IEDA 2540 Statistics for Engineers
Hypothesis Testing

Wei YOU

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Spring, 2025

Introduction

- Hypothesis testing is a critical tool in evidence-based decision-making, a process that begins with posing *a question that demands a clear, data-driven answer*.
- This approach is fundamental in fields where decisions have significant implications, such as healthcare, education, and policy-making.
- Let's elaborate on this concept with two examples.

Example: Hypothesis

Question: "Is this new medication more effective than the current standard?"

- We need to decide between two possibilities:
  - **Possibility 1:** The new medication is *no more* effective than the current standard.
  - **Possibility 2:** The new medication is *more* effective than the current standard.
- As a pharmaceutical company, we wish our newly developed medication to be effective. Hence:
  - **Possibility 1** represents a *conservative stance*, assuming no difference until proven otherwise.
  - **Possibility 2** is the assertion we hope to *support with evidence* to promote our new medication.

Example: Hypothesis

Question: "Does a new webpage design increase the time visitors spend on the webpage?"

- We need to decide between two possibilities:
  - **Possibility 1:** The new design *does not increase* the time spent on the page.
  - **Possibility 2:** The new design *increases* the time spent on the page.
- As a recreational website, we wish our new webpage design to increase visitor engagement. Hence,
  - **Possibility 1** represents a *conservative stance*, assuming no difference until proven otherwise.
  - **Possibility 2** is the assertion we hope to *support with evidence* to promote our new design.

Null Hypothesis and Alternative Hypothesis

- In both examples, there are two *competing possibilities*:
  - The first possibility states that there is *no desirable change or difference*.
  - The second possibility states that there is *an effect*.
- Such a pair of possibilities are called the **null hypothesis** and the **alternative hypothesis**.

Null and alternative hypotheses

The **null hypothesis** (H_0) is the hypothesis to be tested. We set up the hypotheses in hopes of finding evidence against H_0 – usually indicating that "nothing happened" or "no change" occurred.

The **alternative hypothesis** (H_1) is the competing claim and represents our question of interest – usually asserting that "something happened" or there is a change.

Deciding Between the Two Hypotheses

Example: "Does a new webpage design increase the time visitors spend on the webpage?"

- For a data-driven answer, we may deploy both the original and re-designed webpage to several users and observe their **sojourn time** on the webpage.
  - **Original design:** $X_1, X_2, \dots, X_n$ with mean $\mu_x = E[X]$.
  - **New design:** $Y_1, Y_2, \dots, Y_n$ with mean $\mu_y = E[Y]$.
- To test the hypothesis, we compare the two population means:
  - **Null hypothesis** $H_0 : \mu_x \geq \mu_y$ (i.e., the new design does not increase sojourn time).
  - **Alternative hypothesis** $H_1 : \mu_x < \mu_y$ (i.e., the new design increases sojourn time).

Deciding Between the Two Hypotheses

We re-expressed our verbal hypothesis ("new design does not increase the sojourn time" versus "new design increases the sojourn time") as a pair of **statistical hypotheses**:

$$H_0 : \mu_x = \mu_y \quad \text{versus} \quad H_1 : \mu_x < \mu_y.$$

statistical hypotheses are questions/statements about the population parameters.

- How do we decide between H_0 and H_1 when *we do not know the population means* μ_x and μ_y?
- Fortunately, we can calculate the sample means $\bar{X}$ and $\bar{Y}$, and compare them:
  - Should we conclude there is no increase (support H_0) whenever $\bar{X} \geq \bar{Y}$?
- Due to randomness, even if the new design is better, it is possible that $\bar{X} \geq \bar{Y}$ in a given sample.

- For demonstration, assume that $X \sim \mathcal{N}(\mu_x, 1)$ and $Y \sim \mathcal{N}(\mu_y, 1)$.
- Then, the distributions of the sample means are:

$$\bar{X} \sim \mathcal{N}\left(\mu_x, \frac{1}{n}\right) \quad \text{and} \quad \bar{Y} \sim \mathcal{N}\left(\mu_y, \frac{1}{n}\right).$$

- Equivalently, we can test the hypothesis on the difference:

$$H_0 : \mu_x - \mu_y = 0 \quad \text{versus} \quad H_1 : \mu_x - \mu_y < 0.$$

- The difference in sample means is then distributed as:

$$\bar{X} - \bar{Y} \sim \mathcal{N}\left(\mu_x - \mu_y, \frac{1}{n} + \frac{1}{n}\right) = \mathcal{N}\left(\mu_x - \mu_y, \frac{2}{n}\right).$$

Probability of a Wrong Decision under H_0

Suppose $H_0 : \mu_x - \mu_y = 0$ **is true**, then

$$\bar{X} - \bar{Y} \sim \mathcal{N}\left(\mu_x - \mu_y, \frac{2}{n}\right) = \mathcal{N}\left(0, \frac{2}{n}\right).$$

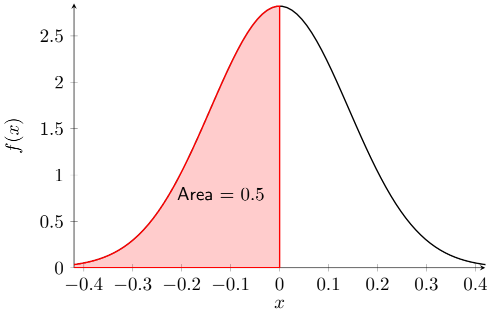
If the sample difference $\bar{X} - \bar{Y} < 0$, we mistakenly reject H_0.

For a symmetric Normal distribution centered at 0, the probability

$$P\left(\bar{X} - \bar{Y} < 0\right) = \text{shaded area} = 0.5.$$

A high chance of making mistake!

