THE COORDINATE-FREE APPROACH TO GAUSS-MARKOV ESTIMATION, AND ITS APPLICATION TO MISSING AND EXTRA OBSERVATIONS

WILLIAM KRUSKAL
UNIVERSITY OF CHICAGO

1. Introduction and summary

The purposes of this paper are (1) to describe the coordinate-free approach to Gauss-Markov (linear least squares) estimation in the context of Model I analysis of variance and (2) to discuss, in coordinate-free language, the topics of missing observations and extra observations.

It is curious that the coordinate-free approach to Gauss-Markov estimation, although known to many statisticians, has infrequently been discussed in the literature on least squares and analysis of variance. The major textbooks in these areas do not use the coordinate-free approach, and I know of only a few journal articles that deal with it ([2], plus some of the references in Dutch that it lists, and, to some extent, [1], [3] and [7]). The coordinate-free viewpoint is implicit in R. A. Fisher's geometrical approach to sampling problems.

The subject of missing observations in Model I analysis of variance is well understood and often discussed. This paper presents no new results here, but it does present a viewpoint different from that usually given. In contrast, the topic of extra observations, although it was briefly considered by Gauss [5], section 35 of *Theoria Combinationis* . . . , has elicited hardly any papers since. (I know only of papers by R. L. Plackett [9] and K. D. Tocher [10].) The problem of extra observations is important in its own right and also in connection with the treatment of so-called outliers. I shall discuss a method of treating extra observations that bears some resemblance to that for missing observations. In particular, it leads to possible methods for treating apparent outliers that I described briefly in [8].

There are two major motivations for emphasizing the coordinate-free approach to Gauss-Markov estimation. First, it permits a simpler, more general, more elegant, and more direct treatment of the general theory of linear estimation

The work leading to this paper was supported in part by the Logistics and Mathematical Statistics Branch of the Office of Naval Research and in part by the National Science Foundation.

than do its notational competitors, the matrix and scalar approaches. Second, it is useful as an introduction to infinite-dimensional spaces, which are important, for example, in the consideration of stochastic processes.

A related point is that more or less coordinate-free treatments of finite-dimensional vector spaces are now more common than they once were and are being taught to students at an earlier stage. With such mathematical background, a student can learn the theoretical side of Model I analysis of variance quickly and efficiently. The treatment in this paper will, however, be compact and without the motivational material and the many examples that would be pedagogically important.

Nonetheless, it may be useful to keep one concrete example before us. Accordingly, I shall illustrate the general theory in terms of a simple illustration, two-way analysis of variance with one observation per cell.

The vector space viewpoint and notation will mostly be taken from P. R. Halmos's text [6]. My own introduction to the coordinate-free approach came from discussions with L. J. Savage and I acknowledge my great debt to him, a debt of which coordinate freedom forms but a part.

2. Gauss-Markov estimation from a coordinate-free viewpoint

We consider a sample point, Y, that ranges over an n-dimensional real vector space, V, on which an inner product, $(,)$, is given. (It would also be possible to start without a given inner product and to define one in terms of the covariance structure of Y.) Perhaps more basically, Y is a function from an underlying probability space onto V such that all sets in the underlying space of form $\{e|(x, Y(e)) \leqq c\}$, where $x \in V$ and c is a real number, are measurable. Of course, Y is the abstract entity usually corresponding, in a particular problem, to the coordinate vector comprising the set of scalar observations; by not writing in terms of coordinates, that is, by not requiring that a basis be specified, we are able to present the general theory succinctly.

All first and second moments discussed will be assumed to exist without further mention.

Clearly, $E(x, Y)$ is a linear functional of $x \in V$, and hence there exists a unique member of V, say μ, such that $E(x, Y) = (x, \mu)$ for all $x \in V$. Call μ the (vector) expectation of Y, EY; this quantity is easily articulated with the vector expectation in coordinate form.

Similarly Cov $[(x, Y), (z, Y)]$, where $x, z \in V$, is clearly a quasi-inner product (like an inner product but possibly nonnegative definite). Hence there exists a unique linear transformation Σ on V such that

$$\text{Cov}\,[(x, Y), (z, Y)] = (x, \Sigma z), \qquad x, z \in V. \tag{2.1}$$

It is easily seen that Σ is nonnegative definite and symmetrical with respect to $(,)$; that is, $(x, \Sigma x) \geqq 0$ and $(x, \Sigma z) = (\Sigma x, z)$ for all $x, z \in V$. Naturally, Var $(x, Y) = (x, \Sigma x)$. Let us say that Y is *weakly spherical* if Σ is a (nonnegative) multiple of I, the identity transformation. Call Σ the covariance operator of Y Cov Y; this quantity is easily articulated with the covariance matrix in coordinate form, especially for an orthonormal basis.

It is easily checked that the definition of EY does not depend on the inner product used, while the definition of Cov Y does depend on the inner product. The usual manipulative rules are immediate. For example, $E(AY) = A(EY)$, where A is a linear transformation; $\text{Cov}(AY) = A(\text{Cov } Y)A'$, where A' is defined by $(x, Az) = (A'x, z)$ for $x, z \in V$.

The standard Gauss-Markov statistical model makes the following assumptions:

(i) $\text{Cov } Y = \sigma^2 I$, where $\sigma^2 \geqq 0$ is unknown, that is, Y is weakly spherical;
(ii) $EY = \mu \in \Omega$, a *given* p-dimensional linear manifold.

About (i): first, it would be possible to assume $\text{Cov } Y = \sigma^2\Sigma_0$, where Σ_0 is a given linear transformation that is nonnegative definite and symmetric with respect to (,), for a linear transformation would then bring us back to the form stated above; second, if σ^2 is known, the Gauss-Markov linear estimation theory, relating to μ, is virtually unchanged (Student distributions become normal and F distributions become chi square).

About (ii): first, if μ is assumed to lie, not in a linear manifold, but in a linear manifold translated so that it does not include the origin, that is, in a p-dimensional flat, a simple translation brings us back to the form stated above; second, other kinds of presentations of Gauss-Markov estimation describe Ω in various specific ways. It is characteristic of the coordinate-free approach that it abstracts from the various ways in which a linear manifold can be described.

In specific applications, Y is usually the coordinate vector of the sample observations, the inner product (,) is the "conventional" sum of cross-products of corresponding coordinates, (ii) says that the coordinates are uncorrelated and with equal variance, and (iii) says that the coordinate expectations are linearly related in a given way, but otherwise unrestricted. A specific kind of example will now be stated and subsequent developments will be applied to it. For clarity, the illustrative material has been inset.

