STA2002: Probability and Statistics II

Hypothesis Testing I

Fangda Song and Ka Wai Tseng

School of Data Science, CUHK(SZ)

September, 2025

Introduction

In this lecture we will introduce

- Important concepts in hypothesis testing
  - Null hypothesis H_0 and alternative hypothesis H_1
  - Test statistic T and critical region C
  - Type I and Type II error
  - Significance level α
  - p-value
- Hypothesis testing for mean
- Suggested reading: Chapter 8.1 of the textbook.

Motivating example

Example

Let X equal the breaking strength of a steel bar. If the bar is manufactured by process I, $X \sim N(50, 36)$. It is hoped that if process II (a new process) is used, $X \sim N(55, 36)$. Given a large number of steel bars manufactured by process II, how could we test whether the five-unit increase in the mean breaking strength was realized?

- Based on the observed the breaking strength of steel bars under process II
- Make a judgment: five-unit increase or not

Formulate H_0 and H_1

Suppose we have $X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, 36)$

- **Null hypothesis** H_0: An initial belief/claim/assumption on the values that μ can take
- **Alternative hypothesis** H_1: The competing belief/claim/assumption against which we would like to test the initial belief/claim/assumption
- **Simple hypothesis**: a hypothesis in which μ only takes on one value.
- **Composite hypothesis**: a hypothesis in which μ takes on a range of values.

Example

Simple null hypothesis $H_0 : \mu = 50$
Composite null hypothesis $H_0 : \mu \leq 50$
Simple alternative hypothesis $H_1 : \mu = 55$
Composite alternative hypothesis $H_1 : \mu > 55$

Formulate H_0 and H_1

Different types of hypothesis

- $H_0 : \mu = 50$, $H_1 : \mu = 55$.
- (One-sided hypothesis test) $H_0 : \mu = 55$, $H_1 : \mu > 55$.
- (One-sided hypothesis test) $H_0 : \mu = 55$, $H_1 : \mu < 55$.
- (Two-sided hypothesis test) $H_0 : \mu = 55$, $H_1 : \mu \neq 55$.
- $H_0 : \mu \leq 55$, $H_1 : \mu > 55$.

Critical Region C and Test Statistic T

- Denote the space of the sample by $\mathcal{D}$, that is,

$$\mathcal{D} := \{(x_1, x_2, \ldots, x_n) | x_i \in S_X, i = 1, 2, \cdots, n\}.$$

- A test of H_0 versus H_1 is based on a subset C of $\mathcal{D}$. This set C that we reject H_0 is called the **critical region**
- The region is usually specified in terms of **test statistic** T.

Example

In the previous example, we have
Test statistics $T = \bar{X} \sim N(\mu, 36/n)$
Critical region $C = \{(x_1, x_2, \cdots, x_n) \in \mathcal{D} | \bar{x} \geq 53\}$

Decision Rule

- Once we formulate H_0 and H_1, we collect data, say $x_1, \ldots, x_n$, to see whether H_0 is favoured or H_1 is favoured.
- Then we can come up with a corresponding decision rule, based on the data $x_1, \ldots, x_n$:
  - Reject H_0 (Accept H_1), if $(x_1, \ldots, x_n) \in C$,
  - Retain H_0 (Reject H_1), if $(x_1, \ldots, x_n) \notin C$.

Example

If $\bar{x} \geq 53$, we reject H_0. Otherwise, if $\bar{x} < 53$, we fail to reject H_0.

Recall the example

Example

Let X equal the breaking strength of a steel bar. If the bar is manufactured by process I, $X \sim N(50, 36)$. It is hoped that if process II (a new process) is used, $X \sim N(55, 36)$. Given a large number of steel bars manufactured by process II, how could we test whether the five-unit increase in the mean breaking strength was realized?

Suppose we have $X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, 36)$

- Simple null hypothesis $H_0 : \mu = 50$
- Simple alternative hypothesis $H_1 : \mu = 55$
- Critical region $C = \{x \geq 53\}$
- Test statistics $\bar{X} \sim N(\mu, 36/n)$
- Decision rule: if $\bar{x} \geq 53$, we reject H_0. Otherwise, if $\bar{x} < 53$, we fail to reject H_0.

Errors in Hypothesis Testing

We may draw incorrect conclusion and commit errors in hypothesis testing. It can happen in two ways:

	H_0 true	H_1 true
Reject H_0	Type I error	Correct
Fail to Reject H_0	Correct	Type II Error

- Type I error: Reject H_0 when H_0 is true.
- Type II error: Fail to reject H_0 when H_0 is false.

