7.1: Large Sample Estimation of a Population Mean

skills to develop

- To become familiar with the concept of an interval estimate of the population mean.
- To understand how to apply formulas for a confidence interval for a population mean.

The Central Limit Theorem says that, for large samples (samples of size \(n \geq 30\)), when viewed as a random variable the sample mean \(\overline{X}\) is normally distributed with mean \(\mu_{\overline{X}} = \mu\) and standard deviation \(\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\). The Empirical Rule says that we must go about two standard deviations from the mean to capture \(95\%\) of the values of \(\overline{X}\) generated by sample after sample. A more precise distance based on the normality of \(\overline{X}\) is \((1.960)\) standard deviations, which is \(E = \frac{1.960 \sigma}{\sqrt{n}}\).

The key idea in the construction of the \((95\%)\) confidence interval is this, as illustrated in Figure \(\PageIndex{1}\), because in sample after sample \((95\%)\) of the values of \(\overline{X}\) lie in the interval \([(\mu - E, \mu + E)]\), if we adjoin to each side of the point estimate \((x-a)\) “wing” of length \(E\), \((95\%)\) of the intervals formed by the winged dots contain \(\mu\). The \((95\%)\) confidence interval is thus \((\bar{x} \pm 1.960\frac{\sigma}{\sqrt{n}})\). For a different level of confidence, say \((90\%)\) or \((99\%)\), the number \((1.960)\) will change, but the idea is the same.

![Figure \(\PageIndex{1}\): When Winged Dots Capture the Population Mean](image)

Figure \(\PageIndex{2}\) shows the intervals generated by a computer simulation of drawing \((40)\) samples from a
normally distributed population and constructing the \((95\%)\) confidence interval for each one. We expect that about \((0.05)(40)=2\) of the intervals so constructed would fail to contain the population mean \(\mu\), and in this simulation two of the intervals, shown in red, do.

![Figure \(\PageIndex{2}\): Computer Simulation of \((40\%)\) \((95\%)\) Confidence Intervals for a Mean](https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Book%3A_Introductory_Statistics_(Shafer_and_Zhang)/07%3A…)

It is standard practice to identify the level of confidence in terms of the area \(\alpha\) in the two tails of the distribution of \(\overline{X}\) when the middle part specified by the level of confidence is taken out. This is shown in Figure \(\PageIndex{3}\), drawn for the general situation, and in Figure \(\PageIndex{4}\), drawn for \((95\%)\) confidence.

![Figure \(\PageIndex{3}\): For \((100(1-\alpha))\%\) confidence the area in each tail is \(\alpha/2\)](https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Book%3A_Introductory_Statistics_(Shafer_and_Zhang)/07%3A…)

Remember from Section 5.4 that the \(z\)-value that cuts off a right tail of area \(\alpha/2\) is denoted \(z_{\alpha/2}\). Thus the number \(1.960\) in the example is \(z_{0.025}\), which is \(z_{\alpha/2}\) for \(\alpha = 1-0.95 = 0.05\).
For \(95\%\) confidence the area in each tail is \(\alpha/2 = 0.025\).

The level of confidence can be any number between \(0\) and \(100\%\), but the most common values are probably \(90\%\) \((\alpha = 0.10)\), \(95\%\) \((\alpha = 0.05)\), and \(99\%\) \((\alpha = 0.01)\).

Thus in general for a \(100(1-\alpha)\%\) confidence interval, \(E = z_{\alpha/2}(\sigma/\sqrt{n})\), so the formula for the confidence interval is \(\bar{x} \pm z_{\alpha/2}(\sigma/\sqrt{n})\). While sometimes the population standard deviation \(\sigma\) is known, typically it is not. If not, for \(n \geq 30\) it is generally safe to approximate \(\sigma\) by the sample standard deviation \(s\).

