Introduction to Statistics and Data Science using *eStat*

Chapter 8 Testing Hypothesis for Two Populations

8.1 Testing hypothesis for two population means - Paired sample -

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

8.1 Testing hypothesis for two population means

8.1.2 Paired Samples

- In some cases, it is difficult to extract samples independently.

  - Typing education to increase the speed of typing

    ⇨ if independent samples are selected, it is difficult to measure effectiveness of education because of individual differences

    ⇨ for a typist who has sampled, if you measure the typing speed before the training and after the training, effect of typing 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 $\dfrac{\bar{d} - D_0}{\dfrac{s_d}{\sqrt{n}}} > t_{n-1;\alpha}$, 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 $\dfrac{\bar{d} - D_0}{\dfrac{s_d}{\sqrt{n}}} < -t_{n-1;\alpha}$, then reject H_0, else accept H_0
3) $H_0 : \mu_1 - \mu_2 = D_0$ $H_1 : \mu_1 - \mu_2 \neq D_0$	$\left\lvert \dfrac{\bar{d} - D_0}{\dfrac{s_d}{\sqrt{n}}} \right\rvert > t_{n-1;\alpha/2}$, 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 a special training to improve the typing speed of eight typists before and after the training.

- Test whether typing speed has increased or not at the 5% significance level. Assume that the speed of typing follows a normal distribution.
- Check the test result 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 Example 8.1.4>

- **Hypothesis**

  $H_0 : \mu_1 - \mu_2 = D_0$

  $H_1 : \mu_1 - \mu_2 < D_0$

- **Decision Rule**

  If $\frac{(\bar{d} - D_0)}{\frac{s_d}{\sqrt{n}}} < -t_{n-1;\,\alpha}$, then Reject H_0

  $\frac{(\bar{d} - D_0)}{\frac{s_d}{\sqrt{n}}} = \frac{-3.5}{\frac{3.16}{\sqrt{8}}}$ = -3.13

  $t_{n-1;\,\alpha}$ = $t_{7;0.05}$ = -1.895

- **Therefore H_0 is rejected**

  Training increased typing speed

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 $\bar{d}=-3.5$ Standard deviation $s_d = 3.16$

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

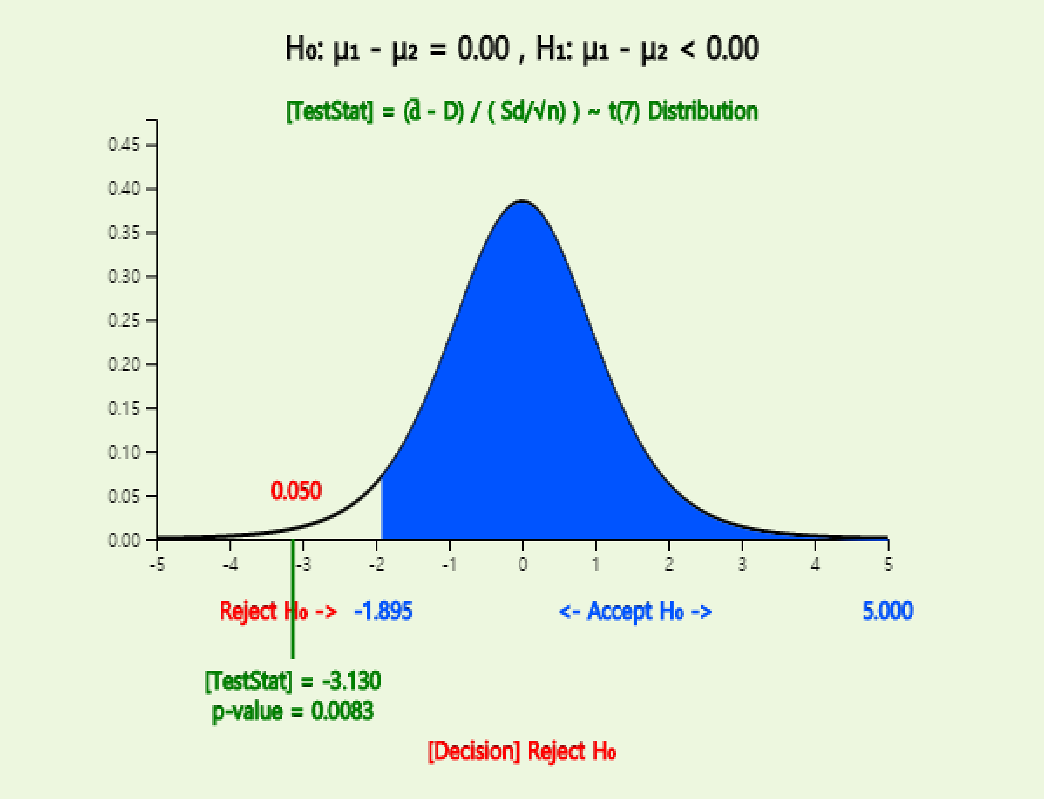
[Sample Data] *Input either sample data using BSV or sample statistics at the next boxes*

