**CH7 Section 1: Confidence Intervals for the Mean When $\sigma$ Is Known**

**Point Estimate** – a specific numerical value used as an estimate of a parameter.

- The sample mean $\bar{X}$ is often the best point estimate of the population mean $\mu$ since the means of samples often vary less than sample medians or modes.

**Estimator** – a specific statistic used to estimate a population parameter.
Three Properties of a Good Estimator

1. The estimator should be an **unbiased estimator**.
   ⇒ The mean of the point estimates obtained from all samples of a given size is equal to the parameter being estimated.

2. The estimator should be a **consistent estimator**.
   ⇒ As the sample size increases, the value of the estimator should approach the value of the parameter being estimated.

3. The estimator should be **relatively efficient**.
   ⇒ The variance of the estimator should be the smallest of all comparable estimators.
**Interval Estimate** – An interval or range of values used to estimate a parameter.

- The interval may or may not contain the true value of the parameter (but we hope it does).
- *We can use statistical knowledge to assign a degree of confidence that our interval contains the true value of the population parameter.*

- A point estimate uses a specific value to estimate a population parameter.
- An interval estimate uses an interval of likely values to estimate the value of the population parameter.
Confidence level – the proportion of interval estimates that will contain the true value of the population parameter when we consider all possible samples of a fixed size $n$ from the population.

- This number doesn’t refer to a specific interval, but to the proportion of intervals that will contain the population parameter.
- We can actually choose this number and construct our interval accordingly. The number is typically chosen to be close to 1, most commonly 0.95 or 95%

Confidence interval – a specific interval estimate of a parameter determined by using data obtained from a sample and by using the specific confidence level of the estimate.
How can we construct a confidence interval?

Suppose we have a population with a normal distribution or we select samples with at least 30 observations. What do we know about the distribution of the sample means?

*Hint:* What are the properties of the sample mean? What does the Central Limit Theorem say?
What is the distribution of \( \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \)?

\[
P\left( -1.96 \leq \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \leq 1.96 \right) =
\]

Do a little algebra . . .

\[
P\left( -1.96 \leq \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \leq 1.96 \right) = P\left( -1.96 \cdot \sigma / \sqrt{n} \leq \bar{X} - \mu \leq 1.96 \cdot \sigma / \sqrt{n} \right)
\]

\[
= P\left( -1.96 \cdot \sigma / \sqrt{n} \leq \mu - \bar{X} \leq 1.96 \cdot \sigma / \sqrt{n} \right)
\]

\[
= P\left( \bar{X} - 1.96 \cdot \sigma / \sqrt{n} \leq \mu \leq \bar{X} + 1.96 \cdot \sigma / \sqrt{n} \right)
\]
We are 95% confident that the interval

\[
(\bar{X} - 1.96 \cdot \sigma / \sqrt{n}, \quad \bar{X} + 1.96 \cdot \sigma / \sqrt{n})
\]

will contain the value of the population mean.

- 95% of the intervals constructed using this method from random samples of the population will contain the mean.
- A specific interval either will contain the population parameter or it will not.
- The probability of making an error is 5%.

If we want a 90% confidence interval or a 99% confidence interval then we should change 1.96 to the appropriate number.
Let \( z_{\alpha/2} \) be the percentile such that \( P(Z \leq z_{\alpha/2}) = 1 - \alpha/2 \) (which is the same as \( P(Z > z_{\alpha/2}) = \alpha/2 \)). i.e., \( z_{\alpha/2} \) is the \((1 - \alpha/2) \times 100\%\) percentile of the standard normal distribution.

What are other common values of \( z_{\alpha/2} \)?

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>Error Probability(( \alpha ))</th>
<th>( z_{\alpha/2} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>90% (0.90)</td>
<td>10% (0.10)</td>
<td></td>
</tr>
<tr>
<td>95% (0.95)</td>
<td>5% (0.05)</td>
<td></td>
</tr>
<tr>
<td>98% (0.98)</td>
<td>2% (0.02)</td>
<td>2.33</td>
</tr>
<tr>
<td>99% (0.99)</td>
<td>1% (0.01)</td>
<td></td>
</tr>
</tbody>
</table>
98%:

- $\alpha = 0.02$
- $\alpha/2 = 0.01$
- Look in table for Z-score with $1 - 0.01 = 0.99$

![Z-score diagram](image)

Table E: The Standard Normal Distribution

<table>
<thead>
<tr>
<th>$z$</th>
<th>.00</th>
<th>.01</th>
<th>.02</th>
<th>.03</th>
<th>...</th>
<th>.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.9901</td>
</tr>
</tbody>
</table>
Formula for the Confidence Interval of the Mean when $\sigma$ is known.

