9.3 Tests about a Population Mean

WHAT YOU WILL LEARN

By the end of the section, you should be able to:

- State and check the Random, 10%, and Normal/Large Sample conditions for performing a significance test about a population mean.
- Perform a significance test about a population mean.
- Use a confidence interval to draw a conclusion for a two-sided test about a population parameter.
- Perform a significance test about a mean difference using paired data.

Confidence intervals and significance tests for a population proportion p are based on z-values from the standard Normal distribution. Inference about a population mean μ uses a t distribution with $n - 1$ degrees of freedom, except in the rare case when the population standard deviation σ is known. We learned how to construct confidence intervals for a population mean in Section 8.3. Now we'll examine the details of testing a claim about a population mean μ.

Zap tax

9.3.1 Carrying Out a Significance Test for μ

In an earlier example, a company claimed to have developed a new AAA battery that lasts longer than its regular AAA batteries. Based on years of experience, the company knows that its regular AAA batteries last for 30 hours of continuous use, on average. An SRS of 15 new batteries lasted an average of 33.9 hours with a standard deviation of 9.8 hours. Do these data give convincing evidence that the new batteries last longer on average? To find out, we perform a test of

$$H_0: \mu = 30 \text{ hours}$$
$$H_a: \mu > 30 \text{ hours}$$

where μ is the true mean lifetime of the new deluxe AAA batteries.

Conditions In Chapter 8, we introduced conditions that should be met before we construct a confidence interval for a population mean: Random, 10% when sampling without replacement, and Normal/Large Sample. These same three conditions must be verified before performing a significance test about a population mean.

As in the previous chapter, the Normal/Large Sample condition for means is

population distribution is Normal or sample size is large ($n \geq 30$)

We often don't know whether the population distribution is Normal. But if the sample size is large ($n \geq 30$), we can safely carry out a significance test. If the sample size is small, we should examine the sample data for any obvious departures from Normality, such as strong skewness and outliers.

CONDITIONS FOR PERFORMING A SIGNIFICANCE TEST ABOUT A MEAN

- **Random:** The data come from a well-designed random sample or randomized experiment.
  - **10%:** When sampling without replacement, check that $n \leq \frac{1}{10}N$.
- **Normal/Large Sample:** The population has a Normal distribution or the sample size is large ($n \geq 30$). If the population distribution has unknown shape and $n < 30$, use a graph of the sample data to assess the Normality of the population. Do not use t procedures if the graph shows strong skewness or outliers.

Here's an example that shows how to check the conditions.

Example 14 Better Batteries

Checking conditions

Figure 9.11 on the next page shows a dotplot, boxplot, and Normal probability plot of the battery lifetimes for an SRS of 15 batteries.

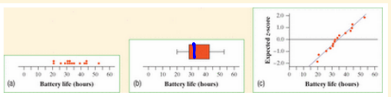


Figure 9.11 (a) A dotplot, (b) a boxplot, and (c) a Normal probability plot of the lifetimes of a simple random sample of 15 AAA batteries.

PROBLEM: Check the conditions for carrying out a significance test of the company's claim about its deluxe AAA batteries.

Random ✓ SRS n=15
10% ✓ produce > 150 total batteries
Large Sample / normal
Dotplot & Boxplot show slight right skewness & no apparent outliers
Normal prob. plot relatively linear
Safe to assume normality
(only need 1 graph, must sketch)

Calculations: Test Statistic and *P*-value When performing a significance test, we do calculations assuming that the null hypothesis H_0 is true. The test statistic measures how far the sample result diverges from the parameter value specified by H_0, in standardized units. As before,

$$\text{test statistic} = \frac{\text{statistic} - \text{parameter}}{\text{standard deviation of statistic}}$$

For a test of $H_0 : \mu = \mu_0$, our statistic is the sample mean $\bar{x}$. Its standard deviation is

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

In an ideal world, our test statistic would be

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Because the population standard deviation σ is usually unknown, we use the sample standard deviation s_x in its place. The resulting test statistic has the standard error of $\bar{x}$ in the denominator

$$t = \frac{\bar{x} - \mu_0}{s_x/\sqrt{n}}$$

As we saw earlier, when the Normal/Large Sample condition is met, this statistic has approximately a t distribution with $n - 1$ degrees of freedom.

In Section 8.3, we used Table B to find critical values from the t distributions when constructing confidence intervals about an unknown population mean μ. Once we have calculated the test statistic, we can use Table B to find the P-value for a significance test about μ. The following example shows how this works.

