Chapter 218

# One-Sample T-Test for Equivalence

## Introduction

This procedure provides reports for making inference about the equivalence of a single variable compared to a known standard value. The question of interest is whether the variable is actually equivalent to the standard, that is, differs on average by a small margin, at most. This is tested by the TOST (two one-sided tests) equivalence test. This test allows you to obtain a p-value so that you can quantify the results of hypothesis test.

## Technical Details

Suppose you want to evaluate the equivalence of a continuous random variable, $X$, as compared to a standard value, $S$. Assume that $n$ observations, $X_k$, $k$ = 1, 2, …, $n$ are available.

### TOST Equivalence Test

Schuirmann's (1987) two one-sided tests (TOST) approach is used to test equivalence. The equivalence test essentially reverses the roles of the null and alternative hypothesis. Assume that $\mu$ represents the population mean of the response, $S$ is a standard reference value, and $M$ is the so-called *margin of equivalence*. The null and alternative hypotheses are

$$H_0: \mu < S - M \quad \text{or} \quad \mu > S + M$$
$$H_1: S - M < \mu < S + M$$

The null hypothesis is made up of two simple one-sided hypotheses:

$$H_{01}: \mu < S - M$$
$$H_{02}: \mu > S + M$$

If both of these one-sided tests are rejected, we conclude H1 that the response is equivalent to the standard value (their difference is confined within a small margin). Schuirmann showed that if we want the alpha level of the equivalence test to be α, then each of the one-sided tests should be $α$ as well (not $α/2$ as you might expect). The probability level (p-value) of the equivalence test is equal to the maximum of the probability levels of the two one-sided tests. These tests are conducted using the standard formulas for the one-sample t-test.

## Assumptions

This section describes the assumptions that are made when you use each of the tests of this procedure. The key assumption relates to normality or non-normality of the data. One of the reasons for the popularity of the t-test is its robustness in the face of assumption violation. Unfortunately, in practice it often happens that more than one assumption is not met. Hence, take the steps to check the assumptions before you make important decisions based on these tests. There are reports in this procedure that permit you to examine the assumptions, both visually and through assumptions tests.

### One-Sample T-Test Assumptions

The assumptions of the one-sample t-test are:

1.  The data are continuous (not discrete).

2.  The data follow the normal probability distribution.

3.  The sample is a simple random sample from its population. Each individual in the population has an equal probability of being selected in the sample.

### Wilcoxon Signed-Rank Test Assumptions

The assumptions of the Wilcoxon signed-rank test are as follows:

1.  The data are continuous (not discrete).

2.  The distribution of the data is symmetric.

3.  The data are mutually independent.

4.  The data all have the same median.

5.  The measurement scale is at least interval.

# Data Structure

For this procedure, the data are entered as a single column and specified as a response variable.

| X1 |
| --- |
| 26 |
| 48 |
| 19 |
| 38 |
| 16 |
| 16 |
| 54 |
| 56 |
| 87 |

# Example 1 – TOST Equivalence Test for a Single Sample

This section presents an example of how to test equivalence for a single sample. Suppose a new and cheaper treatment method is being investigated and a measurement was made on each of 100 subjects using the new treatment (Method1). The data are in the **Bland-Altman** dataset. The researchers wish to determine if the response is equivalent to the known standard value of 50 on average with an equivalence margin of 6.

## Setup

To run this example, complete the following steps:

**1   Open the Bland-Altman example dataset**
- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Bland-Altman** and click **OK**.

**2   Specify the One-Sample T-Test for Equivalence procedure options**
- Find and open the **One-Sample T-Test for Equivalence** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Response Variable ...........................................**Method1**
Equivalence Bounds ........................................**Symmetric**
Standard Value ................................................**50**
Equivalence Margin ........................................**6**

Reports Tab

Wilcoxon Signed-Rank Test.............................**Checked**

**3   Run the procedure**
- Click the **Run** button to perform the calculations and generate the output.

# Descriptive Statistics and Confidence Interval for the Mean (μ)

**Descriptive Statistics and Confidence Interval for the Mean (μ)**

| Variable | N | Mean | Standard Deviation of the Data | Standard Error of the Mean | T* | 95% Confidence Interval Limits for the Mean (μ) | |
|---|---|---|---|---|---|---|---|
| | | | | | | Lower | Upper |
| Method1 | 100 | 50.72 | 28.30893 | 2.830893 | 45.10289 | 56.3371 | |

This report provides basic descriptive statistics and confidence interval for the response variable.

## Variable

The name of the variable for which the descriptive statistics are listed here.

## N

The count gives the number of non-missing values. This value is often referred to as the group sample size or count.

## Mean

This is the average of the data values.

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

## Standard Deviation of the Data

The sample standard deviation is the square root of the variance. It is a measure of dispersion based on squared distances from the mean for the variables listed.

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}$$

## Standard Error of the Mean

This is the estimated standard deviation of the distribution of sample means for an infinite population.

$$SE_{\overline{x}} = \frac{s}{\sqrt{n}}$$

## T*

This is the t-value used to construct the confidence limits. It is based on the degrees of freedom and the confidence level.

## 95% Confidence Interval Limits for the Mean (μ) (Lower and Upper)

These are the lower and upper limits of a confidence interval estimate of the mean based on a Student's t distribution with *n* - 1 degrees of freedom. This interval estimate assumes that the population standard deviation is not known and that the data are normally distributed. The confidence interval formula is

$$\bar{x} \pm T_{df} SE_{\bar{x}}$$

# One-Sample T-Test for Equivalence using TOST (Two One-Sided Tests)

**One-Sample T-Test for Equivalence using TOST (Two One-Sided Tests)**
───────────────────────────────────────────────────────────────
Null Hypothesis (H0):　　　　　(Mean of Method1) ≤ 50 - 6　or　(Mean of Method1) ≥ 50 + 6
Equivalence Hypothesis (H1):　　50 - 6 < (Mean of Method1) < 50 + 6
───────────────────────────────────────────────────────────────

| Test | Alternative Hypothesis | Mean | Standard Error | T-Statistic | DF | P-Value | Reject H0 at α = 0.05? |
|------|------------------------|------|----------------|-------------|-----|---------|------------------------|
| Lower Boundary | μ > 44 | 50.72 | 2.830893 | 2.3738 | 99 | 0.00977 | Yes |
| Upper Boundary | μ < 56 | 50.72 | 2.830893 | -1.8651 | 99 | 0.03256 | Yes |
| Equivalence | 44 < μ < 56 | 50.72 | | | | 0.03256 | Yes |

This report shows the equivalence test (the third row) and the two one-sided tests (first and second rows). Since the *p*-value for the equivalence test (the maximum of the two one-sided tests) is less than the designated value of alpha (0.05), we can reject the null hypothesis and conclude that the mean is equivalent to the standard value of 50 for an equivalence margin of 6. This test assumes that the data are normal.

### Test

The two one-sided tests are shown on the first two rows and the equivalence test is shown at the end.

### Alternative Hypothesis

Assume that $\mu$ represent the population mean of the response, *S* is a standard reference value, and that *M* is the positive *equivalence margin*. The null and alternative hypotheses are

$$H_0: \mu < S - M \quad \text{or} \quad \mu > S + M$$
$$H_1: S - M < \mu < S + M$$

### Mean Difference

This is the average of the data values.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

### Standard Error

This is the estimated standard deviation of the distribution of sample means for an infinite population.

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

## T-Statistic

The T-Statistic is the value used to produce the *p*-value (Prob Level) based on the *T* distribution. The formula for the T-Statistic is:

$$T = \frac{\bar{x} - (S - M)}{SE_{\bar{x}}}$$

## DF

The degrees of freedom define the T *distribution* upon which the probability values are based. The formula for the degrees of freedom is:

$$df = n - 1$$

## P-Value

The *p*-value, also known as the probability level or significance level, is the probability that the test statistic will take a value at least as extreme as the observed value, assuming that the null hypothesis is true. If the *p*-value is less than the prescribed *α*, in this case 0.05, the null hypothesis is rejected in favor of the alternative hypothesis. Otherwise, there is not sufficient evidence to reject the null hypothesis. The Prob Level for the equivalence test is equal to the maximum of the two one-sided tests.

### Reject H0 at α = 0.05?

This column indicates whether or not the null hypothesis is rejected, in favor of the alternative hypothesis, based on the *p*-value and chosen *α*. A test in which the null hypothesis is rejected is sometimes called *significant*.

# Wilcoxon Signed-Rank Test for Equivalence using TOST (Two One-Sided Tests)

**Wilcoxon Signed-Rank Test for Equivalence using TOST (Two One-Sided Tests)**

| | |
|---|---|
| Null Hypothesis (H0): | (Median of Method1) ≤ 50 - 6   or   (Median of Method1) ≥ 50 + 6 |
| Equivalence Hypothesis (H1): | 50 - 6 < (Median of Method1) < 50 + 6 |

**Test Details**

| Test | Sum of Ranks (W) | Mean of W | Standard Deviation of W | Number of Zeros | Number of Sets of Ties | Multiplicity Factor |
|---|---|---|---|---|---|---|
| Lower Boundary | 3160 | 2524.5 | 290.8021 | 1 | 26 | 1026 |
| Upper Boundary | 2013 | 2523.5 | 290.8028 | 2 | 27 | 960 |

**Test Results**

| Test Type | Test | Alternative Hypothesis | Z-Value | P-Value | Reject H0 at α = 0.05? |
|---|---|---|---|---|---|
| Exact* | Lower Boundary | Median > 44 | | | |
| Exact* | Upper Boundary | Median < 56 | | | |
| Exact* | Equivalence | 44 < Median < 56 | | | |
| | | | | | |
| Normal Approximation | Lower Boundary | Median > 44 | 2.1853 | 0.01443 | Yes |
| Normal Approximation | Upper Boundary | Median < 56 | -1.7555 | 0.03959 | Yes |
| Normal Approximation | Equivalence | 44 < Median < 56 | | 0.03959 | Yes |
| | | | | | |
| Normal Approx. with C.C. | Lower Boundary | Median > 44 | 2.1836 | 0.01450 | Yes |
| Normal Approx. with C.C. | Upper Boundary | Median < 56 | -1.7538 | 0.03974 | Yes |
| Normal Approx. with C.C. | Equivalence | 44 < Median < 56 | | 0.03974 | Yes |

* The Exact Test is provided only when there are no ties.

This report shows the nonparametric Wilcoxon Signed-Rank Test for equivalence. Since the *p*-value for the equivalence test (the maximum of the two one-sided tests) is less than the designated value of alpha (0.05), we can reject the null hypothesis in favor of the alternative for an equivalence margin of 6. This test required no assumption that the differences are normal.

## Sum of Ranks (W)

The basic statistic for this test is the sum of the positive ranks, $\sum R_+$ (The sum of the positive ranks is chosen arbitrarily. The sum of the negative ranks could equally be used). This statistic is called $W$.

$$W = \sum R_+$$

## Mean of W

This is the mean of the sampling distribution of the sum of ranks for a sample of $n$ items.

$$\mu_W = \frac{n(n+1) - d_0(d_0+1)}{4}$$

where $d_0$ is the number of zero differences.

## Standard Deviation of W

This is the standard deviation of the sampling distribution of the sum of ranks. Here $t_i$ represents the number of times the $i^{th}$ value occurs.

$$s_W = \sqrt{\frac{n(n+1)(2n+1) - d_0(d_0+1)(2d_0+1)}{24} - \frac{\sum t_i^3 - \sum t_i}{48}}$$

where $d_0$ is the number zero differences, $t_i$ is the number of absolute differences that are tied for a given non-zero rank, and the sum is over all sets of tied ranks.

## Number of Zeros

This is the number of times that the difference between the observed data and the hypothesized value is zero. The zeros are used in computing ranks but are not considered positive ranks or negative ranks.

## Number of Sets of Ties

The treatment of ties is to assign an average rank for the particular set of ties. This is the number of sets of ties that occur in the data, including ties at zero.

## Multiplicity Factor

This is the correction factor that appeared in the standard deviation of the sum of ranks when there were ties.

## Test Type

This is the type of test that is being reported on the current row. The Exact Test is provided only when there are no ties.

## Alternative Hypothesis

For the Wilcoxon signed-rank test, the null and alternative hypotheses relate to the median. The left-tail alternative is represented by Median < S + M (i.e., $H_a$: Median < S + M) while the right-tail alternative is depicted by Median > S - M.

## Exact Probability: P-Value

This is an exact $p$-value for this statistical test, assuming no ties. The $p$-value is the probability that the test statistic will take on a value at least as extreme as the actually observed value, assuming that the null hypothesis is true. If the $p$-value is less than $\alpha$, say 5%, the null hypothesis is rejected. If the $p$-value is greater than $\alpha$, the null hypothesis is accepted.

## Exact Probability: Reject H0 at α = 0.05?

This is the conclusion reached about the null hypothesis. It will be to either fail to reject $H_0$ or reject $H_0$ at the assigned level of significance.

## Approximations with (and without) Continuity Correction: Z-Value

Given the sample size is at least ten, a normal approximation method may be used to approximate the distribution of the sum of ranks. Although this method does correct for ties, it does not have the continuity correction factor. The $z$ value is as follows:

$$z = \frac{W - \mu_W}{\sigma_W}$$

If the correction factor for continuity is used, the formula becomes:

$$z = \frac{W - \mu_W \pm \frac{1}{2}}{\sigma_W}$$

## Approximations with (and without) Continuity Correction: P-Value

This is the *p*-value for the normal approximation approach for the Wilcoxon signed-rank test. The *p*-value is the probability that the test statistic will take a value at least as extreme as the actually observed value, assuming that the null hypothesis is true. If the *p*-value is less than *α*, say 5%, the null hypothesis is rejected. If the *p*-value is greater than *α*, the null hypothesis is accepted.

## Approximations with (and without) Continuity Correction: Reject H0 at α = 0.05?

This is the conclusion reached about the whether to reject null hypothesis. It will be either Yes or No at the given level of significance.

# Test of the Normality Assumption

**Test of the Normality Assumption**

| Normality Test | Test Statistic | P-Value | Reject the Assumption of Normality at α = 0.05? |
|---|---|---|---|
| Shapiro-Wilk | 0.9580 | 0.00292 | Yes |

The main assumption when using the t-test is that the data come from a normal distribution. The normality assumption can be checked statistically by the Shapiro-Wilk normality test and visually by the histogram or normal probability plot.

This section reports the results of a diagnostic test to determine if the data are normal. In this case, they are not, probably because of the outliers that were present. This would indicate that the Wilcoxon Signed-Rank Test would be the better test to use.

# Histogram
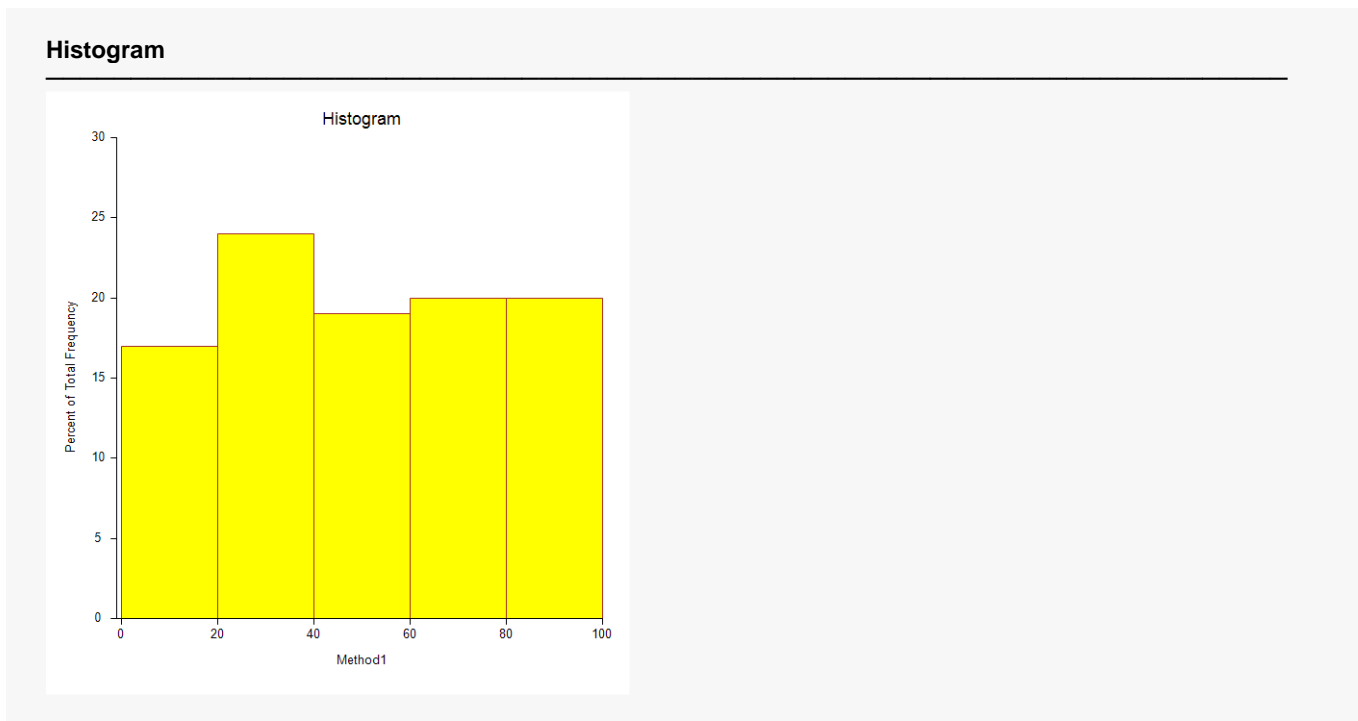
**Histogram**
_____



The nonparametric tests need the assumption of symmetry, and these two graphic tools can provide that information. Since the histogram's shape is impacted by the number of classes or bins and the width of the bins, the best choice is to trust the density trace, which is a smoothed histogram. If the distribution of data is symmetrical but not normal, proceed with the nonparametric test.

# Probability Plot

**Probability Plot**
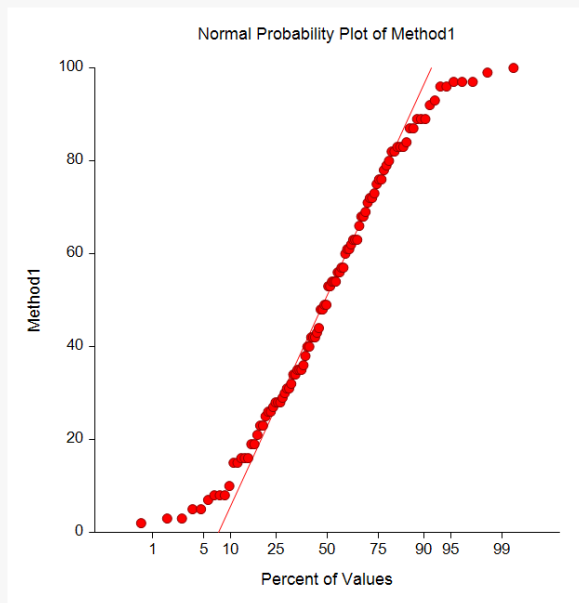


If any of the observations fall outside the confidence bands (if shown), the data are not normal. The goodness-of-fit test mentioned earlier should confirm this fact statistically. If only one observation falls outside the confidence bands and the remaining observations hug the straight line, there may be an outlier. If the data were normal, we would see the points falling along a straight line. Note that the confidence bands are based on large-sample formulas. They may not be accurate for small samples.

# Average vs Difference Plot

**Average vs Difference Plot**
_____
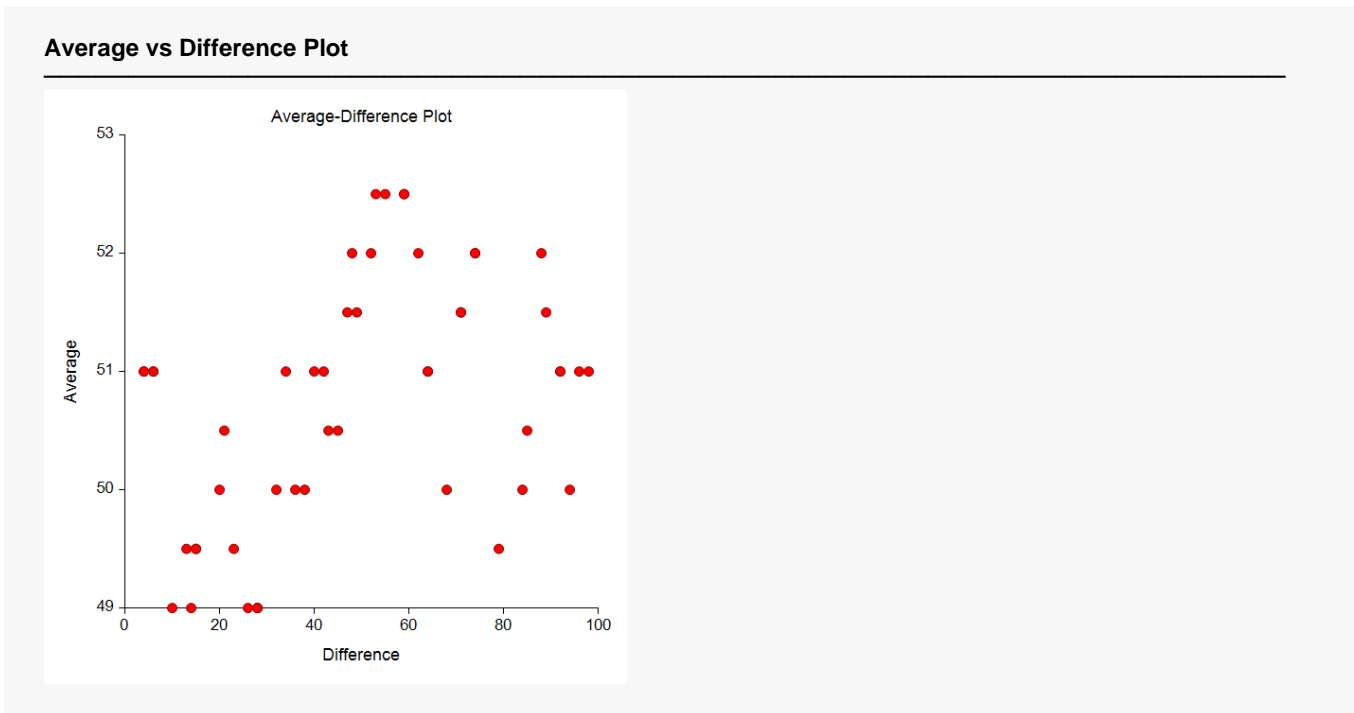


This average-difference plot is designed to detect a lack of symmetry in the data. This plot is constructed from the original data. Let D($i$) represent the $i^{th}$ ordered difference. Pairs of these sorted differences are considered, with the pairing being done as you move toward the middle from either end. That is, consider the pairs D(1) and D(n), D(2) and D(n-1), D(3) and D(n-2), etc. Plot the average versus the difference of each of these pairs. Your plot will have about n/2 points, depending on whether n is odd or even. If the data are symmetric, the average of each pair will be the median and the difference between each pair will be zero.

Symmetry is an important assumption for the t-test. A perfectly symmetric set of data should show a vertical line of points hitting the horizontal axis at the value of the median. Departures from symmetry would deviate from this standard.