Estimating a Population Mean

Colby Community College

Point Estimate

The sample mean $\bar{x}$ is the best point estimate of the population mean μ.

Point Estimate

The sample mean $\bar{x}$ is the best point estimate of the population mean μ.

Note

It's rare to need to estimate the population mean μ but we somehow know the population standard deviation σ. Let us assume we don't know σ.

Point Estimate

The sample mean $\bar{x}$ is the best point estimate of the population mean μ.

Note

It's rare to need to estimate the population mean μ but we somehow know the population standard deviation σ. Let us assume we don't know σ.

Confidence Interval Requirements

1. The sample is a simple random sample.

Point Estimate

The sample mean $\bar{x}$ is the best point estimate of the population mean μ.

Note

It's rare to need to estimate the population mean μ but we somehow know the population standard deviation σ. Let us assume we don't know σ.

Confidence Interval Requirements

1. The sample is a simple random sample.
2. The population is normally distributed or $n > 30$.

Point Estimate

The sample mean $\bar{x}$ is the best point estimate of the population mean μ.

Note

It's rare to need to estimate the population mean μ but we somehow know the population standard deviation σ. Let us assume we don't know σ.

Confidence Interval Requirements

1. The sample is a simple random sample.
2. The population is normally distributed or $n > 30$.
   - The method we will use is robust against departure from normality.

Point Estimate

The sample mean $\bar{x}$ is the best point estimate of the population mean μ.

Note

It's rare to need to estimate the population mean μ but we somehow know the population standard deviation σ. Let us assume we don't know σ.

Confidence Interval Requirements

1. The sample is a simple random sample.
2. The population is normally distributed or $n > 30$.
   - The method we will use is robust against departure from normality.
   - If the distribution is approximately normal, a sample size of 15 to 30 may be acceptable.

Definition

If a population has a normal distribution, then the distribution of

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

is a **Student *t* distribution** for all sample sizes n.

Definition

If a population has a normal distribution, then the distribution of

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

is a **Student *t* distribution** for all sample sizes n.

Note

A Student t distribution is commonly called a ***t* distribution**.

Definition

The **degrees of freedom** (or **df**) for a collection of sample data is the number of sample values that can vary after certain restrictions have been imposed on all data values.

For this section use one less than the sample size:

$$df = n - 1$$

Definition

The **degrees of freedom** (or **df**) for a collection of sample data is the number of sample values that can vary after certain restrictions have been imposed on all data values.

For this section use one less than the sample size:

$$df = n - 1$$

Example 1

If 10 test scores must have mean 80, then their sum must be 800.

Definition

The **degrees of freedom** (or **df**) for a collection of sample data is the number of sample values that can vary after certain restrictions have been imposed on all data values.

For this section use one less than the sample size:

$$df = n - 1$$

Example 1

If 10 test scores must have mean 80, then their sum must be 800.

We can freely assign values to the first 9 scores, but the 10th score would need to be:

$$\text{score}_{10} = 800 - \text{score}_1 - \text{score}_2 - \text{score}_3 - \cdots - \text{score}_8 - \text{score}_9$$

Definition

The **degrees of freedom** (or **df**) for a collection of sample data is the number of sample values that can vary after certain restrictions have been imposed on all data values.

For this section use one less than the sample size:

$$df = n - 1$$

Example 1

If 10 test scores must have mean 80, then their sum must be 800.

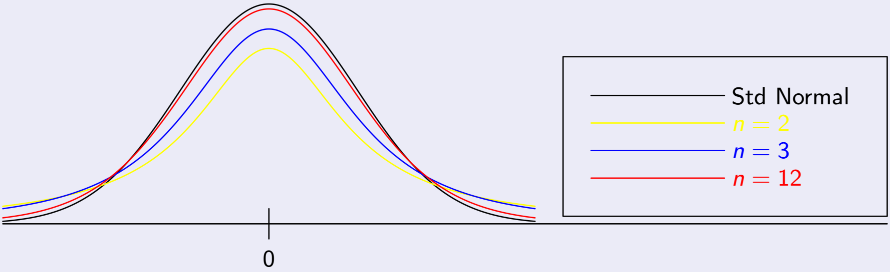
We can freely assign values to the first 9 scores, but the 10th score would need to be:

$$\text{score}_{10} = 800 - \text{score}_1 - \text{score}_2 - \text{score}_3 - \cdots - \text{score}_8 - \text{score}_9$$

Hence 9 degrees of freedom.

