

**Chapters 9-12**

Stat 128: Elementary Statistics

**Nathan Warner**



**Northern Illinois  
University**

Computer Science  
Northern Illinois University  
United States  
July 27, 2023

## Contents

<b>Chapter 9</b>	3
0.1 9.1 Estimating a Population Proportion	3

## Chapter 9

---

### 0.1 9.1 Estimating a Population Proportion

#### *Learning Objectives For This Section:*

1. Obtain a Point Estimate for the Population Proportion
2. Construct and Interpret a Confidence Interval for the Population Proportion
3. Determine the Sample Size Necessary for Estimating a Population Proportion within a Specified Margin of Error

#### Vocab:

- A **point estimate** is the value of a statistic that estimates the value of a parameter.
- A **Confidence Interval** for an unknown parameter consists of an interval of numbers base on a point estimate.
- The **level of confidence** represents the expected proportion of intervals that will contain the parameter if a large number of different samples is obtained.

#### Formulas/Notation:

- **Point Estimate = Sample Proportion**

$$\hat{p} = \frac{x}{n}.$$

- The **Margin of Error** is denoted:

$$E.$$

- **Confidence Intervals for a Proportion** are of the form:

$$\text{point estimate} \pm \text{margin of error}$$

That is:

$$\hat{p} \pm E.$$

We have 95% confidence that the sampling proportion will lie between:

$$\hat{p} \pm 1.96\sigma_{\hat{p}}.$$

- The **level of confidence** is denoted:

$$(1 - \alpha) \cdot 100\%.$$

- **Standard Error:**

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

- **Critical Value**,  $z_{\frac{\alpha}{2}}$  represents the number of standard deviations the sample statistic can be from the parameter and still result in an interval that includes the parameter.

- **Constructing a  $(1-\alpha) \cdot 100\%$  Confidence Interval for a Population Proportion:**

- Suppose that a simple random sample of size  $n$  is taken from a population or the data are the result of a randomized experiment. A  $(1-\alpha)\cdot 100\%$  confidence interval for  $p$  is given by the following quantities:

$$\text{Lower bound : } \hat{p} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\text{Upper bound : } \hat{p} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

**Note:** It must be the case that  $n\hat{p}(1-\hat{p}) \geq 10$  and  $n \leq 0.05N$  to construct this interval. Use  $\hat{p}$  in place of  $p$  in the standard deviation. This is because  $p$  is unknown, and  $\hat{p}$  is the best point estimate of  $p$ .

- The **margin of error**,  $E$ , in a  $(1-\alpha)\cdot 100\%$  confidence interval for a population proportion is given by

$$E = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

or :  $\frac{\text{Upper Limit} - \text{Lower Limit}}{2}.$

- **Sample Size Needed for Estimating the Population Proportion**

$$n = \hat{p} \cdot (1 - \hat{p}) \cdot \left( \frac{z_{\alpha/2}}{E} \right)^2.$$

(rounded up to the next integer) where  $\hat{p}$  is a prior estimate of  $p$ .

If a prior estimate of  $p$  is unavailable, the sample size required is

$$n = 0.25 \cdot \left( \frac{z_{\alpha/2}}{E} \right)^2.$$

rounded up to the next integer. The margin of error should always be expressed as a decimal when using these formulas.

### Introduction

#### Two Types of Inferential Statistics:

1. Estimation
2. Hypothesis Testing

### Obtain a Point Estimate for the Population Proportion

Suppose we want to estimate the proportion of adult Americans who believe that the amount they pay in federal income taxes is fair. It is unreasonable to expect that we could survey every adult American. Instead, we use a sample of adult Americans to arrive at an estimate of the proportion. We call this estimate a point estimate.

**Example:****Problem:**

The Gallup Organization conducted a poll in April 2017 in which a simple random sample of 1019 Americans aged 18 and older were asked, “Do you regard the income tax that you will have to pay this year as fair?” Of the 1019 adult Americans surveyed, 620 said yes. Obtain a point estimate for the proportion of Americans aged 18 and older who believe that the amount of income tax they pay is fair.

**Approach:** The point estimate of the population proportion is  $\hat{p} = \frac{x}{n}$ , where  $x = 620$  and  $n = 1019$ .

**Solution:**

Substituting into the formula, we get  $\hat{p} = \frac{x}{n} = \frac{620}{1019} = 0.608 = 60.8\%$ .

We estimate that 60.8% of Americans aged 18 and older believe that the amount of income tax they pay is fair.

### Construct and Interpret a Confidence Interval for the Population Proportion

What if we conducted a different random sample of 1019 Americans aged 18 years or older? Would we get the same sample proportion who believe the amount of income tax they pay is fair? Probably not. Why? Because statistics such as  $\hat{p}$  vary from sample to sample. So a different random sample of adult Americans might result in a different point estimate of the population proportion, such as  $\hat{p} = 0.593$ .

