HYPOTHESIS TESTING

INFERENTIAL STATISTICS

prepared by:

Gyro A. Madrona

Electronics Engineer

@gyromadrona @gyro-madrona @OrbitGNX @OrbitGNX @gyronics

TOPIC OUTLINE

Hypothesis Test

Rejection Region

Critical Value and Z-score

p-Value

HYPOTHESIS TEST

HYPOTHESIS

A hypothesis is an initial **assumption** formed before collecting data, and it serves as a statement about a **population** parameter rather than about the sample data.

HYPOTHESIS TEST

A hypothesis test is simply comparing reality to an assumption and asking, "**Did things change?**"

Null Hypothesis (H_o)

Represents **no change**, no effect, or the status quo.

Alternative Hypothesis (H_a)

Represents the possibility that things did change or that there is a **significant difference**.

IS YOUR DATA GUILTY?

Hypothesis testing is like a legal system where the defendant is assumed **innocent** until proven guilty.

		The Truth	
		Innocent	Guilty
H_o: Innocent	Accept	✓	✗
	Reject	✗	✓

TYPES OF ERROR

1. **Type I Error** :

Rejecting the null hypothesis when it is actually true. The risk of making type I error is denoted by $\boldsymbol{\alpha}$ (e.g., 0.05).

2. **Type II Error** :

Failing to reject the null hypothesis when it is actually false. The risk of making a type II error is denoted by $\boldsymbol{\beta}$ (e.g. 0.20)

		The Truth	
		Innocent	Guilty
H_0: Innocent	Accept	✓	**Type II Error** (False negative)
	Reject	**Type I Error** (False positive)	✓

POWER OF THE TEST

Power $(\boldsymbol{1-\beta})$ is the probability that a test correctly **rejects** the null hypothesis $\boldsymbol{H_o}$. It is determined by alpha $\boldsymbol{\alpha}$ and sample size $\boldsymbol{n}$.

Increasing $\boldsymbol{\alpha}$ reduces $\boldsymbol{\beta}$ but raises false positives.

		The Truth	
		Innocent	Guilty
H_0: Innocent	Accept	✓	**Type II Error** (False negative)
	Reject	**Type I Error** (False positive)	✓

EXERCISE

A fitness tracker company claims their device measures heart rate with **95%** accuracy compared to medical-grade monitors. An independent lab wants to verify this claim.

Null Hypothesis

$$\boldsymbol{H_o: \mu_o = 95}$$

The average accuracy is 95%.

Alternative Hypothesis

$$\boldsymbol{H_a: \mu_o \neq 95}$$

The average accuracy differs from 95%.

EXERCISE

A manufacturer claims that their new energy-efficient LED bulbs have an average lifespan of **at least 25,000 hours**. A consumer group suspects that the actual lifespan is shorter and decides to test this claim.

<u>Null Hypothesis</u>

$$\boldsymbol{H_o: \mu_o = 25,000}$$

The average lifespan of the LED is 25,000 hours.

<u>Alternative Hypothesis</u>

$$\boldsymbol{H_a: \mu_o < 25,000}$$

The average lifespan of the LED is less than 25,000 hours.

EXERCISE

A study suggests that storing apples in a controlled atmosphere **extends** their shelf life beyond **30 days**. A food scientist wants to verify if this method truly increases shelf life compared to conventional storage.

Null Hypothesis

$$H_o: \mu_o = 30$$

Controlled-atmosphere storage shelf life is 30 days.

Alternative Hypothesis

$$H_a: \mu_o > 30$$

Controlled-atmosphere storage increases shelf life beyond 30 days.

REJECTION REGION

SIGNIFICANCE LEVEL

The **significance level** ($\boldsymbol{\alpha}$) determines the threshold for deciding whether to **reject** the null hypothesis ($\boldsymbol{H_o}$).

