Cluster Sampling: A Pervasive, Yet Little Recognized Survey Design in Fisheries Research

Article *in* Transactions of the American Fisheries Society · July 2014

DOI: 10.1080/00028487.2014.901252

CITATIONS
19

READS
191

1 author:

Gary A Nelson
Commonwealth of Massachusetts
37 PUBLICATIONS **368** CITATIONS

SEE PROFILE

In press. Transactions of the American Fisheries Society

Cluster Sampling: A Pervasive, Yet Little Recognized Survey Design in Fisheries Research

Gary A. Nelson

Massachusetts Division of Marine Fisheries

Annisquam River Marine Fisheries Field Station

30 Emerson Avenue

Gloucester, MA 01930

Email: gary.nelson@state.ma.us

ABSTRACT

Cluster sampling is a common survey design used pervasively in fisheries research to sample fish populations, but it is not widely recognized by researchers. Because fish collected via cluster sampling are not independent of each other, standard simple random sampling estimators and statistical tests that assume independence cannot be used to make inferences about fish populations. If the clustered nature of fisheries data is ignored, the main consequence is that the Type I error rate of common statistical tests will be severely inflated and significant differences will often be found in group comparisons where none exist. The goal of this paper is to provide an introduction to the estimation of population attributes and analysis of fisheries data collected via cluster sampling. The article addresses the nature of clustered fisheries data, reviews the random cluster sampling estimators of population attributes, explores the implications of violating the assumption of independence in hypothesis testing, and reviews current statistical approaches that can be used to analyze appropriately clustered data.

What is right is not always popular and what is popular is not always right – Albert Einstein (1879-1955)

Introduction

The estimation of biological parameters of fish populations such as average size, maturity-at-age, etc. is a major task of many state, federal and academic fisheries researchers. Such information is often used in the stock assessment process where model estimates of management values (e.g., fishing mortality) are used by regulatory boards to control the harvesting of fish population or in ecological studies devised to test some statistical hypothesis about populations (e.g., growth differences). If population attributes are estimated incorrectly, the results, and conclusions on which they are based, will be misleading and may impact unnecessarily the livelihoods of fishers if used in management.

Survey sampling is used in fisheries research to estimate fish population attributes after observing a sample of individuals from the population. The sampling design (the procedure by which the individuals are selected) is usually chosen to produce the most accurate and precise (low-variance) estimates given the monetary resources available to the researcher (Cochran 1977). Sampling designs and associated estimators used commonly in fisheries research are probability-based which means assumptions regarding the selection of individuals must be met to produce unbiased estimates of population attributes (Thompson 2002).

When estimating biological attributes of fish populations, researchers often assume that fish are collected by using *simple random sampling* (SRS), a well-known sampling design with standard estimators for mean, proportion, and ratios and corresponding measures of variance found in most textbooks on elementary statistics and survey sampling (e.g., Zar 1999; Lohr 1999) regardless of the original design used. The assumptions of SRS require that each

individual selected for the sample has the same nonzero probability of occurring in the sample, and that the selection of one individual is not influenced by other individuals already selected (Lohr 1999). Independence among individuals is the backbone of estimation in SRS as well as common statistical methods used in hypothesis testing (Sokal and Rohlf 1995).

To meet the assumptions of SRS in fisheries research, fish would have to be captured individually and at random (Figure 1A) to appropriately apply the SRS estimators and use common statistical methods. This type of selection is nearly impossible in fisheries research because fish populations are distributed over large areas and the types of gear generally used (e.g. seines, trawls, electroshockers[1], etc.) capture fish in groups or *clusters* (Figure 1B), not individually. In this case, the appropriate design for estimating population attributes is *random cluster sampling* (RCS) since the inclusion of an individual in a sample is based on the probability of selecting a cluster, not an individual (Pennington and Volstad 1994). Therefore, RCS estimators must be used to estimate population attributes and statistical tests must be adjusted to account for lack of independence among individuals to obtain unbiased results (Galbraith et al. 2010). Unfortunately, it appears that RCS is not often recognized by fisheries researchers. In a review of all articles published in *Fishery Bulletin* from 2008-2012, only authors of 1 of 54 papers that estimated population attributes or conducted hypothesis testing using samples collected via cluster sampling identified clustering and analyzed the data appropriately; the authors of the remaining papers assumed SRS.

The most likely reason for improper analysis of RCS data is lack of awareness. Cluster sampling for estimating population attributes is usually not taught in fisheries courses except in the context of estimating abundance (e.g., American Fisheries Society's *Fisheries Techniques* book editions) and has been described rarely in publications on designs of trawl surveys (e.g.,

Fogarty, 1985). Because they are unaware, researchers improperly associate the sampling design used to measure fish abundance (e.g., simple or stratified random designs) as the design used to collect individuals. The clear implications of this mistake are that the estimates of population attributes, results of hypothesis testing, and conclusions drawn from statistical analyses in many published papers may be incorrect because clustering was not taken into account in the estimation process or statistical analyses.

As an aid to improving data analysis in fisheries research, this paper provides an introduction to the estimation of population attributes and analysis of fisheries data collected via cluster sampling. I address the nature of clustered fisheries data, review the RCS estimators of population attributes, explore the implications of violating the assumption of independence in hypothesis testing, and review current statistical approaches that can be used to analyze appropriately clustered data.

OVERVIEW OF CLUSTER SAMPLING

In fisheries research, gear surveys are routinely used to measure relative abundance of fish populations by using sampling designs such as stratified sampling that incorporate SRS in site selection (ASMFC 1994). A standardized haul or tow with known swept-area is used to sample fish and individual sites of size equal to the known swept-area are selected randomly. The number of fish caught in each haul is used to develop indices of relative abundance usually by calculating the mean number per standardized tow or swept-area (ASMFC 1994). In this case, use of SRS-based estimators is appropriate because the primary sampling unit is the standardized swept-area site and the number of fish caught is an attribute of the site.

When the research interests turn towards making inferences about fish populations caught in the gear survey, the underlying sampling design becomes RCS-based. A cluster (the group of fish in a tow or haul) is the primary sampling unit which has been selected randomly in association with the gear survey. In RCS, inclusion of an individual in a sample is based on the probability of selecting a cluster; thus, the selection of one individual is dependent on the selection of another individual (Lohr 1999; Thompson 2002). Because of underlying factors that affect membership, individuals of a cluster tend to be more similar than members of other clusters (Pennington and Volstad 1994). This is especially true for gregarious fishes as individuals of similar size often aggregate together (Pitcher and Parrish 1993) and their spatial distribution is often related to size (Milikin 1993; Osenberg et al., 1994). Because of this *within-cluster* similarity, the information content of clustered data is not the same as data collected via SRS sampling because the same information is partially collected from individuals in a cluster and not from other members of the population (Lohr 1999). It is the level of similarity among members of clusters that creates the degree of non-independence.

Estimation of Population Attributes, Variance and Confidence Intervals

Estimators for *simple random cluster sampling* (SRCS) should be used in the estimation of fish attributes. In SRCS, the clusters are chosen randomly via the abundance survey and attributes of individual cluster members are measured. If a stratified random design is used, SRCS is still assumed because the number of sites allocated to a stratum is usually proportional to stratum area and it is essential to choose a unit size that is sufficient for the area of interest rather than for specific subareas (Pennington and Volstad 1994). If the allocation of sites is not proportional to stratum area, stratified RCS estimators are available (Cochran 1977; Fogarty 1985).

