Paper 240-26

Power and Sample Size Determination for Linear Models

John M. Castelloe, SAS Institute Inc., Cary, NC
Ralph G. O'Brien, Cleveland Clinic Foundation, Cleveland, OH

Abstract

This presentation describes the steps involved in performing sample size analyses for a variety of linear models, both univariate and multivariate. As an analyst you must gather and synthesize the information needed, but you should be able to rely on the analytical tools to accommodate the numerous ways in which you can characterize and solve problems. Examples illustrate these principles and review relevant methods. User-written, SAS® software-based programs already handle a wide variety of problems in linear models. Now, SAS Institute itself is developing software that will handle a rich array of sample size analyses, including all those discussed in this paper.

Introduction

Power and sample size computations for linear models present a level of complexity greater than that required for simple hypothesis tests. A number of steps are involved to gather the required information to perform these computations. After settling on a clear research question, the analyst must (1) define the *study design*, (2) posit a *scenario model*, a mathematical model proposing a general explanation for the nature of the data to be collected, and (3) make specific conjectures about the parameters of that model, the magnitudes of the *effects and variability*. Because developing the scenario model is typically a technically difficult and subjective process, various strategies and simplifying formulas exist to make matters more feasible, and software for sample size analysis should exploit them. Once the scenario modeling is done, you must still (4) delineate the primary *statistical methods* that will best address the research question. Finally, the (5) *aim of assessment* must be clearly expressed to ensure that the power and sample size computations accomplish the intended goal in study planning. In hypothesis testing, you typically want to compute the powers for a range of sample sizes or vice-versa. All of this work has strong parallels to ordinary data analysis. The section "Components of a Sample Size Analysis" explains these steps in more detail.

User-developed SAS-based applications, such as UnifyPow (O'Brien 1998) and the SAS/IML® program of Keyes and Muller (1992), already handle a wide variety of problems in linear models. SAS Institute is developing new software for power and sample size analyses to cover the methods discussed in this paper, along with a variety of other models discussed in Castelloe (2000).

This paper describes different strategies for power and sample size analysis for linear models in a series of examples, starting with the *t*-test and progressing through one-way analysis of variance (ANOVA), multiple regression, and multi-way ANOVA. In each example, you will first learn about the specific ingredients required for the power or sample size computation for the linear model being considered. Then the example will proceed to illustrate the implementation of a power or sample size analysis following the five-component strategy. Later sections describe unified approaches for multivariate models with fixed effects and suggest guidelines for extensions such as multiple comparisons, mixed models, and retrospective analyses.

A Review of Power Concepts

Before explaining more about the five components of a sample size analysis and proceeding through examples in linear models, a brief review of terminology used in power and sample size analysis is in order. Refer to Castelloe (2000) for a more thorough treatment of these concepts.

In statistical hypothesis testing, you typically express the belief that some effect exists in a population by specifying an alternative hypothesis H_1. You state a null hypothesis H_0 as the assertion that the effect does *not* exist and attempt to gather evidence to reject H_0 in favor of H_1. Evidence is gathered in the form of sample data, and a statistical test is used to assess H_0. If H_0 is rejected but there really is *no* effect, this is called a *Type I error*. The probability of a Type I error is usually designated "alpha" or α, and statistical tests are designed to ensure that α is suit-

ably small (for example, less than 0.05).

If there really is an effect in the population but H_0 is *not* rejected in the statistical test, then a *Type II error* has been made. The probability of a Type II error is usually designated "beta" or β. The probability $1 - \beta$ of avoiding a Type II error, that is, correctly rejecting H_0 and achieving statistical significance, is called the *power*. An important goal in study planning is to ensure an acceptably high level of power. Sample size plays a prominent role in power computations because the focus is often on determining a sufficient sample size to achieve a certain power, or assessing the power for a range of different sample sizes. Because of this, terms like *power analysis*, *sample size analysis*, and *power computations* are often used interchangeably to refer to the investigation of relationships among power, sample size, and other factors involved in study planning.

Components of a Sample Size Analysis

Even when the research questions and study design seem straightforward, the ensuing sample size analysis can seem technically daunting. It is often helpful to break the process down into five components:

Study Design: What is the structure of the planned design? This must be clearly and completely specified. What groups and treatments ("cells" and "factors" of the design) are going to be assessed, and what will be the relative sizes of those cells? How is each case going to be studied, i.e., what are the primary outcome measures ("dependent variables"), and when will they be measured? Will covariates be measured and included in the statistical model?

Scenario Model: What are your beliefs about patterns in the data? Imagine that you had unlimited time and resources to execute the study design, so that you could gather an "infinite data set." Characterize that infinite data set as best you can using a mathematical model, realizing that it will be a simplification of reality. Alternatively, you may decide to construct an "exemplary" data set that mimics the infinite data set. However you do this, your scenario model should capture the key features of the study design and the main relationships among the primary outcome variables and study factors.

Effects & Variability: What exactly are the "signals and noises" in the patterns you suspect? Set specific values for the parameters of your scenario model, keeping at most one unspecified. It is often enlightening to consider a variety of realistic possibilities for the key values by performing a sensitivity analysis. Alternatively, construct two or three exemplary data sets that capture the competing views on what the infinite data set might look like. For linear models and their extensions an important component is the "residual" term that captures unexplained variation. The standard deviation (SD) of this term plays a critical role in sample size analysis. What is this value? A sensitivity analysis is usually called for in positing SD.

Statistical Method: How will you cast your model in statistical terms and conduct the eventual data analysis? Define the statistical models and procedures that will be used to embody the study design and estimate/test the effects central to the research question. What tests will be done? What significance levels will be used? Will one- or two-tailed tests be used?

Aim of Assessment: Finally, what needs to be determined in the sample size analysis? Most often you want to examine the statistical powers obtained across the various scenarios for the effects, the statistical procedures (tests) to be used, and the feasible total sample sizes. Some analysts find sample size values that provide given levels of power, say 80%, 90%, or 95%. Other analysts compute the value for some key effect parameter (e.g., a given treatment mean) that will provide a given level of power at a given sample size. You might even want to find the α-level that will provide a given power at a given sample size for a given effect scenario.

The following examples illustrate how these components can provide a guiding structure to facilitate more rigorous planning of studies involving linear models.

Preview of Examples

The examples are organized by different types of linear models. The main distinction is in the type of predictors (or independent variables) in the model. All examples are univariate, involving only one continuous response variable. The first two examples are simple cases involving only categorical predictors, the *t*-test and one-way ANOVA. Multiple regression involves predictors treated as continuous variables, although some may be dummy (0/1) variables representing categories. The multi-way ANOVA with covariates contains categorical predictors of interest *and* additional continuous predictors used as covariates to reduce excess variability. Finally, power computations for the multivariate general linear model (GLM) are discussed, without a detailed example, and guidelines are given for some common linear models extensions not covered by examples.

