# Hypothesis Testing

# Agenda

- Introduction to Hypothesis Testing
- Types of Hypotheses
- The Hypothesis Testing Process
- Significance Level and P-Value
- Z-Tests and T-Tests
- Choosing the Right Test
- Conducting Hypothesis Tests and Interpreting Results
- Type I and Type II Errors
- Common Mistakes and Pitfalls
- Ethical Considerations

# Introduction to Hypothesis Testing

- Hypothesis testing is a statistical method used to make inferences about population parameters based on sample data.
- The primary purpose of hypothesis testing is to evaluate whether there is enough evidence in a sample to draw conclusions or make decisions about a population.
- It helps researchers and decision-makers assess claims, test hypotheses, and make informed choices based on data.
- Hypothesis testing plays a crucial role in research by providing a structured and objective method to determine if observed differences or effects are statistically significant or if they could have occurred by random chance.
- Decision-makers in various fields, such as business, healthcare, and public policy, use hypothesis testing to assess the impact of decisions, test the effectiveness of interventions, and make data-driven choices that can lead to improved outcomes.

# Types of Hypotheses

**1. Null Hypothesis (H0):**

- The null hypothesis, often denoted as H0, is a statement of no effect or no difference.
- It represents the status quo, the default assumption, or the hypothesis that there is no significant change or effect in a population.
- For example, in a drug effectiveness study, the null hypothesis might state that the drug has no impact on the condition being treated.

# Types of Hypotheses

**2. Alternative Hypothesis (Ha):**

- The alternative hypothesis, often denoted as Ha or H1, is a statement that contradicts the null hypothesis.
- It represents the researcher's claim or the hypothesis that there is a significant change or effect in a population.
- Using the previous example, the alternative hypothesis could assert that the drug is effective and does have a significant impact on the condition.

# Types of Hypotheses

**3. One-Tailed Hypothesis:**

- A one-tailed hypothesis, also known as a directional hypothesis, specifies the direction of the effect being tested.
- It is used when researchers are interested in determining if a parameter is greater than or less than a certain value, but not both.
- For instance, in a one-tailed test, you might be testing whether a new manufacturing process increases product quality (greater than a certain standard) or decreases it (less than a certain standard).

# Types of Hypotheses

**4. Two-Tailed Hypothesis:**

- A two-tailed hypothesis, also known as a non-directional hypothesis, does not specify the direction of the effect being tested.
- It is used when researchers want to determine if a parameter is different from a certain value, without specifying whether it is greater or less than that value.
- In a two-tailed test, you might be testing whether a new product's weight is different from the claimed weight, without assuming it's either heavier or lighter.

# One-Tailed vs Two-Tailed

**Example 1: One-Tailed Hypothesis**

Scenario: A pharmaceutical company develops a new painkiller and wants to test if it's more effective than the existing painkiller in reducing pain duration.

In this one-tailed test, the researchers are specifically interested in whether the new painkiller is more effective (i.e., reduces pain duration more) than the existing one. They are not interested in whether it's less effective.

**Example 2: Two-Tailed Hypothesis**

Scenario: A manufacturer produces light bulbs and claims that the average lifespan of their bulbs is 1000 hours.

In this two-tailed test, researchers want to determine if the average lifespan of the bulbs is significantly different from the manufacturer's claim of 1000 hours. They are not specifying whether it's longer or shorter, just different.

# The Hypothesis Testing Process

## Step 1: State the Null and Alternative Hypotheses

**Null Hypothesis (H0):** This is the default or status quo hypothesis. It often states that there is no effect, no difference, or no change in the population parameter being tested. It represents the hypothesis that you want to challenge or investigate.

Example: H0: The average weight of a certain breed of dogs is 30 kilograms.

**Alternative Hypothesis (Ha):** This is the hypothesis you are testing or trying to support. It can be directional (one-tailed) or non-directional (two-tailed), depending on whether you are looking for a specific change in one direction or simply any change from the null hypothesis.

Example (one-tailed): Ha: The average weight of a certain breed of dogs is greater than 30 kilograms.

Example (two-tailed): Ha: The average weight of a certain breed of dogs is not equal to 30 kilograms.

# The Hypothesis Testing Process

## Step 2: Choose a Significance Level (Alpha, α)

The significance level, denoted as α, represents the probability of making a Type I error (rejecting a true null hypothesis). Commonly used values for α are 0.05 and 0.01, but it can vary depending on the study's requirements and conventions.

