STA 2023

Module 7

Confidence Intervals for Proportions

Learning Objectives

Upon completing this module, you should be able to:

1. Find and interpret a large-sample confidence interval for a population proportion.
2. Compute the margin of error for the estimate of a population proportion.
3. Describe the relationship between the sample size, confidence interval, and margin of error for a confidence interval for a population proportion.
4. Determine the sample size required for a specified confidence level and margin of error for the estimate of a population proportion.

What is Population Proportion?

As we have learned from the last module, a population proportion, *p* is simply the percentage of a population that has a specified attribute. A population proportion is also known as a true proportion.

Population Proportion and Sample Proportion

Population Proportion and Sample Proportion

Consider a population in which each member either has or does not have a specified attribute. Then we use the following notation and terminology.

Population proportion, p: The proportion (percentage) of the entire population that has the specified attribute.

Sample proportion, $\hat{p}$: The proportion (percentage) of a sample from the population that has the specified attribute.

In short, a sample proportion is obtained by dividing the number of members sampled that have the specified attribute (x) by the total number of members sampled (n).

Sometimes, we refer to "x" as the number of successes and "n-x" as the number o failures.

The Sampling Distribution of the Sample Proportion

In order to make inferences about a population proportion, we must know the sampling distribution of the sample proportion.

The Sampling Distribution of the Sample Proportion

For samples of size n,

- the mean of $\hat{p}$ equals the population proportion: $\mu_{\hat{p}} = p$;
- the standard deviation of $\hat{p}$ equals the square root of the product of the population proportion and one minus the population proportion divided by the sample size: $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$; and
- $\hat{p}$ is approximately normally distributed for large n.

In short, if n is large, the possible sample proportions for samples of size n have approximately a normal distribution with *mean p* and *standard deviation* as above.

The Sampling Distribution Model of Sample Proportions

- The sampling distribution model of $\hat{p}$ is centered at *p*, with standard deviation $\sqrt{\frac{pq}{n}}$, where *q* = *1* - *p*.
- Since we don't know *p*, we can't find the true standard deviation (population standard deviation) of the sampling distribution model, so we need to find the standard error: $SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$

Note that $\hat{p}$ (read as p hat) is the sample proportion, whereas ***p*** is the true population proportion (true proportion).

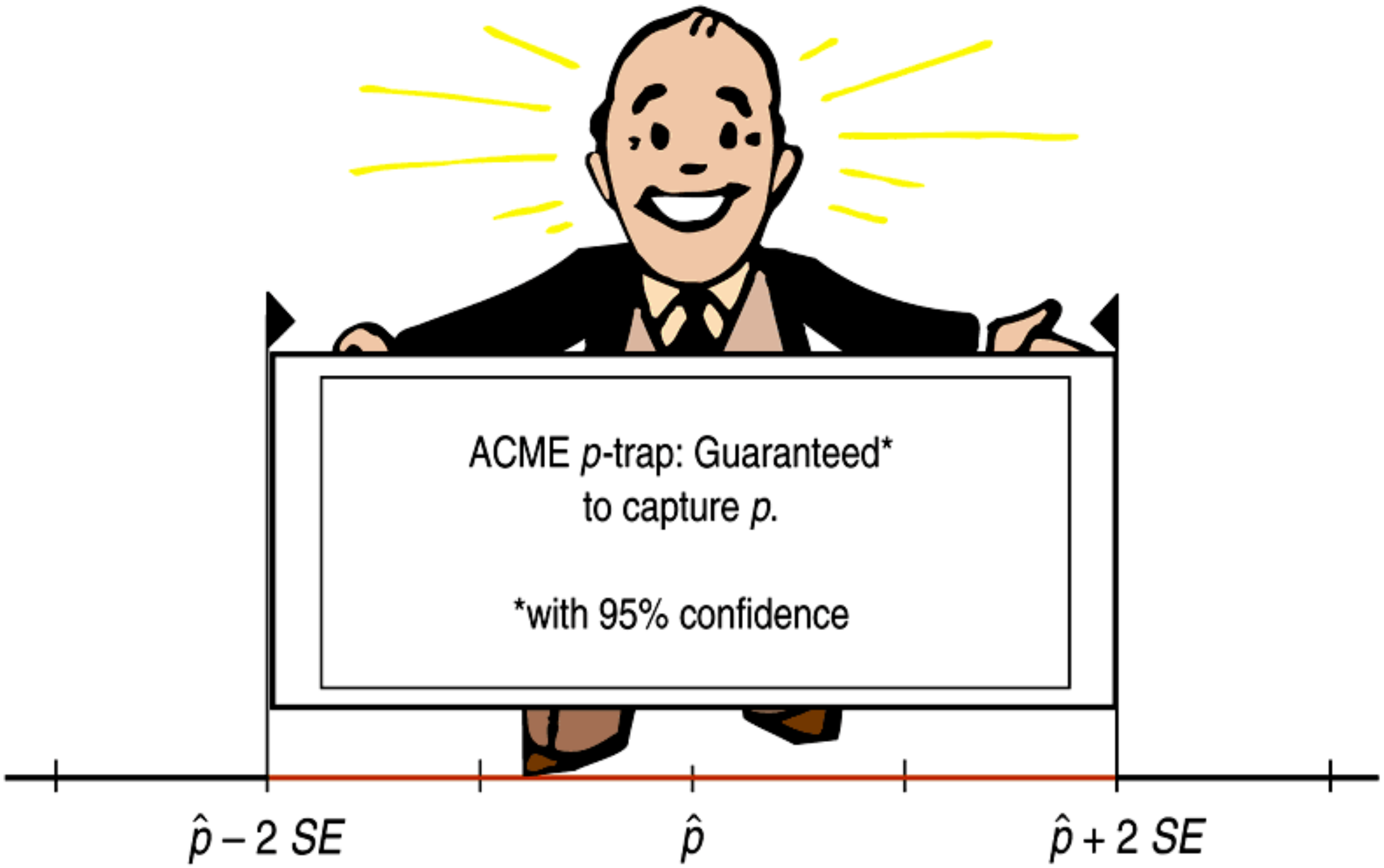
Confidence Interval and 68-95-99.7% Rule

