STAT 423/523 Statistical Methods for Engineers and Scientists

Lecture 4: Hypothesis Testing I

Chencheng Cai

Washington State University

Hypothesis Testing

A **statistical hypothesis** is a statement or assumption about one or more population parameters.

- The average height of WSU students is 68 inches.
- The proportion of defective items produced by a machine is greater than 0.10.
- The average commute time of WSU students is the same as the UI students.

A **hypothesis-testing procedure** usually involves two contradictory hypotheses.

- The average height of WSU students is 68 inches.
- v.s. The average height of WSU students is NOT 68 inches.

Hypothesis Testing

The **null hypothesis** H_0 is the claim that is initially assumed to be true.
The **alternative hypothesis** H_a is the claim that we are trying to find evidence for.

Example:
Let μ be the average height of WSU students.
To test whether μ is exactly 68 inches:

- H_0: $\mu = 68$ inches.
- H_a: $\mu \neq 68$ inches.

To test whether μ is greater than or equal to 68 inches:

- H_0: $\mu \geq 68$ inches.
- H_a: $\mu < 68$ inches.

Hypothesis Testing

A **test of hypothesis** is a procedure for deciding between two contradictory claims.

Two possible outcomes of a hypothesis test:

- **Reject** H_0: There is enough evidence against H_0 to support H_a.
- **Fail to reject** H_0: There is not enough evidence to support H_a.

Note: We never "accept" the null hypothesis. We either reject it or fail to reject it.

The decision is based on the sample data. (So the decision is a random variable.)

Hypothesis Testing

	H_0 is true	H_0 is false
Reject H_0	Type I error	Correct decision
Fail to reject H_0	Correct decision	Type II error

- A **Type I error** occurs when we reject H_0 when H_0 is true.
- A **Type II error** occurs when we fail to reject H_0 when H_0 is false.

- Usually, Type I error and Type II error are inversely related.
- Type I error can be controlled by the **significance level** α.

$$P(\text{reject} \mid H_0) = \alpha$$

- α is also called the **size** of the test.

Example

Suppose the heights of WSU students are normally distributed with mean μ and a standard deviation of 2 inches.
Let $X_1, \ldots, X_{100}$ be a random sample from the population.
We want to test whether the average height of WSU students is **greater than** 68 inches.

Step 1: Set up the null and alternative hypotheses.

- H_0: $\mu = 68$ inches.
- H_a: $\mu > 68$ inches.

Step 2: Formulate a decision based on the sample.

$$\text{reject null if } \bar{X} > 68.6$$

Step 3: Collect the sample data and make a decision.

The significance level of this test is

$$\alpha = P(\text{reject}|H_0) = P(N(68, 0.04) > 68.5) = 1 - \Phi(3).$$

Example

Now consider another test: whether the average height of WSU students is **exactly** 68 inches.

Step 1: Set up the null and alternative hypotheses.

- H_0: $\mu = 68$ inches.
- H_a: $\mu \neq 68$ inches.

Step 2: Formulate a decision based on the sample.

$$\text{reject null if } \bar{X} > 68.4 \text{ or } \bar{X} < 67.6$$

Step 3: Collect the sample data and make a decision.

The significance level of this test is

$$\alpha = P(\text{reject}|H_0) = P(N(68, 0.04) > 68.4 \text{ or } N(68, 0.04) < 67.6) = 0.05.$$

Test Statistics

In our previous sample, our decision is based on the sample mean $\bar{X}$, which is called the test statistic.

A **test statistic** is a function of the sample data whose value is used to make a decision in a hypothesis test.

The set of values of the test statistic for which the null hypothesis is rejected is called the **rejection region**.

Example:

- The rejection region for the test of whether the average height of WSU students is greater than 68 inches is $(68.6, \infty)$.
- The rejection region for the test of whether the average height of WSU students is exactly 68 inches is $(-\infty, 67.6) \cup (68.4, \infty)$.

