

SH  
11  
.A2  
S664  
no.90-03  
c.2

# SOUTHWEST FISHERIES CENTER

NATIONAL MARINE FISHERIES SERVICE    SOUTHWEST FISHERIES CENTER    3150 PARADISE DRIVE    TIBURON, CA 94920

JUNE 1990

## AN INTERACTIVE PORT SAMPLING MODEL FOR CALIFORNIA GROUND FISH

by

Keith R. Parker  
and  
Alec D. MacCall





"This report is used to insure prompt dissemination of preliminary results, interim reports, and special studies to the scientific community. The material is not ready for formal publication since the paper may later be published in a modified form to include more recent information or research results. Abstracting, citing, or reproduction of this information is not allowed. Contact author if additional information is required."

SH  
11  
.A2  
5664  
NO.90-03  
C,2

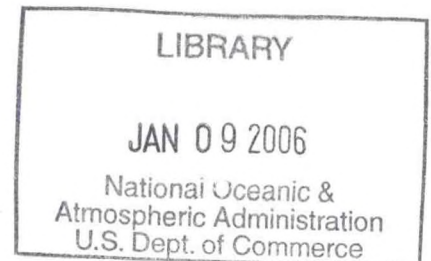
AN INTERACTIVE PORT SAMPLING MODEL FOR CALIFORNIA GROUND FISH

KEITH R. PARKER\*  
and  
ALEC D. MacCALL

TIBURON LABORATORY  
SOUTHWEST FISHERIES SCIENCE CENTER  
3150 PARADISE DRIVE  
TIBURON, CA 94920

SWFSC TIBURON LABORATORY ADMINISTRATIVE REPORT: T-90-03\*\*

JUNE 1990\*\*



\* Data Analysis Group  
3201 Hanson H Road  
georgetown, CA 95634

\*\* A few copies of this Administrative Report were erroneously released as SWFSC Administrative Report T-89-02 in June 1989. The previous version is otherwise identical to the present version, but the previous number and date should not be used.



## AN INTERACTIVE PORT SAMPLING MODEL FOR CALIFORNIA GROUND FISH

This SWFC Tiburon Laboratory Administrative Report consists of a report written by Keith R. Parker pursuant to a contract let by the National Marine Fisheries Service. Alec MacCall of the SWFC Tiburon Laboratory developed the concept and contract specifications, acted as Contracting Officer's Technical Representative, and worked with Keith Parker during its development and implementation. The contract was funded by the NMFS Southwest Region. The contract report is being released as a SWFC Tiburon Laboratory Administrative Report in order to facilitate its distribution and/or citation.

**NOTE:** This report includes a 360K diskette containing the model implemented as a LOTUS 1-2-3 or SYMPHONY compatible spreadsheet.<sup>1</sup> The diskette is available upon request from Alec MacCall, SWFC Tiburon Laboratory, 3150 Paradise Drive, Tiburon, CA 94920.

The model was constructed to prove feasibility and to determine the level of effort needed to produce a working model which could guide allocation of sampling effort. It was successful in both respects. At the present level of implementation, the model already provides useful advice with regard to groundfish port sampling activity in California. However, the model is incomplete in some respects.

- The present model accounts for number of samples, gross landings and average landing size for individual port-areas and calendar quarters. The model accounts indirectly for priorities given to various types of landings (e.g. which boat to sample when two boats arrive simultaneously, one with sablefish and the other with widow rockfish). With the present model, exploration of alternative sampling instructions is possible but tedious.
- The spreadsheet is based on data only for calendar year 1987. More recent data would provide a more current description of sampling performance. Also, a spreadsheet based on a larger base of information could avoid computational limitations caused by very low levels of sampling effort in some port-quarters.



- This spreadsheet does not include sablefish. The information necessary to implement the model for sablefish resides in a different database than that for the rockfish and flatfish, and retrieval was beyond the scope of the present effort. There was also some question of whether the present sampling intensity for sablefish is high enough to support statistical calculations such as these (J. Hightower, NMFS Tiburon Laboratory, personal communication). An attempt to include sablefish in this model is nonetheless warranted.
- Until uniform sampling methods are adopted, different models will have to be constructed for different states. These models would serve the individual states' needs for internal planning, and would also improve documentation of statistical methods.
- A comprehensive model, linking the state models, should be implemented for the entire west coast groundfish sampling. Fishery analyses require coastwide data; accordingly, sampling should be coordinated coastwide.
- The contract report contains other recommendations, including possibilities of simplifying the model based on statistical patterns observed in the data.

The reader is urged to load the spreadsheet on a computer and experiment with the results of changing fishery and sampling parameters. The automatic calculate feature is turned off, so the "calc" key must be used to update the spreadsheet after a change. If a change causes error messages appear, refer to the model documentation; most likely sample size has been altered to be less than 2 or exceeds the number of possible landings for the port/quarter.

1. The National Marine Fisheries Service (NMFS) does not approve, recommend or endorse any proprietary product or proprietary material mentioned in this publication, nor is reference to proprietary products made with the intent to cause the product to be used or purchased because of this NMFS publication.

