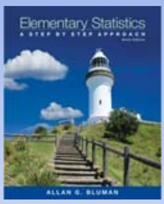


Chapter 9
(part-1)

Testing the difference between two populations

- 區間估計及檢定
  - 常態母體
    - 平均數
      - 一個母體
      - 二個母體(平均數差)
        - 獨立母體
        - 相依母體
    - 變異數
      - 一個母體
      - 二個母體(變異數比)
  - 點二項母體
    - 一個母體比例
    - 二個母體比例差

Chapter 9 Objectives

1. Test the difference between **two means**
   - Independent samples
   - Dependent samples.
2. Test the difference between **two proportions**.
3. Test the difference between **two variances** or **standard deviations**.

Testing the Difference Between Two Means

Dependent- or independent-samples

Time spent in physical activity

Dependent- or independent-samples

Product lifetime

Dependent- or independent-samples

Difference between two proportions

Statistics Today

To Vaccinate or Not to Vaccinate? Small or Large?

Influenza is a serious disease among the elderly, especially those living in nursing homes. Those residents are more susceptible to influenza than elderly persons living in the community because the former are usually older and more debilitated, and they live in a closed environment where they are exposed more so than community residents to the virus if it is introduced into the home. Three researchers decided to investigate the use of vaccine and its value in determining outbreaks of influenza in small nursing homes.

These researchers surveyed 83 licensed homes in seven counties in Michigan. Part of the study consisted of comparing the number of people being vaccinated in small nursing homes (100 or fewer beds) with the number in larger nursing homes (more than 100 beds). Unlike the statistical methods presented in Chapter 8, these researchers used the techniques explained in this chapter to compare two sample proportions to see if there was a significant difference in the vaccination rates of patients in small nursing homes compared to those in large nursing homes. See Statistics Today—Revisited.

rden, Arnold S. Monto, and Suzanne E. Ohmit, "Vaccine Use and the Risk of Outbreaks in a Sample es During an Influenza Epidemic," *American Journal of Public Health* 85, no. 3 (March 1995), pyright 1995 by the American Public Health Association.

Nursing Home

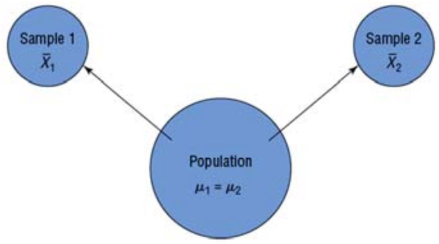
Dependent- or independent-samples

樣本獨立性

- 獨立樣本

若從一個母體抽出的樣本不影響從另一個母體抽出的樣本，則這兩個樣本為獨立樣本。

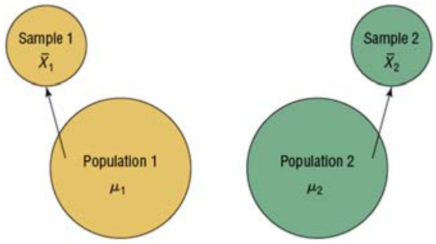
Hypothesis Testing Situations in the Comparison of Means



(a) Difference is not significant

Do not reject H_0: $\mu_1 = \mu_2$ since $\bar{X}_1 - \bar{X}_2$ is not significant.

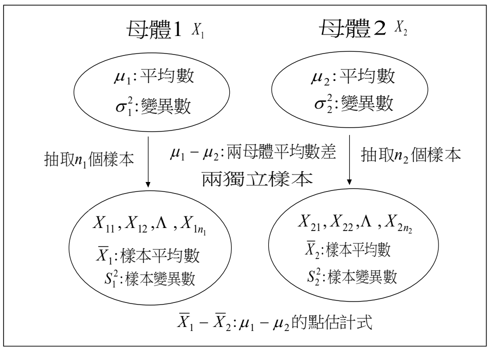
Hypothesis Testing Situations in the Comparison of Means

