

Practical Assessment, Research, and Evaluation

Volume 14 *Volume 14, 2009*

Article 10

2009

Using Power Tables to Compute Statistical Power in Multilevel Experimental Designs

Spyros Konstantopoulos

Follow this and additional works at: <https://scholarworks.umass.edu/pare>

Recommended Citation

Konstantopoulos, Spyros (2009) "Using Power Tables to Compute Statistical Power in Multilevel Experimental Designs," *Practical Assessment, Research, and Evaluation*: Vol. 14 , Article 10.

DOI: <https://doi.org/10.7275/xdz3-p084>

Available at: <https://scholarworks.umass.edu/pare/vol14/iss1/10>

This Article is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Practical Assessment, Research, and Evaluation by an authorized editor of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Copyright is retained by the first or sole author, who grants right of first publication to the *Practical Assessment, Research & Evaluation*. Permission is granted to distribute this article for nonprofit, educational purposes if it is copied in its entirety and the journal is credited.

Volume 14, Number 10, May 2009

ISSN 1531-7714

Using Power Tables to Compute Statistical Power in Multilevel Experimental Designs

Spyros Konstantopoulos, *Boston College*

Power computations for one-level experimental designs that assume simple random samples are greatly facilitated by power tables such as those presented in Cohen's book about statistical power analysis. However, in education and the social sciences experimental designs have naturally nested structures and multilevel models are needed to compute the power of the test of the treatment effect correctly. Such power computations may require some programming and special routines of statistical software. Alternatively, one can use the typical power tables to compute power in nested designs. This paper provides simple formulae that define expected effect sizes and sample sizes needed to compute power in nested designs using the typical power tables. Simple examples are presented to demonstrate the usefulness of the formulae.

To compute statistical power in experimental studies that use simple random samples researchers typically use power tables for one- and two-sample t-tests such as those reported in Cohen's book (1988). In one-level experimental designs that use simple random samples statistical power is a direct function of the sample size (number of individuals) and the effect size (the magnitude of the treatment effect). Larger sample sizes and effect sizes increase statistical power. However, many populations of interest in education and the social sciences have multilevel structure. In education for example, students are nested within classrooms, and classrooms are nested within schools. This nested structure produces an intraclass correlation structure that needs to be taken into account in study design.

In experimental studies that involve nested population structures, one may assign treatment conditions either to individuals such as students or to entire groups (clusters) such as schools. Designs that assign intact groups to treatments are often called cluster or group randomized designs (Bloom, 2005; Donner & Klar, 2000; Kirk, 1995; Murray, 1998). When treatments are assigned to entire subgroups (subclusters) such as classrooms or to individuals within subgroups the designs are called block randomized designs. In nested

designs statistical power computations should take into account the clustering of the design which is typically expressed via intraclass correlations and the sample sizes at each level of the hierarchy. Statistical theory for computing power in two- and three-level balanced designs has been documented (e.g., Hedges & Hedberg, 2007; Konstantopoulos, 2008a, 2008b; Murray, 1998; Raudenbush, 1997; Raudenbush & Liu, 2000). In two- and three-level cluster or block randomized designs the power of the test of the treatment effect is affected heavily and positively by the number of clusters such as schools and to a lesser extent by the number of smaller units such as classrooms or students (Konstantopoulos, 2008a, 2008b; Raudenbush & Liu, 2000). The covariates and the effect size also affect power positively and considerably. In contrast, clustering expressed via intraclass correlations affects the power estimates inversely, that is, larger intraclass correlations result in smaller power other things being equal.

The computation of power in nested experimental designs typically requires the use of the noncentral *F*- or *t*-distribution. Some programming and the use of specific routines and functions of statistical software is typically required for such computations. Alternatively and equivalently, one can use the power tables for one- and

two-sample *t*-tests presented in Cohen (1988) to compute the power of the test of the treatment effect in two- and three-level experimental designs (Barcikowski, 1981; Hedges & Hedberg, 2007). Because power tables are easy to use, the power computations of the test of the treatment effect in nested designs are greatly simplified. To achieve this, one simply needs to select the appropriate sample size and effect size, because typically power values in power tables are provided on the basis of sample sizes and effect sizes. This paper provides ways of selecting sample sizes and effect sizes in two- and three-level cluster and block randomized designs that simplify power computations by making use of power tables. First, I discuss clustering in multilevel designs. Second, I define the effect size and sample size for a two-sample two-tailed *t*-test in two- and three-level cluster randomized designs. Then, I define the effect size and sample size for a one-sample two-tailed *t*-test in two- and three-level block randomized designs. For simplicity, I discuss balanced designs with one treatment and one control group. To illustrate the methods I use examples from education that involve students, classrooms, and schools.

