Adaptive Designs
IMS Lecture Notes - Monograph Series (1995) Volume 25

ADAPTATION IN THE DESIGN OF EXPERIMENTS FOR LINEAR MODELS

BY RAINER SCHWABE

Freie Universität Berlin

Abstract

We demonstrate that even in the classical experimental situation of a linear model the performance of the experiment can be improved by applying the concept of adaptive designs.

1. Introduction. In the classical theory of optimum designs in linear models planning and inference are regarded as two separate successive steps of an experiment. After a design has been fixed all observations will be collected and afterwards the information will be extracted from the outcomes of the experiments [see, for example, the monographs of Bandemer (1977), Fedorov (1972), Pázman (1986), and Silvey (1980)]. If in contrast the experiments are realized successively, the information obtained from the outcomes of the first experiments can be used to design the following ones. Such a scheme will be called an adaptive design.

The aim of this paper is to propose adaptive designs for linear problems in linear models where the corresponding experiments result in more information than could be obtained from the experiments according to any predetermined design.

So far adaptive designs have been considered in the literature for nonlinear situations because there the performance of the estimator and

Received October 1992; revised March 1993.
AMS 1991 subject classifications. Primary 62L05; secondary 62K05.
Key words and phrases. Adaptive design, two-stage procedures.

hence the quality of a design depend on the unknown parameters [for a recent survey we refer to Ford, Titterington, and Kitsos (1989)].

We will demonstrate, however, that even in linear situations and for a fixed number of experiments the performance can be improved by applying the concept of adaptive designs. In the last section we deal with a different problem where the adaptation is necessary because of a lack of knowledge about the underlying variance-covariance structure of the observations. This situation includes for example the well-known Behrens-Fisher problem. Some attempts have been made to attack these problems sequentially and asymptotic results have been obtained [see, for example, Robbins, Simons, and Starr (1967)]. Here we present a two-stage procedure based on an idea of Stein (1945); in the first stage the variances are estimated, and in the second stage, allocation is arranged according to these estimates.

2. Classical design of experiments in linear models. We will start with the formal description of a linear model:

(2.1)
$$X(t) = \sum_{i=1}^{K} a_i(t)\beta_i + Z = a(t)'\beta + Z,$$

where $t \in \mathcal{T}$ is the value of the controlled factors in the design region $\mathcal{T}$, $a = (a_1, \dots, a_K)' : \mathcal{T} \to \mathbf{R}^K$ is the vector of known regression functions, $\beta = (\beta_1, \dots, \beta_K)' \in \mathrm{IR}^K$ is a vector of unknown parameters, $X(t)$ is the observation, and $Z = Z(t)$ is some random noise.

A (deterministic) design d of size N is given by

(2.2)
$$d = (t^{(1)}, \dots, t^{(N)}), \qquad t^{(N)} \in \mathcal{T}.$$

According to (2.1) the nth observation in a designed experiment is

$$X_n(t^{(n)}) = a(t^{(n)})'\beta + Z_n.$$

If we denote by $X_d = (X_1(t^{(1)}), \dots, X_N(t^{(N)}))'$ and $Z_d = (Z_1, \dots, Z_N)'$ the vectors of observations and random noise, respectively, the whole experimental situation can be written in matrix notation:

(2.3)
$$X_d = A_d\beta + Z_d,$$

where $A_d = (a(t^{(1)}), \ldots, a(t^{(N)}))'$ is the design matrix.

EXAMPLE 2.1. *(One-way layout.)* For K different groups, the mean effects $\beta_1, \ldots, \beta_K$ are of interest. The observed responses can be modeled by $X(t) = \beta_t + Z$, $t \in \mathcal{T} = \{1, \ldots, K\}$. This formula fits in the general model (2.1) with indicators $a_i(t) = \mathbf{1}_{\{i\}}(t)$ (equal to 1 if $t = i$ and zero otherwise) as regression functions.

EXAMPLE 2.2. *(Linear regression.)* The expected response is linear in the real explanatory variable $X(t) = \beta_1 + \beta_2 t + Z$, $t \in \mathcal{T} \subseteq \mathrm{R}$ and the regression function a is given by $a(t) = (1, t)'$.

In what follows, we will assume that the random errors $Z_1, \ldots, Z_N$ are independent with zero mean and common finite variance $\sigma^2 > 0$.

