Estimating a Population Proportion

Colby Community College

Definition

A **point estimate** is a single value used to estimate a population parameter.

Definition

A **point estimate** is a single value used to estimate a population parameter.

Notation

The sample proportion $\hat{p}$ is the best point estimate of the population proportion p.

Definition

A **point estimate** is a single value used to estimate a population parameter.

Notation

The sample proportion $\hat{p}$ is the best point estimate of the population proportion p.

Note

We use the sample proportion $\hat{p}$ to estimate the population portion p because it is unbiased and the most consistent of the estimators that could be used.

Definition

A **point estimate** is a single value used to estimate a population parameter.

Notation

The sample proportion $\hat{p}$ is the best point estimate of the population proportion p.

Note

We use the sample proportion $\hat{p}$ to estimate the population portion p because it is unbiased and the most consistent of the estimators that could be used.

Recall

An unbiased estimator is a statistic that targets the value of the corresponding population parameter in the sense that the sampling distribution of the statistic has a mean that is equal to the corresponding population parameter.

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 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 is the best point estimate of the population proportion.

The sample proportion is 0.43, so the best estimate of p is 0.43.

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 is the best point estimate of the population proportion.

The sample proportion is 0.43, so the best estimate of p is 0.43.

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 used to estimate the true value of a population parameter. A confidence interval is abbreviated as CI.

Definition

A **confidence interval** is a range of values used to estimate the true value of a population parameter. A confidence interval is abbreviated as CI.

Definition

The **confidence level** is the probability $1 - \alpha$ 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 used to estimate the true value of a population parameter. A confidence interval is abbreviated as CI.

Definition

The **confidence level** is the probability $1 - \alpha$ that the confidence interval actually does contain the population parameter, assuming that the estimation process is repeated a large number of times.

Note

The confidence is sometimes called the **degree of confidence** or the **confidence coefficient**.

Definition

A **confidence interval** is a range of values used to estimate the true value of a population parameter. A confidence interval is abbreviated as CI.

Definition

The **confidence level** is the probability $1 - \alpha$ that the confidence interval actually does contain the population parameter, assuming that the estimation process is repeated a large number of times.

Note

The confidence is sometimes called the **degree of confidence** or the **confidence coefficient**.

Rounding

Round the confidence interval limits for p to three significant digits.

Interpreting a Confidence Interval

For the confidence interval $0.405 < p < 0.455$ there is one correct interpretation and many creatively incorrect interpretations.

Interpreting a Confidence Interval

For the confidence interval $0.405 < p < 0.455$ there is one correct interpretation and many creatively incorrect interpretations.

Correct: "We are 95% confident that the interval from 0.405 to 0.455 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 $0.405 < p < 0.455$ there is one correct interpretation and many creatively incorrect interpretations.

Correct: "We are 95% confident that the interval from 0.405 to 0.455 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 95% chance that the true value of p will fall between 0.405 and 0.455."

Reason: The population proportion p is a fixed value, it is not a random variable.

Interpreting a Confidence Interval

For the confidence interval $0.405 < p < 0.455$ there is one correct interpretation and many creatively incorrect interpretations.

Correct: "We are 95% confident that the interval from 0.405 to 0.455 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 95% chance that the true value of p will fall between 0.405 and 0.455."

Reason: The population proportion p is a fixed value, it is not a random variable.

Incorrect: "95% of sample proportions will fall between 0.405 and 0.455."

Reason: The values 0.405 and 0.455 result from one sample, they are not parameters describing the behavior of all samples.

