Name Key

8A; Algebra 1

Date ____________

Period____________

Calculating Standard Deviation

Standard deviation is used to tell how far on average any data point is from the mean. The smaller the standard deviation, the closer the scores are on average to the mean. When the standard deviation is large, the scores are more widely spread out on average from the mean. (like MAD)

The standard deviation is the square root of the variance. $\sqrt{\text{variance}}$ = standard deviation

Variance measures how far a set of data is spread out. A variance of zero indicates that all of the data values are identical. All non-zero variances are positive (B/C x^2). A small variance indicates that the data points tend to be very close to the mean, and to each other. A high variance indicates that the data points are very spread out from the mean, and from one another.

Variance is the **average of the squared distances** from each point to the mean. $(\text{standard deviation})^2$ = variance

Population vs. Sample Standard Deviation

When we are working with **every possible data point** of interest, we call this a **population** and use the population standard deviation, σ (lower case sigma (population)). When we have only a **sample** of all possible values we use the sample standard deviation, S (capital S (sample)). The formulas for these two differ very slightly, so their values tend to be slightly fferent. (see below)

What type of standard deviation should be used?

a) In a study of the heights of Merrick's teens, 100 students' heights were recorded.

Sample

b) In a study of the land areas of the states of the United States, the area of each state is used.

population

There will be a lot of new symbols used in this lesson. Here is a list of what some of them mean

Capital sigma → $\sum$ = "the sum of ..."

n = number of pieces of data (population)

$n-1$ = number of pieces of data (sample)

$\overline{x}$ = mean (average) of data

x_i = each of the values in the data

$x_1, x_2, x_3, x_4, \ldots x_n$ (as i goes from 1 to n)

- Standard Deviation is better than MAD B/C it's more specific to population vs. sample
- Standard Deviation is used for symmetrical data

Practice Problem #1: Calculate the standard deviation of the following test data by hand. Use the chart below to record the steps.

Test Scores: 22, 99, 102, 33, 57, 75, 100, 81, 62, 29

Mean: 66

n: 10
population (# of data values)

$n - 1$: 10-1 = 9
sample

Test Score (x)	Difference from the mean $(x - \bar{x})$	(Difference from the mean)^2 $(x - \bar{x})^2$
22	22-66 = -44	$(-44)^2 = 1936$
29	29-66 = -37	$(-37)^2 = 1369$
33	33-66 = -33	$(-33)^2 = 1089$
57	57-66 = -9	$(-9)^2 = 81$
62	62-66 = -4	$(-4)^2 = 16$
75	75-66 = 9	$(9)^2 = 81$
81	81-66 = 15	$(15)^2 = 225$
99	99-66 = 33	$(33)^2 = 1089$
100	100-66 = 34	$(34)^2 = 1156$
102	102-66 = 36	$(3\cancel{4})^2 = 1296$
	Sum of (Difference from the mean)^2 $\sum(x - \bar{x})^2$	+ 8338

Sum of (Difference from the Mean)^2 divided by n = $\sum\frac{(x-\bar{x})^2}{n} = \frac{8338}{10} = 833.8$ (This is Population Variance)

Sum of (Difference from the Mean)^2 divided by $(n - 1)$ = $\sum\frac{(x-\bar{x})^2}{(n-1)} = \frac{8338}{9} = 926.4$ (This is Sample Variance)

Standard Deviation Squared

Final Step: Standard deviation = square root of what you just calculated (variance).

Population Standard deviation = $\sqrt{\sum\frac{(x-\bar{x})^2}{n}} = \sqrt{\frac{8338}{10}} = \sqrt{833.8} = 28.88$

Sample Standard deviation = $\sqrt{\sum\frac{(x-\bar{x})^2}{(n-1)}} = \sqrt{\frac{8338}{9}} = \sqrt{926.4} = 30.44$

Average distance from mean

Show calc

Measures of Variation describe the dispersion or spread of a set of data. The VARIANCE and STANDARD DEVIATION describe how closely a set of data clusters about the mean. Referring back to the "Do Now", do you predict that Lisa or Melissa will have a higher variance value?

We use this formula: $\frac{\sum(x_i - \bar{x})^2}{n}$ or $\frac{\sum(x_i - \bar{x})^2}{n-1}$ to find the variance and we use this chart to help.

Population Sample

Calculator for standard deviation:
STAT 1: Edit put #'s into L_1
STAT → CALC 1: 1-var stats Calculate
* look for Sx (sample) or σx (population)

Lisa's data:

x_i	$\bar{x}$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
50	84	50-84	$(-34)^2 = 1156$
98	84		
100	84		
94	84		
78	84		

* population: B/c we are using all of their test grades

② Variance = $(\sigma x)^2 = 348.80$

population

Do 1st ① Standard Deviation = $\sigma x = 18.68$

* must figure S.D. out 1st so it is stored in the calculator

Round to the nearest hundreth unless stated otherwise

** Don't just square this # to get the variance B/c it will make the variance off by a little

Let's let the calculator do the work for us for Melissa's data! (85, 84, 83, 86, 82)

Calculator:

clear out L1 and L2
enter x_i data values into L1
STAT → CALC → 1-Var Stats L1

Do 1st → ① **Standard deviation** = $\sigma x = 1.41$
② **Variance** = $(\sigma x)^2 = 2.00$
population

The calculator will give us the standard deviation. So, how would we find the variance?