The survey statistics of cluster sampling that are of interest to fisheries researchers are the mean attribute (e.g., mean length, mean stomach weight, etc.) and proportion (e.g., sex composition) and their associated variances. Because the fish attribute and cluster size are random variables in SRCS, the mean (r) and proportion (p) are estimated by using ratio estimators. The general estimators for r and p are:

$$(1)\quad \hat{r} = \frac{\sum_{i=1}^{n} M_i \hat{\mu}_i}{\sum_{i=1}^{n} M_i} \qquad (2)\quad \hat{p} = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} M_i}$$

where M_i is the total number of fish (cluster size) in cluster i, $\hat{\mu}_i$ is the mean attribute of fish in cluster i and a_i is the number of fish of a given condition in cluster i (Cochran 1977). The estimate of $\hat{\mu}_i$ in r is $\hat{\mu}_i = \sum_{j=1}^{M_i} y_{ij} \Big/ M_i$ when one-stage sampling of individuals (all fish are measured in each cluster) occurs, and $\hat{\mu}_i = \sum_{j=1}^{m_i} y_{ij} \Big/ m_i$ when two-stage sampling (a sub-sample of fish is taken) occurs where y_{ij} is the attribute for fish j in cluster i and m_i is the number of fish in the cluster sub-sample. The estimator of p when a sub-sample of fish is measured is

$$\hat{p} = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} m_i}$$

The approximate variance estimator for r is

$$(3)\qquad \mathrm{var}(\hat{r})_u = \frac{N-n}{N} \cdot \frac{\sum_{i=1}^{n} \left(M_i/\overline{M}\right)^2 \left(\hat{\mu}_i - \hat{r}\right)^2}{n(n-1)}$$

and for p is

$$(4)\qquad \mathrm{var}(\hat{p})_u = \frac{N-n}{N} \cdot \frac{\sum_{i=1}^{n} \left(M_i/\overline{M}\right)^2 \left(\hat{p}_i - \hat{p}\right)^2}{n(n-1)}$$

where N is the total number of clusters with fish available for sampling, $\bar{M}$ is the mean number of fish per haul (cluster), n is the total number of clusters with fish, and p_i is the proportion for cluster i. If two-stage sampling is used, a variance term to account for sub-sampling must be added to the variance estimates above. For r, the second term (V2) is

$$V2_r = \frac{n}{N} \cdot \frac{1}{n^2 \bar{M}} \cdot \sum_{i=1}^{n} \frac{M_i^2}{m_i} \cdot \frac{M_i - m_i}{M_i} \cdot \frac{\sum_{j=1}^{m_i} (y_{ij} - \mu_i)^2}{m_i - 1}$$

and for p,

$$V2_p = \frac{n}{N} \cdot \frac{1}{n^2 \bar{M}} \cdot \sum_{i=1}^{n} \frac{M_i^2}{m_i} \cdot \frac{M_i - m_i}{M_i} \cdot \frac{\hat{p}_i(1-\hat{p}_i)}{m_i - 1}$$

The distribution of any discrete attribute of fish such as length or age may be estimated (assuming a multinomial distribution) by using the proportion equations applied to data for discrete bins. For example, with one-stage sampling, the proportion (p) of fish in the k^{th} bin class is estimated by:

$$\hat{p}_k = \frac{\sum_{i=1}^{n} a_{ik}}{\sum_{i=1}^{n} M_i}$$

The variance estimate would be the same as equation 4 except the k subscript would be added to designate the variance associated with the k bin class.

In practice, the finite population correction factor (*(N-n)/N*) may be dropped from the variance estimate because the number of clusters sampled is usually small compared to the total number available, and the variance term for sub-sampling may be negligible because *n/N* is usually small. In general, these estimates of mean, proportion and their variances are not unbiased; however, the bias should become negligible as cluster sample size increases (Cochran 1977).

A confidence interval for the estimate of r (and p) in SRCS is calculated by

$$\hat{r} - t_{\alpha/2,df} \cdot \sqrt{\text{var}(r)} \leq R \leq \hat{r} + t_{\alpha/2,df} \cdot \sqrt{\text{var}(r)}$$

where t is the two-tailed student's t critical value for α (the allowable probability of error) which provides 100(1- α)% confidence intervals, and df are the degrees of freedom associated with the estimate. The degrees of freedom will not be based on the number of individuals ($\sum M_i$ for one-stage sampling), as is often wrongly assumed, but will range between the number of clusters minus one and the number of individuals minus one depending on the level of within-cluster similarity. The df may be approximated by using the df equations derived by Hedges (2007) for a two-sample t-test for unequal cluster sizes corrected for clustering (Dr. L. Hedges, Northwestern University, pers. comm.). For fisheries data, the df obtained may be slightly optimistic because the design effect equation used to correct for clustering may under-estimate the true clustering effect for clusters of unequal size (Eldridge et al. 2006).

An alternate method for calculating confidence intervals is the *percentile bootstrap method* (Haddon 2001). Bootstrapping proceeds by generating b bootstrap samples, each consisting of data from n clusters drawn randomly with replacement from the original n clusters, and estimating r (or p) for each bootstrap sample. The estimate of the 100(1-α) % confidence intervals is determined by taking the 100(α/2) and 100(1-α/2) percentiles of the b bootstrap replicate samples sorted in ascending order (Haddon 2001). Advantages of this method are that asymmetrical confidence intervals may be produced and the df are not needed. A disadvantage is that bootstrapping may produce confidence intervals wider than they should be since the variance may be over-estimated for the sample sizes typically observed in fisheries research (see *Improved Variance Estimates* below).

196 **Components of Variance**

197 Variances of population attributes estimated by using cluster sampling tend to be larger
198 than variances estimated by using SRS because, not only are the attribute characteristics of the
199 population embedded in the estimate, but characteristics related to the population's spatial
200 distribution are as well. As long as attribute size and variance are not related to the density of
201 fish at a station and the level of within-cluster similarity does not depend on the total number of
202 fish caught, Pennington and Volstad (1994) demonstrated that the variance formula of *r* can be
203 written to show its variance components,

204 $$\mathrm{var}(\hat{r})_{PV} = \frac{\sigma_y^2}{\bar{M}n} \cdot \left(1 + \left(\left(1 + CV^2\right)\bar{M} - 1\right)\rho\right)$$

205 where σ^2_y is the population variance of attribute *y*, *CV* is the coefficient of variation for cluster
206 sizes (substituted for $s_m/\bar{M}$ in their equation 2.2), ρ is the intra-cluster correlation coefficient (a
207 measure of within-cluster similarity; Donner, 1986), and s_m is SRS sample standard deviation for
208 M_i values>0. The above equation shows that the variance of *r* is composed of $\sigma^2_y/\bar{M}\,n$, the
209 variance of the mean under SRS, multiplied by $1+((1+CV^2)*\,\bar{M}\,-1)\rho$, a variance inflation factor
210 (VIF) due to the effect of clustering. The VIF is related to the average cluster size, the variability
211 in size among clusters (*CV*) and the within-cluster similarity of individuals (ρ) (Pennington and
212 Volstad 1994). The relationship between the VIF and these parameters are shown in Figure 2 for
213 a range of parameter values. If ρ=0, the variance of the ratio estimator is equivalent to the
214 variance of the mean under SRS because there is no inflation (VIF = 1). When ρ>0, the variance
215 of the mean becomes greater than the variance under SRS because VIF increases as $\bar{M}$,*CV* and ρ
216 increase (Figure 2). Even if ρ is small, the effect of clustering still can be large because $\bar{M}$ and
217 *CV* may be very large for gear surveys (Pennington and Volstad 1994).

Parameters of $var(r)_{PV}$ may be estimated from observed attribute data following methods of Pennington et al. (2002) and Donner (1986). σ^2_y is estimated by

$$\hat{\sigma}_y^2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i}(M_i / m_i)\left(y_{ij} - \hat{r}\right)^2}{\left(\sum_{i=1}^{n} M_i\right) - 1}$$

and s_M for the calculation of *CV* is estimated by using the SRS estimator of the standard deviation (e.g., Sokal and Rohlf, 1995). ρ is estimated by using the results of a one-way analysis of variance (ANOVA) applied to attribute data where the cluster index is treated as a factor (Donner 1986). Given the between (BMS) and within (WMS) mean squares from the ANOVA table and assuming an underlying random effects model, ρ is calculated by:

$$\hat{\rho} = \frac{BMS - WMS}{BMS + (\overline{M}_{adj} - 1)WMS}$$

where $\overline{M}_{adj}$ is the adjusted mean cluster size for unequal cluster sizes defined by

$$\overline{M}_{adj} = \frac{1}{n-1}\left(\sum_{i=1}^{n} M_i - \frac{\sum_{i=1}^{n} M_i^2}{\sum_{i=1}^{n} M_i}\right)$$

Estimates of parameters of $var(r)_{PV}$ for length data of the top-five most abundant juvenile and adult fish and invertebrate species caught in common stratified random trawl and seine surveys from five states along the U. S. Atlantic coast are shown in Table 1. The values of intra-cluster correlation (range: 0.12-0.88; median: 0.40), $\overline{M}$ (range: 13.3-3,618; median: 49.0) and *CV* (range: 0.57-3.49; median: 1.38) are typical for species caught in these types of gear surveys (e.g., Pennington and Volstad, 1994). The moderate to high VIFs (range: 3.8-10,469; median:

75.3) produced from these parameters further demonstrates that clustering is a large contributor to the total variance of mean body sizes of most species (Table 1).