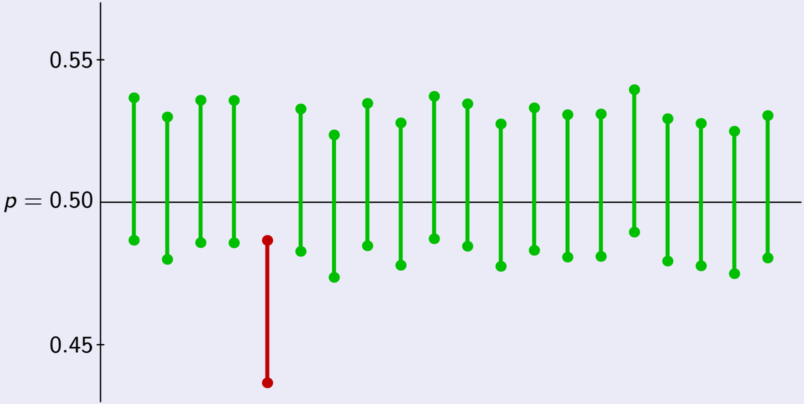
The Process Success Rate

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.

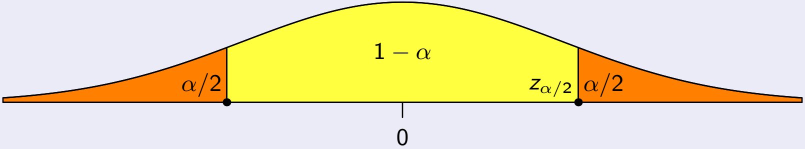
The Process Success Rate

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 about 19 out of 20 confidence intervals to contain the true value of p.

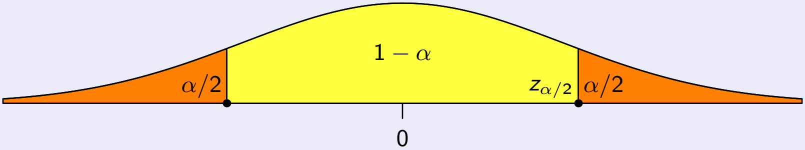


A Few Observations



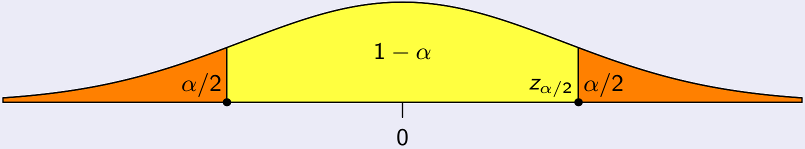
- When requirements are met, the sampling distribution of sample proportions can be approximated by a normal distribution.

A Few Observations



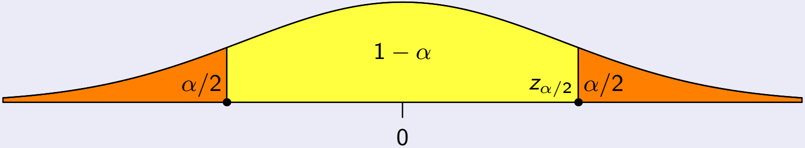
- When requirements are met, the sampling distribution of sample proportions can be approximated by a normal distribution.
- A z score associated with a sample proportion has a probability of $\alpha/2$ of falling in the right tail portion.

A Few Observations



- When requirements 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 of $\alpha/2$ of falling in the right tail portion.
- The z score at the boundary of the right-tail region is commonly denoted by $z_{\alpha/2}$ and is the borderline separating values that significantly high.

A Few Observations



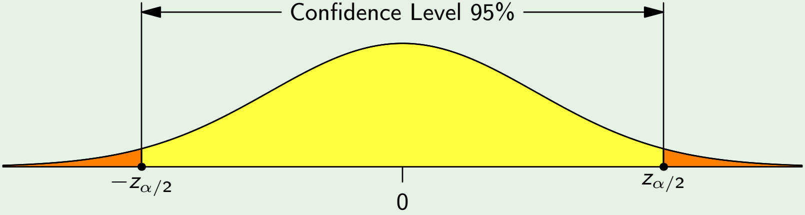
- When requirements are met, the sampling distribution of sample proportions can be approximated by a normal distribution.
- A z score associated with a sample proportion has a probability of $\alpha/2$ of falling in the right tail portion.
- The z score at the boundary of the right-tail region is commonly denoted by $z_{\alpha/2}$ and is the borderline separating values that significantly high.

