9 Size Structure

Robert M. Neumann and Micheal S. Allen

9.1 INTRODUCTION

Size structure analysis is one of the most commonly used fisheries assessment tools. The size structure of a fish population at any point in time can be considered a snapshot that reflects the interactions of the dynamic rates of recruitment, growth, and mortality. Thus, length-frequency data provide valuable insight into the dynamics of fish populations and help identify problems such as inconsistent year-class strength, slow growth, or excessive mortality (Anderson and Neumann 1996). In most cases, a thorough interpretation of size structure data is complemented by other population assessment tools, such as catch per unit effort (C/f), age-and-growth analysis, recruitment analysis, mortality, and body condition.

Proper analysis and interpretation of size structure data should begin with a clear understanding of how, when, and where data were collected. Specifically, a fisheries scientist should know how size structure data are influenced by the sampling gear, time of the year, and location where fish were sampled. The fisheries scientist should also consider whether an appropriate sample size was obtained to estimate size structure reliably.

Fisheries scientists use several techniques to analyze size structure data. In the simplest case, a length-frequency histogram (see section 9.2) is constructed or a size structure index is calculated. Oftentimes, the primary objective is to compare size structure among samples. In these cases, a fisheries scientist may be interested in answering several questions. For example, does the size structure of white crappie populations differ among water bodies? Did the size structure of a rainbow trout population change over time in response to a management action? Are the size structures obtained from a channel catfish population different between two or more sampling gears? What factors influence the size structure of walleye populations?

9.2 PRESENTATION OF SIZE STRUCTURE DATA

Three common measures of fish length include total, fork, and standard length. Total length is measured from the anterior-most part of the fish to the tip of the longest caudal fin ray when the caudal fin is compressed. In this chapter, all lengths are reported as total length. Fork length is measured from the anterior-most part of the fish to the median caudal fin rays, which typically make up the concave portion in a forked caudal fin. Standard length is measured from the anterior-most part of the fish to the end of the caudal peduncle. Anderson and Neumann (1996) described measurements of fish length in detail.

Size structure data are most commonly reported using length-frequency histograms and stock density indices (Anderson and Neumann 1996). Length-frequency histograms show the number or proportion of fish collected in various length categories. The most commonly used length-frequency histogram is the absolute length frequency, which shows the number of fish collected in various length categories (Figure 9.1A). A relative-frequency distribution shows the proportion of all fish that are represented in each length category (Figure 9.1B). For example, in Figure 9.1A, 28 fish are in the 7–9-cm length-group (labeled 8 cm), which represent about 10% of the total number of fish collected (Figure 9.1B). Relative-frequency distributions are useful for comparing length categories that contain different sample sizes, which may result from variable sampling effort or population abundance. An alternative length-frequency distribution is based on *C/f* (Figure 9.1C), which is used to indicate relative abundance of fish in each length category (see Chapter 7 for treatment of *C/f* data).

Selection of interval widths is important for interpretation of length-frequency histograms. Anderson and Neumann (1996) suggested using 1-cm intervals for species that reach 30 cm, 2-cm intervals for 60-cm species, and 5-cm intervals for 150-cm species. Effects of interval width on the characteristics of a length-frequency histogram are demonstrated in Figure 9.2. In Figure 9.2, a 1-cm length interval shows more detail with a clear mode at 10 cm, which likely represents age-0 fish collected during fall. The 2-cm interval width shows the mode of young fish less clearly, and 4-cm interval widths mask the first mode completely.

Cumulative-frequency distributions provide an alternate view of length-frequency histograms and are used in some statistical tests comparing two or more distributions. In Figure 9.3, length-frequency histograms of age-0 walleye from three populations are presented. The respective cumulative-frequency distributions are shown in Figure 9.4. Differences in the size structure of walleye among populations are apparent in the length-frequency histograms and in the cumulative-frequency distributions. In Island Lake, the cumulative-frequency line approaches 100% at a shorter length than does the cumulative-frequency line in Lake Thompson because the Lake Thompson sample contains larger walleye (>200 mm) than does Island Lake. The cumulative-frequency lines clearly show that the sample for Lake Mitchell contains longer walleye overall than do the other lakes, but the maximum length of walleye is the same between Lake Mitchell and Lake Thompson (i.e., the cumulative-frequency lines both approach 100% at 210 mm). Another interpretation is that approximately 50% of the walleye in Island Lake and Lake Thompson are shorter than 160 mm, whereas 50% of the walleye in Lake Mitchell are shorter than 190 mm.

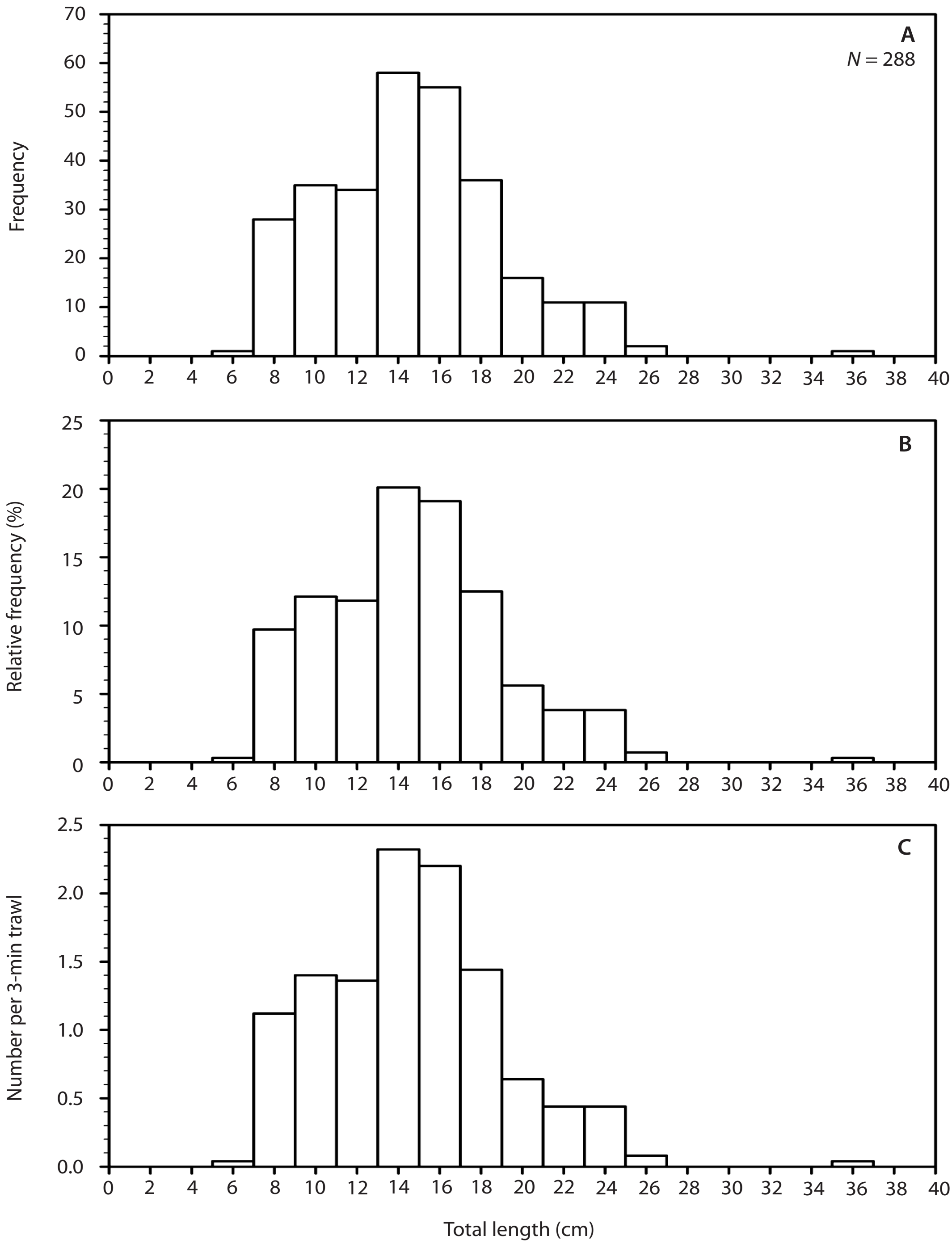

Figure 9.1 Length-frequency histograms for black crappie collected from Lake Jeffords, Florida. Data are displayed using (A) absolute length frequency, (B) relative length frequency, and (C) catch per unit effort.

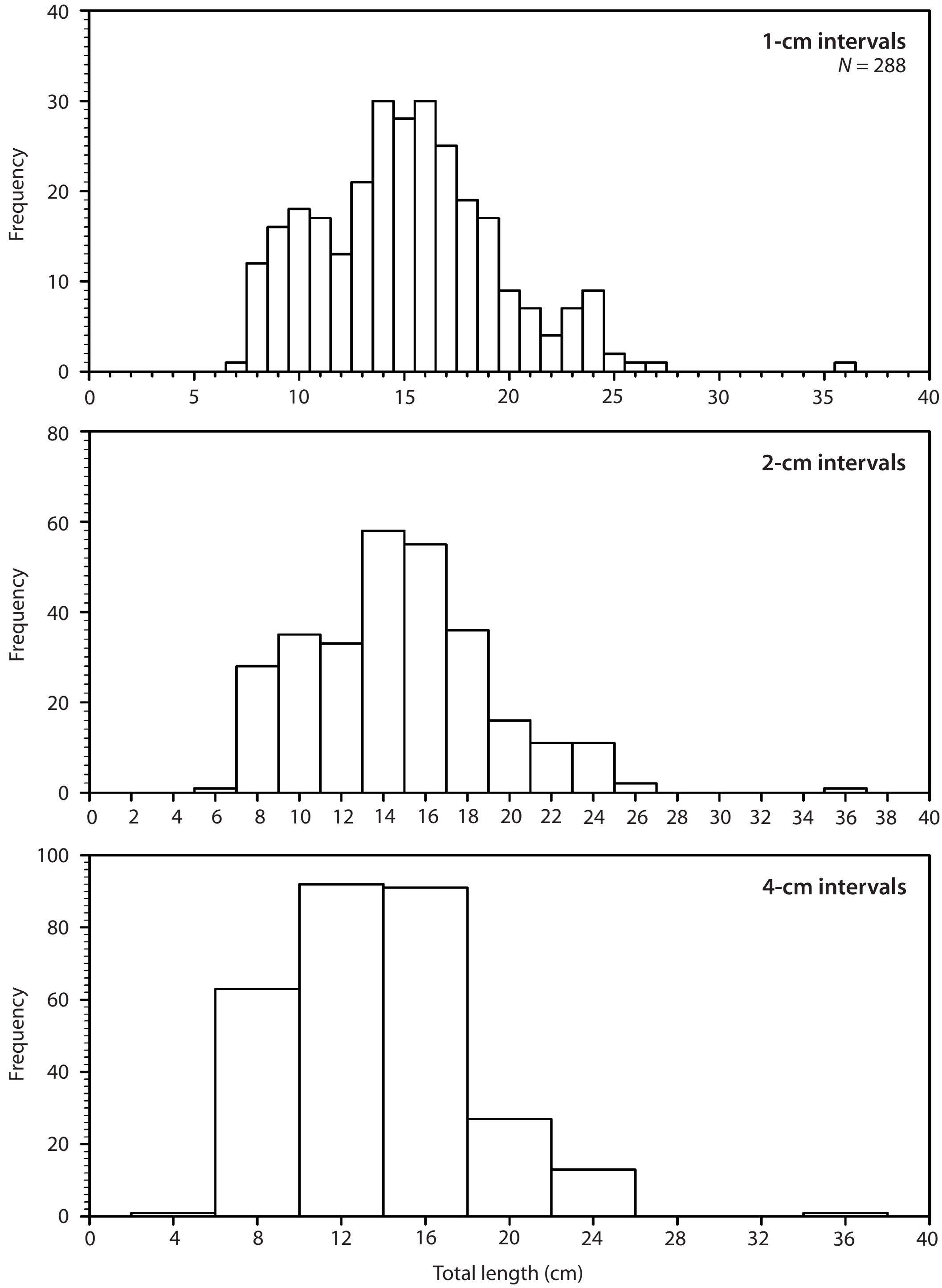

Figure 9.2 Absolute- length-frequency histograms constructed with length interval widths of 1, 2, and 4 cm for black crappie from Lake Jeffords, Florida.

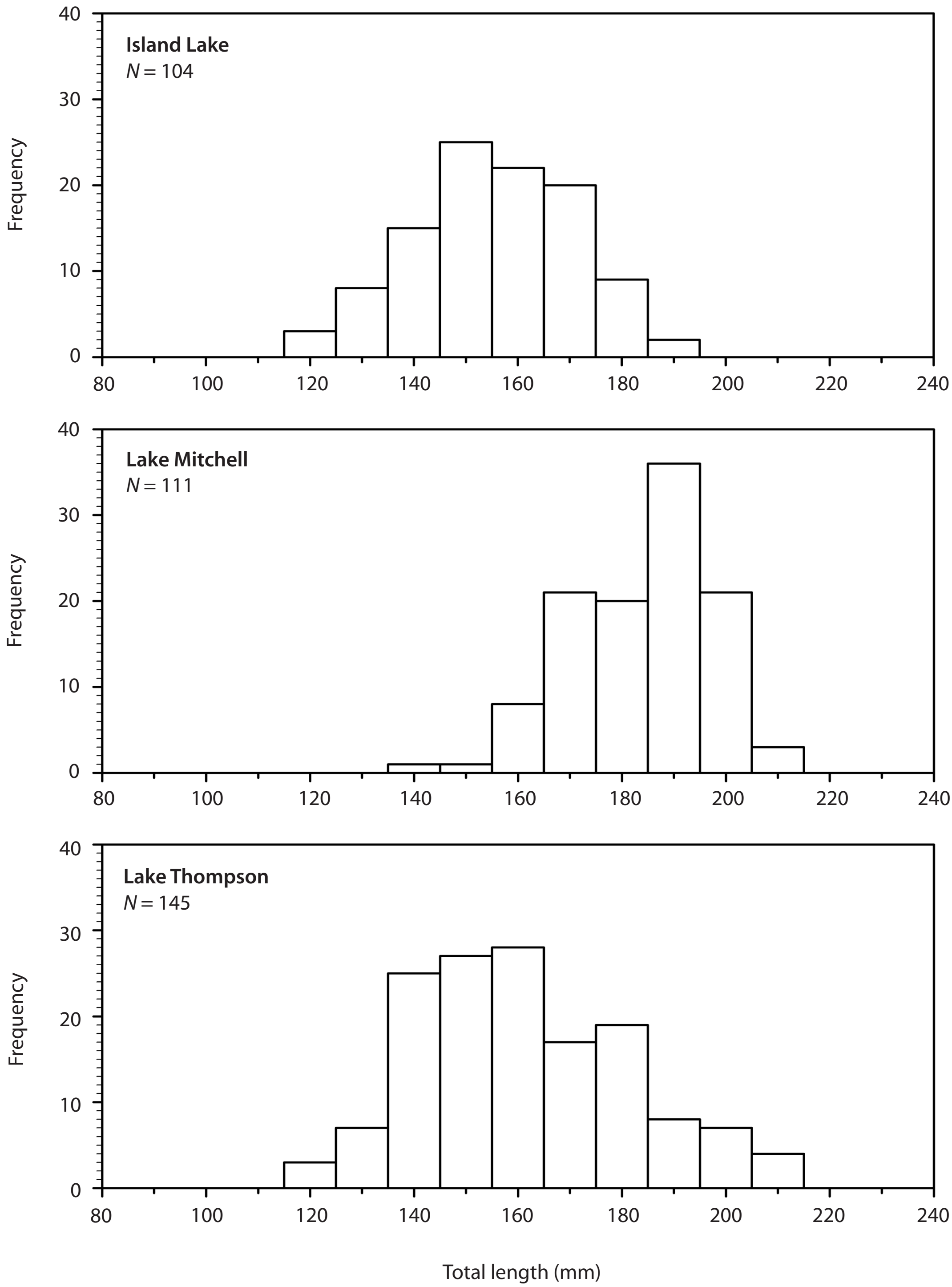

Figure 9.3 Absolute-length-frequency histograms for age-0 walleye collected from three South Dakota lakes (data courtesy of the South Dakota Department of Game, Fish and Parks).