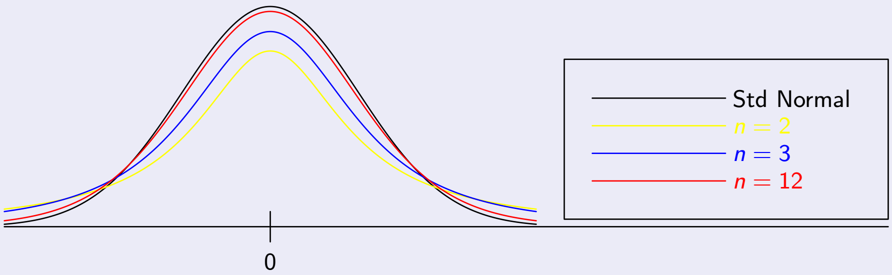
Note

The Student t distribution is different for different sample sizes.



Note

The Student t distribution is different for different sample sizes.



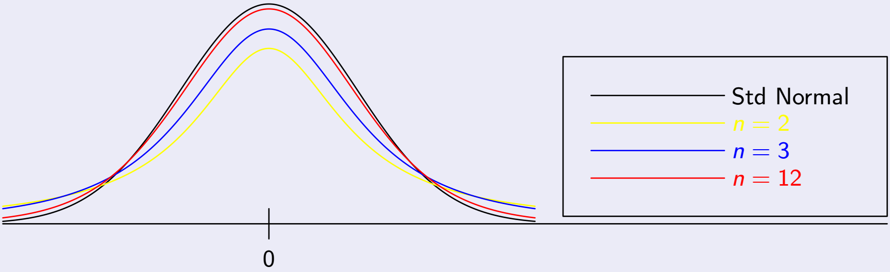
Note

The Student t distribution has a mean of $t = 0$

The standard deviation varies with the sample size, but is greater than 1.

Note

The Student t distribution is different for different sample sizes.



Note

The Student t distribution has a mean of $t = 0$

The standard deviation varies with the sample size, but is greater than 1.

Note

As the sample size n gets larger, the Student t distribution gets closer to the standard normal distribution.

Example 2

Let us find the critical value corresponding to a 95% confidence level, given that the sample size is $n = 15$.

Example 2

Let us find the critical value corresponding to a 95% confidence level, given that the sample size is $n = 15$.

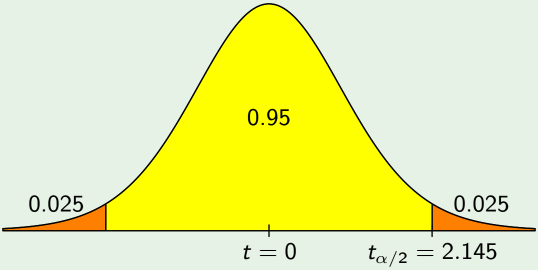
The degrees of freedom is $n - 1 = 15 - 1 = 14$.

Example 2

Let us find the critical value corresponding to a 95% confidence level, given that the sample size is $n = 15$.

The degrees of freedom is $n - 1 = 15 - 1 = 14$.

We can then use technology to find the critical value.



Procedure for Constructing a Confidence Interval for μ

1. Verify that the following requirements are satisfied:
    - The sample is a simple random sample.
    - Population is normally distributed or $n > 30$.

Procedure for Constructing a Confidence Interval for μ

1. Verify that the following requirements are satisfied:
   - The sample is a simple random sample.
   - Population is normally distributed or $n > 30$.
2. Use technology to find the critical value $t_{\alpha/2}$.
   - When σ is unknown, use $n - 1$ degrees of freedom.

Procedure for Constructing a Confidence Interval for μ

1. Verify that the following requirements are satisfied:
   - The sample is a simple random sample.
   - Population is normally distributed or $n > 30$.
2. Use technology to find the critical value $t_{\alpha/2}$.
   - When σ is unknown, use $n - 1$ degrees of freedom.
3. Calculate the margin of error.

