8-2 Basics of Hypothesis Testing

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 choose the sampling distribution that is relevant
- How to identify the $P$-value or identify the critical value(s)
- How to state the conclusion about a claim in simple and nontechnical terms

Definitions

A **hypothesis** is a claim or statement about a property of a population.

A **hypothesis test** is a procedure for testing a claim about a property of a population.

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

- 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 in the sense that we assume it is true and reach a conclusion to either reject $H_0$ or fail to reject $H_0$.

Alternative Hypothesis

- The **alternative hypothesis** (denoted by $H_1$ 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.
Steps 1, 2, 3
Identifying $H_0$ and $H_1$

1. Identify the specific claim or hypothesis to be tested, and put it in symbolic form.

2. Give the symbolic form that must be true when the original claim is false.

3. 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 $\lt$ or $\gt$. Let the null hypothesis $H_0$ be the symbolic expression that the parameter equals the fixed value being considered.

Example

Assume that 100 babies are born to 100 couples treated with the XSORT method of gender selection that is claimed to make girls more likely.

We observe 58 girls in 100 babies. Write the hypotheses to test the claim that "with the XSORT method, the proportion of girls is greater than the 50% that occurs without any treatment".

\[ H_0 : p = 0.5 \]
\[ H_1 : p > 0.5 \]

Step 4
Select the Significance Level $\alpha$

Select the significance level $\alpha$ based on the seriousness of a type 1 error. Make $\alpha$ small if the consequences of rejecting a true $H_0$ are severe. The values of 0.05 and 0.01 are very common.

Common choices for $\alpha$ are 0.05, 0.01, and 0.10.

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 (making the mistake of rejecting the null hypothesis when it is true).

This is the same $\alpha$ introduced in Section 7-2.

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.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Sampling Distribution</th>
<th>Requirements</th>
<th>Test Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion $p$</td>
<td>Normal($\mu$, $\sigma$)</td>
<td>$np \geq 5$ and $n(1-p) \geq 5$</td>
<td>$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$</td>
</tr>
<tr>
<td>Mean $\mu$</td>
<td>Normal($\mu$, $\sigma$)</td>
<td>$\sigma$ known and normally distributed population or $\sigma$ not known and $n &gt; 30$</td>
<td>$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$</td>
</tr>
<tr>
<td>Mean $\mu$</td>
<td>Normal($\mu$, $\sigma$)</td>
<td>$\sigma$ known and normally distributed population or $\sigma$ not known and $n &gt; 30$</td>
<td>$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$</td>
</tr>
<tr>
<td>St. dev. or std. variance $\sigma$</td>
<td>, $\sigma^2$</td>
<td>Same requirements; normally distributed population</td>
<td>$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$</td>
</tr>
</tbody>
</table>
Step 6
Find the Value of the Test Statistic, Then Find Either the P-Value or the Critical Value(s)

First transform the relevant sample statistic to a standardized score called the test statistic. Then find the P-Value or the critical value(s).

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 100 couples who gave birth to 58 girls and 42 boys.

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.

Example - Continued

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:

\[ H_0 : p = 0.5 \]
\[ H_1 : p > 0.5 \]

We work under the assumption that the null hypothesis is true with \( p = 0.5 \).

The sample proportion of 58 girls in 10 births results in:

\[ \hat{p} = \frac{58}{100} = 0.58 \]

Example – Convert to the Test Statistic

\[ z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0.58 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{100}}} = 1.60 \]

We know from previous chapters that a z score of 1.60 is not “unusual”.

At first glance, 58 girls in 100 births does not seem to support the claim that the XSORT method increases the likelihood a having a girl (more than a 50% chance).

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

\[ H_0 : \neq \]
\[ H_1 : \neq \]

\( \alpha \) is divided equally between the two tails of the critical region.
**Left-tailed Test**

\[ H_0 : \quad \text{All } \alpha \text{ in the left tail} \]

\[ H_1 : \quad < \]

**Right-tailed Test**

\[ H_0 : \quad \text{All } \alpha \text{ in the right tail} \]

\[ H_1 : \quad > \]

---

**P-Value**

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

---

**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:

\[ H_0 : \quad p = 0.5 \]
\[ H_1 : \quad p > 0.5 \]

The test statistic was:

\[
\begin{align*}
    z &= \frac{\hat{p} - p}{\sqrt{pq/n}} \\
    &= \frac{0.58 - 0.5}{\sqrt{(0.5)(0.5)/100}} \\
    &= 1.60
\end{align*}
\]
Procedure for Finding $P$-Values

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 figures.

