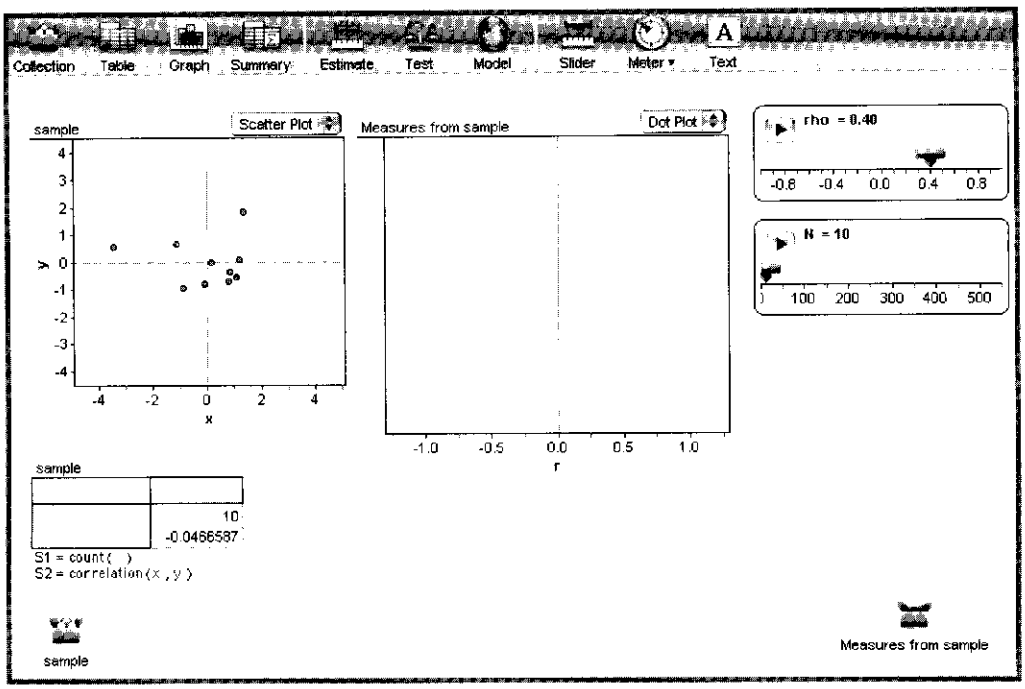
Demo 9: Correlation Coefficients of Samples

How samples from a correlated population yield different values for the correlation •
How sample size affects that sampling distribution

Note: This activity was inspired by *Activity-Based Statistics,* "Relating to Correlation" (first-edition Instructor Resources, page 330).



What To Do

▹ Open **Correlation in Samples.ftm.** It will look something like the illustration.

In the upper left, you see a graph of **N** points (here, **N** is 10, but in general, you control **N** with the slider at the right) drawn from a population of points that have a correlation coefficient of **rho** (here, 0.40; see the slider at right again). Of course, this sample will probably not have the population correlation coefficient 0.40. The *sample's* correlation (here, –0.047) and the number of points in the sample appear in the summary table below the graph.

We'll get to the large, blank, central graph and its collection (**Measures from sample,** lower right) later.

We begin by exploring.

▹ With nothing selected, choose **Rerandomize** from the **Collection** menu to draw a new sample. See how the points change. Rerandomize repeatedly; get a feel for the variation in the sample correlation. Do you ever see a *negative* correlation?

⇨ The shortcut for **Rerandomize** is **⌘+Y** on the Mac or **Control+Y** in Windows.

▹ Play with the **N** slider. Move it to about 200 points. Rerandomize some more to see how that correlation changes and how the impression of correlation is more stable with a larger sample. Do you ever see a negative correlation?

▹ Play with the **rho** slider to see how different correlations look with 200 points.

▹ Then do the same with a smaller sample.

Now let's collect data on repeated samples. If we leave the settings constant and rerandomize, what will the distribution of correlation coefficients **rho** look like?

- ▷ Set the sliders at **N = 30** and **rho = 0.4.**
  - ⇨ You can type these numbers into the sliders to set them exactly.
- ▷ Click once on the **Measures from sample** collection (the box at lower right) to select it.
- ▷ Choose **Collect More Measures** from the **Collection** menu. Fathom will resample 40 times and display the correlation coefficients in the large, formerly empty middle graph.
- ▷ Repeat this procedure many times, for different combinations of **rho** and **N.** Be sure to use large and small samples, and correlations near 0, +1, and −1.

