The Annals of Mathematical Statistics
1970, Vol. 41, No. 6, 1917–1934

THE OPTIMUM DESIGN OF A TWO-FACTOR EXPERIMENT USING PRIOR INFORMATION[1]

BY R. J. OWEN

University College of Wales

0. Summary. This paper is concerned with a two-factor analysis of variance situation, the factors being conveniently referred to as blocks and treatments. One of the treatments has the role of a control. Attention is focused on inference about the treatment parameters, the block parameters being regarded as nuisance parameters. With a general multivariate normal form for the distribution of errors and for the prior distribution on the block and treatment parameters, the posterior distribution of the treatment parameters is derived. With a quadratic loss function an algorithm is derived for the optimum allocation of treatments over a given sample with known blocking. In special cases the optimum allocation can be written down immediately and the algorithm need not be resorted to.

1. Introduction. In the classical mold a study has been made of the analysis of variance situation in which the experimental units are stratified according to two factors, blocks and treatments say. Let one of the treatments be called the control treatment, let I ($\geqq 1$) denote the number of the remaining treatments and Q ($\geqq 1$) the number of blocks. It is easy to see that for given block sizes, with a "minimum variance" criterion for optimality and uncorrelated error assumptions, the optimum design for comparing each treatment individually with the control is (ignoring integer difficulties) to allocate in each block $I^{\frac{1}{2}}$ times as many units to the control treatment as to each of the other treatments. That this allocation has optimal properties in the case of independent observations has been known for some time, see for example Fieller (1947) and Dunnett (1955). Work related to the present paper includes that of Ericson (1965) where, in the one-factor case, using a Bayesian decision theoretic approach (quadratic loss and a linear cost function), an optimal design is derived for estimation of a linear combination of population means. Bechhofer, in a recent unpublished report, also considered the one-factor case and was concerned with estimating the contrasted effects of the treatments with the control. With a confidence interval criterion for optimality, the optimum design was derived and shown to reduce to the above allocation in special cases.

The present paper is also concerned with inference about the contrasted effects of the individual treatments with the control treatment, the block effects being of no interest but here prior information about the parameters is assumed and the errors are not necessarily uncorrelated. Wald's decision theory and the Bayesian framework is adopted with a linear, no interaction model and multivariate normal prior and error distributions. This model is formalized more explicitly in the next

Received October 13, 1969; revised March 30, 1970.

[1] Research supported by the Science Research Council in Wales and completed at Educational Testing Service, Princeton.

section. In Section 3, the posterior distribution of the treatment parameters is derived. In Section 4, interest is focused on estimation and with a quadratic loss function a generalization of the above classical result is derived in the form of a computational algorithm which yields the optimum allocation of treatments in at most Q steps. In Section 5, special cases are found in which the computational algorithm simplifies to give an algebraic expression for the answer. A corollary to one of the special case results is that in the classical mold the "square root rule" mentioned above is still optimum (in a sense) when correlated errors are allowed. Section 6 contains an illustrative example.

NOTATION. $(\mathbf{A})_{ij}$ denotes the element in the ith row and jth column of $\mathbf{A}$ and $||a_{ij}||$ denotes the matrix with (i,j)th element a_{ij}. $\mathbf{I}_n$ denotes the $n \times n$ identity matrix and $\mathbf{1}_n$ denotes the $n \times 1$ column of 1's. $\mathbf{dg}(\mathbf{A}_1, \cdots, \mathbf{A}_k)$ of course denotes the partition matrix which has the matrices $\mathbf{A}_1, \cdots, \mathbf{A}_k$ down the diagonal of partitions and zeros elsewhere and for the vector $\mathbf{a}$, $\mathbf{dg}\,\mathbf{a}$ denotes the diagonal matrix with $(\mathbf{a})_1, (\mathbf{a})_2, \cdots, (\mathbf{a})_n$ down the diagonal.

For an expression R_{iq}, $\{R_{iq}\}$ is shorthand for $(R_{iq}\ i = 1, \cdots, I,\ q = 1, \cdots, Q)$. Either one of i or q may be absent.

$\mathbf{x} \mid \boldsymbol{\theta} \sim N(\mathbf{m}, \mathbf{V})$ means given $\boldsymbol{\theta}$, $\mathbf{x}$ has a multivariate normal distribution with mean $\mathbf{m}$ and dispersion matrix $\mathbf{V}$.

Prior densities are denoted by $\pi(\cdot)$, thus $\pi(\boldsymbol{\theta})$ is the prior density on $\boldsymbol{\theta}$.

2. Specification of the model. Let P denote the population of experimental units which are classified in two mutually exclusive and exhaustive ways, say according to treatments and blocks. Let $\mathscr{T}_0, \cdots, \mathscr{T}_I$ denote the control treatment and I treatments respectively and $\mathscr{B}_1, \cdots, \mathscr{B}_Q$ denote the blocks. For each treated unit there is an observable scalar response, y say.

Let $\tau_0, \tau_1, \cdots, \tau_I$, $\beta_1, \cdots, \beta_Q$ be parameters referring to $\mathscr{T}_0, \mathscr{T}_1, \cdots, \mathscr{T}_I$, $\mathscr{B}_1, \cdots, \mathscr{B}_Q$ respectively. It is assumed that for a unit in $\mathscr{B}_q$ which received $\mathscr{T}_i$

$$y = \tau_i + \beta_q + \varepsilon \qquad (i = 0, 1, \cdots, I, q = 1, \cdots, Q). \tag{1}$$

In an experiment on a given blocked and treated sample, the parameters $\{\tau_i - \lambda, \beta_q + \lambda\}$ clearly fit any possible response vector $\mathbf{y}$ equally well for all λ. The condition that τ_0 is fixed and without loss of generality

$$\tau_0 = 0 \tag{2}$$

expresses the "origin of reference" role $\mathscr{T}_0$ plays in the experiment making the parameters identifiable

Let the units of the sample be arranged in some order and let N denote the sample size. For the nth unit let y_n denote the observed scalar and ε_n the error term. Write $\mathbf{y} = (y_1, \cdots, y_N)'$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_N)'$. Denote by $\mathbf{F}$ the $N \times I$ matrix with the nth $(n = 1, \cdots, N)$ row having a 1 in the ith position and zeros elsewhere if the nth experimental unit receives $\mathscr{T}_i$ and consisting entirely of zeros if the nth experimental unit receives $\mathscr{T}_0$. Similarly define $\mathbf{G}$ to be the $N \times Q$ matrix with the nth $(n = 1, \cdots, N)$ row having a 1 in the qth position and zeros elsewhere if the nth

unit is in $\mathscr{B}_q$. The *design* of an experiment is the choice of the numbers of units from the blocks and subsequent allocation of treatments, i.e., the design is the choice of $(\mathbf{F}, \mathbf{G})$. Denoting $(\tau_1, \cdots, \tau_I)'$, $(\beta_1, \cdots, \beta_Q)'$ by $\boldsymbol{\tau}$ and $\boldsymbol{\beta}$ respectively and referring to (2) the equation corresponding to (1) for the sample is

$$\mathbf{y} = \mathbf{F}\boldsymbol{\tau} + \mathbf{G}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{3}$$

In the above setup let H denote the experimenter's knowledge or belief about $\boldsymbol{\tau}$, $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$ prior to the observation of $\mathbf{y}$. A normal distributional assumption is taken, namely

$$\boldsymbol{\varepsilon} \,|\, \boldsymbol{\tau}, \boldsymbol{\beta}, H \sim N(\mathbf{0}, \mathbf{E}), \tag{4}$$

the dispersion matrix $\mathbf{E}$ being known, positive definite and not dependent on $\boldsymbol{\tau}$ or $\boldsymbol{\beta}$.

The prior information on $\boldsymbol{\tau}$ and $\boldsymbol{\beta}$ (i.e., given only H) is taken to be of multivariate normal form,

$$\begin{pmatrix} \boldsymbol{\tau} \\ \boldsymbol{\beta} \end{pmatrix} \Bigg| H \sim N\left\{ \begin{pmatrix} \mathbf{t} \\ \mathbf{b} \end{pmatrix}, \begin{pmatrix} \mathbf{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix} \right\} \tag{5}$$

where $\mathbf{t}(I \times 1)$ and $\mathbf{b}$ are known and $\mathbf{T}(I \times I)$ and $\mathbf{B}$ are known and positive definite.

For the remainder of this paper H will be omitted but is always to be taken as given.

3. Basic results for inference about the treatment parameters.

3.1. *The posterior distribution of* $\boldsymbol{\tau}$. It is required to eliminate the nuisance parameter $\boldsymbol{\beta}$. By Bayes' theorem, since the prior attitude to $\boldsymbol{\tau}$ is not affected by choice of $(\mathbf{F}, \mathbf{G})$ and since $p(\mathbf{y} \,|\, \mathbf{F}, \mathbf{G})$ does not depend on $\boldsymbol{\tau}$,

$$p(\boldsymbol{\tau} \,|\, \mathbf{y}, \mathbf{F}, \mathbf{G}) \propto p(\mathbf{y} \,|\, \boldsymbol{\tau}, \mathbf{F}, \mathbf{G}) \pi(\boldsymbol{\tau}). \tag{6}$$

The equation for the sample, (3), may be written

$$\mathbf{y} = \mathbf{F}\boldsymbol{\tau} + (\mathbf{G}, \mathbf{I}_N) \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\varepsilon} \end{pmatrix}.$$

Referring to (4) and (5)

$$\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\varepsilon} \end{pmatrix} \Bigg| \boldsymbol{\tau}, \mathbf{F}, \mathbf{G} \sim N\left\{ \begin{pmatrix} \mathbf{b} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{E} \end{pmatrix} \right\}$$

and since $(\mathbf{G}, \mathbf{I}_N)$ has full rank

$$\mathbf{y} \,|\, \boldsymbol{\tau}, \mathbf{F}, \mathbf{G} \sim N(\mathbf{F}\boldsymbol{\tau} + \mathbf{G}\mathbf{b}, \mathbf{V}) \tag{7}$$

where $\mathbf{V} = \mathbf{G}\mathbf{B}\mathbf{G}' + \mathbf{E}$ has full rank.

