Measurement and Research Department Reports 91-8

The Optimum Allocation Of Measurements In Balanced Random Effects Models

P.F. Sanders
T.J.H.M. Eggen

Cito

The Optimum Allocation Of Measurements In Balanced Random Effects Models

P.F. Sanders
T.J.H.M. Eggen

Cito
Arnhem, 1991

Abstract

The problem of maximizing a ratio of variance components given a budget constraint is considered. The proposed solution employs analytical and numerical methods and is an improvement of two existing procedures. The current procedure is exemplified for three- and four-way designs.

Keywords: generalizability theory; ratios of variance components; balanced designs; Lagrange method; branch-and-bound algorithm.

Introduction

Consider the balanced three-way random effects model with one observation per cell

$$Y_{ijk} = \mu + A_i + B_j + C_k + AB_{ij} + AC_{ik} + BC_{jk} + E_{ijk} \,,$$
$$i = 1, \ldots, n_i,\; j = 1, \ldots, n_j,\; k = 1, \ldots, n_k,$$

where μ is a constant and the $\{A_i\}$, $\{B_j\}$, $\{C_k\}$, $\{AB_{ij}\}$, $\{AC_{ik}\}$, $\{BC_{jk}\}$, and $\{E_{ijk}\}$ are independent normals with zero means and variances σ_A^2, σ_B^2, σ_C^2, σ_{AB}^2, σ_{AC}^2, σ_{BC}^2, σ_e^2, respectively.

Parameters expressing the proportion of total variability are used in all fields of applications in which the variance component model is employed. In genetics, ratios of variance components are used as estimates of genetic heritability (e.g., Graybill & Wang 1979). In psychometrics, the variance components ratio

$$\rho_A = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_{AB}^2/n_j + \sigma_{AC}^2/n_k + \sigma_e^2/n_j n_k} \tag{1}$$

is an important parameter (e.g., Cronbach, Gleser, Nanda & Rajaratnam 1972). This parameter is an indicator of the degree of precision provided by a measurement instrument in which factor A constitutes the objects of measurement, that is, examinees, whereas factor B and C each constitute sets of conditions of measurement, for example, n_j questions and n_k raters. The number of measurements per examinee equals $n_j n_k$.

In (1), σ_{AB}^2, σ_{AC}^2 and σ_e^2 constitute the measurement error. From (1), it can be inferred that measurement precision can be improved by increasing n_j and n_k. However, because in practice the administration of measurements has to be conducted under the constraint of limited resources, the optimal allocation of n_j and n_k is of critical importance. A procedure for maximizing ρ_A under the constraint of a fixed number of measurements was first proposed by Woodward and Joe (1973). With this procedure, however, non-optimal solutions are likely to occur because of the equality constraint on

$n_j n_k$. Moreover, the generalization of Woodward and Joe's procedure as well as the procedure developed by Sanders, Theunissen, and Baas (1991), which does provide optimal solutions under a budget constraint, to four-way and other designs becomes very complex.

Addressing the same problem of maximizing ρ_A under a budget constraint, an alternative procedure is presented in this paper. In section 4.2, the maximization problem of the three-way random effects model is formulated as an optimization problem. In section 4.3, a two-step procedure for solving this problem is proposed and illustrated. The solution for the maximization problem of the four-way random effects model is presented in section 4.4.

Optimization Problem

Maximizing (1) is equivalent to minimizing the objective-function

$$f(n_j, n_k) = \frac{\sigma^2_{AB}}{n_j} + \frac{\sigma^2_{AC}}{n_k} + \frac{\sigma^2_e}{n_j n_k} \, . \tag{2}$$

Minimization statement (2) refers to the value of the objective- function with different values for n_j and n_k.