Example

A company producing Brand D yogurt would like to increase its market share.
The company sends out a survey to 100 Brand C consumers asking whether they would like to switch brand.
Let p be the proportion of Brand C consumers who would like to switch to Brand D.
Step 1: Set up the null and alternative hypotheses.

- H_0: $p = 0.5$.
- H_a: $p < 0.5$.

Step 2: Formulate a decision based on the sample. Let X be the number of consumers who would like to switch. Rejection region:

$$X \in [0, 37]$$

Step 3: Collect the sample data and make a decision.

The significance level of this test is

$$\alpha = P(\text{reject}|H_0) = P(\text{Binom}(100, 0.5) \leq 37) = 0.006$$

Design Rejection Region

Let $X_1, \ldots, X_n$ be a random sample from the normal population with unknown mean μ and known variance σ^2.
Consider a hypothesis testing:

- H_0: $\mu = 0$.
- H_a: $\mu \neq 0$.

We should reject null when the sample mean is too far away from 0.

Suppose we would like to design a rejection region in the following format:

$$\text{reject null if } \bar{X} \in (-\infty, a) \cup (b, \infty),$$

for some constant $a < b$ to be determined later.

Design Rejection Region

$$\text{reject null if } \bar{X} \in (-\infty, a) \cup (b, \infty),$$

The significance level of this test is

$$\begin{aligned}
\alpha = P(\text{reject}|H_0) &= P(\bar{X} < a) + P(\bar{X} > b) \\
&= P\left(N\left(0, \frac{\sigma^2}{n}\right) < a\right) + P\left(N\left(0, \frac{\sigma^2}{n}\right) > b\right) \\
&= P\left(N(0,1) < \frac{a\sqrt{n}}{\sigma}\right) + P\left(N(0,1) > \frac{b\sqrt{n}}{\sigma}\right) \quad (*)
\end{aligned}$$

For a given significance level α, we can find a and b that satisfy $(*)$.

- Choice I: $a = -z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ and $b = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.
- Choice II: $a = -z_{\alpha}\frac{\sigma}{\sqrt{n}}$ and $b = +\infty$.
- Choice III: $a = -\infty$ and $b = z_{\alpha}\frac{\sigma}{\sqrt{n}}$.

Design Rejection Region

- Choice I: reject null if $\bar{X} < -z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ or $\bar{X} > z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.
- Choice II: reject null if $\bar{X} < -z_{\alpha}\frac{\sigma}{\sqrt{n}}$.
- Choice III: reject null if $\bar{X} > z_{\alpha}\frac{\sigma}{\sqrt{n}}$.

Properties:

- All three choices have the same significance level α.
- Choice I is a **two-tailed** or **two-sided** test.
- Choice II is a **lower-tailed** or **left-sided** test.
- Choice III is a **upper-tailed** or **right-sided** test.
- The three tests have different **power**s on the alternative hypothesis.
- **Power** of a test is the probability of rejecting the null hypothesis when the alternative hypothesis is true.

Two-sided Z-test

Suppose $X_1, \ldots, X_n$ is a random sample from the normal population with unknown mean μ and known variance σ^2.

Consider a hypothesis testing:

- H_0: $\mu = \mu_0$.
- H_a: $\mu \neq \mu_0$.

The level-α **two-sided Z-test** rejects null hypothesis if

$$\bar{X} < \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \text{ or } \bar{X} > \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Example

Suppose the height of WSU students is normally distributed with a standard deviation of 2 inches.
A random sample of 100 students has an average height of 68.5 inches.
We want to test whether the average height of WSU students is 68 inches.

Step 1: Set up the null and alternative hypotheses.

- H_0: $\mu = 68$ inches.
- H_a: $\mu \neq 68$ inches.

Step 2: Use level-0.05 two-sided Z-test: reject null if

$$\bar{X} < 68 - 1.96\frac{2}{\sqrt{100}} = 67.608 \text{ or } \bar{X} > 68 + 1.96\frac{2}{\sqrt{100}} = 68.395.$$

Step 3: the observed sample mean is $68.5 > 68.395$, so we reject the null.
Step 4: Conclusion:
The average height of WSU students is different from 68 inches with a significance level of 0.05.

