

CEE 260/MIE 273: Probability and Statistics in Civil Engineering

Lecture M4d: Hypothesis Testing and p-values

Jimi Oke

UMassAmherst

College of Engineering

September 8, 2026

Outline

- ① Hypothesis testing
- ② Steps in hypothesis testing
- ③ p -values

Today's objectives

- Learn how to conduct a hypothesis test on the mean of a population
- Know when to use a lower-tailed, upper-tailed or two-tailed test
- Understand Type I errors and their relationship to p -values
- Learn how to use p -values to conduct a hypothesis test

Hypothesis testing

- Hypothesis testing provides a framework for evaluating parameter(s) of a population with respect to a desired or known outcome.
- Given that in most cases, we can only estimate these parameters, hypothesis testing allows us to determine if the estimate supports a **research hypothesis**.
- The results of this testing is useful for **decision-making**.

Formulating a hypothesis test

A hypothesis is a statement regarding a parameter.

In a test, there are usually two competing hypotheses:

- H_0 : the **null** hypothesis
- H_1 : the **alternative** hypothesis (H_A is also used to denote this)

The null hypothesis is usually framed as an equality, i.e.:

$$H_0 : p = p_0 \quad (1)$$

where p_0 is the specified standard.

The alternative is given by

$$H_1 : p \neq p_0 \quad (2)$$

Outcomes of a hypothesis test

The null hypothesis is presumed unless there is sufficient evidence to discard it. The alternative hypothesis, however, is what we hope to support.

In experimental design, we frame the null hypothesis in such a way as to reject it.

Thus there are **two outcomes** of a hypothesis test:

- **Reject H_0** : because of sufficient sample evidence in support of H_1
- **Fail to reject H_0** : because of insufficient evidence in support of H_1

No truth test for the null hypothesis

The failure to reject H_0 does not mean that H_0 is true.

Further explanation of hypothesis test outcomes

Example 1: Outcome of a trial

In a jury trial, the hypotheses are:

- H_0 : defendant is innocent
- H_1 : defendant is guilty (not innocent)

The null hypothesis H_0 is **rejected** if there evidence beyond reasonable doubt that the defendant is guilty.

However, **failure to reject** H_0 does not imply the defendant is innocent, only that there is **insufficient evidence to prove otherwise**.

Hypothesis testing in practice

Example 2: Chip manufacturing

A company manufacturing RAM chips claims the defective rate of the population is 5%. Using a 500-chip sample from production, formulate a hypothesis test to evaluate the validity of the company's claim.

Let p denote the *true* defective probability.

We structure our hypothesis test as follows:

$$H_0 : p = 0.05$$

$$H_1 : p > 0.05$$

Note: this is a one-sided hypothesis test (testing in one direction only)

Hypothesis testing in practice

Example 2: Chip manufacturing (cont.)

In order to test the hypotheses, we must choose a **test statistic**.

Here, we let X denote the number of defective chips in the sample of 1000.

Then in order to determine whether or not to reject H_0 , we must decide on a **critical value**.

We note that this is a Bernoulli process. Thus, if $p = 0.05$, then the expected number of defective chips is

$$\bar{X} = np = 1000 \times 0.05 = 50$$

Say the critical value were $p^* = .1$, then $p \geq .1$ could then be considered as strong evidence that $p > 0.05$.

Hypothesis testing in practice

Example 2: Chip manufacturing (cont.)

Thus, we would reject H_0 .

Errors in hypothesis testing

Since we are working with finite samples, errors are bound to occur in decision-making.

The decision matrix is:

	H_0 is true	H_1 is true
Fail to reject H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

Type I error

Definitions

- The incorrect rejection of H_0 is a Type I error.
- Also known as a *false positive*
- The probability of a Type I error is the **level of significance, α**

Examples of Type I error

- Convicting a defendant of a crime when they are innocent (Example 1)
- Diagnosing a patient with a disease when in fact they do not have it (i.e. the null hypothesis is that the disease is NOT present)

Level of significance

A Type I error is less likely as α reduces.

We revisit Example 2.

Example 2: Chip manufacturing (cont.)

Find the level of significance α when the critical value of p is .1.

In other words, what is the probability of incorrectly rejecting H_0 when it is true?

$$\alpha = P(\text{Type I error}) \quad (3)$$

$$= P(p > 0.1) \quad (4)$$

$$= 1 - \Phi\left(\frac{p^* - p_0}{SE_{p_0}}\right) = 1 - \Phi\left(\frac{.1 - .05}{\sqrt{.05(.95)/1000}}\right) \quad (5)$$

$$= \boxed{0} \quad (6)$$

Type II errors

Definitions

- Failure to reject H_0 when in fact H_1 is true is a Type II error.
- Also known as a *false negative*
- The probability of a Type II error is denoted β

Note

We cannot compute β except the alternative hypothesis H_1 is specified.
Much of our focus will be on dealing with the level of significance, α .