If the design matrix A_d has full column rank, then $A_d'A_d$ is non-singular and β can be estimated by the least squares estimator

(2.4)
$$\widehat{\beta}_d = (A_d'A_d)^{-1}A_d'X_d$$

which is known to be the best linear unbiased estimator for β. The mean squared error matrix

(2.5)
$$E((\widehat{\beta}_d - \beta)(\widehat{\beta}_d - \beta)') = \sigma^2(A_d'A_d)^{-1}$$

equals the covariance matrix of $\widehat{\beta}_d$ and is independent of β.

For comparing different designs d we will be interested in the mean squared Euclidean distance of their resulting least squares estimators $\widehat{\beta}_d$ from β:

(2.6)
$$E(\left\|\widehat{\beta}_d - \beta\right\|^2) = E((\widehat{\beta}_d - \beta)'(\widehat{\beta}_d - \beta)) = \text{trace } (E((\widehat{\beta}_d - \beta)(\widehat{\beta}_d - \beta)')),$$

or, alternatively, in the maximal mean squared deviation

$$\max_{i=1,\ldots,K} E((\widehat{\beta}_{d,i} - \beta_i)^2)$$

of the components $\widehat{\beta}_{d,i}$ of $\widehat{\beta}_d$ from the corresponding components β_i of β. In view of (2.5),

$$E(\|\widehat{\beta}_d - \beta\|) \text{ and } \max_{i=1,\ldots,K} E((\widehat{\beta}_{d,i} - \beta_i)^2)$$

are constant in β for any (deterministic) design d.

3. Adaptive designs. If experiments are made successively we may allow the choice $T^{(n)}$ of the explanatory variables for the nth experiment to depend on the previous observations (and on the previous choices of the explanatory variables). Hence

(3.1)

$$T^{(n)} = f_n(T^{(1)}, \ldots, T^{(n-1)}, X_1(T^{(1)}), \ldots, X_{n-1}(T^{(n-1)}))$$

is a random variable which is completely determined by the history of the experimental situation up to time $(n-1)$. As in (2.2), we define an adaptive design D of size N by

(3.2)

$$D = (T^{(1)}, \ldots, T^{(N)})$$

where the $T^{(n)}$ satisfies (3.1).

We may replace d by D in the definitions of section 2 to obtain a notation that can handle both fixed and adaptive designs. In particular, the whole experimental situation can be expressed in matrix notation as

(3.3)

$$X_D = A_D\beta + Z_D.$$

But now the design matrix $A_D = (a(T^{(1)}), \ldots, a(T^{(N)}))'$ may be random and may depend on the observations. This dependence has crucial effects on the properties of the "least squares estimator"

(3.4)
$$\widehat{\beta}_D = (A_D' A_D)^{-1} A_D' X_D$$

which is defined in accordance with (2.4).

In general, $\widehat{\beta}_D$ is no longer unbiased nor is it linear in the observations. Moreover the mean squared error matrix $E((\widehat{\beta}_D - \beta)(\widehat{\beta}_D - \beta)')$ is dependent on β, which implies that

$$E(\|\widehat{\beta}_d - \beta^2\|)$$

is no longer constant in β. The results of the following sections will justify the use of $\widehat{\beta}_D$.

4. Local results. We investigate the situation where different deterministic designs of size N exist for which the minimum of

$$E(\|\widehat{\beta}_d - \beta^2\|)$$

is attained. This happens for example in the *one-way layout* if N is not a multiple of the number K of groups and in the *linear regression* if N is odd. We are now looking for a rule based on the outcomes of the experiments already made which helps us to decide which one of this equivalent designs to pick.

We begin with an initial deterministic design $d_{N-1} = (t^{(1)}, \ldots, t^{(N-1)})$ of size $(N-1)$ such that for any t in a subset $\mathcal{T}^* \subseteq \mathcal{T}$ the designs $d(t) = (t^{(1)}, \ldots, t^{(N-1)}, t)$ minimize

$$E(\|\widehat{\beta}_d - \beta^2\|)$$

within the class of all deterministic designs. The explanatory variable $T^{(N)}$ for the last experiment will be chosen adaptively out of $\mathcal{T}^*$. We thus consider an adaptive design

(4.1)
$$D(T^{(N)}) = d(T^{(N)}) = (t^{(1)}, \ldots, t^{(N-1)}, T^{(N)}).$$

By the Theorem in Schwabe (1990), we obtain for the present case:

THEOREM 4.1. *Let $d_{N-1} = (t^{(1)}, \ldots, t^{(N-1)})$ be a deterministic design such that $A_{d_{N-1}}' A_{d_{N-1}}$ is regular and for all $t \in \mathcal{T}^* \subseteq \mathcal{T}$ let*

A1) $a(t)'a(t) = c_1$,

A2) $(A'_{d_{N-1}} A_{d_{N-1}})^{-1} a(t) = c_2 a(t)$,

for some c_1 *and* c_2. *Then*

$$E(\left\|\widehat{\beta}_{D(T^{(N)})} - \beta\right\|^2)$$

is minimized within the class of all adaptive designs

$$D(T^{(n)}) = (t^{(1)}, ..., t^{(N-1)}, T^{(N)})$$

with $T^{(N)}$ *taking values in* $\mathcal{T}^*$ *if* $T^{(N)}$ *satisfies*

(4.2)

$$\left|a(T^{(N)})'(\widehat{\beta}_{d_{N-1}} - \beta)\right| = \max_{t \in \mathcal{T}^*} \left|a(t)'(\widehat{\beta}_{d_{N-1}} - \beta)\right|.$$

This rule can be simply stated as

Choose for $T^{(N)}$ that point in the design region (restricted to $\mathcal{T}^*$) for which the predicted value $a(t)'\widehat{\beta}_{d_{N-1}}$ of the response function differs most from the true value $a(t)'\beta$.

In particular, if the number N of observations is small, the gain in the efficiency for the whole experiment is quite large using the procedure (4.2). Explicit formulae for the efficiency in the one-way layout are given in Schwabe (1991).

5. Global results. The rule proposed in the previous section has the main disadvantage that it depends on the unknown (!) parameter β (similar to the locally optimum designs in nonlinear settings). We may get around this problem by replacing the unknown parameter β with an initial guess $\beta^{(0)}$ which has to be fixed before the experiments start. This means that the rule (4.2) is modified so that $T^{(N)}$ is selected according to

(5.1)

$$\left|a(T^{(N)})'(\widehat{\beta}_{d_{N-1}} - \beta^{(0)})\right| = \max_{t \in \mathcal{T}^*} \left|a(t)'(\widehat{\beta}_{d_{N-1}} - \beta^{(0)})\right|$$

which can be paraphrased as:

Choose for $T^{(N)}$ that point in the design region (restricted to $\mathcal{T}^*$) for which the predicted value $a(t)'\widehat{\beta}_{d_{N-1}}$ of the response function differs most from the initial guess $a(t)'\beta^{(0)}$.

If the random noise Z_n is symmetrically distributed it has been shown by Schwabe (1991) that for the examples under consideration the rule (5.1) results in an estimator $\widehat{\beta}_{D(T^{(N)})}$ which gives a smaller value of the criterion function

$$E(\|\widehat{\beta}_D - \beta\|^2)$$

than for an optimum deterministic design of size N uniformly in β.

EXAMPLE 5.1. *(One-way layout.)* If $N-1 = MK$ is a multiple of the number of groups K we are able to allocate an equal number M of experiments to each group in the initial design d_{N-1}. Then $\mathcal{T}^* = \mathcal{T}$ and according to (4.2) we have to perform the last experiment in that group for which the estimated mean $\widehat{\beta}_{d_{N-1},t}$ differs most from the initial guess $\beta_t^{(0)}$.

THEOREM 5.1 *If $T^{(N)}$ is chosen according to (5.1) in the setting of Example 5.1, and if the random errors are independent and identically distributed according to a symmetric distribution, then*

(5.2)
$$E(\|\widehat{\beta}_D - \beta\|^2) \leq \left(\frac{K-1}{M} + \frac{1}{M+1}\right)\sigma^2.$$

Additionally, strict inequality holds for Gaussian random noise. Similar results can be obtained for Example 2.2 [see Schwabe (1987)]. Unfortunately, these positive results are restricted to the situation in which only the last observation is chosen according to the adaptive rule (5.1). If more than one observation will be determined adaptively, the local behaviour can be improved for $\beta^{(0)}$ close to the true value β. However, globally the performance will become much worse than for a deterministic optimal design if

$$\|\beta^{(0)} - \beta\|$$

is large. For an example we refer to the simulation results of Gebhardt and Heckendorff (1983).

