5gg][ba Ybh Standard Deviation and Variance

Remember that the standard deviation and variance can only be used when the mean is used as the measure of center.

The variance is the average of the squares of the deviations of the observations from their mean. Therefore, the variance doesn't have the same unit of measurement as the original data. However, since the standard deviation is the square root of the variance, it *does* have the same unit of measurement as the original data. This is why we often prefer talking about the standard deviation as opposed to the variation. The standard deviation can be thought of as "the average distance of each data point from its mean." Although this is somewhat imprecise, it's a good way of think about the standard deviation.

Be sure that you know when to use the sample standard deviation formula (s) and the population standard deviation formula (σ). This table may help:

Situation	Sample Std. Deviation Formula (s)	Population Std. Deviation Formula (σ)
Calculating std. deviation with all members of the population		**x**
Estimating population standard deviation using sample	**x**	

Remember that the sample standard deviation formula does **not** give you the standard deviation of your sample; it *estimates* the standard deviation of the population.

For example, suppose you have the following 5 test scores out of a population of 30: {87, 92, 74, 70, 66}.

The sample standard deviation

$$s = \sqrt{\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + ... + (x_n - \overline{x})^2}{n-1}} = \sqrt{\frac{(87 - 77.8)^2 + (92 - 77.8)^2 + ... + (66 - 77.8)^2}{4}} = 11.189$$

> is the statistic that you'd use to estimate the standard deviation of the entire population of 30 test scores. The standard deviation of the observation set {87, 92, 74, 70, 66},

$$\sigma = \sqrt{\frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + ... + (x_{30} - \mu)^2}{30}}$$

would give you the standard deviation of all 30 data points in the population.

To avoid ambiguity when you're communicating with others, don't call a data set a sample unless you intend to use it to infer something about the population from which it's drawn.

5gg][ba Ybh Standard Deviation and Variance

For each calculation, use the symbol for the measure you are calculating. For example, if you're calculating a sample mean, use the symbol $\bar{x}$ when referring to your answer.

When referring to the variance, use the symbols s^2 and σ^2.

Don't forget to include units in your answers.

1. **Do <u>not</u> use your calculator for this one. Write out the formulas and show your work. Round your answers to 3 decimal places. (You'll be able to use your calculator on the problems after this one.)** (4 points)

   You're given a sample with *n*=8 measurements: 3, 1, 5, 6, 4, 4, 3, 5.

   A. Calculate the range (maximum value - minimum value).

   B. Calculate the sample mean.

   C. Calculate the sample variance and sample standard deviation.

   D. The range is approximately how many sample standard deviations?

For the rest of the questions, you can use your calculator. Instructions for calculating summary statistics and parameters (such as σ and $\bar{x}$) are in the Statistics chapter of your TI-83 cf H=, (calculator's manual. If you only need a refresher on how to do it, Zc``ck these instructions:

1. Enter the numbers into a list by pressing ON STAT 1 and entering the numbers into one of the lists. Key in a number, press ENTER, then key in the next number, and so on until you're done. Be sure to check your list once you're done to make sure you entered the correct numbers. (Details are in your calculator manual if you need more help.)

2. To calculate summary statistics for your data set, press STAT, use a blue arrow key to move the cursor over to **CALC**, then press ENTER to select **1-Var Stats**.

3. Press 2nd and then the key that shows the name of your data list printed in yellow. For example, if your data is in L1, press 2nd 1 (the 1 key has [L1] as its 2nd function).

4. Then press ENTER. You'll soon see a list of summary statistics and parameters for your data. Use the down arrow key to scroll down the list.

2. Consider the following set of data, which represents a simple random sample of heights (in inches) of full-grown corn plants in Elephanteyetown, WI: (65.8, 62.1, 61.5, 66.6, 65.1, 60.3, 70.2, 69.1, 68.3, 62.1, 65.1, 63.7, 68.5, 71.4, 65.9, 63.8, 66.7, 63.3, 71.9, 56.8) Round your answers to 2 decimal places. (4 points)

A. Calculate the range.

B. Calculate the statistic that gives an estimate of the population mean.

C. Calculate sample statistics to estimate the population standard deviation and variance of all full-grown corn plants in Elephanteyetown, WI.

3. Consider the following set of data, which represents a simple random sample of heights (in inches) of full-grown corn plants in Horseyetown, WI: (78.5, 88.8, 76.2, 73.0, 79.5, 66.9, 72.4, 86.6, 67.6, 73.5, 72.6, 72.6, 80.0, 69.1, 79.1, 73.9, 72.3, 83.3, 64.3, 69.1) Round your answers to 2 decimal places. (4 points)

   A. Calculate the range.

   B. Calculate the sample mean.

   C. Calculate sample statistics to estimate the population standard deviation and variance of all full-grown corn plants in Horseyetown, WI.

4. In Goateyeville, MN, they use a new type of fish fertilizer. A random sampling of the heights of 20 full-grown corn plants in Goateyeville yields the following data: (heights are in inches) (41.0, 51.1, 65.8, 57.3, 35.8, 61.5, 62.5, 65.8, 72.9, 62.7, 76.4, 81.9, 77.7, 79.0, 73.2, 52.1, 72.0, 69.7, 76.4, 73.4) Round your answers to 2 decimal places. (4 points)

   A. Calculate the range.

   B. Calculate the sample mean.

   C. Calculate sample statistics to estimate the population standard deviation and variance of all full-grown corn plants in Goateyeville, MN.

5. Draw histograms of the distributions from Questions 2 and 4. (4 points)

A. How would you characterize their shapes?

B. How do their means compare?

C. How do their standard deviations compare?

D. How do skew and range affect the standard deviation; does one affect it more than the other?

6. To the data set in Question 3, add two stunted plants: **4 inches and 3 inches**.
**(The researcher decided to not throw them out as outliers.) For this new data
set (containing 22 observations) calculate each of the following and in each
case comment on the difference between the results from the data set
containing the outliers and the data set without them. Round your answers to
2 decimal places.** (5 points)

   A. The range

   B. The sample mean

   C. Compute the sample statistic you would use to estimate the population standard
   deviation and variance of all full-grown corn plants in Horseyetown, WI.

   D. If you were a researcher and you wanted to describe the variation in the population
   of corn plants in Horseyetown, how would you decide whether to throw out the
   stunted plants as outliers? Describe a scenario in which you'd toss them out of your
   data set and a scenario where you would **not** toss them out. In each case explain
   your reasoning. For the scenario in which you'd keep the stunted plants in your data
   set, would you want to consider also using the median as your measure of center?