

# air pollution training institute



## introduction to environmental statistics

self-instructional course

SI 473

MODULE III

*t* TESTS

3.



UNIT 1 - ONE-TAILED, ONE-SAMPLE  $t$  TEST

1

This module covers the  $t$  test of statistical significance. The  $t$  test is used in testing the significance of differences between means, either between a sample mean and a postulated mean, or between two population means. This unit will handle the one-tailed  $t$  test as it would be used to test a sample mean against a postulated mean (or a standard). Pay close attention to the concept of degrees of freedom (this was mentioned earlier in the course, but not explained).

---

Answer: 14

15a

---

What is the critical value listed in the intersection of  $t_{.050}$  and  $d.f. = 14$ ? \_\_\_\_\_

16

Let's review a minute:

The  $t$  test is used to test which of the following null hypotheses:

- a.  $H_0: \sigma_1^2 = \sigma_2^2$
- b.  $H_0: \mu_1 = \mu_2$
- c.  $H_0: \mu = \text{any no. (any sample size)}$
- d.  $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

Answer: 1.761

16a

Like the  $z$  distribution, the  $t$  distribution is symmetrical and therefore the table doesn't list negative values. In the one-tailed test with the following alternative

$$H_1: \mu > \text{any number}$$

we are interested only in the positive critical value.

$$\text{If } H_1: \mu < \text{any number}$$

we would be interested only in the \_\_\_\_\_ value.

17

Answer: b.  $H_0: \mu_1 = \mu_2$

2a

c.  $H_0: \mu = \text{any number (any sample size)}$

Notice that b. is a two-sample hypothesis, while c. is a one-sample hypothesis.

---

In this unit, we will be concentrating on the  $t$  test for testing:

3

$H_0: \mu = \text{any number (any sample size)}$  against either of the following alternatives

$H_1: \mu > \text{that number}$   
or

$H_1: \mu < \text{that number}$

In either of these situations we will use the (two/one) tailed, one-sample  $t$  test.

---

Answer: negative

17a

In our present example we have

18

$H_1: \mu > 100$

so we are only interested in exceeding (+/-) 1.761.

4

---

Since the use of the  $t$  test requires that we can determine the *degrees of freedom*, let's consider that concept now.

If you wish to know what the values are in a sample of size 10 whose mean is 20, how many of those values are free to vary? (No answer necessary.)

In other words, given the following 10 values:

$x_1$   $x_2$   $x_3$   $x_4$   $x_5$   $x_6$   $x_7$   $x_8$   $x_9$   $x_{10}$

and we know  $\bar{x} = 20$ , how many of those 10 values can we arbitrarily choose, before we no longer have a choice (the remaining values are determined by the fact that the mean of the values equals 20)? (No answer necessary.)

Let's try it and see.

---

Answer: +

18a

---

Any obtained  $t$  in this situation greater than 1.761 would result in (acceptance/rejection) of the null hypothesis.

19

Let's assign  $x_1$  the value 15.

Can we now arbitrarily set  $x_2$ ? \_\_\_\_\_

---

Answer: rejection

19a

---

20

Back to our problem.

Step 6: Make the appropriate decision. Since our obtained  $t$  of 2.53 is greater than the tabled  $t$  value of 1.761, we reject  $H_0: \mu = 100$ . We can now say that our sample mean of 110 is significantly higher than the postulated mean at the 5% level of significance.

If the null hypothesis were true, we would expect to obtain a mean of 110 by chance alone less than 5% of the time, so our difference is more likely to be the result of something other than chance.

Answer: yes

5a

---

O. K., then let's say  $x_2 = 16$ .

6

Can we arbitrarily assign  $x_3$  and still end up with  $\bar{x} = 20$ ? \_\_\_\_\_

---

Suppose the problem had been to see whether the sample mean is significantly lower than 100:

21

With  $\alpha = .05$  and 14 *d.f.*,

What is the critical value for the following hypotheses?

$$H_0: \mu = 100$$

$$H_1: \mu < 100$$

Answer: yes

6a

---

Let's skip a few to get to the answer of how many values are free to vary.

7

Assume:  $x_1 = 15$   
 $x_2 = 16$   
 $x_3 = 17$   
 $x_4 = 18$   
 $x_5 = 19$   
 $x_6 = 21$   
 $x_7 = 22$   
 $x_8 = 23$   
 $x_9 = 24$

Can we arbitrarily set the value of  $x_{10}$  and still have  $\bar{x} = 20$ ? \_\_\_\_\_

---

Answer: -1.761

21a

---

Any obtained  $t$  in this situation less than -1.761 would result in (acceptance/rejection) of the null hypothesis.

22

Answer: No. Given the values of  $x_1$  through  $x_9$ , we have a sum of 175. To have a mean of 20,  $x_{10}$  must equal 25.

7a

---

To get back to the original question, how many of the 10 values were free to vary, before the remaining values had to be a certain number?

8

# \_\_\_\_\_

---

Answer: rejection

22a

---

With  $\alpha = .01$  and 20 *d.f.*

23

What is the critical value for the following hypotheses?

$$H_0: \mu = 100$$

$$H_1: \mu > 100$$

Answer: 9

8a

---

This then is the number of degrees of freedom.

9

degrees of freedom (*d.f.*) =  $n - 1$  for the one-sample  $t$  test.

---

Answer: +2.528

23a

---

Any obtained  $t$  in this situation greater than +2.528 would result in (acceptance/rejection) of the null hypothesis.

24

Let's consider the one-tailed, one-sample  $t$  test.

Assume we have a sample of 15 readings of TSP concentrations, and we want to see if the mean of these readings is significantly higher, at the 5% level, than the assumed mean of  $100 \mu\text{g}/\text{m}^3$ . The readings are:

120	85	115	100	112	94	90	110
124	117	109	127	131	88	128	

First, find the mean and standard deviation, using the following formulas:

$$\bar{x} = \frac{\sum x_i}{n}$$

$$s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n - 1}}$$

---

Answer: rejection

24a

---

25

Applying the statistical problem solving procedure, you will now solve three problems. The six steps of the procedure are presented on page 11 of the Guide. You should have this page in front of you while you do the problems.

Answer:  $\bar{x} = 110$

$$s = \sqrt{\frac{184,794 - \frac{(1650)^2}{15}}{14}} = \sqrt{235.3} = 15.34$$

---

26

Problem I: A recent study of air quality in a large urban area showed a mean  $\text{NO}_2$  concentration of  $90 \mu\text{g}/\text{m}^3$  with a standard deviation of  $12.7 \mu\text{g}/\text{m}^3$  ( $n = 25$ ). At a 5% level of significance, can we conclude that the average  $\text{NO}_2$  concentration in this area is less than the standard of  $100 \mu\text{g}/\text{m}^3$ ?

Step 1: Define the problem.

$$H_0: \mu = 100$$

$$H_1: \mu > 100$$

Step 2: Select test and set  $\alpha$  level.

One-tailed, one-sample  $t$  test

$$\alpha = .05$$

Step 3: Organize data (not necessary for  $t$  test)

Step 4: Compute the statistic.

Answer:

26a

Step 1: Define the problem.

$$H_0: \mu = 100$$

$$H_1: \mu < 100$$

Step 2: Select technique and set  $\alpha$  level.

One-tailed, one-sample  $t$  test,  $\alpha = .05$

Step 3: Organize and arrange data.

Not necessary

Step 4: Compute the statistic.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$
$$= \frac{90 - 100}{12.7/\sqrt{25}}$$

$$= \frac{-10}{2.54}$$

$$= -3.94$$

Step 5: Determine critical values.

$$\text{Critical value} = -1.711$$

Step 6: Make the appropriate decision.

Reject  $H_0$ , since the obtained value is beyond the critical value.

Here is the formula for the one-tailed, one-sample  $t$  test:

$$t_{n-1} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

where:  $t_{n-1}$  = the  $t$  statistic with  $n - 1$  degrees of freedom

$\bar{x}$  = sample mean

$\mu$  = postulated mean

$s$  = sample standard deviation

$n$  = sample size

---

27

Problem II: A local manufacturer has tested the air around his factory for TSP concentrations and obtained a mean of  $300 \mu\text{g}/\text{m}^3$  ( $s = 27.8 \mu\text{g}/\text{m}^3$ ,  $n = 9$ ). He would like to know if this exceeds the standard of  $260 \mu\text{g}/\text{m}^3$  significantly at the 5% level.

Step 4: Compute the statistic

$$\begin{aligned}
 t_{n-1} &= \frac{110 - 100}{15.34/\sqrt{15}} \\
 &= \frac{10}{15.34/3.87} \\
 &= \frac{10}{3.96} \\
 &= 2.53
 \end{aligned}$$

Answer:

27a

Step 1: Define the problem.

$$H_0: \mu = 260$$

$$H_1: \mu > 260$$

Step 2: Select technique and set  $\alpha$  level.

