Chapter 1

Describing a Population

Figure 1.1 A fish population is sampled by seining.

INTRODUCTION

Ecology is the ambitious attempt to understand life on a grand scale. We know that the mechanics of the living world are too vast to see from a single vantage point, too gradual to observe in a single lifetime, and too complex to capture in a single narrative. This is why ecology has always been a quantitative discipline. Measurement empowers ecologists because our measuring instruments extend our senses, and numerical records extend our capacity for observation. With measurement data, we can compare the growth rates of trees across a continent, through a series of environmental conditions, or over a period of years. Imagine trying to compare from memory the water clarity of two streams located on different continents visited in separate years, and you can easily appreciate the value of measurement.

Numerical data extend our capacity for judgment too. Since a stream runs muddier after a rain and clearer in periods of drought, how could you possibly wade into two streams in different seasons and hope to develop a valid comparison? In a world characterized by change, data sets provide reliability unrealized by single observations. Quantitative concepts such as averages, ratios, variances, and probabilities reveal ecological patterns that would otherwise remain unseen and unknowable. Mathematics, more than any cleverly crafted lens or detector, has opened our window on the universe. It is not the intention of this text to showcase math for its own sake, but we will take measurements and make calculations because this is the simplest and most powerful way to examine populations, communities, and ecosystems.

Sampling

To demonstrate the power of quantitative description in ecology, you will use a series of measurements and calculations to characterize a **population**. In biology, a population is defined as a group of individuals of the same species living in the same place and time. Statisticians have a more general definition of a population, that is, all of the members of any group of people, organisms, or things under investigation. For the ecologist, the biological population is frequently the subject of investigation, so our biological population can be a statistical population as well.

Think about a population of red-ear sunfish in a freshwater lake. Since the population's members may vary in age, physical condition, or genetic characteristics, we must observe more than one representative before we can say much about the sunfish population as a group. When the population is too large to catch every fish in the lake, we must settle for a **sample** of individuals to represent the whole. This poses an interesting challenge for the ecologist: how many individuals must we observe to ensure that we have adequately addressed the variation that exists in the entire population? How can this sample be collected as a fair representation of the whole? Ecologists try to avoid **bias**, or sampling flaws that overrepresent individuals of one type and underrepresent others. If we caught our sample of sunfish with baited hooks, for example, we might selectively capture individuals large enough to take the bait, while leaving out smaller fish. Any estimates of size or age we made from this biased sample would poorly represent the population we are trying to study.

After collecting our sample, we can measure each individual and then use these measurements to develop an idea about the population. If we are interested in fish size, we could measure each of our captured sunfish from snout to tail (Figure 1.2). Reporting every single measurement in a data table would be truthful, but not very useful, because the human mind cannot easily take in long lists of numbers. A more fruitful approach is to take all the measurements of our fish and systematically construct a composite numerical description, or **statistic**, which conveys information about the population in a more concise form. The average length (also called the **mean** length) is a familiar way to represent the size of a typical individual. We might find, for instance, that the mean length of sunfish in this lake is 12.07 centimeters, based on a sample of 80 netted fish. The symbol μ is used for the mean of all fish in the population, which we are trying to estimate in our study. The symbol $\bar{\mathbf{x}}$ is used for the mean of our sample, which we hope to be close to μ.

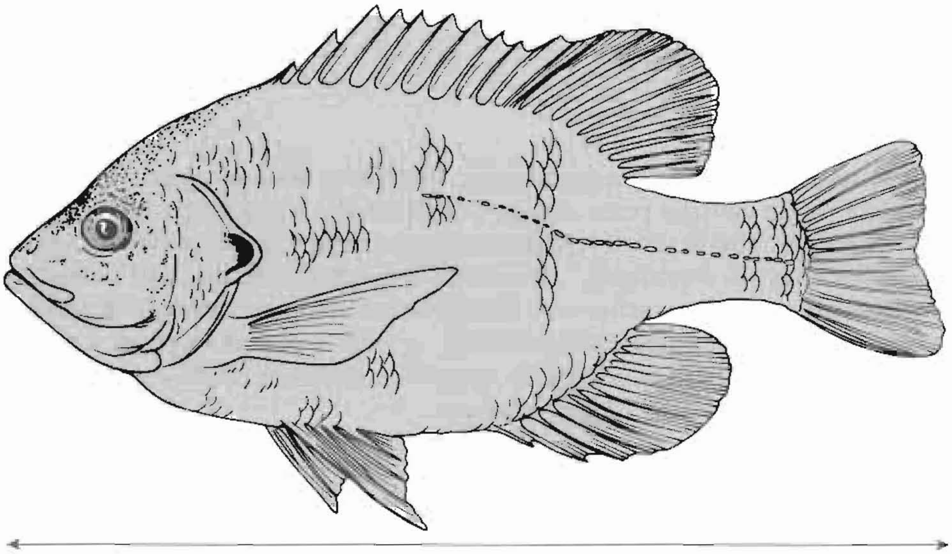


Figure 1.2 Measuring length of red-ear sunfish.

Means are useful, but they can be misleading. If a population were made up of small one-year-old fish and much larger two-year-old fish, the mean we calculate may fall somewhere between the large and small size classes—above any of the small fish, but below any of the large ones. A mean evokes a concept of the "typical" fish, but the "typical" fish may not actually exist in the population (Figure 1.3).