PDF of $\bar{X} - \bar{Y}$ with $n = 100$ under $H_0 : \mu_x = \mu_y$



Question: Can we devise a more reliable decision process?

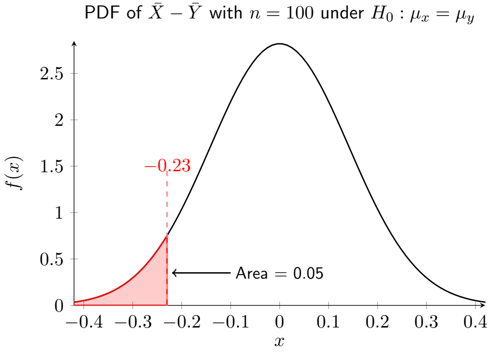
Conservative Decision Rule

- We want the probability of making an incorrect decision to be small, say $\alpha = 0.05$.
- To be conservative, instead of rejecting H_0 when $\bar{X} - \bar{Y} < 0$, we reject H_0 (i.e., claim H_1) *only when*

$$\bar{X} - \bar{Y} < -0.23.$$

- Under this rule, the probability of a wrong decision is small:

$$P\left(\bar{X} - \bar{Y} < -0.23\right) = 0.05.$$



How to Decide Between the Two Hypotheses?

How can we make decisions so that we *don't claim the wrong hypothesis with high probability*?

- In particular, how do we determine the critical threshold (the "magical number" -0.23 from the previous slide)?
- This is the central challenge of hypothesis testing.

This topic

- Upon observing the data, we decide wisely which hypothesis is more consistent with the evidence, **with controlled probability of making any mistake**.

Null and Alternative Hypotheses

Null and alternative hypotheses

The **null hypothesis** (H_0) is the hypothesis to be tested. We set up the hypotheses in hopes of finding evidence against H_0 – usually indicating that "nothing happened" or "no change" occurred.

The **alternative hypothesis** (H_1) is the competing claim and represents our question of interest – usually asserting that "something happened" or there is a change.

- The alternative hypothesis is what we hope to support with evidence.
- Null hypothesis it the hypothesis that we want to reject!
- *Both hypothesis are about the population properties*, not the sample!

Example: Website Design

Statistical hypotheses are often expressed *in terms of the (unknown) population parameters*.

Parameters:

- μ_x: *population mean* sojourn time for the old design.
- μ_y: *population mean* sojourn time for the new design.

Research Question: Does the new design increase the visitor sojourn time?

Statistical Hypotheses:

$$H_0 : \mu_x \geq \mu_y \quad \text{versus} \quad H_1 : \mu_x < \mu_y.$$

Example: Cardiovascular Risk Factors

- Suppose we are interested in high cholesterol levels as an indicator of cardiovascular risk.
- The "average/typical" cholesterol level is 175 mg/dL.
- A group of men who have died from heart disease within the past year are identified, and their cholesterol levels are collected.

Parameters: Let μ be the population mean of the cholesterol levels of all men who have died from heart disease within the past year.

Research Question: Do men who have died from heart disease have higher than the usual cholesterol level?

Statistical Hypotheses:

$$H_0 : \mu = 175 \text{ mg/dL} \quad \text{versus} \quad H_1 : \mu > 175 \text{ mg/dL}.$$

Decisions of a Hypothesis Test

Once hypotheses have been formulated, we need a method for using the sample data to determine whether H_0 should be rejected or not.

Hypothesis test

A hypothesis test is a rule to decide between two competing statistical hypotheses using the sample data.

Possible decisions of a hypothesis test

- **Reject H_0 in favor of H_1:** This indicates that there is sufficient statistical evidence to support that something unusual has happened.
- **Fail to reject H_0 (accept H_0):** This indicates that there is insufficient statistical evidence to support that something unusual has happened.

Types of Errors

In a hypothesis test, the goal is to see if there is sufficient statistical evidence to reject a presumed null hypothesis in favor of a conjectured alternative hypothesis.

Four possible outcomes of a hypothesis test

		Truth	
		H_0	H_1
Decision	Fail to reject H_0	✓	**Type II error**
	Reject H_0	**Type I error**	✓

Type I error = Reject H_0 when H_0 is true.

Type II error = Fail to reject H_0 when H_1 is true (so H_0 is false).

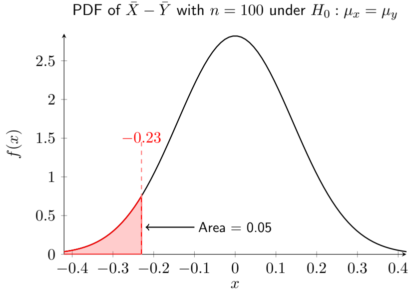
Inevitable Errors in Hypothesis Testing

Ideally, we want to incur no error at all, but due to experimental randomness, some error is inevitable.

Example: Webpage Design Assume H_0 is true (no increase in sojourn time).

- Recall that we claim H_1 when $\bar{X} - \bar{Y} < -0.23$.
- Thus, a Type I error (incorrectly rejecting H_0) occurs whenever $\bar{X} - \bar{Y} < -0.23$.
- The probability of a Type I error is

$$P\left(\bar{X} - \bar{Y} < -0.23\right) = 0.05.$$



No matter how you choose your threshold δ, $P\left(\bar{X} - \bar{Y} < \delta\right) > 0$. There is always a positive chance of making a Type I error!

Rejection (Critical) Region: Website Example

Critical/Rejection region

In hypothesis testing, the **rejection region** is the set of test statistic values for which we reject the null hypothesis H_0.

