American University of Beirut

STAT 210: Elementary Statistics for Sciences

2022–2023 Fall

Siamak Taati

Chapter 9

Testing hypotheses about mean and proportion

Part 2

Statistical hypothesis testing

Testing hypotheses

Very often, we are in a situation that we would like to judge between competing claims in absence of complete information.

Courtroom

Is the suspect guilty?

Medical research

Is a certain treatment effective?

Quality control

Does a certain product have the claimed specification?

Policy making

Is a certain policy effective?

Testing hypotheses

Very often, we are in a situation that we would like to judge between competing claims in absence of complete information.

Objective

Make a rational judgment based on the available evidence.

In statistical hypothesis testing, the judgment is made based on statistical evidence.

Statistical hypothesis testing

Competing hypotheses

H_0 The null hypothesis (the default model)

H_1 The alternative hypothesis (the alternative claim)

Statistical hypothesis testing

Competing hypotheses

- H_0 The null hypothesis (the default model)
- H_1 The alternative hypothesis (the alternative claim)

Examples

null hypothesis	alternative hypothesis

Statistical hypothesis testing

Competing hypotheses

H_0 The null hypothesis (the default model)

H_1 The alternative hypothesis (the alternative claim)

Examples

null hypothesis	alternative hypothesis
The suspect is innocent	The suspect is guilty

Statistical hypothesis testing

Competing hypotheses

H_0 The null hypothesis (the default model)

H_1 The alternative hypothesis (the alternative claim)

Examples

null hypothesis	alternative hypothesis
The suspect is innocent The coin is fair	The suspect is guilty The coin is biased

Statistical hypothesis testing

Competing hypotheses

H_0 The null hypothesis (the default model)

H_1 The alternative hypothesis (the alternative claim)

Examples

null hypothesis	alternative hypothesis
The suspect is innocent	The suspect is guilty
The coin is fair	The coin is biased
The treatment is not effective	The treatment is effective

Statistical hypothesis testing

Competing hypotheses

H_0 The null hypothesis (the default model)

H_1 The alternative hypothesis (the alternative claim)

Objective

Decide whether we should abandon H_0 in favor of H_1.

Examples

null hypothesis	alternative hypothesis
The suspect is innocent	The suspect is guilty
The coin is fair	The coin is biased
The treatment is not effective	The treatment is effective

Statistical hypothesis testing

Competing hypotheses

H_0 The null hypothesis (the default model)

H_1 The alternative hypothesis (the alternative claim)

Objective

Decide whether we should abandon H_0 in favor of H_1.

Evidence

A random sample

Statistical hypothesis testing

Competing hypotheses

H_0 The null hypothesis (the default model)

H_1 The alternative hypothesis (the alternative claim)

Objective

Decide whether we should abandon H_0 in favor of H_1.

Evidence

A random sample

Strategy

We use a statistic T and compare [e.g., sample mean]

Statistical hypothesis testing

Competing hypotheses

H_0 The null hypothesis (the default model)

H_1 The alternative hypothesis (the alternative claim)

Objective

Decide whether we should abandon H_0 in favor of H_1.

Evidence

A random sample

Strategy

We use a statistic T and compare [e.g., sample mean]

1. The observed value of T

Statistical hypothesis testing

Competing hypotheses

- H_0 The null hypothesis (the default model)
- H_1 The alternative hypothesis (the alternative claim)

Objective

Decide whether we should abandon H_0 in favor of H_1.

Evidence

A random sample

Strategy

We use a statistic T and compare [e.g., sample mean]

1. The observed value of T
2. The distribution of T as suggested by H_0,

Statistical hypothesis testing

Competing hypotheses

H_0 The null hypothesis (the default model)

H_1 The alternative hypothesis (the alternative claim)

Objective

Decide whether we should abandon H_0 in favor of H_1.

Evidence

A random sample

Strategy

We use a statistic T and compare [e.g., sample mean]

1. The observed value of T
2. The distribution of T as suggested by H_0,

We ask ourselves:

Statistical hypothesis testing

Competing hypotheses

H_0 The null hypothesis (the default model)

H_1 The alternative hypothesis (the alternative claim)

Objective

Decide whether we should abandon H_0 in favor of H_1.

