

UC San Diego

Fish Bulletin

Title

Fish Bulletin No. 120. Estimating Absolute Age Composition of California Salmon Landings

Permalink

<https://escholarship.org/uc/item/3r6120x8>

Author

Kutkuhn, Joseph H

Publication Date

1962-05-01

**THE RESOURCES AGENCY OF CALIFORNIA
DEPARTMENT OF FISH AND GAME
FISH BULLETIN No. 120
Estimating Absolute Age Composition of California Salmon Landings**



JOSEPH H. KUTKUHN
1963

TABLE OF CONTENTS

	Page
INTRODUCTION -----	3
Limitations of the Processed Data -----	4
DRAWING THE SAMPLE -----	5
Distribution of Landings and Sampling Effort -----	5
Access to Individual Catches: Choice of Vessel -----	6
Consolidation of Raw Data -----	7
Age Determination -----	7
Definition of the Parent Population -----	8
AGE STRUCTURE ESTIMATES: PORT-MONTH LANDINGS --	9
Two-phase Sampling for Stratification -----	10
Age Sample Size for a Fixed Length Sample -----	13
Choice of Stratification -----	15
Two-phase Versus Simple Random Sampling with Varying Age Sample Costs -----	17
ESTIMATING OVER-ALL AGE COMPOSITION -----	19
SUMMARY -----	27
ACKNOWLEDGMENTS -----	28
REFERENCES -----	28
APPENDIXES -----	31

1. INTRODUCTION

Drawing valid inferences about parent populations is directly dependent upon the precision or reliability of estimated parameters. Self-evident as this statement is, however, situations frequently arise wherein unwarranted conclusions may be drawn from sample information whose reliability is imperfectly known. Thus it is clear that wherever possible, estimates of any sort should be supplemented with some measure of their precision. This implies, of course, that assumptions underlying any application of modern sampling theory have been reasonably satisfied.

The present paper represents an attempt to assess estimates that permit delineating the absolute age structure of fish catches, assuming the appropriate attributes are observed in a manner that at least approximates probability sampling. Such procedure further provides a basis for modifying sampling designs so future estimates derived therefrom will possess some *a priori* degree of precision imposed in accordance with available manpower and monetary resources.

Although specifically concerned with determining age composition of California landings of king salmon, *Oncorhynchus tshawytscha* (Walbaum), the techniques discussed herein could be applied to other species whose catches are handled in similar fashion. Moreover, they would be appropriate where dealing with any of several attributes commonly descriptive of fish populations. Besides the number of scale annuli, constituting the attribute of interest in the present paper, these include: (1) characteristics defining each kind of fish subject to capture; or, perhaps, for a given kind of fish, characteristics that would identify a member of a particular subpopulation, aggregations of which may comprise a commonly fished stock; (2) any artificially applied "mark"; (3) a particular class of lengths or weights; and, (4) sex.

Designed to provide a basis for predicting year-to-year king salmon yield potential, the parent study from which part of my data were selected evolved during the recovery phase of a large-scale salmon marking program, instituted in 1949 by the California Department of Fish and Game in cooperation with the Pacific Marine Fisheries Commission (Hallock, *et al*, 1952). In addition to sampling the California commercial salmon landings for marked fish, it was necessary to sample concurrently for other attributes (Fry and Hughes, 1953). During the program's early stages, it was further decided that still others capable of yielding potentially valuable information, *viz.*, length and age, could be sampled incidentally with little added effort and expense. Fry and Hughes (1952) briefly outlined the commercial catch sampling program set up prior to the age study's outset.

To supplement those secured from the above source, I have used additional data collected during a concomitant study of the ocean sport salmon catch (Dingell-Johnson Project California F-12-R: "Ocean Salmon Study") at various points to help develop the methods discussed.

1.1. LIMITATIONS OF THE PROCESSED DATA

Before discussing procedures involved in ascertaining the age structure of an aggregate of fish, it might be well to acknowledge the limitations of such information as determined for king salmon caught by trolling off California. Notwithstanding the fact that age-structure knowledge is important if not necessary for managing any exploited fish population, for salmon, several factors minimized the value of the information derived even though the sampling techniques may have led to very precise descriptions of catch age structure.

Of these, primary consideration must be given the fact that none of the several West Coast king salmon subpopulations is definable in the ocean catch. Vital statistics such as year-class abundance and fishing rates, obtained by periodically examining a population's age structure, have little meaning when they relate to a fished stock whose component subpopulations may, with respect to one another, be subject to considerable fluctuation per unit time. It is clear then that prior to considering a salmon subpopulation's dynamics (acknowledged as necessary for implementing a sound management program for that subpopulation), all members thereof must be readily identifiable at any stage of their development.

Assuming that any member of a salmon subpopulation can be recognized in the catch and that a very good picture of that subpopulation's age structure can be obtained (by means of the procedures outlined herein), its utility will be greatly restricted unless it can be complemented with "effective" effort data. Theoretically, computed "catch-effort" or "catch-intensity" ratios for each year-class comprising a subpopulation should constitute efficient measures of the relative abundance of each. Unfortunately, these cannot be materialized because of an inability to obtain, let alone define, fishing effort derivatives representing true indices of the probability a salmon will be encountered and will succumb to a unit of trolling gear. The only way we can measure the effort expended by the California salmon trolling fleet is by troller-days—but only by troller-days wherein a marketable yield is realized. Such data obviously have limited usefulness.

Finally, mention should be made of the representativeness of the sample observations intended to help predict year-to-year salmon yields. Obviously, any imposition as to size (length) of fish that can be legally retained renders the total catch unrepresentative of the population or stock. Since only two age groups, 3-year-olds and 4-year-olds (unpublished data; and Appendixes I, II, IX, and XI), consistently make up the bulk of California's king salmon landings, even if all biases were removed from catch sample data, using the relative strength of a year-class to project its future contribution to the fishery over such a short time span appears questionable. Involved is not only the rate at which any king salmon year-class is fished but the differential rates at which it continually excludes itself from the fishery to reproduce. To be of any practicable value, predictions (in relative terms) of future salmon yields based upon the classical process of following year-classes should, in the least, entail full knowledge of the pre-recruitment as well as the fully recruited phases of each subpopulation. At present, such information could only be secured through systematic offshore sampling employing research vessels.

A much more positive and realistic approach to the general problem of prediction appears to lie within the framework of what might be termed the "modern" concept of fishery management. Here all measurable manifestations of a population's dynamic nature (as influenced directly by the environment and fishing) are taken into account and collectively provide a basis for predicting, in absolute terms, that quantity of the end product that can be removed per unit time on a sustained basis, leaving the brood stock more or less static.

Because of the shortcomings just pointed out, I have not attempted to integrate information derived herein and available elsewhere to analyze any aspect of king salmon population dynamics. My main concern was to describe and apply a tool designed to extract and evaluate certain information (required by the theory underlying the above-mentioned concept) which together with its complements would theoretically comprise the framework of a sound salmon management program.

2. DRAWING THE SAMPLE

Sampling for fishery landing attributes is quite unlike randomly selecting variegated balls from an urn. Accordingly, it was with some reservation that I applied contemporary sampling theory in an attempt to assess the estimates derived as well as the estimation procedures. Implied nonetheless is the blanket assumption that random selection prevailed at all stages of the sampling designs employed; selection and estimation biases are presumed minimal throughout.

Several features of the California oceanic salmon fisheries had to be considered in implementing a catch sampling program.

2.1. DISTRIBUTION OF LANDINGS AND SAMPLING EFFORT

King and silver salmon, *O. kisutch* (Walbaum), are commonly sought by sport and commercial fishermen in California's coastal waters from Avila to the Oregon border. The area of potential exploitation embraces some 12,000 square nautical miles. Generally speaking, however, spatial distribution of fishing effort thereon is nowhere near uniform, both sport and commercial fleets largely confining their activities to areas immediately adjacent to their home ports. Fortuitously, the north-central California coast has relatively few, widely scattered ports out of which operate varying proportions of the salmon fleet. Thus while monthly or seasonal landings collectively may not be representative of all portions of the coastal salmon populations (or subpopulations), gaining access to all landings to sample them is no problem.

All salmon are taken from coastal waters by trolling gear whether it is powered (commercial) or manipulated by rod and reel (sport). Individual fishing trips infrequently exceed 12 hours and, consequently, are made in the vicinity of a vessel's home port. Some commercial trollers fitted with adequate refrigeration may make extended trips. Pro-rating their catches with respect to origin and time of capture often is a problem.

Principal ports serving both sport and commercial salmon fishing fleets are Avila, Morro Bay, Monterey, Moss Landing, Santa Cruz,

Princeton, San Francisco, Point Reyes, Bodega Bay, Fort Bragg, Eureka, and Crescent City. To insure as complete a coverage of the total annual landings as was practicable, sampling personnel were strategically placed at from three to five of the larger ports (Fry and Hughes, 1952). Permanent (seasonal) samplers were initially (1952) located in the Monterey-Santa Cruz; San Francisco-Point Reyes-Bodega Bay; Fort Bragg; and Eureka-Crescent City areas. In later years, additional personnel frequently assisted in the Monterey vicinity, especially in the outlying Morro Bay-Avila area.

Temporal distribution of salmon fishing effort is, of course, dictated to a large extent by weather. Other factors such as nonavailability of marketable fish and prolonged price disputes often may play major roles in determining when and how much effort is expended. Notwithstanding inconveniences due to uncontrollable factors and the situations they created, time variability in a total annual landing's structure was accounted for by sampling all ports systematically throughout every season. Landings at major commercial ports such as Point Reyes and Eureka usually were sampled five days a week while those sporadically available at smaller ports typically were sampled less often.

2.2. ACCESS TO INDIVIDUAL CATCHES: CHOICE OF VESSEL

Having accounted for as much space-time variability as was practicable by distributing sampling effort, the problem then became one of sampling specific catches. Two more sources of variability had to be dealt with here, *viz.*, that between vessels and that within. Within-vessel variability, where it pertained to sampling for marked fish, length frequencies, and average weight was eliminated by examining and measuring all salmon taken by individual trollers (commercial) or charter boats and skiffs (sport). When particularly numerous, vessels themselves were selected in a random-systematic manner; *i.e.*, the first vessel was taken objectively with a constant multiple of the remainder rounding out a port-day sample. When vessels were minimal, all catches usually entered the sample.

Salmon were sampled for age, when possible, at the rate of 50 per week. At the risk of not being able to fill a "quota" by systematically sampling throughout a weekly period, these were taken at the first opportunity while sampling for other attributes. Thus, while some weekly age samples represented a series of small catches made throughout a week, many did not, often having been taken during the first one or two days from comparatively large catches of a few vessels.

Although perhaps wanting in a few respects, the foregoing scheme was believed the best that could be implemented under prevailing conditions. It may have been desirable to place more emphasis on minimizing any between-vessel error but because of manpower limitations, this frequently would have necessitated reducing the number of salmon sampled per boat, risking a reduction in weekly (and monthly) sample size. Due to uncontrollable factors which normally govern the frequency and magnitude of salmon catches, age sampling had to be carried out on a day-to-day hit-or-miss basis, thereby obviating any attempt to execute a fully systematic scheme.

2.3. CONSOLIDATING RAW DATA

A single criterion usually constitutes the basis for deciding how timeseries data should be consolidated, namely, the rate of change of those parameters being estimated. Frequently, data secured for each time increment in a sampling design are insufficient for precisely estimating an attribute's status during each increment. When those for successive increments are combined, they represent larger increments in which it is presumed the rate of change of the parameter being estimated is either very slight or the parameter itself is effectively constant.

In the present case, estimates of absolute age composition were desired for monthly king salmon landings at each port-area and for over-all annual coastwise landings. Although age samples were obtained at some ports on a daily basis and at practically all on at least a weekly basis, their sizes were often rather small and generally fluctuated widely from period to period and from port to port. This suggested they should be consolidated into larger time units and stratified geographically by general port-areas rather than by individual ports. The idea, of course, was to provide data as amenable to analysis as practicable yet arranged so underlying assumptions as to the spatial and temporal nature of the parameters of interest were not greatly disturbed.