# PORT SAMPLING MODEL

by

Keith R. Parker  
Data Analysis Group  
3201 Hanson H Rd.  
Georgetown, CA 95634  
(916) 333 - 4001

Ref. Order No. 40ABNF803597  
Req. No. NFFSR300810023



# CONTENTS

	Page Number
1.0 INTRODUCTION.....	1
2.0 ESTIMATING VARIANCE, EXPECTED LANDINGS, AND COEFFICIENT OF VARIATION .....	2
2.1 Estimating Variance .....	2
2.2 Estimating Landings .....	4
2.3 Coefficient of Variation .....	4
2.4 Covariance.....	4
2.5 Number of Sampled Clusters Equal to One .....	6
2.6 Number of Sampled Trips Less Than Four .....	7
3.0 ESTIMATING VARIANCE FOR THE SPREADSHEET MODEL.....	8
3.1 Variance Formulae .....	8
3.2 The Within to Between Ratio $> 0.20$ .....	9
3.3 Dover Sole.....	10
4.0 USING THE SPREADSHEET MODEL.....	11
4.1 Command Screen.....	11
4.2 Sample-Size Allocation Screens .....	13
4.3 Data Summary Screens .....	13
4.4 Analysis Screens .....	15
4.5 Spreadsheet Calculations.....	16
4.6 Error Messages.....	18
5.0 RECOMMENDATIONS .....	19
5.1 More Efficient Programming.....	19
5.2 Are Replicate Clusters Within a Sampled Trip Necessary? .....	20
5.3 Regression Based Estimates of Variance .....	20
5.4 Enumerating All Sampled Landings.....	21

## 1.0 INTRODUCTION

This contract provides a working model to use in examining the relationship between sampling effort and variance in landings of selected fish taxa. The model has been programmed into a Lotus 1-2-3 spreadsheet. By altering inputs of (1) number of sampled landings, (2) total pink ticket landings and/or (3) the average weight of sampled landings, the user can compute estimates of variance and changes in estimated coefficients of variation.

This report is organized as follows. Formulae for estimating landings, variance in landings and coefficient of variation are documented in Section (2). These procedures are based on Sen (1984, 1986). In Section (3) are documented the formulae used for variance estimation in the spreadsheet model. Section (4) documents the spreadsheet model. The spreadsheet is described in terms of screens -- those parts of the spreadsheet viewed by the user. Documentation of the spreadsheet model, using row headings as variable descriptions, also is given in Section (4). Section (5) follows with recommendations based on observations made while analyzing data and building the spreadsheet model.



## 2.0 ESTIMATING VARIANCE, EXPECTED LANDINGS, AND COEFFICIENT OF VARIATION:

### 2.1 Estimating Variance:

Sen's (1986) equations (9) through (13) were used for estimating landings and variance. See Sen's argument (7), page 420, for his justification. Sampled landings were stratified into port-quarters. Table 1 lists the six ports and port groups; the species analyzed in this report are listed in Table 2. Within each port-quarter strata, boat trips were randomly sampled. And within each sampled boat trip at least two clusters, of approximately 50 lbs. each, were sampled from a particular market category. Estimated landings and variances were computed by market category and then summed over ports and quarters to estimate yearly landings and variances, by market category. Using Sen's equation (12), based on his equations (4) through (7), the variance for market category  $j$ , for a port-quarter strata is:

$$(1) \quad V(Y_j) = N_j(N_j - n_j)/(n_j(n_j - 1)) \sum_i (M_{ij}^2 (Y_{ij} - \bar{Y}_j)^2 + (N_j/n_j) \sum_i (M_{ij} (M_{ij} - m_{ij})) s_i^2/m_{ij}$$

$N_j$  is the estimated number of landings for market category  $j$ .  $N_j$  is the quotient of the total pink ticket landings ( $W_j$ ) and the mean weight of a sampled landing ( $w_j$ ):  $N_j = W_j/w_j$ .  $W_j$  is read from COMSUM reports.

$n_j$  is the number of randomly sampled trips for market category  $j$ . Sen suggests  $n_j \geq 4$  (See his comments on page 412). With rare exception, Sen's recommendation was met. These exceptions and their consequence are noted below.

$M_{ij}$  is the estimated number of clusters for market category  $j$  on trip  $i$ .  $M_{ij}$  is the quotient of the weight in landings on trip  $i$  ( $W_{ij}$ ) and the average weight of a sampled cluster ( $\bar{w}$ ).  $M_{ij} = W_{ij}/\bar{w}$ . I used  $w = 50$  lbs. The weight of sampled clusters, for all 1987 clusters, rounded to 50 lbs with a standard error was less than 1 lb. Dover sole,  $w = 30$ , was the only exception.

$Y_{ij}$  is the mean catch per cluster for market category  $j$  on trip  $i$  and is either in pounds or number of fish. Usually  $Y_{ij}$  is the mean of two clusters.

$Y_j^*$  is the mean catch per clusters for market category  $j$ , over all sampled trips.  $Y_j^*$  is in either pounds or numbers of fish.

$M_{ij}^2(Y_{ij} - Y_j^*)^2$  is the weighted between trip sums of squares.  $M_{ij}$  is the weighting factor and, again, is the estimated number of clusters for market category  $j$  on trip  $i$ .

$N_j(N_j - n_j) / (n_j(n_j - 1)) \sum(M_{ij}^2 (Y_{ij} - Y_j^*)^2)$  is the between trip component of the total variance. The ratio  $N_j(N_j - n_j) / (n_j(n_j - 1))$ , and consequently the variance, change inversely to the number of sampled trips ( $n_j$ ). The greater the number of sampled trips, the smaller the variance.

$m_{ij}$  is the number of clusters sampled for market category  $j$  on trip  $i$ . From the beginning of 1985,  $m_{ij}$  usually = 2.

$s_i^2$  is the variance in catch per cluster for market category  $j$ , trip  $i$ .  $y_{ijk}$  is the number of fish (or pounds) per cluster.

$s_i^2 = \sum_k (y_{ijk} - Y_{ij})^2 / (m_{ij} - 1)$ .  $k$  is the counter for clusters within a trip. Usually  $k = 1, 2$ .



$(N_j/n_j) \sum_i (M_{ij} (M_{ij} - m_{ij})) s_i^2/m_{ij}$  is the within trip component of variance. There is no within component of variance for Dover sole since clusters were not replicated.

## 2.2 Estimating Landings:

The estimated landings for market category  $j$  ( $Y_j$ ) is the product of the average weighted catch per trip ( $Y^*$ ) and the estimated number of clusters for market category  $j$  (Sen 1986, equation (11)).

$Y_j^* = \sum_i (W_{ij} * Y_{ij}) / \sum_i W_{ij}$ , or the weighted average number of fish (or pounds) per cluster. Again,  $Y_{ij}$  is the mean catch per cluster for market category  $j$  on trip  $i$  and  $W_{ij}$  is the weight in landings for market category  $j$  on trip  $i$ .