Example: Webpage Design. Hypotheses:

$$H_0 : \mu_x \geq \mu_y \quad \text{versus} \quad H_1 : \mu_x < \mu_y.$$

- We choose a critical value $\delta = -0.23$ so that if $\bar{X} - \bar{Y} < -0.23$, we reject H_0.
- The rejection region is

$$\bar{X} - \bar{Y} < -0.23.$$

- This threshold is selected to control the probability of Type I error at $\leq \alpha = 0.05$.

How to design a good rejection region?

Key quantities

$$\textbf{Significance level} = \alpha = \mathbb{P}(\text{Type I error})$$
$$\beta = \mathbb{P}(\text{Type II error})$$
$$\textbf{Power} = 1 - \beta = \mathbb{P}(\text{Reject } H_0 \mid H_1 \text{ is true})$$

- Ideally, we want the error rates small and the power high.
- The power is interpreted as the ability to detect the alternative hypothesis and reject the null hypothesis.

Alternative terminology

		Truth	
		H_0	H_1
Decision	Fail to reject H_0	**true negative** (prob. $= 1-\alpha$)	**false negative** (prob. $= \beta$)
	Reject H_0	**false positive** (prob. $= \alpha$)	**true positive** (prob. $=$ power)

Example: Flip a coin 10 times, let X_i denote the outcome of the ith flip.
$H_0 : p = 0.5.$

- Rejection region $\sum_{i=1}^{10} X_i \leq 2$ or ≥ 8. What is α_1?
- Rejection region $\sum_{i=1}^{10} X_i \leq 1$ or ≥ 9. What is α_2?
- Rejection region $\sum_{i=1}^{10} X_i \leq 2$ or ≥ 8, and $H_1 : p = 0.25$. What is β_1 and power?
- Rejection region $\sum_{i=1}^{10} X_i \leq 1$ or ≥ 9, and $H_1 : p = 0.25$. What is β_2 and power?

Solution: Under H_0, we have $X = \sum_{i=1}^{10} X_i \sim \text{Bin}(10, 0.5)$:

$$\alpha_1 = \mathbb{P}(X \leq 2)+\mathbb{P}(X \geq 8) = \binom{0}{10}0.5^0 0.5^1 0+\binom{1}{10}0.5^1 0.5^9+\binom{2}{10}0.5^2 0.5^8+\binom{8}{10}0.5^8 0.5^2+\binom{9}{10}0.5^9 0.5^1+\binom{10}{10}0.5^{10} 0.5^0 \approx 0.1094$$

$$\alpha_2 = \mathbb{P}(X \leq 1) + \mathbb{P}(X \geq 9) = \binom{0}{10}0.5^0 0.5^{10} + \binom{1}{10}0.5^1 0.5^9 + \binom{9}{10}0.5^9 0.5^1 + \binom{10}{10}0.5^1 0 0.5^0 \approx 0.0215.$$

Under H_1, we have $X = \sum_{i=1}^{10} X_i \sim \text{Bin}(10, 0.25)$:

$$\beta_1 = \mathbb{P}(2 < X < 8) = \sum_{i=3}^{7} \binom{i}{10} 0.25^i (1-0.25)^{10-i} \approx 0.4740 \quad \text{and} \quad \text{power} = 1 - \beta_1 \approx 0.5260.$$

$$\beta_2 = \mathbb{P}(1 < X < 9) = \sum_{i=2}^{8} \binom{i}{10} 0.25^i (1-0.25)^{10-i} \approx 0.7559 \quad \text{and} \quad \text{power} = 1 - \beta_2 \approx 0.2441.$$

Observation from the previous example

- When H_1 is not an equality, the power depends on the true parameter, under which you calculate the probability.
- As the rejection region shrinks, the probability of type I error reduces.
  - In a test with small α, rejecting $H_0 =$ VERY strong evidence against H_0.
- As the rejection region shrinks, the probability of type II error increases, and so the power decreases.
  - However, small α also result in increased risk of type II error.

Ideally, we want the probability of both error to be small at the same time. Unfortunately, this is not achievable.

There is a trade off between α and β!

Neyman-Person Paradigm

With the trade off between α and β. What we usually do is:

- Control the type I error rate (significance level, α) by setting a tolerance level.
  - We usually target $\alpha \leq 0.1$, 0.05 or even 0.01.
- *Minimize type II error rate when possible. (In advanced statistic course.)
  - Instead of rigorously find the "most powerful", we present intuitive ways to construct good enough tests.

Type I error is usually more serious

The reason for the primary focus to be on controlling the type I error is because we are more concerned about making a false positive claim (claiming there is an effect when there isn't) than making a false negative claim (failing to detect an effect when there is one).

Type I and Type II Errors in Different Contexts

What are the type I and type II errors in the following situations?

- "H_0 Not Guilty versus H_1 Guilty" or "H_0 Guilty versus H_1 Not Guilty"?
- "H_0 Spam versus H_1 Not Spam" or "H_0 Not Spam versus H_1 Spam"?
- "H_0 Healthy versus H_1 Sick" or "H_0 Sick versus H_1 Healthy"?
- "H_0 Drug is not safe versus safe" or "H_0 is safe versus H_1 not safe"?

Type I Error is More Serious Than Type II Error

Example: Criminal Justice: H_0: Not Guilty, H_1: Guilty.

- **Type I Error:** Convicting an innocent person (rejecting "Not Guilty" when it is true).
- **Type II Error:** Acquitting a guilty person (failing to reject "Not Guilty" when it is false).

Example: Spam Detection: If we set H_0: Email is Not Spam and H_1: Email is Spam:

- **Type I Error:** Misclassifying a legitimate email as spam.
- **Type II Error:** Failing to detect an actual spam email.

Example: Health Diagnosis: If H_0: Healthy and H_1: Sick:

- **Type I Error:** Diagnosing a healthy person as sick (false positive).
- **Type II Error:** Failing to diagnose a sick person (false negative).