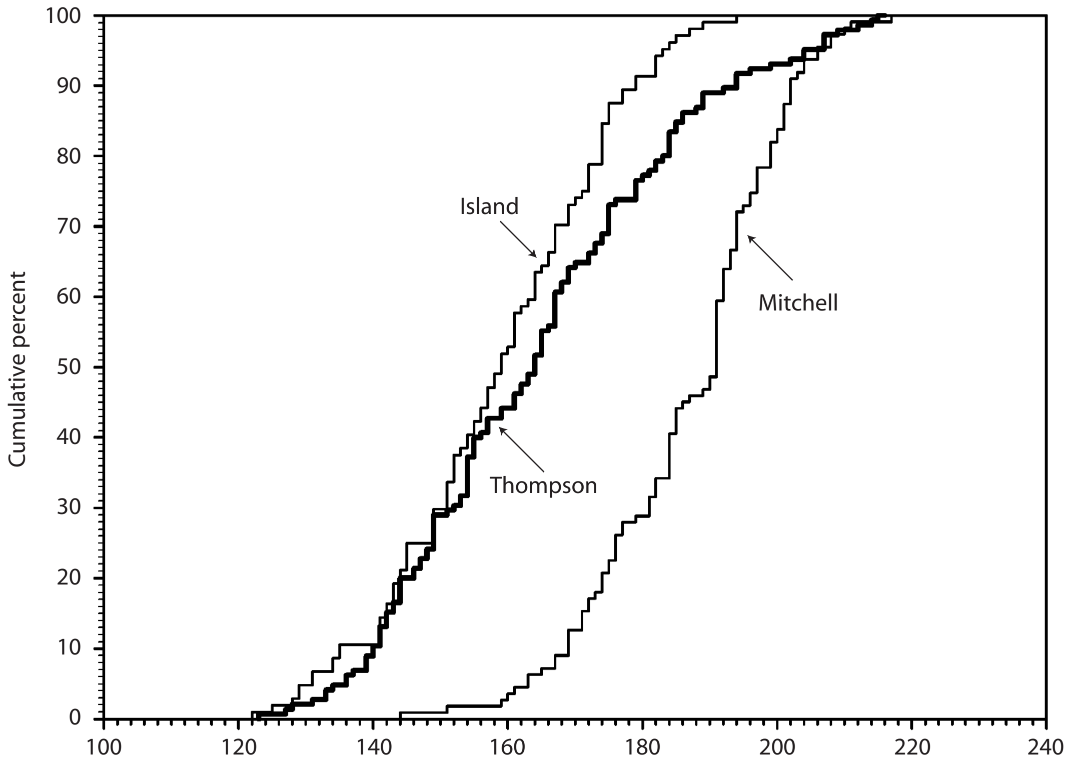

Figure 9.4 Cumulative-frequency distributions for age-0 walleye collected from three South Dakota lakes (data courtesy of the South Dakota Department of Game, Fish and Parks). The respective length-frequency distributions are shown in Figure 9.3.

Stock density indices are used to describe size structure. A detailed review of stock density indices and their calculation can be found in Anderson and Neumann (1996). Proportional stock density (PSD) is calculated as

$$\text{PSD} = \frac{\text{Number of fish} \geq \text{quality length}}{\text{Number of fish} \geq \text{stock length}} \times 100. \tag{9.1}$$

Relative stock density (RSD) is expressed as

$$\text{RSD} = \frac{\text{Number of fish} \geq \text{specified length}}{\text{Number of fish} \geq \text{stock length}} \times 100, \tag{9.2}$$

where the specified length often refers to preferred, memorable, or trophy length. Relative stock densities of preferred-, memorable-, and trophy-length fish are reported as RSD-P, RSD-M, and RSD-T, respectively. The standard convention is to report stock density index values to the nearest whole number without a percent symbol. Minimum stock, quality, preferred, memorable, and trophy lengths for many species are provided in Anderson and Neumann (1996) and Bister et al. (2000). In traditional stock density index calculations, it is important to emphasize that stock, quality, preferred, memorable, and trophy lengths are minimum lengths. For example, stock and quality lengths for largemouth bass are 20 and 30 cm, respectively. Thus, in a sample of largemouth bass, all fish greater than 20 cm are stock length, and all fish greater than 30 cm are quality length. Length-frequency data can also be indexed using incremental stock density indices (Anderson and Neumann 1996).

The use of PSD alone to index size structure can often lead to loss of data sensitivity. For example, two largemouth bass populations can have PSD values of 60, even though one population may be quite different from the other when the length-frequency histograms are inspected. This is because quality length includes all fish greater than or equal to quality, preferred, memorable, and trophy length. Consider two populations that both have 30 quality-length fish. In one population, all 30 quality-length fish may be between quality and preferred length, whereas in the other, 20 may be between quality and preferred length, and 10 may be between preferred and memorable length. This example illustrates the importance of calculating other stock density indices (e.g., RSD-P) to index size structure precisely. Fisheries scientists should calculate the stock density index for the largest length category of interest, given an appropriate sample size.

Gustafson (1988) provided a formula and easy-to-use tables for determining 80% and 95% confidence intervals around stock density index values (Tables 9.1, 9.2). Confidence interval widths depend on sample size and the magnitude of the stock density index value. Although confidence intervals provide a measure of variation around stock density index values, they should not be used as a test for determining statistically significant differences between two or more values, primarily because confidence intervals for index values with unequal sample sizes were derived from distributions with unequal variances. Trippel and Hubert (1990) cautioned against the use of confidence interval overlap as a test for differences between means unless variance is pooled. Statistical treatment of stock density index values is presented in section 9.4.

9.3 COLLECTION OF SIZE STRUCTURE DATA

In an ideal situation, the size structure of a fish population determined from samples would be the same as the true size structure of the fish population. However, when fisheries scientists collect a sample of fish, the size structure obtained from that sample is often different from the true size structure of the fish population. Size structure from samples can be misrepresentative of the true population because the lengths of fish collected may depend on the type of sampling gear used, the season in which the fish were collected, and the location chosen to collect the fish. To overcome these effects, fisheries scientists use standardized sampling so that changes in size structure over time and comparisons of size structure among water bodies can be adequately assessed.

Table 9.1 Approximate confidence intervals (plus or minus) for proportional stock densities (PSD) as a function of sample size (*N*) of stock-length fish at the 80% confidence interval. Values have been omitted when sample sizes are insufficient for a normal approximation to the binomial distribution (from Gustafson 1988).

	Estimated PSD																		
N	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95
10										30									
15							22	22	22	22	22	22	22						
20					16	17	18	18	18	19	18	18	18	17	16				
25				13	14	15	15	16	16	16	16	16	15	15	14	13			
30				12	13	13	14	14	14	14	14	14	14	13	13	12			
35			10	11	12	12	13	13	13	13	13	13	13	12	12	11	10		
40			9	10	11	11	12	12	12	12	12	12	12	11	11	10	9		
45			8	9	10	10	11	11	11	11	11	11	11	10	10	9	8		
50		7	8	9	9	10	10	10	11	11	11	10	10	10	9	9	8	7	
55		6	7	8	9	9	10	10	10	10	10	10	10	9	9	8	7	6	
60		6	7	8	8	9	9	9	9	10	9	9	9	9	8	8	7	6	
65		6	7	7	8	8	9	9	9	9	9	9	9	8	8	7	7	6	
70		6	6	7	8	8	8	9	9	9	9	9	8	8	8	7	6	6	
75		5	6	7	7	8	8	8	8	8	8	8	8	8	7	7	6	5	
80		5	6	7	7	7	8	8	8	8	8	8	8	7	7	7	6	5	
85		5	6	6	7	7	7	8	8	8	8	8	7	7	7	6	6	5	
90		5	6	6	7	7	7	7	8	8	8	7	7	7	7	6	6	5	
95		5	5	6	6	7	7	7	7	7	7	7	7	7	6	6	5	5	
100	3	5	5	6	6	7	7	7	7	7	7	7	7	7	6	6	5	5	3
120	3	4	5	5	6	6	6	6	6	6	6	6	6	6	6	5	5	4	3

9.3.1 Standardized Sampling for Size Structure Data

Willis and Murphy (1996) emphasized the importance of standardized sampling methods because of the numerous gear-, season-, and location-related effects on sampling data for many fishes. They recommended that standardized sampling should consider the use of an effective gear for the fish species being sampled, that the gear be used during an effective time of the year, and that gears be set in standard locations from year to year. Thus, long-term data sets can be established, and trends in sample variables can be monitored over time.

Many fishery management activities focus on the adult portion of a population, whether the goal is to increase abundance of adult fish, increase size structure, or manipulate the adult stock to influence predator–prey dynamics. In cases in which changes in the adult portion of a population are being investigated, the use of a gear that effectively samples fishes that are stock length (see section 9.2) and greater is recommended. Rarely does one gear type effectively sample all lengths of fish in a population. Thus, investigations focusing on recruitment and year-class strength, for which the capture of juvenile fishes is necessary, may require a gear different than that used to capture adult fish. Because each gear

Table 9.2 Approximate confidence intervals (plus or minus) for PSD as a function of sample size (*N*) of stock-length fish at the 95% confidence interval. Values have been omitted when sample sizes are insufficient for a normal approximation to the binomial distribution (from Gustafson 1988).

	Estimated PSD																		
N	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95
10										48									
15							34	35	35	36	35	35	34						
20					26	27	28	29	29	29	29	29	28	27	26				
25				21	22	23	24	25	25	25	25	25	24	23	22	21			
30				19	20	21	21	22	22	22	22	22	21	21	20	19			
35			15	17	18	19	20	20	20	20	20	20	20	19	18	17	15		
40			14	15	17	17	18	18	19	19	19	18	18	17	17	15	14		
45			13	14	15	16	17	17	17	17	17	17	17	16	15	14	13		
50		11	12	13	14	15	16	16	16	16	16	16	16	15	14	13	12	11	
55		10	12	13	14	14	15	15	15	15	15	15	15	14	14	13	12	10	
60		9	11	12	13	14	14	14	15	15	15	14	14	14	13	12	11	9	
65		9	10	12	12	13	13	14	14	14	14	14	13	13	12	12	10	9	
70		9	10	11	12	12	13	13	13	13	13	13	13	12	12	11	10	9	
75		8	10	11	11	12	12	13	13	13	13	13	12	12	11	11	10	8	
80		8	9	10	11	12	12	12	12	12	12	12	12	12	11	10	9	8	
85		8	9	10	11	11	12	12	12	12	12	12	12	11	11	10	9	8	
90		7	9	10	10	11	11	11	12	12	12	11	11	11	10	10	9	7	
95		7	8	9	10	10	11	11	11	11	11	11	11	10	10	9	8	7	
100	5	7	8	9	10	10	11	11	11	11	11	11	11	10	10	9	8	7	5
120	5	6	7	8	9	9	9	10	10	10	10	10	9	9	9	8	7	6	5

type, and configurations within a specific gear type, may select for different sizes of fish, combining data from several gears to determine size structure is not recommended. Rather, standard gears should be used that will allow for comparisons of size structure over time or among water bodies.

9.3.2 Effects of Gear Type on Size Structure

Size selectivity of a particular gear can be related to the physical dimensions of a mesh, reaction of fish to a gear, and the location in a water body where the gear is used. Much research has been conducted to achieve a better understanding of the size selectivity of various sampling gears. Hubert (1996) provided information on the size and seasonal biases of many passive gears used in fisheries research.

With gill nets, only a limited range of lengths are sampled within a given mesh size, with fish of a particular length being held most securely. Fisheries scientists should be aware of mesh-size selectivity and mesh-size efficiency to interpret size structure data collected with gill nets properly. Hubert (1996) provided a detailed explanation of mesh-size efficiency and selectivity and referenced methods to correct size structure data from gill nets. Experimental gill nets, which include several mesh sizes, are often used to sample a broad length range of the species under consideration. The use of experimental gill nets does not ensure that the size structure of the collected fish will be representative of the true size structure of the fish population because mesh-size selectivity and efficiency may still influence the sample size structure. Size selectivity may be reduced when the mesh size and twine diameter complement is carefully chosen and evaluated.

In gears such as gill nets, hoop nets, trap nets, and trawls, the mesh size will determine the minimum length of fish captured (Hubert 1996). Other aspects of net construction, such as mesh material, frame dimensions, and mouth size, also influence size selectivity. Laarman and Ryckman (1982) found that trap nets were selective for larger sizes of some fish species but not others. Holland and Peters (1992) compared length distributions of channel catfish captured from the Platte River, Nebraska, in hoop nets with three different mesh sizes and found that both the minimum and maximum lengths of fish increased as mesh size increased. The effects of mesh size on size selectivity for catfishes has been well documented, and, in general, samples collected with larger mesh sizes produce larger mean lengths (Vokoun and Rabeni 1999).

Electrofishing has also been shown to have size-selective properties. Reynolds and Simpson (1978) demonstrated that in Midwestern ponds, electrofishing efficiency increased as a function of total length for largemouth bass. For bluegill, electrofishing efficiency was higher for 8–15-cm bluegills compared with bluegills less than 8 cm or greater than 15 cm (Reynolds and Simpson 1978). Milewski and Willis (1991) found that compared with trap nets, electrofishing resulted in smaller size structure for smallmouth bass. Robinson (1994) found that large flathead catfish (≥75 mm) were rarely captured when pulsed DC electrofishing was used. Santucci et al. (1999) determined that for channel catfish in a small impoundment, AC electrofishing selected for smaller fish in the population.

