NOAA Technical Memorandum NMFS

**JULY 1984**

# SAMPLING COMMERCIAL ROCKFISH LANDINGS IN CALIFORNIA

A. R. Sen

NOAA-TM-NMFS-SWFC-45

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

NOAA Technical Memorandum NMFS

**NOAA Technical Memorandum NMFS**

This TM series is used for documentation and timely communication of preliminary results, imterim reports, or special purpose information, and have not received complete formal review, editorial control, or detailed editing.

**JULY 1984**

# SAMPLING COMMERCIAL ROCKFISH LANDINGS IN CALIFORNIA

A. R. Sen
Department of Statistics and Applied Probability
University of Alberta
Edmonton, Alberta, Canada

NOAA-TM-NMFS-SWFC-45

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

This report was prepared for the National Marine Fisheries Service, Southwest Fisheries Center, by A. R. Sen under cooperative agreement 83-ABH-00025 with the Humboldt Foundation, Humboldt State University. The statements, findings, conclusions and recommendations herein are those of the author and do not necessarily reflect the views of the National Marine Fisheries Service. Norman J. Abramson, Tiburon Laboratory, Southwest Fisheries Center served as Technical Representative for this contract.

CONTENTS

PAGE

1. INTRODUCTION 1

2. DATA COLLECTION 2

3. DESIGN OF THE SURVEY 3

4. COLLECTION OF REPRESENTATIVE SAMPLE DATA FROM LANDINGS AND CLUSTERS 4

5. ESTIMATION PROCEDURE 7

5.1 Notation 8

5.2 Estimates of Mean, Total and Errors 9

5.3 Estimation Based on Random Categories 11

5.4 Estimation Based on Post-Stratification 13

5.4.1 Cluster Size 14

5.5 General Case: Unequal Cluster Size 16

5.5.1 Random Categories 16

5.5.2 Post-Stratification into Sort Groups 17

5.5.3 Cluster Weight Constant Within Sort Group 20

5.5.4 Estimation of Ratio to Another Variable 21

5.6 Comparison of Methods. Simple Random Sampling vs. Post-Stratification 22

6. OPTIMUM SAMPLING AND SUB-SAMPLING FRACTIONS 25

6.1 Cost Function 27

6.2 Optimum Allocation: Survey Data 29

6.3 Optimum Allocation: Experimental data 30

6.4 Variance Components: Species-sex-age and length Groups 36

7. RELATIVE EFFICIENCY OF ESTIMATORS USING POST-STRATIFICATION 40

CONTENTS - Continued

8. RELIABILITY OF THE ESTIMATES: COEFFICIENT OF VARIATION 44

8.1 Coefficient of Variation (c.v.) of Species Catch by Port-Year Groups 44

8.2 Coefficient of Variation (c.v.) of Species Weight by Port-Year-Sort Groups 46

8.3 Coefficient of Variation of Species by Sex-Age for Port-Year Groups 46

9. RELIABILITY OF THE ESTIMATES: SAMPLE SIZE 62

9.1 Sample Size for Species Catch by Port-Year Groups 63

9.2 Sample Size for Species Weight by Port-Year Sort Groups 65

9.3 Sample size for Species Catch by Sex and Age for Port-Year Groups 66

9.4 Sample Size for Species Catch by Ports on Quarterly Basis 66

10. AGE COMPOSITION 79

10.1 Number of Strata 81

10.2 Strata Boundaries 82

10.3 Optimum Allocation Plan 83

11. SUMMARY AND *RECOMMENDATIONS* 87

ACKNOWLEDGEMENTS 93

LITERATURE CITED 94

## 1. INTRODUCTION

This study was undertaken during 1983 (May 15-October 15) under agreement between the present author, the Humboldt State University Foundation and the Department of Commerce, National Oceanic and Atmospheric Administration (NOAA), National Marine Fisheries Service Tiburon Laboratory. The purpose of the investigation was in the first place to determine if the present sampling plan for the estimation of parameters such as age and species composition of California rockfish landings is workable; i.e., if the operational procedures being utilized during the execution of the plan meet its requirements. If not, whether a revised plan could be designed which would conform to operationally feasible sampling procedures. It was also intended to obtain, under certain assumptions, the precision of the estimates of total catch of rockfish by species and by sex-age groups stratified by type of fishery and port and time of landing in California north of Point Arguello; and compare these with the precision of alternative estimators with a view to find the most efficient estimator. Finally, it was required to determine the optimum sample size for estimating the parameters as accurately as possible within the usual limitations of budget and personnel.

The fishery can be broadly divided into two types--commercial and sport. Commercial fishing can be classified into fishing by trawlers and by nontrawlers. Sport fishing is generally done by party and by non-party boats, from piers, and from shore other than piers.

In view of the greater importance and complexity of the sampling and operational problems involved in commercial fishing, the present study will be confined to an examination of the problems outlined above in the commercial fishery and to the important species, Widow (*Sebastes entomelas*)

Bocaccio (*Sebastes paucispinis*) and Chilipepper (*Sebastes goodei*). It is proposed to take up the study of sport fishing as a separate project at a later date.

We will now briefly review the data collection process and the principal features of the current design to understand the extent to which the assumptions made at the data collection and estimation stage are justified and the changes suggested in the sampling plan when the assumptions are not met.

## 2. DATA COLLECTION

Rockfish are being landed at 14 ports on the California coast. Of these, three cater only to commercial fishing, four to sport fishing and seven to both sport and commercial fishing. The ten commercial ports are grouped into six port groups with a sampler (six in all) assigned to each--Eureka, Fort Bragg, Bodega Bay, San Francisco, Monterey and Morro Bay. In the current plan a sampler is expected to visit each port in his jurisdiction at least once a week for gathering information on sport and commercial fishing. The samplers employed hold temporary jobs which range from 4 to 24 months in duration.

Trawlers make trips varying in length from one to eight days. These vessels maintain logbooks to keep records of area fished and appropriate catch for each tow. Sampling by tow is generally not feasible since the sampler has to be on board during haul time. For the same reasons no estimates of fish being rejected and returned to the sea are obtained since this would involve collection of discarded fish from randomly selected tows within sampled trips. Also, sampling by area of catch is not practicable unless all tows are in a limited area and the sampler is on board at haul time.

## 3. DESIGN OF THE SURVEY

A two-stage stratified random sampling plan with port-month group as stratum and boat trips within a stratum as first-stage sampling units was adopted. To take advantage of the sorting at sea provided by market categories, the first-stage sampling units are post-stratified into sort groups and at least one cluster of a given weight is subsampled within each sort type. Cluster size is either 25 or 50 pounds. Twenty-five pounds cluster is taken when sampling small fish like *Sebastes aurora*, *Sebastes saxicola* or *Sebastes diploproa* or any time small rockfish are landed such that there would be more than 20 fish in the fifty pounds (lbs.) cluster. In all other cases 50 lbs. standared cluster size is selected. A cluster is next separated by number of each species and its weight, which are recorded along with sex, total length and otoliths from *Sebastes* and *Sebastolobus*.

The instructions are to "sample all market categories from a boat, and from as many boats as possible and select:

(i) 1 cluster per 20,000 lbs. of widow rockfish landed by each boat, up to 4 clusters,

(ii) 1 cluster for all other species, if less than 5,000 lbs. landed and

(iii) 2 clusters for all other species if more than 5,000 lbs. are landed. The second cluster should not be taken if this precludes sampling another boat."

Besides, the sampler has to obtain from the skipper the total weight of each category of a sampled landing, as would be clear from the discussion in Section 5.6.

4. COLLECTION OF REPRESENTATIVE SAMPLE DATA FROM LANDINGS AND CLUSTERS

Owing to the uncertainty of arrival times and varying unloading procedures, no objective method is available to ensure random sampling of the trips. When the vessels return to port, they are usually available for sampling except when the landing is transshipped immediately due to inclement weather or lack of processing facilities, uncooperative buyers, or due to unscheduled deliveries at short notice. It is, however, not unreasonable to assume that a representative set of sample landings can be obtained from a port-month stratum.

Although rockfish are landed by sort groups which are mostly determined by market agreement based on size, composition and condition of the catch, the number of sorts per delivery can not be predetermined. This number would vary from delivery to delivery and from dealer to dealer. There are no guarantees that a complete boat sample, covering clusters from each sort group can be taken on any sampling day and some of the categories may be missed in sampling. Some of the possible reasons for missing the categories are (i) when landing would not occur during regular hours (i.e., early morning or late nights) one of the sorts may have already been shipped before the sampler could arrive at the spot (ii) often one of the sorts may be quite small and there may be a buyer at the dock waiting for the fish to be taken away; and (iii) while the sampler is working on a sort, the other sort/s will have either been processed or shipped away and finally (iv) the sampler may be prevented from taking a sample from another sort since the skipper may not like some of his fish being cut. This may, generally, happen at ports where either processing facilities may not be adequate, and a large portion of the landings are shipped, or are bought by local merchants immediately after landing. The question arises if failure to sample from all the categories of a sample landing as required in the current plan would cause appreciable bias and consequent loss in efficiency in the current estimates of species number and its weight. It is proposed to examine this as well as the feasibility of the current plan based on actual data collected in the past.

It may be pointed out that the current technique of selecting a cluster (box) of fish as second-stage sampling unit was rightly preferred to the 'grab technique' based on the assumption of random selection of fish by the sampler since in practice the potential of personal bias of the sampler could be considerable. Tomlinson (1971) feels that a sampler may have a tendency to choose fish with certain qualities and thus may introduce procedural bias. Tomlinson sees no way to avoid the conclusion that "a simple random sample of individual fish is operationally impracticable."

The selection of a representative cluster would depend upon whether samples after sorting on the vessel come from bins, strap boxes or off conveyor belts. At Monterey they are mostly unloaded into metal bins which are either placed in a large cooler or transferred to a conveyor belt for transport to the fillet line. Buyers from small markets occasionally select fish from the top of bins. Hence, to avoid bias, it is preferable to select the cluster from the conveyor belt which exposes unsorted fish from the lower portion of the bin. I noticed this practice being rightly followed at Monterey during my visit to the place. However, where small-market buyers do not buy fish, a cluster may be selected from a bin. Where many bins are present, a systematic sample of two clusters one from each separated in time and preferably from the beginning and end of the trip should be selected. It may be pointed out that an efficient method of subsampling is to assign a number to all individuals in the population and select from it a sample of the required size with the aid of a table of random numbers. In practice, this procedure is not feasible since it is too expensive and time-consuming. Where fish are graded on a conveyor belt before they enter the plant (e.g., Fields Landing at Eureka), the sampler should try to intercept the landings prior to secondary sorting or obtain separate weights for each sub-sort category. Whichever method is relatively easier in practice should be followed.

As has been pointed out earlier, bias may also arise through personal selection of fish from within a cluster. If the sampler were to select a number of clusters with too few fish per cluster (e.g., 25 lbs. cluster for medium and large fish) a cluster will, on the average, contain more of big fish. Thus, although the procedure will reduce the sampling error, it will tend to increase considerably the non-sampling error due to selection bias which could be serious. Sometimes, the top few fish in a bin are selected and put there to impress small buyers. The resulting bias in selection can be avoided by taking all the fish in a cluster (e.g., 50 lbs.) from one side so that the sample measured consists of all the fish originally to one side of the box. At Monterey, Eureka (where sampling of cluster is either from conveyor belt or from bins) and at San Francisco (where clusters are generally selected from strap boxes) we observed the samplers rightly selecting the clusters (without looking at the fish) from one side in a cylindrical section thus collecting fish of different sizes and kinds over the section.

To summarize, random sampling of boat trips is not practicable owing to the uncertainty of their arrival times and the most reasonable assumption is that the boats arrive at a port during a month in a random order. Selection of a random and representative cluster (box) of fish from a boat trip would depend to a large extent on the expertise and experience of the sampler. Hence, the need to have permanent staff at least at the important ports to build the experience.

In general, selection of a cluster for a market category should be done before any presorting is done at the port either on conveyor belt or in bins or from strap boxes; it is felt that fish landed from strap boxes are likely to be subjected to greater sorting by length, etc. than ones from bins. Clusters should be selected for each market category separated in time, e.g., at the beginning and termination of loading from a box, bin or off conveyor belt. As far as practicable, selection should be made from one side of a box, including fish from the top all the way down to the bottom, and fish selected should not be seen in the process.

For obtaining reliable, comprehensive and complete information on population characteristics, it is necessary that good relationships be maintained by the sampler with both the skipper and the buyer. This depends to a large extent on the experience of the sampler gained in the course of his field work over years. This emphasizes again, the need for permanent staff at least at the important ports which have too many problems to handle within a short time.

## 5. ESTIMATION PROCEDURE

Consider the problem of estimating the total catch of a given species for a port-month stratum. The formulas for estimation of other characteristics are straight forward and can be obtained by substituting the value of the characteristic for the catch of the species.

## 5.1. Notation

Let

$N$ = total number of trips,

$n$ = number of randomly sampled trips,

$W$ = total weight of fish caught from all trips,

$W_i$ = weight of fish caught on trip $i$,

$W_{ij}$ = weight of sort $j$ caught in trip $i$,

$m_{ij}$ = number of clusters sampled from sort $j$ of trip $i$,

$W_i = \sum_{j}^{L_i} W_{ij}$ where $L_i$ is number of sorts in trip $i$,

$y_{ijk}$ = number of the species in cluster $k$ from sort $j$ of trip $i$,

$Y_{ij}$ = total number of the species caught from sort $j$ of trip $i$,

$Y$ = total number of the species caught from all trips,

$\bar{y}_{ij} = \sum_{k}^{m_{ij}} y_{ijk}/m_{ij}$ = unbiased estimate of $\bar{Y}_{ij}$,

$w_{ijk}$ = weight of the $k^{th}$ cluster from the $j^{th}$ sort of the $i^{th}$ trip,

$\bar{w} = \sum_i \sum_j \sum_k w_{ijk}/\sum_i \sum_j m_{ij}$

= average weight of all clusters sampled,

= constant (say),

$\hat{M}_i = \dfrac{W_i}{\bar{w}}$ = estimate of $M_i$ .

5.2. Estimates of Mean, Total and Rrrors

$$\bar{\bar{y}}_R \doteq \sum_i^n W_i \bar{y}_i \Big/ \sum_i^n W_i = \text{ratio estimate of mean } \bar{\bar{Y}} \text{ catch per cluster} \tag{1}$$

where $\bar{y}_i = \sum_j M_{ij} \bar{y}_{ij} \Big/ \sum_j M_{ij}$

$$\approx \sum_j^n W_{ij} \bar{y}_{ij} \Big/ \sum_j^n W_{ij}$$

and

$$\hat{Y}_R \doteq \frac{W}{\bar{w}} \bar{\bar{y}}_R = \text{ratio estimate of total catch } Y \tag{2}$$

The above estimates are biased and are based on the assumption that the average weight per cluster is a constant. The estimates are consistent and more efficient than the corresponding unbiased estimates (as would be seen later) and take into account the size of the landing. These satisfy the important requirement that trips larger in size should receive higher weights at the estimation stage. In view of the lack of information (in advance) on the size of the landing, the approach to sample trips proportionate to trip size is not practicable.

Approximate estimates of $V(\hat{Y}_R)$ and $V(\bar{\bar{y}}_R)$ are respectively given by

$$v(\hat{Y}_R) \doteq \underbrace{\frac{N(N-n)}{n(n-1)} \sum^n M_i^2 (\bar{y}_i - \bar{\bar{y}}_R)^2}_{\text{Between}} + \underbrace{\frac{N}{n} \sum_i^n M_i (M_i - m_i) \frac{s_{2i}^2}{m_i}}_{\text{Within}} \tag{3}$$

$$v(\bar{\bar{y}}_R) \doteq (\frac{\bar{w}}{\bar{W}})^2 \, v(\hat{Y}_R) \quad (4)$$

In practice, both N and $M_i$ are not known. These are estimated by

$$\hat{N} = \frac{nW}{\sum^{n} W_i} = \frac{W}{\hat{\bar{W}}}$$

$$\hat{M}_i = W_i/\bar{w} \quad \text{and} \quad (4a)$$

$$s^2_{2i} = \sum_j \sum_k (y_{ijk} - \bar{y}_{ij})^2 / \sum_j (m_{ij} - 1)$$

$\hat{N}$ will be subject to high errors, when $W_i$'s are highly variable and n is small. $\hat{N}$ should be replaced by the true values N, wherever these are available after the season.

The above estimators which were recommended for use in the current design are, however, not workable since these pose serious operational problems in data collection from all the sort types within sample landings as was obvious from an examination of basic data for 1982 available with the California Fish and Game. Data for ten sample landings for Eureka during January 1, 1982 to September 10, 1982 are reproduced to illustrate the point (Table 1).

The sampler failed in all cases to sub-sample from more than one category in the samples (trips) where the landing weight from a boat trip comprised more than one category. The reasons for failure to collect the data were discussed in Section 4.

