NAFO Sci. Coun. Studies, 28: 25–53

Analysis of Data from Bottom Trawl Surveys

Stephen J. Smith
Department of Fisheries and Oceans, Bedford Institute of Oceanography
P.O. Box 1006, Dartmouth, Nova Scotia, Canada B2Y 4A2

Introduction

Bottom trawl surveys provide a major source of fisheries independent information on abundance, species composition and basic biological data for the groundfish communities in the NAFO area (Doubleday, 1981). Most of these surveys use a stratified random design with strata boundaries defined by depth ranges, species-specific distributions and management areas.

The statistical properties of quantities commonly estimated from stratified random trawl surveys such as mean number caught and total numbers in the population are derived from finite population or design-based theory (Cochran, 1977; Smith, 1990; Thompson, 1992). These statistical properties do not require that the observations or estimates made from the observations follow any particular frequency distribution. All properties such as bias and standard errors are developed assuming repeated sampling from a finite population of sample units (e.g. trawl sites) and therefore are functions solely of the survey design.

The design basis for survey estimates has not always been appreciated, especially by those who criticize the use of these estimates for survey data. There is a mistaken belief by many (e.g. Simard *et al.*, 1992; Ecker and Heltshe, 1994) that the assumption of spatial independence is required for these estimates, especially the variance estimates, to be appropriate. The following example reproduced from Smith and Robert (1997) illustrates the design basis for deriving properties of estimates from finite populations. Consider a survey site which can be characterized as a 40×40 grid giving a population of 1600 possible trawl stations. For this site I have generated two very different populations. In the first case a highly skewed distribution without spatial structure was used to provide simulated catch values for each trawl site. In the second case the "catches" were given a strong spatial structure by using a spherical variogram (sill = 4, range =10) and Gaussian noise to generate the data. Random samples of size 10 and 30 were taken from the population of 1 600 trawl sites. Means and their standard errors were calculated for the simulated catches from each of the two populations. For an exact numerical evaluation of the bias and standard error of the mean all possible combinations of sample size 10 and 30 would need to be used. The number of combinations for sample size 10 exceeds 10^{28} and therefore an approximation of randomly choosing 5 000 samples was used for each population and sample size. The population means and standard errors in design-based theory are defined here as the mean and standard errors for the 1 600 population values. Comparisons of the sample values derived from the 5 000 replications with the population values in Table 1 show that the underlying distribution and "spatial" structure of the data have no effect on the unbiasedness of the mean or the accuracy of its standard error.

Typically, means, totals and their respective variances from survey data have been calculated using the survey design but when these data have been analysed for other purposes, model-based methods (e.g. regression, ANOVA) which ignore the complex sampling structure are generally used. The stochastic basis for an analysis method assuming independently and identically distributed observations will not be appropriate when the survey design implies different sampling probabilities over strata. Problems associated with applying standard model-based methods to survey data and methods for incorporating the design into the analysis are discussed in Skinner *et al*. (1989).

There is also a very practical reason for emphasizing the survey design in the analysis of the survey data. The stratified mean or total abundance is used to track changes in a fish stock. Given that these estimates have the design built into them, it only makes sense that this design also appear in the methods that we use to analyse the survey data for other purposes so that findings are comparable with the very abundance index we are trying to say something about.

The fact that the problems with ignoring survey design have only received attention recently means that there is a paucity of methods for the analysis of survey data which incorporate the survey design. This section of the course presents not only the standard descriptive estimates such as means and variances

TABLE 1. Results of the simulation comparing estimates and population values from two finite populations of N = 1 600 sampling units Smith and Robert (1997). The first population was generated as 1 600 random numbers from a Weibull distribution (shape = 30, scale = 1) multiplied by 1 600 Bernoulli random variables with P (z = 1) = 0.75. The second population was constructed as a normally distributed random field on a 40× 40 grid with a spherical variogram function (sill = 4, range = 10 units). Each finite population was resampled 5 000 times (without replacement) for sample sizes of 10 and 30.

	Weibull Model		Spherical Model	
	n = 10	n = 30	n = 10	n=30
Population Quantities				
$\bar{Y}$	28.92	28.92	0.00	0.00
SE ($\bar{Y}$)	9.04	5.18	0.62	0.36
Sample Estimates				
Mean $[\bar{y}]$	28.84	28.97	0.00	0.00
Mean$[\text{se } \bar{y}]$	9.05	5.19	0.62	0.36

but also introduces design-based analytic methods for survey data. These latter methods aid the researcher in evaluating the survey design, exploring for patterns in the survey data and investigating associations between the survey catches for a particular species and ancillary variables such as near-bottom temperature or depth. The unifying theme in all of the above is the incorporation of the survey design into the analysis methods.

This presentation is structured so that the basic theory for each of the methods is described followed by examples. These examples were analyzed using an S-PLUS library (Statistical Sciences, 1995) written by the author. Information on this library is presented in Appendix A and B.

Estimation/Analysis

All of the data contained here are from Canadian groundfish surveys which use stratified random designs. Histories of surveys in the NAFO area, including the Canadian surveys are given by Doubleday (1981), while Halliday and Koeller (1981) discuss surveys on the Scotian Shelf. The sample unit for the survey is defined as the area over the bottom covered by a trawl of a specific width towed at 3.5 knots for a distance of 1.75 nautical miles. The positions of these sample units or sets are selected randomly before the cruise for each stratum. An example of a stratification scheme is that used for the Scotian Shelf (Fig. 1). These strata are primarily based on depth boundaries of 91, 183 and 366 m (originally 50, 100 and 200 fathoms) with further delineation of the strata boundaries reflecting species/stock distributions or management areas (Doubleday, 1981; Halliday and Koeller, 1981).

Means and Variances

The following definitions will be needed for quantities associated with the trawl surveys in any one year.

n_h = the number of hauls or sets sampled in stratum h (h = 1, ..., L),

n = $\sum_{h=1}^{L} n_h$, the total number of sets sampled.

N_h = the total number of possible sets in stratum h,

N = $\sum_{h=1}^{L} N_h$, the total number of possible sets in the survey area.

f_h = n_h / N_h, the sampling fraction in stratum h.

W_h = N_h / N, the proportion of the area in stratum h.

y_{hi} = the number of fish caught in set i and stratum h,

$\bar{y}_h$ = $\sum_{i=1}^{n_h} y_{hi} / n_h$, the estimated mean abundance in stratum h,

and

s_h^2 = $\sum_{h=1}^{L} (y_{hi} - \bar{y}_h)^2 / (n_h - 1)$, the estimated variance in stratum h.

Quantities such as N_h are usually defined to be constant over a survey series except when new trawl nets are introduced or new definitions of swept area are derived from research on trawl dynamics and fish behaviour.

The example data set used here consist of catches of haddock from the July 1988 ground fish survey of the eastern Scotian Shelf (strata 40–66; Fig. 1). A summary of the information on the number of haddock caught for this cruise by stratum is presented in Table 2. This survey is typical of many in the Western

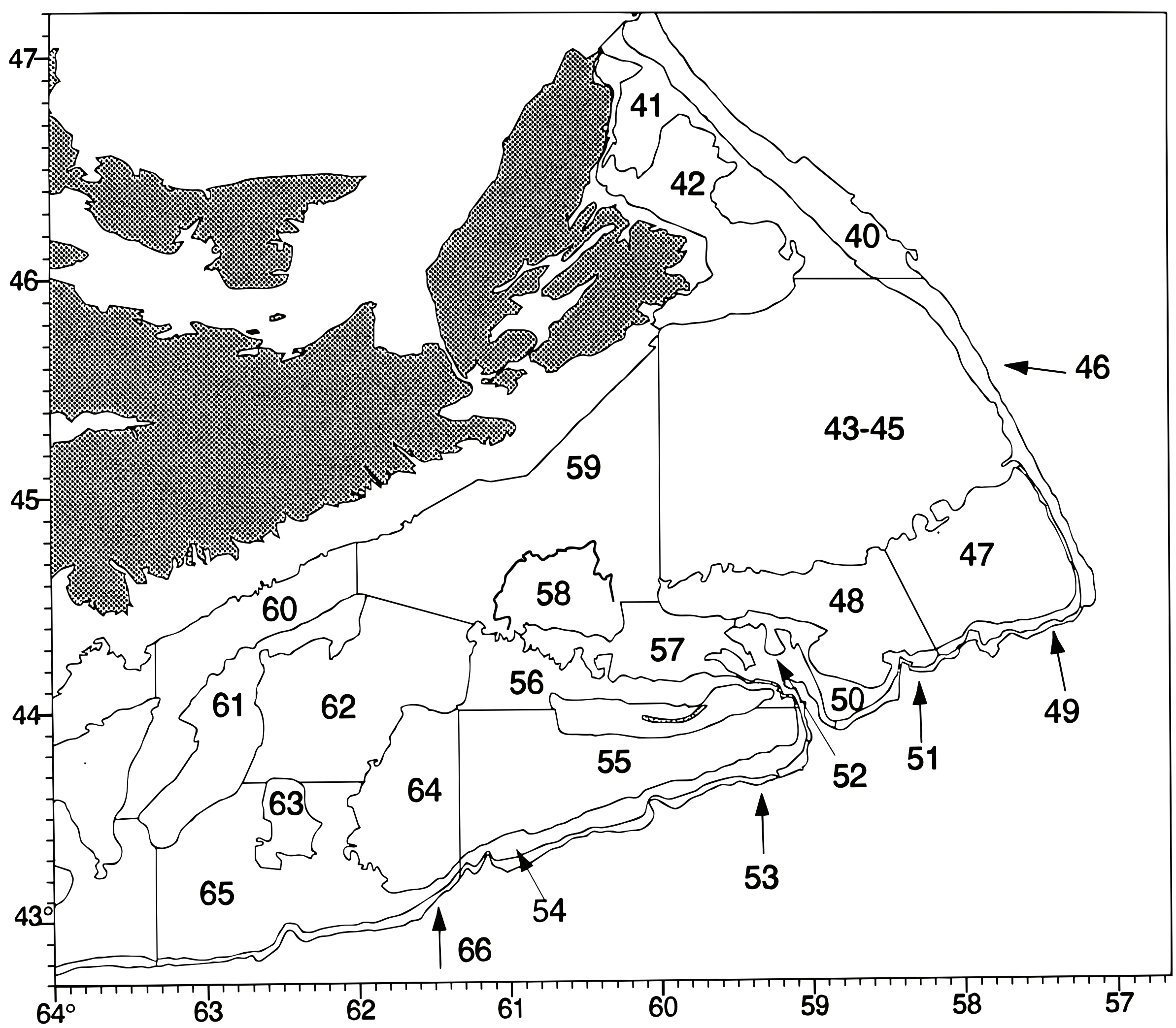

Fig. 1. Stratification map for the Scotian Shelf surveys conducted in July (1970–present). Stratum boundaries are based primarily on depth ranges (originally measured in fathoms). The numbers on the map identify the individual strata.

TABLE 2. Summary of quantities used in calculating the stratified mean and variance for the number of haddock caught in the 1988 Eastern Scotian Shelf Survey. The sample size, mean and standard deviation for each stratum are given by n_h, $\bar{y}_h$ and s_h, respectively. The proportion of the total survey area in each stratum is given in the column headed by W_h.

Strata (h)	n_h	W_h	$\bar{y}_h$	s_h
40	6	0.0294	0.34	0.84
41	4	0.0318	7.40	5.50
42	7	0.0457	0.77	0.86
43	4	0.0419	0.26	0.51
44	4	0.1247	1.29	2.57
45	4	0.0325	0.00	0.00
46	3	0.0156	0.00	0.00
47	6	0.0513	24.67	45.03
48	5	0.0460	0.00	0.00
49	2	0.0046	17.73	11.32
50	3	0.0122	75.72	85.38
51	2	0.0047	2.06	2.91
52	2	0.0110	55.70	77.40
53	2	0.0082	0.00	0.00
54	2	0.0159	45.80	44.40
55	7	0.0674	94.90	59.60
56	6	0.0303	985.90	2212.80
57	2	0.0258	18.00	25.50
58	3	0.0209	77.90	118.90
59	6	0.1000	24.90	48.80
60	3	0.0427	18.60	17.20
61	2	0.0367	0.00	0.00
62	4	0.0672	2.70	4.80
63	2	0.0096	89.70	26.30
64	5	0.0412	109.30	110.50
65	8	0.0757	63.90	65.60
66	2	0.0072	1.50	2.20

North Atlantic by having many strata and few sets per stratum (2–8). Additionally, the haddock appear to be abundant in just a few of the strata with the largest catches (and variance) coming from just one stratum (56).

