Confidence Intervals for a Proportion

Colby Community College

Example 1

In a Gallup poll of 1487 adults, 43% of them said that they have a Facebook page.

Based on this result, what is the best point estimate of the proportion of all adults who have a Facebook page.

Example 1

In a Gallup poll of 1487 adults, 43% of them said that they have a Facebook page.

Based on this result, what is the best point estimate of the proportion of all adults who have a Facebook page.

The sample proportion, 0.43, is the best point estimate of the population proportion.

Example 1

In a Gallup poll of 1487 adults, 43% of them said that they have a Facebook page.

Based on this result, what is the best point estimate of the proportion of all adults who have a Facebook page.

The sample proportion, 0.43, is the best point estimate of the population proportion.

Note

We have no indication of how *good* of an estimate 0.43 is, just that it is the best of the available options.

Definition

A **confidence interval** is a range of values around the point estimate used to estimate the true value of a population parameter.

$$[\text{point estimate} - \text{some value}, \text{point estimate} + \text{some value}]$$

A confidence interval is sometimes abbreviated as CI.

Definition

A **confidence interval** is a range of values around the point estimate used to estimate the true value of a population parameter.

$$[\text{point estimate} - \text{some value}, \text{point estimate} + \text{some value}]$$

A confidence interval is sometimes abbreviated as CI.

Definition

The **confidence level** is the probability that the confidence interval actually does contain the population parameter, assuming that the estimation process is repeated a large number of times.

Definition

A **confidence interval** is a range of values around the point estimate used to estimate the true value of a population parameter.

$$[\text{point estimate} - \text{some value}, \text{point estimate} + \text{some value}]$$

A confidence interval is sometimes abbreviated as CI.

Definition

The **confidence level** is the probability that the confidence interval actually does contain the population parameter, assuming that the estimation process is repeated a large number of times.

Note

Round the confidence interval limits to three significant digits.

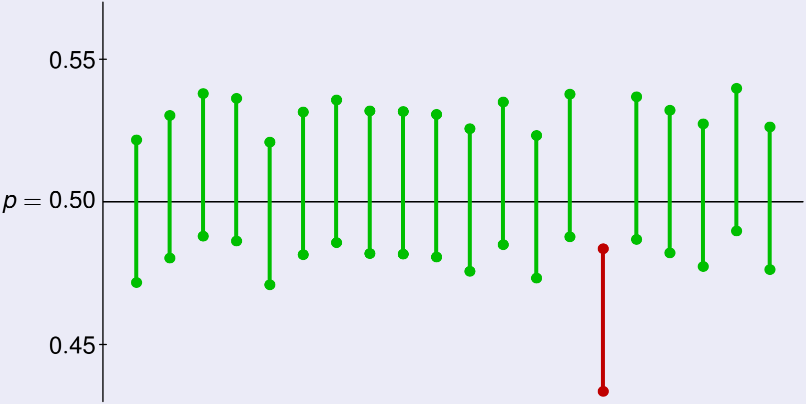
The Process Success Rate

A confidence level of 95% tells us that the process we use should, given enough iterations, result in a confidence interval that contains the true population proportion 95% of the time.

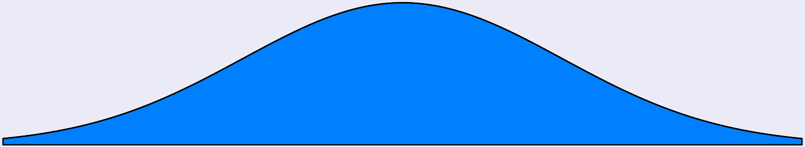
The Process Success Rate

A confidence level of 95% tells us that the process we use should, given enough iterations, result in a confidence interval that contains the true population proportion 95% of the time.

If the true population proportion is $p = 0.5$, then we expect around 19 of 20 confidence intervals to contain the true value of p.

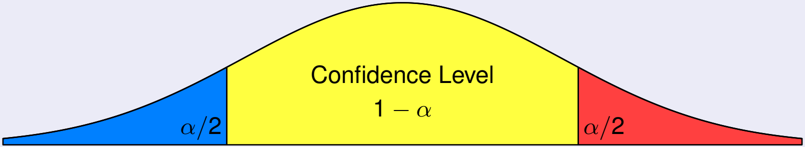


A Few Observations



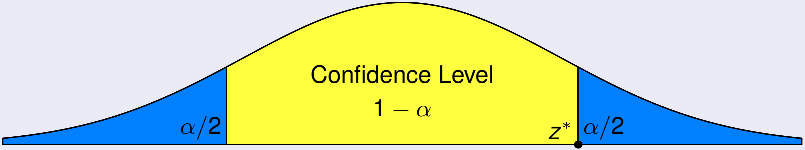
- When the requirements of the Central Limit Theorem are met, the sampling distribution of sample proportions can be approximated by a normal distribution.

