WLHS / AP Bio / Monson

KEY

Name________________________
Date__________________ Per________

PACKET: Statistics Practice Problems - 1

This packet provides you with practice working with standard deviation and standard error for different data sets. For extra background or review on these topics, Mr. Anderson (Bozeman Science) has a few podcasts that will help you (links below). For review of the **mean** and **median** and their computation, watch "Statistics for Science." We use the **Standard Deviation** to compute the **Standard Error**, so you can review that in "Standard Deviation." And for a review of Standard Error, watch "Standard Error."

LINKS to Mr. Anderson's Tutorials:
Statistics for Science: http://www.bozemanscience.com/statistics-for-science
Standard Deviation: http://www.bozemanscience.com/standard-deviation
Standard Error: http://www.bozemanscience.com/standard-error

DATA SET #1: You are helping out at a veterinary office with a litter of new puppies. The birth weights of the puppies are shown below.

Table 1: Birth Weights of Puppies

Birth Order:	1	2	3	4	5	6
Puppy Weight (lbs):	1.2	1.7	1.6	1.5	1.0	0.8

A) Calculate the **mean** or average birth weight for this litter of puppies.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\frac{(1.2+1.7+1.6+1.5+1.0+0.8)}{6} = \boxed{1.3 \text{ lbs}}$$

B) Order the puppies' birth weights from lowest to highest and calculate the **median** birth weight for the litter of puppies.

$0.8,\ 1.0,\ \underbrace{1.2,\ 1.5},\ 1.6,\ 1.7$

$$\text{median} = \frac{(1.2+1.5)}{2} = \boxed{1.35 \text{ lbs}}$$

C) Calculate the **standard deviation** of the birth weights for this litter of puppies.

$$s = \sqrt{\frac{[(1.2-1.3)^2+(1.7-1.3)^2+(1.6-1.3)^2+(1.5-1.3)^2+(1.0-1.3)^2+(.8-1.3)^2]}{(6-1)}} = \sqrt{\frac{.64}{5}} = \boxed{0.36}$$

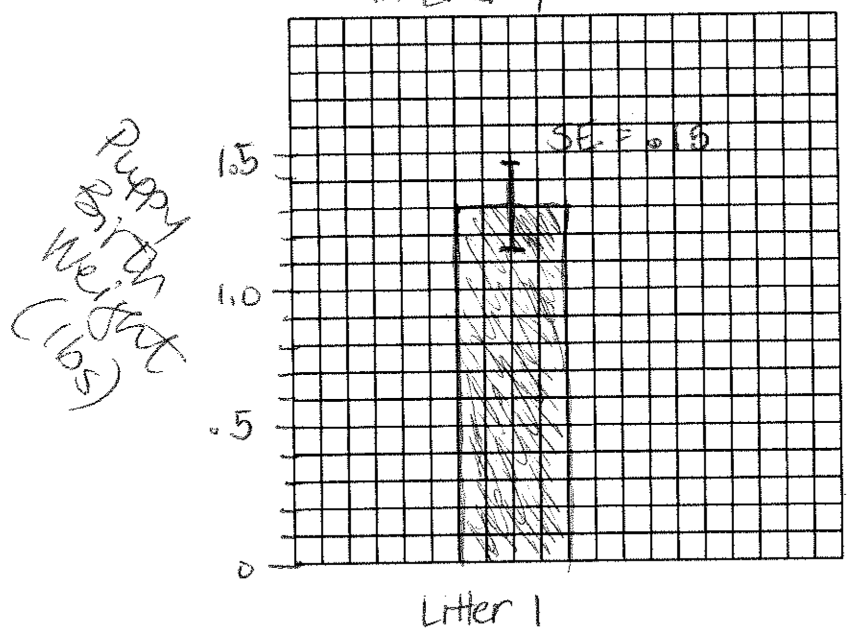
D) Calculate the **standard error** in the birth weights for this litter of puppies.

$$SE = \frac{s}{\sqrt{n}} = \frac{0.36}{\sqrt{6}} = \boxed{0.15}$$

E) Explain, in your own words, what the (slight) difference in median and mean /average means.

Mean/average includes outlying data pts. (i.e. smallest/largest), whereas median is the middle value, so it is not sensitive to outlying data pts.

F) Make a (very simple) bar graph with the mean of the puppies' birth weights. Draw the error bars (using Standard Error) on the graph.



