PINT

(Power IN Two-level designs)

Estimating standard errors
of regression coefficients
in hierarchical linear models
for power calculations

USER'S MANUAL

Version 2.1

Roel J. Bosker
Tom A.B. Snijders
Henk Guldemond

PINT (Power IN Two-level designs) is a program for estimating standard errors of regression coefficients in hierarchical linear models. For more detailed information the user is referred to:

> Snijders, T.A.B. & R.J. Bosker (1993). Standard errors and sample sizes in two-level research. *Journal of Educational Statistics, 18,* 3, 237-260.

If you use PINT in a paper, proposition, or other report, you are requested to refer to this paper and/or to this manual.

The PINT program and this manual can be downloaded from web site http://stat.gamma.rug.nl/snijders/

Questions concerning PINT can be addressed to :

Tom A. B. Snijders
Department of Sociology
University of Groningen
Grote Rozenstraat 31
9712 TG Groningen
The Netherlands
Email T.A.B.Snijders@PPSW.RUG.NL
fax +31 - (0)50 - 3636304

or to

Roel J. Bosker
Department of Educational Sciences
University of Groningen
Grote Rozenstraat 38
9712 TJ Groningen
The Netherlands
Email R.J.Bosker@PPSW.RUG.NL
fax +31-(0)50-3636521

This research was supported by the Netherlands Organisation for Scientific Research (NWO), grant no. 500-278-203.

Groningen/Enschede, first version December 1996; this update April 30, 2003.

Contents

1. General Introduction

Multilevel and, in particular, two-level designs are used frequently in educational and social research. Hierarchical linear models incorporating both random and fixed effects provide a useful statistical paradigm for situations where nesting is an obvious and direct consequence of multistage sampling as well as situations with nested sources of random variability. Introductory texts are Bryk & Raudenbush (1992), Goldstein (1995), Hox (2002), and Snijders & Bosker (1999).

When a researcher is designing a multistage sampling scheme to assess, for instance, the effects of schools on the achievement of students, or, to give another example, to test the hypothesis that citizens in impoverished neighbourhoods are more often victims of crime than other citizens, important decisions must be made with respect to the sample sizes at the various levels. For the two-level design in the first example the question might be phrased like: should one investigate many schools with few students per school or few schools with many students per school? Or, for the second example: should we sample many neighbourhoods with only few citizens per neighbourhood or many citizens per neighbourhood and only few neighbourhoods? In both cases we assume, of course, that there are budgetary constraints for the research to be conducted.

To phrase this question somewhat differently: how should researchers choose sample sizes at the macro- and micro-level in order to ensure a desired level of power given a relevant (hypothesized) effect size and a chosen significance level α?

A general introduction to power analysis can be found in the standard work by Cohen (1988), or, for a quick introduction, Cohen's power primer (Cohen, 1992). The basic idea is that we would like to find support for a research hypothesis (H_1) stating that a certain effect exists, and for that reason we test a hypothesis about the absence of this effect (H_0) using a sample from the population we are interested in. The significance level α represents the risk of mistakenly rejecting H_0. This mistake is known as a Type I error. Vice versa, β is the risk of disappointingly not rejecting H_0, in the case that the effect does exist in the population. This mistake is known as a Type II error. The statistical power of a significance test is the probability of rejecting H_0, given the effect size in the population, the significance level α, and the sample size N. Power is therefore given by $1-\beta$. As a rule of thumb, Cohen suggests that power is moderate when it is .50 and high when it is at least .80. Power increases as α increases, and also as the sample size and/or the effect size increase. The effect size can be conceived as the researcher's idea about "the degree to which the H_0 is believed to be false" (Cohen, 1992, 156).

The relation between effect size, power, significance level, and sample size can be presented in one formula. This formula is an approximation that is valid for practical use when the test in question is a one-sided t-test with a reasonably large number of degrees of freedom (say, d.f. $\geq$ 10). We suppose that the effect size is expressed by a parameter that we can estimate with a certain standard error. Bear in mind that the size of the standard error is a monotone decreasing function of the sample size: the larger the sample size the smaller the standard error! The formula is

$$\text{(effect size / standard error)} \approx (z_{1-\alpha} + z_{1-\beta}) = (z_{1-\alpha} - z_{\beta}) \qquad (1)$$

where z_α, z_β and $z_{1-\beta}$ are the z-scores (values from the standard normal distribution) associated with the indicated α- and β-values. If, for instance, α is chosen at .05 and 1-β at .80 (β thus being .20), and a medium effect size of .50 is what we expect, then we can derive that we are searching for a minimum sample size that satisfies:

$$\text{standard error} \leq [\ .50 / (1.64 + 0.84)\ ]$$
$$\Leftrightarrow$$
$$\text{standard error} \leq 0.20\ .$$

Formula (1) contains 4 "unknowns": this means that if 3 of these are given, then we can compute the fourth. In most applications that we have in mind, the significance level α is given and the effect size is hypothetically considered (or guessed) to have a given value; either the standard error is also known and the power 1-β is calculated, or the power is known and the standard error calculated.

For many types of design one can choose the sample size necessary to achieve a certain level of power on the basis of Cohen's work. For nested designs, however, there are two kinds of sample sizes: the sample size of the micro-units within each macro-unit (n) and the sample size of the macro-units (N), with $N \times n$ being the total sample size for the micro-units. For hierarchically structured ("multilevel") data, where the hypothesis tested refers to one of the regression coefficients ("fixed parameters") in the model, the recipes for standard errors and sample sizes are presented in Snijders & Bosker (1993). The formulae being complex, however, the present software, PINT, was developed to help researchers in designing their studies in the case of hierarchical designs.

In these calculations, it is assumed that for each macro-unit the number of sampled micro-units is equal (the n mentioned above). In practice, this is not often the case. However, the aim of power calculations is to have insight in the order of magnitude of the desired sample sizes and the achieved power, and not to have perfectly precise figures for these quantities. Therefore, n may be interpreted in practice as the average sample size per macro-unit.

The determination of smaple sizes in two-level designs is discussed also in chapter 10 of Snijders & Bosker (1999) (where also parameters of the random part are considered, which are not treated by PINT) and in Snijders (2001). Other papers about determination of sample sizes in multilevel designs are Afshartous (1995) and Mok (1995). These papers both investigate the effect of sample sizes on standard errors by taking one existing data set and simulating (using the Monte Carlo method) two-stage samples from this data set using various sample sizes. Another aspect, the allocation of treatments to subjects or groups, and the gain in precision obtained by including a covariate, is discussed in Raudenbush (1997).

2. The two-level linear model

2.1. A brief introduction to the Hierarchical Linear Model

Since we assume that the user of this software is going to apply hierarchical linear models to test his hypotheses within a multilevel framework, a short introduction to two-level linear models will be presented. For a full-fledged introduction the user is referred to Bryk & Raudenbush (1992), Goldstein (1995), Hox (2002) or Snijders & Bosker (1999). Some of the symbols used are different from those in Snijders and Bosker (1993), because for explaining this computer program the present symbols are simpler.

Model assumptions

We assume a two-level structure. The first level is also the called the micro level, the second level the macro level. At macro level there are N macro-units (for instance, 100 schools), that are assumed to be randomly sampled from an (assumed) infinite population of units. These macro-units are also called groups. At micro level there are n micro-units per macro-unit (for instance, 50 students per school), so that the total sample size is $N \times n$ (in the example $100 \times 50 = 5000$ students).

In practically all research, the number of micro-units per macro-unit is not a fixed number. Since the determination of optimal sample sizes is in practice always an approximate matter, it is not a serious drawback that we have to assume constant group sizes.

The dependent variable is defined at the micro level and is denoted $\boldsymbol{y_{ij}}$ (e.g. the score of student i in school j on an achievement test). The independent variables at micro level are denoted x_{1ij} to x_{kij} and at macro level we have a set of predictor variables z_{1j} to z_{mj}. The two-level model can now be formulated as

$$\boldsymbol{y_{ij}} = \boldsymbol{\beta_{0j}} + \boldsymbol{\beta_{1j}} x_{1ij} + \ldots + \boldsymbol{\beta_{hj}} x_{hij} + \beta_{h+1} x_{(h+1)ij} + \ldots + \beta_k x_{kij} + \boldsymbol{r_{ij}} . \qquad (2)$$

Formula (2) is the micro level model, where a random variable $\boldsymbol{y}$ (e.g. achievement) is regressed on a set of micro level predictors and/or covariates (e.g. aptitude, pretest-score, socio-economic status, motivation, gender, etc.). Note that there are two types of regression coefficients: $\boldsymbol{\beta_{0j}}$ as well as $\boldsymbol{\beta_{1j}}$ to $\boldsymbol{\beta_{hj}}$ (bold faced and indexed with the subscript j) are random coefficients and β_{h+1} to β_k are fixed coefficients. The interpretation of the regression coefficients is straightforward: for every unit increase in x, $\boldsymbol{y}$ is predicted to increase with β units. The fact that $\boldsymbol{\beta_{0j}}$ and $\boldsymbol{\beta_{1j}}$ to $\boldsymbol{\beta_{hj}}$ are random regression coefficients indicates that these coefficients in the regression of $\boldsymbol{y}$ on x may vary between the macro-units.

Next to the micro level model, a macro level model is formulated:

$$\boldsymbol{\beta_{0j}} = \gamma_{00} + \gamma_{01} z_{1j} + \ldots + \gamma_{0m} z_{mj} + \boldsymbol{u_{0j}} \qquad (3a)$$

$$\boldsymbol{\beta_{1j}} = \gamma_{10} + \gamma_{11} z_{1j} + \ldots + \gamma_{1m} z_{mj} + \boldsymbol{u_{1j}} \qquad (3b)$$

$$\vdots$$

$$\boldsymbol{\beta_{hj}} = \gamma_{h0} + \gamma_{h1} z_{1j} + \ldots + \gamma_{hm} z_{mj} + \boldsymbol{u_{hj}} . \qquad (3h)$$

In (3a) the macro-unit specific intercept (i.e., the expected value for the dependent variable in macro-unit *j* in case all the *x*'s are zero) is modelled as a function of the macro level predictors and/or covariates. If the *x*-variables all have a mean of zero, then $\boldsymbol{\beta_{0j}}$ is the expected score for the average micro-unit in the population corresponding to macro-unit *j*. Parameter γ_{00} is sometimes called the GRAND MEAN: it is not really the mean (except in exceptional cases), but it is the expected value for $\boldsymbol{y}_{ij}$ in case all the *x*'s and *z*'s are zero. Parameter γ_{01} is the average regression coefficient associated with the regression of $\boldsymbol{\beta_{0j}}$ on the macro level variable z_1 (for each unit increase in z_1 the predicted increase in $\boldsymbol{\beta_{0j}}$ is γ_{01} units), and similarly for the other γ's. The regression coefficients γ are also called the fixed parameters of the multilevel model.

