Fundamentals of Hypothesis Testing: One-Sample Tests

Prof. dr. Siswanto Agus Wilopo, M.Sc., Sc.D.
Department of Biostatics, Epidemiology and Population Health
Faculty of Medicine
Universitas Gadjah Mada
Learning Objectives

In this lecture, you learn:

• The basic principles of hypothesis testing

• How to use hypothesis testing to test a mean or proportion

• The assumptions of each hypothesis-testing procedure, how to evaluate them, and the consequences if they are seriously violated

• How to avoid the pitfalls involved in hypothesis testing

• The ethical issues involved in hypothesis testing
What is a Hypothesis?

- A hypothesis is a claim (assumption) about a population parameter:
  - population mean
    
    Example: The mean of systolic blood pressure among adults of this city is \( \mu = 120 \) mmHg
  
  - population proportion
    
    Example: The proportion of adults in this city with hypertension is \( \pi = 0.118 \)
The Null Hypothesis, $H_0$

• States the claim or assertion to be tested

Example: The average number of children in NTT woman is equal to three ($H_0 : \mu = 3$)

• Is always about a population parameter, not about a sample statistic

$H_0 : \mu = 3$

$H_0 : \bar{X} = 3$
The Null Hypothesis, $H_0$

(continued)

- Begin with the assumption that the null hypothesis is true
  - Similar to the notion of innocent until proven guilty
- Refers to the status quo
- Always contains “=” , “≤” or “≥” sign
- May or may not be rejected
The Alternative Hypothesis, $H_1$

- Is the opposite of the null hypothesis
  - e.g., The average number of children in NTT woman is not equal to 3 ($H_1: \mu \neq 3$)
- Challenges the status quo
- Contains the “=”, “≤” or “≥” sign
- May or may not be proven
- Is generally the hypothesis that the researcher is trying to prove
Claim: the population mean age is 50. (Null Hypothesis $H_0: \mu = 50$)

Is $\bar{X} = 20$ likely if $\mu = 50$?

If not likely, REJECT Null Hypothesis

Suppose the sample mean age is 20: $\bar{X} = 20$
Reason for Rejecting $H_0$

If it is unlikely that we would get a sample mean of this value...

... if in fact this were the population mean...

... then we reject the null hypothesis that $\mu = 50$. 

Sampling Distribution of $\bar{X}$

$\bar{X}$

If $H_0$ is true

$\mu = 50$

20
Level of Significance, $\alpha$

• Defines the unlikely values of the sample statistic if the null hypothesis is true
  – Defines rejection region of the sampling distribution

• Is designated by $\alpha$, (level of significance)
  – Typical values are 0.01, 0.05, or 0.10

• Is selected by the researcher at the beginning

• Provides the critical value(s) of the test
Level of Significance and the Rejection Region

Level of significance = $\alpha$

<table>
<thead>
<tr>
<th>Null Hypothesis</th>
<th>Alternative Hypothesis</th>
<th>Test Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0: \mu = 3$</td>
<td>$H_1: \mu \neq 3$</td>
<td>Two-tail test</td>
</tr>
<tr>
<td>$H_0: \mu \leq 3$</td>
<td>$H_1: \mu &gt; 3$</td>
<td>Upper-tail test</td>
</tr>
<tr>
<td>$H_0: \mu \geq 3$</td>
<td>$H_1: \mu &lt; 3$</td>
<td>Lower-tail test</td>
</tr>
</tbody>
</table>

$\alpha$ Represents critical value

Rejection region is shaded
Errors in Making Decisions

• **Type I Error**
  – Reject a true null hypothesis
  – Considered a serious type of error

The probability of Type I Error is $\alpha$

• Called *level of significance* of the test
• Set by the researcher in advance
Errors in Making Decisions

• **Type II Error**
  – Fail to reject a false null hypothesis

The probability of Type II Error is $\beta$
### Outcomes and Probabilities

#### Possible Hypothesis Test Outcomes

<table>
<thead>
<tr>
<th>Decision</th>
<th>Actual situation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H₀ True</td>
</tr>
<tr>
<td>Do Not Reject</td>
<td>No error</td>
</tr>
<tr>
<td>H₀</td>
<td>(1 - α)</td>
</tr>
<tr>
<td>Reject H₀</td>
<td>Type I Error</td>
</tr>
<tr>
<td></td>
<td>(α)</td>
</tr>
</tbody>
</table>

**Key:**

- **Outcome (Probability)**
  - No Error: (1 - β)
  - Type I Error: (α)
  - Type II Error: (β)

---

9/22/2017

Biostatistics I: 2017-18
Type I & II Error Relationship

- Type I and Type II errors cannot happen at the same time
  - Type I error can only occur if $H_0$ is true
  - Type II error can only occur if $H_0$ is false