The stratified mean abundance and its associated variance (Cochran, 1977; Thompson, 1992) are estimated as, respectively,

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h \tag{1}$$

and

$$\widehat{Var}(\bar{y}) = \sum_{h=1}^{L} \frac{N_h}{N^2} (N_h - n_h) \frac{s_h^2}{n_h} \tag{2}$$

The square root of the variance of an estimate is referred to as the standard error of that estimate.

There are two important aspects of equation 2 that may not be readily apparent from the formula. In the first place, the variance that is being measured here concerns the effectiveness of estimating the mean catch over all N_h units, $\bar{Y}_h$, with the sample mean. That is, the expected value over all strata of $(\bar{y}_h - \bar{Y}_h)^2$ for all distinct samples of size n_h repeatedly chosen from the N_h possible sample units within each stratum. This formulation is unaffected by any spatial structure that may exist for the species being captured, although temporal stability of the fish distribution is implicitly assumed. Secondly, this estimate gives the variance of the sample mean as a predictor of what may be expected to be caught in those sites where trawls were not made (see for e.g. Smith, 1990). The more sample units that are observed, the smaller $(N_h - n_h)$ will be and the better the sample mean will be as a predictor of fish catch for those trawl set sites that were not sampled.

The stratified mean of 56.151 haddock per tow was one of the highest in the series to date (Zwanenburg *et al.*, MS 1995) but this estimate also had one of the highest standard errors (27.733). Total abundance $\hat{Y}$ is estimated by assuming that the stratified mean will predict the catch made at each of the N total possible trawl sites. Therefore, $\hat{Y}$ is estimated as $N\bar{y}_{st}$ with variance $N^2 \widehat{Var}(\bar{y})$. In the case of the haddock dataset, the estimated total abundance (in numbers) was 149 772 466 haddock with variance $(73972677)^2$.

Sampling texts generally suggest using parametric confidence intervals for the mean which are constructed by assuming that under repeated sampling the $\bar{y}_{st}$ have a normal or Student's t distribution (Cochran, 1977, p. 95–96). Due to the stratified design, the overall or effective degrees of freedom for the Student's-t multiplier are estimated as,

$$df_e = \frac{\left(\sum_{h=1}^{L} g_h s_h^2\right)^2}{\sum_{h=1}^{L} \frac{g_h^2 s_h^4}{n_h - 1}} \tag{3}$$

where $g_h = N_h (N_h - n_h) / n_h$. This method is valid even if the variances differ between strata. However, the method does require the very strong assumption of the y_{hi} being normally distributed so that the individual terms of the sum in the denominator are estimates of the variance of s_h^2. This assumption is rarely met in practice.

The large amount of variability resulted in the effective degrees of freedom for the haddock example being quite small (5.2409) despite the fact that a total of 106 trawl sets were made. The large standard error and small degrees of freedom resulted in an extremely wide confidence interval with limits –14.164 and 126.466. The parametric confidence interval is constructed assuming that the stratified mean will have a symmetric distribution. Therefore, the very large upper limit resulted in the lower limit being less than zero.

Evaluating the Design

The precision of any statistical estimate is proportional to the amount of information that we have on the process that we are measuring. In the case of simple random sampling those estimates of the mean based on larger sample sizes will tend to have greater precision than those based on smaller sample sizes. In this case the amount of information available is measured by sample size.

For complex survey designs, information is also contained in the design variables (e.g. strata). Therefore, the precision of the stratified mean or efficiency of the stratified design is evaluated by comparing it to a situation where there are no strata, that is, simple random sampling.

Gavaris and Smith (1987) and Smith and Gavaris (1993a) evaluated stratified random trawl surveys by taking the difference between the variance of the mean from the stratified random design with that assuming a simple random sample for the same data. A positive difference between the two variances indicates that the stratified design resulted in a smaller variance for the mean and hence the stratification contained useful information about the process being measured.

The difference between the two variances is estimated by:

$$\widehat{Var}(\bar{y}_{srs}) - \widehat{Var}(\bar{y}_{st}) = \sum_{h=1}^{L}\left(\frac{1}{n} - \frac{W_h}{n_h}\right) W_h s_h^2 + \left(\frac{N-n}{n(N-1)}\right)\left(\sum_{h=1}^{L} W_h(\bar{y}_h - \bar{y}_{st})^2 - \sum_{h=1}^{L} W_h (1 - W_h) \frac{s_h^2}{n_h}\right) \tag{4}$$

Gavaris and Smith (1987) expressed the difference as a percentage of the simple random sampling variance which is estimated by,

$$\widehat{Var}(\bar{y}_{srs}) = \sum_{h=1}^{L}\left(\frac{1}{n} - \frac{1}{N}\right) W_h s_h^2 + \left(\frac{N-n}{n(N-1)}\right)\left(\sum_{h=1}^{L} W_h (\bar{y}_h - \bar{y}_{st})^2 - \sum_{h=1}^{L} W_h (1 - W_h) \frac{s_h^2}{n_h}\right) \tag{5}$$

The difference between the two variances in equation 4 can be decomposed into two components. The first term on the right is termed the allocation component and measures the contribution of the scheme for allocating the number of trawl sets to each stratum. This term will be positive, zero or negative depending upon whether the numbers of sets were allocated in proportion to the stratum variance, stratum size or in an arbitrary manner.

The second term, the strata component determines whether the variance between strata is larger than that within strata. The larger this difference, the larger the amount of information that the strata boundaries contain with respect to the distribution of the fish.

In the case of the haddock data the efficiency was positive with the variance of the stratified mean being 47.2% smaller than the variance of the simple random sample mean. Most of this difference came from the allocation component (44.7%). A look at the stratum details in Table 2 shows that many of the more variable strata did receive the larger sample sizes and allocation was probably close to being proportional to the strata variances (also referred to as optimal or Neyman allocation).

The strata component (2.5%) was quite small indicating that the strata boundaries contained very little information about the distribution of haddock. Smith (MS 1991) found that in general the stratified design was efficient for haddock in the 1980–90 groundfish trawl surveys of the eastern Scotian Shelf. In addition, the allocation component was almost always much larger than the generally small strata component.

Smith and Gavaris (1993a) report on redesigning the March eastern Scotian Shelf survey for cod using historical spatial distributions to design the strata boundaries. Evaluation of this new survey design after five years of use indicated that the new design was more efficient than the previous design. However, most of the gain came from the allocation scheme and not the stratification scheme, despite using the cod distribution to design the stratification. The new design consisted of fewer strata than the old design (11 *versus* 24) allowing for more flexibility in the set allocation scheme.

How does the improvement in the precision using modifications to the survey design comparing to just increasing the sample size? Doubling the sample size for the haddock example assuming all other things being equal would result in decreasing the original standard error by 29%. However, if we knew *a priori* which strata accounted for what proportion of the total variance and assigned the stratum sample sizes accordingly, then again assuming all other things equal, the standard error would be reduced by 67% (using equation 5.27 in Cochran, 1977). That is, a much larger gain without the increasing the cost of the survey itself but requiring much more information about the distribution of the population.

With the exception of the efficiency estimates, all of the calculations presented so far are standard for stratified surveys in the NAFO areas. The main focus of these calculations is to estimate quantities from the survey for assessing abundance trends and for further application such as in tuning sequential population analysis. The efficiency estimates are more in the spirit of evaluating the survey design. In the next subsection methods are presented for exploring the survey data which incorporate the survey design.

Exploratory Functions

While no particular statistical distribution need be assumed for survey data, the empirical cumulative frequency distribution of the observations may be of interest in that the resulting curve may thought of, in the predictive sense, as giving some indication of where the observations from the unsampled trawl sites might lie (Jones and Bradbury, 1993). In standard applications where there are n observations and simple random sampling, each observation is assigned a probability of 1/n when constructing the empirical cumulative distribution function (cdf). However, for stratified random designs the probability depends upon what stratum the observation is in and therefore the cdf is calculated as (Chambers and Dunstan, 1986),

$$F(t) = \sum_{h=1}^{L} \sum_{i=1}^{n_h} \frac{W_h}{n_h} I(y_{hi}) \tag{6}$$

where

$$I(y_{hi}) = \begin{cases} 1 \text{ when } y_{hi} \leq t; \\ 0 \text{ otherwise} \end{cases}$$

and the t are the y_{hi} in increasing order.

The cdf for the haddock data shows a highly skewed distribution with a large proportion of zeros (greater than 40% of the observations) and a very long right-hand tail (Fig. 2). This right-hand tail points out an extremely large catch of 5 496 haddock. In fact, this is the largest catch of haddock ever in the history of this survey to the present day. This catch was made in stratum 56 where the largest mean and variance also occurred. The next largest catch of 309 haddock was made in the same stratum. The fact that no sets had between 309 and 5 496 haddock in them raises doubts about the usefulness of assuming a smooth curve such as suggested in Fig. 2 to predict possible catches in the unsampled areas. Instead, it looks like we could be dealing with a mixture here with most catches being in the 0 to 309 range and possibly a very small number of schools up in the 5 400 range. Smith (1997) used non-parametric density estimates to compare the effects of different hypothetical mixture distributions on the resultant estimates of the mean, standard error and confidence intervals.

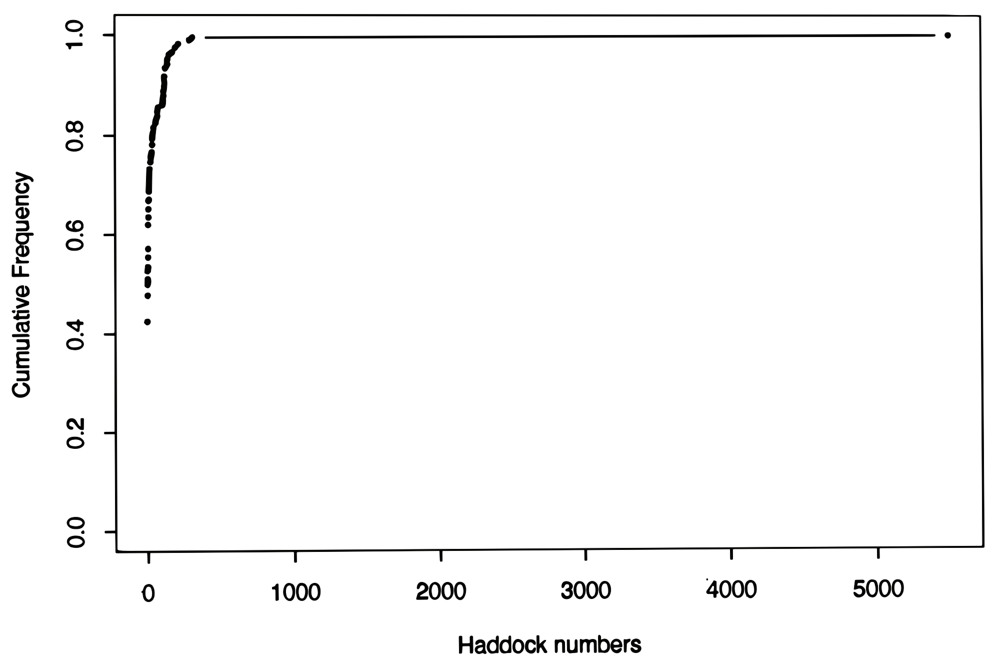

Fig. 2. Empirical cumulative distribution plots for numbers of haddock caught from each tow in eastern Scotian Shelf (July 1988). Note that the maximum observation for the eastern Scotian Shelf survey was 5 496 haddock.

What impact will this large set have on the estimates? Recall from equation 1 that each observation contributes $(W_h y_{hi})/n_h$ to the stratified mean. Hence this factor expressed as a proportion of the $\bar{y}_{st}$ could be considered a measure of the influence of any one observation on the stratified mean. If any of the $(W_h y_{hi} / n_h)/\bar{y}_{st}$ are unusually large, this will indicate that the respective y_{hi} are influential on the magnitude of the stratified mean. If none of the observations are particularly influential then the $(W_h y_{hi} / n_h)/\bar{y}_{st}$ should be roughly equal. The quickest way to get an appreciation of the distribution of the $(W_h y_{hi} / n_h)/\bar{y}_{st}$ with respect to the y_{hi} is to construct a simple scatter plot of the former *versus* the latter. Such an "influence" plot (Fig. 3) for the haddock data[1] highlights the large catch of 5 496 which is shown to account for 49% of the stratified mean. The strata labels for the five most influential sets are indicated on the plot. Removing or replacing the largest catch with the next largest catch (Smith, 1981) or a function of the remaining observations (Moyer and Geissler, 1991) could result in a less variable estimate of the mean but this estimate of the mean will have unknown bias.

Bootstrap Confidence Intervals

Confidence intervals based on applying the Central Limit Theorem to the distribution of means from design-based theory are not always very useful, depending upon the range of sample sizes and observa-

[1] The y_{hi} are cube-root transformed when plotted to scale the plot.

