
LESSON: PAIRED T-TEST

This lesson includes an overview of the subject, instructor notes, and example exercises using Minitab.

Paired t -Test for Population Means

Lesson Overview

We will construct confidence intervals and conduct hypothesis tests on the true mean difference μ_d between two sets of paired observations. Two sets of sample data are considered to be **paired** or **dependent** when the observations are matched or linked in some way.

Prerequisites

This lesson requires knowledge from the **Normal Distribution, Sampling Distribution of \bar{X}** (e.g. Central Limit Theorem), and **Population Mean Inference for Normally Distributed Data and Unknown Population Standard Deviation** lessons. The **paired t -test**, which you can think of as the **dependent sample t -test**, is a statistical procedure used to determine whether the mean difference between two sets of observations is zero or some other target value. In a paired t -test, each subject or entity is measured twice, resulting in pairs of observations. When we have paired observations, a paired t -test is used to make inferences on the population mean difference. A **1-sample t -test on the paired differences** can be used rather than a paired t test. Not only is the 1-sample t -test on the differences conceptually easier to understand, we already know how to build confidence intervals and conduct hypothesis tests for one sample of data.

Learning Targets

This lesson teaches students to:

- Recognize paired (dependent) samples versus two independent samples.
- Calculate paired differences in Minitab.
- Use Minitab to construct a confidence interval and conduct a hypothesis test for paired differences.

Time Required

It will take the instructor at least 40 minutes to do this lesson. The most difficult part of this lesson for students is the difference between paired samples and independent (unpaired) samples. Distinguishing between paired and unpaired data always seems to create confusion. Once students can recognize paired data, then they can proceed to carry out confidence interval or hypothesis testing computations using a 1-sample t -test. In this lesson, they have an opportunity to review how to check for normally distributed data and conduct 1-sample t -test inference.

Materials Required

- Minitab desktop (20 or higher) or Minitab web app
- Minitab worksheet of sample data, entitled *PairedtTest_Lesson.mtw*

Assessment

The activity sheet contains exercises for students to assess their understanding of the learning targets for this lesson.

Possible Extensions

The instructor may want to do the lesson *Two-Sample t -Test for Population Means* after this lesson so that students can perform inference on two independent samples.

Instructor Notes with Examples

Paired Versus Independent Sample Data

Two sets of sample data are considered to be **paired** or **dependent** when the observations in one set are paired to observations in the other, and these two observations are “linked” with respect to some entity or item. The website [Stat Trek](#) has more eloquent definition, stating that:

Two data sets are **paired** when the following **one-to-one relationship** exists between values in the two data sets.

- Each data set has the same number of data points.
- Each data point in one data set is related to one, and only one, data point in the other data set.

We can sometimes detect pairing or dependency according to the way in which the data was collected. In a paired t -test, each subject or entity is measured twice, resulting in pairs of observations. Here are some examples of dependent samples.

- Students going to graduate school take the GRE twice, *once before* and *once after* a prep course. Each student has a before and after score.
- Each student in your class measures the length of their right ring finger and left ring finger. The finger lengths are paired by student.
- Ten identical cars are filled with 87 octane grade gas from the same station and their gas mileages are recorded. Then the same cars go back to the same station and are filled with 89 octane grade gas and again their gas mileages are recorded. The gas mileages are paired by car.

When we have two sets of data, ask students to determine if entry 1 in data set 1 is linked uniquely to entry 1 in data set 2. More generally, is entry i in data set 1 linked to entry i in data set 2, where $i = 1, 2, \dots, n$? If the answer is “yes,” then we have paired data. Otherwise, we have independent data. For example, below are two sets of data for the amount of weight (lbs.) that ten athletes can lift before and after three weeks of training. Consider **entry 2** (bolded below) that represents *Athlete 2*. Is the entry 230 in data set 1 linked to the entry 236 in data set 2? Yes! It’s the amount that *Athlete 2* bench pressed before and after the training sessions. Would it make sense to link *Athlete 2*’s bench pressing ability before the training, 230 lbs., with *Athlete 3*’s bench pressing ability after the training, 179 lbs.? No way! The data below is paired by athlete.

Athlete	Max Amt Bench Press Weight Before (lbs.)	Max Amt Bench Press Weight After (lbs.)
1	210	219
2	230	236
3	182	179
4	205	204
5	262	270
6	253	250
7	219	222
8	216	216
9	175	180
10	190	198

Example 1. For each of the following scenarios, determine if the two samples are independent or paired.

(a) A computer scientist is investigating the usefulness of two different design languages in improving programming tasks. Twelve expert programmers, familiar with both languages, are asked to code a standard function in both languages, and the time (in minutes) is recorded.

Solution: Paired. Each programmer does the same task twice with different programming languages.

(b) The melting temperature for two different types of plastic pipe is being investigated. A random sample of 15 pipe specimens of each type is tested, and the melting temperatures are recorded.

Solution: Independent. Notice that we have two random samples of 15 pipe specimens. Since there is no obvious connection between pipes of different types, these are independent.

(c) Two types of fish attractors, one made from vitrified clay pipes and the other from cement blocks and brush, were used during the same 16 time periods spanning four years at a specific lake. The number of fish caught per fishing day in each of the time periods was recorded.

Solution: Paired. Here, the common element is the time period. For each of the 16 time periods, both types of fish attractors were used.

Confidence Intervals

When two populations have paired data, a paired t -test is used to construct confidence intervals about the population mean difference. As mentioned in this lesson's prerequisites, you can use a 1-sample t -test on their differences rather than a paired t -test. Both tests are doing the same thing, but conceptually, the test of their differences is easier to understand after covering the 1-sample t -test lesson.

Reasoning behind using a 1-sample t -test on the difference in population means

When data is **paired**, any hypothesis about $\mu_1 - \mu_2$ is equivalent to a hypothesis about μ_d , where d is used to indicate "difference." For example, testing whether $\mu_1 - \mu_2 = 3$ is the same as testing $\mu_d = 3$. Assuming that the differences are normally distributed, we can use a 1-sample t -test on $\mu_d = 3$. The reasoning behind why the 1-sample t -test is appropriate follows.

When we have paired data, entry i in data set 1 is paired with entry i in data set 2. For each pair, let x_{Bi} represent the "before" component of entry i and x_{Ai} represent the "after" component of entry i . For example, for bench pressing *Athlete 2*, $x_{B2} = 230$ and $x_{A2} = 236$.

Athlete	Max Amt Bench Press Weight Before (lbs.)	Max Amt Bench Press Weight After (lbs.)
1	210	219
2	230	236
...
10	190	198

We are interested in the difference of the “before” and “after” paired values. For each entry i , the difference d_i must be calculated using the same ordering of the pairs; that is, they must all be calculated as “after – before” or “before – after.” The differences d_i 's for paired data of sample size n are:

$$\begin{aligned}d_1 &= X_{A1} - X_{B1} \\d_2 &= X_{A2} - X_{B2} \\&\dots \\d_i &= X_{Ai} - X_{Bi} \\&\dots \\d_n &= X_{An} - X_{Bn}.\end{aligned}$$

An estimate for the true mean difference μ_d is the average or sample mean of the paired differences: $\bar{x}_d = \frac{\sum_{i=1}^n d_i}{n}$. The sample variance of the differences is $s_d^2 = \frac{\sum_{i=1}^n (d_i - \bar{x}_d)^2}{n-1}$. As long as the differences are normally distributed, we can use the confidence interval formula from the 1-sample t -test lesson to obtain a confidence interval formula for the population mean of paired differences μ_d .

Confidence Interval for Mean of Paired Differences

A **two-tailed 100(1 - α)% confidence interval for μ_d** when the differences are normally distributed and σ_d is unknown is:

$$\left[\bar{x}_d - t_{\alpha/2} \cdot \frac{s_d}{\sqrt{n}}, \bar{x}_d + t_{\alpha/2} \cdot \frac{s_d}{\sqrt{n}} \right]$$

\bar{x}_d = sample mean of the differences,

n = sample size,

s_d = sample standard deviation of the differences,

$t_{\alpha/2}$ = t -value from the t table corresponding to half of the desired α level with $n - 1$ degrees of freedom

Example 2 (Bench Pressing Differences). Ten athletes are randomly selected from a certain high school's soccer team, and they asked to record the maximum amount that they can bench press before and after three weeks of training. Construct a 90% confidence interval for the mean difference μ_d in the maximum amount that team members can bench press before and after three weeks of training. The entire table of data is on page 3.

Athlete	Max Amt Bench Press Weight Before (lbs.)	Max Amt Bench Press Weight After (lbs.)
1	210	219
2	230	236
...
10	190	198

Solution: Since the data is paired, we first want to determine the differences for each athlete. We can do this in Minitab by using its **Calculator**.

- 1 Choose **Calc > Calculator**.
- 2 In the **Store result in variable**, enter an unused column number (for example, *C4*) or a pre-defined column name (for example, *Difference: Bench Press Amounts*).
- 3 In **Expression**, enter variable names and operations. Here compute the Difference as '*Max Amt Bench Press After*' - '*Max Amt Bench Press Before*'
- 4 Click **OK**.

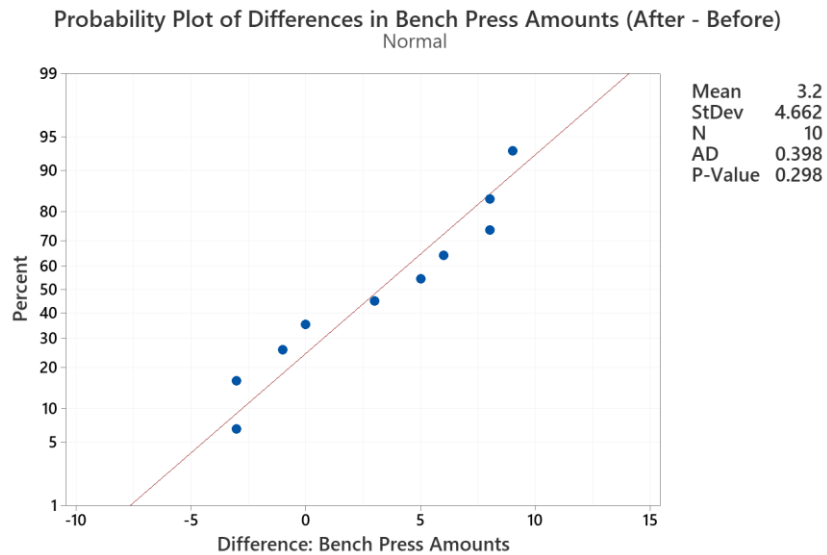
Now column *C4* contains the *After – Before* bench press data, labeled *Difference: Bench Press Amounts*.

C1	C2	C3	C4
Athlete	Max Amt Bench Press Before	Max Amt Bench Press After	Difference: Bench Press Amounts
1	210	219	9
2	230	236	6
3	182	179	-3
4	205	204	-1
5	262	270	8
6	253	250	-3
7	219	222	3
8	216	216	0
9	175	180	5
10	190	198	8

The sample size of the *Difference* data is $n = 10$, which is not large enough for us to assume that the distribution of the sample means is normally distributed. We need to use a **normality test** to determine if the *Difference* data is from a normal distribution. Recall the Minitab instructions for producing an AD test statistic and probability plot.

- 1 Choose **Stat > Basic Statistics > Normality Test**.
- 2 In the **Variable** text box, select the column containing the sample data of interest.
- 3 Under **Test for Normality**, the **Anderson-Darling** test is selected by default.
- 4 Click **OK**.

For the “Difference: Bench Press Amounts” column, Minitab returns the following graphic along with the AD test statistic and its p -value.



Given that the resulting p -value is 0.298, we would not reject that the “Difference: Bench Press Amounts” sample data comes from a normal distribution at level of significance $\alpha = 0.05$. Knowing that \bar{X} is normally distributed allows us to construct a 90% confidence interval for the mean difference μ_d in the maximum amount that team members can bench press before and after three weeks of training. Using Minitab, the 90% 1-Sample t confidence interval is:

Descriptive Statistics

N	Mean	StDev	SE Mean	90% CI for μ
10	3.20	4.66	1.47	(0.50, 5.90)

μ : population mean of Difference: Bench Press Amounts

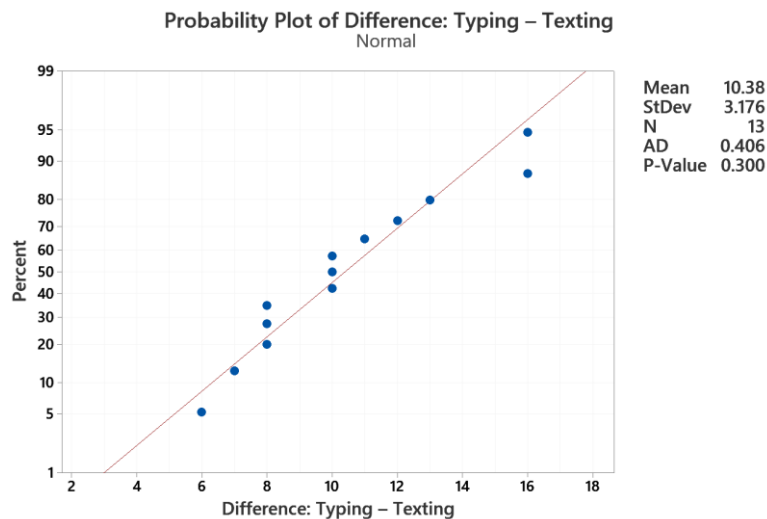
Before moving onto another example, it is worth asking the following question: *At significance level $\alpha = 0.1$, did the mean difference in the athletes’ maximum bench press amounts change after the three weeks of training?* Using the 90% confidence interval, we can say that we’re 90% certain that the interval (0.50, 5.90) contains the true mean difference in the maximum amount that the athletes can bench press from before to after the three weeks of training. At $\alpha = 0.10$, we have evidence to suggest that μ_d is not equal to 0, meaning the amount did change. Since we constructed the difference data as “after – before,” this means that they could lift more after the training rather than before.

Example 3 (Typing vs Texting Speed). A new study reported that people are now texting on their phones nearly as fast as they can type on a computer keyboard. The study consisted of 37,000 volunteers in 160 countries that took the “text” test for speed and accuracy. The average person texted 36 words per minute.

Can College A's incoming freshmen type on their computer keyboard *faster* than they text on their phone (using the two-thumb method)? A random sample of $n = 13$ freshmen are selected from College A and their texting and keyboard typing speeds per minute are recorded.

Freshman at College A	Keyboard Typing Speed	Text Typing Speed	Difference: Typing – Texting
Tom	52	39	13
George	51	44	7
Samantha	48	40	8
Heather	51	39	12
Troy	51	35	16
Michael	49	33	16
Tammy	49	41	8
Mary	47	37	10
John	50	44	6
Joe	50	39	11
Julia	49	41	8
Brenda	50	40	10
Patrick	47	37	10

Solution: Since the two speeds, *Keyboard Typing Speed* and *Text Typing Speed*, are paired by college freshman, we can use Minitab's **Calculator** to determine the column of differences, *Difference: Typing – Texting*. The sample size of the *Difference* data is $n = 13$, which is not large enough for us to assume that the distribution of the sample means is normally distributed. I used Minitab's **normality test** to determine that the *Difference* data is from a normal distribution since the *AD* test statistic's p -value of 0.300 is greater than $\alpha = 0.05$.



Knowing that \bar{X} is normally distributed allows us to construct a 1-Sample t confidence interval in Minitab for the mean difference μ_d in keyboard typing speed versus the text typing speed.

Descriptive Statistics

N	Mean	StDev	SE Mean	90% CI for μ
13	10.385	3.176	0.881	(8.814, 11.955)

μ : population mean of Difference: Typing – Texting



Since we are 90% certain that the interval (8.814, 11.955) words/minute contains the true mean difference μ_d for “typing – texting” speed, then at significance level $\alpha = 0.05$, the mean typing speed is *greater than* the mean texting speed. Why can we say *greater than* at $\alpha = 0.05$ after creating a two-sided 90% confidence interval? Because the “greater than” one-sided 95% confidence interval (8.814, $+\infty$) corresponds to the “two-sided” 90% confidence interval (8.814, 11.955). We know this because there is 0.05 area to the left of 8.814 and 0.05 area to the right of 11.955.

We can use Minitab and the **Options** in the **1-Sample t** dialog box to directly calculate the 95% one-sided lower confidence bound.

Descriptive Statistics

N	Mean	StDev	SE Mean	95% Lower Bound
				for μ
13	10.385	3.176	0.881	8.814

μ : population mean of Difference: Typing – Texting

At $\alpha = 0.05$, does the lower confidence bound also suggest that the mean difference in “typing – texting” speed is greater than 10 words/minute? No! Although the one-sided 95% confidence bound for μ_d contains values greater than 10, it also includes the value 10. Since we are 95% certain that the interval (8.814, $+\infty$) contains μ_d , it is possible for μ_d to be the value 10. We’ll return to this problem in the hypothesis testing section.

* **Reference:** Aalto University. (2019, October 2). Smartphone typing speeds catching up with keyboards. *ScienceDaily*. Retrieved July 8, 2020 from www.sciencedaily.com/releases/2019/10/191002075925.htm

Example 4 (Parallel Parking). A study was conducted in which $n = 45$ subjects were asked to parallel park each of two cars having different wheelbases and turning radii. For each subject, the time (in seconds) required to park each of the two automobiles was recorded. A summary of the data is available below.

	N	Mean	Std. Dev.	Median
Automobile 1	45	36.23	18.58	35.10
Automobile 2	45	28.01	19.45	25.50

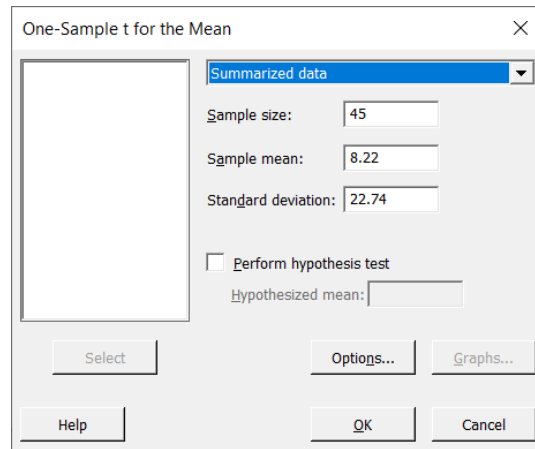
Difference (A1 – A2) 45 8.22 22.74 8.90

The study's interest is in determining if, on average, Automobile 1 requires more than 5 additional seconds to park compared to Automobile 2.

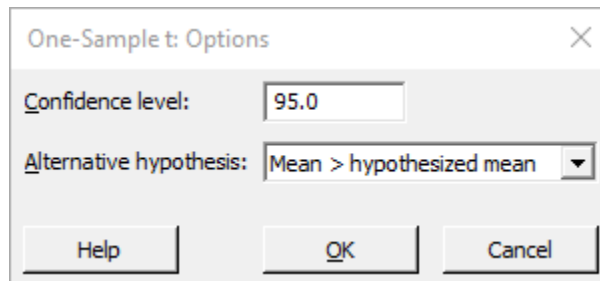
Solution: Unlike the other two examples, we are given summary statistics instead of the actual data. The times to park the two automobiles are paired by subject. Since there are $n = 45$ subjects, the sample size is large enough to assume that the distribution of the sample means for the Difference variable is normally distributed.

We are not given α , so we will assume it to be 0.05. Since we want to determine if the time to park Automobile 1 is *greater than* the time to park Automobile 2, we will construct a one-sided 95% 1-Sample t confidence lower bound for the mean difference in Minitab.

Since we did not have data in this example, we used the 1-Sample t 's drop-down box to select **Summarized Data**.



To change the confidence interval to a 95% lower bound, click **Options** and change the alternative hypothesis to greater than.



Descriptive Statistics

N	Mean	StDev	SE Mean	95% Lower Bound for μ
45	8.22	22.74	3.39	2.52

μ : population mean of Sample

Since the 95% lower confidence bound for the mean difference in parking times for the two automobiles is 2.52, then we don't have enough evidence to suggest that μ_d is greater than 5 seconds at $\alpha = 0.05$. Although there are values greater than 5 seconds in the interval, the value 5 seconds is also captured in it. Thus, the true mean difference could be 5 seconds.

If the study asked if the time to park Automobile 1 required more than 2 additional seconds to park than Automobile 2, on average, then our confidence would have supported the 2 second difference at $\alpha = 0.05$.

Hypothesis Testing

When two populations have paired data, a paired t -test is used to perform hypothesis testing on the population mean difference. When performing a hypothesis test on paired data, we can use a 1-sample t -test on their differences rather than a paired t -test.

Recall that the differences d_i 's for paired data of sample size n are:

$$\begin{aligned}d_1 &= X_{A1} - X_{B1} \\d_2 &= X_{A2} - X_{B2} \\&\dots \\d_i &= X_{Ai} - X_{Bi} \\&\dots \\d_n &= X_{An} - X_{Bn}\end{aligned}$$

where X_{Ai} is the "after" value and X_{Bi} is the "before" value for the i^{th} data pair.

Recall that the estimate for the true mean difference μ_d is the average or sample mean of the paired differences: $\bar{x}_d = \frac{\sum_{i=1}^n d_i}{n}$. The sample variance of the differences is $s_d^2 = \frac{\sum_{i=1}^n (d_i - \bar{x}_d)^2}{n-1}$. If the differences are normally distributed, we can use the formula from the 1-sample t -test lesson to obtain a standardized test statistic for a hypothesis test on μ_d .

Hypothesis test for population mean paired difference μ_d

The null and alternative hypotheses for testing whether the true mean difference μ_d of data pairs is μ_{d_0} are:

Null hypothesis: $H_0: \mu_d = \mu_{d_0}$

Alternative hypothesis: $H_a: \mu_d \neq \mu_{d_0}$

Note: The alternative hypothesis can also be $>$ or $<$. If we went to determine if the true mean difference μ_d of the pairs is equal, then let $\mu_{d_0} = 0$.

Assuming that the differences d_i 's are normally distributed and that the null hypothesis $H_0: \mu_d = \mu_{d_0}$ is true, the standardized test statistic is

$$t_0 = \frac{\bar{x}_d - \mu_{d_0}}{\frac{s_d}{\sqrt{n}}} \sim t_{n-1}, \text{ where}$$

\bar{x}_d = sample mean of the differences,

s_d = sample standard deviation of the differences,

n = number of paired observations,

t_{n-1} = t -distribution with $n - 1$ degrees of freedom,

μ_{d_0} = hypothesized mean of the differences.

Example 3 (Typing vs Texting Speed continued). Can College A's incoming freshmen type on their computer keyboard *faster* than they text on their phone? Are they more than 5 words/minute faster keyboard typing than texting? The full table of the $n = 13$ randomly selected freshmen from College A with their texting and keyboard typing speeds per minute are on page 8.

Freshman at College A	Keyboard Typing Speed	Text Typing Speed	Difference: Typing – Texting
Tom	52	39	13
George	51	44	7
...
Patrick	47	37	10

Solution: The hypothesis that we are testing for $d = \text{Keyboard Typing Speed} - \text{Text Typing Speed}$ is:

$$H_0: \mu_d = 0 \text{ versus } H_a: \mu_d > 0$$

As displayed in the first part of Example 3 earlier, the *Difference* data is normally distributed. By-hand, we can compute the 1-Sample t test statistic for the *Differences*:

$$t_0 = \frac{10.385 - 0}{\frac{3.176}{\sqrt{13}}} \cong 11.790.$$

It's not even necessary to use Minitab's **Graph > Probability Distribution Plot** to determine this test statistic's p -value. The mean difference of 10.385 is 11.790 standard deviations away from 0. If the null hypothesis were true; that is, if the mean typing and texting speeds were equal, then the probability of getting a standardized test statistic of 11.790 on a t distribution with 12 degrees of freedom is essentially 0.

