Name: ________________________________________ Period: _________

# Properties of Water Lab with Stats

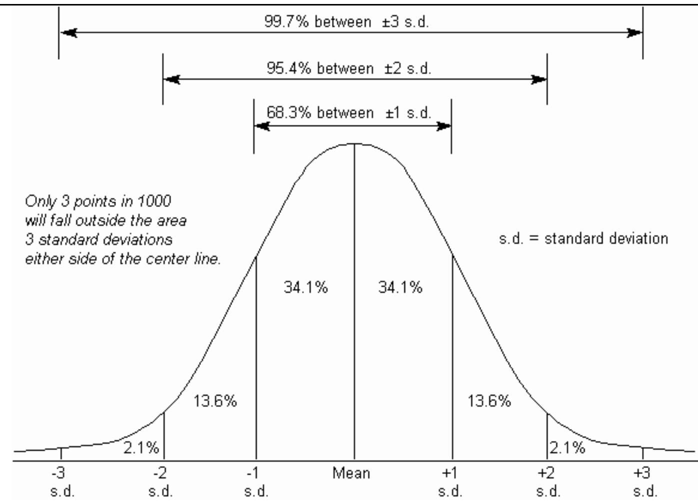
(modified from a lab created by Crystal Jenkins Stawiery)

## Part I. Standard Deviation and Standard Error

When you measure something (like the number of drops of water that can stay on a penny without overflowing), how do you know whether that measurement is valid? Is it representative of *all* pennies? All droppers? All water?

One way to improve the quality of collected data is to increase the sample size, calculate a mean (an average), and then to determine the ***standard deviation***. Standard deviation (often reported as "+/-" ) shows how much variation there is from the mean. When data points are *close* together, the standard deviation will be *small*. If data points are *spread out*, the standard deviation will be *larger.*

Typical data will show a ***normal distribution*** (a bell-shaped curve). In a normal distribution, about 68% of values are within one standard deviation of the mean, 95% of values are within two" standard deviations of the mean, and 99% of the values are within three standard deviations of the mean. The formula for standard deviation is shown below, where $\bar{x}$ is the mean, $x_i$ is any given data value, and n is the sample size. Consider the following sample problem.




**Grades on the most recent AP Biology quiz were as follows: 96, 96, 93, 90, 88, 86, 86, 84, 80, 70.**

**Step 1**: Find the **Mean ( )**. _____________
**Step 2**: Determine the **Deviation ($x_i$ - )$^2$** from the mean for each value and square it, then add up all of the total values. ______________
**Step 3**: Calculate the **Degrees of Freedom (n-1)**. ______________
**Step 4**: Put it all together to find **s**. _______________

Standard Deviation

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

If you're like me, you're going to want to organize this into a table. If you can, try to solve what's below without looking at the answers on the bottom (or, at least, try to do that as much as possible).

| | | | | | | | | | | | Total | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Scores | 96 | 96 | 93 | 90 | 88 | 86 | 86 | 84 | 80 | 70 | | |
| $x_i - \bar{x}$ | | | | | | | | | | | | |
| $(x_i - \bar{x})^2$ | | | | | | | | | | | | |

| | | | |
|---|---|---|---|
| $\Sigma (x_i - \bar{x})^2$ = _____ | n-1 = ___. | $\Sigma (x_i - \bar{x})^2$/ n-1=______ | s=_____ |

In problem above, here are the values you should be getting:

- the mean is 87
- The sum of $x_i - \bar{x}$ is 556.9
- The number of samples is 10, so n - 1 = 9
- 556.9/9= 61.87
- The square root of 61.87 is 7.8. Working with significant figures, that rounds to 8. That's our standard deviation.

So, in this sample, one standard deviation would be (87- 8) through (87 + 8), or 79 to 95. That means that 68% of the data should fall between these numbers. Two standard deviations would be (87 - 16) through (86 + 16), or 71 to 103. At two standard deviations, 95% of the data should fall between these numbers. Three standard deviations would be (87 - 24) through (87 + 24), or 63 to 111. At three standard deviations, 99% of the data should fall between these numbers.

*Standard error of the mean* is used to represent our uncertainty in estimating the mean. It accounts for both sample size and variability. The formula used to calculate standard error of the mean is shown below. As standard error grows smaller, the likelihood that the sample mean is an accurate estimation of the population increases.

Using the data from the standard deviation example above, the mean is 87 and the standard deviation is 8. Plug in the numbers (remembering that ***n*** is 10).

Standard Error

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

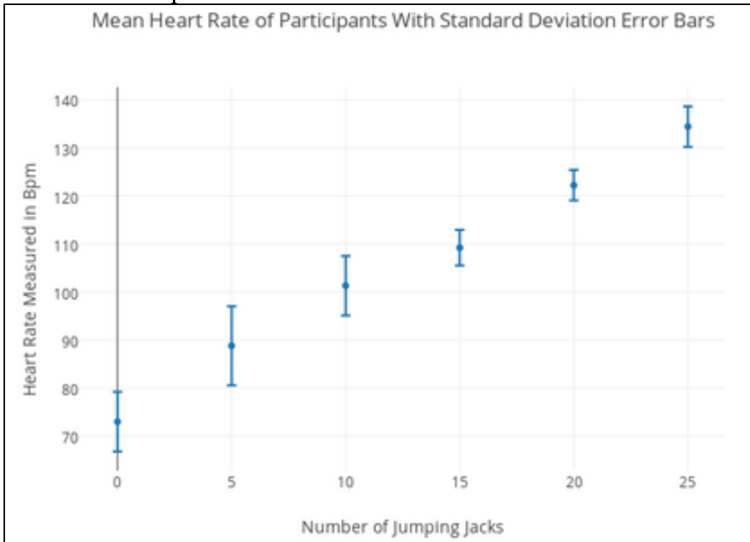
8/3.16 = 2.5.