Effective Sample Size

The *effective sample size* (ESS) is a useful statistic for quantifying the amount of information in clustered data and is defined as the size of an independent sample that would equal the amount of information in the actual correlated sample (Faes et al 2009). In fisheries, ESS is the equivalent number of fish that would need to be sampled randomly to obtain the same level of precision observed for a cluster sampling estimator (Pennington and Volstad 1994; Pennington et al. 2002). In RCS, ESS is related to the level of intra-cluster correlation. When $\rho=0$, all observations are independent within a cluster and ESS equals the total number of observations. As ρ approaches 1 (perfect correlation), the ESS approaches the number of clusters in a sample (Faes et al, 2009). Pennington et al. (2002) showed that the ESS can be estimated by using the estimate of population variance, σ^2_y, and the variance of r:

$$\hat{m}_{eff} = \frac{\hat{\sigma}_y^2}{\mathrm{var}(\hat{r})}$$

where m_{eff} is the effective sample size. The ESS for fish length data are typically much smaller than the total number of fish measured and often range between 0.1% to 12% of the total number measured (Table 2; Pennington et al., 2002; Zhang and Cadrin, 2013). These ESS levels are equivalent to only collecting about one fish per tow or haul (m_{eff}/n; Table 2; Pennington et al. 2002; Zhang and Cadrin 2013).

Improved Variance Estimates

The usual variance estimators for *r* (Eq. 3) and *p* (Eq. 4) are approximations based on the delta method and may under-estimate the true variance when the sample sizes are small (Cochran 1977; Pennington and Volstad 1994). Improved estimates of variance can be obtained by using the computer-intensive jackknife or ordinary bootstrap methods (Haddon 2001) where resampling is made at the cluster level. Pennington and Volstad (1994) compared the usual approximation and jackknife estimators of standard error (square-root of variance) of *r* for mean length via simulation and found that the latter produced consistently more accurate estimates than the former. Kovar et al. (1988) found in a simulation study of stratified simple random sampling that bootstrap-based estimators tend to over-estimate the variance of ratios and this appears true for *r* at cluster samples sizes typically observed in fisheries surveys (Table 3).

HYPOTHESIS TESTING

The frequent application of statistics in fisheries research is to test some scientific hypothesis. Statistical techniques are important because results are not always clear-cut and tests are required to make decisions between alternative hypotheses. For statistical methods, a null hypothesis of no difference is defined before a test is performed and a significance level (e.g., 5%, 1%, etc.) is chosen that is used to decide if the null hypothesis should be rejected. Unfortunately, some samples may be very aberrant due to chance and will mislead us into rejecting a true null hypothesis, otherwise known as a Type I error. The chance of making a Type I error is equal to the level of significance (α) chosen expressed as a probability (Sokal and Rohlf 1995).

The implication of ignoring clustering of fisheries data is to inflate the nominal Type I error rate of statistical methods and is due to two common mistakes made by researchers. First, the variance of a population parameter is under-estimated (greatly in many cases) because the SRS variance formula is used, and this directly impacts standard statistical tests that use variance by inflating the calculated test statistics (e.g., t-test statistic; Sokal and Rohlf, 1995). Second, the *df* are calculated incorrectly by using the number of individuals rather than basing the *df* on the level of independent information (e.g., Hedges, 2007), and this results in the selection of a test's critical value used to accept or rejected the null hypothesis that is lower than it should be. Because of these common errors, the null hypothesis of no difference in statistical tests may be rejected more often than it should be for a specified α level.

The inflation of Type I error caused by these common errors can be severe. This is demonstrated in Figure 3 where mean lengths of two identical, simulated populations with specific intra-cluster correlation levels were compared by using the t-test for unequal variances (Sokal and Rohlf 1995) over a range of cluster sample sizes. A single population had 2,492 clusters of random sizes with 886,190 fish generated from a negative binomial distribution fitted to number per tow data for pinfish off Tampa Bay, Florida (Nelson 2002). For each intra-cluster correlation level, length data of each cluster were generated by drawing randomly individual lengths from a normal distribution parameterized with a randomly-drawn mean length (from a normal distribution with mean 147 mm TL and standard deviation adjusted to produce a desired intra-cluster correlation) and standard deviation of 18.14 mm. The second, identical population was created by simply copying the original population data. For each cluster sample size, an equal number of clusters was selected randomly from each population and the mean lengths compared by using the t-test. Type I error was calculated from the number of significant tests in 5,000

simulations. In Figure 3A, the SRS estimator of variance and the *df* based on the total individuals were used in the test. Theoretically, Type I error should not exceed 5% (α=0.05) at each cluster sample size because there should be no difference in mean length between the two populations except due to chance. This was only true when ρ=0.0 (Figure 3A). For ρ>0, Type I error increased over small (10-100) cluster sample sizes collected commonly in fisheries surveys (this is related to over-estimation of variance at low sample sizes) and it reached an asymptote at larger sizes. Even at low intra-cluster correlation levels (e.g., 0.02), Type I error was inflated by 14 times the nominal level (Figure 3A). When the *df* and test statistics were corrected for clustering (Hedges 2007), the Type I error did not exceed the nominal level at any cluster sample size or intra-cluster correlation level (Figure 3B). This example clearly shows that it is highly likely that significant differences will be found when there are no differences at all if clustering is ignored.

STATISTICAL ANALYSIS OF CLUSTERED DATA IN FISHERIES RESEARCH

When the goal of analysis is to compare statistically fisheries data among two or more groups of observations, the clustering of data must be taken into account to produce unbiased results and expected Type I error rates. There are three general approaches available for analyzing statistically clustered data: 1) cluster-level analysis with observations within a cluster reduced to single summary statistic (e.g., cluster mean or proportion), 2) individual level analysis of observations with standard formulae adjusted for clustering, and 3) multilevel models that explicitly take into account clustering. The choice of method will depend on research objectives, available data, and software availability. Galbraith et al. (2010) and Picquelle and Mier (2011) reviewed the performance of several statistical methods for clustered data described below.

Cluster-Level Analysis

A valid approach to analyzing clustered data is to reduce the multiple observations in each cluster to a single summary statistic, such as the mean or proportion, and use the resulting values in statistical tests. This cluster-level approach reduces the observations to the number of independent clusters and eliminates the violation of independence associated with clustered data (*df* are calculated using the number of clusters). For clusters with unequal sizes, clusters with more observations should be given more weight in the analysis (Galbraith et al. 2010), and Kerry and Bland (2001) recommend weighting the individual cluster summary values by the minimum variance (*w*) calculated as

$$w_i = \frac{M_i}{1+(M_i-1)\rho}$$

where *i* is the cluster index, ρ is assumed known, and the remaining definitions are as described previously. A common ρ for two or more groups may be estimated by using a one-way nested random effects model for unequal cluster sizes (Donner, 1986). Applying weights to data points is accomplished easily in most statistical software packages. Picquelle and Mier (2011) recommended cluster-level analysis as one of the best of five ANOVA methods they examined for comparing fish attributes. An advantage of cluster-level analysis is that it is straight-forward and standard methods found in most statistics books can be used. A disadvantage is that, by reducing the observations in each cluster to a single value, information regarding the individuals is lost and an analysis may not be as powerful as an approach that incorporates individual observations, particularly when cluster sample size is small.

Individual-Level Analysis Adjusted For Clustering Effects

In this approach, standard statistical methods are applied to the individual observations, but the calculated test statistics are adjusted to take into account intra-cluster correlation. Tests are available for normally-distributed data (two-sample t-test: Kish 1965; Hedges 2007; Wald test: Faes et al. 2009; ANOVA: Hedges and Rhoads 2011; Nested ANOVA: Hedges 2009), binary data (χ^2: Rao and Scott 1987; Rao and Scott 1992), and non-normally distributed data (Mann-Whitney U test: Rosner and Grove 1999; rank-sum test: Datta and Satten 2005). Typically, test statistics are adjusted by using the VIF to inflate the variance used in the calculation of the test statistic (Wears 2001), but for some tests, the choice of *df* was not always clear (Hedges 2007). Recently, Hedges (2007; 2009) and Hedges and Rhoads (2011) developed t-test and ANOVA methods for unequal cluster sizes that adjust the test statistics and compute the *df* based on a common intra-cluster correlation coefficient among groups. An advantage of these methods is that the *df* are determined for the level of independent information rather than assuming the *df* are based on the number of individuals (Blair and Higgins 1986) or the number of clusters (Wears 2001). Disadvantages of the above methods are that the VIF formula used to correct the test statistics may under-estimate the true clustering effect for clusters of unequal size (Eldridge et al. 2006).

