Section 9.2 Confidence Intervals for Proportions

- We will learn to use a sample to say something about the *world at large*.
- This process (statistical inference) is based on our understanding of sampling models, and will be our focus for the rest of the course.
- In this section we learned how to construct a confidence interval for a population proportion.

Standard Deviation → Standard Error

- **Standard Deviations** for Normal Sampling Distributions are:
  - For proportions

    $$SD(\hat{p}) = \sqrt{\frac{pq}{n}}$$

  - For means

    $$SD(\bar{x}) = \frac{\sigma}{\sqrt{n}}$$

- **When we don't know *p* or *σ*, we're stuck, right?**
  - Nope. We will use sample statistics to estimate these population parameters.
- **Whenever we estimate the standard deviation of a sampling distribution, we call it a Standard Error:**

- Standard Error for proportions

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

- Standard Error for means

  $$SE(\bar{x}) = \frac{s}{\sqrt{n}}$$

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.

Slide 3

EXAMPLE: Teens Say Sex Can Wait

Confidence interval for *p*

The Gallup Youth Survey asked a random sample of 439 U.S. teens aged 13 to 17 whether they thought young people should wait to have sex until marriage. Of the sample, 246 said "Yes." Construct and interpret a 95% confidence interval for the proportion of all teens who would say "Yes" if asked this question.

- **Check conditions:**

- **Find the z value:**

- **Find $\hat{p}$:**
- **Calculate the confidence interval:**

- **Conclude (in context):**

Slide 4

Choosing Your Sample Size

- The question of how large a sample to take is an important step in planning any study.
- Choose a Margin or Error (ME) and a Confidence Interval Level.
- The formula requires $\hat{p}$ which we don't have yet because we have not taken the sample.
  - When possible do a pilot study.
  - A good estimate for $\hat{p}$, which will yield the largest value for $\hat{p}\hat{q}$ (and therefore for *n*) is 0.50.
- Solve the formula for *n*.

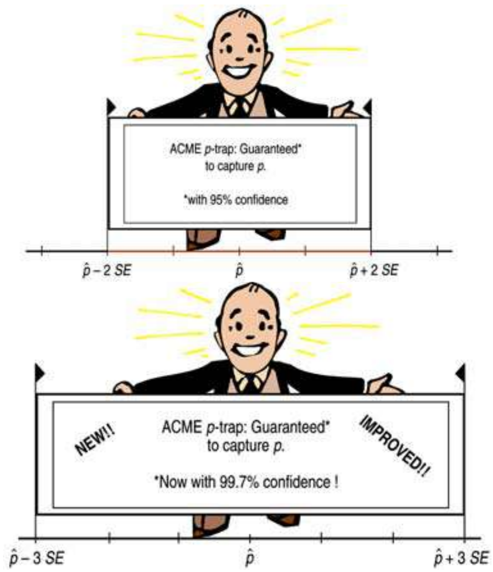
$$ME = z^{*}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

EXAMPLE: Choosing Your Sample Size

A company has received complaints about its customer service. The managers intend to hire a consultant to carry out a survey of customers. Before contacting the consultant, the company president wants some idea of the sample size that they will be required to pay for. One critical question is the degree of satisfaction with the company's customer service, measured on a 5-point scale. The president wants to estimate the proportion *p* of customers who are satisfied (that is, who choose either "satisfied" or "very satisfied," the 2 highest levels on the 5-point scale).

The president wants the estimate to be within 3% (.03) at a 95% confidence level. How large a sample is needed?

Margin of Error: Certainty vs. Precision



Slide 7

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.
- 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).

Slide 8

Key Point

- We've learned to interpret a confidence interval by *Telling* what we believe is true in the entire population from which we took our random sample.
- Of course, *we can't be certain…* **but we can be confident.**

Review the following
Key Points for HW
See me with questions →

Slide 9

Slide 10

Review Assumptions and Conditions

- All statistical models make upon assumptions.
  - Different models make 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.
- 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.
    - **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?
  - Normal Condition for Proportions - Sample Size Assumption: The sample needs to be large enough for us to be able to use the CLT and use a Normal model.
    - **Success/Failure Condition:** We must expect at least 10 "successes" and at least 10 "failures."