Calculator: Vars 5: STATS 4: σx ^{2} or 3: Sx ^{2}
Population Sample

Examples:

1: Ten Merrick students are chosen at random to report their number of text messages in a half hour. Their numbers of text messages were: 15, 13, 12, 10, 9, 7, 5, 4, 3, 2

a) Determine if this data set is a population or a sample:

Sample

b) Find the mean:

$\bar{x} = 8$ $\frac{\sum x}{n}$ $\frac{80}{10} = 8$

* nearest hundredth

c) Standard deviation:

$Sx = 4.50$ → sample standard deviation

d) Variance:

$(Sx)^2 = 20.22$

* Can only do shortcut with regu...

opt 4 ☆☆ L_1, L_2 shortcut in calc → put grade into L_1 + put Freq # into L_2 then press STAT CALC 1 + put L_2 into Freq. List. To get L_2 press 2nd 2

2: The following table shows English test scores for a whole class: A

a) Determine if this data set is a population or a sample:

Population

b) Find the mean:

$\bar{X} = 79.44$ $\quad \frac{\Sigma x}{n} \frac{2860}{36} = 79.44$

c) Standard deviation to the nearest tenth:

population standard dev. $\sigma x = 8.4$

d) Variance:

$L_1 L_2$ shortcut in calc doesn't work w/ variance

$(\sigma x)^2 = 70.52$

Grade (L_1)	Frequency (L_2)	
95	4	= 380
85	13	= 1105
75	11	= 825
70	6	= 420
65	2	= 130
	T.F = 36	2860

OR opt 3 ☆ make sure to put every grade into L_1 ex: 95 put in 4 times 85 put in 13 times ← OR →

☆ opt 1 is to list all the #'s out by hand

opt 2 A "by hand" shortcut

3. Ten people attended a talk at a conference. At the end of the talk, the attendees were given a questionnaire that consisted of four questions. The questions were optional, so it was possible that some attendees might answer none of the questions while others might answer 1, 2, 3, or all 4 of the questions (so the possible numbers of questions answered are 0, 1, 2, 3, and 4).

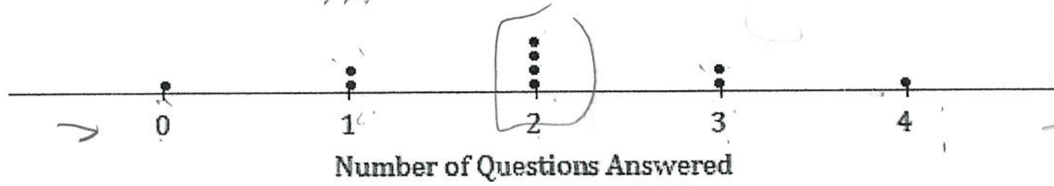
Suppose that the numbers of questions answered by each of the ten people were as shown in the dot plot below.

10 dots = population

Bell curve symmetric

☆ 6 dots are away from the mean

Quantitative



0, 1, 1, 2, 2, 2, 2, 3, 3

shortcut L_1	L_2
0	1
1	2
2	4
3	2
4	1

☆ Remember L_2

Use the statistical features of your calculator to find the mean and the standard deviation, to the nearest hundredth, of the data set. Would you use the population or sample standard distribution for this data?

☆ Population ☆ all 10 people questioned gave an answer

Must be in the Freq List if doing the shortcut

Mean: $\bar{X} = 2.00$ $\quad \frac{\Sigma x}{n} \quad \frac{20}{10}$

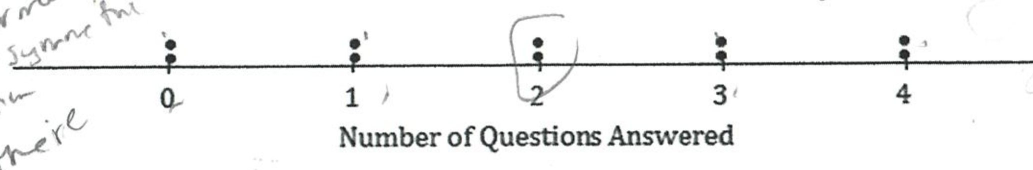
Standard Deviation: $\sigma x = 1.10$

population standard deviation

8 dots are away from the mean

4. Suppose the dot plot looked like this:

normal - symmetric



* population B/C there are 10 dots

Shortcut

L_1	L_2
0	2
1	2
3	2
4	2

* L_2 must be in FreqList

a. Use your calculator to find the mean and the standard deviation, to the nearest hundredth, of this distribution.

