Chapter 8. Data Analysis

Lyman L. McDonald
Christina D. Vojta
Kevin S. McKelvey

8.1 Objective

Perhaps the greatest barrier between monitoring and management is data analysis. Data languish in drawers and spreadsheets because those who collect or maintain monitoring data lack training in how to effectively summarize and analyze their findings. This chapter serves as a first step to surmounting that barrier by empowering any monitoring team with the basic understanding of how to get data out of the drawer and onto the management table. Even if a statistician will complete the task of data analysis, monitoring team members need to have sufficient knowledge about the data analysis process to effectively work with a statistician. This chapter outlines the basic steps involved in data analysis at specific milestones in a monitoring effort.

We begin with key concepts related to data analysis and then provide Internet links and references to statistical textbooks and methods that are designed specifically for natural resource data users. A complete discussion of statistical methods is not possible within the scope of this chapter, so we encourage readers to become acquainted with the wealth of assistance that is available online and in well-written texts.

The remainder of the chapter describes the process of data analysis at various stages in a monitoring program, beginning with the evaluation of pilot data. We then follow with sections that address how to analyze inventory and baseline data, compare data between two points in time, and analyze multiyear data. Throughout the chapter, our emphasis is on field-sampled data, whether it originates from an existing program such as the Forest Inventory and Analysis (FIA) Program, or whether it has been collected for a specific habitat monitoring program. This chapter does not reiterate aspects of sampling design that are addressed in chapter 3. Decisions about the sampling design, however, will ultimately affect the type of data analysis. For example, a systematic sampling design may result in spatial dependence among sampling units and may require a **time series** analysis or other analytical technique that adjusts for spatially correlated data.

8.2 Key Concepts

8.2.1 Planning for Data Analysis

In the temporal sequence of a monitoring study, data analysis follows data collection. Discussions about possible data analysis methods, however, are an important part of the planning phase. Too often, decisions about data analysis methods are left until the data are collected, resulting in highly complex analyses in order to overcome flaws in the initial design. By addressing the data analysis methods and plotting hypothetical data during the planning phase, a monitoring team can evaluate whether the proposed data-collection and data-analysis methods will actually result in the desired information outcome of the monitoring objective. Moreover, by knowing what statistical analyses are planned, the monitoring team can create a sampling design that will meet the assumptions of the statistical methods (Elzinga et al. 1998). Although this chapter on data analysis follows the sequence of process steps for conducting a monitoring study (table 10.1), we encourage monitoring teams to consider data analysis when planning the monitoring design (chapter 3).

8.2.2 Statistical Versus Biological Significance

Over the last century, wildlife research studies have typically used hypothesis testing as the framework for conducting research and reaching defensible conclusions. The basic concept of hypothesis testing is to compare a research hypothesis with a null hypothesis by performing a statistical test and generating a significance level (p-value; i.e., probability of observing the data if the null hypothesis is true [Popper 1959]). More recently, wildlife research studies frequently use multimodel inference (Burnham and Anderson 2002) to compare several competing research hypotheses and choose the model that best supports the underlying data. Hypothesis testing and multimodel inference rely on statistical indices to objectively evaluate support for a research hypothesis.

Statistical significance can serve as a baseline quality standard that allows research results to pass a test of research objectivity. Many statisticians have shown that statistical significance can be meaningless (Johnson 1999, Simberloff 1990, Tukey 1969), however, especially if the null hypothesis is nonsensical or known to be false before data are collected (Johnson 1999). For example, the standard null hypothesis that one tries to disprove is that no effect or change occurred over time and that any observed patterns can be explained through random processes. Because one seldom explores phenomena that are unlikely to have an effect, nearly any study can result in statistical significance if the sample size is extremely large (Berger and Sellke 1987). In general, the failure to detect significant results has more to do with small sample size, which results in a low power to detect change (Nunnally 1960). For landscape analyses, it can be easy to find a statistical difference in some landscape pattern metrics for two points in time, simply because of differences in classification and mapping methods used for the two time periods. Even with simulated landscapes that contain no classification error, small differences in land cover proportion can yield statistically significant differences in landscape pattern indices (Remmel and Csillag 2003).

A key aspect of data analysis is to evaluate whether the results of a study are biologically significant and can be used to inform management actions. Monitoring studies, like other research studies, must consider biological significance when evaluating monitoring results. If the results are statistically significant, the monitoring team needs to ask whether the magnitude of change indicates a meaningful change in either habitat quantity or quality. For example, if the results of a monitoring study indicate a statistically significant increase in canopy cover from 33 to 35 percent, it is important to evaluate whether that difference has changed the quality of habitat for the emphasis species. If the emphasis species can successfully use a vegetation type when canopy cover is 20 percent or greater, the observed increase from 33 to 35 percent does not have biological significance.

Conversely, if the results are insignificant, the monitoring team needs to evaluate whether the sampling design was sufficient to detect a level of change that is biologically meaningful. Elzinga et al. (1998) describe how to calculate the minimum detectable change that is possible for the data, given the sample size, sample standard deviation, threshold significance level for the test, and an acceptable level of power. If the minimum detectable change is larger than a change that is considered to be biologically meaningful, it is possible that a biologically significant change has occurred but that the design was not adequate to detect it and that this change will not be detectable until better methods are devised.

