

## A UNIFIED APPROACH TO COORDINATE-FREE MULTIVARIATE ANALYSIS

M. STONE

(Received May 21, 1973; revised Dec. 11, 1974)

### 1. Introduction and summary

By *coordinate-free* multivariate analysis, we mean that form in which the basic elements of analysis (items, variables) and derived elements, such as estimates, are treatable as points in vector spaces without explicit bases. Multivariate statistical procedures can then be described geometrically and can often be usefully represented in 2-dimensional diagrams. The advantages that may be claimed for the coordinate-free approach are:

- (i) It reveals the natural simplicity of operations that can appear rather forbidding in algebraic terms, especially when the heavy investment that mathematical education makes in matrix algebra is taken into account.
- (ii) The pedagogical use of matrices often lacks a sense of direction and may, with advantage, be deferred up to the point where actual numbers have to be fed into a computer for analysis.
- (iii) The geometrical approach can suggest new but obviously desirable lines of enquiry.

Following Dempster [2] we employ: (i) **variable space**,  $\mathcal{V}$ , a  $p$ -dimensional vector space of **variables**  $U, V, \dots$  over the real field, (ii) **item space**,  $\mathcal{X}$ , the dual vector space of linear functions (**items**,  $x, y, \dots$ ) on  $\mathcal{V}$  with the bilinear product  $[x, V]$ .

For example, with  $p=2$  and the directly measurable variables  $M_1 = \text{WEIGHT}$ ,  $M_2 = \text{HEIGHT}$ , we have  $\mathcal{V} = \left\{ \sum_{i=1}^2 \alpha_i M_i \mid \alpha \in R^2 \right\}$ . Three items in  $\mathcal{X}$  are  $^1x = \text{JIR}\bar{O}$ ,  $^2x = \text{ICHIR}\bar{O}$  and  $^3x = \text{TAR}\bar{O}$ . To state that  $[^1x, M_1] = 62$  kilograms is to state the measurement made for JIR $\bar{O}$ 's WEIGHT; while the hybrid item  $(1/3)\text{JIR}\bar{O} + (1/3)\text{ICHIR}\bar{O} + (1/3)\text{TAR}\bar{O}$  has an analytical utility.

The variable  $M_1 + M_2$ , for example, is not directly measurable but has, for item  $x$ , the implied measurement  $[x, M_1] + [x, M_2]$ . In general, it must be supposed that there are  $p$  directly measurable variables  $M_1, \dots, M_p$ .

$\dots, M_p$  spanning  $\mathcal{V}$  and that items are in 1-1 correspondence with the ordered set of measurements  $[x, M_1], \dots, [x, M_p]$ . The concept of error of measurement is therefore not applicable to items as we consider them in this paper; the "error" is built-in.

Section 2 introduces some useful coordinate-free notation and definitions associated with probability distributions and random samples in  $\mathcal{X}$ . Section 3 is concerned principally with the coordinate-free treatment of multivariate multiple regression, including the Gauss-Markov theory. Section 4 considers covariance adjustment in the estimation of the mean of a probability distribution on  $\mathcal{X}$ . Section 5 relates the present approach to that of other authors. The Appendix establishes some technical results (A1, A2,  $\dots$ ) that are needed.

## 2. Notation and definitions

The following table introduces some nearly standard bilinear operators in simplified notation:

Input	Operator	Defining equation
$x \in \mathcal{X}$	$x^2: \mathcal{V}^2 \rightarrow R$	$x^2(U, V) = [x, U][x, V]$
$x, y \in \mathcal{X}$	$xy: \mathcal{V}^2 \rightarrow R$	$xy(U, V) = [x, U][y, V]$
$\lambda \in R^n, \{({}^k x, {}^k y);$ $k=1, \dots, n\}$	$\sum \lambda_k {}^k x {}^k y: \mathcal{V}^2 \rightarrow R$	$(\sum \lambda_k {}^k x {}^k y)(U, V) = \sum \lambda_k [{}^k x, U][{}^k y, V]$
$\nu$ a measure on $\mathcal{X}$	$\int x^2 d\nu: \mathcal{V}^2 \rightarrow R$	$(\int x^2 d\nu)(U, V) = \int [x, U][x, V] d\nu$

(integrals existing)

Occasionally  $\int x^2 d\nu$  will be employed as a linear transformation  $\mathcal{V} \rightarrow \mathcal{X}$  defined by  $(\int x^2 d\nu)(U) = \int x[x, U] d\nu$  (integrals existing); the usage will be clear from the context. In the  $\mathcal{V}^2 \rightarrow R$  form, it is easy to show that all inner products (i.p.'s) and semi i.p.'s on  $\mathcal{V}$  may be represented in the form  $\int x^2 d\nu$ . The latter is an i.p. or semi i.p. according as the support of  $\nu$  'spans' or does not necessarily 'span'  $\mathcal{X}$  or, equivalently, according as the corresponding linear transformation  $\mathcal{V} \rightarrow \mathcal{X}$  is invertible or not necessarily invertible. For a sample  $s = \{{}^1 x, \dots, {}^n x\}$  of items in  $\mathcal{X}$ , the **sample mean** is  $m = \sum {}^k x / n$ . When  $\{{}^k x - m, k=1, \dots, n\}$ , span  $\mathcal{X}$ ,  $S =_{\text{def}} (n-1)^{-1} \sum ({}^k x - m)^2$  is an i.p. we will call the **sample variance**. Examples of identities that illustrate the simple 'univariate' form of multivariate analysis of variance in the present notation in-

clude

$$\sum {}^k x^2 = nm^2 + \sum ({}^k x - m)^2$$

for a single sample and

$$\sum \sum {}^{ij} x^2 = nm^2 + \sum n_i ({}^i m - m)^2 + \sum \sum ({}^{ij} x - {}^i m)^2$$

for several samples (with the obvious symbolisation).

An i.p. on  $\mathcal{V}$ ,  $I = \int x^2 d\nu$  has a **dual** i.p. on  $\mathcal{X}$  whose associated linear transformation  $\mathcal{X} \rightarrow \mathcal{V}$  is the inverse  $I^{-1}$  of  $I: \mathcal{V} \rightarrow \mathcal{X}$ . We have  $I(U, V) \equiv [IU, V] \equiv I^{-1}(IU, IV)$ . The **sample concentration**,  $C$ , is defined by  $C = S^{-1}$ . A semi i.p.  $I_1$  is **larger** than another,  $I_2$ , if  $I_1(V, V) \geq I_2(V, V)$  for all  $V$ , with inequality for some  $V$ . A semi i.p. is **minimum** in a class if all in the class are larger than or equal to it. Given linear  $T: \mathcal{V} \rightarrow \mathcal{V}$  and a semi i.p.  $I$  on  $\mathcal{V}$ , the **shadow** semi i.p.  $I^T$  is defined by  $I^T(V, V) \equiv I(TV, TV)$ . As a transformation,  $I^T = T'IT$  where  $T'$  denotes the **dual linear transformation** of  $T$  defined by  $[x, TV] \equiv [T'x, V]$ .

Given a random item  $x$  with probability distribution  $P$ , its **mean**  $\mu$  and **variance**  $\mathcal{V}$  are defined by  $\mu = \int x dP$  and  $\mathcal{V} = \int (x - \mu)^2 dP$ , when the integrals exist. We will consider only distributions  $P$  whose support is not confined to any proper manifold (translated subspace) of  $\mathcal{X}$ ; in which case  $\mathcal{V}$  is an i.p. and we write  $\Delta = \mathcal{V}^{-1}$ . Suppose that  $P$  is such that, for a subspace  $S \subset \mathcal{X}$ , conditional probability distributions are definable with respect to the condition  $x \in \mathcal{M} \in \mathcal{X}/S$  where  $\mathcal{X}/S$  is the vector space of cosets of  $S$  in  $\mathcal{X}$ . For simplicity, we suppose the definition is unambiguous, in which case we write  $\mu(\mathcal{M}) = E(x|x \in \mathcal{M})$  and  $\mathcal{V}(\mathcal{M}) = E((x - \mu(\mathcal{M}))^2 | x \in \mathcal{M})$  (integrals existing) for the conditional mean and variance respectively. If  $S$  is proper,  $\mathcal{V}(\mathcal{M})$  will be a semi i.p. We say that

(a)  $P$  has **linear regression on  $\mathcal{X}/S$**  if

$$\mathcal{R}(S) =_{\text{def}} \{\mu(\mathcal{M}) | \mathcal{M} \in \mathcal{X}/S\}$$

is a manifold complementary to  $S$  and

(b)  $P$  is **homoscedastic on  $\mathcal{X}/S$**  if  $\mathcal{V}(\mathcal{M})$  is independent of  $\mathcal{M}$ .

The following lemma reinforces the understanding of this definition of linear regression and, at the same time, illustrates coordinate-free techniques of proof.

**LEMMA 2.1.** *Linear regression on  $\mathcal{X}/S$  is equivalent to  $\mathcal{M}$ -affinity of  $\mu(\mathcal{M})$  for  $\mathcal{M} \in \mathcal{X}/S$ , that is,  $\mu\left(\sum_{i=1}^k \lambda_i \mathcal{M}^i\right) \equiv \sum_{i=1}^k \lambda_i \mu(\mathcal{M}^i)$  for  $\sum \lambda_i = 1$  and  $\mathcal{M}^i \in \mathcal{X}/S$ ,  $i=1, \dots, k$ , for arbitrary  $k$ .*

PROOF. Given the linear regression, let  $\mathcal{T}$  denote the subspace of  $\mathcal{X}$  of which  $\mathcal{R}(S)$  is a translate. Then we may make the resolution  $\mu(\mathcal{M}) = \mu(S) + t(\mathcal{M})$  where  $t(\mathcal{M}) = \mathcal{M} \cap \mathcal{T}$ . Since  $\sum \lambda_i t(\mathcal{M}^i) \in \mathcal{T}$  ( $\mathcal{T}$  is a vector space) and  $\sum \lambda_i t(\mathcal{M}^i) \in \sum \lambda_i \mathcal{M}^i$  (by definition of the latter) we must have  $t(\sum \lambda_i \mathcal{M}^i) = \sum \lambda_i t(\mathcal{M}^i)$  whence  $\mu$  is  $\mathcal{M}$ -affine. The converse is established by showing that  $\{\mu(\mathcal{M}) - \mu(S) \mid \mathcal{M} \in \mathcal{X}/S\}$  span a vector space complementary to  $S$ .

The following result is proved in A3:

THEOREM 2.1. *If  $P$  has linear regression on  $\mathcal{X}/S$  and  $\nabla$  is an i.p. then  $\mu(\mathcal{M})$  is the  $\Delta$ -orthogonal projection of  $\mu$  on  $\mathcal{M}$ , that is,  $\Delta(\mu - \mu(\mathcal{M}), x) \equiv 0$  for  $x \in S$ .*

### 3. Single-sample estimation

Suppose (i)  $^1x, \dots, ^nx, x$  constitute a random sample from  $P$  (ii)  $s = \{^1x, \dots, ^nx\}$  have been completely determined by measurement of  $[^ix, M_j]$ ,  $i=1, \dots, n$  and  $j=1, \dots, p$  but (iii) knowledge of  $x$  is incomplete in that only  $[x, M_j]$ ,  $j=1, \dots, r < p$ , have been measured.

An immediate consequence of (iii) is that knowledge about  $x$  can be expressed in the form  $x \in \mathcal{M}$  for some  $\mathcal{M} \in \mathcal{X}/S$  with  $S$  the bi-orthogonal complement of the subspace  $\mathcal{U}$  of  $\mathcal{V}$  spanned by  $M_1, \dots, M_r$ , that is,  $[t, u] \equiv 0$  for  $t \in S$ ,  $u \in \mathcal{U}$ . The problem to be considered here is that of obtaining an estimator  $\hat{x} = \hat{x}(\mathcal{M}, s)$  of  $x$  based on information about  $P$ , the sample  $s$  and the knowledge that  $x \in \mathcal{M}$ .

*Restriction.* We consider only  $\hat{x}$ 's for which  $\hat{x}(\mathcal{M}, s) \in \mathcal{M}$ ; in which case, estimation of  $x$  is equivalent to prediction of  $[x, M_{r+1}], \dots, [x, M_p]$ .

DEFINITION 3.1.  $\hat{x}(\mathcal{M}, s)$  is  $\mathcal{M}$ -affine if

$$\hat{x}\left(\sum_{i=1}^k \lambda_i \mathcal{M}^i, s\right) = \sum \lambda_i \hat{x}(\mathcal{M}^i, s)$$

identically in  $k$ ,  $\mathcal{M}^1, \dots, \mathcal{M}^k \in \mathcal{X}/S$  if  $\sum \lambda_i = 1$ .

*Remark.* For the case of known  $P$ ,  $\hat{x}(\mathcal{M}) =_{\text{def}} \mu(\mathcal{M})$  is  $\mathcal{M}$ -affine when  $P$  has linear regression. So, in a rough sense, a requirement that  $\hat{x}$  be  $\mathcal{M}$ -affine expresses a belief that  $P$  has linear regression.

By analogy with Lemma 2.1, we see that  $\hat{x}(\mathcal{M}, s)$  is  $\mathcal{M}$ -affine iff  $\hat{x}(\mathcal{M}, s) \equiv \mathcal{M} \cap \mathcal{L}(S, s)$  where  $\mathcal{L}(S, s)$  is a manifold complementary to  $S$ .

DEFINITION 3.2.  $\hat{x}(\mathcal{M}, s)$  is  $S$ -affine if  $\hat{x}(\mathcal{M}, s) \equiv \lambda \hat{x}(\mathcal{M}, s') + (1-\lambda) \cdot \hat{x}(\mathcal{M}, s'')$  for  $^kx = \lambda ^kx' + (1-\lambda) ^kx''$  and  $^kx'' - ^kx' \in S$ ,  $k=1, \dots, n$ .

DEFINITION 3.3.  $\hat{x}(\mathcal{M}, s)$  is equivariant if, for all 1-1 affine trans-

formations  $A, \mathcal{X} \rightarrow \mathcal{X}$ , we have

$$\hat{x}(A\mathcal{M}, As) \equiv A\hat{x}(\mathcal{M}, s).$$

DEFINITION 3.4.  $\hat{x}(\mathcal{M}, s)$  is **conditionally unbiased** if  $E(\hat{x}(\mathcal{M}, s) | {}^{(n)}\mathcal{M}) \equiv \mu(\mathcal{M})$  for  $\mathcal{M} \in \mathcal{X}/S$  where  $E(\cdot | {}^{(n)}\mathcal{M})$  denotes expectation conditional on  ${}^k\mathcal{M} = {}_{\text{def}} {}^kx + S$  fixed,  $k=1, \dots, n$ .

The reader will probably find it useful to examine at this stage the parallels that exist between our formulation and a matrix formulation such as that of Rao ([7], p. 459 *et seq.*). With slight changes of notation, Rao's "Gauss-Markov set-up", incorporating the restriction to an  $X$  whose first column has elements all unity, becomes

$$(3.1) \quad y_k = b + Bx_k + u_k \quad (k=1, \dots, n)$$

with  $u_1, \dots, u_n$  independent and  $E(u_k) = 0$ ,  $E(u_k u_k') = V$ . The correspondence is given by:

$$\begin{aligned} \mathcal{M} \sim \begin{pmatrix} [x, M_1] \\ \vdots \\ [x, M_r] \end{pmatrix} = x, \quad x(\text{in } S) \sim \begin{pmatrix} [x, M_{r+1}] \\ \vdots \\ [x, M_p] \end{pmatrix} = y, \quad {}^kx \sim \begin{pmatrix} x_k \\ y_k \end{pmatrix}, \\ \mu(\mathcal{M}) (\mathcal{M}\text{-affine}) \sim \begin{pmatrix} x \\ b + Bx \end{pmatrix}, \quad \hat{x}(\mathcal{M}, s) \sim \begin{pmatrix} x \\ \hat{y}(x) \end{pmatrix}, \end{aligned}$$

$\mathcal{M}$ -affinity of  $\hat{x}(\mathcal{M}, s) \sim \hat{y}(x)$  an affine transformation of  $x$  (the usual linearity of multiple regression),

$S$ -affinity of  $\hat{x}(\mathcal{M}, s) \sim \hat{y}(x)$  an affine transformation of  $(y'_1, \dots, y'_n)$ , (generalisation to affinity of the usual linearity-in-observations of Gauss-Markov theory),

$V(S)$  (the homoscedastic variance)  $\sim V$ .

Rao's formulation treats  $x_1, \dots, x_n$  as constants whereas we have introduced a probability distribution  $P$  over the whole of  $\mathcal{X}$ ; however the results of this section are conditional on  ${}^{(n)}\mathcal{M}$  so that the distinction is immaterial.

We now introduce an estimator dependent on  $m$  and  $C$  only.

DEFINITION 3.5. The **least squares estimator**,  $\hat{x}_c$  say, is the  $C$ -orthogonal projection of  $m$  onto  $\mathcal{M}$ .

Notice that this estimator has a certain inevitability about it. There is no other estimator that is a function of  $m$  and  $C$  only and is as geometrically natural. It will therefore not be surprising to find that  $\hat{x}_c$  has an important role in statistics; it is, in fact, the usual least squares estimator of the matrix approach so that our terminology is justified.

For a general  $\hat{x}$ ,  ${}^k\hat{x} = {}_{\text{def}}\hat{x}({}^k\mathcal{M}, s)$  is the  $k$ th fitted item. The deviation  ${}^k r = {}^k x - {}^k\hat{x}$  is the  $k$ th residual. The semi i.p.  $R = \sum {}^k r^2$  is the residual sum of squares.

**THEOREM 3.1.** (i)  $\hat{x}_c$  is the  $\mathcal{M}$ -affine estimator minimizing the residual sum of squares;

(ii)  $\hat{x}_c$  is equivariant;

(iii)  $\hat{x}_c(\mathcal{M}, s) = m + \sum \mu_k {}^k x$  where  $\mu_k = (n-1)^{-1} C_{\mathcal{M}}({}^k\mathcal{M} - \bar{\mathcal{M}}, \mathcal{M} - \bar{\mathcal{M}})$  and  $C_{\mathcal{M}}$  denotes the concentration of the sample  ${}^{(n)}\mathcal{M}$  in the vector space  $\mathcal{X}/S$ ;

(iv)  $\hat{x}_c$  is  $S$ -affine;

(v)  $\hat{x}_c$  is conditionally unbiased if  $P$  has linear regression on  $\mathcal{X}/S$ ;

(vi) when  $P$  is homoscedastic on  $\mathcal{X}/S$

$$(3.2) \quad V(\hat{x}_c(\mathcal{M}, s) | {}^{(n)}\mathcal{M}) = \left[ 1 + \frac{n}{n-1} C_{\mathcal{M}}(\mathcal{M} - \bar{\mathcal{M}}, \mathcal{M} - \bar{\mathcal{M}}) \right] \frac{V(S)}{n};$$

(vii) when  $P$  is homoscedastic on  $\mathcal{X}/S$ ,  $\hat{x}_c$  has minimum conditional variance, given by (3.2), among  $S$ -affine estimators that are conditionally unbiased on  $S$  for  $P$  having linear regression on  $\mathcal{X}/S$ ;

(viii) when  $P$  is homoscedastic on  $\mathcal{X}/S$ ,  $\hat{x}_c$  has minimum conditional mean square error, given by the right-hand side of (3.2), among  $S$ -affine estimators that have bounded conditional mean square error for  $P$  having linear regression on  $\mathcal{X}/S$ .

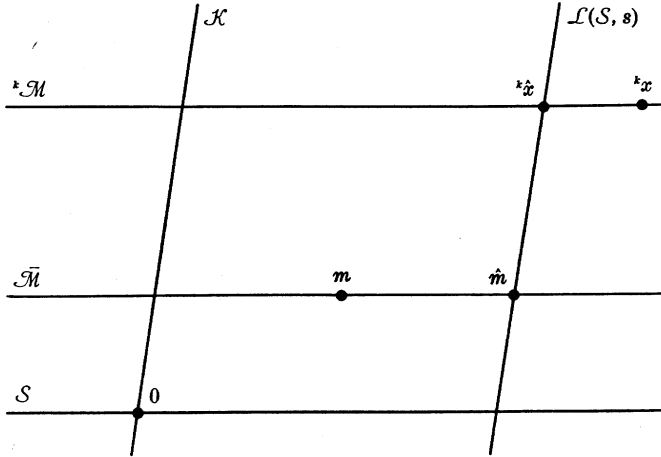
*Remarks.* The result (i) gives an alternative definition that is also a coordinate-free multivariate generalisation of the process of fitting a straight line by least squares. The result (ii) reveals the "robustness" or "indifference" of  $\hat{x}_c$  to all aspects of the estimation problem except its affine invariants. The result (iii) shows that  $\hat{x}_c(\mathcal{M}, s)$  is a scalar combination of  $\{{}^k x\}$  with coefficients which are functions of  ${}^{(n)}\mathcal{M}$  and  $\mathcal{M}$ . This is an interesting fact not immediately apparent from other approaches. Equation (3.2) is the coordinate-free multivariate generalisation of the well-known formula for the variance of the prediction from a least-squares fitted straight line

$$\text{Var}(\hat{y}|x) = \left[ 1 + \frac{n}{n-1} \frac{(x - \bar{x})^2}{s^2} \right] \frac{\sigma_{y|x}^2}{n}$$

where  $s^2 = \sum (x_i - \bar{x})^2 / (n-1)$ . Result (vi) is a version of the multivariate Gauss-Markov theorem while (vii) is an alternative to it in generalisation of Barnard [1]. The relationship with estimation based on (3.1) is that minimum variance estimation of the elements of  $\mathbf{b}$  and  $\mathbf{B}$  by unbiased linear functions of the coordinates of the  $\{\mathbf{y}_k\}$  is equivalent to minimum variance estimation of  $\mu(\mathcal{M})$ , for all  $\mathcal{M} \in \mathcal{X}/S$ , by an  $\mathcal{M}$ -affine,

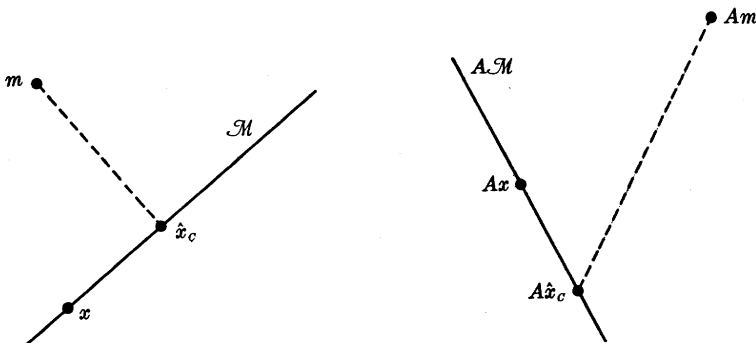
$S$ -affine and conditionally unbiased estimator  $\hat{x}(\mathcal{M}, s)$ . The interpretation of the estimator as equivalent to  $C$ -projection of  $m$  is, we claim, not easily uncovered in the matrix approach.

PROOF OF THEOREM. (i) With  $\hat{x}(\mathcal{M}, s) = \mathcal{M} \cap \mathcal{L}(S, s)$ , let  $\mathcal{K}$  be the subspace complementary to  $S$  and parallel to  $\mathcal{L}(S, s)$ . Let  $\Pi$  denote projection onto  $S$  with kernel  $\mathcal{K}$ :

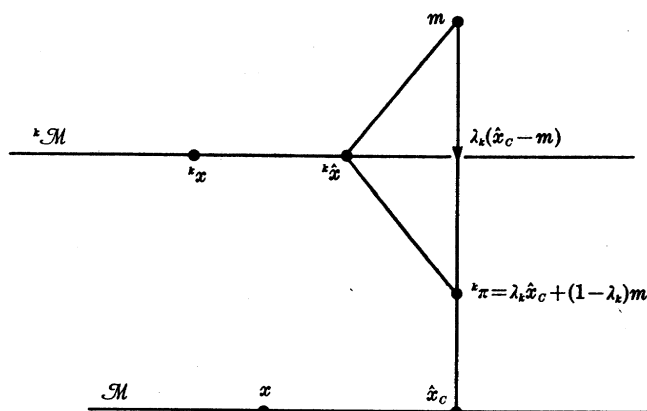


With  $\hat{m} = \text{def } \sum ^k \hat{x}/n$ , we have  $\hat{m} \in \mathcal{L}(S, s)$  and  $\hat{m} - m \in S$  so that  $^k x - ^k \hat{x} = \Pi(^k x - m) - (\hat{m} - m)$  and, by A2,  $R = (n-1)S'' + n(\hat{m} - m)^2$ . Whence, by A1,  $R$  is minimized when  $\hat{m} = m$  and  $\mathcal{K}$  is  $C$ -orthogonal to  $S$ , that is, when  $\mathcal{L}(S, s)$  passes through  $m$  and is  $C$ -orthogonal to  $S$ .

(ii) By the identity  $((S^T)^{-1})^T = C$  for 1-1  $T$ ,  $\hat{x}_c - m$  and  $x - \hat{x}_c$  being  $C$ -orthogonal implies that  $T\hat{x}_c - Tm$  and  $Tx - T\hat{x}_c$  (that is,  $A\hat{x}_c - Am$  and  $Ax - A\hat{x}_c$ ) are  $(S^T)^{-1}$  orthogonal. But  $S^{T'} = \sum (A^k x - Am)^2 / (n-1)$ , the variance of the sample  $As$ , establishing the result.



(iii)  $^k \pi$  is defined as the  $C$ -orthogonal projection of  $^k \hat{x}_c$  on the line through  $m$  and  $\hat{x}_c$ . Now  $\lambda_k(\hat{x}_c - m)$  is the  $C$ -orthogonal projection of  $^k x - m$  on the one-dimensional subspace,  $\mathcal{I}$  say, parallel to  $\hat{x}_c - m$ .



So using A4 with  $I=S$ , we obtain  $\sum [{}^kx - m - \lambda_k(\hat{x}_c - m)]\lambda_k(\hat{x}_c - m) = 0$  which implies  $\sum \lambda_k[{}^kx - m - \lambda_k(\hat{x}_c - m)] = 0$  or since  $\sum \lambda_k = 0$ ,  $\hat{x}_c = m + \sum \mu_k {}^kx$  where  $\mu_k = \lambda_k / \sum \lambda_k^2$ . From the figure, we see that  ${}^kx - m - \lambda_k(\hat{x}_c - m)$  is  $C$ -orthogonal to  $\hat{x}_c - m$ . Whence  $\lambda_k = C({}^kx - m, \hat{x}_c - m) / C(\hat{x}_c - m, \hat{x}_c - m)$ . So

$$\begin{aligned} \sum \lambda_k^2 &= \sum C({}^kx - m, \hat{x}_c - m)^2 / C(\hat{x}_c - m, \hat{x}_c - m)^2 \\ &= (n-1) / C(\hat{x}_c - m, \hat{x}_c - m) \end{aligned}$$

by A5.2. Hence  $\mu_k = (n-1)^{-1} C({}^kx - m, \hat{x}_c - m) = (n-1)^{-1} C({}^k\hat{x}_c - m, \hat{x}_c - m)$ . We can now use A6 with  $\Pi(x - m) = \hat{x}_c - m$  and the fact that a sample concentration is invariant under isomorphism to obtain

$$\mu_k = (n-1)^{-1} C_{\mathcal{M}}({}^k\mathcal{M} - \bar{\mathcal{M}}, \mathcal{M} - \bar{\mathcal{M}}).$$

(iv) This follows from (iii).

(v) This follows from (iii) and A5.1.

(vi) When  $P$  is homoscedastic on  $\mathcal{X}/S$ , we obtain from (iii)

$$\begin{aligned} V(\hat{x}_c(\mathcal{M}, s) | {}^{(n)}\mathcal{M}) &= \sum (\mu_k + n^{-1})^2 V(S) \\ &= [1 + n(n-1)^{-1} C_{\mathcal{M}}(\mathcal{M} - \bar{\mathcal{M}}, \mathcal{M} - \bar{\mathcal{M}})] V(S) / n \end{aligned}$$

using A5.2.

(vii) By the algebraic theory to be found in MacLane and Birkhoff ([6], p. 426), we see that  $S$ -affinity of  $\hat{x}$  is equivalent to

$$(3.3) \quad \hat{x}(\mathcal{M}, s) \equiv x_0 + \sum T_k({}^kx - x_k)$$

where  $x_0, x_1, \dots, x_n$  are arbitrary fixed points in  $\mathcal{M}, {}^1\mathcal{M}, \dots, {}^n\mathcal{M}$ , respectively, while  $T_1, \dots, T_n$  are linear,  $S \rightarrow S$ , with  $x_0, \dots, x_n$  and  $T_1, \dots, T_n$  possibly functions of  ${}^{(n)}\mathcal{M}$ . Conditional unbiasedness of  $\hat{x}$  is then equivalent to

$$(3.4) \quad \mu(\mathcal{M}) \equiv x_0 + \sum T_k[\mu({}^k\mathcal{M}) - x_k]$$



for all linear regressions  $\mu$ . Now

$$(3.5) \quad \mathcal{V}(\hat{x}(\mathcal{M}, s) | {}^{(n)}\mathcal{M}) = E[\{\sum T_k [{}^k x - \mu({}^k \mathcal{M})]\}^2 | {}^{(n)}\mathcal{M}] = \int \sum (T_k x)^2 d\nu$$

if  $\mathcal{V}(S) = \int x^2 d\nu$ . Writing  $q_k = \mu_k + n^{-1}$ ,

$$(3.6) \quad \hat{x}_c(\mathcal{M}, s) = \sum q_k {}^k x$$

and

$$(3.7) \quad \mathcal{V}(\hat{x}_c(\mathcal{M}, s) | {}^{(n)}\mathcal{M}) = \int \sum (q_k x)^2 d\nu.$$

But

$$(3.8) \quad \sum (T_k x)^2 - \sum (q_k x)^2 = \sum [(T_k - q_k I)x]^2 + x(Tx) + (Tx)x$$

where  $T = (\sum q_k T_k) - (\sum q_k^2)I$ . In (3.4), take  $x_k = \mu_0({}^k \mathcal{M})$  (and hence  $x_0 = \mu_0(\mathcal{M})$ ) for some linear regression  $\mu_0$ . Then (3.4), the relation  $\mathcal{M} = \sum q_k {}^k \mathcal{M}$  (a consequence of A5.1) and (iii) imply

$$(3.9) \quad \sum (T_k - q_k I)[\mu({}^k \mathcal{M}) - \mu_0({}^k \mathcal{M})] \equiv 0$$

identically in linear regressions  $\mu$ . The choice

$$\mu({}^k \mathcal{M}) = \mu_0({}^k \mathcal{M}) + \left[ \frac{1}{n} + \frac{1}{n-1} C_{\mathcal{M}}({}^k \mathcal{M} - \bar{\mathcal{M}}, \mathcal{M} - \bar{\mathcal{M}}) \right] s$$

$k=1, \dots, n$ , where  $s \in S$  is arbitrary, is a linear regression and gives  $\mu({}^k \mathcal{M}) - \mu_0({}^k \mathcal{M}) = q_k s$ . Whence, substituting in (3.9),  $\sum q_k (T_k - q_k I)s = 0$  or  $T=0$ . Substitution in (3.8) and comparison of (3.5) and (3.7) then establishes the result.

(viii) With (3.3), the conditional mean square  $E[\{\hat{x}(\mathcal{M}, s) - x\}^2 | {}^{(n)}\mathcal{M}, \mathcal{M}] = \mathcal{V}(\mathcal{M}) + \mathcal{V}(\hat{x}(\mathcal{M}, s) | {}^{(n)}\mathcal{M}) + \left[ x_0 + \sum_{k=1}^n T_k \{\mu({}^k \mathcal{M}) - x_k\} - \mu(\mathcal{M}) \right]^2$ . The transformation

$$(3.10) \quad \mu(\cdot) \rightarrow \sum_{k=1}^n T_k \{\mu({}^k \mathcal{M}) - x_k\} - \mu(\mathcal{M})$$

is an affine transformation from the affine space of linear regressions  $\mu$  to the affine space  $\mathcal{M}$ , whence the right-hand side of (3.10) is necessarily of the form  $x^* + T[\mu(\cdot) - \mu^*(\cdot)]$  for some fixed  $x^* \in \mathcal{M}$  and  $T$  linear from the vector space  $\{\mu(\cdot) - \mu^*(\cdot) | \text{fixed } \mu^*\}$  to  $S$ . Unless  $T=0$ ,  $[x_0 + x^* + T\{\mu(\cdot) - \mu^*(\cdot)\}]^2(V, V)$  will be unbounded for at least one  $V$ , while the condition of minimum mean square error requires  $x^* = -x_0$  whence  $x_0 + \sum T_k \{\mu({}^k \mathcal{M}) - x_k\} \equiv \mu(\mathcal{M})$ , that is,  $\hat{x}(\mathcal{M}, s)$  must be conditionally unbiased. The result then follows from (vii).

*Remark.* The form of  $\hat{x}_c(\mathcal{M}, s)$  given by (iii) can be reexpressed by giving  $C_n$  its transformation role. We have

$$\begin{aligned} \sum (n-1)^{-1} C_n({}^k\mathcal{M} - \bar{\mathcal{M}}, \mathcal{M} - \bar{\mathcal{M}}) {}^k x \\ = \sum (n-1)^{-1} [C_n({}^k\mathcal{M} - \bar{\mathcal{M}}), \mathcal{M} - \bar{\mathcal{M}}] {}^k x = B(\mathcal{M} - \bar{\mathcal{M}}) \end{aligned}$$

where  $B = \sum_k {}^k x (\sum_h ({}^h\mathcal{M} - \bar{\mathcal{M}})^2)^{-1} ({}^k\mathcal{M} - \bar{\mathcal{M}})$  may be called the *regression transformation* from  $\mathcal{X}/S$  to  $\mathcal{X}$ . (We adopt the convention  $(\sum {}^k x {}^k z)U =_{\text{def}} \sum {}^k x [{}^k z, U]$ .) Then

$$(3.11) \quad \hat{x}_c(\mathcal{M}, s) = m + B(\mathcal{M} - \bar{\mathcal{M}}).$$

The fact that  $\hat{x}_c$  is the coordinate-free form of the usual least squares estimator is now explicit.

#### 4. Covariance adjustment

Suppose  $\{{}^1x, \dots, {}^{n+c}x\}$  is a random sample from  $P$  with unknown mean  $\mu$  and unknown variance  $\mathcal{V}$ . As in Section 3 we suppose that  $s = \{{}^1x, \dots, {}^n x\}$  have been completely measured but that, for  $\{{}^{n+1}x, \dots, {}^{n+c}x\}$ , we know only that  ${}^k x \in {}^k\mathcal{M} \in \mathcal{X}/S$ ,  $k = n+1, \dots, n+c$ . Suppose  $P$  has linear regression and is homoscedastic on  $\mathcal{X}/S$  and that it is required to estimate  $\mu$ . The estimator  $\hat{\mu}_n = m = \sum_{k=1}^n {}^k x / n$  has conditional mean and variance  $\mu(\bar{\mathcal{M}}_n)$  and  $\mathcal{V}(S)/n$ , respectively, where  $\bar{\mathcal{M}}_n = \sum_{k=1}^n {}^k\mathcal{M}/n$ . The correlated estimator  $\hat{\mu}_c = \sum_{k=n+1}^{n+c} {}^k \hat{x}_c / c$ , where  $C$  is the concentration of the sample  $s$  and  ${}^k \hat{x}_c$  is the  $C$ -orthogonal projection of  $m$  onto  ${}^k\mathcal{M}$ , has conditional mean  $\mu(\bar{\mathcal{M}}_c)$ , where  $\bar{\mathcal{M}}_c = \sum_{k=n+1}^{n+c} {}^k\mathcal{M}/c$ . It is natural to consider the estimator

$$\hat{\mu}_i = \lambda \hat{\mu}_n + (1-\lambda) \hat{\mu}_c = m - (1-\lambda) B(\bar{\mathcal{M}}_n - \bar{\mathcal{M}}_c)$$

by (3.11). By Theorem 3.1-(v),  $\hat{\mu}_i$  has conditional expectation  $\lambda \mu(\bar{\mathcal{M}}_n) + (1-\lambda) \mu(\bar{\mathcal{M}}_c)$  and, using Theorem 3.1-(iii) and A5, we find

$$\mathcal{V}(\hat{\mu}_i | {}^{(n+c)}\mathcal{M}) = \left[ 1 + \frac{n}{(n-1)} (1-\lambda)^2 C_{\mathcal{M}}(\bar{\mathcal{M}}_n - \bar{\mathcal{M}}_c, \bar{\mathcal{M}}_n - \bar{\mathcal{M}}_c) \right] \mathcal{V}(S)/n.$$

The unconditional mean square error of  $\hat{\mu}_i$  as estimator of  $\mu$  is then

$$\begin{aligned} E[\lambda \{\mu(\bar{\mathcal{M}}_n) - \mu\} + (1-\lambda) \{\mu(\bar{\mathcal{M}}_c) - \mu\}]^2 \\ + \left\{ 1 + \frac{n}{n-1} (1-\lambda)^2 E[C_{\mathcal{M}}(\bar{\mathcal{M}}_c - \bar{\mathcal{M}}_n, \bar{\mathcal{M}}_c - \bar{\mathcal{M}}_n)] \right\} \mathcal{V}(S)/n. \end{aligned}$$

The first expectation equals  $[\lambda^2/n + (1-\lambda)^2/c]\mathcal{V}^{T'}$  where  $T: \mathcal{X} \rightarrow \mathcal{X}$  denotes projection parallel to  $\mathcal{S}$  onto  $\mathcal{I}$ , the complementary subspace parallel to  $\mathcal{R}(S)$ . [If  $P$  has linear regression and is also homoscedastic on  $\mathcal{X}/\mathcal{I}$  then  $\mathcal{V}^{T'} = \mathcal{V}(\mathcal{I})$ .] With interest concentrated on a particular variable  $U$  say, it is easily verified that the mean square error of  $[\hat{\mu}_i, U]$  is minimised by the choice

$$\lambda = \frac{(\delta/c + \epsilon)}{(\delta/n + \delta/c + \epsilon)}$$

where  $\delta = \mathcal{V}(T'U, T'U)/\mathcal{V}(S)(U, U)$  and

$$\epsilon = (n-1)^{-1} E [C_{\mathcal{M}}(\bar{\mathcal{M}}_c - \bar{\mathcal{M}}_n, \bar{\mathcal{M}}_c - \bar{\mathcal{M}}_n)] .$$

$\delta$  may be more easily interpretable when written as  $\mathcal{V}(U_2, U_2)/\mathcal{V}(U_1, U_1)$  where  $U_1$  and  $U_2$  are the components of  $U$  in the resolution  $\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_2$  where (i)  $\mathcal{V}_1$  is the "error variable" subspace of  $\mathcal{V}$ , that is, the subspace of variables that are "regression free" with respect to  $\mathcal{R}(S)$ , namely,  $[\mu(\mathcal{M}), U_1]$  is independent of  $\mathcal{M} \in \mathcal{X}/S$  for  $U_1 \in \mathcal{V}_1$  while (ii)  $\mathcal{V}_2$  is the subspace of variables on which  $^{n+1}x, \dots, ^{n+c}x$  were measured, that is, the biorthogonal complement of  $S$ . This form of  $\delta$  leads with little analysis to the identity  $\delta = R_v^2/(1-R_v^2)$  where  $R_v^2$  is the multiple correlation coefficient between  $[x, U]$  and  $[x, M_1], \dots, [x, M_r]$  for the distribution  $P$ . Without knowledge of  $P$ , it is necessary to estimate  $\delta$  and  $\epsilon$  before  $[\hat{\mu}_i, U]$  can be calculated. Estimation could be based on  $m$  and  $C$ .

Note that the dependence of the choice of  $\lambda$  in  $\hat{\mu}_i$  on the variable of interest  $U$  means that  $\hat{\mu}_i$  is not, in general, a linear function on  $\mathcal{V}$ . So we are not, strictly speaking, estimating  $\mu$  *per se* but providing estimation of each  $[\mu, U]$ .

A special case of some importance is when the distribution,  $P_{\mathcal{M}}$ , of  $\mathcal{M}$  in  $\mathcal{X}/S$  induced by  $P$ , is  $\Delta_{\mathcal{M}}$ -spherical about  $E(\mathcal{M})$ , where  $\Delta_{\mathcal{M}} = [E(\mathcal{M} - E(\mathcal{M}))^2]^{-1}$ . (That is,  $P_{\mathcal{M}}[A] = P_{\mathcal{M}}[E(\mathcal{M}) + T(A - E(\mathcal{M}))]$  where  $T$  is any  $\Delta_{\mathcal{M}}$ -orthogonal transformation of  $\mathcal{X}/S$  and  $A - E(\mathcal{M})$  denotes  $\{\mathcal{M} - E(\mathcal{M}) | \mathcal{M} \in A\}$ .) By an argument based on conformal shells, it can be seen that  $\epsilon$  has the same value for all such spherical distributions and, in particular, the same value as for a normal distribution. Another argument shows that, for  $P_{\mathcal{M}}$  normal,

$$E C_n(\bar{\mathcal{M}}_n - \bar{\mathcal{M}}_c, \bar{\mathcal{M}}_n - \bar{\mathcal{M}}_c) = (n-1)(n+c)d/[nc(n-d-2)]$$

where  $d = \dim(\mathcal{X}/S)$ ; whence  $\epsilon = (n+c)d/[nc(n-d-2)]$ . In the case  $c = \infty$ , we get

$$\lambda = \frac{d(1-R_v^2)}{d(1-R_v^2) + (n-d-2)R_v^2} .$$

## 5. Relationship to other work

*Dempster.* The key relationship is the identity of the least squares estimator,  $\hat{x}_c$ , and Dempster's "augmented best linear predictor" ([2], p. 148). Although Theorem 3.1 gives us implicit assurance of this identity, it is readily established by direct argument. We have to translate from Dempster's definition in  $\mathcal{V}$  to our definition in  $\mathcal{X}$ . From Dempster's equation (8.2.12), the "predicted value" for item  $x$  on variable  $V$  is, in our notation,  $[m, V] + [x - m, \dot{V}]$  in which  $\dot{V} = TV$  where  $T$  is  $S$ -orthogonal projection onto  $\mathcal{V}_2$ . But  $[x - m, \dot{V}] = [T'(x - m), V]$  where  $T'$  is  $C$ -projection onto the  $C$ -orthogonal complement of  $\mathcal{S}$ . Moreover  $T'(x - m) = \hat{x}_c - m$  so that the predicted value using the augmented best linear predictor is  $[m, V] + [\hat{x}_c - m, V] = [\hat{x}_c, V]$ , our least squares estimate of the value of item  $x$  on variable  $V$ .

*Kruskal and Eaton.* Kruskal ([4], [5]) discusses a general 'univariate' model in which our single multivariate sample may be accommodated, albeit somewhat artificially. It is convenient to restate Kruskal's theorem in the terminology developed in this paper. We set up an  $N$ -dimensional vector space  $\mathcal{Q}$ , which we may call data space, and, implicitly, its dual vector space  $\mathcal{W}$ , which we may call combination space, with the bilinear product  $[Y, W]$ ,  $Y \in \mathcal{Q}$ ,  $W \in \mathcal{W}$ . We have a probability distribution on  $\mathcal{Q}$  such that  $EY = \mu$  and the objective is estimation of  $\mu$ , supposing  $\mu \in \Omega$  a subspace (linear manifold) of  $\mathcal{Q}$ . Suppose  $E(Y - \mu)^2 = V = \int y^2 dm$  is an i.p. on  $\mathcal{W}$  and  $A = V^{-1}$ . The Gauss-Markov estimator of  $\mu$ ,  $\hat{\mu}$ , is the  $A$ -projection of  $Y$  on  $\Omega$ . (It is easily verified that  $\hat{\mu}$  is the minimum variance, linear, unbiased estimator of  $\mu$ .) For a given i.p.  $J$  on  $\mathcal{Q}$ , the least  $J$ -squares estimator,  $\mu^*$ , is the  $J$ -orthogonal projection of  $Y$  on  $\Omega$ . The covariance operator  $\mathcal{X}: \mathcal{Q} \rightarrow \mathcal{Q}$  (with respect to  $J$ ) is defined by

$$\mathcal{X}z = \int J(y, z)y dm.$$

It is easily shown that  $\mathcal{X}$  is 1-1. Kruskal's Theorem 1, [5], " $\mu^* = \hat{\mu}$  if and only if  $\mathcal{X}\Omega = \Omega$ ", is proved in this formulation by firstly using A5.1 to establish that

$$A(Y - \mu, \mathcal{X}z) = J(Y - \mu, z) \quad \text{for } \mu, z \in \Omega.$$

(We pay a penalty in proof-length for implicitly distinguishing between data and combination space, a distinction that Kruskal does not make).

To accommodate our single multivariate sample in this version, we fix  $\mathcal{S}$  and  $\{\mathcal{M}\}$ . Take  $N = np$  and identify the sample  $s$  with  $Y \in \mathcal{Q}$  by

$$Y = [{}^1x - \mu_0({}^1\mathcal{M}), \dots, {}^nx - \mu_0({}^n\mathcal{M})]$$

where  $\mu_0$  is an arbitrary linear regression.  $\mathcal{Y}$  is thus isomorphic to  $S^n$ . We have

$$\mu = [\mu({}^1\mathcal{M}) - \mu_0({}^1\mathcal{M}), \dots, \mu({}^n\mathcal{M}) - \mu_0({}^n\mathcal{M})]$$

and  $\{\mu \mid \text{all linear regressions } \mu(\mathcal{M}) \text{ on } \mathcal{X}/S\}$  is a subspace  $\Omega$ .

We have  $\mathcal{E}\mu = \int J(y, \mu) y dm$  where  $dm$  gives the components  $y_1, \dots, y_n$  of  $y$  independent distribution  $P_1, \dots, P_n$ , say, respectively, with zero mean (corresponding to the conditional distributions of  ${}^kx - \mu({}^k\mathcal{M})$  in  ${}^k\mathcal{M}$ ,  $k=1, \dots, n$ ). Hence

$$\mathcal{E}\mu = \left[ \int J_k(y_k, \mu({}^k\mathcal{M}) - \mu_0({}^k\mathcal{M})) y_k dP_k \right]$$

where  $J_k$  is the restriction of  $J$  to the  $k$ th component of  $\mathcal{Y}$ . For a basis  $\{\mathcal{M}_i\}$  of  $\mathcal{X}/S$  with  ${}^k\mathcal{M} = \sum_i {}^k\lambda_i \mathcal{M}_i$ , we have

$$\mathcal{E}\mu = \left[ \sum_i {}^k\lambda_i \mathcal{E}_k(\mu(\mathcal{M}_i) - \mu_0(\mathcal{M}_i)) \right]$$

where  $\mathcal{E}_k = \int J_k(s, \cdot) s dP_k$ . If  $\mathcal{E}_k$  is independent of  $k$  then  $\mathcal{E}\Omega = \Omega$  and  $\hat{\mu} \equiv \mu^*$ . A sufficient condition that  $\mathcal{E}_k$  be independent of  $k$  is that  $\int s^2 dP_k$  and  $J_k$  be independent of  $k$ , when the components of  $\mathcal{Y}$  are identified with  $S$ . This is equivalent to homoscedasticity on  $\mathcal{X}/S$ ; and, if all such constant  $J_k$  are considered, the least  $J$ -squares estimator corresponds to the  $\mathcal{M}$ -affine estimator minimizing  $\sum ({}^kx - {}^k\hat{x})^2$ , that is to  $({}^1\hat{x}_c, \dots, {}^n\hat{x}_c)$ . This insight thus gives us an alternative proof of Theorem 3.1-(vii).

The argument just described may be regarded as providing an alternative to Eaton's [3] embedding of the 'standard multivariate linear model' in the Kruskal setup, at least when the single restriction is made that makes that model equivalent to our conditioned single multivariate sample.

## Appendix

A1. For the class of projections  $\{\Pi : \mathcal{X} \rightarrow \mathcal{X}\}$  with fixed range and an i.p.  $I$  on  $\mathcal{V}$ , the shadow semi i.p.  $I''$  is minimized when  $\Pi$  is  $I^{-1}$ -orthogonal. For  $\Pi'$  has a fixed kernel  $\mathcal{K}$ , say, and  $I(\Pi'V, \Pi'V)$  is, for fixed  $V$ , minimized when  $\Pi'V$  is  $I$ -orthogonal to  $\mathcal{K}$ . But  $I(\Pi'V, \Pi'V) = I''(V, V)$  so that  $I''$  is minimized when  $\Pi'$  is  $I$ -orthogonal to  $\mathcal{K}$  which implies the result.

A2. If  $I = \int x^2 d\nu$  and  $T: \mathcal{V} \rightarrow \mathcal{V}$  is linear,  $I^T = \int (T'x)^2 d\nu$ . For  $I^T(U, V) = \int [x, TU][x, TV] d\nu = \int [T'x, U][T'x, V] d\nu = \left( \int (T'x)^2 d\nu \right) (U, V)$ .

A3. Let  $\Pi_S$  denote projection onto  $S$  parallel to  $\mathcal{R}(S)$ . Then, by A2,  $\mathcal{V}^{\Pi'_S} = E(\Pi_S(x - \mu))^2 = E_{\mathcal{M}} E_{x \in \mathcal{M}} (x - \mu(\mathcal{M}))^2$  with obvious notation. If  $\mu_A(\mathcal{M})$  is the  $A$ -orthogonal projection of  $\mu$  onto  $\mathcal{M}$ , we have

$$\mathcal{V}^{\Pi'_A} = E_{\mathcal{M}} E_{x \in \mathcal{M}} (x - \mu_A(\mathcal{M}))^2.$$

By A1,  $\mathcal{V}^{\Pi'_A} \leq \mathcal{V}^{\Pi'_S}$  or  $E_{\mathcal{M}} E_{x \in \mathcal{M}} (x - \mu_A(\mathcal{M}))^2 \leq E_{\mathcal{M}} E_{x \in \mathcal{M}} (x - \mu(\mathcal{M}))^2$ . But  $\mu(\mathcal{M})$  is the value of  $x^* \in \mathcal{M}$  minimising  $E_{x \in \mathcal{M}} (x - x^*)^2$  so that the identity  $\mu(\mathcal{M}) \equiv \mu_A(\mathcal{M})$  follows.

A4. If  $\mathcal{X} = \mathcal{I} \oplus \mathcal{U}$  where  $\mathcal{I} \cap \mathcal{U} = 0$  and  $I = \int x^2 d\nu$  is an i.p. on  $\mathcal{V}$ , a necessary and sufficient condition that  $\mathcal{I}$  and  $\mathcal{U}$  be  $I^{-1}$ -orthogonal is  $\int t(x)u(x)d\nu = 0$  where  $t(x) \in \mathcal{I}$ ,  $u(x) \in \mathcal{U}$  and  $x = t(x) + u(x)$ . For if  $\Pi_1$  and  $\Pi_2$  are projections defined by  $\Pi_1 x \equiv t(x)$  and  $\Pi_2 x \equiv u(x)$ , we have  $\left( \int \Pi_1 x \cdot \Pi_2 x d\nu \right) (V, V) \equiv \int [x, \Pi'_1 V][x, \Pi'_2 V] d\nu \equiv I(\Pi'_1 V, \Pi'_2 V) \equiv 0$  if and only if  $\Pi'_1$  and  $\Pi'_2$  are  $I$ -orthogonal, equivalent to  $I^{-1}$ -orthogonality of  $\Pi_1$  and  $\Pi_2$ ; whence the result.

A5. If  ${}^1y, \dots, {}^ny$  span  $\mathcal{X}$  so that  $F = \sum {}^ky^2$  is an i.p. and  $G = F^{-1}$  then:

$$A5.1. \quad \sum G({}^ky, x){}^ky \equiv x$$

$$A5.2. \quad \sum G({}^ky, x)G({}^ky, z) \equiv G(x, z)$$

$$A5.3. \quad \sum G({}^ky, {}^ky) = p.$$

For A5.1,  $[\sum G({}^ky, x){}^ky, V] = \sum [{}^ky, Gx][{}^ky, V] = F(Gx, V) = [x, V]$ ; while A5.2 is an immediate consequence. For A5.3, let  $G$  denote the  $n \times n$  symmetric matrix with  $(i, j)$ th element  $G({}^iy, {}^jy)$ . A5.1 shows that  $G$  has rank  $p$  while A5.2 implies its idempotency, whence A5.3 follows.

A6. If  $\mathcal{K}$  is a subspace of  $\mathcal{X}$  and  $\Pi$  denotes  $C$ -orthogonal projection onto  $\mathcal{K}$  then, for  $x \in \mathcal{K}$ ,  $C(x, x) = C(\Pi s) (x, x)$  where  $\Pi s$  denotes the sample  $\{\Pi^1 x, \dots, \Pi^n x\}$  in  $\mathcal{K}$  and  $C(\Pi s)$  is its sample concentration (defined in  $\mathcal{K}$  as vector space).

The proof is fairly straightforward.

## Acknowledgement

I am grateful to A. P. Dawid for his interest in and comments on this work.

UNIVERSITY COLLEGE LONDON

## REFERENCES

- [1] Barnard, G. A. (1963). The logic of least squares, *J. R. Statist. Soc.*, A **25**, 124-127.
- [2] Dempster, A. P. (1969). *Elements of Continuous Multivariate Analysis*, Addison Wesley, Reading, Mass.
- [3] Eaton, Morris L. (1970). Gauss-Markov estimation for multivariate linear models: A coordinate-free approach, *Ann. Math. Statist.*, **41**, 528-538.
- [4] Kruskal, W. H. (1961). The coordinate-free approach to Gauss-Markov estimation and its application to missing and extra observations, *Proc. Fourth Berkeley Symp.*, I, 435-451.
- [5] Kruskal, W. H. (1968). When are Gauss-Markov and least squares estimators identical? A coordinate-free approach, *Ann. Math. Statist.*, **39**, 70-75.
- [6] MacLane, S. and Birkhoff, G. (1967). *Algebra*, Macmillan, New York, Collier-Macmillan, London.
- [7] Rao, C. R. (1965). *Linear Statistical Inference and its Applications*, John Wiley, New York.