> Let V consist of all $I \times J$ arrays $\{y_{ij}\}$ for $i = 1, 2, \cdots, I; j = 1, 2, \cdots, J$. Both I and J must be $\geqq 2$. Scalar multiplication and vector addition are defined in the obvious ways: $c\{y_{ij}\} = \{cy_{ij}\}$ and $\{y_{ij}\} + \{z_{ij}\} = \{y_{ij} + z_{ij}\}$. Let (,) be the conventional $(\{y_{ij}\}, \{z_{ij}\}) = \sum\sum y_{ij}z_{ij}$. The sample point is $\{Y_{ij}\}$. We suppose that $\text{Var } Y_{ij} = \sigma^2$ and $\text{Cov}(Y_{ij}, Y_{i'j'}) = 0$ for $(i, j) \neq (i', j')$.
>
> Let $EY_{ij} = \mu_{ij}$. Then our Ω is defined by the statement: all second differences of the μ_{ij} are zero, that is,
>
> (2.2) $$\mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'} = 0$$
>
> for all i, j, i', j'. Alternatively, one might describe Ω as follows: there exist numbers $\bar{\mu}$, α_i, and β_j such that $\mu_{ij} = \bar{\mu} + \alpha_i + \beta_j$ for $i = 1, \cdots, I; j = 1, \cdots, J$.

It is often thought desirable to add side conditions so that $\bar{\mu}$, the α_i, and the β_j are uniquely determined by the μ_{ij} (estimable, identifiable). A popular set of side conditions is $\sum \alpha_i = \sum \beta_j = 0$ for all i, j. Then it follows that $\bar{\mu} = \sum \sum \mu_{ij}/(IJ)$, that $\alpha_i = (\sum_j \mu_{ij}/J) - \bar{\mu}$, and that $\beta_j = (\sum_i \mu_{ij}/I) - \bar{\mu}$.

What we have described is the model for two-way Model I analysis of variance, with one observation per cell and no interaction. If, in addition, we require joint normality of the Y_{ij}, then the Y_{ij} are independent.

A basic fact is that, under conditions (i) and (ii), the orthogonal projections, $P_\Omega Y$ of Y on Ω and $Y - P_\Omega Y = Q_\Omega Y$ on the orthogonal complement of Ω, are uncorrelated and have weakly spherical distributions in their *own* subspaces with (restricted) covariance transformations $\sigma^2 I$, where σ^2 is the same as that for Y itself. To say that AY and BY are uncorrelated is to say that Cov $[(x, AY), (z, BY)] = 0$ for all $x, z \in V$. This immediately extends to orthogonal decompositions of Y to more than two components. Since P_Ω and Q_Ω are orthogonal projections, they are idempotent (for example, $P_\Omega P_\Omega = P_\Omega$) and symmetric [for example, $P'_\Omega = P_\Omega$ or, equivalently, $(x, P_\Omega z) = (P_\Omega x, z)$ for all $x, z \in \Omega$] with respect to (,).

If we require normality of Y, that is, if we require that (x, Y) be normal for all $x \in V$, then it is almost immediate that $\|P_\Omega Y - \mu\|^2/\sigma^2$ and $\|Q_\Omega Y\|^2/\sigma^2$ have independent chi-square distributions with $p = \dim \Omega$ and $n - p = \dim \Omega^\perp$ degrees of freedom respectively. In any case, these quantities have expectations p and $n - p$.

The vector Gauss-Markov estimator of μ is $P_\Omega Y$ and the scalar Gauss-Markov estimator of a linear functional (x, μ) of μ is $(x, P_\Omega Y) = (P_\Omega x, Y)$. (Commentary on the historical accuracy or inaccuracy of the designation "Gauss-Markov" appears in the discussion of [10].) These Gauss-Markov estimators are characterized by the following well-known properties. To avoid trivialities, we assume that $\sigma^2 > 0$.

(a) $P_\Omega Y$ is the unique linear transformation of Y that is an unbiased estimator of μ and leads to minimum variance for all derived estimators of linear functionals. In other words,

(a_1) $E(P_\Omega Y) = \mu$ for all $\mu \in \Omega$,

(a_2) For all x, and for linear transformations $D \neq P_\Omega$ satisfying $E(DY) = \mu$, all $\mu \in \Omega$, it follows that Var $(x, P_\Omega Y) <$ Var (x, DY).

(b) For all x, the unique minimum variance linear functional of Y that is unbiased for (x, μ) is $(P_\Omega x, Y) = (x, P_\Omega Y)$.

(c) For all x, the unique minimum variance linear functional of Y that has bounded mean square error in μ is $(P_\Omega x, Y)$.

(d) The unique vector $\tilde{\mu}$ in Ω minimizing $\|Y - \tilde{\mu}\|^2$ is $P_\Omega Y$. This is the least squares characterization.

(e) For all x, the unique linear functional of Y whose "coefficient vector," when the functional is expressed in the form (z, Y), lies in Ω and which estimates (x, μ) unbiasedly is $(P_\Omega x, Y)$; that is, for $w \in \Omega$, (w, Y) is the Gauss-Markov estimator of its expectation (w, μ). This characterization often leads to an easy method of obtaining Gauss-Markov estimators when there is high symmetry or when Ω is similar to another manifold onto which we know how to project orthogonally. For if we guess $(P_\Omega x, Y)$ by choosing a vector $w \in \Omega$, we need only compute (w, μ) to see if our guess is right and, if it is nearly right, we can often see immediately how it should be modified. A similar vector characterization for $P_\Omega Y$ is readily written down.

(f) When Y is normal, $P_\Omega Y$ and $(x, P_\Omega Y)$ are the maximum likelihood estimators of μ and (x, μ) respectively. Further, $(x, P_\Omega Y)$ is *the* minimum variance unbiased estimator of (x, μ).

Various other characterizations and properties of $P_\Omega Y$ can be stated, for example, in terms of invariance under relevant linear transformations. Note that Gauss-Markov estimation is linear, that is, that $\alpha(x, P_\Omega Y) + \beta(z, P_\Omega Y) = (\alpha x + \beta z, P_\Omega Y)$, the Gauss-Markov estimator of $(\alpha x + \beta z, \mu)$. Further, in terms of a fixed basis, the coordinates of $P_\Omega Y$ are the Gauss-Markov estimators of the respective coordinates of μ. The conventional (unbiased) estimator of σ^2 is $||Q_\Omega Y||^2/(n - p)$.

Turn now to our example, with the description of Ω in the form $\mu_{ij} = \bar{\mu} + \alpha_i + \beta_j$, where $\sum \alpha_i = \sum \beta_j = 0$. Let a bar denote simple averaging with respect to dotted subscripts and consider linear functionals of Y as follows: $\bar{Y}_{..}$, $\bar{Y}_{i.} - \bar{Y}_{..}$, and $\bar{Y}_{.j} - \bar{Y}_{..}$. Note that there are I of the second kind and J of the third. The coefficient vectors of these functionals are easily seen to be in Ω. Hence the functionals are the Gauss-Markov estimators of their own expectations, $\bar{\mu}$, α_i, and β_j. Hence $P_\Omega Y$ is the coordinate vector with (i, j) component $\bar{Y}_{..} + (\bar{Y}_{i.} - \bar{Y}_{..}) + (\bar{Y}_{.j} - \bar{Y}_{..}) = \bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..}$. There is a standard orthogonal decomposition of $P_\Omega Y$ into three vectors corresponding to "over-all mean," "row effects," and "column effects," which I do not discuss here. The dimension of Ω is readily seen to be $I + J - 1$.

