2015

Data and Error Analysis in Science

A Beginner's Guide

Paul Strode, PhD
Fairview High School
Boulder, Colorado

Ann Brokaw
Rocky River High School
Rocky River, Ohio

August 2015

Mathematics and Statistics in Science

Part 1: Statistical Symbols and Basic Equations

Part 2: Descriptive Statistics Used in Science
> Measures of Average: Mean, Median, and Mode
>> Mean
>> Median
>> Mode
>> When to Use Which One
> Measures of Variability: Range, Standard Deviation, and Variance
>> Range
>> Standard Deviation and Variance
>> Understanding Degrees of Freedom
> Measures of Confidence: Standard Error of the Mean and 95% Confidence Interval

About This Guide

Many state science standards encourage the use of mathematics and statistics in science education, including the newly designed AP and IB sciences courses, the Next Generation Science Standards, and the Common Core.

Two high school biology teachers, Paul Strode and Ann Brokaw, conceived this guide to give first-year secondary science students a general foundation for performing the data, error, and uncertainty analysis they will need in their science education in high school and beyond. This brief guide is not meant to be a textbook of statistics. However, this guide will function as a starting point for the inferential statistical tests that students may learn as they move into more advanced science courses like biology, chemistry, physics, and earth and space science.

The guide is organized in two parts:

- Part 1 is a "cheat sheet" of most of the statistical symbols and equations students will need to perform **descriptive statistics** on the data they generate in their lab activities and the inquiry based investigations they perform.

- Part 2 includes the methods to describe the typical or average value of the data, to determine the spread of the data, and to quantify the error and uncertainty that are inherent in a sample that is taken from a population of possible measurements. Part 2 also gives examples of how students can illustrate error and uncertainty in their data in graphical form.

Part 1: Statistical Symbols and Basic Equations

Listed below are the universal statistical symbols and equations used in this guide. The calculations can all be done using scientific calculators or the formula function in spreadsheet programs.

N: Total number of individuals in a population (i.e., the total number of butterflies of a particular species)

n: Total number of individuals in a sample of a population (i.e., the number of butterflies in a net)

df: The number of measurements in a sample that are free to vary once the sample mean has been calculated; in a single sample, df = $n - 1$

x_i: A single measurement

i: The ith observation in a sample

Σ: Summation

$\bar{x}$: Sample mean $\qquad \bar{x} = \frac{\sum x_i}{n}$

s^2: Sample variance $\qquad s^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1}$

s: Sample standard deviation $\qquad s = \sqrt{s^2}$

SE: Sample standard error, or standard error of the mean (SEM) $\qquad$ SE = $\frac{s}{\sqrt{n}}$

95% CI: 95% confidence interval $\qquad$ 95% CI = $\frac{1.96s}{\sqrt{n}}$

Part 2: Descriptive Statistics Used in Science

It is rarely practical for scientists to record every event or individual in a population of measurements. Instead, we typically collect data on a sample of a population and use these data to draw conclusions (or make inferences) about the entire population. Statistics is a mathematical discipline that relates to this type of analysis.

One of your first steps in analyzing a small data set is to graph the data and examine the distribution. **Figure 1** shows two graphs of beak measurements taken from two samples of medium ground finches that lived on the island of Daphne Major, one of the Galápagos Islands, during a major drought in 1977. The graph on the top shows beak measurements of finches that died during the drought. The graph on the bottom shows beak measurements of finches that survived the drought.

Beak Depths of 50 Medium Ground Finches That Did Not Survive the Drought
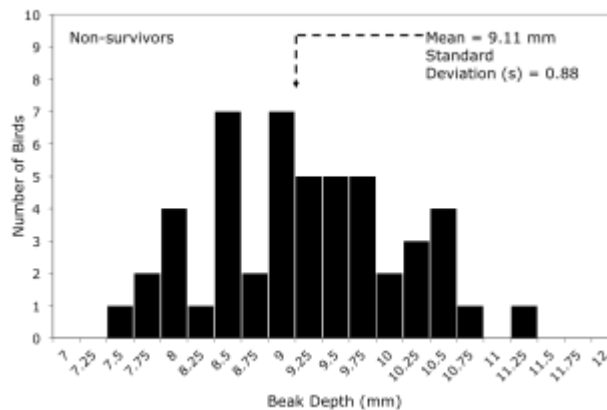
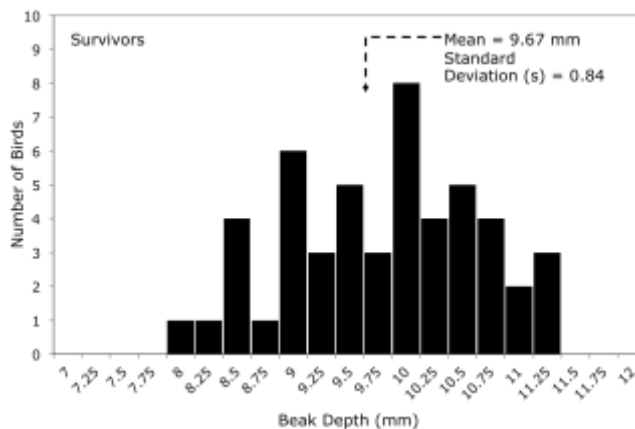**Beak Depths of 50 Medium Ground Finches That Survived the Drought**