Example: Drug Safety: If H_0: Drug is Safe and H_1: Drug is Not Safe:

A Normal Example

Example: Samples $X_1, \ldots, X_n$ are generated from $\mathcal{N}(\mu, \sigma^2)$ with unknown μ and known σ^2. <u>We suspect that the mean is not μ_0</u>.

- Hypotheses

$$H_0 : \mu = \mu_0. \quad vs. \quad \underline{H_1 : \mu \neq \mu_0}.$$

- Consider the statistic below

Z-statistic

Assume that X is normally distributed with unknown mean μ and known variance σ.
Under $H_0 : \mu = \mu_0$

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1) \quad \text{and observed value} \quad z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

The Z-statistic quantifies how far $\bar{x}$ is from μ_0 in its standard deviation units.

When should we reject $H_0 : \mu = \mu_0$?

- Intuitively, if we observe a $\bar{x}$ that is far smaller or far greater than μ_0, we should to reject H_0.
- This is equivalent to observing a

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

  that is far smaller or far greater than 0.
- The rejection region should be

$$R = \{|z| > c\},$$

  for critical value c to be determined.

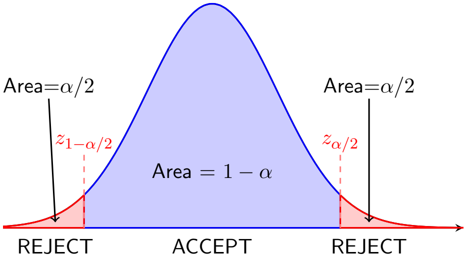
Suppose the rejection region is $R = \{|z| > c\}$.

- When the probability of type I error equals α, we can calculate the critical value:

$$\alpha = \mathbb{P}\left(Z \in R | H_0\right) = \mathbb{P}\left(|Z| > c) | H_0\right)$$

Results in $c = z_{\alpha/2}$.

- The rejection region is then



$$R = \left\{|z| > z_{\alpha/2}\right\} = \left\{\bar{x} \geq \mu_0 + z_{\alpha/2}\sigma/\sqrt{n}\right\} \cup \left\{\bar{x} \leq \mu_0 - z_{\alpha/2}\sigma/\sqrt{n}\right\}$$

Observations

The type I error rate α can always be reduced by appropriate selection of the critical values, i.e., $z_{\alpha/2} \leftrightarrow \alpha$.

Decision Procedure and Critical Region

In summary, we choose the following form of the critical region:

$$R = \left\{x_1, x_2, \ldots, x_n : \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_{\alpha/2}\right\} = \{z : z > z_{\alpha/2}\},$$

- $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ is called the **observed test statistic**.
- The critical value $z_{\alpha/2}$ is the **cutoff point** that defines the boundary of the rejection region.
- The critical region R has a Type I error probability of exactly α.

Testing Procedure

1. Collect observations $x_1, x_2, \ldots, x_n$ and compute observed test statistic $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$.
2. Check if $x_1, x_2, \ldots, x_n \in R$, i.e., if $|z| > z_{\alpha/2}$:
   - If $|z| > z_{\alpha/2}$, reject H_0.
   - If $|z| < z_{\alpha/2}$, accept H_0.

Example: Effective Substance in a Medicine

The effective substance in a medicine needs to be 8.5 mg per tablet. Suppose you sample 5 tablets and obtain a sample mean of 8.8 mg. Assume that the amount follows a normal distribution with mean μ and variance 1. We want to test the hypothesis

$$H_0 : \mu = 8.5 \quad \text{versus} \quad H_1 : \mu \neq 8.5,$$

at a significance level of $\alpha = 0.05$.

Solution:

- Compute $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{8.8-8.5}{\sqrt{1/5}} = \frac{0.3}{\sqrt{0.2}} \approx 0.671$.
- Compare z with $z_{0.025} = 1.96$. Since $0.671 < 1.96$, we do not reject H_0.
- **Conclusion:** There is insufficient evidence to reject the claim. We accept H_0 and conclude that the effective substance in the medicine meets the standard of 8.5 mg per tablet.

Example: (Cont'd) Effect of Sample Size

- What if the sample mean of 8.8 mg is obtained from 100 sampled tablets?
- Compute the observed test statistic:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{8.8 - 8.5}{\sqrt{1/100}} = \frac{0.3}{0.1} = 3.$$

- Compare with the critical value $z_{0.025} = 1.96$ we see that $z = 3 > 1.96$.
- **Conclusion:** Since $z > z_{\alpha/2}$, we reject H_0 and claim that the effective substance in the medicine fails to meet the standard of 8.5 mg per tablet.

Observations

With the same distance between the sample mean of 8.8 mg and the hypothesize value of 8.5 mg, the conclusion of the test may change when sample size changes. A larger sample size make it easier to reject when observing the same sample mean.

Some Intuitions

We decide by comparing test statistic $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ with the critical value $z_{\alpha/2}$.

- **Effect of the Sample Mean:**
  If the sample mean deviates more from μ_0, then z becomes larger in absolute value. Consequently, it is more likely that z will exceed the critical value $z_{\alpha/2}$, leading to rejection of H_0.
- **Effect of the Sample Size:**
  With all else fixed, a larger sample size n decreases the denominator $\sigma/\sqrt{n}$. This increases the absolute value of z, making it easier to reject H_0.
- **Effect of the Significance Level:**
  Increasing α makes $z_{\alpha/2}$ larger, which in turn makes the rejection region larger and the test less conservative (i.e., easier to reject H_0). Conversely, a smaller α leads to a smaller $z_{\alpha/2}$, making the test more conservative.