Defining Clustering Via Intraclass Correlations

The clustering in multilevel designs is typically defined via intraclass correlations. In two-level designs, where for example students are nested within schools, the total variance in the outcome is decomposed into two parts: a between-level-2 units (schools) variance $\tilde{\omega}^2$ and a between level-1 units (students) within-level-2 units variance $\tilde{\sigma}_e^2$, so that the total variance is $\sigma_T^2 = \tilde{\sigma}_e^2 + \tilde{\omega}^2$. Then,

$$\tilde{\rho} = \tilde{\omega}^2 / \sigma_T^2$$

is the intraclass correlation and indicates the proportion of the variance in the outcome between level-2 units or how similar or homogeneous the level-1 units within each level-2 unit are (Cochran, 1977; Lohr, 1999; Raudenbush & Bryk, 2002). For example, suppose that the total variance in achievement is 1. If the between school variance is 0.2 then the intraclass correlation is $0.2/1 = 0.2$ and indicates that 20 percent of the variance in achievement is between schools and 80 percent of the variance is within schools between students. In education, a recent paper provided a comprehensive collection of intraclass correlations for achievement data based on national representative samples (Hedges & Hedberg, 2007). Specifically, the authors provided an

array of plausible values of intraclass correlations for achievement outcomes using recent large-scale studies that surveyed national probability samples of elementary and secondary students in America. This compilation of intraclass correlations is useful for planning two-level designs. The values of intraclass correlations ranged from 0.1 to 0.25 for typical samples and were smaller than 0.1 for more homogeneous samples (low achieving schools)

The clustering in three-level designs, where for example students are nested within classrooms, and classrooms are nested within schools, is defined via two intraclass correlations because nesting can occur at the classroom and at the school level. In this case the total variance in the outcome is decomposed into three components: the between level-1 units within level-2 units variance, σ_e^2 , the between level-2 and within level-3 units variance, τ^2 , and the between level-3 units variance, ω^2 . The total variance in the outcome is defined as $\sigma_T^2 = \sigma_e^2 + \tau^2 + \omega^2$. In this case there is a level-2 (classroom) intraclass correlation

$$\rho_2 = \tau^2 / \sigma_T^2$$

and a level-3 (school) intraclass correlation

$$\rho_3 = \omega^2 / \sigma_T^2.$$

For example, suppose again that the total variance in achievement is 1, that the between school variance is 0.2 and that the between classroom variance within schools is 0.1. Then the intraclass correlation at the classroom level is $0.1/1 = 0.1$ and the intraclass correlation at the school level is $0.2/1 = 0.2$. This indicates that 20 percent of the variance in achievement is between schools, 10 percent of the variance is between classrooms within schools, and 70 percent of the variance is within classrooms between students. A recent study by Nye, Konstantopoulos, and Hedges (2004) provided empirical estimates of intraclass correlations for achievement data that are useful for planning three-level designs. The values of the intraclass correlations ranged from 0.1 to 0.2 and on average the level-2 intraclass correlations were about two-thirds as large as the level-3 intraclass correlations.

MULTILEVEL EXPERIMENTAL DESIGNS

Multilevel designs have nested structures and power analysis of such designs must take into account the clustering expressed frequently in intraclass correlations,

the effect size, and the sample sizes at each level. Hence, the methods discussed here provide formulae that incorporate intraclass correlations, effect sizes, and sample sizes. The researchers need to have some knowledge of the clustering effects and the treatment effects in order to determine the sample sizes that will result in high levels of power.

Two-Level Cluster Randomized Designs

First consider a two-level design where, for example, students are nested within schools and schools are randomly assigned to a treatment or a control group. The nesting structure of the design is illustrated graphically in the upper panel of Figure 1. Specifically, Figure 1a shows that students are nested within schools, and in turn schools are randomly assigned to treatment 1 (treatment) or treatment 2 (control).

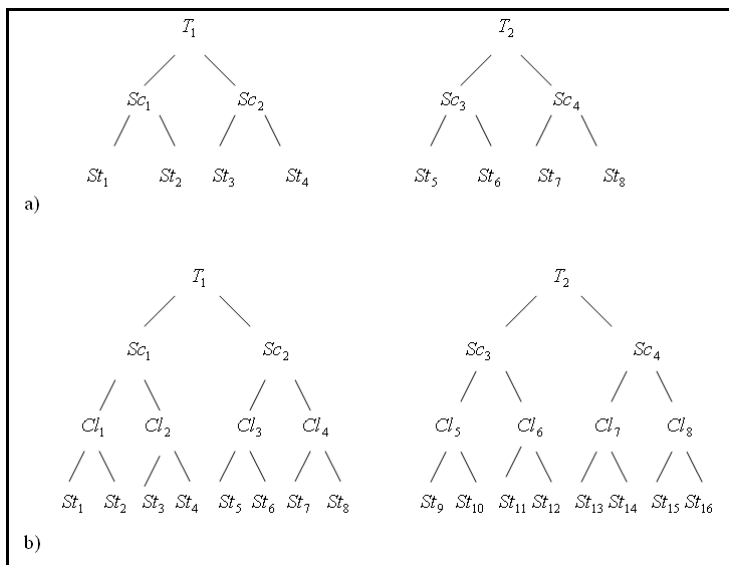


Figure 1: Nesting Structure in Two- and Three-Level Cluster Randomized Designs

Following Barcikowski (1981) and Hedges and Hedberg (2007) the expected effect size to look up in power tables for two-sample two-tailed *t*-tests at the .05 level assuming no covariates at any level is