An alternate approach to comparing clustered data of groups that is applicable to fisheries data is to use randomization methods to develop tests for null hypotheses (Haddon 2001). Randomization methods are very flexible, have no distributional assumptions, and can be developed to test null hypotheses concerning clustered data from one to many groups (Manly 1997). In a general randomization test, an observed test statistic, such as the difference in means between two groups, is compared to an empirical probability density function (*pdf*) of the test

statistic under the null hypothesis of no difference. The empirical *pdf* is created by the repeated ($\geq$1,000) randomization (without replacement) of clusters from all groups into groups with samples sizes equal to the number of clusters in original groups and calculating the test statistic for each randomization. The null hypothesis of no difference is rejected if the original test statistic exceeds the empirical value at which the proportion of all empirical values greater than or equal to original test statistic is less than or equal to a specified significance level (e.g., α=0.05). Some advantages of randomization tests are that the test statistic can be any single, smooth, continuous valued function of the data, intra-cluster correlation is taken directly into account in the null *pdf* because resampling takes place at the cluster level, and the *df* are not needed to select a critical value. Some disadvantages of randomization tests are that they do not work well for small sample sizes (like most statistical tests), there is no straight-forward method to correct for comparison-wise error rates in multiple comparisons, and some level of programming will be required.

Multilevel Modeling Approaches

When analysis of clustered data requires more complex methods, linear models that take intra-cluster correlation directly into account are available in most statistical software packages. The two approaches reviewed below, linear mixed effects models (LME) and generalized estimating equations (GEE), handle intra-cluster correlation in different ways. These methods are particularly useful when there are covariates that need to be included in the analysis, but they may require a minimum of 25 observations in each of 25 clusters with a single level of clustering to achieve their asymptotic performance (Ukoumunne et al. 1999).

Linear mixed effects models are generalized linear models that induce the within-cluster correlation by incorporating random cluster-specific effects (Zuur et al. 2009). Covariates are treated as either fixed or random effects. The LME models assume normally-distributed errors. Like general linear models, the error distribution must be assessed for adequate model fit via standardized residuals plots and other diagnostic measures. Since the model may contain a mixture of fixed and random effects, optimal model selection is accomplished through a top-down strategy using restricted maximum likelihood ratio tests if nested, or by comparing Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) values among models if not nested (Pinheiro and Bates 2004; Zuur et al. 2009). For analyses comparing fish attribute data, Picquelle and Mier (2011) recommended using equal-variance and unequal-variance nested mixed effect models with restricted maximum likelihood estimation if the number of fish are unbalanced between levels of a fixed covariate and if there is also heterogeneity among hauls within each level.

Generalized estimating equations are generalized linear models with extensions that accommodate non-normal error distributions, mixed and fixed effects, a degree of nonlinearity in the model structure, and correlations among observations within independent clusters (Hardin and Hilbe 2003). The GEE models separately the mean response and treats the dependence between observations as a nuisance parameter. The GEE estimates parameter coefficients and a correlation matrix with a user-specified structure (for fisheries data, the *exchangeable or compound-symmetry* correlation structure is useful since a common correlation parameter is assumed for all clusters) using quasi-likelihood methods (i.e., mean and variance function are not from the exponential family). Like GLMs, link and variance functions and selection of error distribution must be assessed for adequate model fit via standardized residuals plots and other

diagnostic measures. The best correlation structure is chosen by using the Quasi-likelihood under Independence model information criterion (QIC) and the QICu is used in model selection (Hardin and Hilbe 2003). Model comparisons can also be made via naïve likelihood ratio tests.

Non-linear Modeling Approaches

Non-linear mixed effects models are available for fitting non-linear mechanistic models (e.g., von Bertalanffy growth equation) that incorporate random cluster-specific effects, and for testing relationships between model parameters and covariates such as group membership (Pinheiro and Bates 2004). Model parameters and covariates are treated as either fixed or random effects where random effects represent deviations of the individual parameters from the fixed effect. Like general linear models, normally-distributed errors are assumed, and model fit must be assessed via standardized residuals plots. Since the model may contain a mixture of fixed and random effects, optimal model selection is accomplished through a similar top-down strategy as linear mixed effects models (Pinheiro and Bates 2004: 365-368) using maximum likelihood ratio tests if nested, or by comparing AIC or BIC values among models if not nested (Pinheiro and Bates 2004).

An alternate approach to model fitting is to use bootstrap replicates of the clusters and estimate parameters of a non-linear model for each bootstrap sample using any non-linear least-squares estimation method. The mean and standard deviation of the bootstrap parameters would be the estimates of model parameters and their associated standard errors. Confidence intervals for the parameters can be obtained by using the percentile bootstrap method (Haddon 2001).

IMPROVING PRECISION OF ESTIMATES AND POWER OF STATISTICAL ANALYSES

Regardless of the type of data measured in fisheries research, high variation will most likely be an inherent feature of it when collected via cluster sampling. However, there are changes to the survey design and sampling protocols that can be made to improve the precision of estimates. The general recommendation for improving precision in fisheries research is to increase the number of tows (or hauls) made during a survey because it is the variation among clusters that determines mostly the precision of estimates when within-cluster similarity is high (Pennington and Volstad 1994; Bogstad et al. 1995; Crone, 1995; Pennington et al. 2002; Zhang and Cadrin 2013). This can be done without increasing survey costs much or altering attribute composition by reducing tow duration and using the time saved to increase the number of stations (Godo et al. 1990; Pennington and Volstad 1991; Pennington and Volstad 1994). Shorter tow duration would collect fewer individuals, but the resulting sample would be more representative of the entire population (Pennington et al. 2002). In addition, the sampling strategy for individuals that produces the most precise estimate of an attribute in conjunction with increases in the number of stations can be determined through scenario analysis by using the $var(r)_{PV}$ equation (see Bogstad et al. 1995) or simulation and, in several cases, the optimal sampling strategy has been to measure fewer individuals at each location (Horppila and Peltonen 1992; Bogstad et al. 1995; Zhang and Cadrin 2013) or collect fewer individuals from each commercial trip when sampling catch (Aanes and Pennington, 2003) .

If a survey is conducted to test hypotheses about populations, the type of analyses and power of the statistical tests must be considered in the planning of the sampling design. If the intention is to analyze cluster-level variates, sample size and power calculations may be

performed to estimate the required number of clusters required to achieve a specified difference (between means or proportions) with a specified power by using the standard sample size equations for SRS given in most statistical textbooks (e.g., Zar 1999). If use of statistical tests adjusted for clustering is planned for group comparisons, standard sample size equations may still be used to determine number of individuals required to achieve a given power but they must be adjusted for clustering by using the VIF (Ukoumunne et al 1999). If the intention is to analyze data by using LME or GEE models, sample size and power calculations may be determined, but analysis will be rather complex if covariates are to be included (Liu and Liang 1997; Berger and Tan 2004). Of course, these types of analyses should be conducted in concert with analyses used to determine changes in survey design needed to improve the precision of estimates (discussed above) since sample size determination will depend ultimately on the level of variation in data used in the statistical tests.

In reality, the ability to make changes to a survey design may be limited. If the survey is well-established or was originally designed with multiple objectives in mind, changes to the design (e.g., tow duration, number of stations, etc.) will likely be met with resistance by the powers-that-be because of concerns over historical continuity or impacts on data collected for other species. Changes to sampling protocols (how fish are sub-sampled at each station) may be met with less resistance, but are less likely to make big improvements in the precision of estimates or power of statistical tests because it is the variation in data among stations that determines mostly the precision of estimates. Unless the survey is designed from the start with estimation of population attributes and hypothesis testing in mind, there may be little the researcher can actually do to improve precision of estimates and power of statistical analyses in fisheries research.

