

8.4: A Population Proportion

During an election year, we see articles in the newspaper that state confidence intervals in terms of proportions or percentages. For example, a poll for a particular candidate running for president might show that the candidate has 40% of the vote within three percentage points (if the sample is large enough). Often, election polls are calculated with 95% confidence, so, the pollsters would be 95% confident that the true proportion of voters who favored the candidate would be between 0.37 and 0.43: $(0.40 - 0.03, 0.40 + 0.03)$.

Investors in the stock market are interested in the true proportion of stocks that go up and down each week. Businesses that sell personal computers are interested in the proportion of households in the United States that own personal computers. Confidence intervals can be calculated for the true proportion of stocks that go up or down each week and for the true proportion of households in the United States that own personal computers.

The procedure to find the confidence interval, the sample size, the error bound, and the confidence level for a proportion is similar to that for the population mean, but the formulas are different. How do you know you are dealing with a proportion problem? First, the underlying distribution is a binomial distribution. (There is no mention of a mean or average.) If X is a binomial random variable, then

$$X \sim B(n, p) \quad (8.4.1)$$

where n is the number of trials and p is the probability of a success.

To form a proportion, take X , the random variable for the number of successes and divide it by n , the number of trials (or the sample size). The random variable \hat{p} (read "p hat") is that proportion,

$$\hat{p} = \frac{X}{n} \quad (8.4.2)$$

When n is large and p is not close to zero or one, we can use the normal distribution to approximate the binomial.

$$X \sim N(np, \sqrt{npq}) \quad (8.4.3)$$

If we divide the random variable, the mean, and the standard deviation by n , we get a normal distribution of proportions with \hat{P} , called the estimated proportion, as the random variable. (Recall that a proportion is the number of successes divided by n .)

$$\frac{X}{n} = \hat{p} \sim N\left(\frac{np}{n}, \frac{\sqrt{npq}}{n}\right) \quad (8.4.4)$$

Using algebra to simplify:

$$\frac{\sqrt{npq}}{n} = \sqrt{\frac{pq}{n}} \quad (8.4.5)$$

\hat{p} follows a normal distribution for proportions:

$$\frac{X}{n} = \hat{p} \sim N\left(p, \sqrt{\frac{pq}{n}}\right) \quad (8.4.6)$$

The confidence interval has the form

$$(\hat{p} - E, \hat{p} + E). \quad (8.4.7)$$

where

- E is error bound for the proportion.
- $\hat{p} = \frac{x}{n}$
- \hat{p} = the estimated proportion of successes (\hat{p} is a point estimate for p , the true proportion.)
- x = the number of successes
- n = the size of the sample

The error bound (E) for a proportion is

$$E = \left(z_{\frac{\alpha}{2}} \right) \left(\sqrt{\frac{\hat{p}\hat{q}}{n}} \right) \quad (8.4.8)$$

where $\hat{q} = 1 - \hat{p}$.

This formula is similar to the error bound formula for a mean, except that the "appropriate standard deviation" is different. For a mean, when the population standard deviation is known, the appropriate standard deviation that we use is $\frac{\sigma}{\sqrt{n}}$. For a proportion, the appropriate standard deviation is

$$\sqrt{\frac{pq}{n}}. \quad (8.4.9)$$

However, in the error bound formula, we use

$$\sqrt{\frac{\hat{p}\hat{q}}{n}} \quad (8.4.10)$$

as the standard deviation, instead of

$$\sqrt{\frac{pq}{n}}. \quad (8.4.11)$$

In the error bound formula, the sample proportions (\hat{p} and \hat{q}) are estimates of the unknown population proportions p and q . The estimated proportions \hat{p} and \hat{q} are used because p and q are not known. The sample proportions \hat{p} and \hat{q} are calculated from the data: \hat{p} is the estimated proportion of successes, and \hat{q} is the estimated proportion of failures.

The confidence interval can be used only if the number of successes $n\hat{p}$ and the number of failures $n\hat{q}$ are both greater than five.