Accordingly, all age samples and resulting data were realigned and grouped on a port-area-month basis. Each such grouping is referred to hereinafter as a port-month cluster. This procedure generally yielded data having increased flexibility computationally but may have resulted in minor losses of information concerning daily or weekly variation patterns in the age structure of landings accumulated at specific coastal points.

The general port-areas recognized are (1) Monterey, which includes all data emanating from sport and commercial landings made between Avila and Santa Cruz; (2) San Francisco, accommodating all sampling data from landings between Half Moon Bay (Princeton) and Bodega Bay; (3) Fort Bragg, with data from the Albion-Fort Bragg-Shelter Cove area; (4) Eureka-Trinidad and (5) Crescent City, to which were assigned all data from catches made in the vicinity of and landed at each. Information from sport catches made in waters adjoining either of the latter two areas was grouped under the heading Crescent City-Eureka. Typically for any given sampling year (1952–57), king salmon age data were made available for 25 commercial and 36 to 40 sport, port-month clusters. During all years covered in this study, the commercial salmon season extended for five months (May-September) and the sport season for nine.

2.4. AGE DETERMINATION

Mid-lateral salmon scales selected for aging were cleaned and mounted dry between microscope slides. During the period 1952–57, only one scale per fish was used. It was selected under low magnification from several available to lessen the likelihood of mounting regenerated material. This procedure minimized processing materials and labor and greatly facilitated both scale examination and storage since five fish could be represented on a single mount. Intensive examination

of scales secured during 1952–56 showed that a specimen's age could not always be determined satisfactorily from a single scale so an additional scale was mounted beginning with mid-season 1957.

All scales were read independently by at least two persons and disagreements were resolved by a third. This system proved quite workable, and, considering the large number of scales examined (15 to 18 thousand), disagreements occurred surprisingly seldom (less than two percent). Most of these were eventually resolved.

Assuming correctness in interpreting scale sculpture, ages were recorded on the basis of the number of effective annuli observed. A 3-year-old, for example, would refer to a salmon captured during its third year of life whose scales displayed two effective annuli.

Many scales exhibiting "ocean-type" nuclei (Gilbert, 1914) possessed initial winter marks having obviously been formed a comparatively short time after eclosion. Such scales were construed to represent progeny of early-run subpopulations which had hatched during or just prior to the period of annulus formation. To maintain correspondence with all reproduction during a given brood year, these true winter marks (ineffective annuli) had to be overlooked when assigning ages.

While such characteristics might constitute clues that would assist in separating oceanic salmon stocks into their component subpopulations, the spatial and temporal relationships involved are not clearly understood let alone defined. The same could be said for salmon whose scales indicate measurable first year stream growth. Not being in a position to evaluate or utilize these characteristics properly, we must be satisfied with our ability to determine with reasonable accuracy any specimen's age, and with age structure estimates based upon systematic but composite samples of total landings, each member being standardized according to its brood year.

2.5. DEFINITION OF THE PARENT POPULATION

Acquiring the sample and delineating its component's attributes are followed by sample estimation of those population parameters representing the attributes defined. In this case, a parent population comprised the totality of fishery landings (per unit time) from which the required samples were drawn. Whether these landings truly reflected the structure of their universe (*i.e.*, the fish population(s) available for exploiting) depended upon the distribution of fishing effort relative to that of the universe, and any biases introduced during exploitation.

Certain restrictions imposed on the California king salmon fisheries significantly curtailed the amount of information about their populations off California during 1952–1957 that we could obtain from the statistics collected. All commercial landings were composed of king salmon generally having over-all lengths of not less than 26 inches while all taken by the sport fleet had to be at least 22 inches long. Prior to 1955, sport fishermen could have a limited number of smaller king salmon.

Gear selectivity and varying availability-vulnerability of population components were sources of possible bias which, incapable of being measured, may have further delimited the parent populations (landings).

3. AGE STRUCTURE ESTIMATES: PORT-MONTH LANDINGS

The most widely employed means of describing catch age structure entails computing percentage contribution of age groups represented in simple random catch samples. Patently, such information by itself is of relative value and has no utility in eliciting further details concerning, for example, mortality parameters. On the other hand, when projected to the total landings and effort generating them, it can help estimate absolute numbers or mass contributed by each exploited age group, these data in turn providing estimates of each age group's (relative) abundance in the fishery, and rather good measures of the total mortality suffered by fished portions of the population under scrutiny. Admittedly, total yield and effort data are not always available and relative age structure measures must therefore suffice—though their use is generally quite limited.

Deserving consideration, however, is the fact that simple direct estimates of the proportions of each age group comprising a landing, even when supplemented with total catch and effort data, may not always constitute the most efficient estimates. One of the first to recognize this was Fridriksson (1934), who, working with cod, proposed that information concerning a closely related and easily measured attribute such as length be used to augment oft-meager age data. Since costs of extracting age measurements (counting otolith annuli) ran quite high and consequently curtailed the number of age samples that could be processed (in turn minimizing the precision of resulting estimates), he reasoned and subsequently established that for a given number of age determinations, more information generally accrued if the age composition estimates for a given landing were composites of those from small age subsamples, each selected from lots of fish having common length. Conversely, for the same level of precision that would have been attained had length stratification not been introduced, fewer age determinations would have been required. Not fully realizing the implications involved in attempting to apply a stratified estimator to commercial fishery statistics, Fridriksson merely had as his objective reducing the size of age samples needed to derive relative (percentage) age composition estimates.

With some modification (but still no assessment), Hodgson (1939) extended the procedure to secure improved estimates of absolute numbers of fish in each year-class entering East-Anglian herring landings. The method was assessed in detail by Gulland (1955) who employed it to derive mortality estimates of certain North Sea bottomfishes. Described briefly, it involves acquiring fairly large samples which are ultimately (if not already) sorted into length categories or strata, these in turn yielding smaller samples for age determination. The subsample proportions of each age group are then applied to the number of the estimated total landed that falls into the corresponding length stratum. Numbers in each age group are finally summed for all length strata to provide estimates of the total number of each age landed during a given time increment. Gulland (1955) showed (using a specific set of data) that this "age-length key" approach resulted in better estimates

of absolute age composition than could have been attained using the same number of direct-sample age determinations. The frequent exceptions that have occurred cast doubt on his age-length key's general applicability and warrant investigating its limitations further. Moreover, executing certain field sampling operations involved in its construction leave much to be desired from a practical standpoint. In fact, except in situations where price differentials may demand that wholesalers sort landings in accordance with some size criterion (whereupon it is assumed that all size lots would be equally available to sampling personnel), the method has little adaptability.

A similar technique, differing significantly in the procedure of selecting the age subsample, was pursued by Tanaka (1953) when estimating absolute age structure of Japanese sardine landings. It is often referred to as the method of double or two-phase sampling for stratification. It differs from stratified subsampling in that a subsample for age is randomly taken from a larger unsorted length sample; that is, stratification occurs after subsampling for age. Since it seemed best suited to the conditions encountered, this method was used to ascertain the absolute age structure of port-month salmon landings.

3.1. TWO-PHASE SAMPLING FOR STRATIFICATION

In delineating the absolute contribution of each age group to a port-month's king salmon landings, the sample was defined as the composite of all specimens measured for length during the period and at the location of interest—subject to certain restrictions. Individual fish constituted the sample elements.

From salmon catches accumulating at a given point, the procedure was to measure fork lengths of as many specimens as practicable, assuming representative selection at all times. Lesser numbers were randomly selected from among those measured for length and scales were taken from them for effecting age determinations. Both first- and second-stage samples were ultimately sorted into 2.5-cm length strata and the second-stage samples were further separated into their component age groups.

Introducing some notation, we have then an initial sample of n fish each measured for length, from which a subsample of m fish is to be aged. Both are subdivided into i length strata, n_i and m_i representing, respectively, the numbers sampled for length and age falling into the i th length stratum. The latter are further separated on the basis of age representation into m_{ij} , the number of m_i falling into the j th age group; then

$$\sum_j m_{ij} = m_i, \sum_i m_i = m, \text{ and } \sum_i n_i = n.$$

EQUATION

Treating each subsampled specimen in the i th length stratum as a binomial or Poisson variate, x_{ijk} , taking the values 1 or 0 depending upon whether its scales possess $(j - 1)$ effective annuli

$$(\sum_k x_{ijk} = m_{ij})$$

EQUATION

, an unbiased estimate of the proportion of age j salmon (P_j) among the total landed is

$$p_{j(st)} = \sum_{i=1}^L \frac{n_i}{n} \cdot \frac{1}{m_i} \sum_{k=1}^{m_{ij}} x_{ijk} = \sum_{i=1}^L p_i \bar{x}_{ij} = \sum_{i=1}^L p_i p_{ij} \quad (1)$$

($j = 1, 2, 3, \dots$ years)

EQUATION

where L denotes the number of length strata constructed and k identifies an individual x_{ijk} .

Regarding p_i (an unbiased estimate of $P_i = N_i/N$) as a multinomial variate, $p_{j(st)}$ variance is, assuming an infinite population, given by

$$\text{var}(p_{j(st)}) = \frac{n}{n-1} \sum_{i=1}^L \left[\left(p_i^2 - \frac{p_i}{n} \right) \frac{p_{ij} q_{ij}}{m_i - 1} + \frac{p_i (p_{ij} - p_{j(st)})^2}{n} \right]$$

(see Cochran 1953, p. 274)

However, a more usable form (approximately equivalent) is

$$\text{var}(p_{j(st)}) = \sum_{i=1}^L \left[p_i^2 \frac{p_{ij} q_{ij}}{m_i} + \frac{p_i (p_{ij} - p_{j(st)})^2}{n} \right] \quad (2)$$

EQUATION

where $q_i = (1 - p_i)$ and $q_{ij} = (1 - p_{ij})$. In either form, the first term on the right side represents that portion of the total variance due to variation within length strata, the second that due to variation between strata.

A point estimate of the total age j salmon in a port-month's landings is, simply,

$$\hat{N}_{j(st)} = N_{at} p_{j(st)}$$

with variance

$$\text{var}(\hat{N}_{j(st)}) = N_{at}^2 \text{var}(p_{j(st)})$$

where N_{at} is an independent estimate of the number of fish landed at the a th port during the t th month. If desired, an approximate interval estimate can be constructed for $\hat{N}_{j(st)}$. Thus,

$$\hat{N}_{j(st)} = t_\alpha \sqrt{\text{var}(\hat{N}_{j(st)})}$$

EQUATION

where t is a normal deviate corresponding to an appropriate (α) confidence probability.

To apply the estimation methods described, two sets of port-month data (Appendixes I and II) were selected from our 1952–1957 data. Because of space limitations, those for all port-months sampled could not be included.

Besides presenting initial sampling data, Appendixes I and II list those obtained at intermediate stages in the process of formulating estimates of $p_{j(st)}$ for each age group represented. Appendixes III, IV, and V (Parts i), respectively, contain the results of intermediate computations leading to estimates of $\text{var}(p_{j(st)})$ for each $p_{j(st)}$ calculated from the data in Appendix II. Table 1 summarizes the results of all computations to which the selected data were subjected, giving also estimates of P_j corresponding to those obtained but derived via the simple random or direct system rather than a stratified estimator. The

simple random estimate, again assuming an infinite population, is denoted by

$$p_{j(ran)} = \frac{\sum_{i=1}^L m_{ij}}{m} = \frac{m_{.j}}{m}$$

with variance (assuming the m_i 's are proportionally allocated)

$$\text{var}(p_{j(ran)}) = \frac{p_{j(ran)} q_{j(ran)}}{m}$$

EQUATION

where $q_{j(ran)} = (1 - p_{j(ran)})$. $[N]_{j(ran)}$ and its variance are subsequently obtained in the same manner as $[N]_{j(st)}$ and $\text{var}([N]_{j(st)})$.

The stratified estimator's variance is nearly always much reduced below that of the simple random estimator, indicating a greater degree of efficiency. Comparing the coefficients of variation of the estimated number in each age group,

$$\text{CV}(\hat{N}_j) = \sqrt{\frac{\text{var}(\hat{N}_j)}{\hat{N}_j^2}}$$

EQUATION

suggested, however, that the over-all improvement using the stratified

TABLE 1

Comparison of Age Structure Estimators and Their Variances Derived via Stratified and Simple Random Sampling Techniques. Raw Data Are in Appendixes I and II. Stratification Is by 2.5-cm Intervals.

