



A Population Proportion

By:
OpenStaxCollege

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 $X \sim B(n, p)$ 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 P' (read "P prime") is that proportion,

$$P' = \frac{X}{n}$$

(Sometimes the random variable is denoted as \hat{P} , read "P hat".)

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

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

A Population Proportion

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

$$\frac{X}{n} = P' \sim N\left(\frac{np}{n}, \frac{\sqrt{npq}}{n}\right)$$

Using algebra to simplify: $\frac{\sqrt{npq}}{n} = \sqrt{\frac{pq}{n}}$

P' follows a normal distribution for proportions: $\frac{X}{n} = P' \sim N\left(\frac{np}{n}, \frac{\sqrt{npq}}{n}\right)$

The confidence interval has the form $(p' - EBP, p' + EBP)$. EBP is error bound for the proportion.

$$p' = \frac{x}{n}$$

p' = the **estimated proportion** of successes (p' is a **point estimate** for p , the true proportion.)

x = the **number** of successes

n = the size of the sample

The error bound for a proportion is

$$EBP = \left(z_{\frac{\alpha}{2}}\right)\left(\sqrt{\frac{p'q'}{n}}\right) \text{ where } q' = 1 - p'$$

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 $\sqrt{\frac{pq}{n}}$.

However, in the error bound formula, we use $\sqrt{\frac{p'q'}{n}}$ as the standard deviation, instead of $\sqrt{\frac{pq}{n}}$.

In the error bound formula, the **sample proportions p' and q' are estimates of the unknown population proportions p and q** . The estimated proportions p' and q' are used because p and q are not known. The sample proportions p' and q' are calculated from the data: p' is the estimated proportion of successes, and q' is the estimated proportion of failures.

A Population Proportion

The confidence interval can be used only if the number of successes np' and the number of failures nq' are both greater than five.

Note

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

If $P' \sim N(p, \sqrt{\frac{pq}{n}})$ then the z -score formula is $z = \frac{p' - p}{\sqrt{\frac{pq}{n}}}$

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

To calculate the confidence interval, you must find p' , q' , and EBP .

$$n = 500$$

$$x = \text{the number of successes} = 421$$

$$p' = \frac{x}{n} = \frac{421}{500} = 0.842$$

$p' = 0.842$ is the sample proportion; this is the point estimate of the population proportion.

$$q' = 1 - p' = 1 - 0.842 = 0.158$$

Since $CL = 0.95$, then $\alpha = 1 - CL = 1 - 0.95 = 0.05$ ($\frac{\alpha}{2}$) = 0.025.

Then $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$

A Population Proportion

Use the TI-83, 83+, or 84+ calculator command $\text{invNorm}(0.975,0,1)$ 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.

$$EBP = \left(z_{\frac{\alpha}{2}}\right)\sqrt{\frac{p'q'}{n}} = (1.96)\sqrt{\frac{(0.842)(0.158)}{500}} = 0.032$$

$$p' - EBP = 0.842 - 0.032 = 0.81$$

$$p' + EBP = 0.842 + 0.032 = 0.874$$

The confidence interval for the true binomial population proportion is $(p' - EBP, p' + EBP) = (0.810, 0.874)$.

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

Explanation of 95% Confidence LevelNinety-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 **x** and enter 421.

Arrow down to **n** 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).

Try It

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.

(0.3315, 0.4525)

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

A Population Proportion

are registered voters. Compute a 90% confidence interval for the true percent of students who are registered voters, and interpret the confidence interval.

- 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 $n = 500$

$$p' = \frac{x}{n} = \frac{300}{500} = 0.600$$

$$q' = 1 - p' = 1 - 0.600 = 0.400$$

Since $CL = 0.90$, then $\alpha = 1 - CL = 1 - 0.90 = 0.10\left(\frac{\alpha}{2}\right) = 0.05$

$$z_{\frac{\alpha}{2}} = z_{0.05} = 1.645$$

Use the TI-83, 83+, or 84+ calculator command $\text{invNorm}(0.95,0,1)$ 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.

$$EBP = \left(z_{\frac{\alpha}{2}}\right) \sqrt{\frac{p'q'}{n}} = (1.645) \sqrt{\frac{(0.60)(0.40)}{500}} = 0.036$$

$$p' - EBP = 0.60 - 0.036 = 0.564$$

$$p' + EBP = 0.60 + 0.036 = 0.636$$

The confidence interval for the true binomial population proportion is $(p' - EBP, p' + EBP) = (0.564, 0.636)$.

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.

A Population Proportion

Solution B

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

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

Arrow down to x and enter 300.

Arrow down to n 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).

Try It

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.

(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.