Figure 1: The distributions of beak depth measurements in two groups of medium ground finches

Notice that the measurements tend to be more or less symmetrically distributed across a range, with most measurements around the center of the distribution. This is a characteristic of a **normal distribution**. The statistical methods covered in this guide, except for the measures of average and the range, apply to data that

are normally distributed, like the beak measurements above; other types of distributions require either different kinds of statistics or transforming data to make them normally distributed.

Also, in the two graphs, the center and spread of each distribution is different. **Descriptive statistics** allows you to describe and quantify these differences.

Measures of Average: Mean, Median, and Mode

A description of a group of observations can include a value for the mean, median, or mode. These are all measures of **central tendency**—in other words, they represent a number close to the center of the distribution.

Mean

You calculate the mean (also referred to as the **average** or **arithmetic mean**) by summing all the data points in a data set (ΣX) and then dividing this number by the total number of data points (*N*):

$$\bar{X} = \frac{\Sigma X}{N}$$

What we want to understand is the mean of the entire population, which is represented by μ. We use the **sample mean**, represented by $\bar{x}$, as an estimate of μ.

Application in physical science

Students in a physical science class measured the velocities of eight toy cars of the same type as they travelled down a ramp that was angled at 10°. The students recorded their results in **Table 1**.

Table 1. Velocity of eight cars travelling down a 10° ramp (± 0.05 cm/s)

Car No.	1	2	3	4	5	6	7	8
Velocity (cm/s)	7.5	10.1	8.3	9.8	5.7	10.3	9.2	8.8

To determine the mean of car velocities, follow these steps:

    I.   Find the sum of the velocities:
        7.5 + 10.1 + 8.3 + 9.8 + 5.7 + 10.3 + 9.2 + 8.8 = **69.7 cm/s**

    II.   Count the number of velocity measurements:
        There are eight velocity measurements.

    III.   Divide the sum of the velocities by the number of measurements to compute the mean:
        mean = 69.7 cm/8 = 8.7 cm/s

The mean velocity for this sample of eight cars is 8.7 cm/s and serves as an estimate for the true mean of a population of toy cars of this type under these conditions (the 10° ramp). Thus, if students collected data from hundreds of toy cars of the same type and graphed the data, the center of the distribution might be around 8.7 centimeters per second. *Note: The above calculation on a calculator will equal 8.7125 cm/s. Since your least precise measurement is to the nearest 0.1 cm/s, the mean you report must be to the nearest 0.1 cm/s. However, if you plan to do additional calculations with the mean, you can carry all the digits through.*

Median

When the data are ordered from the largest to the smallest, the median is the midpoint of the data. It is not distorted by extreme values, or even when the distribution is not normal. For this reason, it may be more useful for you to use the median as the main descriptive statistic for a sample of data in which some of the measurements are extremely large or extremely small.

To determine the median of a set of values, you first arrange them in numerical order from lowest to highest. The middle value in the list is the median. If there is an even number of values in the list, then the median is the mean of the middle two values.

Application in Chemistry

Students in a chemistry class were studying clock reactions. The 26 students in the class organized themselves into 13 groups of two. Each group mixed a colorless solution of hydrogen peroxide and sulfuric acid with a colorless solution containing potassium iodide, sodium thiosulfate, and starch. The groups recorded the time it took for their new solution to turn black. Their results are in **Table 2**.

Table 2. Time for the clock reaction to occur

Group No.	1	2	3	4	5	6	7	8	9	10	11	12	13
Time (sec.)	31	33	81	33	28	29	33	27	27	34	35	28	32

To determine the median time for the clock reaction to occur, follow these steps:

I. Arrange the time values in numerical order from lowest to highest:

27, 27, 28, 28, 29, 31, **32**, 33, 33, 33, 34, 35, 81

II. Find the middle value. This value is the median:

median = 32 seconds

In this case, the median is 32 seconds, but the mean is 35 seconds. The mean is larger than the median because a very slow reaction distorts the mean in the direction of that large value. In this case, we call that very large value an *outlier* because it lies so far away from the rest of the measurements. Why might one group of students have recorded such a large value? *Note: If the researchers can determine that the outlier was produced naturally and not because of an error in the methods of the investigation, there may be something truly unique and interesting going on with that measurement. Sometimes, by investigating the outliers, scientists make some of their most interesting discoveries.*

Mode

The mode is another measure of the average. It is the value that appears most often in a sample of data. In the example shown in **Table 2**, the mode is 33 seconds.

