

Chapter 11: Statistical Analysis and Data Management

MICHAEL C. QUIST, KIMBERLY I. BONVECHIO, AND MICHEAL S. ALLEN

11.1 INTRODUCTION

Collecting fisheries data requires extensive time and financial commitments. Given this high level of investment by management agencies and individual biologists, data storage, summarization, and analysis should be a high priority to ensure that the integrity and accessibility of collected information are maximized over time. Arguably, the most important aspect of sampling is completing an appropriate and thorough analysis of collected data. Standard sampling procedures help ensure that data analysis and database management are appropriate and efficient, but even when data are collected using standardized methods, data structure will vary among species and systems making standardized analyses difficult and sometimes impossible. Different agencies may also have different database management needs and structures, which can influence how data are stored and later accessed for analysis. The purpose of this chapter is to provide an overview of data summarization and analysis techniques, sample-size estimators, and principles of database management. Whenever appropriate, we guide readers to more detailed sources of information because many of these topics have been discussed extensively in the fisheries literature.

11.2 DATA ANALYSIS

Fisheries scientists typically collect large amounts of data during standard sampling surveys, and properly interpreting these data is necessary for making sound management decisions. Data analysis is often hierarchical in that some analyses focus on the estimation of simple summary statistics (e.g., mean, variance), while others rely on inferential statistical tests or modeling. Although fisheries data can be summarized using a variety of approaches, a few techniques have become standard in the profession, the most common of which are discussed in this chapter. Many texts are available on statistical techniques for fisheries data (e.g., Everhart et al. 1975; Jongman et al. 1995; Murphy and Willis 1996; Guy and Brown 2007), so these efforts are not reproduced in this chapter. Rather, we provide the most current references for analyzing fisheries data and encourage biologists to use these resources for more detailed presentations of analytical methods for fisheries data. In particular, we recommend Brown and Austen (1996) as a starting point for readers not familiar with the issues and idiosyncrasies of analyzing such data.

11.2.1 Abundance

Estimating the abundance of fish in a system is probably one of the most common objectives of sampling because estimates of fish abundance are critical for assessing management activities. Abundance estimates can be categorized as those focusing on either absolute or relative abundance. Absolute abundance refers to the total number of organisms in a system (i.e., density or population estimates), whereas relative abundance provides an index (e.g., catch per unit effort [CPUE]) of absolute abundance. Relative abundance should not be confused with relative species composition, which is the proportional (percentage) numerical or gravimetric abundance of a species within a collection of species.

Various techniques are used to estimate absolute abundance of fish in different systems, and they vary from simple techniques that can be executed using a pocket calculator to complex techniques requiring a high level of statistical expertise. Techniques used to estimate population density include hydroacoustic, mark-recapture, and depletion techniques. Broad overviews of these techniques can be found in Everhart et al. (1975), Van Den Avyle and Hayward (1999), and Pine et al. (2003). Detailed discussions can be found in Williams et al. (2002) and Hayes et al. (2007).

Estimating population density is time-consuming and expensive; consequently, it is often excluded from standard fish sampling protocols in lakes, ponds, and larger rivers. Multiple-pass depletion methods are commonly used in small streams for estimating density, but the use of these methods varies depending on the species of interest and the sampling goals. Instead, CPUE is often used as an index of density. Catch per unit effort is simply the number of fish sampled per unit of effort and is assumed to be directly proportional to density (Hubert and Fabrizio 2007). Although the concept and estimation of CPUE appears simple, estimating CPUE has numerous assumptions that must be considered. Using CPUE as a measure of fish abundance assumes that changes in CPUE reflect a proportional change in abundance, which is often not the case. Additionally, a common question is whether or not different gears can be combined to provide an estimate of CPUE (e.g., fyke nets and gill nets). Given differences in catchability (i.e., probability of catching an individual fish in one unit of effort; Van Den Avyle and Hayward 1999) associated with different gear-species combinations, sampling gears should remain separate when estimating CPUE. We suggest that managers choose the gear most appropriate for the target species and system (Chapters 2–10).

Another common question focuses on the computation of CPUE. Two methods can be used. First, the total number of fish can be divided by the total amount of effort. Second, CPUE can be estimated for each sampling unit (e.g., net set, electrofishing transect) and then averaged. If effort is equal among samples, then the estimate of the mean will be equal for both methods. However, differences arise when effort varies among samples. Suppose that three electrofishing runs are conducted and the catches are as shown in Table 11.1. In both scenarios, the total number of fish is 2,310 and total effort is 90 min. If CPUE is estimated using totals, then CPUE is 25.7 fish per minute in both scenarios. However, if CPUE is estimated for each run and averaged, different results are obtained,

Table 11.1 Two scenarios of estimating catch per unit effort (CPUE = catch per unit effort [number of fish per minute]) from three hypothetical electrofishing runs.

| Run | Effort (min) | Scenario A | | Scenario B | |
|-----|--------------|----------------|-------|----------------|------|
| | | Number of fish | CPUE | Number of fish | CPUE |
| 1 | 15 | 2,250 | 150.0 | 45 | 3.0 |
| 2 | 30 | 15 | 0.5 | 15 | 0.5 |
| 3 | 45 | 45 | 1.0 | 2,250 | 50.0 |

with a mean CPUE of 50.5 fish per minute for scenario A and 17.8 fish per minute for scenario B. Differences between the two techniques are particularly dramatic when no fish are sampled in a series of nets or electrofishing runs. The question then becomes which estimate is correct. Using total catch and total effort, the estimate is simply overall CPUE, whereas the other procedure yields a mean CPUE and associated variance. Biologists should use mean CPUE. Such estimates are necessary when inferential statistical analyses or sample-size estimation procedures are conducted. Hubert and Fabrizio (2007) provided a detailed discussion of issues, such as basic assumptions and statistical techniques associated with CPUE data. In addition, biologists should follow standard statistical texts (e.g., Scheaffer et al. 1996) when estimating parameters (e.g., means, variance) from samples collected using designs other than simple random sampling (e.g., stratified or systematic sampling designs; Hansen et al. 2007).

11.2.2 Length-Frequency Distributions

Length-frequency distributions (i.e., plots of frequency, numbers, or catch rates as a function of fish length; Brown and Austen 1996; Neumann and Allen 2007) are used in most fish population evaluations. Communicating the results of a length-frequency distribution to other managers is difficult unless some type of summarization procedure is used. Techniques have been developed to summarize length-frequency distributions (e.g., Bonar 2002) and the most common technique is to use a length-structure index, specifically proportional size distribution (PSD; formerly proportional stock density, Guy et al. 2007) and PSDs for various length categories (formerly relative stock densities; see Willis et al. 1993; Anderson and Neumann 1996; and Neumann and Allen 2007 for additional information on length-structure indices). Length-structure indices quantify length-frequency histograms and promote quick and efficient communication among fisheries professionals (Willis et al. 1993; Anderson and Neumann 1996). Length-structure indices are also used as targets for management activities or as a measure of change in fish populations resulting from anthropogenic disturbance or management activities (e.g., Bauer 2002; Michaletz 2006).

Similar to other fisheries data (e.g., CPUE), length-frequency and length-structure data have a number of analytical issues. Detailed descriptions of those issues and statistical methods associated with analyzing length-frequency data are provided in Anderson and Neumann (1996), and Neumann and Allen (2007).

11.2.3 Body Condition

Body condition can be measured using numerous techniques, including physiological measures of lipid content or stress hormones (e.g., Wedemeyer et al. 1990; Simpkins et al. 2003) and indices based on length–weight relationships (e.g., Blackwell et al. 2000). Physiological analyses are costly and require sophisticated equipment and high levels of technical expertise. However, for most routine fisheries work, length–weight data provide readily available sources of information on fish condition (Anderson and Neumann 1996; Blackwell et al. 2000). Although a number of techniques have been proposed, relative condition (K_n) and relative weight (W_r) are the most common indices. Relative condition and relative weight are obtained using species-specific length–weight relationships that include fish from a small region (i.e., K_n) or from across the entire distribution of the species (i.e., W_r). Whenever possible, W_r should be used to estimate condition. Although W_r is commonly used in the fisheries profession, its utility and the derivation of standard weight equations, as well as alternative methodologies, have been the focus of extensive discussion (e.g., Cone 1989; Murphy and Willis 1992; Gerow et al. 2005). Despite the issues surrounding condition indices, they provide standardized methods of summarizing length–weight data that are efficient and can be easily communicated among biologists in different geographic locations. Statistical analysis of condition indices is discussed in Brenden et al. (2003) and Pope and Kruse (2007).

