximation underestimates the required sample size. The degree of understatement ranges by a factor of 1.7 to 2.4 as the number of categories increases from 3 to 10.

Several authors have approached the related problem of optimal allocation of resources in two-stage sampling designs when

Table 8.—Total catch per week (stratum), weighting factor per week, mean age proportion and variance for ages 1, 2, and 3+ by week, and the overall 1979 seasonal mean proportions ($\bar{P}_{ST}$), variances, and approximate 95% confidence intervals for Port Monmouth, N.J.

Week	Total catch ($\times10^6$)	Weighting factor	Age 1		Age 2		Age 3+	
			$\bar{P}_1$(%)	v($\bar{P}_1$)	$\bar{P}_2$(%)	v($\bar{P}_2$)	$\bar{P}_{3+}$(%)	v($\bar{P}_{3+}$)
6/16	9.35	0.09154	—	—	62.2	4.66×10^{-3}	37.8	4.66×10^{-3}
6/23	6.84	0.06697	—	—	46.0	6.60×10^{-3}	54.0	6.60×10^{-3}
6/30	6.84	0.06697	—	—	70.0	5.00×10^{-3}	30.0	5.00×10^{-3}
7/ 7	10.44	0.10221	—	—	95.0	5.00×10^{-4}	5.0	5.00×10^{-4}
7/14								
7/21	13.68	0.13393	—	—	79.0	4.77×10^{-3}	21.0	4.77×10^{-3}
7/28	6.24	0.06109	—	—	39.3	1.27×10^{-2}	60.7	1.27×10^{-2}
8/ 4	9.66	0.09458	1.4	2.04×10^{-4}	27.1	1.61×10^{-2}	71.4	1.59×10^{-2}
8/11	8.32	0.08146	—	—	30.0	1.27×10^{-2}	70.0	1.27×10^{-2}
8/18	8.01	0.07842	—	—	20.4	1.35×10^{-3}	79.6	1.35×10^{-3}
8/25								
9/ 1	2.35	0.02301	—	—	90.0	5.00×10^{-3}	10.0	5.00×10^{-3}
9/ 8	3.95	0.03867	2.5	6.25×10^{-4}	52.0	3.56×10^{-2}	45.5	4.25×10^{-2}
9/15	4.15	0.04063	—	—	2.5	6.25×10^{-4}	97.5	6.25×10^{-4}
9/22	5.85	0.05727	—	—	31.7	1.89×10^{-2}	68.3	1.89×10^{-2}
9/29	2.05	0.02007	—	—	64.2	1.40×10^{-2}	35.8	1.40×10^{-2}
10/ 6	2.17	0.02125	—	—	35.0	1.75×10^{-2}	65.0	1.75×10^{-2}
10/27	2.24	0.02193	—	—	22.8	2.07×10^{-3}	77.2	2.07×10^{-3}

		$\bar{P}_{ST}$	v($\bar{P}_{ST}$)	≅ 95% Confidence Interval
Seasonal estimates	Age 1	0.23%	2.76×10^{-6}	±0.33%
	Age 2	51.29%	5.99×10^{-4}	±4.89%
	Age 3+	48.47%	6.08×10^{-4}	±4.93%

Table 9.—Total catch per week (stratum), weighting factor per week, mean age proportion and variance for ages 0, 1, 2, and 3+ by week, and the overall 1979 seasonal mean proportions ($\bar{P}_{ST}$), variances, and approximate 95% confidence intervals for Reedville, Va.