Errors in Hypothesis Testing

The probability of a type I error

$$\alpha = \Pr(\text{Test makes Type I Error}) = \Pr(T \in C; H_0).$$

The probability of a type II error

$$\beta = \Pr(\text{Test makes Type II Error}) = \Pr(T \notin C; H_1).$$

Example

$X_1, \dots, X_n \overset{i.i.d.}{\sim} N(\mu, 36)$ with $n = 16$. We take the critical region $C = \{\bar{x} \geq 53\}$. What is the probability of a type I error and a type II error in the following test

$$H_0 : \mu = 50 \text{ against } H_1 : \mu = 55?$$

Errors in Hypothesis Testing

Solution: We know test statistic $T = \bar{X} \sim N(\mu, 36/n)$. Under H_0,

$$\bar{X} \sim N(50, 36/16),$$

and under H_1,

$$\bar{X} \sim N(55, 36/16).$$

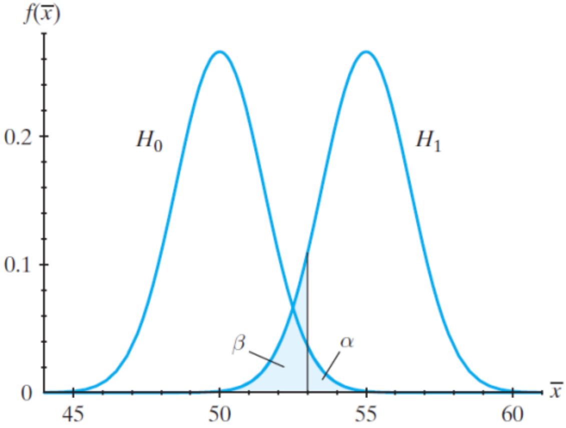
As a result,

$$\alpha = \Pr(\bar{X} \geq 53; H_0) = \Pr\left(Z \geq \frac{53-50}{\sqrt{36/16}}\right) = 0.0228.$$

$$\beta = \Pr(\bar{X} < 53; H_1) = \Pr\left(Z < \frac{53-55}{\sqrt{36/16}}\right) = 0.0913.$$

Note: by changing the critical region C, α increases (decreases) while β decreases (increases), so both errors cannot be reduced at the same time.

Errors in Hypothesis Testing



PDF of $\bar{X}$ under $H_0 : \mu = 50$ and $H_1 : \mu = 55$.

Significance Level

- Most of the time we fix the probability of type I error α, for example, we take $\alpha = 0.05 = 5\%$, and determine the corresponding critical region.
- The value of α is called the **significance level** of the test.
- We can also be more conservative, and for example, take $\alpha = 0.01$.

Example

Let

$$\Pr(\bar{X} \geq c; H_0) = \Pr\left(Z \geq \frac{c - 50}{\sqrt{36/16}}\right) = 0.05,$$

we have $\frac{c-50}{\sqrt{36/16}} = z_{0.05} \Rightarrow c = 52.48$

p-value: Observed Significance Level

- p-value is the probability of observing a more extreme test statistic (in the direction that favours H_1), given that H_0 is true.
- A small p-value indicates the data is unlikely coming from H_0, and so the data is perhaps better explained by H_1.
- We reject H_0 if the p-value $\leq \alpha$, where α is the significance level.
- p-value is also called the observed significance level.

Example

Given the observed test statistics $\bar{x} = 56$, "more extreme" means:

- $\bar{X} \geq 56$, for $H_0 : \mu = 50$ versus $H_1 : \mu = 55$;
- $\bar{X} \geq 56$, for $H_0 : \mu = 55$ versus $H_1 : \mu > 55$,
- $\bar{X} \leq 56$, for $H_0 : \mu = 55$ versus $H_1 : \mu < 55$,
- $\bar{X} \geq 56$ or $\bar{X} \leq 54$, for $H_0 : \mu = 55$ versus $H_1 : \mu \neq 55$,

Hypothesis Testing for Normal Mean

Suppose that $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$. We would like to conduct hypothesis testing on μ, and the discussion is separated into the following two cases:

- Case 1: σ^2 is known
- Case 2: σ^2 is unknown

Three Approaches to Hypothesis Testing

- P-value approach:
  - Compute p-value and compare it to significance level α
  - Reject H_0 if p-value $\leq \alpha$
- Critical region approach
  - Define rejection region under H_0 and calculate test statistic
  - Reject H_0 if statistic falls in critical region
- Confidence interval approach
  - Construct $(1 - \alpha)$ confidence interval and check if null value is included in CI
  - Reject H_0 if null value is not included in CI

These three approaches are equivalent!

