W1211 Introduction to Statistics
Lecture 24

Wei Wang

Nov 28, 2012

What we talked about last lecture

- Confidence Intervals for population mean μ based on t distribution. What is the key assumption for using t distribution?
- Basic Concepts of Hypothesis Testing: the form; null hypothesis and alternative hypothesis.

Hypotheses

- The null hypothesis that we are dealing with in this course will be of the following form:
$$H_0 : \theta = \theta_0$$

Hypotheses

- The null hypothesis that we are dealing with in this course will be of the following form:

$$H_0 : \theta = \theta_0$$

- The alternative hypothesis will possibly take three forms

$$\begin{aligned} H_a :& \theta > \theta_0 \\ H_a :& \theta < \theta_0 \\ H_a :& \theta \neq \theta_0 \end{aligned}$$

Hypotheses

- The null hypothesis that we are dealing with in this course will be of the following form:

$$H_0 : \theta = \theta_0$$

- The alternative hypothesis will possibly take three forms

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

- Null hypotheses and alternative hypotheses are not treated equally. In constructing Testing Procedures, we try to protect the null hypothesis, i.e., setting a more stringent standard for rejecting H_0

Motivating Example

- Suppose we have a coin, we want to test whether it is unbiased or biased in favor of head, $H_0 : p = 0.5$ v.s. $H_a : p > 0.5$. We flip the coin for several times, and record the number of heads.

Motivating Example

- Suppose we have a coin, we want to test whether it is unbiased or biased in favor of head, $H_0 : p = 0.5$ v.s. $H_a : p > 0.5$. We flip the coin for several times, and record the number of heads.
- Intuitively, how should we conduct the test?

Testing Procedures

- A test procedure is specified by the following:
  - Find a test statistic, a function of the sample data on which the decision (reject H_0 or do not reject H_0) is based.
  - Construct a rejection region, the set of all test statistic values for which H_0 will be rejected.
- The null hypothesis will then be rejected if and only if **the observed or computed test statistic value falls in the rejection region**.

Example Cont'd

- Following the aforementintinoed procedures, we can conduct the test by first selecting a test statistic, and then construct a rejection region.
  - The natural test statistic is the sample proportion $\bar{X}$.
  - And we will reject the null hypothesis $p = 0.5$ in favor of the alternative hypothesis $H_a : p > 0.5$ if $\bar{X}$ is too large. So the rejection region will look like $\{\bar{X} > a\}$.
- To determine a, we need to utilize the sampling distribution of the test statistic as well as finer analysis of the errors.

Example Cont'd

- Following the aforementintinoed procedures, we can conduct the test by first selecting a test statistic, and then construct a rejection region.
  - The natural test statistic is the sample proportion $\bar{X}$.
  - And we will reject the null hypothesis $p = 0.5$ in favor of the alternative hypothesis $H_a : p > 0.5$ if $\bar{X}$ is too large. So the rejection region will look like $\{\bar{X} > a\}$.
- To determine a, we need to utilize the sampling distribution of the test statistic as well as finer analysis of the errors.

Example Cont'd

- Go back to the coin flipping example, what are the two type of errors?
  - Type I Error: We reject the null hypothesis when in fact it is true, i.e., the event that
$$\bar{X} > a \text{ when } p = 0.5$$
  - Type II Error: We failed to reject the null hypothesis when the alternative is true, i.e., the event that
$$\bar{X} < a \text{ when } p = p_1 \neq 0.5$$
- We can calculate the value of α and β based on the sampling distribution of $\bar{X}$.

Example 8.2 from the Textbook

- It is known the drying time of a certain type of paint follows a normal distribution with mean 75 min and standard deviation 9 min. A new additive is added to the paint which is believed to lower the mean drying time.

- If we assume the standard deviation stays the same, then the appropriate Hypotheses are $H_0 : \mu = 75$ versus $H_1 : \mu < 75$. If we use the sample mean of 25 test specimens as our test statistic, and $\{\bar{X} < c\}$ with cutoff point $c = 70.8$ as our rejection region.