Week	Total catch ($\times10^6$)	Weighting factor	Age 0		Age 1		Age 2		Age 3+	
			$\bar{P}_0$(%)	v($\bar{P}_0$)	$\bar{P}_1$(%)	v($\bar{P}_1$)	$\bar{P}_2$(%)	v($\bar{P}_2$)	$\bar{P}_{3+}$(%)	v($\bar{P}_{3+}$)
5/26	36.80	0.02646	—	—	2.5	2.68×10^{-4}	96.3	6.92×10^{-4}	1.3	1.56×10^{-4}
6/ 2	55.30	0.03976	—	—	—	—	98.0	1.78×10^{-4}	2.0	1.78×10^{-4}
6/ 9	55.36	0.03980	—	—	—	—	97.0	4.56×10^{-4}	3.0	4.56×10^{-4}
6/16	69.01	0.04961	—	—	3.0	4.56×10^{-4}	97.0	4.56×10^{-4}	—	—
6/23	59.66	0.04289	—	—	2.0	1.78×10^{-4}	97.0	2.33×10^{-4}	1.0	1.00×10^{-4}
6/30	59.36	0.04267	—	—	23.0	1.16×10^{-2}	77.0	1.16×10^{-2}	—	—
7/ 7	62.74	0.04511	—	—	22.9	7.00×10^{-3}	77.1	7.00×10^{-3}	—	—
7/14	59.87	0.04304	—	—	18.0	4.40×10^{-3}	82.0	4.40×10^{-3}	—	—
7/21	42.28	0.03040	—	—	18.2	5.77×10^{-3}	81.8	5.77×10^{-3}	—	—
7/28	71.71	0.05155	—	—	8.0	1.07×10^{-3}	92.0	1.07×10^{-3}	—	—
8/ 4	47.42	0.03409	—	—	9.2	1.61×10^{-3}	90.8	1.61×10^{-3}	—	—
8/11	39.18	0.02817	—	—	11.2	3.01×10^{-3}	79.4	3.92×10^{-3}	9.3	3.90×10^{-3}
8/18	67.21	0.04832	—	—	12.1	1.73×10^{-3}	86.9	2.23×10^{-3}	1.0	1.00×10^{-4}
8/25	30.19	0.02170	—	—	7.0	9.00×10^{-4}	88.0	1.07×10^{-3}	5.0	5.00×10^{-4}
9/ 1	44.76	0.03218	—	—	12.5	3.12×10^{-3}	75.0	2.50×10^{-3}	12.5	3.84×10^{-3}
9/ 8	48.73	0.03503	—	—	47.2	6.10×10^{-3}	52.8	6.10×10^{-3}	—	—
9/15	51.22	0.03682	—	—	65.0	3.57×10^{-3}	35.0	3.57×10^{-3}	—	—
9/22	58.02	0.04171	2.0	4.00×10^{-4}	32.1	1.03×10^{-2}	58.7	8.81×10^{-3}	7.2	1.86×10^{-3}
9/29	48.03	0.03453	5.0	1.17×10^{-3}	66.7	4.44×10^{-3}	28.3	4.94×10^{-3}	—	—
10/ 6	48.14	0.03461	—	—	38.4	8.21×10^{-3}	60.6	8.12×10^{-3}	1.0	1.00×10^{-4}
10/13	59.09	0.04248	1.2	1.56×10^{-4}	65.0	1.46×10^{-2}	33.8	1.46×10^{-2}	—	—
10/20	147.84	0.10628	—	—	78.7	6.84×10^{-3}	21.3	6.84×10^{-3}	—	—
10/27	51.43	0.03697	—	—	68.6	1.70×10^{-2}	28.6	1.32×10^{-2}	2.7	7.72×10^{-4}
11/ 3	45.16	0.03247	—	—	57.1	1.57×10^{-2}	42.9	1.57×10^{-2}	—	—
11/10	32.47	0.02334	2.5	2.68×10^{-4}	82.5	1.53×10^{-2}	11.2	7.66×10^{-3}	3.8	1.41×10^{-3}

		$\bar{P}_{ST}$	v($\bar{P}_{ST}$)	≅ 95% Confidence Interval
Seasonal estimates	Age 0	0.37%	2.51×10^{-6}	±0.32%
	Age 1	32.28%	2.62×10^{-4}	±3.24%
	Age 2	65.73%	2.54×10^{-4}	±3.19%
	Age 3+	1.62%	1.40×10^{-5}	±0.75%

Table 10.—Total catch per week (stratum), weighting factor per week, mean age proportion and variance for ages 0, 1, 2, and 3+ by week, and the overall 1979 seasonal mean proportions ($\bar{P}_{ST}$), variances, and approximate 95% confidence intervals for Beaufort, N.C.