A Few Observations



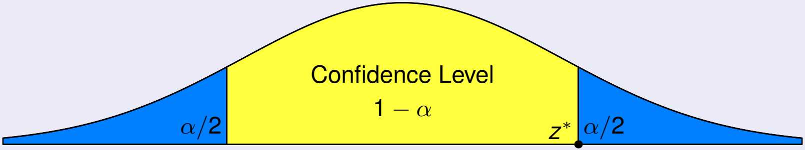
- When the requirements of the Central Limit Theorem are met, the sampling distribution of sample proportions can be approximated by a normal distribution.
- A z score associated with a sample proportion has a probability $\alpha/2$ of falling in the right tail portion.

A Few Observations



- When the requirements of the Central Limit Theorem are met, the sampling distribution of sample proportions can be approximated by a normal distribution.
- A z score associated with a sample proportion has a probability $\alpha/2$ of falling in the right tail portion.
- The z score at the boundary of the right-tail region is commonly denoted by z^*.

A Few Observations



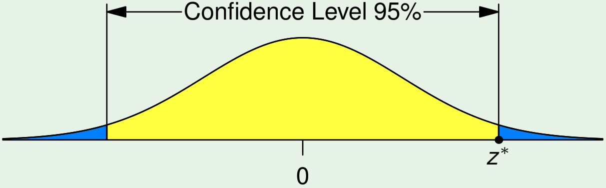
- When the requirements of the Central Limit Theorem are met, the sampling distribution of sample proportions can be approximated by a normal distribution.
- A z score associated with a sample proportion has a probability $\alpha/2$ of falling in the right tail portion.
- The z score at the boundary of the right-tail region is commonly denoted by z^*.

Definition

The value z^* is called a **critical value**.

Example 2

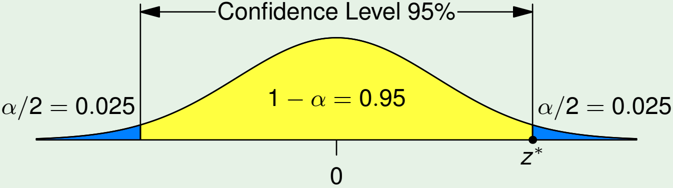
Let us find the critical value corresponding to a 95% confidence level.



Example 2

Let us find the critical value corresponding to a 95% confidence level.

A 95% confidence interval gives $\alpha = 0.05$ and $\alpha/2 = 0.025$.

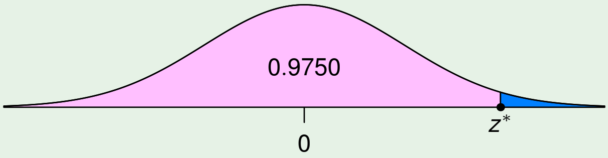


Example 2

Let us find the critical value corresponding to a 95% confidence level.

A 95% confidence interval gives $\alpha = 0.05$ and $\alpha/2 = 0.025$.

To find the z value using the inverse normal distribution, we need to know the area to the left of the right tail, $0.025 + 0.95 = 0.9750$.



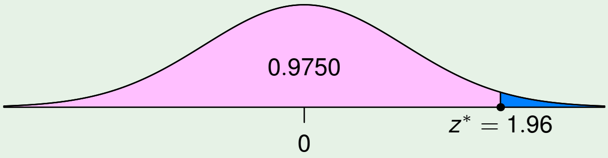
Example 2

Let us find the critical value corresponding to a 95% confidence level.

A 95% confidence interval gives $\alpha = 0.05$ and $\alpha/2 = 0.025$.

To find the z value using the inverse normal distribution, we need to know the area to the left of the right tail, $0.025 + 0.95 = 0.9750$.

Using technology we get $z^* = 1.96$.



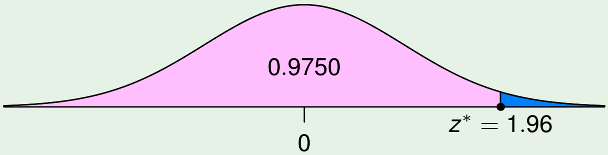
Example 2

Let us find the critical value corresponding to a 95% confidence level.

