How to calculate p-values

Statistics and Big Data

Niccolò Salvini, PhD

UCSC

Academic Year 2025-2026

Course: Statistics and Big Data

Overview

What Are P-Values?

Consider a scenario where we are testing two drugs, Drug A and Drug B.

What Are P-Values?

Consider a scenario where we are testing two drugs, Drug A and Drug B.
How can we determine if Drug A is truly more effective than Drug B?

What Are P-Values?

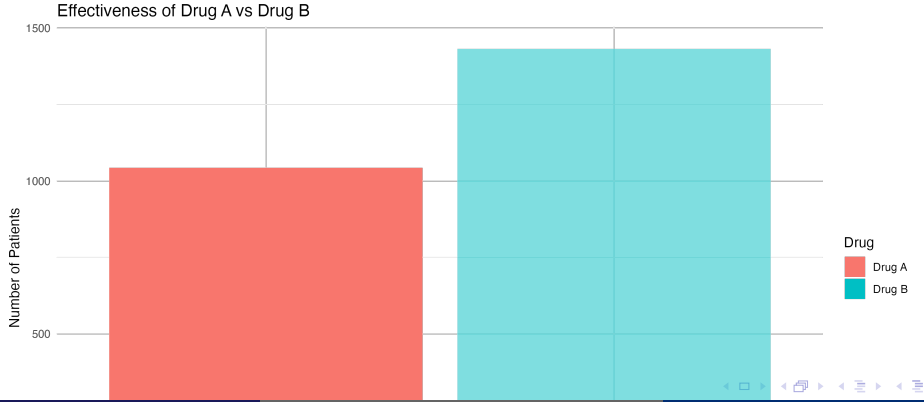
Consider a scenario where we are testing two drugs, Drug A and Drug B.
How can we determine if Drug A is truly more effective than Drug B?
This question leads us to the concept of p-values, a critical tool in statistical hypothesis testing.

Example of Drug Testing

Let's illustrate this with a practical example:
Imagine we conduct an experiment where Drug A is administered to 1,046 individuals, resulting in 1,043 cures, while Drug B is given to 1,434 individuals, yielding only 2 cures.

Example of Drug Testing

Let's illustrate this with a practical example:
Imagine we conduct an experiment where Drug A is administered to 1,046 individuals, resulting in 1,043 cures, while Drug B is given to 1,434 individuals, yielding only 2 cures. This stark difference raises the question: Is Drug A significantly better than Drug B?



Understanding the Results

Now that we have observed the results, we can analyze the effectiveness:

- Drug A: 99.7% cure rate (1,043 cured out of 1,046)
- Drug B: 0.1% cure rate (2 cured out of 1,434)

Understanding the Results

Now that we have observed the results, we can analyze the effectiveness:

- Drug A: 99.7% cure rate (1,043 cured out of 1,046)
- Drug B: 0.1% cure rate (2 cured out of 1,434)

Key insight: The large disparity suggests that the results are unlikely to be due to random chance.

Understanding the Results

Now that we have observed the results, we can analyze the effectiveness:

- Drug A: 99.7% cure rate (1,043 cured out of 1,046)
- Drug B: 0.1% cure rate (2 cured out of 1,434)

Key insight: The large disparity suggests that the results are unlikely to be due to random chance.

Mathematical formulation: The observed proportions can be expressed as:

$$p_A = \frac{1043}{1046} \quad \text{and} \quad p_B = \frac{2}{1434}$$

What If the Results Were Different?

Now, consider an alternative scenario:
Suppose Drug A cures 37% of patients while Drug B cures 31%.

What If the Results Were Different?

Now, consider an alternative scenario:
Suppose Drug A cures 37% of patients while Drug B cures 31%.
How confident can we be that Drug A is superior? This is where p-values come into play.

Defining P-Values

Definition

A p-value quantifies the probability of observing the data, or something more extreme, assuming that the null hypothesis is true.

Defining P-Values

Definition

A p-value quantifies the probability of observing the data, or something more extreme, assuming that the null hypothesis is true.

Mathematical Formulation

$$P(\text{data}|H_0)$$

where H_0 is the null hypothesis stating that there is no difference between Drug A and Drug B.