$Y_j = Y^* (W_j/w)$ , or the estimated landings for market category  $j$  for a port-quarter. Again  $W_j$  is the pink ticket landings for market category  $j$ .  $w$  is the weight of a cluster, 50 lbs.

## 2.3 Coefficient of Variation:

$CV_j = (\text{Var}(Y_j))^{1/2} / Y_j * 100\%$ , the standard formula for the coefficient of variation.

## 2.4 Covariance:

Estimated landings and variances were computed by market category (within a port and quarter) and then summed over ports and quarters to estimate yearly landings and variances. When different market categories are sampled on the same

trip, the covariance term derived by Sen (1986; equations 12 and 13) needs to be added to the summed variances:

$$(2) \quad 2 \sum_j \sum_k (\text{Cov}(Y_j, Y_k)), \text{ for market categories } j \text{ and } k.$$

Sen (page 412) suggests that the covariance term be "ignored when the subsamples from different categories are from different boat trips and are, therefore, independent." For 1983 data Sen found no need to estimate covariances "because the sampler failed in almost all cases to subsample from more than one category." Based on a complete analysis of 1987 data I found no instance of different market categories being sampled on the same trip (for any one species). I made the same observation from 25% subsamples of 1985 and 1986 data.

Although not noted by Sen, landings on the same day (of the same species) from different market categories might also lead to covariances. For 1987 data, the only year for which the data was completely analyzed, this never occurred for the species addressed in this report. I made the same observation for 25% subsamples of 1985 and 1986 data.

To carry it a step further, it is possible that landings of a particular species on different days from different market categories might covary. Only in the 3<sup>rd</sup> quarter of 1987 at Morro Bay did landings meet this description. In this case different market categories were landed on different days: both chilipepper and bocaccio were sampled from both the general rockfish category (#250) and the combined bocaccio/chilipepper category (#956). Pink ticket landings ( $W_j$ ) combined these two categories and therefore there is insufficient data to test for



(and adjust for via equation 2) covariances. In estimating variances, I necessarily combined these two categories. It is unlikely that covariances are a problem for landings on different days from different market categories.

There are at least three logistical reasons why covariance did not pose, nor is likely to pose, a serious problem. First, for a particular rockfish species, most trips are landed in one market category only. Second, Sen recommends there be at least 2 subsamples (clusters) from a market category for each sampled trip; after sampling two clusters samplers move on to another boat i.e., trip. Third, as discussed in the previous paragraph, pink-ticket landings are often combined for different market categories, eliminating the opportunity to estimate variances separately.

## **2.5 Number of Clusters Sampled Equal to One:**

Sen (1984; 1986) recommended that market categories on a sampled trip be sampled for a minimum of two clusters ( $m_{ij} \geq 2$ ). With rare exception, this recommendation has been followed since 1985. The exceptions usually occurred when the landings dropped below 1,000 lbs. Since the sampler took only one cluster, the within component of variance was inestimable. Still, I used these data to estimate between trip variance and consequently, total variance. I elected not to eliminate sampled landings with only 1 cluster for the following two reasons. (1) Using the between component of variance (only) to estimate total variance increased sample size and decreased the variance and coefficient of variation. (2) I sacrificed little in eliminating the within component of variance, which is usually less than 20% of total variance.

## 2.6 Number of Sampled Trips Less Than Four:

Sen (1986) suggests  $n_j$  less than 4 (See his comments on page 412) and in 1987 this suggestion was not met 6 out of 12 times for widow rockfish, 3 out of 24 times for chilipepper and bocaccio, and 4 out of 24 times for bank rockfish. It is likely that for  $n_j \leq 4$  estimates of between and within variance are poor. However, note that these low  $n_j$  had little effect on yearly/coast-wide estimates of landings and its variance since low  $n_j$  usually indicated relatively low landings and estimates of variances. Thus, these relatively low estimates of variance comprised only a small fraction of the total (summed) variance and covariance.



### 3.0 ESTIMATING VARIANCE FOR THE SPREADSHEET MODEL

#### 3.1 Variance Formulae:

The spreadsheet model, described in Section 4.0, requires that variances be estimable for different hypothetical numbers of sampled trips ( $n_j$ ). To accomplish this, the variance by market category,  $V(Y_j)$ , was first divided into two parts which are invariant to  $n_j$ : (1) a between factor, defined presently, and (2) a ratio of within variance to between variance. These two parts were combined in the spreadsheet model so that variance could be computed for different numbers of sampled trips ( $n_j$ ), described in the following.

Between factor: the between factor is the average weighted between sums of squares, where the averaging is over the number of sampled trips:

$$\text{Between Factor} = M_{ij}^2(Y_{ij} - Y_j)^2 / n_j$$

The Between Factor is invariant to the number of sampled trips ( $n_j$ ), as noted above. (However, higher  $n_j$  will result in better estimates of the Between Factor.) The between component of variance under any proposed sample size ( $n_j$ ) can be computed as:

$$(3) \quad N_j(N_j - n_j) / (n_j - 1) * \text{Between Factor}$$

The within trip component of variance was on average found to be a smaller component of total variance than the between trip component, usually less than 20% of the total variance. Exceptions are addressed in the following section

(Section 3.2). The within/between ratio is simply the ratio of the within and between components of variance:

$$\text{Within/Between} = \frac{[(N_j/n_j) \sum_i (M_{ij} - m_{ij})^2 s_i^2/m_{ij}]}{[N_j(N_j - n_j)/(n_j(n_j - 1)) \sum_i (M_{ij}^2 (Y_{ij} - Y_j)^2)]}$$

For computing the total variance,  $V(Y_j)$  under varying numbers of sampled trips ( $n_j$ ) the estimated between variance is computed as follows:

$$\text{Hypothetical } V(Y_j) = \frac{N_j(N_j - n_j)}{(n_j - 1)} * \text{Between Factor} (1 + \text{within/between})$$