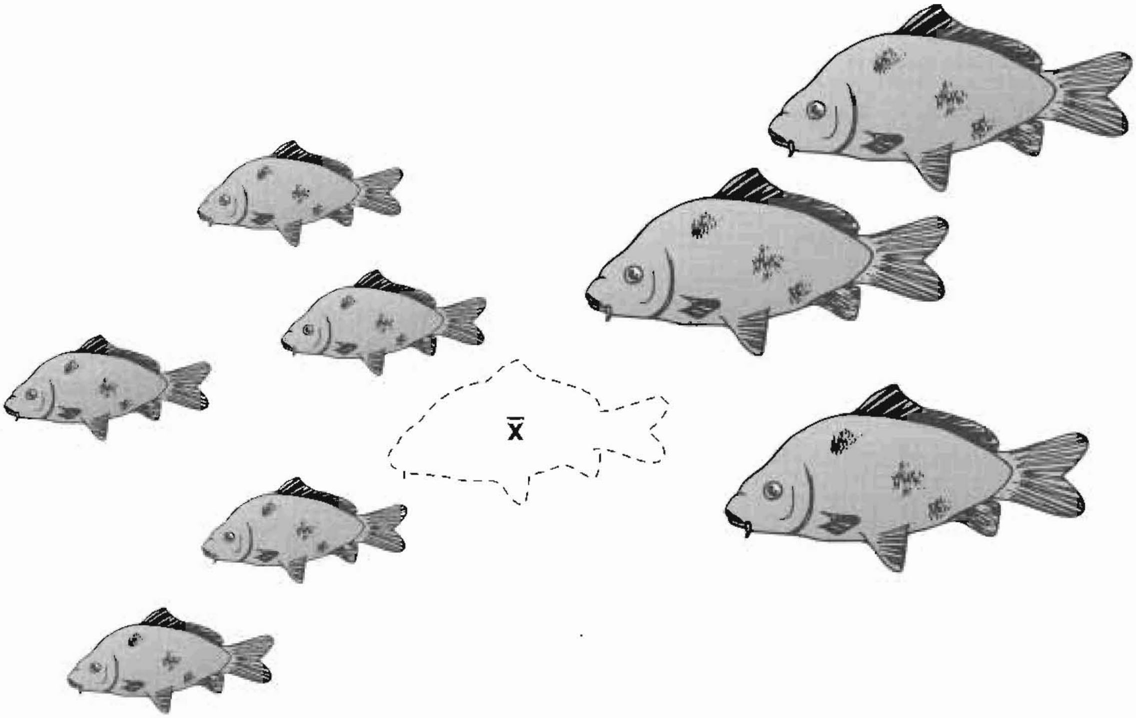


Figure 1.3 The calculated mean describes a "typical" fish that does not actually exist in a population composed of two size classes.

For this reason, it is often helpful to use more than one statistic in our description of the typical member of a population. One useful alternative is the **median**, which is the individual ranked at the 50th percentile when all data are arranged in numerical order. Another is the **mode**, which is the most commonly observed length of all fish in the sample.

Picturing Variation

After calculating statistics to represent the typical individual, it is still necessary to consider variation among members of the population. A **histogram** is a simple graphic representation of the way individuals in the population vary.

Check your progress:

If you wanted to determine the average height of students on your campus, how would you select a sample of students to measure?

Hint: Avoid statistical bias by making sure every student on campus has an equal chance of inclusion in the sample.

To produce a histogram:

1. Choose a measurement variable, such as length in our red-ear sunfish. Assume we have collected 80 sunfish and measured each fish to the nearest millimeter.
2. On a number line, mark the longest and shortest measurements taken from the population (Figure 1.4). The distance on the number line between these points, determined by subtracting the smallest from the largest, is called the **range**. In our example, the longest fish measures 16.3 cm, and the shortest fish measures 8.5 cm, so the range is 7.8 cm.

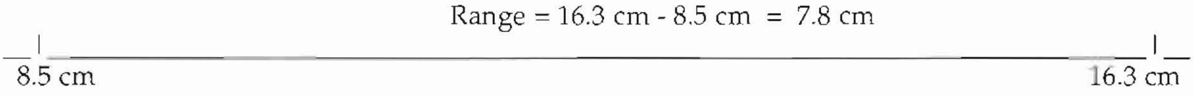


Figure 1.4 Number line indicating range.

3. Next, divide the range into evenly spaced divisions (Figure 1.5). In our example, each division of the number line represents a **size class**. It is customary to use between 10 and 20 size classes in a histogram. For our example, we will divide the range of sunfish sizes into 16 units of 0.5 cm each. The first size class includes fish of sizes 8.5 through 8.9 cm. The next size class includes fish of sizes 9.0 through 9.4 cm, and so forth to the last size class, which includes sizes 16.0–16.4.

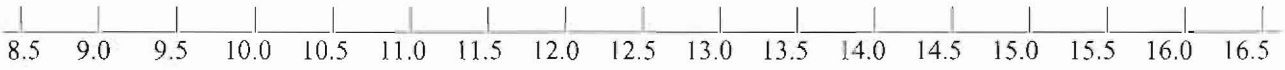


Figure 1.5 Number line divided into size classes of 0.5 cm.

4. Having established these size classes, it is possible to look back at our measurement data and count how many fish fall into each class. The number of individuals falling within a class is the **frequency** of that class in the population. By representing each measurement with an X, as shown in Figure 1.6, we can illustrate frequencies on the number line.
5. On the completed histogram illustrated in Figure 1.7, the stack of X-marks is filled in as a vertical bar. The height of each bar represents the proportion of the entire population that falls within a given size class.

Check your progress:

In the sample described by the histogram (Figure 1.7), how many fish measured between 10.0 and 10.4 cm?

Answer: 4

Figure 1.6 Counting frequencies.