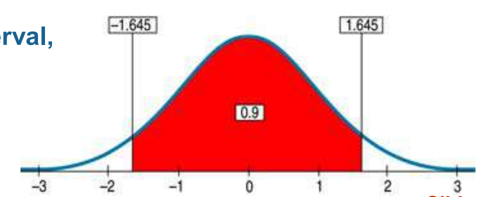
Slide 11

Review Finding Critical Values

- The '2' in $\hat{p} \pm 2SE(\hat{p})$ (our 95% confidence interval) came from the 68-95-99.7% Rule.
- Using Table A or technology, we find that a more exact value for our 95% confidence interval is 1.96 instead of 2.
  - We call 1.96 the critical value and denote it z^*.
- For any confidence level, we can find the corresponding critical value (the number of SEs that corresponds to our confidence interval level).

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

- **How do you find this number?**
  - **Table A or invNorm(.05,0,1)**



Slide 12

Key Points: 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 about the parameter (not the statistic).
- Don't claim to know too much.
- Do take responsibility (for the uncertainty).
- Do treat the whole interval equally.

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…

Slide 13

Key Points: What Can Go Wrong? (cont.)

Choosing Your Sample Size:

- In general, the sample size needed to produce a confidence interval with a given margin of error at a given confidence level is:

$$n = \frac{\left(z^*\right)^2 \hat{p}\hat{q}}{ME^2}$$

  where z^* is the critical value for your confidence level.
- To be safe, round up the sample size you obtain.

Violations of Assumptions:

- Watch out for biased samples—keep in mind what you learned in Chapter 2.
- Think about independence.

Slide 14

APPENDIX

EXAMPLE: Teens Say Sex Can Wait

Confidence interval for *p*

- **Check conditions:**
- **Random**: Gallup surveyed a random sample of 439 U.S. teens.
- **Normal**: We check the counts of success and failures

$$n\hat{p} = 439 * .56 = 246 \text{ and } n(1-\hat{p}) = 439 * .44 = 193$$

The counts of successes and failures are both ≥ 10

- **Independent**: Since Gallup sampled without replacement, we need to check the 10% condition. At least 10(439) = 4390 U.S. teens aged 13 to 17.

- **Find the 95% z value:** Calculator command: invNorm(.025,0,1) = -1.96 or invNorm(.975,0,1) = 1.96
- **Find**: $\hat{p}$ =246/439=.56
- **Calculate the confidence interval:** $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.56 \pm 1.96\sqrt{\frac{(0.56)(0.44)}{439}} = 0.56 \pm 0.046$

$= (0.514, 0.606)$

- **Conclude (in context):** We are 95% confident that the interval from .514 to .606 captures the true proportion of 13-to 17-year-olds in the United States who would say that teens should wait until marriage to have sex.

Slide 15

APPENDIX

EXAMPLE: Choosing Your Sample Size

The Customer Service Problem Here is how to determine the sample size needed to estimate *p* within 0.03 with 95% confidence.

✓ The critical value for 95% confidence is $z^* = 1.96$.

$$ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

✓ Since the company president wants a margin of error of no more than 0.03, we need to solve the equation

$$1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le 0.03$$

Multiply both sides by square root *n* and divide both sides by 0.03.

$$\frac{1.96}{0.03}\sqrt{\hat{p}(1-\hat{p})} \le \sqrt{n}$$

Square both sides.

$$\left(\frac{1.96}{0.03}\right)^2 \hat{p}(1-\hat{p}) \le n$$

Substitute 0.5 for the sample proportion to find the largest *ME* possible.

$$\left(\frac{1.96}{0.03}\right)^2 (0.5)(1-0.5) \le n$$

$$1067.111 \le n$$

We round up to 1068 respondents to ensure the margin of error is no more than 0.03 at 95% confidence.

Slide 16