Solution A

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

$$p' = 0.68$$

$$q' = 1 - p' = 1 - 0.68 = 0.32$$

Since $CL = 0.97$, we know $\alpha = 1 - 0.97 = 0.03$ and $\frac{\alpha}{2} = 0.015$.

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

Using the TI 83, 83+, or 84+ calculator function $\text{InvNorm}(.985, 0, 1)$,

$$z_{0.015} = 2.17$$

A Population Proportion

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

$$p' - EPB = 0.68 - 0.0269 = 0.6531$$

$$p' + EPB = 0.68 + 0.0269 = 0.7069$$

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*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).

“Plus Four” Confidence Interval for p

There is a certain amount of error introduced into the process of calculating a confidence interval for a proportion. Because we do not know the true proportion for the population, we are forced to use point estimates to calculate the appropriate standard deviation of the sampling distribution. Studies have shown that the resulting estimation of the standard deviation can be flawed.

Fortunately, there is a simple adjustment that allows us to produce more accurate confidence intervals. We simply pretend that we have four additional observations. Two of these observations are successes and two are failures. The new sample size, then, is $n + 4$, and the new count of successes is $x + 2$.

Computer studies have demonstrated the effectiveness of this method. It should be used when the confidence level desired is at least 90% and the sample size is at least ten.

A random sample of 25 statistics students was asked: “Have you smoked a cigarette in the past week?” Six students reported smoking within the past week. Use the “plus-four” method to find a 95% confidence interval for the true proportion of statistics students who smoke.

Solution A

A Population Proportion

Solution A Six students out of 25 reported smoking within the past week, so $x = 6$ and $n = 25$. Because we are using the “plus-four” method, we will use $x = 6 + 2 = 8$ and $n = 25 + 4 = 29$.

$$p' = \frac{x}{n} = \frac{8}{29} \approx 0.276$$

$$q' = 1 - p' = 1 - 0.276 = 0.724$$

Since $CL = 0.95$, we know $\alpha = 1 - 0.95 = 0.05$ and $\frac{\alpha}{2} = 0.025$.

$$z_{0.025} = 1.96$$

$$EPB = \left(z_{\frac{\alpha}{2}}\right) \sqrt{\frac{p'q'}{n}} = (1.96) \sqrt{\frac{0.276(0.724)}{29}} \approx 0.163$$

$$p' - EPB = 0.276 - 0.163 = 0.113$$

$$p' + EPB = 0.276 + 0.163 = 0.439$$

We are 95% confident that the true proportion of all statistics students who smoke cigarettes is between 0.113 and 0.439.

Solution B

Press STAT and arrow over to TESTS.
Arrow down to A:1-PropZint. Press ENTER.

Reminder

Remember that the plus-four method assume an additional four trials: two successes and two failures. You do not need to change the process for calculating the confidence interval; simply update the values of x and n to reflect these additional trials.

Arrow down to x and enter eight.
Arrow down to n and enter 29.
Arrow down to C-Level and enter 0.95.
Arrow down to Calculate and press ENTER.
The confidence interval is (0.113, 0.439).

Try It

A Population Proportion

Out of a random sample of 65 freshmen at State University, 31 students have declared a major. Use the “plus-four” method to find a 96% confidence interval for the true proportion of freshmen at State University who have declared a major.

Solution A

Using “plus four,” we have $x = 31 + 2 = 33$ and $n = 65 + 4 = 69$.

$$p' = \frac{33}{69} \approx 0.478$$

$$q' = 1 - p' = 1 - 0.478 = 0.522$$

Since $CL = 0.96$, we know $\alpha = 1 - 0.96 = 0.04$ and $\frac{\alpha}{2} = 0.02$.

$$z_{0.02} = 2.054$$

$$EPB = \left(z_{\frac{\alpha}{2}}\right) \sqrt{\frac{p'q'}{n}} = (2.054) \left(\sqrt{\frac{(0.478)(0.522)}{69}}\right) \approx 0.124$$

$$p' - EPB = 0.478 - 0.124 = 0.354$$

$$p' + EPB = 0.478 + 0.124 = 0.602$$

We are 96% confident that between 35.4% and 60.2% of all freshmen at State U have declared a major.

Solution B

Press STAT and arrow over to TESTS.

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

Arrow down to x and enter 33.

Arrow down to n and enter 69.

Arrow down to C-Level and enter 0.96.

Arrow down to Calculate and press ENTER.

The confidence interval is (0.355, 0.602).

The Berkman Center for Internet & Society at Harvard recently conducted a study analyzing the privacy management habits of teen internet users. In a group of 50 teens, 13 reported having more than 500 friends on Facebook. Use the “plus four” method to find a 90% confidence interval for the true proportion of teens who would report having more than 500 Facebook friends.