11.2.4 Diversity Indices

Although population-level measures will remain a focus of most fisheries professionals, an increased focus on the conservation of native species and their associated fish assemblages is emerging (e.g., Clarkson et al. 2005). A number of summarization and analysis techniques are available, and in-depth discussions are provided in many ecology texts. Summarization techniques have focused on species richness and diversity indices (e.g., Wolda 1981; Magurran 2004). Statistical analyses of fish assemblages are often conducted using multivariate analyses, such as cluster analysis or ordination techniques (e.g., Jongman et al. 1995). Additional information on the summarization and analysis of assemblage data can be found in Wolda (1981), Jongman et al. (1995), and Magurran (2004). Crowder (1990) and Kwak and Peterson (2007) provided excellent discussions of issues and statistical techniques to quantify community metrics in fish assemblages.

11.3 SAMPLE-SIZE ESTIMATION

Estimating the number of samples required to detect changes in fish population characteristics is important for ensuring efficient use of time and financial resources. For example, fisheries biologists often need to determine whether a management action has resulted in a change in a fish population between two points in time. Information on the number of samples required before and after the management action provides insight on whether or not a change can be detected at some level of sampling effort. Sample-size esti-

mates also provide guidance when establishing standardized sampling protocols and should be considered along with other aspects of study design during the planning process (e.g., fixed versus random samples, simple random versus stratified random sampling; Hansen et al. 2007). All sample-size estimation procedures utilize parameter estimates (e.g., variance, mean) specific to a particular system. If existing data are available, those data should be used to estimate the number of samples required to meet management objectives. If data are unavailable, samples should be collected using standard methods to obtain the necessary parameter estimates, and resulting estimates can then be used to calculate required sample sizes. The utility of sample-size estimates is often questioned because estimates are often beyond the sampling capabilities of most biologists. However, sample-size estimates are useful because they provide biologists with realistic expectations of what differences may be detected with a reasonable amount of effort. For example, if sampling variability is high, fishery managers may only be able to detect a $\pm 50\%$ difference in population metrics.

A number of sample-size estimators are available to fisheries biologists, but we discuss those most commonly used. These estimators can be categorized based on their level of complexity and ease of use. In general, more complex estimators are more difficult to use but provide more realistic estimates of needed sample sizes.

11.3.1 Rules of Thumb

The following sections provide fisheries scientists with different techniques for estimating needed sample sizes, all of which require some level of mathematical computation. While these estimators are useful, summary tables and rules of thumb are commonly used to guide sampling efforts. Summary tables and rules of thumb are particularly useful when estimators are highly complex or when research has not yet established clear guidelines for sample-size requirements. The following discussion focuses on length-structure, body-condition, and growth data. Although these are presented as general guidelines, we believe that scientists should strive to use more complex (and realistic) sample-size estimators and to adhere to the tenet that "more data are always better."

11.3.1.1 Length-structure indices

Describing the length-frequency distribution of samples from fish populations is a primary goal of standardized sampling. Anderson and Neumann (1996) suggested that at least 100 stock-length fish should be measured to describe length-frequency distributions. However, these guidelines and others (e.g., Gilliland 1987) are largely based on professional judgment rather than statistical estimation. Gustafson (1988) provided a series of tables that can be used to estimate sample-size requirements. Specifically, the tables provide 80% or 95% confidence intervals (CIs) for length-structure index estimates (e.g., PSD) at various sample sizes. If an estimate of PSD (or PSD-P, PSD-M, etc.) is obtained, biologists can then use the table to gain insight on the number of stock-length fish required to obtain some level of confidence. Alternatively, Gustafson (1988) provided an equation that biologists can use to develop their own tables for dif-

ferent levels of statistical confidence. The tables provided by Gustafson (1988) focus on different levels of precision (i.e., number of samples and associated CIs for a sample) and do not focus on the number of samples required to detect a statistical difference between PSDs in different systems or through time in a single system. In response to these limitations, Miranda (1993) developed equations and tables that allow biologists to estimate the number of samples necessary to detect differences in length-structure indices. More recent research using resampling techniques (see Section 11.3.4) have provided additional insight on required sample sizes (Vokoun et al. 2001; Miranda 2007). Results of these studies suggest that required sample sizes associated with examining length-structure indices and length-frequency distributions may be as low as 75 fish and exceed 400 fish depending on the goals of data analysis and the species of interest. Thus, tables in Gustafson (1988) and Miranda (1993), or the numbers provided in Vokoun et al. (2001) and Miranda (2007), can be used depending on the goals and objectives of sampling. For the most basic length-structure index, PSD, a minimum sample size of 125 would be appropriate in most instances. However, we recommend measuring at least 400 fish to maximize the flexibility of what can be accomplished with the data, including comparison of more detailed size-structure descriptors like length frequency.

11.3.1.2 Body condition indices

Although a number of indices are available for describing the body condition of fishes, we focus this discussion on W_r due to its popularity and widespread use by biologists within natural resource agencies (Blackwell et al. 2000). The first sample size recommendation for describing W_r was made by Wege and Anderson (1978) who suggested that, in lakes with densities greater than 50 largemouth bass/ha, a sample of 10–20 largemouth bass was required to estimate W_r within three units of true W_r with 90% confidence. They further recommended that a sample of more than 20 largemouth bass was required in lakes with less than 50 largemouth bass/ha. More recently, Brown and Murphy (1991) provided an equation to estimate the 95% CI for mean W_r values that can then be used to construct tables similar to those provided by Gustafson (1988) and Miranda (1993) for length-structure indices. Unfortunately, few studies have examined sample-size requirements for mean W_r estimates, likely due to statistical issues associated with the calculation and distribution of W_r (Brenden et al. 2003; but see Hyatt and Hubert 2001). Scientists should measure and weigh at least 100 fish, greater than the minimum length for application of standard weight equations for a given species. If data are available to provide adequate estimates of variance, empirical data should be used to construct tables of CIs using equations provided by Brown and Murphy (1991) to help guide sampling.

11.3.1.3 Growth

Similar to length distributions and body condition, information on age and growth can help answer a number of management questions. Unfortunately, little guidance is available on the number of samples required for the estimation of population statistics com-

monly derived from age and growth information (e.g., mortality rates, growth models, mean length at age, mean back-calculated lengths). Until such studies are conducted, a common rule of thumb is to subsample at least five fish per centimeter-length-group for management purposes and at least 10 fish per centimeter-length-group for research projects. When data are used to evaluate population dynamics associated with age-structure data, the subsample should be extrapolated to the entire sample using an age-length key (DeVries and Frie 1996; Bettoli and Miranda 2001).