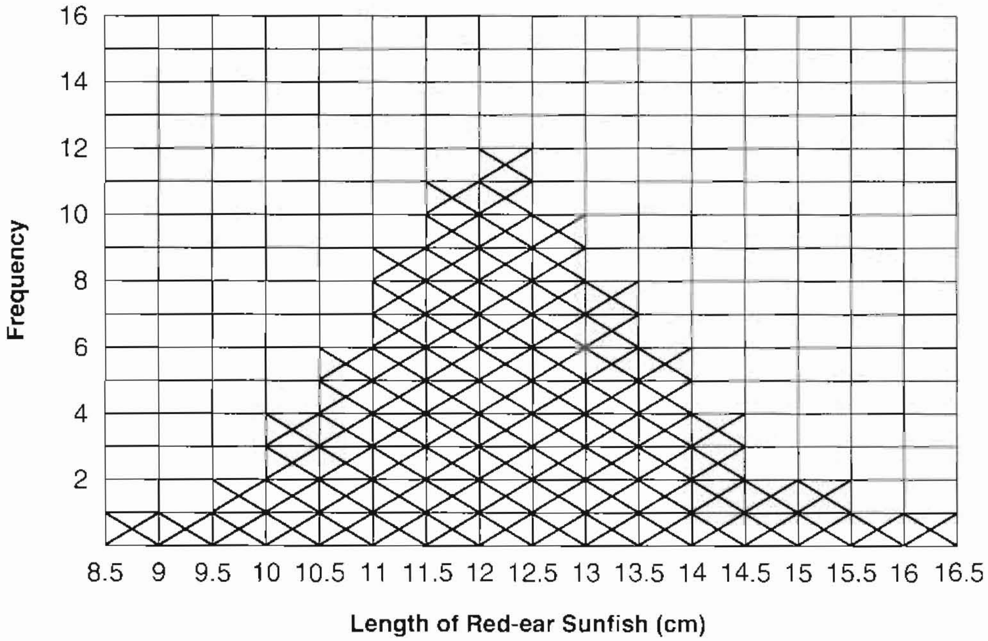
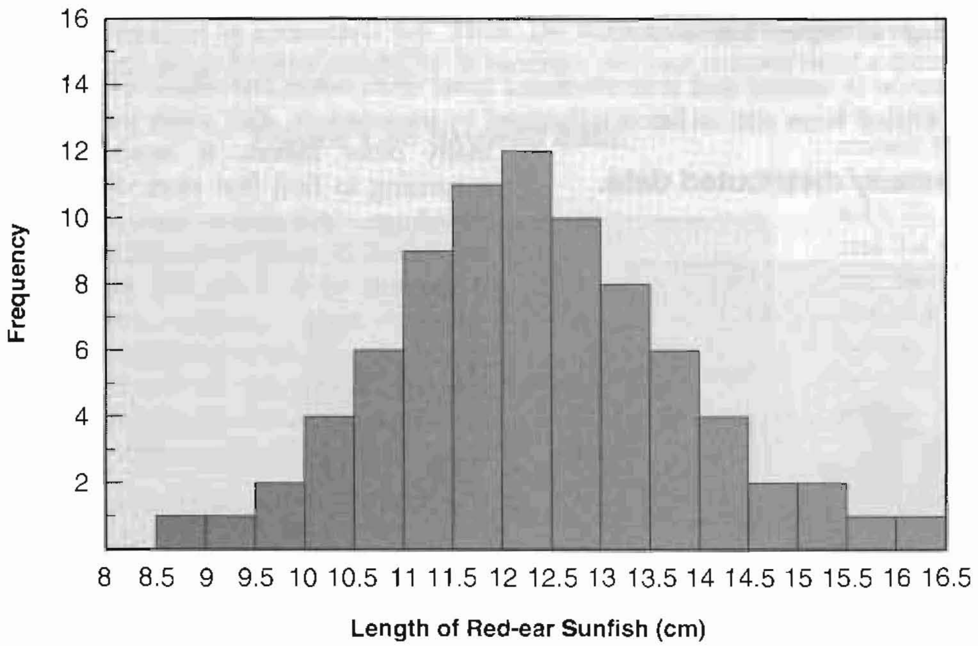


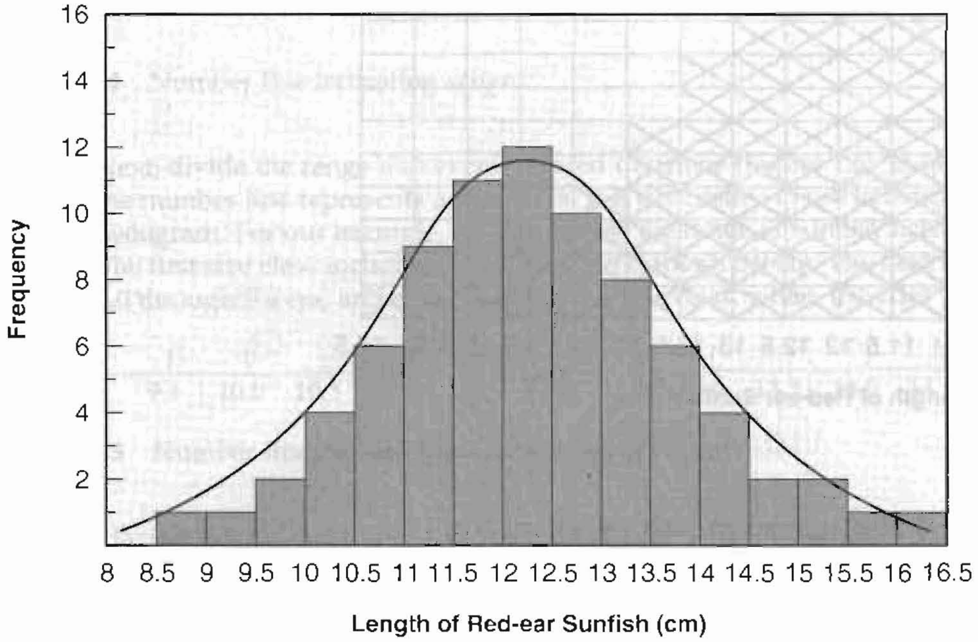
Figure 1.7 Histogram of fish lengths.



Describing the Pattern of Variation

Notice that in our sample histogram, the most common size classes are near the middle of the distribution (Figure 1.7). Extremely large and extremely small fish are rare, while intermediate sizes are more common. The "skyline" of the histogram fits under a bell-shaped curve that is symmetrical, and has characteristic rounded "shoulders" and "tails" that taper to the ends of the range in a predictable way. Statisticians call this shape a **normal distribution** (Figure 1.8).

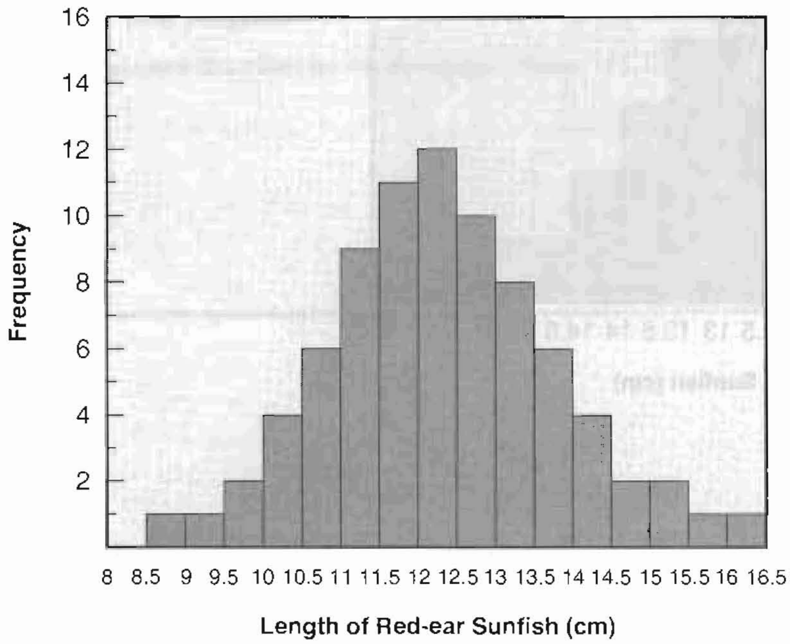
Figure 1.8 Normally distributed data.