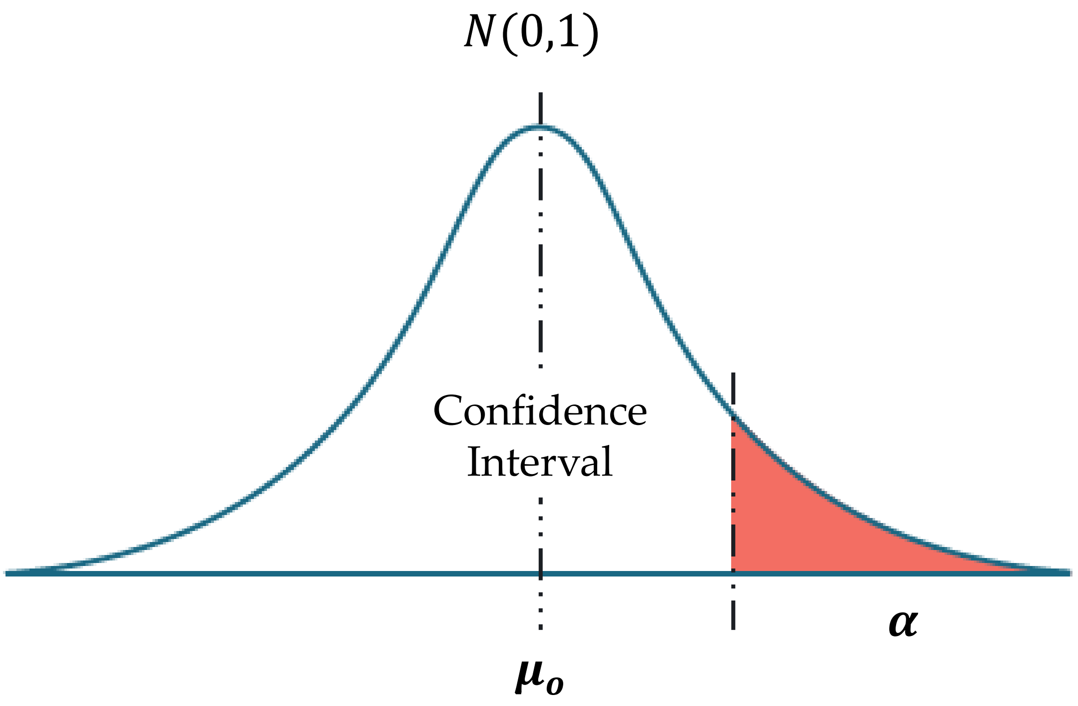
Typical values for $\boldsymbol{\alpha}$:

