Chapter 19 Summary
Confidence Intervals for Proportions

What have we learned?

- Finally we have learned 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 book.
- In this chapter we learned how to construct a confidence interval for a population proportion.
- And, we learned that interpretation of our confidence interval is key—we can't be *certain*, but we can be confident.

A Confidence Interval

- Recall that the sampling distribution model of $\hat{p}$ is centered at *p*, with standard deviation $\sqrt{\frac{pq}{n}}$.
- Since we don't know *p*, we can't find the true standard deviation of the sampling distribution model, so we need to find the standard error: $SE(\hat{p})=\sqrt{\frac{\hat{p}\hat{q}}{n}}$
- 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…
- Consider the 95% level:
  - There's a 95% chance that *p* is no more than 2 *SE*s away from .
  - 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 , we can be 95% confident that this interval contains the true proportion.
- This is called a 95% confidence interval.

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.
- 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

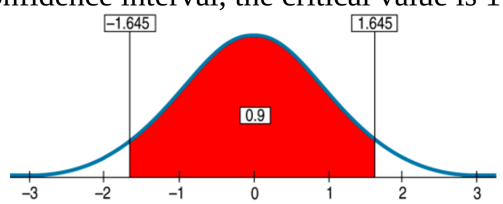
- We can claim, with 95% confidence, that the interval $\hat{p}\pm 2SE(\hat{p})$ contains the true population proportion.
- The extent of the interval on either side of $\hat{p}$ is called the margin of error (*ME*).
- In general, confidence intervals have the form *estimate* ± *ME*.
- The more confident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

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

- 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).

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.
  - We call 1.96 the critical value and denote it z^*.
- For any confidence level, we can find the corresponding critical value.
- Example: For a 90% confidence interval, the critical value is 1.645:



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.
- Here are the assumptions and the corresponding conditions you must check before creating a confidence interval for a proportion:
- Independence Assumption: The data values are assumed to be independent from each other. We check three conditions to decide whether independence is reasonable.
  - Plausible Independence Condition: Is there any reason to believe that the data values somehow affect each other? This condition depends on your knowledge of the situation—you can't check it with data.
- 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 10 "successes" and at least 10 "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.

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).
- 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…
- 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 12.
  - Think about independence.