

## Chapter 9: Hypothesis Testing with One Sample

### Introduction

#### What is a hypothesis?

A **hypothesis** is a statement (or claim) about a property/characteristic of a population.

**Hypothesis testing** is a procedure, based on sample evidence and probability, for testing claims about a property/characteristic of a population.

#### Example 1. Does a Majority Favor a Lower Limit for Drunk Driving?

Suppose that your state legislature is considering a proposal to lower the legal limit for the blood alcohol concentration (BAC) that constitutes drunk driving. A legislator wants to determine whether a majority of adults in her district favor this proposal. To gather information, she surveys 200 randomly selected individuals from her district and learns that 59% of these people favor the proposal. This sample information will be used to decide whether more than 50% of the population in her district favors the proposal.

Can the legislator conclude that a majority of all adults in her district favor this proposal? Because the result is based on a sample, there is the possibility that the observed majority may have occurred just by the “luck of the draw.” If a majority of the whole population actually is opposed to the proposal (so 50% or less favor it), how likely is it that 59% of a random sample from this population would favor it?

There are six basic steps for any hypothesis test:

Step 1 Determine the null and alternative hypotheses.

Step 2 Verify all conditions have been met and state the level of significance.

Step 3 Summarize the data into an appropriate test statistic.

Step 4 Find the  $p$ -value by comparing the test statistic to the possibilities expected if the null hypothesis were true OR determine the critical value.

Step 5 Decide whether the result is statistically significant based on the  $p$ -value.

Step 6 Report the conclusion in the context of the situation.

Because we use sample data to test hypotheses, we cannot state with 100% certainty that the statement is true; we can only determine whether the sample data support the statement or not. In fact, because the statement can be either true or false, hypothesis testing is based on two types of hypotheses.

## 9.1: Null and Alternative Hypotheses

### Null Hypothesis ( $H_0$ ):

- A statement of no change, no effect, or no difference.
- Assumed true until evidence indicates otherwise.
- We either reject or fail to reject  $H_0$ . WE NEVER 'ACCEPT'  $H_0$  TO BE TRUE!!!

**Alternative Hypothesis ( $H_a$ ):** A statement that we are trying to find evidence to support; contradictory to  $H_0$ .

Three ways to set up the null and alternative hypotheses are:

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

$$H_0 : \text{parameter} = \text{null value}$$

$$H_1 : \text{parameter} \neq \text{null value}$$

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

$$H_0 : \text{parameter} = \text{null value}$$

$$H_1 : \text{parameter} < \text{null value}$$

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

$$H_0 : \text{parameter} = \text{null value}$$

$$H_1 : \text{parameter} > \text{null value}$$

Left- and right-tailed tests are referred to as **one-tailed tests**. Only our alternative hypothesis changes; the null hypothesis remains the same in all three tests.

Example 2. Determine the null and alternative hypotheses and the tailed-ness.

- (a) According to a Gallup poll conducted in 1995, 74% of Americans felt that men were more aggressive than women. A researcher wonders if the percentage of Americans that feel men are more aggressive than women is different today.
  
- (b) The packaging on a light bulb states that the bulb will last 500 hours under normal use. A consumer advocate would like to know if the mean lifetime of a bulb is less than 500 hours.
  
- (c) The standard deviation of the rate of return for a certain class of mutual funds is 0.08. A mutual fund manager believes the standard deviation of the rate of return for his fund is less than 0.08.

Example 3. Determine the null and alternative hypotheses and the tailed-ness.

- (a) The Medco pharmaceutical company has just developed a new antibiotic for children. Among the competing antibiotics, 2% of children who take the drug experience headaches as a side effect. A researcher for the Food and Drug Administration wishes to know if the percentage of children taking the new antibiotic who experience headaches as a side effect is more than 2%.
  
- (b) The *Blue Book* value of a used 3-year-old Chevy Corvette is \$37,500. Grant wonders if the mean price of a used 3-year-old Chevy Corvette in the Miami metropolitan area is different from \$37,500.
  
- (c) The standard deviation of the content of a 64-ounce bottle of detergent using an old filling machine was known to be 0.23 ounces. The company purchased a new filling machine and wants to know if there is less variability with the new filling machine.

## 9.2: Outcomes and the Type I and Type II Errors

Everytime we make a decision (either reject  $H_0$  or fail to reject  $H_0$ ), we either made the correct decision or we made an error.

### Four Outcomes from Hypothesis Testing

1. We reject the null hypothesis when the alternative hypothesis is true. This decision would be correct; also called **Power**.
2. We do not reject the null hypothesis when the null hypothesis is true. This decision would be correct.
3. We reject the null hypothesis when the null hypothesis is true. This decision would be incorrect - **Type I error**.
4. We do not reject the null hypothesis when the alternative hypothesis is true. This decision would be incorrect - **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

The numbering of the errors indicates which of the two hypotheses, (1) null or (2) alternative, is actually true. For example, a Type I error is the error that occurs when the first hypothesis, the null, is really true, but we decide in favor of the alternative.

$$\alpha = P(\text{Type I Error}) = P(\text{Rejecting } H_0 \text{ when } H_0 \text{ is true})$$

$$\beta = P(\text{Type II Error}) = P(\text{Failing to Reject } H_0 \text{ when } H_0 \text{ is false})$$

**Power** = probability we correctly reject the null hypothesis, which occurs with probability  $1 - \beta$ .

The logic of statistical hypothesis testing is similar to the “presumed innocent until proven guilty” principle of the U.S. judicial system.

- We assume the null hypothesis is a possible truth until the sample data conclusively demonstrate otherwise
- The alternative hypothesis is chosen only when the data shows us that we can reject the null hypothesis

Think of a criminal trial. In any trial, the defendant is assumed to be innocent. (We give the defendant the benefit of the doubt). The district attorney must collect and present evidence proving that the defendant is guilty beyond all reasonable doubt.

Because we are seeking evidence for guilt, it becomes the alternative hypothesis. Innocence is assumed, so it is the null hypothesis. The hypotheses for a trial are written

$H_0$  : the defendant is innocent

$H_1$  : the defendant is guilty

Using this court analogy, the two correct decisions are to:

- conclude that an innocent person is not guilty or
- conclude that a guilty person is guilty.

The two incorrect decisions are to:

- convict an innocent person (a Type I error) or
- to let a guilty person go free (a Type II error).

Example 4: Suppose you have the following hypothesis:

$H_0$ : Frank's rock climbing equipment is safe.

$H_1$ : Frank's rock climbing equipment is not safe.

1. What is the Type I error?

2. What is the Type II error?

Example 5: Suppose you have the following hypothesis:

$H_0$ : The blood cultures contain no traces of pathogen X.

$H_1$ :

1. What is the Type I error?

2. What is the Type II error?

## Power of a Hypothesis Test

Rather than focusing on the risk of making a mistake, many investigators prefer to focus on the chance that their sample will provide the evidence necessary to make the right choice.

The **power** of a hypothesis test is the probability  $(1 - \beta)$  of rejecting a false null hypothesis. That is, that we decide in favor of the alternative hypothesis given a specific truth about the population. When the alternative is actually true, power is the probability that we do not make a Type II error.

There are two features of power that apply to all hypothesis tests and that researchers should keep in mind when they plan a study:

- *The power increases when the sample size is increased.* The sample statistic is a more accurate estimate of the population value, making it easier to detect a difference between the true population value and the null value (Law of Large Numbers).
- *The power increases when the difference between the true population value and the null hypothesis value increases.* This makes sense because the probability of detecting a large difference is higher than the probability of detecting a small difference.

For a specified true value and level of significance, we can either calculate the power for a given sample size or we can compute the sample size required to achieve a desired power.

Example 6. A University is considering a plan to offer regular classes year-round by making the summer session a regular term but will do so only if there is sufficient student interest. A university administrator is planning a student survey to determine whether a majority of all students at the school would attend summer session under that structure. The plan will be implemented only if a majority would attend, so the status quo is to assume that the proportion who would attend is not a majority. With  $p$  representing the proportion of all students who would attend in the summer, a hypothesis testing structure is as follows:

$$\begin{aligned} H_0 : p \leq 0.50 & \quad (\text{the proportion who would attend is not a majority}) \\ H_a : p > 0.50 & \quad (\text{a majority would attend}) \end{aligned}$$

The table below shows the *power* for three different “true” proportions and for three different sample sizes, for a test with a significance level = 0.05. As you look at the table, keep in mind that the power is the probability the sample evidence will lead us to conclude that a majority of the student population would attend a regular summer term.

		True Population Proportion		
		0.52	0.60	0.65
Sample Size	$n = 50$	0.09	0.41	0.69
	$n = 100$	0.11	0.64	0.92
	$n = 400$	0.20	0.99	Nearly 1

- The power increases when the sample size is increased
- The power increases when the difference between the true population value and the null hypothesis value increases.

Researchers should evaluate power before they collect data that will be used to do hypothesis tests to make sure they have sufficient power to make the study worth while.

Example 7. Suppose that a test for ESP has four choices, and that the probability of a correct guess by chance on each trial is 0.25. A researcher believes that the true probability of a correct guess is 0.33. The following output shows the power of the one-sided test for this situation for three possible sample sizes:

Table 1: Test for One Proportion

Sample Size	Power
50	0.3776
100	0.5740
400	0.9705

Testing proportion = 0.25 (versus  $> 0.25$ )  
 Calculating power for proportion = 0.33

- (a) What is the power of the test if  $n = 50$  trials are used?
- (b) Write a sentence providing the power of the test for  $n = 100$  and explain its meaning.
- (c) If the researcher wants to have at least a 0.95 probability of detecting ESP in the study, and is correct that the true probability of a success is 0.33, would a sample of size 400 be sufficient? Explain.
- (d) If the true probability of success is actually 0.40 on each trial, would the power for each sample size be higher or lower than that shown in the output? Explain.



Example 8. Determine the null and alternative hypotheses; explain what it would mean to make a Type I error; and explain what it would mean to make a Type II error.

(a) According to *Giving and Volunteering in the United States, 2001 Edition*, the mean charitable contribution per household in the United States in 2000 was \$1,623. A researcher believes that the level of giving has changed since then.

(b) According to the Centers for Disease Control and Prevention, 16% of children aged 6 to 11 years are overweight. A school nurse thinks that the percentage of 6- to 11-year-olds who are overweight is higher in her school district.

Example 9. According to the Centers for Disease Control and Prevention, in 2005, 15.2% of tenth-grade students had tried marijuana. The Drug Abuse and Resistance Education (DARE) program underwent several major changes to keep up with technology and issues facing students in the 21st century. After the changes, a school resource officer (SRO) thinks that the proportion of tenth-grade students who have tried marijuana has decreased from the 2005 level.

(a) Determine the null and alternative hypotheses.

(b) Assume you fail to reject  $H_0$ , and suppose, in fact, that the proportion of tenth-grade students who have tried marijuana is 14.7%. Was a Type I or Type II error committed?

### 9.3: Distribution Needed for Hypothesis Testing

Earlier in the course, we discussed sampling distributions. Particular distributions are associated with hypothesis testing. Perform tests of a population mean using a **Student's t-distribution**. (Remember, use a Student's t-distribution when the population standard deviation is unknown and the distribution of the sample mean is approximately normal.) We perform tests of a population proportion using a normal distribution (usually  $n$  is large).

If you are testing a single population mean, the distribution for the test is for means:

$$\bar{X} \sim t_{df}$$

The population parameter is  $\mu$ . The estimated value (point estimate) for  $\mu$  is  $\bar{x}$ , the sample mean.

If you are testing a single population proportion, the distribution for the test is for proportions or percentages:

$$\hat{p} \sim N \left( p, \sqrt{\frac{p(1-p)}{n}} \right)$$

The population parameter is  $p$ . The estimated value (point estimate) for  $p$  is  $\hat{p}$ .  $\hat{p} = \frac{x}{n}$  where  $x$  is the number of successes and  $n$  is the sample size.

Step 2: Verify that all conditions are met

- Always check to see if the sampling method is sound
  
- For means
  - $\bar{X} \sim N$  if  $X \sim N$
  
  - $\bar{X} \sim N$  if the Central Limit Theorem holds
    - \* Check sample size,  $n \geq 30$
  
- For proportions
  - $X$  must follow a binomial distribution
  - $np$  and  $nq$  must both be at least 5

## 9.4: Rare Events, the Sample, Decision and Conclusion

### Step 3: Test Statistic

After determining the null and alternative hypotheses, the next step is to calculate the data summary called a *test statistic* that measures the difference between the sample result and the null value.

- Test statistic for a single proportion:  $z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$
- Test statistic for a single mean:  $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$

We, then, compute a probability, called a ***p-value***, which is the probability of getting a value of the test statistic that is at least as extreme as the test statistic obtained from the sample data, assuming the null hypothesis is true.

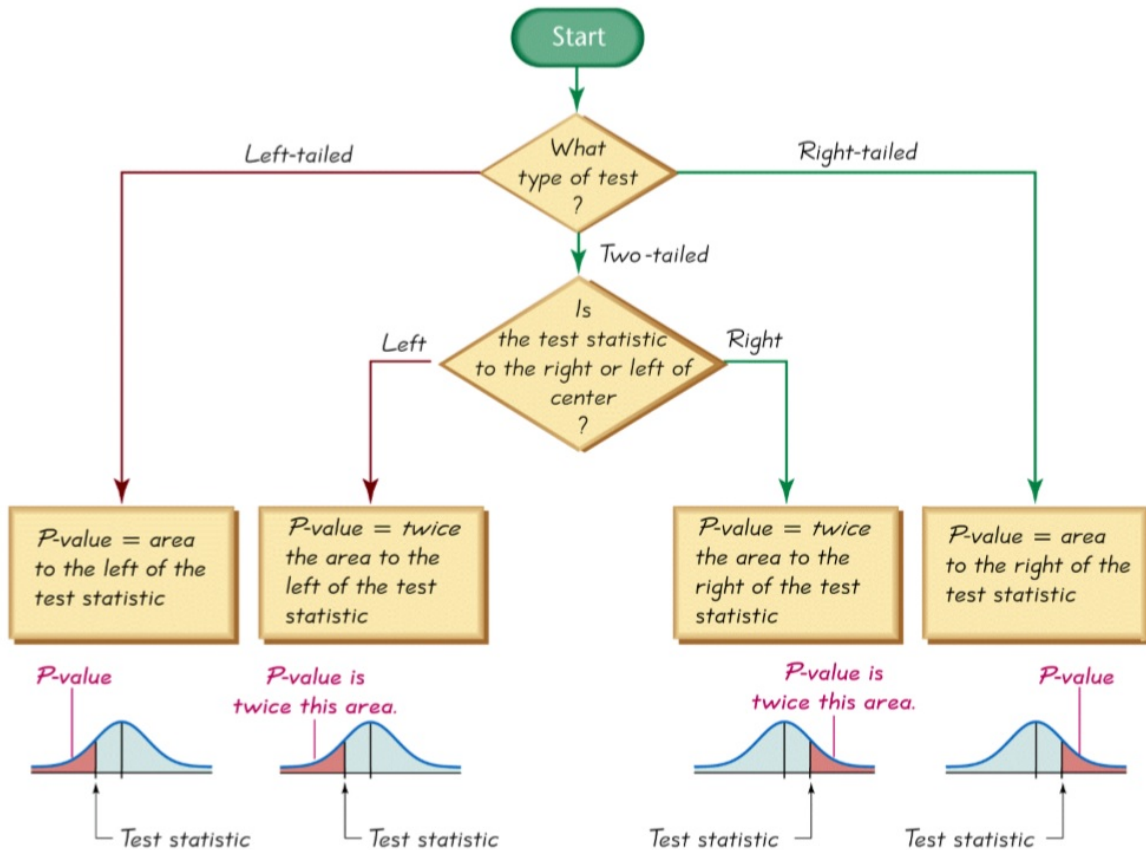
### Step 4: Computing the *p*-Value for the Test

In hypothesis testing, the objective is to decide if we should reject the null hypothesis in favor of the alternative. We do this by comparing the *p*-value to a designated standard called the *level of significance* for the test.

The details of how to find the *p*-value - the probability of a test statistic as extreme as or more extreme than the observed test statistic - depend on the direction specified in the alternative hypothesis:

- For a *less than* alternative hypothesis, find the probability that the test statistic *z* could have been *equal to or less than* what it is.
- For a *greater than* alternative hypothesis, find the probability that the test statistic *z* could have been *equal to or greater than* what it is.
- For a *two-tailed* alternative hypothesis, the *p*-value includes the probability areas in both extremes of the distribution of the test statistic *z*.

A visual is provided below.



