**MAKING MEANING OF DATA: THE VALUE OF STATISTICS**

It is important to organize and analyze a collection of numerical data correctly in order to make meaning of your information. The process of doing this is called STATISTICS. Since biologists almost never have ALL of the data for an entire population of organisms, statistical analysis can help us reach conclusions about the entire population, based on the sample for which data was collected.

Like many other forms of communication, statistics has a written “language” that uses symbols to represent ideas. Learning the meaning of these symbols takes practice, but the more you practice, the easier it gets.

The first two important concepts to analyze a set of numerical data are:

* **MEAN**: often also called the “average.” It is calculated by adding up all of your data points and dividing the total by the number of data points that you have.

Symbols used to calculate mean:

n: the total number of individuals in a population (ex.: the number of butterflies in a net)

xi: the ith observation in a sample (ex.: your first butterfly observation is x1; your 20th butterfly observation is x20)

∑: Summation. This symbol instructs you to add up a series of numbers.

$\overline{x}$: the sample mean.

Formula used to calculate mean:

 $\overline{x}$ = $\frac{Σx\_{i}}{n}$

Formula in words:

**Mean** equals the **sum** of all of the **individual observations**, **divided by** the **total number of individuals.**

* **STANDARD DEVIATION**: Standard deviation is the most widely used measure of variability in a sample. It essentially measures how different each individual data point is from the mean. The next page shows you the SIX EASY STEPS needed to calculate Standard Deviation.

**STEPS TO CALCULATE STANDARD DEVIATION**

  

| **Column 1:****Data Point #** | **Column 2:****Measurement (**$x\_{i})$ | **Column 3:**$x\_{i}-\overline{x}$ | **Column 4:**$(x\_{i}-\overline{x})^{2}$ |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
| ***Step 1: use Column 2 to calculate the mean →***  | $\overline{x} =$ | ***Step 4: Add up all the values in Column 4. →***$Σ(x\_{i}-\overline{x})^{2}$ |  |
|  |  | ***Step 5: divide Step 4 by your total number of data points, minus 1. This is called the VARIANCE and is symbolized by*** $s^{2}$ ***→***$s^{2}=$$\frac{Σ(x\_{i}-\overline{x})^{2}}{n-1}$ |  |
|  |  | ***Step 6: take the square root of Step 5. This is your Standard Deviation!***$s=\sqrt{s^{2}}$***→*** |  |

**Example**:

  

| **Column 1:****Plant #** | **Column 2:****Plant Height (mm)** **(**$x\_{i})$ | **Column 3:**$x\_{i}-\overline{x}$ | **Column 4:**$(x\_{i}-\overline{x})^{2}$ |
| --- | --- | --- | --- |
| 1 | 112 | 112-110 = 2 | 4 |
| 2 | 102 | 102-110 = -8 | 64 |
| 3 | 106 | 106-110 = -4 | 16 |
| 4 | 120 | 120-110 = 10 | 100 |
| ***Step 1: use Column 2 to calculate the mean →***  | $\overline{x} =$**110** | ***Step 4: Add up all the values in Column 4. →***$Σ(x\_{i}-\overline{x})^{2}$ | **214** |
|  |  | ***Step 5: divide Step 4 by your total number of data points, minus 1. This is called the VARIANCE and is symbolized by*** $s^{2}$ ***→***$s^{2}=$$\frac{Σ(x\_{i}-\overline{x})^{2}}{n-1}$ | **214/3 = 71.33** |
|  |  | ***Step 6: take the square root of Step 5. This is your Standard Deviation!***$s=\sqrt{s^{2}}$***→*** | **8.44** |

Check your work: A convenient way to check your work is by using an online statistical calculator such as [Calculator.net](https://www.calculator.net/standard-deviation-calculator.html).

**STANDARD ERROR OF THE MEAN**

As you have learned, standard deviation measures how different each individual data point is from the mean. It’s important to remember, though, that your data almost certainly does not include every possible measurement on an entire population. For example, the mean in the example above is based on the height of only four plants; this species of plant possibly includes anywhere from hundreds to billions of individuals! How different is the mean that you calculate from your data from the mean that you would get if you could measure all of the individuals in a population? To estimate an answer to this question, you calculate the **Standard Error of the Mean.** *(Note: the word “error”, in this context, does NOT suggest that there is a mistake in your data or your calculations!).*

Formula used to calculate Standard Error:

 $SE\_{\overline{x}}$ = $\frac{s}{\sqrt{n}}$

Formula in words:

**Standard Error of the Mean** equals **standard deviation**, **divided by** the **square root of the total number of individuals.**

What Standard Error of the Mean tells you:

Imagine that you collect samples of a population an infinite number of times, and calculate the sample mean for each population. Standard Error of the Mean says that about two-thirds (68.3%) of your sample means would be within one standard error of the mean for the entire population. About 95% of your sample means would be within two standard errors of the mean for the entire population; almost all of your sample means would be within three standard errors.

This usually makes more sense when you see sample means graphed together with their standard errors of the mean, which are shown as standard error bars. Here’s an example:

**Importance of Selfie Sticks, By Gender**



What this graph tells you is that when you, the scientist, asked a sample of females to rate the importance of selfie sticks on a scale from 1 to 5, the mean importance of selfie sticks to females is 2.72. However, if you repeated your test, about 68% of the time, your means would fall within the standard error bracket. Your mean would be as high as 3.77 (2.72+1.05) or as low as 3.67 92.72-1.05) about 68% of the time.

In summary: Standard Error of the Mean is one way to evaluate how accurately your sample mean reflects what’s going on with an entire population.