Normal Distribution of Proportions

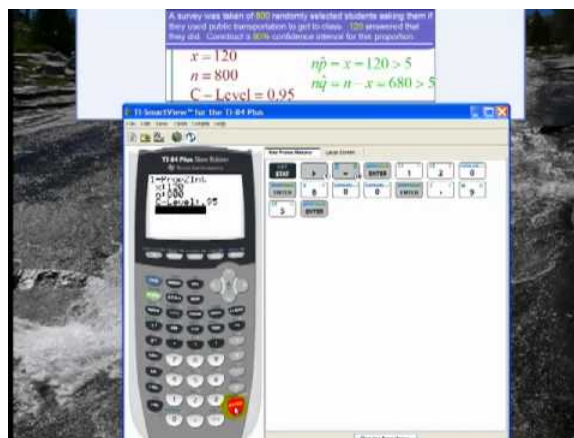
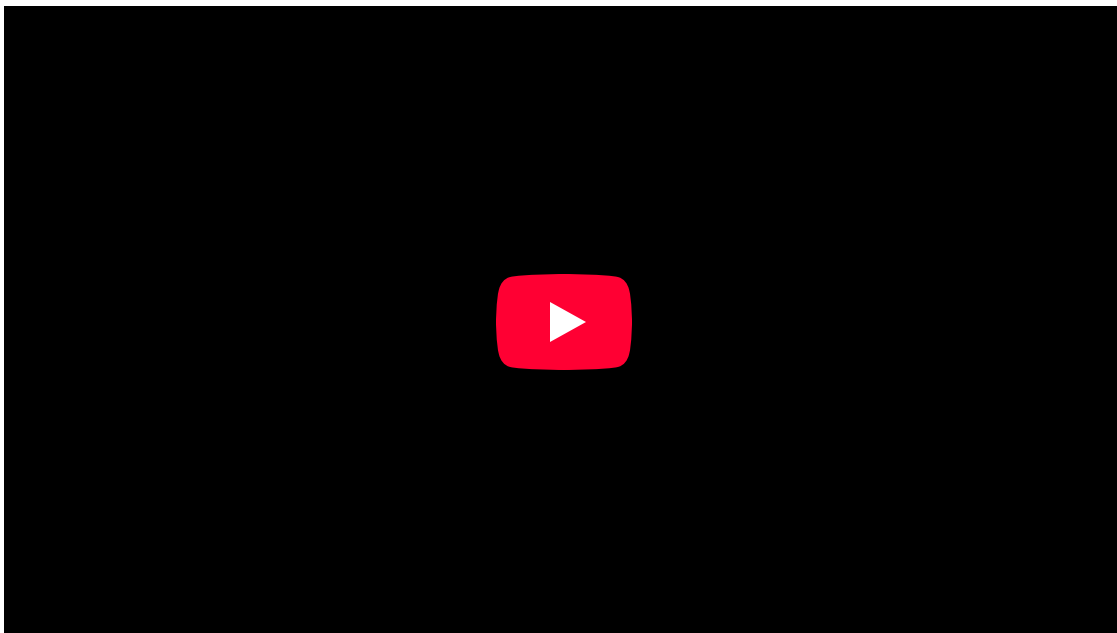
For the normal distribution of proportions, the z -score formula is as follows.

If

$$\hat{p} \sim N \left(p, \sqrt{\frac{pq}{n}} \right) \quad (8.4.12)$$

then the z -score formula is

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \quad (8.4.13)$$



CONFIDENCE INTERVAL for proportions Calculator

Fill in the sample size (n), the number of successes (x), and the confidence level (CL). Write the confidence level as a decimal. For example, for a 95% confidence level, enter 0.95 for CL. Then hit Calculate and assuming np and nq are large enough, the correct confidence interval will be calculated for you.

n:	x:	CL:	Calculate	Lower Bound	Upper Bound	\hat{p}
<input type="text"/>	<input type="text"/>	<input type="text"/>		<input type="text"/>	<input type="text"/>	<input type="text"/>

Example 8.4.1

Suppose that a market research firm is hired to estimate the percent of adults living in a large city who have cell phones. Five hundred randomly selected adult residents in this city are surveyed to determine whether they have cell phones. Of the 500 people surveyed, 421 responded yes - they own cell phones. Using a 95% confidence level, compute a confidence interval estimate for the true proportion of adult residents of this city who have cell phones.

Solution A

- The first solution is step-by-step (Solution A).

- The second solution uses a function of the TI-83, 83+ or 84 calculators (Solution B).

Let X = the number of people in the sample who have cell phones. X is binomial.

$$X \sim B(500, \frac{421}{500}). \quad (8.4.14)$$

To calculate the confidence interval, you must find \hat{p} , \hat{q} , and E .