Method	Two-phase sampling for stratification			Simple random sampling		
Age groups	2	3	4	2	3	4
From data in Appendix I:						
p_j	0.02081	0.96393	0.01536	0.04762	0.92857	0.02381
$\hat{N}_j$	177	8193	130	405	7893	202
$\text{var}(p_j)$	0.000086	0.000251	0.000173	0.000540	0.000790	0.000277
$\text{var}(\hat{N}_j)$	6214	18135	12499	39015	57078	20013
$\text{CV}(\hat{N}_j)$	0.445	0.016	0.860	0.488	0.030	0.700
From data in Appendix II:						
p_j	0.04221	0.62315	0.33464	0.08462	0.50769	0.40769
$\hat{N}_j$	450	6643	3567	902	5412	4346
$\text{var}(p_j)$	0.000109	0.001140	0.001045	0.000596	0.001923	0.001857
$\text{var}(\hat{N}_j)$	12386	129545	118750	67727	218521	211021
$\text{CV}(\hat{N}_j)$	0.247	0.054	0.097	0.288	0.086	0.106

TABLE 1

Comparison of Age Structure Estimators and Their Variances Derived via Stratified and Simple Random Sampling Techniques. Raw Data Are in Appendixes I and II. Stratification Is by 2.5-cm Intervals

estimator was frequently of such small magnitude as to question its general utility.

Implementation of the stratified (two-phase) design is advantageous only when gains more than compensate for the loss in precision of the principal attribute's (age) estimate. Such loss is brought about by a decrease in age sample size when part of the total sampling effort is used to measure a secondary attribute (length). In many, if not most,

comparable situations, any sampling technique is restricted by the funds or manpower available. Thus to assess truly the two-phase stratified estimator one must bear in mind not only the quantity of available resources, but more important, the costs of measuring the primary attribute compared to those associated with measuring one related to it but of lesser concern. Intrinsically involved here are the size of the secondary sample where a fixed length-sample size is established, and the manner in which the first-, and hence, second-stage samples are stratified. Understanding such factors is prerequisite to determining for a fixed cost what first- to second-stage sample ratio, with varying stratification, provides a "best" estimate of the principal attribute's absolute status.

Over and above any theoretical advantages ascribed to stratified sampling, Gulland (1955) has pointed out one other with practical implications where age structure estimates are concerned. Significantly, a design incorporating stratification yields estimates of numbers in adjacent age groups which tend to be negatively correlated, a desirable condition not attained when employing the simple random sampling approach. Using simple random sampling, a positive correlation could prevail between estimates of numbers of fish falling into adjacent age groups whenever samples destined for aging are biased by length, by weight or by both. Thus if large fish were over-represented in a particular sample, the numbers in each of the older age groups would be over-estimated. Stratifying by length tends to obviate untoward effects of age sampling bias.

3.2. AGE SAMPLE SIZE FOR A FIXED LENGTH SAMPLE

Assuming comparatively large numbers of salmon can be measured for length at relatively little expense, it is now desirable to determine approximately how many should be sampled for age to provide reasonably precise estimates of P_j . The problem, for a fixed n and prescribed level of precision, is to find the minimum m for each age group, letting the largest serve as an estimate of the required overall m .

If m is randomly taken from n and the expected value of p_j is P_j , then (for any age group) the size of each m_i will be approximately proportional to p_i , that is,

$$\frac{m_i}{m} \cong \frac{n_i}{n} \cong \frac{N_i}{N}$$

$$m_i \cong m p_i$$

EQUATION

Substituting this value in (2) and assuming an infinite population, gives, for salmon in the j th age group,

$$\text{var}(p_{j(st)})_{pro} = \frac{\sum_{i=1}^L p_i p_{ij} q_{ij}}{m} + \frac{\sum_{i=1}^L p_i (p_{ij} - p_{j(st)})^2}{n} \quad (3)$$

EQUATION

which, solving for m , leads to

$$\hat{m}_{pro} \cong \frac{n \left(\sum_{i=1}^L p_i p_{ij} q_{ij} \right)}{nV - \left[\sum_{i=1}^L p_i (p_{ij} - p_{j(st)})^2 \right]} \quad (4)$$

EQUATION

the formula for age sample size with a fixed length sample where V is the desired variance of $p_{j(st)}$ and the subscript *pro* indicates that estimating a proportion is involved. When interest lies mainly in estimating the number in a given landing which fall into the j th age group, i.e., $N_{at} P_j$, this modification can be employed:

$$\hat{m}_{tot} \cong \frac{N_{at}^2 n \left(\sum_{i=1}^L p_i p_{ij} q_{ij} \right)}{nV - N_{at}^2 \left[\sum_{i=1}^L p_i (p_{ij} - p_{j(st)})^2 \right]} \quad (5)$$

EQUATION

where V is the desired variance of $[N]_{j(st)}$ and the subscript *tot* indicates that estimating a total is involved. The maximum m_{pro} (or m_{tot}) determined for a particular set of data then represents an approximation to m_{pro} (or m_{tot}), the over-all age sample size required for estimating all P_j 's at, or greater than, some predesignated level of precision. For future sampling at corresponding port-months whose landings could be expected to exhibit the same age structure, m_{pro} constitutes a reasonable guide to the required magnitude of m .

To apply these formulas by adducing the extent to which additional age sampling was indicated for a particular situation, and, perhaps, suggest how large the projected samples should be in similar situations, data and subsequent estimates in Appendixes II to V are used. Thus to estimate the proportion of 3-year-old king salmon (± 15 percent of the true value) in the Crescent City-Eureka July 1955 sport landings, we have, assuming a length sample size of 2,875.

$$\hat{m}_{pro} \cong \frac{2875(0.11376)}{2875[(0.15)(0.62315)]^2 - (0.12106)} \cong 14$$

EQUATION

If the level of precision for $p_{3(st)}$ were increased to ± 10 or 5 percent, m_{pro} would increase to 30 and 123 fish respectively. To estimate with the same precision the number of 3-year-old king salmon in the same landings, we have by means of (5) and using the same $n, m_{pro} = 9$ specimens. The same procedure would be followed for all other age groups and the largest m_{pro} or m_{tot} calculated would determine the desired over-all age sample size.

Table 2 gives approximate values of m_{pro} for estimating P_j at two levels of precision with varying length sample sizes. It will be noted that little would be gained by merely increasing n provided each age group is reasonably represented, $p_{j(st)}$, say, being $[>] 0.15$. The size of m depends critically either upon how many age groups occur in a given set of landings or upon the extent to which each overlaps the others.

Obviously, the more age groups represented, the smaller, on the average, will be $p_{j(st)}$, and the larger the over-all age sample size needed for estimating all P_j 's at a given level of precision. This points up the difficulty in predicting m 's size; hence without *a priori* information, the usual alternative is to take the maximum age sample permitted by the available resource, acknowledging in advance that the status of certain age groups will be determined very precisely while others will be determined with an untenably low degree of reliability (cf. Table 1).

Though $p_{j(st)}$ can be expected to vary—often quite widely—in space and time, it is nonetheless suggested that if not too stringent precision requirements are imposed in P_j 's estimation (i.e., with not more than ± 20 percent allowable error), an m of scales from 200 king salmon per port-month should suffice in practically all cases. For other species exhibiting greater age ranges, m would have to be increased accordingly or the precision requirements for $p_{j(st)}$ relaxed.

TABLE 2
Age Sample Sizes With Fixed Length Samples for Estimating $p_{j(st)}$
at Two Levels of Precision
(Predicted Variances Were Derived From Data in Appendix II)

Length Sample Size (n)	Desired Coefficient of Variation (%)	Age Sample Size ($\hat{m}$)			
		Two-year olds only	Three-year olds only	Four-year olds only	All ages
1000.....	15	1052*	14	44	1052*
	20	230	8	25	230
2875.....	15	304	14	43	304
	20	149	8	24	149

* Would have to age all specimens in length sample.

TABLE 2
Age Sample Sizes With Fixed Length Samples for Estimating $p_{j(st)}$ at Two Levels of Precision

3.3. CHOICE OF STRATIFICATION

Since the age sample, if randomly taken from the length sample, can be expected to distribute itself among length strata in proportion to the latter's size, the number of strata, L , employed can be as large as desired so long as each contains at least one second-phase element. As is usually the case, relatively little is known of the variability of a population's age structure with respect to that of its length. Thus when implementing a scheme such as described, the size of L must be chosen arbitrarily at the outset. The tendency often is to over-stratify, which in turn perhaps leads to unnecessarily high costs of computing the desired estimates and measures of their precision. But while stratification with proportional allocation of sample elements always results in estimates having greater precision than those obtained by simple random sampling, the rate of precision gain with successive stratum division may decrease markedly after a certain point is reached, all gains thereafter being increasingly trifling.

An answer to the question of how many strata to employ in a two-phase sampling program defies generalization. Preliminarily, any decision

as to L 's size would have to be based upon some compromise made in accordance with the expected age structure of the population under scrutiny. And although an optimum or best stratification, could be determined *a posteriori* for a particular situation, it would have little value from the standpoint of general projected utility. Rather, assuming that a large length sample can be secured, a more practical approach would be to weigh the size of over-all age sample commensurate with available resources, against some predesignated minimum m_i which, collectively over i strata, could be expected to yield reasonably precise estimates of the absolute magnitude of most age groups represented.

One possible approach, assuming m_i/m has expectation N_i/N , is

$$\frac{m_i}{m} = \frac{n_i}{n} = \frac{N_i}{N}$$

and

$$\frac{\bar{m}}{m} = \frac{\bar{n}}{n} = \frac{\bar{N}}{N} = \frac{\bar{N}}{\bar{N}L} = \frac{1}{L}$$

EQUATION

where $[m]$ is some specified minimum average m_i , and $[n]$ and $[N]$ are the sample and population mean stratum sizes respectively. Thus, if n is large, relative to m ,

$$L = \frac{m}{\bar{m}}$$

EQUATION

represents a criterion for establishing the stratification to be employed where $[m]$, the estimated over-all age sample size determined earlier, is substituted for m .

The size of $[m]$ can be decided upon once the desired limits of error in estimating each P_{ij} are set. If a coefficient of variation of 0.20 is stipulated, an m_i of 25 to 30 would yield estimates of any P_{ij} being in error by more than ± 20 percent of the actual value only once in 20 times. Being conservative and letting $[m] = 30$ specimens, we obtain where 200 specimens can be aged,

$$L = \frac{200}{30} = 6 + (\text{rounding upward to } 7)$$

EQUATION

as a rough approximation to the number of length strata needed to obtain estimates having the desired precision. Note that $[m]$ is an average and most strata will be allocated this number of aged specimens at least, but several will ordinarily be under-represented. This probably would not constitute a serious detriment in the case of California king salmon because individuals at the extremities of observed length distributions rarely fall into more than one age group. Such would not be true with species represented by a greater number of age groups or in situations where a particular age group is markedly dominant. Regardless of the sampling technique, estimates of the total numbers in the infrequently observed age groups occupying the upper extreme of any exploited population's length distribution, will almost never be as precise

as might be desired since the attendant sample size will normally be nowhere near that required to attain such precision.

To determine the effects of reducing the number of length strata (or of reducing strata widths), the length and age distributions given in Appendix II on the basis of 2.5-cm length strata were realigned into 10-cm strata (Appendix VI). Alternative estimates of P_j and its variance, using formulas (1) and (2), are given along with intermediate calculations in Appendix VII. Comparing these with those secured by two-phase sampling for 2.5-cm length strata and simple random sampling (Appendix VIII) suggests that very little would have been lost by 10-cm stratifying. This would be expected, since not only had the initial double-sampling process involving 2.5-cm strata already exhibited little superiority over single-phase sampling (Table 1), but the progressive diminution of L itself leads ultimately to the latter situation. Discounting the deviation reflected by the four-year-olds (which can be attributed to the vagaries of reallocating age data among the larger length strata), note that $\text{var } [N]_j$ calculated with 10-cm strata lies between the values for $\text{var } [N]_j$ computed by double sampling with 2.5-cm strata and simple random or single-phase sampling (Appendix VIII).