Case 1: σ^2 is known

- We consider the test statistic

$$T := \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}.$$

- Under $H_0 : \mu = \mu_0$, $T \sim N(0, 1)$.
- Once we observe the data $x_1, \ldots, x_n$, we can compute

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

Case 1: σ^2 is known

For the one-sided test $H_0 : \mu = \mu_0$, $H_1 : \mu > \mu_0$

- p-value is
$$p = \Pr(T > z; H_0) = \Pr(Z > z),$$
- Reject H_0 at significance level α when,

$$\begin{aligned}
&p\text{-value} = \Pr(Z > z) \leq \alpha \quad (p\text{-value approach}) \\
\Leftrightarrow &z \geq z_\alpha \quad \text{(critical region approach)} \\
\Leftrightarrow &\mu_0 \notin \left(\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}}, \infty\right) \quad \text{(confidence interval approach)}
\end{aligned}$$

Case 1: σ^2 is known

For the one-sided test $H_0 : \mu = \mu_0$, $H_1 : \mu < \mu_0$

- p-value is
$$p = \Pr(T < z; H_0) = \Pr(Z < z),$$
- Reject H_0 at significance level α when,

$$\begin{aligned}
&p\text{-value} = \Pr(Z < z) \leq \alpha \\
\Leftrightarrow & z \leq -z_\alpha \\
\Leftrightarrow & \mu_0 \notin \left(-\infty, \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}\right)
\end{aligned}$$

Case 1: σ^2 is known

For two-sided test $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$

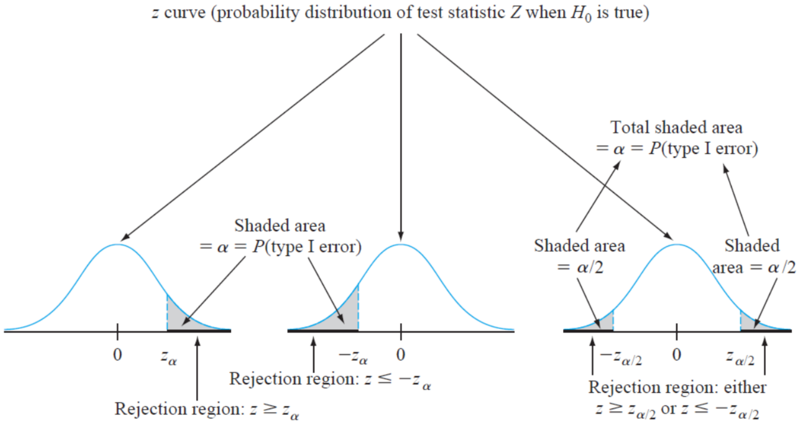
- p-value is

$$
\begin{aligned}
p &= \Pr(|T| > |z|; H_0) \\
&= \Pr(Z > |z|) + \Pr(Z < -|z|) \\
&= 2\Pr(Z > |z|).
\end{aligned}
$$

- Reject H_0 at significance level α when,

$$
\begin{aligned}
&p\text{-value} = 2\Pr(Z > |z|) \leq \alpha \\
\Leftrightarrow &|z| \geq z_{\alpha/2} \\
\Leftrightarrow &\mu_0 \notin \left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)
\end{aligned}
$$

Case 1: σ^2 is known



Case 2: σ^2 is unknown

- Essentially, we replace the standard normal distribution by t-distribution with df$= (n-1)$ and σ^2 by S^2 in Case 1.
- We consider the test statistic

$$T := \frac{\bar{X} - \mu_0}{S/\sqrt{n}}.$$

- Under $H_0 : \mu = \mu_0$, $T \sim t(n-1)$, t distribution with $n-1$ degrees of freedom.
- Once we observe the data $x_1, \ldots, x_n$, we can compute

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}.$$

Case 2: σ^2 is unknown

For the one-sided test $H_0 : \mu = \mu_0$, $H_1 : \mu > \mu_0$,

- p-value is

$$p = \Pr(T > t; H_0) = \Pr(T > t; T \sim t(n-1)).$$

- Reject H_0 at significance level α when,

$$\begin{aligned} & p\text{-value} = \Pr(T > t; T \sim t(n-1)) \leq \alpha \\ \Leftrightarrow & t \geq t_\alpha(n-1) \\ \Leftrightarrow & \mu_0 \notin \left(\bar{x} - t_\alpha(n-1)\frac{s}{\sqrt{n}}, \infty\right) \end{aligned}$$

Case 2: σ^2 is unknown

For the one-sided test $H_0 : \mu = \mu_0$, $H_1 : \mu < \mu_0$,

- p-value is

$$p = \Pr(T < t; H_0) = \Pr(T < t; T \sim t(n-1)).$$

- Reject H_0 at significance level α when,

$$\begin{aligned} & p\text{-value} = \Pr(T < t; T \sim t(n-1)) \leq \alpha \\ \Leftrightarrow & t \leq -t_\alpha(n-1) \\ \Leftrightarrow & \mu_0 \notin \left(-\infty, \bar{x} + t_\alpha(n-1)\frac{s}{\sqrt{n}}\right) \end{aligned}$$