This pattern of variation is common in nature, and is encountered quite often when one effect is influenced by many independently acting causes. Since size in fish is influenced by temperature, diet, water purity, and many other factors, it would not be surprising to find that sizes of fish in a mixed-age population are normally distributed. Because the normal distribution is encountered so commonly, many of the statistical tools ecologists use to test hypotheses assume that variations in their data are distributed in this bell-shaped form. Models and tests based on this kind of distribution are called **parametric statistics**.

If the histogram of variation is lopsided, has more than one peak, or is too broad or too narrow, then parametric tests should not be used (Figure 1.9). **Non-parametric tests** have been developed for these kinds of data. Because the nature of variation in your measurements is critical to further analysis, it is always a good idea to draw a histogram and compare your data to a normal distribution before taking your analysis any farther.

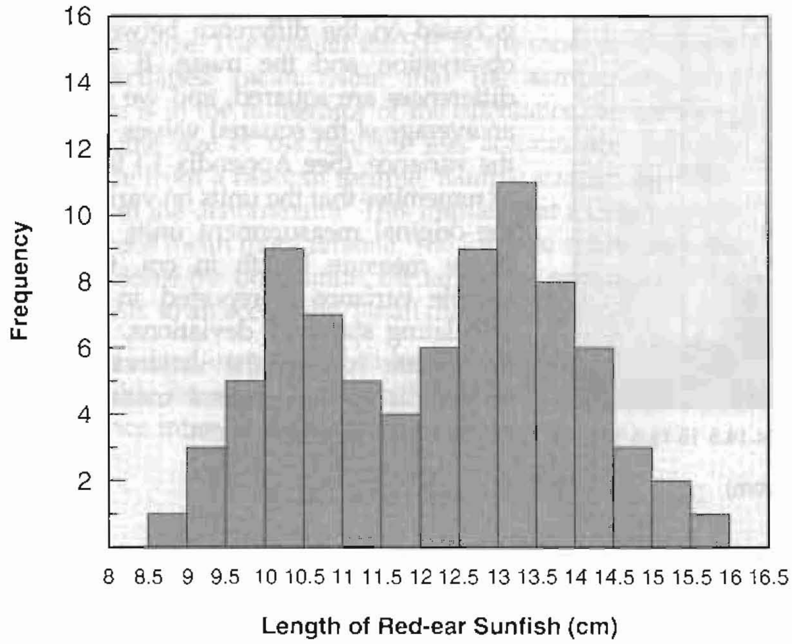
Figure 1.9a Normally distributed data.



Measuring Variation

How trustworthy is the mean that we calculate from our sample of fish? Two factors come into play. First, the **sample size** is critical. If fish in the lake vary a lot in size, a mean calculated from a small sample (say 10 fish) might be significantly off the mark. By chance, the 10 fish you caught might be larger or smaller than the average population size you are trying to describe. If the sample is expanded to 1000, it is much more likely that your calculated mean will accurately reflect the population average. *A fundamental principle of data collection is that the sample size must be large enough to eliminate sampling errors due to chance departures from the population mean.* To keep our thoughts straight, we use n = **size of the sample**, and **N = size of the entire population**. N is usually unknown, but can be estimated in a number of ways. (See Chapter 4.)

Figure 1.9b Bimodally distributed data.



How large, then, must a sample be? This depends on the amount of variation in the population. Samples taken from a fish farm where all the fish are nearly the same size will give reliable estimates, even if the sample is small. In a natural population with a great range of sizes, the sample has to be expanded to ensure that the larger variation is accounted for. Thus, *the more variable the population, the larger the sample must be* to achieve the same level of reliability. It becomes obvious that we need a statistic to measure variation.

To measure the amount of variation around the mean, we use a statistic called the **standard deviation (abbreviated s.d.)**. To distinguish between our sample and the entire population, we define s = **the standard deviation of the sample**, and σ = **the standard deviation of the whole population**. The standard deviation is expressed in the same units as the original measurements, which would be cm in our hypothetical fish study. A standard deviation can thus be shown as a portion of the range on a number line. In normally distributed populations, 95% of all individuals fall within 1.96 standard deviations from the mean. This means that the X-axis of a histogram can be marked off in four standard deviation units (two above the mean, and two below), and roughly 95% of the observations will fall within that region (Figure 1.10).

Figure 1.9c Skewed data.