A 95% confidence interval gives $\alpha = 0.05$ and $\alpha/2 = 0.025$.

To find the z value using the inverse normal distribution, we need to know the area to the left of the right tail, $0.025 + 0.95 = 0.9750$.

Using technology we get $z^* = 1.96$.



Common Confidence Levels

Confidence Level	α	Critical Value
90%	0.10	1.645
95%	0.05	1.960
99%	0.01	2.575

Definition

Anytime a point estimate is used to estimate a parameter, there will be some amount of error.

Definition

Anytime a point estimate is used to estimate a parameter, there will be some amount of error.

The maximum likely amount of error is called the **margin of error**

Definition

Anytime a point estimate is used to estimate a parameter, there will be some amount of error.

The maximum likely amount of error is called the **margin of error**

When the point estimate closely follows a normal model, the margin of error is found by multiplying the critical value and the standard error:

$$z^* \cdot SE$$

Definition

Anytime a point estimate is used to estimate a parameter, there will be some amount of error.

The maximum likely amount of error is called the **margin of error**

When the point estimate closely follows a normal model, the margin of error is found by multiplying the critical value and the standard error:

$$z^* \cdot SE$$

Note

The margin of error for $\hat{p}$ is:

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

Procedure for Constructing a Confidence Interval for p

1. Verify that $\hat{p}$ is nearly normal:

Procedure for Constructing a Confidence Interval for p

1. Verify that $\hat{p}$ is nearly normal:
    - The observations are independent.

Procedure for Constructing a Confidence Interval for p

1. Verify that $\hat{p}$ is nearly normal:
    - The observations are independent.
    - $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$.

Procedure for Constructing a Confidence Interval for p

1. Verify that $\hat{p}$ is nearly normal:
   - The observations are independent.
   - $n\hat{p} \geq 10$ and $n(1 - \hat{p}) \geq 10$.
2. Find the critical value z^*.

Procedure for Constructing a Confidence Interval for p

1. Verify that $\hat{p}$ is nearly normal:
   - The observations are independent.
   - $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$.
2. Find the critical value z^*.
3. Evaluate the margin of error: $E = z^* \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

Procedure for Constructing a Confidence Interval for p

1. Verify that $\hat{p}$ is nearly normal:
   - The observations are independent.
   - $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$.
2. Find the critical value z^*.
3. Evaluate the margin of error: $E = z^* \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$
4. Construct the interval: $(\hat{p} - E, \hat{p} + E)$

Procedure for Constructing a Confidence Interval for p

1. Verify that $\hat{p}$ is nearly normal:
   - The observations are independent.
   - $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$.
2. Find the critical value z^*.
3. Evaluate the margin of error: $E = z^* \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$
4. Construct the interval: $(\hat{p} - E, \hat{p} + E)$
5. Interpret the confidence interval in the context of the problem.

Procedure for Constructing a Confidence Interval for p

1. Verify that $\hat{p}$ is nearly normal:
   - The observations are independent.
   - $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$.
2. Find the critical value z^*.
3. Evaluate the margin of error: $E = z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
4. Construct the interval: $(\hat{p} - E, \hat{p} + E)$
5. Interpret the confidence interval in the context of the problem.

Note

Statistics software and graphing calculators, can calculate the confidence interval for you.

Procedure for Constructing a Confidence Interval for p

1. Verify that $\hat{p}$ is nearly normal:
   - The observations are independent.
   - $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$.
2. Find the critical value z^*.
3. Evaluate the margin of error: $E = z^* \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$
4. Construct the interval: $(\hat{p} - E, \hat{p} + E)$
5. Interpret the confidence interval in the context of the problem.

Note

Statistics software and graphing calculators, can calculate the confidence interval for you.

Note

Round the confidence interval limits to four digits.

Example 3

A 2018 Pew Research poll found that 88.7% of a random sample of 1000 American adults supported expanding solar power. Let's compute a 90% confidence interval.

Example 3

A 2018 Pew Research poll found that 88.7% of a random sample of 1000 American adults supported expanding solar power. Let's compute a 90% confidence interval.

1. $n\hat{p} =$

Example 3

A 2018 Pew Research poll found that 88.7% of a random sample of 1000 American adults supported expanding solar power. Let's compute a 90% confidence interval.

