Demo 28: The Confidence Interval of a Proportion

Defining the confidence interval • Looking at sample results in terms of plausibility

This demo is about the meaning of a confidence interval (CI). The CI is a difficult concept; it's easy to get confused. We can talk about how the CI measures confidence in a *process*—the process of constructing the interval—and that the confidence level is the probability that if the process were to be repeated, such-and-such a percentage of the intervals so constructed would contain the true population value. While that's accurate, it sounds circular and is mired in subjunctive, subtle language.

I prefer this definition, which is the basis of our demo:

> **The confidence interval is the range of possible parameter values for which the observed value is *plausible*.**

What do we mean by "plausible"? Let's look at the rest of the statement in context. Suppose we have poll results that say 56% of people support Measure Q. This poll is of a particular number of people N, the sample size. We assume that it's a simple random sample and that N is much smaller than the population.

The population parameter we're interested in is p, the proportion of people who favor Measure Q. (In other contexts, we might be interested in the *mean* height of students or the *median* income of dock workers. But this is the confidence interval of a *proportion;* the logic will be the same.) Suppose we polled 50 people and 28 of them said they favored Q. That gives us the sample proportion 0.56—and is where the 56% came from. We call the sample proportion $\hat{p}$, pronounced *p-hat*. We're interested in the population parameter p, *but we will never know its value.*

Instead, we ask, "Would we be surprised if the true population parameter were 54%?" Certainly not. It would be easy to get 28 positive responses out of 50 if the population had 54% in favor. But would it be plausible for only 10% of the population to be in favor? No. The chances of getting 28 out of 50 are too small. Traditionally, the border between plausible and implausible is a 5% probability, though you can change that depending on your situation. All that's left is to find that range of values for which the population parameter is plausible.

That's a long introduction; let's see what it means:

What To Do

▹ Open **CI of a Proportion.ftm.** It will look something like the illustration.

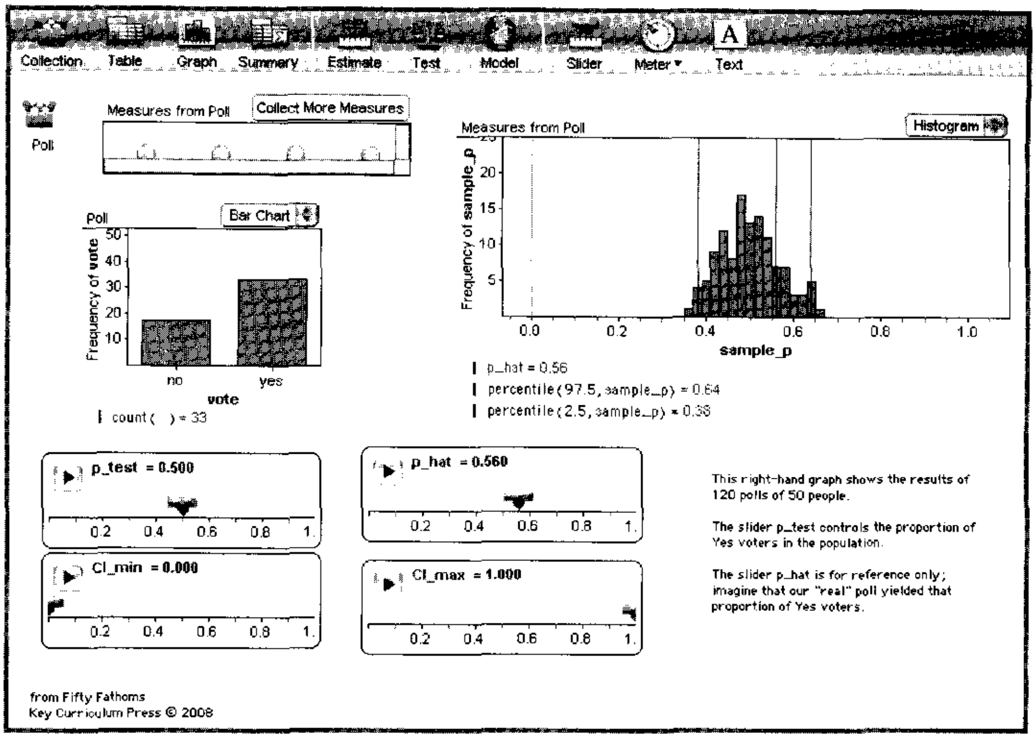
This document simulates many repeated polls. The **Poll** collection (upper left) contains 50 cases, **yes** or **no**, randomly controlled by the **p_test** slider. That is, **p_test** is the probability that a respondent in **Poll** will vote **yes.** The bar graph shows the results of *one* of those polls.

The **Measures from Poll** collection and graph show the results of 120 simulated polls of 50 people. In the graph, the two "outside" vertical lines show the 2.5% and 97.5% percentiles for the polls in the graph; that is, 95% of the polls are between the two lines. We will call all of those results *plausible.*

The **p_hat** slider controls a line on the graph; it represents the results of our real-world poll. So the graph in the illustration, with a population proportion of **p_test = 0.50**, shows that our poll, **p_hat,** is a plausible result.

Note that the histogram of **sample_p**—the simulated poll results—is very much like the graph in Demo 20, "The Distribution of Sample Proportions."