The micro level residual $\boldsymbol{r_{ij}}$ is assumed to be normally distributed with zero mean and some unknown variance σ^2. This residual is uncorrelated with the macro level residuals $\boldsymbol{u_{0j}}$ to $\boldsymbol{u_{hj}}$. These latter residuals are multivariate normally distributed, with zero mean and variance-covariance $\boldsymbol{\tau}$.

A simple example

Suppose we want to assess the effect of a school policy to enhance the achievement of students from low socio-economic status families, while taking into account aptitude differences between students. Then formulae (2) to (3h) read as follows:

$$\textbf{ACHIEVEMENT}_{ij} = \boldsymbol{\beta_{0j}} + \boldsymbol{\beta_{1j}} SES_{ij} + \beta_2 IQ_{ij} + \boldsymbol{r_{ij}} \qquad (2)$$

$$\boldsymbol{\beta_{0j}} = \gamma_{00} + \gamma_{01}\,\text{POLICY}_j + \boldsymbol{u_{0j}} \qquad (3a)$$

$$\boldsymbol{\beta_{1j}} = \gamma_{10} + \gamma_{12}\,\text{POLICY}_j + \boldsymbol{u_{1j}} \qquad (3b)$$

with $\boldsymbol{r_{ij}} \sim N(0,\sigma^2)$ and $(\boldsymbol{u_{0j}}, \boldsymbol{u_{1j}})' \sim N(\mathbf{0},\boldsymbol{\tau})$, where $\mathbf{0}$ is a vector and $\boldsymbol{\tau}$ is a 2×2 matrix.

The formulae presented above are in line with the notation used in Bryk & Raudenbush (1992). In Snijders & Bosker (1993) a slightly different notation is used to facilitate the specific matrix-partitionings helpful to express the formulae for the standard errors of the regression coefficients. The example written in the Snijders - Bosker notation would be:

$$\textbf{ACHIEVEMENT}_{ij} = \gamma_{11}\,\text{IQ}_{ij} + \boldsymbol{\gamma_{21j}} + \boldsymbol{\gamma_{23j}}\,\text{SES}_{ij} + \boldsymbol{R}_{ij} \qquad (2)$$

$$\boldsymbol{\gamma_{21j}} = \gamma_{21} + \gamma_{22}\,\text{POLICY}_j + \boldsymbol{U}_{21j} \qquad (3a)$$

$$\boldsymbol{\gamma_{23j}} = \gamma_{23} + \gamma_{24}\,\text{POLICY}_j + \boldsymbol{U}_{22j} \qquad (3b)$$

with $\boldsymbol{R}_{ij} \sim N(0,\sigma^2)$ and $(\boldsymbol{u_{21j}}, \boldsymbol{u_{22j}})' \sim N(\mathbf{0},\boldsymbol{\tau})$.

Formulae (2) to (3b) can be combined to one equation by substituting (3a) and (3b) into (2). The result then reads, for this example,

$$\mathbf{ACHIEVEMENT}_{ij} = \gamma_{11}\,\mathrm{IQ}_{ij} + \gamma_{21} + \gamma_{22}\,\mathrm{POLICY}_j + \gamma_{23}\,\mathrm{SES}_{ij}$$

$$+ \gamma_{24}\,\mathrm{SES}_{ij} \times \mathrm{POLICY}_j + \boldsymbol{U}_{21j} + \mathrm{SES}_{ij} \times \boldsymbol{U}_{22j} + \boldsymbol{R}_{ij}\,. \qquad (4)$$

PINT is aimed at the calculation of the standard errors of the estimates of the regression coefficients γ. The formulae given in Snijders and Bosker (1993), that are computed by PINT, are approximations that are sufficiently accurate if the sample sizes are not too small. As a rule of thumb one should require that n is at least 6, and N is at least 10. For testing hypotheses about these regression coefficients, standard errors may be translated into power, and vice versa, using formula (1).

2.2. Information required for PINT

Three types of predictor variables

For the purposes of calculating standard errors, three categories of predictor variables must be distinguished:

(1) micro level variables with only a fixed effect;
(2) micro level variables with a fixed and a random effect;
(3) macro level variables explaining between-group variability.

In the derivation of the formulae for the calculation of optimal sample sizes, it appeared to be convenient to take the micro level predictors with fixed and random effects ("category 2") as deviates from their group means. In other words, for the original X_{ij} the centered variable $X'_{ij} = X_{ij} - X_{.j}$ (with $X_{.j}$ denoting the mean of X in macro-unit j) is used instead. This centering of course leads to a model modification, which might partially be compensated, however, by introducing X_j as a macro level predictor into the model. Returning to our example, SES_{ij} is taken as the deviate from the original school mean SES_j , so it represents the *relative* socio-economic status score for student i in school j.

For simplicity we also assume that every macro level predictor ("category 3") is entered as a predictor for all between-group variability into the model: so it is assumed that when one is interested in a cross-level interaction effect ($\mathrm{SES}_{ij} \times \mathrm{POLICY}_j$ in the example), one is also interested in the main effects of these variables; and that when one is interested in the main effect of a macro level variable, one is also interested in its interaction effect with all micro level predictors having random effects. In actual practice this does not cause problems since in using PINT the user may ignore the standard errors of the regression coefficients for those parts of the equation in which there is no interest. Further, in transforming a priori information into program-input, the researcher may assume that these effects are zero (not affecting residual between school variation).

Information needed for calculating standard errors

PINT calculates approximate standard errors of the fixed regression coefficients, which are denoted γ in the model specifications above. To make these calculations, information is needed about the means, variances and covariances of the predictor variables, as well as about the variances and covariances of the random effects. (If the standard errors are to be translated into statistical power values by formula (1), then we also need an estimate for the effect size.) Such information can be gathered from earlier research or a reasonable guess can be made. This section indicates which information is required.

Numbers of predictor variables

The numbers of variables of "categories 1, 2, and 3", as defined on the top of this page, are indicated in the program description by L_1, L_2, and L_3, respectively. (In the formulae in Section 2.1, k corresponds to L_1+L_2, h corresponds to L_2, and m corresponds to L_3.) The `constant' variable, which represents the intercept, is included anyway in the model, and not included in these numbers L. For example, L_2 indicates the number of random slopes. It is allowed that some or all of these numbers are 0, corresponding to absence of variables in the corresponding category or categories.

Means, variances and covariances of predictor variables

The whole idea of multi-level modelling is, that we are modelling at the micro level as well as at the macro level. In other words: we are modelling within-macro-units variation as well as between-macro-units variation. In determining optimal sample sizes we consequently should be careful about the distinction of these two kinds of variability.

It is very convenient, but not necessary, to assume that all variables are standardized scores (with mean 0 and variance 1). Because of the two-level structure, however, more information is required: also the decomposition of variability into within-group and between-group variability must be given, as well as various correlation coefficients. The distinction between the three categories of variables is important here.

- The variables of "category 1" (micro level variables with only a fixed effect) are assumed to have within-group as well as between-group variability (but it is allowed that some of these variables have only within-group variability).
- Variables of "category 2" (micro level variables with a fixed and a random effect) are assumed to be within-group deviation variables, and have therefore no between-group variability.
- Variables of "category 3" (macro level variables explaining between-group variability) are macro level variables, so that they have no within-group variability.

In the first place, we need the expectations (i.e., population means) of the predictor variables. The vector μ_1 contains the means over schools of the micro level variables with fixed effects only ("category 1"). (If we use z-scores these means are 0). Vector μ_3 contains the means over schools of the macro level predictors ("category 3"). If we are using z-scores this vector again has all elements equal to 0.

In the next place, information is needed about the within- as well as the between-groups covariance matrix of the predictors. This information is given in matrices Σ^W and Σ^B. For input into PINT these are constructed as follows.

Σ^W is a matrix of order $L_1 + L_2$ (the total number of micro level variables). The first elements of the diagonal of this matrix contain the within-group variances of the micro level predictors ("category 1") having fixed effects only. The remaining elements of the diagonal give the within-group variances of the micro level predictors having random effects (if we rescaled all variables to z-scores, these latter variances are 1). The remaining part of this symmetric matrix contains the within-group covariances between the corresponding variables.

There are a total of $L_3 + L_1$ macro level variables: L_3 "original" macro level variables (of "category 3") but also the group means of the L_1 micro level variables of "category 1". The

corresponding covariance matrix Σ^B is therefore a matrix of order $L_3 + L_1$. The first L_3 elements of the diagonal of this matrix contain the between-group variances of the original macro level predictors ("category 3") (if all variables were rescaled to *z*-scores, this variance is 1). The remaining elements of the diagonal refer to the between-group variances of the predictor variables of "category 1" (and if we rescaled all variables to *z*-scores, the sum of this figure and the corresponding within-group variance can be assumed to be 1). The remaining part of this symmetric matrix contains the between-group covariances between these variables.

Variances and covariances of residuals (the random part of the model)
Finally, we need information on the residual variance σ^2 and the covariance matrix $\boldsymbol{\tau}$ of the random intercept and random slopes.

Although it will usually be hard to have all this information on within- and between group variances and covariances, the user of PINT can do a number of runs with the software, providing PINT with various reasonable looking estimates of these parameters, so as to find lower and upper limits for the sample sizes at micro- and macro level.

2.3. Remarks about parameters in the two-level model[1]

Some of the parameters mentioned in Section 2.2, especially the means, variances and covariances of the predictor variables, may be obtained from existing knowledge, combined with intelligent guesswork. Other parameters, especially the parameters of the random part of the model, may be harder to specify. It may be helpful to have some insight into how the parameters jointly imply a decomposition of the variance of the dependent variable.

