

10.3 Hypothesis Tests for a Population Mean

Preparing for This Section Before getting started, review the following:

- Sampling distribution of \bar{x} (Section 8.1, pp. 423–431)
- The t -distribution (Section 9.2, pp. 463–467)
- Using probabilities to identify unusual results (Section 5.1, p. 277)
- Confidence intervals for a mean (Section 9.2, pp. 467–469)

Objectives

- ① Test hypotheses about a mean
- ② Understand the difference between statistical significance and practical significance

① Test Hypotheses about a Mean

In Section 8.1, we learned that the distribution of \bar{x} is approximately normal with mean $\mu_{\bar{x}} = \mu$ and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ provided the population from which the sample was drawn is normally distributed or the sample size is sufficiently large (because of the Central Limit Theorem). So $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$ follows a standard normal distribution.

However, it is unreasonable to expect to know σ without knowing μ . This problem was resolved by William Gosset, who determined that $t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$ follows Student's

t -distribution with $n - 1$ degrees of freedom. We use this distribution to perform hypothesis tests on a mean.

Testing hypotheses about a mean follows the same logic as testing a hypothesis about a population proportion. The only difference is that we use Student's t -distribution, rather than the normal distribution.

Testing Hypotheses Regarding a Population Mean

To test hypotheses regarding the population mean, use the following steps, provided that

- the sample is obtained using simple random sampling or from a randomized experiment.
- the sample has no outliers and the population from which the sample is drawn is normally distributed, or the sample size, n , is large ($n \geq 30$).
- the sampled values are independent of each other.

Step 1 Determine the null and alternative hypotheses. The hypotheses can be structured in one of three ways:

Two-Tailed	Left-Tailed	Right-Tailed
$H_0: \mu = \mu_0$	$H_0: \mu = \mu_0$	$H_0: \mu = \mu_0$
$H_1: \mu \neq \mu_0$	$H_1: \mu < \mu_0$	$H_1: \mu > \mu_0$

Note: μ_0 is the assumed value of the population mean.

Step 2 Select a level of significance, α , depending on the seriousness of making a Type I error.

Classical Approach

Step 3 Compute the **test statistic**

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

which follows Student's *t*-distribution with $n - 1$ degrees of freedom.

Use Table VII to determine the critical value.

	Two-Tailed	Left-Tailed	Right-Tailed
Critical value(s)	$-t_{\frac{\alpha}{2}}$ and $t_{\frac{\alpha}{2}}$	$-t_{\alpha}$	t_{α}

Step 4 Compare the critical value to the test statistic.

Two-Tailed	Left-Tailed	Right-Tailed
If $t_0 < -t_{\frac{\alpha}{2}}$ or $t_0 > t_{\frac{\alpha}{2}}$, reject the null hypothesis.	If $t_0 < -t_{\alpha}$, reject the null hypothesis.	If $t_0 > t_{\alpha}$, reject the null hypothesis.

P-Value Approach

By Hand Step 3 Compute the **test statistic**

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

which follows Student's *t*-distribution with $n - 1$ degrees of freedom.

Use Table VII to approximate the *P*-value.

Two-Tailed	Left-Tailed	Right-Tailed
The sum of the area in the tails is the <i>P</i> -value $P\text{-value} = 2P(t > t_0)$	The area left of t_0 is the <i>P</i> -value $P\text{-value} = P(t < t_0)$	The area right of t_0 is the <i>P</i> -value $P\text{-value} = P(t > t_0)$

Technology Step 3 Use a statistical spreadsheet or calculator with statistical capabilities to obtain the *P*-value. The directions for obtaining the *P*-value using the TI-83/84 Plus graphing calculators, Minitab, Excel, and StatCrunch are in the Technology Step-by-Step on page 528.

Step 4 If the *P*-value $< \alpha$, reject the null hypothesis.

Step 5 State the conclusion.