If Type I error probability ($\alpha$) \uparrow, then
Type II error probability ($\beta$) \downarrow
Factors Affecting Type II Error

• All else equal,
  - $\beta$ when the difference between hypothesized parameter and its true value
  - $\beta$ when $\alpha$
  - $\beta$ when $\sigma$
  - $\beta$ when $n$
Hypothesis Tests for the Mean

- \( \sigma \) Known (Z test)
- \( \sigma \) Unknown (t test)
Z Test of Hypothesis for the Mean (\(\sigma\) Known)

- Convert sample statistic (\(\bar{X}\)) to a Z test statistic

\[
Z = \frac{\bar{X} - \mu}{\sigma \sqrt{n}}
\]

The test statistic is:

Hypothesis Tests for \(\mu\)

- \(\sigma\) Known (Z test)
- \(\sigma\) Unknown (t test)
Critical Value Approach to Testing

• For a two-tail test for the mean, $\sigma$ known:

• Convert sample statistic ($\bar{X}$) to test statistic ($Z_{statistic}$)

• Determine the critical $Z$ values for a specified level of significance $\alpha$ from a table or computer

• Decision Rule: If the test statistic falls in the rejection region, reject $H_0$; otherwise do not reject $H_0$
Two-Tail Tests

- There are two cutoff values (critical values), defining the regions of rejection.

\[ H_0: \mu = 3 \]
\[ H_1: \mu \neq 3 \]
6 Steps in Hypothesis Testing

1. State the null hypothesis, $H_0$ and the alternative hypothesis, $H_1$

2. Choose the level of significance, $\alpha$, and the sample size, $n$

3. Determine the appropriate test statistic and sampling distribution

4. Determine the critical values that divide the rejection and nonrejection regions
6 Steps in Hypothesis Testing

(continued)

5. Collect data and compute the value of the test statistic

6. Make the statistical decision and state the managerial conclusion.

- If the test statistic falls into the non-rejection region, do not reject the null hypothesis $H_0$.
- If the test statistic falls into the rejection region, reject the null hypothesis.
- Express the managerial conclusion in the context of the problem.
Hypothesis Testing Example

Test the claim that the true mean of children in NTT woman is \( \neq \) equal to 3. (Assume \( \sigma = 0.8 \))

1. State the appropriate null and alternative hypotheses
   - \( H_0: \mu = 3 \quad H_1: \mu \neq 3 \) (This is a two-tail test)

2. Specify the desired level of significance and the sample size
   - Suppose that \( \alpha = 0.05 \) and \( n = 100 \) are chosen for this test
Hypothesis Testing Example

3. Determine the appropriate technique
   - \( \sigma \) is known so this is a Z test.

4. Determine the critical values
   - For \( \alpha = 0.05 \) the critical Z values are \( \pm 1.96 \)

5. Collect the data and compute the test statistic
   - Suppose the sample results are
     \( n = 100, \ \bar{X} = 2.84 \) (\( \sigma = 0.8 \) is assumed known)

So the test statistic is:

\[
Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{2.84 - 3}{0.8/\sqrt{100}} = \frac{-0.16}{0.08} = -2.0
\]
Hypothesis Testing Example

6. Is the test statistic in the rejection region?

\[ \alpha = 0.05/2 \]

Reject \( H_0 \) if \( Z < -1.96 \) or \( Z > 1.96 \); otherwise do not reject \( H_0 \).

Here, \( Z = -2.0 \) < -1.96, so the test statistic is in the rejection region.
Hypothesis Testing Example

6(continued). Reach a decision and interpret the result

Since $Z = -2.0 < -1.96$, we **reject the null hypothesis** and conclude that there is sufficient evidence that the mean number of children of NTT woman is not equal to 3.
p-Value Approach to Testing

• p-value: Probability of obtaining a test statistic more extreme (≤ or ≥) than the observed sample value given H₀ is true
  – Also called observed level of significance
  – Smallest value of α for which H₀ can be rejected
p-Value Approach to Testing

- Convert Sample Statistic (e.g., $\bar{X}$) to Test Statistic (e.g., Z statistic)

- Obtain the p-value from a table or computer

- Compare the p-value with $\alpha$
  - If p-value < $\alpha$, reject $H_0$
  - If p-value $\geq \alpha$, do not reject $H_0$
**p-Value Example**

• **Example:** How likely is it to see a sample mean of 2.84 (or something further from the mean, in either direction) if the true mean is $\mu = 3.0$?

