

# **OVERVIEW AND EXAMPLES OF POWER AND SAMPLE SIZE DETERMINATION USING SAS PROC POWER**

Design, Epidemiology, and Biostatistics Core  
Pennington Biomedical Research Center  
Baton Rouge, LA

*Supported by 1 U54 GM104940 from the National Institute of General Medical Sciences of  
the National Institutes of Health which funds the Louisiana Clinical and Translational  
Science Center*

# TABLE OF CONTENTS

<b>Introduction to Power Analysis</b> . . . . .	<b>1</b>
<b>Power and Sample Size in SAS</b> . . . . .	<b>2</b>
<b>Test for Correlation</b> . . . . .	<b>3</b>
<b>One-Sample Test for Proportion</b> . . . . .	<b>6</b>
<b>One-sample <math>t</math>-Test</b> . . . . .	<b>8</b>
<b>Paired <math>t</math>-Test</b> . . . . .	<b>10</b>
<b>Two-Sample <math>t</math>-Test</b> . . . . .	<b>12</b>
<b>One-Way ANOVA Overall <math>F</math>-Test</b> . . . . .	<b>14</b>

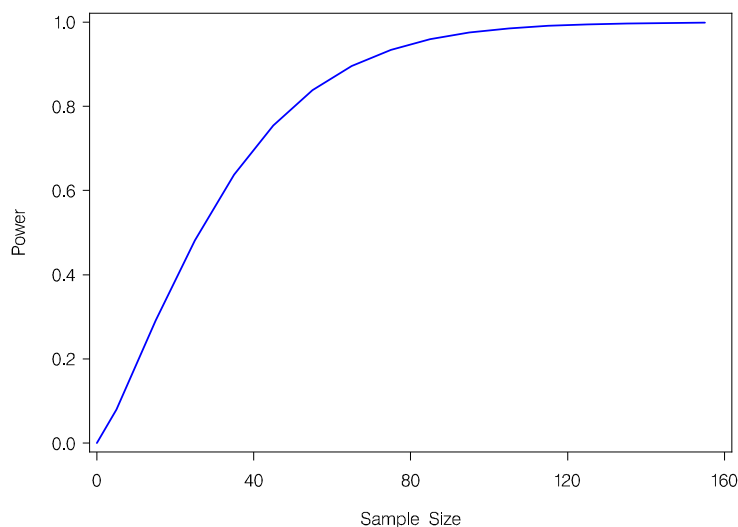
## Introduction to Power Analysis

Let us first provide a general definition of power. Power can be thought of as the probability of detecting an effect, given that the effect actually exists. Statistically, this is the probability of rejecting the null hypothesis given that it is false.

Power and sample size analyses are performed in the design phase of an experiment. It is typically of interest to find what sample size is required for the study in order to achieve a certain level of power to detect an effect of a certain size. "Effect" is defined differently for different experiments. One example is the difference in means of some outcome between a treatment and control group.

In some instances, it may be desired to find power given particular sample size. For instance, consider the case in which the maximum possible sample size is restricted by some study criteria. In this case, it is of interest to find out what level of power can be achieved given this allowable sample size.

Last, let us consider the relationship between power and sample size. For a fixed effect size, power always increases with increasing sample size. Conversely, an increase in power requires a larger sample size. See the following graph for clarification:



## Power and Sample Size in SAS

The following are guidelines for performing power and sample size using the POWER procedure in SAS. The POWER procedure can be utilized to obtain sample size configurations for given levels of statistical power and vice versa.

The general syntax for this procedure is:

```
PROC POWER;  
<ANALYSIS TYPE> <options>;  
RUN;
```

Available options for this procedure are:

- ONECORR - power analysis for test of correlation between two variables
- ONESAMPLEFREQ - power analysis for a one-sample test of a proportion
- ONESAMPLEMEANS - power analysis for a one-sample t-test
- PAIREDMEANS - power analysis for a paired t-test
- TWOSAMPLEMEANS - power analysis for a two-sample t-test
- ONEWAYANOVA - power analysis for overall F-test in a one-way ANOVA

Provided below is the SAS syntax required to run power analyses for simple cases of these common statistical designs. There are, however, many other options available in SAS PROC POWER for situations outside of those described in this tutorial.

## **ONECORR: *Test for correlation***

Example: A researcher is interested in seeing whether a significant positive correlation exists between reading speed and IQ in adolescents. Before beginning the study, the researcher would like to know what sample size would be required to detect a positive correlation of 0.5 with power of 80%.

First we set up the null and alternative hypotheses:

$$H_0 : \rho = 0$$

$$H_1 : \rho > 0$$

Next, we set the level of significance (0.05 is typically the default value). Finally, to obtain the required sample size, we employ the following code:

```
PROC POWER;  
ONECORR SIDES = 1  
    ALPHA = 0.05  
    CORR = 0.5  
    NTOTAL = .  
    POWER = 0.8;  
RUN;
```

In the PROC POWER code above, the **SIDES=1** option indicates that the test being performed is one-tailed. Notice in the alternative hypothesis above, the inequality is only in one direction. The option **ALPHA=0.05** specifies the significance level used and **CORR=0.5** specifies the correlation we are looking to detect in the data. The final two options, **NTOTAL** and **POWER** refer to sample size and power respectively. Power and sample size are the two values that an investigator is attempting to determine in a power analysis. In every run of PROC POWER, regardless of the type of analysis,

there will be options for sample size and power specifications. When running PROC POWER, one of these will be specified and the other left blank. In this example, power is specified at 80% and the required sample size is needed. Specifically, `NTOTAL` is left blank so that the output will contain the total sample size required at 80% power. The output is given below. Based on this, we see that a total sample size of 23 subjects is required in order to detect a correlation of 0.5 at 80% power.

Correlation	Actual Power	N Total
0.5	0.808	23

It may be of interest to investigate what sample sizes are required for detecting various levels of correlation. To do so, a list is used. Notice, in the SAS code below, that three values for the correlation coefficient (0.2, 0.5, and 0.8) are provided. The resulting output gives the estimated sample sizes for the correlations at the specified 80% power.

```

PROC POWER;
ONECORR SIDES = 1
ALPHA = 0.05
CORR = 0.5 0.5 0.8
NTOTAL = .
POWER = 0.8;

RUN;

```

The results are given in the table below:

Correlation	Actual Power	N Total
0.2	0.801	153
0.5	0.808	23
0.8	0.843	8

It turns out that a large sample size ( $N=153$ ) is required to detect only a small correlation between IQ and reading speed, but much smaller sample sizes are required to detect the larger correlations.

## **ONESAMPLEFREQ: *One-sample test for proportion***

Example: A survey claims that 9 out of 10 dentists recommend a particular brand of toothpaste for their patients suffering with sensitive teeth. A researcher decides to test this claim by taking a random sample of 80 dentists, but wants to first find out if this sample size is large enough to achieve 80% power.

First we set up the null and alternative hypotheses:

$$H_0 : p = 0.9$$

$$H_1 : p \neq 0.9$$

If we want to test the hypothesis at a 0.05 level of significance (typically the default value), and want to detect a difference of 0.1, the following code will output the required sample size.

```
PROC POWER;  
ONESAMPLEFREQ TEST  
    SIDES = 2  
    NULLPROPORTION = 0.9  
    ALPHA = 0.05  
    PROPORTION = 0.8  
    NTOTAL = 80  
    POWER = .;  
RUN;
```

Unlike in the example given previously for a correlation analysis, we the option **SIDES=2**. This specifies a two-tailed test based on the form of the alternative hypothesis. **NULLPROPORTION=0.9** specifies the proportion value in the null hypothesis and **PROPORTION=0.8** is the expected proportion in the sample (the difference between the null proportion and this value is 0.1). Notice that in this example, **NTOTAL** is spec-



ified and **POWER** is set equal to missing. If the sample size option is specified and **POWER** is missing, the procedure will produce resulting power based on the given sample size in the output window when executed.

The resulting output from the example above gives us the following information:

N Total	Power
80	0.654

The power based on a sample size of 80 is only 65.4%. In order to achieve 80% power, the sample size must be increased. As mentioned in the previous example, certain options allow ranges of values to be investigated. For instance, using the above example, we can investigate what power would be obtained for total sample size set at values ranging from 80 to 110. If looking over a range of equally spaced values, **TO** and **BY** statements may be used to simplify coding. The syntax is the same as in the box above, but the **NTOTAL** option will now be:

**NTOTAL = 80 to 100 by 10**

The SAS output window will then display the power estimates associated with each sample size specified in the PROC POWER statement.

N Total	Power
80	0.654
90	0.740
100	0.808

It is apparent that a sample size of 100 dentists will be sufficient to achieve 80% power.

## **ONESAMPLEMEANS: *One-sample t-test***

Example: A researcher is planning a pharmacokinetic study on a new formulation of a drug. The current formulation has an average elimination rate of 0.06. The researcher hypothesizes that the elimination rate for the new formulation is higher than 0.06. Wanting to be confident, the researcher would like to see how large the sample size must be to achieve 90% power. A standard deviation of 0.02 will be used based on studies of the original formulation of the drug.

First we set up the null and alternative hypotheses:

$$H_0 : \mu = 0.06$$

$$H_1 : \mu > 0.06$$

We want to determine necessary sample sizes to detect various mean differences in elimination rate at power of 90% and a 0.05 level of significance. The following code will output the required sample sizes.

```
PROC POWER;  
ONESAMPLEMEANS SIDES = 1  
    NULLMEAN = 0.06  
    ALPHA = 0.05  
    MEAN = 0.07 to 0.1 by 0.01  
    STDDEV = 0.02  
    NTOTAL = .  
    POWER = 0.9;  
RUN;
```

The setup for this type of test is similar to that of the one-sample proportion. Here, **NULLMEAN** specifies the mean in the null hypothesis and **MEAN** specifies the mean one would expect to see in the sample. In this example, we look at a range of means

greater than the null mean of 0.06. The **STDDEV** option specifies the standard deviation to use in power and sample size calculations.

Mean	Actual Power	N Total
0.07	0.903	36
0.08	0.924	11
0.09	0.931	6
0.10	0.908	4

From this table, it appears that 90% power can be achieved with relatively manageable sample sizes at each level of mean elimination rate specified. Of course, as the difference between observed and hypothesized mean increases, required sample size decreases.

## **PAIREDMEANS: *Paired t-test***

Example: A researcher is interested in investigating whether BMI changes in males aged 55-65 years after spending four weeks on a novel diet and exercise program. The researcher plans to take BMI measurements on a random sample of men before and after the intervention and see whether there was a change. An 80% level of power is desired and a standard deviation of 2.0 based on past studies of weight loss and BMI change is used for calculations.

First we set up the null and alternative hypotheses:

$$H_0 : \mu_d = 0$$

$$H_1 : \mu_d \neq 0$$

We want to figure out the number of pairs (with pair defined as a set of before and after scores for a single subject) needed to detect a range of mean differences greater than 0.5. The following code will output the required sample size.

```
PROC POWER;  
PAIREDMEANS SIDES = 2  
      NULLDIFF = 0  
      ALPHA = 0.05  
      MEANDIFF = 0.5 to 3.0 by 0.5  
      CORR = 0.4  
      STDDEV = 2.0  
      NPAIRS = .  
      POWER = 0.8;  
RUN;
```

The syntax for this type of t-test is a bit different than that for the one-sample t-test. Here, the options **NULLDIFF** and **MEANDIFF** replace **NULLMEAN** and **MEAN**.

In addition, we need to specify a value for the **CORR** option, which designates the correlation between pairs being investigated. The final difference in the code above is the **NPAIRS** option rather than **NTOTAL**.

Below is the displayed output:

Mean Diff	Actual Power	N Pairs
0.5	0.801	153
1.0	0.804	40
1.5	0.806	19
2.0	0.821	12
2.5	0.849	9
3.0	0.853	7

As with other tests discussed previously, larger sample sizes are needed to detect smaller changes in the mean.

## **TWOSAMPLEMEANS: *Two-sample t-test***

Example: A researcher wants to study how daily protein supplementation in the elderly population will affect baseline liver fat. The study budget will allow enrollment of 50 patients. Half will be randomized to a placebo group and half to the protein supplement treatment group and the trial will be carried out over one month. It is desired to see whether the mean change in percentage of liver fat from baseline to the end of the study differs between the two groups in the study. The researcher needs to know what power can be obtained under the sample size restrictions to identify a change in mean percent liver fat of 0.10. Based on results observed in past similar studies, a common standard deviation of 0.12 will be used for each treatment group in the power analysis.

First we set up the null and alternative hypotheses:

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 - \mu_2 \neq 0$$

If we want to test the hypothesis at a 0.05 level of significance (typically the default value) with a total sample size of 40, and want to detect a difference of at least 0.15, the following code will output the resulting power.

```
PROC POWER;  
TWOSAMPLEMEANS SIDES = 2  
                  NULLDIFF = 0  
                  ALPHA = 0.05  
                  MEANDIFF = 0.1 to 3.0 by 0.5  
                  STDDEV = 0.12  
                  NPERGROUP = 25  
                  POWER = .;  
RUN;
```

Notice the change in syntax above when specifying sample size. The option **NPERGROUP** is used to specify that 25 subjects are randomized to each of two treatment groups. This option can only be used if the design is balanced (equal number of subjects in each group). If the design is to be unbalanced, the option **GROUPNS** may be used to specify the number allocated to each group. For instance, in the above example, the researcher may have wanted 35 subject in the treatment group and only 15 in the placebo group. The following SAS code would be used:

```
PROC POWER;
TWOSAMPLEMEANS SIDES = 2
NULLDIFF = 0
ALPHA = 0.05
MEANDIFF = 0.1 to 3.0 by 0.5
STDDEV = 0.12
GROUPNS = (35 15)
POWER = .;

RUN;
```

The following are the resulting estimates of power for the balanced and unbalanced sample size configurations just described.

N Per Group	Power
25	0.823

N Group 1	N Group 2	Power
35	15	0.754

With the unbalanced case, power is lower. Power can be improved by using the balanced sample size configuration or adding subjects to the placebo group.

## **ONEWAYANOVA: *One-way ANOVA overall F-test***

Example: A researcher is interested in investigating the effects of three different diets on percent weight loss when implemented along with a 5-day per week cardio exercise program. The diets include a low carbohydrate diet, a high protein diet, and a control diet (just exercise). Before beginning the study, sample size determinations must be made. The researcher would like to achieve power of 80%.

First we set up the null and alternative hypotheses:

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$H_1$  : *At least one pairwise difference between means exists*

If we want to test the overall hypothesis at a 0.05 level of significance (typically the default value) with power at 80% the following code will output the resulting power. For the values of mean percentage of weight loss in each group and for the standard deviation specified in the code below, estimates obtained in a small preliminary study are used.

```
PROC POWER;  
ONEWAYANOVA TEST = OVERALL  
    ALPHA = 0.05  
    GROUPMEANS = 8 | 10 | 6  
    STDDEV = 3.0  
    NPERGROUP = .  
    POWER = 0.8;  
RUN;
```

The option **TEST=OVERALL** indicates that the power analysis is based on the overall  $F$ -test for differences between groups. The **GROUPMEANS** option requires values of expected mean outcome for each group in the analysis. The option **NPERGROUP** is



used to request the group sample size estimates for a balanced design.

Below is the resulting sample size required for each of the three diet groups in order to achieve power of at least 80%.

N Per Group	Power
12	0.804