Chapter 10 Hypothesis Tests Regarding a Parameter
Section 10.1 The Language of Hypothesis Testing

Objectives

1. Determine the null and alternative hypotheses
2. Explain Type I and Type II errors
3. State conclusions to hypothesis tests

EXAMPLE An Easy "A"?

For each student in the class, I am going to flip a coin. If the coin comes up heads, your grade in the class is the grade earned based on the syllabus. However, if the coin comes up tails, you automatically earn an A.

Suppose the outcome of the first five students is H, H, H, H, H. What might you conclude?

Based on the analysis above, you can make one of two conclusions:

1. The coin is fair, but just happened to come up five heads in a row.
2. The coin is not fair.

Are you willing to accuse your instructor of using a coin that is not fair (that is, biased toward flipping heads)?

This is at the heart of *hypothesis testing*. An assumption is made about reality. We then look at sample evidence to determine if it contradicts or is consistent with our assumption.

1 Determine the Null and Alternative Hypotheses

A **hypothesis** is a statement regarding a characteristic of one or more populations.

Consider the following:

(A) According to the American Time Use Survey, the mean amount of time employed male Americans spend working on weekends 5.46 hours. A researcher wonders if males employed part-time work more hours, on average, on weekends.
(B) Harris Interactive reports that 55% of adult Americans (aged 18 or over) prefer purchasing name brand coffee over generic brands. A marketing manager with Starbux wonders if the percentage of seniors who prefer name brand coffee differs from that all adult Americans.
(C) Using an old manufacturing process, the standard deviation of the amount of wine put in a bottle was 0.23 ounces. With new equipment, the quality control manager believes the standard deviation has decreased.

Hypothesis testing is a procedure, based on sample evidence and probability, used to test statements regarding a characteristic of one or more populations.

Steps in Hypothesis Testing

1. Make a statement regarding the nature of the population.
2. Collect evidence (sample data) to test the statement.
3. Analyze the data to assess the plausibility of the statement.

The **null hypothesis**, denoted H_0 (read "H-naught"), is a statement to be tested. The null hypothesis is a statement of no change, no effect, or no difference and is assumed true until evidence indicates otherwise.

The **alternative hypothesis**, denoted H_1 (read "H-one"), is a statement that we are trying to find evidence to support.

In this chapter, there are three ways to set up the null and alternative hypotheses.

1. Equal hypothesis versus not equal hypothesis **(two-tailed test)**

$$H_0\text{: parameter} = \text{some value}$$
$$H_1\text{: parameter} \neq \text{some value}$$

2. Equal versus less than **(left-tailed test)**

$$H_0\text{: parameter} = \text{some value}$$
$$H_1\text{: parameter} < \text{some value}$$

3. Equal versus greater than **(right-tailed test)**

$$H_0\text{: parameter} = \text{some value}$$
$$H_1\text{: parameter} > \text{some value}$$

EXAMPLE Forming Hypotheses

Determine the null and alternative hypothesis for each of the following. State whether the test is two-tailed, left-tailed, or right-tailed.

(a) According to the American Time Use Survey, the mean amount of time employed male Americans spend working on weekends 5.46 hours. A researcher wonders if males employed part-time work more hours, on average, on weekends.
(b) Harris Interactive reports that 55% of adult Americans (aged 18 or over) prefer purchasing name brand coffee over generic brands. A marketing manager with Starbux wonders if the percentage of seniors who prefer name brand coffee differs from that all adult Americans.
(c) Using an old manufacturing process, the standard deviation of the amount of wine put in a bottle was 0.23 ounces. With new equipment, the quality control manager believes the standard deviation has decreased.

2 Explain Type I and Type II Errors

Four Outcomes from Hypothesis Testing

1. Reject the null hypothesis when the alternative hypothesis is true. This decision would be correct.
2. Do not reject the null hypothesis when the null hypothesis is true. This decision would be correct.
3. Reject the null hypothesis when the null hypothesis is true. This decision would be incorrect. This type of error is called a **Type I error**.
4. Do not reject the null hypothesis when the alternative hypothesis is true. This decision would be incorrect. This type of error is called a **Type II error**.

		Reality	
		H_0 Is True	H_1 Is True
Conclusion	Do Not Reject H_0	Correct Conclusion	Type II Error
	Reject H_0	Type I Error	Correct Conclusion

EXAMPLE Type I and Type II Errors

Explain what it would mean to make a Type I error for the following hypothesis test. What would it mean to make a Type II error?

According to the American Time Use Survey, the mean amount of time employed male Americans spend working on weekends 5.46 hours. A researcher wonders if males employed part-time work more hours, on average, on weekends.