The mode is not typically used as a measure of central tendency in science, but it can be useful in describing some distributions. For example, **Figure 2** shows a distribution of a sample of body lengths of weaver ant workers with two peaks, or modes—called a **bimodal distribution**. Describing these data with a measure of central tendency like the mean or median would obscure this fact.
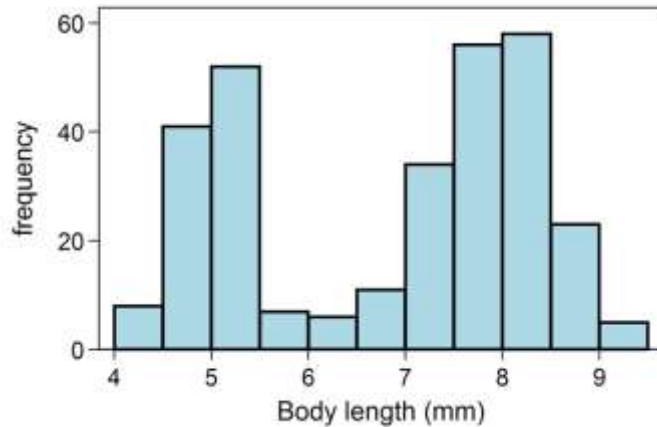
Figure 2: Graph of body lengths of weaver ant workers (Reproduced from http://en.wikipedia.org/wiki/File:BimodalAnts.png.)

When to Use Which One

The purpose of these statistics is to characterize "typical" data from a data set. You use the mean most often for this purpose, but it becomes less useful if the data in the data set are not normally distributed. When the data are *not* normally distributed, then other descriptive statistics can give a better idea about the typical value of the data set. The median, for example, is a useful number if the distribution is heavily skewed, like the mass of the planets in our solar system. Indeed, Jupiter has a mass of 1.9×10^{27} kg and Saturn has a mass of 5.7×10^{26} kg. Taken together, these two planets make up 92.5% of the mass of all eight planets. Since there is an even number of planets in our solar system, the median is determined by calculating the mean of Earth and Uranus (the 4th and 5th most massive planets, respectively), which is 4.6×10^{25} kg. Again, the mode is not used very frequently in science, but it may be useful in describing some types of distributions—for example, ones with more than one peak.

Measures of Variability: Range, Standard Deviation, and Variance

Variability describes the extent to which numbers in a data set diverge from the central tendency. It is a measure of how "spread out" the data are. The most common measures of variability are **range**, **standard deviation**, and **variance**.

Range

The simplest measure of variability in a sample of normally distributed data is the range, which is the difference between the largest and smallest values in a set of data. You can use the range for data that are not normally distributed.

Application in Earth Science

Students in an earth science class obtained data on the thickness of the Earth's ozone layer from monitoring sites around the world and recorded their results in **Table 3**. The thickness of the ozone layer is measured in Dobson units (DU). A measurement of less than 220 DU is considered an ozone hole.

Table 3. Thickness of the ozone layer at eight global monitoring stations

Station No.	1	2	3	4	5	6	7	8
Thickness (DU)	283	245	204	256	260	300	290	305

To determine the range of ozone thickness, follow these steps:

    I.   Identify the largest and smallest values in the data set:
        largest = 305 DU, smallest = 204 DU

    II.   To determine the range, subtract the smallest value from the largest value:
        range = 305 DU – 204 DU = 101 DU

Which station has an ozone hole?

For any data, a larger range value indicates a greater spread of the data—in other words, the larger the range, the greater the variability. However, an extremely large or small value in the data set will make the variability appear high. The standard deviation provides a more reliable measure of the "true" spread of the data.

Standard Deviation and Variance

The standard deviation is the most widely used measure of variability. The **sample standard deviation** (*s*) is essentially the average of the difference between each measurement in the sample and the sample mean ($\overline{x}$). The sample standard deviation is an estimate of the standard deviation in the larger population.

The formula for calculating the sample standard deviation follows:

$$s = \sqrt{\frac{\sum(x_i - \overline{x})^2}{(n-1)}}$$

Calculation Steps

1. Calculate the mean ($\overline{x}$) of the sample.

2. Find the difference between each measurement (x_i) in the data set and the mean ($\overline{x}$) of the entire set: $(x_i - \overline{x})$

3. Square each difference to remove any negative values: $(x_i - \overline{x})^2$

4. Add up (sum, Σ) all the squared differences: $\Sigma(x_i - \overline{x})^2$

5. Divide by the **degrees of freedom**, which is 1 less than the sample size ($n - 1$) (*degrees of freedom are explained in its own section on Page 12*):

$$\frac{\Sigma(x_i - \overline{x})^2}{(n - 1)}$$

Note that the number calculated at this step provides a statistic called **variance** (s^2). Variance is an important measure of variability that is used in certain statistical methods. It is the square of the standard variation.