Evidence

A random sample

Strategy

We use a statistic T and compare [e.g., sample mean]

1. The observed value of T
2. The distribution of T as suggested by H_0,

We ask ourselves:

Q Is the observed value of T too extreme for H_0 to be plausible?

Statistical hypothesis testing

Competing hypotheses

H_0 The null hypothesis (the default model)

H_1 The alternative hypothesis (the alternative claim)

Objective

Decide whether we should abandon H_0 in favor of H_1.

Possible outcomes

Statistical hypothesis testing

Competing hypotheses

H_0 The null hypothesis (the default model)

H_1 The alternative hypothesis (the alternative claim)

Objective

Decide whether we should abandon H_0 in favor of H_1.

Possible outcomes

- Reject H_0

  (Meaning: The evidence is statistically significant in favor of H_1.)

Statistical hypothesis testing

Competing hypotheses

H_0 The null hypothesis (the default model)

H_1 The alternative hypothesis (the alternative claim)

Objective

Decide whether we should abandon H_0 in favor of H_1.

Possible outcomes

- Reject H_0
  (Meaning: The evidence is statistically significant in favor of H_1.)
- Fail to reject H_0
  (Meaning: The evidence is inconclusive.)

Statistical hypothesis testing

Competing hypotheses

H_0 The null hypothesis (the default model)

H_1 The alternative hypothesis (the alternative claim)

Objective

Decide whether we should abandon H_0 in favor of H_1.

Possible outcomes

- Reject H_0
  (Meaning: The evidence is statistically significant in favor of H_1.)
- Fail to reject H_0
  (Meaning: The evidence is inconclusive.)

Remark

The statistical evidence is always meant to be against H_0.
Even if the evidence is weak, it can never lead us to accept H_0.

Statistical hypothesis testing

Competing hypotheses

- H_0 The null hypothesis (the default model)
- H_1 The alternative hypothesis (the alternative claim)

Remark

Since the evidence is limited, our judgment can be erroneous!

Statistical hypothesis testing

Competing hypotheses

H_0 The null hypothesis (the default model)

H_1 The alternative hypothesis (the alternative claim)

Remark

Since the evidence is limited, our judgment can be erroneous!

Two types of errors

Statistical hypothesis testing

Competing hypotheses

H_0 The null hypothesis (the default model)

H_1 The alternative hypothesis (the alternative claim)

Remark

Since the evidence is limited, our judgment can be erroneous!

Two types of errors

- Type I (false rejection):

  H_0 is true but we incorrectly reject it.

Statistical hypothesis testing

Competing hypotheses

H_0 The null hypothesis (the default model)

H_1 The alternative hypothesis (the alternative claim)

Remark

Since the evidence is limited, our judgment can be erroneous!

Two types of errors

- Type I (false rejection):

  H_0 is true but we incorrectly reject it.

- Type II (failure to reject):

  H_0 is not true but we fail to reject it.

Statistical hypothesis testing

Competing hypotheses

- H_0 The null hypothesis (the default model)
- H_1 The alternative hypothesis (the alternative claim)

Remark

The hypotheses can often be phrased in terms of the parameters of the population (or populations) they are about.

Statistical hypothesis testing

Competing hypotheses

H_0 The null hypothesis (the default model)

H_1 The alternative hypothesis (the alternative claim)

Remark

The hypotheses can often be phrased in terms of the parameters of the population (or populations) they are about.

For instance, the hypothesis

H_1 The coin is unfair.

Statistical hypothesis testing

Competing hypotheses

H_0 The null hypothesis (the default model)

H_1 The alternative hypothesis (the alternative claim)

Remark

The hypotheses can often be phrased in terms of the parameters of the population (or populations) they are about.

For instance, the hypothesis

H_1 The coin is unfair.

can be rephrased as

H_1 $p \neq 1/2$.

where p is the probably of getting heads in one flip of the coin.