Other options besides conducting significance tests are available for evaluating monitoring data. In section 8.6, we illustrate the use of **confidence intervals** as an alternative to significance tests and describe how confidence intervals can be used to evaluate the biological significance of the data. In chapter 3, we recommend comparing monitoring data with a threshold that is based on ecological information rather than arbitrary statistical considerations.

8.3 Statistical Resources for Monitoring

An excellent source of information is *Measuring and Monitoring Plant Populations* (Elzinga et al. 1998; http://www.blm.gov/nstc/library/pdf/MeasAndMon.pdf). This publication specifically describes monitoring of plant populations; however, the study designs and analysis procedures are directly applicable to wildlife habitat monitoring studies. Chapter 7 addresses sampling design, and chapter 11 describes data analysis methods. In addition, the following appendixes of Elzinga et al. (1998) address statistical analysis: appendix 7 (Sample Size Equations), appendix 8 (Terms and Formulas Commonly Used in Statistics), appendix 14 (Introduction to Statistical Analysis Using Resampling Methods), and appendix 18 (Estimating the Sample Size Necessary to Detect Changes Between Two Time Periods in a Proportion When Using Permanent Sampling Units [based on data from only the first year]).

For complete coverage of basic statistical principles and analyses, we recommend *Biometry: The Principles and Practices of Statistics in Biological Research, 3rd Edition* (Sokal and Rolf 1994), and *Biostatistical Analysis, 5th Edition* (Zar 2010). Both texts cover a broad range of topics and use numerous natural resource examples. For an overview of design of biological field-sampling procedures and statistical analyses, we recommend two texts: *Wildlife Study Design*, *2nd Edition* (Morrison et al. 2008), and *Statistics for Environmental Science and Management, 2nd Edition* (Manly 2009). Manly's book emphasizes sampling of the environment with special attention to monitoring, impact assessment, reclamation assessment, and basic analysis methods for time series and spatially correlated data. For information on graphical presentation of data, we recommend *Graphical Methods for Data Analysis* (Chambers et al. 1983) and *Creating More Effective Graphs* (Robbins 2005). See also Collier (2008) for examples of using graphs when describing and analyzing wildlife research data and Friendly (1995) for a discussion on the use of visual models for categorical data.

Forest Service statisticians developed a Web site that is internally available to Forest Service employees at http://statistics.fs.fed.us. The link named Statistical How-To's takes readers to a number of statistical analysis procedures prepared by statisticians at each Forest Service research station and listed under the acronym for that research station.

8.4 Evaluating Data From a Pilot Study

A monitoring team can choose to run a pilot study for a number of reasons, such as to test a field protocol, determine logistical constraints, estimate costs, or train field personnel (chapter 3). We recommend that any pilot study gather enough samples (usually more than 20) to evaluate the expected variability associated with each selected habitat attribute. Low variability can suggest that the attribute may not be a good indicator of habitat change, or that it is being measured at a resolution that is too coarse to detect differences over time. High variability can indicate that the attribute has a wide range in potential values because of factors such as elevation, slope, or soil type. High variability can also be a forewarning of possible measurement error, however, and could motivate the need to test for differences among observers in the attribute values measured at the same sites.

In addition, an evaluation of variability is necessary to determine whether it will be possible to meet the monitoring objectives at the desired power and precision. The amount of variability observed in pilot data can lead to important decisions such as changing the resolution of measurements, providing additional training for field personnel, or dropping the habitat attribute altogether and replacing it with one that is more or less sensitive, easier to measure, or easier to interpret.

The team can begin to evaluate spatial variability by displaying the data in a simple three-dimensional scatter plot for each habitat attribute that shows the values obtained for the attribute (vertical axis) plotted against the longitude and latitude of each sampling unit (horizontal axes). Also, the team can visually evaluate the effects of environmental factors in two-dimensional scatter plots by plotting the attribute against individual factors of interest (e.g., elevation) on a single horizontal axis. This information may come into play as part of adaptive management if changes in habitat are disproportionate across different elevations, management units, or other subunits of the area of interest.

If an evaluation of variability reveals only minor differences among sampling units, the monitoring team needs to address whether sufficient differences are likely between years to make the attribute worth sampling. Each attribute should have some sensitivity related to environmental conditions or management actions for managers to detect meaningful change over time. By contrast, an attribute that manifests high variability among sampling units may warrant further investigation to ensure that the variability is not largely because of measurement errors or differences among surveyors. During the pilot study, we recommend that each surveyor take measurements at the same set of sampling units so that the values obtained from each surveyor can be graphically compared. The monitoring team should avoid field methods that result in unacceptable measurement error or require subjective input from surveyors, because these deficiencies will reduce the value of the habitat attribute for monitoring.

The variability observed in a pilot study can be used in a power analysis to estimate the sample size needed for the full monitoring program (see also chapter 3, section 3.3.3, Estimate the number of sampling units required). In the context of **frequentist statistics,** a power analysis evaluates the power of a given statistical test to reject a null hypothesis when the null hypothesis is false. The generalized null hypothesis for most monitoring efforts is that the monitored attributes will not change between monitoring periods or over the course of a monitoring program. Because the power analysis is based on a specific statistical test, the monitoring team must have a fairly good idea of the type of test that eventually will be used to analyze the data. For example, different forms of power analysis exist for t-tests, paired t-tests, analysis of variance, and comparisons of proportions that take into consideration the sampling design and distribution of the data (e.g., normal, chi-square, Poisson).

