Chapter 564

Equivalence Tests for One-Way Analysis of Variance Allowing Unequal Variances

Introduction

This procedure computes power and sample size of equivalence tests of multiple means which are analyzed using an extension of the Welch test.

The results in this chapter come from Jan and Shieh (2019), Wellek (2010), and Welch (1951).

Technical Details

Background

Suppose G groups each have a normal distribution and with means $\mu_1, \mu_2, \ldots, \mu_G$ and standard deviations $\sigma_1, \sigma_2, \ldots, \sigma_G$. Let $n_1, n_2, \ldots, n_G$ denote the sample size of each group and let N denote the total sample size of all groups. The multigroup equivalence problem requires one to show that the standardized means are sufficiently close to each other. Wellek (2010) accomplished this by defining a set of standardized equivalence means (μ_i/σ_i), that are as far apart as possible and can still be termed *equivalent*. These means are denoted as $\mu_{01}, \mu_{02}, \ldots, \mu_{0G}$. Jan and Shieh (2019) present power formulas for the equivalence extension of the ADF-Welch test.

Jan and Shieh (2019) suggest testing the equivalence of the standardized means using the statistical hypotheses

$$H_0: \omega^2 \geq \omega_0^2 \ versus \ H_1: \omega^2 < \omega_0^2$$

Let ω_1^2 represent the variation in the standardized means under the alternative hypothesis of equivalence and ω_0^2 represent the variation in the standardized means under the null hypothesis of non-equivalence.

Test Statistic

ANOVA F-Test

Assuming homogeneity of variance among the groups, the most popular procedure for analyzing a set of G means is the ANOVA F-Test which is calculated as follows.

$$F^* = \frac{SSM/(G-1)}{SSE/(N-G)}$$

where *SSM* is the sum of squares of treatment means, *SSE* is the sum of squares of error, and *N* is the total sample size.

Welch's Test

If variance heterogeneity is suspected, a common approach is to use Welch's procedure which is calculated as follows.

$$W = \frac{\sum_{i=1}^{G} W_i(\bar{X}_i - \bar{X})/(G-1)}{1 + 2(G-2)Q/(G^2-1)}$$

where

$W_i = n_i/S_i^2$, $S_i^2 = \sum_{j=1}^{n_i}\left(X_{ij} - \bar{X}_i\right)^2/(n_i - 1)$, $\bar{X}_i = \sum_{j=1}^{n_i} X_{ij}/n_i$, $\bar{X} = \sum_{i=1}^{G} W_i\bar{X}_i/U$, $U = \sum_{i=1}^{G} W_i$, and $Q = \sum_{i=1}^{G}\left(1 - \frac{W_i}{U}\right)^2/(n_i - 1)$.

Under the null hypothesis, Welch (1951) gave the approximate distribution of *W* as an *F* distribution with degrees of freedom *G* – 1 and *v*, where

$$v = \frac{G^2 - 1}{3Q}$$

Jan and Shieh's Extension of Welch's Test

Jan and Shieh (2019) proposed testing the equivalence of the *G* group means in the face of variance heterogeneity using the hypotheses

$$H_0: \omega^2 \geq \omega_0^2 \ versus \ H_1: \omega^2 < \omega_0^2$$

where

$\omega^2 = \sum_{i=1}^{G} w_i(\mu_i - \mu^*)^2$, $w_i = \frac{n_i}{N\sigma_i^2}$, $\mu^* = \sum_{i=1}^{G}\frac{w_i\mu_i}{v}$, $v = \sum_{i=1}^{G} w_i$, $\omega_0^2 = \sum_{i=1}^{G} w_i(\mu_{0i} - \mu_0^*)^2$, and $\mu_0^* = \sum_{i=1}^{G}\frac{w_i\mu_{0i}}{v}$.

Under H_0, *W* is assumed to follow the noncentral F distribution $W \sim F'_{G-1,v,\Omega_0}$, where $\Omega_0 = N\omega_0^2$. Note that Ω_0 does not depend on data. Rather, it depends on the user specified set of equivalence values.

The null hypothesis is rejected at the significance level α if $W < F'_{1-\alpha,G-1,v,\Omega_0}$.

Power

The power function of the extended Welch's test computed at a particular set of means, $\mu_{11}, \mu_{12}, \dots, \mu_{1G}$, is given by

$$\text{Power} = \Pr\left[F'_{G-1,N-G,\Omega_1} < F'_{1-\alpha,G-1,\eta,\Omega_0}\right]$$

where

$\Omega_1 = N\omega_1^2$, $\eta = (G^2-1)/(3\tau)$, $\tau = \sum_{i=1}^{G}\left(1-\frac{w_i}{v}\right)^2/(n_i-1)$, $\omega_1^2 = \sum_{i=1}^{G} w_i(\mu_{1i}-\mu_1^*)^2$, and $\mu_1^* = \sum_{i=1}^{G}\frac{w_i\mu_{1i}}{v}$.

You have to be careful that $\omega_1^2 < \omega_0^2$ when testing equivalence. When a sample size is desired, it can be determined using a standard binary search algorithm.

Example 1 – Finding Sample Size

An experiment is being designed to assess sample size needed for an equivalence test of four means using the extended Welch test with a significance level of 0.05 and a power of 0.9. Previous studies have shown that the standard deviation in the control group (group 1) is 2. The standard deviations in the three treatment groups are all 4.

The variation allowed in equivalent means is represented by the values {15, 12, 12, 13}. Three sets of alternative treatment means are to be compared: {15, 14, 14, 14}, {15, 13, 13, 13}, and {15, 12, 13, 14}. 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	**Sample Size**
Power	**0.90**
Alpha	**0.05**
G (Number of Groups)	**4**
Group Allocation Input Type	**Equal (n1 = ⋯ = nG)**
μ0 Input Type	**Enter Columns Containing Sets of μ0's**
Columns Containing Sets of μ0's	**5**
μ1 Input Type	**Enter Columns Containing Sets of μ1's**
Columns Containing Sets of μ1's	**2 3 4**
σ Input Type	**Enter Columns Containing Sets of σ's**
Columns Containing Sets of σ's	**1**

Input Spreadsheet Data

Row	C1	C2	C3	C4	C5
1	2	15	15	15	15
2	4	14	13	12	12
3	4	14	13	13	12
4	4	14	13	14	13

Annotated Output

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

Numeric Results

Numeric Results
Number of Groups 4

	Sample Size		Group Means		Grp Std Dev	Std Dev of Standardized Means		
	Total	Grp	H0	H1		H0	H1	
Power	N	ni	μ0	μ1	σ	ω0	ω1	Alpha
0.90580	132	33	C5(1)	C2(1)	C1(1)	0.448	0.164	0.05
0.90114	680	170	C5(1)	C3(2)	C1(1)	0.448	0.327	0.05
0.90004	1716	429	C5(1)	C4(3)	C1(1)	0.448	0.372	0.05

Value Lists

Name	Value
C5(1)	15, 12, 12, 13
C2(1)	15, 14, 14, 14
C3(2)	15, 13, 13, 13
C4(3)	15, 12, 13, 14
C1(1)	2, 4, 4, 4

Group Sample Size Details

n	N	Group Sample Sizes	Group Allocation Proportions
n(1)	132	33, 33, 33, 33	0.25, 0.25, 0.25, 0.25
n(2)	680	170, 170, 170, 170	0.25, 0.25, 0.25, 0.25
n(3)	1716	429, 429, 429, 429	0.25, 0.25, 0.25, 0.25

References
Jan, S-L and Shieh, G. 2019. 'On the Extended Welch Test for Assessing Equivalence of Standardized Means'. Statistics in Biopharmaceutical Research. DOI:10.1080/19466315.2019.1654915
Welch, B.L. 1951. 'On the Comparison of Several Mean Values: An Alternative Approach'. Biometrika,38,330-336.
Wellek, Stefan. 2010. Testing Statistical Hypotheses of Equivalence and Noninferiority, 2nd Edition. CRC Press. New York.

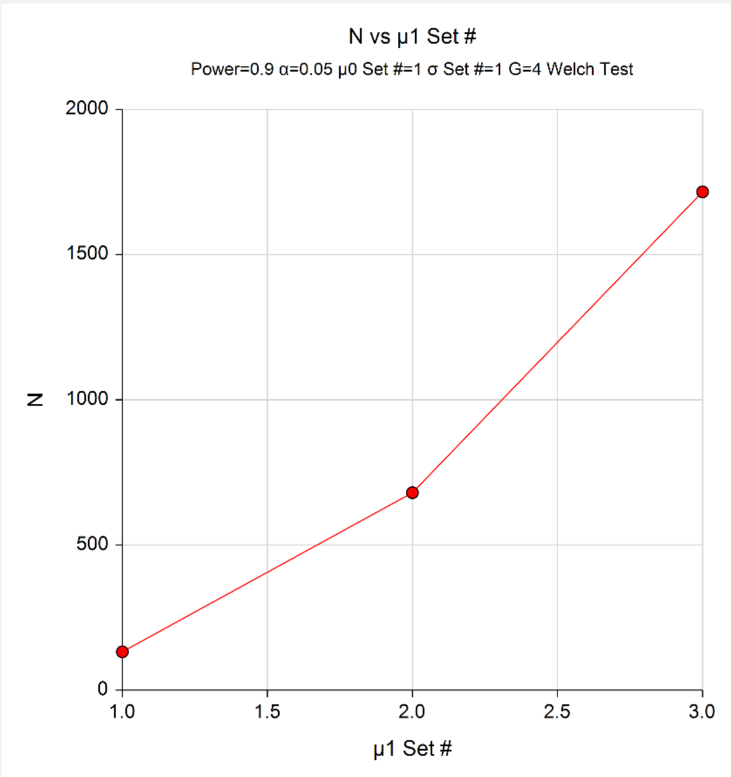
Report Definitions
Power is the probability of rejecting a false null hypothesis of non-equivalence in favor of the alternative hypothesis of equivalence.
N, the Total Sample Size, is the total number of subjects in the study.
ni, the Sample Size per Group, is the number of items sampled from each group in the study.
μ0, the Group Means | H0, is the column name and set number of the group means under the null hypothesis. These values are used to form the equivalence boundary ω0.
μ1, the Group Means | H1, is the column name and set number of the group means under the alternative hypothesis. This is the set of means at which the power is calculated.
σ, the Group Standard Deviations, is the column name and set number of the group standard deviations.
ω0, the Std Dev of Standardized Means, is the population standard deviation of the standardized means, μ0(i)/σ(i), assumed by the null hypothesis, H0. This is the upper bound of equivalence. Note that you must have ω1 < ω0.
ω1, the Std Dev of Standardized Means, is the population standard deviation of the standardized means, μ1(i)/σ(i), assumed by the alternative hypothesis, H1. Note that you must have ω1 < ω0.
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 an equivalence, one-way ANOVA study that allows for unequal group variances, a sample of 132 subjects, divided among 4 groups, achieves a power of 91%. This power assumes the data will be analyzed with Welch's test with a significance level of 0.05. The group subject counts are 33, 33, 33, 33. The group means under the null hypothesis are 15, 12, 12, 13. These means define the equivalence boundary. The group means under the alternative hypothesis are 15, 14, 14, 14. The group standard deviations are 2, 4, 4, 4. The standard deviation of the standardized means under the null hypothesis is 0.448. This is the equivalence boundary. The standard deviation of the standardized means under the alternative hypothesis is 0.164.

This report shows the numeric results of this study.