3.4. Two-phase Versus Simple Random Sampling with Varying Age Sample Costs

As previously indicated, a sampling design can best be appreciated only after taking into account unit costs as opposed to fixed resources and comparing results with those of some predesignated standard technique. We have considered the special case involving two-phase sampling for stratification where the first- to second-stage sampling cost ratio was assumed to be unity, *i.e.*, length and age measurements were equally costly. In the case of salmon, length stratification improved the required estimates only slightly. In fact, if the effort and expense required to make the length measurements noted (Appendixes I and II) had not served other purposes, they would have been virtually wasted.

It is desirable now to compare two-phase and simple random sampling procedures where, in the former, the second-phase sample is always more costly than the first. With either approach, we assume resources can provide an age sample of size m .

As pointed out by Cochran (1953), the values of m_i and n that yield a minimum variance in a particular situation are unavoidably intractable. In approximating a solution, it is suggested that m_i be proportional to the weighted within-stratum standard deviation, *i.e.*, to the root of the numerator of the first term on the right side of (2). This gives

$$m_i = \frac{m p_i \sqrt{p_{ij} q_{ij}}}{\sum_{i=1}^L p_i \sqrt{p_{ij} q_{ij}}}$$

EQUATION

whose substitution in (2) yields, approximately,

$$\text{var}(p_j)_{opt} = \frac{\left(\sum_{i=1}^L p_i \sqrt{p_{ij} q_{ij}} \right)^2}{m} + \frac{\sum_{i=1}^L p_i (p_{ij} - p_{j(st)})^2}{n} \quad (6)$$

EQUATION

the optimum variance of p_j . This can be simplified notationally to

$$\text{var}(p_j)_{opt} = \frac{V_m}{m} + \frac{V_n}{n} \quad (7)$$

EQUATION

where, assuming optimum sampling, V_m is the variance within length strata and V_n is the variance between. From a practical standpoint allocating a second-phase sample under conditions such as those met with salmon would be highly unrealistic. It can be shown, however, that in two-phase sampling for stratification where the second-phase sample is randomly taken from the first, proportional and optimum allocation are theoretically the same (Deming 1950, p. 240). Hence (3) could have been used instead of (6) though to emphasize that maximum efficiency in stratified sampling is attained when the sample is allocated optimally among strata, formula (7) shall be invoked.

It is appropriate now to introduce a simple cost function of the form

$$\text{Cost} = C = mc_m + nc_n \quad (8)$$

EQUATION

which represents the total cost of securing the age (m) and length (n) samples where the cost of m (c_m) is appreciably greater than that of n (c_n). Under conditions of optimum sampling, the variance of p_j , (7), is now minimized by the appropriate choice of m and n for a given total cost, C . Thus differentiation of

$$\text{var}(p_j)_{opt} + \lambda(mc_m + nc_n)$$

EQUATION

(where λ is a Lagrange multiplier) with respect to m and n leads to

$$\frac{m}{n} = \sqrt{\frac{V_m c_n}{V_n c_m}} \quad (9)$$

EQUATION

which, after appropriate manipulation and substitution in (7) (see Cochran 1953, p. 272), yields

$$\text{var}(p_j)_{opt} = \frac{(\sqrt{V_m c_m} + \sqrt{V_n c_n})^2}{C}$$

EQUATION

After several years of collecting and processing king salmon length and age data, it was determined that, on the average, the cost of eliciting a specimen's age was roughly ten times that of measuring its length. Using sampling data emanating from the July 1955 Crescent City-Eureka sport landings (Appendix II), V_m and V_n were calculated for the three age groups and are listed in Parts ii of Appendixes III, IV, and V, and in Appendix VII. Assuming that resources would permit aging no more than 130 specimens regardless of the sampling approach, their substitution in (10) [or in (8) and (9) jointly] gave predicted values for $\text{var}(p_j)_{opt}$ (Table 3, line 3).

For comparative purposes, the same procedure was employed to secure predicted values for $\text{var}(p_j)_{opt}$ where the cost of age samples and determinations was presumed to be twice and five times that of the corresponding length sample measurements (Table 3, lines 1 and 2). All predicted $\text{var}(p_j)_{opt}$ values were compared with appropriate predicted values for $\text{var}(p_j)_{ran}$, the variance resulting in the event a double or two-phase sampling scheme is not employed (i.e., $n = 0$).

Five salient pieces of information can be gleaned from Table 3: (1) Two-phase sampling for stratification becomes increasingly more efficient for estimating age as the cost of the age sample increases relative to the length sample. (2) For a given cost relationship, diminution of the number of length strata results in a marked increase in variance depending upon the sample age group distribution. (3) Assuming representativeness of these data, there seems to be little point in implementing a two-phase length-age sampling scheme unless the age sample is at least five times costlier than the length samples [cf. $\text{var}(p_j)_{opt}$ and $\text{var}(p_j)_{van}$]. (4) When determining age is much costlier than length (e.g., line 3), age sample sizes required of the two-phase scheme for estimating $j(st)$ (and $N_{at}p_j(st)$) with the same precision as attained by the simple random or direct approach, are always reduced (often greatly); conversely, at high second-phase sample costs and for a fixed second-phase sample size, greater precision can be expected to accrue from double sampling. (5) Very little added precision can be expected by increasing first-stage sample sizes (cf. n in Appendix II and Table 3).

4. ESTIMATING OVER-ALL AGE COMPOSITION

Considering the commercial catch only, over-all seasonal estimates of the absolute number of a given age landed could be obtained by merely summing the $N_{j(st)}$'s over all ports and months—provided sampling data are available for all port-month clusters. Since this is usually not the case, some alternative scheme is in order.

A possible estimator appropriate to cluster sampling wherein cluster size is variable is a ratio estimator using the stratified estimates $N_{j(st)}$ obtained for each port-month sampled. For the j th age group it is

$$N_s \hat{P}_{j(r)}$$

where

$$\hat{P}_{j(r)} = \frac{\sum_h^k \hat{N}_{j(st)}}{\sum_h^k N_{at}} \quad (h = 1, 2, 3, \dots k)$$

EQUATION

is a simple unweighted estimate of $P_{j(r)}$, N_s is the independently estimated over-all total number of fish landed during a season, and k is the

TABLE 3
Comparison of Variances of p_j Estimated by Two-phase Sampling at Two Levels of Stratification Presuming Optimum Allocation, and by Simple Random Sampling. Raw Data Are in Appendixes II and VI. Intra-strata Variances, V_n and V_m Are Derived in Appendixes III-V; (parts ii) and Appendix VII.
(See Text for Formulas)

Age group (j)	Two-year olds			Three-year olds			Four-year olds		
Cost function	$\text{var}(p)_{\text{opt}}$ 21 strata	6 strata	$\text{var}(p)_{\text{sr}}$	$\text{var}(p)_{\text{opt}}$ 21 strata	6 strata	$\text{var}(p)_{\text{sr}}$	$\text{var}(p)_{\text{opt}}$ 21 strata	6 strata	$\text{var}(p)_{\text{sr}}$
1. $C = 130 = m + 0.5n$	0.00018	0.00029		0.00205	0.00253		0.00184	0.00212	
m	21	40		69	80		66	73	
n	219	181		124	102		129	117	
2. $C = 130 = m + 0.2n$	0.00008	0.00015	0.00060 $CV(p) = 0.288$	0.00177	0.00187	0.00192 $CV(p) = 0.086$	0.00123	0.00149	0.00186 $CV(p) = 0.106$
m	30	54		83	93		81	87	
n	501	384		237	188		249	221	
3. $C = 130 = m + 0.1n$	0.00002	0.00007		0.00083	0.00125		0.00071	0.00093	
m	106	118		125	137		125	126	
n	2497	1203		507	362		546	455	

FISH BULLETIN NO. 120

TABLE 3
Comparison of Variances of p_j Estimated by Two-phase Sampling at Two Levels of Stratification Presuming Optimum Allocation, and by Simple Random Sampling. Raw Data Are in Appendixes II and VI. Intra-strata Variances, V_n and V_m Are Derived in Appendixes III-V; (parts ii) and Appendix VII

number of h port-month clusters sampled. Assuming an infinite population, its estimated variance (Cochran, 1953, p. 125) is approximately

$$\text{var}(\hat{P}_{j(r)}) = \frac{1}{k(k-1)\bar{N}^2} \left[\sum_h^k \hat{N}_{j(st)}^2 + \hat{P}_{j(r)}^2 \sum_h^k N_{at}^2 - 2\hat{P}_{j(r)} \sum_h^k \hat{N}_{j(st)} N_{at} \right]$$

EQUATION

where

$$\bar{N} = \frac{\sum_h^k N_{at}}{k}$$

EQUATION

is the average cluster size and the remaining notation is as defined earlier. It follows that the variance of $N_S [P]_{j(r)}$ is simply

$$\text{var}(N_S \hat{P}_{j(r)}) = N_S^2 \text{var}(\hat{P}_{j(r)})$$

EQUATION

Providing sample sizes are large enough so that the normal approximation applies, confidence limits for $N_S \hat{P}_{j(r)}$ can be calculated

$$N_S \hat{P}_{j(r)} \pm t_\alpha \sqrt{N_S^2 \text{var}(\hat{P}_{j(r)})}$$

EQUATION

where t is a normal deviate corresponding to an appropriate (α) confidence probability.

Using the stratified (two-phase sampling) age estimates calculated for individual clusters (Appendix IX) and treating them as weighted

TABLE 4
Age Composition of California's 1952 Commercial King Salmon Landings.
Basic Data Are Total Numbers Estimated by Two-phase Stratified
Sampling Within Each Port-month Cluster.
(See Appendixes IX and X)

Age (Years)	Estimated Proportion	Standard Error	Estimated Total	Interval Estimate*	
				for $N_S \hat{P}_{j(r)}$	
(j)	$\hat{P}_{j(r)}$	$\sqrt{\text{Var}(\hat{P}_{j(r)})}$	$N_S \hat{P}_{j(r)}$	Lower Limit	Upper Limit
2	0.0007	0.0005	330	0	810
3	0.6000	0.0570	284770	230660	338880
4	0.3681	0.0561	174710	121460	227960
5	0.0312	0.0068	14810	8360	21270
Total	1.0000		474620		

* 0.95 confidence probability level.

TABLE 4
Age Composition of California's 1952 Commercial King Salmon Landings. Basic Data Are Total Numbers Estimated by Two-phase Stratified Sampling Within Each Port-month Cluster

sample observations, $[P]_{j(r)}$'s for the four age groups in the 1952 commercial landings are given in the lower right hand corner of Appendix X. In Appendix X the row values opposite the ages correspond to the $[N]_{j(s)}$'s calculated as described earlier, and over-all $[P]_j$'s for ports ($[P]_{j(a)}$) and months ($[P]_{j(t)}$), computed from the $[N]_{j(st)}$'s summed over ports and months, are in the right and lower margins. These are given merely to indicate on a seasonal basis, relative distribution of the fished ages in the landings by port and month. Their reliability is questionable,

however, because of the small number of sampled clusters upon which they are based (*re* Appendix X, a minimum of one and a maximum of three clusters serve to compute the marginal $[P]_{j(r)}$'s). Results of applying the estimator of main concern, $[P]_{j(r)}$, are given in Table 4. Here, too, interval estimates of $N[P]_{j(r)}$ have questionable utility since the variance (and standard error) of the ratio of two random variables is biased for small samples.