$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

Procedure for Constructing a Confidence Interval for μ

1. Verify that the following requirements are satisfied:
   - The sample is a simple random sample.
   - Population is normally distributed or $n > 30$.
2. Use technology to find the critical value $t_{\alpha/2}$.
   - When σ is unknown, use $n - 1$ degrees of freedom.
3. Calculate the margin of error.

$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

4. Using the value of the calculated margin of error E and the value of the sample mean $\bar{x}$, find the values of the confidence interval limits $\bar{x} - E$ and $\bar{x} + E$.

Procedure for Constructing a Confidence Interval for μ

1. Verify that the following requirements are satisfied:
   - The sample is a simple random sample.
   - Population is normally distributed or $n > 30$.
2. Use technology to find the critical value $t_{\alpha/2}$.
   - When σ is unknown, use $n - 1$ degrees of freedom.
3. Calculate the margin of error.

$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

4. Using the value of the calculated margin of error E and the value of the sample mean $\bar{x}$, find the values of the confidence interval limits $\bar{x} - E$ and $\bar{x} + E$.
5. Round the resulting confidence interval limits:
   - With an original data set, round to three significant digits.
   - When using summary statistics, round to the same number of decimal places.

Example 3

The weights (in hectograms, hg) of randomly selected girls at birth are

33	28	33	37	31	32	31	28	34	28	33	26	30	31	28

(Based on data from the National Center for Health Statistics.)

Example 3

The weights (in hectograms, hg) of randomly selected girls at birth are

33	28	33	37	31	32	31	28	34	28	33	26	30	31	28

(Based on data from the National Center for Health Statistics.)

The summary statistics for this sample are

$$n = 15 \qquad \bar{x} = 30.9 \qquad s = 2.9$$

Example 3

The weights (in hectograms, hg) of randomly selected girls at birth are

33	28	33	37	31	32	31	28	34	28	33	26	30	31	28

(Based on data from the National Center for Health Statistics.)

The summary statistics for this sample are

$$n = 15 \qquad \bar{x} = 30.9 \qquad s = 2.9$$

Let's construct a 95% confidence interval for the mean birth weight of girls.

Example 3

The weights (in hectograms, hg) of randomly selected girls at birth are

33	28	33	37	31	32	31	28	34	28	33	26	30	31	28

(Based on data from the National Center for Health Statistics.)

The summary statistics for this sample are

$$n = 15 \qquad \bar{x} = 30.9 \qquad s = 2.9$$

Let's construct a 95% confidence interval for the mean birth weight of girls.

The margin of error is

$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

Example 3

The weights (in hectograms, hg) of randomly selected girls at birth are

33 28 33 37 31 32 31 28 34 28 33 26 30 31 28

(Based on data from the National Center for Health Statistics.)

The summary statistics for this sample are

$$n = 15 \qquad \bar{x} = 30.9 \qquad s = 2.9$$

Let's construct a 95% confidence interval for the mean birth weight of girls.

The margin of error is

$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 2.145 \cdot \frac{2.9}{\sqrt{15}}$$

Example 3

The weights (in hectograms, hg) of randomly selected girls at birth are

33	28	33	37	31	32	31	28	34	28	33	26	30	31	28

(Based on data from the National Center for Health Statistics.)

The summary statistics for this sample are

$$n = 15 \qquad \bar{x} = 30.9 \qquad s = 2.9$$

Let's construct a 95% confidence interval for the mean birth weight of girls.

The margin of error is

$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 2.145 \cdot \frac{2.9}{\sqrt{15}} = 1.606126$$

Example 3

The weights (in hectograms, hg) of randomly selected girls at birth are

33	28	33	37	31	32	31	28	34	28	33	26	30	31	28

(Based on data from the National Center for Health Statistics.)

The summary statistics for this sample are

$$n = 15 \qquad \bar{x} = 30.9 \qquad s = 2.9$$

Let's construct a 95% confidence interval for the mean birth weight of girls.

The margin of error is

$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 2.145 \cdot \frac{2.9}{\sqrt{15}} = 1.606126$$

The confidence interval is

$$\bar{x} - E \qquad < \quad \mu \quad < \qquad \bar{x} + E$$

Example 3

The weights (in hectograms, hg) of randomly selected girls at birth are

33	28	33	37	31	32	31	28	34	28	33	26	30	31	28

(Based on data from the National Center for Health Statistics.)