Chart Section

Chart Section



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

Example 2 – Validation using Jan and Shieh (2019)

Jan and Shieh (2019) page 5 presents an example in which alpha = 0.05, G = 4, the sample sizes are {35, 45, 55, 65}, the standard deviations are {2, 2.82843, 3.46410, 4}, the null means are {0, 0.708, 1.416, 2.124}, and the alternative means are {0, 0.25, 0.5, 0.75}. The resulting power is given as 0.7071.

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	**Power**
Alpha	**0.05**
G (Number of Groups)	**4**
Group Allocation Input Type	**Enter n (Group Sample Sizes)**
n (Group Sample Sizes)	**35 45 55 65**
μ0 Input Type	**Enter μ0 (Group Means|H0)**
μ0 (Group Means|H0)	**0 0.708 1.416 2.124**
μ1 Input Type	**Enter μ1 (Group Means|H1)**
μ1 (Group Means|H1)	**0 0.25 0.5 0.75**
σ Input Type	**Enter σ (Group Standard Deviations)**
σ (Group Standard Deviations)	**2 2.82843 3.4641 4**

Output

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

Numeric Results

Numeric Results
Number of Groups 4

	Sample Size		Group Means		Grp	Std Dev of Standardized Means		
	Total	Grp	H0	H1	Std Dev	H0	H1	
Power	N	n	μ0	μ1	σ	ω0	ω1	Alpha
0.70712	200	n(1)	μ0(1)	μ1(1)	σ(1)	0.269	0.095	0.05

Value Lists

Name	Value
μ0(1)	0, 0.708, 1.416, 2.124
μ1(1)	0, 0.25, 0.5, 0.75
σ(1)	2, 2.82843, 3.4641, 4

Group Sample Size Details

n	N	Group Sample Sizes	Group Allocation Proportions
n(1)	200	35, 45, 55, 65	0.175, 0.225, 0.275, 0.325

PASS also found the power to be 0.7071. The procedure is validated.

Example 3 – Using Patterns and Multipliers

One of the novel features of this and similar procedures is the ability to compare the results of designs with similar mean patterns, but with differing mean magnitudes. This example will show how this is accomplished.

Suppose an experiment with four groups is being designed to assess the equivalence the group means. The analysis will use the Welch-ADF equivalence test. The significance level is 0.05 and per group sample size of 50.

The researchers determine that it will be useful to study only sets of means that follow a pattern in which the means of the first two groups are equal and of the last two groups are equal. Hence the basic pattern of the means is 0, 0, 1, 1 for the four groups. The group standard deviations will be set to either 4, 4, 8, 8 or 5, 5, 10, 10.

Table of Group Means

Groups	1	2	3	4
Basic Pattern	0	0	1	1
H0 Means				
Pattern x 9	0	0	9	9
H1 Means				
Pattern x 8	0	0	8	8
Pattern x 6	0	0	6	6
Pattern x 4	0	0	4	4

These sets of means could be entered on the spreadsheet, but the reports and plots will be easier to interpret if the input uses the *patterns and multipliers* options.

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	**Power**
Alpha	**0.05**
G (Number of Groups)	**4**
Group Allocation Input Type	**Equal to ni (Sample Size per Group)**
ni (Sample Size per Group)	**50**
μ0 Input Type	**Enter μ0 (Group Means|H0)**
μ0 (Group Means|H0)	**0 0 9 9**
μ1 Input Type	**Enter μ1 (Group Means|H1), K1 (μ1 Multiplier)**
μ1 (Group Means|H1)	**0 0 4 4**
K1 (μ1 Multiplier)	**1 1.5 2**
σ Input Type	**Enter σ (Standard Deviations), Ks (σ Multiplier)**
σ (Group Standard Deviations)	**4 4 8 8**
Ks (σ Multiplier)	**1 1.25**

Output

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

Numeric Results

Numeric Results

Number of Groups	4
μ1 (Group Means|H1)	0, 0, 4, 4
σ (Group Standard Deviations)	4, 4, 8, 8