- $n = 500$
- x = the number of successes = 421

$$\hat{p} = \frac{x}{n} = \frac{421}{500} = 0.842 \quad (8.4.15)$$

- $\hat{p} = 0.842$ is the sample proportion; this is the point estimate of the population proportion.

$$\hat{q} = 1 - \hat{p} = 1 - 0.842 = 0.158 \quad (8.4.16)$$

Since $CL = 0.95$, then

$$\alpha = 1 - CL = 1 - 0.95 = 0.05 \left(\frac{\alpha}{2} \right) = 0.025. \quad (8.4.17)$$

Then

$$z_{\frac{\alpha}{2}} = z_{0.025} = 1.96 \quad (8.4.18)$$

Use the TI-83, 83+, or 84+ calculator command `invNorm(0.975,0,1)` to find $z_{0.025}$. Remember that the area to the right of $z_{0.025}$ is 0.025 and the area to the left of $z_{0.025}$ is 0.975. This can also be found using appropriate commands on other calculators, using a computer, or using a Standard Normal probability table.

$$E = \left(z_{\frac{\alpha}{2}} \right) \sqrt{\frac{\hat{p}\hat{q}}{n}} = (1.96) \sqrt{\frac{(0.842)(0.158)}{500}} = 0.032 \quad (8.4.19)$$

$$\hat{p} - E = 0.842 - 0.032 = 0.81 \quad (8.4.20)$$

$$\hat{p} + E = 0.842 + 0.032 = 0.874 \quad (8.4.21)$$

The confidence interval for the true binomial population proportion is $(\hat{p} - E, \hat{p} + E) = (0.810, 0.874)$.

Interpretation

We estimate with 95% confidence that between 81% and 87.4% of all adult residents of this city have cell phones.

Explanation of 95% Confidence Level

Ninety-five percent of the confidence intervals constructed in this way would contain the true value for the population proportion of all adult residents of this city who have cell phones.

Solution B

Press **STAT** and arrow over to **TESTS** .

Arrow down to **A:1-PropZint** . Press **ENTER** .

Arrow down to **xx** and enter 421.

Arrow down to **nn** and enter 500.

Arrow down to **C-Level** and enter .95.

Arrow down to **Calculate** and press **ENTER** .

The confidence interval is (0.81003, 0.87397).

Exercise 8.4.1

Suppose 250 randomly selected people are surveyed to determine if they own a tablet. Of the 250 surveyed, 98 reported owning a tablet. Using a 95% confidence level, compute a confidence interval estimate for the true proportion of people who own

tablets.

Answer

(0.3315, 0.4525)

Example 8.4.2

For a class project, a political science student at a large university wants to estimate the percent of students who are registered voters. He surveys 500 students and finds that 300 are registered voters. Compute a 90% confidence interval for the true percent of students who are registered voters, and interpret the confidence interval.

Answer

- The first solution is step-by-step (Solution A).
- The second solution uses a function of the TI-83, 83+, or 84 calculators (Solution B).

Solution A

- $x = 300$ and
- $n = 500$

$$\hat{p} = \frac{x}{n} = \frac{300}{500} = 0.600 \quad (8.4.22)$$

$$\hat{q} = 1 - \hat{p} = 1 - 0.600 = 0.400 \quad (8.4.23)$$

Since $CL = 0.90$, then

$$\alpha = 1 - CL = 1 - 0.90 = 0.10 \left(\frac{\alpha}{2} \right) = 0.05 \quad (8.4.24)$$

$$z_{\frac{\alpha}{2}} = z_{0.05} = 1.645 \quad (8.4.25)$$

Use the TI-83, 83+, or 84+ calculator command `invNorm(0.95,0,1)` to find $z_{0.05}$. Remember that the area to the right of $z_{0.05}$ is 0.05 and the area to the left of $z_{0.05}$ is 0.95. This can also be found using appropriate commands on other calculators, using a computer, or using a standard normal probability table.

$$E = \left(z_{\frac{\alpha}{2}} \right) \sqrt{\frac{\hat{p}\hat{q}}{n}} = (1.645) \sqrt{\frac{(0.60)(0.40)}{500}} = 0.036 \quad (8.4.26)$$

$$\hat{p} - E = 0.60 - 0.036 = 0.564 \quad (8.4.27)$$

$$\hat{p} + E = 0.60 + 0.036 = 0.636 \quad (8.4.28)$$

The confidence interval for the true binomial population proportion is $(\hat{p} - E, \hat{p} + E) = (0.564, 0.636)$.

