Both "g" Power
calculations are
required.

GPower 3.1.3

Central and noncentral distributions | Protocol of power analyses

Test family
F tests ▼

Statistical test
ANOVA: Fixed effects, omnibus, one-way ▼

Type of power analysis
A priori: Compute required sample size – given α, power, and effect size ▼

Input Parameters

Determine =>	Effect size f	0.25
	α err prob	0.05
(Power (1–β err prob))	0.95
	Number of groups	5

Output Parameters

Noncentrality parameter λ	?
Critical F	?
Numerator df	?
Denominator df	?
Total sample size	?
Actual power	?

X–Y plot for a range of values | Calculate

$1-\beta > 0.95$

Select procedure
Effect size from means ▶

Number of groups   5

SD σ within each group   1

Group	Mean	Size
1	0	5
2	0	5
3	0	5
4	0	5
5	0	5

Equal n   5

Total sample size   20

Effect size f   ?

Calculate | Calculate and transfer to main window | Close

ONE-WAY ANOVA POWER ANALYSIS | G*POWER DATA ANALYSIS EXAMPLES

NOTE: This page was developed using G*Power version 3.0.10. You can download the current version of G*Power from http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/ (http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/) . You can also find help files, the manual and the user guide on this website.

Introduction

Power analysis is the name given to the process for determining the sample size for a research study. The technical definition of power is that it is the probability of detecting a "true" effect when it exists. Many students think that there is a simple formula for determining sample size for every research situation. However, the reality it that there are many research situations that are so complex that they almost defy rational power analysis. In most cases, power analysis involves a number of simplifying assumptions, in order to make the problem tractable, and running the analyses numerous times with different variations to cover all of the contingencies.

In this unit we will try to illustrate the power analysis process using a simple four group design.

Description of the experiment

We wish to conduct a study in the area of mathematics education involving different teaching methods to improve standardized math scores in local classrooms. The study will include four different teaching methods and use fourth grade students who are randomly sampled from a large urban school district and are then random assigned to the four different teaching methods.

Here are the four different teaching methods which will be examined: 1) The traditional teaching method where the classroom teacher explains the concepts and assigns homework problems from the textbook; 2) the intensive practice method, in which students fill out additional work sheets both before and after school; 3) the computer assisted method, in which students learn math concepts and skills from using various computer based math learning programs; and, 4) the peer assistance learning method, which pairs each fourth grader with a fifth grader who helps them learn the concepts followed by the student teaching the same material to another student in their group.

Students will stay in their math learning groups for an entire academic year. At the end of the Spring semester all students will take the Multiple Math Proficiency Inventory (MMPI). This standardized test has a mean for fourth graders of 550 with a standard deviation of 80.

The experiment is designed so that each of the four groups will have the same sample size. One of the important questions we need to answer in designing the study is, how many students will be needed in each group?

The power analysis

In order to answer this question, we will need to make some assumptions and some educated guesses about the data. First, we will assume that the standard deviation for each of the four groups will be equal and will be equal to the national value of 80. Further, because of prior research, we expect that the traditional teaching group (Group 1) will have the lowest mean score and that the peer assistance group (Group 4) will have the highest mean score on the MMPI. In fact, we expect that Group 1 will have a mean of 550 and that Group 4 will have mean that is greater by 1.2 standard deviations, i.e., the mean will equal at least 646. For the sake of simplicity, we will assume that the means of the other two groups will be equal to the grand mean.

To begin, the program should be set to the F family of tests, to a one-way ANOVA, and to the 'A Priori' power analysis necessary to identify sample size. From there we need the following information: the alpha level, the power, the number of groups and the effect size.

Test family	Statistical test	
F tests	ANOVA: Fixed effects, omnibus, one-way	

Type of power analysis:
A priori: Compute required sample size – given α, power, and effect size

| Input Parameters			Output Parameters	
Determine =>	Effect size f		Noncentrality parameter λ	?
	α err prob	.05	Critical F	?
	Power (1-β err prob)	0.8	Numerator df	?
	Number of groups	4	Denominator df	?
			Total sample size	?
			Actual power	?

X-Y plot for a range of values | Calculate

The latter can be determined via the 'Determine' button, which calls up a menu requesting the number of groups, their shared standard deviation, and the mean of each group. All of our known variables can now be inputted. As stated above, there are four groups, a=4. We will set alpha = 0.05. We already have the mean = 550 for the lowest group and the mean = 646 for the highest group. We will first set the means for the two middle groups to be the grand mean. Based on this setup and the assumption that the common standard deviation is equal to 80, we can do some simply calculation to see that the grand mean will be 598.

Let's set the power to be .8 and calculate the corresponding sample size. A click of 'Calculate and transfer to main window', followed by the main window's 'Calculate' button produces the following result.
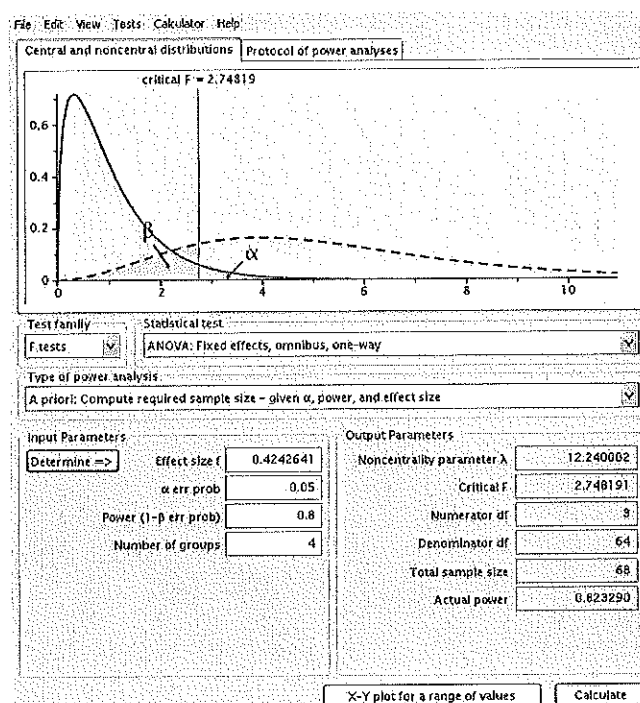
A total of 68 students will be required for the test; 17 for each class. Now, if we want to see how sample size affects power, we can click 'X-Y plot for a range of values', provide a range of sample sizes, and follow a graph with power as the dependent variable. Simply set power as a function of sample size with an appropriate set of sizes, here 40 students through 200 in steps of 10.
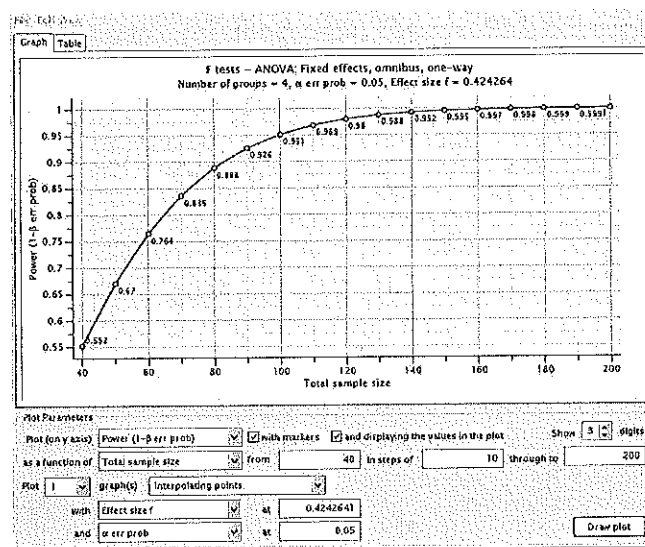
So we see that when we have 100 subjects (25 in each group), we will have power of .951.

In the setup above, we have arranged so that the two middle groups will have means equal to the grand mean. In general, the means for the two middle groups can be anything in between the extreme values. If you have a good idea on what these means should be, you might want to make use of this piece of information in your power analysis. Let's say, for instance, that the means for the two middle groups should be 575 and 635. We will compute the power for a sequence of sample sizes as we did earlier.


Inputting the new effect size into the plot, we get:
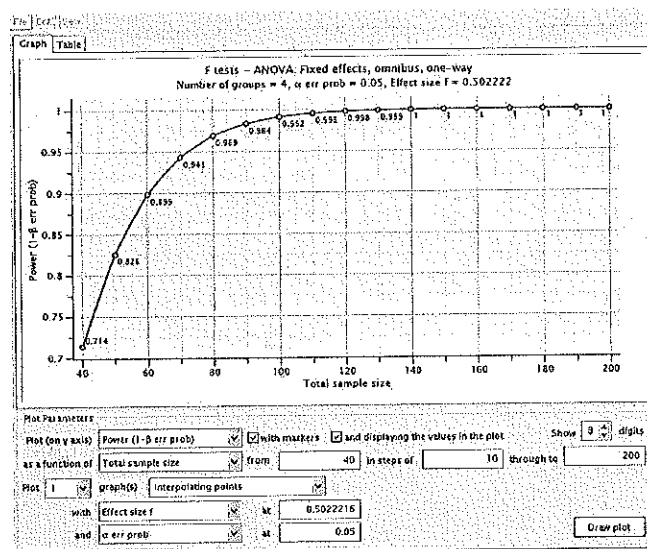
So we see that to produce a power of .8 we need fewer subjects than in the earlier case when the two middle groups have the grand mean as their means. This should be expected since the power here is the overall power of the F test for ANOVA, and since the means are more polarized towards the two extreme ends, it is easier to detect the group effect.

Effect size

The difference of the means between the lowest group and the highest group over the common standard deviation is a measure of effect size. In the calculation above, we have used 550 and 646 with common standard deviation of 80. This gives effect size of $(646-550)/80 = 1.2$. This is considered to be a large effect size. Let's say now we have a medium effect size of .75. What does this translate into in terms of groups means? Well, we can always use 550 for the lowest group. The mean for the highest group will be $.75*80 + 550 = 610$. Let's assume the two middle groups have the means of grand mean, say g. Then we have $(550 + g + g + 610) / 4 = g$. This gives us $g = (550 + 610)/2 = 580$. Let's now redo our sample size calculation with this set of means.
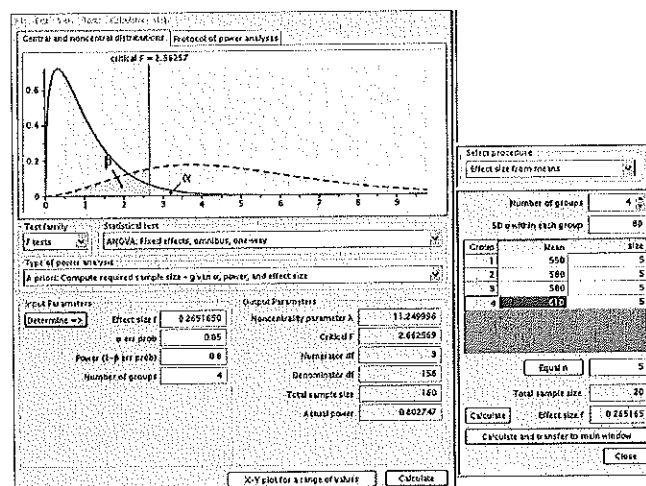
So we see that at a power of .8, we have a sample size of 160, or 40 for each group.

What about a small effect size; say, .25? We can do the same calculation as we did previously. The mean for each of the groups will be 550 , 560, 560 and 570.
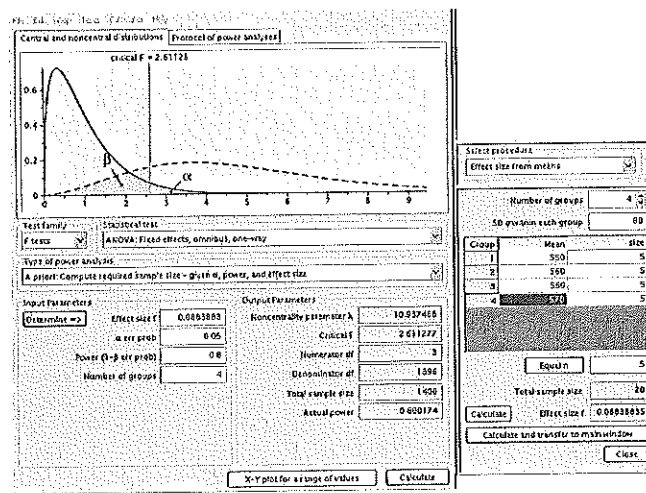
Now the sample size goes way up.

Discussion

The sample size calculation is based a number of assumptions. One of these is the normality assumption for each group. We also assume that the groups have the same common variance. As our power analysis calculation is rooted in these assumptions it is important to remain aware of them.

We have also assumed that we have knowledge of the magnitude of effect we are going to detect which is described in terms of group means. When we are unsure about the groups means, we should use more conservative estimates. For example, we might not have a good idea on the two means for the two middle groups, then setting them to be the grand mean is more conservative than setting them to be something arbitrary.

Here are the sample sizes per group that we have come up with in our power analysis: 17 (best case scenario), 40 (medium effect size), and 350 (almost the worst case scenario). Even though we expect a large effect, we will shoot for a sample size of between 40 and 50. This will help ensure that we have enough power in case some of the assumptions mentioned above are not met or in case we have some incomplete cases (i.e., missing data).

Click here to report an error on this page or leave a comment

How to cite this page (https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-how-do-i-cite-web-pages-and-programs-from-the-ucla-statistical-consulting-group/)