0.01

0.05

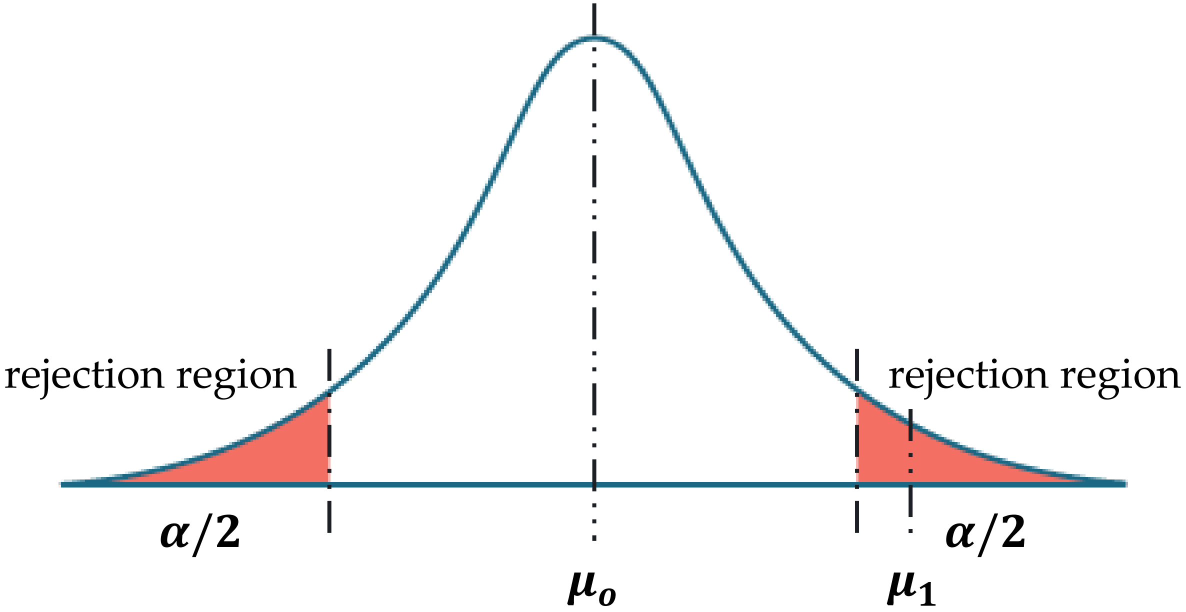
0.1

Standard Normal Distribution



REJECTION REGION

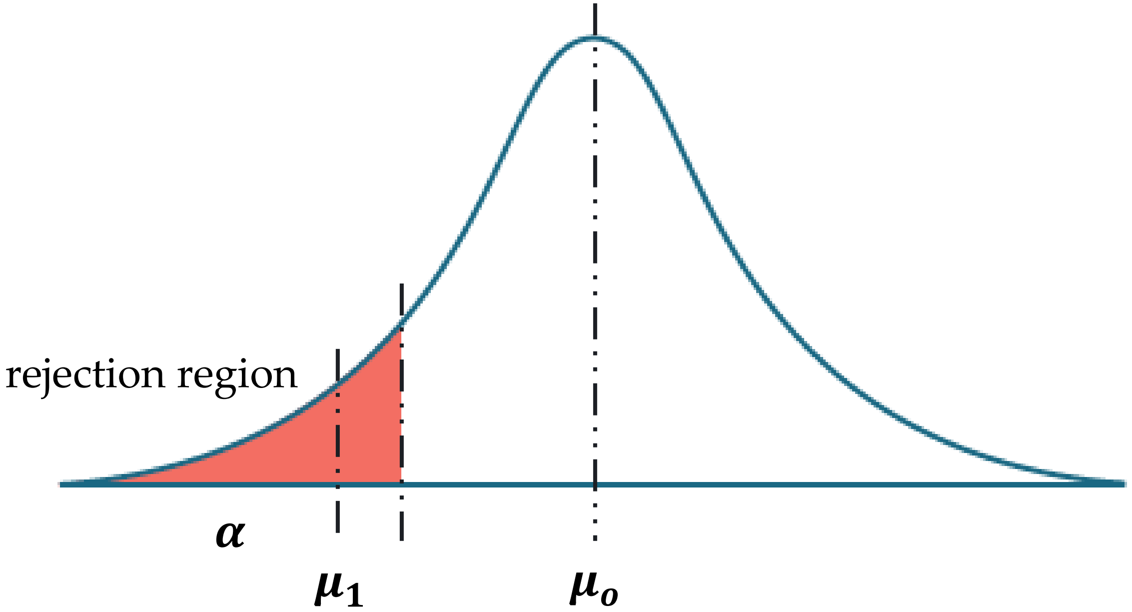
Two-Tailed Test



$$H_o\colon \mu_1 = \mu_o$$

$$H_a\colon \mu_1 \neq \mu_o$$

One-Tailed Test

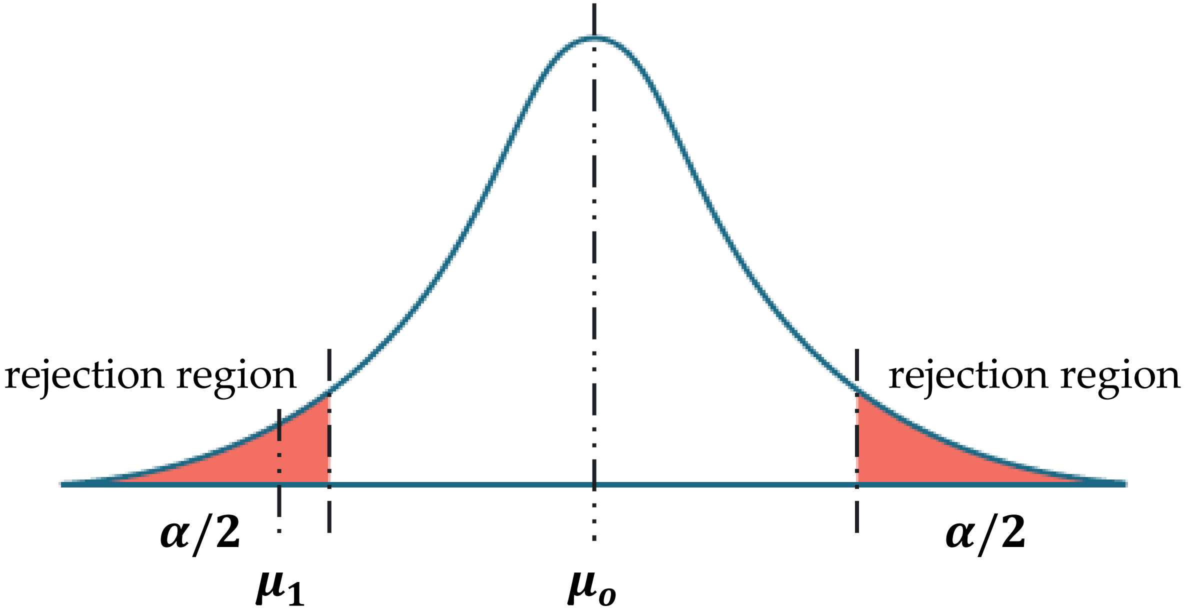


$$H_o\colon \mu_1 = \mu_o$$

$$H_a\colon \mu_1 < \mu_o$$

