6.1 Large Sample CT for a Mean

A confidence interval for a parameter is a data-based interval of numbers likely to include the true value of the parameter with a probability-based confidence.

A 95% confidence interval for \( \mu \) is an interval which was constructed in a manner such that 95% of such intervals contain the true value of \( \mu \).

<table>
<thead>
<tr>
<th>Desired Confidence</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>80%</td>
<td>1.28</td>
</tr>
<tr>
<td>90%</td>
<td>1.645</td>
</tr>
<tr>
<td>95%</td>
<td>1.96 ( \approx 2 )</td>
</tr>
<tr>
<td>99%</td>
<td>2.58</td>
</tr>
</tbody>
</table>

If the population is normal, then a 95% CI for \( \mu \) is

\[
\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}
\]

- In section 6.3 we will find using the t-table if we have a sample standard deviation, \( s \), instead of \( \sigma \).
- For now the book, like most books, uses Z values for simplicity in its initial description.
- This gets a bit confusing where the book does a problem differently before section 6.3 than the same problem in section 6.3.
- The right way is following section 6.3.

The rule for exams:

- Use \( z \) if you are given \( \sigma \); use the t-table if you have a sample standard deviation, \( s \).
- Given data, we have a sample standard deviation, \( s \).
- In practice the only time you use \( z \) is before collecting the data
  - when you compute the number of samples required
  - with a “what if” value of \( \sigma \).
  - See the \( n= \) formula below.
The methods we are studying assume normal populations.

If the population is not normal, then we can

1. Transform to normal
   For example using the log scale values

2. Use other methods for a distribution that match that data, e.g. Weibull

3. Rely on the Central Limit Theorem which says that for large enough \( n \) the distribution of \( \bar{x} \) is nearly normal even if the population isn’t normal.

   - Option (3) is less efficient then (1) or (2) in that more samples are required to achieve the same level of information.
   - By relying on (3) rather than getting things to match a particular distribution, you are making your job harder.
   - You need to collect more data to accomplish the same level of precision in your measurement.

   - We would use normal plots to see if the normality assumption is warranted for the data at hand.
   - The book also describes plotting strategies for exponential and Weibull distributions.
The two-sided confidence interval comes from

$$P \left( -1.96 \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq 1.96 \right) = 0.95$$

$$P \left( \bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}} \right) = 0.95$$

For large \( n \) \( s \approx \sigma \)

$$P \left( \bar{x} - 1.96 \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.96 \frac{s}{\sqrt{n}} \right) = 0.95$$

Suppose \( X = \) strength \( \bar{x} = 120, \ \sigma = 15, \ n = 100, \ s \approx 15, \ s/\sqrt{n} \approx 1.5 \)

$$120 \pm 1.96 \frac{15}{\sqrt{100}} = 120 \pm 1.96 \times 1.5 \approx 120 \pm 3$$

117 to 123

This does not say that 95% of the time the measured strength is between 117 and 123. The standard deviation is 15, so an interval containing 95% of the strengths would be much wider.
One-sided Confidence Bounds

Find a 95% lower confidence bound for $\mu$, $\mu_{\text{lower}}$ to $\infty$.

$$P \left( \mu > \bar{x} - 1.645 \frac{s}{\sqrt{n}} \right) = 0.95$$

$$120 - 1.645 \frac{15}{\sqrt{100}} = 120 - 1.645 \times 1.5 \approx 120 - 2.47 = 117.53$$

$117.53$ to $\infty$.

Find a 95% upper confidence bound for $\mu$, $-\infty$ to $\mu_{\text{upper}}$.

$$P \left( \mu < \bar{x} + 1.645 \frac{s}{\sqrt{n}} \right) = 0.95$$
**Required Sample Size (n)**

We are 95% certain $\bar{X}$ will be within $\pm 1.96 \frac{\sigma}{\sqrt{n}} = \pm 1.96$ SE’s of $\mu$.

In planning the experiment, if we want to be 95% certain that $\bar{X}$ is off by no more than $\Delta$, we need

$$\Delta = 1.96 \frac{\sigma}{\sqrt{n}}$$

$$n \approx \left( \frac{1.96 \sigma}{\Delta} \right)^2$$

Before the experiment we have a guess at $\sigma$ from:

- Previous data – $s$ from earlier
- Engineering judgment

More generally

$$n = \left( \frac{z\sigma}{\Delta} \right)^2$$

Suppose $X =$ distance for tennis balls for fixed propellant and charge.

We don’t know $\sigma$. We just have a guess, “what if.”

Suppose we anticipate $\sigma \approx 2$

Suppose we want 95% confidence $\bar{x}$ differs from true $\mu$ by no more than 0.5.

$$n \approx \left( \frac{1.96 \times 2}{0.5} \right)^2 \approx 61$$
6.2 Large Sample Significance Tests for a Mean

“The reason students have trouble understanding hypothesis testing may be that they are trying to think.” *Deming*

We have 2 possibilities to choose from,

\[ H_0, \text{ null hypothesis} \]
\[ H_a, \text{ alternative hypothesis} \]

In the end we “reject the null hypothesis” or “fail to reject the null hypothesis.”