A complete problem specification includes two other constraints. First, the constraint that specifies the available resources, that is, the budget. As costs in general vary in proportion to $n_j n_k$, the cost function is defined as

$$c_{jk} \, n_j \, n_k \leq C,$$

where c_{jk} denotes the cost of one measurement, for example, the rating by one rater of the answers of all examinees in the sample to one question, and C the budget. Assuming that the cost of one measurement is the same for each examinee, the cost function can be written as

$$n_j \, n_k \leq \frac{C}{c_{jk}} = M, \tag{3}$$

n_k be integer values and that each factor has at least one condition of measurement, a lower bound integer constraint is specified

$$n_j \text{ and } n_k \text{ integer} \geq 1 \tag{4}$$

Proposed Solution Three-Way Design

Optimal integer solutions for n_j and n_k of the optimization problem defined by (2), (3) and (4) are obtained in two steps. In the first, solutions are derived for a continuous relaxation of constraint (4), that is, $n_j,\ n_k > 0$. These optimal continuous solutions are used in the second step as the bounds in a branch-and-bound procedure (see Papadimitriou & Steiglitz 1982, p. 443).

4.3.1. Continuous Solution

The continuous problem can be solved by using Lagrange multipliers method. The Lagrange function is given by

$$L\ (n_j, n_k, \lambda) = \frac{\sigma^2_{AB}}{n_j} + \frac{\sigma^2_{AC}}{n_k} + \frac{\sigma^2_e}{n_j n_k} + \lambda\,(n_j n_k - M)\,. \tag{5}$$

The necessary conditions for the minimum of $f(n_j, n_k)$ are

$$\frac{\delta L}{\delta n_j} = -\frac{\sigma^2_{AB}}{n_j^2} - \frac{\sigma^2_e}{n_j^2 n_k} + \lambda n_k = 0, \tag{6}$$

$$\frac{\delta L}{\delta n_k} = -\frac{\sigma^2_{AC}}{n_k^2} - \frac{\sigma^2_e}{n_j n_k^2} + \lambda n_j = 0, \tag{7}$$

$$\frac{\delta L}{\delta \lambda} = n_j n_k - M = 0. \tag{8}$$

From (6), (7) and (8) the optimal continuous solutions are derived as

$$n_j^* = \left(M\,\frac{\sigma^2_{AB}}{\sigma^2_{AC}} \right)^{1/2}, \text{ and} \tag{9}$$

$$n_k^* = \left(M \frac{\sigma_{AC}^2}{\sigma_{AB}^2} \right)^{1/2} . \tag{10}$$

Note that because of constraint (3), the minimization of (2) without the error term will give the same solutions.

4.3.2. Integer Solution

The optimal continuous solutions, derived in the preceding section, are used as a starting point for a branch-and-bound algorithm to obtain the optimal integer solutions. Assume that n_j^* and n_k^* are the optimal continuous solutions, and both values are not integer. If $\lfloor n_j^* \rfloor$ denotes n_j^* rounded down and $\lceil n_j^* \rceil$ denotes n_j^* rounded up, two problems have to be solved. The first problem is

$$\text{minimize} \quad f(n_j, n_k) = \frac{\sigma_{AB}^2}{n_j} + \frac{\sigma_{AC}^2}{n_k} + \frac{\sigma_e^2}{n_j n_k} ,$$

subject to $n_j\, n_k \le M$,

$n_j,\ n_k \ge 1,$ and

$n_j \le \lfloor n_j^* \rfloor$.

In a solution $(\hat{n}_j,\ \hat{n}_k)$ for this problem, $\hat{n}_j = \lfloor n_j^* \rfloor$ must hold, because if $\hat{n}_j \ne \lfloor n_j^* \rfloor$, the rounding down constraint would be irrelevant and a non-integer solution would result. Hence, the solution for the first problem is $\hat{n}_j = \lfloor n_j^* \rfloor$ and $\hat{n}_k = M/\hat{n}_j$, or $\hat{n}_j = 1$ and $\hat{n}_k = M$. The second problem differs from the first in that the rounding down constraint is replaced by the rounding up constraint $n_j \ge \lceil n_j^* \rceil$. The solution for the second problem is $\hat{n}_j = \lceil n_j^* \rceil$ and $\hat{n}_k = M/\hat{n}_j$. Each of these two problems yield two new problems with additional constraints regarding n_k whenever the value $\hat{n}_k$ is non-integer. If, for instance, $\hat{n}_k$ is non-integer in the first problem, the constraint sets for the following two problems are given by (1) $n_k \ge 1$, $n_j \ge \lceil \hat{n}_j \rceil$, and (2) $n_k \ge 1$, $1 \le n_j \le \lfloor \hat{n}_j \rfloor$. Branching continues until either an integer solution or an infeasible constraint set is found. Each time an integer solution is obtained during the branching process, its value is compared with that of the previous best solution and accepted as the new best solution or rejected as such. At the end of the process, the current best solution is the optimal integer solution.