Definition

A **critical value** is the number on the borderline separating sample statistics that are significantly high or low from those that are not significant.

Example 2

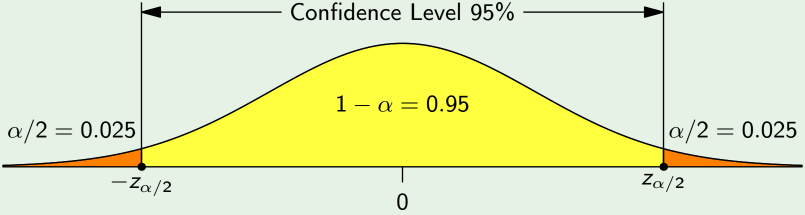
Let us find the critical value $z_{\alpha/2}$ corresponding to a 95% confidence level.



Example 2

Let us find the critical value $z_{\alpha/2}$ corresponding to a 95% confidence level.

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

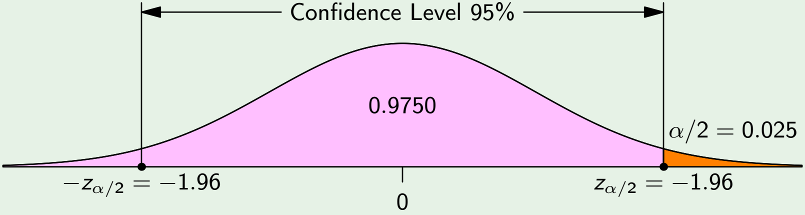


Example 2

Let us find the critical value $z_{\alpha/2}$ 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 cumulative area to the left of the right tail, $0.025 + 0.95 = 0.9750$.

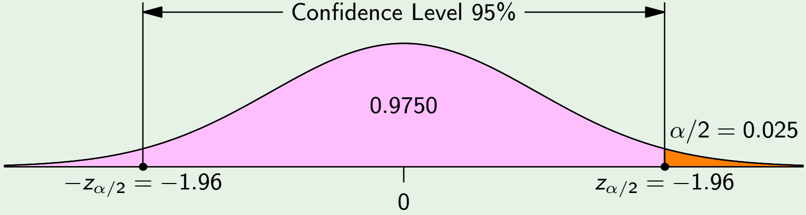


Example 2

Let us find the critical value $z_{\alpha/2}$ 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 cumulative area to the left of the right tail, $0.025 + 0.95 = 0.9750$.



Common Confidence Levels

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

Definition

When data from a simple random sample are used to estimate a population proportion p, the difference between the sample proportion $\hat{p}$ and the population proportion p is an error.

Definition

When data from a simple random sample are used to estimate a population proportion p, the difference between the sample proportion $\hat{p}$ and the population proportion p is an error.

The maximum likely amount of error is the **margin of error**, denoted by $\boldsymbol{E}$.

Definition

When data from a simple random sample are used to estimate a population proportion p, the difference between the sample proportion $\hat{p}$ and the population proportion p is an error.

The maximum likely amount of error is the **margin of error**, denoted by $\boldsymbol{E}$.

There is a probability of $1-\alpha$ that $\hat{p}-E<p<\hat{p}+E$.

Definition

When data from a simple random sample are used to estimate a population proportion p, the difference between the sample proportion $\hat{p}$ and the population proportion p is an error.

The maximum likely amount of error is the **margin of error**, denoted by $\boldsymbol{E}$.

There is a probability of $1 - \alpha$ that $\hat{p} - E < p < \hat{p} + E$.

Note

The margin of error E is also called the **maximum error of the estimate** and can be found by multiplying the critical value and the estimated standard deviation of sample proportions.

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

Procedure for Constructing a Confidence Interval for p

1. Verify that the following requirements are satisfied:

Procedure for Constructing a Confidence Interval for p

1. Verify that the following requirements are satisfied:
   - The sample is a simple random sample.

Procedure for Constructing a Confidence Interval for p

1. Verify that the following requirements are satisfied:
   - The sample is a simple random sample.
   - The conditions for the binomial distribution are satisfied.