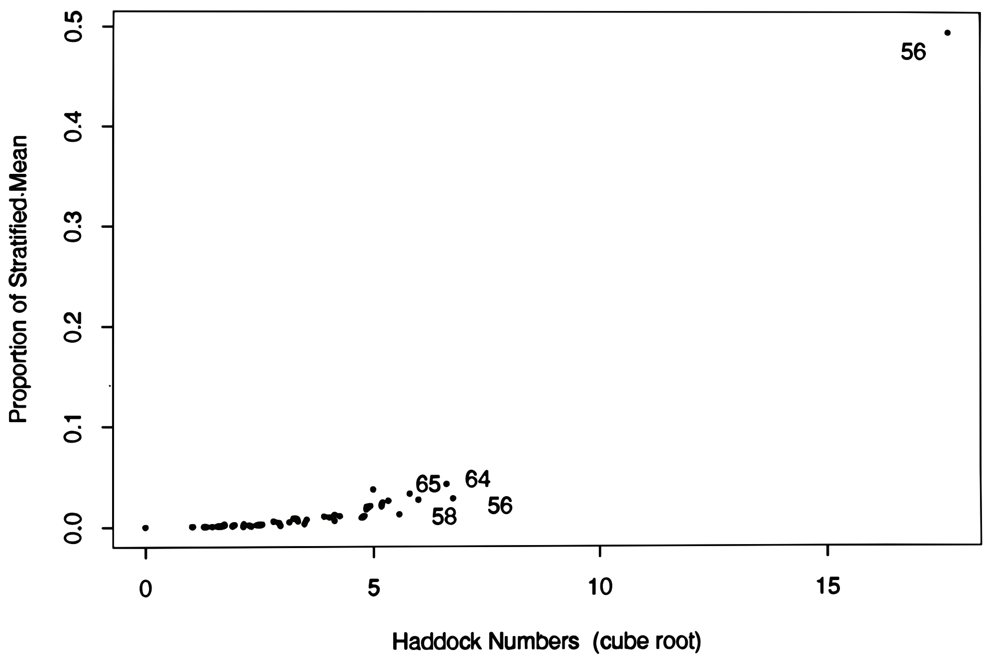

Fig. 3. Influence plot for the number of haddock from eastern Scotian Shelf survey, July 1988. The five most influential sets are identified by their respective stratum label.

tions (Smith, MS 1988; Smith, 1997). The haddock example presented in the previous section shows how poorly such confidence intervals can do when sample sizes are small and variances are large within strata.

The bootstrap technique developed by Efron (1982) offers an alternative approach to constructing confidence intervals. Bootstrap methods have been used in a number of fisheries survey applications (e.g. Kimura and Balsiger, 1985; Sigler and Fujioka, 1988; Robotham and Castillo, 1990; Pelletier and Gros, 1991; Buckland *et al.*, 1992; Stanley, 1992; Smith and Gavaris, 1993b) as a means of substituting computational power for theoretical analysis in situations where complex survey designs have been used. The bootstrap offers a natural way of modelling survey estimates given that its basis is very similar to that of the randomization basis for finite population theory. The basic idea of the bootstrap is to treat the original sample of size as the target population and the original estimated statistic (e.g. mean, ratio) as the population parameter to be estimated. Repeated sampling with replacement of size from the original data set is used to create a large number of new pseudo-samples. Estimates of the parameter of interest are made for each of the pseudo-samples and the empirical distribution of these resultant estimates are used to characterize the distribution of the original statistic. Bias is evaluated with respect to the difference between the average of all of the bootstrap estimates and the original estimate. In cases where it is known that the original statistic and its standard error are unbiased, the bootstrap estimate and its standard error estimate must also be so.

The bootstrap method was originally introduced for the simple random sampling case where all sample units had an equal probability of being chosen. In complex survey designs, such as stratified random designs, the sample units have the same probability of being chosen within any one stratum but different strata may have very different sampling intensities. Therefore, modifications need to be made to the bootstrap procedure to reflect the survey design (Rao and Wu, 1988; Kovar *et al.*, 1988; Smith and Gavaris, 1993b; Smith, 1997).

An obvious modification is to simply independently resample n_h observations with replacement from each of the h strata. This method, referred to here as the **Naïve** method, has been shown to result in biased estimates of the variance of the stratified mean (Rao and Wu, 1988; Smith, 1997). Two other methods have been proposed to eliminate this kind of bias. The **Rescale** method introduced by Rao and Wu

(1988) resamples m_h $(m_h < n_h)$ observations and then rescales the resulting observations to obtain the correct variance. Rao and Wu (1988) suggested setting $m_h = n_h - 3$, based upon comparing the bootstrap third moment with the unbiased estimate of the third moment of $\bar{y}_{st}$. In a recent study, Kovar *et al.* (1988) compared the merits of setting $m_h = n_h - 3$ to $m_h = n_h - 1$ and found that their results favoured the latter over the former. Smith and Gavaris (1993b) reported similar results for a limited example.

While the rescaling method has been found to provide more unbiased estimates of the variance than the Naïve method, it has the disadvantage of being a more computer-intensive method.

The Bootstrap-with-replacement (**BWR**) method is a special case of the mirror-match method introduced by Sitter (1992). The BWR is similar to the Naïve method with the addition of a randomization step for choosing either n_h or $n_h - 1$ resamples within each stratum. Smith (1997) compared all three methods on trawl survey data and found that the **BWR** and **Rescale** methods performed equally well in generating bootstrap estimates with variances very close to the stratified variance of the mean.

Smith (1997) found that for the case of the haddock survey used here, at least 750 replications were required before the variance estimates converged to their expected values for either **BWR** or **Rescale** methods. Reference levels for bootstrap resample sizes have been given as 20 to 50 for estimating standard errors (p. 273, Efron and Tibshirani, 1993), however the results in Smith (1997) indicate that much higher levels are required when dealing with complex survey designs.

Bootstrap estimates of the stratified mean and variance from the haddock data using each of the three resampling methods are presented in Table 3. A total of 1 000 replications was used for each estimate. While the bootstrap estimates of the stratified mean provided unbiased estimates of the observed stratified mean, note that the variance estimate for the Naïve method resulted in a biased estimate as advertised.

Bootstrap confidence intervals do not require a distributional assumption for their construction and thus can be used to evaluate the standard normal theory intervals. If the bootstrap estimates exhibit the correct variance then it is assumed that the confidence intervals calculated from the bootstrap estimates

TABLE 3. Summary of results from calculating the stratified mean, median of the bootstrap distribution, variance and 95% confidence intervals for the number of haddock caught in the 1988 eastern Scotian Shelf survey (from Smith, 1997). The confidence intervals for the Original method were calculated assuming a Student's-t distribution (St). Confidence intervals for the bootstrap mean were calculated using the percentile (PC), bias-corrected (BC) and bias-corrected accelerated (BC_a) methods. The Length column refers to the length of the confidence intervals. The Shape column refers to a measure of symmetry (symmetric: Shape = 0) given by Efron (1992). Note that the expected values for the variance of the Naïve bootstrap for the eastern Scotian Shelf was 640.30.

					95% Confidence Interval			
Method	Mean	Median	Variance	Type	Lower	Upper	Length	Shape
Original	56.15	*	769.1	St:	-14.20	126.50	140.70	0.00
Naïve Bootstrap	56.19	54.89	646.4	PC:	23.69	112.20	88.51	0.61
				BC:	24.64	115.71	91.08	0.56
				BC_a:	25.88	139.62	113.70	0.94
Rescale Bootstrap								
$(m_h = n_h - 1)$	56.54	58.93	768.6	PC:	21.94	124.82	102.88	0.58
				BC:	20.35	99.84	79.49	1.40
				BC_a:	22.26	127.12	104.90	1.87
BWR Bootstrap	56.41	58.21	769.7	PC:	22.36	125.46	103.10	0.63
				BC:	20.96	102.79	81.84	1.09
				BC_a:	22.83	128.28	105.40	1.45

* by definition equivalent to the mean for the Student-t distribution.

will also be correct.

Three of the more common bootstrap confidence interval methods are the percentile method (**PC**), the bias-corrected method (**BC**), and the bias-corrected and accelerated method (**BC_a**) (Efron and Tibshirani, 1993).

The PC method assumes that the frequency distribution of the bootstrap estimates fully describes the distribution function of some estimator $\hat{\theta}$, $G(s) = \text{Prob}\,(\hat{\theta} \le s)$. That is,

$$\widehat{G}(s) = \sum_{i=1}^{B} I_i / B,$$

where

$$I_i = \begin{cases} 1, \text{ if } \hat{\theta}_i^* \le s; \\ 0, \text{ otherwise.} \end{cases}$$

and $\hat{\theta}_i^*$ denotes the i th bootstrap estimate of θ (e.g. stratified mean) and B (i = 1, ..., B) denotes the number of bootstrap replications. Upper and lower α confidence intervals are calculated as $\widehat{G}^{-1}(1-\alpha/2)$, and $\widehat{G}^{-1}(\alpha/2)$, respectively.

The BC method (Efron, 1981) introduces a correction to the PC method to account for differences between $\hat{\theta}$ and the median of the frequency distribution. The α upper and lower confidence intervals for this bias-corrected method are obtained as $\widehat{G}^{-1}\{\Phi(z^{(1-\alpha/2)} + 2z_0)\}$, $\widehat{G}^{-1}\{\Phi(z^{(\alpha/2)} + 2z_0)\}$, respectively; where Φ is the standard normal distribution function, z^t is the t-th percentile of the standard normal distribution and $z_0 = \Phi^{-1}\left\{\widehat{G}^{-1}\left(\frac{\#\{\hat{\theta}_i^* < \hat{\theta}\}}{B}\right)\right\}$ (where #{ } refers to a count of how many times the condition within the braces is true). The term z_0 will be equal to zero when the bootstrap estimate and the median of the $\hat{\theta}_i^*$ are equal; the bias-corrected and percentile methods are equivalent in this case.

Finally, the normal approximation often used in constructing confidence intervals assumes that the mean is independent of the variance. However, in many cases and certainly for trawl survey catch data this assumption doesn't appear to hold. A further correction factor, a, referred to as the acceleration is introduced as a measure of the rate of change of the standard error of $\hat{\theta}$ with respect to the true parameter value θ measured on a normalized scale. The α upper and lower confidence intervals for the BC_a method are given as, respectively,

$$\widehat{G}^{-1}\left\{\Phi\left(z_0 + \frac{z_0 + z^{(1-\alpha/2)}}{1 - \hat{a}\,(z_0 + z^{(1-\alpha/2)})}\right)\right\}$$

and

$$\widehat{G}^{-1}\left\{\Phi\left(z_0 + \frac{z_0 + z^{(\alpha/2)}}{1 - \hat{a}\,(z_0 + z^{(\alpha/2)})}\right)\right\}.$$

Note that setting both z_0 and $\hat{a}$ to zero would result in the formula for the percentile confidence intervals. The acceleration is estimated here using the jackknife-based estimate (page 186, Efron and Tibshirani, 1993) modified for stratified random surveys by Smith (1997).

The 95% upper and lower limits are provided in Table 3 for the each of the three confidence interval methods and three resampling methods. The haddock data exhibited considerable skew in their distribution (Fig. 2). Citing general findings in the literature Cochran (1977) suggested that $1-\alpha$ confidence intervals for the mean based on the normal assumption will behave as follows when the original data have a skewed distribution. First the area covered between the upper and lower limits will be less than $1-\alpha$. Additionally, the probability of the mean being less than the lower limit will be less than $\alpha/2$ while the probability of being greater than the upper limit will be greater than $\alpha/2$. If the bootstrap limits are to be an improvement then their respective upper and lower limits should be greater than those from the Student-t distribution. The **BC_a** limits appear to be more reasonable than those given by the Student-t method and correct the Student-t limits in the expected way. However, simulation results reported by Smith (1997) suggest that **BC_a** method may over-correct when used with the resampling methods presented here.

The **Length** column in the table refers to the length of the confidence interval. The bootstrap confidence intervals are narrower than that from the Student-t method (140.63 = 126.466 – (–14.164)). The right-most column presents a measure of **Shape** which is calculated as the natural log of the ratio of the upper limit minus the median to the median minus the lower limit (Efron, 1992). A confidence interval which is symmetric around the median will have a shape measure of zero while a shape greater than zero indicates distributions skewed to the right. The bootstrap confidence intervals are highly skewed, no doubt to accommodate the large catch of 5 496 haddock in stratum 56.