$$\Delta = \delta \sqrt{\frac{m\tilde{n}}{2\tilde{N}}} \sqrt{\frac{1}{1+(\tilde{n}-1)\tilde{\rho}}} = \delta \sqrt{\frac{\tilde{n}}{1+(\tilde{n}-1)\tilde{\rho}}} \quad (1)$$

where $\tilde{N} = m/2$, m is the number of level-2 units (schools) in the treatment or the control group, \tilde{n} is the number of level-1 units within each level-2 unit, δ is the effect size parameter, and $\tilde{\rho}$ is the intraclass correlation at the second level. The degrees of freedom of the *t*-test

assuming simple random samples are $N_t + N_c - 2$ and $\tilde{N} = N_t N_c / (N_t + N_c)$, where N indicates sample size and the subscripts *t* and *c* represent treatment and control groups respectively (Cohen, 1988). In two-level cluster randomized designs the degrees of freedom of the *t*-test are $m_t + m_c - 2$ (Hedges & Hedberg, 2007; Raudenbush, 1997). When the sample sizes in the treatment and control groups are equal, $m_t = m_c = m$, the natural choice for N_t or N_c is m , and as a result $\tilde{N} = m/2$. In this case the expected sample size is m .

When covariates are included in the model the expected effect size to look up in power tables is

$$\Delta = \delta \sqrt{\frac{m\tilde{n}}{2\tilde{N}^*}} \sqrt{\frac{1}{\eta_1 + (\tilde{m}_2 - \eta_1)\tilde{\rho}}} = \delta \sqrt{\frac{\tilde{n}}{2(m-q)}} \sqrt{\frac{\tilde{n}}{\eta_1 + (\tilde{m}_2 - \eta_1)\tilde{\rho}}} \quad (2)$$

where $\tilde{N}^* = (m^2 - mq)/(2m - q)$, and η_1, η_2 indicate the proportions of the level-1 and level-2 residual variances to the total variances at the corresponding levels, q is the number of level-2 covariates (school characteristics), and the other terms were defined previously (Hedges & Hedberg, 2007; Murray, 1998). For example, if the covariates at the first and second level explain 30 percent of the variance at the corresponding level then $\eta_1 = \eta_2 = 0.70$. The degrees of freedom of the *t*-test are $m_t + m_c - 2 - q$, and if $N_t = m_t - q$, $N_c = m_c$, and $m_t = m_c = m$, then it follows that $\tilde{N}^* = (m^2 - mq)/(2m - q)$. In this case, the expected sample size is $m - q/2$.

In order to compute the expected effect size one needs to have some knowledge of the intraclass correlation and the treatment effect (effect size). Plausible values of intraclass correlations for achievement data in two-level designs are reported in a recent study by Hedges and Hedberg (2007). Plausible estimates of the treatment effect can be obtained from previous empirical or meta-analytic work. To illustrate the simplicity of the computations suppose that 20 schools are randomly assigned to two conditions ($m = 10$ in each condition) and that 40 students ($\tilde{n} = 40$) are sampled within each school for a total sample size of 800 students. Assume that there are no covariates at any level, that the effect size is $\delta = 0.5$ standard deviations, and that the intraclass correlation at the second level is $\tilde{\rho} = 0.2$. Then for a two-sample two-tailed *t*-test at the .05 level the expected effect size using equation 1 is

$$\Delta = 0.5 \sqrt{\frac{40}{(1 + (40 - 1)0.2)}} = 1.07.$$

$$\Delta = 0.5 \sqrt{\frac{(2 * 10 - 1)}{2(10 - 1)}} \sqrt{\frac{40}{(0.75 + (40 * 0.75 - 0.75)0.2)}} = 1.27.$$

This effect size estimate is represented in Cohen’s (1988) Table 2.3.5 by *d* (the *x* axis). The number of schools in each condition is *m* = 10 and is represented in Cohen’s Table 2.3.5 by *n* (the *y* axis). Table 1 essentially replicates Cohen’s Table 2.3.5. The expected effect size is very close to 1.1 and according to Table 1 the power is 0.64 when the effect size *d* = 1.1 and the number of schools per condition is *n* = 10. This indicates that there

The expected sample size is *m* – *q*/2 = 10 – 0.5 = 9.5. An effect size of 1.27 is very close to 1.3 and hence according to Table 1 when *d* = 1.3 and *n* = 9 the power is 0.74 and when *d* = 1.3 and *n* = 10 the power is 0.78. Since the expected sample size is 9.5 which is halfway between 9 and 10 the power would be 0.76 (halfway between 0.74 and 0.78).