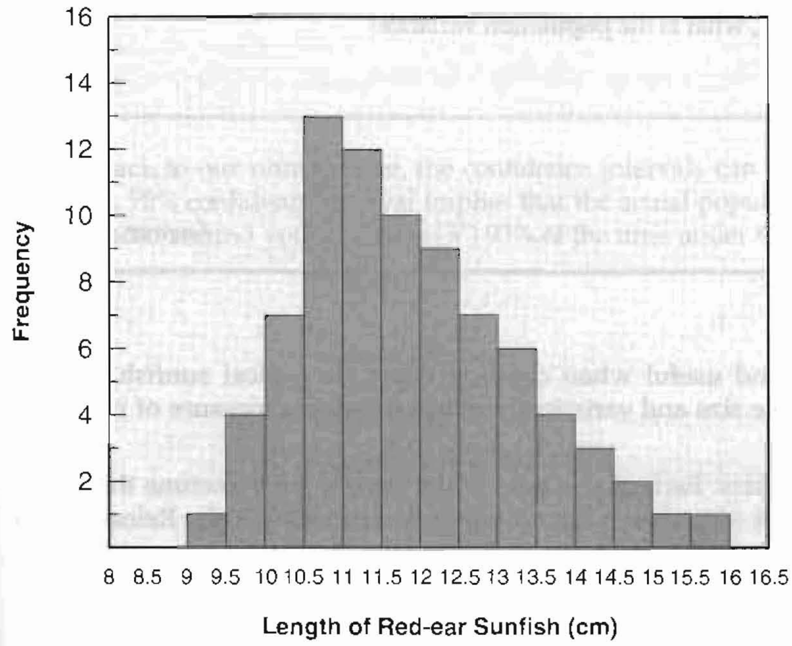
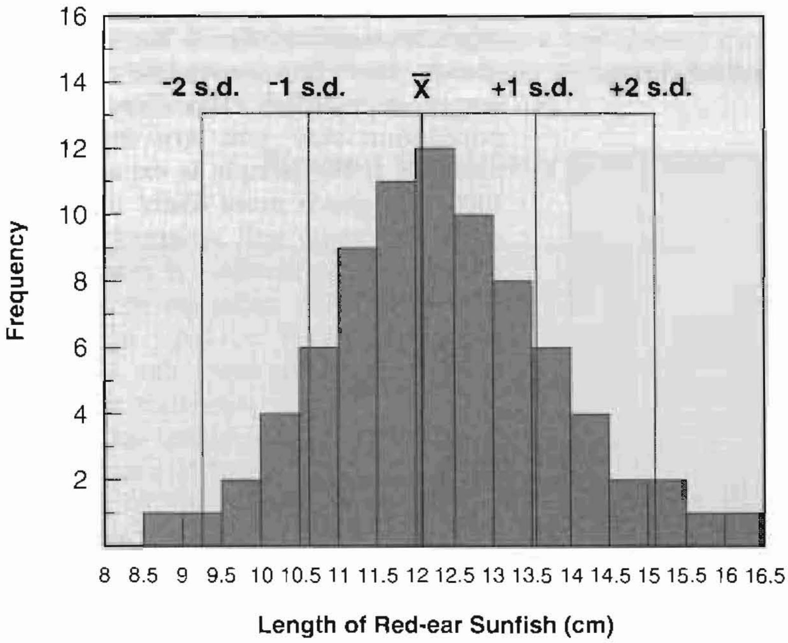


Figure 1.10 Standard deviations.



To calculate the size of a standard deviation, it is actually easier first to calculate a related statistic called the **variance**. The variance is the square of the standard deviation, so we use s^2 = **the sample variance**, and σ^2 = **the population variance.** Calculation of the variance is based on the difference between each observation and the mean. If all these differences are squared, and we calculate an average of the squared values, we have the variance. (See Appendix 1.) It is good to remember that the units on variance are the original measurement units squared. If we measure length in cm, then the sample variance is reported in cm^2. In calculating standard deviations, we take the square root of the variance, which returns us to our original measurement units, which is length in cm.

$$s = \sqrt{s^2} \qquad \text{and} \qquad \sigma = \sqrt{\sigma^2}$$

s = sample s.d.

s^2 = sample variance

σ = population s.d.

σ^2 = population variance

Check your progress:

If the standard deviation of a population is 9.5, what is the population variance?

Answer: 90.25

Confidence Intervals

There is one more statistic that you will find useful when characterizing the typical sunfish in your population with a mean. Incorporating sample size and variation, you can develop a measure of reliability of the mean called the **standard error (S.E.)**.

Assume there are 25 people in your ecology class. Each of you goes to the same pond sometime this week, nets a sample of sunfish in the same way, and measures 100 randomly selected individuals. Releasing the fish unharmed, you return to the lab and calculate a mean and standard deviation from your data.

Everyone else does the same. Would your 25 sample means be identical? No, but the variation in means would be considerably smaller than the total variation among the fish in the pond. Repeating a sampling program 25 times is usually impractical. Fortunately, statistics gives us a way to measure reliability when we have only one mean developed from one sample. The variation among all possible sample means can be predicted from the sample size and the variation in the pond's sunfish with the following formula:

$$S.E. = s/\sqrt{n}$$

S.E. = standard error of the mean
s = standard deviation of sample
n = sample size

Looking at the formula, you can see the relationship between error in our estimate, the variability of sunfish, and the sample size. The smaller the S.E. is, the more trustworthy your calculated mean. Note that the sample standard deviation is in the numerator of the calculation, so the more variable the size of the fish, the less accurate the estimate you made from a random sample. Sample size, on the other hand, is in the denominator. This implies that a large sample makes your mean more reliable. The formula shows that the more variable the population, the larger our sample must be to hold S.E. to an acceptably small margin of error.

Since standard errors tend to be normally distributed, it is a safe assumption that 95% of the variation in all possible means will fall within 1.96 S.E. of the actual mean. This fact can be used to calculate a 95% confidence interval as follows:

$$\text{95\% Confidence interval} = \bar{x} \pm 1.96\ \text{S.E.}$$

$\bar{x}$ = sample mean
S.E. = standard error

Check your progress:

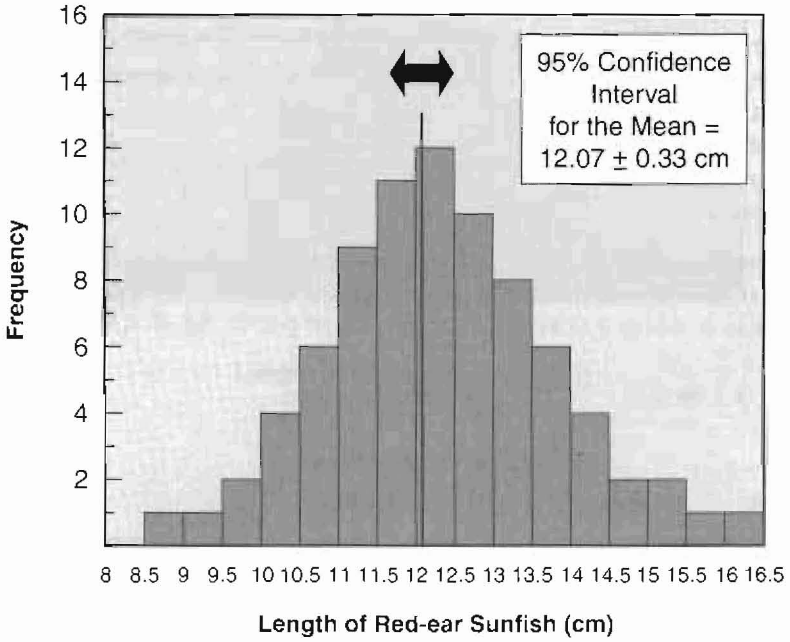
Calculate the 95% confidence interval for a mean of 14.3, derived from a sample of 25, where the standard deviation is 4.2. What are the upper and lower limits?