REJECTION REGION

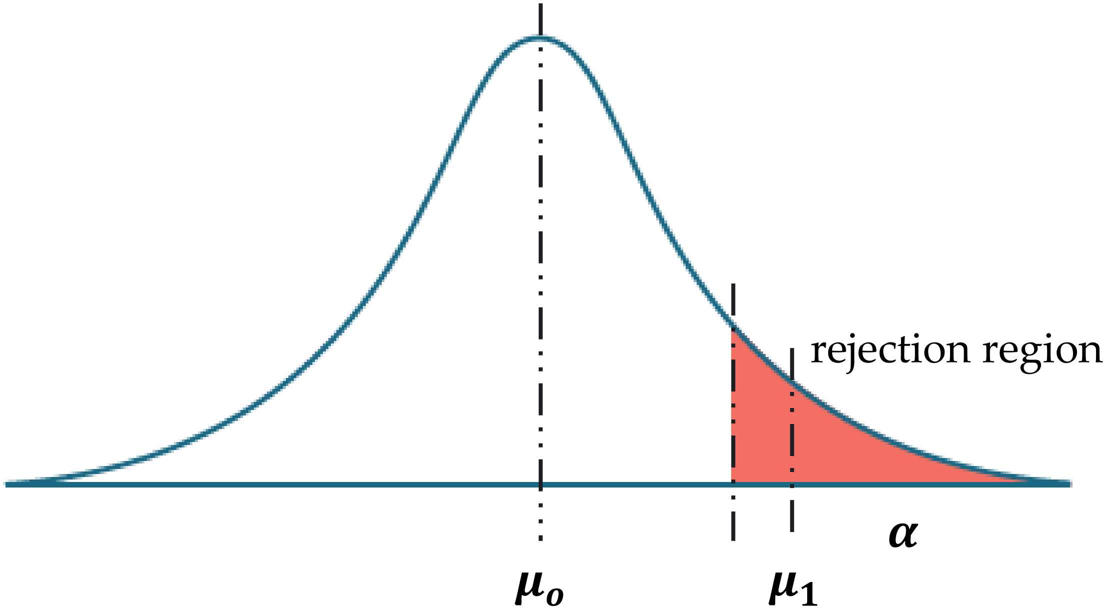
Two-Tailed Test



H_o: $\mu_1 = \mu_o$

H_a: $\mu_1 \neq \mu_o$

One-Tailed Test



H_o: $\mu_1 = \mu_o$

H_a: $\mu_1 > \mu_o$

CRITICAL VALUE AND Z-SCORE

CRITICAL VALUE AND Z-SCORE

lowercase z

z refers to the **critical value** obtained from the standard normal distribution table (z-table).

