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 Concept 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 four 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 four forms
$$\begin{aligned} H_a :&\theta > \theta_0 \\ H_a :&\theta < \theta_0 \\ H_a :&\theta \neq \theta_0 \end{aligned}$$
- Null hypothesis and alternative hypothesis are not treated equally. In constructing Testing Procedures, we try to protect 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 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).
  - 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 $p = 0.5$ 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.

Two types of errors

- Definition

A type I error α consists of rejecting the null hypothesis H_0 when it is true.

A type II error β involves not rejecting H_0 when H_0 is false.

	Decide to accept	Decide to reject
Null is true	Right	Type I
Alternative is true	Type II	Right

Example 8.2 from the Textbook

- It is known the dying 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

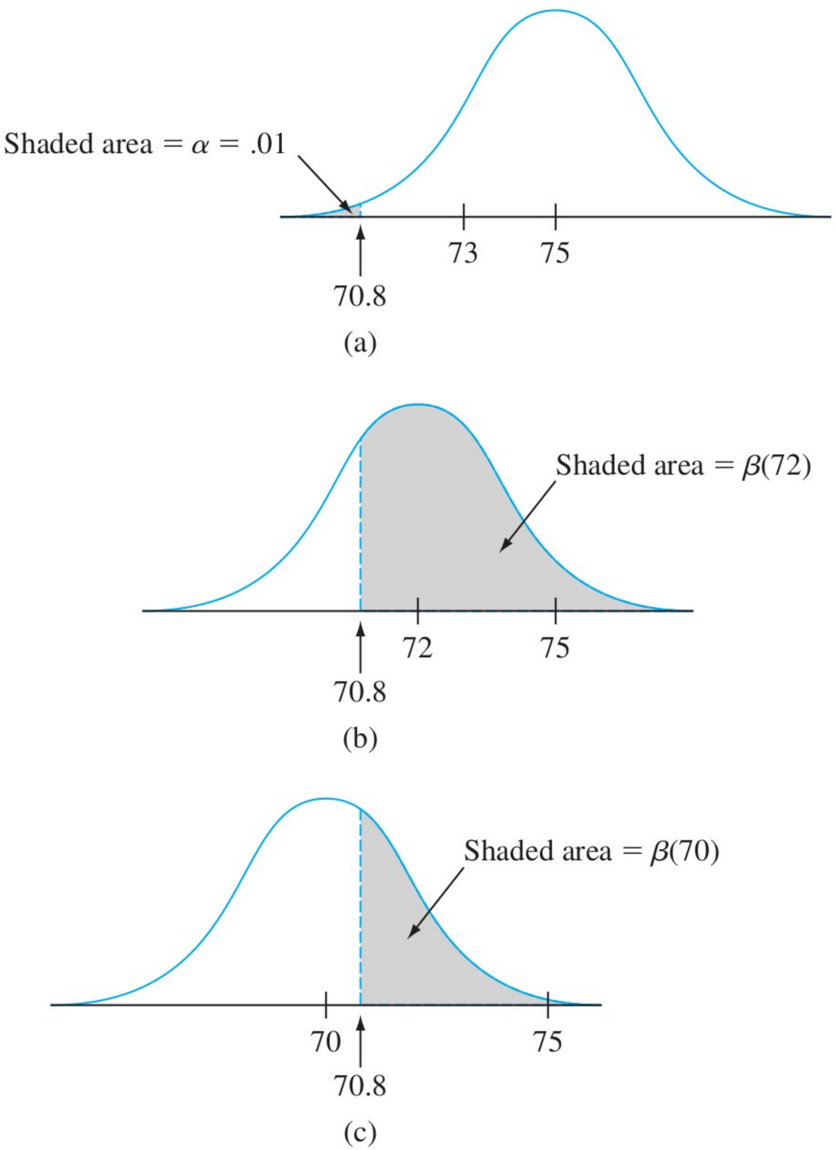

Figure: Illustrations of α and β for the testing procedure:(a) $\mu = 75$; (b) $\mu = 72$; (c) $\mu = 70$.

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 also used. It is simply $1 - \beta$, which means the probability of rejecting null hypothesis when it is false.