Math Revision Session

Statistics (7): Hypothesis Testing

Jukina HATAKEYAMA

The University of Osaka, Department of Economics

April 10, 2026

Introduction to Hypothesis Testing

- **Hypothesis testing** is a statistical method for making inferences about a population using sample data.
- We usually compare two competing hypotheses:
  - **Null hypothesis** (H_0): the baseline or status quo.
  - **Alternative hypothesis** (H_1): the claim for which we seek evidence.
- We choose a **significance level** α, which is the probability of rejecting H_0 when H_0 is true.
- Two kinds of errors may occur:
  - **Type I error**: rejecting H_0 when it is true.
  - **Type II error**: failing to reject H_0 when it is false.

Basic Procedure

Procedure

1. **Assume H_0 is true.**
2. **Collect sample data.**
3. **Compute a test statistic** whose distribution under H_0 is known.
4. **Compare the statistic with a critical value**, or use a p-value.
5. **Make a decision:**
   - if the evidence is strong, reject H_0;
   - otherwise, fail to reject H_0.

Testing a Population Mean

We illustrate hypothesis testing by testing whether a population mean is zero.

- Suppose we have a random sample of size n from a normal distribution with known variance σ^2.
- The sample mean $\bar{X}$ estimates the population mean μ.
- Suppose the realised sample mean is

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = 0.01.$$

- We test

$$H_0 : \mu = 0 \qquad \text{against} \qquad H_1 : \mu \neq 0.$$

Steps in Hypothesis Testing

1. **Set up hypotheses:**

$$H_0 : \mu = 0, \qquad H_1 : \mu \neq 0.$$

2. **Choose a significance level:**

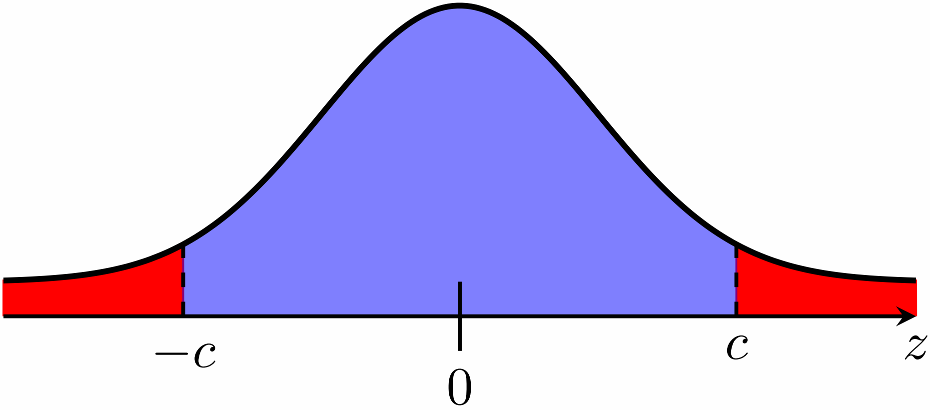
$$\alpha = 0.05 \qquad \text{so} \qquad 1 - \alpha = 0.95.$$

3. **Compute the test statistic:**

$$Z = \frac{\bar{X} - 0}{\sigma/\sqrt{n}} \sim N(0,1) \quad \text{under } H_0.$$

4. **Find the rejection region.**
5. **Make a decision.**

Rejection Region in a Two-Sided Test



$$\Pr(Z < -c) + \Pr(Z > c) = \alpha.$$

The red areas are the rejection regions.

How to Interpret the Result

- If $H_0 : \mu = 0$ is true, then the test statistic should usually be close to 0.
- Values far from 0 are unlikely under H_0.
- Hypothesis testing asks whether the observed test statistic is **too far from 0** to be plausibly explained by random sampling alone.

Reject or Fail to Reject

- In hypothesis testing, we usually say:
  - **reject** H_0, or
  - **fail to reject** H_0.
- We do **not** usually say that we **accept** H_0.
- Failing to reject H_0 only means that the data do not provide sufficiently strong evidence against it.

Introduction to Test Statistics

- A **test statistic** is a function of the sample data used to test a hypothesis.
- It summarises the information in the sample that is relevant for the test.
- The test statistic is chosen so that its distribution under H_0 is known.
- Large deviations from what is expected under H_0 provide evidence against H_0.