Example 15 Better Batteries

Computing the test statistic and P-value

The battery company wants to test H_0: $\mu = 30$ versus H_a: $\mu > 30$ based on an SRS of 15 new AAA batteries with mean lifetime $\bar{x} = 33.9$ hours and standard deviation $s_x = 9.8$ hours. The test statistic is

$$\text{test statistic} = \frac{\text{statistic} - \text{parameter}}{\text{standard deviation of statistic}}$$

$$t = \frac{\bar{x} - \mu_0}{s_x/\sqrt{n}} = \frac{33.9 - 30}{9.8/\sqrt{15}} = 1.54$$

The P-value is the probability of getting a result this large or larger in the direction indicated by H_a, that is, $P(t \geq 1.54)$. Figure 9.12 shows this probability as an area under the t distribution curve with df = 15 − 1 = 14. We can find this P-value using Table B.

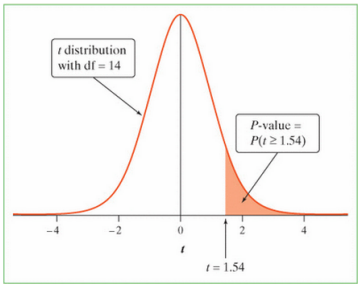


Figure 9.12 The P-value for a one-sided test with $t = 1.54$.

Go to the df = 14 row. The t statistic falls between the values 1.345 and 1.761. If you look at the top of the corresponding columns in Table B, you'll find that the "Upper-tail probability p" is between 0.10 and 0.05. (See the excerpt from Table B below.) Because we are looking for $P(t \geq 1.54)$, this is the probability we seek. That is, the P-value for this test is between 0.05 and 0.10.

goes w/ chart on next slide

Upper-tail probability p			
df	.10	.05	.025
13	1.350	1.771	2.160
14	1.345	1.761	2.145
15	1.341	1.753	2.131
	80%	90%	95%
	Confidence level C		

$.05 \leq \text{pval} \leq .10$

As you can see, Table B gives an interval of possible P-values for a significance test. We can still draw a conclusion from the test in much the same way as if we had a single probability. In the case of the new AAA batteries, for instance, we would fail to reject H_0: $\mu = 30$ because the P-value exceeds our default $\alpha = 0.05$ significance level. We don't have convincing evidence that the company's new AAA batteries last longer than 30 hours, on average.

Table B has other limitations for finding P-values. It includes probabilities only for t distributions with degrees of freedom from 1 to 30 and then skips to df = 40, 50, 60, 80, 100, and 1000. (The bottom row gives probabilities for df = ∞, which corresponds to the standard Normal distribution.) Also, Table B shows probabilities only for positive values of t. To find a P-value for a negative value of t, we use the symmetry of the t distributions. The next example shows how we deal with both of these issues.

Example 16 Two-Sided Tests, Negative *t*-Values, and More

Using Table B wisely

What if you were performing a test of H_0: $\mu = 5$ versus H_a: $\mu \neq 5$ based on a sample size of $n = 37$ and obtained $t = -3.17$? Because this is a two-sided test, you are interested in the probability of getting a value of t less than or equal to -3.17 or greater than or equal to 3.17. **Figure 9.13** shows the desired P-value as an area under the t distribution curve with 36 degrees of freedom. Notice that $P(t \leq -3.17) = P(t \geq 3.17)$ due to the symmetric shape of the density curve. **Table B** shows only positive t-values, so we will focus on $t = 3.17$.

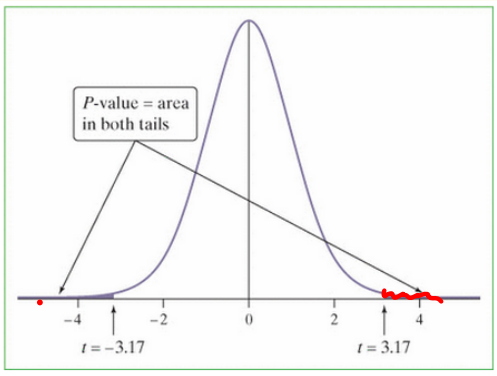


Figure 9.13 The P-value for a two-sided test with $t = -3.17$.

df=36 round down →

Upper-tail probability p			
df	.005	.0025	.001
29	2.756	3.038	3.396
30	2.750	3.030	3.385
40	2.704	2.971	3.307
	99%	99.5%	99.8%
Confidence level C			

← only positive t's

.001 ≤ one tailed prob ≤ .0025

.002 ≤ pval ≤ .005

Because df = 37 − 1 = 36 is not available on the table, use df = 30. You might be tempted to use df = 40, but doing so would result in a smaller P-value than you are entitled to with df = 36. (In other words, you'd be cheating!) Move across the df = 30 row, and notice that $t = 3.17$ falls between 3.030 and 3.385. The corresponding "Upper-tail probability p" is between 0.001 and 0.0025. (See the excerpt from **Table B**.) For this two-sided test, the corresponding P-value would be between 2(0.001) = 0.002 and 2(0.0025) = 0.005.