Statistical hypothesis testing

Competing hypotheses

H_0 The null hypothesis (the default model)

H_1 The alternative hypothesis (the alternative claim)

Remark

The hypotheses can often be phrased in terms of the parameters of the population (or populations) they are about.

For instance, the hypothesis

H_1 On average, gazelles run faster than lions.

Statistical hypothesis testing

Competing hypotheses

H$_0$ The null hypothesis (the default model)

H$_1$ The alternative hypothesis (the alternative claim)

Remark

The hypotheses can often be phrased in terms of the parameters of the population (or populations) they are about.

For instance, the hypothesis

H$_1$ On average, gazelles run faster than lions.

can be rephrased as

H$_1$ $\mu_1 > \mu_2$.

where μ_1 is the population mean running speed of gazelles and μ_2 is the population mean running speed of lions.

Tests about a population parameter

Suppose θ is a parameter of the population. [e.g., μ or σ]

Tests about a population parameter

Suppose θ is a parameter of the population. [e.g., μ or σ]

Depending on the form of the alternative hypothesis, we can have three different scenarios:

Tests about a population parameter

Suppose θ is a parameter of the population. [e.g., μ or σ]

Depending on the form of the alternative hypothesis, we can have three different scenarios:

Competing hypotheses

H_0	$\theta = \theta_0$
H_1	$\theta \neq \theta_0$

Tests about a population parameter

Suppose θ is a parameter of the population. [e.g., μ or σ]

Depending on the form of the alternative hypothesis, we can have three different scenarios:

Competing hypotheses

H_0 $\theta = \theta_0$

H_1 $\theta \neq \theta_0$

Competing hypotheses

H_0 $\theta = \theta_0$

H_1 $\theta > \theta_0$

Tests about a population parameter

Suppose θ is a parameter of the population. [e.g., μ or σ]

Depending on the form of the alternative hypothesis, we can have three different scenarios:

Competing hypotheses	Competing hypotheses	Competing hypotheses
H_0 $\theta = \theta_0$	H_0 $\theta = \theta_0$	H_0 $\theta = \theta_0$
H_1 $\theta < \theta_0$	H_1 $\theta \neq \theta_0$	H_1 $\theta > \theta_0$

Tests about a population parameter

Suppose θ is a parameter of the population. [e.g., μ or σ]

Depending on the form of the alternative hypothesis, we can have three different scenarios:

Competing hypotheses	Competing hypotheses	Competing hypotheses
H_0 $\theta = \theta_0$	H_0 $\theta = \theta_0$	H_0 $\theta = \theta_0$
H_1 $\theta < \theta_0$	H_1 $\theta \neq \theta_0$	H_1 $\theta > \theta_0$
	two-tailed test	

Tests about a population parameter

Suppose θ is a parameter of the population. [e.g., μ or σ]

Depending on the form of the alternative hypothesis, we can have three different scenarios:

Competing hypotheses	Competing hypotheses	Competing hypotheses
H_0 $\theta = \theta_0$	H_0 $\theta = \theta_0$	H_0 $\theta = \theta_0$
H_1 $\theta < \theta_0$	H_1 $\theta \neq \theta_0$	H_1 $\theta > \theta_0$
	two-tailed test	right-tailed test

Tests about a population parameter

Suppose θ is a parameter of the population. [e.g., μ or σ]

Depending on the form of the alternative hypothesis, we can have three different scenarios:

Competing hypotheses	Competing hypotheses	Competing hypotheses
H_0 $\theta = \theta_0$	H_0 $\theta = \theta_0$	H_0 $\theta = \theta_0$
H_1 $\theta < \theta_0$	H_1 $\theta \neq \theta_0$	H_1 $\theta > \theta_0$
left-tailed test	two-tailed test	right-tailed test

Tests about a population parameter

Suppose θ is a parameter of the population. [e.g., μ or σ]

Depending on the form of the alternative hypothesis, we can have three different scenarios:

Competing hypotheses	Competing hypotheses	Competing hypotheses
H_0 $\theta \geq \theta_0$	H_0 $\theta = \theta_0$	H_0 $\theta \leq \theta_0$
H_1 $\theta < \theta_0$	H_1 $\theta \neq \theta_0$	H_1 $\theta > \theta_0$
left-tailed test	two-tailed test	right-tailed test

Tests about a population parameter

Suppose θ is a parameter of the population. [e.g., μ or σ]

Depending on the form of the alternative hypothesis, we can have three different scenarios:

Competing hypotheses	Competing hypotheses	Competing hypotheses
H_0 $\theta \geq \theta_0$	H_0 $\theta = \theta_0$	H_0 $\theta \leq \theta_0$
H_1 $\theta < \theta_0$	H_1 $\theta \neq \theta_0$	H_1 $\theta > \theta_0$
left-tailed test	two-tailed test	right-tailed test

Example (A ...... test)

H_0 $p = 1/2$ (the coin is fair)

H_1 $p \neq 1/2$ (the coin is unfair)

Tests about a population parameter

Suppose θ is a parameter of the population. [e.g., μ or σ]

Depending on the form of the alternative hypothesis, we can have three different scenarios:

Competing hypotheses	Competing hypotheses	Competing hypotheses
H_0 $\theta \geq \theta_0$	H_0 $\theta = \theta_0$	H_0 $\theta \leq \theta_0$
H_1 $\theta < \theta_0$	H_1 $\theta \neq \theta_0$	H_1 $\theta > \theta_0$
left-tailed test	two-tailed test	right-tailed test

Example (A two-tailed test)

H_0 $p = 1/2$ (the coin is fair)

H_1 $p \neq 1/2$ (the coin is unfair)

Tests about a population parameter

Suppose θ is a parameter of the population. [e.g., μ or σ]

Depending on the form of the alternative hypothesis, we can have three different scenarios:

Competing hypotheses	Competing hypotheses	Competing hypotheses
H_0 $\theta \geq \theta_0$	H_0 $\theta = \theta_0$	H_0 $\theta \leq \theta_0$
H_1 $\theta < \theta_0$	H_1 $\theta \neq \theta_0$	H_1 $\theta > \theta_0$
left-tailed test	two-tailed test	right-tailed test

Example (A ...... test)

H_0 $\mu = 10\,\mathrm{g}$ (the average sugar content of a drink is as claimed)

H_1 $\mu > 10\,\mathrm{g}$ (the average sugar content of a drink is more than claimed)

Tests about a population parameter

Suppose θ is a parameter of the population. [e.g., μ or σ]

Depending on the form of the alternative hypothesis, we can have three different scenarios:

Competing hypotheses	Competing hypotheses	Competing hypotheses
H_0 $\theta \geq \theta_0$	H_0 $\theta = \theta_0$	H_0 $\theta \leq \theta_0$
H_1 $\theta < \theta_0$	H_1 $\theta \neq \theta_0$	H_1 $\theta > \theta_0$
left-tailed test	two-tailed test	right-tailed test

Example (A right-tailed test)

H_0 $\mu = 10\,\mathrm{g}$ (the average sugar content of a drink is as claimed)

H_1 $\mu > 10\,\mathrm{g}$ (the average sugar content of a drink is more than claimed)

Tests about a population parameter

Competing hypotheses	Competing hypotheses	Competing hypotheses
H_0 $\theta = \theta_0$	H_0 $\theta = \theta_0$	H_0 $\theta = \theta_0$
H_1 $\theta < \theta_0$	H_1 $\theta \neq \theta_0$	H_1 $\theta > \theta_0$
left-tailed test	two-tailed test	right-tailed test

Test statistic

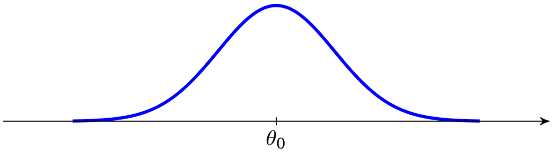
A point estimator $\widehat{\theta}$ for θ. [e.g., sample mean $\widehat{\mu} = \overline{X}$]