Relation to Confidence Interval

The rejection condition for a two-sided test is

$$\begin{aligned} &\bar{X} \in \left(-\infty, \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \cup \left(\mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \infty\right) \\ \iff &\bar{X} - \mu_0 \in \left(-\infty, -z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \cup \left(z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \infty\right) \\ \iff &\mu_0 - \bar{X} \in \left(-\infty, -z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \cup \left(z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \infty\right) \\ \iff &\mu_0 \in \left(-\infty, \bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \cup \left(\bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \infty\right) \\ \iff &\mu_0 \notin \left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \end{aligned}$$

The right hand side is the $1-\alpha$ (normal) confidence interval for μ (ignoring the boundaries)!

Relation to Confidence Interval

For hypothesis testing of normal population mean with known variance:

- H_0: $\mu = 0$.
- H_a: $\mu \neq 0$.

The level-α two-sided Z-test rejects null hypothesis if μ_0 is not in the $1-\alpha$ (normal) confidence interval for μ, that is, if

$$\mu_0 \notin \left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right].$$

Example

Back to the WSU student height example.
The 95% confidence interval for the average height of WSU students is

$$68.5 \pm 1.96 \frac{2}{\sqrt{100}} = (68.395, 68.605).$$

The null hypothesis is $\mu = 68$, which is not in the confidence interval for μ.
Therefore, we reject the null hypothesis with a significance level of 0.05.

P-values

P-value is a single number computed from the sample data that provides a measure of how much evidence we have against the null hypothesis.

Def I: The **p-value** is the probability, computed assuming that the null hypothesis is true, that the test statistic would take a value as extreme as or more extreme than the one actually observed.

Def II: The **p-value** is the smallest significance level at which the null hypothesis can be rejected by the sample data.

P-value for Two-sided Z-test

The rejection rule for the two-sided Z-test:

$$
\begin{aligned}
&\bar{X} \in \left(-\infty, \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \cup \left(\mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \infty\right) \\
\iff &\bar{X} - \mu_0 \in \left(-\infty, -z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \cup \left(z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \infty\right) \\
\iff &\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \in \left(-\infty, -z_{\alpha/2}\right) \cup \left(z_{\alpha/2}, \infty\right) \\
\iff &P\left(N(0,1) > \left|\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}\right|\right) + P\left(N(0,1) < -\left|\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}\right|\right) < \alpha
\end{aligned}
$$

We call $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ the **standardized test statistic** or simply **Z-statistic**.
The rejection region for Z is $(-\infty, -z_{\alpha/2}) \cup (z_{\alpha/2}, \infty)$.

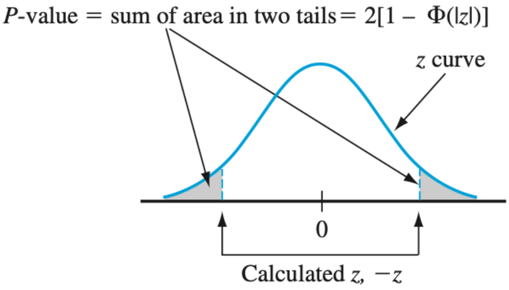
P-value for Two-sided Z-test

The decision can be made by comparing the following quantity with the significance level α:

$$P(N(0,1) > |Z|) + P(N(0,1) < -|Z|) = 2P(N(0,1) > |Z|) = 2(1 - \Phi(|Z|))$$

We call $2(1 - \Phi(|Z|))$ the **two-sided p-value** for the two-sided Z-test.

We reject null if the p-value is less than the significance level α.



Example

Back to the WSU student height example.
The Z-statistic for the null hypothesis $\mu = 68$ is

$$Z = \frac{68.5 - 68}{2/\sqrt{100}} = 2.5.$$

The two-sided p-value for the two-sided Z-test is

$$\text{p-value} = 2(1 - \Phi(2.5)) = 0.0124 < 0.05.$$

Therefore, we reject the null:
The average height of WSU students is different from 68 inches with a significance level of 0.05.