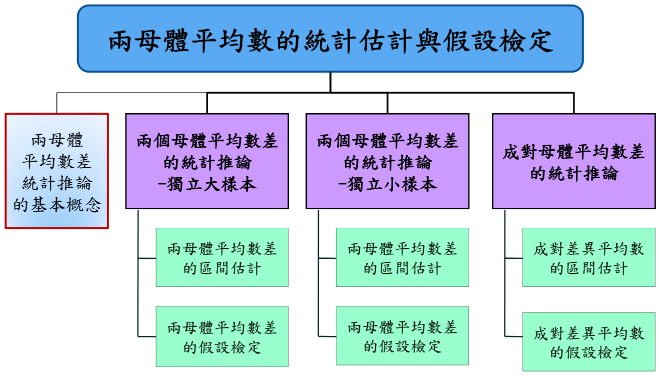


(b) Difference is significant

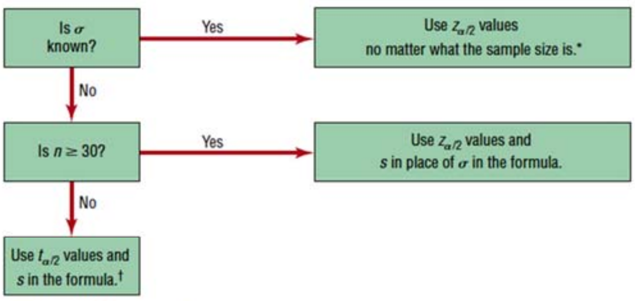
Reject H_0: $\mu_1 = \mu_2$ since $\bar{X}_1 - \bar{X}_2$ is significant.

兩個獨立母體平均數差的區間估計





Outline



*Variable must be normally distributed when $n < 30$.
†Variable must be approximately normally distributed.

Section 9-1

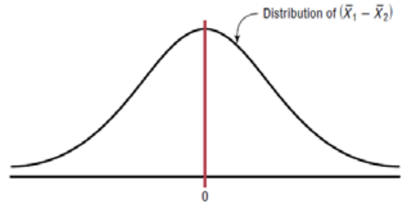
Testing the Difference Between Two Means Using the *z* Test (independent samples)

Introduction

Question:
Whether there is a **difference** in the average age of students who enroll at a community college and those who enroll at Chang Gung University?

$$H_0: \mu_1 = \mu_2 \qquad H_0: \mu_1 - \mu_2 = 0$$
$$H_1: \mu_1 \neq \mu_2 \qquad H_1: \mu_1 - \mu_2 \neq 0$$

$$H_0: \mu_1 = \mu_2 \qquad H_0: \mu_1 - \mu_2 = 0$$
$$H_1: \mu_1 \neq \mu_2 \qquad H_1: \mu_1 - \mu_2 \neq 0$$



Testing the difference between two means: using the *z* test

Assumptions:

1. Both samples are random samples.
2. The samples must be independent of each other. That is, there can be no relationship between the subjects in each sample.
3. The standard deviations of both populations must be known, if the sample sizes are less than 30, the populations must be normally or approximately normally distributed.

$$\text{Test value} = \frac{(\text{observed value}) - (\text{expected value})}{\text{standard error}}$$

- The variance of the difference $\bar{X}_1 - \bar{X}_2$ is equal to the sum of the individual variances of $\bar{X}_1$ and $\bar{X}_2$. That is,

$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = \sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2}$$

where $\sigma^2_{\bar{X}_1} = \frac{\sigma_1^2}{n_1}$ and $\sigma^2_{\bar{X}_2} = \frac{\sigma_2^2}{n_2}$

So the standard deviation of $\bar{X}_1 - \bar{X}_2$ is

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$H_0: \mu_1 - \mu_2 = 0$$

0

when the null hypothesis is true

$$\text{Test value} = \frac{\overbrace{(\text{observed value})}^{\bar{X}_1 - \bar{X}_2} - (\text{expected value})}{\underbrace{\text{standard error}}_{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}}$$

Testing the Difference Between Two Means: Large Samples

Formula for the *z* test for comparing two means from independent populations

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Types of tests

Right-tailed		
$H_0: \mu_1 \le \mu_2$ $H_1: \mu_1 > \mu_2$	or	$H_0: \mu_1 - \mu_2 \le 0$ $H_1: \mu_1 - \mu_2 > 0$

Left-tailed		
$H_0: \mu_1 \ge \mu_2$ $H_1: \mu_1 < \mu_2$	or	$H_0: \mu_1 - \mu_2 \ge 0$ $H_1: \mu_1 - \mu_2 < 0$

The procedure of testing

Step 1 State the hypotheses and identify the claim.

Step 2 Find the critical value(s).

Step 3 Compute the test value.

Step 4 Make the decision.

Step 5 Summarize the results.

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Example 9-1: 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.

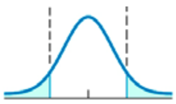
At α = 0.05, can it be concluded that there is a significant difference in the average times each group spends on leisure activities?