Answer: 12.65 to 15.95

To go back to our number line, the confidence intervals can be represented by brackets around the sample mean. A 95% confidence interval implies that the actual population mean (μ) will fall within the brackets you have placed around your estimate ($\bar{x}$) 95% of the time under these experimental conditions (Figure 1.11).

The importance of confidence limits cannot be overstated. Scientifically, it is dishonest to report a sample mean by itself. Without sharing with your readers the sampling methods, sample size, and the variability of the population, there is no way for them to know how accurately the sample mean represents the population. *Always report sample size with a mean, and add some measure of variation.* Any of the statistics representing variation (s, s^2, or S.E.) can be reported, since any one of these can be used to calculate the other two.

Figure 1.11 95% Confidence interval.



METHOD A: SEED WEIGHTS IN LEGUMES

[Laboratory activity]

Research Question

How does a population of bean seeds vary in weight around the typical individual?

Preparation

At a grocery or health food store, find uncooked dried beans of several types. Bags of approximately one pound are ideal. Lima beans, pinto beans, navy beans, and great northern beans are good varieties to choose from. If possible, buy unsorted beans; these best represent the population in the field.

Materials (per laboratory team)

1-pound bag of beans (Different teams can use different varieties.)

Analytical balance or electronic balance sensitive to 0.01 g (.001 g is preferable).

Plastic weighing tray, as large as the balance will accommodate.

Electronic calculator

Procedure

1. *Recognize that each seed in your bag is a living organism,* harvested from a mature population of annual plants. If sprouted and allowed to mature, these seeds would produce a population of bean plants. Seed weight is of obvious interest in cultivated plants like beans, but is also biologically important in wild plants, because seed weight affects the distribution, growth rate, and survivorship of the seedling. (For example, see Rees, 1995.) Since the maternal plant must expend more resources to produce larger seeds, plants making larger seeds cannot produce as many. In most plant species, the tradeoff between larger seeds vs. more numerous seeds is influenced by genetics, subject to **natural selection**, and varying from one plant type to another. Identifying mean and variance for seed weights is therefore a biologically important description of **reproductive strategy** in a plant population.
2. *Develop a sampling plan.* Your sample size will be 80 beans. If you were to cut the bag open and take only the beans on top, would your "grab sample" represent the population fairly? If larger and smaller beans settled differently during shipment, this approach might result in a biased sample. Spreading beans out on a lab bench, mixing them to randomize your sample, and selecting a sample of 80 closest to your side of the bench is a much better way to ensure that your sample is random.
3. *Weigh and record observations.* Weigh each of the 80 beans in your sample. How accurate is your balance? To simplify your data analysis, record weights in milligrams. (For example, rather than 0.087 g, record 87 mg. If your balance measures only to two decimal places, record 0.09 g as 90 mg. Enter your measurements on the calculation page at the end of this chapter.
4. *Produce a histogram.* Following the example in the introduction, produce a histogram of seed weights.
5. *Calculate descriptive statistics.* Calculate a mean, standard deviation, standard error, and confidence limits for this population, following the directions in the introduction and appendices.
6. *Interpret your data.* What does the histogram show about the variation of seed weights in this species? Answer the Questions for Discussion that follow the calculation pages. Considering the way natural selection works, would you always expect symmetrical frequency distributions of seed weights?
7. *Check your accuracy.* Working with your lab partners, count every bean in the bag. Then weigh the whole population by weighing the beans in a plastic weighing tray and subtracting the weight of the tray. If your balance cannot accommodate all the beans at once, divide the population into parts, weigh each part, and then sum the separate measurements. Divide the weight of the whole population by the number of beans in the population to calculate the population mean. How does the population mean compare with the sample mean you calculated in step 3? Does the population mean fall within the 95% confidence intervals you calculated?

METHOD B: NEEDLE LENGTH IN CONIFERS

[Outdoor/indoor activity]

Research Question

How do pine needles vary in length, within and among individual trees?

Preparation

Locate several pine trees of the same species. Needles of other conifers may be used if they have needles long enough to be measured with a mm ruler. If students cannot collect needles themselves, pine branches can be collected elsewhere, or "pine straw" can be obtained from garden supply firms. Laboratory teams can be larger if greater effort is required to collect and measure 80 needles.

Materials (per laboratory team)

Metric ruler, marked in mm

Electronic calculator

Procedure

1. *Recognize that each needle is a plant organ, developing according to a genetic program influenced by local conditions.* A pine needle performs the critically important job of photosynthesis, producing chemical energy for the tree. A needle's length may affect how well it functions. If the needle is too short, it may lack sufficient photosynthetic tissue to produce an adequate supply of food. On the other hand, needles that are too long may fail to transport fluids adequately to the tip, or may accumulate too much ice and break limbs in the winter. Limits on size and shape affecting the performance of a biological trait are called **functional constraints**, and they help explain why many species' characteristics remain within predictable ranges.

   Although needles on the same tree might be expected to conform to a genetically determined size, differences in leaf age, sun or shade, exposure to wind, temperature, or moisture supplied through a particular branch could influence needle length. Within a tree, we can recognize many sources of variation. Among a population of trees, the variation is probably even larger because different trees have different genes, and probably experience a broader range of environmental conditions.
2. *Develop a sampling plan.* Your sample size will be 80 needles. If you have access to a grove or row of pine trees, spread out your needle collection to include roughly equal numbers of needles from each of the trees. A sampling plan that includes the same number of needles from each tree in your research area is a **stratified sample.** Decide whether you will pull a needle from a live branch, or pick a fallen needle from the ground. Since pine needles decay slowly, they can be collected long after they fall from the tree. If you collect live needles from the tree, will you always collect from a low branch, or will you try to collect equal numbers of needles from high, mid-height, and low branches?