- By the 68-95-99.7% Rule, we know
  - about 68% of all samples will have $\hat{p}$'s within 1 *SE* of *p*
  - about 95% of all samples will have $\hat{p}$'s within 2 *SE*s of *p*
  - about 99.7% of all samples will have $\hat{p}$'s within 3 *SE*s of *p*
- We can look at this from $\hat{p}$'s point of view…

Confidence Interval and 68-95-99.7% Rule (cont.)

- Consider the 95% level:
  – There's a 95% chance that *p* is no more than 2 *SE*s away from $\hat{p}$.
  – So, if we reach out 2 *SE*s, we are 95% sure that *p* will be in that interval. In other words, if we reach out 2 *SE*s in either direction of $\hat{p}$, we can be 95% confident that this interval contains the true proportion (population proportion).
- This is called a 95% confidence interval.
- One side note: When we use a 95% confidence interval, we will be having a 5% significance level. We will learn more about it later in the course.

Confidence Interval (cont.)

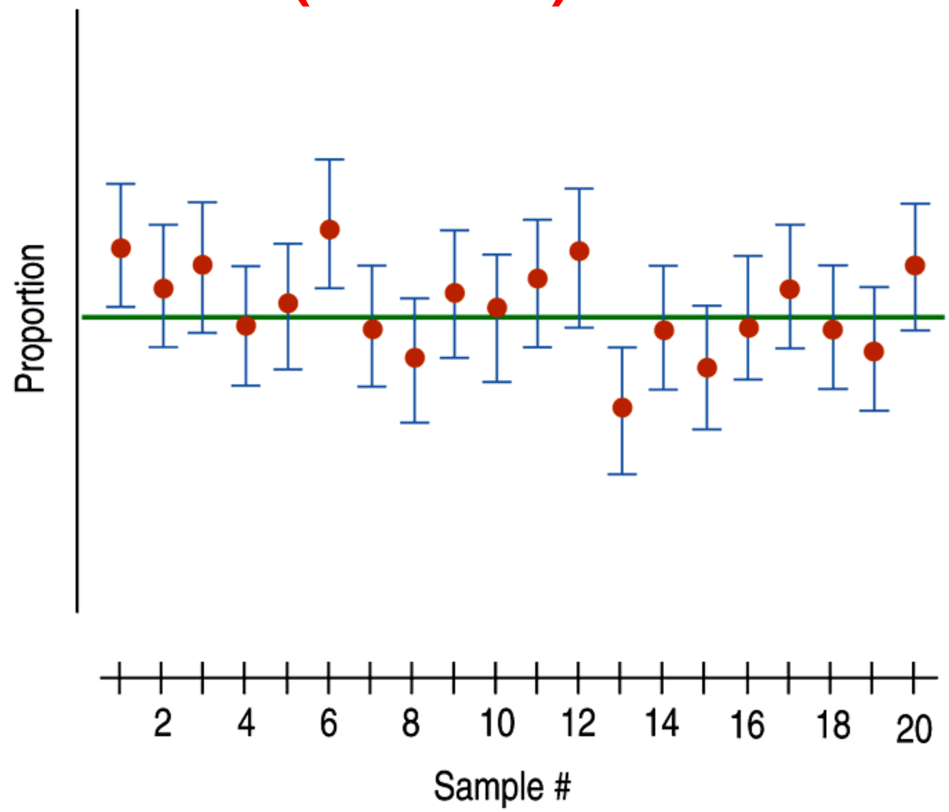


What Does "95% Confidence" Really Mean?

- Each confidence interval uses a sample statistic to estimate a population parameter.
- But, since samples vary, the statistics we use, and thus the confidence intervals we construct, vary as well.

What Does "95% Confidence" Really Mean? (cont.)

- The figure to the right shows that some of our confidence intervals capture the true proportion (the green horizontal line), while others do not:



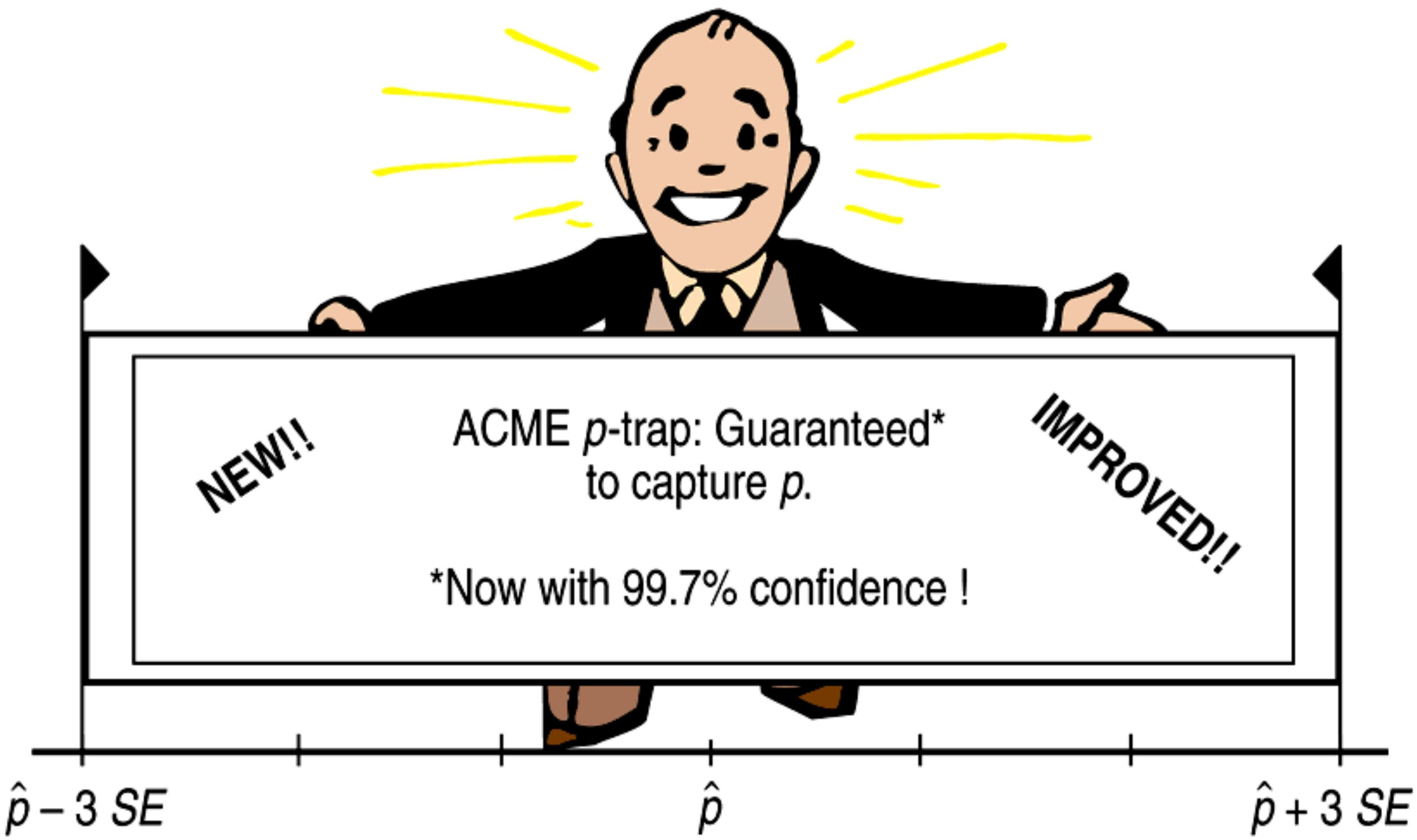
What Does "95% Confidence" Really Mean? (cont.)