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 $\alpha$.

Example

For the XSORT birth hypothesis test, the critical value and critical region for an $\alpha = 0.05$ test are shown below:

Caution

Don’t confuse a $P$-value with a proportion $p$. Know this distinction:

$P$-value = probability of getting a test statistic at least as extreme as the one representing sample data

$p$ = population proportion

Step 7: Make a Decision: Reject $H_0$ or Fail to Reject $H_0$

The methodologies depend on if you are using the $P$-Value method or the critical value method.
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 \( H_0 \).

---

**Critical Value 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 \( H_0 \).

---

**Example**

For the XSORT baby gender test, the test had a test statistic of \( z = 1.60 \) and a \( P \)-Value of 0.0548. We tested:

\[
H_0 : p = 0.5 \\
H_1 : p > 0.5
\]

Using the \( P \)-Value method, we would fail to reject the null at the \( \alpha = 0.05 \) level.

Using the critical value method, we would fail to reject the null because the test statistic of \( z = 1.60 \) does not fall in the rejection region.

(You will come to the same decision using either method.)

---

**Step 8 : Restate the Decision Using Simple and Nontechnical Terms**

State a final conclusion that addresses the original claim with wording that can be understood by those without knowledge of statistical procedures.

**Example**

For the XSORT baby gender test, there was not sufficient evidence to support the claim that the XSORT method is effective in increasing the probability that a baby girl will be born.

---

**Wording of Final Conclusion**

- **Yes:** There is sufficient evidence to support the claim that \( p > 0.5 \).
- **No:** There is not sufficient evidence to support the claim that \( p > 0.5 \).
- **Yes:** The sample data supports the claim that \( p \) is greater than 0.5.
- **No:** The sample data does not support the claim that \( p \) is greater than 0.5.
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 that the available evidence is not strong enough to warrant rejection of the null hypothesis.

Type I Error

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

Type I and Type II Errors

<table>
<thead>
<tr>
<th>True State of Nature</th>
<th>Type I Error</th>
<th>Type II Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>The null hypothesis is true</td>
<td>$\alpha$</td>
<td>$\beta$</td>
</tr>
<tr>
<td>The null hypothesis is false</td>
<td>$\beta$</td>
<td>$\alpha$</td>
</tr>
</tbody>
</table>

Decision

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

Type I error (rejecting a true null hypothesis) $P(\text{Type I error}) = \alpha$

Type II error (failing to reject a false null hypothesis) $P(\text{Type II error}) = \beta$

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 girls is $p > 0.5$.

Here are the null and alternative hypotheses:

$H_0 : p = 0.5$

$H_1 : p > 0.5$

a) Identify a type I error.
b) Identify a type II error.
Example - Continued

a) A type I error is the mistake of rejecting a true null hypothesis:

We conclude the probability of having a girl is greater than 50%, when in reality, it is not. Our data misled us.

b) A type II error is the mistake of failing to reject the null hypothesis when it is false:

There is no evidence to conclude the probability of having a girl is greater than 50% (our data misled us), but in reality, the probability is greater than 50%.

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.

Part 2: Beyond the Basics of Hypothesis Testing: The Power of a Test

Definition

The power of a hypothesis test is the probability $1 - \beta$ of rejecting a false null hypothesis.

The value of the power 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.

Power and the Design of Experiments

Just as 0.05 is a common choice for a significance level, a power of at least 0.80 is a common requirement for determining that a hypothesis test is effective. (Some statisticians argue that the power should be higher, such as 0.85 or 0.90.)

When designing an experiment, we might consider how much of a difference between the claimed value of a parameter and its true value is an important amount of difference.

When designing an experiment, a goal of having a power value of at least 0.80 can often be used to determine the minimum required sample size.