One point from the example deserves repeating: *if the df you need isn't provided in **Table B**, use the next lower df that is available.* It's no fair "rounding up" to a larger df. This is like pretending that your sample size is larger than it really is. Doing so would give you a smaller P-value than is true and would make you more likely to incorrectly reject H_0 when it's true (make a Type I error).

Given the limitations of **Table B**, our advice is to use technology to find P-values when carrying out a significance test about a population mean.

19. TECHNOLOGY CORNER COMPUTING P-VALUES FROM *t* DISTRIBUTIONS ON THE CALCULATOR

You can use the tcdf command on the TI-83/84 and TI-89 to calculate areas under a t distribution curve. The syntax is tcdf(lower bound,upper bound,df).

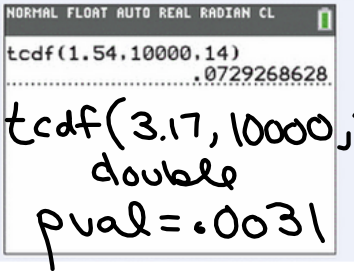
Let's use the tcdf command to compute the P-values from the last two examples.

Better batteries: To find $P(t \geq 1.54)$,

TI-83/84

- Press 2nd VARS (DISTR) and choose tcdf(.

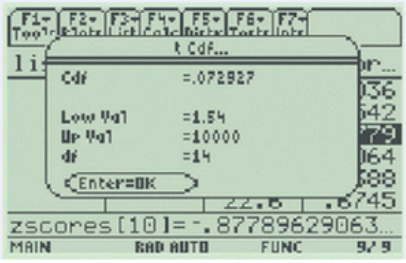
OS 2.55 or later: In the dialog box, enter these values: lower:1.54, upper: 10000, df:14, choose Paste, and then press ENTER. **Older OS:** Complete the command tcdf(1.54,10000,14) and press ENTER.



tcdf(3.17, 10000, 36)
double
pval = .0031

TI-89

- In the Stats/List Editor, press F5 (Distr) and choose t Cdf . . . .
- In the dialog box, enter these values: Lower value:1.54, Upper value:10000, Deg of Freedom, df:14, and then choose ENTER.



CHECK YOUR UNDERSTANDING

The makers of Aspro brand aspirin want to be sure that their tablets contain the right amount of active ingredient (acetylsalicylic acid). So they inspect a random sample of 36 tablets from a batch in production. When the production process is working properly, Aspro tablets have an average of μ = 320 milligrams (mg) of active ingredient. The amount of active ingredient in the 36 selected tablets has mean 319 mg and standard deviation 3 mg.

1. State appropriate hypotheses for a significance test in this setting.

μ = true mean amount of active ingredient in this batch of Aspro tablets

$H_0: \mu = 320$ $H_a: \mu \neq 320$

2. Check that the conditions are met for carrying out the test.

Random ✓ SRS $n = 36$ 10% ✓ more than 360 in batch

Large Counts ✓ $n \geq 30$

3. Calculate the test statistic. Show your work.

$$t = \frac{319 - 320}{3/\sqrt{36}} = -2$$

4. Use Table B to find the *P*-value. Then use technology to get a more accurate result. What conclusion would you draw?

two tailed, double area

tcdf(lower, upper, df) tcdf(-10000, -2, 35) = .0267

pval = 2(.0267) = .0534 $\alpha = .05$

pval > α, Fail to reject H_0

No convincing evidence that the average mg of active ingred. in this batch is different than 320.

(17) $n \geq \left(\frac{z^* \cdot \sigma}{ME}\right)^2$

from CI $z^* \frac{\sigma}{\sqrt{n}} \leq ME$

$n \geq \left(\frac{1.96(7200)}{200}\right)^2$

$z^* = 1.96$ for 95%

$n \geq 4978.71 \longrightarrow C$

(16) D

(15) E $\uparrow \alpha$ & $\uparrow n \Rightarrow \uparrow$ power

9.3.2 The One-Sample *t* Test

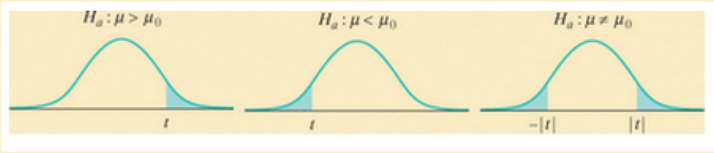
When the conditions are met, we can test a claim about a population mean μ using a **one-sample *t* test for a mean**. Here are the details.

ONE-SAMPLE *t* TEST FOR A MEAN

Suppose the conditions are met. To test the hypothesis H_0: $\mu = \mu_0$ compute the one-sample t statistic

$$t = \frac{\bar{x} - \mu_0}{s_x/\sqrt{n}}$$

Find the P-value by calculating the probability of getting a t statistic this large or larger in the direction specified by the alternative hypothesis H_a in a t distribution with df = $n - 1$:



Now we are ready to test a claim about an unknown population mean. Once again, we follow the four-step process.

