Hypothesis Testing and The Null Hypothesis

Statistics and Big Data

Niccolò Salvini, PhD

UCSC

Academic Year 2025-2026

Overview

What is Hypothesis Testing?

Imagine we are testing two different drugs to see which one helps patients recover from a virus more effectively.

What is Hypothesis Testing?

Imagine we are testing two different drugs to see which one helps patients recover from a virus more effectively. How do we determine if one drug is truly better than the other, or if the difference we observe is just due to random chance?

What is Hypothesis Testing?

Imagine we are testing two different drugs to see which one helps patients recover from a virus more effectively. How do we determine if one drug is truly better than the other, or if the difference we observe is just due to random chance? This leads us to the concept of hypothesis testing: a formal procedure for using sample data to evaluate a claim about a population.

Roadmap of Hypothesis Testing in this Course

In this course, we will focus on building hypothesis tests for the mean. We will explore different scenarios, each with its own specific tools.

Topics We Will Cover

- The logic of hypothesis testing.
- Hypothesis testing on **one mean**.
- Hypothesis testing on **two means** (paired and independent samples).
- Hypothesis testing on **more than two means** (ANOVA).
- Hypothesis testing in **regression**.

Concrete Example of Drug Testing

Let's consider a practical scenario:

Drug A Recovery Times

- Patient 1: 10 hours
- Patient 2: 15 hours
- Patient 3: 12 hours

Concrete Example of Drug Testing

Let's consider a practical scenario:

Drug A Recovery Times

- Patient 1: 10 hours
- Patient 2: 15 hours
- Patient 3: 12 hours

Next, we administer Drug B to another three patients:

Drug B Recovery Times

- Patient 4: 20 hours
- Patient 5: 18 hours
- Patient 6: 22 hours

Concrete Example of Drug Testing

Let's consider a practical scenario:

Drug A Recovery Times

- Patient 1: 10 hours
- Patient 2: 15 hours
- Patient 3: 12 hours

Next, we administer Drug B to another three patients:

Drug B Recovery Times

- Patient 4: 20 hours
- Patient 5: 18 hours
- Patient 6: 22 hours

What do these recovery times suggest about the effectiveness of the drugs?

Observing Differences

From our example, we observe that:

$$\bar{x}_A = \frac{10 + 15 + 12}{3} = 12.33 \text{ hours},$$
$$\bar{x}_B = \frac{20 + 18 + 22}{3} = 20 \text{ hours}.$$

Observing Differences

From our example, we observe that:

$$\bar{x}_A = \frac{10 + 15 + 12}{3} = 12.33 \text{ hours},$$
$$\bar{x}_B = \frac{20 + 18 + 22}{3} = 20 \text{ hours}.$$

This indicates that patients taking Drug A seem to recover faster on average.

Observing Differences

From our example, we observe that:

$$\bar{x}_A = \frac{10 + 15 + 12}{3} = 12.33 \text{ hours},$$
$$\bar{x}_B = \frac{20 + 18 + 22}{3} = 20 \text{ hours}.$$

This indicates that patients taking Drug A seem to recover faster on average. But can we conclude that Drug A is definitively better than Drug B?

The Role of Random Variation

What happens if we repeat the experiment and find:

$$\bar{x}_A = 35 \text{ hours},$$
$$\bar{x}_B = 30 \text{ hours}.$$

The Role of Random Variation

What happens if we repeat the experiment and find:

$$\bar{x}_A = 35 \text{ hours},$$
$$\bar{x}_B = 30 \text{ hours}.$$

This is contrary to our initial findings. The differences in recovery times could be attributed to uncontrolled variables, such as individual health or lifestyle factors.

The Role of Random Variation

What happens if we repeat the experiment and find:

$$\bar{x}_A = 35 \text{ hours},$$
$$\bar{x}_B = 30 \text{ hours}.$$

This is contrary to our initial findings. The differences in recovery times could be attributed to uncontrolled variables, such as individual health or lifestyle factors. This variability, or **random variation**, raises an important question: How can we be sure our initial result wasn't just a fluke?

Formalizing the Hypotheses

To test our observation formally, we structure our question around two competing hypotheses:

The Null Hypothesis (H_0)

This is the hypothesis of "no effect" or "no difference." It represents the status quo, the idea that any observed difference is just due to random chance.

Example: The two drugs have the same true mean effect on recovery time ($\mu_A = \mu_B$).

Formalizing the Hypotheses

To test our observation formally, we structure our question around two competing hypotheses:

The Null Hypothesis (H_0)

This is the hypothesis of "no effect" or "no difference." It represents the status quo, the idea that any observed difference is just due to random chance.

Example: The two drugs have the same true mean effect on recovery time ($\mu_A = \mu_B$).

The Alternative Hypothesis (H_1 or H_a)

This is the claim we want to find evidence for. It's the idea that there is a real effect or a real difference.

