p.87

Covariance Adjustment and Related Problems in Multivariate Analysis

C. RADHAKRISHNA RAO
RESEARCH AND TRAINING SCHOOL
INDIAN STATISTICAL INSTITUTE
CALCUTTA, INDIA

1. SUMMARY

In this paper are considered a number of problems arising out of discrimination between two populations when it is known that they do not differ with respect to a given subset of characters. Such characters may be called, for convenience, *concomitant variables* (although ancillary variables would be a better term) to distinguish them from other (main) variables in which the populations may differ. The concomitant variables, although they by themselves do not discriminate between the populations, may provide additional discrimination in the presence of other variables in which the populations differ. This happens when the concomitants are correlated with the main variables. However, in practice, when the correlations are unknown and have to be estimated from data (for utilizing the concomitants through the technique of covariance adjustment), there may, indeed, be loss of information, unless the correlations are high. Therefore, some caution is needed in the choice and use of concomitant variables.

The methods are extended to the general case of analysis of dispersion under the Gauss-Markoff setup for multiple variables, given that certain linear functions of parameters occurring in the expectations of some functions of the variables have known values. Such functions of the variables are chosen as concomitants and the conditional expectations of the others given the concomitants is considered, which is also of the Gauss-Markoff type. The multivariate least-squares theory applied to the conditional expectations provides a generalization of analysis of covariance which is well known in the context of univariate analysis of variance (A.V.). Such an analysis may be called analysis of dispersion (A.D.) with covariance adjustment.

It is shown that a model for multiple variables recently considered by Potthoff and Roy [3] can be reduced to a Gauss-Markoff model of the

conditional type. Thus an appropriate analysis of their model is provided which is different from the one proposed by them.

2. DISCRIMINATION BETWEEN TWO POPULATIONS

Consider a $(p+q)$-dimensional random variable $(\mathbf{Y}, \mathbf{Z})$, where $\mathbf{Y}$ is p-dimensional and $\mathbf{Z}$ is q-dimensional. We shall refer to $\mathbf{Y}$ as the main variable and $\mathbf{Z}$ as the concomitant variable. Let E_1 and E_2 denote expectations with respect to two $p+q$ variate normal distributions with the same dispersion matrix and possibly different mean vectors. Some hypotheses of interest are as follows:

$$\mathrm{H}_{01}: \quad E_1(\mathbf{Y}) = E_2(\mathbf{Y}), \quad E_1(\mathbf{Z}) = E_2(\mathbf{Z}).$$

$$\mathrm{H}_{02}: \quad E_1(\mathbf{Y}\,|\,\mathbf{Z}) = E_2(\mathbf{Y}\,|\,\mathbf{Z}), \quad \text{whether } E_1(\mathbf{Z}) = E_2(\mathbf{Z}) \text{ or not.}$$

$$\mathrm{H}_{03}: \quad E_1(\mathbf{Y}) = E_2(\mathbf{Y}) \quad \text{given } E_1(\mathbf{Z}) = E_2(\mathbf{Z}).$$

$$\mathrm{H}_{04}: \quad E_1(\mathbf{Y}) = E_2(\mathbf{Y}).$$

The interpretations of these hypotheses and their tests based on independent samples of sizes n_1 and n_2 from the two distributions, although well known, need some re-examination in the light of numerous applications made during recent years.

The hypotheses H_{01} and H_{04} can be tested by using Hotelling's T^2 or Mahalanobis's D^2. The test for H_{02} for a general p was first given by the author [4]; the special case for $p=1$ was well known in the context of analysis of variance (of a single variable) with covariance adjustment (for a number of concomitants). A test for H_{03}, different from that of H_{02}, was developed in [5]. Let D_{p+q}, D_p^2, and D_q^2 be the estimated Mahalanobis' distances based on all the $p+q$ variables $(\mathbf{Y}, \mathbf{Z})$, the main variable $(\mathbf{Y})$, and the concomitant $(\mathbf{Z})$, respectively. Further let $c = n_1 n_2/(n_1+n_2)$ and $N = n_1 + n_2$. Then the suggested test criteria for the hypotheses H_{01} to H_{04} are as follows:

Hypothesis	*Test criterion*	*Null distribution*
H_{01}	$\mathrm{T}_1 = \dfrac{c(N-p-q-1)}{(p+q)(N-2)} D_{p+q}^2$	$F(p+q, N-p-q-1)$
H_{02}	$\mathrm{T}_2 = \dfrac{N-p-q-1}{p} \dfrac{c(D_{p+q}^2 - D_q^2)}{N-2+cD_q^2}$	$F(p, N-p-q-1)$
H_{03}	$\mathrm{T}_3 = \dfrac{c}{N-2}(D_{p+q}^2 - D_q^2)$	Given in (2.1)
H_{04}	$\mathrm{T}_4 = \dfrac{c(N-p-1)}{p(N-2)} D_p^2$	$F(p, N-p-1)$

In the case of T_3 it is convenient to use the related statistic $W = T_3/(1 + T_3)$, with large values of W indicating significance, which has the density

$$P(W) = \left[\Gamma\left(\frac{N+p-q-1}{2}\right)\Gamma\left(\frac{N-1}{2}\right)\Big/\Gamma\left(\frac{N-p-q-1}{2}\right)\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{N+q-1}{2}\right)\right]$$
$$\times W^{(p/2)-1}(1-W)^{[(N-p-q-1)/2]-1}$$
$$\times {}_2F_1\left(\frac{q}{2}, \frac{N-q-1}{2}; \frac{N+p-1}{2}, W\right), \tag{2.1}$$

where ${}_2F_1$ is the hypergeometric function of the second kind.

Let Δ^2_{p+q}, $\Delta_p{}^2$, and $\Delta_q{}^2$ be the true Mahalanobis' distances based on the variables (**Y**, **Z**), **Y**, and **Z** only. Then the hypotheses H_{01} to H_{04} can be stated in alternative forms:

H_{01}: $\Delta^2_{p+q} = 0$

H_{02}: $\Delta^2_{p+q} - \Delta_q{}^2 = 0$ whether $\Delta_q{}^2$ is zero or not

H_{03}: $\Delta^2_{p+q} = 0$ given $\Delta_q{}^2 = 0$, equivalent to $\Delta_p{}^2 = 0$ given $\Delta_q{}^2 = 0$

H_{04}: $\Delta_p{}^2 = 0$

It is, however, customary to use the same criterion T_2 for both the hypotheses H_{02} and H_{03}, as H_{03} implies H_{02} under the condition $E_1(\mathbf{Z}) = E_2(\mathbf{Z})$. See, for instance, the papers by Cochran and Bliss [2] and Cochran [1]. The author has shown elsewhere [5] that there is a slight advantage in using T_3 for testing H_{03}. Such a conclusion was reached by computing the variances of estimates of Δ^2_{p+q} based on the statistics T_2 and T_3, when $\Delta_q{}^2$ is given to be zero, and observing that the variance is smaller for the latter. The percentage points of the distribution (2.1) are, however, not available, but it is hoped to provide the necessary tables in the near future. A good approximation is provided by the use of

$$F = \frac{N-q-1}{N-1}\,\frac{N-p-q-1}{p}\,T_3 \tag{2.2}$$

as a variance ratio on p and $N - p - q - 1$ d.f. The nonnull distribution of W is also derived in [5] in a form suitable for the computation of the power function of W. A tabulation of the power function of W may be needed to examine more fully the relative performances of the tests T_2 and T_3.

When $E_1(\mathbf{Z}) = E_2(\mathbf{Z})$, that is, $\Delta_q{}^2 = 0$, the hypotheses H_{01} to H_{04} are equivalent and, therefore, T_1 and T_4 may also be considered as alternatives to T_2 or T_3. It is expected that T_1 would be inefficient compared to T_2 (or T_3), since the degrees of freedom of the numerator for the corresponding variance ratio test is not based on the effective number p but on the larger number $p+q$. A rough computation of the loss of efficiency in using T_1 instead of T_2 has been made recently by Cochran [1].

But a real competitor to T_2 (or T_3) may be T_4, the test based on D_p^2 ignoring the observations on the q characters. It is clear that when the two sets of variables **Y** and **Z** are uncorrelated, the observations on **Z** do not provide any information on the distribution of **Y**, and consequently the test T_4 should be better than T_2 (or T_3). If the correlations are small, T_4 will still be better, unless the sample sizes n_1 and n_2 are large. In practice we usually have a situation where a test of the type T_2 (or T_3) using a subset of the available concomitants would be more efficient. An examination of the estimated correlations between the sets of variables **Y** and **Z** may enable us to choose such a subset $\mathbf{Z}_1$ (of **Z**). The set $\mathbf{Z}_1$ may be empty, in which case T_4 is the appropriate test.

3. DISCRIMINANT FUNCTIONS

A closely related problem is the estimation of the discriminant function in the following situations

(a) $E_1(\mathbf{Y}) \neq E_2(\mathbf{Y}), \quad E_1(\mathbf{Z}) \neq E_2(\mathbf{Z})$.
(b) $E_1(\mathbf{Y} \mid \mathbf{Z}) \neq E_2(\mathbf{Y} \mid \mathbf{Z})$, using conditional distributions only.
(c) $E_1(\mathbf{Y}) \neq E_2(\mathbf{Y})$ given that $E_1(\mathbf{Z}) = E_2(\mathbf{Z})$.

Let $\boldsymbol{\delta}_{\mathbf{Y}} = E_1(\mathbf{Y}) - E_2(\mathbf{Y})$, $\boldsymbol{\delta}_{\mathbf{Z}} = E_1(\mathbf{Z}) - E_2(\mathbf{Z})$, and let the dispersion matrix of $(\mathbf{Y}, \mathbf{Z})$ be

$$\boldsymbol{\Sigma} = \begin{pmatrix} D(\mathbf{Y}) & C(\mathbf{Y}, \mathbf{Z}) \\ C(\mathbf{Y}, \mathbf{Z}) & D(\mathbf{Z}) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{YY}} & \boldsymbol{\Sigma}_{\mathbf{YZ}} \\ \boldsymbol{\Sigma}_{\mathbf{ZY}} & \boldsymbol{\Sigma}_{\mathbf{ZZ}} \end{pmatrix}.$$

The discriminant function in terms of population parameters in the case of (a) is

$$(\boldsymbol{\delta}_{\mathbf{Y}}' \vdots \boldsymbol{\delta}_{\mathbf{Z}}')\boldsymbol{\Sigma}^{-1}\begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix} = (\boldsymbol{\delta}_{\mathbf{Y}}'\boldsymbol{\Sigma}'' + \boldsymbol{\delta}_{\mathbf{Z}}'\boldsymbol{\Sigma}^{21})\mathbf{Y} + (\boldsymbol{\delta}_{\mathbf{Y}}'\boldsymbol{\Sigma}^{12} + \boldsymbol{\delta}_{\mathbf{Z}}'\boldsymbol{\Sigma}^{22})\mathbf{Z}$$

and in the case of (c),

$$(\boldsymbol{\delta}_{\mathbf{Y}}' \vdots \mathbf{0})\boldsymbol{\Sigma}^{-1}\begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix} = \boldsymbol{\delta}_{\mathbf{Y}}'\boldsymbol{\Sigma}^{11}\mathbf{Y} + \boldsymbol{\delta}_{\mathbf{Y}}'\boldsymbol{\Sigma}^{12}\mathbf{Z}, \tag{3.2}$$

where

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}^{11} & \boldsymbol{\Sigma}^{12} \\ \boldsymbol{\Sigma}^{21} & \boldsymbol{\Sigma}^{22} \end{pmatrix}.$$

The discriminant functions (3.1) and (3.2) can be estimated by substituting for δ_Y, δ_Z, and Σ^{-1} their estimates from the sample. The discriminant function in the case (b) obtained from the likelihood ratio of the conditional distributions of **Y** given **Z** is

$$L(\mathbf{Y} \mid \mathbf{Z}) = (\delta_Y' \vdots \delta_Z')\Sigma^{-1}\begin{pmatrix}\mathbf{Y}\\ \mathbf{Z}\end{pmatrix} - \delta_Z'\Sigma_{22}^{-1}\mathbf{Z}, \tag{3.3}$$

which is the difference between the discriminant functions based on (**Y**, **Z**) and **Z**. For estimating the function (3.3), we substitute for δ_Y, δ_Z, Σ^{-1}, and Σ_{22}^{-1} their estimates.

