Hypothesis Testing

With one sample

Noara Razzak

August 8, 2025

Basic Concepts

- **Null Hypothesis (H_0)**: Default assumption (e.g., $\mu = \mu_0$)
- **Alternative Hypothesis (H_1)**: Counter-claim (e.g., $\mu \neq \mu_0$)

Types of Errors Possible Outcomes in Hypothesis Testing

	H_0 True	**H_0 False**
Reject H_0	Type I Error (α)	Correct Decision
Fail to Reject H_0	Correct Decision	Type II Error (β)

Visualizing error types

Type I Error ()

False positive

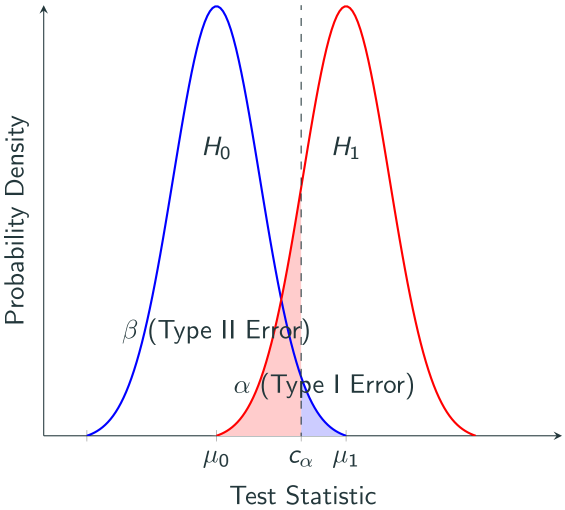
Rejecting H_0 when it's true

Type II Error ()

False negative

Failing to reject H_0 when it's false

Visualizing error types



Explanation:

- The blue curve represents the null hypothesis (H_0) distribution
- The red curve represents the alternative hypothesis (H_1) distribution
- The dashed line marks the critical value for rejecting H_0
- The blue-shaded area (α) shows the probability of Type I error - rejecting H_0 when it's actually true
- The red-shaded area (β) shows Type II error - failing to reject H_0 when H_1 is true

Type I error occurs when we observe a result in the blue-shaded region (extreme under H_0) leading us to incorrectly reject the null hypothesis.

Real world implications

Legal System Parallel

- H_0: Defendant is innocent
- H_1: Defendant is guilty

Type I Error

Convicting an innocent person (False positive)

Type II Error

Freeing a guilty person (False negative)

Real World Implications

Disease Screening

- H_0: Patient is healthy
- H_1: Patient has disease

Type I Error

False alarm
Healthy patient diagnosed as sick
(Leads to unnecessary treatment)

Type II Error

Missed detection
Sick patient diagnosed as healthy
(Leads to lack of treatment)

Recap for previous lecture

- **Test t-Statistic**: $T_n = \frac{\bar{X}_n - \mu_0}{s/\sqrt{n}}$ when population α is unknown
- **Test z-Statistic**: $Z_n = \frac{\bar{X}_n - \mu_0}{\alpha/\sqrt{n}}$ when population α is known
- **Significance Level (α)**: Probability of Type I error (typically 0.05)

Normal Distribution vs. t-Distribution for text statistic

When population α is known use the Normal Distribution

When population α is not known use the t-Distribution

Choosing Significance Level (α)

- α = Probability of Type I error
- Common choices: $0.01, 0.05, or 0.10$
- Selection depends on consequences of errors

Factors to Consider

- More serious consequences of Type I error $\rightarrow$ Choose smaller
- More serious consequences of Type II error $\rightarrow$ May choose larger
- Field standards (0.05 common in social sciences)

Example 1

In an issue of U. S. News and World Report, an article on school standards stated that about half of all students in France, Germany, and Israel take advanced placement exams and a third pass. The same article stated that 6.6% of U.S. students take advanced placement exams and 4.4% pass. Test if the percentage of U.S. students who take advanced placement exams is more than 6.6%. State the null and alternative hypotheses.

Example 1 - Solution

$H_o : p \leq 0.066$ $H_a : p > 0.066$

Example 2

Statistics students believe that the mean score on the first statistics test is 65. A statistics instructor thinks the mean score is higher than 65. He samples ten statistics students and obtains the scores 65; 65; 70; 67; 66; 63; 63; 68; 72; 71. He performs a hypothesis test using a 5% level of significance. The data are assumed to be from a normal distribution.

Example 2 - solution

A 5% level of significance means that $\alpha = 0.05$. This is a test of a single population mean.

$$H_o : \mu = 65$$

$$\mathrm{H}_a : \mu > 65$$

Random variable: $\bar{X}$ = average score on the first statistics test.
Distribution for the test: If you read the problem carefully, you will notice that there is no population standard deviation given.

You are only given $n = 10$ sample data values. Notice also that the data come from a normal distribution. This means that the distribution for the test is a student's t.

Example 2 - solution

Use the t-distribution.

Therefore, the distribution for the test is

$$t_9 \ where \ n = 10 \ and \ df = 10 - 1 = 9$$

.

Calculate the p-value using the Student's t-distribution:
$p - value = P(\bar{x} > 67) = 0.0396$ where the sample mean and sample standard deviation are calculated as 67 and 3.1972 from the data.

Next class we will cover Hypothesis Testing with two samples