Step 5: Make a decision based on the  $p$ -value

The **level of significance**,  $\alpha$  (alpha), is the value that is the borderline between when a  $p$ -value is small enough to choose the alternative hypothesis, and when it is not small enough.

- If  $p\text{-value} \leq \alpha$ , we reject the null hypothesis. (“If the  $P$  is low, the null must go!”)
- If  $p\text{-value} > \alpha$ , we fail to reject the null hypothesis.

The level of significance is chosen by the researcher BEFORE the experiment/study begins.

The phrase **statistically significant** is used to describe the results when the researcher has decided that the  $p$ -value is small enough to decide in favor of the alternative hypothesis.

#### Step 4: Computing the Critical Value for the Test

Before technology became so widespread, it was difficult or impossible to compute the  $p$ -value in many circumstances, so researchers used a method called the *rejection region approach*.

The **critical region**, or **rejection region**, in a hypothesis test is the region of possible values for the test statistic that would lead to rejection of the null hypothesis.

A boundary of a rejection region is called a **critical value**. The critical value is denoted as  $z^*$  or  $t^*$  (similar to what we found when computing confidence intervals), while the rejection region is the area more extreme than this value.

We compare the test statistic to the rejection region. If the test statistic falls in the rejection region, the null hypothesis is rejected.

#### Step 5: Make a decision based on the rejection region

Possible decisions:

- We reject the null hypothesis
- We fail to reject the null hypothesis

We never ‘ACCEPT’ the null hypothesis.

- If the test statistic is in the shaded region (rejection region), reject  $H_0$
- If the test statistic is NOT in the shaded region (rejection region), fail to reject  $H_0$

The  $p$ -value method and the rejection region method give equivalent conclusions.

Recall our criminal trial. The trial is the process whereby the jury obtains information (sample data). The jury then deliberates about the evidence (the data analysis). Finally, the jury either convicts the defendant (rejects the null hypothesis) or declares the defendant not guilty (fails to reject the null hypothesis).

Note: the defendant is never declared innocent. That is, we never conclude that the null hypothesis is true.

#### Step 6: State your conclusion in terms of the problem

- If you Reject  $H_0$ : There is sufficient evidence to conclude [statement in  $H_a$ ].
- If you Fail to Reject  $H_0$ : There is not sufficient evidence to conclude [statement in  $H_a$ ].

## 9.5: Additional Information and Full Hypothesis Test

Notes:

- Remember that the hypotheses are always statements about a parameter, not about a statistic. The whole point is to see what you can infer about an unknown parameter value based on a sample statistic.
- The  $p$ -value is NOT the probability that the null hypothesis is true. Rather, it is the probability of obtaining such an extreme sample result (or one even more extreme) if the null hypothesis were true.
- Remember we do NOT *accept* the null hypothesis, we *fail to reject* the null.

## 9.6a: Hypothesis Testing for a Single Proportion

### Testing Hypotheses Regarding a Single Population Proportion, $p$

1. Set up **hypothesis**:

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

2. Verify the following two **requirements** are satisfied:

- (a) The sample is a representative sample.
- (b)  $X$  must follow a binomial distribution
- (c)  $np_0$  and  $n(1 - p_0)$  should be at least 5.

3. Compute the **test statistic**

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

4. Find the  **$p$ -Value**. Use the  $z$ -table.

Refer to the previous pages for illustration.

4. OR determine the **critical value**,  $z^*$ .

Refer to the previous pages for illustration.

5. **Decision Rule**: If the  $p$ -value  $< \alpha$ , reject the null hypothesis.

5. OR **Decision Rule**: If the test statistic is MORE EXTREME than the critical value, reject the null hypothesis.

6. **Conclusion** - State your decision and your conclusion in terms of the problem.

If you Reject  $H_0$ : There is sufficient evidence to conclude [statement in  $H_a$ ].

If you Fail to Reject  $H_0$ : There is not sufficient evidence to conclude [statement in  $H_a$ ].

Example 10. There are two major college entrance exams that a majority of colleges accept for admission, the SAT and ACT. ACT looked at historical records and established 22 as the minimum score on the ACT math portion of the exam for a student to be considered prepared for college mathematics. (Note: “Being prepared” means there is a 75% probability of successfully completing College Algebra in college.) An official with the Illinois State Department of Education wonders whether less than half of the students in her state are prepared for College Algebra. She obtains a simple random sample of 500 records of students who have taken the ACT and finds that 219 are prepared for college mathematics (that is, 219 scored at least 22 on the math portion of the ACT). Does this represent significant evidence that less than half of the students in the state of Illinois are prepared for college mathematics upon graduation from high school? Use the  $\alpha = 0.05$  level of significance.