The format of each example is as follows. First the type of linear model is briefly explained, and the problem situation of the study planner is revealed. The five-component strategy explained in the "Components of a Sample Size Analysis" section is applied

to solve the problem, and then the power computation details are explained, including relevant equations and references. The equations are used to solve the problem at hand, but they are sufficiently general for the variety of different possible goals (e.g., solving for power, sample size, etc.).

Ordinary Two-Sample *t*-test

The two-sample (pooled) *t*-test is equivalent to an ANOVA with two groups and thus is a special case of a linear model. Although power computations for *t*-tests are widely understood and implemented, a characterization following the five-component strategy provides a useful framework for more complicated examples. In addition, the last part of the example illustrates the consideration of unbalanced designs and costs.

Suppose an industrial chemist is researching whether her firm should switch to a new grade of ammonium chloride when producing an organic compound, SHHS-01. A more expensive, finely ground grade is touted to give a higher yield than the standard coarse grade, but this needs to be tested. A 5% increase in yield would offset the extra cost. The chemist's goal is to determine an appropriate sample size to have adequate power in a comparison of the two grades using a *t*-test.

Study Design: Under laboratory conditions designed to mimic production conditions, equal numbers of mini-batches of SHHS-01 will be made using either the coarse or fine grade of ammonium chloride. The primary outcome measure will be the yield, measured simply as weight in grams.

Scenario Model: The conjectured "infinite data set" has two groups of yields, each containing independent and normally distributed data. The mean for the fine group exceeds the mean for the coarse group. There is no reason to suspect that the variability differs between groups.

Effects & Variability: Based on numerous previous laboratory studies with the coarse ammonium chloride, the chemist knows that this yield averages about 160g/batch. Because this is an organic process, there is significant variability as well, with a standard deviation of 20g. For the fine ammonium chloride, biological modeling predicts that the yield could be at least 10% greater, at 176g. The 20g standard deviation should apply. This scenario is illustrated in Figure 1.

Statistical Method: The statistical question is whether the fine grade ammonium chloride will produce at least 5% (8g) more SHHS-01 than the coarse grade (to offset the extra cost). This situation conforms to an ordinary, one-tailed *t*-test, with hypotheses $H_0 : \mu_2 - \mu_1 = \mu_{\text{diff}} < 8$ and $H_1 : \mu_2 - \mu_1 = \mu_{\text{diff}} \geq 8$. Making the mini-batches in the lab is inexpensive, but making a Type I error could lead to establishing a production process that is substantially more costly. Hence, alpha will be set at 0.001 or, at most, 0.005.

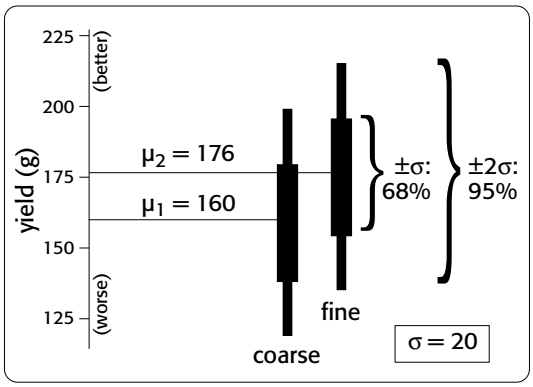

Figure 1. Conjectured Scenario for Coarse and Fine Grades

Aim of Assessment: If the fine grade is really 10% more effective, as believed, then it will save the company a lot of money. Thus, failing to discover this would be so costly that the chemist decides to produce enough mini-batches to achieve a statistical power of 99%. Remember, making and weighing the mini-batches is relatively inexpensive. So the chemist wants to determine the required sample size to provide 99% power.

This required sample size is determined by solving the following equation for n:

$$\text{power} = P\left(t(2n-2,\delta) \geq t_{1-\alpha;2n-2}\right)$$

where $t(u,\delta)$ is distributed as noncentral t with u d.f. and noncentrality $\delta = \sqrt{n/2}(\mu_{\text{diff}} - \mu_0)/\sigma$, and $t_{p;u}$ is the p^{th} quantile of the central t distribution with u d.f.

With 99% power, the required sample size per group is n = 303 for α = 0.005 and n = 370 for α = 0.001. Hence, using N = 740 total will achieve outstanding control of Type I and Type II errors. N = 606 is also feasible.

Unbalanced Designs and Cost Suppose that in the lab it costs 75% more to produce a mini-batch using the experimental fine grade of ammonium chloride. The chemist wonders if an unbalanced design with fewer fine mini-batches than coarse ones would produce as much power but at less cost. Consider a 3:2 sampling ratio, i.e., cell weights of w_1 = 3/5 and w_2 = 2/5.

Power computations can be performed in terms of group weights and total sample size:

$$\text{power} = P\left(t(N-2,\delta) \geq t_{1-\alpha;N-2}\right)$$

where $\delta = \sqrt{N w_1 w_2}(\mu_{\text{diff}} - \mu_0)/\sigma$.

To achieve 99% power using α = 0.001 requires 462 + 308 = 770 cases versus 370 + 370 = 740 cases for the balanced design. The unbalanced study design is 4.1% larger, but it would cost about 1.6% less to run. While the ordinary two-group *t*-test has optimum *statistical* efficiency with a balanced design, it can have sub-optimum *budgetary* efficiency if the cost per sampling unit differs between the groups.

Note that two-tailed versions of the above formulas are available, using the noncentral $F = t^2$ distribution (with noncentrality $\lambda = \delta^2$).

One-Way ANOVA with One-d.f. Contrast

This example extends the previous one by increasing the number of groups from two to three. An appropriate sample size will be determined for a comparison (represented this time as a linear contrast) between one group and the average of the other two.

Suppose the chemist introduced in the previous section implements the comparison between the fine and coarse grades of ammonium chloride and concludes that the fine grade is advantageous, giving a yield of about 176g as predicted. Plans to purchase the fine grade chemical from Producer A are interrupted when Producer B offers a package deal of "special-fine" bundled with "super-fine" in a 1:1 ratio, for about the same cost. Engineers at Producer B claim that the special-fine grade yields 172g of SHHS-01, while the super-fine yields 190g, with the same variability as the fine grade from Producer A (SD = 20g). The chemist is asked to compare the yield using Producer A's fine grade to the average yield of special fine and super fine (used separately) supplied by Producer B, with enough mini-batches to achieve 90% power if the engineers are correct.

Study Design: The chemist will conduct experiments using the three different grades of ammonium chloride, with the two varieties from Producer B used equally often and each twice as often as the Producer A variety. In other words, the weights are 2 special-fine : 2 super-fine : 1 fine (w_1 = 0.4, w_2 = 0.4, w_3 = 0.2). The dependent variable is the yield in grams.

Scenario Model: The chemist will assume that the chemicals from the two producers will produce mean yields as previously surmised for Producer A and claimed by engineers for Producer B.

Effects & Variability: The mean yields are conjectured to be μ_1=172g (special-fine), μ_2=190g (super-fine), and μ_3=176g (fine). The standard deviation is assumed to be about 20g for each grade.

Statistical Method: A 1-way ANOVA will be conducted to test a contrast of fine with average over special-fine and super-fine, using the usual F statistic with $\alpha = 0.05$. The contrast can be written as a CONTRAST statement for PROC GLM, for example, as

```
contrast grade -1 -1 2
```

Aim of Assessment: The chemist wishes to calculate the required sample size to achieve 90% power.

The required total sample size N can be calculated from the following equation:

$$\text{power} = P\left(F(1, N-G, \lambda) \geq F_{1-\alpha;1;N-G}\right)$$

where

$$\lambda = N \frac{\left(\sum_{i=1}^{G} c_i \mu_i - c_0\right)^2}{\sigma^2 \sum_{i=1}^{G} \frac{c_i^2}{w_i}}$$

and $\{c_i\}$ are the contrast coefficients. The required sample size is found to be 735 (rounded up from 732 to avoid fractional group sizes).

Multiple Linear Regression

Instead of categorical predictors as in the *t*-test and 1-way ANOVA, multiple regression involves continuous and dummy independent variables. Although as a special case multiple regression could be used with dummy variables to conduct an ANOVA, the intention here is to demonstrate the more typical usage focusing on tests of individual predictors controlling for other predictors. Such a test is planned in this example, and the goal is to compute the power.

One of the important considerations in multiple regression and correlation analysis is whether to treat the predictors as fixed or random. There are also many alternative ways to characterize the effects, using various forms of correlations and regression coefficients. The example uses fixed predictors and involves an effect specification in terms of partial correlation. Following the example is a discussion of the ramifications of the distinction between fixed and random predictors and a collection of equations showing the alternative ways to specify effects.

A team of preventive cardiologists is investigating whether elevated serum homocysteine levels are

linked to atherosclerosis (plaque build-up in coronary arteries). The analysis will use ordinary least squares regression to assess the relationship between total homocysteine level (tHcy) and a plaque burden index (PBI), adjusting for six other variables: age, gender, and plasma levels of folate, vitamins B_6 and B_{12}, and a serum cholesterol index. The group wonders whether 100 subjects will provide adequate statistical power.

Using the five components, the power analysis breaks down as follows:

Study Design: This is a correlational study at a single time. Subjects will be screened so that about half will have had a heart problem. All eight variables will be measured during one visit.

Scenario Model: Most clinicians are familiar with simple correlations between two variables, so the collaborating statistician decides to pose the statistical problem in terms of estimating and testing the partial correlation between X_1 = tHcy and Y = PBI, controlling for the six other predictor variables ($R_{YX_1|X_{-1}}$). This greatly simplifies matters, especially the elicitation of the conjectured effect.

The statistician uses partial regression plots like that shown in Figure 2 to teach the team that the partial correlation between PBI and tHcy is the correlation of two sets of residuals obtained from ordinary regression models, one from regressing PBI on the six covariates and the other from regressing tHcy on the same covariates. Thus each subject has "expected" tHcy and PBI values based on the six covariates. The cardiologists believe that subjects who are relatively higher than expected on tHcy will also be relatively higher than expected on PBI. The partial correlation quantifies that adjusted association just like a standard simple correlation does with the unadjusted linear association between two variables.

Effects & Variability: Based on previously published studies of various coronary risk factors and after viewing a set of scatterplots showing various correlations, the team surmises that the true partial correlation is likely to be at least 0.35.

Statistical Method: Regress PBI on tHcy and the six other predictors, plus the intercept. Use an ordinary *F* test to assess whether tHcy is a significant predictor in this model with seven predictors. The test presumes that the residuals come from a normal distribution.

Aim of Assessment: Compute the statistical powers associated with N = 80 and 100, using α = 0.05 and 0.01.

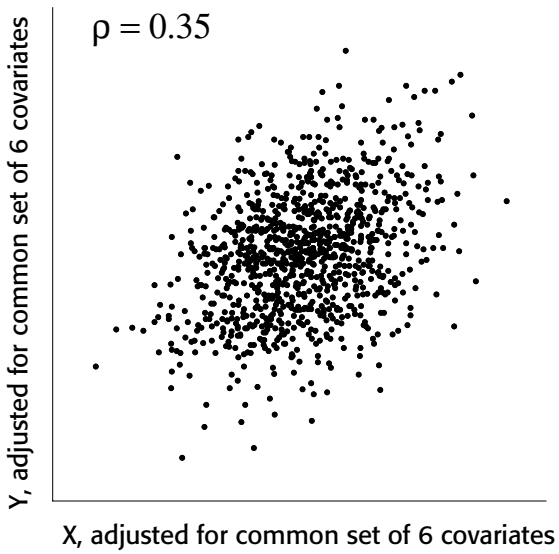

Figure 2. Partial Correlation Plot

The exact power can be computed from the equation

$$\text{power} = P\left(F(p_1, N-p-1, \lambda) \geq F_{1-\alpha;p_1;N-p-1}\right) \quad (1)$$

where p is the total number of predictors in the model (including the predictor of interest, but not the intercept), p_1 is the number of predictors being tested simultaneously (here, p_1 = 1), and

$$\lambda = N\frac{R^2_{YX_1|X_{-1}}}{1-R^2_{YX_1|X_{-1}}} \quad (2)$$

The calculated powers range from 75% (N = 80, α = 0.01) to 96% (N = 100, α = 0.05). The latter result is almost balanced with respect to Type I and Type II error rates. The study seems well designed at N = 100.

Fixed vs. Random Predictors The computations in the example assume a *conditional* model, as typically used in multiple linear regression. The predictors (represented collectively as X) are assumed to be fixed, and the responses Y are assumed to be independently normally distributed conditional on X. The usual test statistic considered is the Type III F test where the null hypothesis states that all coefficients of the p_1 predictors of interest are zero.

A related approach is the *unconditional* model, typically used in multiple correlation analysis, in which predictors are assumed to be random. The variables in Y and X are taken to have a joint multivariate normal distribution. Power computations differ for the conditional and unconditional models. Gatsonis and Sampson (1989) outline an exact power computa-

tion method for the unconditional model due to Lee (1972).

It is important to note, however, that the usual test statistics for conditional and unconditional models are equivalent, having exactly the same null distribution. "The conceptual difference between them is primarily one of interpretation and generalizability of the conclusions" (Gatsonis and Sampson 1989, p. 516). Thus the strategies for describing effects in each of the two approaches can be used interchangeably in sample size analysis. For example, the cardiologists conjectured effects in terms of partial correlation but planned to use multiple regression.

Alternative Effect Specifications The remainder of this section describes the various ways you can describe effects using different types of correlations and regression coefficients. You can use the same parameterizations for either conditional or unconditional models in power computations. The well-known method for the conditional framework is outlined explicitly here, and you can refer to Gatsonis and Sampson (1989) for analogous computations in the unconditional framework.

Consider the general situation in which you are interested in testing that the coefficients of $p_1 \geq 1$ predictors in a set X_j are zero, controlling for all of the other predictors X_{-j} (comprised of $p - p_1 \geq 0$ variables). For the conditional model, the power can be computed using equation (1), where the noncentrality λ is defined differently for various alternative specifications of the effects. You can choose whichever one is most convenient for expressing the conjectured effects in your situation.

One such specification involves the multiple partial correlation $R_{YX_j|X_{-j}}$:

$$\lambda = N\frac{R^2_{YX_j|X_{-j}}}{1 - R^2_{YX_j|X_{-j}}}$$

You can also express the effects in terms of the multiple correlations in full ($R_{Y|(X_j,X_{-j})}$) and reduced ($R_{Y|X_{-j}}$) nested models:

$$\lambda = N\frac{R^2_{Y|(X_j,X_{-j})} - R^2_{Y|X_{-j}}}{1 - R^2_{Y|(X_j,X_{-j})}} \quad (3)$$

The numerator of (3) is equivalent to the squared multiple semipartial correlation $R^2_{Y|(X_j|X_{-j})}$. Thus

$$\lambda = N\frac{R^2_{Y|(X_j|X_{-j})}}{1 - R^2_{Y|(X_j,X_{-j})}} \quad (4)$$

You may find it easier to work in terms of standard (zero-order) correlations, even though there are more parameter values to specify. A form of λ involving the correlations between Y and variables in $X = \{X_j, X_{-j}\}$ (labeled as vectors ρ_{XY} and $\rho_{X_{-j}Y}$), and between pairs of variables in X (labeled as correlation matrices S_{XX} and $S_{X_{-j}X_{-j}}$), is given by the following:

$$\lambda = N\frac{\rho'_{XY}S^{-1}_{XX}\rho_{XY} - \rho'_{X_{-j}Y}S^{-1}_{X_{-j}X_{-j}}\rho_{X_{-j}Y}}{1 - \rho'_{XY}S^{-1}_{XX}\rho_{XY}} \quad (5)$$

The remaining specifications apply only to cases in which X_j consists of a single predictor.

You can express λ in terms of the standardized regression coefficient ($\tilde{\beta}_j$) of X_j; the tolerance of X_j, computed as $1 - R^2_{X_j|X_{-j}}$ in a regression of X_j on the other predictors; and the multiple correlation $R_{Y|(X_j,X_{-j})}$ for the full model:

$$\lambda = N\frac{\tilde{\beta}^2_j(1 - R^2_{X_j|X_{-j}})}{1 - R^2_{Y|(X_j,X_{-j})}} \quad (6)$$

Or, you can posit the unstandardized coefficient β_j along with the tolerance of X_j, SD of X_j (σ_{X_j}), and SD of residual (σ):

$$\lambda = N\frac{\beta^2_j(1 - R^2_{X_j|X_{-j}})\sigma^2_{X_j}}{\sigma^2}$$

If an exchangeable correlation structure is deemed reasonable, equation (5) can be simplified to include only the common correlation between Y and each predictor (ρ_{XY}) and the common pairwise correlations between predictors (ρ_{XX}):

$$\lambda = N\frac{\rho^2_{XY}(1 - \rho_{XX})}{[1 + (p-1)\rho_{XX} - p\rho^2_{XY}][1 + (p-2)\rho_{XX}]}$$

A useful compromise between the exchangeable correlation structure and the necessity of specifying all correlations is a *relaxed* exchangeable correlation structure (Maxwell 2000), which allows different correlations ρ_{X_jY} between Y and X_j, and $\rho_{X_jX_{-j}}$ between X_j and the other predictors, in addition to common correlations $\rho_{X_{-j}Y}$ between Y and components of X_{-j}, and $\rho_{X_{-j}X_{-j}}$ between elements of X_{-j}:

$$\lambda = N\frac{R^2_{Y|(X_j,X_{-j})} - R^2_{Y|X_{-j}}}{1 - R^2_{Y|(X_j,X_{-j})}}$$

where

$$R^2_{Y|(X_j,X_{-j})} = \Big\{\rho^2_{X_jY}[1 + (p-2)\rho_{X_{-j}X_{-j}}] +$$

$$(p-1)\rho^2_{X_{-j}Y} - 2(p-1)\rho_{X_jY}\rho_{X_{-j}Y}\rho_{X_jX_{-j}}\Big\} \cdot$$
$$\Big\{1 - \rho_{X_{-j}X_{-j}} + (p-1)\rho_{X_{-j}X_{-j}} - \rho^2_{X_jX_{-j}}\Big\}^{-1}$$

and

$$R^2_{Y|X_{-j}} = \frac{(p-1)\rho^2_{X_{-j}Y}}{1 + (p-2)\rho_{X_{-j}X_{-j}}}$$

If you want to test contrasts of the regression coefficients, you can use the more general formulation discussed in the section "The Univariate GLM with Fixed Effects."

Multi-Way ANOVA with Fixed Effects and Covariates

This next example features an ANOVA model that is an extension of the kind of model considered in the section "One-Way ANOVA with One-d.f. Contrast" in the sense of having two factors (instead of one) and additionally a continuous covariate. The planned tests are also more complicated, involving several contrasts. The goal of the scientists in the example is to assess whether their largest possible sample size will provide adequate (possibly excessive) power for these tests.

The discussion in this section follows the five-component layout as used in the previous examples. Details regarding the mathematical power computations, and other alternative ways of describing the components, are covered in the following section, "The Univariate GLM with Fixed Effects."

Suppose a team of animal scientists hypothesizes that dietary supplements of the trace element sugianimum (fictitious) increase the growth rate in female newborn rabbits. Standard rabbit chow contains 5 ppm sugianimum. The scientists want to study four other supplemental formulations, +10, +20, +40, and +80 ppm (with the standard chow designated as +0). This will allow them to conduct an ANOVA to test models with thresholds, ceiling effects, and/or dose-response effects. They have sufficient facilities and funding to study at most 240 rabbits but would be pleased if fewer would seem to suffice.

Rabbits are eight to nine weeks old when they arrive from the commercial breeder, and their body weight is 1.5 $\pm$ 0.25 kg (mean $\pm$ SD). After eating only standard chow for the next 24 weeks, female rabbits of this breed have a body weight of about 4.2 kg $\pm$ 0.56 kg.

Study Design: The primary outcome measure will be each rabbit's body weight after 24 weeks on the study.

Sugianimum level is represented by a factor called Sugianimum Supplementation Level or SugiSupp, with five levels (+0, +10, +20, +40, and +80 ppm). The scientists will include rabbit feed from all five of the major U.S. manufacturers (Gamma, Epsilon, Zeta, Eta, and Theta) to enable greater generalizability of the results. Call this factor Company. Thus, if all manufacturers supplied all formulations, the design would be a 5 $\times$ 5 factorial, Company $\times$ SugiSupp.

Suppose each company produces only two formulations besides the standard one (+0 ppm), thus making a complete factorial design impossible. The scientists will use a randomized design with cell weights as displayed in Table 1. Thus, they are planning a 2:1:1 ratio for the standard and two supplemental formulations for each company.

		SugiSupp				
Company		+0	+10	+20	+40	+80
<1>	Gamma	2	1	1	0	0
<2>	Epsilon	2	1	0	1	0
<3>	Zeta	2	0	1	0	1
<4>	Eta	2	0	0	1	1
<5>	Theta	2	1	0	0	1

Table 1. Cell Weights for Design

Rabbits that are larger at baseline tend to gain more body weight during the study period. Because of this correlation, the rabbits' initial body weight RabbitWgt00 could serve as a useful covariate by accounting for extra variation in body weight at 24 weeks. So the scientists plan to include the measurement of RabbitWgt00 in the study protocol.

Scenario Model: The scientists envision two different scenarios for the means of body weights at 24 weeks across SugiSupp and Company. Both scenarios assert a monotonically increasing dose-response relationship until 40 ppm but a ceiling effect after that, and the average weight gains differ by company. "Scenario 1" conjectures that the pattern of sugianimum effects is the same across companies, i.e., the Company $\times$ SugiSupp interaction is null. "Scenario 2" involves essentially the same main effects but reflects the suspicion that Gamma's +10 formulation is less effective than its own +0, and Epsilon's +10 formulation is unusually effective compared to its own +0.

Effects & Variability: For scenario 1, the scientists conjecture means for the 5 $\times$ 5 factorial as shown in Table 2. The means are displayed graphically in Figure 3.

One can confirm that these means conform perfectly

to the main-effects-only linear model,

$$\mu_{\{\text{Company,SugiSupp}\}} = 4.2 + A_{\{\text{Company}\}} + B_{\{\text{SugiSupp}\}}$$

where $A_{\{1\}}$ = 0.0, $A_{\{2\}}$ = -0.2, $A_{\{3\}}$ = 0.2, $A_{\{4\}}$ = -0.1, $A_{\{5\}}$ = 0.1; and $B_{\{+0\}}$ = 0, $B_{\{+10\}}$ = 0.1, $B_{\{+20\}}$ = 0.4, $B_{\{+40\}}$ = 0.5, $B_{\{+80\}}$ = 0.5.

		SugiSupp				
Company		+0	+10	+20	+40	+80
<1>	Gamma	4.2	4.3	4.6	4.7	4.7
<2>	Epsilon	4.0	4.1	4.4	4.5	4.5
<3>	Zeta	4.4	4.5	4.8	4.9	4.9
<4>	Eta	4.1	4.2	4.5	4.6	4.6
<5>	Theta	4.3	4.4	4.7	4.8	4.8

Table 2. Conjectured Means for Scenario 1

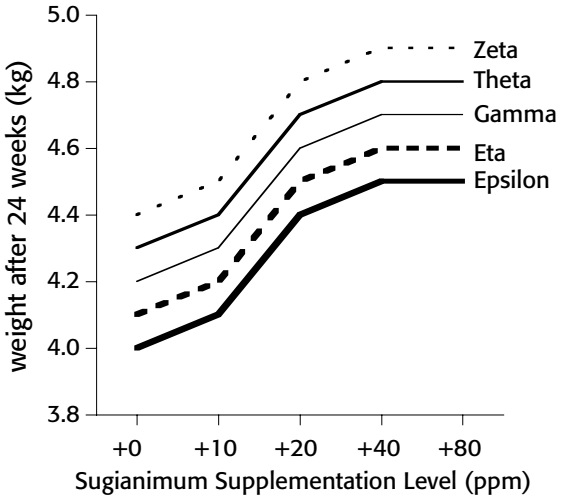

Figure 3. Conjectured Means for Scenario 1

In scenario 2, the scientists consider the same main effects but also a small interaction involving only the 2 $\times$ 2 cells in the top left corner of the table:

$$\mu_{\{1,+0\}} = 4.3,\ \mu_{\{1,+10\}} = 4.2,$$
$$\mu_{\{2,+0\}} = 3.9,\ \mu_{\{2,+10\}} = 4.2$$

All of the specifications for this problem can be incorporated into a single SAS data set, as follows.

```
proc plan ordered;
  factors Company=5 SugiSupp=5 / noprint;
  output out=Design Company  cvals=('Gamma'
    'Epsilon' 'Zeta' 'Eta' 'Theta')
    SugiSupp nvals=(0 10 20 40 80);
run;

data CellWeights;
   input CellWgt @@;
datalines;
2 1 1 0 0
2 1 0 1 0
2 0 1 0 1
2 0 0 1 1
2 1 0 0 1
;

data CellMeans; keep Scenario1 Scenario2;
  array A{5} (0.0 -0.2 0.2 -0.1 0.1);
  array B{5} (0.0  0.1 0.4  0.5 0.5);
  do i = 1 to 5; do j = 1 to 5;
    Scenario1 = 4.2 + A{i} + B{j};
    Scenario2 = Scenario1;
    if      ((i=1)&(j=1))
      then Scenario2 = Scenario2 + 0.1;
    else if ((i=1)&(j=2))
      then Scenario2 = Scenario2 - 0.1;
    else if ((i=2)&(j=1))
      then Scenario2 = Scenario2 - 0.1;
    else if ((i=2)&(j=2))
      then Scenario2 = Scenario2 + 0.1;
    output;
  end; end;
run;
data rabbits5x5;
  merge Design CellWeights CellMeans;
run;
```

The scientists consider 0.56 to be a reasonable guess of the error SD, but they would also like to assess the power assuming this SD is as high as 0.73. They believe there is a correlation of about ρ=0.45 between baseline body weight and body weight after 24 weeks. The design is randomized, and so there is no underlying relationship between RabbitWgt00 and the design factors, Company and SugiSupp. The team also presumes that the Company and SugiSupp effects are not moderated by RabbitWgt00, i.e., there is no RabbitWgt00 $\times$ Company or RabbitWgt00 $\times$ SugiSupp interaction. Accordingly, the only effect of adding RabbitWgt00 to the linear model will be to reduce the error standard deviation to $(1-\rho^2)^{\frac{1}{2}}$ of its original value. Thus, ρ = 0.45 reduces the SD values by $100\left[1-(1-\rho^2)^{\frac{1}{2}}\right]$% = 10.7%. So the conjectured values for error SD (originally 0.56 and 0.73) become 0.5 and 0.65.

Statistical Method: The team assumes normality for the distribution of rabbits' body weights (conditional on the explanatory variables). Variables such as body weight tend to be positively skewed and may need to be transformed prior to analysis. In addition, without such a transform, the SDs for the two groups may not be equal, because there tends to be a positive relationship between groups' means and their SDs. Both problems are often greatly mitigated by using a log transform, i.e., by assuming that the original data is lognormal in distributional form. But for the purposes of this example, assume that the normality assumption for the body weights is reasonable.

This study could be analyzed in numerous ways. The

strategy chosen should be incorporated into a sample size analysis that conforms to the data analysis plan. The scientists decide to compare each of the four supplemental formulations with the control in an ANOVA, using a Bonferroni correction for multiple testing with overall α = 0.05, or α = 0.05/4 = 0.0125 per test. Alternatively, they could use Dunnett's test. They will also test for a dose-response relationship, assuming that the essential component of that relationship is captured using just the linear trend across the five levels of SugiSupp. Formally, this assumes that +0, +10, +20, +40, and +80 ppm of sugianimum are equally spaced in terms of the potential effect on body weight. The appropriate contrast is

```
contrast "linear trend" SugiSupp -2 -1 0 1 2
```

The model is a two-way ANOVA with main effects only and a covariate, which can typically be specified with SAS code as

```
freq CellWgt;
class Company SugiSupp;
model Scenario1 Scenario2 = Company SugiSupp
                            RabbitWgt00;
```

Note that scenario 2, with its small interaction effect, does not satisfy this statistical model. But power computations are still perfectly valid. It may be of interest to investigate how a model misspecification affects power.

Contrasts between the standard formulation and each of the alternatives can typically be specified with SAS code as

```
contrast "+0 vs +10" SugiSupp 1 -1;
contrast "+0 vs +20" SugiSupp 1 0 -1;
contrast "+0 vs +40" SugiSupp 1 0 0 -1;
contrast "+0 vs +80" SugiSupp 1 0 0 0 -1;
contrast "linear trend" SugiSupp -2 -1 0 1 2;
```

Significance will be judged at α = 0.0125 for the pairwise comparisons and α = 0.05 for the test of a linear trend in dose response.

Aim of Assessment: The scientists want to ascertain whether 240 rabbits are sufficient to provide adequate power for their planned tests, according to their two scenario models. They wonder whether fewer rabbits might suffice. To investigate the effect of sample size on power, they will also consider a design with only 160 rabbits.

The approach used to calculate power for this situation is explained in the next section, "The Univariate GLM with Fixed Effects." The results computed using UnifyPow (O'Brien 1998) are displayed in Table 3.

α = .0125 for comparisons and .05 for linear trend		Standard Deviation 0.5		0.65	
		Total N		Total N	
Scenario	Test	160	240	160	240
1	+0 vs +10	.047	.067	.032	.043
2	+0 vs +10	.047	.067	.032	.043
1	+0 vs +20	.573	.788	.332	.515
2	+0 vs +20	.529	.746	.301	.473
1	+0 vs +40	.804	.948	.532	.749
2	+0 vs +40	.833	.961	.566	.782
1	+0 vs +80	.942	.994	.737	.912
2	+0 vs +80	.942	.994	.737	.912
1	linear trend	.996	.999	.941	.991
2	linear trend	.996	.999	.946	.992

Table 3. Power Values

So, the small degree of interaction (in scenario 2) barely affects the power. Assuming these scenarios are reasonable, a main-effects-only model should suffice. The study is likely to find significance for the linear trend in dose/respose and the higher formulations of SugiSupp, but it is also very likely that +10 will be deemed "below threshold." The scientists should be conservative in reporting non-significant results for the threshold comparisons, since the power is quite low until +40. If the SD is 0.65, then perhaps only +80 would see a threshold effect. Given the mediocrity of the power values, the scientists realize that they should use all 240 rabbits.

As a side note, a power analysis ignoring the covariate RabbitWgt00 reveals that the maximum possible power with SD = 0.73, for tests other than the linear trend contrast (which has very high power in all cases), is 82.4% for the contrast between +0 and +80 with N = 240.

The next two sections outline power computations for the general framework of linear models with fixed effects (univariate and then multivariate) and alternative strategies for specifying the relevant components.

The Univariate GLM with Fixed Effects

The example in the previous section involves an ANOVA with two factors and a covariate, a special case of the univariate GLM with fixed effects.

Methods for computing power for the general linear model with fixed effects have been developed in a series of papers, providing exact results for univariate models, as well as good approximations for both multivariate models and univariate approaches to repeated measures. Muller et al. (1992) summarize results for these situations, and O'Brien and Shieh

(1992) develop a slightly improved power formula for multivariate models.

The univariate GLM is discussed in this section, with special emphasis on the alternative ways in which you can specify the quantities involved in power computations. These quantities are encompassed by the five-component layout as demonstrated in the examples in this paper.

The multivariate GLM is discussed in the next section. Computations for the univariate approach to repeated measures (with sphericity or without, using Greenhouse-Geisser or Huynh-Feldt corrections) are not discussed here but are similar in spirit to the ones outlined in this section; details can be found in Muller and Barton (1989) and Muller et al. (1992).

The univariate GLM is represented as follows:

$$Y = XB + \epsilon \quad \text{where} \quad \epsilon \sim \mathrm{N}(0, \sigma^2)$$

where Y is the vector of responses, X is the design matrix, B is the vector of effect coefficients, and ϵ is the vector of errors.

The independent variables represented in X may be either categorical or continuous. Consequently, the univariate GLM covers *t*-tests, fixed-effects ANOVA and ANCOVA, and multiple linear regression, which have been discussed along with examples in previous sections. This section outlines a more general framework and expounds on the various ways of expressing the components required for a sample size analysis.

Typically, hypotheses of interest in these models have the general form of a linear contrast

$$H_0 : CB = \theta_0$$

where C is a matrix of contrast coefficients and θ_0 is the null contrast value. Note that this formulation covers the overall test and tests of individual effects as special cases.

The components involved in power computations can be broken down as follows, showing alternative formats for how some of the quantities can be specified:

Study Design:

- design profiles: {essence design matrix} or {exemplary X} or {empirical mean and covariance of X rows}
- sample size: {total sample size and weights of design profiles} or {number of replications of design profiles}

Effects & Variability:

- model parameters: {cell means} or {model parameters using another coding scheme} or {exemplary X and Y}
- error variance: {error standard deviation (root MSE)} or {exemplary X and Y}
- multiple correlation between Y and continuous covariates (if applicable)

Statistical Method:

- model equation
- contrast coefficients
- test statistic (including multiple comparison information, if applicable)
- null contrast value
- significance level

Aim of Assessment:

- various (compute power, sample size, etc.)

The (exact) computation of power is intuitive in the sense that it involves the noncentral F distribution whose noncentrality is computed in exactly the same way as the F test statistic except with estimates ($\hat{B}$ and $\hat{\sigma}$) replaced by conjectured true values. For the equations and other computational details, see O'Brien and Shieh (1992). Although the equations express power as a function of the other components, solutions for sample size and other quantities can be obtained via iteration.

There are several different ways in which you can specify the components required to compute power. You may choose to specify some or all quantities directly, such as the design matrix (X), error standard deviation (σ), and model parameters (B). Recall that in the rabbit example in the previous section, SD was posited directly.

Instead of the full design matrix, you can provide the *essence* design matrix (the collection of unique rows in X) along with the weights or frequencies of each row. This has the benefit of expressing the design profiles and sample sizes independently of each other, since the number of rows in the full design matrix X varies with sample size.

Even if you don't code the X matrix as a cell-means model, you can express the model parameters B as cell means, the collection of mean response values at each factor-level combination. This is often the most familiar coding scheme.

The CONTRAST statement in GLM and other SAS/STAT® procedures can be a handy shorthand way for specifying contrasts of interest in complicated models.

As an alternative to specifying quantities directly, you can formulate an *exemplary data set*, a hypothetical data set having the same format as the one that will eventually be used in the data analysis. Instead of gathering real data, however, you fill the exemplary data set with "pretend" observations that are representative of the scenario for which you want to perform power computations. It summarizes the mean scenarios and cell weights under consideration. Often this approach is easier (than direct parameter specification) for inferring the design, effect values, and (optionally) error standard deviation. Recall that in the rabbit example in the previous section, an exemplary data set was used to specify the design and means.

To provide a minimally useful amount of information, an exemplary data set must contain each design profile that will be used, and the response values must be indicative of conjectured effects. This allows the design structure and effect values to be inferred. If the design profiles occur in the same proportion as they will in the actual study, then the profile weights and error standard deviation can also be inferred. Since the model and statistical test cannot be inferred from exemplary data, they must be specified separately.

Special considerations apply in the presence of continuous independent variables ("covariates"), depending on whether they are involved in the statistical tests. In a randomized design where the covariates X_c are measured at baseline (before randomization) and are *not* included in the contrast, you can compute an approximate power as demonstrated in the rabbit example in the previous section. Conjecture the (multiple) correlation $R^2_{Y|X_c}$ between Y and X_c. Reduce the standard deviation of the residual term to $\sigma(1 - R^2_{Y|X_c})^{\frac{1}{2}}$. Proceed as if the covariates are not in the model, except that the degrees of freedom for the residual is reduced by the number of covariates. This simplification holds only if X_c is uncorrelated with the variables already in the model.

If the covariate distribution differs across groups, then the contrasts apply to the least square means (LSMEANS) rather than to the simple means.

If covariates are included in the statistical tests, then you have two feasible (albeit complicated) strategies to choose from. For contrasts amounting to tests of individual effects, you can re-cast the contrast and effects in terms of correlations and use one of the approaches described in the "Multiple Linear Regression" section. Or for any contrast, you can specify X in its full form or in terms of its empirical mean and covariance.

The Multivariate GLM with Fixed Effects

The multivariate GLM is an extension of the univariate GLM in the sense of having more than one response variable, i.e., Y is a matrix instead of a vector. Important special cases include repeated measures and MANOVA. Although exact power computations are not availiable except in the case of one-d.f. contrasts, O'Brien and Shieh (1992) develop good approximate formulas.

As an example, the model used for the rabbits in the "Multi-Way ANOVA with Fixed Effects and Covariates" section could be extended to a multivariate model by including body weight measurements at a number of different times, say, 12, 24, and 36 weeks.

The multivariate GLM is represented as follows:

$$Y = XB + \epsilon \quad \text{where} \quad \epsilon_i \sim \mathrm{N}(0, \Sigma)$$

where Y is the matrix of responses, X is the design matrix, B is the matrix of effect coefficients, ϵ is the matrix of errors (with rows $\{\epsilon_i\}$), and Σ is the covariance matrix of the Y columns (varying over "within" factor levels). The matrix Σ is often referred to as the covariance of repeated measures.

The hypothesis under consideration is the contrast

$$H_0 : CBA = \theta_0$$

where C is a "between" contrast matrix (involving effects specified by X), and A is a "within" contrast matrix (involving the columns of Y).

All of the sample size analysis components discussed for the univariate GLM also apply for the multivariate model. In addition, you must specify the test statistic, the covariance of repeated measures, and the within contrast matrix. Special forms of the within contrast matrix give rise to special cases such as classical MANOVA, growth profile analysis with time polynomials, and between-trend analysis. Here is a summary of the *additional* required components in a multivariate GLM sample size analysis:

Study Design:

- number of repeated measurements (i.e., number of columns in Y)

Effects & Variability:

- covariance of repeated measures: {covariance matrix} or {type of covariance matrix and relevant parameters (for example, compound symmetry or AR(1))} or {exemplary X and Y}

Statistical Method:

- "within" contrast coefficients
- test statistic: Wilk's likelihood ratio, Hotelling-Lawley trace, Pillai trace, etc.

A Survey of Other Situations

There are many types of linear models, and other approaches to the multivariate models with fixed effects, not covered in the previous sections. This section presents a brief summary for sample size analysis with other popular types of linear models.

Lognormal Data When the data are lognormally distributed in an ANOVA, you can specify effects in terms of mean ratios, supply a conjecture for the coefficient of variation (assumed common across design profiles), and proceed with the same approaches already developed for standard ANOVA models. Lognormal outcomes occur in many situations, such as when the response variable is a probability or growth measurement. Often you can express cell means conveniently in this paradigm, as a fraction of the reference or baseline level.

Multiple Comparisons When a study involves multiple inferences or multiple comparisons, power considerations require specifying precisely which inferences you want power for. Westfall et al. (1999) discuss the issue and give some computational tools.

Mixed Models Currently there is no accepted general standard for power computations in *mixed* linear models, with both fixed and random effects, although methods have been developed for some special cases. It is an active area of research, and currently simulation remains the recommended approach.

Some classes of models with random effects (for example, simple split-plot designs) can be re-cast as multivariate linear models, with the random effect modeled instead as multiple response values. Power analysis can thus be performed using the methods discussed in this paper.

O'Brien and Muller (1993, section 8.5.2) show an exact power computation for a one-way random-effects ANOVA using a multiple of a central F distribution. Lenth (2000) computes approximate power for a wide variety of balanced ANOVA designs with fixed and random effects (where all of the random effects are mutually independent, inducing a compound symmetry correlation structure) by constructing F tests as ratios of expected mean squares and applying Satterthwaite corrections for degrees of freedom. Effects are specified as variance components for random factors and sums of squares for fixed factors.

Power analysis is particularly complicated for mixed models, due to the wide variety of statistical tests that are available. Helms (1992) develops a method for computing approximate power for contrasts of fixed effects, for the approximate F test involving REML estimators of the model coefficients and covariance. The non-null distribution is approximated by a noncentral F with noncentrality estimated much in the same way as O'Brien and Shieh (1992) do for the multivariate GLM, by replacing estimates with conjectured true values.

Simulation Regardless of the availablility of exact or approximate formulas for power computations, simulation remains a viable approach for conducting a sample size analysis for any linear model (indeed, any statistical model). You must be able to simulate realizations of the model (in other words, data sets generated according to the conjectured model, design, effects, and variability), compute the test statistic, and determine when the null hypothesis is rejected. You can repeat this process and estimate power as the percentage of rejections.

Retrospective Analysis This paper has focused on *prospective* power calculations, performed as part of study planning and not based directly on actual data. *Retrospective* calculations attempt to infer the power of a study already performed or estimate power from pilot data. Bias corrections should be used in such retrospective analyses. In addition, since the variability in the observed data can be characterized and propagated through the power analysis, confidence intervals for power can be constructed. These issues are discussed thoroughly in Muller and Pasour (1997), Taylor and Muller (1996), and O'Brien and Muller (1993).

Conclusion

Power and sample size determination has been illustrated for several varieties of linear models, ranging from simple *t*-tests to multivariate models. For any of these situations, you can gather the information required for power computations by considering five aspects of study planning: the design, scenario models representing beliefs about the data, specific conjectures about the effects and variability, the statistical method to be used in data analysis, and the aim of the assessment. The ensuing power computations reveal important aspects about the planned study, such as adequate choices for sample sizes or the likelihood of significant results. Future SAS software will provide analytical tools to help you characterize and solve such problems in power and sample size analy-

sis.

Acknowledgments

We are grateful to Virginia Clark, Keith Muller, Bob Rodriguez, Maura Stokes, and Randy Tobias for valuable assistance in the preparation of this paper.

References

Castelloe, J.M. (2000), "Sample Size Computations and Power Analysis with the SAS® System," *Proceedings of the Twenty-Fifth Annual SAS Users Group International Conference,* Paper 265-25, Cary, NC: SAS Institute Inc.

Gatsonis, C. and Sampson, A.R. (1989), "Multiple Correlation: Exact Power and Sample Size Calculations," *Psychological Bulletin*, 106, 516–524.

Helms, R.W. (1992), "Intentionally Incomplete Longitudinal Designs: I. Methodology and Comparison of Some Full Span Designs," *Statistics in Medicine*, 11, 1889–1913.

Keyes L.L. and Muller, K.E. (1992), "IML Power Program," available at ftp.bios.unc.edu/ pub/ faculty/ muller/power01/distrib/.

Lee, Y.S. (1972), "Tables of the Upper Percentage Points of the Multiple Correlation," *Biometrika*, 59, 175–189.

Lenth, R.V. (2000), "Java Applets for Power and Sample Size," www.stat.uiowa.edu/ ˜rlenth/Power/.

Maxwell, S.E. (2000), "Sample Size and Multiple Regression Analysis," *Psychological Methods*, 5, 434–458.

Muller, K.E. and Barton, C.N. (1989), "Approximate Power for Repeated Measures ANOVA Lacking Sphericity," *Journal of the American Statistical Association*, 84, 549–555 (with correction in volume 86 (1991), 255–256).

Muller, K.E., LaVange, L.M., Ramey, S.L., and Ramey, C.T. (1992), "Power Calculations for General Linear Multivariate Models Including Repeated Measures Applications," *Journal of the American Statistical Association*, 87, 1209–1226.

Muller, K.E. and Pasour, V.B. (1997), "Bias in Linear Model Power and Sample Size Due to Estimating Variance," *Communications in Statistics – Theory and Methods*, 26, 839–851.

Muller, K.E. and Peterson, B.L. (1984), "Practical Methods for Computing Power in Testing the Multivariate General Linear Hypothesis," *Computational Statistics and Data Analysis*, 2, 143–158.

O'Brien, R.G. (1998), "A Tour of UnifyPow: A SAS Module/Macro for Sample-Size Analysis," *Proceedings of the Twenty-Third Annual SAS Users Group International Conference*, Cary, NC: SAS Institute Inc., 1346–1355. Software and updates to this article can be found at www.bio.ri.ccf.org/ UnifyPow/.

O'Brien, R.G. and Muller, K.E. (1993), "Unified Power Analysis for t-Tests Through Multivariate Hypotheses," In Edwards, L.K., ed. (1993), *Applied Analysis of Variance in Behavioral Science*, New York: Marcel Dekker, Chapter 8, 297–344.

O'Brien, R.G. and Shieh, G. (1992). "Pragmatic, Unifying Algorithm Gives Power Probabilities for Common F Tests of the Multivariate General Linear Hypothesis." Poster presented at the American Statistical Association Meetings, Boston, Statistical Computing Section. Also, paper in review, downloadable in PDF form from www.bio.ri.ccf.org/UnifyPow.

SAS Institute Inc. (1999a). *SAS/IML® User's Guide, Version 8*, Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1999b). *SAS/STAT® User's Guide, Version 8*, Cary, NC: SAS Institute Inc.

Taylor, D.J. and Muller, K.E. (1996), "Bias in Linear Model Power and Sample Size Calculation Due to Estimating Noncentrality," *Communications in Statistics –Theory and Methods*, 25, 1595–1610.

Westfall, P.H., Tobias, R.D., Rom, D. Wolfinger, R.D. and Hochberg, Y. (1999), *Multiple Comparisons and Multiple Tests Using the SAS® System*, Cary, NC: SAS Institute Inc.

Contact Information

John M. Castelloe, SAS Institute Inc., SAS Campus Drive, Cary, NC 27513. Phone (919) 531-5728, FAX (919) 677-4444, Email John.Castelloe@sas.com.

Ralph G. O'Brien, Department of Biostatistics and Epidemiology/Wb4, 9500 Euclid Avenue, Cleveland, Ohio, 44195. Phone (216) 445-9451, FAX (216) 444-8023, Email robrien@bio.ri.ccf.org.

Version 1.2