Interpretation

- We estimate with 90% confidence that the true percent of all students that are registered voters is between 56.4% and 63.6%.
- Alternate Wording: We estimate with 90% confidence that between 56.4% and 63.6% of ALL students are registered voters.

Explanation of 90% Confidence Level

Ninety percent of all confidence intervals constructed in this way contain the true value for the population percent of students that are registered voters.

Solution B

Press `STAT` and arrow over to `TESTS`.

Arrow down to `A:1-PropZint`. Press `ENTER`.

Arrow down to `xx` and enter 300.

Arrow down to `nn` and enter 500.

Arrow down to **C-Level** and enter 0.90.
 Arrow down to **Calculate** and press **ENTER** .
 The confidence interval is (0.564, 0.636).

Exercise 8.4.2

A student polls his school to see if students in the school district are for or against the new legislation regarding school uniforms. She surveys 600 students and finds that 480 are against the new legislation.

a. Compute a 90% confidence interval for the true percent of students who are against the new legislation, and interpret the confidence interval.

Answer

(0.7731, 0.8269); We estimate with 90% confidence that the true percent of all students in the district who are against the new legislation is between 77.31% and 82.69%.

b. In a sample of 300 students, 68% said they own an iPod and a smart phone. Compute a 97% confidence interval for the true percent of students who own an iPod and a smartphone.

Answer

Solution A

Sixty-eight percent (68%) of students own an iPod and a smart phone.

$$\hat{p} = 0.68 \quad (8.4.29)$$

$$\hat{q} = 1 - \hat{p} = 1 - 0.68 = 0.32 \quad (8.4.30)$$

Since $CL = 0.97$, we know

$$\alpha = 1 - 0.97 = 0.03 \quad (8.4.31)$$

and

$$\frac{\alpha}{2} = 0.015. \quad (8.4.32)$$

The area to the left of $z_{0.05}$ is 0.015, and the area to the right of $z_{0.05}$ is $1 - 0.015 = 0.985$.

Using the TI 83, 83+, or 84+ calculator function $\text{InvNorm}(0.985, 0, 1)$,

$$z_{0.05} = 2.17 \quad (8.4.33)$$

$$E = \left(z_{\frac{\alpha}{2}} \right) \sqrt{\frac{p'q'}{n}} = 2.17 \sqrt{\frac{0.68(0.32)}{300}} \approx 0.0269 \quad (8.4.34)$$

$$\hat{p} - E = 0.68 - 0.0269 = 0.6531 \quad (8.4.35)$$

$$\hat{p} + E = 0.68 + 0.0269 = 0.7069 \quad (8.4.36)$$

We are 97% confident that the true proportion of all students who own an iPod and a smart phone is between 0.6531 and 0.7069.

Solution B

Press **STAT** and arrow over to **TESTS**.

Arrow down to **A:1-PropZint**. Press **ENTER**.

Arrow down to **x** and enter $300 \cdot 0.68$.

Arrow down to **n** and enter 300.

Arrow down to **C-Level** and enter 0.97.

Arrow down to **Calculate** and press **ENTER**.

The confidence interval is (0.6531, 0.7069).