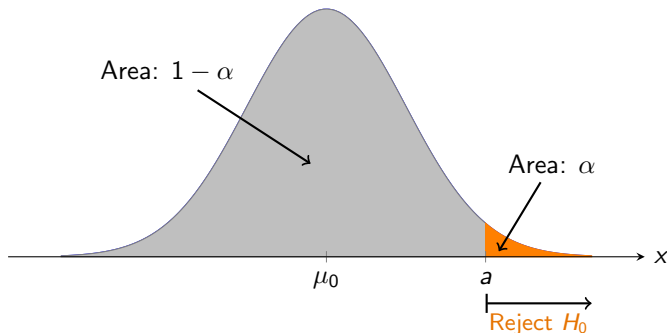
Summary of hypothesis testing approach

- 1 *Define* the **null** (H_0) and **alternative** (H_1) hypotheses
- 2 *Determine* the appropriate **test statistic** (and distribution)
- 3 *Estimate* the test statistic from the sample data
- 4 *Specify* or *identify* the **level of significance** (α)
- 5 *Define* the **region of rejection/critical region** of the null hypothesis by choosing the **critical value**.
- 6 *Decide*. If the test statistic is in the critical region, reject H_0 . If not, do not reject H_0 (fail to reject it)

One-sided tests

Case A: upper tail

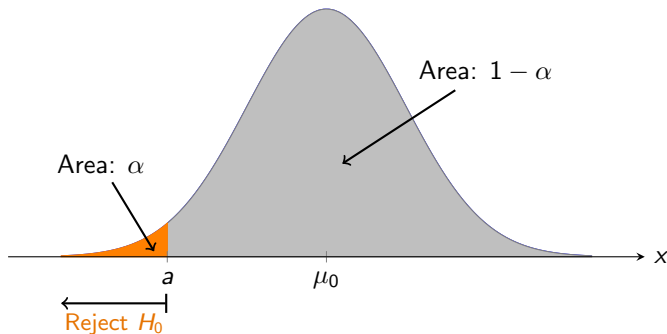
- $H_0 : \mu = \mu_0$
- $H_1 : \mu > \mu_0$



One-sided tests (cont.)

Case B: lower tail

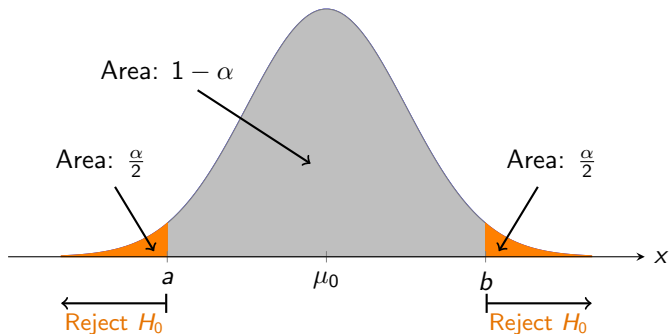
- $H_0 : \mu = \mu_0$
- $H_1 : \mu < \mu_0$



Two-sided tests

Case C: both tails

- $H_0 : \mu = \mu_0$
- $H_1 : \mu \neq \mu_0$



Distribution of the test statistic

In this lecture, the test statistic is the **sample proportion**.

We will assume the normal distribution is the success-failure condition holds.

The sample proportion is **normally** distributed and its variance is :

$$\mathbb{V}(p) = \frac{p(1-p)}{n} \quad (7)$$

And thus, the standard error is:

$$SE_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} \quad (8)$$

Thus, to compute the probability (area under curve) of the test statistic, we use the z-score:

$$z^* = \frac{p - p_0}{SE_p} \quad (9)$$

which is **normally** distributed.

What is a p -value?

Definition

The p -value is the smallest level of significance at which H_0 would be rejected when a specified test procedure is used on a given dataset.
Equivalently, this is the minimum probability of a Type I error.

Motivating the usage of p -values

Example 3: Nicotine content

Based on data from a sample of cigarettes, the Z statistic is $z = 2.10$. You want to verify if the true nicotine content (measured in proportion of tobacco weight) is $p = .015$ (H_0) versus the alternative hypothesis that is greater: $H_1 : p > .015$.

This is an **upper-tailed** hypothesis test.