The summary statistics for this sample are

$$n = 15 \qquad \bar{x} = 30.9 \qquad s = 2.9$$

Let's construct a 95% confidence interval for the mean birth weight of girls.

The margin of error is

$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 2.145 \cdot \frac{2.9}{\sqrt{15}} = 1.606126$$

The confidence interval is

$$\begin{array}{rcccl} \bar{x} - E & < & \mu & < & \bar{x} + E \\ 30.9 - 1.606126 & < & \mu & < & 30.9 + 1.606126 \end{array}$$

Example 3

The weights (in hectograms, hg) of randomly selected girls at birth are

33	28	33	37	31	32	31	28	34	28	33	26	30	31	28

(Based on data from the National Center for Health Statistics.)

The summary statistics for this sample are

$$n = 15 \qquad \bar{x} = 30.9 \qquad s = 2.9$$

Let's construct a 95% confidence interval for the mean birth weight of girls.

The margin of error is

$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 2.145 \cdot \frac{2.9}{\sqrt{15}} = 1.606126$$

The confidence interval is

$$\begin{array}{ccccc} \bar{x} - E & < & \mu & < & \bar{x} + E \\ 30.9 - 1.606126 & < & \mu & < & 30.9 + 1.606126 \\ 29.3 \text{ hg} & < & \mu & < & 32.5 \text{ hg} \end{array}$$

Example 3

The weights (in hectograms, hg) of randomly selected girls at birth are

33	28	33	37	31	32	31	28	34	28	33	26	30	31	28

(Based on data from the National Center for Health Statistics.)

The summary statistics for this sample are

$$n = 15 \qquad \bar{x} = 30.9 \qquad s = 2.9$$

Let's construct a 95% confidence interval for the mean birth weight of girls.

The margin of error is

$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 2.145 \cdot \frac{2.9}{\sqrt{15}} = 1.606126$$

The confidence interval is

$$\begin{array}{ccccc} \bar{x} - E & < & \mu & < & \bar{x} + E \\ 30.9 - 1.606126 & < & \mu & < & 30.9 + 1.606126 \\ 29.3 \text{ hg} & < & \mu & < & 32.5 \text{ hg} \end{array}$$

We are 95% confident that the interval from 29.2 hg to 32.5 hg actually does contain the true value of μ.

Finding $\bar{x}$ from a Confidence Interval

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

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

Finding $\bar{x}$ from a Confidence Interval

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

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

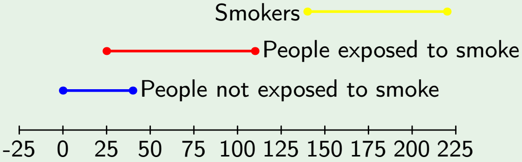
Finding E from a Confidence Interval

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

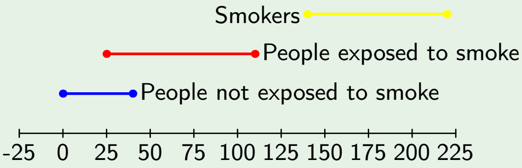
We can compare the confidence intervals of the mean cotinine level in each of three samples (Data Set 12).



Because cotinine is produced in the body when nicotine is absorbed, cotinine is a good indication of nicotine intake.

Example 4

We can compare the confidence intervals of the mean cotinine level in each of three samples (Data Set 12).

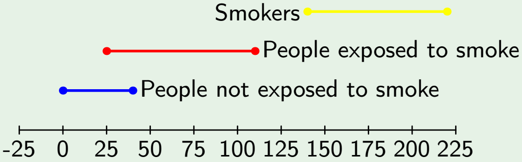


Because cotinine is produced in the body when nicotine is absorbed, cotinine is a good indication of nicotine intake.

We see that the confidence interval for smokers does not overlap the other confidence intervals, so it appears that the mean cotinine level of smokers is different from that of the other two groups.

Example 4

We can compare the confidence intervals of the mean cotinine level in each of three samples (Data Set 12).



Because cotinine is produced in the body when nicotine is absorbed, cotinine is a good indication of nicotine intake.

We see that the confidence interval for smokers does not overlap the other confidence intervals, so it appears that the mean cotinine level of smokers is different from that of the other two groups.