### 3.2 The Within to Between Ratio > 0.20:

The ratio of the within component of variance to the between component of variance was usually less than 0.20. Exceptions occurred when the mean catch per cluster ( $Y_{ij}$  -- in either numbers or pounds), was high. This usually occurred when the target species or age comprised 50% or more of the sampled clusters. Under these circumstances, fractional deviations in catch per cluster ( $y_{ijk}$ ) are at a maximum and push up the variance in catch per cluster ( $s_i^2$ ). Since under these circumstances, where the within component makes up a large fraction of the variance, sampling additional trips has less of an effect on  $[V(Y_j)]$  than when the ratio is small. The computation of variance for the spreadsheet (above, Section 3.1) model accurately reflects this: increasing the number of sampled trips has little effect on reducing variance and the coefficient of variation.



### 3.3 Dover Sole:

Dover sole made for three exceptions. First, data on aged fish and COMSUM reports were available only for the Eureka, Crescent City, Fields port group. The spreadsheet model only includes this port group. Second, clusters were not sampled. Only the between component of variance was estimable. The estimate of variance in the spreadsheet model is on the between component only. Third, the average weight per sampled cluster equals 30 lbs., and not 50 lbs. as for the other species in this report.

## 4.0 USING THE SPREADSHEET MODEL

The spreadsheet model is divided into four interlinked parts: (1) the command screen, (2) sample size allocation screens, (3) data summary screens and (4) analysis screens. On the command screen the user inputs and views the effect of different sample sizes on coefficients of variation. The sample size allocation screens -- one screen for each species -- partition the total number of samples into the number of samples taken for each species. The data summary screens present estimates of landings, sample size, variance and coefficient of variation by port and quarter. Detailed computations are displayed on the analysis screens. I elaborate on each of these sections in the following. In a final section I document spreadsheet calculations. In the following, standard Lotus 1-2-3 field ranges, such as (B1..Q24) which means all cells in the rectangle from B1 to Q24, are used. In this documentation, ??? indicates landings in either pounds or numbers of fish.

### 4.1 Command Screen:

Figure 1 shows the command screen.

Upon retrieving the spreadsheet file (1987.wk1) the command screen is the first screen seen; it is the top-most screen (A1..G20) in the file. From the command screen the user can change the total sampling efforts by port and quarter and view the effect of these changed efforts on the estimated coefficients of variation. The command screen also shows the total estimated landings for the year over all ports. The command screen lists macro keys which are used to page directly to analysis screens.



1987 efforts in number of sampled trips by port and quarter are listed at the top of the screen (B1..G8): **EFFORT -- NUMBER OF SAMPLES**. This is the number of randomly sampled trips for rockfish, sablefish and dover sole plus 5% of their total (to account for undocumented sampled trips; see Section 5.4). This part of the screen is unprotected and it is here that the user inputs hypothetical efforts. Sampling effort can be either increased or decreased, and the effect of this new effort seen on the estimated CV and the ratio of **New CV/Old CV** -- these fields will be discussed presently and are shown in the lower right corner of the screen.

In the lower right of the screen **New CV** and **Old CV** are shown (E11..G20). The **Old CV** has been estimated on the efforts shown at the top of the screen (B1..G8). The **Old CV** is fixed. The **New CV** differs from the **Old CV** in that it is re-estimated each time sampling efforts are changed. The **New CV/Old CV** fields (G11..G20) are ratios of **New** to **Old CV**'s. Increase efforts and the ratios decrease; decrease efforts and the ratios increase. Note, efforts, in terms of number of sampled trips, have an upper limit and a lower limit. The upper limit is the estimated number of landings,  $N_j$ . The lower limit is 2, that necessary to compute variance. Inputs of efforts that exceed this range result in error messages in **New CV** and **New CV/Old CV** fields (Section 4.6).

Fields (B11..B20) present the taxa names. In field B20 a Template has been added. Information for additional species can be entered into Template. Template can be copied to add additional species. I recommend the user become familiar with the spreadsheet before attempting to add additional species, however.

Fields (C11..C20) and fields (D11..D20) show the age and estimated landings. Details on how landings are estimated are addressed in the sections on Data Summary and Analysis Screens. Fields (A11..A20) show the macro keys. Simultaneously pressing **Alt** and one of the displayed macro keys pages the screen automatically to the species indicated. For instance **Alt C** pages directly to bank rockfish. At **Alt S** a template for new entries has been added.

#### 4.2 Sample-Size Allocation Screens:

Figure 2 shows a sample-size allocation screen.

These screens are directly below the command screen. There is a separate screen for each species. Each species occupies an entire screen. Species are reached by paging down from the command screen. Each screen is divided into two parts. The upper half lists the proportion of total effort (**Proportion of effort**). This is the proportion of the effort shown on the command screen and is the proportion of the total effort that was used for sampling the species being viewed. the proportion of effort is unprotected. The bottom half lists the actual number of sampled landings ( $n_j$ ), the **Number of samples**. The  $n_j$  are connected directly to data summary screens.

#### 4.3 Data Summary Screens:

Figure 3 shows a Data Summary Screen.

Each species has a separate Data Summary Screen. The data summary screens are reached by the menu of macro keys on the command screen (or by paging down). The "Menu - Alt M" at the upper left corner of the data summary screen shows that the main menu (or command screen) is reached by simultaneously depressing **Alt** and **M**.

In the upper center of the screen is shown the species and age. Below this are **Total CV** and **Total Landings**.

**Total CV** = the coefficient of variation for the entire year over all ports.

**Total Landings** = the estimated total landings for the entire year over all ports. Total Landings are in 1,000s of either pounds (lbs) or fish.

The remainder of the Data Summary Screen is divided into information by (1) port and quarter, (2) yearly summaries by port, over quarters (3) yearly summaries by quarter, over ports and (4) a yearly summary over ports and quarters.

