Chapter 8
Hypothesis Testing

8-1 Review and Preview
8-2 Basics of Hypothesis Testing
8-3 Testing a Claim about a Proportion
8-4 Testing a Claim About a Mean: σ Known
8-5 Testing a Claim About a Mean: σ Not Known
8-6 Testing a Claim About a Standard Deviation or Variance

S4.D2.Chapter 8 Hypothesis Testing
These notes cover the following topics: Basics of Hypothesis Testing; Testing Claim about Proportion; Testing Claim of Mean, σ known; Testing Claim of Mean, σ unknown; Testing Claim about Standard Deviation.

In Chapters 2 and 3 we used “descriptive statistics” when we summarized data using tools such as graphs, and statistics such as the mean and standard deviation. Methods of inferential statistics use sample data to make an inference or conclusion about a population. The two main activities of inferential statistics are using sample data to (1) estimate a population parameter (such as estimating a population parameter with a confidence interval), and (2) test a hypothesis or claim about a population parameter. In Chapter 7 we presented methods for estimating a population parameter with a confidence interval, and in this chapter we present the method of hypothesis testing.

Definitions
In statistics, a **hypothesis** is a claim or statement about a property of a population.

A **hypothesis test** (or test of significance) is a standard procedure for testing a claim about a property of a population.

Main Objective
The main objective of this chapter is to develop the ability to conduct hypothesis tests for claims made about a population proportion $p$, a population mean $\mu$, or a population standard deviation $\sigma$.

Examples of Hypotheses that can be Tested

- **Genetics**: The Genetics & IVF Institute claims that its XSORT method allows couples to increase the probability of having a baby girl.
- **Business**: A newspaper headline makes the claim that most workers get their jobs through networking.
- **Medicine**: Medical researchers claim that when people with colds are treated with echinacea, the treatment has no effect.
- **Aircraft Safety**: The Federal Aviation Administration claims that the mean weight of an airline passenger (including carry-on baggage) is greater than 185 lb, which it was 20 years ago.
- **Quality Control**: When new equipment is used to manufacture aircraft altimeters, the new altimeters are better because the variation in the errors is reduced so that the readings are more consistent. (In many industries, the quality of goods and services can often be improved by reducing variation.)
Caution

When conducting hypothesis tests as described in this chapter and the following chapters, instead of jumping directly to procedures and calculations, be sure to consider the context of the data, the source of the data, and the sampling method used to obtain the sample data.

Key Concept

This section presents individual components of a hypothesis test. We should know and understand the following:

- How to identify the null hypothesis and alternative hypothesis from a given claim, and how to express both in symbolic form
- How to calculate the value of the test statistic, given a claim and sample data
- How to identify the critical value(s), given a significance level
- How to identify the P-value, given a value of the test statistic
- How to state the conclusion about a claim in simple and nontechnical terms

Part 1: The Basics of Hypothesis Testing

Rare Event Rule for Inferential Statistics

If, under a given assumption, the probability of a particular observed event is exceptionally small, we conclude that the assumption is probably not correct.

Null Hypothesis: \( H_0 \)

- The null hypothesis (denoted by \( H_0 \)) is a statement that the value of a population parameter (such as proportion, mean, or standard deviation) is equal to some claimed value.
- We test the null hypothesis directly.
- Either reject \( H_0 \) or fail to reject \( H_0 \).

Alternative Hypothesis: \( H_1 \)

- The alternative hypothesis (denoted by \( H_1 \) or \( H_a \) or \( H_A \)) is the statement that the parameter has a value that somehow differs from the null hypothesis.
- The symbolic form of the alternative hypothesis must use one of these symbols: \( \neq, <, > \).
Note about Forming Your Own Claims (Hypotheses)

If you are conducting a study and want to use a hypothesis test to support your claim, the claim must be worded so that it becomes the alternative hypothesis.

Example:
Consider the claim that the mean weight of airline passengers (including carry-on baggage) is at most 195 lb (the current value used by the Federal Aviation Administration). Follow the three-step procedure outlined in Figure 8-2 to identify the null hypothesis and the alternative hypothesis.

Note about Identifying $H_0$ and $H_1$

Figure 8-2

Step 1: Express the given claim in symbolic form. The claim that the mean is at most 195 lb is expressed in symbolic form as $\mu \leq 195$ lb.

Step 2: If $\mu \leq 195$ lb is false, then $\mu > 195$ lb must be true.