It is convenient to work with standardized predictor variables, i.e., to have them scaled so that their mean is 0 and their variances 1. In this case, the variance of the dependent variable can be decomposed as follows.

First consider a random intercept model, i.e., a model without random slopes. This is formulated as a special case of (2) (namely, with $h = 0$):

$$\boldsymbol{y_{ij}} = \boldsymbol{\gamma_{00}} + \beta_1 x_{1ij} + \ldots + \beta_k x_{kij} + \boldsymbol{u_{0j}} + \boldsymbol{r_{ij}} .$$

If the predictor variables also are uncorrelated, then the variance of the dependent variable can be decomposed as

$$\mathrm{var}(\boldsymbol{y}_{ij}) = \beta_1^{\,2} + \ldots + \beta_k^{\,2} + \tau_{00} + \sigma^2 , \qquad (5)$$

i.e., the sum of squared regression coefficients, plus the intercept variance, plus the residual level-1 variance. If the predictor variables are correlated, one has to add to this two times the sum of $\beta_r \beta_s \times \rho(x_r, x_s)$, summed over all r and s with $1 \le r < s \le k$ (with $\rho(x_r, x_s)$ being the correlation between the two predictor variables x_r and x_s).

[1] It is best to skip this section, unless you need more insight into the parameters of the multilevel model in order to specify their values for input in PINT. If you do wish to know more about the decomposition of variance of $\boldsymbol{y}$ implied by the multilevel model, you can consult Section 7.2 of Snijders & Bosker (1999).

Second consider a random slope model. Such a model is obtained by substituting (3a) through (3h) into (2), and the result is

$$\begin{aligned} y_{ij} &= \gamma_{00} + \gamma_{10} x_{1ij} + \ldots + \gamma_{h0} x_{hij} + \beta_{h+1} x_{(h+1)ij} + \ldots + \beta_k x_{kij} \\ &\quad + \gamma_{01} z_{1j} + \ldots + \gamma_{0m} z_{mj} + \gamma_{11} x_{1ij} z_{1j} + \ldots + \gamma_{hm} x_{hij} z_{mj} \\ &\quad + u_{0j} + u_{1j} x_{1ij} + \ldots u_{hj} x_{hij} + r_{ij} \quad . \end{aligned}$$

Assume that the variables with random slopes have no between-group variability (i.e., their group means are always 0), and all variables x_r and z_s again have a mean of 0 and a variance of 1. This implies that the product variables $x_r z_s$ also have mean 0 and variance 1.

To keep away from too great complexity, it is also assumed that all variables are uncorrelated. This implies that the variance of y can be decomposed as a sum of squared regression coefficients, plus the intercept variance, plus a sum of squared random slope variances, plus the residual variance:

$$\begin{aligned} \mathrm{var}(y_{ij}) &= \gamma_{10}^2 + \ldots + \gamma_{h0}^2 + \beta_{h+1}^2 + \ldots + \beta_k^2 \\ &\quad + \gamma_{01}^2 + \ldots + \gamma_{0m}^2 + \gamma_{11}^2 + \ldots + \gamma_{hm}^2 \\ &\quad + \tau_{00} + \tau_{11} + \ldots + \tau_{hh} + \sigma^2 \quad . \end{aligned} \tag{6}$$

Since the variables with random slopes are assumed to be pure level-1 variables, the sum of the squared slope variances represents part of the within-group variability.

3. Permitted sample sizes: cost function or all combinations

In the PINT software there are two options for limiting the sample sizes at the two levels. The first option is based on a budget constraint. The second option uses all combinations of sample sizes at either level between certain minimum and maximum values.

Budget constraints option

In the first option, the program takes into account that it may be (but it does not have to be) more costly to sample one extra macro-unit (with n micro-units) than to sample n extra micro-units from already sampled macro-units. E.g., when administering aptitude and achievement tests to students, travelling costs and salary are needed for research assistants and these will be increased by having to visit extra schools.

To give an idea: assume that observation costs are composed of salary, traveling costs, and the material required. Assume that the costs of contacting one school and the travel for the visit is \$150. Further assume that the salary and material costs for testing one students are \$5. E.g., investigating 25 students at one school costs a total of \$150 + 25×\$5 = \$275. More generally, this means that the cost function can be taken as $\$150N + \$5Nn = \$\ 5N(n+30)$.

In general, PINT assumes that the cost function is proportional to $N(n+c)$ for some value c. The program requires that c is an integer number. The number c is the ratio, indicating how much more costly it is to sample one extra macro-unit (without changing the overall total number of micro-units sampled), than it is to sample one extra micro-unit within an already chosen macro-unit. In the example above, $c = 30$. Usually for mail or telephone surveys using a two-stage sampling design $c = 0$: there is no efficiency gain in using a two-stage sample as compared to using a simple random sample. But for research studies in which face-to-face interviewing or supervised testing is required efficiency gains can be made by using a two-stage sampling design, which is reflected by a positive value for c.

Further, PINT assumes that one wants to find sample sizes at micro and macro level that satisfy the following inequality:

$$N(n+c) \leq K\,, \tag{7}$$

with N being the sample size of macro-units, n the sample size of micro-units per macro-unit, and K the budget expressed in monetary units equal to the cost of sampling one additional micro level unit. In the example above, c would be 30 and K would be the budget in dollars divided by \$5. Equation (7) is called a linear budget constraint.

For a number of values of n, PINT calculates the largest value of N that satisfies the budget constraint (7), and computes the standard errors of the parameter estimates for these combinations of n and N. This implies, if n increases with steps of 1, a kind of saw-tooth pattern for the total sample size Nn, because of the requirement that n as well as N be integer numbers. The actual costs $N(n+c)$ can be (slightly) less than the allowed budget K, for some (usually: most) combinations of n and N.

One should realize that the groups are independent replications of each other and therefore the number of groups, N, has the usual effect of a sample size on the standard errors: these are inversely proportional to $\sqrt{N}$, when n remains constant.

All combinations option

The second option is that the user specifies minimum and maximum numbers of the level-1 sample size n, as well as minimum and maximum numbers of the level-2 sample size N. All combinations of n and N between these bounds are in principle permitted; but to limit the quantity of output obtained, the step sizes with which n and N are increased may be less than 1. The incremental step size for n is specified by the user; the incremental step size for N is determined internally by PINT according to the following rule. The minimum and maximum values for N are denoted, respectively, N_{2min} and N_{2max} .

- If $N_{2max} - N_{2min} \leq 20$ then the step size is 1;
- If $20 < N_{2max} - N_{2min} \leq 50$ then the step size is 2;
- If $50 < N_{2max} - N_{2min} \leq 100$ then the step size is 5;
- If $N_{2max} - N_{2min} > 100$ then the step size is 10.

4. Running the PINT-program

4.1. Introduction: PINT for Windows

PINT computes the expressions provided by Snijders and Bosker (1993) for the standard errors of the estimated regression coefficients in the two-level model. This is done for parameter values that have to be provided by the user. The purpose is not to compute standard errors for data that have already been collected and analyzed; such standard errors are provided by the multilevel software used for the data analysis. The purpose is to estimate the standard errors that would be obtained in future research about populations, parameter values of which (means, variances, covariances) can be estimated or guessed, if one would employ given sample sizes.

If one starts with given sample sizes at levels 1 and 2, and an estimated effect size, these standard errors can then be used with formula (1) to compute the statistical power of the design. On the other hand, if one wishes to achieve a given power for a given effect size, one can use formula (1) to calculate the required standard error, let PINT calculate standard errors for a variety of sample sizes, and determine the sample size that yields this value for the standard error.

Parameter file

The Windows version of PINT uses a *parameter file* that contains all parameters values and the constraints to the sample sizes described in Section 3. The user interactively specifies the values used in this parameter file, and this file is then saved for future reference. The parameter file is written in ASCII code (also called a DOS text file). The default name is AX.DAT, but any other name can be specified. If an existing parameter input file is specified, then PINT will read the values from this file, and the user is requested to confirm or change these values.

The normal way to write and change this parameter file is through PINT itself. However, the parameter file can also be written and changed with any editor that can produce ASCII files, provided the format indicated below is used. (For those used to version 1.6 of PINT, the older DOS program PPINT can also still be used, provided that one uses thet “Budget Constraint” option from Section 3.)

4.2. File names

Given that one chooses the parameter file name, e.g., AX.DAT, output is written by default to files AX.OUT (main output: standard errors) and AX.COV (secondary output: covariance matrices of the regression coefficients, which sometimes are useful but often are superfluous). These files are introduced in the sequel of this chapter.

PINT asks the user interactively for the name of the main output file; if the root name (e.g., AX) differs from the root name of the parameter file, e.g., BX.OUT, then the secondary output file will be BX.COV. If the names selected for the input and output files are already in use, PINT will ask for a confirmation, because the files will be overwritten. For the secondary output file, PINT will try to find an unused extension names (.CO1, or if this is in use already then .CO2, etc.). The main output file will indicate the name used for the secondary output file.

PINT writes the names of the output files used at the end of the constructed input file, so you will know later which output files to look at for a given input file. This last line of the input file is not read by PINT, and serves only as information to the user.

4.3. Specifying the input

Sections 2 and 3 indicated the input that is required for calculating the standard errors. This input can be summarized as follows. (The 'categories' mentioned refer to Section 2.2.)

- Number of variables in the three categories (cf. section 2.2): L_1, L_2, and L_3.
- Means of predictor variables: μ_1 and μ_3.
- Within-groups covariance matrix Σ^W of order L_1+L_2 (for predictor variables of categories 1 and 2: those in category 3 have no within-group variance).
- Between-groups covariance matrix Σ^B of order L_1+L_3 (for predictor variables of categories 1 and 3: those in category 2 have no between-group variance). Note that the between-group variances and covariances of the variables in category 1 (which are level-1 variables) are the variances and covariances of their group means.
- Residual variance at level 1, σ^2.
- Covariance matrix of the random coefficients (random intercept and random slopes): covariance matrix $\boldsymbol{\tau}$ of order L_2+1. In this covariance matrix, the first variable corresponds to the random intercept, the last L_2 to the random slopes.
- Choice between the option of a linear budget constraint, or all combinations of level-1 and level-2 sample sizes between certain bounds.
- For the first option: cost parameter c and budget constraint K.
- For the second option: minimum and maximum level-2 sample sizes N_{2min} and N_{2max}.