11.3.2 "Simple" Estimators

"Simple" sample-size estimators are those that do not require extensive knowledge or expertise in using statistical or modeling software, but require some mathematical calculations. Sample-size requirements can be estimated using a spreadsheet or pocket calculator. Although these estimators are computationally simple, sample-size estimates are limited to simple sample-size questions and may not provide realistic sample-size estimates, at least relative to more sophisticated techniques. As such, these estimates should be used with the understanding that they are rough estimates of sample-size requirements. One of the most basic sample-size estimators was provided by Merritt et al. (1984):

$$n = \frac{t^2 \times s^2}{(\bar{x} \times d)^2},$$

where n is the estimated sample size, t is the t -distribution deviate, s^2 is the sample variance, \bar{x} is the sample mean, and d is the desired difference expressed as a proportion (e.g., 25% change in the mean or 0.25). The t -distribution deviate is obtained by consulting t -tables in statistical textbooks and is chosen based on the degrees of freedom ($n - 1$). Because the number of samples is unknown, a value of 1.96 ($\alpha = 0.05$ or 95% probability) is commonly used in the equation. However, this practice may underestimate the required number of samples. Although an iterative process (i.e., solve for n with some t -value, refit with another t -value based on the estimate of n) is often used to address this issue (see Section 11.3.3), using a t -value of 1.96 is probably adequate for those requiring only a rough approximation of sample sizes. Using this equation, an estimate of the variance is obtained from a sample (e.g., previous sampling) and sample size is estimated to ensure that a difference (d) can be detected with some level of confidence (i.e., by manipulating the t -distribution deviate). This equation is analogous to the equation presented by Snedecor and Cochran (1989):

$$n = \frac{t^2 \times CV^2}{d^2},$$

where n , t , and d are as above and CV is the coefficient of variation (SD/mean). For examples of the application of these equations, see Merritt et al. (1984) or Peterson and Rabeni (1995).

Fisheries professionals are often interested in maintaining a high level of statistical power (i.e., $1 - \beta$; Peterman 1990; McAllister and Peterman 1992). A common formula

used to estimate sample size with a consideration for statistical power is provided by Snedecor and Cochran (1989):

$$n = \frac{2 \times s^2 \times (z_\alpha + z_\beta)^2}{\delta^2},$$

where n and s are as above, δ is the desired difference expressed as an actual value (difference between mean 1 and mean 2), and z_α and z_β are values from the standard normal distribution that account for the chance of false significance (z_α) and the chance of falsely failing to reject the null hypothesis (z_β). For example, with $\alpha = 0.05$, $z_\alpha = 1.96$ for a two-tailed test and $z_\alpha = 1.65$ for a one-tailed test, if the desired statistical power ($1 - \beta$) is 0.80, then $z_\beta = 0.84$.

11.3.3 Estimators Incorporating Statistical Power

The following estimators have been used extensively in fisheries research and management and are similar to the estimator provided by Snedecor and Cochran (1989), discussed in the previous section. These estimators are best implemented using spreadsheet programs or preprogrammed statistical software. A common estimator is

$$n = \frac{2 \times s^2 \times (t_\alpha + t_\beta)^2}{\delta^2},$$

where n is the estimated sample size, t_α is the t -distribution deviate for a two-tailed test at a given α , t_β is the t -distribution deviate for the given level of statistical power, s^2 is the sample variance, and δ is as above (Zar 1984). One of the complications of using this equation is that estimates of t_α and t_β are required, and estimating t_α and t_β require an estimate of the sample size (as previously discussed). Consequently, an iterative process is used to estimate required sample sizes. Online statistical software provided by Simple Interactive Statistical Analysis (SISA; www.quantitativeskills.com/sisa) makes computation of this equation relatively straightforward and has been used to answer fisheries sample-size questions (e.g., Tate et al. 2003; Paukert 2004). One of the benefits of using a spreadsheet or preprogrammed software package is that different levels of statistical power and varying differences can be rapidly simulated to provide a distribution of sample sizes. For instance, Tate et al. (2003) estimated the number of samples required to detect a 25% and 50% change in CPUE of largemouth bass at four levels of statistical power (i.e., 60%, 70%, 80%, and 90%). When such an analysis is conducted, the resulting distribution of sample-size estimates not only provides insight on the number of samples required to detect a difference at a given level of statistical power, but the analysis also provides insight on what differences and levels of statistical power might be expected with different sample sizes. Additional examples of this analysis framework are provided in Simonson et al. (1994), Lester et al. (1996), Kreuger et al. (1998), and Paukert (2004).

All of the sample-size estimators discussed thus far are appropriate for estimating sample-size requirements for most fisheries management questions. Most estimators assume

that sample sizes are equal among time periods or populations, variances are equal among populations or time periods, samples are collected completely at random (i.e., samples are completely independent), and data are normally distributed. Numerous scenarios can be envisioned where variance changes among time periods, sample size varies among time periods, or paired or fixed sites are sampled through time. In addition, many fisheries data are not normally distributed (e.g., Hubert and Fabrizio 2007). Gerow (2007) recently developed a sample-size estimator to help address some of these concerns:

$$n = \frac{C(t_{\alpha} + t_{\beta})^2 (\lambda_0^2 + \lambda_1^2 - 2r\lambda_0\lambda_1)}{(\lambda_0 - \lambda_1)^2},$$

where n is the estimated sample size, C is a variance inflation constant (standard deviation of the mean/mean) that accounts for a non-Poisson distribution, t_{α} is the t -distribution deviate for a one-tailed test given α , t_{β} is the t -distribution deviate for a one-tailed test for the given level of statistical power, λ_0 is the mean at time 0, λ_1 is the mean at time 1, and r is the correlation among samples between two time periods ($r = 0$ if samples are not paired). The equation allows estimation of the number of samples (i.e., equal or unequal samples sizes) necessary for detecting an increase or decrease in a mean value and allows for the incorporation of paired- or fixed-site sampling designs (see Gerow 2007 for a detailed description and Quist et al. 2006 for an application). Although the estimator is relatively complex, a spreadsheet that allows for simple application of the equation is available online (www.statsalive.com).

Many standard sampling protocols use a stratified sampling design to allocate effort among different strata (e.g., littoral and pelagic habitat in lakes; pools, riffles, and runs in streams). Although sample-size estimators for such designs may be highly complex, a relatively simple estimator is provided by Scheaffer et al. (1996):

$$n = \frac{\sum \left(\frac{N_i^2 \times \sigma_i^2}{w_i} \right)}{(N^2 \times D) + \sum (N_i \times \sigma_i^2)}$$

where n is the estimated number of samples (across all strata), N_i is the number of samples in stratum i , σ_i^2 is the estimated variance from stratum i , w_i is the allocation fraction (i.e., allocation fraction of the final design), N is the total sample size, and D is equal to $B^2/4$ (B is the desired number of measurements units from the sample mean; see Box 11.1 for an example). Alternative forms of this sample-size estimator are available in Scheaffer et al. (1996), as are various methods for allocating samples (e.g., proportional allocation, allocation based on cost, allocation based on variance).

11.3.4 Resampling Methods

Thus far, we have focused on sample-size estimators used to detect a change in the mean (e.g., mean CPUE) between time periods or populations. While such questions will

Box 11.1 Example of estimating sample size using a “moderately complex” estimator and allocating samples among strata.

Suppose that a biologist has pre-existing sampling data for a fish species sampled from three different strata in a river system and wants to estimate the number of samples required to estimate catch per unit effort (CPUE) within 5% of the mean CPUE. A total of 50 samples (i.e., $N_1 = 10$, $N_2 = 15$, $N_3 = 25$) resulted in a mean CPUE estimate of 25 fish/net-night and strata variances as follows: $\sigma_1^2 = 25$, $\sigma_2^2 = 5$, and $\sigma_3^2 = 50$. In this example, D would be $B^2/4 \left[\frac{(0.05 \times 25)^2}{4} = 0.39 \right]$. Also, assume that the biologist wants to allocate the sampling effort proportional to the variance: stratum 1 = 31% of the samples, stratum 2 = 6%, and stratum 3 = 63%. Sample size would then be estimated as

$$n = \frac{\sum \left(\frac{N_i^2 \times \sigma_i^2}{w_i} \right)}{(N^2 \times D) + \sum (N_i \times \sigma_i^2)}$$

$$n = \frac{\left(\frac{10^2 \times 25}{0.31} \right) + \left(\frac{15^2 \times 5}{0.06} \right) + \left(\frac{25^2 \times 50}{0.63} \right)}{(50^2 \times 0.39) + [(10 \times 25) + (15 \times 5) + (25 \times 30)]}$$

