Activity #1: Correlation, Regression, and Outliers

1. Choose the *Correlation and Regression* applet at www.stapplet.com/spa.

2. Create a scatterplot with about 10 points and a correlation around $r = 0.50$ in the lower-left, as shown here. Check the boxes to show the least-squares line and the mean x and y lines.

3. Imagine putting an additional point on the least-squares line in the upper-right corner. How do you think this point will affect the correlation? The least-squares line?

4. Now, add the point to see if your answer was correct.

5. What do you think would happen if you dragged the outlier to the lower-right corner? Don't do it yet!

6. Now, move the outlier to the lower-right and see if you were correct.

7. Move the outlier to the $(\bar{x}, \bar{y})$ point. What do you think will happen if you drag this point straight up along the $\bar{x}$ line? Don't do it yet!

8. Now, drag the point up the $\bar{x}$ line to see if you were correct.

9. Summarize the effects of an outlier on the correlation and the least-squares line.

Activity #2: Guess the correlation—March Madness style!

Guess the correlation applet available at www.rossmanchance.com/applets.

Also check out their new Guess the *P*-value applet!

Activity #3: The Sampling Distribution of a Sample Mean $\bar{x}$

1. Go to the following website: onlinestatbook.com/rvls.html. Choose: Simulations/Demonstrations then Sampling Distribution Simulation and press Begin.

2. The display will show 4 graphs. The first graph shows the population. Currently, the population should be Normal. The second graph is called "Sample Data." This will show the observations that are sampled from the population in the top graph. On the third and fourth graph, there are choices for sample statistics and sample sizes. Set graph 3 to "mean" with "n = 5" and set graph 4 to "none." Then click "Animated Sample" a few times from graph 2 and clearly describe what is happening.

3. Now, click on "100,000 Samples" and clearly describe what is happening.

4. Now, in graph 3, change the sample size to "n = 2" and generate 100,000 samples. Repeat this for each of the sample size choices (clear the bottom 3 graphs each time!) and clearly describe what happens to the distribution's shape, center, and spread as the sample size increases.

5. Now, change the population from Normal to Uniform and repeat step #4.

6. Now, change the population from Uniform to Skewed and repeat step #4.

7. Now, change the population from Skewed to "Custom" and paint a bimodal (double-peaked) distribution. Repeat step #4.

8. Summarize your responses to questions 4–7. Do you see anything in common? Under what conditions will the sampling distribution of the sample mean be approximately Normal?

To do a similar activity for the sampling distribution of the sample proportion $\hat{p}$, see the Reese's Pieces applet at www.rossmanchance.com/applets.

Activity #4: Confidence Interval BINGO!

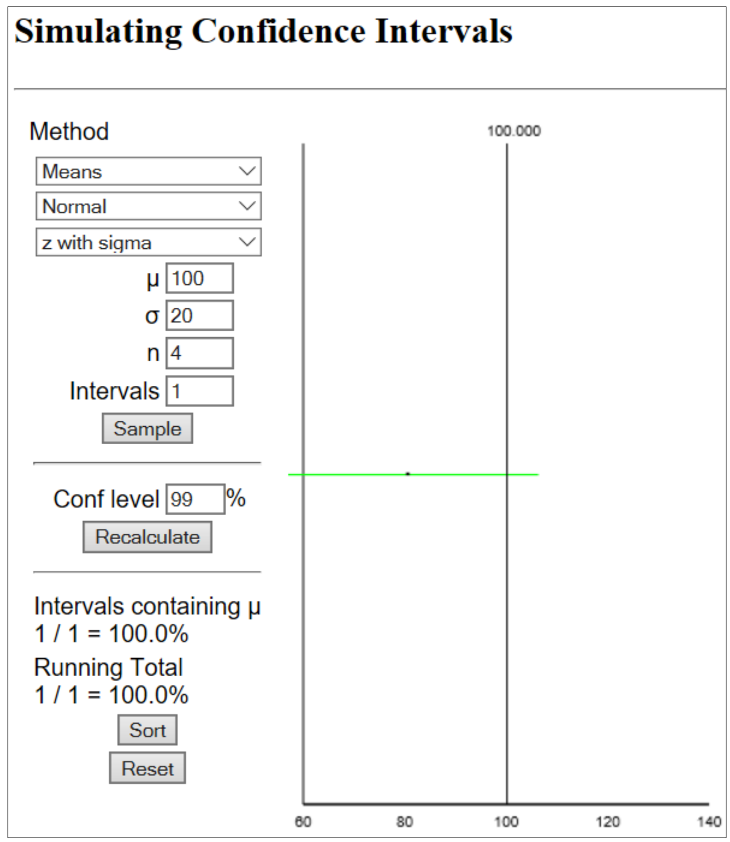
A farmer wants to estimate the mean weight (in grams) of all tomatoes grown on his farm. To do so, he will select a random sample of 4 tomatoes, calculate the mean weight (in grams), and use the sample mean $\bar{x}$ to create a 99% confidence interval for the population mean μ. Suppose that the weights of all tomatoes on his farm are approximately normally distributed with a mean of 100 grams and a standard deviation of 20 grams.

Let's use an applet to simulate taking an SRS of n = 4 tomatoes and calculating a 99% confidence interval for μ using three different methods.

Method 1 (assuming σ is known):

Formula:

1. Launch the *Simulating Confidence Intervals for a Population Parameter* applet at www.rossmanchance.com/applets. Choose "Means" from the drop-down menu and leave the others menus as "Normal" and "z with sigma." Then enter μ = 100, σ = 20, n = 4 and confidence level = 99% as shown in the screen shot.



2. Press "Sample." The applet will select an SRS of n = 4 tomatoes and calculate a confidence interval for μ. This interval will be displayed as a horizontal line segment, along with a vertical line at μ = 100. If the interval captures μ = 100, the interval will be green. If the interval misses μ = 100, the interval will be red.

3. Press "Sample" many times, shouting out "BINGO!" whenever you get an interval that misses μ = 100 (i.e., a red interval).

4. How well did Method 1 work? Compare the running total in the lower-left corner with the stated confidence level of 99%.

Method 2 (using s_x as an estimate for σ):

Formula:

1. Press the Reset button in the lower-left. Then, in the third drop-down menu, choose "z with s." Keep everything else the same.

2. Press "Sample" many times, shouting out "BINGO!" whenever you get an interval that misses μ = 100 (i.e., a red interval).

3. Did Method 2 work as well as Method 1? How do you know?

4. Now, change the number of intervals to 50 and press "Sample" 10 times, for a total of more than 1000 intervals. Compare the running total in the lower-left corner with the stated confidence level of 99%.

5. What do you notice about the length of the intervals that missed?

To increase the capture rate of the intervals to 99%, we need to make the intervals longer. We can do this by using a different critical value, called a t^* critical value. You'll learn how to calculate this number soon.

Method 3 (using s_x as an estimate for σ and a t^* critical value instead of a z^*critical value):

Formula:

1. Press the Reset button in the lower-left. Change "Intervals" to 1 and the third drop-down menu to "t." Keep everything else the same.

2. Press "Sample" many times, shouting out "BINGO!" whenever you get an interval that misses μ = 100 (i.e., a red interval).

3. Did Method 3 work better than Method 2? How does it compare to Method 1?

4. Now, change the number of intervals to 50 and press "Sample" 10 times, for a total of at least 1000 intervals. Compare the running total in the lower-left corner with the stated confidence level of 99%.

5. What do you notice about the length of the intervals compared to Method 2?

6. When should we use a t^* critical value rather than a z^* critical value for calculating a CI for a population mean?