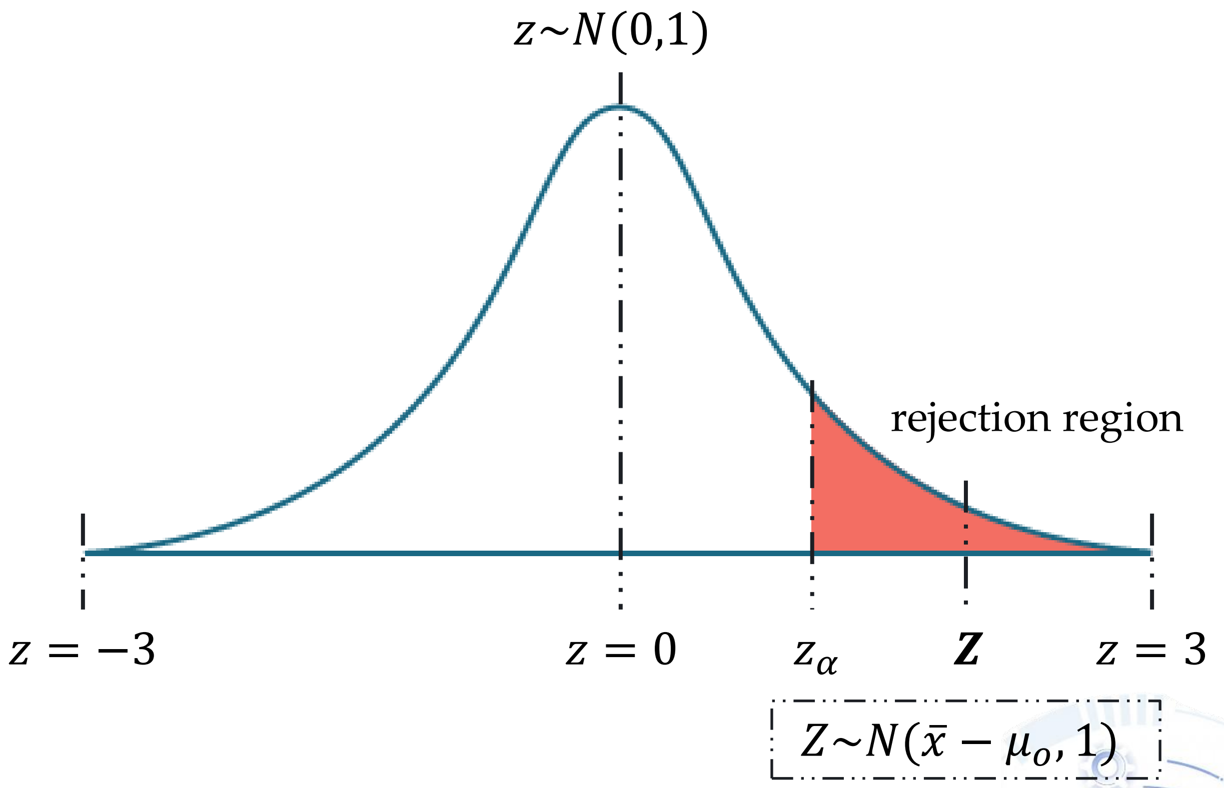
uppercase Z

Z is a standardized variable associated with the test called the **Z-score**.

Formula:

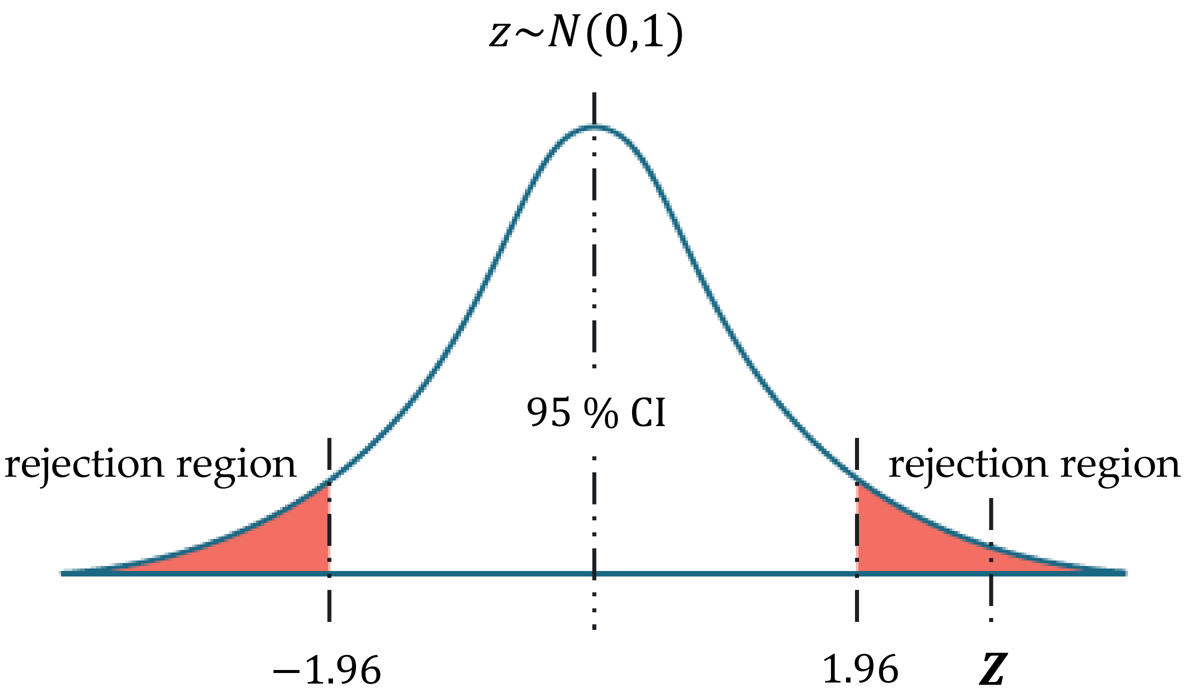
$$Z = \frac{\bar{x} - \mu_o}{\sigma/\sqrt{n}}$$

One-Tailed Test



EXERCISE

Two-Tailed Test



$\alpha = 0.05$

$z_{0.025} = 1.96$

Null Hypothesis

$\boldsymbol{H_o: \mu_o = 95}$

The average accuracy is 95%.