The quantiles of the bootstrap estimates can be plotted against those for a standard normal distribution to assess how the distribution of the bootstrap deviates from a symmetric distribution. This kind of plot referred to as a Q-Q plot (Chambers *et al.*, 1983) is presented in Fig. 4 for the results from applying the BWR method to the haddock data. The diagonal line indicates where the observed points should be if the bootstrap estimates exhibited a normal distribution. However, these estimates show much heavier tails than expected for a normal distribution. In addition, the points also exhibit discontinuities where few or no bootstrap estimates of the stratified mean were obtained. These discontinuities separate segments which going from the bottom to the top of the figure give the distribution of the stratified mean when the catch of 5 496 haddock was not chosen or chosen once, twice or three or more times for the bootstrap sample in its stratum. The fact that this one catch would occur more than once in a bootstrap sample is not too surprising given that there were only six sets taken in strata 56 and resampling was done with replacement. Therefore, the bootstrap method is not immune from the effects of very large catches. For surveys where there are more sets per stratum and no extreme catches the Q-Q plot will tend to fall along the diagonal line (See Georges Bank haddock survey example in Smith, 1997).

Extremely large catches in a survey may simply be due to chance events but also may be due to unusually large aggregations of fish forming for predictable biological reasons. In the next section, techniques for investigating for environmental reasons for large catches are presented.

Analysis of Environmental Associations

The idea that relationships between environmental conditions and fish catch may underlie some of the patterns we see in abundance indices from surveys, is not a new one, but has received a lot of attention lately. Recent studies have shown that a number of groundfish species exhibit strong and consistent

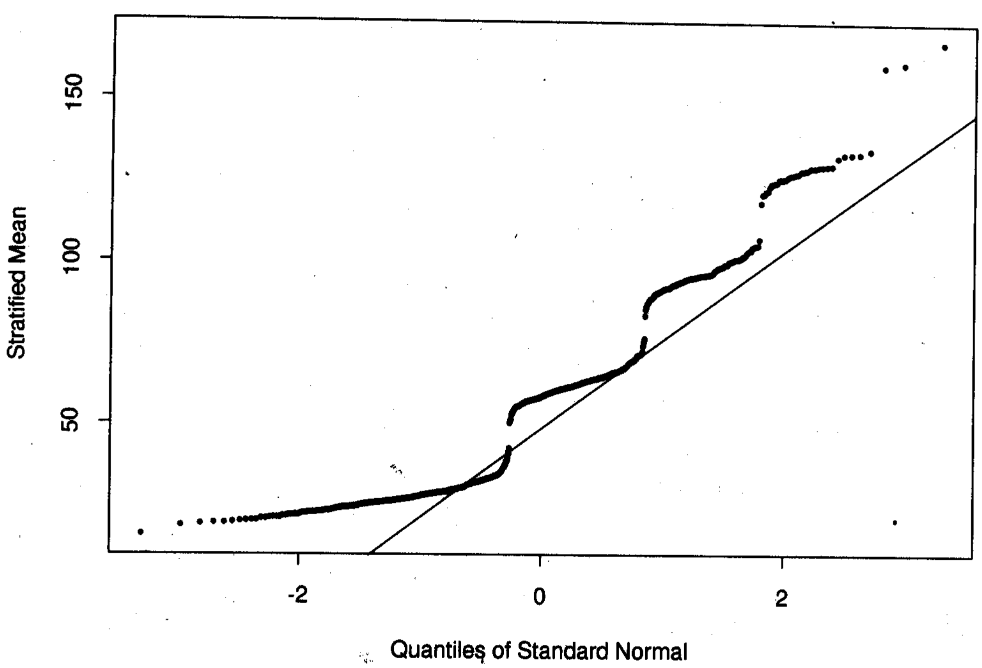

Fig. 4. Quantile-quantile plot for bootstrap estimates of stratified mean number of haddock from eastern Scotian Shelf survey, July 1988. BWR resampling method was used.

associations for a particular range of temperatures or salinities or both (Scott, 1982; Smith *et al.*, 1991; Sinclair, 1992; D'Amours, 1993; Page *et al.*, 1994; Perry and Smith, 1994; Smith *et al.*, 1994). There is also ample evidence that the amount of water impinging on the bottom that exhibits the seemingly preferred characteristics can fluctuate over time and may therefore affect the availability or catchability of the species being surveyed to the trawl (Smith *et al.*, 1991; Page *et al.*, 1994; Smith *et al.*, 1994; Smith and Page, 1994). Unsuitable conditions in the water near the bottom may keep fish off-bottom and unavailable to the trawl, or affect the catchability of the fish through metabolic considerations (Swain and Kramer, 1995) or fish swimming speed/trawl towing speed interactions (Smith and Page, 1996). In addition to these studies Perry *et al*. (1994) provide evidence for associations between bottom type and fish distribution.

In this section the catch-weighted cumulative distribution function method is presented to explore for associations between fish catch and concurrently measured environmental variables (Smith, 1990; Perry and Smith, 1994). The first step in the method involves characterizing the general frequency distribution of environmental variable, x_{hi} (e.g. near-bottom temperature) as observed during the survey by constructing its empirical cumulative frequency distribution. This is done using the same approach in equation 6 for fish catch, i.e.,

$$G(t) = \sum_{h=1}^{L} \sum_{i=1}^{n_h} \frac{W_h}{n_h} I(x_{hi}) \tag{7}$$

where

$$I(x_{hi}) = \begin{cases} 1 & \text{when } x_{hi} \le t \\ 0 & \text{otherwise} \end{cases}$$

Next, we determine what proportion of the stratified mean was associated with each of the points of G(t).

$$K(t) = \sum_{h=1}^{L} \sum_{i=1}^{n_h} \frac{W_h}{n_h} \frac{y_{hi}}{\bar{y}_{st}} I(x_{hi}) \tag{8}$$

If large proportions of the stratified mean are associated with a narrow range of environmental conditions, then this suggests a strong association between the distribution of the fish species and those conditions. In this case G(t) and K(t) would show strong differences between each other. On the other hand if the proportions of the stratified mean were randomly distributed with respect to the entire range of the environmental conditions then we could assume that there was no association and G(t) and K(t) would be almost identical.

The empirical cumulative distribution function G(t) and the catch-weighted function K(t) are given in Fig. 5, with the Habitat line referring to the cumulative distribution function for temperature. Haddock seem to have a definite aversion to cold water. Note that while approximately 50% of the water sampled had near-bottom temperatures less than or equal to 4°C, less than 18% of the stratified mean was found at these temperatures. Most of the haddock were found at temperatures warmer 4°C, with the large catch of 5 496 haddock (recall that this catch accounted for 49% of the mean) being associated with a temperature of 10.98°C.

How strong is the association between haddock and temperature? Could the differences that we observe in Fig. 5 be due to random chance? Perry and Smith (1994) developed a test statistic similar to the Kolmogorov-Smirnov test statistic to measure the difference between G(t) and K(t). That is, calculate the maximum vertical difference between the two curves as,

$$\max|G(t) - F(t)| = \max\left|\sum_h \sum_i \frac{W_h}{n_h} \left(\frac{y_{hi} - \bar{y}_{st}}{\bar{y}_{st}}\right) I(x_{hi})\right| \tag{9}$$

Given the complex sample design it is unlikely that tabled values for the Kolmogorov-Smirnov test would be appropriate here. Instead, Perry and Smith (1994) designed a randomization test to evaluate the significance of the difference between the two curves as measured by equation 9.

The maximum vertical distance between the two curves was 0.4899. The randomization test using 4 000 replications to create the null hypothesis distribution suggests that a difference this large is fairly significant (p = 0.055).

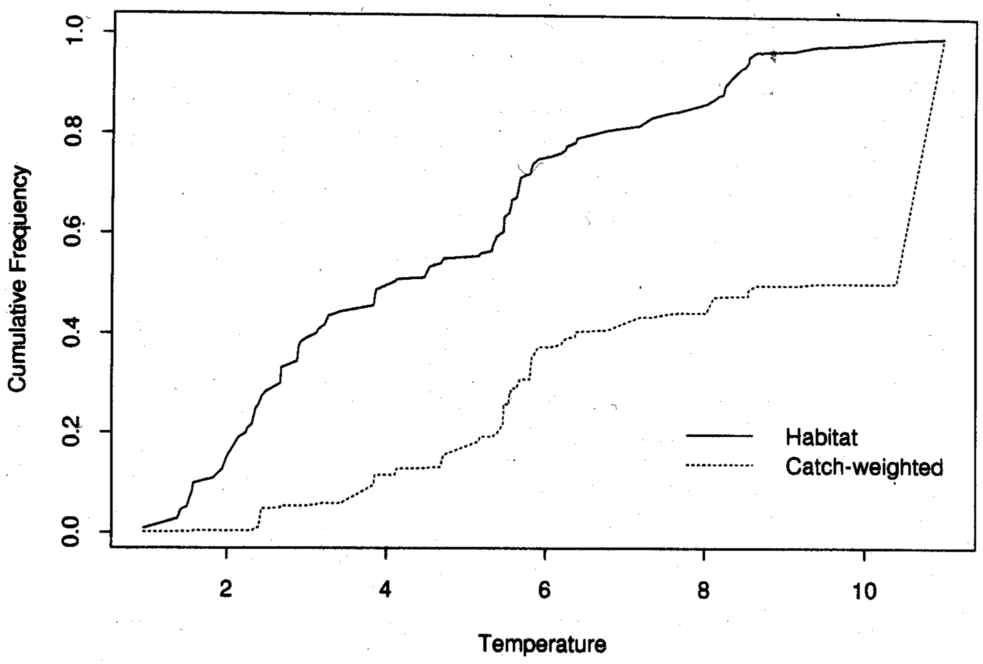

Fig. 5. Association plot for temperature and stratified mean number of haddock from eastern Scotian Shelf survey, July 1988. Habitat refers to cumulative frequency of temperature during survey. Catch-weighted represents the proportion of stratified mean associated with each temperature observation.

Perry and Smith (1994) found that (for 2000 replications) p-levels for association tests between haddock and temperature ranged from <0.01 to 0.09 for surveys in July (1979–84). The temperature of the bottom water in NAFO Div. 4VW has ranged from less than 0.0 up to 12.0°C during the March and July survey cruises. An analysis of haddock catches for all of the surveys in the 1970 to 1993 period showed that haddock are consistently caught in water with temperatures greater than 2.0°C (Smith *et al.*, 1994).

Summary

Most presentations of survey data in stock assessment reports provide the stratified mean or total abundance and their respective standard errors. When virtual population analyses are used to estimate fishing mortality, the stratified mean or total is generally used as a tuning index through ADAPT or some other method (Mohn and Cook, 1993). However, there is much more information in the survey than simply the stratified mean. The main points to be gleaned from the analysis of the haddock survey data are:

1. The stratified mean number caught has a high variance associated with it,

2. The survey design is efficient for haddock relative to a simple random sample design but all of the gain in precision came from the allocation scheme and not the stratification,

3. The survey data are highly skewed with one very large catch which accounted for almost half of the stratified mean,

4. Normal theory confidence intervals are not of much use given the degree of skew and variability in the original data,

5. Bootstrap confidence intervals appear to be more useful but there is a suggestion of a mixture(s) in the data which may complicate the interpretation,

6. Haddock show a strong association with warm temperatures and the large catch in stratum 56 was associated with the highest temperature observed in the survey. If we had information on the extent of this warm water, we could evaluate how large the high abundance area was. Given the extent of this area, the large catch could be reweighted both at the stratified mean and bootstrap stage. Unfortunately, temperature observations are only available where trawl sets were made in the surveys of Div. 4VW.

One major implication of item 6 is that the area of very warm water may actually be less extensive than implied by the 1/6 weighting given to the large catch in the present survey. Therefore, any reweighting for the true extent of the warm water would probably result in giving less weight to the large catch which in turn would result in a lower stratified mean and could change our perception of the status of this stock in 1988.

Given that trawl surveys are quite expensive, more benefit should be obtained from the data collected. The methods presented in this chapter are offered as a took kit for estimation, exploration and evaluation of survey data. These methods all operate on the same basis as the survey itself – design-based inference where the survey design, not a statistical probability model, describes the probability basis for the sample units.

Acknowledgments

I thank Hans Lassen for inviting me to participate in the Workshop in St. Petersburg, Russia. This invitation not only provided me with an interesting venue and an interested audience for this material but also motivated me to organize my work on surveys along a common theme. Doug Swain, Stratis Gavaris and Bob O'Boyle provided very useful comments on earlier drafts of this and I thank them for their time and help. The students in the Workshop also provided very useful feedback on the text and the S-PLUS library. Finally, I must acknowledge Gordie Swartzman's FishIdeas computer package as being the prototype for my menu/dialogue approach for constructing a S-PLUS library for the analysis of survey data.

References

BECKER, R. A., J. M. CHAMBERS, and A. R. WILKS. 1988. The new S language. Wadsworth and Brooks/Cole, Pacific Grove, CA.

BUCKLAND, S. T., K. L. CATTANACH and A. A. ANGANUZZI. 1992. Estimating trends in abundance of dolphins associated with tuna in the eastern Tropical Pacific Ocean, using sightings data collected on commercial tuna vessels. *Fish. Bull.,* U.S., **90**: 1–12.

