Chapter 583

Studentized Range Tests for Equivalence

Introduction

This procedure computes power and sample size of equivalence tests of the means of two or more groups which are analyzed using a studentized range tests. It turns out that the question of equivalence among a set of means is often more meaningful than the question of equality.

Methodology for testing the equivalence of the means of two groups has received much attention. However, testing equivalence among three or more groups has not received much attention. The article by Shieh (2016) gives results for two competing test procedures: The F-test and the studentized range test. Results for the F-test are available in **PASS** in another procedure. This procedure provides power and sample size results the studentized range test.

While the F-test is by far the most commonly used method for testing the equality of two or more means, Shieh (2016) showed that when testing for mean equivalence, neither test is always optimal. In fact, the studentized range test is more powerful when the actual range is close to the equivalence boundary.

Technical Details for the Studentized Range Test

Suppose G groups each have a normal distribution and with means $\mu_1, \mu_2, \ldots, \mu_G$ and common variance σ^2. Let $N_1, N_2, \ldots, N_G = N_i$ denote the common sample size of all groups and let N denote the total sample size. In this case of equal group sizes, $N = GN_i$. The multigroup equivalence problem requires one to show that the means are sufficiently close to each other. Shieh (2016) considering whether the difference between the minimum and maximum means (the range of the means) is sufficiently small so that the differences among the means can be regarded as of no practical importance.

The One-Way Model

Consider the usual one-way fixed-effects model

$$Y_{gj} = \mu_g + \varepsilon_{gj}$$

where Y_{gj} is response, μ_g are the treatment means, and ε_{gj} are the independent, normally distributed error with zero mean and common variance σ^2. Here the subscript g indexes the G groups, and the subscript j indexes the N_i subjects in each group.

Cohen (1988) showed that hypotheses about the G means may be obtained using either the variance of the means in terms of the F-test or their range in terms of the studentized range.

Equivalence Hypothesis

The hypothesis of mean equivalence is

$$H_0: \frac{\delta}{\sigma} \geq \frac{\delta_0}{\sigma} \ versus \ H_1: \frac{\delta}{\sigma} < \frac{\delta_0}{\sigma}$$

where $\delta = \mu_{Max} - \mu_{Nin}$ represents the range and δ_0 is the equivalence bound.

Studentized Range Statistic

The studentized range statistic is defined as follows

$$Q = \frac{\left[\max_{g=1\ to\ G}(\bar{Y}_g) - \min_{g=1\ to\ G}(\bar{Y}_g)\right]\sqrt{N_i}}{S}$$

where $\bar{Y}_g$ are the sample means, and S is the sample variance.

It turns out that the distribution of Q is a function of the pairwise mean differences $\mu_g - \mu_h$, not just the range (the maximum of these differences).

The cumulative distribution function, from which the power can be computed, is given by

$$\Theta(q) = P\{Q \leq q\} = E_K\left\{\sum_{g=1}^{G} E_{Z_g}\left[\prod_{\substack{h=1 \\ h \neq g}}^{G}\left(\Phi\{Z_g + \delta_{gh}\sqrt{N_i}\} - \Phi\{Z_g + \delta_{gh}\sqrt{N_i} - q\sqrt{K/(N-G)}\}\right)\right]\right\}$$

where $\delta_{gh} = \mu_g - \mu_h$, K is a chi-square random variable with $N - G$ degrees of freedom, $\Phi\{z\}$ is the CDF of a standard normal distribution, Z_g are independent standard normal random variables, $E_K\{x\}$ is the expectation with respect to K, and $E_{Z_g}\{x\}$ is the expectation with respect to Z_g.

Note that the critical value is based on the set of group means. It cannot be determined from just δ_0. When only δ_0 is specified, the least favorable configuration (LFC) of the means is used. This is given by

$$\{\mu_1, \dots, \mu_G\} = \left\{-\frac{\delta_0}{2}, \frac{\delta_0}{2}, 0, \dots, 0\right\}$$

If a sample size is desired, it can be determined using a standard binary search algorithm.

Example 1 – Finding Power

An experiment is being designed to assess the equivalence of the means of four groups using a studentized range test with a significance level of 0.05. Previous studies have shown a standard deviation of 2. The maximum range of the four means allowed in equivalent means is 2. Power calculations assume that the actual range is 1. To better understand the relationship between power and sample size, the researcher wants to compute the power for several group sample sizes between 20 and 100. The sample sizes will be equal across all groups.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the procedure window. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Power**
Alpha	**0.05**
G (Number of Groups)	**4**
Ni (Sample Size Per Group)	**20 40 60 80 100**
μ1i's Input Type	**Enter Range of Means|H1**
δ1 (Range of Means|H1)	**1**
μ0i's Input Type	**Enter Range of Means|H0**
δ0 (Range of Means|H0 or Equiv Limit)	**2**
σ (Standard Deviation)	**2**

Annotated Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results
Number of Groups 4

Power	Total Sample Size N	Sample Size Per Group Ni	H1 Range of Means δ1	H0 Range of Means δ0	Std Dev σ	Alpha
0.3921	80	20	1	2	2	0.05
0.6847	160	40	1	2	2	0.05
0.8463	240	60	1	2	2	0.05
0.9270	320	80	1	2	2	0.05
0.9659	400	100	1	2	2	0.05

References
Shieh, G. 2016. 'A comparative appraisal of two equivalence tests for multiple standardized effects'. Computer Methods and Programs in Biomedicine, Vol 126, Pages 110-117. http://dx.doi.org/10.1016/j.cmpb.2015.12.004
Wellek, Stefan. 2010. Testing Statistical Hypotheses of Equivalence and Noninferiority, 2nd Edition. CRC Press. New York.
Cohen, Jacob. 1988. Statistical Power Analysis for the Behavioral Sciences. Lawrence Erlbaum Associates. Hillsdale, New Jersey

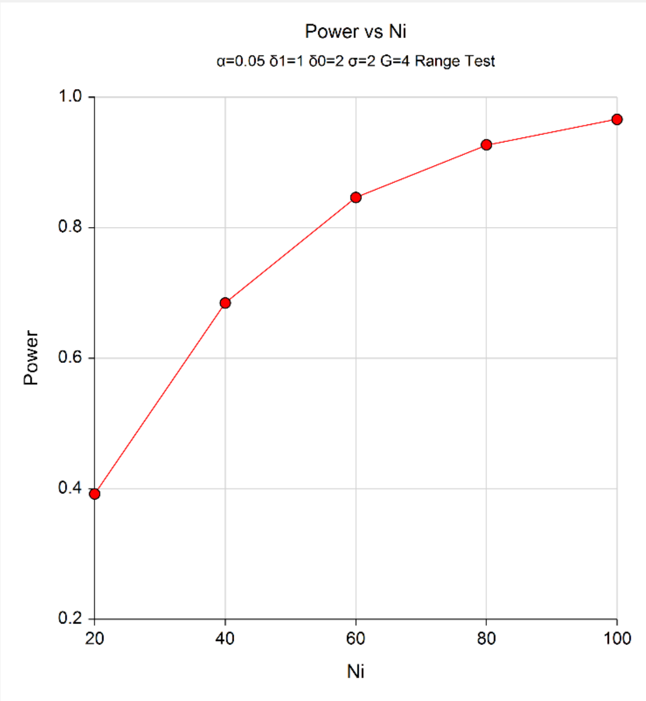
Report Definitions
Power is the probability of rejecting a false null hypothesis of non-equivalence in favor of the alternative hypothesis of equivalence.
Total Sample Size N is the total number of subjects in the study.
Sample Size Per Group Ni is the number of items sampled per group.
H1 Range of Means δ1 is the range of the group means assumed by the alternative hypothesis. It is the value at which the power is computed. Note that you must have δ1 < δ0.
H0 Range of Means δ0 is the range of the group means assumed by the null hypothesis. This value is the equivalence limit (bound). Note that you must have δ1 < δ0.
Std Dev σ is the standard deviation of the responses for all groups.
Alpha is the significance level of the test: the probability of rejecting the null hypothesis of non-equivalent means when it is actually true.

Summary Statements
In a one-way equivalence study, a sample of 80 subjects, divided among 4 groups, achieves a power of 39%. This power assumes a studentized range test of equivalence with a significance level of 0.05. The group subject counts are 20. The range boundary for declaring equivalence is 2. The range at which the power is computed is 1. The standard deviation of all groups is 2.

This report shows the numeric results of this power study.

Chart Section

Chart Section



This plot gives a visual presentation of the results in the Numeric Report.

Example 2 – Finding the Sample Size Necessary to Reject

Continuing with the last example, we will determine how large the sample size would need to have been for alpha = 0.05 and power = 0.80 or 0.9.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the procedure window. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Sample Size**
Power	**0.8 0.9**
Alpha	**0.05**
G (Number of Groups)	**4**
μ1i's Input Type	**Enter Range of Means|H1**
δ1 (Range of Means|H1)	**1**
μ0i's Input Type	**Enter Range of Means|H0**
δ0 (Range of Means|H0 or Equiv Limit)	**2**
σ (Standard Deviation)	**2**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results
Number of Groups 4

Power	Total Sample Size N	Sample Size Per Group Ni	H1 Range of Means δ1	H0 Range of Means δ0	Std Dev σ	Alpha
0.8014	212	53	1	2	2	0.05
0.9013	288	72	1	2	2	0.05

This report shows the necessary sample sizes for achieving powers of 0.8 and 0.9.

Example 3 – Validation using Shieh (2016)

Shieh (2016) page 115 presents an example in which alpha = 0.05, G = 4, σ = 7.47583, δ1 = 0.1821(7.47583) = 1.36135, δ0 = 7.47583, and power = 0.8. The resulting sample size is 27 per group for a total of 108.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the procedure window. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Sample Size**
Power	**0.8**
Alpha	**0.05**
G (Number of Groups)	**4**
μ1i's Input Type	**Enter Range of Means|H1**
δ1 (Range of Means|H1)	**1.36135**
μ0i's Input Type	**Enter Range of Means|H0**
δ0 (Range of Means|H0 or Equiv Limit)	**7.47583**
σ (Standard Deviation)	**7.47583**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results
Number of Groups 4

Power	Total Sample Size N	Sample Size Per Group Ni	H1 Range of Means δ1	H0 Range of Means δ0	Std Dev σ	Alpha
0.811	108	27	1.361	7.476	7.476	0.05

PASS also found Ni = 27 and N = 108.

Example 4 – Comparing the Impact of Mean Configuration on Power and Sample Size

This example will look at the impact the configuration of the means assumed by the null hypothesis on the required sample size. The three configurations suggested by Cohen (1988) will be studied.

Let alpha = 0.05, G = 5, σ = 1, δ1 = 0.4, and power = 0.8. The following three configurations will be entered in the spreadsheet and used to specify δ0. Note that all three configurations result in δ0 = 0.8.

Column	Means|H0	Comments
C1	-0.4, -0.4, 0.4, 0.4, 0.4	Minimum mean variation
C2	-0.4, -0.2, 0, 0.2, 0.4	Intermediate mean variation
C3	-0.4, 0, 0, 0, 0.4	Maximum mean variation

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the procedure window. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Sample Size**
Power	**0.8**
Alpha	**0.05**
G (Number of Groups)	**5**
μ1i's Input Type	**Enter Range of Means|H1**
δ1 (Range of Means|H1)	**0.4**
μ0i's Input Type	**Enter Columns Containing Sets of μ0i's**
Columns Containing Sets of μ0i's	**1-3**
σ (Standard Deviation)	**1**

Input Spreadsheet Data

Row	C1	C2	C3
1	-0.4	-0.4	-0.4
2	-0.4	-0.2	0.0
3	0.4	0.0	0.0
4	0.4	0.2	0.0
5	0.4	0.4	0.4

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results
Number of Groups 5

Power	Total Sample Size N	Sample Size Per Group Ni	H1 Range of Means δ1	H0 Group Means Set μ0i	H0 Range of Means δ0	Std Dev σ	Alpha
0.8106	145	29	0.4	C1(1)	0.8	1	0.05
0.8032	350	70	0.4	C2(2)	0.8	1	0.05
0.8028	415	83	0.4	C3(3)	0.8	1	0.05

Name	Values
C1(1)	-0.4, -0.4, 0.4, 0.4, 0.4
C2(2)	-0.4, -0.2, 0, 0.2, 0.4
C3(3)	-0.4, 0, 0, 0, 0.4

These results show that the sample size required varies from 145 to 415 depending on the configuration.