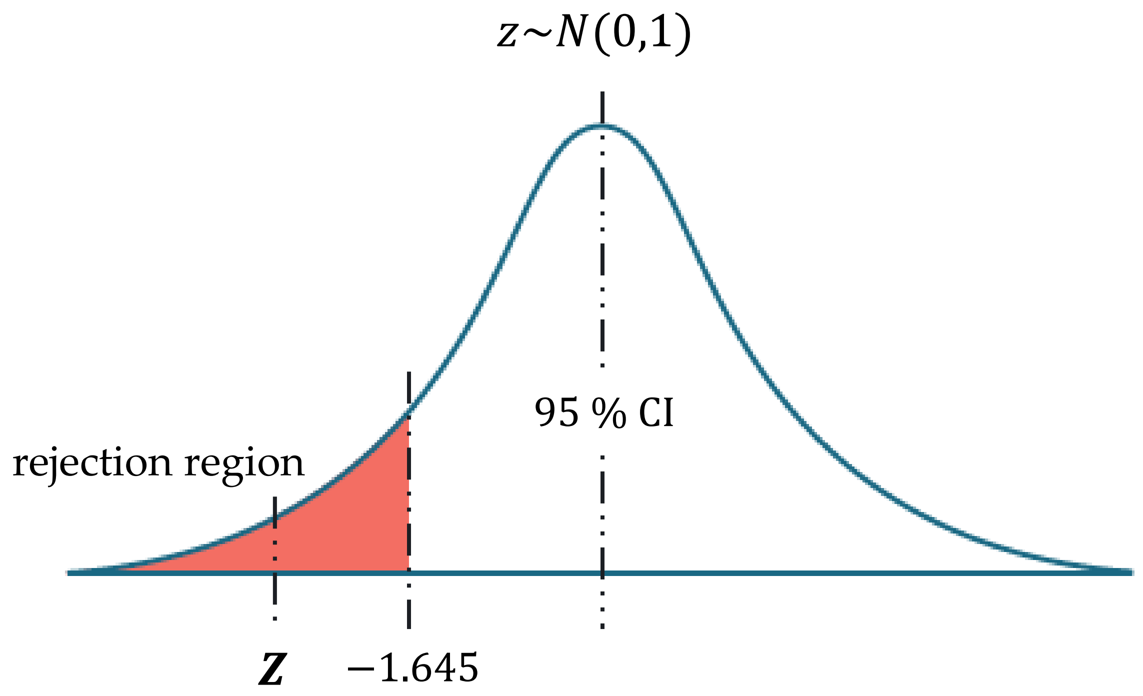
Alternative Hypothesis

$\boldsymbol{H_a: \mu_o \neq 95}$

The average accuracy differs from 95%.

EXERCISE

One-Tailed Test



$\alpha = 0.05$

$z_{0.05} = 1.645$

Null Hypothesis

$\boldsymbol{H_o: \mu_o = 25,000}$

The average lifespan of the LED is 25,000 hours.