Step 3: Of the two symbolic expressions $\mu \leq 195$ lb and $\mu > 195$ lb, we see that $\mu > 195$ lb does not contain equality, so we let the alternative hypothesis $H_1$ be $\mu > 195$ lb. Also, the null hypothesis must be a statement that the mean equals 195 lb, so we let $H_0$ be $\mu = 195$ lb.

Note that the original claim that the mean is at most 195 lb is neither the alternative hypothesis nor the null hypothesis. (However, we would be able to address the original claim upon completion of a hypothesis test.)
Test Statistic

The test statistic is a value used in making a decision about the null hypothesis, and is found by converting the sample statistic to a score with the assumption that the null hypothesis is true.

Example:
Let’s again consider the claim that the XSORT method of gender selection increases the likelihood of having a baby girl. Preliminary results from a test of the XSORT method of gender selection involved 14 couples who gave birth to 13 girls and 1 boy. Use the given claim and the preliminary results to calculate the value of the test statistic. Use the format of the test statistic given above, so that a normal distribution is used to approximate a binomial distribution. (There are other exact methods that do not use the normal approximation.)

Test Statistic -

Test statistic for proportion
\[ z = \frac{\hat{p} - p}{\sqrt{pq/n}} \]

Test statistic for mean
\[ z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \quad \text{or} \quad t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \]

Test statistic for standard deviation
\[ \chi^2 = \frac{(n-1)s^2}{\sigma^2} \]

Example:
The claim that the XSORT method of gender selection increases the likelihood of having a baby girl results in the following null and alternative hypotheses: \( H_0: p = 0.5 \) and \( H_1: p > 0.5 \). We work under the assumption that the null hypothesis is true with \( p = 0.5 \). The sample proportion of 13 girls in 14 births results in \( \hat{p} = 13/14 = 0.929 \). Using \( p = 0.5 \), \( \hat{p} = 0.929 \), and \( n = 14 \), we find the value of the test statistic as follows:

\[ z = \frac{\hat{p} - p}{\sqrt{pq/n}} = \frac{0.929 - 0.5}{\sqrt{(0.5)(0.5)/14}} = 3.21 \]
Example:

We know from previous chapters that a z score of 3.21 is "unusual" (because it is greater than 2). It appears that in addition to being greater than 0.5, the sample proportion of 13/14 or 0.929 is significantly greater than 0.5.

Critical Region

The critical region (or rejection region) is the set of all values of the test statistic that cause us to reject the null hypothesis. For example, see the red-shaded region in the figure below.

Example:

The figure below shows that the sample proportion of 0.929 does fall within the range of values considered to be significant because they are so far above 0.5 that they are not likely to occur by chance (assuming that the population proportion is $p = 0.5$).

Significance Level

The significance level (denoted by $\alpha$) is the probability that the test statistic will fall in the critical region when the null hypothesis is actually true. This is the same $\alpha$ introduced in Section 7-2. Common choices for $\alpha$ are 0.05, 0.01, and 0.10.

Critical Value

A critical value is any value that separates the critical region (where we reject the null hypothesis) from the values of the test statistic that do not lead to rejection of the null hypothesis. The critical values depend on the nature of the null hypothesis, the sampling distribution that applies, and the significance level $\alpha$. See the previous figure where the critical value of $z = 1.645$ corresponds to a significance level of $\alpha = 0.05$.
**P-Value**

The *P-value* (or *p-value* or *probability value*) is the probability of getting a value of the test statistic that is **at least as extreme** as the one representing the sample data, assuming that the null hypothesis is true.

- **Critical region in the left tail:** \( P\text{-value} = \text{area to the left of the test statistic} \)
- **Critical region in the right tail:** \( P\text{-value} = \text{area to the right of the test statistic} \)
- **Critical region in the two tails:** \( P\text{-value} = \text{twice the area in the tail beyond the test statistic} \)

The null hypothesis is rejected if the *P*-value is very small, such as 0.05 or less.

Here is a memory tool useful for interpreting the *P*-value:

- If the *P* is low, the null must go.
- If the *P* is high, the null will fly.

**Procedure for Finding P-Values**

![Figure 8-5](image)

**Caution**

Don't confuse a *P*-value with a proportion *p*. Know this distinction:

\[
P\text{-value} = \text{probability of getting a test statistic at least as extreme as the one representing sample data}
\]

\[
p = \text{population proportion}
\]
Example

Consider the claim that with the XSORT method of gender selection, the likelihood of having a baby girl is different from $p = 0.5$, and use the test statistic $z = 3.21$ found from 13 girls in 14 births.