One-tailed, one-sample  $t$  test,  $\alpha = .05$

Step 3: Organize and arrange data.

Not necessary

Step 4: Compute the statistic.

$$\begin{aligned}
 t &= \frac{\bar{x} - \mu}{s/\sqrt{n}} \\
 &= \frac{300 - 260}{27.8/\sqrt{9}} \\
 &= \frac{40}{9.26} \\
 &= 4.32
 \end{aligned}$$

Step 5: Determine critical values.

Critical value = 1.860

Step 6: Make the appropriate decision.

Reject  $H_0$ , since the obtained value is greater than the critical value.

Step 5: Consult the appropriate table.

14

Turn to Table F.

This table is titled \* \_\_\_\_\_

---

28

Problem III: For years, the air in an urban area has been thought to be very clean (mean TSP concentration =  $50 \mu\text{g}/\text{m}^3$ ). But a local citizens' group, campaigning for more stringent regulations, feels that the air quality has deteriorated lately and the current mean TSP concentration is higher than  $50 \mu\text{g}/\text{m}^3$ . They sample the air ( $n = 25$ ) and obtain a sample mean of  $54 \mu\text{g}/\text{m}^3$  with a standard deviation of  $5.2 \mu\text{g}/\text{m}^3$ . At the 5% level, can they conclude that the TSP concentration is significantly higher than  $50 \mu\text{g}/\text{m}^3$ ?

Answer: The  $t$  Distribution

14a

Answer:

28a

Step 1: Define the problem.

$$H_0: \mu = 50$$

$$H_1: \mu > 50$$

Step 2: Select technique and set  $\alpha$  level.

One-tailed, one-sample  $t$  test,  $\alpha = .05$

Step 3: Organize and arrange data.

Not necessary.

Step 4: Compute the statistic.

$$\begin{aligned} t &= \frac{\bar{x} - \mu}{s/\sqrt{n}} \\ &= \frac{54 - 50}{5.2/\sqrt{25}} \\ &= \frac{4}{1.04} \\ &= 3.85 \end{aligned}$$

Step 5: Determine critical values.

Critical value = 1.711

Step 6: Make the appropriate decision.

Reject  $H_0$ , since obtained value exceeds the critical value.

Let's see how to use the table. Across the top we see (Note: The subscript now refers to  $\alpha$  level instead of degrees of freedom.)

15

$t_{.100}$      $t_{.050}$      $t_{.025}$      $t_{.010}$      $t_{.005}$

In the table we are given the critical value ( $t_{\alpha}$ ) above which  $\alpha$  (in this case .05) of the values fall for specific degrees of freedom. In this example, we had 15 values, so our degrees of freedom =           .

Go to:

Frame 15a

Page III-1

---

29

This unit demonstrated the use of the one-tailed, one-sample  $t$  test. This test should only be used in response to the alternative hypotheses that the mean is either larger or smaller than the postulated mean (or standard). In turn, these alternative hypotheses should only be used in the case where a deviation opposite to the one postulated would have no practical significance. For example, if a minimum value is required, we would only be interested in knowing if our sample mean is significantly lower than the minimum. We would not be concerned with values higher than the minimum.

The formula for the statistic is:

$$t_{n-1} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Proceed to Unit 2.

UNIT 2 - TWO-TAILED, ONE-SAMPLE  $t$  TEST

1

The previous unit described the technique of the one-tailed, one-sample  $t$  test. This unit will continue by explaining the two-tailed, one-sample  $t$  test. The only difference between the two is the selection of the critical values, due to the different alternative hypothesis. In the one-tailed test, the alternative hypotheses are  $\mu >$  the postulated mean or  $\mu <$  the postulated mean. As a result, we were only concerned with extreme values in one direction.

The two-tailed test, however, is concerned with extremes in both tails of the distribution, since its alternative hypothesis is  $\mu \neq$  the postulated mean. This unit will explain how to determine the critical values for the two-tailed, one-sample  $t$  test.

12a

Answer: reject

Our obtained  $t$  was larger than the tabled critical value.  
 $(t_{\text{OBTAINED}} > 2.145)$

13

Perform the appropriate  $t$  test for the following 5 situations.

$$t_{n-1} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

- I.  $H_0: \mu = 100$                        $\bar{x} = 105$   
 $H_1: \mu > 100$                        $s = 10$   
 $\alpha = .01$                        $n = 9$

What is critical value? # \_\_\_\_\_

Accept or reject  $H_0$ ? \_\_\_\_\_ Why? + \_\_\_\_\_

When presented with the following sets of hypotheses, what test do we use (one- or two-tailed, one-sample  $t$  test)?

A)  $H_0: \mu = \text{any value}$

$H_1: \mu > \text{that value}$  \* \_\_\_\_\_

B)  $H_0: \mu = \text{any value}$

$H_1: \mu \neq \text{that value}$  \* \_\_\_\_\_

C)  $H_0: \mu = \text{any value}$

$H_1: \mu < \text{that value}$  \* \_\_\_\_\_

Answer: one-tailed, one-sample  $t$  test

13a

$$t_{n-1} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$t = \frac{105 - 100}{10/\sqrt{9}}$$

$$= \frac{5}{10/3}$$

$$= \frac{5}{3.33}$$

$$= 1.5$$

Critical value = 2.896.

Accept  $H_0$  because obtained  $t < 2.896$

- Answer: A) one-tailed, one-sample  $t$  test  
 B) two-tailed, one-sample  $t$  test  
 C) one-tailed, one-sample  $t$  test
- 

3

In this unit we will concentrate on the  $t$  test for testing:

$H_0: \mu = \text{any number (any sample size)}$

against the alternative

$H_1: \mu \neq \text{that number}$

Using this  $H_1$  means we will be using the (two/one)-tailed  $t$  test.

---

II.  $H_0: \mu = 50$                        $\bar{x} = 60$   
 $H_1: \mu \neq 50$                        $s = 3$   
 $\alpha = .05$                                $n = 16$

14

What is critical value(s)? # \_\_\_\_\_

Accept or reject  $H_0$ ? \_\_\_\_\_ Why? + \_\_\_\_\_

---

Answer: two-tailed

3a

---

The formula for the two-tailed test is exactly the same as the one-tailed:

$$t_{n-1} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

The difference comes in the determination of the critical values.

---

4

Answer: two-tailed, one-sample  $t$  test

14a

$$\begin{aligned} t_{n-1} &= \frac{\bar{x} - \mu}{s/\sqrt{n}} \\ &= \frac{60 - 50}{3/\sqrt{16}} \\ &= \frac{10}{3/4} \\ &= \frac{10}{.75} \\ &= 13.3 \end{aligned}$$

Critical value = +2.131  
-2.131

Reject  $H_0$  because obtained  $t$  is  
larger than +2.131

Let's see how to use Table F in finding the critical values for a two-tailed test. Please turn now to Table F in the Guide. Across the top we see:

$t_{.100}$        $t_{.050}$        $t_{.025}$        $t_{.010}$        $t_{.005}$

In the table we are given the critical values for the point above which  $\alpha$  (in this case .05) of the values fall. When using a two-tailed test, we divide  $\alpha$  by 2 and look in the column headed by that  $t_{\alpha/2}$  value. For example, given  $\alpha = .05$  and  $n = 15$ , we use the column  $t_{.025}$  (since  $\alpha = .05$ ).

Down the left side is a column of degrees of freedom. In this case,  $d.f. = 15 - 1 = 14$ .

At the intersection of  $t_{.025}$  and  $d.f. = 14$ , we find the critical value # \_\_\_\_\_.

III.       $H_0: \mu = 75$                        $\bar{x} = 68$   
           $H_1: \mu \neq 75$                        $s = 10$   
           $\alpha = .01$                            $n = 25$

What is critical value? # \_\_\_\_\_

Accept or reject  $H_0$ ? \_\_\_\_\_ Why? + \_\_\_\_\_

Answer: 2.145

5a

6

We will accept  $H_0: \mu = 100$ , if our obtained  $t$  falls between # \_\_\_\_\_ and # \_\_\_\_\_.

Using the following format, write the critical values.

# \_\_\_\_\_ <  $t_{OBTAINED}$  < # \_\_\_\_\_

Answer:

two-tailed, one-sample  $t$  test

15a

$$\begin{aligned} t_{n-1} &= \frac{\bar{x} - \mu}{s/\sqrt{n}} \\ &= \frac{68 - 75}{10/\sqrt{25}} \\ &= \frac{-7}{10/5} \\ &= \frac{-7}{2} \\ &= -3.5 \end{aligned}$$

Critical value -  $\pm 2.797$

Reject  $H_0$  because obtained  $t$  is beyond the critical values.

