Sampling Words

Introduction:
Random sampling is one of the most important concepts in statistics. Independent random samples from a population are the basis for most statistical theory. This activity leads you through a series of exercises dealing with sampling words from the Gettysburg address.

To Begin:
1. Access the Sampling Words Applet

http://www.rossmanchance.com/applets/OneSample.html?population=gettysburg

2. Click the box next to "Show Sampling Options" on the right hand side of the page. You should now see three plots.
3. The plot on the far left shows the distribution of the original population.

Other two plots will be empty. The one in the middle will show the sample values for the most recent sample, once we get to that part. The plot on the right will show the distribution of a sampling statistic (the mean, median, or t-statistic) when we get to that point.

Exercises (not required):
We are not actually sampling words; we are sampling word lengths. To get familiar with the applet, do the following exercises:

1. Choose sample size = 5 and number of samples = 1. Click on the "draw samples" button.

Write down the words and their corresponding lengths that you sampled. Find the mean of the word lengths. To describe variability, we will use the range, which is the largest value (the maximum word length in this case) minus the smallest value (the minimum word length, in this case).

Mean from one sample of size 5:

Range of one sample of size 5:

What would be an "unusual" value of word lengths for a sample of size 5?

2. Click the "reset" button. Now set the number of samples to 5. Click "draw samples". Write down the sample MEANS of each of the five samples. To obtain the mean of each of the five samples, move your mouse over each of the blue rectangles on plot at the bottom right.

What is the mean of the means from samples of size 5?

What is the range of the means from samples of size 5?

What would be an unusual value for a mean from a sample of size 5?

Simulation (required):
1. Now, take 500 random samples, each of size 5. What is the mean of the 500 means?
2. What is the range of the 500 sample means(smallest value and largest value)?
3. What is the standard deviation of the 500 sample means? You can find this value in the upper right corner of the graph on the lower right.

4. Describe the distribution of 500 sample means from samples of size 5 (think shape, center, and spread).
5. Run the following simulations using the applet. Click "reset" before each simulation!! For each simulation, obtain the mean and standard deviation of the 500 sample means.
    a. Number of samples = 500, sample size = 10
    b. Number of samples = 500, sample size = 20
    c. Number of samples = 500, samples size = 50
6. For each of a, b, and c, above, describe the shape of the distribution of the 500 sample means relative to the original population distribution.
7. For each of a,b, and c, calculate sigma divided by the square root of n and compare this to the standard deviation of the distribution of the sample means from your simulation.

For Discussion in Live Session / Please make a slide or slides for you answer in the designated section of the For LIve Session Assignment (FSL):

1. How do the distributions of simulated samples of the same sample size compare to one another (think center, shape, and spread). These are the samples that you see in the middle histogram.
2. Describe the distributions of the sample means for samples of size 5, 10, 20, and 50. How do these distributions compare to one another, and to the population distribution?
3. What would you expect to happen to the distribution of the sample means if we changed the sample size again, to a larger number?
4. What is the pattern that you see?
5. Do you think the same pattern would result if we took samples and made distributions of sampling statistics for other types of populations (not word lengths)? Try it with "HasE" and "Is a Noun". What do you see?
6. Do you think the same pattern would result if we took samples and made distributions of other sampling statistics (not the mean)? Try it by punching the radio buttons for the median and standard deviation over the plot at the bottom right.