1. $n\hat{p} = 1000 \cdot 0.887$

Example 3

A 2018 Pew Research poll found that 88.7% of a random sample of 1000 American adults supported expanding solar power. Let's compute a 90% confidence interval.

1. $n\hat{p} = 1000 \cdot 0.887 = 887 \geq 10$

Example 3

A 2018 Pew Research poll found that 88.7% of a random sample of 1000 American adults supported expanding solar power. Let's compute a 90% confidence interval.

1. $n\hat{p} = 1000 \cdot 0.887 = 887 \geq 10$
   $n(1 - \hat{p}) =$

Example 3

A 2018 Pew Research poll found that 88.7% of a random sample of 1000 American adults supported expanding solar power. Let's compute a 90% confidence interval.

1. $n\hat{p} = 1000 \cdot 0.887 = 887 \geq 10$
   $n(1 - \hat{p}) = 1000 \cdot (1 - 0.887)$

Example 3

A 2018 Pew Research poll found that 88.7% of a random sample of 1000 American adults supported expanding solar power. Let's compute a 90% confidence interval.

1. $n\hat{p} = 1000 \cdot 0.887 = 887 \geq 10$
   $n(1 - \hat{p}) = 1000 \cdot (1 - 0.887) = 113 \geq 10$

Example 3

A 2018 Pew Research poll found that 88.7% of a random sample of 1000 American adults supported expanding solar power. Let's compute a 90% confidence interval.

1. $n\hat{p} = 1000 \cdot 0.887 = 887 \geq 10$
   $n(1 - \hat{p}) = 1000 \cdot (1 - 0.887) = 113 \geq 10$
2. The critical value for 90% confidence interval is z^* is 1.645.

Example 3

A 2018 Pew Research poll found that 88.7% of a random sample of 1000 American adults supported expanding solar power. Let's compute a 90% confidence interval.

1. $n\hat{p} = 1000 \cdot 0.887 = 887 \geq 10$
   $n(1 - \hat{p}) = 1000 \cdot (1 - 0.887) = 113 \geq 10$
2. The critical value for 90% confidence interval is z^* is 1.645.
3. The margin of error is

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

Example 3

A 2018 Pew Research poll found that 88.7% of a random sample of 1000 American adults supported expanding solar power. Let's compute a 90% confidence interval.

1. $n\hat{p} = 1000 \cdot 0.887 = 887 \geq 10$
   $n(1-\hat{p}) = 1000 \cdot (1 - 0.887) = 113 \geq 10$
2. The critical value for 90% confidence interval is z^* is 1.645.
3. The margin of error is

$$E = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.645 \sqrt{\frac{0.887(1-0.887)}{1000}}$$

Example 3

A 2018 Pew Research poll found that 88.7% of a random sample of 1000 American adults supported expanding solar power. Let's compute a 90% confidence interval.

1. $n\hat{p} = 1000 \cdot 0.887 = 887 \geq 10$
   $n(1-\hat{p}) = 1000 \cdot (1 - 0.887) = 113 \geq 10$
2. The critical value for 90% confidence interval is z^* is 1.645.
3. The margin of error is

$$E = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.645 \sqrt{\frac{0.887(1-0.887)}{1000}} = 0.016469$$

Example 3

A 2018 Pew Research poll found that 88.7% of a random sample of 1000 American adults supported expanding solar power. Let's compute a 90% confidence interval.

1. $n\hat{p} = 1000 \cdot 0.887 = 887 \geq 10$
   $n(1-\hat{p}) = 1000 \cdot (1-0.887) = 113 \geq 10$
2. The critical value for 90% confidence interval is z^* is 1.645.
3. The margin of error is

$$E = z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.645\sqrt{\frac{0.887(1-0.887)}{1000}} = 0.016469$$

4. The confidence interval is

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

Example 3

A 2018 Pew Research poll found that 88.7% of a random sample of 1000 American adults supported expanding solar power. Let's compute a 90% confidence interval.

1. $n\hat{p} = 1000 \cdot 0.887 = 887 \geq 10$
   $n(1-\hat{p}) = 1000 \cdot (1-0.887) = 113 \geq 10$
2. The critical value for 90% confidence interval is z^* is 1.645.
3. The margin of error is
   $$E = z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.645\sqrt{\frac{0.887(1-0.887)}{1000}} = 0.016469$$
4. The confidence interval is
   $$\begin{array}{ccccc} (& \hat{p}-E & , & \hat{p}+E &) \\ (& 0.887-0.016469 & , & 0.887+0.016469 &) \end{array}$$

Example 3

A 2018 Pew Research poll found that 88.7% of a random sample of 1000 American adults supported expanding solar power. Let's compute a 90% confidence interval.

