Chapter 8

Testing Hypothesis for Two Population Parameters

Professor Jung Jin Lee

Table of Contents

8.1 Testing hypothesis for two population means
- 8.1.1 Independent sample
- 8.1.2 Paired sample

8.2 Testing hypothesis for two population variances

8.3 Testing hypothesis for two population proportions

8.1 Testing hypothesis for two population means

- There are a lot of examples comparing the mean of the two populations.

  - Is there a difference between the starting salary of male and female for this year's college graduates?
  - Is there a difference in the weight of the products produced in the two production lines?
  - Did the special training given to the typist to increase the speed of typing really bring about an increase in the speed of typing?

- Comparison of the two population means is possible by testing the hypothesis that t he difference in the population means is greater than or equal to zero.

- The comparison of these two population means is depending on whether samples t aken from each population are independently or not (paired comparison).

8.1 Testing hypothesis for two population means

8.1.1 Two Independent Samples

- Generally, testing hypothesis for two population means can be divided into three types, depending on the type of the alternative hypothesis:

1) $H_0: \mu_1 - \mu_2 = D_0$
   $H_1: \mu_1 - \mu_2 > D_0$

2) $H_0: \mu_1 - \mu_2 = D_0$
   $H_1: \mu_1 - \mu_2 > D_0$

3) $H_0: \mu_1 - \mu_2 = D_0$
   $H_1: \mu_1 - \mu_2 \neq D_0$

Here D_0 is the value for the difference in population means to be tested. When samples are selected independently from each other in the population, the estimator of the difference of the population means $\mu_1 - \mu_2$ is the difference in sample means $\overline{X}_1 - \overline{X}_2$, and the distribution of all possible sample mean difference is approximately equal to the normal distribution with mean $\mu_1 - \mu_2$ and variance $\sigma_1^2/n_1 + \sigma_2^2/n_2$ if the sample is large enough.

8.1 Testing hypothesis for two population means

Table 8.1.1 Testing hypothesis of two population means
- independent samples, populations are normal distributions, case of two population variances are equal

Type of Hypothesis	Decision Rule
1) $H_0: \mu_1 - \mu_2 = D_0$ $H_1: \mu_1 - \mu_2 > D_0$	If $\dfrac{(\overline{X}_1 - \overline{X}_2) - D_0}{\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}} > t_{n_1+n_2-2;\alpha}$, then reject H_0, else accept H_0
2) $H_0: \mu_1 - \mu_2 = D_0$ $H_1: \mu_1 - \mu_2 < D_0$	If $\dfrac{(\overline{X}_1 - \overline{X}_2) - D_0}{\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}} < -t_{n_1+n_2-2;\alpha}$, then reject H_0, else accept H_0
3) $H_0: \mu_1 - \mu_2 = D_0$ $H_1: \mu_1 - \mu_2 \neq D_0$	If $\left\lvert \dfrac{(\overline{X}_1 - \overline{X}_2) - D_0}{\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}} \right\rvert > t_{n_1+n_2-2;\alpha/2}$, then reject H_0, else accept H_0

※ If sample sizes are large enough ($n_1 > 30$, $n_2 > 30$), t distribution is approximately close to the standard normal distribution and the decision rule may use the standard normal distribution.

8.1 Testing hypothesis for two population means

[Ex 8.1.1] Two machines produce a cookie at a factory and a cookie package has a static capacity of 270 grams. Samples were extracted from each of the packages by two machines to examine the weight of package.

The weight average of 15 packages extracted from machine 1 was 275g, the standard deviation was 12g, and the weight average of 14 packages extracted from machine 2 was 269g and the standard deviation was 10g.

Test the weights of the cookie package produced by the two machines are different at a 1% significant level. Check the test results using 『eStatU』.

8.1 Testing hypothesis for two population means

<Answer of Ex 8.1.1>

- The hypothesis of this problem is $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 \neq \mu_2$. Hence the decision rule is as follows.

$$\text{If } \left| \frac{(\overline{X}_1 - \overline{X}_2) - D_0}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} \right| > t_{n_1+n_2-2;\,\alpha/2}, \text{ then reject } H_0$$

The information in the example can be summarized as follows.

$$n_1 = 15,\ \overline{X}_1 = 275,\ s_1 = 12,$$
$$n_2 = 14,\ \overline{X}_2 = 269,\ s_2 = 10$$

Therefore,

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$$
$$= \frac{(15-1)12^2 + (14-1)10^2}{15 + 14 - 2} = 122.815$$
$$\left| \frac{275-269}{\sqrt{\frac{122.815}{15} + \frac{122.815}{14}}} \right| = 1.457$$
$$t_{15+14-2;\,0.01/2} = t_{27;\,0.005} = 2.7707$$

Since $1.457 < 2.7707$, H_0 cannot be rejected.

8.1 Testing hypothesis for two population means

<Answer of Ex 8.1.1>

Testing Hypothesis μ1, μ2 Menu

[Hypothesis] $H_o : \mu_1 - \mu_2 = D$ [0]

◉ $H_1 : \mu_1 - \mu_2 \neq D$ ○ $H_1 : \mu_1 - \mu_2 > D$ ○ $H_1 : \mu_1 - \mu_2 < D$

[Test Type] t test , Variance Assumption ◉ $\sigma_1^2 = \sigma_2^2$ ○ $\sigma_1^2 \neq \sigma_2^2$

Significance Level α = ○ 5% ◉ 1%

Sampling Type ◉ independent sample ○ paired sample

[Sample Data] *Input either sample data using BSV or sample statistics at the next boxes*

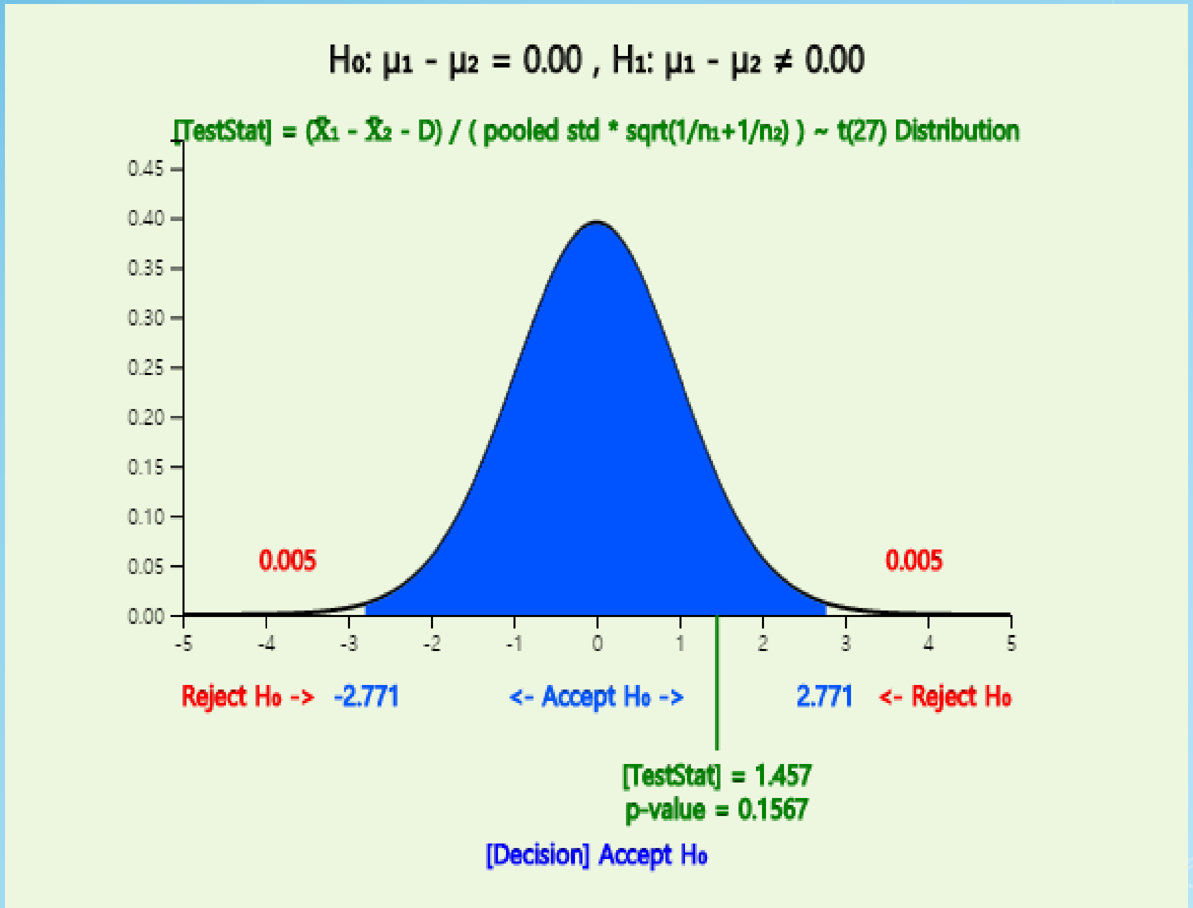
Sample 1 [ ]

Sample 2 [ ]

[Sample Statistics]

Sample Size	n_1 =	15	n_2 =	14		
Sample Mean	$\bar{x}_1$ =	275	$\bar{x}_2$ =	269	$\bar{x}_d$ =	
Sample Variance	s_1^2 =	144	s_2^2 =	100	s_d^2 =	

[Execute]



8.1 Testing hypothesis for two population means

8.1.1 Two Independent Samples

- If the variances of the two populations are different, the test statistic

$$\frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

do not follow the t distribution even if the populations are normally distributed. The testing hypothesis of the two population means when their population variances are different is called a Behrens-Fisher problem. Several methods for solving this problem have been studied. The Satterthwaite method approximate the degree of freedom ϕ of the t distribution, in the decision rule in Table 8.1.1 as follows.

$$\phi = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

8.1 Testing hypothesis for two population means

[Example 8.1.2] If the two population variances are different in [Example 8.1.1], test the weight of the cookie packages produced in Machine 1 and Machine 2 at a 1% significant level. Check the test results using 「eStatU」.

<Answer>

- Since the population variances are different, the degree of freedom ϕ of t distribution is approximated as follows.

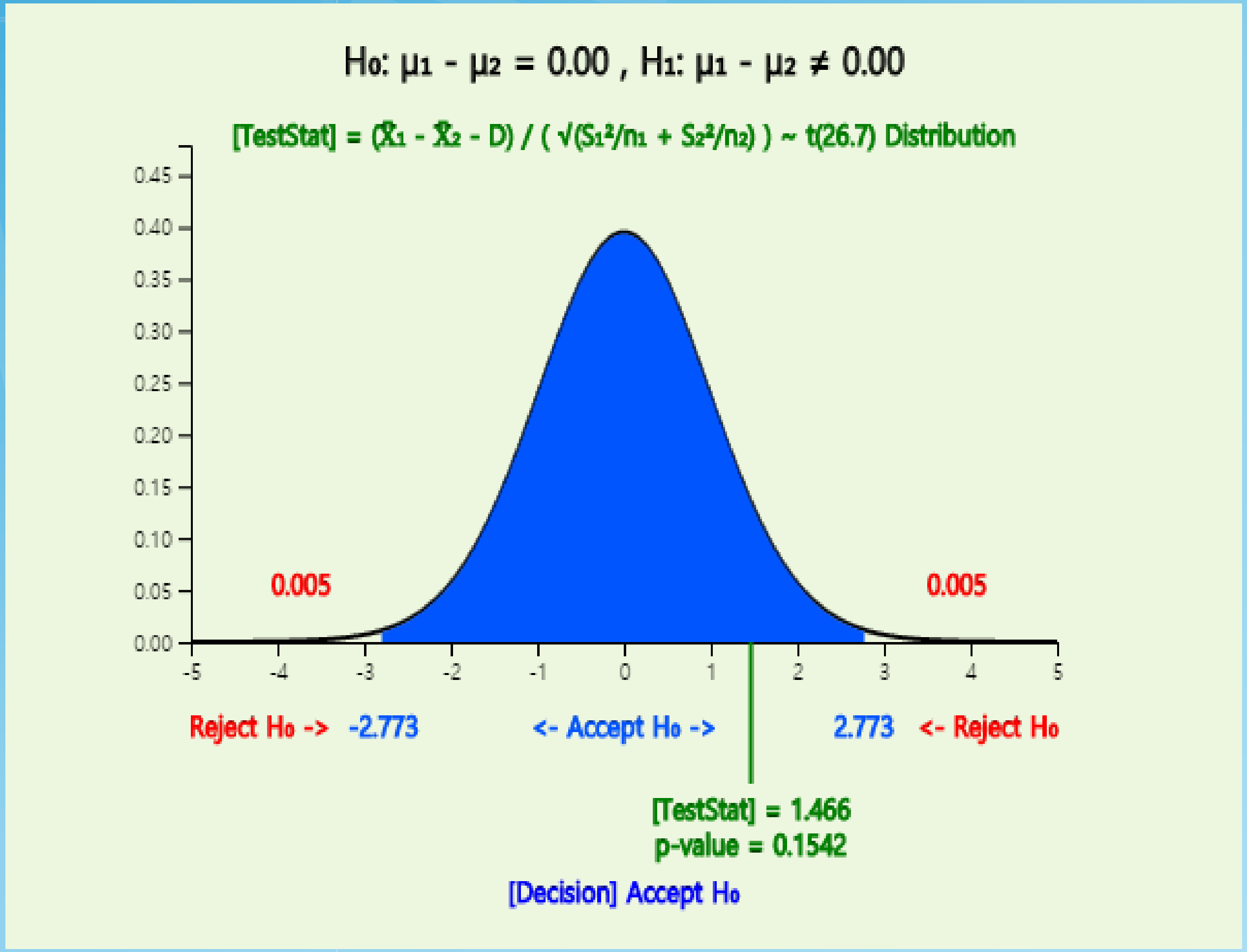
$$\phi = \frac{\left[\frac{12^2}{15} + \frac{10^2}{14}\right]^2}{\frac{\left(\frac{12^2}{15}\right)^2}{15-1} + \frac{\left(\frac{10^2}{14}\right)^2}{14-1}} = 26.67$$

$$t_{26.7\,;\,0.01/2} = 2.773$$

Since 1.457 < 2.773, H_0 cannot be rejected. .

8.1 Testing hypothesis for two population means

<Answer of Ex 8.1.2>



8.1 Testing hypothesis for two population means

[Example 8.1.3] (『eStat』 exercise)
A sample of 10 men and women in the male and female populations of college graduates this year was taken and the monthly average wage was examined as follows. (Unit 10,000 KRW)

Men	272 255 278 282 296 312 356 296 302 312
Women	276 280 369 285 303 317 290 250 313 307

1) If the population variances are the same, test the hypothesis at a significant level of 5% whether the average monthly wage for male and female is the same.
2) If you assume that the population variances are different, test the hypothesis at a significant level of 5% whether the average monthly wage for male and female is the same.

8.1 Testing hypothesis for two population means

<Answer of Ex 8.1.3>

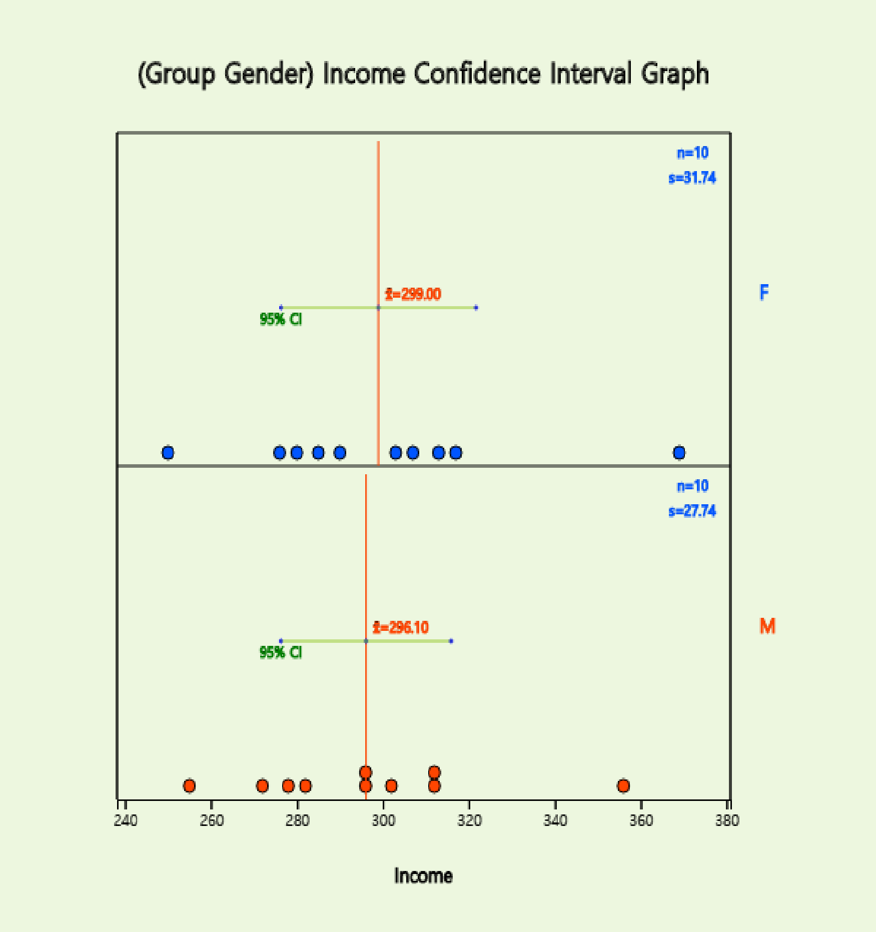
File Ex813IncomByGender.csv

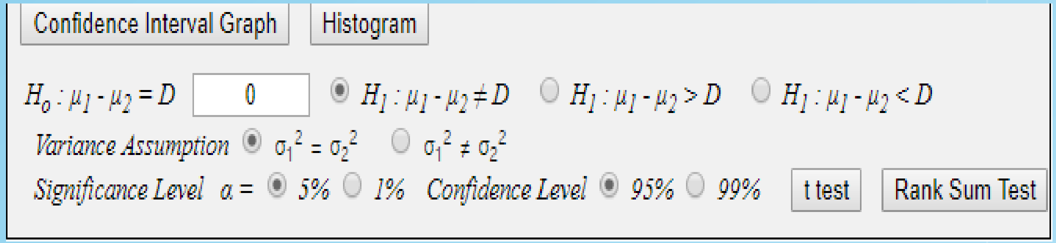
Analysis Var: 2: Income — by Group: 1: Gender

(Selected data: Raw Data) (or Paired Var)

SelectedVar V2 by V1,

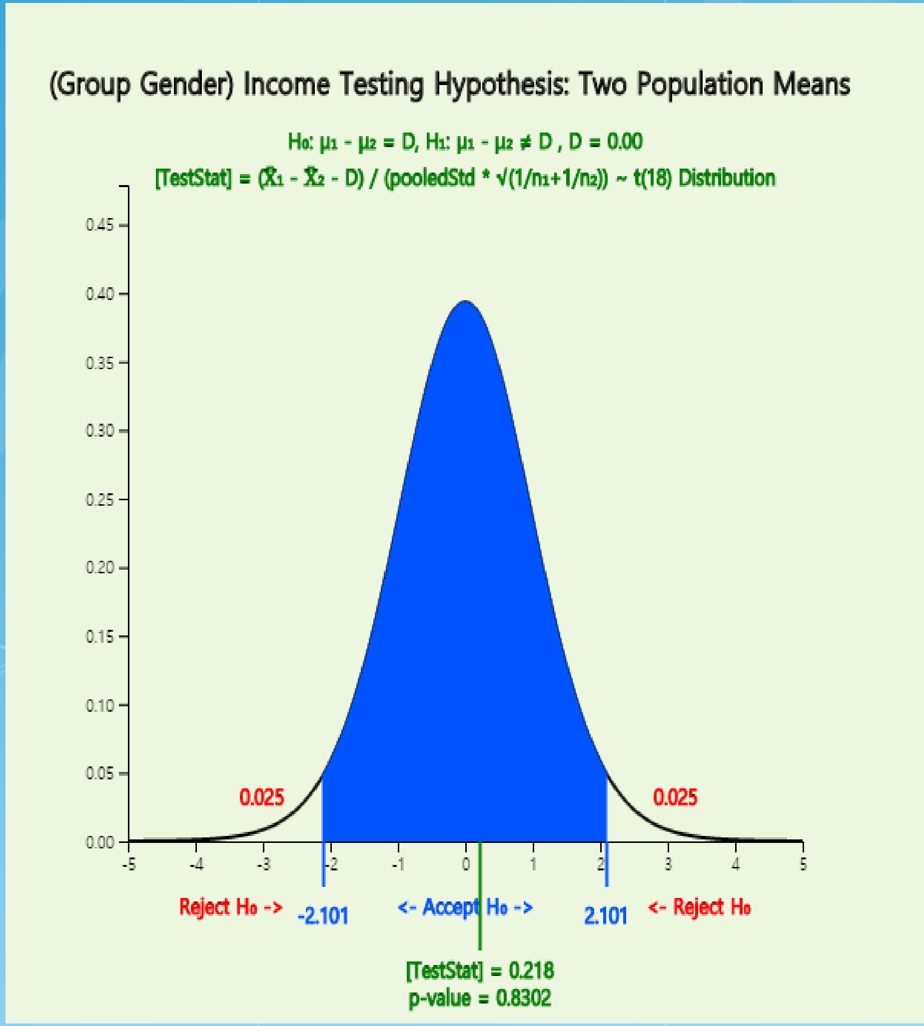
	Gender	Income	V3	V4
1	M	272		
2	M	255		
3	M	278		
4	M	282		
5	M	296		
6	M	312		
7	M	356		
8	M	296		
9	M	302		
10	M	312		
11	F	276		
12	F	280		
13	F	369		
14	F	285		
15	F	303		
16	F	317		
17	F	290		
18	F	250		
19	F	313		
20	F	307		





8.1 Testing hypothesis for two population means

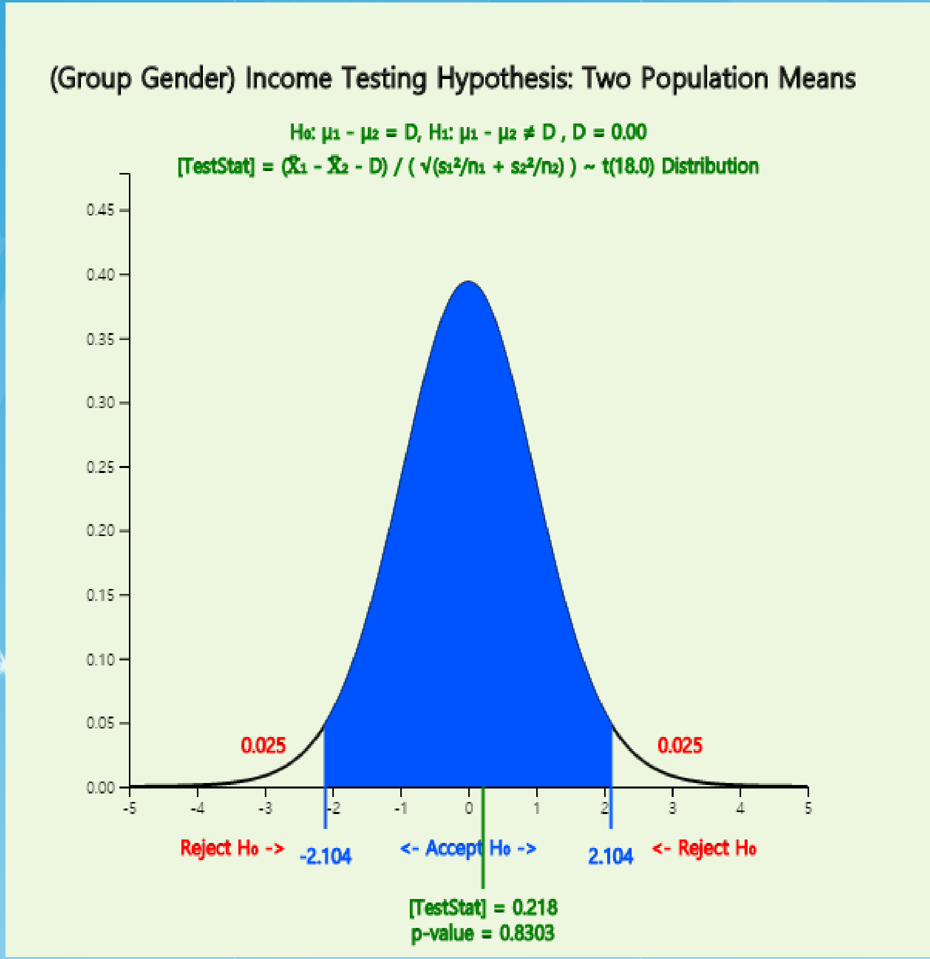
<Answer of Ex 8.1.3>



Testing Hypothesis: Two Population Means	Analysis Var	Income	Group Name	Gender	
Statistics	Observation	Mean	Std Dev	std err	Population Mean 95% Confidence Interval
1 (F)	10	299.000	31.742	10.038	(276.293, 321.707)
2 (M)	10	296.100	27.739	8.772	(276.257, 315.943)
Total	20	297.550	29.051	6.496	(283.954, 311.146)
Missing Observations	0				
Hypothesis	Variance Assumption	$\sigma_1^2 = \sigma_2^2$			
$H_0 : \mu_1 - \mu_2 = D$	D	[TestStat]	t value	p-value	μ_1-μ_2 95% Confidence Interval
$H_1 : \mu_1 - \mu_2 \neq D$	0.00	Difference of Sample Means	0.218	0.8302	(-25.106, 30.906)

8.1 Testing hypothesis for two population means

<Answer of Ex 8.1.3>



Testing Hypothesis: Two Population Means	Analysis Var	Income	Group Name	Gender	
Statistics	Observation	Mean	Std Dev	std err	Population Mean 95% Confidence Interval
1 (F)	10	299.000	31.742	10.038	(276.293, 321.707)
2 (M)	10	296.100	27.739	8.772	(276.257, 315.943)
Total	20	297.550	29.051	6.496	(283.954, 311.146)
Missing Observations	0				
Hypothesis	Variance Assumption	$\sigma_1^2 \neq \sigma_2^2$			
$H_0 : \mu_1 - \mu_2 = D$	D	[TestStat]	t value	p-value	μ_1-μ_2 95% Confidence Interval
$H_1 : \mu_1 - \mu_2 \neq D$	0.00	Difference of Sample Means	0.218	0.8303	(-25.142, 30.942)

8.1 Testing hypothesis for two population means

8.1.2 Paired Comparison

- In some cases it is difficult to extract them independently, or when they are extracted independently, the resulting analysis may be meaningless because the characteristics of each sample object differ too much.

- For example, let's say that you want to give a typist a special education to increase the speed of typing and see if this training has been effective in the speed of typing. At this time, it is difficult to measure the effectiveness of education because individual differences are severe if different samples are extracted before and after education. In this case, for a typist who has sampled, if you measure the typing speed before the training and after the training, the effect of special education can be well understood.

8.1 Testing hypothesis for two population means

8.1.2 Paired Comparison

Table 8.1.2 Data for a paired comparison

Sample of population 1 (x_{i1})	Sample of population 2 (x_{i2})	Difference $d_i = x_{i1} - x_{i2}$
x_{11} x_{21} .... x_{n1}	x_{12} x_{22} .... x_{n2}	$d_1 = x_{11} - x_{12}$ $d_2 = x_{21} - x_{22}$ .... $d_n = x_{n1} - x_{n2}$
	Mean of d_i Variance d_i	$\bar{d} = \sum d_i / n$ $s_d^2 = \sum (d_i - \bar{d})^2 / (n-1)$

8.1 Testing hypothesis for two population means

Table 8.1.3 Testing hypothesis of two population means (paired comparison)
- two populations are normal, and paired sample case

Type of Hypothesis	Decision Rule
1) $H_0 : \mu_1 - \mu_2 = D_0$ $H_1 : \mu_1 - \mu_2 > D_0$	If $\dfrac{\bar{d} - D_0}{\frac{s_d}{\sqrt{n}}} > t_{n-1;\alpha}$, then reject H_0, else accept H_0
2) $H_0 : \mu_1 - \mu_2 = D_0$ $H_1 : \mu_1 - \mu_2 < D_0$	If $\dfrac{\bar{d} - D_0}{\frac{s_d}{\sqrt{n}}} < -t_{n-1;\alpha}$, then reject H_0, else accept H_0
3) $H_0 : \mu_1 - \mu_2 = D_0$ $H_1 : \mu_1 - \mu_2 \neq D_0$	$\left\lvert \dfrac{\bar{d} - D_0}{\frac{s_d}{\sqrt{n}}} \right\rvert > t_{n-1;\alpha/2}$, then reject H_0, else accept H_0

8.1 Testing hypothesis for two population means

[Example 8.1.4] The following is the result of training to speed up the typing speed of eight typists before and after the training. Test whether or not typing speed has increased at a 5% significant level. Assume that the speed of typing is nomal. Check the test results using 「eStat」 and 「eStatU」

id	Typing speed before training (unit: words/min)	Typing speed after training (unit: words/min)
1	52	58
2	60	62
3	63	62
4	43	48
5	46	50
6	56	55
7	62	68
8	50	57

8.1 Testing hypothesis for two population means

<Answer of Ex 8.1.4>

- This problem is for testing the null hypothesis $H_0: \mu_1 - \mu_2 = 0$ to the alternative hypothesis $H_1: \mu_1 - \mu_2 < 0$ to compare the typing speed of typists before training (population 1) and after training (population 2) using paired samples. Therefore, the decision rule is as follows:

  If $\dfrac{\bar{d} - D_o}{\frac{s_d}{\sqrt{n}}} < -t_{n-1;\alpha}$, then reject H_0.

8.1 Testing hypothesis for two population means

<Answer of Ex 8.1.4>

id	Typing speed before training (unit: words/min)	Typing speed after training (unit: words/min)	Difference d_i
1	52	58	−6
2	60	62	−2
3	63	62	1
4	43	48	−5
5	46	50	−4
6	56	55	1
7	62	68	−6
8	50	57	−7
			Mean $\overline{d} = -3.5$ Standard deviation $s_d = 3.16$

- The test statistic is as follows.

$$\frac{\overline{d} - D_o}{\dfrac{s_d}{\sqrt{n}}} = \frac{-3.5}{\dfrac{3.16}{\sqrt{8}}} = -3.13$$

$$-t_{n-1\,;\,\alpha} = -t_{8-1\,;\,0.05} = -t_{7\,;\,0.05} = -1.8946,$$

Therefore H_0 is rejected and concludes that the training increased the typing speed.

8.1 Testing hypothesis for two population means

<Answer of Ex 8.1.4>

Testing Hypothesis μ_1, μ_2 Menu

[Hypothesis] $H_o : \mu_1 - \mu_2 = D$ 0

○ $H_1 : \mu_1 - \mu_2 \neq D$ ○ $H_1 : \mu_1 - \mu_2 > D$ ◉ $H_1 : \mu_1 - \mu_2 < D$

[Test Type] t test , Variance Assumption ◉ $\sigma_1^2 = \sigma_2^2$ ○ $\sigma_1^2 \neq \sigma_2^2$

Significance Level α = ◉ 5% ○ 1%

Sampling Type ○ independent sample ◉ paired sample

[Sample Data] *Input either sample data using BSV or sample statistics at the next boxes*

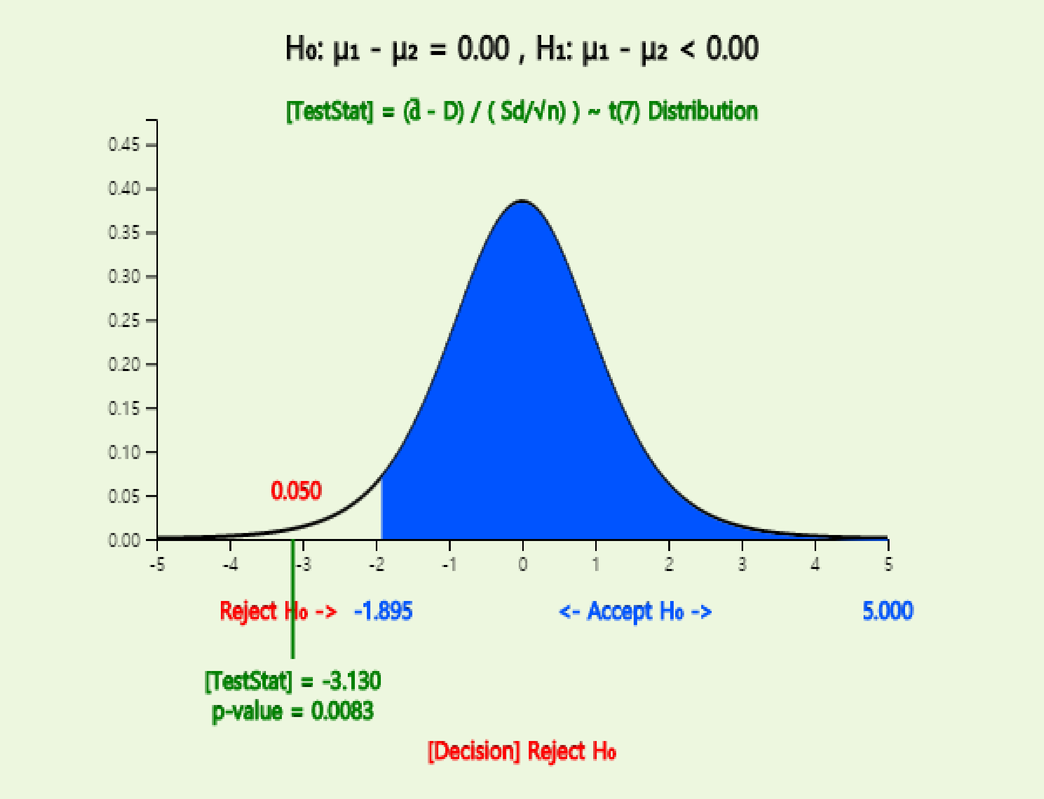
Sample 1: 52 60 63 43 46 56 62 50

Sample 2: 58 62 62 48 50 55 68 57

[Sample Statistics]

Sample Size	n_1 = 8	n_2 = 8			
Sample Mean	$\bar{x}_1$ = 54.00	$\bar{x}_2$ = 57.50	$\bar{x}_d$ = -3.500		
Sample Variance	s_1^2 = 55.71	s_2^2 = 43.43	s_d^2 = 10.000		

Execute



8.1 Testing hypothesis for two population means

<Answer of Ex 8.1.4>

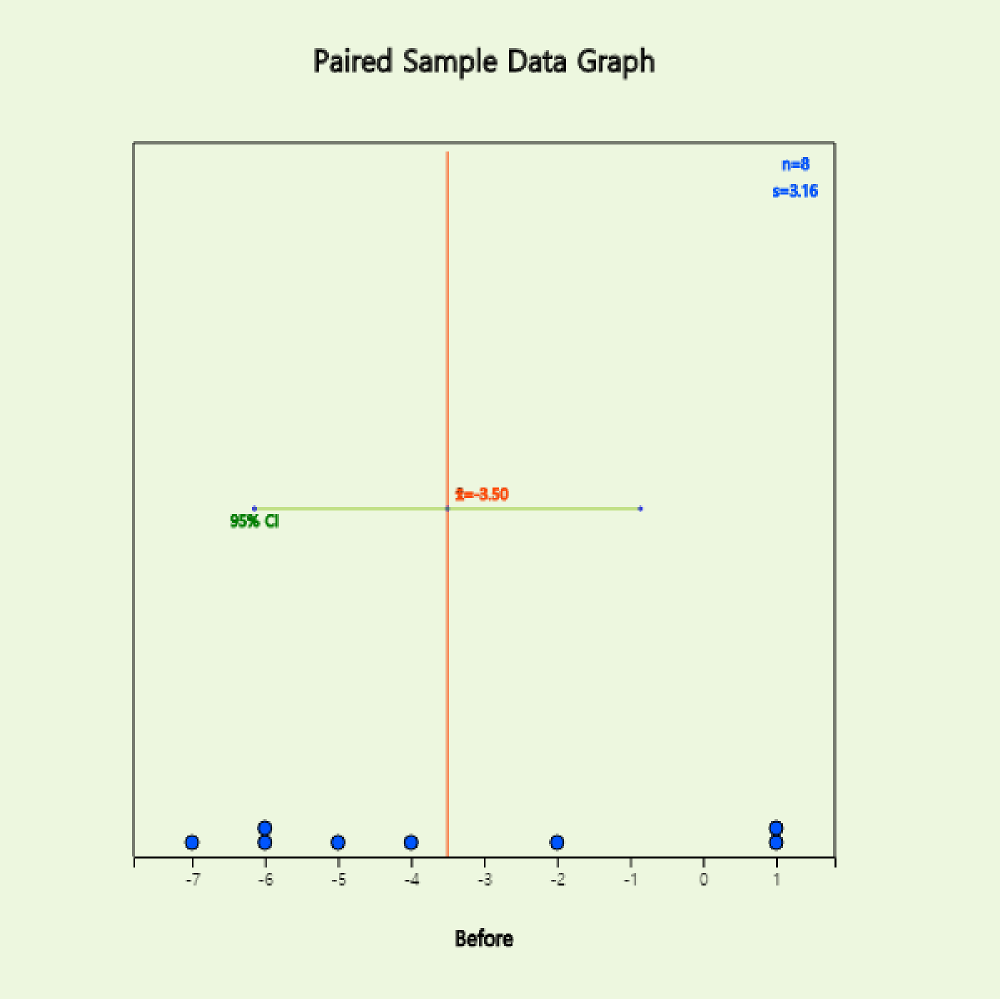
File: Ex814TypingSpeedEducation.csv

Analysis Var: 1: Before — by Group: 2: After

(Selected data: Raw Data) (or Paired Var)

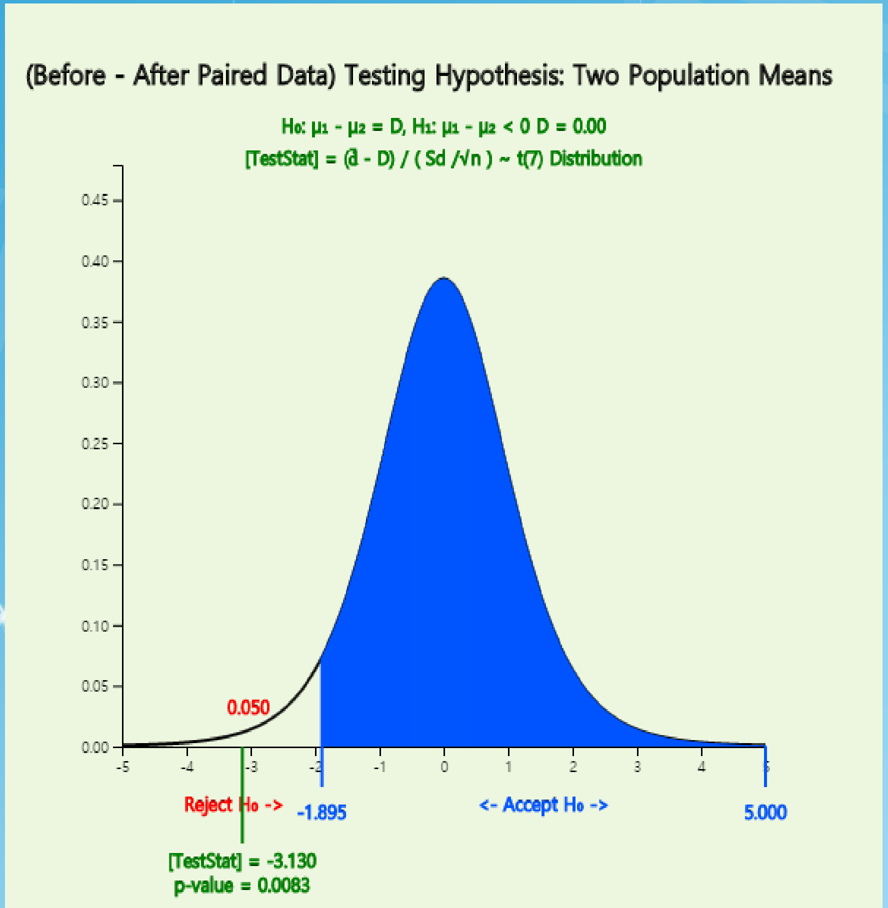
SelectedVar: V1 by V2,

	Before	After	V3	V4	V5
1	52	58			
2	60	62			
3	63	62			
4	43	48			
5	46	50			
6	56	55			
7	62	68			
8	50	57			



8.1 Testing hypothesis for two population means

<Answer of Ex 8.1.4>



Testing Hypothesis: Two Population Means	Analysis Var	(Before - After)			
Statistics	Observation	Mean	Std Dev	std err	Population Mean 95% Confidence Interval
	8	-3.500	3.162	1.118	(-6.144, -0.856)
Missing Observations	0				
Hypothesis	Variance Assumption	$\sigma_1^2 = \sigma_2^2$			
$H_0 : \mu_1 - \mu_2 = D$	D	[TestStat]	t value	p-value	μ_1-μ_2 95% Confidence Interval
$H_1 : \mu_1 - \mu_2 < D$	0.00	Difference of Sample Means	-3.130	0.0083	(-6.144, -0.856)

8.2 Testing hypothesis for two population variances

- Consider the following examples of comparing the two population variances.

- When comparing the two population means in the previous section, we found that if the sample was small, the decision rule for testing hypothesis were different depending on whether the two population variances were the same or different. So how can we test if the two unknown population variances are the same?

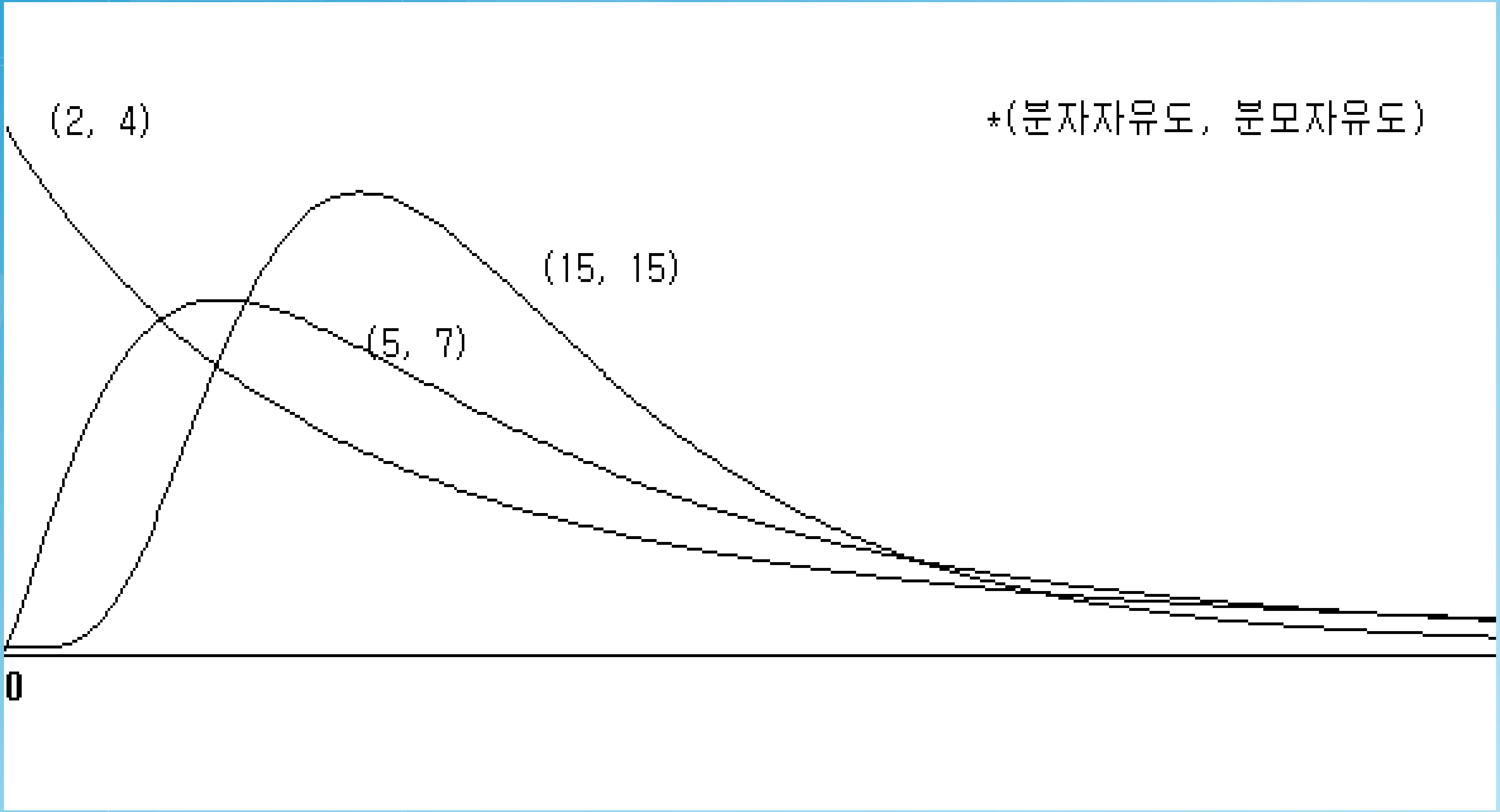
- The quality of bolts used to assemble cars depends on the strict specification for their diameter. The average diameter of these bolts is said to be the same for the two companies. So, how can you compare the variances when you think of smaller variances as superior?

8.2 Testing hypothesis for two population variances

$$\frac{\left(\frac{S_1^2}{\sigma_1^2}\right)}{\left(\frac{S_2^2}{\sigma_2^2}\right)}$$

follows the F distribution with numerator degree of freedom $n_1 - 1$ and denominator degree of freedom $n_2 - 1$. Using this fact, we can perform testing hypothesis on the ratio of population variances.

8.2 Testing hypothesis for two population variances



8.2 Testing hypothesis for two population variances

Table 8.2.1 Testing hypothesis for two population variances
- Two populations are normally distributed-

Type of Hypothesis	Decision Rule
1) $H_0 : \sigma_1^2 = \sigma_2^2$ $H_1 : \sigma_1^2 > \sigma_2^2$	If $\dfrac{S_1^2}{S_2^2} > F_{n_1-1,n_2-1\,;\,\alpha}$, then reject H_0, else accept H_0
2) $H_0 : \sigma_1^2 = \sigma_2^2$ $H_1 : \sigma_1^2 < \sigma_2^2$	If $\dfrac{S_1^2}{S_2^2} < F_{n_1-1,n_2-1\,;\,1-\alpha}$, then reject H_0, else accept H_0
3) $H_0 : \sigma_1^2 = \sigma_2^2$ $H_1 : \sigma_1^2 \neq \sigma_2^2$	If $\dfrac{S_1^2}{S_2^2} < F_{n_1-1,n_2-1\,;\,1-\alpha/2}$ or $\dfrac{S_1^2}{S_2^2} > F_{n_1-1,n_2-1\,;\,\alpha/2}$, then reject H_0, else accept H_0

8.2 Testing hypothesis for two population variances

[Example 8.2.1] A company that produces a bolt has two plants. One day, ten bolts produced in Plant 1 were sampled and the variance of diameter is 0.11^2. 12 bolts produced in Plant 2 were sample and the variance of diameter is 0.13^2. Test the variances of the bolts from two plants are the same or not with a 5% significant level. Check the test results using 「eStatU」

<Answer>

- The hypothesis of this problem is $H_0 :\ \sigma_1^2 = \sigma_2^2 \quad H_1 :\ \sigma_1^2 \neq \sigma_2^2$ and its decision rule is as follows.

If $\dfrac{S_1^2}{S_2^2} < F_{n_1-1, n_2-1\,;\,1-\alpha/2}$ or $\dfrac{S_1^2}{S_2^2} > F_{n_1-1, n_2-1\,;\,\alpha/2}$, then reject H_0, else accept H_0

The test statistic using two sample variances s_1^2, s_2^2 and percentile of F distribution is as follows.

$$\frac{s_1^2}{s_2^2} = \frac{0.0121}{0.0169} = 0.716$$

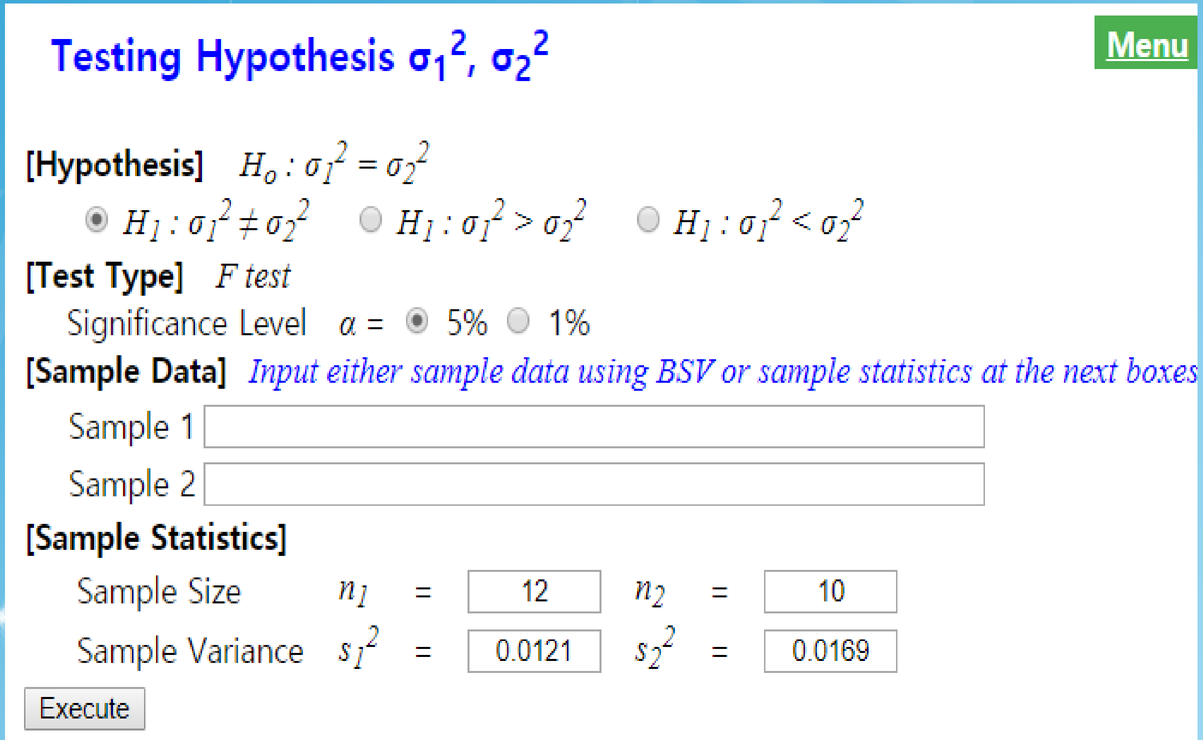
$$F_{n_1-1, n_2-1\,;\,1-\alpha/2} = F_{11,9\,;\,0.975} = 0.279$$

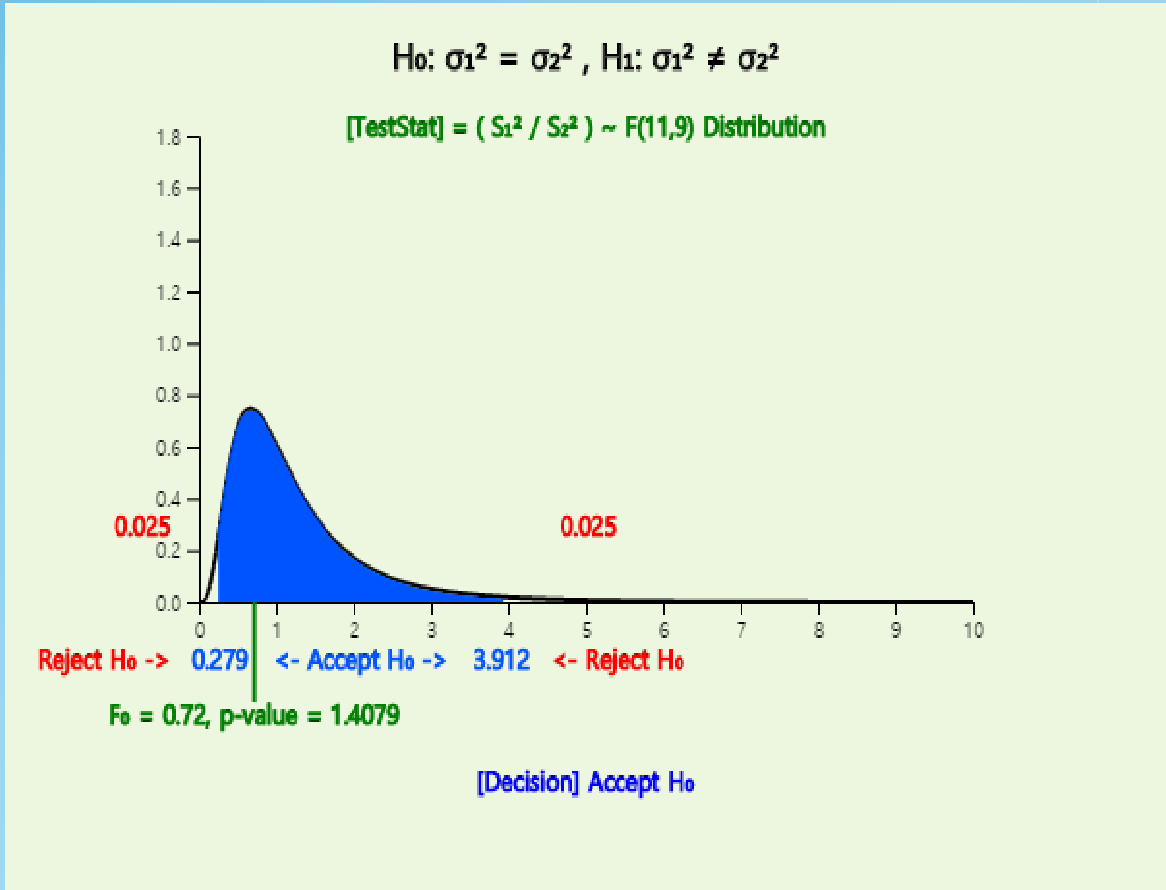
$$F_{n_1-1, n_2-1\,;\,\alpha/2} = F_{11,9\,;\,0.025} = 3.912$$

Hence the hypothesis H_0 cannot be rejected and conclude that two variances are equal.

8.2 Testing hypothesis for two population variances

<Answer of Ex 8.2.1>





8.2 Testing hypothesis for two population variances

[Example 8.2.2] (『eStat』) [Example 8.1.3] data
A sample of 10 male and 10 female of college graduates this year was taken and the monthly average income was examined as follows. (Unit 1000 KRW) Test if the variances of the two populations are equal.

Male	**272 255 278 282 296 312 356 296 302 312**
Female	**276 280 369 285 303 317 290 250 313 307**

8.2 Testing hypothesis for two population variances

<Answer of Ex 8.2.2>

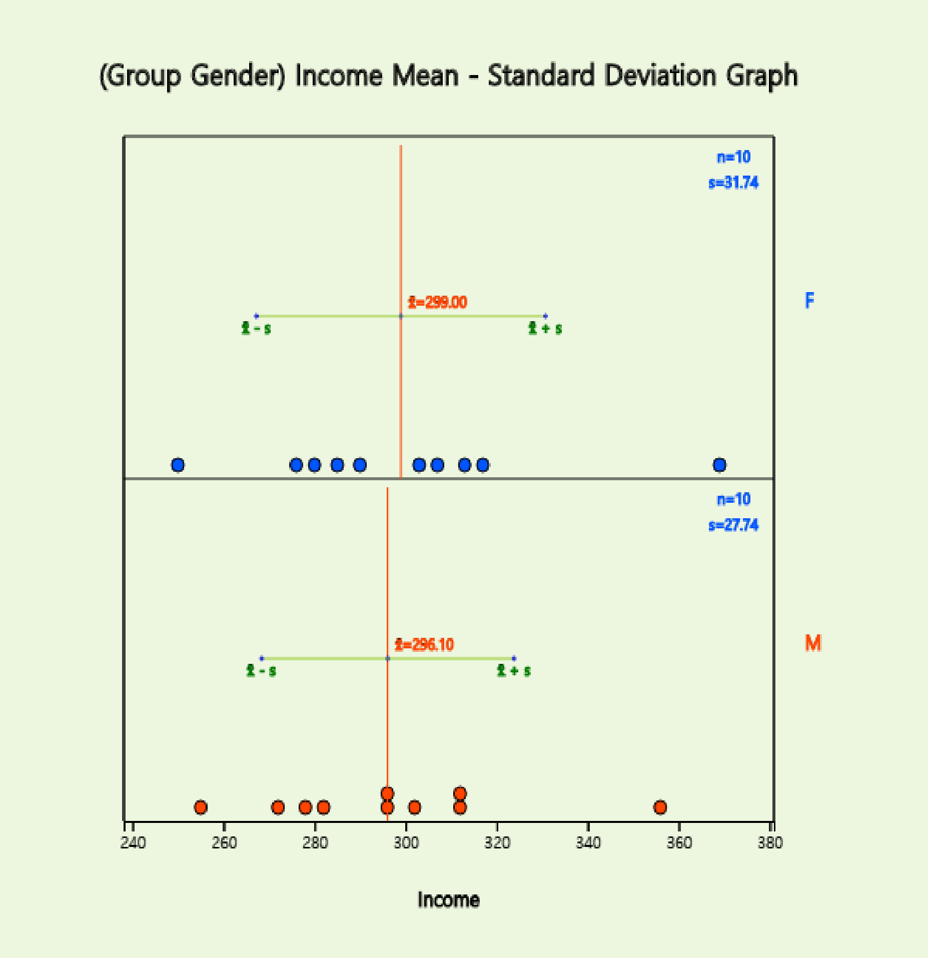
File: Ex813IncomByGender.csv

Analysis Var: 2: Income — by Group: 1: Gender

(Selected data: Raw Data) (or Paired Var)

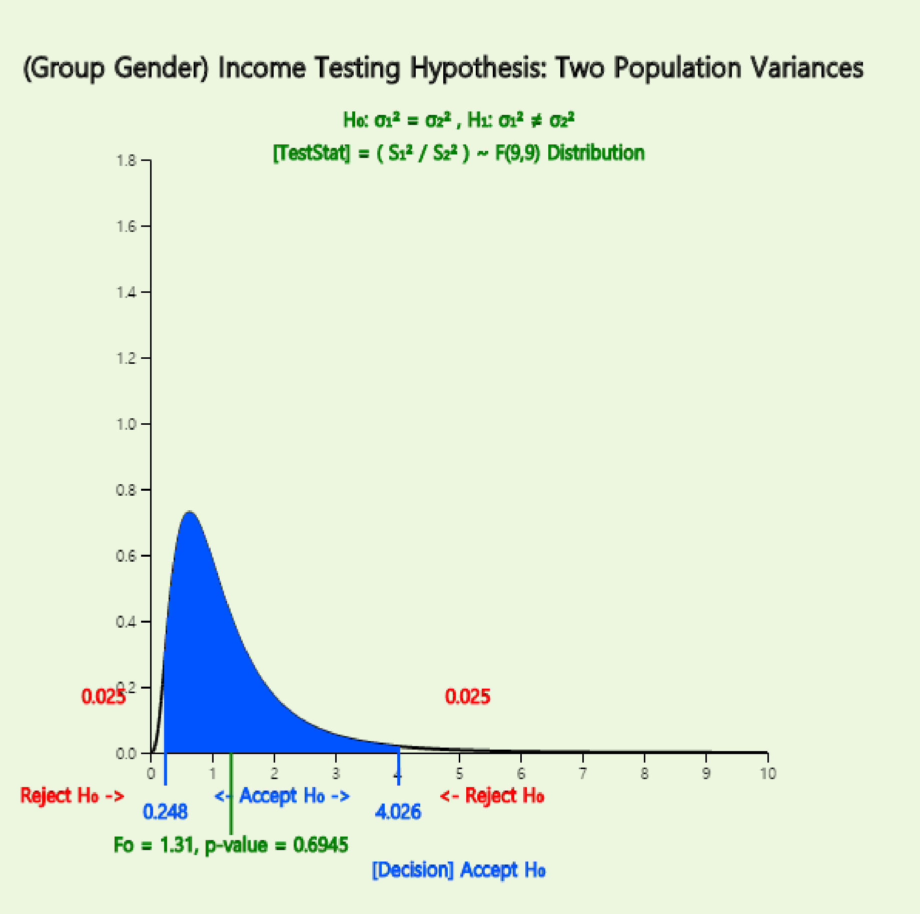
SelectedVar: V2 by V1,

	Gender	Income	V3	V4
1	M	272		
2	M	255		
3	M	278		
4	M	282		
5	M	296		
6	M	312		
7	M	356		
8	M	296		
9	M	302		
10	M	312		
11	F	276		
12	F	280		
13	F	369		
14	F	285		
15	F	303		
16	F	317		
17	F	290		
18	F	250		
19	F	313		
20	F	307		



