Math Revision Session

Statistics (7): Hypothesis Testing

Jukina HATAKEYAMA

The University of Osaka, Department of Economics

June 18, 2025

1. The Idea of Hypothesis Testing
    - Example: Hypothesis Testing for a Population Mean
    - Steps in Hypothesis Testing
    - Interpreting Hypothesis Testing Results
2. Test Statistics
    - Introduction
    - Normal Distribution
3. One-Sided vs Two-Sided Tests
    - Introduction
    - One-Sided
    - Two-Sided
4. Testing the Difference of Means
    - When the Population Variances Are Equal
    - When the Population Variances Are Unequal
5. Type I and Type II Errors

1. The Idea of Hypothesis Testing
   - Example: Hypothesis Testing for a Population Mean
   - Steps in Hypothesis Testing
   - Interpreting Hypothesis Testing Results
2. Test Statistics
   - Introduction
   - Normal Distribution
3. One-Sided vs Two-Sided Tests
   - Introduction
   - One-Sided
   - Two-Sided
4. Testing the Difference of Means
   - When the Population Variances Are Equal
   - When the Population Variances Are Unequal
5. Type I and Type II Errors

Introduction to Hypothesis Testing

- **What is Hypothesis Testing?** Hypothesis testing is a statistical method used to make inferences about a population based on sample data. It helps determine whether there is enough evidence to reject a given hypothesis.
- **Null and Alternative hypothesis** In hypothesis testing, we set up two competing hypothesis:
  - **Null Hypothesis (H_0)**: Represents the status quo or a statement of no effect.
  - **Alternative Hypothesis (H_1)**: Represents what we seek evidence for.
- **Significance Level and Decision Rule** We set a significance level (α), which defines the probability of rejecting H_0 when it is actually true. The decision to reject or not reject H_0 is based on statistical evidence.
- **Types of Errors**
  - **Type I Error**: Rejecting H_0 when it is true (false positive).
  - **Type II Error**: Failing to reject H_0 when it is false (false negative).

Procedure

1. **Assume H_0 is True** Just as a defendant is presumed innocent until proven guilty, we assume H_0 is true until the data provides sufficient evidence against it.
2. **Collect Evidence (Sample Data)** We gather data and use statistical methods to evaluate whether the observed results are likely under H_0.
3. **Evaluate the Strength of Evidence**
   - If the evidence is **strong**, we reject H_0 and accept H_1 (alternative hypothesis).
   - If the evidence is **weak**, we do not reject H_0, meaning we do not have enough proof to support H_1.
4. **Make a Decision** The conclusion is based on the probability of observing the sample data assuming H_0 is true. If this probability (p-value) is too small, we reject H_0.

Hypothesis Testing for a Population Mean (Known Variance)

We illustrate hypothesis testing using an example where we test whether the 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 $\bar{X}$ serves as an estimator of the population mean.
- The sample mean is calculated as $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 0.01$.
- This sample mean varies across different samples.
- We test the null hypothesis:

$$H_0 : \mu = 0$$

against the alternative hypothesis:

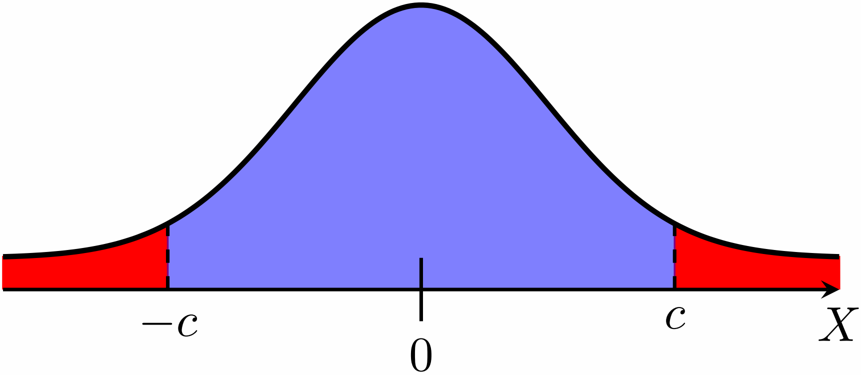
$$H_1 : \mu \neq 0$$

Steps in Hypothesis Testing

1. **Set up hypothesis:** Define the null hypothesis H_0 and the alternative hypothesis H_1.
   $\Rightarrow \quad H_0 : \mu = 0, \quad H_1 : \mu \neq 0$
2. **Choose a significance level:** Select α, the probability of rejecting H_0 when it is true.
   $\Rightarrow \quad 0.95 = 1 - \alpha$
3. **Compute the test statistic:** Standardise the sample mean under H_0:

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

4. **Determine the critical region:** Find the critical values based on the chosen α.
5. **Make a decision:** If the test statistic falls in the rejection region, we can reject H_0; otherwise, cannot reject H_0.



$$Pr(T < -c) + Pr(T > c) = \alpha = 0.05$$

- The probability of obtaining a value in the red area is low.
- The sample mean $\bar{X}$ is more likely to take values near zero.

- The true value of μ is unknown, but if the null hypothesis $H_0 : \mu = 0$ is correct, the sample mean $\bar{X}$ is likely to take values close to zero.
- Under this assumption, the test statistic T, which serves as a measure of deviation from the null hypothesis, should also take values near zero with high probability.
- The probability of observing a test statistic far from zero is small if H_0 is true.
- This can be visualised in the previous figure, where values closer to zero occur more frequently, while values further away are less probable.
- Hypothesis testing evaluates whether the observed value of T is unusually far from zero, leading us to either reject or fail to reject H_0.

Interpreting Hypothesis Testing Results

- In hypothesis testing, we do not **accept** the null hypothesis H_0; rather, we assess whether we have enough evidence to **reject** it.
- A failure to reject H_0 does not mean H_0 is true—only that the data do not provide sufficient evidence against it.
- This is because statistical tests assess whether the observed data are **unlikely** under H_0, not whether H_0 is definitively correct.
- Even if H_0 is false, a test might fail to reject it due to insufficient sample size or variability in the data.
- Therefore, hypothesis testing conclusions are framed in terms of rejection or failure to reject, rather than acceptance.

1 The Idea of Hypothesis Testing
Example: Hypothesis Testing for a Population Mean
Steps in Hypothesis Testing
Interpreting Hypothesis Testing Results

2 Test Statistics
Introduction
Normal Distribution

3 One-Sided vs Two-Sided Tests
Introduction
One-Sided
Two-Sided

4 Testing the Difference of Means
When the Population Variances Are Equal
When the Population Variances Are Unequal

5 Type I and Type II Errors

Introduction to Test Statistics

- A **test statistic** is a function of the sample data used to assess the validity of a hypothesis.
- It summarises the information in the sample relevant to testing the null hypothesis H_0.
- The test statistic is designed so that its distribution under H_0 is known, allowing us to quantify how extreme the observed value is.
- Large deviations of the test statistic from its expected value under H_0 provide evidence against H_0.
- The choice of test statistic depends on the hypothesis being tested and the underlying data distribution.

When the Population Distribution is Normal

Assume X_i follows $N_{\mathbb{R}}(\mu, \sigma^2)$. Then, the sample mean $\bar{X}$ satisfies below:

$$\bar{X} \sim N_{\mathbb{R}}(\mu, \frac{\sigma^2}{n}) \quad \Rightarrow \quad \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim N_{\mathbb{R}}(0, 1)$$

If we do not know the population variance, then we replace σ^2 by sample variance S^2 which is the unbiased estimator of population variance:

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

By the distribution assumption, we have:

$$\Rightarrow \frac{\bar{X} - \mu}{\sqrt{S^2/n}} \sim t(n-1)$$

If the null hypothesis $H_0 : \mu = 0$ is correct, then the **test statistic** should be as follows:

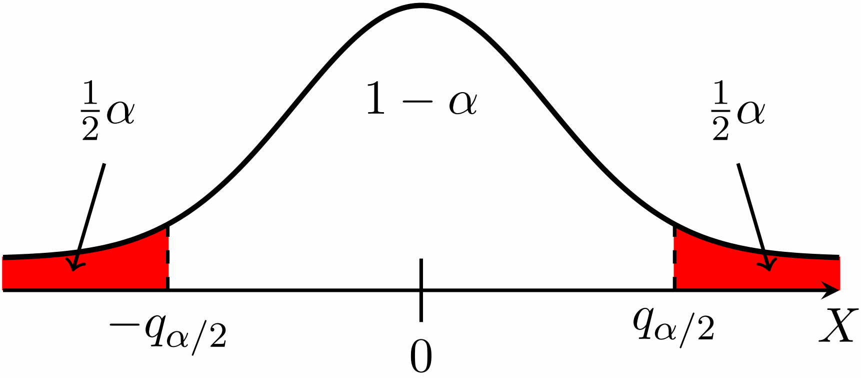
$$\frac{\bar{X}}{\sqrt{S^2/n}} \sim t(n-1)$$

Now, we consider the rejection area. First, we set the confidence level:

$$\alpha = Pr\left(\frac{\bar{X}}{\sqrt{S^2/n}} < -q_{\alpha/2}\right) + Pr\left(\frac{\bar{X}}{\sqrt{S^2/n}} > q_{\alpha/2}\right)$$

or

$$1 - \alpha = Pr\left(-q_{\alpha/2} < \frac{\bar{X}}{\sqrt{S^2/n}} < q_{\alpha/2}\right)$$



The red area is the rejection area. If the test statistic is included in the rejection area, then we can reject the null hypothesis.

Brief Example

We set null hypothesis, $H_0 : \mu = 0$, and alternative hypothesis, $H_1 : \mu \neq 0$. In addition, we decide the confidence level, $\alpha = 0.05$. If we have 10 observations, the sample mean is 0.01, and sample variance is 0.007, then we have:

$$-q_{(t,0.025)} < -2.262 < \frac{0.01}{\sqrt{\frac{0.007}{10}}} = 0.38 < q_{(t,0.025)} < 2.262$$

which means that the test statistic is not included in the rejection area. Therefore we cannot reject the null hypothesis.

1. The Idea of Hypothesis Testing
    - Example: Hypothesis Testing for a Population Mean
    - Steps in Hypothesis Testing
    - Interpreting Hypothesis Testing Results
2. Test Statistics
    - Introduction
    - Normal Distribution
3. One-Sided vs Two-Sided Tests
    - Introduction
    - One-Sided
    - Two-Sided
4. Testing the Difference of Means
    - When the Population Variances Are Equal
    - When the Population Variances Are Unequal
5. Type I and Type II Errors

One-Sided vs Two-Sided Tests

In hypothesis testing, we aim to make inferences about a population based on a sample. There are two main types of tests: one-sided tests and two-sided tests.

- **One-Sided Test:** This test examines if a parameter is either greater than or less than a certain value. It focuses on detecting an effect in one direction.
- **Two-Sided Test:** This test evaluates if a parameter is significantly different from a specific value, without specifying a direction. It considers both possibilities: greater than and less than.

The choice between one-sided and two-sided tests depends on the research question and the hypotheses being tested.

Example: One-Sided Test

Problem Statement:
A new medicine is believed to have a lowering effect on blood pressure. The average blood pressure before treatment is known to be 130 mmHg. We want to test whether the blood pressure of patients treated with the new medicine is significantly lower.
Hypotheses:

- Null Hypothesis (H_0): $\mu \geq 130$ (The medicine has no effect or blood pressure is 130 mmHg or higher)
- Alternative Hypothesis (H_1): $\mu < 130$ (The medicine lowers blood pressure to below 130 mmHg)

Data: Blood pressure measurements (in mmHg) from 10 patients treated with the new medicine: 125, 128, 126, 130, 129, 127, 124, 123, 132, 126
Significance Level: $\alpha = 0.05$

1. Calculate the sample mean ($\bar{X}$) and sample standard Error (SE).

$$\Rightarrow \quad \bar{X} = 126.0, \quad S^2 \simeq 8.89$$

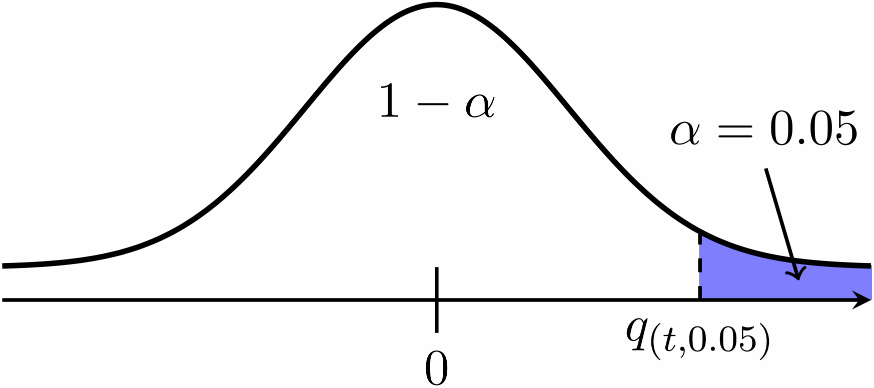
$$\Rightarrow \quad SE \simeq 0.94$$

2. Calculate the test statistic:

$$\frac{\bar{x} - 130}{\sqrt{S^2/n}} = \frac{126.0 - 130}{0.94} \simeq -4.26$$

3. Comper the test statistic and the quantile $q_{(t,0.05)}$:

$$\frac{\bar{x} - 130}{\sqrt{S^2/n}} \simeq -4.26 < 1.833 \simeq q_{(t,0.05)}$$



The blue area represents the rejection area for the null hypothesis.

