3. Standard Deviation and Standard Error[1]

This worksheet provides you with practice working with standard deviation and standard error for different data sets. For some extra background or review on these topics before you practice, Mr. Anderson has some videos that will help you. For review of the mean and median and their computation, watch the video entitled *Statistics for Science*. We use the standard deviation to compute the standard error, and you can review standard deviation in the video entitled *Standard Deviation*. For review of standard error, watch the video entitled *Standard Error*. The links are provided below.

(a) *Statistics for Science*:
`https://www.youtube.com/watch?v=jf9VT4V4aRI`

(b) *Standard Deviation*:
`https://www.youtube.com/watch?v=09kiX3p5Vek`

(c) *Standard Error*:
`https://www.youtube.com/watch?v=BwYj69LAQOI`

Complete the questions in the Warm Up section before going on to the Further Practice section. The questions in the Warm Up section review the basics and intend to prepare you for more practice in computing and interpreting standard deviations and standard errors.

Warm Up

1. The formula for the standard deviation, S, is:

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

Identify what each of the following parts of the formula mean by explaining it in words.

(a) n _____

(b) $\bar{x}$ _____

(c) $\sum$ _____

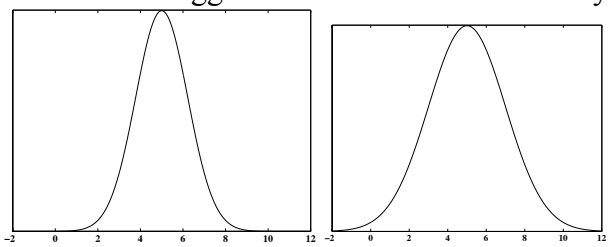
(d) A name for the quantity $n - 1$ is _____[2]

---

[1] Standard Error here refers to Standard Error of the mean (as is commonly used).

[2] You use this quantity in many calculations, including the χ^2 test.

2. Consider the two figures below. Each shows a distribution of data with a mean, $\bar{x}$, of 5. Which has a bigger standard deviation and why?



3. The formula for the standard error is:

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

Identify what each of the following parts mean by explaining it in words.

(a) S _____

(b) n _____

4. Look at the formulas for standard deviation and for standard error.

(a) Explain, in words, the difference between the standard deviation and the standard error.

(b) You have three data sets with the same standard deviation, $S = 3.298$. *Data Set 1* has ten observations in it (n=10), *Data Set 2* has twenty (n=20), and *Data Set 3* has fifty observations in it (n=50). For each of these calculate the standard error.

Data Set 1: _____ *Data Set 2*: _____ *Data Set 3*: _____

(c) Explain how the standard error changes when the sample size changes (but the standard deviation stays the same). Then, explain how the formula for the standard error justifies this change.

WARM UP SOLUTIONS

1. ***(a)*** The variable ***n*** stands for the sample size, i.e. the number of observations, in the data set. ***(b)*** The $\bar{x}$ represents the mean or average value for the data set. ***(c)*** This symbol is a summation sign. It means we need to calculate the quantity $x - \bar{x}$ for every value in the data set. Then we add all of these differences together. ***(d)*** The quantity ***n – 1*** is called the degrees of freedom.

2. The graph to the reader's right has a bigger standard deviation. You can tell this because the peak is broader than the peak in the other graph. This demonstrates that more of the values in that data set are farther away from the mean than in the data set on the left. Such a set will have a bigger standard deviation.

3. ***(a)*** In the formula for standard error, the ***S*** represents the standard deviation***.*** ***(b)*** The ***n*** once again represents the number of observations in the data set, just as it did in question *1*.