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>

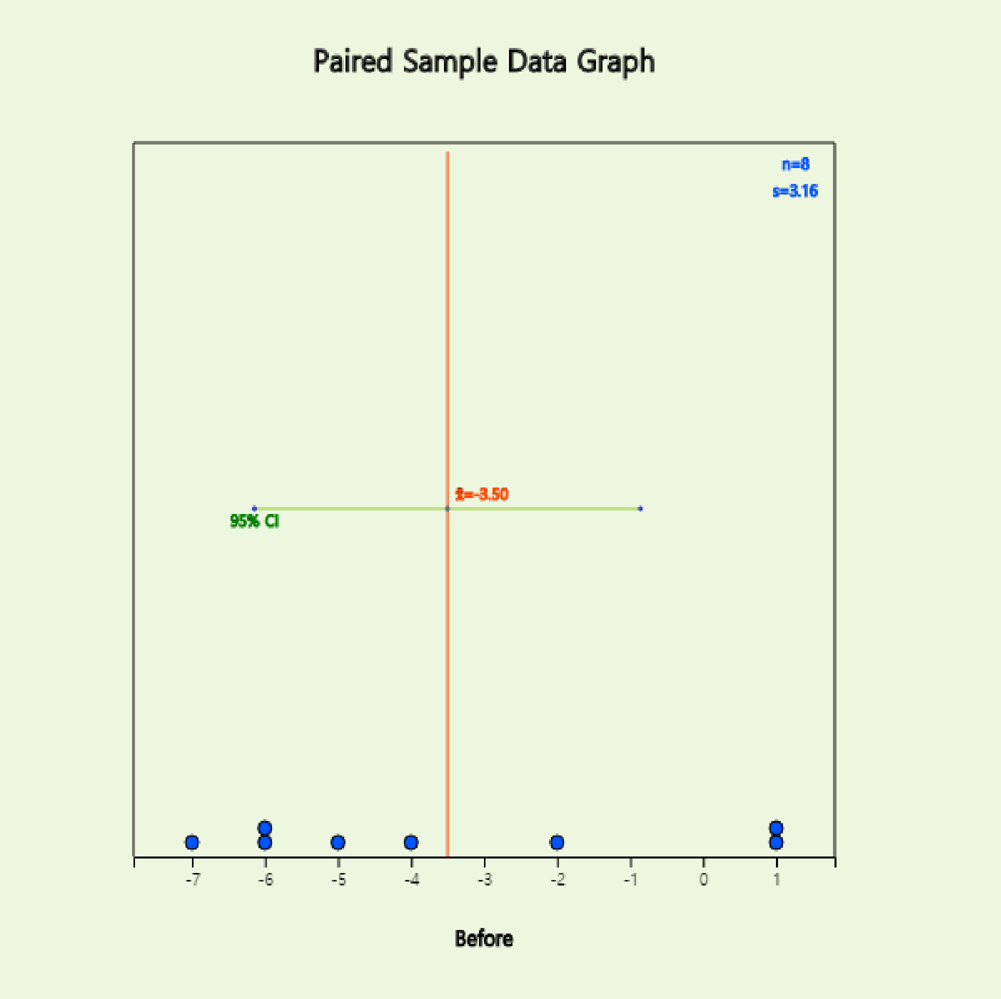
File: Ex814TypingSpeedEducation.csv

Analysis Var: 1: Before (Selected data: Raw Data)

by Group: 2: After (or Paired Var)

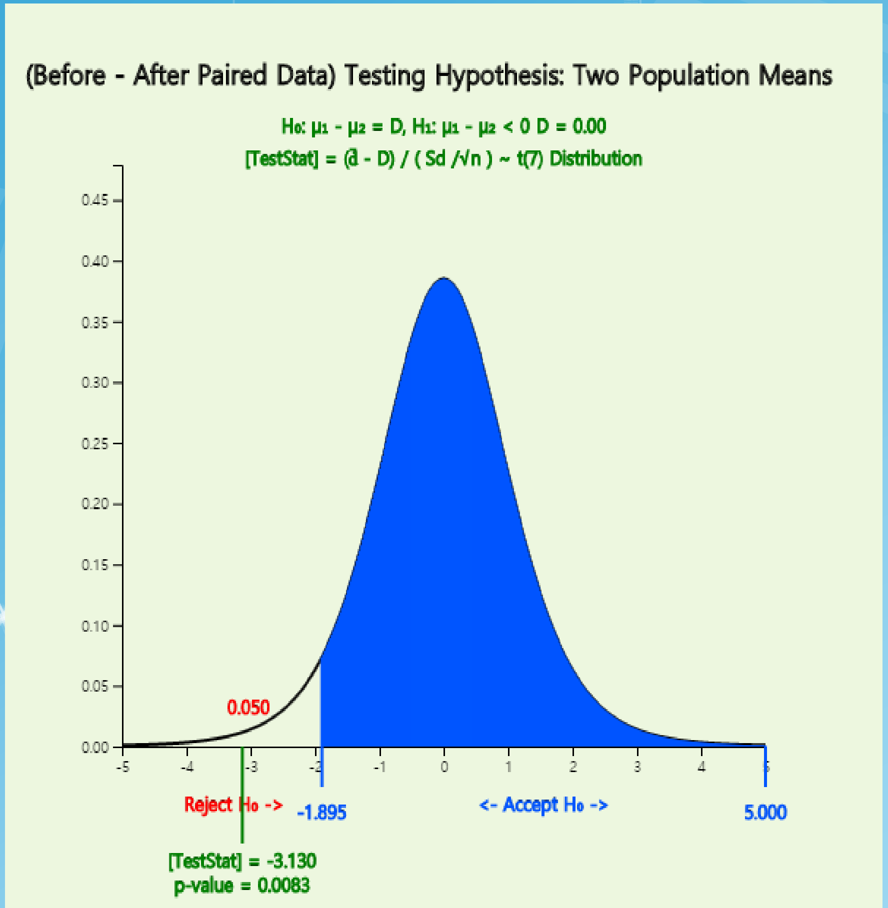
SelectedVar: V1 by V2,

	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	$\mu_1 - \mu_2$ 95% Confidence Interval
$H_1 : \mu_1 - \mu_2 < D$	0.00	Difference of Sample Means	-3.130	0.0083	(-6.144, -0.856)

Thank you