Demo 40: On the Equivalence of Tests and Estimates

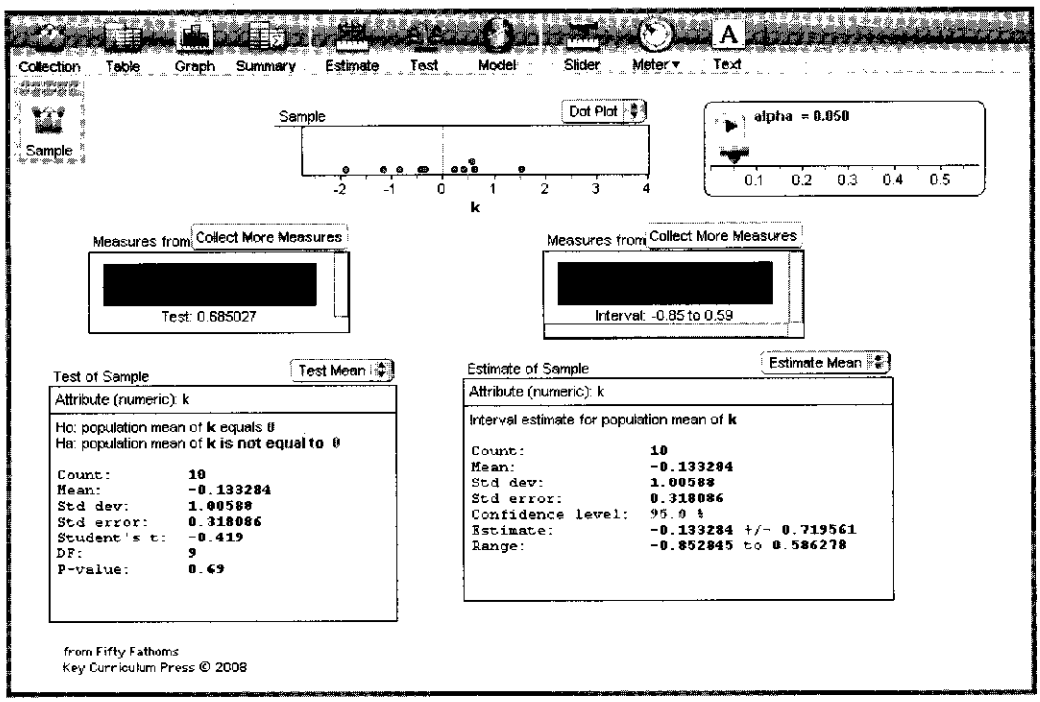
How a hypothesis test and a confidence interval are really the same

In ancient times, statistical tables gave only a critical value—the value for a statistic at the edge of significance for some significance level, usually 5%. If your t or χ^2 was bigger than the critical value, your result was significant. When you tabulated your research results in a journal, you placed an asterisk next to each one that was significant.

Unfortunately, the reader had no way of knowing *how* significant it was (unless you put in a lot of asterisks).

Nowadays, we can calculate the actual P-values easily, or give confidence intervals; that way, readers can make up their own minds whether the result is significant to their satisfaction.

But which is better—the confidence interval or the P-value? This demo shows that, in one sense at least, they are equivalent. Suppose you're testing whether the mean of a population is zero. You construct 95% confidence intervals of the mean. You also perform a t-test and set the significance level—the critical P-value α—at 0.05. Then a significant result—$P < \alpha$—occurs whenever the confidence interval does not contain zero, and vice versa.



What To Do

▷ Open **Tests and CIs.ftm.** It will look something like the illustration.

At the center top, you see our data, called **Sample**, displayed in a dot plot. Below the data are two black bars. Below them are, on the left, a hypothesis test that the mean is not equal to zero, and on the right, an estimate showing the confidence interval for the mean. At the upper right, a slider controls the significance level **alpha.**

Now to the black bars: They show the test (left) and estimate (right) results. The size of the bars doesn't matter. But they change to red if the test shows significance, either because $P < \alpha$ or because the confidence interval no longer contains zero.

▷ Drag points in the graph to make it clear that the mean of the population is not zero. As you pass the critical values, note that the two bars change color at the same time. Repeat as necessary.

▷ Move the points so that the rectangles just barely turn from red to black.

- ▷ Now drag the **alpha** slider so that the rectangles turn red again. (If, while you are dragging **alpha,** one of the black boxes turns into a giant squashed gold ball, don't be alarmed.)
- ▷ Play with both effects as long as you like!

Questions

1 Which direction do you have to move the **alpha** slider to make black rectangles turn red? Why does that make sense?

2 Which way do you have to move to make red rectangles turn black?

3 As you move the **alpha** slider, the number under the left-hand box does not change. Why not?

4 Why does the right-hand rectangle go nuts (for example, turn into a fat gold ball) sometimes? **Sol**

5 Why doesn't the left-hand rectangle go nuts at the same time? **Sol**

6 What happens when the sample mean is zero? (Note: You will probably have to make a case table to get it to be exactly zero.)

Challenges

7 Based on what confidence intervals, the null hypothesis, and hypothesis tests *mean,* explain why it makes sense that this demo should work.

8 Make a case that one or the other inferential tool—the hypothesis test or the confidence interval—is better, even though they are equivalent in this sense of this demo.