TABLE 1. Distribution of Landing Weights From All Categories and From the Sampled Category for Eureka for 1982

| Sample No. (Boat trip) | No. of Clusters Sampled ($m_i$) | Market Category Sampled | Landing Wt (in lbs) From All Categories ($W_i$) | Landing Wt (in lbs) for the Categories Sampled ($W_i'$) | $\sum_k w_{ik}$ (j=1) | $\bar{w}_i = (\sum_k \frac{w_{ik}}{k})$ |
|---|---|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| 1528 | 1 | 269 | 26,550 | 24,176 | 51 | 51 |
| 1529 | 1 | 250 | 4,133 | 445 | 44 | 44 |
| 1530 | 2 | 269 | 59,218 | 58,239 | 104 | 52 |
| 1531 | 1 | 269 | 20,511 | 15,987 | 51 | 51 |
| 1533 | 1 | 269 | 35,022 | 14,661 | 49 | 49 |
| 1534 | 1 | 269 | 20,757 | 20,705 | 54 | 54 |
| 1535 | 1 | 269 | 15,812 | 8,436 | 50 | 50 |
| 1536 | 1 | 250 | 1,975 | 1,010 | 52 | 52 |
| 1537 | 1 | 250 | 16,055 | 1,075 | 53 | 53 |
| 1541 | 3 | 269 | 65,837 | 65,837 | 145 | 48 |

5.3. Estimation Based on Random Categories

Assuming that the clusters sampled from a category from within a sampled trip are simple random samples from all possible clusters in the trip, more valid ratio estimates $\bar{\bar{y}}_{1R}$ of $\bar{\bar{Y}}$ and $\hat{Y}_{1R}$ of $Y$ are respectively given by

$$\bar{\bar{y}}_{1R} = \frac{\sum_{i}^{n} W_i \bar{y}_i}{\sum_{i}^{n} W_i} \; ; \qquad \hat{Y}_{1R} = (\frac{W}{\bar{w}})\bar{\bar{y}}_1 \tag{5}$$

where $\bar{y}_i = \frac{\sum_k \sum_j y_{ijk}}{\sum_j m_{ij}}$ and $W_i$ is the total landing weight from all categories for the $i^{th}$ boat trip in the sample $(W = \sum_i W_i)$. In practice, the sampler would tend to sample from a category which is preponderant and is accessible. Hence, the estimate may be slightly biased, though its contribution to the total error will be negligible, since this would occur at the second stage of sampling.

Equations (3) and (4) will be replaced by

$$v(\hat{Y}_{1R}) \doteq \left[\frac{1}{n}\left(1 - \frac{\sum^{n} W_i}{W}\right) s_b^2 + \frac{f_1(1-\bar{f}_2)}{n\bar{m}} s_w^2\right] (\frac{W}{\bar{W}})^2 \tag{6}$$

$$v(\bar{\bar{Y}}_{1R}) \doteq (\frac{\bar{w}}{W})^2 \, v(\hat{Y}_{1R}) \tag{7}$$

$$\text{where} \quad s_b^2 = \sum_{i}^{n} \left(\frac{W_i}{\hat{\bar{W}}}\right)^2 \frac{(\bar{y}_i - \bar{\bar{Y}}_{1R})^2}{n-1} \; ; \qquad s_w^2 = \sum^{n} \frac{\bar{m}}{n} \left(\frac{W_i}{\hat{\bar{W}}}\right)^2 \frac{s_{2i}^2}{m_i} \tag{8}$$

$$\text{and} \qquad s_{2i}^2 = \sum_{k}^{m_i} (y_{ik} - \bar{y}_i)^2 / (m_i - 1) \; ; \qquad \hat{\bar{W}} = \frac{\sum^{n} W_i}{n}$$

$$f_1 = \sum_i^n W_i/W \quad ; \quad \bar{f}_2 = \frac{\bar{w}\sum_i^n \frac{m_i}{W_i}}{n} \quad ; \quad \bar{w} = \sum_i^n \sum_k^{m_i} w_{ik} / \sum m_i \qquad (9)$$

$$\bar{m} = \sum^n m_i/n \; ;$$

$$W = \sum_{i=1}^N W_i = \text{Total of all rockfish landings in the port-month stratum.}$$

5.4. Estimation Based on Post-Stratification

We will now consider an operationally feasible plan based on post-stratification by sort groups. Thus, when a sampler sub-samples from a category from a boat trip at a port during a month, these can be classified into the categories into which the port-month group has been post-stratified. This method is almost as precise as proportional stratified sampling if within each port-month stratum (a) a minimum of four landings ($n_j \geq 4$) is selected for each category, and (b) the landing weights are available by sort groups after the season to serve as weights at the estimation stage. The procedure has four advantages over the one currently being used: (i) it would provide estimates by market categories for each port-month stratum, (ii) it would account for the bias in estimation due to missing categories, (iii) would enable sampling from another boat when sub-sampling from another category would preclude its selection, and (iv) finally it would be more efficient than the current method. The analysis of the data will show that is is preferable to sample as many boat trips as possible.

Recall that

$W_{ij}$ = pounds of sort j caught in trip i, in a port-month stratum,

$W_j = \sum_i W_{ij}$ = pounds of sort j caught in a port-month stratum.

5.4.1. Cluster Size

Assume as before (4.2)

$$w_{ijk} = \text{constant, i.e.,}$$

cluster weight for all trips, sorts within trips is a constant and is estimated by

$$\overline{w} = \sum_i \sum_j \sum_k w_{ijk} / \sum_i \sum_j m_{ij}$$

In practise, a 50 lbs. cluster should be within 50 ± 5 lbs. in weight i.e., individual cluster weights should be between 45 and 55 lbs. at the maximum and care should be taken to ensure that a cluster should be as close to 50 lbs. in weight as possible; for a 25 lbs. cluster in small fish category, e.g., rose fish, individual clusters should be between 23 and 27 lbs. at the maximum. For j = 1,2,...,L sorts we have ratio estimate of the mean per cluster and of total given by

$$\overline{\overline{y}}_{2R} = \sum_j^L W_j \overline{\overline{y}}_j / \sum_j^L W_j \; ; \quad \hat{Y}_{2R} = \sum_j^L \hat{Y}_j \tag{10}$$

where $$\overline{\overline{y}}_j = \sum_i^{n_j} W_{ij} \overline{y}_{ij} / \sum_i W_{ij} \tag{11}$$

$$\text{and}\quad \hat{Y}_j = \frac{\sum_i^{n_j} W_{ij}\bar{y}_{ij}}{\sum_i^{n} W_{ij}} \frac{W_j}{\bar{w}} \tag{12}$$

In a few cases, where 2 or more categories are sub-sampled from the same trip, it is reasonable to assume that the sub-samples from the categories are independent since the categories are likely to be different in their species composition and weights. If the assumption is incorrect, this may result in some bias in the estimates of error which will not be substantial.

Further

$$v(\bar{\bar{y}}_{2R}) = \sum_j \frac{W_j^2}{W^2} v(\bar{\bar{y}}_j) \tag{13}$$

and

$$v(\hat{Y}_{2R}) = \sum_j v(\hat{Y}_j) \tag{14}$$

where $v(\hat{Y}_j)$ and $v(\bar{\bar{y}}_j)$ can be obtained as in (3) and (4) by substituting $N_j$, $n_j$, $M_{ij}$, $\bar{y}_{ij}$, $\bar{\bar{y}}_j$, $s^2_{2ij}$, $m_{ij}$ and $W_j$ for $N$, $n$, $M_i$, $\bar{y}_i$, $\bar{\bar{y}}_R$, $s^2_{2i}$, $m_i$ and $W$ respectively. Since $n_j$ would occur in the denominator, $n_j > 0$. For efficient estimates $n_j \geq 4$ (see page 13).

Emphasis on "25 lbs. weight for small rock fish including widows and 50 lbs. for all other clusters" led in some cases to wide variability among samples (boat trips) resulting in violation of the assumption of equal cluster weight required in the current method of estimation of parameters. Formulas for the estimation of parameters and their errors have been developed for the general case of variable cluster weight in Section 5.5.2. Since exact formula for the estimates of error variance are rather complicated, approximate workable expressions are provided. However, exact expressions are provided in Section 5.5.3 for the more practical case when cluster weight can be treated as a constant within a sort group.

## 5.5. General Case: Unequal Cluster Size

We will consider an estimating procedure, when clusters of variable weights are used and it may not be possible to take two 50 lbs. cluster when sampling landings from sorts of small fish like rose fish or small widows either because of small landings or buyer not cooperative to let the sampler select more than one cluster.

### 5.5.1. Random Categories

Assume that the weight per cluster varies over boat trips but remains the same within a trip, i.e.,

$$\bar{w}_i = \sum_j \sum_k w_{ijk} / \sum_j m_{ij}$$

As discussed above, the cluster weight may vary more among sort groups but is approximately a constant within a sort group, so that in practise $w_j$ will be the weight of a cluster for the $j^{th}$ sort group.

Consider the two ratio estimators

$$\hat{Y}_{3R} = \frac{W}{\sum_i^n W_i} \left[ \sum_i^n \sum_k^{m_i} \frac{y_{ik}}{m_i} \frac{W_i}{w_{ik}} \right]$$

$$= \frac{W}{\sum_i^n W_i} \left[ \sum_i^n W_i \, \bar{r}_i \right] \qquad (15)$$

$$\text{where} \quad \bar{r}_i = \frac{1}{m_i} \sum_{k}^{m_i} \frac{y_{ik}}{w_{ik}} = \frac{1}{m_i} \sum_{k}^{m_i} r_{ik}$$

$$\hat{Y}_{4R} = \frac{W}{\sum_i^n W_i} \left[ \sum_i W_i \frac{\sum_k^{m_i} y_{ik}}{\sum_k w_{ik}} \right]$$

$$= W \left[ \frac{\sum^n W_i \hat{R}_i}{\sum^n W_i} \right] \qquad (16)$$

$$\text{where} \qquad \hat{R}_i = \frac{\sum^{m_i} y_{ik}}{\sum_k w_{ik}} = \frac{\bar{y}_i}{\bar{w}_i}$$

$$\text{and} \qquad \bar{\bar{y}}_{4R} = \frac{\sum^n W_i \hat{R}_i}{\sum^n W_i/\bar{w}_i} \qquad (17)$$

In this study we will use the estimator $\hat{Y}_{4R}$ since only a small sample is selected from each boat trip and also since $\hat{R}_i$ is likely to be more stable than $\bar{r}_i$ .

$v(\hat{Y}_{4R})$ and $v(\bar{\bar{y}}_{4R})$ can be obtained similar to (6) and (7) using (16) and (17) and noting $w_i$ varies over samples.

5.5.2. Post-Stratification into Sort Groups

Assume as before that the weight per cluster varies over trips, i.e., $w_i$ varies with i, as observed among cluster weights during 1977 and due to size sorting of widow and rock fish in small fish categories.

For any sort category (say $j^{th}$), let

$$\hat{R}_j = \sum_i^{n_j} \hat{R}_{ij} W_{ij} / \sum_i^{n_j} W_{ij} \tag{18}$$

where $$\hat{R}_{ij} = \sum_k^{m_{ij}} y_{ijk} / \sum_k^{m_{ij}} w_{ijk}$$

$$= \bar{y}_{ij} / \bar{w}_{ij} \tag{19}$$

= estimated number of fish (per pound)

of a species for the $j^{th}$ category from the $i^{th}$ boat in the sample and

$$\bar{w}_{ij} = \sum \frac{w_{ijk}}{m_{ij}} \; ; \quad \bar{y}_{ij} = \sum^{m_{ij}} y_{ijk}/m_{ij}$$

If $n_j$ is small compared to $N_j$ and if the same subsampling strategy is applied to each of the $n_j$ sampled landings, we have, ignoring contribution due to second-stage units,

$$v_1(\hat{R}_j) \doteq \frac{1}{n_j(n_j - 1)\bar{x}_j^2} \left[ \sum_{i=1}^{n_j} z_{ij}^2 - 2\hat{R}_j \sum_{i=1}^{n_j} z_{ij} x_{ij} + \hat{R}_j^2 \sum_i^{n_j} x_{ij}^2 \right] \tag{20}$$

approximately where $z_{ij} = \hat{R}_{ij} \cdot W_{ij}$ and $x_{ij} = W_{ij}$ . (20) reduces to

$$v_1(\hat{R}_j) \doteq \frac{1}{n_j(n_j - 1)} \sum_i^{n_j} \left( \frac{W_{ij}}{\hat{\bar{W}}_{ij}} \right)^2 (\hat{R}_{ij} - \hat{R}_j)^2 \tag{21}$$

Another estimator of $V(\hat{R}_j)$ is given by

$$v_2(\hat{R})_j \doteq \frac{n_j - 1}{n_j} \sum_i^{n_j} (\hat{R}'_{ij} - \hat{R}'_j)^2 \tag{22}$$

where

$$\hat{R}'_{ij} = \frac{\hat{R}_{1j}W_{1j} + \hat{R}_{2j}W_{2j} + \dots + \hat{R}_{(i-1)j}W_{(i-1)j} + \hat{R}_{(i+1)j}W_{(i+1)j} + \dots + \hat{R}_{nj}W_{nj}}{W_{1j} + W_{2j} + \dots + W_{(i-1)j} + W_{(i+1)j} + \dots + W_{nj}} \quad (23)$$

and $$\hat{R}'_j = \frac{1}{n_j} \sum_{i=1}^{n_j} R'_{ij} \quad (24)$$

Thus $\hat{R}'_{ij}$ is obtained by omitting trip i from the sample for sort j and calculating $\hat{R}'_{ij}$ instead of $\hat{R}_{ij}$ as in (19).

For estimating mean and total for a sort j for a species

$$\bar{\bar{y}}_j = \sum_i^{n_j} W_{ij}\hat{R}_{ij} / \sum_i^{n_j} W_{ij}/\bar{w}_{ij} \quad (25)$$

$$\hat{Y}_j = \left[ \sum_i^{n_j} W_{ij}\hat{R}_{ij} / \sum_i^{n} W_{ij} \right] W_j \quad (26)$$

Also, $$v(\hat{Y}_j) = W_j^2 v_1(\hat{R}_j) \quad (27)$$

or $$v(\hat{Y}_j) = W_j^2 v_2(\hat{R}_j)$$

where $v_1(\hat{R}_j)$ and $v_2(\hat{R}_j)$ are given by (21) and (22).

For estimates of total over all sort groups

$$\hat{Y}_{5R} = \sum_j^{L} \hat{Y}_j = \sum_i \sum_j \left[ W_{ij}\hat{R}_{ij} / \sum_i W_{ij} \right] W_j \quad (28)$$

$$v(\hat{Y}_{5R}) = \sum_j v(\hat{Y}_j) + 2 \sum_{j<k}\sum \text{cov}(\hat{Y}_j, \hat{Y}_k)$$

$$\doteq \sum_j v(\hat{Y}_j) \text{ approximately} \quad (29)$$

assuming that sub-samples from two or more categories from within a boat-trip sample are independent.

5.5.3. Cluster Weight Constant Within Sort Group

It is more reasonable to assume that cluster weight would be approximately a constant within a sort group (e.g., rose fish category) but would vary among sort groups. If so, the estimates of the cluster mean and total for a species are given by

$$\bar{\bar{y}}_{6R} = \sum_{j}^{L} W_j \bar{\bar{y}}_j \Big/ \sum_{j}^{L} W_j \; ; \qquad \hat{Y}_{6R} = \sum_{j}^{L} \hat{Y}_j \tag{30}$$

where

$$\bar{\bar{y}}_j = \frac{\sum_i W_{ij} \bar{y}_{ij}}{\sum_i W_{ij}} \tag{31}$$

and

$$\hat{Y}_j = \frac{\sum_i W_{ij} \bar{y}_{ij}}{\sum_i W_{ij}} \; \frac{W_j}{\bar{w}_j} \tag{32}$$

$L$ is the number of strata with $\bar{w}_j$ the sample mean weight of clusters in the $j^{th}$ group.

It follows that

$$v(\bar{\bar{y}}_{6R}) = \sum_j A_j^2 v(\bar{\bar{y}}_j) \tag{33}$$

and

$$v(\hat{Y}_{6R}) = \sum_j \left( \frac{W_j}{\bar{w}_j} \right)^2 v(\bar{\bar{y}}_j) \tag{34}$$

where

$$A_j = \frac{W_j}{\sum_j^L W_j}$$

### 5.5.4. Estimation of Ratio to Another Variable

Estimation of the ratio of number of fish in a age, sex or age-sex group to the total number for the species for a sort group can often be obtained as a ratio of two variables in a two-stage sampling procedure. Thus, if $\bar{z}_i$ and $\bar{y}_i$ represent respectively the mean of the number of fish in an age-sex and sex group respectively, we have

$$\hat{R} \doteq \frac{\sum^{n} W_i \bar{z}_i}{\sum^{n} W_i \bar{y}_i} \qquad (35)$$

where $R = Z/Y$, $z_{ij}$ is the number (or weight of a given species for a given age) in the $j^{th}$ cluster of the $i^{th}$ trip.

$$v(\hat{R}) \doteq \frac{(1-f_1)}{n} \sum_{i}^{n} \frac{n^2 W_i^2}{(\sum W_i \bar{y}_i)^2} \frac{(\bar{z}_i - \hat{R}\bar{y}_i)^2}{(n-1)} + \sum_{i}^{n} W_i \sum_{i=1}^{n} \left( \frac{W_i}{\sum_{i}^{n} W_i \bar{y}_i} \right)^2 s^2_{d_{2i}} (1-f_{2i}) \qquad (36)$$

and

$$v(\hat{z}_R) \doteq \frac{N^2(1 - \frac{n}{N})}{n} \sum_{i=1}^{n} \frac{M_i^2(\bar{z}_i - \hat{R}\hat{Y}_i)^2}{n-1} + \frac{N}{n} \sum_{i}^{n} \frac{M_i^2 s^2_{d_{2i}}}{m_i} (1-f_{2i}) \qquad (37)$$

where $s^2_{d_{2i}} = \sum_{j=1}^{m_i} \frac{(z_{ij} - \hat{R}y_{ij})^2}{m_i - 1}$, $f_{2i} \doteq \frac{m_i}{M_i}$; $f_1 = \frac{n}{N} = \frac{\sum^{n} W_i}{W}$ and $N$ and $M_i$ are given by (4a).