Power is a function of four factors: sample size, effect size, variance, and alpha level (the critical value set for rejecting the null hypothesis). By presetting the power to a desired level, the power equation can be rearranged to solve for the estimated sample size needed, using a specified effect size, a selected alpha level, and the variance estimated from the pilot study. In case the variance from the full monitoring program turns out to be greater than the variance from the pilot study, it is always wise to boost the final sample size upward from the estimated sample size derived from the power analysis.

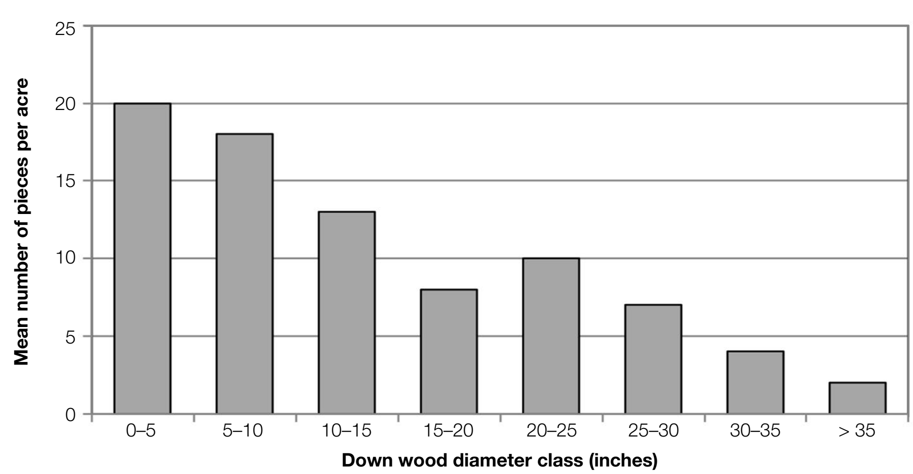
Given that properly designed pilot studies are costly, the desire is often to get as much out of pilot data as possible. A monitoring team should not report descriptive statistics from a pilot study as part of the monitoring results or as preliminary guidance to managers, however, especially if the pilot was based on a much smaller sample size than will be used for the full study. A number of factors could cause pilot study values to differ substantially from the values obtained during a full monitoring effort: (1) insufficient representation of the entire area of interest, (2) inadequate sample size to determine actual variability in the attribute, (3) higher or lower attribute values if collected outside the intended monitoring period (e.g., the autumn or spring before the monitoring period), and (4) either higher or lower attribute values during the training period. Pilot data provide valuable input for refining a monitoring design but should not be treated as monitoring outcomes for adjusting management actions.

8.5 Analysis of Inventory Data and Baseline Monitoring

As with a pilot study, the first step in the analysis of inventory or baseline data is to plot the data so that it can be visually evaluated. Even if the sample size is based on variability estimated from the pilot study, a habitat attribute likely will exhibit greater variability when all sites are sampled, especially if samples are collected over a larger spatial extent than the pilot. As with pilot data, a simple three-dimensional scatter plot of each habitat attribute, plotted against longitude and latitude of sampling units on the horizontal axes and values of the attribute on the vertical axis may quickly reveal a great deal about the properties of the collected data. A second way to visually evaluate the data is to create two-dimensional scatter plots with values of one habitat attribute on the vertical axis and values of another attribute (or other gradient such as elevation) on the horizontal axis.

Another useful exercise is to place continuous data of an attribute into classes (e.g., diameter classes of down wood) and then create a histogram that shows the distribution of the class data across all sampling units (figure 8.1). Often a histogram is more meaningful than a simple average value because it shows the proportion of the total sample that falls into each class. For example, knowing that 50 percent of sampled down wood diameters are in the range of 10 to 20 inches (in) could be more meaningful in evaluating a species' habitat quality than knowing that the average diameter across all plots is 14 in.

Figure 8.1.—*A simple histogram showing the mean number of pieces of down wood per acre in eight diameter classes.*



A way to evaluate whether the full suite of attributes needs to be monitored is to develop a correlation matrix between all combinations of habitat attributes (table 8.1). A high correlation (e.g., $|r|$ is greater than 0.60) between any two attributes could indicate that both are providing similar information, so that it is unnecessary to continue measuring both attributes in the next monitoring period. If two variables are highly correlated, we recommend selecting the attribute that has the least potential for measurement error for continued monitoring, while also considering ease of measurement and understanding by managers. For example, basal area and canopy cover are often highly correlated, and canopy cover based on field estimates can be highly variable. If the first year of data collection confirms a high degree of variability in measures of canopy cover, and if the two attributes are indeed highly correlated, we recommend dropping the canopy cover estimates and measuring only basal area.

The next step in the analysis of inventory or baseline data is to calculate descriptive statistics (e.g., mean, median, standard deviation, coefficient of variation, skewness) for each habitat attribute. The choice of an average statistic will depend on how the data are distributed, and, as mentioned previously, the frequency distribution or **box-and-whisker plots** are more meaningful for some attributes than any form of average value.

