Basic Properties of Confidence Intervals

MATH 3342
Section 7.1

Example

• A sample of 100 soda cans, from a population with soda volume being Normally distributed having \( \sigma = 0.20 \) oz, produced a sample mean equal to 12.09 oz.

• What is a likely range of values for the population mean \( \mu \) of soda volume?
Confidence Intervals

- Provides a range of plausible values for a parameter.
- Developed from a sample
- For a given confidence level

- A **confidence level C** is a measure of the degree of reliability of the interval.
  - Gives the probability that the interval will capture the true parameter in repeated samples.
  - C is usually 90% or higher.
Assumptions for This Section

Suppose $\mu$ is the parameter of interest:

1. The data is from a SRS. There are no non-sampling errors.
2. The variable of interest is exactly Normal distributed.
3. We don’t know the population mean $\mu$, but we do know the population standard deviation $\sigma$.
   - Not necessarily a practical assumption.

Construction of a 95% CI

- We build our interval starting from the formula for a z-score and a probability statement about a standard normal RV.

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$P(-1.96 < Z < 1.96) = 0.95$$
Construction of a 95% CI

\[ 0.95 = P(-1.96 < Z < 1.96) \]
\[ = P\left(-1.96 \cdot \frac{X - \mu}{\sigma/\sqrt{n}} < 1.96 \right) \]
\[ = P\left(-1.96 \cdot \frac{\sigma}{\sqrt{n}} < X - \mu < 1.96 \cdot \frac{\sigma}{\sqrt{n}} \right) \]
\[ = P\left(-X - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < -\mu < -X + 1.96 \cdot \frac{\sigma}{\sqrt{n}} \right) \]
\[ = P\left(X - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < X + 1.96 \cdot \frac{\sigma}{\sqrt{n}} \right) \]

Construction of a 95% CI

\[ \left( \bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}} \right) \] is a Random Interval.

The width is not random, only the center is.

*Interpretation*: The probability that this random interval includes the true value of \( \mu \) is 95%.
95% CI for $\mu$

Suppose we observe $X_1 = x_1, \ldots, X_n = x_n$ and compute $\bar{x}$ and use it in place of $\bar{X}$, then:

$$\left( \bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}} \right)$$

is a 95% CI for $\mu$.

Alternatively,

$$\bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

with 95% confidence.

Interpretation

• When we substitute in the observed sample mean, the interval is no longer random!

• Therefore:
  • The probability statement is not about any one observed interval
  • Rather, a statement about the long-term relative frequency if we repeatedly sample and compute intervals in this manner.
A sample of 100 soda cans, from a population with soda volume being Normally distributed having $\sigma = 0.20$ oz, produced a sample mean equal to 12.09 oz. A 95% confidence interval would be:

$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

$$12.09 \pm 1.96 \frac{0.20}{\sqrt{100}}$$

$$12.09 \pm 0.039$$

12.051 ounces ———————— 12.129 ounces
100(1 - \alpha \%) \%
Confidence Intervals

\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}

where:
- \( z_{\alpha/2} \) = Standard Normal Critical Value
- \( \sigma \) = Population standard deviation
- \( n \) = Sample size

Common Critical Values

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>( z_{\alpha/2} )</td>
<td>1.645</td>
<td>1.960</td>
<td>2.576</td>
</tr>
</tbody>
</table>
Example

A sample of 100 soda cans, from a population with soda volume being Normally distributed having \( \sigma = 0.20 \) oz, produced a sample mean equal to 12.09 oz. A 99% confidence interval would be:

\[
\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}
\]

\[
12.09 \pm 2.576 \frac{0.20}{\sqrt{100}}
\]

12.09 \pm 0.052

<table>
<thead>
<tr>
<th>12.051 ounces</th>
<th>12.129 ounces</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.038 ounces</td>
<td>12.142 ounces</td>
</tr>
</tbody>
</table>

How do we reduce width?

- Lower the confidence level. This results in a lower critical value \( z_{\alpha/2} \).
- Increase the sample size \( n \). This reduces the variability of the sampling distribution.
Example: Impact of $n$

If instead of sample of 100 cans, suppose a sample of 400 cans, from a population with $\sigma = 0.20$, produced a sample mean equal to 12.09. A 95% confidence interval would be:

$$\bar{x} \pm z \frac{\sigma}{\sqrt{n}}$$

$$12.09 \pm 1.96 \frac{0.20}{\sqrt{400}}$$

$$12.09 \pm 0.0196$$

<table>
<thead>
<tr>
<th>n = 400</th>
<th>12.0704 ounces</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 100</td>
<td>12.051 ounces</td>
</tr>
</tbody>
</table>

The Minimum $n$ for a Desired Width

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

where:

$z_{\alpha/2} = $ Critical value

$w = $ Desired width

$\sigma = $ Population standard deviation
Always round up for $n$ !!!

- Suppose the calculation says that 58.2 observations are needed.
- 58 observations would produce a slightly wider interval than is wanted.
- 59 observations would produce a slightly narrower interval than desired.
- Estimate would still then be within the desired margin of error with the desired level of confidence

Example

The manager of the Georgia Timber Mill wishes to construct a 90% confidence interval within 0.50 inches in estimating the mean diameter of logs. Assume a population standard deviation of 4.8 inches.

$$n = \left( 2 \cdot 1.645 \frac{4.8}{1} \right)^2 = 249.38 \approx 250$$