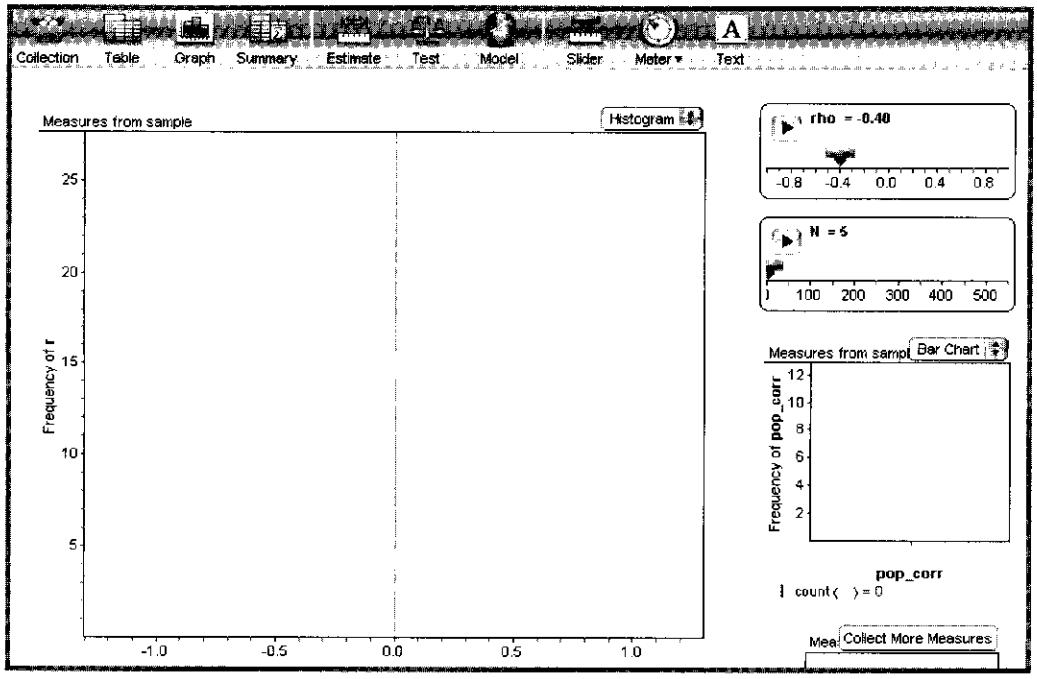
Questions

1. If you leave the correlation coefficient **rho** the same, what effect does **N** have on the center and spread of the distribution?
2. If you leave the sample size **N** the same, what effect does the correlation coefficient **rho** have on the center and spread of the distribution?
3. With a sample size of 10 and a true correlation of 0.5, about how well do you think you could know that true correlation based on a single sample?
4. With a sample size of 500 an a true correlation of 0.5, about how well do you think you could know that true correlation based on a single sample?

Extension

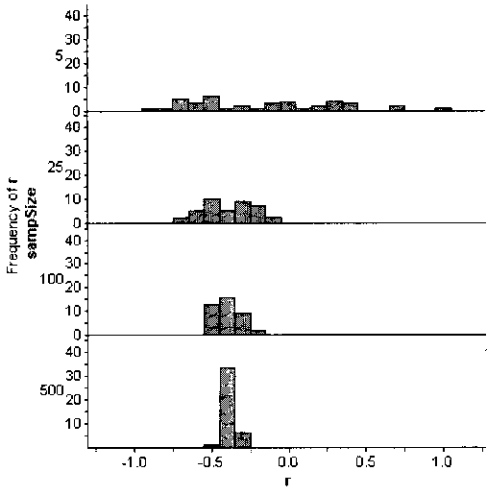
Let's see a lot of these graphs side by side.

- ▷ Open **Correlation in Samples 2.ftm.** It will look something like this:



It looks practically empty, but really it's pretty much the same as **Correlation in samples.ftm**, just rearranged and slightly enhanced. You can see that we begin with **rho = –0.40** and **N = 5**. There is also a new bar graph (currently empty) of **pop_corr**.

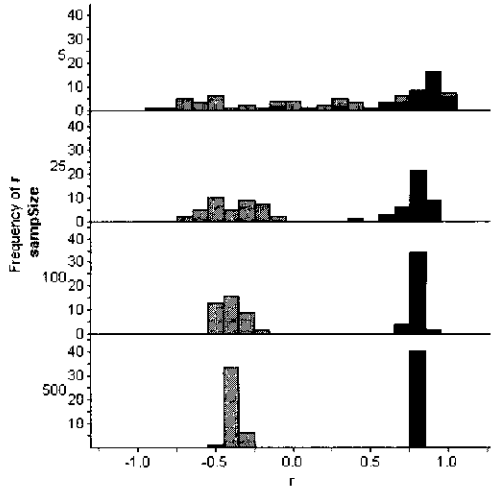
- ▷ Press the **Collect More Measures** button on the collection in the lower right. The graph gets some data as before.
- ▷ Change **N** to **25**. *Do so by typing in the slider, not dragging;* simply edit the number and press **Enter** when you're done.
  - ⇨ Try to predict what you will see.
- ▷ Press **Collect More Measures** again. Instead of replacing the data, Fathom adds the data to the graph. It will look like the illustration below.
- ▷ Do the same for **N = 100** and **N = 500**. The graph now has four panes, one for each sample size.



At this point, it should be clear from the graph that if you increase the sample size, you get a tighter correlation: You'll have a better idea of the population correlation based on a single sample.

If the histogram bars are a little chunky, adjust their widths, and maybe rescale the vertical axis. You can do this by dragging *carefully* or by double-clicking on the graph and typing into the inspector you get. A bin width of 0.05 is fine.

- ▷ Now, leave **N** at 500 and set **rho** at +0.80. Press **Collect More Measures.** Fathom adds the points to the graph; now it has two spikes in the bottom pane.
- ▷ Work your way back down the sample sizes, leaving **rho** at +0.80, but collecting measures at **N = 100**, **N = 25**, and **N = 5**. Be sure to type in the numbers—if you just drag the sliders, you may be off by a bit.
- ▷ Now the final step. In the bar chart—the small graph—click on the *bar* for **pop_corr = 0.8.** This selects all of the data that we took at that higher correlation, turning them all red in the big graph. You should see a graph like the one in the illustration.



Do you see how the distributions in the **N = 5** case overlap? That means that if you see a sample correlation of, say +0.3, the population correlation could be anything from –0.4 to +0.8.

Another Question

5 The spreads for **pop_corr = +0.8** are smaller than the ones for **pop_corr = –0.4.** Why is that? **Sol**