Solution A Using “plus-four,” we have $x = 13 + 2 = 15$ and $n = 50 + 4 = 54$.

A Population Proportion

$$p' = \frac{15}{54} \approx 0.278$$

$$q' = 1 - p' = 1 - 0.278 = 0.722$$

Since $CL = 0.90$, we know $\alpha = 1 - 0.90 = 0.10$ and $\frac{\alpha}{2} = 0.05$.

$$z_{0.05} = 1.645$$

$$EPB = (z_{\frac{\alpha}{2}}) \left(\sqrt{\frac{p'q'}{n}} \right) = (1.645) \left(\sqrt{\frac{(0.278)(0.722)}{54}} \right) \approx 0.100$$

$$p' - EPB = 0.278 - 0.100 = 0.178$$

$$p' + EPB = 0.278 + 0.100 = 0.378$$

We are 90% confident that between 17.8% and 37.8% of all teens would report having more than 500 friends on Facebook.

Solution B

Press STAT and arrow over to TESTS.

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

Arrow down to x and enter 15.

Arrow down to n and enter 54.

Arrow down to C-Level and enter 0.90.

Arrow down to Calculate and press ENTER.

The confidence interval is (0.178, 0.378).

Try It

The Berkman Center Study referenced in [\[link\]](#) talked to teens in smaller focus groups, but also interviewed additional teens over the phone. When the study was complete, 588 teens had answered the question about their Facebook friends with 159 saying that they have more than 500 friends. Use the “plus-four” method to find a 90% confidence interval for the true proportion of teens that would report having more than 500 Facebook friends based on this larger sample. Compare the results to those in [\[link\]](#).

Solution A

Using “plus-four,” we have $x = 159 + 2 = 161$ and $n = 588 + 4 = 592$.

$$p' = \frac{161}{592} \approx 0.272$$

A Population Proportion

$$q' = 1 - p' = 1 - 0.272 = 0.728$$

Since $CL = 0.90$, we know $\alpha = 1 - 0.90 = 0.10$ and $\frac{\alpha}{2} = 0.05$

$$EPB = \left(z_{\frac{\alpha}{2}}\right) \left(\sqrt{\frac{p'q'}{n}}\right) = (1.645) \left(\sqrt{\frac{(0.272)(0.728)}{592}}\right) \approx 0.030$$

$$p' - EPB = 0.272 - 0.030 = 0.242$$

$$p' + EPB = 0.272 + 0.030 = 0.302$$

We are 90% confident that between 24.2% and 30.2% of all teens would report having more than 500 friends on Facebook.

Solution B

Press STAT and arrow over to TESTS.

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

Arrow down to x and enter 161.

Arrow down to n and enter 592.

Arrow down to C-Level and enter 0.90.

Arrow down to Calculate and press ENTER.

The confidence interval is (0.242, 0.302).

Conclusion: The confidence interval for the larger sample is narrower than the interval from [\[link\]](#). Larger samples will always yield more precise confidence intervals than smaller samples. The “plus four” method has a greater impact on the smaller sample. It shifts the point estimate from 0.26 (13/50) to 0.278 (15/54). It has a smaller impact on the EPB , changing it from 0.102 to 0.100. In the larger sample, the point estimate undergoes a smaller shift: from 0.270 (159/588) to 0.272 (161/592). It is easy to see that the plus-four method has the greatest impact on smaller samples.

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.

A Population Proportion

The error bound formula for a population proportion is

- $EBP = \left(z_{\frac{\alpha}{2}} \right) \left(\sqrt{\frac{p'q'}{n}} \right)$
- Solving for n gives you an equation for the sample size.
- $n = \frac{\left(z_{\frac{\alpha}{2}} \right)^2 (p'q')}{EBP^2}$

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.

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

However, in order to find n , we need to know the estimated (sample) proportion p' . Remember that $q' = 1 - p'$. But, we do not know p' yet. Since we multiply p' and q' together, we make them both equal to 0.5 because $p'q' = (0.5)(0.5) = 0.25$ 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.

$$n = \frac{z_{\frac{\alpha}{2}}^2 p'q'}{EBP^2} \text{ gives } n = \frac{1.645^2(0.5)(0.5)}{0.03^2} = 751.7$$

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.

Try It

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?

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

References

Jensen, Tom. "Democrats, Republicans Divided on Opinion of Music Icons." Public Policy Polling. Available online at <http://www.publicpolicypolling.com/Day2MusicPoll.pdf> (accessed July 2, 2013).