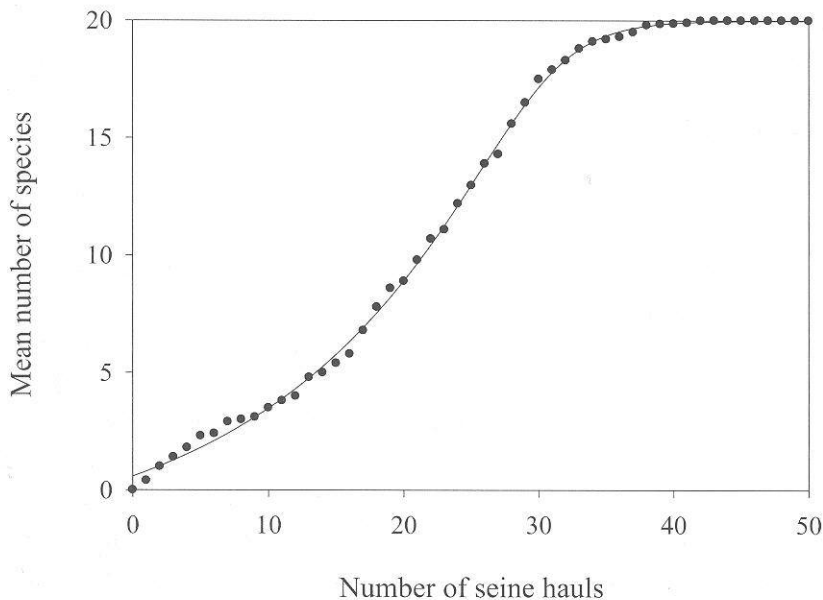
$$n = \frac{76,418}{2,550} = 30$$

Thus, 30 samples would be required and would be allocated as $N_1 = 30 \times 0.31 = 9$; $N_2 = 30 \times 0.06 = 2$; and $N_3 = 30 \times 0.63 = 19$. This example illustrates the value of preliminary sampling, which can provide investigators with an idea of the variance they will encounter and allow an appropriate sample allocation.

remain important in fisheries management and research, other sample-size questions are common. In particular, biologists are often interested in the number of samples required to capture some proportion of the species present in a system with some level of confidence (e.g., Lyons 1992; Angermeier and Smogor 1995; Walsh et al. 2002). Questions such as these do not conform to standard sample-size equations. The most common technique for estimating sample size in these situations is to use a resampling simulation (e.g., Colwell and Coddington 1994). Resampling procedures are conceptually straightforward, but require a relatively high level of programming skill. Because resampling methods are often tailored to specific sample-size questions and standard formulae are not used, an example is provided in Box 11.2 to illustrate how a resampling procedure might be used to estimate sample-size requirements. Although the example is specific to species richness, it provides a framework that can be used to answer a number of questions common to sampling fish populations (i.e., questions focused on CPUE, length structure, condition, or other parameters). MacKenzie et al. (2002) introduced the use of site occupancy models to estimate species richness, a

Box 11.2 Example of using a resampling method for estimating species richness.

Suppose a biologist is interested in estimating the number of seine hauls needed to capture all of the species present in a river segment (e.g., Lyons 1992; Patton et al. 2000; Walsh et al. 2002). Using pre-existing data (e.g., 50 seine hauls from the system), replicate random samples (e.g., 500 replicate samples or iterations) are drawn with varying numbers of seine hauls (e.g., 1, 2, 3, 4...50 seine hauls). For a sample size of one, one seine haul is randomly sampled 500 times and the metric of interest (e.g., species richness) is estimated from each iteration. For a sample size of two, two seine hauls are randomly sampled (without replacement) 500 times and species richness is estimated from each iteration (i.e., total number of species across all transects), and so forth up to 50 seine hauls. Thus, for each number of seine hauls there are 500 random samples, each with some number of collected fish species. One method for estimating the required sample size is to develop a species "accumulation curve" by plotting the cumulative number of species (e.g., minimum, mean, median, or maximum number of species across the 500 iterations) against the number of seine hauls. The resulting curve will likely reach an asymptote and provide an estimate of the required sample size. Other methods for developing species accumulation curves are provided in Kwak and Peterson (2007) and applications are provided in Lyons (1992), Angermeier and Smogor (1995), and Walsh et al. (2002).

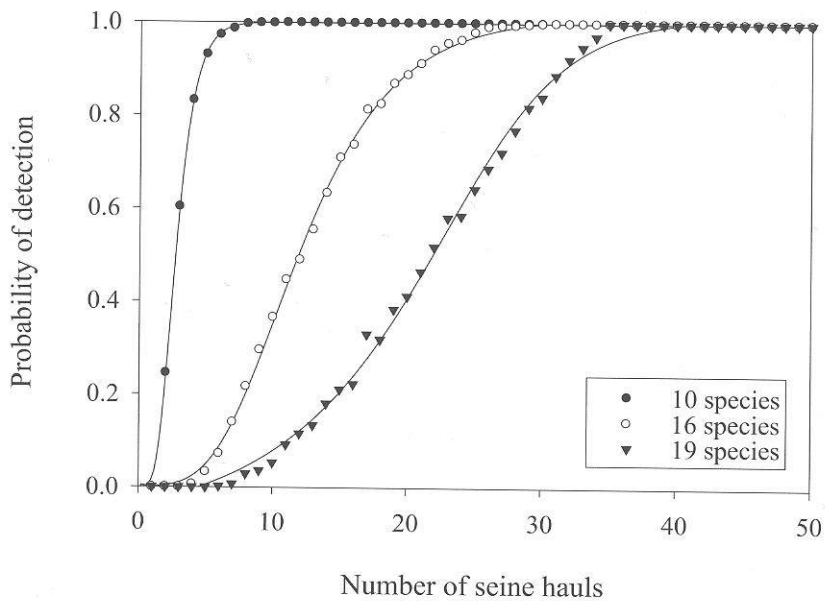


Relationship between mean species richness (i.e., mean of 500 random samples for each number of seine hauls) and the number of seine hauls. Based on this relationship, around 40 seine hauls should be sufficient to sample most of the species present.

(Box continues)

Box 11.2 (continued)

An alternative method is to estimate the probability of sampling a specified proportion of species. For example, if a total of 20 fish species is sampled in the 50 seine hauls, a biologist may be interested in the number of samples required to capture 50% (i.e., 10 species), 80% (i.e., 16 species), or 95% (i.e., 19 species) of the species present in the study area. The probability of sampling 50% of the species can be estimated for each sample size (i.e., number of seine hauls) by tallying the number of times out of 500 that 10 or more species were sampled, and then dividing that tally by 500. The probabilities can then be plotted against the number of seine hauls and a regression model (e.g., logistic regression model) can be fit to the probabilities. Fitting a regression model helps “smooth” the relationship and allows one to estimate the probability of detecting a defined percentage of the species present (e.g., how many samples are needed to have a 95% probability of sampling 50% of the species; see Bailey and Gerow 2005 for a detailed example).



Relationship between the probability of sampling a specified number of species and the number of seine hauls.

method which has been used for estimating animal abundance and has gained popularity in wildlife studies. Some software packages are available online that will conduct portions of statistical simulations. One example of such software is EstimateS that can be used to derive species accumulation curves (<http://viceroy.eeb.uconn.edu/EstimateS>; see Walsh et al. 2002 for an application). Another straightforward package for conducting

resampling methods is PopTools (www.cse.csiro.au/poptools), which is a free add-in to Microsoft (MS) Excel (Microsoft Corporation, Redmond, Washington).