- Our confidence is in the *process of* constructing the interval, not in any one interval itself.
- Thus, we expect 95% of all 95% confidence intervals to contain the true parameter that they are estimating.

Margin of Error: Certainty vs. Precision

- We can claim, with 95% confidence, that the interval $\hat{p} \pm 2SE(\hat{p})$ contains the true population proportion.
  - The extent of the interval on either side of $\hat{p}$ is called the margin of error (*ME*).
- In general, confidence intervals have the form *estimate ± ME*.
- The more confident we want to be, the larger our *ME* needs to be.

Margin of Error: Certainty vs. Precision (cont.)



Margin of Error: Certainty vs. Precision (cont.)

- To be more confident, we wind up being less precise.
  - We need more values in our confidence interval to be more certain.
- Because of this, every confidence interval is a balance between certainty and precision.
- The tension between certainty and precision is always there.
  - Fortunately, in most cases we can be both sufficiently certain and sufficiently precise to make useful statements.

Margin of Error: Certainty vs. Precision (cont.)

- The choice of confidence level is somewhat arbitrary, but keep in mind this tension between certainty and precision when selecting your confidence level.
- The most commonly chosen confidence levels are 90%, 95%, and 99% (but any percentage can be used).
- Another side note: 95% confidence interval will give us a 5% significance level. Similarly, a 90% confidence interval will give us a 10% significance level. Thus, the most commonly chosen significance level would be 10%, 5%, and 1%.

Critical Values

- The '2' in $\hat{p} \pm 2SE(\hat{p})$ (our 95% confidence interval) came from the 68-95-99.7% Rule.
- Using a table or technology, we find that a more exact value for our 95% confidence interval is 1.96 instead of 2.

[**Calculator hint**: [2nd][Distr] 3:invNorm(0.025), where 0.025 (2.5%) is coming from dividing the significance level by two.]

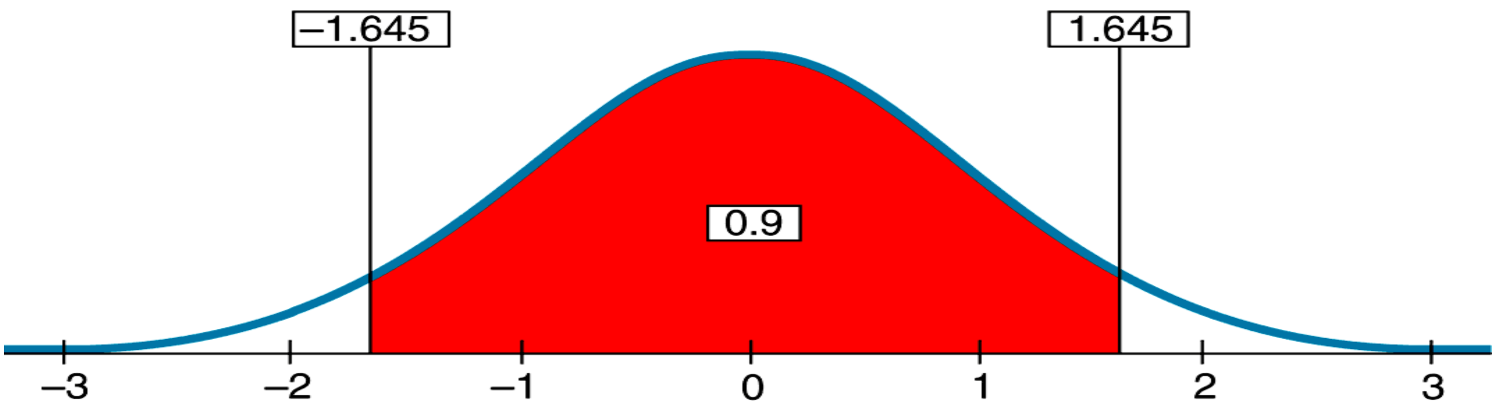
Critical Values (cont.)