4. ***(a)*** The standard error formula is a modification of the standard deviation formula. To get the standard error, one takes the standard deviation formula and divides it by the square root of the sample size in the data set. This indicates that if two data sets have the same standard deviation but a different number of observations, they will have different standard errors. ***(b)*** For *Data Set 1*, $SE = \frac{3.298}{\sqrt{10}} = 1.043$. For *Data Set 2*, $SE = \frac{3.298}{\sqrt{20}} = 0.737$. For *Data Set 3*, $SE = \frac{3.298}{\sqrt{50}} = 0.466$. ***(c)*** When the sample size gets larger the standard error gets smaller because the standard deviation is the same. This is clear from the standard error formula as described in part *(a)*. The standard error is just the standard deviation divided by the square root of the sample size. Larger numbers have larger square roots. Dividing the same number by a bigger number results in a smaller number. This is exactly what we see in part *(b)*.

Further Practice

1. In relation to AP Biology Investigation #11, you and your lab partner record the following counts of stomata in sunflower leaves.

Table 1: Stomata per Examination Area

Sunflower Plant	**1**	**2**	**3**	**4**	**5**	**6**
Stomata (*per examination area*)	**88**	**93**	**90**	**92**	**75**	**78**

(a) Calculate the mean or average number of stomata for these sunflower leaves, $\bar{x}$.

(b) Order the number of stomata from lowest to highest and calculate the median number of stomata for the sunflower leaves.

(c) Calculate the standard deviation of the number of stomata for the sunflower leaves.

(d) Calculate the standard error in the number of stomata for the sunflower leaves.

(e) Explain in words what the difference in median and mean/average values means.

(f) Make a (very simple) bar graph with the mean of the number of stomata for the sunflower leaves. Draw the error bars on the graph. Interpret the standard error values.

2. Repeat the exercises in *1(a) – (f)* using a spreadsheet for the data set listed below.

Sunflower Plant	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Stomata (*per examination area*)	**67**	**85**	**90**	**100**	**72**	**79**	**99**	**84**	**95**	**103**	**88**	**93**	**90**	**92**	**75**	**78**

3. Related to AP Biology Investigation #12, three classes study when fruit fly populations choose Food A over Food B. Each pair of partners in the classes record the number of times that Food A is chosen in preference to Food B in a total of 25 trials. The data from each pair for the three classes is shown below. Use the data to answer the questions below.

	Group 1	***Group 2***	***Group 3***	***Group 4***	***Group 5***	***Group 6***	***Group 7***	***Group 8***
Class 1	12	16	11	11	13	14	12	15
Class 2	10	9	18	8	16	18	13	12
Class 3	4	19	6	20	12	13	23	7

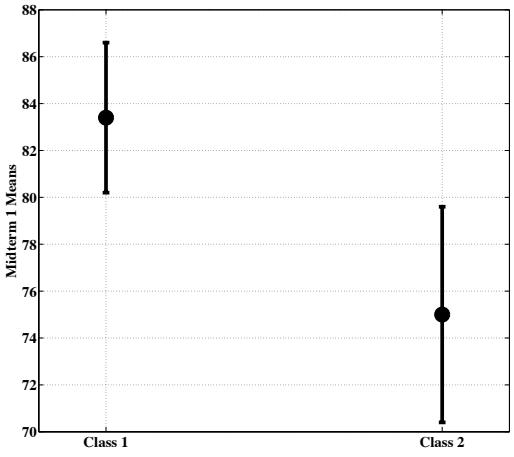
(a) Verify that each class' data set has the same mean and median.

(b) Based on the data that is given, which do you expect to have a bigger standard deviation? Which do you expect to have the smallest standard deviation? *Explain* your answer.

(c) Find the standard deviation for each data set. If your answers do not match your predictions, make sure to go back and explain how you can predict the ranking (i.e. smallest and largest) standard deviations from the data that are given.

4. Two different AP Biology instructors compute the means and standard errors for the first exam score for their two different AP Biology classes. The means and the SE bars are shown in the graphs below. For each of the teacher's sets determine whether the difference between the means of the two classes is (a) definitely significantly different, (b) definitely not significantly different, or (c) unknown based on the graph whether they are significantly different or not. As usual, explain your answer.

Teacher 1



Teacher 2