The Probability of Making a Type I or Type II Error

$$\alpha = P(\text{Type I error}) = P(\text{rejecting } H_0 \text{ when } H_0 \text{ is true})$$
$$\beta = P(\text{Type II error}) = P(\text{not rejecting } H_0 \text{ when } H_1 \text{ is true})$$

The probability of making a Type I error, α, is chosen by the researcher before the sample data is collected.

The **level of significance**, α, is the probability of making a Type I error.

❸ State Conclusions to Hypothesis Tests

CAUTION!

We never "accept" the null hypothesis, because, without having access to the entire population, we don't know the exact value of the parameter stated in the null. Rather, we say that we do not reject the null hypothesis. This is just like the court system. We never declare a defendant "innocent", but rather say the defendant is "not guilty".

EXAMPLE Stating the Conclusion

According to the American Time Use Survey, the mean amount of time employed male Americans spend working on weekends 5.46 hours. A researcher wonders if males employed part-time work more hours, on average, on weekends.

(a) Suppose the null hypothesis is rejected. State the conclusion.

(b) Suppose the null hypothesis is not rejected. State the conclusion.

10.2A Hypothesis Tests for a Population Proportion with Simulation

Objectives

1. Explain the logic of the simulation method
2. Test the hypotheses about a population proportion using the simulation method

The material in Section 10.1 introduced the language of hypothesis testing. Now, we discuss a method for testing hypotheses about a population proportion. In these tests, the null hypothesis will be H_o: $p = p_o$ versus one of three alternative hypotheses:

(1) H_1: $p \neq p_0$	(2) H_1: $p < p_0$	(3) H_1: $p > p_0$
Two-tailed	Left-tailed	Right-tailed

In each of these hypothesis tests, the value of p_o is a proportion between 0 and 1 and is the *status quo* (no change or no effect) value of the population proportion. Remember, the statement in the null hypothesis is assumed to be true and we are looking for evidence in support of the statement in the alternative hypothesis.

Consider the following scenario. At Joliet Junior College, the proportion of students who pass Grant Alexander's Intermediate Algebra is 0.5. In an effort to increase the pass rate, Grant decided to start "flipping" his class. Flipping refers to the idea that lectures are recorded on video and watched at home by the student, so that class time is utilized for homework, group learning activities, and explorations. In Professor Alexanders Intermediate Algebra course, 29 of the 50 students passed the course. Do the results suggest that the pass rates in Alexander's Intermediate Algebra course increased?

What are the null and alternative hypotheses for this study? What is the sample proportion? Is it possible that the proportion of students who passed is still 0.5 and the class just happened to have a higher proportion of students passing? What would be convincing, or *statistically significant*, evidence to you?

DEFINITION
When observed results are unlikely under the assumption that the null hypothesis is true, we say the result is **statistically significant** and reject the statement in the null hypothesis.

To determine if the results are statistically significant, we build a model that generates data randomly under the assumption the statement in the null hypothesis is true (the proportion of students who pass is 0.5). Call this model the **null model**. Then compare the results of the randomly generated data from the null model to those observed to see if the observed results are unusual.

How do we generate the random data? If the pass rate is still 0.5, then a coin could be used to represent a random student. Flip a coin – if the coin comes up heads, then the student passes the class; if it comes up tails, the student does not pass.

Activity The Logic of the P-value Approach to Hypothesis Testing

For the Alexander Intermediate Algebra scenario, answer the following.

(a) Go to StatCrunch and open the Coin-Flipping applet (Applets > Simulation > Coin flipping). Enter 0.5 for the probability of heads and 50 for the number of coins. Set Tally heads in tosses to "Number", select the >= inequality in the drop-down menu, and enter 29 in the cell (29 represents the observed number of students who passed the class). Click Comptue!. Click 1 run. Explain what this result represents.

(b) Repeat part (a). Did you get the same results? Explain.

(c) Now, flip the 50 coins at least 5000 times. What proportion of the simulations resulted in 29 or more heads (29 or more students who pass)? What does this result suggest?

A ***P*-value** is the probability of observing a sample statistic as extreme or more extreme than one observed under the assumption that the statement in the null hypothesis is true. Put another way, the *P*-value is the likelihood or probability that a sample will result in a statistic such as the one obtained if the null hypothesis is true.

(d) Click Options, then Edit. Select the Proportion radio button under Tally type 1 balls. Change the direction of the inequality to >= and enter 29/50 = 0.58 in the cell. Click Compute! Click 1000 runs 5 times for a total of 5000 simulations. Describe the shape of the distribution of sample proportions. Approximately where is the center of the distribution? Is this surprising?

In the Intermediate Algebra study, a coin was used to simulate outcomes. Next, we use a different method of simulating outcomes.

Activity – Using the Urn Applet to Build a Null Model

In a Pew Research poll conducted in 2017, it was reported that 43% of adult Americans (aged 18 and older) believe the reason a person is rich is because the individual "had advantages in life." In a random sample of 80 Republicans, it was found that 23 believe the reason a person is rich is because the individual "had advantages in life." Does the sample evidence suggest a lower proportion of Republicans believe the reason a person is rich is because the individual "had advantages in life"? What are the null and alternative hypotheses?

(a) Go to StatCrunch and open the Urn sampling applet (Applets > Simulation > Urn sampling). Assume there are 80 million registered Republicans in the country. Build an Urn with 0.43(80,000,000) = 34,400,000 green balls (to represent the Republicans who believe the reason a person is rich is because the individual "had advantages in life.") and 0.57(80,000,000) = 45,600,000 red balls to represent the Republicans who do not believe the reason a person is rich is because the individual "had advantages in life." . Set Number of balls to draw to 80. Set Tally type to "Number", select the <= inequality in the drop-down menu, and enter 23 in the cell. Click Compute!. Click 1 run. Explain what the results represent.

(b) Repeat part (a). Did you get the same results? Explain.

(c) Now, draw 80 balls at least 5000 times. What proportion of the simulations resulted in fewer than 23 green balls (23 Republicans who believe the reason a person is rich is because the individual "had advantages in life.")? What does this result suggest?

In simulating results to approximate the P-value, you need to decide whether the coin-flipping applet or the urn applet is a better choice. In most cases, either applet may be used. However, there are some guidelines that may be used to help decide.

- Is the scenario a sequence of events? For a sequence of events, the coin-flipping applet may be better. For example, in the Grant Alexander study, we had a sequence of 40 events, so the coin-flipping applet makes sense.
- Is it easy to build a population from the statement in the null hypothesis? If so, the urn applet may be better. In the redistricting study, we were able to build a population of voters, so the urn applet makes sense.

Hypothesis Testing Using the P-value Approach

If the probability of getting a sample statistic as extreme or more extreme than the one obtained is small under the assumption the statement in the null hypothesis is true, reject the null hypothesis.

A Rule of Thumb

If the P-value < 0.05 (the level of significance), we reject the statement in the null hypothesis.

Testing Hypotheses Regarding a Population Proportion Using Simulation

Step 1. Verify that the variable of interest in the study is qualitative with two possible outcomes.

Step 2. Determine the null and alternative hypotheses. The hypotheses can be structured in one of three ways.

Two-Tailed	Left-Tailed	Right-Tailed
$H_0: p = p_0$	$H_0: p = p_0$	$H_0: p = p_0$
$H_1: p \neq p_0$	$H_1: p < p_0$	$H_1: p > p_0$

Note: p_0 is the assumed value of the population proportion.

Step 3. Build a null model that generates data randomly under the assumption the statement in the null hypothesis is true.

Step 4. Estimate the P-value from the model in Step 3.

Step 5. State the conclusion.

EXAMPLE

Do 10.2A.7 from MyLabStats

Section 10.2B Hypothesis Testing on a Population Proportion Using the Normal Model

Objectives

1. Explain the Logic of Hypothesis Testing Using the Normal Model
2. Test Hypotheses about a Population Proportion Using the Normal Model

In Section 10.2A, simulation was used to test a hypothesis regarding a population proportion. In both simulations, you may have noticed the distribution of simulated results was bell-shaped (approximately normal).

Activity – Why Can We Use the Normal Model to Estimate P-values?

In a Pew Research poll conducted in 2017, it was reported that 43% of adult Americans (aged 18 and older) believe the reason a person is rich is because the individual "had advantages in life." In a random sample of 80 Republicans, it was found that 23 believe the reason a person is rich is because the individual "had advantages in life." Does the sample evidence suggest a lower proportion of Republicans believe the reason a person is rich is because the individual "had advantages in life"? What are the null and alternative hypotheses?

(a) Go to StatCrunch and open the Urn sampling applet (Applets > Simulation > Urn sampling). Assume there are 80 million registered Republicans in the country. Build an Urn with 0.43(80,000,000) = 34,400,000 green balls (to represent the Republicans who believe the reason a person is rich is because the individual "had advantages in life.") and 0.57(80,000,000) = 45,600,000 red balls to represent the Republicans who do not believe the reason a person is rich is because the individual "had advantages in life." Set Tally type to Proportion; set the drop-down menu to <=, and enter 23/80 = 0.2875 in the cell. Click Compute! Run 5000 simulations (click 1000 runs five times). What is the shape of the distribution of sample proportions?