Example: Let's choose α = 0.05 for our dog weight study.

## Step 3: Collect Data or Sample

This step involves gathering data from your study or experiment. The data should be representative of the population you are investigating.

Example: You collect weight measurements from a random sample of 50 dogs of the specified breed.

# The Hypothesis Testing Process

## Step 4: Conduct a Statistical Test

In this step, you perform a statistical test using the collected data. The choice of the test depends on the nature of your data (e.g., t-test, chi-square test, ANOVA, etc.) and the type of hypothesis you are testing (e.g., comparing means, proportions, associations).

Example: You conduct a one-sample t-test to determine if the average weight of the sampled dogs is significantly different from 30 kilograms.

# The Hypothesis Testing Process

## Step 5: Interpret the Results

Based on the statistical test, you obtain results such as a test statistic and a p-value. The p-value represents the probability of observing the data if the null hypothesis is true.

To make a decision, compare the p-value to the chosen significance level (α):

- If **p-value ≤ α**: Reject the null hypothesis. There is enough evidence to support the alternative hypothesis.
- If **p-value > α**: Fail to reject the null hypothesis. There isn't enough evidence to support the alternative hypothesis.

Example: If the calculated p-value is 0.03 (which is less than α = 0.05), you would reject the null hypothesis and conclude that the average weight of the sampled dogs is significantly different from 30 kilograms.

# Significance Level and P-Value

- The significance level, denoted as α (alpha), is a **predetermined threshold** used in hypothesis testing. It is the maximum allowable probability of making a Type I error, which is the error of rejecting a true null hypothesis.
- In simpler terms, α represents the level of tolerance for false positives in a hypothesis test. It sets the bar for how strong the evidence must be against the null hypothesis for you to reject it.
- Commonly used values for α include 0.05 (5%) and 0.01 (1%), but the choice of α can vary depending on the context and the level of risk associated with making a Type I error. A smaller α indicates a lower tolerance for false positives but may require stronger evidence to reject the null hypothesis.

# Significance Level and P-Value

- The p-value is a statistical measure that quantifies the strength of evidence against the null hypothesis based on the observed data from a hypothesis test.
- The p-value represents the probability of obtaining a result as extreme or more extreme than the one observed, assuming that the null hypothesis is true. In other words, it tells you how likely it is to see the data you have if the null hypothesis is correct.
- A **small p-value (typically less than α) suggests strong evidence against the null hypothesis**. It indicates that the observed data is unlikely to occur by random chance alone, leading to the rejection of the null hypothesis in favor of the alternative hypothesis.
- A **large p-value (greater than α) suggests weak evidence against the null hypothesis**. It indicates that the observed data could plausibly occur under the assumptions of the null hypothesis, leading to a failure to reject the null hypothesis.

# Significance Level and P-Value

The significance level (α) and the p-value are closely related in hypothesis testing:

- If the p-value is less than or equal to α ($p \leq \alpha$), **you reject the null hypothesis.** This indicates that the observed data provides strong evidence against the null hypothesis, and you conclude in favor of the alternative hypothesis.
- If the p-value is greater than α ($p > \alpha$), **you fail to reject the null hypothesis.** This suggests that the observed data is not strong enough to warrant rejecting the null hypothesis, and you do not conclude in favor of the alternative hypothesis.

# Z-Tests and T-Tests

**Z-Test for Population Means (Known Population Standard Deviation):**

The Z-test is a statistical test used to compare a sample mean to a known population mean when you know the population standard deviation (σ).

It is typically used when you have a large sample size (usually $n > 30$) and can assume that the sample mean follows a normal distribution.

Example:

Suppose you want to test if the average height of students in your school is different from the national average height (with a known standard deviation) of 65 inches. You have a sample of 100 students' heights

# Z-Tests and T-Tests

**T-Test for Population Means (Unknown Population Standard Deviation)**

The T-test is used to compare a sample mean to a hypothesized population mean when you do not know the population standard deviation (σ) and have a small to moderate sample size.

It accounts for the additional uncertainty introduced by estimating the population standard deviation from the sample standard deviation.

Suppose you want to test if a new teaching method significantly improves students' test scores compared to the traditional method. You have two groups of students, each taught with one of the methods, and you want to compare the means. Since you don't know the population standard deviation, you use a T-test.

