MAXIMUM LIKELIHOOD ESTIMATION OF A LINEAR FUNCTIONAL RELATIONSHIP

By C. Villegas

Instituto de Matemática y Estadística, Universidad de la República, Montevideo, Uruguay

1. Introduction and summary. We shall consider the problem of estimating a linear functional relationship

$$\alpha + \beta_1\tau^1 + \cdots + \beta_p\tau^p = 0 \tag{1.1}$$

among p variables $\tau^1, \cdots, \tau^p$ when the observed values do not satisfy it because all of them are subject to errors or fluctuations (superscripts will, in general, be indexing symbols, not powers, in this paper). Geometrically, the problem is equivalent to fitting straight lines or planes to a series of q observed points when all the coordinates are subject to error. This problem has a long history. R. J. Adcock, in two papers written in 1877 [2] and 1878 [3], solves it by minimizing the sum of squares of the orthogonal distances from the points to the hyperplane (1.1). Adcock and many other authors used the model

$$y_i = \tau_i + \epsilon_i \qquad (i = 1, \cdots, q), \tag{1.2}$$

where y_i and τ_i are column vectors representing the observed and true points, and the errors ϵ_i are independent random vectors with mean value zero. Since the τ_i are points lying on the hyperplane (1.1), we have in matrix notation

$$\alpha + \beta\tau_i = 0 \qquad (i = 1, \cdots, q) \tag{1.3}$$

where β is a row vector with components $\beta_1, \cdots, \beta_p$. If we assume that the τ_i are independently drawn from a probability distribution, then the estimate of β obtained by Adcock is not consistent. In fact, in 1937, J. Neyman [21] pointed out that if the distribution of the true vectors τ_i and the errors ϵ_i is normal, then the distribution of the observed vectors y_i is also normal and, being determined by moments of the first two orders, it is not sufficient to determine the parameters α and β; the functional relationship (1.1) is, therefore, nonidentifiable. Several methods have been proposed to overcome this difficulty, for which the reader is referred to a recent general survey of the literature by A. Madansky [18], which also contains an extensive bibliography. One approach is to assume that we know the covariance matrix of the errors up to a numerical factor. As was shown, in general, by C. F. Gauss [8], [9], in the case of independently and normally distributed observations whose variances are known up to a numerical factor, the maximum likelihood estimate is simply the weighted least-squares estimate. This estimate of the linear functional relationship was obtained as early as 1879

Received January 25, 1960; revised April 18, 1961.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

by C. H. Kummell [15], for the case in which the components ϵ_i^h of the vectors ϵ_i are independently distributed with variances $\kappa^{ii}\sigma_{hh}$, where the κ^{ii} are known constants and the σ_{hh} are known only up to a numerical factor. Kummell found that his estimate coincides with the estimate proposed by Adcock only in the case in which all the variances are equal.

M. J. van Uven [24] considered the case in which the errors ϵ_i are independent and have the same multivariate normal distribution with a covariance matrix Σ which is known only up to a numerical factor. His method is essentially the following. He considers $\tau^1, \cdots, \tau^p$ as skew coordinates in a new, "isotropic" space in which the rectilinear orthogonal coordinates are independent and have the same variance. In the new space he then uses Adcock's principle of adjustment, namely, he chooses as the estimate the hyperplane which minimizes the sum of orthogonal distances. Later, T. Koopmans [14] showed that van Uven's estimate is the maximum likelihood estimate for the case considered. If the τ_i are assumed independently drawn from a probability distribution, the estimate of the linear functional relationship thus obtained is consistent, but the estimate of Σ converges in probability to $p^{-1}\Sigma$ (see also [16]). B. M. Dent [7] solved the maximum likelihood equations in the case in which Σ is not known, but, as was shown later by D. V. Lindley [16], her estimates are not consistent, and should, therefore, be rejected. More recently, J. Kiefer and J. Wolfowitz [13] showed that, under certain conditions of identifiability, when the τ_i have a probability distribution, the method of maximum likelihood, if properly applied, yields consistent estimates of both the linear functional relationship and the probability distribution of the τ_i. However, Kiefer and Wolfowitz do not give explicit expressions for the maximum likelihood estimates.

No difficulties with respect to the identifiability of the functional relationship or with respect to the consistency of the estimates arise if we can replicate the observations. The model is now, in matrix notation,

$$y_{ij} = \tau_i + \epsilon_{ij}. \tag{1.4}$$

In general we have n_i observed points for each of the true points τ_i, and it is assumed that not all the τ_i lie on a translated subspace of dimension smaller than $p - 1$. Obviously this implies that $q \geqq p$. This model has been considered previously by G. W. Housner and J. F. Brennan [12], J. W. Tukey [23] and by F. S. Acton [1] (see also [11] and [25]). If we assume that the errors ϵ_{ij} are independently and normally distributed with a known covariance matrix Σ, we lose nothing if we consider only the averages $y_{i\cdot} = n_i^{-1}\sum_j y_{ij}$. We have then the same model (1.2), with the only difference that y_i is replaced by $y_{i\cdot}$. If, however, the covariance matrix is not known, we can now obtain in the usual way an estimate S of Σ. F. S. Acton [1] suggested the use of S instead of Σ in the estimate obtained by the method of maximum likelihood in the case of known Σ. In this paper it will be shown that the estimate thus obtained is the maximum likelihood estimate when Σ is unknown.