$X = 2.84$ is translated to a Z score of $Z = -2.0$

$P(Z < -2.0) = 0.0228$

$P(Z > 2.0) = 0.0228$

$p$-value

$= 0.0228 + 0.0228 = 0.0456$
p-Value Example

- Compare the p-value with $\alpha$
  - If $p$-value < $\alpha$, reject $H_0$
  - If $p$-value $\geq \alpha$, do not reject $H_0$

Here: $p$-value = 0.0456
$\alpha = 0.05$

Since 0.0456 < 0.05, we reject the null hypothesis.
Connection to Confidence Intervals

- For $\bar{X} = 2.84$, $\sigma = 0.8$ and $n = 100$, the 95% confidence interval is:

$$2.6832 \leq \mu \leq 2.9968$$

- Since this interval does not contain the hypothesized mean (3.0), we reject the null hypothesis at $\alpha = 0.05$.
One-Tail Tests

• In many cases, the alternative hypothesis focuses on a particular direction

\[ H_0: \mu \geq 3 \]
\[ H_1: \mu < 3 \]

This is a lower-tail test since the alternative hypothesis is focused on the lower tail below the mean of 3

\[ H_0: \mu \leq 3 \]
\[ H_1: \mu > 3 \]

This is an upper-tail test since the alternative hypothesis is focused on the upper tail above the mean of 3
Lower-Tail Tests

- There is only one critical value, since the rejection area is in only one tail

\[ H_0: \mu \geq 3 \]
\[ H_1: \mu < 3 \]

- Reject \( H_0 \)
- Do not reject \( H_0 \)

Critical value
Upper-Tail Tests

- There is only one critical value, since the rejection area is in only one tail

\[ H_0: \mu \leq 3 \]
\[ H_1: \mu > 3 \]

There is only one critical value, since the rejection area is in only one tail.
Example: Upper-Tail Z Test for Mean ($\sigma$ Known)

An investigator thinks that annual health expenditure have increased, and now average over $52 per year. The investigator wishes to test this claim. (Assume $\sigma = 10$ is known)

Form hypothesis test:

$H_0: \mu \leq 52$  the average is not over $52 per month

$H_1: \mu > 52$  the average is greater than $52 per year

(i.e., sufficient evidence exists to support the investigator’s claim)
Example: Find Rejection Region

(continued)

- Suppose that $\alpha = 0.10$ is chosen for this test

Find the rejection region:

- Reject $H_0$ if $Z > 1.28$
- Do not reject $H_0$ if $Z \leq 1.28$

$\alpha = 0.10$
Review: One-Tail Critical Value

What is Z given $\alpha = 0.10$?

Critical Value = 1.28

Standardized Normal Distribution Table (Portion)

<table>
<thead>
<tr>
<th>Z</th>
<th>0.07</th>
<th>0.08</th>
<th>0.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>.8790</td>
<td>.8810</td>
<td>.8830</td>
</tr>
<tr>
<td>1.2</td>
<td>.8980</td>
<td>.8997</td>
<td>.9015</td>
</tr>
<tr>
<td>1.3</td>
<td>.9147</td>
<td>.9162</td>
<td>.9177</td>
</tr>
</tbody>
</table>
Example: Test Statistic

Obtain sample and compute the test statistic

Suppose a sample is taken with the following results: \( n = 64, \bar{X} = 53.1 \) \((\sigma=10 \text{ was assumed known})\)

– Then the test statistic is:

\[
Z = \frac{\bar{X} - \mu}{\sigma} = \frac{53.1 - 52}{10} = \frac{0.88}{\sqrt{64}} = 0.88
\]
Example: Decision

Reach a decision and interpret the result:

\[ Z = 0.88 \]

Do not reject \( H_0 \) since \( Z = 0.88 \leq 1.28 \)

i.e.: there is not sufficient evidence that the mean bill is over $52
Calculate the p-value and compare to $\alpha$ (assuming that $\mu = 52.0$)

$P(\bar{X} \geq 53.1) = P\left(Z \geq \frac{53.1 - 52.0}{10/\sqrt{64}}\right)
= P(Z \geq 0.88) = 1 - 0.8106
= 0.1894$

Do not reject $H_0$ since p-value = 0.1894 > $\alpha = 0.10$
t Test of Hypothesis for the Mean (σ Unknown)

- Convert sample statistic (\( \bar{X} \)) to a t test statistic

Hypothesis Tests for \( \mu \)

\( \sigma \) Known (Z test)

\( \sigma \) Unknown (t test)

The test statistic is:

\[
t_{n-1} = \frac{\bar{X} - \mu}{S / \sqrt{n}}
\]
Example: Two-Tail Test
\( (\sigma \text{ Unknown}) \)

The average height of male in Indonesia is said to be 168 cm. A random sample of 25 male resulted in
\[
\bar{X} = 172.50 \text{ cm and } S = 15.40 \text{ cm.}
\]
Test at the \( \alpha = 0.05 \) level.

