

Chapter 563

Factorial Analysis of Variance using Effect Size

Introduction

This routine calculates power or sample size for F tests from a multi-factor analysis of variance design using only Cohen's (1988) effect sizes as input.

A common task in research is to compare the average response across levels of one or more factor variables. Examples of factor variables are income level of two regions, nitrogen content of three lakes, or drug dosage. The factorial analysis of variance compares the means of two or more factors. *F* tests are used to determine statistical significance of the factors and their interactions. The tests are non-directional in that the null hypothesis specifies that all means are equal, and the alternative hypothesis simply states that at least one mean is different. This **PASS** module performs power analysis and sample size estimation for an analysis of variance design with up to three fixed factors.

In the following example, the responses of a weight loss experiment are arranged in a two-factor, fixed-effect, design. The first factor is diet (D1 and D2) and the second factor is dose level of a dietary drug (low, medium, and high). The twelve individuals available for this study were assigned at random to one of the six treatment groups (cells) so that there were two per group. The response variable was an individual's weight loss after four months.

Table of Individual Weight Losses			
	Dietary Drug Dose Level		
Diet	Low	Medium	High
D1	14, 16	15, 18	23, 28
D2	18, 21	18, 22	38, 39

Important features to note are that each table entry represents a different individual, and that the response variable (weight loss) is continuous, while the factors (Diet and Dose) are discrete.

Means can be calculated for each cell of the table. These means are shown in the table below. Note that we have added an additional row and column for the row, column, and overall means. The six means in the interior of this table are called the *cell means*.

Table of Means				
	Dietary Drug Dose Level			
Diet	Low	Medium	High	Total
D1	15.00	16.50	25.50	19.00
D2	19.50	20.00	38.50	26.00
Total	17.25	18.25	32.00	22.50

The Linear Model

A mathematical model may be formulated that underlies this experimental design. This model expresses each cell mean, μ_{ij} , as the sum of parameters called *effects*. A common linear model for a two-factor experiment is

$$\mu_{ij} = m + a_i + b_j + (ab)_{ij}$$

where $i = 1, 2, \dots, I$ and $j = 1, 2, \dots, J$. This model expresses the value of a cell mean as the sum of four components:

- m the grand mean.
- a_i the effect of the i^{th} level of factor A. Note that $\sum a_i = 0$.
- b_j the effect of the j^{th} level of factor B. Note that $\sum b_j = 0$.
- ab_{ij} the combined effect of the i^{th} level of factor A and the j^{th} level of factor B. Note that $\sum (ab)_{ij} = 0$.

Another way of stating this model for the two-factor case is

$$\text{Cell Mean} = \text{Overall Effect} + \text{Row Effect} + \text{Column Effect} + \text{Interaction Effect.}$$

Since this model is the sum of various constants, it is called a *linear model*.

Calculating the Effects

We will now calculate the effects for our example. We will let Drug Dose correspond to factor A and Diet correspond to factor B.

Step 1 - Remove the Grand Mean

Remove the grand mean from the table of means by subtracting 22.50 from each entry. The values in the margins are the *effects* of the corresponding factors.

Table of Mean Weight Losses After Subtracting the Grand Mean				
	Dietary Drug Dose Level			
Diet	Low	Medium	High	Overall
D1	-7.50	-6.00	3.00	-3.50
D2	-3.00	-2.50	16.00	3.50
Overall	-5.25	-4.25	9.50	22.50

Step 2 – Remove the Effects of Factor B (Diet)

Subtract the Diet effects (-3.50 and 3.50) from the entries in those rows.

Table of Mean Weight Losses After Subtracting the Diet Effects				
	Dietary Drug Dose Level			
Diet	Low	Medium	High	Overall
D1	-4.00	-2.50	6.50	-3.50
D2	-6.50	-6.00	12.50	3.50
Overall	-5.25	-4.25	9.50	22.50

Step 3 – Remove the Effects of Factor A (Drug Dose)

Subtract the Drug Dose effects (-5.25, -4.25, and 9.50) from the rest of the entries in those columns. This will result in a table of effects.

Table of Effects				
	Dietary Drug Dose Level			
Diet	Low	Medium	High	Overall
D1	1.25	1.75	-3.00	-3.50
D2	-1.25	-1.75	3.00	3.50
Overall	-5.25	-4.25	9.50	22.50

We have calculated a table of effects for the two-way linear model. Each cell mean can be calculated by summing the appropriate entries from this table.

