

Power and Sample Size Calculation

By Gayla Olbricht and Yong Wang

Definition and Application

Statistical power is defined as the probability of rejecting the null hypothesis while the alternative hypothesis is true. Factors that affect statistical power include the sample size, the specification of the parameter(s) in the null and alternative hypothesis, i.e. how far they are from each other, the precision or uncertainty the researcher allows for the study (generally the confidence or significance level) and the distribution of the parameter to be estimated. For example, if a researcher knows that the statistics in the study follow a Z or standard normal distribution, there are two parameters that he/she needs to estimate, the population mean (μ) and the population variance (σ^2). Most of the time, the researcher know one of the parameters and need to estimate the other. If that is not the case, some other distribution may be used, for example, if the researcher does not know the population variance, he/she can estimate it using the sample variance and that ends up with using a T distribution.

In research, statistical power is generally calculated for two purposes.

1. It can be calculated before data collection based on information from previous research to decide the sample size needed for the study.
2. It can also be calculated after data analysis. It usually happens when the result turns out to be non-significant. In this case, statistical power is calculated to verify whether the non-significant result is due to really no relation in the sample or due to a lack of statistical power.

Statistical power is positively correlated with the sample size, which means that given the level of the other factors, a larger sample size gives greater power. However, researchers are also faced with the decision to make a difference between statistical difference and scientific difference. Although a larger sample size enables researchers to find smaller difference statistically significant, that difference may not be large enough be scientifically meaningful. Therefore, as consultants, we would like to recommend that our clients have an idea of what they would expect to be a scientifically meaningful difference before doing a power analysis to determine the actual sample size needed.

Calculation of Statistical Power

The power is a probability and it is defined to be the probability of rejecting the null hypothesis when the alternative hypothesis is true. After plugging in the required information, a researcher can get a function that

describes the relationship between statistical power and sample size and the researcher can decide which power level they prefer with the associated sample size. The choice of sample size may also be constrained by factors such as the financial budget the researcher is faced with. But generally consultants would like to recommend that the minimum power level is set to be 0.80.

In some occasions, calculation of power is simple and can be done by hand. Statistical software packages such as SAS also offers a way of calculating power and sample size.

The researchers must have some information before they can do the power and sample size calculation. The information includes previous knowledge about the parameters (their means and variances) and what confidence or significance level is needed in the study.

Hand Calculation.

We will use an example to illustrate how a researcher can calculate the sample size needed for a study. Given that a researcher has the null hypothesis that $\mu = \mu_0$ and alternative hypothesis that $\mu = \mu_1 \neq \mu_0$, and that the population variance is known as σ^2 . Also, he knows that he wants to reject the null hypothesis at a significance level of α which gives a corresponding Z score, called it $Z_{\alpha/2}$. Therefore, the power function will be

$$P\{Z > Z_{\alpha/2} \text{ or } Z < -Z_{\alpha/2} | \mu_1\} = 1 - \Phi[Z_{\alpha/2} - (\mu_1 - \mu_0)/(\sigma/\sqrt{n})] + \Phi[-Z_{\alpha/2} - (\mu_1 - \mu_0)/(\sigma/\sqrt{n})].$$

That is a function of the power and sample size given other information known and the researcher can get the corresponding sample size for each power level.

For example, if the researcher learns from literature that the population follows a normal distribution with mean of 100 and variance of 100 under the null hypothesis and he/she expects the mean to be greater than 105 or less than 95 under the null hypothesis and he/she wants the test to be significant at 95% level, the resulting power function would be:

Power = $1 - \Phi[1.96 - (105 - 100)/(\sqrt{100}/\sqrt{n})] + \Phi[-1.96 - (95 - 100)/(\sqrt{100}/\sqrt{n})]$, which is,

Power = $1 - \Phi[1.96 - \sqrt{n}/2] + \Phi[-1.96 + \sqrt{n}/2]$.