Notice that the procedure just presented requires either that the population from which the sample was drawn be normal or that the sample size be large ($n \geq 30$). The procedure is robust, so minor departures from normality will not adversely affect the results of the test. However, if the data include outliers, the procedure should not be used.

We will verify these assumptions by constructing normal probability plots (to assess normality) and boxplots (to discover whether there are outliers). If the normal probability plot indicates that the data do not come from a normal population or if the boxplot reveals outliers, nonparametric tests should be performed (Chapter 15).

Before we look at a couple of examples, it is important to understand that we cannot find exact *P*-values using the *t*-distribution table (Table VII) because the table provides *t*-values only for certain areas. However, we can use the table to calculate lower and upper bounds on the *P*-value. To find exact *P*-values, use statistical software or a graphing calculator with advanced statistical features.

EXAMPLE 1 Testing a Hypothesis about a Population Mean: Large Sample

Problem The mean height of American males is 69.5 inches. The heights of the 43 male U.S. presidents* (Washington through Obama) have a mean 70.78 inches and a standard deviation of 2.77 inches. Treating the 43 presidents as a simple random sample, determine if there is evidence to suggest that U.S. presidents are taller than the average American male. Use the $\alpha = 0.05$ level of significance.

Approach Assume that all U.S. presidents come from a population whose height is 69.5 inches (that is, there is no difference between heights of U.S. presidents and the general American male population). Then determine the likelihood of obtaining a sample mean of 70.78 inches or higher from a population whose mean is 69.5 inches.

(continued)

*Grover Cleveland was elected to two non-consecutive terms, so there have technically been 44 presidents of the United States.

If the result is unlikely, reject the assumption stated in the null hypothesis in favor of the more likely notion that the mean height of U.S. presidents is greater than 69.5 inches. However, if obtaining a sample mean of 70.78 inches from a population whose mean is assumed to be 69.5 inches is not unusual, do not reject the null hypothesis (and attribute the difference to sampling error). Assume the population of potential U.S. presidents is large (for independence). Because the sample size is large, the distribution of \bar{x} is approximately normal. Follow Steps 1 through 5.

Solution

Step 1 We want to know if U.S. presidents are taller than the typical American male who is 69.5 inches. We assume there is no difference between the height of a typical American male and U.S. presidents, so

$$H_0: \mu = 69.5 \text{ inches} \quad \text{versus} \quad H_1: \mu > 69.5 \text{ inches}$$

Step 2 The level of significance is $\alpha = 0.05$.

Classical Approach

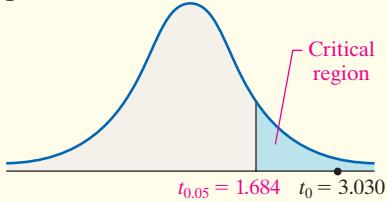
Step 3 The sample mean is $\bar{x} = 70.78$ inches and the sample standard deviation is $s = 2.77$ inches.

The test statistic is

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{70.78 - 69.5}{\frac{2.77}{\sqrt{43}}} = 3.030$$

Because this is a right-tailed test, determine the critical value at the $\alpha = 0.05$ level of significance with $43 - 1 = 42$ degrees of freedom to be $t_{0.05} = 1.684$ (using 40 degrees of freedom since this is closest to 42). The critical region is shown in Figure 14.

Figure 14



Step 4 The test statistic, $t_0 = 3.030$, is labeled in Figure 14. Because the test statistic lies in the critical region, reject the null hypothesis.

P-Value Approach

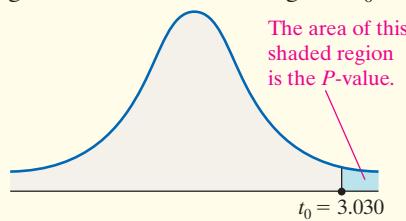
By Hand Step 3 The sample mean is $\bar{x} = 70.78$ inches and the sample standard deviation is $s = 2.77$ inches.

