Data Analytics

ECON 1008, Semester 1, 2019

Giulio Zanella
University of Adelaide
School of Economics

Chapter 12

Hypothesis testing

Introduction

Hypothesis testing: to determine whether there is enough statistical evidence in favour of a certain belief about an unknown population parameter.

It’s related to the construction of confidence intervals, which we studied last week.

Hypothesis testing is an important tool in statistical inference and has important **social and economic applications.**

Introduction…

Examples

1. In a criminal trial, a jury must decide whether the defendant is innocent or guilty based on the evidence presented at the court.

Introduction…

Examples

1. In a criminal trial, a jury must decide whether the defendant is innocent or guilty based on the evidence presented at the court.
2. Is there statistical evidence in a random sample of potential customers, that supports the hypothesis that more than 20% of potential customers will purchase a new product?

Introduction…

Examples

1. In a criminal trial, a jury must decide whether the defendant is innocent or guilty based on the evidence presented at the court.
2. Is there statistical evidence in a random sample of potential customers, that supports the hypothesis that more than 20% of potential customers will purchase a new product?
3. Is a new drug effective in curing a certain disease? A sample of patients is randomly selected. Half of them are given the drug, and the other half a placebo. The improvement in the patients' conditions is then measured and compared.

Key concepts of hypothesis testing

Five components of a hypothesis test

1. Null Hypothesis (H_0)
2. Alternative Hypothesis (H_A)
3. Test statistic
4. Rejection region
5. Decision rule

In a criminal trial

In a criminal trial, a jury must decide whether the defendant is innocent or guilty based on the evidence presented at the court.

The null hypothesis H_0 is

H_0: The defendant is innocent.

The alternative hypothesis H_A is

H_A: The defendant is guilty.

In a criminal trial…

The jury does not know which hypothesis is true. They must make a decision on the basis of evidence presented.

Statistically, convicting the defendant is called

rejecting the null hypothesis (the defendant is innocent) in favor of the alternative (the defendant is guilty).

That is, the jury is saying that there is enough evidence to conclude that the defendant is guilty (i.e., there is enough evidence to support the alternative hypothesis).

In a criminal trial…

If the jury acquits it is stating that

there is not enough evidence to support the alternative hypothesis.

Notice: the jury is not saying that the defendant is innocent, only that there is not enough evidence to support the alternative hypothesis.

We never say that ‘we accept the null hypothesis’ (that the defendant is innocent), we can only reject it.

Four possible outcomes

Figure 12.1 Results of a test of hypothesis

	H_0 is true	H_0 is false
Do not reject H_0	Correct decision	Type II error P(Type II error) = β
Reject H_0	Type I error P(Type I error) = α	Correct decision

Four possible outcomes	Decision
1. Do not reject H_0 when H_0 is true	Correct
2. Reject H_0 when H_0 is true	Incorrect
3. Do not reject H_0 when H_0 is false	Incorrect
4. Reject H_0 when H_0 is false	Correct

Type I and Type II errors

Two possible errors can be made in any test.

A Type I error occurs when we reject a true null hypothesis (i.e. reject H_0 when H_0 is true). In the criminal trial, a Type I error occurs when the jury convicts an innocent person.

A Type II error occurs when we don't reject a false null hypothesis (i.e. do not reject H_0 when H_0 is false). In a criminal trial, a Type II error occurs when a guilty defendant is acquitted.

Type I and Type II errors…

The probability of a Type I error is denoted as α (Greek letter *alpha*). The probability of a Type II error is β (Greek letter *beta*).

P (making Type I error) = α

P (making Type II error) = β

α is called the level of significance.

The two probabilities are inversely related. Decreasing one increases the other (effect of being more conservative in judicial decisions)

In a criminal trial…

In our judicial system, Type I errors are regarded as more serious. We try to avoid convicting innocent people. Therefore, we are more willing to acquit a guilty person.

We arrange to make α small by requiring the prosecution to prove its case and instructing the jury to find the defendant guilty only if there is ‘evidence beyond a reasonable doubt’.

Now let’s move byond this example, to the general case

Concepts of hypothesis testing…

Null and alternative hypotheses

There are **two** hypotheses. One is called the ***null hypothesis*** and the other the ***alternative hypothesis***. The usual notation is:

pronounced
H ‘nought’

H_0: — *the ‘null’ hypothesis*

H_A: — *the ‘alternative’ hypothesis*

The null hypothesis (H_0) will always state that the ***parameter equals the value*** specified in the alternative hypothesis (H_A).

Concepts of hypothesis testing…

Null and alternative hypotheses

Test on population means:

H_0: $\mu = \mu_0$ (μ_0 is a given value for μ)

H_A: $\mu \neq \mu_0$ or H_A: $\mu < \mu_0$ or H_A: $\mu > \mu_0$

Concepts of hypothesis testing…

Test statistics

We need to use a sample statistic to test a hypothesis.

Test on population mean, μ:

a) If population variance σ^2 is known

Test statistic: $\bar{X}$; standardised test statistic: $Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$ ~N(0,1)

b) If population variance σ^2 is unknown

Test statistic: $\bar{X}$; standardised test statistic: $t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}}$ ~t_{n-1}

Concepts of hypothesis testing…

A rejection region of a test consists of all values of the test statistic for which H_0 is rejected.

An acceptance region of a test consists of all values of the test statistic for which H_0 is not rejected.

The critical value is the value that separates the acceptance and rejection region.

The decision rule defines the range of values of the test statistic for which H_0 is rejected in favour of H_A.

Concepts of hypothesis testing: A summary

1. There are two hypotheses, the null and the alternative hypotheses.
2. The procedure begins with the assumption that the null hypothesis is true.
3. The goal is to determine whether there is enough evidence to infer that the alternative hypothesis is true.

Concepts of hypothesis testing…

4. There are two possible decisions:
   - Conclude that there is enough evidence to support the alternative hypothesis.
   - Conclude that there is not enough evidence to support the alternative hypothesis.
5. Two possible errors can be made.
   - Type I error: Reject a true null hypothesis.
   - Type II error: Do not reject a false null hypothesis.

   P(making a Type I error) = α = level of significance

   P(making a Type II error) = β

Concepts of hypothesis testing…

Consider an example where we want to know **whether the population mean is different from 130 units**. We can rephrase this request into a test of the hypothesis:

H_0: $\mu = 130$

Thus, our alternative hypothesis becomes:

H_A: $\mu \neq 130$

This is what we are interested in determining…

Concepts of hypothesis testing…

The testing procedure begins with the ***assumption that the null hypothesis is true***.

Thus, until we have further statistical evidence, we will **assume**:

H_0: $\mu = 130$ (assumed to be TRUE)

Concepts of hypothesis testing…

The **goal** of the process is to determine ***whether there is enough evidence*** to infer that the alternative hypothesis is true.

That is, is there sufficient statistical information to determine if this statement is true?

H_A: $\mu \neq 130$

This is what we are interested in determining…

Concepts of hypothesis testing…

There are **two** possible decisions that can be made:

Conclude that there ***is enough evidence*** to support the alternative hypothesis (also stated as: rejecting the null hypothesis in favour of the alternative).

Conclude that there ***is not enough evidence*** to support the alternative hypothesis (also stated as: **not** rejecting the null hypothesis in favour of the alternative).

NOTE: we **do not** say that we **accept** the null hypothesis!

Concepts of hypothesis testing…

Once the null and alternative hypotheses are stated, the next step is to select a random sample from the population and calculate a ***test statistic*** (in this example, the sample mean).

If the test statistic's value is inconsistent with the null hypothesis ***we reject the null hypothesis*** and ***infer that the alternative hypothesis is true***.

12.2 Testing the population mean when the variance σ^2 is known

For example, if we're trying to decide whether the mean is not equal to 130, if the selected sample gives a large value of $\bar{X}$ (say, 300) that would provide enough evidence to reject H_0: $\mu = 130$ and to conclude there is some evidence to support H_A: $\mu \neq 130$, μ is different from 130.

On the other hand, if the selected sample gives a value for $\bar{X}$ close to 130 (say, 132) we could not say that this provides a great deal of evidence to infer that the population mean is different from 130.

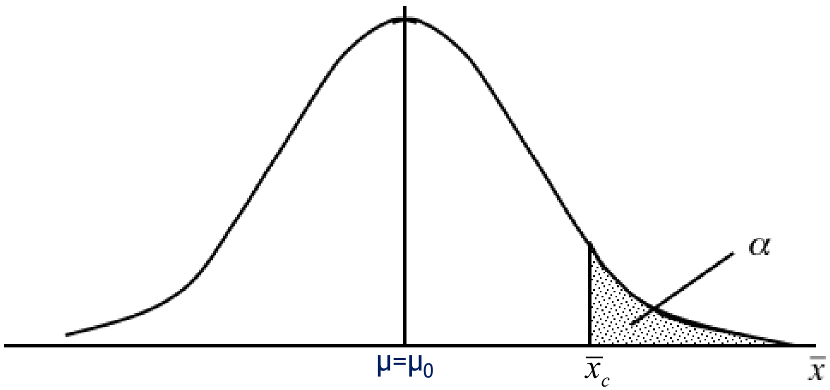
Right-tail test

Hypothesis to test:

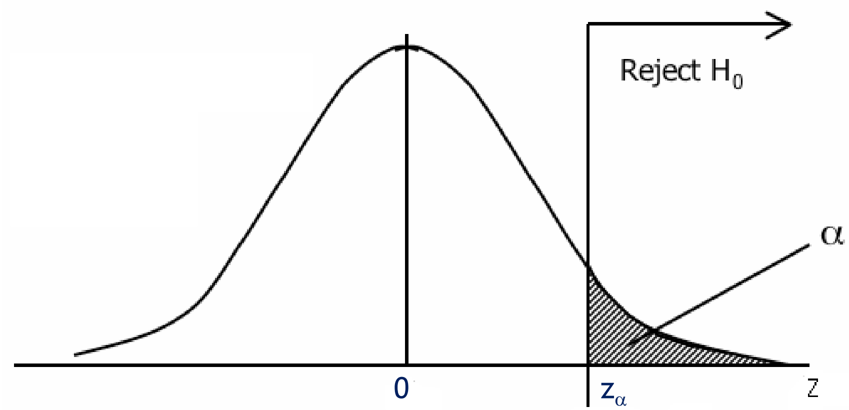
H_0: $\mu=\mu_0$

H_A: $\mu>\mu_0$

Level of significance = α



Rejection region: $\bar{x} > \bar{x}_c$



If the selected sample gives a sample mean value very much larger than μ_0, then we should reject H_0 for larger values of $\bar{X}$. That is, the rejection region in this case will be on the right tail of the sampling distribution of the test statistic. Since P(reject H_0|H_0 is true) = α, the area of the whole right tail will be equal to α.

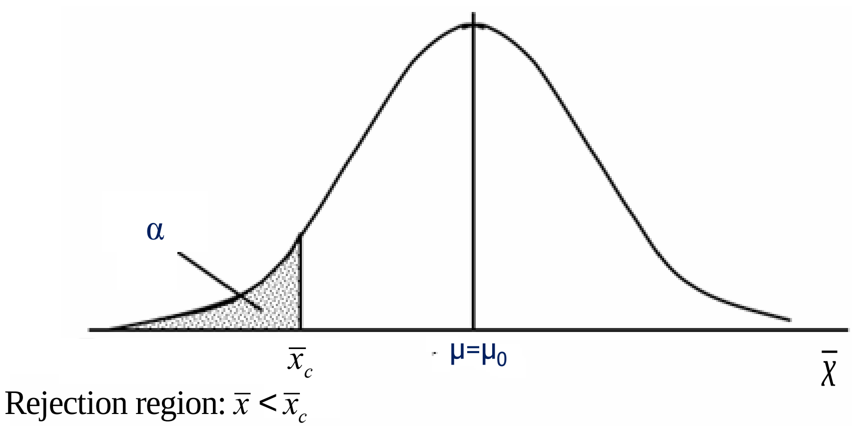
Left-tail test

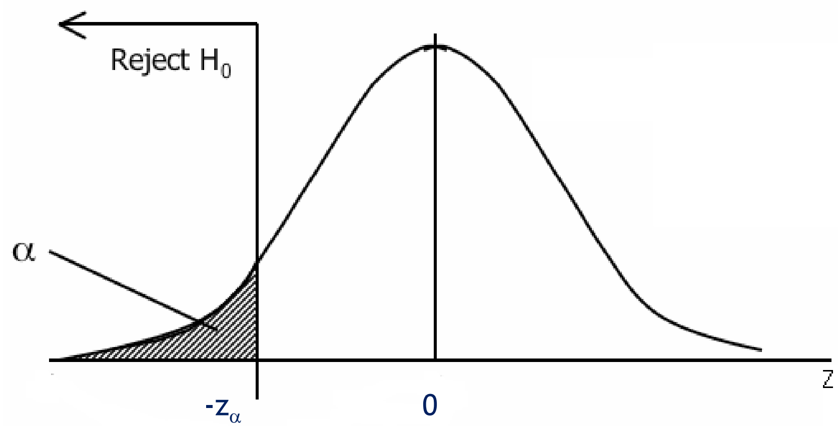
Hypothesis to test:

H_0: $\mu=\mu_0$

H_A: $\mu<\mu_0$

Level of significance = α

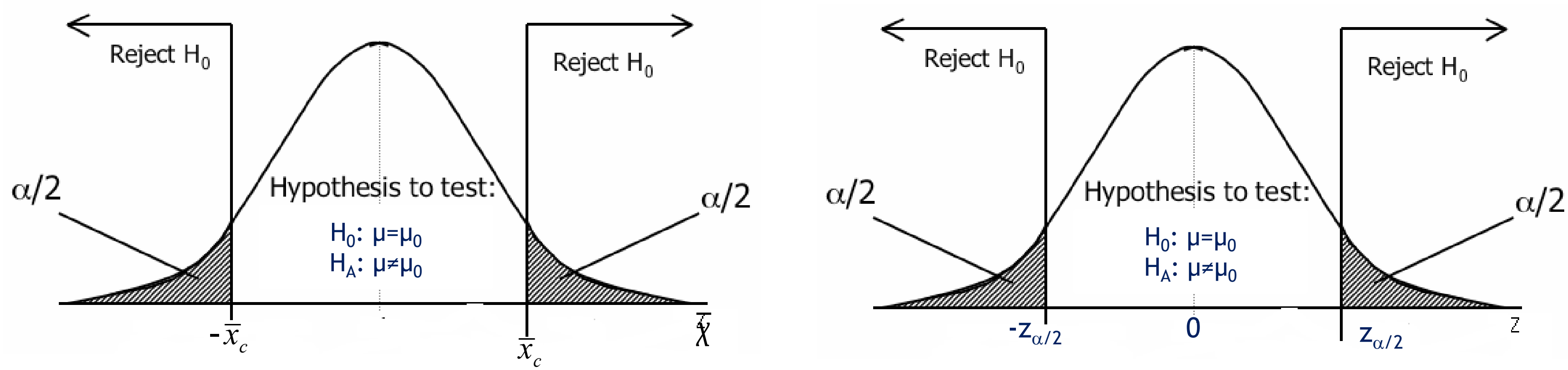




If the selected sample gives a sample mean value very much smaller than μ_0, then we should reject H_0 for smaller values of $\bar{X}$. That is, the rejection region in this case will be on the left tail of the sampling distribution of the test statistic. Since P(reject $H_0|H_0$ is true) = α, the area of the whole left tail will be equal to α.

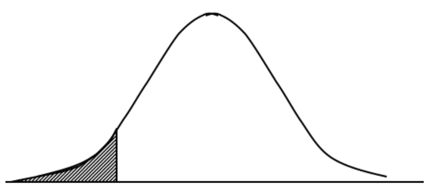
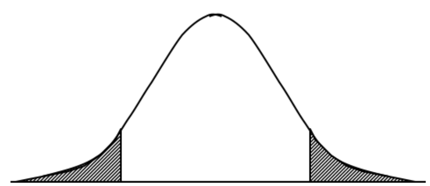
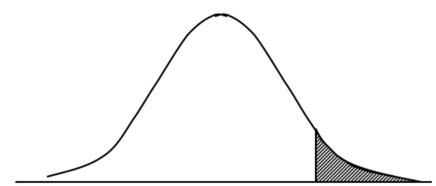
Two-tail test

Hypothesis to test:
H_0: $\mu=\mu_0$
H_A: $\mu\neq\mu_0$
Level of significance = α



If the selected sample gives a sample mean value either very much larger or very much smaller than μ_0, then we should reject H_0 for either larger or smaller values of $\bar{X}$. That is, the rejection region in this case will be both tails of the sampling distribution of the test statistic. Since P(reject $H_0|H_0$ is true) = α and the distribution is symmetric, the area of each tail will be equal to $\alpha/2$.

One- and two-tail tests: A summary

One-Tail Test (left tail)	Two-Tail Test	One-Tail Test (right tail)
H_o: $\mu = \mu_o$ H_A: $\mu < \mu_o$	H_o: $\mu = \mu_o$ H_A: $\mu \neq \mu_o$	H_o: $\mu = \mu_o$ H_A: $\mu > \mu_o$
		

Six-step process for testing hypothesis

Step 1: Set up the null and alternative hypotheses.

Note: Since the alternative hypothesis answers the question, set this one up first. The null hypothesis will automatically follow.

Step 2: Determine the test statistic and its sampling distribution.

Step 3: Specify the significance level.

Note: We usually set $\alpha = 0.01$*, 0.05 or 0.10, but other values are possible.*

Six-step process for testing hypothesis…

Step 4: Define the decision rule.

Note: This involves using the appropriate statistical table from Appendix B to determine the critical value(s) and the rejection region.

Step 5: Calculate the value of the test statistic.

Note: Non-mathematicians need not fear. Only simple arithmetic is needed.

Step 6: Make a decision and answer the question.

Note: Remember to answer the original question. Making a decision about the null hypothesis is not enough.

Example 1: Time required to complete an assembly line

The mean of the amount of time required to complete a critical part of a production process on an assembly line is believed to be 130 seconds. To test if this belief is correct, a sample of 100 randomly selected assemblies is drawn and the processing time recorded. The sample mean is 126.8 seconds. If the process time is normally distributed with a standard deviation of 15 seconds, can we conclude that the belief regarding the mean is *incorrect*?

Example 1: Solution

1. Hypotheses: $H_0: \mu = 130$
   $H_A: \mu \neq 130$ Two-tailed test

2. Test statistic: $\bar{X}$ or standardised test statistic:

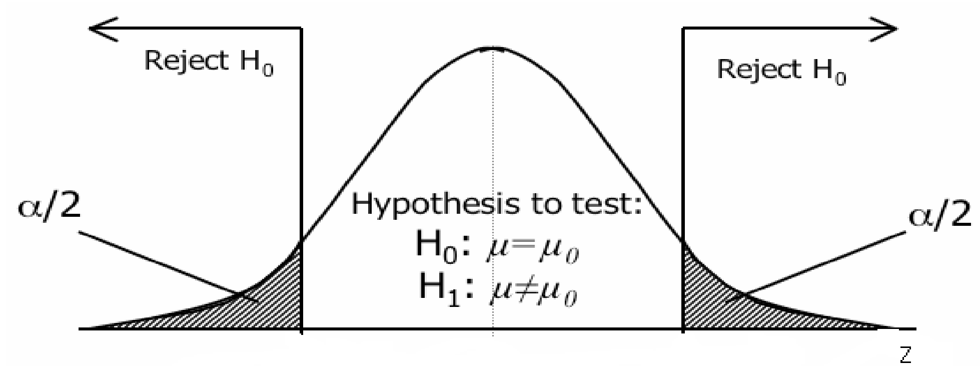
$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim \text{Normal}(0,1)$$

3. Level of significance: $\alpha=0.05$

Example 1: Solution…

4. Decision rule

The rejection region is set up so we can reject the null hypothesis when the test statistic is large ***or*** when it is small.

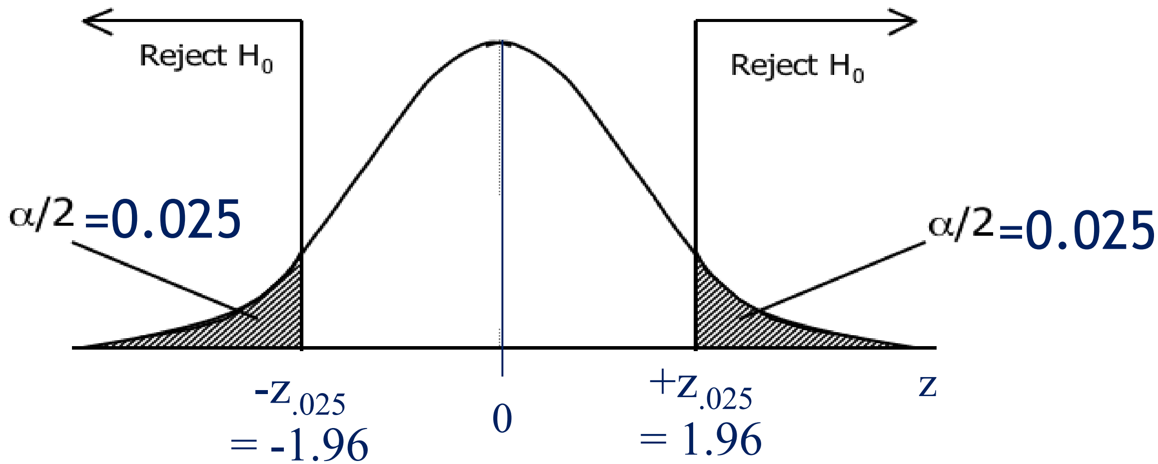


Stat is 'very small' Stat is 'very large'

That is, we set up a two-tail rejection region. The total area in the rejection region must sum to α, so we divide this probability by 2.

Example 1: Solution…

As α = 0.05, we have $\alpha/2$ = 0.025 and $z_{0.025}$ = 1.96 and our rejection region is $z < -1.96$ **or** $z > 1.96$.



Decision rule (z-value method):
Do not reject H_0 if $-1.96 \le z \le 1.96$. Otherwise, reject H_0.

Example 1: Solution…

5. Value of the test statistic

   From the sample data, $\bar{X}$ = 126.8.

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{126.8 - 130}{15 / \sqrt{100}} = -2.13$$

6. Conclusion

   Since z = -2.13 < -1.96, we reject H_o.

12.3 The p-value of a test of hypothesis

The p-value of a test is the minimum level of significance that is required to reject the null hypothesis.

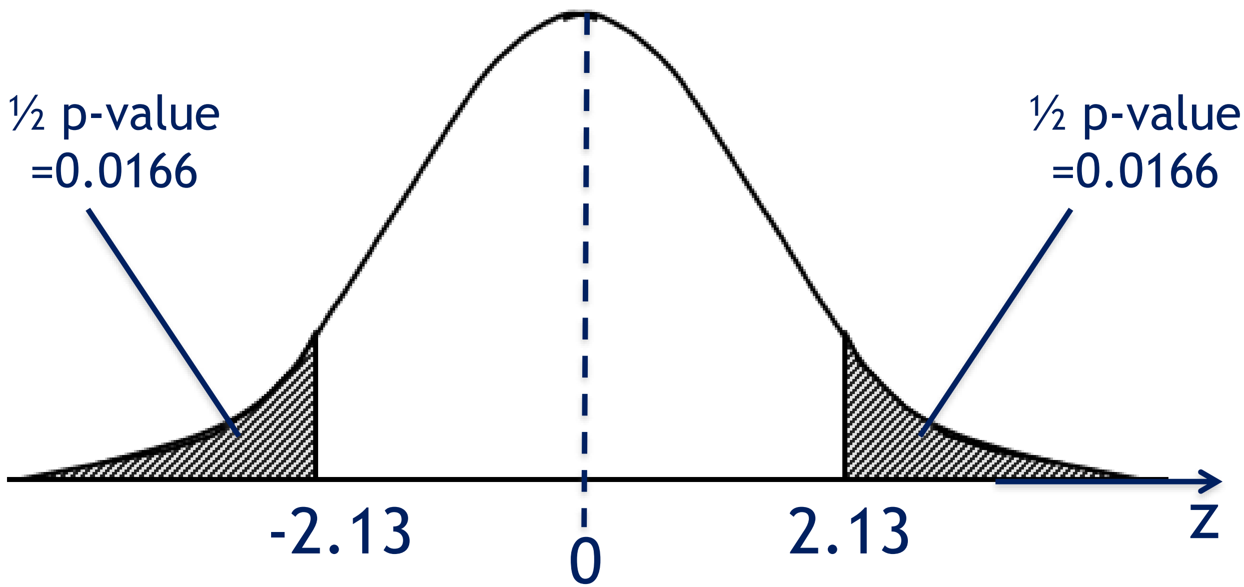
In Example 1, we obtained

$$z = -2.13$$

Since it was a two-tail test,

$$\begin{aligned} \text{p-value} &= P(Z < -2.13) + P(Z > 2.13) \\ &= 2(0.0166) = 0.0332 \end{aligned}$$

The p-value of a test of hypothesis…



The p-value of a test of hypothesis…

Therefore, if we select a level of significance larger than 0.0332 (3.32%), the conclusion would be to reject H_0.

For example, if $\alpha = 0.05$ (5%), *p*-value = $0.0332 < \alpha = 0.05$, so the conclusion is to reject H_0 in favour of H_A: $\mu \neq 130$.

On the other hand, if we choose $\alpha = 0.03$ (3%), *p*-value = $0.0332 > \alpha = 0.03$, so the conclusion would be do not reject H_0.

The p-value of a test of hypothesis…

<u>*p*-value calculation</u>

In general:

For a right tail test (H_A: $\mu > \mu_0$), *p*-value = $P(z > z_0)$

For a left tail test (H_A: $\mu < \mu_0$), *p*-value = $P(z < -z_0)$

For a two-sided test (H_A: $\mu \neq \mu_0$), *p*-value = $2P(z > |z_0|)$

<u>Decision rule</u>

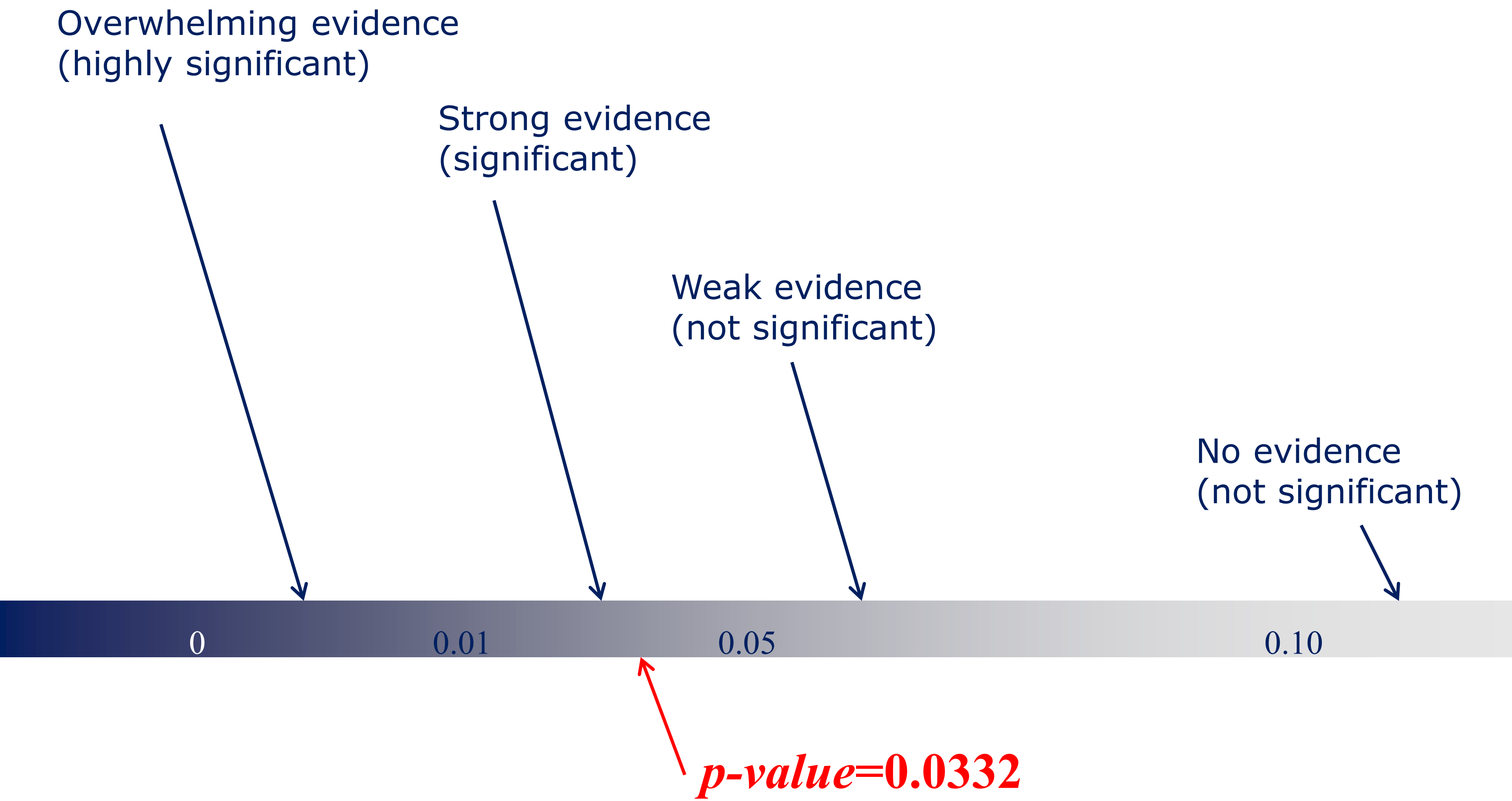
If *p*-value < α, we reject H_0.

If *p*-value > α, we do not reject H_0.

Interpreting the *p*-value

- The smaller the p-value, the more statistical evidence exists to support the alternative hypothesis.
- If the p-value is less than 1%, there is ***overwhelming evidence*** that supports the alternative hypothesis.
- If the p-value is between 1% and 5%, there is ***strong evidence*** that supports the alternative hypothesis.
- If the p-value is between 5% and 10% there is ***weak evidence*** that supports the alternative hypothesis.
- If the p-value exceeds 10%, there is ***no evidence*** that supports the alternative hypothesis.
- ***In Example 1, we observed a p-value of 0.0332, hence there is strong evidence to support H_A: $\mu \neq 130$.***

Interpreting the *p*-value…



Interpreting the *p*-value…

Compare the *p*-value with the selected value of the significance level:

If the *p*-value is less than α, we judge the p-value to be small enough to reject the null hypothesis.

If the *p*-value is greater than α, we do not reject the null hypothesis.

In Example 1, since p-value = 0.0332 $<$ α = .05, we reject H_0 in favour of H_A at the 5 percent level of significance.

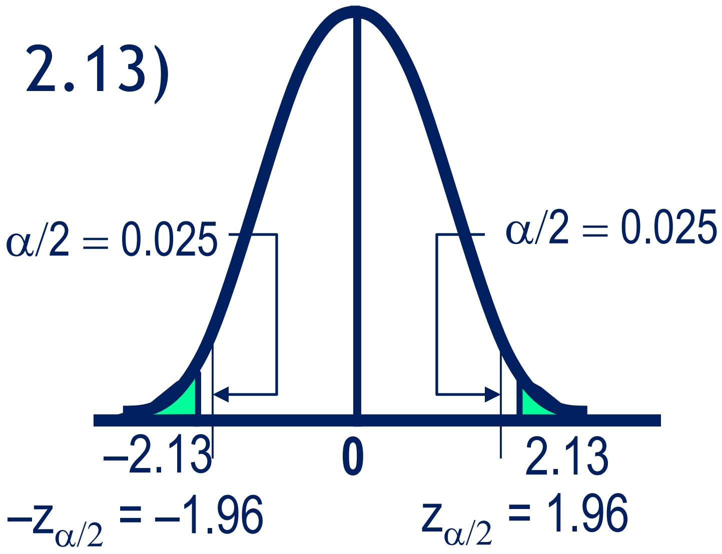
Example 1: Solution…

COMPUTE

In summary, in Example 1, given α=0.05,

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{126.8 - 130}{15/\sqrt{100}} = -2.13$$

p-value = P(Z < -2.13) + P(Z > 2.13)
= 2(0.0166) = 0.0332



Conclusion: As p-value = 0.0332 < 0.05 = α, we reject H_o.

INTERPRET

Since the value of the test statistic falls in the rejection region, we reject the null hypothesis in favour of the alternative hypothesis.

OR (alternatively)

Since the *p*-value=0.0332 is less than the level of significance $\alpha = 0.05$, we reject the null hypothesis in favour of the alternative hypothesis.

There is sufficient evidence to infer that the mean is not 130.

Example 2: Is the product label acceptable?

XM12-02 There are a variety of government agencies devoted to ensuring that food producers package their products in such a way that the weight or volume of the contents listed on the label is correct. Recently, a number of customer complaints were received regarding the weight of the 500g garlic packs imported from overseas by a particular wholesale distributor, Ausvege Ltd. For example, garlic packs whose labels state that the contents have a net weight of 500g must have a net weight of at least 500g. However, it is impossible to check all garlic packs sold in the country. As a result, statistical techniques are used. A random sample of the product is selected and its contents measured. If the mean of the sample provides sufficient evidence to infer that the mean weight of all 500g

Example 2…

Ausvege garlic packs is less than 500g, the product label is deemed to be unacceptable. Suppose that a government inspector weighs the contents of a random sample of 25 garlic packs labelled ‘Net weight: 500 grams’ distributed by Ausvege Ltd and records the measurements below. Using a 5% significance level, can the inspector conclude that the product label is unacceptable? (Assume that the inspector knows from previous experiments that the weight of all 500g garlic packs distributed by Ausvege Ltd is normally distributed with a standard deviation of 10g.)

Net weight of 25 random ‘500-gram’ garlic packs

500	502	505	498	501	496	504	505	500	499	498	503	510
503	505	499	497	502	504	507	510	495	470	501	480	

Example 2: Solution

Data type: Numerical, single population (σ known)

Problem objective: To draw a conclusion about the mean weights of 500g garlic packs.

We investigate whether the mean weight is *less than 500 grams* (that is, whether the product label is unacceptable).

Then

$H_0: \mu = 500$

$H_A: \mu < 500$

Left one-tail test

Based on the given data, sample mean $\bar{X}$ = 499.76.

Example 2: Solution…

1. Our hypotheses

H_0: $\mu = 500$
H_A: $\mu < 500$

Left one-tail test

2. The test statistic is:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \text{Normal}(0,1)$$

3. Level of significance: $\alpha = 0.05$

4. Decision rule

Reject H_o if $Z < -z_\alpha = -1.645$
(or Reject H_o if p-value $< \alpha = 0.05$)

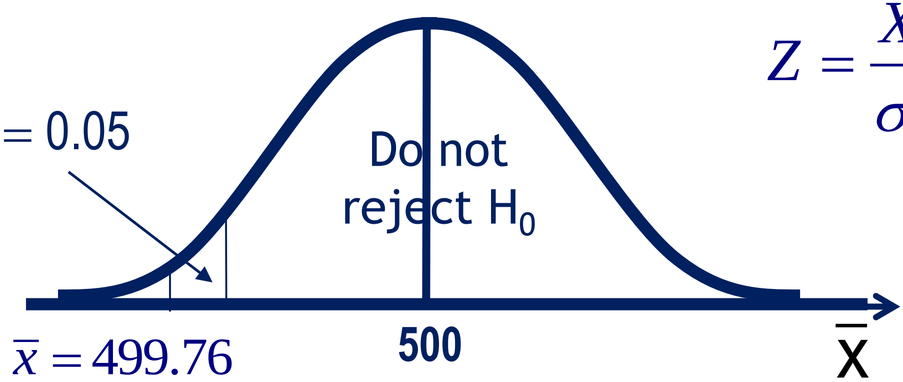
5. Value of the test statistic under H_o:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{499.76 - 500}{10/\sqrt{25}} = -0.12$$

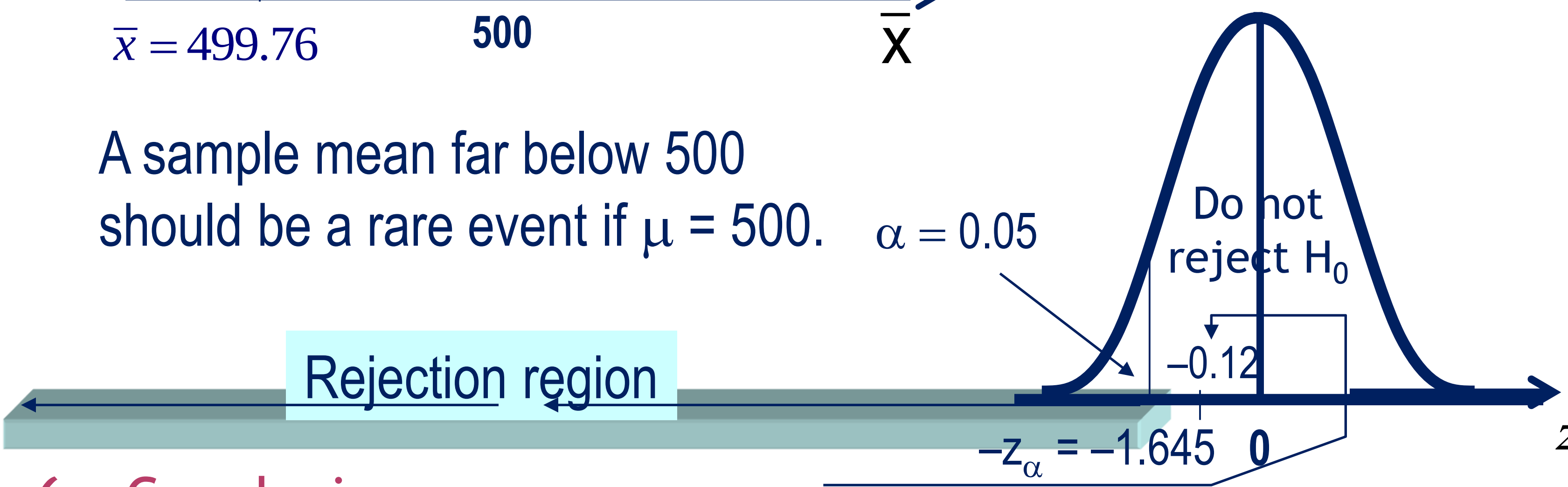
Example 2: Solution…

z-value method:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{499.76 - 500}{10 / \sqrt{25}} = -0.12$$



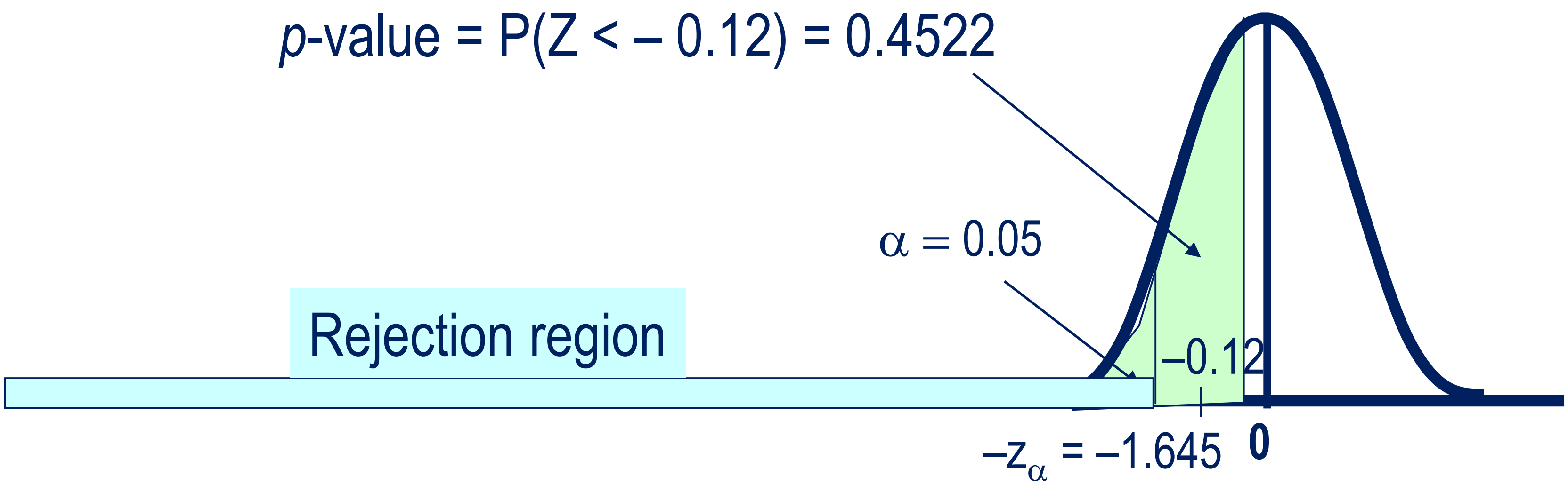
A sample mean far below 500 should be a rare event if μ = 500.



6. Conclusion:

As z = -0.12 > -1.645 = $z_{critical}$, we do not reject H_o.

p-value Method



Conclusion: As p-value = P(Z < – 0.12) = 0.4522 > 0.05, we do not reject H_o.

INTERPRET

Since the value of the test statistic does not fall in the rejection region, we do not reject the null hypothesis in favour of the alternative hypothesis.

OR (alternatively)

Since the p-value is greater than the level of significance, we do not reject the null hypothesis in favour of the alternative hypothesis.

There is *insufficient evidence* to infer that the mean is less than 500 grams.

12.4 Testing the population mean when the variance σ^2 is unknown

Recall that when σ is known, the test statistic

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

is normally distributed

- if the sample is drawn from a normal population, or
- if the population is not normal but the sample is sufficiently large.

When σ is unknown, we estimate σ by its point estimator *s*, and the z test statistic is replaced by the *t*-statistic.

Inference with variance unknown

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \quad \Longrightarrow \quad t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

When σ is unknown, we use its point estimator **s** and the z-statistic is replaced by the t-statistic, where the number of “degrees of freedom” ν, is n–1.

Testing the population mean when the variance σ^2 is unknown…

If the population is normally distributed, the test statistic for μ when σ is unknown is t.

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

This statistic is Student t-distributed with (n–1) degrees of freedom.

Example 3 – Has production declined due to new government regulations?
(Example 12.5, p496)

XM12-05 A manufacturer of television picture tubes has a production line that used to produce an average of 100 tubes per day. because of recently introduced government regulations, a new safety device is installed, which the manufacturer believes will reduce the average daily output. After installation of the safety device, a random sample of 15 days' production was recorded, as follows:

93 103 95 101 91 105 96 94 101 88 98 94 101 92 95

Assuming that the daily output is normally distributed, is there sufficient evidence to allow the manufacturer to conclude that average daily output has decreased following installation of the safety device? (use α = 0.05.)

Example 3: Solution

IDENTIFY

Identifying the technique

Data type: Numerical, single population (σ known)

Problem objective: To describe the population of daily output (X). We investigate whether the mean daily output is *less than 100 tubes* (that is, whether output has decreased following installation of the safety device).

Parameter of interest: Population mean μ

Population variance: σ^2 unknown, use s^2 to estimate σ^2.

Distribution of X: Assume normal.

Example 3: Solution…

IDENTIFY

1. Hypotheses: H_0: μ = 100
   H_A: μ < 100 [Left one tail test]
2. Test statistic: $t = \frac{\bar{X} - \mu}{s / \sqrt{n}} \sim t_{n-1}$
3. Level of significance: $\alpha = 0.05$
4. Critical value: $-t_{\alpha,n-1} = -t_{0.05,14} = -1.761$

   Decision rule:

   Reject H_0 if $t < -1.761$; otherwise do not reject H_0.

CALCULATE

Example 3: Solution…

5. Value of the test statistic:

   n=14, $\bar{X}$ = 96.47, s^2 = 23.55, s = 4.85

$$t = \frac{96.47 - 100}{4.85 / \sqrt{15}} = -2.82$$

6. Conclusion:

   t-value method

   As t = -2.82 < -1.1761, reject H_0.

There is enough evidence to conclude that mean daily production has decreased after the installation of the safety device.

CALCULATE

Example 3: Solution…

6. Conclusion (cont.):

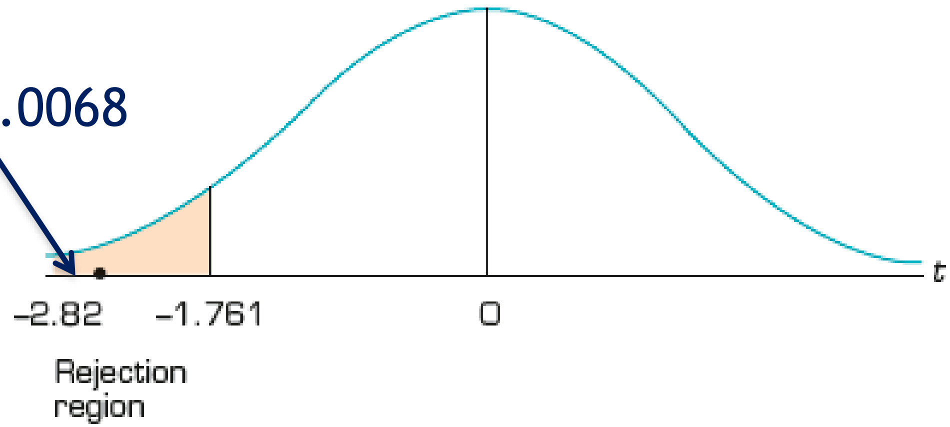
   <u>*p-value method*</u>

   From the Excel output (see below),

   p-value (one tail) = 0.0068

   Since *p*-value = 0.0068 < α = 0.05, we reject H_0.

p-value = 0.0068



INTERPRET

Example 3: Solution…

Interpreting the results

The manufacturer would be advised to look for ways to restore productivity with the safety device in place. Perhaps developing another safety device would help. We note that the results are valid only if the assumption that daily output is normally distributed is true.

Example 3: Solution…

Checking the required condition

As before in Chapter 10, we check the required condition that the population is normal. The histogram of the data below indicates that the population is approximately normal. Hence, the t-test and the conclusion is valid.