Answer: -2.145

+2.145

$$-2.145 < t_{\text{OBTAINED}} < +2.145$$

7

Let's run through the statistical problem-solving procedure on the same data used in Unit 1, with a new alternative hypothesis.

We have a sample of 15 readings of TSP concentrations, and we want to see if the mean of these readings differs significantly at the 5% level from the assumed mean of 100  $\mu\text{g}/\text{m}^3$ . We already know:

$$\bar{x} = 110$$

$$s = 15.34$$

Step 1: Define the problem.

$$H_0: \mu = 100$$

$$H_1: \underline{\hspace{2cm}}$$

16

IV. We wish to determine if the mean CO concentration in a certain area is significantly different from the standard of 10  $\text{mg}/\text{m}^3$  at the 5% level of significance. We obtain a sample of 16 CO concentration readings over an 8-hour period with a mean of 12  $\text{mg}/\text{m}^3$  and standard deviation of 4  $\text{mg}/\text{m}^3$ . Using the six-step statistical problem-solving procedure, can we conclude that the mean CO concentration differs significantly from the standard.

Answer:  $H_1: \mu \neq 100$  (since it can differ in both directions)

7a

---

Answer: Step 1:  $H_0: \mu = 10$

16a

$H_1: \mu \neq 10$

Step 2: Two-tailed, one-sample  $t$  test  $\alpha = .05$

Step 3: Not necessary

Step 4:  $t_{n-1} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$

$$= \frac{12 - 10}{4/\sqrt{16}}$$

$$= \frac{2}{4/4}$$

$$= \frac{2}{1}$$

$$= 2.00$$

Step 5: Critical value =  $\pm 2.131$

Step 6: Accept  $H_0$ , since the obtained  $t$  falls between the critical values.

The sample mean does not differ significantly from the standard.

Step 2: Select test and set  $\alpha$  level.

test is \* \_\_\_\_\_

$\alpha = \#$  \_\_\_\_\_

---

17

- V. We wish to know if the mean TSP concentration in a suburban area is significantly lower than the postulated mean of  $150 \mu\text{g}/\text{m}^3$  at the 1% level of significance. We obtain a sample of 9 readings of TSP concentrations with a mean of  $135 \mu\text{g}/\text{m}^3$  and a standard deviation of  $15 \mu\text{g}/\text{m}^3$ . Using the six-step statistical problem-solving procedure, answer the question. (Be careful in your selection of the alternative hypothesis.)

Answer: two-tailed, one-sample  $t$  test

$$\alpha = .05$$

Step 3: Organize data

9

(necessary or not necessary?)

Answer: Step 1:  $H_0: \mu = 150$

17a

$$H_1: \mu < 150$$

Step 2: One-tailed, one-sample  $t$  test  $\alpha = .01$

Step 3: Not necessary

Step 4:

$$t_{n-1} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$= \frac{135 - 150}{15/\sqrt{9}}$$

$$= \frac{-15}{15/3}$$

$$= \frac{-15}{5}$$

$$= -3.00$$

Step 5: Critical value = -2.896

Step 6: The sample mean is significantly lower than the standard.

To review:

18

To test the null hypothesis that the population mean is equal to a certain value (against the alternative hypothesis that the mean does not equal that value) we use the \* \_\_\_\_\_.

Answer: not necessary in  $t$  test

9a

---

Step 4: Compute the statistic.

10

The two-tailed  $t$  is computed exactly as the one-tailed  $t$ .

Find  $t_{n-1}$

$$t_{n-1} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

---

Answer: two-tailed, one-sample  $t$  test

18a

---

This test is expressed:

19

$H_0$ : \* \_\_\_\_\_

$H_1$ : \* \_\_\_\_\_

10a

Answer: 
$$t_{n-1} = \frac{110 - 100}{15.34/\sqrt{15}}$$
$$= \frac{10}{15.34/3.87}$$
$$= \frac{10}{3.96}$$
$$t_{n-1} = 2.52$$

---

Step 5: Consult the appropriate table for critical values.

11

Critical values = # \_\_\_\_\_

(Caution: remember that this is where  
the difference lies)

---

Answer:  $H_0: \mu = \text{any number}$

19a

$H_1: \mu \neq \text{that number}$

Answer: Since  $\alpha = .05$  we look at  $t_{\alpha/2}$  column ( $t_{.025}$ ) and  $d.f. = 14$ .

Critical values =  $\pm 2.145$

12

Step 6: Make the appropriate decision.

Accept or reject  $H_0$ ? \_\_\_\_\_

Why? + \_\_\_\_\_

Go to:

Frame 12a

Page III-18

20

The test is performed through use of the formula:

$$t_{n-1} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

where  $t_{n-1}$  = the  $t$  statistic with  $n - 1$  degrees of freedom

$\bar{x}$  = the sample mean

$\mu$  = the postulated mean or standard

$s$  = the sample standard deviation

$n$  = the sample size

The result is then compared with the critical values found in Table F. In using this table with a two-tailed test, we must be careful to select the  $t_{\alpha}$  level that reflects an  $\alpha/2$  value. For  $\alpha = .05$  we need to use the column  $t_{.025}$  and for  $\alpha = .01$  we use  $t_{.005}$  when dealing with a two-tailed test.

Proceed to Unit 3.

UNIT 3 - TWO-SAMPLE  $t$  TEST

1

The preceding two units demonstrated the use of  $t$  tests to test the null hypothesis:

$$H_0: \mu = \text{any value}$$

against one of the following alternative hypotheses:

$$H_1: \mu \neq \text{that value}$$

$$H_1: \mu > \text{that value}$$

$$H_1: \mu < \text{that value}$$

36a

Answer:  $H_0: \mu_1 = \mu_2$

$$H_1: \mu_1 \neq \mu_2$$

37

Step 2: Select appropriate test statistic and  $\alpha$  level.

Test is # \_\_\_\_\_ -tailed, # \_\_\_\_\_ -sample test

$\alpha = \#$  \_\_\_\_\_

This unit will explain another use of the  $t$  test, one in which sample means are compared to decide if they are drawn from the same population. In this latter case, the null hypothesis takes the form of:

$$H_0: \mu_1 = \mu_2$$

Again, the alternative hypothesis,  $\mu_1 \neq \mu_2$ , requires a two-tailed test, and the other two,  $\mu_1 > \mu_2$  and  $\mu_1 < \mu_2$ , require one-tailed tests.

This unit will present both the two-tailed test and the one-tailed test.

---

Answer: two-tailed, two-sample  $t$  test

37a

 $\alpha = .01$

The two-tailed, two-sample  $t$  test is used to test:

$H_0$ : \_\_\_\_\_

$H_1$ : \_\_\_\_\_

---

Step 3: Organize and arrange the data.

38

\* \_\_\_\_\_

Answer:  $\mu_1 = \mu_2$

3a

$\mu_1 \neq \mu_2$

---

In testing  $H_0: \mu_1 = \mu_2$  against  $H_1: \mu_1 \neq \mu_2$ , there are three questions that determine the correct formula to use:

4

Question 1: Are the two samples independent or paired?

Question 2: If the samples are independent, are the sample sizes equal?

Question 3: If the sample sizes are not equal, are the variances homogeneous? (i.e., are the population variances equal?)

We will discuss each question individually and present the appropriate formulas in the following frames.

---

Answer: Not necessary

38a

Question 1: Are the two samples independent or paired?

The answer to this question is determined by whether or not each item in one sample is related to a corresponding item in the other sample.

An example of independent samples:

Two samples from two different stacks at different locations

If, however, we have two samples from the same stack under two different treatments whose effects are being studied, we would have paired samples.

Step 4: Compute the statistic.

Set up the table.

$\frac{x_1}{n_1}$	$\frac{x_1^2}{n_1}$	$\frac{x_2}{n_2}$	$\frac{x_2^2}{n_2}$
-------------------	---------------------	-------------------	---------------------

Find  $\bar{x}_1$  and  $\bar{x}_2$  and add the columns in the table.

$$\bar{x} = \frac{\sum x_i}{n}$$

$$t_{N-2} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\left[ \sum x_1^2 - \frac{(\sum x_1)^2}{n_1} \right] + \left[ \sum x_2^2 - \frac{(\sum x_2)^2}{n_2} \right]}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

Which of the following samples is independent and which is paired?

