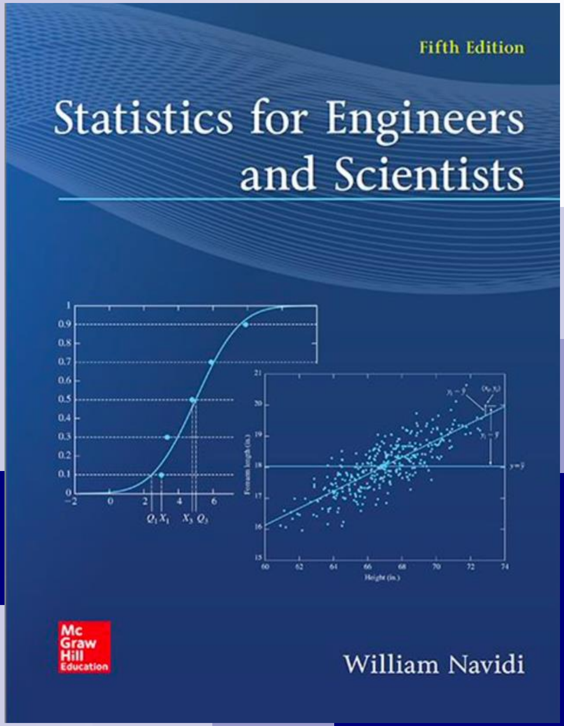


Chapter 5

Confidence Intervals (part 3)

Ch. 5 Overview (required sections)

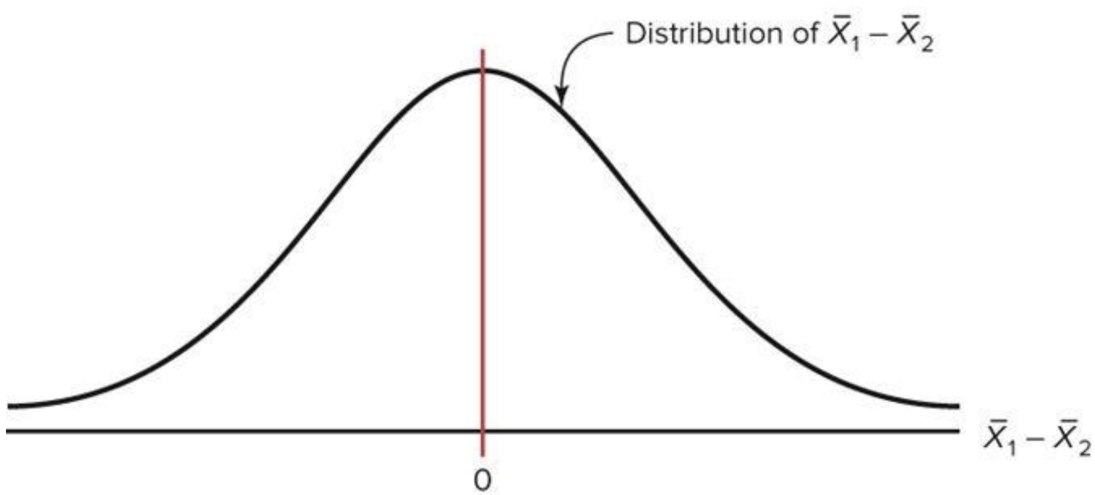
- **5-1** Large-Sample Confidence Intervals for a Population Mean
- **5-2** Confidence Intervals for Proportions
- **5-3** Small-Sample Confidence Intervals for a Population Mean
- **5-4 Confidence Intervals for the Difference Between Two Means**
- **5-7 Confidence Intervals with Paired Data**

Introduction

- We now investigate examples in which we wish to estimate the difference between the means of two populations.
- The data will consist of two samples, one from each population.
- We will compute the difference of the sample means and the standard deviation of that difference.

Introduction…

- The theory behind testing the difference between two means is based on selecting pairs of samples and comparing the means of the pairs.
- The population means do not need to be known.
- Distribution of differences of means of pairs of samples is shown below.