Week	Total catch ($\times 10^6$)	Weighting factor	Age 0 $\bar{P}_0$(%)	Age 0 $v(\bar{P}_0)$	Age 1 $\bar{P}_1$(%)	Age 1 $v(\bar{P}_1)$	Age 2 $\bar{P}_2$(%)	Age 2 $v(\bar{P}_2)$	Age 3+ $\bar{P}_{3+}$(%)	Age 3+ $v(\bar{P}_{3+})$
5/12	11.43	0.00737	—	—	—	—	96.7	1.11×10^{-3}	3.3	1.11×10^{-3}
5/19	14.44	0.00931	—	—	5.0	2.50×10^{-3}	95.0	2.50×10^{-3}	—	—
5/26	6.94	0.00447	—	—	3.3	1.11×10^{-3}	96.7	1.11×10^{-3}	—	—
6/ 2	12.43	0.00801	—	—	12.0	5.40×10^{-3}	86.0	5.60×10^{-3}	2.0	4.00×10^{-4}
6/ 9	11.16	0.00719	—	—	—	—	97.5	6.25×10^{-4}	2.5	6.25×10^{-4}
6/16	18.57	0.01197	—	—	8.3	2.28×10^{-3}	88.3	1.61×10^{-3}	3.3	4.44×10^{-4}
6/23	15.68	0.01010	—	—	12.5	3.96×10^{-3}	87.5	3.96×10^{-3}	—	—
6/30	18.16	0.01170	—	—	12.0	6.40×10^{-3}	86.0	9.60×10^{-3}	2.0	4.00×10^{-4}
7/ 7	22.34	0.01440	—	—	40.0	1.60×10^{-1}	60.0	1.60×10^{-1}	—	—
7/14	27.76	0.01789	—	—	15.0	2.50×10^{-3}	85.0	2.50×10^{-3}	—	—
7/21	6.18	0.00399	—	—	7.5	2.29×10^{-3}	90.0	1.67×10^{-3}	2.5	6.25×10^{-4}
7/28	2.12	0.00137	—	—	10.4	1.37×10^{-5}	86.3	1.00×10^{-3}	3.3	1.11×10^{-3}
8/ 4	3.05	0.00196	—	—	15.0	2.25×10^{-2}	65.0	2.50×10^{-3}	20.0	4.00×10^{-2}
8/11	0.95	0.00061	—	—	5.0	2.50×10^{-3}	95.0	2.50×10^{-3}	—	—
8/18										
8/25										
9/ 1	5.76	0.00371	50.0	2.50×10^{-1}	—	—	40.0	1.60×10^{-1}	10.0	1.00×10^{-2}
9/ 8										
9/15										
9/22										
9/29										
10/ 6	2.33	0.00150	—	—	76.3	7.30×10^{-3}	23.7	7.30×10^{-3}	—	—
10/13										
10/20	1.22	0.00078	—	—	37.5	2.06×10^{-2}	62.5	2.06×10^{-2}	—	—
10/27	3.83	0.00247	40.0	6.00×10^{-2}	8.0	6.40×10^{-3}	44.0	3.76×10^{-2}	8.0	6.40×10^{-3}
11/ 3	4.76	0.00307	55.0	3.18×10^{-2}	21.7	1.16×10^{-2}	23.3	2.58×10^{-2}	—	—
11/10	3.86	0.00249	42.9	4.08×10^{-2}	—	—	54.3	3.76×10^{-2}	2.8	8.16×10^{-4}
11/17										
11/24	14.82	0.00955	—	—	10.9	2.81×10^{-3}	79.6	2.92×10^{-3}	9.5	1.63×10^{-3}
12/ 1	6.83	0.00440	—	—	17.2	9.53×10^{-3}	75.6	8.29×10^{-3}	7.2	9.69×10^{-4}
12/ 8	65.48	0.04220	28.7	8.61×10^{-3}	41.4	8.15×10^{-3}	26.8	5.51×10^{-3}	3.2	4.76×10^{-4}
12/15	91.76	0.05913	59.2	1.44×10^{-2}	7.4	1.09×10^{-3}	23.8	9.03×10^{-3}	9.6	3.91×10^{-3}
12/22	409.01	0.26359	92.1	1.56×10^{-3}	7.9	1.56×10^{-3}	—	—	—	—
12/29										
1/ 5	595.74	0.38394	100.0	0.00	—	—	—	—	—	—
1/12	175.05	0.11282	100.0	0.00	—	—	—	—	—	—

		$\bar{P}_{ST}$	$v(\bar{P}_{ST})$	≅ 95% Confidence Interval
Seasonal estimates	Age 0	79.22%	1.78×10^{-4}	±2.67%
	Age 1	6.12%	1.64×10^{-4}	±2.56%
	Age 2	13.59%	8.14×10^{-5}	±1.80%
	Age 3+	1.06%	1.53×10^{-5}	±0.78%

Table 11.—Estimated age composition for the combined catch of Port Monmouth, N.J., Reedville, Va., and Beaufort, N.C., during the 1979 fishing season. Variances and approximate 95% confidence intervals are given for the age proportion estimates.

