8.2
ESTIMATING A POPULATION PROPORTION

HW: P. 496 (27, 31-37 ODD, 41, 43, 47, 49-52)

OVERVIEW…

- In the last section, you learned the basic ideas behind confidence intervals.
- In the next two sections, you will learn how to construct and interpret confidence intervals for proportions and means.

CONDITIONS FOR INFERENCE ABOUT A POPULATION PROPORTION

- Random Sample
  - The data are a random sample from the population of interest.
- 10% Rule (Independent)
  - The sample size is no more than 10% of the population size: $n \le \frac{1}{10}N$
- Large Counts (Normal)
  - Counts of successes and failures must be 10 or more: $n\hat{p} \ge 10$ and $n(1-\hat{p}) \ge 10$
  - $n\hat{p}$ is the number of successes and $n(1-\hat{p})$ are the number of failures

CHECK YOUR UNDERSTANDING P. 487

In each of the following settings, check whether the conditions for calculating a confidence interval for the population proportion p are met.

1. An AP Statistics class at a large high school conducts a survey. They ask the first 100 students to arrive at school one morning whether or not they slept at least 8 hours the night before. Only 17 students say "Yes."

   Random: not met. This was a convenience sample.

   Normal: met. $n\hat{p} = 17$ and $n(1 - \hat{p}) = 83$. Both are at least 10.

   Independent: met. A large high school has more than $10(100) = 1000$ students.

CHECK YOUR UNDERSTANDING P. 487

2. A quality control inspector takes a random sample of 25 bags of potato chips from the thousands of bag filled in an hour. Of the bags selected, 3 had too much salt.

Random: met.

Normal: not met. $n\hat{p} = 3$ is not at least 10.

Independent: met. There are thousands of bags produced per hour, so the sample is less than 10% of the population.

