HYPOTHESIS TESTING ERRORS: TYPE I, TYPE II, AND POWER

AP Statistics · Mr. Merrick · February 4, 2026

In hypothesis testing, we make a decision (reject or fail to reject H_0), but reality could be different (H_0 might be true or false). This packet focuses on two possible mistakes:

$$\alpha = P(\text{reject } H_0 \mid H_0 \text{ true}) \qquad \beta = P(\text{fail to reject } H_0 \mid H_0 \text{ false})$$
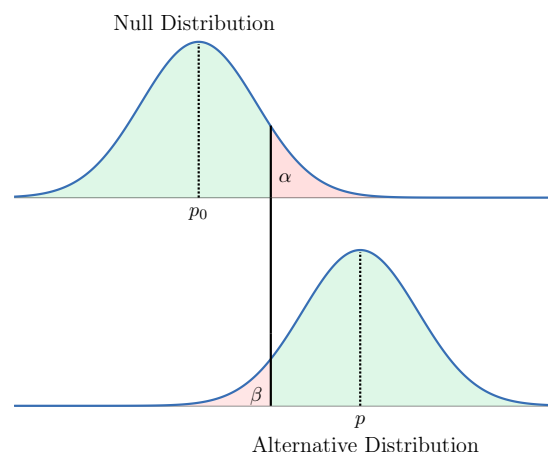
and on **power**, the probability we correctly reject H_0 when H_0 is false:

$$\text{power} = 1 - \beta.$$

Decision vs. Reality

- In a hypothesis test, we choose a rule that tells us when to reject H_0 (a *cutoff*).

- If H_0 is true, rejecting it is a Type I error (probability α).

- If H_0 is false, failing to reject it is a Type II error (probability β).

- When H_0 is false, the chance we correctly reject is power $= 1 - \beta$.

About the decision

- Rejecting H_0 does *not* prove H_a. It means the result would be unusually rare if H_0 were true.

- Failing to reject H_0 does *not* mean H_0 is true. It means the data is reasonably consistent with the null model.

How changes affect α, β, and power

Tradeoff (same sample size, same effect size)

- If you make α smaller (stricter cutoff), β usually gets larger (harder to reject).

- If you make α larger (easier cutoff), β usually gets smaller.

How to increase power

- Increase the sample size n (curves get narrower).

- Have a larger effect size (means are farther apart).

- Use a larger α (move the cutoff left), but this increases Type I risk.

Think of the cutoff as the line between "Fail to Reject" and "Reject." Moving the cutoff changes the size of the shaded regions.

- Make the cutoff *harder* to reach (more extreme to reject):

$$\alpha \downarrow \quad \beta \uparrow \quad \text{power } (1 - \beta) \downarrow$$

- Make the cutoff *easier* to reach (less extreme to reject):

$$\alpha \uparrow \quad \beta \downarrow \quad \text{power } (1 - \beta) \uparrow$$

Connection to p-values: choosing a smaller α means requiring a smaller p-value to reject H_0.

When is a Type I error worse vs. a Type II error worse?

Whether Type I or Type II errors are "more favorable" depends on the context (the consequences).

When Type I errors are worse (so we prefer a smaller α):

- Court trials: convicting an innocent person.

- Medical approval: approving a medication that does not actually work (or has no real benefit).

- Accusations/cheating investigations: punishing someone who didn't do it.

When Type II errors are worse (so we prefer smaller β / higher power):

- Disease screening: missing a real disease (sending a sick person home as "healthy").

- Safety systems: failing to detect a real problem (fire alarm, carbon monoxide detector).

- Quality control: shipping defective products as "good."

AP-style practice: interpreting errors

Problem 1: Airport Security Screening

An airport uses a screening test to detect prohibited items. Let H_0 be "a passenger does not have a prohibited item" and H_a be "a passenger has a prohibited item."

1. Describe a Type I error in context.

2. Describe a Type II error in context.

3. Which error would the airport consider more serious? Explain.

Problem 2: New Tutoring Program

A school claims a new tutoring program increases the pass rate on a final exam above 70%. Let p be the true pass rate for students in the program.

$$H_0 : p = 0.70 \qquad H_a : p > 0.70$$

1. Interpret $\alpha = P(\text{reject } H_0 \mid H_0 \text{ true})$ in context.

2. Interpret $\beta = P(\text{fail to reject } H_0 \mid H_0 \text{ false})$ in context.

3. Interpret power in context.