If the design is a (in general incomplete) block design, we have, if the treatment i is applied on the block j,

$$y_{ij} = \tau_i + b_j + \epsilon_{ij}, \tag{1.5}$$

where b_j is a column vector representing the block effect. Considering the block effects b_j as unknown constants, we get in the usual way the intrablock estimates t_i of the treatment effects τ_i. Then, the same equation (1.2) still holds if y_i is replaced by t_i, but in general the errors ϵ_i and consequently also the estimates t_i will no longer be independent. If the design consists of r replications of a basic design, then the covariance of two errors ϵ_i, $\epsilon_{i'}$ will be given by

$$\text{cov}(\epsilon_i, \epsilon_{i'}) = \frac{\kappa^{ii'}\Sigma}{r} \tag{1.6}$$

where $\kappa^{ii'}$ are known coefficients and the matrix Σ is unknown.

In this paper maximum likelihood estimates for Σ and the parameters of the linear functional relationship will be found for the case in which (1.6) holds. It will be shown that the maximum likelihood estimates $\hat{\alpha}$, $\hat{\beta}$ in the case of unknown Σ are obtained from the corresponding estimates in the case of known Σ by simply replacing Σ by the linear regression estimate S. In the last Section it will be shown that if the maximum likelihood method is applied directly to the variables y_{ij} instead of the variables t_i and S, then the same estimate $\hat{\beta}$ is obtained, but the estimate of Σ is multiplied by $1 - k^{-1} + N^{-1}$, where k is the number of experimental units in each block and N is the total number of experimental units. All of the estimates obtained are consistent, with the exception only of the estimate of Σ obtained by the direct approach in the last Section, which converges to $(1 - k^{-1})\Sigma$.

2. The model. We shall consider now in more detail the intrablock analysis of a (in general incomplete) block design to which the additive model (1.5) applies. We shall assume that errors coming from different experimental units are independent, and that the errors coming from a single experimental unit have a multivariate normal distribution with zero means and covariance matrix $\Sigma = \{\sigma_{hh'}\}$. Therefore,

$$\text{cov}(\epsilon_{ij}^{h}, \epsilon_{ij}^{h'}) = \sigma_{hh'}, \tag{2.1}$$

where $\epsilon_{ij}^{h} (h = 1, \cdots, p)$ are the components of ϵ_{ij}. If we do not consider the linear functional relationship (1.1), the estimation of τ_i and Σ is simply a linear regression problem. In order to arrive at a unique solution $t_1, \cdots, t_q$, it is usual to add some arbitrary linear restriction, say,

$$\sum_i \omega_i t_i = 0, \qquad \sum_i \omega_i \neq 0. \tag{2.2}$$

where κ^{ii} are known coefficients and the matrix Σ is unknown.

It is known ([5] Section 8.2) that the linear regression estimates t_i are linear combinations of the observed vectors y_{ij}. If the design consists of r replications of a basic design, then the covariance of two estimates t_i, $t_{i'}$ is given by (1.6), where the $\kappa^{ii'}$ are known coefficients that depend on the basic design and the linear restriction (2.2). Since the errors are assumed to be normally distributed, it follows that the t_i have a multivariate distribution with means τ_i and covariances given by (1.6). Finally, the linear regression estimate of $\sigma_{hh'}$ is

$$s_{hh'} = \frac{1}{\nu} \sum_{ij}{}' e_{ij}^{h} e_{ij}^{h'} \tag{2.3}$$

where $\nu = N - q - b + 1$, b is the number of blocks, e_{ij}^{h} is the linear regression estimate of the error ϵ_{ij}^{h}, and the prime over the summation sign indicates that the sum must be extended over all pairs (i, j) such that treatment i appears on block j. It is known ([5], Section 8.2) that the estimated covariance matrix $S = \{s_{hh'}\}$ is independent of the t_i and has a Wishart distribution with mean value Σ and ν degrees of freedom.

We have then the following linear functional relationship model. The p-dimensional random variables $t_1, \cdots, t_q$ have a multivariate normal distribution, with means $\tau_1, \cdots, \tau_q$ that satisfy (1.3) and covariances

$$\operatorname{cov}(t_i, t_{i'}) = \frac{\kappa^{ii'}}{r} \Sigma \tag{2.4}$$

where r and the $\kappa^{ii'}$ are known coefficients, and Σ is unknown. The matrix S is an unbiased estimate of Σ, is independent of the t_i and has a Wishart distribution with a number of degrees of freedom ν which tends to infinity when $r \to \infty$; the quotient r/ν converging to a positive limit.

The matrix $K = \{\kappa^{ii'}\}$ is always nonnegative because, for a given h, $r^{-1}\sigma_{hh}K$ is the covariance matrix of the hth components $t_1^h, \cdots, t_q^h$ of $t_1, \cdots, t_q$. If the t_i are not subject to any linear restriction like (2.2), then for any h the distribution of $t_1^h, \cdots, t_q^h$ is of rank q (see for example [6], p. 297) and the matrix K is positive definite. If there is only one linear restriction (2.2), the matrix K is of rank $q - 1$ and

$$K\omega = 0, \tag{2.5}$$

where ω is the column vectors the components of which are $\omega_1, \cdots, \omega_q$ (see for instance [17]).

3. Covariance matrix known. We shall consider in the first place the case in which the covariance matrix Σ is a known positive definite matrix and the t_i are not subject to any linear restriction (and consequently K is a positive definite matrix).