	$\bar{P}_{ST}$	$v(\bar{P}_{ST})$	≅ 95% Confidence Intervals
Age 0	40.54%	4.69×10^{-5}	± 1.37%
Age 1	17.87%	9.72×10^{-5}	± 1.97%
Age 2	38.67%	7.48×10^{-5}	± 1.73%
Age 3+	2.91%	7.57×10^{-6}	± 0.55%

more than one variable is measured. Saila et al. (1976) pointed out that the traditional approach requires solutions for the optimal primary and secondary sample size for each variable separately, followed by a compromise to reach a final allocation decision. Saila et al. (1976) acknowledged that this solution is not necessarily optimal and may incur excessive costs in large surveys, but their final recommendation appears to be an inefficient compromise. Schweigert and Sibert (1983) determined the optimal allocation of sampling resources for a multiage fishery, but their method is similar to the one proposed by Saila et al. (1976). The more correct approach would be a simultaneous solution for all age groups that minimizes cost subject to several variance constraints. This approach relies on an iterative solution (Kelley 1971) that may incur significant computer costs; a graphical solution may be more practical.

The allocation of sampling effort when two age classes occur is demonstrated for the Port Monmouth data (Table 5). The theoretical variance of the population is:

$$v(\bar{p}_j) = \frac{\sigma_a^2}{n} + \frac{\sigma_w^2}{nm},$$

where σ_a^2 and σ_w^2 may best be estimated by s_a^2 and s_w^2. These estimates are derived by analogy:

$$s_a^2 = s_1^2 - \frac{s_2^2}{m} = \frac{0.33556}{8} - \frac{(10)\,(1.78)}{(9)\,(9)\,(10)} = 0.01997$$

$$s_w^2 = s_2^2 = \frac{(10)\,(1.78)}{(9)\,(9)} = 0.219753.$$

For a new sample we would, therefore, predict the variance of the average proportion of age 2 fish to be:

$$\mathrm{v}(\bar{p}_2) = \frac{s_a^2}{n} + \frac{s_w^2}{nm} = \frac{0.01997}{9} + \frac{0.219753}{(9)\,(10)} = 0.004661.$$

Different values of n and m may be computed to satisfy this relationship for any specified variance. Assuming the same overall cost function as before (c_1:c_2 = 6), optimal subsample size may be found by:

$$m_{\mathrm{opt}} = \sqrt{\frac{c_1 s_w^2}{c_2 s_a^2}} = \sqrt{\frac{(6)\,(0.219753)}{(1)\,(0.01997)}} = 9.$$

The optimal number of primary units would be found based on the total budget available for sampling or from the predetermined variance requirement. The calculations for m_{opt} are only computed for the one example shown, but for that case the optimal subsample size is larger for age proportions than for average weight. Again, this conclusion depends on the actual age structure of the menhaden population at different times and places. Individual sets at Port Monmouth appear to be nonuniform with respect to age distribution, so larger subsamples are indicated. This result will hold in all but cases of single (or almost single) age populations.

When more than two age groups are present, the optimal allocation of sampling effort is found by minimizing the cost of the sampling program, provided that predetermined levels of variance are satisfied for each proportion. As Tortora (1978), Saila et al. (1976), and Kelley (1971) illustrated, this problem is difficult, and I will be addressing it in a future report.

Number of fish at age.—To conclude this section, I estimate the number of fish in each age group and generate an approximate solution to the variance function. The present method requires that the total number of fish landed, C_{pw}, be multiplied by the proportion of fish in age group j, $\bar{p}_j$:

$$C_{\mathrm{pw}j} = (C_{\mathrm{pw}})\,(\bar{p}_j), \tag{25}$$

where $C_{\mathrm{pw}j}$ is the number of age j fish landed per port week. An estimate of the variance of $C_{\mathrm{pw}j}$ requires an expression for the variance of the product of two random variables. A first approximation of this relationship that has been previously applied to fishery data (Southward 1976) is:

$$\mathrm{v}(C_{\mathrm{pw}j}) = (C_{\mathrm{pw}})^2\,\mathrm{v}(\bar{p}_j) + (\bar{p}_j)^2\,\mathrm{v}(C_{\mathrm{pw}}). \tag{26}$$

This equation ignores any possible covariance between the estimates of catch and age proportions. Clearly, however, the estimate of catch is not independent of age structure, since, within any one set, average weight and age are related. That is, a greater proportion of young fish generally implies a lower average weight, which in turn leads to a higher estimate of total catch. The covariance terms are included in the following modification of the exact representation of the variance proposed by Goodman (1960):