Review of Standard Deviation and Standard Error (of the Mean):

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 = # of samples taken / data points

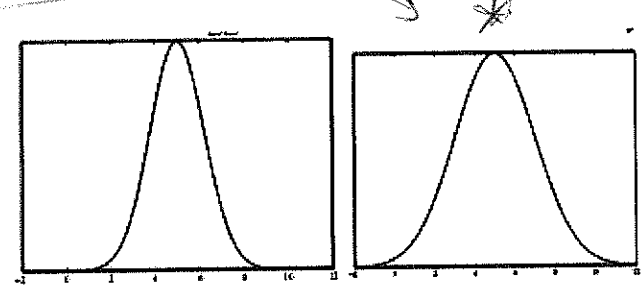
B) $\bar{x}$ = mean / average of all data points / measurement

C) Σ = sum of

D) A name for the quantity n-1 is: degrees of freedom

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

* second graph has a bigger Std. deviation. This is because the "normal" curve is wider, reflecting a larger "spread" of data



3) The formula for the **standard error** (of the mean) 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 = standard deviation

B) n = # of samples / data points / measurements taken

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.**

Std. deviation is the "spread" of values in a sample, stays same when sample size ↑

Std. error takes pop/sample size into account; as n ↑, SE ↓

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.

$\frac{3.298}{\sqrt{10}}$ **Data Set 1:** 1.04 $\frac{3.298}{\sqrt{20}}$ **Data Set 2:** 0.74 **Data Set 3:** 0.47 $\frac{3.298}{\sqrt{50}}$

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.

As sample size goes up, SE goes down. The formula for SE takes Std. dev. & divides it by $\sqrt{n}$.

Further Practice:

1) In an AP Biology investigation, 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, $\bar{x}$.

$\bar{x} = \frac{88+93+90+92+75+78}{6}$

$= \boxed{86 \text{ stomata}}$

B) Order the number of stomata from lowest to highest and identify the **median** number of stomata for the sunflower leaves.

75, 78, 88, 90, 92, 93

median $= \frac{88+90}{2} = \boxed{89}$

C) Calculate the **standard deviation** of the number of stomata for the sunflower leaves.

$S = \sqrt{\frac{(88-86)^2+(93-86)^2+(90-86)^2+(92-86)^2+(75-86)^2+(78-86)^2}{(6-1)}} = \boxed{7.6}$

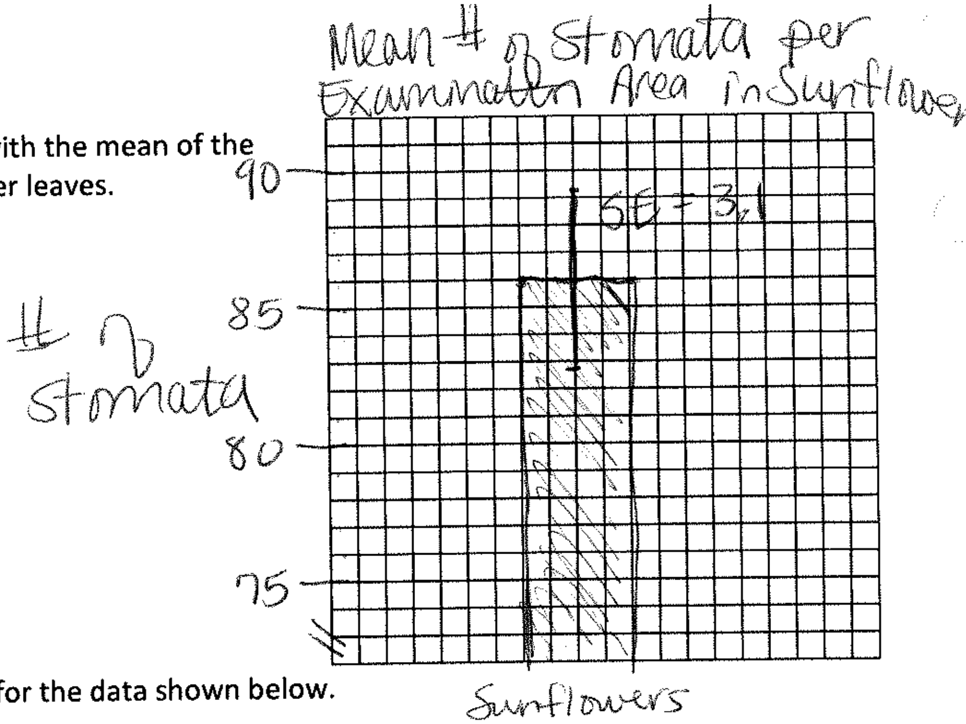
D) Calculate the **standard error** in the number of stomata for the sunflower leaves.

$SE = \frac{S}{\sqrt{n}} = \frac{7.6}{\sqrt{6}} = \boxed{3.1}$

E) 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.