COMPUTER SOFTWARE

Many statistical techniques discussed in this paper are implemented in the *R* statistical environment (*R* Development Core Team 2013). Package *fishmethods* (Nelson 2013) provides functions that estimate intra-cluster correlation, calculate attribute mean and different variances (approximate, jackknife and bootstrap) for simple random cluster sampling, calculate the two-sample t-test adjusted for clustering (Hedges, 2007), and use randomization tests to compare length frequencies for simple and stratified random cluster sampling. Package *survey* (Lumley, 2012) provides functions for estimation of survey statistics for more complex designs like stratified random cluster sampling. Jackknifing can be performed by using package *bootstrap* (Tibshirani 2012) and bootstrapping can be performed using either packages *bootstrap* or *boot* (Canty and Ripley 2013). The linear and non-linear mixed effects models are implemented in package *nlme* (Pinhero and Bates 2013), and GEEs are implemented in package *geepack* (Yan et al. 2012). The *R* code to implement the rank-sum test for clustered data is available from Galbraith et al. (2010). Methods for LME and GEE are also implemented in other statistical packages such as SAS (SAS 2013) or STATA (Rabe-Hesketh and Skrondal 2012).

CONCLUSIONS

Cluster sampling is a common survey design used pervasively in fisheries research to sample on fish populations, but it is not widely recognized. Because fish collected via cluster sampling are not independent of each other, standard SRS estimators and statistical tests that assume independence cannot be used to make inference about fish populations. If the clustered nature of fisheries data is ignored, the results and conclusions of analyses may be severely biased because statistical tests will often indicate significant differences in group comparisons when there are none (Type I error rate inflation). Three general methods for analyzing clustered data

(cluster-level analysis, individual-level analysis adjusted for clustering effects and statistical modeling approaches) take into account the clustered nature of fisheries data and should be used to avoid inflation of Type I error. The precision of attribute estimates and power of statistical tests may be improved by adding more stations to current surveys since it is the variation in data among stations that determines mostly the precision of estimates.

ACKNOWLEDGEMENTS

Funding for this study was provided by the USFWS Sportfish Restoration Program Grant F-57-R. I wish to thank Jon Volstad and Michael Pennington for their aid with the MSE simulations. Paul Kostovick of the National Marine Fisheries Service provided the original survey data used in Pennington and Volstad's paper to test the MSE simulations. Thanks go to Jeremy King of the Massachusetts Division of Marine Fisheries, Kurt Gottschall of Connecticut Department of Environmental Protection, Linda Barry of the New Jersey Division of Fish and Wildlife, Jason Rock of the North Carolina Division of Marine Fisheries, and Deborah Leffler of the Florida Fish and Wildlife Conservation Commission for providing survey data. The analyses and conclusions drawn using the survey data are those of the author and do not necessarily represent the views of contributing agencies. Thanks to Chris Legault of the National Marine Fisheries Service, Mike Bednarski of the Massachusetts Division of Marine Fisheries and two anonymous reviewers for comments that helped improve the quality of the manuscript. This is Massachusetts Division of Marine Fisheries Contribution Number 44.

REFERENCES

Aanes, S. and M. Pennington. 2003. On estimating the age composition of the commercial catch of Northeast Arctic cod from a sample of clusters. ICES Journal of Marine Science 60: 297-303.

Atlantic States Marine Fisheries Commission. 1994. Proceedings of the workshop on the collection and use of trawl survey data for fisheries management. Special Report No. 35, December 1994. 187 p.

Berger, M. P. F., and F. E. S. Tan. 2004. Robust designs for linear mixed effects models. Journal of the Royal Statistical Society, Series C (Applied Statistics) 53: 569-581.

Blair,R. C., and J. J. Higgins. 1986. Comment on "Statistical power of group mean as the units of analysis". Journal of Educational Statistics 11: 161-169.

Bogstad, B., and M. Pennington, J. H. Volstad. 1995. Cost-efficient survey designs for estimating consumption by fish. Fisheries Research 23: 37-46.

Canty, A., and B. Ripley. 2013. Package *boot*: Bootstrap R (S-Plus) functions, version 1.3.9. http://cran.r-project.org/web/packages/boot/boot.pdf.

Cochran, W. G. 1977. Sampling techniques, 3nd edition. John Wiley and Sons, Inc., New York.

Crone, P. 1995. Sampling design and statistical considerations for the commercial groundfish fishery of Oregon. Canadian Journal fo Fisheries and Aquatic Sciences: 716-732.

553 Datten, S., and G. A. Satten. 2005. Rank-sum tests for clustered data. Journal of the American
554 Statistical Association 100: 908-915.

555 Donner, A. 1986. A review of inference procedures for the intraclass correlation coefficient in
556 the one-way random effects model. International Statistical review 54: 67-82/

557 Eldridge, S. M., D. A. Ashby and S. Kerry. 2006. Sample size for cluster randomized trials:
558 effect of coefficient of variation of cluster size and analysis method. International
559 Journal of Epidemiology 35: 1292-1300.

560 Faes, C., G., Molenberghs, M. Aerts, G. Verbeke and M. G. Kenward. 2009. The effective
561 sample size and an alternative small-sample degrees of freedom method. American
562 Statistical Association 63: 389-399.

563 Fogarty, M. J. 1985. Statistical considerations in the design of trawl surveys. FAO Fisheries
564 Circular Number 786 FIRM/FIPL/C786.

565 Galbraith, S., J. A. Daniel and B. Vissel. 2010. A study of clustered data and approaches to its
566 analysis. The Journal of Neuroscience 30: 10601-10608.

567 Godo, O. R., M. Pennington, and J. H. Volstad. 1990. Effect of tow duration on length
568 composition of trawl catches. Fisheries Research 9: 165-179.

569 Haddon, M. 2001. Modelling and quantitive methods in fisheries. Chapman & Hall/CRC, Boca
570 Raton.

571 Hardin, J. W., and J. M. Hilbe. 2003. Generalized estimating equations. Chapman & Hall/CRC,
572 Boca Raton.

573 Hedges, L. V. 2007. Correcting a significance test for clustering. Journal of Educational and
574 Behavioral Statistics. 32: 151-179.

575 Hedges, L. V. 2009. Adjusting a significance test for clustering in designs with two levels of
576 nesting. Journal of Educational and Behavioral Statistics. 34: 464-490.
577
578 Hedges, L. V. and C. H. Rhoads. 2011. Correcting an analysis of variance for clustering. British
579 Journal of Mathematical and Statistical Psychology. 62: 20-37.

580 Horppila, J. and H. Peltonen. 1992. Optimizing sampling from trawl catches: contemporaneous
581 multistage sampling for age and length structures. Canadian Journal of Fisheries and
582 Aquatic Sciences. 49: 1555-1559.

583 Kerry, S. M. and J. M. Bland. 2001. Unequal cluster sizes for trials in English and Welsh general
584 practice: implications for sample size calculations. Statistics in Medicine 20:377-390.

585 Kish, L. 1965. Survey sampling. John Wiley and Sons, Inc. New York.

586 Kovar, J., J. N. K. Rao, C. F. J. Wu. 1988. Bootstrap and other methods to measure errors in
587 survey estimates. Canadian Journal of Statistics (suppl) 16: 25-45.

588 Lohr, S. L. 1999. Sampling: design and analysis. Duxbury Press, Boston.

589 Lui, G., and K. Liang. 1997. Sample size calculations for studies with correlated observations.
590 Biometrics 53: 937-947.

591 Lumley, T. 2012. Package *survey*: analysis of complex survey samples. http://cran.r-
592 project.org/web/packages/survey/survey.pdf

593 Manly, B. F. J. 1997. Randomization, bootstrap and Monte Carlo methods in biology, 2nd edition.
594 Chapman and Hall, New York.

595 Milikin, M. 1993. Predation risk and feeding behavior. Pages 285-306 *in* T. J. Pitcher, editor.
596 Behavior of teleost fishes, 2nd Edition. Chapman and Hall, Fish and Fisheries Series 7,
597 New York.

598 Nelson, G. A. 2002. Age, growth, mortality, and distribution of pinfish (*Lagodon rhomboides*)
599 in Tampa Bay and adjacent Gulf of Mexico waters. Fishery Bulletin 100: 582-592.

600 Nelson, G. A. 2013. Package *fishmethods*: Fisheries Science Methods and Models in R.
601 http://cran.r-project.org/web/packages/fishmethods/fishmethods.pdf.
602

603 Osenberg, C. W., M. H. Olson and G. G. Mittelbach. 1994. Stage structure in fishes: resource
604 productivity and competition gradients. Pages 151-170 *in* D.J. Stouder, K. L. Fresh and
605 R. J. Feller, editors. Theory and application in fish feeding ecology. Belle W. Baruch
606 Library in Marine Science 18, University of South Carolina Press, Columbia, South
607 Carolina..

