Hypothesis Testing

Lesson 12

Digital Finance

Outline

What is Hypothesis Testing?

Goal: Make decisions based on data

Examples:

- Is this drug effective?
- Does this trading strategy beat the market?
- Is the coin fair?

Framework:

- State competing hypotheses
- Collect data
- Decide which hypothesis is supported

Hypothesis testing formalizes statistical decision-making.

Null and Alternative Hypotheses

Null hypothesis H_0:

- Default/status quo position
- What we assume unless evidence says otherwise
- Example: $H_0 : \mu = 0$ (no effect)

Alternative hypothesis H_1 (also written H_a):

- What we're trying to show
- Example: $H_1 : \mu \neq 0$ (there is an effect)

Analogy: Innocent until proven guilty

We either reject H_0 or fail to reject it; we never "accept" H_0.

Types of Errors

Two ways to be wrong:

- **Type I (false alarm):** Convict an innocent person. Reject H_0 when it's true.
- **Type II (missed detection):** Let a guilty person go free. Don't reject H_0 when it's false.

We control Type I error with significance level α (usually 5%).
Power $= 1 - \beta =$ ability to detect a real effect.

	H_0 True	H_0 False
Reject H_0	Type I (false alarm)	Correct!
Don't Reject	Correct!	Type II (miss)

Trade-off: Lowering false alarms increases missed detections.

Test Statistic

Question: How surprising is our data if H_0 is true?
Answer: The test statistic measures this.
Intuition:

$$\text{Test stat} = \frac{\text{What we observed} - \text{What } H_0 \text{ says}}{\text{Typical random variation}}$$

Interpretation:

- Test stat near 0: Data is close to what H_0 predicts
- Test stat far from 0: Data is surprising if H_0 is true

Large absolute test statistic = evidence against H_0.

P-value

P-value: Probability of observing result as extreme (or more) if H_0 is true

Interpretation:

- Small p-value: Evidence against H_0
- Large p-value: Data consistent with H_0

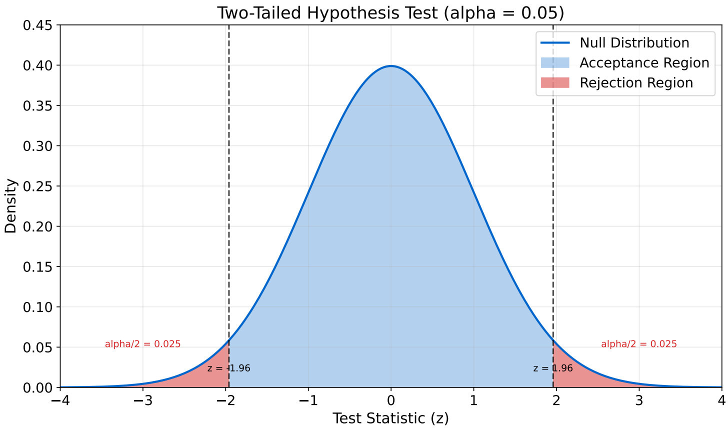
Decision rule:

- If p-value $\leq \alpha$: Reject H_0
- If p-value $> \alpha$: Fail to reject H_0

Common α: 0.05 (5%), 0.01 (1%), 0.10 (10%)

P-value is NOT the probability that H_0 is true!

Rejection Regions



Extreme values lead to rejection of the null hypothesis.

One-tailed vs Two-tailed Tests

Two-tailed (most common):

- $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$
- Reject for extreme values in either direction

Right-tailed:

- $H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$
- Reject only for large positive values

Left-tailed:

- $H_0 : \mu \geq \mu_0$ vs $H_1 : \mu < \mu_0$
- Reject only for large negative values

Direction of alternative determines rejection region.

One-Sample t-Test

Test whether μ equals a specific value:

Hypotheses: $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$

Test statistic:

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

Reject H_0 if: $|t| > t_{\alpha/2,n-1}$

Finance example: Test if average return equals zero

Most common test for a single mean.