Example: Two-Sided Test

Problem Statement:
A new medicine is believed to have an effect on blood pressure. The average blood pressure before treatment is known to be 130 mmHg. We want to test whether the blood pressure of patients treated with the new medicine is significantly different from 130 mmHg.
Hypotheses:

- Null Hypothesis (H_0): $\mu = 130$ (The medicine has no effect on blood pressure)
- Alternative Hypothesis (H_1): $\mu \neq 130$ (The medicine has an effect on blood pressure)

Data: Blood pressure measurements (in mmHg) from 10 patients treated with the new medicine: 125, 128, 126, 130, 129, 127, 124, 123, 132, 126
Significance Level: $\alpha = 0.05$

1. Calculate the sample mean $(\bar{X})$ and sample standard error (SE).

$$\Rightarrow \quad \bar{X} = 126.0, \quad S^2 \simeq 8.89$$

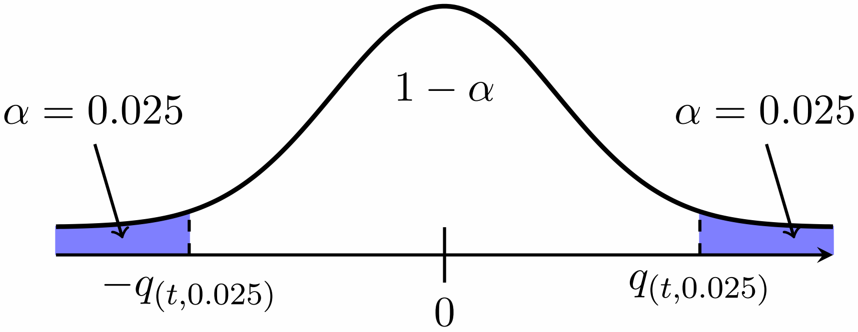
$$\Rightarrow \quad SE \simeq 0.94$$

2. Calculate the test statistic:

$$t = \frac{\bar{X} - 130}{SE} = \frac{126.0 - 130}{0.943} \approx -4.24$$

3. Compare the test statistic and the critical values $q_{(t,0.025)}$:

$$|t| \approx 4.24 > 1.833 \quad (q_{(t,0.025)})$$



The blue areas represent the rejection areas for the null hypothesis.

Understanding the Null Hypothesis:

- In the context of our test, the null hypothesis is defined as:

$$H_0 : \mu = 130$$

- If the test statistic indicates that the sample mean is significantly different from 130, we reject the null hypothesis.
- However, if the sample mean is exactly 130, we do not have enough evidence to reject the null hypothesis.
- This means:
  - A sample mean greater than 130 could lead to rejection of H_0.
  - A sample mean less than 130 could also lead to rejection of H_0.
  - But a sample mean equal to 130 indicates that H_0 cannot be rejected.

The null hypothesis serves as a baseline assumption, and we look for evidence to determine if there is a significant difference from this baseline.

1 The Idea of Hypothesis Testing
Example: Hypothesis Testing for a Population Mean
Steps in Hypothesis Testing
Interpreting Hypothesis Testing Results

2 Test Statistics
Introduction
Normal Distribution

3 One-Sided vs Two-Sided Tests
Introduction
One-Sided
Two-Sided

4 Testing the Difference of Means
When the Population Variances Are Equal
When the Population Variances Are Unequal

5 Type I and Type II Errors

Setting

We assume that X and y are random variables such that:

$$X \sim N_{\mathbb{R}}(\mu_x, \sigma_x^2), \quad Y \sim N_{\mathbb{R}}(\mu_y, \sigma_y^2).$$

Our interest is whether there exists population mean differences. According to this motivation, we set the null hypothesis and alternative hypothesis:

$$H_0 : \mu_x = \mu_y, \quad H_1 : \mu_x \neq \mu_y.$$

We can rewrite above hypotheses as:

$$H_0 : \mu_x - \mu_y = 0, \quad H_1 : \mu_x - \mu_y \neq 0.$$

This expression is closer to our interest.