Tests about a population parameter

Competing hypotheses	Competing hypotheses	Competing hypotheses
H$_0$ $\theta = \theta_0$	H$_0$ $\theta = \theta_0$	H$_0$ $\theta = \theta_0$
H$_1$ $\theta < \theta_0$	H$_1$ $\theta \neq \theta_0$	H$_1$ $\theta > \theta_0$
left-tailed test	two-tailed test	right-tailed test

Test statistic

A point estimator $\widehat{\theta}$ for θ. [e.g., sample mean $\widehat{\mu} = \overline{X}$]



Distribution of $\widehat{\theta}$ assuming H$_0$

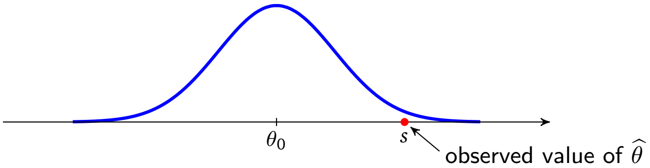
Tests about a population parameter

Competing hypotheses	Competing hypotheses	Competing hypotheses
H_0 $\theta = \theta_0$	H_0 $\theta = \theta_0$	H_0 $\theta = \theta_0$
H_1 $\theta < \theta_0$	H_1 $\theta \neq \theta_0$	H_1 $\theta > \theta_0$
left-tailed test	two-tailed test	right-tailed test

Test statistic

A point estimator $\widehat{\theta}$ for θ.

[e.g., sample mean $\widehat{\mu} = \overline{X}$]



Distribution of $\widehat{\theta}$ assuming H_0

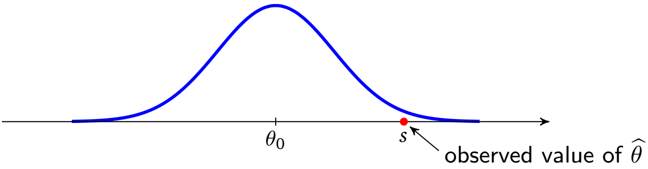
Tests about a population parameter

Competing hypotheses	Competing hypotheses	Competing hypotheses
H$_0$ $\theta = \theta_0$	H$_0$ $\theta = \theta_0$	H$_0$ $\theta = \theta_0$
H$_1$ $\theta < \theta_0$	H$_1$ $\theta \neq \theta_0$	H$_1$ $\theta > \theta_0$
left-tailed test	two-tailed test	right-tailed test

Test statistic

A point estimator $\widehat{\theta}$ for θ.

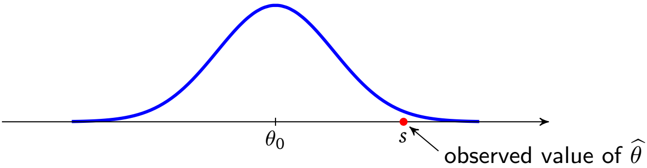
[e.g., sample mean $\widehat{\mu} = \overline{X}$]



Distribution of $\widehat{\theta}$ assuming H$_0$

Q Is the observed value $\widehat{\theta} = s$ too extreme for H$_0$ to be plausible?

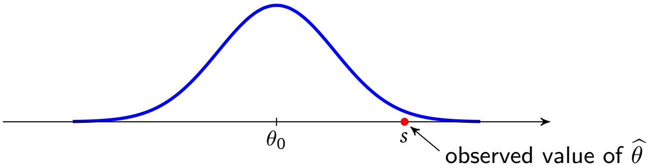
Tests about a population parameter



Distribution of $\widehat{\theta}$ assuming H$_0$

Q Is the observed value $\widehat{\theta} = s$ too extreme for H$_0$ to be plausible?

Tests about a population parameter

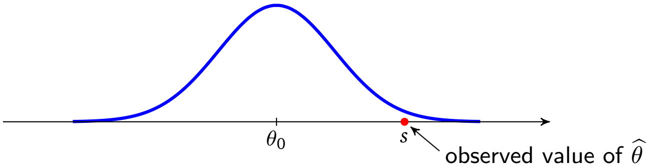


Distribution of $\widehat{\theta}$ assuming H_0

Q Is the observed value $\widehat{\theta} = s$ too extreme for H_0 to be plausible?