From the standard multivariate normal theory, it is easy to verify that the posterior distribution of $\boldsymbol{\tau}$, given $(\mathbf{y}, \mathbf{F}, \mathbf{G})$ is a multivariate normal distribution with mean vector $\mathbf{m}$ and dispersion matrix $\mathbf{D}$ where

$$\mathbf{D}^{-1} = \mathbf{F}'\mathbf{V}^{-1}\mathbf{F} + \mathbf{T}^{-1} \tag{8}$$

and

$$\mathbf{m} = \mathbf{D}[\mathbf{F}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{G}\mathbf{b}) + \mathbf{T}^{-1}\mathbf{t}] \tag{9}$$

where recall $\mathbf{V} = \mathbf{GBG}' + \mathbf{E}$. Notice that none of $\mathbf{y}$, $\mathbf{t}$ and $\mathbf{b}$ appear in the expression for $\mathbf{D}$ given in (8). This completes the derivation of

THEOREM 1. *The distribution of τ posterior to the experiment has a multivariate normal distribution with mean $\mathbf{m}$ and dispersion matrix $\mathbf{D}$ expressions for which are given in* (8) *and* (9).

An important corollary to this theorem will now be derived, part of which is a self-contained classical result.

Referring to (7) and **A1**(viii) (i.e. the part of appendix **A1** specified in the parentheses),

$$\mathbf{V}^{-1} = \mathbf{E}^{-1} - \mathbf{E}^{-1}\mathbf{G}(\mathbf{G}'\mathbf{E}^{-1}\mathbf{G} + \mathbf{B}^{-1})^{-1}\mathbf{G}'\mathbf{E}^{-1}. \tag{10}$$

Define $\mathbf{V}_0^{-1}$ to be the value of $\mathbf{V}^{-1}$ when $\mathbf{B}^{-1}$ is the zero matrix. It follows that

$$\mathbf{V}_0^{-1} \text{ is singular and } \mathbf{V}_0^{-1}\mathbf{G} = 0. \tag{11}$$

It is seen, referring to (8) and (9), that

$$\mathbf{D}^{-1} \text{ and } \mathbf{m} \text{ are continuous in } \mathbf{T}^{-1} \text{ and } \mathbf{B}^{-1} \text{ even at } \mathbf{B}^{-1} = 0. \tag{12}$$

By virtue of the form of the likelihood [(3), (4)] it follows that when $\mathbf{T}^{-1}$ and $\mathbf{B}^{-1}$ are the zero matrices (vague prior information on the parameters) the posterior distribution of τ is just $N(\hat{\tau}, \operatorname{Var}\hat{\tau})$ where $\hat{\tau}$ is the maximum likelihood or weighted least squares estimator of τ. Theorem 1, (10), (11), (12) and this result establish

COROLLARY 1. *Let $\hat{\tau}$ denote the maximum likelihood (or equivalently, the weighted least squares) estimator of τ and let $(\mathbf{F}, \mathbf{G})$ be of full rank, then in the notation of Theorem* 1:

$$\hat{\tau} = \lim_{\mathbf{T}^{-1} \to 0 \,\&\, \mathbf{B}^{-1} \to 0} \mathbf{m} = (\mathbf{F}'\mathbf{V}_0^{-1}\mathbf{F})^{-1}\mathbf{F}'\mathbf{V}_0^{-1}\mathbf{y}$$

and

$$\operatorname{Var}\hat{\tau} = \lim_{\mathbf{T}^{-1} \to 0 \,\&\, \mathbf{B}^{-1} \to 0} \mathbf{D} = (\mathbf{F}'\mathbf{V}_0^{-1}\mathbf{F})^{-1}$$

where

$$\mathbf{V}_0^{-1} = \mathbf{E}^{-1} - \mathbf{E}^{-1}\mathbf{G}(\mathbf{G}'\mathbf{E}^{-1}\mathbf{G})^{-1}\mathbf{G}'\mathbf{E}^{-1}.$$

Of course $\hat{\tau}$ has a multivariate normal distribution with mean τ (because $\hat{\tau}$ is linear in $\mathbf{y}$). Note also that "$(\mathbf{F}, \mathbf{G})$ is of full rank" is a necessary, as well as a sufficient condition for the inverses in the above expressions to exist. This follows from **A1**(v) and the equation $|(\mathbf{F}, \mathbf{G})'\mathbf{E}^{-1}(\mathbf{F}, \mathbf{G})| = |\mathbf{G}'\mathbf{E}^{-1}\mathbf{G}|\,|\mathbf{F}'\mathbf{V}_0^{-1}\mathbf{F}|$ which holds if $\mathbf{G}'\mathbf{E}^{-1}\mathbf{G}$ is nonsingular.

The following definition is required for the next subsection.

(13) Let a_{iq} $(i = 0, 1, \cdots, I, q = 1, \cdots, Q)$ denote the number of units of the sample which received $\mathscr{T}_i$ and are from $\mathscr{B}_q$. The $(I+1) \times Q$ array $\{a_{0q}, a_{iq}\}$ is

called the allocation for the sample and the $I \times Q$ matrix $||a_{iq}\, i = 1, \cdots, I, q = 1, \cdots, Q||$ is denoted by **A**. **A** describes the allocation $\{a_{0q}, a_{iq}\}$ if the block sizes $s_1, \cdots, s_Q$ are known.

3.2. *An expression for* $\mathbf{D}^{-1}$ *in terms of the allocation.* Without loss of generality and for convenience, let the arrangement of the units in the sample be such that all those of $\mathscr{B}_1$ occupy the first places, then all those of $\mathscr{B}_2$ and so on. Denoting the number of elements of the sample in $\mathscr{B}_q$ by s_q, with this ordering of the sample

$$\mathbf{G} = \mathbf{dg}(\mathbf{1}_{s_1}, \mathbf{1}_{s_2}, \cdots, \mathbf{1}_{s_Q}). \tag{14}$$

Specialization. Here and for the remainder of this paper the block sizes $\{s_q\}$ are taken to be positive and the distribution of errors is specialized to within block homoscedasticity and constant correlation between different errors in a given pair of (coincident or different) blocks. In the notation of **2** and 3.2 the specialization may be expressed (introducing notation) by the equation

$$\mathbf{E} = \mathbf{G}\hat{\mathbf{E}}\mathbf{G}' + \mathbf{\Delta} \tag{15}$$

where $\mathbf{\Delta}$ denotes $\mathbf{dg}(e_1 \mathbf{I}_{s_1}, \cdots, e_Q \mathbf{I}_{s_Q})$ where $e_1, \cdots, e_Q$ are positive. Of course $\hat{\mathbf{E}}$ has to be symmetric and $\hat{\mathbf{E}}$ and $e_1, \cdots, e_Q$ have to be such that **E** is positive definite. Denote $\hat{\mathbf{E}}$ by $||e_{qr}||$ so $\mathbf{G}\hat{\mathbf{E}}\mathbf{G}' = ||e_{qr} J_{qr}\, q, r = 1, \cdots, Q||$ where $J_{qr} = \mathbf{1}_{s_q}\mathbf{1}'_{s_r} (q, r = 1, \cdots, Q)$.

With this specialization (7) becomes $\mathbf{V} = \mathbf{G}(\mathbf{B} + \hat{\mathbf{E}})\mathbf{G}' + \mathbf{\Delta}$ and so $\mathbf{V}^{-1} = -\mathbf{\Delta}^{-1}\mathbf{G}[\mathbf{G}'\mathbf{\Delta}^{-1}\mathbf{G} + (\mathbf{B} + \hat{\mathbf{E}})^{-1}]^{-1}\mathbf{G}'\mathbf{\Delta}^{-1} + \mathbf{\Delta}^{-1}$ [**A1**(viii)]. Noting that $\mathbf{G}'\mathbf{\Delta}^{-1}\mathbf{G} = \mathbf{dg}(s_1 e_1^{-1}, \cdots, s_Q e_Q^{-1})$ and that in the first term the block partitioned matrix between the outside $\mathbf{\Delta}^{-1}$'s is still a block partitioned matrix when pre- and post-multiplied by $\mathbf{\Delta}^{-1}$ it follows that

$$\mathbf{V}^{-1} = -\mathbf{GCG}' + \mathbf{\Delta}^{-1} \quad \text{where} \tag{16}$$

$$\mathbf{C}^{-1} = \mathbf{dg}(s_1 e_1, \cdots, s_Q e_Q) + \mathbf{dg}(e_1, \cdots, e_Q)(\mathbf{B} + \hat{\mathbf{E}})^{-1}\mathbf{dg}(e_1, \cdots, e_Q). \tag{17}$$

It is easy to see that $\mathbf{F}'\mathbf{G} = \mathbf{A}$ and $\mathbf{F}'\mathbf{\Delta}^{-1}\mathbf{F} = \mathbf{dg}\,\mathbf{Ap}$ where $\mathbf{p} = (e_1^{-1}, \cdots, e_Q^{-1})'$ so from (16) and Theorem 1 there follows the required expression for $\mathbf{D}^{-1}$ in terms of the allocation for given block sizes $s_1, \cdots, s_Q$. For the second part note that $(\mathbf{B} + \hat{\mathbf{E}})^{-1} \to \mathbf{0}$ as $\mathbf{B}^{-1} \to \mathbf{0}$ and that "full rank for $(\mathbf{F}, \mathbf{G})$" is a necessary and sufficient condition for existence of the limit of **D** [Corollary 1]. Now $(\mathbf{F}, \mathbf{G})$ has the same rank as $(\mathbf{F}, \mathbf{G})'(\mathbf{F}, \mathbf{G})$ which is expressed in terms of **A** and $\{s_q\}$ on noting that $\mathbf{F}'\mathbf{F} = \mathbf{dg}(\mathbf{A}\mathbf{1}_Q)$ and $\mathbf{G}'\mathbf{G} = \mathbf{dg}(s_1, \cdots, s_Q)$. The final sufficient condition is established from the observation that the sum along the ith row of $\lim \mathbf{D}^{-1}$ is $\sum_{q=1}^{Q} a_{iq} a_{0q}/(s_q e_q)$ and the Lévy–Desplanques theorem [see Marcus–Minc (1964)].

