## Learning Outcome

By the end of this lecture, you will be able to:

- Identify how to construct and interpret a confidence interval estimate for the difference between two independent population means when the standard deviations are known

- Identify how to construct and interpret a confidence interval estimate for the difference between two independent population means when the standard deviations are unknown

- Identify how to construct and interpret a confidence interval estimate for the difference between two population means from paired samples

## Confidence Interval Estimate for the Difference Between Two Means

Confidence interval estimate for the difference between two means is constructed based on the nature of samples.

The samples may be:

| | |
|---|---|
| **Independent Samples** | If the sample selected from one population has no effect on the sample selected from the other population, then the **two samples are independent**. |
| **Dependent Samples** | If the samples that are selected in such a way that values in one sample are matched with values in the second sample, then the **two samples are dependent**. |

In this lecture, you will learn about estimation of population means.

**Construct and Interpret a Confidence Interval**

There are four different cases to construct and interpret a confidence interval estimate for the difference between two independent population means. Let's explore these four cases one by one, beginning with case one.

**Case 1: Standard Deviations are Known**

Say a $(1-\alpha)100\%$ C.I for the difference between two population means, $\mu_1 - \mu_2$ is to be found.

Here are the assumptions:

- the two populations are normally distributed
- the two samples are independent
- $\sigma_1$ and $\sigma_2$ are known

Then, $(1-\alpha)100\%$ C.I is given by:

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad (1)$$

Where,

The point estimate for $\mu_1 - \mu_2$ is : $(\bar{x}_1 - \bar{x}_2)$.

The standard error of $(\bar{x}_1 - \bar{x}_2)$ is: $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

The margin of error is: $e = \pm\, z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

**Case 2: Standard Deviations are Unknown**

Say a $(1-\alpha)100\%$ C.I for the difference between two population means, $\mu_1 - \mu_2$ is to be found.

The assumptions are:

- The two populations are normally distributed
- The two samples are independent
- $\sigma_1$ and $\sigma_2$ are unknown
- Both sample sizes are $\geq 30$

Then, $(1-\alpha)100\%$ C.I is given by:

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \qquad (2)$$

Where,

The point estimate for $\mu_1 - \mu_2$ is : $(\bar{x}_1 - \bar{x}_2)$.

The standard error of $(\bar{x}_1 - \bar{x}_2)$ is: $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$.

The margin of error is: $e = \pm z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$.

**Case 3: Standard Deviations are Unknown and Equal**

A $(1-\alpha)100\%$ C.I for the difference between two population means, $\mu_1 - \mu_2$ is to be found.

The assumptions are:

- the two populations are normally distributed
- the two samples are independent
- $\sigma_1$ and $\sigma_2$ are unknown and equal $(\sigma_1 = \sigma_2)$
- $n_1$ or $n_2$ or both are small ($n_1 < 30$ OR $n_2 < 30$ OR both less than 30)

Then, $(1-\alpha)100\%$ C.I is: $(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}, n_1+n_2-2}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

Where,

The pooled standard deviation is given by:

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} \qquad (3)$$

and $t_{\frac{\alpha}{2}}$ has $n_1 + n_2 - 2$ degrees of freedom.

The point estimate for $\mu_1 - \mu_2$ is : $(\bar{x}_1 - \bar{x}_2)$.

The standard error of $(\bar{x}_1 - \bar{x}_2)$ is: $\sigma_{\bar{x}_1 - \bar{x}_2} = S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$.

The margin of error is: $e = \pm z_{\alpha/2}\, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$.

**Case 4: Standard Deviations are Unknown and Unequal**

A $(1-\alpha)100\%$ C.I for the difference between two population means, $\mu_1 - \mu_2$ is to be found.

The assumptions are:

- the two populations are normally distributed
- the two samples are independent
- $\sigma_1$ and $\sigma_2$ are unknown and not equal $(\sigma_1 \neq \sigma_2)$
- $n_1$ or $n_2$ is small ($n_1 < 30$ OR $n_2 < 30$)

Then, $(1-\alpha)100\%$ C.I is given by: $(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}, v} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$.

Where,

$$t_{\frac{\alpha}{2}} \; has \; v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} \; degrees \; of \; freedom \qquad (4)$$

The point estimate for $\mu_1 - \mu_2$ is : $(\bar{x}_1 - \bar{x}_2)$.

The standard error of $(\bar{x}_1 - \bar{x}_2)$ is: $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$.

The margin of error is: $e = \pm t_{\frac{\alpha}{2}, v}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$.

**Example**

Given: $n_1 = 36$, $n_2 = 45$, $\bar{x}_1 = 2456$, $\bar{x}_2 = 2460$, $s_1 = 32$, $s_2 = 80$

Determine the 95% C.I estimate for the difference between population means. Based on the example, is there is a difference between the two means? Explain.

*Solution:*

Assume that the populations are normal and the samples are independent.

Since both $\sigma_1$ *and* $\sigma_2$ are unknown and both sample sizes are large $n_1 > 30$ and $n_2 > 30$.

A $(1 - \alpha)100\%$ C.I. for $\mu_1 - \mu_2$ is:

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$1 - \alpha = 0.95 \rightarrow z_{0.05/2} = z_{0.025} = 1.96$$

A 95% C.I. for $\mu_1 - \mu_2$ is:

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$= (2456 - 2460) \pm 1.96\sqrt{\frac{32^2}{36} + \frac{80^2}{45}}$$

$$= -4 \pm 25.60533$$

$$= [-29.605,\ 21.605]$$

The confidence interval that we found in the example gave us an estimate for the difference between the two means (the difference may be equal to any value in the interval).

Since $0 \in [-29.605, 21.605]$, we can be 95% confident that $\mu_1 - \mu_2$ (no difference between the two means).

**Example**

Given: $n_1 = 100$, $\bar{x}_1 = 50$, $\sigma_1 = 6$ and $n_2 = 150$, $\bar{x}_2 = 65$, $\sigma_2 = 8$

Determine the 90% C.I estimate for the difference between population means. Do you think that there is a difference between the two means? Explain.

*Solution:*

Assume that the populations are normal and the samples are independent.

Since both $\sigma_1$ *and* $\sigma_2$ are known, $\alpha$ $(1 - \alpha)100\%$ C.I. for $\mu_1 - \mu_2$ is:

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$1 - \alpha = 0.9 \rightarrow z_{0.1/2} = z_{0.05} = 1.645$$

A 90% C.I. for $\mu_1 - \mu_2$ is:

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$= (50 - 65) \pm 1.645\sqrt{\frac{6^2}{100} + \frac{8^2}{150}}$$

$$= -15 \pm 1.45902$$

$$= [-16.459, \quad -13.541]$$

The confidence interval that we found gave us an estimate for the difference between the two means (the difference may be equal to any value in the interval).

Since $0 \notin [-16.459, -13.541]$, we can be 95% confident that $\mu_1 \neq \mu_2$ (there is a difference between the two means).

**Example 3**

A random sample of 15 bulbs produced by an old machine was tested and found to have a mean life span of 40 hours with a standard deviation of 5 hours. Also, a random sample of 10 bulbs produced by a new machine was found to have a mean life span of 45 hours with a standard deviation of 6 hours. Assume that the life span of a bulb has a normal distribution for both machines, and true variances are the same. Construct a 95% confidence interval for the difference between the mean lives of the bulbs produced by two machines.

*Solution:*

$n_1 = 15$, $n_2 = 10$, $\bar{x}_1 = 40$, $\bar{x}_2 = 45$, $s_1 = 5$, $s_2 = 6$

The assumptions are:

- Both populations are normally distributed
- Samples are independent
- Both standard deviations are unknown
- Small sample sizes: $n_1 = 15 < 30$ and $n_2 = 15 < 30$.
- The true variances are equal

A $(1 - \alpha)100\%$ C.I. for $\mu_1$ - $\mu_2$ is:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}, n_1+n_2-2}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$1 - \alpha = 0.95 \rightarrow t_{0.05/2, 10+15-2} = t_{0.025,23} = 2.0687$$

And

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{14(25) + 9(36)}{23}}$$

$$= 5.41334904$$

A 95% C.I. for $\mu_1$ - $\mu_2$ is:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}, n_1+n_2-2}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= (40 - 45) \pm 2.0687(5.41334904)\sqrt{\frac{1}{15} + \frac{1}{10}}$$

$$= -5 \pm 4.571807$$

$$= [-9.57181, \quad -0.42819]$$

**Confidence Intervals**

There are situations in which we would want to use paired samples to control the source of variation (could be extraneous factors) that might otherwise distort the conclusion of the study.