One-Sided Hypothesis Test for Mean with Known Variance

We wish to test whether the population mean μ **exceeds** a specified value μ_0, i.e.,

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu > \mu_0.$$

- Intuitively, if the sample mean $\bar{x}$ is much higher than μ_0, we reject H_0.
- We define the critical region as

$$R = \{x_1, x_2, \ldots, x_n : \bar{x} - \mu_0 > c\}.$$

- In standardized form, let $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$, then the critical region becomes (one-sided)

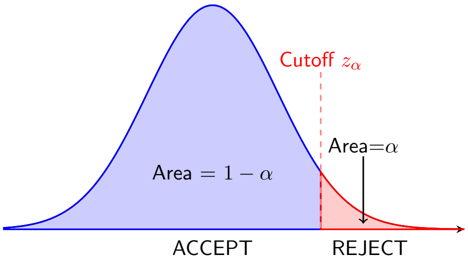
$$R = \{z : z > c/(\sigma/\sqrt{n})\}.$$

- How to determine c in order to control the Type I error at significance level α?

Similar to the two-sided test, we seek to control the probability of a Type I error (rejecting H_0 when it is true) to be less than the significance level α.

- Under the null hypothesis, the test statistic is

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1).$$



- Therefore, the probability of a Type I error is

$$P\Big(\bar{X}-\mu_0 > c\Big) = P\Big(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{c}{\sigma/\sqrt{n}}\Big) = P\Big(Z > \frac{c}{\sigma/\sqrt{n}}\Big).$$

To ensure this probability equals α, we set

$$\frac{c}{\sigma/\sqrt{n}} = z_\alpha \quad \Longrightarrow \quad \boxed{c = z_\alpha \frac{\sigma}{\sqrt{n}}.}$$

Decision Procedure and Critical Region (One-Sided Test, Right-tailed)

In testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$, we choose the following critical region:

$$R = \left\{x_1, x_2, \ldots, x_n : \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha\right\} = \{z : z > z_\alpha\}.$$

- $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ is called the **observed test statistic**.
- The critical region C is chosen such that the probability of a Type I error (rejecting H_0 when it is true) is exactly α.

Testing Procedure

1. Collect observations $x_1, x_2, \ldots, x_n$ and compute observed test statistic $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$.
2. Check whether the computed z falls in the critical region R, i.e., if $z > z_\alpha$:
   - If $z > z_\alpha$, reject H_0.
   - If $z \leq z_\alpha$, fail to reject (accept) H_0.

Decision Procedure and Critical Region (One-Sided Test, Left-tailed)

In testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$, we choose the following critical region:

$$R = \left\{ x_1, x_2, \ldots, x_n : \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -z_\alpha \right\} = \{ z : z < -z_\alpha \}.$$

- $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ is called the **observed test statistic**.
- The critical region C is chosen such that the probability of a Type I error (rejecting H_0 when it is true) is exactly α.

Testing Procedure

1. Collect observations $x_1, x_2, \ldots, x_n$ and compute observed test statistic $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$.
2. Check whether the computed z falls in the critical region R, i.e., if $z < -z_\alpha$:
   - If $z < -z_\alpha$, reject H_0.
   - If $z \geq -z_\alpha$, fail to reject (accept) H_0.

Measuring Evidence Against H_0

Example: Speeding Offense: Decide whether a driver commits a speeding offense:

$$H_0 : s = s_0 \quad \text{(not speeding)} \quad \text{versus} \quad H_1 : s > s_0 \quad \text{(speeding)},$$

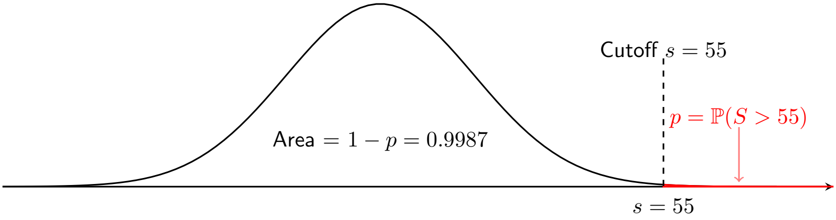
where $s_0 = 40$ km/h is the speed limit.

- Assume driving speed follows a normal distribution: $S \sim \mathcal{N}(40, 25)$.
- For a driver with speed 55 km/h, standardize:

$$Z = \frac{S - 40}{5} \sim \mathcal{N}(0,1) \quad \Rightarrow \quad \mathbb{P}(S \geq 55) = P\left(Z \geq \frac{15}{5}\right) = \mathbb{P}(Z \geq 3) \approx 0.0013.$$

- Only 0.13% of drivers would be faster than 55 km/h.
- $\Rightarrow$ Strong evidence against not speeding.
- **Conclusion:** With such strong evidence against H_0, a speeding ticket is warranted.

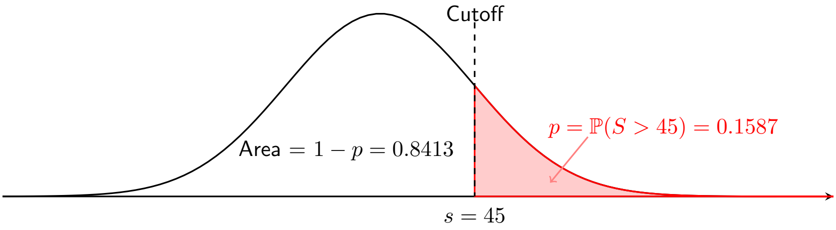
p-Value



- For a driver traveling at 55 km/h, the p-value is given by

$$p = \mathbb{P}(S \geq 55) = P\left(\frac{S-40}{5} \geq \frac{15}{5}\right) = \mathbb{P}(Z \geq 3) \approx 0.0013.$$

- This means that only 0.13% of all drivers would be driving "**more extreme**(faster)" than the driver at question.
- This probability p is called **the p-value**.