The test statistic is

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{70.78 - 69.5}{\frac{2.77}{\sqrt{43}}} = 3.030$$

Because this is a right-tailed test, the P -value is the area under the t -distribution with 42 degrees of freedom to the right of $t_0 = 3.030$ as shown in Figure 15.

Figure 15



Using Table VII, find the row that corresponds to 40 degrees of freedom (we use 40 degrees of freedom because it is closest to the actual degrees of freedom, $43 - 1 = 42$). The value 3.030 lies between 2.971 and 3.307. The value of 2.971 has an area of 0.0025 to the right under the t -distribution with 40 degrees of freedom. The area under the t -distribution with 40 degrees of freedom to the right of 3.307 is 0.001.

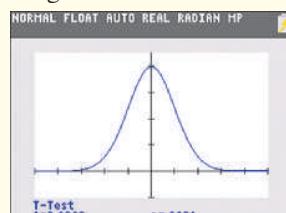
Because 3.030 is between 2.971 and 3.307, the P -value is between 0.001 and 0.0025. So

$$0.001 < P\text{-value} < 0.0025$$

There is an alternate form of Table VII (see Table XVII in Appendix A) that is useful for finding more accurate P -values. The table is set up similarly to Table V (the standard normal table). To use this alternate version, find the column that corresponds to the degrees of freedom, and the row that corresponds to the test statistic (rounded to the nearest tenth). The intersection of the row and column represents the area under the t -distribution to the right of the test statistic. Using 40 degrees of freedom (because 42 df is not in the table) with a test statistic of 3.0, the P -value is 0.002.

Technology Step 3 Using a TI-84 Plus C graphing calculator, the P -value is 0.0021. See Figure 16.

Figure 16



Note

If you are using Table XVII to find P -values for a left-tailed test, use the symmetry of the t -distribution. That is,

$$P(t < -t_0) = P(t > t_0)$$

Now Work Problem 13

Step 4 The P -value of 0.0021 [by hand: $0.001 < P\text{-value} < 0.0025$] means that, if the null hypothesis that $\mu = 69.5$ inches is true, we expect a sample mean of 70.78 inches or higher in about 2 out of 1000 samples. The results we obtained do not seem to be consistent with the assumption that the mean height of this population is 69.5 inches. Put another way, because the P -value is less than the level of significance, $\alpha = 0.05$ ($0.0021 < 0.05$), we reject the null hypothesis.

Step 5 There is sufficient evidence at the $\alpha = 0.05$ level of significance to conclude that U.S. presidents are taller than the typical American male.

EXAMPLE 2 Testing a Hypothesis about a Population Mean: Small Sample

Table 1

19.68	20.66	19.56
19.98	20.65	19.61
20.55	20.36	21.02
21.50	19.74	

Source: Michael Carlisle, student at Joliet Junior College

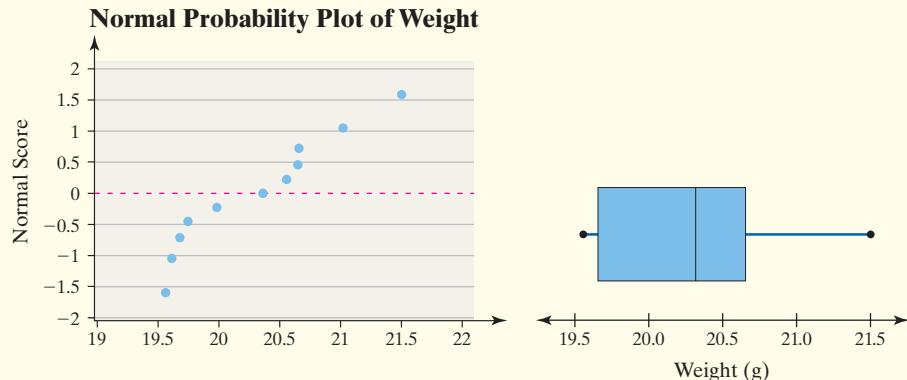
Problem The “fun size” of a Snickers bar is supposed to weigh 20 grams. Because the penalty for selling candy bars under their advertised weight is severe, the manufacturer calibrates the machine so the mean weight is 20.1 grams. The quality-control engineer at M&M–Mars, the Snickers manufacturer, is concerned about the calibration. He obtains a random sample of 11 candy bars, weighs them, and obtains the data shown in Table 1. Should the machine be shut down and calibrated? Because shutting down the plant is very expensive, he decides to conduct the test at the $\alpha = 0.01$ level of significance.