4.3.3. Example

For the balanced three-way random effects model crossed design the example from Sanders et al. (1991) with variance components $\hat{\sigma}^2_A = 5.435$, $\hat{\sigma}^2_{AB} = 3.421$, $\hat{\sigma}^2_{AC} = 1.140$, and $\hat{\sigma}^2_e = 11.850$ is used. The objective- function for this example is

$$\text{minimize} \quad \frac{3.421}{n_j} + \frac{1.140}{n_k} + \frac{11.850}{n_j n_k} .$$

Assuming that conditions of factor B, for example, questions, and factor C, for example, raters, cost nothing, but that the rating by one rater of the answers of all examinees in the sample to one question costs 80 dollars, and that the budget is limited to 3000 dollars, the budget constraint for this example is stated as

$$80 n_j n_k \leq 3000. \tag{11}$$

The optimal integer solutions $\hat{n}_j$ and $\hat{n}_k$ for this optimization problem are derived in two steps. First, using equations (9) and (10), the optimal continuous solutions $n_j^* = 10.6$ and $n_k^* = 3.5$ are obtained. Second, because both solutions are non-integer, a branch-and-bound algorithm is employed to obtain the optimal integer solutions. For the problem with additional constraint $n_j \leq \lfloor n_j^* \rfloor = 10$, one finds a solution $\hat{n}_j = 10$, $\hat{n}_k = 3.7$ and for the problem with $n_j \geq \lceil n_j^* \rceil = 11$, a solution $\hat{n}_j = 11$, $\hat{n}_k = 3.4$. Further branching yields four problems with constraint sets, including n_j, $n_k \geq 1$, (1) $n_j \leq 10$, $n_k \leq 3$; (2) $n_j \leq 10$, $n_k \geq 4$; (3) $n_j \geq 11$, $n_k \leq 3$, and (4) $n_j \geq 11$, $n_k \geq 4$.

The search process is shown in Figure 1.