- A) Samples of  $\text{SO}_2$  concentrations for a specific time period at City 1 and City 2. (independent/paired)
- B) Samples of two identical stacks, each having a different scrubbing device, from the same industrial furnace taken at the same time. (independent/paired)

Answer:  $\bar{x}_1 = 100$   
 $\bar{x}_2 = 110$

39a

$x_1$	$x_1^2$	$x_2$	$x_2^2$
120	14400	130	16900
100	10000	110	12100
110	12100	120	14400
90	8100	100	10000
80	6400	90	8100
100	10000	110	12100
$\Sigma x_1 = 600$	$\Sigma x_1^2 = 61000$	$\Sigma x_2 = 660$	$\Sigma x_2^2 = 73600$

Answer: A) independent

B) paired

7

In answer to Question 1, "Are the samples independent or paired?", if the answer is "paired," the appropriate formula is Formula A. (We will get to the individual formulas after we've discussed all three questions.)

Formula A is used on \_\_\_\_\_ samples

40

Substitute the values in the formula:

$$\begin{aligned}
 t_{N-2} &= \frac{100 - 110}{\sqrt{\frac{\left[61,000 - \frac{(600)^2}{6}\right] + \left[73,600 - \frac{(660)^2}{6}\right]}{6 + 6 - 2}} \left(\frac{1}{6} + \frac{1}{6}\right)} \\
 &= \frac{-10}{\sqrt{\frac{1000 + 1000}{10}} \left(\frac{1}{3}\right)} \\
 &= \frac{-10}{\sqrt{66.67}} \\
 t_{N-2} &= -1.22
 \end{aligned}$$

-2  
Answer: paired

7a

---

8

If the answer to Question 1 is "independent," we need to ask Question 2.

Question 2: Are the sample sizes equal?

This is easy. If  $n_1 = 20$  and  $n_2 = 20$ , we have equal sample sizes. If we have equal sample sizes, we use Formula B.

Formula B is used when we have two \_\_\_\_\_ (independent/paired) samples of \_\_\_\_\_ size.

---

41

Step 5: Determine critical values.

$$\alpha = .01$$

Two-tailed test

$$d.f. = N-2 = 10$$

Critical values =  $t$  \_\_\_\_\_

Answer: independent

8a

equal

---

If we do not have equal sample sizes, we need to ask Question 3.

9

Question 3: Are the variances of the two samples homogeneous?

In order to decide whether or not the variances are homogeneous, we need to compute an  $F$  statistic.

---

Answer: Critical values =  $\pm 3.169$

41a

---

Step 6: Make the appropriate decision.

42

(Accept/Reject)  $H_0: \mu_1 = \mu_2$

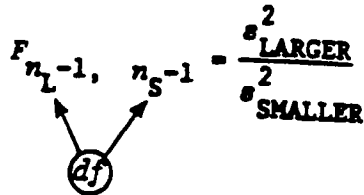
Why? + \_\_\_\_\_

The mean of sample 1 (is/is not) significantly different from sample 2 at the 1% level.

To compute  $F$  we find the ratio of the variances of the two samples. 10

A: Calculate the variance ( $s^2$ ) of the two samples. (You know how to calculate  $s$ ; calculating  $s^2$  is easier since you don't need to extract the square root. If you have trouble finding  $s^2$ , refer to page I-3 in the Guide.)

B: Place the larger variance over the smaller in the formula

$$F_{n_L-1, n_S-1} = \frac{s^2_{\text{LARGER}}}{s^2_{\text{SMALLER}}}$$


C: Turn to Table G. (Notice that Table G is really two tables; one for the 1% level and one for the 5% level.) Be sure to look at the table for the 5% level of significance. This table represents the 10% level for the homogeneity test (used by convention in this test). The tabled values of  $F$  are for the right tail of the  $F$  distribution. But we are now interested in whether  $s^2_{\text{LARGER}}$  and  $s^2_{\text{SMALLER}}$  differ significantly, and this is a two-tailed test. Therefore, we have 5% in each tail when using this table yielding a total significance level of 10%. The values listed across the top are the degrees of freedom of the numerator (the larger variance) and down the side are the degrees of freedom for the denominator (the smaller variance). Again degrees of freedom =  $n - 1$  for each sample. At the intersection is the critical value, which if exceeded means the variances are not homogeneous.

Answer: Accept.

Because obtained  $t$  falls between the critical values  
is not

42a

Let's try the homogeneity test.

Given two samples,  $n_1 = 16$        $n_2 = 25$

$$s_1^2 = 27 \quad s_2^2 = 12$$

A: The variance has already been calculated. (If you don't know how to calculate the variance re-read Module I, Unit 3. Remember,  $s$  is the standard deviation and  $s^2$  is the variance.)

B: Put the larger variance over the smaller to calculate  $F$ .

$$F_{n_L-1, n_S-1} = \frac{s_{\text{LARGER}}^2}{s_{\text{SMALLER}}^2}$$

$$F_{15,24} = \frac{27}{12}$$

$$= 2.25$$

C: Go to Table G. Read across the top to column 15 and down to 24. At the intersection we find 2.11.

Since 2.25 is larger than 2.11, we conclude that the variances are not homogeneous.

On occasion you may be given summary data instead of raw data. If you have the values of  $\bar{x}_1$ ,  $\bar{x}_2$ ,  $s_1$ , and  $s_2$ , then the formula changes slightly:

$$t_{N-2} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

(the original formula from now on will be called B-1 and this one B-2)

Try this one on your own.

$$n_1 = 13$$

$$n_2 = 21$$

$$s_1^2 = 4$$

$$s_2^2 = 11$$

A: The variance has already been calculated

$$F_{n_L-1, n_S-1} = \frac{s_{\text{LARGER}}^2}{s_{\text{SMALLER}}^2}$$

B:  $F_{\# \_, \# \_} = \frac{\# \_}{\# \_}$

C: Critical value =  $\# \_$

Are variances homogeneous?  $\_$

Let's try this one. Given the following independent sample data:

44

$$\bar{x}_1 = 100$$

$$\bar{x}_2 = 90$$

$$s_1 = 3$$

$$s_2 = 4$$

$$n_1 = 15$$

$$n_2 = 15$$

We want to know if the sample means differ significantly at the 5% level.

Insert values into the formula and compute the statistic.

$$t_{N-2} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Answer: Step 2:  $F_{20,12} = \frac{11}{4} = 2.75$

Step 3: Critical value = 2.54

No

Getting back to Question 3: "Are the variances homogeneous?"

If the answer is yes, we should use Formula B again; if the answer is no, we should use Formula C.

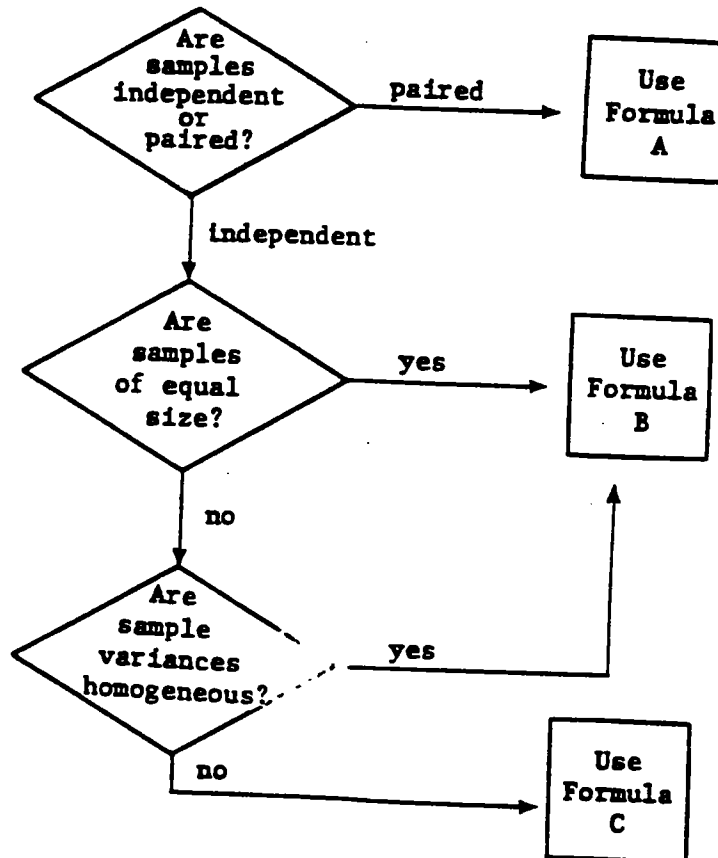
$$\begin{aligned}
 t_{N-2} &= \frac{100 - 90}{\sqrt{\frac{(15-1)(9) + (15-1)(16)}{15+15-2} \left(\frac{1}{15} + \frac{1}{15}\right)}} \\
 &= \frac{10}{\sqrt{\frac{126 + 224}{28} \left(\frac{2}{15}\right)}} \\
 &= \frac{10}{\sqrt{1.67}} \\
 &= \frac{10}{1.29} \\
 &= 7.75
 \end{aligned}$$

Go to:

Frame 45

Page III-45

So now we have a procedure. Ask these questions until you reach a formula:



Now that we know when to use Formulas A, B, C, let's find out what these formulas are and how to use them.