1. $n\hat{p} = 1000 \cdot 0.887 = 887 \geq 10$
   $n(1-\hat{p}) = 1000 \cdot (1 - 0.887) = 113 \geq 10$
2. The critical value for 90% confidence interval is z^* is 1.645.
3. The margin of error is
$$E = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.645 \sqrt{\frac{0.887(1-0.887)}{1000}} = 0.016469$$
4. The confidence interval is
$$\begin{array}{ccccc} (& \hat{p} - E & , & \hat{p} + E &) \\ (& 0.887 - 0.016469 & , & 0.887 + 0.016469 &) \\ (& 0.8705 & , & 0.9035 &) \end{array}$$

Example 3

A 2018 Pew Research poll found that 88.7% of a random sample of 1000 American adults supported expanding solar power. Let's compute a 90% confidence interval.

1. $n\hat{p} = 1000 \cdot 0.887 = 887 \geq 10$
   $n(1-\hat{p}) = 1000 \cdot (1-0.887) = 113 \geq 10$
2. The critical value for 90% confidence interval is z^* is 1.645.
3. The margin of error is
   $$E = z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.645\sqrt{\frac{0.887(1-0.887)}{1000}} = 0.016469$$
4. The confidence interval is
   $$\begin{array}{lcccr} (& \hat{p}-E & , & \hat{p}+E &) \\ (& 0.887-0.016469 & , & 0.887+0.016469 &) \\ (& 0.8705 & , & 0.9035 &) \end{array}$$
5. We are 90% confident that between 87.1% and 90.4% of American adults supported the expansion of solar power in 2018.

Interpreting a Confidence Interval

For the confidence interval $(0.8705, 0.9035)$ there is one correct interpretation and many creatively incorrect interpretations.

Interpreting a Confidence Interval

For the confidence interval $(0.8705, 0.9035)$ there is one correct interpretation and many creatively incorrect interpretations.

Correct: "We are 90% confident that the interval from 0.8705 to 0.9035 actually does contain the true value of the population proportion p."

Reason: The confidence level 95% refers to the success rate of the process used to estimate the population proportion.

Interpreting a Confidence Interval

For the confidence interval $(0.8705, 0.9035)$ there is one correct interpretation and many creatively incorrect interpretations.

Correct: "We are 90% confident that the interval from 0.8705 to 0.9035 actually does contain the true value of the population proportion p."

Reason: The confidence level 95% refers to the success rate of the process used to estimate the population proportion.

Incorrect: "There is a 90% chance that the true value of p will fall between 0.8705 and 0.9035."

Reason: The population proportion p is a fixed value.

Interpreting a Confidence Interval

For the confidence interval $(0.8705, 0.9035)$ there is one correct interpretation and many creatively incorrect interpretations.

Correct: "We are 90% confident that the interval from 0.8705 to 0.9035 actually does contain the true value of the population proportion p."

Reason: The confidence level 95% refers to the success rate of the process used to estimate the population proportion.

Incorrect: "There is a 90% chance that the true value of p will fall between 0.8705 and 0.9035."

Reason: The population proportion p is a fixed value.

Incorrect: "90% of sample proportions will fall between 0.8705 and 0.9035."

Reason: The values 0.8705 and 0.9035 result from one sample, they are not parameters describing the behavior of all samples.

Analyzing Polls

When analyzing results from polls, consider the following:

- The sample should be a simple random sample.

Analyzing Polls

When analyzing results from polls, consider the following:

- The sample should be a simple random sample.
- The confidence interval should be provided.

Analyzing Polls

When analyzing results from polls, consider the following:

- The sample should be a simple random sample.
- The confidence interval should be provided.
- The sample size should be provided.

Analyzing Polls

When analyzing results from polls, consider the following:

- The sample should be a simple random sample.
- The confidence interval should be provided.
- The sample size should be provided.
- Except for relatively rare cases, the quality of the poll results depends on the sampling method and the size of the sample, but the size of the population is usually not a factor.

Analyzing Polls

When analyzing results from polls, consider the following:

- The sample should be a simple random sample.
- The confidence interval should be provided.
- The sample size should be provided.
- Except for relatively rare cases, the quality of the poll results depends on the sampling method and the size of the sample, but the size of the population is usually not a factor.