Madden, Mary, Amanda Lenhart, Sandra Coresi, Urs Gasser, Maeve Duggan, Aaron Smith, and Meredith Beaton. "Teens, Social Media, and Privacy." PewInternet, 2013. Available online at <http://www.pewinternet.org/Reports/2013/Teens-Social-Media-And-Privacy.aspx> (accessed July 2, 2013).

Prince Survey Research Associates International. "2013 Teen and Privacy Management Survey." Pew Research Center: Internet and American Life Project. Available online at http://www.pewinternet.org/~media/Files/Questionnaire/2013/Methods%20and%20Questions_Teens%20and%20Social%20Media.pdf (accessed July 2, 2013).

Saad, Lydia. "Three in Four U.S. Workers Plan to Work Pas Retirement Age: Slightly more say they will do this by choice rather than necessity." Gallup® Economy, 2013. Available online at <http://www.gallup.com/poll/162758/three-four-workers-plan-work-past-retirement-age.aspx> (accessed July 2, 2013).

The Field Poll. Available online at <http://field.com/fieldpollonline/subscribers/> (accessed July 2, 2013).

Zogby. "New SUNYIT/Zogby Analytics Poll: Few Americans Worry about Emergency Situations Occurring in Their Community; Only one in three have an Emergency Plan; 70% Support Infrastructure 'Investment' for National Security." Zogby Analytics, 2013. Available online at <http://www.zogbyanalytics.com/news/299-americans-neither-worried-nor-prepared-in-case-of-a-disaster-sunyit-zogby-analytics-poll> (accessed July 2, 2013).

"52% Say Big-Time College Athletics Corrupt Education Process." Rasmussen Reports, 2013. Available online at http://www.rasmussenreports.com/public_content/lifestyle/sports/may_2013/52_say_big_time_college_athletics_corrupt_education_process (accessed July 2, 2013).

Chapter Review

Some statistical measures, like many survey questions, measure qualitative rather than quantitative data. In this case, the population parameter being estimated is a proportion.

A Population Proportion

It is possible to create a confidence interval for the true population proportion following procedures similar to those used in creating confidence intervals for population means. The formulas are slightly different, but they follow the same reasoning.

Let p' represent the sample proportion, x/n , where x represents the number of successes and n represents the sample size. Let $q' = 1 - p'$. Then the confidence interval for a population proportion is given by the following formula:

$$(\text{lower bound, upper bound}) = (p' - EBP, p' + EBP) = \left(p' - z\sqrt{\frac{p'q'}{n}}, p' + z\sqrt{\frac{p'q'}{n}} \right)$$

The “plus four” method for calculating confidence intervals is an attempt to balance the error introduced by using estimates of the population proportion when calculating the standard deviation of the sampling distribution. Simply imagine four additional trials in the study; two are successes and two are failures. Calculate $p' = \frac{x+2}{n+4}$, and proceed to find the confidence interval. When sample sizes are small, this method has been demonstrated to provide more accurate confidence intervals than the standard formula used for larger samples.

Formula Review

$p' = x / n$ where x represents the number of successes and n represents the sample size. The variable p' is the sample proportion and serves as the point estimate for the true population proportion.

$$q' = 1 - p'$$

$p' \sim N\left(p, \sqrt{\frac{pq}{n}}\right)$ The variable p' has a binomial distribution that can be approximated with the normal distribution shown here.

$$EBP = \text{the error bound for a proportion} = z_{\frac{\alpha}{2}} \sqrt{\frac{p'q'}{n}}$$

Confidence interval for a proportion:

$$(\text{lower bound, upper bound}) = (p' - EBP, p' + EBP) = \left(p' - z\sqrt{\frac{p'q'}{n}}, p' + z\sqrt{\frac{p'q'}{n}} \right)$$

$n = \frac{z_{\frac{\alpha}{2}}^2 p' q'}{EBP^2}$ provides the number of participants needed to estimate the population proportion with confidence $1 - \alpha$ and margin of error EBP .

A Population Proportion

Use the normal distribution for a single population proportion $p' = \frac{x}{n}$

$$EBP = \left(z_{\frac{\alpha}{2}}\right) \sqrt{\frac{p'q'}{n}} \quad p' + q' = 1$$

The confidence interval has the format $(p' - EBP, p' + EBP)$.

\bar{x} is a point estimate for μ

p' is a point estimate for p

s is a point estimate for σ

Use the following information to answer the next two exercises: Marketing companies are interested in knowing the population percent of women who make the majority of household purchasing decisions.

When designing a study to determine this population proportion, what is the minimum number you would need to survey to be 90% confident that the population proportion is estimated to within 0.05?

If it were later determined that it was important to be more than 90% confident and a new survey were commissioned, how would it affect the minimum number you need to survey? Why?