5.6. Comparison of Methods. Simple Random Sampling Versus Post Stratification

In this section we will compare the efficiencies of the estimators (5) based on random categories, i.e., clusters selected from all the sort groups from within a sampled trip with (10) based on post-stratification of a port-year stratum into sort groups.

The analysis will be based on 1982 data (January-September) for Eureka and Monterey for which total landing weights by categories were available from the California Fish and Game. The sample size by individual months was not large enough to provide estimates with any reasonable degree of accuracy.

The distribution of the samples (Table 2) by categories show that the number of landings are too few for certain categories. This number should be raised by either increasing the number of sample landings within each category for a port-month group or by decreasing the number of strata by suitably combining some of the categories or both.

The coefficient of variation (c.v.) of mean catch per cluster for the species based on the methods (1) random categories, i.e., ignoring stratification by categories and (2) using post-stratification by categories (Table 3) show that the estimates of mean catch of widow were estimated with less than 8 percent c.v. by method (2) at both Eureka and Monterey though the precision was considerably less using method (1). Again, using method (2) both Chilipepper and Bocaccio were estimated with higher precision (13.9 and 10.3 respectively) at Monterey though at Eureka the precision was somewhat lower. The estimated c.v. for Chilipepper at Eureka based on the same number of samples (40) as at Monterey was 21.55 which was much higher than the corresponding figure (13.92) at Monterey. The c.v.'s for short-spine thorny head not shown in the table were much lower by method (2).

TABLE 2. Distribution of Landing Weights (lbs.) in the Sample and Population by Categories During January 1 to September 30, 1982

Eureka

| Category | Mean Cluster wt | Landing wt in Sample ($W_i'$) | Landing wt in Population (W) | Number in Sample (Boat trips) |
|---|---|---|---|---|
| 250 | 49.11 | 154,777 | 2,821,222 | 59 |
| 262 | 50.50 | 4,390 | 1,903,258 | 2 |
| 269 | 51.49 | 694,924 | 5,519,313 | 21 |
| 270 | 27.83 | 1,249 | 57,895 | 6 |
| Total | | 855,340 | 10,301,688 | 88 |

Monterey

| Category | Mean Cluster wt | Landing wt in Sample ($W_i'$) | Landing wt in Population (W) | Number in Sample (Boat trips) |
|---|---|---|---|---|
| 250 | 47.71 | 241,696 | 1,150,164 | 29 |
| 253 | 47.96 | 177,165 | 1,607,824 | 16 |
| 262 | 50.00 | 3,655 | 144,177 | 2 |
| 269 | 50.80 | 395,137 | 1,352,110 | 6 |
| 270 | 32.50 | 10,330 | 104,070 | 1 |
| Total | | 827,983 | 4,358,345 | 54 |

TABLE 3. Coefficient of Variation (in Percent) of Mean Catch by Species for Eureka and Monterey based on the two Methods During January 1 to September 30, 1982

Eureka

| Species | Sample Size | Coefficient of Variation (%) | |
|---|---|---|---|
| | | Method 1* | Method 2** |
| Widow (2316) | (88) | 11.48 | 7.33 |
| Chilipepper (2320) | (88) | 30.83 | 32.12 |
| Bocaccio (2334) | (88) | 26.01 | 24.40 |

Monterey

| Species | Sample Size | Coefficient of Variation (%) | |
|---|---|---|---|
| | | Method 1* | Method 2** |
| Widow (2316) | (54) | 18.31 | 6.62 |
| Chilipepper (2320) | (54) | 15.68 | 13.92 |
| Bocaccio (2334) | (54) | 12.57 | 10.32 |

*Method 1. Based on random categories.

**Method 2. Based on post-stratification by categories.

Thus the estimates of catch of a species based on post-stratification by market categories were generally more efficient than those based by ignoring such stratification.

The coefficient of variation for the species at Eureka and Monterey during 1982 by sex-age groups for which the number of sample landings was greater than or equal to ten (Table 4) show that in almost all cases method (2) is more precise than method (1) for estimating the number of fish of a species in a particular sex-age group.

In summary, method (2) based on post-stratification by sort groups is preferred to method (1) for estimating the number of a species based on random categories; also, method (2) is recommended for estimating the number of a species in a sex-age group.

## 6. OPTIMUM SAMPLING AND SUB-SAMPLING FRACTIONS

The instructions on the number and distribution of clusters to be selected from sample landings in the current plan are not specific enough in the absence of information on the between and within sample component of variation in species number, weight for obtaining the optimum allocation for a given cost. In the following sections, we will obtain the cost function and the optimum number of clusters for a given cost using both the survey data from sample landings during 1978 (when four clusters were mostly available per sample) and data based on an experimental study undertaken during 1983 at some of the ports.

TABLE 4. Coefficient of Variation (in Percent) of Mean Catch per Cluster by Species for Important* Age-Sex Groups based on the two Methods for Eureka and Monterey During 1982 (January 1 - September 30)

| Eureka | | | | Monterey | | | |
|---|---|---|---|---|---|---|---|
| Widow (2316) | | | | Widow (2316) | | | |
| Age (years) | Sex | C.V. (%) Method 1 | Method 2 | Age (years) | Sex | C.V. (%) Method 1 | Method |
| 12 | M | 15.49 | 14.84 | | | | |
| 7 | M | 19.71 | 18.83 | | | | |
| 16 | F | 15.82 | 16.08 | 13 | F | 39.98 | 24.29 |
| 12 | F | 15.14 | 14.10 | 12 | F | 35.16 | 20.49 |
| 7 | F | 13.50 | 10.94 | 9 | F | 31.22 | 27.02 |
| Chilipepper (2320) | | | | Chilipepper (2320) | | | |
| Age (years) | Sex | C.V. (%) Method 1 | Method 2 | Age (years) | Sex | C.V. (%) Method 1 | Method |
| 13 | F | 39.98 | 24.89 | 9 | F | 18.48 | 7.63 |
| 12 | F | 34.77 | 31.21 | 8 | F | 27.84 | 9.00 |
| | | | | 7 | F | 22.09 | 9.81 |
| | | | | 6 | F | 25.67 | 8.86 |

*Age-sex groups for which primary sampling units (landings) are ≥ 10

TABLE 4. (Continued) Coefficient of Variation (in Percent) of Mean Catch Per Cluster by Species for Important Age-Sex Groups Based on the Two Methods for Eureka and Monterey During 1982 (Jan. 1 - Sept. 30)

| Eureka Bocaccio | | | | Monterey Bocaccio | | | |
|---|---|---|---|---|---|---|---|
| Age (years) | Sex | C.V. (%) Method 1 | C.V. (%) Method 2 | Age (years) | Sex | C.V. (%) Method 1 | C.V. (%) Method 2 |
| 6 | M | 30.10 | 19.82 | 7 | M | 27.46 | 12.45 |
| 7 | M | 30.74 | 31.91 | 6 | M | 23.23 | 9.75 |
| 4 | M | 39.22 | 41.71 | 5 | M | 23.27 | 14.38 |
| 6 | F | 35.87 | 32.45 | 4 | M | 25.35 | 10.65 |
| 5 | F | 22.65 | 22.80 | 7 | F | 24.34 | 10.06 |
| | | | | 6 | F | 23.98 | 8.03 |
| | | | | 5 | F | 22.32 | 9.65 |
| | | | | 4 | F | 27.23 | 12.34 |

## 6.1. Cost Function

From equations (6) and (7) we have

$$v(\bar{\bar{y}}_{1R}) \doteq \frac{(1-f_1)}{n} \sum^{n} \frac{W_i^2}{(\hat{\bar{W}})^2} \frac{(\bar{y}_i - \bar{\bar{y}}_{1R})^2}{n-1} + \frac{(1-\bar{f}_2)}{n\bar{m}} \sum_i^n \frac{\bar{m}}{n} \left(\frac{W_i^2}{\hat{\bar{W}}^2}\right) \frac{s_{2i}^2}{m_i} \qquad (38)$$

where $\hat{\bar{W}} = \frac{\sum^n W_i}{n}$ ; $f_1' = \frac{\sum^n W_i}{W}$ ; $\bar{f}_2 = \frac{\bar{w}}{n} \sum_i^n \frac{m_i}{W_i}$ and $\bar{m} = \frac{\sum_i^n m_i}{n}$ .

Putting

$$\sum_{i}^{n} \left(\frac{W_i}{\hat{\bar{W}}}\right)^2 \frac{(\bar{y}_i - \bar{\bar{y}}_{1R})^2}{n-1} = s_b^2 \tag{39}$$

and

$$\sum_{i}^{n} \frac{\bar{m}}{n} \left(\frac{W_i}{\hat{\bar{W}}}\right)^2 \frac{s_{2i}^2}{m_i} = s_w^2 \tag{40}$$

we have

$$\bar{m}_{opt} = \frac{s_w}{\sqrt{s_b^2 - \frac{s_w^2}{\bar{\bar{m}}}}} \sqrt{\frac{c_1}{c_2}} \tag{41}$$

for the simple cost function (ignoring travel costs between sample landings)

$$C = c_1 n + c_2 n\bar{m} \tag{42}$$

where $c_1$ is the average cost (in minutes) per boat trip due to transport, contact and delay in making a contact, $c_2$ is the average cost (in minutes) of data collection (e.g., identification of species, sex, length, otoliths, etc.) per cluster and $C$ is the total cost involved in visiting the primary sampling units (boat trips) and collecting data from the $n$ boats with an average of $\bar{m}$ clusters per boat sampled. From data collected at Tiburon as a result of the cooperative program between California Dept. of Fish and Game and the National Marine Fisheries Service we have

$$c_1 = 111.80 \text{ minutes}, \quad c_2 = 58.3 \text{ minutes}$$

so that $\frac{c_1}{c_2} = 1.9177 = 2$ approximately. However, from some more recent data collected we have approximately

$$c_1/c_2 = 3$$

Actually, the components of $c_1$ and $c_2$ were estimated at

| Activity | Percent | Mean (in minutes) |
|---|---|---|
| Transport | 50 | 81.7 |
| Contact | 5 | 8.7 |
| Delay (off-loading, etc.) | 13 | 21.4 |
| | 68 | 111.8 minutes |

| Data Collection | Percent | Mean (in minutes) |
|---|---|---|
| Species* | 7.7 | 14.0 |
| Sex, length | 5.8 | 10.6 |
| Otolith | 10.8 | 19.7 |
| Preparation Time | 7.7 | 14.0 |
| | 32.0 | 58.3 |

*excluding samples dominated by single species.

## 6.2. Optimum Allocation - Survey Data

We will now estimate $m_{opt}$ at the different ports on the basis of survey data collected by California Fish and Game department based on its cooperative program with the National Marine Fisheries Service for estimating the mean or total catch of a species at a port during a season.

It would be seen (Tables 5 and 6) that the optimum number of clusters per boat in the sample for estimating species number or species weight was unity in almost all cases. Since a minimum of two clusters is needed to provide an estimate of the component of cluster variation required for obtaining an estimate of error, it is recommended that two clusters per primary sampling unit (boat trip) be randomly selected. In practice, it is preferable to select a systematic sample of clusters separated in time. This will provide an unbiased estimate of $V(\bar{\bar{y}}_R)$, if $f_1$ is small. However, if $f_1$ is not negligible, the estimate of the error will be slightly biased.

6.3. Optimum allocation: Experimental data

We will now consider the case where data from an experimental study using 4 clusters per sample were analysed to find the relative contribution of the between and within cluster variation to throw further light on the problem.

Four clusters each of approximately 50 lbs. in weight (in a few cases when the total fish caught was small in weight, clusters of 25 lbs. were selected) were selected at random from a sort group of a sample landing (boat trip) on different landing days of June and July 1983 at Morro Bay, Monterey, Moss Landing, San Francisco, Fort Bragg and Eureka. Details of the method of random sampling are given in the section on data collection. A cluster is next separated by number of each species and its weight which are recorded along with sex.

The proportion of the number of each species to the total for all species in a cluster from a sample landing was next analysed using the arc sine transformation. Where the total number $n$ in a cluster was $< 50$, a zero proportion was counted as $1/4n$ and a 100% proportion as

TABLE 5. Optimum Values of $m$ for Estimating Species Number per Cluster by Sort Groups for Different Variance and Cost Ratios

Eureka: 1978

| Species (number) | Sort | n | $s_b^2$ | $s_w^2$ | $\bar{m}$ | $m_{opt}$ $c_1/c_2 = 2$ | $m_{opt}$ $c_1/c_2 = 3$ |
|---|---|---|---|---|---|---|---|
| Bocaccio | 250 | 25 | 1.80 | 3.01 | 2.16 | 3.86 | 4.73 |
| Chilipepper | 250 | 13 | 24.45 | 3.13 | 1.92 | 0.52 | 0.64 |
| Widow | 250 | 11 | 59.49 | 8.71 | 2.46 | 0.56 | 0.68 |

Monterey: 1978

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Bocaccio | 253 | 31 | 95.15 | 4.20 | 1.97 | 0.63 | 0.77 |
| Chilipepper | 253 | 33 | 43.71 | 4.16 | 1.94 | 0.45 | 0.55 |
| Widow | 253 | 12 | 22.38 | 4.66 | 2.00 | 0.68 | 0.84 |

San Francisco: 1978

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Bocaccio | 253 | 20 | 17.99 | 5.74 | 2.30 | 0.86 | 1.05 |
| Chilipepper | 253 | 15 | 10.79 | 14.70 | 2.65 | 2.31 | 2.82 |

Fort Bragg: 1978

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Bocaccio | 253 | 86 | 14.81 | 2.37 | 1.26 | 0.61 | 0.74 |

TABLE 6. Optimum Values of m for Estimating Species Weight per Cluster by Sort Groups for Different Variance and Cost Ratios

Eureka: 1978

| Species (weight) | Sort | n | $s_b^2$ | $s_w^2$ | $\bar{m}$ | $m_{opt}$ $c_1/c_2 = 2$ | $m_{opt}$ $c_1/c_2 = 3$ |
|---|---|---|---|---|---|---|---|
| Chilipepper | 250 | 12 | 181.64 | 18.67 | 2 | 0.47 | 0.57 |
| Widow | 250 | 11 | 231.15 | 48.09 | 2.4 | 0.67 | 0.83 |

Monterey: 1978

| Species (weight) | Sort | n | $s_b^2$ | $s_w^2$ | $\bar{m}$ | $m_{opt}$ $c_1/c_2 = 2$ | $m_{opt}$ $c_1/c_2 = 3$ |
|---|---|---|---|---|---|---|---|
| Bocaccio | 253 | 36 | 63.17 | 34.56 | 2 | 1.23 | 1.50 |
| Chilipepper | 253 | 33 | 86.45 | 21.31 | 1.94 | 0.75 | 0.92 |
| Widow | 253 | 12 | 27.07 | 37.08 | 2 | 2.95 | 3.61 |

San Francisco: 1978

| Species (weight) | Sort | n | $s_b^2$ | $s_w^2$ | $\bar{m}$ | $m_{opt}$ $c_1/c_2 = 2$ | $m_{opt}$ $c_1/c_2 = 3$ |
|---|---|---|---|---|---|---|---|
| Bocaccio | 253 | 20 | 117.27 | 62.43 | 2.3 | 1.18 | 1.44 |

$(n-\frac{1}{4})/n$ before transforming to angles. This transformation was done to improve the equality of variance in the angles. The F values based on the analysis of variance of the angles at the ports for the species Chilipepper, Widow and Bocaccio are shown in Table 7.

The F values for variation due to samples are large in almost all the cases though those due to the variation between clusters within samples are much too small (assuming that all the within-cluster variation is binomial). This suggests that we should have more samples and fewer clusters per sample to estimate mean number of a species per cluster with a high degree of precision. Thus, the optimum value of the number of clusters for Chilipepper at Morro Bay will be given by

$$\frac{s_w}{\sqrt{s_b^2 - (s_w^2/\bar{m})}} \sqrt{c_1/c_2}$$

assuming $c = c_1 n + c_2 \bar{m} n$ .

In this case from the analysis of variance table we have

$$s_w = \sqrt{54} = 7.3485$$

Hence

$$m_{opt} = \frac{7.3485}{17.3133} \sqrt{2} \quad \text{if} \quad c_1/c_2 = 2$$

$$= 0.6002$$

$$= 1 \quad \text{approximately.}$$

In practice, we should take $m_{opt} = 2$ to provide estimate of within component of total variance.

