Chapter 8: Estimating with Confidence

Section 8.2

Estimating a Population Proportion

The Practice of Statistics, 4th edition – For AP*

STARNES, YATES, MOORE

Chapter 8
Estimating with Confidence

- 8.1 Confidence Intervals: The Basics
- **8.2 Estimating a Population Proportion**
- 8.3 Estimating a Population Mean

Section 8.2
Estimating a Population Proportion

Learning Objectives

After this section, you should be able to…

- ✓ CONSTRUCT and INTERPRET a confidence interval for a population proportion
- ✓ DETERMINE the sample size required to obtain a level *C* confidence interval for a population proportion with a specified margin of error
- ✓ DESCRIBE how the margin of error of a confidence interval changes with the sample size and the level of confidence *C*

Activity: The Beads

Your teacher has a container full of different colored beads. Your goal is to estimate the actual proportion of red beads in the container.

- ✓ Form teams of 3 or 4 students.
- ✓ Determine how to use a cup to get a simple random sample of beads from the container.
- ✓ Each team is to collect one SRS of beads.
- ✓ Determine a point estimate for the unknown population proportion.
- ✓ Find a 90% confidence interval for the parameter p. Consider any conditions that are required for the methods you use.
- ✓ Compare your results with the other teams in the class.

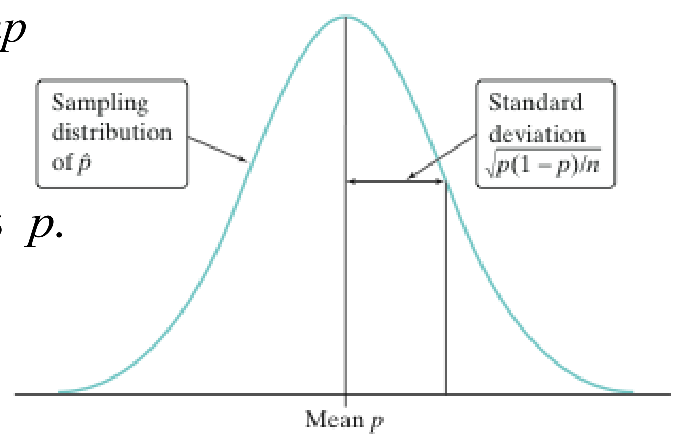
Conditions for Estimating *p*

Suppose one SRS of beads resulted in 107 red beads and 144 beads of another color. The point estimate for the unknown proportion *p* of red beads in the population would be

$$\hat{p} = \frac{107}{251} = 0.426$$

How can we use this information to find a confidence interval for *p*?

- If the sample size is large enough that both np and $n(1-p)$ are at least 10, the sampling distribution of $\hat{p}$ is approximately Normal.
- The mean of the sampling distribution of $\hat{p}$ is p.
- The standard deviation of the sampling distribution of $\hat{p}$ is $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.



In practice, we do not know the value of p. If we did, we would not need to construct a confidence interval for it! In large samples, $\hat{p}$ will be close to p, so we will replace p with $\hat{p}$ in checking the Normal condition.

Conditions for Estimating *p*

Check the conditions for estimating *p* from our sample. $\hat{p} = \frac{107}{251} = 0.426$

Random: The class took an SRS of 251 beads from the container.

Normal: Both np and $n(1 - p)$ must be greater than 10. Since we don't know p, we check that

$$n\hat{p} = 251\left(\frac{107}{251}\right) = 107 \text{ and } n(1-\hat{p}) = 251\left(1 - \frac{107}{251}\right) = 144$$

The counts of successes (red beads) and failures (non-red) are both ≥ 10.

Independent: Since the class sampled without replacement, they need to check the 10% condition. At least 10(251) = 2510 beads need to be in the population. The teacher reveals there are 3000 beads in the container, so the condition is satisfied.

Since all three conditions are met, it is safe to construct a confidence interval.

Alternate Example – The pennies

- Ms. Smith's class wants to construct a confidence interval for the proportion *p* of pennies more than 10 years old in their collection. Their sample had 57 pennies more than 10 years old and 45 pennies that were at most 10 years old.
- **Problem:** Check that the conditions for constructing a confidence interval for *p* are met.
- **Solution:**
- *Random*: The class took an SRS of 102 pennies from the collection.

$$\bullet\ \text{N}ormal: n\hat{p} = 102\left(\frac{57}{102}\right) = 57 \text{ and } n(1-\hat{p}) = 102\left(\frac{45}{102}\right) = 45$$

Both the number of successes and the number of failures are at least 10.