If the method used to select the adult Americans was done appropriately, both point estimates would be good guesses of the population proportion. Due to variability in the sample proportion, we report a range (or interval) of values, including a measure of the likelihood that the interval includes the unknown population proportion.

Consider the following:

$$\hat{p} = 0.54 \quad E = 0.034.$$

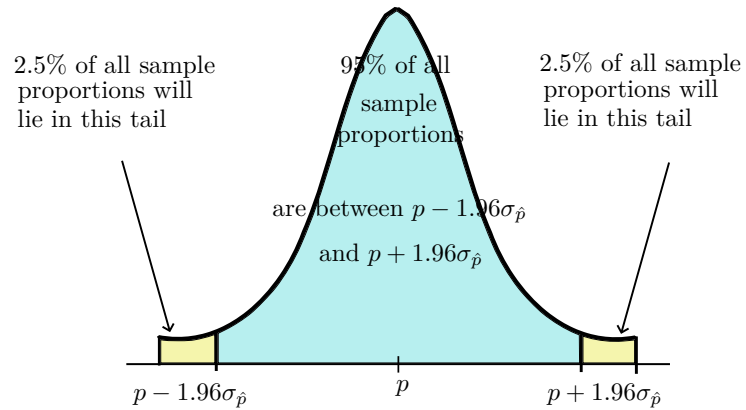
Then:

$$0.54 - 0.034 = 0.506$$

$$0.54 + 0.034 = 0.574.$$

With this we can say “We are 95% confident that the proportion is between 0.506 and 0.574”

Figure:



Where does the 1.96 comes from?

$$z_{\alpha}$$

$$z_{0.025} = 1.96.$$

This graphic also infers that 95% of all sample proportions will lie between:

$$p - 1.96\sigma_{\hat{p}} < \hat{p} < p + 1.96\sigma_{\hat{p}}.$$

By solving this inequality such that  $p$  is in the center:

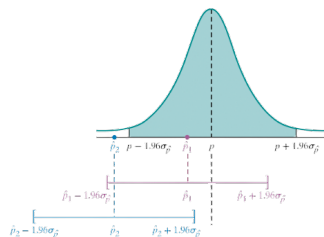
$$\hat{p} - 1.96\sigma_{\hat{p}} < p < \hat{p} + 1.96\sigma_{\hat{p}}.$$

And we can write this in shorthand form:

$$\hat{p} \pm 1.96\sigma_{\hat{p}}$$

$$\text{or : } \hat{p} \pm E.$$

figure:



**Summary:**

- for a 95% confidence interval, any sample proportion that lies within 1.96 standard errors of the population proportion will result in a confidence interval that includes  $p$ . This will happen in 95% of all possible samples.
- Any sample proportion that is more than 1.96 standard errors from the population proportion will result in a confidence interval that does not contain  $p$ . This will happen in 5% of all possible samples (those sample proportions in the tails of the distribution).
- A confidence interval for an unknown parameter consists of an interval of numbers based on a point estimate.
- The level of confidence represents the expected proportion of intervals that will contain the parameter if a large number of different samples is obtained. The level of confidence is denoted  $(1-\alpha) \cdot 100\%$ .
- Whether a confidence interval contains the population parameter depends solely on the value of the sample statistic. Any sample statistic that is in the tails of the sampling distribution will result in a confidence interval that does not include the population parameter. See Figure 1. Notice that  $\hat{p}_1$  results in a confidence interval that includes the population proportion,  $p$ . However,  $\hat{p}_2$  results in a confidence interval that does not include the population proportion,  $p$ , because  $\hat{p}_2$  is in the tails of the distribution.

**Question 1**

The horizontal axis in the sampling distribution of represents all possible sample proportions from a simple random sample of size  $n$ .

- What percent of sample proportions results in a 75% confidence interval that includes the population proportion?
- What percent of sample proportions results in a 75% confidence interval that does not include the population proportion?

**Solution::**



- 75%
- 25%

**Caution:**

95% confident does not mean 95% probability. Probability describes the likelihood of undetermined event.

It does not make sense to talk about the probability that the interval contains the parameter, because the parameter is an unknown, but fixed value. Thus, it is either in the interval, or not.

Consider a coin flip, if a coin has been flipped but its outcome has not been revealed, the flip resulting in a head is not  $\frac{1}{2}$ , because the outcome has already been determined. Instead the probability is either 0, or 1

Therefore, the probability that any confidence interval contains a parameter is either 0, or 1.

We do say "we are 95% confident the interval contains the parameter" because the method "works" in 95% of all samples.

### Constructing any $(1 - \alpha) \cdot 100\%$ Confidence interval

We need a method for constructing any  $(1 - \alpha) \cdot 100\%$  confidence interval. When  $\alpha = 0.05$ , we are constructing a 95% confidence interval.