It would decrease, because the z-score would decrease, which reducing the numerator and lowering the number.

Use the following information to answer the next five exercises: Suppose the marketing company did do a survey. They randomly surveyed 200 households and found that in 120 of them, the woman made the majority of the purchasing decisions. We are interested in the population proportion of households where women make the majority of the purchasing decisions.

Identify the following:

1. $x =$ _____
2. $n =$ _____
3. $p' =$ _____

Define the random variables X and P' in words.

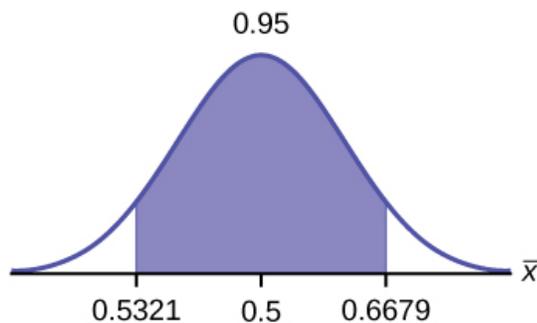
A Population Proportion

X is the number of “successes” where the woman makes the majority of the purchasing decisions for the household. P' is the percentage of households sampled where the woman makes the majority of the purchasing decisions for the household.

Which distribution should you use for this problem?

Construct a 95% confidence interval for the population proportion of households where the women make the majority of the purchasing decisions. State the confidence interval, sketch the graph, and calculate the error bound.

CI: (0.5321, 0.6679)



EBM: 0.0679

List two difficulties the company might have in obtaining random results, if this survey were done by email.

Use the following information to answer the next five exercises: Of 1,050 randomly selected adults, 360 identified themselves as manual laborers, 280 identified themselves as non-manual wage earners, 250 identified themselves as mid-level managers, and 160 identified themselves as executives. In the survey, 82% of manual laborers preferred trucks, 62% of non-manual wage earners preferred trucks, 54% of mid-level managers preferred trucks, and 26% of executives preferred trucks.

We are interested in finding the 95% confidence interval for the percent of executives who prefer trucks. Define random variables X and P' in words.

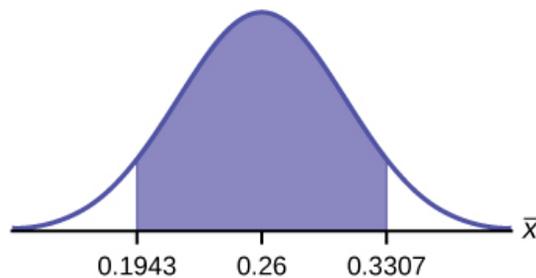
X is the number of “successes” where an executive prefers a truck. P' is the percentage of executives sampled who prefer a truck.

Which distribution should you use for this problem?

A Population Proportion

Construct a 95% confidence interval. State the confidence interval, sketch the graph, and calculate the error bound.

CI: (0.19432, 0.33068)



EBM: 0.0707

Suppose we want to lower the sampling error. What is one way to accomplish that?

The sampling error given in the survey is $\pm 2\%$. Explain what the $\pm 2\%$ means.

The sampling error means that the true mean can be 2% above or below the sample mean.

Use the following information to answer the next five exercises: A poll of 1,200 voters asked what the most significant issue was in the upcoming election. Sixty-five percent answered the economy. We are interested in the population proportion of voters who feel the economy is the most important.

Define the random variable X in words.

Define the random variable P' in words.

P' is the proportion of voters sampled who said the economy is the most important issue in the upcoming election.

Which distribution should you use for this problem?

Construct a 90% confidence interval, and state the confidence interval and the error bound.

CI: (0.62735, 0.67265)

EBM: 0.02265

A Population Proportion

What would happen to the confidence interval if the level of confidence were 95%?

Use the following information to answer the next 16 exercises: The Ice Chalet offers dozens of different beginning ice-skating classes. All of the class names are put into a bucket. The 5 P.M., Monday night, ages 8 to 12, beginning ice-skating class was picked. In that class were 64 girls and 16 boys. Suppose that we are interested in the true proportion of girls, ages 8 to 12, in all beginning ice-skating classes at the Ice Chalet. Assume that the children in the selected class are a random sample of the population.

What is being counted?

The number of girls, ages 8 to 12, in the 5 P.M. Monday night beginning ice-skating class.

In words, define the random variable X .

Calculate the following:

1. $x =$ _____
2. $n =$ _____
3. $p' =$ _____

1. $x = 64$
2. $n = 80$
3. $p' = 0.8$

State the estimated distribution of X . $X \sim$ _____

Define a new random variable P' . What is p' estimating?

p

In words, define the random variable P' .