- *Independent*: Since we are sampling without replacement, the number of pennies in the population must be at least 10(102) = 1020. Since there are more than 2000 pennies in Ms. Smith's collection, the 10% condition is met.

Constructing a Confidence Interval for *p*

We can use the general formula from Section 8.1 to construct a confidence interval for an unknown population proportion *p*:

statistic $\pm$ **(critical value)** $\cdot$ **(standard deviation of statistic)**

The sample proportion $\hat{p}$ is the statistic we use to estimate p. When the Independent condition is met, the standard deviation of the sampling distibution of $\hat{p}$ is

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Since we don't know p, we replace it with the sample proportion $\hat{p}$. This gives us the **standard error (SE)** of the sample proportion:

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

Definition:

When the standard deviation of a statistic is estimated from data, the results is called the **standard error** of the statistic.

Finding a Critical Value

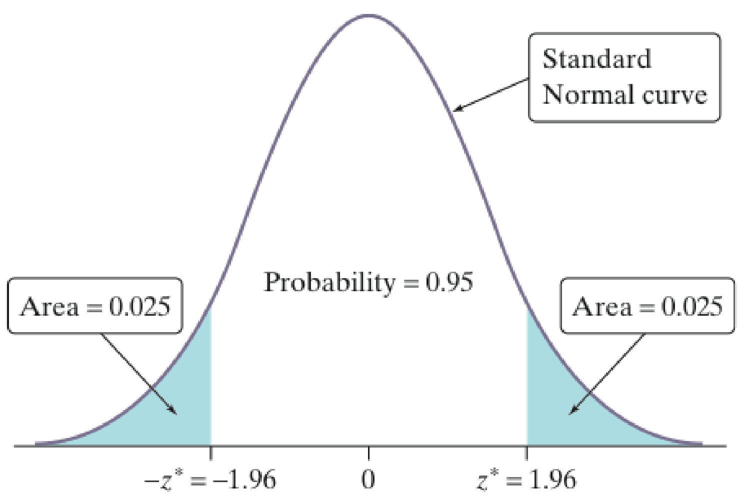
How do we find the critical value for our confidence interval?

statistic ± (critical value) · (standard deviation of statistic)

If the Normal condition is met, we can use a Normal curve. To find a level *C* confidence interval, we need to catch the central area *C* under the standard Normal curve.

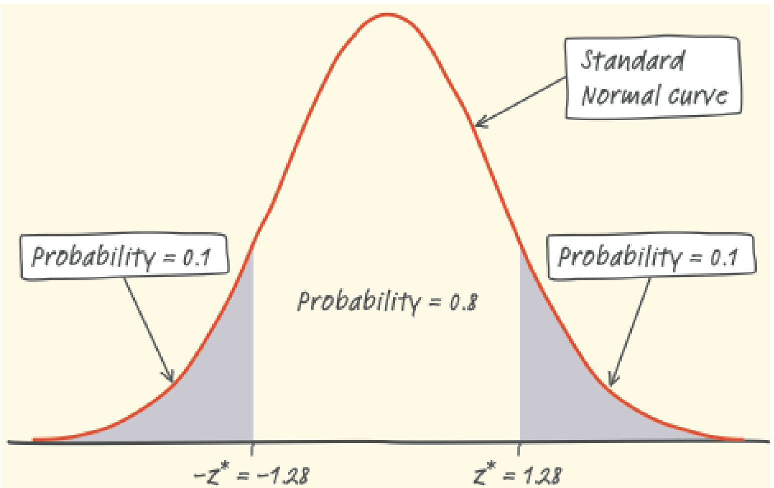
For example, to find a 95% confidence interval, we use a critical value of 2 based on the 68-95-99.7 rule. Using Table A or a calculator, we can get a more accurate critical value.

Note, the **critical value *z**** is actually 1.96 for a 95% confidence level.



Finding a Critical Value

Use Table A to find the critical value z^* for an 80% confidence interval. Assume that the Normal condition is met.