Caution

Never think that poll results are unreliable if the sample size if a small percentage of the population size.

Finding $\hat{p}$ from a Confidence Interval

If you know the confidence interval, (upper limit, lower limit), we can calculate the point estimate:

$$\hat{p} = \frac{(\text{upper limit}) + (\text{lower limit})}{2}$$

Finding $\hat{p}$ from a Confidence Interval

If you know the confidence interval, $(\text{upper limit}, \text{lower limit})$, we can calculate the point estimate:

$$\hat{p} = \frac{(\text{upper limit}) + (\text{lower limit})}{2}$$

Finding E from a Confidence Interval

If you know the confidence interval, $(\text{upper limit}, \text{lower limit})$, we can calculate the margin of error:

$$E = \frac{(\text{upper limit}) - (\text{lower limit})}{2}$$

Example 4

The article “High-Dose Nicotine Patch Therapy”, by Dale, Hurt, et al. (*Journal of the American Medical Association*, Vol 274, No. 17) includes the statement:

> Of the 71 subjects, 70% were abstinent from smoking at 8 weeks (95% confidence interval [CI], 58% to 81%).

Example 4

The article “High-Dose Nicotine Patch Therapy”, by Dale, Hurt, et al. (*Journal of the American Medical Association*, Vol 274, No. 17) includes the statement:

> Of the 71 subjects, 70% were abstinent from smoking at 8 weeks (95% confidence interval [CI], 58% to 81%).

The point estimate $\hat{p}$ is

$$\hat{p} = \frac{\text{(upper limit)} + \text{(lower limit)}}{2}$$

Example 4

The article “High-Dose Nicotine Patch Therapy”, by Dale, Hurt, et al. (*Journal of the American Medical Association*, Vol 274, No. 17) includes the statement:

> Of the 71 subjects, 70% were abstinent from smoking at 8 weeks (95% confidence interval [CI], 58% to 81%).

The point estimate $\hat{p}$ is

$$\hat{p} = \frac{\text{(upper limit)} + \text{(lower limit)}}{2} = \frac{0.81 + 0.58}{2}$$

Example 4

The article “High-Dose Nicotine Patch Therapy”, by Dale, Hurt, et al. (*Journal of the American Medical Association*, Vol 274, No. 17) includes the statement:

> Of the 71 subjects, 70% were abstinent from smoking at 8 weeks (95% confidence interval [CI], 58% to 81%).

The point estimate $\hat{p}$ is

$$\hat{p} = \frac{(\text{upper limit}) + (\text{lower limit})}{2} = \frac{0.81 + 0.58}{2} = 0.695$$

Example 4

The article “High-Dose Nicotine Patch Therapy”, by Dale, Hurt, et al. (*Journal of the American Medical Association*, Vol 274, No. 17) includes the statement:

> Of the 71 subjects, 70% were abstinent from smoking at 8 weeks (95% confidence interval [CI], 58% to 81%).

The point estimate $\hat{p}$ is

$$\hat{p} = \frac{(\text{upper limit}) + (\text{lower limit})}{2} = \frac{0.81 + 0.58}{2} = 0.695$$

The margin of error E is

$$E = \frac{(\text{upper limit}) - (\text{lower limit})}{2}$$

Example 4

The article "High-Dose Nicotine Patch Therapy", by Dale, Hurt, et al. (*Journal of the American Medical Association*, Vol 274, No. 17) includes the statement:

> Of the 71 subjects, 70% were abstinent from smoking at 8 weeks (95% confidence interval [CI], 58% to 81%).

The point estimate $\hat{p}$ is

$$\hat{p} = \frac{(\text{upper limit}) + (\text{lower limit})}{2} = \frac{0.81 + 0.58}{2} = 0.695$$

The margin of error E is

$$E = \frac{(\text{upper limit}) - (\text{lower limit})}{2} = \frac{0.81 - 0.58}{2}$$

Example 4

The article "High-Dose Nicotine Patch Therapy", by Dale, Hurt, et al. (*Journal of the American Medical Association*, Vol 274, No. 17) includes the statement:

> Of the 71 subjects, 70% were abstinent from smoking at 8 weeks (95% confidence interval [CI], 58% to 81%).

