Chapter 8

Introduction to Univariate Analysis

"To be is to be the value of a variable." – Willard van Orman Quine

Remember the good ol' days when you were first learning algebra? The equations with only a single variable were always the easiest to solve. Turns out the same is true with statistics: **univariate** (single-variable) analyses are the easiest to set up for hypothesis testing, and they also result in very obvious, straightforward conclusions.

Univariate comparisons are at the very heart of scientific inquiry. After all, the cardinal rule of science is to limit the number of variables in an experiment so as to limit any sources of bias, while at the same time limit the confounding effects of multiple influences on the outcome. So the ultimate experiment should have only one variable. Many, many experimental groups, but only one variable.

Unfortunately for us, Mother Nature can rarely be convinced to keep things that simple; at least, not in a natural field setting. But even if your field program is designed to measure lots and lots of different variables, we can still choose to analyze each variable, one by one, when we make our comparisons between our sampled populations. Sticking to one variable makes things easy, but "easy" can still yield very important results.

Key Concepts

- Investigations of single-variant dynamics require that only a single variable be analyzed, separate from the rest.
- The *t* test is the most common statistical test used to make single-variable comparisons between different populations of data.
- The *t* test can be used to compare measures of central tendency (that is, the mean, median, or mode) between two populations.
- The *t* test can also be used to compare measures of the standard deviation or variance between two populations.
- Comparisons between multiple (three or more) populations can be done simultaneously using the one-way analysis of variance (ANOVA).

The Foundations of Univariate Analysis

As we have seen in the previous chapters, great attention has been paid to the many different kinds of variables that are deemed important in the aquatic sciences. And while we have also explored the practical methods by which those variables can (and should) be measured, it is critical that we look beyond the mere numbers on the page and engage ourselves in the effort to analyze those measurements. In most circumstances, the data we have gathered throughout the course of our laboratory and/or field efforts are most useful to us in the context of "comparative analysis." Univariate analyses represent the simplest, most straightforward method for making comparisons between populations.

If we wished to use inferential statistics (Chapter 2) to make those comparisons, there are a variety of statistical tests designed to determine equality (or inequality) between two or more sets of measurements of the same variable. For example, if we were to focus our attention on a single variable, we could use the unique distribution of our data (using central tendency, standard deviation, or variance) to compare our measurements against some other theoretical value or distribution.

As we learned in Chapter 2, the basic test of equality is accomplished using a two-tailed test, where the null hypothesis H_0 assumes the compared measures are not significantly different from each other (in other words, they are equal to each other). In the course of our statistical tests, if we are able to confirm the alternative hypothesis H_a that the compared measures are in fact not equal, then we can then follow up with a right- or left-tailed test to determine which population is significantly greater than the other.

As elementary as those concepts seem, don't let their simplicity fool you—statistics need not be complicated in order to be powerful or insightful. There is sublime elegance in the ability to determine equality (or the degree of inequality) based on the comparative measures using a single variable. Well, that may be romanticizing things a bit, but the point should be well taken: always opt for simple when simple will do.

That being said, there are definite limits to what information we can glean from single-variable analyses. One of the most notable weaknesses of univariate analysis is the fact that without the use of more than one variable, we are also limited in our ability to determine correlations. By definition, any associative relationship between one variable and another usually requires a minimum of two variables. Likewise, if we are interested in establishing a directional association that implies causation, we typically have to use two or more variables in that analysis.

For example, let us assume that we are using biometric data from juvenile fishes collected from a seine net survey (**Figure 8.1**). From our collection, we have measured the standard length SL and standard depth SD among all Atlantic menhaden (*Brevoortia tyrannus*) and Atlantic croaker (*Micropogonias undulatus*) specimens collected (**Table 8.1**).

If we were interested to explore the growth dynamics of Atlantic menhaden, it would be reasonable for us to expect that as the fish grows, that growth would affect each fish's length (SL) and its depth (SD), and that the two measures are in some way related to (correlated with) each other. Thus, univariate analysis would not be possible, because we would need to compare two variables in order to establish that relationship. In most circumstances,

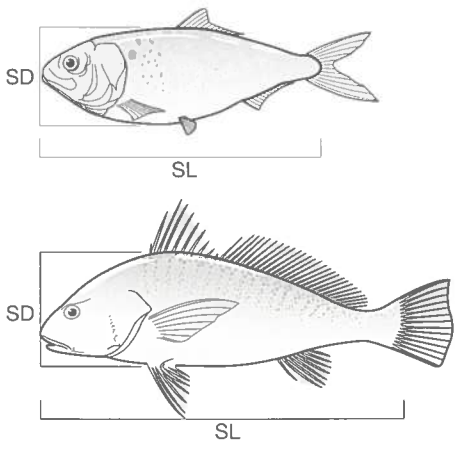


Figure 8.1 Biometric dimensions of Atlantic menhaden (*Brevoortia tyrannus*, top) and Atlantic croaker (*Micropogonias undulatus*, bottom), measured as standard length (SL) and standard depth (SD) for each fish collected in the sample. (Courtesy of the Maryland Department of Natural Resources.)

Table 8.1 Standard Length (SL) and Standard Depth (SD) Measurements Taken from Juvenile Atlantic Menhaden (*B. tyrannus*) and Atlantic Croaker (*M. undulatus*)

	B. tyrannus		*M. undulatus*	
	SL (mm) $N = 14$	SD (mm) $N = 14$	SL (mm) $N = 14$	SD (mm) $N = 14$
	102	44	89	36
	97	41	62	29
	78	36	90	36
	52	28	68	31
	112	46	54	27
	91	39	88	35
	36	18	65	26
	44	20	78	33
	27	17	65	25
	60	31	58	28
	89	38	47	24
	118	47	102	45
	98	41	86	36
	107	45	74	32
Mean ($\bar{X}$) =	79	35	73	32
Std Dev (s) =	30	11	16	5.7

correlations and causative relationships within our data can only be investigated within the context of multivariate analysis (which we shall explore in the next chapter).

Instead, let's say you wanted to compare the standard lengths SL between the two species in your collection. Then it would be a relatively simple matter to determine the central tendency, standard deviation, and variance of SL among your menhaden and croaker specimens and make your comparisons. In this case, you would be using only one variable to make your comparisons (that is, the standard length of each species), so univariate analysis would be appropriate.

You may be asking yourself—wouldn't "species type" be considered a second variable in this example? The answer is: it depends on how the comparisons are structured, and whether that second variable is intended to be used as a grouping variable or as a scaled variable. Let's consider this more closely.

Univariate Comparisons Must Be Structured to Possess Only One Scaled Variable

Recall from our initial discussion in Chapter 2 that there were two fundamental types of measurements: scaled measures and nominal measures. Scaled measures are those that can be ordered according to a continuous scale (which would also include ordinal measures). In our earlier example using menhaden and croaker biometrics (see Table 8.1), it is easy to see that the standard length and/or depth of a particular specimen will vary

Table 8.2 Standard Length (SL) Measurements Taken from Juvenile Atlantic Menhaden (*B. tyrannus*), Ranked According to the SL Ordination Scheme of Tiny (1–20 mm), Small (21–50 mm), Medium (51–80 mm), Large (81–110 mm), and Subadult (111+ mm)

B. tyrannus SL (mm)—scale $N = 14$	*B. tyrannus* SL (mm)—ranked $N = 14$	*B. tyrannus* SL—ordination $N = 14$
102	27	0 (Tiny)
97	36	3 (Small)
78	44	3 (Medium)
52	52	6 (Large)
112	60	2 (Subadult)
91	78	
36	89	total = 14
44	91	
27	97	
60	98	
89	102	
118	107	
98	112	
107	118	

according to a continuous length scale (in mm). Even if we decided to rank the juvenile menhaden in size, according to some ordinal scheme of our own choosing for SL:

Tiny:	1–20 mm
Small:	21–50 mm
Medium:	51–80 mm
Large:	81–110 mm
Subadult:	111+ mm

the data would still be ordered according to an underlying scaled measurement (in this example, according to SL). So let's take another look at our SL data from Table 8.1 for Atlantic menhaden (*B. tyrannus*), but this time let's rank all the SL data according to the ordinal size scheme (**Table 8.2**).

