

Chapter Twenty One: Paired Data

A teacher is testing out a new program to help high school students learn math. A random sample of 40 students was selected. The teacher wants to determine if the new program helps the students perform better on exams. Consider the following two scenarios:

1. The students were randomly divided into two groups, each of size 20. One group of students were enrolled in the new program, and the other group was not. All students took the same exam.
2. All 40 students were enrolled in the new program. Each student completed the same math exam before entering the program and another (very similar) exam after completing the program.

What is different about the two studies?

How would we proceed to analyze each of these studies?

For the first scenario:

-

-

For the second scenario:

-

-

Paired Data

Sometimes we collect data from two groups that are dependent on one another. We refer to this as _____. For example:

-

-

The general data structure of paired data is as follows:

Pair	Group 1	Group 2	Differences
1	y_{11}	y_{21}	$d_1 = y_{11} - y_{21}$
2	y_{12}	y_{22}	$d_2 = y_{12} - y_{22}$
\vdots	\vdots	\vdots	\vdots
n	y_{1n}	y_{2n}	$d_n = y_{1n} - y_{2n}$
Means	\bar{y}_1	\bar{y}_2	\bar{d}
SDs	s_1	s_2	s_d

We define \bar{d} and s_d as follows

Now let's look at a few examples of scenarios of experiments and determine whether it is paired data:

1. We would like to determine if students taking an ACT prep course will score better than students not taking the course. A random sample of 25 students was chosen who took the course and a random sample of another 25 students was chosen who did not take the course. At the end of the prep course, both groups were given the ACT. Is there paired data, two independent sample data, or one sample data?
2. We would like to determine if students can improve their ACT score by taking a prep course. A random sample of 25 students was chosen. They first took the ACT test. Then they spent 6 weeks taking the prep course. At the end of the 6 weeks, they took the ACT test again. Is there paired data, two independent sample data, or one sample data?
3. We would like to determine ACT scores for students that take a prep course. A random sample of 50 students was selected. All 50 students took the prep course. At the end of the prep course, students took the ACT test. Is there paired data, two independent sample data, or one sample data?

Notation for Paired Data

μ_D :

n :

\bar{d} :

s_D :

Inference for Paired Data

Like the last two chapters we will consider two types of inference: Confidence Intervals and Hypothesis tests. The process for both of these types of inferences are very similar to previous chapters.

Assumptions for CIs and Hypothesis Tests

The following assumptions need to be met in order to construct a confidence interval or perform a hypothesis test for paired data.

- i. **Randomization Condition:**
- ii. **10% Condition:**
- iii. **Nearly Normal Condition:**

Confidence Interval

If the assumptions are met, then the confidence interval for paired data can be computed using the formula below. Note how similar it is for CIs for one sample mean:

Paired Data

One Sample Mean

In both cases, t^* has a t distribution with $n - 1$ degrees of freedom.

Hypothesis Test Procedure

Step 1: Hypotheses

- Null Hypothesis:
- Alternative Hypothesis:

Step 2: Assumptions See previous pages.

Step 3: Test Statistic

Calculate the test statistic below (or find with JMP). Note the similarity to the Chapter 18 one sample mean formula.

Paired Data:

One Sample Mean:

Step 4: Find P-value The p-value is found using a t-distribution with $n-1$ degrees of freedom. Remember, the way the p-value is found depends on the alternative hypothesis.

-

-

-

Step 5: List your Decision

<u>P-value</u>	<u>Evidence (against H_0)</u>
Greater than .10	Little to no evidence
Between .05 and .10	Weak evidence
Between .01 and .05	Moderate Evidence
Less than .01	Strong evidence

Step 6: Conclusion

Make a statement about μ_D (the population difference in means with pairing) given the information from the hypothesis test.

Make sure to include:

-

-

-

A teacher is testing out a new program to help high school students learn math. A random sample of 40 students was selected. The teacher wants to determine if the new program helps the students perform better on exams. Each student completed the same math exam before entering the program and another (very similar) exam after completing the program. Construct a 95% confidence interval and conduct a hypothesis test to see if there is a difference in score before and after completing the program, and if so how much.