When the Population Distribution is Normal

Assume

$$X_i \sim N(\mu, \sigma^2).$$

Then

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \qquad \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

If σ^2 is unknown, we replace it by the sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

and then

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1).$$

Under $H_0 : \mu = 0$, the test statistic becomes

$$T = \frac{\bar{X}}{S/\sqrt{n}} \sim t(n-1).$$

Critical Region

For a two-sided t-test,

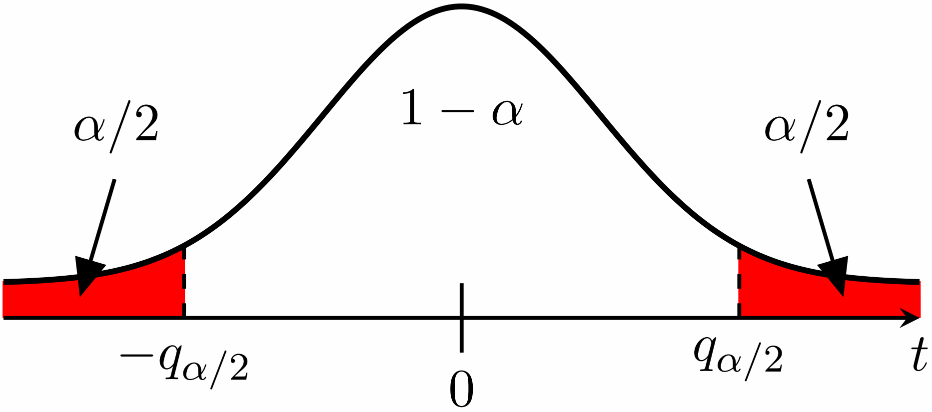
$$\alpha = \Pr(T < -q_{\alpha/2}) + \Pr(T > q_{\alpha/2}),$$

or equivalently,

$$1 - \alpha = \Pr(-q_{\alpha/2} < T < q_{\alpha/2}).$$

If the observed value of T lies outside this interval, we reject H_0.

Visualising the Rejection Region



The red areas are the rejection regions.

Brief Example

Suppose

$$H_0 : \mu = 0, \qquad H_1 : \mu \neq 0, \qquad \alpha = 0.05.$$

If $n = 10$, $\bar{x} = 0.01$, and $S^2 = 0.007$, then

$$T = \frac{0.01}{\sqrt{0.007/10}} \approx 0.38.$$

With 9 degrees of freedom, the two-sided critical value is about

$$t_{0.025,9} \approx 2.262.$$

Since

$$|0.38| < 2.262,$$

we fail to reject H_0.

One-Sided vs Two-Sided Tests

- A **one-sided test** looks for an effect in one specific direction.
- A **two-sided test** looks for any difference, regardless of direction.

The choice depends on the research question.

Example: One-Sided Test

Problem: A new medicine is believed to lower blood pressure. The known benchmark is 130 mmHg. We want to test whether the mean blood pressure after treatment is lower.

Hypotheses:

$$H_0 : \mu \geq 130, \qquad H_1 : \mu < 130.$$

Data: 125, 128, 126, 130, 129, 127, 124, 123, 132, 126
Significance level:

$$\alpha = 0.05.$$

Calculation for the One-Sided Test

1. Sample mean and sample variance:

$$\bar{X} = 127.0, \qquad S^2 \approx 7.56.$$

2. Standard error:

$$SE = \frac{S}{\sqrt{n}} \approx \frac{\sqrt{7.56}}{\sqrt{10}} \approx 0.87.$$

3. Test statistic:

$$T = \frac{\bar{X} - 130}{SE} = \frac{127.0 - 130}{0.87} \approx -3.45.$$

4. Critical value for a left-sided t-test with 9 degrees of freedom:

$$t_{0.05,9} \approx -1.833.$$

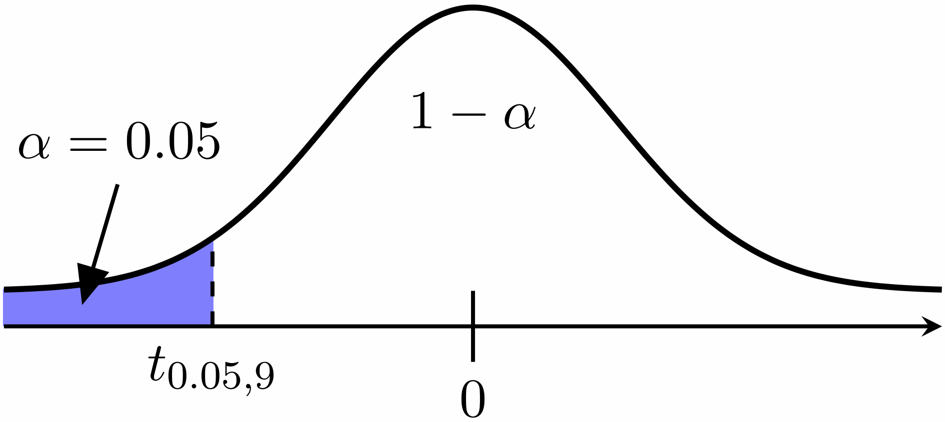
Decision for the One-Sided Test

Since

$$-3.45 < -1.833,$$

the test statistic falls in the rejection region.
Therefore, we reject H_0.
This suggests that the medicine lowers blood pressure.

Graph of a One-Sided Test



The blue area is the rejection region.

Example: Two-Sided Test

Problem: We now test whether the medicine changes blood pressure in either direction.

Hypotheses:

$$H_0 : \mu = 130, \qquad H_1 : \mu \neq 130.$$

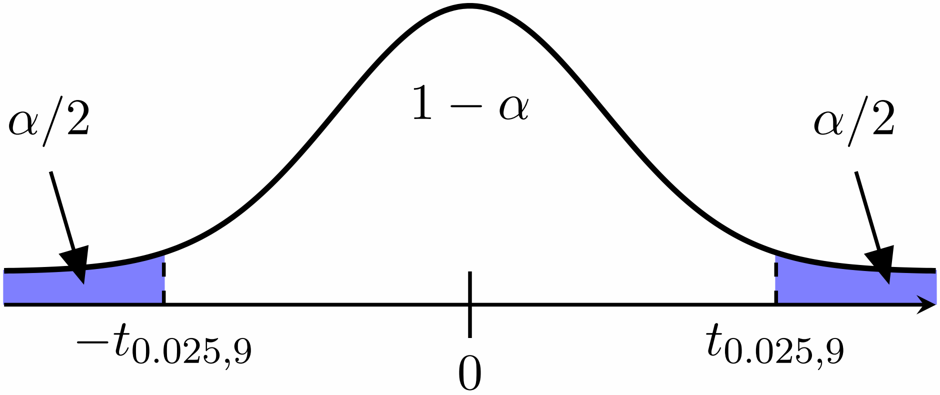
Data: 125, 128, 126, 130, 129, 127, 124, 123, 132, 126
Significance level:

$$\alpha = 0.05.$$

Calculation for the Two-Sided Test

1. $$\bar{X} = 127.0, \qquad S^2 \approx 7.56, \qquad SE \approx 0.87.$$
2. Test statistic:
$$T = \frac{127.0 - 130}{0.87} \approx -3.45.$$
3. Two-sided critical value with 9 degrees of freedom:
$$t_{0.025,9} \approx 2.262.$$
4. Since
$$|T| \approx 3.45 > 2.262,$$
we reject H_0.

Graph of a Two-Sided Test



The blue areas are the rejection regions.

Interpretation of a Two-Sided Test

In a two-sided test,

$$H_0 : \mu = 130$$

is rejected when the sample evidence suggests that μ is either significantly smaller or significantly larger than 130.
So both tails matter.

Setting

Assume

$$X \sim N(\mu_X, \sigma_X^2), \qquad Y \sim N(\mu_Y, \sigma_Y^2).$$

We are interested in whether the population means differ:

$$H_0 : \mu_X = \mu_Y, \qquad H_1 : \mu_X \neq \mu_Y.$$

Equivalently,

$$H_0 : \mu_X - \mu_Y = 0, \qquad H_1 : \mu_X - \mu_Y \neq 0.$$

Difference of Sample Means

Let

$$\bar{X} = \frac{1}{n_X}\sum_{i=1}^{n_X} X_i, \qquad \bar{Y} = \frac{1}{n_Y}\sum_{j=1}^{n_Y} Y_j.$$

Then

$$\mathbb{E}[\bar{X} - \bar{Y}] = \mu_X - \mu_Y,$$

and, assuming independence,

$$\mathrm{Var}(\bar{X} - \bar{Y}) = \frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}.$$