From (2.4) it follows that the covariance matrix of all the variables t_i^h is $r^{-1}K \otimes \Sigma$, where the symbol $\otimes$ denotes the Kronecker product of two matrices. The determinant of this covariance matrix is $r^{-pq}|K|^p|\Sigma|^q$. Therefore the probability density function for $t_1^1, \cdots, t_q^p$ is, up to a numerical factor, equal to

$$|K|^{-\frac{1}{2}p}|\Sigma|^{-\frac{1}{2}q}e^{-\frac{1}{2}rQ}, \tag{3.1}$$

where, if $K^{-1} = \{\kappa_{ii'}\}$,

$$Q = \sum_{i,i'} \kappa_{ii'}(t_i - \tau_i)'\Sigma^{-1}(t_{i'} - \tau_{i'}).$$

We shall denote the trace of a matrix X by tr X. Then, in matrix notation,

$$Q = \operatorname{tr} \Sigma^{-1}(t - \tau)K^{-1}(t - \tau)', \tag{3.2}$$

where t is the $p \times q$ matrix the ith column of which is t_i, and similarly, τ is the $p \times q$ matrix the ith column of which is τ_i. The maximum likelihood estimates of α, β and τ are the values $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\tau}$ that minimize (3.2) subject to the conditions (1.3) which may be written

$$\alpha u + \beta\tau = 0, \tag{3.3}$$

where u is the row vector the q components of which are equal to 1. This problem was solved by Koopmans [14] for the particular case $K = I$. In what follows, unless otherwise specified, the symbols α, β and τ will denote mathematical variables (not true values or true parameters). In the first place we shall find the minimum of (3.2) subject to the condition (3.3), for given values of α and β. Since Σ and K are positive definite, uniquely defined positive square roots $\Sigma^{\frac{1}{2}}$ and $K^{\frac{1}{2}}$ exist (see [10], p. 166). Consider the change of variable

$$\delta = \Sigma^{-\frac{1}{2}}(t - \tau)K^{-\frac{1}{2}}. \tag{3.4}$$

Since tr XY = tr YX, (3.2) may be written, if δ_i is the ith column of δ,

$$Q = \operatorname{tr} \delta\delta' = \sum_i \delta_i'\delta_i, \tag{3.5}$$

that is, Q is the sum of the squares of the distances from the origin to the points δ_i. The condition (3.3) is, in the new variables,

$$\gamma_i - \beta\Sigma^{-\frac{1}{2}}\delta_i = 0, \qquad (i = 1, \cdots, q) \tag{3.6}$$

where γ_i is the ith component of $(\alpha u + \beta t)K^{-\frac{1}{2}}$. We have to minimize the sum of squares of the distances from the origin to the points δ_i, subject to the condition that each δ_i lies on the corresponding hyperplane (3.6). Note that these are, in general, q parallel hyperplanes. This was the principle from which M. J. van Uven [24] derived his estimates for the case $K = I$. Obviously, the minimum is reached when δ_i is on the perpendicular from the origin to the hyperplane (3.6) and, therefore,

$$\delta_i = \gamma_i \frac{\Sigma^{\frac{1}{2}}\beta'}{\beta\Sigma\beta'}.$$

Going back to the old variables, we have

$$t - \hat{\tau} = \frac{\Sigma\hat{\beta}'}{\hat{\beta}\Sigma\hat{\beta}'}(\hat{\alpha}u + \hat{\beta}t). \tag{3.7}$$

By substitution of (3.7) into (3.2) it follows that the minimum value of Q, for given values of α and β, is

$$Q_1 = \frac{(\alpha u + \beta t)\mathrm{K}^{-1}(\alpha u + \beta t)'}{\beta\Sigma\beta'} . \tag{3.8}$$

We shall now find the minimum value of Q_1 for a given β. By differentiation we have $(\alpha u + \beta t)\mathrm{K}^{-1}u' = 0$. If we set

$$w = \frac{\mathrm{K}^{-1}u'}{u\mathrm{K}^{-1}u'}$$

and $t. = tw = \sum_i t_i w_i$, where $w_1, \cdots, w_q$ are the components of w, we get

$$\hat{\alpha} = -\hat{\beta}t. . \tag{3.9}$$

Since Q_1 tends to $+\infty$ when $\alpha \to \pm\infty$, it follows that the value α given by (3.9) minimizes Q_1. From the definition of w we have $uw = \sum w_i = 1$, so that if, as happens with the usual matrices, K, the w_i are nonnegative, then $t.$ is a weighted average of the vectors t_i, with weights, w_i. If we write $\Delta t_i = t_i - t.$ and $\Delta t = t - t.u$, we have at the minimum $\alpha u + \beta t = \beta\Delta t$, and, therefore, the minimum value of Q_1 is

$$Q_2 = \frac{\beta F \beta'}{\beta\Sigma\beta'} \tag{3.10}$$

where

$$F = \Delta t\mathrm{K}^{-1}(\Delta t)' \tag{3.11}$$

is a nonnegative matrix.

We now have to find the vector $\hat{\beta}$ which minimizes Q_2. Consider the change of variable $\beta = \mu\Sigma^{-\frac{1}{2}}$. Substituting into (3.10)

$$Q_2 = \frac{\mu\Sigma^{-\frac{1}{2}}F\Sigma^{-\frac{1}{2}}\mu'}{\mu\mu'} . \tag{3.12}$$

The vector $\hat{\mu}$ which minimizes this expression is obviously any proper vector of the smallest proper value of the nonnegative matrix $\Sigma^{-\frac{1}{2}}F\Sigma^{-\frac{1}{2}}$; that is, $\hat{\mu}$ is given by

$$\hat{\mu}(\Sigma^{-\frac{1}{2}}F\Sigma^{-\frac{1}{2}} - \lambda I) = 0,$$

where λ is the smallest root of the equation,

$$|\Sigma^{-\frac{1}{2}}F\Sigma^{-\frac{1}{2}} - \lambda I| = 0.$$

In computations the following equivalent equations may be preferred. The maximum likelihood estimate of β is given by the equation

$$\hat{\beta}(F - \lambda\Sigma) = 0, \tag{3.13}$$

where λ is the smallest root of

$$|F - \lambda\Sigma| = 0. \tag{3.14}$$

4. Covariance matrix unknown. We assume now that the covariance matrix Σ is unknown, but that we have an estimate S which is independent of t and has a Wishart distribution with mean value Σ and ν degrees of freedom, the probability density of which is proportional to

$$|\Sigma|^{-\frac{1}{2}\nu}|S|^{\frac{1}{2}(\nu-p-1)} \exp -\tfrac{1}{2}\nu \operatorname{tr} \Sigma^{-1}S \tag{4.1}$$

for S positive definite and 0 otherwise, where $S = \{s_{hh'}\}$. It is assumed, as before, that Σ is positive definite, and that the t_i are not subject to any linear restriction (and, therefore, K is positive definite). We shall consider only the case, which happens with probability 1, in which S is positive definite.