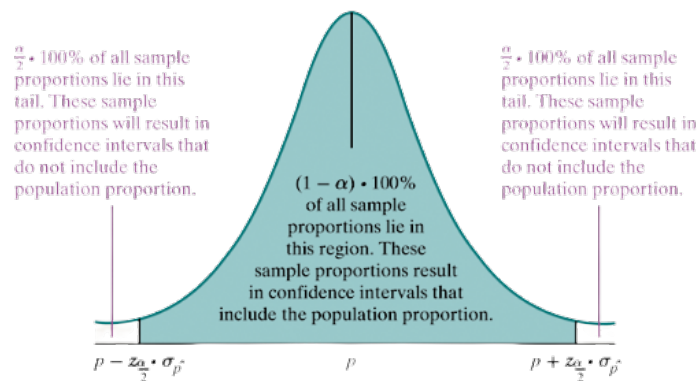
We generalize

$$p - 1.96\sigma_{\hat{p}} < \hat{p} < p + 1.96\sigma_{\hat{p}}$$

$$\text{parameter} - 1.96 \text{ standard error} < \text{point estimate} < \text{parameter} + 1.96 \text{ standard error}$$

by first noting that  $(1 - \alpha) \cdot 100\%$  of all sample proportions are in the interval as shown in Figure 2.

Figure 2



Rewrite this inequality with  $p$  in the middle and obtain

$$\hat{p} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} < p < \hat{p} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}.$$

So  $(1 - \alpha) \cdot 100\%$  of all sample proportions will result in confidence intervals that contain the population proportion. The sample proportions that are in the tails of the distribution in Figure 2 will not result in confidence intervals that contain the population proportion.



Table 1 shows some of the common critical values used in the construction of confidence intervals. Notice that higher levels of confidence correspond to higher critical values. After all, if your level of confidence that the interval includes the unknown parameter increases, the width of your interval (through the margin of error) should increase.

Level of confidence $(1 - \alpha) \cdot 100\%$	Area in each tail $\frac{\alpha}{2}$	Critical Value $z_{\frac{\alpha}{2}}$
90%	0.05	1.645
95%	0.025	1.96
99%	0.005	2.575

### Interpretation of a Confidence Interval

A  $(1 - \alpha) \cdot 100\%$  confidence interval indicates that  $(1 - \alpha) \cdot 100\%$  of all simple random samples of size  $n$  from the population whose parameter is unknown will result in an interval that contains the parameter.

For example, a 90% confidence interval for a parameter suggests that 90% of all possible samples will result in an interval that includes the unknown parameter and 10% of the samples will result in an interval that does not capture the parameter.

#### Example: Interpreting a Confidence Interval

**Problem:** The Gallup Organization conducted a poll in April 2017 in which a simple random sample of 1019 Americans aged 18 and older were asked, “Do you regard the income tax that you will have to pay this year as fair?” We learned from Example 1 that the proportion of those surveyed who responded yes was 0.608. Gallup reported its “survey methodology” as follows:

**Approach:** Confidence intervals for a proportion are of the form point estimate  $\pm$  margin of error. So add and subtract the margin of error from the point estimate to obtain the confidence interval. Interpret the confidence interval, “We are 95% confident that the proportion of Americans aged 18 and older who believe that the income tax they will have to pay this year is fair is between lower bound and upper bound.”

#### Solution:

The point estimate is 0.608, and the margin of error is 0.04. The confidence interval is  $0.608 \pm 0.04$ . Therefore, the lower bound of the confidence interval is  $0.608 - 0.04 = 0.568$  and the upper bound of the confidence interval is  $0.608 + 0.04 = 0.648$ . We are 95% confident that the proportion of Americans aged 18 and older who believe that the income tax they will have to pay this year is between 0.568 and 0.648.

We are now prepared to present a method for constructing a confidence interval about the population proportion,  $p$

### Constructing a $(1 - \alpha)$ 100% Confidence Interval for a Population Proportion:

Suppose that a simple random sample of size  $n$  is taken from a population or the data are the result of a randomized experiment. A  $(1 - \alpha)$  100% confidence interval for  $p$  is given by the following quantities:

$$\begin{aligned} \text{Lower bound} &: \hat{p} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \\ \text{Upper bound} &: \hat{p} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}. \end{aligned}$$

#### Note:-

It must be the case that  $n\hat{p}(1 - \hat{p}) \geq 10$  and  $n \leq 0.05N$  to construct this interval. Use  $\hat{p}$  in place of  $p$  in the standard deviation. This is because  $p$  is unknown, and  $\hat{p}$  is the best point estimate of  $p$ .