In addition, to indicate for which values of n the standard errors are to be calculated, required input is

- the smallest and largest values of n, to be called $n_{\min}$ and $n_{\max}$, and the step size n_{step} with which n is to be increased in going from $n_{\min}$ to $n_{\max}$.

With respect to the covariance matrices Σ^B and Σ^W, the first blocks of these, of dimensions $L_1 \times L_1$, refer to the variables in category 3. The overall covariance matrix of these variables is the sum of the two corresponding blocks. This implies, e.g., that if all variables are z-scores and have variances equal to 1, then it holds for each $i = 1, \ldots, L_1$ that the sum $(\Sigma^B)_{i\,i} + (\Sigma^W)_{(L3+i,\,L3+i)}$ is equal to 1.

The various parameters have to be determined from existing knowledge, supplemented by intelligent guesswork. What is said in Section 2.3 about the decomposition of the variance of the dependent variable may be used to get some insight into the values that the various parameters could have.

4.3.1. Parameter input

PINT asks the values of these parameters in the following order. Logical constraints are required for some of the parameters (e.g., variances may not be negative, etc.).

- Number of predictor variables:
  the total number of level-1 variables, $L_1 + L_2$;
  the number of variables with only a fixed effect, L_1;
  the number of level-2 variables, L_3.
- Means of predictor variables:
  the means of the level-1 variables with fixed effects only, i.e., L_1 numbers together forming the vector μ_1;
  then the means of the level-2 variables, i.e., L_3 numbers together forming the vector μ_3.
- Within-groups covariance matrix Σ^W:
  The variables must be ordered so that first come the L_1 variables in category 1, then the L_2 variables in category 2.
  Since the matrix is symmetric, only the sub-diagonal part needs to be given. Each row of the covariance matrix must be given on a separate input line. A total of $L_1 + L_2$ input lines are required, the first containing one number, the second 2 numbers, etc., up to the last one containing $L_1 + L_2$ numbers.
- Between-groups covariance matrix Σ^B:
  The variables must be ordered so that first come the L_3 variables in category 3, then the L_1 variables in category 1.
  Again, only the sub-diagonal part needs to be given and each row of the covariance matrix must be given on a separate input line. A total of $L_3 + L_1$ input lines are required, the first containing one number, the second two numbers, etc., up to the last one containing $L_3 + L_1$ numbers.
- Residual variance at level 1, σ^2.
- Covariance matrix of the random coefficients (first random intercept, then random slopes): covariance matrix τ of order $L_2 + 1$. It must be given in the same way as the other covariance matrices.
- Choice between the option of a linear budget constraint, or all combinations of level-1 and level-2 sample sizes between certain bounds.
- For the first option: cost parameter c and budget constraint K; and the smallest and largest values of n, to be called $n_{\min}$ and $n_{\max}$, and the step size n_{step} with which n is to be increased in going from $n_{\min}$ to $n_{\max}$.
- For the second option: minimum and maximum level-2 sample sizes N_{2min} and N_{2max}; and here also the smallest and largest values of n, to be called $n_{\min}$ and $n_{\max}$, and the step size n_{step} with which n is to be increased in going from $n_{\min}$ to $n_{\max}$.

After having given in all these values, the user has to choose one of the following optons to proceed:

1. Save the parameter file and calculate the standard errors.
2. Save the parameter file without calculating the standard errors.
3. Quit the program without saving the parameter file or doing any calculations (in this case, all specifications given will be lost).

4.3.2. Format of the parameter file

If you wish to write a parameter file for PINT directly (using some text editor), the required input order is the following.

In the option of a budget constraint:

L_1+L_2 L_1 L_3
$n_{\min}$ n_{step} $n_{\max}$
K c
σ^2
$\boldsymbol{\tau}$
$\boldsymbol{\Sigma}^{\mathbf{W}}$
$\boldsymbol{\Sigma}^{\mathbf{B}}$
μ_1
μ_3

In the option of all combinations of sample sizes between certain bounds:

L_1+L_2 L_1 L_3
$n_{\min}$ $-n_{\text{step}}$ $n_{\max}$
N_{2min} N_{2max}
σ^2
$\boldsymbol{\tau}$
$\boldsymbol{\Sigma}^{\mathbf{W}}$
$\boldsymbol{\Sigma}^{\mathbf{B}}$
μ_1
μ_3

(note that the step size n_{step} now is given with a minus sign, i.e., it is multiplied by –1).

The matrices $\boldsymbol{\tau}$, $\boldsymbol{\Sigma}^{\mathbf{W}}$, and $\boldsymbol{\Sigma}^{\mathbf{B}}$ may be given as entire matrices, or as lower diagonal matrices; only the lower diagonal part is read. They are read as one matrix row per input line. (In other words: after reading the diagonal element, the program passes on to reading the next line.)
All other input elements are read in free format. This implies that it is not an error if the first 8 integer numbers (L_1+L_2 up to c) are given in a total of more than 3 lines, provided the order of these numbers is as indicated. Comments are not allowed in the input file before μ_3. After this, you may give any comments you like.
The first 8 input variables (until the value of c) must be integers (no decimal point). The other variables are read as real numbers (may, but need not, contain a decimal point.)
The matrix $\boldsymbol{\tau}$ must be of order L_2+1, matrix $\boldsymbol{\Sigma}^{\mathbf{W}}$ must be of order L_1+L_2, and matrix $\boldsymbol{\Sigma}^{\mathbf{B}}$ of order L_3+L_1. The vector μ_1 has L_1 elements, while μ_3 has L_3 elements.

A possible input file AX.DAT is as follows:

```
   2 1     1
  10 5    60
1000 5
  0.5
  0.09
 -0.01    0.0075
  0.8
  0.3     1
  1
 -0.13    0.2
  0
  0
```

4.3.3. An example

We will illustrate the input by proceeding with the example provided earlier that was also treated in Snijders and Bosker (1993).
In this example the effect of a POLICY variable on ACHIEVEMENT of students is assessed after taking into account IQ and SES differences between students. Furthermore this variable may affect the within-school relation between ACHIEVEMENT and SES, i.e. we hypothesize a cross-level interaction effect of SES×POLICY on ACHIEVEMENT. The two-level model was expressed in Section 2.1 by the formula

$$\mathbf{ACHIEVEMENT}_{ij} = \gamma_{11}\,\mathrm{IQ}_{ij} + \gamma_{21} + \gamma_{22}\,\mathrm{POLICY}_j + \gamma_{23}\,\mathrm{SES}_{ij}$$
$$+ \gamma_{24}\,\mathrm{SES}_{ij}\times\mathrm{POLICY}_j + \boldsymbol{U}_{21j} + \mathrm{SES}_{ij}\times\boldsymbol{U}_{22j} + \boldsymbol{R}_{ij}\,. \qquad (4)$$

We suppose that the standard errors demanding our interest are those that are associated with the regression coefficients γ_{22} and γ_{24}. The various parameters for input into PINT are determined as follows.

- Number of variables in the three categories: L_1, L_2, and L_3.
  In our example there are two level-1 predictor variables (IQ and SES) so that $L_1 + L_2 = 2$.
  The only micro level predictor with a fixed effect is IQ, so $L_1 = 1$.
  There is only one level-2 predictor in the model (POLICY), so $L_3 = 1$.
  Recall that for each level-2 predictor it is assumed that it predicts intercept-differences between macro-units as well as differences in (random modeled) regression coefficients between macro-units.

- Means of predictor variables with fixed effects only: μ_1 and μ_3.
  All variables are used as *z*-scores, thus both $\mu_1 = 0$ (a vector of $L_1 = 1$ element) and likewise $\mu_3 = 0$ ($L_3 = 1$ element).

- Within-groups covariance matrix Σ^W of order $L_1 + L_2$.

  In this case, the level-1 variables are IQ and SES (in this order: IQ is in "category 1" while SES is in "category 2"; see page 7). Since we are dealing with a hierarchical design, the predictor variables at level 1 will usually have within-group as well as between-group (co)variation. The within-group covariance matrix is the matrix Σ^W. Since all variables are *z*-scores, the only problems in specifying this matrix deal with the following:
  - how much of the variance in the fixed effect level-1 variables is located at level 1?
  - what is the within-group covariance between these variables?
  - what is the within-group covariance between these variables and the random effects variables?
  - what is the within-group covariance between the random effects variables?

  For the present example IQ is the fixed effect variable at level 1, and we assume that 80 percent of its variance is located at level 1, and the remaining variance (20 percent) is located at level 2. Since random effects variables should be taken as within-group deviates and are *z*-scores, their variance can be taken to be 1. The only specification problem that remains is the within-group covariance between IQ and SES. For the present example we assume it to be 0.30, so that Σ^W is

  0.80 0.30
  0.30 1.00

  This matrix indicates the values of

  level-1 var in IQ covar between IQ and SES
  covar between IQ and SES level-1 var in SES .

- Between-groups covariance matrix Σ^B of order $L_3 + L_1$.

  The variable in category 1 is IQ, the variable in category 3 is POLICY. Here we need some a priori information, or an educated guess, on the level-2 (between-group) variances and covariances. The order of variables is: first original level-2 variables (POLICY), then group means of level-1 variables with fixed effects (IQ).
  The variance at level-2 of the IQ-variable is already implied by what was decided about the amount of level-1 variance in this variable: the remaining 20% is its between-group variance. The POLICY variable is a *z*-score, so has variance 1. The specification of the between-group covariance between IQ and POLICY, which is the covariance between the group-average $IQ_{.j}$ and the POLICY variable, can be deduced from our guess about the correlation since

$$\text{correlation}(\mathbf{x}, \mathbf{y}) = \frac{\text{covariance}(\mathbf{x}, \mathbf{y})}{\text{s.d.}(\mathbf{x}) \times \text{s.d.}(\mathbf{y})} .$$

  In the present example we assumed a negative correlation (because the POLICY was aimed at low-SES and therefore on average lower-IQ students) of -.30, which, given standard deviations of $\sqrt{0.20} = 0.45$ and $\sqrt{1.00} = 1.00$, results in a covariance of -.13.