We have to decide on a criterion before observing the evidence.

Tests about a population parameter



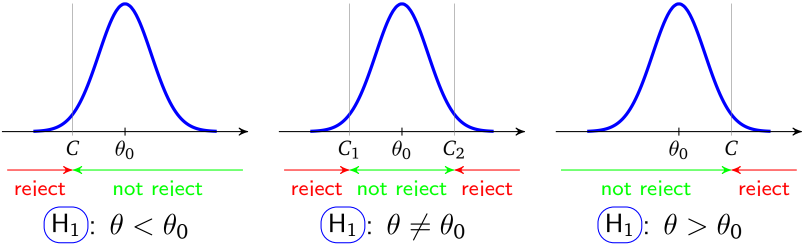
Distribution of $\widehat{\theta}$ assuming H$_0$

Q Is the observed value $\widehat{\theta} = s$ too extreme for H$_0$ to be plausible?

We have to decide on a criterion before observing the evidence.

The main idea is to set a threshold.

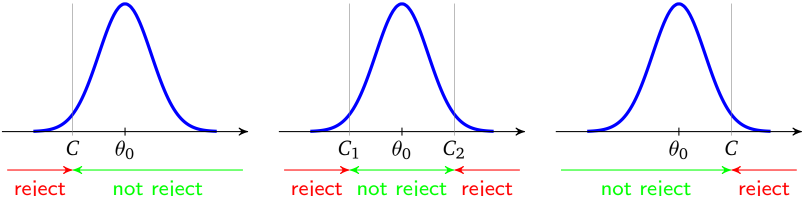
Tests about a population parameter



Q How should we choose the critical values?

Looking at the probability of Type I error could help.

Tests about a population parameter



H$_1$: $\theta < \theta_0$ H$_1$: $\theta \neq \theta_0$ H$_1$: $\theta > \theta_0$

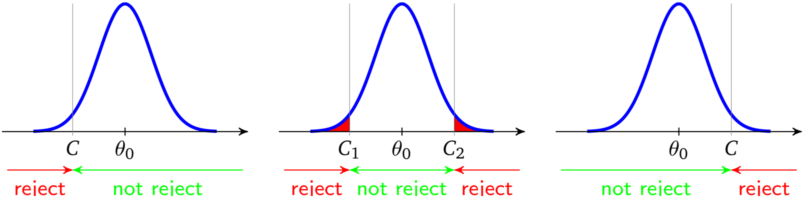
Significance level

The probability of Type I error

$$\alpha := \mathbb{P}(\text{H}_0 \text{ is rejected} \mid \text{H}_0 \text{ is true})$$

is called the significance level.

Tests about a population parameter



H$_1$: $\theta < \theta_0$ H$_1$: $\theta \neq \theta_0$ H$_1$: $\theta > \theta_0$

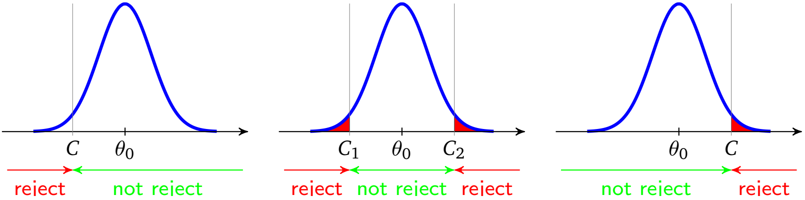
Significance level

The probability of Type I error

$$\alpha := \mathbb{P}(\text{H}_0 \text{ is rejected} \mid \text{H}_0 \text{ is true})$$

is called the significance level.