Law case
\[ H_0: \text{Innocent} \]
\[ H_a: \text{Guilty} \]

In order to reject \( H_0 \) we need sufficient evidence. The burden of proof is on disproving \( H_0 \). “Innocent until proven guilty”

Example: Discharge limit = 2 ppm

Possibly \( H_0: \mu = 2 \text{ppm} \) (or maybe \( \mu > 2 \text{ppm} \))
\[ H_a: \mu < 2 \text{ppm} \leftarrow \text{burden of proof} \]

Possibly \( H_0: \mu = 2 \text{ppm} \) (or maybe \( \mu < 2 \text{ppm} \))
\[ H_a: \mu > 2 \text{ppm} \leftarrow \text{burden of proof} \]

Possibly \( H_0: \mu = 2 \text{ppm} \)
\[ H_a: \mu \neq 2 \text{ppm} \leftarrow \text{burden of proof} \]

The null hypothesis (in this book) always has =.

Set \( H_a \) first. Then \( H_0 \) is =.
Steps (p. 350)

**Steps 1 and 2:** State the null and alternative hypothesis.

**Step 3:** State the test criteria. That is, give the formula for the test statistic (plugging in only the hypothesized value from the null hypothesis but not any sample information) and the reference distribution. Then state in general terms what observed values of the test statistic constitute evidence against the null hypothesis.

**Step 4:** Show the sample based calculations.

**Step 5:** Report an observed level of significance, p-value, and (to the extent possible) state its implications in the context of the real engineering problem.

Example: Bearings are intended to be 5 cm in diameter.

**Steps 1 and 2:** State the null and alternative hypothesis.

The null hypothesis in this book always has $=.

\[ H_0: \mu = 5, H_a: \mu \neq 5 \text{ or maybe } H_0: \mu = 5, H_a: \mu < 5 \]

**Step 3:** “State the test criteria.” The test statistic is some statistic for which we can find probabilities such as

\[ P(\bar{x} > 5.014|\mu = 5) = P\left(Z > \frac{5.014-5}{\sigma/\sqrt{n}}\right) \]

\[ = P(Z > \frac{5.014-5}{0.95/\sqrt{100}}) = P(Z > 2.8) \approx 0.0026 \]

The reason is that we want to qualify how unlikely the data are if $H_0$ is true.

If it’s unlikely to have seen this kind of data if $H_0$ is true, then we reject $H_0$. 
“Give the formula for the test statistic.”

Suppose we are taking a sample of n=100 ball bearings.

Test statistic $Z = \frac{\bar{x} - 5}{\sigma/\sqrt{100}}$

The reference distribution is standard normal.

“Then state in general terms what observed values of the test statistic constitute evidence against the null hypothesis.”

Since $H_a: \mu \neq 5$, $Z$ values far from 0 in either direction are good evidence against $H_0: \mu = 5$.

If $H_a: \mu > 5$, $Z$ values far from 0 in a positive direction are good evidence against $H_0: \mu = 5$.

If $H_a: \mu < 5$, $Z$ values far from 0 in a negative direction are good evidence against $H_0: \mu = 5$.

**Step 4: Show the sample based calculations.**

Suppose our sample data turn out to be

$n = 100 \quad \bar{x} = 5.006 \quad \sigma \approx 0.05 \quad \frac{\sigma}{\sqrt{n}} \approx 0.005$

Then $z = \frac{\bar{x} - 5}{0.005} = \frac{0.006}{0.005} = 1.2$

**Step 5: Report the p-value or level of significance.**

p-value: probability of this much or more evidence against $H_0$ if $H_0$ is true.

$p$-value $= 2 \times 0.1151 = 0.2302$
Step 6: Interpret the Results

If the p-value or level of significance is small,
- This type of data are unlikely if $H_0$ is true.
- The null hypothesis looks bad $\Rightarrow$ reject $H_0$.

For example in a law case,
- If there is a small chance the prosecution could have found this much evidence against a guilty person,
- The null hypothesis of innocent looks bad.
- Small p-values are bad for the null hypothesis, $H_0$.

The significance level or p-value is
- The probability of a result as least as extreme (away from what the null hypothesis would have predicted)
- If in fact the null hypothesis is true.
- So if the data are extremely unlikely when the null hypothesis is true,
  - The p-value is small and
  - The null hypothesis looks bad.

In the example, if $H_0: \mu = 5$ is true and $H_a: \mu \neq 5$.
- P-value = 0.23
- There would be a 23% chance of this much or more evidence against $H_0: \mu = 5$ in favor of $H_a: \mu \neq 5$.
- There is a 23% chance of getting data this far from expected if $H_0$ is true, $\mu = 5$.
- This resulting data are not extremely unlikely if $\mu = 5$,
  - So the evidence against $\mu = 5$ is not overwhelming.
  - The data are not obviously inconsistent with $H_0$.
  - We do not reject $H_0: \mu = 5$. 
If, however, suppose the sample mean is \( \bar{x} = 5.014 \)

\[
z = \frac{\bar{x} - 5}{0.005} = \frac{5.014 - 5}{0.005} = \frac{0.014}{0.005} = 2.8
\]

- The p-value is now \( 2 \times 0.0026 = 0.0052 \)
- Very good evidence against \( H_0 \).
- We would have a very small chance (\( \approx 5/1000 \)) of observing this sort of data if \( H_0 \) is true.