11.3.5 Monitoring Spatial and Temporal Trends

The purpose of standardized sampling is to monitor whether the magnitude of some quantity (e.g., fish density, diversity, body condition, mortality) is increasing or decreasing through time or space. While examining trends is critical for effective management and conservation, changes in a quantity over short temporal and spatial scales can lead to erroneous conclusions because these changes may be due to natural variation rather than representation of a true trend. Consequently, investigating changes between a few sampling events often requires more samples than for investigating a trend over long time periods (e.g., Larsen et al. 2004). Sections 11.3.1 through 11.3.4 largely focus on sample-size requirements associated with detecting changes in two points in time or space. Sample-size questions focused on long-term trends are somewhat different than those related to a comparing two values, but progress has been made in the development of monitoring designs that allow scientists to maximize statistical power by allocating samples over time and space based on the variance structure of initial samples. A detailed description of these techniques is beyond the scope of this chapter but can be found in the published literature (e.g., Gerrodette 1987; Gibbs et al. 1998; Urquhart et al. 1998; Larsen et al. 2004).

A number of software programs have been developed to assist scientists with sample-size calculations. These programs generally focus on allocating samples in time and space while maintaining a specified level of statistical power. A useful characteristic of these programs is that they allow the user to evaluate the benefits and limitations of different designs by simulating a variety of sampling scenarios. Popular programs include MONITOR (available from the Illinois Natural History Survey, Clearinghouse for Ecology Software, Champaign; <http://nhsbig.inhs.uiuc.edu/wes/populations.html>) and TRENDS (available from NOAA Fisheries Service, Southwest Fisheries Science Center; <http://swfsc.noaa.gov>).

11.4 DATA MANAGEMENT

According to the Data Management Association, data management refers to the “development and execution of architectures, policies, and procedures that properly manage the full data lifecycle of an enterprise.” For our purposes, data management can be defined as the mechanism by which data are handled, which involves the acquisition, storage, and retrieval of information for later use by an individual or group of individuals. As we develop standardized sampling methods, we are essentially predicting what assessments will be important in the future, and that, in turn, influences which data and how data are collected (Schnute and Richards 1994). The same logic applies to data management, in that how data are stored places constraints on how data can be retrieved and used in the future.

A database is simply a collection of related data or a data repository. Fisheries scientists have been using databases since the first datum was recorded on a datasheet and placed in a filing cabinet. Currently, data are stored in a variety of locations and formats, from pa-

per records that are only accessible at the local or regional level to electronic data that are accessible by a user across a large geographic area. Most fisheries sampling data collected over long temporal and large spatial scales are stored in relational databases. Although many biologists continue to store data in flat file formats, (i.e., where data are stored in single, nonrelational tables or layers), such as MS Excel, this format is disadvantageous for large data sets because file space is quickly consumed by redundant data. Data stored in flat files can also be cumbersome to update, particularly if data are distributed in multiple worksheets. Relational databases were first developed in 1970 and are simply data stored in multiple tables that are linked or related by their content (Codd 1970). Therefore, a large database stored in a single table in MS Excel can be separated into several related tables that reduce data redundancy, data entry errors, and time spent entering data while, at the same time, improving the ease of maintaining the database and increasing the efficiency of data retrieval. For instance, if time or effort was incorrectly entered into the database depicted in Figure 11.1, many rows of data would need to be corrected in the flat file structure (table on left) versus one line in the relational database structure (top right table). The probability of making an error is reduced and database maintenance is more efficient if databases are organized in a relational structure.

In 1998, the first National Freshwater Fisheries Database Summit was held in San Diego, California. Representatives from state, federal, and private organizations met to discuss issues related to the management and sharing of fisheries information (Loftus 1998). Additional summits were organized in 2002 (Loftus and Faibisch 2003) and 2006. Some of the data issues presented at these database summits are addressed here, particularly those related to data standards and sharing.

11.4.1 Data Standards

Although some data are unique to a region or state, much of the information collected on fisheries resources is common among agencies and geographic locations. These include date and time notations, taxonomic information, geographic data, weather data (e.g., wind speed, temperature), habitat data (e.g., aquatic vegetation type, substrate composition), and biological data (e.g., fish length and weight). To facilitate sharing of information and standardization of data management procedures, agencies must adopt data standards (Loftus 1998). Many standards have been created and adopted by federal agencies, including the U.S. Fish and Wildlife Service (www.fws.gov/stand) and the U.S. Environmental Protection Agency (www.epa.gov/edr). The creation of common species, habitat, and geographic "look-up tables" and standard reporting forms increases the ability of integrating and sharing information across jurisdictional boundaries (Beard et al. 1998). Furthermore, the adoption of other data standards, including defined measurement units and sampling techniques, can facilitate data sharing (Loftus 1998). Due to their broad scope, many federal data standards (Table 11.2) have been adopted by state and local agencies. Other data standards, such as measurement units, precision standards, and equipment codes, have only been standardized at the state level. Even if standards

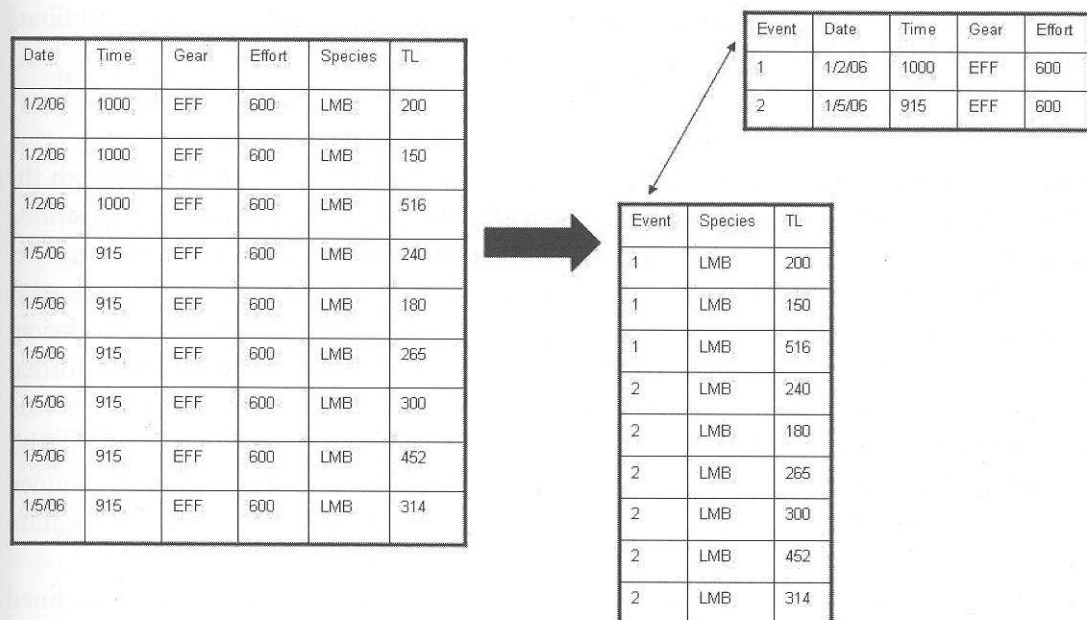


Figure 11.1 Example of data stored in a flat file format (e.g., MS Excel spreadsheet; left table) and the same data stored in a relational format (e.g., MS Access; right tables).

Table 11.2 Common data standards established by federal agencies that, if used by all levels, could facilitate the sharing of fisheries information among agencies.

| Data type | Standard code | Source | Web site |
|---|--|---|---|
| Taxonomic identifiers | Taxonomic serial numbers (TSN) | Integrated Taxonomic Information System | www.itis.gov |
| Geographic identifiers | Geographic Names Information System (GNIS) | U.S. Board on Geographic Names | geonames.usgs.gov |
| | Reach codes from EPA's Reach File Version 3 (RF3) | National Hydrography Dataset (NHD) | nhd.usgs.gov |
| Wetland and deepwater habitat classifications | National Wetlands Inventory (NWI) code lists | U.S. Fish and Wildlife Service | www.fws.gov/nwi |
| Drainage basins | Hydrologic unit code (HUC) | U.S. Geological Survey | water.usgs.gov/GIS/huc.html |
| National vegetation classifications | National Vegetation and Information Standard (NVCS) code lists | Federal Geographic Data Committee | http://biology.usgs.gov/npsveg/nvcs.html |

are specific to a state agency, these standards should be clearly documented to facilitate potential data sharing in the future.