Now we are ready to test a claim about an unknown population mean. Once again, we follow the four-step process.

Example 17 Healthy Streams

Performing a significance test about μ

The level of dissolved oxygen (DO) in a stream or river is an important indicator of the water's ability to support aquatic life. A researcher measures the DO level at 15 randomly chosen locations along a stream. Here are the results in milligrams per liter (mg/l):

4.53	5.04	3.29	5.23	4.13	5.50	4.83	4.40
5.42	6.38	4.01	4.66	2.87	5.73	5.55	

A dissolved oxygen level below 5 mg/l puts aquatic life at risk.

PROBLEM:

(a) Do we have convincing evidence at the $\alpha = 0.05$ significance level that aquatic life in this stream is at risk?

(b) Given your conclusion in part (a), which kind of mistake—a Type I error or a Type II error—could you have made? Explain what this mistake would mean in context.

SOLUTION:

(a) We will follow the four-step process.

STATE: We want to test a claim about the true mean dissolved oxygen level μ in this stream at the $\alpha = 0.05$ level. Our hypotheses are

$$H_0: \mu = 5$$
$$H_a: \mu < 5$$

PLAN: If the conditions are met, we should do a one-sample t test for μ.

- *Random:* The researcher measured the DO level at 15 randomly chosen locations.
  - 10%: There is an infinite number of possible locations along the stream, so it isn't necessary to check the 10% condition.
- *Normal/Large Sample:* We don't know whether the population distribution of DO levels at all points along the stream is Normal. With such a small sample size (n = 15), we need to graph the data to see if it's safe to use t procedures. **Figure 9.14** shows our hand sketches of a calculator histogram, boxplot, and Normal probability plot for these data. The histogram looks roughly symmetric; the boxplot shows no outliers; and the Normal probability plot is fairly linear. With no outliers or strong skewness, the t procedures should be pretty accurate.

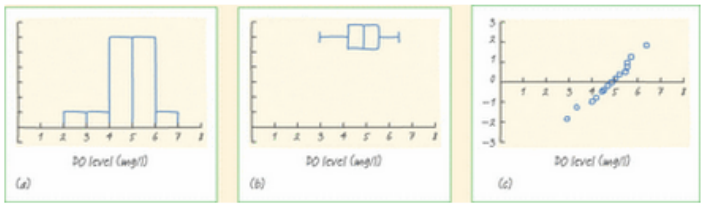
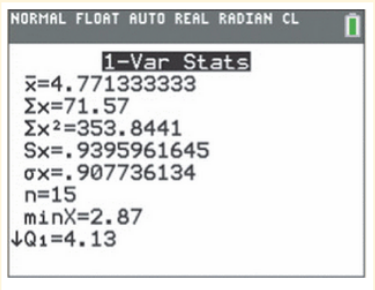


Figure 9.14 Sketches of (a) a histogram, (b) a boxplot, and (c) a Normal probability plot for the dissolved oxygen (DO) readings in the sample, in mg/l.

AP® EXAM TIP It is not enough just to make a graph of the data on your calculator when assessing Normality. You must *sketch* the graph on your paper to receive credit. You don't have to draw multiple graphs—any appropriate graph will do.

DO: We entered the data into our calculator and did 1-Var Stats (see screen shot). The sample mean is $\bar{x}$ = 4.771 and the sample standard deviation is s_x = 0.9396.



- *Test statistic*

$$t = \frac{\bar{x} - \mu_0}{s_x/\sqrt{n}} = \frac{4.771 - 5}{0.9396/\sqrt{15}} = -0.94$$

- *P*-value The *P*-value is the area to the left of $t = -0.94$ under the t distribution curve with degrees of freedom $df = 15 - 1 = 14$. **Figure 9.15** shows this probability. *Using the table:* **Table B** shows only areas in the upper tail of the distribution. Because the t distributions are symmetric, $P(t \le -0.94) = P(t \ge 0.94)$. Search the df = 14 row of **Table B** for entries that bracket t = 0.94 (see the excerpt below). Because the observed t lies between 0.868 and 1.076, the *P*-value lies between 0.15 and 0.20.

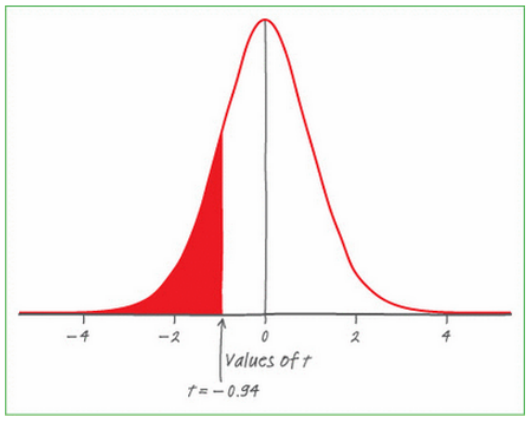


Figure 9.15 The *P*-value for a one-sided test when $t = -0.94$.