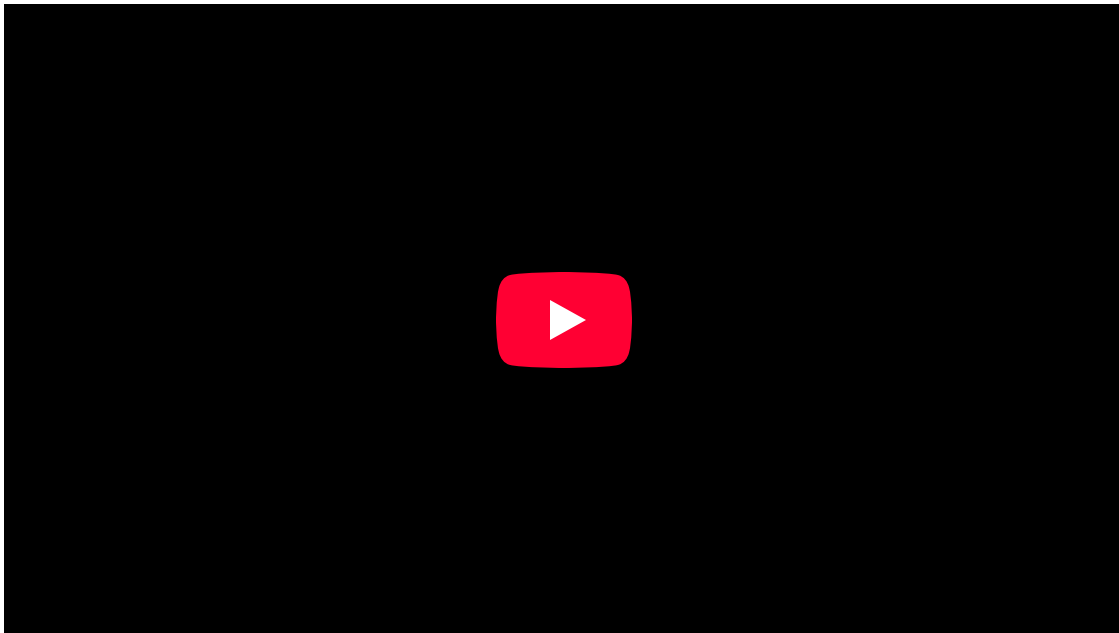
Calculating the Sample Size n

If researchers desire a specific margin of error, then they can use the error bound formula to calculate the required sample size. The error bound formula for a population proportion is

$$E = \left(z_{\frac{\alpha}{2}} \right) \left(\sqrt{\frac{\hat{p}\hat{q}}{n}} \right) \quad (8.4.37)$$

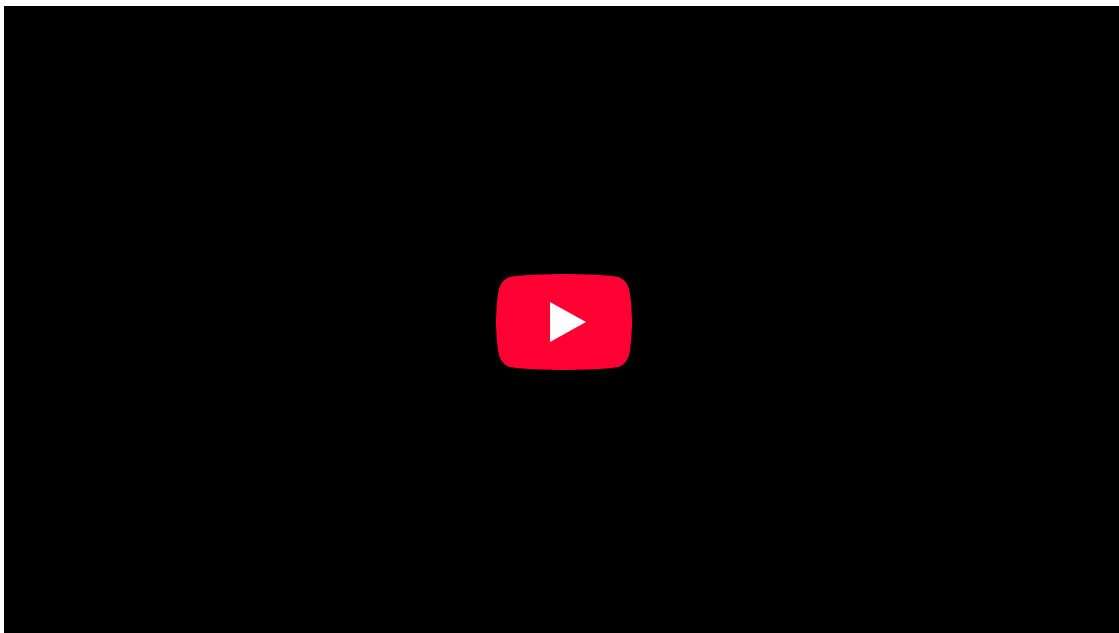
Solving for n gives you an equation for the sample size.

$$n = \frac{\left(z_{\frac{\alpha}{2}} \right)^2 (\hat{p}\hat{q})}{E^2} \quad (8.4.38)$$



32% of all Americans are obese. If you want to find a 95% confidence interval for the proportion of obese people in your city and you want an error bound of no more than 0.03, how many residents do you need to survey?