If p-value < 0.05,
- this is commonly considered “statistically significant” evidence against the null hypothesis at the \( \alpha = 0.05 \) level.

\( \alpha \) is
- How small the p-value has to be in order to reject \( H_0 \)
- The chance that we are willing to take of rejecting \( H_0 \) when \( H_0 \) is in fact true.
- The chance of convicting an innocent person or innocent \( H_0 \).
One-sided Alternatives

If we had $H_0: \mu = 5, H_a: \mu < 5$ and

- $z = \frac{\bar{x} - 5}{0.005} = \frac{5.014 - 5}{0.005} = \frac{0.014}{0.005} = 2.8$

- the p-value would be $1 - 0.0026 = 0.9974$.
- Only very negative values of $z$, $\bar{x}$ far enough below 5, constitute evidence against $H_0$.
- We have no indication that $H_0$ is false, so the p-value is large.
- **Small** p-values are bad for $H_0$
  - because this sort of data are unlikely if $H_0$ is true.

If we had $H_0: \mu = 5, H_a: \mu > 5$ and

- $z = \frac{\bar{x} - 5}{0.005} = \frac{0.014}{0.005} = 2.8$
- the p-value would be $0.0026$.
- Only very positive values of $z$ constitute evidence against $H_0$.
- We have good indication that $H_0$ is false, so the p-value is small.
- **Small** p-values are bad for $H_0$
  - because then this sort of data are unlikely if $H_0$ is true.
P-values and hypothesis testing are widely used.

- **However**, in my opinion and some others’ opinions (see author’s comments later in the chapter), more often than not, such significance tests are not useful summaries. See Deming quote earlier.

Generally, confidence intervals are more useful summaries.

\[
\begin{align*}
  n &= 100 \quad \bar{x} = 5.006 \quad \sigma \approx 0.05 \quad \frac{\sigma}{\sqrt{n}} \approx 0.005 \\
  5.006 \pm 1.96(\text{SE}) & \quad 5.006 \pm 1.96(0.005) \quad 5.006 \pm 0.0098 \\
  4.996 \text{ to } 5.016.
\end{align*}
\]

A mean diameter of 5.0 is included in these values,

- So it’s one of the potentially likely values for the mean.
- A null hypothesis of \( \mu = 5 \) is not rejected.
  - For a two-sided alternative
  - Using this 2-sided confidence interval
- But the mean could be as high as 5.016.
- Whether the mean is **close enough** to 5.0 for our purposes depends on the engineering considerations.

Likely the real question isn’t, “Is the mean equal to 5.0?”

- For one thing, we know the mean isn’t **exactly** 5.000000.
- In addition, it doesn’t do us much good if the average is 5.0 if the bearing diameters are really variable.
- We are likely more concerned how often the bearings are meeting specifications.
  - Consideration of the purposes for the data.
Often
- We know the null hypothesis is false.
- The real question isn’t whether the null hypothesis is true.

Suppose we are comparing our standard glues and a new glue for plywood.
- Our question isn’t really whether \( \mu_1 = \mu_2 \).
- We don’t just want to know if the two glues are equivalent.
- We want to know how different the strength of plywood is using the two glues,
  - not just whether there is a difference.

Is the improvement worth the cost?

A confidence interval for \( \mu_{new} - \mu_{old} \) would be useful in addressing this question.

Two-sided alternative \( \leftrightarrow \) two-sided confidence interval
- Reject \( H_a: \mu = 5 \) in favor of \( H_a: \mu \neq 5 \) if
  - Two-sided interval fails to include \( \mu = 5 \).

One-sided alternative \( \leftrightarrow \) One-sided confidence interval
- Reject \( H_a: \mu = 5 \) in favor of \( H_a: \mu > 5 \) if
  - Lower bound interval fails to include \( \mu = 5 \).
  - The interval is completely above 5.
  - We have evidence that \( \mu > 5 \).
- Reject \( H_a: \mu = 5 \) in favor of \( H_a: \mu < 5 \) if
  - Upper bound bound interval fails to include \( \mu = 5 \).
  - The interval is completely below 5.
  - We have evidence that \( \mu < 5 \).
What if the confidence interval is wider than we want?

- We want to know more precisely how different the glues are.
- We need a smaller standard error.
- Either
  - Increase sample size, \( n \).
  - Decrease variance, \( \sigma^2 \)
    - Strategies such as pairing or blocking are one way of reducing variability in comparisons.
    - Controlling variables better
    - Better, more precise measurements
    - Strategies from Chapter 2.