Example 8.2 Cont'd

- We know the sampling distribution of $\bar{X}$ is $N(\mu, \frac{9}{25} = 1.8^2)$.
- Type I Error

$$\begin{aligned}\alpha &= P(\text{type I error}) = P(H_0 \text{ is rejected when it is true}) \\ &= P(\bar{X} < 70.8 \text{ when } \bar{X} \sim N(75, 1.8^2)) \\ &= P(Z < \frac{70.8 - 75}{1.8}) = 0.01\end{aligned}$$

- Type II Errors for some values of μ

$$\begin{aligned}\beta(72) &= P(\text{type II error when } \mu = 72) \\ &= P(\bar{X} > 70.8 \text{ when } \bar{\sim} N(72, 1.8^2)) \\ &= 1 - P(Z < \frac{70.8 - 72}{1.8}) = 0.7486\end{aligned}$$

$$\beta(70) = 0.33 \qquad \beta(67) = 0.0174$$

Example 8.2 Cont'd

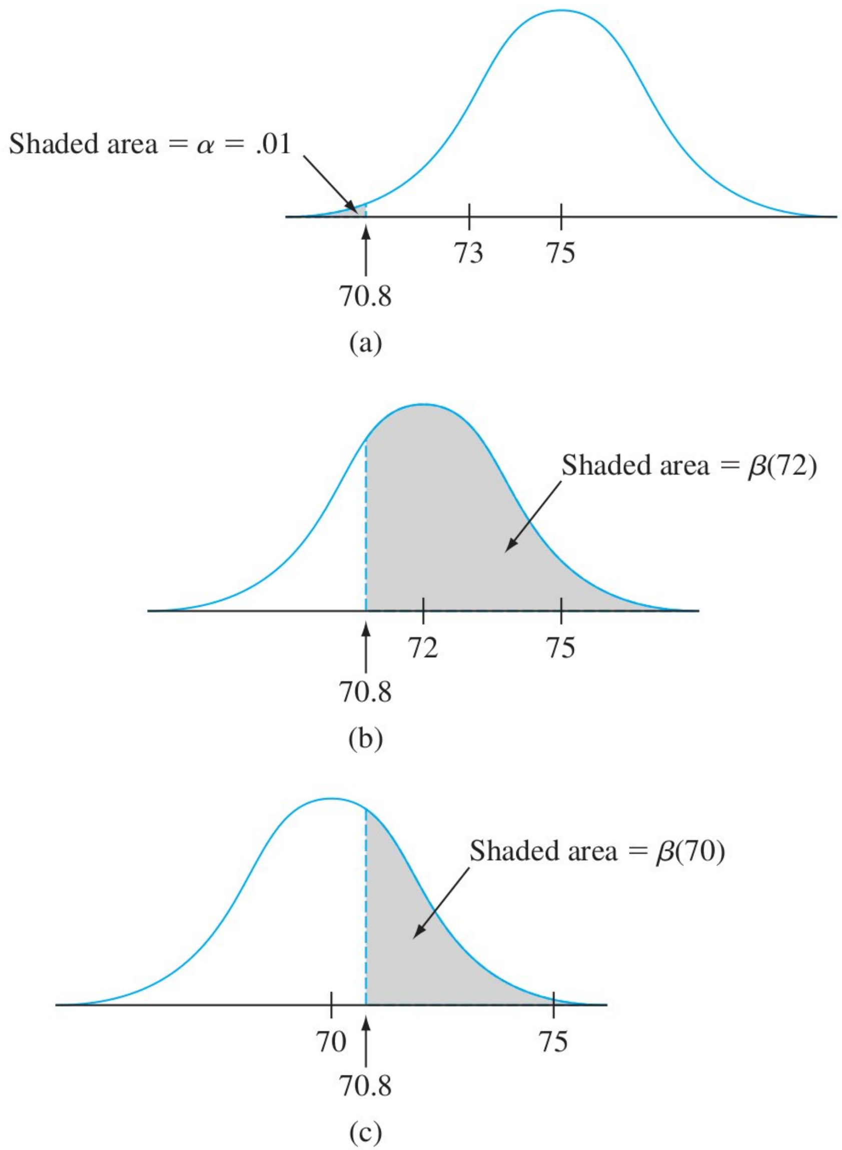


Figure: Illustrations of α and β for the testing procedure:(a) $\mu = 75$; (b) $\mu = 72$; (c) $\mu = 70$.

Example 8.2 Cont'd

- If we change the cutoff point to 72, α and β will change correspondingly
- Type I Error

$$\begin{aligned}\alpha &= P(\text{type I error}) = P(H_0 \text{ is rejected when it is true}) \\ &= P(\bar{X} < 72 \text{ when } \bar{X} \sim N(75, 1.8^2)) \\ &= P(Z < \frac{72-75}{1.8}) = 0.05\end{aligned}$$

- Type II Errors for some values of μ

$$\begin{aligned}\beta(72) &= P(\text{type II error when } \mu = 72) \\ &= P(\bar{X} > 72 \text{ when } \bar{\sim} N(72, 1.8^2)) \\ &= 1 - P(Z < \frac{72-72}{1.8}) = 0.5\end{aligned}$$

$$\beta(70) = 0.1335 \qquad \beta(67) = 0.0027$$

Balancing Two Types of Errors

- A good test will be aimed to make two types of errors, both α and β, as small as possible. But simultaneously minimizing the two is impossible once a test statistic is given, so we need to construct a rejection region that effects a good compromise between α and β.