Upper-tail probability p			
df	.25	.20	.15
13	.694	.870	1.079
14	.692	.868	1.076
15	.691	.866	1.074
	50%	60%	70%
Confidence level C			

pval =.1816

Using technology: We can find the exact *P*-value using a calculator: tcdf(lower: −100, upper: −0.94, df: 14) = 0.1816.

CONCLUDE: Because the *P*-value, 0.1816, is greater than our $\alpha = 0.05$ significance level, we fail to reject H_0. We don't have convincing evidence that the mean DO level in the stream is less than 5 mg/l.

(b) Because we decided not to reject H_0 in part (a), we could have made a Type II error (failing to reject H_0 when H_0 is false). If we did, then the mean dissolved oxygen level μ in the stream is actually less than 5 mg/l, but we didn't find convincing evidence of that with our significance test.

Because the *t* procedures are so common, all statistical software packages will do the calculations for you. **Figure 9.16** shows the output from Minitab for the one-sample *t* test in the previous example. Note that the results match!

Session

One-Sample T: DO (mg/L)

Test of mu = 5 vs < 5

Variable	N	Mean	StDev	SE Mean	T	P
DO (mg/L)	15	4.771	0.940	0.243	-0.94	0.181

Figure 9.16 Minitab output for the one-sample *t* test from the dissolved oxygen example.

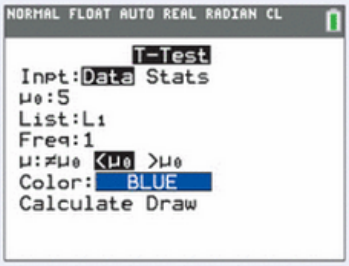
You can also use your calculator to carry out a one-sample *t* test. But be sure to read the AP® exam tip at the end of the Technology Corner.

20. TECHNOLOGY CORNER ONE-SAMPLE *t* TEST FOR A MEAN ON THE CALCULATOR

You can perform a one-sample *t* test using either raw data or summary statistics on the TI-83/84 or TI-89. Let's use the calculator to carry out the test of H_0: $\mu = 5$ versus H_a: $\mu < 5$ from the dissolved oxygen example. Start by entering the sample data in L1/list1. Then, to do the test:

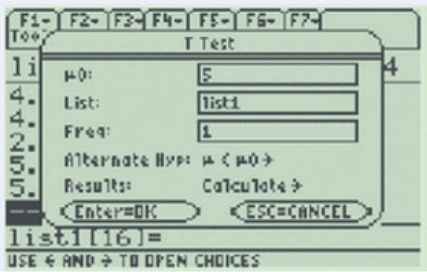
TI-83/84

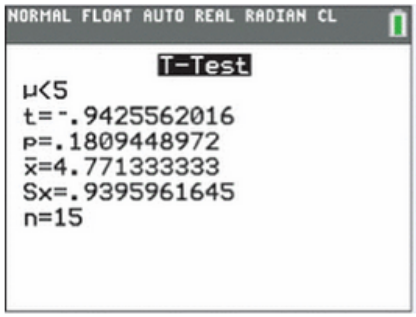
- Press STAT, choose TESTS and T-Test.
- Adjust your settings as shown.

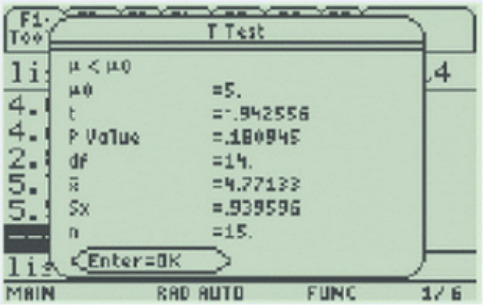


TI-89

- Press 2nd F1 ([F6]) and choose T-Test.
- Adjust your settings as shown.

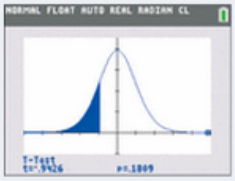


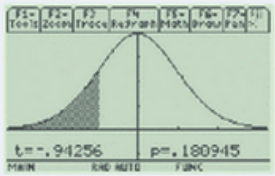




The test statistic is $t = -0.94$ and the P-value is 0.1809.

If you specify "Draw," you see a t distribution curve (df = 14) with the lower tail shaded.