If we choose to rank our scale data (see Table 8.2, middle column), we have not altered the data in any way; we've simply organized it in a way that makes it easier for us to apply our ordinal ranking scheme. So even if we transform our scale data into ordinal data (see Table 8.2, right column), the ranks are still based on a continuous scale of SL, and the counts themselves vary according to a continuous scale. We could do the same for our juvenile Atlantic croaker (*M. undulatus*) biometrics (**Table 8.3**)

Keep in mind that the only restriction placed on univariate analysis is that we constrain our comparisons to a single scaled (or ordinal) variable. In this case, we could use SL as our single scaled variable and compare menhaden SL versus croaker SL. Or we could opt to use our ordinal ranks and compare menhaden size ranks versus croaker size ranks (**Figure 8.2**).

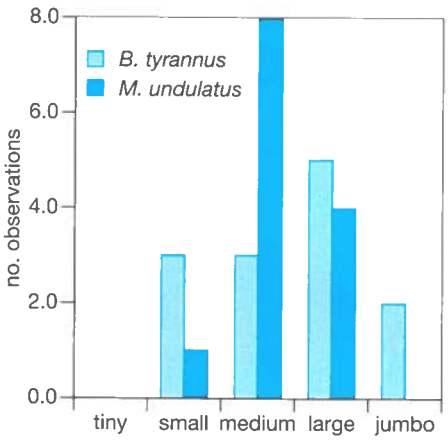


Figure 8.2 Histogram of the measured standard lengths (SL) among Atlantic menhaden (*B. tyrannus*) and Atlantic croaker (*M. undulatus*), indicating frequency of occurrence within each size class according to the ordinal scheme of Tiny (1–20 mm), Small (21–50 mm), Medium (51–80 mm), Large (81–110 mm), and Subadult (111+ mm).

Table 8.3 Standard Length (SL) Measurements Taken from Juvenile Atlantic Croaker (*M. undulatus*), Ranked According to the SL Ordination Scheme of Tiny (1–20 mm), Small (21–50 mm), Medium (51–80 mm), Large (81–110 mm), and Subadult (111+ mm)

M. undulatus SL (mm)—scale $N = 14$	*M. undulatus* SL (mm)—ranked $N = 14$	*M. undulatus* SL—ordination $N = 14$
89	47	0 (Tiny)
62	54	1 (Small)
90	58	8 (Medium)
68	62	5 (Large)
54	65	0 (Subadult)
88	65	
65	68	total = 14
78	74	
65	78	
58	86	
47	88	
102	89	
86	90	
74	102	

The key to remember here is that we are not using "species" as a variable that can be measured according to a continuous scale; we are using it instead as a way to categorize the populations we are comparing. So that means it is appropriate to use univariate analysis only if our second variable is nominal. In this example, our second variable represents categorical data that are used to discriminate between the species being compared.

Sometimes it can get a little tricky trying to decide when the variable you are using to define your groups is truly nominal and would therefore call for univariate analysis (**Technical Box 8.1**). Consider a typical field study, where

TECHNICAL BOX 8.1

How to Decide When Univariate Analysis Is Appropriate

Can't decide whether to group your comparisons using a nominal (categorical) scheme? Here are some examples of very common situations for comparison, and whether univariate or multivariate analyses are appropriate.

	Grouping variable
Nominal (use univariate analysis)	**Scaled/ordinal (use multivariate analysis)**
Time 1 vs. Time 2	Time elapsed between Time 1 and Time 2; or time series analysis
Before vs. after	Time elapsed between before and after; or time series analysis
Single treatment vs. control	Multiple treatments vs. control; or multiple effects measured over time
Location 1 vs. Location 2	Distance measured between Location 1 and Location 2; or distance measured from a fixed geographic point
Depth 1 vs. Depth 2	Depth measured between Depth 1 and Depth 2; or scaled depth from surface/bottom

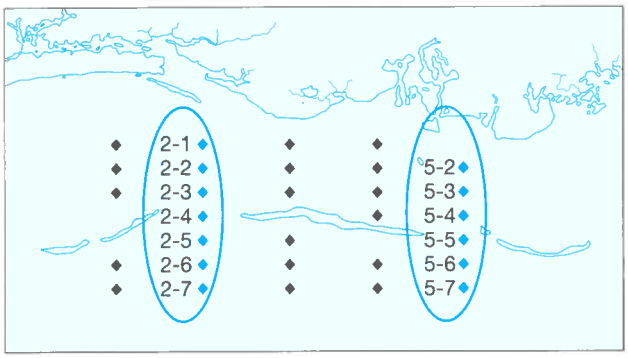


Figure 8.3 Site-to-site comparisons can be performed using univariate analysis as long as the stations are discriminated from each other using a nominal (categorical) measure, like Transect # - Station #. In that case, pairwise comparisons can be made between individual stations (for example, Station 2-1 versus Station 2-7, or Station 2-7 versus Station 5-7). Comparisons can also be made between entire transects (for example, Transect 2 versus Transect 5).

you have established a number of field stations along a number of cross-shelf transects (**Figure 8.3**). If you wanted to use Transect 2 and compare the SL of menhaden captured at the nearshore station (Station 2-1) against those from the offshore station (Station 2-7), the use of a nominal variable like "Station #" would permit univariate analysis because "Station #" is being used as a category, not as a scale measurement. However, if you decided to discriminate your stations by using something like "distance from the coastline," your comparisons between Station 2-1 and Station 2-7 would instead be based on a continuous scale measurement (that is, distance from the shore). In that case, you would have to use multivariate analyses (see Chapter 9) to look at the correlation between menhaden SL (scaled variable #1) and distance from shore (scaled variable #2).

Common Statistical Methods for Univariate Analysis

Just because we are focusing our attention on a single variable does not mean that the many, many measurements of that variable will be uniform. Quite the contrary, we expect there to be differences in those measurements. As we learned in Chapter 2, one of the most fundamental ways to describe a nonuniform dataset is by its central tendency; that is, by its mode, median, or mean.

When performing comparisons between two datasets, it is a common practice to compare the central tendencies of those datasets; chief among these are comparisons of the mean ($\bar{X}$). Conceptually, this makes the most sense to investigators, as the mean represents the "average condition" of the measured variable. The central tendency is usually the best way to quantify the variable in question, so it naturally follows that it would also be the best way to make comparisons between populations.

However, it is also possible (and in some cases more desirable) to compare the distribution of data rather than the central tendency of that data. Recall that the standard deviation s represents the basic tendency of the measured data to depart from the mean. Thus, a small standard deviation is indicative of minimal variability within the dataset. Another important descriptor of data distribution is the variance s^2, which is simply defined as the square of the standard deviation (so it is an even more sensitive method to describe departure from the mean).

Typically, comparisons of central tendency are used when investigators wish to test how accurately the datasets represent the populations from which the measurements were taken. Thus, comparisons of central tendency are

thought to best represent the population from which the data were taken. For example, the mean standard lengths of our juvenile menhaden and croaker from Table 8.1 are 79 mm and 73 mm, respectively. These values represent the central tendency (in this case, the mean SL) for juvenile menhaden and croaker, so any comparisons made between the two are comparisons made between the "average condition" of each sampled population. Comparisons between the standard deviations (or variances) can also be made, but they are only concerned with how the data are distributed about the central tendency, not what the data actually represent.

Although it would at first seem preferable to compare central tendency rather than the data distribution, it is important to consider both. After all, it is conceivable that two populations might share the same mean as a coincidence, but have very different distributions. Likewise, two populations might have identical distributions, but very different central tendencies (**Figure 8.4**). Each will have its own meaning relative to the other, so it is important that you carefully consider which strategy offers the most value to your analyses. For example, you may not care about the average standard length of menhaden or croaker juveniles; it may be more important to study the overall variability within your measures of SL. If that were the case, you might be more interested to compare the standard deviations (or the variances) between the two populations.

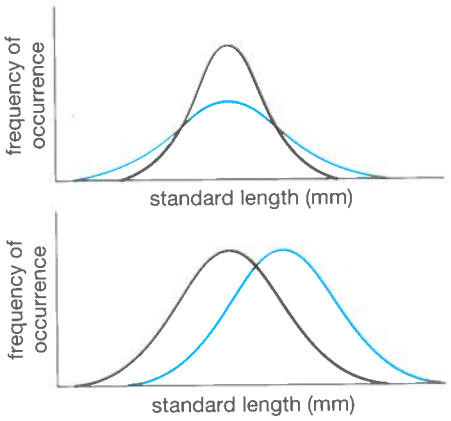


Figure 8.4 It is possible for two populations to share the same central tendency despite having very different distributions (top), just as it is possible for two populations to exhibit identical distributions but with different central tendencies (bottom).