If the monitoring program is based on FIA data, the team can generate tabular and spatial summaries based on user-defined inputs using tools available at http://fia.fs.fed.us/tools-data/default.asp (see chapter 4, section 4.4.1). Within the Natural Resource Manager (NRM) environment, a monitoring team can use the Graphical Interface tool to perform queries (referred to as views) that summarize the raw data collected at sites, in fixed area plots, along line intercepts, or in quadrats. Data summaries include descriptive statistics for continuous and categorical variables. These views will extract summary or raw data from NRM into either an Access database or an Excel spreadsheet. From that point, data can be analyzed in any statistical software program.

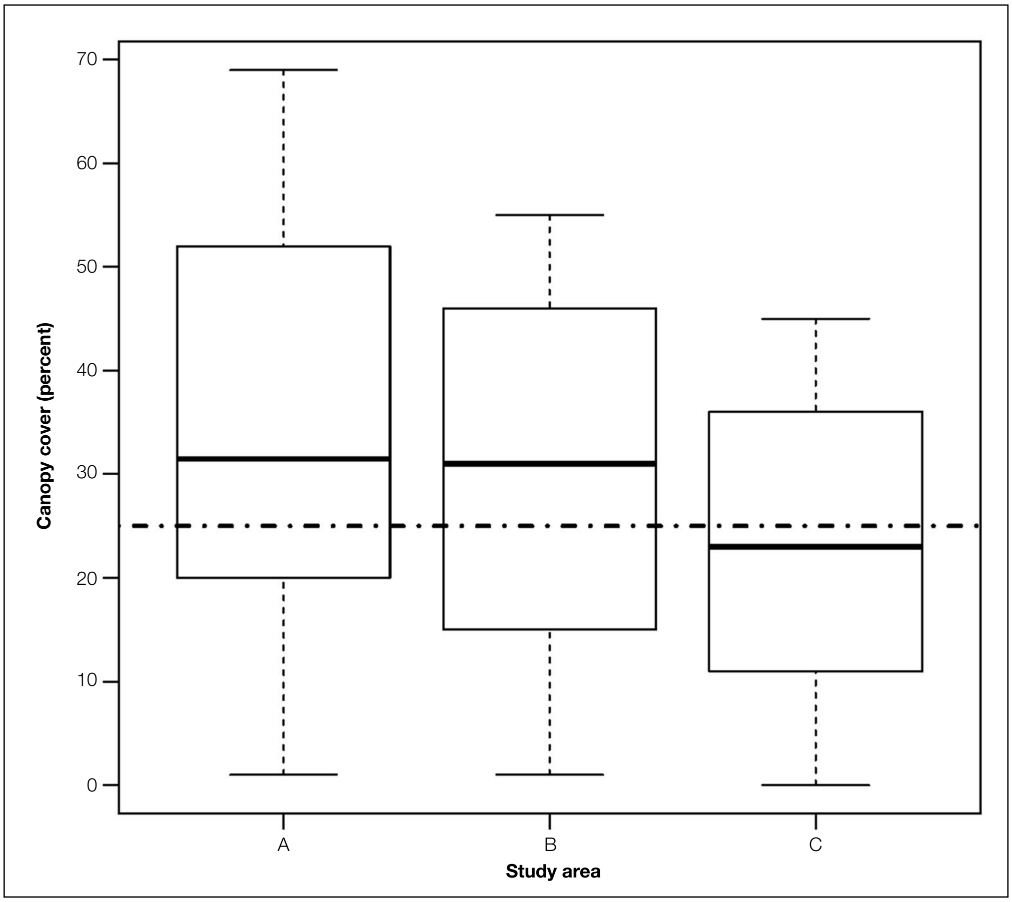
For an inventory or baseline study, descriptive statistics and box-and-whisker plots may be the only form of data analysis and presentation needed. If the objective of a monitoring program is to compare the status of an attribute with a desired condition or threshold, however, the final step of the analysis is to compare the outcomes of the

Table 8.1.—*Correlation matrix of potential habitat attributes for monitoring. Tree canopy cover and basal area have a correlation coefficient (r) greater than 0.60 (shown in bold) and are therefore considered highly correlated.*

Attribute	Elevation	Slope	Tree canopy cover	Basal area	Down wood volume	Distance to road	Distance to well pad
Elevation	1.000	0.273	0.294	0.483	0.327	0.225	0.483
Slope	0.273	1.000	0.201	0.366	– 0.218	0.475	0.530
Tree canopy cover	0.294	0.201	1.000	**0.713**	0.398	0.030	0.162
Basal area	0.483	0.366	**0.713**	1.000	0.366	0.426	0.488
Down wood volume	0.327	– 0.218	0.398	0.366	1.000	0.269	0.287
Distance to road	0.225	0.475	0.030	0.426	0.269	1.000	0.560
Distance to well pad	0.483	0.530	0.162	0.488	0.287	0.560	1.000

sampling effort with the desired condition or threshold value. Using raw data, the relationship of the threshold value to the median and range of the distribution of data can be easily evaluated graphically by plotting the threshold value on a box-and-whisker plot (figure 8.2). Statistically, this comparison is generally accomplished by evaluating the relationship between the threshold value and an approximate 95-percent confidence interval about the mean; i.e., the mean plus or minus two standard errors.