$$\begin{aligned}\mathrm{v}(C_{\mathrm{pw}j}) = {} & C_{\mathrm{pw}}^2\,\mathrm{v}(\bar{P}_j) + \bar{P}_j^2\,\mathrm{v}(C_{\mathrm{pw}}) \\ & + 2\,C_{\mathrm{pw}}\,\bar{P}_j\,\frac{\Sigma(C_{\mathrm{pw}}-\bar{C}_{\mathrm{pw}})\,(P_{ij}-\bar{P}_j)}{n(n-1)} \\ & + 2\,C_{\mathrm{pw}}\,\frac{\Sigma(C_{\mathrm{pw}}-\bar{C}_{\mathrm{pw}})\,(P_{ij}-\bar{P}_j)^2}{n(n-1)} \\ & + 2\,\bar{P}_j\,\frac{\Sigma(C_{\mathrm{pw}}-\bar{C}_{\mathrm{pw}})^2\,(P_{ij}-\bar{P}_j)}{n(n-1)} \\ & + \frac{\Sigma(C_{\mathrm{pw}}-\bar{C}_{\mathrm{pw}})^2\,(P_{ij}-\bar{P}_j)^2}{n(n-1)} - \left[\frac{\Sigma(C_{\mathrm{pw}}-\bar{C}_{\mathrm{pw}})\,(P_{ij}-\bar{P}_j)}{n(n-1)}\right]^2.\end{aligned} \tag{27}$$

The method of solution for the Port Monmouth data is given in Table 12. In this case, the 95% confidence region extended to ± 26.3% of the number of age 2 fish caught (5,815,054 ± 1,526,988 fish). As in the case of total catch, various age-specific catches and their variances may be summed to obtain estimates and confidence intervals over time and space. We have accomplished this by port over the entire 1979 fishing season and arrive at confidence intervals that are very similar to those obtained by multiplying the upper and lower confidence bounds of the port-season age proportion estimate by total number of fish landed.

CONCLUSIONS AND RECOMMENDATIONS

In this paper I have attempted to evaluate constructively the effectiveness of the present menhaden sampling scheme within the framework of probability sampling theory. The derivation of the basic population statistics and their associated variance functions has been accomplished, within the constraints of stated assumptions. In so doing, I have examined the correspondence of those who have addressed this problem previously, explored the literature available on related fishery sampling problems, interviewed several people who have been involved with data collection and analysis, and analyzed a limited body of data to gain a preliminary feel for the sampling variability.

In general, a theoretical approach is outlined which is consistent with present estimation techniques and can be applied to the raw data. Although it appears, from the limited amount of data considered, that the current number of fish sampled per set is adequate for estimating mean weight and total catch, it may be too small for estimating age composition. It also appears, at least for the 1979 data, that a greater number of sets should be sampled. The specific number depends on the area and time concerned, but generally no less than 10 sets/port-week should be sampled, since fewer degrees of freedom will raise the t-table value used to calculate confidence intervals.

Ultimately it is up to fishery planners to decide what level of uncertainty is acceptable. This, of course, depends on the actual use of the data and the sensitivity of management decisions. If information for every port-week combination is critical, and if 1979 data are representative, then we probably need to sample more intensively. If summary information over the fishing season is most important, then our level of sampling may be sufficient. For example, on a weekly basis, 95% confidence intervals about numbers of fish landed average ± 14.7% (range 3.3-33.3%), ±

Table 12.—Estimate of the number of age 2 fish landed at Port Monmouth, N.J., during the week ending 6/16/79, with method to calculate approximate 95% confidence intervals. Data taken from Table 5.

Sampled set no.	Mean weight ($\bar{y}_i$)	Total catch based on $\bar{y}_i$ (C_{pwi})	Proportion of age 2 fish (P_{i2})
1	284.1	8.296×10^6	0.6
2	321.8	7.324×10^6	0.2
3	255.7	9.217×10^6	0.4
4	247.4	9.527×10^6	0.8
5	224.3	10.508×10^6	0.7
6	234.8	10.038×10^6	0.7
7	220.5	10.669×10^6	0.8
8	248.7	9.477×10^6	0.6
9	232.4	10.141×10^6	0.8

$\bar{P}_2 = 0.62222 \quad v(\bar{P}_2) = 0.00466 \quad SE = 0.06827$

Number of age 2 fish landed $= C_{pw}\bar{P}_2 = (9.346\times10^6)(0.62222) = 5.815\times10^6$ fish

$$v(C_{pwj}) = C_{pw}^2\, v(\bar{P}_j) + \bar{P}_j^2\, v(C_{pw}) + 2\, C_{pw}\, \bar{P}_j\, \frac{\Sigma(C_{pw}-\bar{C}_{pw})(P_{ij}-\bar{P}_j)}{n(n-1)}$$