Formula A: [Used with (independent/paired) samples?]

$$t_{n'-1} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\Sigma d^2}{n'(n'-1)}}$$

$t_{n'-1}$  =  $t$  with  $n'-1$  degrees of freedom ( $n'$ , read "n prime"  
= no. of pairs of data).

$\bar{x}_1$  = mean of sample 1

$\bar{x}_2$  = mean of sample 2

$d$  = deviation (we will explain this concept shortly)

Next we will explain this formula and show you how to use it.

---

All other steps in the statistical problem-solving procedure remain the same. In this case, we would reject  $H_0$  since our obtained  $t$  (7.75) is larger than the table value (2.048).

By now you see how helpful it is to have summary data of this sort, since this formula is considerably easier to use than the previous one given.

In this formula:

$$t_{n'-1} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\Sigma d^2}{n'(n'-1)}}$$

you should know how to find all the terms except  $\Sigma d^2$  (the sum of the squared deviations). Here's how you go about finding  $\Sigma d^2$ :

A. Construct the following table:

<u>Values in Sample 1</u>	<u>Paired Value in Sample 2</u>	<u>(D) Difference</u>	<u>(D<sup>2</sup>) Difference<sup>2</sup></u>
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮

B.  $\Sigma d^2 = \Sigma D^2 - \frac{(\Sigma D)^2}{n'}$

C. Sum the  $D$  column and square the sum. Yields  $(\Sigma D)^2$

D. Sum the  $D^2$  column. Yields  $\Sigma D^2$ .

E. Put them in the above formula. The result is  $\Sigma d^2$ .

Formula C: Used with independent samples of unequal sizes and having "non-homogeneous" (heterogeneous) variances.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Go to:

Frame 47

Page III-48

Let's digress a moment and consider what a deviation is. Remember we have the following table:

<u>Values in Sample 1</u>	<u>Paired Value in Sample 2</u>	<u>(D) Difference</u>	<u>(D<sup>2</sup>) Difference<sup>2</sup></u>
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.

If we were to add the difference (D) column and divide by the number of differences, we would have a "mean difference" ( $\bar{D}$ ). Then, if we calculated the difference between each individual difference ( $D_i$ ) and this mean difference, we would have a set of deviations. Let's call each such deviation  $d$ . If we were then to square each deviation ( $d$ ) and add them together we would have  $\Sigma d^2$ . Symbolically this can be noted as:

$$\Sigma d^2 = \Sigma (D_i - \bar{D})^2$$

We use the formula:

$$\Sigma d^2 = \Sigma D^2 - \frac{(\Sigma D)^2}{n'}$$

instead because it's a shorter, easier method to determine "the sum of the squared deviations." (As in the standard deviation formulas, these formulas can be proven to be equal, but we are not going to discuss the proof.)

Let's try one:

A. Construct table.

<u>Test #</u>	<u>Sample 1</u>	<u>Sample 2</u>	<u>D</u>	<u>D<sup>2</sup></u>
1	15	17		
2	20	16		
3	22	18		
4	<u>23</u>	<u>17</u>		
	$\bar{x}_1 = 20$	$\bar{x}_2 = 17$		

Fill in D and D<sup>2</sup> columns.

In order to use this formula you need to find both the mean and variance for each sample, and then compute the statistic.

For example,

$\bar{x}_1$	$\bar{x}_2$
2	1
4	3
6	6
8	9
5	12
5	15
6	
7	

Find:  $\bar{x}_1$ ,  $\bar{x}_2$ ,  $s_1^2$ ,  $s_2^2$

Recall,

$$\bar{x} = \frac{\sum x_i}{n}$$

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n - 1}$$

Answer:

<u>Test #</u>	<u>Sample 1</u>	<u>Sample 2</u>	<u>D</u>	<u>D<sup>2</sup></u>
1	15	17	-2	4
2	20	16	4	16
3	22	18	4	16
4	<u>23</u>	<u>17</u>	<u>6</u>	<u>36</u>
	$\bar{x}_1 = 20$	$\bar{x}_2 = 17$		

Answer:

$$\bar{x}_1 = 5.38$$

$$\bar{x}_2 = 7.67$$

$$s_1^2 = 3.41$$

$$s_2^2 = 28.67$$

47a

Calculate F.

48

$$F_{n_L-1, n_S-1} = \frac{s_{\text{LARGER}}^2}{s_{\text{SMALLER}}^2}$$

Critical value for  $F_{n_L-1, n_S-1} = F_{5,7} = \#$  \_\_\_\_\_.

B. Sum both the  $D$  and  $D^2$  columns.

<u>Test #</u>	<u>Sample 1</u>	<u>Sample 2</u>	<u>D</u>	<u>D<sup>2</sup></u>
1	15	17	-2	4
2	20	16	4	16
3	22	18	4	16
4	<u>23</u>	<u>17</u>	<u>6</u>	<u>36</u>
	$\bar{x}_1 = 20$	$\bar{x}_2 = 17$		

$$\Sigma D = \# \underline{\hspace{2cm}}$$

$$\Sigma D^2 = \# \underline{\hspace{2cm}}$$

$$F = \frac{28.67}{3.41} = 8.41$$

48a

Critical value for  $F_{5,7} = 3.97$ . Since our obtained  $F$  is greater than the critical value we conclude the variances are not homogeneous.

Answer:  $\Sigma D = 12$

19a

$$\Sigma D^2 = 72$$

---

C. Square  $\Sigma D$ .

20

$$(\Sigma D)^2 = \# \underline{\hspace{2cm}}$$

---

49

Substitute into the formula

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$= \frac{5.38 - 7.67}{\sqrt{\frac{3.41}{8} + \frac{28.67}{6}}}$$

$$= \frac{-2.29}{\sqrt{5.21}}$$

$$= \frac{-2.29}{2.28}$$

$$= -1.00$$

Answer:  $(\Sigma D)^2 = 144$

D. Compute  $\Sigma d^2$

21

$$\Sigma d^2 = \Sigma D^2 - \frac{(\Sigma D)^2}{n'}$$

$$= 72 - \frac{144}{4}$$

$$= 72 - 36$$

$$= 36$$

$$\Sigma d^2 = \# \underline{\hspace{2cm}}$$

50

To determine the critical values for this test, you must do something a little different. You look up the critical values for  $t_{n_1-1}$  and  $t_{n_2-1}$ . If the obtained  $t$  value falls between these two, you have to interpolate using this formula:

$$\text{Critical Value} = \frac{t_{n_1-1} \left( \frac{s_1^2}{n_1} \right) + t_{n_2-1} \left( \frac{s_2^2}{n_2} \right)}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where  $t_{n_1-1}$  and  $t_{n_2-1}$  are the critical values for those degrees of freedom.