Information by port (or port group) and quarter:

**# Sampled Trips** = the  $n_j$  computed in the Sample-Size Allocation Screen.

**Ave Landings (000)lbs** = the mean landings in thousands of pounds (lbs),  $w_j$ . This is unprotected and can be changed by the user.

**Gross Landings (000)lbs** = the pink ticket landings for the market category landed in thousands of pounds. This is unprotected and can be changed by the user.



**Var / Grand Total Var** = the ratio of the variance for the port/quarter and the total variance for the entire year, over all ports and quarters.

**Target Landings (000)???** = the estimated landings for the species in either pounds (lbs) or numbers of fish.

**Est. CV** = the estimated coefficient of variation for the port/quarter.

Information by port (or port group) are summarized over quarter and are shown in the final column on the right.

Information by quarter are summarized over ports for the entire year and are shown in the final set of rows on the bottom.

Information for the entire year, summarized over ports and quarters, is shown in the right most column, bottom set of rows. This information is repeated on the top of the data summary screen and on the command screen.

#### 4.4 Analysis Screens:

Figure 4 shows an Analysis Screen.

Each species has a separate Analysis Screen. Analysis screens are reached by paging right from the Data Summary Screens.

The Analysis Screen is similar to the Data Summary Screen in that it is organized by ports (or port groups) and quarters. The first three rows of the

Analysis Screen can be changed by the user -- they do not result from formulae linked to other cells. However these cells are protected and will need to be unprotected before they can be changed.

**W/in to Between Ratio** = the ratio of the within component of variance to the between component of variance, and is estimated by the user. See 3.1.

**Between Factor (000)** = is described in 3.1 and is estimated by the user.

**Wtd Ave ???/Cluster** = the weighted average pounds (lbs) or number of fish per cluster is estimated by the user. See 2.1.

**Est. Nj** = is computed as in Section 2.1.

**Est. W/in Var (000,000)** = the **W/in Between Ratio** \* **Est Between Var (000)**, following.

**Est. Between Var (000,000)** = the estimated between variance and is computed as shown in Section 2.1.

**Total Var (000,000)** = **Est. W/in Var (000,000)** + **Est. Between Var (000,000)**

#### 4.5 Spreadsheet Calculations:

Several calculations are made in both Data Summary and Analysis Screens. For a given port (or port group), calculations are made by quarter: data from a given quarter for both Data Summary and Analysis Screens are used to calculate estimates of variance, coefficient of variation and catch. In documenting these calculations I use the row headings as they appear on the Data Summary and

Analysis Screens. These calculations are also documented in two other ways: (1) by formulae in Section 2.1, previous and (2) in the spreadsheet cells themselves.

On the Analysis Screen:

$$\text{Gross Total Var (000,000)} = \Sigma \text{ Total Var (000,000)}_{\text{port, quarter}}$$

$$\text{Est. } N_j = \text{Gross Landings (000)lbs} / \text{Ave Landings(000)lbs}$$

Gross landings and Ave Landings come from the Data Summary Screen.

$$\text{Est. W/in Var (000,000)} = \text{W/in to Between Ratio} * \text{Est. Between Var(000,000)}$$

$$\text{Est. Between Var(000,000)} = \text{Between Factor(000,000)} * \frac{[\text{Est. } N_j (\text{Est } N_j - \# \text{ Sampled Trips}) / (\# \text{ sampled trips} - 1) / 1000}$$

The # Sampled Trips comes from the Data Summary Screen.

$$\text{Total Var (000,000)} = \text{Est. Between Var(000,000)} + \text{Est. W/in Var(000,000)}$$

On the Data Summary Screen:

$$\text{Var} / \text{Grand Total Var} = \text{Total Var (000,000)} / \text{Grand Total Var (000,000)}$$

Both Total Var (000,000) and Grand Total Var (000,000) come from the Analysis Screen.

$$\text{Target Landings ??? (000)} = (\text{Wtd Ave ???/Cluster}) * \text{Gross (Landings(000))} / 50$$



The Wtd Ave ???/Cluster, either in lbs or numbers of fish, comes from the Analysis Screen.

$$\text{Est. CV} = (\text{Total Var (000,000)})^{1/2} / \text{Target Landings ???(000)}$$

#### 4.6 Error Messages:

A Lotus 1-2-3 error message (ERR) results when either the number of sampled trips ( $n_j$ ) falls below 2 or the estimated number of landings ( $N_j$ ) rises above that initially calculated in the spreadsheet. In both cases, the variance [ $V(Y_j)$ , equation 1] becomes inestimable. This error message serves as a warning to users that allocations of sample size (computed on the Sample-Size Allocation Screens) have exceeded practical limits, as a direct result of efforts entered on the Command Screen. The Analysis Screen can be consulted to determine where (ports) and when (quarters) the error occurs; the Total Variance will read ERR.

## 5.0 RECOMMENDATIONS

### 5.1 More Efficient Programming:

All computer programs should be linked; all data should be on computer accessible files. This contract concentrated on feasibility and little effort was invested in developing efficient procedures for data analysis. As a consequence, analytical steps consumed time and were prone to transcription error. Checking results -- necessary because of data transmittal -- consumed additional time. In outline, these programming steps were followed:

- (1) Convert a (Fortran generated) formatted ASCII, with headers and subheaders, into a more efficiently analyzable file. Check for data entry errors.
- (2) Read pink ticket data from an unformatted ASCII file. Transmit by hand to analysis program.
- (3) Run analysis program. Output results to hard copy for checking and transmittal.
- (4) Transmit result of analysis program to a Lotus 1-2-3 spreadsheet model.

If estimating the effect of efforts on the variance is to be routine, I strongly recommend that software be developed to accomplish the above steps within a single program. This will take time to develop, but will be worth the effort. Some of the DBMS packages that work directly with Lotus 1-2-3 (like Oracle) might be useful.