For the present example this leads to the following specification of Σ^B:

1.00 -0.13
-0.13 0.20

This matrix indicates the values of

var in POLICY	covar between POLICY and $IQ_{.j}$
covar between POLICY and $IQ_{.j}$	var in $IQ_{.j}$.

- Residual variance at level 1, σ^2.
  Since we have assumed that all variables were *z*-scores, the residual variance at the micro level is the amount of variation accounted for neither by the variables in the model nor by the grouping (the random intercept). The range is from 0 (if there is no residual variance) to 1 (no explanatory variables in the model). For the present example we take σ^2 = 0.5, meaning that half of the variation in the dependent variable (ACHIEVEMENT) is accounted for by variables in the model, whether these are the fixed effects (main effects of IQ, SES, and POLICY; interaction effect SES×POLICY) or the random effects (the schools that the students are in).

- Covariance matrix of the random coefficients (random intercept and random slopes): covariance matrix $\boldsymbol{\tau}$ of order $L_2 + 1$.
  Here the relevant sizes of the residual variance components at level 2 will be specified. To start with, var($\boldsymbol{U_{21j}}$)=τ_{00}, the random intercept variance, which is the amount of residual between-group variance in the dependent variable given that all level-1 variables with a random slope have a value of 0, has to be specified. Since all variables with random slopes are within-group deviation variables, this can be identified with the amount of variance in the dependent variable located at level 2 that is left after the fixed effects have been controlled for. We suppose that 20% of the variance of achievement is a school effect, half of which is explained by $IQ_{.j}$ (schools having students with higher IQ-scores having higher ACHIEVEMENT-scores) and that about 10% of the residual variation in school effects may be explained by the POLICY variable. There remains a random intercept variance of 0.09.
  We also assume that the regression coefficient of ACHIEVEMENT on SES has a variance of 0.01 (i.e., a standard deviation of 0.1) and that about 25% of this variance is accounted for by the regression of this slope on POLICY. There remains an unexplained random slope variance of 0.0075. Finally, the residual intercept-slope covariance is estimated at -0.01, corresponding to a correlation of -0.39. (This coefficient is rather hard to specify; fortunately, the resulting standard errors as calculated by PINT are not very sensitive to it.)
  The resulting covariance matrix is

  0.09 -0.01
  -0.01 0.0075

This matrix indicates the values of

residual intercept variance	intercept-slope covariance
intercept-slope covariance	residual slope variance.

- The smallest and largest values of n, called n_{min} and n_{max}, and the step size n_{step} with which n is to be increased in going from n_{min} to n_{max}.
  The smallest practical group size here is n_{min}=10. To get a first idea about the behaviour of the standard errors, we set n_{step}=5.
  The level-2 units of course mostly set a ceiling to the number of level-1 units that are available. When investigating neighbourhood effects on political affiliation of the inhabitants, the number is rather large, but in educational research, schools will have a limited total number of, e.g., 30 (primary schools) or 200 students (secondary schools) per grade. This then sets the limit. For the present example we assume that n_{max}=60. Since n_{step} is given as a positive number, PINT knows that the option of a linear budget constraint is used.

- Cost parameter c and budget constraint K.
  These amounts are expressed as multiples of the marginal cost of having one extra student; in other words, this marginal cost is used as the unit of cost.
  For the present example we take K=1000, implying that in any case we cannot have a total sample size of more than 1000 students.
  Parameter c represents the extra costs involved in sampling an extra macro-unit. For this example we take c=5, meaning that sampling the same total amount of students in one extra school is 5 times more expensive than taking one student extra in one already sampled school.

The corresponding parameter file AX.DAT for PINT, that was already given at the end of Section 4.3.2, then reads:

```
   2 1     1
  10 5    60
1000 5
  0.5
  0.09
 -0.01    0.0075
  0.8
  0.3     1
  1
 -0.13    0.2
  0
  0
```

4.4. Output

The output is written to two files. The first (default name AX.OUT) contains the standard errors of the estimated regression coefficients. The second (default name AX.COV) contains the variance-covariance matrices of the estimates.

4.4.1. Primary output on AX.OUT

The primary output is written to AX.OUT. In discussing the output we will proceed with the output as produced by the example given. Italicised remarks are explanations that are not part of the output.

(First comes the program header.)

```
  PPPPPP       III    NNN      NNN  TTTTTTTTTTTTT
  PPP   PPP           NNNN     NNN       TTT
  PPP   PP     III    NNN NN   NNN       TTT
  PPPPPP       III    NNN  NN  NNN       TTT
  PPP          III    NNN   NN NNN       TTT
  PPP          III    NNN    NNNNN       TTT
  PPP          III    NNN     NNNN       TTT
  PPP          III    NNN      NNN       TTT

            power in two-level designs
            version 2.1       april 2003

copyright:   roel bosker    (RUG)  & tom snijders (RUG)
programming: henk guldemond (RUG)  & tom snijders (RUG)

This programs performs calculations corresponding to the paper
"Standard errors and Sample Sizes for Two-Level Research",
by Tom A.B. Snijders and Roel J. Bosker,
Journal of Educational Statistics, Vol. 18, 1993, p. 237-259.

Date and time: 4/27/2003 2:18:00 PM

Input parameters read from file C:\Tom\Mulev\Pint\Examples\ax1.dat.

File C:\Tom\Mulev\Pint\Examples\ax1.cov contains secondary output
(entire covariance matrices).
```

(The first section of the output summarizes the input as specified on AX.DAT)

```
Design:
(Between parentheses, the symbol is mentioned that is used for this parameter;
 if the PINT manual uses a different symbol than the Snijders-Bosker paper,
 then the symbol from the paper is given between parentheses,
 the symbol from the manual between square brackets.)

NUMBER OF FIXED EFFECTS              (K_1)           [L_1 + L_2 + 1] :    1
NUMBER OF RANDOM EFFECTS INCL. CONST (K_2)           [L_2 + 1]       :    2
NUMBER OF LEVEL-2 VARS INCL. CONST   (length of W_3j) [L_3 + 1]      :    2
TOTAL COSTS                          (k)             [K]             : 1000
RELATIVE COST PER LEVEL-2 UNIT       (c)                             :    5
SMALLEST VALUE OF n                                  [n_min]         :   10
STEP SIZE FOR n                                      [n_step]        :    5
LARGEST VALUE FOR n                                  [n_max]         :   60
```

(The parameters given next are in expanded form, also containing elements corresponding to the "constant" as one of the explanatory variables. This is to give exact correspondence with Snijders and Bosker (1993). This gives extra rows and columns of zeros in the covariance matrices, an extra zero in the vector μ_3 of expected values for the level-2 variables, and an

extra vector μ_2 consisting of a one followed by zeros. You may forget about all this, unless you want to study very precisely the correspondence with the paper.)

```
Parameters:

WITHIN-GROUPS COVARIANCE MATRIX  (SIGMA-W)

      0.80000      0.00000      0.30000
      0.00000      0.00000      0.00000
      0.30000      0.00000      1.00000


BETWEEN-GROUPS COVARIANCE MATRIX (SIGMA-B)

      0.00000      0.00000      0.00000
      0.00000      1.00000     -0.13000
      0.00000     -0.13000      0.20000


RESIDUAL VARIANCE (sigma-squared)

          0.50000

COVARIANCE MATRIX OF RANDOM EFFECTS (tau_2)

      0.09000     -0.01000
     -0.01000      0.00750


EXPECTATION OF LEVEL-1 VARIABLES WITH FIXED EFFECTS (mu_1)

      0.00000


EXPECTATION OF LEVEL-2 VARIABLES (mu_3)

      1.00000
      0.00000


CONSTANT MEAN VECTOR (mu_2 = e)

      1.00000
      0.00000
```

(An explanation of the produced table of standard errors follows.)

```
The following table contains the standard errors (s.e.):
Fixed:   s.e. of regr. coeff.s of level-1 variables with a fixed effect only.
Const:   s.e. of the intercept.
Group:   s.e. of regr. coeff.s of level-2 variables.
Random:  s.e. of regr. coeff.s of level-1 variables with a random effect.
Cross-L: s.e. of regr. coeff.s of cross-level interactions
         (product of "Group" with "Random effect" variables).
```

(And now for the real output.)

```
    Sample sizes      costs   Standard errors

    N*n     N     n (n+c)*N    Fixed   Const   Group   Random Cross-L
    660    66    10     990   0.03124 0.04606 0.04624 0.03097 0.02952
    750    50    15    1000   0.02961 0.04967 0.04981 0.02993 0.02858
    800    40    20    1000   0.02885 0.05362 0.05375 0.02979 0.02850
    825    33    25     990   0.02853 0.05774 0.05785 0.03011 0.02887
    840    28    30     980   0.02837 0.06172 0.06183 0.03059 0.02938
    875    25    35    1000   0.02786 0.06459 0.06469 0.03068 0.02952
    880    22    40     990   0.02783 0.06826 0.06835 0.03129 0.03015
    900    20    45    1000   0.02756 0.07110 0.07119 0.03161 0.03051
```

```
   900    18    50     990   0.02760 0.07454 0.07462 0.03226 0.03118
   880    16    55     960   0.02794 0.07870 0.07878 0.03327 0.03220
   900    15    60     975   0.02765 0.08097 0.08105 0.03353 0.03249
```

The first three columns contain information on the sample sizes; since in this example $c > 0$, the total sample size will tend to increase when less macro-units are sampled. The next five columns contain information on the standard errors; generally the standard error of the level-1 fixed effect will decrease as the total sample size ($N \times n$) increases whereas for the constant (i.e. the intercept) and the macro-level variable the standard error will increase with a decrease in the sample size of the macro-units N (at least when the amount of residual variance in the dependent level at level-2 exceeds 0). The 6th and 8th column contain the standard errors of primary interest. The standard errors of the main effect of the POLICY-variable as well as the standard error of the cross-level interaction effect POLICY×SES appear to have an optimum somewhere between $10 < n < 40$. We take these boundaries rather loosely since N and n have to be integer numbers such that $N\ (n + c) \leq K$, which implies that $N \times n$ can vary substantially when the number of micro-units increases.