If ω is a q-dimensional linear manifold within Ω, then the standard F statistic for the null hypothesis $\mu \in \omega$ against all alternatives is

$$\frac{(p - q)^{-1}(||Q_\omega Y||^2 - ||Q_\Omega Y||^2)}{(n - p)^{-1}||Q_\Omega Y||^2} = \frac{(p - q)^{-1}||P_{\Omega - \omega} Y||^2}{(n - p)^{-1}||Q_\Omega Y||^2}, \tag{2.3}$$

with large values critical. Here $Q_\omega Y = Y - P_\omega Y$, and $\Omega - \omega$ is the orthogonal complement of ω with respect to Ω. This test statistic has, under the null hypothesis, the central F distribution with $p - q$ and $n - p$ degrees of freedom. The geometrical interpretation of the F statistic is well known.

The purpose of this paper is to discuss estimation, not testing. Hence, except for a few scattered remarks, we leave testing with the above brief paragraph. A good bit of the standard literature on Gauss-Markov estimation and Model I analysis of variance may be interpreted in terms of forming and combining orthogonal projections, especially when Ω is regarded as the direct sum of several linear manifolds with interesting statistical meanings.

3. Missing observations

Several topics within the general framework of Gauss-Markov estimation take the following form. Suppose that we have an explicit expression for $P_\Omega Y$ (or its components) in a specific problem; now suppose that the actual linear manifold of interest is not Ω, but another manifold nearly the same as Ω. Can we find an explicit expression for the orthogonal projection onto the new manifold by using our knowledge of $P_\Omega Y$, and without considering the new problem ab initio?

In terms of systems of linear equations, like the normal equations that one considers in a scalar approach to Gauss-Markov estimation, the kind of problem considered here is: if we know explicitly how to solve a set of linear equations, can we easily find the solution of a slightly modified set without treating it ab initio? Naturally, one may also phrase this in terms of inverting slightly modified matrices ([9], [10], and [12]).

The analysis of covariance may be regarded from this viewpoint, among others. Here, V stays the same but the modified Ω is spanned by the vectors of the old Ω, plus (hopefully a few) new vectors not in the old Ω.

Our present interest, however, is in another problem of the same general nature, that of missing observations. Here both V and Ω change.

We suppose that Y is incompletely observed in the following sense: write V as the direct sum of two orthogonal linear manifolds, V_1 and V_2, that is, $V = V_1 + V_2$. Let $Y^1 = P_{V_1}Y$ and $Y^2 = P_{V_2}Y$, where, as before, P_{V_i} means orthogonal projection on V_i. The interpretation is that only Y^1 will be observed, and the approach will primarily be useful when dim V_2 is small.

Thus, if Y is a coordinate vector of scalar observations, and some of these observations are missing, we are in the situation described above, with V_1 the coordinate manifold corresponding to the nonmissing observations. The description, however, covers more general cases that may be called instances of *mixed-up* observations. Suppose, for example, that Y is the coordinate vector of observations and that, although some observations are missing, their sum is known. Then Y^1 may be taken as the coordinate vector with each of the missing observations replaced by their mutual arithmetic average. In this case, V_1 is spanned by the coordinate axes for the known observations, together with the equiangular diagonal for the remaining coordinate axes.

It is assumed that the marginal distribution of Y^1 is the same with and without the missing or mixed-up observations. This is an important, but seldom discussed, matter in applications for, in some experimental setups where observations are randomly missing, the conditional distribution of the remaining observations, given that some observations are missing, may depend on the missing ones. For example, if the observations are weight gains in litters of baby pigs, the accidental death of one pig may alter the amount of maternal milk available to the others and hence may affect their weight gains.

A related point is that, in some cases, the probability that an observation will be missing depends on the value it would have had if it were present. For example, the sickly piglet is probably more likely to die accidentally before the end of the experiment than is his robust brother. In such a case we are dealing with a complicated selection process.

I do not intend to discuss these important problems further here. Instead we shall assume that the experimental design is for Y^1 from the start and that, if observations are missing by chance, it is meaningful to carry out statistical analyses conditionally on the "observed" V_1.

Some notation is needed. Let $\Omega_i = P_{V_i}\Omega$ for $i = 1, 2$. Let

$$\mu^i = P_{V_i}\mu = EY^i, \qquad i = 1, 2, \tag{3.1}$$

so that $\Omega_1 \perp \Omega_2$ and $\mu^1 \perp \mu^2$. In general, $\Omega \neq \Omega_1 + \Omega_2$, although $\Omega \subset \Omega_1 + \Omega_2$.

We make the basic assumption that

$$\dim \Omega = \dim \Omega_1, \tag{3.2}$$

and we ask how to find $P_{\Omega_1}Y^1$, or its equivalent, in terms of P_Ω, which we suppose known explicitly. (In the following, I permit myself the customary ambiguity of using "μ^1," say, both to denote the unknown true $\mu^1 = EY^1$ and to serve as a running variable over Ω_1.) Since $\dim \Omega = \dim \Omega_1$, to each $\mu^1 \in \Omega_1$ there corresponds a unique $\mu^2 \in \Omega_2$, say $A\mu^1$, such that $(I + A)\mu^1 \in \Omega$, $P_{\Omega_1}(I + A)\mu^1 = \mu^1$, and $P_{\Omega_2}(I + A)\mu^1 = A\mu^1$. In short, to each μ^1, there corresponds a unique $\mu \in \Omega$ with $P_{\Omega_1}\mu = \mu^1$; $A\mu^1$ is that μ minus μ^1. For completeness, take $Ax = 0$ when $x \perp \Omega_1$ so that A is a well-defined transformation. It is readily shown to be linear.

Instead of seeking $P_{\Omega_1}Y^1 = \hat{\mu}^1$ as such, it is equivalent and more convenient to seek its analogue $(I + A)P_{\Omega_1}Y^1 = \hat{\mu}$ in Ω. Note that, if we know $\hat{\mu}^2 = AP_{\Omega_1}Y^1$, we could easily obtain $\hat{\mu}$ since

$$\hat{\mu} = P_\Omega(Y^1 + \hat{\mu}^2) \tag{3.3}$$

and since we know P_Ω explicitly. The proof is simple, for

$$\begin{aligned} P_\Omega(Y^1 + \hat{\mu}^2) &= P_\Omega(P_{\Omega_1}Y^1 + \hat{\mu}^2) + P_\Omega(Y^1 - P_{\Omega_1}Y^1) \\ &= P_\Omega(\hat{\mu}^1 + \hat{\mu}^2) = \hat{\mu}, \end{aligned} \tag{3.4}$$

where the last expression on the right of the first line is zero since $Y^1 - P_{\Omega_1}Y^1$ is obviously orthogonal to both Ω_1 and Ω_2 and hence to Ω.

Next, note that $\hat{\mu}^2$ can be obtained via the following consistency condition which amounts, in any specific case, to a set of simultaneous linear equations in $\dim \Omega_2$ scalar unknowns:

$$\hat{\mu}^2 = P_{\Omega_2}P_\Omega(Y^1 + \hat{\mu}^2). \tag{3.5}$$

This condition has great intuitive appeal for it says that $\hat{\mu}^2$ is that element of Ω_2 which, when added to Y^1, followed by orthogonal projection on Ω and then on Ω_2, gives us $\hat{\mu}^2$ back again. It is easy to see that (3.5) holds, for $\hat{\mu}^2 = P_{\Omega_2}\hat{\mu}$, and we need only substitute for $\hat{\mu}$ from (3.3). That $\hat{\mu}^2$ is determined uniquely by (3.5) may be seen thus. Suppose that u and u_* both are in Ω_2 and satisfy (3.5).

Then $u - u_* = P_{\Omega_2} P_\Omega (u - u_*)$. But $||u - u_*|| > ||P_{\Omega_2} P_\Omega (u - u_*)||$ unless both $u - u_* \in \Omega$ and $P_\Omega(u - u_*) \in \Omega_2$. Hence $P_\Omega(u - u_*)$ is in both Ω and Ω_2; it is therefore zero and hence $u = u_*$.

The use of (3.5) in practice is simple for the kind of application envisaged since P_Ω is known explicitly and P_{Ω_2} presents little difficulty if $\dim \Omega_2$ is small.

Let us turn to our example and suppose that Y_{IJ} is missing, that is, that Y^1 has components like Y except that the (I, J) component is zero. Further, μ^1 has (i, j) components $\bar{\mu} + \alpha_i + \beta_j$ as before, except that the (I, J) component is zero; μ^2 has all components zero except the (I, J) one which is $\bar{\mu} + \alpha_I + \beta_J$. It is readily checked that $\dim \Omega = \dim \Omega_1 = I + J - 1$.

To find $\hat{\mu}^2$ is to find $\hat{\mu}_{IJ}$, which I shall call y for brevity. Then $P_{\Omega_2} P_\Omega [Y^1 + \hat{\mu}^2]$ has all components zero but the (I, J) one, which we know from the previous discussion is

$$\frac{Y_{I\odot} + y}{J} + \frac{Y_{\odot J} + y}{I} - \frac{Y_{\odot\odot} + y}{IJ}, \tag{3.6}$$

where a circled dot subscript means summation over that subscript for all possible nonmissing observations. The basic identity (3.5) here says that the above displayed quantity is y. Hence, solving the resulting linear equation,

$$y = \frac{I}{I-1} \frac{J}{J-1} \left(\frac{Y_{I\odot}}{J} + \frac{Y_{\odot J}}{I} - \frac{Y_{\odot\odot}}{IJ} \right). \tag{3.7}$$

As a check, we may compute the expectation of this quantity, $\bar{\mu} + \alpha_I + \beta_J$. Having found $y = \hat{\mu}_{IJ}$, by (3.3) we may use this value to "complete" Y^1 and apply P_Ω to the completion, thus getting $\hat{\bar{\mu}}$, $\hat{\alpha}_i$, and $\hat{\beta}_j$. It is straightforward to write down explicit descriptions of these quantities.

Other methods of treating missing observations are (i) minimization of the quadratic form $||Q_\Omega [Y^1 + \hat{\mu}^2]||^2$ in $\hat{\mu}^2$, which leads to (3.5) again, and (ii) the use of traditional covariance analysis with dummy covariate vectors, each having a coordinate one for the missing observation to which it corresponds and zero coordinates elsewhere. (This last method requires caution and modification if $\dim \Omega_2 < \dim V_2$, that is, if the expectations of the missing observations are linearly related.) The covariance method is based on the easily shown identity $P_{\Omega + \Omega_2} Y^1 = P_{\Omega_1} Y^1$.

The major discussion of this section, in which the normal equations (3.5) for $\hat{\mu}^2$ are obtained directly, may be regarded as the coordinate-free analogue of a paper by Wilkinson [11].

If we are concerned with F testing and want to use the approach of this section toward missing observations, then we require $\dim \omega = \dim P_{V_1} \omega$ as well as (3.2). Let $\omega_1 = P_{V_1} \omega$. Equation (3.5) must be worked out separately for $||Q_{\Omega_1} Y||^2$ and $||Q_{\omega_1} Y||^2$ in (2.3). A suggestion first made, I believe, by Yates is to approximate the F statistic by solving (3.5) for $\hat{\mu}^2$ under Ω_1 only and using this quantity throughout the F statistic. The resulting $||Q_\omega Y||^2$ is too large so that, if the "Yates approximation" leads to an F value that is not statistically significant, one can be sure that the correct F value would also not be statistically significant. The converse is certainly not true. In some problems the "Yates approximation" is very good and in others it is poor. I know of no general way of deciding in advance.

4. Extra observations

Now, using the same notation, suppose it is $P_{\Omega_1}Y^1$ that is known explicitly and $P_\Omega Y$ or, equivalently, $P_{\Omega_1}P_\Omega Y$ that is wanted. This might happen if we knew the Gauss-Markov estimators explicitly for some experimental design, and wanted to know the Gauss-Markov estimators for a design related to the prior one by the addition of a few observations. Thus Y is observable and $P_\Omega Y$ is our goal. We assume no change in the dimension of Ω (3.2), and so to know $P_{\Omega_1}P_\Omega Y$ is to know $P_\Omega Y$. Often the former expression is more convenient.

It is useful to partition Ω_1 into cosets. Recall that A takes μ^1 into the corresponding element of Ω_2. Let U be the subspace of Ω_1 such that $AU = 0$ so that U is the intersection of Ω_1 and the null space of A. Let $\Omega_1 - U$ be the orthogonal complement of U with respect to Ω_1. Thus A represents a one to one linear transformation between $\Omega_1 - U$ and Ω_2. We have partitioned Ω_1 into cosets of the form $U + z$ for $z \in \Omega_1 - U$, where each coset contains those members of Ω_1 going into a specific member of Ω_2.

In concrete examples, $\Omega_1 - U$ is usually easy to describe explicitly. For $u \in \Omega_1$ will usually be described parametrically and to say $Au = 0$ is to say that dim Ω_2 linear equations of form $(x_i, u) = 0$ for $i = 1, \cdots,$ dim $\Omega_2 =$ dim $(\Omega_1 - U)$ hold. Thus the $P_{\Omega_1}x_i$ will span $\Omega_1 - U$.

