2014

Using BioInteractive Resources to Teach

Mathematics and Statistics in Biology

Paul Strode, Ph.D.
Fairview High School
Boulder, Colorado

Ann Brokaw
Rocky River High School
Rocky River, Ohio

Last Update: April 2014

Using BioInteractive Resources to Teach

Mathematics and Statistics in Biology

Measures of Variability: Range, Standard Deviation, Variance

Variability describes the extent to which numbers in a data set diverge from the central tendency or average. 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.

Application in Biology

Students in a biology class measured the width of eight leaves from eight different maple trees and recorded their results in **Table 3.**

Table 3. Width of Maple Tree Leaves

Plant #	1	2	3	4	5	6	7	8
Width (cm)	7.5	10.1	8.3	9.8	5.7	10.3	9.2	8.7

To determine the range of leaf widths:

I. Identify the largest and smallest values in the data set:
Largest = 10.3 cm, Smallest = 5.7 cm

II. To determine the range, subtract the smallest value from the largest value:
Range = 10.3 cm – 5.7 cm = 4.6 cm

Standard Deviation

The standard deviation is the most widely applied measure of variability. The sample mean ($\overline{x}$) provides a measure of the central tendency of the sample; the **sample standard deviation** (s) measures the average deviation between each measurement in the sample and the mean ($\overline{x}$). The value of s is not absolute but varies from sample to sample. It is an estimate of the standard deviation in the larger population.

To calculate sample standard deviation:

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

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 of the squared differences $\Sigma(x_i - \overline{x})^2$.

5. Divide by the **degrees of freedom**, which is one less than the sample size ($n - 1$).

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

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*) will grow 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 $n = 17$). Each plant is measured from the tips of the roots to the top of the tallest stem. The measurements are recorded in **Table 4**, along with the steps for calculating standard deviation.

Table 4. Plant Measurements and Steps for Calculating Standard Deviation.

Plant #	Plant Height (mm)	*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* $\sum(x_i - \overline{x})^2$ = 2205
		Variance (s^2)	*Step 5* $\sum(x_i - \overline{x})^2/(n-1)$ = 2205/16 = 138
		Standard Deviation (s)	*Step 6* $\sqrt{s^2}$ = $\sqrt{138}$ = 11.7

Summary:

The mean height of the bean plants in this sample is 103 mm $\pm$ 11.7 mm. What does this mean? We estimate that this sample of 17 observations is drawn from a population with a standard deviation of 11.7 mm. In a data

Measures of Confidence: Standard Error and 95% Confidence Intervals

The **sample mean** provides an estimate of the **population mean**, but it is not necessarily identical to the population mean. The uncertainty of the estimate can be expressed by calculating the standard error of the mean ($SE_{\bar{x}}$) or by calculating the 95% confidence interval (95% CI).

Both standard error of the mean (SE) and the 95% confidence interval (95% CI) use the sample standard deviation (s) in their calculations.

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

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

$$95\%\ CI = \frac{1.96s}{\sqrt{n}}$$

Note: The equation for 95% CI is $\frac{1.96s}{\sqrt{n}}$ but typically rounded to $\frac{2s}{\sqrt{n}}$.

Standard error of the mean is different from standard deviation (s): standard deviation measures the distribution of individual data points around the mean, while standard error of the mean provides an estimate of how precisely the sample mean represents the population mean. (It is also known as the standard error of the sample means distribution.)

The 95% confidence interval (95% CI) is equivalent to 1.96 (typically rounded to 2) standard errors above and below the mean. This does not mean that there is a 95% chance that the true population mean falls within these values but rather that we are 95% confident that the population mean falls within about 2 SE of a sample mean. Another way of stating this is, that if we were to calculate a 95% *CI* for each of a large number of samples in a population, then 95% of these 95% *CI*s would contain the true population mean.

Note: Both $SE_{\bar{x}}$ 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 ($SE_{\bar{x}}$) or the 95% CI as error bars in a bar graph provides a clear visual clue as to whether the means of two samples are likely to represent a significant difference between the means of two populations.

Application in Biology—Example 1

Seeds of many weed species germinate best in recently disturbed soil that lacks a light-blocking canopy of vegetation. Students in a biology 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. Half of the Petri dishes were placed in the dark and half placed 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 standard error of the mean and 95% confidence intervals (**Figure 3**).

Table 5. Number of Crofton (*Ageratina adenophora*) Seeds that Germinated After 72 Hours in the Dark and in the Light. (The number of replicates (i.e., sample size; *n*) equals 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 ($\bar{x}$)	$\bar{x}_1 = 9.6$	$\bar{x}_2 = 18.4$	$\frac{\sum (x_i - \bar{x}_1)2}{n-1} = \frac{158.4}{9}$	$\frac{\sum (x_i - \bar{x}_2)2}{n-1} = \frac{98.4}{9}$
		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 $SE_{\bar{x}} = \frac{s}{\sqrt{n}}$			$SE_{\bar{x}} = \frac{4.20}{\sqrt{10}} = 1.33$	$SE_{\bar{x}} = \frac{3.31}{\sqrt{10}} = 1.05$
			$95\%\ CI = \frac{2(4.20)}{\sqrt{10}} = 2.7$	$95\%\ CI = \frac{2(4.74)}{\sqrt{10}} = 2.1$

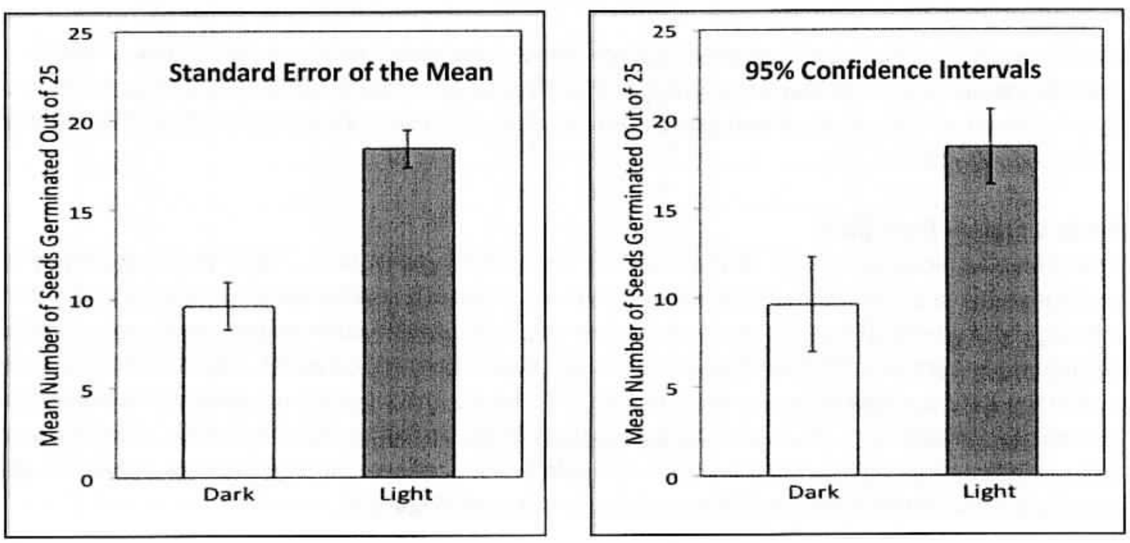

Figure 3. Mean number of Crofton (*Ageratina adenophora*) seeds that germinated after 72 hours in the dark or in the light. The Standard Error of the Mean graph shows the $SE_{\bar{x}}$ as error bars, and the 95% Confidence Interval graph shows the 95% CI as error bars.