Confidence Interval for the Difference of 2 Population Means

Ha Khanh Nguyen (hknguyen)

\[ X_1, X_2, \ldots, X_{n_1} \] from population 1
mean \( \mu_1 \), std \( \sigma_1 \)

\[ Y_1, Y_2, \ldots, Y_{n_2} \] from population 2
mean \( \mu_2 \), std \( \sigma_2 \)

When one of these 2 conditions are satisfied:
- \( n_1 \) and \( n_2 \) are large \((\geq 30)\)
- Population 1 and population 2 are approximately normal

Then, \( (\bar{X} - \bar{Y}) \) is approximately normal with mean \( \mu_1 - \mu_2 \) and standard deviation \( \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \).

A \((1 - \alpha)\)100\% confidence interval for \( \mu_1 - \mu_2 \) is

\[ (\bar{X} - \bar{Y}) \pm z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \]

If \( \sigma_1 \) and \( \sigma_2 \) are unknown, use:

\[ (\bar{X} - \bar{Y}) \pm t_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \]

where the number of degrees of freedom of the \( t \)-dist = smaller of \( n_1 - 1 \) and \( n_2 - 1 \).

Notes: If \( n_1 \) and \( n_2 \) are large, \( t_{\alpha/2} \) can be approximated by \( z_{\alpha/2} \).

Example 1:
Dr. Statman claims that his new revolutionary study method “Study While You Sleep” (SWYS) is more effective than the traditional study methods. In an experiment, 250 students enrolled in the same section of STAT 400 at UIUC were divided into 2 groups. 100 students volunteered to study using SWYS method, and the other 150 students did whatever the students usually do. At the end of the semester, the average of the total number of points (out of 500) were compared for the 2 groups.

Note: This is not a good experiment design!

<table>
<thead>
<tr>
<th></th>
<th>SWYS 1</th>
<th>Traditional 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average total points</td>
<td>450</td>
<td>410</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>20</td>
<td>45</td>
</tr>
</tbody>
</table>

Construct a 95\% confidence interval for the difference in the average total points for SWYS and the traditional study methods.
Step 0: Check the assumptions
\[
\begin{align*}
&n_1 = 100 > 30 \quad \text{AND} \quad n_2 = 150 > 30 \\
\Rightarrow &\quad \text{we can use the CI formula above}
\end{align*}
\]

Step 1: Compute the \((1 - \alpha)100\%\) confidence interval

A 95\% CI for \(\mu_1 - \mu_2:\) (\(\alpha = 0.05\))

\[
(\bar{x}_1 - \bar{x}_2) \pm Z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 450 - 410 \pm Z_{0.05/2} \sqrt{\frac{20^2}{100} + \frac{45^2}{150}}
\]

when \(n_1\) and \(n_2\) are large

AND \(\sigma_1\) and \(\sigma_2\) are unknown

\[
\approx (31.8009, 48.1991)
\]