CHAMBERS, J. M., W. S. CLEVELAND, B. KLEINER, and P. A. TUKEY 1983. Graphical methods for data analysis. Wadsworth Statistics/Probability Series, Belmont, CA.

CHAMBERS, R. L. and DUNSTAN, R. 1986. Estimating distribution functions from survey data. *Biometrika*, **73**: 597–604.

COCHRAN, W. G. 1977. Sampling techniques. John Wiley & Sons, New York, N.Y.

D'AMOURS, D. 1993. The distribution of cod (*Gadus morhua*) in relation to temperature and oxygen level in the Gulf of St. Lawrence. *Fish. Ocean.*, **2**: 24–29.

DOUBLEDAY, W. G. (ed.) 1981. Manual on groundfish surveys in the Northwest Atlantic. *NAFO Sci. Coun. Studies,* **2**: 55 p.

ECKER, M. D., and J. F. HELTSHE. 1994. Geostatistical estimates of scallop abundance, *In*: Case studies in biometry. N. Lange, L. Ryan, L. Billard, D. Brillinger, L. Conquest and J. Greenhouse (eds), John Wiley & Sons, Inc., chapter 6, p. 125–144.

EFRON, B. 1981. Nonparametric standard errors and confidence intervals. *Can. J. Stat.*, **9**: 139–172.

EFRON, B.1982. The jackknife, the bootstrap and other resampling plans. Society for Industrial and Applied Mathematics, Philadelphia. 38: 92p.

EFRON, B. 1992. Jackknife-after-bootstrap standard errors and influence functions (with discussion). *J. Royal Stat. Soc. Ser. B*, **54**: 83–127.

EFRON, B. and R. J. TIBSHIRANI. 1993. An introduction to the bootstrap. Chapman & Hall, London.

GAVARIS, S., and S. J. SMITH. 1987. Effect of allocation and stratification strategies on precision of survey abundance estimates for Atlantic cod (*Gadus morhua*) on the Eastern Scotian Shelf. *J. Northw. Atl. Fish. Sci.,* **8**: 137–145.

HALLIDAY, R. G. and P. A. KOELLER, 1981. A history of Canadian groundfish trawling surveys and data usage in ICNAF Divisions 4TVWX. *In* Bottom Trawl Surveys. W. G. Doubleday and D. Rivard (eds.): *Can. Spec. Publ. Fish. Aquat. Sci.,* **8**: 27–41.

JONES, M. C. and I. S. BRADBURY. 1993. Kernel smoothing for finite populations. *Stat. Comput.*, **3**: 45–50.

KIMURA, D. K. and J. W. BALSIGER. 1985. Bootstrap methods for evaluating sablefish pot index surveys. *N. Am. J. Fish. Manage.,* **5**: 47–56.

KOVAR, J. G., RAO, J. N. K. and C. F. J. WU. 1988. Bootstrap and other methods to measure errors in survey estimates. *Can. J. Stat.*, **16** (suppl.): 25–45.

MOHN, R. K. and R. COOK. 1993. Workbook: Introduction to sequential population analysis, *NAFO Sci. Coun. Studies*, **17**: 110 p.

MOYER, L. M. and P. H. GEISSLER. 1991. Accommodating outliers in wildlife surveys. *Wild. Soc. Bull.*, **19**: 267–270.

PAGE, F. H., R. J. LOSIER, S. J. SMITH and K. HATT. 1994. Associations between cod, and temperature, salinity and depth within the Canadian groundfish bottom trawl surveys (1970–1993) conducted in NAFO Divisions 4VWX and 5Z. *Can. Tech. Rep. Fish. Aquat. Sci.,* **1958**: vii + 160 p.

PELLETIER, D., and P. GROS. 1991. Assessing the impact of sampling error on model-based management advice: comparison of equilibrium yield per recruit variance estimators. *Can. J. Fish. Aquat. Sci.*, **48**: 2129–2139.

PERRY, R. I., and S. J. SMITH. 1994. Identifying habitat associations of marine fishes using survey data: an application to the Northwest Atlantic. *Can. J. Fish. Aquat. Sci.*, **51**: 589–602.

PERRY, R. I., M. STOCKER, and J. FARGO. 1994. Environmental effects on the distribution of groundfish in Hecate Strait, British Columbia. *Can. J. Fish. Aquat. Sci.,* **51**: 1401–1409.

RAO, J. N. K. and C. F. J. WU. 1988. Resampling inference with complex survey data. *J. Amer. Stat. Assoc.,* **83**: 231–241.

ROBOTHAM, H. and J. CASTILLO. 1990. The bootstrap method: An alternative for estimating confidence intervals of resources surveyed by hydroacoustic techniques. *ICES Rapp. Proc.-Verb.,* **189**: 421–424.

SCOTT, J. S. 1982. Depth, temperature and salinity preferences of common fishes of the Scotian Shelf. *J. Northw. Atl. Fish. Sci.*, **3**: 29–40.

SIGLER, M. F. and J. T. FUJIOKA. 1988. Evaluation of variability in sablefish, *Anoplopoma fimbria*, abundance indices in the Gulf of Alaska using the bootstrap method. *Fish. Bull. U.S.,* **86**: 445–452.

SIMARD, Y., P. LEGENDRE, G. LAVOIE, and D. MARCOTTE. 1992. Mapping, estimating biomass and optimizing sampling programs for spatially autocorrelated data: case study of the northern shrimp (*Pandalus borealis*). *Can. J. Fish. Aquat. Sci.,* **48**: 32–45.

SINCLAIR, A. 1992. Fish distribution and partial recruitment: the case of eastern Scotian Shelf cod. *J. Northw. Atl. Fish. Sci.*, **13**: 15–24.

SITTER, R. R. 1992. A resampling procedure for complex survey data., *J. Amer. Stat. Assoc.*, **87**: 755–765.

SKINNER, C. J., D, HOLT, and T. M. F. SMITH (eds). 1989. Analysis of complex surveys. John Wiley & Sons., New York.

SMITH, S. J.1981. A comparison of estimators of location for skewed populations, with application to groundfish trawl surveys. *In* Bottom Trawl Surveys. W. G. Doubleday and D. Rivard (eds). *Can. Spec. Pub. Fish. Aquat. Sci.*, **58**: 154–163.

SMITH, S. J. MS 1988. Abundance indices from research survey data. *In*: Collected papers on Stock Assessment Methods. D. Rivard (ed.). *CAFSAC Res. Doc***.,** No. 61: 16–43.

SMITH, S. J. 1990. Use of statistical models for the estimation of abundance from groundfish trawl surveys. *Can. J. Fish. Aquat. Sci.*, **47**: 894–903.

SMITH, S. J. MS 1991. Assessing the efficiency of the groundfish survey design for the Scotian Shelf summer surveys – 1980–1990. *CAFSAC Res. Doc.*, No. 39: 15 p.

SMITH, S. J. 1997. Bootstrap confidence limits for groundfish trawl survey estimates of mean abundance. *Can. J. Fish. Aquat. Sci.,* **54**: (in press).

SMITH, S. J. and S. GAVARIS. 1993a. Improving the precision of fish abundance estimates of Eastern Scotian Shelf cod from bottom trawl surveys. *N. Am. J. Fish. Manage.*, **13**: 35–47.

SMITH, S. J. and S. GAVARIS. 1993b. Evaluating the accuracy of projected catch estimates from sequential population analysis and trawl survey abundance estimates. *In*: Risk evaluation and biological reference points for fisheries management. S. J. Smith, J. J. Hunt and D. Rivard (eds.). *Can. Spec. Pub. Fish. Aquat. Sci.*, **120**: 163–172.

SMITH, S. J., and F. H. PAGE. MS 1994. Implications of temperature and haddock associations on survey abundance trends. *DFO Atl. Fish. Res. Doc.*, No. 21: 34 p.

SMITH, S. J., and F. H. PAGE. 1996. Associations between Atlantic cod (*Gadus morhua*) and hydrographic variables: Implications for the management of the 4VsW cod stock. *ICES J. Mar. Sci.*, **53**: 597–614.

SMITH, S. J., and G. ROBERT. 1997. Getting more out of your survey designs: an application to Georges Bank scallops (*Placopecten magellanicus*). *In* North Pacific symposium on invertebrate stock assessment and management, G. Jamieson and A. Campbell (eds). *Can. Spec. Pub. Fish. Aquat. Sci.*, (in press).

SMITH, S. J., R. L. LOSIER, F. H. PAGE, and K. HATT. 1994. Associations between haddock, and temperature, salinity and depth within the Canadian groundfish bottom trawl surveys (1970–1993) conducted in NAFO Divisions 4VWX and 5Z. *Can. Tech. Rep. Fish. Aquat. Sci.*, **1959**: vi + 70 p.

SMITH, S. J., R. I. PERRY, and L. P. FANNING. 1991. Relationships between water mass characteristics and estimates of fish population abundance from trawl surveys. *Environ. Monitor. Assess.*, **17**: 227–245.

STANLEY, R. D. 1992. Bootstrap calculation of catch-per-unit-effort variance from trawl logbooks: Do fisheries generate enough observations for stock assessments? *N. Am. J. Fish. Manage.*, **12**: 19–27.

STATISTICAL SCIENCES, INC. 1995. S-PLUS user's manual, version 3.3 for windows. Statistical Science, Inc., Seattle, Washington.

SWAIN, D. P., and D. L. KRAMER. 1995. Annual variation in temperature selection by Atlantic cod *Gadus morhua* in the southern Gulf of St. Lawrence, Canada and its relation to population size. *Mar. Ecol. Prog. Ser.,* **116**: 11–23.

THOMPSON, S. K. 1992. Sampling. John Wiley & Sons, New York, N.Y.

ZWANENBURG, K., G. A. P. BLACK, and G. YOUNG. 1995. Haddock in Division 4TVW in 1994. *DFO Atl. Fish. Res. Doc.*, No. 112, 69 p.

Appendix A: S-PLUS Library

All of the analysis methods in this presentation are available from the author as an S-PLUS library. S is an object-oriented command line language developed by researchers at AT&T (Becker *et al*., 1988). S-PLUS is a superset of this language with enhanced data manipulation, analysis and graphics functions. The purpose of this course on survey design and analysis was not to teach S-PLUS; the methods presented were either written in and/or developed using S-PLUS; because of the power and versatility of the language. Instead the users can access the S-PLUS functions through a series of dialogue screens developed by the author[2] and supported under version 3.3 for Microsoft Windows. The dialogue screens are themselves accessible through a series of custom-made pull-down menus.

A readme file is provided with the library describing how to install the library and create a course specific icon. To start up the "hands-on" session on survey design and analysis, double-click the cursor on the nafocourse icon in the S-PLUS group. Once the screen has loaded, control is returned to the user when the cursor arrow and the prompt (usually a > sign) appears. This session contains the haddock data set used in this presentation plus some other data sets for use by the students. The analyses and estimation methods are contained in a library (nafolib) which is loaded automatically. In addition, the **Survey Tools** menu is added to the top of the S-PLUS screen between the standard **Options** and **Windows** options. Put the cursor on the **Survey Tools** item and single-click on it. The pull-down menu will list four major survey analyses items: **Describe**, **Estimate**, **Bootstrap** and **Association** (Fig. A1). Note that when the first letter of a menu item is underlined, this indicates that the item may be directly accessed using that letter on the keyboard instead of the cursor.

The survey data are stored as *list* objects with class *strata data*. S-PLUS *list* objects refer to data objects which can contain other types of data objects such as character data, numerical data, matrices, vectors and even other *list* objects. Data objects with a class designation usually have a method associated with them – this is the object-oriented aspect of the S-PLUS language. In our case, data from a stratified random survey design contain structure corresponding to what stratum the observation (or trawl set) was obtained from. All of the methods contained herein use this structure and therefore the data objects have to contain and make available this structure. That is, all methods associated with the stratified random design will be appropriate for data objects with the *strata data* class.

The **Describe** menu has only one function associated with it called **Names** (Fig. A2). Single-click on **Names** to access the associated dialogue screen. The resulting display contains a list of *strata data* class objects available in the session workspace (Fig. A3). Four data sets are available for analysis. The names of the first three identify the species (cod and haddock), year (83, 88 and 89) with "j" indicating July surveys, management area (4VsW, 4VW and gb indicating Georges Bank). The fourth data set contains data on four species (cod, haddock, yellowtail and silver hake) from the July 1982 survey in NAFO Div. 4VsW.