The status of the discriminant function (3.3) must be clearly understood. If $L(\mathbf{Y}, \mathbf{Z})$ is the discriminant function based on (**Y**, **Z**) and $L(\mathbf{Z})$ that based on **Z** alone, then we have the decomposition of $L(\mathbf{Y}, \mathbf{Z})$,

$$L(\mathbf{Y}, \mathbf{Z}) = L(\mathbf{Z}) + L(\mathbf{Y} \mid \mathbf{Z}), \tag{3.4}$$

into two independent components. In using $L(\mathbf{Y} \mid \mathbf{Z})$ we are considering only the information provided by **Y** independent of **Z**. There may be practical situations where it may be necessary to do so.

Cochran and Bliss [2] and Cochran [1] recommend the use of an estimate of $L(\mathbf{Y} \mid \mathbf{Z})$ substituting $\mathbf{d}_Z$ (the observed differences) for δ_Z even in the case of (c), which specifies that the true values of δ_Z is zero. A more appropriate discriminant function for the case (c) is the expression (3.2).

4. THE EFFECT OF INCREASING THE NUMBER OF CHARACTERS

Suppose there are two experimental conditions (treatments), and on each individual subjected to a treatment multiple measurements have been obtained. The differences between treatments in such a situation can be tested by using Hotelling's T^2 or Mahalanobis's D^2, taking into consideration a certain number of measurements (variables). But it is not uncommon in practice to find significant differences at a given level by applying Student's t test on each individual measurement, whereas the T^2 or D^2 test utilizing all the measurements simultaneously fails to indicate significance at the same level. An example from an anthropometric investigation is given in Table I. The F statistic (square of t) for each character is significant at the 5% level. The T_4 statistic for testing the joint hypothesis (H_{04}) that the population means of femur and humerus are the same for both the communities is 2.685, which as a variance ratio on 2 and 44 d.f. is not significant at the 5% level. Here is a dangerous situation where the inclusion of an extra character decreases the discriminatory power of the test.

To study the effect of increase in the number of characters on tests of

TABLE I

Tests of Differences between Communities by Individual Characters

	Sample size	Mean length of	
		Femur	Humerus
Community 1	27	460.4	335.1
Community 2	20	444.3	323.2
F statistic for each character		5.301 (1, 45 d.f.)	4.901 (1, 45 d.f)

significance we have to compute the power function of the test T_4 for different value of p. For given sample sizes n_1 and n_2, the power depends only on p, the number of characters, and $\Delta_p{}^2$, the true distance between populations. Figures 1, 2, and 3 give the power functions for values of $p = 2(1)9$ when the sample sizes are equal and the common sample size $N = 10$, 20, and 30, respectively. It is seen from these charts that for a given sample size, the power can decrease with increase in the number of characters from q to $p + q$ unless the increase in the true Mahalanobis distance $\Delta_{p+q}^2 - \Delta_q{}^2$ is of a certain order of magnitude. The difference in Mahalanobis distances necessary to maintain the same power decreases, however, as the sample size increases for any given p and q and also as q increases for given p and sample size.

For instance, in the numerical example, the D^2 for femur alone is $D_1{}^2 = 0.4614$ and that due to femur and humerus is $D_2{}^2 = 0.4777$, so that the increase in D^2 due to the inclusion of humerus is 0.0163, indicating that the increase in the population Δ^2 is small. Such a small increase is not of value in samples of sizes 20 and 27 from the two populations. Perhaps with 10 more observations on the total and equal distribution of sample sizes, the inclusion of humerus would have increased the power of discrimination.

5. ANALYSIS OF DISPERSION WITH COVARIANCE ADJUSTMENT

The Gauss-Markoff model in the multivariate case can be written

$$E(\mathbf{X}) = \mathbf{AT}_0, \tag{5.1}$$

where $\mathbf{X}$ is an $n \times p$ matrix of observations, $\mathbf{A}$ is an $n \times m$ matrix of known coefficients, and $\mathbf{T}_0$ is an $m \times p$ matrix of unknown parameters. The rows of $\mathbf{X}$ are independently distributed, while the components of any row have a p-variate distribution with a dispersion matrix $\mathbf{\Sigma}$. Let $\mathbf{C}$ be an $n \times q$ matrix