The estimated linear effects are:

$$m = 22.50$$

$$a_1 = -5.25 \quad a_2 = -4.25 \quad a_3 = 9.50$$

$$b_1 = -3.50 \quad b_2 = 3.50$$

$$ab_{11} = 1.25 \quad ab_{21} = 1.75 \quad ab_{31} = -3.00$$

$$ab_{12} = -1.25 \quad ab_{22} = -1.75 \quad ab_{32} = 3.00.$$

The six cell means are calculated from these effects as follows:

$$15.00 = 22.50 - 5.25 - 3.50 + 1.25$$

$$19.50 = 22.50 - 5.25 + 3.50 - 1.25$$

$$16.50 = 22.50 - 4.25 - 3.50 + 1.75$$

$$20.00 = 22.50 - 4.25 + 3.50 - 1.75$$

$$25.50 = 22.50 + 9.50 - 3.50 - 3.00$$

$$38.50 = 22.50 + 9.50 + 3.50 + 3.00$$

Analysis of Variance Hypotheses

The hypotheses that are tested in an analysis of variance table concern the effects, so in order to conduct a power analysis you must have a firm grasp of their meaning. For example, we would usually test the following hypotheses:

1. Are there differences in weight loss among the three drug doses? That is, are the drug dose effects all zero? This hypothesis is tested by the F test for factor A , which tests whether the standard deviation of the a_i is zero.
2. Is there a difference in weight loss between the two diets? That is, are the diet effects all zero? This hypothesis is tested by the F test for factor B , which tests whether the standard deviation of the b_j is zero.
3. Are there any diet-dose combinations that exhibit a weight loss that cannot be explained by diet and/or drug dose singly? This hypothesis is tested by the F test for the AB interaction, which tests whether the standard deviation of the $(ab)_{ij}$ is zero.

Each of these hypotheses can be tested at a different alpha level and different precision. Hence each can have a different power. One of the tasks in planning such an experiment is to determine a sample size that yields necessary power values for each of these hypothesis tests. This is accomplished using this program module.

Definition of Terms

Factorial designs evaluate the effect of two or more categorical variables (called *factors*) on a response variable by testing hypotheses about various averages. These designs are popular because they allow experimentation across a wide variety of conditions and because they evaluate the *interaction* of two or more factors. Interaction is the effect that may be attributed to a combination of two or more factors, but not to one factor singly.

A *factor* is a variable that relates to the response. Either the factor is discrete by nature (as in location or gender) or has been made discrete by collapsing a continuous variable (as in income level or age group). The term *factorial* implies that all possible combinations of the factors being studied are included in the design.

A *fixed* factor is one in which all possible *levels* (categories) are considered. Examples of fixed factors are gender, dose level, and country of origin. They are different from *random* factors which represent a random selection of individuals from the population described by the factor. Examples of random factors are people living within a region, a sample of schools in a state, or a selection of labs. Again, a fixed factor includes the range of interest while a random factor includes only a sample of all possible levels.

A factorial design is analyzed using the analysis of variance. When only fixed factors are used in the design, the analysis is said to be a *fixed-effects analysis of variance*.

Suppose a group of individuals have agreed to be in a study involving six treatments. In a *completely randomized factorial design*, each individual is assigned at random to one of the six groups and then the treatments are applied. In some situations, the randomization occurs by randomly selecting individuals from the populations defined by the treatment groups. The designs analyzed by this module are completely randomized factorial designs.

Power Calculations

The calculation of the power of a particular test proceeds as follows

1. Determine the critical value, $F_{df1,df2,\alpha}$ where $df1$ is the numerator degrees of freedom, $df2$ is the denominator degrees of freedom, and α is the probability of a type-I error (significance level). Note that the F test is a two-tailed test as no logical direction is assigned in the alternative hypothesis.
2. Calculate the standard deviation of the hypothesized effects, using the formula:

$$\sigma_m = \sqrt{\frac{\sum_{i=1}^k (e_i - \bar{e})^2}{k}}$$

where the e_i are effect values and k is the number of effects. Note that the average effect will be zero by construction, so this formula reduces to

$$\sigma_m = \sqrt{\frac{\sum_{i=1}^k (e_i)^2}{k}}$$

3. Compute the noncentrality parameter λ using the relationship:

$$\lambda = N \frac{\sigma_m^2}{\sigma^2} = Nf^2$$

where N is the total number of subjects.

4. Compute the power as the probability of being greater than $F_{df1,df2,\alpha}$ on a noncentral-F distribution with noncentrality parameter λ .

Effect Size

The effect size, f , which is defined as

$$f = \frac{\sigma_m}{\sigma}$$

where σ is the standard deviation of values within a cell. Cohen (1988) has designated values of f less than 0.1 as *small*, values around 0.25 to be *medium*, and values over 0.4 to be *large*. That is, a σ_m that is 10% of the underlying σ is *small*, 25% is *medium*, and 40% is *large*.

When you are analyzing the power of an existing analysis of variance table, you can compute the values of σ_m for each term from its mean square or F ratio using the following formulas:

$$\sigma_m = \sqrt{\frac{df_{\text{numerator}} MS_{\text{numerator}}}{N}} \quad \text{or} \quad \sigma_m = \sqrt{\frac{df_{\text{numerator}}(F)(MSE)}{N}}$$

where N is the total number of observations, MSE is the mean square error, df is the numerator degrees of freedom, MS is the mean square of the term, and F is the F ratio of the term.

Factorial Analysis of Variance using Effect Size

Similarly, you can calculate the value of σ using

$$\sigma = \sqrt{MSE}$$

And finally, you can compute f using the ratio of these quantities.

Note on Interactions

The general formula for the calculation of the standard deviation of the interaction effects is

$$\sigma_m = \sqrt{\frac{\sum_{i=1}^k (e_i)^2}{k}}$$

where k is the number of effects. In the case of a two-way interaction, the standard deviation is calculated using the formula:

$$\sigma_m(AB) = \sqrt{\frac{\sum_{i=1}^I \sum_{j=1}^J (\mu_{ij} - \mu_{i.} - \mu_{.j} + \bar{\mu})^2}{IJ}}$$

where i is the factor A index (from 1 to I), j is the factor B index (from 1 to J), μ_{ij} is the mean in the ij^{th} cell, $\mu_{i.}$ is the i^{th} mean of factor A across all levels of other factors, $\mu_{.j}$ is the j^{th} mean when factor B across all levels of other factors, and $\bar{\mu}$ is the overall mean of the means.

To see how this works, consider the following table of means from an experiment with $I = 2$ and $J = 3$:

		<i>i</i>		
		1	2	
	1	2.0	4.0	3.0
<i>j</i>	2	4.0	6.0	5.0
	3	6.0	11.0	8.5
	---	---	---	---
	Total	4.0	7.0	5.5

Now, if we subtract the factor A means, subtract the factor B means, and add the overall mean, we get the interaction effects:

0.5	-0.5
0.5	-0.5
-1.0	1.0

Next, we sum the squares of these six values:

$$(0.5)^2 + (-0.5)^2 + (0.5)^2 + (-0.5)^2 + (-1.0)^2 + (1.0)^2 = 3$$

Factorial Analysis of Variance using Effect Size

Next, we divide this value by the number of cells which is $(2)(3) = 6$:

$$3/6 = 0.5$$

Finally, we take the square root of this value:

$$\sqrt{0.5} = 0.7071$$

Hence, for this configuration of means,

$$\sigma_m(AB) = 0.7071.$$

Notice that the average of the absolute values of the interaction effects is:

$$[0.5 + 0.5 + 0.5 + 0.5 + 1.0 + 1.0]/6 = 0.6667$$

We see that $SD(\text{interaction})$ is close to the average absolute interaction effect. That is, 0.7071 is close to 0.6667. This will usually be the case. Hence, one way to interpret the interaction-effects standard deviation is as a number a little larger than the average absolute interaction effect.

Example 1 – Power of an Interaction after a Study

This example will explain how to calculate the power of F tests from data that have already been collected and analyzed.

Analyze the power of the experiment that was given at the beginning of this chapter. These data were analyzed using the analysis of variance procedure in **NCSS** and the following results were obtained.

Analysis of Variance Table

Source Term	DF	Sum of Squares	Mean Square	F-Ratio	P-Value
A: Dose	2	543.5	271.75	50.95	0.000172
B: Diet	1	147	147	27.56	0.001920
AB	2	54.5	27.25	5.11	0.050629
S	6	32	5.333333		
Total (Adjusted)	11	777			
Total	12				

Suppose we want the power of the test of the AB interaction using the f of from the data. Using the formula for $\sigma_m(AB)$, the value of f is calculated as

$$f = \sqrt{\frac{2(27.25)/12}{5.333333}} = 0.922801$$

We will set the f values of the factors to *large* values.

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	Power
Alpha for All Terms	0.05
Number of Factors	2
A - Levels	3
A - f	0.4
B - Levels	2
B - f	0.4
A*B	Checked
A*B - f	0.922801
n (Size Per Group)	2

Output

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

Numeric Results

Solve For: **Power**

Term	Power	Sample Size		Degrees of Freedom		Effect Size f	Alpha
		Average Group n	Total N	df1	df2		
A(3)	0.1499	2	12	2	6	0.400	0.05
B(2)	0.2162	2	12	1	6	0.400	0.05
A*B	0.5889	2	12	2	6	0.923	0.05

Power	The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
n	The average sample size of per group (treatment combination).
Total N	The total sample size of all groups combined.
df1	The numerator degrees of freedom.
df2	The denominator degrees of freedom.
σ_m	The standard deviation of the group means or effects.
σ	The subject-to-subject standard deviation.
f	Effect Size. $f = \sigma_m / \sigma$.
Alpha	The probability of rejecting a true null hypothesis.

Summary Statements

A (fixed factor) factorial design with two factors (one with 3 levels and the other with 2 levels, resulting in 6 treatment combinations) will be used to test whether there are differences among the levels of the factors. Each term will be tested using a linear model analysis of variance F-test with a Type I error rate (α) of 0.05. With a sample size of 2 subjects per treatment combination, the total number of subjects is 12. For factor A, to detect an effect size of 0.4, the power is 0.1499. For factor B, to detect an effect size of 0.4, the power is 0.2162. For the A*B interaction effect, to detect an effect size of 0.923, the power is 0.5889.

Dropout-Inflated Sample Size

Group	Dropout Rate	Sample Size n	Dropout-Inflated Enrollment Sample Size n'	Expected Number of Dropouts D
1 - 6	20%	2	3	1
Total		12	18	6

Group	Lists the group numbers.
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 for each group at which power is computed (as entered by the user). 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 number of subjects that should be enrolled in each group in order to obtain n evaluable subjects, based on the assumed dropout rate. 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 in each group. $D = n' - n$.

Factorial Analysis of Variance using Effect Size

Dropout Summary Statement

Anticipating a 20% dropout rate, group sizes of 3, 3, 3, 3, 3, and 3 subjects should be enrolled to obtain final group sample sizes of 2, 2, 2, 2, 2, and 2 subjects.

References

- Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates. Hillsdale, New Jersey
- Neter, J., Kutner, M., Nachtsheim, C., and Wasserman, W. 1996. *Applied Linear Statistical Models*. Richard D. Irwin, Inc. Chicago, Illinois.
- Winer, B.J. 1991. *Statistical Principles in Experimental Design*. Third Edition. McGraw-Hill. New York, NY.
-

This report shows the power for each of the three terms. It is apparent that the cell sample size of 2 does not achieve a reasonable power for any of the tests.

Term

This is the term (main effect or interaction) from the analysis of variance model being displayed on this line.

Power

This is the power of the F test for this term. Note that since adding and removing terms changes the denominator degrees of freedom (df_2), the power depends on which other terms are included in the model.

n

This is the sample size per group (treatment combination). Fractional values indicate an unequal allocation among the cells.

N

This is the total sample size for the complete design.

df1

This is the numerator degrees of freedom of the F test.

df2

This is the denominator degrees of freedom of the F test. This value depends on which terms are included in the AOV model.

F (Effect Size = σ_m/σ)

This is the standard deviation of the means divided by the standard deviation of subjects. It provides an index of the magnitude of the difference among the means that can be detected by this design.

Alpha

This is the significance level of the F test. This is the probability of a type-I error given the null hypothesis of equal means and zero effects.

Example 2 – Finding the Sample Size

In this example, we will determine the sample size necessary to achieve a power of 0.80 for all three F-tests in Example 1 to detect a medium value of f .

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 **Sample Size**
 Based on Term **All**
 Minimum Power **0.80**
 Alpha for All Terms **0.05**
 Number of Factors **2**
 A - Levels **3**
 A - f **0.4**
 B - Levels **2**
 B - f **0.4**
 A*B **Checked**
 A*B - f **0.4**

Output

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

Numeric Results

Solve For: [Sample Size](#)

Term	Power	Sample Size		Degrees of Freedom		Effect Size f	Alpha
		Average Group n	Total N	df1	df2		
A(3)	0.8171	11	66	2	60	0.4	0.05
B(2)	0.8920	11	66	1	60	0.4	0.05
A*B	0.8171	11	66	2	60	0.4	0.05

The necessary sample size is 11 per group for a total of 66 subjects.

Example 3 – Validation using Winer (1991)

Example 4 in the *Factorial Analysis of Variance* procedure in **PASS** presents a validation run based on Winer (1991) pages 428-429 for a two-way design in which factor *A* has two levels and factor *B* has three levels. In this example, $f_A = 0.2404$, $f_B = 0.4377$, and $f_{AB} = 0.8923$. The corresponding power values are 0.1558, 0.2918, and 0.8534. The cell sample size is 3 and all alpha levels are 0.05.

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 3** 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
Alpha for All Terms	0.05
Number of Factors	2
A - Levels	2
A - f	0.2404
B - Levels	3
B - f	0.4377
A*B	Checked
A*B - f	0.8923
n (Size Per Group)	3

Output

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

Numeric Results							
Solve For: Power							
Term	Power	Sample Size		Degrees of Freedom		Effect Size f	Alpha
		Average Group n	Total N	df1	df2		
A(2)	0.1558	3	18	1	12	0.240	0.05
B(3)	0.2918	3	18	2	12	0.438	0.05
A*B	0.8534	3	18	2	12	0.892	0.05

The power values match the results in Winer (1991), so the procedure is validated.