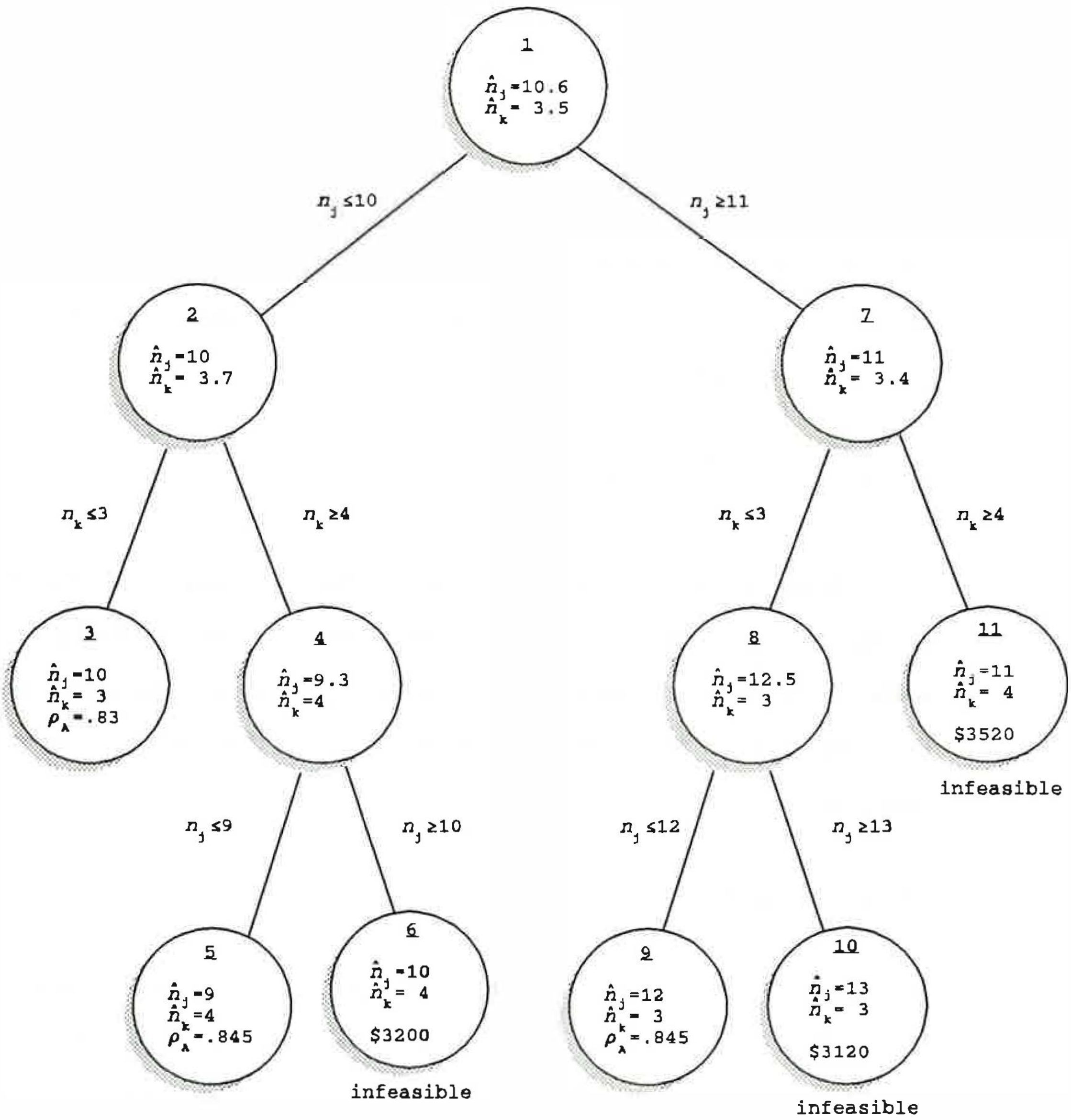

FIGURE 1

Search-Tree for the Three-Way Design Example

Starting from the optimal continuous solution in Node 1, the strategy to traverse the search-tree is depth-first and from left to right, as indicated by the numbering of the nodes. Node 3 produces the first candidate solution, which is improved by the solution found in Node 5. The solution in Node 6 does not satisfy the budget constraint of 3000 dollars and is therefore an infeasible solution. Other infeasible solutions are found in Node 10 and 11. The solution found in Node 9 is equal to the solution found in Node 5. From the search process in Figure 1 and the solutions for this cost function presented in Table 1, it can be concluded that a more exhaustive search will result in either infeasible solutions or solutions that do not improve the solutions in Node 5 and 9. Therefore, the solution in Node 5, $(\hat{n}_j, \hat{n}_k) = (9,4)$, that is, nine questions for which the answers of all examinees have to be rated by four raters, or the solution in Node 9, $(\hat{n}_j, \hat{n}_k) = (12,3)$, that is twelve questions rated by three raters, are the optimal integer solutions for this problem.