The Most Common Statistical Tool for Univariate Analysis Is the *t* Test

As we have discussed in Chapter 2, there are a wide variety of statistical tests that are available to the researcher, depending on whether the data are normally distributed or not (see Figures 2.10–2.11). Although it is impossible to offer a "one size fits all" statistical test for the reader, what can be offered is a clear, concise review of one of the most common statistical methods used in the analysis of single-variant datasets. This of course would be the ***t* test**.

In the parlance of statistics, the *t* test is warranted for the analysis of single-variant data from one or more sampled populations, provided that the measures being compared are (1) normally distributed, (2) homoscedastic (exhibit equal variances), and (3) sampled independently from the two populations being compared. Hence, the *t* test is a parametric test, and a very powerful one at that. The power of the *t* test lies not only in its mathematical robustness, but more importantly, in its simplicity (thus making it rather easy to interpret the results). Essentially, when the *t* test is performed, a test statistic *t* is calculated from the sample mean ($\overline{X}$, using Equation 2.2) and analyzed for the population's deviation *d*, relative to the estimated standard error *SE*, as defined in Equation 8.1:

$$t = \frac{\overline{X} - d}{SE}, \quad \text{and} \quad SE = \frac{s}{\sqrt{N}} \tag{8.1}$$

where *N* represents the total number of observations in the dataset and assumes that the standard deviation *s* can be estimated from Equation 2.3. However, it is unnecessary for the reader to perform these calculations longhand using Equation 8.1, as all modern statistical software programs are capable of performing the *t* test statistic calculations as a routine operation.

In order for the *t* test to be of any value to the researcher, it is critical that an appropriate null and alternative hypothesis (H_0 and H_a, respectively) are defined prior to performing the statistical test. In Chapter 2, we learned that these hypotheses are always predefined for us, so we don't need to worry

about making any mistakes. For a two-tailed test (that is, a *t* test of equality), the hypotheses are always:

Two tailed $H_o: \bar{X}_n = \bar{X}_o$
$H_a: \bar{X}_n \neq \bar{X}_o$

For a one-tailed test (that is, a *t* test of inequality), we have to exercise a little more grey matter and decide which of the two possible states of inequality we'd like to test. They are:

Right tailed $H_o: \bar{X}_n \leq \bar{X}_o$
$H_a: \bar{X}_n > \bar{X}_o$
Left tailed $H_o: \bar{X}_n \geq \bar{X}_o$
$H_a: \bar{X}_n < \bar{X}_o$

Using these hypotheses, we are able to test the statistical significance of the mathematical expressions, comparing the mean of the *n*th population ($\bar{X}_n$) against the mean of some "original" population ($\bar{X}_o$). The beauty of the *t* test is that the test statistic will always return a probability value (*p*-value) between 0.00 and 1.00, the significance of which is established by your choice of acceptable probability level (α), so our choice is clear:

If $(p < \alpha)$: Accept H_a
Reject H_o
If $(p \geq \alpha)$: Accept H_o
Reject H_a

Recall that our comparisons need not be restricted to the mean. The *t* test is equally effective at testing the distribution of data by comparing either the standard deviation *s* or the variance s^2 of two populations.

Some of the best statistical software packages will automatically report *p*-values for all three tests (two tailed, right tailed, and left tailed); if so, you need only choose the *p*-value that corresponds to the type of tailed test in which you were most interested. Unfortunately, most statistical programs (particularly the freeware versions) only report the "two-tailed significance" by default.

If your statistical software only reports the two-tailed *p*-value, you must use the sign (+/−) of the test statistic *t* in order to calculate the left-tailed and right-tailed *p*-values (**Technical Box 8.2**). Although it is not a difficult task,

TECHNICAL BOX 8.2

How to Calculate *p*-Values for One-Tailed *t* Tests

Converting *p*-values: Two tailed → One tailed

	Test Statistic $(t) < 0$	Test Statistic $(t) > 0$
$H_a: \bar{X}_n > \bar{X}_o$ (Right tailed)	$p = 1.00 - \left(\frac{\textit{Two-tailed p-value}}{2}\right)$	$p = \left(\frac{\textit{Two-tailed p-value}}{2}\right)$
$H_a: \bar{X}_n < \bar{X}_o$ (Left tailed)	$p = \left(\frac{\textit{Two-tailed p-value}}{2}\right)$	$p = 1.00 - \left(\frac{\textit{Two-tailed p-value}}{2}\right)$

it requires that you have defined your right- and left-tailed alternate hypotheses H_a very meticulously.

The Appropriate Use of the *t* Test Depends on Three Critical Assumptions

Although the fundamental structure of the *t* test is always the same, the type of data comparisons you wish to perform will have an impact on which *t* test is most appropriate for your analysis. Remember that the *t* test can only test two populations at a time, and that the comparisons made between those two populations can only be done for a single variable. Fortunately, it is easy for us to determine the proper *t* test to use (**Figure 8.5**) once we have answered three simple questions:

1. How many sample populations are contained within the dataset?
2. Are the sampled populations independent of each other?
3. Are the standard deviations of the sampled populations equal?

Number of Sampled Populations: Remember that the fundamental goal of field-based research is to collect a subset of measurements that faithfully represent the larger population from which those measurements were taken. Most field research programs will conduct measurements of a few variables taken from multiple populations; however, it is also common practice to study one particular location (or population) in very great detail, with measurements conducted on a wide variety of different variables. Although the *t* test can only be used on a single variable at a time, it is important that you take the time to carefully consider whether your dataset truly represents a single population or more.

Independence of Sampled Populations: As we originally discussed in Chapter 2, an independent sample is one that, when its value is measured,

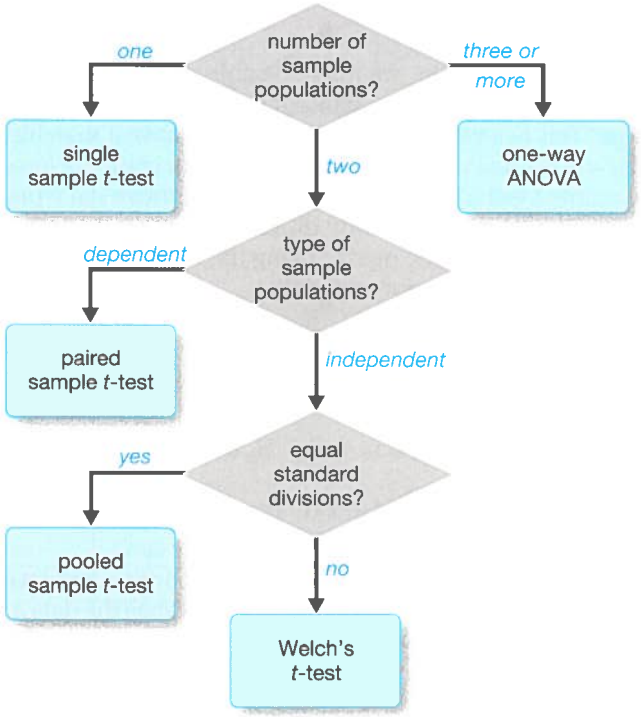


Figure 8.5 Flowchart to determine the proper statistical test for univariate analysis.

will not influence (or be influenced by) any other sample. In the context of population independence, you must carefully consider whether your populations might have a significant influence on each other. For example, if we were measuring wind speed at a number of ocean buoys within a few kilometers from each other, there might be some dependence there (such that the wind at one buoy might be blowing toward another buoy, thereby influencing the wind speed measured there). If we wanted to investigate winds at one buoy compared to another, we might have to assume those "populations" are dependent on one another. However, if we wanted to compare the winds on April 1st to those on May 1st, it's probably a safe bet to consider those populations as independent.

Standard Deviation Equality: If two different populations share the same standard deviation, they must also share the same variance, and therefore the exact same probability distribution. But just because two populations have the same distribution does not mean that they also share the same central tendency. That means we can still test for differences in the mean, median, and mode of our two populations.

The Single-Sample *t* Test Is Used to Compare Measurements Against a Historical or Theoretical Value

If the *t* test is used to make statistical comparisons between populations, it might seem a bit odd to consider using a *t* test on a single population. The single-sample *t* test is quite useful when you have data from only a single population but you wish to compare that population to some idealized example or to some theoretical value for which you do not have data.