Size structure data collected by underwater observation (i.e., snorkel or scuba) were shown to be overestimated because of underwater magnification (Griffith 1981). Mullner et al. (1998) found that length frequencies of three trout species and their hybrids were significantly different between snorkeling and electrofishing samples, and they used an underwater magnification factor of 1.25 to adjust length frequencies. In contrast, Wildman and Neumann (2003) found that when broad length categories were used, size structure estimated by snorkeling and electrofishing were not substantially different for brook trout and brown trout in Connecticut streams.

Size structure is underestimated for most species of fish captured in cove rotenone samples (Hayne et al. 1967). Bayley and Austen (1990) tested the sampling efficiency of rotenone in ponds and coves and found that efficiency was high for large fish in warm water and low for small fish in cool water. Typically, cove rotenone sampling is conducted during mid to late summer, when large individuals of many species move offshore (Willis et al. 1993; Bettoli and Maceina 1996). Thus, summer cove rotenone samples may be selective for small fish.

The use of size structure data obtained from competitive fishing events and angler diaries is becoming more common. As with any collection technique, caution must be given to how angler data are interpreted because angler data may be selective for larger fish compared with data from traditional sampling gears (Willis et al. 1993). Gabelhouse and Willis (1986) found that tournament anglers in Kansas selected for larger sizes of largemouth bass than did electrofishing, and stock density indices (see section 9.2) calculated from angler data were higher than those based on electrofishing samples. Jacobs et al. (1995) found that in Connecticut lakes, the proportion of largemouth bass greater than 38 cm was usually greater for electrofishing samples compared with tournament samples. In contrast, smallmouth bass greater than 38 cm tended to be underestimated in electrofishing samples compared with tournament samples. Green et al. (1986) found differences in size structure of largemouth bass and smallmouth bass based on data collected by anglers using diaries versus by electrofishing. They provided empirical adjustment factors to predict the size structure of largemouth bass and smallmouth bass in electrofishing samples from the length distribution of the angler catch. In contrast, Ebbers (1987) found that largemouth bass size structure estimated from angler diaries and electrofishing samples were similar in Minnesota. Thus, angler behavior may vary by geographic location and affect data used to determine size structure.

9.3.3 Effects of Sampling Time on Size Structure

Size structure of samples can differ among seasons of the year even when a standard gear is being used. Seasonal changes in size structure occur because of size-dependent changes in fish behavior and physiology throughout the year (Pope and Willis 1996). For example, Carline et al. (1984) found that for largemouth bass sampled by electrofishing in an Ohio impoundment, samples contained larger fish in spring and fall compared to summer. Largemouth bass greater than 30 cm apparently moved offshore after spawning and were not as vulnerable to capture during summer; as water temperature cooled during the fall, large fish returned to inshore areas. Gilliland (1987) and Bettross and Willis (1988) have reported similar seasonal changes in size structure for largemouth bass.

Pope and Willis (1996) provided a review of several studies that documented seasonal changes in size structure. Spring and fall peaks in size structure have been observed for several species, including bluegill captured in trap nets (Bettross and Willis 1988) and yellow perch (Lott and Willis 1991), walleye, and sauger (Mero and Willis 1992) captured in experimental gill nets. Boxrucker and Ploskey (1989) found that greater proportions of larger and older white crappies were captured in trap nets during spring than fall in Oklahoma impoundments. Seasonal patterns in size structure other than spring and fall peaks have also been observed. In a South Dakota lake, size structure of northern pike sampled with experimental gill nets was highest during winter and declined into the summer; significant inverse correlations between size structure and water temperature were observed (Neumann and Willis 1995).

Size structure has also been shown to differ even within a single season and between day and night samples. Across a 1-month period during spring, size structure of largemouth bass captured by electrofishing increased substantially, apparently due to a greater proportion of largemouth bass greater than 30 cm moving to inshore areas in preparation for spawning (Carline et al. 1984). Paragamian (1989) found that the size structure determined from samples of smallmouth bass was higher at night than that during the day in an Iowa river. In Oklahoma reservoirs, largemouth bass size structure was similar between day and night electrofishing samples in spring, but during fall, day samples produced a narrower range of fish lengths and contained mostly smaller individuals (Gilliland 1987). Size structure of sauger captured by electrofishing during the day in a turbid main-stem reservoir was consistently higher than at night, and sauger greater than 51 cm were collected only during the day (Van Zee et al. 1996).

9.3.4 Effects of Sample Location on Size Structure

Biologists often choose subjective sampling sites based on the likelihood of capturing a large sample size of the target species (Willis et al. 1993). Hubbard and Miranda (1988) found that the size structure of largemouth bass collected by electrofishing from subjective sites was greater than was the size structure obtained from random sites. King et al. (1981) compared sample parameters for several fish species collected by electrofishing from fixed and random sites. They found few statistical differences in population parameters between the two types of sampling sites. However, the fixed sites they sampled over time were initially chosen at random.

Sampling fish from fixed or random sites should depend on the experimental design being used. Sampling at fixed sites is often used to track changes in population characteristics within a single water body, whereas sampling at random sites is more suitable for comparing population characteristics among water bodies. The use of fixed or random sites may also depend on the need to continue standard sampling designs previously developed.

9.3.5 Sample Size Considerations

The sample size necessary to describe the size structure of a fish population adequately is quite large. Anderson and Neumann (1996) recommended that for general stock assessment purposes, at least 100 fish greater than stock length (see section 9.2) should be sampled. Gilliland (1987) compared length frequencies based on various sample sizes of largemouth bass that were sampled by electrofishing in Oklahoma reservoirs and concluded that a sample size of 150 largemouth bass was adequate to estimate size structure, whereas a sample of 50 was not adequate. More recently, Vokoun et al. (2001) estimated the sample size necessary to construct a length-frequency distribution with a given accuracy and precision for bluegill and channel catfish. They compared the length frequency histogram from a known sample to computer generated length frequency histograms by means of bootstrapping methods. Their results demonstrated the importance of using at least 300–400 individuals whenever possible.

Weithman et al. (1980) developed a sequential sampling method that allows a biologist to monitor continuously how many stock-length and quality-length fish are necessary to obtain a reliable estimate of PSD while sampling is being conducted. Miranda (1993) developed a method by which biologists can approximate the sample size required for estimating PSD before collection begins. These sampling methods are further described in Anderson and Neumann (1996). Sample size requirements discussed in this section are recommendations based on existing information. Clearly, the sample size necessary to describe size structure reliably will depend on the species, population structure, sampling constraints, and study objectives.

9.4 STATISTICAL ANALYSES FOR SIZE STRUCTURE DATA

Analysis of size structure data should begin with an exploratory analysis by constructing length-frequency histograms or calculating stock density index values. There are also many statistical tests available to analyze size structure data. In this section, we review several statistical techniques commonly applied to size structure data. Experimental design considerations and statistical assumptions are reviewed in Chapters 2 and 3.

Fisheries scientists are often interested in comparing size structure between two or more samples. For example, comparisons of size structure are often made between different gear types, water bodies, or time periods. Consider the comparison of two hypothetical length-frequency data sets. Many commonly applied statistical tests, such as *t*-tests and analysis of variance (ANOVA), assume that data are normally distributed and, as such, are typically not appropriate for tests of length-frequency data (Brown and Austen 1996). When a broad length range of fish is sampled, length-frequency data are often multimodal, highly skewed, and contain extreme observations. In these cases, nonparametric tests may be more appropriate for comparing length-frequency distributions. Conditions favorable for nonparametric statistics and cautions about their use are described in Brown and Austen (1996) and Chapter 1. Given sufficient sample sizes, and when data approximate a normal distribution, most commonly employed parametric tests are sufficiently robust and can perform well (Zar 1996). Methods to evaluate normality of data and considerations for data transformations are provided in Chapter 3.

9.4.1 Parametric Tests

Assuming a normal distribution, size structure data are commonly compared between two samples using a *t*-test or by an ANOVA in the case of comparing more than two samples. These tests are used to compare the estimated means (e.g., means of length) to determine whether or not the samples come from the same population (Koopmans 1987). The fisheries scientist may use these tests to determine if mean lengths of samples are significantly different. An ANOVA is typically followed by a multiple-comparison test to determine which means are significantly different from one another. An example of using an ANOVA to compare mean length among three length-frequency samples is provided in Box 9.1.

9.4.2 Nonparametric Tests for Comparing Size Structure

Several nonparametric statistical tests are useful for comparing size structures from two or more samples. Nonparametric tests are usually applied to length-frequency data, primarily because of concerns regarding the distribution of the data. Nonparametric tests commonly applied to length-frequency data include the Kolmogorov–Smirnov two sample, Wilcoxon's rank sum, Kruskal–Wallis, and the chi-square. The Kolmogorov–Smirnov two-sample test is used to determine whether the distribution of a variable (e.g., length) is the same across different groups (e.g., lakes). The test statistic is calculated as the largest absolute distance between the distribution functions (cumulative frequency distributions) associated with the samples (Zar 1996; SAS 1999). This test is often used to determine whether length-frequency distributions are different between samples (Box 9.2). Examples of the application of the Kolmogorov–Smirnov test to examine differences among length-frequency distributions can be found in Cornelius and Margenau (1999), Underwood (2000), Unmuth et al. (2001), Isermann et al. (2002), and Tate et al. (2003).

When applying two-sample tests (such as the Kolmogorov–Smirnov), pairwise tests are performed rather than multiple comparisons. Under these circumstances, the significance level for comparisons should be adjusted using the Bonferroni correction in order to maintain the predetermined experimentwise error rate (Koopmans 1987). This can be achieved by setting the significance level for each subtest equal to the experimentwise error rate divided by the number of subtests. For example, if the experimentwise error rate was $\alpha = 0.05$ and there were three subtests performed, then the significance level for each subtest would be $\alpha = 0.05/3 = 0.017$.

Wilcoxon's rank-sum test for two samples and the Kruskal–Wallis test for several samples are rank-testing procedures and sometimes are considered nonparametric counterparts to the *t*-test and ANOVA, respectively. In fact, the Kruskal–Wallis test is often called "ANOVA by ranks" (Zar 1996). For these tests, the observations from all samples are combined, ordered, and assigned a rank value, and the test statistic is calculated based on rank scores. These tests are used to test for differences in location and scale based on rank scores. Fisheries scientists often use these tests to determine whether length-frequency distributions are different among samples (i.e., does one population tend to yield larger or smaller values than the other). An example of the Kruskal–Wallis test applied to length-frequency data is provided in Box 9.3; additional applications of the Kruskal–Wallis test to length-frequency data can be found in Neumann et al. (1995) and Neal et al. (1999). Several nonparametric multiple-comparison tests are available for use with tests such as the Kruskal–Wallis test (Conover 1980; Zar 1996). Examples of multiple-comparison testing procedures are provided in Box 9.4.

According to Conover (1980), an advantage of the Kolmogorov–Smirnov two-sample test over rank tests (e.g., Wilcoxon-s rank sum and Kruskal–Wallis) is that the Kolmogorov–Smirnov test is sensitive to detecting differences in location (magnitude of observations) and shape (variance) between distribution functions. Methods based on ranks are sensitive to differences in the magnitude of ranked data among samples, but they may not detect differences in variances or shape of the distributions. Thus, fisheries scientists should visually inspect length-frequency histograms and use statistical tests cautiously when analyzing length-frequency data.

Chi-square tests are commonly used to test for differences in length frequency among samples. Examples of the application of the chi-square test to length-frequency data can be found in Michaletz et al. (1995), Van Den Avyle et al. (1995), Roni and Fayram (2000), and Wildman and Neumann (2003). The chi-square test is used to test that the frequencies of observations among length-groups is independent of the treatment (e.g., water body, gear type, or time period). Chi-square tests are often applied, but not limited to, length-frequency data for which the length-groups are rather large. For example, length data are often categorized using stock density index length categories rather than by more detailed length intervals (Box 9.5)

When size structure is indexed using stock density indices, a fisheries scientist may be interested in statistically comparing stock density index values between two or more samples. Because stock density index values are frequently calculated from a more detailed length-frequency histogram, statistical procedures (as described above) can be applied to the raw length-frequency data, and the outcome of those tests along with stock density index values can be reported. An alternate approach may be to use a chi-square test (Box 9.5) in which stock density index length categories are used as length intervals. Fisheries scientists are often involved in studies in which a treatment (e.g., an experimental harvest regulation) is applied to several water bodies, and additional water bodies are used as a control group. In this case, stock density indices can be calculated for each water body, and the fisheries scientist can test for differences in the mean stock density index values (e.g., mean PSD) between treatments. For example, Margenau and AveLallemant (2000) used two-sample *t*-tests to compare mean stock density index values of muskellunge populations before and after a special harvest regulation was implemented. Proportions (such as PSD) form a binomial distribution rather than a normal distribution (Zar 1996). Thus, PSD values may require a data transformation (e.g., arcsine-root) before analyses (see Chapter 3 for discussion of data transformations).

9.4.3 The Experimental Unit

In each of the examples presented in Boxes 9.1–9.5, catches of fish in each unit of effort were pooled into a single sample, and statistical tests were performed on pooled length-frequency data. By far, this is the most commonly used approach to treating and testing size structure data. In some instances, performing tests on pooled length-frequency data can result in tests with inflated power, resulting in significant differences in length-frequency distributions even though there may be only slight differences between distributions. This is especially the case when large sample sizes are created by pooling length-frequency data. For example, in

Box 9.1 Testing for Differences in Mean Length By Means of Analysis of Variance (ANOVA)

Age-0 walleye were sampled from three eastern South Dakota lakes (Island Lake, Lake Mitchell, and Lake Thompson) by biologists from the South Dakota Department of Game, Fish and Parks in September 2001. In each lake, six 20-min-standardized sites were sampled at night with an electrofishing boat. Because the distributions of lengths in each sample were considered normal, ANOVA was chosen to analyze these data. The analysis was performed using the general linear model procedure (PROC GLM) in SAS (SAS 1999). The purpose of this analysis was to compare mean length of age-0 walleyes among the three lakes. Differences in mean length of age-0 walleye among lakes in fall should indicate differences in growth achieved during the first year of life. The null hypothesis is that there is no difference in mean length among lakes.

Data
The length-frequency histograms for each population are presented in Figure 9.3. All walleye were measured to the nearest millimeter.

Program
```
DATA ONE;
INPUT LAKE $ LENGTH;
CARDS;
ISLAND         122
ISLAND         126
ISLAND         129
[Data input continued]
MITCHELL       145
MITCHELL       152
MITCHELL       160
[Data input continued]
THOMPSON       123
THOMPSON       128
THOMPSON       129
[data input continued]
;
PROC SORT;
BY LAKE LENGTH;
PROC GLM;
CLASS LAKE;
MODEL LENGTH=LAKE;
RUN;
```

Output

Table General linear model (GLM) procedure for length of age-0 walleyes (dependent variable) compared among three South Dakota lakes. The data included 360 observations.