Statistical vs Practical Significance

Statistical significance:

- $p < 0.05$ means unlikely under H_0
- Does NOT mean the effect is large or important

Practical significance:

- Is the effect size meaningful?
- With large n, tiny effects become significant

Always report:

- Effect size (how big?)
- Confidence interval (how precise?)
- P-value (how unlikely under H_0?)

A statistically significant result may be practically irrelevant.

Type I and Type II Errors

Decision Outcomes in Hypothesis Testing

	H_0 True (No Effect)	**H_0 False (Real Effect)**
Reject H_0	TYPE I ERROR (alpha = False Positive)	CORRECT (Power = 1-beta)
Fail to Reject H_0	CORRECT (1-alpha)	TYPE II ERROR (beta = False Negative)

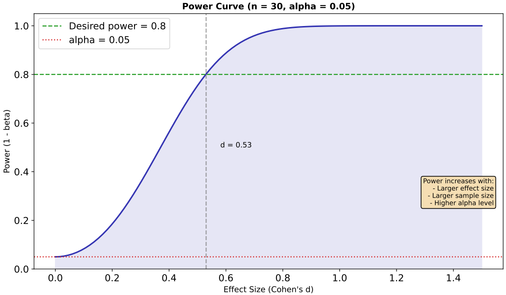
alpha = P(Type I Error) = P(Reject H_0 | H_0 True)

beta = P(Type II Error) = P(Fail to Reject H_0 | H_0 False)

Balance false positives and false negatives.

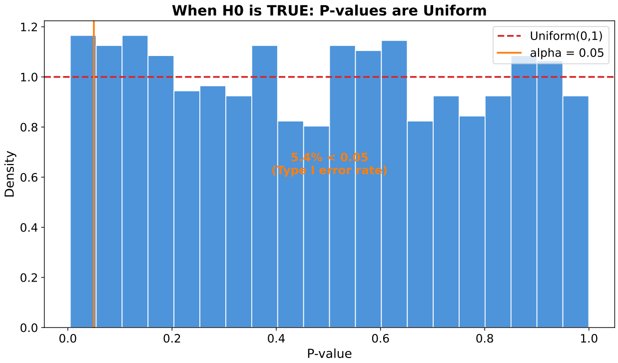
Statistical Power

Power = Probability of detecting a real effect (if it exists).
Low power = high risk of missing real effects!



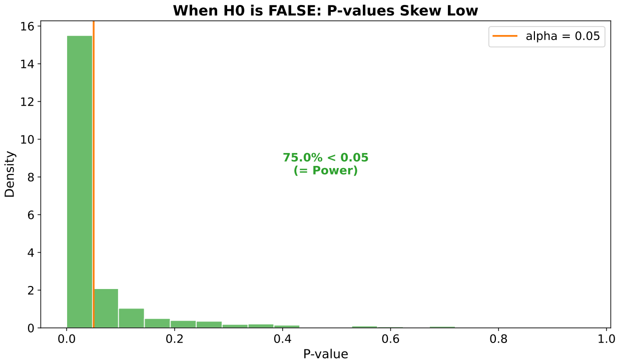
Power increases with larger sample size and larger effect size.

P-values When H0 is True



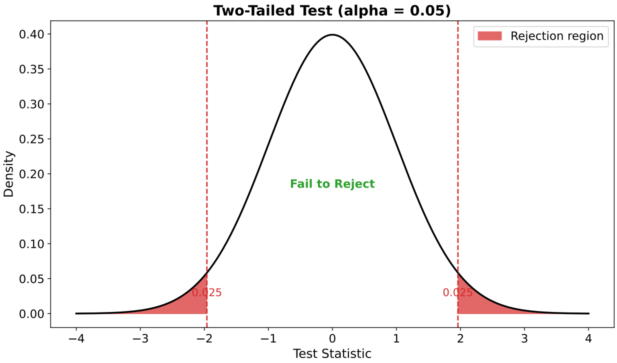
Under H0, p-values are uniformly distributed.