- Because we try to protect the null hypothesis, the Type I Error is considered more serious than the Type II Error. So minimizing α is more important.

Balancing Two Types of Errors

- A good test will be aimed to make two types of errors, both α and β, as small as possible. But simultaneously minimizing the two is impossible once a test statistic is given, so we need to construct a rejection region that effects a good compromise between α and β.
- Because we try to protect the null hypothesis, the Type I Error is considered more serious than the Type II Error. So minimizing α is more important.
- In practice, people often fix the value of α, typically at levels such as 0.1, 0.05 and 0.01, which is called **significance level** of the test, and then minimize β subject to the constraint of significance level. The corresponding test procedure is called a **level α test**.

Balancing Two Types of Errors

- A good test will be aimed to make two types of errors, both α and β, as small as possible. But simultaneously minimizing the two is impossible once a test statistic is given, so we need to construct a rejection region that effects a good compromise between α and β.

- Because we try to protect the null hypothesis, the Type I Error is considered more serious than the Type II Error. So minimizing α is more important.

- In practice, people often fix the value of α, typically at levels such as 0.1, 0.05 and 0.01, which is called **significance level** of the test, and then minimize β subject to the constraint of significance level. The corresponding test procedure is called a **level α test**.

- In applied statistics, another criterion called **power** is often used. It is simply $1 - \beta$, which means the probability of rejecting null hypothesis when it is false.

Hypothesis Testing for a Population Mean

- In this section, the null hypothesis is about a population mean $H_0 : \mu = \mu_0$ and there are three possible Alternative Hypotheses $H_a : \mu > \mu_0$ or $H_a : \mu < \mu_0$ or $H_a : \mu \neq \mu_0$.
- We will discuss three cases which parallel our discussion about Confidence Interval for a Population Mean.
-
  - Case I: Normal Distribution and Known σ (z Test)
  - Case II: General Distribution, Unknown σ but Large Sample (z Test)
  - Case III: Normal Distribution and Unknown σ (t Test)

Case I: Normal Distribution and Known σ (z Test)

- Under the null hypothesis, the test statistic

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

follow a standard normal distribution.

Case I: Normal Distribution and Known σ (z Test)

- Under the null hypothesis, the test statistic

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

  follow a standard normal distribution.
- If the Alternative Hypothesis is $H_a : \mu > \mu_0$, then the Rejection Region is something like $\{z \geq c\}$, where c is a constant to be determined.

Case I: Normal Distribution and Known σ (z Test)

- Under the null hypothesis, the test statistic
$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$
follow a standard normal distribution.
- If the Alternative Hypothesis is $H_a : \mu > \mu_0$, then the Rejection Region is something like $\{z \geq c\}$, where c is a constant to be determined.
- c is determined by the level of the test α, if we set c as z critical value z_α then
$$\begin{aligned} P(\text{type I error}) &= P(H_0 \text{ is rejected when } H_0 \text{ is true}) \\ &= P(Z > z_\alpha \text{ when } Z \sim N(0,1)) = \alpha \end{aligned}$$

Case I: Normal Distribution and Known σ (z Test)

Null hypothesis: $H_0: \mu = \mu_0$

Test statistic value: $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

Alternative Hypothesis	Rejection Region for Level α Test
$H_a: \mu > \mu_0$	$z \geq z_\alpha$ (upper-tailed test)
$H_a: \mu < \mu_0$	$z \leq -z_\alpha$ (lower-tailed test)
$H_a: \mu \neq \mu_0$	either $z \geq z_{\alpha/2}$ or $z \leq -z_{\alpha/2}$ (two-tailed test)

▶

Case I: Normal Distribution and Known σ (z Test)

- Null hypothesis: $H_0: \mu = \mu_0$

  Test statistic value: $z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$

Alternative Hypothesis	Rejection Region for Level α Test
$H_a: \mu > \mu_0$	$z \geq z_\alpha$ (upper-tailed test)
$H_a: \mu < \mu_0$	$z \leq -z_\alpha$ (lower-tailed test)
$H_a: \mu \neq \mu_0$	either $z \geq z_{\alpha/2}$ or $z \leq -z_{\alpha/2}$ (two-tailed test)