Tests about a population parameter



H$_1$: $\theta < \theta_0$ H$_1$: $\theta \neq \theta_0$ H$_1$: $\theta > \theta_0$

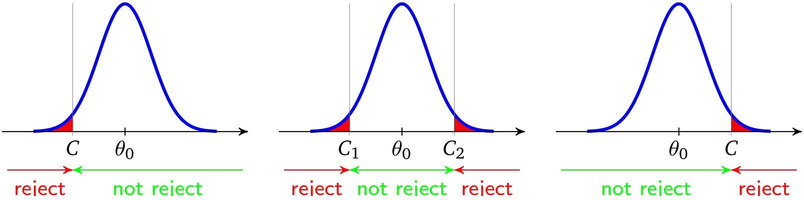
Significance level

The probability of Type I error

$$\alpha := \mathbb{P}(\text{H}_0 \text{ is rejected} \mid \text{H}_0 \text{ is true})$$

is called the significance level.

Tests about a population parameter



H_1: $\theta < \theta_0$ H_1: $\theta \neq \theta_0$ H_1: $\theta > \theta_0$

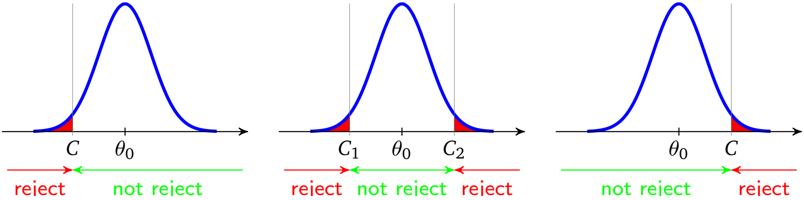
Significance level

The probability of Type I error

$$\alpha := \mathbb{P}(\mathrm{H}_0 \text{ is rejected} \mid \mathrm{H}_0 \text{ is true})$$

is called the significance level.

Tests about a population parameter



H$_1$: $\theta < \theta_0$ H$_1$: $\theta \neq \theta_0$ H$_1$: $\theta > \theta_0$

Significance level

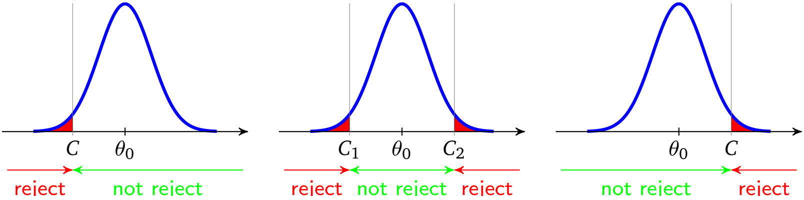
The probability of Type I error

$$\alpha := \mathbb{P}(\text{H}_0 \text{ is rejected} \mid \text{H}_0 \text{ is true})$$

is called the significance level.

Given the significance level, we can find the critical values for each type of test.

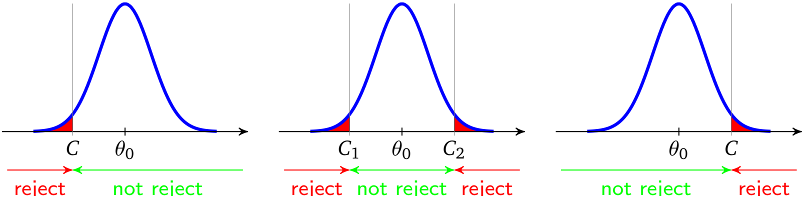
Tests about a population parameter



H$_1$: $\theta < \theta_0$ H$_1$: $\theta \neq \theta_0$ H$_1$: $\theta > \theta_0$

Q How should we choose the critical values?

Tests about a population parameter

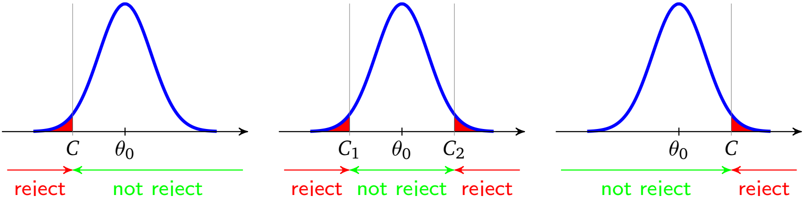


H$_1$: $\theta < \theta_0$ H$_1$: $\theta \neq \theta_0$ H$_1$: $\theta > \theta_0$

Q How should we choose the critical values?