$$+ 2\, C_{pw}\, \frac{\Sigma(C_{pw}-\bar{C}_{pw})(P_{ij}-\bar{P}_j)^2}{n(n-1)} + 2\, \bar{P}_j\, \frac{\Sigma(C_{pw}-\bar{C}_{pw})^2(P_{ij}-\bar{P}_j)}{n(n-1)}$$

$$+ \frac{\Sigma(C_{pw}-\bar{C}_{pw})^2(P_{ij}-\bar{P}_j)^2}{n(n-1)} - \left[\frac{\Sigma(C_{pw}-\bar{C}_{pw})(P_{ij}-\bar{P}_j)}{n(n-1)}\right]^2$$

$$v(C_{pwj}) = (9.346\times10^6)^2(0.00466) + (0.622)^2(1.014\times10^{11})$$

$$+ 2(9.346\times10^6)(0.622)\left(\frac{1.459\times10^6}{(8)(9)}\right) + 2(9.346\times10^6)\left(\frac{-3.239\times10^5}{(8)(9)}\right)$$

$$+ 2(0.622)\left(\frac{-1.532\times10^{12}}{(8)(9)}\right) + \left(\frac{8.936\times10^{11}}{(8)(9)}\right) - \left(\frac{1.459\times10^6}{(8)(9)}\right)^2$$

$v(C_{pwj}) = 5.829\times10^{11} \qquad SE = 763{,}494$ fish

95% Confidence Interval $\cong 5{,}815{,}054 \pm 2(763{,}494)$
$\cong 5{,}815{,}054 \pm 26.3\%$

19.2% (5.7-45.4%), and ± 94.4% (11.3-395.0%) for Port Monmouth, Reedville, and Beaufort, respectively. Corresponding confidence intervals by season, however, are ± 5.1%, ± 4.3%, and ± 11.8% (Table 7). Similarly, confidence bounds about mean age proportions are much narrower for the entire season than for individual weeks (Tables 8-11).

Throughout this discussion, the assumptions basic to probability sampling theory have been emphasized, and potential breaches have been identified. The most serious violations stem from 1) the method of set selection, and 2) the use of sets as the primary sampling unit.

With regard to set selection there are at least two areas of concern. First, it appears that when trips occur both nearshore and offshore (e.g., inside and outside Chesapeake Bay) the last set taken is more likely to occur close to shore. Consequently, these nearshore sets are prone to oversampling since they are the ones available to port samplers. Given the current sampling design, this flaw is difficult to correct, but the introduced bias may be small, since a minority of vessel trips (35% for Reedville) are involved. It may also be possible to estimate the bias by stratifying the sample into nearshore and offshore sets and comparing estimates of average age and size.

The second area of concern is that vessels themselves are not being sampled at random. This nonrandom behavior can occur on the level of plant or individual vessel. For example, at Reedville, plant B appears to be preferentially sampled over plant A for the 1979 season. In addition, within plants, certain vessels are more or less likely to be sampled (Table 3). This problem may be avoided by providing port samplers with explicit instructions dictating the method to be used for vessel selection. As is true for nearshore vs. offshore sets, the degree of bias introduced by nonrandom selection depends on actual population differences between sets selected for and sets selected against.

The second major problem involves the implicit assumption that set size is uniform with respect to numbers of fish. When set size does vary, current sampling theory requires a knowledge of these sizes. Paradoxically, that is the very quantity we want to estimate. As a consequence, the range of available procedures is narrowed. It is essential that we explore the effect of unknown and unequal set size on the parameter estimates. In most cases, for the data studied, violating the assumption of equal set size does not appear to appreciably bias the estimates, except perhaps for the North Carolina fall fishery (Table 4). Even with unequal set sizes it is still possible to estimate some population characters if accurate estimates of each set's approximate weight is recorded, but the calculation can become rather circular. Use of captains' daily fishing reports coupled with weight data from the sampled set can provide the estimate of relative set size necessary to a weighted calculation (ratio-to-size estimate). Appropriate variance formulations may be derived from Cochran (1977).

The advantages of a different sampling design might also be considered. For example, fewer assumptions would be needed if, instead of sampling sets at random, we could sample full boatloads. To accomplish this would require that fish be sampled at various and randomly selected times throughout the unloading process. This sort of sampling plan might offer at least one other advantage. Instead of estimating numbers of specific age fish indirectly from total catch and age composition, we could calculate the numbers directly. Tomlinson (1971) offered a similar plan for schooling pelagic fish and applied it to the California anchovy fishery. Basically, primary units are taken to be boatloads and secondary units are now defined as clusters of fish, each cluster weighing a known amount. By taking two or more random clusters from each boatload sampled, we could obtain replicate counts of the numbers of specific age fish. Since we know (approximately) the weight of each boatload and the weight of a cluster, we can estimate the number of secondary units in a boatload (assuming nonoverlapping clusters). It is then possible to devise a sampling scheme whereby boatloads are sampled with probabilities proportional to estimated catch size, a method which often yields the smallest sampling variance.