AP® EXAM TIP Remember: if you just give calculator results with no work, and one or more values are wrong, you probably won't get any credit for the "Do" step. If you opt for the calculator-only method, name the procedure (t test) and report the test statistic ($t = -0.94$), degrees of freedom (df = 14), and P-value (0.1809).

Note: If you are given summary statistics instead of the original data, you would select the option "Stats" instead of "Data" in the first screen.

CHECK YOUR UNDERSTANDING

A college professor suspects that students at his school are getting less than 8 hours of sleep a night, on average. To test his belief, the professor asks a random sample of 28 students, "How much sleep did you get last night?" Here are the data (in hours):

9 6 8 6 8 8 6 6.5 6 7 9 4 3 4 5 6 11 6 3 6 6 10 7 8 4.5 9 7 7

1. Do these data provide convincing evidence at the $\alpha = 0.05$ significance level in support of the professor's suspicion?

State: μ = true mean amount of sleep that students at this school get in a night

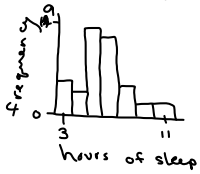
$H_0: \mu = 8$ $H_a: \mu < 8$

Plan: 1 Sample t test for μ

Random ✓ SRS $n = 28$

10% ✓ >280 students at this college

Normal/Large



Not much skewness
no apparent outliers
Safe to assume normality

Do: $\bar{x} = 6.6$ $S_x = 1.98$ $df = 27$

$t = -3.625$ $pval = 5.9 \times 10^{-4}$

$\approx$ almost 0

Conclusion:
We reject H_0 (pvalue $< \alpha = .05$)
Evidence supports that students at this school, on average, get less than 8 hrs of sleep a night.

9.3.3 Two-Sided Tests and Confidence Intervals

Now let's look at an example involving a two-sided test.

Example 18 Juicy Pineapples

A two-sided test

At the Hawaii Pineapple Company, managers are interested in the sizes of the pineapples grown in the company's fields. Last year, the mean weight of the pineapples harvested from one large field was 31 ounces. A different irrigation system was installed in this field after the growing season. Managers wonder how this change will affect the mean weight of future pineapples grown in the field. To find out, they select and weigh a random sample of 50 pineapples from this year's crop. The Minitab output below summarizes the data.

Descriptive Statistics: Weight (oz)

Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Weight (oz)	50	31.935	0.339	2.394	26.491	29.990	31.739	34.115	35.547

PROBLEM:

(a) Do these data give convincing evidence that the mean weight of pineapples produced in the field has changed this year?

(b) Can we conclude that the different irrigation system caused a change in the mean weight of pineapples produced? Explain your answer.

SOLUTION:

(a) STATE: We want to perform a test of

$$H_0: \mu = 31$$
$$H_a: \mu \neq 31$$

where μ = the mean weight (in ounces) of all pineapples grown in the field this year. Because no significance level is given, we'll use $\alpha = 0.05$.

PLAN: If the conditions are met, we should conduct a one-sample t test for μ.

- *Random:* The data came from a random sample of 50 pineapples from this year's crop.
  - 10%: There need to be at least 10(50) = 500 pineapples in the field because managers are sampling without replacement. We would expect many more than 500 pineapples in a "large field."
- *Normal/Large Sample:* We don't know whether the population distribution of pineapple weights this year is Normally distributed. But $n = 50 \geq 30$, so the large sample size makes it OK to use t procedures.

DO: From the Minitab output, $\bar{x} = 31.935$ ounces and $s_x = 2.394$ ounces.

- *Test statistic*

$$t = \frac{\bar{x} - \mu_0}{s_x/\sqrt{n}} = \frac{31.935 - 31}{2.394/\sqrt{50}} = 2.762$$

- *P-value* **Figure 9.17** displays the *P*-value for this two-sided test as an area under the *t* distribution curve with 50 − 1 = 49 degrees of freedom.

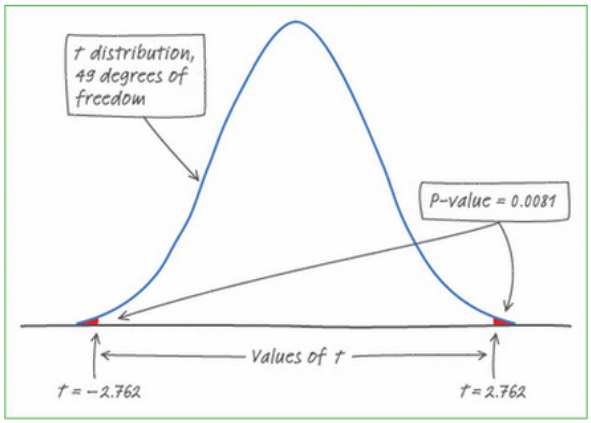


Figure 9.17 The *P*-value for the two-sided test with $t = 2.762$.

