

10.6: Matched or Paired Samples

In most cases of economic or business data we have little or no control over the process of how the data are gathered. In this sense the data are not the result of a planned controlled experiment. In some cases, however, we can develop data that are part of a controlled experiment. This situation occurs frequently in quality control situations. Imagine that the production rates of two machines built to the same design, but at different manufacturing plants, are being tested for differences in some production metric such as speed of output or meeting some production specification such as strength of the product. The test is the same in format to what we have been testing, but here we can have matched pairs for which we can test if differences exist. Each observation has its matched pair against which differences are calculated. First, the differences in the metric to be tested between the two lists of observations must be calculated, and this is typically labeled with the letter "d." Then, the average of these matched differences, \bar{x}_d is calculated as is its standard deviation, s_d . We expect that the standard deviation of the differences of the matched pairs will be smaller than unmatched pairs because presumably fewer differences should exist because of the correlation between the two groups.

When using a hypothesis test for matched or paired samples, the following characteristics may be present:

1. Simple random sampling is used.
2. Sample sizes are often small.
3. Two measurements (samples) are drawn from the same pair of individuals or objects.
4. Differences are calculated from the matched or paired samples.
5. The differences form the sample that is used for the hypothesis test.
6. Either the matched pairs have differences that come from a population that is normal or the number of differences is sufficiently large so that distribution of the sample mean of differences is approximately normal.

In a hypothesis test for matched or paired samples, subjects are matched in pairs and differences are calculated. The differences are the data. The population mean for the differences, μ_d , is then tested using a Student's-t test for a single population mean with $n-1$ degrees of freedom, where n is the number of differences, that is, the number of pairs not the number of observations.

The null and alternative hypotheses for this test are:

$$H_a : \mu_d \neq 0$$

the test statistic is

$$t_c = \frac{\bar{x}_d - \mu_d}{\left(\frac{s_d}{\sqrt{n}}\right)}$$

? Exercise 10.6.1

A company has developed a training program for its entering employees because they have become concerned with the results of the six-month employee review. They hope that the training program can result in better six-month reviews. Each trainee constitutes a "pair", the entering score the employee received when first entering the firm and the score given at the six-month review. The difference in the two scores were calculated for each employee and the means for before and after the training program was calculated. The sample mean before the training program was 20.4 and the sample mean after the training program was 23.9. The standard deviation of the differences in the two scores across the 20 employees was 3.8 points. Test at the 10% significance level the null hypothesis that the two population means are equal against the alternative that the training program helps improve the employees' scores.

Answer

The first step is to identify this as a two sample case: before the training and after the training. This differentiates this problem from simple one sample issues. Second, we determine that the two samples are "paired." Each observation in the first sample has a paired observation in the second sample. This information tells us that the null and alternative hypotheses should be:

$$H_0 : \mu_d \leq 0 \quad (10.6.1)$$

$$H_a : \mu_d > 0 \quad (10.6.2)$$

This form reflects the implied claim that the training course improves scores; the test is one-tailed and the claim is in the alternative hypothesis. Because the experiment was conducted as a matched paired sample rather than simply taking scores from people who took the training course those who didn't, we use the matched pair test statistic:

$$\text{Test Statistic: } t_c = \frac{\bar{X}_d - \mu_d}{\frac{S_d}{\sqrt{n}}} = \frac{(23.9 - 20.4) - 0}{\left(\frac{3.8}{\sqrt{20}}\right)} = 4.12 \quad (10.6.3)$$

In order to solve this equation, the individual scores, pre-training course and post-training course need to be used to calculate the individual differences. These scores are then averaged and the average difference is calculated:

$$\bar{X}_d = \bar{x}_1 - \bar{x}_2 \quad (10.6.4)$$

From these differences we can calculate the standard deviation across the individual differences:

$$S_d = \frac{\sum (d_i - \bar{X}_d)^2}{n - 1} \text{ where } d_i = x_{1i} - x_{2i} \quad (10.6.5)$$

We can now compare the calculated value of the test statistic with the critical value. The critical value is a Student's t with degrees of freedom equal to the number of pairs, not observations, minus 1. In this case 20 pairs and at 90% confidence level $t_{\alpha/2} = \pm 1.729$ at $df = 20 - 1 = 19$. The calculated test statistic is most certainly in the tail of the distribution and thus we cannot accept the null hypothesis that there is no difference from the training program. Evidence seems indicate that the training aids employees in gaining higher scores.

? Exercise 10.6.1

A study was conducted to investigate the effectiveness of hypnotism in reducing pain. Results for randomly selected subjects are shown in the table below. A lower score indicates less pain. The "before" value is matched to an "after" value and the differences are calculated. Are the sensory measurements, on average, lower after hypnotism? Test at a 5% significance level.

Subject:	A	B	C	D	E	F	G	H
Before	6.6	6.5	9.0	10.3	11.3	8.1	6.3	11.6
After	6.8	2.4	7.4	8.5	8.1	6.1	3.4	2.0

Answer

Corresponding "before" and "after" values form matched pairs. (Calculate "after" – "before.")