A more useful and realistic means of estimating over-all catch age structure entails fully utilizing the sampling scheme possible with portmonth clusters. Providing reasonably adequate sampling obtained in all clusters, four possible estimators are available for each age landed. They are (for fish of any age j):

$$\hat{P}_{j(unw)} = \frac{\sum_a^u \sum_t^v m_{at} \frac{\hat{P}_{j(at)}}{m_{at}}}{\sum_a^u \sum_t^v m_{at}} = \frac{\sum_a^u \sum_t^v m_{at} \hat{P}_{j(at)}}{\sum_a^u \sum_t^v m_{at}} \quad (\text{unweighted}) \quad (11)$$

$$\hat{P}_{j(pts)} = \frac{\sum_a^u N_{a.} \frac{m_{a.} \hat{P}_{j(a.)}}{m_{a.}}}{\sum_a^u N_{a.}} = \frac{\sum_a^u N_{a.} \hat{P}_{j(a.)}}{\sum_a^u N_{a.}} \quad (\text{weighted by port-areas}) \quad (12)$$

$$\hat{P}_{j(mon)} = \frac{\sum_t^v N_{.t} \frac{m_{.t} \hat{P}_{j(.t)}}{m_{.t}}}{\sum_t^v N_{.t}} = \frac{\sum_t^v N_{.t} \hat{P}_{j(.t)}}{\sum_t^v N_{.t}} \quad (\text{weighted by months}) \quad (13)$$

$$\hat{P}_{j(a-t)} = \frac{\sum_a^u \sum_t^v N_{at} \frac{m_{at} \hat{P}_{j(at)}}{m_{at}}}{\sum_a^u \sum_t^v N_{at}} = \frac{\sum_a^u \sum_t^v N_{at} \hat{P}_{j(at)}}{\sum_a^u \sum_t^v N_{at}} \quad (\text{weighted by ports and months}) \quad (14)$$

$$(a = 1, 2, 3, \dots, u)$$

$$(t = 1, 2, 3, \dots, v)$$

EQUATION

TABLE 5
Estimated Proportions and Total Numbers of Age j King Salmon Entering California's 1955 Commercial Landings. For Comparison, Four Estimators Were Employed. Basic Data Are in Appendix XI.

Estimator	(11) Unweighted	(12) Weighted by ports	(13) Weighted by months	(14) Weighted by ports and months
$\begin{matrix} \text{Age} \\ (j) \end{matrix}$	$\sum_a \sum_t m_{at} \hat{P}_{j(at)} \hat{P}_{j(unw)}$	$\sum_a N_{a.} \hat{P}_{j(a.)} \hat{P}_{j(pw)}$	$\sum_t N_{.t} \hat{P}_{j(.t)} \hat{P}_{j(mon)}$	$\sum_a \sum_t N_{at} \hat{P}_{j(at)} \hat{P}_{j(a-t)}$
2	58 0.0142	7508 0.0101	9891 0.0133	5692 0.0076
3	2153 0.5261	386299 0.5191	394968 0.5307	391791 0.5265
4	1775 0.4338	328617 0.4416	319054 0.4287	323938 0.4353
5	106 0.0259	21754 0.0292	20265 0.0272	22757 0.0306

$$\sum_a \sum_t N_{at} = 744178^*$$

$$\sum_a \sum_t m_{at} = 4092$$

* Does not include $N_{a.}$ for Monterey: September (No sampling).

TABLE 5

Estimated Proportions and Total Numbers of Age j King Salmon Entering California's 1955 Commercial Landings. For Comparison, Four Estimators Were Employed. Basic Data Are in Appendix XI

Note that each estimator uses basic (untreated) cluster data and, except in the case of (11), the independently estimated catch totals. Estimated port-month totals secured by two-phase sampling were not considered.

To illustrate the role of the two-way cluster scheme in obtaining best point and interval estimates of over-all catch age structure, I used the 1955 commercial age sampling data. Appendix XI contains raw age data and estimated catch totals for each sampled cluster, as well as intermediate calculations needed with estimators (11), (12), and (13). Individual cluster $[P]_{j(at)}$'s used with estimator (14) are recorded in Appendix XII. Each of the four estimators compared in Table 5 gives approximately the same results. This is to be expected if there is little variability in the proportion of age j fish between ports and months. Estimator (14) should be the least biased since it takes into account the effects (if any) of both ports and months. It is with (14) that we shall be primarily concerned in the following discussion.

Although $N_s [P]_{j(at)}$ would be expected to yield the best point estimate of the total number of age j fish landed seasonally, a measure of its precision may not be too useful since the available variance estimator for a ratio $[P]$ is subject to bias in small samples. To secure an appropriate interval estimate (and hence measure of precision) of $N_s [P]_{j(at)}$ might call for constructing such an interval using the least biased variance estimator, viz., the conventional (but weighted) binomial estimator, and then adjusting it for cluster effects by employing an appropriate correction factor. Such an estimator would be appropriate in the unrealistic case where

$$\sum_a \sum_t m_{at}$$

EQUATION

is regarded as a simple random sample of m individual fish.

The correction factor used is simply the ratio

$$'G_j' = \frac{\sqrt{\text{var}(\hat{P}_{j(a-t)})}}{\sqrt{\text{var}(\hat{P}_{j(a-t)}\text{bin})}}$$

EQUATION

where the numerator and denominator are standardized with respect to a common level of P , a ratio representing a measure of the relative change in standard error caused by using the two-way cluster sampling scheme.

The basic variance of $P_{j(a-t)}$, $\text{var}(P_{j(a-t)})$, is, assuming an infinite population and regarding the individual $[P]_{j(at)}$'s as continuous variates, the mean square for port-month interaction derivable by ordinary analysis of variance techniques; i.e.,

Source of Variation	Degrees of Freedom	Mean Square	Expected Mean Square
Between ports	$(u - 1)$	MSP	$\text{var}(\hat{P}_{j(a-t)}) + v \text{var}(\hat{P}_{j(a..)})$
Between months	$(v - 1)$	MSM	$\text{var}(\hat{P}_{j(a-t)}) + u \text{var}(\hat{P}_{j(.t)})$
Interaction	$(u - 1)(v - 1)$	MSE	$\text{var}(\hat{P}_{j(a-t)})$
Total	$(uv - 1)$		

EQUATION

For any age group j , $\text{var} P_{j(a-t)}$ is due to the failure of differences between successive port and month $P_{j(at)}$ to follow a definite pattern from port to port and month to month (Appendix XII).

Since cluster size (m_{at}) varied considerably from one to the next, the interaction sum of squares could not be obtained in the conventional manner. It was necessary therefore to compute a weighted interaction sum of squares by first obtaining the expected P (or regression) value for each port-month cell. Assuming they were independent, the deviations of the expected (or regression) values from their respective observed values (Appendix XII) were used to calculate the approximate sum of squares, and thence an approximation to the interaction mean square, employing the relationship

$$\text{MSE} = \text{var}(\hat{P}_{(a-t)}) = \frac{\sum_a \sum_t m_{at}^2 [\hat{P}_{at} - E(P)]^2}{k(k-1)\bar{\bar{m}}^2}$$

EQUATION

where k is the number of clusters sampled and

$$\bar{\bar{m}} = \frac{\sum_a \sum_t m_{at}}{k}.$$

EQUATION

When all clusters are sampled, $k = uv$. Original and intermediate data used in computing this value for each age group encountered in the 1955 landings are in Appendixes XI, XII, and XIII. In the foregoing procedure, data in the two-way table must be complete or nearly so. Values for the data missing from the Monterey-September cluster (see Appendixes XI and XII) were obtained using the method described by Anderson and Bancroft (1952, p. 246) where the estimated variates are determined such that the resulting interaction sum of squares will be minimized. This can be seen in Appendix XII (Monterey-September) where, for all practical purposes, the substituted $[P_{at}]$'s are their own expected values.

Regarding $\text{var}([P_{(a-t)}])$ as standardized with respect to the over-all mean $[P]$ for any corresponding age group j , the binomial variance for the same level of $[P]$, given by

$$\text{var}(\bar{P})_{bin} = \frac{\bar{P}(1 - \bar{P})}{k\bar{m}_{wd}}$$

where

$$\bar{P} = \frac{\sum_a^u \sum_t^v \hat{P}_{at}}{uv} = \frac{\hat{P}_{..}}{uv}$$

EQUATION

was then estimated for each age group. Its square root served as the denominator in the ratio 'G' (Appendix VIII). An approximation to the binomial variance of $[P]_{(a-t)}$ was made using the weighted estimator

$$\text{var}(\hat{P}_{(a-t)})_{bin} = \frac{\hat{P}_{(a-t)}(1 - \hat{P}_{(a-t)})}{k\bar{m}_{wd}}$$

where

$$\bar{m}_{wd} = \frac{\sum_a^u \sum_t^v N_{at} m_{at}}{\sum_a^u \sum_t^v N_{at}}$$

EQUATION

is the weighted mean cluster size. It follows that an approximation to the least biased variance estimator of the total in any age group j , i.e., $N_S [P]_{j(a-t)}$, is $N_S \text{var}([P]_{j(a-t)})$ with basic interval estimate for $N_S [P]_{j(a-t)}$ given by

$$N_S \hat{P}_{j(a-t)} \pm t_\alpha \sqrt{\text{var}(N_S \hat{P}_{j(a-t)})_{bin}}$$

EQUATION

The desired interval estimate which takes into account the effects of clusters is then simply

$$N_S \hat{P}_{j(a-t)} \pm t_\alpha 'G_j' \sqrt{\text{var}(N_S \hat{P}_{j(a-t)})_{bin}}$$

EQUATION

Point and corrected interval estimates of the numbers of king salmon in each of four age groups entering the 1955 commercial catch are given in Table 6.

Although somewhat complex computationally, I feel this system provides a convenient means for obtaining best estimates (in absolute terms) of the characters in question, besides permitting some generalizing as to the average size of cluster to employ in future censuses. If it is determined that the required minimum cluster size would be too unwieldy, it might be well to examine critically the between ports variability as regards the proportions of each age group entering the catch (Appendix XI) and, in situations where two-phase (stratified) sampling is used to estimate individual cluster age composition, the age apportionment within corresponding length classes. If over a period of years this variability is generally small (as in the case of certain ages represented in the 1952 and 1955 landings), a reallocation of age sampling effort would be in order. By eliminating some ports and increasing age sample sizes (i.e., cluster sizes) in the remainder, more precise overall estimates of the absolute contributions of each age group to the fishery could be obtained with little or no increase in overall sampling effort.

By making preliminary estimates of mean cluster size using the binomial two-way estimator for $[P]_{j(a-t)}$ and then inflating it by squaring the 'G' factor to determine the mean size of cluster for a given level of precision (Cochran 1953, p. 204), I found that the cluster sizes for 1955 were generally more than adequate for determining precisely (10 percent

TABLE 6
Point and Interval Estimates of Total 2-, 3-, 4-, and 5-year-old King Salmon
Entering the 1955 California Commercial Landings. (Point Estimates
Were Derived Using the Two-way (Port-month) Estimator.
Corresponding Interval Estimates Were Corrected for
Cluster Effects by an Appropriate 'G' Factor.)

Age (Years)	Estimated Proportion	Standard Error	'G'	Estimated Total	Corrected Interval Estimate*	
(j)	$\hat{P}_{(a-t)}$	$\sqrt{\text{Var}(\hat{P}_{(a-t)})}$		$Ns\hat{P}_{(a-t)}$	for $Ns\hat{P}_{(a-t)}$	
					Lower Limit	Upper Limit
2	0.0076	0.00123	2.82	5679	524	10836
3	0.5265	0.00709	1.96	393428	376391	414202
4	0.4353	0.00705	2.08	325279	303384	347174
5	0.0306	0.00245	1.40	22866	17710	28022
Total	1.0000			747252		

* 0.95 confidence probability level

TABLE 6
Point and Interval Estimates of Total 2-, 3-, 4-, and 5-year-old King Salmon Entering the 1955 California Commercial Landings. (Point Estimates Were Derived Using the Two-way (Port-month) Estimator. Corresponding Interval Estimates Were Corrected for Cluster Effects by an Appropriate 'G' Factor.)

coefficient of variation) the total number of 3- and 4-year-old king salmon in the catch. This was not the case for 2- and 5-year-old fish, however. In the former, a mean cluster size of about 3,290 fish would have been needed to provide an estimate of the total number in that age group having a correspondingly high precision. For 5-year-old salmon

at about the same level of precision, a mean cluster size of approximately 260 fish would have been required. The actual mean cluster size $[m]$ was 164.