First determine whether the given conditions result in a critical region in the right tail, left tail, or two tails, then use Figure 8-5 to find the P-value. Interpret the P-value.

Example

The claim that the likelihood of having a baby girl is different from $p = 0.5$ can be expressed as $p \neq 0.5$ so the critical region is in two tails. Using Figure 8-5 to find the P-value for a two-tailed test, we see that the P-value is twice the area to the right of the test statistic $z = 3.21$. We refer to Table A-2 (or use technology) to find that the area to the right of $z = 3.21$ is 0.0007. In this case, the P-value is twice the area to the right of the test statistic, so we have:

$$P\text{-value} = 2 \times 0.0007 = 0.0014$$

Example

The P-value is 0.0014 (or 0.0013 if greater precision is used for the calculations). The small P-value of 0.0014 shows that there is a very small chance of getting the sample results that led to a test statistic of $z = 3.21$. This suggests that with the XSORT method of gender selection, the likelihood of having a baby girl is different from 0.5.

Types of Hypothesis Tests:

Two-tailed, Left-tailed, Right-tailed

The tails in a distribution are the extreme regions bounded by critical values.

Determinations of P-values and critical values are affected by whether a critical region is in two tails, the left tail, or the right tail. It therefore becomes important to correctly characterize a hypothesis test as two-tailed, left-tailed, or right-tailed.
### Two-tailed Test

- **$H_0$:** $\alpha$ is divided equally between the two tails of the critical region
- **$H_1$:** $\neq$

Means less than or greater than

#### Left-tailed Test

- **$H_0$:** $\leq$
- **$H_1$:** $<$

Points Left

#### Right-tailed Test

- **$H_0$:** $\geq$
- **$H_1$:** $>$

Points Right

### Conclusions in Hypothesis Testing

We always test the null hypothesis. The initial conclusion will always be one of the following:

1. Reject the null hypothesis.
2. Fail to reject the null hypothesis.
**Decision Criterion**

*P*-value method:  
Using the significance level $\alpha$:

If $P$-value $\leq \alpha$, reject $H_0$.  
If $P$-value $> \alpha$, fail to reject $H_0$.

---

**Decision Criterion**

Another option:  
Instead of using a significance level such as 0.05, simply identify the $P$-value and leave the decision to the reader.

---

**Decision Criterion**

Traditional method:  
If the test statistic falls within the critical region, reject $H_0$.  
If the test statistic does not fall within the critical region, fail to reject $H_0$.

---

**Decision Criterion**

Confidence Intervals:  
A confidence interval estimate of a population parameter contains the likely values of that parameter.  
If a confidence interval does not include a claimed value of a population parameter, reject that claim.
Wording of Final Conclusion

![Flowchart showing the process of hypothesis testing]

Caution

Never conclude a hypothesis test with a statement of “reject the null hypothesis” or “fail to reject the null hypothesis.”

Always make sense of the conclusion with a statement that uses simple nontechnical wording that addresses the original claim.

Accept Versus Fail to Reject

- Some texts use “accept the null hypothesis.”
- We are not proving the null hypothesis.
- Fail to reject says more correctly
- The available evidence is not strong enough to warrant rejection of the null hypothesis (such as not enough evidence to convict a suspect).

Type I Error

- A Type I error is the mistake of rejecting the null hypothesis when it is actually true.
- The symbol $\alpha$ (alpha) is used to represent the probability of a type I error.

$P(\text{type I error}) = \alpha$
Type II Error

- A Type II error is the mistake of failing to reject the null hypothesis when it is actually false.
- The symbol $\beta$ (beta) is used to represent the probability of a type II error.

\[ P(\text{type II error}) = \beta \]

Type I and Type II Errors

<table>
<thead>
<tr>
<th>True State of Nature</th>
<th>The null hypothesis is true</th>
<th>The null hypothesis is false</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision</td>
<td>We decide to reject the null hypothesis</td>
<td>Type I error (rejecting a true null hypothesis) $P(\text{Type I error}) = \alpha$</td>
</tr>
<tr>
<td>We fail to reject the null hypothesis</td>
<td>Correct decision</td>
<td>Type II error (failing to reject a false null hypothesis) $P(\text{Type II error}) = \beta$</td>
</tr>
</tbody>
</table>

Example:

Assume that we are conducting a hypothesis test of the claim that a method of gender selection increases the likelihood of a baby girl, so that the probability of a baby girl is $p > 0.5$.