Another avenue of research is the optimal allocation of sampling resources in the case of a multiage fishery. As indicated earlier, we have been exploring this problem and feel that significant progress can be made, at least in generating a general solution. Initially we will examine the behavior of the age-specific variance function for representative menhaden samples. After specifying a cost function, our goal will be to derive graphical and analytical procedures for estimating the optimal number of primary and secondary sampling units.

ACKNOWLEDGMENTS

Robert Chapoton, Dean Ahrenholz, and William Nicholson made valuable contributions to this project by acquainting me with the realities and practicalities of the menhaden survey and by acting as sounding boards for many of the ideas expressed in this report. I would also like to acknowledge James Waters for the many hours of discussion on sampling theory and optimal allocation. Many of his ideas are represented here and we hope to collaborate further on the project. Mayo Judy contributed to the report by providing summaries from daily fishing reports used to investigate random sampling assumptions. Finally, I thank Charles Krouse for developing computer programs to estimate sampling statistics from raw bioprofiles and catch information.

LITERATURE CITED

COCHRAN, W. G.
1977. Sampling techniques. 3d ed. Wiley, N.Y. 428 p.
GOODMAN, L. A.
1960. On the exact variance of products. J. Am. Stat. Assoc. 55:708-713.
GULLAND, J. A.
1966. Manual of sampling and statistical methods for fisheries biology. Part 1. Sampling methods. FAO Man. Fish. Sci. 3, 87 p.
JUNE, F. C.
1972. Variations in size and length composition of Atlantic menhaden groupings. Fish. Bull., U.S. 70:699-713.
JUNE, F. C., and J. W. REINTJES.
1959. Age and size composition of the menhaden catch along the Atlantic coast of the United States, 1952-55; with a brief review of the commercial fishery. U.S. Fish Wildl. Serv., Spec. Sci. Rep. Fish. 317, 65 p.
KELLEY, J. C.
1971 Multivariate oceanographic sampling. Math. Geol. 3:43-50.
KENNEY, B. C.
1982. Beware of spurious self-correlations! Water Resour. Res. 18:1041-1048.
NELSON, W. R., M. C. INGHAM, and W. E. SCHAAF.
1977. Larval transport and year-class strength of Atlantic menhaden, *Brevoortia tyrannus*. Fish. Bull., U.S. 75:23-41.
NICHOLSON, W. R.
1971a. Coastal movements of Atlantic menhaden as inferred from changes in age and length distributions. Trans. Am. Fish. Soc. 100:708-716.
1971b. Changes in catch and effort in the Atlantic menhaden purse-seine fishery 1940-68. Fish. Bull., U.S. 69:765-781.
1972. Population structure and movements of Atlantic menhaden, *Brevoortia tyrannus,* as inferred from back-calculated length frequencies. Chesapeake Sci. 13:161-174.
POPE, J. A.
1956. An outline of sampling techniques. Rapp. P.-V. Réun. Cons. Int. Explor. Mer 140:11-20.
SAILA, S. B., R. A. PIKANOWSKI, and D. S. VAUGHAN.
1976. Optimum allocation strategies for sampling benthos in the New York Bight. Estuarine Coastal Mar. Sci. 4:119-128.
SCHAAF, W. E.
1975. Status of the Gulf and Atlantic menhaden fisheries and implications for resource management. Mar. Fish. Rev. 37(9):1-9.
1979. An analysis of the dynamic population response of Atlantic menhaden, *Brevoortia tyrannus,* to an intensive fishery. Rapp. P.-V. Réun. Cons. Int. Explor. Mer 177:243-251.
SCHAAF, W. E., and G. R. HUNTSMAN.
1972. Effects of fishing on the Atlantic menhaden stock: 1955-1969. Trans. Am. Fish. Soc. 101:290-297.
SCHAAF, W. E., J. E. SYKES, and R. B. CHAPOTON.
1975. Forecasts of Atlantic and Gulf menhaden catches based on the historical relation of catch and fishing effort. Mar. Fish. Rev. 37(10):5-9.
SCHWEIGERT, J. F., and J. R. SIBERT.
1983. Optimizing survey design for determining age structure of fish stocks: An example from British Columbia Pacific herring (*Clupea harengus pallasi*). Can. J. Fish. Aquat. Sci. 40:588-597.
SNEDECOR, G. W., and W. G. COCHRAN.
1967. Statistical methods. 6th ed. Iowa State Univ. Press, Ames, 593 p.
SOKAL, R. R., and F. J. ROHLF.
1981. Biometry. 2d ed. Freeman, San Franc., 859 p.
SOUTHWARD, G. M.
1976. Sampling landings of halibut for age composition. Int. Pac. Halibut Comm. Sci. Rep. 58, 31 p.
SUKHATME, P. V., and B. V. SUKHATME.
1970. Sampling theory of surveys with applications. 2d ed. Iowa State Univ. Press, Ames, 452 p.
TOMLINSON, P. K.
1971. Some sampling problems in fishery work. Biometrics 27:631-641.
TORTORA, R. D.
1978. A note on sample size estimation for multinomial populations. Am. Stat. 32:100-102.
WILLIAMS, W. H.
1978. A sampler on sampling. Wiley, N.Y., 254 p.

