Section 7.4

Paired T-Test

7.4 Paired Comparisons:

- In this section, we are interested in comparing the means of two related (non-independent/dependent) normal populations.
- In other words, we wish to make statistical inference for the difference between the means of two related normal populations.
- Paired t-Test concerns about testing the equality of the means of two related normal populations.

Examples of related populations are:

1. Height of the father and height of his son.
2. Mark of the student in MATH and his mark in STAT.
3. Pulse rate of the patient before and after the medical treatment.
4. Hemoglobin level of the patient before and after the medical treatment.

Example: (effectiveness of a diet program)
Suppose that we are interested in studying the effectiveness of a certain diet program. Let the random variables X and Y are as follows:

X = the weight of the individual before the diet program

Y= the weight of the same individual after the diet program

We assume that the distributions of these random variables are normal with means μ_1 and μ_2, respectively.

These two variables are related (dependent/non-independent) because they are measured on the same individual.

Populations:

1-st population (X): weights before a diet program

mean = μ_1

2-nd population (Y): weights after the diet program

mean = μ_2

Confidence Interval for the Difference between the Means of Two Related Normal Populations ($\mu_D = \mu_1 - \mu_2$):

In this section, we consider constructing a confidence interval for the difference between the means of two related (non-independent) normal populations. As before, let us define the difference between the two means as follows:

$$\mu_D = \mu_1 - \mu_2$$

where μ_1 is the mean of the first population and μ_2 is the mean of the second population. We assume that the two normal populations are not independent.

Result:

A $(1-\alpha)100\%$ confidence interval for $\mu_D = \mu_1 - \mu_2$ is:

$$\bar{D} \pm t_{1-\frac{\alpha}{2}} \frac{S_D}{\sqrt{n}}$$

$$\bar{D} - t_{1-\frac{\alpha}{2}} \frac{S_D}{\sqrt{n}} < \mu_D < \bar{D} + t_{1-\frac{\alpha}{2}} \frac{S_D}{\sqrt{n}}$$

where:

$$\bar{D} = \frac{\sum_{i=1}^{n} D_i}{n}, \quad S_D = \sqrt{S_D^2}, \quad S_D^2 = \frac{\sum_{i=1}^{n} (D_i - \bar{D})^2}{n-1}, \quad df = \nu = n-1.$$

where:

$$\mu_D = \mu_1 - \mu_2$$

- We calculate the following quantities:
  - The differences (D-observations):

$$D_i = X_i - Y_i \quad (i=1, 2, \ldots, n)$$

  - Sample mean of the D-observations (differences):

$$\bar{D} = \frac{\sum_{i=1}^{n} D_i}{n} = \frac{D_1 + D_2 + \ldots + D_n}{n}$$

  - Sample variance of the D-observations (differences):

$$S_D^2 = \frac{\sum_{i=1}^{n} (D_i - \bar{D})^2}{n-1} = \frac{(D_1 - \bar{D})^2 + (D_2 - \bar{D})^2 + \cdots + (D_n - \bar{D})^2}{n-1}$$

  - Sample standard deviation of the D-observations:

$$S_D = \sqrt{S_D^2}$$

- We select a random sample of n individuals. At the beginning of the study, we record the individuals' weights before the diet program (X). At the end of the diet program, we record the individuals' weights after the program (Y). We end up with the following information and calculations:

Individual i	Weight before X_i	Weight after Y_i	Difference $D_i = X_i - Y_i$
1	X_1	Y_1	$D_1 = X_1 - Y_1$
2	X_2	Y_2	$D_2 = X_2 - Y_2$
. .	. .	. .	

Example:

Consider the data given in the previous numerical example:

Individual (i)	1	2	3	4	5	6	7	8	9	10
Weight before (X_i)	86.6	80.2	91.5	80.6	82.3	81.9	88.4	85.3	83.1	82.1
Weight after (Y_i)	79.7	85.9	81.7	82.5	77.9	85.8	81.3	74.7	68.3	69.7

Find a 95% confidence interval for the difference between the mean of weights before the diet program (μ_1) and the mean of weights after the diet program (μ_2).

Solution:

Calculations:

i	X_i	Y_i	$D_i = X_i - Y_i$
1	86.6	79.7	6.9
2	80.2	85.9	-5.7
3	91.5	81.7	9.8
4	80.6	82.5	-1.9
5	82.3	77.9	4.4
6	81.9	85.8	-3.9
7	88.4	81.3	7.1
8	85.3	74.7	10.6
9	83.1	68.3	14.8
10	82.1	69.7	12.4
sum	$\sum X$=842	$\sum Y$=787.5	$\sum D$=54.5

$$\overline{D} = \frac{\sum_{i=1}^{n} D_i}{n} = \frac{54.5}{10} = 5.45$$

$$S_D^2 = \frac{\sum_{i=1}^{n}(D_i - \overline{D})^2}{n-1} = \frac{(6.9-5.45)^2 + \ldots + (12.4-5.45)^2}{10-1} = 50.3283$$

$$S_D = \sqrt{S_D^2} = \sqrt{50.3283} = 7.09$$

We need to find a 95% confidence interval for $\mu_D = \mu_1 - \mu_2$:

$$\overline{D} \pm t_{1-\frac{\alpha}{2}} \frac{S_D}{\sqrt{n}}$$

We have found:

$$\overline{D} = 5.45 \quad , \quad S_D^2 = 50.3283 \quad , \quad S_D = \sqrt{S_D^2} = 7.09$$

The value of the reliability coefficient $t_{1-\frac{\alpha}{2}}$ $(df = \nu = n-1 = 9)$ is

$t_{1-\frac{\alpha}{2}} = t_{0.975} = 2.262$.

Therefore, a 95% confidence interval for $\mu_D = \mu_1 - \mu_2$ is

$$5.45 \pm (2.262)\frac{7.09}{\sqrt{10}}$$

$$5.45 \pm 5.0715$$

$$0.38 < \mu_D < 10.52$$

$$0.38 < \mu_1 - \mu_2 < 10.52$$

Testing hypotheses for two related population

1. Hypotheses:

we have three cases

Case I :

H_0: $\mu 1 = \mu 2 \rightarrow \mu_1 - \mu_2 = 0 \rightarrow \mu_D = 0$

H_A: $\mu_1 \neq \mu 2 \rightarrow \mu_1 - \mu_2 \neq 0 \rightarrow \mu_D \neq 0$

Case II :

H_0: $\mu 1 \leq \mu 2 \rightarrow \mu_1 - \mu_2 \leq 0 \rightarrow \mu_D \leq 0$

H_A: $\mu_1 > \mu 2 \rightarrow \mu_1 - \mu_2 > 0 \rightarrow \mu_D > 0$

Case III :

H_0: $\mu 1 \geq \mu 2 \rightarrow \mu_1 - \mu_2 \geq 0 \rightarrow \mu_D \geq 0$

H_A: $\mu_1 < \mu 2 \rightarrow \mu_1 - \mu_2 < 0 \rightarrow \mu_D < 0$

2.Test statistic (T.S)

$$T = \frac{\overline{D}}{S_D / \sqrt{n}}$$

where:

$$\mu_D = \mu_1 - \mu_2$$

- We calculate the following quantities:
  - The differences (D-observations):
    $$D_i = X_i - Y_i \quad (i=1, 2, \ldots, n)$$
  - Sample mean of the D-observations (differences):
    $$\overline{D} = \frac{\sum_{i=1}^{n} D_i}{n} = \frac{D_1 + D_2 + \ldots + D_n}{n}$$
  - Sample variance of the D-observations (differences):
    $$S_D^2 = \frac{\sum_{i=1}^{n} (D_i - \overline{D})^2}{n-1} = \frac{(D_1 - \overline{D})^2 + (D_2 - \overline{D})^2 + \cdots + (D_n - \overline{D})^2}{n-1}$$
  - Sample standard deviation of the D-observations:
    $$S_D = \sqrt{S_D^2}$$

3. Rejection Region(R.R):

As mentioned before.

(Use T-table where df =n-1)

4.Decision:

Reject H_0 or accept H_0 .

Numerical Example:

In the previous example, suppose that the sample size was 10 and the data were as follows:

Individual (i)	1	2	3	4	5	6	7	8	9	10
Weight before (X_i)	86.6	80.2	91.5	80.6	82.3	81.9	88.4	85.3	83.1	82.1
Weight after (Y_i)	79.7	85.9	81.7	82.5	77.9	85.8	81.3	74.7	68.3	69.7

Does these data provide sufficient evidence to allow us to conclude that the diet program is effective? Use α=0.05 and assume that the populations are normal.

Solution:

μ_1 = the mean of weights before the diet program

μ_2 = the mean of weights after the diet program

Hypotheses:

H_o: $\mu_1 = \mu_2$ (H_o: the diet program is not effective)

H_A: $\mu_1 \neq \mu_2$ (H_A: the diet program is effective)

Equivalently,

H_o: $\mu_D = 0$

H_A: $\mu_D \neq 0$ (where: $\mu_D = \mu_1 - \mu_2$)

We have found:

$\overline{D} = 5.45$, $S_D^2 = 50.3283$, $S_D = \sqrt{S_D^2} = 7.09$

Degrees of freedom:

$$df = \nu = n-1 = 10-1 = 9$$

Significance level: $\alpha = 0.05$

Rejection Region of H_o:

Critical values: $t_{0.025} = -2.262$ and $t_{0.975} = -t_{0.025} = 2.262$

Critical Region: $t < -2.262$ or $t > 2.262$