Since we want to capture the central 80% of the standard Normal distribution, we leave out 20%, or 10% in each tail.
Search Table A to find the point z^* with area 0.1 to its left.

z	.07	.08	.09
– 1.3	.0853	.0838	.0823
– 1.2	.1020	.1003	.0985
– 1.1	.1210	.1190	.1170

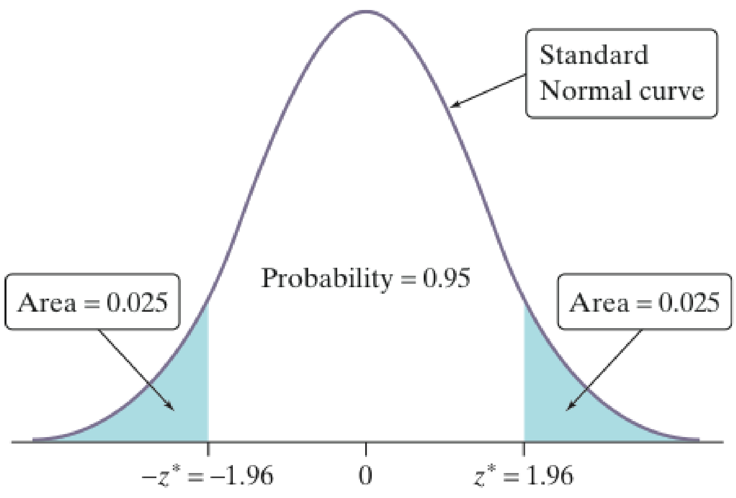
The closest entry is $z = -1.28$.

So, the **critical value z^*** for an 80% confidence interval is $z^* = 1.28$.

Alternate Example – 96% confidence

- **Problem:** Use Table A to find the critical value z^* for a 96% confidence interval. Assume that the Normal condition is met.

Solution: For a 96% confidence interval, we need to capture the middle 96% of the standard Normal distribution. This leaves out 2% in each tail. So, we want to find the z-score with an area of 0.02 to its left. The closest entry is $z = -2.05$, so the critical value we want is $z^* = 2.05$.



One-Sample *z* Interval for a Population Proportion

Once we find the critical value z^*, our confidence interval for the population proportion p is

statistic $\pm$ (critical value) $\cdot$ (standard deviation of statistic)

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

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**

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

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.

One-Sample *z* Interval for a Population Proportion

Calculate and interpret a 90% confidence interval for the proportion of red beads in the container. Your teacher claims 50% of the beads are red. Use your interval to comment on this claim.

z	.03	.04	.05
– 1.7	.0418	.0409	.0401
– 1.6	.0516	.0505	.0495
– 1.5	.0630	.0618	.0606

- ✓ sample proportion = 107/251 = 0.426
- ✓ We checked the conditions earlier.
- ✓ For a 90% confidence level, z^* = 1.645

statistic ± (critical value) • (standard deviation of the statistic)

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

$$= 0.426 \pm 1.645\sqrt{\frac{(0.426)(1-0.426)}{251}}$$

$$= 0.426 \pm 0.051$$

$$= (0.375,\ 0.477)$$

We are 90% confident that the interval from 0.375 to 0.477 captures the actual proportion of red beads in the container.

Since this interval gives a range of plausible values for p and since 0.5 is not contained in the interval, we have reason to doubt the claim.

Alternate Example – The pennies

- **Problem:** Ms. Smith's class took an SRS of 102 pennies and discovered that 57 of the pennies were more than 10 years old.

  (a) Calculate and interpret a 99% confidence interval for *p* = the true proportion of pennies from the collection that are more than 10 years old.

  The proportion of pennies more than 10 years old in the sample was = 57/102 = 0.559. The critical value for a 99% confidence interval can be found by looking for the point that has an area of 0.005 to the left. The calculator's invNorm(0.005,0,1) gives –2.576 so the appropriate critical value for 99% confidence is *z** = 2.576. The 99% confidence interval is:

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.559 \pm 2.576\sqrt{\frac{0.559(1-0.559}{102}} = 0.559 \pm 0.127 = (0.432, 0.686)$$

We are 99% confident that the interval from 0.432 to 0.686 captures the actual proportion of pennies in the collection that are more than 10 years old.

(b) Is it plausible that exactly 60% of all the pennies in the collection are more than 10 years old? Explain.

Yes, since 0.6 is included in the confidence interval, it is plausible that 60% of all the pennies in the collection are more than 10 years old.

The Four-Step Process

We can use the familiar four-step process whenever a problem asks us to construct and interpret a confidence interval.

Confidence Intervals: A Four-Step Process

State: What *parameter* do you want to estimate, and at what confidence level?

Plan: *Identify* the appropriate inference *method*. Check *conditions*.

Do: If the conditions are met, perform *calculations*.

Conclude: *Interpret* your interval in the context of the problem.

Alternate Example: Kissing the right way?