6. Testing. While the choice of the initial guess $\beta^{(0)}$ in estimation problems seems to be rather arbitrary there is a natural choice in test situations. Because the shrinkage effect of $\widehat{\beta}_{D(T^{(N)})}$ in the direction of

$\beta^{(0)}$ is most evident for $\beta^{(0)} = \beta$, the initial guess should be chosen from the null-hypothesis. In particular, if we want to test the hypothesis $\beta = \beta_0$ against the alternative $\beta \neq \beta_0$ then set $\beta^{(0)} = \beta_0$.

We illustrate this procedure using the example of the one-way layout with observations from two groups because it is the most simple, and also the most elucidating, non-trivial design problem [see, for instance, Kiefer (1958)]. To keep the situation as easy as possible we look at three observations. Without loss of generality, we may set $t^{(1)} = 1$ and $t^{(2)} = 2$ to ensure estimability. Additionally, there is no restriction in assuming $\beta_0 = 0$. According to (5.1), the allocation of the third observation can now be chosen adaptively: $T^{(3)} = 1$ if $|X_1(1)| \geq |X_2(2)|$ and $T^{(3)} = 2$ if $|X_1(1)| < |X_2(2)|$.

The interpretation of the allocation rule is self-evident: we are going to make the next observation in that group where the largest deviation from the hypothesis null $\beta = \beta_0$ occurs. If we assume independent identically distributed Gaussian noise with known variance $\sigma^2 = 1$, then it is possible to determine the conditional and unconditional densities of the distribution of $\widehat{\beta} = \widehat{\beta}_{D(T^{(N)})}$. The adaptive analysis of the observations is based on conditional "χ^2-statistics"

(6.1)

$$S_1 = \begin{cases} 2\widehat{\beta}_1^2 + \widehat{\beta}_2^2 & \text{if } T^{(3)} = 1 \\ \\ \widehat{\beta}_1^2 + 2\widehat{\beta}_2^2 & \text{if } T^{(3)} = 2. \end{cases}.$$

Since the unconditional distribution of $\widehat{\beta}$ is symmetric under the hypothesis one can consider alternatively the unconditional "χ^2-statistics":

(6.2)

$$S_2 = \widehat{\beta}_1^2 + \widehat{\beta}_2^2.$$

Both statistics do not follow a χ^2-distribution (even not conditionally) and the distributions cannot be given in closed form. Hence the critical values and the power have to be calculated by means of numerical integration on the basis of the conditional densities $f_{\widehat{\beta}|T^{(3)}=i}$.

In Figure 1 we show the computed power of both tests with level $\alpha = 0.05$ compared to the power of a χ^2-test based on the usual randomized design where the allocation of $T^{(3)}$ is made with equal probability $1/2$ independent of the observations. Similar pictures arise for $\alpha = 0.10$.

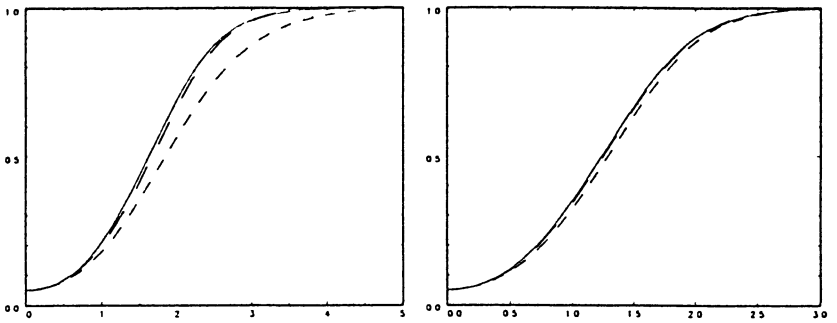

FIGURE 1: Power of $S1$(—) and $S2$(– –) compared to a χ^2-test based on a randomized design(- - -); a) $\beta_1 > 0, \beta_2 = 0$; b) $\beta_1 = \beta_2 > 0$.

We notice that the conditional test based on the adaptive design (6.1) is reasonably more powerful than the unconditional one (6.2) which is still better than the χ^2-test based on the randomized design.