11.4.2 Data Sharing

Most fisheries management issues are not restricted to a single system or species. With the onset of ecosystem-based management and activities that involve cooperative efforts among local, state, and federal agencies, it is important that databases are constructed in a manner that facilitates sharing of information (Hamm 1993; Schnute and Richards 1994; Beard et al. 1998). Data stored on individual hard drives or in filing cabinets are no longer sufficient to meet the needs of the public. For example, the Multistate Aquatic Resources Information System (MARIS) was developed in 1994 and represents a collaborative effort between multiple state and federal natural resource agencies (Beard et al. 1998; see Box 11.3). This project is focused on providing access of state-collected freshwater fisheries data obtained using various methodologies and stored in multiple formats via a single online location (Beard et al. 1998).

Most fisheries data collected by state agencies are available to the public as defined by the Freedom of Information Act. However, many professionals are concerned that data can be misused or incorrectly analyzed and then used to argue inappropriately for opposing views. Although states are obligated to make these data available to the public, introducing a time lag in the data availability, creating readily accessible summary reports, or providing a mechanism by which raw data are requested can be used to limit the release of sensitive data and monitor where information is distributed. For example, the Western Pacific Fishery Information Network (WPACFIN) warehouses data collected by multiple fishery agencies of Samoa, Hawaii, Guam, and the Northern Mariana Islands (Hamm 1993). Requests for confidential or raw data from WPACFIN are submitted to the donor agency and must be approved by that agency before being distributed. However, summary and nonconfidential data are available without restriction (Hamm 1993). Undoubtedly, some data will be used incorrectly, but fisheries professionals should be open to different interpretations of their data, including interpretations by members of the public (Schnute and Richards 1994). Furthermore, providing information in a widely used interface, such as through online sources, may enhance communication with the public, and can be used as an educational tool and as a mechanism for increasing support of management programs (Loftus 1998).

A primary component associated with sharing data is documentation and availability of metadata. Metadata are data about the data or information describing a particular data set, and they can include information such as location, species, methodology, precision, accuracy, and measured variables included in the database (Loftus 1998). Metadata can be used to determine whether data are appropriate to meet a particular objective and to identify potential constraints associated with a data set (Loftus 1998). Because the individual collecting and entering data is rarely the same individual who conducts the data analysis, proper and consistent documentation of metadata is critical (Gray et al. 2005). Furthermore, metadata

Box 11.3 The Multistate Aquatic Resource Information System (MARIS): sharing data across agency boundaries (Andrew J. Loftus and T. Douglas Beard, Jr.).

The Multistate Aquatic Resources Information System (MARIS) was initiated in the 1990s to facilitate data sharing across political boundaries. The initial goals of MARIS were to reduce the duplication of data collection efforts, promote sharing of technical expertise among member agencies, and better utilize the millions of dollars worth of existing information annually collected by states (Beard et al. 1998).

The focus of MARIS is to allow agencies to share information related to the status and trends of aquatic resources. Utilizing a single Web interface, users access data from multiple fisheries management jurisdictions (currently 10 states) via a common query structure. MARIS is designed to provide select elements that the contributing agencies determine will serve the intended goal of the information system (currently, status and trends information). The information provided in MARIS is not raw data, but rather first- or second-order synthesis of data collected as part of fisheries sampling events. MARIS is designed to serve data at the water body level but is flexible enough to allow analysis of such data at larger geographic scales. Quantitative data elements reflecting status and trends currently in the system include population estimates and catch per unit effort.

MARIS is designed as a distributed transactional information system. Agencies that contribute data to the system house the subset of data that they will provide to MARIS on their own internal agencies' servers. The MARIS server periodically queries each agency server to obtain the most current data available and creates a centralized cache of data for increased efficiency and accessibility. Thus, agencies that collect the data retain ultimate control over them, updates to the data can be conducted automatically by the agency, and the most current available data in the state can be made available to MARIS. This approach also allows states to maintain control over the availability of sensitive data such as the location of threatened and endangered species.

An important feature shaping MARIS development is the adoption of existing data standards to facilitate data sharing. While many states have developed their own internal data standards, MARIS does not require them to change those standards. A series of lookup tables will convert their data to the MARIS standard system. To the extent possible, MARIS is using existing standards, such as the integrated Taxonomic Information System for species coding.

MARIS is an example of a current system that offers a mechanism through which nearly any type of fisheries information can be shared. Fisheries professionals who desire to exchange data between jurisdictions on a continuing basis should strongly consider working with an existing effort, such as MARIS, to meet their information sharing needs. The ability to integrate data across borders and to use these data for multiple purposes increases the value of every data point collected.

allow agencies to inventory their data (e.g., by location) and provide guidance on the proper interpretation of data by other parties (Loftus 1998).

11.4.3 Database Platform

Ultimately, the primary focus of any fisheries database is the end user. Depending on the nature and scope of the data, desktop applications such as MS Access can be used, but in more complex situations (e.g., when multiple users must access the database from remote locations), a full enterprise product such as MS SQL Server or Oracle (Oracle Corporation, Redwood Shores, California) is more appropriate. In most cases, the end user is unaware of the database structure or program used; rather, they just need it to work for their needs. Therefore, the database designer and manager are responsible for ensuring that the database structure and application optimize performance while catering to the needs of end users. Unfortunately, database personnel often lack the biological expertise or background information about the data to make decisions on their own. Creating and maintaining an open channel of communication with end users, or persons familiar with the data and the needs of end users, is vital to the success of any database because objectives can be clearly identified and pertinent applications can be developed. For instance, questions such as what kinds of data are or will be collected, how the data will be used, who will need access to the data, what kinds of reports will need to be generated, and what queries are most useful are all important for ensuring that the database meets the needs of end users. In some situations, appointing a liaison between the database developers and end users may be necessary. The liaison would be responsible for reviewing the database during its development and ensuring that important questions are adequately addressed so that the needs of the end users are met.

In addition to sound database design, the structure and layout of the user interface should also be considered because the ease at which data can be entered and retrieved will have a large impact on the future use of the database. If the interface is cumbersome, the database will not be used and biologists will return to the "old way of doing things." For example, when entering fish lengths and weights from a standard sampling survey, biologists would rather not type or select from a drop-down menu the species name or code for each measurement (Figure 11.2; top panel). Instead, a more time efficient solution is to create a single drop-down menu for a species, under which multiple length and weight records can be entered (Figure 11.2; bottom panel). The user interface can further be optimized by standardizing the layout, such as creating standard navigation tools, presentation of prompts and forms, and output reports.

After the database and associated applications have been designed, database personnel must maintain and improve the database based on its performance and feedback from end users. As part of this process, quality-control procedures and processes should be established to maintain the integrity of the data. If water temperature was entered as 83°C for a natural Florida lake, a quality-control procedure (e.g., an acceptable range of 0–40 for the temperature field) could alert the user that a data-entry error occurred. Many

| Number of Fish | Species | Size Class | Total Length | Total Weight |
|----------------|---|------------|--------------|--------------|
| 1 | LMB - largemouth bass - Micropterus salmoides | | | |
| 1 | LMB - largemouth bass - Micropterus salmoides | | | |
| 1 | LMB - largemouth bass - Micropterus salmoides | | | |
| 1 | LMB - largemouth bass - Micropterus salmoides | | | |
| 1 | LMB - largemouth bass - Micropterus salmoides | | | |
| 1 | LMB - largemouth bass - Micropterus salmoides | | | |
| 1 | LMB - largemouth bass - Micropterus salmoides | | | |



Select Species

LMB - largemouth bass - Micropterus salmoides

| Size Class | Total Count | Total of Batch Weights (g) | Total Weight (g) |
|------------|-------------|----------------------------|------------------|
| 1 mm | 7 | | 0 |
| 1 cm | 0 | | 0 |
| 2 cm | 0 | | 0 |
| Combined | 7 | | 0 |

| 1mm Group | | 1cm Group | | 2cm Group | |
|----------------|------------|--------------|------------------|---|---|
| Number of Fish | Size Class | Total Length | Total Weight (g) | <input checked="" type="checkbox"/> Tag # | <input checked="" type="checkbox"/> Sex |
| 1 | 1 mm | | | | |
| 1 | 1MM | | | | |
| 1 | 1MM | | | | |
| 1 | 1MM | | | | |
| 1 | 1MM | | | | |
| 1 | 1MM | | | | |
| 1 | 1MM | | | | |

Figure 11.2 Design of a user interface that minimizes time associated with data entry.

quality-control parameters cannot be defined by database personnel, but rather by those knowledgeable about the data. Therefore, communication between database personnel and end users is critical throughout the postdevelopment period. Open dialogue is also helpful when creating a user's manual or troubleshooting guide that addresses common database issues encountered by end users.