# Z-Tests and T-Tests

**Paired Samples T-Test:**

The paired samples T-test, also known as the dependent samples T-test, is used when you want to compare the means of two related groups or conditions.

It is often used in before-and-after studies or when comparing matched pairs of data.

Example:

You want to determine if a weight loss program is effective. You measure the weight of the same group of individuals before and after the program. The paired samples T-test helps you assess if there is a significant difference in weight before and after the program.

# When to Choose Z-Tests vs. T-Tests:

Use a Z-test when:

- You have a large sample size (typically $n > 30$).
- You know the population standard deviation ($\sigma$).
- Your data approximately follows a normal distribution.

Use a T-test when:

- You have a small to moderate sample size (usually $n < 30$).
- You do not know the population standard deviation ($\sigma$).
- Your data may not perfectly follow a normal distribution, but it should be reasonably close.

Use a paired samples T-test when:

- You have paired or matched data, such as before-and-after measurements.
- You want to compare the means of two related groups or conditions.

# When to Use a Z-Test:

1. **Large Sample Size with Known Population Standard Deviation:**

Research Question: "Is the average salary of employees in my company different from the national average salary, with a known national average and population standard deviation?"

Use a Z-test because you have a large sample size and know the population standard deviation.

2. **Proportions Testing:**

Research Question: "Is the proportion of customers who prefer Product A different from 0.50?"

Use a Z-test for proportions when you're comparing a sample proportion to a known or hypothesized population proportion.

# When to Use a T-Test:

1. **Small to Moderate Sample Size with Unknown Population Standard Deviation:**

Research Question: "Does a new drug result in a significant change in blood pressure compared to a control group?"

Use an independent samples T-test because you have a small to moderate sample size and don't know the population standard deviation.

2. **Paired Data (Before-and-After Comparisons):**

Research Question: "Is there a significant difference in students' test scores before and after a tutoring program?"

Use a paired samples T-test to compare the means of paired or dependent data.

# When to Use a T-Test:

3. **Comparing Two Independent Groups with Equal Variances:**

Research Question: "Is there a significant difference in the time taken to complete a task between two different training methods when you believe the variances in both groups are equal?"

Use an independent samples T-test, assuming equal variances.

4. **Comparing Two Independent Groups with Unequal Variances:**

Research Question: "Is there a significant difference in the scores of two groups of students when you suspect that the variances in the two groups are different?"

Use an independent samples T-test with Welch's correction for unequal variances.

# Conducting Hypothesis Tests:

1. Formulas for calculating test statistics.
2. Critical values and rejection regions.
3. Calculating and interpreting the test statistic.

# 1. Hypothesis Tests

**Z-Test for Population Means (Known Population Standard Deviation):**

Formula for Test Statistic (Z):

Test Statistic when $\sigma$ known

$$z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$$

$\bar{x}$ : sample mean

$\mu$ : population mean

$\sigma$ : population standard deviation

$n$ : sample size

# 1. Hypothesis Tests

**T-Test for Population Means (Unknown Population Standard Deviation):**

Formula for Test Statistic (t):

Where $\bar{x}$ is the sample mean, μ is the hypothesized population mean, s is the sample standard deviation, and n is the sample size.

Test Statistic when $\sigma$ unknown

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)}, \quad df = n - 1$$

$\bar{x}$ : sample mean
$\mu$ : population mean
$s$ : sample standard deviation
$n$ : sample size

# 1. Hypothesis Tests

**Z-Test for Population Proportions:** When you want to compare a single sample proportion to a known or hypothesized population proportion.

where p̂ is the sample proportion, p is the hypothesized population proportion, and n is the sample size.

$$\textit{test statistic}: \quad z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

$$\hat{p} = \textit{sample proportion}$$

$$p = \textit{population proportion}$$

$$n = \textit{sample size}$$

# 1. Hypothesis Tests

Z-Test for Sample Proportions:This test is used when you want to determine whether there is a significant difference between two proportions.