7. **Two-stage procedures.** For the example of a one-way layout with possibly different variances we indicate a two-stage approach. Let X_{ij} be the j-th observation in group $i = 1, \ldots, K$ and assume that all observations are independent and that all the observation within a group are identically distributed according to a Gaussian law with mean β_i and variance σ_i^2 respectively. Our objective is to estimate $\beta_1, \ldots, \beta_K$ by the means

$$\widehat{\beta_i} = \frac{1}{N_i} \sum_{j=1}^{N_i} X_{ij},$$

where N_i is the number of experiments in the ith group. To measure the quality of our estimates, we consider again the mean squared Euclidean distance

$$\sum_{i=1}^{K} E((\widehat{\beta}_i - \beta_i)^2),$$

and alternatively, the maximal variance

$$\max_{i=1,\ldots,K} E((\widehat{\beta}_i - \beta_i)^2).$$

If the total number N of experiments is fixed then

$$\sum_{i=1}^{K} E((\widehat{\beta}_i - \beta_i)^2)$$

is minimized if the numbers N_i of observations are proportional to the standard deviations σ_i, i. e.

$$N_i \approx N\sigma_i / \sum_{k=1}^{K} \sigma_k.$$

In this case,

$$\sum_{i=1}^{K} E((\widehat{\beta}_i - \beta_i)^2) = \sum_{i=1}^{K} \sigma_i^2/N_i \approx (\sum_{i=1}^{K} \sigma_i)^2/N.$$

Alternatively, if N_i is proportional to σ_i^2, i. e.

$$N_i \approx N\sigma_i^2 / \sum_{k=1}^{K} \sigma_k^2,$$

then the maximal variance is minimized and attains the value

$$\max_{i=1,\ldots,K} E((\widehat{\beta}_i - \beta_i)^2) = \max_{i=1,\ldots,K} \sigma_i^2/N_i \approx \sum_{i=1}^{K} \sigma_i^2/N.$$

However, in general, the relative magnitude of the variances σ_i^2 is unknown such that the optimum allocation schemes cannot be constructed.

Let us now turn from the fixed sample size problem to the alternative design problem of a given prescribed accuracy, i. e. a prespecified upper bound for

$$\sum_{i=1}^{K} E((\widehat{\beta}_i - \beta_i)^2)$$

and for

$$\max_{i=1,\ldots,K} E((\widehat{\beta}_i - \beta_i)^2).$$

In the case of a totally known variance structure, the solutions to these design problems are proportional to the fixed sample size situation. But already in the classical case of an unknown scaling factor σ^2 a prespecified accuracy cannot be attained by an experiment based on a deterministic design. For sampling from one group, Stein's (1945) two-stage procedure shows a way out of this problem. We will make use of this idea for sampling from K groups in a one-way layout:

We start with a preliminary sample $X_{i_1}, \ldots, X_{i,N_0}$ of equal size $N_0 \geq 2$ in each group $i = 1, \ldots, K$. The choice of an initial equal allocation rule can be justified by the fact that it is minimax over all possible combinations of variances $\sigma_1^2, \ldots, \sigma_K^2$ with respect to relative efficiency [for a more detailed description in a similar situation we refer to Page (1990)]. Then we can estimate σ_i^2 by the empirical variances

$$\widehat{\sigma}_i^2 = \frac{1}{N_0 - 1} \sum_{j=1}^{N_0} (X_{i,j} - \frac{1}{N_0} \sum_{k=1}^{N_0} X_{ik})^2$$

and substitute $\widehat{\sigma}_i^2$ for σ_i^2 into the allocation rules.

In particular, for the design criterion of achieving a value of the mean squared error

$$\sum_{i=1}^{K} E((\widehat{\beta}_i - \beta_i)^2)$$

which is smaller than a prespecified bound, the sample sizes N_i are determined according to

(7.1)
$$N_i = \max\left([\widehat{\sigma}_i(\widehat{\sigma}_1 + \cdots + \widehat{\sigma}_K)/z] + 1, N_0\right),$$

where $[x]$ denotes the largest integer less than x and where z has to be specified appropriately to guarantee the accuracy desired (see below). Then

$$\widehat{\beta} = \frac{1}{N_i} \sum_{j=1}^{N_i} X_{ij}$$

is an unbiased estimator of β_i and $N_i/N_k \approx \widehat{\sigma}_i/\widehat{\sigma}_k$. When N_i is chosen according to (7.1),

$$\sum_{i=1}^{K} Var(\widehat{\beta}_i) \leq z(N_0 - 1)^2/(N_0 - 3)^2,$$

and with an appropriate z, any prespecified accuracy can be achieved.