Step 1: State the hypotheses and identify the claim.

$H_0: \mu_1 = \mu_2$ and $H_1: \mu_1 \ne \mu_2$ (claim)

Step 2: Find the critical value.

The critical value is $z = \pm 1.96$.



Example 9-1: Leisure time

Step 3: Compute the test value.

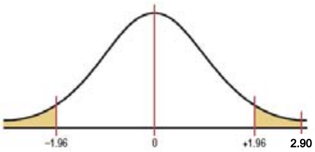
$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{(39.6 - 35.4) - 0}{\sqrt{\frac{6.3^2}{35} + \frac{5.8^2}{35}}} = 2.90$$

Example 9-1: Leisure time

Step 4: Make the decision.

Reject the null hypothesis at α = 0.05, since 2.90 > 1.96.



Step 5: Summarize the results.

There is enough evidence to support the claim that the means are not equal. Hence, the average of the times spent on leisure activities is different for the groups.

Example 9-2: College Sports Offerings

A researcher hypothesizes that the average number of sports that colleges offer for males is greater than the average number of sports that colleges offer for females. A sample of the number of sports offered by colleges is shown. At α = 0.10, is there enough evidence to support the claim? Assume σ_1 and σ_2 = 3.3.

Males					Females				
6	11	11	8	15	6	8	11	13	8
6	14	8	12	18	7	5	13	14	6
6	9	5	6	9	6	5	5	7	6
6	9	18	7	6	10	7	6	5	5
15	6	11	5	5	16	10	7	8	5
9	9	5	5	8	7	5	5	6	5
8	9	6	11	6	9	18	13	7	10
9	5	11	5	8	7	8	5	7	6
7	7	5	10	7	11	4	6	8	7
10	7	10	8	11	14	12	5	8	5

Example 9-2: College Sports Offerings

Step 1: State the hypotheses and identify the claim.

H_0: $\mu_1 = \mu_2$ and H_1: $\mu_1 \neq \mu_2$ (claim)

Step 2: Compute the test value.

For the males: $\bar{X}_1 = 8.6$ and $\sigma_1 = 3.3$

For the females: $\bar{X}_2 = 7.9$ and $\sigma_2 = 3.3$

Substitute in the formula.

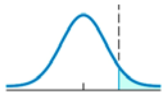
$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \frac{(8.6 - 7.9) - (0)}{\sqrt{\dfrac{3.3^2}{50} + \dfrac{3.3^2}{50}}} = 1.06$$

Example 9-2: College Sports Offerings

Step 3: Find the *P-value.*

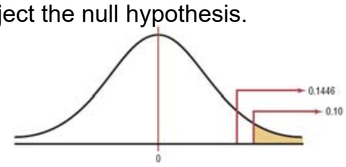
For z = 1.06, the area is 0.8554.

The P-value is 1.0000 - 0.8554 = 0.1446.



Step 4: Make the decision.

Do not reject the null hypothesis.



Step 5: Summarize the results.

There is not enough evidence to support the claim that colleges offer more sports for males than they do for females.

Question:

The students at a community college are, on average, 5 years older than those at Chang Gung University.

$$\begin{cases} H_0 : \mu_1 - \mu_2 \le 5 \\ H_1 : \mu_1 - \mu_2 > 5 \end{cases}$$

$H_0 : \mu_1 - \mu_2 \le 5$ 5

$\bar{X}_1 - \bar{X}_2$

$$\text{Test value} = \frac{(\text{observed value}) - (\text{expected value})}{\text{standard error}}$$

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Confidence Intervals for the Difference Between Two Means

Formula for the z confidence interval for the difference between two means from independent populations

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

Example 9-3: Confidence Intervals

Find the 95% confidence interval for the difference between the means for the data in Example 9–1.

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

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

$$1.36 < \mu_1 - \mu_2 < 7.03$$

Section 9-2

Testing the Difference Between Two Means:

Using the t Test

(independent samples)

Testing the Difference Between Two Means of Independent Samples: Using the *t* Test

Assumptions

1. The samples are random samples.
2. The sample data are independent of one another.
3. When the two sample sizes both are less than 30, the populations must be normally or approximately normally distributed.

Testing the difference between two means: using the *t* test

$$\text{Test value} = \frac{(\text{observed value}) - (\text{expected value})}{\text{standard error}}$$

$$t = \frac{\left(\bar{X}_1 - \bar{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

where the **degrees of freedom** are equal to **the smaller of**

$\boxed{(n_1 - 1 \text{ or } n_2 - 1)}$ or $(n_1 + n_2 - 2)$.

Example 9-4: Weights of Newborn Infants

A researcher wishes to see if the average weights of newborn male infants are different from the average weights of newborn female infants. She selects a random sample of 10 male infants and finds the mean weight is 3300 grams and the standard deviation of the sample is 20 grams. She selects a random sample of 8 female infants and finds that the mean weight is 3100 grams and the standard deviation of the sample is 150 grams. Can it be concluded at α = 0.05 that the mean weight of the males is different from the mean weight of the females? Assume that the variables are normally distributed.

Step 1: State the hypotheses and identify the claim.

H_0: $\mu_1 = \mu_2$ and H_1: $\mu_1 \neq \mu_2$ (claim)

Example: Weights of Newborn Infants

Step 2: Find the critical values.

The two-tailed with α = 0.05, the degrees of freedom are the smaller of $n_1 - 1$ or $n_2 - 1$.

➔ d.f.= 8 – 1 = 7.

➔ the critical values are -2.365 and 2.365.

Step 3: Find the test value.

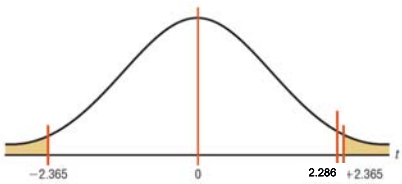
$$t = \frac{\left(\bar{X}_1 - \bar{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \frac{(3300 - 3100) - 0}{\sqrt{\dfrac{220^2}{10} + \dfrac{150^2}{8}}}$$

$$= 2.286$$

Example: Weights of Newborn Infants

Step 4: Make the decision.

Do not reject the null hypothesis.



Step 5: Summarize the results.

There is not enough evidence to support the claim that the mean of the weights of the male infants is different from the mean of the weights of the female infants.

Confidence intervals for the difference between two means

- Formula for the *t* confidence interval for the difference between two means from independent populations with unequal variances

d.f. smaller value of $(n_1 - 1$ or $n_2 - 1)$ or $(n_1 + n_2 - 2)$.

$$\left(\bar{X}_1 - \bar{X}_2\right) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \left(\mu_1 - \mu_2\right)$$

$$< \left(\bar{X}_1 - \bar{X}_2\right) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Example: Confidence Intervals

Find the 95% confidence interval for the difference between the means for the data in last example.

$$\left(\bar{X}_1-\bar{X}_2\right)-t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}<\mu_1-\mu_2<\left(\bar{X}_1-\bar{X}_2\right)+t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

$$(3300-3100)-2.365\sqrt{\frac{220^2}{10}+\frac{150^2}{8}}<\mu_1-\mu_2<(3300-3100)+2.365\sqrt{\frac{220^2}{10}+\frac{150^2}{8}}$$

$$-6.89<\mu_1-\mu_2<406.89$$

p. 502

- About degree of freedom

Section 9-3

Testing the Difference Between Two Means: Dependent Samples

Testing the difference between two means: dependent samples

- Examples for dependent samples
  - 成對資料(paired data)
  - 減肥藥的效用: (Data:減肥前&減肥後)
  - 新藥物療效: (Data:服用前&服用後)
  - 增高器效用: (Data:使用前&使用後)
  - ...

Testing the difference between two means of dependent samples: using the *t* test

Assumptions

1. The samples are random samples.
2. The sample data are dependent.
3. When the two sample sizes both are less than 30, the populations must be normally or approximately normally distributed.

Testing the difference between two means: dependent samples

$$D=X_1-X_2$$

$$\bar{D}=\frac{\sum D}{n} \text{ and } S_D=\sqrt{\frac{\sum(D_i-\bar{D})^2}{n-1}}$$

$$\text{Test value}=\frac{(\text{observed value})-(\text{expected value})}{\text{standard error}}$$

$$t=\frac{\bar{D}-\mu_D}{s_D/\sqrt{n}} \qquad \text{with d.f.}=n-1$$

Example: Vitamin for Strength

A physical education director claims by taking a special vitamin, a weight lifter can increase his strength. Eight athletes are selected and given a test of strength, using the standard bench press. After 2 weeks of regular training, supplemented with the vitamin, they are tested again. Test the effectiveness of the vitamin regimen at $\alpha = 0.05$. Each value in the data represents the maximum number of pounds the athlete can bench-press. Assume that the variable is approximately normally distributed.