5. SUMMARY

Because conditions under which king salmon catches must be sampled vary considerably in space and time, it is impossible to generalize as to the width of length strata (i) and sizes of length (n) and age (m) samples which, properly integrated, could be expected to yield satisfactory age composition estimates on a port-month basis. For a fixed stratum width, to improve two-phase sampling estimates requires larger age samples. Only slight improvement can be had by taking larger length samples (Tanaka 1953; Gulland 1955, 1957). Age sample components should adequately represent all length classes involved. Since proper allocation may not be possible, age-within-length samples (m_{ij}) could be fixed at some convenient level for all length strata, admitting wasted effort in those strata exhibiting minimal variability. Under certain conditions (assuming sufficiently large length samples can be obtained), a reasonable compromise might be to sample differentially on the basis of experience; that is, secure an m of any fixed size by judiciously taking for age determination fewer fish from length classes which would be expected to exhibit slight variation (p_{ij} relatively large) than from classes in which greater variability would be expected (p_{ij} relatively small).

In future programs utilizing the two-phase estimator, port-month estimates of the total king salmon catch (by number) in any age group should be based upon age samples consisting of scales from at least 200 specimens. From the standpoint of the labor involved in processing and reading salmon scales, experience has shown this is a reasonable quantity. Also, length strata should have a minimum width of 5 cm and those containing king salmon 80 cm long and longer should be sampled heavier than those with smaller fish. If 5-cm length classes are employed, an average of eight effective strata would likely be involved in commercial salmon landings, four of which would probably include fish 80 cm long and longer. Age samples from large-length strata, twice the size (if possible) of those from small-length strata, should provide, with minimum wasted effort, reasonably precise estimated total numbers of fish landed in each age group.

Both indirect- and direct-estimation age samples are dependent upon the number and distribution of ages in the catch. For a fixed m , both approaches appear to give comparable results where there are few ages, as with salmon. Indirect estimation does, however, attain increasing precedence over direct as the number of ages increases; such was the case with Japanese sardines (Tanaka 1953) and North Sea plaice (Gulland 1955). In light of the relatively slight gains exhibited for salmon by the two-phase technique over the simple random method, unless age samples are more than five times as costly as length samples, the direct method which is less expensive and less complicated would be preferred over the indirect in future sampling programs. Except

where conditions definitely warrant its use, the two-phase method is at least as costly as the simple direct method, is more complex computationally and is more difficult to execute insofar as actual sampling is concerned.

of even more concern than deciding which of several estimators is the most efficient for determining the absolute status of any attribute within individual groups (or clusters) of fish, is the problem of accounting for space-time variability between groups in attempts to secure best estimates for a management unit as a whole. I made such an attempt for salmon employing a two-way cluster sampling scheme that accounted for both port and month effects. My basic data were from direct estimation of individual cluster parameters (P_{at}); data obtained this way are more amenable to the techniques involved than those complicated by using length as an adjunct. Correcting for cluster effects, I determined that an average cluster size of about 250 age determinations could, in future programs, be expected to yield fairly precise (CV = 10 percent) over-all estimates of the number of king salmon in each age group commonly represented.

6. ACKNOWLEDGMENTS

I am grateful for the generous assistance of Raymond J. Jessen, C-E-I-R, Inc., whose critical review of the manuscript, subsequent comments, and many helpful suggestions added measurably to it. Errors of any sort, however, remain my full responsibility.

Acknowledgment is also given all personnel who helped carry out the California Department of Fish and Game's extensive catch-sampling program described herein.

6.1. REFERENCES

- Anderson, R. L., and T. A. Bancroft 1952. Statistical theory in research. New York, McGraw-Hill Book Co., 399 pp.
- Cochran, W. G. 1953. Sampling techniques. New York, J. Wiley & Sons, 330 pp.
- Deming, W. E. 1950. Some theory of sampling. New York, J. Wiley & Sons, 602 pp.
- Fridriksson, A. 1934. On the calculation of age-distribution within a stock of cod by means of relatively few age-determinations as a key to measurements on a large scale. Rapp. Cons. Explor. Mer, vol. 86, no. 6, pp. 1–14.
- Fry, D. H., Jr., and E. P. Hughes 1952. A sampling program for recovery of marked king and silver salmon. Calif. Fish and Game, vol. 38, no. 4, pp. 535–540.
1953. Proportion of king and silver salmon in California's 1952 landings. Calif. Dept. Fish and Game, Fish. Bull. No. 95, pp. 7–16.
- Gilbert, C. H. 1914. Age at maturity of the Pacific coast salmon of the genus *Oncorhynchus*. Bull. U. S. Bur. Fish., vol. 32, pp. 1–22.
- Gulland, J. A. 1955. Estimation of growth and mortality in commercial fish populations. Min. Agric. and Fish., Fish. Invest., ser. 2, vol. 18, no. 9, 46 pp.
1957. Sampling problems and methods in fisheries research. FAO Fish Bull., vol. 10, pp. 157–181.

- Hallock, R. J., G. H. Warner, and D. H. Fry, Jr. 1952. California's part in a three-state salmon fingerling marking program. *Calif. Fish and Game*, vol. 38, no. 3, pp. 301–332.
- Hodgson, W. C. 1939. An improved method of sampling herring shoals. *Rapp. Cons. Explor. Mer*, vol. 110, pp. 31–80.
- Tanaka, S. 1953. Precision of age-composition of fish estimated by double sampling method using the length for stratification. *Bull. Japan. Soc. Sci. Fish.*, vol. 19, no. 5, pp. 657–670.

APPENDIXES

APPENDIX I

32

FISH BULLETIN NO. 120

APPENDIX I

Length-age Data—Two-phase Stratified Sampling From California King
Salmon Sport Landings at San Francisco, February 1955. Estimated
Total Number Landed (N_{at}) : 8,500. Stratification
Is by 2.5-cm Intervals.

Length strata*	Length phase		Age phase							
	n_i	p_i	m_i	m_{i1}	p_{i1}	m_{i2}	p_{i2}	m_{i4}	p_{i4}	
30.0										
32.5	3	0.00462	1	1	1.0000					
35.0										
37.5	4	0.00616	1			1	1.0000			
40.0	8	0.01233	3	2	0.6667	1	0.3333			
42.5	11	0.01694	1			1	1.0000			
45.0	25	0.03853	2			2	1.0000			
47.5	61	0.09400	5			5	1.0000			
50.0	88	0.13560	17	1	0.0588	16	0.9412			
52.5	125	0.19262	14			14	1.0000			
55.0	123	0.18954	11			11	1.0000			
57.5	89	0.13712	10			9	0.9000	1	0.1000	
60.0	49	0.07550	10			10	1.0000			
62.5	32	0.04931	5			5	1.0000			
65.0	11	0.01693	1			1	1.0000			
67.5										
70.0	19	0.02926	2			2	1.0000			
72.5										

87.5	1	0.00154	1	—		—		1	1.0000	
Total..	649 = n	1.00000	84 = m	4		78		2		

* Mid-point, caudal fork length

Length-age Data—Two-phase Stratified Sampling From California King Salmon Sport Landings at San Francisco,
February 1955. Estimated Total Number Landed (N_{at}) : 8,500. Stratification Is by 2.5-cm Intervals

APPENDIX II

ESTIMATING SALMON AGE

33

APPENDIX II

Length-age Data—Two-phase Stratified Sampling From California King
Salmon Sport Landings at Crescent City-Eureka, July 1955. Estimated
Total Number Landed (N_{at}) : 10,660. Stratification
Is by 2.5-cm Intervals.

Length strata*	Length phase		Age phase							
	n_i	p_i	m_i	m_{i2}	p_{i2}	m_{i3}	p_{i3}	m_{i4}	p_{i4}	
40.0 }	7	0.00244	1	1	1.0000					
42.5 }										
45.0	5	0.00174	1	1	1.0000					
47.5	9	0.00313	2	2	1.0000					
50.0	15	0.00522	2	2	1.0000					
52.5	55	0.01913	4	4	1.0000					
55.0	91	0.03165	1			1	1.0000			
57.5	182	0.06330	6	1	0.1667	5	0.8333			
60.0	311	0.10817	11			11	1.0000			
62.5	333	0.11582	15			15	1.0000			
65.0	314	0.10920	13			11	0.8462	2	0.1538	
67.5	289	0.10050	8			7	0.8750	1	0.1250	
70.0	226	0.07860	10			5	0.5000	5	0.5000	
72.5	234	0.08138	6			3	0.5000	3	0.5000	
75.0	235	0.08173	7			2	0.2857	5	0.7143	
77.5	215	0.07478	13			4	0.3077	9	0.6923	
80.0	143	0.04974	13			1	0.0769	12	0.9231	
82.5	85	0.02966	7			1	0.1429	6	0.8571	
85.0 }	81	0.02817	1					1	1.0000	
87.5 }										
90.0	22	0.00765	4					4	1.0000	
92.5	13	0.00452	4					4	1.0000	
95.0 }										
97.5 }	10	0.00347	1					1	1.0000	
100.0										
Total ..	2875 = n	1.00000	130 = m	11		66		53		

* Mid-point, caudal fork length

Length-age Data—Two-phase Stratified Sampling From California King Salmon Sport Landings at Crescent City-Eureka, July 1955. Estimated Total Number Landed (N_{at}) : 10,660. Stratification Is by 2.5-cm Intervals

APPENDIX III

34

FISH BULLETIN NO. 120

APPENDIX III

Variance Calculations Employing Two-phase Stratified Sampling for Estimating Proportion of Two-year-old King Salmon From Data in Appendix II.
(Formulas and Explanation in Text.)

Effective length strata	(i) Terms in $\text{var}(p_{j(st)})$:		p_i	(ii) Terms in $\text{var}(p_{j(st)})$:		
	$p_i^2 p_{ij} q_{ij}$	$p_i (p_{ij} - p_{j(st)})^2$		$p_i q_{ij}$	$\sqrt{p_i q_{ij}}$	$p_i (p_{ij} - p_{j(st)})^2$
m_i	n					
40.0\}	-0-	0.000001	0.00244	-0-	--	0.00224
42.5\}						
45.0	-0-	0.000001	0.00174	-0-	--	0.00160
47.5	-0-	0.000001	0.00313	-0-	--	0.00287
50.0	-0-	0.000002	0.00522	-0-	--	0.00479
52.5	-0-	0.000006	0.01913	-0-	--	0.01755
55.0	-0-	--	0.03165	-0-	--	0.00006
57.5	0.000093	--	0.06330	0.13889	0.37268	0.00098
All other strata	-0-	0.000005	--	--	--	0.00155
$\sum_i$	0.000093	0.000016				$V_n = 0.03164$

$$p_{j(st)} = 0.04221$$

$$\text{var}(p_{j(st)}) = 0.000109$$

$$\sum_i p_i p_{ij} q_{ij} = 0.00879$$

$$\sum_i p_i \sqrt{p_i q_{ij}} = 0.02359$$

$$\left(\sum_i p_i \sqrt{p_i q_{ij}} \right)^2 = V_n = 0.00056$$

Variance Calculations Employing Two-phase Stratified Sampling for Estimating Proportion of Two-year-old King Salmon From Data in Appendix II

APPENDIX IV

ESTIMATING SALMON AGE

35

APPENDIX IV

Variance Calculations Employing Two-phase Stratified Sampling for Estimating Proportion of Three-year-old King Salmon From Data in Appendix II.
(Formulas and Explanation in Text.)