The joint probability density of all of the variables t_i^h, $s_{hh'}$ is proportional to the product of (3.1) and (4.1). The maximum likelihood estimates are the values $\hat{\tau}$, $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\Sigma}$ that maximize

$$|\Sigma|^{-\frac{1}{2}(q+\nu)} \exp -\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}[\nu S + r(t - \tau)\mathrm{K}^{-1}(t - \tau)'] \tag{4.2}$$

subject to the condition (3.3). Instead of maximizing (4.2) we can minimize

$$\operatorname{tr} \Sigma^{-1}[\nu S + r(t - \tau)\mathrm{K}^{-1}(t - \tau)'] - (q + \nu) \log |\Sigma^{-1}|. \tag{4.3}$$

We may in the first place keep α, β and τ fixed and find the value of Σ which minimizes (4.3). By the Lemma 3.2.2 of Anderson's book [4] the maximum likelihood estimate of Σ is

$$\hat{\Sigma} = (q + \nu)^{-1}[\nu S + r(t - \hat{\tau})\mathrm{K}^{-1}(t - \hat{\tau})'] \tag{4.4}$$

where $\hat{\tau}$ is the maximum likelihood estimate of τ (based on maximum likelihood estimates of α and β and the restraint (3.3)). Substituting this estimate of Σ (as a function of $\hat{\alpha}$, $\hat{\beta}$, $\hat{\tau}$) into (4.3) we see that $\hat{\alpha}$, $\hat{\beta}$, $\hat{\tau}$ must minimize

$$|\nu S + r(t - \tau)\mathrm{K}^{-1}(t - \tau)'| \tag{4.5}$$

subject to the aforementioned restraint. Consider the change of variable

$$\delta = S^{-\frac{1}{2}}(t - \tau)\mathrm{K}^{-\frac{1}{2}} \tag{4.6}$$

We have then to minimize

$$|\nu I + r\delta\delta'| \tag{4.7}$$

subject to the following conditions, similar to (3.5),

$$\gamma_i - \beta S^{\frac{1}{2}}\delta_i = 0 \qquad (i = 1, \cdots, q), \tag{4.8}$$

where, as in Section 3, γ_i is the ith component of $(\alpha u + \beta t)\mathrm{K}^{-\frac{1}{2}}$ and δ_i is the ith column of δ. We shall find in the first place the minimum of (4.7) for fixed values of α and β. The expression (4.7) may be written

$$\nu^p + D_1\nu^{p-1}r + D_2\nu^{p-2}r^2 + \cdots + D_pr^p$$

where D_h is the sum of all the principal minors of order h of the matrix $\delta\delta'$. The elements of the matrix $\delta\delta'$ are the product-moments with respect to the origin of the system of points δ_i, and consequently all the principal minors of $\delta\delta'$, and a fortiori all the coefficients D_h, are non-negative. In particular,

$$D_1 = \operatorname{tr} \delta\delta' = \sum_i \delta_i'\delta_i \tag{4.9}$$

is the sum of the squares of the distances from the origin to the points δ_i. This is a minimum when all the points δ_i are on the perpendicular from the origin to the hyperplanes (4.8). Since these points are on a straight line which goes through the origin, it is easily seen that at these points all of the principal minors of $\delta\delta'$ of order $\geqq 2$, and consequently, also $D_2, \cdots, D_p$, vanish simultaneously. Therefore, minimizing (4.7) is equivalent to minimizing (4.9) subject to the conditions (4.8). The same problem was solved in Section 3, with the only difference that we now have S instead of Σ. Therefore, the maximum likelihood estimate of β is given by the equation

$$\hat{\beta}(F - lS) = 0 \tag{4.10}$$

where l is the smallest root of

$$|F - lS| = 0. \tag{4.11}$$

Equivalently, $\hat{\beta} = \hat{m}S^{-\frac{1}{2}}$, where $\hat{m}$ is any proper vector of the minimum proper value l of the nonnegative matrix $S^{-\frac{1}{2}}FS^{-\frac{1}{2}}$. The maximum likelihood estimate of α is given by the same equation (3.9) as before, the maximum likelihood estimate of τ is

$$\hat{\tau} = t - \frac{S\hat{\beta}'\hat{\beta}\Delta t}{\hat{\beta}S\hat{\beta}'} \tag{4.12}$$

and the maximum likelihood estimate of Σ is given by (4.4), or, equivalently, by

$$\hat{\Sigma} = \frac{1}{q + \nu}\left(\nu S + rl\,\frac{S\hat{\beta}'\hat{\beta}S}{\hat{\beta}S\hat{\beta}'}\right). \tag{4.13}$$

