

# Chapter 4

---

---

## Part II. Hypothesis Testing

---

---

# LEARNING OBJECTIVES

- Explain why hypothesis testing is important
- Describe the role of sampling in hypothesis testing
- Identify Type I and Type II errors and how they conflict with each other
- Interpret the confidence level, the significance level and the power of a test
- Compute and interpret  $p$ -values
- Determine the sample size and significance level for a given hypothesis test
- Plot power curves and operating characteristic curves

# Using Statistics

- A hypothesis is a statement or assertion about the state of nature (about the true value of an unknown population parameter):
  - ✓ The accused is innocent
  - ✓  $\mu = 100$
- Every hypothesis implies its contradiction or alternative:
  - ✓ The accused is guilty
  - ✓  $\mu \neq 100$
- A hypothesis is either true or false, and you may fail to reject it or you may reject it on the basis of information:
  - ✓ Trial testimony and evidence
  - ✓ Sample data

# Decision-Making

- One hypothesis is maintained to be true until a decision is made to reject it as false:
  - ✓ Guilt is proven “beyond a reasonable doubt”
  - ✓ The alternative is highly improbable
- A decision to fail to reject or reject a hypothesis may be:
  - ✓ Correct
    - A true hypothesis may not be rejected
      - » An innocent defendant may be acquitted
    - A false hypothesis may be rejected
      - » A guilty defendant may be convicted
  - ✓ Incorrect
    - A true hypothesis may be rejected
      - » An innocent defendant may be convicted
    - A false hypothesis may not be rejected
      - » A guilty defendant may be acquitted

# Statistical Hypothesis Testing

- A **null hypothesis**, denoted by  $H_0$ , is an assertion about one or more population parameters. This is the assertion we hold to be true until we have sufficient statistical evidence to conclude otherwise.

✓  $H_0: \mu = 100$

- The **alternative hypothesis**, denoted by  $H_1$ , is the assertion of all situations *not* covered by the null hypothesis.

✓  $H_1: \mu \neq 100$

- $H_0$  and  $H_1$  are:
  - ✓ Mutually exclusive
    - Only one can be true.
  - ✓ Exhaustive
    - Together they cover *all* possibilities, so one or the other *must* be true.

# Hypothesis about other Parameters

- Hypotheses about other parameters such as population proportions and and population variances are also possible. For example

✓  $H_0: p \geq 40\%$

✓  $H_1: p < 40\%$

✓  $H_0: \sigma^2 \leq 50$

✓  $H_1: \sigma^2 > 50$

# 1-Tailed and 2-Tailed Tests

The tails of a statistical test are determined by the need for an action. If action is to be taken if a parameter is greater than some value  $a$ , then the alternative hypothesis is that the parameter is greater than  $a$ , and the test is a **right-tailed** test.

$$H_0: \mu \leq 50$$

$$H_1: \mu > 50$$

If action is to be taken if a parameter is less than some value  $a$ , then the alternative hypothesis is that the parameter is less than  $a$ , and the test is a **left-tailed** test.

$$H_0: \mu \geq 50$$

$$H_1: \mu < 50$$

If action is to be taken if a parameter is either greater than or less than some value  $a$ , then the alternative hypothesis is that the parameter is not equal to  $a$ , and the test is a **two-tailed** test.

$$H_0: \mu = 50$$

$$H_1: \mu \neq 50$$

# The Null Hypothesis, $H_0$

- The **null hypothesis**:
  - ✓ Often represents the status quo situation or an existing belief.
  - ✓ Is maintained, or held to be true, until a **test** leads to its rejection in favor of the alternative hypothesis.
  - ✓ Is accepted as true or rejected as false on the basis of a consideration of a **test statistic**.



# The Concepts of Hypothesis Testing

- A **test statistic** is a sample statistic computed from sample data. The value of the test statistic is used in determining whether or not we may reject the null hypothesis.
- The **decision rule** of a statistical hypothesis test is a rule that specifies the conditions under which the null hypothesis may be rejected.

Consider  $H_0: \mu = 100$ . We may have a decision rule that says: “Reject  $H_0$  if the sample mean is less than 95 or more than 105.”

In a courtroom we may say: “The accused is innocent until proven guilty beyond a reasonable doubt.”

# Errors in Hypothesis Testing

- A decision may be incorrect in two ways:
  - ✓ Type I Error: Reject a true  $H_0$ 
    - The Probability of a Type I error is denoted by  $\alpha$ .
    - $\alpha$  is called the **level of significance** of the test
  - ✓ Type II Error: Accept a false  $H_0$ 
    - The Probability of a Type II error is denoted by  $\beta$ .
    - $1 - \beta$  is called the power of the test.
- $\alpha$  and  $\beta$  are conditional probabilities:
  - $\alpha = P(\text{Reject } H_0 \mid H_0 \text{ is true})$
  - ✓  $\beta = P(\text{Accept } H_0 \mid H_0 \text{ is false})$

# Type I and Type II Errors

A **contingency table** illustrates the possible outcomes of a statistical hypothesis test.

	<b>State of Nature</b>	
<b>Decision</b>	$H_0$ True	$H_0$ False
Do not Reject $H_0$	Correct	Type II Error ( $\beta$ )
Reject $H_0$	Type I Error ( $\alpha$ )	Correct

# The Power of a Test

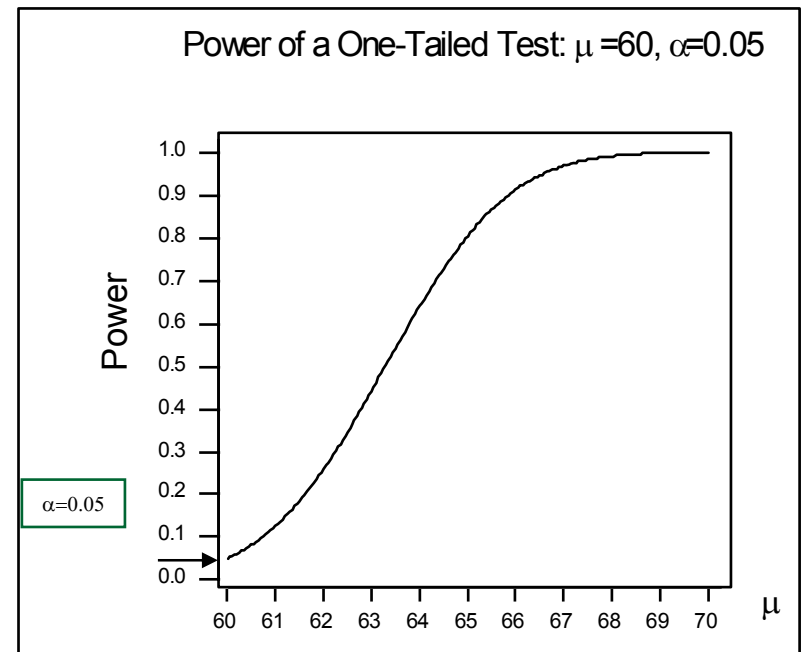
The **power** of a statistical hypothesis test is the probability of rejecting the null hypothesis when the null hypothesis is false.

$$\text{Power} = (1 - \beta)$$

# The Power Function

The probability of a type II error, and the power of a test, depends on the actual value of the unknown population parameter. The relationship between the population mean and the power of the test is called the **power function**.

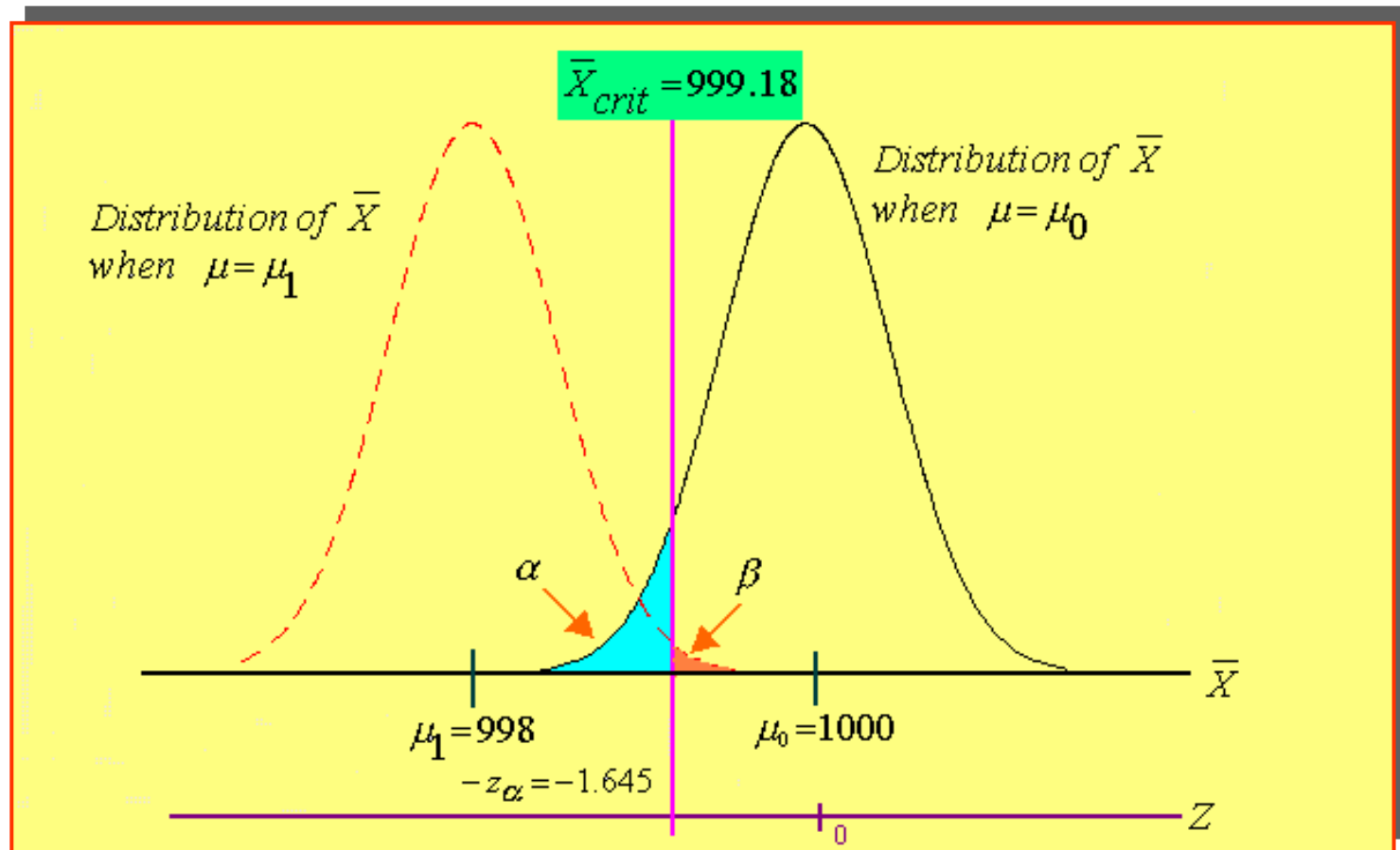
Value of $\mu$	$\beta$	Power = $(1 - \beta)$
61	0.8739	0.1261
62	0.7405	0.2695
63	0.5577	0.4423
64	0.3613	0.6387
65	0.1963	0.8037
66	0.0877	0.9123
67	0.0318	0.9682
68	0.0092	0.9908
69	0.0021	0.9972



# Factors Affecting the Power Function

- The power depends on the distance between the value of the parameter under the null hypothesis and the *true value* of the parameter in question: *the greater this distance, the greater the power.*
- The power depends on the population standard deviation: *the smaller the population standard deviation, the greater the power.*
- The power depends on the sample size used: *the larger the sample, the greater the power.*
- The power depends on the level of significance of the test: *the smaller the level of significance,  $\alpha$ , the smaller the power.*

# Significant level $\alpha$ and power $1-\beta$



# The $p$ -Value

The ***p-value*** is the probability of obtaining a value of the test statistic as extreme as, or more extreme than, the actual value obtained, when the null hypothesis is true.

The  $p$ -value is the smallest level of significance,  $\alpha$ , at which the null hypothesis may be rejected using the obtained value of the test statistic.

**Policy:** When the  $p$ -value is less than  $\alpha$ , reject  $H_0$ .

**NOTE:** More detailed discussions about the  $p$ -value will be given later in the chapter when examples on hypothesis tests are presented.



## Example

A company that delivers packages within a large metropolitan area claims that it takes an average of 28 minutes for a package to be delivered from your door to the destination. Suppose that you want to carry out a hypothesis test of this claim.

Set the null and alternative hypotheses:

$$H_0: \mu = 28$$

$$H_1: \mu \neq 28$$

Collect sample data:

$$n = 100$$

$$\bar{x} = 31.5$$

$$s = 5$$

Construct a 95% confidence interval for the average delivery times of *all* packages:

$$\begin{aligned}\bar{x} \pm z_{.025} \frac{s}{\sqrt{n}} &= 31.5 \pm 1.96 \frac{5}{\sqrt{100}} \\ &= 31.5 \pm .98 = [30.52, 32.48]\end{aligned}$$

We can be 95% sure that the average time for all packages is between 30.52 and 32.48 minutes.

**Since the asserted value, 28 minutes, is not in this 95% confidence interval, we may reasonably reject the null hypothesis.**

# Example

An automatic bottling machine fills cola into two liter (2000 cc) bottles. A consumer advocate wants to test the null hypothesis that the average amount filled by the machine into a bottle is at least 2000 cc. A random sample of 40 bottles coming out of the machine was selected and the exact content of the selected bottles are recorded. The sample mean was 1999.6 cc. The population standard deviation is known from past experience to be 1.30 cc. Compute the  $p$ -value for this test.

# Computing $\beta$ (for a left-tailed test)

Consider the following null and alternative hypotheses:

$$H_0: \mu \geq 1000$$

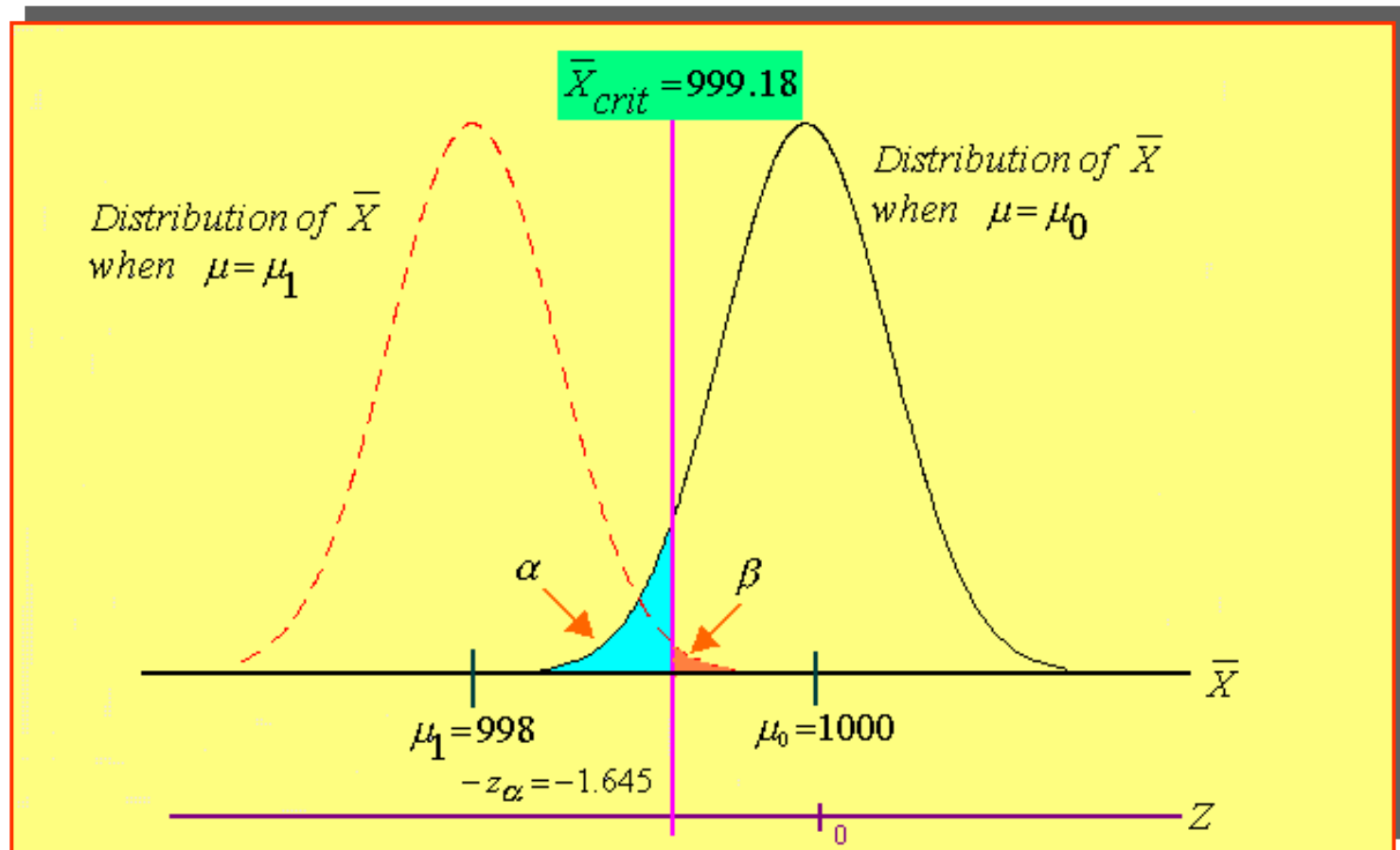
$$H_1: \mu < 1000$$

Let  $\sigma = 5$ ,  $\alpha = 5\%$ , and  $n = 100$ . We wish to compute  $\beta$  when  $\mu = \mu_1 = 998$ .

## Refer to next slide

- The figure shows the distribution of  $\bar{x}$  when  $\mu = \mu_0 = 1000$ , and when  $\mu = \mu_1 = 998$ .
- Note that  $H_0$  will be rejected when  $\bar{x}$  is less than the critical value given by  $(\bar{x})_{crit} = \mu_0 - z_\alpha * \sigma/\sqrt{n} = 1000 - 1.645*5/\sqrt{100} = 999.18$ .
- Conversely,  $H_0$  will not be rejected whenever  $\bar{x}$  is greater than  $(\bar{x})_{crit}$ .

# Computing $\beta$ (continued)



# The Concepts of Hypothesis Testing

- A **test statistic** is a sample statistic computed from sample data. The value of the test statistic is used in determining whether or not we may reject the null hypothesis.
- The **decision rule** of a statistical hypothesis test is a rule that specifies the conditions under which the null hypothesis may be rejected.

- The ***p-value*** is the probability of obtaining a value of the test statistic as extreme as, or more extreme than, the actual value obtained, when the null hypothesis is true.

