

# W1211 Introduction to Statistics

## Lecture 25

Wei Wang

Dec 3, 2012

# What we talked about last lecture

- ▶ Test Statistics
- ▶ Rejection Region
- ▶ Type I Error  $\alpha$
- ▶ Type II Error  $\beta$
- ▶ Power

# 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  $\sigma$  (z Test)
  - ▶ Case II: General Distribution, Unknown  $\sigma$  but Large Sample (z Test)
  - ▶ Case III: Normal Distribution and Unknown  $\sigma$  (t Test)

# Case I: Normal Distribution and Known $\sigma$ (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 $\sigma$ (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 $\sigma$ (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  $\alpha$ , if we set  $c$  as  $z$  critical value  $z_\alpha$  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 $\sigma$ (z Test)

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

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

Alternative Hypothesis

Rejection Region for Level  $\alpha$  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 $\sigma$ (z Test)

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

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

Alternative Hypothesis

$H_a: \mu > \mu_0$

$H_a: \mu < \mu_0$

$H_a: \mu \neq \mu_0$

Rejection Region for Level  $\alpha$  Test

$z \geq z_\alpha$  (upper-tailed test)

$z \leq -z_\alpha$  (lower-tailed test)

either  $z \geq z_{\alpha/2}$  or  $z \leq -z_{\alpha/2}$  (two-tailed test)

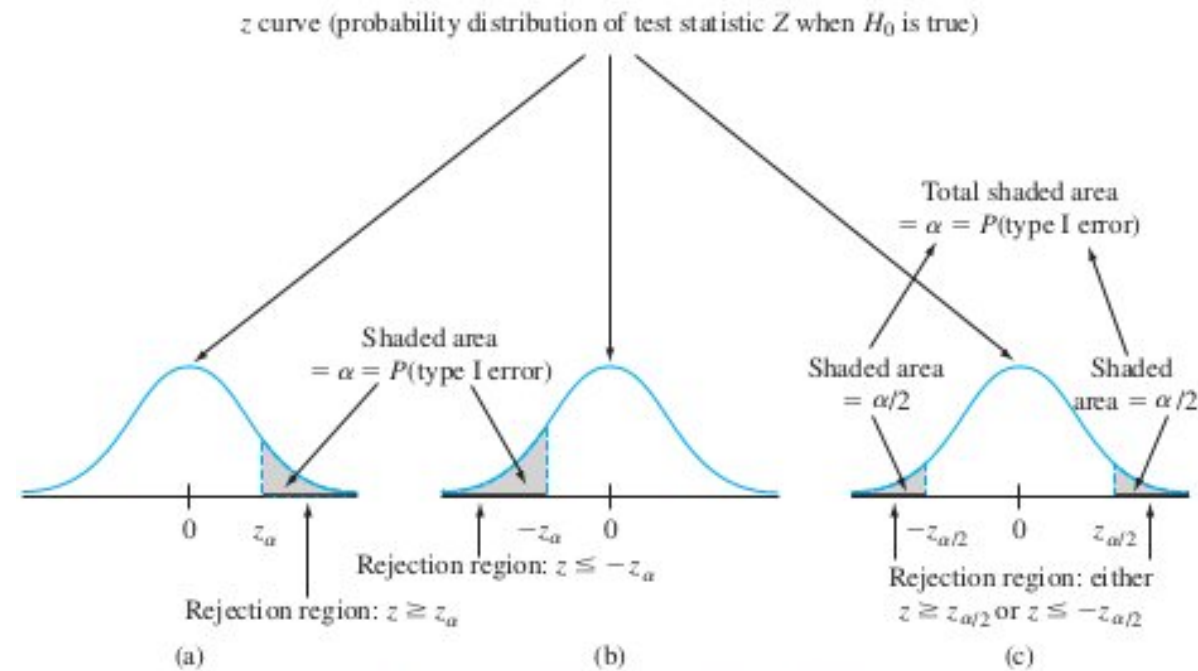


Figure 8.2 Rejection regions for z tests: (a) upper-tailed test; (b) lower-tailed test; (c) two-tailed test





# Case I: Normal Distribution and Known $\sigma$ (z Test)

- ▶ We can also compute Type II Error  $\beta$  and sample size  $n$ . Still we consider the upper-tailed test as a demonstration.
- ▶ Type II Error  $\beta$  will be a function of any particular number  $\mu'$  that is larger than the null value  $\mu_0$ .

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

$\Phi()$  is the CDF of standard normal.

- ▶ What is the power of the test?

# Case I: Normal Distribution and Known $\sigma$ (z Test)

- ▶ For a given True Value  $\mu'$ , Type I Error level  $\alpha$  and Type II Error  $\beta$ , we can determine the sample size  $n$ . We can equate the last expression to the desired  $\beta$  level.

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

# Case I: Normal Distribution and Known $\sigma$ (z Test)

- ▶ For a given True Value  $\mu'$ , Type I Error level  $\alpha$  and Type II Error  $\beta$ , we can determine the sample size  $n$ . We can equate the last expression to the desired  $\beta$  level.