6. Take the square root to calculate the standard deviation (s) for the sample.

Application in Biology

You are interested in knowing how tall bean plants (*Phaseolus vulgaris*) grow in two weeks after planting. You plant a sample of 20 seeds ($n = 20$) in separate pots and give them equal amounts of water and light. After two weeks, 17 of the seeds have germinated and have grown into small seedlings (now $n = 17$). You measure each plant from the tips of the roots to the top of the tallest stem. You record the measurements in **Table 4** on the next page, along with the steps for calculating the standard deviation.

Table 4. Plant measurements and steps for calculating the standard deviation

Plant No.	Plant Height (mm ± 0.5)	Step 2: $(x_i - \overline{x})$	Step 3: $(x_i - \overline{x})^2$
1	112	(112 − 103) = 9	9^2 = 81
2	102	(102 − 103) = (−1)	$(−1)^2$ = 1
3	106	(106 − 103) = 3	3^2 = 9
4	120	(120 − 103) = 17	17^2 = 289
5	98	(98 − 103) = (−5)	$(−5)^2$ = 25
6	106	(106 − 103) = 3	3^2 = 9
7	80	(80 − 103) = (−23)	$(−23)^2$ = 529
8	105	(105 − 103) = 2	2^2 = 4
9	106	(106 − 103) = 3	3^2 = 9
10	110	(110 − 103) = 7	7^2 = 49
11	95	(95 − 103) = (−8)	$(−8)^2$ = 64
12	98	(98 − 103) = (−5)	$(−5)^2$ = 25
13	74	(74 − 103) = (−29)	$(−29)^2$ = 841
14	112	(112 − 103) = 9	9^2 = 81
15	115	(115 − 103) = 12	12^2 = 144
16	109	(109 − 103) = 6	6^2 = 36
17	100	(100 − 103) = (−3)	$(−3)^2$ = 9
Step 1: Calculate mean.	$\overline{x}$ = 103		*Step 4:* $\sum(x_i - \overline{x})^2$ = 2,205
		Variance (s^2)	*Step 5:* $\sum(x_i - \overline{x})^2/(n-1)$ = 2,205/16 = 138
		Standard deviation (s)	*Step 6:* $\sqrt{s^2}$ = $\sqrt{138}$ = 11.7

The mean height of the bean plants in this sample is 103 millimeters $\pm$ 12 (11.7) millimeters. What does this tell us?

In a data set with a large number of continuous measurements that are normally distributed, 68% of the measurements are expected to fall within 1 standard deviation from the mean and around 95% of all data points lie within 2 standard deviations of the mean on either side (**Figure 3**). Thus, in this example, if you assume that this sample of 17 observations is drawn from a population of measurements that are normally distributed, 68% of measurements in the population should fall between 91 and 115 millimeters and 95% of the measurements should fall between 79 millimeters and 127 millimeters.
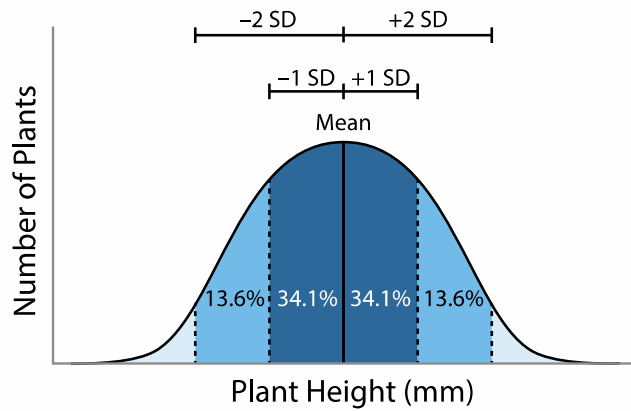
Figure 3: For normally distributed data, 68% of data points lie between ±1 standard deviation of the mean and 95% of data points lie between ±2 standard deviations of the mean.