4.4.2. Secondary output on AX.COV

AX.COV contains the variance-covariance matrix of the estimated coefficients. The square root of the diagonal elements are the standard errors of the primary output. The covariance matrices can be used if it is necessary to calculate standard errors of linear combinations of the estimated coefficients. If you have messed up your files, you can look at the date and time when the output was produced to check the correspondence between the AX.COV and the AX.OUT files.

The variance-covariance matrix of the estimated coefficients has the same structure as the produced table of standard errors in the primary output. So in this example the structure is, according to the head of this table:

Fixed:	s.e. of regr. coeff.s of level-1 variables with a fixed effect only.
Const:	s.e. of the intercept.
Group:	s.e. of regr. coeff.s of level-2 variables.
Random:	s.e. of regr. coeff.s of level-1 variables with a random effect.
Cross-L:	s.e. of regr. coeff.s of cross-level interactions (product of "Group" with "Random effect" variables).

```
4/27/2003 2:18:00 PM

N = 66 ; n     = 10

Covariance matrix of estimated coefficients

   0.00097599   0.00000000   0.00012688  -0.00029280   0.00000000
   0.00000000   0.00212121   0.00000000  -0.00015152  -0.00000000
   0.00012688   0.00000000   0.00213771  -0.00003806  -0.00015152
  -0.00029280  -0.00015152  -0.00003806   0.00095905  -0.00000000
   0.00000000   0.00000000  -0.00015152  -0.00000000   0.00087121

N = 50 ; n     = 15
```

Covariance matrix of estimated coefficients

0.00087663	-0.00000000	0.00011396	-0.00026299	0.00000000
-0.00000000	0.00246667	0.00000000	-0.00020000	-0.00000000
0.00011396	0.00000000	0.00248148	-0.00003419	-0.00020000
-0.00026299	-0.00020000	-0.00003419	0.00089556	0.00000000
0.00000000	-0.00000000	-0.00020000	0.00000000	0.00081667

N = 40 ; n = 20

Covariance matrix of estimated coefficients

0.00083234	-0.00000000	0.00010820	-0.00024970	-0.00000000
0.00000000	0.00287500	0.00000000	-0.00025000	0.00000000
0.00010820	0.00000000	0.00288907	-0.00003246	-0.00025000
-0.00024970	-0.00025000	-0.00003246	0.00088741	0.00000000
-0.00000000	-0.00000000	-0.00025000	0.00000000	0.00081250

N = 33 ; n = 25

Covariance matrix of estimated coefficients

0.00081413	-0.00000000	0.00010584	-0.00024424	-0.00000000
-0.00000000	0.00333333	-0.00000000	-0.00030303	0.00000000
0.00010584	-0.00000000	0.00334709	-0.00003175	-0.00030303
-0.00024424	-0.00030303	-0.00003175	0.00090660	0.00000000
-0.00000000	0.00000000	-0.00030303	0.00000000	0.00083333

N = 28 ; n = 30

Covariance matrix of estimated coefficients

0.00080463	0.00000000	0.00010460	-0.00024139	0.00000000
0.00000000	0.00380952	0.00000000	-0.00035714	0.00000000
0.00010460	0.00000000	0.00382312	-0.00003138	-0.00035714
-0.00024139	-0.00035714	-0.00003138	0.00093551	-0.00000000
0.00000000	-0.00000000	-0.00035714	-0.00000000	0.00086310

```
N = 25 ; n     = 35

Covariance matrix of estimated coefficients

   0.00077615   0.00000000   0.00010090  -0.00023284   0.00000000
   0.00000000   0.00417143  -0.00000000  -0.00040000   0.00000000
   0.00010090   0.00000000   0.00418455  -0.00003027  -0.00040000
  -0.00023284  -0.00040000  -0.00003027   0.00094128  -0.00000000
   0.00000000  -0.00000000  -0.00040000  -0.00000000   0.00087143


N = 22 ; n     = 40

Covariance matrix of estimated coefficients

   0.00077464   0.00000000   0.00010070  -0.00023239   0.00000000
   0.00000000   0.00465909  -0.00000000  -0.00045455   0.00000000
   0.00010070  -0.00000000   0.00467218  -0.00003021  -0.00045455
  -0.00023239  -0.00045455  -0.00003021   0.00097881   0.00000000
   0.00000000   0.00000000  -0.00045455   0.00000000   0.00090909


N = 20 ; n     = 45

Covariance matrix of estimated coefficients

   0.00075973   0.00000000   0.00009877  -0.00022792  -0.00000000
   0.00000000   0.00505556   0.00000000  -0.00050000   0.00000000
   0.00009877   0.00000000   0.00506840  -0.00002963  -0.00050000
  -0.00022792  -0.00050000  -0.00002963   0.00099893   0.00000000
   0.00000000   0.00000000  -0.00050000  -0.00000000   0.00093056

N = 18 ; n     = 50

Covariance matrix of estimated coefficients

   0.00076164   0.00000000   0.00009901  -0.00022849   0.00000000
   0.00000000   0.00555556   0.00000000  -0.00055556   0.00000000
   0.00009901   0.00000000   0.00556843  -0.00002970  -0.00055556
  -0.00022849  -0.00055556  -0.00002970   0.00104077  -0.00000000
   0.00000000   0.00000000  -0.00055556   0.00000000   0.00097222


N = 16 ; n     = 55

Covariance matrix of estimated coefficients

   0.00078059  -0.00000000   0.00010148  -0.00023418   0.00000000
  -0.00000000   0.00619318  -0.00000000  -0.00062500  -0.00000000
   0.00010148   0.00000000   0.00620637  -0.00003044  -0.00062500
  -0.00023418  -0.00062500  -0.00003044   0.00110719   0.00000000
   0.00000000  -0.00000000  -0.00062500  -0.00000000   0.00103693


N = 15 ; n     = 60

Covariance matrix of estimated coefficients

   0.00076462   0.00000000   0.00009940  -0.00022938  -0.00000000
   0.00000000   0.00655556  -0.00000000  -0.00066667   0.00000000
   0.00009940   0.00000000   0.00656848  -0.00002982  -0.00066667
  -0.00022938  -0.00066667  -0.00002982   0.00112437   0.00000000
  -0.00000000   0.00000000  -0.00066667   0.00000000   0.00105556
```

5. Examples

This section contains a number of examples. For the examples of Sections 5.1 and 5.3, the exact methods of Snijders (2001, Sections 11.4 and 11.5) could also be applied (but those methods have a much smaller scope than the approximations used in PinT). Another extensive example of PinT is given in Snijders (2001, Section 11.9); but this example has an error in the between-groups covariance between X_1 and X_3, which should be –0.043 instead of +0.043. Still another example is given in Snijders & Bosker (1999, Section 10.4).

5.1. An assessment study with a two-stage sample

In this first example we have an international assessment study on mathematics achievement in secondary schools. The mathematics achievement variable is a *z*-score, and within each country the mean should be estimated with a standard error of .02. If a random sample would be taken it can readily be deduced that the sample size should be:

$$\text{s.e.} = \text{s.d.}/\sqrt{n}$$
$$\Leftrightarrow$$
$$n = 1/.02^2 = 2{,}500.$$

What will happen to the standard error if a two-stage sampling scheme is employed (first schools then students), in case the between-school variance is 0.20, and assuming that there are no direct extra budgetary consequences of sampling schools (this might be the case where one is estimating costs, when the standard errors are imposed by some international board)?
The model can be formulated as follows:

$$\mathbf{ACHIEVEMENT}_{ij} = \boldsymbol{\beta}_{0j} + \boldsymbol{r}_{ij} \qquad (2)$$

$$\boldsymbol{\beta}_{0j} = \gamma_0 + \boldsymbol{u}_{0j} \qquad (3a)$$

with $\boldsymbol{r}_{ij} \sim N(0,\sigma^2)$ and $\boldsymbol{u}_{0j} \sim N(\mathbf{0},\boldsymbol{\tau})$. Note that we have no explanatory variables at all. In multilevel terminology, this is called the empty model.

The parameter file would have to be specified as follows:

```
0      0      0
```

(0 level-1 predictor variables; none of these has a fixed effect; 0 level-2 variables)

```
1      1     30
```

(start with 1 micro-unit per macro-unit, then run with an increase of 1 micro-unit per macro-unit until a maximum of 30 micro-units per macro-unit)

```
 2500      0
```

(total budget available is 2500 units and it will cost 0 units extra when sampling an additional school)

```
       0.80
```

(the residual variance at level 1)

```
0.2
```

(the residual covariance matrix at level 2, which is simply the variance in mathematics achievement between schools).

Thus AX.DAT would look like:

```
   0     0     0
   1     1    30
2500     0
    0.80
    0.20
```

The primary output file AX.OUT then contains the following output concerning the standard errors:

```
 Sample sizes      costs   Standard errors

 N*n     N     n (n+c)*N     Const
2500  2500     1    2500   0.02000
2500  1250     2    2500   0.02191
2499   833     3    2499   0.02367
2500   625     4    2500   0.02530
2500   500     5    2500   0.02683
2496   416     6    2496   0.02831
2499   357     7    2499   0.02967
2496   312     8    2496   0.03101
2493   277     9    2493   0.03229
2500   250    10    2500   0.03347
2497   227    11    2497   0.03466
2496   208    12    2496   0.03581
2496   192    13    2496   0.03691
2492   178    14    2492   0.03801
2490   166    15    2490   0.03907
2496   156    16    2496   0.04003
2499   147    17    2499   0.04100
2484   138    18    2484   0.04209
2489   131    19    2489   0.04299
2500   125    20    2500   0.04382
2499   119    21    2499   0.04473
2486   113    22    2486   0.04574
2484   108    23    2484   0.04663
2496   104    24    2496   0.04737
2500   100    25    2500   0.04817
2496    96    26    2496   0.04903
2484    92    27    2484   0.04996
2492    89    28    2492   0.05068
2494    86    29    2494   0.05144
2490    83    30    2490   0.05226
```

Here you can clearly see the saw-tooth pattern in the total sample size $N \times n$. For $n = 1$, as expected, the standard error is exactly the desired 0.2. It can be seen that, since the between-school variance is larger than 0, the standard errors deteriorate as the total number of schools in the sample increases.

Suppose now that practical restrictions lead to sampling one class of, on average, $n = 30$ pupils per school. For $n = 30$, the standard error in the output above is seen to be 0.05226 for $N = 83$ groups. Since the groups are independent replications, multiplying the number of groups by some factor f will multiply the standard error by a factor $1/\sqrt{f}$. This suggests that to arrive at a standard error of 0.02, which amounts to multiplication by $1/\sqrt{f} = 0.02/0.05226$, the number of schools should be multiplied by $f = (0.05226/0.02)^2 = 6.83$. The total sample size then would be approximately ($6.83 \times 83 \times 30 =$) 17,000. Running PINT with a budget of $K = 17000$ for n from 25 to 35 leads to the following output:

```
   N*n     N     n     Const
 17000   680    25   0.01847
 16978   653    26   0.01880
 16983   629    27   0.01911
 16996   607    28   0.01941
 16994   586    29   0.01971
 16980   566    30   0.02001
 16988   548    31   0.02030
 16992   531    32   0.02058
 16995   515    33   0.02087
 17000   500    34   0.02114
 16975   485    35   0.02144
```

This shows that sampling $N = 566$ schools with each $n = 30$ pupils will yield a standard error of 0.02, as desired.

Because this design, without explanatory variables, is so simple, we could also have obtained this result from existing formulae. Cochran (1977, Chapter 9) provides formulae to calculate sample sizes in case of two-stage sampling. On p. 242, he indicates that the "design effect" for a two-stage sample, which is the factor by which the variance of an estimate is increased because of using a two-stage sample rather than a simple random sample, is given by

$$\text{design effect} = 1 + (n - 1) \times \rho ,$$

where ρ is the intra-class correlation (ratio of between-group variance to total variance, given by $\tau_{00}/(\tau_{00} + \sigma^2)$). The design effect is also the factor by which the total sample size has to be multiplied to obtain the same standard error.
In this case, $\rho = 0.2$. Using Cochran's formulae and assuming that per school one class of students will be sampled, so that $n = 30$, the total sample size for a two-stage random sample that is equivalent to a simple random sample of size 2,500 should be

$$Nn = 2{,}500 \times (1 + (30 - 1) \times 0.2) = 17{,}000.$$

5.2. An association between level-1 variables using a two-stage sample

Suppose one wants to assess the association between, for instance, income and total number of years spent in school as part of a larger national survey using face-to-face interviews. Since interviewers have to travel to the respondents it seems worthwile to reduce travelling costs and to take a two-stage sample: randomly select neighbourhoods and within each neighbourhood select a number of respondents. The model (a random intercept model) would be as follows:

$$\mathbf{INCOME}_{ij} = \gamma_{00} + \gamma_{10}\text{YEARS}_{ij} + \boldsymbol{u_{0j}} + \boldsymbol{r_{ij}}\ .$$

Let us assume that INCOME as well as YEARS are *z*-scores (mean 0, total variance 1), that YEARS has a between-neighbourhood variance of 0.10 and therefore a within-neighborhood variance of 0.90, that 20% of INCOME is explained by YEARS (so that the correlation between YEARS and INCOME is 0.447), and that the 80% unexplained variance of INCOME is distributed over individuals and neighbourhoods in the ratio of 70:10. Let us assume that sampling one extra neigbourhood has a price attached to it - over and above the price associated with sampling individuals within this neighbourhood - that is the equivalent of sampling 8 individuals.

The following might be a specification of the PINT-input on the parameter file AX.DAT:

```
   1   1   0
   8   2  30
1000   8
0.70
0.10
0.90
0.10
   0
```

AX.OUT will contain the following information on the standard errors:

```
   N*n     N     n (n+c)*N    Fixed   Const
   496    62     8     992   0.03861 0.05499
   550    55    10     990   0.03677 0.05560
   600    50    12    1000   0.03529 0.05627
   630    45    14     990   0.03450 0.05774
   656    41    16     984   0.03387 0.05921
   684    38    18     988   0.03321 0.06046
   700    35    20     980   0.03286 0.06211
   726    33    22     990   0.03230 0.06320
   744    31    24     992   0.03193 0.06455
   754    29    26     986   0.03175 0.06616
   756    27    28     972   0.03172 0.06804
   780    26    30     988   0.03125 0.06887
```

Since the association between years of schooling and income is the object of investigation, the relevant standard errors are those in the column "Fixed" (this column gives the standard errors of γ_{10}). It turns out that, given the cost parameter c = 8, and given that there are reasons not to sample more than n = 30 persons per neigbourhood, it is best to sample the lowest possible number of neighbourhoods, N = 26. The resulting standard error will be 0.03125. If we were interested in the power of the statistical tests to be employed in this study, we could apply (1). Suppose we would set α at .01, then the power would be derived to be:

$.03125 = [.447/(2.33+z_{1-ß})] \Leftrightarrow z_{1-ß} = 11.97 \Leftrightarrow 1-ß = 1.00$

Bear in mind that this is the power of a test on the correlation against 0, which is not very informative. No wonder that in this case the power of the test is perfect. More informative would be a power calculation in case we were 0.10 off from the 'true' value 0.547. In that case the power would be:

$.03125 = [.10/(2.33+z_{1-ß})] \Leftrightarrow z_{1-ß} = 0.87 \Leftrightarrow 1-ß = .81.$

5.3. Main effects of a macro level variable

Let us now turn to an example in which a macro level variable is supposed to have an effect on a micro level variable. A case might be an experiment with class size, in which a class size reduction of 6 pupils per class is compared to a usual class size of 26 in its effect on achievement of young pupils.

$$\mathbf{ACHIEVEMENT}_{ij} = \gamma_{00} + \gamma_{01}\text{EXPERIMENT}_j + \boldsymbol{u_{0j}} + \boldsymbol{r_{ij}}\,.$$

Let us assume that achievement is measured in *z*-scores, and that the experiment is set up to detect an effect of 0.20 (or larger) in the population of interest. Transforming this effect size into a correlation coefficient *r* using (Rosenthal, 1994, p.239)

$$d = 2r / \sqrt{(1-r^2)}$$

results in $r = 0.10$. The astonishing thing about this example is, that most researchers would test all pupils within a class and would take the within class size of 23 (being the average of 20 and 26) as given. There is no need at all to do so! A straigthforward use of the formulae immediately points us to the clue that if $c = 0$, it is optimal (from a statistical point of view) to take as many classes as possible with only one pupil per class. It is only because of budgetary constraints that a two-stage sample may be preferred. Let us take *c* to be 23 (taking an extra school is 23 times the price of taking one extra student within an already sampled school).

The following might be a specification of the PINT parameter file on AX.DAT:

```
    0     0     1
    1     1    23
 1000    23
     0.80
     0.19
     1.00
     0.00
```

The file AX.OUT will contain the following information on the standard errors:

Sample sizes			costs	Standard errors	
N*n	N	n	(n+c)*N	Const	Group
41	41	1	984	0.15539	0.15539
80	40	2	1000	0.12145	0.12145
114	38	3	988	0.10962	0.10962
148	37	4	999	0.10267	0.10267
175	35	5	980	0.10000	0.10000
204	34	6	986	0.09752	0.09752
231	33	7	990	0.09602	0.09602
256	32	8	992	0.09520	0.09520
279	31	9	992	0.09485	0.09485
300	30	10	990	0.09487	0.09487
319	29	11	986	0.09518	0.09518
336	28	12	980	0.09574	0.09574
351	27	13	972	0.09652	0.09652
378	27	14	999	0.09567	0.09567
390	26	15	988	0.09674	0.09674
400	25	16	975	0.09798	0.09798
425	25	17	1000	0.09738	0.09738
432	24	18	984	0.09884	0.09884
437	23	19	966	0.10046	0.10046
460	23	20	989	0.10000	0.10000
462	22	21	968	0.10182	0.10182
484	22	22	990	0.10144	0.10144
483	21	23	966	0.10346	0.10346

The optimum for the standard error associated with the effect of the treatment appears to be when we take a sample of 9 pupils per school and 31 schools. In that case the standard error of interest reaches its minimum of .09485. It will be clear from the outset that this standard error is too big to reach a satisfactory level of power. Even with α as high as .10, the power will be .42 only. The first and most easy solution to this problem would be to increase the budget. Straigthforward application of (1) shows that to reach a power of .80 with α=.10, the standard error should be half of the size that was found. This implies a multiplication of the number of schools by a factor 2^2=4, which amounts to 4*31=124 schools with 9 pupils in each of them. To make it even more worse: politicians do not like to spend money on interventions that may not work, and for that reason α may be put to be as low as .01. Applying the formulae will show that the original standard error of .09485 should be brought down to one third of its value, implying a sample size of schools of 279 (=3^2*31) schools. Raudenbush (1997) shows how much the standard error for the experimental effect can be decreased (and thus the power increased) if a covariate is added that is strongly related to the outcome variable. In this case this would imply, in practice, to administer a pretest or intelligence test.