5. Consistency. When the number r of replications tends to infinity, S converges in probability to Σ and F converges in probability to $\Phi = \Delta\tau K^{-1}(\Delta\tau)'$ where $\Delta\tau = \tau - \tau.u$ and $\tau. = \tau w$. The direction of the true vector β is the only direction orthogonal to all the vectors $\Delta\tau_i = \tau_i - \tau.$, because all points τ_i lie on the hyperplane (1.1) but do not lie on any translated subspace of smaller dimension. Consequently, up to an arbitrary factor, the true vector β is the only vector such that $\beta\Delta\tau = 0$, and, since K^{-1} is a positive definite matrix, it is also the only vector such that $\beta\Phi\beta' = 0$. But, since Φ is a nonnegative matrix, β is the only vector (up to an arbitrary factor) such that $\beta\Phi = 0$. Let μ be defined by $\beta = \mu\Sigma^{-\frac{1}{2}}$. Then μ is the only vector (up to an arbitrary factor) such that $\mu\Sigma^{-\frac{1}{2}}\Phi\Sigma^{-\frac{1}{2}} = 0$. Therefore, the matrix $\Sigma^{-\frac{1}{2}}\Phi\Sigma^{-\frac{1}{2}}$ is singular, the smallest proper value is 0, and its only proper vector is μ. Since the proper vector is a continuous function of the matrix, $\hat{m}$ converges in probability[1] to μ and consequently $\hat{\beta}$ converges in probability to β (up to an arbitrary factor). Since $\beta\Delta t$ converges in probability to $\beta\Delta\tau = 0$, it follows from (4.12) that $\hat{\tau}$ converges in probability to the true matrix τ. It follows then easily from (3.9) that $\hat{\alpha}$ converges to α and from (4.4) that $\hat{\Sigma}$ converges to Σ (since by assumption, the quotient r/ν converges to a positive limit).

6. Homogeneous linear functional relationship. We assume as before that there is no linear restriction (2.2), and, therefore, K is positive definite. If it is known that $\alpha = 0$, the equation (3.8) may be written simply

$$Q_1 = \frac{\beta t \mathrm{K}^{-1} t' \beta'}{\beta \Sigma \beta'}. \tag{6.1}$$

Instead of defining F by (3.11) we shall define F by

$$F = t\mathrm{K}^{-1}t', \tag{6.2}$$

and we have

$$Q_1 = \frac{\beta F \beta'}{\beta \Sigma \beta'}, \tag{6.3}$$

which is similar to (3.10). We can then proceed by the same method that was employed in Section 3. The results and formulae obtained there and in Section 4 will also apply to the homogeneous linear functional relationship case, provided that it is understood that F is given by (6.2) and not by (3.11).

7. Linear restrictions. We shall now assume that the t_i are subject to a known single linear restriction (2.2). In matrix notation

$$t\omega = 0. \tag{7.1}$$

In this section we shall assume that the coefficients ω_i are normalized so that

$$u\omega = 1. \tag{7.2}$$

Since t is assumed to verify (7.1), it follows that it is impossible that $t = u$. It follows also that we have

$$\tau\omega = 0. \tag{7.3}$$

If we multiply (3.3) on the right by ω we obtain, by (7.2) and (7.3),

$$\alpha = 0. \tag{7.4}$$

By (3.3) we have then

$$\beta\tau = 0. \tag{7.5}$$

---

[1] Because, if the vector valued function f is continuous at the point a, and if the random vector x converges in probability to a, then $f(x)$ converges in probability to $f(a)$. This result follows as a special case (y_N constant) from Corollary 2 of [20]. (See also Lemma V of [19]).

Since there is only one linear restriction (2.2), the matrix K is of rank $q - 1$, as was pointed out in Section 2, and consequently, the previous theory cannot be applied directly in this case. Let $L = \{L_{ii'}\}$ be a nonsingular $q \times q$ matrix whose last column is the vector ω, *i.e.*,

$$L_{iq} = \omega_i \qquad (i = 1, \cdots, q), \tag{7.6}$$

and such that all other columns have a sum equal to zero, *i.e.*,

$$\sum_{i'} L_{ii'} = 0 \qquad (i = 1, \cdots, q - 1). \tag{7.7}$$

Consider the new variables

$$t_i^{*h} = \sum_{i'} t_{i'}^h L_{i'i}. \tag{7.8}$$

By (2.2) and (7.6)

$$t_q^{*h} = 0 \qquad (h = 1, \cdots, p). \tag{7.9}$$

If we denote by t^* the $p \times (q - 1)$ matrix the elements of which are the t_i^{*h} with $i \neq q$, and by τ^* and τ_i^{*h} the corresponding true values, we have in matrix notation

$$(t^* \mid 0) = tL, \qquad (\tau^* \mid 0) = \tau L. \tag{7.10}$$

Therefore, if we multiply (7.5) on the right by L, we obtain

$$\beta\tau^* = 0. \tag{7.11}$$

Consequently, the new variables $\tau_i^{*h}(i \neq q)$ satisfy an homogeneous linear functional relationship with the same parameter β. It can be easily seen that the covariance matrix of the pq variables t_i^{*h} is $r^{-1}L'\mathrm{K}L \otimes \Sigma$. From (2.5) and (7.6) it follows that

$$L'\mathrm{K}L = \begin{pmatrix} \mathrm{K}^* & 0 \\ 0 & 0 \end{pmatrix}, \tag{7.12}$$

where K^* is a $(q - 1) \times (q - 1)$ matrix. Since L is nonsingular, and K has rank $q - 1$, it follows that $L'\mathrm{K}L$ has also rank $q - 1$ and, therefore, K^* is nonsingular. Since $r^{-1}\mathrm{K}^* \otimes \Sigma$ is the covariance matrix of the $p(q - 1)$ variables $t_i^{*h}(i \neq q)$, in order to find the maximum likelihood estimates, in the case of unknown Σ, we have to minimize

$$\mathrm{tr}\, \Sigma^{-1}(\nu S + r(t^* - \tau^*)\mathrm{K}^{*-1}(t^* - \tau^*)'] - (q + \nu)\log |\Sigma^{-1}| \tag{7.13}$$

