

BASIC SIMULATION*

David Lane

This work is produced by The Connexions Project and licensed under the Creative Commons Attribution License †

1 General Instructions

This simulation illustrates the concept of a sampling distribution.

Depicted on the top graph is the population from which we are going to sample. There are 33 different values in the population: the integers from 0 to 32 (inclusive). You can think of the population as consisting of having an extremely large number of balls with each 0's, an extremely large number with 1's, etc. on them. The height of the distribution shows the relative number of balls of each number. There is an equal number of balls for each number, so the distribution is a rectangle.

The second graph shows the sampling processes as it might happen in the physical world. After you push the "animated sampling" button, five balls are selected and are plotted on the second graph. The mean of this sample of five is then computed and plotted on the third graph. If you push the "animated sampling" button again, another sample of five will be taken, and again plotted on the second graph. The mean will be computed and plotted on the third graph. This third graph is labeled "Distribution of Sample Means, $N = 5$ " because each value plotted is a sample mean based on a sample of five. At this point, you should have two means plotted in this graph.

The mean is depicted graphically on the distributions themselves by a blue vertical bar below the X-axis. For Graphs 1 and 3, a red line starts from this mean value and extends one standard deviation in length in both directions. The values of both the mean and the standard deviation are given to the left of the graph. Notice that the numeric form of a property matches its graphical form.

The sampling distribution of a statistic is the relative frequency distribution of that statistic that is approached as the number of samples (not the sample size!) approaches infinity. To approximate a sampling distribution, click the "5,000 samples" button several times. The bottom graph is then a relative frequency distribution of the thousands of means. It is not truly a sampling distribution because it is based on a finite number of samples. Nonetheless, it is a very good approximation.

The simulation has been explained in terms of the sampling distribution of the mean for $N = 5$. All statistics, not just the mean, have sampling distributions. Moreover, there is a different sampling distribution for each value of N . For the sake of simplicity, this simulation only uses $N = 5$. Finally, the default is to sample from a distribution for which each value has an equal chance of occurring. Other shapes of the distribution are possible. In this simulation, you can make the population normally distributed as well.

In this simulation, you can specify a sample statistic (the default is mean) and then sample a sufficiently large number of samples until the sampling distribution stabilizes. Make sure you understand the difference between the sample size (which here is 5) and the number of samples included in a distribution. You should also compare the value of a statistic in the population and the mean of the sampling distribution of that statistic. For some statistics, the mean of the sampling distribution will be very close to the corresponding

*Version 1.4: Jul 14, 2003 4:29 pm -0500

†<http://creativecommons.org/licenses/by/1.0>

population parameter; for at least one, there will be a large difference. Also note how the overall shape of sampling distribution differs from that of the population.

2 Step by Step Instructions

Show Questions¹

1. With the default setting, (uniform population, sample statistic set to mean), click the "Animated Sample" a couple time. Notice how the sample means from each random sample accumulate in the bottom graph to gradually form a distribution. Then click "5 samples" and "500 samples" a couple times.
2. Click "10,000 samples" a couple of times until the total number of samples exceeds 50,000 and the sampling distribution stabilizes. Notice its shape and compare it with the population. Compare the mean of the sampling distribution with the mean of the population.
3. Select "Median" as the sample statistic. Draw 50,000 samples and take note of resulting distribution. Compare the mean of the sampling distribution with the median of the population.
4. Select "Range" as the sample statistic. Draw 50,000 samples and take note of resulting distribution. Compare the mean of the sampling distribution with the range of the population.
5. Select "Variance" as the sample statistic. Draw 50,000 samples and take note of resulting distribution. Compare the mean of the sampling distribution with the variance of the population.
6. Set the population to be "Normal". Repeat steps 2-5.

This is a Java Applet. To view, please see <http://cnx.org/content/m11204/latest/>

3 Summary

The distribution of a sample statistic (mean, median, etc.) from an infinite number of samples is called a "sampling distribution". In this simulation, the sampling distribution is approximated by including a sufficiently large number of samples in the distribution. Each sample, on the other hand, consists of a fixed number of data points from the population (called "sample size"), and in turn, contributes only one data point to the sampling distribution. The standard deviation of a sampling distribution is called the "standard error", as opposed to the standard deviation in the population.

¹"Basic Demo" <<http://cnx.org/content/m11203/latest/>>