Chapter 8
Hypothesis Testing

8-1 Overview
8-2 Basics of Hypothesis Testing (TODAY)
8-3 Testing a Claim about a Proportion
8-4 Testing a Claim About a Mean: \( \sigma \) Known
8-5 Testing a Claim About a Mean: \( \sigma \) Not Known
8-6 Testing a Claim About a Standard Deviation or Variance

Key Concept
This section presents individual components of a hypothesis test, and the following sections use those components in comprehensive procedures. The role of the following should be understood:

- null hypothesis
- alternative hypothesis
- test statistic
- critical region
- significance level
- critical value
- \( P \)-value
- Type I and II error

Section 8-2 Objectives

- Given a claim, identify the null hypothesis and the alternative hypothesis, and express them both in symbolic form.
- Given a claim and sample data, calculate the value of the test statistic.
- Given a significance level, identify the critical value(s).
- Given a value of the test statistic, identify the \( P \)-value.
- State the conclusion of a hypothesis test in simple, non-technical terms.

Example: Let’s again refer to the Gender Choice product that was once distributed by ProCare Industries. ProCare Industries claimed that couples using the pink packages of Gender Choice would have girls at a rate that is greater than 50% or 0.5. Let’s again consider an experiment whereby 100 couples use Gender Choice in an attempt to have a baby girl; let’s assume that the 100 babies include exactly 52 girls, and let’s formalize some of the analysis.

Under normal circumstances the proportion of girls is 0.5, so a claim that Gender Choice is effective can be expressed as \( p > 0.5 \).

Using a normal distribution as an approximation to the binomial distribution, we find \( P(52 \text{ or more girls in 100 births}) = 0.3821 \).

Example: Let’s again refer to the Gender Choice product that was once distributed by ProCare Industries. ProCare Industries claimed that couples using the pink packages of Gender Choice would have girls at a rate that is greater than 50% or 0.5. Let’s again consider an experiment whereby 100 couples use Gender Choice in an attempt to have a baby girl; let’s assume that the 100 babies include exactly 52 girls, and let’s formalize some of the analysis.

Figure 8-1, following, shows that with a probability of 0.5, the outcome of 52 girls in 100 births is not unusual.

We do not reject random chance as a reasonable explanation. We conclude that the proportion of girls born to couples using Gender Choice is not significantly greater than the number that we would expect by random chance.
Observations

- Claim: For couples using Gender Choice, the proportion of girls is \( p > 0.5 \).
- Working assumption: The proportion of girls is \( p = 0.5 \) (with no effect from Gender Choice).
- The sample resulted in 52 girls among 100 births, so the sample proportion is \( \hat{p} = 52/100 = 0.52 \).
- Assuming that \( p = 0.5 \), we use a normal distribution as an approximation to the binomial distribution to find that \( P(\text{at least 52 girls in 100 births}) = 0.3821 \).
- There are two possible explanations for the result of 52 girls in 100 births: Either a random chance event (with probability 0.3821) has occurred, or the proportion of girls born to couples using Gender Choice is greater than 0.5.
- There isn’t sufficient evidence to support Gender Choice’s claim.

Components of a Formal Hypothesis Test

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.

Note about Identifying \( H_0 \) and \( H_1 \)

Figure 8-2
- Identify the specific claim or hypothesis to be tested and express it in symbolic form.
- Give the symbolic form that must be true when the original claim is false.
- Of the two symbolic expressions obtained so far, let the alternative hypothesis \( H_1 \) be the one not containing equality, so that \( H_0 \) uses the symbol \( \neq \) or \( < \) or \( > \). Let the null hypothesis \( H_0 \) be the symbolic expression that the parameter equals the fixed value being considered.
Example: Identify the Null and Alternative Hypothesis. Refer to Figure 8-2 and use the given claims to express the corresponding null and alternative hypotheses in symbolic form.

a) The proportion of drivers who admit to running red lights is greater than 0.5.

b) The mean height of professional basketball players is at most 7 ft.

c) The standard deviation of IQ scores of actors is equal to 15.