Other terms for dependent samples:

- Paired or matched samples
- Repeated measures (before/after)

**Confidence Intervals**

The assumptions determine confidence intervals are:

- The samples are dependent
- Both populations are normally distributed
- Or, if not normal, large samples are used

Use the difference between paired values: The $i^{th}$ paired difference is $d_i = x_{1i} - x_{2i}$.

The point estimate for the population mean paired difference is:

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n} \qquad (5)$$

The sample standard deviation is:

$$s_d = \sqrt{\frac{\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2}{n-1}} = \sqrt{\frac{\sum d_i^2 - n\left(\bar{d}\right)^2}{n-1}} \qquad (6)$$

Where, $n$ is the number of pairs in the paired sample

When a $(1-\alpha)100\%$ C.I for $d$ is:

$$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}} \qquad (7)$$

Where $t_{\alpha/2}$ has $n-1$ d.f.

**Solved Problem**

Assume you send your salespeople to a "customer service" training workshop. Find 95% C.I for the difference between the number of complaints.

Given:

Customer service

| | **Number of Complaints:** | | **Differences** |
|---|---|---|---|
| **Salesperson** | **Before (1)** | **After (2)** | **d** |
| 1 | 6 | 4 | 2 |
| 2 | 20 | 6 | 14 |
| 3 | 3 | 2 | 1 |
| 4 | 0 | 0 | 0 |
| 5 | 4 | 0 | 4 |

*Solution:*

The assumptions are:

- Both populations are normally distributed
- Samples are dependent (related)

A $(1-\alpha)100\%$ C.I. for $\mu_1$ - $\mu_2$ is:

$$\bar{d} \pm t_{\frac{\alpha}{2},n-1} \frac{s_d}{\sqrt{n}}$$

$$\bar{d} = \frac{\sum d_i}{n} = \frac{21}{5} = 4.2$$

$$s_d = \sqrt{\frac{\sum d_i^2 - n\bar{d}^2}{n-1}}$$

$$= \sqrt{\frac{217 - 5(4.2)^2}{4}}$$

$$= 5.674504384$$

$$1-\alpha = 0.95 \rightarrow t_{0.05/2,5-1} = t_{0.025,4} = 2.7764$$

A 95% C.I. for $\mu_1$ - $\mu_2$ is:

$$\bar{d} \pm t_{\frac{\alpha}{2},n-1} \frac{s_d}{\sqrt{n}}$$

$$= 4.2 \pm (2.7764)\frac{5.674504384}{\sqrt{5}}$$

$$= 4.2 \pm 7.045713$$

$$= [-2.84571, \ \ 11.24571]$$

**Recap**

In this lecture, you have learned that:

- When the two populations are normally distributed, the two samples are independent and $\sigma_1$ and $\sigma_2$ are known, then A $(1-\alpha)100\%$ C.I is:

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- When the two populations are normally distributed, the two samples are independent, $\sigma_1$ and $\sigma_2$ are unknown and both sample sizes are $\geq 30$, then A $(1-\alpha)100\%$ C.I is:

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- When the two populations are normally distributed, the two samples are independent, $\sigma_1$ and $\sigma_2$ are unknown and equal, $n_1$ or $n_2$ or both are small ($n_1 < 30$ OR $n_2 < 30$ OR both less than 30), then A $(1-\alpha)100\%$ C.I is:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}, n_1+n_2-2}\, s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

- When the two populations are normally distributed, the two samples are independent, $\sigma_1$ and $\sigma_2$ are unknown and not equal, $\sigma_1 \neq \sigma_2$ unknown and $n_1$ or $n_2$ is small ($n_1 < 30$ OR $n_2 < 30$, then A$(1-\alpha)100\%$ C.I is:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}, v}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$