CONSTRUCTING A CONFIDENCE INTERVAL FOR *P*

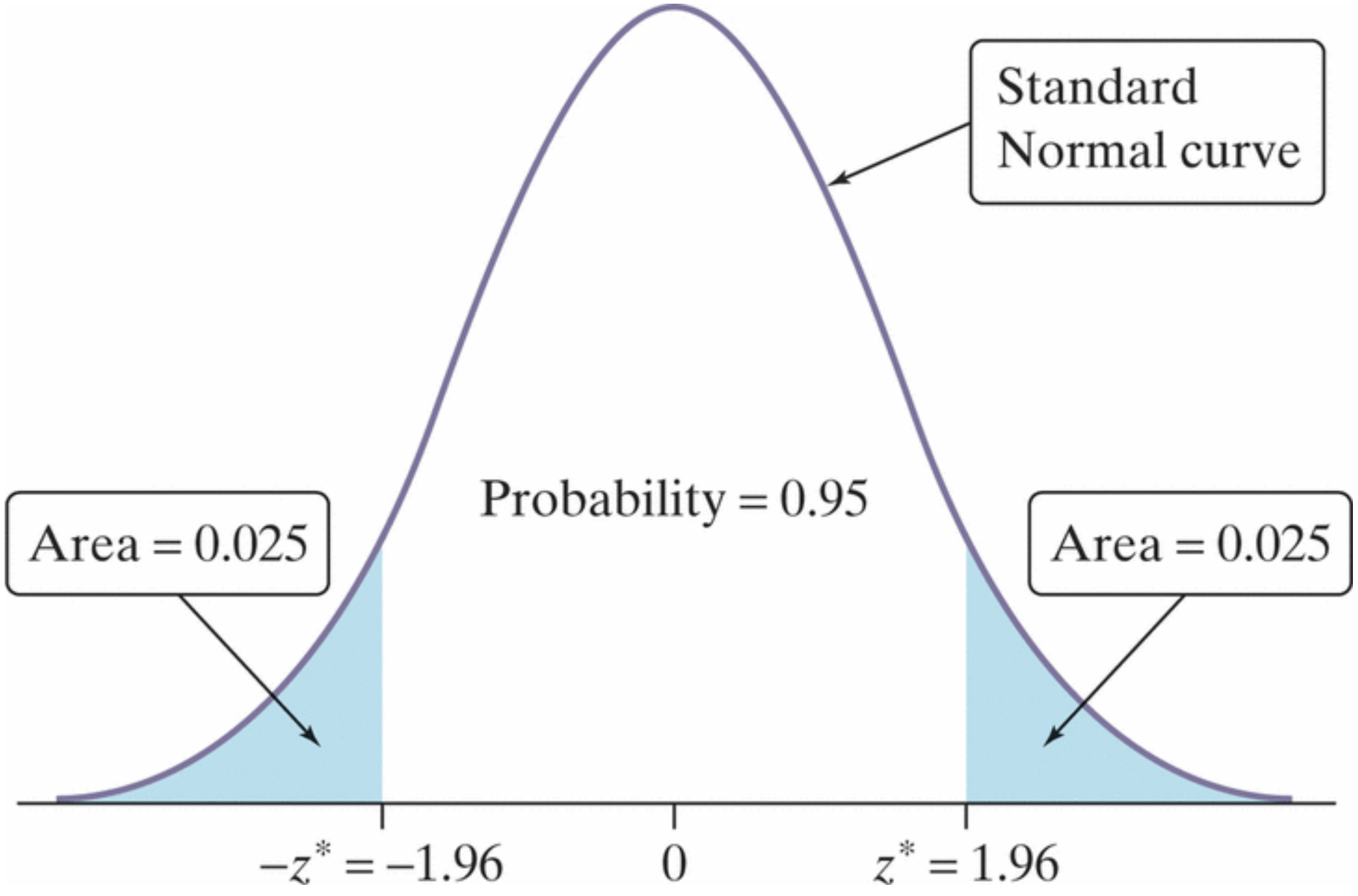
STANDARD ERROR

- When the **standard deviation** of a statistic is estimated from data, the result is called the **standard error (SE)** of the statistic.
- Since we don't know the value of p, we replace it with the sample proportion $\hat{p}$.
- $SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ Standard error

CRITICAL VALUE z^*

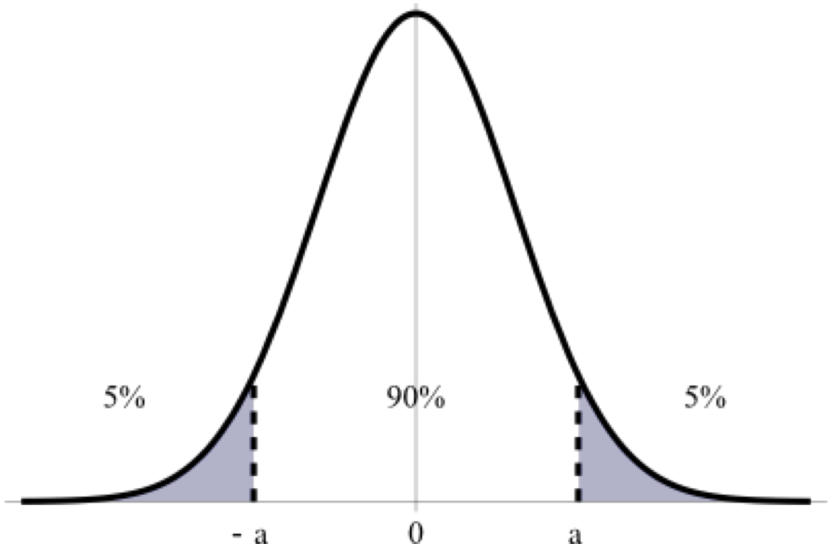
- When we use a table or calculator to get a critical value, we use $*$. This reminds us that the value is a critical value, not a standardized score that has been calculated from data.
- To find a level C confidence interval, we need to catch the central area C under the standard Normal curve. The next slide shows an example that shows how to get the **critical value** z^* for a different confidence level.

CRITICAL VALUE z^*



CRITICAL VALUE z^*

- Let's practice!
- Find the **Critical Value z^*** for a 90% and 80% confidence interval.
- *Hint: use invNorm
- 90%:
  - $invnorm(.05, 0, 1)$
  - $z^* = \pm 1.645$
- 80%:
  - $invnorm(.1, 0, 1)$
  - $z^* = \pm 1.28$



ONE-SAMPLE z INTERVAL FOR A POPULATION PROPORTION

Choose an SRS of size *n* from a large population that contains an unknown proportion *p* of successes. An approximate level *C* **confidence interval for *p* is**

Statistic

$$\hat{p} \pm \boxed{z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

Margin of error

Where z^* is the critical value for the standard Normal curve with area *C* between $-z^*$ and z^*.

Use this interval only when the numbers of successes and failures in the sample are both at least 10 and the population is at least 10 times as large as the sample.

4 STEP PROCESS

To construct a confidence interval for a population proportion p, you should follow the four step process introduced in Chapter 1.

- **State** the parameter you want to estimate and at what confidence level.
- **Plan** which confidence interval you will construct and verify that the conditions have been met.
- **Do** the actual construction of the interval using the basic idea from Section 8.1
- **Conclude** by interpreting the interval in the context of the problem.