Step 1: Hyp

Step 2: Req

Step 3: TS

Step 4:  $p$ -val

Step 5: Dec

Step 6: Conc

Step 4: CV

Step 5: Dec



Example 11. The drug Prevnar is a vaccine meant to prevent meningitis. It is typically administered to infants. In clinical trials, the vaccine was administered to 710 randomly sampled infants between 12 and 15 months of age. Of the 710 infants, 121 experienced a loss of appetite. Is there significant evidence to conclude that the proportion of infants who receive Prevnar and experience a loss of appetite is different from 0.135, the proportion of children who experience a loss of appetite with competing medications? Use the  $\alpha = 0.01$  level of significance.

Step 1: Hyp

Step 2: Req

Step 3: TS

Step 4:  $p$ -val

Step 5: Dec

Step 6: Conc

Step 4: CV

Step 5: Dec

## 9.6b: Hypothesis Testing for a Single Mean

### Testing Hypotheses Regarding a Single Population Mean, using $t$

1. Set up **hypothesis**:

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

2. Verify the following **requirements** are satisfied:

- (a) Representative Sample
- (b) One of the following:
  - i. The population must be normally distributed, OR
  - ii. The sample size needs to be *large* enough,  $n \geq 30$ , OR
  - iii. Check the Q-Q plot and boxplot

3. Compute the **test statistic**

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

where  $df = n - 1$

4. Find the  **$p$ -Value**. Use the  $t$ -table.

Refer to the previous pages for illustration.

4. OR determine the **critical value**,  $t^*$ .

Refer to the previous pages for illustration.

5. **Decision Rule**: If the  $p$ -value  $< \alpha$ , reject the null hypothesis.

5. OR **Decision Rule**: If the test statistic is MORE EXTREME than the critical value, reject the null hypothesis.

6. **Conclusion** - State your decision and your conclusion in terms of the problem.

If you Reject  $H_0$ : There is sufficient evidence to conclude [statement in  $H_a$ ].

If you Fail to Reject  $H_0$ : There is not sufficient evidence to conclude [statement in  $H_a$ ].

Example 12. Find the  $t$ -values for a right-tailed hypothesis test if we have 15 degrees of freedom, and  $\alpha = 0.10$ . What if the alternative hypothesis is left-tailed? Two-tailed?

The critical value for  $z$  with an area to the right of 0.10 is approximately 1.28. Notice that the critical value for  $t$  is bigger than the corresponding critical value of  $z$  with an area to the right of 0.10. This is because the  $t$ -distribution has more spread than the  $z$ -distribution.

If the degrees of freedom we desire are not available in the  $t$ -table, we follow the practice of choosing the closest number of degrees of freedom available in the table. For example, if we have 43 degrees of freedom, we use 45 degrees of freedom from the  $t$ -table.

In addition, the last row of the  $t$ -table provides the  $z$ -values from the standard normal distribution. We use these values for situations where the degrees of freedom are more than 2,000. This is acceptable because the  $t$ -distribution starts to behave like the standard normal distribution as  $n$  increases.

Example 13. According to the Centers for Disease Control, the mean number of cigarettes smoked per day by individuals who are daily smokers is 18.1. Do retired adults who are daily smokers smoke less than the general population of daily smokers? To answer this question, we obtain a random sample of 40 retired adults who are current daily smokers and record the number of cigarettes smoked on a randomly selected date. The data result in a sample mean of 16.8 cigarettes and a standard deviation of 4.7 cigarettes. Is there sufficient evidence at the  $\alpha = 0.1$  level of significance to conclude that retired adults who are daily smokers smoke less than the general population of daily smokers?