For the maximum-variance criterion

(7.2)
$$N_i = \max\left(\left[\widehat{\sigma}_i^2/z\right] + 1, N_0\right)$$

which ensures $N_i/N_k \approx \widehat{\sigma}_i^2/\widehat{\sigma}_k^2$. In this case,

$$\max_{i=1,\ldots,K} Var(\widehat{\beta}_i) \leq z(N_0 - 1)/(N_0 - 3)$$

and the accuracy can be predetermined by the choice of z.

As could be expected the equal allocation rule for obtaining the same accuracy will result in a larger expected total number of observations. For example in case of the maximum-variance criterion the equal allocation rule requires

$$\widetilde{N}_1 = \cdots = \widetilde{N}_K = \max\left(\left[\max_{i=1,\ldots,K} \widehat{\sigma}_i^2/z\right] + 1, N_0\right)$$

for attaining the same accuracy. In this situation the relative efficiency with respect to the expected sample size

$$E(\textstyle\sum_{i=1}^{K} N_i)/E(\sum_{i=1}^{K} \widetilde{N}_i) \approx (\sigma_1^2 + \cdots + \sigma_K^2)/(K \max_{i=1,\ldots,K} \sigma_i^2) < 1.$$

Similar results can be obtained for the Behrens-Fisher problem of comparing the means of two populations with possible different variances [see, for example, Schwabe (1993) and additional references given therein]. In case of multivariate observations it would be interesting to develop analogous results based on the two-stage procedures proposed by Healy (1956), Chatterjee (1959), and Dudewicz, Hyakutake and Taneja (1991).

Acknowledgement. The author wishes to express his thanks to the referees for their helpful comments.

References

BANDEMER, H., ed. (1977). *Theorie und Anwendung der optimalen Versuchsplanung I. Handbuch zur Theorie.* Berlin: Akademie.

CHATTERJEE, S.K. (1959). On an extension of Stein's two-sample procedure to the multi-normal problem. *Bulletin of the Calcutta Statistical Association* **8** 121-148.

DUDEWICZ, E.J., HYAKUTAKE, H. and TANEJA, V.S. (1991). The heteroscedastic method II: simplified sampling. *American Journal of Mathematics and Management Sciences* **11** 263-276 .

FEDOROV, V.V. (1972). *Theory of Optimal Experiments.* New York: Academic .

FORD, I., TITTERINGTON, D.M. and KITSOS, C.P. (1989). Recent advances in nonlinear experimental design. *Technometrics* **31** 49-60 .

GEBHARDT, R. and HECKENDORFF, H. (1983). Zur sequentiellen Versuchsplanung für das Regressionsproblem. *Math. Operationsforsch. Statist., ser. statist .* **14** 355-366.

HEALY, W.C. (1956). Two-sample procedures in simultaneous estimation . *The Annals of Mathematical Statistics* **27** 687-702 .

KIEFER, J. (1958). On the nonrandomized optimality and randomized nonoptimality of symmetrical designs. *The Annals of Mathematical Statistics* **29** 675-699 .

PAGE, C.F. (1990). Allocation proportional to coefficients of variation when estimating the product of parameters. *Journal of the American Statistical Association* **85** 1134-1139.

PÁZMAN, A. (1986). *Foundations of Optimum Experimental Design.* Dordrecht: Reidel .

ROBBINS, H., SIMONS, G., and STARR, N. (1967). A sequential analogue to the Behrens-Fisher problem. *The Annals of Mathematical Statistics* **38** 1384-1391 .

SCHWABE, R. (1987). On an adaptive design in regression . *Statistics* **18** 521-525 .

SCHWABE, R. (1990). Adaptive designs for linear problems . *Stochastic Methods in Experimental Sciences* (Kasprzak, W. and Weron, A., eds) . Singapore: Scientific, 407-420.

SCHWABE, R. (1991). Adaptive design in the one-way layout. *Journal of Statistical Planning and Inference* **28** 319-338 .

SCHWABE, R. (1993). Some two-stage procedures for treating the Behrens-Fisher problem. *Model-Oriented Data Analysis* (Müller, W.G., Wynn, H.P. and Zhigljavsky, A.A., eds). Heidelberg: Physica, 81-89. .

SILVEY, S.D. (1980) . *Optimal Design.* London: Chapman and Hall .

STEIN, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. *Annals of Mathematical Statistics* **16** 243-258 .

MATHEMATISCHES INSTITUT
FREIE UNIVERSITÄT BERLIN
ARNIMALLEE 2–6
D-14 195 BERLIN
GERMANY