608 Pennington, M., and J. H. Volstad. 1991. Optimum size of sampling unit for estimating the
609 density of marine populations. Biometrics 47: 717-723.

610 Pennington, M., and J. H. Volstad. 1994. Assessing the effect of intra-haul correlation and
611 variable density on estimates of population characteristics from marine surveys.
612 Biometrics 50: 725-732.

613 Pennington, M., L. M. Burmeister, and V. Hjellvik. 2002. Assessing the precision of frequency
614 distributions estimated from trawl-survey samples. Fishery Bulletin 100: 74-80.

615 Pinheiro, J. C., and D. M. Bates. 2004. Mixed-effects models in S and S-plus. Statistics and
616 Computing Series, Springer Science and Business Media, LLC, New York.

617 Pinheiro, J. C., and D. M. Bates. 2013. Linear and nonlinear mixed effects models in R.
618 http://cran.r-project.org/web/packages/nlme/nlme.pdf.

619 Picquelle, S. J. and K. L. Mier. 2011. A practical guide to statistical methods for comparing
620 means from two-stage sampling. Fisheries Research 107: 1-13.

621 Pitcher, T. J., and J. K. Parrish. 1993. Functions of shoaling behavior in teleosts. Pages 363-439
622 *in* T. J. Pitcher, editor. Behavior of teleost fishes, 2nd Edition. Chapman and Hall, Fish
623 and Fisheries Series 7, New York.

624 Rabe-Hesketh, S., and A. Skrondal. 2012. Multilevel and longitudinal modeling using Stata,
625 Third Edition. Stata Press. 974 p.
626
627 Rao, J. N. K., and A. J. Scott. 1984. On chi-squared tests for multi-way contingency tables with
628 cell proportions estimated from survey data. The Annals of Statistics 12: 46-60.
629
630 Rao, J. N., and A. J. Scott. 1992. A simple method for the analysis of clustered binary data.
631 Biometrics 48: 577-585.
632

R Development Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org/.

Rosner, B., and D. Grove. 1999. Use of the Mann-Whitney U-test for clustered data. Statistics in Medicine 18:1387-1400.

SAS Institute. 2004. SAS/Stat 9.1 user's guide volumes 1-7. SAS Inst., Inc. Cary, NC.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry, 3rd Edition. W. H. Freeman and Company, New York.

Thompson, S. K. Sampling, 2nd Edition. John J. Wiley and Sons, Inc., New York.

Tibshirani, R. 2012 (1992). S original, from StatLib, R port by Friedrich Leisch.. *bootstrap*: functions for the Book "An Introduction to the Bootstrap". R package version 2012.04-0. http://cran.r-project.org/web/packages/bootstrap/bootstrap.pdf

Ukoumunne, O. K., M. C. Gulliford, S. Chinn, J. A. C. Sterne and P. G. J. Burney. 1999. Methods for evaluating area-wide and organization-based interventions in health and health care: a systematic review. Health Technology Assessment 3(5): 98 p.

Wears, R. L. 2002. Advanced statistics: statistical methods for analyzing cluster and cluster-randomized data. Academic Emergency Medicine 9: 330-340.

Yan, J., S., Hojsgaard, and U. Halekoh. 2012. Package *geepack*: Generalized Estimating Equations package. http://cran.r-project.org/web/packages/geepack/geepack.pdf.

Zar, J. H. 1999. Biostatistical analysis, 4th Edition. Prentice Hall, Upper Saddle River, New Jersey.

Zhang, Y., and S. X. Cadrin. 2013. Estimating effective sample size for monitoring length distributions: a comparative study of Georges Bank groundfish. Transactions of the American Fisheries Society 142: 59-67.

Zuur, A. F., E. N. Ieno, N. J. Walker, A. A. Saveliev, and G. M. Smith. 2009. Mixed effects models and extensions in ecology with R. , Springer Science and Business Media, LLC, New York.

FOOTNOTE

[1] Although capable of selecting individuals, gears such as rod-and-reel and spear still produce clustering because sampling efforts are usually limited to several small areas within the range of a population.

Table 1. Estimates of mean body length (r) and variance components for the top five most abundant species caught in Atlantic state coastal stratified random surveys during spring 2008. n_T = total hauls made during the survey, n is the number of hauls with fish, $\bar{M}$ is the mean cluster size for $M_i>0$, σ_y is the standard deviation of the population length distribution, ρ is the intra-cluster correlation coefficient, *CV* is the coefficient of variation of positive cluster sizes, and *VIF* is the approximate variance inflation factor due to clustering. Only species with $n>3$ are shown.

State	Gear	n_T	Species	n	$\bar{M}$	r(cm)	σ_y	ρ	*CV*	*VIF*
MA	Trawl[a]	103	Atlantic Cod	98	254.0	6.1	6.7	0.84	3.09	2254.3
			Sand Lance	32	316.9	11.9	2.1	0.68	2.81	1918.1
			Scup	31	217.3	20.2	3.4	0.31	1.65	251.8
			Longhorn Sculpin	52	114.0	23.6	3.8	0.28	1.78	134.9
			Yellowtail Flounder	51	66.7	31.4	3.7	0.24	2.06	85.9
CT	Trawl[b]	40	Scup	39	677.2	18.5	4.9	0.40	1.38	798.6
			Butterfish	25	174.8	15.3	2.2	0.13	1.20	56.6
			Longfin Squid	28	66.0	10.4	5.2	0.21	2.11	75.3
			Winter Flounder	37	39.5	16.3	5.0	0.16	1.00	13.3
			Windowpane Flounder	23	15.6	22.3	5.5	0.12	0.76	3.8
NJ	Trawl[c]	39	Butterfish	39	3,618.1	7.9	2.5	0.50	2.19	10469.7
			Scup	33	1,032.4	9.0	0.9	0.21	3.11	2305.0
			Longfin Squid	39	655.6	6.2	3.8	0.16	0.96	200.0
			Round Herring	14	363.3	8.9	0.8	0.88	3.49	4231.4
			Northern Searobin	38	124.1	20.3	3.3	0.37	3.30	545.8
NC	Trawl[d]	54	Spot	46	117.4	9.9	2.9	0.29	0.57	45.7
			Atlantic Croaker	39	103.5	10.8	3.4	0.34	0.67	52.1
			Blue Crab	53	34.2	7.0	2.8	0.14	1.33	14.0
			Brown Shrimp	27	49.0	9.4	1.5	0.26	0.99	25.8
			Pink Shrimp	37	30.0	11.6	1.6	0.19	1.30	16.3
FL	Trawl[e]	15	Portunus Crab	13	32.8	4.1	0.9	0.71	1.62	85.2
			Pinfish	8	17.0	8.0	2.7	0.67	1.24	29.0

				Pigfish	6	18.2	6.8	2.0	0.58	1.56	36.5
				Leopard Searobin	12	9.1	8.4	3.5	0.50	1.13	10.9
				Silver Jenny	8	9.9	8.0	1.7	0.68	1.51	22.5
	FL	Seine[f]	34	Pinfish	10	176.1	3.8	1.1	0.39	1.20	168.0
				Silver Perch	8	46.2	3.5	1.1	0.41	1.33	52.5
				Rainwater Killifish	7	47.4	2.7	0.4	0.45	2.02	109.0
				Tidewater Mojarra	7	28.4	2.4	0.6	0.81	1.62	83.4
				Clown Goby	6	30.3	3.1	0.7	0.52	0.68	23.5
	FL	Seine[g]	20	Pinfish	15	137.0	9.9	2.5	0.40	1.36	156.8
				Hardhead Catfish	11	42.9	23.3	5.4	0.81	2.34	224.0
				Tidewater Mojarra	9	13.9	9.5	0.8	0.36	0.68	8.0
				Silver Jenny	7	17.3	8.4	0.8	0.43	1.62	27.3
				Pigfish	6	13.3	9.8	2.4	0.60	1.03	16.9

[a] Massachusetts Division of Marines Fisheries, coastal state waters, May, 15.5-m footrope length, 20 min tow duration.

[b] Connecticut Department of Environmental Protection, Long Island Sound, June, 14.0-m footrope length, 30 min tow duration.

[c] New Jersey Department of Fish, Game and Wildlife, coastal state waters, June, 30.5- m footrope length, 20 min tow duration.