For example, if you had collected daily precipitation data from a weather station in your hometown over an entire year, you would likely have enough data to ensure that they are normally distributed, thanks to the central limit theorem. Based on your sampling strategy, it is clear that we only have one population in our dataset: the daily precipitation data from a single location. From our rainfall data, we might conclude that the mean daily precipitation for the entire year was 0.76 cm day^{-1}. Of course, with 365 measurements of daily rainfall, we could also easily calculate the standard deviation and variance if we were so inclined.

If we look in the record books, we might be able to find a historic value that indicates that the mean daily precipitation, for our particular location, is 0.62 cm day^{-1}. This is a value for which we do not have a matching dataset; however, it is a historical value against which we can compare our data using the single-sample *t* test to determine whether our mean daily precipitation value (0.76 cm day^{-1}) is significantly different from the historical value of 0.62 cm day^{-1}. Thus, we define our $\bar{X}_o$ using the historical value and our $\bar{X}_a$ using the value derived from our own data:

$\bar{X}_o$: 0.62 cm day^{-1}
$\bar{X}_a$: 0.76 cm day^{-1}

Then it is a relatively simple task to define H_o and H_a for hypothesis testing.

The Paired-Sample *t* Test Is Used to Test the Dependence Between Two Populations

As we have already discussed, populations that are dependent on each other will have that dependence reflected in the data. Because the data are in some ways "related" to each other, we must first attempt to separate the effects of one population on the other prior to our analysis. This is most effectively

SINGLE-SAMPLE *t* TEST ASSUMPTIONS

1. Independence: The result of one observation must not influence the result of another observation in the dataset.
2. Single Population: Every observation within the dataset was collected from the same population.
3. Normal Distribution: The distribution of values within the dataset follows a normal (Gaussian) curve.

done by determining the magnitude of the differences between the two samples and conducting our test on those differences (rather than on the original measurements themselves).

Earlier in our discussion of population independence, we used the example of wind speed data collected at a variety of ocean buoys and theorized that the wind field at one buoy might influence the wind field at a neighboring buoy, thereby making our populations dependent. To investigate that potential dependence, we might calculate the difference between the average hourly wind speed ($\bar{v}$) measured at buoy 1 versus buoy 2 (**Figure 8.6**).

If we define $\Delta = \bar{v}_2 - \bar{v}_1$, we can perform our t test on the calculated variable Δ to determine whether the difference between $\bar{v}_1$ and $\bar{v}_2$ is significant or not. What we're really doing here is a single-sample t test on the difference between $\bar{v}_1$ and $\bar{v}_2$ (as $\bar{X}_\Delta$) and determining whether $\bar{X}_\Delta = 0$. Our hypotheses would then be stated as:

H_o: $\bar{X}_\Delta = 0$
H_a: $\bar{X}_\Delta \neq 0$

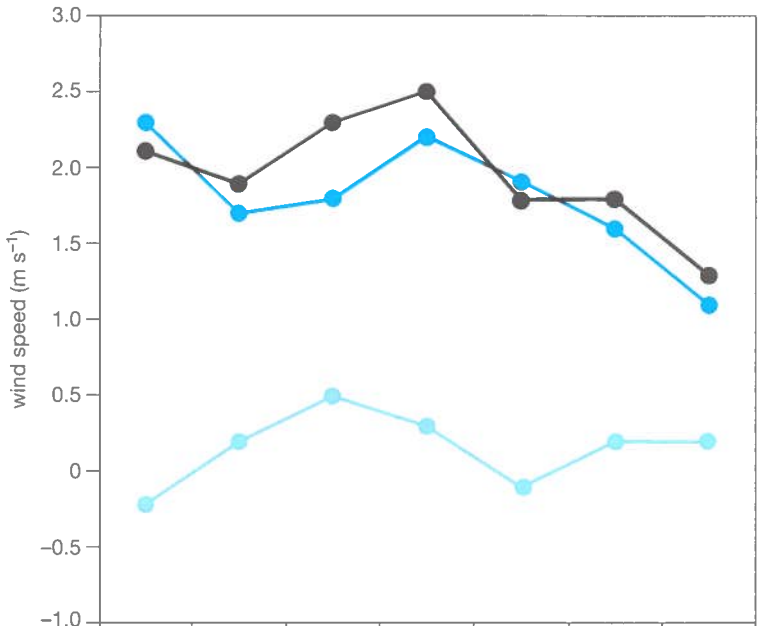
Of course, the example given in Figure 8.6 is meant to illustrate the general application of the paired t test. In reality, a sample size of $N = 7$ at each of our buoys is far too few for statistical analysis, and we would most likely be in violation of the assumption of normality.

Variance Is Used to Determine Whether to Use the Pooled-Sample or Welch's *t* Test

The classic pooled-sample t test and Welch's variation are each different from the previous t tests in that they both use the sample variance s^2, rather

PAIRED-SAMPLE *t* TEST ASSUMPTIONS

1. Independence: The data pairs used to calculate Δ must not influence (or be influenced by) any other data pairs in the dataset.
2. Single Population: Each pair of observations within the dataset was collected from the same population.
3. Normal Distribution: The distribution of values for Δ follows a normal (Gaussian) curve.



local time	12:00	13:00	14:00	15:00	16:00	17:00	18:00
buoy 1 (m s^{-1})	2.3	1.7	1.8	2.2	1.9	1.6	1.1
buoy 2 (m s^{-1})	2.1	1.9	2.3	2.5	1.8	1.8	1.3
Δ (m s^{-1})	−0.2	0.2	0.5	0.3	−0.1	0.2	0.2

Figure 8.6 The wind fields at buoy 1 and buoy 2 can be tested for dependence by taking the difference (Δ) between the average hourly wind speeds recorded at the two sites and performing a paired t test. If the winds recorded at buoy 1 have no influence on the winds recorded at buoy 2 (and vice versa), the paired t test should return a result indicating that $\Delta = 0$ (no dependence).

than central tendency, to test hypotheses of population equality. However, before you can decide which of these two tests is most applicable, you must first determine whether the standard deviations s in your populations are equal ($s_1 = s_2$). Fortunately, we already have all the tools we need to proceed.

Recall that the single-sample t test can be used to test the central tendency, standard deviation, or variance of any population against a theoretical value. If we have two populations, all that is necessary is to gather the descriptive statistics of population 1 in order to calculate its standard deviation (s_1) using Equation 2.3. This value can then be used as your theoretical value, which you can use to determine whether $s_2 = s_1$ using a single-sample t test to test the hypotheses:

H_o: $s_2 = s_1$ ($s_2^2 = s_1^2$)
H_a: $s_2 \neq s_1$ ($s_2^2 \neq s_1^2$)

If the resultant p-value $\geq \alpha$, you must accept H_o (that is, the standard deviations are indeed equal) and therefore proceed with the pooled-sample t test. Of course, if the p-value $< \alpha$, you must reject the null hypothesis H_o and accept H_a instead, which states that $s_2 \neq s_1$. If this is the case, you must proceed using Welch's t test.

The Pooled-Sample t Test Assumes Equal Variances Between the Populations Being Compared

The pooled-sample t test, sometimes called the "independent-samples t test," simply pools the standard deviation from both populations in order to calculate the test statistic t using Equation 8.2:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{s_P \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \tag{8.2}$$

Note that the means $\overline{X}_i$ and the total number of observations (N_i) of each population are used in conjunction with the pooled standard deviation s_P, calculated in Equation 8.3 as:

$$s_P = \sqrt{\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{(N_1 - N_2) - 2}} \tag{8.3}$$

Fortunately, the solutions to these equations are typically calculated by statistical software, which will automatically provide the user with a value for the test statistic t and the associated p-value; Equations 8.2–8.3 are provided merely for the reader's interest and information.

Remember that the only reason we can pool the standard deviations of both populations is because we just confirmed that $s_1 = s_2$; that is, the standard deviations are mathematically indistinguishable from each other. This means that the variances s^2 of each population are also equal. Hence, the only real functionality of the pooled-sample t test is to test the equality of the central tendency (usually $\overline{X}$) of population 1 versus population 2.

POOLED-SAMPLE t TEST ASSUMPTIONS

1. Independence: Every observation recorded in population 1 is independent from (that is, not influenced by) the observations recorded in population 2.
2. Equal Variance: The variances s^2 of both sampled populations are equal to each other.
3. Normal Distribution: Both populations exhibit normal (Gaussian) curves.