Class Level Information		
Class	Levels	Values
Lake	3	Island Mitchell Thompson

GLM Procedure					
Source	*df*	Sum of squares	Mean square	*F*-value	$P > F$
Model	2	49324.1151	24662.0575	81.73	<0.0001
Error	357	107724.5072	301.7493		
Corrected total	359	157048.6222			

Results
Results of the ANOVA indicated that there was a significant ($F = 81.73$; $P < 0.0001$) difference in mean length among lakes, leading to the rejection of the null hypothesis.

Next, a multiple-comparison test was performed to determine which mean lengths (lakes) were different from one another. In this example, the Tukey's multiple-comparison test was used; it can be invoked using the following code. The program also calls for calculation of mean length for each lake.

Program
```
PROC SORT;
BY LAKE LENGTH;
PROC GLM;
CLASS LAKE;
MODEL LENGTH=LAKE;
MEANS LAKE/TUKEY;
PROC MEANS;
BY LAKE;
VAR LENGTH;
RUN;
```

Output

Table The GLM procedure for Tukey's studentized range (HSD) test for length. This test controls the type I experimentwise error rate. Comparisons significant at the 0.05 level are indicated by ***.

Test Statistics	
Alpha	0.05
Error *df*	357
Error mean square	301.7493
Critical value of studentized range	3.32840

Means Comparisons

Lake comparison	Difference between means	Simultaneous 95% confidence limits	
Mitchell–Thompson	22.210	17.054	27.366***
Mitchell–Island	28.315	22.736	33.894***
Thompson–Mitchell	–22.210	–27.366	–17.054***
Thompson–Island	6.105	0.852	11.359***
Island–Mitchell	–28.315	–33.894	–22.736***
Island–Thompson	–6.105	–11.359	–0.852***

The MEANS Procedure

Lake	*N*	Mean	SD	Minimum	Maximum
Island	104	159.4326923	15.8853687	122.0000000	195.0000000
Mitchell	111	187.7477477	13.7136087	145.0000000	218.0000000
Thompson	145	165.5379310	20.5895809	123.0000000	216.0000000

Results
According to this test, all mean lengths are significantly ($P \leq 0.05$) different from one another. Mean length is greatest in Lake Mitchell (188 mm) followed by Lake Thompson (166 mm) and Island Lake (159 mm).

Box 9.2 Testing for Differences among Length-Frequency Distributions by Means of the Kolmogorov–Smirnov Two-Sample Test

The same walleye data analyzed in Box 9.1 (and shown in Figure 9.3) are used in this example. The purpose of this analysis is to compare length-frequency distributions of walleyes among the three lakes by means of a Kolmogorov–Smirnov two-sample test. The analysis was performed using the NPAR1WAY procedure in SAS (SAS 1999). The null hypothesis is that there are no differences in length-frequency distributions (i.e., distribution functions) among lakes. This is a popular nonparametric method to determine differences in length frequencies, as length-frequency data oftentimes deviate substantially from normal. Because this is a two-sample test, only two lakes can be compared simultaneously. Thus, a total of three comparisons (between Mitchell and Thompson, between Island and Thompson, and between Island and Mitchell) were made. In the SAS code shown, Island Lake was deleted from the analysis for the comparison between Lake Mitchell and Lake Thompson. To maintain an experimentwise error rate of $\alpha = 0.05$, the significance level for each comparison ($P = 0.017$) was established by dividing α (0.05) by the number of comparisons (3).

Data

See Box 9.1 and Figure 9.3.

Program

```
DATA ONE;
INPUT LAKE $ LENGTH;
CARDS;
ISLAND        122
ISLAND        126
ISLAND        129
[Data input continued]
MITCHELL      145
MITCHELL      152
MITCHELL      160
[Data input continued]
THOMPSON      123
THOMPSON      128
THOMPSON      129
[Data input continued]
;
DATA TWO; SET ONE;
IF LAKE = "ISLAND" THEN DELETE;
PROC SORT;
BY LAKE LENGTH;
RUN;
PROC NPAR1WAY;
CLASS LAKE;
VAR LENGTH;
RUN;
```

Output

Table Comparison of Lake Mitchell and Lake Thompson. Kolmogorov–Smirnov test for variable LENGTH classified by variable LAKE. The EDF is the empirical distribution function; KS represents the Kolmogorov–Smirnov statistic and KS_a the asymptotic KS; D is the two-sample KS statistic; and $P > KS_a$ is the asymptotic P-value of KS_a, which equals $P > D$.

Kolmogorov–Smirnov Test			
Lake	*N*	EDF at maximum	Deviation from mean at maximum
Mitchell	111	0.090090	–3.166332
Thompson	145	0.620690	2.770345
Total	256	0.390625	

Maximum deviation occurred at observation	201
Value of LENGTH at maximum	169.0

Kolmogorov–Smirnov Two-Sample Test (Asymptotic)			
KS	0.262950	*D*	0.530600
KSa	4.207193	$P > KS_a$	<0.0001

Table Comparison of Island Lake and Lake Thompson. Kolmogorov–Smirnov test for variable LENGTH classified by variable LAKE.

Kolmogorov–Smirnov Test			
Lake	*N*	EDF at maximum	Deviation from mean at maximum
Island	104	0.846154	0.929386
Thompson	145	0.689655	–0.787098
Total	249	0.755020	

Maximum deviation occurred at observation	204
Value of LENGTH at maximum	175.0

Kolmogorov–Smirnov Two-Sample Test (Asymptotic)			
KS	0.077181	*D*	0.156499
KSa	1.217900	$P > KS_a$	0.1029

Table Comparison of Island Lake and Lake Mitchell. Kolmogorov–Smirnov test for variable LENGTH classified by variable LAKE.

Kolmogorov–Smirnov Test			
Lake	*N*	EDF at maximum	Deviation from mean at maximum
Island	104	0.875000	3.421086
Mitchell	111	0.225225	–3.311458
Total	215	0.539535	

Maximum deviation occurred at observation	129
Value of LENGTH at maximum	176.0

Kolmogorov–Smirnov Two-Sample Test (Asymptotic)			
KS	0.324715	*D*	0.649775
KSa	4.761259	$P > KS_a$	<0.0001

Results

Results of these tests indicate that differences in the length-frequency distributions (i.e., distribution functions) were found among the three lakes, leading to the rejection of the null hypothesis. The length-frequency distribution of age-0 walleye in Lake Mitchell was significantly ($P < 0.0001$) greater than that in Island Lake and Lake Thompson. No difference was observed between Island Lake and Lake Thompson ($P = 0.1037$). Thus, the fisheries scientist may conclude that growth of age-0 walleye was fastest in Lake Mitchell.

Box 9.3 Testing for Differences among Length-Frequency Distributions by Means of the Kruskal–Wallis test

A Kruskal–Wallis test was applied to the same walleye data used in Boxes 9.1 and 9.2. The objective of this analysis was to test whether length-frequency distributions were different among samples (i.e., does one population tend to yield larger or smaller values than the other) based on rank scores. The null hypothesis was that there was no difference among the length-frequency distributions. The Kruskal–Wallis test is an extension of Wilcoxon's rank-sum test for two samples. Results of the Kruskal–Wallis and Wilcoxon's rank-sum tests are provided in the output through execution of the NPAR1WAY procedure in SAS (SAS 1999). The input data are the same as used in Box 9.1 and presented in Figure 9.3.

Program

```
DATA ONE;
INPUT LAKE $ LENGTH;
CARDS;
[See data input in Box 9.1]
;
DATA TWO; SET ONE;
PROC SORT;
BY LAKE LENGTH;
RUN;
PROC NPAR1WAY;
CLASS LAKE;
VAR LENGTH;
RUN;
```

Output

Table Wilcoxon scores (rank sums) for the variable length classified by the variable lake.

Wilcoxon Scores					
Lake	N	Sum of scores	Expected under H_0	SD under H_0	Mean score
Island	104	12969.00	18772.00	894.755226	124.701923
Mitchell	111	29767.50	20035.50	911.651297	268.175676
Thompson	145	22243.50	26172.50	968.212806	153.403448

Kruskal–Wallis Test	
Chi-square	118.5671
df	2
P > chi-square	<0.0001

Results

The output indicates that there is a significant ($P < 0.0001$) difference among the three length-frequency distributions, and thus, the null hypothesis is rejected. The mean ranks for Island Lake (124.7), Lake Mitchell (268.2), and Lake Thompson (153.4) are provided in the output under the mean score column.

By default, the NPAR1WAY procedure in SAS provides approximated P-values based on asymptotic methods (SAS 1999). Exact P-values can be calculated by using the EXACT statement in the NPAR1WAY procedure. Asymptotic methods may not be valid when sample sizes are very small and when data are sparse, skewed, or heavily tied (SAS 1999). When sample sizes are large, asymptotic P-values approach exact P-values. The EXACT statement in SAS can be computationally time-consuming depending on the sample size and the number of groups. Exact P-values for this example can be obtained by using the following code.

```
PROC NPAR1WAY;
CLASS LAKE;
VAR LENGTH;
EXACT;
RUN;
```

Program

In SAS, the Kruskal–Wallis test can also be performed by using a combination of the RANK and GLM procedures (SAS 1990). The overall *F*-test is asymptotically equivalent to the Kruskal–Wallis test in SAS. The program below will perform an ANOVA based on ranked data.

```
PROC RANK OUT=RANKS;
RANKS RLENGTH;
VAR LENGTH;
RUN;
PROC GLM DATA=RANKS;
CLASS LAKE;
MODEL RLENGTH=LAKE;
RUN;
```

Output

Table The GLM procedure for the dependent variable RLENGTH, the rank for the variable length. Abbreviations are as follows: mean square error (MSE); coefficient of variation (CV); and sum of squares (SS).

Analysis of Variance					
Source	*df*	Sum of squares	Mean square	*F*-value	$P > F$
Model	2	1283518.268	641759.134	88.03	<0.0001
Error	357	2602744.232	7290.600		
Corrected total	359	3886262.500			
R^2	0.330271	Root MSE	85.38501		
CV	47.30472	RLENGTH mean	180.5000		

Source	*df*	Type I SS	Mean square	*F*-value	$P > F$
Lake	2	1283518.268	641759.134	88.03	<0.0001

Source	*df*	Type III SS	Mean square	*F*-value	$P > F$
Lake	2	1283518.268	641759.134	88.03	<0.0001

In this ANOVA, note that $(n - 1)R^2 = 118.57$ and is the same as the chi-square statistic provided for the Kruskal–Wallis test (SAS 1999). There is a significant ($P < 0.0001$) difference in mean ranked length among the three lakes, leading to the rejection of the null hypothesis.

Box 9.4 Performing Multiple Comparisons of Length-Frequency Data

In Box 9.3, three length-frequency distributions were compared using the Kruskal–Wallis test. The distributions were found to be significantly different. Once the null hypothesis is rejected, the fisheries scientist usually will want to determine between which of the samples the significant differences exist. An example of a nonparametric multiple-comparison test (Zar 1996) based on the walleye data presented in Box 9.3 is illustrated below. This particular multiple-comparison test is appropriate in the case of several tied ranks and unequal sample sizes, which are typical characteristics of length-frequency data, especially when fish are measured and reported to the nearest length-group (e.g., centimeters). However, several other multiple-comparison tests are available, depending on the characteristics of the data being analyzed (Conover 1980; Zar 1996). Information in the summary table below can be found in the SAS output for the Kruskal–Wallis test in Box 9.3.

Parameter	Island	Thompson	Mitchell
Mean rank ($\bar{R}$)	124.70	153.40	268.18
Sample size (*n*)	104	145	111

After the entire data set was rank ordered, the number of groups (lengths) with tied ranks (*m*) was determined to be 69. Next, calculate *T*, the tied-rank statistic,

$$T = \sum_{i=1}^{m} (t_i^3 - t_i) ,$$

where t = the frequency of observations with tied ranks in the *i*th group (length). For example, if in a data set there were two groups (lengths) with tied ranks (three 247-mm fish and two 248-mm fish), T would equal $(3^3 - 3) + (2^3 - 2) = 30$. For the walleye example used in this box, there were many ties, and $T = 20{,}490$.

Next, SEs are calculated for each comparison. The SE for the comparison of Lake Mitchell with Island Lake is calculated as

$$\mathrm{SE} = \sqrt{\left(\frac{N(N+1)}{12} - \frac{T}{12(N-1)}\right)\left(\frac{1}{n_{\mathrm{Mitchell}}} + \frac{1}{n_{\mathrm{Island}}}\right)}$$

$$= \sqrt{\left(\frac{360(361)}{12} - \frac{20{,}490}{12(359)}\right)\left(\frac{1}{111} + \frac{1}{104}\right)} = 14.20$$

For the comparison of Lake Mitchell and Lake Thompson, the SE = 13.12.
For the comparison of Lake Thompson and Lake Island Lake, the SE = 13.37.

The test statistic (*Q*) for each comparison is calculated as the difference in mean ranks divided by the associated SE. Critical values for *Q* can be obtained from tables for nonparametric multiple comparisons (e.g., Zar 1996). In this example, the critical value of *Q* at $\alpha = 0.05$ for three samples is 2.394. For each comparison, the null hypothesis of no difference between length-frequency distributions is rejected if the calculated *Q* exceeds the critical value of *Q*.

Table Nonparametric multiple comparison among three lakes of length-frequency distributions of walleye.

Comparison	$\bar{R}_x - \bar{R}_y$	SE	Q	$Q_{0.05,3}$	Conclusion
Mitchell and Island	268.18 – 124.70 = 143.48	14.20	10.10	2.394	Reject H_0
Mitchell and Thompson	268.18 – 153.40 = 114.78	13.12	8.75	2.394	Reject H_0
Thompson and Island	153.40 – 124.70 = 28.70	13.37	2.14	2.394	Accept H_0

The fisheries scientist can conclude that the length-frequency distribution from Lake Mitchell is significantly greater than that of Island Lake and Lake Thompson. Length-frequency distributions were not significantly different between Lake Thompson and Island Lake.