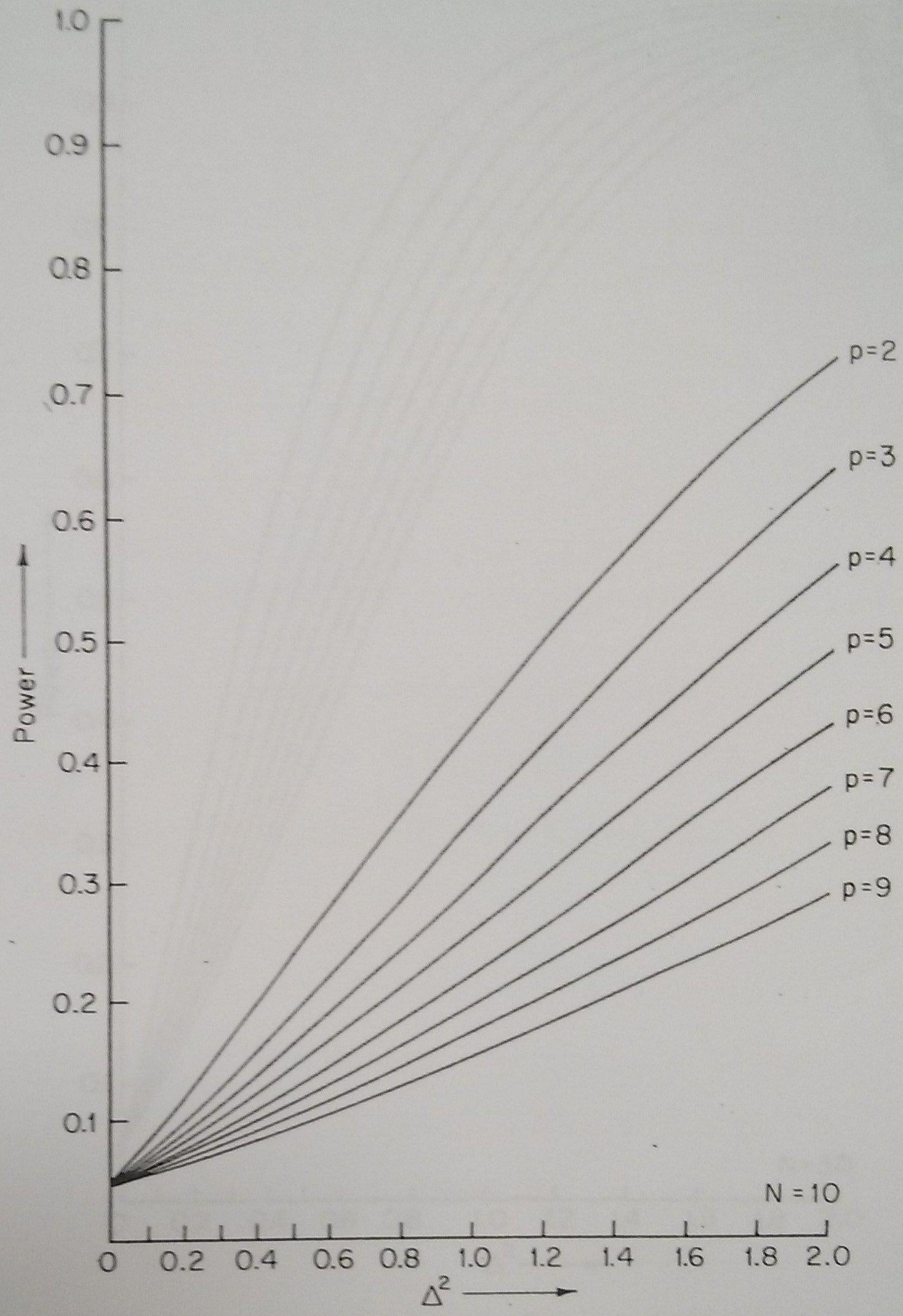

Fig. 1. Power function of the D^2 statistic with $N = 10$.

of concomitant observations, and assume that the conditional expectation of **X** given **C** is

$$E(\mathbf{X}|\mathbf{C}) = \mathbf{AT} + \mathbf{CB}, \tag{5.2}$$

where **B** is a $q \times p$ matrix of unknown regression coefficients. The conditional model (5.2) is again of the Gauss-Markoff type, involving the unknown parameters (**T**, **B**), the known coefficients (**A**, **C**), and the observations **X**, and therefore no new problem arises if conditional inference is needed.

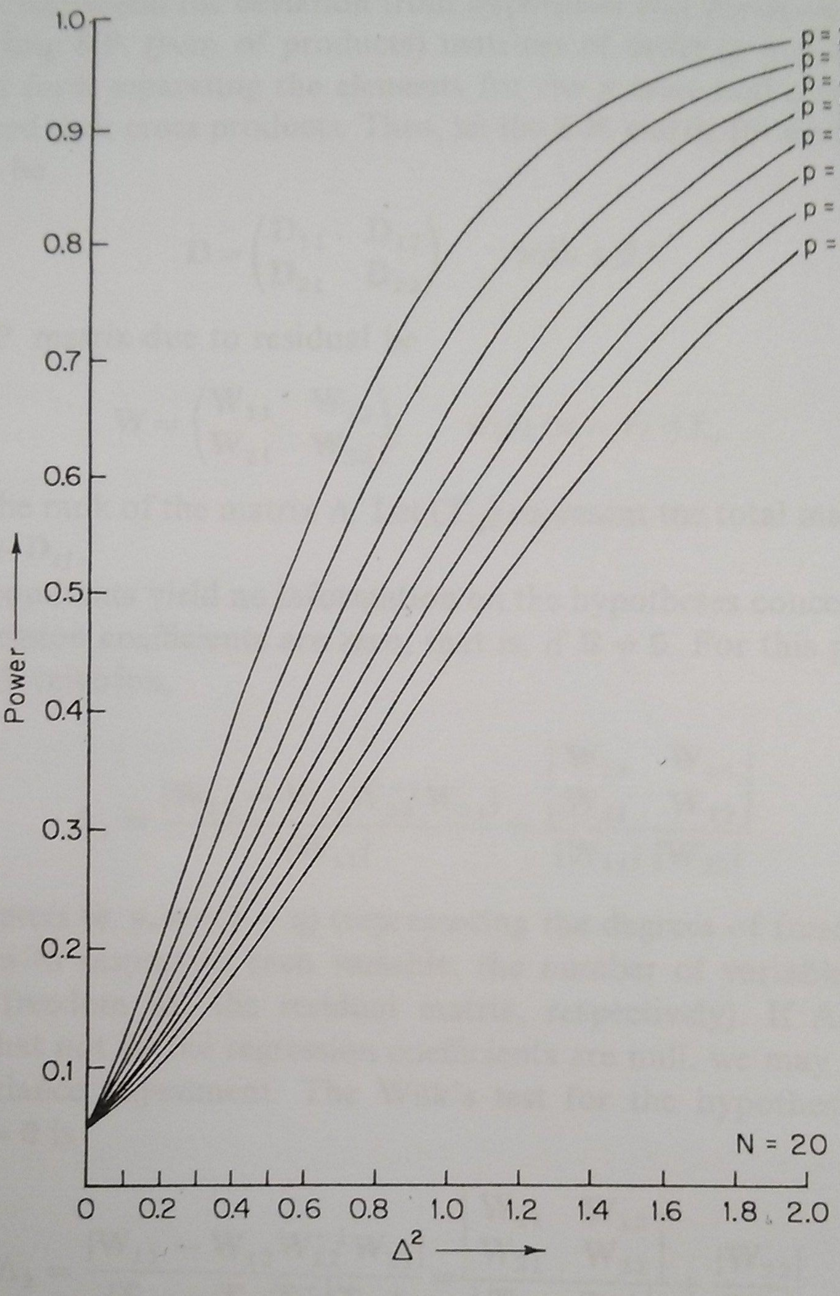

Fig. 2. Power function of the D^2 statistic with $N = 20$.

The parameters $\mathbf{T}$ and $\mathbf{T}_0$ may not, however, be the same. In the context of multivariate normal distributions, if $E(\mathbf{C}) = \mathbf{AT}_1$, then

$$\mathbf{T} = \mathbf{T}_0 - \mathbf{T}_1\mathbf{B}. \tag{5.3}$$

Hence a homogeneous linear hypothesis involving $\mathbf{T}$ is not the same as that on $\mathbf{T}_0$ unless the corresponding linear hypothesis on $\mathbf{T}_1$ is given to be true. Thus for testing linear hypotheses on $\mathbf{T}_0$ when the corresponding linear hypotheses on $\mathbf{T}_1$ are given to be true, we have two alternative models (5.1) and (5.2), of which the former ignores the observations $\mathbf{C}$. We should

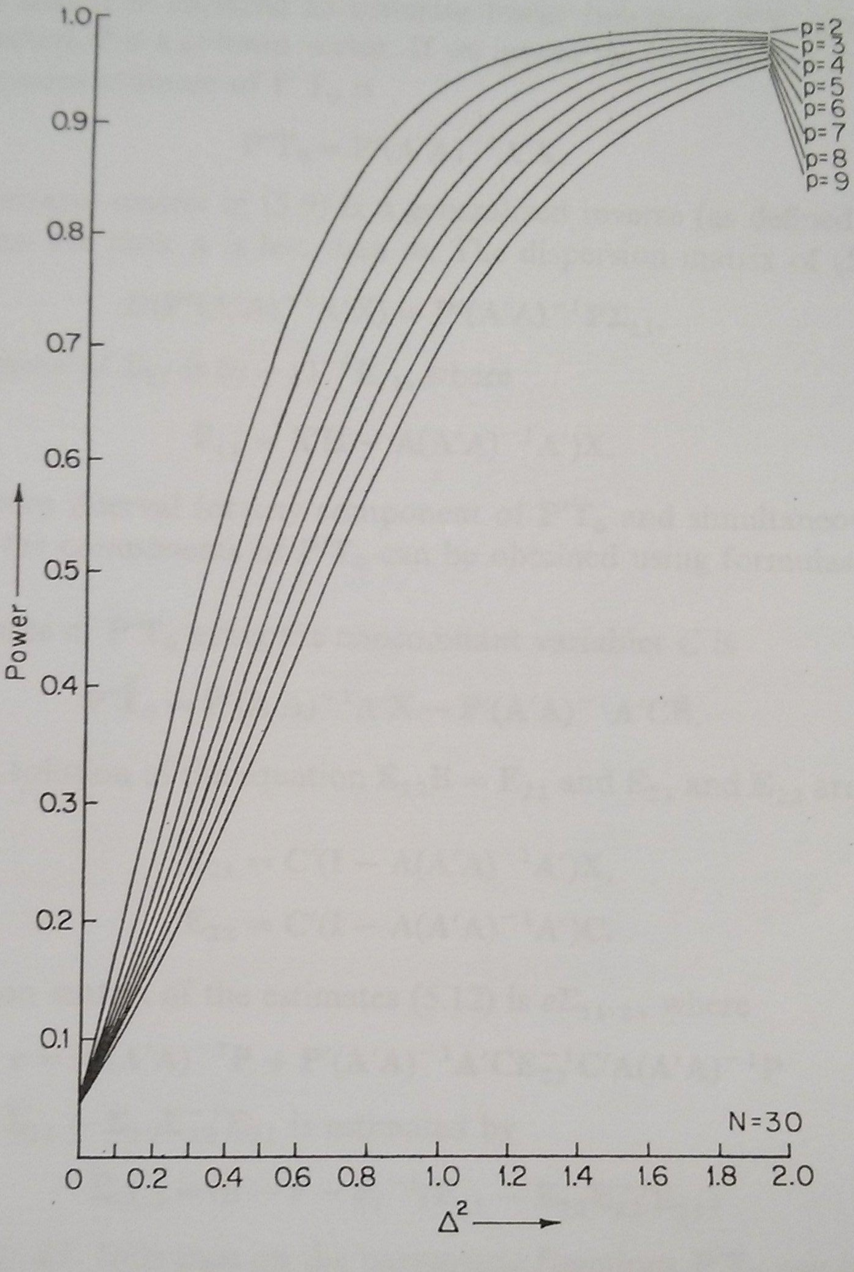

Fig. 3. Power function of the D^2 statistic with $N = 30$.

therefore first examine whether the information provided by **C** is useful for the purpose of examining linear hypotheses on $\mathbf{T}_0$.

For this purpose we first consider the Gauss-Markoff model

$$E(\mathbf{X}) = \mathbf{A}\mathbf{T}_0, \qquad E(\mathbf{C}) = \mathbf{A}\mathbf{T}_1 \tag{5.4}$$

in terms of the matrices $(\mathbf{X} \vdots \mathbf{C})$ of observations and parameters $(\mathbf{T}_0 \vdots \mathbf{T}_1)$:

$$E(\mathbf{X} \vdots \mathbf{C}) = \mathbf{A}(\mathbf{T}_0 \vdots \mathbf{T}_1). \tag{5.5}$$

Let us consider the hypothesis, $\mathbf{R}(\mathbf{T}_0 \vdots \mathbf{T}_1) = \mathbf{0}$, where the rank of $\mathbf{R}$ is s.

Starting with the model (5.5) for $p+q$ correlated variables, let us obtain an analysis of dispersion for deviation from hypothesis and due to residual. The corresponding S.P. (sum of products) matrices of order $(p+q)$ are given in partitioned form separating the elements for the p main and q concomitant variables and their cross products. Thus, let the S.P. matrix for deviation from hypothesis be

$$\mathbf{D}=\begin{pmatrix}\mathbf{D}_{11} & \mathbf{D}_{12}\\ \mathbf{D}_{21} & \mathbf{D}_{22}\end{pmatrix} \qquad \text{with } s \text{ d.f.}$$

and the S.P. matrix due to residual be

$$\mathbf{W}=\begin{pmatrix}\mathbf{W}_{11} & \mathbf{W}_{12}\\ \mathbf{W}_{21} & \mathbf{W}_{22}\end{pmatrix} \qquad \text{with } (n-r) \text{ d.f.,}$$

where r is the rank of the matrix $\mathbf{A}$. Let $(\mathbf{T}_{ij})$ represent the total matrix, where $\mathbf{T}_{ij}=\mathbf{W}_{ij}+\mathbf{D}_{ij}$.