Using the cursor, single-click on haddock88j.4vw and then single-click on **OK** and the following will appear on the command screen.

```
>
 [1]  "vessel"   "cruise.no"    "set.no"    "strata"    "day"
 [6]  "month"    "year"         "tow.dist"  "species"   "haddock"
[11]  "depth"    "temperature"  "salinity"
>
```

This is a list of the items within the *strata data* object haddock88j.4vw. This data object contains many different data items but almost all of the necessary information for stratified random surveys is contained in strata, haddock and tow.dist. The item strata contains the stratum membership for each observation and in this case haddock contains the numbers of haddock caught for each trawl set. The item tow.dist refers to the actual distance towed for each trawl set. Recall that for these surveys a standard distance of 1.75 nm was assumed, although the actual length of the tow may vary. The current practice is to standardize the numbers of fish caught for the actual distance towed to that which would have been caught over 1.75 nm.

[2] These dialogue screens are still under development and all comments and suggestions are welcome.

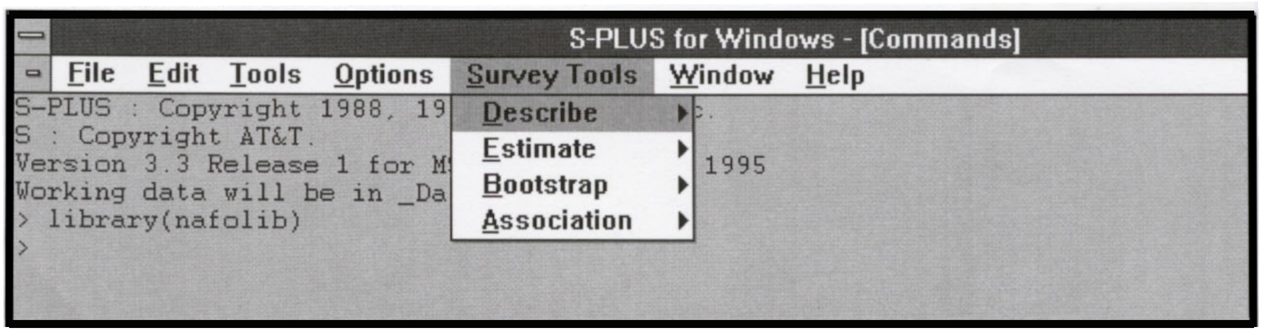

Fig. A1: Survey Tools pull-down menu with the four major survey analyses items.

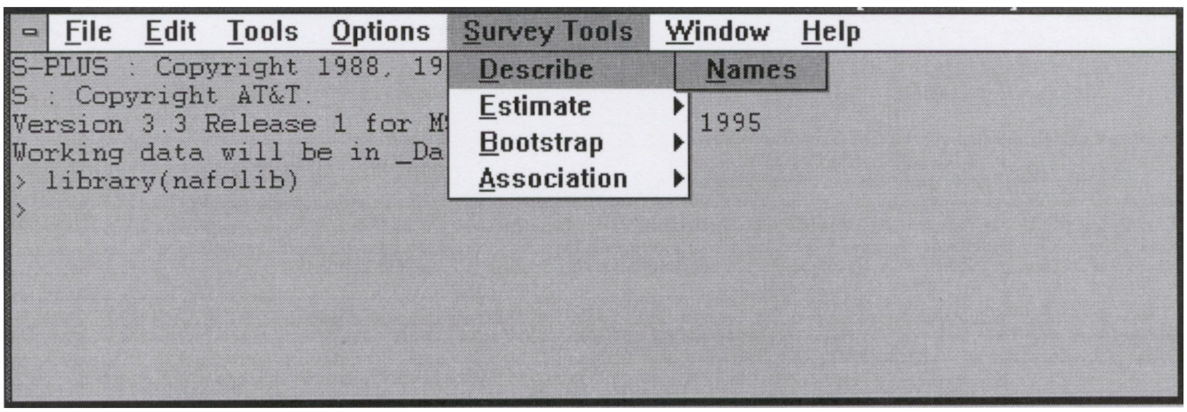

Fig. A2: **Describe** pull down menu with **Names** item shown.

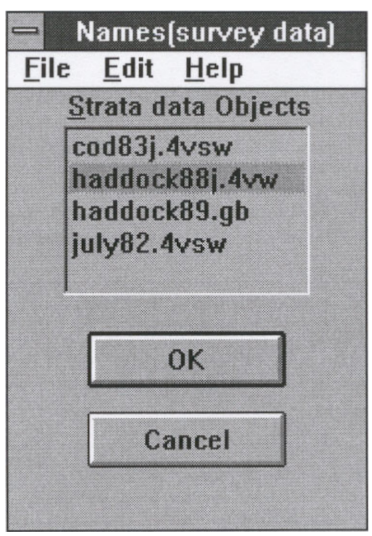

Fig. A3: Dialogue screen for **Names** item of the **Describe** menu.

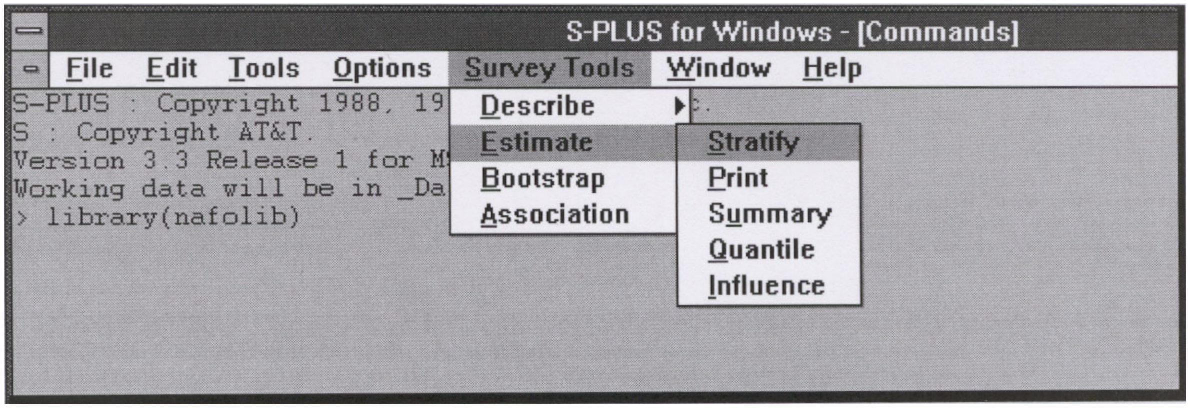

Fig. A4: **Estimate** pull down menu with the five major items contained therein.

The **Estimate** pull-down menu has five items associated with it (Fig. A4). The first three (**Stratify**, **Print** and **Summary**) are used to conduct a standard stratified analysis of the survey data. The latter two items are used for exploratory analyses of the survey data.

The purpose of the **Stratify** function (Fig. A5) is to extract necessary quantities from haddock88j.4vw and strata.4vw to produce a *strata* class object. Given that a *strata data* object can contain data on many species, the **Stratify** function prepares only the data specified for the analysis for one species.

The **Stratify** dialogue lists *strata data* objects and the *strata area* objects. Choose one object from each list by single-clicking on the item in the list. In the example haddock88j.4vw and strata.4vw were chosen. The latter object contains information on the N_h for the strata in NAFO area 4VW. The radiobutton Fish Catch has also been selected. When Fish Catch has been selected, tow distance corrections (if tow.dist is available) are applied to each catch. Such a correction is not appropriate for hydrographic quantities and this option should be turned off when analysing variables other than fish catches.

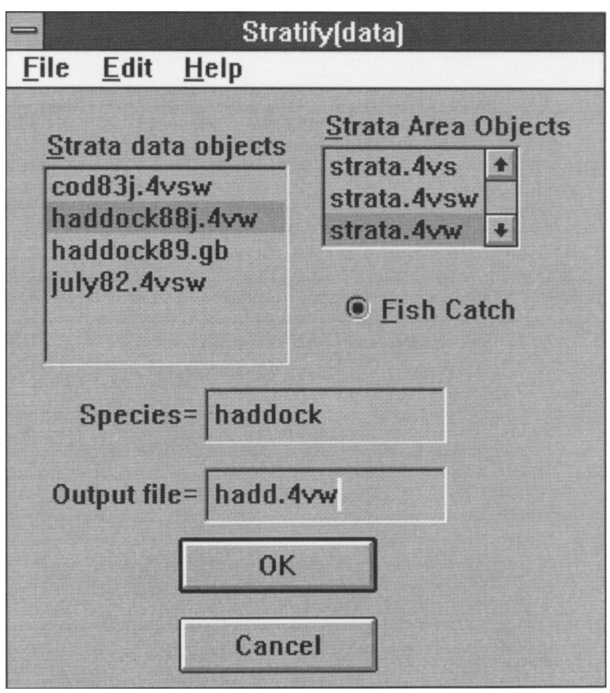

Fig. A5: Dialogue screen for **Stratify** function of the **Estimate** pull-down menu.

Enter the variable to be analysed in the **Species=** and the output file name in the **Output file=** text fields. The latter field allows for a *strata* object (in this case hadd.4vw) to be passed on to the other analysis functions. Single-click on **OK** and the necessary quantities are stored in hadd.4vw.

Choose the **Print** item from the **Estimate** menu and you will see two *strata* objects listed (Fig. A6). The first, codage4.4vsw, was created from the cod83j.4vsw object for users to experiment with, while the second hadd.4vw is the object created above with the **Stratify** function. Single click on hadd.4vw and then on **OK** and the following will appear on the command screen:

>

	Strata	nh	Wh	Mean	Std. Dev.
[1,]	40	6	0.029356	0.3431	0.8405
[2,]	41	4	0.031770	7.3953	5.4986
[3,]	42	7	0.045654	0.7745	0.8587
[4,]	43	4	0.041873	0.2574	0.5147
[5,]	44	4	0.124698	1.2868	2.5735
[6,]	45	4	0.032501	0.0000	0.0000
[7,]	46	3	0.015599	0.0000	0.0000
[8,]	47	6	0.051341	24.6678	45.0335
[9,]	48	5	0.046035	0.0000	0.0000
[10,]	49	2	0.004575	17.7288	11.3230
[11,]	50	3	0.012168	75.7218	85.3797
[12,]	51	2	0.004670	2.0588	2.9116
[13,]	52	2	0.010961	55.7491	77.3853
[14,]	53	2	0.008228	0.0000	0.0000
[15,]	54	2	0.015853	45.8088	44.4021
[16,]	55	7	0.067416	94.9112	59.6268
[17,]	56	6	0.030340	985.8548	2212.7907
[18,]	57	2	0.025765	18.0469	25.5221
[19,]	58	3	0.020907	77.8731	118.9076
[20,]	59	6	0.100012	24.8962	48.8229
[21,]	60	3	0.042699	18.6152	17.2183
[22,]	61	2	0.036663	0.0000	0.0000
[23,]	62	4	0.067226	2.7183	4.7750
[24,]	63	2	0.009595	89.6553	26.3410
[25,]	64	5	0.041206	109.2778	110.5316
[26,]	65	8	0.075708	63.9120	65.6185
[27,]	66	2	0.007180	1.5441	2.1837

>

Fig. A6: Dialogue screen for **Print** *strata* object function of the **Estimate** pull-down menu.

The column designated **Strata** gives the strata labels (h = 1, ..., 27) (ranging from 40 to 66) for the 4VW area of the eastern Scotian Shelf. The next column, **nh**, lists the number of trawl sets (n_h) for each of the strata in this survey. The proportion of the area in each stratum (W_h) is given in the column headed by **Wh**. Finally, the mean (**Mean**) and standard deviation (**Std. Dev.**) for each stratum are listed in the last two columns. This particular presentation of the data is for information purposes only and not indicative of how the data appear in **hadd.4vw.** The **Print** function performs some processing on the *strata* object.

The stratified mean, variance and many other quantities are obtained using the **Summary** item (Fig. A7). This dialogue lists *strata* objects in addition to a slider labeled **alpha=0.05** and a radiobutton labelled **Efficiency Estimates**. An output name for the summary can also be specified.

The slider allows the user to choose their own level $(1 - \alpha)$ for parametric confidence intervals for the stratified mean. Choosing efficiency estimates will result in the gain in efficiency due to allocation and strata components being calculated.

Within the **Summary** dialogue choose **hadd.4vw** and efficiency estimates. Selecting **OK** produces the following output on the command screen:

```
>
$yst:
[1] 56.151

$se.yst:
[1] 27.733

$Yst:
[1] 149772466

$df.yst:
[1] 5.2409

$alpha:
[1] 0.05

$ci.yst:
[1] -14.164   126.466

$effic.alloc:
[1] 44.701

$effic.str:
[1] 2.4552

$descrip:
[1] "Stratified Analysis"
>
```

This output simply produces a listing of the **Summary** of object **hadd.4vw**. The individual elements are:

$yst:	estimate of the stratified mean.
$se.yst:	estimate of standard error of mean (square root of equation 2).
$Yst:	estimate of stratified total, $N\bar{y}_{st}$.
$df.yst:	effective degrees of freedom as per equation 3.
$alpha:	α level for confidence interval.
$ci.yst:	lower and upper limits for confidence interval.