TABLE 7. Analysis of Variance in Angles

Morro Bay 1983

| Source | | Chilipepper | | | Widow | | | Bocaccio | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | DF | MS | F | P | MS | F | P | MS | F | P |
| Samples | 4 | 1307 | 24.2 | <0.001 | 72 | 10.3 | <0.001 | 444 | 14.21 | <0.01 |
| Clusters (within samples) | 14 | 54 | | | 7 | | | 31 | | |
| Within Clusters | 330 | 101 | | | 63 | | | 67 | | |

Monterey 1983

| Source | | Chilipepper | | | Widow | | | Bocaccio | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | DF | MS | F | P | MS | F | P | MS | F | P |
| Samples | 4 | 1308 | 8.9 | <0.001 | 449 | 15.2 | <0.001 | 1671 | 12.6 | <0.001 |
| Clusters (within samples) | 15 | 147 | | | 30 | | | 132 | 1.7 | <0.05 |
| Within Clusters | 437 | 82 | | | 61 | | | 80 | | |

San Francisco 1983

| Source | | Chilipepper | | | Widow | | | Bocaccio | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | DF | MS | F | P | MS | F | P | MS | F | P |
| Samples | 4 | 254 | 7.7 | <0.005 | 5.6 | | | 233 | | |
| Clusters (within samples) | 15 | 33 | | | 5.7 | | | 118 | | |
| Within Clusters | 259 | 160 | | | 68 | | | 159 | | |

TABLE 7. (Continued) Analysis of Variance in Angles

Fort Bragg 1983

| Source | DF | Chilipepper MS | Chilipepper F | Chilipepper P | Widow MS | Widow F | Widow P | Bocaccio MS | Bocaccio F | Bocaccio P |
|---|---|---|---|---|---|---|---|---|---|---|
| Samples | 8 | 1512 | 9.1 | <0.001 | 55 | 2.9 | <0.025 | 1416 | 6.3 | <0.001 |
| Clusters (within samples) | 24 | 165 | 1.8 | <0.05 | 18 | | | 223 | 2.3 | <0.01 |
| Within Clusters | 606 | 93 | | | 40 | | | 95 | | |

Eureka 1983

| Source | DF | Chilipepper MS | Chilipepper F | Chilipepper P | Widow MS | Widow F | Widow P | Bocaccio MS | Bocaccio F | Bocaccio P |
|---|---|---|---|---|---|---|---|---|---|---|
| Samples | 5 | 556 | 30.7 | <0.001 | 124 | 11.9 | <0.001 | 590 | 12.0 | <0.001 |
| Clusters (within samples) | 13 | 18 | | | 10 | | | 49 | | |
| Within Clusters | 230 | 338 | | | 82 | | | 167 | | |

## 6.4. Variance components: Species - age - sex and length groups

We will now obtain the relative contribution of variation in length/age of fish due to sample landings, clusters within landings to within-cluster variation for a given species within a sort group. This will, incidentally, throw light on the extent to which sorting by length and or age is done between and within samples for a given market category.

It would be seen (Tables 8(a) to 8(i)) that the variation due to length and age was generally high among sample landings compared to that within samples. For widow most of the variation was due to sample landings at the centres (for which adequate data was available); also, the variation between clusters was small and was of the same order as that within clusters. It would be seen that the optimum number of clusters was $\leq 2$.

Further, most of the significant variation for the species Bocaccio/ Chilipepper was due to samples and was uniformly high for length and age; in a few cases, however, (e.g., Chilipepper at San Francisco 1979, Monterey 1979; Bocaccio at Monterey 1979) the variation between clusters was significant but not consistent.

The variation in length was generally high among sample landings than in age for the species under study.

On the whole, both the variation in species as well as in length-age for a species was high among sample landings relative to that between clusters within landings. Hence, for precise estimation of species number, it is recommended that data be collected from a large number of sample landings and from few clusters within each of the landings.

To summarize the optimum plan for efficient estimates of number and weight of a species is to have two clusters of 25 lbs. each from sample landings in 'small fish' category (e.g., rose fish) and two of 50 lbs. each

TABLE 8. Two-Level Nested ANOVA of Length/Age for a Species with Unequal Sample Sizes by Ports and Years

(a) Analysis of Variance (ANOVA) for Widows at Bodega Bay, 1981.

| | | AGE | | | LENGTH | | |
|---|---|---|---|---|---|---|---|
| Source | DF | MS | F | P | MS | F | P |
| Samples | 22 | 87.24 | 8.724 | <0.001 | 197.72 | 4.700 | <0.005 |
| Clusters (within samples) | 12 | 10.00 | 1.197 | $\sim$0.25 | 42.07 | 1.314 | $\sim$0.20 |
| Within Clusters | 636 | 8.352 | | | 32.01 | | |

(b) Analysis of Variance (ANOVA) for Chilipepper at Fort Bragg, 1979.

| | | AGE | | | LENGTH | | |
|---|---|---|---|---|---|---|---|
| Source | DF | MS | F | P | MS | F | P |
| Samples | 37 | 44.02 | 3.12 | $\sim$0.008 | 172.13 | 2.10 | $\sim$0.05 |
| Clusters (within samples) | 11 | 14.12 | 1.80 | | 81.78 | 2.36 | $\sim$0.01 |
| Within Clusters | 508 | 7.85 | | | 34.65 | | |

(c) Analysis of Variance (ANOVA) for Widow at Eureka, 1979.

| | | AGE | | | LENGTH | | |
|---|---|---|---|---|---|---|---|
| Source | DF | MS | F | P | MS | F | P |
| Samples | 15 | 34.45 | 4.75 | <0.005 | 37.86 | 3.09 | <0.025 |
| Clusters (within samples) | 13 | 7.25 | 1.19 | 0.35 | 12.27 | 1.43 | $\sim$0.18 |
| Within Clusters | 320 | 6.09 | | | 8.58 | | |

Only variation between samples among lengths was significant at Eureka during 1978.

TABLE 8. (Cont.) Two-Level Nested ANOVA of Length/Age for a Species with Unequal Sample Size by Ports and Years

(d) ANOVA for Chilipepper at San Francisco, 1979.

| | AGE | | | | LENGTH | | | |
|---|---|---|---|---|---|---|---|---|
| Source | DF | MS | F | P | DF | MS | F | P |
| Samples | 9 | 23.92 | 1.46 | ∿0.10 | 9 | 136.17 | 2.82 | ∿0.025 |
| Clusters | 15 | 16.36 | 2.29 | <0.01 | 16 | 48.22 | 1.78 | ∿0.05 |
| Within Clusters | 185 | 7.14 | | | 214 | 27.08 | | |

(e) ANOVA for Chilipepper at Monterey, 1978.

| | AGE | | | | LENGTH | | | |
|---|---|---|---|---|---|---|---|---|
| Source | DF | MS | F | P | DF | MS | F | P |
| Samples | 38 | 32.10 | 8.04 | <0.001 | 42 | 221.42 | 7.21 | <0.001 |
| Clusters | 29 | 3.99 | 1.16 | ∿0.25 | 33 | 30.71 | 1.28 | ∿0.15 |
| Within Clusters | 536 | 3.43 | | | 617 | 24.07 | | |

(f) ANOVA for Chilipepper at Monterey, 1979.

| | AGE | | | | LENGTH | | | |
|---|---|---|---|---|---|---|---|---|
| Source | DF | MS | F | P | DF | MS | F | P |
| Samples | 43 | 31.74 | 4.05 | <0.001 | 48 | 145.20 | 4.02 | <0.001 |
| Clusters | 39 | 7.84 | 1.80 | ∿0.001 | 44 | 36.10 | 1.43 | ∿0.035 |
| Within Clusters | 859 | 4.35 | | | 971 | 25.25 | | |

TABLE 8. (Cont.) Two-Level Nested ANOVA of Length/Age for a Species with Unequal Sample Size by Ports and Years

(g) ANOVA for Bocaccio at San Francisco, 1979.

| | AGE | | | | LENGTH | | | |
|---|---|---|---|---|---|---|---|---|
| Source | DF | MS | F | P | DF | MS | F | P |
| Samples | 10 | 84.97 | 6.95 | <0.001 | 10 | 317.88 | 6.98 | <0.001 |
| Clusters | 15 | 12.23 | 1.20 | ∿0.30 | 16 | 45.55 | 0.80 | ∿0.75 |
| Within Clusters | 225 | 10.20 | | | 227 | 57.11 | | |

(h) ANOVA for Bocaccio at Monterey, 1978.

| | AGE | | | | LENGTH | | | |
|---|---|---|---|---|---|---|---|---|
| Source | DF | MS | F | P | DF | MS | F | P |
| Samples | 37 | 49.17 | 14.32 | <0.001 | 41 | 1167.33 | 14.43 | <0.001 |
| Clusters | 33 | 3.43 | 1.33 | ∿0.10 | 37 | 80.91 | 1.38 | ∿0.06 |
| Within Clusters | 579 | 2.59 | | | 645 | 58.79 | | |

(i) ANOVA for Bocaccio at Monterey, 1979.

| | AGE | | | | LENGTH | | | |
|---|---|---|---|---|---|---|---|---|
| Source | DF | MS | F | P | DF | MS | F | P |
| Samples | 45 | 59.07 | 9.36 | <0.001 | 50 | 918.70 | 11.74 | <0.001 |
| Clusters | 41 | 6.31 | 1.48 | ∿0.025 | 46 | 78.27 | 1.25 | ∿0.01 |
| Within Clusters | 899 | 4.25 | | | 1015 | 62.55 | | |

in all other categories; for small landings of rockfish like widows, one or two clusters of 50 lbs. each should be taken depending upon availability. For a boat landing more than one important category, the sampler should subsample from as many of such categories as possible; additional sort groups should not be sampled at the cost of gathering less than 2 clusters from a sort group. If there is more than one landing on a given day, attempt should be made to cover the other landing after completing the requirements of the first landing.

If a sort group is infrequently landed, sampling should be directed towards the infrequent sort group, as long as the number of landings for the sort group is less than four per month.

7. RELATIVE EFFICIENCY OF ESTIMATORS USING POST-STRATIFICATION

We will now compare the efficiency of three important estimators using different estimation methods but the same selection procedure with post-stratification by sort groups.

For any sort group, the unweighted mean of the unit means is given by

$$\bar{\bar{y}}' = \frac{1}{n} \sum_{i=1}^{n} \bar{y}_i \qquad (43)$$

where $\bar{y}_i$ is the simple mean of the species number (or weight) per cluster for the sort from the $i^{th}$ sample landing and n is the number of landings at a port-year stratum. Consider the ratio estimator $\bar{\bar{y}}_{R_1}$ for a sort group given by

$$\bar{\bar{y}}_{R_1} = \sum_i^n W_i \bar{y}_i / \sum_i^n W_i \tag{44}$$

which is the same as (31).

Consider a more general estimator (ratio) than (44)

$$\bar{\bar{y}}_{R_2} = \hat{R}\,\bar{w} \tag{45}$$

where $\hat{R}$ is given by (18) and $\bar{w}$ is an estimate of average weight per cluster.

It would be seen that all the above estimators are biased. The first estimator is simpler to work out but may have large bias. The other two are ratio estimators.

The unbiased estimator

$$\bar{\bar{y}}_i = \frac{N}{nM} \sum_{i=1}^{n} M_i \cdot \bar{y}_i \tag{46}$$

cannot be used since $M_i$ and $N$ are unknown. The estimate of $v(\bar{\bar{y}}')$ is given by

$$v(\bar{\bar{y}}') = \underbrace{\left(\frac{1}{n} - \frac{\hat{\bar{W}}}{\bar{W}}\right) s_b^2}_{\text{Between}} + \underbrace{\frac{1}{n}\frac{\hat{\bar{W}}}{\bar{W}} \sum_i^n \left(\frac{1}{m_i} - \frac{\bar{W}}{\hat{\bar{W}}}\right) s_{2i}^2}_{\text{Within}}$$

$$+ \underbrace{\left[\frac{\left(1 - \frac{\hat{\bar{W}}}{\bar{W}}\right)}{\hat{\bar{W}}} \sum^n \frac{(W_i - \hat{\bar{W}})(\bar{y}_i - \bar{\bar{y}})}{n-1}\right]^2}_{\text{(Bias)}} \tag{47}$$

Also

$$v(\bar{\bar{y}}_{R_1}) \doteq \frac{1}{n}(1-f_1)s_b^2 + f_1 \frac{(1-f_2)}{n\bar{m}} s_w^2 \tag{48}$$

where $f_1 = \frac{\sum^n W_i}{W}$ , $f_2 = \bar{w} \sum_i^n \left(\frac{m_i}{W_i}\right)/n$ , $\bar{w} = \sum_i^n \sum_R^{m_i} w_{iR} / \sum^n m_i$

$$v(\bar{\bar{y}}_{R_2}) \doteq (1-f_1)v_2(\hat{R})w^2 \tag{49}$$

where $v_2(\hat{R})$ is given by (22).

Formula (49) is easy to work out but is approximate and the approximation is close if $f_1$ is small.

The coefficient of variation of mean number of Bocaccio, Chilipepper and widow per cluster for different categories by port-year groups and the three estimators, $\bar{\bar{y}}'$, $\bar{\bar{y}}_{R_1}$ and $\bar{\bar{y}}_{R_2}$ (Table 9) show that the estimators (44) and (45) are highly efficient. In all cases, the contribution of bias to total variance of (43) was considerably high. Further, estimator (44) is somewhat more efficient than (45), taking into account the fact that $v(\bar{\bar{y}}_{R_2})$ is likely to be under estimate since it does not take into account the within component of variance. The same trend of results was obtained for c.v.'s of mean weight per cluster for the species.

The empirical evidence supports strongly the use of the estimator $\bar{\bar{y}}_{R_1}$ against $\bar{\bar{y}}'$ and $\bar{\bar{y}}_{R_2}$. We will in the following section present some of the coefficient of variations for estimating (a) number of a species (b) number of a species in a particular age-sex group and

TABLE 9. Coefficient of Variation (in Percent) of Mean Number of Bocaccio, Chilipepper and Widow per Cluster by Port-Year Group and for Different Categories for the Three Estimators

<u>Bocaccio</u>

| Port (1) | Year (2) | Category (3) | Sample Size (4) | $\bar{\bar{y}}_{R_1}$ (5) | $\bar{\bar{y}}'$ (6) | $\bar{\bar{y}}_{R_2}$ (7) |
|---|---|---|---|---|---|---|
| San Francisco | 1978 | 253 | 20 | 10.24 | 13.51 | 11.64 |
| Fort Bragg | 1978 | 250 | 86 | 7.36 | 16.21 | 8.14 |
| | 1979 | 250 | 46 | 25.70 | 47.73 | 27.10 |
| Monterey | 1978 | 253 | 31 | 17.93 | 12.03 | 19.51 |
| | 1979 | 253 | 51 | 7.36 | 8.95 | 9.53 |
| Eureka | 1978 | 250 | 25 | 26.00 | 40.11 | 29.84 |
| | 1982 | 250 | 59 | 25.24 | 39.22 | 28.03 |

<u>Chilipepper</u>

| | | | | | | |
|---|---|---|---|---|---|---|
| Eureka | 1978 | 250 | 13 | 34.52 | 37.66 | 42.33 |

<u>Widow</u>

| | | | | | | |
|---|---|---|---|---|---|---|
| Monterey | 1978 | 250 | 12 | 43.47 | 111.20 | 68.29 |
| Eureka | 1978 | 250 | 11 | 27.81 | 72.69 | 33.90 |

(c) weight of a species per cluster by port-year groups; the results will also be presented for all ports combined for 1981 (for which data were available) since this has important management implications.

## 8. RELIABILITY OF THE ESTIMATES: COEFFICIENT OF VARIATION

### 8.1. Coefficient of Variation (c.v.) of Species Catch by Port-Year Groups

The components of variance, i.e., between and within sample landings and the coefficient of variation of the ratio estimate (10) of average number of a species per cluster

$$\bar{\bar{y}}_{2R} = \sum_{j}^{L} W_j \bar{\bar{y}}_j \Big/ \sum_{j}^{L} W_j$$

are presented in Tables 10 to 10.2 by Port-year groups. The estimates were obtained on the assumption that the cluster weight is a constant for all trips and sorts within trips for the species Widow, Chilipepper and Bocaccio. A more realistic assumption would be to use $\bar{\bar{y}}_{6R}$ and thus assume that cluster weight is a constant within a sort group but varies over sort groups. Although this change will not affect the over-all conclusions, it is recommended that $\bar{\bar{y}}_{6R}$ and $v(\bar{\bar{y}}_{6R})$ be used in such situations.

For 1982 representative data were available for only 2 important port groups Eureka and Monterey and for the period January to September. The variation between sample landings was invariably high in all cases and the variation within landings was mostly negligible suggesting that for obtaining efficient estimates of total number of a species we should take more landings and too few clusters from each landing.

During 1982 the c.v. of the number of Widows per cluster at Eureka summed over the sort groups was reasonably low (7.33) when 88 sample landings were used in the study. For the other two species Bocaccio and Chilipepper the precision was low though for Bocaccio the c.v. was lower (24.40) compared to Chilipepper (32.12, not shown in the table). This is because Bocaccio was more abundant being available in 45 of the 88 landings compared to only 18 landings for Chilipepper. For Yellowtail (results not shown), the c.v. was as high as 54 percent. Here again, only 18 out of the 88 sample landings contained the species. For some of the sort groups the sampled clusters from the landings did not contain any of the rockfish under study (e.g., the sample with the sort groups 262, 270 did not have any Widows). If this is known apriori with a high probability, estimation made by eliminating such sort groups and selecting more samples from other sort groups should result in increased over-all efficiency.