Welch's *t* Test Is Used When the Variances of Two Populations Are Not Equal

Welch's t test is very similar to the pooled-sample t test, except for the very important distinction that Welch's t test should be used for independent samples that do not share equal standard deviations or variances. For Welch's t test, the test statistic t is calculated using Equation 8.4:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \tag{8.4}$$

Remember that the applicability of Welch's t test presupposes that neither the standard deviations s nor the variances s^2 of our populations are equal. So, if we already know that $s_1 \neq s_2$ (and therefore $s_1^2 \neq s_2^2$), we can use Welch's t test to take these conclusions one step further and perform one-tailed tests on the standard deviations and variances to investigate:

Right tailed	H_o: $\bar{X}_1 \leq \bar{X}_2$, or $s_1 \leq s^2$, or $s_1^2 \leq s_2^2$
	H_a: $\bar{X}_1 > \bar{X}_2$, or $s_1 > s_2$, or $s_1^2 > s_2^2$
Left tailed	H_o: $\bar{X}_1 \geq \bar{X}_2$, or $s_1 \geq s_2$, or $s_1^2 \geq s_2^2$
	H_a: $\bar{X}_1 < \bar{X}_2$, or $s_1 < s_2$, or $s_1^2 < s_2^2$
Two tailed	H_o: $\bar{X}_n = \bar{X}_o$
	H_a: $\bar{X}_n \neq \bar{X}_o$

WELCH'S *t* TEST ASSUMPTIONS

1. Independence: All observations in population 1 are independent of all observations in population 2.
2. Single Population: All observations taken from population 1 belong to a single population, which may or may not be the same as population 2. But all observations taken from population 2 must also belong to a single population.
3. Normal Distribution: The distribution of values in both populations follow normal (Gaussian) curves.

The One-Way Analysis of Variance (ANOVA) Is Used to Compare Multiple Populations Simultaneously

Regardless of the actual name of the test, the one-way analysis of variance has absolutely nothing to do with the variance. Nor does it have anything to do with the standard deviation. In fact, the one-way analysis of variance (or one-way ANOVA) should actually be called the analysis of means (ANOME?), since that's exactly what it's used for.

The good news is that as far as statistical tests go, the ANOVA is remarkably easy to set up and perform. Essentially, the one-way ANOVA is used to test whether multiple sample means are in fact equal to each other. Thus, the ANOVA tests the hypotheses:

H_o: $\bar{X}_1 = \bar{X}_2 = \bar{X}_3 = \cdots = \bar{X}_n$

H_a: Not all $\bar{X}$ are equal

Essentially, the ANOVA is used to assess the differences among several means to investigate whether those differences are due to random variations in the populations (H_o) or whether those differences are significant enough to be attributed to something other than natural variation (H_a).

The one-way ANOVA can be a real time-saver when it comes to hypothesis testing multiple populations. If an ANOVA is performed and the p-value $> \alpha$, the result would indicate that the researcher must accept H_o and thereby proclaim that all of the means, from all sampled populations, are mathematically indistinguishable from each other.

However, if the p-value $< \alpha$, the result of the one-way ANOVA would indicate that at least one of the means is different from all the rest. It is also entirely possible that all of the means are different from each other. Unfortunately, the ANOVA cannot be used to determine which of the means are different from the rest, so any confirmation of H_a would dictate that the researcher would then have to compare one population to the next using the appropriate t test, in a stepwise fashion, until all possible comparisons have been made.

As an example, let us assume that we were comparing the means of six different populations. If this were the case, our hypotheses would be:

H_o: $\bar{X}_1 = \bar{X}_2 = \bar{X}_3 = \bar{X}_4 = \bar{X}_5 = \bar{X}_6$
H_a: Not all $\bar{X}$ are equal

If we had adopted $\alpha = 0.05$ and our one-way ANOVA returned a p-value $< \alpha$, we would be forced to accept H_a and would then have to compare all six populations, using an independent-samples t test (pooled or Welch's), for all possible combinations:

$\bar{X}_1 : \bar{X}_2$	$\bar{X}_2 : \bar{X}_3$	$\bar{X}_3 : \bar{X}_5$
$\bar{X}_1 : \bar{X}_3$	$\bar{X}_2 : \bar{X}_4$	$\bar{X}_3 : \bar{X}_6$
$\bar{X}_1 : \bar{X}_4$	$\bar{X}_2 : \bar{X}_5$	$\bar{X}_4 : \bar{X}_5$
$\bar{X}_1 : \bar{X}_5$	$\bar{X}_2 : \bar{X}_6$	$\bar{X}_4 : \bar{X}_6$
$\bar{X}_1 : \bar{X}_6$	$\bar{X}_3 : \bar{X}_4$	$\bar{X}_5 : \bar{X}_6$

That's 15 separate t test comparisons! What's worse, we are accepting a 5% chance of error ($\alpha = 0.05$) for each of the 15 comparisons we are making. If we use Equation 8.5 to consider the accumulation of all of those chances of error together (as an "experimentwise error rate," ε), our error becomes:

$$\varepsilon = 1 - (1 - \alpha)^n \tag{8.5}$$

where the total number of separate comparisons made (n) has a profound effect on ε. For our particular example, our experimentwise error rate would be:

$$\varepsilon = 1 - (1 - 0.05)^{15} \approx 1 - (0.46) \approx 54\% \tag{8.6}$$

A 54% error rate seems positively lousy.

ONE-WAY ANOVA ASSUMPTIONS

1. Independence: Every observation, regardless of the population from which it came, is independent from every other observation in the dataset.
2. Equal Variance: The variances s^2 of all sampled populations are equal to each other.
3. Normal Distribution: All sampled populations exhibit normal (Gaussian) curves.

Actually, ε refers to the probability that at least one of the comparisons would lead the researcher to mistakenly claim a significant difference between the means when in fact there was no significant difference (a Type I error). In that context, $\varepsilon = 54\%$ is not as disastrous as it sounds. Keep in mind that $\varepsilon = \alpha$ for each separate comparison, so on a case-by-case basis, our error rate is constrained to α.

Of course, if you wanted to define a smaller acceptable experimentwise error rate, it would require that you dramatically reduce your chosen value for α. In our example, if we wanted to limit ε to 10%, we simply rearrange Equation 8.5 and solve for α:

$$\alpha = 1 - (1 - \varepsilon)^{1/n} = 1 - (1 - 0.10)^{1/15} \approx 0.007 \tag{8.7}$$

Community Comparisons Using Univariate Analysis

As we have discussed in earlier chapters, there are a great variety of methods that can be used to collect physicochemical data from the substrate and/or aquatic medium. As long as we focus on a single variable at a time, these data can be analyzed using univariate analysis. Likewise, if we wished to focus on a single biometric variable (such as standard length), we could use univariate analysis for those comparisons as well. However, when we start looking at entire communities of organisms, those assemblages become quite complex, and the organisms themselves may have very different biometric variables. For example, the standard length SL may be a perfectly appropriate biometric for a variety of fish species, but our seagrass species would require a different biometric (such as leaf-blade length). If we wanted to investigate the entire community using univariate analysis, we would have to pick a single variable that would have universal applicability.

Generally, this is why ecologists use measures that include either counts (the number of individuals per species) or biomass (the amount of organic mass per species) to assess the biological community under investigation. It's a simple but very important concept: no matter what kind of critter we encounter, their "presence" can be measured as a count, and the relative "quantity" of that critter can be measured as biomass. These measures can then be easily related in terms of the two-dimensional (areal) or three-dimensional (volumetric) space they occupy, which gives us some idea of their species density within the community:

Density based on counts: $$\frac{\#\ individuals\ (unitless)}{area\ or\ volume\ occupied} \Rightarrow \frac{\#}{m^2} \text{ or } \frac{\#}{m^3}$$

Density based on mass: $$\frac{biomass\ (g)}{area\ or\ volume\ occupied} \Rightarrow \frac{g}{m^2} \text{ or } \frac{g}{m^3}$$

Species density (our scaled variable) can then be calculated for each species (our nominal variable). Since we're only analyzing for one scaled variable in this scenario, we can do so using univariate analysis.