(b) Click Analyze to export the results of the 5000 simulations to the StatCrunch spreadsheet. Compute the mean and standard deviation of the 5000 sample proportions. Now, compute the theoretical mean and standard deviation of the distribution of the sample proportion based on the results found in Section 8.2.

(c) Verify the normal model may be used to describe the sampling distribution of the sample proportion. Use the normal model to determine the probability of observing 23 or fewer Republicans (out of 80) who believe the reason a person is rich is because the individual "had advantages in life." Assuming the population proportion is 0.43. Compare the result to that found in Section 10.2A.

Hypothesis Testing Using the *P*-Value Approach

If the probability of getting a sample statistic as extreme or more extreme than the one obtained is small under the assumption the statement in the null hypothesis is true, reject the null hypothesis.

Testing Hypotheses Regarding a Population Proportion, p

Use Steps 1 through 5, provided that

- the sample is obtained by simple random sampling or the data result from a randomized experiment.
- $np_0(1 - p_0) \geq 10$.
- 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
$H_0: p = p_0$	$H_0: p = p_0$	$H_0: p = p_0$
$H_1: p \neq p_0$	$H_1: p < p_0$	$H_1: p > p_0$

Note: p_0 is the assumed value of the population proportion.

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

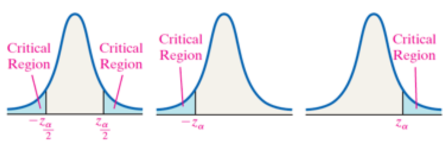
Classical Approach

Step 3 Compute the **test statistic**

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

Use Table V to determine the critical value.

	Two-Tailed	Left-Tailed	Right-Tailed
Critical value	$-z_{\frac{\alpha}{2}}$ and $z_{\frac{\alpha}{2}}$	$-z_{\alpha}$	z_{α}



Step 4 Compare the critical value with the test statistic.

Two-Tailed	Left-Tailed	Right-Tailed
If $z_0 < -z_{\frac{\alpha}{2}}$ or $z_0 > z_{\frac{\alpha}{2}}$, reject the null hypothesis.	If $z_0 < -z_{\alpha}$, reject the null hypothesis.	If $z_0 > z_{\alpha}$, reject the null hypothesis.

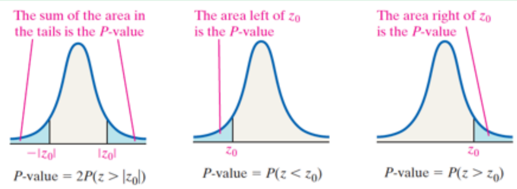
P-Value Approach

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

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

Use Table V to determine the *P*-value.

Two-Tailed	Left-Tailed	Right-Tailed



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 pages 489–490.

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

Step 5 State the conclusion.

EXAMPLE

PunditTracker is a website that tracks predictions of various political, sports, and pop culture "experts." It tracked 15 of Fox News anchor Bill O'Reilly and found that his predictions were correct 9 times.

a. Does this suggest O'Reilly is correct a majority of the time? Test this hypothesis using simulation.

b. Suppose the prediction rate Bill O'Reilly had in part (a) continues and over the course of 75 predictions, 45 were correct. Explain why the normal model could be used to approximate the *P*-value and find the *P*-value using the normal model.

c. Test the hypothesis from part (b) using simulation. Compare the results to the normal model.

EXAMPLE

In 1994, 21% of mothers aged 40 to 44 had exactly one child. Pew Research conducted a survey of 500 mothers aged 40 to 44 and found that 112 had exactly one child. Is this evidence to suggest the proportion of mothers aged 40 to 44 with one child has changed since 1994?

Two-Tailed Hypothesis Testing Using Confidence Intervals

When testing H_0: $p = p_0$ versus H_1: $p \neq p_0$, if a $(1 - \alpha) \cdot 100\%$ confidence interval contains p_0, do not reject the null hypothesis. However, if the confidence interval does not contain p_0, conclude that $p \neq p_0$ at the level of significance, α.

10.3 Hypothesis Tests for a Population Mean
Objectives

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

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

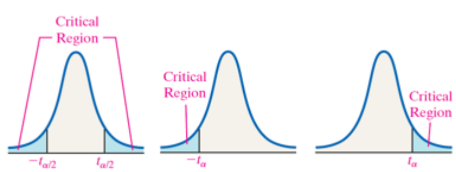
Step 3 Compute the **test statistic**

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

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

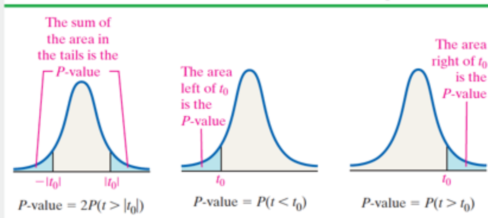
By Hand Step 3 Compute the **test statistic**

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

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



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 500.