Example: The two drugs have different true mean effects on recovery time ($\mu_A \neq \mu_B$).

The Logic of Rejection

The core logic of hypothesis testing is:

1. Assume the null hypothesis (H_0) is true.
2. Collect data.
3. Ask: **"If the null hypothesis were true, how likely is it that we would observe data this extreme (or more extreme)?"**

The Logic of Rejection

The core logic of hypothesis testing is:

1. Assume the null hypothesis (H_0) is true.
2. Collect data.
3. Ask: **"If the null hypothesis were true, how likely is it that we would observe data this extreme (or more extreme)?"**

Key Principle

If the observed data is very unlikely under the assumption that H_0 is true, we have strong evidence against H_0 and we **reject** it in favor of H_1.

Errors, Confidence, and Power

When we make a decision, we can be wrong. It's crucial to understand the two types of errors and the two types of correct decisions.

Decision	**H_0 is True**	**H_0 is False**
Accept H_0	**Correct Decision** (Prob $= 1 - \alpha$)	**Type II Error** (Prob $= \beta$)
Reject H_0	**Type I Error** (Prob $= \alpha$)	**Correct Decision** (Prob $= 1 - \beta$)

- **Type I Error:** Rejecting H_0 when it is true. The probability is α, the **level of significance.**
- **Type II Error:** Failing to reject H_0 when it is false. The probability is β.
- **Level of Confidence $(1 - \alpha)$:** The probability of correctly not rejecting a true H_0.
- **Power of the Test $(1 - \beta)$:** The probability of correctly rejecting a false H_0. This is the test's ability to detect a real effect.

An Intuitive Example: A Courtroom Trial

The Setup

- **Null Hypothesis (H_0):** The defendant is innocent.
- **Alternative Hypothesis (H_1):** The defendant is guilty.

The system assumes innocence (H_0) until proven guilty. The prosecutor must provide strong evidence to reject H_0.

An Intuitive Example: A Courtroom Trial

The Setup

- **Null Hypothesis (H_0):** The defendant is innocent.
- **Alternative Hypothesis (H_1):** The defendant is guilty.

The system assumes innocence (H_0) until proven guilty. The prosecutor must provide strong evidence to reject H_0.

The Possible Errors

- **Type I Error:** Rejecting H_0 (innocence) when it's true $\rightarrow$ **Condemning an innocent person.**
- **Type II Error:** Failing to reject H_0 (innocence) when it's false $\rightarrow$ **Letting a guilty person go free.**

Because the social cost of a Type I error is so high, we set the bar for rejecting H_0 very high (by choosing a small α).

Failing to Reject the Hypothesis

In our drug example, what if the results are close? Maybe the difference is small and could plausibly be due to random chance.

Failing to Reject the Hypothesis

In our drug example, what if the results are close? Maybe the difference is small and could plausibly be due to random chance. In this case, we don't have enough evidence to confidently reject the null hypothesis.

Failing to Reject the Hypothesis

In our drug example, what if the results are close? Maybe the difference is small and could plausibly be due to random chance. In this case, we don't have enough evidence to confidently reject the null hypothesis.

Conclusion

We **fail to reject the null hypothesis**.

Failing to Reject the Hypothesis

In our drug example, what if the results are close? Maybe the difference is small and could plausibly be due to random chance. In this case, we don't have enough evidence to confidently reject the null hypothesis.

Conclusion

We **fail to reject the null hypothesis**.

This does not mean we've proven H_0 is true! It only means we don't have sufficient evidence to say it's false. It's like a "not guilty" verdict in court, which is different from a verdict of "innocent."

Summary of Key Concepts

1. Hypothesis testing uses sample data to decide between a null (H_0) and an alternative (H_1) hypothesis.
2. We start by assuming H_0 (no effect) is true.
3. If our data is too surprising under this assumption, we reject H_0.
4. The decision can lead to Type I (rejecting a true H_0, prob=α) or Type II (not rejecting a false H_0, prob=β) errors.
5. We control the Type I error by setting a significance level, α, before the experiment.
6. If evidence is not strong enough, we "fail to reject" H_0, which is not the same as proving it is true.

Exercises

Exercise 1

A pharmaceutical company develops a new drug to lower blood pressure. They want to test if it's more effective than the existing drug.

- What is the null hypothesis (H_0)?
- What is the alternative hypothesis (H_1)?
- Describe what a Type I and a Type II error would mean in this context. Which one is more dangerous for the patient?

Exercises

Exercise 1

A pharmaceutical company develops a new drug to lower blood pressure. They want to test if it's more effective than the existing drug.

- What is the null hypothesis (H_0)?
- What is the alternative hypothesis (H_1)?
- Describe what a Type I and a Type II error would mean in this context. Which one is more dangerous for the patient?

Exercise 2

Given the following recovery times, calculate the mean for each drug. Do you think the difference is significant just by looking at it?

- Drug A: 10, 12, 14 hours
- Drug B: 15, 18, 20 hours