mean = $\boxed{\bar{x} = 2.00}$ $\frac{\Sigma x}{n} = \frac{20}{10}$

population standard deviation → $\boxed{\sigma x = 1.41}$

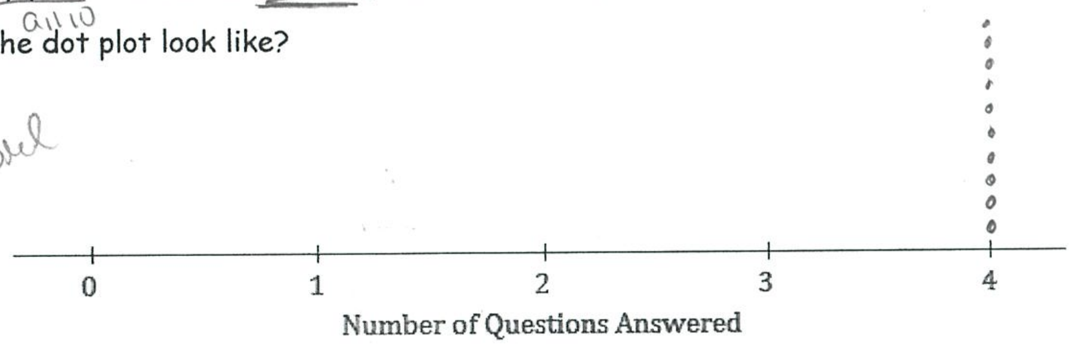
b. Remember that the size of the standard deviation is related to the size of the deviations from the mean. Explain why the standard deviation of this distribution is greater than the standard deviation in Example 3.

There are more points that are farther away from the mean than in Ex #3. (#'s are more spread out)

5. Suppose that every person answers all four questions on the questionnaire.

a. What would the dot plot look like? (all 10)

unimodal



Shortcut

L_1	L_2
4	10

L_2 must be in FreqList

b. What is the mean number of questions answered? (You should be able to answer without doing any calculations!)

$\boxed{\bar{x} = 4}$ $\frac{\Sigma x}{n} = \frac{40}{10} = 4$

c. What is the standard deviation? (Again, don't do any calculations!)

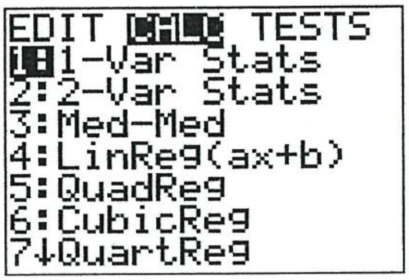
$\boxed{\sigma x = 0}$ * B/C all of the data values are the same as the mean

population standard deviation

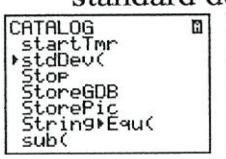
To find standard deviation:

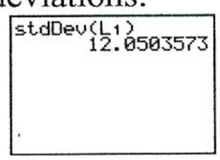
To find standard deviation: *Since this question deals with the complete set, we will be using "population" form, not sample form.*

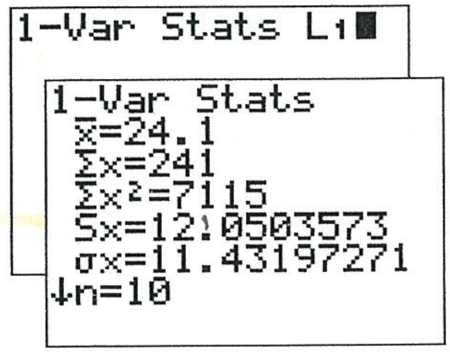
Go to one-variable stats for "population" standard deviation. STAT → CALC #1 1-Var Stats



NOTE! The standard deviations found in the CATALOG, stdDev, and also found by 2nd LIST → MATH #7 stdDev are both **Sample** standard deviations.







$\bar{x}$ = mean

$\sum x$ = sum of the data

$\sum x^2$ = sum of squares of the data

Sx = sample standard deviation

σx = population standard deviation

n = sample size (# of pieces of data)

min X = smallest data entry

Q_1 = first quartile

med = median (second quartile)

Q_3 = third quartile

max X = largest data entry

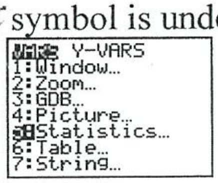
Population Standard Deviation = 11.43

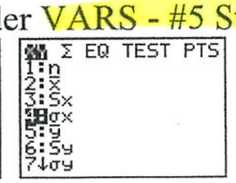
FYI: Using the lists, the calculator can simulate a spreadsheet style "by hand" computation of standard deviation. Click here to see the spreadsheet-style approach.

To find variance:

To find variance: The "population" variance is the square of the population standard deviation.

The σx symbol is under VARS - #5 Statistics





NOTE! The variance found in the CATALOG and also found by 2nd List → MATH #8 variance are both **Sample** variances.