This is to certify that the

dissertation entitled

Computation of power in the nested random effects models

presented by

Xiaofeng Liu

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Measurement & Quantitative Methods

Stephen W. Raudenbush
Major professor

Date 12-14-99

Computation of Power in the Nested Random Effects Models

By

Xiaofeng Liu

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling, Educational Psychology, and Special Education

1999

ABSTRACT

Computation of Power in the Nested Random Effects Models

By

Xiaofeng Liu

Nested random effects models contain random effects due to the nested sampling units. Such models used to be framed as mixed analysis of variance (ANOVA). Nested data are now often analyzed by means of more flexible models such as hierarchical linear models (HLM) because HLMs accommodate continuous covariates and unbalanced designs. However, few people understand how to compute power in HLMs. This dissertation utilizes the relation between ANOVA and HLM as a basis for the computation of power in HLM. In essence, hierarchical linear models without continuous covariates are mixed ANOVA models. Power functions in ANOVA can be derived through ANOVA tables, though it is typically difficult to obtain ANOVA tables for complex designs. The dissertation simplifies the derivation of the ANOVA table through the structural representation of the models in HLM. The derived ANOVA tables can then be translated into similar HLM tables for deriving power functions in HLM. Knowledge of those power functions in HLM allows us to choose appropriate sample sizes for prospective studies using HLM. Two hypothetical examples are

provided to illustrate the application of power functions in planning educational studies.

To My Parents

ACKNOWLEDGMENTS

The dissertation is the culmination of my intellectual pursuit for the past three years. I am grateful for all the people who had helped me finish this dissertation. I thank Dr. Stephen Raudenbush for his long term academic guidance and unlimited support in my doctoral studies. Dr. Raudenbush is a very knowledgeable and generous person. He first introduced me into the computation of power two years ago, and I have been working very closely with him in this area. His insight and wisdom have greatly shaped my thinking and have tremendously helped found the basis of the dissertation. It is really hard to enumerate all his ideas in my dissertation, and I am sure that I will inadvertently overlook giving him due credit many places in the dissertation. I am thankful for Dr. Kenneth Frank, Dr. Mark Reckase, and Dr. Aaron Pallas for agreeing to serve on my dissertation committee and giving me valuable feedback.

I am also grateful for my parents who provided me the good education and supported me in any of my endeavors. Last but not the least, I should thank all my relatives for their unfailing belief in my intellectual ability since my childhood, and their help that made my first trip to the United States possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS

HLM Hierarchical Linear Model

CRT Cluster Randomized Trial

MST Multi-site Clinical Trial

ANOVA Analysis of Variance

INTRODUCTION

A random effects model specifies more than one random term. The random factors contain levels that are randomly selected from a population of possible levels. As the number of possible levels of some factors may be very large, it is impossible to assess all of them. Random inclusion of some levels in the model becomes economical and convenient. Suppose that we are studying the effect of schools on implementing one instruction method. There are hundred of thousands of schools. It would be impossible to examine all the schools. A sensible strategy is to identify all the schools and randomly sample a few of them to assess school effects (Littell et al., 1997).

A complete random effects model contains only a mean; the rest of the terms are random factors. The key interest surrounds the estimation of the grand mean and the variance of each random component. The random effects may be crossed in some cases, but nesting of random effects occurs in most real sampling schemes. Large experimental units are first selected, and then the smaller units are randomly selected from each large experimental unit. In education, school districts are the natural sampling units from which individual schools may be randomly chosen. The institutional hierarchical structure often determines the design of our studies, be it of experimental or survey nature. In fact, the hierarchical structure of experimental units leads to the nesting of random effects, which correspond to different units of various sizes

The random effects models can be analyzed through two frameworks. In the past they were treated as mixed analysis of variance (ANOVA) models. The mixed ANOVA approach always assumes a balanced design and no continuous covariates. It therefore restricts its application to many real data analysis. Data tend to come in with some missing values; and the continuous covariates are common. People are now often analyzing the data by more flexible models like hierarchical linear models (HLM). HLMs can include continuous covariates in the models and can accommodate missing data.

In HLM the models are arranged in a few levels based on the hierarchy of the sampling units. Corresponding to the experimental units, the hierarchical linear model may be represented by a sub-linear model at each level. For simplicity we assume one fixed effect at each level. The generic presentation at each level is as follows:

Outcome = mean + coefficient*fixed effect + random effect.

The "mean" and "coefficient" may be the outcome variables for the next level. Each level must contain a "mean", while the "fixed effect" and "random effect" may be optional at the higher levels. To generalize the model further we may include a continuous covariate for each level. The formula of each level changes into

Outcome = mean + coefficient*fixed effect +

coefficient*covariate + random effect.

The hierarchical formulation reflects the structure of the design. It appeals to people's intuition, and, therefore, the hierarchical linear model has gained popularity among researchers in various disciplines.

Much research in hierarchical linear model so far has been focused on estimation theory and algorithmic implementation of the parameter estimation. Much is yet to be known about the performance of the test of those parameters in the model. The power of the test of the parameters in the model is rarely computed for two reasons. First, the complexity of the model itself deters people from computing the power of the tests. It requires some mathematical sophistication to carry out the computation of power of relevant tests. Second, it is hard to derive the power function of key parameters in HLM.

Power functions are usually required for determining sample size in planning a study. Many researchers who plan a study using HLM want to choose an appropriate sample size. For example, a researcher may want to compare two types of counseling practice in schools. It is logistically easy to have one school practice one type of counseling; so the researcher decides to randomly assign half of schools to one type of counseling and the other half of schools to the other type of counseling. At the end of the study, students in schools will be examined

on certain criterion outcomes, and data will be analyzed as a cluster randomized trial by HLM. The researcher is interested in knowing how many students from each school should be recruited in the study. The question can be easily answered if we have the power function for the test of the main effect of treatment in the corresponding hierarchical linear model. In this case the power function may be derived analytically if we know the variance of the estimate of the treatment (see Raudenbush, 1997). However, there is not a general way to derive power functions for the general HLM when the design of the study becomes complex.

Stroup (1998) provides a general way to compute the power for mixed linear models. Since HLM is a subset of mixed linear models, his approach is applicable to the power calculation of HLM. If we formulate the hierarchical linear model in the framework of a linear mixed model, it can be expressed as

Model : $y = X\beta + Zu + e$ (1)

$$\begin{bmatrix} u \\ e \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}\right)$$

The vector β contains the parameters of fixed effects; and the vector u represents the random effects whose covariance matrix is G; and e is a vector of individual random error whose covariance matrix R is a diagonal matrix with the diagonal elements being σ^2.

It is noted that

$$E(y) = X\beta; \qquad Var(y) = V = ZGZ' + R.$$

Test Ho: $K'\beta = 0$,

where K is a matrix containing contrast constants.

$$(K'\hat{\beta})'[K'(X'V^{-1}X)K]^{-1}(K'\hat{\beta}) \quad \sim \chi^2_{rank(K)}(\lambda) \qquad (2)$$

where λ is the non-centrality parameter.

$\lambda = (K'\beta)'[K'(X'V^{-1}X)K]^{-1}(K'\beta)$ (Stroup, 1998).

We essentially treat the model as a case of generalized linear model. In practice this can not be used because the V will not be known exactly. However, we may substitute for the V its estimate $\hat{V}$ from the previously collected data. If $\hat{V}$ is estimated from a large sample, the computed power will be a good approximation.

The dilemma lies in the fact that unless the same study is replicated, we barely know V and β. In addition, people are interested in one or two key parameters in their research. They want to have a power function in a closed form to work with. The matrix representation rarely helps to assess the power of individual parameters.

The way out of this dilemma is to divide the model into each level. At each level we look at the corresponding test, standardize a few parameters, and make the power calculation feasible and meaningful. The practice of standardization is both convenient and necessary. Standardization of elements in V and β allows us to explain parameters meaningfully. They make sense to educational researchers because their studies do not use a common measurement scale. Standardization allows us to disregard any measurement scale in a particular study and evaluate a prospective study in a general sense. Raudenbush and Liu (1999) have created a scheme to standardize the parameters for cluster randomized trial (CRT) and multi-site clinical trial (MST).

The same approach can be used for HLM in general. Structure-wise, every pair of levels is either like a cluster randomized trial or a multi-site trial. The same standardization principles (see Raudenbush, 1997, and Raudenbush & Liu, 1999) used in CRT and MST can apply across different levels.

The key problem revolves around how to construct the test for each parameter at each level. The construction of the test of parameters at each level requires knowledge of the standard error of its estimate. When the model is very simple, there are ready formulas to derive those standard errors. When the model is complex, standard errors are normally expressed through matrices. There are not any simple procedures on how to write down those standard errors

algebraically. However, this problem may be solved with the aid of an ANOVA table.

Mixed ANOVA models are often used in experimental design, although educational researchers and many other social scientists are now increasingly replacing them with flexible models like HLM in data analysis. In fact these models are equivalent if we take continuous covariates out of HLM. Raudenbush (1993) shows that the nested random effects ANOVA is equivalent to HLM. If restricted maximum likelihood is used for estimation, HLM duplicates the ANOVA estimates of the same parameters. Since the power computation of the mixed ANOVA is known, we may capitalize on this to come up with a general scheme to compute the power for HLM.

We may translate HLM without continuous covariates into a mixed ANOVA to derive the power functions, and then we may extend the computation of power to HLM with continuous covariates. The computation of power in ANOVA is based on the expectation of the mean squares in the ANOVA table. In fact the ANOVA table for even moderately complicated HLMs are hard to obtain. Although there are procedures by Scheffe (1959), etc., to derive ANOVA tables, these procedures are too unwieldy to be used. Most of the time, following these procedures will not allow us to derive the correct ANOVA table.

This dissertation provides a different set of rules and procedures to derive the mixed ANOVA table. The rules hinge on the structural representation of the mixed ANOVA model in its HLM format. The structural representation allows us to identify the relation among all the mean squares in the ANOVA table. The expectations of those mean squares may easily be obtained once we know which mean squares are tested against which for a certain test.

The provided rules basically have two implications for methodology. First, they highlight the fact that ANOVA and HLM without continuous covariates are related in a systematic way. Once the relation is illustrated, it is easy for people with training of either HLM or ANOVA to learn the other. Also, the rules enable people to derive the ANOVA expected mean square table more easily, compared to other available rules. The new rules complement those old ones by providing a way to write down complicated models and check the derived results. They are easy to use. Second, ANOVA originates from experimental design. Planning an ANOVA design is not new. Power computation is already known in ANOVA. The relation between HLM and ANOVA allows us to design a HLM study with all the planning tools from ANOVA.

The dissertation will start with a general introduction to power in the 1st chapter. In the 2nd chapter it will identify the above-mentioned rules and illustrate the derivation of ANOVA table of a generic example. In the 3rd chapter we provide algebraic proof to justify these rules. To simplify the computation of power in

HLM we replace the ANOVA table with a similar HLM table. The HLM table bears similar features as the ANOVA table except that all the parameters are in HLM notation. In the 4th chapter four HLM tables are provided for four complex HLM designs. In the 5th chapter the application of power function is shown with two examples. In particular, the power functions are used to choose sample sizes for two hypothetical studies using HLM. The choice between two different designs is discussed with respect of the performance of statistical power. The conclusion discusses the power in HLM with continuous covariates and with missing data.

Chapter 1

STATISTICAL POWER AND ITS COMPUTATION

Hypothesis testing and statistical power

Hypothesis testing usually sets up two complementary hypotheses. One is the null hypothesis; the other is the alternative hypothesis. It is the null hypothesis that we usually test; and its rejection establishes the plausibility of the alternative hypothesis. Normally we put whatever we wish to prove as the alternative hypothesis and its complementary opposite as the null hypothesis. For example, to test the differential effectiveness of two teaching methods we state the null hypothesis that the two methods are equally effective and the alternative that they are not.

In any case the rejection of the null hypothesis is of our great interest. If the null is true, its rejection is a type I error. If the alternative is true, our ultimate goal is to reject the null. The probability of rejecting the null is called the power of the test. When the null is false, failure to reject the null results in the type II error.

In the following we provide the power functions for T test, F test, and Chi square tests. In addition, the power function is shown for the test of a random component in a random effects model, and power functions are also derived for linear model and mixed linear models.

T test

The two-sample t test is a widely used t test. We might suppose that responses from the experiment group $y_{e1},\ldots, y_{en}$ are i.i.d. $N(\mu_e, \sigma^2)$ and that responses from the control group $y_{c1},\ldots, y_{cn}$ are i.i.d. $N(\mu_c, \sigma^2)$. If we try to produce evidence that there is a difference in responses between the two groups because of the treatment administered to the experiment group, we may set the model as follows:

$$y_{ji}=\alpha+\beta X_{ji}+e_{ji} \; ; \qquad i:1,2...n; j=e \text{ (experimental)}, c \text{ (control)}$$

X_{ji} is ½ if it is the experimental condition or -½ if the control condition; and $\beta = \mu_E - \mu_C$.

The hypotheses are

H_0: $\beta = 0$

H_1: $\beta > 0$,

Here σ^2 is assumed to be unknown, and the test is a T statistic,

$$T=\sqrt{\frac{n}{2}}\frac{\hat{\beta}}{\sqrt{\frac{\sum_{i=1}^{n}(y_{ei}-\bar{y}_{ei})^2+\sum_{i=1}^{n}(y_{ci}-\bar{y}_{ci})^2}{2n-2}}}. \qquad (4)$$

The power function of the test at the 5 percent significance level is

1 - probt(tinv(0.95, 2n-2), 2n-2, $\delta\sqrt{\frac{n}{2}}$)[1], (5)

where probt is the cumulative distribution function for the non-central T; and tinv is the quantile function for the central T (see appendix A for the definition of the functions); 0.95 is equivalent to 1 - the alpha level of the test (we assume 0.05 alpha level from thereon); 2n-2 is the degrees of freedom for the central and non-central T distribution; δ is the standardized effect size $\frac{\beta}{\sigma}$ (Cohen, 1988); $\delta\sqrt{\frac{n}{2}}$ is the non-centrality parameter of the non-central T distribution.

F test

The multi-site clinical trial is a popular two-factor design. The treatment factor is a fixed effect; and its power function can be expressed in terms of the cumulative distributive function of a noncentral F. The random factor, i.e. the site, is a random effect, and its power function can be formulated in terms of the cumulative distributive function of a central F.

The model can be written in the ANOVA notation:

$$y_{ijk} = \mu + \alpha_k + \pi_j + (\alpha\pi)_{jk} + \varepsilon_{i(jk)}, \quad (6)$$

$$\pi_j \sim N(0, \tau^2),\ (\alpha\pi)_{jk} \sim N(0, \sigma^2_{\alpha\pi}),\ \varepsilon_{i(jk)} \sim N(0, \sigma^2)$$

For simplicity we only consider one-sided test and α level of 0.05 throughout the dissertation, though the two-sided test can be derived similarly. All the cumulative distribution functions and their inverse functions from here on use the same notation as SAS. Their definitions are provided in the Appendix A.

where Y_{ijk} is the outcome for the *i*th participant nested within the *j*th site and receiving treatment *k* (i = 1,..., n; j = 1,...,J, k=1,...K). Here μ is the grand mean, α_k is the main effect of treatment k, π_j is the main effect of site j, $\alpha\pi_{jk}$ is the interaction effect between site j and treatment k. Note that π and $\alpha\pi$ are viewed as random effects, and that they are independent.

The test of the treatment effect assumes the null hypothesis that

$$\alpha_1 = \alpha_2 = \cdots = \alpha_k .$$

The test statistic is the ratio of the mean square for the treatment over the mean square for the treatment-by-site interaction. It assumes an F distribution with df for the numerator as k-1, df for the denominator as J-1, and the non-centrality parameter

$$\frac{nJ\sum_{k}^{2}\alpha_k^2}{\sigma^2 + n\sigma_{\alpha\pi}^2} .$$

The power function for the test at α=0.05 is

$$1 - \text{probf}(\text{finv}(0.95, \text{K-1}, \text{J-1}), \text{K-1}, \text{J-1}, \frac{nJ\sum_{k}^{2}\alpha_k^2}{\sigma^2 + n\sigma_{\alpha\pi}^2}), \qquad (7)$$

where probf is the cumulative distribution function for the non-central F; finv is the quantile function for the central F; 0.95 is equal to 1 – alpha level of the test; K-1 is the df for the numerator for the central and non-central F distribution; J-1 is the

df for the denominator for the central and non-central F distribution; $\frac{nJ\sum_{k}^{2}\alpha_k^2}{\sigma^2+n\sigma_{\alpha\pi}^2}$ is the non-centrality parameter for the non-central F distribution.

The test of the random site effect assumes a central F distribution after transformation (see Raudenbush and Liu, 1999). The power function can be expressed as follows:

$$1 - \text{probf}(\text{finv}(0.95, J-1, (n-1)*K*J)* \frac{\sigma^2}{\sigma^2+n\sigma_{\alpha\pi}^2}, J-1, (n-1)*K*J), \qquad (8)$$

where probf is the cumulative distribution function for the central F; finv is the quantile function for the central F; 0.95 is equal to 1 – alpha level of the test; J-1 is the df for the numerator for the central F distribution; (n-1)*K*J is the df for the denominator for the central F distribution; and $\frac{\sigma^2}{\sigma^2+n\sigma_{\alpha\pi}^2}$ times finv function is the quantile parameter for probf.

χ^2 test

The Wald test is a χ^2 test, although it is seldom used. Suppose that

$$Y = X\beta + e, \qquad (9)$$

where $e \sim N(0, \sigma^2 I)$.

The null test can usually be set as

H_0: $A\beta = 0$

$A\hat{\beta}$ assumes a normal distribution, i.e. $A\hat{\beta} \sim N(A\beta, A(X'X)^{-1}A'\sigma^2)$; and the test statistics

$$Q = (A\hat{\beta})'[A(X'X)^{-1}A'\sigma^2]^{-1}(A\hat{\beta}) \sim \chi^2_{rank(A)}(\delta), \qquad (10)$$

where the non-centrality parameter δ is $(A\beta)'[A(X'X)^{-1}A'\sigma^2]^{-1}(A\beta)$.

If we do not know σ^2, we usually substitute its estimate $\hat{\sigma}^2$. The estimate $\hat{\sigma}^2$ has a chi square distribution times a constant. Therefore, the new statistics $\frac{[Q/rank(A)]}{\hat{\sigma}^2/\sigma^2}$ follows a F distribution with non-centrality parameter $\delta = (A\beta)'[A(X'X)^{-1}A'\sigma^2]^{-1}(A\beta)$. It can be proved that the new statistic is a monotonic function of the likelihood ratio statistic (see pp. 110, Stapleton, 1995). It is also noted that $\hat{\beta}$ can be the maximum likelihood estimate or the least square estimate, and that they are identical if e follows the normality and independence assumption.

The power functions can be expressed as:

1 – probchi(cinv(0.95, rank(A)), rank(A), $(A\beta)'[A(X'X)^{-1}A'\sigma^2](A\beta)$), (11)

where probchi is the cumulative distribution function for non-central chi square; and the cinv is the quantile function for central chi square.

The test of parameters in the mixed model provides another example. We can formulate the general model as:

$$y = X\beta + Zu + e \qquad (12)$$

$$\begin{bmatrix} u \\ e \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}\right)$$

$$E(y) = X\beta; \qquad Var(y) = V = ZGZ' + R$$

Test Ho: $K'\beta = 0$

$$(K'\hat{\beta})'[K'(X'V^{-1}X)^{-1}K](K'\hat{\beta}) \quad \sim \chi^2_{rank(K)}(\lambda) \quad \text{(Stroup, 1998)}$$

Therefore, the power function can be expressed as:

1 – probchi(cinv(0.95, rank(K)), rank(K), $(K'\beta)'[K'(X'V^{-1}X)^{-1}K](K'\beta)$), (13)

where V is usually not known but may be replaced by its estimate from a previous study. When the estimate is based on a large sample, the chi square test is approximately true and the formulas is still applicable.

Chapter 2

COMPUTATION OF POWER IN THE HIERARCHILCAL LINEAR MODEL THROUGH THE ANOVA TABLE

Researches in education and mental health often involve the use of people as subjects. People are usually situated within some social and clinical settings. For example, they may be nested in schools or in community health centers. Randomly assigning individual people into experimental conditions is sometimes not ethically and logistically possible. However, groups of people who are geographically or socially related may be randomly assigned into experimental conditions. In school-based experiments, classrooms are often assigned into treatment or control conditions; in mental health research clusters of patients who attend the same clinic are assigned into a new therapy or control treatment (Raudenbush, 1997). Those experiments can be analyzed using the HLM, and they are closely related to mixed ANOVA.

Under the analytical framework of ANOVA the effect of a factor may be considered as either fixed or random. Mean squares may be computed for the factors in the model, and the test of each factor may be constructed according to its expected mean square and those of other factors. The power functions can then be constructed from their respective tests from the ANOVA table.

The idea of using the expected mean square ANOVA table assumes a balanced design. For a given sample size with equal cost and variance per treatment a balanced design yields the maximum power. In planning a study a balanced design is usually chosen. If the cost of the study is at issue, then an unbalanced design may achieve better power through optimal sample allocation. For example, we may enlarge the total sample by recruiting more subjects that cost less. Of course, the optimality issue is beyond the scope of the dissertation; and we limit our inquiry to balanced designs. We will discuss unbalanced designs, which result from missing data in the concluding section. In short, the assumption of a balanced design makes it possible to relate HLM to ANOVA and to plan a study using HLM with the aid of the power functions from ANOVA tables.

Scheffe (1959), Bennett and Franklin (1954), and Searle (1971) all provide general rules on deriving the ANOVA table of an experiment be it fixed factorial or mixed design. The correctness of the derived table crucially depends on including every effect term and the correct subscripts of each effect term in the model. Yet, it is very easy to miss an effect term or some subscripts as the design gets complex. The derivation of the ANOVA table for high-order mixed designs requires extreme meticulousness and patience. People commonly are unsuccessful. Moreover, the only way to check the correctness of the final results is to repeat the same process.

In the following a new set of rules will be introduced to simplify the procedure in deriving ANOVA tables of mixed design and the statistical power functions of the relevant tests. The rules are largely based on the equivalence between mixed linear model (hierarchical linear model) and mixed ANOVA model. Raudenbush (1993) shows that a hierarchical linear model produces through the restricted maximum likelihood method the same results as the traditional ANOVA when the nested random effects design is balanced.

For simplicity we will consider only the typical case of a hierarchical linear model, that is, each level has a constant and a random effect with the option of a categorical covariate. The constant and categorical covariate become the fixed terms in ANOVA; and the random effect of each level transforms into a random term or random interaction term in the ANOVA model. The definition of fixed and random terms carries the same meaning for both HLM and ANOVA. The random term in HLM usually results from randomly sampling units at a particular level. It denotes the random variation among the units at that level after accounting for the effect of the covariate of that level. Each random term in the HLM corresponds to a unique random effect in the ANOVA model. In the latter model the random term may change into a random interaction term. Its meaning becomes less clear than its counterpart in the HLM model. The fixed effect in the HLM model can either be an experimental condition or a classifying variable which takes a finite number of values. When translated into the ANOVA model those fixed effect terms may become fixed interaction terms. Although the

following-stated rules apply to the HLM with a constant, an optional covariate, and a random effect at each level, extension to a more general case can easily be made with slight modification.

Procedures on writing the terms and their subscripts in an ANOVA model

The steps and rules are stated with reference to an example in which the cluster randomized trial is replicated across a number of sites and sites are classified by a dichotomous characteristic. A hypothetical educational study may be constructed for this design. The site can be the school and the cluster can be the classroom. At each school classrooms can be randomly assigned into two different counseling types. Schools are then classified as public or catholic.

This particular design is chosen for two reasons. First, it includes both crossing and nesting and bears the feature of a split-plot design. The site is the "plot." "Plots" are classified by a fixed effect, and there is a randomized experiment in each "plot". Second, the inclusion of cluster randomized trials at each "plot" complicates the design and illustrates the generality of those rules.

Step 1: write down the design in terms of a hierarchical linear model (HLM)

Experimental units are specified as levels in HLM. A fixed effect term is included at the level where random assignment into treatment conditions occurs or a

covariate is introduced (For simplicity all fixed factors and covariates are treated as dichotomous. The results generalize.) At each level by default an observation is a mean value plus its random error of that level.

The "site" is the largest randomly sampled unit and is situated at the highest level. It might be schools in educational research. The "cluster" can be classrooms in each school. The "individual" can be students in a certain classroom. The subscript denotes either *k*th, or *j*th, or *l*th unit at the corresponding level.

Level	subscript	Linear model at each level[2]
Site	k	$\beta_{00k} = \gamma_{000} + \gamma_{001} W + u_{00k}$ $\beta_{01k} = \gamma_{010} + \gamma_{011} W + u_{01k}$
Cluster	j	$\pi_{0jk} = \beta_{00k} + \beta_{01k} X + r_{0jk}$
individual	l	$Y_{ljk} = \pi_{0jk} + e_{ljk}$

Table 1: HLM formulation of the example

Combined model:

$$Y_{ljk} = \gamma_{000} + \gamma_{001} W + u_{00k} + \gamma_{010} X + \gamma_{011} WX + u_{01k} X + r_{0jk} + e_{ljk} \qquad (14)$$

[2] Since the HLM terms are only used to identify their ANOVA counterparts, we omit the subscripts for the HLM terms for simplicity.

Step 2: write down ANOVA model with reference to the terms in HLM notation

Rules on naming terms in HLM: The random effect at each level is named after the variable name of that level. For example: r is the cluster effect and U is the site effect. Fixed effects are determined by dichotomous variables in the model, i.e. a covariate at each level, and coefficients do not have bearing on naming the terms in the model. Ignoring the coefficients, we can easily identify the interaction term by looking at the presence of random terms and fixed terms in the model. The term map is provided for this example.

Term	Name	ANOVA term
$\gamma_{001}W$	Effect of site characteristic	η
u_{00k}	Effect of site	γ
$\gamma_{010}X$	Effect of treatment	α
$\gamma_{011}WX$	Effect of treatment*site characteristic	$\alpha\eta$
$u_{01k}X$	Effect of treatment*site	$\alpha\gamma$
r_{0jk}	Effect of cluster	β
e_{ijk}	Within cell error	ε

Table 2: translation between HLM and ANOVA terms

Each term in the map table has its corresponding counterpart in the ANOVA model. The order of the terms in HLM is kept the same in ANOVA. The letter for the ANOVA terms may be different to conform to convention; the subscripts in HLM can be adopted to show the relation between HLM and ANOVA. The ANOVA model for this example is as below:

$$Y = \mu + \eta + \gamma + \alpha + \alpha\eta + \alpha\gamma + \beta + e. \qquad (15)$$

Step 3: add subscripts to the terms in the ANOVA model

Rule 1: Attach subscript i to the treatment effect and any letter, say, h other than the already used subscript to site characteristic. Attach the same subscripts as in the HLM notation to the single letter terms.

$$Y = \mu + \eta_h + \gamma_k + \alpha_i + \alpha\eta + \alpha\gamma + \beta_j + e_l. \qquad (16)$$

Rule 2: If there is treatment or covariate effect at a level in the HLM, then the units of that level are nested within the treatment or covariate, and the units of the lower levels are said to be nested within the units and its covariate at the higher levels.

For this example, the cluster is nested within the treatment because level 2 has a treatment term in the HLM; site is nested in site characteristic because site characteristic is the covariate at the site level. In addition, the cluster is said to

be nested within the site and the site characteristic because the cluster is at a level below the site level. In summary, the cluster is nested within the treatment, the site, and the site characteristic. The site is nested in the site characteristic. The individual is nested in the cluster, the treatment, the site, and the site characteristic because the individual sits lower than any other levels in the HLM representation.

Rule 3: If an effect is nested within other effects, add the subscripts of the other effects in parenthesis after its own subscript.

$$Y = \mu + \eta_h + \gamma_{k(h)} + \alpha_i + \alpha\eta + \alpha\gamma + \beta_{j(ikh)} + e_{l(ijkh)}. \qquad (17)$$

Rule 4: The subscript for the interaction term is the combination of subscripts of those terms which constitute the interaction.

$$Y = \mu + \eta_h + \gamma_{k(h)} + \alpha_i + \alpha\eta_{ih} + \alpha\gamma_{ik(h)} + \beta_{j(ikh)} + e_{l(ijkh)}. \qquad (18)$$

In addition, if there is a high level above the level where treatment occurs, then the treatment is said to be crossed with the high level.

NOVA

e the mean square of

effect to its right (see

row points at the term

term from which the

(19)

square of the random

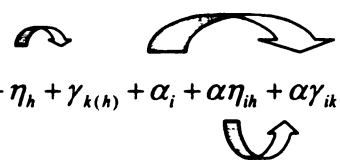
ct on its right without

n square divides the

(20)

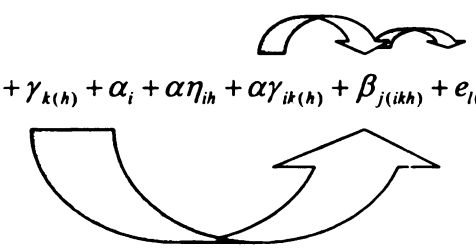
Rules on constructing tests for mixed ANOVA

Rule 5: For the test of a fixed effect or interaction, we divide the mean square of that term by the mean square of the next available random effect to its right (see the arrows on the following equation for illustration). The arrow points at the term whose mean square divides the mean square of the other term from which the arrow comes.



$$Y = \mu + \eta_h + \gamma_{k(h)} + \alpha_i + \alpha\eta_{ih} + \alpha\gamma_{ik(h)} + \beta_{j(ikh)} + e_{l(ijkh)}. \qquad (19)$$

Rule 6: For the test of a random effect, we divide the mean square of the random effect by the mean square of the next closest random effect on its right without the same letter. The arrow points at the term whose mean square divides the mean square of the other term from which the arrow comes.



$$Y = \mu + \eta_h + \gamma_{k(h)} + \alpha_i + \alpha\eta_{ih} + \alpha\gamma_{ik(h)} + \beta_{j(ikh)} + e_{l(ijkh)} \qquad (20)$$

Procedures and rules for deriving the ANOVA table and statistical power functions

With the prior knowledge of which mean square should be tested against which mean square, the expected mean square of each term may be easily derived without resort to the tedious task of establishing a subscript table.

Step 1: Construct a frame of an ANOVA table, add the subscript and its number of levels to the right of a single letter term, write down the df according to their subscripts, namely, the product of the number of levels associated with each subscript in the parenthesis and the number of levels minus one associated with each other subscript(s).

	Name	term	df	EMS
$h:d$	Effect of site characteristic	η_h	$d-1$	
$k:c$	Effect of site	$\gamma_{k(h)}$	$(c-1)d$	
$i:a$	Effect of treatment	α_i	$a-1$	
	Effect of treatment*site characteristic	$\alpha\eta_{ih}$	$(a-1)(d-1)$	
	Effect of treatment*site	$\alpha\gamma_{ik(h)}$	$(a-1)(c-1)d$	
$j:b$	Effect of cluster	$\beta_{j(ikh)}$	$a(b-1)cd$	
$l:n$	Within cell error	$\varepsilon_{l(ijkh)}$	$abcd(n-1)$	

Table 3: ANOVA table of the example

Rule 7: Each term in the ANOVA model has either a variance component (random effect) or fixed factor (fixed effect associated with it). The variance component for the random effect is σ^2_{term}, which is subscripted by its term. The fixed effect is represented by the sum of squares of the model components associated with the factor divided by its degree of freedom (Montgomery, 1996).

Rule 8: The expected mean square of within cell error is always σ^2. The EMS of the random term which should be tested against within cell error is σ^2 plus the variance component of that random term times the product of the numbers of levels of the other subscripts which are not present on the term. Likewise, the previously derived EMS of random term is used to write down the EMS of other random or fixed term whose test uses the MS of the previous random term as the denominator in the F test. The EMS of new random or fixed term is the EMS of the previous random term plus the variance or fixed effect of the new term times the product of the numbers of levels of those subscripts which are not on the new term.

In the case of the current example, the derivation is illustrated as below in 3 steps:

	Name	term	df	EMS
$h:d$	Effect of site characteristic	η_h	$d-1$	
$k:c$	Effect of site	$\gamma_{k(h)}$	$(c-1)d$	
$i:a$	Effect of treatment	α_i	$a-1$	
	Effect of treatment*site characteristic	$\alpha\eta_{ih}$	$(a-1)(d-1)$	
	Effect of treatment*site	$\alpha\gamma_{ik(h)}$	$(a-1)(c-1)d$	
$j:b$	Effect of cluster	$\beta_{j(ikh)}$	$a(b-1)cd$	$\sigma^2+n\sigma_\beta^2$
$l:n$	Within cell error	$\varepsilon_{l(ijkh)}$	$abcd(n-1)$	σ^2

Table 4: derivation of EMS—step 1

	Name	term	df	EMS
$h:d$	Effect of site characteristic	η_h	$d-1$	
$k:c$	Effect of site	$\gamma_{k(h)}$	$(c-1)d$	$\sigma^2+n\sigma^2_\beta+abn\sigma^2_\gamma$
$i:a$	Effect of treatment	α_i	$a-1$	
	Effect of treatment*site characteristic	$\alpha\eta_{ih}$	$(a-1)(d-1)$	
	Effect of treatment*site	$\alpha\gamma_{ik(h)}$	$(a-1)(c-1)d$	$\sigma^2+n\sigma^2_\beta+bn\sigma^2_{\alpha\gamma}$
$j:b$	Effect of cluster	$\beta_{j(ikh)}$	$a(b-1)cd$	$\sigma^2+n\sigma^2_\beta$
$l:n$	Within cell error	$\varepsilon_{l(ijkh)}$	$abcd(n-1)$	σ^2

Table 5: derivation of EMS—step 2

	Name	term	df	EMS
$h:d$	Effect of site characteristic	η_h	$d-1$	
$k:c$	Effect of site	$\gamma_{k(h)}$	$(c-1)d$	$\sigma^2+n\sigma^2_\beta+abn\sigma^2_\gamma$
$i:a$	Effect of treatment	α_i	$a-1$	$\sigma^2+n\sigma^2_\beta+bn\sigma^2_{\alpha\gamma}+bcdn\frac{\sum_i^a \alpha_i^2}{a-1}$
	Effect of treatment*site characteristic	$\alpha\eta_{ih}$	$(a-1)(d-1)$	$\sigma^2+n\sigma^2_\beta+bn\sigma^2_{\alpha\gamma}+bcn\frac{\sum\sum\alpha\eta^2_{ih}}{(a-1)(h-1)}$
	Effect of treatment*site	$\alpha\gamma_{ik(h)}$	$(a-1)(c-1)d$	$\sigma^2+n\sigma^2_\beta+bn\sigma^2_{\alpha\gamma}$
$j:b$	Effect of cluster	$\beta_{j(ikh)}$	$a(b-1)cd$	$\sigma^2+n\sigma^2_\beta$
$l:n$	Within cell error	$\varepsilon_{l(ijkh)}$	$abcd(n-1)$	σ^2

Table 6: derivation of EMS—step 3

	Name	term	df	EMS
$h:d$	Effect of site characteristic	η_h	$d-1$	$\sigma^2+n\sigma_\beta^2+abn\sigma_\gamma^2+abcn\frac{\sum_h^d \eta_h^2}{h-1}$
$k:c$	Effect of site	$\gamma_{k(h)}$	$(c-1)d$	$\sigma^2+n\sigma_\beta^2+abn\sigma_\gamma^2$
$i:a$	Effect of treatment	α_i	$a-1$	$\sigma^2+n\sigma_\beta^2+bn\sigma_{\alpha\gamma}^2+bcdn\frac{\sum_i^a \alpha_i^2}{a-1}$
	Effect of treatment*site characteristic	$\alpha\eta_{ih}$	$(a-1)(d-1)$	$\sigma^2+n\sigma_\beta^2+bn\sigma_{\alpha\gamma}^2+bcn\frac{\sum\sum \alpha\eta_{ih}^2}{(a-1)(h-1)}$
	Effect of treatment*site	$\alpha\gamma_{ik(h)}$	$(a-1)(c-1)d$	$\sigma^2+n\sigma_\beta^2+bn\sigma_{\alpha\gamma}^2$
$j:b$	Effect of cluster	$\beta_{j(ikh)}$	$a(b-1)cd$	$\sigma^2+n\sigma_\beta^2$
$l:n$	Within cell error	$\varepsilon_{l(ijkh)}$	$abcd(n-1)$	σ^2

Table 7: final ANOVA table of the example

The above ANOVA table is the same as directly derived according to the old rules.

With the aid of the ANOVA table, the power function of the tests may be easily constructed. We provide the power functions for this example as below:

The power function for testing the site characteristic is

$$1 - \text{probf}(\text{finv}(0.95,\ d\text{-}1,\ (c\text{-}1)d),\ d\text{-}1,\ (c\text{-}1)d,\ \frac{abcn\sum_{h}^{d}\eta_h^2}{\sigma^2 + n\sigma_\beta^2 + abn\sigma_\gamma^2}). \quad (21)$$

The power function for testing the site effect is

$$1 - \text{probf}(\text{finv}(0.95,\ (c\text{-}1)d,\ a(b\text{-}1)cd)^* \frac{\sigma^2 + n\sigma_\beta^2}{\sigma^2 + n\sigma_\beta^2 + abn\sigma_\gamma^2},$$

$$(c\text{-}1)d,\ a(b\text{-}1)cd). \quad (22)$$

The power function for testing the treatment is

$$1 - \text{probf}(\text{finv}(0.95,\ a-1,\ (a-1)(c-1)d,\ \frac{bcdn\sum_{i}^{a}\alpha_i^2}{\sigma^2 + n\sigma_\beta^2 + bn\sigma_{\alpha\gamma}^2}). \quad (23)$$

The power function for testing the treatment-by-site characteristics is

$$1 - \text{probf}(\text{finv}(0.95,\ (a-1)(d-1),\ (a-1)(c-1)d,\ \frac{bcn\sum\sum\alpha\eta_{ih}^2}{\sigma^2 + n\sigma_\beta^2 + bn\sigma_{\alpha\gamma}^2}). \quad (24)$$

The power function for testing the treatment-by-site variance is

$$1 - \text{probf}(\text{finv}(0.95,\ (a-1)(c-1)d,\ a(b-1)cd)^* \frac{\sigma^2 + n\sigma_\beta^2}{\sigma^2 + n\sigma_\beta^2 + bn\sigma_{\alpha\gamma}^2},$$

$$(a-1)(c-1)d,\ a(b-1)cd). \quad (25)$$

The power function for testing the cluster variance is

$$1 - \text{probf}(\text{finv}(0.95,\ a(b-1)cd,\ abcd(n-1))^* \frac{\sigma^2}{\sigma^2 + n\sigma_\beta^2},$$

$$a(b-1)cd,\ abcd(n-1)). \quad (26)$$

The same power functions in HLM can be constructed from a similar HLM table.

The HLM has all the parameters in HLM notation.

Chapter 3

COMPUTATION OF POWER IN HIERARCHICAL LINEAR MODEL THROUGH HLM TABLE

In this chapter proof will be given to support the rules we have used in the previous chapter. The same rules will be used to construct the HLM table and derive the power functions directly. The sample example in the previous chapter is used to construct a HLM table.

Proof

The rules we have used in the previous chapter follow three principles. First, the parameters of fixed effects are always tested against the random term at the same level in HLM. Second, the random term at the higher level is always tested against the random term at the next lower level. Third, the conditional hierarchical linear model is equivalent to mixed ANOVA model; and each term in the HLM has its corresponding term in the mixed ANOVA.

HLM can be divided into individual levels. At each level it takes the form of either

Outcome = mean + random component

or

Outcome = mean + parameter*fixed effect + random component.

We simplify the proof by having the "fixed effect" take only two levels, i.e. treatment or control, though the proof can be easily adapted to the fixed effect of

more than one level. Essentially, each level of HLM can be either considered as requiring a one sample t test or two sample independent t test. We will first prove that the first principle holds true at the either of the two forms at the 2nd level. We continue to prove that the first principle is true at the n+1 level. Similarly, we will prove the second principle at the 2nd level, and then we generalize it to the n+1 level. We will give reference to the third principle.

Case 1: The 2nd level has a fixed effect

Level 2: $\beta_{0j} = \gamma_0 + \gamma_1 w_j + u_j, \quad u_j \sim N(0,\tau) \quad j = 1,...,J$ (27)

Level 1: $Y_{ij} = \beta_{0j} + r_{ij}, \quad r_{ij} \sim N(0,\sigma^2) \quad i = 1,....n$ (28)

w_j takes -½ or ½ for control and experimental conditions.

$\hat{\beta}_{0j} = \bar{Y}_{.j}$ and $\bar{Y}_{.1}$ ... $\bar{Y}_{.J}$ are independent and have variance $\frac{\sigma^2}{n} + \tau$.

The test of parameter γ_1 is the same as the two sample independent t test. The power of the test of γ_1 is

$$1 - \text{probt}(\text{tinv}(0.95, J\text{-}2), J\text{-}2, \frac{\gamma_1}{\sqrt{\frac{4}{J}(\frac{\sigma^2}{n} + \tau)}}). \quad (29)$$

So the estimated γ_1 is tested against the estimated variance of the error term u_j.

Alternatively,

$$\sum^{J}(\bar{Y}_{.j} - \bar{\bar{Y}}_{..})^2 = \frac{J}{2}\sum_{k=E,C}(\bar{Y}_{K.j} - \bar{\bar{Y}}_{..})^2 + \sum_{k=E,C}\sum_{j}^{J/2}(\bar{Y}_{k.j} - \bar{\bar{Y}}_{E..})^2 \quad (30)$$

and, that is, TSS = SSB + SSE,

where SSB is the between sum of squares and SSE is the within sum of square as in the one-way ANOVA.

$$E(\frac{SSE}{J-2}) = \sigma^2_{\bar{Y}_j} = \frac{\sigma^2}{n} + \tau \text{, and } \frac{SSE}{J-2} \sim (\frac{\sigma^2}{n} + \tau)\chi^2_{J-2}. \quad (31)$$

$$E(\frac{SSB}{2-1}) = \frac{\sigma^2}{n} + \tau + \frac{J\gamma_1}{4} \text{, and } \frac{SSB}{2-1} \sim (\frac{\sigma^2}{n} + \tau)\chi^2_1(\delta = \frac{J\gamma_1^2}{4(\frac{\sigma^2}{n} + \tau)}). \quad (32)$$

The test of γ_1 is F=$\frac{SSB}{SSE/(J-2)}$. It follows a F distribution. The power function is

$$1 - \text{probf(finv(0.95, 1, } J\text{-2), } J\text{-2, } \frac{J\gamma_1^2}{4(\frac{\sigma^2}{n} + \tau)}). \quad (33)$$

It is equivalent to the previous power function (see appendix B). Therefore, the first principle is true for the 2nd level in the first case.

From text books on experimental design (see Montgomery, 1997), the estimated $\hat{\sigma}^2$ always has the expectation σ^2 and follows $\sigma^2\chi^2$ distribution when the design is balanced. We know from the above that

$$E(nSSE/(J-2)) = \sigma^2 + n\tau. \quad (34)$$

The test statistics for τ is $\frac{nSSE/(J-2)}{\hat{\sigma}^2}$. After transformation, it follows a central F distribution. The power function is

$$1 - \text{probf(finv(0.95, } J\text{-2, } Jn\text{-}J)^* \frac{\sigma^2}{\sigma^2 + n\tau}, J\text{-2, } Jn\text{-}J). \quad (35)$$

Therefore, the second principle holds true too for the 2nd level.

Now we can generalize the proof to the n+1 level. The model can be formulated as follows:

Level n+1:

$$\beta_{0j}^{n+1} = \gamma_0^{n+1} + \gamma_1^{n+1} w_j^{n+1} + u_j^{n+1}, \quad u_j \sim N(0, \tau^{n+1}) \quad j = 1,...,J \tag{36}$$

Level n:

$$Y_{ij}^{n} = \beta_{0j}^{n} + r_{ij}^{n}, \quad Var(Y_{ij}^{2}) = \sigma_n^2 \quad i = 1,....n \tag{37}$$

w_j^{n+1} takes -½ or ½ for control and experimental conditions.

$\hat{\beta}_{0j}^{n+1} = \bar{Y}_{.j}^{n+1}, \bar{Y}_{.1}^{n+1} \ \ldots \ \bar{Y}_{.J}^{n+1}$ are independent since the nth level are nested in the n+1 level. The test of γ_1^{n+1} is the same as a two sample independent t test. The test statistic has a T distribution with degree of freedom *J-2* and a non-centrality parameter $\frac{\gamma_1^{n+1}}{\sqrt{\frac{4}{J}(\frac{\sigma_n^2}{n} + \tau^{n+1})}}$. The same results duplicate in the n+1 level. Therefore the first and second principle are proved.

Case 2: The level does not have a fixed effect.

Level 2: $\beta_{0j} = \gamma_{00} + u_{0j}$ (38)

$$\beta_{1j} = \gamma_{10} + u_{1j} \tag{39}$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \tau_{00} & \tau_{10} \\ \tau_{01} & \tau_{11} \end{bmatrix} \right)$$

Level 1:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij} \qquad r_{ij} \sim i.i.d.N(0, \sigma^2) \tag{40}$$

$\hat{\beta}_{0j} = \bar{Y}_{.j}$ and $\bar{Y}_{.1} \ldots \bar{Y}_{.J}$ are independent and have variance $\frac{\sigma^2}{n} + \tau_{00}$.

$\hat{\beta}_{1j} = \bar{Y}_{E.j} - \bar{Y}_{C.j} = \Delta\bar{Y}_{.j}$,

where $\bar{Y}_{E.j}$ is the mean for the experimental group at the jth site; and $\bar{Y}_{C.j}$ is the mean for the control group at the jth site. $\Delta\bar{Y}_{.1}...\Delta\bar{Y}_{.J}$ are independent and have variance $\frac{4\sigma^2}{n} + \tau_{11}$.

The test of γ_{00} and γ_{10} are the same as one sample t test. $\hat{\beta}_{0j}$ and $\hat{\beta}_{1j}$ are the corresponding observed scores. The estimates are:

$$\hat{\gamma}_{00} = \frac{\sum^{J} \bar{Y}_{.j}}{J} \qquad \hat{Var}(\bar{Y}_{.j}) = \frac{\sum^{J} (\bar{Y}_{.j} - \bar{\bar{Y}}_{..})^2}{J-1} \qquad E(\hat{Var}(\bar{Y}_{.j})) = \sigma^2_{\bar{Y}_{.j}} = \frac{\sigma^2}{n} + \tau_{00}; \quad (41)$$

$$\hat{\gamma}_{10} = \frac{\sum^{J} \Delta\bar{Y}_{.j}}{J} \qquad \hat{Var}(\Delta\bar{Y}_{.j}) = \frac{\sum^{J} (\Delta\bar{Y}_{.j} - \Delta\bar{\bar{Y}}_{..})^2}{J-1} \qquad E(\hat{Var}(\Delta\bar{Y}_{.j})) = \sigma^2_{\Delta\bar{Y}_{.j}} = \frac{4\sigma^2}{n} + \tau_{11}$$

(42)

The power function for the test of γ_{10} is

$$1 - \text{probt}(\ \text{tinv}(0.95,\ J\text{-}1),\ J\text{-}1,\ \frac{\gamma_{10}}{\sqrt{(\frac{4\sigma^2}{n} + \tau_{11})/J}}\ ). \qquad (43)$$

Therefore the first principle holds true in this case.

It should be noted that the estimates

$$\hat{Var}(\bar{Y}_{.j}) = \frac{\sum^{J} (\bar{Y}_{.j} - \bar{\bar{Y}}_{..})^2}{J-1}, \quad \hat{\gamma}_{10} = \frac{\sum^{J} \Delta\bar{Y}_{.j}}{J}, \qquad \hat{Var}(\Delta\bar{Y}_{.j}) = \frac{\sum^{J} (\Delta\bar{Y}_{.j} - \Delta\bar{\bar{Y}}_{..})^2}{J-1}$$

are algebraically related to the mean squares for the terms in the ANOVA, which correspond to the parameters u_{0j}, γ_{10}, u_{1j} in the HLM model.

Now we prove that the second principle is true in this case. The estimated $\hat{\sigma}^2$ again has the expectation σ^2 and follows $\sigma^2\chi^2$ distribution.

$\hat{Var}(\bar{Y}_{.j}) = \frac{\sum^{J}(\bar{Y}_{.j} - \bar{\bar{Y}}_{..})^2}{J-1}$ has a $(\frac{\sigma^2}{n} + \tau_{00})\chi^2_{J-1}$ distribution. The test statistics for the parameter u_{0j} is $\frac{n\hat{Var}(\bar{Y}_{.j})}{\hat{\sigma}^2}$. The numerator is equivalent to the mean square for the site effect in the ANOVA model. After transformation, the statistics has a central F distribution. The power function for the test of u_{0j} is

$$1 - \text{probf}(\text{ finv}(0.95, J\text{-}1, nJ\text{-}2J)^* \frac{\sigma^2}{\frac{\sigma^2}{n} + \tau_{00}}, J\text{-}1, nJ\text{-}2J). \qquad (44)$$

$\hat{Var}(\Delta\bar{Y}_{.j}) = \frac{\sum^{J}(\Delta\bar{Y}_{.j} - \Delta\bar{\bar{Y}}_{..})^2}{J-1}$ has a $(\frac{4\sigma^2}{n} + \tau_{11})\chi^2_{J-1}$. The test statistics for the parameter u_{1j} is $\frac{4/n(\hat{Var}(\Delta\bar{Y}_{.j})}{\hat{\sigma}^2}$. It is the same as the mean square ratio of treatment-by-site interaction over within-cell error. After transformation, the test statistics has a central F distribution. The power function is

$$1 - \text{probf}(\text{ finv}(0.95, J\text{-}1, Jn\text{-}2J)^* \frac{\sigma^2}{\sigma^2 + n\tau_{11}/4}, J\text{-}1, Jn\text{-}2J). \qquad (45)$$

For the n+1 level, we add n+1 superscripts to the parameters in the model. Essentially the proof is the same as we have provided in the previous case.

Creating the HLM table

With the aid of the three principles we may construct an HLM table similar to ANOVA table and derive the power functions directly. We use the same example from the previous chapter to illustrate the construction of a HLM table.

First, we put all the terms in the combined model into the table and write the subscripts and their corresponding total number of levels on the leftmost column. Second, we use the same rules to construct the subscript for each term and derive the degrees of freedom as we have done in the ANOVA table. Third, we fill in the parameters for the random and fixed effects. There are three rules to do so (for simplicity, we restrain all the covariates to take two values):

1. If the outcome at one level is the coefficient of the covariate at the next lower level, divide the variance of the random component of that level by 2. Otherwise, we write down the variance of the random component.
2. For the fixed effect we write down the corresponding coefficient, square it , and then divide it by 2.
3. For the fixed interaction effect we write down the corresponding coefficient, square it, and then divide it by 4.

Fourth, we refer to the rule 8 in the previous chapter to write down the numerator or denominator of the non-centrality parameter for the power functions.

With the aid of the HLM table, we can derive power functions for the tests of the fixed and random effects. For the test of the random component the construction of the power function is the same as in the previous chapter. For the fixed effect all the power functions use a non-central t distribution. The non-centrality parameter is the square root of the ratio of the piece for the fixed effect in the last column over the expectation of the random term, against which the fixed effect is tested. In general, non-centrality parameter =sqrt[(numerator of non-centrality parameter for the fixed effect)/(denomenator of non-centrality for the random effect)].

The HLM table for the example in the previous chapter is provided as below. The subscripts are kept the same to highlight the translation between ANOVA and HLM. In the following chapter all the HLM tables adopt the HLM subscripts and notation. Here *h* denotes for site characteristic; *k* for site; *i* for treatments; *j* for cluster; and *l* for individual. *a* is the number of levels for the treatment factor; *b* for the cluster factor; *c* for the site factor; *d* for the site characteristic factor; *n* for the within-cell error.

	Term	Sub-script	df	Param-eter	Numerator or denominator of non-centrality parameter
$h:d$	$\gamma_{001}W$	$h-1$	$d-1$	$\frac{\gamma_{001}^2}{2}$	$abcn\frac{\gamma_{001}^2}{2}$
$k:c$	u_{00k}	$k(h)$	$(c-1)d$	$\tau_{\beta 00}$	$\sigma^2+b\tau_{\pi}+abn\tau_{\beta 00}$
$i:a$	$\gamma_{010}X$	i	$a-1$	$\frac{\gamma_{010}^2}{2}$	$bcdn\frac{\gamma_{010}^2}{2}$
	$\gamma_{011}WX$		$(a-1)(d-1)$	$\frac{\gamma_{011}^2}{4}$	$bcn\frac{\gamma_{011}^2}{4}$
	$u_{01k}X$	$ik(l)$	$(a-1)(c-1)$	$\frac{\tau_{\beta 01}}{2}$	$\sigma^2+n\tau_{\pi}+bn\frac{\tau_{\beta 01}}{2}$
$j:b$	r_{0jk}	$b(kh)$	$a(b-1)cd$	τ_{π}	$\sigma^2+n\tau_{\pi}$
$l:n$	e_{ijk}	$l(ijkh)$	$abcd(n-1)$	σ^2	σ^2

Table 8: HLM table of the example

The power function for testing the treatment is

$$1 - \text{probf}(\text{tinv}(0.95,\ (a-1)(c-1)d),\ (a-1)(c-1)d,\ \sqrt{\frac{bcdn\frac{\gamma_{010}^{2}}{2}}{\sigma^2 + n\tau_{\pi} + bn\frac{\tau_{\beta 01}}{2}}}). \quad (46)$$

The power function for testing the site characteristic is

$$1 - \text{probt}(\text{tinv}(0.95,\ (c-1)d),\ (c-1)d,\ \sqrt{\frac{abcn\frac{\gamma_{001}^{2}}{2}}{\sigma^2 + n\tau_{\pi} + abn\frac{\tau_{\beta 00}}{2}}}). \quad (47)$$

The power function for testing the treatment-by-site characteristics is

$$1 - \text{probt}(\text{tinv}(0.95,\ (a-1)(c-1)d),\ (a-1)(c-1)d,\ \sqrt{\frac{bcn\frac{\gamma^{2}_{011}}{4}}{\sigma^2 + n\tau_{\pi} + bn\tau_{\beta 01}}}). \quad (48)$$

Chapter 4

FOUR EXAMPLES

We provide the HLM tables for three different designs: a 3-level cluster randomized trial, a multi-site clinical trial with a site covariate, and a combination of cluster randomized trial and multi-site clinical trial (cluster randomized trial replicated across multi-sites). For each design the power functions are given for the tests of the parameters in the model. Finally we provide the power function for a potential HLM analysis based on Tennessee classroom size study (Finn & Achilles, 1990).

3-level cluster randomized trial

In school-based intervention studies schools are randomly assigned into treatment and control condition. Classrooms are nested within each school. The design can be formulated as a 3-level HLM model:

Level 3:

$$\beta_{00k} = \gamma_{000} + \gamma_{001} w_k + u_{00k}, \quad u_j \sim N(0, \tau_\beta) \qquad k = 1,..., K \qquad (49)$$

Level 2:

$$\pi_{0jk} = \beta_{00k} + r_{0jk} \qquad r_{0jk} \sim N(0, \tau_\pi) \qquad j = 1,..., J \qquad (50)$$

Level 1:

$$Y_{ijk} = \pi_{0jk} + e_{ijk}, \quad e_{ijk} \sim N(0, \sigma^2) \qquad i = 1,....n \qquad (51)$$

w_k takes -½ or ½ for control and experimental conditions.

	Term	Subscript	df	Parameter	Numerator or denominator of non-centrality parameter
$l:2$	$\gamma_{001} w_k$	l	2-1	$\frac{\gamma_{001}^2}{2}$	$JKn\frac{\gamma_{001}^2}{2}$
$k:K$	u_{00k}	$k(l)$	$2(K-1)$	τ_β	$\sigma^2 + n\tau_\pi + nJ\tau_\beta$
$j:J$	r_{0jk}	$j(kl)$	$2K(J-1)$	τ_π	$\sigma^2 + n\tau_\pi$
$i:n$	e_{ijk}	$i(jkl)$	$2JK(n-1)$	σ^2	σ^2

Table 9: HLM table of 3 level CRT

The power function for the test of γ_{001} is

$$1 - \text{probt}(\text{ tinv}(0.95,\ 2(K-1)),\ 2(K-1),\ \sqrt{\frac{JKn\frac{\gamma_{001}^2}{2}}{\sigma^2 + n\tau_\pi + nJ\tau_\beta}}\ ).\ ^{3} \qquad (52)$$

The power function for the test of τ_β is

$$1 - \text{probf}(\text{ finv}(0.95,\ 2(K-1),\ 2K(J-1))^* \frac{\sigma^2 + n\tau_\pi}{\sigma^2 + n\tau_\pi + nJ\tau_\beta},$$

$$2(K-1),\ 2K(J-1)\ ). \qquad (53)$$

[3] A SAS programs can be used to compute the power value of any listed functions and plot a power curve. The functions input are exactly the same as provided (see Appendix C, the SAS programs).

The power function for the test of τ_π is

$$1 - \text{probf}(\ \text{finv}(0.95,\ 2K(J-1),\ 2JK(n-1)\ )^* \frac{\sigma^2}{\sigma^2 + n\tau_\pi},$$

$$2K(J-1)),\ 2JK(n-1)). \qquad (54)$$

Multi-site clinical trial with a site covariate

The multi-site clinical trial is widely used in mental health research. Patients are randomly assigned into the treatment or control condition at each clinical site. The same study is replicated across a number of clinical sites. The key interests surround the average treatment effect across the sites and the variability of treatment effect among the sites (Raudenbush & Liu, 1999). When the treatment effects vary significantly across sites, it usually implies that the fluctuation of the treatment effect is not simply random but is related to some characteristic of the sites, i.e. a site covariate. The model may be formulated as a 3-level HLM:

Level-2:

$$\beta_{0j} = \theta_{00} + \theta_{01} w_j + u_{0j} \qquad u_{0j} \sim N(0, \tau_{00}) \qquad j = 1,...,J \qquad (55)$$

$$\beta_{1j} = \theta_{10} + \theta_{11} w_j + u_{1j}. \qquad u_{1j} \sim N(0, \tau_{11}) \qquad (56)$$

Level 1:

$$y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij} \qquad r_{ij} \sim i.i.d.N(0, \sigma^2) \qquad i = 1,....n \qquad (57)$$

where $X_{ij} = 1/2$ if subject in the treatment

$X_{ij} = -1/2$ if subject in the control.

	Term	Subscript	df	Parameter	Numerator or denominator of non-centrality parameter
$l:2$	$\theta_{01} w_j$	l	2-1	$\frac{\theta_{01}^2}{2}$	$2\frac{J}{2}\frac{n}{2}\frac{\theta_{01}^2}{2}$
$j:\frac{J}{2}$	u_{0j}	$j(l)$	$2(\frac{J}{2}-1)$	τ_{00}	$\sigma^2+2\frac{n}{2}\tau_{00}$
$k:2$	$\theta_{10} X_{ij}$	k	$2-1$	$\frac{\theta_{10}^2}{2}$	$2\frac{n}{2}\frac{J}{2}\frac{\theta_{10}^2}{2}$
	$\theta_{11} w_j X_{ij}$	kl	$(2-1)(2-1)$	$\frac{\theta_{11}^2}{4}$	$\frac{n}{2}\frac{J}{2}\frac{\theta_{11}^2}{4}$
	$u_{1j} X_{ij}$	$jk(l)$	$2(\frac{J}{2}-1)$	$\frac{\tau_{11}}{2}$	$\sigma^2+\frac{n}{2}\frac{\tau_{11}}{2}$
$i:\frac{n}{2}$	r_{ij}	$i(jkl)$	$4\frac{J}{2}(n/2-1)$	σ^2	σ^2

Table 10: HLM table of MST with a covariate at the site level

The power function for the test of θ_{01} is

$$1 - \text{probt}(\ \text{tinv}(\ 0.95,\ J-2\ ),\ J-2,\ \sqrt{\frac{Jn\theta_{01}^2/4}{\sigma^2+n\tau_{00}}}\ ). \qquad (58)$$

The power function for the test of τ_{00} is

$$1 - \text{probf}(\ \text{finv}(\ 0.95,\ J-2),\ 2J(n/2-1)\ )^* \frac{\sigma^2}{\sigma^2+n\tau_{00}}, \qquad (59)$$

$2(J-1),\ 2J(n/2-1)).$

The power function for the test of θ_{10} is

$$1 - \text{probt}(\ \text{tinv}(\ 0.95,\ J-2),\ J-2,\ \sqrt{\frac{nJ\theta_{10}^2/4}{\sigma^2+\frac{n}{2}\frac{\tau_{11}}{2}}}). \qquad (60)$$

The power function for the test of θ_{11} is

$$1 - \text{probt}(\ \text{tinv}(\ 0.95,\ J-2),\ J-2,\ \sqrt{\frac{\frac{n}{2}\frac{J}{2}\frac{\theta_{11}^2}{4}}{\sigma^2+\frac{n}{2}\frac{\tau_{11}}{2}}}). \qquad (61)$$

The power function for the test of τ_{11} is

$$1 - \text{probf}(\ \text{finv}(\ 0.95,\ J-2,\ 2J(n/2-1))^* \frac{\sigma^2}{\sigma^2+\frac{n}{2}\frac{\tau_{11}}{2}},$$

$$J-2,\ 4J(n/2-1)). \qquad (62)$$

A combination of cluster randomized trial (CRT) and multi-site trial (MST)

This design has the features of both CRT and MST. At each site there is a CRT; and the same CRT is replicated across a number of sites. For example, a school-based intervention CRT can be conducted across a number of different school districts. It then becomes a 3-level HLM; and the model is listed as follows:

Level 3:

$$\beta_{00k} = \gamma_{000} + u_{00k} \quad u_{00k} \sim N(0, \tau_{\beta_{00k}}) \qquad k = 1,\ldots,K \qquad (63)$$

$$\beta_{01k} = \gamma_{010} + u_{01k} \quad u_{01k} \sim N(0, \tau_{\beta_{01k}}) \qquad (64)$$

Level 2:

$$\pi_{0jk} = \beta_{00k} + \beta_{01k} X + r_{0jk} \quad r_{0jk} \sim N(0, \tau_\pi) \qquad j = 1,\ldots,J \tag{65}$$

Level 1:

$$Y_{ijk} = \pi_{0jk} + e_{ijk} \qquad e_{ijk} \sim N(0, \sigma^2) \qquad i = 1,\ldots.n \tag{66}$$

	Term	Subscript	df	Parameter	Numerator or denominator of non-centrality parameter
$k:K$	u_{00k}	k	$K-1$	$\tau_{\beta_{00k}}$	$\sigma^2 + n\tau_\pi + 2\frac{J}{2} n\tau_{\beta_{00k}}$
$l:2$	$\gamma_{010} X$	l	$2-1$	$\frac{\gamma_{010}^2}{2}$	$\frac{J}{2} Kn \frac{\gamma_{010}^2}{2}$
	$u_{01k} X$	kl	$(K-1)(2-1)$	$\tau_{\beta_{01k}}$	$\sigma^2 + n\tau_\pi + \frac{J}{2} n\tau_{\beta_{01k}}$
$j:\frac{J}{2}$	r_{0jk}	$j(kl)$	$2K(\frac{J}{2}-1)$	τ_π	$\sigma^2 + n\tau_\pi$
$i:n$	e_{ijk}	$i(jkl)$	$2\frac{J}{2}K(n-1)$	σ^2	σ^2

Table 11: HLM table of a combination of CRT and MST

The power function for the test of $\tau_{\beta_{00k}}$ is

$$1 - \text{probf}(\text{ finv}(0.95,\ K-1,\ 2K(\frac{J}{2}-1)) * \frac{\sigma^2 + n\tau_\pi}{\sigma^2 + n\tau_\pi + nJ\tau_{\beta_{00k}}},$$

$$K-1,\ K(\frac{J}{2}-1)\ ). \tag{67}$$

The power function for the test of γ_{010} is

$$1-\text{probt}(\ \text{tinv}(0.95,\ (K-1)(2-1)\ ),\ (K-1)(2-1),\ \sqrt{\frac{\frac{J}{2}Kn\frac{\gamma_{010}^2}{2}}{\sigma^2+n\tau_\pi+\frac{J}{2}n\tau_{\beta_{01k}}}}\ ). \tag{68}$$

The power function for the test of $\tau_{\beta_{01k}}$ is

$$1-\text{probf}(\ \text{finv}(0.95,\ (K-1)(2-1),\ 2K(\frac{J}{2}-1))^*\frac{\sigma^2+n\tau_\pi}{\sigma^2+n\tau_\pi+\frac{J}{2}n\tau_{\beta_{01k}}},$$

$$(K-1)(2-1),\ 2K(\frac{J}{2}-1)\ ). \tag{69}$$

The power function for the test of τ_π is

$$1-\text{probf}(\ \text{finv}(0.95,\ 2K(\frac{J}{2}-1),\ JK(n-1)\ )^*\ \frac{\sigma^2}{\sigma^2+n\tau_\pi},$$

$$2K(\frac{J}{2}-1),\ JK(n-1)\ ). \tag{70}$$

Multi-site trial with a continuous covariate and a site characteristic

The studies of school effectiveness relate different types of school policy to students' achievement. Data are often collected on students from a number of schools, which can be classified by their policy types. The students are nested in individual schools, and schools are nested in different policy types. HLM is often used to analyze those nested data. The school policy types are considered as a school-level categorical covariate; the students' background information is

modeled as student-level variables. They can be either categorical or continuous variables. For example, students' gender is a categorical variable, and their scores on achievement tests are continuous variables. A generic model may be constructed as follows:

School-level:

$$\beta_{0j} = \theta_{00} + \theta_{01} W_j + u_{0j} \qquad u_{0j} \sim N(0, \tau_{00}) \qquad j = 1,...,J \qquad (71)$$

$$\beta_{1j} = \theta_{10} + \theta_{11} W_j + u_{1j}. \qquad u_{1j} \sim N(0, \tau_{11}) \qquad (72)$$

$$\beta_{2j} = \theta_{20} + \theta_{21} W_j + u_{2j}. \qquad u_{2j} \sim N(0, \tau_{22}) \qquad (73)$$

where W_j is a school level categorical variable; and it is assumed to be dichotomous for simplicity.

Student-level:

$$y_{ij} = \beta_{0j} + \beta_{1j} X_{1ij} + \beta_{2j} X_{2ij} + r_{ij} \qquad r_{ij} \sim i.i.d. N(0, \sigma_C^2) \qquad i = 1,....n \qquad (74)$$

where X_1 is a categorical variable, and it takes ½ and -½ for a student-level dichotomous characteristic; X_2 is a continuous variable, i.e. a continuous covariate; σ_C^2 is the student-level error variance with the inclusion of a continuous covariate.

Kreft (1993) used the same type of HLM model to study the effect of school selective recruitment on students' success. The sample contains 70 secondary schools in Amsterdam. Some schools selectively admit students based on their scores on achievement tests, and the other schools admit all students regardless of their scores. So the selective policy of schools is the school level covariate,

and it is represented by W_j in the model. The student level variables contain gender, test score on an achievement test, and their interaction. The gender corresponds to X_1 in the model and the test score to X_2. The interaction can be deemed as an additional continuous covariate like X_2 (we limit the number of continuous covariates to one in the model for simplicity, though the results generalize).

If we plan a similar study, we can use the same model. Assuming the model is balanced, we may derive the power function for the test of school types, i.e. the test of parameter θ_{01}. The power function is based on equation (58) except that σ^2 in (58) is replaced by σ_C^2, that is,

$$1 - \text{probt}(\text{ tinv}(\ 0.95,\ J-2\ ),\ J-2\ ,\ \sqrt{\frac{Jn\theta_{01}^2/4}{\sigma_C^2 + n\tau_{00}}}\ ). \qquad (75).$$

This is because we reduce the above-mentioned model to a multi-site trial with a site characteristic. If we move the continuous covarite X_2 to the left side of the equation (73), (73) changes into (76) and (76) is equivalent to equation (57), which is the level 1 model in the multi-site trial with a site characteristic.

$$Y_{ij}^{*} = Y_{ij} - \beta_{2j}X_{2ij} = \beta_{0j} + \beta_{1j}X_{1ij} + r_{ij} \qquad (76)$$

After changing the Y_{ij} into the adjusted Y_{ij}^{*}, we can apply the power functions in the multi-site trial with a site characteristic to the above-mentioned generic model.

Chapter 5

NUMERICAL APPLICATOIN OF POWER FUNCTIONS IN PLANNING EDUCATIONAL STUDIES

The power function evaluates the probability of rejecting the null hypothesis in our study. Since most of the studies are used to reject the null hypothesis, statistical power becomes a natural criterion to evaluate the soundness of a research plan. In the following we examine statistical power in two designs using HLM, i.e. cluster randomized trial and multi-site trial. In each design we pose a research question. Appropriate power functions are then chosen to determine the sample sizes. At the end the two designs are compared in terms of power performance.

Cluster randomized trial

The cluster randomized trial is used widely in educational research. For example, schools are randomly assigned to the treatment or control condition. Students in the same schools tend to share common characteristics; and their responses to the treatment may not be independent of each other. The nesting nature of the design requires HLM analysis (see Raudenbush, 1997).

The model may be formulated as follows:

Level 1:

$$Y_{ij} = \beta_{0j} + r_{ij}; \quad r_{ij} \sim N(0,\sigma^2) \quad I\ (1, 2, \ldots, n)\ j\ (1, 2, \ldots J) \qquad (77)$$

where

Y_{ij} is the individual score;

β_{0j} is the mean of the jth cluster;

r_{ij} is the individual error

n is the number of subjects in each cluster

J is the total number of clusters.

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j}, \quad u_{0j} \sim N(0,\tau_{00}), \qquad (78)$$

where

γ_{00} is the grand mean;

γ_{01} is the treatment effect;

W_j takes ½ for the treatment condition and -½ for the control condition

u_{0j} is the cluster effect.

The combined model is therefore:

$$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + u_{0j} + r_{ij}. \qquad (79)$$

The derived HLM table is as follows:

	Term	Subscript	Df	Parameter	Numerator/denominator of non-centrality parameter
$k:2$	$\gamma_{01}W_j$	k	1	$\frac{\gamma_{01}^2}{2}$	$\frac{Jn\gamma_{01}^2}{4}$
$j:J/2$	u_{0j}	$j(k)$	$J-2$	τ_{00}	$\sigma^2+n\tau$
$i:n$	r_{ij}	$i(jk)$	$J(n-1)$	σ^2	σ^2

The power functions of the test of the main treatment effect is

$$\text{1-probt(tinv(0.95, } J-2\text{), } J-2\text{, } \sqrt{\frac{Jn\gamma_{01}^2}{4(\sigma^2+n\tau)}}\text{).} \qquad (80)$$

The power function for the test of the cluster effect is

$$\text{1-probf(finv(0.95, } J-2\text{, } J(n-1)\text{)*}\frac{\sigma^2}{\sigma^2+n\tau}\text{, } J-2\text{, } J(n-1)\text{).} \qquad (81)$$

The variance components and effect size γ_{01} are real value parameters and are influenced by their measurement scale. In planning a specific study we rarely know those parameters. However, functions of those parameters are available from previous studies of similar nature. In the cluster randomized trial $\rho=\frac{\tau}{\sigma^2+\tau}$ is reported as an intraclass correlation coefficient in most of the previous studies using the same design. It varies from 0 to 1.0. $\delta=\frac{\gamma_{01}}{\sigma^2+\tau}$ is the standardized effect size whose magnitude may easily be evaluated. δ may be assumed to take 0.2, 0.5, 0.8 for small, median, and large effect (Cohen, 1988). It is

therefore natural to translate the variance components and effect size into their functional forms, whose values we can get from previous studies. After reparameterization the power function for the test of the main effect becomes

$$\text{1-probt(tinv(0.95, } J-2\text{), } J-2\text{, } \frac{\delta}{\sqrt{\frac{4}{J}(\frac{1-\rho}{n}+\rho)}}\text{).} \qquad (82)$$

The power function for the test of the cluster effect becomes

$$\text{1-probf(finv(0.95, } J-2\text{, } J(n-1)\text{)*}\frac{1-\rho}{1-\rho+n\rho}\text{, } J-2\text{, } J(n-1)\text{). (83)}$$

We may substitute the hypothesized parameter values into the power function and plot the power against a sample size variable, e.g. *J*, the number of clusters or *n* the number of subjects in each cluster. An appropriate sample size may be found from the power curve to obtain a desired power level. Depending on our research question we may use different power functions in planning the study. In the following we present a typical research problem.

An educational researcher wants to design a school-based intervention study. The researcher is interested in comparing the differential effects of two counseling programs on students morale and academic aspiration. The outcome of students morale and academic aspiration will be a composite score on a continuous scale. It is logistically feasible to administer the same counseling program in a school at one time. So the evaluator decides to use the cluster randomized trial. The schools as clusters are randomly assigned to using either one counseling program or the other. The evaluator has 10 participating schools

and wants to know how many students should be recruited in each school. Since the effect of the counseling programs corresponds to the treatment effect in the model, the power function for the test of the main effect of treatment should be used to choose the sample size. Assume that the researcher gets an intraclass correlation coefficient from previous school-based studies, e.g. ρ=0.05, and a standardized effect size 0.5 from a preliminary study. The power function can be plotted over the sample size *n* in the figure 1 (see table 12 in appendix D for numerical values). If the cluster size *n* is set to be 20, then the power will be 0.75. The choice of sample size of 20 at each school is therefore justified.

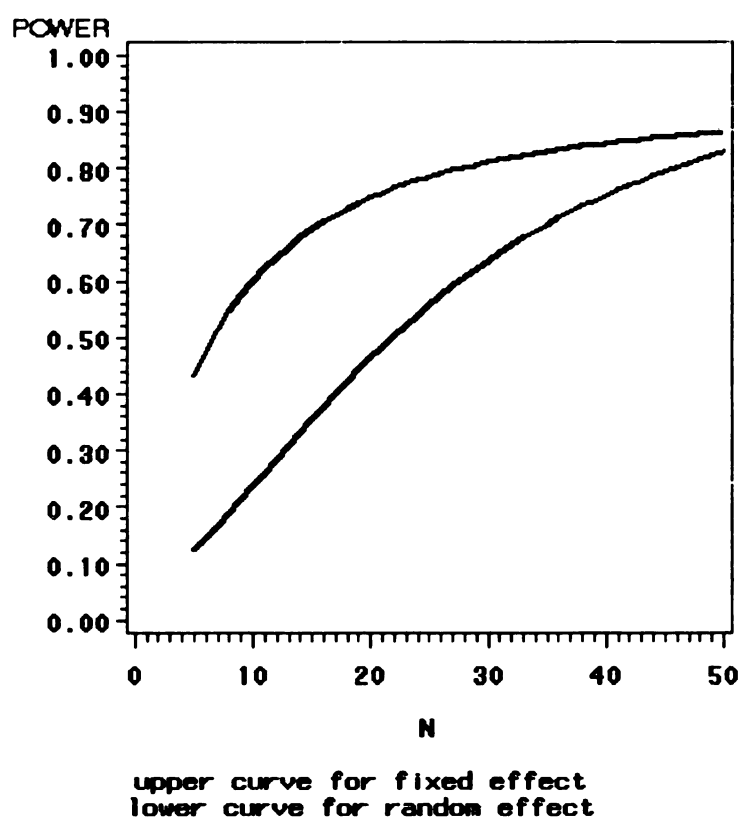


Figure 1: power in the cluster randomized trial

Observing the power function we may notice how the n and J influence the power given the parameters ρ and δ. If ρ is small like 0.05 in the previous case, then most of the variation among students scores occurs within schools. If it is more costly to sample clusters than to sample people within clusters, then increasing n is more efficient to raise the power than increasing J. Increasing n greatly reduces the denominator of the non-centrality parameter in the power function and thus increases the power. It is exactly reflected in the figure 1. On the contrary, if ρ is large, then increasing J is more efficient to get high power than increasing n (see Raudenbush, 1997).

It is noted that high power of one test is achieved at the cost of low power of the other tests. In the cluster randomized trial obtaining the desirable power of the test of the treatment effect does not necessarily guarantee high power for the test of the cluster effect. In the figure 1 the lower curve represents the power for the test of the cluster effect. It is obvious that its power is much lower than the power for the test of the treatment. Such conflict may be easily resolved if the researcher compares the importance of individual tests with reference to research questions they answer and sets them in priority order. The power function of the test, which answers the key research question, is used to choose the sample size. In the current example the main effect of treatment is of keen interest. The test of the treatment effect overweighs the test of the cluster effect; and choice of sample sizes should be made with the power of the test of the main effect of treatment.

Multi-site trial

The multi-site trial is a popular design because it is easy to administer. At each site there is an independent randomized experiment; and the same experiment is replicated across a number of sites (see Raudenbush and Liu, 1999). The model may be formulated as a 2-level HLM:

Level 1:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij}, \quad r_{ij} \sim N(0,\sigma^2) \; I \; (1, 2, \ldots, n) j \; (1, 2, \ldots J) \qquad (84)$$

where

Y_{ij} is the individual score;

β_{0j} is the mean at the j-th site;

β_{1j} is the treatment effect at the j-th site;

r_{ij} is the within-cell error.

Level-2 :

$$\beta_{0j} = \gamma_{00} + u_{0j}, \quad u_{0j} \sim N(0,\tau_{00}); \qquad (85)$$

$$\beta_{1j} = \gamma_{10} + u_{1j}, \quad u_{1j} \sim N(0,\tau_{11}), \qquad (86)$$

where

γ_{00} is the grand mean;

γ_{10} is the main effect of treatment;

$Cov(u_{0j}, u_{1j}) = \tau_{01}$ is the covariance between the site mean and treatment effect.

The combined model is

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + u_{0j} + \gamma_{10} X_{ij} u_{1j} + r_{ij} . \qquad (87)$$

If we express the combined model in the ANOVA notation, it becomes equation (6). The terms in both models are arranged in the same order.

The HLM table for the multi-site trial is as follows:

	Term	Subscript	df	Parameter	Numerator or denominator of non-centrality parameter
$k:2$	$\gamma_{10} X_{ij}$	k	1	$\frac{\gamma_{10}^2}{2}$	$J\frac{n}{2}\frac{\gamma_{10}^2}{2}$
	$u_{1j} X_{ij}$	jk	$J-1$	$\frac{\tau_{11}}{2}$	$\sigma^2 + \frac{n}{2}\frac{\tau_{11}}{2}$
$j:J$	u_{0j}	j	$J-1$	τ_{00}	$\sigma^2 + \frac{n}{2}\tau_{00}$
$i:n/2$	r_{ij}	$i(jk)$	$2J(n/2-1)$	σ^2	σ^2

The power function for the test of main effect of treatment is

$$\text{1-probt(tinv(0.95, } J-1\text{), } J-1\text{, } \sqrt{\frac{Jn\gamma_{10}^2}{4\sigma^2 + n\tau_{11}}}\text{).} \qquad (88)$$

The power function for the test of treatment-by-site interaction is

$$\text{1-probf(finv(0.95, } J-1\text{, } J(n-2)\text{)*}\frac{\sigma^2}{\sigma^2 + \frac{n\tau_{11}}{4}}\text{, } J-1\text{, } J(n-2)\text{).} \qquad (89)$$

Observing the HLM model, we may notice that γ_{10} is the unstandardized treatment effect, and that τ_{11} is the variance of the unstandardized treatment effects across individual sites. As in the cluster randomized trial we translate those parameters into their functional forms whose values can be conjectured. γ_{10} is transformed into $\delta = \frac{\gamma_{10}}{\sigma}$, and it becomes a standardized effect size.

Similarly, τ_{11} becomes $\sigma_\delta^2 = \frac{\tau_{11}}{\sigma^2}$, i.e. the variance of standardized treatment effects across sites; and its value may be set at 0.05, 0.10, 0.15 for small, median, and large (Raudenbush & Liu, 1999). The power functions with the new parameterization are as follows:

$$\text{1-probt(tinv(0.95, } J-1\text{), } J-1\text{, } \frac{\delta}{\sqrt{\frac{4}{Jn}+\frac{\sigma_\delta^2}{J}}}\text{);} \qquad (90)$$

and

$$\text{1-probf(finv(0.95, } J-1\text{, } J(n-2)\text{)*} \frac{1}{1+\frac{n\sigma_\delta^2}{4}}\text{, } J-1\text{, } J(n-2)\text{).} \qquad (91)$$

We may use either of power functions to choose sample size. The choice depends on the research question in the study. If the researcher tries to find out whether one innovative instruction program is better than the routine program, then the main effect of instruction is of great importance. The power function for the test of treatment effect should be used to make sample size choice. If on the contrary, the researcher is concerned about whether the differential treatment effect is related to the administration of those treatments at individual sites, the

power function for the treatment-by-site effect should be used to select a sample size.

Suppose that a researcher is interested in the differential effect of two tutoring methods, and that he or she conjectures a median effect size 0.5, median effect size variability across sites 0.10, and that there are 10 participating schools. He or she wants to know how many students at each school should be recruited to maintain the power of the test of the treatment at 0.75. The power can be plotted over a range of possible sample sizes n (see figure 2 and table 13 in appendix D). The power of the test for the treatment-by-site interaction (random effect) is also plotted over the same range of sample size n. The power arises very quickly with the increase of n; and it reaches 0.76 when n is 14. So the sample size 14 gives the researcher good chance to discover any median treatment effect. It is easy to see that the power for the treatment is much higher than the power for the interaction. This does not affect the adequacy of the research design. Although the interaction is included in the model, it is not considered to be significant. Its inclusion allows us to trace the source of the variances and get a good estimate of each variance components.

Observing the power function (91), we can also see the effect of J on power. For a constant n, increasing J will raise the power because it increases the non-centrality parameter in the power function. This is especially true when the effect size variability is large.

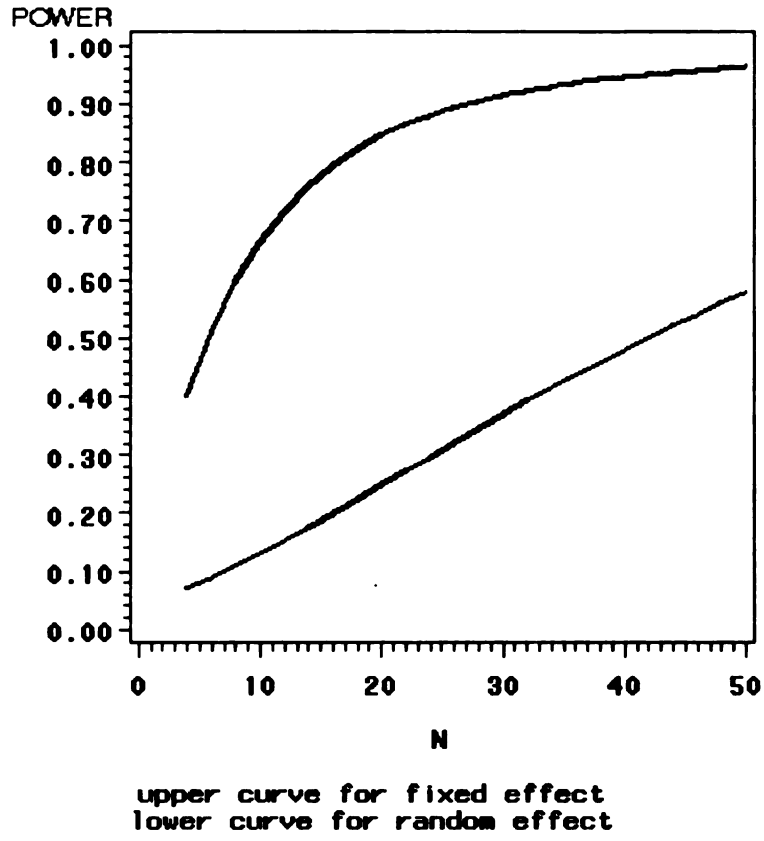


Figure 2: power in the multi-site trial

In the multi-site trial the treatment conditions are crossed with the sites; and the design allows the estimation of treatment-by-site interaction in addition to the estimation of the site random effect. In the cluster randomized trial, the cluster-by-treatment interaction is not estimable and is swept under the cluster random effect. This increases our uncertainty about the source of the variation in subjects' responses to the treatments, and it in turn enlarges the variance of our estimate of the treatment effect. As the variance of the estimate of treatment effect increases, it is less likely to reject the null and have high statistical power. Comparing the two designs in terms of power, we can see that the multi-site trial

is superior to the cluster randomized trial. The same sample size *n* returns higher power in the multi-site trial than in the cluster randomized trial. For example, when *n* is set at 14, power is 0.76 for the multi-site trial and 0.68 for the cluster randomized trial (see table 12 and 13 in appendix D). When *n* is chosen to be 20, power is 0.86 for the multi-site trial and 0.75 for the cluster randomized trial. In addition, the site and cluster variability are unfavorable for the power of the test of main effect of treatment. In the examples multi-site trial accommodates a higher site variability than the cluster randomized trial. In the multi-site trial the site variability is set at a moderate level, i.e. the effect size variability of 0.10, whereas in the cluster randomized trial the variability of clusters over the total variance is 0.05, which is considered low. In short, the multi-site trial outperforms the cluster randomized trial in terms of power even under unfavorable conditions. However, the choice of design may depend on other logistical issues. If the schools can not give differential treatments to the students at one time, cluster randomized trial may become a favorable design. It does not require that the subjects receive different treatments at one place.

In sum, the power functions may be used to assess the statistical adequacy of sample size in a certain design. They can also be used to compare different designs in terms of the power performance. The key is to come up with a reasonable set of parameters, which are meaningful to researchers. Once the parameter values are chosen, the power function can be plotted over a certain

possible range of sample size. It is then easy to determine a desirable power level and sample size for the design.

Conclusion

Sample size issue plays an important role in educational and social research. Prediction studies need to use a large enough sample to make sound generalizations. The larger the sample; the more stable the estimates become. Sample size is related to the extent to which the model can make an accurate prediction in the general case (Brooks, et al, 1996). Other studies, which test a research hypothesis, also involve choice of sample size. The test of the parameter needs a large enough sample to make the final inference defensible. The larger sample the study uses; the more information it can generate, and the more confidently the conclusion can be made about the detected treatment effect. Such confidence in the conclusion is related to the probability with which we can reject the null hypothesis and confirm our belief in the alternative hypothesis. The probability of rejecting the null hypothesis is the statistical power of the test, and the power is related to sample size of the study. The larger the sample is; the higher the statistical power can be achieved.

Sample size determination depends on the power function of the relevant test. When the model becomes complicated, it is hard to derive the power functions. It is especially true in multi-level modeling. The model is complex because the

coefficients at the lower level are considered as random at the higher level. The estimation of parameters follows very sophisticated algorithms, i.e. iterative generalized least squares or the expectation maximization (EM) procedure. It is really hard to estimate the power of all the tests in all cases. However, we can simplify the derivation of the power functions by placing some reasonable constraints on our model. We may impose a balanced design requirement on power analysis. Given the fact that studies are usually planned as taking a balanced design, it is quite practical to apply the constraint of balanced design in the power analysis. Once we limit our investigation to balanced designs. We literally eliminate the difference in many estimation methods of the parameters. They converge on the same estimate when the design is balanced. This gives us an unique solution to power analysis of multi-level models.

However, the power will vary under the unbalanced design. The unbalanced design may either be a result of missing value or unbalanced sampling plan. In the first case the power should be lowered because of information loss in the data. It may spuriously be higher or lower than it should be. This may be true if the data are not missing at random and the imputation methods are not properly used. We will discuss below the logic of power attrition when data are missing. Assume that we use the multiple imputation method. Distributions are first hypothesized for missing values; and then multiple values are generated for each missing value from those distributions to yield multiple complete data sets. The routine analysis is then performed on those multiple data sets to produce multiple

estimates of the same parameter; and the multiple estimates are averaged to give the final estimate of the parameter. The variance of the final estimate consists of two components: the first the average of the imputed estimates' variances; the second the sample variance of those estimates. When the data are complete, only the first component exists. Therefore, the variance of the final estimate from multiple imputation method is larger than it should be if no data are missing (Schafer and Olsen, 1998; Rubin, 1987). The larger the variance of the estimate; the less likely the test will reject the null hypothesis. The power of the test therefore decreases.

The unbalanced design may also arise from a sampling plan. Some sampling units may naturally have more subjects than other units. In general the power will be lower than in an unbalanced design given the total sample size. It is difficult to assess the power change without real cases. There are many procedures to adjust those unbalanced design in the data analysis. Those procedures may vary in their power performance. In addition, the distribution of the test statistics often depends on specifically used procedures. If the departure from balance design is not severe, we may treat it as a balanced design and calculate power by substituting average sample sizes or their harmonic means into the power functions.

Under the balanced design multi-level modeling can be carried out in two approaches: mixed ANOVA's and hierarchical linear models. They are

essentially the same in the planning stage of a study. The dissertation points out very clearly the connection between the two approaches. They can literally be translated from one to the other. In the former approach it is easier to do the power analysis of tests of fixed effects of more than one levels. The second approach (HLM) gives flexibility and advantages in the stage of data analysis because it accommodates missing values and the unbalanced designs of real data. The dissertation shows the power analysis for both approaches. With the HLM approach the dissertation invents a handy HLM table to derive the power functions of parameters in the model.

In HLM the power analysis literally uses the estimates of parameters at the lower level as the outcome for the parameters at the high level. At each level the model is simplified to a linear regression. It takes either the form of a one sample t test or a two sample independent t test. The power functions of the relevant parameters are derived similarly to the case of one sample t test or two sample t test. The expectation of the estimated variance of random component has patterns from the lower level to the higher level. Through algebraic transformations we may use the estimates of variance to test each random component at each level. The estimates of the treatment and variance components are algebraically related to the mean squares of their couterparts in ANOVA. The ANOVA tables provides a scaffold for systematically developing power functions for the key parameters in HLM.

With slight modification the power analysis can be extended to the case of having continuous covariates at each level in HLM. We use the CRT as an example to illustrate the approach and generalize it to any level. We may assume some covariates at the 2nd level and modify equation 27 as follows:

$$\beta_{0j} = \gamma_0 + \gamma_1 w_j + \gamma_2 X_{1j} + \gamma_3 X_{2j} + .. + u_j, \quad u_j \sim N(0,\tau) \quad j = 1,...,J. \qquad (92)$$

To simplify the computation, we assume that the population coefficients of those covariates are known, and that the percentage of variation in β_{0j} due to the covariates are known (we may use empirical estimates from the previous study to substitute), and that the covariates do not have any collinearity with the fixed effects (Randomization or matching subjects on covariates can help to achieve that). If we leave out those covariates in the analysis, we literally force τ to be larger than it should be. We may view the variation due to the covariates are swept under the random error at that level. To assess the power change due to the inclusion of covariates we may adjust the τ parameters in the power function by a percentage score, that is

$$\eta = \frac{\tau_C}{\tau} \qquad (93)$$

where η is the ratio of the reduced τ_c due to inclusion of covariates over the orignal τ. For example, the adjusted power function for testing the treatment effect becomes

$$1 - \text{probt}(\text{ tinv}(0.95, J\text{-}2), J\text{-}2, \frac{\gamma_1}{\sqrt{\frac{4}{J}(\frac{\sigma^2}{n} + \tau\eta)}}). \qquad (94)$$

Of course, the computed power value will be approximate. It should be higher than the real one because it does not assume the estimation of covariate coefficients. The estimation of covariate coefficients consumes some information in the data, which may otherwise be used to gain more precision in estimating the treatment effect. If we consider the collinearity between covariates and fixed effect, then the variance estimate for the treatment effect will be increased correspondingly (see Raudenbush, 1997) and power decreases correspondingly. In short, the real power value falls between the unadjusted power function and the adjusted power function. To generalize the approach to any level, we may hypothesize a percentage score $\eta = \frac{\tau_C}{\tau}$ for each level. τ is the random variance at that level; and τ_C is the reduced random variance due to the inclusion of covariates. We may obtain those percentage scores from previous studies, and then we may adjust the random error parameters in the power function by their corresponding percentage scores.

When planning a study researchers can standardize the parameters in the power functions and bypass the assumption of full knowledge of the key parameters. This makes it easy to plan a study. Raudenbush (1997) and Raudenbush & Liu (1999) have proposed some standardization scheme for cluster randomized trial and multi-site clinical trial. They can be adapted to general cases. This is because the every two levels of HLM essentially assumes a CRT or MST.

The rules in the dissertation may form the basis of a computer software which computes the power of the tests of key parameters in the HLM. Hopefully the dissertation will become a stepping stone to serious investigation of power analysis of general HLM, e.g. categorical outcome and multivariate outcome.

APPENDICES

APPENDIX A

DEFINITIONS OF THE USED PROBABILITY FUNCTIONS

Noncentral T cumulative distributive function:

probt(x, degrees of freedom, non-centrality parameter)

Quantile function for central T distribution:

tinv(cumulated probability, degrees of freedom);

Central F cumulative distributive function:

probf(x, df for the numerator, df for the denominator)

Noncentral F cumulative distributive function:

probf(x, df for the numerator, df for the denominator, non-centrality parameter)

Quantile function for central F:

finv(cumulated probability, df for the numerator, df for the denominator)

Noncentral Chi cumulative distributive function:

probchi(x, df, non-centrality parameter)

Quantile function for central Chi:

cinv(cumulated probability, df)

APPENDIX B

CONVERSION BETWEEN NON-CENTRAL T' AND F'

Definition of non-central T' and F'

$T'_\nu(\delta) \sim \dfrac{U+\delta}{\sqrt{\chi^2_\nu / \nu}}$, where U is a standard normal random variable.

$$F'_{1;\nu}(\delta^2) \sim \frac{\chi^2_1(\delta^2)/1}{\chi^2_\nu/\nu} = \frac{(U+\delta)^2}{\chi^2_\nu/\nu}$$

Therefore

$$T'_\nu(\delta) = \begin{cases} \sqrt{F'_{1;\nu}(\delta^2)} & T'_\nu(\delta) \geq 0 \\ -\sqrt{F'_{1;\nu}(\delta^2)} & T'_\nu(\delta) < 0 \end{cases}.$$

Also we state the following results without proof, since the proof uses the same logic as the following derivation:

$t_{1-\frac{\alpha}{2};\nu} = \sqrt{f_{1-\alpha;1;\nu}}$ and

$t_{\frac{\alpha}{2};\nu} = -\sqrt{f_{1-\alpha;1;\nu}}$.

II. Conversion in power of two-sided test between non-central T' and F'

power= P[$T'_\nu(\delta) \geq t_{1-\frac{\alpha}{2};\nu}$] + P[$T'_\nu(\delta) \leq t_{\frac{\alpha}{2};\nu}$]

= P[$T'_\nu(\delta) \geq \sqrt{f_{1-\alpha;1;\nu}}$] + P[$T'_\nu(\delta) \leq -\sqrt{f_{1-\alpha;1;\nu}}$]

= P[$(T'_\nu(\delta))^2 \geq f_{1-\alpha;1;\nu}$]

= P[$F'_{1;\nu}(\delta^2) \geq f_{1-\alpha;1;\nu}$]

APPENDIX C

SAS PROGRAM TO COMPUTE POWER

```
/*================================================

 THIS SAS PROGRAM IS USED TO COMPUTE THE VALUES
 OF POWER FUNCTIONS IN THE DISSERTATION.
 THE FUNCTIONS SHOULD BE ENTERED AS THEY APPEAR
 IN THE DISSERTATION; AND ALL THE PARAMETERS SHOULD
 BE REAL VALUES.

================================================*/

%KEYDEF F1 'END; PGM; REC; SUB';
%LET P=;
%LET FUN=;
%WINDOW FUNCTION COLOR=CYAN ROWS=30 COLUMNS=70

 GROUP=FIRST
 #5 @4 "INPUT THE POWER FUNCTION"
 #6 @4 "ALL THE PARAMETER INPUTS SHOULD BE REAL VALUES"

 #10 @4 "ENTER POWER FUNCTION BELOW"
 #12 @4  FUN 60 ATTR=UNDERLINE REQUIRED=YES

 GROUP=SECOND
 #5 @4 FUN 60
 #7 @4 'THE ABOVE FUNCTION IS EQUAL TO ' @36 P 8 ATTR=UNDERLINE
 #12 @4 'PRESS' @10 'ENTER' A=UNDERLINE @16 'TO END'
 #13 @4 'OR PRESS FUNCTION KEY' @28 'F1' A=UNDERLINE @32 'TO
CONTINUE'
;

%DISPLAY FUNCTION.FIRST;

DATA DSN1;
POWER=&FUN;
RUN;

DATA NULL;
SET DSN1;
CALL SYMPUT('P',TRIM(LEFT(POWER)) );
```

```
RUN;

%DISPLAY FUNCTION.SECOND;




/*================================================
  THIS PROGRAM TAKES A VARIABLE NAME, ITS RANGE,
  AND A POWER FUNCTION. IT THEN PLOTS THE POWER
  FUNCTION AGAINST THE VARIABLE OVER THE PROVIDED
  RANGE
================================================*/
%KEYDEF F1 'END; PGM; REC; SUB';
%LET X=;
%LET UPBOUND=;
%LET LOWBOUND=;
%LET FUN=;

%WINDOW PWPLOT COLOR=CYAN ROWS=30 COLUMNS=70

 GROUP=FIRST
 #5 @4 "INPUT THE VARIABLE NAME" @36 X 8 ATTR=UNDERLINE
 #6 @4 "AGAINST WHICH POWER SHOULD BE PLOTTED"

 GROUP=SECOND
 #5 @4 "INPUT THE VARIABLE NAME" @36 X 8 ATTR=UNDERLINE
 #6 @4 "AGAINST WHICH POWER SHOULD BE PLOTTED"

 #8 @4 "ENTER THE UPBOUND"
   @29 UPBOUND 8 ATTR=UNDERLINE REQUIRED=YES
   @43 "FOR" @48 X PROTECT=YES

 #10 @4 "ENTER THE LOWBOUND"
   @29 LOWBOUND 8 ATTR=UNDERLINE REQUIRED=YES
   @43 "FOR" @48 X PROTECT=YES

 #13 @4 "ENTER POWER FUNCTION BELOW"
 #14 @4 FUN 60 ATTR=UNDERLINE REQUIRED=YES
;

%DISPLAY PWPLOT.FIRST;
%DISPLAY PWPLOT.SECOND;
```

```
DATA PW (KEEP=POWER &X);
 LOW=SYMGET('LOWBOUND');
 UP =SYMGET('UPBOUND');
 INC=(UP-LOW)/100;

 DO &X=LOW TO UP BY INC;
   POWER=&FUN;
   OUTPUT;
 END;
RUN;

GOPTION HORIGIN=2 VORIGIN=2 VSIZE=5 HSIZE=4;
symbol1 interpol=join width=2;
AXIS1 ORDER=(0 TO 1.0 BY 0.1);
PROC GPLOT DATA=PW;
 PLOT POWER*&X/ VAXIS=AXIS1 FRAME;
RUN;
```

APPENDIX D

SAS PROGRAM FOR FIGURE 1 AND 2 AND TABLE 12 AND 13

```
/*================================================

 THIS PROGRAM PRODUCES FIGURE 1 AND 2 TABLE 12 AND
 13 IN THE DISSERTATION.
 FIGURE 1 AND TABLE 12 ARE FOR CRT;
 FIGURE 2 AND TABLE 13 ARE FOR MST;

================================================*/

FILENAME TABLE1 'C:\liu\dissertation\table1.rtf';
FILENAME TABLE2 'C:\liu\dissertation\table2.rtf';

DATA CRT (KEEP=N POWER_F POWER_R);
FILE TABLE1;

/*========================
 PARAMETERS FOR CRT
========================*/

ALPHA=0.05;  * SIGNICANCE LEVEL;
DELTA=0.5;    * DELTA STANDS FOR STANDARDIZED EFFECT SIZE;
RHO=0.05;   * RHO IS THE INTRACLASS CORRELATION;
J= 10;      * J IS # OF CLUSTERS;

PUT @10  'n'
   @20  'fixed effect'
   @40  'random effect'
   //;


 DO N=5 TO 50;

  /*=====================================
    POWER FUNCTION IS THE SAME AS ( 82, 83) IN CHAPTER 5.
    NC IS THE 4TH PARAMETER IN THE POWER
    FUNCTION FOR FIXED EFFECT;
    OMEGA IS THE SCALE IN TH POWER
```

```
   FUNCTION FOR RANDOM EFFECT

   ========================================*/

  NC=DELTA/SQRT(4*( (1-RHO)/N + RHO )/J );
  POWER_F=1-PROBT(TINV(1-ALPHA,J-2),J-2,NC) ;

  OMEGA=(1-RHO)/( 1-RHO + N*RHO);
  POWER_R=1-PROBF(FINV(1-ALPHA,J-2,J*(N-1))*OMEGA,J-2,J*(N-1) ) ;

  FORMAT POWER_F 8.2 POWER_R 8.2;

  PUT @10 N @20 POWER_F @40 POWER_R;

  OUTPUT;

 END;

PUT //
   @10 'Table 12: power in cluster randomized trial' ;
RUN;

*PROC PRINT DATA=CRT; RUN;


DATA MST (KEEP=N POWER_F POWER_R);
FILE TABLE2;

ALPHA=0.05;
DELTA=0.5;
SIG_DELT=0.10;   * VARIABILITY OF DELTA ACROSS SITES;
J=10;

PUT @10  'n'
   @20  'fixed effect'
   @40  'random effect'
   //;

DO N=4 TO 50 BY 2;  *ASSUME A BALANCED DESIGN ;

/*============================================
  POWER FUNCTIONS ARE THE SAME AS (90,91)
  IN CHAPTER 5.
============================================*/

 NC=DELTA/SQRT(4/(N*J)+SIG_DELT/J);
```

```
 POWER_F=1-PROBT(TINV(1-ALPHA,J-1), J-1, NC);

 OMEGA=1/(1+N*SIG_DELT/4);
 POWER_R=1-PROBF(FINV(1-ALPHA, J-1, J*(N-2))*OMEGA, J-1, J*(N-2) );

 FORMAT POWER_F 8.2 POWER_R 8.2;

 PUT @10 N @20 POWER_F @40 POWER_R;
 OUTPUT;

END;

PUT //
   @10 'Table 13: power in multi-site trial' ;
RUN;

%MACRO PWPLOT(DSN);

 GOPTION HORIGIN=2 VORIGIN=2 VSIZE=5 HSIZE=4;

 SYMBOL1  INTERPOL=JOIN LINE=1 WIDTH=2 ;
 SYMBOL2  INTERPOL=JOIN LINE=2 WIDTH=1;
 FOOTNOTE1 J=C H=1 'upper curve for fixed effect';
 FOOTNOTE2 J=C H=1 'lower curve for random effect';
 AXIS1 ORDER=(0 TO 1.0 BY 0.1)
     LABEL=(FONT=SWISS 'POWER');

 PROC GPLOT DATA=&DSN;
 PLOT POWER_F*N POWER_R*N /OVERLAY VAXIS=AXIS1 FRAME;
 RUN;
%MEND PWPLOT;

%PWPLOT(CRT)
%PWPLOT(MST)
```

APPENDIX D

n	treatment effect	cluster effect
5	0.43	0.12
6	0.48	0.14
7	0.51	0.17
8	0.55	0.19
9	0.57	0.21
10	0.60	0.23
11	0.62	0.26
12	0.64	0.28
13	0.66	0.31
14	0.68	0.33
15	0.69	0.35
16	0.70	0.38
17	0.72	0.40
18	0.73	0.42
19	0.74	0.44
20	0.75	0.46
21	0.76	0.48
22	0.76	0.50
23	0.77	0.52
24	0.78	0.54
25	0.78	0.56
26	0.79	0.57
27	0.80	0.59
28	0.80	0.61
29	0.81	0.62
30	0.81	0.63
31	0.81	0.65
32	0.82	0.66
33	0.82	0.67
34	0.83	0.69
35	0.83	0.70
36	0.83	0.71
37	0.84	0.72
38	0.84	0.73
39	0.84	0.74
40	0.84	0.75
41	0.85	0.76
42	0.85	0.77
43	0.85	0.78
44	0.85	0.79
45	0.85	0.79
46	0.86	0.80
47	0.86	0.81
48	0.86	0.81
49	0.86	0.82
50	0.86	0.83

Table 12: power in cluster randomized trial

n	treatment effect	treatment*site
4	0.40	0.07
6	0.51	0.09
8	0.59	0.11
10	0.66	0.13
12	0.72	0.15
14	0.76	0.17
16	0.79	0.20
18	0.82	0.22
20	0.85	0.25
22	0.86	0.27
24	0.88	0.30
26	0.89	0.32
28	0.90	0.34
30	0.91	0.37
32	0.92	0.39
34	0.93	0.41
36	0.94	0.44
38	0.94	0.46
40	0.95	0.48
42	0.95	0.50
44	0.95	0.52
46	0.96	0.54
48	0.96	0.56
50	0.96	0.58

Table 13: power in multi-site trial

BIBLIOGRAPHY

Bibliography

Bennet, C. A. and N. L. Franklin (1954*). Statistical analysis in chemistry and the chemical industry*. Wiley, New York.

Brooks, Gordon et al. (1996). Precision power and its application to the selection of regression sample sizes. *Mid-Western Educational Researcher*, 9, 10-17.

Kreft, Ita (1993). Using multilevel analysis to assess school effectiveness: a study of Dutch secondary schools. *Sociology of Education*, 66, 104-129.

Littell, et al. (1997). *SAS System for Mixed Models*. SAS, North Carolina.

Montgomery, D. C. (1996). *Design and analysis of experiments*. Wiley, New York.

Raudenbush, S. (1993). Hierarchical linear models and experimental design. In Edwards, L (Ed.), *Applied analysis of variance in behavioral science*. New York: Marcel Dekker, Inc.

Raudenbush, Stephen & Liu, Xiaofeng (1999). Statistical power and optimal design in multi-site clinical trial, revised and resubmitted to *Psychological Methods*

Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys*. J. Wiley & Sons, New York.

Schafer, J.L. and Olsen, M.K. (1998). Multiple imputation for multivariate missing-data problems: a data analyst's perspective. *Multivariate Behavioral Research*, 33 (4), 545-571.

Scheffe, H. (1959). *The analysis of variance*. Wiley, New York.

Searle, S. R. (1971). *Linear models*. Wiley, New York.

Stapleton, James. (1995). *Linear statistical models*. New York: John Wiley.

Stroup, Walter. (1998). *An introduction to mixed model analysis*. course notes.