P-values When H0 is False



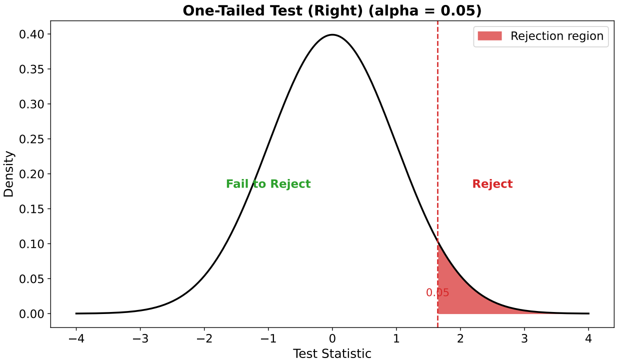
Under H1 (real effect), p-values concentrate near 0.

Two-Tailed Test



Reject for extreme values in either direction.

One-Tailed Test (Right)



Reject only for large positive values.

Summary

Hypothesis testing framework:

- H_0 (null) vs H_1 (alternative)
- Type I and Type II errors

Key concepts:

- Test statistic measures evidence against H_0
- P-value quantifies how surprising the data is
- Reject H_0 if p-value $\leq \alpha$

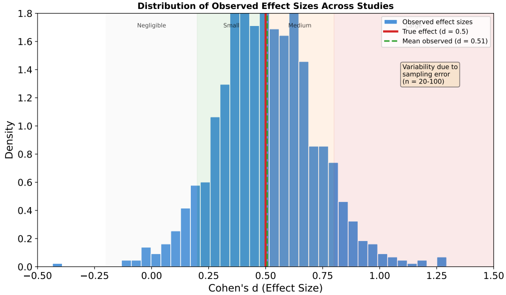
Remember:

- Statistical $\neq$ practical significance
- Report effect sizes and CIs

Course complete! You now have foundations in probability and statistics.

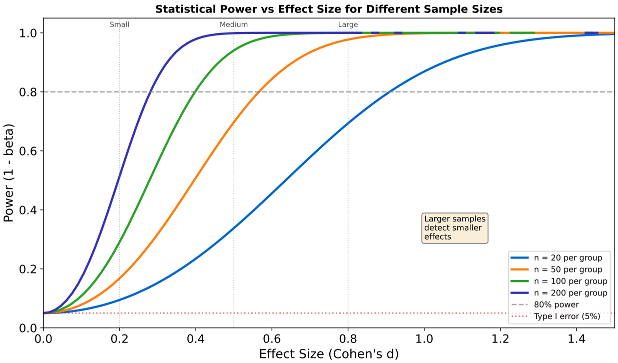
Effect Size Distribution

Effect size = How big is the effect? (standardized, scale-free measure)
Common effect sizes: Cohen's d (means), r (correlation), odds ratio (binary)



Small d ≈ 0.2, medium ≈ 0.5, large ≈ 0.8 (Cohen's benchmarks).

Power vs Sample Size

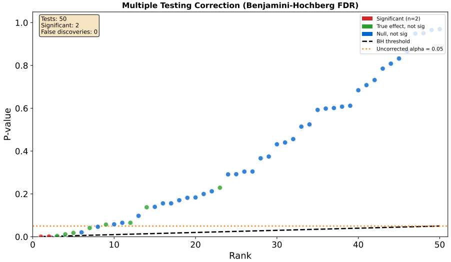


Larger samples increase power to detect true effects.

Multiple Testing Correction (FDR)

Problem: Test 100 hypotheses at $\alpha = 0.05 \rightarrow$ expect 5 false positives by chance!
Solution: Adjust for multiple comparisons.

- Bonferroni: Divide α by number of tests (conservative)
- FDR (Benjamini-Hochberg): Controls expected % of false discoveries (less strict)



Advanced topic. Key: more tests = more corrections needed.