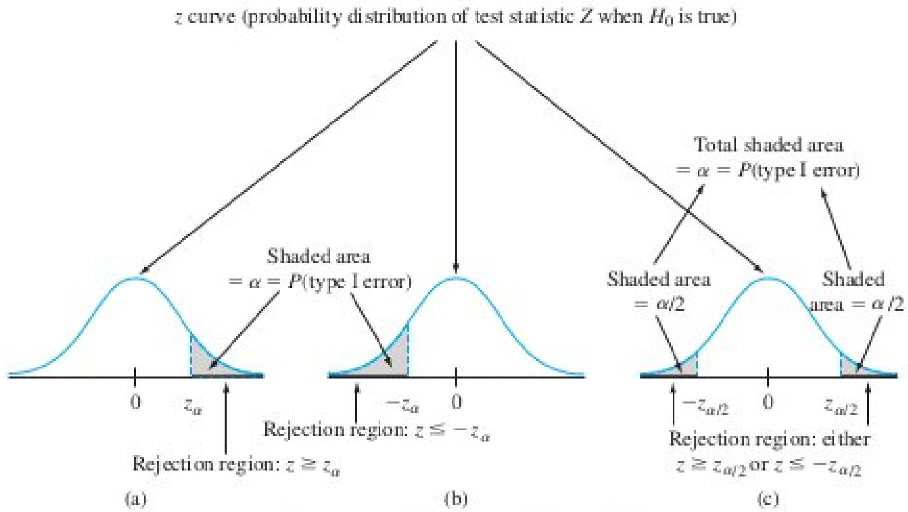
- 

  Figure 8.2 Rejection regions for z tests: (a) upper-tailed test; (b) lower-tailed test; (c) two-tailed test

Case I: Normal Distribution and Known σ (z Test)

- We can also compute Type II Error β and sample size n. Still we consider the upper-tailed test as a demonstration.
- Type II Error β will be a function of any particular number μ' that is larger than the null value μ_0.

$$
\begin{aligned}
\beta(\mu') &= P(Z < z_\alpha \text{ when } \mu = \mu') \\
&= P(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < z_\alpha \text{ when } \mu = \mu') \\
&= P(\frac{\bar{X} - \mu'}{\sigma/\sqrt{n}} < z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}} \text{ when } \mu = \mu') \\
&= \Phi(z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}) \leq 1 - \alpha
\end{aligned}
$$

  $\Phi()$ is the CDF of standard normal.
- What is the power of the test?

Case I: Normal Distribution and Known σ (z Test)

- For a given True Value μ', Type I Error level α and Type II Error β, we can determine the sample size n that we need with

$$\Phi(z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}) = \beta$$

Thus

$$-z_\beta = z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}$$

Case I: Normal Distribution and Known σ (z Test)

Alternative Hypothesis	Type II Error Probability $\beta(\mu')$ for a Level α Test
H_a: $\mu > \mu_0$	$\Phi\left(z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$
H_a: $\mu < \mu_0$	$1 - \Phi\left(-z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$
H_a: $\mu \neq \mu_0$	$\Phi\left(z_{\alpha/2} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right) - \Phi\left(-z_{\alpha/2} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$

where $\Phi(z) =$ the standard normal cdf.

The sample size n for which a level α test also has $\beta(\mu') = \beta$ at the alternative value μ' is

$$n = \begin{cases} \left[\dfrac{\sigma(z_\alpha + z_\beta)}{\mu_0 - \mu'}\right]^2 & \text{for a one-tailed (upper or lower) test} \\ \left[\dfrac{\sigma(z_{\alpha/2} + z_\beta)}{\mu_0 - \mu'}\right]^2 & \text{for a two-tailed test (an approximate solution)} \end{cases}$$

Case I: Normal Distribution and Known σ (z Test)

- Example

Let μ denote the true average tread life of a certain type of tire. Consider testing H_0: $\mu = 30{,}000$ versus H_a: $\mu > 30{,}000$ based on a sample of size $n = 16$ from a normal population distribution with $\sigma = 1500$. A test with $\alpha = .01$ requires $z_\alpha = z_{.01} = 2.33$. The probability of making a type II error when $\mu = 31{,}000$ is

$$\beta(31{,}000) = \Phi\left(2.33 + \frac{30{,}000 - 31{,}000}{1500/\sqrt{16}}\right) = \Phi(-.34) = .3669$$

Since $z_{.1} = 1.28$, the requirement that the level .01 test also have $\beta(31{,}000) = .1$ necessitates

$$n = \left[\frac{1500(2.33 + 1.28)}{30{,}000 - 31{,}000}\right]^2 = (-5.42)^2 = 29.32$$

The sample size must be an integer, so $n = 30$ tires should be used. ■