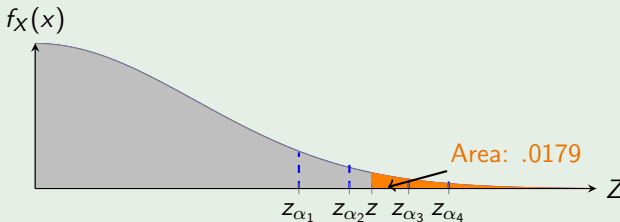
What are your conclusions from testing at the following significance levels:

- $\alpha_1 = 0.05$
- $\alpha_1 = 0.025$
- $\alpha_1 = 0.01$
- $\alpha_1 = 0.005$

Motivating the usage of p -values

Example 3: Nicotine content (cont.)

The p -value $1 - \Phi(2.10)$ (area to the right of z) $\therefore p = 1 - 0.9821 = 0.0179$.



Your conclusions are as follows:

Level of significance α	Rejection Region	Conclusion
$\alpha_1 = 0.05$	$z \geq 1.645$	Reject H_0
$\alpha_2 = 0.025$	$z \geq 1.96$	Reject H_0
$\alpha_3 = 0.01$	$z \geq 2.33$	Fail to reject H_0
$\alpha_4 = 0.005$	$z \geq 2.58$	Fail to reject H_0

Usefulness of p -value

- Provides more information about the strength of a test
- Indicates the smallest level at which the data is significant
- Can be compared with α irrespective of which type of test was used

Alternative definition

The p -value is the probability of obtaining a test statistic value at least as contradictory to H_0 as the value that actually resulted. **The smaller the p -value, the more contradictory are the data to H_0 .**

Hypothesis testing with the *p*-value

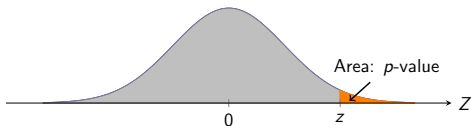
Step 1. Formulate your hypotheses

Step 2. Determine the *p*-value from the test statistic

Step 3. Conclude the test based on a chosen level of significance:

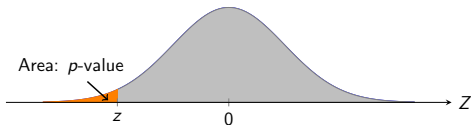
- ① $p\text{-value} \leq \alpha \implies$ reject H_0 at level α .
- ② $p\text{-value} > \alpha \implies$ do not reject H_0 at level α .

p-value for z tests



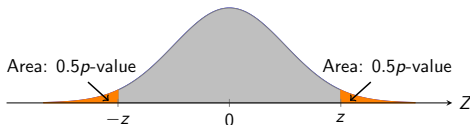
p-value: area in upper tail

$$p = 1 - \Phi(z) \quad (10)$$



p-value: area in lower tail

$$p = \Phi(z) \quad (11)$$



p-value: sum of area in both tails

$$p = 2(1 - \Phi(|z|)) \quad (12)$$

Hypothesis testing using p -value approach

Example 4: Getting enough sleep (OS 5.21)

400 students were randomly sampled from a large university, and 289 said they did not get enough sleep. Conduct a hypothesis test to check whether this represents a statistically significant difference from 50%, and use a significance level of 0.01

Step 1. Parameter of interest: p (proportion of students not getting enough sleep)

Step 2. Null hypothesis: $H_0 : p = 289/400 = .723$.

Step 3. Alternative hypothesis: $H_1 : p \neq .723$.

Step 4. Formula for test statistic value: $z = \frac{p - p_0}{SE_p}$

Hypothesis testing using p -value approach

Example 4: Getting enough sleep (cont.)

Step 5. Calculate test statistic value:

$$z = \frac{.723 - .5}{\sqrt{.5(.5)/400}} = 8.92$$

Step 6. Determine p -value (two-tailed test):

$$p\text{-value} = 2(1 - \Phi(8.92)) = 0.0$$

Step 7. Conclude:

Using a significance level of 0.01, we reject H_0 since $0.0204 > 0.01$. Thus, at the 1% significance level, there is sufficient evidence to conclude that true proportion differs from the target value of 0.5.

Recap of this lecture

- Definition of hypothesis testing
 - Null hypothesis (default/expected outcome) H_0
 - Alternate hypothesis (what we want to test/support; research hypothesis) H_1 or H_A
 - One-tailed or two-tailed
- Types of errors:
 - Type I: false positive
 - Type II: false negative
- Test statistic:
 - Sample proportion with independent observations and large enough sample size (normal distribution); Z-statistic:

$$z^* = \frac{p - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \quad (13)$$

- The p -value is the minimum probability of a Type I error.
 - Upper-tailed test: $p\text{-value} = 1 - \Phi(z)$; MATLAB: `normcdf(z, 'upper')`
 - Lower-tailed test: $p\text{-value} = \Phi(z)$; MATLAB: `normcdf(z)`
 - Two-tailed test: $p\text{-value} = 2(1 - \Phi(|z|))$; MATLAB: `2 * normcdf(abs(z))`