Since $\Omega = U + (I + A)(\Omega_1 - U)$ and since the two summands are orthogonal,

$$(4.1) \qquad \begin{aligned} P_\Omega Y &= P_\Omega Y^1 + P_\Omega Y^2 \\ &= P_U Y^1 + P_{(I+A)(\Omega_1 - U)}Y^1 + P_U Y^2 + P_{(I+A)(\Omega_1 - U)}Y^2. \end{aligned}$$

Since $U \subset V_1$ and $V_1 \perp V_2$, $P_U Y^2 = 0$. Using the relationship $U + (\Omega_1 - U) = \Omega_1$ and recombining, we have

$$(4.2) \qquad P_\Omega Y = P_{\Omega_1}Y^1 - P_{\Omega_1 - U}Y^1 + P_{(I+A)(\Omega_1 - U)}Y,$$

whence, finally,

$$(4.3) \qquad P_{\Omega_1}P_\Omega Y = P_{\Omega_1}Y^1 - P_{\Omega_1 - U}Y^1 + P_{\Omega_1}P_{(I+A)(\Omega_1 - U)}Y.$$

The first summand we already know explicitly, while the second and third require orthogonal projection only onto manifolds of dimension dim Ω_2, which is small in the envisaged applications. The first two summands together form $P_U Y^1$, but is typically easier to compute this in the subtractive form of (4.3).

An interpretation of (4.3) is the following: the first two terms on the right estimate $P_U\mu$ from Y^1, as it seems clear that Y^2 can tell us nothing about $P_U\mu$, while the last term estimates $P_{\Omega_1-U}\mu$ from the full Y. Then the two vector estimators of orthogonal quantities are added.

It is often easier to work in terms of estimating linear functionals (x, μ) of μ and, in many cases, x is already in Ω_1. Then the Gauss-Markov estimator of (x, μ) is

$$(4.4)\qquad (x, P_{\Omega_1}P_\Omega Y) = (P_\Omega x, Y) = (x, Y^1) - (P_{\Omega_1-U}x, Y^1) + (P_{(I+A)(\Omega_1-U)}x, Y),$$

for $x \in \Omega_1$.

If, in addition, $\dim \Omega_2 = 1$, then (4.4) may be simplified further, as follows. Suppose that $\Omega_1 - U$ is spanned by z (and Ω_2 is spanned by Az). Straightforward computation gives

(4.5)

$$(x, P_{\Omega_1}P_\Omega Y) = (x, Y^1) - \frac{(x, z)(z, Y^1)}{||z||^2} + \frac{(x, z)[(z, Y^1) + (Az, Y^2)]}{||z||^2 + ||Az||^2}$$
$$= (x, Y^1) - \frac{(x, z)}{||z||^2 + ||Az||^2}\left[\frac{||Az||^2}{||z||^2}(z, Y^1) - (Az, Y^2)\right]$$

for $x \in \Omega_1$ and z spanning $\Omega_1 - U$. In applications, z is usually obtained easily as the coefficient vector of Y^1 for the Gauss-Markov estimator of the expectation of the single nonzero coordinate of μ^2, this for V_1 and Ω_1, that is, for no extra observation.

Let us illustrate the use of (4.3) to (4.5) in our example, supposing that there is an extra observation Y_{IJ2} in the (I, J) cell. Now V_1 is IJ-dimensional and V is $(IJ + 1)$-dimensional. Let Ω_1 have the same coordinates as before except that the new $(IJ + 1)$st coordinate is zero. Let $\Omega_2 = V_2$ have all coordinates zero except the $(IJ + 1)$st. Then Au for $u \in \Omega_1$ is that vector in Ω_2 whose $(IJ + 1)$st coordinate is the same as the (I, J) coordinate of u. Hence to say that $Au = 0$ is to say that $\bar{\mu}(u) + \alpha_I(u) + \beta_J(u) = 0$. (Here I have kept u in the expression as an argument to emphasize that $\bar{\mu}$, α_I, and β_J are functionals of u.) Hence U is defined by $(x, u) = 0$ where x has all coordinates zero except that the (I, J) coordinate is one. Project x on Ω_1 orthogonally to obtain the vector in Ω_1 with (i, j) coordinate

$$(4.6)\qquad J^{-1}\delta_{iI} + I^{-1}\delta_{jJ} - (IJ)^{-1},$$

where the δ's are Kronecker deltas. This vector z spans $\Omega_1 - U$, which is one-dimensional. The transformation $I + A$, applied to z, takes it into another vector, which is the same except that the $(IJ + 1)$st coordinate 0 of z becomes $J^{-1} + I^{-1} - (IJ)^{-1} = \lambda$, say, equal to the common (I, J) coordinate. It is readily computed that $||z||^2 = \lambda$ and $||(I + A)z||^2 = \lambda(1 + \lambda)$.

Note that z could also be obtained as follows: the Gauss-Markov estimator, with no extra observation, of $\bar{\mu} + \alpha_I + \beta_J$ is $\bar{Y}_{I\cdot} + \bar{Y}_{\cdot J} - \bar{Y}_{\cdot\cdot}$. This linear functional of Y^1 has z as its coefficient vector.

Suppose that we want to find the Gauss-Markov estimator of α_1. We know that $\bar{Y}_{1\cdot} - Y_{\cdot\cdot}$ is that estimator if there were no extra observation, hence we may conveniently write α_1 as (x, μ), where $x \in \Omega_1$ is the vector with (i, j) coordinate $J^{-1}\delta_{i1} - (IJ)^{-1}$ and with $(IJ + 1)$st coordinate zero. Compute directly that

$$
(4.7) \qquad \begin{aligned}
(x, Y^1) &= \bar{Y}_{1\cdot} - \bar{Y}_{\cdot\cdot}, \\
(x, z) &= -(IJ)^{-1} \\
\|z\|^2 + \|Az\|^2 &= \|(I + A)z\|^2 = \lambda(1 + \lambda), \\
\|Az\|^2 &= \lambda^2, \\
(z, Y^1) &= \bar{Y}_{I\cdot} + \bar{Y}_{\cdot J} - \bar{Y}_{\cdot\cdot}, \\
(Az, Y^2) &= \lambda Y_{IJ2}.
\end{aligned}
$$

Hence the Gauss-Markov estimator of α_1 is

$$
(4.8) \qquad \bar{Y}_{1\cdot} - \bar{Y}_{\cdot\cdot} + [(1 + \lambda)IJ]^{-1}[(\bar{Y}_{I\cdot} + \bar{Y}_{\cdot J} - \bar{Y}_{\cdot\cdot}) - Y_{IJ2}].
$$

Note that the last term in square brackets is the difference between the Gauss-Markov estimators of $\bar{\mu} + \alpha_I + \beta_J$ from Y^1 and Y^2 alone, respectively. Call this quantity Δ and observe that $E\Delta = 0$. Since $(1 + \lambda)IJ = IJ + I + J - 1$, we see that

$$
(4.9) \qquad \hat{\alpha}_1 = \bar{Y}_{1\cdot} - \bar{Y}_{\cdot\cdot} + \frac{\Delta}{IJ + I + J - 1}.
$$

A simple interchange of indexes gives us $\hat{\alpha}_2, \cdots, \hat{\alpha}_{I-1}; \hat{\beta}_1, \cdots, \hat{\beta}_{J-1}$. Similar computations provide $\hat{\alpha}_I$, $\hat{\beta}_J$, and $\hat{\mu}$. The final results may be summarized as follows:

$$
(4.10) \qquad \begin{aligned}
\hat{\mu} &= \bar{Y}_{\cdot\cdot} - \frac{\Delta}{IJ + I + J - 1}, \\
\hat{\alpha}_i &= \bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot} - \frac{\Delta(I\delta_{iI} - 1)}{IJ + I + J - 1}, \\
\hat{\beta}_j &= \bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot} - \frac{\Delta(J\delta_{jJ} - 1)}{IJ + I + J - 1}, \\
\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j &= \bar{Y}_{i\cdot} + \bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot} - \frac{\Delta(I\delta_{iI} + J\delta_{jJ} - 1)}{IJ + I + J - 1},
\end{aligned}
$$

thus permitting explicit expression of $P_{\Omega_1}P_{\Omega}Y$ as $P_{\Omega_1}Y^1$ plus a "correction" term.

Note an apparent lack of symmetry between Y_{IJ} and Y_{IJ2} in the above expressions. That this is only apparent may be seen by computing the coefficient of Y_{IJ} in $\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j$; it is the same as that of Y_{IJ2}.

As with missing observations, the requisite manipulations should be carried out twice for F testing purposes, once for each of the two basic sums of squares.

Another approach to extra observations is that of combining $AP_{\Omega_1 - U}Y^1$ and $P_{\Omega_2}Y^2$ matrically with matrix weights given by their covariance structures [4].

5. Replications of a single original observation

The treatment of section 4 is quite general in that the natural coordinates of Y^2 need not have expectations that are the same as any of those of the Y^1 coordinates. As in the special case of the example, however, one might expect some simplification when Y^2 resembles Y^1 in any of various senses. I want now to treat one of those senses, that one in which there are $m - 1$ extra scalar observations, each having the same expectation as the same one of the "original" observations.

Here it is convenient to let V_1 correspond to the original observations *excluding* the replicated one, and to let V_2 correspond to the m-dimensional space of the replicated observation. The situation is distinguished by the special nature of Ω_2, which is one-dimensional regardless of m. Under the usual orthonormal basis, in a specific case, Ω_2 is the equiangular line in V_2. In general, we may say that Ω_2 is spanned by a nonzero vector s in V_2.

It is also convenient to write V_2 as the direct sum of two orthogonal subspaces, V_{21} and V_{22}. Let V_{21} be a one-dimensional subspace corresponding to a single observation of the replicated set and let V_{22} be the orthogonal $(m - 1)$-dimensional subspace corresponding to the rest of the replicated set. Let $s_i = P_{V_{2i}}s$ so that $s = s_1 + s_2$ and $||s||^2 = ||s_1||^2 + ||s_2||^2$. Further, let us write $Y^{21} + Y^{22} = Y^2$; $\Omega_{21} = V_{21}$, spanned by s_1; Ω_{22}, spanned by s_2; et cetera, with obvious intent.

Let A_1 be the transformation taking $\mu^1 = EY^1$ into $\mu^{21} = EY^{21}$ and such that $A_1x = 0$ for x orthogonal to Ω_1. Let A_2 be the corresponding transformation taking μ^1 into $\mu^{22} = EY^{22}$. Then $A = A_1 + A_2$ and the A's are linear. Since Ω_2, Ω_{21}, and Ω_{22} are all one-dimensional, we may write $A\mu^1 = (w, \mu^1)s$ and $A_i\mu^1 = (w, \mu^1)s_i$ for $i = 1, 2$, and an appropriate $w \in \Omega_1$.

The general, coordinate-free features of the present special case are the dimensionalities and orthogonalities described above, together with (5.2) below.

Suppose that we know explicitly how to find $P_{(I+A)\Omega_1}[Y^1 + Y^{21}]$, for this is the basic projection if there were exactly one of the replicated observations, $m = 1$. We ask whether there is a $y^{21} \in \Omega_{21}$ such that

$$P_{(I+A_1)\Omega_1}P_{\Omega}Y = P_{(I+A_1)\Omega_1}[Y^1 + y^{21}], \tag{5.1}$$

that is, whether we can carry out our known projection on a new sample point in $V_1 + V_{21}$, where the whole set of replicated observations is replaced by $y^{21} \in \Omega_{21}$, and thus obtain the projection on $(I + A_1)\Omega_1$ of $P_{\Omega}Y$. We may write $y^{21} = bs_1$ and seek b satisfying (5.1).

We assume that

$$\dim \Omega_1 = \dim (I + A_1)\Omega_1 = \dim (I + A)\Omega. \tag{5.2}$$

Note that $P_{\Omega_1}Y^1$ is easily obtained by the missing observation technique, since $\dim \Omega_{21} = 1$. In fact, $A_1P_{\Omega_1}Y^1$, the Gauss-Markov estimator from Y^1 alone of EY^{21}, is $(w, Y^1)s_1$ since $w \in \Omega_1$ and $E(w, Y^1) = (w, \mu^1) = \mu^{21}$. Further, $A_iw = ||w||^2s_i$ and $Aw = ||w||^2s$.

The relationship $(w, x) = 0$ for $x \in \Omega_1$ determines U. Hence, w spans the one-dimensional $\Omega_1 - U$. Also, $(I + A)w = w + ||w||^2s$ spans $(I + A)(\Omega_1 - U)$. Hence,

$$P_{\Omega_1}Y^1 = P_{\Omega_1}P_{(I+A_1)\Omega_1}[Y^1 + (w, Y^1)s_1],$$

$$P_{\Omega_1 - U}Y^1 = \frac{(w, Y^1)}{||w||^2}\, w, \tag{5.3}$$

$$P_{(I+A)(\Omega_1 - U)}Y = \frac{(w, Y^1) + ||w||^2(s, Y^2)}{||w||^2 + ||w||^4||s||^2}\,[w + ||w||^2s],$$

and from (4.3) or (4.2)

$$\begin{aligned} P_{(I+A_1)\Omega_1}P_{\Omega}Y &= (I + A_1)P_{\Omega_1}P_{\Omega}Y \\ &= P_{(I+A_1)\Omega_1}[Y^1 + (w, Y^1)s_1] \\ &\quad - \frac{(w, Y^1)}{||w||^2}\,[w + ||w||^2s_1] \\ &\quad + \frac{(w, Y^1) + ||w||^2(s, Y^2)}{||w||^2 + ||w||^4||s||^2}\,(w + ||w||^2s_1) \end{aligned} \tag{5.4}$$

since $s \perp \Omega_1$. Hence the desired $y^{21} \in \Omega_{21}$, if it exists, must satisfy

$$P_{(I+A_1)\Omega_1}y^{21} = (w, Y^1)P_{(I+A_1)\Omega_1}s_1 + \frac{(s, Y^2) - ||s||^2(w, Y^1)}{1 + ||s||^2||w||^2}\,(w + ||w||^2s_1). \tag{5.5}$$

Since y^{21} may be written bs_1 and since $(I + A_1)\Omega_1$ may be regarded as the direct sum of U and the manifold spanned by $w + ||w||^2s_1$, (5.5) may be written as the following scalar equation, omitting the common vector $w + ||w||^2s_1$.

$$b\,\frac{||s_1||^2}{1 + ||w||^2||s_1||^2} = \frac{(w, Y^1)||s_1||^2}{1 + ||w||^2||s_1||^2} + \frac{(s, Y^2) - ||s||^2(w, Y^1)}{1 + ||s||^2||w||^2}, \tag{5.6}$$

or, simplifying,

$$b = \frac{-||s_2||^2(w, Y^1) + ||s||^2(1 + ||w||^2||s_1||^2)[(s, Y^2)/||s||^2]}{||s_1||^2(1 + ||w||^2||s||^2)}. \tag{5.7}$$

Observe that this is a weighted average of (w, Y^1), the s_1 coordinate of the missing observation vector estimator from Y^1, and $(s, Y^2)/||s||^2$, the estimator of the same quantity from Y^2. The weights are proportional to $-||s_2||^2$ and $||s||^2(1 + ||w||^2||s_1||^2)$ respectively.

In the usual applications, with coordinates for an orthonormal basis, $||s||^2 = m$ and $||s_1||^2 = 1$ so that the weights become, respectively,

$$\frac{1 - m}{1 + m||w||^2}, \qquad m\,\frac{1 + ||w||^2}{1 + m||w||^2}. \tag{5.8}$$

Further, $||w||^2$ is $1/\sigma^2$ times the variance of the replicated observation's expectation, as estimated from Y^1. So we may summarize the result as follows, for the usual applications.

Let $\hat{y}$ be the Gauss-Markov estimator of the replicated observation's expectation from Y^1 alone. Let Var $\hat{y} = \theta^2\sigma^2$.

Let $\tilde{y}$ be the Gauss-Markov estimator of the same quantity from Y^2 alone. (It will be the arithmetic average of the replicated observations.)

Let y be the weighted average of $\hat{y}$ and $\tilde{y}$ with weights proportional to $(1 - m)$ and $m(1 + \theta^2)$ respectively. For Gauss-Markov estimation based on Y, treat y as if it were a single observation replacing all the replicated ones. Note that this is essentially a prescription for finding Gauss-Markov estimators when weak symmetry does not hold but rather the following conditions hold:

(i) the observations are uncorrelated;

(ii) all the observations have variance σ^2, except that one ($\tilde{y}$) has variance σ^2/m where m is known.

In the special case $m = 1$ (original design), the weights are 0 and 1, as they should be. In the special case $m = 0$, which, strictly speaking, is not covered by the above analysis but can be handled by similar methods, the weights are 1 and 0, again as they should be.

Let us apply this to our example. Here, V may be expressed as the $(IJ + m - 1)$-dimensional space $\{Y_{ij}$ for $i = 1, \cdots, I; j = 1, \cdots, J; (i, j) \neq (I, J); Y_{IJ1}; Y_{IJ2}, \cdots, Y_{IJm}\}$. The three groups of coordinates corresponding to V_1, V_{21}, and V_{22} are separated by semicolons. The Ω's are given via $EY_{ij} = \bar{\mu} + \alpha_i + \beta_j$ and $EY_{IJl} = \bar{\mu} + \alpha_I + \beta_J$ for $l = 1, \cdots, m$. We may take ε_1 as having all zero coordinates, except that the $(I, J, 1)$ coordinate is one; similarly, ε_2 has all zero coordinates, except that the last $m - 1$ coordinates are one.

Now w is given by the coefficient vector for y toward the end of section 3, that is, w is the vector in Ω_1 with (i, j) coordinate, when $(i, j) \neq (I, J)$, $[I/(I - 1)][J/(J - 1) [J^{-1}\delta_{iI} + I^{-1}\delta_{jJ} - (IJ)^{-1}]$; of course, w has zero (IJl) coordinates. Further, (w, Y^1), the estimator of $\bar{\mu} + \alpha_I + \beta_J$ from Y^1 alone, is $[I/(I - 1)][J/(J - 1)][J^{-1}Y_{I\odot} + I^{-1}Y_{\odot J} - (IJ)^{-1}Y_{\odot\odot}]$. Next, observe that $(s, Y^2)/||s||^2 = \sum_{l=1}^{m} Y_{IJl}/m = \overline{Y}_{IJ}$.

We compute that $||w||^2 = [I/(I - 1)][J/(J - 1)]\lambda(1 - \lambda) = (I + J - 1)/[(I - 1)(J - 1)]$ so that the weights are $1 - m$ and $mIJ/[(I - 1)(J - 1)]$ respectively. Hence

$$b = \frac{(1 - m)IJ[J^{-1}Y_{I\odot} + I^{-1}Y_{\odot J} - (IJ)^{-1}Y_{\odot\odot}] + IJ\sum_l Y_{IJl}}{(I - 1)(J - 1) + m(I + J - 1)}, \tag{5.9}$$

and we need only use this in place of all the Y_{IJl} and apply the symmetrical explicit formulas of ordinary Gauss-Markov estimation to the resulting IJ-fold balanced array.

It can be checked that, in the special case $m = 2$, this leads to the same results as in the last section.

Note further that correction type expressions, like those in section 4, are readily obtained when there are m of the Y_{IJl}. In fact, the expressions are just the same except that $\Delta = \bar{Y}_{I\cdot} + \bar{Y}_{\cdot J} - \bar{Y}_{\cdot\cdot} - \sum_2^m Y_{IJl}/(m-1)$ and that the $IJ + I + J - 1$ term in the denominators becomes here $[IJ/(m-1)] + I + J - 1$. As before, but with the present slightly changed notation, $\bar{Y}_{I\cdot}$ is interpreted as $J^{-1}[\sum_{j=1}^{J-1} Y_{Ij} + Y_{IJ1}]$.

These correction type expressions may be simpler for certain computations, but the approach in which b is substituted for all the Y_{IJl} is more nearly in the direction we next take.

6. Extensions of the substitute-observation method

In this section I consider situations in which Y^1 of section 5 is replicated k times, while Y^2 stays the same as it was, corresponding to an m-fold replication of a single scalar observation. From now on, unless otherwise specified, we interpret Y^1 as the k-fold replication of an "original" Y^1 like that considered in section 5.

6.1. *The case* $m \geqq k$. This can be thrown back onto the work of section 5 quite simply. Let V_1 and Y^1 correspond to the k-fold replication of the original Y^1, let V_{21} and Y^{21} correspond to k of the coordinates of Y^2, and let V_{22} and Y^{22} correspond to the remaining $m - k$ coordinates of Y^2. Then this case falls squarely into the general discussion of section 5, except that, under conventional orthonormal coordinates, $||s_1||^2 = k$.

The weights, therefore, become for conventional orthonormal coordinates,

$$\alpha = \frac{1 - (m/k)}{1 + m||w||^2}, \tag{6.1}$$

$$1 - \alpha = \frac{(m/k)(1 + k||w||^2)}{1 + m||w||^2}.$$

This can be reinterpreted as follows. Suppose that we have a design consisting of the *original* Y^1 plus the original Y^{21}, but where the coordinates of Y^1 have common variance σ^2/k and the single nonzero coordinate of Y^{21} has variance σ^2/m. (This amounts to working with the arithmetic average of each group of observations having a common expectation.) Then the weights of (6.1) are applicable to find a "substitute-observation" for the nonzero coordinate of Y^{21}. Since $||w||^2\sigma^2 = k||w||^2(\sigma^2/k)$, we consider $k||w||^2$ as the variance of the usual missing observation estimator, divided by σ^2/k, the common variance of the reinterpreted coordinates of Y^1.