The concomitants yield no information on the hypotheses concerning $\mathbf{T}_0$ if all the regression coefficients are zero, that is, if $\mathbf{B}=\mathbf{0}$. For this purpose we use Wilk's Λ criterion,

$$\Lambda_1=\frac{|\mathbf{W}_{11}-\mathbf{W}_{12}\mathbf{W}_{22}^{-1}\mathbf{W}_{21}|}{|\mathbf{W}_{11}|}=\frac{\begin{vmatrix}\mathbf{W}_{11} & \mathbf{W}_{12}\\ \mathbf{W}_{21} & \mathbf{W}_{22}\end{vmatrix}}{|\mathbf{W}_{11}|\,|\mathbf{W}_{22}|} \tag{5.6}$$

with parameters $(q, p, n-r-q)$ (representing the degrees of freedom of the hypothesis with respect to each variable, the number of variables, and the degrees of freedom for the residual matrix, respectively). If Λ_1 is small, indicating that not all the regression coefficients are null, we may proceed to make covariance adjustment. The Wilk's test for the hypothesis $\mathbf{RT}_0=\mathbf{0}$ given $\mathbf{RT}_1=\mathbf{0}$ is

$$\Lambda_2=\frac{|\mathbf{W}_{11}-\mathbf{W}_{12}\mathbf{W}_{22}^{-1}\mathbf{W}_{21}|}{|\mathbf{T}_{11}-\mathbf{T}_{12}\mathbf{T}_{22}^{-1}\mathbf{T}_{21}|}=\frac{\begin{vmatrix}\mathbf{W}_{11} & \mathbf{W}_{12}\\ \mathbf{W}_{21} & \mathbf{W}_{22}\end{vmatrix}}{\begin{vmatrix}\mathbf{T}_{11} & \mathbf{T}_{12}\\ \mathbf{T}_{21} & \mathbf{T}_{22}\end{vmatrix}}\div\frac{|\mathbf{W}_{22}|}{|\mathbf{T}_{22}|} \tag{5.7}$$

with the parameters $(s, p, n-r-q)$. An alternative test similar to T_3 of Section 2 is the difference

$$\Lambda_3=\frac{\begin{vmatrix}\mathbf{T}_{11} & \mathbf{T}_{12}\\ \mathbf{T}_{21} & \mathbf{T}_{22}\end{vmatrix}}{\begin{vmatrix}\mathbf{W}_{11} & \mathbf{W}_{12}\\ \mathbf{W}_{21} & \mathbf{W}_{22}\end{vmatrix}}-\frac{|\mathbf{T}_{22}|}{|\mathbf{W}_{22}|}, \tag{5.8}$$

whose distribution is unknown except in the case $s=1$.

Suppose that it is required to estimate linear functions $\mathbf{P}'\mathbf{T}_0$ given that $\mathbf{P}'\mathbf{T}_1 = \mathbf{0}$, where $\mathbf{P}$ is a column vector. If we ignore the information $\mathbf{P}'\mathbf{T}_1 = \mathbf{0}$, the least-squares estimate of $\mathbf{P}'\mathbf{T}_0$ is

$$\mathbf{P}'\hat{\mathbf{T}}_0 = \mathbf{P}'(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{X}, \tag{5.9}$$

where the inverse matrix in (5.9) is a generalized inverse (as defined in Rao [7],[8]), when the rank $\mathbf{A}$ is less than m. The dispersion matrix of (5.9) is

$$D(\mathbf{P}'(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{X}) = \mathbf{P}'(\mathbf{A}'\mathbf{A})^{-1}\mathbf{P}\mathbf{\Sigma}_{11}, \tag{5.10}$$

and an estimate of $\mathbf{\Sigma}_{11}$ is $(n-r)^{-1}\mathbf{E}_{11}$, where

$$\mathbf{E}_{11} = \mathbf{X}'(\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')\mathbf{X}. \tag{5.11}$$

The confidence interval for any component of $\mathbf{P}'\mathbf{T}_0$ and simultaneous intervals for all the components of $\mathbf{P}'\mathbf{T}_0$ can be obtained using formulas (5.9) to (5.11).

The estimate of $\mathbf{P}'\mathbf{T}_0$ using the concomitant variables $\mathbf{C}$ is

$$\mathbf{P}'\hat{\hat{\mathbf{T}}}_0 = \mathbf{P}'(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{X} - \mathbf{P}'(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{C}\hat{\mathbf{B}}, \tag{5.12}$$

where $\hat{\mathbf{B}}$ is a solution of the equation $\mathbf{E}_{22}\mathbf{B} = \mathbf{E}_{21}$ and $\mathbf{E}_{21}$ and $\mathbf{E}_{22}$ are defined as

$$\begin{aligned} \mathbf{E}_{21} &= \mathbf{C}'(\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')\mathbf{X}, \\ \mathbf{E}_{22} &= \mathbf{C}'(\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')\mathbf{C}. \end{aligned} \tag{5.13}$$

The dispersion matrix of the estimates (5.12) is $e\mathbf{\Sigma}_{11\cdot 2}$, where

$$e = \mathbf{P}'(\mathbf{A}'\mathbf{A})^{-1}\mathbf{P} + \mathbf{P}'(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{C}\mathbf{E}_{22}^{-1}\mathbf{C}'\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{P} \tag{5.14}$$

and $\mathbf{\Sigma}_{11\cdot 2} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$ is estimated by

$$\hat{\mathbf{\Sigma}}_{11\cdot 2} = (n-r-q)^{-1}(\mathbf{E}_{11} - \mathbf{E}_{12}\mathbf{E}_{22}^{-1}\mathbf{E}_{21}) \tag{5.15}$$

on $(n-r-q)$ d.f. Inference on the parametric functions $\mathbf{P}'\mathbf{T}_0$ can be drawn using formulas (5.12) to (5.15). Such inference is of the conditional type, given the values of the concomitant variables. When variations in the concomitant variables are considered, one uses the distribution (2.1) instead of the F distribution.

To compare the alternative estimators (5.9) and (5.12) of $\mathbf{P}'\mathbf{T}_0$ we examine the corresponding dispersion matrices of the estimators $\mathbf{P}'\mathbf{A}\mathbf{P}\mathbf{\Sigma}_{11}$ and $e\mathbf{\Sigma}_{11\cdot 2}$. It is seen from (5.14) that $e \geq \mathbf{P}'\mathbf{A}\mathbf{P}$, and consequently covariance adjustment leads to higher efficiency only if the elements of $\mathbf{\Sigma}_{11\cdot 2}$ are somewhat smaller than those of $\mathbf{\Sigma}_{11}$. Of course, when $\mathbf{\Sigma}_{11\cdot 2} = \mathbf{\Sigma}_{11}$, that is, when $\mathbf{\Sigma}_{12} = 0$, there is always loss of efficiency.

To investigate this problem a little further, we shall consider the simple case of estimating the mean of a variable adjusting for q concomitant variables which are uncorrelated with the main variable. Let us suppose that n independent observations are available on $(q+1)$ variables, one main variable y, and q concomitants $z_1, \dots, z_q$. Further, let

$$(\bar{y} \vdots \bar{\mathbf{z}}') = (\bar{y}, \bar{z}_1, \dots, \bar{z}_q) \tag{5.16}$$

be the observed mean values and

$$\mathbf{S} = \begin{pmatrix} S_{00} & \mathbf{S}_{01} \\ \mathbf{S}_{10} & \mathbf{S}_{11} \end{pmatrix} \tag{5.17}$$

be $(q+1) \times (q+1)$ matrix of the corrected sum of products. The length of the confidence interval for the mean of y based on the observations on y alone is the square root of

$$c_1^2 = \frac{4S_{00}}{n(n-1)} t_1^2, \tag{5.18}$$

where t_1 is the upper $\alpha/2$ point of the t distribution on $(n-1)$ d.f.

When adjustment is made for the concomitant variables $z_1, \dots, z_q$, the corresponding expression is

$$c_2^2 = \frac{4|\mathbf{S}_{11} + n\bar{\mathbf{z}}\bar{\mathbf{z}}'|}{n|\mathbf{S}_{11}|} \frac{S_{00} - \mathbf{S}_{01}\mathbf{S}_{11}^{-1}\mathbf{S}_{10}}{n-q-1} t_2^2, \tag{5.19}$$

where $\bar{\mathbf{z}}' = (\bar{z}_1, \dots, \bar{z}_q)$ and t_2 is the upper $\alpha/2$ point of the t distribution on $(n-q-1)$ d.f. The confidence interval c_1 will be shorter than c_2 if

$$\frac{S_{00} - \mathbf{S}_{01}\mathbf{S}_{11}^{-1}\mathbf{S}_{10}}{S_{00}} > \frac{(n-q-1)t_1^2}{(n-1)t_2^2} \frac{|\mathbf{S}_{11}|}{|\mathbf{S}_{11} + n\bar{\mathbf{z}}\bar{\mathbf{z}}'|}. \tag{5.20}$$

To compute the probability of the event (5.20), let us observe that

$$\frac{S_{00} - \mathbf{S}_{01}\mathbf{S}_{11}^{-1}\mathbf{S}_{10}}{S_{00}} \sim B\left(\frac{n-q-1}{2}, \frac{q}{2}\right) \tag{5.21}$$

(where the symbol $\sim$ is used for "distributed as") and

$$\frac{|\mathbf{S}_{11}|}{|S_{11} + n\bar{\mathbf{z}}\bar{\mathbf{z}}'|} \sim B\left(\frac{n-q}{2}, \frac{q}{2}\right) \tag{5.22}$$

are independently distributed. Hence the desired probability is $P(X_1 > cX_2)$, where X_1 and X_2 are two independent random variables with beta distributions as in (5.21) and (5.22) and c is the constant $(n-q-1)t_1^2/(n-1)\, t_2^2$. Actual computations show that the probability is of the order of 0.80 for $q = 1$ and increases with q for values of n of practical interest. The probability will be slightly smaller for large values of n.

6. THE MODEL OF POTTHOFF AND ROY

Potthoff and Roy [3] introduced a model of observations $\mathbf{X}$ such that

$$E(\mathbf{X}) = \mathbf{A}\xi\mathbf{F}, \tag{6.1}$$

where $\mathbf{A}$ and $\mathbf{F}$ are known matrices of orders $n \times m$ and $q \times p$ and ξ is a matrix of order $m \times q$ of unknown parameters. The different rows of $\mathbf{X}$ are distributed independently, while the p components in each row have a p-variate normal distribution with the same dispersion matrix. Starting from the model (6.1) the authors consider the estimation of linear functions of ξ and tests of linear hypotheses on ξ. Their approach is summarized below.

Let the rank of $\mathbf{F}$ be q, in which case the rank of $\mathbf{FG}^{-1}\mathbf{F}'$ is q, where $\mathbf{G}$ is any positive definite matrix of order $p \times p$. Multiplying both sides of (6.1) by $\mathbf{G}^{-1}\mathbf{F}'(\mathbf{FG}^{-1}\mathbf{F}')^{-1}$, we have

$$E(\mathbf{Y}) = \mathbf{A}\xi, \tag{6.2}$$

where $\mathbf{Y} = \mathbf{XG}^{-1}\mathbf{F}'(\mathbf{FG}^{-1}\mathbf{F}')^{-1}$. Thus the model (6.1) implies the model (6.2) of the Gauss-Markoff type. The general theory of least squares can then be applied on the model (6.2) for drawing inferences on the parameter ξ.

Such a procedure is unsatisfactory, for two reasons. First, the matrix $\mathbf{G}$ is arbitrary. Second, the matrix of observations $\mathbf{X}$ of order $n \times p$ is reduced to a matrix $\mathbf{Y}$ of order $n \times q$, and, if $q < p$, there will be loss of information unless the dispersion matrix of the variables in any row of $\mathbf{X}$ is known and $\mathbf{G}$ is chosen to be this known dispersion matrix. Potthoff and Roy suggest that $\mathbf{G}$ may be chosen on the basis of prior information or estimated from previous data. From a practical point of view, both the suggestions are subject to criticism.