Approach Assume that the machine is calibrated correctly. So there is no difference between the actual mean weight and the calibrated weight of the candy. We want to know whether the machine is incorrectly calibrated, which would result in a mean weight that is too high or too low. Therefore, this is a two-tailed test.

Before performing the hypothesis test, verify that the data come from a population that is normally distributed with no outliers by constructing a normal probability plot and boxplot. Then proceed to follow Steps 1 through 5.

Solution Figure 17 displays the normal probability plot and boxplot. The correlation between the weights and expected z -scores is 0.967 [Tech: 0.970]. Because $0.967 > 0.923$ (Table VI), the normal probability plot indicates that the data could come from a population that is approximately normal. The boxplot has no outliers. We can proceed with the hypothesis test.

Figure 17



Step 1 The engineer wishes to determine whether the Snickers have a mean weight of 20.1 grams or not. The hypotheses can be written

$$H_0: \mu = 20.1 \text{ grams} \quad \text{versus} \quad H_1: \mu \neq 20.1 \text{ grams}$$

This is a two-tailed test.

Step 2 The level of significance is $\alpha = 0.01$.

(continued)

Classical Approach

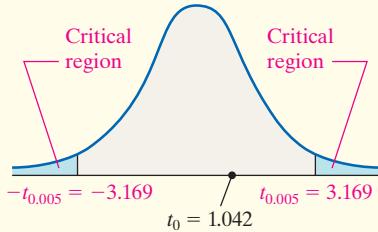
Step 3 From the data in Table 1, the sample mean is $\bar{x} = 20.301$ grams and the sample standard deviation is $s = 0.64$ gram.

The test statistic is

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{20.301 - 20.1}{\frac{0.64}{\sqrt{11}}} = 1.042$$

Because this is a two-tailed test, determine the critical values at the $\alpha = 0.01$ level of significance with $11 - 1 = 10$ degrees of freedom to be $-t_{0.01/2} = -t_{0.005} = -3.169$ and $t_{0.01/2} = t_{0.005} = 3.169$. The critical regions are shown in Figure 18.

Figure 18



Step 4 Because the test statistic, $t_0 = 1.042$, does not lie in the critical region, do not reject the null hypothesis.

P-Value Approach

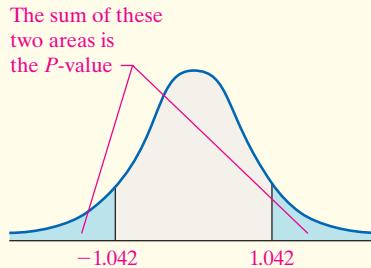
By Hand Step 3 From the data in Table 1, the sample mean is $\bar{x} = 20.301$ grams and the sample standard deviation is $s = 0.64$ gram.

The test statistic is

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{20.301 - 20.1}{\frac{0.64}{\sqrt{11}}} = 1.042$$

Because this is a two-tailed test, the P -value is the area under the t -distribution with $n - 1 = 11 - 1 = 10$ degrees of freedom to the left of $-t_0 = -1.042$ and to the right of $t_0 = 1.042$, as shown in Figure 19. That is, P -value = $P(t < -1.042) + P(t > 1.042) = 2P(t > 1.042)$, with 10 degrees of freedom.

Figure 19



Using Table VII, we find the row that corresponds to 10 degrees of freedom. The value 1.042 lies between 0.879 and 1.093. The value of 0.879 has an area of 0.20 to the right under the t -distribution. The area under the t -distribution to the right of 1.093 is 0.15.