8.2 Testing hypothesis for two population variances

<Answer of Ex 8.2.2>



Testing Hypothesis: Two Population Variances	Analysis Var	Income	Group Name	Gender	
Statistics	Observation	Mean	Std Dev	std err	Population Variance 95% Confidence Interval
1 (F)	10	299.000	31.742	10.038	(476.692, 3358.034)
2 (M)	10	296.100	27.739	8.772	(364.032, 2564.408)
Total	20	297.550	29.051	6.496	(488.092, 1800.362)
Missing Observations	0				
Hypothesis					
$H_0 : \sigma_1^2 = \sigma_2^2$		[TestStat]	F-value	p-value	σ_1^2 / σ_2^2 95% Confidence Interval
$H_1 : \sigma_1^2 \neq \sigma_2^2$		S_1^2 / S_2^2	1.309	0.6945	(0.325, 5.272)

8.3 Testing hypothesis for two population proportions

- **Let's take a look at the example below that compares the two proportions.**

- **Is there a gender gap in the approval rating for a particular candidate in this year's presidential election?**

- **A factory has two machines that make products. Do the two machines have different defect rates?**

8.3 Testing hypothesis for two population proportions

Table 8.3.1 Testing hypothesis for two population proportions
- two independent large samples -

Type of Hypothesis	Decision Rule
1) $H_0 : p_1 = p_2$ $H_1 : p_1 > p_2$	If $\dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n_1} + \dfrac{\overline{p}(1-\overline{p})}{n_2}}} > z_\alpha$, then reject H_0, else accept H_0
2) $H_0 : p_1 = p_2$ $H_1 : p_1 < p_2$	If $\dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n_1} + \dfrac{\overline{p}(1-\overline{p})}{n_2}}} < -z_\alpha$, then reject H_0, else accept H_0
3) $H_0 : p_1 = p_2$ $H_1 : p_1 \neq p_2$	If $\left\| \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n_1} + \dfrac{\overline{p}(1-\overline{p})}{n_2}}} \right\| > z_{\alpha/2}$, then reject H_0, else accept H_0

8.3 Testing hypothesis for two population proportions

[Example 8.3.1] In this year's presidential election, 54 out of 225 samples from male population supported the candidate A and 52 out of 175 samples from female population supported the candidate A where samples are independent. Test whether there is a difference in male and female approval ratings at a 5% significant level.

Check the results using 『eStatU』

8.3 Testing hypothesis for two population proportions

<Answer of Ex 8.3.1>

- The hypothesis of this proble is $H_0 \; : \; p_1 = p_2$, $H_1 \; : \; p_1 \neq p_2$ and its decision rule is as follows.

  If $\left| \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{\bar{p}(1-\bar{p})}{n_1} + \dfrac{\bar{p}(1-\bar{p})}{n_2}}} \right| > z_{\alpha/2}$, then reject H_0, else accept H_0

- Since $\hat{p}_1$= 54/225 = 0.240, $\hat{p}_2$= 52/175 = 0.297, $\bar{p}$ and the test statistic can be calculated as follows.

  $\bar{p}$ = (54+52) / (225 + 175) = 106/400 = 0.265

$$\left| \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{\bar{p}(1-\bar{p})}{n_1} + \dfrac{\bar{p}(1-\bar{p})}{n_2}}} \right| = \left| \frac{0.240 - 0.297}{\sqrt{\dfrac{0.265(1-0.265)}{225} + \dfrac{0.265(1-0.265)}{175}}} \right| = 1.28$$

$z_{\alpha/2}$ = $z_{0.05/2}$ = $z_{0.025}$ = 1.96

Therefore the hypothesis H_0 cannot be rejected and we conclude that there is not enough evidence to say that the approval ratings of certain male and female candidates are different.

8.3 Testing hypothesis for two population proportions

<Answer of Ex 8.3.1>

Testing Hypothesis p_1, p_2

[Hypothesis] $H_0 : p_1 - p_2 = D$ [0]

(•) $H_1 : p_1 - p_2 \neq D$ ( ) $H_1 : p_1 - p_2 > D$ ( ) $H_1 : p_1 - p_2 < D$

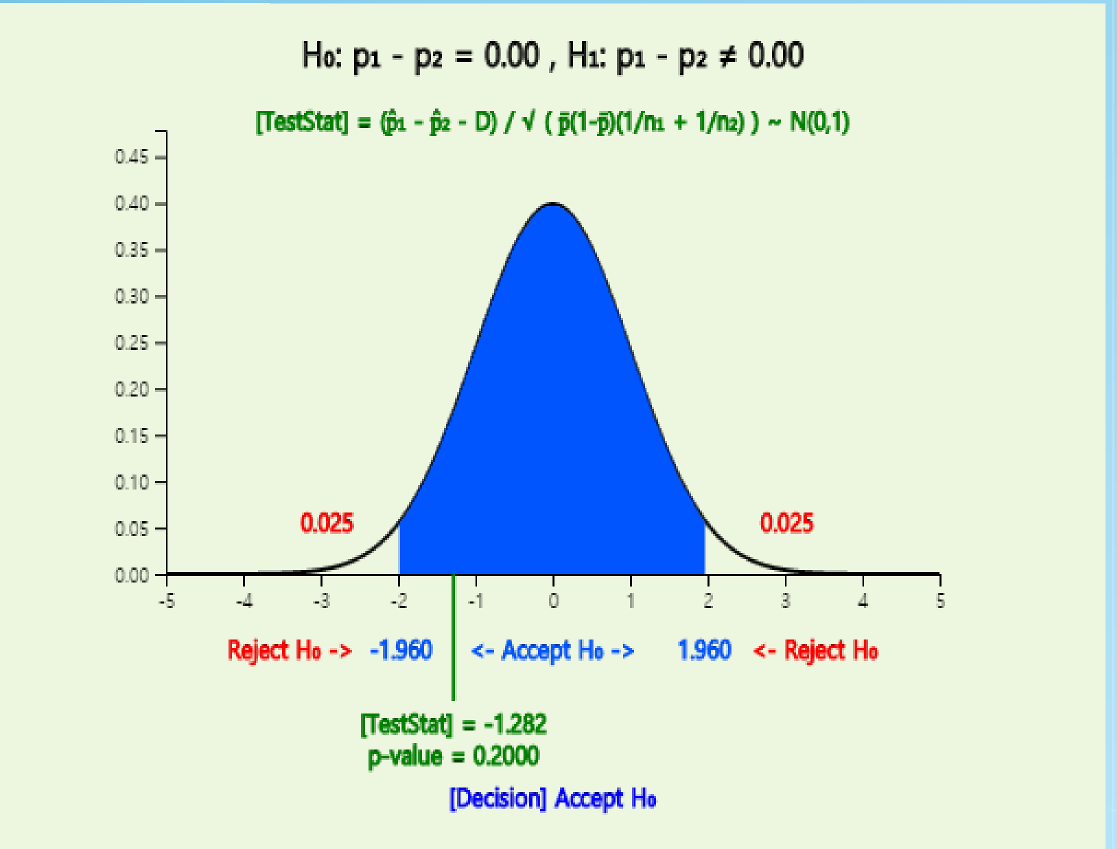
[Test Type] *Z test*

Significance Level $\alpha =$ (•) 5% ( ) 1%

[Sample Data]

Sample Size	n_1 =	225	n_2 =	175
Sample Proportion	$\hat{p}_1$ =	0.240	$\hat{p}_2$ =	0.297

[Execute]



8.4 Summary

- Testing hypothesis for two population means
  - Independent sample
  - Paired sample
- Testing hypothesis for two population variances
- Testing hypothesis for two population proportions

Thank you