Procedure for Constructing a Confidence Interval for p

1. Verify that the following requirements are satisfied:
    - The sample is a simple random sample.
    - The conditions for the binomial distribution are satisfied.
    - There are at least 5 successes and at least 5 failures.

Procedure for Constructing a Confidence Interval for p

1. Verify that the following requirements are satisfied:
   - The sample is a simple random sample.
   - The conditions for the binomial distribution are satisfied.
   - There are at least 5 successes and at least 5 failures.
2. Use technology to find the critical value $z_{\alpha/2}$.

Procedure for Constructing a Confidence Interval for p

1. Verify that the following requirements are satisfied:
   - The sample is a simple random sample.
   - The conditions for the binomial distribution are satisfied.
   - There are at least 5 successes and at least 5 failures.
2. Use technology to find the critical value $z_{\alpha/2}$.
3. Evaluate the margin of error E.

Procedure for Constructing a Confidence Interval for p

1. Verify that the following requirements are satisfied:
    - The sample is a simple random sample.
    - The conditions for the binomial distribution are satisfied.
    - There are at least 5 successes and at least 5 failures.
2. Use technology to find the critical value $z_{\alpha/2}$.
3. Evaluate the margin of error E.
4. Using the value of the calculated margin of error E and the value of the sample proportion $\hat{p}$, find the values of the confidence interval limits $\hat{p} - E$ and $\hat{p} + E$.

Procedure for Constructing a Confidence Interval for p

1. Verify that the following requirements are satisfied:
   - The sample is a simple random sample.
   - The conditions for the binomial distribution are satisfied.
   - There are at least 5 successes and at least 5 failures.
2. Use technology to find the critical value $z_{\alpha/2}$.
3. Evaluate the margin of error E.
4. Using the value of the calculated margin of error E and the value of the sample proportion $\hat{p}$, find the values of the confidence interval limits $\hat{p} - E$ and $\hat{p} + E$.
5. Round the resulting confidence interval limits to three significant digits.

Procedure for Constructing a Confidence Interval for p

1. Verify that the following requirements are satisfied:
   - The sample is a simple random sample.
   - The conditions for the binomial distribution are satisfied.
   - There are at least 5 successes and at least 5 failures.
2. Use technology to find the critical value $z_{\alpha/2}$.
3. Evaluate the margin of error E.
4. Using the value of the calculated margin of error E and the value of the sample proportion $\hat{p}$, find the values of the confidence interval limits $\hat{p} - E$ and $\hat{p} + E$.
5. Round the resulting confidence interval limits to three significant digits.

Note

Statistics software, such as Statdisk, can calculate the confidence interval for you.

Example 3

A Gallup poll of 1487 adults showed that 43% of the respondents have Facebook pages. Let us calculate a 95% confidence interval by hand.

1. The critical value for 95% confidence interval is $z_{\alpha/2}$ is 1.96.

Example 3

A Gallup poll of 1487 adults showed that 43% of the respondents have Facebook pages. Let us calculate a 95% confidence interval by hand.

1. The critical value for 95% confidence interval is $z_{\alpha/2}$ is 1.96.
2. The margin of error is

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

Example 3

A Gallup poll of 1487 adults showed that 43% of the respondents have Facebook pages. Let us calculate a 95% confidence interval by hand.

1. The critical value for 95% confidence interval is $z_{\alpha/2}$ is 1.96.
2. The margin of error is

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96\sqrt{\frac{0.43(1-0.43)}{1487}}$$

Example 3

A Gallup poll of 1487 adults showed that 43% of the respondents have Facebook pages. Let us calculate a 95% confidence interval by hand.

1. The critical value for 95% confidence interval is $z_{\alpha/2}$ is 1.96.
2. The margin of error is

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96\sqrt{\frac{0.43(1-0.43)}{1487}} = 0.0251636$$

Example 3

A Gallup poll of 1487 adults showed that 43% of the respondents have Facebook pages. Let us calculate a 95% confidence interval by hand.