Table 1 Power of two-sample two-tailed t-test at .05 level

n	d														
	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00	1.10	1.20	1.30	1.40	1.50
2	0.05	0.05	0.05	0.06	0.06	0.07	0.07	0.08	0.09	0.10	0.10	0.11	0.13	0.14	0.15
3	0.05	0.05	0.06	0.07	0.08	0.09	0.10	0.12	0.14	0.16	0.18	0.21	0.23	0.26	0.29
4	0.05	0.06	0.07	0.08	0.09	0.11	0.13	0.16	0.19	0.22	0.26	0.30	0.34	0.38	0.43
5	0.05	0.06	0.07	0.09	0.11	0.13	0.16	0.20	0.24	0.29	0.33	0.39	0.44	0.49	0.55
6	0.05	0.06	0.08	0.10	0.12	0.16	0.20	0.24	0.29	0.35	0.41	0.47	0.53	0.59	0.65
7	0.05	0.06	0.08	0.11	0.14	0.18	0.23	0.28	0.34	0.41	0.47	0.54	0.61	0.67	0.73
8	0.05	0.07	0.09	0.12	0.15	0.20	0.26	0.32	0.39	0.46	0.54	0.61	0.68	0.74	0.80
9	0.05	0.07	0.09	0.13	0.17	0.22	0.29	0.36	0.43	0.51	0.59	0.67	0.74	0.80	0.85
10	0.06	0.07	0.10	0.14	0.19	0.25	0.32	0.40	0.48	0.56	0.64	0.72	0.78	0.84	0.89
11	0.06	0.07	0.10	0.15	0.20	0.27	0.35	0.43	0.52	0.61	0.69	0.76	0.83	0.88	0.92
12	0.06	0.08	0.11	0.16	0.22	0.29	0.37	0.47	0.56	0.65	0.73	0.80	0.86	0.91	0.94
13	0.06	0.08	0.11	0.16	0.23	0.31	0.40	0.50	0.60	0.69	0.77	0.84	0.89	0.93	0.96
14	0.06	0.08	0.12	0.17	0.25	0.33	0.43	0.53	0.63	0.72	0.80	0.86	0.91	0.95	0.97
15	0.06	0.08	0.12	0.18	0.26	0.35	0.46	0.56	0.66	0.75	0.83	0.89	0.93	0.96	0.98
16	0.06	0.09	0.13	0.19	0.28	0.38	0.48	0.59	0.69	0.78	0.85	0.91	0.94	0.97	0.98
17	0.06	0.09	0.14	0.20	0.29	0.40	0.51	0.62	0.72	0.81	0.87	0.92	0.96	0.98	0.99
18	0.06	0.09	0.14	0.21	0.31	0.42	0.53	0.65	0.75	0.83	0.89	0.94	0.97	0.98	0.99
19	0.06	0.09	0.15	0.22	0.32	0.44	0.56	0.67	0.77	0.85	0.91	0.95	0.97	0.99	0.99
20	0.06	0.09	0.15	0.23	0.34	0.46	0.58	0.69	0.79	0.87	0.92	0.96	0.98	0.99	1.00
21	0.06	0.10	0.16	0.24	0.35	0.48	0.60	0.72	0.81	0.89	0.94	0.97	0.98	0.99	1.00
22	0.06	0.10	0.16	0.25	0.37	0.49	0.62	0.74	0.83	0.90	0.95	0.97	0.99	0.99	1.00
23	0.06	0.10	0.17	0.26	0.38	0.51	0.64	0.76	0.85	0.91	0.95	0.98	0.99	1.00	1.00
24	0.06	0.10	0.17	0.27	0.40	0.53	0.66	0.77	0.86	0.92	0.96	0.98	0.99	1.00	1.00
25	0.06	0.11	0.18	0.28	0.41	0.55	0.68	0.79	0.88	0.93	0.97	0.99	0.99	1.00	1.00
26	0.06	0.11	0.19	0.29	0.42	0.56	0.70	0.81	0.89	0.94	0.97	0.99	1.00	1.00	1.00
27	0.07	0.11	0.19	0.30	0.44	0.58	0.71	0.82	0.90	0.95	0.98	0.99	1.00	1.00	1.00
28	0.07	0.11	0.20	0.31	0.45	0.60	0.73	0.84	0.91	0.96	0.98	0.99	1.00	1.00	1.00
29	0.07	0.12	0.20	0.32	0.46	0.61	0.75	0.85	0.92	0.96	0.98	0.99	1.00	1.00	1.00
30	0.07	0.12	0.21	0.33	0.48	0.63	0.76	0.86	0.93	0.97	0.99	1.00	1.00	1.00	1.00

is roughly a 64 percent chance of detecting an effect size of 1.1 standard deviations when there are 10 schools per condition.

Suppose now that one covariate is included at each level of the hierarchy and each covariate explains 0.25 percent of the variance at the corresponding level. This indicates that $\eta_2 = \eta_1 = 0.75$ and *q* = 1. Suppose that all other values remain unchanged. Now, using equation 2 the expected effect size is

Three-Level Cluster Randomized Designs

Now consider a three-level design where for example students are nested within classrooms, classrooms are nested within schools, and schools are randomly assigned to a treatment or a control group. The nesting structure of the design is illustrated graphically in the lower panel of Figure 1. Specifically, Figure 1b shows that students are nested within classrooms, classrooms are nested within schools, and in turn schools are randomly assigned to treatment 1 (treatment) or

treatment 2 (control). In this case and assuming no covariates, the expected effect size to look up in power tables is

$$\Delta = \delta \sqrt{\frac{pn}{1+(n-1)\rho_2+(pn-1)\rho_3}} \quad (3)$$

where p is the number of level-2 units within level-3 units, n is the number of level-1 units within each level-2 unit, ρ_3 is the intraclass correlation at the third level, ρ_2 is the intraclass correlation at the second level, and all other terms have been defined previously. The expected sample size is m as in the two-level case. When covariates are included in the model the expected effect size to look up in power tables is

$$\Delta = \delta \sqrt{\frac{2m-q}{2(m-q)}} \sqrt{\frac{pn}{\eta_1+(n\eta_2-\eta_1)\rho_2+(pn\eta_3-\eta_1)\rho_3}} \quad (4)$$

where η_3 is the proportion of the residual variance to the total variance at the third level, and all other terms have been already defined. The expected sample size is $m - q/2$ as in the two-level case.