Athlete	1	2	3	4	5	6	7	8
Before (X_1)	210	230	182	205	262	253	219	216
After (X_2)	219	236	179	204	270	250	222	216

Example: Vitamin for Strength

Athlete	1	2	3	4	5	6	7	8
Before (X_1)	210	230	182	205	262	253	219	216
After (X_2)	219	236	179	204	270	250	222	216

Step 1: State the hypotheses and identify the claim.

H_0: $\mu_D \geqq 0$ and H_1: $\mu_D < 0$ (claim)

Step 2: Find the critical value.

The degrees of freedom are $n - 1 = 8 - 1 = 7$.
The critical value for a left-tailed test with $\alpha = 0.05$ is $t = -1.895$.

Example: Vitamin for Strength

Step 3: Compute the test value.

Before (X_1)	After (X_2)	$D = X_1 - X_2$	D^2
210	219	-9	81
230	236	-6	36
182	179	3	9
205	204	1	1
262	270	-8	64
253	250	3	9
219	222	-3	9
216	216	0	0
		$\Sigma D = -19$	$\Sigma D^2 = 209$

$$\bar{D} = \frac{\sum D}{n} = \frac{-19}{8} = -2.375$$

$$s_D = \sqrt{\frac{n\sum D^2 - \left(\sum D\right)^2}{n(n-1)}} = \sqrt{\frac{8 \cdot 209 - (-19)^2}{8 \cdot 7}} = 4.84$$

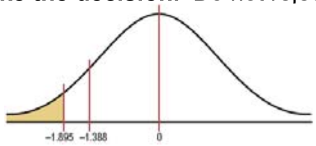
Example: Vitamin for Strength

Step 3: Compute the test value.

$$\bar{D} = -2.375, \quad s_D = 4.84$$

$$t = \frac{\bar{D} - \mu_D}{s_D/\sqrt{n}} = \frac{-2.375 - 0}{4.84/\sqrt{8}} = -1.388$$

Step 4: Make the decision. Do not reject the null.



Step 5: Summarize the results.
There is not enough evidence to support the claim that the vitamin increases the strength of weight lifters.

Example 9-7: 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 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.) Can it be concluded that the cholesterol level has been changed at $\alpha = 0.10$? Assume the variable is approximately normally distributed.

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

Example 9-7: Cholesterol Levels

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

Step 1: State the hypotheses and identify the claim.

H_0: $\mu_D = 0$ and H_1: $\mu_D \neq 0$ (claim)

Step 2: Find the critical value.

The degrees of freedom are 5. At $\alpha = 0.10$, the critical values are ± 2.015.

Example 9-7: Cholesterol Levels

Step 3: Compute the test value.

Before (X_1)	After (X_2)	$D = X_1 - X_2$	D^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
		$\Sigma D = 100$	$\Sigma D^2 = 4890$

$$\bar{D} = \frac{\sum D}{n} = \frac{100}{6} = 16.7$$

$$s_D = \sqrt{\frac{n\sum D^2 - \left(\sum D\right)^2}{n(n-1)}} = \sqrt{\frac{6 \cdot 4890 - (100)^2}{6 \cdot 5}} = 25.4$$

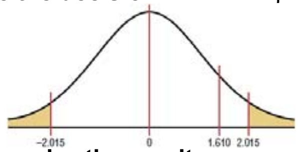
Example 9-7: Cholesterol Levels

Step 3: Compute the test value.

$$\bar{D} = 16.7, \ s_D = 25.4$$

$$t = \frac{\bar{D} - \mu_D}{s_D/\sqrt{n}} = \frac{16.7 - 0}{25.4/\sqrt{6}} = 1.610$$

Step 4: Make the decision. Do not reject the null.



Step 5: Summarize the results.
There is not enough evidence to support the claim that the mineral changes a person's cholesterol level.

Confidence Interval for the Mean Difference

Formula for the *t* confidence interval for the mean difference

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

$$\text{d.f.} = n - 1$$

Example 9-8: Confidence Intervals

Find the 90% confidence interval for the difference between the means for the data in Example 9–7.

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

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

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

$$-4.19 < \mu_D < 37.59$$

Since 0 is contained in the interval, the decision is to not reject the null hypothesis H_0: $\mu_D = 0$.