Example: Lightbulbs (p.356)

- Assume that a new design of lightbulb has been developed that is thought to last longer than an old design.
- A simple random sample of **144 new lightbulbs** has an **average lifetime of 578 hours** and a **standard deviation of 22 hours**.
- A simple random sample of **64 old lightbulbs** has an **average lifetime of 551 hours** and a **standard deviation of 33 hours**.
- The samples are **independent**, in that the lifetimes for one sample do not influence the lifetimes for the other
- We wish to find a **95% confidence interval** for the difference between the mean lifetimes of lightbulbs of the two designs.

Example: Lightbulbs (p.357)

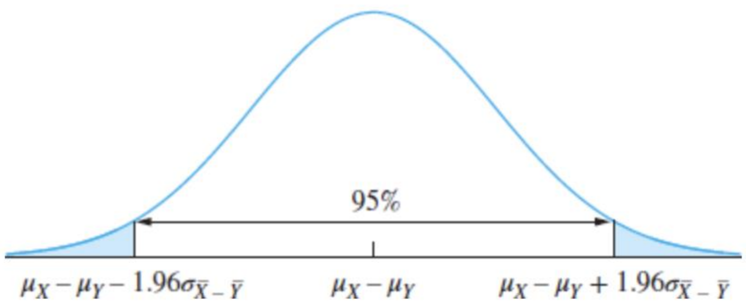


FIGURE 5.13 The observed difference $\overline{X} - \overline{Y} = 27$ is drawn from a normal distribution with mean $\mu_X - \mu_Y$ and standard deviation $\sigma_{\overline{X}-\overline{Y}} = \sqrt{\sigma_X^2/144 + \sigma_Y^2/64}$.

Estimating the population standard deviations σ_X and σ_Y with the sample standard deviations $s_X = 22$ and $s_Y = 33$, respectively, we estimate $\sigma_{\overline{X}-\overline{Y}} \approx \sqrt{22^2/144 + 33^2/64} =$ 4.514. The 95% confidence interval for $\mu_X - \mu_Y$ is therefore 578 − 551 ± 1.96(4.514), or 27 ± 8.85.

Summary

Let $X_1, \ldots, X_{n_X}$ be a *large* random sample of size n_X from a population with mean μ_X and standard deviation σ_X, and let $Y_1, \ldots, Y_{n_Y}$ be a *large* random sample of size n_Y from a population with mean μ_Y and standard deviation σ_Y. If the two samples are independent, then a level $100(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$ is

$$\boxed{\overline{X} - \overline{Y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \tag{5.16}$$

When the values of σ_X and σ_Y are unknown, they can be replaced with the sample standard deviations s_X and s_Y.

Formula for the z Confidence Interval for Difference Between Two Means

$$(\overline{X}_1 - \overline{X}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\overline{X}_1 - \overline{X}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Example: Leisure Time

A study using two random samples of 35 people each found that the average amount of time those in the age group of 26–35 years spent per week on leisure activities was 39.6 hours, and those in the age group of 46–55 years spent 35.4 hours. Assume that the population standard deviation for those in the first age group found by previous studies is 6.3 hours, and the population standard deviation of those in the second group found by previous studies was 5.8 hours.

Find the **95% confidence interval** for the difference between the means.

Example: Leisure Time…

SOLUTION

Substitute in the formula, using $z_{\alpha/2} = 1.96$.

$$(\overline{X}_1 - \overline{X}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\overline{X}_1 - \overline{X}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$(39.6 - 35.4) - 1.96\sqrt{\frac{6.3^2}{35} + \frac{5.8^2}{35}} < \mu_1 - \mu_2 < (39.6 - 35.4) + 1.96\sqrt{\frac{6.3^2}{35} + \frac{5.8^2}{35}}$$

$$4.2 - 2.8 < \mu_1 - \mu_2 < 4.2 + 2.8$$

$$1.4 < \mu_1 - \mu_2 < 7.0$$

(The confidence interval obtained from the TI-84 is $1.363 < \mu_1 - \mu_2 < 7.037$.)

Example 5.23 (p.357)

- The chemical composition of soil varies with depth. Fifty specimens were each taken at depths 50 and 250 cm.

- At a depth of 50 cm, the average NO_3 concentration (in mg/L) was 88.5 with a standard deviation of 49.4.

- At a depth of 250 cm, the average concentration was 110.6 with a standard deviation of 51.5.

- Find a **95% confidence interval** for the difference between the NO3 concentrations at the two depths.

Example 5.23 (p.357)…

Solution

Let $X_1, \ldots, X_{50}$ represent the concentrations of the 50 specimens taken at 50 cm, and let $Y_1, \ldots, Y_{50}$ represent the concentrations of the 50 specimens taken at 250 cm. Then $\overline{X} = 88.5$, $\overline{Y} = 110.6$, $s_X = 49.4$, and $s_Y = 51.5$. The sample sizes are $n_X = n_Y = 50$. Both samples are large, so we can use expression (5.16). Since we want a 95% confidence interval, $z_{\alpha/2} = 1.96$. The 95% confidence interval for the difference $\mu_Y - \mu_X$ is $110.6 - 88.5 \pm 1.96\sqrt{49.4^2/50 + 51.5^2/50}$, or 22.1 ± 19.8.

5-7 Confidence Intervals with Paired Data

- The method discussed so far for finding confidence intervals on the basis of two samples have required that the samples be independent.
- In some cases, it is better to design an experiment so that each item in one sample is paired with an item in the other.

Example (p.372)

- A tire manufacturer wishes to compare the tread wear of tires made of a new material with that of tires made of a conventional material.
- One tire of each type is placed on each front wheel of each of 10 front-wheel-drive automobiles.
- The choice as to which type of tire goes on the right wheel and which goes on the left is made with the flip of a coin.
- Each car is driven for 40,000 miles, then the tires are removed, and the depth of the tread on each is measured.
- The results are presented in Figure 5.15.

Example (p.372)

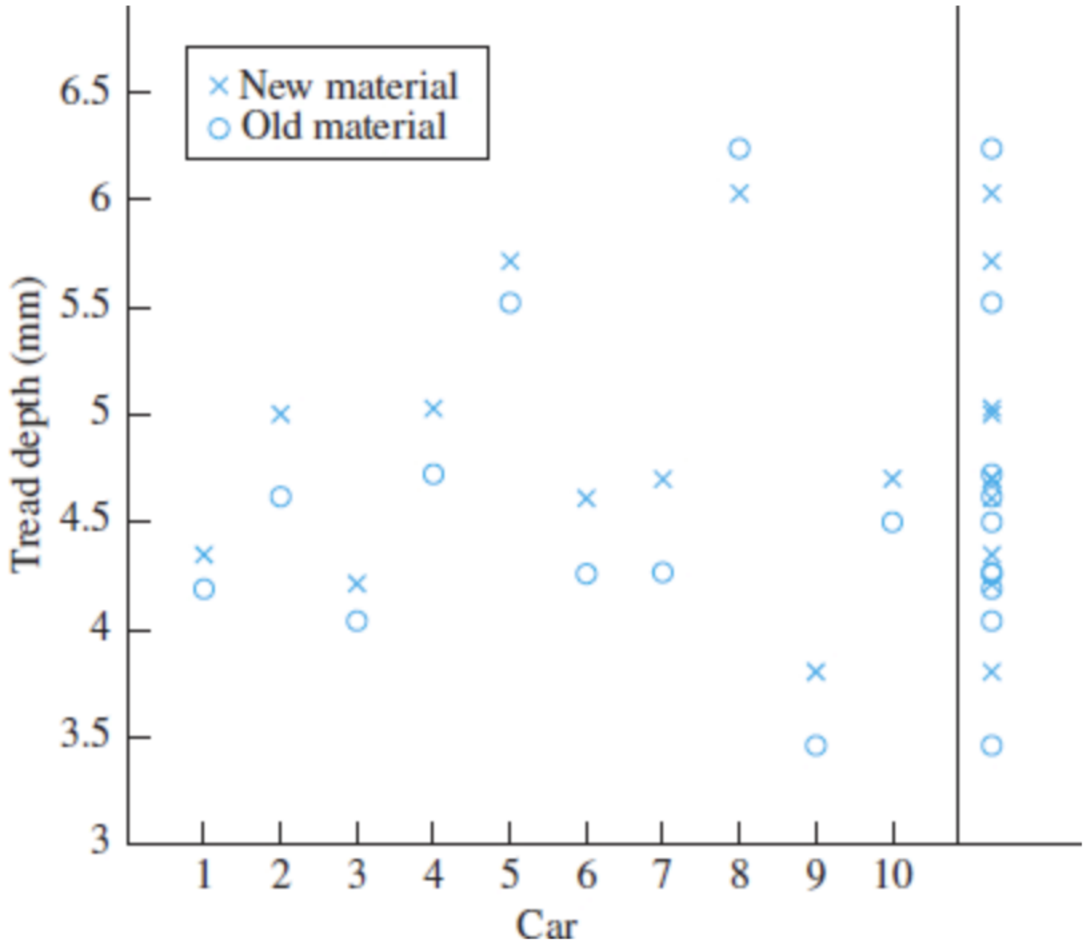


FIGURE 5.15 Tread depth for 10 pairs of tires.

Example (p.372)…

- Table 5.1 presents, for each car, the depths of tread for both the tires as well as the difference between them.

- We wish to find a 95% confidence interval for the mean difference in tread wear between old and new materials in a way that takes advantage of the reduced variability produced by the paired design.

TABLE 5.1 Depths of tread, in mm, for tires made of new and old material

	Car									
	1	2	3	4	5	6	7	8	9	10
New material	4.35	5.00	4.21	5.03	5.71	4.61	4.70	6.03	3.80	4.70
Old material	4.19	4.62	4.04	4.72	5.52	4.26	4.27	6.24	3.46	4.50
Difference	0.16	0.38	0.17	0.31	0.19	0.35	0.43	–0.21	0.34	0.20

Example (p.372)…

- Let $(X_1, Y_1), \ldots, (X_{10}, Y_{10})$ be the 10 observed pairs, with X_i representing the tread on the tire made from the new material on the *i*-th car and Y_i representing the tread on the tire made from the old material on the *i*-th car.
- Let $D_i = X_i - Y_i$ represent the difference between the treads for the tires on the *i*-th car.
- Let ¼$_X$ and ¼$_Y$ represent the population means for X and Y, respectively.
- We wish to find a 95% confidence interval for the difference ¼$_X$ − ¼$_Y$.
- Let ¼$_D$ represent the population mean of the differences.

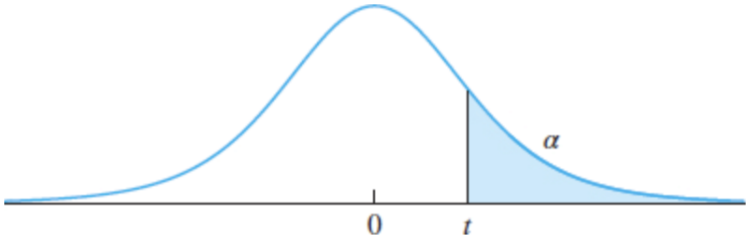
Example (p.372)…

- Since the sample $D_1, \ldots, D_{10}$ is a random sample from a population with mean ¼$_D$, we can use one-sample methods to find confidence intervals for ¼$_D$.
- In this example, since the sample size is small, we use the Student's t-method.
- The observed values of the sample mean and sample standard deviation are

$$\bar{D} = 0.232 \qquad s_D = 0.183$$

- The appropriate t value is $\mathbf{t_{9,0.025} = 2.262}$

TABLE A.3 Upper percentage points for the Student's *t* distribution



	α						
v	**0.40**	**0.25**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947

Example (p.372)...

- The confidence interval using expression (5.9 in Section 5.3) is therefore

$$0.232 \pm (2.262)(0.183)/\sqrt{10}, \text{ or } (0.101,\ 0.363).$$

Summary

Let $D_1,\ldots, D_n$ be a *small* random sample ($n \leq 30$) of differences of pairs. If the population of differences is approximately normal, then a level $100(1-\alpha)\%$ confidence interval for the mean difference ¼$_D$ is given by

$$\boxed{\overline{D} \pm t_{n-1,\alpha/2}\frac{s_D}{\sqrt{n}}} \tag{5.24}$$

where s_D is the sample standard deviation of $D_1,\ldots, D_n$. Note that this interval is the same as that given by expression (5.9).

If the sample size is large, a level $100(1-\alpha)\%$ confidence interval for the mean difference ¼$_D$ is given by

$$\boxed{\overline{D} \pm z_{\alpha/2}\sigma_{\overline{D}}} \tag{5.25}$$

In practice $\sigma_{\overline{D}}$ is approximated with $s_D/\sqrt{n}$. Note that this interval is the same as that given by expression (5.1).

Example: Cholesterol Levels

A dietitian wishes to see if a person's cholesterol level will change if the diet is supplemented by a certain mineral. Six randomly selected subjects were pretested, and then they took the mineral supplement for a 6-week period. The results are shown in the table. (Cholesterol level is measured in milligrams per deciliter.)

Subject	1	2	3	4	5	6
Before (X_1)	210	235	208	190	172	244
After (X_2)	190	170	210	188	173	228

Find the 90% confidence interval for the data

Example: Cholesterol Levels - SOLUTION

- Make a table (*this is a suggestion not a requirement*)

Before (X_1)	After (X_2)	A $D = X_1 - X_2$	B $D^2 = (X_1 - X_2)^2$
210	190		
235	170		
208	210		
190	188		
172	173		
244	228		

- Then complete the table (*this is a suggestion not a requirement*)

Before (X_1)	After (X_2)	A $D = X_1 - X_2$	B $D^2 = (X_1 - X_2)^2$
210	190	20	400
235	170	65	4225
208	210	–2	4
190	188	2	4
172	173	–1	1
244	228	16	256
		$\Sigma D = 100$	$\Sigma D^2 = 4890$

Example: Cholesterol Levels – SOLUTION…

Before (X_1)	After (X_2)	A $D = X_1 - X_2$	B $D^2 = (X_1 - X_2)^2$
210	190	20	400
235	170	65	4225
208	210	–2	4
190	188	2	4
172	173	–1	1
244	228	16	256
		$\Sigma D = 100$	$\Sigma D^2 = 4890$

- Find the mean and the standard deviation of the differences

$$\overline{D} = \frac{\Sigma D}{n} = \frac{100}{6} = 16.7$$

$$s_D = \sqrt{\frac{n\Sigma D^2 - (\Sigma D)^2}{n(n-1)}}$$

$$= \sqrt{\frac{6 \cdot 4890 - 100^2}{6(6-1)}}$$

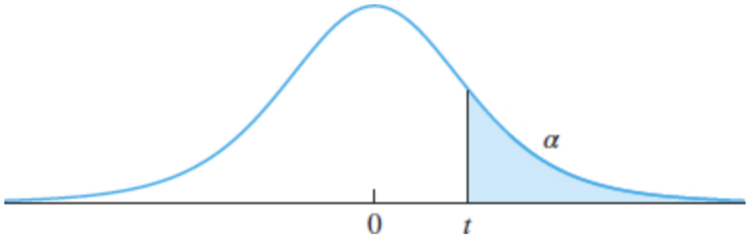
$$= \sqrt{\frac{29{,}340 - 10{,}000}{30}}$$

$$= 25.4$$

Example: Cholesterol Levels – SOLUTION…

- Find $t_{5,0.05} = \pm 2.015$

TABLE A.3 Upper percentage points for the Student's *t* distribution



	α						
v	**0.40**	**0.25**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169

Example: Cholesterol Levels – SOLUTION…

- Substitute into Equation (5.24)

$$\boxed{\overline{D} - t_{\alpha/2}\frac{s_D}{\sqrt{n}} < \mu_D < \overline{D} + t_{\alpha/2}\frac{s_D}{\sqrt{n}}}$$

$$16.7 - 2.015 \cdot \frac{25.4}{\sqrt{6}} < \mu_D < 16.7 + 2.015 \cdot \frac{25.4}{\sqrt{6}}$$

$$16.7 - 20.89 < \mu_D < 16.7 + 20.89$$

$$-4.19 < \mu_D < 37.59$$

$$\boxed{-4.2 < \mu_D < 37.6}$$