W1211 Introduction to Statistics
Lecture 26

Wei Wang

Dec 5, 2012

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)

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

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

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)

►

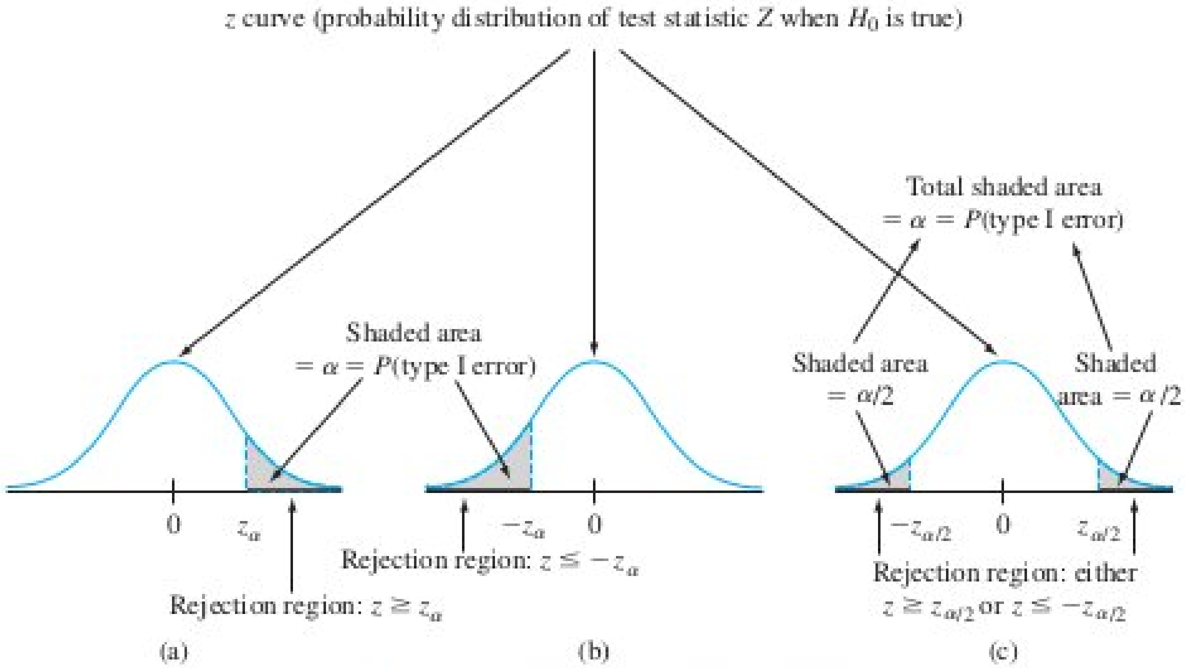


Figure 8.2 Rejection regions for z tests: (a) upper-tailed test; (b) lower-tailed test; (c) two-tailed test

►

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

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 II: General Distribution, Unknown σ but Large Sample (z Test)

- As we discussed in Confidence Interval, under the null hypothesis, the test statistic
$$Z = \frac{\bar{X} - \mu_0}{\hat{\sigma}\sqrt{n}}$$
approximately follow a standard normal distribution.
- The rule of thumb is $n > 40$.
- All the procedure, e.g., Test Statistic, Rejection Region and formula for β and sample size, are the same except for substituting σ with its estimator $\hat{\sigma}$.

Case III: Normal Distribution and Unknown σ (t Test)

- Under the null hypothesis, the test statistic
$$T = \frac{\bar{X} - \mu_0}{\hat{\sigma}\sqrt{n}}$$
follows a t distribution with degrees of freedom $n - 1$

Case III: Normal Distribution and Unknown σ (t Test)

- Under the null hypothesis, the test statistic
$$T = \frac{\bar{X} - \mu_0}{\hat{\sigma}\sqrt{n}}$$
follows a t distribution with degrees of freedom $n - 1$
- Test Procedure

The One-Sample t Test

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

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

Alternative Hypothesis	Rejection Region for a Level α Test
$H_a: \quad \mu > \mu_0$	$t \geq t_{\alpha,n-1}$ (upper-tailed)
$H_a: \quad \mu < \mu_0$	$t \leq -t_{\alpha,n-1}$ (lower-tailed)
$H_a: \quad \mu \neq \mu_0$	either $t \geq t_{\alpha/2,n-1}$ or $t \leq -t_{\alpha/2,n-1}$ (two-tailed)

Case III: Normal Distribution and Unknown σ (t Test)

- The calculation of Type II Error β is much more difficult than z Test.

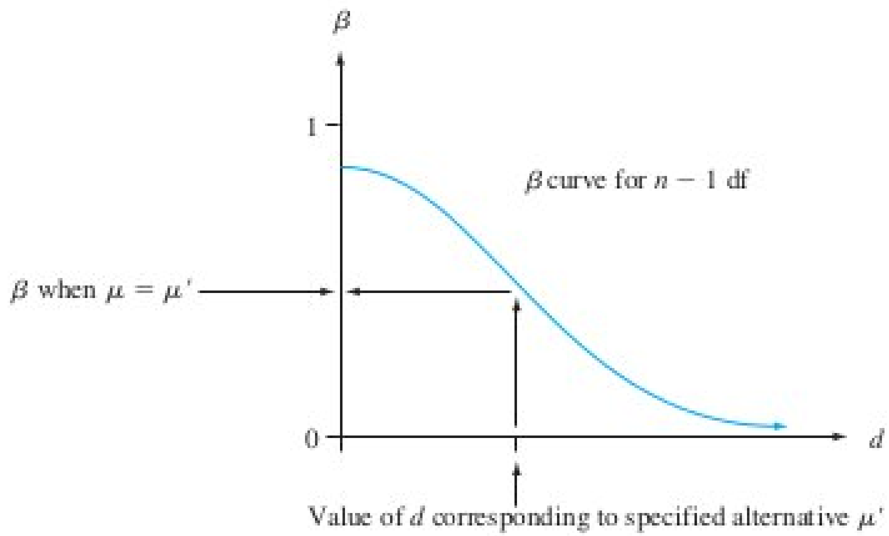
$$\beta(\mu') = P(T < t_{\alpha,n-1} \text{ when } \mu = \mu' \text{ rather than } \mu_0)$$

Case III: Normal Distribution and Unknown σ (t Test)

- The calculation of Type II Error β is much more difficult than z Test.

$$\beta(\mu') = P(T < t_{\alpha,n-1} \text{ when } \mu = \mu' \text{ rather than } \mu_0)$$

- A typical β curve



Hypothesis Testing for a Population Proportion

- Let p denote the proportion of individuals or objects in a population who possess a specified property (probability of success). In order to make inference about p, naturally we would look at the sample proportion, which is X/n. X is the number of Successes in the sample. In practice, X should follow a binomial distribution, and when n is large, it can further be approximated by a normal distribution.
- We will consider large sample tests only.

Large-sample tests

- Thanks to the Central Limit Theorem, we have

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} \sim N(0, 1)$$

  under the null hypothesis.

- Thus the rejection region is determined by

1. H_a: $p > p_0$: $Z > z_\alpha$
2. H_a: $p < p_0$: $Z < -z_\alpha$
3. H_a: $p \neq p_0$: $Z > z_{\alpha/2}$ or $Z_0 < -z_{\alpha/2}$

- The test procedures are valid provided that $np_0 \geq 10$ and $n(1-p_0) \geq 10$.

Example

Ex. (Defective rate cont.) A factory claims that less than 10% of the components they produce are defective. A consumer group is skeptical of the claim and checks a random sample of 300 components and finds that 39 are defective. Is there evidence that 10% of all components made at the factory are defective?

$H_0 : p = 0.10$ $H_a: p > 0.10$

$$\hat{p} = \frac{39}{300} = 0.13 \quad \mathrm{Z} = \frac{0.13 - 0.1}{\sqrt{0.1(1-0.1)/300}} = 1.72$$

$z_{0.05} = 1.645$. $Z > z_{0.05}$, thus we would reject H_0 at level α=0.05.

Type II Error

- We can calculate Type II Error based on the large sample normal approximation

$$\begin{aligned}
\beta(p') &= \mathrm{P}(H_0 \text{ is not rejected when } p = p') \\
&= \mathrm{P}\left(\frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \leq z_\alpha | p = p'\right) \\
&= \mathrm{P}\left(\frac{\hat{p} - p'}{\sqrt{p_0(1-p_0)/n}} \leq z_\alpha + \frac{p_0 - p'}{\sqrt{p_0(1-p_0)/n}} | p = p'\right) \\
&= \mathrm{P}\left(\frac{\hat{p} - p'}{\sqrt{p'(1-p')/n}} \leq \frac{z_\alpha \sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}} + \frac{(p_0 - p')}{\sqrt{p'(1-p')/n}} | p = p'\right) \\
&= \Phi\left(\frac{p_0 - p' + z_\alpha \sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right)
\end{aligned}$$

Determining sample size

- If we specify a particular alternative *p'* and specify a β value that can be tolerated (e.g. 0.1). Then from

$$\beta = \Phi\left(\frac{p_0 - p' + z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right) \Longrightarrow -z_\beta = \frac{p_0 - p' + z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}$$

- Therefore, in order to achieve the specified type I and type II error, one has to have a sample size of at least

$$n = \left(\frac{z_\alpha\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p'(1-p')}}{p' - p_0}\right)^2$$

- For two sided test, we have to change z_α to $z_{\alpha/2}$ in the above formula.