5.4. A cross-level interaction effect

We have already presented an example on a cross-level effect when illustrating the general principles of PINT, and when discussing the input and output of the program. For didactical purposes we will add one more example, which illustrates power calculations in case of studying determinants of growth. Let us think of a (very simple) study on cognitive growth of young children. The treatment to be evaluated is a pre-school program (delivered at the child's home) to stimulate early literacy, and as a control group there are children not in this program. The levels are defined as follows: time is level one and child is level two. The growth curve is supposed to be represented well by a simple linear term. The general problem now is, should we sample many children in the control and experimental condition, or should we increase the number of timepoints on which we assess each child's literacy status? Since this is an intervention study, the costs of sampling an extra child (notice that in this case the child is the macro unit) is very expensive, since each child in the experimental condition should be given the treatment. Suppose now that treating one child costs $200 per two hour session, and that the program lasts for half a year, with two sessions per week. The total costs per child then are equal to ½ × 52 × 2 × 200 = $10,400. We assume that it is desired that the treatment and the control groups are equally large. Since a child in the control group does not receive any (paid) treatment, the costs per pair of children (one in the experimental and one in the control condition) are $10,400 + $0 = $10,400, which amounts to $5,200 as the average costs for a child in this study. Now let us suppose furthermore that administering one test per child at a given timepoint costs $200. Initially the researchers planned to take the tests at 14 timepoints (one per two week period). The total costs, as the study was designed and budgeted to be conducted with 28 children (14 control, 14 experimental), are 28 × 14 × 200 + 28 × 5,200 = $156,000. Or in our usual expression: 28 (14 + 26) = 1280 units of $200 per unit. Given this situation, and having been granted the budget requested to conduct the study, a reconsideration takes place. Is 28 enough, shouldn't the number of children in this study be increased at the expense of taking less measurements per child? Before running PINT plausible parameter values have to be generated. The model to be tested looks like:

$$\mathbf{TESTSCORE}_{ti} = \gamma_{0i} + \gamma_{1i}\text{TIME}_{ti} + \boldsymbol{r_{ti}}$$

$$\gamma_{0i} = \gamma_{00} + \gamma_{01}\,\text{PROGRAM}_i + \boldsymbol{u_{0i}}$$

$$\gamma_{1i} = \gamma_{10} + \gamma_{11}\,\text{PROGRAM}_i + \boldsymbol{u_{1i}}.$$

Bearing in mind what has been said on the decomposition of the variance in the dependent variable (see Section 2.3), we might assume –under the condition that both TIME and PROGRAM are standard normal variables, uncorrelated which each other- that

$$\text{var}(\mathbf{TESTSCORE}_{ti}) = \gamma_{01}^2 + \gamma_{10}^2 + \gamma_{11}^2 + \tau_{00} + \tau_{11} + \sigma^2,$$

with $\tau_{00} = \text{var}(\boldsymbol{u_{0i}})$, $\tau_{11} = \text{var}(\boldsymbol{u_{1i}})$, and $\sigma^2 = \text{var}(\boldsymbol{r_{ti}})$.
Without the PROGRAM variable these variances might be 0.25, 0.25, and 0.25 respectively (with TIME already accounting for half of the variation at level 1, the level of the measurements). Including PROGRAM as a predictor in the model τ_{00} and τ_{11} might reduce to half of their initial value (0.125), implying that both γ_{01} and γ_{11} are $\sqrt{(0.125)} = 0.35$. Now

suppose that the program is considered to be sufficiently effective only if the difference in average time slopes is at least 0.30, corresponding to γ_{11} = 0.15. The difference between the hypothesized and tested parameter values then is 0.35 – 0.15 = 0.20.
Let us assume a correlation between the growth line and the literacy status half way the experiment of 0.72, so that the covariance τ_{01} = 0.72 / ($\sqrt{0.125} \times \sqrt{0.125}$) = 0.09.

The parameter file AX.DAT for PINT might look like (we cannot estimate a linear equation with less than three points!):

```
   1     0     1
   3     1    30
1280     26
    0.25
    0.125
    0.09 0.125
    1
    1
    0
```

The output file AX.OUT contains the following information on the standard errors:

```
Sample sizes      costs   Standard errors

N*n     N     n (n+c)*N    Const   Group   Random Cross-L
132    44     3    1276   0.06881 0.06881 0.06881 0.06881
168    42     4    1260   0.06682 0.06682 0.06682 0.06682
205    41     5    1271   0.06533 0.06533 0.06533 0.06533
240    40     6    1280   0.06455 0.06455 0.06455 0.06455
266    38     7    1254   0.06503 0.06503 0.06503 0.06503
296    37     8    1258   0.06498 0.06498 0.06498 0.06498
324    36     9    1260   0.06514 0.06514 0.06514 0.06514
350    35    10    1260   0.06547 0.06547 0.06547 0.06547
374    34    11    1258   0.06592 0.06592 0.06592 0.06592
396    33    12    1254   0.06648 0.06648 0.06648 0.06648
416    32    13    1248   0.06714 0.06714 0.06714 0.06714
448    32    14    1280   0.06682 0.06682 0.06682 0.06682
465    31    15    1271   0.06760 0.06760 0.06760 0.06760
480    30    16    1260   0.06847 0.06847 0.06847 0.06847
493    29    17    1247   0.06941 0.06941 0.06941 0.06941
522    29    18    1276   0.06920 0.06920 0.06920 0.06920
532    28    19    1260   0.07024 0.07024 0.07024 0.07024
540    27    20    1242   0.07136 0.07136 0.07136 0.07136
567    27    21    1269   0.07121 0.07121 0.07121 0.07121
572    26    22    1248   0.07242 0.07242 0.07242 0.07242
598    26    23    1274   0.07229 0.07229 0.07229 0.07229
600    25    24    1250   0.07360 0.07360 0.07360 0.07360
625    25    25    1275   0.07348 0.07348 0.07348 0.07348
624    24    26    1248   0.07489 0.07489 0.07489 0.07489
648    24    27    1272   0.07479 0.07479 0.07479 0.07479
644    23    28    1242   0.07631 0.07631 0.07631 0.07631
667    23    29    1265   0.07622 0.07622 0.07622 0.07622
660    22    30    1232   0.07785 0.07785 0.07785 0.07785
```

Notice that as a consequence of our standardization procedure the standard errors accompanying the various effects are, for a given sample size, exactly the same. As it turns out,

the smallest standard errors occur for a design with 40 children and 6 timepoints. If we take α at 0.01 in a one-sided test, then the power for this design can be derived from

$$0.20 / 0.06455 \approx 2.33 + z_{\gamma} \quad \Leftrightarrow \quad 0.77 = z_{\gamma} \quad \Leftrightarrow \quad \gamma = .78$$

If this power is considered to be too low for such an important study, then the usual potential manipulations are an increase of α or an increase of the budget to be able to increase the sample size.

Note that the example is rather artificial, since in a normal study on growth one would employ a more sophisticated model. A polynomial of a certain degree to model the curve adequately, would imply a number of timepoints that is at least one more than the degree of the polynomial. Or if one uses a spline function with a certain number of knots and complex curves, more timepoints may be needed.

6. Some general guidelines

6.1. Principles of optimal design

If you have run PINT more than once, the discovery will be made that in most cases the number of macro-units is what really counts, unless there are budgetary constraints. If there are budgetary constraints the balance between the sample size of the macro units, and that of the micro units is not clear in advance, and PINT will hopefully be of help there. Generally for each study the researcher should of course carefully address questions like the reliability of the measures to be taken, and the general principles of good design. One of these we would like to stress, and that is the minimisation of within group variability by including one or more proper covariates and/or a pretest (see Raudenbush, 1997 for a more detailed treatment of this subject).

6.2. Loose ends

Some of the recommendations that you will derive from PINT may be counterintuitive. Examples of such counterintuitive results are that it may be optimal to take as many macro units (e.g. schools) as possible, at the expense of the number of micro units per macro unit (e.g. pupils). Keep in mind that PINT only addresses the question of sample size from the point of view of estimating regression coefficients, i.e. the strength of the fixed effects! If you have in mind to estimate scale values or other relevant characteristics for each macro unit (like: mean achievement level of the students to inform parents on the quality of a school, or a good estimate of each child's personal growth curve) PINT is of no help. We refer to Snijders & Bosker (1999), notably Section 10.3, for guiding principles for these kind of situations.
PINT is of no help either in case you want to set up a study for exploring the variance structure i.e. estimating the intraclass correlation ρ. In that case the standard error of ρ is what we are interested in. This situation is dealt with in Section 10.5 of Snijders & Bosker (1999). Finally you may be specifically interested in estimating the variance components. You will find some guidelines and further references in the same section of the same textbook.

6.3. PINT as shareware

Feel free to use PINT, but please be so kind to inform us if you encounter problems and refer to this manual and to Snijders & Bosker (1993) when publishing results based on power calculations with PINT. In the latter case we would appreciate it very much if you send us a copy of the paper or the article, or at least provide us with a reference.

6.4 Differences between versions 1.61 and 2.1

The only difference between versions 1.61 (April 1999) and 2.1 (April 2003) is that the latter version is a full Windows version (programmed in Delphi 5), thereby more user-friendly, and that the option has been added of giving results for all combinations of n and N between certain bounds. Parameter input files of version 1.61 (called "input files" in the manual of 1.6) can also be used for version 2.1.

References

Afshartous, D. (1995). *Determination of sample size for multilevel model design.* Perspectives on Statistics for Educational Research: Proceedings of the National Institute for Statistical Sciences (NISS). Paper presented at the AERA meeting in San Francisco, April 1995.

Bryk, A.S. & S.W. Raudenbush (1992). *Hierarchical Linear Models. Applications and Data Analysis Methods.* Newbury Park / London / New Delhi: Sage publications.

Cochran, W.G. (1977). *Sampling Techniques* (3rd ed.). New York: Wiley.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). "A power primer". *Psychological Bulletin, 112,* 155-159.

Goldstein, H. (1995). *Multilevel Statistical Models* (2nd ed.). London: Edward Arnold.

Hox, J.J. (2002). *Multilevel Analysis, Techniques and Applications*. Lawrence Erlbaum Associates, 2002.

Mok, M. (1995). "Sample size requirements for 2-level designs in educational research", *Multilevel Modelling Newsletter, 7(2),* 11-15.

Raudenbush, S.W. (1997). "Statistical analysis and optimal design for cluster randomized trials". *Psychological Methods, 2(2),* 173-185.

Rosenthal, R. (1994). "Parametric measures of effect size". In: H. Cooper & L.V. Hedges (Eds.), *Handbook of Research Synthesis.* New York: Russell Sage Foundation. (pp. 231-244)

Snijders, T.A.B. (2001). "Sampling". Chapter 11 (pp. 159-174) in A. Leyland and H. Goldstein (eds.), *Multilevel Modelling of Health Statistics.* Chichester etc.: Wiley.

Snijders, T.A.B. & R.J. Bosker (1993). "Standard errors and sample sizes for two-level research". *Journal of Educational Statistics, 18,* 237-260.

Snijders, T.A.B. & R.J. Bosker (1999). *Multilevel analysis. An introduction to basic and advanced multilevel modeling.* Newbury Park/London/New Delhi: Sage Publications.