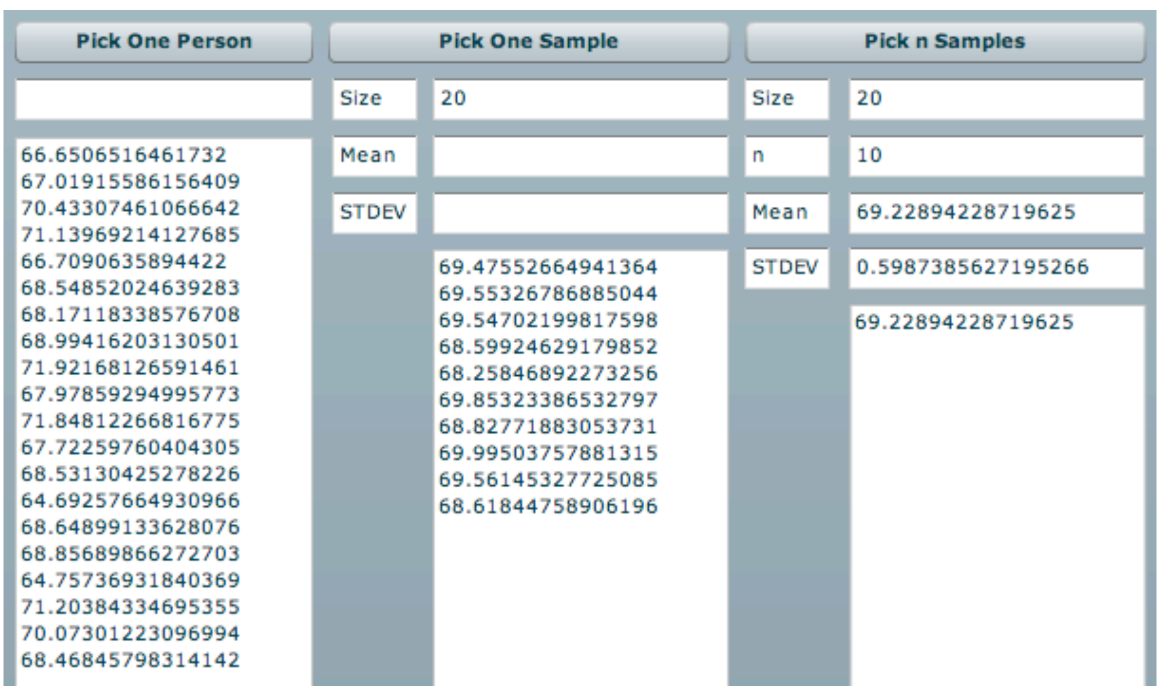
Example – Introduction to Sample Means

Abstract: Many beginning students have difficulty understanding the difference between the population mean of a random variable X and the sample mean

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

of a sample of n independent values of X. This interactive lesson is intended to help students understand the basic ideas before they look at the formal details. The lesson is built around the simulation shown below. This simulation enables students to choose

- a single random value of X by clicking the **Pick One Person** button;
- a random sample of a given size and compute its sample mean and standard deviation by clicking the **Pick One Sample** button; and
- n samples and compute their individual sample means and then the mean of the sample means and the standard deviation of the sample means by clicking the **Pick n Samples** button.



Storyboard: The setting for this lesson is very commonplace. We have a population of people and want to know their (population) mean height. You can get a very rough estimate of the mean height by picking a single person and measuring his height. This is the first experiment we ask students to do. We give them the (population) mean height but emphasize that in practice the population mean is unknown. Indeed, the whole purpose of this example is studying methods of estimating the unknown population mean.

We ask students to obtain ten estimates of the mean height by measuring the heights of ten people chosen at random. We ask them how good these individual estimates are.

Next we ask students to find the mean of their ten estimates. Notice this is a sample mean. The idea is to build on students' intuition that a sample mean is better estimate of the population mean than just a single measurement. Next we ask students to use this basic idea to get a better estimate. Again, most students know intuitively that "larger" samples give "better" estimates. This lesson is intended to be an introduction to the next lesson, which uses the Central Limit Theorem to make the adjectives "larger" and "better" more precise.

The final two experiments ask students to find several sample means using samples of size 40 and then several sample means using samples of size 1000. The simulation automatically computes the mean and standard deviation for these two experiments providing some evidence for how sample size affects how well the sample means approximate the population means.

Details of the Simulation: