

# How can I perform a paired samples t-test using SAS?

Authored by  
**stats writer**

June 24, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can I perform a paired samples t-test using SAS?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=150540>

The paired samples t-test is a statistical method used to compare the means of two related groups. In order to perform this test using SAS, the first step is to input the data into SAS and organize it into two columns, one for each group. Then, using the PROC TTEST procedure, specify the type of test (paired t-test) and the variables to be compared. SAS will then calculate the t-statistic, degrees of freedom, and p-value, allowing for the determination of whether there is a significant difference between the two groups. This formal statistical analysis can help researchers make informed decisions based on their data.

## Paired Samples t Test

The Paired Samples t Test compares the means of two measurements taken from the same individual, object, or related units. These "paired" measurements can represent things like:

A measurement taken at two different times (e.g., pre-test and post-test score with an intervention administered between the two time points)  
A measurement taken under two different conditions (e.g., completing a test under a "control" condition and an "experimental" condition)  
Measurements taken from two halves or sides of a subject or experimental unit (e.g., measuring hearing loss in a subject's left and right ears).

The purpose of the test is to determine whether there is statistical evidence that the mean difference between paired observations is significantly different from zero. The Paired Samples t Test is a parametric test.

This test is also known as:

Dependent t Test  
Paired t Test  
Repeated Measures t Test

The variable used in this test is known as:

Dependent variable, or test variable (continuous), measured at two different times or for two related conditions or units

## Common Uses

The Paired Samples t Test is commonly used to test the following:

Statistical difference between two time points  
Statistical difference between two conditions  
Statistical difference between two measurements  
Statistical difference between a matched pair

**Note:** The Paired Samples t Test can only compare the means for two (and only two) related (paired) units on a continuous outcome that is normally distributed. The Paired Samples t Test is

not appropriate for analyses involving the following: 1) unpaired data; 2) comparisons between more than two units/groups; 3) a continuous outcome that is not normally distributed; and 4) an ordinal/ranked outcome.

To compare unpaired means between two independent groups on a continuous outcome that is normally distributed, choose the Independent Samples t Test. To compare unpaired means between more than two groups on a continuous outcome that is normally distributed, choose ANOVA. To compare paired means for continuous data that are not normally distributed, choose the nonparametric Wilcoxon Signed-Ranks Test. To compare paired means for ranked data, choose the nonparametric Wilcoxon Signed-Ranks Test.

## Data Requirements

Your data must meet the following requirements:

Dependent variable that is continuous (i.e., interval or ratio level)

**Note:** The paired measurements must be recorded in two separate variables. Related samples/groups (i.e., dependent observations)

The subjects in each sample, or group, are the same. This means that the subjects in the first group are also in the second group. Random sample of data from the population Normal distribution (approximately) of the difference between the paired values No outliers in the difference between the two related groups

**Note:** When testing assumptions related to normality and outliers, you must use a variable that represents the difference between the paired values - not the original variables themselves.

**Note:** When one or more of the assumptions for the Paired Samples t Test are not met, you may want to run the nonparametric Wilcoxon Signed-Ranks Test instead.

## Hypotheses

The hypotheses can be expressed in two different ways that express the same idea and are mathematically equivalent:

$H_0: \mu_1 = \mu_2$  ("the paired population means are equal")

$H_1: \mu_1 \neq \mu_2$  ("the paired population means are not equal")

OR

$H_0: \mu_1 - \mu_2 = 0$  ("the difference between the paired population means is equal to 0")

$H_1: \mu_1 - \mu_2 \neq 0$  ("the difference between the paired population means is not 0")

where

$\mu_1$  is the population mean of variable 1, and  $\mu_2$  is the population mean of variable 2.

## Test Statistic

The test statistic for the Paired Samples t Test, denoted  $t$ , follows the same formula as the one sample  $t$  test.

$$t = \frac{\overline{x}_{\text{diff}} - 0}{s_{\overline{x}}}$$

where

$$s_{\overline{x}} = \frac{s_{\text{diff}}}{\sqrt{n}}$$

where

$\overline{x}_{\text{diff}}$  = Sample mean of the differences

$n$  = Sample size (i.e., number of observations)

$s_{\text{diff}}$  = Sample standard deviation of the differences

$s_{\overline{x}}$  = Estimated standard error of the mean ( $s/\sqrt{n}$ )

The calculated  $t$  value is then compared to the critical  $t$  value with  $df = n - 1$  from the  $t$  distribution table for a chosen confidence level. If the calculated  $t$  value is greater than the critical  $t$  value, then we reject the null hypothesis (and conclude that the means are significantly different).

## Data Set-Up

Your data should include two continuous numeric variables (represented in columns) that will be used in the analysis. The two variables should represent the paired variables for each subject (row). If your data are arranged differently (e.g., cases represent repeated units/subjects), simply restructure the data to reflect this format.

## Using PROC TTEST for Paired Samples t Tests

When conducting a Paired Samples t Test, the general form of PROC TTEST is:

```
PROC TTEST DATA=dataset-name ALPHA=.05;
PAIRED V1*V2;
RUN;
```

In the PROC TTEST statement, the DATA option specifies the name of your dataset. The optional

`ALPHA` option specifies the desired significance level. By default, PROC TTEST uses `ALPHA=.05` (i.e., 5% significance), but you can set it to `ALPHA=.01` for 1% significance, or `ALPHA=.10` for 10% significance, etc.

The `PAIRED` statement is where you specify the pair(s) of variables you will test, using an asterisk between the variable names denote a pair. You can specify multiple pairings by separating each pairing with a space.

## Example

### Problem Statement

The sample dataset has placement test scores (out of 100 points) for four subject areas: English, Reading, Math, and Writing. Each student took all four placement tests. Suppose we are particularly interested in the English and Math sections, and want to determine whether English or Math had higher test scores on average. We could use a paired  $t$  test to test if there was a significant difference in the average of the two tests.

### Before the Test

#### State the Null and Alternative Hypotheses

The hypotheses for this example can be expressed as:

$H_0: \mu_{\text{English}} - \mu_{\text{Math}} = 0$  ("the difference between the average English and Math scores is equal to 0")

$H_1: \mu_{\text{English}} - \mu_{\text{Math}} \neq 0$  ("the difference between the average English and Math scores is not 0")

Before we perform our hypothesis tests, we should decide on a *significance level* (denoted  $\alpha$ ). The significance level is the threshold we will use to decide whether a test result is significant. For this example, let's use  $\alpha = 0.05$ , or 5%.

### Data Set-Up

In the sample dataset, each student's responses are recorded on one row. Their English and Math scores are represented in the variables English and Math. This format is already appropriate for the paired samples t-test, so no further restructuring is needed.

	ID Number	Score on English placement test	Score on Math placement test
1	43783	88.24	60.02
2	20278	89.45	70.19
3	20389	96.73	71.2
4	22820	74.06	55.89
5	24559	82.61	65.52
6	28980	70.1	61.4
7	33312	78.98	76.71
8	40274	.	58.64
9	40390	75.65	51.14
10	28942	68.27	55.2
11	29045	86	64.41

## Running the Test

### SAS Program

```
PROC TTEST DATA=sample ALPHA=.05;  
PAIRED English*Math;  
RUN;
```

### Output

#### Tables

After executing the SAS program above, SAS produces the following set of tables:

## The SAS System

### The TTEST Procedure

#### Difference: English - Math

N	Mean	Std Dev	Std Err	Minimum	Maximum
398	17.2973	9.5030	0.4763	-10.6400	41.4900

Mean	95% CL Mean	Std Dev	95% CL Std Dev
17.2973	16.3608 18.2337	9.5030	8.8855 10.2135

DF	t Value	Pr >  t
397	36.31	<.0001

The heading "Difference: English - Math" tells us the order of the subtraction used for these numbers. This is important, since it determines how we interpret positive and negative numbers. Because Math is subtracted from English, positive numbers correspond to higher English scores, and negative numbers correspond to higher Math scores.

In the first table, we have descriptive statistics for the difference scores:

N: The effective sample size (students who had both an English score and a Math score). Mean and St Dev: The average difference between a student's English and Math scores. On average, students had a 17.3-point difference between their English and Math scores (+/- a standard deviation of 9.5). Std Err: The standard error of the difference scores,  $s/\sqrt{n}$ . Minimum: The smallest difference score observed in the sample. Here, the score -10.64 represents a student who scored 10.64 points higher on their Math test than their English test. Maximum: The largest difference score observed in the sample. Here, the score 41.69 represents a student who scored 41.69 points higher on their English test than their Math test.

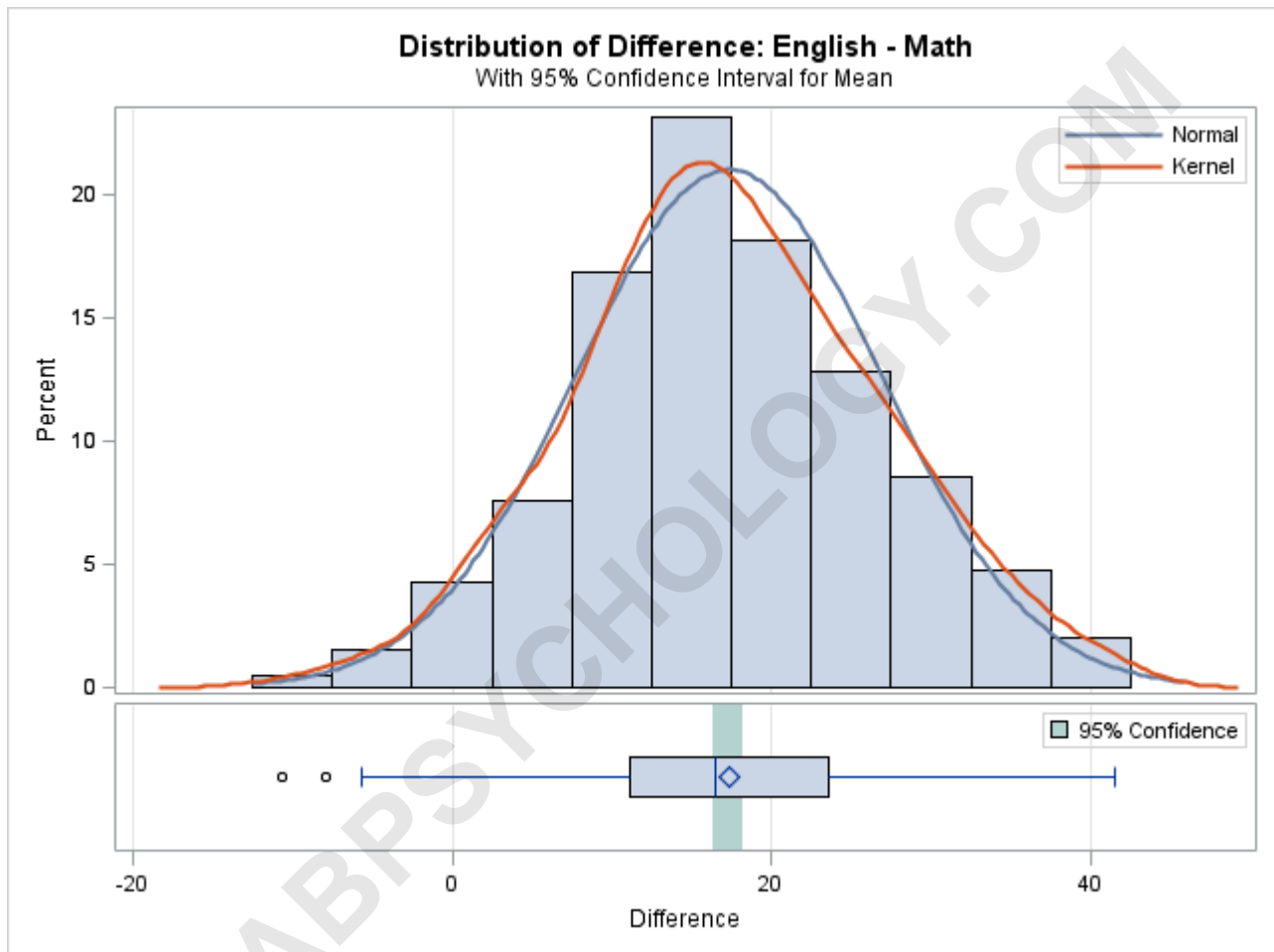
In the second table, we have the 95% confidence intervals for the mean difference and the standard deviation of the differences.

In the third table, we have the actual paired t test results. The p-value is very small ( $p < .0001$ ), so we reject the null hypothesis that the average English and Math scores were the same, and conclude that the English scores had a significantly different average than the Math scores.

## Graphs

### Graph 1: Distribution of Difference Scores

The first graph depicts the distribution of the difference scores, using both a histogram (top panel) and a boxplot (bottom panel).



If this histogram were centered about 0, it would correspond to no difference between the two test scores; however, the highlighted region in the boxplot shows that the center of the distribution is between 17 and 18.

In the histogram:

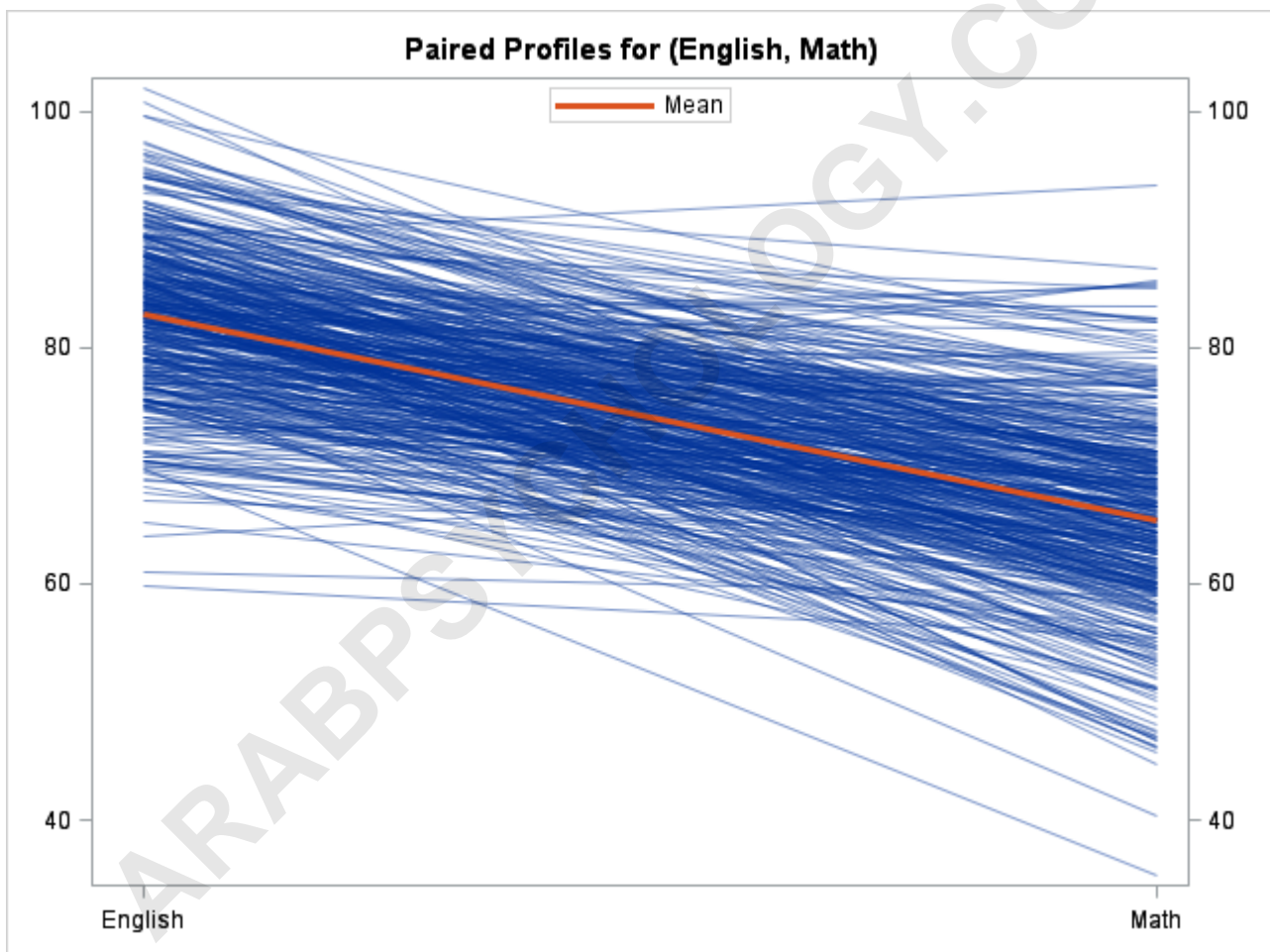
The blue line shows the shape of a normal distribution with the mean and standard deviation from this sample. The red line shows the *kernel density estimate* - a type of approximation for the shape of a distribution. If the scores were perfectly normally distributed, the kernel density estimate would "line up" with the normal approximation. In the boxplot:

The box's center line shows the median, while the diamond shows the mean. There are two outliers on the low end; these represent individuals who scored 7-10 points higher on the Math test

than the English test.

### Graph 2: Profile Plot

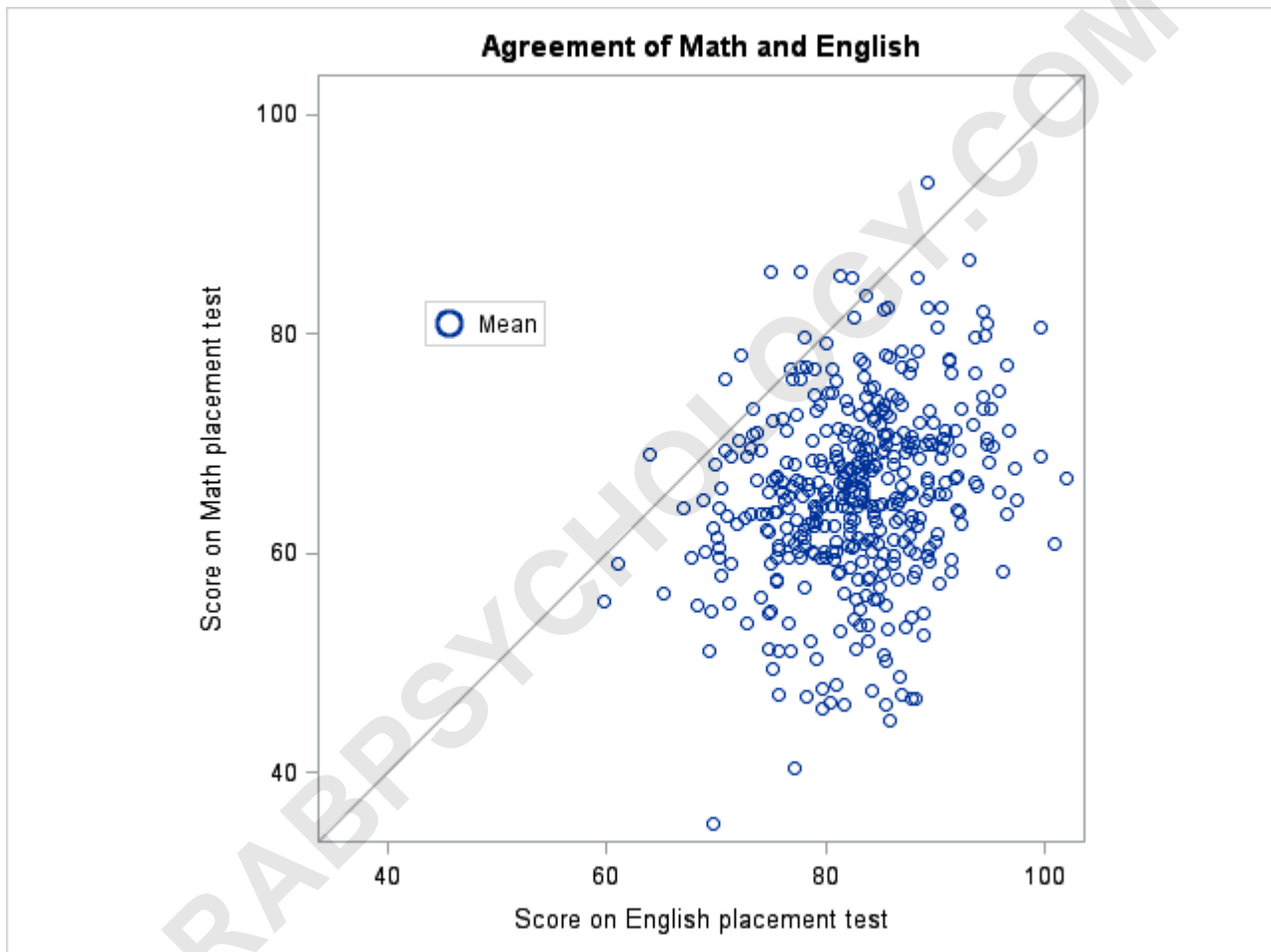
The second graph is a paired *profile plot*. Profile plots depict the "trajectory" of individuals. Each line represents one subject or case. On the left side is the person's English score; on the right side is their Math score. (Note that the axes on both sides have the same range; this is an important feature of profile plots.) By looking at the slope of these lines, we can get a feel for whether the scores are approximately equal (horizontal lines), or if one score was higher than the other (sloping lines).



Although some of the lines are roughly horizontal, most of the lines tend to have a downward slope. Since English is on the left and Math is on the right, this corresponds to most students scoring higher on the English placement test than the Math placement test. The red line showing the average trend confirms this.

### Graph 3: Agreement Plot

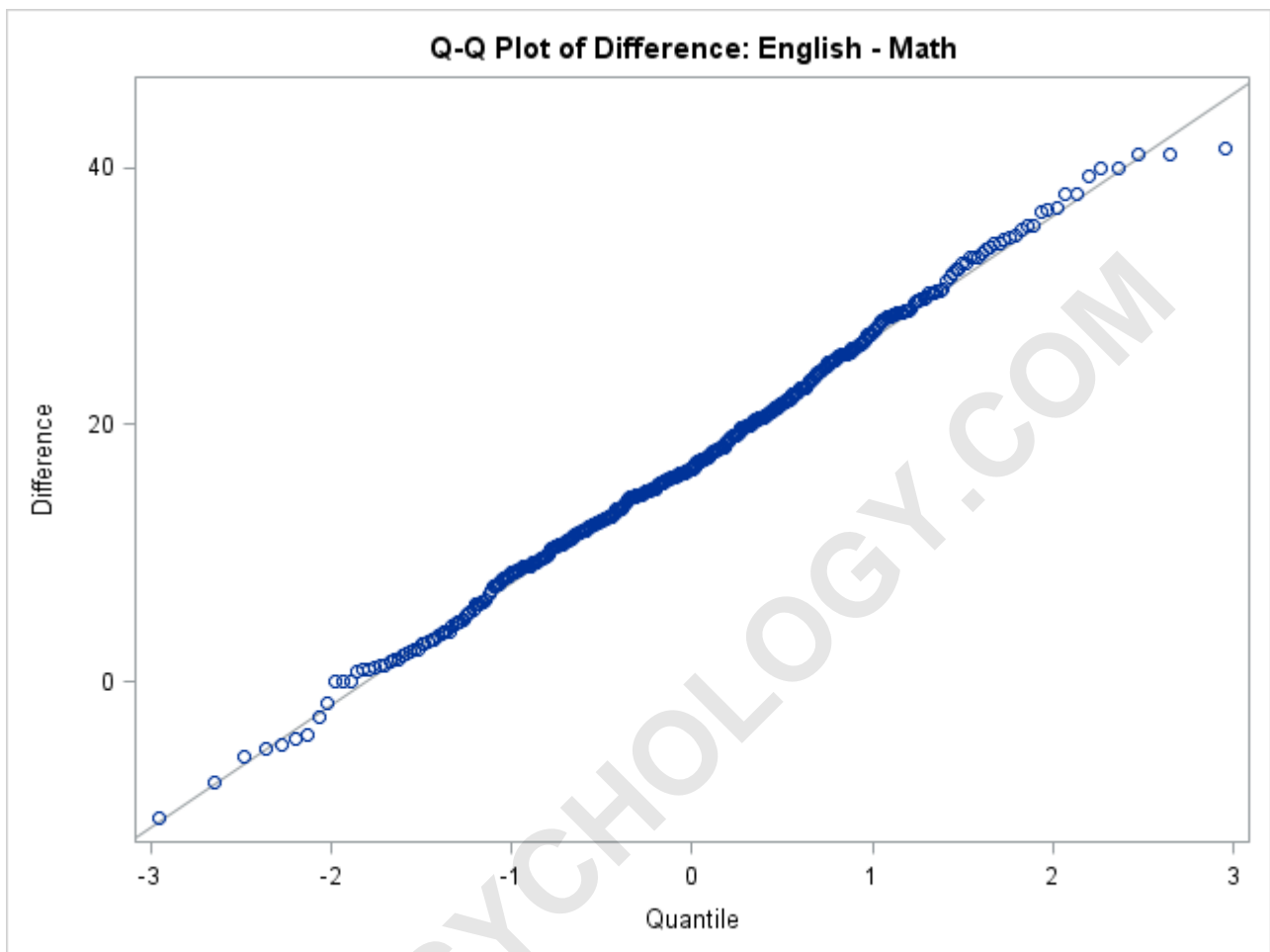
The third graph shows the "agreement" of the two scores. The plot itself is a variation on a simple scatterplot. The diagonal reference line represents identical English and Math scores. Datapoints that fall on this line (or near this line) represent students who scored the same on their English and Math tests. Datapoints above the line represent students who scored higher on the Math test than the English test. Points below the line represent students who scored higher on English than Math. In this graph, we see many more points below the line than above the line, which means that most students scored higher on English than on Math.



**Graph 4: Quantile-Quantile (Q-Q) Plot of Normality for Differences**

The fourth graph is a Q-Q plot, or quantile-quantile plot, of the difference scores. Q-Q Plots are used to inspect whether an observed variable (represented as points) matches what we would expect that variable to look like if it were truly normally distributed (represented as a solid line). To read a Q-Q plot, we look to see if the dots (the observed values) match up with the expected values for a normal distribution (the diagonal line). If the points fall along the line, then the values are consistent with what we would expect them to be if the data were truly normally distributed. Here, we see that the majority of the difference scores fall on the diagonal line, so we can say that

the data appear to be approximately normally distributed.



## Decision and Conclusions

Since  $p < .0001$  is less than our chosen significance level  $\alpha = 0.05$ , we can reject the null hypothesis, and conclude that the English and Math scores were significantly different from each other.

Based on the results, we can state the following:

There was a significant difference in the average English and Math scores ( $t_{397} = 36.31$ ,  $p < .05$ ). On average, students scored 17.3 points higher on the English test than the Math test (95% confidence interval).

## Tutorial Feedback