

## CEE 260/MIE 273: Probability and Statistics in Civil Engineering

### Lecture 6B: Inference for Two Samples

**Jimi Oke**

UMass **Amherst**

---

College of Engineering

November 20, 2025

# Outline

- ① CI for difference of two means
- ② Hyp. Tests for difference of two means
- ③ Paired data
- ④ Outlook

# Confidence intervals for difference of two means

The CI for a difference of two means (with unknown variance) is given by:

$$\langle \mu_1 - \mu_2 \rangle_{1-\alpha} = (\bar{x} - \bar{y} \pm t^* \times SE_{diff}) \quad (1)$$

where:

- The critical  $t$ -value is

$$t^* = F_T^{-1}(1 - \alpha/2, df) \quad (2)$$

given by `t.ppf(1 - alpha/2, df)` in Python

- The standard error of the difference of the two means is

$$SE_{diff} \approx \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (3)$$

where  $s_1^2$  and  $s_2^2$  are the respective sample variances

## Notes

- If the population variances  $\sigma_1^2$  and  $\sigma_2^2$  are known, then you can use the z-score instead
- You can easily derive the formulas for the upper/lower confidence bounds.

# Computing $df$ for difference of two means

To find the critical  $t$ -value, we need to compute the degrees of freedom ( $df$ ) parameter  $df$ .

For a single sample,  $df = n - 1$ . However, when dealing with **two samples**,  $df$  given by:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}} \quad (4)$$

This is a complicated formula.

## df shortcut

A simpler way to estimate  $df$  is to use the formula:

$$df \approx \min(n_1 - 1, n_2 - 1) \quad (5)$$

(i.e. the smaller of the two)

# Example 1: Permeability of textile fabrics

The void volume within a textile fabric affects comfort, flammability and insulation properties. Permeability of a fabric refers to the accessibility of void space to the flow of a gas or liquid. Consider the following permeability ( $\text{cm}^3/\text{cm}^2/\text{sec}$ ) summary data on two different types of plain-weave fabric



Microscopic images of cotton fiber

arrangements. Source:

<https://journals.sagepub.com/doi/full/10.1177/15589250211024225>

Fabric type	Sample size	Sample mean	Sample SD
Cotton	10	51.71	0.79
Triacetate	10	136.14	3.59

Assuming that the permeability distributions for both types of fabric are normal, calculate a CI for the difference between true average permeability for the cotton fabric and that for the triacetate fabric using a 95% confidence level.

# Example 1: Permeability of textile fabrics (cont)

## Solution

First we compute the degrees of freedom:

$$\begin{aligned}
 df &= \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}} \\
 &= \frac{\left(\frac{0.6241}{10} + \frac{12.8881}{10}\right)^2}{\frac{(0.6241/10)^2}{9} + \frac{(12.8881/10)^2}{9}} = \frac{1.8258}{0.1850} = 9.87
 \end{aligned}$$

We round down to nearest integer. Thus, we use:  $df = 9$ .

## Quick back-of-the-hand calculation for $df$

A shortcut for finding  $df$  is

$$df \approx \min(n_1 - 1, n_2 - 1) \quad (6)$$

In this case, we can see that since  $n_1 = n_2$ , we ended up with  $df = 10 - 1 = 9$ .

# Example 1: Permeability of textile fabrics (cont)

## Solution

Next, we compute the critical  $T$ -value, given that

$$1 - \alpha = 0.95 \implies \alpha/2 = 0.025:$$

$$-t^* = t_{(\alpha/2, df)} = t_{0.025, 9} = -2.262 \quad \text{t.ppf(.025, 9)}$$

The margin of error is therefore:

$$\begin{aligned} t^* \times SE &= t_{(1-\alpha/2)} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 2.262 \sqrt{\frac{0.6241}{10} + \frac{12.8881}{10}} \\ ME &= 2.63 \end{aligned}$$

# Example 1: Permeability of textile fabrics (cont)

## Solution

The CI in this case is given by

$$\langle \mu_1 - \mu_2 \rangle_{1-\alpha} = \left( \bar{x}_1 - \bar{x}_2 + t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \bar{x}_1 - \bar{x}_2 + t_{(1-\alpha/2)} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right) \quad (7)$$

But we can rewrite it conveniently as:

$$\langle \mu_1 - \mu_2 \rangle_{1-\alpha} = \bar{x}_1 - \bar{x}_2 \pm ME \quad (8)$$

where *ME* is the margin of error.

Thus, the 95% CI is:

$$\begin{aligned} \langle \mu_1 - \mu_2 \rangle_{.95} &= 51.71 - 136.14 \pm 2.63 \\ &= -84.43 \pm 2.63 = \boxed{(-87.06, -81.80)} \end{aligned}$$

# Example 1: Permeability of textile fabrics (cont)

## Solution

### Interpretation:

With a high degree of confidence, we can say that true average permeability for triacetate fabric specimens exceeds that for cotton specimens by between 81.80 and 87.06  $\text{cm}^3/\text{cm}^2/\text{sec}$ .

# Comparing two populations: difference of two means

## Example cases

- Is there any significant difference between the performances in the 2018 NY marathon and that of 2019?
- Is the hardness of heat-treated steel similar to that of cold-rolled steel?
- How can an engineer test if the proportion of defective batteries in a production batch is similar to that of another batch?

## Key assumptions

- Normality:  $X_1$  is a random normal sample;  $X_2$  is a random normal sample
- Independence:  $X_1$  and  $X_2$  are independent of each other

# Testing normal populations with known variances

**Null hypothesis:**  $H_0 : \mu_1 - \mu_2 = \Delta_0$ .

Often  $\Delta_0$  is often 0, in which case,  $H_0 : \mu_1 = \mu_2$ .

**Test statistic:**

$$z = \frac{\bar{x}_1 - \bar{x}_2 - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (9)$$

where  $\bar{x}_1, \bar{x}_2$  are the sample means,  $\sigma_1^2, \sigma_2^2$  are the respective population variances, and  $n_1, n_2$ , the respective sample sizes.

---

<b>Alternative Hypothesis</b>	<b>Rejection Region for level <math>\alpha</math> test</b>
-------------------------------	--

---

$$H_1 : \mu_1 - \mu_2 > \Delta_0 \quad z \geq z_{1-\alpha} \text{ (upper-tailed)}$$

---

$$H_1 : \mu_1 - \mu_2 < \Delta_0 \quad z \leq z_\alpha \text{ (lower-tailed)}$$

---

$$H_1 : \mu_1 - \mu_2 \neq \Delta_0 \quad z \leq z_{\alpha/2} \text{ or } z \geq z_{(1-\alpha)/2} \text{ (both tails)}$$

---

# Difference between two population means

## Example 1: Cold-rolled vs. galvanized steel

Analysis of a random sample consisting of  $n_1 = 20$  specimens of cold-rolled steel to determine yield strengths resulted in a sample average strength of  $\bar{x} = 29.8$  ksi.

A second random sample of  $n_2 = 25$  two-sided galvanized steel specimens gave a sample average strength of  $\bar{y} = 34.7$  ksi.

Assuming that the two yield-strength distributions are normal with  $\sigma_1 = 4.0$  and  $\sigma_2 = 5.0$ , do the data indicate that the corresponding true average yield strengths  $\mu_1$  and  $\mu_2$  are different?

Carry out a test at significance level  $\alpha = 0.01$ .

# Difference between two population means

## Example 1: Cold-rolled vs. galvanized steel

**Step 1.** Parameter of interest:  $\mu_1 - \mu_2$  (difference between the true average strengths)

**Step 2.** Null hypothesis:  $H_0 : \mu_1 - \mu_2 = \Delta_0 = 0$

**Step 3.** Alternative hypothesis:  $H_1 : \mu_1 - \mu_2 \neq 0$

**Step 4.** Formula for test statistic value:

$$z = \frac{\bar{x} - \bar{y} - (0)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

# Difference between two population means

## Example 1: Cold-rolled vs. galvanized steel (cont.)

**Step 5.** Calculate test statistic value:

$$z = \frac{29.8 - 34.7}{\sqrt{\frac{16.0}{20} + \frac{25.0}{25}}} = \frac{-4.90}{1.34} = -3.66$$

**Step 6.** Find the critical values (two-tailed test):

$$\alpha/2 = 0.01/2 = 0.005$$

$$z^* = z_{(1-\alpha/2)} = z_{0.995} = 2.58 \quad (\text{norm.ppf}(0.995))$$

$$z_{\alpha/2} = z_{0.005} = -2.58$$

# Difference between two population means

## Example 1: Cold-rolled vs. galvanized steel (cont.)

### Step 7. Conclude:

Since  $-3.66 < -2.58$ ,  $z$  falls in the lower tail of the rejection region.  $H_0$  is therefore rejected in favor of the conclusion that  $\mu_1 \neq \mu_2$ .

Furthermore, the  $p$ -value is  $2(1 - \Phi(3.66)) \approx 2(1 - 1) = 0$ .

So,  $H_0$  should be rejected at any reasonable significance level.

# Paired data

Paired data arise when two different observations are made on the **same set** of  $n$  individuals in a sample.

## Examples of paired data

- Prices from two different vendors on a set of undergraduate textbooks
- Monthly average proportion of absent students in a school *before* and *after* an intervention to boost attendance by serving free meals.
- Zinc concentration in six bodies of water collected at the *surface* and at the *bottom*.