To illustrate the computations suppose that 20 schools are randomly assigned to two conditions ($m = 10$), that 2 classrooms are sampled per school ($p = 2$), and 20 students are sampled within each classroom ($n = 20$) for a total sample size of 800 students. Assume that there are no covariates, that the effect size is $\delta = 0.5$ standard deviations, and that the intraclass correlations at the second level and third level are $\rho_2 = 0.1$ and $\rho_3 = 0.2$ respectively. The expected effect size using equation 3 is

$$\Delta = 0.5 \sqrt{\frac{2*20}{1+(20-1)0.1+(2*20-1)0.2}} = 0.97$$

and the expected sample size is 10. The value of the effect size is very close to 1 and hence according to Table 1 when $d = 1$ and $n = 10$ the power is 0.56.

Suppose now that one covariate is included at each level of the hierarchy and each covariate explains 0.25 percent of the variance at the corresponding level. This indicates that $\eta_3 = \eta_2 = \eta_1 = 0.75$ and $q = 1$. Suppose also that all other values remain unchanged. Using equation 4 the expected effect size is

$$\Delta = 0.5 \sqrt{\frac{(2*10-1)}{2(10-1)}} \sqrt{\frac{2*20}{(0.75+(20*0.75-0.75)0.1+(2*20*0.75-0.75)0.2)}} = 1.15$$

The expected sample size is $m - q/2 = 10 - 0.5 = 9.5$. An effect size of 1.15 is halfway between 1.1 and 1.2. According to Table 1 when $d = 1.1$ and $n = 9$ the power is 0.59 and when $d = 1.2$ and $n = 9$ the power is 0.67. Also, according to Table 1 when $d = 1.1$ and $n = 10$ the power is 0.64 and when $d = 1.2$ and $n = 10$ the power is 0.72. The power is essentially the average of all 4 power values and is roughly 0.66.

Two-Level Block Randomized Designs

First consider a two-level design where for example students are nested within schools and students are randomly assigned to a treatment or a control group within schools. The schools in this design serve as blocks. The nesting structure of the design is illustrated graphically in the upper panel of Figure 2. Specifically, Figure 2a shows that students are randomly assigned to treatment 1 (treatment) or treatment 2 (control) within each school. Again, following Barcikowski (1981) and Hedges and Hedberg (2007) the expected effect size to look up in power tables for one-sample two-tailed t -tests at the .05 level assuming no covariates at any level is

$$\Delta = \delta \sqrt{\frac{n^*/2}{1+(n^*\mathcal{G}_2-1)\tilde{\rho}}} \quad (5)$$

where n^* is the number of level-1 units within each condition within each level-2 unit, and \mathcal{G}_2 is the proportion of the between-level-2 units variance of the treatment effect to the total between level-2 units variance (Konstantopoulos, 2008b). The degrees of freedom of the one-sample t -test assuming simple random samples and no nesting are $N - 1$, where N is the total sample size. In two-level designs where level-1 units are randomly assigned to conditions within level-2 units the degrees of freedom of the t -test are $m^* - 1$ where m^* is the total number of level-2 units (schools). In this case the expected sample size is m^* . When covariates are included in the model the expected effect size to look up in the power tables is

$$\Delta = \delta \sqrt{\frac{m^*}{2(m^*-q)}} \sqrt{\frac{n^*}{\eta_1+(n^*\mathcal{G}_{R2}\eta_2-\eta_1)\tilde{\rho}}} \quad (6)$$

where \mathcal{G}_{R2} is the proportion of the between-level-2 units residual variance of the treatment effect to the total between level-2 units residual variance and all other terms have been defined already (Konstantopoulos, 2008b). The expected sample size in this case is $m^* - q$.