Power of Two-sided Z-test

The **power** of a test is the probability of rejecting the null hypothesis when the true parameter is in the alternative hypothesis.

Consider a normal population with unknown mean μ and known variance σ^2, and the hypothesis:

- H_0: $\mu = \mu_0$.
- H_a: $\mu \neq \mu_0$.

The power at $\mu_1 \in H_a$, denoted by $1 - \beta(\mu_1)$, is

$$P(\text{reject} \mid \mu = \mu_1)$$

Power of Two-sided Z-test

When the truth is $\mu = \mu_1$, the Z-statistic

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N\left(\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}, 1\right),$$

because now $\bar{X} \sim N(\mu_1, \sigma^2/n)$.
The power at μ_1 is

$$\begin{aligned}
1 - \beta(\mu_1) &= P(\text{reject} \mid \mu = \mu_1) = P\left(Z \in (-\infty, -z_{\alpha/2}) \cup (z_{\alpha/2}, \infty)\right) \\
&= P\left(N\left(\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}, 1\right) \in (-\infty, -z_{\alpha/2}) \cup (z_{\alpha/2}, \infty)\right) \\
&= P\left(N(0,1) \in \left(-\infty, -z_{\alpha/2} - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right) \cup \left(z_{\alpha/2} - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}, \infty\right)\right) \\
&= \Phi\left(-z_{\alpha/2} - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right) + \Phi\left(-z_{\alpha/2} + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right)
\end{aligned}$$

Sample Size Determination

It is usually desirable to have a test with high power.
In many cases, we need to determine the sample size n to achieve a desired power.

In order to determine the sample size, we need to know the following quantities:

- The significance level α.
- The desired power $1 - \beta^*$.
- The hypothetical true parameter μ_1. (Or in some cases, the effect size $\mu_1 - \mu_0$.)
- The standard deviation σ. (for Z-test only)

The the minimum sample size n can be determined by

$$n^* = \min\left\{n : \Phi\left(-z_{\alpha/2} - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right) + \Phi\left(-z_{\alpha/2} + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right) \geq 1 - \beta^*\right\}$$

The approximate solution is

$$n^* = \left\lceil \left(\frac{\sigma(z_{\alpha/2} - z_{1-\beta^*})}{\mu_1 - \mu_0}\right)^2 \right\rceil$$

Example

For the WSU student height example, suppose we hypothetically assume the true average height is 68.5
We want to determine the sample size to achieve a power of 0.8.

The solution to the equation:

$$\Phi\left(-z_{\alpha/2} - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right) + \Phi\left(-z_{\alpha/2} + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right) = 1 - \beta^*$$

is $n = 125.585$. So the minimum sample size would be 126.

Summary of Two-sided Z-test

- Hypothesis:

$$H_0: \ \mu = \mu_0 \quad \text{v.s.} \quad H_a: \ \mu \neq \mu_0$$

- Test statistic:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

- Rejection region:

$$\text{reject null if } |Z| > z_{\alpha/2}$$

- P-value:

$$\text{p-value} = 2(1 - \Phi(|Z|))$$

- Power:

$$1 - \beta(\mu_1) = \Phi\left(-z_{\alpha/2} - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right) + \Phi\left(-z_{\alpha/2} + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right)$$

- Sample size determination: (approximation)

$$n^* = \left\lceil \left(\frac{\sigma\left(z_{\alpha/2} - z_{1-\beta^*}\right)}{\mu_1 - \mu_0}\right)^2 \right\rceil$$

One-sided Z-test

For hypothesis testing:

$$H_0: \ \mu = \mu_0 \quad \text{v.s.} \quad H_a: \ \mu > \mu_0$$

We should rejet null only if the sample mean is too large.