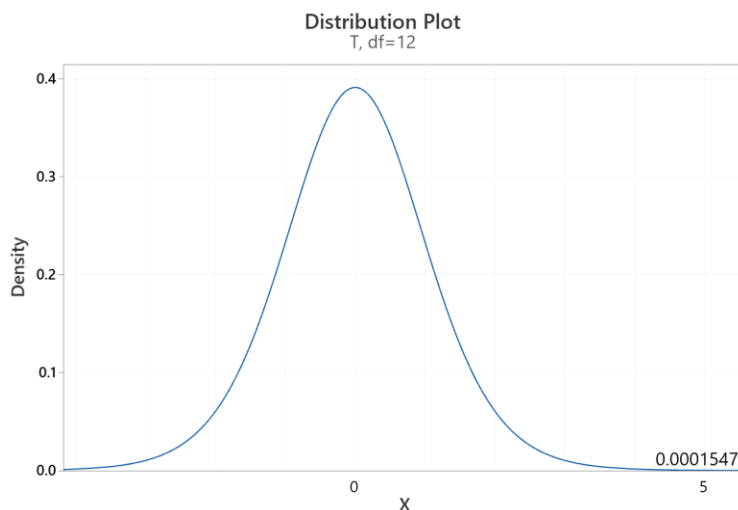
Since the test statistic's p -value is less than $\alpha = 0.05$, then we have enough evidence to reject H_0 , indicating the mean keyboard typing speed is faster than the mean text typing speed.

We continue this example by considering if the mean keyboard typing speed is more than 5 words/minute faster than the mean text typing speed.

The only difference in this part of the problem compared to the last is that the difference that we are trying to detect is 5 words/minute, not simply 0 words/minute. The new test statistic is:

$$t_0 = \frac{10.385 - 5}{\frac{3.176}{\sqrt{13}}} \cong 6.113.$$

We can use Minitab's **Graph > Probability Distribution Plot** to determine that p -value associated with this test statistic, though we know it will also be essentially 0.



We can use either Minitab's **1-Sample t** or **Paired t** to provide us with both the test statistic and p -value at the same time. We will return to using 0 for the difference.

- 1 **Stat > Basic Statistics > 1-Sample t**
- 2 In **One or more samples, each in a column**, enter *Difference: Typing – Texting*.
- 3 Select **Perform hypothesis test**. In **Hypothesized mean**, enter 0.
- 4 Select **Options**.
- 5 In **Alternative hypothesis**, choose **Mean > hypothesized mean**.
- 6 Click **OK** in each dialog box.

Test

Null hypothesis $H_0: \mu = 0$
 Alternative hypothesis $H_1: \mu > 0$

T-Value	P-Value
11.79	0.000

Since the test statistic's p -value is less than $\alpha = 0.05$, then we have enough evidence to reject H_0 , indicating the mean keyboard typing speed is faster than the mean text typing speed. Notice that this agrees with our confidence interval results.

Since the 95% one-sided 1-Sample t lower bound for the mean difference is 8.814, we will not obtain a p -value less than $\alpha = 0.05$ unless we test a value greater than 8.814 words/minute. For example, is the mean keyboard typing speed 10 seconds greater than the mean text typing speed? Because of the 95% one-sided confidence bound for the mean difference contains the value 10, then we will not reject 10 in the following hypothesis test:

$$H_0: \mu_d = 10 \text{ versus } H_a: \mu_d > 10$$

Using Minitab, we obtain:

Test

Null hypothesis $H_0: \mu = 10$
 Alternative hypothesis $H_1: \mu > 10$

T-Value	P-Value
0.44	0.335

We can also the Paired t-test directly using Minitab. Again, we return to using 0 for the difference.

- 1 **Stat > Basic Statistics > Paired t.**
- 2 In **Sample 1**, enter *Keyboard Typing Speed* and in **Sample 2**, enter *Text Typing Speed*.
- 3 Select **Options**.
- 4 In **Alternative hypothesis**, choose **Difference > hypothesized difference**.
- 5 Click **OK** in each dialog box.

Test

Null hypothesis $H_0: \mu_{\text{difference}} = 0$

Alternative hypothesis $H_a: \mu_{\text{difference}} > 0$

T-Value P-Value

11.79 0.000