Upper-tail probability *p*			
df	.005	.0025	.001
30	2.750	3.030	3.385
40	2.704	2.971	3.307
50	2.678	2.937	3.261
	99%	99.5%	99.8%
	Confidence level *C*		

Using the table: **Table B** doesn't have an entry for df = 49, so we have to use the more conservative df = 40. As the excerpt below shows, the upper-tail probability is between 0.0025 and 0.005. So the desired *P*-value for this two-sided test is between 2(0.0025) = 0.005 and 2(0.005) = 0.01.

Using technology: The calculator's T-Test command gives $t = 2.762$ and *P*-value 0.0081 using df = 49.

CONCLUDE: Because the *P*-value, 0.0081, is less than $\alpha = 0.05$, we reject H_0. We have convincing evidence that the mean weight of the pineapples grown this year is not 31 ounces.

(b) No. This was not a comparative experiment, so we cannot infer causation. It is possible that other things besides the irrigation system changed from last year's growing season. Maybe the weather was different this year, and that's why the pineapples have a different mean weight than last year.

The significance test in the previous example gives convincing evidence that the mean weight μ of the pineapples grown in the field this year differs from last year's 31 ounces. Unfortunately, the test doesn't give us an idea of what the actual value of μ is. For that, we need a confidence interval.

Example 19 Juicy Pineapples

Confidence intervals give more information

Minitab output for a significance test and confidence interval based on the pineapple data is shown below. The test statistic and *P*-value match what we got earlier (up to rounding).

One-Sample T: Weight (oz)

```
Test of mu = 31 vs not = 31
Variable      N   Mean  StDev SE Mean      95% CI         T     P
Weight (oz)  50 31.935 2.394   0.339  (31.255, 32.616) 2.76 0.008
```

The 95% confidence interval for the mean weight of all the pineapples grown in the field this year is 31.255 to 32.616 ounces. We are 95% confident that this interval captures the true mean weight μ of this year's pineapple crop.

As with proportions, there is a link between a two-sided test at significance level α and a $100(1 - \alpha)\%$ confidence interval for a population mean μ. For the pineapples, the two-sided test at $\alpha = 0.05$ rejects H_0: $\mu = 31$ in favor of H_a: $\mu \neq 31$. The corresponding 95% confidence interval does not include 31 as a plausible value of the parameter μ. In other words, the test and interval lead to the same conclusion about H_0. But the confidence interval provides much more information: a set of plausible values for the population mean.

The connection between two-sided tests and confidence intervals is even stronger for means than it was for proportions. That's because both inference methods for means use the standard error of $\bar{x}$ in the calculations.

$$\text{test statistic: } t = \frac{\bar{x} - \mu_0}{s_x/\sqrt{n}} \qquad \text{Confidence interval: } \bar{x} \pm t^* \frac{s_x}{\sqrt{n}}$$

When the two-sided significance test at level α rejects H_0: $\mu = \mu_0$, the $100(1 - \alpha)\%$ confidence interval for μ will not contain the hypothesized value μ_0. And when the test fails to reject the null hypothesis, the confidence interval will contain μ_0.

Is there a connection between one-sided tests and confidence intervals for a population mean?

As you might expect, the answer is yes. But the link is more complicated. Consider a one-sided test of H_0: $\mu = 10$ versus H_a: $\mu > 10$ based on an SRS of 30 observations. With df = 30 − 1 = 29, **Table B** says that the test will reject H_0 at $\alpha = 0.05$ if the test statistic t is greater than 1.699. For this to happen, the sample mean $\bar{x}$ would have to exceed $\mu_0 = 10$ by more than 1.699 standardized units.

Table B also shows that $t^* = 1.699$ is the critical value for a 90% confidence interval. That is, a 90% confidence interval will extend 1.699 standardized units on either side of the sample mean $\bar{x}$. If $\bar{x}$ exceeds 10 by more than 1.699 standardized units, the resulting interval will not include 10. And the one-sided test will reject H_0 : $\mu = 10$. There's the link: our one-sided test at $\alpha = 0.05$ gives the same conclusion about H_0 as a 90% confidence interval for μ.

CHECK YOUR UNDERSTANDING

The health director of a large company is concerned about the effects of stress on the company's middle-aged male employees. According to the National Center for Health Statistics, the mean systolic blood pressure for males 35 to 44 years of age is 128. The health director examines the medical records of a random sample of 72 male employees in this age group. The Minitab output displays the results of a significance test and a confidence interval.