▹ Drag the slider called **p_test.** The graph will *slowly* update (it has to simulate 120 polls of 50 people every time). Set it to 0.3. For that value, you



can see that our **p_hat,** 0.56, is outside the bounds of plausibility. If 30% of the true population favored Measure Q, it would be really unlikely that we would get a 56% result in a poll of 50 people.

▷ Experiment with different values of **p_test** to find values so that the upper and lower "boundary" lines on the graph are coincident with the **p_hat** line. (In fact, since Fathom draws the **p_hat** line first, these "percentile" lines will hide the red **p_hat** line when they match exactly.)

⇨ If you set **p_test** to unusual values (such as −1), the histogram may look strange. Just choose **Histogram** again from the menu in the corner of the graph.

Question

1 Why do you suppose the "percentile" values below the graph seem to be mostly multiples of 2%? That is, you might get 0.56 or 0.58, but seldom 0.57, and never 0.572.

Onward!

▷ When you find a proportion that seems the minimum **p_test** for which **p_hat** is plausible, set the slider **CI_min** to that value. When you find the maximum value, set **CI_max.**

▷ Let's draw the theoretical distribution functions for population proportions at those two limits. Select the graph and choose **Plot Function** from the **Graph** menu. The formula editor opens.

▷ Enter **binomialProbability(round(50x), 50,CI_max)*120.** Press **OK** to plot the function and close the editor.

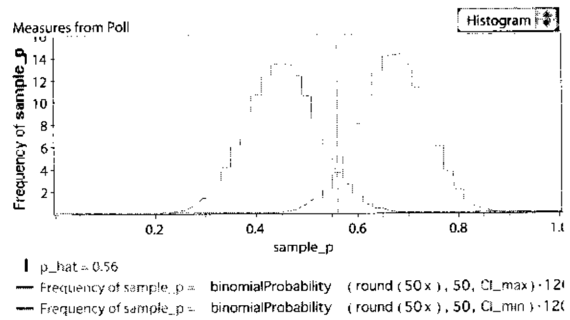
⇨ For help with this function, see "Plotting Binomial Probability—and Other Discrete Distributions" in Appendix A.

▷ Choose **Plot Function** again; this time enter **binomialProbability(round(50x), 50,CI_min)*120** (that is, the same except for "min" instead of "max"). Press **OK** to close the editor.

▷ With the data on the graph, it's pretty busy. Choose **Select All** from the **Edit** menu to select all the cases in the graph; then choose **Delete Cases** from the **Edit** menu to get rid of them.

- ▹ We also have those leftover vertical lines. Select the first one's formula at the bottom of the graph (it starts **percentile(97.5,...)** and choose **Clear Formula** from the **Edit** menu. The line evaporates. Do the same to the other **percentile** line.

Now you have only the functions and **p_hat** left. The graph will look like this:



Note: Your **p_hat** line may not cut the curves exactly at the intersection.

Now our display shows us how the distributions of the minimum and maximum of the interval relate to **p_hat.** The **p_hat** line cuts off two tails of the distributions; each has an area of approximately 2.5%. Those tails represent the chance that we are wrong, that our interval fails to enclose the true population proportion.

If you want less risk of making that mistake, move the **CI_min** and **CI_max** sliders to make the tail bits smaller. Then the peaks will be farther apart, demonstrating in a different way why confidence intervals get wider—less precise—if you increase the confidence level.

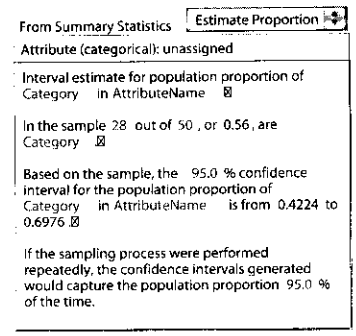
- ▹ Move **p_hat** to a different value, then move the **CI_min** and **CI_max** sliders to cut off about the right amount of tails.

Another Question

2 What happens when you set **p_hat** to a very large or small number, like 0.96 or 0.02? What values of **CI_max** and **CI_min** are appropriate?

Extensions

- ▹ Test your interval by using a traditional confidence interval: drag a new estimate from the shelf. (It looks like a ruler.)
- ▹ Choose **Estimate Proportion** from the pop-up menu in the estimate itself.



- ▹ Enter the data directly into the estimate (do not drag an attribute there). For example, make it look like the illustration.
- ▹ Compare your "plausibility interval" values to the ones you see in the estimate.
- ▹ Change the confidence level to 90% and redo both the plausibility interval (as above) and the one in the estimate.

Challenges

3 Explain clearly why this "plausibility interval" is really the same as the traditional confidence interval ("if you were to draw new samples and construct intervals repeatedly . . .") described at the beginning of this demo. **Sol**

4 Explain why, to get a 95% confidence interval, we use the 2.5 and 97.5 percentiles. That makes sense if you have one distribution (the two tails together make 5%), but here we have two distributions with a combined area of 2, not 1.

5 What if **p_hat** is zero or one? Can you still make a "plausibility interval"? Explain what it means, and relate it to the traditional confidence interval and what you get using Fathom's estimate.