Here are the null and alternative hypotheses:

- $H_0$: $p = 0.5$, and $H_1$: $p > 0.5$.

a) Identify a type I error.
b) Identify a type II error.

Example:

a) A type I error is the mistake of rejecting a true null hypothesis, so this is a type I error: Conclude that there is sufficient evidence to support $p > 0.5$, when in reality $p = 0.5$.

b) A type II error is the mistake of failing to reject the null hypothesis when it is false, so this is a type II error: Fail to reject $p = 0.5$ (and therefore fail to support $p > 0.5$) when in reality $p > 0.5$. 
Controlling Type I and Type II Errors

- For any fixed $\alpha$, an increase in the sample size $n$ will cause a decrease in $\beta$.
  $\alpha$ is fixed: $n \uparrow$ causes $\beta \downarrow$

- For any fixed sample size $n$, a decrease in $\alpha$ will cause an increase in $\beta$. Conversely, an increase in $\alpha$ will cause a decrease in $\beta$.
  $n$ is fixed: $\alpha \downarrow$ causes $\beta \uparrow$; $\alpha \uparrow$ causes $\beta \downarrow$

- To decrease both $\alpha$ and $\beta$, increase the sample size.
  $\text{Increase } n: \alpha \downarrow \text{ and } \beta \downarrow$

Comprehensive Hypothesis Test – P-Value Method

Comprehensive Hypothesis Test – Traditional Method

Comprehensive Hypothesis Test - cont

A confidence interval estimate of a population parameter contains the likely values of that parameter. We should therefore reject a claim that the population parameter has a value that is not included in the confidence interval.

<table>
<thead>
<tr>
<th>Table 5.2</th>
<th>Confidence Level for Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Two-Tailed Test</td>
</tr>
<tr>
<td>Significance</td>
<td>0.01</td>
</tr>
<tr>
<td>Level for</td>
<td>0.05</td>
</tr>
<tr>
<td>Hypothesis Test</td>
<td>0.10</td>
</tr>
</tbody>
</table>
Caution

In some cases, a conclusion based on a confidence interval may be different from a conclusion based on a hypothesis test. See the comments in the individual sections.

Recap

In this section we have discussed:
- Null and alternative hypotheses.
- Test statistics.
- Significance levels.
- P-values.
- Decision criteria.
- Type I and II errors.

Make a decision about the given claim. Use only the rare event rule stated in Section 8-2, and make subjective estimates to determine whether events are likely. For example, if the claim is that a coin favors heads and sample results consist of 11 heads in 20 flips, conclude that there is not sufficient evidence to support the claim that the coin favors heads (because it is easy to get 11 heads in 20 flips by chance with a fair coin).

415f. Claim: The proportion of households with telephones is greater than the proportion of 0.35 found in the year 1920. A recent simple random sample of 2480 households results in a proportion of 0.955 households with telephones (based on data from the U. S. Census Bureau).

\[ H_0: \, \hat{p} \leq 0.35 \quad H_1: \, \hat{p} > 0.35 \quad \hat{p} = 0.955 \]

Since 0.955 is much larger than 0.35, it seems that the claim is true.

415e. Claim: Movie patrons have IQ scores with a standard deviation that is less than the standard deviation of 15 for the general population. A simple random sample of 40 movie patrons results in IQ scores with a standard deviation of 14.8.

\[ H_0: \, \sigma \leq 15 \quad H_1: \, \sigma > 15 \quad \sigma = 14.8 \]

Sample: \( n = 40 \), \( s = 14.8 \)

\( \sigma < 15 \) claim

Sample: \( n = 40 \), \( s = 14.8 \)

\( \sigma < 15 \) claim

\( ? \) either claim is true or claim is not true
Identifying $H_0$ and $H_1$. In Exercises 9–16, examine the given statement, then express the null hypothesis and alternative hypothesis in symbolic form. Be sure to use the correct symbol ($\mu$, $p$, and $\sigma$) for the indicated parameter.

415/10. The proportion of people aged 18 to 25 who currently use illicit drugs is equal to 0.20 (or 20%).