Program

Currently, nonparametric multiple-comparison procedures are not available in SAS. However, some of the required calculations such as a table of ranked lengths, a table of the frequency of observations with tied ranks in the *i*th group, and *T* can be obtained by invoking the following SAS program.

```
PROC RANK OUT=RANKS;
RANKS RLENGTH;
VAR LENGTH;
RUN;
PROC FREQ; TABLES RLENGTH/OUT=FRANK;
RUN;
PROC PRINT DATA=FRANK;
RUN;
DATA CALCT; SET FRANK;
IF COUNT=1 THEN DELETE;
T=((COUNT*COUNT*COUNT)-COUNT);
PROC PRINT;
RUN;
PROC MEANS SUM;
VAR T;
RUN;
```

Multiple comparisons can also be accomplished on the ranked data in the GLM procedure in SAS. The following SAS program performs an ANOVA on the ranked data (see Box 9.3) and uses a Tukey's multiple-range test to determine differences among the mean ranks.

```
PROC RANK OUT=RANKS;
RANKS RLENGTH;
VAR LENGTH;
RUN;
PROC GLM DATA=RANKS;
CLASS LAKE;
MODEL RLENGTH=LAKE;
MEANS LAKE/TUKEY;
RUN;
```

(Box continues)

Box 9.4 *(continued)*

Output

The ANOVA output for this analysis is shown in Box 9.3. The results of the Tukey's multiple-comparison test are shown below.

Table Tukey's studentized range (HSD) test for RLENGTH (the rank for variable length). This test controls the type I experimentwise error rate. Comparisons significant at the 0.05 level are indicated by ***.

Test Statistics	
Alpha	0.05
Error df	357
Error mean square	7290.6
Critical value of studentized range	3.32840

Means Comparisons

Lake comparison	Difference between means	Simultaneous 95% confidence limits	
Mitchell–Thompson	114.77	89.43	140.12***
Mitchell–Island	143.47	116.05	170.90***
Thompson–Mitchell	–114.77	–140.12	–89.43***
Thompson–Island	28.70	2.88	54.52***
Island–Mitchell	–143.47	–170.90	–116.05***
Island–Thompson	–28.70	–54.52	–2.88***

Results

These results show that there is a significant difference in mean ranked length among each of the three lakes.

In these examples, the multiple-comparison-testing methods had different results. The nonparametric multiple-range test was more conservative that the Tukey test. This clearly demonstrates that different multiple-comparison tests can provide different results. The choice of a multiple-comparison test should be made before the analysis is conducted rather than by searching for significance by performing multiple tests.

Box 9.5 Using Contingency Tables to Test for Differences in Length-Frequency Distributions

The chi-square test is commonly used to test for differences in length-frequency distributions. In this example, DC electrofishing at night was used to collect bluegill in 1996, 1998, and 2000 from a private pond in Connecticut. Bluegills were classified into two length-groups: stock to quality length (80–149 mm) and quality length (≥150 mm). Proportional stock density (see section 9.3) was also calculated for each year. The objective of this analysis was to determine whether length-frequency distributions (summarized by PSD values) were different among years. The chi-square analysis was performed using the frequency procedure (FREQ) in SAS (SAS 1999). The null hypothesis is that the frequency of observations among length-groups (stock to quality length and quality length) is independent of year.

Data

Table The number of bluegill collected in each length-group and proportional stock density (PSD).

	Year		
Size category and length index	1996	1998	2000
Stock to quality length (80–149 mm)	77	124	251
Quality length (≥150 mm)	85	44	34
Total stock length (≥80 mm)	162	168	285
PSD	52	26	12

Program

In the following SAS program, LCAT is the length category (S–Q = stock to quality length and Q = greater than or equal to quality length) and NUM is the number of fish.

```
DATA ONE;
INPUT YEAR LCAT $ NUM;
CARDS;
1996 S-Q   77
1996   Q   85
1998 S-Q  124
1998   Q   44
2000 S-Q  251
2000   Q   34
;
PROC SORT;
BY YEAR LCAT NUM;
RUN;
DATA TWO;
SET ONE;
BY YEAR LCAT NUM;
IF FIRST.LCAT THEN DO;
DO I = 1 TO NUM;
LCAT = LCAT;
YEAR = YEAR;
OUTPUT;
END;
END;
RUN;
PROC FREQ;
TABLES YEAR*LCAT / CHISQ;
RUN;
```

(Box continues)

Box 9.5 *(continued)*

Output

Table Summary statistics for chi-square analysis of length category (LCAT) by year. Sample size is 615.

	Length category		
Year and measure	Q	S–Q	Total
1996			
Frequency	85	77	162
Percent	13.82	12.52	26.34
Row %	52.47	47.53	
Column %	52.15	17.04	
1998			
Frequency	44	124	168
Percent	7.15	20.16	27.32
Row %	26.19	73.81	
Column %	26.99	27.43	
2000			
Frequency	34	251	285
Percent	5.53	40.81	46.34
Row %	11.93	88.07	
Column %	20.86	55.53	
Total			
Frequency	163	452	615
Percent	26.50	73.50	100.00

Box 9.1 an *F*-test with 357 error degrees of freedom results in a high level of power to detect differences among length-frequency distributions. Similarly, Kolmogorov–Smirnov two-sample tests are often highly significant when sample sizes are large, even though the distributions can appear similar. Caution should be applied when using individual fish as experimental units, resulting in very high sample sizes.

An alternative approach to comparing length frequencies would be to treat each group of fish caught in a unit of effort (e.g., trap net or electrofishing station) as a sample. In other words, each unit of effort would be considered a sample or "collection event," and individual fish would be considered subsamples. Consider sampling black crappies with 20 trap nets during a single sampling period in a reservoir. If the 20 nets are set according to a particular sampling design, then each net (location) may be adequate to use as an independent experimental unit. Examples of using units of effort as samples to compare size structure are provided in Boxes 9.6 and 9.7.

Table Chi-square statistics of length category by year.

Statistic	*df*	Value	*P*
Chi-square	2	87.1540	<0.0001
Likelihood ratio chi-square	2	85.5173	<0.0001
Mantel–Haenszel chi-square	1	84.8020	<0.0001
Phi coefficient		0.3764	
Contingency coefficient		0.3523	
Cramer's V		0.3764	

Results

According to the chi-square test, there is a significant ($\chi^2 = 87.15$, $P < 0.0001$) difference in the frequency of observations between length-groups, leading to the rejection of the null hypothesis. To test which years were significantly different from each other, a chi-square test was performed for each combination of years (i.e., 1996 and 1998, 1998 and 2000, and 1996 and 2000) based on 2 × 2 contingency tables. Although the results for each comparison are not shown, all pairwise tests showed significant differences ($P < 0.0001$) between years. To maintain an experimentwise error rate of $\alpha = 0.05$, the significance level for each comparison ($P = 0.017$) was established by dividing α by the number of comparisons (3). Size structure declined from 1996 (PSD = 52) to 1998 (PSD = 26) to 2000 (PSD = 12).

In this pond, the decrease in PSD of bluegills over the 3 years was probably due to the reduction in density of chain pickerel in the pond. Mean *C/f* (number/h electrofishing) of chain pickerel declined from 82 in 1996 to 39 in 2000. Declines in chain pickerel abundance probably lead to reduced predation on bluegills, resulting in higher abundance and reduced growth of bluegills.

To test for differences in length-frequency data that are summarized using stock density indices other than PSD (e.g., relative stock density preferred length [RSD-P] or quality-to-preferred length [RSD-Q-P]), simply change the length categories in the analysis. For example, to test for differences in length frequency summarized as RSD-P, test for differences in the frequency of occurrence of stock-to-preferred-length fish and preferred-length fish among treatments.

9.4.4 Analysis of Repeated Measures

Fisheries scientists frequently assess changes in size structure on one population through time (e.g., across years). One consideration is that many of the statistical procedures mentioned above (e.g., chi-square test and Kruskal–Wallis) assume the samples are independent. For example, in Box 9.5 bluegill PSD was tested using samples collected at 2-year intervals, and the chi-square test assumes that those samples are independent. Because samples were at 2-year intervals, this assumption may be realistic. However, samples collected over a number of consecutive years are likely not independent (Maceina et al. 1994) because catch rates or size structure in 1 year may influence the size structure in subsequent years (i.e., the same year-classes are sampled over time).

Repeated-measures ANOVA provides a series of models that incorporate time dependency of the data into the analysis (see also Chapter 7 for discussion of

Box 9.6 Testing for Differences in Size Structure by Treating Groups of Fish Caught in Each Unit of Effort as Samples

Fisheries scientists oftentimes evaluate the effectiveness of alternative sampling methods. However, before alternative sampling methods are implemented into standard sampling programs, the fisheries scientist should understand how data (e.g., size structure) collected by the new sampling method compares to the method currently used. For example, the use of angler-collected data in research and monitoring is becoming more popular due to reliability and reduced costs and effort associated with data collection compared with more traditional methods such as electrofishing.

In this example, size structure of largemouth bass obtained from two sampling methods is compared. Largemouth bass were sampled from Mansfield Hollow Reservoir, Connecticut, in spring 2002. Twelve stations along the lake perimeter were sampled at night by means of DC electrofishing, and size structure data were collected at 12 bass fishing tournaments over the same time period. Individual fish were measured to the nearest centimeter total length at the end of each electrofishing station and fishing tournament.

Catches from each electrofishing station and fishing tournament were considered independent samples. Electrofishing stations did not overlap, and catches in one tournament were considered independent of the others. The null hypothesis tested is that the ratio of the number of preferred-length (i.e., ≥38 cm) fish to the number of quality-length (i.e., ≥30 cm) fish was not different between the two sampling methods.

Data

Table Largemouth bass data from Mansfield Hollow Reservoir, Connecticut, in spring 2002. Catches from each electrofishing station and fishing tournament were considered independent samples.

	Number of fish			Number of fish	
Fishing tournament	≥30 cm	≥38 cm	Electrofishing station	≥30 cm	≥38 cm
1	3	2	1	23	12
2	28	4	2	22	2
3	13	1	3	35	8
4	8	0	4	6	2
5	61	16	5	11	1
6	76	12	6	15	7
7	38	10	7	12	3
8	49	12	8	9	6
9	62	24	9	25	5
10	43	10	10	25	8
11	59	18	11	9	1
12	24	5	12	7	1

Program

Electrofishing (ELEC) and fishing tournaments (TOURN) are the sampling methods used, and QUAL and PREF are the number of fish collected in each length category for each electrofishing station and fishing tournament. The variable LOGIT was created, which is the ratio of the number of preferred-length (≥38 cm) fish to the number of quality-length (≥30 cm) fish in each sample, after a value of 0.5 was added to QUAL and PREF to remove zeros prior to log transformation. From a parametric statistics standpoint, using LOGIT has an advantage over using a proportion because it can exceed one and is more likely to be normally distributed.

The GLM procedure (SAS 1999) was used to conduct a *t*-test to determine whether there was a significant difference in mean LOGIT between the two sampling methods. The WEIGHT statement weights each sample based on the number of fish collected in each sample.

```
DATA BASS;
INPUT METHOD $ QUAL PREF;
CARDS;
ELEC    23 12
ELEC    22  2
ELEC    35  8
ELEC     6  2
[Data input continued]
TOURN   28  4
TOURN   13  1
TOURN    8  0
TOURN   61 16
[Data input continued]
;
DATA BASS2; SET BASS;
LOGIT=LOG((PREF+0.5)/(QUAL+0.5));
PROC PRINT;
PROC SORT; BY METHOD;
PROC MEANS; BY METHOD; VAR LOGIT;
WEIGHT QUAL;
PROC GLM;
CLASS METHOD;
MODEL LOGIT=METHOD;
WEIGHT QUAL;
RUN;
```

Output

Table The number of fish collected in each length category for each sampling method. The variable LOGIT is the ratio of the number of preferred-length (≥38 cm) fish to the number of quality-length (≥30 cm) fish in each sample, after a value of 0.5 was added to QUAL and PREF.

	Number of fish		
Method and observation	QUAL	PREF	LOGIT
TOURN			
1	3	2	–0.33647
2	28	4	–1.84583
3	13	1	–2.19722
4	8	0	–2.83321
5	61	16	–1.31568
6	76	12	–1.81156
7	38	10	–1.29928
8	49	12	–1.37624
9	62	24	–0.93649
10	43	10	–1.42139
11	59	18	–1.16821
12	24	5	–1.49393

(Box continues)

Box 9.6 *(continued)*

Method and observation	Number of fish QUAL	PREF	LOGIT
ELEC			
13	23	12	−0.63127
14	22	2	−2.19722
15	35	8	−1.42947
16	6	2	−0.95551
17	11	1	−2.03688
18	15	7	−0.72594
19	12	3	−1.27297
20	9	6	−0.37949
21	25	5	−1.53393
22	25	8	−1.09861
23	9	1	−1.84583
24	7	1	−1.60944

Table Summary statistics (MEANS procedure) based on LOGIT values for two sampling methods.

Method	*N*	Mean	SD	Minimum	Maximum
ELEC	12	−1.3281428	2.2037466	−2.1972246	−0.3794896
TOURN	12	−1.4280742	2.4069863	−2.8332133	−0.3364722

Table Result of GLM procedure to compare mean LOGIT (*t*-test) between the two sampling methods. The WEIGHT statement weights each sample based on the total number of fish (QUAL) collected in each sample. Analysis is based on 24 observations.

Class Level Information

Class	Levels	Values
METHOD	2	ELEC TOURN

GLM Procedure

Source	*df*	Sum of squares	Mean square	*F*-value	*P* > *F*
Model	1	1.3907895	1.3907895	0.26	0.6144
Error	22	117.1509050	5.3250411		
Corrected total	23	118.5416945			
R^2	0.011732	Root MSE	2.307605		
CV	−165.0553	LOGIT mean	−1.398080		

Results

At the top of the output, LOGIT values for each sample are provided by the PROC PRINT statement. Sample LOGIT values are followed by the mean LOGIT for electrofishing (mean LOGIT = −1.33) and fishing tournaments (mean LOGIT = −1.43) weighted by METHOD based on the number of fish in each sample (QUAL). By calculating the inverse $\log_e$ of these values, the mean ratio of the number of preferred-length (i.e., ≥38 cm) fish to the number of quality-length (i.e., ≥30 cm) fish was 0.26 for electrofishing and 0.24 for fishing tournaments. The output for the GLM procedure indicates that there was not a significant difference (F = 0.26; P = 0.6144) in mean LOGIT between electrofishing and fishing tournaments. The fisheries scientist fails to reject the null hypothesis that the ratio of the number of preferred-length (i.e., ≥38 cm) fish to the number of quality-length (i.e., ≥30 cm) fish was the same between the two sampling methods.

