UnCodemy

# Hypothesis Testing Guide PDF

Hypothesis testing checks whether data provides enough evidence against a default claim, called the null hypothesis. You state both hypotheses, choose a significance level, compute a test statistic and p-value, and compare. A small p-value means the data would be unusual if the null were true, which is not the same as proving the alternative.

## Key takeaways

- The null hypothesis is the default claim of no effect or no difference.
- A p-value is the chance of data at least this extreme if the null is true.
- Failing to reject the null is not proof that it is true.
- Type I error is a false alarm; Type II error is a missed effect.
- Statistical significance does not mean practical importance.

## What is hypothesis testing?

Hypothesis testing is a statistical procedure that uses sample data to decide whether there is enough evidence to reject a null hypothesis in favour of an alternative claim about a population.

## What is the logic behind a hypothesis test?

You begin with a claim you want to challenge, such as "the new delivery process has the same average time as the old one". This is the **null hypothesis (H0)**. The **alternative (H1)** says there is a difference. You collect a sample, compute how far the result is from what H0 predicts, and judge whether the gap is too large to be explained by chance alone.

## What are the steps?

1. State H0 and H1 in words and symbols.
2. Choose a significance level, often 0.05, before looking at results.
3. Pick a test that fits the data type and design.
4. Compute the test statistic and the p-value.
5. If the p-value is below the level, reject H0; otherwise do not reject it.
6. Report the effect size, a confidence interval and a plain-language conclusion.

## How do I interpret a p-value?

The p-value is the probability of getting a result at least as extreme as the one observed, assuming H0 is true. A p-value of 0.03 does not mean a three percent chance that H0 is true. It

means that if H0 were true, such data would appear about three times in a hundred. Always read it with the effect size, because with a huge sample even a trivial difference can be significant.

## Which test should I use?

| Situation | Common test |
|---|---|
| Compare a sample mean with a known value | One-sample t-test |
| Compare means of two independent groups | Two-sample t-test |
| Compare before and after on the same units | Paired t-test |
| Compare means of three or more groups | ANOVA |
| Relationship between two categorical variables | Chi-square test |
| Data far from normal or small samples | Non-parametric tests such as Mann-Whitney |

Check assumptions such as independence, rough normality of means and similar spread.

## What are Type I and Type II errors?

A **Type I error** rejects a true null, a false alarm, and its probability is the significance level. A **Type II error** fails to reject a false null, a missed effect. **Power** is the chance of detecting a real effect and rises with larger samples and larger effects. Plan sample size before collecting data.

## Can I show a small worked example?

```
import statistics as st
a = [12, 14, 11, 13, 15, 12]
b = [15, 17, 14, 16, 18, 16]
mean_diff = st.mean(b) - st.mean(a)
print(round(mean_diff, 2))
```

The difference in means is 3.33 here. A two-sample t-test would then compare this difference with the variation inside each group to decide whether it is likely to be chance.

## What mistakes should I avoid?

Do not try many tests and report only the significant one; this inflates false positives, so use corrections when running several tests. Do not stop collecting data the moment the p-value dips below the limit. Do not confuse significance with importance. Do not ignore assumptions or sample design.

## How should results be reported?

State the question, the test, the sample size, the effect size with a confidence interval, the p-value and a decision in plain words. For example: "Delivery time fell by about three minutes on average; the interval excludes zero, and the change is large enough to matter operationally."

## How does sample size affect the result?

With a small sample, even a real difference can look like noise, so the test may fail to detect it. With a very large sample, tiny and unimportant differences become statistically significant. For this reason, decide in advance the smallest difference that would matter in practice and calculate the sample size needed to detect it with reasonable power. Reporting a confidence interval alongside the p-value shows both the size and the uncertainty of the effect, which readers can judge for themselves.

## What does a plain-language conclusion look like?

Avoid saying only "p is less than 0.05". Say instead, "Orders placed through the new checkout were completed about eight seconds faster on average; the interval for the difference stayed above zero, so we recommend rolling it out to more users while we keep measuring." Notice that the sentence states the effect, the evidence and the action. It also avoids claiming certainty, which is appropriate because a test supports decisions under uncertainty and cannot remove it completely.

## Frequently asked questions

### What is a null hypothesis?

The null hypothesis is the default claim that there is no effect or difference, which a test tries to find enough evidence against.

### What does a p-value of 0.05 mean?

It means that if the null hypothesis were true, results at least this extreme would occur about five times in a hundred.

### Does failing to reject the null prove it is true?

No. It only means the data did not provide enough evidence against it, which may be due to a small sample or noise.

### What is the difference between Type I and Type II errors?

A Type I error rejects a true null as a false alarm, while a Type II error misses a real effect by failing to reject a false null.

### Is a significant result always important?

No. With large samples tiny differences can be significant, so always check effect size and practical meaning alongside the p-value.

## Sources and further reading

- SciPy statistics documentation
- NIST e-Handbook of Statistical Methods

## About this PDF

These PDFs are study material for learning and revision. They are not a certificate, a degree or a promise of a job, and technology changes, so check official documentation for the latest details. Spot an error? Write to info@uncodemy.com and the trainer team will review it.

More free resources and courses: uncodemy.com | info@uncodemy.com | +91 98183 66550