Published in final edited form as:

Int J Knowl Eng Soft Data Paradig. 2009 January 1; 1(1): 40–48. doi:10.1504/IJKESDP.2009.021983.

Statistical Power Calculations for Clustered Continuous Data

A.T. Galecki^{*}, T. Burzykowski^{**}, S. Chen^{*}, J.A. Faulkner^{*}, and J. Ashton-Miller^{*}

^{*}University of Michigan, Ann Arbor, USA

^{**}Hasselt University, Diepenbeek, Belgium

Abstract

To calculate the sample size for a research study it is important to take into account several aspects of the study design. In particular, one needs to take into account the hypotheses being tested, the study design, the sampling design, and the method to be used for the analysis. In this paper we propose a simple method to calculate sample size for clustered continuous data under various scenarios of study design.

1 Introduction

In this paper we focus on sample size calculation for studies leading to clustered continuous data. Study designs that can result in clustered data sets include observational studies involving clusters of observations; cluster-randomized trials, in which a treatment is randomly assigned to all units within a cluster of units; or randomized block design experiments, in which the blocks represent clusters and treatments are assigned to units within blocks. In all of these studies, characteristics of both the clusters and the units forming the clusters are measured. The primary objective of the analysis of clustered data is to quantify the effects of predictors on the dependent variable measured for all units, while accounting for the presence of clusters of units, and for the possible within-cluster correlation of the dependent variable measurements in particular. We propose a simple method to calculate sample size for clustered data under various study design scenarios.

2 Background

2.1 Model specification

The important feature of clustered data is that measurements taken on units within clusters involve two sources of variation, i.e., between and within cluster variation. To properly account for the two sources of variation, mixed-effects models are often used as the analytical method. More specifically, models with two random terms, i.e., a random intercept, to describe the between-cluster variation, and a residual error, to describe the variation within a cluster, are frequently applied.

The model for the j -th unit within the i -th cluster ($i = 1, \dots, n$) is specified as

$$y_{ij} = \mathbf{x}_{ij}\beta + u_i + e_{ij}, \quad (1)$$

where $u_i \sim \mathcal{N}(0, \sigma_c^2)$, $e_{ij} \sim \mathcal{N}(0, \sigma_e^2)$, and $j = 1, \dots, n_i$. The total number of units in all n clusters is equal to $N = \sum_{i=1}^n n_i$

Note that predictors stored in vector $\mathbf{x}_i$ do not depend on index j , and therefore they characterise clusters themselves, not units within a cluster. Vector β , with p elements, contains the fixed effects to be estimated.

In the context of model (1), the following two important concepts are often defined. The first concept involves an intraclass correlation coefficient for clustered data. The coefficient is defined as $\rho = \sigma_c^2 / (\sigma_c^2 + \sigma_e^2)$. The second concept is a marginal variance-covariance matrix, $\mathbf{V}_i$ for cluster i , which results from model (1) after integrating out the random effects u_i . Matrix $\mathbf{V}_i$ has dimension $n_i \times n_i$; its diagonal elements are equal to $\sigma^2 = \sigma_c^2 + \sigma_e^2$ and off-diagonal elements are equal to σ_c^2 . This particular structure of $\mathbf{V}_i$ is called compound-symmetry or an equicorrelational structure with homogenous variances. Both of these concepts, i.e., ρ and $\mathbf{V}_i$, quantify correlation between any two units, i.e., between y_{ij} and $y_{ij'}$ for $j \neq j'$, from the same cluster.

2.2 Hypothesis test for fixed effects

In the analysis of clustered data we are typically interested in testing hypotheses about contrasts of fixed effects β . The hypotheses are specified as

$$H_0: \mathbf{L}\beta = 0 \quad \text{versus} \quad H_A: \mathbf{L}\beta = \Delta_0, \quad (2)$$

where $\mathbf{L}$ is a known $q \times p$ matrix and $\Delta_0 \neq \mathbf{0}$ is a known vector with q elements. Note that H_0 and H_A specify the null and alternative hypotheses, respectively.

Null hypothesis H_0 can be tested in a variety of ways. One possible approach is to use the F-test. The test statistic for H_0 , constructed based on model (1), can be specified as

$$F = \hat{\beta}' \mathbf{L}' \left(\mathbf{L} \left(\sum_i w_i \mathbf{x}_i' \mathbf{x}_i \right)^{-1} \mathbf{L}' \right)^{-1} \mathbf{L} \hat{\beta} / (\hat{\sigma}^2 \text{rank}(\mathbf{L})), \quad (3)$$

where $w_i = n_i / (1 + (n_i - 1)\rho)$ and $\hat{\sigma}^2 = \hat{\sigma}_c^2 + \hat{\sigma}_e^2$. Respective estimates $\hat{\beta}$, $\hat{\sigma}_c$, $\hat{\sigma}_e$ of β , σ_c and σ_e are obtained using likelihood methods. Under H_0 , statistic F follows approximately the central F distribution with numerator and denominator degrees of freedom equal to $\text{rank}(\mathbf{L})$ and $N - n$, respectively. For details related to a more general case of linear mixed effects models see Verbeke and Molenberghs (2000).

2.3 Statistical Power Calculations

A general method to perform power calculations for linear mixed effects models was proposed in Helms (1992). When applied to model (1), the method implies that, under alternative hypothesis H_A , the F-statistic, defined in (3), follows approximately a *non-central F* distribution, with numerator and denominator degrees of freedom $\text{rank}(\mathbf{L})$ and $N - p$, respectively. The non-centrality parameter, δ , is equal to

$$\delta = \Delta_0' \left(\mathbf{L} \left(\sum_i w_i \mathbf{x}_i' \mathbf{x}_i \right)^{-1} \mathbf{L}' \right)^{-1} \Delta_0 / \sigma^2. \quad (4)$$

Assuming that we are testing hypothesis (2) at significance level α , the calculation of power is performed as follows. We first calculate the $(1 - \alpha)$ quantile $F_{1-\alpha}$ of the central F distribution

with $\text{rank}(\mathbf{L})$ and $N-p$ numerator and denominator degrees of freedom, respectively. Then we calculate power as $1-F(F_{1-\alpha}, \text{rank}(\mathbf{L}), N-p, \delta)$, where $F(x, d_1, d_2, \delta)$ is the cumulative distribution function for a non-central F distribution with numerator and denominator degrees of freedom d_1 and d_2 , respectively, and non-centrality parameter δ .

3 Algorithm

In this section we outline a simple algorithm to perform statistical power calculations for clustered data. Based on similarities between the formulae for the F -statistic, specified in (3), and for the non-centrality parameter δ , specified in (4), we propose to perform power calculations in the following three steps:

- Step 1: Create an auxiliary dataset.
- Step 2: Test hypotheses of interest in the auxiliary data.
- Step 3: Calculate statistical power.

Each of these steps is described in more detail in the remainder of this section.

First, we predefine the number of subjects n , and number of units n_i per each cluster. Based on pilot data or based on the information from the literature we set σ_e^2 and σ_c^2 to known values. Then we proceed with Step 1 and prepare an auxiliary dataset containing *one* row per cluster. The created auxiliary dataset is essentially the same as data commonly used for an analysis using a classical linear model. The data contain the predictor variables needed to calculate the elements of vectors $\mathbf{x}_i$ and two additional columns containing the pseudo-dependent variable y_i^* , and weights w_i , used in Step 2 of the algorithm. More specifically, values y_i^* of the pseudo-dependent variable are calculated using the formula $y_i^* = \mathbf{x}_i \beta^*$, where β^* is any value of β fulfilling the equation $\mathbf{L}\beta = \Delta_0$. The second additional variable contains weights, defined as $w_i = n_i / (1 + (n_i - 1)\rho)$, where ρ is the intraclass correlation coefficient defined in Section 2.1.

In Step 2 of the algorithm, we test null hypothesis H_0 in the context of the model

$$y_i = \mathbf{x}_i \beta + \varepsilon_i, \quad (5)$$

Where $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ for $i = 1, \dots, n$. Note that we use vector y_i^* as observed values of the dependent variable y_i . To find estimates $\widehat{\beta}$, we use weighted linear regression with weights w_i . The reader may note that, in contrast to a classical linear model, the variance of ε_i term in (5) is known and it is not being estimated from the data. This implies that any software designed for the linear regression models can be used, provided it allows for a *known* residual variance. In Step 3, we compute the power in the way described in Section 2.3. The idea presented in this section, although similar to that presented in Litell et al (2006), takes advantage of simplifications implied by the specific structure of model (1).

4 Illustration: Michigan Life Science Study

4.1 Study Design and Objectives

The Michigan Life Science (MLS) Corridor Study “Improving muscle power and mobility of elderly men and women” was a four-year study funded by the State of Michigan. The MLS study was designed as a randomised controlled trial, in which healthy young (21–30 years) and older (65–80 years) male and female subjects were randomised into one of two arms of a 12-week progressive resistance training (PRT) exercise intervention. The intervention aimed at strengthening lower extremity muscle strength and power. One group of subjects performed

“fast” PRT of the leg muscles using lighter weights, while the other performed the traditional “slow” PRT of the leg muscles using heavier weights. The two study groups were stratified by gender and age (young/old). The training regimen consisted of training three days per week over three months.

The primary outcome was cross-sectional area (CSA) of single muscle fibers obtained from study subjects during biopsy at the end of the study. Data obtained from this study are an example of clustered data, because during each biopsy multiple fibers were obtained and CSA was measured for all the fibers. The primary objective of the study was to test the hypothesis that there was a difference in the effect of the “fast” PRT on CSA as compared to the “slow” PRT.

4.2 Power Calculations

Power calculations are performed assuming the model

$$y_{ij} = \begin{cases} \mu_F + u_i + e_{ij} & \text{for fast group,} \\ \mu_S + u_i + e_{ij} & \text{for slow group} \end{cases},$$

where μ_F and μ_S are fixed effects for “Fast” and “Slow” PRT study groups. The remaining terms u_i and e_{ij} are defined in Section 2.1. The null and alternative hypotheses are

$$H_0: \mu_F - \mu_S = 0 \quad \text{versus} \quad H_A: \mu_F - \mu_S = \Delta. \quad (6)$$

To perform the power calculations, we needed an estimate of a clinically and physiologically meaningful intervention effect, Δ , that should be detected. Based on previous experience, the value of Δ was set to $3 \mu\text{m}^2$. The values of σ_c^2 and σ_e^2 were estimated from pilot data and were set to $12.4 \mu\text{m}^4$ and $23.6 \mu\text{m}^4$, respectively. The corresponding value of ρ is equal to 0.34.

One of the important issues in designing the MLS study was to find an optimal number of muscle fibers needed per biopsy, given budget constraints. We assumed the following cost function:

$$\text{Total cost} = n \times C + N \times U, \quad (7)$$

where C and U are costs associated with a cluster and with a unit within a cluster, respectively. In the MLS study we assumed that the subject-specific cost is equal to $C = \$2,860$ and the fiber-specific cost is equal to $U = \$170$.

We tentatively set the number of subjects studied in each intervention group to 25. The cost function (7), along with the total (hypothetical) budget that should not exceed \$200,000, implies that the number of fibers taken at the biopsy should be equal to six. Assuming $\alpha = 0.05$, the power is equal to 0.74. In Appendix we present portions of the SAS and R code used to perform calculations.

The reader may note that, for any value of constant c_0 , all sets of values, $c_0\Delta$, $c_0\sigma_c$, and $c_0\sigma_e$ are equivalent, in the sense that they imply the same intraclass correlation coefficient, ρ , and power. Therefore, instead of requesting Δ , σ_c^2 , and σ_e^2 , to perform power calculations it is sufficient to provide value of ρ and the effect size, say ES, defined as $ES = \Delta/\sigma$, where $\sigma^2 = \sigma_e^2 + \sigma_c^2$. The effect size for our study is equal $ES = 3 / \sqrt{12.6 + 23.4} = 0.5$.

Figure 1 presents the relationship between the power and the number of subjects per study group for different values of ES varying from 0.1 to 0.5, under the assumption that the total budget is equal to \$200,000. Note that on the horizontal axis, in addition to the number of subjects, the corresponding number of fibers per subject is given in parentheses. The optimal number of subjects per group is 25 with 6 fibers assessed for every subject. Another combination having power very close to the optimal one is 28 subjects per group and 4 fibers per subject. It is interesting to note that the effect of additional subjects over the optimum number of 25 on statistical power is offset by a lower number of fibers collected for each subject.

5 Discussion

The method most commonly used to perform power calculations for clustered data is based on a correction of the sample size that involves the design effect defined as $1 + \rho(k - 1)$, where k is the average cluster size. The method has some drawbacks. For example, it does not allow one to take into account the presence of covariates, and it does not allow one to properly take into account a varying cluster size. The method presented in this paper allows one to overcome these shortcomings. It can be directly used in designing a study or in simulations.

Acknowledgments

We gratefully acknowledge support from the Claude Pepper Center Grants P30-AG08808 and AG024824 from the National Institute on Aging and the financial support from the IAP Research Network P6/03 of the Belgian Government (Belgian Science Policy).

Appendix

In the Appendix we present portions of the SAS and R code used to perform calculations.

In Step 1 of the algorithm, described in Section 3, we created data named `dt` that contained 50 rows and three variables named `recno`, `grp`, and `y`. Selected rows from these data are given for illustration below

recno	grp	y
1	1	0
...		
25	1	0
26	2	3
...		
50	2	3

Variables `recno`, `grp`, and `y` indicate row number, study group and value of y_i^* variable, respectively. Variable `wght` (not shown) is created separately during the execution.

Calculations using SAS

In the code below we demonstrate how to use SAS to perform Steps 2 and 3 of the algorithm. To fit linear model in Step 2, we use PROC MIXED. Statement PARMS is used to specify σ^2 . Options NOPROFILE and NOITER in PARMS statement are used to keep it constant during calculations. ODS OUTPUT statement is used to save the F-statistic in data named `contrasts`. In the DATA step power is calculated, and Step 3 of the algorithm is completed.

Title "Power calculations: Step 2";

```

ods output contrasts=contrasts;
proc mixed data=dt noprofile;
class grp;
weight wght;
model y=grp;
parms (36)/ noiter;
contrast "grp" grp 1 -1 ;
quit;
ods output close;

Title "Power calculations: Step 3";
data power;
    set contrasts;
    alpha = 0.05;
    n = 50;           * Number of subjects;
    ni = 6;           * Number of fibers per subject
    P = 2;            * Number of fixed effects
    dendf0 = n*ni - n;
    qf = finv(1-alpha, numdf, dendf0);
    dendfA = n*ni - p;
    power = 1 - probf(qf, numdf, dendfA, Fvalue*numdf);
    put power=;       * Equal to 0.7436023;
run;

```

Calculations using R language

The code below demonstrates key commands used in R language to perform power calculations. Note that, in contrast to PROC MIXED in SAS, function `lm()` in R does not allow to specify arbitrary value of σ . Instead, the `{lm()}` function estimates σ from data and, not surprisingly, obtains a value of `sd0` very close to 0. To obtain the desired F-statistic, we calculated scalar `sc` and rescaled the F-statistic. Although the procedure of “rescaling” appears to work fine, some caution is required, because it involves division by `sd0` having value very close to zero.

```

# Step 2 using R:
lm0    <- lm(ym ~ grp, data = dt, weights = wght)
sd0     <- summary(lm0)$sigma
sc      <- sqrt(36)/sd0
Fvalue  <- anova(lm0)[“grp”, “F value”]/(sc*sc) # Rescale F stat

# Step 3
alpha   <- 0.05
n        <- 50
ni       <- 6
P        <- 2
numdf    <- 1
dendf0   <- n*ni - n
qF       <- qf(1-alpha, numdf, dendf0)

```

```
dendfA <- n*ni - p
(power <- 1 - pf(qF,numdf,dendfA, numdf*Fvalue))
[1] 0.7436023
```

References

- [1]. Verbeke, G.; Molenberghs, G. Linear Mixed Models for Longitudinal Data. Springer; 2000.
- [2]. Helms RW. Intentionally incomplete longitudinal designs: I. Methodology and comparison of some full span designs. *Statistics in Medicine* 1992;11:1889–1913. [PubMed: 1480880]
- [3]. Littell, RC.; Milliken, GA.; Stroup, WW.; Wolfinger, RD.; Schabenberger, O. SAS for Mixed Models. SAS Publishing; 2006.

Figure 1.

MLS Study: Power plotted versus number of subjects per group for different values of effect size (ES). (Note: Total budget = \$200, 000, $\rho = 0.34$, $\alpha = 0.05$)