Two-Sample Equivalence Tests for Location

Rev. 1.0, PGM, 10/12/04
Rev. 1.1, PGM, 10/25/04, added graphs and proofread

In a normal two-sample hypothesis or significance test for location, the hypotheses are:

$$
\begin{aligned}
H_0 &: \quad \Delta\mu = 0 \\
H_A &: \quad \Delta\mu \neq 0
\end{aligned}
$$

where $\Delta\mu = \mu_1 - \mu_2$ and acceptable decisions are to either: 1) reject H_0 in favor of H_A or 2) to not reject H_0 or reserve judgment. Such tests are really about the claim made in H_A, which is either accepted or not accepted. When H_A is accepted, that means there is evidence of a statistically significant difference between μ_1 and μ_2.

Sometimes hypotheses tests are performed by constructing the $(1 - \alpha)\,100\%$ confidence interval for $\Delta\mu$ and observing whether the interval does not contain zero, in which case we reject H_0 in favor of H_A, or whether the interval does contain zero, in which case we cannot reject H_0. Both of these methods, the hypothesis test and the confidence interval method, have analogies in equivalence testing.

When the purpose of a test is to demonstrate the equality of two treatment means, we perform an *equivalence test*, where the hypotheses are effectively:

$$
\begin{aligned}
H_0 &: \quad \Delta\mu \neq 0 \\
H_A &: \quad \Delta\mu = 0
\end{aligned}
$$

Since it is impossible to be certain that H_A is *exactly* true, because there could still be a practically insignificant but undetectable bias between μ_1 and μ_2, we use different forms for these hypotheses - forms that formally consider the possibility of such a bias:

$$
\begin{aligned}
H_0 &: \; |\Delta\mu| \geq \delta \\
H_A &: \; |\Delta\mu| < \delta
\end{aligned}
$$

where δ, called the *limit of practical equivalance* (LOPE), has to be sufficiently small so that the claim that $\Delta\mu = 0$ is, for all practical purposes, still justified. We say that if the data support $H_A : |\Delta\mu| < \delta$, then μ_1 and μ_2 are practically equivalent.

The equivalence test is performed using two one-sided tests of means (TOST). The absolute values in the original equivalence test hypotheses are broken up into two separate tests:

$$\begin{aligned} H_{01} &: \quad \Delta\mu \leq -\delta \text{ versus } H_{A1} : \Delta\mu > -\delta \\ H_{02} &: \quad \Delta\mu \geq \delta \text{ versus } H_{A2} : \Delta\mu < \delta \end{aligned}$$

If, on the basis of the sample data, both H_{01} *and* H_{02} can be rejected, then we can accept the claim of the equivalence of μ_1 and μ_2 within the practical limits $-\delta < \Delta\mu < \delta$. For the two-sample t test with unknown but equal variances, the test statistics for H_{01} and H_{02} are, respectively:

$$t_1 = \frac{\Delta\bar{x} + \delta}{s_{\epsilon(\Delta\bar{x})}} \tag{1}$$

and

$$t_2 = \frac{\Delta\bar{x} - \delta}{s_{\epsilon(\Delta\bar{x})}} \tag{2}$$

where, as usual, the standard deviation of the distribution of $\Delta\bar{x}$ is:

$$s_{\epsilon(\Delta\bar{x})} = \sqrt{\frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \tag{3}$$

The rejection conditions for H_{01} and H_{02} are, respectively:

$$t_1 > t_{\alpha,\nu} \tag{4}$$

and

$$t_2 < -t_{\alpha,\nu} \tag{5}$$

where:

$$\nu = n_1 + n_2 - 2 \tag{6}$$

The TOST null hypotheses are shown graphically in Figure 1.

Note that: 1) it is necessary to reject both H_{01} *and* H_{02} to accept $H_A : |\Delta\mu| < \delta$ and 2) despite the fact that both of the one-sided tests use significance , their combination still gives significance for the equivalence test. When it is necessary to report a p value for the equivalence test, the larger of the two p values associated with t_1 and t_2 must be reported.

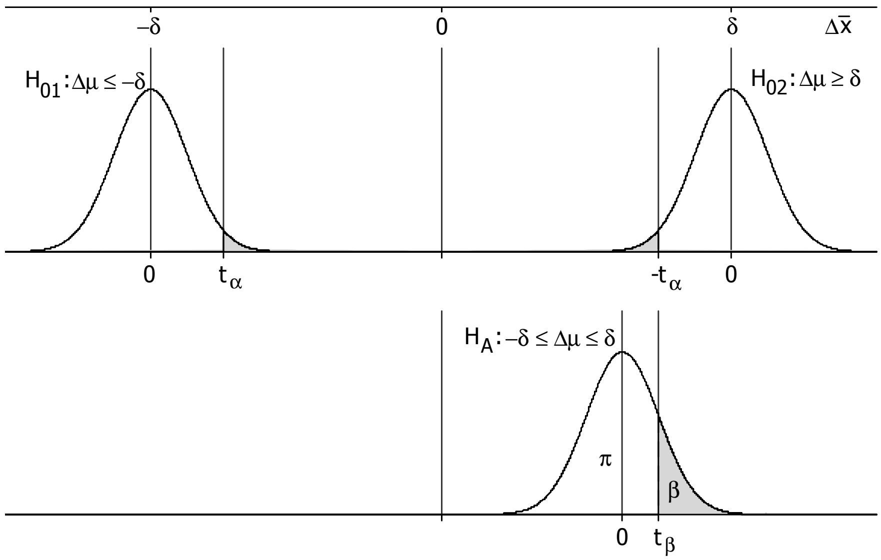

Figure 1: Two one-sided test approach to equivalence testing.

An alternative and equivalent method of performing the equivalence test, which is much easier to understand because it can be presented graphically, involves the consideration of the overlap of the interval $-\delta < \Delta\bar{x} < \delta$ with the $(1-2\alpha)\,100\%$ confidence inteval for $\Delta\mu$ determined from the sample data. The confidence interval is given by:

$$P\left(\Delta\bar{x} - t\ s_{\epsilon(\Delta\bar{x})} < \Delta\mu < \Delta\bar{x} + t\ s_{\epsilon(\Delta\bar{x})}\right) = 1 - 2\alpha \tag{7}$$

If the confidence interval falls completely within the interval $-\delta < \Delta\bar{x} < \delta$, then we can reject H_0 in favor of the claim of practical equivalence H_A: $|\Delta\mu| < \delta$. If, however, the confidence interval falls outside of the interval $-\delta < \Delta\bar{x} < \delta$, in part or totally, then we cannot reject H_0.

The derivation of the equivalence of the TOST method with the confidence interval method is based on the manipulation and combination of Equations 1, 2, 4, and 5 which leads to the following condition under which H_0 can be rejected and H_A can be accepted:

$$-\delta < \Delta\bar{x} - t\ s_{\epsilon(\Delta\bar{x})} < \Delta\bar{x} + t\ s_{\epsilon(\Delta\bar{x})} < \delta \tag{8}$$

The inner pair of terms in this compound inequality represent the end points

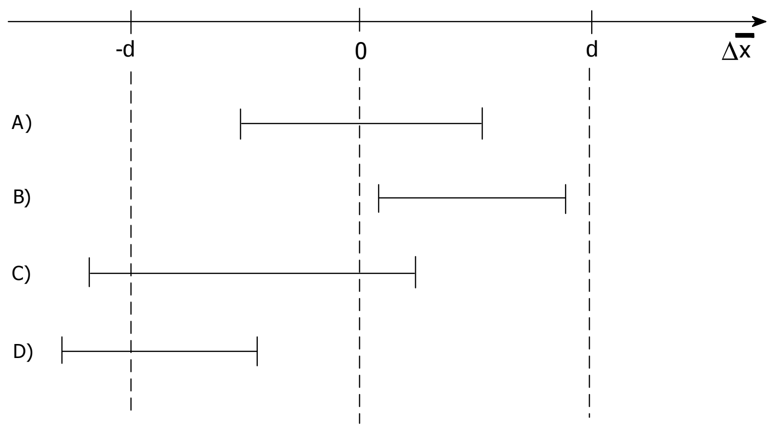

Figure 2: Confidence intervals for significance and equivalence tests.

of the $(1-2\alpha)\,100\%$ confidence interval for $\Delta\mu$ as in Equation 7. To summarize, Equation 8 indicates that we can reject H_0 in favor of H_A when the $(1-2\alpha)\,100\%$ confidence interval for $\Delta\mu$ falls completely inside of the interval $-\delta < \Delta\bar{x} < \delta$.

The easiest way to perform the two-sample location test for equivalence in MINITAB is to construct the necessary confidence interval from the **Stat**> **Basic Stats**> **2-Sample t** menu. Use the **Options** menu to set the appropriate $(1-2\alpha)\,100\%$ confidence level.

Example 1 *Interpret the four confidence intervals in Figure 2 in the context of: 1)* H_0: $\Delta\mu = 0$ *versus* H_A: $\Delta\mu \neq 0$ *and 2)* H_0: $|\Delta\mu| > \delta$ *versus* H_A: $|\Delta\mu| < \delta$.
Solution: *In case 1, which tests for a significant difference between* μ_1 *and* μ_2, *the confidence intervals must be checked to see if they do or don't contain zero. For those cases that are slipped from zero (B and D),* H_0 *may be rejected in favor of* H_A: $\Delta\mu \neq 0$. *In case 2, which tests for the equivalence between* μ_1 *and* μ_2, *the confidence intervals must be checked to see if they do or don't fall within the interval* $-\delta < \Delta\bar{x} < \delta$. *For those cases that fall completely inside of the interval, (A and B),* H_0 *may be rejected in favor of* H_A: $|\Delta\mu| < \delta$.

Example 2 *Revise the analysis of the two-sample location equivalence test for the one-sample case.*

Solution: *The hypotheses for the one-sample equivalence test are:*

$$H_0 : |\Delta\mu| > \delta$$
$$H_A : |\Delta\mu| < \delta$$

where $\Delta\mu = \mu - \mu_0$ *and* μ_0 *is a specified constant. The test statistics are:*

$$t_1 = \frac{\bar{x} + \delta}{s_{\epsilon(\bar{x})}} \tag{9}$$

$$t_2 = \frac{\bar{x} - \delta}{s_{\epsilon(\bar{x})}} \tag{10}$$

where the standard deviation of the distribution of sample $\bar{x}$*s is:*

$$s_{\epsilon(\bar{x})} = \frac{s}{\sqrt{n}} \tag{11}$$

Reject H_0 *in favor of* H_A *if both* $t_1 > t$ *and* $t_2 < -t$ *or if the* $(1 - 2)\,100\%$ *confidence interval for* μ *falls completely inside of the interval* $-\delta < \bar{x} < \delta$*. The test statistics and the confidence interval for* μ *are interpreted the same way as for the two-sample case.*