Two proportion test

- **Pooled** : two populations being compared have the same population proportion/variance
- **Unpooled** : two populations being compared do not have the same population proportion/variance

# Hypothesis Tests

**Two Proportions Z Test with Pooled Approach** :The pooled proportion is the weighted average of the proportions of the two samples. It is used in the two-proportions z-test with a pooled approach to estimate the population proportion when the population variances of the two samples are assumed to be equal.

$$p_{pooled} = \frac{p1 \cdot n1 + p2 \cdot n2}{n1 + n2}$$

$$z = \frac{(p_1 - p_2)}{\sqrt{\frac{p_{pooled}(1-p_{pooled})}{n_1} + \frac{p_{pooled}(1-p_{pooled})}{n_2}}}$$

# Hypothesis Tests

**Two Proportions Z Test with Unpooled Approach:** Where p1 is the proportion of the first sample, p2 is the proportion of the second sample, n1 is the size of the first sample, and n2 is the size of the second sample.

$$z = \frac{(p1 - p2)}{\sqrt{\frac{p1(1-p1)}{n1} + \frac{p2(1-p2)}{n2}}}$$

| Test for | $H_0$ | Test statistic | Use when |
|---|---|---|---|
| Pop. mean $\mu$ | $\mu = \mu_0$ | $z = \dfrac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ | Normal dist. or $n > 30$, $\sigma$ known |
| Pop. mean $\mu$ | $\mu = \mu_0$ | $t = \dfrac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$ | $n < 30$ and/or $\sigma$ unknown |
| Pop. prop. $p$ | $p = p_0$ | $z = \dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$ | $n\hat{p} \geq 10$, $n(1 - \hat{p}) \geq 10$ |

# 2. Critical Values and Rejection Regions:

- Critical values are values from a standard normal (Z) or t-distribution that correspond to the chosen significance level (α) and degrees of freedom (df).
- They define the boundary for rejecting or failing to reject the null hypothesis.
- For one-tailed tests, critical values are typically found in one tail of the distribution.
- For two-tailed tests, they are divided into both tails.

# 3. Calculating and Interpreting the Test Statistic

Calculate the test statistic using the appropriate formula based on the type of test (Z-test or T-test).

Once you have the test statistic, compare it to the critical values or determine if it falls within the rejection region.

Interpret the results:

- If the test statistic falls in the rejection region, reject the null hypothesis.
- If the test statistic does not fall in the rejection region, fail to reject the null hypothesis.

# Example - Z-Test for Population Means:

Suppose you want to test if the average IQ score of students in your school is significantly different from the national average IQ score of 100 with a known population standard deviation of 15. You have a sample of 30 students with a sample mean IQ score ($\bar{X}$) of 105.