That function shows a relationship between power and sample size. For each level of sample size, there is a corresponding sample size. For example, if $n=20$, the corresponding power level would be about 0.97, or, if the power level is 0.95, the corresponding sample size would be 16.

Using Statistical Package (SAS)

Statistical packages like SAS enables a researcher to do the power calculation easily. The procedure in which power and sample size are calculated is specified in the following text.

In SAS, statistical power and sample size calculation can be done either through program editor or by clicking the menu the menu. In the latter, a set of code is automatically generated every time a calculation is done.

PROC POWER and GLMPower

PROC POWER and GLMPower are new additions to SAS as of version 9.0. As of this writing, SAS 9.0 is not currently installed on ITaP machines, but it can be installed on your home computer using disks available in Steward B14. Make sure to bring your Purdue ID.

The table on the following page (taken from the SAS help file) shows the types of analyses offered by PROC POWER. At least one statement is required. The syntax within each statement varies, however, there is some syntax common to all. These common features will be expressed by an example using a paired t-test. More information on each procedure can be found in the SAS help file.

In the example, assume that a pilot study has been done, and that the standard deviation of the difference between the two groups has been found to be 5, with a mean difference of 2. We'd like to calculate the required sample size for an experiment with 80% power.

```
proc power;  
  pairedmeans test=diff  
    meandiff = 2  
    stddev = 5  
    npairs = .  
    power = .80;  
run;
```

Analysis	Statement	Options
Multiple linear regression: Type III F test	MULTREG	
Correlation: Fisher's z test	ONECORR	DIST=FISHERZ
Correlation: t test	ONECORR	DIST=T
Binomial proportion: Exact test	ONESAMPLEFREQ	TEST=EXACT
Binomial proportion: z test	ONESAMPLEFREQ	TEST=Z
Binomial proportion: z test with continuity adjustment	ONESAMPLEFREQ	TEST=ADJZ
One-sample t test	ONESAMPLEMEANS	TEST=T
One-sample t test with lognormal data	ONESAMPLEMEANS	TEST=T DIST=LOGNORMAL
One-sample equivalence test for mean of normal data	ONESAMPLEMEANS	TEST=EQUIV
One-sample equivalence test for mean of lognormal data	ONESAMPLEMEANS	TEST=EQUIV DIST=LOGNORMAL
Confidence interval for a mean	ONESAMPLEMEANS	CI=T
One-way ANOVA: One-degree-of-freedom contrast	ONEWAYANOVA	TEST=CONTRAST
One-way ANOVA: Overall F test	ONEWAYANOVA	TEST=OVERALL

McNemar exact conditional test	PAIREDFREQ	
McNemar normal approximation test	PAIREDFREQ	DIST=NORMAL
Paired <i>t</i> test	PAIREDMEANS	TEST=DIFF
Paired <i>t</i> test of mean ratio with lognormal data	PAIREDMEANS	TEST=RATIO
Paired additive equivalence of mean difference with normal data	PAIREDMEANS	TEST=EQUIV_DIFF
Paired multiplicative equivalence of mean ratio with lognormal data	PAIREDMEANS	TEST=EQUIV_RATIO
Confidence interval for mean of paired differences	PAIREDMEANS	CI=DIFF
Pearson chi-square test for two independent proportions	TWOSAMPLEFREQ	TEST=PCHI
Fisher's exact test for two independent proportions	TWOSAMPLEFREQ	TEST=FISHER
Likelihood ratio chi-square test for two independent proportions	TWOSAMPLEFREQ	TEST=LRCHI
Two-sample <i>t</i> test assuming equal variances	TWOSAMPLEMEANS	TEST=DIFF
Two-sample Satterthwaite <i>t</i> test assuming unequal variances	TWOSAMPLEMEANS	TEST=DIFF_SATT
Two-sample pooled <i>t</i> test of mean ratio with lognormal data	TWOSAMPLEMEANS	TEST=RATIO
Two-sample additive equivalence of mean difference with normal data	TWOSAMPLEMEANS	TEST=EQUIV_DIFF
Two-sample multiplicative equivalence of mean ratio with lognormal data	TWOSAMPLEMEANS	TEST=EQUIV_RATIO
Two-sample confidence interval for mean difference	TWOSAMPLEMEANS	CI=DIFF
Log-rank test for comparing two survival curves	TWOSAMPLESURVIVAL	TEST=LOGRANK
Gehan rank test for comparing two survival curves	TWOSAMPLESURVIVAL	TEST=GEHAN
Tarone-Ware rank test for comparing two survival curves	TWOSAMPLESURVIVAL	TEST=TARONEWARE

Power and Sample Size Calculation Using SAS Menu

Power and sample size can also be calculated using the menu in SAS. When using the menu, the user should specify the chosen design for the underlying project, and then fill in the required parameters needed to do the calculation for each design.

The general procedure of using the menu is as follows:

- 1). Open SAS
- 2). Go to the **enhanced editor** window.
- 3). Click the '**solutions**' button on the menu.
- 4). In the submenu, click '**analysis**'.
- 5). In the next submenu, click '**analyst**', then a new window will pop-up.

6). In the new window, click '**statistics**' button on the menu.

7). Select '**Sample size**', then select the design you want to use. (the designs available in that menu include: one-sample t-test, paired t-test, two sample t-test and one-way ANOVA).

8). After you select the design another window pops-up and asks you to input the needed options and parameters. If you need to know the needed sample size for your research, you can select '**N per Group**', then input **number of treatments**, **corrected sum of square**, the **standard deviation** and the **alpha** level. If the researcher wants to calculate the sample size corresponding to each power level, he/she may want to specify the range and interval of power level in the '**Power**' row in the menu.

The corrected sum of squares (CSS) is calculated as the sum of the squared distance from each treatment mean to the grand-mean. For example, there are two treatments with mean of 10 and 20, respectively. That gives us a grand mean of $(10+20)/2=15$ (assuming equal cell size). Therefore, the corrected sum of squares is: $(10-15)^2+(20-15)^2=50$.

Once the request for calculation is submitted, SAS will pop-up a window which includes a table of power level and corresponding sample size. You can also ask SAS to generate a curve showing the relation between power level and sample size. Another important feature of SAS menu is that you can generate the code by which you use to do the power calculation and it will be displayed in another window.

Example Output

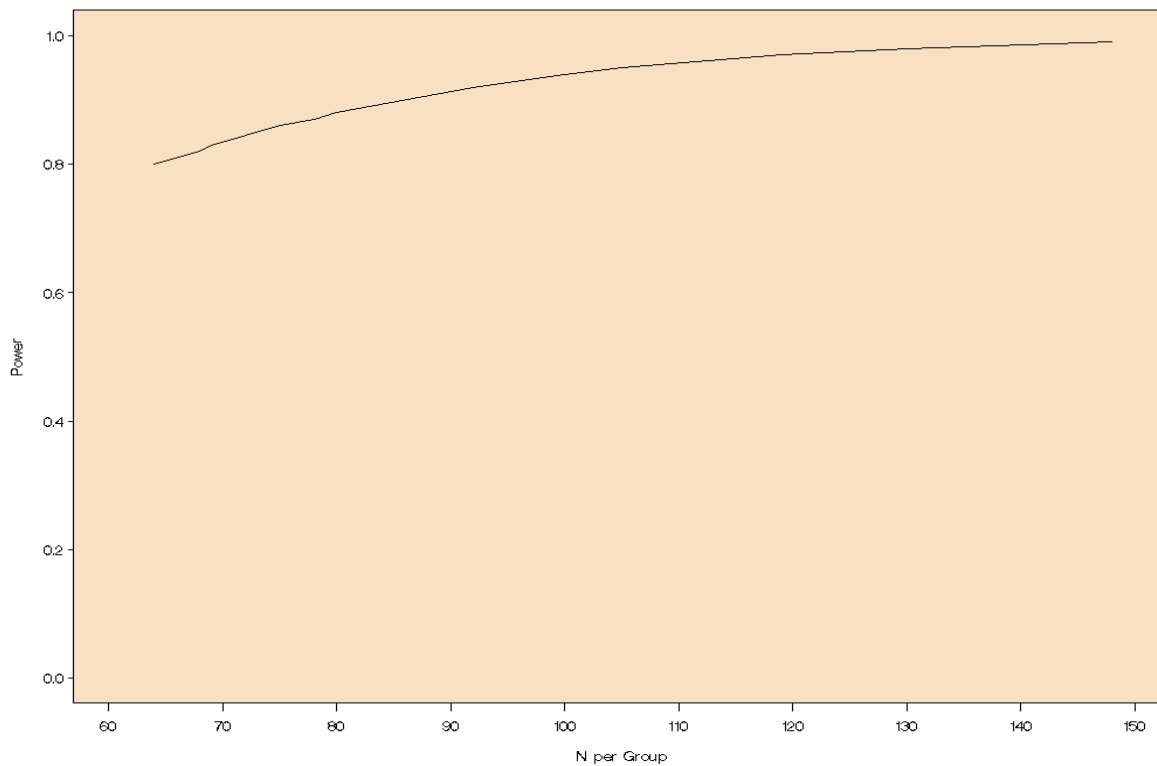
An example is shown below using the CSS mentioned above and assuming a one-way ANOVA design is used. We also assume that the standard deviation is 20 and the alpha is 0.05. We want to find out the corresponding sample size for each power level ranging from 0.8 to 0.99 at 0.01 intervals. The outputs should look like the following:

One-Way ANOVA
Treatments = 2 CSS of Means = 50
Standard Deviation = 20 Alpha = 0.05

	N per
Power	Group
0.800	64
0.810	66
0.820	68
0.830	69

4	0.840	71
5	0.850	73
5	0.860	75
7	0.870	78
8	0.880	80
9	0.890	83
0	0.900	86
1	0.910	89
2	0.920	92
3	0.930	96
4	0.940	100
5	0.950	105
6	0.960	112
7	0.970	119
8	0.980	130
9	0.990	148

One-Way ANOVA
 # Treatments = 2 CSS of Means = 50
 Standard Deviation = 20 Alpha = 0.05



The output above gives the required sample size per group for each power level. For example, if we want a power level of 0.9, we actually need $86 \times 2 = 172$ subjects in the sample.

Example from Consulting Service Clients' Project

Consider a hypothetical study in which the goal is to determine the effectiveness of a certain drug in lowering diastolic blood pressure. A group of men and women will be randomly assigned to either receive the drug or to receive a placebo. This design can be analyzed as a one-way ANOVA with four groups: (1) men not taking the drug, (2) men taking the drug, (3) women not taking the drug, and (4) women taking the drug. A previous study indicates that the means for each of these groups might be 93, 74.6, 86.7, and 76.5 respectively. That study examined a similar question and although the means may not be exact, they are a good estimate. The standard deviation in diastolic blood pressure between subjects was 27 for that study. The researcher planning the study would like to know the total number of subjects that will be needed to detect a practical difference in the diastolic blood pressure between subjects receiving the drug and the subjects not receiving the drug. A significance level of 0.05 and a power of 0.8 are desired.

The following SAS code was used to arrive at an appropriate sample size given these conditions.

```
proc power;
onewayanova
    test=contrast
    groupmeans = 93 | 74.6 | 86.7 | 76.5
    stddev = 27
    alpha = 0.05
    contrast = (-1 1 -1 1)
    ntotal =.
    power = 0.8;
    plot x=power min=0.6 max=1.0;
run;
```

Explanation of code:

onewayanova - Designates the type of design.

test=contrast - Designates the type of test for which the power will be computed. In this case, a contrast which will compare subjects receiving the drug to subjects not receiving the drug is the main test of interest.

groupmeans - Step where each of the four group means are listed. If other magnitudes of mean difference were of interest, these could be modified.

stddev - Step where the standard deviation is specified.

alpha - Step where the significance level is specified.

contrast - Specifies the details of the contrast. In this case, the contrast will be between groups 1 and 3 (men and women not taking the drug) and groups 2 and 4 (men and women taking the drug). If a contrast that compares men and women were of interest, this step could read: contrast= (1 1 -1 -1).

ntotal =. - Specifies that the total sample size is what needs to be calculated. This could be given and the power for that particular sample size could be calculated instead.

power =0.8 - Step where the desired power is specified. This could be calculated (designated with at '.') if the sample size is given.

plot x=power min=0.6 max=1.0 - This statement provides a power curve which will display power ranging from 0.6 to 1.0 on the x-axis and the sample size which corresponds to that power on the y-axis.

ANOVA Power Calculation Results

The POWER Procedure

Single DF Contrast in One-Way ANOVA

Fixed Scenario Elements

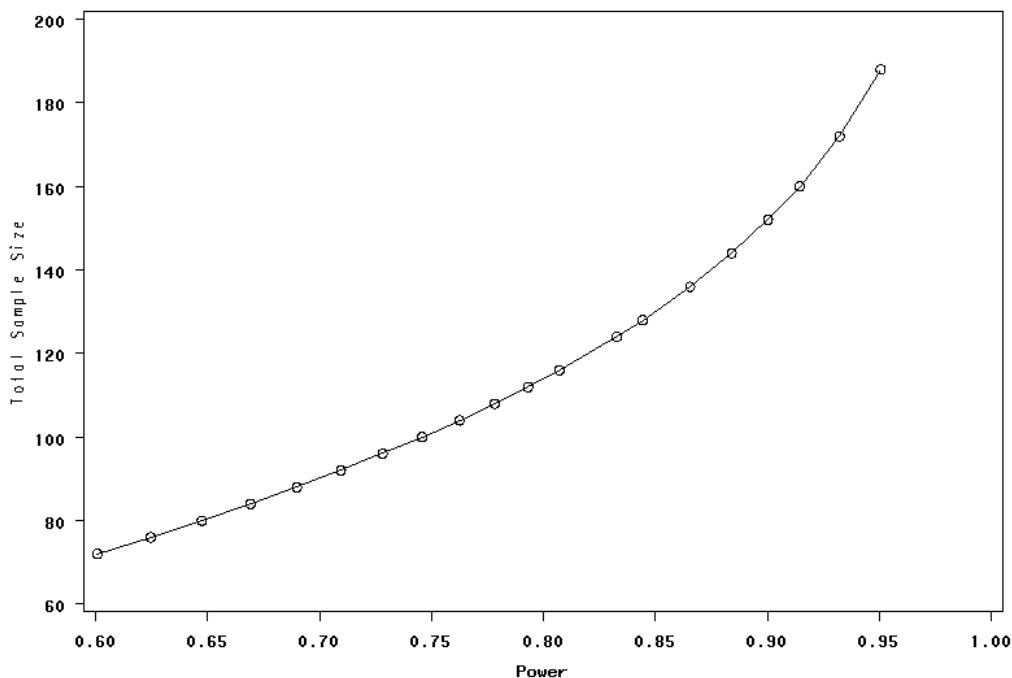
Method	Exact
Contrast Coefficients	-1 1 -1 1
Alpha	0.05
Group Means	93 74.6 86.7 76.5
Standard Deviation	27
Nominal Power	0.8
Number of Sides	2
Null Contrast Value	0
Group Weights	1 1 1 1

Computed N Total

Actual	N
<u>Power</u>	<u>Total</u>
0.807	116

From this output, it was determined that 116 subjects total or 116/4=29 subjects per group will be needed to achieve a power of 0.807 for the specified test.

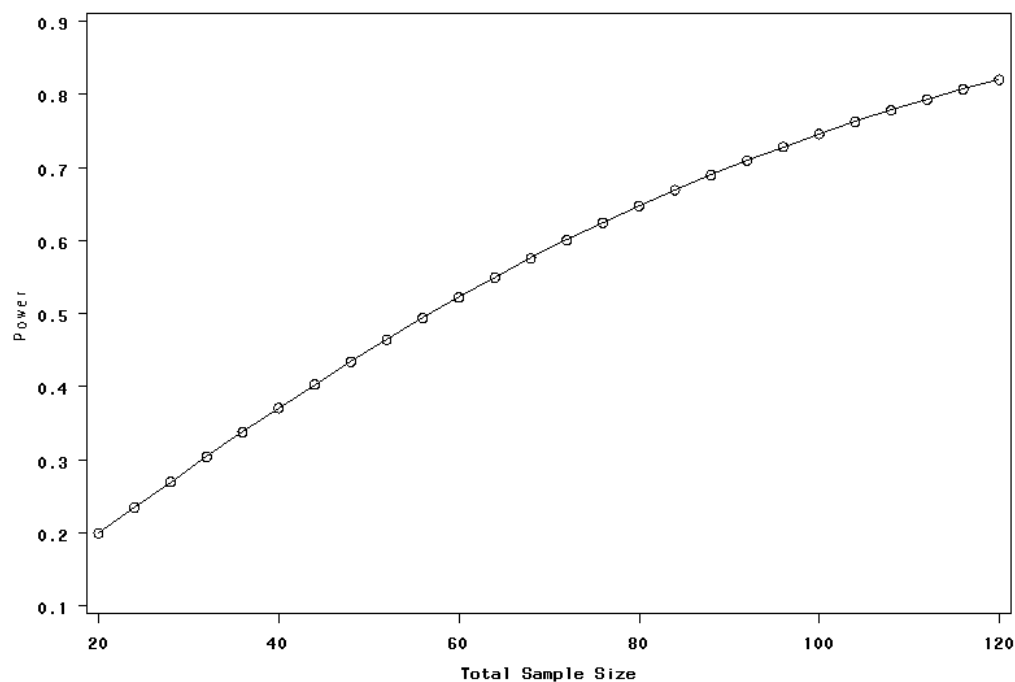
After seeing this result, the researcher may be willing to either recruit more subjects to achieve a higher power or recruit less subjects and sacrifice a small reduction in power. To visualize these kinds of tradeoffs, two power curves were constructed. The first curve (i), plots sample size as a function of power. The SAS code for this plot was given previously. This curve would be useful if the researcher knows a range of power that is desired. From this graph, we can see that lowering the power to 0.75 results in a sample size of around 100, whereas increasing the power to 0.80 results in a sample size of around 130.



i. Power Curve for ANOVA. Sample size versus power.

Alternatively, if the researcher knows a range of sample sizes that is practical in terms of cost and availability of subjects, a different type of power curve might be more useful. This curve (ii) graphs power as a function of sample size. From this curve, we can see that if only 80 subjects complete the study, the power will be reduced to around 0.65. If subjects are likely to withdraw from the study, this curve could also be useful for hypothetical situations involving different numbers of subjects dropping out given a certain number of subjects are recruited in the beginning of the study. The SAS code for this curve (ii) is the same as for the previous

4 curve (i) except a number must be specified for ntotal, a dot must be specified for power, and the plot statement
5 must change to plot x=n min=20 max=120.



ii. Power Curve for ANOVA. Power versus sample size.

Future Plan for the Project

The next step of the project is trying to find out how to do power calculation on different kinds of designs and how to do power analysis on other software packages other than SAS.