$effic.alloc: Allocation component of estimated efficiency as a percentage of simple random sampling variance.

$effic.str: Strata component of estimated efficiency as a percentage of simple random sampling variance.

$descrip: Simple label for summary object.

If an output file name had been given, the individual components (e.g. stratified mean), could be identified as **objectname$yst**.

The **Quantile** dialogue (Fig. A8) on the **Estimate** menu will estimate and plot the cdf for *strata* objects. The fields for this dialogue offer the user the ability to enter a specific x-axis label, the option of immediately plotting the cdf or saving the x and y coordinates to a file. Specific quantiles such as the upper and lower quartiles, can be read right off the graph or from the file.

The **Plot** option will open a new graphics screen where a plot such as that presented in Fig. 2 will appear.

The **Influence** dialogue (Fig. A9) will plot $(W_h y_{hi} / n_h) / \bar{y}_{st}$ against y_{hi} or save these values in an output file. The y_{hi} are cube-root transformed when plotted to scale the plot (Fig. 3). Once the graph appears on the graphics screen, the cursor will be presented as a crosshair. Position the cursor on a point that you wish to identify and click the left mouse button. Up to five points can be identified this way. Clicking the right hand button will cancel the identification function and return control to the user.

The **Bootstrap** menu has three functions under it, **Boot Strata Object**, **Summary** and **QQnorm** (Fig. A10). The dialogue screen for **Boot Strata Object** is presented in Fig. A10 . This function works directly on *strata* objects which are identified in the **Strata Objects** box. The list labelled **Resample Method** allows the user to choose one of three published resampling methods to conduct the bootstrap analysis with. The slider on the right hand side of the dialogue in Fig. A11 refers to the number of bootstrap resamples that will be made. The left hand slider labelled **n-m=0** is for setting the m_h for the **Rescale** resampling method, that is setting **n-m=1** when $n_h - m_h = 1$.

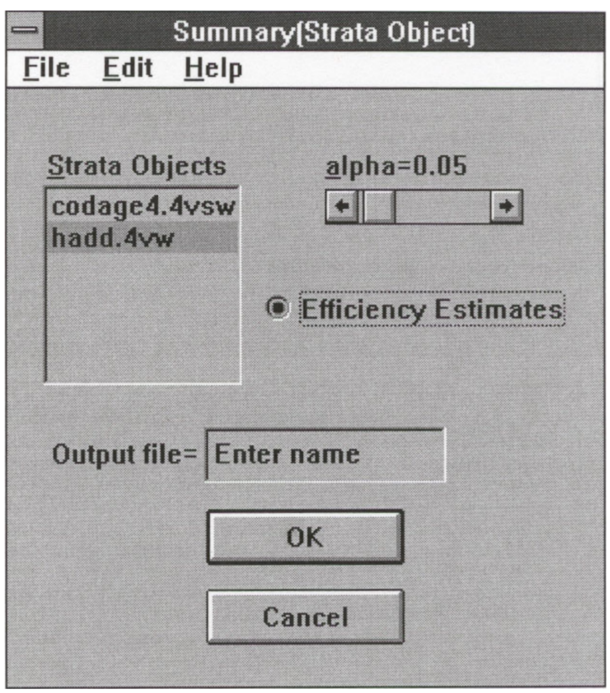

Fig. A7: Dialogue screen for **Summary** strata function of the **Estimate** pull-down menu.

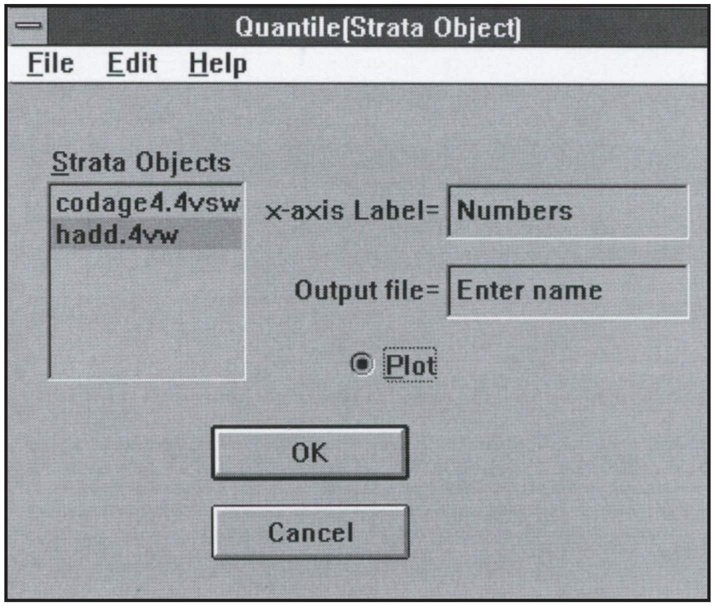

Fig. A8: Dialogue screen for **Quantile** function of the **Estimate** pull-down menu.

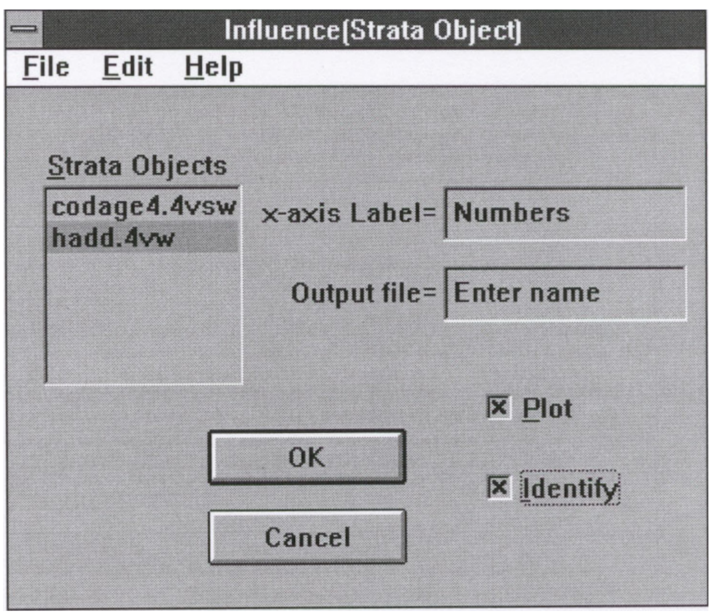

Fig. A9: Dialogue screen for **Influence** function of the **Estimate** pull-down menu.

The results of the bootstrap replications can be stored in a file which will have class *boot* for further analysis. The results from such a *boot* object can be summarized using the **Summary** dialogue shown in Fig. A12. *Boot* objects are identified in the list on the left hand side. There are two such objects available for immediate use, although users are encouraged to generate their own[3]. The first object, haddock88j.bwr

3 WARNING! Although, the bootstrap algorithm has been written as efficiently as possible it is still written in S-PLUS code and may take a few minutes. An implementation using C code with a dynamic load will be considered for a future version.

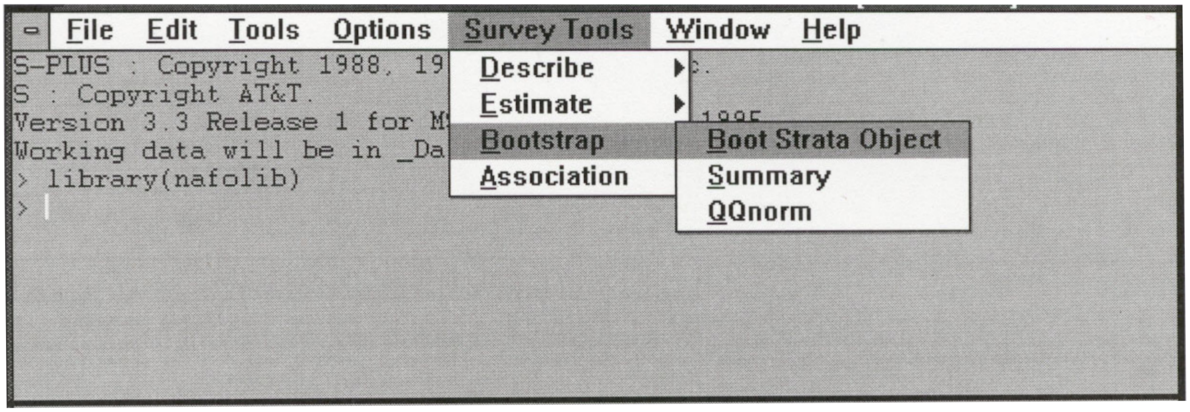

Fig. A10: **Bootstrap** pull down menu with the three major items contained therein.

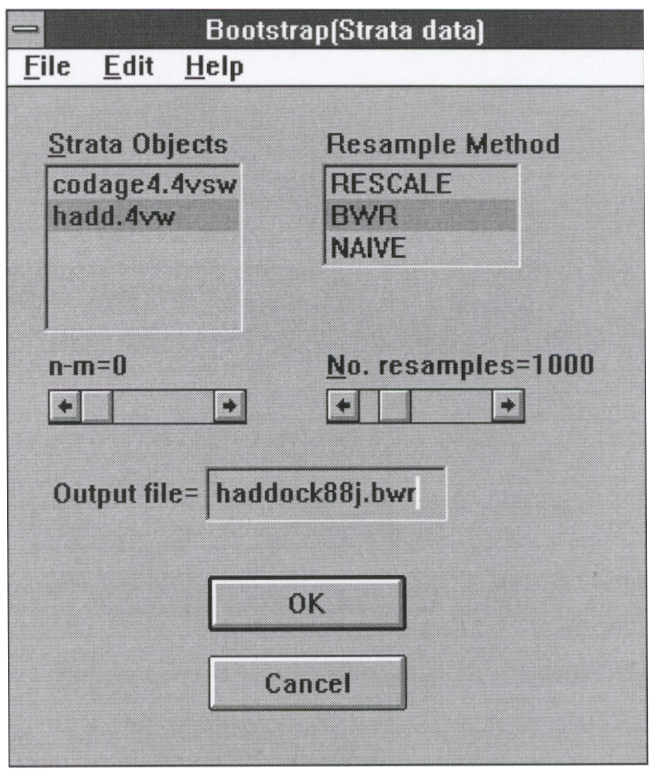

Fig. A11: Dialogue screen for **Boot Strata Object** function of the **Bootstrap** pull-down menu.

was generated with the settings in Fig. A11, while the second was generated using the haddock89.gb *strata data* object, the **Rescale** method and **n-m = 1**. Both objects contain 1 000 bootstrap resamples.

The slider marked alpha=0.05 allows the user to choose their own level $(1-\alpha)$ for bootstrap confidence intervals for the stratified mean.

The list labelled **CI Method** lists three of the more common bootstrap confidence interval methods, the percentile method (PC), the bias-corrected method (BC), and the bias-corrected and accelerated method (BCa). Choosing the haddock88j.bwr object with alpha=0.05 and CI Method = BCa results in the following output.

```
>
Original Mean= 56.15
Original Variance 769.1
Number of bootstraps = 1 000
Bootstrap Mean= 56.41
Variance of Bootstrap Mean= 769.7
BCa CI's for alpha= 0.05 are 22.83 128.28
Length = 105.4
Shape= 1.452
```

Min.	1st Qu.	Median	Mean 3rd	Qu.	Max.
15.73	29.98	58.21	56.41	67	167.8

>

The first two lines of the summary give the original stratified mean and variance of the stratified mean. Next the number of bootstrap resamples is given. The mean of the bootstrap estimates of the stratified mean, 56.41 is very close to the original estimate of 56.15. The next line gives the bootstrap estimate of the variance of the stratified mean which is also very close to the original estimate.

The Length entry refers to the length of the confidence interval. The bootstrap confidence interval is shorter than that from the Student-t (140.63 = 126.466 – (–14.164)). Shape is calculated as the natural log of the ratio of the upper limit minus the median to the median minus the lower limit (Efron, 1992). Therefore, a confidence interval which is symmetric around the median will have a shape measure of zero while a shape greater than zero indicates distributions skewed to the right. The bootstrap confidence is highly skewed, no doubt to accommodate the large catch of 5 496 haddock in stratum 56.