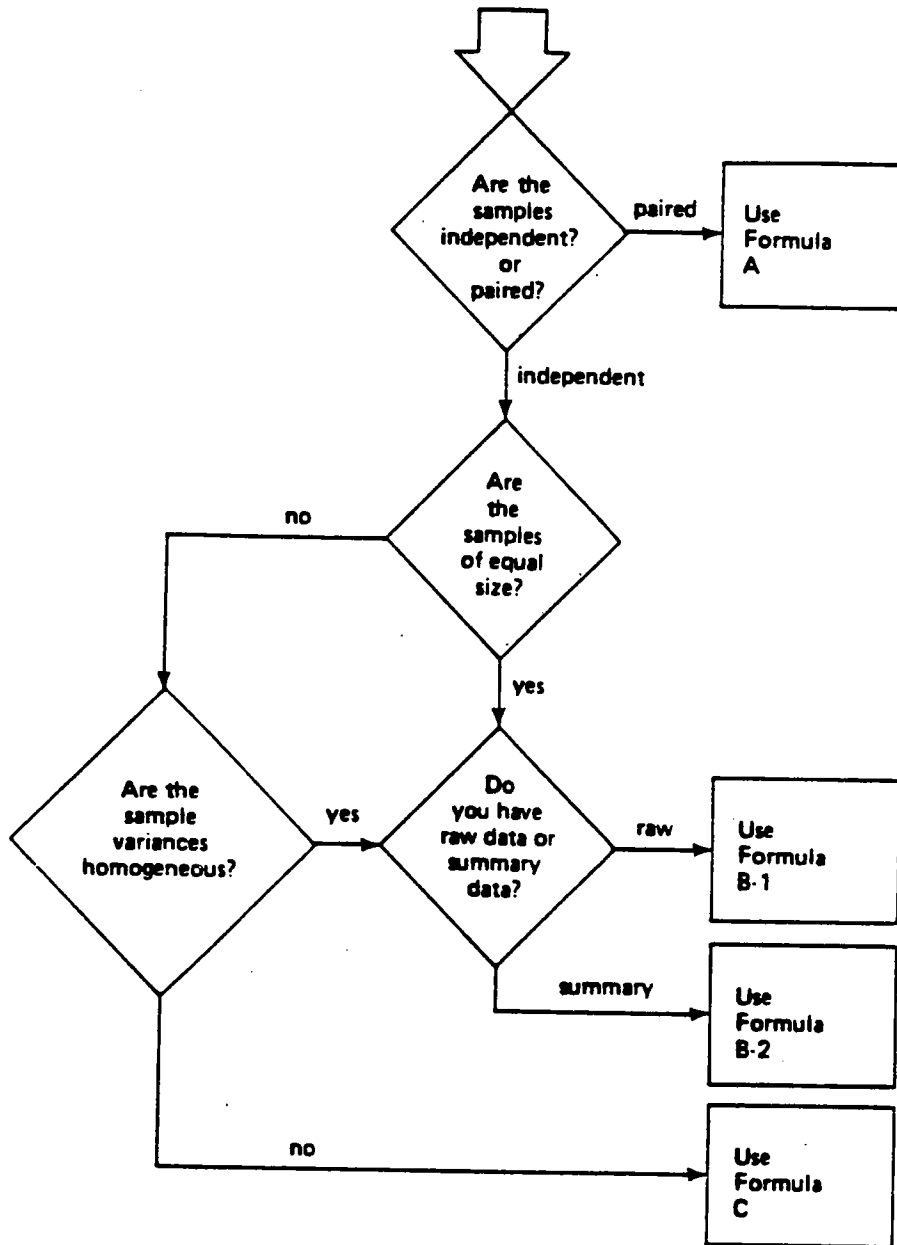
If the obtained  $t$  does not fall between the two tabled critical values, then you base your decision as you would before. For example, in the sample problem we obtained  $t = -1.00$ . The critical values =  $\pm 2.365$  and  $\pm 2.571$ . (Assume  $\alpha = .05$  and two-tailed test.)

Because our obtained  $t$  does not exceed either set of critical values, we accept  $H_0: \mu_1 = \mu_2$ . If our obtained  $t$  fell between  $-2.571$  and  $-2.365$  or between  $+2.571$  and  $+2.365$ , however, we would have had to interpolate to find the "true" critical value, and base our acceptance or rejection on that value.

Let's review the questions:

51

Remember that we now have a total of four questions!



E. Insert the values in the  $t_{n'-1}$  formula (Formula A) and compute the statistic.

$$\begin{aligned}
 t_{n'-1} &= \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\Sigma d^2}{n'(n'-1)}}} \\
 &= \frac{20 - 17}{\sqrt{\frac{36}{4(4-1)}}} \\
 t_{n'-1} &= 1.73
 \end{aligned}$$

Until now we've focused on the two-tailed, two-sample  $t$  test. Let's now consider the one-tailed test. The one-tailed test tests the null hypothesis

$$H_0: \mu_1 = \mu_2$$

against which of the following alternatives?

A.  $H_1: \mu_1 \neq \mu_2$

B.  $H_1: \mu_1 > \mu_2$

C.  $H_1: \mu_1 < \mu_2$

To find the critical values, go to Table F (t distribution) and find the value for  $n'-1$  degrees of freedom and the  $\alpha$  level you selected. For this example  $\alpha = .05$ . What are critical values based on

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Critical values =  $t$  \_\_\_\_\_

---

Answer: B.  $H_1: \mu_1 > \mu_2$

52a

C.  $H_1: \mu_1 < \mu_2$

Answer: Crit. values  $\pm 3.182$   
 (If you answered  $\pm 2.353$ , you forgot that this was a two-tailed test based on

$$H_1: \mu_1 \neq \mu_2$$

As a result, you use  $t_{\frac{.05}{2}}$  or  $t_{.025}$  in the Table).

Based on the results, do you accept  $H_0$ ? \_\_\_\_\_

Why?  $t$  \_\_\_\_\_  
 \_\_\_\_\_

The computation of the one-tailed  $t$  test is exactly the same as the two-tailed. The only difference lies in the determination of critical values. In the two-tailed test we used the column headed by  $t_{\alpha/2}$  ( $t_{.025}$  if  $\alpha = .05$ ,  $t_{.005}$  if  $\alpha = .01$ ). In the one-tailed test, we don't divide  $\alpha$  by 2. So, if  $\alpha = .05$ , we use the column  $t_{\#}$  and for  $\alpha = .01$  we use  $t_{\#}$ .

Answer: Yes. Accept  $H_0$  since the obtained  $t$  was between the critical values.

---

25

Formula B: Used on independent samples of equal sizes, or on independent samples of unequal sizes, provided the variances are homogeneous.

$$t_{N-2} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\left[ \sum x_1^2 - \frac{(\sum x_1)^2}{n_1} \right] + \left[ \sum x_2^2 - \frac{(\sum x_2)^2}{n_2} \right]}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Now let us show you how easy this formula is, despite its appearance.

---

53a

Answer:  $t_{.05}$

$t_{.01}$

To use this formula set up the following table:

$x_1$	$x_1^2$	$x_2$	$x_2^2$
·	·	·	·
·	·	·	·
·	·	·	·
·	·	·	·
·	·	·	·

When added, these columns equal

$\Sigma x_1$	$\Sigma x_1^2$	$\Sigma x_2$	$\Sigma x_2^2$
--------------	----------------	--------------	----------------

Then all you need to do is substitute in the formula.

Find the critical values for the following:

(Remember--

if  $H_1$  contains  $>$ , then we are interested in the positive critical value for  $t_\alpha$ .

if  $H_1$  contains  $<$ , then we are interested in the negative critical value for  $t_\alpha$ .

if  $H_1$  contains  $\neq$ , then we have a two-tailed test and we are interested in both the positive and negative values for  $t_{\alpha/2}$ .)

A.  $H_0: \mu_1 = \mu_2$

$H_1: \mu_1 > \mu_2$

$d.f. = 9$

$\alpha = .05$

Critical value = # \_\_\_\_\_

B.  $H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

$d.f. = 9$

$\alpha = .05$

Critical value = # \_\_\_\_\_

C.  $H_0: \mu_1 = \mu_2$

$H_1: \mu_1 < \mu_2$

$d.f. = 15$

$\alpha = .01$

Critical value = # \_\_\_\_\_

D.  $H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

$d.f. = 15$

$\alpha = .01$

Critical value = # \_\_\_\_\_

Let's try one.

$\bar{x}_1$	$\bar{x}_2$
52	62
50	60
48	58
50	60

A. Set up the table.

$\bar{x}_1$	$\bar{x}_1^2$	$\bar{x}_2$	$\bar{x}_2^2$
52	# _____	62	# _____
50	# _____	60	# _____
48	# _____	58	# _____
50	# _____	60	# _____

Answer:

A. Critical value = 1.833

B. Critical value =  $\pm 2.262$

C. Critical value = -2.602

D. Critical value =  $\pm 2.947$

54a

Answer:	$\underline{x_1}$	$\underline{x_1^2}$	$\underline{x_2}$	$\underline{x_2^2}$
	52	2704	62	3844
	50	2500	60	3600
	48	2304	58	3364
	50	2500	60	3600

---

B. Add the columns.

28

$$\Sigma x_1 = \# \underline{\hspace{2cm}}$$

$$\Sigma x_1^2 = \# \underline{\hspace{2cm}}$$

$$\Sigma x_2 = \# \underline{\hspace{2cm}}$$

$$\Sigma x_2^2 = \# \underline{\hspace{2cm}}$$


---

55

Next we want you to solve some problems and, at the same time, get practice in using the Guide to Statistical Problem Solving. Now open the Guide to section III B on "Two-sample t Tests" and use the six-step statistical problem-solving procedure to solve the problems that follow.