State the estimated distribution of P' . Construct a 92% Confidence Interval for the true proportion of girls in the ages 8 to 12 beginning ice-skating classes at the Ice Chalet.

$$P' \sim N\left(0.8, \sqrt{\frac{(0.8)(0.2)}{80}}\right). (0.72171, 0.87829).$$

How much area is in both tails (combined)?

How much area is in each tail?

A Population Proportion

0.04

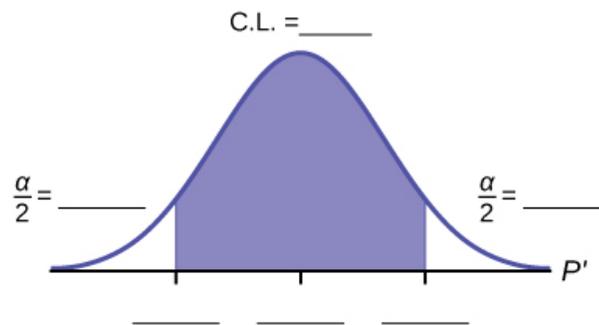
Calculate the following:

1. lower limit
2. upper limit
3. error bound

The 92% confidence interval is _____.

(0.72; 0.88)

Fill in the blanks on the graph with the areas, upper and lower limits of the confidence interval, and the sample proportion.



In one complete sentence, explain what the interval means.

With 92% confidence, we estimate the proportion of girls, ages 8 to 12, in a beginning ice-skating class at the Ice Chalet to be between 72% and 88%.

Using the same p' and level of confidence, suppose that n were increased to 100. Would the error bound become larger or smaller? How do you know?

Using the same p' and $n = 80$, how would the error bound change if the confidence level were increased to 98%? Why?

The error bound would increase. Assuming all other variables are kept constant, as the confidence level increases, the area under the curve corresponding to the confidence level becomes larger, which creates a wider interval and thus a larger error.

If you decreased the allowable error bound, why would the minimum sample size increase (keeping the same level of confidence)?

Homework

Insurance companies are interested in knowing the population percent of drivers who always buckle up before riding in a car.

1. When designing a study to determine this population proportion, what is the minimum number you would need to survey to be 95% confident that the population proportion is estimated to within 0.03?
2. If it were later determined that it was important to be more than 95% confident and a new survey was commissioned, how would that affect the minimum number you would need to survey? Why?
 1. 1,068
 2. The sample size would need to be increased since the critical value increases as the confidence level increases.

Suppose that the insurance companies did do a survey. They randomly surveyed 400 drivers and found that 320 claimed they always buckle up. We are interested in the population proportion of drivers who claim they always buckle up.

1.
 1. $x =$ _____
 2. $n =$ _____
 3. $p' =$ _____
2. Define the random variables X and P' , in words.
3. Which distribution should you use for this problem? Explain your choice.
4. Construct a 95% confidence interval for the population proportion who claim they always buckle up.
 1. State the confidence interval.
 2. Sketch the graph.
 3. Calculate the error bound.
5. If this survey were done by telephone, list three difficulties the companies might have in obtaining random results.

According to a recent survey of 1,200 people, 61% feel that the president is doing an acceptable job. We are interested in the population proportion of people who feel the president is doing an acceptable job.

1. Define the random variables X and P' in words.
2. Which distribution should you use for this problem? Explain your choice.
3. Construct a 90% confidence interval for the population proportion of people who feel the president is doing an acceptable job.
 1. State the confidence interval.
 2. Sketch the graph.
 3. Calculate the error bound.

A Population Proportion

1. X = the number of people who feel that the president is doing an acceptable job;
 P' = the proportion of people in a sample who feel that the president is doing an acceptable job.
2. $N(0.61, \sqrt{\frac{(0.61)(0.39)}{1200}})$
3.
 1. CI: (0.59, 0.63)
 2. Check student's solution
 3. *EBM*: 0.02

An article regarding interracial dating and marriage recently appeared in the Washington Post. Of the 1,709 randomly selected adults, 315 identified themselves as Latinos, 323 identified themselves as blacks, 254 identified themselves as Asians, and 779 identified themselves as whites. In this survey, 86% of blacks said that they would welcome a white person into their families. Among Asians, 77% would welcome a white person into their families, 71% would welcome a Latino, and 66% would welcome a black person.

1. We are interested in finding the 95% confidence interval for the percent of all black adults who would welcome a white person into their families. Define the random variables X and P' , in words.
2. Which distribution should you use for this problem? Explain your choice.
3. Construct a 95% confidence interval.
 1. State the confidence interval.
 2. Sketch the graph.
 3. Calculate the error bound.

Refer to the information in [\[link\]](#).