## 5.2 Are Replicate Clusters Within a Sampled Trip Necessary?

Less variable estimates of catch might result by switching efforts from sampling within trip clusters to sampling more trips. Sen (1986) strongly recommends that replicate clusters be taken within a trip. Replicate clusters allow the between trip variance to be estimated on means and therefore helps to reduce the between trip variance. Sen's recommendation was instituted in 1985. Given that a sufficient number of trips or proportion of total number of landings ( $N_j$ ) are sampled, I believe that Sen's recommendation is a good one.

However, I'm not convinced that taking replicate clusters within a sampled trip is the most efficient way to sample, with regards to achieving minimum variance. I base this recommendation on the following two observations: (1) within trip variance is often relatively small compared to between trip variance and (2) the total number of sampled trips is often small, say less than eight. This recommendation presumes that boats go unsampled because the sampler is busy taking within trip replicates. Interviews with samplers and analysis of historical data might be useful in evaluating this recommendation. This scheme might prove useful at Bodega and San Francisco, where the number of sampled trips per market category most frequently falls below 4.

## 5.3 Regression Based Estimates of Variance:

It may be possible to estimate variance with a simple regression model. Figures 5 and 6 show the relationship between  $\log(\text{landings (000)})$  and  $\log(\text{Variance (000,000)})$  for chilipepper and bank rockfish. Regression coefficients were significant ( $\alpha = 0.05$ ). The 0-intercept  $\beta$ 's were not significantly different,  $\beta \approx 1.4$ .



Figure 7 shows a combined regression for chilipepper and bank rockfish. Only port-quarters with at least 10 sampled trips were used in generating these regression results. The regression for Bocaccio was not significant.

In order to estimate and fully test  $\beta$ , I recommend that data from 1985 onward, with at least 8 sampled trips per quarter, be used. This regression approach may be good enough for purposes of experimental design and allocating sampling effort.

Also, there may be a better way to estimate the Between Factor, or some other sample-size invariant estimate of sums of squares. The Between Factor, as shown in Section 2.7.1, is an intuitive estimate. A maximum likelihood, method of moments, or other formal estimate can probably be derived.

#### **5.4 Enumerating All Sampled Landings:**

The total number of sampled landings used in this report -- shown on the Command Screen, Section 3.1 -- are for rockfish, dover sole (at Eureka/Crescent/Fields) and sablefish, plus 5% of their total. Other trips were sampled, usually at the direction of CDF&G staff. These trips are not easily accounted for and based on discussions with CDF&G staff 5% seems an appropriate adjustment. All trips should be documented so that total sampling efforts can be addressed with regard to allocation.

Note that in 1987, efforts were allocated in appropriate proportions among ports: ports with the highest landings -- those in the north and in the south -- receive

the greatest sampling effort. Thus  $n_j$  taken in the Eureka Area / Fort Bragg, and Morro Bay / Monterey appropriately exceed those taken for the central coast.

## References

- Sen, A. R. 1986. Methodological problems in sampling commercial rockfish landings. U.S. Fish. Bull. 84:409 - 411.
- Sen, A. R. 1984. Sampling Commercial Rockfish Landings in California. NOAA Technical Memorandum NMFS. NOAA-TM-NMFS-SWFC-45.



Table 1. Ports and port groups.

	<u>Port Code(s)</u>
Crescent City / Eureka / Fields	201, 220, 217
Fort Bragg	223
Bodega Bay	473
San Francisco (Area)	440, 452
Santa Cruz / Moss Landing / Monterey	550, 592, 593
Morro Bay	602, 606

Table 2. Species .

	<u>Code(s)</u>
Bank Rockfish ( <u>Sebastes rufus</u> )	2399, 2368
Bocaccio ( <u>S. paucispinis</u> )	2334
Chilipepper ( <u>S. goodei</u> )	2320
Widow Rockfish ( <u>S. entomelas</u> )	2316
Dover Sole ( <u>Microstomus pacificus</u> )	3110

Figure 1. Command Screen. Shaded area is unprotected.

		EFFORT -- NUMBER OF SAMPLES			
PORT:		Q1	Q2	Q3	Q4
Eureka/Fields/Crescent		36	78	76	28
Fort Bragg		39	34	46	30
Bodega Bay		10	6	17	24
S.F./Half Moon Bay		0	8	57	28
S.C./Moss/Monterey		16	18	22	7
Morro Bay		14	19	32	24

		1987 ESTIMATES				
GoTo:	TARGET SPECIES:	AGE:	Landings (000)	New CV:	Old CV:	New CV/ Old CV
Alt Z	Bocaccio	lbs. All	2264.7	5.4%	5.4%	1.00
Alt X	Chilipepper	lbs. All	2251.3	5.9%	5.9%	1.00
Alt C	Bank	lbs. All	1368.4	12.1%	12.1%	1.00
Alt V	Widow	#'s 5	39.0	48.8%	48.8%	1.00
Alt B	Widow	#'s 8	230.5	7.0%	7.0%	1.00
Alt N	Widow	#'s 12	116.3	8.8%	8.8%	1.00
Alt A	Dover Sole	#'s 9	490.8	20.2%	20.2%	1.00
Alt S	Template	lbs. All	1380.0	11.4%	11.4%	1.00

Figure 2. Sample-Size Allocation Screen. Shaded area is unprotected.

## Proportion of effort -- Bocaccio

PORT:	Q1	Q2	Q3	Q4
Eureka/Fields/Crescent	0.224	0.232	0.370	0.616
Fort Bragg	0.206	0.179	0.325	0.427
Bodega Bay	0.286	0.317	0.298	0.455
S.F./Half Moon Bay	0.000	0.357	0.265	0.388
S.C./Moss/Monterey	0.635	0.560	0.544	1.000
Morro Bay	0.806	0.741	0.759	0.870

## Number of samples -- Bocaccio

PORT:	Q1	Q2	Q3	Q4
Eureka/Fields/Crescent	8	18	28	17
Fort Bragg	8	6	15	13
Bodega Bay	3	2	5	11
S.F./Half Moon Bay	0	3	15	11
S.C./Moss/Monterey	10	10	12	7
Morro Bay	11	14	24	21



Figure 3. Data summary screen. Shaded areas are unprotected.