Case 2: σ^2 is unknown:

For the one-sided test $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$,

- p-value is

$$
\begin{aligned}
p &= \Pr(|T| > |t|; H_0) \\
&= \Pr(T > |t|; T \sim t(n-1)) + \Pr(T < -|t|; T \sim t(n-1)) \\
&= 2\Pr(T > |t|; T \sim t(n-1)).
\end{aligned}
$$

- Reject H_0 at significance level α when,

$$
\begin{aligned}
&p\text{-value} = 2\Pr(T > |t|; T \sim t(n-1)) \leq \alpha \\
\Leftrightarrow &|t| \geq t_{\alpha/2}(n-1) \\
\Leftrightarrow &\mu_0 \notin \left(\bar{x} - t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}\right)
\end{aligned}
$$

Example of two-sided test

Example

Let X (in millimeters) equal the growth in 15 days of a tumor induced in a mouse. Assume that the distribution of X is $N(\mu, \sigma^2)$. We shall test the null hypothesis $H_0 : \mu = 4.0$ against the two-sided alternative hypothesis $H_1 : \mu \neq 4.0$. We observe $n = 9$ observations with $\bar{x} = 4.3, s = 1.2$ and set a significance level of $\alpha = 0.10$.

We have

$$t = \frac{4.3 - 4}{1.2/\sqrt{9}} = 0.75$$

and $t_{0.05}(8) = 1.86$

Example of two-sided test

- *p*-**value approach:**

$$p = 2\Pr(T > 0.75;\, T \sim t(8)) = 0.4747 > 0.1,$$

so we fail to reject H_0 at 10% significance level.

- **Critical region approach:**

$$|t| = 0.75 < 1.86,$$

so we fail to reject H_0 at 10% significance level.

- **Confidence interval approach:** The 90% two-sided CI is

$$4.3 \pm 1.86\frac{1.2}{\sqrt{9}} = [3.556, 5.044],$$

which includes 4, so we fail to reject H_0 at 10% significance level.

Example of one-sided test

Example

Suppose that we have $n = 25$, $\bar{x} = 308.8$, $s = 115.15$. We would like to test $H_0 : \mu = 500$ against $H_1 : \mu < 500$. We take the significance level $\alpha = 0.01 = 1\%$.

Then we have

$$t = \frac{308.8 - 500}{115.15/\sqrt{25}} = -8.30$$

and $t_{0.01}(24) = 2.492$

Example of one-sided test

- ***p*-value approach:**

$$p = \Pr(T < -8.30; T \sim t(24)) = 8.17 \times 10^{-9} < 0.01,$$

so we reject H_0 at 1% significance level.

- **Critical region approach:**

$$t = -8.30 < -2.492,$$

so we reject H_0 at 1% significance level.

- **Confidence interval approach:** The 99% one-sided CI is

$$(-\infty, 308.8 + 2.492 \times 115.15/\sqrt{25}] = (-\infty, 366.191],$$

which does not include 500, so we reject H_0 at 1% significance level.

Summary

Table 1: Tests of hypotheses for one mean

H_0	H_1	p-**value**	**Critical Region**
σ^2 **known,** $Z \sim N(0,1), z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$			
$\mu = \mu_0$	$\mu > \mu_0$	$\Pr(Z > z)$	$z \geq z_\alpha$
$\mu = \mu_0$	$\mu < \mu_0$	$\Pr(Z < z)$	$z \leq -z_\alpha$
$\mu = \mu_0$	$\mu \neq \mu_0$	$2\Pr(Z > \lvert z \rvert)$	$\lvert z \rvert \geq z_{\alpha/2}$
σ^2 **unknown,** $T \sim t(n-1), t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$			
$\mu = \mu_0$	$\mu > \mu_0$	$\Pr(T > t)$	$t \geq t_\alpha(n-1)$
$\mu = \mu_0$	$\mu < \mu_0$	$\Pr(T < t)$	$t \leq -t_\alpha(n-1)$
$\mu = \mu_0$	$\mu \neq \mu_0$	$2\Pr(T > \lvert t \rvert)$	$\lvert t \rvert \geq t_{\alpha/2}(n-1)$

- p-value approach: Reject H_0 if p-value $\leq \alpha$
- Critical region approach: Reject H_0 if z or t belongs to the critical region
- Confidence interval approach: Reject H_0 if CI excludes μ_0