Standardised Difference

If the population variances are known, then

$$\frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \sim N(0, 1).$$

Under $H_0 : \mu_X - \mu_Y = 0$,

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \sim N(0, 1).$$

When the Population Variances Are Equal

Now assume

$$\sigma_X^2 = \sigma_Y^2 = \sigma^2.$$

Then we estimate the common variance by the pooled variance

$$S_p^2 = \frac{\sum_{i=1}^{n_X}(X_i - \bar{X})^2 + \sum_{j=1}^{n_Y}(Y_j - \bar{Y})^2}{n_X + n_Y - 2}.$$

The test statistic becomes

$$T = \frac{\bar{X} - \bar{Y}}{S_p\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \sim t(n_X + n_Y - 2).$$

When the Population Variances Are Unequal

If we do not assume equal population variances, we use Welch's t-test:

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}}.$$

Here

$$S_X^2 = \frac{1}{n_X - 1} \sum_{i=1}^{n_X} (X_i - \bar{X})^2, \qquad S_Y^2 = \frac{1}{n_Y - 1} \sum_{j=1}^{n_Y} (Y_j - \bar{Y})^2.$$

Welch–Satterthwaite Degrees of Freedom

In Welch's test, the degrees of freedom are approximated by

$$\nu \approx \frac{\left(\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}\right)^2}{\frac{(S_X^2/n_X)^2}{n_X-1} + \frac{(S_Y^2/n_Y)^2}{n_Y-1}}.$$

This is more reliable when the population variances differ.

Type I and Type II Errors

Two kinds of errors may occur in hypothesis testing:

- **Type I error:** reject H_0 when it is true.
- **Type II error:** fail to reject H_0 when it is false.

Their probabilities are denoted by

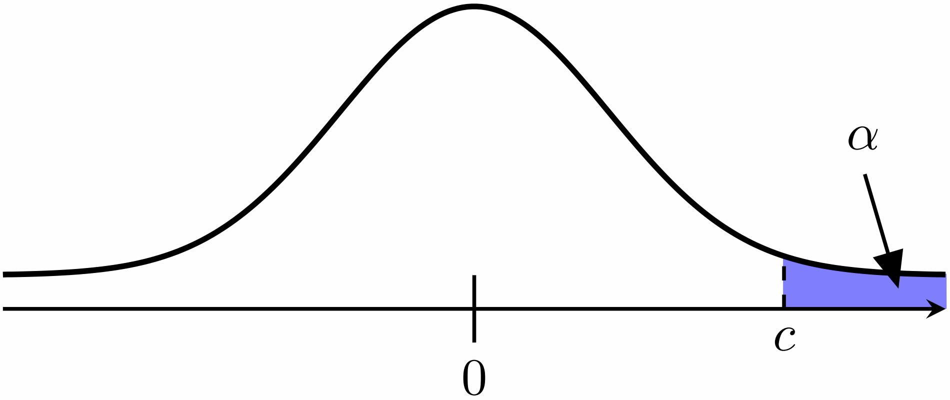
$$\alpha = \Pr(\text{Type I error}), \qquad \beta = \Pr(\text{Type II error}).$$

Trade-off Between Type I and Type II Errors

- Decreasing α makes the test more conservative.
- This usually reduces the probability of a Type I error.
- But it may increase the probability of a Type II error.

So there is often a trade-off between the two types of errors.

Visualising Type I Error



The shaded area represents the probability of rejecting H_0 when H_0 is true.

p-Value

The **p-value** is the probability, under H_0, of observing a test statistic at least as extreme as the one actually observed.

A small p-value means that the observed sample would be unlikely if H_0 were true.

Decision Rule Using the p-Value

- If
$$p\text{-value} < \alpha,$$
reject H_0.
- If
$$p\text{-value} \geq \alpha,$$
fail to reject H_0.

This is equivalent to using the rejection region, but often easier to report.

Interpretation of the p-Value

A p-value is **not** the probability that H_0 is true.
It is the probability of obtaining data at least as extreme as the observed data, assuming that H_0 is true.

Summary

In this lecture, we studied:

- the basic idea of hypothesis testing,
- null and alternative hypotheses,
- rejection regions and test statistics,
- one-sided and two-sided tests,
- tests for a population mean,
- tests for the difference of means,
- goodness-of-fit tests,
- ANOVA and the F-test,
- Type I and Type II errors,
- p-values.