[d] North Carolina Division of Marine Fisheries, Pamlico Sound, June, 10.4-m footrope length, 20 min tow duration,

[e] Florida Fish and Wildlife Conservation Commission, Tampa Bay, June, 6.1-m footrope length, 10 min tow duration.

[f] Florida Fish and Wildlife Conservation Commission, Tampa Bay, June, 21.0 m x 1.8 m center bag seine

[g] Florida Fish and Wildlife Conservation Commission, Tampa Bay, June, 183.0 m x 2.5 m center bag seine

Table 2. Estimates of effective sample size (m_{eff}) for body size of the top five most abundant species caught in Atlantic state coastal stratified random surveys. *n* is the number of hauls with fish, *M* is the total number of fish caught, *m* is the number of fish measured, σ^2_y is the estimate of variance of the population length distribution, $var(r)_J$ is the variance of the mean body size from jackknifing.

State	Species	*n*	*M*	*m*	σ_y^2	$var(r)_J$	m_{eff}	m_{eff}/n
MA	Atlantic Cod	98	24,897	3,620	45.1	0.69	65	0.7
	Sand Lance	32	10,142	967	4.2	1.34	3	0.1
	Scup	31	6,737	1,555	11.7	0.51	23	0.7
	Longhorn Sculpin	52	5,928	2,227	14.1	0.30	47	0.9
	Yellowtail Flounder	51	3,403	1,315	13.9	0.36	39	0.8
CT	Scup	39	26,409	6,773	24.2	0.68	36	0.9
	Butterfish	25	4,369	878	4.7	0.06	81	3.2
	Longfin Squid	28	1,849	965	27.0	0.84	32	1.1
	Winter Flounder	37	1,461	1,279	25.1	0.19	135	3.7
	Windowpane Flounder	23	360	344	30.1	0.27	110	4.8
NJ	Butterfish	39	141,106	2,966	6.3	0.63	10	0.3
	Scup	33	34,071	1,208	0.8	0.01	56	1.7
	Longfin squid	39	25,570	3,828	14.4	0.09	158	4.1
	Round herring	14	5,087	329	0.6	1.11	1	<0.1
	Northern searobin	38	4,716	1,264	11.1	1.48	8	0.2
NC	Spot	46	5,400	1,878	8.6	0.08	109	2.4
	Atlantic Croaker	39	4,035	1,288	11.3	0.14	79	2.0
	Blue Crab	53	1,812	1,695	8.0	0.04	216	4.1
	Brown Shrimp	27	1,323	618	2.2	0.04	53	2.0
	Pink Shrimp	37	1,112	613	2.4	0.04	61	1.6
FL	Portunus Crab	13	426	90	0.7	0.12	6	0.5
	Pinfish	8	136	63	7.2	2.60	3	0.3

		Pigfish	6	109	30	4.2	2.38	2	0.3
		Leopard Searobin	12	109	76	12.3	2.06	6	0.5
		Silver Jenny	8	79	38	2.7	1.80	2	0.2
	FL	Pinfish	10	1761	133	1.2	0.13	9	0.9
		Silver Perch	8	370	84	1.1	0.18	6	0.8
		Rainwater Killifish	7	332	45	0.2	0.09	2	0.3
		Tidewater Mojarra	7	199	49	0.4	0.35	1	0.1
		Clown Goby	6	182	69	0.5	0.11	4	0.7
	FL	Pinfish	15	2055	261	6.3	0.83	8	0.5
		Hardhead Catfish	11	472	100	29.4	43.46	1	0.1
		Tidewater Mojarra	9	125	113	0.6	0.04	15	1.6
		Silver Jenny	7	121	61	0.6	0.14	4	0.6
		Pigfish	6	80	60	5.6	0.91	6	1.0

Table 3. Simulation results for assessing the performance of the usual approximation, the jackknife and the ordinary bootstrap estimators of standard error (square-root of variance) of the ratio estimator of mean length following Pennington and Volstad (1994). For each species, 2,000 samples of cluster sample sizes 10, 20 and 30 were generated from the positive catches. CV is the coefficient of variation for the nonzero catches, and the true mean square error (MSE) is derived from the 2,000 simulations. For each simulation, 200 bootstrap replicates were generated to estimate the standard error. *n* is the cluster sample size.

				True	Average SE			Percent Deviation		
State	Species	CV	*n*	$\sqrt{MSE}$	Appr.	Jack.	Boot.	Appr.	Jack.	Boot.
MA	Atlantic Cod	3.09	10	6.09	3.43	4.96	6.80	-44	-19	12
			20	3.67	2.35	3.42	4.14	-36	-7	13
			30	2.58	1.87	2.65	3.00	-28	3	16
MA	Longhorn Sculpin	1.78	10	1.04	0.73	0.89	1.15	-30	-14	11
			20	0.78	0.64	0.67	0.84	-18	-14	8
			30	0.66	0.57	0.58	0.68	-14	-12	3
MA	Yellowtail Flounder	2.06	10	1.13	0.67	0.90	1.24	-41	-20	10
			20	0.88	0.68	0.74	0.99	-23	-16	13
			30	0.70	0.61	0.66	0.79	-13	-6	13
NC	Spot	0.57	10	0.58	0.51	0.53	0.59	-12	-10	2
			20	0.41	0.40	0.40	0.41	-3	-4	-1
			30	0.34	0.33	0.33	0.34	-1	-3	1
NC	Blue Crab	1.33	10	4.58	4.11	4.43	5.07	-10	-3	11
			20	3.21	3.03	3.08	3.34	-6	-4	4
			30	2.57	2.50	2.52	2.66	-3	-2	4

FIGURE CAPTIONS

Figure 1. Depictions of random sampling (top) and cluster sampling (bottom) of fish populations.

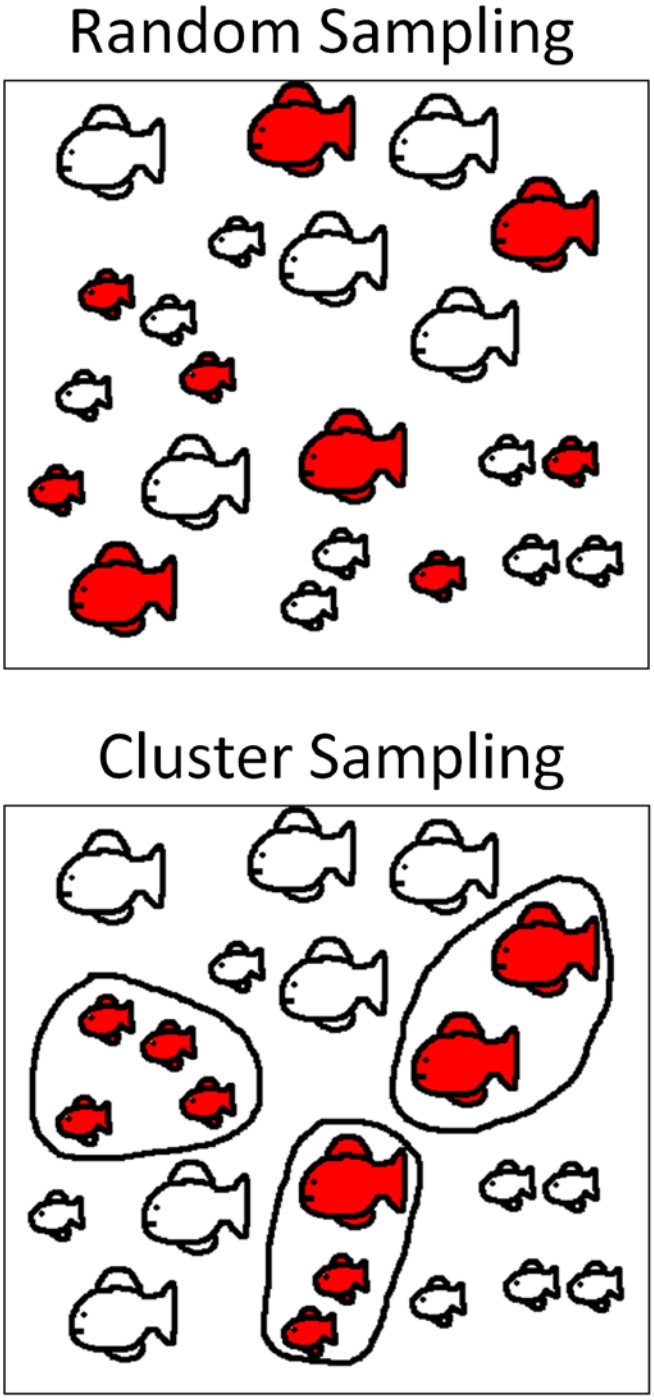
Figure 2. Contour plots of the variance inflation factor (VIF) due to clustering for a range of mean cluster sizes and coefficients of variation at ten intra-cluster correlation levels.