THEOREM 2. *In the notation and specialization of* (15) *the expression for the posterior dispersion matrix of the treatment parameters,* **D**, *in terms of the allocation is* $\mathbf{D}^{-1} = -\mathbf{ACA}' + \mathbf{dg}(\mathbf{Ap}) + \mathbf{T}^{-1}$ *where* **p** *denotes* $(e_1^{-1}, e_2^{-1}, \cdots, e_Q^{-1})'$ *and* $\mathbf{C}^{-1}$ *is given above in* (17).

If $\mathbf{T}^{-1}$ and $\mathbf{B}^{-1}$ tend to zero matrices (vague prior information) then $\mathbf{D}^{-1} \to -\mathbf{A}\,\mathbf{dg}(s_1^{-1}e_1^{-1}, \cdots, s_Q^{-1}e_Q^{-1})\mathbf{A}' + \mathbf{dg}(\mathbf{Ap})$ and equality obtains in the limit if $(\mathbf{B}+\hat{\mathbf{E}})^{-1}$ is defined to be $\mathbf{0}$ when $\mathbf{B}^{-1} = \mathbf{0}$. $\lim \mathbf{D}$ exists iff

$$\begin{bmatrix} \mathbf{dgA1}_Q & \mathbf{A} \\ \mathbf{A}' & \mathbf{dg}(s_1, \cdots, s_Q) \end{bmatrix} \quad \text{is of full rank.}$$

A sufficient condition for $\lim \mathbf{D}$ to exist is simply that "each treatment is taken in some block, together with the control."

4. With given block sizes, the optimum allocation of treatments for estimation of the treatment parameters.

4.1. In the Bayesian decision theoretic framework the optimum criterion for inference about τ is expressed in a loss function. Here it is required that the loss function be appropriate for estimation and it is taken to be the quadratic form

$$L(\mathbf{d}, \tau) \equiv (\mathbf{d}-\tau)'\mathbf{W}(\mathbf{d}-\tau) \tag{18}$$

where $\mathbf{d}$ is an element of the decision (or estimation) space (the set of $I \times 1$ vectors) and $\mathbf{W}$ is an $I \times I$ positive definite and without loss of generality symmetric matrix. It is known that for an experiment with given design (i.e., given $\mathbf{F}, \mathbf{G}$) and the above loss function, Bayes' estimator of τ is $\mathbf{m}$, the posterior expectation of τ [see Theorem 1] and the expected loss with Bayes' estimator is $\operatorname{tr}\mathbf{WD}$ [see Theorem 1 and Theorem 2].

4.2. The expected posterior loss with Bayes' estimator, $\operatorname{tr}\mathbf{WD}$ [4.1] doesn't depend on $\mathbf{y}$ [(8)] and hence the criterion for optimal design is to minimize $\operatorname{tr}\mathbf{WD}$. For given $\mathbf{G}$ it is therefore required to choose $\mathbf{F}$ to minimize $\operatorname{tr}\mathbf{WD}$. Referring to Theorem 2 all designs $(\mathbf{F}, \mathbf{G})$ with the same allocation $\{a_{0q}, a_{iq}\}$ will have the same associated $\mathbf{D}$ and so it is only required to consider the different allocations. The minimization problem is split into two stages according to

$$\min_{\{a_{0q},a_{iq}\} \mid \{s_q\}} \operatorname{tr}\mathbf{WD} = \min_{\{a_{0q}\} \mid \{s_q\}} \min_{\{a_{iq}\} \mid \{s_q, a_{0q}\}} \operatorname{tr}\mathbf{WD}. \tag{19}$$

NOTATION. (i) An $(I+1) \times Q$ array of nonnegative integers $\{a_{0q}, a_{iq}\}$ has been called an allocation. For mathematical convenience it is necessary to consider such an array with the entries not restricted to integer values. This array with nonnegative reals for entries will be called a "continuous allocation." The former, more restricted array will be referred to as an "integer allocation" or simply as an "allocation."

(ii) $\mathbf{D}(\mathscr{A})$ shall denote the dispersion matrix $\mathbf{D}$ under allocation $\mathscr{A}$.

(iii) For square matrices $\mathbf{X}, \mathbf{Y}$ of the same order, $\mathbf{X} \geqq \mathbf{Y}$ means that $\mathbf{X}-\mathbf{Y}$ is nonnegative definite and $\mathbf{X} > \mathbf{Y}$ means that $\mathbf{X}-\mathbf{Y}$ is positive definite.

(20) In this paper the minimization (19) is taken over continuous allocations (except in Theorem 6) though of course only experiments with integer allocations exist. However when the block sizes $\{s_q\}$ are large in comparison with I any integer allocation near the optimum continuous allocation will, because of continuity and stationarity of the loss function at the optimum continuous allocation, have a

hardly larger expected loss. The solution of the above modified problem is usually considered the first priority [see, for example, Kiefer (1959), page 281].

(21) Subsequent results in this paper take **C** [(17), (15)] to be positive definite. Refering to (17) and **A1**(vii) and (iv) it follows by continuity that **C** is positive definite if $\mathbf{B}+\hat{\mathbf{E}}$ is nonnegative definite. Since **B** is positive definite, this always occurs if $\hat{\mathbf{E}}$ is nonnegative definite. Whatever the value of $\hat{\mathbf{E}}$, $\mathbf{B}+\hat{\mathbf{E}}$ is nonnegative definite if, roughly speaking, **B** has sufficiently large characteristic roots, i.e., if the prior information about $\boldsymbol{\beta}$ is sufficiently vague. Notice that $\hat{\mathbf{E}}$ is nonnegative definite in the case of uncorrelated errors (when $\hat{\mathbf{E}} = 0$) and so **C** is always positive definite in this case. More generally **C** is positive definite if $e_{qr} = e$ for $q \neq r$, $e_{qq} = e'$ and $e' \geqq e \geqq 0$. $\hat{\mathbf{E}}$ is not, however, always nonnegative definite, for example, if there are one or more blocks such that their block error covariances (e_{qq} for $\mathscr{B}_q$) are negative.

4.3.

LEMMA 1. *If* **C** *is positive definite* [see (21)], *and* **W** *is diagonal the optimum continuous allocation for given block sizes* $\{s_q\}$ *and given amount of control per block* $\{a_{0q}\}$, *is unique.*

PROOF. For given block sizes $\{s_q\}$ *suppose* the inner minimum of (19) occurs at the different continuous allocations represented (in the notation of (13)) by the $I \times Q$ matrices $\mathbf{A}_1, \mathbf{A}_2$. Write $\mathbf{A}_3 = \lambda\mathbf{A}_1 + (1-\lambda)\mathbf{A}_2$ where $0 < \lambda < 1$. Let $\mathbf{D}_i$ denote $\mathbf{D}(\mathbf{A}_i)$ $(i = 1, 2, 3)$. It may be verified using Theorem 2 that

$$\mathbf{D}_3^{-1} = \lambda\mathbf{D}_1^{-1} + (1-\lambda)\mathbf{D}_2^{-1} + \lambda(1-\lambda)(\mathbf{A}_1 - \mathbf{A}_2)\mathbf{C}(\mathbf{A}_1 - \mathbf{A}_2)'. \tag{22}$$

Referring to **A1**(v):

$$\lambda(1-\lambda)(\mathbf{A}_1 - \mathbf{A}_2)\mathbf{C}(\mathbf{A}_1 - \mathbf{A}_2)' \underset{\neq}{\geqq} \mathbf{0}. \tag{23}$$

Referring to (22) and (23)

$$\mathbf{D}_3^{-1} \underset{\neq}{\geqq} \lambda\mathbf{D}_1^{-1} + (1-\lambda)\mathbf{D}_2^{-1}$$

and therefore

$$\mathbf{D}_3 \underset{\neq}{\leqq} [\lambda\mathbf{D}_1^{-1} + (1-\lambda)\mathbf{D}_2^{-1}]^{-1} \qquad \text{[A1(iii)]}. \tag{24}$$

However

$$[\lambda\mathbf{D}_1^{-1} + (1-\lambda)\mathbf{D}_2^{-1}]^{-1} \leqq \lambda\mathbf{D}_1 + (1-\lambda)\mathbf{D}_2 \qquad \text{[A1(ii)]} \tag{25}$$

and so

$$\mathbf{D}_3 \underset{\neq}{\leqq} \lambda\mathbf{D}_1 + (1-\lambda)\mathbf{D}_2 \qquad \text{[(24), (25)]} \tag{26}$$

which implies that

$$\mathbf{W}\mathbf{D}_3 \underset{\neq}{\leqq} \lambda\mathbf{W}\mathbf{D}_1 + (1-\lambda)\mathbf{W}\mathbf{D}_2 \qquad \text{[A1(vii)]}.$$

Hence $\operatorname{tr}\mathbf{WD}_3 < \lambda \operatorname{tr}\mathbf{WD}_1 + (1-\lambda)\operatorname{tr}\mathbf{WD}_2$ contradicting the above supposition and proving the lemma.

THEOREM 3. *If* $\mathbf{C}$ *is positive definite* [see (21)], $\mathbf{W} = \mathbf{I}_I$ *and (in the notation of* **A1**(i)) $\mathbf{T} = t^2\mathbf{Z}(1, \rho)$ *where* $(-1/(I-1) < \rho < 1)$, *the (unique) optimum continuous allocation for given block sizes* $\{s_q\}$ *and control allocation* $\{a_{0q}\}$ *is* $\{a_{iq} = (s_q - a_{0q})/I\}$, *i.e., that part of each block not allocated to control is equally divided among the* I *treatments.*

PROOF. Let $\{a_{iq1}\}$ be the optimum continuous allocation and *suppose* that $\{a_{1q1} = a_{2q1} = \cdots = a_{Iq1}\}$ is not true. Take a permutation $\alpha(\cdot)$ on I elements such that $\{a_{iq2} \equiv a_{\alpha(i)q1}\}$ is different from $\{a_{iq1}\}$. By symmetry of the treatments over all the blocks w.r.t. $\operatorname{tr}\mathbf{D}$ [(15)], continuous allocation $\{a_{iq2}\}$ is also optimum and Lemma 1 is contradicted. The above supposition must therefore be rejected proving the lemma.