The point estimate $\hat{p}$ is

$$\hat{p} = \frac{(\text{upper limit}) + (\text{lower limit})}{2} = \frac{0.81 + 0.58}{2} = 0.695$$

The margin of error E is

$$E = \frac{(\text{upper limit}) - (\text{lower limit})}{2} = \frac{0.81 - 0.58}{2} = 0.115$$

Example 5

Let us assume a researcher can find no information about the percentage of adults who make online purchases. This information is extremely important to online stores, so the researcher decides to conduct a survey.

Example 5

Let us assume a researcher can find no information about the percentage of adults who make online purchases. This information is extremely important to online stores, so the researcher decides to conduct a survey.

How many adults must be surveyed in order to be 95% confident that the sample percentage is in error by no more than three percentage points?

Example 5

Let us assume a researcher can find no information about the percentage of adults who make online purchases. This information is extremely important to online stores, so the researcher decides to conduct a survey.

How many adults must be surveyed in order to be 95% confident that the sample percentage is in error by no more than three percentage points?

We can solve the margin of error equation for n:

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

Example 5

Let us assume a researcher can find no information about the percentage of adults who make online purchases. This information is extremely important to online stores, so the researcher decides to conduct a survey.

How many adults must be surveyed in order to be 95% confident that the sample percentage is in error by no more than three percentage points?

We can solve the margin of error equation for n:

$$E = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad \Rightarrow \quad n = \frac{(z^*)^2 \hat{p}(1-\hat{p})}{E^2}$$

Example 5

Let us assume a researcher can find no information about the percentage of adults who make online purchases. This information is extremely important to online stores, so the researcher decides to conduct a survey.

How many adults must be surveyed in order to be 95% confident that the sample percentage is in error by no more than three percentage points?

We can solve the margin of error equation for n:

$$E = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad \Rightarrow \quad n = \frac{(z^*)^2 \hat{p}(1-\hat{p})}{E^2}$$

We have $z^* = 1.96$ and $E = 0.03$, but what about $\hat{p}$?

Sample Size Required to Estimate a Population Proportion

The sample must be a simple random sample of independent sample units.

Sample Size Required to Estimate a Population Proportion

The sample must be a simple random sample of independent sample units.

If a reasonable estimate of $\hat{p}$ can be made by using previous samples, a pilot study, or someone's expert knowledge:

$$n = \frac{(z^*)^2 \hat{p}(1-\hat{p})}{E^2}$$

Sample Size Required to Estimate a Population Proportion

The sample must be a simple random sample of independent sample units.

If a reasonable estimate of $\hat{p}$ can be made by using previous samples, a pilot study, or someone's expert knowledge:

$$n = \frac{(z^*)^2 \hat{p}(1-\hat{p})}{E^2}$$

If nothing is known about the value $\hat{p}$:

$$n = \frac{(z^*)^2 0.25}{E^2}$$

Sample Size Required to Estimate a Population Proportion

The sample must be a simple random sample of independent sample units.

If a reasonable estimate of $\hat{p}$ can be made by using previous samples, a pilot study, or someone's expert knowledge:

$$n = \frac{(z^*)^2 \hat{p}(1 - \hat{p})}{E^2}$$

If nothing is known about the value $\hat{p}$:

$$n = \frac{(z^*)^2 0.25}{E^2}$$

Rounding

If the computed sample size n is not a whole number, round the value of n up to the next larger whole number.

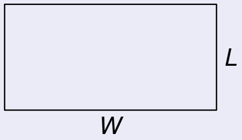
Why 0.25?

If $\hat{p} = 0.5$, then $\hat{p}(1 - \hat{p}) = 0.25$ is the largest possible product.

Why 0.25?

If $\hat{p} = 0.5$, then $\hat{p}(1 - \hat{p}) = 0.25$ is the largest possible product.

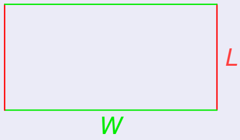
To see why, consider the rectangle:



Why 0.25?

If $\hat{p} = 0.5$, then $\hat{p}(1 - \hat{p}) = 0.25$ is the largest possible product.

To see why, consider the rectangle:



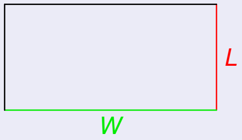
Let us assume that the perimeter of this rectangle is 2, which means:

$$2L + 2W = 2$$

Why 0.25?

If $\hat{p} = 0.5$, then $\hat{p}(1 - \hat{p}) = 0.25$ is the largest possible product.