Effective length strata	(i) Terms in $\text{var}(p_{f(st)})$:		p_i	(ii) Terms in $\text{var}(p_i)_{opt}$:		
	$p_i^2 p_{ij} q_{ij}$ m_i	$p_i(p_{ij} - p_{f(st)})^2$ n		$p_{ij} q_{ij}$	$\sqrt{p_{ij} q_{ij}}$	$p_i(p_{ij} - p_{f(st)})^2$
55.0	-0-	0.000002	0.03165	-0-	--	0.00449
57.5	0.000093	0.000001	0.06330	0.13889	0.37268	0.00280
60.0	-0-	0.000005	0.10817	-0-	--	0.01536
62.5	-0-	0.000006	0.11582	-0-	--	0.01645
65.0	0.000119	0.000002	0.10920	0.13015	0.36076	0.00543
67.5	0.000138	0.000002	0.10050	0.10937	0.33071	0.00637
70.0	0.000156	--	0.07860	0.25000	0.50000	0.00119
72.5	0.000276	--	0.08138	0.25000	0.50000	0.00123
75.0	0.000195	0.000003	0.08173	0.20407	0.45174	0.00931
77.5	0.000092	0.000003	0.07478	0.21302	0.46154	0.00744
80.0	0.000013	0.000005	0.04974	0.07099	0.26644	0.01484
82.5	0.000016	0.000002	0.02966	0.12248	0.34997	0.00684
All other strata	-0-	0.000010	--	--	--	0.02931
$\sum_i$	0.001098	0.000042				$V_n = 0.12106$

$p_{f(st)} = 0.62315$	$\sum_i p_i p_{ij} q_{ij} = 0.11376$
$\text{var}(p_{f(st)}) = 0.001140$	$\sum_i p_i \sqrt{p_{ij} q_{ij}} = 0.27128$
	$\left(\sum_i p_i \sqrt{p_{ij} q_{ij}} \right)^2 = V_n = 0.07359$

Variance Calculations Employing Two-phase Stratified Sampling for Estimating Proportion of Three-year-old King Salmon From Data in Appendix II

APPENDIX V

Variance Calculations Employing Two-phase Stratified Sampling for Estimating Proportion of Four-year-old King Salmon From Data in Appendix II.

(Formulas and Explanation in Text.)

Effective length strata	(i)		p_i	(ii)		
	Terms in $\text{var}(p_{j(st)})$:			Terms in $\text{var}(p_j)_{opt}$:		
	$p_i^2 p_{ij} q_{ij}$	$p_i(p_{ij} - p_{j(st)})^2$		$p_{ij} q_{ij}$	$\sqrt{p_{ij} q_{ij}}$	$p_i(p_{ij} - p_{j(st)})^2$
	m_i	n				
65.0	0.000119	0.000001	0.10920	0.13015	0.36076	0.00357
67.5	0.000138	0.000002	0.10050	0.10937	0.33071	0.00442
70.0	0.000156	--	0.07860	0.25000	0.50000	0.00215
72.5	0.000276	--	0.08138	0.25000	0.50000	0.00222
75.0	0.000195	0.000004	0.08173	0.20407	0.45174	0.01178
77.5	0.000092	0.000003	0.07478	0.21302	0.46154	0.00956
80.0	0.000013	0.000006	0.04974	0.07099	0.26644	0.01722
82.5	0.000016	0.000003	0.02966	0.12248	0.34997	0.00809
85.0	-0-	0.000004	0.02817	-0-	--	0.01247
87.5						
90.0	-0-	0.000001	0.00765	-0-	--	0.00339
92.5	-0-	0.000001	0.00542	-0-	--	0.00200
95.0	-0-	0.000001	0.00347	-0-	--	0.00154
97.5						
All other strata	-0-	0.000014	--	--	--	0.03926
$\sum_i$	0.001005	0.000040				$V_n = 0.11767$

$$p_{j(st)} = 0.33464$$

$$\text{var}(m_{j(st)}) = 0.001045$$

$$\sum_i p_i p_{ij} q_{ij} = 0.10497$$

$$\sum_i p_i \sqrt{p_{ij} q_{ij}} = 0.24769$$

$$\left(\sum_i p_i \sqrt{p_{ij} q_{ij}} \right)^2 = V_n = 0.06135$$

Variance Calculations Employing Two-phase Stratified Sampling for Estimating Proportion of Four-year-old King Salmon From Data in Appendix II

APPENDIX VI

APPENDIX VI

Length-age Data—Two-phase Stratified Sampling From California King Salmon Sport Landings Crescent City-Eureka, July 1955. Estimated Number Landed (N_{st}): 10,660. Stratification Is by Six, 10-cm Intervals.

Length strata	Length phase		Age phase							
	n_i	p_i	m_i	m_{i2}	p_{i2}	m_{i3}	p_{i3}	m_{i4}	p_{i4}	
-48.8	21	0.00730	4	4	1.0000					
49.0-58.8	343	0.11931	13	7	0.5385	6	0.4615			
59.0-68.8	1247	0.43374	47			44	0.9362	3	0.0638	
69.0-78.8	910	0.31652	36			14	0.3889	22	0.6111	
79.0-88.8	309	0.10748	21			2	0.0952	19	0.9048	
89.0-98.8	45	0.01565	9					9	1.0000	
Total----	2875 = n	1.00000	130 = m	11		66		53		

Length-age Data—Two-phase Stratified Sampling From California King Salmon Sport Landings Crescent City-Eureka, July 1955. Estimated Number Landed (N_{st}): 10,660. Stratification Is by Six, 10-cm Intervals

APPENDIX VII

APPENDIX VII
Variance Calculations Employing Two-phase Stratified Sampling for Estimating Proportions of Two-, Three-, and Four-year-old King Salmon From Data in Appendix VI. (Formulas and Explanation in Text.)

Age.....		Two years		Three years		Four years	
Length strata	p_i	$\sqrt{p_i q_i}$	$p_i(p_i - p_{j(u)})^2$	$\sqrt{p_i q_i}$	$p_i(p_i - p_{j(u)})^2$	$\sqrt{p_i q_i}$	$p_i(p_i - p_{j(u)})^2$
48-8	0.00730	--	0.00629	0.00135	--	--	0.00081
49-0-58-8	0.11921	0.49852	0.02601	0.49852	0.00211	--	0.01331
59-0-68-8	0.43374	--	--	0.24440	0.05066	0.24440	0.03166
69-0-78-8	0.31652	--	0.00447	0.48750	0.01337	0.48750	0.02431
79-0-88-8	0.10748	--	--	0.29350	0.02670	0.29350	0.03502
89-0-98-8	0.01565	--	--	--	0.00676	--	0.00694
Total	1.00000		0.03677 = V_{u2}		0.10104 = V_{u3}		0.11205 = V_{u4}
$\frac{p_{j(u)}}{\text{var}(p_{j(u)})}$		0.07155		0.59446		0.33399	
$\sum_i \frac{p_i}{\sqrt{p_i q_i}}$		0.00030		0.00127		0.00097	
$(\sum_i p_i \sqrt{p_i q_i})^2$		0.05948		0.35133		0.29185	
			0.00854		0.12343		0.08518

ESTIMATING SALMON AGE

37

Variance Calculations Employing Two-phase Stratified Sampling for Estimating Proportions of Two-, Three-, and Four-year-old King Salmon From Data in Appendix VI. (Formulas and Explanation in Text.)

APPENDIX VIII

APPENDIX VIII

Comparison of Age Structure Estimates and Their Variances Employing Two-phase Sampling for Two Stratification Levels. Raw Data Are in Appendixes II and VI.

Age group	2	3	4
From data in Appendix II—2.5-cm length strata:			
p_j	0.04221	0.62315	0.33464
$\hat{N}_j$	450	6643	3567
$\text{var}(p_j)$	0.000109	0.001140	0.001045
$\text{var}(\hat{N}_j)$	12386	129545	118750
$\text{CV}(\hat{N}_j)$	0.247	0.054	0.097
From data in Appendix VI—10-cm length strata:			
p_j	0.07155	0.59446	0.33399
$\hat{N}_j$	763	6337	3560
$\text{var}(p_j)$	0.000300	0.001270	0.000970
$\text{var}(\hat{N}_j)$	34091	144317	110226
$\text{CV}(\hat{N}_j)$	0.242	0.060	0.093

Comparison of Age Structure Estimates and Their Variances Employing Two-phase Sampling for Two Stratification Levels. Raw Data Are in Appendixes II and VI

APPENDIX IX

APPENDIX IX

Estimated Numbers by Age ($\hat{N}_{j(st)}$) of King Salmon Landed Commercially at California Ports During 1952. Figures in Parentheses Are Variation Coefficients Expressed as Percentages. Empty Cells Indicate No Sampling.

Age group	(j)	2 years		3 years		4 years		5 years	
Crescent City	May	--	--	--	--	--	--	--	--
	Jun	--	--	6081	(7.5)	1789	(17.8)	--	--
	Jul	--	--	--	--	--	--	--	--
	Aug	--	--	--	--	--	--	--	--
	Sep	--	--	--	--	--	--	--	--
Eureka	May	--	--	772	(15.1)	611	(15.8)	--	--
	Jun	--	--	16409	(6.3)	6792	(13.9)	105	(88.3)
	Jul	--	--	--	--	--	--	--	--
	Aug	--	--	9766	(6.2)	1777	(25.3)	140	(55.8)
	Sep	--	--	--	--	--	--	--	--
Fort Bragg	May	--	--	--	--	--	--	--	--
	Jun	--	--	--	--	--	--	--	--
	Jul	132	(81.2)	23722	(5.2)	9481	(10.7)	406	(63.3)
	Aug	85	(83.5)	12599	(3.6)	3249	(10.5)	454	(34.2)
	Sep	--	--	--	--	--	--	--	--
San Francisco	May	--	--	--	--	--	--	--	--
	Jun	--	--	--	--	--	--	--	--
	Jul	--	--	20876	(8.1)	32994	(5.6)	2178	(27.5)
	Aug	--	--	12706	(6.8)	6595	(12.1)	1540	(23.3)
	Sep	--	--	37946	(5.4)	13667	(12.0)	2717	(16.5)
Monterey	May	--	--	12193	(11.0)	17116	(8.5)	1087	(32.0)
	Jun	--	--	22326	(6.7)	10406	(13.3)	99	(56.4)
	Jul	--	--	5386	(9.2)	6430	(8.3)	678	(27.9)
	Aug	--	--	--	--	--	--	--	--
	Sep	--	--	--	--	--	--	--	--
Sacramento River	Sep	157	(25.7)	7041	(6.0)	5577	(7.5)	353	(34.5)

Estimated Numbers by Age ($N_{j(st)}$) of King Salmon Landed Commercially at California Ports During 1952. Figures in Parentheses Are Variation Coefficients Expressed as Percentages. Empty Cells Indicate No Sampling

APPENDIX X

ESTIMATING SALMON AGE

39

APPENDIX X
Estimated Total Numbers ($\hat{N}_{j(i)}$) and Derived Proportions ($\hat{P}_j$) of Age j King Salmon. Basic Data From 1952 California Commercial Landing Samples. Empty Cells Indicate No Sampling.

Port-area (a)	Age (j)	May	Jun	Month (i) Jul	Aug	Sep	Total	$\hat{P}_j$
Crescent City	2	--	--	--	--	--	--	--
	3	--	6081	--	--	--	6081	0.7727
	4	--	1789	--	--	--	1789	0.2273
	5	--	--	--	--	--	--	--
	N_{st}	--	7870	--	--	--	7870	
Eureka	2	--	--	--	--	--	--	--
	3	772	16409	--	9766	--	26947	0.7409
	4	611	6792	--	1777	--	9180	0.2524
	5	--	105	--	140	--	245	0.0067
	N_{st}	1383	23306	--	11683	--	36372	
Fort Bragg	2	--	--	132	85	--	217	0.0043
	3	--	--	23722	12599	--	36321	0.7246
	4	--	--	9481	3249	--	12730	0.2539
	5	--	--	406	454	--	860	0.0172
	N_{st}	--	--	33741	16387	--	50128	
San Francisco	2	--	--	--	--	--	--	--
	3	--	--	20876	12706	37946	71528	0.5451
	4	--	--	32994	6595	13667	53256	0.4059
	5	--	--	2178	1540	2717	6435	0.0490
	N_{st}	--	--	56048	20841	54330	131219	
Monterey and South	2	--	--	--	--	--	--	--
	3	12193	22326	5386	--	--	39905	0.5270
	4	17116	10406	6430	--	--	33952	0.4484
	5	1087	99	678	--	--	1864	0.0246
	N_{st}	30396	32831	12494	--	--	75721	
All Ports	2	--	--	132	85	--	217	
	3	12965	44816	49984	35071	37946	180782	
	4	17727	18987	48905	11621	13667	110907	
	5	1087	204	3262	2134	2717	9404	
	N_{st}	31779	64007	102283	48911	54330	301310	
$\hat{P}_j$	2	--	--	0.0013	0.0018	--	$\hat{P}_{2(i)} = 0.0007$	
	3	0.4080	0.7002	0.4887	0.7170	0.6985	$\hat{P}_{3(i)} = 0.6000$	
	4	0.5578	0.2966	0.4781	0.2376	0.2515	$\hat{P}_{4(i)} = 0.3681$	
	5	0.0342	0.0032	0.0319	0.0436	0.0500	$\hat{P}_{5(i)} = 0.0312$	

Estimated Total Numbers ($[N]_{j(st)}$) and Derived Proportions ($[P]_j$) of Age j King Salmon. Basic Data From 1952 California Commercial Landing Samples. Empty Cells Indicate No Sampling

APPENDIX XI

40

FISH BULLETIN NO. 120

APPENDIX XI

Age Sampling Data for 1955 California Commercial King Salmon Landings. Data in Right and Lower Margins Used to Derive Estimates in Table 5.