$\bar{X} = 86$ stomata

$SE = 3.1$



2) Repeat the exercises in 1 (A)-(E) for the data shown 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

67 90
72 90
75 92
78 93
79 95
84 99
85 100
88 103

A) $\bar{X} = \frac{1390}{16} = \boxed{87 \text{ stomata}}$

B) median $= \frac{88+90}{2} = \boxed{89 \text{ stomata}}$

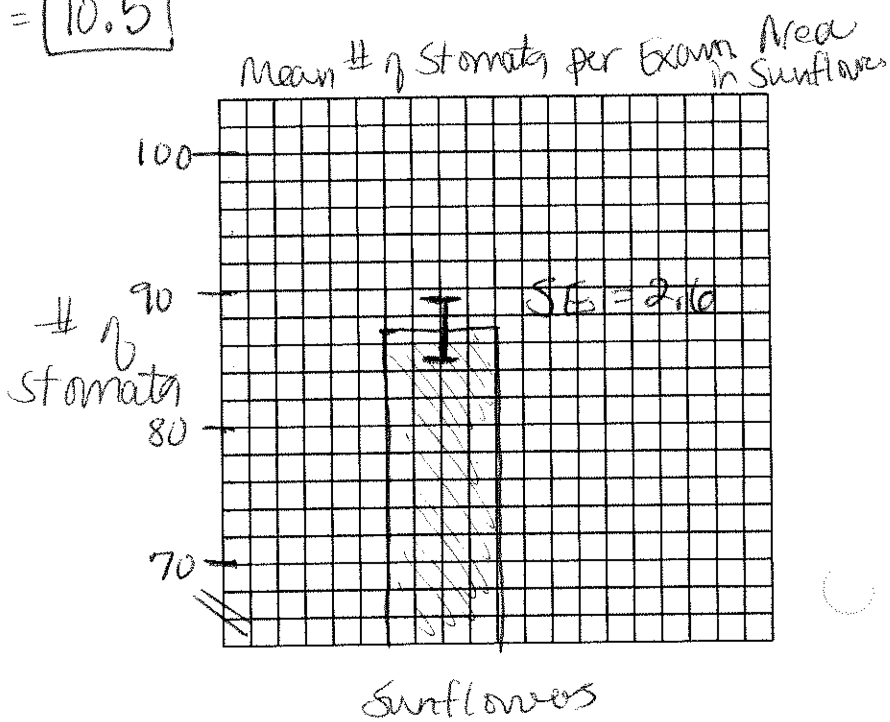
C) $S = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{15}} = \sqrt{\frac{1640}{15}} = \boxed{10.5}$

D) $SE = \frac{10.5}{\sqrt{16}} = \boxed{2.6}$

E) (use graph area provided):

$\bar{X} = 87$ stomata

$SE = 2.6$



3) In an AP Biology investigation, three classes studied how fruit fly populations choose between two different sources: Food A or 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.

<u>Data Table</u>: Fruit Fly Food Selection – Frequency of Food A Selection (each trial = sample size of 25)

	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

11, 11, 12, 12, 13, 14, 15, 16
8, 9, 10, 12, 13, 16, 18, 18
4, 6, 7, 12, 13, 19, 20, 23

A) Verify that each class' data set has the same **mean** and **median**.

Class 1 : $\bar{x} = 13$ median = 12.5
Class 2 : $\bar{x} = 13$ median = 12.5
Class 3 : $\bar{x} = 13$ median = 12.5

B) Based on the data given, which do you expect / predict to have a bigger standard deviation? Which do you expect to have the smallest standard deviation? **EXPLAIN** you answer.

Class 3 will have bigger S.D. → larger "spread" of data (4 → 23)
Class 1 will have smallest S.D. → tighter "spread" of data (11 → 16)

C) **Find the standard deviation for each class / data set.** If your answers do not match your predictions above, make sure to go back and explain how you can predict the ranking (i.e. the smallest and largest) standard deviations from the data that are given. ~~Class 1~~ SD $= \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{7}} =$

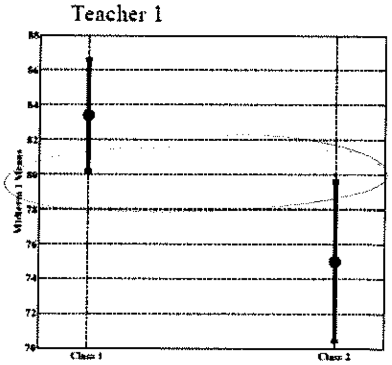
Class 1 SD = 1.85

Class 2 SD = 3.96

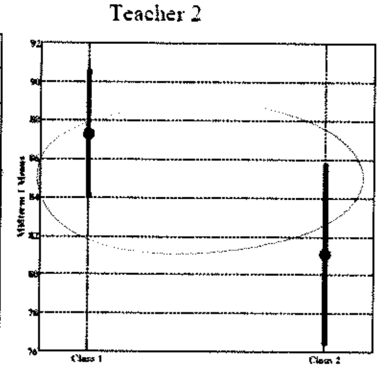
Class 3 SD = 7.1

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. 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. **EXPLAIN** your answer for each teacher's set.

Teacher 1



Teacher 2



Teacher 1 : A) the means are def. significantly different b/c class 1 $\bar{x} = 83.5$ + class 2 $\bar{x} = 75$, + the SE bars don't overlap.

Teacher 2 : B) the means are def. NOT significantly different b/c class 1 $\bar{x} = 87.3$ + class 2 $\bar{x} = 81$, but error bars overlap.