- We call 1.96 the critical value and denote it z^*.
- For any confidence level, we can find the corresponding critical value.
- The critical value can be obtained from the standard-normal table or technology, based on the chosen significance level.

Note: In this case, the significance level is 5% because the ccnfidence level is 95%. To obtain the critical value, we divide the significance level by two for each tail.

Critical Values (cont.)

Example: For a 90% confidence interval, the critical value is 1.645:



Calculator hint: You can obtain the critical value -1.645 from your calculator TI-84+ by performing: [2nd][Distr] 3:invNorm(0.05).

Assumptions and Conditions

- All statistical models depend upon assumptions.
  - Different models depend upon different assumptions.
  - If those assumptions are not true, the model might be inappropriate and our conclusions based on it may be wrong.
- You can never be sure that an assumption is true, but you can often decide whether an assumption is plausible by checking a related condition.

Assumptions and Conditions (cont.)

- Here are the assumptions and the corresponding conditions you must check before creating a confidence interval for a proportion:
- Independence Assumption: We first need to *Think* about whether the Independence Assumption is plausible. It's not one you can check by looking at the data. Instead, we check two conditions to decide whether independence is reasonable.

Assumptions and Conditions (cont.)

- Randomization Condition: Were the data sampled at random or generated from a properly randomized experiment? Proper randomization can help ensure independence.
- 10% Condition: Is the sample size no more than 10% of the population?

- Sample Size Assumption: The sample needs to be large enough for us to be able to use the CLT.
  - Success/Failure Condition: We must expect at least 5 successes and at least 5 failures.

One-Proportion *z*-Interval

- When the conditions are met, we are ready to find the confidence interval for the population proportion, *p*.
- The confidence interval is $\hat{p} \pm z^* \times SE(\hat{p})$

  where $SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$

- The critical value, *z**, depends on the particular confidence level, C, that you specify.

How to Find A Confidence Interval for a Proportion?

One-Proportion z-Interval Procedure

Purpose To find a confidence interval for a population proportion, p

Assumptions

1. Simple random sample
2. The number of successes, x, and the number of failures, $n - x$, are both 5 or greater.

STEP 1 **For a confidence level of $1 - \alpha$, use Table II to find $z_{\alpha/2}$.**

STEP 2 **The confidence interval for p is from**

$$\hat{p} - z_{\alpha/2} \cdot \sqrt{\hat{p}(1 - \hat{p})/n} \quad \text{to} \quad \hat{p} + z_{\alpha/2} \cdot \sqrt{\hat{p}(1 - \hat{p})/n}$$

where $z_{\alpha/2}$ is found in Step 1, n is the sample size, and $\hat{p} = x/n$ is the sample proportion.

STEP 3 **Interpret the confidence interval.**

What Can Go Wrong?

Don't Misstate What the Interval Means:

- Don't suggest that the parameter varies.
- Don't claim that other samples will agree with yours.
- Don't be certain about the parameter.
- Don't forget: It's the parameter (not the statistic).
- Don't claim to know too much.
- Do take responsibility (for the uncertainty).

What Can Go Wrong? (cont.)

Margin of Error Too Large to Be Useful:

- We can't be exact, but how precise do we need to be?
- One way to make the margin of error smaller is to reduce your level of confidence. (That may not be a useful solution.)
- You need to think about your margin of error when you design your study.
  - To get a narrower interval without giving up confidence, you need to have less variability.
  - You can do this with a larger sample…

Violations of Assumptions:

- Watch out for biased sampling.
- Think about independence.

What have we learned?

We have learned to:

1. Find and interpret a large-sample confidence interval for a population proportion.
2. Compute the margin of error for the estimate of a population proportion.
3. Describe the relationship between the sample size, confidence interval, and margin of error for a confidence interval for a population proportion.
4. Determine the sample size required for a specified confidence level and margin of error for the estimate of a population proportion.

Credit

Some of these slides have been adapted/modified in part/whole from the slides of the following textbooks.

- Weiss, Neil A., Introductory Statistics, 8th Edition
- Bock, David E., Stats: Data and Models, 3rd Edition