Port-area (a)	Age (j)	May	Jun	Month (i) Jul	Aug	Sep	Total	$\hat{P}_a$	$N_a \hat{P}_a$
Crescent City	2	--	--	--	--	--	--	--	--
	3	42	54	42	31	12	181	0.4049	36142
	4	128	40	51	15	17	251	0.5615	50121
	5	6	1	4	1	3	15	0.0336	2999
	m_{at}	176	95	97	47	32	447		
Eureka	N_{at}	10908	17795	44762	11893	3904	89262		
	2	--	--	--	--	1	1	0.0013	233
	3	65	163	79	76	37	420	0.5512	98798
	4	111	101	55	35	4	306	0.4016	71983
	5	8	8	10	6	3	35	0.0459	8227
Fort Bragg	m_{at}	184	272	144	117	45	762		
	N_{at}	45743	72988	36741	10242	13527	179241		
	2	--	--	--	1	6	7	0.0108	1614
	3	26	55	67	79	77	304	0.4699	70240
	4	22	77	69	63	80	311	0.4807	71854
San Francisco	5	--	14	7	3	1	25	0.0386	5770
	m_{at}	48	146	143	146	164	647		
	N_{at}	3980	36225	41875	42092	25306	149478		
	2	--	1	1	7	7	16	0.0122	3142
	3	89	131	141	250	112	723	0.5523	142255
Monterey and South	4	104	116	68	140	122	550	0.4202	108231
	5	5	6	3	2	4	20	0.0153	3941
	m_{at}	198	254	213	399	245	1309		
	N_{at}	20036	97854	50269	62328	27082	257569		
	2	--	--	12	22	--	34	0.0367	2519
All Ports	3	104	203	92	126	--	525	0.5663	38864
	4	86	184	44	43	--	357	0.1851	26428
	5	2	8	1	--	--	11	0.0119	817
	m_{at}	192	395	149	191	--	927		
	N_{at}	30440	20074	11264	6850	3074	*68628		
$\hat{P}_a$	2	--	1	13	30	14	58		
	3	326	606	421	562	238	2153		
	4	451	518	287	296	223	1775		
	5	21	37	25	12	11	106		
	m_{at}	798	1162	746	900	486	4092		
$N_a \hat{P}_a$	N_{at}	111107	244936	184911	133405	*69819	*744178		
	2	--	0.0009	0.0174	0.0333	0.0288			
	3	0.4085	0.5215	0.5643	0.6245	0.4897			
	4	0.5652	0.4458	0.3848	0.3289	0.4588			
	5	0.0263	0.0318	0.0335	0.0133	0.0227			
$N_a \hat{P}_a$	2	--	221	3217	4442	2011			
	3	45387	127734	104345	83312	34190			
	4	62798	109192	71154	43887	32033			
	5	2922	7789	6195	1774	1585			

* Does not include N_{at} for Monterey: September (no sampling).

Age Sampling Data for 1955 California Commercial King Salmon Landings. Data in Right and Lower Margins Used to Derive Estimates in Table 5

APPENDIX XII

APPENDIX XII
Observed, $\hat{P}_{at}$, and Deviation From Regression, $\hat{P}_{at} - E(P)$, * Values for Calculating
Port-month Interaction Sum of Squares and Mean Square Error for Two-way
Cluster Sampling. $\hat{P}_{at}$'s are Cluster Estimates of Proportions of King
Salmon in Each of Four Age Groups (j) Landed by the 1955
California Commercial Fishery. Sample Data
Are in Appendix XI.

Port-area (a)	Age (j)	Month (i)										Mean $\hat{P}_{.j}$	ESTIMATING SALMON AGE
		May		Jun		Jul		Aug		Sep			
		$\hat{P}_{at}$	$\hat{P}_{at} - E(P)$	$\hat{P}_{at}$	$\hat{P}_{at} - E(P)$	$\hat{P}_{at}$	$\hat{P}_{at} - E(P)$	$\hat{P}_{at}$	$\hat{P}_{at} - E(P)$	$\hat{P}_{at}$	$\hat{P}_{at} - E(P)$		
Crescent City.....	2		0.0154	--	0.0146	--	-0.0017	--	-0.0125	--	-0.0158	--	
	3	0.2386	-0.1098	0.5684	0.1302	0.4330	-0.0363	0.6596	0.1089	0.3750	-0.0631	0.4549	
	4	0.7273	0.0920	0.4211	-0.1103	0.6238	0.0420	0.3191	-0.0890	0.5312	0.0651	0.5049	
	5	0.0341	0.0024	0.0105	-0.0345	0.0412	-0.0040	0.0213	-0.0074	0.0938	0.0438	0.0402	
Eureka.....	2		0.0110	--	0.0102	--	-0.0061	--	-0.0169	0.0222	0.0020	0.0044	
	3	0.3533	-0.1348	0.5963	0.0214	0.5486	-0.0604	0.6496	-0.0408	0.3222	0.2144	0.5946	
	4	0.6033	0.1240	0.3713	-0.0041	0.3819	0.0541	0.2991	0.0470	0.0889	-0.2212	0.3489	
	5	0.0434	-0.0002	0.0294	-0.0275	0.0695	0.0124	0.0517	0.0107	0.0667	0.0048	0.0521	
Fort Bragg.....	2		0.0067	--	0.0059	--	-0.0104	0.0068	-0.0144	0.0366	0.0121	0.0087	
	3	0.5417	0.1687	0.3767	-0.0961	0.4685	-0.0254	0.5411	-0.0342	0.4695	-0.0232	0.4795	
	4	0.4583	-0.1496	0.5274	0.0234	0.4825	0.0261	0.4315	0.0508	0.4878	0.0491	0.4775	
	5	--	-0.0258	0.0659	0.0568	0.0490	0.0097	0.0206	-0.0022	0.0061	-0.0380	0.0343	
San Francisco.....	2		0.0045	0.0039	0.0076	0.0047	-0.0079	0.0175	-0.0059	0.0286	0.0019	0.0109	
	3	0.4495	0.0138	0.5158	-0.0097	0.6620	0.1054	0.6266	-0.0114	0.4571	-0.0983	0.5422	
	4	0.0253	-0.0531	0.4567	0.0002	0.5192	-0.0897	0.3509	-0.0177	0.4080	0.1068	0.4300	
	5	0.0252	0.0168	0.0236	0.0019	0.0141	-0.0078	0.0050	-0.0004	0.0163	-0.0104	0.0109	
Monterey and South.....	2		-0.0375	--	-0.0383	0.0805	0.0259	0.1152	0.0498	0.0687	0.0000	0.0529	
	3	0.5417	0.0617	0.5139	-0.0559	0.6175	0.0166	0.6597	-0.0226	0.5966	-0.0001	0.5865	
	4	0.4479	-0.0313	0.4658	0.0905	0.2653	-0.0324	0.2251	-0.0269	0.3100	0.0000	0.3488	
	5	0.0104	0.0071	0.0203	0.0037	0.0067	-0.0101	--	-0.0003	0.0217	0.0001	0.0118	
Mean	2		0.0008	--	0.0008	0.0171	0.0279	--	0.0512	--	0.0154	0.0154	
	3	0.4250	0.5148	--	0.5459	--	0.6273	--	0.5447	--	0.5315	0.5315	
	4	0.5524	0.4485	--	0.4069	--	0.3252	--	0.3832	--	0.4220	0.4220	
	5	0.0226	0.0359	--	0.0361	--	0.0196	--	0.0409	--	0.0311	0.0311	
$\hat{P}_{.j}/s$											$\hat{P}_{.j}/s$		

* All row and column deviations do not add to zero because of rounding errors.
u = number of ports; v = number of months;

Observed, $[P]_{at}$, and Deviation From Regression, $[P]_{at} - E(P)$, * Values for Calculating Port-month Interaction
Sum of Squares and Mean Square Error for Two-way Cluster Sampling. $[P]_{at}$'s are Cluster Estimates of Propor-
tions of King Salmon in Each of Four Age Groups (j) Landed by the 1955 California Commercial Fishery. Sample
Data Are in Appendix XI

APPENDIX XIII

42

FISH BULLETIN NO. 120

APPENDIX XIII

Statistics for Calculating Port-month Interaction Mean Square and Correction Factor 'Gj'. Basic Age Data Are From 1955 California Commercial King Salmon Landing Samples. (See Appendixes XI, XII, and Text.)

Port-area (a)	Month (t)	N _{at}	m _{at}	m ² _{at}	Age group (j)			
					2 years $\hat{P}_{at} - E(P)^2$	3 years $\hat{P}_{at} - E(P)^2$	4 years $\hat{P}_{at} - E(P)^2$	5 years $\hat{P}_{at} - E(P)^2$
Crescent City	May	10908	176	30976	0.000237	0.012056	0.008464	0.000006
	Jun	17795	95	9025	0.000213	0.016952	0.012166	0.001190
	Jul	44762	97	9409	0.000003	0.001318	0.001764	0.000016
	Aug	11893	47	2209	0.000156	0.011859	0.007921	0.000055
	Sep	3904	32	1024	0.000250	0.008608	0.004238	0.001918
Eureka	May	45743	184	33856	0.000121	0.017767	0.015376	0.000000
	Jun	72988	272	73984	0.000104	0.000458	0.000017	0.000756
	Jul	36741	144	20736	0.000037	0.003648	0.002927	0.000154
	Aug	10242	117	13689	0.000286	0.001665	0.002209	0.000114
	Sep	13527	45	2025	0.000004	0.045967	0.048929	0.000023
Fort Bragg	May	3980	48	2304	0.000045	0.028460	0.022380	0.000666
	Jun	36225	146	21316	0.000035	0.007413	0.000548	0.003226
	Jul	41875	143	20449	0.000108	0.000645	0.000681	0.000094
	Aug	42092	146	21316	0.000207	0.001170	0.002581	0.000005
	Sep	25306	164	26896	0.000146	0.000538	0.002411	0.001444
San Francisco	May	20036	198	39204	0.000020	0.000190	0.001232	0.000282
	Jun	97854	254	64516	0.000058	0.000094	0.000000	0.000004
	Jul	50269	213	45369	0.000062	0.011109	0.008046	0.000061
	Aug	62328	399	159201	0.000035	0.000130	0.000313	0.000000
	Sep	27082	245	60025	0.000004	0.009663	0.011406	0.000108
Monterey and South	May	30440	192	36864	0.001406	0.003807	0.000980	0.000050
	Jun	20074	395	156025	0.001467	0.003125	0.008190	0.000014
	Jul	11264	149	22201	0.000671	0.000276	0.001050	0.000102
	Aug	6850	191	36481	0.002480	0.000511	0.000724	0.000000
	Sep	3074	0	0	0.000000	0.000000	0.000000	0.000000
Total		747252	4092					
Weighted sum of squares: port-month interaction = $\sum_a \sum_t m^2_{at} (\hat{P}_{at} - E(P))^2$ -----					437	3448	3831	212
Weighted mean square: port-month interaction = (A) -----					0.00002954	0.00023309	0.00025898	0.00001433
var($\hat{P}_{..} / uv$) _{int} = (B) -----					0.00000371	0.00006058	0.00005960	0.00000735
'G' = $\sqrt{(A)/(B)}$ -----					2.82	1.96	2.08	1.40

Statistics for Calculating Port-month Interaction Mean Square and Correction Factor 'Gj'. Basic Age Data Are From 1955 California Commercial King Salmon Landing Samples. (See Appendixes XI, XII, and Text.)