We estimate each μ_x and μ_y by their unbiased estimator:

$$\bar{X} = \frac{1}{n_x} \sum_{i=1}^{n_x} X_i \sim N(\mu_x, \frac{\sigma_x^2}{n_x})$$

$$\bar{Y} = \frac{1}{n_y} \sum_{i=1}^{n_y} Y_i \sim N(\mu_y, \frac{\sigma_y^2}{n_y})$$

To examine the difference, we use $\bar{X} - \bar{Y}$ and standardise it. Its mean and variance are derived as follows:

$$E[\bar{X} - \bar{Y}] = \mu_x - \mu_y,$$

$$Var[\bar{X} - \bar{Y}] = \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}.$$

The standardised difference is:

$$\frac{(\bar{X}-\bar{Y})-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} \sim N(0,1)$$

Under H_0, we have

$$\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} \sim N(0,1)$$

All we need to do is calculate the test statistic, set the confidence level, check the t-distribution table, and comper the test statistic and quantile derived from the table.

When the Population Variances Are Equal

Now, we assume $\sigma^2 = \sigma_x^2 = \sigma_y^2$.Then we have:

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma^2}{n_x} + \frac{\sigma^2}{n_y}}} \sim N(0, 1)$$

The sample variance is:

$$S^2 = \frac{1}{n_x + n_y - 2} \left\{ \sum_{i=1}^{n_x} (X_i - \bar{X})^2 + \sum_{j=1}^{n_y} (Y_j - \bar{Y})^2 \right\}$$

Since we use two sample mean to calculate S^2, denominator should be $n_x + x_y - 2$. Finally, we have:

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S^2}{n_x} + \frac{S^2}{n_y}}} \sim t(n_x + n_y - 2)$$

When the Population Variances Are Unequal

Now, we do not assume $\sigma_x^2 = \sigma_y^2$. In this case, the standardised test statistic follows:

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}} \sim t(v)$$

where S_x^2 and S_y^2 are the sample variances:

$$S_x^2 = \frac{1}{n_x - 1} \sum_{i=1}^{n_x} (X_i - \bar{X})^2, \quad S_y^2 = \frac{1}{n_y - 1} \sum_{j=1}^{n_y} (Y_j - \bar{Y})^2.$$

The degrees of freedom v is approximated using the Welch-Satterthwaite equation:

$$v \approx \frac{\left(\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}\right)^2}{\frac{(S_x^2/n_x)^2}{n_x - 1} + \frac{(S_y^2/n_y)^2}{n_y - 1}}.$$

This is known as Welch's t-test, which is more reliable when population variances are different.

Goodness-of-Fit Test

The Goodness-of-Fit Test is used to determine whether an observed frequency distribution fits an expected distribution.

Hypotheses:

- Null Hypothesis (H_0): The observed data follows the expected distribution.
- Alternative Hypothesis (H_1): The observed data does not follow the expected distribution.

The test is commonly performed using the χ^2 (Chi-Square) test.

Chi-Square Test Statistic

The test statistic is calculated as:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}, \tag{1}$$

where:

- O_i = Observed frequency for category i.
- E_i = Expected frequency for category i.

The test statistic follows a Chi-Square distribution with $(k-1)$ degrees of freedom, where k is the number of categories.

Decision Rule

- Choose a significance level α (e.g., 0.05).
- Find the critical value from the Chi-Square distribution table for $(k-1)$ degrees of freedom.
- If χ^2 exceeds the critical value, reject H_0.
- Otherwise, do not reject H_0.

Test for Independence (Contingency Table)

The chi-square test for independence is used to determine whether two categorical variables are independent of each other.

Hypotheses:

- H_0: The two categorical variables are independent.
- H_1: The two categorical variables are not independent.

Example: Suppose we conduct a survey on 200 individuals about their preference for coffee or tea and whether they are male or female. The data is presented in a contingency table:

	Coffee	Tea	Total
Male	50	30	80
Female	70	50	120
Total	120	80	200

We test whether the preference for coffee or tea is independent of gender.

Chi-Square Test Statistic

The test statistic is calculated as:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

where:

- $O =$ observed frequency from the contingency table
- $E =$ expected frequency under the assumption of independence

The expected frequency is calculated as:

$$E = \frac{\text{row total} \times \text{column total}}{\text{grand total}}$$

The chi-square statistic follows a chi-square distribution with degrees of freedom:

$$df = (r - 1)(c - 1)$$

where r is the number of rows and c is the number of columns.

