

9.3: Comparison of Two Population Means - Paired Samples

Learning Objectives

- To learn the distinction between independent samples and paired samples.
- To learn how to construct a confidence interval for the difference in the means of two distinct populations using paired samples.
- To learn how to perform a test of hypotheses concerning the difference in the means of two distinct populations using paired samples

Suppose chemical engineers wish to compare the fuel economy obtained by two different formulations of gasoline. Since fuel economy varies widely from car to car, if the mean fuel economy of two independent samples of vehicles run on the two types of fuel were compared, even if one formulation were better than the other the large variability from vehicle to vehicle might make any difference arising from difference in fuel difficult to detect. Just imagine one random sample having many more large vehicles than the other. Instead of independent random samples, it would make more sense to select pairs of cars of the same make and model and driven under similar circumstances, and compare the fuel economy of the two cars in each pair. Thus the data would look something like Table 9.3.1, where the first car in each pair is operated on one formulation of the fuel (call it Type 1 gasoline) and the second car is operated on the second (call it Type 2 gasoline).

Table 9.3.1: Fuel Economy of Pairs of Vehicles

Make and Model	Car 1	Car 2
Buick LaCrosse	17.0	17.0
Dodge Viper	13.2	12.9
Honda CR-Z	35.3	35.4
Hummer H 3	13.6	13.2
Lexus RX	32.7	32.5
Mazda CX-9	18.4	18.1
Saab 9-3	22.5	22.5
Toyota Corolla	26.8	26.7
Volvo XC 90	15.1	15.0

The first column of numbers form a sample from Population 1, the population of all cars operated on Type 1 gasoline; the second column of numbers form a sample from Population 2, the population of all cars operated on Type 2 gasoline. It would be incorrect to analyze the data using the formulas from the previous section, however, since the samples were not drawn independently. What is correct is to compute the difference in the numbers in each pair (subtracting in the same order each time) to obtain the third column of numbers as shown in Table 9.3.2 and treat the differences as the data. At this point, the new sample of differences $d_1 = 0.0, \dots, d_9 = 0.1$ in the third column of Table 9.3.2 may be considered as a random sample of size $n = 9$ selected from a population with mean $\mu_d = \mu_1 - \mu_2$. This approach essentially transforms the paired two-sample problem into a one-sample problem as discussed in the previous two chapters.

Table 9.3.2: Fuel Economy of Pairs of Vehicles


Make and Model	Car 1	Car 2	Difference
Buick LaCrosse	17.0	17.0	0.0
Dodge Viper	13.2	12.9	0.3
Honda CR-Z	35.3	35.4	-0.1
Hummer H 3	13.6	13.2	0.4

Make and Model	Car 1	Car 2	Difference
Lexus RX	32.7	32.5	0.2
Mazda CX-9	18.4	18.1	0.3
Saab 9-3	22.5	22.5	0.0
Toyota Corolla	26.8	26.7	0.1
Volvo XC 90	15.1	15.0	0.1

Note carefully that although it does not matter what order the subtraction is done, it must be done in the same order for all pairs. This is why there are both positive and negative quantities in the third column of numbers in Table 9.3.2.

Confidence Intervals

When the population of differences is normally distributed the following formula for a confidence interval for $\mu_d = \mu_1 - \mu_2$ is valid.

 100(1 - α)% Confidence Interval for the Difference Between Two Population Means: Paired Difference Samples

$$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}$$

where there are n pairs, \bar{d} is the mean and s_d is the standard deviation of their differences.

The number of degrees of freedom is

$$df = n - 1.$$

The population of differences must be *normally distributed*.

✓ Example 9.3.1

Using the data in Table 9.3.1 construct a point estimate and a 95% confidence interval for the difference in average fuel economy between cars operated on Type 1 gasoline and cars operated on Type 2 gasoline.

Solution

We have referred to the data in Table 9.3.1 because that is the way that the data are typically presented, but we emphasize that with paired sampling one immediately computes the differences, as given in Table 9.3.2, and uses the differences as the data.