1. The critical value for 95% confidence interval is $z_{\alpha/2}$ is 1.96.
2. The margin of error is

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96\sqrt{\frac{0.43(1-0.43)}{1487}} = 0.0251636$$

3. The confidence interval is

$$\hat{p} - E \quad < \quad p \quad < \quad \hat{p} + E$$

Example 3

A Gallup poll of 1487 adults showed that 43% of the respondents have Facebook pages. Let us calculate a 95% confidence interval by hand.

1. The critical value for 95% confidence interval is $z_{\alpha/2}$ is 1.96.
2. The margin of error is

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96\sqrt{\frac{0.43(1-0.43)}{1487}} = 0.0251636$$

3. The confidence interval is

$$\begin{array}{ccccc} \hat{p} - E & < & p & < & \hat{p} + E \\ 0.43 - 0.0251636 & < & p & < & 0.43 + 0.0251636 \end{array}$$

Example 3

A Gallup poll of 1487 adults showed that 43% of the respondents have Facebook pages. Let us calculate a 95% confidence interval by hand.

1. The critical value for 95% confidence interval is $z_{\alpha/2}$ is 1.96.
2. The margin of error is

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96\sqrt{\frac{0.43(1-0.43)}{1487}} = 0.0251636$$

3. The confidence interval is

$$\begin{array}{ccccc} \hat{p}-E & < & p & < & \hat{p}+E \\ 0.43-0.0251636 & < & p & < & 0.43+0.0251636 \\ 0.405 & < & p & < & 0.455 \end{array}$$

Example 3

A Gallup poll of 1487 adults showed that 43% of the respondents have Facebook pages. Let us calculate a 95% confidence interval by hand.

1. The critical value for 95% confidence interval is $z_{\alpha/2}$ is 1.96.
2. The margin of error is

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96\sqrt{\frac{0.43(1-0.43)}{1487}} = 0.0251636$$

3. The confidence interval is

$$\begin{array}{ccccc} \hat{p} - E & < & p & < & \hat{p} + E \\ 0.43 - 0.0251636 & < & p & < & 0.43 + 0.0251636 \\ 0.405 & < & p & < & 0.455 \end{array}$$

Note

Looking at the confidence interval we can conclude that less than half of adults have a Facebook page.

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 limits, we can calculate the point estimate:

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

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

If you know the confidence interval limits, we can calculate the point estimate:

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

Finding E from a Confidence Interval

If you know the confidence interval limits, we can calculate the margin of error:

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

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

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

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

$$\begin{aligned}\hat{p} &= \frac{(\text{upper confidence interval limit}) + (\text{lower confidence interval limit})}{2} \\ &= \frac{0.81 + 0.58}{2}\end{aligned}$$

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

$$\begin{aligned}\hat{p} &= \frac{(\text{upper confidence interval limit}) + (\text{lower confidence interval limit})}{2} \\ &= \frac{0.81 + 0.58}{2} = 0.695\end{aligned}$$

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

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

The margin of error E is

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

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

$$\begin{aligned}\hat{p} &= \frac{\text{(upper confidence interval limit)} + \text{(lower confidence interval limit)}}{2} \\ &= \frac{0.81 + 0.58}{2} = 0.695\end{aligned}$$

The margin of error E is

$$\begin{aligned}E &= \frac{\text{(upper confidence interval limit)} - \text{(lower confidence interval limit)}}{2} \\ &= \frac{0.81 - 0.58}{2}\end{aligned}$$

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

$$\begin{aligned}\hat{p} &= \frac{(\text{upper confidence interval limit}) + (\text{lower confidence interval limit})}{2} \\ &= \frac{0.81 + 0.58}{2} = 0.695\end{aligned}$$

The margin of error E is

$$\begin{aligned}E &= \frac{(\text{upper confidence interval limit}) - (\text{lower confidence interval limit})}{2} \\ &= \frac{0.81 - 0.58}{2} = 0.115\end{aligned}$$

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:

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

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:

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

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:

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

We have $z_{\alpha/2} = 1.96$ 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:

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

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:

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

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

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

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:

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

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

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

Rounding

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

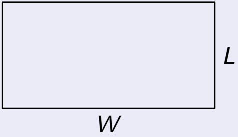
Why 0.25?