With this book and similar initiatives, the fisheries profession is becoming increasingly integrated. Federal and state agencies are beginning to work together to create common standards and to provide a centralized online location where data can be accessed, such as MARIS (see Box 11.3). The Internet has increased the accessibility of fisheries

data to end users, who are often located over large geographic areas. Even so, security and database accessibility are problematic, and these issues should be considered when using this type of interface. As technology and access issues are addressed, agencies will continue the process of integrating data over larger and larger areas. A single online location where fisheries data can be accessed for any state or federal agency is not only possible, but is quickly becoming reality. Such efforts should greatly increase collaboration among agencies and will enable biologists to address management problems that occur over large spatial and temporal scales.

11.5 REFERENCES

- Anderson, R. O., and R. M. Neumann. 1996. Length, weight, and associated structural indices. Pages 447–482 in B. R. Murphy and D. W. Willis, editors. *Fisheries techniques*, 2nd edition. American Fisheries Society, Bethesda, Maryland.
- Angermeier, P. L., and R. A. Smogor. 1995. Estimating number of species and relative abundances in stream-fish communities: effects of sampling effort and discontinuous spatial distributions. *Canadian Journal of Fisheries and Aquatic Sciences* 52:936–949.
- Bailey, P. E., and K. G. Gerow. 2005. Seining effort needed to estimate species richness of small littoral zone fishes: a simulation study in three Wyoming reservoirs. *North American Journal of Fisheries Management* 25:1310–1314.
- Bauer, D. L. 2002. White bass population differences in Nebraska reservoirs with gizzard shad or alewife prey bases. *North American Journal of Fisheries Management* 22:665–670.
- Beard, T. D., Jr., D. Austen, S. J. Brady, M. E. Costello, H. G. Drewes, C. H. Young-Dubovsky, C. H. Flather, T. W. Gengerke, C. Larson, A. J. Loftus, and M. J. Mac. 1998. The Multi-state Aquatic Resources Information System. *Fisheries* 23(5):14–18.
- Bettoli, P. W., and L. E. Miranda. 2001. Cautionary note about estimating mean length at age with subsampled data. *North American Journal of Fisheries Management* 21:425–428.
- Blackwell, B. G., M. L. Brown, and D. W. Willis. 2000. Relative weight (W_r) status and current use in fisheries assessment and management. *Reviews in Fisheries Science* 8:1–44.
- Bonar, S. A. 2002. Relative length frequency: a simple, visual technique to evaluate size structure in fish populations. *North American Journal of Fisheries Management* 22:1086–1094.
- Brenden, T. O., B. R. Murphy, and J. B. Birch. 2003. Statistical properties of the relative weight (W_r) index and an alternative procedure for testing W_r differences between groups. *North American Journal of Fisheries Management* 23:1136–1151.
- Brown, M. L., and B. R. Murphy. 1991. Standard weights (W_s) for striped bass, white bass, and hybrid striped bass. *North American Journal of Fisheries Management* 11:451–467.
- Brown, M. L., and D. J. Austen. 1996. Data management and statistical techniques. Pages 17–62 in B. R. Murphy and D. W. Willis, editors. *Fisheries techniques*, 2nd edition. American Fisheries Society, Bethesda, Maryland.
- Clarkson, R. W., P. C. Marsh, S. E. Stefferud, and J. A. Stefferud. 2005. Conflicts between native fish and nonnative sport fish management in the southwestern United States. *Fisheries* 30(9):20–28.
- Codd, E. F. 1970. A relational model of data for large shared data banks. *Communications of the Association for Computing Machinery* 13(6):377–387.

- Colwell, R. K., and J. A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. *Philosophic Transactions of the Royal Society of London* 345:101–118.
- Cone, R. S. 1989. The need to reconsider the use of condition indices in fishery science. *Transactions of the American Fisheries Society* 118:510–514.
- Crowder, L. B. 1990. Community ecology. Pages 609–632 in C. B. Schreck and P. B. Moyle, editors. *Methods for fish biology*. American Fisheries Society, Bethesda, Maryland.
- DeVries, D. R., and R. V. Frie. 1996. Determination of age and growth. Pages 483–512 in B. R. Murphy and D. W. Willis, editors. *American Fisheries Society*, Bethesda, Maryland.
- Everhart, W. H., A. W. Eipper, and W. D. Youngs. 1975. *Principles of fishery science*. Cornell University Press, Ithaca, New York.
- Gerrodette, T. 1987. A power analysis for detecting trends. *Ecology* 68:1364–1372.
- Gerow, K. G. 2007. Power and sample size estimation techniques for fisheries management: assessment and a new computational tool. *North American Journal of Fisheries Management* 27:397–404.
- Gerow, K. G., R. C. Anderson-Sprecher, and W. A. Hubert. 2005. A new method to compute standard-weight equations that reduces length-related bias. *North American Journal of Fisheries Management* 25:1288–1300.
- Gibbs, J. P., S. Droege, and P. Eagle. 1998. Monitoring populations of plants and animals. *BioScience* 48:935–940.
- Gilliland, E. 1987. Evaluation of Oklahoma's standardized electrofishing in calculating population structure indices. *Proceedings of the Annual Conference Southeastern Association of Fish and Wildlife Agencies* 39(1985):277–287.
- Gray, J., M. Nieto-Santisteban, A. S. Szalay, D. DeWitt, and G. Heber. 2005. Scientific data management in the coming decade. Microsoft Corporation, Microsoft Research, Technical Report MSR-TR-2005-10, Redmond, Washington.
- Gustafson, K. A. 1988. Approximating confidence intervals for indices of fish population size structure. *North American Journal of Fisheries Management* 8:139–141.
- Guy, C. S., and M. L. Brown, editors. 2007. *Analysis and interpretation of freshwater fisheries data*. American Fisheries Society, Bethesda, Maryland.
- Guy, C. S., R. M. Neumann, D. W. Willis, and R. O. Anderson. 2007. Proportional size distribution (PSD): a further refinement of population size structure index terminology. *Fisheries* 32:348.
- Hamm, D. C. 1993. The Western Pacific Fishery Information Network: a fisheries information system. *Marine Fisheries Review* 55:102–108.
- Hansen, M. J., T. D. Beard, Jr., and D. B. Hayes. 2007. Sampling and experimental design. Pages 51–120 in C. S. Guy and M. L. Brown, editors. *Analysis and interpretation of freshwater fisheries data*. American Fisheries Society, Bethesda, Maryland.
- Hayes, D. B., J. R. Bence, T. J. Kwak, and B. E. Thompson. 2007. Abundance, biomass, and production. Pages 327–374 in C. S. Guy and M. L. Brown, editors. *Analysis and interpretation of freshwater fisheries data*. American Fisheries Society, Bethesda, Maryland.
- Hubert, W. A., and M. C. Fabrizio. 2007. Relative abundance and catch per unit effort. Pages 279–325 in C. S. Guy and M. L. Brown, editors. *Analysis and interpretation of freshwater fisheries data*. American Fisheries Society, Bethesda, Maryland.
- Hyatt, M. W., and W. A. Hubert. 2001. Statistical properties of relative weight distributions of four salmonid species and their sampling implications. *North American Journal of Fisheries Management* 21:666–670.