$H_0: p = 0.20$ (claim)

$H_1: p \neq 0.20$

415/12. The majority of college students have credit cards.

$H_0: p \leq 0.5$

$H_1: p > 0.5$ (claim)

415/14. The standard deviation of daily rainfall amounts in San Francisco is 0.66 cm.

$H_0: \sigma = 0.66$ (claim)

$H_1: \sigma \neq 0.66$

415/16. The mean weight of plastic discarded by households in one week is less than 1 kg.

$H_0: \mu \geq 1$

$H_1: \mu < 1$ (claim)
Finding Critical Values. In Exercises 17–24, assume that the normal distribution applies and find the critical z values.

416/18. Two-tailed test; $\alpha = 0.10$.

416/20. Left-tailed test; $\alpha = 0.10$.

416/22. $\alpha = 0.005$; $H_1$ is $p > 0.05$.

416/24. $\alpha = 0.005$; $H_1$: $p \neq 0.45$ mm.
Finding Test Statistics. In Exercises 25–28, find the value of the test statistic $z$ using $z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$.

416/25. Carbon Monoxide Detectors The claim is that less than $\frac{1}{2}$ of adults in the United States have carbon monoxide detectors. A KRC Research survey of 1005 adults resulted in 462 who have carbon monoxide detectors.

$$H_0: p \geq \frac{1}{2}$$
$$H_1: p < \frac{1}{2}$$

Claim: $p < \frac{1}{2}$

$\hat{p} = \frac{462}{1005} = 0.4612$

Test statistic: $z = \frac{0.4612 - 0.5}{\sqrt{\frac{0.5(0.5)}{1005}}} = -2.56$

Finding P-values. In Exercises 29–36, use the given information to find the P-value. (Hint: Follow the procedure summarized in Figure 8-5.) Also, use a 0.05 significance level and state the conclusion about the null hypothesis (reject the null hypothesis or fail to reject the null hypothesis).

416/30. The test statistic in a right-tailed test is $z = 2.50$.

Finding P-values. In Exercises 29–36, use the given information to find the P-value. (Hint: Follow the procedure summarized in Figure 8-5.) Also, use a 0.05 significance level and state the conclusion about the null hypothesis (reject the null hypothesis or fail to reject the null hypothesis).

416/32. The test statistic in a two-tailed test is $z = -0.55$.

You may get instructions how to find the P-value using the TI calculator at http://cfcc.edu/faculty/cmoore/155ClassNotesFa11/TI_P_Value_zTest.pdf
Finding P-values. In Exercises 29–36, use the given information to find the P-value. (Hint: Follow the procedure summarized in Figure 8-5.) Also, use a 0.05 significance level and state the conclusion about the null hypothesis (reject the null hypothesis or fail to reject the null hypothesis).

416/34 With \( H_0: p = 0.74 \), the test statistic is \( z = 0.35 \).

You may get instructions how to find the P-value using the TI calculator at http://cfcc.edu/faculty/cmoore/155ClassNotesFa11/TI_P_Value_zTest.pdf

Finding P-values. In Exercises 29–36, use the given information to find the P-value. (Hint: Follow the procedure summarized in Figure 8-5.) Also, use a 0.05 significance level and state the conclusion about the null hypothesis (reject the null hypothesis or fail to reject the null hypothesis).

416/36 With \( H_1: p < 0.777 \), the test statistic is \( z = -2.95 \).

You may get instructions how to find the P-value using the TI calculator at http://cfcc.edu/faculty/cmoore/155ClassNotesFa11/TI_P_Value_zTest.pdf

Stating Conclusions. In Exercises 37–40, state the final conclusion in simple non-technical terms. Be sure to address the original claim. (Hint: See Figure 8-7.)

417/38 Original claim: The percentage of on-time U.S. airline flights is less than 75%. Initial conclusion: Reject the null hypothesis.

Stating Conclusions. In Exercises 37–40, state the final conclusion in simple non-technical terms. Be sure to address the original claim. (Hint: See Figure 8-7.)

417/40 Original claim: The percentage of Americans who believe in heaven is equal to 90%. Initial conclusion: Reject the null hypothesis.
Identifying Type I and Type II Errors. In Exercises 41–44, identify the type I error and the type II error that correspond to the given hypothesis.

41. The percentage of Americans who believe that life exists only on earth is equal to 20%.

42. The percentage of households with at least two cell phones is less than 60%.