Measures of Species Richness Are the Easiest Way to Compare Communities

Depending on the size of your study site, it may be possible to conduct a complete count, or complete collection, of all the organisms in the area. This method is particularly effective for areas with obvious boundaries or well-contained areas. However, we are seldom fortunate enough to be able to conduct direct and complete counts, as most study areas are far too expansive for us to attempt a complete count. Occasionally, the number of organisms counted or collected is far too large to justify the time commitment to count and/or measure the biometrics of each and every specimen. In such cases, we are often compelled to perform some kind of subsampling regime, such as the more typical plot and plotless methods described in Chapter 3.

Although the simplest census method is to survey a variety of locations and simply determine whether a particular species is present or absent, such qualitative studies are rarely useful because they lack ecological context and

provide no data regarding the relative abundance or biomass of a particular species in relation to any other species in the survey. Still, presence/absence (P/A) data are useful in providing a count of the total number of different species found in the area or system under investigation: this is what's known as **species richness** (S). Of course, richness can be assessed at any taxonomic level, which can be particularly useful if the investigator has difficulty in identifying organisms to the species level.

Simple Community Comparisons Can Be Performed Using Ranked Abundance Data and the DAFOR/SACFOR Methods

Beyond using simple P/A data to calculate species richness, the collection of frequency data (that is, how often a particular species is encountered) can be used to establish the relative abundance of each species within the community. This is typically performed using a ranking scheme that seeks to define the whether each recorded species is dominant, abundant, frequent, occasional, or rare (DAFOR) within the community. Although DAFOR ranks are established using a scaled variable to define the difference(s) between each rank, DAFOR frequencies are still considered to be semiquantitative assessments because it is entirely up to the observer to set the frequency conditions that define DAFOR.

For example, an investigator may choose to define DAFOR based on the frequency or biomass of each species in relation to the total amount assessed (that is, as a proportion of the total):

% of survey	*Rank*
80+	Dominant
40–79	Abundant
15–39	Frequent
5–14	Occasional
1–5	Rare

What this method lacks in objective quantitation, it more than makes up for in simplicity and ease of application in the field (**Figure 8.7**). Comparisons can still be made across study sites, or over time, so long as the observer maintains a consistent survey methodology and does not redefine the DAFOR conditions. This method can be used in either plot or plotless surveys and is well suited for simple, qualitative monitoring projects.

Timed DAFOR surveys can also be used to gather information on species richness and relative abundance, but they cannot be used to determine population densities or most other biometrics. To conduct a timed DAFOR survey, the observer defines a fixed time period (one hour, for example), which is then divided into equal time intervals that are given numerical ranks of decreasing magnitude:

Rank 6 (dominant) = 0–10 minutes
Rank 5 (abundant) = 11–20 minutes
Rank 4 (frequent) = 21–30 minutes
Rank 3 (occasional) = 31–40 minutes
Rank 2 (rare) = 41–50 minutes
Rank 1 (very rare) = 51–60 minutes

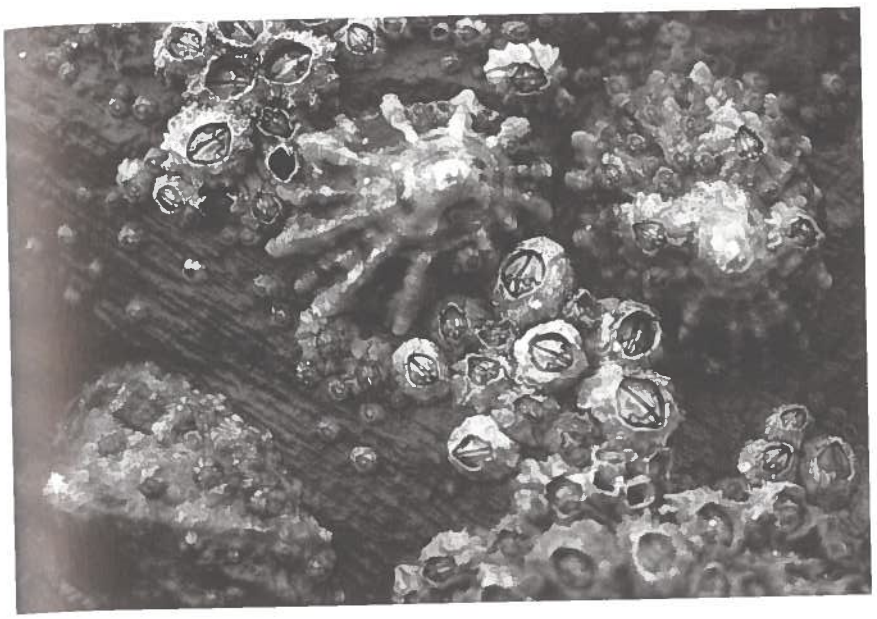


Figure 8.7 A photo-quadrat showing two different species of benthic organisms. We can see that there are 3 *Patella* limpet snails and 160 *Chthalamus* barnacles in the photo-quadrat (N = 163 total individuals, S = 2 species richness). If we were using simple counts for our DAFOR analysis, we would classify *Chthalamus* as Dominant (160 of 163, or 98.2%) and *Patella* as Rare (3 of 163, or 1.8%) in this particular example. (Courtesy of Mark A. Wilson.)

Once the observer begins the survey, the entire period is spent searching for previously unrecorded species that are recorded as belonging to the time interval in which they were first encountered. The fundamental assumption of this method is that the more abundant species are likely to be recorded in the earlier time intervals, whereas rare species will take much longer to find. Of course, this method does not account for organism behaviors, so timid or reclusive species may be recorded as "rare" regardless of whether they are truly less abundant. Taking several replicate surveys in the same area can improve the accuracy of results and is always recommended for DAFOR frequency assessments.

In some cases, it may be preferable to add additional ranks to the DAFOR scale. Another scale commonly used is the SACFOR scheme, where the ranks indicate whether a particular species is superabundant, abundant, common, frequent, occasional, or rare (SACFOR). Which of these is used, and how the ranks are defined, is completely at the discretion of the investigator. However, if previous DAFOR (or SACFOR) assessments have been made for your particular area of interest, you would be well advised to choose the identical scheme—that way, you would at least have the flexibility to compare your own survey results with previous investigations.

Biodiversity and Its Related Measures Are the Cornerstones of Community Comparison Methods

For decades, ecologists have sought to develop quantitative methods for assessing the "ecological health" of a particular area or habitat—an inherently qualitative judgment. In order to study community dynamics, it swiftly became evident that anecdotal observations of changes to the ecosystem were too subjective for scientific applicability. In an effort to objectively define the community structure within a particular ecosystem (and compare those results to a variety of investigations performed at different times or locations), ecologists devised several methods to assess community structure according to strict but simple mathematical protocols.

The most widely used indices of community structure are those that determine **species diversity** as a quantitative measure of the variety of species and

their numerical contribution to the community as a whole. Diversity analyses can take many mathematical forms, and as we have demonstrated, species richness S is the simplest index of diversity for us to measure. Although richness has its uses, it contains no information about the total number of individuals belonging to each of the species encountered, nor does it contain any information about the total number of individuals counted in the entire community. For example, if we can describe a community of fishes collected from trawl nets as having a species richness S of 8, all we would know is that there are 8 different species identified from that survey (**Figure 8.8**); we would know nothing about the number of individuals belonging to each species, nor would we know how many total specimens were collected from the trawls.

So it is clear that a better measure of species diversity (also known as **biodiversity**) should take into account both the number of species in the collection as well as the number of individuals belonging to each species. Just as we saw in the example of DAFOR surveys, some species in a collection will be abundant while others will be less common. If the species were all represented evenly in the collection, we would arguably have a perfectly diverse population. Conversely, when we see that a certain species is

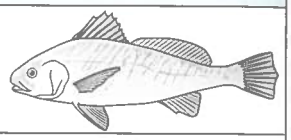
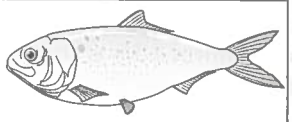
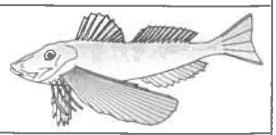
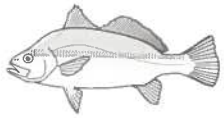
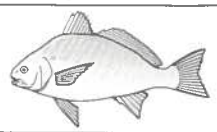
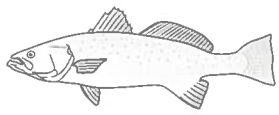
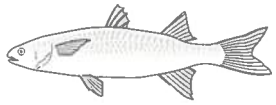
common name	species	
Atlantic croaker (croaker, hardhead)	*Micropogonias undulatus*	
Atlantic menhaden (alewife, bunker, pogy, bugmouth, fat-back)	*Brevoortia tyrannus*	
butterfish	*Peprilus triacanthus*	
northern sea robin (sea robin)	*Prionotus carolinus*	
silver perch (perch, sand perch)	*Bairdiella chrysoura*	
spot (norfolk spot, yellowbelly)	*Leiostomus xanthurus*	
spotted seatrout (speckled trout, spotted trout, speckle)	*Cynoscion nebulosus*	
striped mullet (mullet, jumping mullet)	*Mugil cephalus*	

Figure 8.8 The number of different species identified within a particular community can provide a very simple index of biodiversity, known as the species richness S. This particular collection of fishes from the Atlantic Ocean includes eight ($S = 8$) different species. (Courtesy of the Maryland Department of Natural Resources.)

dominant while others are quite rare (a condition of unequal representation), the population is said to be less diverse.