# Example - Z-Test for Population Means:

1. **Calculate the Test Statistic (Z):**

   $Z = (105 - 100) / (15/\sqrt{30}) = 5 / (15/\sqrt{30}) \approxeq 3.06$

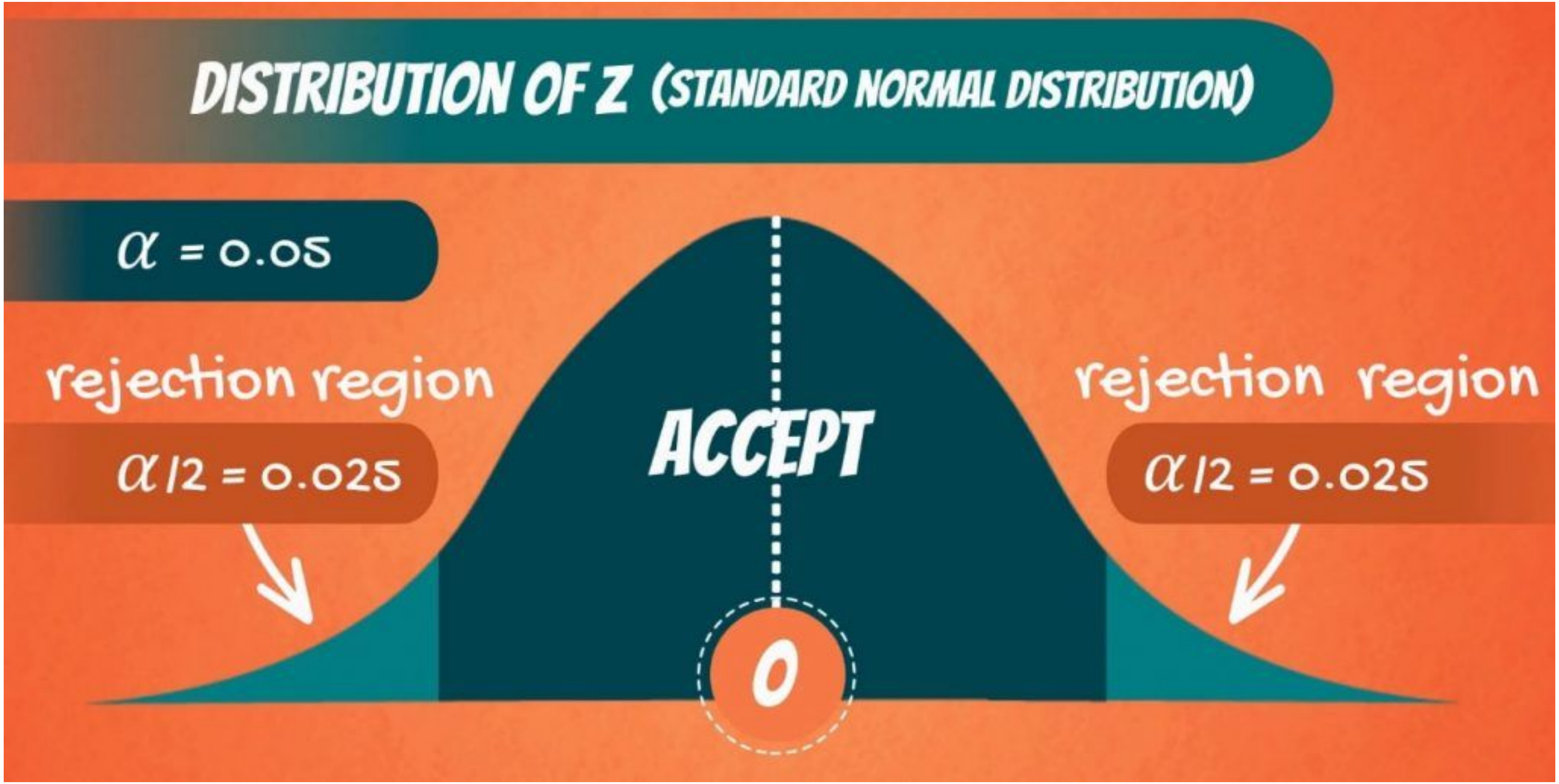
2. **Determine Critical Values ($\alpha$ = 0.05, two-tailed test):**

   Using a standard normal distribution table or calculator, you find the critical values for $\alpha/2$ (0.025 in each tail). For a 95% confidence interval, critical values are approximately ±1.96.

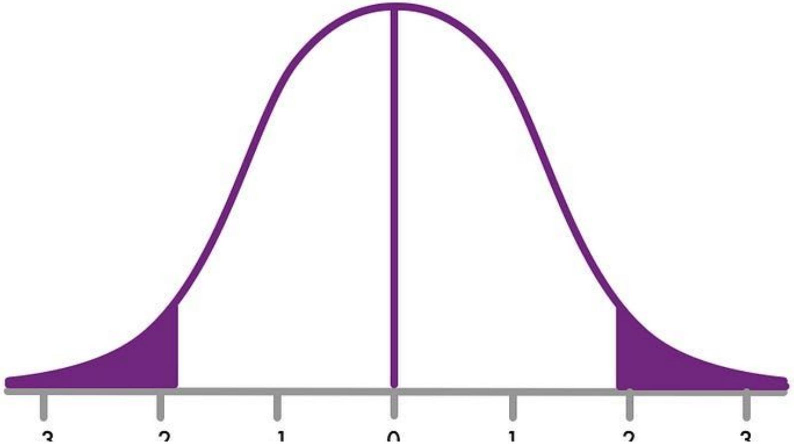
3. **Interpret the Results:**

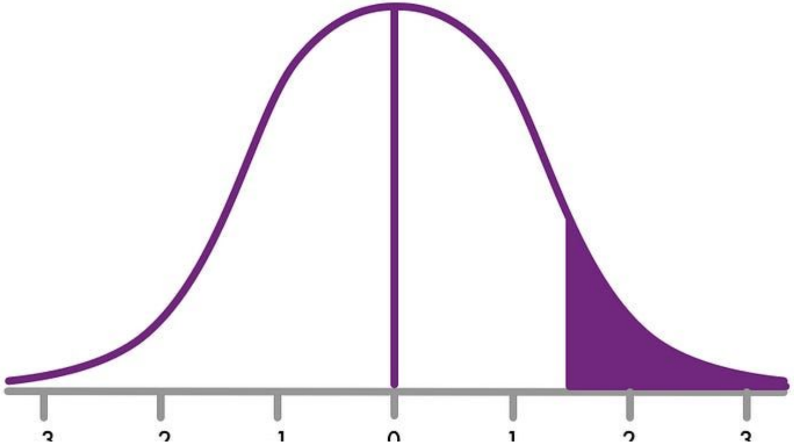
   The calculated test statistic (Z = 3.06) is greater than the critical value (±1.96). Therefore, you reject the null hypothesis.