The two non-smoking groups have overlapping confidence intervals, so it possible that they have the same mean cotinine level.

Caution

Confidence intervals can be used to *informally* compare different data sets, but the overlapping of confidence intervals *should not* be used for making formal and final conclusions about equality of means.

Caution

Confidence intervals can be used to *informally* compare different data sets, but the overlapping of confidence intervals *should not* be used for making formal and final conclusions about equality of means.

Sample Size Required to Estimate a Population Mean

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

Caution

Confidence intervals can be used to *informally* compare different data sets, but the overlapping of confidence intervals *should not* be used for making formal and final conclusions about equality of means.

Sample Size Required to Estimate a Population Mean

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

We must know, or have estimated, the population standard deviation σ.

Caution

Confidence intervals can be used to *informally* compare different data sets, but the overlapping of confidence intervals *should not* be used for making formal and final conclusions about equality of means.

Sample Size Required to Estimate a Population Mean

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

We must know, or have estimated, the population standard deviation σ.

The required sample size is

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

(Recall that $z_{\alpha/2}$ is the critical value for the standard normal distribution.)

Caution

Confidence intervals can be used to *informally* compare different data sets, but the overlapping of confidence intervals *should not* be used for making formal and final conclusions about equality of means.

Sample Size Required to Estimate a Population Mean

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

We must know, or have estimated, the population standard deviation σ.

The required sample size is

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

(Recall that $z_{\alpha/2}$ is the critical value for the standard normal distribution.)

Rounding

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

Dealing with Unknown σ When Finding Sample Size

1. Use the range rule of thumb:

$$\sigma \approx \text{range}/4$$

Dealing with Unknown σ When Finding Sample Size

1. Use the range rule of thumb:

$$\sigma \approx \text{range}/4$$

2. Start the sample collection process without knowing σ and, using the first several values, calculate the sample standard deviation s and use it in place of σ.

Dealing with Unknown σ When Finding Sample Size

1. Use the range rule of thumb:

$$\sigma \approx \text{range}/4$$

2. Start the sample collection process without knowing σ and, using the first several values, calculate the sample standard deviation s and use it in place of σ.
   - The estimate for σ can be improved as more sample data are collected, and the required sample size can be adjusted along the way.

Dealing with Unknown σ When Finding Sample Size

1. Use the range rule of thumb:

$$\sigma \approx \text{range}/4$$

2. Start the sample collection process without knowing σ and, using the first several values, calculate the sample standard deviation s and use it in place of σ.
   - The estimate for σ can be improved as more sample data are collected, and the required sample size can be adjusted along the way.
3. Estimate the value of σ by using the results of some other earlier study.

Dealing with Unknown σ When Finding Sample Size

1. Use the range rule of thumb:

$$\sigma \approx \text{range}/4$$

2. Start the sample collection process without knowing σ and, using the first several values, calculate the sample standard deviation s and use it in place of σ.
   - The estimate for σ can be improved as more sample data are collected, and the required sample size can be adjusted along the way.
3. Estimate the value of σ by using the results of some other earlier study.

Caution

When determining the sample size n, any errors should always be conservative in the sense that they make the sample size too large instead of too small.

Example 5

Suppose that we want the mean IQ for the population of all Colby students.

Example 5

Suppose that we want the mean IQ for the population of all Colby students.

How many Colby students must be randomly selected for IQ tests if we want 95% confidence that the sample mean is within 3 IQ points of the population mean?

Example 5

Suppose that we want the mean IQ for the population of all Colby students.

How many Colby students must be randomly selected for IQ tests if we want 95% confidence that the sample mean is within 3 IQ points of the population mean?

For a 95% confidence interval, we have $\alpha = 0.5$, so $z_{\alpha/2} = 1.96$.

Example 5

Suppose that we want the mean IQ for the population of all Colby students.

How many Colby students must be randomly selected for IQ tests if we want 95% confidence that the sample mean is within 3 IQ points of the population mean?

For a 95% confidence interval, we have $\alpha = 0.5$, so $z_{\alpha/2} = 1.96$.

Our desired margin of error is $E = 3$.

Example 5

Suppose that we want the mean IQ for the population of all Colby students.

How many Colby students must be randomly selected for IQ tests if we want 95% confidence that the sample mean is within 3 IQ points of the population mean?

For a 95% confidence interval, we have $\alpha = 0.5$, so $z_{\alpha/2} = 1.96$.

Our desired margin of error is $E = 3$.

Wechsler IQ tests are designed so that the standard deviation is 15.

Example 5

Suppose that we want the mean IQ for the population of all Colby students.

How many Colby students must be randomly selected for IQ tests if we want 95% confidence that the sample mean is within 3 IQ points of the population mean?

For a 95% confidence interval, we have $\alpha = 0.5$, so $z_{\alpha/2} = 1.96$.

Our desired margin of error is $E = 3$.

Wechsler IQ tests are designed so that the standard deviation is 15.

While we don't know the σ of all Colby students, we can safely assume that they are a more homogeneous group than the general population, and hence have a standard deviation less than 15. Thus, we can use $\sigma = 15$.

Example 5

Suppose that we want the mean IQ for the population of all Colby students.

How many Colby students must be randomly selected for IQ tests if we want 95% confidence that the sample mean is within 3 IQ points of the population mean?

For a 95% confidence interval, we have $\alpha = 0.5$, so $z_{\alpha/2} = 1.96$.

Our desired margin of error is $E = 3$.

Wechsler IQ tests are designed so that the standard deviation is 15.

While we don't know the σ of all Colby students, we can safely assume that they are a more homogeneous group than the general population, and hence have a standard deviation less than 15. Thus, we can use $\sigma = 15$. Using the formula for sample size we get

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

Example 5

Suppose that we want the mean IQ for the population of all Colby students.

How many Colby students must be randomly selected for IQ tests if we want 95% confidence that the sample mean is within 3 IQ points of the population mean?

For a 95% confidence interval, we have $\alpha = 0.5$, so $z_{\alpha/2} = 1.96$.

Our desired margin of error is $E = 3$.

Wechsler IQ tests are designed so that the standard deviation is 15.

While we don't know the σ of all Colby students, we can safely assume that they are a more homogeneous group than the general population, and hence have a standard deviation less than 15. Thus, we can use $\sigma = 15$. Using the formula for sample size we get

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2 = \left(\frac{1.96 \cdot 15}{3}\right)^2$$

Example 5

Suppose that we want the mean IQ for the population of all Colby students.

How many Colby students must be randomly selected for IQ tests if we want 95% confidence that the sample mean is within 3 IQ points of the population mean?

For a 95% confidence interval, we have $\alpha = 0.5$, so $z_{\alpha/2} = 1.96$.

Our desired margin of error is $E = 3$.

Wechsler IQ tests are designed so that the standard deviation is 15.

While we don't know the σ of all Colby students, we can safely assume that they are a more homogeneous group than the general population, and hence have a standard deviation less than 15. Thus, we can use $\sigma = 15$. Using the formula for sample size we get

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2 = \left(\frac{1.96 \cdot 15}{3}\right)^2 = 96.04$$

Example 5

Suppose that we want the mean IQ for the population of all Colby students.

How many Colby students must be randomly selected for IQ tests if we want 95% confidence that the sample mean is within 3 IQ points of the population mean?

For a 95% confidence interval, we have $\alpha = 0.5$, so $z_{\alpha/2} = 1.96$.

Our desired margin of error is $E = 3$.

Wechsler IQ tests are designed so that the standard deviation is 15.

While we don't know the σ of all Colby students, we can safely assume that they are a more homogeneous group than the general population, and hence have a standard deviation less than 15. Thus, we can use $\sigma = 15$. Using the formula for sample size we get

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2 = \left(\frac{1.96 \cdot 15}{3}\right)^2 = 96.04$$

To be 95% confident that out interval contains the true population standard deviation, we would need a sample size of at least 97.

Note

If we somehow know the population standard deviation but do not know the population mean, then we calculate the confidence interval using the methods in section 7.1.

Note

If we somehow know the population standard deviation but do not know the population mean, then we calculate the confidence interval using the methods in section 7.1.

Choosing the Appropriate Distribution

Conditions	**Method**
σ not known and normal population or σ not known and $n > 30$	Student t distribution
σ known and normal population or σ known and $n > 30$	Normal distribution
Population is not normal and $n \leq 30$	Use other methods.