Step 1: Hyp

Step 2: Req

Step 3: TS

Step 4:  $p$ -val

Step 5: Dec

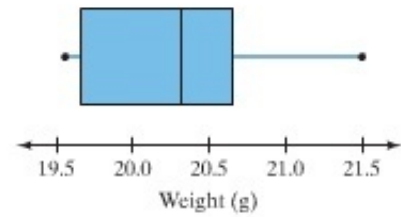
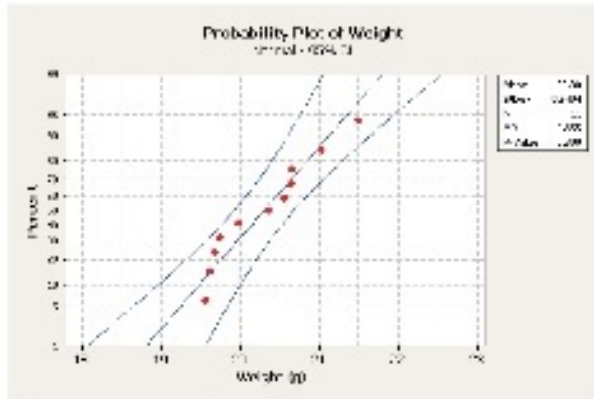
Step 6: Conc

Step 4: CV

Step 5: Dec

Example 14. The “fun size” of a Snickers bar is supposed to weigh 20 grams. Because the punishment for selling candy bars that weigh less than 20 grams is so severe, the manufacturer calibrates the machine so that the mean weight is 20.1 grams. The quality-control engineer at M&M-Mars, the company that manufactures Snickers bars, is concerned that the machine that manufactures the candy is miscalibrated. She obtains a random sample of 11 candy bars, weighs them, and obtains the data below. Should the machine be shut down and calibrated? Because shutting down the plant is very expensive, she decides to conduct a test at the  $\alpha = 0.01$  level of significance.

19.68  
 20.66  
 19.56  
 19.98  
 20.65  
 19.61  
 20.55  
 20.36  
 21.02  
 21.50  
 19.74



Step 1: Hyp

Step 2: Req

Step 3: TS

Step 4:  $p$ -val

Step 5: Dec

Step 6: Conc

Step 4: CV

Step 5: Dec

## Mixed Problems

Example 15. To test  $H_0 : \mu = 100$  versus  $H_1 : \mu \neq 100$ , a simple random sample of size  $n = 23$  is obtained from a population that is known to be normally distributed.

- (a) If  $\bar{x} = 104.8$  and  $s = 9.2$ , compute the test statistic.
  
  
  
  
  
  
  
  
  
  
- (b) If the researcher decides to test this hypothesis at the  $\alpha = 0.01$  level of significance, determine the critical value(s).
  
  
  
  
  
  
  
  
  
  
- (c) What is your decision? Why?
  
  
  
  
  
  
  
  
  
  
- (d) State your conclusions in terms of the problem.
  
  
  
  
  
  
  
  
  
  
- (e) Construct a 99% confidence interval to test the hypothesis.
  
  
  
  
  
  
  
  
  
  
- (f) Interpret the interval in part (f).

Example 16. *Time* magazine reported that in a 1994 survey of 507 randomly selected adult American Catholics, 59% answered yes to the question “Do you favor allowing women to be priests?” Test the hypothesis that more than 55% of adult American Catholics said yes. (*Time*, 26 December - 2 January 1995, pp. 74-76).

1. State your hypotheses.
2. What conditions are necessary for testing this hypothesis? Are those conditions met?
3. Compute the test statistic.
4. Calculate the p-value.
5. What is your decision, using 0.05 as your level of significance?
6. State your conclusion in terms of the problem.
7. Find the rejection region and make a decision.

Example 17. Carl Reinhold August Wunderlich said that the mean temperature of humans is  $98.6^{\circ}\text{F}$ . Researchers Philip Mackowiak, Steven Wasserman, and Myron Levine [*JAMA*, Sept. 23-30, 1992; 268(12): 1578-80] measured the temperatures of 26 females 1 to 4 times daily for 3 days to get a total of 123 measurements. The sample data yielded a sample mean of  $98.4^{\circ}\text{F}$  and a sample standard deviation of  $0.7^{\circ}\text{F}$ . Determine whether the normal temperature of women is less than  $98.6^{\circ}\text{F}$  at the  $\alpha = 0.01$  level of significance.

(a) State the hypotheses.

(b) Determine the test statistic.

(c) Determine the  $p$ -value.

(d) What is your decision? Why?

(e) State your conclusions in terms of the problem.