- Jongman, R. H. G., C. J. F. ter Braak, and O. F. R. van Tongeren. 1995. Data analysis in community and landscape ecology. Cambridge University Press, Cambridge, UK.
- Kreuger, K. L., W. A. Hubert, and R. M. Pierce. 1998. Tandem-set fyke nets for sampling benthic fishes in lakes. *North American Journal of Fisheries Management* 18:154–160.
- Kwak, T. J., and J. T. Peterson. 2007. Community indices, parameters, and comparisons. Pages 677–763 in C. S. Guy and M. L. Brown, editors. Analysis and interpretation of freshwater fisheries data. American Fisheries Society, Bethesda, Maryland.
- Larsen, D. P., P. R. Kaufmann, T. M. Kincaid, and N. S. Urquhart. 2004. Detecting persistent change in the habitat of salmon-bearing streams in the Pacific Northwest. *Canadian Journal of Fisheries and Aquatic Sciences* 61:283–291.
- Lester, N. P., W. I. Dunlop, and C. C. Willox. 1996. Detecting changes in the nearshore fish community. *Canadian Journal of Fisheries and Aquatic Sciences* 53:391–402.
- Loftus, A. J. 1998. Proceedings of the national freshwater fisheries database summit. American Fisheries Society, Computer User Section, Bethesda, Maryland.
- Loftus, A. J., and J. Faibisch. 2003. National fish and wildlife database summit final report-recommendations and findings. Available: www.fishdata.org/summit/2002%20Fish%20&c%20Wildlife%20Data%20Meeting.pdf (January 2009).
- Lyons, J. 1992. The length of stream to sample with a towed electrofishing unit when fish species richness is estimated. *North American Journal of Fisheries Management* 12:198–203.
- MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83:2248–2255.
- Magurran, A. E. 2004. Measuring biological diversity. Blackwell Publishing, Oxford, UK.
- McAllister, M. K., and R. M. Peterman. 1992. Experimental design in the management of fisheries: a review. *North American Journal of Fisheries Management* 12:1–18.
- Merritt, R. W., K. W. Cummins, and V. H. Resh. 1984. Collecting, sampling, and rearing methods for aquatic insects. Pages 11–26 in R. W. Merritt and K. W. Cummins, editors. An introduction to the aquatic insects of North America, 2nd edition. Kendall/Hunt Publishing Company, Dubuque, Iowa.
- Michaletz, P. H. 2006. Do channel catfish stockings affect growth and size structure of bluegills in small impoundments? *Transactions of the American Fisheries Society* 135:1254–1265.
- Miranda, L. E. 1993. Sample sizes for estimating and comparing proportion-based indices. *North American Journal of Fisheries Management* 13:383–386.
- Miranda, L. E. 2007. Approximate sample sizes required to estimate length distributions. *Transactions of the American Fisheries Society* 136:409–415.
- Murphy, B. R., and D. W. Willis. 1992. Proper distinction between relative weight and relative condition factor. *North American Journal of Fisheries Management* 12:665–666.
- Murphy, B. R., and D. W. Willis, editors. 1996. Fisheries techniques, 2nd edition. American Fisheries Society, Bethesda, Maryland.
- Neumann, R. M., and M. Allen. 2007. Size structure. Pages 375–421 in C. S. Guy and M. L. Brown, editors. Analysis and interpretation of freshwater fisheries data. American Fisheries Society, Bethesda, Maryland.
- Patton, T. M., W. A. Hubert, F. J. Rahel, and K. G. Gerow. 2000. Effort needed to estimate species richness in small streams on the Great Plains in Wyoming. *North American Journal of Fisheries Management* 20:394–398.

- Paukert, C. P. 2004. Comparison of electrofishing and trammel netting variability for sampling native fishes. *Journal of Fish Biology* 65:1643–1652.
- Peterman, R. M. 1990. Statistical power analysis can improve fisheries research and management. *Canadian Journal of Fisheries and Aquatic Sciences* 47:2–15.
- Peterson, J. T., and C. F. Rabeni. 1995. Optimizing sampling effort for sampling warmwater stream fish communities. *North American Journal of Fisheries Management* 15:528–541.
- Pine, W. E., K. H. Pollock, J. E. Hightower, T. J. Kwak, and J. A. Rice. 2003. A review of tagging methods for estimating fish population size and components of mortality. *Fisheries* 28(10):10–23.
- Pope, K. L., and C. G. Kruse. 2007. Condition. Pages 423–471 in C. S. Guy and M. L. Brown, editors. *Analysis and interpretation of freshwater fisheries data*. American Fisheries Society, Bethesda, Maryland.
- Quist, M. C., K. G. Gerow, M. R. Bower, and W. A. Hubert. 2006. Random- versus fixed-site sampling when monitoring relative abundance of fishes in headwater streams of the upper Colorado River basin. *North American Journal of Fisheries Management* 26:1011–1019.
- Scheaffer, R. L., W. Mendenhall, III, and R. L. Ott. 1996. *Elementary survey sampling*; 5th edition. Duxbury Press, Wadsworth Publishing Company, Belmont, California.
- Schnute, J. T., and L. J. Richards. 1994. Stock assessment of the 21st century. *Fisheries* 19:10–16.
- Simonson, T. D., J. Lyons, and P. D. Kanehl. 1994. Quantifying fish habitat in streams: transect spacing, sample size, and a proposed framework. *North American Journal of Fisheries Management* 14:607–615.
- Simpkins, D. G., W. A. Hubert, C. Martinez Del Rio, and D. C. Rule. 2003. Physiological responses of juvenile rainbow trout to fasting and swimming activity: effects on body composition and condition indices. *Transactions of the American Fisheries Society* 132:576–589.
- Snedecor, G. W., and W. G. Cochran. 1989. *Statistical methods*, 8th edition. Iowa State University Press, Ames.
- Tate, W. B., M. S. Allen, R. A. Meyers, and J. R. Estes. 2003. Comparison of electrofishing and rotenone for sampling largemouth bass in vegetated areas of two Florida lakes. *North American Journal of Fisheries Management* 23:181–188.
- Urquhart, N. S., S. G. Paulsen, and D. P. Larsen. 1998. Monitoring for policy-relevant regional trends over time. *Ecological Applications* 8:246–257.
- Van Den Avyle, M. J., and R. S. Hayward. 1999. Dynamics of exploited fish populations. Pages 127–166 in C. C. Kohler and W. A. Hubert, editors. *Inland fisheries management in North America*, 2nd edition. American Fisheries Society, Bethesda, Maryland.
- Vokoun, J. C., C. F. Rabeni, and J. S. Stanovick. 2001. Sample-size requirements for evaluating population size structure. *North American Journal of Fisheries Management* 21:660–665.
- Walsh, M. G., D. B. Fenner, and D. L. Winkelman. 2002. Comparison of an electric seine and prepositioned area electrofishers for sampling stream fish communities. *North American Journal of Fisheries Management* 22:77–85.
- Wedemeyer, G. A., B. A. Barton, and D. J. McLeay. 1990. Stress and acclimation. Pages 451–489 in C. B. Schreck and P. B. Moyle, editors. *Methods for fish biology*. American Fisheries Society, Bethesda, Maryland.
- Wege, G. J., and R. O. Anderson. 1978. Relative weight (W_r): a new index of condition for largemouth bass. Pages 79–91 in G. D. Novinger and J. G. Dillard, editors. *New approaches to*

- the management of small impoundments. American Fisheries Society, North Central Division, Special Publication 5, Bethesda, Maryland.
- Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and management of animal populations: modeling, estimation, and decision making. Academic Press, New York.
- Willis, D. W., B. R. Murphy, and C. S. Guy. 1993. Stock density indices: development, use, and limitations. *Reviews in Fisheries Science* 1:203–222.
- Wolda, H. 1981. Similarity indices, sample size and diversity. *Oecologia* 50:296–302.
- Zar, J. H. 1984. Biostatistical analysis. Prentice Hall, Englewood Cliffs, New Jersey.