Because 1.042 is between 0.879 and 1.093, the P -value is between $2(0.15)$ and $2(0.20)$. So

$$0.30 < P\text{-value} < 0.40$$

Using Table XVII we find P -value = $2P(t > 1.0) = 2(0.170) = 0.340$ with 10 degrees of freedom.

Technology Step 3 Using Minitab, the exact P -value is 0.323.

Step 4 The P -value of 0.323 [by-hand: $0.30 < P$ -value < 0.40] means that, if the null hypothesis that $\mu = 20.1$ grams is true, we expect about 32 out of 100 samples to result in a sample mean as extreme or more extreme than the one obtained. The result we obtained is not unusual, so we do not reject the null hypothesis.

Step 5 There is not sufficient evidence to conclude that the Snickers have a mean weight different from 20.1 grams at the $\alpha = 0.01$ level of significance. The machine should not be shut down. •

• Now Work Problem 21

In Other Words

Results are statistically significant if the difference between the observed result and the statement made in the null hypothesis is unlikely to occur due to chance alone.

②

Understand the Difference between Statistical Significance and Practical Significance

When a large sample size is used in a hypothesis test, the results could be statistically significant even though the difference between the sample statistic and mean stated in the null hypothesis may have no *practical significance*.

Definition

Practical significance refers to the idea that, while small differences between the statistic and parameter stated in the null hypothesis are statistically significant, the difference may not be large enough to cause concern or be considered important.

EXAMPLE 3 Statistical versus Practical Significance

Problem According to the American Community Survey, the mean travel time to work in Collin County, Texas, in 2013 was 27.5 minutes. The Department of Transportation reprogrammed all the traffic lights in Collin County in an attempt to reduce travel time. To determine if there is evidence that travel time has decreased as a result of the reprogramming, the Department of Transportation obtains a random sample of 2500 commuters, records their travel time to work, and finds a sample mean of 27.2 minutes with a standard deviation of 8.5 minutes. Does this result suggest that travel time has decreased at the $\alpha = 0.05$ level of significance?

Approach We will use both the classical and P -value approach to test the hypothesis.

Solution

Step 1 The Department of Transportation wants to know if the mean travel time to work has decreased from 27.5 minutes. From this, we have

$$H_0: \mu = 27.5 \text{ minutes} \quad \text{versus} \quad H_1: \mu < 27.5 \text{ minutes}$$

Step 2 The level of significance is $\alpha = 0.05$.

Step 3 The test statistic is

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{27.2 - 27.5}{\frac{8.5}{\sqrt{2500}}} = -1.765$$

Classical Approach

Because this is a left-tailed test, the critical value with $\alpha = 0.05$ and $2500 - 1 = 2499$ degrees of freedom is $-t_{0.05} \approx -1.645$ (use the last row of Table VII when the degrees of freedom is greater than 1000).

Step 4 Because the test statistic is less than the critical value (the test statistic falls in the critical region), we reject the null hypothesis.

Step 5 There is sufficient evidence at the $\alpha = 0.05$ level of significance to conclude the mean travel time to work has decreased.

While the difference between 27.2 minutes and 27.5 minutes is statistically significant, it has no practical meaning. After all, is 0.3 minute (18 seconds) really going to make anyone feel better about his or her commute to work? •

 P -Value Approach

Because this is a left-tailed test, P -value = $P(t_0 < -1.765)$. From Table VII, we find the approximate P -value is $0.025 < P$ -value < 0.05 [Technology: P -value = 0.0389].

Step 4 Because the P -value is less than the level of significance, $\alpha = 0.05$, we reject the null hypothesis.

The reason that the results from Example 3 were statistically significant had to do with the large sample size. The moral of the story is this:

CAUTION!

Beware of studies with large sample sizes that claim statistical significance because the differences may not have any practical meaning.

Large sample sizes can lead to results that are statistically significant, while the difference between the statistic and parameter in the null hypothesis is not enough to be considered practically significant.