   When collecting needles from pines (genus *Pinus*) you will discover that their needles come in bunches. The brown collar of tissue holding the bunch of needles together is actually a dwarf branch, called a fascicle. The number of needles in a bunch is fairly consistent, and is useful for identification. For instance, the Eastern White Pine (*Pinus strobus*) typically has five needles per fascicle, while Red Pine (*Pinus resinosa*) has two.

Make a decision about which needle in the bunch you will measure. The longest one? A randomly selected one? What will you do if you encounter a broken needle? Whatever your method, it would be best to measure only one needle per bunch, so that your data are not clustered into subgroups. Pull the needles apart carefully, so as not to introduce error by breaking the base, and measure one needle according to your predetermined sampling plan.

If the population of pines on campus is not that large, collect all your needles from the same tree. Recognize that this collection is not a population in the biological sense, since only one individual produced all the needles. However, your collection is a population in the statistical sense, because you are measuring a sample of a much larger number of needles. Different lab groups can sample different trees so that you can compare your results.

3. *Measure lengths and record observations.* Measure each of the 80 needles in your sample. How accurate is your ruler? Record needle lengths in mm on the data pages at the end of this chapter. Follow directions on the calculation page to produce a histogram of needle lengths and to calculate a mean, standard deviation, standard error, and confidence limits for this population.
4. *Interpret your data.* What does the histogram show about the variation of needle lengths in this species? Answer the Questions for Discussion that follow the calculation pages. Think about the functional constraints on evergreen needles, and try to explain the distribution of sizes in biological terms.
5. *Check your accuracy.* Compare your results with those of another lab group. Does the other group's calculated mean fall within the 95% confidence limits you calculated for your own mean? If not, how do you interpret the difference between the two estimates? If two groups sampled from the same tree, then significant differences in your calculated means might result from sampling bias, measurement error, or calculation mistakes. If two lab groups sampled different trees, then the data may reflect real biological differences between the two trees.

CALCULATIONS (METHOD A OR B)

1. *Enter your 80 measurements* (x_i) in the second column of the table, recording 20 measurements per page on each of the next four pages. You may wish to split up this task, with each member of your team completing a page.
2. *Sum the measured values* (seed weight or needle length) for each page, and then complete calculation of the mean ($\bar{x}$) in the calculation box at the end of the tables by adding totals from all four pages and dividing by the sample size (n = 80).
3. Subtract the mean from each of the 80 measurements to *obtain the deviation* above or below the average. (Deviation from mean for sample i = $\mathbf{d}_i$.)
4. *Square each deviation* ($\mathbf{d}_i^2$).
5. *Add up all the squared deviations* on each page, then sum the totals for the four pages to compute the sum of squared deviations ($\Sigma \mathbf{d}^2$).
6. Divide the sum of squared deviations by (sample size – 1) to *calculate the sample variance* ($\mathbf{s}^2$).
7. Take the square root of the variance to *calculate the standard deviation* ($\mathbf{s}$).

Data (Methods A or B)—Page 1

Sample number (i)	Measurement for sample i (x_i)		Deviation (d_i)		Squared deviations (d_i^2)
1		$-(\bar{x}) =$		^ 2 =	
2		$-(\bar{x}) =$		^ 2 =	
3		$-(\bar{x}) =$		^ 2 =	
4		$-(\bar{x}) =$		^ 2 =	
5		$-(\bar{x}) =$		^ 2 =	
6		$-(\bar{x}) =$		^ 2 =	
7		$-(\bar{x}) =$		^ 2 =	
8		$-(\bar{x}) =$		^ 2 =	
9		$-(\bar{x}) =$		^ 2 =	
10		$-(\bar{x}) =$		^ 2 =	
11		$-(\bar{x}) =$		^ 2 =	
12		$-(\bar{x}) =$		^ 2 =	
13		$-(\bar{x}) =$		^ 2 =	
14		$-(\bar{x}) =$		^ 2 =	
15		$-(\bar{x}) =$		^ 2 =	
16		$-(\bar{x}) =$		^ 2 =	
17		$-(\bar{x}) =$		^ 2 =	
18		$-(\bar{x}) =$		^ 2 =	
19		$-(\bar{x}) =$		^ 2 =	
20		$-(\bar{x}) =$		^ 2 =	
Page 1 Sum $\Sigma(x_i) =$		**Page 1 Sum of Squared Deviations** $\Sigma(d_i^2) =$			

Data (Methods A or B)—Page 2

Sample number (i)	Measurement for sample i (x_i)		Deviation (d_i)		Squared deviations (d_i^2)
21		$- (\bar{x}) =$		^ 2 =	
22		$- (\bar{x}) =$		^ 2 =	
23		$- (\bar{x}) =$		^ 2 =	
24		$- (\bar{x}) =$		^ 2 =	
25		$- (\bar{x}) =$		^ 2 =	
26		$- (\bar{x}) =$		^ 2 =	
27		$- (\bar{x}) =$		^ 2 =	
28		$- (\bar{x}) =$		^ 2 =	
29		$- (\bar{x}) =$		^ 2 =	
30		$- (\bar{x}) =$		^ 2 =	
31		$- (\bar{x}) =$		^ 2 =	
32		$- (\bar{x}) =$		^ 2 =	
33		$- (\bar{x}) =$		^ 2 =	
34		$- (\bar{x}) =$		^ 2 =	
35		$- (\bar{x}) =$		^ 2 =	
36		$- (\bar{x}) =$		^ 2 =	
37		$- (\bar{x}) =$		^ 2 =	
38		$- (\bar{x}) =$		^ 2 =	
39		$- (\bar{x}) =$		^ 2 =	
40		$- (\bar{x}) =$		^ 2 =	
Page 2 Sum $\Sigma(x_i) =$		**Page 2 Sum of Squared Deviations** $\Sigma(d_i^2) =$			

