



predictum

Equivalence Testing

minimum requirements:
JMP v 7.02 + curiosity
Copyright © 2003 - 2009 Predictum Inc.

Learning Objectives

Difference between statistically significant differences and practical differences

Conducting and interpreting Equivalence tests

The effects of alpha, sample size, variability and difference on Equivalence tests

The meaning of Two One-sided tests (TOST).

Parameters

Difference considered practically zero, $\delta = 2$

Actual $\delta = 1.8$

$\sigma = 1$

Sample Size across both groups, $n = 30$

Two One-Sided Tests (TOST)

Equivalence Test

Test	t Ratio	p-Value
Upper Threshold	-0.4796	0.3182
Lower Threshold	9.11208	<.0001*
Max over both		0.3182

With enough data, any difference, no matter how small, can be found to be statistically significant with a conventional t-test. Equivalence tests are an alternative to the conventional t-test that investigates *practical* differences.

Enter the minimum, practical difference via the slider

Simulate an "actual difference" with this slider

Adjust sigma, the variability, in the data via this slider

See how having more or less data affects the power of the analysis here

Note that each of these parameters can be edited directly by simply clicking and editing the number to the right.

Click here to reset the parameters.

Parameters

Difference considered practically zero, $\delta = 2$

Actual $\delta = 1.8$

$\sigma = 1$

Sample Size across both groups, $n = 30$

Re-initialize Parameters

Simulate Analysis

Generate Data & Analysis

Close Data Table & Analysis

Display Options

Thresholds Confidence Interval

Both Show

Lower Only Hide

Upper Only

Download Instructions

Two One-Sided Tests (TOST)

Equivalence Test

Specified Practical Difference Threshold	2.000	
Actual Difference in Means	1.800	
Std Error of Difference	0.417	
Test	t Ratio	p-Value
Upper Threshold	-0.4796	0.3182
Lower Threshold	9.11208	<.0001*
Max over both		0.3182

Thresholds indicate the practical differences on both sides.

The actual difference. If this difference is "the same" (ie: no practical difference) as one of the thresholds, the Max p-value is non-significant.



PREDICTUM Equivalence Testing TOST v1.70beta
☰

Parameters

☐
Difference considered practically zero, $\delta = 1.79$

☐
Actual $\delta = 1.51$

☐
 $\sigma = 2.78$

☐
Sample Size across both groups, $n = 163$

Re-initialize Parameters

Simulate Analysis

Generate Data & Analysis

Close Data Table & Analysis

Display Options

Thresholds	Confidence Interval
<input checked="" type="radio"/> Both	<input checked="" type="radio"/> Show
<input type="radio"/> Lower Only	<input type="radio"/> Hide
<input type="radio"/> Upper Only	Set α
	☐
	$\alpha = 0.05$

Download Instructions

Two One-Sided Tests (TOST)

Equivalence Test

Specified Practical Difference Threshold	1.790
Actual Difference in Means	1.510
Std Error of Difference	0.435

Test	t Ratio	p-Value
Upper Threshold	-0.6429	0.2606
Lower Threshold	7.57762	<.0001*
Max over both		0.2606

Here, the difference is practically significant because the actual difference (vertical red line) is not significantly different than one of the thresholds.

The p-value for the Upper Threshold is not significant. The "Max over both" is the higher of the two p-values. If it is above alpha, the difference is practically significant.

Parameters

Difference considered practically zero, $\delta = 1.79$

Actual $\delta = 0.66$

$\sigma = 1$

Sample Size across both groups, $n = 30$

Re-initialize Parameters

Simulate Analysis

Generate Data & Analysis

Close Data Table & Analysis

Display Options

Thresholds	Confidence Interval
<input checked="" type="radio"/> Both	<input type="radio"/> Show
<input type="radio"/> Lower Only	<input checked="" type="radio"/> Hide
<input type="radio"/> Upper Only	

Download Instructions

Two One-Sided Tests (TOST)

Equivalence Test

Specified Practical Difference Threshold 1.790		
Actual Difference in Means 0.660		
Std Error of Difference 0.417		
Test	t Ratio	p-Value
Upper Threshold	-2.7096	0.0057*
Lower Threshold	5.87489	<.0001*
Max over both		0.0057*

Here, the difference is not practical. The actual difference is significantly different than either of the thresholds.

The p-value for the Upper Threshold is not significant. The "Max over both" is the higher of the two p-values. If it is above alpha, the difference is practically significant.

Parameters

Difference considered practically zero, $\delta = 1.79$

Actual $\delta = 1.51$

$\sigma = 2.78$

Sample Size across both groups, $n = 163$

Re-initialize Parameters

Simulate Analysis

Generate Data & Analysis

Close Data Table & Analysis

Display Options

Thresholds Confidence Interval

Both Show

Lower Only Hide

Upper Only Set α

Download Instructions

Two One-Sided Tests (TOST)

Equivalence Test

Specified Practical Difference Threshold	1.790
Actual Difference in Means	1.510
Std Error of Difference	0.435
Test	t Ratio p-Value
Upper Threshold	-0.6429 0.2606
Lower Threshold	7.57762 <.0001*
Max over both	0.2606

Reveal and conceal the confidence interval with these radio buttons

When the confidence interval is revealed, a slider appears to adjust alpha

Power is an important consideration. Here, there is a practical difference and the confidence interval suggests that the analysis was sufficiently powerful.

A situation where the confidence interval crosses over the 0 difference line indicates that there was not enough power.



PREDICTUM Equivalence Testing TOST v1.70beta
☰

Parameters

☐
Difference considered practically zero, $\delta = 1.79$

☐
Actual $\delta = 1.51$

☐
 $\sigma = 2.78$

☐
Sample Size across both groups, $n = 30$

Re-initialize Parameters

Simulate Analysis

Generate Data & Analysis

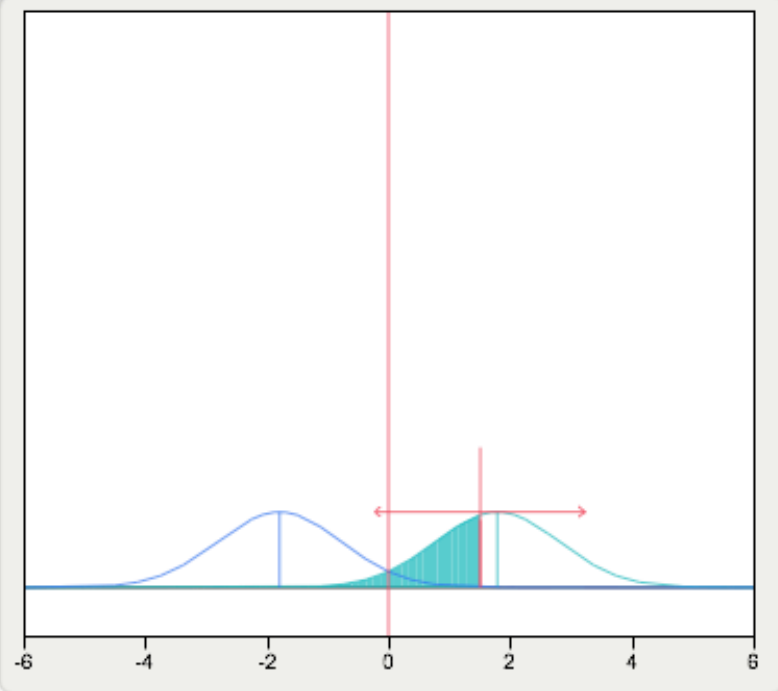
Close Data Table & Analysis

Display Options

Thresholds	Confidence Interval
<input checked="" type="radio"/> Both	<input checked="" type="radio"/> Show
<input type="radio"/> Lower Only	<input type="radio"/> Hide
<input type="radio"/> Upper Only	Set α
	☐ $\alpha = 0.05$

Download Instructions

Two One-Sided Tests (TOST)



Equivalence Test

Specified Practical Difference Threshold	1.790
Actual Difference in Means	1.510
Std Error of Difference	1.015

Test	t Ratio	p-Value
Upper Threshold	-0.2758	0.3924
Lower Threshold	3.25087	0.0015*
Max over both		0.3924

This situation is the same as on the previous page except the sample size has been greatly reduced. As a result, the distributions are much more spread as the standard error is bigger due to the drop in sample size.

Note that although the "Max over both" p-value is not significant, the confidence interval is very wide, stretching over the 0 line.

Parameters

Difference considered practically zero, $\delta = 1.79$

Actual $\delta = 0.85$

$\sigma = 2.78$

Sample Size across both groups, $n = 163$

Two One-Sided Tests (TOST)

Equivalence Test

Specified Practical Difference Threshold	1.790	
Actual Difference in Means	0.850	
Std Error of Difference	0.435	
Test	t Ratio	p-Value
Upper Threshold	-2.1585	0.0162*
Lower Threshold	6.0621	<.0001*
Max over both		0.0162*

Click here to create a data table with simulated data and corresponding analysis generated using the parameters indicated here.

Click here to close the generated table so that you can change the parameters and create another data table/analysis set.



As with any JMP application, click toward either end of the axes and drag to expand or contract the scale

PREDICTUM Equivalence Testing TOST v1.70

Parameters

Difference considered practically zero, $\delta = 2$

Actual $\delta = 1.8$

$\sigma = 1$

Sample Size across both groups, $n = 30$

Two One-Sided Tests (TOST)

Simulate Analysis

- Generate Data & Analysis
- Close Data Table & Analysis

Display Options

Thresholds Confidence Interval

Both Show

Lower Only Hide

Upper Only

Equivalence Test

Specified Practical Difference Threshold 2.000
Actual Difference in Means 1.800
Std Error of Difference 0.365

Test	t Ratio	p-Value
Upper Threshold	-0.5477	0.2941
Lower Threshold	10.4067	<.0001*
Max over both		0.2941

Collect all of our illustrations
by visiting our website.



predictum

click logo to view



BOOST PRODUCTIVITY

JMP SCRIPTING LANGUAGE
DEVELOPMENT SERVICE



APPLIED STATISTICAL ESSENTIALS

a fun and accessible way for analysts
everywhere to learn and apply
statistical methods.

click above to view