subject to the only restriction (7.11). This is an homogeneous linear functional-relationship problem of the type discussed in the previous section. Therefore, the maximum likelihood estimates $\hat{\beta}$, $\hat{\Sigma}$ are given by (4.10) and (4.13), where l is, as before, the smallest root of (4.11) and F is given by

$$F = t^*\mathrm{K}^{*-1}t^{*\prime}. \tag{7.14}$$

We shall now show that we also have

$$F = t(\mathrm{K} + \kappa_0 u'u)^{-1}t', \tag{7.15}$$

where κ_0 is an arbitrary number different from zero. From (7.12) and (7.7) it follows that

$$L'(\mathrm{K} + \kappa_0\, u'\, u)L = \begin{pmatrix} \mathrm{K}^* & 0 \\ 0 & \kappa_0 \end{pmatrix}, \tag{7.16}$$

and, consequently, $\mathrm{K} + \kappa_0 u'u$ is nonsingular. Moreover, the right-hand member of (7.15) is equal to

$$tL[L'(\mathrm{K} + \kappa_0 u'u)L]^{-1}L't'$$

and, therefore, by (7.10) and (7.16) is equal to the right-hand member of (7.14). In practical applications the expression (7.15) will be used with preference to (7.14). Moreover, in the case of balanced designs, all of the elements that are not in the diagonal of K will have a common value $\kappa' \neq 0$. By choosing $\kappa_0 = -\kappa'$, the matrix $\mathrm{K} + \kappa_0 u'u$ is a diagonal matrix and, consequently, the computations are considerably simplified.

A similar argument shows that the maximum likelihood estimates are also the values that minimize

$$\operatorname{tr} \Sigma^{-1}[\nu S + r(t - \tau)(\mathrm{K} + \kappa_0 u'u)^{-1}(t - \tau)'] - (q + \nu)\log |\Sigma^{-1}| \tag{7.17}$$

subject to the conditions (7.3) and (7.5).

8. Intrablock analysis: direct approach. We shall now estimate the linear functional relationship by applying directly the maximum likelihood method to the model (1.5), considering the coefficients b_j as unknown constant vectors (intrablock analysis). In order to arrive at a unique solution $\hat{\tau}$ we add as usual the linear restriction (2.2). It follows then, as was shown in the previous section, that $\alpha = 0$. As was already mentioned in Section 2, we assume that the errors coming from different experimental units are independent, and that the errors coming from a single experimental unit have a multivariate normal distribution with zero means and covariance matrix Σ. If there are N experimental units, then the probability density for all pN variables y_{ij}^h is proportional to

$$|\Sigma|^{-\frac{1}{2}N}\exp -\tfrac{1}{2}\sum_{hh'} \sigma^{hh'} Q_{hh'} \tag{8.1}$$

where

$$Q_{hh'} = \sum_{ij}{}' \epsilon_{ij}^h \epsilon_{ij}^{h'} . \tag{8.2}$$

The maximum likelihood estimates are the values that maximize (8.1) subject to the conditions (1.3) and (2.2), or, equivalently, the conditions (7.3) and (7.5). Let $y_{\cdot j}^h$, $t_{\cdot j}^h$ and $\tau_{\cdot j}^h$ denote the average of yields, estimated and true treatment effects for the hth variable over all experimental units of block j. Define the adjusted yields and adjusted treatment effects by

$$\tilde{y}_{ij}^h = y_{ij}^h - y_{\cdot j}^h, \qquad \tilde{t}_{ij}^h = t_i^h - t_{\cdot j}^h, \qquad \tilde{\tau}_{ij}^h = \tau_i^h - \tau_{\cdot j}^h,$$

where τ_i^h is the hth component of τ_i.

It can be easily shown that, if $\hat{\tau}_i^h$ is the maximum likelihood estimate of τ_i^h, then the maximum likelihood estimate of b_j^h is $\hat{b}_j^h = y_{\cdot j}^h - \hat{\tau}_{\cdot j}^h$, where $\hat{\tau}_{\cdot j}^h$ denotes the average of the $\hat{\tau}_i^h$ for all treatments occurring in block j. By substitution into (8.1) and (8.2) it follows that the maximum likelihood estimates $\hat{\beta}$, $\hat{\tau}$, $\hat{\Sigma}$ are the values that maximize, subject to the conditions (7.3) and (7.5) the expression

$$|\Sigma|^{-\frac{1}{2}N} \exp -\tfrac{1}{2}\sum_{h,h'} \sigma^{hh'} Q'_{hh'} \tag{8.3}$$

where

$$Q'_{hh'} = \sum_{i,j}{}' (\tilde{y}_{ij}^h - \tilde{\tau}_{ij}^h)(\tilde{y}_{ij}^{h'} - \tilde{\tau}_{ij}^h). \tag{8.4}$$

Suppose that we have numbered serially the N experimental units, and let $\tilde{y}_n^h$, $\tilde{\tau}_n^h$, $\tilde{t}_n^h$ denote the adjusted yield, and the adjusted treatment effects (true and estimated) for the nth experimental unit and the hth characteristic being measured. If $\tilde{y}^h$, $\tilde{\tau}^h$, $\tilde{t}^h$ denote the row vectors the N components of which are the corresponding experimental unit values, we have

$$Q'_{hh'} = (\tilde{y}^h - \tilde{\tau}^h)(\tilde{y}^{h'} - \tilde{\tau}^{h'})'. \tag{8.5}$$