Consider another driver traveling at 45 km/h:

- The p-value is

$$p = \mathbb{P}(S \geq 45) = P\left(\frac{S - 40}{5} \geq \frac{45 - 40}{5}\right) = \mathbb{P}(Z \geq 1) \approx 0.1587.$$

- With a p-value of 0.1587, about 15.87% of all drivers would drive **more extremely**.
- If this driver is ticketed, then we would ticket at least 15.87% of all drivers – clearly too many.

p-Value

p-value

Given an observed t (e.g. s in speeding example), the evidence against H_0 can be measured by **the p-value**, defined as the probability of observing a *more extreme (unlikely)* value of the test statistic T (e.g. S in speeding example) outcome than t, under the null hypothesis $H_0 : \mu = \mu_0$.

- The p-value depends on the **observed sample** t.
- It measures, for an observation, the strength of evidence against H_0.
- The smaller the p-value, the stronger the evidence against H_0. (Because fewer people would be expected to act "more extremely".)

Meaning of "More Extreme" Depends on H_1

If $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$:

- When z is large, $\bar{x}$ is much larger than μ_0. This suggests that μ is more likely to be larger than μ_0, providing evidence in support of H_1.
- The p-value is $p = \mathbb{P}(Z > z), \quad Z \sim N(0,1)$.

If $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$:

- When z is small, $\bar{x}$ is much smaller than μ_0. This suggests that μ is more likely to be smaller than μ_0, providing evidence in support of H_1.
- The p-value is $p = \mathbb{P}(Z < z), \quad Z \sim N(0,1)$.

If $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$:

- When $|z| = \left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right|$ is large, $|\bar{x} - \mu_0|$ is large, indicating that μ is likely to differ from μ_0. Hence, a large $|z|$ is considered more "extreme."
- The p-value is $p = P\left(|Z| > |z|\right), \quad Z \sim N(0,1)$.

Connection to the Rejection Region

Consider the one-sided test as an example:

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu > \mu_0.$$

- The p-value is defined as $p = P\Big(Z > z\Big), \quad Z \sim N(0,1)$, where $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ is the observed test statistic.
- The rejection region is given by

$$R = \{z : z > z_\alpha\}.$$

- Notice that $p \leq \alpha$ is equivalent to $z > z_\alpha$.

Decision rule using p-values

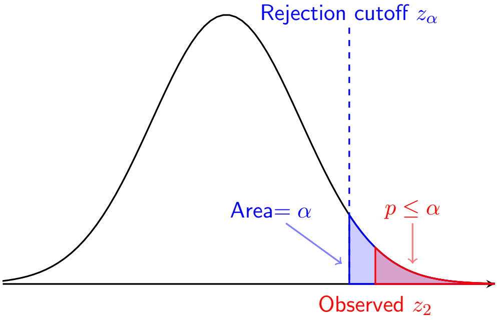
- If $p \leq \alpha$ (i.e., $z \in C$), reject H_0.
- If $p > \alpha$, fail to reject (accept) H_0.

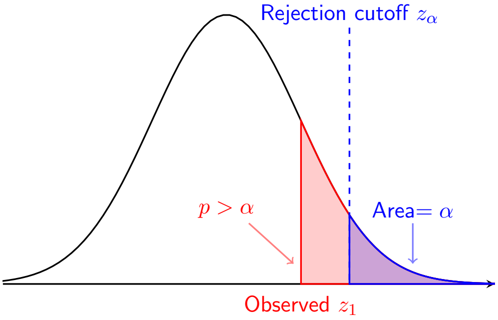
Equivalence of rejection rule and p-value decision rule

- If $p \leq \alpha$ if and only if $z \in C$, then we reject H_0.
- If $p > \alpha$ if and only if $z \notin C$, then we fail to reject (accept) H_0.

$p \leq \alpha$ implies rejection of H_0

$p > \alpha$ implies failure to reject H_0





Exercise: Verification of Rejection Rules

1. **One-Sided Test:** $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$
   - The test statistic is
   $$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}.$$
   - The p-value is defined as
   $$p = \mathbb{P}(Z < z), \quad Z \sim N(0,1).$$
   - To reject H_0 at significance level α, we require $p \leq \alpha$.
   - Since $\mathbb{P}(Z < -z_\alpha) = \alpha$, the condition $p \leq \alpha$ is equivalent to
   $$z < -z_\alpha.$$
   - Hence, rejecting H_0 when $z < -z_\alpha$ is equivalent to $p \leq \alpha$.

Exercise: Verification of Rejection Rules

2. Two-Sided Test: $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$

- The test statistic is

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}.$$

- The p-value is defined as

$$p = \mathbb{P}(|Z| > |z|), \quad Z \sim N(0,1).$$

- To reject H_0 at significance level α, we require $p \leq \alpha$.
- Since $\mathbb{P}(|Z| > z_{\alpha/2}) = \alpha$, the condition $p \leq \alpha$ is equivalent to

$$|z| > z_{\alpha/2}.$$

- Thus, rejecting H_0 when $|z| > z_{\alpha/2}$ is equivalent to $p \leq \alpha$.

Hypothesis Testing using p-values

Procedure:

1. Collect observations $x_1, x_2, \ldots, x_n$ and compute $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$.
2. Compute the p-value associated with the observed z:
   - $p = \mathbb{P}(Z > z), \quad Z \sim N(0,1)$, if $H_1 : \mu > \mu_0$.
   - $p = \mathbb{P}(Z < z), \quad Z \sim N(0,1)$, if $H_1 : \mu < \mu_0$.
   - $p = \mathbb{P}(|Z| > |z|), \quad Z \sim N(0,1)$, if $H_1 : \mu \neq \mu_0$.
3. **Decision:** For a prespecified significance level α:
   - If $p \leq \alpha$, reject the null hypothesis H_0.
   - If $p > \alpha$, accept (fail to reject) H_0.