Data (Methods A or B)—Page 3

Sample number (i)	Measurement for sample i (x_i)		Deviation (d_i)		Squared deviations (d_i^2)
41		− ($\bar{x}$) =		^ 2 =	
42		− ($\bar{x}$) =		^ 2 =	
43		− ($\bar{x}$) =		^ 2 =	
44		− ($\bar{x}$) =		^ 2 =	
45		− ($\bar{x}$) =		^ 2 =	
46		− ($\bar{x}$) =		^ 2 =	
47		− ($\bar{x}$) =		^ 2 =	
48		− ($\bar{x}$) =		^ 2 =	
49		− ($\bar{x}$) =		^ 2 =	
50		− ($\bar{x}$) =		^ 2 =	
51		− ($\bar{x}$) =		^ 2 =	
52		− ($\bar{x}$) =		^ 2 =	
53		− ($\bar{x}$) =		^ 2 =	
54		− ($\bar{x}$) =		^ 2 =	
55		− ($\bar{x}$) =		^ 2 =	
56		− ($\bar{x}$) =		^ 2 =	
57		− ($\bar{x}$) =		^ 2 =	
58		− ($\bar{x}$) =		^ 2 =	
59		− ($\bar{x}$) =		^ 2 =	
60		− ($\bar{x}$) =		^ 2 =	
Page 3 Sum $\Sigma(x_i)$ =		**Page 3 Sum of Squared Deviations** $\Sigma(d_i^2)$ =			

Data (Methods A or B)—Page 4

Sample number (i)	Measurement for sample i (x_i)		Deviation (d_i)		Squared deviations (d_i^2)
61		$-(\bar{x}) =$		^ 2 =	
62		$-(\bar{x}) =$		^ 2 =	
63		$-(\bar{x}) =$		^ 2 =	
64		$-(\bar{x}) =$		^ 2 =	
65		$-(\bar{x}) =$		^ 2 =	
66		$-(\bar{x}) =$		^ 2 =	
67		$-(\bar{x}) =$		^ 2 =	
68		$-(\bar{x}) =$		^ 2 =	
69		$-(\bar{x}) =$		^ 2 =	
70		$-(\bar{x}) =$		^ 2 =	
71		$-(\bar{x}) =$		^ 2 =	
72		$-(\bar{x}) =$		^ 2 =	
73		$-(\bar{x}) =$		^ 2 =	
74		$-(\bar{x}) =$		^ 2 =	
75		$-(\bar{x}) =$		^ 2 =	
76		$-(\bar{x}) =$		^ 2 =	
77		$-(\bar{x}) =$		^ 2 =	
78		$-(\bar{x}) =$		^ 2 =	
79		$-(\bar{x}) =$		^ 2 =	
80		$-(\bar{x}) =$		^ 2 =	
Page 4 Sum $\Sigma(x_i) =$		**Page 4 Sum of Squared Deviations** $\Sigma(d_i^2) =$			

Calculation of Variance and Standard Deviation

Page 1 Sum $\Sigma(x_i)$ =		**Page 1 Sum of Squared Deviations ⇒** $\Sigma(d_i^2)$ =	
Page 2 Sum $\Sigma(x_i)$ =		**Page 2 Sum of Squared Deviations ⇒** $\Sigma(d_i^2)$ =	
Page 3 Sum $\Sigma(x_i)$ =		**Page 3 Sum of Squared Deviations ⇒** $\Sigma(d_i^2)$ =	
Page 4 Sum $\Sigma(x_i)$ =		**Page 4 Sum of Squared Deviations ⇒** $\Sigma(d_i^2)$ =	
Grand Total $\Sigma(x_i)$ =		**Grand Total Sum of Squared Deviations ⇒** $\Sigma(d_i^2)$ =	
Mean $\Sigma(x_i) / n$ =		**Sample Variance ⇒** $s^2 = \Sigma(d_i^2)/(n-1)$ =	
	Standard Deviation ⇒ $\sqrt{s^2}$ =		

(Methods A or B)

Summary of Results (Method A or B)

Sample Size (n) =

Sample Mean ($\bar{x}$) =

Sample Variance (s^2) =

Standard Deviation (s) =

Standard Error (S.E.) =

95% Confidence Interval for Mean =

Histogram of Group Data

FREQUENCY

MEASUREMENT CLASSES

Questions (Method A or B)

1. In comparing your histogram to the description of a normal distribution, did you seem to get a fairly good fit to the bell-shaped curve, or did you notice a different pattern? Describe these differences: was the histogram bimodal? skewed? flattened or sharply peaked? If you performed statistical tests using these data, would you be comfortable using parametric statistics, or would you seek a non-parametric alternative? Do you think you made enough measurements to make a certain judgment on this question, or do you think more data may be needed?

2. The mean, median, and mode are three different statistical approaches to describe the "typical" individual in a population. Recall that the median is the middle observation when all data are arranged in numerical order, and the mode is the most commonly observed measurement. Based on your data, does it matter very much which of these three statistics is used? Explain how this answer is related to your answer for the previous question.

3. What do the variance, S.E., or standard deviation estimates tell you about your population that the mean does not tell you? Why is it important to report some measure of variation, along with the sample size, whenever you report a calculated mean?

4. Variation among members of a population can lead to natural selection, but only if two conditions are met: First, the trait must be relevant to an individual's survival and/or reproductive rate. Second, variation in this trait must be **heritable**, that is, at least partly controlled by genes. How would you design experiments to determine the importance of this trait in determining survival and reproduction? How would you test the extent to which this trait is heritable?

FOR FURTHER INVESTIGATION

1. In a field guide to the trees in your region, identify the species of evergreen you sampled in this exercise. What does the book say about the needle length of this species? Does the description include a range of needle lengths, or just a mean length? If keys are included in the guide, is needle length used to distinguish this species from others? Based on your data, how often might a single measurement of a randomly selected needle lead to an incorrect species identification?
2. Are some varieties of beans inherently more variable than others? Calculate variances for different types of beans, or from the same variety purchased from different suppliers. Is genetic variety a good predictor of seed weight, or are other considerations such as growing conditions in a given crop year more significant?

FOR FURTHER READING

Howe, H.F. and J. Smallwood. 1982. Ecology of seed dispersal. *Annual Reviews of Ecology and Systematics* 13:201–228.

Preston, Richard J. Jr. 1989. *North American Trees.* Iowa State University Press, Ames.

Rees, Mark. 1995. Community structure in sand dune annuals: Is seed weight a key quantity? *The Journal of Ecology* 83(5):857–863.