The estimated c.v.'s at Monterey (1982) were reasonably accurate, being respectively 6.62, 13.92 and 10.31 for Widow, Chilipepper and Bocaccio. Although the total number of sample landings was smaller (54) compared to that at Eureka (88), the species were more abundant at Monterey measured in terms of the proportion of landings containing the species.

During 1981 the c.v.'s varied considerably between species among port groups. The estimate of total catch for each species based on 232 landings for 'all ports' obtained by combining the port groups Monterey, San Francisco, Fort Bragg, Morro Bay, and Eureka was sufficient enough for management purposes (c.v.'s ranging between 7 to 11.6 percent). However, for Yellowtail (not shown in table) the overall c.v. was high (25%), presumably because only 31 out of the 232 landings had Yellowtail.

8.2. Coefficient of Variation of Species Weight by Port-Year Sort Groups

The coefficients of variation (c.v.) of mean weight per cluster during 1978 are given by sort groups (Table 10.3).

We have used $\bar{\bar{y}}_{6R}$ as the estimate of mean cluster weight within a sort group which is a reasonable assumption to make. The c.v.'s for estimating corresponding mean number of a species per cluster are shown in brackets in the last column of the table which shows that for the same sample size, the estimates of species weight are generally more efficient than the corresponding estimates for species number.

The estimated c.v.'s of weight are higher for Chilipepper and Bocaccio at Eureka than at Monterey which is due to Monterey being abundant in the species so that a larger proportion of the samples contained the species. The c.v. for Widow at Eureka based on 11 landings was rather high (24.79) during 1978 and is a suspect in view of the small sample size on which the estimate is based. However, during 1980 (based on 54 landings not shown in the table) it was as low as 2 percent and the corresponding figure for catch was 2.2 percent. This latter figure agrees with estimates during 1981 and 1982 as discussed in the preceding pages. San Francisco gave a higher precision (c.v. = 7.45) for Bocaccio than other centres.

8.3. Coefficient of Variation of Species by Sex-Age for Port-Year Groups

The c.v.'s along with the between, within and total component of variance of mean catch by sex-age for a species (1981-82), are shown in Tables 10.4 to 10.6. Sex-age groups for which c.v.'s were greater than 25% are not shown in the table.

In general, the estimates of cluster mean for a species by sex-age groups are less precise than for the species as a whole without such classification. Thus for Widow at Eureka and Monterey (1982), the c.v.'s were estimated at 7.33 and 6.62 respectively (Table 10), though the corresponding figures for a specific sex-age group (Table 10.4) were considerably higher. For Chilipepper and Bocaccio only a few of the sex-age groups showed lower c.v.'s (higher precision) than the corresponding figures for estimating the species number.

The coefficients of variation for estimating the number of Bocaccio for various Port-year groups were lower than for 4- year-old female Bocaccio (Table 10.7). Thus, to ensure equal precision we would generally need a larger number of sample landings for estimating the total catch of a species in a given sex-age group than for estimating the catch for the species as a whole. This is further discussed in the section on sample size.

The c.v.'s for estimating total catch of a species vary considerably among port-year groups. Thus during 1982, both Chilipepper and Bocaccio were estimated with much higher precision at Monterey than at Eureka although the estimates at the latter port were based on a larger sample size (88) than at Monterey (54). This is because Monterey was more abundant in the species than Bocaccio. Hence, the reliability of the estimates of a catch of a species at a port will depend on how important the port is with regard to the species.

The combined estimates for 'all ports' for 1981 based on 232 sample landings were reliable enough for estimating total catch of a species (c.v.'s ranging between 7 to 11.6 percent) and even for some of the dominant sex-age groups (12-year-old male and female Widows, 8-year-old female Chilipepper and 5- year-old male and female Bocaccio) for which the c.v.'s were reasonably small.

TABLE 10. Coefficient of Variation (in Percent) of Mean Number of Species Per Cluster Using $\bar{\bar{y}}_{2R}$ at

Eureka - 1982 (January - September)

(88)

| Species | Variance | | | |
|---|---|---|---|---|
| | Between | Within | Total | c.v. (%) |
| Widow | 0.5695 | 0.0022 | 0.5717 | 7.33 |
| Chilipepper | 0.0361 | 0.0002 | 0.0363 | — |
| Bocaccio | 0.0184 | 0.0000 | 0.0185 | 24.40 |

Monterey - 1982 (January - September)

(54)

| Species | Variance | | | |
|---|---|---|---|---|
| | Between | Within | Total | c.v. (%) |
| Widow | 0.2078 | 0.0132 | 0.2210 | 6.62 |
| Chilipepper | 0.0937 | 0.0041 | 0.0978 | 13.92 |
| Bocaccio | 0.1354 | 0.0054 | 0.1408 | 10.31 |

[1]Figures within brackets show sample landings on which estimates are based.

TABLE 10.1. Coefficient of Variation (in Percent) of Mean Number of Species per Cluster Using $\bar{\bar{y}}_{2R}$ by Port-Year Groups

Morro Bay - 1981

(40)

| Species | Variance | | | |
|---|---|---|---|---|
| | Between | Within | Total | c.v. (%) |
| Widow | 0.3317 | 0.0000 | 0.3317 | – |
| Chilipepper | 1.1329 | 0.0000 | 1.1329 | 14.21 |
| Bocaccio | 0.8489 | 0.0000 | 0.8489 | 12.28 |

Fort Bragg - 1981

(44)

| Species | Variance | | | |
|---|---|---|---|---|
| | Between | Within | Total | c.v. (%) |
| Widow | 0.4385 | 0.0000 | 0.4385 | – |
| Chilipepper | 0.3219 | 0.0000 | 0.3219 | 14.70 |
| Bocaccio | 0.4658 | 0.0000 | 0.4658 | 14.36 |

Eureka - 1981

(101)

| Species | Variance | | | |
|---|---|---|---|---|
| | Between | Within | Total | c.v. (%) |
| Widow | 0.1582 | 0.0003 | 0.1585 | 3.45 |
| Chilipepper | 0.0004 | 0.0000 | 0.0024 | – |
| Bocaccio | 0.1176 | 0.0000 | 0.1176 | – |

TABLE 10.1. (Continued) Coefficient of Variation (in Percent) of Mean Number of Species per Cluster Using $\bar{\bar{y}}_{2R}$ by Port-Year Groups

Monterey - 1981

(34)

| Species | Variance | | | |
|---|---|---|---|---|
| | Between | Within | Total | c.v. (%) |
| Widow | 0.1659 | 0.0004 | 0.1663 | — |
| Chilipepper | 0.7091 | 0.0002 | 0.7093 | 20.98 |
| Bocaccio | 0.5959 | 0.0006 | 0.5965 | 12.14 |

San Francisco - 1981

(13)

| Species | Variance | | | |
|---|---|---|---|---|
| | Between | Within | Total | c.v. (%) |
| Widow | 0.6552 | 0.0000 | 0.6552 | 23.64 |
| Chilipepper | 2.6615 | 0.0000 | 2.6615 | 19.45 |
| Bocaccio | 0.6304 | 0.0000 | 0.6304 | — |

TABLE 10.2. Coefficient of Variation (in Percent) of Mean Number of Species per Cluster Using $\bar{\bar{y}}_{2R}$ for

All Ports* - 1981

(232)

| Species | Variance | | | |
|---|---|---|---|---|
| | Between | Within | Total | c.v. (%) |
| Widow | 0.2625 | 0.0001 | 0.2626 | 6.99 |
| Chilipepper | 0.0949 | 0.0000 | 0.0949 | 11.60 |
| Bocaccio | 0.1103 | 0.0000 | 0.1103 | 10.72 |

*Includes the port-groups Monterey, San Francisco, Fort Bragg, Morro Bay and Eureka.

Dash ( — ) indicates that the corresponding c.v.'s are greater than 25%.

TABLE 10.3. Coefficient of Variation (in Percent) of Mean Weight of a Species per Cluster Using $\bar{\bar{y}}_{6R}$ by Port-Year Sort Groups

Eureka - 1978

| Species | Sort | Sample Landings | Variance | | | c.v. (%) |
|---|---|---|---|---|---|---|
| | | | Between | Within | Total | |
| Widow | 250 | 11 | 25.1717 | 0.5668 | 35.7385 | 24.79 (27.81) |
| Chilipepper | 250 | 13 | 20.4407 | 0.0665 | 20.5071 | — |
| Bocaccio | 250 | 25 | 1.6132 | 0.9451 | 2.5583 | 21.81 (26.00) |

Monterey - 1978

| Species | Sort | Sample Landings | Variance | | | c.v. (%) |
|---|---|---|---|---|---|---|
| | | | Between | Within | Total | |
| Widow | 253 | 12 | 33.2973 | 0.2329 | 33.5302 | — |
| Chilipepper | 253 | 33 | 2.5313 | 0.0204 | 2.5517 | 13.86 (16.63) |
| Bocaccio | 253 | 31 | 10.4338 | 0.0375 | 10.4713 | 13.84 (17.93) |

San Francisco - 1978

| Species | Sort | Sample Landings | Variance | | | c.v. (%) |
|---|---|---|---|---|---|---|
| | | | Between | Within | Total | |
| Bocaccio | 253 | 20 | 5.6482 | 0.0447 | 5.6929 | 7.45 (10.24) |

1. Dash ( — ) indicates that the corresponding c.v.'s are greater than 25%.
2. Figures within brackets ( ) show the corresponding c.v.'s for estimating mean number of a species.

TABLE 10.4. Coefficient of Variation (in Percent) and Components of Variance of Mean Catch per Cluster During 1982 (January - September) by Species-Age-Sex Groups Using $\bar{\bar{y}}_{2R}$ at Eureka

| Sex | Age (in years) | Variance Between | Variance Within | Total | c.v. (%) |
|---|---|---|---|---|---|
| | | Widow[1] (88) | | | |
| M | 16 | 0.0013 | 0.0002 | 0.0015 | 20.37 |
| M | 12 | 0.0080 | 0.0003 | 0.0083 | 14.84 |
| M | 11 | 0.0050 | 0.0002 | 0.0052 | 18.77 |
| M | 9 | 0.0010 | 0.0002 | 0.0012 | 19.92 |
| M | 7 | 0.0042 | 0.0002 | 0.0044 | 18.83 |
| F | 18 | 0.0019 | 0.0003 | 0.0022 | 19.68 |
| F | 17 | 0.0013 | 0.0004 | 0.0017 | 14.81 |
| F | 16 | 0.0083 | 0.0006 | 0.0089 | 16.08 |
| F | 15 | 0.0045 | 0.0002 | 0.0047 | 20.88 |
| F | 12 | 0.0190 | 0.0014 | 0.0204 | 14.10 |
| F | 11 | 0.0027 | 0.0002 | 0.0029 | 19.70 |
| F | 7 | 0.0051 | 0.0008 | 0.0059 | 10.94 |
| M | 8 | 0.0022 | 0.0001 | 0.0023 | 23.04 |
| F | 19 | 0.0037 | 0.0001 | 0.0038 | 23.22 |
| F | 14 | 0.0126 | 0.0003 | 0.0129 | 22.20 |
| M | 5 | 0.0050 | 0.0000 | 0.0050 | 24.14 |
| | | Chilipepper[1] (88) | | | |
| F | 13 | 0.0005 | 0.0000 | 0.0005 | 24.89 |
| F | 11 | 0.0006 | 0.0000 | 0.0006 | 23.03 |
| | | Bocaccio[1] (88) | | | |
| M | 6 | 0.0004 | 0.0002 | 0.0004 | 19.82 |
| F | 5 | 0.0002 | 0.0000 | 0.0002 | 22.65 |
| M | 5 | 0.0034 | 0.0000 | 0.0034 | 31.91 |
| F | 5 | 0.0002 | 0.0000 | 0.0002 | 22.65 |

1. M → Male; F → Female.
2. c.v. $\leq$ 25% are shown.

TABLE 10.4. Coefficient of Variation (in Percent) and Components of Variance of Mean Catch per Cluster by Species-Age-Sex Using $\bar{\bar{y}}_{2R}$ at Monterey During 1982 (January - September)

Widow[1] (54)

| Sex | Age (in years) | Variance Between | Variance Within | Variance Total | c.v. (%) |
|---|---|---|---|---|---|
| M | 11 | 0.0010 | 0.0005 | 0.0015 | 18.11 |
| M | 10 | 0.0011 | 0.0013 | 0.0024 | 25.38 |
| F | 13 | 0.0102 | 0.0007 | 0.0109 | 24.29 |
| F | 12 | 0.0090 | 0.0006 | 0.0096 | 20.49 |
| F | 11 | 0.0016 | 0.0003 | 0.0019 | 21.00 |

Chilipepper[1] (54)

| Sex | Age (in years) | Variance Between | Variance Within | Variance Total | c.v. (%) |
|---|---|---|---|---|---|
| M | 9 | 0.0002 | 0.0000 | 0.0002 | 18.36 |
| M | 8 | 0.0009 | 0.0000 | 0.0000 | 16.14 |
| M | 7 | 0.0019 | 0.0000 | 0.0019 | 12.30 |
| M | 6 | 0.0007 | 0.0000 | 0.0007 | 15.55 |
| M | 5 | 0.0002 | 0.0000 | 0.0002 | 15.67 |
| M | 3 | 0.0001 | 0.0000 | 0.0001 | 22.56 |
| F | 13 | 0.0005 | 0.0001 | 0.0006 | 21.40 |
| F | 11 | 0.0010 | 0.0001 | 0.0011 | 24.90 |
| F | 10 | 0.0006 | 0.0001 | 0.0007 | 16.61 |
| F | 9 | 0.0023 | 0.0002 | 0.0025 | 7.63 |
| F | 8 | 0.0028 | 0.0003 | 0.0031 | 9.00 |
| F | 7 | 0.0033 | 0.0001 | 0.0034 | 9.81 |
| F | 6 | 0.0006 | 0.0001 | 0.0007 | 8.86 |
| F | 5 | 0.0002 | 0.0000 | 0.0002 | 21.87 |
| F | 4 | 0.0007 | 0.0001 | 0.0008 | 18.34 |

TABLE 10.4. (Continued) Coefficient of Variation (in Percent) and Components of Variance of Mean Catch per Cluster by Species-Age-Sex Using $\bar{\bar{y}}_{2R}$ at Monterey During 1982 (January-September)

Bocaccio[1] (54)

| Sex | Age (in years) | Variance Between | Variance Within | Variance Total | c.v. (%) |
|---|---|---|---|---|---|
| M | 10 | 0.0004 | 0.0000 | 0.0004 | 23.22 |
| M | 9 | 0.0004 | 0.0000 | 0.0004 | 19.39 |
| M | 8 | 0.0002 | 0.0000 | 0.0002 | 7.43 |
| M | 7 | 0.0006 | 0.0001 | 0.0007 | 12.45 |
| M | 6 | 0.0015 | 0.0001 | 0.0016 | 9.75 |
| M | 5 | 0.0122 | 0.0001 | 0.0123 | 14.38 |
| M | 4 | 0.0032 | 0.0001 | 0.0033 | 10.65 |
| M | 3 | 0.0005 | 0.0000 | 0.0005 | 20.28 |
| F | 15 | 0.0002 | 0.0000 | 0.0002 | 24.64 |
| F | 13 | 0.0008 | 0.0001 | 0.0009 | 21.03 |
| F | 12 | 0.0002 | 0.0000 | 0.0002 | 8.96 |
| F | 11 | 0.0001 | 0.0001 | 0.0002 | 6.90 |
| F | 10 | 0.0003 | 0.0000 | 0.0003 | 13.57 |
| F | 9 | 0.0003 | 0.0001 | 0.0004 | 16.24 |
| F | 8 | 0.0003 | 0.0001 | 0.0004 | 10.55 |
| F | 7 | 0.0047 | 0.0003 | 0.0050 | 10.06 |
| F | 6 | 0.0013 | 0.0002 | 0.0015 | 8.03 |
| F | 5 | 0.0066 | 0.0003 | 0.0069 | 9.65 |
| F | 4 | 0.0022 | 0.0002 | 0.0024 | 12.34 |
| F | 3 | 0.0005 | 0.0001 | 0.0006 | 12.41 |

[1]Coefficient of variation $\leq$ 25%.