You can graph the mean and the standard deviation of this sample of bean plants using a bar graph with standard deviation as bars on either side of the mean to show the variation of your measurements around your estimate of the mean (**Figure 4**). Standard deviation bars do not reflect the error in the measurements—this is a common misconception. On the other hand, true error bars based on the standard error of the mean and 95% confidence interval reflect the ***uncertainty*** in the sample mean when estimating the population mean, and they are dependent on sample size. (These statistics are discussed further in "Measures of Confidence: Standard Error of the Mean and 95% Confidence Intervals.")
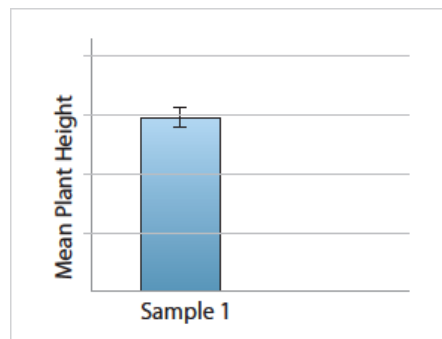

Figure 4: This bar graph shows the mean plant height of a sample of bean plants and an error bar that represents ±1 standard deviation. Sixty-eight percent of measurements in this population fall in the range indicated by the bar.

Note: Another common misconception is that standard deviation decreases with increasing sample size. Standard deviation can either increase or decrease as sample size increases; it just depends on the measurements in the sample. If there is a lot of variation in a population, the standard deviation will be large.

Understanding Degrees of Freedom

Calculations of sample estimates, such as the standard deviation and variance, use degrees of freedom instead of sample size. The way you calculate degrees of freedom depends on the statistical method you are using, but for calculating the standard deviation, it is defined as 1 less than the sample size ($n - 1$).

To illustrate what this number means, consider the following example. Biologists are interested in the variation in leg lengths among grasshoppers. They catch five grasshoppers ($n = 5$) in a net and prepare to measure the left legs. As the scientists pull grasshoppers one at a time from the net, they have no way of knowing the leg lengths until they measure them all. In other words, all five leg lengths are "free" to vary within some general range for this particular species. The scientists measure all five leg lengths and then calculate the mean to be 10 millimeters. They then place the grasshoppers back in the net and decide to pull them out one at a time to measure them again. This time, since the biologists already know the mean to be 10, only the first four measurements are free to vary within a given range. If the first four measurements are 8, 9, 10, and 12 millimeters, then there is no freedom for the fifth measurement to vary; it has to be 11. Thus, once they know the sample mean, the number of degrees of freedom is 1 less than the sample size, df $= 4$.

Measures of Confidence: Standard Error of the Mean and 95% Confidence Interval

The standard deviation provides a measure of the spread of the data from the mean. A different type of statistic reveals the *uncertainty* in the calculation of the mean. The **sample mean** is not necessarily identical to the mean of the entire population. In fact, every time you take a sample and calculate a sample mean, you would expect a slightly different value. In other words, the sample means themselves have variability. This variability can be expressed by calculating **the standard error of the mean** (abbreviated as $SE_{\bar{x}}$ or SEM) or **the 95% confidence interval** (95% CI). Indeed, a critical step in making claims in science is to statistically determine the level of uncertainty in a claim.

To calculate $SE_{\bar{x}}$, divide the standard deviation by the square root of the sample size:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{(n-1)}}$$

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

$$95\% \text{ CI} = \frac{1.96s}{\sqrt{n}} \quad \text{(typically rounded to } \frac{2s}{\sqrt{n}}.\text{)}$$

Assume that there is a population of a species of anole lizards living on an island of the Caribbean. If you were able to measure the length of the hind limbs of every individual in this population and then calculate the mean, you would know the value of the population mean. However, there are thousands of individuals, so you take a sample of 10 anoles and calculate the mean hind limb length for that sample. Another researcher working on that island might catch another sample of 10 anoles and calculate the mean hind limb length for this sample, and so on. The sample means of many different samples would be normally distributed. **The standard error of the mean represents the standard deviation of such a distribution** and estimates how close the sample mean is to the population mean. The greater the sample size (i.e., 50 rather than 10 anoles), the more closely the sample mean will estimate the population mean, and therefore the standard error of the mean becomes smaller.

The 95% confidence interval (95% CI) is equivalent to around 2 standard errors of the mean. Because the sample means are assumed to be normally distributed, 95% of all sample means should fall between 2 standard deviations above and below the population mean, estimated by the 95% CI.

What these calculations tell you is that you can be 68% confident that the population mean falls within ±1 standard error of the sample mean and 95% confident that the population mean falls within approximately ±2 standard errors of the sample mean.