Example: Identify the Null and Alternative Hypothesis. Refer to Figure 8-2 and use the given claims to express the corresponding null and alternative hypotheses in symbolic form.

a) The proportion of drivers who admit to running red lights is greater than 0.5. In Step 1 of Figure 8-2, we express the given claim as \( p > 0.5 \). In Step 2, we see that if \( p > 0.5 \) is false, then \( p \leq 0.5 \) must be true. In Step 3, we see that the expression \( p > 0.5 \) does not contain equality, so we let the alternative hypothesis \( H_1 \) be \( p > 0.5 \), and we let \( H_0 \) be \( p = 0.5 \).

b) The mean height of professional basketball players is at most 7 ft. In Step 1 of Figure 8-2, we express "a mean of at most 7 ft" in symbols as \( \mu \leq 7 \). In Step 2, we see that if \( \mu \leq 7 \) is false, then \( \mu > 7 \) must be true. In Step 3, we see that the expression \( \mu > 7 \) does not contain equality, so we let the alternative hypothesis \( H_1 \) be \( \mu > 0.5 \), and we let \( H_0 \) be \( \mu = 7 \).

c) The standard deviation of IQ scores of actors is equal to 15. In Step 1 of Figure 8-2, we express the given claim as \( \sigma = 15 \). In Step 2, we see that if \( \sigma = 15 \) is false, then \( \sigma \neq 15 \) must be true. In Step 3, we let the alternative hypothesis \( H_1 \) be \( \sigma \neq 15 \), and we let \( H_0 \) be \( \sigma = 15 \).

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.

\[
\begin{align*}
z &= \frac{\hat{p} - p}{\sqrt{\frac{pq}{ n}}} & \text{Test statistic for proportions} \\
z &= \frac{\bar{x} - \mu_x}{\frac{\sigma}{\sqrt{n}}} & \text{Test statistic for mean} \\
\chi^2 &= \frac{(n-1)s^2}{\sigma^2} & \text{Test statistic for standard deviation}
\end{align*}
\]
Example: A survey of $n = 880$ randomly selected adult drivers showed that 56% (or $p = 0.56$) of those respondents admitted to running red lights. Find the value of the test statistic for the claim that the majority of all adult drivers admit to running red lights. (In Section 8-3 we will see that there are assumptions that must be verified. For this example, assume that the required assumptions are satisfied and focus on finding the indicated test statistic.)

Solution: The preceding example showed that the given claim results in the following null and alternative hypotheses: $H_0: p = 0.5$ and $H_1: p > 0.5$. Because we work under the assumption that the null hypothesis is true with $p = 0.5$, we get the following test statistic:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0.56 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{880}}} = 3.56$$

Interpretation: We know from previous chapters that a $z$ score of 3.56 is exceptionally large. It appears that in addition to being “more than half,” the sample result of 56% is significantly more than 50%. See figure following.

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 previous figure.

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 \( \alpha = 0.05 \).

Two-tailed, Right-tailed, Left-tailed Tests

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

Two-tailed Test

- \( H_0: = \) (null hypothesis)
- \( H_1: \neq \) (alternative hypothesis)

\( \alpha \) is divided equally between the two tails of the critical region.

Means less than or greater than

Right-tailed Test

- \( H_0: = \) (null hypothesis)
- \( H_1: > \) (alternative hypothesis)

Points Right

Left-tailed Test

- \( H_0: = \) (null hypothesis)
- \( H_1: < \) (alternative hypothesis)

Points Left

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

Traditional method:

Reject $H_0$ if the test statistic falls within the critical region.

Fail to reject $H_0$ if the test statistic does not fall within the critical region.

Decision Criterion - cont

$P$-value method:

Reject $H_0$ if the $P$-value $\leq \alpha$ (where $\alpha$ is the significance level, such as 0.05).

Fail to reject $H_0$ if the $P$-value $> \alpha$.

Decision Criterion - cont

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 - cont

Confidence Intervals:

Because a confidence interval estimate of a population parameter contains the likely values of that parameter, reject a claim that the population parameter has a value that is not included in the confidence interval.