1. Construct three 95% confidence intervals.
 1. percent of all Asians who would welcome a white person into their families.
 2. percent of all Asians who would welcome a Latino into their families.
 3. percent of all Asians who would welcome a black person into their families.
2. Even though the three point estimates are different, do any of the confidence intervals overlap? Which?
3. For any intervals that do overlap, in words, what does this imply about the significance of the differences in the true proportions?
4. For any intervals that do not overlap, in words, what does this imply about the significance of the differences in the true proportions?
 1. (0.72, 0.82)
 2. (0.65, 0.76)

3. (0.60, 0.72)
2. Yes, the intervals (0.72, 0.82) and (0.65, 0.76) overlap, and the intervals (0.65, 0.76) and (0.60, 0.72) overlap.
3. We can say that there does not appear to be a significant difference between the proportion of Asian adults who say that their families would welcome a white person into their families and the proportion of Asian adults who say that their families would welcome a Latino person into their families.
4. We can say that there is a significant difference between the proportion of Asian adults who say that their families would welcome a white person into their families and the proportion of Asian adults who say that their families would welcome a black person into their families.

Stanford University conducted a study of whether running is healthy for men and women over age 50. During the first eight years of the study, 1.5% of the 451 members of the 50-Plus Fitness Association died. We are interested in the proportion of people over 50 who ran and died in the same eight-year period.

1. Define the random variables X and P' in words.
2. Which distribution should you use for this problem? Explain your choice.
3. Construct a 97% confidence interval for the population proportion of people over 50 who ran and died in the same eight-year period.
 1. State the confidence interval.
 2. Sketch the graph.
 3. Calculate the error bound.
4. Explain what a “97% confidence interval” means for this study.

A telephone poll of 1,000 adult Americans was reported in an issue of Time Magazine. One of the questions asked was “What is the main problem facing the country?” Twenty percent answered “crime.” We are interested in the population proportion of adult Americans who feel that crime is the main problem.

1. Define the random variables X and P' in words.
2. Which distribution should you use for this problem? Explain your choice.
3. Construct a 95% confidence interval for the population proportion of adult Americans who feel that crime is the main problem.
 1. State the confidence interval.
 2. Sketch the graph.
 3. Calculate the error bound.
4. Suppose we want to lower the sampling error. What is one way to accomplish that?
5. The sampling error given by Yankelovich Partners, Inc. (which conducted the poll) is $\pm 3\%$. In one to three complete sentences, explain what the $\pm 3\%$ represents.

A Population Proportion

1. X = the number of adult Americans who feel that crime is the main problem; P' = the proportion of adult Americans who feel that crime is the main problem
2. Since we are estimating a proportion, given $P' = 0.2$ and $n = 1000$, the distribution we should use is $N\left(0.2, \sqrt{\frac{(0.2)(0.8)}{1000}}\right)$.
3.
 1. CI: (0.18, 0.22)
 2. Check student's solution.
 3. *EBM*: 0.02
4. One way to lower the sampling error is to increase the sample size.
5. The stated " $\pm 3\%$ " represents the maximum error bound. This means that those doing the study are reporting a maximum error of 3%. Thus, they estimate the percentage of adult Americans who feel that crime is the main problem to be between 18% and 22%.

Refer to [\[link\]](#). Another question in the poll was "[How much are] you worried about the quality of education in our schools?" Sixty-three percent responded "a lot". We are interested in the population proportion of adult Americans who are worried a lot about the quality of education in our schools.

1. Define the random variables X and P' in words.
2. Which distribution should you use for this problem? Explain your choice.
3. Construct a 95% confidence interval for the population proportion of adult Americans who are worried a lot about the quality of education in our schools.
 1. State the confidence interval.
 2. Sketch the graph.
 3. Calculate the error bound.
4. The sampling error given by Yankelovich Partners, Inc. (which conducted the poll) is $\pm 3\%$. In one to three complete sentences, explain what the $\pm 3\%$ represents.

Use the following information to answer the next three exercises: According to a Field Poll, 79% of California adults (actual results are 400 out of 506 surveyed) feel that "education and our schools" is one of the top issues facing California. We wish to construct a 90% confidence interval for the true proportion of California adults who feel that education and the schools is one of the top issues facing California.

A point estimate for the true population proportion is:

1. 0.90
2. 1.27
3. 0.79
4. 400

A Population Proportion

c

A 90% confidence interval for the population proportion is _____.

1. (0.761, 0.820)
2. (0.125, 0.188)
3. (0.755, 0.826)
4. (0.130, 0.183)

The error bound is approximately _____.