APPENDIX

Derivation of Variance Formulae in Two-Stage Sampling for Age Proportions

According to Cochran (1977), the among set and within set components of variance for the mean proportion of age j fish in a two-stage design are:

$$S_1^2 = \frac{\sum_{i=1}^{n} (P_{ij} - \bar{P}_j)^2}{n-1}$$

$$S_2^2 = \frac{m}{n\,(m-1)} \sum_{i=1}^{n} P_{ij}\, q_{ij}.$$

The best way to appreciate these formulae is by utilizing the device of coding fish in age class j as 1 and all other fish as 0 and perform an ANOVA. For example in Table 3 for age 2 fish:

	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1
	1	0	1	1	1	1	1	1	1
	1	0	1	1	1	1	1	1	1
	1	0	0	1	1	1	1	1	1
	1	0	0	1	1	1	1	1	1
	0	0	0	1	1	1	1	0	1
	0	0	0	1	0	0	1	0	1
	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0
P_{i2}	.6	.2	.4	.8	.7	.7	.8	.6	.8

ANOVA table

Source	df	SS	MS
Total	89	21.15556	
Sets	8	3.35556	0.41944
Error	81	17.80000	0.21975

$$S_1^2 = \frac{\Sigma(P_{ij}-\bar{P}_j)^2}{n-1} = \frac{0.33556}{8} = 0.04195$$

$$\text{SET MS} = S_1^2 * m = (0.04195)\,(10) = 0.41945$$

$$S_2^2 = S_W^2 = \text{ERROR MS} = \frac{m}{n(m-1)} \Sigma P_{ij}\, q_{ij}$$

$$= \frac{10\ (1.78)}{(9)\ (9)} = 0.21975$$

$$S_2^2 = \frac{m}{n(m-1)} \Sigma P_{ij}\, q_{ij} = \frac{\sum_i \sum_j (X_{ij}-\bar{X})^2}{(m-1)\ (n)}.$$

Proof that Cochran's (1977) formula for S_2^2 is equal to the error mean square (S_W^2) of the ANOVA.

$$\frac{m}{n(m-1)} \Sigma P_{ij}\, q_{ij} = \frac{\Sigma\, \Sigma\, (X_{ij}-\bar{X}_i)^2}{(m-1)\ (n)}$$

$m\ \Sigma\ P_{ij}\, q_{ij} = \Sigma\ \Sigma\ (X_{ij} - \bar{X}_i)^2 =$ sum of each set's sum of squares

let $X_{ij} = 1$ if in class j and 0 if not in class j

$$m\ \Sigma \left(\frac{\Sigma X_i}{m} * \frac{m-\Sigma X_i}{m} \right) = \Sigma\ \Sigma\ (X_{ij} - \bar{X}_i)^2$$

$$\frac{1}{m}\ \Sigma\ \left[(\Sigma X_i)m - (\Sigma X_i)^2\right] = \Sigma\ \Sigma\ (X_{ij} - \bar{X}_i)^2$$

$$\Sigma\ [\ \Sigma X_i - \frac{(\Sigma X_i)^2}{m}\ ] = \Sigma\ \Sigma\ (X_{ij} - \bar{X}_i)^2$$

$$\Sigma\ [\ \Sigma X_i - \frac{(\Sigma X_i)^2}{m}\ ] = \Sigma\ (\Sigma X_i^2 - \frac{[\Sigma X_i]^2}{m}\ )$$

this is true because when $X_i = 1,\ X_i^2 = 1$
$\therefore\ \Sigma X_i = \Sigma X_i^2.$