TABLE 10.5. Coefficient of Variation (in Percent) and Components of Variance of Mean Catch per Cluster by Species-Age-Sex Groups Using $\bar{\bar{y}}_{2R}$ at Ports During 1981

## Morro Bay

Chilipepper (40)

| Sex | Age (in years) | Variance: Between | Variance: Within | Variance: Total | c.v. (%) |
|---|---|---|---|---|---|
| M | 8 | 0.0191 | 0.0000 | 0.0191 | 24.16 |
| F | 11 | 0.0139 | 0.0000 | 0.0139 | 25.84 |
| F | 9 | 0.0939 | 0.0000 | 0.0939 | 20.41 |
| F | 8 | 0.0592 | 0.0000 | 0.0592 | 16.00 |
| F | 7 | 0.0367 | 0.0000 | 0.0367 | 22.77 |

Bocaccio (40)

| Sex | Age (in years) | Variance: Between | Variance: Within | Variance: Total | c.v. (%) |
|---|---|---|---|---|---|
| M | 6 | 0.0026 | 0.0000 | 0.0026 | 25.43 |
| M | 5 | 0.0608 | 0.0000 | 0.0608 | 11.99 |
| F | 6 | 0.0051 | 0.0000 | 0.0051 | 21.92 |
| F | 5 | 0.0897 | 0.0000 | 0.0897 | 12.67 |

## Fort Bragg

Chilipepper (44)

| Sex | Age (in years) | Variance: Between | Variance: Within | Variance: Total | c.v. (%) |
|---|---|---|---|---|---|
| F | 9 | 0.0196 | 0.0000 | 0.0196 | 25.00 |

Bocaccio (44)

| Sex | Age (in years) | Variance: Between | Variance: Within | Variance: Total | c.v. (%) |
|---|---|---|---|---|---|
| M | 5 | 0.0866 | — | 0.0866 | 22.27 |
| F | 5 | 0.0558 | — | 0.0558 | 18.39 |

TABLE 10.5. (Continued) Coefficient of Variation (in Percent) and Components of Variance of Mean Catch per Cluster by Species-Age-Sex Groups Using $\bar{\bar{y}}_{2R}$ at Ports During 1981

Monterey

Chilipepper (34)

| Sex | Age (in years) | Variance Between | Variance Within | Total | c.v. (%) |
|---|---|---|---|---|---|
| F | 8 | 0.0154 | 0.0000 | 0.0154 | 25.80 |
| F | 7 | 0.0276 | 0.0000 | 0.0276 | 18.39 |
| | | Bocaccio (34) | | | |
| M | 5 | 0.0310 | 0.0003 | 0.0313 | 13.30 |
| F | 5 | 0.2065 | 0.0005 | 0.2070 | 17.88 |

San Francisco

Widow (13)

| Sex | Age (in years) | Variance Between | Variance Within | Total | c.v. (%) |
|---|---|---|---|---|---|
| M | 12 | 0.0247 | 0.0000 | 0.0247 | 24.33 |
| M | 7 | 0.0056 | 0.0000 | 0.0056 | 25.58 |
| F | 14 | 0.0002 | 0.0000 | 0.0002 | 14.93 |
| | | Chilipepper (13) | | | |
| M | 10 | 0.0123 | 0.0000 | 0.0123 | 24.10 |
| M | 9 | 0.0341 | 0.0000 | 0.0341 | 22.44 |
| M | 5 | 0.0083 | 0.0000 | 0.0083 | 23.03 |
| F | 9 | 0.0949 | 0.0000 | 0.0949 | 24.36 |
| F | 7 | 0.4019 | 0.0000 | 0.0000 | 21.75 |
| F | 6 | 0.0118 | 0.0000 | 0.0118 | 19.84 |
| | | Bocaccio (13) | | | |
| F | 5 | 0.0373 | 0.0000 | 0.0373 | 21.25 |

TABLE 10.5. (Continued) Coefficient of Variation (in Percent) and Components of Variance of Mean Catch per Cluster by Species-Age-Sex Groups Using $\bar{\bar{y}}_{2R}$ at Ports During 1981

Eureka

Widow (101)

| Sex | Age (in years) | Variance Between | Variance Within | Total | c.v. (%) |
|---|---|---|---|---|---|
| M | 12 | 0.0148 | 0.0003 | 0.0151 | 14.04 |
| M | 11 | 0.0221 | 0.0004 | 0.0225 | 20.46 |
| F | 17 | 0.0074 | 0.0002 | 0.0076 | 15.52 |
| F | 16 | 0.0048 | 0.0000 | 0.0048 | 12.88 |
| F | 15 | 0.0110 | 0.0001 | 0.0111 | 20.98 |
| F | 14 | 0.0142 | 0.0001 | 0.0143 | 15.94 |
| F | 13 | 0.0171 | 0.0001 | 0.0172 | 13.56 |
| F | 12 | 0.0316 | 0.0009 | 0.0325 | 10.96 |
| F | 11 | 0.0074 | 0.0001 | 0.0075 | 14.03 |

TABLE 10.6. Coefficient of Variation (in Percent) and Components of Variance of Mean Catch per Cluster by Species-Age-Sex Group Using the Ratio Estimate $\bar{\bar{y}}_{2R}$ for

All Ports* - 1981

Widow (232)

| Sex | Age (in years) | Variance Between | Variance Within | Total | c.v. (%) |
|---|---|---|---|---|---|
| M | 13 | 0.0091 | 0.0000 | 0.0091 | 20.22 |
| M | 12 | 0.0070 | 0.0001 | 0.0071 | 13.94 |
| M | 11 | 0.0092 | 0.0001 | 0.0093 | 20.59 |
| M | 8 | 0.0012 | 0.0000 | 0.0012 | 21.48 |
| M | 7 | 0.0029 | 0.0001 | 0.0030 | 24.24 |
| F | 17 | 0.0036 | 0.0001 | 0.0037 | 18.04 |
| F | 16 | 0.0028 | 0.0000 | 0.0028 | 15.57 |
| F | 15 | 0.0045 | 0.0001 | 0.0046 | 20.69 |
| F | 14 | 0.0061 | 0.0000 | 0.0061 | 17.58 |
| F | 13 | 0.0081 | 0.0000 | 0.0081 | 15.23 |
| F | 12 | 0.0145 | 0.0002 | 0.0147 | 12.09 |
| F | 11 | 0.0035 | 0.0000 | 0.0035 | 15.42 |
| F | 8 | 0.0026 | 0.0000 | 0.0026 | 23.02 |

Chilipepper (232)

| Sex | Age (in years) | Variance Between | Variance Within | Total | c.v. (%) |
|---|---|---|---|---|---|
| M | 8 | 0.0010 | 0.0000 | 0.0010 | 21.14 |
| M | 7 | 0.0016 | 0.0000 | 0.0016 | 25.50 |
| F | 12 | 0.0009 | 0.0000 | 0.0009 | 24.43 |
| F | 11 | 0.0011 | 0.0000 | 0.0011 | 19.75 |
| F | 10 | 0.0012 | 0.0000 | 0.0012 | 24.80 |
| F | 9 | 0.0055 | 0.0000 | 0.0055 | 17.34 |
| F | 8 | 0.0036 | 0.0000 | 0.0036 | 15.81 |
| F | 7 | 0.0077 | 0.0000 | 0.0077 | 19.61 |
| F | 6 | 0.0006 | 0.0000 | 0.0006 | 17.80 |

*Includes the port groups Monterey, San Francisco, Fort Bragg, Morro Bay and Eureka.

TABLE 10.6. (Continued) Coefficient of Variation (in Percent) and Components of Variance of Mean Catch per Cluster by Species-Age-Sex Group Using the Ratio Estimate $\bar{\bar{y}}_{2R}$ for

All Ports - 1981

Bocaccio (232)

| Sex | Age (in years) | Variance Between | Variance Within | Total | c.v. (%) |
|---|---|---|---|---|---|
| M | 6 | 0.0008 | 0.0000 | 0.0008 | 22.94 |
| M | 5 | 0.0127 | 0.0000 | 0.0127 | 14.29 |
| M | 4 | 0.0019 | 0.0000 | 0.0019 | 22.78 |
| F | 6 | 0.0018 | 0.0000 | 0.0018 | 18.70 |
| F | 5 | 0.0122 | 0.0000 | 0.0122 | 11.28 |
| F | 4 | 0.0003 | 0.0000 | 0.0003 | 25.30 |

TABLE 10.7. Coefficient of Variation (in Percent) of Mean Catch per Cluster of Bocaccio and of Four-Year-Old Female Bocaccio by Port-Year Groups

| | | Bocaccio | | 4 Year Old Female Bocaccio | |
|---|---|---|---|---|---|
| Port | Year | Sample Size | c.v.(%) | Sample Size | c.v.(%) |
| San Francisco | 1978 | 20 | 11.59 | 20 | 39.14 |
| Fort Bragg | 1978 | 86 | 7.36 | 86 | 15.00 |
| | 1979 | 46 | 27.10 | 46 | 70.81 |
| Monterey | 1978 | 31 | 19.55 | 27 | 36.69 |
| | 1979 | 51 | 9.53 | 45 | 18.53 |
| | 1981 | 34 | 12.14 | 34 | 32.93 |
| | 1982 | 54 | 10.31 | 54 | 12.34 |
| Eureka[1] | 1978 | 25 | 30.03 | 25 | 53.05 |
| | 1982 | 88 | 24.40 | 88 | 35.35 |

[1]c.v. for 1981 (not shown) are high for species and considerably higher for 4-year-old females.

9. RELIABILITY OF THE ESTIMATES: SAMPLE SIZE

This section will deal with the number of landings required to ensure a specified level of c.v. for estimating the total catch (or weight) at a port during a year; sample sizes for obtaining quarterly estimates will also be indicated.

We will consider c.v.'s of 10, 15 and 20 percent for estimating total catch of a species (and weight) and of 10 and 20 percent for estimating catch in a particular sex-age group. Sample size to ensure 5 percent c.v. was generally too high for estimating species catch with the present staff; even a 10 percent c.v. could not be realized in some cases, e.g., Chilipepper and Bocaccio at Eureka, Widow at Fort Bragg and Morro Bay during 1981. Sample size to estimate the catch in a species-sex-age group with a 10 percent c.v. was even high in a number of cases and a twenty percent c.v. is recommended in such cases.

For a population, the sample size needed to ensure a coefficient of variation e of estimated total number (or weight) of a species is given by

$$n = \frac{Nc^2}{(N-1)e^2 + c^2} \approx \frac{n_o}{1 + \frac{n_o}{N}} \qquad (51)$$

$$\text{where } n_o = (c/e)^2$$

where c is the population coefficient of variation, N, the total number of landings in the population. As a first approximation we take $n_o = (c/e)^2$, by substituting an advance estimate of c and knowing e, the desired c.v. of the sample estimate.

In practice, N is not known and is estimated by

$$\hat{N} = \frac{W}{\hat{\bar{W}}}$$

It is essential we have a good estimate of N. One possibility is to base $\hat{\bar{W}}$ on a large sample. In the absence of such an estimate of N or when $\hat{N}$ is less than $n_o$, no adjustment is possible as in (51) and $n_o$ is our best choice. If the number of landings is proportional to the number in each weight group, $\hat{N}$ will be an unbiased estimate of N.

Sampling within sort groups can not be realized in practice since no sampling frame by sort groups is available in a port-time stratum. However, an improved estimate of the population coefficient of variation (and hence of the sample size) can be obtained by using a weighted mean of the variances by sort groups at the post-stratification stage after the samples are selected. Also, the estimate of sample size will be more precise when a minimum number of samples per sort group (i.e. 4 per month) is selected at each port during a month as recommended elsewhere in this report.

9.1. Sample Size for Species Catch by Port-Year Groups

The sample landings needed for estimating total catch by port-year groups (1981-1982) show that, generally speaking, the number needed is very high, if a 5 percent c.v. is aimed at in obtaining the estimates. Hence, it is decided to present more workable estimates using 10, 15 and 20 percent.

A smaller number of landings is needed at Monterey (where the species variability is small) than at Eureka for estimating Chilipepper and Bocaccio; in fact, the number needed at Eureka during 1981 for estimating Chilipepper and Bocaccio with a c.v. of 10 percent is over 1000 sample landings, which certainly cannot be managed even if the staff were doubled for the port-group. At Fort Bragg and Morro Bay, the sample landings required for estimating the species with a 10 percent c.v. are reasonable and can be managed by the present staff assuming that a minimum of four landings per month per sort group are covered at a port. At San Francisco, the sample size required for estimating catch of Widow and Chilipepper with a ten percent c.v. are 56 and 41, and it should be possible for the present staff to complete this task working at the desired speed. The estimates should, however, be used with caution since the estimates of population c.v.'s are based on too few landings (13).

For 'all ports' which comprised the port-groups Monterey, San Francisco, Fort Bragg, Morro Bay and Eureka, the sample landings (not shown in the table) needed to estimate the catch of Widows (using 1981 data to estimate population c.v.) with 5 percent c.v. is about 50 percent more than the sample landings (232) used in estimating with 7 percent c.v. For Bocaccio and Chilipper, the sample size needed to estimate with 10 percent c.v. is slightly more than used for 'all ports', which the present staff should be able to manage. In fact, with a minimum of 4 samples per sort per port-month group as recommended elsewhere, the total number of samples needed for 'all ports' is estimated at 432 landings per year which should provide annual estimate of total catch of Widow for 'all ports' with less than 5 percent c.v. and of Bocaccio and Chilipper with c.v.'s ranging between 5 and 10 percent. Additional staff will be needed at least at the important port groups (e.g., Eureka) to provide more reliable estimates for the latter species.

## 9.2 Sample Size for Species Weight by Port-Year Sort Groups

The sample size needed for estimating total weight (Table 11.2) is based on the analysis of data for 1978 for Eureka, Monterey and San Francisco and for 2 sort groups. The sort groups and sample sizes on which estimates of population coefficient of variation are based are the same as those for estimating coefficient of variation (Table 10.3).

The number of landings needed for estimating weight of Widow at Eureka (using 1978 data to estimate population c.v.) with a 10 percent c.v. was high compared to those using 1981 and 1982 data. As stated earlier this estimate may not be reliable enough, because of the small sample (eleven landings) on which it is based. Using 1980 data (figures not shown) to estimate population c.v. a sample size of 54 would provide an estimate of weight with a 2 percent c.v. and of species catch with a 2.2 percent c.v. It is, therefore, possible for the present staff to estimate annual weight of Widow landings at Eureka with a five percent c.v.

The sample size needed to estimate weight of Chilipepper and Bocaccio at Monterey during a year with a ten percent c.v. is between 50 and 60 landings which, again lies within the workload of existing staff as per recommendations of landings for the port group during a year.

The sample size needed for estimating total weight of a species for a port-year group is generally less than that for estimating total catch. For example, the number of landings needed to estimate with a ten percent c.v. the catch of Bocaccio at Eureka, Monterey and San Francisco (using 1978 data) are 109, 87 and 21 respectively as against 86, 53 and 11 for estimating total weight. This supports the conclusion in the earlier section that the variation (c.v.) among samples for species weight is less than that for less than species catch (Table 10.3) for the ports considered in this study.

## 9.3. Sample Size for Species Catch by Sex and Age for Port-Year Groups

The number of landings at a port during a year needed for estimating catch of a species in a sex-age group (Tables 11.3 - 11.4) with a ten percent c.v. is generally high and cannot be managed by the present staff except for some of the dominant sex-age groups. These are seven- and twelve-year old female Widow at Eureka, eleven-year-old male Widow at Monterey, six- seven- eight-year-old male and female Chilipepper and Bocaccio at Monterey, and five-year-old male and female Bocaccio at Morro Bay.

For 'all ports' the annual catch for mostly dominant sex-age groups e.g., 12-year-old male and female Widow, 8-year-old female Chilipepper and 5-year-old male and female Bocaccio were estimated with a ten percent c.v. For all other sex-age groups the estimates are subject to high error (20 percent c.v. or more) and the current staff is inadequate for obtaining estimates with a reasonable degree of accuracy.

## 9.4. Sample Size for Species Catch by Ports on Quarterly Basis

We have seen that for port-year estimates for species catch (or weight) sample sizes should be generally based on a ten percent c.v. and for sex-age groups on a 20 percent c.v. except where these are available in abundance when a smaller sample size might serve the purpose. For 'all ports', estimates during a year will be measurable with smaller error i.e., 5 to 7 percent c.v.

For quarterly estimates, the recommended sample size can at best provide estimates of catch with 20 percent c.v. assuming that the population c.v. would be approximately the same for quarterly and annual estimates. For estimating species catch with a smaller error during a quarter we would need more samples and hence additional staff, at least at the important port groups. In view of the very low precision of quarterly estimates it is recommended that either these be not published or published with the note that their use is restricted to serve merely as indicators rather than provide a measure of reliability for species catch.

To sum up, the present staff should be able to provide estimates of total catch of species by 'all ports' during a year with a c.v. between 5 and 7 percent and at port-year level with a c.v. of 10 percent. Sample size for estimating total weight would be somewhat smaller to provide the same level of accuracy.

To estimate species catch (or weight) by sex and age at the port-year level with a reasonable degree of accuracy, we would need more samples and hence extra staff for their collection, at least at the important port groups such as Eureka and Monterey.

TABLE 11. Sample Size Needed for Specified c.v. Levels for Estimating Total Catch of a Species by Port-Year Groups

Eureka - 1982 (January - September)

(88)

| Species | Number of Landings Required For | | |
|---|---|---|---|
| | 10% c.v. | 15% c.v. | 20% c.v. |
| Widow | 43 | 20 | 12 |
| Chilipepper | 489 | 292 | 187 |
| Bocaccio | 351 | 191 | 117 |

Monterey - 1982 (January - September)

(54)

| Species | Number of Landings Required For | | |
|---|---|---|---|
| | 10% c.v. | 15% c.v. | 20% c.v. |
| Widow | 22 | 11 | 6 |
| Chilipepper | 77 | 40 | 24 |
| Bocaccio | 47 | 24 | 13 |

[1]Figures in brackets show 'sample landings'.