CHECK YOUR UNDERSTANDING P. 490

Alcohol abuse has been described by college presidents as the number one problem on campus, and it is an important cause of death in young adults. How common is it? A survey of 10,904 randomly selected U.S. college students collected information on drinking behavior and alcohol-related problems. The researchers defined "frequent binge drinking" as having five or more drinks in a row three or more times in the past two weeks. According to this definition, 2486 students were classified as frequent binge drinkers.

1. Identify the population and the parameter of interest.

   Population: U.S. college students.

   Parameter: true proportion who are classified as binge drinkers.

2. Check conditions for constructing a confidence interval for the parameter.

   Random: students were chosen randomly.

   Normal: 2486 successes and 8418 failures, both are at least 10.

   Independent: 10,904 is clearly less than 10% of all U.S. college students. All conditions are met.

CHECK YOUR UNDERSTANDING P. 490

3. Find the critical value for a 99% confidence interval. Show your method. Then calculate the interval.

Find z*:

$invnorm(.005, 0, 1)$

z* = 2.576.

Find $\hat{p}$:

$$\hat{p} = \frac{2486}{10{,}904} \approx 0.228$$

Calculate Interval:

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.228 \pm (2.576)\sqrt{\frac{.228(.772)}{10904}} = 0.228 \pm 0.0103$$

(0.218, 0.238)

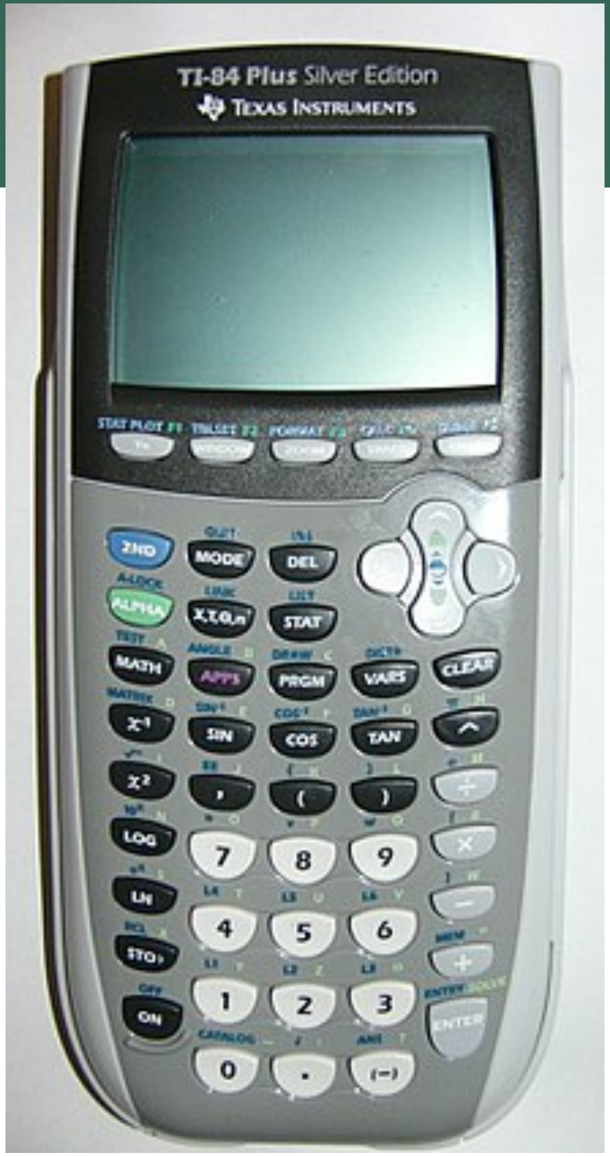
4. Interpret the interval in context.

We are 99% confident that the interval from 21.8% to 23.8% captures the true proportion of U.S. college students who would be classified as binge drinkers.

CALCULATOR SHORTCUT

- Press STAT → Test
- A:1-PropZInt
  - x: number of success
  - n: sample size
  - C-Level: in decimal form
  - Calculate

Let's try the previous question!



CHOOSING THE SAMPLE SIZE

- Our goal is to estimate the parameter as precisely as possible. We want high confidence and low margin of error.
- To achieve that, we can determine how large a sample size we need before proceeding with the data collection.
- To calculate the sample size necessary to achieve a set margin of error at a confidence level *C*, we simply solve the following inequality for *n*:

$$z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq ME$$

where $\hat{p}$ is estimated based on a previous study or set to 0.5 to maximize the possible margin of error.

EXAMPLE:

Suppose you wanted to estimate p = the true proportion of students at your school who have a tattoo with 95% confidence and a margin of error of no more than 0.10.

Determine how many students should be surveyed.

$$z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq ME$$

$$1.96 \sqrt{\frac{0.5(1-0.5)}{n}} \leq 0.10$$

$$\boldsymbol{n \geq 96.04}$$