Large Sample \(100(1-\alpha)\%\) Confidence Interval for a Population Mean

- If \(\sigma\) is known: \(\bar{x} \pm z_{\alpha/2}(\sigma/\sqrt{n})\)
- If \(\sigma\) is unknown: \(\bar{x} \pm z_{\alpha/2}(s/\sqrt{n})\)

A sample is considered large when \(n \geq 30\).

As mentioned earlier, the number
\[E = z_{\alpha/2}(\sigma/\sqrt{n})\]

or
\[E = z_{\alpha/2}(s/\sqrt{n})\]

is called the margin of error of the estimate.

Example \(\PageIndex{1}\)

Find the number \(z_{\alpha/2}\) needed in construction of a confidence interval:

1. when the level of confidence is \(90\%\);
2. when the level of confidence is \(99\%\).

using the tables in Figure \(\PageIndex{5}\) below.
Solution:

1. For confidence level \( \alpha = 0.90 \), \( z_{\alpha/2} = z_{0.05} \). Since the area under the standard normal curve to the right of \( z_{0.05} \) is 0.05, the area to the left of \( z_{0.05} \) is 0.95. We search for the area 0.9500 in Figure \( \PageIndex{5} \). The closest entries in the table are 0.9495 and 0.9505, corresponding to \( z \)-values 1.64 and 1.65. Since 0.95 is halfway between 0.9495 and 0.9505 we use the average 1.645 of the \( z \)-values for \( z_{0.05} \).

2. For confidence level \( \alpha = 0.01 \), \( z_{\alpha/2} = z_{0.005} \). Since the area under the standard normal curve to the right of \( z_{0.005} \) is 0.01, the area to the left of \( z_{0.005} \) is 0.9950. We search for the area 0.9950 in Figure \( \PageIndex{5} \). The closest entries in the table are 0.9949 and 0.9951, corresponding to \( z \)-values 2.57 and 2.58. Since 0.995 is halfway between 0.9949 and 0.9951 we use the average 2.575 of the \( z \)-values for \( z_{0.005} \).

Example \( \PageIndex{2} \)

Use Figure \( \PageIndex{5} \) below to find the number \( z_{\alpha/2} \) needed in construction of a confidence interval:

1. when the level of confidence is \( \alpha = 0.90 \);
2. when the level of confidence is \( \alpha = 0.01 \).
### Critical Values of $t$

<table>
<thead>
<tr>
<th>df</th>
<th>0.200</th>
<th>0.100</th>
<th>0.050</th>
<th>0.025</th>
<th>0.010</th>
<th>0.005</th>
<th>0.0025</th>
<th>0.001</th>
<th>0.0005</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.376</td>
<td>2.062</td>
<td>2.571</td>
<td>3.290</td>
<td>4.032</td>
<td>4.842</td>
<td>5.991</td>
<td>6.632</td>
<td>7.173</td>
</tr>
<tr>
<td>4</td>
<td>2.431</td>
<td>3.000</td>
<td>3.446</td>
<td>4.115</td>
<td>4.821</td>
<td>5.847</td>
<td>6.849</td>
<td>7.940</td>
<td>8.926</td>
</tr>
</tbody>
</table>

---

https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Book%3A_Introductory_Statistics_(Shafer_and_Zhang)/07%3A...
Solution:

1. In the next section we will learn about a continuous random variable that has a probability distribution called the Student \(t\)-distribution. Figure \(\PageIndex{6}\) gives the value \(t_c\) that cuts off a right tail of area \(c\) for different values of \(c\). The last line of that table, the one whose heading is the symbol \(\infty\) for infinity and \(z\), gives the corresponding \(z\)-value \(z_c\) that cuts off a right tail of the same area \(c\). In particular, \(z_{0.05}\) is the number in that row and in the column with the heading \(t_{0.05}\). We read off directly that