TABLE 11.1. Sample Size Needed for Specified c.v. Levels for Estimating Total Catch of a Species by Port-Year Groups

Morro Bay - 1981
(40)

| Species | Number of Landings Required For | | |
|---|---|---|---|
| | 10% c.v. | 15% c.v. | 20% c.v. |
| Widow | 565* | 251* | 87 |
| Chilipepper | 60 | 31 | 18 |
| Bocaccio | 48 | 24 | 4 |

Fort Bragg - 1981
(44)

| Species | Number of Landings Required For | | |
|---|---|---|---|
| | 10% c.v. | 15% c.v. | 20% c.v. |
| Widow | – | 274 | 189 |
| Chilipepper | 83 | 39 | 23 |
| Bocaccio | 80 | 38 | 22 |

Eureka - 1981
(101)

| Species | Number of Landings Required For | | |
|---|---|---|---|
| | 10% c.v. | 15% c.v. | 20% c.v. |
| Widow | 12 | 5 | 3 |
| Chilipepper | – | – | 803* |
| Bocaccio | – | 845* | 283 |

[1]Figures in brackets show 'sample landings'.

[2]Dash (–) indicates over 1,000 landings.

*Unadjusted estimates.

TABLE 11.1. (Continued) Sample Size Needed for Specified c.v. Levels for Estimating Total Catch of a Species by Port-Year Groups

Monterey - 1981
(34)

| Species | Number of Landings Required For | | |
|---|---|---|---|
| | 10% c.v. | 15% c.v. | 20% c.v. |
| Widow | 200 | 120 | 77 |
| Chilipepper | 111 | 58 | 34 |
| Bocaccio | 45 | 21 | 13 |

San Francisco - 1981
(13)

| Species | Number of Landings Required For | | |
|---|---|---|---|
| | 10% c.v. | 15% c.v. | 20% c.v. |
| Widow | 56 | 28 | 17 |
| Chilipepper | 41 | 20 | 11 |
| Bocaccio | 102 | 59 | 37 |

All Ports* - 1981
(232)

| Species | Number of Landings Required For | | |
|---|---|---|---|
| | 10% c.v. | 15% c.v. | 20% c.v. |
| Widow | 107 | 49 | 28 |
| Chilipepper | 272 | 130 | 75 |
| Bocaccio | 237 | 112 | 65 |

[1]Figures in brackets show 'sample landings'.

*Includes the port groups Monterey, San Francisco, Fort Bragg, Morro Bay, and Eureka.

TABLE 11.2. Sample Size Needed for Estimating Total Weight of Catch of a Species for Specified c.v. (%) Levels

Eureka - 1978

| Species | Sort | Number of Landings Required For | | |
|---|---|---|---|---|
| | | 10% c.v. | 15% c.v. | 20% c.v. |
| Widow (11) | 250 | 49 | 27 | 15 |
| Chilipepper (13) | 250 | 233* | 77 | 44 |
| Bocaccio (25) | 250 | 86 | 45 | 27 |

Monterey - 1978

| Species | Sort | Number of Landings Required For | | |
|---|---|---|---|---|
| | | 10% c.v. | 15% c.v. | 20% c.v. |
| Widow (12) | 253 | 208 | 105 | 67 |
| Chilipepper (33) | 253 | 56 | 26 | 15 |
| Bocaccio (31) | 253 | 53 | 18 | 14 |

San Francisco - 1978

| Species | Sort | Number of Landings Required For | | |
|---|---|---|---|---|
| | | 10% c.v. | 15% c.v. | 20% c.v. |
| Bocaccio (26) | 253 | 11 | 5 | 3 |

*Unadjusted estimates

Number within brackets show the number of landings for estimating population c.v.

TABLE 11.3. Sample Size Needed for Estimating Total Catch of a Species for Specified c.v. (%) Levels by Age, Sex, Port and Year Groups

Eureka - 1982 (January - September)

Widow (88)

| Sex | Age (in years) | Number of Landings Required For 10% c.v. | 20% c.v. |
|---|---|---|---|
| M | 16 | 271 | 84 |
| M | 12 | 164 | 46 |
| M | 11 | 240 | 72 |
| M | 9 | 262 | 80 |
| M | 17 | 241 | 73 |
| F | 18 | 258 | 79 |
| F | 7 | 163 | 46 |
| F | 16 | 187 | 54 |
| F | 15 | 282 | 88 |
| F | 12 | 150 | 42 |
| F | 11 | 258 | 79 |
| F | 7 | 95 | 25 |

Chilipepper (88)

| | | | |
|---|---|---|---|
| F | 13 | 360 | 120 |
| F | 11 | 324 | 105 |

Bocaccio (88)

| | | | |
|---|---|---|---|
| M | 6 | 261 | 80 |
| M | 5 | 486 | 185 |
| F | 5 | 316 | 102 |

TABLE 11.3. (Continued) Sample Size Needed for Estimating Total Catch of a Species for Specified c.v. (%) Levels by Age, Sex, Port and Year Groups

Monterey - 1982 (January - September)

Widow (54)

| Sex | Age (in years) | Number of Landings Required For 10% c.v. | 20% c.v. |
|---|---|---|---|
| M | 11 | 109 | 38 |
| M | 10 | 346* | 67 |
| F | 13 | 319* | 62 |
| F | 12 | 126 | 47 |
| F | 11 | 129 | 49 |

Chilipepper (54)

| Sex | Age (in years) | 10% c.v. | 20% c.v. |
|---|---|---|---|
| M | 9 | 109 | 38 |
| M | 8 | 94 | 31 |
| M | 7 | 64 | 19 |
| M | 6 | 89 | 30 |
| M | 5 | 90 | 30 |
| M | 3 | 140 | 56 |
| F | 13 | 132 | 51 |
| F | 11 | 335* | 65 |
| F | 10 | 98 | 33 |
| F | 9 | 28 | 8 |
| F | 8 | 38 | 11 |
| F | 7 | 44 | 12 |
| F | 6 | 36 | 10 |
| F | 5 | 135 | 53 |
| F | 4 | 106 | 39 |

*Not adjusted

TABLE 11.3. (Continued) Sample Size Needed for Estimating Total Catch of a Species for Specified c.v. (%) Levels by Age, Sex, Port and Year Groups

Monterey - 1982 (January - September)

Bocaccio (54)

| Sex | Age (in years) | Number of Landings Required For 10% c.v. | 20% c.v. |
|---|---|---|---|
| M | 10 | 291* | 58 |
| M | 9 | 118 | 43 |
| M | 8 | 27 | 7 |
| M | 7 | 65 | 19 |
| M | 6 | 43 | 12 |
| M | 5 | 80 | 25 |
| M | 4 | 50 | 14 |
| M | 3 | 124 | 46 |
| F | 15 | 328* | 64 |
| F | 13 | 130 | 49 |
| F | 12 | 37 | 10 |
| F | 11 | 24 | 6 |
| F | 10 | 88 | 23 |
| F | 9 | 95 | 32 |
| F | 8 | 49 | 14 |
| F | 7 | 46 | 13 |
| F | 6 | 31 | 9 |
| F | 5 | 42 | 11 |
| F | 4 | 63 | 19 |

TABLE 11.4. Sample Size Needed for Estimating Total Catch of a Species for Specified c.v. (%) Levels by Age, Sex, Port and Year Groups

**Morro Bay - 1981**

Chilipepper (40)

| Sex | Age (in years) | Number of Landings Required For | |
|---|---|---|---|
| | | 10% c.v. | 20% c.v. |
| M | 8 | 233* | 46 |
| F | 11 | 267* | 52 |
| F | 9 | 97 | 35 |
| F | 8 | 70 | 23 |
| F | 7 | 109 | 42 |

Bocaccio (40)

| Sex | Age (in years) | 10% c.v. | 20% c.v. |
|---|---|---|---|
| M | 6 | 259* | 51 |
| M | 5 | 46 | 13 |
| F | 6 | 105 | 40 |
| F | 5 | 50 | 37 |

**Fort Bragg - 1981**

Chilipepper (44)

| Sex | Age (in years) | Number of Landings Required For | |
|---|---|---|---|
| | | 10% c.v. | 20% c.v. |
| F | 9 | 194 | 62 |

Bocaccio (44)

| Sex | Age (in years) | 10% c.v. | 20% c.v. |
|---|---|---|---|
| M | 5 | 164 | 50 |
| F | 5 | 121 | 35 |

TABLE 11.4. Sample Size Needed for Estimating Total Catch of a Species for Specified c.v. (%) Levels by Age, Sex, Port and Year Groups

## Monterey - 1981

Chilipepper (34)

| Sex | Age (in years) | Number of Landings Required For 10% c.v. | 20% c.v. |
|---|---|---|---|
| F | 8 | 148 | 49 |
| F | 7 | 112 | 27 |

Bocaccio (34)

| Sex | Age (in years) | 10% c.v. | 20% c.v. |
|---|---|---|---|
| M | 5 | 53 | 14 |
| F | 5 | 87 | 25 |

## San Francisco - 1981

Widow (13)

| Sex | Age (in years) | Number of Landings Required For 10% c.v. | 20% c.v. |
|---|---|---|---|
| M | 12 | 59 | 18 |
| M | 7 | 63 | 19 |
| F | 14 | 26 | 7 |

Chilipepper (13)

| Sex | Age (in years) | 10% c.v. | 20% c.v. |
|---|---|---|---|
| M | 10 | 57 | 18 |
| M | 9 | 51 | 15 |
| M | 5 | 54 | 16 |
| F | 9 | 59 | 18 |
| F | 7 | 49 | 14 |
| F | 6 | 42 | 12 |

Bocaccio (13)

| Sex | Age (in years) | 10% c.v. | 20% c.v. |
|---|---|---|---|
| F | 5 | 48 | 14 |

TABLE 11.4. (Continued) Sample Size Needed for Estimating Total Catch of a Species for Specified c.v. (%) Levels by Age, Sex, Port and Year Groups

Eureka - 1981

Widow (101)

| Sex | Age (in years) | Number of Landings Required For | |
|---|---|---|---|
| | | 10% c.v. | 20% c.v. |
| M | 12 | 155 | 47 |
| M | 11 | 263 | 92 |
| F | 17 | 180 | 56 |
| F | 16 | 92 | 40 |
| F | 15 | 271 | 96 |
| F | 14 | 188 | 59 |
| F | 13 | 147 | 43 |
| F | 12 | 103 | 29 |
| F | 11 | 155 | 47 |

All Ports* - 1981

Widow (232)

| Sex | Age (in years) | Number of Landings Required For | |
|---|---|---|---|
| | | 10% c.v. | 20% c.v. |
| M | 13 | 657 | 231 |
| M | 12 | 373 | 107 |
| M | 11 | 674 | 221 |
| M | 8 | 713 | 238 |
| M | 7 | 833 | 294 |
| F | 17 | 558 | 174 |
| F | 16 | 445 | 132 |
| F | 15 | 678 | 222 |
| F | 14 | 537 | 165 |
| F | 13 | 430 | 126 |
| F | 12 | 293 | 82 |
| F | 11 | 439 | 130 |
| F | 8 | 781 | 268 |

TABLE 11.4. (Continued) Sample Size Needed for Estimating Total Catch of a Species for Specified c.v. (%) Levels by Age, Sex, Port and Year Groups

All Ports* - 1981

Chilipepper (232)

| Sex | Age (in years) | Number of Landings Required For 10% c.v. | 20% c.v. |
|---|---|---|---|
| M | 8 | 699 | 231 |
| M | 7 | 886 | 321 |
| F | 12 | 841 | 298 |
| F | 11 | 636 | 204 |
| F | 10 | 857 | 306 |
| F | 9 | 527 | 161 |
| F | 8 | 456 | 136 |
| F | 7 | 630 | 202 |
| F | 6 | 547 | 169 |

Bocaccio (232)

| Sex | Age (in years) | Number of Landings Required For 10% c.v. | 20% c.v. |
|---|---|---|---|
| M | 6 | 778 | 267 |
| M | 5 | 388 | 112 |
| M | 4 | 771 | 264 |
| F | 6 | 588 | 185 |
| F | 5 | 259 | 72 |
| F | 4 | 877 | 316 |

*Includes the port groups Monterey, San Francisco, Fort Bragg, Morro Bay and Eureka.

## 10. AGE COMPOSITION: DOUBLE SAMPLING

For most theoretical population work and for management purposes, the knowledge of the age composition is important to predict the status of the stock in future years. When sample sizes are large, as in commercial fishing it requires considerable time and expense to age each individual in the sample. This can be done from otoliths with a fair degree of accuracy. The length of a fish gives a good guide to its age. Because many length measurements may be done relatively quicker than an age determination, the age composition is often easily obtained by double sampling using a large sample of length measurements and a relatively few age determinations as a sub-sample from the large sample (Fridriksson, 1934).

Ketchen (1950) improved over Fridriksson's approach by sub-sampling from large samples sorted into length categories or strata which provided more accurate results for age groups at the extremities of the distribution.

Kutkuhn (1963) mentions the limitations of this (age-length key) approach except in situations where price differentials may demand sorting of landings by size criteria. Westrheim and Ricker (1978) point out that the age-length key approach will almost always give biased estimates. Following the method due to Tanaka (1953) in which stratification occurs after sub-sampling for age, Kutkuhn estimated absolute age composition of California salmon landings by port-month groups. He showed that the sampling procedure is **not** effective unless the age sample is at least five times costlier than the length sample.

Mackett (1963) found double sampling more efficient than simple random sampling with fixed sampling costs for estimating relative age composition of Pacific albacore landings.

Southward (1976) found that a sample of otoliths proportional to the length frequency of sampled fish from each port was preferable to fixed sample size procedure for estimating age composition of Pacific Halibut. Kimura (1977) arrived at the same conclusion as Southward by following a somewhat different approach.

In a personal communication (1983), Lenarz examines the use of double sampling to estimate the age composition of fish landings for a fixed cost by allocating effort between length and age sampling and among length strata.

We will present in this report some of the important statistical considerations in sampling for estimation of age composition with illustration from recent Widow rock fish data from the California coast.

The previous studies have shown that since aging from otoliths of each individual fish in a sample is more expensive than an easily measured quantity such as length, it may pay to measure the length of each individual and then either (a) choose a random sub-sample from the whole sample, or (b) stratify the sample according to length classes and choose a sub-sample from each class for age determination.

This technique is profitable only if the correlation between the length and age is fairly high for use of ratio or regression estimates or stratification.

In the construction of length strata for selection of the sub-sample the following additional questions arise.

(i) How many strata,

(ii) How to decide boundaries between strata and

(iii) determination of the allocation plan among strata

for deriving maximum gain from double sampling.

## 10.1 Number of Strata

I present below the value of $V(\bar{y}_{st})/V_{\bar{y}}$ as a function of $L$, number of strata using the linear model

$$y = \alpha + \beta x + \varepsilon$$

where $y$ is the length and $x$ the age of a fish and

| | Data Set | | |
|---|---|---|---|
| L | 1 | 2 | 3 |
| 2 | 0.6041 | 0.5114 | 0.4747 |
| 3 | 0.5308 | 0.4209 | 0.3774 |
| 4 | 0.5052 | 0.3892 | 0.3434 |
| 5 | 0.4933 | 0.3746 | 0.3276 |

Type of Data

| Set | Data | x | y | Source |
|---|---|---|---|---|
| | | age (years) | length (cm) | |
| 1. | Female widow rockfish (328) landed at Eureka during April - December (1980) | 1980 | 1980 | California Fish and Game and Tiburon Laboratory |
| 2. | Female widow rockfish (444) landed at Eureka during January - September (1981) | 1981 | 1981 | " " |
| 3. | Female widow rockfish (532) landed at Monterey, San Francisco and Bodega Bay during January - March (1982) | 1982 | 1982 | " " |

$$\frac{V(\bar{y}_{st})}{V(\bar{y})} = \left[\frac{\rho^2}{L^2} + (1 - \rho^2)\right] \tag{52}$$

where $\rho$ is the correlation between length and age in the unstratified population and L is the number of strata. It can be shown that with this model unless $\rho > 0.95$ there is hardly any gain due to stratification if L exceeds 6. Increasing L would however, result in somewhat lower precision owing to reduction in sample size in different strata. The improvement is highest for data set 3 for which $r^2 = 0.7004$ and lowest for set 1 for which $\rho^2 = 0.5278$.

10.2 Strata Boundaries

Using the length-age data on 239 female widow rockfish for 1982 for San Francisco and the rule based on the cumulative of $\sqrt{f(y)}$ (Cochran 1977 , pp. 127-129), where $y$ denotes length in centimeters, the nearest available points for 2 and 4 strata are

| | Stratum 1 | Stratum 2 |
|---|---|---|
| Boundaries | (36 - 47) cm | (48 - 55) cm |
| Intervals on Cum $\sqrt{f}$ | 18.70 | 23.72 |

| | Stratum 1 | Stratum 2 | Stratum 3 | Stratum 4 |
|---|---|---|---|---|
| Boundaries | (36 - 43) cm | (44-47) cm | (48-51) cm | (52-55) cm |
| Intervals on Cum $\sqrt{f}$ | 9.30 | 9.40 | 16.25 | 7.47 |

It turns out that the overleaf division point is approximately the same as as for young and old widow. This should give efficient stratification for length-age which has linear regression with high correlation. The correlation $(\rho_h)$ will be moderate when a number of strata are used.

