

[Illustrative Mathematics](#)

7.SP Estimating the Mean State Area

[Alignment 1: 7.SP.A](#)

The table below gives the areas (in thousands of square miles) for each of the “lower 48” states. This serves as the population for this study. Your task involves taking small samples from this population and using the sample mean to estimate the mean area for the population of states by following the steps indicated below.

Your challenge is to discover some important properties of random samples, properties that illustrate why random sampling is the key to getting good statistical information about a population. In this task, unlike “real life” situations, you will have all of the population data at hand, and will use it to see how random sampling works. Your classmates will have the same task, and you will be combining your data with theirs. In the first part of the task, you will select a sample of states by a method of your choice. For the second part you will follow a specified procedure.

Procedure #1: Choose your own sample

- a. By any quick method you like, select 5 states that you think represent the 48 (perhaps by tossing 5 grains of sand on a map of the states and selecting the states on which they fall; shutting your eyes and pointing your finger at a spot on the map, repeating the process until 5 states are selected; systematically selecting a state from the northeast, the south, the mid-west, all east of the Mississippi, and two states from west of the Mississippi).
- b. Find the areas of these 5 states and calculate the mean for your sample.
- c. As a class, construct a dot plot of the sample means.

Procedure #2: Use random sampling

- d. Number the 48 states from 1 to 48. Then use a random number table or a random number generator to obtain 5 random numbers between 1 and 48, and then find the states corresponding to these numbers.
- e. Find the areas of these 5 states and calculate the mean for your random sample.
- f. As a class, construct a dot plot of the sample means from the random samples.
- g. Compare the plots produced in steps c and f. Where are the centers? Which has greater spread?
- h. Repeat steps 1-6 for random samples of size 10 and compare the plots. What differences, if any, do you see in the plots? What feature or features appear to stay the same?
- i. Find the actual mean state area using the data from all the states. Summarize at least two important points concerning the value of random sampling.

State	Area
Texas	269
California	164
Montana	147
New Mexico	122
Arizona	114
Nevada	111
Colorado	104
Oregon	98
Wyoming	98
Michigan	97
Minnesota	87
Utah	85
Idaho	84
Kansas	82
Nebraska	77
South Dakota	77
Washington	71
North Dakota	71
Oklahoma	70
Missouri	70
Florida	66
Wisconsin	65
Georgia	59
Illinois	58
Iowa	56
New York	55
North Carolina	54
Arkansas	53
Alabama	52
Louisiana	52
Mississippi	48
Pennsylvania	46
Ohio	45
Virginia	43
Tennessee	42
Kentucky	40
Indiana	36
Maine	35
South Carolina	32
West Virginia	24
Maryland	12
Massachusetts	11
Vermont	10
New Hampshire	9
New Jersey	9
Connecticut	6
Delaware	2

Commentary:

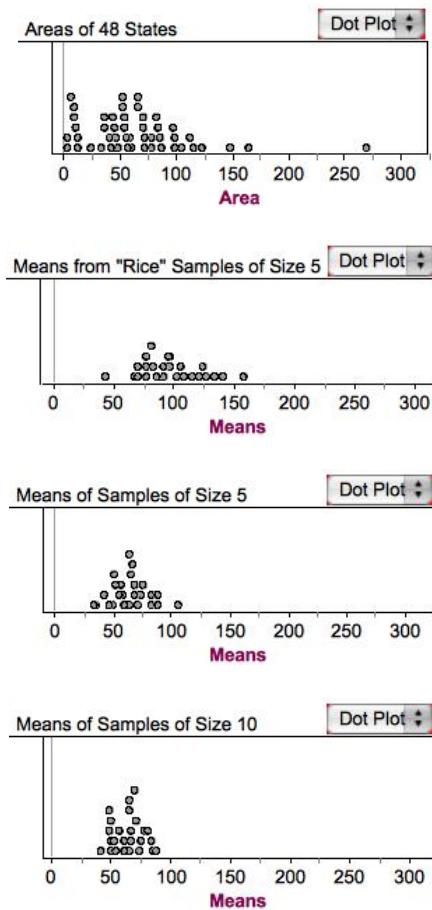
The task is designed to show that random samples produce distributions of sample means that center at the population mean, and that the variation in the sample means will decrease noticeably as the sample size increases. Random sampling (like mixing names in a hat and drawing out a sample) is not a new idea to most students, although the terminology is likely to be new. Most students readily grasp this as a “fair” way to select the sample because each item in the population gets an equal chance of being selected. Standard 1 uses the term “representative,” which has no technical definition in statistics but might now be interpreted as “unbiased” in the sense that the distribution of sample means centers right where you want it to □ at the population mean.

Solution: Possible Solution

The plots displayed below show the areas of the population of 48 states, followed by three sets of 25 sample means each collected, first, through a process similar to the sand throwing idea given above, second, by random samples of size 5 and, third, by random samples of size 10. The population mean is 65 and the three means of the distributions of sample means are 97.5, 63.4 and 64.1, respectively. The two random samples center at the population mean, but the “sand” sample means do not. They are “biased” by the fact that a grain of sand has a greater chance of landing on a large state than on a small one.

The “sand” sample means also have larger variation than either or the random sampling methods. The random sampling means for samples of size 5 have much less variability than the population; the random sampling means for samples of size 10 have, in turn, less variability than do those for samples of size 5. (In fact, the variability of sample means scales down by a factor of $\frac{1}{\sqrt{n}}$, where n denotes the sample size. Samples of size 5 have less than half the variability of the population and samples of size 10 have less than a third of the variability of the population.)

In summary, random sampling produces sample means whose distributions center at the population mean and have variation that decreases as the sample size increases.





[7.SP Estimating the Mean State Area](#) is licensed by [Illustrative Mathematics](#) under a [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License](#)