Answer:  $\Sigma x_1 = \underline{200}$

28a

$$\Sigma x_1^2 = \underline{10008}$$

$$\Sigma x_2 = \underline{240}$$

$$\Sigma x_2^2 = \underline{14408}$$

---

C. Square  $\Sigma x_1$  and  $\Sigma x_2$

29

$$(\Sigma x_1)^2 = \# \underline{\hspace{2cm}}$$

$$(\Sigma x_2)^2 = \# \underline{\hspace{2cm}}$$

---

Problem I: To test two different scrubbing devices, a manufacturer installs one in a stack at Plant #1 and the other in a stack at Plant #2. These stacks are from similar processes and have nearly identical characteristics. The manufacturer takes 15 readings at both stacks and obtains the following:

56

$$\underline{\text{Plant \#1}} \quad \bar{x} = 137$$

$$s = 18.6$$

$$\underline{\text{Plant \#2}} \quad \bar{x} = 104$$

$$s = 21.7$$

Can he conclude with 95% confidence that these results actually differ significantly?

Answer:  $(\Sigma x_1)^2 = \underline{40000}$

$(\Sigma x_2)^2 = \underline{57600}$

56a

Answer:

Step 1: Define the problem.

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Step 2: Select technique and set  $\alpha$  level.

Two-tailed  $t$  test  $\alpha = .05$

Step 3: Organize and arrange data.

Not necessary

Step 4: Compute the statistic.

We have independent samples of equal size, and are given summary data. We therefore use Formula B-2.

$$\begin{aligned}
 t_{N-2} &= \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \\
 &= \frac{137 - 104}{\sqrt{\frac{(14)(18.6)^2 + (14)(21.7)^2}{15 + 15 - 2} \left( \frac{1}{15} + \frac{1}{15} \right)}} \\
 &= \frac{33}{\sqrt{54.46}} \\
 &= \frac{33}{7.38} \\
 &= 4.47
 \end{aligned}$$

Step 5: Determine critical values.

Critical values =  $\pm 2.048$

Step 6: Make the appropriate decision.

Reject  $H_0$ , since the obtained value is greater than the critical value.

D. Substitute all values in the formula

$$t_{N-2} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\left[ \sum x_1^2 - \frac{(\sum x_1)^2}{n_1} \right] + \left[ \sum x_2^2 - \frac{(\sum x_2)^2}{n_2} \right]}{n_1 + n_2 - 2}} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$$t_{N-2} = \frac{50 - 60}{\sqrt{\frac{\left[ 10008 - \frac{40000}{4} \right] + \left[ 14408 - \frac{57600}{4} \right]}{4 + 4 - 2}} \left( \frac{1}{4} + \frac{1}{4} \right)}$$

$$t_{N-2} = -8.67$$

57

Problem II: The manufacturer still wasn't satisfied, and set up an experiment in which the scrubbers were installed in two identical stacks on the exact same furnace, and tests were made at the same time. This time the results looked like this:

<u>Measurement</u>	<u>Stack A</u>	<u>Stack B</u>
1	151	107
2	129	92
3	145	71
4	172	143
5	104	112
6	140	98

Again, at the 5% level, can he conclude that these values differ significantly?

To determine critical values ( $\alpha = .05$ , two-tailed), the degrees of freedom are  $N-2$ . ( $N$  indicates total of both sample sizes, i.e.,  $n_1 + n_2$ )

$$d.f. = \# \underline{\hspace{2cm}}$$

57a

Answer:

Step 1: Define the problem.

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Step 2: Select technique and set  $\alpha$  level.Two-tailed  $t$  test  $\alpha = .05$ 

Step 3: Organize and arrange data.

Not necessary.

Step 4: Compute the statistic.

This time we have paired samples, so we use Formula A.

$$t_{n'-1} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\Sigma d^2}{n'(n'-1)}}$$

<u>Measurement</u>	<u>Stack A</u>	<u>Stack B</u>	<u>D</u>	<u>D<sup>2</sup></u>
1	151	107	44	1936
2	129	92	37	1369
3	145	71	74	5476
4	172	143	29	841
5	104	112	-8	64
6	140	98	42	1764
	841	623	$\Sigma D = 218$	$\Sigma D^2 = 11450$

$$\bar{x}_1 = 140.16 \quad \bar{x}_2 = 103.83 \quad (\Sigma D)^2 = 47524$$

Answer:  $8 - 2 = 6$

32

In Table F, find the critical values for  $\alpha = .05$ , two-tailed test and  $d.f. = 6$ .

Critical values = # \_\_\_\_\_

57a continued

$$\Sigma d^2 = \Sigma D^2 - \frac{(\Sigma D)^2}{n'}$$

$$= 11450 - \frac{47524}{6}$$

$$= 3529.34$$

$$t_{n'-1} = \frac{140.16 - 103.83}{\sqrt{\frac{3529.34}{6(6-1)}}}$$

$$= \frac{36.33}{\sqrt{117.64}}$$

$$= \frac{36.33}{10.84}$$

$$= 3.35$$

Step 5: Determine critical values.

Critical values =  $\pm 2.571$

Step 6: Make the appropriate decision.

Reject  $H_0$ , since the obtained value is greater than the critical value.

Answer: Critical values =  $\pm 2.447$

32a

Go to:

Frame 33

Page III-68

---

58

Problem III: You want to compare two cities'  $\text{SO}_2$  concentrations, so you take 6 readings at each. You get these data.

<u>City A</u>	<u>City B</u>
20	7
8	12
7	5
29	5
5	9
22	8

Can you conclude that their mean  $\text{SO}_2$  concentrations differ significantly? (Use  $\alpha = .01$ )

Step 1: Define the problem.

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Step 2: Select technique and set  $\alpha$  level.

Two-sample  $t$  test  $\alpha = .01$

Step 3: Organize and arrange data.

Not necessary

Step 4: Compute the statistic.

Since we have independent samples of equal size and raw data are given, we use Formula B-1.

$$t_{N-2} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\left[ \sum x_1^2 - \frac{(\sum x_1)^2}{n_1} \right] + \left[ \sum x_2^2 - \frac{(\sum x_2)^2}{n_2} \right]}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$\frac{x_1}{22}$	$\frac{x_1^2}{484}$	$\frac{x_2}{8}$	$\frac{x_2^2}{64}$
20	400	7	49
8	64	12	144
7	49	5	25
29	841	5	25
5	25	9	81
<u>22</u>	<u>484</u>	<u>8</u>	<u>64</u>
$\sum x_1 = 91$	$\sum x_1^2 = 1863$	$\sum x_2 = 46$	$\sum x_2^2 = 388$

$$\bar{x}_1 = 15.17$$

$$\bar{x}_2 = 7.67$$

$$t_{N-2} = \frac{15.17 - 7.67}{\sqrt{\frac{\left[ (1863) - \frac{(8281)}{6} \right] + \left[ (388) - \frac{(2116)}{6} \right]}{6 + 6 - 2} \left( \frac{1}{6} + \frac{1}{6} \right)}}$$

(continued on page III-68)



Answer: reject  $H_0$  since  $-8.67$  is less than  $-2.447$

Our conclusion then is that the two sample means differ significantly at the 5% level.

Answer:

Step 1: Define the problem.

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Step 2: Select technique and set  $\alpha$  level.

Test is a two-tailed, two-sample  $t$ -test.  $\alpha = .01$

Step 3: Organize and arrange data.

Not necessary.

Step 4: Compute the statistic.

In order to determine the correct formula, ask the four questions.

- A) Are the samples independent or paired? Independent.
- B) Are the samples of equal size? No.
- C) Are the sample variances homogeneous? Yes.

$$F_{27,28} = \frac{(27.4)^2}{(23.7)^2}$$

$$= 1.34$$

Critical value = 1.89 (by interpolation)

D) Raw or summary data? Summary

So go to Formula B-2.

(continued on page III-70)

Next we will try this one, using the statistical problem-solving procedure:

Do the means from these two independent samples differ significantly at the 1% level?

$\bar{x}_1$	$\bar{x}_2$
120	130
100	110
110	120
90	100
80	90
<u>100</u>	<u>110</u>

We'll work this problem in the following frames.

$$\begin{aligned}
 t_{N-2} &= \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \\
 &= \frac{103 - 144}{\sqrt{\frac{(28)(23.7)^2 + (27)(27.4)^2}{29 + 28 - 2} \left( \frac{1}{29} + \frac{1}{28} \right)}} \\
 &= \frac{-41}{\sqrt{654.5 (.0702)}} \\
 &= \frac{-41}{\sqrt{45.95}} \\
 &= \frac{-41}{6.78} \\
 &= -6.05
 \end{aligned}$$

59a continued

Step 5: Determine critical values.

$$\text{Critical values} = \pm 2.576$$

Step 6: Make the appropriate decision.

Reject  $H_0$ , since the obtained value exceeds the tabled value. In addition, since the sample mean concentration in Pittsburgh was higher than that in New York, we conclude, with 99% confidence, that based on these sample data, Pittsburgh has the "dirtier" air.

Step 1: Define the problem.