**Example: Constructing a Confidence Interval for a Population Proportion**

**Problem:** In the Parent–Teen Cell Phone Survey conducted by Princeton Survey Research Associates International, 800 randomly sampled 16- to 17-year-olds living in the United States were asked whether they have ever used their cell phone to text while driving. Of the 800 teenagers surveyed, 272 indicated that they text while driving. Obtain a 95% confidence interval for the proportion of 16- to 17-year-olds who text while driving.

**Approach:** It is important to verify the requirements for constructing the confidence interval first. It must be the case that  $n\hat{p}(1-\hat{p}) \geq 10$  and the sample size is no more than 5% of the population size ( $n \leq 0.05N$ ). Then, construct the confidence interval either by hand or using technology.

**Solution:**

First, we obtain  $\hat{p}$

$$\hat{p} = \frac{272}{800} = .34.$$

Verify Necessary Conditions:

$$\begin{aligned} n\hat{p}(1-\hat{p}) &\geq 10 \\ 800(.34)(.66) &= 179.52 \geq 10. \end{aligned}$$

Also, we know that  $n \leq 0.05N$ :

now in statcrunch:

1. Stat > Proportion Stats > One Sample > with summary
2. Input no. of successes
3. Input no. of observations
4. Input confidence level
5. Method: Standard-Wald
6. Compute

after doing the StatCrunch steps, we get a result of 0.307 and 0.373, therefore, we can conclude that we are 95% confident that the proportion of 16 to 17-year-olds who text while driving is between 0.307 and 0.373

## The Effect of Level of Confidence on the Margin of Error

### Example: The Role of the Level of Confidence in the Margin of Error

**Problem:** In the Parent–Teen Cell Phone Survey conducted by Princeton Survey Research Associates International, 800 randomly sampled 16- to 17-year-olds living in the United States were asked whether they have ever used their cell phone to text while driving. Of the 800 teenagers surveyed, 272 indicated that they text while driving. From the last example, we concluded that we are 95% confident that the proportion of 16- to 17-year-olds who text while driving is between 0.307 and 0.373. Determine the effect on the margin of error by increasing the level of confidence from 95% to 99%.

**Approach:** We would expect the margin of error to increase with a larger level of confidence. Construct the confidence interval either by hand or using technology.

#### Solution:

The margin of error for the 95% confidence interval found in Example 3 is 0.033, and the margin of error for the 99% confidence interval is 0.043. So increasing the level of confidence increases the margin of error, resulting in a wider confidence interval.

## The Effect of Sample Size on the Margin of Error

We know that larger sample sizes produce more precise estimates (the Law of Large Numbers). Given that the margin of error is  $E = z_{\alpha/2} \cdot \hat{p} \cdot \sqrt{\frac{(1-\hat{p})}{n}}$ , we can see that increasing the sample size  $n$  decreases the standard error, so the margin of error decreases. Therefore, larger sample sizes will result in narrower confidence intervals.

### What If We Do Not Satisfy the Normality Condition?

When the normality condition is not satisfied, the proportion of intervals that capture the parameter is below the level of confidence.

*Determine the Sample Size Necessary for Estimating a Population Proportion within a Specified Margin of Error*

**Example: Determining Sample Size**

**Problem:** An economist wants to know if the proportion of the U.S. population who commutes to work via car-pooling is on the rise. What size sample should be obtained if the economist wants an estimate within 2 percentage points of the true proportion with 90% confidence?

Assume that the economist uses the estimate of 10% obtained from the American Community Survey.

**Approach:** Since  $1 - \alpha = 0.9$ , we know that  $\alpha = 0.10$ . Use  $E = 0.02$ ,  $z_{\alpha/2} = z_{0.12} = z_{0.05} = 1.645$ , and  $\hat{p} = 0.10$  (the prior estimate).

**Solution:** Using the formula assuming that a prior estimate is available,  $n = \hat{p}(1 - \hat{p}) \left( \frac{z_{\alpha/2}}{E} \right)^2$ , we obtain

$$n = \hat{p}(1 - \hat{p}) \left( \frac{z_{\alpha/2}}{E} \right)^2 = 0.10 \times (1 - 0.10) \times \left( \frac{1.645}{0.02} \right)^2 = 608.9.$$

Round this value up to 609. So the economist must survey 609 randomly selected residents of the United States.

**statcrunch steps:**

First:

$$E = 0.02 \text{ (The 2 percentage points)}$$

$$W = 2(E) = 0.04$$

$$C.Level = 0.9$$

$$Target \ Proportion = 0.1.$$

1. Stats > Proportion Stats > One Sample > Width/Sample Size
2. Input confidence level
3. Input Target Proportion
4. Compute

And we find that the minimum sample size would be 609

Now consider that the economist does not use any prior estimates (target proportion), in this case, our target proportion will be 0.5

With this target proportion we get 1691