From the definition of $\tilde{\tau}^h$ it follows that

$$\tilde{\tau}^h = \sum_i c_i \tau_i^h, \qquad \tilde{t}^h = \sum_i c_i t_i^h, \tag{8.6}$$

where the c_i are row vectors that depend only on the experimental design. Since the t_i^h are the values that minimize Q'_{hh} subject to the only conditions

$$\sum_i \omega_i t_i^h = 0 \qquad (h = 1, \cdots, p) \tag{8.7}$$

by differentiation with respect to τ_i^h we have, if λ is a Lagrange multiplier,

$$c_i(\tilde{y}^h - \tilde{t}^h)' + \lambda\omega_i = 0.$$

If we multiply this expression by t_i^h and add for $i = 1, \cdots, q$ we have by (8.6) and (8.7)

$$\tilde{t}^h(\tilde{y}^h - \tilde{t}^h)' = 0. \tag{8.8}$$

If we multiply instead by $\tilde{\tau}_i^h$ we have

$$\tilde{\tau}^h(\tilde{y}^h - \tilde{t}^h)' = 0. \tag{8.9}$$

By (8.8) and (8.9) it follows then from (8.5) that

$$Q'_{hh'} = (\tilde{y}^h - \tilde{t}^h)(\tilde{y}^{h'} - \tilde{t}^{h'})' + (\tilde{t}^h - \tilde{\tau}^h)(\tilde{t}^h - \tilde{\tau}^{h'})'$$

and by (2.3)

$$Q'_{hh'} = \nu s_{hh'} + (\bar{t}^h - \bar{\tau}^h)(\bar{t}^{h'} - \bar{\tau}^{h'})'. \tag{8.10}$$

Let $\lambda_{ii'}$ be the number of blocks where both treatments i and i' are applied (and consequently λ_{ii} will be the number of replications of the treatment i). Let $\Lambda = \{\lambda_{ii'}\}$ and let Λ_d be the diagonal matrix whose diagonal is $\lambda_{11}, \cdots, \lambda_{qq}$. Assume that all blocks contain k experimental units and define the matrix $\tilde{K} = \{\tilde{\kappa}_{ii'}\}$ by

$$\tilde{K} = \omega\omega' + \frac{1}{r}\left(\Lambda_d - \frac{\Lambda}{k}\right). \tag{8.11}$$

It can be easily shown that

$$(\bar{t}^h - \bar{\tau}^h)(\bar{t}^{h'} - \bar{\tau}^{h'})' = r\sum_{i,i'} (\tilde{\kappa}_{ii'} - \omega_i\omega_{i'})(t_i^h - \tau_i^h)(t_{i'}^{h'} - \tau_{i'}^{h'}).$$

Therefore, in matrix notation we have, by (7.1) and (7.3),

$$\sum_{h,h'} \sigma^{hh'}Q'_{hh'} = \operatorname{tr} \Sigma^{-1}[\nu S + r(t - \tau)\tilde{K}(t - \tau)']. \tag{8.12}$$

Let $\tilde{Y}$ be the matrix of the adjusted total yields $\tilde{Y}_i^h = \sum_j' \tilde{y}_{ij}^h$. It can be shown that (see for instance [4], p. 251) $\tilde{Y} = t(\Lambda_d - \Lambda/k)$. Then, by (8.11),

$$\tilde{Y} = rt(\tilde{K} - \omega\omega'). \tag{8.13}$$

We shall show that the system of equations (7.1) and (8.13) is equivalent to

$$rt\tilde{K} = \tilde{Y}. \tag{8.14}$$

It is obvious that (8.14) is a consequence of (7.1) and (8.13). From the definitions of Λ and $\tilde{Y}$ it follows that $\Lambda u' = k\Lambda_d u'$ and $\tilde{Y}u' = 0$. By (8.11) we have then

$$\tilde{K}u' = \omega u\omega. \tag{8.15}$$

Therefore, if we multiply (8.14) on the right by u', we obtain the equation (7.1). From (7.1) and (8.14) the equation (8.13) follows immediately. We assume now that the design of the experiment is such that the system of equations (7.1) and (8.13) has a unique solution, t. Then it follows that (8.14) has a unique solution, and, therefore, that $\tilde{K}$ is nonsingular. If $\tilde{K}^{-1} = \{\tilde{\kappa}^{ii'}\}$ is the inverse matrix, then

$$t_i^h = \frac{1}{r}\sum_{i'} \tilde{\kappa}^{ii'}\, \tilde{Y}_{i'}^h,$$

and, therefore,

$$\operatorname{cov}(t_i^h, t_{i'}^{h'}) = \frac{1}{r^2}\sum_{j,j'} \tilde{\kappa}^{ij}\,\tilde{\kappa}^{i'j'} \operatorname{cov}(Y_j^h, \; Y_{j'}^h).$$

It can be shown that

$$\operatorname{cov}(Y_j^h, Y_{j'}^{h'}) = r(\tilde{\kappa}_{jj'} - \omega_j\,\omega_{j'})\sigma_{hh'}$$

and, therefore, (1.6) holds, with

$$\kappa^{ii'} = \tilde{\kappa}^{ii'} - (\sum_j \tilde{\kappa}^{ij}\omega_j)(\sum_{j'} \tilde{\kappa}^{i'j'}\omega_{j'}).$$

In matrix notation $K = \tilde{K}^{-1} - \tilde{K}^{-1}\omega(\tilde{K}^{-1}\omega)'$ and by (8.15)

$$\tilde{K}^{-1} = K + \kappa_0 u'u, \tag{8.16}$$

where $\kappa_0 = (u\omega)^{-2}$ is an arbitrary positive number. By substitution into (8.12) we have

$$\sum_{h,h'} \sigma^{hh'} Q'_{hh'} = \operatorname{tr} \Sigma^{-1}[\nu S + r(t-\tau)(K + \kappa_0 u'u)^{-1}(t-\tau)'].$$