   Alternatively, use rejection regions:
   - If $z > z_\alpha$, reject the null hypothesis H_0 for $H_1 : \mu > \mu_0$.
   - If $z < -z_\alpha$, reject the null hypothesis H_0 for $H_1 : \mu < \mu_0$.
   - If $|z| > z_{\alpha/2}$, reject the null hypothesis H_0 for $H_1 : \mu \neq \mu_0$.

Two Equivalent Procedures for Rejection

Two equivalent procedures for rejection

1. Specified a rejection region with significance level α, observe the data, and reject if the data falls in the rejection region.
2. **[The usual way]** Observe the data, calculate the p-value, and reject if the $p \leq \alpha$.

H_1	Rejection region	p-value
$\mu \neq \mu_0$	$\lvert z \rvert > z_{\alpha/2}$	$\mathbb{P}(\lvert Z \rvert > \lvert z \rvert \mid H_0 \text{ true})$
$\mu > \mu_0$	$z > z_\alpha$	$\mathbb{P}(Z > z \mid H_0 \text{ true})$
$\mu < \mu_0$	$z < z_\alpha$	$\mathbb{P}(Z < z \mid H_0 \text{ true})$

*$Z \sim N(0,1)$

One-Sample Tests for Normal

- One-sample t-Test for normal mean with unknown variance.
- One-sample χ^2-test for normal variance and standard deviation

One-Sample t-Test – Normal Mean with Unknown Variance

Consider a normal random sample where μ is unknown and σ is unknown.

Recall from Confidence Interval (Slide 38) that we introduced the t-distribution as the distribution of standardized sample mean when the variance is unknown.

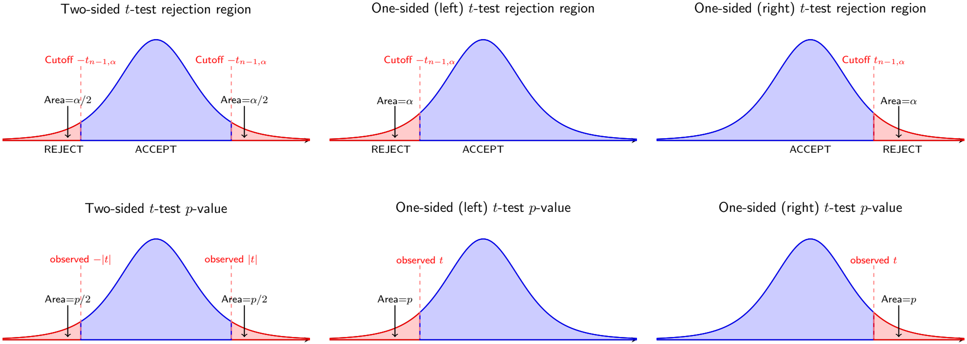
T-statistic

Assume that X is normally distributed with unknown mean μ and unknown variance σ^2. Under $H_0 : \mu = \mu_0$,

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{(\bar{X} - \mu_0)/(\sigma/\sqrt{n})}{S/\sigma} = \frac{N(0,1)}{\sqrt{\chi^2_{n-1}/(n-1)}} \sim t_{n-1}$$

where t_{df} denotes a t-distribution with degree of freedom df.

One-Sample t-Test – Normal Mean with Unknown Variance



H_1	Rejection region	p-value
$\mu \neq \mu_0$	$\lvert t\rvert > t_{n-1,\alpha/2}$	$\mathbb{P}(\lvert T\rvert > \lvert t\rvert \mid H_0)$
$\mu > \mu_0$	$t > t_{n-1,\alpha}$	$\mathbb{P}(T > t \mid H_0)$
$\mu < \mu_0$	$t < -t_{n-1,\alpha}$	$\mathbb{P}(T < t \mid H_0)$

*$T \sim t_{n-1}$, $\mathbb{P}(T > t_{n-1,\alpha}) = \alpha$.

Example – One-Sample t-Test (One-Sided)

Example: In the smart student example, assume that the standard deviation of population IQ is unknown. From the random sample of 9 students, we can calculate $\bar{x} = 112.8$ and $s = 12.7$.

- **Null and alternative hypothesis**

$$H_0 : \mu = 100. \quad vs. \quad H_1 : \mu > 100.$$

- **Test statistic**

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1} \text{ under } H_0, n = 9$$

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{112.8 - 100}{12.7/\sqrt{9}} \approx 3.02.$$

- p**-value**

$$p = \mathbb{P}(T > 3.02 \mid H_0) = 0.0083 < 0.05 = \alpha.$$

- **Conclusion:** Reject the H_0 at significance level 0.05.
- Alternatively, find the rejection region $\{t : t > t_{n-1,\alpha} = 1.8595\}$.

Example – One-Sample t-Test (Two-Sided)

Example: An insurance company is reviewing its current policy rates. When originally setting the rates, they believed that the average claim amount was \$1,800. They are concerned that the true mean is actually different from this:

- If the mean is much higher than \$1,800, they lose money because of the claims.
- If the mean is much lower than \$1,800, they lose clients because they may be charging too much for the insurance.

We assume the claim follows a normal distribution with a unknown variance.

They randomly select 100 claims, and calculate a sample mean of \$1,650 and sample standard deviation of \$700.

Question: Consider a test at significance level $\alpha = 0.05$ to see if the insurance company should be concerned.