Menu - Alt M

TARGET: BANK -- TOTAL					
Total CV =					12.1%
Total Landings lbs (000) =					1368.4
Eureka/Fields/Crescent:	Q1	Q2	Q3	Q4	Year
# of Sampled Trips:	0	18	28	17	-
Ave Landing (000)lbs:	0	10.6	9.6	11.0	-
Gross Landings (000)lbs:	0	1210.9	1705.6	813.9	-
Var / Grand Total Var =	0	0.000	0.000	0.264	0.265
Target Landings lbs (000) =	0	4.1	5.5	128.3	137.8
Est. CV =	0	70.7%	65.7%	66.1%	61.6%
Fort Bragg:	Q1	Q2	Q3	Q4	Year
# of Sampled Trips:	8	0	15	13	-
Ave Landing (000)lbs:	12.0	0	11.2	13.2	-
Gross Landings (000)lbs:	359.5	0	574.3	596.8	-
Var / Grand Total Var =	0.000	0	0.012	0.002	0.014
Target Landings lbs (000) =	3.5	0	19.2	10.1	32.9
Est. CV =	64.0%	0	95.5%	68.3%	60.0%
Bodega Bay:	Q1	Q2	Q3	Q4	Year
# of Sampled Trips:	3	0	5	11	-
Ave Landing (000)lbs:	3.8	0	0.7	16.3	-
Gross Landings (000)lbs:	294.4	0	138.1	244.1	-
Var / Grand Total Var =	0.222	0	0.002	0.000	0.224
Target Landings lbs (000) =	186.5	0	18.5	0.0	205.1
Est. CV =	41.7%	0	42.5%	0.0%	38.1%
S.F. / Half Moon Bay:	Q1	Q2	Q3	Q4	Year
# of Sampled Trips:	0	3	15	11	-
Ave Landing (000)lbs:	0	16.1	15.6	16.5	-
Gross Landings (000)lbs:	0	386.7	424.1	674.6	-
Var / Grand Total Var =	0	0.047	0.006	0.004	0.057
Target Landings lbs (000) =	0	49.2	24.9	23.6	97.7
Est. CV =	0	72.5%	53.1%	41.6%	40.2%
S.C. / Moss / Monterey:	Q1	Q2	Q3	Q4	Year
# of Sampled Trips:	10.0	10.0	12.0	7.0	-
Ave Landing (000)lbs:	8.3	9.3	9.9	8.7	-
Gross Landings (000)lbs:	356.3	319.1	513.5	414.1	-
Var / Grand Total Var =	0.022	0.017	0.151	0.166	0.357
Target Landings lbs (000) =	47.1	40.6	165.7	306.3	559.7
Est. CV =	52.0%	53.0%	38.8%	22.0%	17.6%
Morro Bay:	Q1	Q2	Q3	Q4	Year
# of Sampled Trips:	11.0	14.0	27.0	21.0	-
Ave Landing (000)lbs:	7.5	9.7	9.4	5.7	-
Gross Landings (000)lbs:	270.1	260.1	257.8	233.5	-
Var / Grand Total Var =	0.048	0.020	0.000	0.015	0.083
Target Landings lbs (000) =	88.9	121.1	11.2	113.9	335.2
Est. CV =	40.4%	19.4%	11.4%	17.8%	14.2%
Over All Ports:	Q1	Q2	Q3	Q4	Year
Var / Grand Total Var =	0.292	0.084	0.173	0.451	1.000
Target Landings lbs (000) =	326.1	215.0	245.0	582.3	1368.4
Est. CV =	27.3%	22.3%	28.0%	19.0%	12.1%

Figure 4. Analysis screen.

Grand Total Var (000,000):		27205		
	Q1	Q2	Q3	Q4
W/in to Between Ratio:	0.00	0.14	0.21	0.00
Between Factor (000):	0.00	11.40	10.90	26982.50
Wtd Ave lbs/Cluster:	0.00	0.17	0.16	7.88
Est. Nj:	0.00	114.41	176.84	74.28
Est. W/in Var (000,000):	0.00	1.07	2.23	7.18
Est. Between Var (000,000):	0.00	7.40	10.63	7175.37
Total Var (000,000):	0.00	8.46	12.86	7182.55
	Q1	Q2	Q3	Q4
W/in to Between Ratio:	0.47	0.00	0.47	0.06
Between Factor (000):	36.40	0.00	1704.50	371.00
Wtd Ave lbs/Cluster:	0.49	0.00	1.67	0.85
Est. Nj:	30.08	0.00	51.42	45.33
Est. W/in Var (000,000):	1.63	0.00	107.38	2.72
Est. Between Var (000,000):	3.45	0.00	227.99	45.31
Total Var (000,000):	5.09	0.00	335.37	48.03
	Q1	Q2	Q3	Q4
W/in to Between Ratio:	0.00	0.00	0.01	1.77
Between Factor (000):	2087.30	0.00	6.13	0.07
Wtd Ave lbs/Cluster:	31.68	0.00	6.71	0.01
Est. Nj:	77.58	0.00	202.76	14.94
Est. W/in Var (000,000):	6.04	0.00	0.61	0.00
Est. Between Var (000,000):	6037.88	0.00	61.44	0.00
Total Var (000,000):	6043.92	0.00	62.06	0.00
	Q1	Q2	Q3	Q4
W/in to Between Ratio:	0.00	0.04	1.06	0.49
Between Factor (000):	0.00	4841.20	3597.30	528.50
Wtd Ave lbs/Cluster:	0.00	6.36	2.94	1.75
Est. Nj:	0.00	24.04	27.18	40.90
Est. W/in Var (000,000):	0.00	48.97	90.19	31.66
Est. Between Var (000,000):	0.00	1224.31	85.08	64.61
Total Var (000,000):	0.00	1273.28	175.27	96.27
	Q1	Q2	Q3	Q4
W/in to Between Ratio:	0.04	0.04	0.00	0.00
Between Factor (000):	3652.50	4841.00	21856.70	13899.90
Wtd Ave lbs/Cluster:	6.61	6.36	16.13	36.99
Est. Nj:	43.01	34.17	51.91	47.81
Est. W/in Var (000,000):	23.05	17.77	4.12	4.52
Est. Between Var (000,000):	576.23	444.23	4116.56	4519.88
Total Var (000,000):	599.28	462.00	4120.68	4524.40
	Q1	Q2	Q3	Q4
W/in to Between Ratio:	0.01	0.08	5.24	0.01
Between Factor (000):	14397.20	19385.40	479.50	10181.50
Wtd Ave lbs/Cluster:	16.46	23.29	2.17	24.39
Est. Nj:	35.81	26.76	27.51	40.69
Est. W/in Var (000,000):	12.79	40.71	1.36	4.08
Est. Between Var (000,000):	1279.45	508.93	0.26	407.94
Total Var (000,000):	1292.24	549.65	1.62	412.02