4.4. In this subsection the minimization (19) is completed, i.e., the outer minimization of (19) is effected.

For the remainder of this paper the following conditions are taken:

(27) $\mathbf{E}$ is of the form (15), $\mathbf{B}+\hat{\mathbf{E}}$ is positive definite, $\mathbf{T}$ is of the form specified in Theorem 3 and $\mathbf{W}$ is $\mathbf{I}_I$.

Theorem 3 shows that for given block sizes $\{s_q\}$ and allocation of control, $\{a_{0q}\}$, the optimum continuous allocation of the other treatments is

$$\{a_{1q} = a_{2q} = \cdots = a_{Iq} = (s_q - a_{0q})/I \equiv x_q, \text{ say}\}. \tag{28}$$

Let $\mathbf{D}(\mathbf{x})$ denote $\mathbf{D}(\{a_{iq} = x_q\})$, $\lambda_1(\mathbf{x}) = \mathbf{p}'\mathbf{x} + \delta$ and $\lambda_2(\mathbf{x}) = \mathbf{p}'x - I\mathbf{x}'\mathbf{C}\mathbf{x} + \gamma$ where $\mathbf{x} = (x_1, \cdots, x_Q)'$, $\mathbf{p} = (e_1^{-1}, \cdots, e_Q^{-1})'$, $\gamma = [1+(I-1)\rho]^{-1}t^{-2}$ and $\delta = (1-\rho)^{-1}t^{-2}$. By Theorem 2 and **A1**(i).

(29) $[\mathbf{D}(\mathbf{x})]^{-1}$ is *positive* definite for $\{0 < x_q < s_q/I\}$ (even if $\mathbf{B}^{-1}$ and $\mathbf{T}^{-1}$ are zero matrices) has $(I-1)$ fold root $\lambda_1(\mathbf{x})$ and simple root $\lambda_2(\mathbf{x})$. Hence

$$\operatorname{tr}\mathbf{D}(\mathbf{x}) = (I-1)/\lambda_1(\mathbf{x}) + 1/\lambda_2(\mathbf{x}). \tag{30}$$

To find the complete solution of the optimum continuous allocation of treatments for given block sizes it is required to find the optimum continuous allocation of control treatment when (28) holds and this is equivalent to finding $\mathbf{x}$ over the region $\{0 \leqq x_q \leqq s_q/I\}$, R say, which minimizes $\operatorname{tr}\mathbf{D}(\mathbf{x})$.

Let U_π denote the hyperplane $\mathbf{p}'\mathbf{x} = \pi$ and $\mathbf{s}$ denote $(s_1/I, \cdots, s_Q/I)'$, the minimization of $\operatorname{tr}\mathbf{D}(\mathbf{x})$ is split according to

$$\min_{\mathbf{x}\in R} \operatorname{tr}\mathbf{D}(\mathbf{x}) = \min_{0 \leqq \pi \leqq \mathbf{p}'\mathbf{s}} \min_{R \cap U_\pi} \operatorname{tr}\mathbf{D}(\mathbf{x}). \tag{31}$$

(32) Notice that, because of (29), the inner minimum of (31) occurs where $\mathbf{x}'\mathbf{C}\mathbf{x}$ is least over $R \cap U_\pi$.

(33) It is necessary to introduce some further notation. Let Ω denote the set of integers $1, 2, \cdots, Q$, Λ a subset of Ω and $\bar{\Lambda}$ the complement of Λ. For a $Q \times 1$ vector $\mathbf{a}$, let $\mathbf{a}_\Lambda$ denote the column vector consisting of elements $(\mathbf{a})_q$ for $q \in \Lambda$ in the same order they occur in $\mathbf{a}$. Similarly, for subsets Λ_1, Λ_2 of Ω and a $Q \times Q$ matrix $\mathbf{A}$, let $\mathbf{A}_{\Lambda_1,\Lambda_2}$ denote the matrix which remains if the rows of $\mathbf{A}$ corresponding to $\bar{\Lambda}_1$ and the columns corresponding to $\bar{\Lambda}_2$ are deleted. By R_Λ and $U_{\pi\Lambda}$ are meant the subsets of R and U_π respectively where $\mathbf{x}_{\bar{\Lambda}}$ is fixed at $\mathbf{s}_{\bar{\Lambda}}$ but the rest of $\mathbf{x}$ is left free to vary. Λ may be used just as a suffix for a scalar quantity.

Before returning to the main argument it is most economical to derive here two subsidiary results for a general subset Λ of Ω.

By the method of Lagrange Multipliers there is only one local minimum point of $\mathbf{x}'\mathbf{C}\mathbf{x}$ over $U_{\pi\Lambda}$ and it is the global minimum point because $\mathbf{x}'\mathbf{C}\mathbf{x}$ is a convex function [**A2**]. This global minimum occurs at

$$\mathbf{x}_\Lambda \equiv \phi(\Lambda)\pi + \psi(\Lambda) \tag{34}$$

where $\phi(\Lambda) = \mathbf{C}_{\Lambda,\Lambda}^{-1}\mathbf{p}_\Lambda / \mathbf{p}_\Lambda'\mathbf{C}_{\Lambda,\Lambda}^{-1}\mathbf{p}_\Lambda$ and

$$\psi(\Lambda) = -\mathbf{C}_{\Lambda,\Lambda}^{-1}\left[\frac{\mathbf{p}_{\bar{\Lambda}}'\mathbf{s}_{\bar{\Lambda}} - \mathbf{p}_\Lambda'\mathbf{C}_{\Lambda,\Lambda}^{-1}\mathbf{C}_{\Lambda,\bar{\Lambda}}\mathbf{s}_{\bar{\Lambda}}}{\mathbf{p}_\Lambda'\mathbf{C}_{\Lambda,\Lambda}^{-1}\mathbf{p}_\Lambda}\mathbf{p}_\Lambda + \mathbf{C}_{\Lambda,\bar{\Lambda}}\mathbf{s}_{\bar{\Lambda}}\right].$$

(35) Define $\mathbf{x}(\pi, \Lambda)$ by $\mathbf{x}(\pi, \Lambda)_\Lambda = \phi(\Lambda)\pi + \psi(\Lambda)$ and $\mathbf{x}(\pi, \Lambda)_{\bar{\Lambda}} = \mathbf{s}_{\bar{\Lambda}}$ so $\mathbf{x}(\pi, \Lambda)$ is the above global minimum point in Q space of $\mathbf{x}'\mathbf{C}\mathbf{x}$ over $U_{\pi\Lambda}$.

Let $\phi(\Lambda)_q$ and $\psi(\Lambda)_q$ for $q \in \Lambda$ be the components of $\phi(\Lambda)$ and $\psi(\Lambda)$ respectively corresponding to $\mathbf{x}(\pi, \Lambda)_q$ so $\mathbf{x}(\pi, \Lambda)_q = \phi(\Lambda)_q\pi + \psi(\Lambda)_q$ for $q \in \Lambda$.

Consider now $\operatorname{tr}\mathbf{D}(\mathbf{x}(\pi, \Lambda))$ regarded as a function of π. Clearly $\lambda_1(\mathbf{x}(\pi, \Lambda)) = \pi + \delta$. Let

$$\begin{aligned} &u_\Lambda \equiv \mathbf{p}_\Lambda'\mathbf{C}_{\Lambda,\Lambda}^{-1}\mathbf{p}_\Lambda / I, \qquad v_\Lambda \equiv I\mathbf{s}_{\bar{\Lambda}}'[(\mathbf{C}^{-1})_{\bar{\Lambda},\bar{\Lambda}}]^{-1}\mathbf{s}_{\bar{\Lambda}}, \\ &w_\Lambda \equiv \mathbf{p}_{\bar{\Lambda}}'\mathbf{s}_{\bar{\Lambda}} - \mathbf{p}_\Lambda'\mathbf{C}_{\Lambda,\Lambda}^{-1}\mathbf{C}_{\Lambda,\bar{\Lambda}}\mathbf{s}_{\bar{\Lambda}}, \qquad a_\Lambda = 1/u_\Lambda, \\ &b_\Lambda = 1 + 2w_\Lambda/u_\Lambda \quad \text{and} \quad c_\Lambda = \gamma - v_\Lambda - w_\Lambda^2/u_\Lambda \qquad \text{then} \\ &\lambda_2(\mathbf{x}(\pi, \Lambda)) = -a_\Lambda\pi^2 + b_\Lambda\pi + c_\Lambda. \end{aligned} \tag{36}$$

Only subregions R_Λ of R_Ω such that $\mathbf{x}(\pi, \Lambda) \in R_\Lambda$ for some value of π are of interest. For such a subregion R_Λ and suitable value of π, $\lambda_2(\mathbf{x}(\pi, \Lambda)) > 0$ and so

$$\lambda_2(\mathbf{x}(\pi, \Lambda)) = 0 \tag{37}$$

must have two real roots, $\pi_{1,\Lambda} < \pi_{2,\Lambda}$, say.

(38) For such a subregion R_Λ and for $\pi \in (\max(-\delta, \pi_{1,\Lambda}), \pi_{2,\Lambda})$ $\operatorname{tr}\mathbf{D}(\mathbf{x}(\pi, \Lambda))$ is a positive convex function of π [**A2**] which tends to ∞ as π approaches either end of this interval from inside.

(39) Define $f_\Lambda(\pi) = (2a_\Lambda\pi - b_\Lambda)(\pi + \delta)^2$ and $g_\Lambda(\pi) = (I-1)(-a_\Lambda\pi^2 + b_\Lambda\pi + c_\Lambda)^2$. $\partial/\partial\pi \operatorname{tr}\mathbf{D}(\mathbf{x}(\pi, \Lambda)) = 0$ iff $f_\Lambda(\pi) = g_\Lambda(\pi)$ and $\pi \neq -\delta, \pi_{1,\Lambda}$ or $\pi_{2,\Lambda}$. $f_\Lambda(\pi)$ is continuous

everywhere in π, vanishes at $-\delta$ and $b_\Lambda/2a_\Lambda$ and has a turning point at $-\delta$ and $\frac{1}{3}(b_\Lambda/a_\Lambda-\delta)$. Now $f_\Lambda''(-\delta)=f_\Lambda''(\frac{1}{3}(b_\Lambda/a_\Lambda-\delta))=2(2a_\Lambda\delta+b_\Lambda)$, so the stationary point at the higher value of π is a minimum and the one at the lower value of π is a maximum. The two possibilities for the graph are:

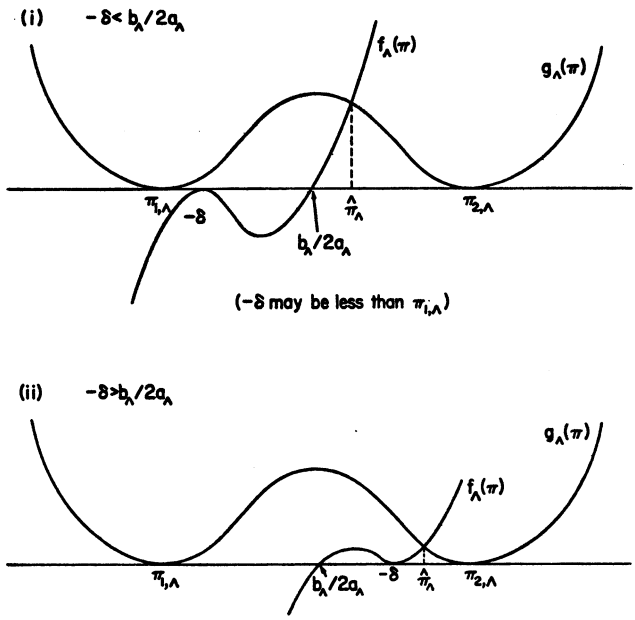

FIG. 1

In either case there is one root of $f_\Lambda(\pi)=g_\Lambda(\pi)$ between $\max(-\delta, b_\Lambda/2a_\Lambda)$ and $\pi_{2,\Lambda}$, $\hat{\pi}_\Lambda$ say. Notice that $b_\Omega(=1)>0$ and so for $\Lambda=\Omega$ graph (i) must obtain and

$$\hat{\pi}_\Omega > 0. \tag{40}$$

(41) It is concluded therefore that $\operatorname{tr}\mathbf{D}(\mathbf{x}(\pi,\Lambda))$ is positive and convex over $\pi\in(\max(-\delta,\pi_{1,\Lambda}),\pi_{2,\Lambda})$, tends to infinity as π tends to either end point from within and is minimum at $\hat{\pi}_\Lambda$ which is the unique root of $f_\Lambda(\pi)=g_\Lambda(\pi)$ between $\max(-\delta, b_\Lambda/2a_\Lambda)$ and $\pi_{2,\Lambda}$.

(42) The condition $\mathbf{C}^{-1}\mathbf{p}>0$ is required; this is a weak condition which holds if, for example, $\mathbf{C}$ is diagonal and by continuity if $\mathbf{C}$ is near enough to being diagonal. The condition therefore holds if $\{s_q/e_q$ is large enough$\}$ or if $(\mathbf{B}+\hat{\mathbf{E}})$ is near enough to being diagonal.

Returning to the main theme of the argument at (32), let $\mathbf{x}(\pi)$ denote the point where $\operatorname{tr}\mathbf{D}(\mathbf{x})$ is least over $R\cap U_\pi$ ($\pi\in[0,\mathbf{p}'\mathbf{s}]$). By (32) and (35)

(43) $\mathbf{x}(\pi)=\mathbf{x}(\pi,\Omega)$ provided $\mathbf{x}(\pi,\Omega)\in R$ which is so, referring to (42), iff $\pi\in[0,\pi_Q]$ where $\pi_\Omega=\min_{q\in\Omega}(s_q/I)/(\phi_\Omega)_q$.

(44) Denote the set of integers in Ω where this minimum occurs by $\overline{\Lambda}_1$.

In the notation of (37)

(45) $\lambda_2(\mathbf{x}(\pi,\Omega))$ is only positive when $\pi\in(\pi_{1,\Omega},\pi_{2,\Omega})$. For each $\pi\in[0,\mathbf{p's}]$ there is a point in $R\cap U_\pi$ for which $\mathbf{x'Cx}$ is not smaller than it is at $\mathbf{x}(\pi,\Omega)$ and since by (29) $\lambda_2(\mathbf{x})$ is positive for each point of R it follows that

(46) $\lambda_2(\mathbf{x}(\pi,\Omega))$ is positive when $\pi\in[0,\mathbf{p's}]$.

From (45) and (46),

(47) $$\pi_{1,\Omega}<0<\mathbf{p's}<\pi_{2,\Omega}.$$

(48) If $\hat{\pi}_\Omega\leqq\pi_\Omega$ [c.f. graph (i), (43) and (40)] then $\mathbf{x}(\hat{\pi}_\Omega,\Omega)$ is in R and it must be the required optimum point because if one takes any other point $\mathbf{x}\in R$ and denotes the value of $\mathbf{p'x}$ by π then by (35) and (41) $\operatorname{tr}\mathbf{D}(\mathbf{x})\geqq\operatorname{tr}\mathbf{D}(\mathbf{x}(\pi,\Omega))\geqq\operatorname{tr}\mathbf{D}(\mathbf{x}(\hat{\pi}_\Omega,\Omega))$.

If $\hat{\pi}_\Omega>\pi_\Omega$ a generalization of the condition (42) is required, namely that for all subsets Λ of Ω

(49) $$\mathbf{C}_{\Lambda,\Lambda}^{-1}\mathbf{p}_\Lambda>\mathbf{0}.$$

Because $\operatorname{tr}\mathbf{D}$ is a continuous and strongly convex function over R which is a convex region, $\mathbf{x}(\pi)$ is a continuous curve. It is known [see (43), (35) and (42)] that for $\pi\in[0,\pi_\Omega]$ the components of $\mathbf{x}(\pi)$ are strictly increasing functions of π and it is required to show that each component of $\mathbf{x}(\pi)$ continues to increase as π increases to $\mathbf{p's}$ until it reaches its maximum value (s_q/I for the qth component). Suppose that as π increases through π_1, say, a component of $\mathbf{x}(\pi)$ stops increasing but is not at its maximum value. Since for some subset Λ of Ω, $\mathbf{x}(\pi_1)$ is an internal point of $R_\Lambda\cap U_{\pi_1}$ (none of its components being zero), $\mathbf{x}(\pi_1)$ must be given by (35), viz. $\mathbf{x}(\pi_1,\Lambda)$ and the contradiction is established referring to (34) by condition (49). A component which has reached its maximum value is clearly precluded from decreasing by the above argument. It is therefore established that

(50) the components of $\mathbf{x}(\pi)$ are nondecreasing in π and those components which have not attained their maximum value (s_q/I for the qth component) are strictly increasing in π.

Because of the above property of $\mathbf{x}(\pi)$ and since $\mathbf{x}(\pi_\Omega)\in R_{\Lambda_1}$ [(43),(44)] then $\mathbf{x}(\pi)\in R_{\Lambda_1}$ for $\pi\geqq\pi_\Omega$. Since $\hat{\pi}_\Omega>\pi_\Omega$ and $\operatorname{tr}\mathbf{D}(\mathbf{x}(\pi,\Omega))$ is a convex function of π, $\mathbf{x}(\pi_\Omega,\Omega)$ is the best point of $R\cap(U_\pi\,|\,0\leqq\pi\leqq\pi_\Omega)$ hence

(51) If $\hat{\pi}_\Omega>\pi_\Omega$ the required optimum point must be in $R_{\Lambda_1}\cap(U_\pi\,|\,\pi\in[\pi_\Omega,\mathbf{p's}])$.

Let π_{Λ_1} denote $\min_{q\in\Lambda_1}[s_q/I-\psi(\Lambda_1)_q]/\phi(\Lambda_1)_q$. Referring to (35), $\mathbf{x}(\pi,\Lambda_1)\in R_{\Lambda_1}$ for $\pi\in[\pi_\Omega,\pi_{\Lambda_1}]$ (and possibly for smaller values of π too) so

(52) $\mathbf{x}(\pi)=\mathbf{x}(\pi,\Lambda_1)$ for $\pi\in[\pi_\Omega,\pi_{\Lambda_1}]$ where π_{Λ_1} is defined above.

In the same way that (47) was derived it follows that

(53) $$\pi_{1,\Lambda_1}<\pi_\Omega\leqq\mathbf{p's}<\pi_{2,\Lambda_1}.$$

It cannot be that $\hat{\pi}_{\Lambda_1} < \pi_\Omega$ for if it were so, $\mathbf{x}(\hat{\pi}_{\Lambda_1}, \Lambda_1)$ would be better than $\mathbf{x}(\pi_\Omega, \Lambda_1)(=\mathbf{x}(\pi_\Omega, \Omega))$ and worse than $\mathbf{x}(\hat{\pi}_{\Lambda_1}, \Omega)$ which in turn is worse than $\mathbf{x}(\pi_\Omega, \Omega)$, a contradiction, so

$$\hat{\pi}_{\Lambda_1} \geqq \pi_\Omega. \tag{54}$$

If $\hat{\pi}_{\Lambda_1} < \pi_{\Lambda_1}$ then referring to (51) and by the same argument which derived (48)

(55) $\mathbf{x}(\hat{\pi}_{\Lambda_1}, \Lambda_1)$ is the optimum point of R. If $\hat{\pi}_{\Lambda_1} > \pi_{\Lambda_1}$ let Λ_2 denote the integers of Λ_1 which do not minimize $[s_q/I - \psi(\Lambda_1)_q]/\phi(\Lambda_1)_q$ over $q \in \Lambda_1$. The optimum point is in $R_{\Lambda_2} \cap (U_\pi \,|\, \pi \in [\pi_{\Lambda_1}, \mathbf{p}'\mathbf{s}])$ and the procedure continues. The procedure does of course terminate in at most Q steps.

Because of (29) the preceding argument goes through in the case $\mathbf{T}^{-1}$ and $\mathbf{B}^{-1}$ are zero matrices where $(\mathbf{B}+\hat{\mathbf{E}})^{-1}$ is defined to be $\mathbf{0}$ when $\mathbf{B}^{-1} = \mathbf{0}$.

ALGORITHM 1. Let $\mathbf{E}$, the error dispersion matrix for the sample (be positive definite) and take the form of (15), and let $\mathbf{B}+\hat{\mathbf{E}}$ be positive definite (a weak condition). Let $\mathbf{T}$ (the prior dispersion matrix for the treatment parameters) take the form $t^2\mathbf{Z}(1, \rho)$ (in the notation of **A1**(i)) where ρ (the prior correlation between any two treatment parameters) is between $-1/(I-1)$ and 1. Let $\mathbf{W}$ (see 4.1) be the identity matrix so the loss function is specialized to the sum of squared errors of estimation.

The optimum continuous allocation [4.3] of treatments over the blocks of given sizes $\{s_q\}$ is $\{\hat{a}_{1q} = \hat{a}_{2q} = \cdots = \hat{a}_{Iq} = \hat{x}_q\,(\hat{a}_{0q} = s_q - I\hat{x}_q)\}$ where $(\hat{x}_1, \cdots, \hat{x}_Q)' \equiv \hat{\mathbf{x}}$ is found by the procedure described below.

First some definitions. Let $\mathbf{p} = (e_1^{-1}, \cdots, e_Q^{-1})'$, $\mathbf{s} = (s_1/I, \cdots, s_Q/I)'$, $\gamma = [1+(I-1)\rho]^{-1}t^{-2}$, $\delta = (1-\rho)^{-1}t^{-2}$(*1) and $\mathbf{C}^{-1} = \mathbf{dg}(e_1 s_1, \cdots, e_Q s_Q) + \mathbf{dg}(e_1, \cdots, e_Q)(\mathbf{B}+\hat{\mathbf{E}})^{-1}\mathbf{dg}(e_1, \cdots, e_Q)$. In the partitioning notation of (33) let Ω denote the set of integers $(1, \cdots, Q)$, Λ denote a subset of Ω and $\bar{\Lambda}$ denote the complement of Λ w.r.t. Ω. Let $u_\Lambda = \mathbf{p}_\Lambda' \mathbf{C}_{\Lambda,\Lambda}^{-1}\mathbf{p}_\Lambda/I, v_\Lambda = I\mathbf{s}_{\bar{\Lambda}}'[(\mathbf{C}^{-1})_{\bar{\Lambda},\bar{\Lambda}}]^{-1}\mathbf{s}_{\bar{\Lambda}}$, $w_\Lambda = \mathbf{p}_{\bar{\Lambda}}'\mathbf{s}_{\bar{\Lambda}} - \mathbf{p}_\Lambda'\mathbf{C}_{\Lambda,\Lambda}^{-1}\mathbf{C}_{\Lambda,\bar{\Lambda}}\mathbf{s}_{\bar{\Lambda}}$, $\phi(\Lambda) = \mathbf{C}_{\Lambda,\Lambda}^{-1}\mathbf{p}_\Lambda/Iu_\Lambda$, $\psi(\Lambda) = -\mathbf{C}_{\Lambda,\Lambda}^{-1}[w_\Lambda/(Iu_\Lambda)\mathbf{p}_\Lambda + \mathbf{C}_{\Lambda,\bar{\Lambda}}\mathbf{s}_{\bar{\Lambda}}]$ and let $\mathbf{x}(\pi, \Lambda)$ be defined by $\mathbf{x}_{\bar{\Lambda}}(\pi, \Lambda) = \mathbf{s}_{\bar{\Lambda}}$ and $\mathbf{x}_\Lambda(\pi, \Lambda) = \phi(\Lambda)\pi + \psi(\Lambda)$. Let $\phi(\Lambda)_q$ and $\psi(\Lambda)_q$ be the components of $\phi(\Lambda)$ and $\psi(\Lambda)$ respectively corresponding to $\mathbf{x}(\pi, \Lambda)_q$. Let π_Λ denote $\min_{q\in\Lambda}[s_q/I - \psi(\Lambda)_q]/\phi(\Lambda)_q$. (*2) Let $a_\Lambda = 1/u_\Lambda$, $b_\Lambda = 1 + 2w_\Lambda/u_\Lambda$ and $c_\Lambda = \gamma - v_\Lambda - w_\Lambda^2/u_\Lambda$. Let $\pi_{2,\Lambda} = (b_\Lambda + [b_\Lambda^2 + 4a_\Lambda c_\Lambda]^{\frac{1}{2}})/2a_\Lambda$, if $\pi_{2,\Lambda}$ is real let $\hat{\pi}_\Lambda$ denote the unique root between $\max(-\delta, b_\Lambda/2a_\Lambda)$ and $\pi_{2,\Lambda}$ of the quartic equation in π:

$$(2a_\Lambda\pi - b_\Lambda)(\pi+\delta)^2 = (I-1)(-a_\Lambda\pi^2 + b_\Lambda\pi + c_\Lambda)^2.$$

This root is indicated graphically above. Notice that in the case $I = 1$ (one treatment and the control treatment), $\hat{\pi}_\Lambda = b_\Lambda/2a_\Lambda$.

The first step in the procedure is to evaluate $\hat{\pi}_\Omega$, the root of the quartic equation in π

$$(2u_\Omega^{-1}\pi - 1)(\pi+\delta)^2 = (I-1)(-u_\Omega^{-1}\pi^2 + \pi + \gamma)^2$$

which lies between $\frac{1}{2}u_\Omega$ and $\frac{1}{2}(1+[1+4u_\Omega^{-1}\gamma]^{\frac{1}{2}})u_\Omega$ and to evaluate (*3) $\pi_\Omega \equiv \min_{q\in\Omega} s_q/I(\phi_\Omega)_q$. If $\mathbf{C}^{-1}\mathbf{p} > 0$ and $\hat{\pi}_\Omega \leqq \pi_\Omega$ the required optimum point is

$$\hat{\mathbf{x}} = (\mathbf{C}^{-1}\mathbf{p})/(\mathbf{p}'\mathbf{C}^{-1}\mathbf{p})\hat{\pi}_\Omega.$$

If $\hat{\pi}_\Omega > \pi_\Omega$ the condition $\mathbf{C}_{\Lambda,\Lambda}^{-1}\mathbf{p}_\Lambda > \mathbf{0}$ for all subsets Λ of Ω is required for the following procedure. (This condition holds if, for example, $\{s_q/e_q$ is large enough$\}$ or $\mathbf{B}+\hat{\mathbf{E}}$ is near enough to being diagonal.) Denote by $\bar{\Lambda}_1$ the set of values $q\in\Omega$ for which $s_q/\phi(\Omega)_q$ is least. Find π_{Λ_1} and $\hat{\pi}_{\Lambda_1}$ (π_{2,Λ_1} is real), if $\hat{\pi}_{\Lambda_1} \leqq \pi_{\Lambda_1}$ the optimum point is $\mathbf{x}(\hat{\pi}_{\Lambda_1}, \Lambda_1)$. If $\hat{\pi}_{\Lambda_1} > \pi_{\Lambda_1}$, denote by Λ_2 the set of values $q\in\Lambda_1$ for which $[s_q/I-\psi(\Lambda_1)_q]/\phi(\Lambda_1)_q$ is not least. Now evaluate π_{Λ_2} and $\hat{\pi}_{\Lambda_2}$ (π_{2,Λ_2} is real) and the procedure continues.

(*4) The procedure terminates in at most Q steps.

Having found the optimum continuous allocation, an integer allocation is taken near to it and if this is not the optimum integer allocation it will usually be nearly optimum (see (20)). To check how good his integer allocation is, the experimenter may compare its associated expected loss with the expected loss using the optimum continuous allocation. If the "nearby" integer allocation is taken to be a "balanced" allocation, i.e., of the form $\{a_{iq} = (s_q - a_{0q})/I\}$, (30) may be used to calculate both expected losses.

5. Special cases. In various special cases Algorithm 1 simplifies. In the first case to be taken a simpler procedure obtains and in the other cases there is an explicit expression for the optimum continuous allocation.

The first case to be considered is as follows.

(56) $\mathbf{E} = \mathbf{dg}(e_1\mathbf{I}_{s_1}, \cdots, e_Q\mathbf{I}_{s_Q})$ (so $\hat{\mathbf{E}} = \mathbf{0}$ and the errors are uncorrelated) and $\mathbf{B} = \mathbf{dg}(b_{11}, \cdots, b_{QQ})$ (block parameters are a priori uncorrelated).

In this case and the notation of Algorithm 1, the following simplified formulae obtain.