- According to an article in the *San Gabriel Valley Tribune* (2-13-03), "Most people are kissing the 'right way'." That is, according to the study, the majority of couples tilt their heads to the right when kissing. In the study, a researcher observed a random sample 124 couples kissing in various public places and found that 83/124 (66.9%) of the couples tilted to the right. Construct and interpret a 95% confidence interval for the proportion of all couples who tilt their heads to the right when kissing.
- *State:* We want to estimate *p* = the true proportion of couples that tilt their heads to the right when kissing at the 95% confidence level.
- *Plan:* We will use a one-sample *z* interval for *p* if the following conditions are satisfied.

- Random: The researcher observed a random sample of couples.

$\cdot$ Normal: $n\hat{p} = 83 \geq 10 \text{ and } n(1-\hat{p}) = 41 \geq 10$

- Independent: The number of couples in the population is more than 10(124) = 1240.

- *Do:* $\hat{p} \pm z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.669 \pm 1.96\sqrt{\frac{0.669(1-0.669)}{124}} = 0.669 \pm 0.083 = (0.586, 0.752)$

Conclude: We are 95% confident that the interval from 0.586 to 0.752 captures the true proportion of couples that tilt their heads to the right when kissing.

Choosing the Sample Size

In planning a study, we may want to choose a sample size that allows us to estimate a population proportion within a given margin of error.

The margin of error (*ME*) in the confidence interval for *p* is

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

✓ *z** is the standard Normal critical value for the level of confidence we want.

Because the margin of error involves the sample proportion $\hat{p}$, we have to guess the latter value when choosing n. There are two ways to do this:

- Use a guess for $\hat{p}$ based on past experience or a pilot stud
- Use $\hat{p} = 0.5$ as the guess. ME is largest when $\hat{p} = 0.5$

Sample Size for Desired Margin of Error

To determine the sample size *n* that will yield a level *C* confidence interval for a population proportion *p* with a maximum margin of error *ME*, solve the following inequality for *n*:

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

where $\hat{p}$ is a guessed value for the sample proportion. The margin of err will always be less than or equal to ME if you take the guess $\hat{p}$ to be 0.5.

Example: Customer Satisfaction

Read the example on page 493. **Determine the sample size needed to estimate *p* within 0.03 with 95% confidence.**

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

Alternate Example: Tattoos

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

- **Problem:** Determine how many students should be surveyed to estimate p within 0.10 with 95% confidence.

- **Solution:** Since we don't have any previous knowledge of the proportion of students with a tattoo, we will use = 0.5 to estimate the sample size needed.

$$1.96\sqrt{\frac{0.5(1-0.5)}{n}} \le 0.10 \rightarrow \left(\frac{1.96}{0.10}\right)^2 (0.5)(1-0.5) \le n \rightarrow n \ge 96.04$$

So, we need to survey at least 97 students to estimate the true proportion of students with a tattoo with 95% confidence and a margin of error of at most 0.10.

Section 8.2
Estimating a Population Proportion

Summary

In this section, we learned that…

- ✓ Confidence intervals for a population proportion p are based on the sampling distribution of the sample proportion $\hat{p}$. When n is large enough that both np and $n(1-p)$ are at least 10, the sampling distribution of $\hat{p}$ is approximately Normal.

- ✓ In practice, we use the sample proportion $\hat{p}$ to estimate the unknown parameter p. We therefore replace the standard deviation of $\hat{p}$ with its standard error when constructing a confidence interval.

  The level C confidence interval for p is: $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Section 8.2
Estimating a Population Proportion

Summary

In this section, we learned that…

- When constructing a confidence interval, follow the familiar four-step process:
  - STATE: What parameter do you want to estimate, and at what confidence level?
  - PLAN: Identify the appropriate inference method. Check conditions.
  - DO: If the conditions are met, perform calculations.
  - CONCLUDE: Interpret your interval in the context of the problem.
- The sample size needed to obtain a confidence interval with approximate margin of error *ME* for a population proportion involves solving

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

for n, where $\hat{p}$ is a guessed value for the sample proportion, and z^* is the critical value for the level of confidence you want. If you use $\hat{p} = 0.5$ in this formula, the margin of error of the interval will be less than or equal to ME.

+

Looking Ahead…

In the next Section…

We'll learn how to estimate a population mean.

We'll learn about

- ✓ **The one-sample *z* interval for a population mean when *σ* is known**
- ✓ **The *t* distributions when *σ* is unknown**
- ✓ **Constructing a confidence interval for *μ***
- ✓ **Using *t* procedures wisely**