

Chapter 596

2x2 Factorial Analysis of Variance Allowing Unequal Variances

Introduction

This procedure computes power and sample size of tests of the means from a 2x2 factorial design which will be analyzed using the Welch-Satterthwaite t-test. This method is recommended when the four group variances are not equal. The results in this chapter come from Jan and Shieh (2016).

The 2x2 factorial design may be used when tests of two factors and their interaction are desired. This procedure allows both the group variances and the sample sizes to be unequal. It also allows the tests to be made in the presence of a non-zero null distribution.

Technical Details

Suppose data from the four groups of a 2x2 factorial design each have a normal distribution with means $\mu_{1,1}, \mu_{1,2}, \mu_{2,1}, \mu_{2,2}$, standard deviations $\sigma_{1,1}, \sigma_{1,2}, \sigma_{2,1}, \sigma_{2,2}$, and sample sizes $n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2}$. Let N denote the total sample size of all groups. The common ANOVA model for this design is

$$Y_{i,j,k} = \mu_0 + A_i + B_j + AB_{i,j} + \varepsilon_{i,j,k}, \text{ where } i = 1, 2 \text{ and } j = 1, 2.$$

Since the variances are unequal, the usual F-test from the ANOVA cannot be used. It is used to show that various comparisons of the means are significantly different from each other. Since each of the three terms that may need to be tested only have one degree of freedom, the results can be analyzed using appropriate contrasts. This procedure provides results for contrasts used in an ANOVA design when the variances are not necessarily equal and a non-zero null value is known.

A contrast of the means may be defined as

$$\delta = \sum_{i=1}^2 \sum_{j=1}^2 c_{i,j} \mu_{i,j}$$

Here, the $c_{i,j}$ are the known contrast coefficients. An unbiased estimate of δ is obtained by replacing the population means by the corresponding sample means.

The three terms in the ANOVA model may be tested using one of the following three sets of contrast coefficients:

c_t	$c_{1,1}$	$c_{1,2}$	$c_{2,1}$	$c_{2,2}$
c_A	-0.5	-0.5	0.5	0.5
c_B	-0.5	0.5	-0.5	0.5
c_{AB}	0.5	-0.5	-0.5	0.5

2x2 Factorial Analysis of Variance Allowing Unequal Variances

A hypothesis test of each term uses the appropriate version of c_t . The null and alternative hypotheses for each model term are given by $H_0: \delta = \delta_0$ versus $H_1: \delta \neq \delta_0$. Here, δ_0 is the value of the contrast for the means at the null hypothesis, H_0 .

The corresponding test statistics are given by

$$T_t = \frac{\hat{\delta} - \delta_0}{\widehat{\sigma(\delta)}}$$

where

$$\delta_0 = \sum_{i=1}^2 \sum_{j=1}^2 c_{i,j} \mu_{0,i,j}$$

$$\hat{\delta} = \sum_{i=1}^2 \sum_{j=1}^2 c_{i,j} \bar{Y}_{i,j}$$

$$\widehat{\sigma^2(\delta)} = \sum_{i=1}^2 \sum_{j=1}^2 c_{i,j}^2 S_{i,j}^2 / n_{i,j}$$

$$\bar{Y}_{i,j} = \sum_{k=1}^{n_{i,j}} Y_{i,j,k} / n_{i,j}$$

$$S_{i,j}^2 = \sum_{k=1}^{n_{i,j}} (Y_{i,j,k} - \bar{Y}_{i,j})^2 / (n_{i,j} - 1)$$

Here $\widehat{\sigma^2(\delta)}$ is an estimator of $\sigma^2(\delta) = \text{Var}(\hat{\delta})$ where

$$\text{Var}(\hat{\delta}) = \sum_{i=1}^2 \sum_{j=1}^2 c_{i,j}^2 \sigma_{i,j}^2 / n_{i,j}$$

Under the null hypothesis, Satterthwaite (1946) and Welch (1947) showed that T is approximately distributed as a Student's t with ν degrees of freedom, where

$$\nu = \frac{(\sum_{i=1}^2 \sum_{j=1}^2 c_{i,j}^2 \sigma_{i,j}^2 / n_{i,j})^2}{\sum_{i=1}^2 \sum_{j=1}^2 c_{i,j}^4 \sigma_{i,j}^4 / [n_{i,j}^2 (n_{i,j} - 1)]}$$

This value, ν , is estimated as

$$\hat{\nu} = \frac{(\sum_{i=1}^2 \sum_{j=1}^2 c_{i,j}^2 S_{i,j}^2 / n_{i,j})^2}{\sum_{i=1}^2 \sum_{j=1}^2 c_{i,j}^4 S_{i,j}^4 / [n_{i,j}^2 (n_{i,j} - 1)]}$$

2x2 Factorial Analysis of Variance Allowing Unequal Variances

The test statistic T has the approximate distribution

$$T \sim t_{\hat{v}}$$

The two-sided, Welch-Satterthwaite test rejects H_0 at a significance level α if $|T| > t_{\hat{v}, 1-\frac{\alpha}{2}}$.

Power

Shieh and Jan (2015) noted that T has the general approximate distribution

$$T \sim t_{\hat{v}, \Delta}$$

where $t_{\hat{v}, \Delta}$ is a noncentral t with \hat{v} degrees of freedom and noncentrality parameter Δ . Here, Δ is defined as

$$\Delta = \frac{\delta_1 - \delta_0}{\sigma(\delta)}$$

where

$$\delta_1 = \sum_{i=1}^2 \sum_{j=1}^2 c_{i,j} \mu_{1,i,j}$$

is the value of the contrast for the means at the alternative hypothesis, H_1 .

Hence, the power can be approximated as

$$Power = P\left(|t_{\hat{v}, \Delta}| > t_{\hat{v}, 1-\frac{\alpha}{2}}\right)$$

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

Example 1 – Finding Sample Size

Suppose an experiment is being designed to assess the sample size needed for a 2x2 design that will be analyzed with the extended Welch test at a significance level of 0.05 and a power of 0.9. The null means are all 0. The alternative means are {48, 62, 66, 64}. The standard deviations are {3, 5, 4, 6}. Two types of group allocations will be used: equal (Eq) and proportional to σ (SD).

The following table shows the setting each of the parameters.

A	B	μ_1	σ	Eq	SD
1	1	48	3	1	3
1	2	62	5	1	5
2	1	66	4	1	4
2	2	64	6	1	6

The group allocation values are entered into the spreadsheet as follows.

Eq	SD
1	3
1	5
1	4
1	6

To better understand the pattern of means that we have chosen, press the **Sm** button to the right of the **μ_1 (H1 Group Means)** section. This will bring up the **Standard Deviation of Means Calculator** window. Enter the four means in the **Input** table at the top, two per row. When you are finished, the **Input** table should appear as follows:

	B1	B2
A1	48	62
A2	66	64

The **Output** table shows the **Effects** for each term. Note that the **Effects** of factor A are -5 and 5, those of factor B are -3 and 3, and those of the interaction are -4 and 4 (shown in the 2-by-2 table in the upper left of the table).

Notice that the **Standard Deviation of Means (Sm)** report at the bottom presents the standard deviations of each of the terms: $Sm(A) = 5$, $Sm(B) = 3$, and $Sm(AB) = 4$. Note that these values are equal to the absolute values of the **Effects** shown in the **Output** table. This will always be true for 2-by-2 tables.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters 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.

Design Tab

Solve For **Sample Size**
 Alternative Hypothesis **Two-Sided ($H_1: \delta \neq \delta_0$)**
 Power..... **0.90**
 Alpha..... **0.05**
 Group Allocation Input Type **Enter Columns of Allocation Patterns**
 Columns of Allocation Patterns **1-2**
 Model Terms..... **A, B, AB**
 μ_0 Input Type..... **μ_0 (H_0 Group Means)**
 $\mu_0(1,1)$ **0**
 $\mu_0(1,2)$ **0**
 $\mu_0(2,1)$ **0**
 $\mu_0(2,2)$ **0**
 μ_1 Input Type..... **Enter μ_1 (H_1 Group Means)**
 $\mu_1(1,1)$ **48**
 $\mu_1(1,2)$ **62**
 $\mu_1(2,1)$ **66**
 $\mu_1(2,2)$ **64**
 σ Input Type..... **σ (Group Standard Deviations)**
 $\sigma(1,1)$ **3**
 $\sigma(1,2)$ **5**
 $\sigma(2,1)$ **4**
 $\sigma(2,2)$ **6**

Input Spreadsheet Data

Row	Eq	SD
1	1	3
2	1	5
3	1	4
4	1	6

2x2 Factorial Analysis of Variance Allowing Unequal Variances

Output

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

Numeric Reports

Numeric Results

Solve For: [Sample Size](#)

Model: $Y(i, j, k) = \mu + A(i) + B(j) + AB(i, j) + e(i, j, k)$, $i = 1, 2$, $j = 1, 2$

Hypotheses: $H_0: \delta = \delta_0$ vs. $H_1: \delta \neq \delta_0$

Model Term	Power	Sample Size			Group Means		Differences among Means		Group Standard Deviations σ	Standard Error of δ 's $\sigma(\delta)$	Noncentrality Parameter Δ	Alpha
		Total N	Allocation r	Group n	H0 μ_0	H1 μ_1	H0 δ_0	H1 δ_1				
1: A	0.97150	16	Eq(1)	n(1)	$\mu_0(1)$	$\mu_1(1)$	0	10	$\sigma(1)$	2.318	4.313	0.05
2: B	0.90184	28	Eq(1)	n(2)	$\mu_0(1)$	$\mu_1(1)$	0	6	$\sigma(1)$	1.753	3.424	0.05
3: AB	0.94549	20	Eq(1)	n(3)	$\mu_0(1)$	$\mu_1(1)$	0	-8	$\sigma(1)$	2.074	-3.858	0.05
1: A	0.91419	12	SD(2)	n(4)	$\mu_0(1)$	$\mu_1(1)$	0	10	$\sigma(1)$	2.606	3.837	0.05
2: B	0.91081	27	SD(2)	n(5)	$\mu_0(1)$	$\mu_1(1)$	0	6	$\sigma(1)$	1.735	3.458	0.05
3: AB	0.93828	18	SD(2)	n(6)	$\mu_0(1)$	$\mu_1(1)$	0	-8	$\sigma(1)$	2.121	-3.771	0.05

Item	(1,1)	(1,2)	(2,1)	(2,2)
C_A*	-0.5	-0.5	0.5	0.5
C_B*	-0.5	0.5	-0.5	0.5
C_AB*	0.5	-0.5	-0.5	0.5
Eq(1)	1	1	1	1
SD(2)	3	5	4	6
n(1)	4	4	4	4
n(2)	7	7	7	7
n(3)	5	5	5	5
n(4)	2	3	3	4
n(5)	4	8	6	9
n(6)	3	5	4	6
$\mu_0(1)$	0	0	0	0
$\mu_1(1)$	48	62	66	64
$\sigma(1)$	3	5	4	6

* C_A is the contrast for testing term A. C_B is the contrast for testing term B. C_AB is the contrast for testing term AB.

Term	The number and name of the model term being reported on this report line. Each term is associated with a different array of contrast coefficients: C_A, C_B, or C_AB. Note that the term number is the number that is plotted along the horizontal axis, in the legend, or in the plot title.
Power	The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N	The total number of subjects in the study. N is the sum of the group sample sizes.
r	The name and number of the set containing the Group Allocation Pattern $\{r(1,1), r(1,2), r(2,1), r(2,2)\}$. The pattern values are rescaled to sum to one and thus become the Group Allocation Proportions.
n	The Group Sample Size is the name and number of the set containing the sample size of each group.
μ_0	The name and number of the set containing the group means under H0. Note that $\delta_0 = \mu_0'C$, where C is the appropriate array of contrast coefficients.
μ_1	The H1 Group Means is the name and number of the set containing the group means under H1. This is the set of means at which the power is calculated using $\delta_1 = \mu_1'C$.
δ_0	The Diff's Among Means H0 is the dot product of μ_0 and C, where C is the appropriate set of contrast coefficients (C = C_A, C_B, or C_AB). The dot product is the sum of the products of the corresponding entries of the two sets of numbers. Note that you must have $\delta_0 \neq \delta_1$. Also note that when $\delta_0 \neq 0$, this test is called a "non-zero null" test.
δ_1	The Diff's Among Means H1 is the dot product of μ_1 and C, where C is the appropriate set of contrast coefficients (C = C_A, C_B, or C_AB). The dot product is the sum of the products of the corresponding entries of the two sets of numbers. Thus, δ_1 is the difference between the means at the two levels of the term. Note that $(\delta_1)/2$ is sometimes called the "effect" of with this term. Also note that you must have $\delta_0 \neq \delta_1$.
σ	The name and number of the set containing the standard deviation of each group.
$\sigma(\delta)$	The standard error of the δ 's is used in the calculation of Δ . Note that $\sigma(\delta)^2 = \sum_i \{ \sum_j [c(i,j) \sigma(i,j)]^2 / n(i,j) \}$, where the $c(i,j)$ are the individual contrast coefficients of the term being reported on this row.

2x2 Factorial Analysis of Variance Allowing Unequal Variances

- Δ The NCP (noncentrality parameter) is used with the adjusted noncentral t-distribution to calculate the power. Note that $\Delta = (\delta_1 - \delta_0) / \sigma(\delta)$.
- Alpha The probability of rejecting a true null hypothesis.

Group Sample Size Details

n	N	Group Sample Sizes				Group Allocation Proportions			
		n(1,1)	n(1,2)	n(2,1)	n(2,2)	p(1,1)	p(1,2)	p(2,1)	p(2,2)
n(1)	16	4	4	4	4	0.250	0.250	0.250	0.250
n(2)	28	7	7	7	7	0.250	0.250	0.250	0.250
n(3)	20	5	5	5	5	0.250	0.250	0.250	0.250
n(4)	12	2	3	3	4	0.167	0.250	0.250	0.333
n(5)	27	4	8	6	9	0.148	0.296	0.222	0.333
n(6)	18	3	5	4	6	0.167	0.278	0.222	0.333

Summary Statements

A 2x2 factorial ANOVA design will be used to test whether the contrast of the means for the main-effect A (δ_A) is different from 0 ($H_0: \delta_A = 0$ versus $H_1: \delta_A \neq 0$, where the null value 0 is the value of the contrast applied to the null means '0, 0, 0, 0'). The comparison will be made using a two-sided Welch t-test using the contrast coefficients -0.5, -0.5, 0.5, 0.5, with a Type I error rate (α) of 0.05. The within-group standard deviations are assumed to be 3, 5, 4, 6. The noncentral t-distribution noncentrality parameter is 4.313. To detect group means of 48, 62, 66, 64 (or a contrast of means value of 10), with 90% power, group subject counts of 4, 4, 4, 4 (totaling 16 subjects) will be needed.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout- Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	16	20	4
20%	28	35	7
20%	20	25	5
20%	12	15	3
20%	27	34	7
20%	18	23	5

- Dropout Rate The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
- N The evaluable sample size at which power is computed. If N subjects are evaluated out of the N' subjects that are enrolled in the study, the design will achieve the stated power.
- N' The total number of subjects that should be enrolled in the study in order to obtain N evaluable subjects, based on the assumed dropout rate. After solving for N, N' is calculated by inflating N using the formula $N' = N / (1 - DR)$, with N' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. (2018) pages 32-33.)
- D The expected number of dropouts. $D = N' - N$.

Dropout Summary Statements

Anticipating a 20% dropout rate, 20 subjects should be enrolled to obtain a final sample size of 16 subjects.

2x2 Factorial Analysis of Variance Allowing Unequal Variances

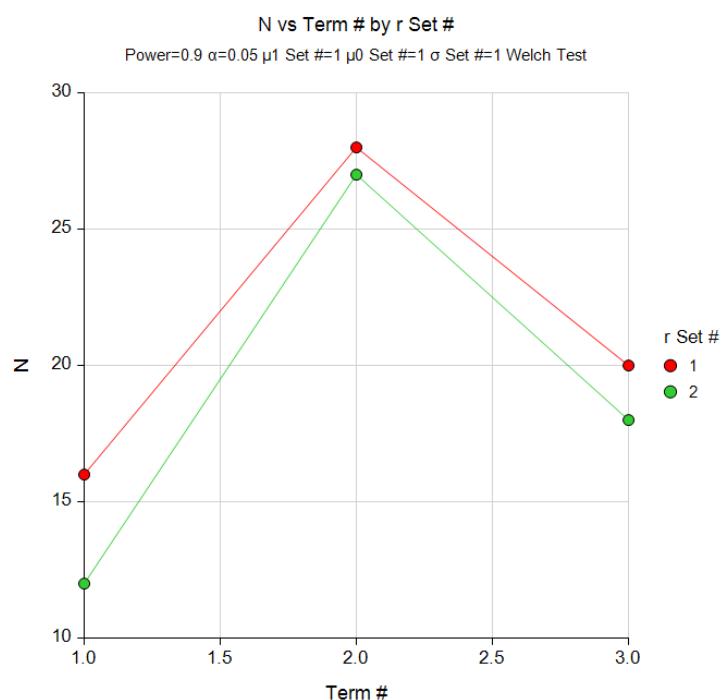
References

- Jan, S.L. and Shieh, G. 2016. 'A systematic approach to designing statistically powerful heteroscedastic 2 x 2 factorial studies while minimizing financial costs.' BMC Medical Research Methodology, 16:114.
- Kirk, Roger E. 2013. Experimental Design: Procedures for the Behavioral Sciences, 4th Edition. Sage. Washington, D.C.
- Luh, W.M. and Guo, J.H. 2016. 'Allocating sample sizes to reduce budget for fixed-effect 2 x 2 heterogeneous analysis of variance.' Journal of Experimental Education, 84:197-211.
- Satterthwaite, F.E. 1946. 'An approximate distribution of estimate of variance components,' Biometric Bulletin, 2:110-114.
- Shieh, G. and Jan, S-L. 2015. 'Power and sample size calculations for testing linear combinations of group means under variance heterogeneity with applications to meta and moderation analysis'. Psicologica, 36:367-390.
- Welch, B.L. 1951. 'On the Comparison of Several Mean Values: An Alternative Approach'. Biometrika, 38, 330-336.

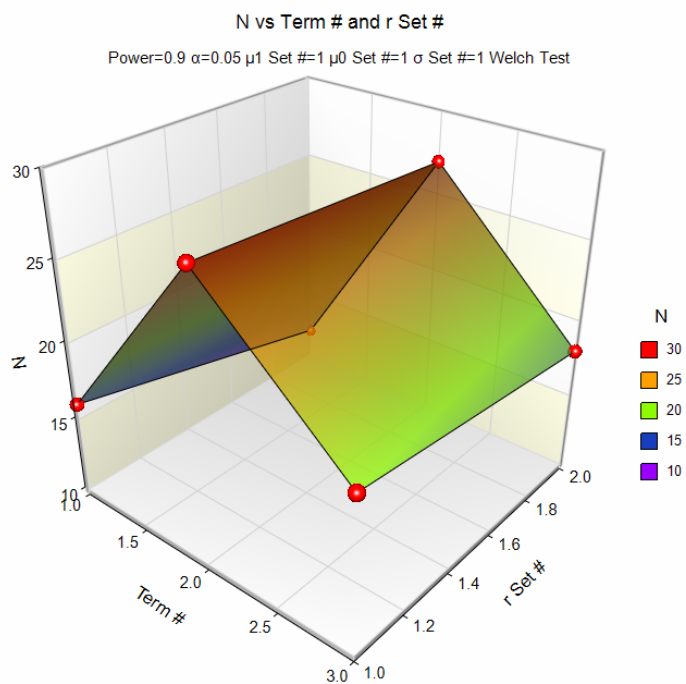
This report shows the numeric results of this study along with the report definitions.

Plots Section

Plots



2x2 Factorial Analysis of Variance Allowing Unequal Variances



These plots give a visual presentation of the results in the Numeric Report.

Note that the three values of the Term # axis are 1, 2, and 3. These values correspond to the terms A, B, and AB.

Example 2 – Validation using Jan and Shieh (2016)

Jan and Shieh (2016) page 6, Table 1, presents an example in which $\alpha = 0.05$, the sample sizes are {16, 14, 7, 15}, the standard deviations are {0.83, 0.72, 0.34, 0.77}, the null means are {0, 0, 0, 0}, and the alternative means are {1.23, 0.42, 0.13, 0.38}. The resulting power for testing the interaction is given as 0.8038.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For	Power
Alternative Hypothesis	Two-Sided ($H_1: \delta \neq \delta_0$)
Alpha.....	0.05
Group Allocation Input Type	Enter n (Group Sample Sizes)
n(1,1)	16
n(1,2)	14
n(2,1)	7
n(2,2)	15
Model Terms.....	AB
μ_0 Input Type.....	μ_0 (H0 Group Means)
$\mu_0(1,1)$	0
$\mu_0(1,2)$	0
$\mu_0(2,1)$	0
$\mu_0(2,2)$	0
μ_1 Input Type.....	Enter μ_1 (H1 Group Means)
$\mu_1(1,1)$	1.23
$\mu_1(1,2)$	0.42
$\mu_1(2,1)$	0.13
$\mu_1(2,2)$	0.38
σ Input Type.....	σ (Group Standard Deviations)
$\sigma(1,1)$	0.83
$\sigma(1,2)$	0.72
$\sigma(2,1)$	0.34
$\sigma(2,2)$	0.77

2x2 Factorial Analysis of Variance Allowing Unequal Variances

Output

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

Numeric Results

Solve For: [Power](#)

Model: $Y(i, j, k) = \mu + A(i) + B(j) + AB(i, j) + e(i, j, k)$, $i = 1, 2$, $j = 1, 2$

Hypotheses: $H_0: \delta = \delta_0$ vs. $H_1: \delta \neq \delta_0$

Model Term	Power	Sample Size		Group Means		Differences among Means		Group Standard Deviations σ	Standard Error of δ 's $\sigma(\delta)$	Noncentrality Parameter Δ	Alpha
		Total N	Group n	H0 μ_0	H1 μ_1	H0 δ_0	H1 δ_1				
1: AB	0.80376	52	n(1)	$\mu_0(1)$	$\mu_1(1)$	0	0.53	$\sigma(1)$	0.184	2.873	0.05

Item	(1,1)	(1,2)	(2,1)	(2,2)
C_AB*	0.5	-0.5	-0.5	0.5
n(1)	16	14	7	15
$\mu_0(1)$	0	0	0	0
$\mu_1(1)$	1.23	0.42	0.13	0.38
$\sigma(1)$	0.83	0.72	0.34	0.77

* C_AB is the contrast for testing term AB.

Group Sample Size Details

n	N	Group Sample Sizes				Group Allocation Proportions			
		n(1,1)	n(1,2)	n(2,1)	n(2,2)	p(1,1)	p(1,2)	p(2,1)	p(2,2)
n(1)	52	16	14	7	15	0.308	0.269	0.135	0.288

PASS found the power to be 0.80376 which rounds to 0.8038. The procedure is validated.