Defining P-Values

Definition

A p-value quantifies the probability of observing the data, or something more extreme, assuming that the null hypothesis is true.

Mathematical Formulation

$$P(\text{data}|H_0)$$

where H_0 is the null hypothesis stating that there is no difference between Drug A and Drug B.

Interpretation

A smaller p-value indicates stronger evidence against the null hypothesis.

The Threshold for Decision-Making

What threshold should we use to determine significance?

The Threshold for Decision-Making

What threshold should we use to determine significance?
A commonly accepted threshold is 0.05. This means that if the p-value is less than 0.05, we reject the null hypothesis.

The Threshold for Decision-Making

What threshold should we use to determine significance?
A commonly accepted threshold is 0.05. This means that if the p-value is less than 0.05, we reject the null hypothesis.
In practical terms, this implies that if we repeated the experiment many times, we would expect to see a false positive in 5% of cases.

Exploring False Positives

Consider a scenario where the p-value is 0.9.

Exploring False Positives

Consider a scenario where the p-value is 0.9.
This indicates a high probability that any observed difference is due to random variation, leading us to fail to reject the null hypothesis.

Exploring False Positives

Consider a scenario where the p-value is 0.9.
This indicates a high probability that any observed difference is due to random variation, leading us to fail to reject the null hypothesis.
Conversely, a p-value of 0.01 suggests a significant difference, even if both groups received the same treatment.

Understanding False Positives

Terminology Alert

A small p-value when there is no actual difference is termed a false positive.

Understanding False Positives

Terminology Alert

A small p-value when there is no actual difference is termed a false positive.

With a threshold of 0.05, we expect 5% of experiments to yield a p-value below this threshold purely by chance.

Understanding False Positives

Terminology Alert

A small p-value when there is no actual difference is termed a false positive.

With a threshold of 0.05, we expect 5% of experiments to yield a p-value below this threshold purely by chance.
This highlights the importance of selecting an appropriate threshold based on the context of the study.

The Importance of Effect Size

While a small p-value indicates a statistically significant difference, it does not convey the magnitude of that difference.

The Importance of Effect Size

While a small p-value indicates a statistically significant difference, it does not convey the magnitude of that difference.
For example, a p-value of 0.04 may arise from a small difference of 1 unit, while a p-value of 0.24 could correspond to a larger difference of 6 units.

The Importance of Effect Size

While a small p-value indicates a statistically significant difference, it does not convey the magnitude of that difference.
For example, a p-value of 0.04 may arise from a small difference of 1 unit, while a p-value of 0.24 could correspond to a larger difference of 6 units.
Key insight: Statistical significance does not equate to practical significance.

Summary of Key Concepts

1. P-values help us assess the evidence against the null hypothesis.
2. A common threshold for significance is 0.05.
3. Small p-values indicate statistical significance but do not reflect effect size.

Summary of Key Concepts

1. P-values help us assess the evidence against the null hypothesis.
2. A common threshold for significance is 0.05.
3. Small p-values indicate statistical significance but do not reflect effect size.

Understanding these concepts is crucial for interpreting experimental results accurately.

Exercises

Exercise 1

Define a p-value and explain its significance in hypothesis testing.

Exercises

Exercise 1

Define a p-value and explain its significance in hypothesis testing.

Exercise 2

Given the following data, calculate the p-value: Drug A cures 30 out of 100 patients, and Drug B cures 25 out of 100 patients.

Exercises

Exercise 1

Define a p-value and explain its significance in hypothesis testing.

Exercise 2

Given the following data, calculate the p-value: Drug A cures 30 out of 100 patients, and Drug B cures 25 out of 100 patients.

Exercise 3

Discuss a scenario where a false positive could have serious implications in a medical study.

Exercises

Exercise 1

Define a p-value and explain its significance in hypothesis testing.

Exercise 2

Given the following data, calculate the p-value: Drug A cures 30 out of 100 patients, and Drug B cures 25 out of 100 patients.

Exercise 3

Discuss a scenario where a false positive could have serious implications in a medical study.

Exercise 4

Analyze the difference between statistical significance and practical significance in the context of drug effectiveness.