Box 9.7 Using Repeated-Measures ANOVA to Test for Size Structure Differences with Time-Dependent Data

Repeated-measures ANOVA is commonly used to test for differences in a population response when samples are not independent, often because they are collected through time. In this example, we assessed differences in the size structure of a largemouth bass population at Lake Jackson, Florida, after implementation of a 330–431-mm protected-slot-length limit. Data were collected by Florida Fish and Wildlife Conservation Commission biologists. The population had no size limit prior to 1991, and the slot-length limit was enacted in July of 1990. Daytime electrofishing samples were collected at 12 fixed sites during April from 1988 to 1996. Fixed sites in the analysis were treated as subjects sampled through time. The size distribution of largemouth bass is likely to be dependent through time (i.e., size structure of fish present in previous sampling influences size structure at later time intervals). Thus, the analysis should consider that the size structure of fish at a given site is not independent through time. In this example, we assessed whether the size structure of largemouth bass differed before the slot-length limit ($N = 3$ years of data) compared with after the slot limit ($N = 6$ years of data). Largemouth bass were classified into three groups (based on total length): below 200 mm, 200–329 mm, and 330 mm and larger. The objective of this analysis was to test whether the ratio of fish 330 mm and larger to fish between 200 and 329 mm differed before and after the slot-length limit was enacted. Fish below 200 mm were removed from the analysis because the slot limit was not expected to influence abundance of fish below 200 mm. The null hypothesis is that the ratio of fish 330 mm and larger to subslot-size fish (i.e., fish > 200 mm but less than 330 mm) was not different before and after the slot limit was enacted. The mixed-models procedure (PROC MIXED; SAS 1999) was used to conduct the test.

Data: Part I

In the data table below, COUNT is the number of fish in each size-group, and size-groups are given as UND (200–329 mm) and SLOT (330 mm and longer). A SITE was included if a fish was collected in at least one size-group. However, sites that did not contain fish in either size-group were removed from the analysis because collection of no fish provides no information about size structure (i.e., if both the UND and SLOT size groups had a COUNT of zero, the site was not included in the analysis).

Table Size-group data for largemouth bass fishery in Lake Jackson, Florida, before and after slot-length limit implementation.

Year and site	Size-group	Count
1988		
1	UND	4
1	SLOT	1
2	UND	9
2	SLOT	2
[Data continued]		
1996		
11	UND	4
11	SLOT	4
12	UND	9
12	SLOT	4
[Data continued]		

Program: Part I

In the following SAS program, the data were rearranged using PROC TRANSPOSE prior to creating the dependent variable for the test. This procedure changes columns to rows or rows to columns. In this example the column COUNT was changed to rows for both size-groups.

(Box continues)

Box 9.7 *(continued)*

```
DATA A;
INPUT YEAR SITE    SIZEGRP $ COUNT;
CARDS;
1988        1        bund       4
1988        1        slot       1
1988        2        bund       9
1988        2        slot       2
[Data input continued]
;
DATA B; SET A;
IF YEAR LE 1990 THEN PERIOD = 'APRE';
IF YEAR GT 1990 THEN PERIOD = 'BPOST';
RUN;

PROC SORT;
BY PERIOD YEAR SITE;
PROC TRANSPOSE OUT=C;
BY PERIOD YEAR SITE;
VAR COUNT;

DATA D; SET C;
RENAME COL2 = UNDER;
RENAME COL3 = SLOT;
DATA E; SET D;
UNDERT=UNDER+0.5;
SLOTT = SLOT+0.5;
TOTAL =UNDERT+SLOTT;
LOGIT=LOG(SLOTT/(UNDERT));
```

In step DATA E, 0.5 was added to each count to remove zeros prior to the log transformation.

The variable TOTAL is used to weight each transect in PROC MIXED below. Remember that fish shorter than 200 mm were removed, and the ratio of the number of fish greater than or equal to 330 mm to the number of fish between 200 and 329 mm was used.

The variable LOGIT is the log of the ratio of fish greater than or equal to 330 mm relative to fish between 200 and 329 mm. It was predicted that after the slot limit is in place (i.e., anglers cannot keep fish between 330 and 431 mm) the ratio, and thus, the LOGIT, would increase.

```
PROC PRINT;
VAR YEAR PERIOD SITE UNDER UNDERT SLOT SLOTT LOGIT;
RUN;
```

Data: Part II
From the above program, the final data set was created prior to analysis.

Table Final data set for analysis of largemouth bass fishery before and after slot-length limit implementation. Size-groups are UNDER (200–329 mm), SLOT (330 mm and longer), and those two categories transformed (UNDERT and SLOTT).

Period, year, and observation	Site	UNDER	UNDERT	SLOT	SLOTT	LOGIT
Pre-slot limit						
1988						
1	1	4	4.5	1	1.5	–1.09861
2	2	9	9.5	2	2.5	–1.33500
3	3	11	11.5	1	1.5	–2.03688

[Data continued]

Period, year, and observation	Site	UNDER	UNDERT	SLOT	SLOTT	LOGIT
Post-slot limit						
1996						
1	10	4	4.5	2	2.5	–0.58779
2	11	4	4.5	4	4.5	0.00000
3	12	9	9.5	4	4.5	–0.74721
[Data continued]						

Program: Part II

```
PROC UNIVARIATE PLOT NORMAL;
BY PERIOD;
VAR LOGIT;
```

The Wilk's lambda in PROC UNIVARIATE, specified with the NORMAL option, was used to assess whether the dependent variable (LOGIT) was normally distributed for each PERIOD (pre- versus post-slot limit years). In this case the assumptions of normality were met ($P > 0.05$ for both PERIODS).

```
PROC MIXED;
CLASS PERIOD YEAR SITE;
MODEL LOGIT=PERIOD;
WEIGHT TOTAL;
RANDOM YEAR(PERIOD);
REPEATED YEAR/SUBJECT=SITE TYPE=AR(1);
LSMEANS PERIOD/PDIFF;
RUN;
```

The model tests whether the mean LOGIT differs significantly between periods (pre- versus post-slot limit regulation). The WEIGHT statement weights each site based on the number of fish collected (i.e., sites with large catch influence the test proportionally more). The RANDOM statement assumes that among-year variation within each PERIOD was random. The REPEATED statement indicates the consecutive years of sampling (i.e., time variable), and the SUBJECT statement assigns each site as an individual station sampled through time. In this case, sites were not chosen at random so sites are treated as a fixed effect in the model. In SAS, the TYPE statement allows the researcher to investigate various covariance structures to model the time-dependence of the data (discussed below). The LSMEANS statement is a means separation option that will give the overall least-squares means for each period and their significance level (PDIFF).

Output

Table The following output is from the PROC MIXED statement presented above that tests whether the mean LOGIT differs significantly between periods (PRE = pre-slot limit and POS = post-slot limit).

Model Information	
Data Set	WORK.E
Dependent variable	LOGIT
Weight variable	TOTAL
Covariance structures	Variance components, Autoregressive
Subject effect	SITE
Estimation method	REML
Residual variance method	Profile

(Box continues)

Box 9.7 *(continued)*

Fixed effects SE method	Model-based
Degrees of freedom method	Containment

Class Level Information

Class	Levels	Values
PERIOD	2	PRE POS
YEAR	9	1988 1989 1990 1991 1992 1993 1994 1995 1996
SITE	12	1 2 3 4 5 6 7 8 9 10 11 12

Dimensions

Covariance parameters	3
Columns in X	3
Columns in Z	9
Subjects	1
Maximum observations per subject	101
Observations used	101
Observations not used	0
Total observations	101

Iteration History

Iteration	Evaluations	–2Residual log likelihood	Criterion[a]
0	1	269.51966315	
1	2	268.00863114	0.00000795
2	1	268.00828432	0.00000000

Covariance Parameter Estimates

Covariance parameter	Subject	Estimate
Year (period)		0.04188
AR(1)	Transect	0.09518
Residual		7.7798

Fit Statistics

–2Residual log likelihood	268.0
AIC (smaller is better)	274.0
AICc (smaller is better)	274.3
BIC (smaller is better)	274.6

Type III Test of Fixed Effects

Effect	Numerator df	Denominator df	*F*-value	*P* > *F*
PERIOD	1	7	4.60	0.0691

Least Squares Means

Effect	Period	Estimate	Standard error	*df*	*t*-value	*P* > |*t*|

PERIOD	PRE	–0.8506	0.1829	7	-4.65	0.0023
PERIOD	POS	–0.3602	0.1396	7	-2.58	0.0365

Differences of Least Squares Means							
Effect	Period	Period	Estimate	SE	*df*	*t*-value	*P* > |*t*|
PERIOD	PRE	POS	–0.4905	0.2286	7	–2.15	0.0691

[a] The convergence criteria were met.

Results

The "Model Information" and "Class Level Information" output show the model configuration and levels of each class going into the model. The "Iteration History" reveals whether the model converged on a solution. The series of "Fit Statistics" allows one to compare various covariance matrix structures to one's data. The Akaike's Information Criteria (AIC), small sample corrected AIC (AICc), and Bayesian Information Criteria (BIC) are model fit statistics commonly used for many ecological modeling applications (see Guthery et al. [2005] for a review and critique). In this example, the AIC statistic is used to assess how well the time-dependent structure of one's data fit the chosen TYPE covariance structure specified in the model (SAS 1999). In this case, TYPE = AR(1) was used, which is the first-order autoregressive structure. The AR(1) structure models correlations between time periods that are linear and decline with the distance in time that observations are made (e.g., assumes years 1 and 2 are more closely related than years 1 and 4; Littell et al. 1996). Littell et al. (1996) described various options for covariance structures in PROC MIXED, and the investigator can choose the structure type with the lowest AIC score (i.e., lowest deviance between the data and the specified structure type). The AR(1) structure is one option for data sampled at regular time intervals, which in this example was appropriate because electrofishing occurred in April of each year. The AR(1) model also obtained the lowest AIC score of several covariance structures considered.

However, we note that the time dependency of the data were not strong based on covariance parameter estimates of 0.095 for the AR(1) variable relative to a high residual value (7.78). Analyses showing strong time dependence typically exhibit covariance parameter estimates of equal or greater magnitude compared with the residual values (authors, personal observation). The lack of a relationship between size structure data in successive years is not atypical given variation around electrofishing data, and would be grounds to ignore time dependency and use a regular one-way ANOVA to test for differences in LOGIT between PERIODS. Thus, samples collected in successive years do not automatically require repeated-measures analyses! Here we'll continue with the output interpretation as an example of the analysis.

Results of this analysis showed that the LOGIT approached significance between pre- and post-size-limit time periods ($P = 0.069$) at an $\alpha = 0.05$. The LSMEANS procedure output the least-squares means of the LOGIT (–0.85 and –0.36). By taking the inverse $\log_e$ of these values, we find that the ratio of fish 330 mm and larger to fish between 200 and 329 mm averaged 0.427 before the slot limit and 0.698 after the slot limit was enacted. Thus, at an α level of 0.10, the ratio increased after the slot was enacted, suggesting that the size structure increased. It is important to note that although there is a significant difference, variables other than the slot limit (e.g., strong year-classes or changes in large fish catchability) could have also influenced the result. This example shows how time dependency in the data can be included in assessment of fish size structure.

repeated-measures data). Maceina et al. (1994) described how a split-plot ANOVA could be used to conduct repeated-measures tests. More recently, mixed-model ANOVA provides multiple options to handle repeated-measures data. The advantage of mixed models over split-plot analyses is that the split-plot ANOVA assumes compound symmetry (Littell et al. 1996). Compound symmetry is defined as constant dependence; in other words, each time period is assumed to be equally related to all other time periods. Mixed-model ANOVA allows the investigator to specify covariance matrix structures other than compound symmetry (Littell et al. 1996). For example, you might expect samples collected in consecutive years to be more highly related than are samples collected 5 years apart. Box 9.7 provides an example of using a repeated-measures ANOVA to test for size structure differences based on time-dependent data. When assessing population size structure on one population through time, or across multiple populations sampled through time, use of repeated-measures designs is recommended.

9.5 INTERPRETATION OF SIZE STRUCTURE

9.5.1 Length-Frequency Distributions

Length-frequency distributions reflect an interaction of the rates of recruitment, growth, and mortality of a fish population. Length-frequency data can provide insight into the dynamics of fish populations and identify problems such as inconsistent year-class strength, slow growth, and excessive mortality (Anderson and Neumann 1996). In most instances, a thorough assessment of a fish population requires other population assessment tools, such as *C/f,* age and growth, or body condition, in addition to length-frequency data.

Length-frequency data for black crappie collected with a trawl from two Florida lakes are presented in Figure 9.5. Based on the length-frequency distribution for Lake Jackson, a fisheries scientist may conclude that the black crappie population is balanced. A balanced population is one that has moderate rates of recruitment, growth, and mortality compared with what is expected for populations in the same geographic region. A length-frequency histogram from a balanced fish population will have a stable decline from the shorter to longer lengths, reflecting a stable age structure produced by consistent recruitment and consistent, moderate rates of mortality among successive age-classes. In exploited populations, the term balance has also been referred to a population that produces sustainable yields of harvestable-size fish. However, balanced populations can also occur in unexploited water bodies (Anderson and Neumann 1996).

The length-frequency histogram for Alligator Lake (Figure 9.5) may indicate that this population is unbalanced. The most striking difference between Lake Jackson and Alligator Lake is that Alligator Lake does not show a stable decline in the numbers of fish with increasing length. Instead, the length-frequency histogram is "interrupted" by length-groups with many individuals bounded by length-groups with fewer individuals. If the strong and weak interruptions corresponded