Fortunately, there are several methods for calculating the unequal representation of species within a collection, as species diversity. One of the most widely used indices of diversity is the Shannon–Weaver index (see Equation 8.8), which provides a quantification of species diversity (H') as:

$$H' = -\sum \left(\frac{N_i}{N}\right) \cdot log_{10}\left(\frac{N_i}{N}\right) \tag{8.8}$$

where N_i is the number of individuals belonging to the ith species and N is the total number of individuals in the collection. As an alternative to H', **species dominance** calculates the probability that two individuals randomly selected from the community will actually belong to the same species. This probability is typically calculated using Equation 8.9, which represents the Simpson index of species dominance (ℓ):

$$\ell = \sum \left(\frac{N_i}{N}\right)^2 \tag{8.9}$$

Simpson diversity (D_S), which is analogous to H', can be easily calculated using Equation 8.10, once ℓ is known:

$$D_S = 1 - \ell \tag{8.10}$$

Since both values of H' and D_S are sensitive to the overall species richness S of the collection, maximum species diversity is reached as $N_i \rightarrow (N/S)$. This is an important point, because the values for H' and D_S will be meaningless unless you calculate just how diverse the population is (H' and D_S) when compared to the theoretical maximum diversity (H'_{max} and D_{max}), defined in Equations 8.11–8.12 as:

$$H'_{max} = log_{10}\, S \tag{8.11}$$

$$D_{max} = 1 - \left(\frac{1}{S}\right) \tag{8.12}$$

Once the diversity measures (H' and D_S) and their theoretical maxima (H'_{max} and D_{max}) are calculated, it is a relatively simple task to determine **species evenness** using Equations 8.13–8.14, which define species evenness as the mathematical "nearness" of the observed diversity to the theoretical maximum diversity:

$$J' = \frac{H'}{H'_{max}} \tag{8.13}$$

$$E_S = \frac{D_S}{D_{max}} \tag{8.14}$$

where J' and E_S are the calculated evenness values for the Shannon–Weaver (H') and Simpson (D_S) indices of diversity, respectively.

EXAMPLE BOX 8.1

Using Specimen Counts to Compute Species Richness, Diversity, Dominance, and Evenness

Let's assume we had conducted a number of trawls at three different coastal sites and carefully identified and enumerated each collected specimen to yield the data below. If we were interested in characterizing the diversity and evenness of each site, we would then be able to make some univariate comparisons between the sites. Not only would we be able to compare the sites immediately, if we returned to the same locations at a later date and conducted the same trawl surveys, we would then be able to make some "before-and-after" comparisons as well. Consider our data in **Table 8.4**.

In the context of univariate analysis, diversity and evenness calculations are extremely valuable because they represent an entire community of organisms that have been mathematically reduced to a single scaled variable. As long as the method of data collection is consistent, diversity and evenness measures can be compared between different sites, or between different time periods, using univariate statistical methods like the *t* test or ANOVA (assuming of course that our data do not violate any of the critical assumptions of those parametric tests). Even if statistical comparisons are not possible, diversity and evenness measures still represent summary data of complex biological communities, ripe for comparison.

Table 8.4 Identification and Enumeration of Nekton Specimens Collected from Coastal Research Stations A–C

Common name	Species name	N_i at site A	N_i at site B	N_i at site C
Atlantic croaker	*Micropogonias undulatus*	51	27	11
Atlantic menhaden	*Brevoortia tyrannus*	8	32	17
butterfish	*Peprilus triacanthus*	9	0	0
northern sea robin	*Prionotus carolinus*	0	2	5
silver perch	*Bairdiella chrysoura*	0	14	18
spot croaker	*Leiostomus xanthurus*	44	21	28
spotted seatrout	*Cynoscion nebulosus*	9	2	0
striped mullet	*Mugil cephalus*	3	1	2
N		124	99	81
S		6	7	6
H'		0.600	0.664	0.679
H'_{max}		0.778	0.845	0.778
J'		0.771	0.786	0.873
ℓ		0.310	0.245	0.236
D_S		0.690	0.755	0.764
D_{max}		0.833	0.857	0.833
E_S		0.828	0.881	0.917

Although site A had the greatest abundance of fish (N), it was more heavily dominated by a few particular species, as evidenced by the largest value for Simpson dominance ($\ell = 0.320$). Indeed, site A has the lowest values for Shannon-Weaver ($H' = 0.600$) and Simpson ($D_S = 0.690$) diversity, compared to the other sites.

Although site B had a slightly higher species richness ($S = 7$) than the other locations, site C actually exhibited the highest species diversity, regardless of whether the Shannon-Weaver ($H' = 0.679$) or Simpson ($D_S = 0.764$) diversity indices were used. This is particularly evident if we consider that the evenness calculation at site C estimates that the biodiversity measured there was within 87.3% ($J' = 0.873$) to 91.7% ($E_S = 0.917$) of the theoretical maximum, the using Shannon-Weaver and Simpson diversity measures, respectively.

Using the Before-After-Control-Impact (BACI) Approach

One of the simplest and most widely used methods for performing univariate comparisons is the before-after-Control-Impact (BACI) design. This method involves the collection of data before some event (to serve as the baseline), followed by subsequent measures of the same variable immediately after the event occurs. In this way, the measures of a single variable can be assessed in pairwise fashion to detect the "before-and-after" differences using a two-sample. However, one of the major difficulties in this type of experimental setup is that it is unlikely you will be able to anticipate exactly when (or where) the event will occur. In order to compensate for this, it is often necessary to establish at least two study sites: a Control site and an Impact site. Of course, the impact site serves as the location where the event is anticipated to occur. Your Control site should be as similar as possible to the Impact site, but located where it is highly unlikely, in any circumstance, to be affected by the event you wish to study. This way, you have all your bases covered: time-sensitive changes will be detected in the before versus after data, whereas spatially sensitive changes will be detected in the Control versus Impact data.

As an example, let's assume there are plans to build a titanium dioxide (TiO_2) plant near the coast, and there are some concerns that TiO_2 pollution may adversely affect the nearshore environment. If you suspect there may be an Impact to the environment after the factory is built and becomes operational, it would be wise to analyze the coastal environment for various pollutants before the plant is built, to serve as your baseline data. To make sure there are no spatial differences in the baseline pollutant data, you decide to locate your Control site 10 km upcoast from the anticipated Impact site. Once the baseline data at both sites are measured, any event-related changes (for example, one year after the TiO_2 plant became operational) should be relatively easy to quantify and test for significance against the baseline (**Figure 8.9**).

Event-related changes (sometimes called perturbations) to the system are essentially the same as performing an experiment in the laboratory, or delivering a treatment to a test subject—in either case, you are essentially looking to quantify the "before-and-after" changes. Although our pollution example is most certainly a negative effect on the ecosystem, keep in mind that not all perturbations are negative. It is possible to perform the same kind of perturbation analysis on positive effects, like increased survivability against infection post-treatment with antibiotics, or increased water clarity after the creation of managed oyster reefs.