	Sample Size		Group Means			Std Dev		Std Dev of Standardized Means		
	Total	Grp	H0	H1				H0	H1	
Power	N	ni	μ0	μ1	K1	σ	Ks	ω0	ω1	Alpha
0.99921	200	50	μ0(1)	μ1(1)	1.0	σ(1)	1.00	0.712	0.316	0.05
0.99124	200	50	μ0(1)	μ1(1)	1.0	σ(2)	1.25	0.569	0.253	0.05
0.89212	200	50	μ0(1)	μ1(2)	1.5	σ(1)	1.00	0.712	0.474	0.05
0.78057	200	50	μ0(1)	μ1(2)	1.5	σ(2)	1.25	0.569	0.379	0.05
0.24667	200	50	μ0(1)	μ1(3)	2.0	σ(1)	1.00	0.712	0.632	0.05
0.20113	200	50	μ0(1)	μ1(3)	2.0	σ(2)	1.25	0.569	0.506	0.05

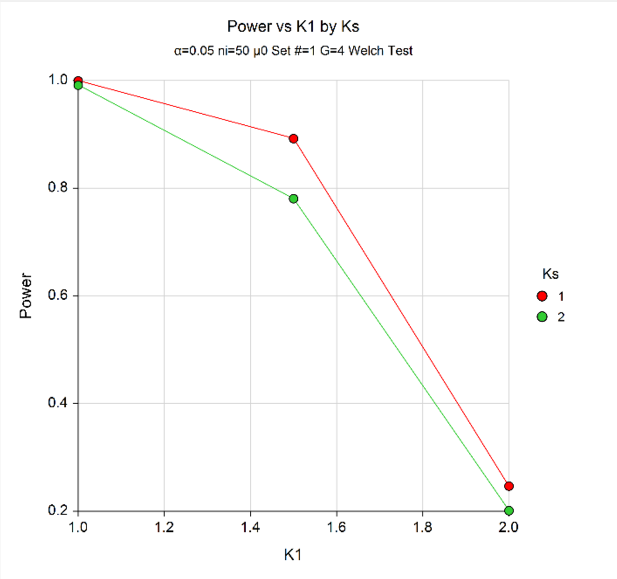
Value Lists

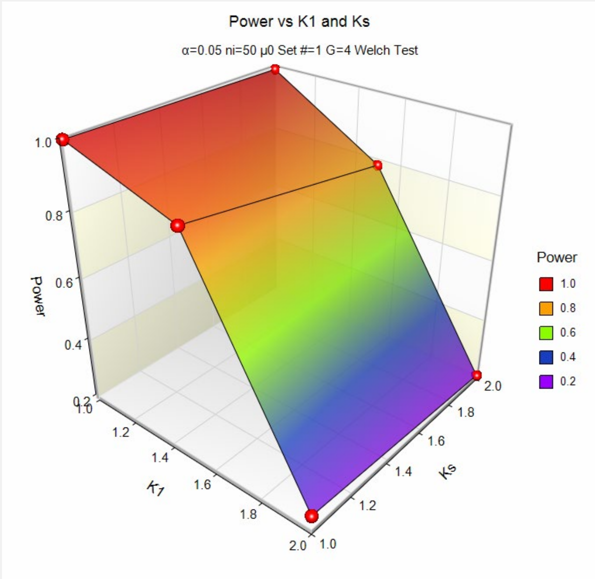
Name	Value
μ0(1)	0, 0, 9, 9
μ1(1)	0, 0, 4, 4
μ1(2)	0, 0, 6, 6
μ1(3)	0, 0, 8, 8
σ(1)	4, 4, 8, 8
σ(2)	5, 5, 10, 10

Group Sample Size Details

n	N	Group Sample Sizes	Group Allocation Proportions
n(1)	200	50, 50, 50, 50	0.25, 0.25, 0.25, 0.25

Chart Section





Notice how easy it is to interpret the reports and plots. The power is pretty good until K1 equals 2. Then it is very low.