TABLE 1

Values of n_j, n_k, Variance Components, ρ_A and C for Two Cost Functions

n_j	$c_j n_j$	n_k	$n_j n_k$	$c_{jk} n_j n_k$	$\hat{\sigma}^2_A$	$\hat{\sigma}^2_{AB}/n_j$	$\hat{\sigma}^2_{AC}/n_k$	$\hat{\sigma}^2_e/n_j n_k$	ρ_A	C
1. c_{jk} = 80 Dollars, C = 3000 Dollars										
10.6	0	3.5	37.5	3000	5.435	.322	.322	.316	.850	3000
9	0	4	36	2880	5.435	.380	.285	.329	.845	2880
11	0	4	44	3520	5.435	.311	.285	.269	.863	3520
10	0	3	30	2400	5.435	.342	.380	.395	.830	2400
11	0	3	33	2640	5.435	.311	.380	.359	.838	2640
10	0	4	40	3200	5.435	.342	.285	.296	.855	3200
12	0	3	36	2880	5.435	.285	.380	.329	.845	2880
13	0	3	39	3120	5.435	.263	.380	.304	.852	3120
2. c_j = 40 Dollars, c_{jk} = 80 Dollars, and C = 3000 Dollars										
8	320	4	32	2560	5.435	.428	.285	.370	.834	2880
9	360	4	36	2880	5.435	.380	.285	.329	.845	3240
10	400	3	30	2400	5.435	.342	.380	.395	.830	2800
11	440	3	33	2640	5.435	.311	.380	.359	.838	3080

Although the budget constraint employed here was not taken from an actual measurement problem, the example is realistic since in most cases, differences between the number of measurements will have more impact on the budget than on ρ_A. This means that in many cases nearly the same measurement precision can be obtained for a considerably smaller budget. Table 1 also contains the solutions for a second cost function

$$40n_j + 80n_jn_k \leq 30000, \tag{12}$$

where ($c_j =$) 40 dollars denotes the cost of one question. For all linear cost functions with the n_jn_k measurements taking up the major part of the budget, the optimal continuous solution for a problem with cost function (11) can be used as a starting point of a branch-and-bound algorithm for a problem with cost function (12). The optimal continuous solution $\hat{n}_j$ for a problem with cost function (11) is the upper bound of $\hat{n}_j$ for a problem with cost function (12). It can be inferred from the results presented in Table 1, that a simple branch-and-bound algorithm will produce the optimal integer solution for problem (12), that is, eight questions rated by four raters.

Proposed Solution Four-Way Design

Consider the balanced four-way random effects model with one observation per cell

$$Y_{ijkm} = \mu + A_i + B_j + C_k + D_m + AB_{ij} + AC_{ik} + AD_{im} + BC_{jk} + BD_{jm} + CD_{km} + ABC_{ijk} + ABD_{ijm} + ACD_{ijm} + BCD_{jkm} + E_{ijkm},$$
$$i = 1, \ldots, n_i,\ j = 1, \ldots, n_j,\ k = 1, \ldots, n_k,\ m = 1, \ldots, n_m,$$

where μ is a constant and the $\{A_i\}$, $\{B_j\}$, $\{C_k\}$, $\{D_m\}$, $\{AB_{ij}\}$, $\{AC_{ik}\}$, $\{AD_{im}\}$, $\{BC_{jk}\}$, $\{BD_{jm}\}$, $\{CD_{km}\}$, $\{ABC_{ijk}\}$, $\{ABD_{ijm}\}$, $\{ACD_{ijm}\}$, $\{BCD_{jkm}\}$, and $\{E_{ijkm}\}$ are independent normals with zero means and variances σ_A^2, σ_B^2, σ_C^2, σ_D^2, σ_{AB}^2, σ_{AC}^2, σ_{AD}^2, σ_{BC}^2, σ_{BD}^2, σ_{CD}^2, σ_{ABC}^2, σ_{ABD}^2, σ_{ACD}^2, σ_{BCD}^2, and σ_e^2, respectively. The maximization of the generalizability coefficient for the four-way design, amounts to minimizing

$$f(n_j n_k n_m) = \frac{\sigma^2_{AB}}{n_j} + \frac{\sigma^2_{AC}}{n_k} + \frac{\sigma^2_{AD}}{n_m} + \frac{\sigma^2_{ABC}}{n_j n_k} + \frac{\sigma^2_{ABD}}{n_j n_m} + \frac{\sigma^2_{ACD}}{n_k n_m} + \frac{\sigma^2_e}{n_j n_k n_m},$$

subject to $C_{jkm}\, n_j n_k n_m \le C \rightarrow n_j n_k n_m \le \frac{C}{C_{jkm}} = M.$