Figure 3. Type I error rates for a t-test with unequal variances used to compare mean lengths of two identical populations with specific intra-cluster correlation levels over a range of cluster sample sizes A) ignoring clustering and B) corrected for clustering. See text for more information.

795

796 Figure 1.

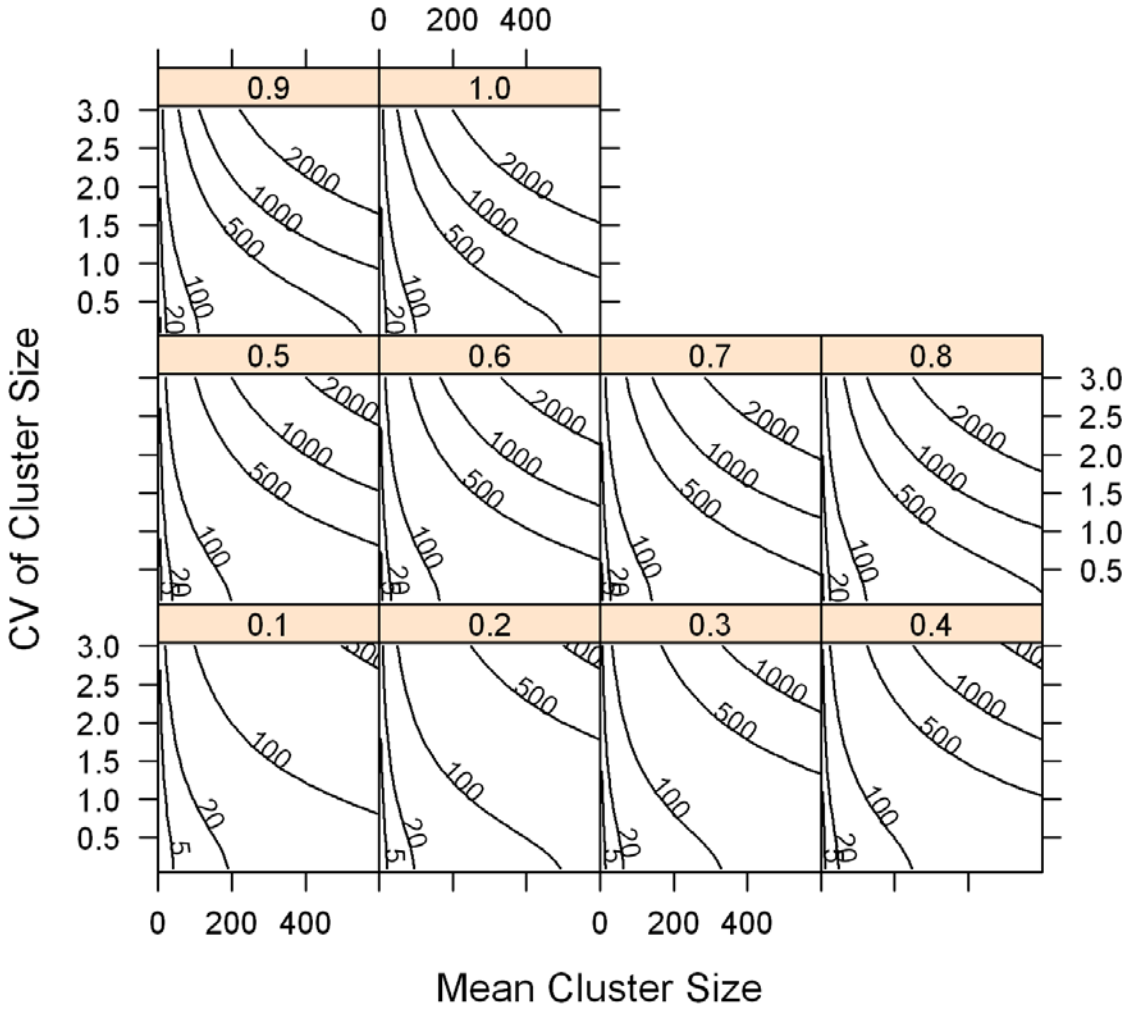
797



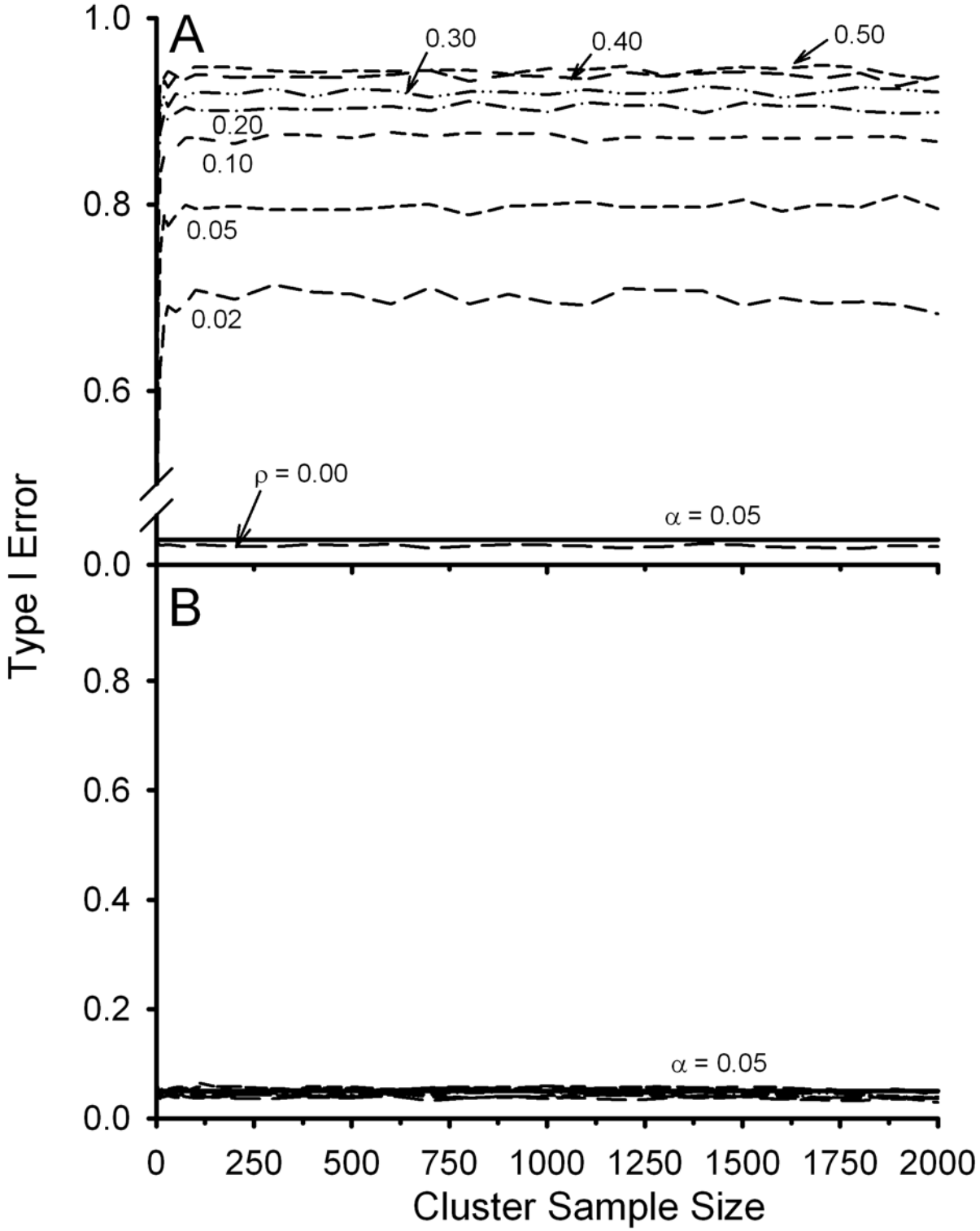
798 Figure 2.

799

800



801 Figure 3



802