Conclusion: There is enough evidence to suggest that the average IQ score of students in your school is significantly different from the national average IQ score of 100.


DISTRIBUTION OF Z (STANDARD NORMAL DISTRIBUTION)
α = 0.05
rejection region
α/2 = 0.025
ACCEPT
rejection region
α/2 = 0.025
0

**Two - tailed?** or **One - tailed?**





1. **Significance Level (α):** You choose a significance level, denoted as α, which represents the maximum allowable probability of making a Type I error (i.e., rejecting a true null hypothesis). Commonly used values for α are 0.05 (5%) and 0.01 (1%), but it can vary depending on the study's requirements.

2. **Critical Values:** Critical values are values from a specific probability distribution (e.g., standard normal distribution for Z-tests or t-distribution for T-tests) that correspond to the chosen significance level (α) and the degrees of freedom (df) if applicable. Critical values define the boundary for rejecting or failing to reject the null hypothesis.

- For a one-tailed test, critical values are typically located in one tail of the distribution.
- For a two-tailed test, critical values are divided into both tails.

3. **Rejection Region:** The rejection region is the area under the probability distribution curve where the test statistic must fall in order to reject the null hypothesis. It's determined by the critical values and is typically the extreme portion of the distribution(s) where values are considered extreme enough to reject the null hypothesis.

- For One-Tailed Tests: If the test statistic falls in the tail of the distribution (either to the right or left, depending on the direction of the test), it is in the rejection region, and you reject the null hypothesis.
- For Two-Tailed Tests: If the test statistic falls outside the range defined by the critical values in both tails (i.e., it is significantly higher or lower), it is in the rejection region, and you reject the null hypothesis.

# Type I and Type II Errors

**Type I Error (False Positive):**

**Definition**: A Type I error occurs when you reject the null hypothesis when it is actually true. In other words, you conclude that there is a significant effect, difference, or relationship when there isn't one in reality.

**Symbol**: Denoted as α (alpha), it represents the significance level or the probability of making a Type I error.

**Example**: Concluding that a new drug is effective when, in fact, it has no effect on the condition being treated.

# Type I and Type II Errors

**Type II Error (False Negative):**

**Definition**: A Type II error occurs when you fail to reject the null hypothesis when it is actually false. In other words, you conclude that there is no significant effect, difference, or relationship when there is one in reality.

**Symbol**: Often denoted as β (beta), it represents the probability of making a Type II error.

**Example**: Concluding that a new drug is not effective when, in fact, it does have a significant effect on the condition being treated.d.

# Trade-off Between Type I and Type II Errors:

1. **Significance Level (α):**

- Increasing the significance level (e.g., from $\alpha = 0.05$ to $\alpha = 0.10$) decreases the risk of a Type II error but increases the risk of a Type I error.
- Conversely, decreasing the significance level (e.g., from $\alpha = 0.05$ to $\alpha = 0.01$) decreases the risk of a Type I error but increases the risk of a Type II error.

2. **Power of a Test (1 - β):**

- The power of a test represents the probability of correctly rejecting the null hypothesis when it is false (i.e., the ability to detect a true effect).
- Increasing the power of a test reduces the risk of a Type II error but may increase the risk of a Type I error.
- Power is influenced by factors like sample size, effect size, and the chosen significance level.

## Practical Considerations:

In practice, the choice of significance level (α) and power should be balanced based on the specific goals and consequences of the hypothesis test:

- Lowering α (e.g., using $\alpha = 0.01$) reduces the risk of making false positive claims but increases the risk of missing true effects (Type II errors).
- Increasing power (e.g., by using a larger sample size) reduces the risk of missing true effects (Type II errors) but may lead to a higher chance of false positives (Type I errors).