(57) $$\mathbf{C} = \mathbf{dg}(c_1, \cdots, c_Q) \quad \text{where} \quad \{c_q = b_{qq}/e_q(e_q + s_q b_{qq})\}.$$

Because $\mathbf{C}$ is diagonal and $\{c_q > 0\}$ the restriction (49) is trivially satisfied. (56) of course implies that $\mathbf{B}+\hat{\mathbf{E}}$ is positive definite. Then in the notation of Algorithm 1,

$$u_\Lambda = I^{-1}\textstyle\sum_{q\in\Lambda} e_q^{-2}c_q^{-1}, \qquad v_\Lambda = I^{-1}\sum_{q\in\bar{\Lambda}} s_q^2 c_q,$$
$$w_\Lambda = I^{-1}\textstyle\sum_{q\in\bar{\Lambda}} e_q^{-1}s_q, \qquad \pi_\Lambda = u_\Lambda \min_{q\in\Lambda} s_q e_q c_q + w_\Lambda,$$
$$\mathbf{x}_{\bar{\Lambda}}(\pi, \Lambda) = \mathbf{s}_{\bar{\Lambda}}, \qquad [\mathbf{x}_\Lambda(\pi, \Lambda)]_q = [(\pi - w_\Lambda)/Iu_\Lambda]\, e_q^{-1}c_q^{-1}$$

for $q\in\Lambda$. The nested subsets of $\Omega, \Lambda_1 \supset \Lambda_2 \supset \cdots$ are easily defined in this case: compute the Q numbers $s_1 e_1 c_1, \cdots, s_Q e_Q c_Q$, strike out from Ω the (one or more) integers corresponding to the smallest value of these numbers. The set of integers which remains is denoted by Λ_1. Strike out from Λ_1 the (one or more) integers corresponding to the next smallest value of these numbers, the set which remains is denoted by Λ_2, and so on.

There follows the simplified version of Algorithm 1:

ALGORITHM 1A. With the specialization (56) read to (*1) of Algorithm 1 and then refer to definitions (57). Now read from (*2) to (*3) and then: $\pi_\Omega = \min_{q\in\Omega} s_q e_q c_q$. If $\pi_\Omega \leqq \hat{\pi}_\Omega$ the required optimum point is

$$\{(\hat{\mathbf{x}})_q = (\hat{\pi}_\Omega / I u_\Omega) c_q^{-1} e_q^{-1}\}.$$

If $\hat{\pi}_\Omega > \pi_\Omega$ find π_{Λ_1} and $\hat{\pi}_{\Lambda_1}$ (π_{2,Λ_1} is real), if $\hat{\pi}_{\Lambda_1} \leqq \pi_{\Lambda_1}$ the optimum point is $\mathbf{x}(\hat{\pi}_{\Lambda_1}, \Lambda_1)$. If $\hat{\pi}_{\Lambda_1} > \pi_{\Lambda_1}$ evaluate π_{Λ_2} and $\hat{\pi}_{\Lambda_2}$ (π_{2,Λ_2} is real) and the procedure continues. The remarks after (*4) should now be read.

In the following theorem a simple analytic expression obtains for the optimum continuous allocation.

THEOREM 4. *In the setting described in the first paragraph of Algorithm* 1, *let*

(a) $\{0 \leqq s_q + e_q \sum_{r=1}^{Q} [(\mathbf{B}+\hat{\mathbf{E}})^{-1}]_{qr}$ *all* $q = 1, \cdots, Q\}$ (*a weak condition*). *Let* u *denote* $(1/I)\sum_{q=1}^{Q} s_q/e_q + (1/I)\sum_{q=1}^{Q}\sum_{r=1}^{Q} [(\mathbf{B}+\hat{\mathbf{E}})^{-1}]_{qr}$, δ *denote* $(1-\rho)^{-1}t^{-2}$, γ *denote* $[1+(I-1)\rho]^{-1}t^{-2}$ *and* $\hat{\lambda}$ *denote the unique root between* $\frac{1}{2}$ *and* $\frac{1}{2}(1+(1+4\gamma/u)^{\frac{1}{2}})$ *of the quartic equation in* λ: $(2\lambda-1)(\lambda+\delta u^{-1})^2 = (I-1)(-\lambda^2+\lambda+\gamma u^{-1})^2$. *Let* $\{x_q \equiv I^{-1}\hat{\lambda}(s_q + e_q \sum_{r=1}^{Q} [(\mathbf{B}+\hat{\mathbf{E}})^{-1}]_{qr})\}$, *then if*

(b) $\{x_q \leqq s_q/I$, *all* $q = 1, \cdots, Q\}$ *the optimum continuous allocation for fixed blocking* $\{s_q\}$ *is* $\{\hat{a}_{iq} = x_q\}$.

(Of course (a) and (b) together simply say that $\{0 \leqq x_q \leqq s_q/I\}$.) If $\{x_q \leqq s_q/I$, all $q = 1, \cdots, Q\}$ is not true, the experimenter must proceed to the second stage of Algorithm 1.

Define $(\mathbf{B}+\hat{\mathbf{E}})^{-1}$ to be $\mathbf{0}$ when $\mathbf{B}^{-1} = \mathbf{0}$ then $\{x_q\}$ is continuous w.r.t. $\mathbf{B}$, t and ρ even at $\mathbf{B}^{-1} = \mathbf{0}$ and $t^{-1} = 0$, i.e., $\{x_q\}$ is continuous w.r.t the prior information even at the case of vagueness.

Limiting cases. At $\mathbf{T}^{-1} = 0$ (vague prior information on the treatment parameters)

$$\left\{x_q = \frac{1}{(1+I^{\frac{1}{2}})I^{\frac{1}{2}}}\left(s_q + e_q \sum_{r=1}^{Q} [(\mathbf{B}+\hat{\mathbf{E}})^{-1}]_{qr}\right)\right\}.$$

If both $\mathbf{T}^{-1} = 0$ and $\mathbf{B}^{-1} = \mathbf{0}$ (vague prior information on all the parameters)

$$\left\{x_q = \frac{1}{(1+I^{\frac{1}{2}})I^{\frac{1}{2}}} s_q\right\} \quad \left(\text{notice} \quad \frac{\hat{a}_{0q}}{\hat{a}_{iq}} = I^{\frac{1}{2}}\right).$$

In the case $\mathbf{T}^{-1} = 0$, condition (b) may be written:

(b′) $$\{e_q \textstyle\sum_{r=1}^{Q} [(\mathbf{B}+\hat{\mathbf{E}})^{-1}]_{qr} \leqq s_q/I^{\frac{1}{2}} \quad \text{all} \quad q = 1, \cdots, Q\}.$$

If $\mathbf{B}^{-1}$ is near enough to $\mathbf{0}$ conditions (a) and (b′) are both satisfied.

Notice that (a) holds for example if $\hat{\mathbf{E}} = \mathbf{0}$ (uncorrelated errors) and $\mathbf{B}$ is diagonal (a priori uncorrelated block parameters). Notice too that for a given error distribution and prior distribution of the block parameters, (a), (b) and (b′) hold if the block sizes ($\{s_q\}$) are large enough.

PROOF. For the first part refer to Algorithm 1.

For the limiting case of vague prior information, see the last sentence of the proof of Algorithm 1.

For the continuity of $\{x_q\}$ w.r.t. $\mathbf{B}$, t and ρ, notice $(\mathbf{B}+\hat{\mathbf{E}})^{-1}$ is continuous w.r.t. $\mathbf{B}$ even at $\mathbf{B}^{-1}=\mathbf{0}$. The coefficient of $\hat{\lambda}$ in the expression for $\{x_q\}$ therefore has the property of continuity and, referring to (41), $\hat{\lambda}$ is a simple finite root of the quartic equation above and is therefore a continuous function of the coefficients of the quartic which in turn are continuous functions of $\mathbf{B}$, t and ρ even at $\mathbf{B}^{-1}=\mathbf{0}$ and $t^{-1}=0$.

Finally note that in the case $t^{-1}=0$, $I^{\frac{1}{2}}/(1+I^{\frac{1}{2}})$ is the unique root between $\frac{1}{2}$ and 1 of the above quartic and $\hat{\lambda}\to I^{\frac{1}{2}}/(1+I^{\frac{1}{2}})$ as one or more of the block sizes tends to infinity.

From Theorem 4 and Corollary 1 there follows:

COROLLARY 4A. *In the setting described in the first paragraph of Algorithm* 1, *as* $\mathbf{B}^{-1}\to\mathbf{0}$ (*through positive definite values*) *and* $t^{-1}\to 0$ (*i.e., as the prior information tends to vagueness*) *the optimum continuous allocation for fixed blocking* $\{s_q\}$:

$$\{\hat{a}_{iq}\to(\textit{and}=\textit{when}\quad \mathbf{B}^{-1}=\mathbf{0}\quad \textit{and}\quad \mathbf{T}^{-1}=\mathbf{0})[(1+I^{\frac{1}{2}})I^{\frac{1}{2}}]^{-1}s_q\}.$$

This continuous allocation also minimizes tr Var $\hat{\boldsymbol{\tau}}$ (where $\hat{\boldsymbol{\tau}}$ is the maximum likelihood estimator of $\hat{\boldsymbol{\tau}}$) and is the limiting optimum continuous allocation as $\{s_q\to\infty\}$ for fixed prior information.

The following result is established from Theorem 4 and (30).

COROLLARY 4B. *In the case of Theorem* 4 *let conditions* (a) *and* (b) *obtain. The expected posterior loss under optimum continuous allocation is*

$$u^{-1}\left(\frac{I-1}{\hat{\lambda}+u^{-1}\delta}+\frac{1}{\hat{\lambda}-\hat{\lambda}^2+u^{-1}\gamma}\right)$$

which reduces to

$$\frac{I(1+I^{\frac{1}{2}})^2}{\sum_{q=1}^{Q}s_q e_q^{-1}+\sum_{q=1}^{Q}\sum_{r=1}^{Q}[(\mathbf{B}+\hat{\mathbf{E}})^{-1}]_{qr}}$$

in the case $t^{-1}=0$.

Note on convergence of the posterior distribution.

One would certainly expect that the posterior distribution of $\boldsymbol{\tau}$ under optimum allocation tends to a point distribution as one or more of the block sizes tend to infinity and this may be verified by reference to Corollary 4B.

