Paper SA18-2011

A Macro to Estimate Sample-size Using the Non-Centrality Parameter of a Non-Central F-Distribution

Matthew C. Fenchel, M.S., Cincinnati Children's Hospital Medical Center, Cincinnati, OH
Raouf S. Amin, M.D., Cincinnati Children's Hospital Medical Center, Cincinnati, OH
Rhonda D. VanDyke, Ph.D., Cincinnati Children's Hospital Medical Center, Cincinnati, OH

Abstract

An approximation of the power of an F-test can be estimated using the non-centrality parameter calculated from the observed F-statistic. Often though, investigators planning a study are interested in an "opposite" question: "With the given statistical model and data parameters, what *sample-size* is needed to obtain a desired power?" This macro expands the approach and SAS® code described by Littell (2006) and Stroup (1999) in order to create a user-friendly method of sample-size calculation for user-defined general linear models (including mixed and repeated measures). The program is designed for calculating sample-size needed to obtain a desired power for group contrasts.

Introduction

An approximation of the power of an F-test can be estimated using the non-centrality parameter calculated from the observed F-statistic. This has been well described by Littell (2006), Stroup (1999), Rosa (2006) and Guenther (1979). Littell (2006) and Stroup (1999) also included SAS® code to output the power from a statistical model, using a fixed sample-size. This answers the question: "With the given statistical model, data parameters and sample-size, what is the *power* (for the specified contrast, etc.)?"

Often though, investigators planning a study are interested in an "opposite" question: "With the given statistical model and data parameters, what *sample-size* is needed to obtain a desired power (for a specific test, etc.)?" Holding a statistical model fixed, the relationship between power and sample-size is one-to-one. Based on this, the "non-central F" method for calculating power (using a fixed sample-size) can be "reversed" and used to calculate the sample-size needed to obtain a desired (fixed) power.

This macro adopts and expands the approach and SAS® code described by Littell (2006) and Stroup (1999) in order to create a user-friendly method of sample-size calculation for user-defined general linear models. The program is designed for group contrasts -- either independent groups, or time-points with non-independent observations. Ideally, this program is designed for ANOVA, mixed and longitudinal models. The MIXED procedure is used to run the user-defined model. The output includes two tables. The first table lists the contrast(s) requested and statistics for those contrasts that were used for the power and sample-size calculations. The second table produces the sample-size required (by cell or group) in order to reach the given power level.

This program can be run "as is" with no initial input from the user. It calculates power for a "Pre-Post" contrast and for a "Case-Control" contrast. This will give the user an idea of the output produced, etc.

To use this program, the user is mostly asked to input details about the hypothesized data and model, along with a few lesser details. The applications described in this paper can be used by someone with an intermediate knowledge of ANOVA, mixed or longitudinal models, and intermediate knowledge of SAS® procedures, including macros and macro variables.

Overview of the Concept

The underlying programming concept behind this SAS® code is described in detail by Littell (2006) and Stroup (1999). Essentially, the user implements the same basic statistical model/design that he anticipates using for analysis. With hypothesized estimates of the cell-means (perhaps from previous studies or pilot data), a "dummy" data set is created with replicates of cell-means taking the place of actual data values. With hypothesized estimates of the variance parameters, using a variance/covariance structure of the user's choice, the model is run on the "dummy" data set, holding the variance parameters fixed. Key statistics are then generated that are used for sample-size, power and effect size. (Effect size is defined as the following calculation: $(nc / N)^{0.5}$, where “nc” is the non-centrality parameter of the F-distribution and “N” is the total sample-size of the two groups in the contrast. See “FAQ” [2011] and Liu [2004].)

In Brief: Power and the Non-Centrality Parameter of a Non-Central F Distribution

Stroup (1999) provides a thorough explanation of linear and mixed-linear model theory behind the estimation of power using the non-centrality parameter of a non-central F distribution. The purpose of this paper is not to duplicate this excellent write-up, but rather to describe a user-ready program where this estimation can be used. However, we thought a brief overview would be helpful.

Power (in a research context) is, of course, the probability of rejecting the null hypothesis (H_0) when the alternative hypothesis (H_a) is true. A Type I error is rejecting H_0 when actually H_a is false. The well-known p-value is simply an estimate of the probability of committing a Type I error. A Type II error is rejecting H_a when H_a is actually true. Thus, power equates to one minus the probability of committing a Type II error.

The rate of a Type I error is most often denoted by α, and the rate of a Type II error is denoted by β. Thus, power = 1 - β. A common table illustrating the relationships between decisions and “reality” in hypothesis testing is shown below:

	H_0 is true	**H_0 is false**
Reject H_0	Type I error *[Pr(Type I error) = α]*	Correct decision[1]
Fail to reject H_0	Correct decision	Type II error *[Pr(Type II error) = β]*

[1]Probability of this correct decision = 1 - β = power.

In general, the estimation of power is dependent on a number of factors: the sample-size (in specific, of the groups being compared), the *a priori* α-level (where α equals the probability of committing a Type I error that the researcher is willing to allow -- often α = 0.05), the level of variation within the data and the ability to account (allocate or assign) that variation to measured effects, and the lowest level of difference between groups or treatments that the researcher wishes to detect.

It is not difficult to see that -- holding all but one of these factors fixed -- one can alter the power by changing the remaining factor. Of interest to most researchers is the ability to increase power by increasing sample-size (assuming all other factors can be held constant). Of course, practical limitations (availability of subjects, cost, time, etc.) usually put a “brake” on this.

As Stroup (1999) insightfully discusses though, careful consideration of treatment design and experimental design is critical. This can possibly keep the need for an unaffordable sample-size at bay by preventing unwanted variation from entering the data through experimental units or experimental time-points (repeated measures). As a corollary to this, the statistician has the challenge (at the end of the experiment) of modeling data in the most statistically appropriate (yet practical) way possible, considering the limitations of the experimental design and data, in order to account for as much random variation as possible. This will hopefully lead to an accurate estimation of the research parameters of interest -- and to the detection the anticipated difference (assuming such a difference exists).

The concept proposed by Stroup (1999) and Littell (2006) (and applied in this paper) is to use the same statistical model (that will be used to analyze the data) to calculate *a priori* power -- in specific, with regards to mixed and repeated-measures models.

Notation that is often used for the general linear mixed model is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\mu} + \mathbf{e}$$

where

$$\boldsymbol{\mu} \sim N(\mathbf{0}, \mathbf{G})$$
$$\mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$$
$$\mathbf{Y} \sim MVN(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$$
$$\mathbf{V} = \mathbf{ZGZ}` + \mathbf{R}$$

Briefly, $\mathbf{Y}$ is the vector of observations for the response variable; $\mathbf{X}$ is the design matrix for fixed effects; $\boldsymbol{\beta}$ is the unknown coefficients (parameter) vector for fixed effects; $\mathbf{Z}$ is the design matrix for random effects; $\boldsymbol{\mu}$ is the unknown coefficients vector for random effects; $\mathbf{e}$ is the vector of random residuals (errors); $\mathbf{G}$, $\mathbf{R}$ and $\mathbf{V}$ are the variance / covariance matrixes for the random effects, random residuals and $\mathbf{Y}$, respectively.

The variance / covariance matrices $\mathbf{G}$ and $\mathbf{R}$ can take on numerous valid structures, and PROC MIXED offers a large variety these forms. Please see SAS/STAT® 9.2 User's Guide (The MIXED Procedure) for a listing and descriptions of these structures.

Given an estimable contrast of H_0: $\mathbf{L}`\boldsymbol{\beta} = 0$ then

$$F_{L\beta} \sim F_{\{rank(L),\ ddf,\ \lambda\}}$$
$$\lambda = (L`\beta)`[L`(X`V^{-1}X)^{-1}L]^{-1}\ (L`\beta)$$
$$ddf = \text{denominator degrees of freedom}$$

Here rank(L) represents the rank of the contrast matrix and λ is the non-centrality parameter of the above F-distribution. Under H_0, $\lambda = 0$; however, under H_a, $\lambda > 0$. We thus come to the following power estimation:

$$\text{Power} = \Pr[F_{\{rank(L),\ ddf,\ \lambda\}} > F_{\{rank(L),\ ddf,\ 0,\ \alpha\}}]$$

$F_{\{rank(L),\ ddf,\ 0,\ \alpha\}}$ is also referred to as the "critical F" (Littell) or simply "F_{crit}" (Stroup) and is the value of F from the central F-distribution at α.

This summary is abbreviated, but hopefully it will give the reader the main idea how power -- which, in reality, is simply a probability -- is derived. The formula for power will be referred to again as the SAS® code is being discussed.

Overview of the Program

As mentioned earlier, this program is designed to calculate power for estimable contrasts, i.e. differences between different levels of a group, block, time, or some interaction of these effects. There are two separate variables in the program that can be used for the group, block or time variables -- "Group" and "Block". How those two variables are modeled is up to the user. Both of these variables have the possibility of having up to four (4) levels each, for a total of 16 different cell-means. The models can be almost any form of the general linear model, including models with random effects, random coefficients and repeated measures.

The program (as given in the Appendix) can be run almost "as is" -- in other words, values are given for most of the inputs, so that the user can see how the program functions. (The only thing the user would need to change is the name and path of the output file in Step 5.)

The user is able to specify five (5) levels of power for which sample-size will be calculated. The program will use cell-means (given by the user) to create a "dummy" data set with replicates of these cell-means. Holding the user-defined variance/covariance estimates fixed, the program will use PROC MIXED to generate the necessary statistics to calculate power for the desired contrast. If a desired level of power is not reached, the number of replications in the "dummy" data set will be increased until the specified power is reached or exceeded. Once this occurs, the program either moves on to the next level of power, or comes to an end.

The program assumes a balanced design among groups ($n_1 = n_2 = n_3$ and so on). In other words, as the size of the data set is increased to achieve the desired power level, each “Group” or “Block” is added to equally.

After sample-sizes have been calculated for all desired levels of power, a final *.rtf file is created using ODS output. As mentioned before, the output includes two tables. The first table lists the contrast(s) requested, and statistics for those contrasts. The second table produces the sample-size required (by cell or group) in order to reach the given power level.

The next several sections describe what the user will need to input. After that, each part of the program is discussed.

Step 1: Initial Sample-Size, α-Level and Desired Power

The first input allows the user to provide an initial sample-size. While this should always be lower than the expected sample-size, it does not need to be “ultra-low.” In other words, if a user realistically expects the sample-size to come close to 200, you don’t need to have the “dummy” data set start out at N = 10. It will eventually reach the 200-level, but it will obviously take more time to do so. Starting at N = 150 (in this case) would save some time. That being said, this program does not take long to run.

The second item is to set the alpha-level (α) -- i.e. the *a priori* level at which the researcher decides that statistical significance is reached. Of course, the range of α should be $0 < \alpha < 1$.

The final inputs are the desired power levels: dp1 - dp5. These do not need to be in any type of order. However, if sample-size calculations for less than five (5) levels of power are desired, the remaining, unused macro variables should be set to 0. (The program in the Appendix has an example of this.) Of course, once again, the range of power should be $0 < dp < 1$.

Step 2: Estimated Cell Means

In this section, the user inputs the hypothesized (or estimated) means of the response variable for each level of “Group” or “Block” or Group x Block combination. In this example below (corresponding to the program in the Appendix), “Group” is used as a treatment variable and “Block” is used as a time variable. There are two levels of “Group” (treatment) and two levels of “Block” (time). Thus, we have the following input:

```
%let G1B1M = 63;                    %let G2B1M = 45;
%let G1B2M = 86.5;                  %let G2B2M = 64;
```

G1B1M is the mean for treatment = 1 at baseline (time = 1), while G2B1M is the mean for treatment = 2 at baseline. In a similar fashion, the means at the second time-point are represented by G1B2M and G2B2M.

What is most important about these numbers is that differences in their values reflect the investigator’s hypothesized difference for which he needs a sample-size calculation. In this case, we are interested in a difference of 23.5 between time-points for Group = 1 and a difference of 22.5 between treatments at Time (Block) = 2. As long as we obtain these differences in the contrasts, we could substitute the numbers above with almost any other values, since we are interested in sample-sizes based on the power of the contrasts -- not based on the power for the main effects.

Of course, as described in the program, simply having either a “Group” or a “Block” variable (not both) is entirely fine.

Step 3: Output -- Yes or No

In this step, the user can specify whether or not he wants to see all of the output in the output window -- which will primarily come from PROC MIXED. No matter which option the user chooses (“No” or “Yes”), only final tables will be sent to the ODS .rtf file that the user specifies.

Step 4: Specify Details of the PROC MIXED Code

Now the user will need to specify most of the PROC MIXED code -- everything within the %STR() macro function. Required code are the MODEL, PARMS, CONTRAST and ESTIMATE statements. Optional would be any REPEATED or RANDOM statements.

The MODEL statement must be in some form of modeling “Resp” as the response variable against some combination of “Group”, “Block” or “Group*Block”. These are the names of the respective variables that are generated in the “dummy” data set that PROC MIXED will use.

For this run, (as described above and in the Appendix), this section is as follows:

```
%let ProcMixedCode =
    %str( model Resp=Group*Block  / solution noint;
          repeated Block / subject = rep(Group) type = csh;
          parms (547.56) (2180.89) (.5)  / noiter;

           contrast 'G1: B2 - B1' Group*Block -1 1 0 0;
           estimate 'G1: B2 - B1' Group*Block -1 1 0 0;

           contrast 'B2: G1 - G2' Group*Block 0 1 0 -1;
           estimate 'B2: G1 - G2' Group*Block 0 1 0 -1; );
```

Since for the purposes of this program we were only interested in the Group*Block (Treatment*Time) contrasts, the MODEL statement could be simple -- i.e. modeling an interaction fixed effect only with no intercept. Doing this (and choosing “Yes” in Step 3 above), we could verify the cell-means of each group from the PROC MIXED output, as shown below in the “Estimate” column:

Solution for Fixed Effects							
Effect	**Group**	**Block**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Group*Block	G1	B1	63.0000	7.3997	16	8.51	<.0001
Group*Block	G1	B2	86.5000	14.7678	16	5.86	<.0001
Group*Block	G2	B1	45.0000	7.3997	16	6.08	<.0001
Group*Block	G2	B2	64.0000	14.7678	16	4.33	0.0005

At this point, we need to mention that the experimental unit or subject variable in our “dummy” data set is “Rep” (or replicate). That value though, is not unique. In other words, G1 has a “Rep” = 1 and G2 has a “Rep” = 1. In many models, this won’t matter. However, it is important to know if using the REPEATED statement, as discussed below.

Since we have repeated measures over the “Block” (Time) variable, we use the REPEATED statement, as given above. In the SUBJECT option, we needed to specify “Rep(Group)” to let SAS know that the values for “Rep” are unique only within each level of “Group”. Otherwise PROC MIXED will read duplicates of “Rep” within each time-period (“Block”) and PROC MIXED will crash. The TYPE option is used to specify the variance/covariance structure. (In our case, CSH or heterogeneous compound-symmetry, which is appropriate since we only have two time-points.) This is where the user will need to input the variance/covariance structure he plans to use to model the residual (R-side) variances and covariances (if a repeated measures model is being used). Of course, a similar option is available in the RANDOM statement to model the G-side variance/covariance structure, if that is the model that needs to be run.

In the PARMS statement, you will need to insert the covariance parameter estimates that you wish to use -- and hold fixed -- in running the model. Critical here is the NOITER option, which prevents PROC MIXED from running Newton-Raphson iterations. This option -- combined with the NOPROFILE option in the PROC statement (see Section B below) -- fixes the user-defined values in the PARMS statement at those values, without attempting to update them through iteration estimation.

(A side note here: there is also a HOLD= option in the PARMS statement which would seem to duplicate what was described above in using the NOITER and NOPROFILE options together. At the time of writing though, the authors did not have the opportunity to fully investigate how this option might differ from the NOITER / NOPROFILE combination.)

The contrasts of interest then need to be defined in both the CONTRAST and ESTIMATE statements. While only statistics from the CONTRAST statement are used in the calculation of power and sample-size, results from the ESTIMATE statement are also produced, so that the user can verify that the differences tested within PROC MIXED were indeed at the levels that he intended. (These will be available in the ODS output.)

The content of this %STR() macro function -- and &ProcMixedCode macro variable -- is inserted later in the program (see Section B below).

Step 5: Specify Path and Name of RTF File

Lastly, the user specifies the path and name of the *.rtf file where the final output will be stored. After doing this, there is nothing more for the user to do in order to run the program as designed.

Some of the sections below are not that complicated; however, we divided the program up into sections for ease of discussion.

Section A: Initial "Dummy" Data Set

In this section, through the use of DATA steps, the initial "dummy" data set is created with the various combinations of "Group" and "Block" and these are assigned the respective cell-means, as specified by the user. The size of this data set is initially determined by the user (as described earlier). Macro %DO-%TO and %DO-%UNTIL loops are employed in order to a) create results for each level (up to five) of desired power, and b) increase the size of the data set until the respective level of power is reached.

A small sub-section of the data set (produced by the code in the Appendix) is shown below.

Group	Block	Resp	Rep
G1	B1	63.0	1
G2	B1	45.0	1
G1	B2	86.5	1
G2	B2	64.0	1
G1	B1	63.0	2
G2	B1	45.0	2
G1	B2	86.5	2
G2	B2	64.0	2
G1	B1	63.0	3
G2	B1	45.0	3
G1	B2	86.5	3
G2	B2	64.0	3

Section B: Calculating Power with the Current Sample-Size

The FREQ procedure and the DATA step that are used at the beginning of this section are simply to obtain a count of the number of replicates each combination of Group*Block has, in order to calculate effect size later. This count is saved to the macro variable &CellCount.

Next comes the remainder of the PROC MIXED code. Please notice the macro variable -- &ProcMixedCode -- that reads the statements that were inserted by the user in Step 4.

```
proc mixed data = temp2 noprofile order=data;
  class Group Block Rep;
  &ProcMixedCode
  ods output Contrasts = Cont Estimates = Est;
  run;
```

In the following DATA step, the power is calculated with the current sample-size.

```
data power; set Cont;
  Alpha=&alpha;
  nc=numdf*fvalue;
  fcrit = finv(1-Alpha, numdf, dendf, 0);
  Power=round((1-probf(fcrit, numdf, dendf, nc)), .0001);
  Order = _N_;
  run;
```

The syntax for this DATA step is similar to that published by Littell (2006) and Stroup (1999).

Section C: Determining Whether a Desired Power Level has been Reached

At the beginning of this section, the data set containing the statistics from the CONTRAST statement and power calculations is merged with the data set containing the statistics from the ESTIMATE statement. If the desired level of power has been reached, it is this data set (with a couple of minor modifications) that will be sent to the ODS RTF file.

In a second DATA step (same data set), the difference between the achieved power and the desired power is determined. Depending on this difference, a macro variable is created whose value will decide how many replications will be added to the "dummy" data set: four, two, one or zero. (Of course, if the value comes up zero, it means the desired power has been reached or exceeded and this iteration of the program will come to an end.) If replications need to be added, obviously most of the previous elements of the program will run again until the desired level of power is reached or exceeded.

A logical question may arise at this point: What happens if the user is testing two or more different contrasts (within the same run of this program -- which is certainly acceptable to do) and these contrasts need different sample-sizes to reach the desired level of power? Which contrast will the program look at in order to determine whether the desired level of power has been reached or not?

The program is designed to end (or move on to the next level of desired power) as soon as one of the contrasts reaches or exceeds the current desired level. For example, let us say your first desired level of power is 0.80. The power of your first contrast sits at 0.71 with a sample-size of 50. With the same sample-size (since the number of "Reps" in all groups or blocks are increased equally), the second contrast has reached a power of 0.8152. The program would now end (or move onto another level of desired power, if one has been entered) since one contrast has reached the desired level.

If the user wants to know what sample-size is needed for the first contrast (in order to reach, for example, a power of 0.80), you could re-run the program again, using only that contrast. If the user wants to know what level of power this new sample-size would produce for the second contrast, you could also re-run the program, setting the initial sample-size (in Step 1) to the sample-size that produced the desired power for the first contrast.

After the above DATA steps, there are two PRINT procedures invoked that will show preliminary power and sample-size output -- if the user selected “Yes” in Step 3. Otherwise, only the final results for each level of desired power will be shown when the program is finished.

Section D: Final Output

In this section, some minor steps are completed in order to produce the final output. The number of Groups and Blocks are assigned to macro variables, to be used in a TITLE statement. A final %DO-%TO loop is used to print final results for all desired power levels, and save results in separate data sets.

As has been mentioned earlier, one of the tables produced will show the sample-size needed for each Group-Block combination. Both tables will be shown in the examples below.

Example 1: Completely Randomized Design with Three Replications

The first example is based on hypothetical data from Stroup (1999), and is a completely random design (CRD) with three treatments, with means of 26, 20 and 20, respectively. The estimated residual variance (in the PARMS statement) is set at 5. The main part of the PROC MIXED code is given below:

```
%let ProcMixedCode =
  %str( model Resp=Group  / solution noint;
        parms (5)  / noiter;
        contrast 'Example1' Group 2 -1 -1;
        estimate 'Example1' Group 2 -1 -1; );
```

This example was run for desired power levels of 0.85, 0.90 and 0.99. Our initial sample-size was three (3) per group, since that is what was used in the original example from Stroup (1999) in which power was simply calculated for that sample-size. Here is the actual output from our sample-size program.

FINAL RESULTS
Power Analysis & Effect Size
Alpha = 0.05; Desired Power = 0.85
Groups = 3; Blocks = 1

Label	Estimate	Standard Error	t Value	Num DF	Den DF	F Value	Pr > F	Alpha	Power	Effect Size
Stroup1	12.0000	3.1623	3.79	1	6	14.40	0.0090	0.05	0.8824	1.54919

Sample Size, By Group

Group	Block	Group_n
G1	B1	3
G2	B1	3
G3	B1	3

FINAL RESULTS
Power Analysis & Effect Size
Alpha = 0.05; Desired Power = 0.9
Groups = 3; Blocks = 1

Label	Estimate	Standard Error	t Value	Num DF	Den DF	F Value	Pr > F	Alpha	Power	Effect Size
Stroup1	12.0000	2.7386	4.38	1	9	19.20	0.0018	0.05	0.9726	1.54919

Sample Size, By Group

Group	Block	Group_n
G1	B1	4
G2	B1	4
G3	B1	4

FINAL RESULTS
Power Analysis & Effect Size
Alpha = 0.05; Desired Power = 0.99
Groups = 3; Blocks = 1

Label	Estimate	Standard Error	t Value	Num DF	Den DF	F Value	Pr > F	Alpha	Power	Effect Size
Stroup1	12.0000	2.4495	4.90	1	12	24.00	0.0004	0.05	0.9941	1.54919

Sample Size, By Group

Group	Block	Group_n
G1	B1	5
G2	B1	5
G3	B1	5

Results from the first desired power level (0.85) match exactly those from Stroup (1999) -- the only sample-size used in that paper for the CRD example. Since we had set the initial sample-size (per group) = 3, and the resulting power was calculated at 0.88, obviously no increase in the sample-size was made.

Results from the second desired power level (0.90) show that -- in order to reach or exceed that level -- an additional replication per group was needed, i.e. sample-size per group needed to increase to four (4). In a similar fashion, for the final desired power level (0.99), sample-size per group needed to be five (5).

Example 2: Mixed Model

The next example uses mouse data from Littell (2006). In the original data, there are four Conditions and three Diets for these mice, with weight gain as the response variable. For our use in this program, our “Block” variable will be used for Condition and our “Group” variable will be used for Diet. These data had 36 subjects, divided into six cages.

In Littell (2006, p. 480), the following PROC MIXED code was used on the actual data:

```
proc mixed data = mice nobound cl;
  class cage condition diet;
  model gain= condition|diet / solution ddfm=satterth;
  random cage cage*condition;
  ods output tests3=t3;
  run;
```

After running the data using the above PROC MIXED statements, the results for the covariance parameter estimates and the Type 3 tests of fixed effects are shown in the tables below:

Type 3 Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
diet	2	16	0.82	0.4561
condition	3	4.72	5.16	0.0589
diet*condition	6	16	1.52	0.2334

Covariance Parameter Estimates	
Cov Parm	**Estimate**
cage	5.0292
cage*condition	-6.2415
Residual	31.5567

Our program does not, though, use actual data. Therefore, we first calculated all Diet*Condition (Group*Block) means from the original data, in order to run the program. The entries were as follows (Step 2 in the program):

```
%let G1B1M = 55.333;     %let G2B1M = 55.333;      %let G3B1M = 57.333;
%let G1B2M = 54.667;     %let G2B2M = 52;          %let G3B2M = 53.333;
%let G1B3M = 61.667;     %let G2B3M = 61.333;      %let G3B3M = 55.667;
%let G1B4M = 60;         %let G2B4M = 53.333;      %let G3B4M = 66.333;
```

Going back a moment to Step 1, we set our initial sample-size as three (3) per Group*Block combination, which would give us a total sample-size of 36 -- equal to the original data. One purpose of this was to see how close the results for the Type 3 tests would be, compared to using the original data.

For Step 4 in our program, the input was:

```
%let ProcMixedCode =
    %str(model Resp=Group|Block  / solution intercept ddf = 23.5, 16, 4.72, 16;
          random rep rep*block;
          parms (5.0292) (-6.2416) (31.5567)  / noiter;

           contrast 'Group: 1 - 2' Group 1 -1 0;
           estimate 'Group: 1 - 2' Group 1 -1 0;
           contrast 'Block: 4 - 1' Block 1 0 0 -1;
           estimate 'Block: 4 - 1' Block 1 0 0 -1;);
```

We used the DDF= option in the model statement, in order to duplicate the degrees-of-freedom for the intercept and the fixed effects. (This was deleted for actual runs of the program to obtain sample-size estimates.) The results for the Type 3 tests were very close to those from the original data for the original run of 36 subjects:

Type 3 Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Group	2	16	0.82	0.4561
Block	3	4.72	6.82	0.0357
Group*Block	6	16	1.52	0.2334

(Note: for our example, we are using “Rep” in the RANDOM statement in place of “Cage”. In the original data, the 36 mice were in six cages. In the above example, there are only three levels of “Rep”.)

For arbitrary reasons only, we chose the contrasts listed in the above statements. We first ran them with a desired power level = 0.80. The results are shown on the next page.

FINAL RESULTS
Power Analysis & Effect Size
Alpha = 0.05; Desired Power = 0.8
Groups = 3; Blocks = 4

Label	Estimate	Standard Error	t Value	Num DF	Den DF	F Value	Pr > F	Alpha	Power	Effect Size
Group: 1-2	2.4170	1.1467	2.11	1	88	4.44	0.0379	0.05	0.5497	0.43026
Block: 4-1	3.8890	1.3241	2.94	1	33	8.63	0.0060	0.05	0.8135	0.59955

Sample Size, By Group

Group	Block	Group_n
G1	B1	12
G1	B2	12
G1	B3	12
G1	B4	12
G2	B1	12
G2	B2	12
G2	B3	12
G2	B4	12
G3	B1	12
G3	B2	12
G3	B3	12
G3	B4	12

We then ran the program with the “Group” contrast only, to see what level of sample-size was required to reach power = 0.80. The results are below:

FINAL RESULTS
Power Analysis & Effect Size
Alpha = 0.05; Desired Power = 0.8
Groups = 3; Blocks = 4

Label	Estimate	Standard Error	t Value	Num DF	Den DF	F Value	Pr > F	Alpha	Power	Effect Size
Group: 1-2	2.4170	0.8469	2.85	1	168	8.15	0.0049	0.05	0.8099	0.43026

Sample Size, By Group

Group	Block	Group_n
G1	B1	22
G1	B2	22
G1	B3	22
G1	B4	22
G2	B1	22
G2	B2	22
G2	B3	22
G2	B4	22
G3	B1	22
G3	B2	22
G3	B3	22
G3	B4	22

Finally, out of curiosity, we wanted to see what level of power the “Block” contrast would have if we set the sample-size (by Group*Block) to 22. This was done by setting initial sample-size in Step 1 to 22.

FINAL RESULTS
Power Analysis & Effect Size
Alpha = 0.05; Desired Power = 0.7
Groups = 3; Blocks = 4

Label	Estimate	Standard Error	t Value	Num DF	Den DF	F Value	Pr > F	Alpha	Power	Effect Size
Group: 1-2	2.4170	0.8469	2.85	1	168	8.15	0.0049	0.05	0.8099	0.43026
Block: 4-1	3.8890	0.9779	3.98	1	63	15.82	0.0002	0.05	0.9747	0.59955

Sample Size, By Group

Group	Block	Group_n
G1	B1	22
G1	B2	22
G1	B3	22
G1	B4	22
G2	B1	22
G2	B2	22
G2	B3	22
G2	B4	22
G3	B1	22
G3	B2	22
G3	B3	22
G3	B4	22

Example 3: Random Slopes Model

Our third and final example is based on pilot data that analyzes lung function over time in children with Congenital Muscular Dystrophy. The lung function measurement is forced expiratory volume over one second percent predicted, which we abbreviate as FEV_1%. A small cohort of children was measured over time to see how much a certain treatment would improve lung function. Values for FEV_1% were modeled over time -- using a random slopes model -- to obtain variance parameter estimates. Using these estimates then, we wanted to determine what sample-sizes were necessary to detect differences in annual FEV_1% slope of 1, 2 and 3 at different levels of power.

For these calculations, our “Group” variable represented literally four different Groups, each of which would have a different annual slope. “Block” became a *continuous* time variable, which means we had to change some of the internal code to reflect that (in Sections A and C). Our cell means were as follows:

```
%let G1B1M = 80;  %let G2B1M = 80;    %let G3B1M = 80;    %let G4B1M = 80;
%let G1B2M = 79;  %let G2B2M = 78;    %let G3B2M = 77;    %let G4B2M = 76;
%let G1B3M = 78;  %let G2B3M = 76;    %let G3B3M = 74;    %let G4B3M = 72;
%let G1B4M = 77;  %let G2B4M = 74;    %let G3B4M = 71;    %let G4B4M = 68;
```

We let B1 through B4 be four different years. Thus, G1 would have an annual slope of -1, G2 an annual slope of -2, G3 an annual slope of -3 and G4 an annual slope of -4. As one can see, we can then set up contrasts for differences of 1, 2 or 3 in annual slope. While this may seem like only a few values per group, that is sufficient since we are using fixed variance / covariance values estimated from the actual pilot data.

We ran the program three times (one for each contrast of interest), at desired power levels of 0.8, 0.85 and 0.9. One example of the PROC MIXED code from Step 4 is:

```
%let ProcMixedCode =
  %str(model Resp=Block(Group)  / solution;
       random Block / subject = rep(Group) type = un;
       parms (0.9148) (46.2685)  / noiter;
       contrast 'Slope Diff = 1' Block(Group) 1 0 0 -1;
       estimate 'Slope Diff = 1' Block(Group) 1 0 0 -1; );
```

Because of the length, we chose not to include the actual output here, but rather summarize the 9 different results in the table below.

Comparison	Estimate (Difference)	Standard Error	t Value	F Value	Pr > F	Alpha	Power	Number of Subjects per Group
Group1 - Group 2	1.00	0.3549	2.82	7.94	0.0052	0.05	0.8016	67
Group1 - Group 2	1.00	0.3311	3.02	9.12	0.0027	0.05	0.8534	77
Group1 - Group 2	1.00	0.3062	3.27	10.66	0.0012	0.05	0.9027	90
Group1 - Group 3	2.00	0.6857	2.92	8.53	0.0047	0.05	0.8211	18
Group1 - Group 3	2.00	0.6496	3.08	9.48	0.0029	0.05	0.8599	20
Group1 - Group 3	2.00	0.6057	3.30	10.90	0.0014	0.05	0.9041	23
Group1 - Group 4	3.00	1.0271	2.92	8.53	0.0068	0.05	0.805	8
Group1 - Group 4	3.00	0.9684	3.10	9.60	0.0040	0.05	0.8517	9
Group1 - Group 4	3.00	0.8759	3.43	11.73	0.0014	0.05	0.9165	11

Conclusion

Several authors have written about estimating the power of an F-test through use of the non-centrality parameter of a non-central F-distribution. As mentioned extensively, Stroup (1999) and Littell (2006) proposed SAS® code to efficiently obtain this power (with a fixed sample-size) from general and mixed linear models using PROC MIXED. Our purpose here was to build on the work already done and create a user-friendly program to address the "reverse" issue of obtaining the sample-size needed for a desired level of power in linear contrasts.

Including the example used in the code given in the Appendix, we have demonstrated the use of this program through four different examples: completely randomized design, mixed effects, random slopes and repeated measures. It is our hope that interested users can apply and modify these examples for a wide range of use in their own work.

Future Work

1) Create the capacity to have the user specify how many levels in each "Group" or "Block" are needed, without limiting the number of levels to four per variable.
2) Generalize the program so that sample-size can be calculated for a desired level of power for main effects, not just contrasts.
3) Build in the ability for different sample-sizes for different levels of "Group" or "Block". For example, if Group1 is always to have twice as many subjects as Group2, then the program would make that incremental increase until the desired power is reached.

References

Littell, Ramon C. et al. (2006). *SAS for Mixed Models, Second Edition*. Cary, NC: SAS Institute Inc. (pp. 479-495)

Stroup, Walter W. (1999), "Mixed Model Procedures to Assess Power, Precision, and Sample-size in the Design of Experiments," 1999 Proceedings of the Biopharmaceutical Section, Alexandria, VA: American Statistical Association. (pp. 15-24).

Rosa, Guilherme J.M., et al. (2006), "Reassessing Design and Analysis of Two-Colour Microarray Experiments Using Mixed Effects Models." *Comp. Funct. Genom.* 6:123-131.

Guenther, William C. (1979), “The Use of Noncentral F Approximations for Calculation of Power and Sample Size.” *The American Statistician*. 33(4):209-210.

"FAQ: How is effect size used in power analysis?" UCLA: Acad. Tech. Serv., Stat. Consulting Group. http://www.ats.ucla.edu/stat/mult_pkg/faq/general/effect_size_power/effect_size_power.htm (accessed June 15, 2011).

Liu, Xiaofeng and Raudenbush, Stephen (2004), "A Note on the Noncentrality Parameter and Effect Size Estimates for the F Test in ANOVA," *Journal of Educational and Behavioral Statistics.* 29(2):251-255

SAS/STAT® 9.2 User’s Guide (The MIXED Procedure), Second Edition. Copyright 2009 SAS Institute Inc.

Contact Information

Comments and questions are welcome. The author may be contacted at:

Matthew Fenchel
Division of Biostatistics and Epidemiology
Cincinnati Children’s Hospital Medical Center
MLC 5041, 3333 Burnet Avenue
Cincinnati, OH 45229-3039
E-mail: Matthew.Fenchel@cchmc.org

Appendix: Complete Program

```
/*******************************************************************************************
PROJECT:              PROGRAM FOR CALCULATING MODEL-BASED SAMPLE SIZE FOR A TWO GROUP CONTRAST
REVISION DATE:        08-25-2011
STATISTICIAN:         Matthew Fenchel, M.S.


BACKGROUND: An approximation of the power of an F-test can be estimated, using the non-centrality
parameter calculated from the observed F-statistic. This has been well described by Littell
(2006), Stroup (1999) and others. Holding a given statistical model fixed, the relationship
between power and sample size is one-to-one. Based on this, the "noncentral-F" approach to
calculating power (using a fixed sample size) can be "reversed" and used to calculate the sample
size needed to obtain a certain (fixed) power.

PURPOSE: The purpose of this program is to calculate sample size for up to five (5) different
levels of power, using the noncentrality parameter calculated from the observed F-statistic of a
user-defined statistical model / design. The program is designed for a two group comparison --
either two independent groups, or two time-points with non-independent observations.

CONCEPT: The underlying concept behind this program is described in detail by Littell (2006) and
Stroup (1999). Essentially, the user implements the same basic statistical model/design that he
anticipates using for analysis. With hypothesized estimates of the means (perhaps from previous
studies or pilot data), a data set is created with replicates of mu taking the place of actual
data values.  With hypothesized estimates of the variance parameters, using a variance/covariance
structure of the user's choice, the model is run on the "dummy" data set, holding the variance
parameters fixed. Key statistics are then generated that are used for sample size, power and
effect size.

DESIGN:  This program can be run "as is" with no initial input from the user.  It calculates
power for a "Pre-Post" comparison and for a "Case-Control" comparison.  This will give the user
an idea of the output produced, etc.

There are two categorical predictors available -- "Group" and "Block" -- and either of them can
be used in the MODEL, REPEATED or RANDOM statements.  Both have the possibility of having up to
four (4) levels.

The continuous response variable is "Resp".  The subject variable is "Rep".

USER INPUT:  There are five (5) steps at the beginning of the program where the user must input
information.  In brief, they are as follows:

STEP 1:  Input an initial sample size for the exemplary data set, desired power levels and alpha
level;
STEP 2:  Input hypothesized means for the levels and combinations of "Group" and "Block".
STEP 3: Specify whether preliminary output is desired.
STEP 4:  Specify details of the PROC MIXED model, including covariance structure and parameters.
STEP 5:  Specify path and name of output file, containing the final results.

OUTPUT: The final results are uploaded to a .rtf file using ODS OUTPUT.  The final results
include the sample size needed for each group to reach the power specified.  Other key statistics
used in the computations are also included in this file.

IMPORTANT NOTES:

a)  All data used in this program is generated by the program.  No actual data sets are used.
b)  Effect size (f) is computed from the noncentrality parameter (nc) of the F distribution,
using the following formula:

                      f = sqrt(nc/N)

where N is the total sample size (n1 + n2) of the two "Groups", "Blocks" or combinations being
compared.  See "FAQ" (2011) and Liu (2004).

c)  The program assumes a balanced (n1 = n2) design between "Groups" or "Blocks".
d)  The program moves onto the next power level -- or finishes -- when any of the specified
contrasts reaches the specified power level.  If you specify more than one contrast (which is
possible), then the other contrasts will have a lower power level than the contrast that reached
the desired level.
```

REFERENCES

Littell, Ramon C. et al. (2006). SAS for Mixed Models, Second Edition. Cary, NC: SAS Institute Inc. (479-495)

Stroup, Walter W. (1999), "Mixed Model Procedures to Assess Power, Precision, and Sample Size in the Design of Experiments," 1999 Proceedings of the Biopharmaceutical Section, Alexandria, VA: American Statistical Association. (15-24).

"FAQ: How is effect size used in power analysis?" UCLA: Academic Technology Services, Statistical Consulting Group. From

http://www.ats.ucla.edu/stat/mult_pkg/faq/general/effect_size_power/effect_size_power.htm (accessed June 15, 2011).

Liu, Xiaofeng and Raudenbush, Stephen (2004), "A Note on the Noncentrality Parameter and Effect Size Estimates for the F Test in ANOVA," Journal of Educational and Behavioral Statistics. 29(2):251-255


```
*********************************************************************************************/

/********************************************************************************************

                              STEPS TO BE COMPLETED BY USER

*********************************************************************************************/


/*      STEP 1:
        a)  Set initial sample-size -- nn -- by Group or Block.  Should be lower than needed to
               obtain the lowest desired power.
        b)  Set the desired alpha-level -- alpha.
        c)  Set the desired power -- dp1 through dp5 -- for which sample-size should be computed.
               If less than five power levels are desired, then set the remaining = 0.    */

        %let nn = 10;           %let alpha = 0.05;

        %let dp1 = 0.8;
        %let dp2 = 0.85;
        %let dp3 = 0.87;
        %let dp4 = 0.90;
        %let dp5 = 0;


/*      STEP 2: Insert means for the response variable, by the given Group(G) by Block(B)
        combinations (cells). If you only have one predictor variable, you either use G1B1, G2B1,
        G3B1 -OR- G1B1, G1B2, G1B3.                                                         */

        %let G1B1M = 63;      %let G2B1M = 45;      %let G3B1M = .;       %let G4B1M = .;
        %let G1B2M = 86.5;    %let G2B2M = 64;      %let G3B2M = .;       %let G4B2M = .;
        %let G1B3M = .;       %let G2B3M = .;       %let G3B3M = .;       %let G4B3M = .;
        %let G1B4M = .;       %let G2B4M = .;       %let G3B4M = .;       %let G4B4M = .;


/*      STEP 3: If all preliminary output is desired, then set Outp = Yes.  Preliminary output
        will not be saved in the final output file, though. */

        %let Outp = No;


/*      STEP 4:  Specify the details of the PROC MIXED code.  Some things that may need to be
        specified, depending upon the model you are using.  The model you specify will typically
        be the model you plan to use for analyses.

        a)  If your comparison is between two independent groups, then remove or change the
               REPEATED and/or TYPE statements.
        b)  You will need to change the values in the PARMS statement.
        c)  You may want to change the CONTRAST and ESTIMATE statements, depending upon your
               comparisons of interest.
        d)  Either "Group" or "Block" may be used in a RANDOM statement as well.
```

```
       NOTE:  "Resp" must be the response variable.  "Group" and "Block" are the possible
              predictor or random variables. Identical "Contrast" and "Estimate" statements must
              be in the model for the program to run.            */

              %let ProcMixedCode =
                     %str(
                            model Resp=Group*Block  / solution noint;
                            repeated Block / subject = rep(Group) type = csh;
                            parms (547.56) (2180.89) (.5)  / noiter;

                            contrast 'G1: B2 - B1' Group*Block -1 1 0 0;
                            estimate 'G1: B2 - B1' Group*Block -1 1 0 0;
                            contrast 'B2: G1 - G2' Group*Block 0 1 0 -1;
                            estimate 'B2: G1 - G2' Group*Block 0 1 0 -1;
                            );


/*  STEP 5: Completely specify the path and name of the file for the final results.  */

%let outfile = N:\Research\VanDyke\Shared_VanDyke\MWSUG 2011\MWSUG 2011 Power NonCenF OUTPUT.doc;


/*************************************************************************************************

       START OF PROGRAMMING TO GENERATE SAMPLE SIZES.  NOTHING MORE REQUIRED OF USER.

*************************************************************************************************/

       options nodate nonumber formdlim = "*";


/**************  SECTION A  *****************/

       data temp;                          /* Initial data set with Group and Time variables. */
              input Group $ Block $ @@;
              informat Group $2. Block $2.;
              datalines;
G1 B1 G2 B1 G3 B1 G4 B1 G1 B2 G2 B2 G3 B2 G4 B2 G1 B3 G2 B3 G3 B3 G4 B3 G1 B4 G2 B4 G3 B4 G4 B4
              ; run;


       %macro power;

       %do j = 1 %to 5;                              /*  Loop for each level of desired power. */

       %if &&dp&j = 0 %then %goto exit1;     /*  Goes to end of program if desired power = 0. */


/*     Programming below is setting initial values for data sets and macro variables.  */

       data temp2; informat Group $2.; informat Block $2.; run;

       %let power = 0;               %let n = &nn;


/*     The programming below creates the datasets used in the PROC MIXED calculations.  The size
       of these datasets increase until the desired power is reached.  */

       %do %until (&power ge &&dp&j);

       %do i = 1 %to &n;
              data temp2;
              set temp2 temp;
              if Group = "" then delete;
              if Group = "G1" and Block = "B1" then Resp = &G1B1M;
              if Group = "G2" and Block = "B1" then Resp = &G2B1M;
              if Group = "G3" and Block = "B1" then Resp = &G3B1M;
              if Group = "G4" and Block = "B1" then Resp = &G4B1M;
              if Group = "G1" and Block = "B2" then Resp = &G1B2M;
              if Group = "G2" and Block = "B2" then Resp = &G2B2M;
              if Group = "G3" and Block = "B2" then Resp = &G3B2M;
              if Group = "G4" and Block = "B2" then Resp = &G4B2M;
              if Group = "G1" and Block = "B3" then Resp = &G1B3M;
              if Group = "G2" and Block = "B3" then Resp = &G2B3M;
              if Group = "G3" and Block = "B3" then Resp = &G3B3M;
```

```
            if Group = "G4" and Block = "B3" then Resp = &G4B3M;
            if Group = "G1" and Block = "B4" then Resp = &G1B4M;
            if Group = "G2" and Block = "B4" then Resp = &G2B4M;
            if Group = "G3" and Block = "B4" then Resp = &G3B4M;
            if Group = "G4" and Block = "B4" then Resp = &G4B4M;
            if Resp = . then delete;
            run;
            %end;

        proc sort data=temp2 out=test1 nodupkey; /*  Obtain count of Group*Block combinations. */
            by Group Block;
            run;

        data test1b; set test1;
            call symput("TC", _N_);
            run;

        data temp2; set temp2;               /*  Assigning values of "Rep".  */
            Rep = int((_N_+&TC-1)/&TC);
            run;


/*********  SECTION B  ***************/

/*      The procedures below will eventually output the final sample size that was needed for the
        desired power. */

        proc freq data=temp2 noprint;
            tables Group*Block / out=count&j cumcol;
            run;

        data count2_&j; set count&j;
            Group_n = COUNT;
            call symput("CellCount", Count);
            drop PERCENT COUNT;
            run;


/*      The PROC MIXED code below generates the necessary statistics to calculate power and
        sample size.  */

        %if &Outp = No %then %do; ods select none; %end;

        proc mixed data = temp2 order=data noprofile;
            class Group Block Rep;
            &ProcMixedCode
            ods output Contrasts = Cont Estimates = Est;
            run;


/*      The programming below obtains the necessary information from the Proc Mixed output, to
        calculate the critical value of the non-central F distribution and power.    */

        data power; set Cont;
            Alpha=&alpha;
            nc=numdf*fvalue;
            fcrit = finv(1-Alpha, numdf, dendf, 0);
            Power=round((1-probf(fcrit, numdf, dendf, nc)), .0001);
            Order = _N_;
            run;

        proc sort data=Est; by Label; run;

        proc sort data=power; by Label; run;


/************  SECTION C  ********************/

/*  The code below combines the estimate calculations with the power calculations.  */

        data final&j;
            merge Est power;
            by Label;
            run;
```

```
        proc sort data=final&j; by Power; run;

        data final&j; set final&j;
                EffSize = (nc/(&CellCount*2))**0.5;
                label EffSize = "Effect Size";

                PowerDiff = &&dp&j - Power;
                if PowerDiff ge 0.2 then n = 4;
                else if PowerDiff ge 0.1 then n = 2;
                else if PowerDiff < 0.1 and power < &&dp&j then n = 1;
                call symput('n', n);

                drop probt nc fcrit df order n PowerDiff;
                run;


/*  Using Contrast with highest power for comparison against desired power.  */

        proc means data=final&j noprint;
                var power;
                output out=maxpower max = ;
                run;

        data maxpower; set maxpower;
                call symput('power',power);
                run;

        %put test=&n; %put temp=&power;

/*  Preliminary output. */

        proc print data=final&j noobs; run;

        proc print data=count2_&j; run;

        %end;                         /*  End of one loop through desired power levels.   */
        %exit1: %end;                 /*  End of all loops through desired power levels.  */


/************  SECTION D  ********************/

        ods select all;

        proc sort data=temp2 out=temp3 nodupkey;  /*  Counts of Group for output titles.  */
                by Group;
                run;

        data temp3b; set temp3;
                n = trim(left(put(_N_, 1.)));
                call symput("GCount", n);
                run;

        proc sort data=temp2 out=temp4 nodupkey;  /*  Counts of Block for output titles.  */
                by Block;
                run;

        data temp4b; set temp4;
                n = trim(left(put(_N_, 1.)));
                call symput("BCount", n);
                run;


        /* Final output to a file.  */

        ods rtf file = "&outfile" bodytitle style=sansprinter startpage=no;

        %do j = 1 %to 5;
                %if &&dp&j = 0 %then %goto exit2;

        title1 "FINAL RESULTS";
        title2 "Power Analysis & Effect Size";
        title3 "Alpha = &alpha;  Desired Power = &&dp&j";
        title4 "Groups = &GCount; Blocks = &BCount";

        proc print data=final&j noobs label; run;
```

```
title1 "Sample Size, By Group"; title2; title3; title4;

proc print data=count2_&j noobs; run;

%exit2: %end;

ods rtf close;

%mend;
%power;


/*  END OF PROGRAM  */
```