The mean and standard deviation of the differences are

$$\bar{d} = \frac{\sum d}{n} = \frac{1.3}{9} = 0.14$$

$$s_d = \sqrt{\frac{\sum d^2 - \frac{1}{n}(\sum d)^2}{n-1}} = \sqrt{\frac{0.41 - \frac{1}{9}(1.3)^2}{8}} = 0.16$$

The point estimate of $\mu_1 - \mu_2 = \mu_d$ is

$$\bar{d} = 0.14$$

In words, we estimate that the average fuel economy of cars using Type 1 gasoline is 0.14 mpg greater than the average fuel economy of cars using Type 2 gasoline.

To apply the formula for the confidence interval, we must find $t_{\alpha/2}$. The 95% confidence level means that $\alpha = 1 - 0.95 = 0.05$ so that $t_{\alpha/2} = t_{0.025}$. From Figure 7.1.6, in the row with the heading $df = 9 - 1 = 8$ we read that $t_{0.025} = 2.306$. Thus

$$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}} = 0.14 \pm 2.306 \left(\frac{0.16}{\sqrt{9}} \right) \approx 0.14 \pm 0.13$$

We are 95% confident that the difference in the population means lies in the interval $[0.01, 0.27]$ in the sense that in repeated sampling 95% of all intervals constructed from the sample data in this manner will contain $\mu_d = \mu_1 - \mu_2$. Stated differently, we are 95% confident that mean fuel economy is between 0.01 and 0.27 mpg greater with Type 1 gasoline than with Type 2 gasoline.

Hypothesis Testing

Testing hypotheses concerning the difference of two population means using paired difference samples is done precisely as it is done for independent samples, although now the null and alternative hypotheses are expressed in terms of μ_d instead of $\mu_1 - \mu_2$. Thus the null hypothesis will always be written

$$H_0 : \mu_d = D_0$$

The three forms of the alternative hypothesis, with the terminology for each case, are:

Form of H_a	Terminology
$H_a : \mu_d < D_0$	Left-tailed
$H_a : \mu_d > D_0$	Right-tailed
$H_a : \mu_d \neq D_0$	Two-tailed

The same conditions on the population of differences that was required for constructing a confidence interval for the difference of the means must also be met when hypotheses are tested. Here is the standardized test statistic that is used in the test.

Standardized Test Statistic for Hypothesis Tests Concerning the Difference Between Two Population Means: Paired Difference Samples

$$T = \frac{\bar{d} - D_0}{s_d / \sqrt{n}}$$

where there are n pairs, \bar{d} is the mean and s_d is the standard deviation of their differences.

The test statistic has Student's t -distribution with $df = n - 1$ degrees of freedom.

The population of differences must be normally distributed.

✓ Example 9.3.2: using the critical value approach

Using the data of Table 9.3.2 test the hypothesis that mean fuel economy for Type 1 gasoline is greater than that for Type 2 gasoline against the null hypothesis that the two formulations of gasoline yield the same mean fuel economy. Test at the 5% level of significance using the critical value approach.

Solution

The only part of the table that we use is the third column, the differences.

- Step 1.** Since the differences were computed in the order Type 1 mpg – Type 2 mpg, better fuel economy with Type 1 fuel corresponds to $\mu_d = \mu_1 - \mu_2 > 0$. Thus the test is

$$\begin{aligned} H_0 : \mu_d &= 0 \\ \text{vs.} \\ H_a : \mu_d &> 0 @ \alpha = 0.05 \end{aligned}$$

(If the differences had been computed in the opposite order then the alternative hypotheses would have been $H_a : \mu_d < 0$.)

- Step 2.** Since the sampling is in pairs the test statistic is

$$T = \frac{\bar{d} - D_0}{s_d / \sqrt{n}}$$

- **Step 3.** We have already computed \bar{d} and s_d in the previous example. Inserting their values and $D_0 = 0$ into the formula for the test statistic gives

$$T = \frac{\bar{d} - D_0}{s_d/\sqrt{n}} = \frac{0.14}{0.16/\sqrt{3}} = 2.600$$

- **Step 4.** Since the symbol in H_a is ">" this is a right-tailed test, so there is a single critical value, $t_\alpha = t_{0.05}$ with 8 degrees of freedom, which from the row labeled $df = 8$ in Figure 7.1.6 we read off as 1.860. The rejection region is $[1.860, \infty)$
- **Step 5.** As shown in Figure 9.3.1 the test statistic falls in the rejection region. The decision is to reject H_0 . In the context of the problem our conclusion is:

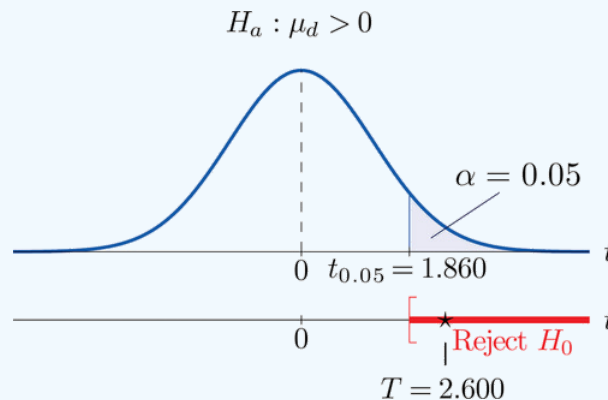


Figure 9.3.1: Rejection Region and Test Statistic for "Example 9.3.2"

The data provide sufficient evidence, at the 5% level of significance, to conclude that the mean fuel economy provided by Type 1 gasoline is greater than that for Type 2 gasoline.

✓ Example 9.3.3: using the p-value approach

Perform the test in Example 9.3.2 using the p-value approach.

Solution

The first three steps are identical to those 9.3.2.

- **Step 4.** Because the test is one-tailed the observed significance or p -value of the test is just the area of the right tail of Student's t -distribution, with 8 degrees of freedom, that is cut off by the test statistic $T = 2.600$. We can only approximate this number. Looking in the row of Figure 7.1.6 headed $df = 8$, the number 2.600 is between the numbers 2.306 and 2.896, corresponding to $t_{0.025}$ and $t_{0.010}$. The area cut off by $t = 2.306$ is 0.025 and the area cut off by $t = 2.896$ is 0.010. Since 2.600 is between 2.306 and 2.896 the area it cuts off is between 0.025 and 0.010. Thus the p -value is between 0.025 and 0.010. In particular it is less than 0.025. See Figure 9.3.2.

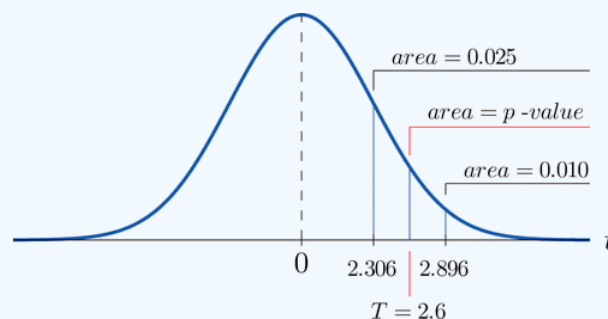


Figure 9.3.2: P-Value for "Example 9.3.3"

- **Step 5.** Since $0.025 < 0.05$, $p < \alpha$ so the decision is to reject the null hypothesis:

The data provide sufficient evidence, at the 5% level of significance, to conclude that the mean fuel economy provided by Type 1 gasoline is greater than that for Type 2 gasoline.

The paired two-sample experiment is a very powerful study design. It bypasses many unwanted sources of “statistical noise” that might otherwise influence the outcome of the experiment, and focuses on the possible difference that might arise from the one factor of interest.

If the sample is large (meaning that $n \geq 30$) then in the formula for the confidence interval we may replace $t_{\alpha/2}$ by $z_{\alpha/2}$. For hypothesis testing when the number of pairs is at least 30, we may use the same statistic as for small samples for hypothesis testing, except now it follows a standard normal distribution, so we use the last line of Figure 7.1.6 to compute critical values, and p -values can be computed exactly with Figure 7.1.5, not merely estimated using Figure 7.1.6.

Key Takeaway

- When the data are collected in pairs, the differences computed for each pair are the data that are used in the formulas.
- A confidence interval for the difference in two population means using paired sampling is computed using a formula in the same fashion as was done for a single population mean.
- The same five-step procedure used to test hypotheses concerning a single population mean is used to test hypotheses concerning the difference between two population means using pair sampling. The only difference is in the formula for the standardized test statistic.

This page titled [9.3: Comparison of Two Population Means - Paired Samples](#) is shared under a [CC BY-NC-SA 3.0](#) license and was authored, remixed, and/or curated by [Anonymous](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.