Both $\text{SE}_{\bar{x}}$ and 95% CI can be illustrated as error bars in a bar graph of the means of two or more samples that are being compared. Depicting $\text{SE}_{\bar{x}}$ or the 95% CI as error bars in a bar graph provides a clear visual clue to the uncertainty of the calculations of the sample means. *Note: Unlike standard deviation, both standard error and 95% CI decrease with increasing sample size. In other words, the greater the sample size the less the uncertainty in the calculations of the sample means.*

Application in Agriculture—Example 1

Seeds of many weed species germinate best in recently disturbed soil that lacks a light-blocking canopy of vegetation. Students in an environmental science class **hypothesized** that weed seeds germinate in response to light. To test this hypothesis, the students placed 25 seeds of crofton weed (*Ageratina adenophora*, an invasive species on several continents) in each of 20 petri dishes and covered them with distilled water. They placed half the petri dishes in the dark and half in the light. After 72 hours, the students counted the number of seeds that had germinated in each dish. The data and calculations of variance, standard deviation, standard error of the mean, and 95% confidence interval are shown in **Table 5**. The students plotted the data as two bar graphs showing the standard error of the mean and 95% confidence interval (**Figure 5**).

Table 5. Number of crofton seeds that germinated after 72 hours in the dark and in the light (number of replicates [i.e., sample size, n] = 10)

| Petri Dishes	Dark (x_1)	Light (x_2)	Dark $(x_i - \bar{x}_1)^2$	Light $(x_i - \bar{x}_2)^2$
1 and 2	12	18	$(12 - 9.6)^2$ = 5.8	$(18 - 18.4)^2$ = 0.16
3 and 4	8	22	$(8 - 9.6)^2$ = 2.6	$(22 - 18.4)^2$ = 12.96
5 and 6	15	17	$(15 - 9.6)^2$ = 29.1	$(17 - 18.4)^2$ = 1.96
7 and 8	13	23	$(13 - 9.6)^2$ = 11.5	$(23 - 18.4)^2$ = 21.16
9 and 10	6	16	$(6 - 9.6)^2$ = 13.0	$(16 - 18.4)^2$ = 5.76
11 and 12	4	18	$(4 - 9.6)^2$ = 31.4	$(18 - 18.4)^2$ = 0.16
13 and 14	13	22	$(13 - 9.6)^2$ = 11.6	$(22 - 18.4)^2$ = 12.96
15 and 16	14	12	$(14 - 9.6)^2$ = 19.3	$(12 - 18.4)^2$ = 40.96
17 and 18	5	19	$(5 - 9.6)^2$ = 21.1	$(19 - 18.4)^2$ = 0.36
19 and 20	6	17	$(6 - 9.6)^2$ = 13.0	$(17 - 18.4)^2$ = 1.96
			$\sum (x_i - \bar{x}_1)^2$ = 158.4	$\sum (x_i - \bar{x}_2)^2$ = 98.4
Mean ($\bar{x}$)	$\bar{x}_1$ = 9.6	$\bar{x}_2$ = 18.4	$\frac{\sum(x_i - \bar{x})^2}{(n-1)} = \frac{158.4}{9}$	$\frac{\sum(x_i - \bar{x})^2}{(n-1)} = \frac{98.4}{9}$
		Variance (s^2)	s_1^2 = 17.6	s_2^2 = 10.93
Standard deviation $s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{(n-1)}}$			s = 4.20	s = 3.31
Standard error of the mean $\text{SE}_{\bar{x}} = \frac{s}{\sqrt{n}}$			$\text{SE}_{\bar{x}} = \frac{4.20}{\sqrt{10}}$ = 1.33	$\text{SE}_{\bar{x}} = \frac{3.31}{\sqrt{10}}$ = 1.05
95% CI = $\frac{2s}{\sqrt{n}}$			95% CI = $\frac{2(4.20)}{\sqrt{10}}$ = 2.7	95% CI = $\frac{2(4.74)}{\sqrt{10}}$ = 2.1
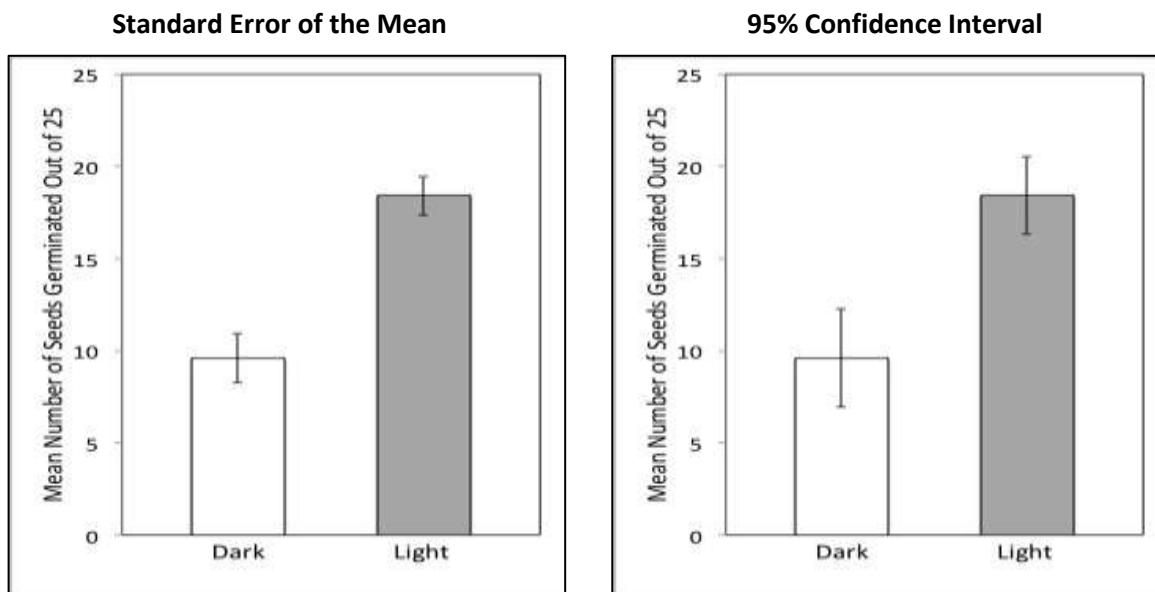
Figure 5: Mean number of crofton seeds that germinated after 72 hours in the dark or in the light. The standard error of the mean graph shows the $\mathrm{SE}_{\bar{x}}$ as error bars, and the 95% confidence interval graph shows the 95% CI as error bars.