<table>
<thead>
<tr>
<th>(df)</th>
<th>0.020</th>
<th>0.010</th>
<th>0.005</th>
<th>0.0025</th>
<th>0.001</th>
<th>0.0005</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.849</td>
<td>1.298</td>
<td>1.675</td>
<td>2.008</td>
<td>2.402</td>
<td>2.676</td>
</tr>
<tr>
<td>10</td>
<td>0.849</td>
<td>1.298</td>
<td>1.675</td>
<td>2.007</td>
<td>2.400</td>
<td>2.674</td>
</tr>
<tr>
<td>20</td>
<td>0.849</td>
<td>1.298</td>
<td>1.674</td>
<td>2.006</td>
<td>2.399</td>
<td>2.672</td>
</tr>
<tr>
<td>30</td>
<td>0.848</td>
<td>1.297</td>
<td>1.674</td>
<td>2.005</td>
<td>2.397</td>
<td>2.670</td>
</tr>
<tr>
<td>50</td>
<td>0.848</td>
<td>1.297</td>
<td>1.673</td>
<td>2.004</td>
<td>2.396</td>
<td>2.668</td>
</tr>
<tr>
<td>100</td>
<td>0.848</td>
<td>1.297</td>
<td>1.673</td>
<td>2.003</td>
<td>2.395</td>
<td>2.667</td>
</tr>
</tbody>
</table>

https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Book%3A_Introductory_Statistics_(Shafer_and_Zhang)/07%3A...
\(z_{0.05}=1.645\).

2. In Figure \(\PageIndex{6}\) \(z_{0.005}\) is the number in the last row and in the column headed \(t_{0.005}\), namely \(2.576\).

Figure \(\PageIndex{6}\) can be used to find \(z_c\) only for those values of \(c\) for which there is a column with the heading \(t_c\) appearing in the table; otherwise we must use Figure \(\PageIndex{5}\) in reverse. But when it can be done it is both faster and more accurate to use the last line of Figure \(\PageIndex{6}\) to find \(z_c\) than it is to do so using Figure \(\PageIndex{5}\) in reverse.

Example \(\PageIndex{3}\)

A sample of size \(49\) has sample mean \(35\) and sample standard deviation \(14\). Construct a \(98\%\) confidence interval for the population mean using this information. Interpret its meaning.

Solution:

For confidence level \(98\%\), \(\alpha =1-0.98=0.02\), so \(z_{\alpha/2}=z_{0.01}\). From Figure \(\PageIndex{6}\) we read directly that \(z_{0.01}=2.326\). Thus

\[
\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}} = 35 \pm 2.326 \left( \frac{14}{\sqrt{49}} \right) = 35 \pm 4.652 \approx 35 \pm 4.7
\]

We are \(98\%\) confident that the population mean \(\mu\) lies in the interval \((30.3, 39.7)\), in the sense that in repeated sampling \(98\%\) of all intervals constructed from the sample data in this manner will contain \(\mu\).

Example \(\PageIndex{4}\)

A random sample of \(120\) students from a large university yields mean GPA \(2.71\) with sample standard deviation \(0.51\). Construct a \(90\%\) confidence interval for the mean GPA of all students at the university.

Solution:

For confidence level \(90\%\), \(\alpha =1-0.90=0.10\), so \(z_{\alpha/2}=z_{0.05}\). From Figure \(\PageIndex{6}\) we read directly that \(z_{0.05}=1.645\). Since \(n=120\), \(\bar{x}=2.71\), and \(s=0.51\),

\[
\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}} = 2.71 \pm 1.645 \left( \frac{0.51}{\sqrt{120}} \right) = 2.71 \pm 0.0766
\]

One may be \(90\%\) confident that the true average GPA of all students at the university is contained in the interval \((2.63, 2.79)\).

Key Takeaway

- A confidence interval for a population mean is an estimate of the population mean together with an indication of reliability.
- There are different formulas for a confidence interval based on the sample size and whether or not the population standard deviation is known.
- The confidence intervals are constructed entirely from the sample data (or sample data and the population standard deviation, when it is known).
Contributor

- Anonymous