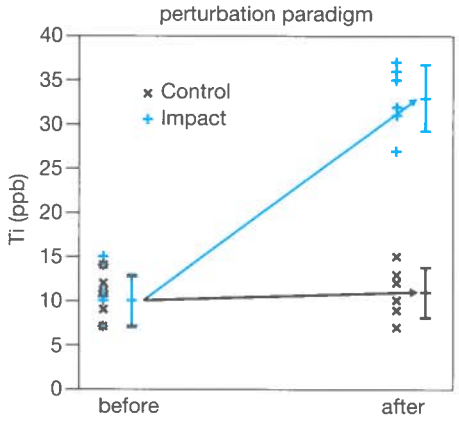


Figure 8.9 The classic BACI perturbation paradigm, where the Control and Impact sites exhibit very similar baseline measures, before the system is disturbed by some event. Time-related changes, unrelated to the disturbance event, can be measured by making comparisons between the "before-and-after" conditions within the Control site. Spatially related changes (presumably from the perturbation event) can be measured by comparing differences between the Control and Impact sites. In this example, the concentration of dissolved titanium (Ti, in parts per billion) measured in the coastal waters near a TiO_2 plant strongly indicates an Impact to the system when compared to the Control.

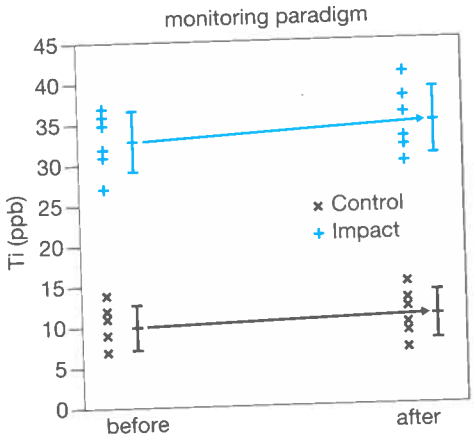


Figure 8.10 The classic BACI monitoring paradigm, where the Control and Impact sites may exhibit very different baseline measures, but the time-related changes measured at each site can provide context to the "natural" state of change at each site. Monitoring efforts are also valuable because they essentially establish the new baseline for each site under investigation. In this example, the concentration of dissolved titanium (Ti, in parts per billion) measured at the Impact site exceeds the contamination measured at the Control site, but the time-related changes at each site are relatively stable. If a new perturbation event occurred, these data would serve as the "new" baseline for all future analyses. Without the benefit of baseline data acquired from monitoring programs, many temporal and spatial comparisons would not be possible.

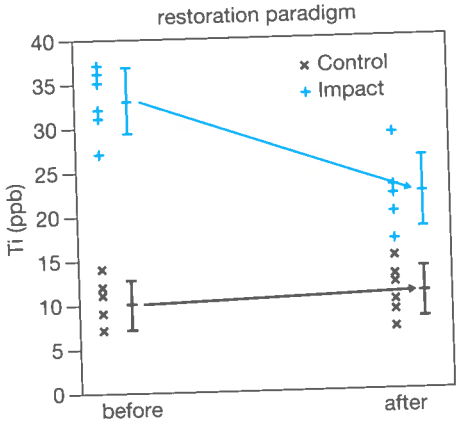


Figure 8.11 The classic BACI restoration paradigm, where the Control and Impact sites may exhibit very different baseline measures, but the effectiveness of restoration efforts (or natural recovery) can be quantified as the Impact site returns to its original, pristine state (represented by the Control site). In this example, coastal clean-up efforts (and reductions in waste outfalls) have significantly reduced the titanium contamination (Ti, in parts per billion) at the Impact site. Although the Impact site has not yet fully recovered (using the Control site as our standard), this data could be used to test the effectiveness of the clean-up efforts, or provide an estimate of how much longer it will take before the Impact site is fully recovered.

The BACI design is also used quite heavily in monitoring programs, where the intent is not to describe event-related changes between the Control and Impact site, but to simply monitor the sites over time. Such studies may be appropriate for locations where a perturbation event has already occurred, but it is still necessary to monitor how the system is responding over time. Monitoring programs are often underappreciated, because there is the misconception that there is no research going on; that data are merely being collected for the sake of collection. However, monitoring studies do provide a means to demonstrate how similarly (or dissimilarly) the Control and Impact sites will change over time, under their respective "natural" conditions (**Figure 8.10**). Monitoring programs are also beneficial because we can rarely anticipate when (or where) perturbation events will occur next, so data collected from monitoring programs can be used as the critical baseline when new perturbations occur.

The BACI design is also quite effective in measuring the effectiveness of restoration efforts and/or ecosystem recovery after a perturbation event. Whether they benefit from natural or anthropogenic intervention, natural systems are capable of recovering from all but the most grievous perturbation events. Measuring the capacity for ecosystem recovery, either as a function of time or location, can be accomplished using the BACI design by simply analyzing the differences between the control and impact sites before recovery, and then again at some future time (**Figure 8.11**). Such analyses are used to estimate the amount of time necessary for the impact site to return to its natural state. If specific restoration efforts were employed in an effort to return the site to its original, pristine condition, the effectiveness of those methods can also be analyzed in this fashion.

What is critical to the BACI design, whether it is applied as a perturbation, monitoring, or restoration paradigm, is that all comparisons must be conducted in pairwise fashion (control vs. impact, or before vs. after) to provide the necessary univariate evidence of correlation and/or causation. The inclusion of multiple sites, multiple treatments, or multiple time series does not make such analyses impossible. Quite the contrary: multivariate analyses often provide much stronger (and more useful) evidence of correlation and causation. But those are methods that involve more than one scaled variable and, as such, require a completely different class of statistical tests for hypothesis testing.

Moving Beyond the Simplicity of Univariate Analysis

As we have seen, univariate analyses can be quite useful, especially if we are interested in making simple comparisons between one population and

another. These comparison strategies can be applied to any field within the natural sciences, whether you're comparing grain size between two different sediment cores, the concentration of mercury within two different estuaries, the significant wave heights measured at two different ocean buoys, or the number of fish species collected in April compared to November. Such pairwise comparisons can be used to establish the statistical foundation on which more thoughtful and more rigorous comparisons can be made.

As we have already discussed, the greatest weaknesses of univariate comparisons are twofold: (1) we cannot compare more than one variable at a time; and (2) it is more difficult to establish association or causation without the use of multivariate comparisons. Think about that second point for a moment. How can we possibly hope to form even the simplest association (if A then B) if we cannot analyze A and B in tandem as scaled variables?

For example, with univariate analyses, we might be able to demonstrate mathematically that the concentration of mercury in estuary 1 is greater than estuary 2. We might follow up our analyses and demonstrate that disease rates among fishes in estuary 1 are also greater than in estuary 2. We might be tempted to infer that those increased disease rates are coincident with increased mercury levels, but without the capacity to analyze both variables simultaneously, our conclusions would be little more than innuendo. It would be a reasonable assumption, to be sure; but there would be no direct statistical evidence of that correlation. For that, we will need to expand our horizons to include multivariate analyses and the power to resolve such associations.

As we will discover in the next chapter, multivariate analyses ultimately allow us to establish both association and causation. And by the very act of demonstrating cause and effect in the language of mathematics (that is, statistics), we shall discover that we are at the same time embarking on the process of modeling our natural world.

References

Jones ER (1996) Statistical Methods in Research. Edward R. Jones.

Kanji GK (1999) 100 Statistical Tests. SAGE Publications.

Keeping ES (1995) Introduction to Statistical Inference. Dover Publications.

Kenny DA (1979) Correlation and Causality. John Wiley and Sons.

Linton M, Gallo Jr PS, & Logan CA (1975) The Practical Statistician: Simplified Handbook of Statistics. Wadsworth Publishing Company.

Mandel J (1964) The Statistical Analysis of Experimental Data. Dover Publications.

Salkind NJ (2007) Statistics for People Who (Think They) Hate Statistics: The Excel Edition. SAGE Publications.

Steiner F (ed) (1997) Optimum Methods in Statistics. Akadémiai Kiadó.

Thompson SK (1992) Sampling. John Wiley and Sons.

Further Reading

Jaisingh LR (2006) Statistics for the Utterly Confused, 2nd ed. McGraw-Hill.

Kachigan SK (1986) Statistical Analysis: An Interdisciplinary Introduction to Univariate and Multivariate Methods. Radius Books.

Keller DK (2006) The Tao of Statistics. SAGE Publications.

Newman I & Newman C (1977) Conceptual Statistics for Beginners. University Press of America.

Timm NH & Mieczkowski TA (1997) Univariate and Multivariate General Linear Models. SAS Institute.

Turner JR & Thayer JF (2001) Introduction to Analysis of Variance: Design, Analysis, and Interpretation. Sage Publications.

Wei WWS (2005) Time Series Analysis: Univariate and Multivariate Methods, 2nd ed. Pearson.