$$\begin{aligned} p &= 0.32 \\ z_{\alpha/2} &= 1.96 \\ EB &= 0.03 \end{aligned} \quad \begin{aligned} n &= p(1-p) \left(\frac{z_{\alpha/2}}{EB} \right)^2 \\ &= 0.32(1-0.32) \left(\frac{1.96}{0.03} \right)^2 \\ &= 928.81 \end{aligned}$$



If you want to obtain a 90% confidence interval for the proportion of all entering college students who do not know what to major in, and if you want an error bound of no more than 0.05, how many entering students should you survey?

$$z_{\alpha/2} = 1.645$$

$$EB = 0.05$$

Sample Size Calculator

Fill in the error bound (E) and the confidence level (CL) written as a decimal, for example a 95% confidence level is 0.95. If there is an estimate for the population proportion, check the "Have Estimate for p" radio button and then fill in the estimate in the p box that appears. Then hit "Calculate n" and the needed sample size will be calculated for you.

<input checked="" type="radio"/> No Estimate for p	<input type="radio"/> Have Estimate for p	Estimate for p:	E:	CL:	Calculate n	n:
			<input type="text"/>	<input type="text"/>		<input type="text"/>

Estimate for p:

Example 8.4.5

Suppose a mobile phone company wants to determine the current percentage of customers aged 50+ who use text messaging on their cell phones. How many customers aged 50+ should the company survey in order to be 90% confident that the estimated

(sample) proportion is within three percentage points of the true population proportion of customers aged 50+ who use text messaging on their cell phones.

Answer

From the problem, we know that $E = 0.03$ (3%=0.03) and $z_{\alpha/2} = z_{0.05} = 1.645$ because the confidence level is 90%.

However, in order to find n , we need to know the estimated (sample) proportion \hat{p} . Remember that $\hat{q} = 1 - \hat{p}$. But, we do not know \hat{p} yet. Since we multiply \hat{p} and \hat{q} together, we make them both equal to 0.5 because $\hat{p}\hat{q} = (0.5)(0.5) = 0.25$ results in the largest possible product. (Try other products: $(0.6)(0.4) = 0.24$; $(0.3)(0.7) = 0.21$; $(0.2)(0.8) = 0.16$ and so on). The largest possible product gives us the largest n . This gives us a large enough sample so that we can be 90% confident that we are within three percentage points of the true population proportion. To calculate the sample size n , use the formula and make the substitutions.

$$n = \frac{z^2 \hat{p} \hat{q}}{E^2} \quad (8.4.39)$$

gives

$$n = \frac{1.645^2 (0.5)(0.5)}{0.03^2} = 751.7 \quad (8.4.40)$$

Round the answer to the next higher value. The sample size should be 752 cell phone customers aged 50+ in order to be 90% confident that the estimated (sample) proportion is within three percentage points of the true population proportion of all customers aged 50+ who use text messaging on their cell phones.

Exercise 8.4.5

Suppose an internet marketing company wants to determine the current percentage of customers who click on ads on their smartphones. How many customers should the company survey in order to be 90% confident that the estimated proportion is within five percentage points of the true population proportion of customers who click on ads on their smartphones?

Answer

271 customers should be surveyed. Check the Real Estate section in your local

Glossary

Binomial Distribution

a discrete random variable (RV) which arises from Bernoulli trials; there are a fixed number, n , of independent trials.

“Independent” means that the result of any trial (for example, trial 1) does not affect the results of the following trials, and all trials are conducted under the same conditions. Under these circumstances the binomial RV X is defined as the number of successes in n trials. The notation is: $X \sim B(n, p)$. The mean is $\mu = np$ and the standard deviation is $\sigma = \sqrt{npq}$. The probability of exactly x successes in n trials is $P(X = x) = \binom{n}{x} p^x q^{n-x}$.

Error Bound for a Population Proportion (E)

the margin of error; depends on the confidence level, the sample size, and the estimated (from the sample) proportion of successes.

Contributors

- Barbara Illowsky and Susan Dean (De Anza College) with many other contributing authors. Content produced by OpenStax College is licensed under a Creative Commons Attribution License 4.0 license. Download for free at <http://cnx.org/contents/30189442-699...b91b9de@18.114>.

This page titled [8.4: A Population Proportion](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [OpenStax](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.