To see why, consider the rectangle:



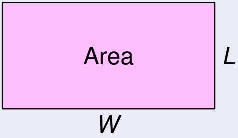
Let us assume that the perimeter of this rectangle is 2, which means:

$$2L + 2W = 2 \qquad \Rightarrow \qquad L + W = 1$$

Why 0.25?

If $\hat{p} = 0.5$, then $\hat{p}(1 - \hat{p}) = 0.25$ is the largest possible product.

To see why, consider the rectangle:



Let us assume that the perimeter of this rectangle is 2, which means

$$2L + 2W = 2 \qquad \Rightarrow \qquad L + W = 1$$

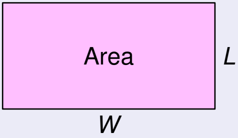
We can then write the area of this rectangle as

$$\text{Area} = LW = L(1 - L) = -L^2 + L$$

Why 0.25?

If $\hat{p} = 0.5$, then $\hat{p}(1 - \hat{p}) = 0.25$ is the largest possible product.

To see why, consider the rectangle:



Let us assume that the perimeter of this rectangle is 2, which means

$$2L + 2W = 2 \qquad \Rightarrow \qquad L + W = 1$$

We can then write the area of this rectangle as

$$\text{Area} = LW = L(1 - L) = -L^2 + L$$

Since this is a parabola that opens down, we know that the vertex, $(0.5, 0.5)$, is the maximum value.

Example 6

Let us return to Example 5, finding the sample size for a survey.

Example 6

Let us return to Example 5, finding the sample size for a survey.

We have $z^* = 1.96$ and $E = 0.03$, but we know nothing about $\hat{p}$.

Example 6

Let us return to Example 5, finding the sample size for a survey.

We have $z^* = 1.96$ and $E = 0.03$, but we know nothing about $\hat{p}$.

$$n = \frac{(z^*)^2 0.25}{E^2}$$

Example 6

Let us return to Example 5, finding the sample size for a survey.

We have $z^* = 1.96$ and $E = 0.03$, but we know nothing about $\hat{p}$.

$$n = \frac{(z^*)^2 0.25}{E^2} = \frac{(1.96)^2 0.25}{(0.03)^2}$$

Example 6

Let us return to Example 5, finding the sample size for a survey.

We have $z^* = 1.96$ and $E = 0.03$, but we know nothing about $\hat{p}$.

$$n = \frac{(z^*)^2 0.25}{E^2} = \frac{(1.96)^2 0.25}{(0.03)^2} = 1067.11$$

Example 6

Let us return to Example 5, finding the sample size for a survey.

We have $z^* = 1.96$ and $E = 0.03$, but we know nothing about $\hat{p}$.

$$n = \frac{(z^*)^2 0.25}{E^2} = \frac{(1.96)^2 0.25}{(0.03)^2} = 1067.11$$

So, we need at least 1068 adults in our survey.

Example 6

Let us return to Example 5, finding the sample size for a survey.

We have $z^* = 1.96$ and $E = 0.03$, but we know nothing about $\hat{p}$.

$$n = \frac{(z^*)^2 0.25}{E^2} = \frac{(1.96)^2 0.25}{(0.03)^2} = 1067.11$$

So, we need at least 1068 adults in our survey.

Caution

1. Don't make the mistake of using $E = 3$ as the margin of error corresponding to "three percentage points."

Example 6

Let us return to Example 5, finding the sample size for a survey.

We have $z^* = 1.96$ and $E = 0.03$, but we know nothing about $\hat{p}$.

$$n = \frac{(z^*)^2 0.25}{E^2} = \frac{(1.96)^2 0.25}{(0.03)^2} = 1067.11$$

So, we need at least 1068 adults in our survey.

Caution

1. Don't make the mistake of using $E = 3$ as the margin of error corresponding to "three percentage points."
2. Be sure to substitute correct the critical z score for z^* for the confidence level.

Example 6

Let us return to Example 5, finding the sample size for a survey.

We have $z^* = 1.96$ and $E = 0.03$, but we know nothing about $\hat{p}$.

$$n = \frac{(z^*)^2 0.25}{E^2} = \frac{(1.96)^2 0.25}{(0.03)^2} = 1067.11$$

So, we need at least 1068 adults in our survey.

Caution

1. Don't make the mistake of using $E = 3$ as the margin of error corresponding to "three percentage points."
2. Be sure to substitute correct the critical z score for z^* for the confidence level.
3. Be sure to *round up to the next highest integer*, do not round using the usual rounding rules.