1. 1.581
2. 0.791
3. 0.059
4. 0.030

d

Use the following information to answer the next two exercises: Five hundred and eleven (511) homes in a certain southern California community are randomly surveyed to determine if they meet minimal earthquake preparedness recommendations. One hundred seventy-three (173) of the homes surveyed met the minimum recommendations for earthquake preparedness, and 338 did not.

Find the confidence interval at the 90% Confidence Level for the true population proportion of southern California community homes meeting at least the minimum recommendations for earthquake preparedness.

1. (0.2975, 0.3796)
2. (0.6270, 0.6959)
3. (0.3041, 0.3730)
4. (0.6204, 0.7025)

The point estimate for the population proportion of homes that do not meet the minimum recommendations for earthquake preparedness is _____.

1. 0.6614
2. 0.3386
3. 173
4. 338

a

A Population Proportion

On May 23, 2013, Gallup reported that of the 1,005 people surveyed, 76% of U.S. workers believe that they will continue working past retirement age. The confidence level for this study was reported at 95% with a $\pm 3\%$ margin of error.

1. Determine the estimated proportion from the sample.
2. Determine the sample size.
3. Identify CL and α .
4. Calculate the error bound based on the information provided.
5. Compare the error bound in part d to the margin of error reported by Gallup. Explain any differences between the values.
6. Create a confidence interval for the results of this study.
7. A reporter is covering the release of this study for a local news station. How should she explain the confidence interval to her audience?

A national survey of 1,000 adults was conducted on May 13, 2013 by Rasmussen Reports. It concluded with 95% confidence that 49% to 55% of Americans believe that big-time college sports programs corrupt the process of higher education.

1. Find the point estimate and the error bound for this confidence interval.
2. Can we (with 95% confidence) conclude that more than half of all American adults believe this?
3. Use the point estimate from part a and $n = 1,000$ to calculate a 75% confidence interval for the proportion of American adults that believe that major college sports programs corrupt higher education.
4. Can we (with 75% confidence) conclude that at least half of all American adults believe this?

1. $p' = \frac{(0.55 + 0.49)}{2} = 0.52$; $EBP = 0.55 - 0.52 = 0.03$

2. No, the confidence interval includes values less than or equal to 0.50. It is possible that less than half of the population believe this.

3. $CL = 0.75$, so $\alpha = 1 - 0.75 = 0.25$ and $\frac{\alpha}{2} = 0.125$ $z_{\frac{\alpha}{2}} = 1.150$. (The area to the right of this z is 0.125, so the area to the left is $1 - 0.125 = 0.875$.)

$$EBP = (1.150)\sqrt{\frac{0.52(0.48)}{1,000}} \approx 0.018$$

$$(p' - EBP, p' + EBP) = (0.52 - 0.018, 0.52 + 0.018) = (0.502, 0.538)$$

Alternate Solution

STAT TESTS A: 1-PropZinterval with $x = (0.52)(1,000)$, $n = 1,000$, $CL = 0.75$.

Answer is (0.502, 0.538)

A Population Proportion

4. Yes – this interval does not fall less than 0.50 so we can conclude that at least half of all American adults believe that major sports programs corrupt education – but we do so with only 75% confidence.

Public Policy Polling recently conducted a survey asking adults across the U.S. about music preferences. When asked, 80 of the 571 participants admitted that they have illegally downloaded music.

1. Create a 99% confidence interval for the true proportion of American adults who have illegally downloaded music.
2. This survey was conducted through automated telephone interviews on May 6 and 7, 2013. The error bound of the survey compensates for sampling error, or natural variability among samples. List some factors that could affect the survey's outcome that are not covered by the margin of error.
3. Without performing any calculations, describe how the confidence interval would change if the confidence level changed from 99% to 90%.

You plan to conduct a survey on your college campus to learn about the political awareness of students. You want to estimate the true proportion of college students on your campus who voted in the 2012 presidential election with 95% confidence and a margin of error no greater than five percent. How many students must you interview?

$CL = 0.95$ $\alpha = 1 - 0.95 = 0.05$ $\frac{\alpha}{2} = 0.025$ $z_{\frac{\alpha}{2}} = 1.96$. Use $p' = q' = 0.5$.

$$n = \frac{z_{\frac{\alpha}{2}}^2 p' q'}{EBP^2} = \frac{1.96^2 (0.5)(0.5)}{0.05^2} = 384.16$$

You need to interview at least 385 students to estimate the proportion to within 5% at 95% confidence.

In a recent Zogby International Poll, nine of 48 respondents rated the likelihood of a terrorist attack in their community as “likely” or “very likely.” Use the “plus four” method to create a 97% confidence interval for the proportion of American adults who believe that a terrorist attack in their community is likely or very likely. Explain what this confidence interval means in the context of the problem.