In this tutorial, we investigate relationships and differences among the notions of "(overall, or population) mean," "sample means," and "mean of the sample means."

Consider the data set

$$X = \{-93, -42, -33, 51, 81\}.$$

Think of the data points in X as being measurements on some overall population (of size 5) that we're interested in studying.

1. Compute the mean, let's call it μ, and the standard deviation, call it σ, of this data set. Use the formulas for mean and standard deviation that we studied in class. Write your answers to four decimal places.

$\mu =$ -7.2000 $\sigma =$ 71.4227

2. Write down all three-element subsets of X. Hint: there are ten of these subsets; call them $S_1, S_2, \ldots, S_{10}$. The first two have been done for you, to get you started. (You can write down the remaining ones in any order.)

   Remark. Think of each of the subsets $S_1, S_2, \ldots, S_{10}$ as being a size-3 *sample* from the overall population X.

$S_1 = \{-93, -42, -33\}$ $S_2 = \{-93, -42, 51\}$

$S_3 = \{-93, -42, 81\}$ $S_4 = \{-93, -33, 51\}$

$S_5 = \{-93, -33, 81\}$ $S_6 = \{-93, 51, 81\}$

$S_7 = \{-42, -33, 51\}$ $S_8 = \{-42, -33, 81\}$

$S_9 = \{-42, 51, 81\}$ $S_{10} = \{-33, 51, 81\}$

3. Each set S_k has a mean $\overline{x}_k$: that is, S_1 has mean $\overline{x}_1$, S_2 has mean $\overline{x}_2$, and so on. Each $\overline{x}_k$ is called a (size-3) *sample mean* from the overall population X.

   Compute the $\overline{x}_k$'s and write your answers in the spaces below (the first two have been done for you).

$$\overline{x}_1 = \frac{-93-42-33}{3} = -56 \qquad \overline{x}_2 = \frac{-93-42+51}{3} = -28$$

$$\overline{x}_3 = \frac{-93-42+81}{3} = -18 \qquad \overline{x}_4 = \frac{-93-33+51}{3} = -25$$

$$\overline{x}_5 = \frac{-93-33+81}{3} = -15 \qquad \overline{x}_6 = \frac{-93+51+81}{3} = 13$$

$$\overline{x}_7 = \frac{-42-33+51}{3} = -8 \qquad \overline{x}_8 = \frac{-42-33+81}{3} = 2$$

$$\overline{x}_9 = \frac{-42+51+81}{3} = 30 \qquad \overline{x}_{10} = \frac{-33+51+81}{3} = 33$$

4. Let $\overline{X}$ denote the population (that is, the set) of all of the above sample means; that is,

$$\overline{X} = \{\overline{x}_1, \overline{x}_2, \overline{x}_3, \overline{x}_4, \overline{x}_5, \overline{x}_6, \overline{x}_7, \overline{x}_8, \overline{x}_9, \overline{x}_{10}\}.$$

   Compute the mean, let's call it $\overline{\mu}$, and the standard deviation, call it $\overline{\sigma}$, of $\overline{X}$. Again, use the formulas for mean and standard deviation that we studied in class. Write your answers to four decimal places.

$\overline{\mu} =$ -7.2000 $\qquad$ $\overline{\sigma} =$ 27.4906

5. How do μ and $\overline{\mu}$ compare? That is: which is smaller, or are they equal?

   $\mu = \overline{\mu}$.

6. How do σ and $\overline{\sigma}$ compare? That is: which is smaller, or are they equal?

   $\overline{\sigma} < \sigma$.

7. Fill in the blanks to complete this tutorial. Use your answers above to guide you. Each blank should be filled with one of the words/phrases "smaller than," "equal to," "average," "extreme," or "spread."

   Consider a set X of data points corresponding to some overall population; suppose X has mean μ and standard deviation σ.

   Fix a sample size n, and compute the mean of each size-n sample from X. Let $\overline{X}$ denote the set of all of these sample means. Then:

   - The mean $\overline{\mu}$ of $\overline{X}$ is __equal to__ the mean μ of X. This reflects the intuitively plausible fact that "the __average__ of the averages equals the __average__."
   - The standard deviation $\overline{\sigma}$ of $\overline{X}$ is __smaller than__ the standard deviation σ of X. Why should this be true? It's because the process of taking averages tends to mitigate the effect of outliers, or __extreme__ values, in your data. That is: an __extreme__ data value, meaning one that's far away from the mean, can result in a relatively large standard deviation, or __spread__, in your data. But if this __extreme__ value is averaged against other data values, then its impact on the spread in the data will not be as __extreme__, since the other data values involved in the computation will tend to pull things back towards the __average__. Consequently, averaging data will tend to yield numbers with a smaller __spread__, and consequently a smaller standard deviation, than the original data itself. Or, to summarize (and repeat): the standard deviation $\overline{\sigma}$ of the set $\overline{X}$ of size-n sample means from a data set X is __smaller than__ the standard deviation σ of the data set X itself.