**Policy:** When the *p-value* is less than  $\alpha$  , reject  $H_0$ .

# Testing Population Means

- Cases in which the **test statistic** is **Z**
  - ✓  $\sigma$  is known and the population is normal.
  - ✓  $\sigma$  is known and the sample size is at least 30. (The population need not be normal)

*The formula for calculating Z is :*

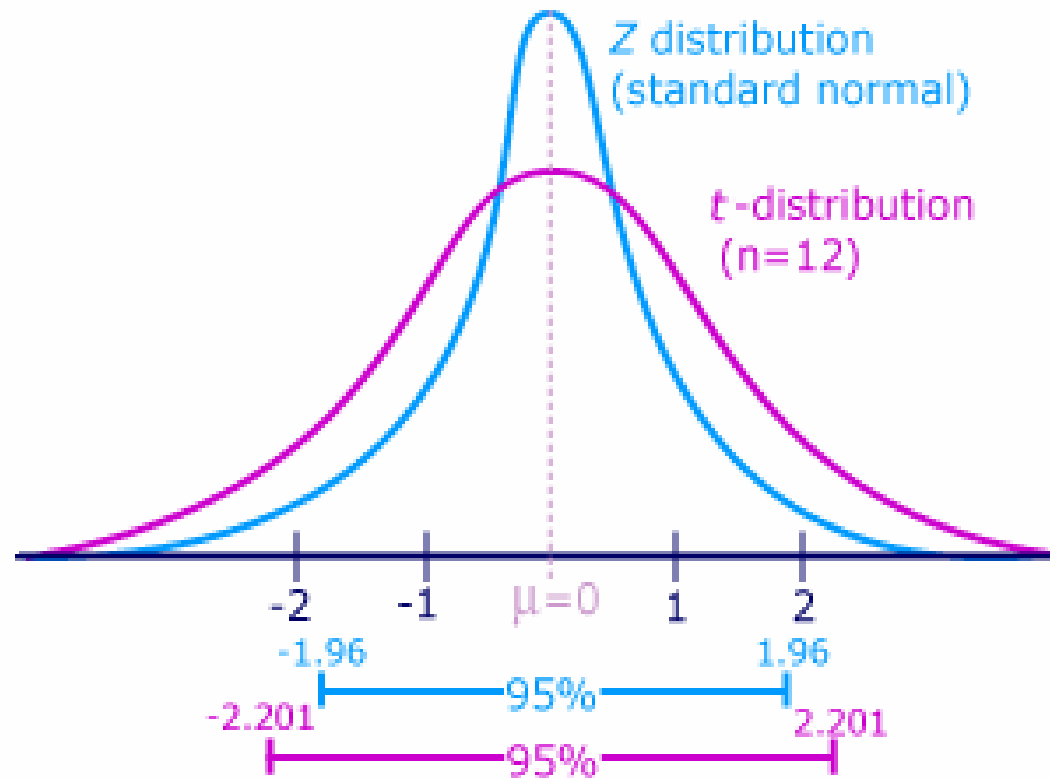
$$z = \frac{\bar{x} - \mu}{\left(\sigma / \sqrt{n}\right)}$$

- Cases in which the **test statistic** is **t**
  - ✓  $\sigma$  is unknown but the sample standard deviation is known and the population is normal.

*The formula for calculating t is :*

$$t = \frac{\bar{x} - \mu}{\left(s / \sqrt{n}\right)}$$

# Normal distribution and t-distribution



Test Statistic for population mean :  $\frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$

---

If the null hypothesis  $\mu = \mu_0$  is true, then  $\frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1)$

A random sample of size  $n$  with sample mean  $\bar{x}$  leads to an observed test statistic  $z_0 = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$

Given the significance level  $\alpha$ :

Case 1.  $H_1: \mu > \mu_0$ , the rejection region is  $\{z > z_\alpha\}$

---

$$\text{p-value} = P\{Z > z_0\}$$

Case 2.  $H_1: \mu < \mu_0$ , the rejection region is  $\{z < -z_\alpha\}$

$$\text{p-value} = P\{Z < z_0\}$$

Case 3.  $H_1: \mu \neq \mu_0$ , the rejection region is  $\{|z| > z_{\alpha/2}\}$

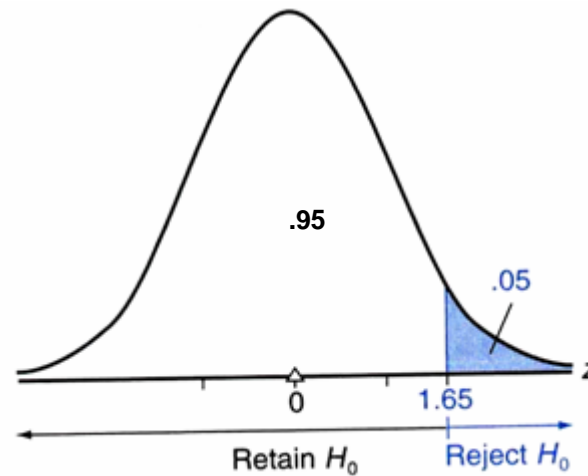
$$\text{p-value} = 2 \cdot P\{Z > |z_0|\}$$



# Critical Region

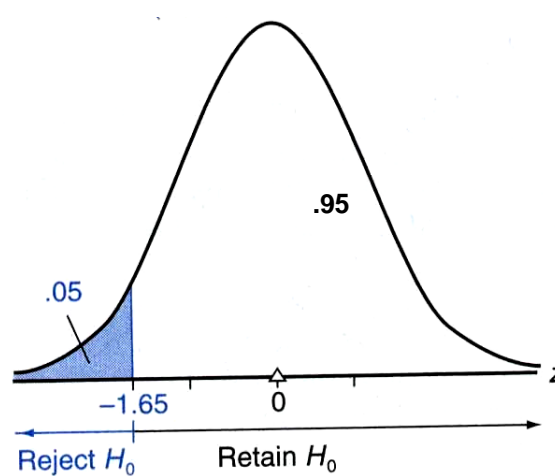
$H_0: \mu = \mu_0$  vs.  $H_1: \mu > \mu_0$

Right-tail Critical Region



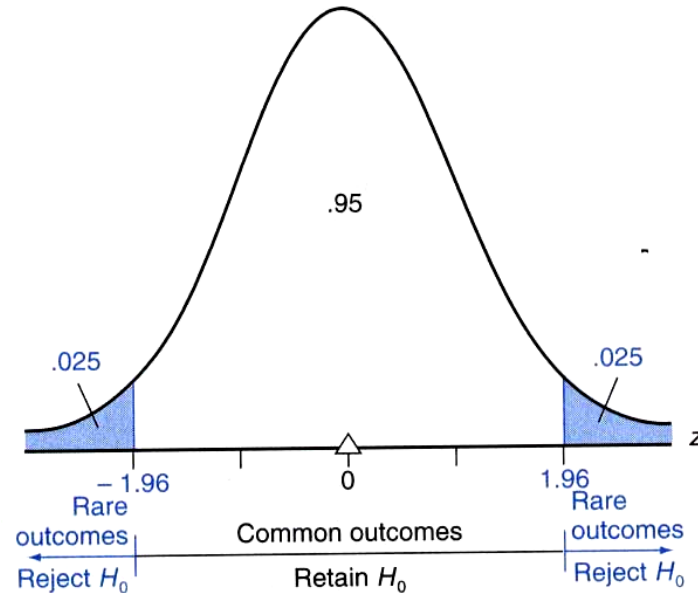
$H_0: \mu = \mu_0$  vs.  $H_1: \mu < \mu_0$

Left-tail Critical Region



$H_0: \mu = \mu_0$  vs.  $H_1: \mu \neq \mu_0$

Two-tailed Critical Region



Test Statistic for population mean if  $\sigma$  is unknown :  $\frac{\bar{X} - \mu_0}{s / \sqrt{n}}$

---

If the null hypothesis  $\mu = \mu_0$  is true, then  $\frac{\bar{X} - \mu_0}{s / \sqrt{n}} \sim t(n-1)$

A random sample of size  $n$  with sample mean  $\bar{x}$  leads to an observed test statistic  $t_0 = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$

Given the significance level  $\alpha$ :

Case 1.  $H_1: \mu > \mu_0$ , the rejection region is  $\{t > t(\alpha, n-1)\}$

---

$$\text{p-value} = P\{t(n-1) > t_0\}$$

Case 2.  $H_1: \mu < \mu_0$ , the rejection region is  $\{t < -t(\alpha, n-1)\}$

$$\text{p-value} = P\{t(n-1) < t_0\}$$

Case 3.  $H_1: \mu \neq \mu_0$ , the rejection region is  $\{|t(n-1)| > t(\alpha/2, n-1)\}$

$$\text{p-value} = 2 \cdot P\{t(n-1) > |t_0|\}$$

## Example

A floodlight is said to last an average of 65 hours. A competitor believes that the average life of the floodlight is less than that stated by the manufacturer and sets out to prove that the manufacturer's claim is false. A random sample of 21 floodlight elements is chosen and shows that the sample average is 62.5 hours and the sample standard deviation is 3. Using  $\alpha=0.01$ , determine whether there is evidence to conclude that the manufacturer's claim is false.

## Testing Population Proportion $p$

If the null hypothesis  $p = p_0$  is true, then 
$$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$$

A random sample of size  $n$  with sample proportion  $\hat{p}$  leads to an observed test statistic

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Given the significance level  $\alpha$ :

Case 1.  $H_1: p > p_0$ , the rejection region is  $\{z > z_\alpha\}$

$$\text{p-value} = P\{Z > z_0\}$$

Case 2.  $H_1: p < p_0$ , the rejection region is  $\{z < -z_\alpha\}$

$$\text{p-value} = P\{Z < z_0\}$$

Case 3.  $H_1: p \neq p_0$ , the rejection region is  $\{|z| > z_{\alpha/2}\}$

$$\text{p-value} = 2 \cdot P\{Z > |z_0|\}$$

## Example

A coin is to be tested for fairness. It is tossed 200 times and only 80 Heads are observed. Test if the coin is fair at an  $\alpha$  of 5% (significance level).

Let  $p$  denote the probability of a Head

$H_0: p = 0.5$       vs.       $H_1: p \neq 0.5$        $\alpha = 5\%$

Test statistic:

Critical region:

$p$ -value :

Confidence interval

## Example

- “After looking at 1349 hotels nationwide, we’ve found 13 that meet our standards.” This statement by the Small Luxury Hotels Association implies that the proportion of all hotels in the United States that meet the association’s standards is  $13/1349=0.0096$ .
- The management of a hotel that was denied acceptance to the association wanted to prove that the standards are not as stringent as claimed and that, in fact, the proportion of all hotels in the United States that would qualify is higher than 0.0096.
- The management hired an independent research agency, which visited a random sample of 600 hotels nationwide and found that 7 of them satisfied the exact standards set by the association.
- Is there evidence to conclude that the population proportion of all hotels in the country satisfying the standards set by the Small Luxury hotels Association is greater than 0.0096 given level  $\alpha=0.10$ ?

$H_0$ : \_\_\_\_\_ vs.  $H_1$ : \_\_\_\_\_

For  $\alpha = 0.10$  the critical value

Reject  $H_0$  if { \_\_\_\_\_ }

The test statistic is:

Decision:

p-Value =