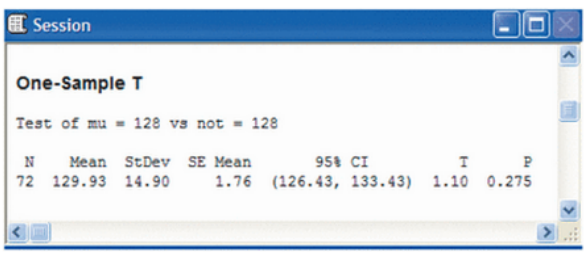
Check assumptions:
Random ✓ SRS n=72
10% ✓ >720 middle aged male employees at this company
Large ✓ n ≥ 30

1. Do the results of the significance test give convincing evidence that the mean blood pressure for all the company's middle-aged male employees differs from the national average? Justify your answer.

Fail to Reject H_0
Pvalue = .275 > α = .05
Not convincing evidence the mean blood pressure of the company's middle aged male employees is different than 128.

2. Interpret the 95% confidence interval in context. Explain how the confidence interval leads to the same conclusion as in Question 1.

We are 95% confident the true mean blood pressure for male middle-aged employees is between 126.43 and 133.43 (128 is on this interval so it is a plausible value).



Session

One-Sample T

Test of mu = 128 vs not = 128

N	Mean	StDev	SE Mean	95% CI	T	P
72	129.93	14.90	1.76	(126.43, 133.43)	1.10	0.275

9.3.4 Inference for Means: Paired Data

Study designs that involve making two observations on the same individual, or one observation on each of two similar individuals, yield **paired data.** When paired data result from measuring the same quantitative variable twice, we can make comparisons by analyzing the differences in each pair. If the conditions for inference are met, we can use one-sample t procedures to perform inference about the mean difference μ_d. (These methods are sometimes called **paired *t* procedures**). An example should help illustrate what we mean.

Example 20 Is Caffeine Dependence Real?

Paired data and one-sample t procedures

Researchers designed an experiment to study the effects of caffeine withdrawal. They recruited 11 volunteers who were diagnosed as being caffeine dependent to serve as subjects. Each subject was barred from coffee, colas, and other substances with caffeine for the duration of the experiment. During one 2-day period, subjects took capsules containing their normal caffeine intake. During another 2-day period, they took placebo capsules. The order in which subjects took caffeine and the placebo was randomized. At the end of each 2-day period, a test for depression was given to all 11 subjects. Researchers wanted to know whether being deprived of caffeine would lead to an increase in depression.[18]

The table below contains data on the subjects' scores on the depression test. Higher scores show more symptoms of depression. For each subject, we calculated the difference in test scores following each of the two treatments (placebo − caffeine). We chose this order of subtraction to get mostly positive values.

Results of a caffeine-deprivation study

Subject	Depression (caffeine)	Depression (placebo)	Difference (placebo − caffeine)
1	5	16	11
2	5	23	18
3	4	5	1
4	3	7	4
5	8	14	6
6	5	24	19
7	0	6	6
8	0	3	3
9	2	15	13
10	11	12	1
11	1	0	−1

PROBLEM:

(a) Why did researchers randomly assign the order in which subjects received placebo and caffeine?

(b) Carry out a test to investigate the researchers' question.

SOLUTION:

(a) Researchers want to be able to conclude that any statistically significant change in depression score is due to the treatments themselves and not to some other variable. One obvious concern is the order of the treatments. Suppose that caffeine were given to all the subjects during the first 2-day period. What if the weather were nicer on these 2 days than during the second 2-day period when all subjects were given a placebo? Researchers wouldn't be able to tell if a large increase in the mean depression score is due to the difference in weather or due to the treatments. Random assignment of the caffeine and placebo to the two time periods in the experiment should help ensure that no other variable (like the weather) is systematically affecting subjects' responses.

(b) We'll follow the four-step process.

STATE: If caffeine deprivation has no effect on depression, then we would expect the actual mean difference in depression scores to be 0. We therefore want to test the hypotheses

$$H_0: \mu_d = 0$$
$$H_a: \mu_d > 0$$

where μ_d is the true mean difference (placebo − caffeine) in depression score for subjects like these. Because no significance level is given, we'll use $\alpha = 0.05$.

PLAN: If the conditions are met, we should conduct a paired t test for μ_d

- *Random:* Researchers randomly assigned the treatments—placebo then caffeine, caffeine then placebo—to the subjects.
  - 10%: We aren't sampling, so it isn't necessary to check the 10% condition.
- *Normal/Large Sample:* We don't know whether the actual distribution of difference in depression scores (placebo–caffeine) for subjects like these is Normal. With such a small sample size ($n = 11$), we need to graph the data to see if it's safe to use t procedures. **Figure 9.18** shows hand sketches of a calculator histogram, boxplot, and Normal probability plot for these data. The histogram has an irregular shape with so few values; the boxplot shows some right skewness but no outliers; and the Normal probability plot is slightly curved, indicating mild skewness. With no outliers or strong skewness, the t procedures should be fairly accurate.