- Difference between the sample size calculation formula in chapter 7 and the one above.

Type II Error and Sample Size calculation

- In general Type II Error and Sample Size formulas are give below

Alternative Hypothesis	$\beta(p')$
H_a: $p > p_0$	$\Phi\left[\dfrac{p_0 - p' + z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right]$
H_a: $p < p_0$	$1 - \Phi\left[\dfrac{p_0 - p' - z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right]$
H_a: $p \neq p_0$	$\Phi\left[\dfrac{p_0 - p' + z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right] - \Phi\left[\dfrac{p_0 - p' - z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right]$

The sample size n for which the level α test also satisfies $\beta(p') = \beta$ is

$$n = \begin{cases} \left[\dfrac{z_\alpha\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p'(1-p')}}{p' - p_0}\right]^2 & \text{one-tailed test} \\ \left[\dfrac{z_{\alpha/2}\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p'(1-p')}}{p' - p_0}\right]^2 & \text{two-tailed test (an approximate solution)} \end{cases}$$

Example

<u>Ex.</u> A package-delivery service advertises that at least 90% of all packages brought to its office by 9 a.m. for delivery in the same city are delivered by noon that day. Let p denote the true proportion of such packages that are delivered as advertised and consider the hypothesis $H_0 : p = 0.9$ versus $H_a : p < 0.9$. If only 80% of the packages are delivered, how likely is it that a level .01 test based on n=225 packages will detect such departure from H_0? What should the sample size be to ensure that $\beta(0.8) = 0.01$? With $\alpha = .01$, $p_0 = .9$, $p' = .8$, and $n = 225$.

Type II error:
$$\begin{aligned}\beta(p') &= 1 - \Phi\left(\frac{p_0 - p' - z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right) \\ &= 1 - \Phi\left(\frac{.9 - .8 - 2.33\sqrt{(.9)(.1)/225}}{\sqrt{(.8)(.2)/225}}\right) \\ &= 1 - \Phi(2.00) = .0228\end{aligned}$$

Example cont.

- Using $z_{.01}$=2.33, the sample size can then be calculated from

$$\begin{aligned} n &= \left(\frac{z_\alpha \sqrt{p_0(1-p_0)/n} + z_\beta \sqrt{p'(1-p')/n}}{p' - p_0} \right)^2 \\ &= \left(\frac{2.33\sqrt{(.9)(.1)} + 2.33\sqrt{(.8)(.2)}}{.8 - .9} \right)^2 \approx 266 \end{aligned}$$

- 1-β is often referred to as the power of a test. It is the probability that the test can actually detect the alternative given the alternative is true! For α-level tests, the bigger the power the better!

P-Value

- To report the result of a hypothesis-testing analysis is to simply say whether the null hypothesis was rejected at a specified level of significance. This type of statement is somewhat inadequate because it says nothing about whether the conclusion was a very close call or quite clear cut.

- P-value is a quantity that conveys much information about the strength of evidence against H_0 and allows an individual decision maker to draw a conclusion at any specified level α.

- The P-value (*observed significance level*) is the probability, under the null hypothesis, that the test statistic is more ***extreme*** than the observed statistic.

Clarifications on the Concept of P-value

- What P-value is
  - The P-value is a probability.
  - This probability is calculated assuming **the null hypothesis is true**.

Clarifications on the Concept of P-value

- What P-value is
  - The P-value is a probability.
  - This probability is calculated assuming **the null hypothesis is true**.
- What P-Value is not
  - The P-value is not the probability that H_0 is true.
  - The P-value is not Type I Error α.
  - The P-value is not the significance level.
  - The P-value is not Type II Error β

Calculating P-value

- Example 8.14 in P329 in the textbook.

Calculating P-value

- Example 8.14 in P329 in the textbook.
- P-value=P(Test Statistic is more extreme than observed Test Statistic Value under Null Hypothesis)
- P-value provides a measure of the strength of the evidence.

Conducting Hypothesis Testing using P-value

- **The smaller the P-value, the more evidence there is in the sample data against the null hypothesis and for the alternative hypothesis.**
- P-values can be seen as a more flexible procedure of Hypothesis Testing. The practical advantage is that it is easier to switch to a test of different significance level
- The decision rule based on P-values

Decision rule based on the *P*-value

Select a significance level α (as before, the desired type I error probability). Then

$$\text{reject } H_0 \text{ if } P\text{-value} \leq \alpha$$
$$\text{do not reject } H_0 \text{ if } P\text{-value} > \alpha$$

Remarks

- The P-value is the smallest significance level α at which the null hypothesis can be rejected.
- P-value is a feature of the observed sample together with the test, whereas Type I Error α is a feature of a test, not related to any specific sample.

P-values and Tails

- To calculate P-values under an upper-tailed, lower-tailed and two-tailed test.