$H_0$ : \_\_\_\_\_

$H_1$ : \_\_\_\_\_

Go to:

Frame 36a

Page III-31

---

60

In review--statisticians often have a need to determine if the means of two samples differ significantly from each other. In testing this difference, the statistician uses as his null hypothesis:

$$H_0: \mu_1 = \mu_2$$

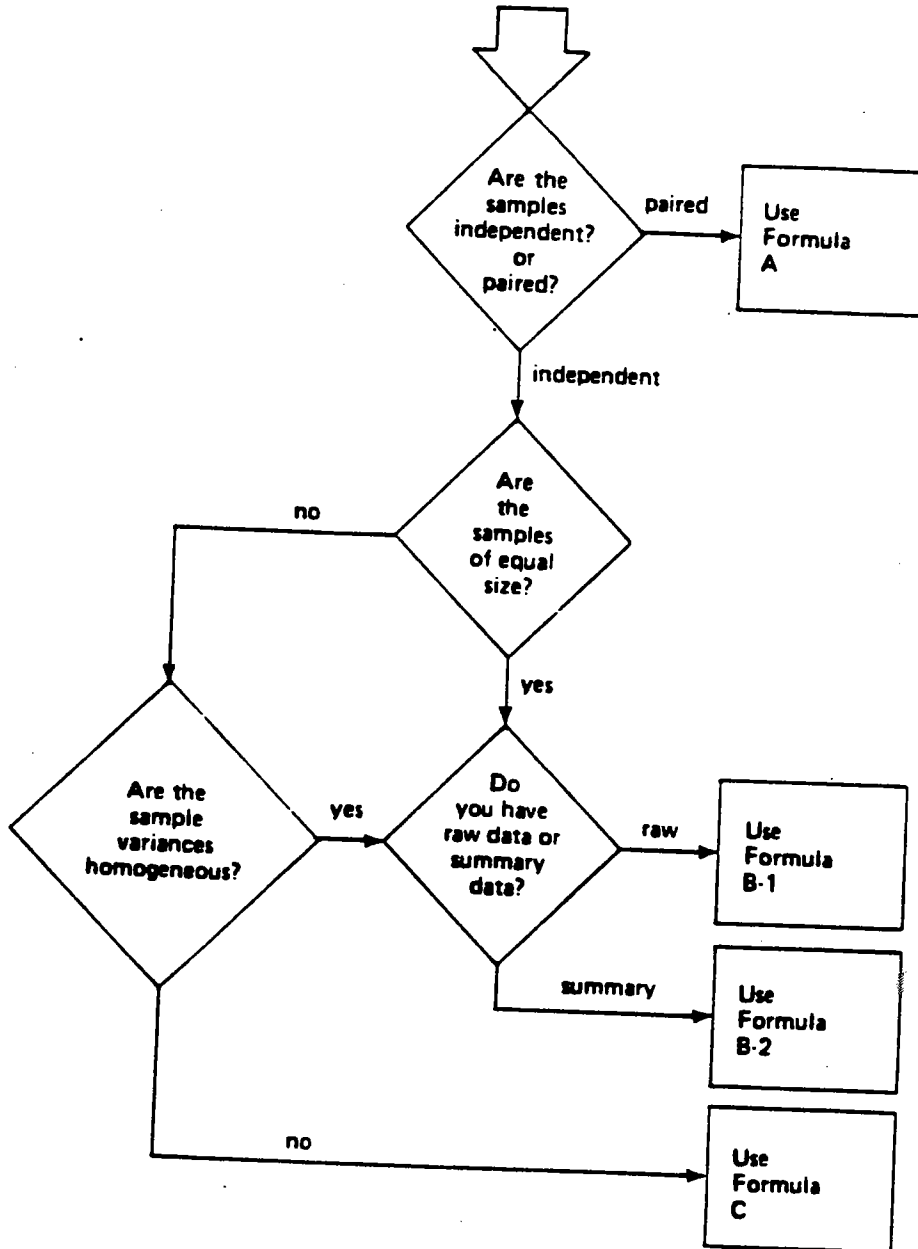
He then tests this hypothesis against one of the following hypotheses:

$$H_1: \mu_1 \neq \mu_2$$

$$H_1: \mu_1 > \mu_2$$

$$H_1: \mu_1 < \mu_2$$

Before running a *t* test the statistician needs to ask four questions about the data in order to determine the appropriate test formula. The sequence is:



Formula A: Used on paired samples.

$$t_{n'-1} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\Sigma d^2}{n'(n'-1)}}$$

Where  $t_{n'-1}$  = the  $t$  value with  $n'-1$  degrees of freedom  
( $n'$  = no. of pairs)

$\bar{x}_1$  = mean of sample 1

$\bar{x}_2$  = mean of sample 2

$\Sigma d^2$  = sum of squares of deviation scores found by  
the formula:

$$\Sigma d^2 = \Sigma D^2 - \frac{(\Sigma D)^2}{n'}$$

Where  $D$  = the difference between the two paired sample  
raw scores, use a table like this:

<u><math>x_1</math></u>	<u><math>x_2</math></u>	<u><math>D = (x_1 - x_2)</math></u>	<u><math>D^2</math></u>
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
		<u><math>\Sigma D</math></u>	<u><math>\Sigma D^2</math></u>

Formulas B-1 and B-2: Used on independent samples of equal sizes, or independent samples of unequal sizes provided that the variances of the samples are homogeneous.

B-1 Given Raw Data

$$t_{N-2} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\left[ \sum x_1^2 - \frac{(\sum x_1)^2}{n_1} \right] + \left[ \sum x_2^2 - \frac{(\sum x_2)^2}{n_2} \right]}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Where  $t_{N-2}$  = the  $t$  value with  $N-2$  degrees of freedom

$$(N = n_1 + n_2)$$

$\bar{x}_1$  = mean of sample 1

$\bar{x}_2$  = mean of sample 2

$n_1$  = size of sample 1

$n_2$  = size of sample 2

To determine  $\sum x_1^2$ ,  $(\sum x_1)^2$ ,  $\sum x_2^2$ ,  $(\sum x_2)^2$ , use the following table:

$\underline{x_1}$	$\underline{x_1^2}$	$\underline{x_2}$	$\underline{x_2^2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\underline{\quad}$	$\underline{\quad}$	$\underline{\quad}$	$\underline{\quad}$
$\sum x_1$	$\sum x_1^2$	$\sum x_2$	$\sum x_2^2$

$$(\sum x_1) (\sum x_1) = (\sum x_1)^2$$

$$(\sum x_2) (\sum x_2) = (\sum x_2)^2$$

$$t_{N-2} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Where  $t_{N-2}$  = the  $t$  value with  $N-2$  degrees of freedom

$\bar{x}_1$  = mean of sample 1

$\bar{x}_2$  = mean of sample 2

$n_1$  = size of sample 1

$n_2$  = size of sample 2

$s_1^2$  = variance of sample 1

$s_2^2$  = variance of sample 2

If  $n_1 \neq n_2$ , then a test of homogeneity is necessary.

$$F_{n_L-1, n_S-1} = \frac{s_{\text{LARGER}}^2}{s_{\text{SMALLER}}^2}$$

$s_{\text{LARGER}}^2$  = the larger variance of the two samples.

$s_{\text{SMALLER}}^2$  = the smaller variance of the two samples.

$F_{n_L-1, n_S-1}$  =  $F$  value with  $n_L-1$  and  $n_S-1$  degrees of freedom.

If obtained  $F$  exceeds the critical value, then variances are not homogeneous and Formula C should be used.

If obtained  $F$  does not exceed the critical value, then variances are homogeneous and Formula B should be used.

Formula C: Used with independent samples of unequal size and having heterogeneous variances.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Where  $t = t$  value

$\bar{x}_1$  = mean of sample 1

$\bar{x}_2$  = mean of sample 2

$s_1^2$  = variance of sample 1

$s_2^2$  = variance of sample 2

$n_1$  = size of sample 1

$n_2$  = size of sample 2

If obtained  $t$  value does not fall between the critical values in the table for  $t_{n_1-1}$  and  $t_{n_2-1}$  degrees of freedom, you base your decision to accept or reject on the basis of whether BOTH critical values are exceeded or not.

If, however, the obtained  $t$  does fall between the critical values, then a "true" critical value needs to be interpolated, using this formula:

$$\text{"True" critical value} = \frac{t_{n_1-1} \left( \frac{s_1^2}{n_1} \right) + t_{n_2-1} \left( \frac{s_2^2}{n_2} \right)}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where  $t_{n_1-1}$  = the critical value for  $t_{n_1-1}$  degrees of freedom

$t_{n_2-1}$  = the critical value for  $t_{n_2-1}$  degrees of freedom

Proceed to Module IV.