(Assume the population distribution is normal)

\[
H_0: \mu = 168 \\
H_1: \mu \neq 168
\]
**Example Solution: Two-Tail Test**

\[ H_0: \mu = 168 \]
\[ H_1: \mu \neq 168 \]

- \( \alpha = 0.05 \)
- \( n = 25 \)
- \( \sigma \) is unknown, so use a t statistic
- Critical Value:

\[ t_{n-1} = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{172.50 - 168}{15.40/\sqrt{25}} = 1.46 \]

Do not reject \( H_0 \): not sufficient evidence that true mean height is different than 168 cm
Connection to Confidence Intervals

- For $\bar{X} = 172.5$, $S = 15.40$ and $n = 25$, the 95% confidence interval is:

$$172.5 - (2.0639) \frac{15.4}{\sqrt{25}} \quad \text{to} \quad 172.5 + (2.0639) \frac{15.4}{\sqrt{25}}$$

$$166.14 \leq \mu \leq 178.86$$

- Since this interval contains the Hypothesized mean (168), we do not reject the null hypothesis at $\alpha = 0.05$
Hypothesis Tests for Proportions

• Involves categorical variables

• Two possible outcomes
  – “Success” (possesses a certain characteristic)
  – “Failure” (does not possesses that characteristic)

• Fraction or proportion of the population in the “success” category is denoted by $\pi$
Proportions

• Sample proportion in the success category is denoted by \( p \)

\[
p = \frac{X}{n} = \frac{\text{number of successes in sample}}{\text{sample size}}
\]

• When both \( n\pi \) and \( n(1-\pi) \) are at least 5, \( p \)
can be approximated by a normal distribution with mean and standard deviation

\[
\mu_p = \pi \\
\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}
\]
Hypothesis Tests for Proportions

- The sampling distribution of $p$ is approximately normal, so the test statistic is a $Z$ value:

$$Z = \frac{p - \pi}{\sqrt{\pi(1-\pi)n}}$$

- Hypothesis Tests for $p$
  - $n\pi \geq 5$ and $n(1-\pi) \geq 5$
  - $n\pi < 5$ or $n(1-\pi) < 5$
  - Not discussed in this chapter
Z Test for Proportion in Terms of Number of Successes

- An equivalent form to the last slide, but in terms of the number of successes, $X$:

$$Z = \frac{X - n\pi}{\sqrt{n\pi(1-\pi)}}$$

Hypothesis Tests for $X$

- $X \geq 5$ and $n-X \geq 5$
- $X < 5$ or $n-X < 5$

Not discussed in this chapter
Example: Z Test for Proportion

A researcher claims that it receives 8% responses from its mailing. To test this claim, a random sample of 500 were surveyed with 25 responses. Test at the $\alpha = 0.05$ significance level.

Check:

$n\pi = (500)(.08) = 40$
$n(1-\pi) = (500)(.92) = 460$
Z Test for Proportion: Solution

$H_0: \pi = 0.08$

$H_1: \pi \neq 0.08$

$\alpha = 0.05$

$n = 500, \ p = 0.05$

Test Statistic:

$$Z = \frac{p - \pi}{\sqrt{\pi(1-\pi)/n}} = \frac{0.05 - 0.08}{\sqrt{0.08(1-0.08)/500}} = -2.47$$

Decision:

Reject $H_0$ at $\alpha = 0.05$

Conclusion:

There is sufficient evidence to reject the researcher claim of 8% response rate.
p-Value Solution

Calculate the p-value and compare to $\alpha$
(For a two-tail test the p-value is always two-tail)

Do not reject $H_0$

$\alpha/2 = .025$

0.0068

-1.96

$Z = -2.47$

Reject $H_0$

$p$-value = 0.0136:

\[ P(Z \leq -2.47) + P(Z \geq 2.47) = 2(0.0068) = 0.0136 \]

Reject $H_0$ since $p$-value = 0.0136 < $\alpha = 0.05$
Potential Pitfalls and Ethical Considerations

- Use randomly collected data to reduce selection biases
- Do not use human subjects without informed consent
- Choose the level of significance, $\alpha$, and the type of test (one-tail or two-tail) before data collection
- Do not employ “data snooping” to choose between one-tail and two-tail test, or to determine the level of significance
- Do not practice “data cleansing” to hide observations that do not support a stated hypothesis
- Report all pertinent findings
Chapter Summary

• Addressed hypothesis testing methodology

• Performed Z Test for the mean (\(\sigma\) known)

• Discussed critical value and p–value approaches to hypothesis testing

• Performed one-tail and two-tail tests
Chapter Summary

(continued)

- Performed t test for the mean ($\sigma$ unknown)
- Performed Z test for the proportion
- Discussed pitfalls and ethical issues