6.2. *The case* $m < k$. We obtain the same end result here, but the details must be changed slightly. Let V_1 and Y^1 be the same as in the preceding case, let V_{21} and Y^{21} correspond to the m nonzero coordinates of the original Y^2, and let V_{22} be a $(k - m)$-dimensional subspace orthogonal to $V_1 + V_{21}$. This means enlarging the space with which we started. Think of a (nonobservable) Y^{22} in V_{22} with expectation in $\{s_2\}$, the one-dimensional subspace of V_{22}

spanned by a single vector that is set up to correspond with s_1, that is, $EY^{21} = (w, \mu^1)s_1$ and $EY^{22} = (w, \mu^1)s_2$.

If we could observe $Y^1 + Y^{21} + Y^{22}$, orthogonal projection on its Ω would be easy for we would have a k-fold replication of a design for which we have the Gauss-Markov estimators. Although we do not have Y^{22}, we can ask whether there is a $y^2 \in \Omega_2$ such that

$$P_{(I+A_1)\Omega_1}P_{\Omega}(Y^1 + y^2) = P_{(I+A_1)\Omega_1}(Y^1 + Y^{21}), \tag{6.2}$$

where $A_1\mu^1 = \mu^{21} = (w, \mu^1)s_1$. In this case, with the orthonormal coordinates we have in mind, $||s_1||^2 = m$ and $||s||^2 = k$.

The estimator of EY^2 from Y^1 alone is $(w, Y^1)s$ and from Y^{21} alone it is $[(s_1, Y^{21})/||s_1||^2]s$. We can write Ω as the direct sum of orthogonal subspaces, $U + (I + A)(\Omega_1 - U)$, just as before and express y^2 in the form bs. Carrying out the operations indicated by (6.2), we obtain

$$\alpha = \frac{||s_2||^2}{||s||^2(1 + ||w||^2||s_1||^2)}. \tag{6.3}$$

Then, putting in $||s_1||^2 = m$ along with $||s_2||^2 = k - m$ and $||s||^2 = k$, a legitimate operation for orthonormal coordinates, we finally obtain exactly (6.1) again. The final paragraph of the discussion of the case $m \geqq k$ holds verbatim.

6.3. *Summary of the preceding work of this section.* We may summarize as follows. Suppose that $P_{\Omega}[Y^1 + Y^2]$ is known explicitly, but that $Y = Y^1 + Y^2$ is not weakly spherical. Suppose further (i) that Y^1 ranges over an $(n - 1)$-dimensional space and in that space has the covariance transformation $\sigma^2 I$, and (ii) that Y^2 ranges over a one-dimensional space orthogonal to that of Y^1 and in that space has the covariance transformation $\tau^2\sigma^2 I$, where τ^2 is known [and rational].

Let $\hat{y}$ be the Gauss-Markov estimator, based on Y^1 alone, of the coordinate of EY^2 with respect to a unit vector t in the space of Y^2. Let Var $\hat{y} = \theta^2\sigma^2$.

Let $\tilde{y}$ be the coordinate of Y^2 with respect to the same unit vector.

Let y be the weighted average of $\hat{y}$ and $\tilde{y}$ with weights proportional to $1 - \tau^{-2}$ and $\tau^{-2}(1 + \theta^2)$ respectively.

Then $P_{\Omega}[Y^1 + yt]$ is the Gauss-Markov estimator of EY.

The special cases $\tau^2 = 1$ (original design) and $\tau^2 = \infty$ (Y^2 is missing) work out as they should. The special case $\tau^2 = 0$ (corresponding to $m = \infty$ or EY^2 known a priori) may be useful in some circumstances. The bracketed words, "and rational," above may be omitted by continuity of Gauss-Markov estimation in Σ.

A related discussion is given by Gauss [5], section 36 of *Theoria Combinationis. . . .*

7. Application to apparent outliers

One approach to apparently outlying observations is to apply some criterion and then either decide that the suspect observation is not an outlier and handle it in the usual way for analysis, or else decide that the observation *is* an outlier and omit it completely from analysis.

One might, however, consider intermediate positions in which a suspect observation is treated with a lower weight than the rest, that is, has an imputed variance higher than that of the other observations. To completely omit the observation is, in effect, to give it an infinite variance, but why go that far?

If there is only one such suspect observation, the method described above permits the relatively simple incorporation of the suspect observation into Gauss-Markov estimation, provided that the ratio of its imputed variance to the variance of the other observations is given.

REFERENCES

[1] M. S. Bartlett, "The vector representation of a sample," *Proc. Cambridge Philos. Soc.*, Vol. 30 (1933–34), pp.327–340.

[2] L. C. A. Corsten, "Vectors, a tool in statistical regression theory," *Meded. Landbouwhogeschool Wageningen*, Vol. 58 (1958), pp. 1–92.

[3] J. Durbin and M. G. Kendall, "The geometry of estimation," *Biometrika*, Vol. 38 (1951), pp. 150–158.

[4] D. A. S. Fraser, "On the combining of interblock and intrablock estimates," *Ann. Math. Statist.*, Vol. 28 (1957), pp. 814–816.

[5] C.-F. Gauss, *Méthode des Moindres Carrés*, Paris, Mallet-Bachelier, 1855; translation into French by J. Bertrand of Gauss's works on least squares. (A translation into English by H. F. Trotter is Technical Report No. 5, Statistical Techniques Research Group, Princeton, 1957.)

[6] P. R. Halmos, *Finite-Dimensional Vector Spaces*, Princeton, Van Nostrand Press, 1958 (2nd ed.).

[7] A. N. Kolmogorov, "On the proof of the method of least squares," *Uspehi Mat. Nauk*, Vol. 1 (1946), pp. 57–70. (In Russian.)

[8] W. Kruskal, "Discussion of the papers of Messrs. Anscombe and Daniel," *Technometrics*, Vol. 2 (1960), pp. 157–166. (Pages indicated also include discussion by T. S. Ferguson, J. W. Tukey, and E. J. Gumbel.)

[9] R. L. Plackett, "Some theorems in least squares," *Biometrika*, Vol. 37 (1950), pp. 149-157.

[10] K. D. Tocher, "The design and analysis of block experiments," *J. Roy. Statist. Soc., Ser. B*, Vol. 14 (1952), pp. 45–91. Discussion, pp. 91-100.

[11] G. N. Wilkinson, "Estimation of missing values for the analysis of incomplete data," *Biometrics*, Vol. 14 (1958), pp. 257–286.

[12] M. A. Woodbury, "Modified matrix functions and applications," unpublished paper presented at the IBM Seminar for Directors of Computing Centers, June 26, 1959.