Decision Rule

- Set the significance level α (e.g., 0.05).
- Compare the computed χ^2 statistic with the critical value from the chi-square distribution table.
- If χ^2 exceeds the critical value, reject H_0 (variables are not independent).
- Otherwise, do not reject H_0 (no sufficient evidence to conclude dependence).

Analysis of Variance (ANOVA)

Analysis of Variance (ANOVA) is a statistical method used to compare means among multiple groups to determine if at least one group mean is significantly different from the others.

Hypotheses:

- Null Hypothesis (H_0): All group means are equal.
- Alternative Hypothesis (H_1): At least one group mean is different.

Applications:

- Comparing the effectiveness of different treatments.
- Evaluating differences among different population groups.
- Used in experimental design and quality control.

One-Way ANOVA

Model:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

where:

- Y_{ij} is the observation from group i and sample j.
- μ is the overall mean.
- α_i is the effect of group i.
- ε_{ij} is the random error.

Assumptions:

- Independence of observations.
- Normally distributed errors.
- Homogeneity of variances across groups.

F-Test in ANOVA

Test Statistic:

$$F = \frac{\text{Between-group variance}}{\text{Within-group variance}}$$

Decision Rule:

- Compare the calculated F value with the critical value from the F-distribution.
- If F is greater than the critical value, reject H_0.

Interpretation:

- If H_0 is rejected, at least one group mean is significantly different.
- Post-hoc tests (e.g., Tukey's test) can identify which groups differ.

1 The Idea of Hypothesis Testing
Example: Hypothesis Testing for a Population Mean
Steps in Hypothesis Testing
Interpreting Hypothesis Testing Results

2 Test Statistics
Introduction
Normal Distribution

3 One-Sided vs Two-Sided Tests
Introduction
One-Sided
Two-Sided

4 Testing the Difference of Means
When the Population Variances Are Equal
When the Population Variances Are Unequal

5 Type I and Type II Errors

Type I and Type II Errors

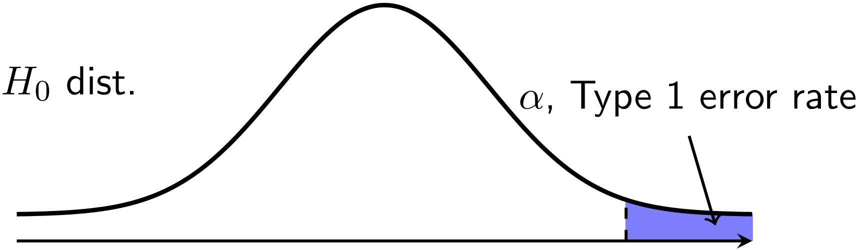
When performing hypothesis testing, there are two types of errors that can occur:

- **Type I Error (False Positive)**
  - Occurs when we reject the null hypothesis H_0 when it is actually true.
  - The probability of making a Type I error is denoted by α (significance level).
  - Example: Concluding that a new drug is effective when it actually has no effect.
- **Type II Error (False Negative)**
  - Occurs when we fail to reject the null hypothesis H_0 when it is actually false.
  - The probability of making a Type II error is denoted by β.
  - Example: Concluding that a new drug has no effect when it is actually effective.

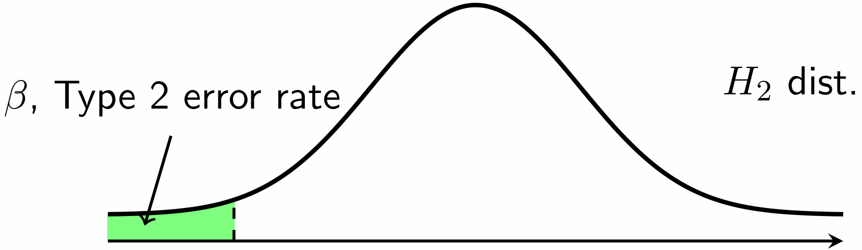
Trade-off Between Type I and Type II Errors

- Decreasing α (making the test more stringent) reduces the likelihood of a Type I error but increases the likelihood of a Type II error.
- Increasing α (making the test more lenient) reduces the likelihood of a Type II error but increases the likelihood of a Type I error.
- The choice of α depends on the context and the consequences of each type of error.

Probability of making a Type 1 error:



Probability of making a Type 2 error:



Probability of making a Type 2 error: