Introduction to Statistics and Data Science using *eStat*

Chapter 7 Testing Hypothesis for Single Population

7.1 Testing Hypothesis for a Population Mean

Jung Jin Lee
Professor of Soongsil University, Korea
Visiting Professor of ADA University, Azerbaijan

7.1 Testing Hypothesis for a Population Mean

- Examples of testing hypothesis for a population mean.

  - Capacity of a cookie bag is indicated as 200g. Will there be enough cookies in the indicated capacity?
  - At a light bulb factory, a newly developed light bulb advertises a longer bulb life than the past. Is this propaganda reliable?
  - In this year's academic test, students said that there will be an average English score of 5 points higher than last year. How can you investigate if this is true?

- **Testing hypothesis** is to decide which hypothesis is to use for the two hypotheses about the unknown population parameter using the sample.
  - test of population mean
  - test of population variance
  - test of population proportion

7.1 Testing Hypothesis for a Population Mean

[Ex 7.1.1] At a light bulb factory, the average life expectancy of a light bulb made by a conventional production method is known to be 1,500 hours and the standard deviation is 200 hours.

- Recently, the company is trying to introduce a new production method, with the average life expectancy of 1,600 hours for light bulbs.
- To confirm this argument, 30 samples were taken by simple random sampling and the sample mean was 1555 hours. Can you tell me that the new type of light bulb has a life of 1600 hours?

<Answer>

- A statistical approach to the question of this issue is first to make two assumptions about the different arguments for the population mean μ.

$$H_0 : \mu = 1500, \quad H_1 : \mu = 1600$$

- H_0 is called a null hypothesis and H_1 is an alternative hypothesis

7.1 Testing Hypothesis for a Population Mean

<Ex 7.1.1 Answer>

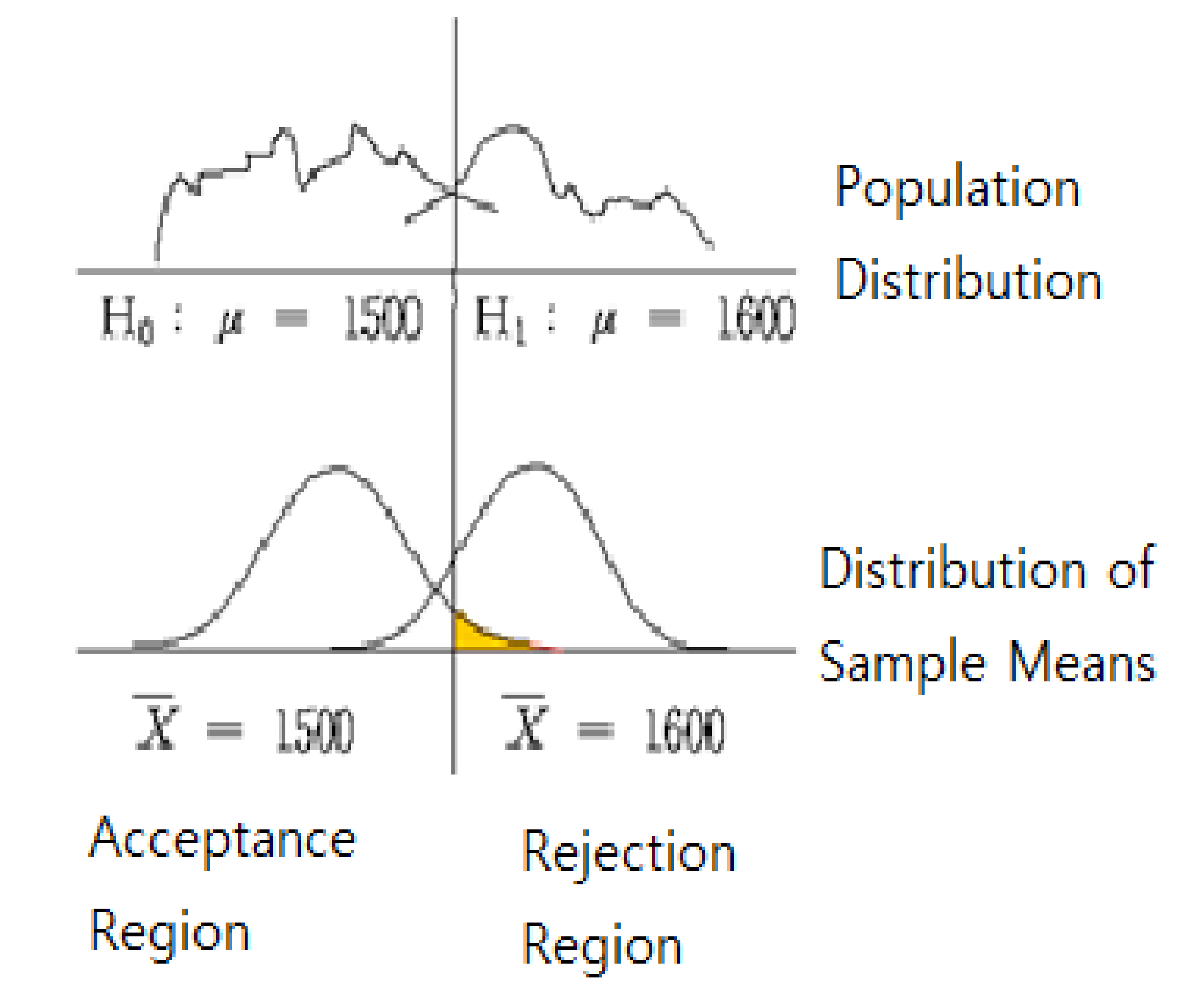
- Null hypothesis : ‘existing known fact’,
  Alternative hypothesis : ‘new facts or changes in current belief’.
  ⇨ 'unless there is a significant reason, we keep null hypothesis
  ⇨ ‘conservative decision making’
- Common sense criterion to choose hypothesis
  ⇨ 'which population mean of two hypothesis is closer to sample mean’.
  ⇨ $\overline{X}$ = 1555 is closer to H_1 : μ = 1600 so H_1 will be chosen.

- Testing hypothesis is based on the sampling distribution of $\overline{X}$.
  ⇨ select a critical value C based on the sampling distribution and
  make a decision rule:

‘If $\overline{X}$ is smaller than C, then accept H_0 will be chosen, else reject H_0’

7.1 Testing Hypothesis for a Population Mean

<Ex 7.1.1 Answer>



7.1 Testing Hypothesis for a Population Mean

<Ex 7.1.1 Answer>

Table 7.1.1 Two types of errors in testing hypothesis

	Actual	
	H_0 is true	H_1 is true
Decision: H_0 is true	Correct	Type 2 Error
H_1 is true	Type 1 Error	Correct

7.1 Testing Hypothesis for a Population Mean

<Ex 7.1.1 Answer>

- If H_0 : μ = 1500 is true, sampling distribution of $\overline{X} \sim N(1500, \frac{200^2}{30})$.
- If H_1 : μ = 1600 is true, sampling distribution of $\overline{X} \sim N(1600, \frac{200^2}{30})$.
- Decision rule becomes as follows:
  'If $\overline{X}$ < C, then accept H_0, else accept H_1 (i.e. reject H_0)'

- If we set the significance level is 5%, C can be calculated by finding the percentile of $N(1500, \frac{200^2}{30})$

$$C = 1500 + 1.645 \sqrt{\frac{200^2}{30}} = 1560.06$$

- Decision rule can be written as follows:
  'If $\overline{X}$ < 1560.06, then accept H_0, else reject H_0'

7.1 Testing Hypothesis for a Population Mean

<Ex 7.1.1 Answer>

- The decision rule is often written.

  ‘If $\frac{\overline{X} - 1500}{\frac{200}{\sqrt{30}}}$ < 1.645, then accept H_0, else reject H_0’

- Since $\overline{X}$ = 1555, and $\frac{1555 - 1500}{\frac{200}{\sqrt{30}}}$ = 1.506, it is less than 1.645.

  ⇨ accept H_0.

- p-value is probability of type 1 error when the observed sample mean value is considered as the critical value for decision
  ⇨ p-value indicates where the observed sample mean is located among all possible sample means

7.1 Testing Hypothesis for a Population Mean

Table 7.1.2 Testing hypothesis for population mean - known σ case

Type of Hypothesis	Decision Rule
1) $H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	If $\dfrac{\overline{X} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}} > z_\alpha$, then reject H_0
2) $H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	If $\dfrac{\overline{X} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}} < -z_\alpha$, then reject H_0
3) $H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	If $\left\lvert \dfrac{\overline{X} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}} \right\rvert > z_{\alpha/2}$, then reject H_0

Note: The H_0 of 1) can be written as $H_0 : \mu \leq \mu_0$, 2) as $H_0 : \mu \geq \mu_0$

7.1 Testing Hypothesis for a Population Mean

- **Decision rule using p-value**
  'If p-value is less than the significance level, then reject H_0 else accept H_0'

Table 7.1.3 Calculation of p-value

Type of Hypothesis	p-value
1) $H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	$P(\overline{X} > \overline{x}_{obs})$
2) $H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$P(\overline{X} < \overline{x}_{obs})$
3) $H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	$If\ \overline{X}_{obs} > \mu_o\ ,\ 2P(\overline{X} > \overline{x}_{obs})\quad else\ 2P(\overline{X} < \overline{X}_{obs})$

Note : $\overline{x}_{obs}$ is the observed sample mean.

7.1 Testing Hypothesis for a Population Mean

Table 7.1.4 Testing hypothesis for population mean - unknown σ case (population is a normal distribution)

Type of Hypothesis	Decision Rule
1) $H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	$If \ \dfrac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}} > t_{n-1;\alpha}$, $reject\ H_0$
2) $H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$If \ \dfrac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}} < -t_{n-1;\alpha}$, $reject\ H_0$
3) $H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	$If \ \left\lvert \dfrac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}} \right\rvert > t_{n-1;\alpha/2}$, $reject\ H_0$

Note: The H_0 of 1) can be written as $H_0 : \mu \leq \mu_0$, 2) as $H_0 : \mu \geq \mu_0$

7.1 Testing Hypothesis for a Population Mean

[Example 7.1.2] The weight of a bag of cookies is supposed to be 250 grams. Suppose the weight of all bags of cookies is a normal distribution.

- In the survey of 100 samples of bags which were randomly selected, the sample mean was 253g and the standard deviation was 10 grams.

1) Test hypothesis whether the weight of the bag of cookies is 250g or larger and find the p-value. α = 1%

2) Test hypothesis whether or not the weight of the bag of cookies is 250g and find the p-value. α = 1%

3) Use『eStatU』to test the hypothesis above.

7.1 Testing Hypothesis for a Population Mean

<Answer of Example 7.1.2>

1) Hypothesis is $H_0 : \mu = 250$, $H_1 : \mu > 250$.

- Since sample size is large (n =100), we can use Z distribution instead of t distribution. Decision rule is as follows:

  'If $\frac{\overline{X} - \mu_0}{\frac{s}{\sqrt{n}}} > z_\alpha$, then reject H_0'

  'If $\frac{253 - 250}{\frac{10}{\sqrt{100}}} > z_\alpha$, then reject H_0'

  Since (253-250) / (10/10) = 3 and $z_{0.01}$ = 2.326, H_0 is rejected.

- Since p-value is the probability of Type 1 error when $\overline{X}$ is critical value. the probability of P($\overline{X}$ > 253).

  $\overline{X}$ is $N(250, \frac{100}{100})$ when $H_0 : \mu = 250$ is true

  p-value = P($\overline{X}$ > 253 | H_0 is true) = P(Z > $\frac{253-250}{\frac{10}{10}}$) = P(Z > 3) = 0.0013

7.1 Testing Hypothesis for a Population Mean

<Answer of Example 7.1.2>

2) Hypothesis is $H_0 : \mu = 250$, $H_1 : \mu \neq 250$.

- Since n is large, we can use the Z distribution instead of t distribution. Decision rule is as follows:

  'If $\left| \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} \right| > z_{\frac{\alpha}{2}}$, then reject H_0'

  'If $\left| \frac{253 - 250}{\frac{10}{\sqrt{100}}} \right| > z_{0.005}$, then reject H_0'

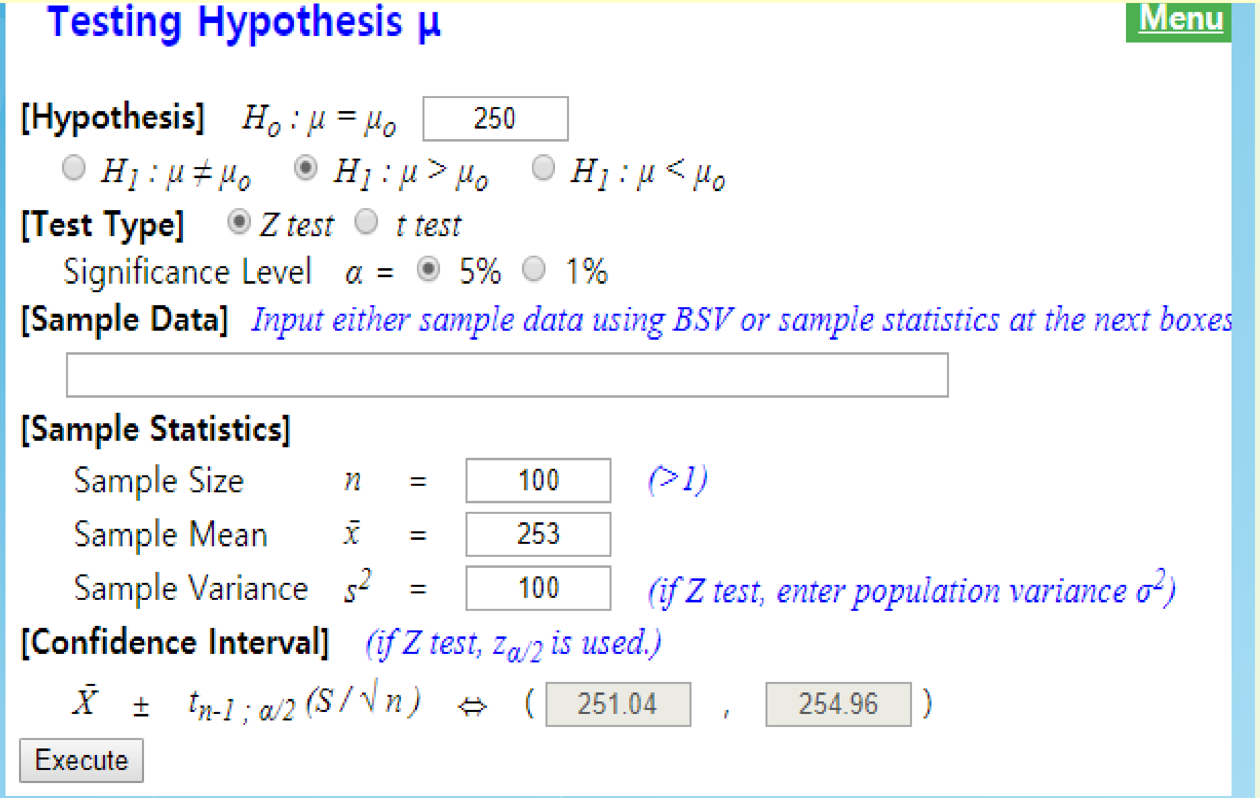
  Since $\frac{253-250}{\frac{10}{10}} = 3$ and $z_{0.005} = 2.575$, H_0 is rejected.

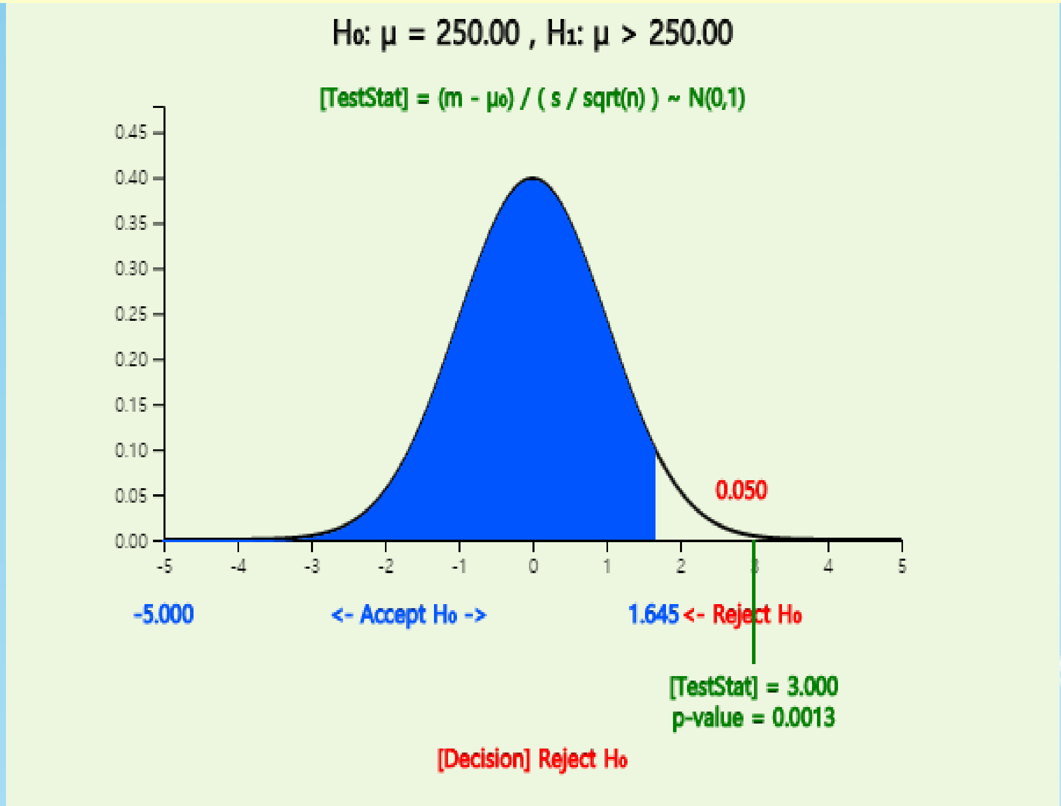
- p-value = 2 P($\bar{X}$ > 253) = P(Z > $\frac{253-250}{\frac{10}{10}}$) = 2 P(Z > 3) =0.0026

7.1 Testing Hypothesis for a Population Mean

3) In『eStatU』menu, select 'Testing Hypothesis ', enter 250 at the box of on [Hypothesis] and select the alternative hypothesis as the right test.

- Check [Test Type] as Z test and check the significance level at 5%.
- At the [Sample Statistics], enter sample size 100, sample mean 253, and sample variance 100.





7.1 Testing Hypothesis for a Population Mean

[Example 7.1.3] When sample size is 16 and sample variance is 100 in [Example 7.1.2], test whether the average weight of the cookie bags is 250g or greater and obtain the p-value. Check the result using 『eStatU』

<Answer>

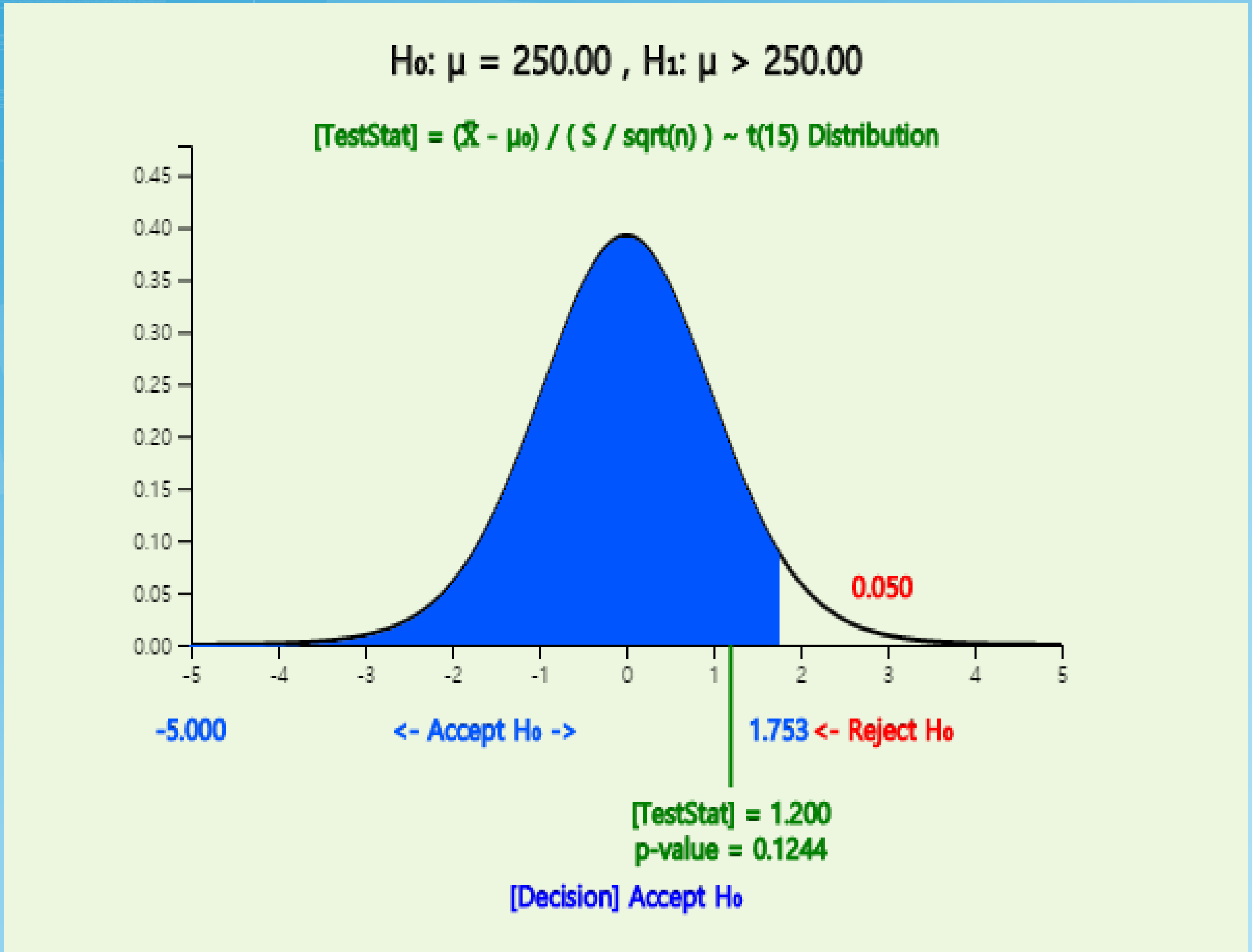
- Since the population standard deviation is unknown and the sample size is small, the decision rule is as follows:

  'If $\frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}} > t_{n-1;\, \alpha}$, then reject H_0'

  'If $\frac{253 - 250}{\frac{10}{\sqrt{16}}} > t_{16-1; 0.01}$, then reject H_0'

  Since (253-250) / ($\frac{10}{\sqrt{16}}$) = 1.2 and $t_{15;\, 0.01}$ = 2.602, H_0 is accepted.
- Since p-value is probability that $t_{15;}$ is greater than the test statistics 1.200, p-value is 0.124 by using the module of t-distribution in 『eStatU』.

7.1 Testing Hypothesis for a Population Mean



7.1 Testing Hypothesis for a Population Mean

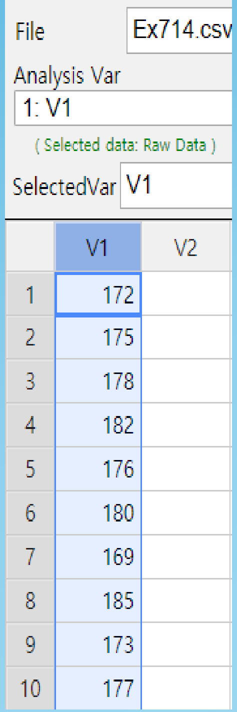
[Example 7.1.4] 10 male students are sampled in a university and examined their heights as follows:

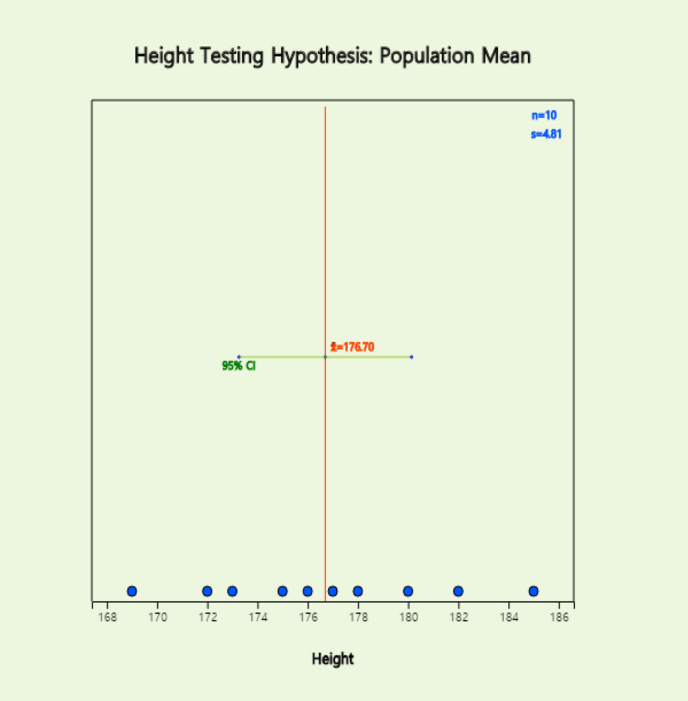
172 175 178 182 176

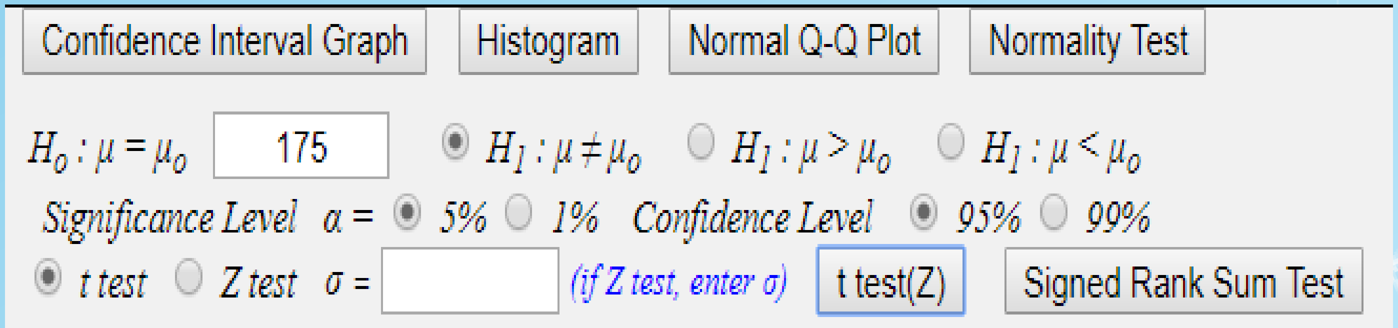
180 169 185 173 177 (Unit cm)

[Ex] ⇨ eBook ⇨ EX070104_Height.csv.

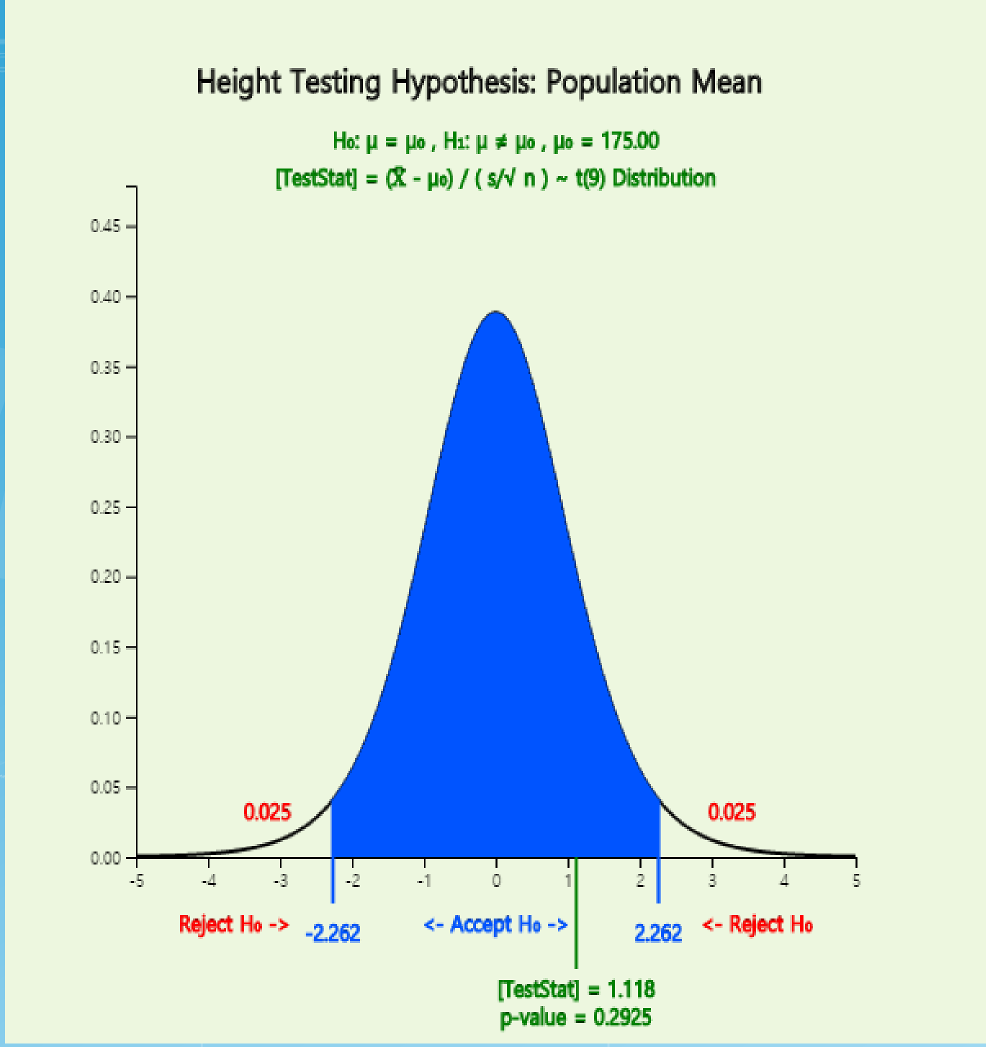
Test the hypothesis whether the population mean is 175cm or greater with the significance level of 5%.







7.1 Testing Hypothesis for a Population Mean



Testing Hypothesis: Population Mean	Analysis Var	Height			
Statistics	Observation	Mean	Std Dev	std err	Population Mean 95% Confidence Interval
	10	176.700	4.809	1.521	(173.260, 180.140)
Missing Observations	0				
Hypothesis					
$H_0 : \mu = \mu_0$	μ_0	[TestStat]	t value	p-value	
$H_1 : \mu \neq \mu_0$	175.00	sample mean	1.118	0.2925	

Thank you