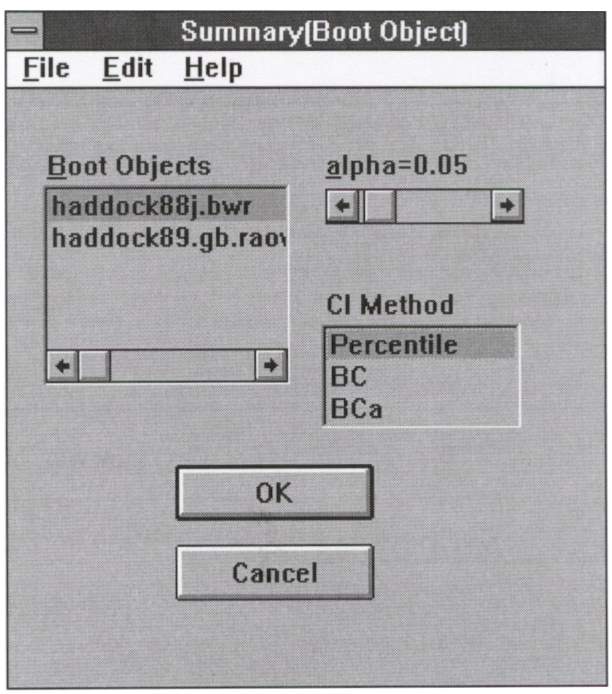

Fig. A12: Dialogue screen for **Summary** function of the **Bootstrap** pull-down menu.

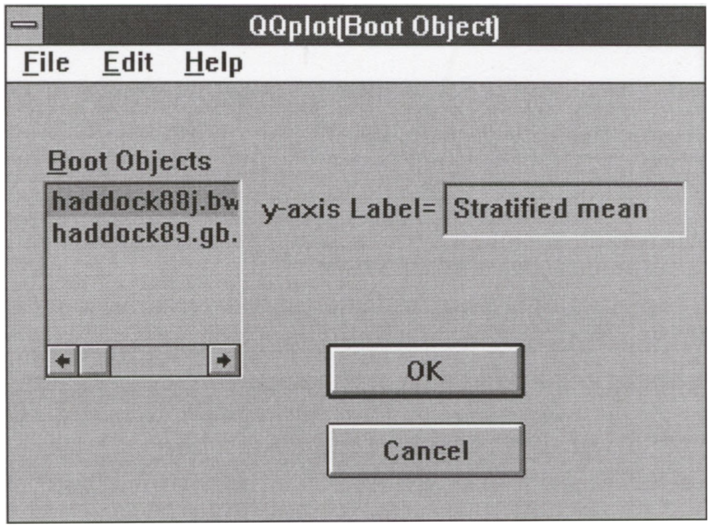

Fig. A13: Dialogue screen for **QQnorm** function of the **Bootstrap** pull-down menu.

Finally, key percentiles of the distribution of the bootstrap estimates are given. The distribution of these bootstrap estimates can be plotted as quantile-quantile plot using the **QQnorm** dialogue in Fig. A13. Choosing the haddock88j.bwr *boot* object (along with option to enter y -axis label) results in the plot in Fig. 4.

The environmental association methods are accessed through the **Association** pull-down menu shown in Fig. A14. The empirical cumulative distribution function G(t) and the catch-weighted function K(t) are obtained and plotted using the **Association Plot** dialogue (Fig. A15). This dialogue is similar to that for **Stratify** (Fig. A5) in that *strata data* and *strata area* objects have to be chosen. In addition, the fish species and environmental variable must be specified. Once the plot appears on the graphics screen the cursor is presented as a crosshair. Choose a location for the legend to appear and click on the cursor using the left button of the mouse. Clicking the right button cancels the legend placing action, returning control to the user. The resulting association plot is given in Fig. 5, with the **Habitat** line referring to the cumulative distribution function for temperature.

The test statistic and randomization test[4] can be produced using the dialogue **Association Test** shown in Fig. A16. In addition to the **KS-Test** (Kolmogorov-Smirnov), this dialogue also offers the option of choosing a form of the Cramer-von Mises test (**CVM-Test**) which is defined as the sum of the absolute vertical differences between G(t) and K(t). A slider is provided for choosing the number of replications for the randomization test.

If the results of the **Association Test** dialogue are stored in an output file, they can be summarized by using the **Summary** dialogue shown in Fig. A17. The results of this summary for the haddock data are given below including key percentiles of the distribution of the randomized statistic under the null hypothesis of no association.

```
>
Kolmorogorov-Smirnov Type test
Test Statistic = 0.4899
```

[4] WARNING! Although, the randomization algorithm has been written as efficiently as possible it is still written in S-PLUS code and may take a few minutes. An implementation using C code with a dynamic load will be considered for a future version.

Randomization Test
P-level for randomization test = 0.05499 Summary of distribution of 4000 test statistics from randomization simulation.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.228	0.309	0.372	0.373	0.439	0.59

>

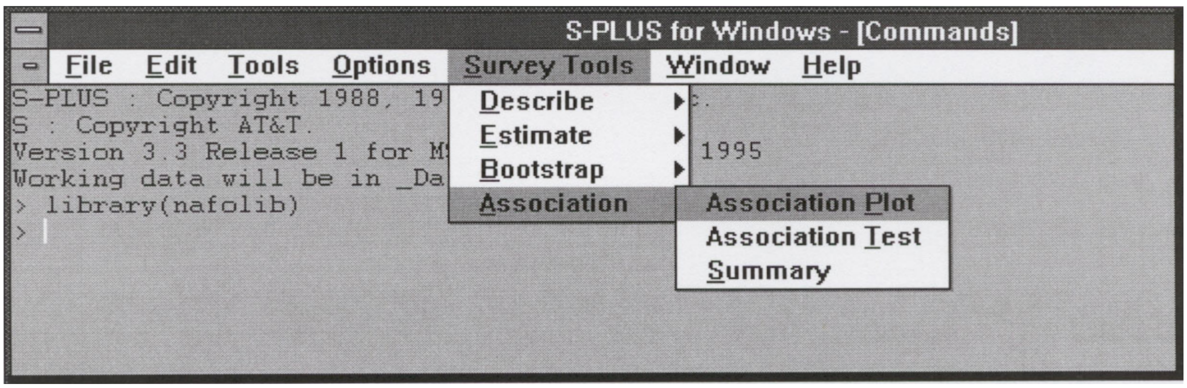

Fig. A14: **Association** pull down menu with the three major items contained therein.

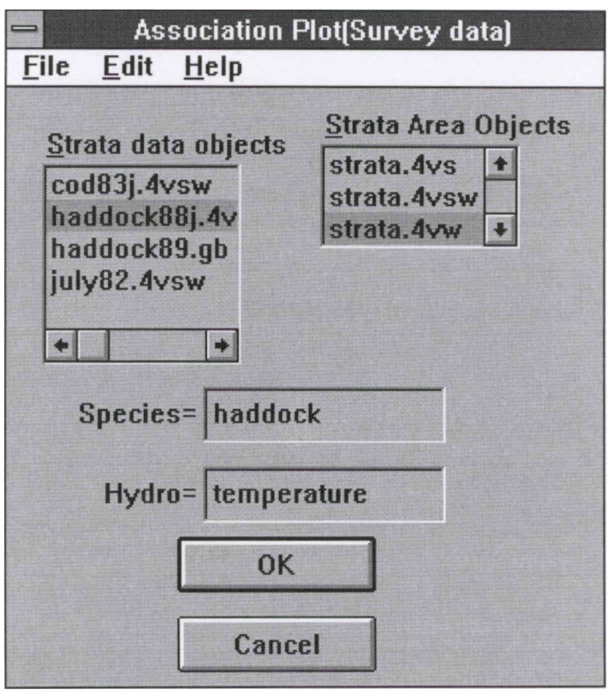

Fig. A15: Dialogue screen for **Association Plot** function of the **Association** pull-down menu.

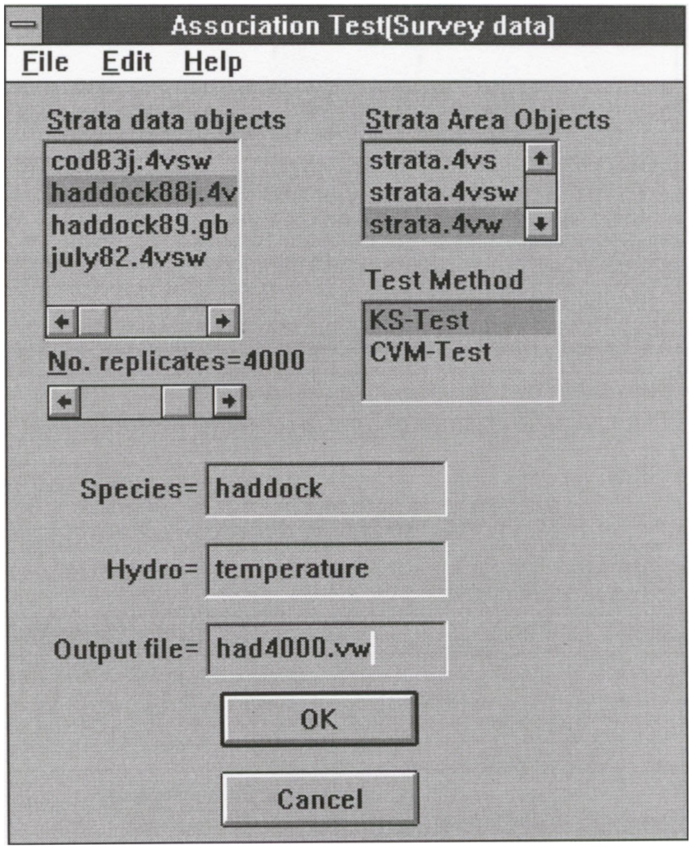

Fig. A16: Dialogue screen for **Association Test** function of the **Association** pull-down menu.

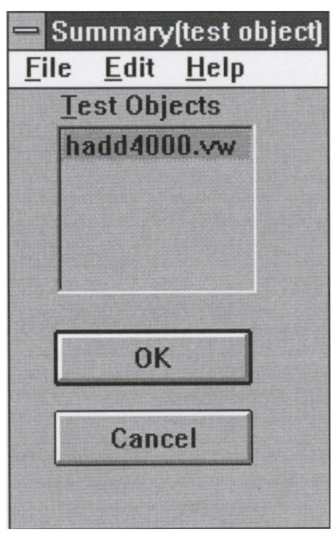

Fig. A17: Dialogue screen for **Summary** function of the **Association** pull-down menu.

Appendix B: S-PLUS Line Commands

The menus and screen dialogues presented above were developed for users unfamiliar with the S-PLUS command language. The command for each dialogue function is given below for those users who wish to use S-PLUS commands directly. These commands are shown with the necessary settings to give the examples presented above. More detailed help on each of these functions is available as windows help files by double clicking on the **NAFO Library help** entry in the S-PLUS **Help** menu item.

The **Names** dialogue simply implements the S-PLUS native function names():

```
>names(haddock88j.4vw)
```

The **Stratify** dialogue implements the custom-made function stratify(). To reproduce the haddock example enter:

```
>hadd.4vw<-stratify(haddock88j.4vw,strata.4vw,species=haddock)
```

To obtain results for a hydrographic variable such as temperature, declare species=temperature and set hydro=T.

```
>temperature.4vw<-stratify(haddock88j.4vw,strata.4vw,species=temperature, hydro=T)
```

The object hadd.4vw has been given class *strata* and therefore the native functions print() and summary() can be used, i.e., >print(hadd.4vw) and summary(hadd.4vw,effic=T). The second argument in the summary command refers to the efficiency calculations.

The S-PLUS commands to produce the **Quantile and Influence**[5] plots are, respectively:

```
>quantile.strata(hadd.4vw,Plot=T)
>
>influence.strata(hadd.4vw,Plot=T)
```

The following commands are for the **Bootstrap** menu item. The first command provides 1 000 resample (nresamp=1000) estimates of the stratified mean (stored in the object haddock88j.bwr) using the Bootstrap-with-replacement method (BWR).

```
>haddock88j.bwr<-boot.strata(hadd.4vw,nresamp=1000,method = "BWR")
```

The summary (**Summary**) and quantile (**QQnorm**) commands are straightforward:

```
>summary(haddock88j.bwr, CI.method = "BCa", alpha.b = 0.05)

>quantile.boot(haddock88j.bwr, ylab = "Haddock Numbers")
```

The plot generated by the **Association Plot** dialogue requires two applications of the quantile.prefer() command. The first (with implicit plot=T) presents the empirical cumulative distribution plot for the hydrographic variable (e.g., temperature).

```
>quantile.prefer(haddock88j.4vw,hydro=temperature, strata.group=strata.4vw+,ylab="temperature")
```

The line for catch-weighted function is added to the plot as:

```
>lines(quantile.prefer(haddock88j.4vw, hydro=temperature,
+ strata.group=strata.4vw,species=haddock,plot=F),lty=2)
```

The association test and its summary are produced by the following command.

```
> had4vw.test<-prefer.test(haddock88j.4vw, hydro=temperature,
+ strata.group=strata.4vw,species=haddock,nreps = 500, method = "KS-Test")

> summary (had4vw.test)
```

[5] The points can be individually labelled using identify( ).