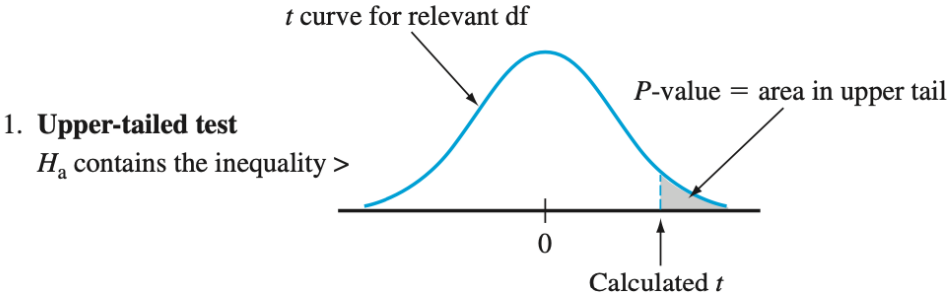
The two-sided Z-test is not suitable for this case. We need to modify the rejection region to

$$\text{reject null if } \bar{X} > \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}.$$

Or equivalently, for the test statistic $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$, the rejection region is

$$\text{reject null if } Z > z_\alpha.$$

One-sided Z-test



P-value for One-sided Z-test

The reject region for Z is

$$\text{reject null if } Z > z_\alpha.$$

The p-value for the one-sided Z-test is the smallest significance level at which the null hypothesis can be rejected by the sample data.

The p-value for the one-sided Z-test is

$$\text{p-value} = 1 - \Phi(Z) = 1 - \Phi\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}\right).$$

Power for One-sided Z-test

Consider a hypothetical true parameter μ_1 in the alternative hypothesis (i.e. $\mu_1 > \mu_0$). The power at μ_1 is

$$
\begin{aligned}
1 - \beta(\mu_1) &= P(\text{reject} \mid \mu = \mu_1) = P(Z > z_\alpha \mid \mu = \mu_1) \\
&= P\left(N\left(\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}, 1\right) > z_\alpha\right) \\
&= P\left(N(0,1) > z_\alpha - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right) \\
&= 1 - \Phi\left(z_\alpha - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right)
\end{aligned}
$$

Sample Size Determination

Suppose we want to determine the sample size n to achieve a desired power $1 - \beta^*$ for the one-sided Z-test at μ_1.

The minimum sample size n can be determined by

$$n^* = \min \left\{ n : 1 - \Phi\left(z_\alpha - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} \right) \geq 1 - \beta^* \right\}$$

The solution is

$$n^* = \left\lceil \left(\frac{\sigma(z_\alpha - z_{1-\beta^*})}{\mu_1 - \mu_0} \right)^2 \right\rceil$$

Summary of One-sided Z-test

- Hypothesis:

$$H_0: \ \mu = \mu_0 \quad \text{v.s.} \quad H_a: \ \mu > \mu_0$$

- Test statistic:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

- Rejection region:

$$\text{reject null if } Z > z_\alpha$$

- P-value:

$$\text{p-value} = 1 - \Phi(Z)$$

- Power:

$$1 - \beta(\mu_1) = 1 - \Phi\left(z_\alpha - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right)$$

- Sample size determination:

$$n^* = \left\lceil \left(\frac{\sigma(z_\alpha - z_{1-\beta^*})}{\mu_1 - \mu_0}\right)^2 \right\rceil$$

Extension of One-sided Z-test

- It also works for the hypothesis testing:

$$H_0 : \ \mu \leq \mu_0 \quad \text{v.s.} \quad H_a : \ \mu > \mu_0$$

- For the hypothesis testing with opposite direction:

$$H_0 : \ \mu = \mu_0 \quad \text{v.s.} \quad H_a : \ \mu < \mu_0$$

  all quantiles and p-values should be reversed:
  - Rejection region: $Z < -z_\alpha$.
  - P-value: $\Phi(Z)$.
  - Power: $\Phi\left(-z_\alpha - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right)$.
  - Sample size determination: $\left\lceil \left(\frac{\sigma(z_\alpha - z_{1-\beta^*})}{\mu_1 - \mu_0}\right)^2 \right\rceil$.