9. Large-Sample Tests of Hypotheses

Adams Wai Kin Kong
School of Computer Science and Engineering
Nanyang Technological University, Singapore
adamskong@ntu.edu.sg

9.1 A Statistical Test of Hypothesis

A Statistical Test of Hypothesis (1 of 8)

A statistical test of hypothesis consists of five parts:

1. The null hypothesis, denoted by H_0
2. The alternative hypothesis, denoted by H_a
3. The test statistic and its p-value
4. The significance level and the rejection region
5. The conclusion

A Statistical Test of Hypothesis (2 of 8)

Definition:

The two competing hypotheses are the **alternative hypothesis** $\boldsymbol{H_a}$—generally the hypothesis that the researcher wishes to support—and the **null hypothesis** $\boldsymbol{H_0}$**,** a contradiction of the alternative hypothesis.

A Statistical Test of Hypothesis (3 of 8)

A statistical test of hypothesis begins by assuming that the null hypothesis, H_0, is true. If the researcher wants to show support for the alternative hypothesis, H_a, he needs to show that H_0 is false, using the sample data to decide whether the evidence favors *H*a rather than H_0. The researcher then draws one of two **conclusions:**

- Reject H_0 and conclude that H_a is true.
- Do not reject H_0.

Example 9.1

You wish to show that the average hourly wage of electricians in the state of California is different from $21, which is the national average.

This is the alternative hypothesis, written as

$$H_a : \mu \neq 21$$

The null hypothesis is

$$H_0 : \mu = 21$$

If you reject the null hypothesis, you can conclude that the average hourly wage is not equal to $21.

Example 9.2 (1/2)

A die cutting process for sheet metal currently produces an average of 3% defectives. You are interested in showing that a simple adjustment on a machine will decrease p, the proportion of defectives produced in the die cutting process.

Thus, the alternative hypothesis is

$$H_a : p < .03$$

and the null hypothesis is

$$H_0 : p = .03$$

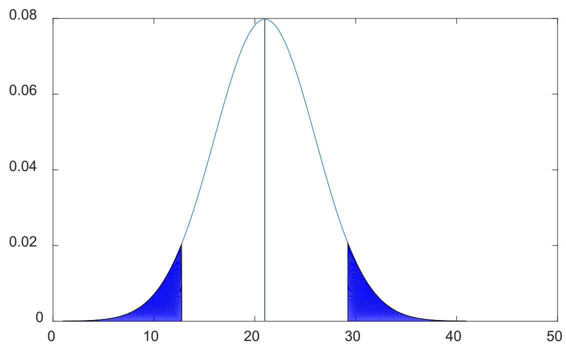
https://en.wikipedia.org/wiki/Die_cutting_(web)

Example 9.2 (2 of 2)

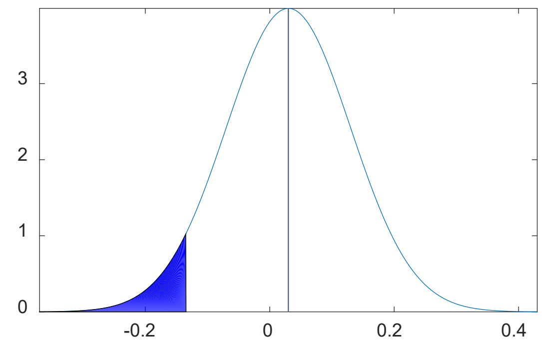
If you reject H_0, you will conclude that the adjusted process produces fewer than 3% defectives.

A Statistical Test of Hypothesis (4 of 8)

Example 9.1 is an example of **two-tailed test of hypothesis.** $(H_a\colon \mu \neq 21)$



Example 9.2 is an example of **one-tailed test of hypothesis.** $(H_a\colon p < 0.03)$



A Statistical Test of Hypothesis (5 of 8)

To decide whether to reject or accept H_0, we can use two pieces of information calculated from a sample, drawn from the population of interest:

- The **test statistic:** a single number calculated from the sample data. The test statistic is generally based on the best estimator for the parameter to be tested.
- The ***p*-value:** a probability calculated using the test statistic

Example 9.3 (1 of 4)

The test of hypothesis in Example 9.1 involves the average hourly wage of California electricians, in the form:

$$H_0 : \mu = 21 \text{ versus } H_a : \mu \neq 21$$

Assume that the null hypothesis H_0 is true and that $\mu = 21$. We take a random sample of 100 California electricians and find $\bar{x} = 22$ with a standard deviation of $s = 2$. Is this sample result unusual, given that H_0 is true? You can use two different measures to find out.

Example 9.3 (2 of 4)

- $\bar{x}$ approximately follows normal distribution (CLT, n>30)

- The standard error (standard deviation) of $\bar{x}$ is

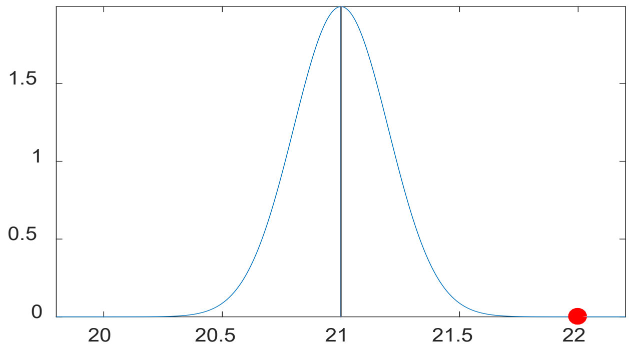
$$SE = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}} = \frac{2}{\sqrt{100}} = .2$$

Example 9.3 (3 of 4)

- Converting to z, we obtain the **test statistic**,

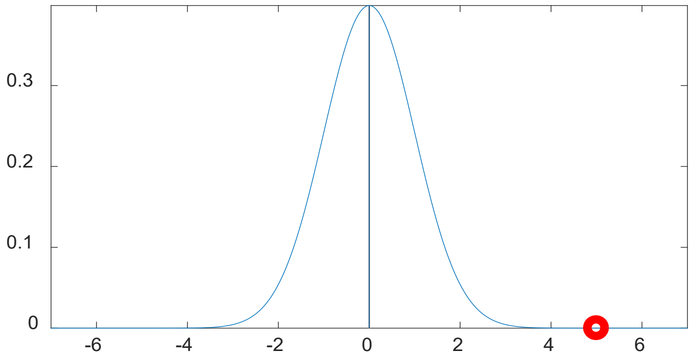
$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \approx \frac{22 - 21}{.2} = 5$$

- The sample mean $\bar{x} = 22$, lies 5 standard deviations above the mean—a *very* unlikely event if H_0 is true.



Distribution of $\bar{x}$

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$



Distribution of z

Example 9.3 (4 of 4)

- What is the probability of observing $\bar{x} = 22$ or something even more unlikely if μ = 21? The value $\bar{x} = 22$ lies 5 standard deviations above μ = 21, but an equally "unlikely" value of $\bar{x}$ would be one lying 5 standard deviations below μ = 21.

This probability is called the ***p*-value**, calculated as

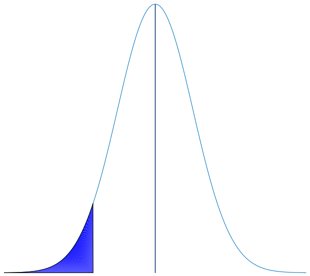
$$p\text{-value} = P(z > 5) + P(z < -5) \approx 0$$

- Again, this is a very unlikely event, if indeed H_0 is true and μ = 21. Note that we compute P(z>5) and P(z<-5) because it is a two-side test.

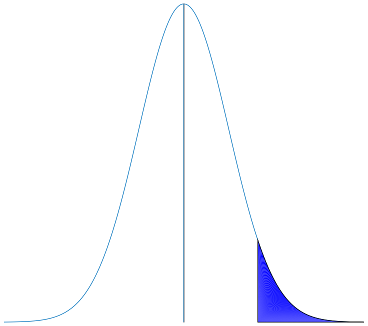
NANYANG TECHNOLOGICAL UNIVERSITY

A Statistical Test of Hypothesis (7 of 8)

- When the rejection region is in the left tail of the distribution, the test is called a **left-tailed test**.
  (e.g., $H_a : p < .03$)



- A test with its rejection region in the right tail is called a **right-tailed test**. (e.g., $H_a : \mu > 21$)



A Statistical Test of Hypothesis (8 of 8)

Definition:

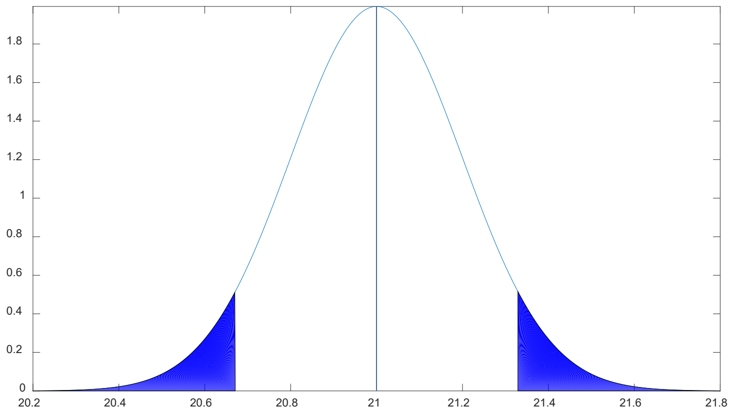
The **level of significance (significance level) $\boldsymbol{\alpha}$** for a statistical test of hypothesis is

$$\alpha = P(\text{falsely rejecting } H_0) = P(\text{rejecting } H_0 \text{ when it is true})$$

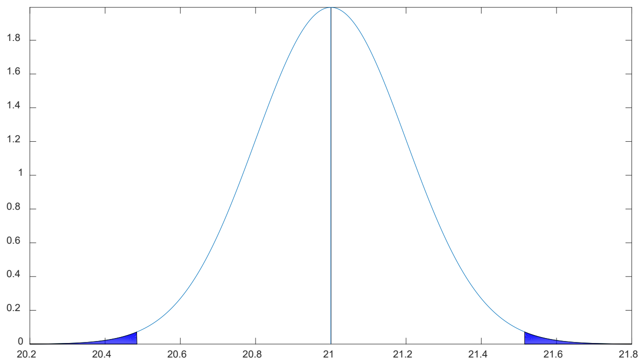
The value α represents the *maximum tolerable risk* of incorrectly rejecting H_0. Once this significance level is fixed, the rejection region can be set to allow the researcher to reject H_0 with a fixed degree of confidence in the decision.

A Statistical Test of Hypothesis

- $\alpha = 0.1, \mu = 21, SE = 0.2$, two-tailed test
- $\alpha = 0.01, \mu = 21, SE = 0.2$, two-tailed test



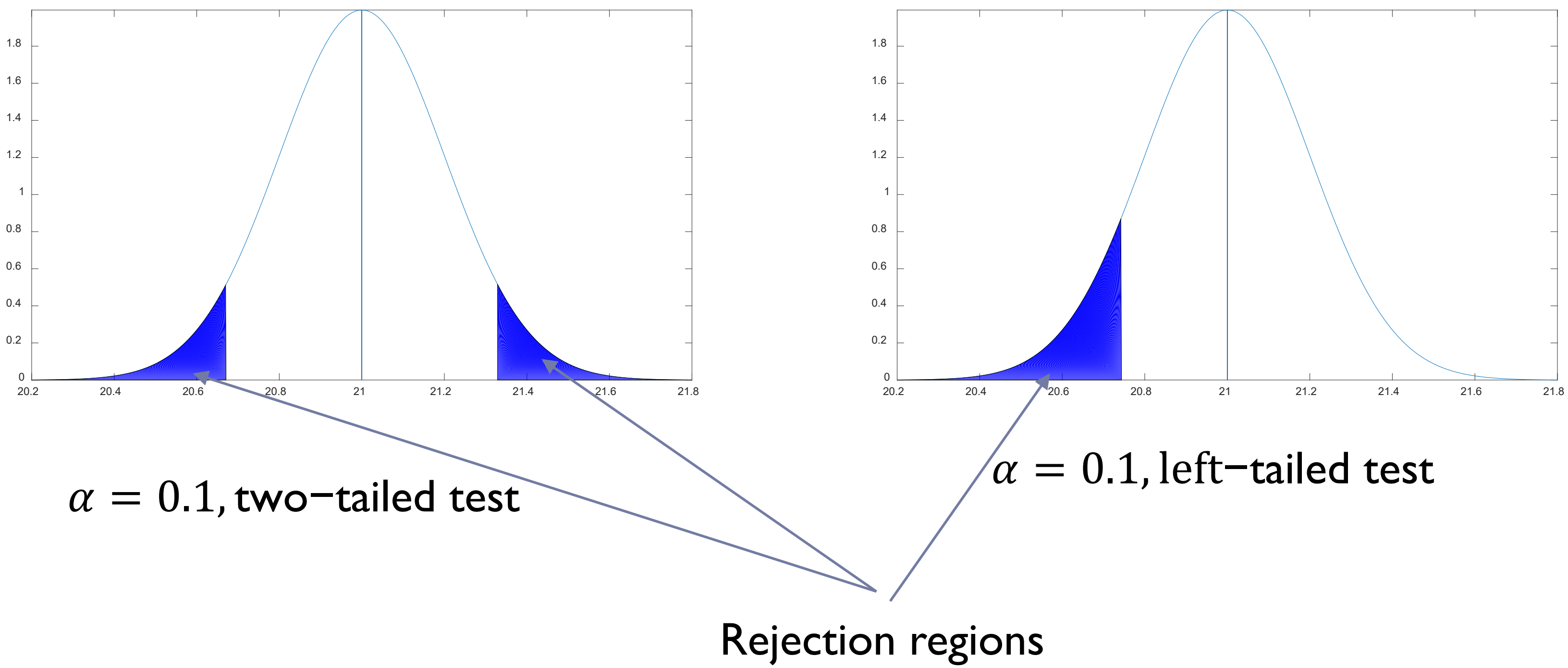
$\alpha = 0.1$



$\alpha = 0.01$

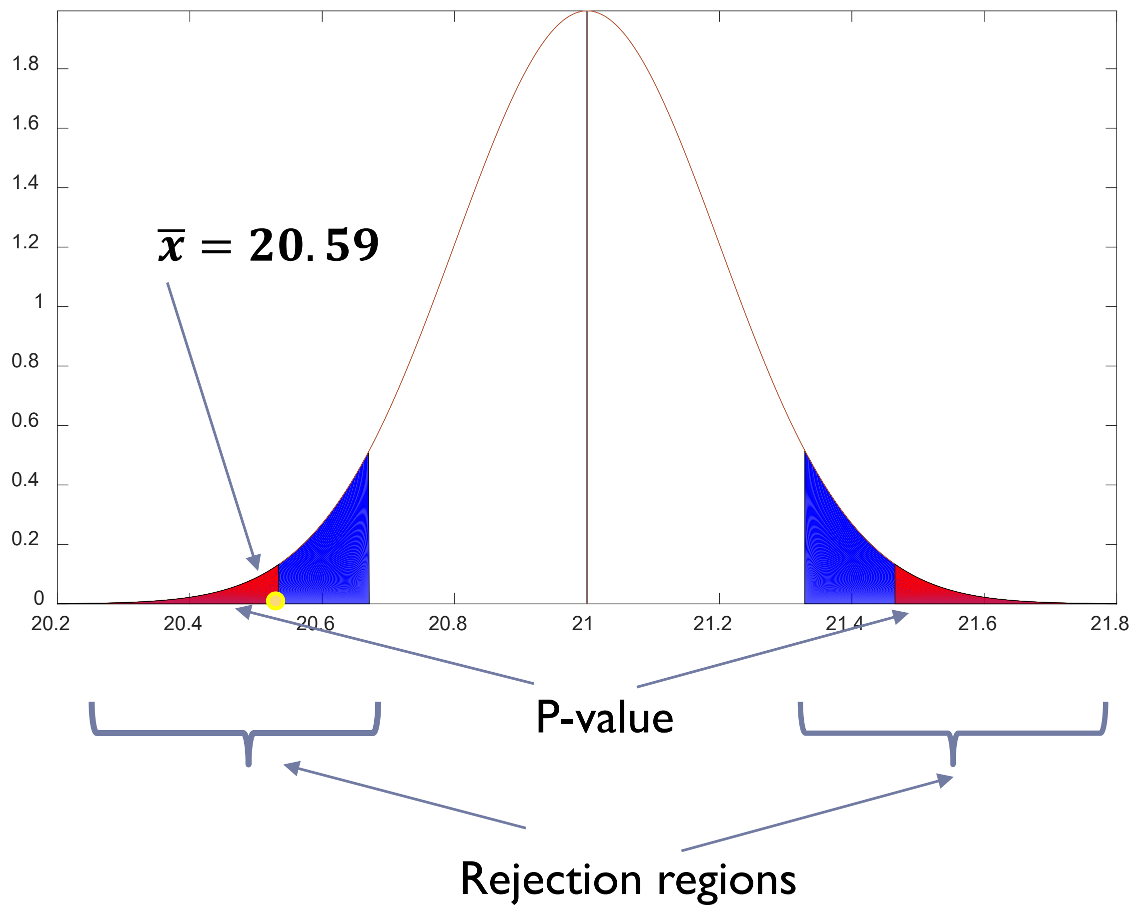
A Statistical Test of Hypothesis

- $\alpha = 0.1, \mu = 21, SE = 0.2,$ **two-tailed test**
- $\alpha = 0.1, \mu = 21, SE = 0.2,$ **left-tailed test**



A Statistical Test of Hypothesis

- $\alpha = 0.1, \mu = 21, SE = 0.2, \bar{x} = 20.59,$ **two-tailed test**



9.2 A Large-Sample Test About a Population Mean

The Essentials of the Test

The Essentials of the Test (1 of 6)

The sample mean $\bar{x}$ is the best estimate of the actual value of μ, which is presently in question. If H_0 is true and $\mu = \mu_0$, then $\bar{x}$ should be fairly close to μ_0. But if $\bar{x}$ is much *larger or smaller* than μ_0, this would indicate that H_a might be true.

Hence, you should reject H_0 in favor of a H_a if $\bar{x}$ is much larger or smaller than expected.

The Essentials of the Test (2 of 6)

- $\bar{x}$ follows approximately normal when n **is large.**
- The number of standard deviations that $\bar{x}$ lies from μ_0 can be measured using the **test statistic,**

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

which has an approximate standard normal distribution when H_0 is true and $\mu = \mu_0$.

The Essentials of the Test (3 of 6)

The rejection region, shown in figure, consists of values of z, which are much larger than expected.

Since the significance level α is defined as P(rejecting H_0 when H_0 is true), it is the area under the curve above the rejection region—the shaded area under the curve in figure. The critical value of z cutting off area α in the right tail is called z_α

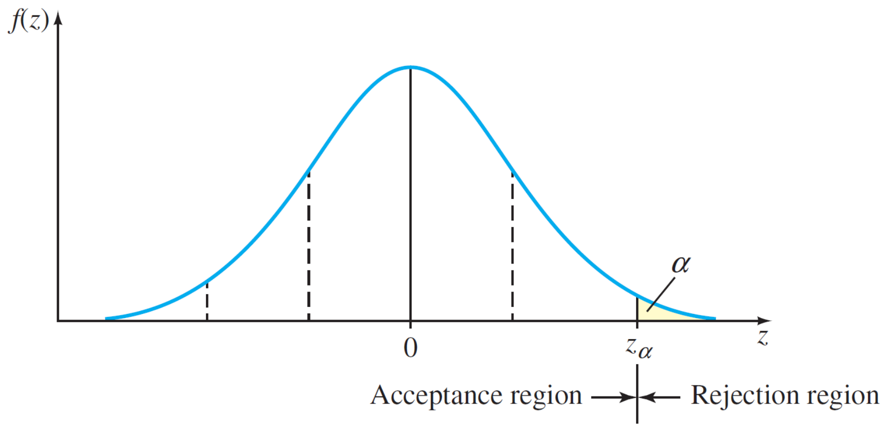


Figure 9.3

The rejection region for a right-tailed test with significance level α

Example 9.4

The average weekly earnings for female social workers is $670. Do men in the same positions have average weekly earnings that are higher than those for women? A random sample of $n = 40$ male social workers showed $\bar{x} = \$725$ and s=$102.

Test the appropriate hypothesis using $\alpha = 0.01$

Example 9.4 – Solution (1 of 4)

You would like to show that the average weekly earnings for men are higher than \$670, the women's average. Hence, if μ is the average weekly earnings for male social workers, you can set out the formal test of hypothesis in steps:

Null and alternative hypotheses:

$$H_0 : \mu = 670 \quad \text{versus} \quad H_a : \mu > 670$$

Test statistic: Using the sample information, with *s* as an estimate of the population standard deviation, calculate

$$z \approx \frac{\bar{x} - 670}{s/\sqrt{n}} = \frac{725 - 670}{102/\sqrt{40}} = 3.41$$

NANYANG TECHNOLOGICAL UNIVERSITY

Example 9.4 – Solution (2 of 4)

Rejection region: When α = 0.01, $z_{0.01} = 2.33$. Rejection region: $z > 2.33$. Note it is a **right-tailed test**.

Since z=3.41>2.33, H_0 is rejected.

Confidence Coefficient, $(1-\alpha)$	α	$\alpha/2$	$z_{\alpha/2}$
.90	.10	.05	1.645
.95	.05	.025	1.96
.98	.02	.01	2.33
.99	.01	.005	2.58

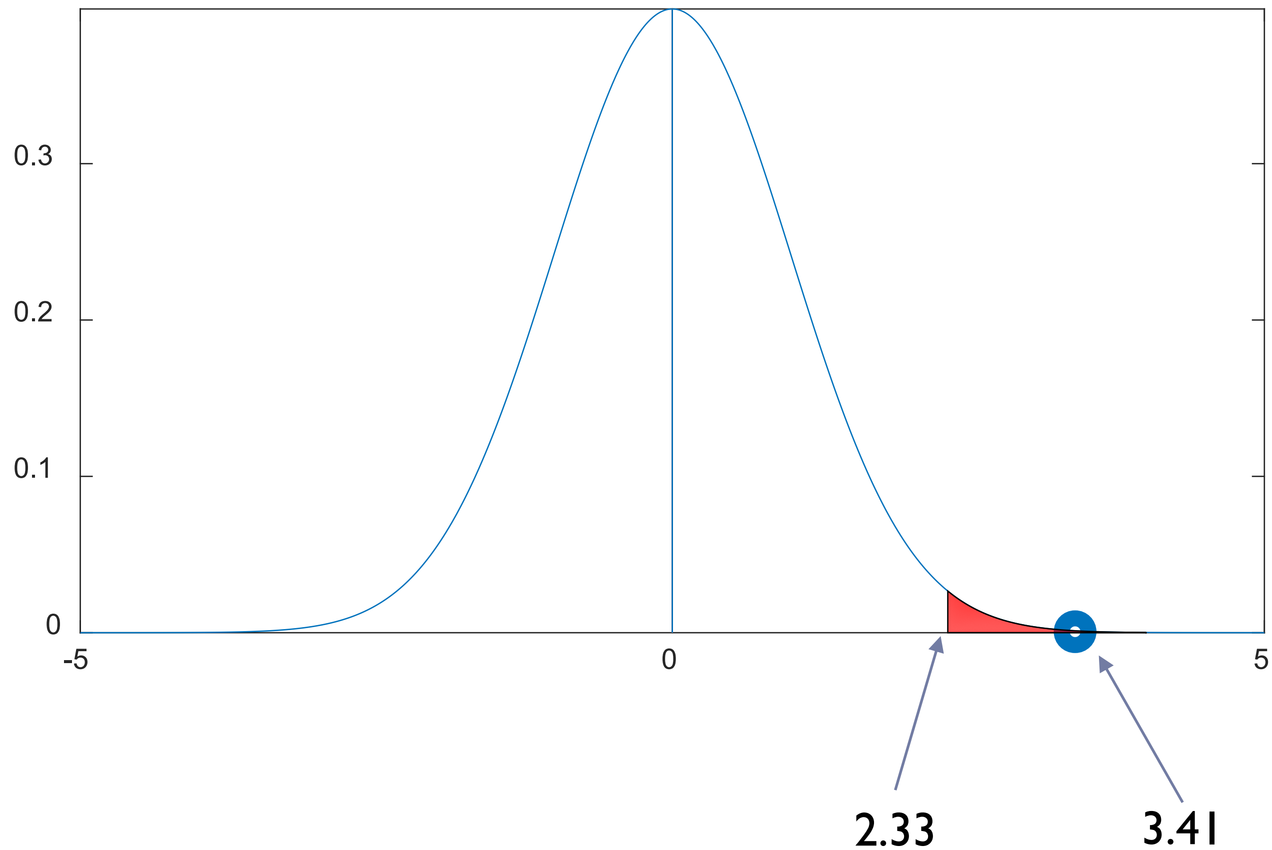
It is one-sided test.

Solution (3 of 4)

Conclusion: the average weekly earnings for male social workers are higher than the average for female social workers.

The probability that you have made an incorrect decision is α =0 .01.

Solution (4 of 4)



Large-Sample Statistical Test for μ

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

2. Alternative hypothesis:

One-Tailed Test	Two-Tailed Test
$H_a : \mu > \mu_0$	$H_a : \mu \neq \mu_0$
(or, $H_a : \mu < \mu_0$)	

3. Test statistic: $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ estimated as $z = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

The Essentials of the Test (5 of 6)

4. Rejection region: Reject H_0 when

One-Tailed Test

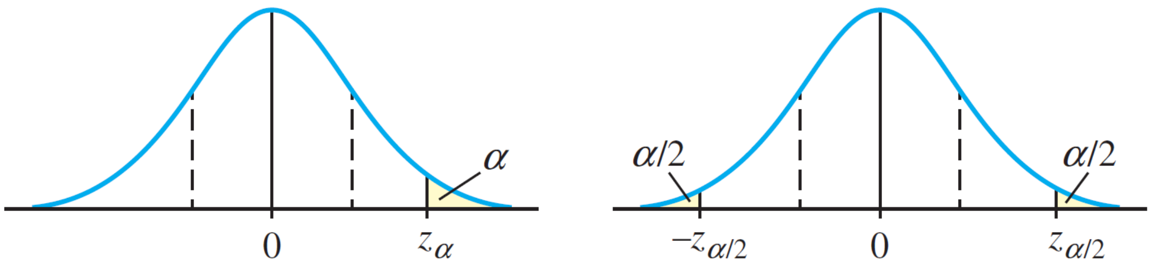
$z > z_\alpha$

(or, $z < -z_\alpha$ when the alternative hypothesis is $H_a : \mu < \mu_0$)

Two-Tailed Test

$z > z_{\alpha/2}$ or $z < -z_{\alpha/2}$

The Essentials of the Test (6 of 6)



Assumptions: The *n* observations in the sample are randomly selected from the population and *n* is large—say, $n \geq 30$.

Calculating the p-Value

Calculating the p-Value (1 of 6)

In the previous examples, the decision to reject or accept H_0 was made by comparing the calculated value of the test statistic with a critical value of z based on the significance level α of the test.

However, different significance levels may lead to different conclusions.

In example 9.4 If $\alpha = 0.01, Z_\alpha$=2.33. Since $Z > Z_\alpha$ (3.41>2.33), rejecting H_0.

If $\alpha = 0.0001, Z_\alpha = 3.72$. Since $Z < Z_\alpha$ ($3.41 < 3.72$), not rejecting H_0.

NANYANG TECHNOLOGICAL UNIVERSITY

Calculating the p-Value (2 of 6)

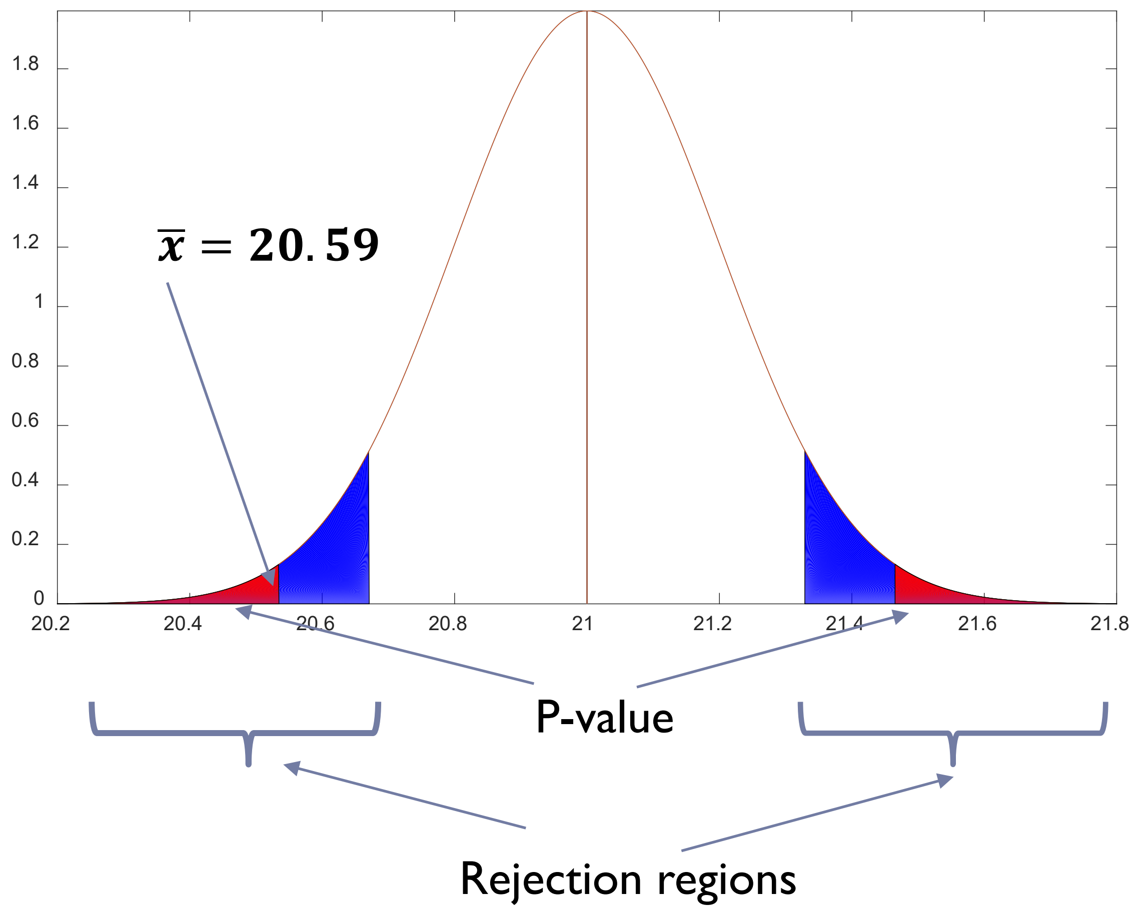
To avoid any ambiguity in their conclusions, some experimenters prefer to use a variable level of significance called the ***p*-value** for the test.

Definition

The ***p*-value** or observed significance level of a statistical test is the smallest value of α for which H_0 can be rejected. It is the *actual risk* of committing a Type I error, if H_0 is rejected based on the observed value of the test statistic. The p-value measures the strength of the evidence against H_0.

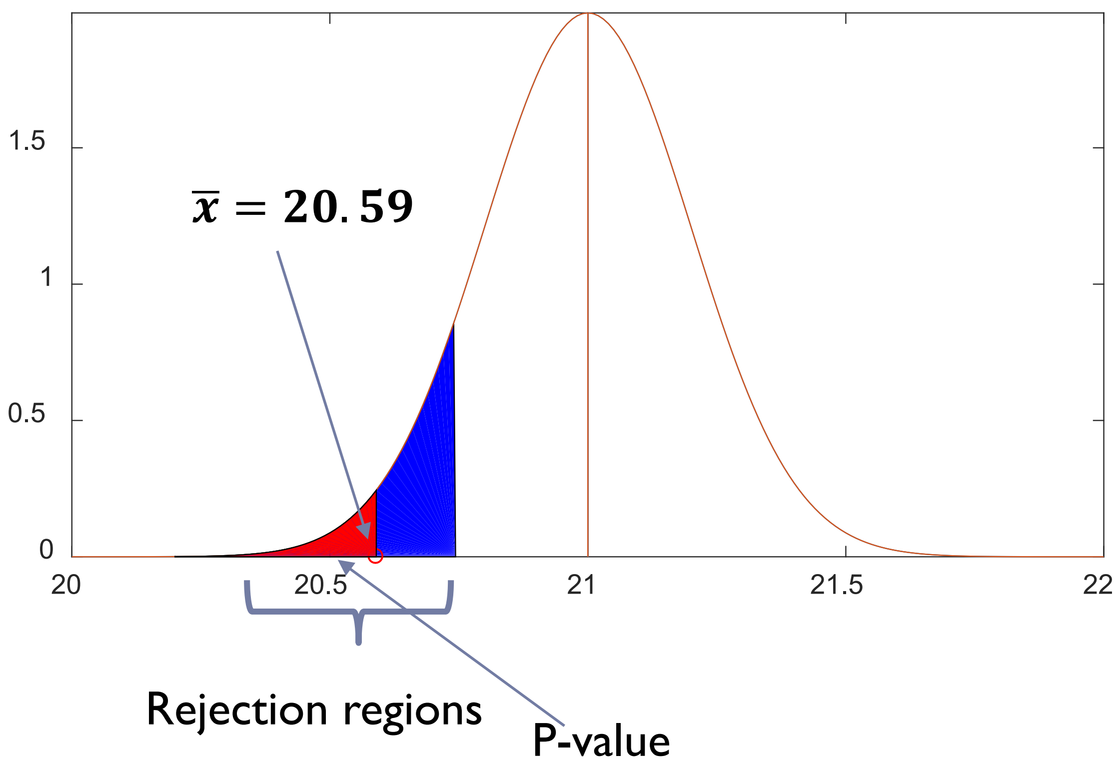
Calculating the *p*-Value (2 of 6) (Illustration)

$\alpha = 0.1, \mu = 21, SE = 0.2, \bar{x} = 20.59$, **two-tailed test**



Calculating the *p*-Value (2 of 6) (Illustration)

- $\alpha = 0.1, \mu = 21, SE = 0.2, \bar{x} = 20.59,$ **left-tailed test**



Calculating the p-Value (3 of 6)

A *small p-value* indicates that the observed value of the test statistic lies far away from the hypothesized value of μ. This presents strong evidence that H_0 is false and should be rejected.

Large p-values indicate that the observed test statistic is not far from the hypothesized mean and does not support rejection of H_0.

Calculating the p-Value (4 of 6)

Definition:

If the p-value is less than or equal to a preassigned significance level α (p-value≤ α), then the null hypothesis can be rejected, and you can report that the results are **statistically significant** at level α.

Example 9.6

The quality control manager wants to know whether the daily yield at a local chemical plant—which has averaged 880 tons for the last several years—has changed in recent months.

A random sample of 50 days gives an average yield of 871 tons with a standard deviation of 21 tons. Calculate the *p*-value for this two-tailed test of hypothesis. Use the *p*-value to draw conclusions regarding the statistical test.

Example 9.6 – Solution (1 of 2)

$\bar{x}$ follows approximately a normal distribution (n>50)

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{871 - 880}{21/\sqrt{50}} = -3.03$$

Since it is two-tailed test, the smallest rejection region that you can use and still reject H_0 is $|z| > 3.03$.

p-value = $P(z > 3.03) + P(z < -3.03)$

$= (1 - .9988) + .0012 = .0024$

NANYANG TECHNOLOGICAL UNIVERSITY

Example 9.6 – Solution (2 of 2)

Notice that the two-tailed p-value is actually twice the tail area corresponding to the calculated value of the test statistic.

If this p-value = .0024 is less than or equal to the preassigned level of significance α, H_0 can be rejected. For this test, you can reject H_0 at either the 1% or the 5% level of significance.

Calculating the *p*-Value (5 of 6)

Use the following "sliding scale" to classify the results.

- If the *p*-value is less than .01, H_0 is rejected. The results are **highly significant.**
- If the *p*-value is between .01 and .05, H_0 is rejected. The results are **statistically significant.**
- If the *p*-value is between .05 and .10, H_0 is usually not rejected. The results are only **tending toward statistical significance.**
- If the *p*-value is greater than .10, H_0 is not rejected. The results are **not statistically significant.**

Example 9.7

Standards set by government agencies indicate that Americans should not exceed an average daily sodium intake of 3300 milligrams (mg).

To find out whether Americans are exceeding this limit, a sample of 100 Americans is selected, and the mean and standard deviation of daily sodium intake are found to be 3400 mg and 1100 mg, respectively. Use $\alpha = .05$ to conduct a test of hypothesis.

Example 9.7 – Solution (1 of 3)

The hypotheses to be tested are

$$H_0 : \mu = 3300 \quad \text{versus} \quad H_a : \mu > 3300$$

and the test statistic is

$$z \approx \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{3400 - 3300}{1100/\sqrt{100}} = .91$$

Example 9.7 – Solution (2 of 3)

The two approaches developed in this section yield the same conclusions.

- **The critical value approach:** Since the significance level is α = .05 and the test is one-tailed, the rejection region is determined by a critical value with tail area equal to α = .05; that is, H_0 can be rejected if z > 1.645

Since z = .91 is not greater than the critical value, H_0 is not rejected.

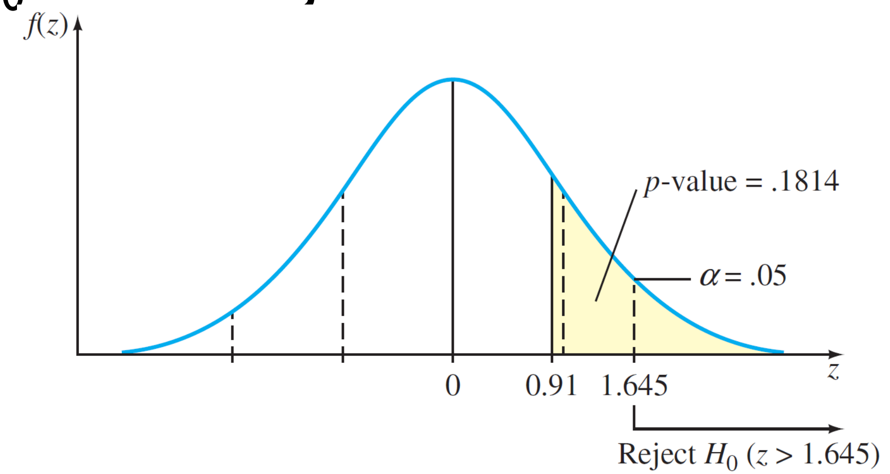


Figure 9.8

Example 9.7 – Solution (3 of 3)

- **The *p*-value approach:** Calculate the *p*-value, the probability that z is greater than or equal to $z = .91$:

  $p\text{-value} = P(z > .91) = 1 - .8186 = .1814$

The null hypothesis can be rejected only if the *p-value is less than or equal to the specified* 5% *significance level.* Therefore, H_0 is not rejected and the results are *not statistically significant.*

There is not enough evidence to indicate that the average daily sodium intake exceeds 3300 mg.

Calculating the *p*-Value (6 of 6)

The *p*-value approach does have two advantages:

- Statistical output from computer software packages usually reports the *p*-value of the test.
- Based on the *p*-value, your test results can be evaluated using any significance level you wish to use. Many researchers report the smallest possible significance level for which their results are *statistically significant.*

Two Types of Errors

Two Types of Errors (1 of 2)

For a statistical test, these two types of errors are defined as **Type Ⅰ** and **Type Ⅱ** errors, shown in Table 9.1(b).

(b) Statistical Test of Hypothesis		
	Null Hypothesis	
Decision	**True**	**False**
Reject H_0	Type I Error	Correct
Accept H_0	Correct	Type II Error

Decision Table

Table 9.1

Type I error: reject H_0 when it is really true (α).

Type II error: accept H_0 when it is really false

Two Types of Errors (4 of 4)

Definition:

A **Type Ⅰ error** for a statistical test happens if you reject the null hypothesis when it is true. The probability of making a Type Ⅰ error is denoted by the symbol α.

A **Type Ⅱ error** for a statistical test happens if you accept the null hypothesis when it is false and some alternative hypothesis is true. The probability of making a Type Ⅱ error is denoted by the symbol β.

The Power of a Statistical Test

The Power of a Statistical Test (1 of 1)

Definition:

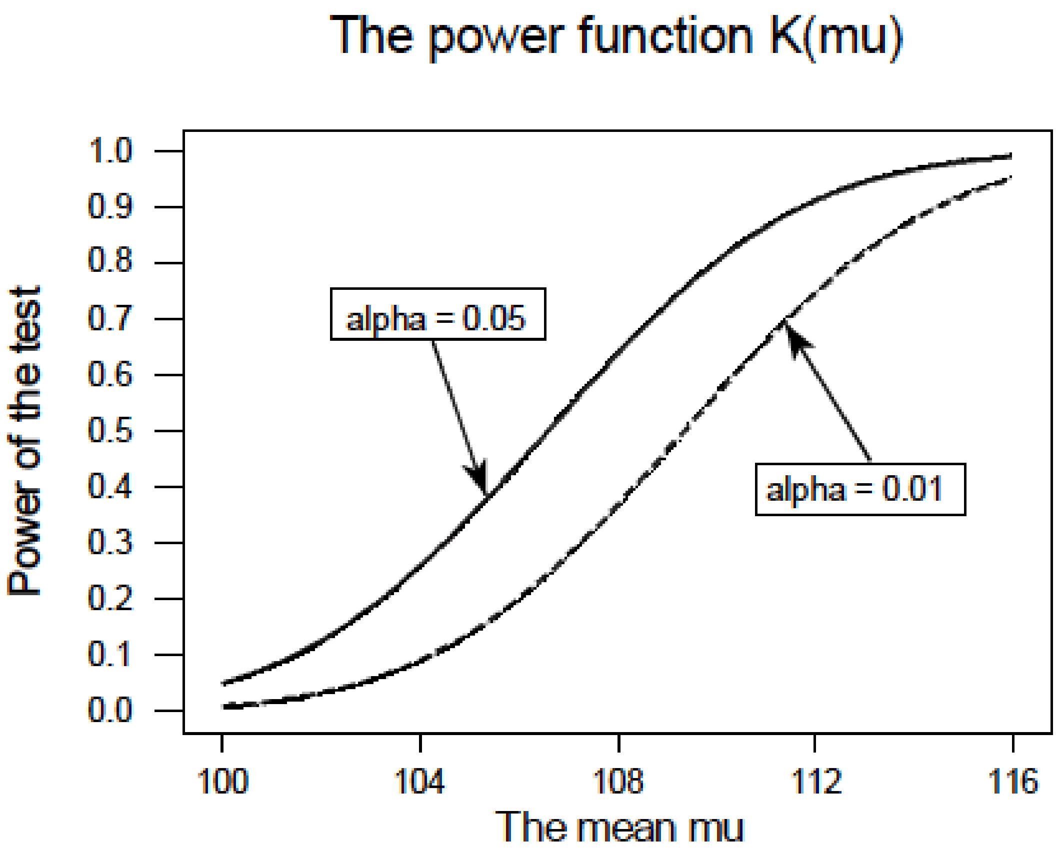
The **power of a statistical test**, given as

$1 - \beta = P(\text{reject } H_0 \text{ when } H_a \text{ is true})$

measures the ability of the test to perform as required.

A graph of $(1 - \beta)$, the probability of rejecting H_0 when in fact H_0 is false, as a function of the true value of the parameter of interest is called the **power curve** for the statistical test. Ideally, you would like α to be small and the *power* $(1 - \beta)$ to be large.

An Example of Power Curve



https://online.stat.psu.edu/stat415/lesson/25/25.2

Example 9.8

The daily yield for a local chemical plant has averaged 880 tons for the last several years. The quality control manager would like to know whether this average has changed in recent months. She randomly selects 50 days from the computer database and calculates the average and standard deviation of the $n = 50$ yields as $\bar{x} = 871$ tons and $s = 21$ tons, respectively.

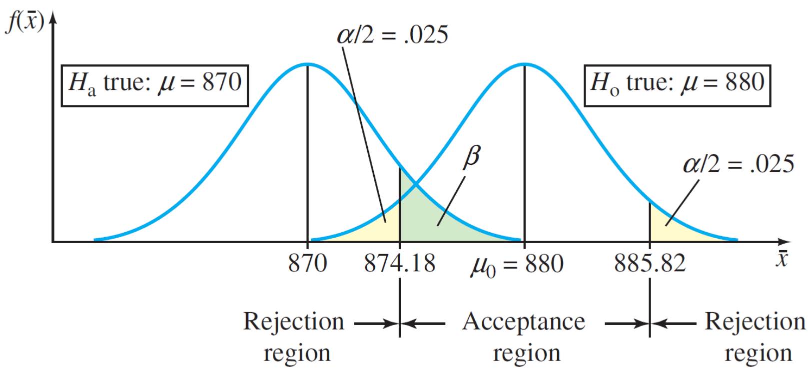
Calculate β and the power of the test $(1 - \beta)$ when μ is actually equal to 870 tons.

Example 9.8 – Solution (2 of 5)

This implies that the acceptance region is

$$-1.96 < \frac{\bar{x} - 880}{21\sqrt{50}} < 1.96 \quad \text{or} \quad 874.18 < \bar{x} < 885.82$$

shown along the horizontal axis in Figure 9.10.



Example 9.8 – Solution (3 of 5)

When H_0 is false and μ = 870, the sampling distribution of $\bar{x}$ is actually represented by the left-hand curve in Figure 9.10, a normal distribution with μ = 870 and $SE = 21/\sqrt{50} = 2.97.$

Then β, the probability of accepting H_0 when μ = 870, is the area under the left-hand normal curve located between 874.18 and 885.82 (see Figure 9.10).

Example 9.8 – Solution (4 of 5)

Calculating the z-values corresponding to 874.18 and 885.82, you get

$$z_1 \approx \frac{\overline{x}-\mu}{s/\sqrt{n}} = \frac{874.18-870}{21/\sqrt{50}} = 1.41$$

$$z_2 \approx \frac{\overline{x}-\mu}{s/\sqrt{n}} = \frac{885.82-870}{21/\sqrt{50}} = 5.33$$

Then

$$\beta = P(\text{accept } H_0 \text{ when } \mu = 870) = P(874.18 < \overline{x} < 885.82 \text{ when } \mu = 870)$$
$$= P(1.41 < z < 5.33)$$

Example 9.8 – Solution (5 of 5)

You can see from Figure 9.10 that the area under the normal curve with μ = 870 above $\bar{x} = 885.82$ (or $z = 5.33$) is negligible. Therefore,

$$\beta = P(z > 1.41)$$

From the normal distribution you can find

$$\beta = 1 - .9207 = .0793$$

Hence, the power of the test is $1 - \beta = 1 - .0793 = .9207$

The probability of correctly rejecting H_0, given that μ is really equal to 870, is .9207, or approximately 92 chances in 100.