## Assumptions

- $n$  independently selected pairs of observations:  $\mathbb{E}(X_i) = \mu_1, \mathbb{E}(Y_i) = \mu_2$ .
- Differences between pairs  $D_i = X_i - Y_i$  are normally distributed with mean  $\mu_D$  and variance  $\sigma_D^2$ 
  - If  $n$  is small, conduct *t*-test
  - If  $n$  is large, conduct *z*-test

# Paired $t$ -test

**Null hypothesis:**  $H_0 : \mu_D = \mu_0$ .

**Test statistic:**

$$t = \frac{\bar{d} - \mu_0}{s_D / \sqrt{n}} \quad (10)$$

where  $\bar{d}$  is the sample difference,  $\mu_0$  is the null difference and  $s_D$  is the sample standard deviation

---

<b>Alternative Hypothesis</b>	<b>Rejection Region for level <math>\alpha</math> test</b>
-------------------------------	--

---

$$H_1 : \mu_D > \mu_0 \quad t \geq t_{\alpha, n-1} \text{ (upper-tailed)}$$

---

$$H_1 : \mu_D < \mu_0 \quad t \leq t_{\alpha, n-1} \text{ (lower-tailed)}$$

---

$$H_1 : \mu_D \neq \mu_0 \quad t \leq t_{\alpha/2, n-1} \text{ or } t \geq t_{(1-\alpha/2), n-1} \text{ (both tails)}$$

---

# Paired *t*-test example

## Example 2: Intervention for ergonomic improvements

Musculoskeletal neck-and-shoulder disorders are all too common among office staff who perform repetitive tasks using visual display units. A study was conducted to determine whether more varied work conditions would have any impact on arm movement.<sup>a</sup>

The accompanying data was obtained from a sample of  $n = 16$  subjects. Each observation is the amount of time, expressed as a proportion of total time observed, during which arm elevation was  $30^\circ$ . The two measurements from each subject were obtained 18 months apart.

<sup>a</sup> "Upper-Arm Elevation During Office Work"  
(*Ergonomics*, 1996: 1221-1230)



Illustration of upper-arm elevation. Source:  
<https://www.sciencedirect.com/science/article/pii/S0003687018300590>

# Paired *t*-test in practice

## Example 2: Intervention for ergonomic improvements (cont.)

During this period, work conditions were changed, and subjects were allowed to engage in a wider variety of work tasks. Do the data suggest that true average time during which elevation is below  $30^\circ$  differs after the change from what it was before the change? Let  $\mu_D$  denote the true average difference between elevation time before the change in work conditions and time after the change.

**Step 1.**  $H_0 : \mu_D = 0$  (i.e. there is no difference between true average time before the change and true average time after the change)

**Step 2.**  $H_1 : \mu_D \neq 0$

**Step 3.** Compute the parameters:

$$\begin{aligned} n &= 16 \\ \sum d_i &= 108 \\ \sum d_i^2 &= 1746 \\ \bar{d} &= 6.75 \\ s_D &= 8.234 \end{aligned}$$

# Paired *t*-test in practice

## Example 2: Intervention for ergonomic improvements (cont.)

**Step 5.** Compute the *T*-statistic:

$$\begin{aligned} t &= \frac{\bar{d} - 0}{s_D / \sqrt{n}} \\ &= \frac{6.75}{8.234 / \sqrt{16}} = 3.28 \approx 3.3 \end{aligned}$$

**Step 6.** Find the *p*-value:

$$\begin{aligned} p\text{-value} &= 2 \cdot (1 - F_T(3.3, 15)) \quad (2 * \text{t.sf}(3.3, 15)) \\ &= 2(0.0024) \approx 0.005 \end{aligned}$$

# Paired *t*-test in practice

## Example 2: Intervention for ergonomic improvements (cont.)

### Step 7. Conclude:

Since  $0.005 < 0.01$ , the null hypothesis can be rejected at either significance level 0.05 or 0.01. Thus, this test indicates that the true average time after the change is different from that before the change.

# Summary

- Confidence intervals and hypothesis tests can be used to compare two population means.
- When population variances are unknown, we use the  $t$ -distribution with degrees of freedom computed using a complicated formula or a simple shortcut.
- Paired data arise when two different observations are made on the same set of individuals in a sample.
- The paired  $t$ -test is used to test hypotheses about the mean difference between paired observations.

# Key equations

- CI for difference of two means (unknown variances):

$$\langle \mu_1 - \mu_2 \rangle_{1-\alpha} = (\bar{x} - \bar{y} \pm t^* \times SE_{diff})$$

where  $SE_{diff} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

- If variances are known, use z-score instead of t-score; and  $\sigma$  instead of  $s$ .
- Degrees of freedom:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}} \quad \text{or} \quad df \approx \min(n_1 - 1, n_2 - 1)$$

- Two-sample z-test (known variances):

$$z = \frac{\bar{x}_1 - \bar{x}_2 - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

# Key equations (cont.)

- Paired data arise when two different observations are made on the **same set** of  $n$  individuals in a sample.
- Paired data differences  $D_i = X_i - Y_i$  are normally distributed with mean  $\mu_D$  and variance  $\sigma_D^2$ .
- Paired t-test:

$$t = \frac{\bar{d} - \mu_0}{s_D / \sqrt{n}}$$

where  $\bar{d} = \bar{x}_1 - \bar{x}_2$  is the sample mean difference and  $s_D$  is the sample standard deviation of differences