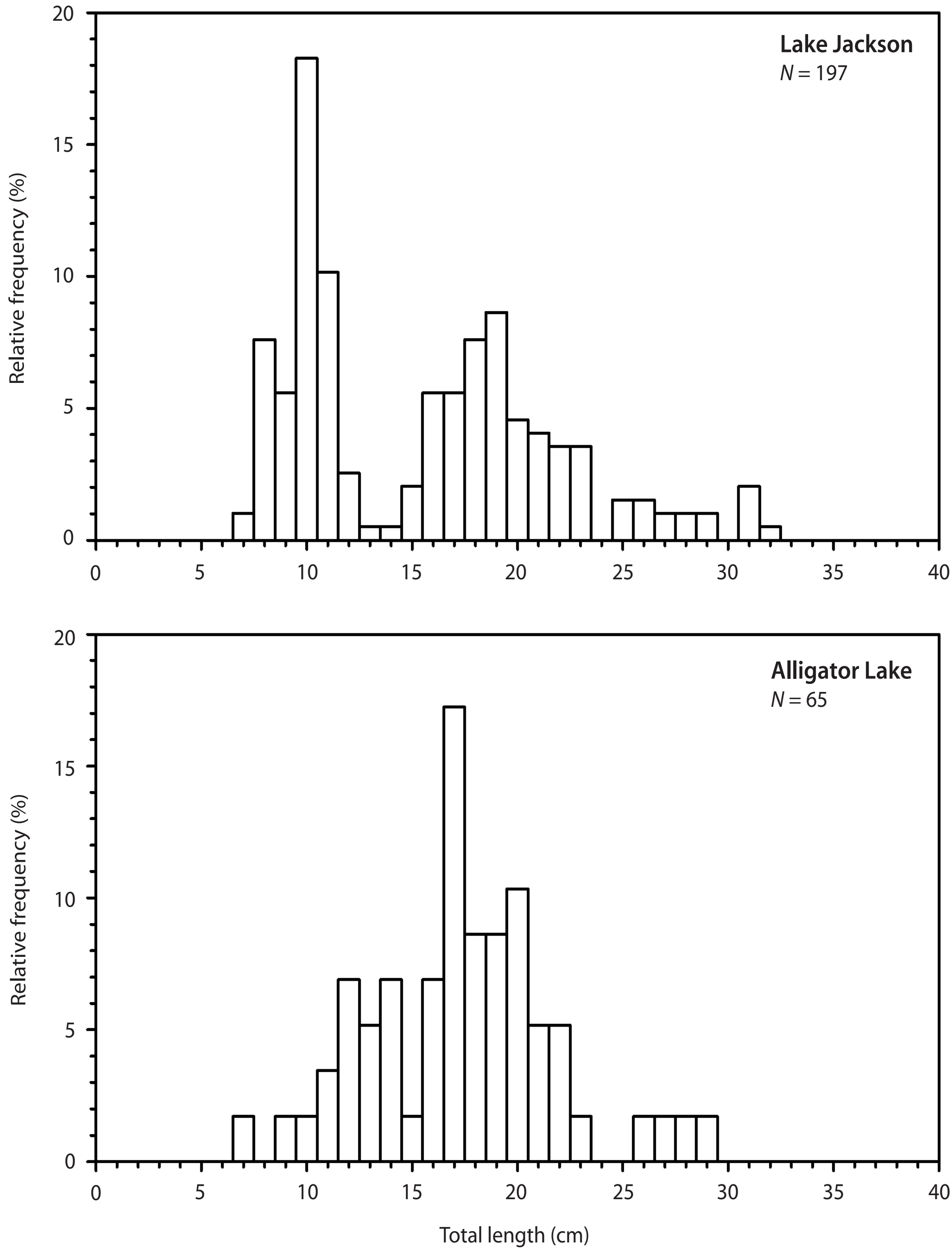

Figure 9.5 Relative-frequency histograms for black crappie from Lake Jackson and Alligator Lake, Florida, collected by means of a trawl. Data were provided by the Florida Fish and Wildlife Conservation Commission.

to age-groups, and if all lengths represented were vulnerable to the sampling gear, then a fisheries scientist might conclude that year-class strength at Alligator Lake is inconsistent compared with Jackson Lake. However, the clearest indication of variable year-class strength would be determined from age-frequency analysis (see Chapter 4).

The length-frequency distribution for largemouth bass collected by means of night electrofishing in a South Dakota pond is presented in Figure 9.6. Note that all largemouth bass sampled were less than quality length; thus PSD = 0. Mortality in this population possibly is high, demonstrated by the lack of largemouth bass greater than quality length. When examining the size structure, a fisheries scientist might arrive at one of several conclusions about the status of this population: (1) low recruitment, slow growth, and moderate to high mortality due to poor habitat; (2) overharvest of largemouth bass greater than quality length; or (3) high density of small, slow-growing largemouth bass due to excessive recruitment. The last condition is often referred to as stunting. In this example, length-frequency information alone could not be interpreted to arrive at the cause for the poor population structure. Other information such as *C/f*, growth, or body condition assessment would be necessary. In Knox Pond, *C/f* was 306 stock-length largemouth

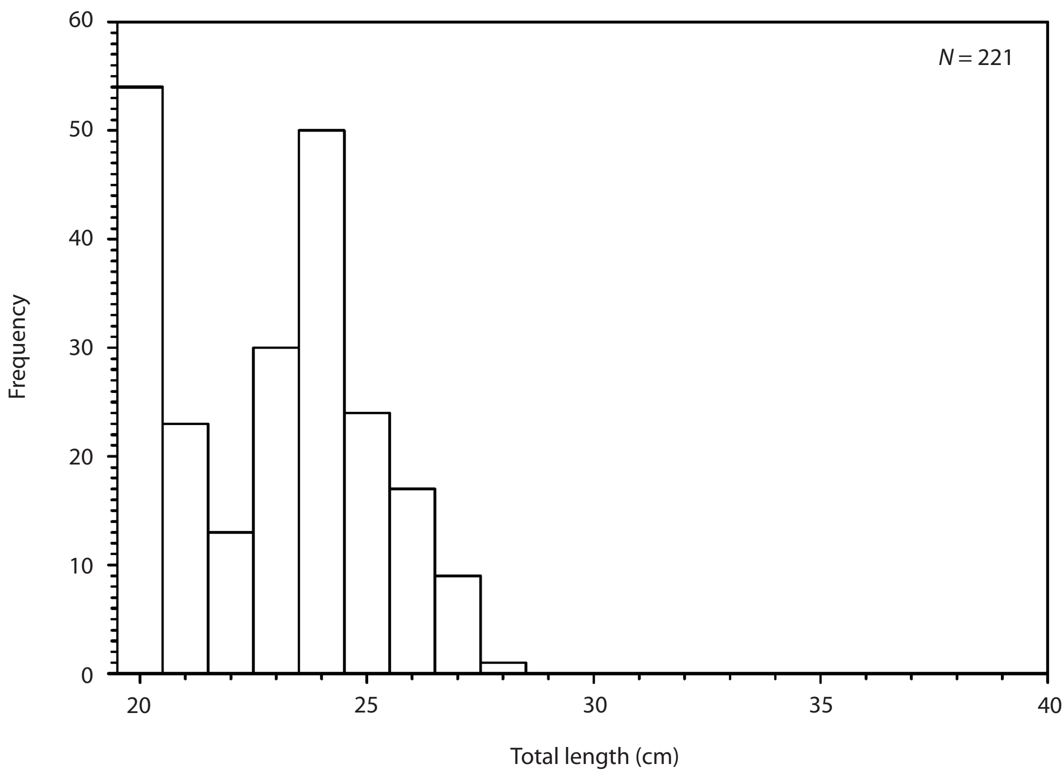

Figure 9.6 Absolute-length-frequency histogram of stock-length (≥20 cm) largemouth bass sampled by electrofishing from Knox Pond, South Dakota.

bass per hour of electrofishing (Neumann et al. 1994), which was high compared with other populations in the state. Mean relative weight (W_r) was 77, and growth rate was well below the state average. Thus, this population represented condition 3 listed above. Condition 1 might be confirmed if *C/f* was low, growth was slow, and poor habitat was documented. Condition 2 might be confirmed if growth was moderate to fast, habitat conditions were favorable, and creel statistics showed a high harvest of quality-length fish. This example also demonstrates the value of statewide or regional summaries of sampling data for comparative purposes.

9.5.2 Stock Density Indices

The use of stock density indices in size structure assessment should be thought of as a complement, and not a replacement, to other methods of length-frequency analysis. Any size structure assessment should begin with a thorough inspection of length-frequency histograms, as they can provide detail that may be lost when length data are summarized in wide length categories or by an index. A benefit of calculating stock density indices is that the index values can be used to test correlations between size structure and other factors. An appropriate question concerning the use of stock density indices is whether the index value (i.e., size structure) reflects density and dynamics of fish populations (Willis et al. 1993). As the density of a population increases, PSD tends to decrease; declines in size structure can be attributed to slowing of growth and increased mortality as resources become scarce. However, a low PSD value may also occur at low population densities due to overharvest or poor habitat. Negative correlations between PSD and density, *C/f,* or biomass have been observed for many species, including largemouth bass (Reynolds and Babb 1978; Gabelhouse 1984a; Boxrucker 1987; Guy and Willis 1990; Saffel et al. 1990; Hill and Willis 1993), black crappie (RSD-P; Guy and Willis 1995), black bullhead (Brown et al. 1999), and brook trout (Johnson et al. 1992). Such negative correlations are more likely in small water bodies with simple fish communities.

As growth increases, there is a tendency for PSD to increase. Low density may result in fast growth, whereas high density may result in slow growth. Correlations between stock density indices and growth have been observed for largemouth bass (Miranda 1983; Jacobs and O'Donnell 1996), smallmouth bass (Jacobs and O'Donnell 1996), bluegill (Novinger and Legler 1978; Paukert and Willis 2000), northern pike (Willis and Scalet 1989), yellow perch (Willis et al. 1991; Paukert and Willis 2000; Paukert et al. 2002), and black crappie (Guy and Willis 1995; Paukert and Willis 2000; Paukert et al. 2002).

Several studies have demonstrated that body condition is positively correlated to growth rate (see Chapter 10). Individuals from low-density populations in which PSD is high tend to have high body condition values, and individuals from high-density populations in which PSD is low tend to have low body condition values. Positive correlations between PSD and W_r for species such as largemouth bass (Wege and Anderson 1978), white crappie and black crappie (Gabelhouse 1984a), northern pike (Willis and Scalet 1989), walleye (Murphy et al. 1990), sauger (Guy et al. 1990), yellow perch (Willis et al. 1991), and brook trout (Johnson et al. 1992) have been observed. However, body condition is an instantaneous measure, and slow-growing fish may exhibit high body condition at times of the year when food is abundant or when gonads are mature during the spawning period.

As total annual mortality increases, there is a tendency for PSD to decrease. In situations in which recruitment is high, as in the Knox Pond example (Figure 9.6), mortality tends to be high and PSD tends to be low. High mortality due to overharvest and poor habitat also results in low PSD values. Negative correlations between PSD and mortality have been observed in largemouth bass (Reynolds and Babb 1978; Miranda 1983; Jacobs and O'Donnell 1996) and smallmouth bass (Jacobs and O'Donnell 1996).

Correlations between stock density indices and density or dynamic rate functions are often moderate in strength, and there is a wide variability in the strength of correlations observed among studies. One reason for this may be that stock density indices may lack sensitivity in some cases; two populations can have the same stock density index value and actually have different length-frequency distributions. Variations in factors such as productivity and growing season can affect establishment of a clear relationship between stock density indices and population parameters (Willis et al. 1993). Additionally, variability in PSD may be related to water body size. For example, largemouth bass in small impoundments may be more recruitment driven than recruitment limited. Jakes (1987) found that size structure of largemouth bass increased in three impoundments ranging in size from 9 to 1,100 ha. Stock density indices also provide more interpretive information when populations are relatively steady state, (i.e., when recruitment, growth, and mortality remain somewhat constant) (Willis et al. 1993). For example, PSD will provide little interpretive information for populations with highly variable recruitment. Willis et al. (1993) provided an example in which the PSD of a black crappie population increased from 3 in spring to 100 in fall. This was the result of a single cohort of black crappies growing over the course of one season. Allen and Pine (2000) found that PSD would often not change significantly in response to minimum length limits if recruitment was highly variable (e.g., coefficients of variation in recruits to age 1 that are greater than 70–90%).

Correlations between predator and prey stock density index values are listed in Table 9.3. Because largemouth bass is a common predator in ponds and small impoundments, most examples listed deal with largemouth bass as the predator, although several examples of prey are listed. In ponds and small impoundments, predator PSD tends to decline as predator density increases. As predator density increases, prey fish density decreases. Thus, prey PSD tends to increase as predator density increases, resulting in an inverse correlation between predator PSD and prey PSD (Willis et al. 1993).

The likelihood of an inverse relationship between predator PSD and prey PSD tends to decline in large water bodies. Carline et al. (1984) suggested that in Ohio impoundments, inverse relationships between size structure of largemouth bass and bluegills may not be expected in impoundments greater than 15 ha in size. In some instances, inverse relationships have been observed in impoundments larger

Table 9.3 Summary of correlation coefficients (*r*) between stock density indices of predator and prey species and other parameters. Parameters compared are proportional stock density (PSD); relative stock density of preferred-length fish (RSD-P); and catch-per-unit-effort (*C/f*).

Predator	Parameter	Prey	Parameter	*r*	Reference
Largemouth bass	PSD	Black bullhead	Mean length	−0.81	Saffel et al. (1990)
	C/f	Bluegill	PSD	0.71	Guy and Willis (1990)
	PSD		PSD	−0.83	Guy and Willis (1990)
	RSD-P		Growth	−0.64	Guy and Willis (1990)
	PSD		PSD	−0.49	Paukert and Willis (2000)
	C/f		PSD	0.52	Paukert and Willis (2000)
	PSD	Crappie[a]	PSD	−0.85	Gabelhouse (1984a)
	RSD-P		PSD	−0.84	Gabelhouse (1984a)
	PSD		*C/f*	0.73	Boxrucker (1987)
	RSD-P		*C/f*	0.88	Boxrucker (1987)
	C/f		PSD	0.56	Boxrucker (1987)
	PSD		PSD	−0.56	Boxrucker (1987)
	C/f	Yellow perch	PSD	0.81	Guy and Willis (1991)
	PSD		PSD	−0.82	Guy and Willis (1991)
	PSD		Growth	−0.95	Guy and Willis (1991)
	C/f		PSD	0.82	Paukert and Willis (2000)
Northern pike	*C/f*	Black bullhead	PSD	−0.54	Brown et al. (1999)

[a] Includes white crappie and black crappie.

than 15 ha (Gabelhouse 1984b; Boxrucker 1987; Guy and Willis 1991; Paukert and Willis 2000; Paukert et al. 2002).

Stock density indices are useful tools not only to report size structure but also to reflect density and population dynamics in certain situations. However, because of the variability in correlations and confounding factors, stock density indices should be used in association with other assessment tools to evaluate fish populations properly.

9.6 CONCLUSIONS

Factors influencing the accuracy and precision of size structure data such as gear selectivity and time of collection should be considered prior to data analysis and interpretation. Standardized sampling that allows for relative comparisons through time or across water bodies provides the most powerful inferences, and planning the study design prior to data collection is imperative. Traditionally, fisheries scientists pool individual fish data from multiple collection events (e.g., electrofishing runs or nets) to develop and test length-frequency histograms. In this chapter, we provided alternative analysis methods that consider the collection event as the experimental unit rather than individual fish. Using the collection event as the experimental unit has advantages because the analysis considers among-sample variation in size structure rather than among-individual-fish variation. Additionally, using individual fish as the experimental unit often causes the error degrees of freedom to be very high, resulting in significant differences when distributions appear quite similar (e.g., Kolmogorov–Smirnov two-sample test). Thus, the use of collection events as the experimental unit results in a more conservative test of size structure differences, and we recommend these methods when possible. The fisheries scientist should understand the advantages of various statistical tests and match analysis methods as best as possible to design experiments properly. Size structure data can be analyzed as categorical (e.g., chi-square), proportional (e.g., PSD), or continuous (e.g., LOGIT) data, depending on the study design and sample size. Examples in this chapter provide guidance for comparisons across systems, through time, or both, depending on the study objectives. Experimental design and hypothesis testing methods for analyses of length-frequency data will continue to improve.

9.7 REFERENCES

Allen, M. S., and W. E. Pine, III. 2000. Detecting fish population responses to a minimum length limit: effects of variable recruitment and duration of evaluation. North American Journal of Fisheries Management 20:672–682.

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.

Bayley, P. B., and D. J. Austen. 1990. Modeling the sampling efficiency of rotenone in impoundments and ponds. North American Journal of Fisheries Management 10:202–208.

Bettoli, P. W., and M. J. Maceina. 1996. Sampling with toxicants. Pages 303–333 *in* B. R. Murphy and D. W. Willis editors. Fisheries techniques, 2nd edition. American Fisheries Society, Bethesda, Maryland.

Bettross, E. A., and D. W. Willis. 1988. Seasonal patterns in sampling data for largemouth bass and bluegills in a northern Great Plains impoundment. Prairie Naturalist 20:193–202.

Bister, T. J., D. W. Willis, M. L. Brown, S. M. Jordan, R. M. Neumann, M. C. Quist, and C. S. Guy. 2000. Development of standard weight (*Ws*) equations and standard length categories for 18 warmwater nongame and riverine fish species. North American Journal of Fisheries Management 20:570–574.

Boxrucker, J. 1987. Largemouth bass influence on size structure of crappie populations in small Oklahoma impoundments. North American Journal of Fisheries Management 7:273–278.

Boxrucker, J., and G. Ploskey. 1989. Gear and seasonal biases associated with sampling crappies in Oklahoma. Proceedings of the Annual Conference of the Southeastern Association of Fish and Wildlife Agencies 42(1988):89–97.

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.

Brown, M. L., D. W. Willis, and B. G. Blackwell. 1999. Physicochemical and biological influences on black bullhead populations in eastern South Dakota glacial lakes. Journal of Freshwater Ecology 14:47–60.

Carline, R. F., B. L. Johnson, and T. J. Hall. 1984. Estimation and interpretation of proportional stock density for fish populations in Ohio impoundments. North American Journal of Fisheries Management 4:139–154.

Conover, W. J. 1980. Practical nonparametric statistics, 2nd edition. Wiley, New York.

Cornelius, R. R., and T. L. Margenau. 1999. Effects of length limits on muskellunge in Bone Lake, Wisconsin. North American Journal of Fisheries Management 19:300–308.

Ebbers, M. A. 1987. Vital statistics of a largemouth bass population in Minnesota from electrofishing and angler supplied data. North American Journal of Fisheries Management 7:252–259.

Gabelhouse, D. W., Jr. 1984a. An assessment of crappie stocks in small Midwestern private impoundments. North American Journal of Fisheries Management 4:371–384.

Gabelhouse, D. W., Jr. 1984b. An assessment of largemouth bass slot length limits in five Kansas lakes. Kansas Fish and Game Commission, Comprehensive Planning Option Project FW-9-P-3, Pratt.

Gabelhouse, D. W., Jr., and D. W. Willis. 1986. Biases and utility of angler catch data for assessing size structure and density of largemouth bass. North American Journal of Fisheries Management 6:481–489.

Gilliland, E. 1987. Evaluation of Oklahoma's standardized electrofishing in calculating population structure indices. Proceedings of the Annual Conference of the Southeastern Association of Fish and Wildlife Agencies 39(1985):277–287.

Green, D. M., B. J. Schonhoff III, and W. D. Youngs. 1986. The New York state bass study 1977–1980: use of angler collected data to determine population dynamics. New York State Department of Environmental Conservation, Albany.

Griffith, J. S. 1981. Estimation of the age-frequency distribution of stream dwelling trout by underwater observation. Progressive Fish-Culturist 43:51–53.

Gustafson, K. A. 1988. Approximating confidence intervals for indices of fish population size structure. North American Journal of Fisheries Management 8:139–141.

Guthery, F. S., L. A. Brennan, M. J. Peterson, and J. L. Lusk. 2005. Information theory in wildlife science: critique and viewpoint. Journal of Wildlife Management 69:457-465.

Guy, C. S., E. A. Bettross, and D. W. Willis. 1990. A proposed standard weight (*Ws*) equation for sauger. Prairie Naturalist 22:41–48.

Guy, C. S., and D. W. Willis. 1990. Structural relations of largemouth bass and bluegill populations in South Dakota ponds. North American Journal of Fisheries Management 10:338–343.

Guy, C. S., and D. W. Willis. 1991. Evaluation of largemouth bass–yellow perch communities in small South Dakota impoundments. North American Journal of Fisheries Management 1143–49.

Guy, C. S., and D. W. Willis. 1995. Population characteristics of black crappies in South Dakota waters: a case for ecosystem specific management. North American Journal of Fisheries Management 15:754–765.

Hayne, D. W., G. E. Hall, and H. M. Nichols. 1967. An evaluation of cove sampling of fish populations in Douglas Reservoir, Tennessee. Pages 244–297 *in* Reservoir Committee, editors. Reservoir fishery resources symposium. American Fisheries Society, Southern Division, Reservoir Committee, Bethesda, Maryland.

Hill, T. D., and D. W. Willis. 1993. Largemouth bass biomass, density, and size structure in small South Dakota impoundments. South Dakota Academy of Science 72:31–39.

Holland, R. S., and E. J. Peters. 1992. Differential catch rates by hoop nets of three different mesh sizes in the lower Platte River. North American Journal of Fisheries Management 12:237–243.

Hubbard, W. D., and L. E. Miranda. 1988. Competence of non-random electrofishing sampling in assessment of structural indices. Proceedings of the Annual Conference of the Southeastern Association of Fish and Wildlife Agencies 40(1986):79–84.

Hubert, W. A. 1996. Passive capture techniques. Pages 157–192 *in* B. R. Murphy and D. W. Willis, editors. Fisheries techniques, 2nd edition. American Fisheries Society, Bethesda, Maryland.

Isermann, D. A., P. A. Hanchin, and D. W. Willis. 2002. Comparison of two mesh sizes for collecting larval yellow perch in surface trawls. North American Journal of Fisheries Management 22:585–589.

Jacobs, R. P., and E. B. O'Donnell. 1996. An electrofishing survey of selected Connecticut lakes. Connecticut Department of Environmental Protection, Fisheries Division, Federal Aid in Sport Fish Restoration Project F-57-R-14, Final Report, Hartford.

Jacobs, R. P., E. B. O'Donnell, and A. P. Petrillo. 1995. Assessment of experimental length limits on largemouth bass and Lake Saltonstall fisheries investigations. Connecticut Department of Environmental Protection, Fisheries Division, Federal Aid in Sport Fish Restoration Final Report F-57-R-12, Hartford.

Jakes, D. S. 1987. Population structures and prey exploitation of largemouth bass in three unexploited impoundments. Master's thesis. University of Georgia, Athens.

Johnson, S. L., F. J. Rahel, and W. A. Hubert. 1992. Factors influencing the size structure of brook trout in beaver ponds in Wyoming. North American Journal of Fisheries Management 12:118–124.

King, T. A., J. C. Williams, W. D. Davies, and W. L. Shelton. 1981. Fixed versus random sampling of fishes in a large reservoir. Transactions of the American Fisheries Society 110:563–568.

Koopmans, L. H. 1987. Introduction to contemporary statistical methods. PWS Publishers, Duxbury Press, Boston.

Laarman, P. W., and J. R. Ryckman. 1982. Relative size selectivity of trap nets for eight species of fish. North American Journal of Fisheries Management 2:33–37.

Littell, R. C., G. A. Milliken, W. W. Stroup, and R. D. Wolfinger. 1996. SAS system for mixed models. SAS Institute, Cary, North Carolina.

Lott, J. P., and D. W. Willis. 1991. Gill net mesh size efficiency for yellow perch. Prairie Naturalist 23:139–144.

Maceina, M. J., P. W. Bettoli, and D. R. DeVries. 1994. Use of a split-plot analysis of variance design for repeated-measures fishery data. Fisheries (Bethesda) 19(3):14–20.

Margenau, T. L., and S. P. AveLallemant. 2000. Effects of a 40-inch minimum length limit on muskellunge in Wisconsin. North American Journal of Fisheries Management 20:986–993.

Mero, S. W., and D. W. Willis. 1992. Seasonal variation in sampling data for walleye and sauger collected with gill nets from Lake Sakakawea, North Dakota. Prairie Naturalist 24:231–240.

Michaletz, P., J. Boxrucker, S. Hale, and J. R. Jackson. 1995. Comparison of four trawls for sampling juvenile shad. North American Journal of Fisheries Management 15:918–923.

Milewski, C. L., and D. W. Willis. 1991. Smallmouth bass size structure and catch rates in five South Dakota lakes as determined from two sampling gears. Prairie Naturalist 23:53–60.

Miranda, L. E. 1983. Relation between growth, mortality, and proportional stock density of largemouth bass in Texas reservoirs. Annual Proceedings of the Texas Chapter, American Fisheries Society 5:29–38.

Miranda, L. E. 1993. Sample sizes for estimating and comparing proportion-based indices. North American Journal of Fisheries Management 13:383–386.

Mullner, S. A., W. A. Hubert, and T. A. Wesche. 1998. Snorkeling as an alternative to depletion electrofishing for estimating abundance and length-class frequencies of trout in small streams. North American Journal of Fisheries Management 18:947–953.

Murphy, B. R., M. L. Brown, and T. A. Springer. 1990. Evaluation of the relative weight index with new applications to walleye. North American Journal of Fisheries Management 10:85–97.

Neal, J. W., R. L. Noble, and J. A. Rice. 1999. Fish community response to hybrid striped bass introduction in small warmwater impoundments. North American Journal of Fisheries Management 19:1044–1053.

Neumann, R. M., C. S. Guy, and D. W. Willis. 1995. Precision and size structure of juvenile Percichthyid samples collected with various gears from Lake Texoma. North American Journal of Fisheries Management 15:956–962.

Neumann, R. M., and D. W. Willis. 1995. Seasonal variation in gill-net sample indexes for northern pike collected from a glacial prairie lake. North American Journal of Fisheries Management 15:838–844.

Neumann, R. M., D. W. Willis, and D. D. Mann. 1994. Evaluation of largemouth bass slot length limits in two small South Dakota impoundments. Prairie Naturalist 26:15–32.

Novinger, G. D., and R. E. Legler. 1978. Bluegill population structure and dynamics. Pages 37–49 *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.

Paragamian, V. L. 1989. A comparison of day and night electrofishing: size structure and catch per unit effort for smallmouth bass. North American Journal of Fisheries Management 9:500–503.

Paukert, C. P., and D. W. Willis. 2000. Factors affecting panfish populations in Sandhill lakes. Nebraska Game and Parks Commission, Federal Aid in Sport Fish Restoration Project F-11-R, Completion Report, Lincoln.

Paukert, C. P., D. W. Willis, and J. A. Klammer. 2002. Effects of predation and environment on quality of yellow perch and bluegill populations in Nebraska Sandhill lakes. North American Journal of Fisheries Management 22:86–95.

Pope, K. L., and D. W. Willis. 1996. Seasonal influences on freshwater fisheries sampling data. Reviews in Fisheries Science 4:57–73.

Reynolds, J. B., and L. R. Babb. 1978. Structure and dynamics of largemouth bass populations. Pages 50–61 *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.

Reynolds, J. B., and D. E. Simpson. 1978. Evaluation of fish sampling methods and rotenone census. Pages 11–24 *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.

Robinson, J. W. 1994. Sampling procedures to capture flathead catfish and channel catfish. Missouri Department of Conservation, Federal Aid in Sport Fish Restoration Project F-1-R-43, Final Report, Columbia.

Roni, P., and A. Fayram. 2000. Estimating winter salmonid abundance in small western Washington streams: a comparison of three techniques. North American Journal of Fisheries Management 20:683–692.

Saffel, P. D., C. S. Guy, and D. W. Willis. 1990. Population structure of largemouth bass and black bullheads in South Dakota ponds. Prairie Naturalist 22:113–118.

Santucci, V. J., Jr., D. H. Wahl, and D. F. Clapp. 1999. Efficiency and selectivity of sampling methods used to collect channel catfish in impoundments. Pages 317–328 *in* E. R. Irwin, W. A. Hubert, C. F. Rabeni, H. L. Schramm, Jr., and T. Coon, editors. Catfish 2000: Proceedings of the international ictalurid symposium. American Fisheries Society, Symposium 24, Bethesda, Maryland.

SAS. 1999. SAS statistics user's guide. SAS Institute, Cary, North Carolina.

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.

Trippel, E. A., and J. J. Hubert. 1990. Common statistical errors in fishery research. Pages 93–102 *in* J. Hunter, editor. Writing for fishery journals. American Fisheries Society, Bethesda, Maryland.

Underwood, T. J. 2000. Abundance, length composition, and migration of spawning inconnu in the Selawik River, Alaska. North American Journal of Fisheries Management 20:386–393.

Unmuth, J. M. L., M. J. Hansen, P. W. Rasmussen, and T. D. Pellett. 2001. Effects of mechanical harvesting of Eurasian water milfoil on angling for bluegills in Fish Lake, Wisconsin. North American Journal of Fisheries Management 21:448–454.

Van Den Avyle, M. J., G. R. Ploskey, and P. W. Bettoli. 1995. Evaluation of gill-net sampling for estimating abundance and length frequency of reservoir shad populations. North American Journal of Fisheries Management 15:898–917.

Van Zee, B. E., D. W. Willis, and C. C. Stone. 1996. Comparison of diel sampling data for sauger collected by electrofishing. Journal of Freshwater Ecology 11:139–143.

Vokoun, J. C., and C. F. Rabeni. 1999. Catfish sampling in rivers and streams: a review of strategies, gears, and methods. Pages 271–286 *in* E. R. Irwin, W. A. Hubert, C. F. Rabeni, H. L. Schramm, Jr., and T. Coon, editors. Catfish 2000: proceedings of the international ictalurid symposium. American Fisheries Society, Symposium 24, 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.

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.

Weithman, A. S., J. B. Reynolds, and D. E. Simpson. 1980. Assessment of structure of largemouth bass stocks by sequential sampling. Proceedings of the Annual Conference of the Southeastern Association of Fish and Wildlife Agencies 33(1979):415–424.

Wildman, T. L., and R. M. Neumann. 2003. Comparison of snorkeling and electrofishing for estimating abundance and size structure of brook trout and brown trout in two southern New England streams. Fisheries Research 60:131–139.

Willis, D. W., C. S. Guy, and B. R. Murphy. 1991. Development and evaluation of a proposed standard weight (*Ws*) equation for yellow perch. North American Journal of Fisheries Management 11:374–380.

Willis, D. W., and B. R. Murphy. 1996. Planning for sampling. Pages 1–16 *in* B. R. Murphy and D. W. Willis, editors. Fisheries techniques, 2nd edition. American Fisheries Society, Bethesda, Maryland.

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.

Willis, D. W., and C. G. Scalet. 1989. Relations between proportional stock density and growth and condition of northern pike populations. North American Journal of Fisheries Management 9:488–492.

Zar, J. H. 1996. Biostatistical analysis, 3rd edition. Prentice-Hall, Upper Saddle River, New Jersey.