

What Can We Say about a Proportion?

So what can we really say about p ? Of course, I'm not *sure* that my interval catches p . And I don't know its true value, but I can state a probability that I've covered the true value in an interval. Here's a list of things we'd like to be able to say and the reasons we can't say most of them:

1. **"42.0% of all U.S. adults thought the economy was improving."** It would be nice to be able to make absolute statements about population values with certainty, but we just don't have enough information to do that. There's no way to be sure that the population proportion is the same as the sample proportion; in fact, it almost certainly isn't. Observations vary. Another sample would yield a different sample proportion.
2. **"It is *probably* true that 42.0% of all U.S. adults thought the economy was improving."** No. In fact, we can be pretty sure that whatever the true proportion is, it's not exactly 42.0%, so the statement is not true.
3. **"We don't know exactly what proportion of U.S. adults thought the economy was improving, but we know that it's within the interval $42.0\% \pm 2 \times 0.8\%$. That is, it's between 40.4% and 43.6%."** This is getting closer, but we still can't be certain. We can't know for sure that the true proportion is in this interval—or in any particular range.

4. **"We don't know exactly what proportion of U.S. adults thought the economy was improving, but the interval from 40.4% to 43.6% *probably* contains the true proportion."** Close! Now, we've fudged twice—first by giving an interval and second by admitting that we only think the interval "probably" contains the true value.

That last statement is true, but it's a bit wishy-washy. We can tighten it up by quantifying what we mean by "probably." We saw that 95% of the time when we reach out 2 SEs from \hat{p} , we capture p , so we can be 95% confident that this is one of those times. After putting a number on the probability that this interval covers the true proportion, we've given our best guess of where the parameter is and how certain we are that it's within some range.

5. **"We are 95% confident that between 40.4% and 43.6% of U.S. adults thought the economy was improving."** Statements like this are called **confidence intervals**. They don't tell us everything we might want to know, but they're the best we can do.

Above Image taken from: Sharpe, De Veaux, Velleman. *Business Statistics*. Pearson, edition III, 2015. Pp 279.

For Prediction of means:

Cautions about Interpreting Confidence Intervals

Confidence intervals for means offer new, tempting, wrong interpretations. Here are some ways to keep from going astray:

- **Don't say**, "95% of all the policies sold by this sales rep have profits between \$942.48 and \$1935.32." The confidence interval is about the *mean*, not about the measurements of individual policies.
- **Don't say**, "We are 95% confident that a randomly selected policy will have a net profit between \$942.48 and \$1935.32." This false interpretation is also about individual policies rather than about the *mean* of the policies. We are 95% confident that the *mean* profit of all (similar) policies sold by this sales rep is between \$942.48 and \$1935.32.
- **Don't say**, "The mean profit is \$1438.90 95% of the time." That's about means, but still wrong. It implies that the true mean varies, when in fact it is the confidence interval that would have been different had we gotten a different sample.
- **Finally, don't say**, "95% of all samples will have mean profits between \$942.48 and \$1935.32." That statement suggests that *this* interval somehow sets a standard for every other interval. In fact, this interval is no more (or less) likely to be correct than any other. You could say that 95% of all possible samples would produce intervals that contain the true mean profit. (The problem is that because we'll never know what the true mean profit is, we can't know if our sample was one of those 95%.)

So, what *should* you say? Since 95% of random samples yield an interval that captures the true mean, you should say:

- "I am 95% confident that the interval from \$942.48 to \$1935.32 contains the mean profit of all policies sold by this sales representative." It's also okay to make this a little less formal by saying something like:
- "I am 95% confident that the mean profit for all policies sold by this sales rep is between \$942.48 and \$1935.32."

Remember: Your uncertainty is about the interval, not the true mean. The interval varies randomly. The true mean profit is neither variable nor random—just unknown.

Above image taken from: Sharpe, De Veaux, Velleman. *Business Statistics*. Pearson, edition III, 2015. Pp 347-348.