A Based on the significance level.

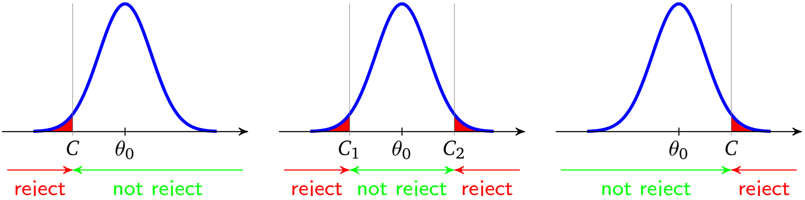
Tests about a population parameter



H$_1$: $\theta < \theta_0$ H$_1$: $\theta \neq \theta_0$ H$_1$: $\theta > \theta_0$

Q How should we choose the significance level?

Tests about a population parameter

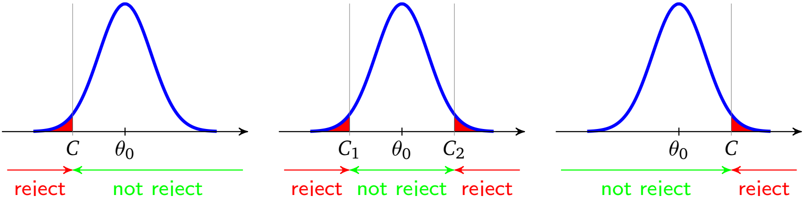


H$_1$: $\theta < \theta_0$ H$_1$: $\theta \neq \theta_0$ H$_1$: $\theta > \theta_0$

Q How should we choose the significance level?

A Not too large, not too small.

Tests about a population parameter



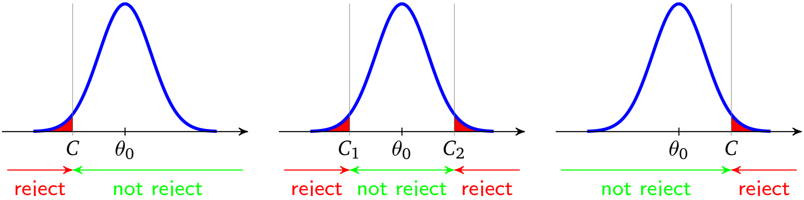
H$_1$: $\theta < \theta_0$ H$_1$: $\theta \neq \theta_0$ H$_1$: $\theta > \theta_0$

Q How should we choose the significance level?

A Not too large, not too small.

Trade-off

Tests about a population parameter



H$_1$: $\theta < \theta_0$ H$_1$: $\theta \neq \theta_0$ H$_1$: $\theta > \theta_0$

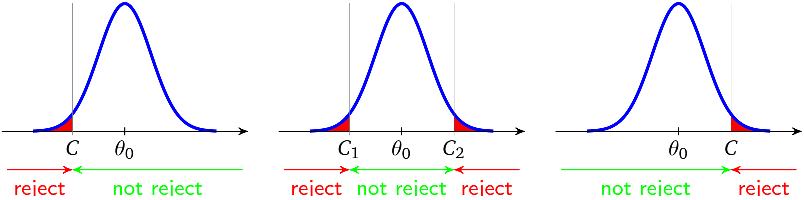
Q How should we choose the significance level?

$\boxed{\text{A}}$ Not too large, not too small.

Trade-off

- If α is too large, then the chance of false rejection (Type I error) is high.

Tests about a population parameter



H_1: $\theta < \theta_0$ H_1: $\theta \neq \theta_0$ H_1: $\theta > \theta_0$

Q How should we choose the significance level?

A Not too large, not too small.

Trade-off

- If α is too large, then the chance of false rejection (Type I error) is high.
- If α is too small, then there is a higher chance of missing subtle differences (Type II error).