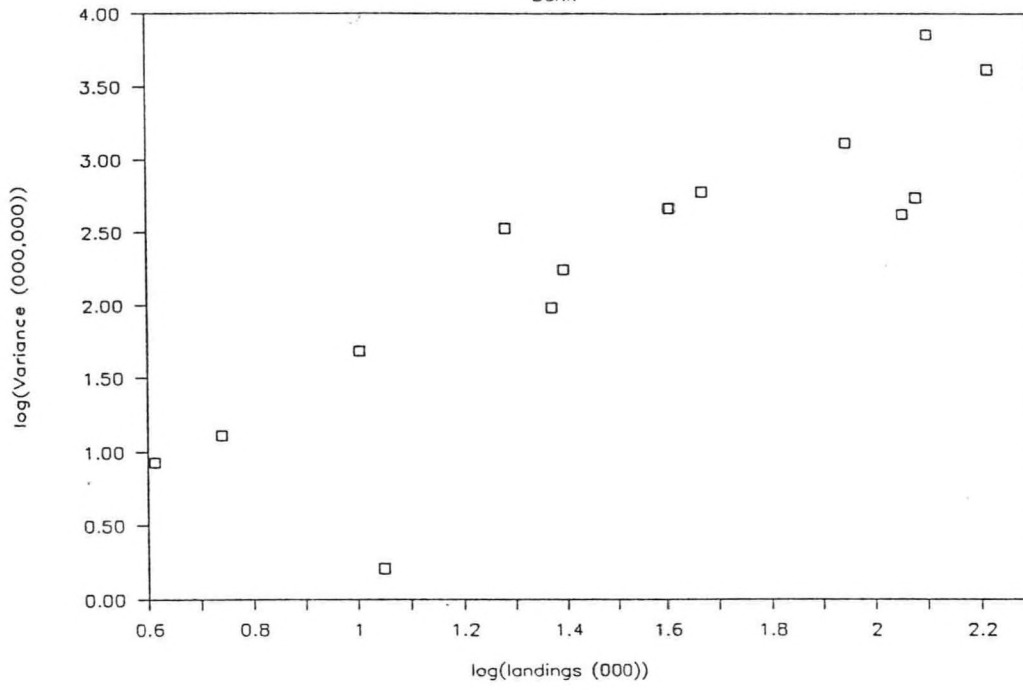
Figure 5  
Bank

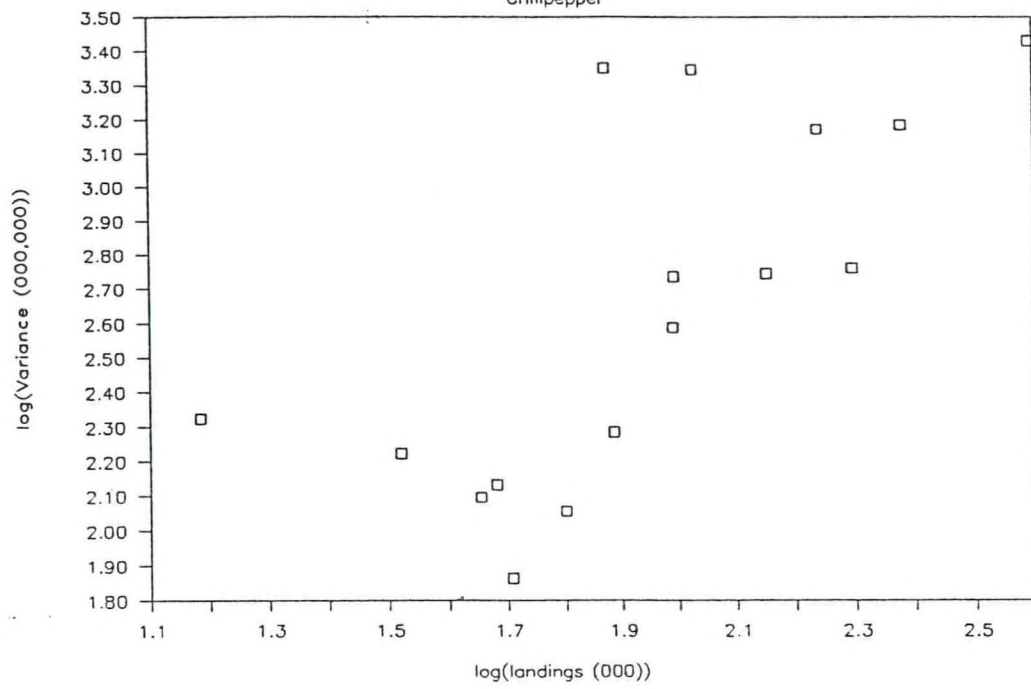
Figure 6  
Chilipepper



Figure 7

Bank and Chilipepper