To maximize (8.3) subject to the conditions (7.3) and (7.5) is then equivalent to minimize

$$\operatorname{tr} \Sigma^{-1}[\nu S + r(t-\tau)(K + \kappa_0 u'u)^{-1}(t-\tau)'] - N \log |\Sigma^{-1}| \tag{8.17}$$

subject to the same conditions. But in Section 7 the same problem was solved, with the only difference being that we now have N instead of $q + \nu = N - b + 1$, where b is the number of blocks. The estimate $\hat{\beta}$ is, therefore, the same as before, but instead of (4.13) we have now

$$\hat{\Sigma} = \frac{1}{N}\{\nu S + rl\,[S\hat{\beta}'\hat{\beta}S/\hat{\beta}S\hat{\beta}']\}. \tag{8.18}$$

This expression is equal to the estimate (4.13) multiplied by $1 - k^{-1} + N^{-1}$. Therefore, since (4.13) converges in probability to the true value Σ, the estimate (8.18) converges in probability to $(1 - k^{-1})\Sigma$, and consequently, it is inconsistent. This fact is explained by the existence in the model (1.5) of an indefinitely increasing number of incidental parameters b_j.

The same inconsistency is found in linear regression analysis, *i.e.*, when we drop the restriction (1.1). The maximum likelihood estimate of Σ is then $(\nu/N)S$. When r tends to infinity, $\nu/N \to 1 - k^{-1}$ and, therefore, since S is a consistent estimate of Σ, it follows that the maximum likelihood estimate of Σ in linear regression converges also to $(1 - k^{-1})\Sigma$. Obviously, this happens also in the ordinary univariate analysis of block designs, as was pointed out by J. Neyman and E. L. Scott ([22], Example 2) in the case of a block design with the same treatment applied to all experimental units.

Acknowledgment. The author wishes to acknowledge the valuable help of J. J. Schäffer and the referee, who have materially contributed to the final form of this paper.

REFERENCES

[1] Forman S. Acton, *Analysis of Straight-line Data,* John Wiley and Sons, New York, 1959.

[2] R. J. Adcock, "Note on the method of least squares," *Analyst,* Vol. 4 (1877), pp. 183–184.

[3] R. J. ADCOCK, "A problem in least squares," *Analyst*, Vol. 5 (1878), pp. 53–54.
[4] R. L. ANDERSON AND T. A. BANCROFT, *Statistical Theory in Research*, McGraw Hill Book Co., New York, 1952.
[5] T. W. ANDERSON, *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, New York, 1958.
[6] HARALD CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, N. J., 1946.
[7] BERYL M. DENT, "On observation of points connected by a linear relation," *Proc. Physical Soc. London*, Vol. 47 (1935), pp. 92–108.
[8] CAROLO FRIDERICO GAUSS, "Theoria combinationis observationum erroribus minimis obnoxiae," *Carl Friedrich Gauss Werke*, Vol. 4, pp. 1–108, Gottingen, 1880.
[9] CAROLO FRIDERICO GAUSS, "Theoria motus corporum coelestium (1809)," *Carl Friedrich Gauss Werke*, Vol. 7, pp. 240–257, Leipzig, 1906.
[10] PAUL R. HALMOS, *Finite Dimensional Vector Spaces*, 2nd ed., D. Van Nostrand, Princeton, N. J., 1958.
[11] MAX HALPERIN, "Fitting of straight lines and prediction when both variables are subject to error," *Biometrics*, Vol. 15 (1959), p. 491.
[12] G. W. HOUSNER AND J. F. BRENNAN, "The estimation of linear trends," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 380–388.
[13] J. KIEFER AND J. WOLFOWITZ, "Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 887–906.
[14] T. KOOPMANS, *Linear Regression Analysis of Economic Time Series*, De Erven F. Bohn, Haarlem, 1937.
[15] CHAS. H. KUMMELL, "Reduction of observed equations which contain more than one observed quantity," *Analyst*, Vol. 6 (1879), pp. 97–105.
[16] D. V. LINDLEY, "Regression lines and the linear functional relationship," *Suppl. J. Roy. Stat. Soc.*, Vol. 9 (1947), pp. 218–244.
[17] J. LUKOMSKI, "On some properties of multidimensional distributions," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 236–246.
[18] ALBERT MADANSKY, "The fitting of straight lines when both variables are subject to error," *J. Amer. Stat. Assn.*, Vol. 54 (1959), pp. 173–206.
[19] WILLIAM G. MADOW, "Limiting distributions of quadratic and bilinear forms," *Ann. Math. Stat.*, Vol. 19 (1940), pp. 125–146.
[20] H. B. MANN AND A. WALD, "On stochastic limit and order relationships," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 217–226.
[21] J. NEYMAN, "Remarks on a paper by E. C. Rhodes," *J. Roy. Stat. Soc.*, Vol. 100 (1937), pp. 50–57.
[22] J. NEYMAN AND ELIZABETH L. SCOTT, "Consistent estimates based on partially consistent observations," *Econometrica*, Vol. 16 (1948), pp. 1–32.
[23] JOHN W. TUKEY, "Components in regression," *Biometrics*, Vol. 7 (1951), pp. 33–70.
[24] M. J. VAN UVEN, "Adjustment of N points (in n-dimensional space) to the best linear $(n - 1)$ dimensional space," *Koninklijke Akademie van Wetenschappen te Amsterdam, Proceedings of the Section of Sciences*, Vol. 33 (1930), pp. 143–157 and 307–326.
[25] CESÁREO VILLEGAS, "Sobre la estimación de una relación funcional lineal en el caso de un modelo a valores fijos," *Revista de la Unión Matemática Argentina y de la Asociación Física Argentina*, Vol. 18 (1958), pp. 179–180.