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

- ▶ Thus

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

which gives

$$n = \left[ \frac{\sigma(z_{\alpha} + z_{\beta})}{\mu_0 - \mu'} \right]^2$$

# Case I: Normal Distribution and Known $\sigma$ (z Test)

Alternative Hypothesis    Type II Error Probability  $\beta(\mu')$  for a Level  $\alpha$  Test

$$\begin{aligned} H_a: \quad \mu &> \mu_0 && \Phi\left(z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right) \\ H_a: \quad \mu &< \mu_0 && 1 - \Phi\left(-z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right) \\ H_a: \quad \mu &\neq \mu_0 && \Phi\left(z_{\alpha/2} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right) - \Phi\left(-z_{\alpha/2} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right) \end{aligned}$$

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

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

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

# Case I: Normal Distribution and Known $\sigma$ (z Test)

## ► Example

Let  $\mu$  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. 

# Case II: General Distribution, Unknown $\sigma$ 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  $\beta$  and sample size, are the same except for substituting  $\sigma$  with its estimator  $\hat{\sigma}$ .

# Case III: Normal Distribution and Unknown $\sigma$ (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 $\sigma$ (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:  $t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$

### Alternative Hypothesis

$H_a: \mu > \mu_0$

$H_a: \mu < \mu_0$

$H_a: \mu \neq \mu_0$

### Rejection Region for a Level $\alpha$ Test

$t \geq t_{\alpha, n-1}$  (upper-tailed)

$t \leq -t_{\alpha, n-1}$  (lower-tailed)

either  $t \geq t_{\alpha/2, n-1}$  or  $t \leq -t_{\alpha/2, n-1}$  (two-tailed)



# Case III: Normal Distribution and Unknown $\sigma$ (t Test)

- ▶ The calculation of Type II Error  $\beta$  is much more difficult than z Test.

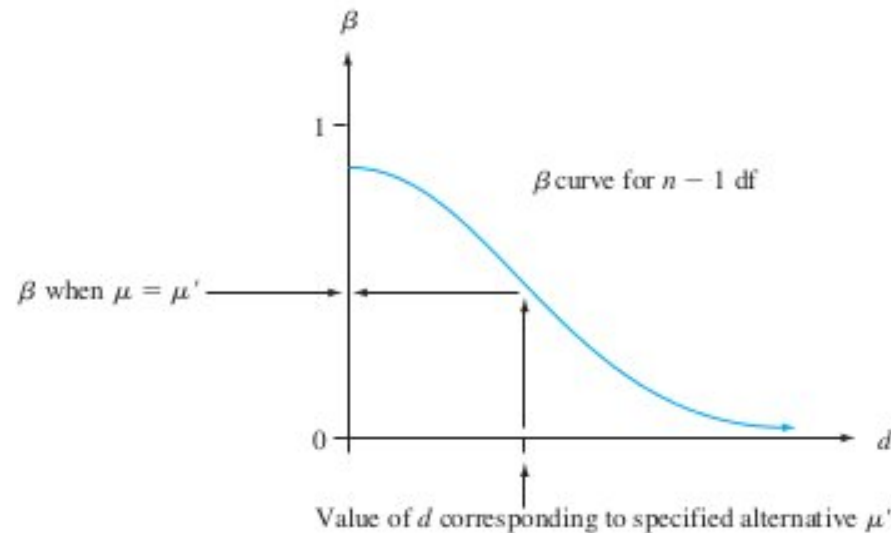
$$\beta(\mu') = P(T < t_{\alpha, n-1} \text{ when } \mu = \mu' \text{ rather than } \mu_0)$$

# Case III: Normal Distribution and Unknown $\sigma$ (t Test)

- ▶ The calculation of Type II Error  $\beta$  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  $\beta$  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.  $H_a: p \neq p_0: Z > z_{\alpha/2} \text{ or } Z < -z_{\alpha/2}$

- 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?

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

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

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

# Type II Error

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

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

# Determining sample size

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

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

- 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  $z_\alpha$  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

$\beta(p')$

$$\begin{aligned} H_a: p > p_0 & \quad \Phi \left[ \frac{p_0 - p' + z_\alpha \sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}} \right] \\ H_a: p < p_0 & \quad 1 - \Phi \left[ \frac{p_0 - p' - z_\alpha \sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}} \right] \\ H_a: p \neq p_0 & \quad \Phi \left[ \frac{p_0 - p' + z_{\alpha/2} \sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}} \right] \\ & \quad - \Phi \left[ \frac{p_0 - p' - z_{\alpha/2} \sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}} \right] \end{aligned}$$

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

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



# Example

Ex. 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  $n = 225$ .

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

## Example cont.

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

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

- $1-\beta$  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  $\alpha$ -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  $\alpha$ .
- 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  $\alpha$ .
  - ▶ The P-value is not the significance level.
  - ▶ The P-value is not Type II Error  $\beta$

# Calculating P-value

- ▶ Example 8.14 in P329 in the textbook.

# Calculating P-value

- ▶ Example 8.14 in P329 in the textbook.
- ▶  $P\text{-value} = P(\text{Test Statistic is more extreme than observed Test Statistic Value under Null Hypothesis})$
- ▶ 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  $\alpha$  (as before, the desired type I error probability).  
Then

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



# Remarks

- ▶ The P-value is the smallest significance level  $\alpha$  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  $\alpha$  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.