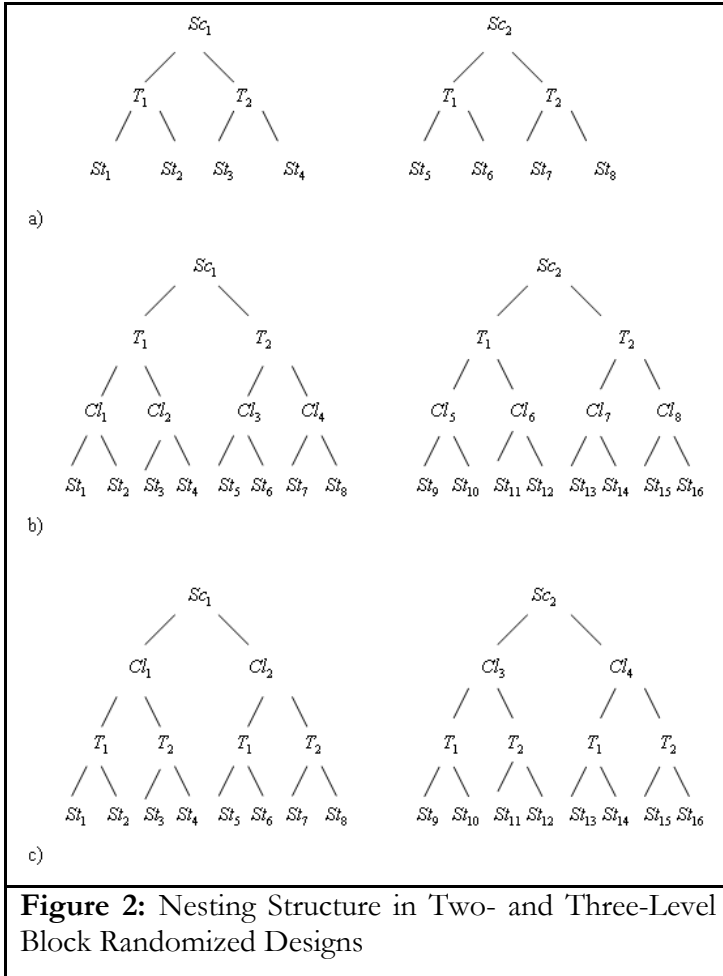


Figure 2: Nesting Structure in Two- and Three-Level Block Randomized Designs

To illustrate the computations consider a two-level design where level-1 units are randomly assigned to conditions within level-2 units. Suppose that there are 20 schools overall ($m^* = 20$) and 20 students per condition per school ($n^* = 20$) for a total sample size of 800 students. Suppose that no covariates are included at any level, that the effect size is $\delta = 0.25$ standard deviations, that $\mathcal{G}_2 = 1/9$, and that the intraclass correlation at the school level is $\tilde{\rho} = 0.20$. Using equation 5 the expected effect size is

$$\Delta = 0.25 \sqrt{\frac{20}{2(1+(20(1/9)-1)0.2)}} = 0.71,$$

and the expected sample size is 20. The expected effect size 0.71 is very close to 0.7 and according to Table 2 when $d = 0.7$ and $n = 20$ the power is 0.84. Suppose now that one covariate is included at each level of the hierarchy and that each covariate explains 0.25 percent of the variance at the corresponding level and everything

else remains the same. This indicates that $\eta_2 = \eta_1 = 0.75$ and $q = 1$. Then, using equation 6 the expected effect size is

$$\Delta = 0.25 \sqrt{\frac{20}{2(20-1)}} \sqrt{\frac{20}{(0.75 + (20 * 0.75 * (1/9) - 0.75)0.2)}} = 0.84$$

and the expected sample size is $20 - 1 = 19$. The expected effect size is roughly halfway between 0.8 and 0.9. According to Table 2 when $d = 0.8$ and $n = 19$ the power is 0.91 and when $d = 0.9$ and $n = 19$ the power is 0.96. Then, the power estimate is halfway between 0.91 and 0.96 and hence it is roughly 0.94.

Three-Level Block Randomized Designs

Now consider a three-level model where for example students are nested within classrooms and classrooms are nested within schools and classrooms within schools are randomly assigned to conditions. The nesting structure of the design is illustrated graphically in the middle panel of Figure 2. Specifically, Figure 2b shows that students are nested within classrooms and classrooms are randomly assigned to treatment 1 (treatment) or treatment 2 (control) within each school. In this case the expected effect size to look up in power tables is

$$\Delta = \delta \sqrt{\frac{p^* n / 2}{1 + (n-1)\rho_2 + (p^* n \mathcal{G}_3 - 1)\rho_3}} \quad (7)$$

where p^* is the number of classrooms randomly assigned to conditions within schools, n is the number of level-1 units within level-2 units, and \mathcal{G}_3 is the proportion of the between level-3 units variance of the treatment effect to the total variance at the third level. The expected sample size is m^* as in the two-level case. When covariates are included at each level the expected effect size to look up in power tables is

$$\Delta = \delta \sqrt{\frac{m^*}{2(m^* - q)}} \sqrt{\frac{p^* n}{\eta_e + (n\eta_2 - \eta_1)\rho_2 + (p^* n \mathcal{G}_{R3} \eta_3 - \eta_1)\rho_3}} \quad (8)$$

where m^* is the total number of level-3 units, \mathcal{G}_{R3} is the proportion of the between level-3 units residual variance of the treatment effect to the total residual variance at the third level, and all other terms have been defined already. The expected sample size is $m^* - q$ as in the two-level case.

Table 2 Power of one-sample two-tailed t-test at .05 level

n	d														
2	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00	1.10	1.20	1.30	1.40	1.50
3	0.05	0.05	0.05	0.06	0.06	0.07	0.07	0.08	0.09	0.09	0.10	0.11	0.12	0.13	0.13
4	0.05	0.06	0.06	0.07	0.08	0.10	0.12	0.14	0.16	0.18	0.20	0.23	0.26	0.29	0.32
5	0.05	0.06	0.07	0.09	0.11	0.14	0.17	0.21	0.25	0.29	0.34	0.38	0.43	0.48	0.53
6	0.05	0.06	0.08	0.11	0.14	0.18	0.23	0.28	0.34	0.40	0.47	0.53	0.59	0.65	0.71
7	0.06	0.07	0.09	0.13	0.17	0.22	0.29	0.36	0.43	0.51	0.58	0.66	0.72	0.78	0.83
8	0.06	0.07	0.10	0.15	0.20	0.27	0.35	0.43	0.52	0.60	0.68	0.75	0.82	0.87	0.91
9	0.06	0.08	0.11	0.17	0.23	0.31	0.40	0.50	0.59	0.68	0.76	0.83	0.88	0.92	0.95
10	0.06	0.08	0.13	0.19	0.26	0.35	0.46	0.56	0.66	0.75	0.82	0.88	0.93	0.96	0.98
11	0.06	0.09	0.14	0.21	0.29	0.40	0.51	0.62	0.72	0.80	0.87	0.92	0.95	0.98	0.99
12	0.06	0.09	0.15	0.23	0.32	0.44	0.55	0.67	0.77	0.85	0.91	0.95	0.97	0.99	0.99
13	0.06	0.10	0.16	0.25	0.35	0.48	0.60	0.71	0.81	0.88	0.93	0.97	0.98	0.99	1.00
14	0.06	0.10	0.17	0.26	0.38	0.51	0.64	0.75	0.85	0.91	0.95	0.98	0.99	1.00	1.00
15	0.06	0.11	0.18	0.28	0.41	0.55	0.68	0.79	0.88	0.93	0.97	0.99	0.99	1.00	1.00
16	0.07	0.11	0.19	0.30	0.44	0.58	0.71	0.82	0.90	0.95	0.98	0.99	1.00	1.00	1.00
17	0.07	0.12	0.20	0.32	0.47	0.61	0.75	0.85	0.92	0.96	0.98	0.99	1.00	1.00	1.00
18	0.07	0.12	0.21	0.34	0.49	0.64	0.77	0.87	0.94	0.97	0.99	1.00	1.00	1.00	1.00
19	0.07	0.13	0.23	0.36	0.52	0.67	0.80	0.89	0.95	0.98	0.99	1.00	1.00	1.00	1.00
20	0.07	0.13	0.24	0.38	0.54	0.70	0.82	0.91	0.96	0.98	1.00	1.00	1.00	1.00	1.00
21	0.07	0.14	0.25	0.40	0.57	0.72	0.84	0.92	0.97	0.99	1.00	1.00	1.00	1.00	1.00
22	0.07	0.14	0.26	0.42	0.59	0.74	0.86	0.94	0.98	0.99	1.00	1.00	1.00	1.00	1.00
23	0.07	0.15	0.27	0.43	0.61	0.77	0.88	0.95	0.98	0.99	1.00	1.00	1.00	1.00	1.00
24	0.07	0.15	0.28	0.45	0.63	0.79	0.89	0.96	0.99	1.00	1.00	1.00	1.00	1.00	1.00
25	0.08	0.16	0.29	0.47	0.65	0.80	0.91	0.96	0.99	1.00	1.00	1.00	1.00	1.00	1.00
26	0.08	0.16	0.30	0.48	0.67	0.82	0.92	0.97	0.99	1.00	1.00	1.00	1.00	1.00	1.00
27	0.08	0.17	0.31	0.50	0.69	0.84	0.93	0.98	0.99	1.00	1.00	1.00	1.00	1.00	1.00
28	0.08	0.17	0.32	0.52	0.71	0.85	0.94	0.98	0.99	1.00	1.00	1.00	1.00	1.00	1.00
29	0.08	0.18	0.33	0.53	0.72	0.86	0.95	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00
30	0.08	0.18	0.35	0.55	0.74	0.88	0.95	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
30	0.08	0.19	0.36	0.56	0.75	0.89	0.96	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00

To illustrate the computations consider a three-level design where for example classrooms within schools are randomly assigned to conditions, there are 20 schools ($m^* = 20$), one classroom per condition per school ($p^* = 1$), and 20 students per classroom ($n = 20$) for a total sample size of 800 students. Suppose that no covariates are included at any level, that the effect size is $\delta = 0.25$ standard deviations, that $\mathcal{G}_{R3} = 1/9$, and that the intraclass correlations at the classroom and school level are $\rho_2 = 0.10$ and $\rho_3 = 0.20$ respectively. Using equation 7 the expected effect size is

$$\Delta = 0.25 \sqrt{\frac{1 * 20}{2(1 + (20 - 1)0.1 + (1 * 20 * (1/9) - 1)0.2)}} = 0.45,$$

and the expected sample size is 20. When $d = 0.45$ and $n = 20$ according to Table 2 the power is halfway between 0.40 ($d = 0.4$) and 0.57 ($d = 0.5$), that is, it is roughly 0.49. The effects of covariates in power can also be incorporated in the expected effect size using equation 8.

Finally, consider a three-level design where level-1 units are randomly assigned to conditions within level-2 units within level-3 units. The nesting structure of the design is illustrated graphically in the lower panel of Figure 2. Specifically, Figure 2c shows that students are randomly assigned to treatment 1 (treatment) and treatment 2 (control) within each classroom, and classrooms are nested within schools. Assuming no covariates at any level the expected effect size to look up in the power tables is

$$\Delta = \delta \sqrt{\frac{pn^*/2}{1 + (n^*\mathcal{G}_2 - 1)\rho_2 + (pn^*\mathcal{G}_3 - 1)\rho_3}} \quad (9)$$

where \mathcal{G}_2 , \mathcal{G}_3 are the proportions of the between level-2 and between level-3 unit variance of the treatment effect to the total variance at the second and third level respectively. The expected sample size in this case is m^* .

When covariates are included the effect size to look up in the power tables is

$$\Delta = \delta \sqrt{\frac{m^*}{2(m^* - q)}} \sqrt{\frac{pn^*}{\eta_1 + (n^* \mathcal{G}_{R2} \eta_2 - \eta_1) \rho_2 + (pn^* \mathcal{G}_{R3} \eta_3 - \eta_1) \rho_3}} \quad (10)$$

where \mathcal{G}_{R2} , \mathcal{G}_{R3} are the proportions of the between level-2 and between level-3 unit residual variance of the treatment effect to the residual variance at the second and third level respectively. The expected sample size in this case is $m^* - q$.

To illustrate the computations consider a three-level design where for example students within classrooms are randomly assigned to conditions, there are 20 schools ($m^* = 20$), two classrooms per school ($p = 2$), and 10 students per condition per classroom ($n^* = 10$) for a total sample size of 800 students. Suppose that no covariates are included at any level, that the effect size is $\delta = 0.25$ standard deviations, that $\mathcal{G}_3 = \mathcal{G}_2 = 1/9$, and that the intraclass correlations at the classroom and school level are $\rho_2 = 0.1$ and $\rho_3 = 0.2$ respectively. Using equation 9 the expected effect size is

$$\Delta = 0.25 \sqrt{\frac{2 * 10}{2(1 + (10 * (1/9) - 1)0.1 + (2 * 10 * (1/9) - 1)0.2)}} = 0.71,$$

and the expected sample size is 20. The effect size value of 0.71 is very close to 0.7 and according to Table 2 when $d = 0.7$ and $n = 20$ the power is 0.84. The effects of covariates in power can also be incorporated in the expected effect size using equation 10.

CONCLUSION

This paper showed how conventional power tables such as those reported in Cohen (1988) can be used to compute statistical power in two- and three-level cluster and block randomized designs. Once the expected effect size and the expected sample size are derived, the computation of statistical power using power tables is straightforward. Such computations provide an easier alternative to computing power in nested designs using complicated formulae of the non-central F- or t-distribution. All computations can be easily performed using a calculator or excel.

It should be noted that methods for *a priori* power computations during the design phase of an experimental study as those provided here are intended to serve simply as useful guides for study design. That is, *a priori* power estimates although informative should be

treated as approximate and not exact (Kraemer & Thieman, 1987). The reason is that the power estimates are as accurate as the estimates of effect sizes and intraclass correlations, which are typically educated guesses.

REFERENCES

- Barcikowski, R. S. (1981). Statistical power with group mean as the unit of analysis. *Journal of Educational Statistics*, 6, 267-285.
- Bloom, H. S. (2005). *Learning more from social experiments*. New York: Russell Sage.
- Cochran, W. G. (1977). *Sampling techniques*. New York: Wiley.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.
- Donner, A., & Klar, N. (2000). *Design and analysis of cluster randomization trials in health research*. London: Arnold.
- Hedges, L. V., & Hedberg, E. (2007). Intraclass correlation values for planning group randomized trials in Education. *Educational Evaluation and Policy Analysis*, 29, 60-87.
- Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole Publishing.
- Konstantopoulos, S. (2008a). The power of the test for treatment effects in three-level cluster randomized designs. *Journal of Research on Educational Effectiveness*, 1, 66-88.
- Konstantopoulos, S. (2008b). The power of the test for treatment effects in three-level block randomized designs. *Journal of Research on Educational Effectiveness*, 1, 265-288.
- Kraemer, H. C., & Thieman, S. (1987). *How many subjects? Statistical power analysis in research*. Newbury Park, CA: Sage.
- Lohr, S. L. (1999). *Sampling: Design and analysis*. Pacific Grove, CA: Duxbury Press.
- Murray, D. M. (1998). *Design and analysis of group-randomized trials*. New York: Oxford University Press.
- Nye, B, Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, 26, 237-257.
- Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trails. *Psychological Methods*, 2, 173-185.
- Raudenbush, S. W., & Liu, X. (2000). Statistical power and optimal design for multisite randomized trails. *Psychological Methods*, 5, 199-213.
- Raudenbush, S. W, & Bryk, A. S. (2002). *Hierarchical Linear Models*. Thousand Oaks, CA: Sage.

Note

The author is indebted to Chen Ann for creating the figures used in this report.

Citation

Konstantopoulos, Spyros (2009). Using Power Tables to Compute Statistical Power in Multilevel Experimental Designs. *Practical Assessment, Research & Evaluation*, 14(10). Available online: <http://pareonline.net/getvn.asp?v=14&n=10>.

Author

Spyros Konstantopoulos
Lynch School of Education
Boston College
Chestnut Hill, MA 02467

<http://www.bc.edu/schools/lsoe/facultystaff/faculty/konstantopoulos.html>

Konstans [at] bc.edu