After data	Before data	Difference
6.8	6.6	0.2
2.4	6.5	-4.1
7.4	9	-1.6
8.5	10.3	-1.8
8.1	11.3	-3.2

After data	Before data	Difference
6.1	8.1	-2
3.4	6.3	-2.9
2	11.6	-9.6

The data **for the test** are the differences: $\{0.2, -4.1, -1.6, -1.8, -3.2, -2, -2.9, -9.6\}$

The sample mean and sample standard deviation of the differences are: $\bar{x}_d = -3.13$ and $s_d = 2.91$. Verify these values. Let μ_d be the population mean for the differences. We use the subscript d to denote "differences."

Random variable: $\bar{X}_d =$ the mean difference of the sensory measurements

$$H_0 : \mu_d \geq 0 \quad (10.6.6)$$

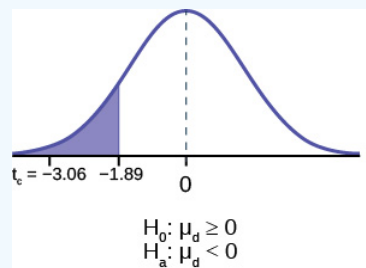
The null hypothesis is zero or positive, meaning that there is the same or more pain felt after hypnotism. That means the subject shows no improvement. μ_d is the population mean of the differences.)

$$H_a : \mu_d < 0 \quad (10.6.7)$$

The alternative hypothesis is negative, meaning there is less pain felt after hypnotism. That means the subject shows improvement. The score should be lower after hypnotism, so the difference ought to be negative to indicate improvement.

Distribution for the test: The distribution is a Student's t with $df = n - 1 = 8 - 1 = 7$. Use t_7 . (Notice that the test is for a single population mean.)

Calculate the test statistic and look up the critical value using the Student's- t distribution: The calculated value of the test statistic is 3.06 and the critical value of the t -distribution with 7 degrees of freedom at the 5% level of confidence is 1.895 with a one-tailed test.



\bar{X}_d is the random variable for the differences.

The sample mean and sample standard deviation of the differences are:

$$\begin{aligned} \bar{x}_d &= -3.13 \\ \bar{s}_d &= 2.91 \end{aligned} \quad (10.6.8)$$

Compare the critical value for alpha against the calculated test statistic.

The conclusion from using the comparison of the calculated test statistic and the critical value will give us the result. In this question the calculated test statistic is 3.06 and the critical value is 1.895. The test statistic is clearly in the tail and thus we cannot accept the null hypothesis that there is no difference between the two situations, hypnotized and not hypnotized.

Make a decision: Cannot accept the null hypothesis, H_0 . This means that $\mu_d < 0$ and there is a statistically significant improvement.

Conclusion: At a 5% level of significance, from the sample data, there is sufficient evidence to conclude that the sensory measurements, on average, are lower after hypnotism. Hypnotism appears to be effective in reducing pain.

✓ Example 10.6.1

A college football coach was interested in whether the college's strength development class increased his players' maximum lift (in pounds) on the bench press exercise. He asked four of his players to participate in a study. The amount of weight they could each lift was recorded before they took the strength development class. After completing the class, the amount of weight they could each lift was again measured. The data are as follows:

Weight (in pounds)	Player 1	Player 2	Player 3	Player 4
Amount of weight lifted prior to the class	205	241	338	368
Amount of weight lifted after the class	295	252	330	360

The coach wants to know if the strength development class makes his players stronger, on average.

Record the **differences** data. Calculate the differences by subtracting the amount of weight lifted prior to the class from the weight lifted after completing the class. The data for the differences are: {90, 11, -8, -8}.

$$\bar{x}_d = 21.3, s_d = 46.7 \quad (10.6.9)$$

Using the difference data, this becomes a test of a single mean.

Define the random variable: \bar{X}_d mean difference in the maximum lift per player.

The distribution for the hypothesis test is a student's t with 3 degrees of freedom.

$$H_0 : \mu_d \leq 0, H_a : \mu_d > 0 \quad (10.6.10)$$

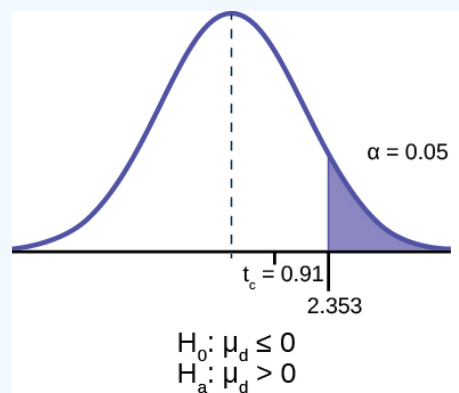


Figure 10.10

Calculate the test statistic look up the critical value: Critical value of the test statistic is 0.91. The critical value of the student's t at 5% level of significance and 3 degrees of freedom is 2.353.

Decision: If the level of significance is 5%, we cannot reject the null hypothesis, because the calculated value of the test statistic is not in the tail.

What is the conclusion?

At a 5% level of significance, from the sample data, there is not sufficient evidence to conclude that the strength development class helped to make the players stronger, on average.

Solution

Add example text here.