- **Null and alternative hypothesis**

$$H_0 : \mu = 1800. \quad vs. \quad H_1 : \mu \neq 1800.$$

- **Test statistic**

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1} \text{ under } H_0, n = 100,$$

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{1650 - 1800}{700/\sqrt{100}} \approx -2.14.$$

- p-**value**:

$$p = \mathbb{P}(|T| > |-2.14| \mid H_0) = 0.0348 < 0.05 = \alpha.$$

- **Conclusion:** Reject the H_0 at significance level 0.05. There are significant evidence that the mean of the claims is not \$1,800.
- Alternatively, find the rejection region $\{t : t > t_{100-1,0.05} \approx z_{0.05} = 1.96\}$. Since $|-2.14| > 1.96$ reject H_0 at significance level 0.05.

One-sample χ^2-test – Normal Variance and Standard Deviation

Now, we want to test hypotheses regarding the variance of a normal population.

Assumptions

- Normal random sample or large sample.
- Both mean μ and the variance σ^2 are unknown.

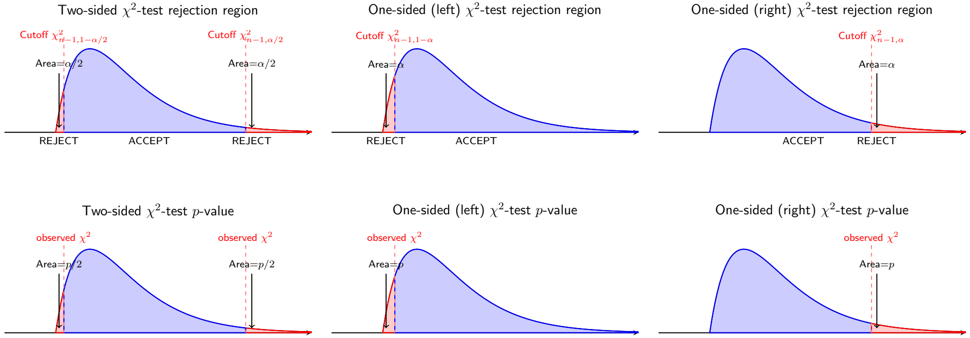
Hypotheses

$$H_0 : \sigma^2 = \sigma_0^2. \quad vs. \quad H_1 : \sigma^2 \neq \sigma_0^2.$$

Test statistic

$$X^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1} \text{ under } H_0.$$

where χ^2_{df} denotes a χ^2-distribution with degree of freedom df.



H_1	Rejection region	p-value
$\sigma^2 \neq \sigma_0^2$	$x^2 > \chi^2_{n-1,\alpha/2}$ or $x^2 < \chi^2_{n-1,1-\alpha/2}$	$2 \times \min\{\mathbb{P}(X > x^2 \mid H_0), \mathbb{P}(X^2 < x^2 \mid H_0)\}$
$\sigma^2 > \sigma_0^2$	$x^2 > \chi^2_{n-1,\alpha}$	$\mathbb{P}(X^2 > x^2 \mid H_0)$
$\sigma^2 < \sigma_0^2$	$x^2 < \chi^2_{n-1,1-\alpha}$	$\mathbb{P}(X^2 < x^2 \mid H_0)$

*$X^2 \sim \chi^2_{n-1}$, $\mathbb{P}(X^2 > \chi^2_{n-1,\alpha}) = \alpha$.

Example – One-Sample χ^2-Test for Normal Standard Deviation

Example: The lapping process is used to grind certain silicon wafers to the proper thickness. It is acceptable only if σ, the population standard deviation is at most 0.5 mil (thousandth of an inch). Suppose we have a sample of size 15 and observed a sample standard deviation of 0.64 mil.

Question: Set $\alpha = 0.05$, is the sample significant enough to support the claim that the population variance $\sigma > 0.5$?

- **Null and alternative hypotheses**

$$H_0 : \sigma = \sigma_0 = 0.5. \quad vs. \quad H_1 : \sigma > 0.5.$$

- **Test statistic**

$$X^2 = (n-1)S^2/\sigma_0^2 \sim \chi^2_{n-1} \text{ under } H_0, n = 15$$

$$x^2 = (n-1)s^2/\sigma_0^2 = (15-1) \times 0.64^2/0.5^2 \approx 22.94.$$

- p-**value**

$$p = \mathbb{P}(X^2 > x^2 \mid H_0) = \mathbb{P}(X^2 > 22.94 \mid H_0) = 0.061 > 0.05 = \alpha.$$

- **Conclusion:** We fail to reject the null hypothesis "no difference" at significance level 0.05.
- Alternatively, find the rejection region: $\{x^2 : x^2 > \chi^2_{n-1,\alpha} = 23.68\}$. Note that $22.94 < 23.68$, fail to reject H_0.

Connection Between Confidence Intervals and Hypothesis Tests

Confidence intervals and hypothesis tests are two sides of the same coin.

- Confidence intervals provide a range of plausible values for a population parameter.
- Hypothesis tests offer a formal procedure to decide whether to accept or reject a specific claim about that parameter.
- In many cases, the results of a hypothesis test at a given significance level α can be directly interpreted through the corresponding $100(1-\alpha)\%$ confidence interval.

Connection Between Confidence Intervals and Hypothesis Tests

Example:

1. For testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ at level α
   - We construct a two-sided $100(1-\alpha)\%$ confidence interval for μ.
   - If the interval contains μ_0, we do not reject H_0.
2. For testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$ at level α
   - We construct a one-sided $100(1-\alpha)\%$ confidence **lower bound** for μ.
   - If the bound contains μ_0, we do not reject H_0.
3. For testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$ at level α,
   - We construct a one-sided $100(1-\alpha)\%$ confidence **upper bound** for μ.
   - If the bound contains μ_0, we do not reject H_0.

Hypothesis Testing and Confidence Interval

Connections

- Typically, conclusion from a hypothesis test can be reached using a CI.
- If the observed $(1-\alpha)\times 100\%$ CI does not contain the hypothesized value specified in the null hypothesis, we reject H_0 in favor of a two-sided H_1 at significance level α.