Figure 8.2.—*Relationship of estimates of percentage canopy cover at three study sites to a threshold value of 25 percent. For two of the study sites (A and B), the median values are above the threshold, but for the third study site (C) the median is below the threshold, indicating that at this location a change in management or more intensive sampling is warranted.*



8.6 Comparing Data Between Two Monitoring Periods

Although the ultimate goal of any habitat monitoring program is to be able to detect change in habitat quantity or quality over time, change will usually not be evident after two monitoring periods unless (1) the temporal span between periods is large or (2) habitat has undergone substantial change because of management actions, large-scale natural disturbances, or changes in land use. If the conceptual model (chapter 2) suggests that habitat will likely change as a consequence of vegetation establishment or growth, a span of 5 or more years may be needed to detect any change. Even when environmental or management factors are expected to cause rapid change (e.g., intensive management actions, wildfire, or insect outbreaks), these factors often affect small spatial extents relative to the monitoring area, and change is therefore difficult to quantify.

In spite of these limitations, the merit in comparing data between two monitoring periods is apparent for several reasons. First, in the context of adaptive management, early evaluation of data can hint at eventual loss of habitat that could be prevented by making immediate adjustments in current management. For example, the establishment of invasive plants after certain types of burn prescriptions may not create a significant reduction in habitat between two monitoring periods, but the immediate implementation of a weed control plan following prescribed fire could prevent invasive species from having a more significant effect on habitat over time.

Second, early evaluation of monitoring data demonstrates to managers that data are indeed being collected for a purpose and that the effort is worth the financial investment. Although data from 2 years should never be overinterpreted, summary statistics and visual graphs can serve to promote the continuation of a monitoring program when it is still in its infancy.

Third, monitoring of land management plan objectives and desired conditions may be fiscally limited to only two time steps, one at the beginning and one at the end of each planning period. Thus, a meaningful approach for comparing these two periods will be necessary as part of the planning process. Even a simple qualitative comparison between monitoring periods that uses graphs and other visual aids can be important for communicating changes in wildlife habitat to the public and stakeholders.

As with pilot and first year data, the first step in comparing data from two monitoring periods is to plot the data. Following up on our previous suggestions, a possible approach is to plot attribute data in three dimensions, with longitude and latitude on the x and y axes and measured values of an attribute on the third axis, using different colors for each year. Alternatively, one could construct scatter plots in two dimensions of two habitat attributes, or one habitat attribute against a gradient such as elevation, with the two time periods in different symbols or colors. All these approaches will qualitatively reveal the amount of overlap and the degree of difference between the 2 years, although, in general, less variation is expected in the scatter plots if data are paired (i.e., repeated measures exist on the same plot for both time periods) relative to variation with nonpaired data. If attribute values from the 2 years overlap substantially on the scatter plots, it is unlikely that statistical tests will show any difference between years. These plots will also show the amount of variability associated with each attribute and whether the variability was similar between years.

Another useful technique, assuming repeated measurements on the same sampling units, is to plot the first year of data (x) against the second year of data (y) and insert a diagonal line (i.e., y = x). The plotted data will indicate a general increase in values of the attribute if most points fall above the line and a general decrease if most points fall below the line (figure 8.3).

The team should then generate descriptive statistics for each attribute, whether or not these statistics will be used in a formal test comparing the two monitoring periods. The team can visually evaluate differences between years by plotting two box-and-whisker plots side by side, one for each monitoring period, similar to figure 8.2. For some attributes (e.g., attributes derived from the diameters of trees, snags, or logs), a frequency distribution of size classes can be more meaningful than mean values. The team can compare histograms for each monitoring period, either side by side or one inverted under the other in a mirror image (figure 8.4). This comparison will show whether changes in attribute values between monitoring periods were evenly distributed across classes or more pronounced for certain size classes.

If scatter plots indicate that a statistical test is warranted, the team will want statistical assistance to ensure that the test performed is appropriate for the data. We caution, however, that statistical tests of hypotheses are easy to misinterpret and are not as useful for analysis of monitoring data as they are often promoted to be (Johnson 1999). Environmental data tend to have high variance, so the outcome of the test may not show statistical significance, although the change in habitat may be ecologically significant. For example, annual production of forbs is often highly variable across landscapes even with

Figure 8.3.—*A scatter plot of monitoring data from Year 1 and Year 10 for 18 landscapes in which mean patch size was measured. The relationship of attribute values to the diagonal line (x = y) indicates that no change in mean patch size exists for landscapes with smaller patches, but a decline in mean patch size occurred for larger patches in the second monitoring period.*