The calculations show that although the students don't know how many seeds of the entire population of crofton seeds would germinate in the dark, it is likely to be a number around the sample mean of 9.6 seeds. But how confident are they of this estimate? They are 68% confident that the dark treatment population mean lies between 8.3 (9.6 – 1.3) and 10.9 (9.6 + 1.3) and that the light treatment population mean lies between 17.35 and 19.45 seeds. In addition, they are 95% confident that the dark treatment population mean lies between 7 and 12.2 seeds and the light treatment population mean lies between 16.3 and 21.5 seeds.

Note: By looking at the bar graphs, you can see that the means for the light and dark treatments are different. Can you tell if the differences are statistically significant? If the standard error bars and 95% CI bars overlap, then it is likely that the differences are not statistically significant. In this case, the standard error bars do not overlap; however, this does not necessarily mean that the difference between the means is statistically significant (truly different). On the other hand, the fact that the 95% confidence interval error bars do not overlap strongly suggests that the two sample means are *significantly* different. An additional statistical test, the Student's *t*-test, is required to be sure.

Application in Physiology—Example 2
A teacher had five students in a physical education class write their names on the board, first with their dominant hands and then with their nondominant hands. The rest of the class observed that the students wrote more slowly and with less precision with the nondominant hand than with the dominant hand. The teacher then asked the class to explain their observations by developing testable hypotheses. They hypothesized that the dominant hand was better at performing fine motor movements than the nondominant hand. The class tested this hypothesis by timing (in seconds) how long it took each student to break 20 toothpicks with each hand. The results of the experiment and the calculations of variance, standard deviation, standard error of the mean, and 95% confidence interval are presented in **Table 6**. The students then illustrated the data and uncertainty with two bar graphs, one showing the standard error of the mean and the other showing the 95% confidence interval (**Figure 6**).

Table 6. Number of seconds it took for students to break 20 toothpicks with their nondominant (ND) and dominant hands (D) (number of replicates [n] = 14)