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

Why 0.25?

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

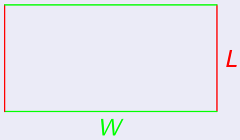
To see why, consider the rectangle:



Why 0.25?

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

To see why, consider the rectangle:



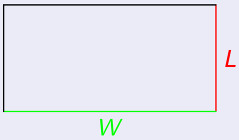
Let us assume that the perimeter of this rectangle is 2, which means:

$$2L + 2W = 2$$

Why 0.25?

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

To see why, consider the rectangle:



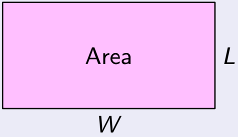
Let us assume that the perimeter of this rectangle is 2, which means:

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

Why 0.25?

If $\hat{p} = 0.5$ then the product $\hat{p}(1 - \hat{p}) = 0.25$ 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 the product $\hat{p}(1 - \hat{p}) = 0.25$ 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 $z_{\alpha/2} = 1.96$ 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 $z_{\alpha/2} = 1.96$ and $E = 0.03$, but we know nothing about $\hat{p}$.

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

Example 6

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

We have $z_{\alpha/2} = 1.96$ and $E = 0.03$, but we know nothing about $\hat{p}$.

$$n = \frac{(z_{\alpha/2})^2 0.25}{E^2} = \frac{(1.96)^2 0.25}{(0.03)^2}$$

Example 6

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

We have $z_{\alpha/2} = 1.96$ and $E = 0.03$, but we know nothing about $\hat{p}$.

$$n = \frac{(z_{\alpha/2})^2 0.25}{E^2} = \frac{(1.96)^2 0.25}{(0.03)^2} = 1067.11$$

Example 6

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

We have $z_{\alpha/2} = 1.96$ and $E = 0.03$, but we know nothing about $\hat{p}$.

$$n = \frac{(z_{\alpha/2})^2 0.25}{E^2} = \frac{(1.96)^2 0.25}{(0.03)^2} = 1067.11$$

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 $z_{\alpha/2} = 1.96$ and $E = 0.03$, but we know nothing about $\hat{p}$.

$$n = \frac{(z_{\alpha/2})^2 0.25}{E^2} = \frac{(1.96)^2 0.25}{(0.03)^2} = 1067.11$$

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 $z_{\alpha/2} = 1.96$ and $E = 0.03$, but we know nothing about $\hat{p}$.

$$n = \frac{(z_{\alpha/2})^2 0.25}{E^2} = \frac{(1.96)^2 0.25}{(0.03)^2} = 1067.11$$

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 $z_{\alpha/2}$ for the confidence level.

Example 6

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

We have $z_{\alpha/2} = 1.96$ and $E = 0.03$, but we know nothing about $\hat{p}$.

$$n = \frac{(z_{\alpha/2})^2 0.25}{E^2} = \frac{(1.96)^2 0.25}{(0.03)^2} = 1067.11$$

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 $z_{\alpha/2}$ for the confidence level.
3. Be sure to *round up to the next highest integer*, do not round using the usual rounding rules.

Note

We have been discussing what is called the **Wald** confidence interval.

Note

We have been discussing what is called the **Wald** confidence interval.

Plus Four Method

An easy improvement of the Wald Confidence Interval is:

1. Add two to the number of successes.
2. Add two to the number of failures
3. Calculate the Wald confidence interval.

Note

We have been discussing what is called the **Wald** confidence interval.

Plus Four Method

An easy improvement of the Wald Confidence Interval is:

1. Add two to the number of successes.
2. Add two to the number of failures
3. Calculate the Wald confidence interval.

Note

There are other types of confidence intervals:

- The Wilson score.
- The Clopper-Pearson Method.