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

Step 5 State the conclusion.

EXAMPLE Testing a Hypothesis about a Population Mean – Large Sample

A fast-food restaurant manager wants to see receipts per car traveling through the drive-thru increase. Historically, cars typically spend \$13.20. After implementing a new sales process to be used for cars using the drive-through, the manager randomly selects 35 cars and finds the mean bill is \$13.96 with a standard deviation of \$3.69. Does the sample evidence suggest the new process results in higher receipts per car? Use an $\alpha = 0.05$ level of significance.

EXAMPLE Testing a Hypothesis about a Population Mean – Small Sample

22. Reading Rates Michael Sullivan, son of the author, decided to enroll in a reading course that allegedly increases reading speed and comprehension. Prior to enrolling in the class, Michael read 198 words per minute (wpm). The following data represent the words per minute read for 10 different passages read after the course.

206	217	197	199	210
210	197	212	227	209

Does the evidence suggest the class was effective?

❷ Understand the Difference between Statistical Significance and Practical Significance

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.

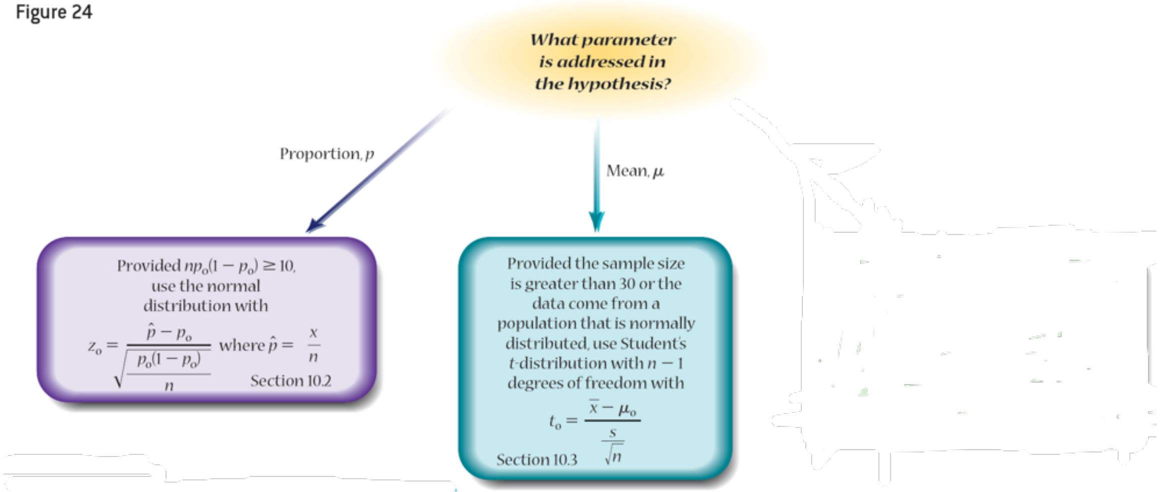
EXAMPLE Statistical versus Practical Significance

In 2003, the average age of a mother at the time of her first childbirth was 25.2. To determine if the average age has increased, a random sample of 1200 mothers is taken and is found to have a sample mean age of 25.5 with a standard deviation of 4.8, determine whether the mean age has increased using a significance level of $\alpha = 0.05$.

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.

Putting It Together: Which Method Do I Use?

Figure 24



10. The Atomic Bomb In October 1945, the Gallup organization asked 1487 randomly sampled Americans, "Do you think we can develop a way to protect ourselves from atomic bombs in case other countries tried to use them against us?" with 788 responding yes. Did a majority of Americans feel the United States could develop a way to protect itself from atomic bombs in 1945? Use the $\alpha = 0.05$ level of significance.

16. Sleepy? According to the National Sleep Foundation, children between the ages of 6 and 11 years should get 10 hours of sleep each night. In a survey of 56 parents of 6 to 11 year olds, it was found that the mean number of hours the children slept was 8.9 with a standard deviation of 3.2. Does the sample data suggest that 6 to 11 year olds are sleeping less than the required amount of time each night? Use the 0.01 level of significance.

Which type of test would be most appropriate for the following study?

28. Researchers measured regular testosterone levels in a random sample of athletes and then measured testosterone levels prior to an athletic event. They wanted to know whether testosterone levels increase prior to athletic events.