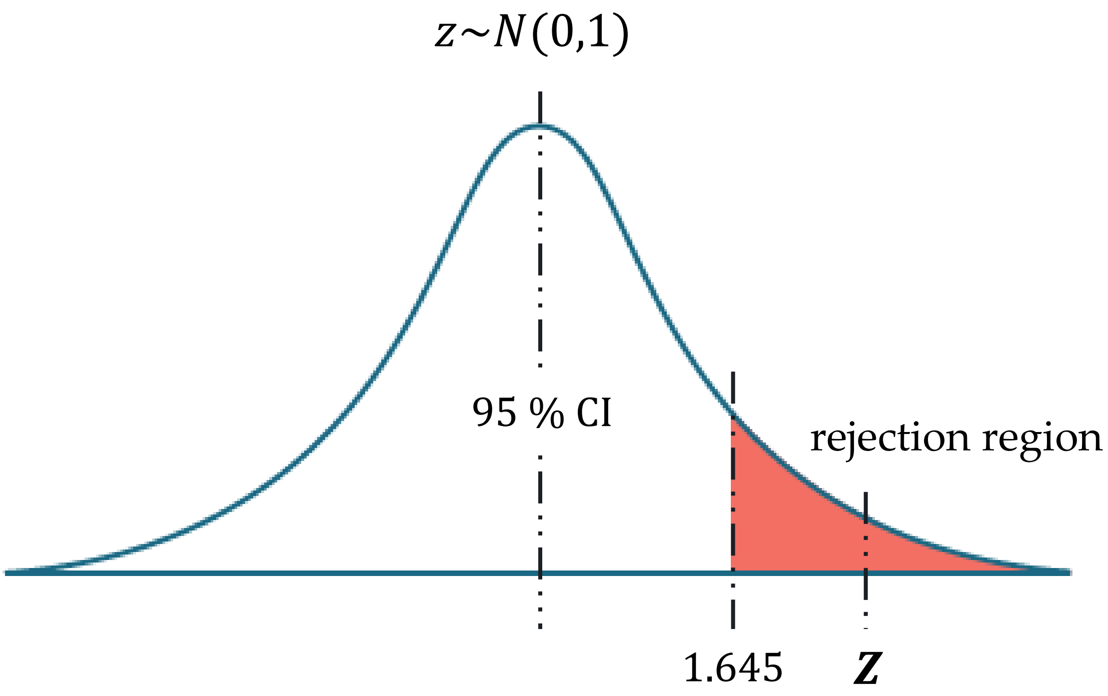
Alternative Hypothesis

$\boldsymbol{H_a: \mu_o < 25,000}$

The average lifespan of the LED is less than 25,000 hours.

EXERCISE

One-Tailed Test



$\alpha = 0.05$

$z_{0.05} = 1.645$

Null Hypothesis

$\boldsymbol{H_o}$: $\boldsymbol{\mu_o = 30}$

Controlled-atmosphere storage shelf life is 30 days.

Alternative Hypothesis

$\boldsymbol{H_a}$: $\boldsymbol{\mu_o > 30}$

Controlled-atmosphere storage increases shelf life beyond 30 days.

EXERCISE

A manufacturing process is claimed to have an average defect rate of **10.32** units, with a known standard deviation of **3.17** units. The Statistical Process Control (SPC) department suspects this claim may no longer be valid and collects a **random sample** of **30** production units to test whether the true average **defect rate differs** significantly from **10.32**.

Dataset:

defects-data-30-samples.csv

Solution

P-VALUE

P-VALUE

The **p-value** (probability value) is the **smallest level of significance** at which we can still reject the null hypothesis, given the observed sample statistic.

Formula:

One-sided Test

$$\textbf{p-value} = \mathbf{1} - \text{value from the table}$$

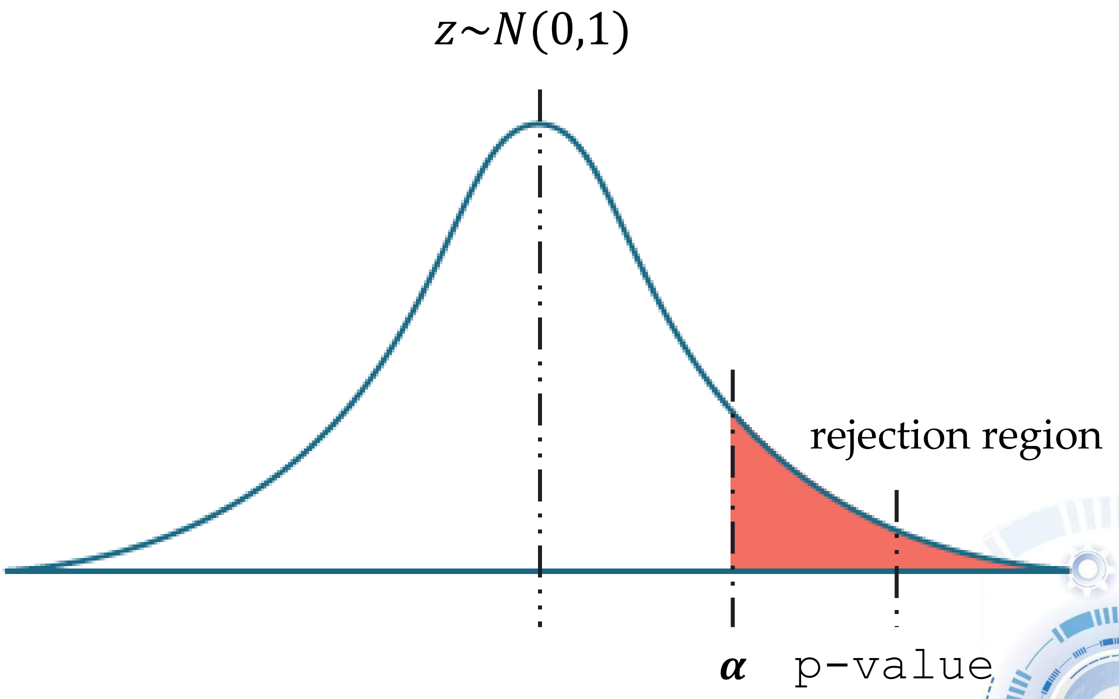
Two-sided Test

$$\textbf{p-value} = (\mathbf{1} - \text{value from the table}) \times \mathbf{2}$$

Hypothesis Test

Reject H_o if **p-value** $\leq \alpha$

Fail to reject H_o if p-value $> \alpha$



EXERCISE

A manufacturing process is claimed to have an average defect rate of **10.32** units, with a known standard deviation of **3.17** units. The Statistical Process Control (SPC) department suspects this claim may no longer be valid and collects a **random sample** of **30** production units to test whether the true average **defect rate differs** significantly from **10.32**.

Dataset:

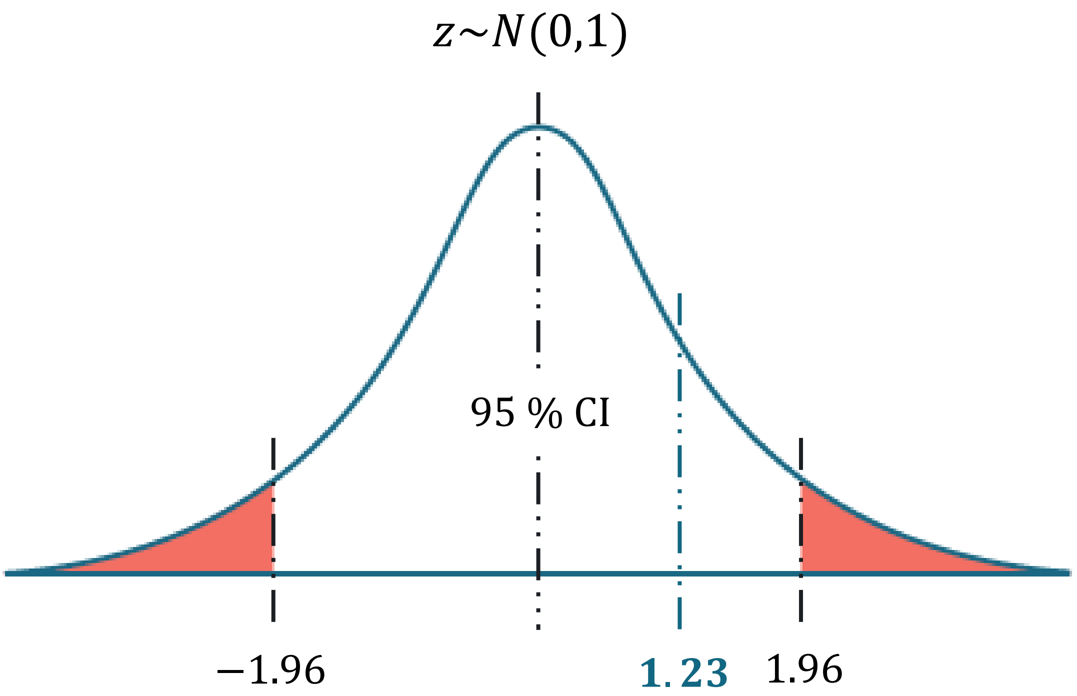
defects-data-30-samples.csv

Solution

EXERCISE

Two-Tailed Test

Solution



$\alpha = 0.05$

$z_{0.025} = 1.96$

STEPS IN HYPOTHESIS TESTING

State $\boldsymbol{H_o}$ and $\boldsymbol{H_a}$

Set confidence level (pick $\boldsymbol{\alpha}$)

Select the hypothesis test

Calculate sample size or the power $(\mathbf{1} - \boldsymbol{\beta})$

Collect data and run the test

Translate to practical solution and verify

LABORATORY