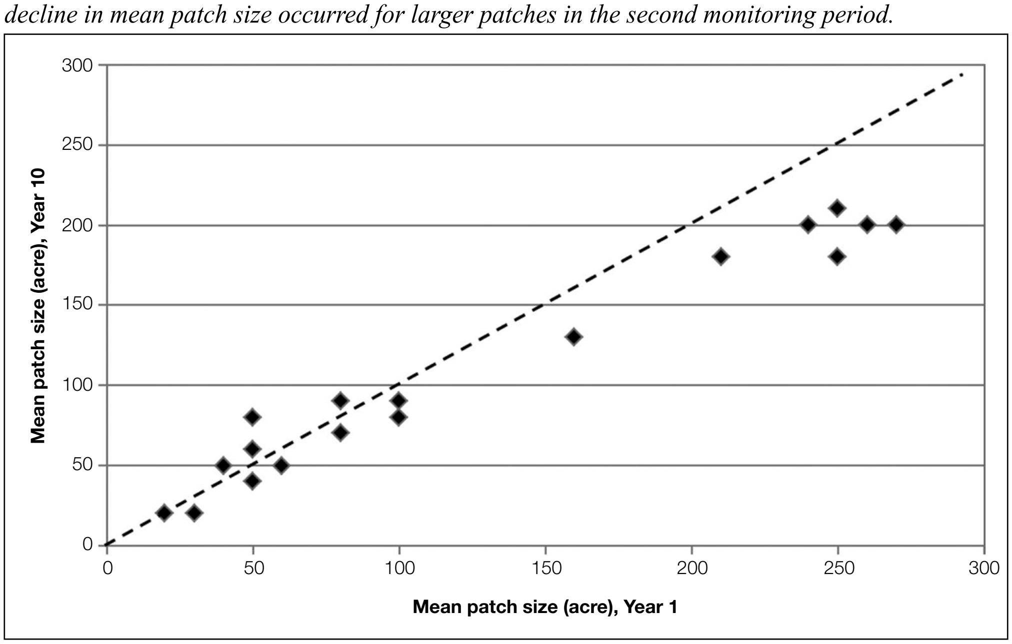
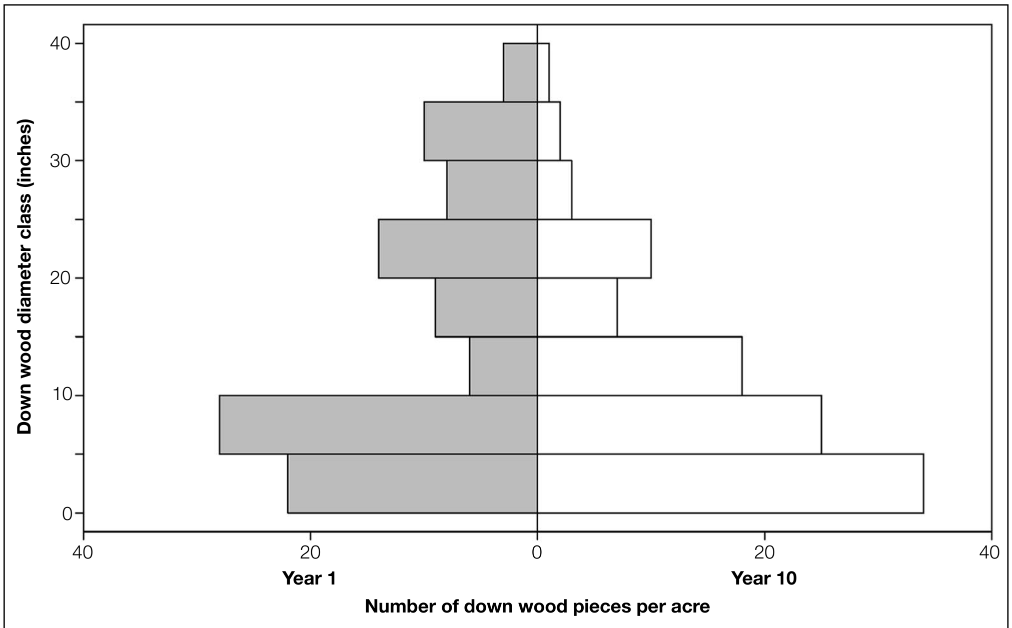


Figure 8.4.—*A comparison of histograms from two monitoring periods showing the distribution of down wood pieces per acre in eight diameter classes. Note that in Year 10 the number of pieces in the larger diameter classes (top of graph) has declined relative to Year 1, whereas the number of pieces in the smaller diameter classes has increased.*



repeated measurements on the same sampling units, so a test for differences may appear statistically insignificant when in actuality the reduction in forbs is ecologically significant for the emphasis species (see sidebar).

In contrast to statistical tests, confidence intervals are hard to misinterpret, both for comparing 2 years and for evaluating change over time. The usual presentation of a confidence interval is a graph showing the full range of possible attribute values on the vertical axis; the mean value of the attribute is indicated on the graph as a small horizontal bar, and the confidence interval is shown as a vertical line extending above and below the horizontal bar. The length of the vertical line is set to the values that represent two standard errors of the mean above and below the mean value.

In the sidebar example, the mean difference between the two monitoring periods was 100 lb/acre (ac) but was statistically not significant. The approximate 95-percent confidence interval on the mean difference is 100 ± (1.96)(300) lb/ac, or from 0.0 to 688 lb/ac.

The failure to reject the null hypothesis (no change in a metric before and after a point in time at a specified *p*-value) does not prove that there has been no important biological change. For example, assuming normality of the data, the estimated difference in production of forbs per acre might be $\overline{X}_1 - \overline{X}_2 = 100$ pounds (lb). If the estimated variance of the difference is 90,000, a z-statistic will be z = 100/√90,000 = 0.333 and the *p*-value will be much larger than the nominal value. A difference of 100 lb/acre, however, might be very important from a biological point of view; therefore, it is misleading to state that no statistically significant difference exists. Likewise, an observed difference can be statistically significant but have no biological significance. Always be skeptical of statistical results stated simply as significant or not significant with a given *p*-value (Johnson 1999).

(Note that the lower value of the confidence interval is 0 rather than -488, because the latter is computationally correct, but biologically implausible.) This confidence interval indicates not only that the mean difference is 100 lb/ac, but also that evidence indicates that the difference may be substantially greater than 100 lb. Ecologically, reporting the confidence interval is more meaningful than reporting that the variance is too great to show a significant difference between the mean values.

Thus, we recommend using confidence intervals rather than statistical tests when comparing two monitoring periods, especially when the data are highly variable and when the actual values of the data are more meaningful than the *p*-value associated with the test. Confidence intervals are also useful for assessing whether attributes are approaching or getting further from a threshold or desired condition value after the second monitoring period. The visual presentation of confidence intervals along with graphical plots of data from both periods provides an excellent tool for communicating with managers.

8.7 Analysis of Multiyear Data

A primary objective for most long-term monitoring studies is to look for a change or trend in either the means or proportions of the selected attribute(s) over a specified time period. A trend is an average annual increase or decrease in the attribute from year X to year Y. Analysis of trend can be challenging because it is difficult to distinguish a directional trend from multiyear cyclical variation, within-year seasonal variation, or erratic fluctuations (Dixon et al. 1998). Moreover, attribute values may show an abrupt increase or decrease caused by a single event (e.g., a management action or natural disturbance) rather than a gradual change over multiple time periods. Therefore, it is sometimes better to think of the multiyear analysis as a detection of change rather than a detection of trend.

A multiyear analysis begins with exploratory data analysis in the form of scatter plots and histograms, similar to that conducted with 1- and 2-year data. Two types of plots will usually be of interest: (1) a scatter plot of all data with lines connecting the dots for each sampling unit (e.g., plot, point, transect) and (2) a scatter plot of average values over time. In addition, we recommend using confidence intervals and box-and-whisker plots to evaluate the variance associated with each monitoring period.

Numerous methods of statistical trend analysis exist, and the use of any given method depends on the nature of the monitoring data. Olsen et al. (1999) distinguish between design-based and model-based statistical inferences. A design-based analysis requires a probability sample and uses properties of the sampling design to make inferences about changes or trend (Edwards 1998, Olsen et al. 1999). For example, long-term changes in the means and proportions of FIA data can be inferred through design-based approaches because the data were (and continue to be) collected with a probability sample. Throughout this technical guide, we have advocated the use of sampling designs for habitat monitoring, so we assume that most analyses of change or trend will be design based.

In contrast, a model-based analysis makes inferences about trend using a model that assumes a relationship between the mean attribute values and a set of covariates. A good example of a model-based analysis of trend is the Breeding Bird Survey, in which data on bird abundance are collected along transects that have been selected through professional judgment and therefore do not constitute a probability sample. The statisticians who analyze these data use complex models that take into account a number of factors that could influence estimates of mean counts of different species of birds (Link and Sauer 1998).

In addition to knowing whether the analysis will be design based or model based, other factors contribute to selection of the analysis method. For example, it is important to know whether the study area was stratified and how sampling units that are inaccessible were handled (McDonald et al. 2009). The statistician will also need to consider whether spatial and temporal correlation will cause problems during analysis.

Randomly sampled units are considered independent, whereas systematically sampled units may be spatially dependent, requiring a form of analysis that accounts for the spatial correlation between adjacent sampling units. If systematically sampled units are treated as if they were randomly sampled (i.e., independent), the analysis will usually be overly conservative, meaning that the reported standard errors may be larger and confidence intervals wider than would be obtained if the spatial correlation were taken into account (Manly 2009). On occasion, systematically sampled units can result in standard errors that are smaller than expected, if the systematic design has inadvertently duplicated a cyclic pattern on the landscape (e.g., every unit falls on a ridgetop). This latter possibility is usually recognized in the design stage and is easily avoided.

Many statistical approaches for detecting **monotonic** trend use the number of time periods as the sample size and fit a linear regression line through the yearly mean values, a strategy that may be appropriate for long time series of monitoring data but is less effective over short periods of time. It is clear that, with only a few time periods of data, the sample size will be small, with little power to make statistical inferences that important changes have occurred on a study area.

For monotonic trends that are assumed to be linear or log-linear, the basic unit analysis (McDonald et al. 2009) is to fit a line to the data collected on each unit. In the simplest form of analysis, **least squares regression** is used to estimate the slope of a line through the data collected on a single sampling unit over time. Each line yields an estimated slope, S_i, for each of the sample units (n), i = 1 through n. After the slopes are estimated, the average slope is computed from the random or systematic sample of size n. The statistician then computes the standard error of the mean slope and a confidence interval on the mean slope. If the confidence interval does not contain 0.0 then the conclusion is that a significant increasing (or decreasing) linear trend exists. This approach is equivalent to testing the null hypothesis that the mean slope is 0.0; however the width of the confidence interval provides more information than a simple yes or no answer from hypothesis testing with a specified *p*-value.

McDonald et al. (2009) refer to this approach as a unit analysis, where unit refers to each sampling unit and a measure of change or trend is obtained on each unit. In the previous paragraph, the measure of trend was the slope computed over time (i.e., average annual increase or decrease in the attribute) for each unit. The measure of change could be the difference in attribute values between two time periods on each unit. Depending on the data, the analyst can then report the mean or median of the measure of change or trend with accompanying confidence intervals and graphical presentations. From these summaries, statements can be made about whether the average or median slope is significantly different from 0, as well as the proportion of units in the population experiencing trends greater (or less) than a specific some standard value.

A unit analysis is clearly inappropriate for more complex study designs and long time series of data; however, it is simple and powerful in the sense that the number of sampling units, rather than the number of monitoring periods, is the sample size used in the analysis. In most studies, the number of sampling units will be fairly large, and the mean statistic for change or trend will be approximately normal. As a result, the confidence interval will be approximately correct.

The unit slope analysis is only one of many approaches to testing for trends. We highly recommend chapter 11 of Elzinga et al. (1998), which contains a wealth of information on statistical analyses of trend data.

8.8 Using Ancillary Data To Interpret Monitoring Outcomes

Every competent habitat monitoring program is designed to eventually produce defensible data that reveal whether the attributes of interest have substantially changed over time. Management decisions are often needed before sufficient time has elapsed, however, to reliably demonstrate results from a monitoring study. Even after the passage of several years, natural variability or issues of scale can inhibit the ability to detect whether changes in habitat have occurred, leaving the manager with the sense that the monitoring data are inconclusive and therefore not valuable.

Two basic understandings can help with the interpretation of monitoring data. The first is that any apparent trend in a habitat attribute is likely not spurious, because it takes a tremendous amount of management and natural disturbances to substantially change most habitat attributes over large spatial scales. Therefore, if data analysis indicates a weak trend, this result should be taken seriously, especially if it is a downward trend in a desired attribute or in habitat quantity or quality as a whole. Changes in management should not wait for a statistically significant trend.

The second understanding is that a lack of change in selected habitat attributes does not necessarily indicate that the habitat of the emphasis species is static. Environmental conditions are continually changing, so it is unlikely that habitat has remained unchanged.

It is possible that the selected attributes are less sensitive to change than originally thought and that other, unmeasured components of habitat are being lost or gained. For example, average tree diameter may show little change over the short term, implying that late seral conditions are still present for the emphasis species. In reality, a loss of older, larger trees may be occurring, but this loss is not captured by the diameter data because of the recruitment of small-diameter trees into medium-diameter size classes.

We recommend using a variety of ancillary data to either support or refute weak results from a monitoring program because this practice will enhance the utility of the monitoring program in an adaptive management context. In general, trends for one attribute do not occur in isolation, so the monitoring team can look for changes in related attributes that were not a part of the original design. The ancillary data may originate from a larger spatial scale, may have been collected for a different purpose, or may be from a data source not used in the original monitoring design. For example, changes in FIA core variables may mirror the changes seen on aerial photos or satellite imagery that were used for the monitoring program. Data collected for several range allotments that are only partially within the monitoring area may indicate similar changes as those found on the probability sample associated with the monitoring design. By aggregating a variety of data, both those directly derived from the monitoring and ancillary data from multiple sources, a monitoring team can arrive at a reasoned position on the likelihood that a weak trend (or conversely, a conclusion of no change) is true and not simply a sampling artifact. The aggregation of data from several sources will often enable managers to consider adaptive approaches early in the monitoring program, rather than waiting for more conclusive results that could come too late for management options.

The preceding discussion brings us to the inescapable conclusion that managers will have to work with levels of uncertainty that are far greater than those associated with scientific standards. The presence of uncertainty does not undermine the value of a carefully designed monitoring program, but managers need to become comfortable with using data of various qualities. Although managers cannot rely on weak data to the extent that they rely on rigorous evaluations of trends or targets, they should not ignore data simply because they are weak. It is often these data that enable managers to evaluate and achieve desired objectives.

8.9 Conclusions

Statistical analyses constitute a fundamental step in making monitoring information useful to management. Unless a monitoring team has conducted a complete census of each habitat attribute, statistical analyses are needed to make inferences from the sampled data to the entire area of interest. It is essential to address data analysis during the planning stage to ensure that the monitoring results will meet the monitoring objectives.

In this chapter, we present recommendations for conducting data analyses and displaying data at various stages of a monitoring program—as part of the pilot study, after the first sampling period (inventory and baseline data), after two points in time, and after a period of years. During each stage, the analysis team needs to use scatterplots and other forms of graphical presentation to become familiar with the data. Early and frequent analyses help build support for a monitoring program and can indicate whether a change in the monitoring design is needed.

When evaluating data at any point in time, it is important to look for biological significance as well as statistical significance. Statistical significance can serve as a baseline quality standard to establish confidence in the monitoring results, but statistical significance does not always imply biological significance and biological significance is what affects management actions. A statistical change in the value of one or more habitat attributes may be interesting, but a biological change can trigger a discussion of whether a change in management actions is warranted.

We advocate the use of confidence intervals to evaluate the data. Often the variance in data is equally or more meaningful than mean values, and the variance could shed light on biological significance. Confidence intervals are useful at all time periods, from the analysis of baseline to multiyear data. The visual presentation of confidence intervals provides an excellent tool for communicating with managers.

We recommend using a variety of ancillary data to either support or refute weak results from a monitoring program. The small spatial extent of most restoration projects, coupled with slow changes in vegetation from natural succession, make it difficult to show significant changes in habitat over short time periods. Ancillary data collected for other purposes may help indicate when an apparent trend has biological significance.