| Students	ND (x_1) Sec. (± 0.5)	D (x_2) Sec. (± 0.5)	ND $(x_i - \bar{x}_1)^2$	D $(x_i - \bar{x}_2)^2$
Josh	33	37	$(33 - 33)^2 = 0$	$(37 - 35)^2 = 4$
Bobby	24	22	$(24 - 33)^2 = 81$	$(22 - 35)^2 = 169$
Qing	35	37	$(35 - 33)^2 = 4$	$(37 - 35)^2 = 4$
Julie	33	28	$(33 - 33)^2 = 0$	$(28 - 35)^2 = 49$
Lisa	42	50	$(42 - 33)^2 = 81$	$(50 - 35)^2 = 225$
Akash	36	36	$(36 - 33)^2 = 9$	$(36 - 35)^2 = 1$
Hector	31	36	$(31 - 33)^2 = 4$	$(36 - 35)^2 = 1$
Viviana	40	46	$(40 - 33)^2 = 49$	$(46 - 35)^2 = 121$
Brenda	28	26	$(28 - 33)^2 = 25$	$(26 - 35)^2 = 81$
Jane	24	28	$(24 - 33)^2 = 81$	$(28 - 35)^2 = 49$
Asa	23	22	$(23 - 33)^2 = 100$	$(22 - 35)^2 = 169$
Eli	44	52	$(44 - 33)^2 = 121$	$(52 - 35)^2 = 289$
Adee	35	29	$(35 - 33)^2 = 4$	$(29 - 35)^2 = 36$
Jenny	36	37	$(36 - 33)^2 = 9$	$(37 - 35)^2 = 4$
			$\sum (x_i - \bar{x}_1)^2 = 568$	$\sum (x_i - \bar{x}_2)^2 = 1,200$
Mean ($\bar{x}$)	$\bar{x}_1 = 33$	$\bar{x}_2 = 35$	$\frac{\sum (x_i - \bar{x}_1)2}{n-1} = \frac{568}{13}$	$\frac{\sum (x_i - \bar{x}_2)2}{n-1} = \frac{1,200}{13}$
		Variance (s^2)	$s_1^2 = 44$	$s_2^2 = 92$
Standard deviation $s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{(n-1)}}$			$s = 6.6$	$s = 9.6$
Standard error of the mean $\text{SE}_{\bar{x}} = \frac{s}{\sqrt{n}}$			$\text{SE}_{\bar{x}} = \frac{6.6}{\sqrt{14}} = 1.8$	$\text{SE}_{\bar{x}} = \frac{9.6}{\sqrt{14}} = 2.6$
95% CI $= \frac{2s}{\sqrt{n}}$			95% CI $= \frac{2(6.84)}{\sqrt{14}} = 3.5$	95% CI $= \frac{2(9.6)}{\sqrt{14}} = 5.1$
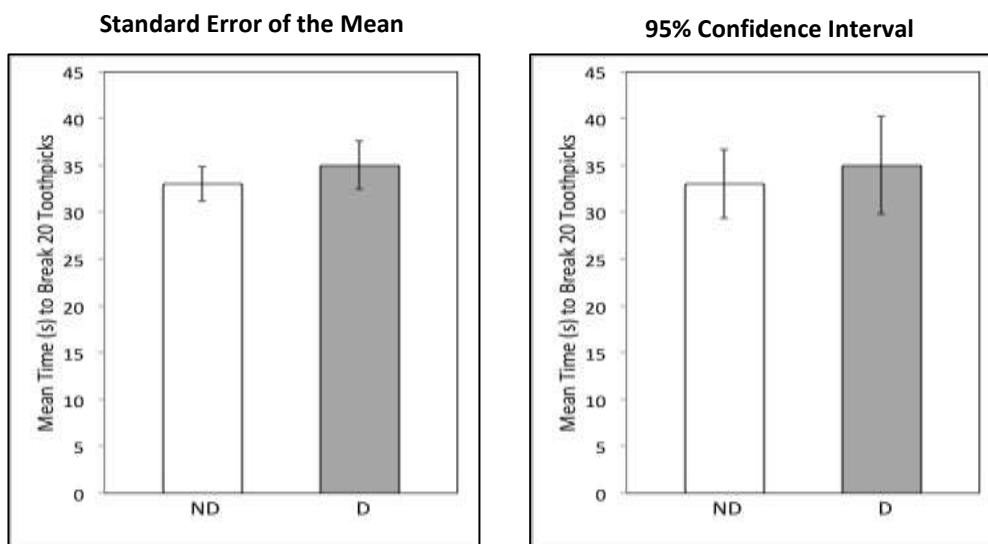
Figure 6: Mean number of seconds for students to break 20 toothpicks with their nondominant hands (ND) and dominant hands (D). The standard error of the mean graph shows the $\text{SE}_{\bar{x}}$ as error bars, and the 95% confidence interval graph shows the 95% CI as error bars.

The calculations show that students can be 68% confident that the population mean for the nondominant hand treatment lies within 31.2 seconds (33 – 1.8) and 34.8 seconds (33 + 1.8). Likewise, for the dominant hand, students can be 68% confident that the population mean falls within 32.4 and 37.6 seconds.

Students are 95% confident that the population mean for the nondominant hands falls somewhere between 29.5 seconds (33 – 3.5) and 36.5 seconds (33 + 3.5) and that the population mean for the dominant hands falls somewhere between 29.9 seconds (35 – 5.1) and 40.1 seconds (35 + 5.1).

Note: The fact that the 95% confidence interval error bars for the two sample means overlap, so much so that they include the means, suggests that the two means may not be significantly different. Indeed, if a student reported that she broke 20 toothpicks with one hand, we could not conclude with confidence whether she used her dominant or her nondominant hand because 30 seconds is encompassed by both 95% confidence interval bars.

Edited by Laura Bonetta, PhD, Howard Hughes Medical Institute
Reviewed by Brad Williamson, PhD, University of Kansas; Sandra Blumenrath, PhD, and Satoshi Amagai, PhD, Howard Hughes Medical Institute
Copyedited by Barbara Resch, Howard Hughes Medical Institute
Graphics by Heather McDonald, PhD, and Bill Pietsch, Howard Hughes Medical Institute

Student Notes