Assumptions:

1. Our sample is a random sample from some population.
2. The population from which the sample came is normally distributed or the sample size $n \geq 30$.
3. The standard deviation $\sigma$ is known.

Formula: $\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$
The right side of the previous formula, $z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$, is called the maximum error of estimate or the margin of error.

- $(1 - \alpha) \times 100\%$ of the means will fall within this interval.
Rounding Rule for a Confidence Interval for a Mean

When working from raw data, round one more decimal place than the original data.

When working with a sample mean and standard deviation, round to the same number of decimal places as the sample mean.
Example: A random sample of reading scores for 35 fifth graders has a mean of 82. Assume the standard deviation of the population of scores is 15. Find the 95% confidence interval of the mean reading score for all fifth graders.

Assumptions: Fifth graders come from the same population. The standard deviation is known.

\[
\overline{X} = 82 \quad \sigma = 15 \quad n = 35 \quad z_{\alpha/2} = 1.96
\]

\[
\overline{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 82 \pm 1.96 \frac{15}{\sqrt{35}} = 82 \pm 4.969507 = (77.87)
\]
Interpretation:

**We are 95% confident that the mean reading score for fifth graders falls between 77 and 87.**

What’s the probability that this interval contains the value of the population mean?

The interval either does or does not contain the parameter. The 95% corresponds to “the process.” That is, the probability that the interval will contain the parameter, assuming that a large number of samples are selected and that the estimation process on the same parameter is repeated is 0.95.
**Example:** A survey of 40 adults found that the mean age of a person’s primary vehicle is 5.6 years. Assuming the standard deviation of the population is 0.8 years, find the 90% confidence interval of the population mean.

**Assumptions:**

\[
\bar{X} = \quad \sigma = \quad n = \quad z_{\alpha/2} =
\]

\[
\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} =
\]
Interpretation:

What’s the probability that this interval contains the value of the population mean?
Sample Size

Often, we need to decide how big of a sample we need to make a precise estimate of the population parameter. This answer relies on 3 things:

1. the maximum error of estimate
2. the population standard deviation, and
3. the degree of confidence.
Formula for the Minimum Sample Size needed for an Interval Estimate of the Population Mean

\[ n = \left( \frac{Z_{\alpha/2} \cdot \sigma}{E} \right)^2 \] where \( E \) is the maximum error of estimate.

If the answer is not a whole number be sure to round up to the next whole number (\textit{never round down}).

Where did this formula come from?
**Example**: A university dean wishes to estimate the average number of hours students spend doing homework per week. The standard deviation from a previous study is 6.2 hours. How large a sample must be selected if he wants to be 99% confident of finding whether the true mean differs from the sample mean by no more than 1.5 hours?

**Known**: \( \sigma = 6.2, \alpha = 0.01, E = 1.5 \)

We need to find \( z_{\alpha/2} \) for \( \alpha = 0.01 \) which corresponds to having 0.005 in each tail so \( z_{\alpha/2} = 2.575 \).

And,

\[
n = \left( \frac{z_{\alpha/2} \cdot \sigma}{E} \right)^2 = \left( \frac{2.575 \cdot 6.2}{1.5} \right)^2 = 113.2805 \quad \Rightarrow \quad 114 \text{ (always round UP)}
\]
**Example:** A scientist wishes to estimate the average depth of a river within 2 feet with 98% confidence. From a previous study the standard deviations of the depths measured was 4.38 feet. What sample size does he need to achieve his goals?
Important concept: Confidence level

1. It is NOT the probability that our interval contains the population parameter. A specific interval either does contain the parameter or it does not.

2. It IS the proportion of all intervals constructed using this mechanism that will contain the population parameter.

**Example:** If we constructed 100 confidence intervals from random samples using a confidence level of 95%, then about 95% of the intervals will contain the population parameter.

- Why “about 95%” instead of exactly 95?
Example: The mean of a certain population is 15. A researcher took 50 random samples and constructed 95% confidence intervals for the population mean.

About how many of the intervals should contain the value of the population mean?

About how many of the intervals should NOT contain the value of the population mean?

The next page visually shows the confidence intervals from our samples.