Procedure for Finding $P$-Values

Figure 8-6
Example: Finding $P$-values. First determine whether the given conditions result in a right-tailed test, a left-tailed test, or a two-tailed test, then find the $P$-values and state a conclusion about the null hypothesis.

a) A significance level of $\alpha = 0.05$ is used in testing the claim that $p > 0.25$, and the sample data result in a test statistic of $z = 1.18$.
b) A significance level of $\alpha = 0.05$ is used in testing the claim that $p \neq 0.25$, and the sample data result in a test statistic of $z = 2.34$.

Example: Finding $P$-values. First determine whether the given conditions result in a right-tailed test, a left-tailed test, or a two-tailed test, then find the $P$-values and state a conclusion about the null hypothesis.

a) With a claim of $p > 0.25$, the test is right-tailed. Because the test is right-tailed, Figure 8-6 shows that the $P$-value is the area to the right of the test statistic $z = 1.18$. We refer to Table A-2 and find that the area to the right of $z = 1.18$ is 0.1190. The $P$-value of 0.1190 is greater than the significance level $\alpha = 0.05$, so we fail to reject the null hypothesis. The $P$-value of 0.1190 is relatively large, indicating that the sample results could easily occur by chance.
b) With a claim of $p \neq 0.25$, the test is two-tailed. Because the test is two-tailed, and because the test statistic of $z = 2.34$ is to the right of the center, Figure 8-6 shows that the $P$-value is twice the area to the right of $z = 2.34$. We refer to Table A-2 and find that the area to the right of $z = 2.34$ is 0.0096, so $P$-value $= 2 \times 0.0096 = 0.0192$. The $P$-value of 0.0192 is less than or equal to the significance level, so we reject the null hypothesis. The small $P$-value of 0.0192 shows that the sample results are not likely to occur by chance.

Accept Versus Fail to Reject

- Some texts use “accept the null hypothesis.”
- We are not proving the null hypothesis.
- The sample evidence is not strong enough to warrant rejection (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 true.
- The symbol $\alpha$ (alpha) is used to represent the probability of a type I error.
Type II Error

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

Example: Assume that we are conducting a hypothesis test of the claim $p > 0.5$. 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: Assume that we are conducting a hypothesis test of the claim $p > 0.5$. Here are the null and alternative hypotheses: $H_0: p = 0.5$, and $H_1: p > 0.5$.

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 $p > 0.5$.

Type I and Type II Errors

<table>
<thead>
<tr>
<th>Decision</th>
<th>True State of Nature</th>
<th>Type I Error</th>
<th>Type II Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>We decide to reject the null hypothesis (Type I error)</td>
<td>The null hypothesis is true</td>
<td>Correct decision</td>
<td>Type II error</td>
</tr>
<tr>
<td>We fail to reject the null hypothesis (Type II error)</td>
<td>The null hypothesis is false</td>
<td>Incorrect decision</td>
<td>Correct decision</td>
</tr>
</tbody>
</table>

Controlling Type I and Type II Errors

- For any fixed $\alpha$, an increase in the sample size $n$ will cause a decrease in $\beta$.
- 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 $\beta$.
- To decrease both $\alpha$ and $\beta$, increase the sample size.
Definition

The **power of a hypothesis test** is the probability \((1 - \beta)\) of rejecting a false null hypothesis, which is computed by using a particular significance level \(\alpha\) and a particular value of the population parameter that is an alternative to the value assumed true in the null hypothesis. That is, the power of the hypothesis test is the probability of supporting an alternative hypothesis that is true.

Comprehensive Hypothesis Test – Traditional Method

1. Identify the specific research hypothesis to be tested and express it as a null hypothesis, \(H_0\), and an alternative hypothesis, \(H_a\).
2. Select the significance level \(\alpha\) to be used in the test.
3. Select the appropriate statistical test based on the form of the research hypothesis and the type of data involved.
4. Specify the rejection region based on the significance level \(\alpha\) and the type of test.
5. Collect the sample data and compute the test statistic.
6. Compare the test statistic to the critical value for the rejection region.
7. Make the decision to either reject or fail to reject the null hypothesis.
8. Interpret the results in the context of the original research question.

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 2: Confidence Level for Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Significance</td>
</tr>
<tr>
<td>0.05</td>
</tr>
<tr>
<td>Two-Tailed Test</td>
</tr>
<tr>
<td>99%</td>
</tr>
<tr>
<td>95%</td>
</tr>
<tr>
<td>90%</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.