So, our standard error of the mean equals 2.5 This means that measurements vary by +/- 2.5 from the mean.

(The next section is slightly adapted from [https://www.biologyforlife.com/interpreting-error-bars.html](https://www.biologyforlife.com/interpreting-error-bars.html))

We use standard error of the mean to draw error bars. An error bar is a line through a point on a graph, parallel to one of the axes, which represents the uncertainty or variation of the corresponding coordinate of the point.

Here's an example:

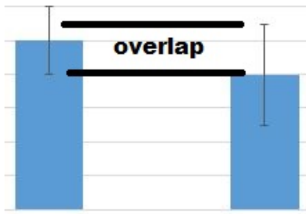
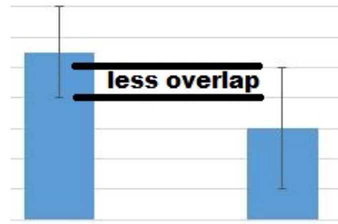
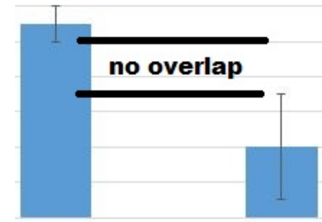



Error bars can communicate the following information about your data:

- How spread the data are around the mean value (small SD bar = low spread, data are clumped around the mean; larger SD bar = larger spread, data are more variable from the mean).
- The reliability of the mean value as a representative number for the data set. In other words, how accurately the mean value represents the data (small SD bar = more reliable, larger SD bar = less reliable). It's important to note that just because you have a larger SD, it does not indicate your data is not valid. Biological measurements are notoriously variable.
- The likelihood of there being a significant difference between data sets.

A "significant difference" means that the results that are seen are most likely not due to chance or sampling error. In any experiment or observation that involves sampling from a population, there is always the possibility that an observed effect would have occurred due to sampling error alone. But if result is "significant," then the investigator may conclude that the observed effect actually reflects the characteristics of the population rather than just sampling error or chance.

The standard deviation error bars on a graph can be used to get a sense for whether or not a difference is significant. Look for overlap between the standard deviation bars:

| | | |
|---|---|---|
|   |   |   |
| When standard deviation errors bars overlap quite a bit, it's a clue that the difference is *not* statistically significant. | When standard deviation errors bars overlap even less, it's a clue that the difference is *probably not* statistically significant. | When standard deviation error bars do not overlap, it's a clue that the difference may be significant. |

**<u>Standard Deviation Practice</u>**

| | | |
|---|---|---|
| Standard Deviation<br>$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$ | Standard Error<br>$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$ | A group of students if measuring the number of stomata (a type of pore used to let in carbon dioxide and control water loss) / $cm^2$ on the bottom surface of sunflower leaves. They took six measurements. Figure out the standard deviation and the standard error. |

| **Sunflower Plant** | 1 | 2 | 3 | 4 | 5 | 6 | TOTAL | Mean ($\bar{x}$) | |
|---|---|---|---|---|---|---|---|---|---|
| **Stomata** (*per $cm^2$*) | 88 | 93 | 90 | 92 | 75 | 78 | | | |
| $x_i - \bar{x}$ | | | | | | | | | |
| $(x_i - \bar{x})^2$ | | | | | | | | | |
| **Σ** $(x_i - \bar{x})^2 =$ _____ | | n-1 = ___. | | | **Σ** $(x_i - \bar{x})^2/$ n-1=_______ | | | s=_____ | $SE_x$= |

**TEACHER PAGE: (solutions)**

| Problem 1, quiz scores | | | | | | | | | | | TOTAL | MEAN | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| scores | 96 | 96 | 93 | 90 | 88 | 86 | 86 | 84 | 80 | 70 | 869 | 86.9 | |
| sample - mean | 9.1 | 9.1 | 6.1 | 3.1 | 1.1 | -0.9 | -0.9 | -2.9 | -6.9 | -16.9 | | | |
| (sample-mean) squared | 82.81 | 82.81 | 37.21 | 9.61 | 1.21 | 0.81 | 0.81 | 8.41 | 47.61 | 285.61 | 556.9 | | |
| | | | | | | | | | | | 61.8777 | divide by n-1 | |
| | | | | | | | Total | Mean | | | 7.86624 | take square root | |
| sunflower problem | 88 | 93 | 90 | 92 | 75 | 78 | 516 | 86 | | | | | |
| sample - mean | 2.00 | 7.00 | 4.00 | 6.00 | -11.0 | -8.0 | | | | | | | |
| (sample-mean) squared | 4 | 49 | 16 | 36 | 121 | 64 | 290 | | | | | | |
| | | | | | | | 58 | divide by n-1 | | | | | |
| | | | | | | | 7.62 | take square root | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | 2.45 | sqrt n) | | | | | |
| | | | | | | | 3.1091 | SE x | | | | | |