The application of the Lagrange multipliers method to this minimization problem results in the equations

$$n_m = n_k \left(\frac{\sigma^2_{AD} + \frac{\sigma^2_{ABD}}{n_j}}{\sigma^2_{AC} + \frac{\sigma^2_{ABC}}{n_j}} \right), \quad n_m = n_j \left(\frac{\sigma^2_{AD} + \frac{\sigma^2_{ACD}}{n_k}}{\sigma^2_{AB} + \frac{\sigma^2_{ABC}}{n_k}} \right), \text{ and}$$

$$n_k = n_j \left(\frac{\sigma^2_{AC} + \frac{\sigma^2_{ACD}}{n_m}}{\sigma^2_{AB} + \frac{\sigma^2_{ABD}}{n_m}} \right).$$

As second-order interaction variance components are generally very small, the above equations can be written as

$$\frac{n_m}{n_k} = \frac{\sigma^2_{AD}}{\sigma^2_{AC}}, \quad \frac{n_m}{n_j} = \frac{\sigma^2_{AD}}{\sigma^2_{AB}}, \text{ and } \frac{n_k}{n_j} = \frac{\sigma^2_{AC}}{\sigma^2_{AB}}.$$

From these equations and the cost function, nearly optimal continuous solutions for n_j, n_k and n_m can be derived. These solutions are used as the starting point for a branch-and-bound algorithm analogous to that of the three-way design.

Conclusions and Discussion

The results from this and other studies (Sanders, Theunissen & Baas 1989, 1991) show how integer optimization methods can be applied to find the optimal design for the maximization of a ratio of variance components. In general, these methods are ultimately suitable for solving problems for which analytical solutions are hard to obtain. One of these problems - already discussed by Scheffé in 1959 (pp. 236-238) - is to find the optimal design for the estimation of (ratios of) variance components. It was only recently, however, that Mukerjee and Huda (1988, pp. 78-79) suggested the use of enumeration methods for solving this problem under restrictions on total cost and other practical constraints. In this article a procedure is presented which can be used to solve some specific design problems.

References

Cronbach, L.J., Gleser, G.C., Nanda, H. & Rajaratnam, N. (1972). *The dependability of behavioral measurements*. New York: Wiley.

Graybill, F.A. & Wang, C.M. (1979). Confidence intervals for proportions of variability in two-factor nested variance component models. Journal of the American Statistical Association 74, 368-374.

Mukerjee, R. & Huda, S. (1988). Optimal design for the estimation of variance components. *Biometrika 75*, 75-80.

Papadimitriou, Ch.H. & Steiglitz, K. (1982). *Combinatorial optimization: Algorithms and complexity*. Englewood Cliffs: Prentice-Hall.

Sanders, P.F., Theunissen, T.J.J.M. & Baas, S. (1989). Minimizing the number of observations: A generalization of the Spearman-Brown formula. *Psychometrika 54*, 587-598.

Sanders, P.F., Theunissen, T.J.J.M. & Baas, S. (1991). Maximizing the coefficient of generalizability under the constraint of limited resources. *Psychometrika 56*, 87-96.

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

Woodward, J.A. & Joe, G.W. (1973). Maximizing the coefficient of generalizability in multi-facet decision studies. *Psychometrika 38*, 173-181.

Recent Measurement and Research Department Reports:

91-1 N.D. Verhelst & N.H. Veldhuijzen. A New Algorithm For Computing Elementary Symmetric Functions And Their First And Second Derivatives.

91-2 C.A.W. Glas. Testing Rasch Models For Polytomous Items: With An Example Concerning Detection Of Item Bias.

91-3 C.A.W. Glas & N.D. Verhelst. Using The Rasch Model For Dichotomous Data For Analyzing Polytomous Responses.

91-4 N.D. Verhelst & C.A.W. Glas. A Dynamic Generalization Of The Rasch Model.

91-5 N.D. Verhelst & H.H.F.M. Verstralen. The Partial Credit Model With Non-Sequential Solution Strategies.

91-6 H.H.F.M. Verstralen & N.D. Verhelst. The Sample Strategy Of A Test Information Function In Computerized Test Design.

91-7 H.H.F.M. Verstralen & N.D. Verhelst. Decision Accuracy In IRT Models.