THEOREM 5. *In the setting described in the first paragraph of Algorithm* 1, *take the case*

$$\{(e_q/s_q)\textstyle\sum_{r=1}^{Q}[(\mathbf{B}+\hat{\mathbf{E}})^{-1}]_{qr}=h-1\}$$

where h does not depend on q. Let $\hat{\lambda}$ *be defined as in Theorem* 4 ($u=I^{-1}h\sum_{q=1}^{Q}s_q/e_q$).

The optimum continuous allocation for fixed blocking $\{s_q\}$ *is*

$$\{\hat{a}_{iq} = I^{-1}s_q \min[h\hat{\lambda}, 1]\}.$$

PROOF. This result is evident from the proof of Algorithm 1 on noticing that the above condition ensures that the locus of minima of $\operatorname{tr}\mathbf{D}$ over $R \cap U_\pi$ as π varies is the "hyperdiagonal" of R joining $\{0\}$ to $\{s_q/I\}$.

The remarks in the last paragraph of Algorithm 1 should be appended to the preceding three results.

In the case $I = 1$, $\mathbf{B}$ diagonal and $\mathbf{E}$ diagonal, $\operatorname{tr}\mathbf{D}$ takes such a simple form that Algorithm 1 need not be resorted to. In this case both the integer and continuous optimum allocations have simple expressions. Here

$$(\operatorname{tr}\mathbf{D})^{-1} = \sum_{q=1}^{Q} e_q^{-1}[a_{1q} - a_{1q}^2 b_{qq}/(e_q + s_q b_{qq})] + t^{-2}. \tag{58}$$

If, for $q = 1, \cdots, Q$, $\hat{a}_{1q}$ maximizes $a_{1q} - a_{1q}^2 b_{qq}/(e_q + s_q b_{qq})$, then each term of the above sum is maximized and $\hat{a}_{1q}$ is the optimum allocation. The next theorem follows from this observation and the symmetry of the parabola.

THEOREM 6. *In the case of one treatment and control (i.e.,* $I = 1$*),* $\mathbf{E} = \mathbf{dg}(e_1 \mathbf{I}_{s_1}, \cdots, e_Q \mathbf{I}_{s_Q})$ *and* $\mathbf{B} = \mathbf{dg}(b_{11}, \cdots, b_{QQ})$*, the optimum integer allocation is* $\{a_{1q} = \min[s_q,$ *nearest integer to* $\frac{1}{2}(s_q + e_q/b_{qq})]\}$ *and the optimum continuous allocation is* $a_{1q} = \{\min[s_q, \frac{1}{2}(s_q + e_q/b_{qq})]\}$*. Notice neither allocation depends on* t*, the prior variance for the treatment parameter.*

6. Example. Four blocks of sizes 100, 120, 130 and 140 experimental units, nine treatments and the control treatment. Normally distributed errors, within block error homoscedasticity with error variances 10, 20, 30 and 40 respectively and covariances expressed (in the notation of (15)) by

$$\hat{\mathbf{E}} = \begin{matrix} 1 & -0.2 & 0.2 & -0.1 \\ & 3 & 0 & 0.1 \\ & & 2 & 0.3 \\ & & & 5 \end{matrix}$$

where, for example, 1 is the covariance between (different) errors in the first block and -0.2 the covariance between errors in the first and second blocks. The prior distribution of the parameters is multivariate normal with a priori zero correlation between each block parameter and each treatment parameter. The prior variance of each of the treatment parameters is $\frac{1}{4}$ and the prior correlation coefficient between each pair of treatment parameters is 0.11. In the notation of (5) the prior dispersion matrix on the block parameters is

$$\mathbf{B} = 0.4 \begin{matrix} & -0.2 & 0.3 & 0.2 \\ & 0.6 & 0.1 & 0 \\ & & 0.8 & -0.1 \\ & & & 1.0. \end{matrix}$$

In the notation of this paper

$$Q = 4, \quad I = 9, \quad t = \tfrac{1}{2}, \quad \rho = 0.11,$$
$$s_1 = 100, \quad s_2 = 120, \quad s_3 = 130, \quad s_4 = 140,$$
$$e_1 = 10, \quad e_2 = 20, \quad e_3 = 30, \quad e_4 = 40.$$

Computing $\{x_q\}$ from Theorem 4 it is found that $u = 2.80660$, $\hat{\lambda} = 0.926851$ and hence $x_1 = 10.98$, $x_2 = 13.00$, $x_3 = 13.93$, $x_4 = 14.91$. Notice that $\{0 < x_q < s_q/I\}$ and so by Theorem 4 the optimum continuous allocation of treatments is $\{\hat{a}_{iq} = x_q\}$. The expected loss associated with this allocation is the least possible and by Corollary 4B is 1.5589 and the nearby integer allocation

$$\{a_{i1} = 11(a_{01} = 1), \quad a_{i2} = 13(a_{02} = 3), \quad a_{i3} = 14(a_{03} = 4), \quad a_{i4} = 15(a_{04} = 5)\}$$

is very nearly optimum (this may be checked using (30)).

If the prior information is ignored and the corresponding optimum allocation (which by Theorem 4 is the familiar $\{a_{0q}/a_{iq} = 3\}$) is taken the expected loss (by Corollary 4B) is 6.042, nearly four times the optimum expected loss.

APPENDIX

The following known results are used in the text.

A1 (i) Let $n \geqq 2$ and $Z_n(a, b)$ denote the $n \times n$ matrix with a's down the diagonal and b's elsewhere. If $a \neq b$ or $-(n-1)b$ it may be verified that $[\mathbf{Z}_n(a, b)]^{-1} = \mathbf{Z}_n(a', b')$ where,

$$a' = \frac{a+(n-2)b}{(a-b)[a+(n-1)b]} \quad \text{and} \quad b' = -\frac{b}{(a-b)[a+(n-1)b]}.$$

The characteristic roots of $Z_n(a, b)$ are: $n-1$ fold root $a-b$ and simple root $a+(n-1)b$. Notice that if $n = 1$ the result still holds if b is interpreted to be zero.

(ii) If $\mathbf{X}_1$ and $\mathbf{X}_2$ are positive definite and $0 < \lambda < 1$ then $\lambda \mathbf{X}_1^{-1} + (1-\lambda)\mathbf{X}_2^{-1} \geqq [\lambda \mathbf{X}_1 + (1-\lambda)\mathbf{X}_2]^{-1}$ with equality iff $\mathbf{X}_1 = \mathbf{X}_2$ (Kiefer, 1959).

(iii) If $\mathbf{X}_1 \geqq \mathbf{X}_2 > 0$ then $\mathbf{X}_2^{-1} \geqq \mathbf{X}_1^{-1} > \mathbf{0}$.

(iv) If $\mathbf{X}_1 \geqq \mathbf{X}_2$ and $\mathbf{Y}_1 \geqq \mathbf{Y}_2$ then $\mathbf{X}_1 + \mathbf{X}_2 \geqq \mathbf{Y}_1 + \mathbf{Y}_2$ and equality only obtains for the sums if it obtains for both pairs of matrices.

(v) Let $\mathbf{A}, \mathbf{B}$ be matrices of orders $m \times n$ and $n \times n$ respectively, then $\mathbf{ABA}'$ is nonnegative definite and if $\mathbf{B}$ is positive definite $\mathbf{ABA}'$ has the same rank as $\mathbf{A}$.

(vi) If $\mathbf{X}$ is a square, symmetric, positive definite matrix, there exists nonsingular $\mathbf{Y}$ such that $\mathbf{X} = \mathbf{YY}'$.

(vii) The product of a positive (nonnegative) definite matrix with a diagonal matrix of positive numbers is positive (nonnegative) definite respectively.

(viii) Let $\mathbf{S}^{rxr}$, $\mathbf{B}^{nxn}$ be nonsingular matrices and $\mathbf{U}, \mathbf{V}$ $n \times r$ matrices then provided the inverses exist: $(\mathbf{B}-\mathbf{USV}')^{-1} = \mathbf{B}^{-1} + \mathbf{B}^{-1}\mathbf{U}(\mathbf{S}^{-1} - \mathbf{V}'\mathbf{B}^{-1}\mathbf{U})^{-1}\mathbf{V}'\mathbf{B}^{-1}$.

A2 (i) A local minimum of a convex differentiable function over a closed convex set is a global minimum.

(ii) Hyperplanes and hyperrectangles are closed convex sets and intersection preserves closedness and convexity.

(iii) If $\mathbf{A}$ is positive definite, $\mathbf{x}'\mathbf{A}\mathbf{x}$ is a convex function over $\mathbf{x}$ space.

(iv) If $f(\mathbf{x})$ is a concave (convex) function over a convex set then $-f(\mathbf{x})$ is convex (concave) respectively over the set.

(v) If $f(\mathbf{x})$ is a concave and positive function over a convex set then $1/f(\mathbf{x})$ is convex over the set.

(vi) If $f_1(\mathbf{x})$, $f_2(\mathbf{x})$ are concave (convex) functions over a convex set then so (respectively) is $f_1(\mathbf{x})+f_2(\mathbf{x})$.

Acknowledgment. I should like to express my gratitude to Professor D. V. Lindley for suggesting this research topic and for his valuable criticisms and comments on an earlier version of this work.

I am also indebted to Mrs. Dorothy Thayer whose computer program reduced the labor of computation in the example.

REFERENCES

[1] BECHHOFER, R. E. (1968). Optimal allocation of observations when comparing several treatments with a control. Tech. Report No. 55, Department of Operations Research, Cornell Univ.

[2] DUNNETT, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *J. Amer. Statist. Assoc.* **50** 1096–1121.

[3] ERICSON, W. A. (1965). Optimum stratified sampling using prior information. *J. Amer. Statist. Assoc.* **60** 750–771.

[4] FIELLER, E. C. (1947). Some remarks on the statistical background in bio-assay. *Analyst.* **72** 37–43.

[5] KIEFER, J. (1959). Optimum experimental design. *J. Roy. Statist. Soc. Ser. B.* **21** 272–304.

[6] MARCUS, M. and MINC, H. (1964). *A Survey of Matrix Theory and Matrix Inequalities.* Allyn and Bacon, Boston.