For length-age data (1981) based on 444 female widow rock fish landings at Eureka, the boundaries using 2, 3, and 4 strata are

| | Stratum 1 | Stratum 2 |
|---|---|---|
| Boundaries | (31.5 - 46) cm | (46.5 - 55) cm |
| Intervals on Cum $\sqrt{f}$ | 17.70 | 29.01 |

| | Stratum 1 | Stratum 2 | Stratum 3 |
|---|---|---|---|
| Boundaries | (31.5 - 46) cm | (46.5 - 49.0) cm | (49.5 - 55) cm |
| Intervals on Cum $\sqrt{f}$ | 17.70 | 13.12 | 15.89 |

| | Stratum 1 | Stratum 2 | Stratum 3 | Stratum 4 |
|---|---|---|---|---|
| Boundaries | (31.5 - 43) cm | (43.5-46) cm | (46.5-49) cm | (49.5-55) cm |
| Intervals on Cum $\sqrt{f}$ | 9.34 | 8.36 | 13.12 | 15.9 |

10.3 Optimum Allocation Plan

We have seen in the earlier section that one of the requirements of double sampling is that the correlation between the length and age of fish should be high. In fact it is known that double sampling is more efficient than single sampling (when the first sample is measured for age alone) for the same cost if

$$\rho^2 > \frac{4(c/c')}{(1+c/c')^2} \tag{53}$$

where $\rho$ is the correlation between length and age of fish and $c$ and $c'$ are respectively the costs of aging and measuring a fish.

Assuming that the average cost of aging fish (including small and large) is 6 minutes and of measuring is 1.2 minutes (estimates based on measurements by W. Lenarz of Tiburon Laboratory).

We have from (53)

$$\rho^2 > 0.5555$$

$$\text{or} \quad \rho > 0.7453$$

For the three data sets (p.81) the values of $\rho^2$ are respectively 0.5278, 0.6515 and 0.7004, so that (53) is approximately satisfied. However, neither $\rho$ nor $c/_{c'}$ are large enough to suggest that double sampling will be much more efficient than single sampling. On the other hand, with possible improvement in economics of aging techniques $c/c'$ would be smaller so that single sampling could be used with advantage.

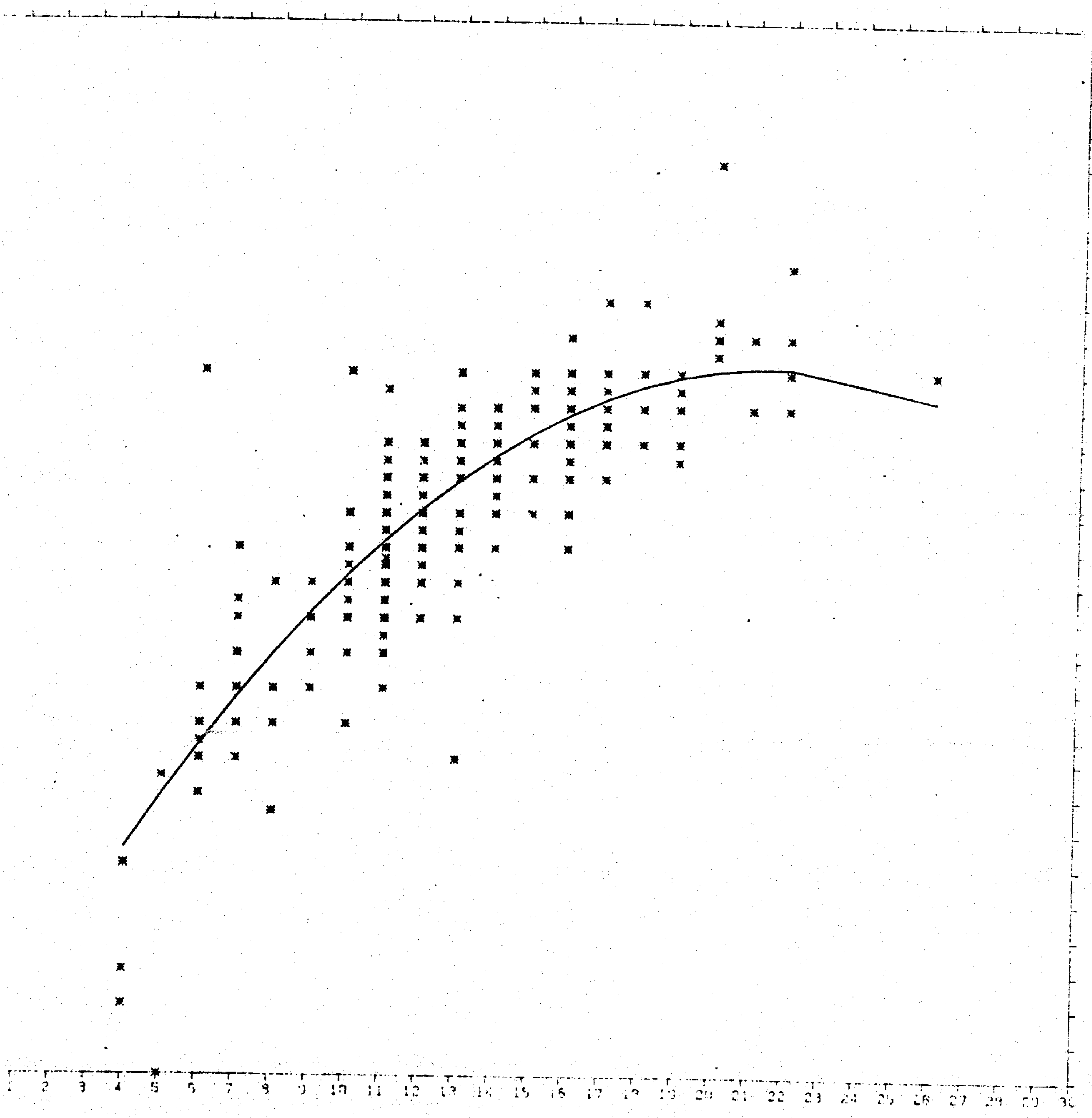
We will illustrate the use of double sampling by analysing 1981 length-age data to estimate the proportion of female widow rock fish in age group eleven at Eureka based on a sample of 444 fish.

Consider the 3 length strata $h = 1, 2, 3$ with stratum boundaries based on quadratic fits of length on age (Figs. D2), 31.5 - 43 cm, 43.5 - 49 cm and 49.5 - 55 cm. (Note this is different than boundaries based on length only.)

$C_o$ = 1.2 mins, $C_1$ = 3.8 mins, $C_2$ = 3.8 mins., $C_3$ = 8 mins.

$w_1 = 0.0653$, $w_2 = 0.5451$, $w_3 = 0.3896$

$s_1 = 0.1825$, $s_2 = 0.4966$, $s_3 = 0.1503$



Age (in years)
Sen D. 2 Regression (1981)

SEN.D2 Regression

where $w_1$, $w_2$ and $w_3$ are the proportion of fish in the sample and $c_o$ is the cost of measuring a fish, $c_1$, $c_2$, $c_3$ are respectively the costs of aging them in the 3 length groups. From Cochran (1977, p. 331) we have

$$v \min(p_{st}) = \frac{1}{c^*} \left[ \sum w_h s_h \sqrt{c_h} + (S^2 - \Sigma w_h s_h^2)^{\frac{1}{2}} \sqrt{c'} \right]^2 \text{ approximately} \quad (5$$

$$= 1.1225/c^*$$

where $p_{st}$ is the estimated proportion and $c^* = E(c_0 n + \sum c_h n_h) = 3179.79$ with $n_1 = 14$, $n_2 = 120$ and $n_3 = 48$, $n = 444$.

The efficiency of double sampling with respect to single sampling is given by

$$v \text{ SRS } (p)/v \min (p_{st}) = 120$$

i.e. double sampling is 20 percent more efficient than single sampling.

Our studies have shown that the best length-age relationship (logarithmic or quadratic) do not change significantly if every other fish is sub-sampled systematically over time. Hence the sample size for estimation of age distribution in a port-year group can be reduced appreciably (without loss in efficiency) by selecting fish for age from length-time strata with the number in each cell proportional to the product of the number of fish and standard deviation of age estimated from earlier studies conducted in the region.

## 11. SUMMARY AND *RECOMMENDATIONS*

1. Random sampling of boat trips is not practicable owing to the uncertainty of their arrival times and the most reasonable assumption is that boats arrive at a port during a month in a random order (pp. 6-7).

2. The current design of post-stratification of a boat sample and sub-sampling clusters from each sort type is not feasible (unless the present strength of staff responsible for data collection is doubled) since some of the categories may be missed in sampling (pp. 4, 11).

3. Estimates of species catch (Widow, Chilipepper and Bocaccio) and by sex and age based on post-stratification of each sample landing by sorts at a port during a month are less efficient than ones based on post-stratification by sort types of arrivals at port-month level (Tables, pp. 24, 26 and 27).

4. *Post-stratification of landings by sorts is recommended by port-month groups.*

5.1. *As far as practicable, selection of a cluster for a market category should be done before any presorting is done at the port either from bins, strap boxes or off conveyor belts.*

5.2. *Clusters should be selected from the conveyor belt where small-market buyers would select fish from the top of bins.*

5.3. *For large-market buyers, clusters should be selected systematically, separated in time from the beginning and end of loading.*

5.4. *A cluster should be selected from one side (of a box all the way down to the bottom) without looking at fish being selected in the process.*

6. *Selection of a random and representative cluster (box) of fish from a boat trip would depend to a large extent on the expertise and experience of the sampler. Hence, the need to have permanent staff at least at the important ports to build up the experience.*

7. Variation (within sort group) in length and age for a species was considerably higher among sample landings than among clusters within landings: also, variation among clusters was not significant compared to variation within clusters (Tables pp. 37-39). This suggests that some presorting is done in the boats before landing.

8. Using the cost function $c = c_1 n + c_2 n\bar{m}$ where $c_1$ is the average cost and $c_2$, the cost of data collection, $n$ and $\bar{m}$ the number of sampled boats and clusters per boat (p.28), the optimum number of clusters per sampled boat for a fixed cost for a sort type is two (pp. 29-35).

9. Emphasis on "25 lbs. cluster weight for small rock fish including widows and 50 lbs. for all others" in the instructions on data collection led in several cases to wide variability among samples and hence to biased and inefficient estimates of species catch owing to violation of the assumption of equal cluster weight required in the current method of estimation of parameters (pp. 8,15).

10. *Two clusters of 25 lbs. each in "small fish" in the rose fish category and 50 lbs. each for all others should be selected from sample landings.*

10.1. *For small landings of widows one or two clusters of 50 lbs. each should be selected depending upon availability (pp. 36,40).*

*The current instructions on the selection of clusters should be* *modified accordingly.*

11.1. *The principal contribution of this report is that a minimum of 4 landings should be sampled for each category from a port-month stratum i.e., approximately one per week (p. 13) and two clusters (as in recommendation 10 above) of 50 lbs. (or 25 lbs.) each should be sampled to provide port-year estimates of species catch (or weight) with a reasonable degree of accuracy. Effort should be directed to sampling more landings for an important category as judged by prior knowledge of the share of its landing to total landing weight for the port-month group.*

11.2. *For a boat landing with more than one important category, the sampler should subsample from as many of such categories as possible; additional sort groups should not be sampled at the cost of gathering less than 2 clusters from a sort group.*

11.3. *If there is more than one landing on a given day, attempt should be made to cover the other landing after completing the requirements of the first landing (p. 40).*

11.4. *If a sort group is infrequently landed, sampling should be directed towards the infrequent sort group, as long as the number of landings for the infrequent sort group is less than four per month (p. 40).*

12. Formulas for the estimation of parameters and their errors have been developed for the general case of unequal cluster size (section 5.5.2) and for the more practical case when cluster weight can be treated as a constant within a sort type (section 5.5.3).

13. The efficiency of the ratio estimator [pp. 41, 42, equation, (44), (48) based on post-stratification by sort groups at port-year level was compared with two others (the unbiased estimator and the ratio estimator based on Jack Knife). The empirical evidence (Table 9) indicates that the ratio estimator of total catch (32) of a species has the smallest coefficient of variation and is, therefore, most efficient amongst others for estimating species catch (and weight).

14.1. The between component of total variation was considerably higher than the within component which was almost negligible (Tables pp. 48-59) indicating that for obtaining efficient estimates of the parameters, a large number of landings with few clusters (two per sort group) per landing should be selected.

14.2. The coefficient of variation (c.v.) for species catch (or weight) varied considerably among port-year groups.

14.3. The c.v.'s for species weight by port-year sort groups were slightly smaller than the corresponding c.v.'s for species catch (p. 52, Table 10.3).

14.4. The c.v.'s for species catch by sex-age group are generally higher than for species without such classification. Thus, for Widow at Eureka and Monterey (1982), the c.v.'s were 7.33 and 6.62 (Table 10) respectively as against higher figures for sex-age groups. Similarly, estimates of 4-year-old Bocaccio (Table 10.7) were subject to very high error compared to estimates of Bocaccio by port-year groups under similar conditions.

14.5. The c.v.'s of combined estimates for species catch by sex-age group for "All ports (Monterey, San Francisco, Fort Bragg, Morro Bay and Eureka)" were reasonably small for some of the dominant sex-age groups (Table 10.6).

15.1. The sample size required to estimate species catch during a year at "All ports" level with a 5 to 7 percent c.v. can be managed by the present staff with a work load as in recommendations (10) and (11) above; at port-year level the accuracy will be somewhat lower i.e., 10 percent (p. 70).

15.2. Sample size for estimating total weight would be somewhat smaller than for estimating total catch to ensure the same level of accuracy (p. 71, Table 11.2).

15.3. Species catch (or weight) by sex and age at the 'All port-year' level can be estimated with a 20 percent c.v. (except for some of the dominant species); to estimate with a 10 percent c.v., we would need more staff, at least at the important ports.

15.4. To estimate total catch of species by sex and age at port-year level with a 10 percent c.v., we would certainly need additional staff at almost all the port groups.

16. Quarterly estimates of catch (or weight) can be obtained by the present staff with a 20 percent c.v., assuming that the population c.v. would be approximately the same for quarterly and annual estimates. In view of the very low precision of quarterly estimates these can at best serve as indicators.

17. For estimating proportion of fish in an age group by double sampling a sample of 444 Widow fish landed at Eureka during 1981, was divided into 3 lenth strata and optimum allocation for age was adopted within strata. It showed that the estimate of proportion in eleven year age group was 20 percent more efficient than single sampling for age when average cost for aging is about six times that for length measurement.

    With possible improvements in technology, single sampling may, however, prove more efficient than at present.

18. *Double sampling for age with the first sample stratified by length and time is recommended for estimation of age composition.*

19. *For obtaining reliable and valid estimates of population characteristics it is essential that good relationship be maintained by the sampler with both the skipper and the buyer. This would depend to a large extent on the experience of the sampler gained in the course of his field work. This emphasizes the need for having permanent staff at least at the important port groups which have too many problems to be handled within a short time.*

## ACKNOWLEDGEMENTS

Acknowledgements are due to the valuable assistance by employees of Tiburon Laboratory during the period of field work (May - September 1983). My special thanks are due to William Lenarz who provided me with information on the problems related to rock fish landings on the California coast, to Candis Cooperrider and to Mark Allen of Humboldt State University for the computations done on the data collected, to the field staff of the California Department of Fish and Game responsible for collection of relevant data and to Norman Abramson, Director of the Laboratory for all the assistance rendered to me during my work at the Laboratory.

My thanks are also due to Vivian Spak of the Department of Statistics and Applied Probability, University of Alberta for assistance in typing the report.

# LITERATURE CITED

Cochran, W.G. 1977. Sampling Techniques, third edition, Wiley and Sons, New York.

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. P.V. Reun. Cons. Int. Explor. Mer. 86: 1-14.

Ketchen, K.S. 1950. Stratified subsampling for determining age distributions. Trans. Am. Fish. Soc. 79 ; 205-212.

Kimura, D.K. 1977. Statistical Assessment of the Age-Length Key. J. Fish. Res. Board. Can. 34; 317-324.

Kutkuhn, J.H. 1963. Estimating absolute age composition of California salmon landings. California Department of Fish and Game, Fish Bulletin No. 120; 47 p.

Mackett, D.J. 1963. A Method of Sampling the Pacific Albacore (Thunnus Germo) Catch for Relative Age Composition F.A.O. Fisheries Reports. No. 6, vol. 3; pp. 1355-1366.

Quinn II, Best, E.A., Bijsterveld & R. McGregor. Sampling Pacific Halibut Landings for Age Compostion: History, Evaluation and Estimation. International Pacific Halibut Commission, P.O. Box 95009, Univ. Station, Seattle, Washington, 98105, U.S.A.

Southward, G.M. 1976. Sampling landings of halibut for age composition. Int. Pacific Halibut Commission Sci. Rep. 58; 1-31.

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.

Tomlinson, Patrick K. 1971. Some Sampling Problems in Fishery Work. Biometrics 27, Sept. pp. 631-41.

Westrheim, S.J., and W.E. Ricker. 1978. Bias in using an age-length key to estimate age-frequency distributions. J. Fish. Res. Board. Can. 35: 184-189.