We shall give an alternative reduction of the model (6.1) leading to a conditional model of the type (5.2), in which case the general theory of Section 5 will apply. Construct a $p \times p$ nonsingular matrix $\mathbf{H} = (\mathbf{H}_1 \vdots \mathbf{H}_2)$ such that $\mathbf{FH}_2 = \mathbf{0}$ and the columns of $\mathbf{H}_1$ form a basis of the vector space generated by the rows of $\mathbf{F}$. Such a matrix $\mathbf{H}$ is not necessarily unique. Let r be the number of columns in $\mathbf{H}_1$. Multiplying both sides of (6.1) by $\mathbf{H}$, we find

$$E(\mathbf{XH}_1) = \mathbf{A}\xi\mathbf{FH}_1, \qquad E(\mathbf{XH}_2) = \mathbf{0}. \tag{6.3}$$

The rank of $\mathbf{FH}_1$ is obviously r, and hence $\xi\mathbf{FH}_1$ can be replaced by an $(m \times r)$-order matrix $\boldsymbol{\eta}$ of independent parameters, so that the setup (6.1) is equivalent to

$$E(\mathbf{Y}) = \mathbf{A}\boldsymbol{\eta}, \qquad E(\mathbf{Z}) = \mathbf{0}, \tag{6.4}$$

where $\mathbf{Y} = \mathbf{XH}_1$ and $\mathbf{Z} = \mathbf{XH}_2$. Hence the conditional expectation of $\mathbf{Y}$ given $\mathbf{Z}$ can be written

$$E(\mathbf{Y}|\mathbf{Z}) = \mathbf{F}\boldsymbol{\eta} + \mathbf{ZB}, \tag{6.5}$$

introducing the matrix $\mathbf{B}$ of unknown regression parameters. Thus we have the observations $\mathbf{Y}$ and the expectation matrix (6.5) of the Gauss-Markoff type as in (5.2). Then the general theory of analysis of dispersion with adjustment for concomitant variables as given in Section 5 is applicable.

It may be noted that the matrix $\mathbf{H}$ is not unique but the estimates of parametric functions and test criteria based on analysis of dispersion will be the same for all choices of $\mathbf{H}$, satisfying the stated conditions.

When the rank of $\mathbf{F}$ is q, we can choose $\mathbf{H}_1$ as $\mathbf{G}^{-1}\mathbf{F}'(\mathbf{F}\mathbf{G}^{-1}\mathbf{F}')^{-1}$ and $\mathbf{H}_2$ such that $\mathbf{F}\mathbf{H}_2 = 0$, where $\mathbf{G}$ is an arbitrary positive-definite matrix. In such a case the conditional expectation (6.5) is

$$E(\mathbf{Y}|\mathbf{Z}) = \mathbf{A}\boldsymbol{\xi} + \mathbf{Z}\mathbf{B}, \tag{6.6}$$

in which the original parameters are retained.

In the method of Potthoff and Roy, the information contained in $\mathbf{Z}$ is ignored.

7. AN ILLUSTRATIVE EXAMPLE

Let $\mathbf{Y}' = (y_1, \dots, y_p)$ be observations on a growth curve at p points of time such that

$$E(y_i) = \beta_0 + \beta_1\varphi_{1i} + \cdots + \beta_k\varphi_{ki}, \quad D(\mathbf{Y}) = \boldsymbol{\Sigma} \qquad (i = 1, \dots, p), \tag{7.1}$$

where φ_{ji} is the value of the jth orthogonal polynomial (in time) at the ith time point. If n independent growth curves each satisfying condition (7.1) are considered, we have an $n \times p$ matrix $\mathbf{X}$ of observations with the expectation

$$E(\mathbf{X}) = \mathbf{U}\boldsymbol{\beta}\boldsymbol{\Phi}, \tag{7.2}$$

where $\mathbf{U}' = (1, \dots, 1)$ with all unities, $\boldsymbol{\beta} = (\beta_0, \dots, \beta_k)$, and $\boldsymbol{\Phi}$ is a $(1 + k) \times p$ matrix of values of orthogonal polynomials of orders 0 to k. The model (7.2) is a special case of (6.1). Let $\boldsymbol{\Phi}_0$ be a $(p - 1 - k) \times p$ matrix of values of orthogonal polynomials of orders $(k + 1)$ to p. Then

$$E[\mathbf{X}(\boldsymbol{\Phi}' \vdots \boldsymbol{\Phi}_0')] = \mathbf{U}\boldsymbol{\beta}\boldsymbol{\Phi}(\boldsymbol{\Phi}' \vdots \boldsymbol{\Phi}_0') = \mathbf{U}\boldsymbol{\beta}(\mathbf{I} \vdots \mathbf{0}), \tag{7.3}$$

giving

$$E(\mathbf{X}\boldsymbol{\Phi}') = \mathbf{U}\boldsymbol{\beta}, \qquad E(\mathbf{X}\boldsymbol{\Phi}_0') = \mathbf{0},$$

a model of the type (6.3). The transformation (7.3) simply implies that if, instead of $(y_1, \dots, y_p)$ in (7.1), we consider the equivalent variables $(b_0, \dots, b_{p-1})$ such that

$$b_i = y_1\varphi_{i1} + \cdots + y_p\varphi_{ip}, \tag{7.4}$$

we have a p-dimensional variable $\mathbf{b}$ such that

$$\begin{aligned} E(b_i) &= \beta_i, \qquad i = 0, 1, \ldots, k, \\ &= 0, \qquad i = k+1, \ldots, p-1. \end{aligned} \tag{7.5}$$

We are required to draw inferences on β_i given n sets of independent observations on $\mathbf{b}$ with expectation as in (7.5) and with an unknown dispersion matrix.

We recognize the variables $b_{k+1}, \ldots, b_{p-1}$ as concomitants and proceed to estimate the coefficients β_i using the method outlined in Section 5. Before doing so, it is worthwhile examining whether it is profitable to consider the concomitants. We shall illustrate the procedure by an example given in [3] with $p = 4$ and $n = 16$. The first step is to obtain b_0, b_1, b_2, and b_3 from $y_1, \ldots, y_4$ on each curve and compute the averages and the corrected sum of squares and products from the 16 sets of four b values. The results are summarized in Table II, which also contains individual F tests for testing the significance of each coefficient, and also the correlation matrix in the lower half of matrix on the right side. ($b_0, \ldots, b_3$ are computed using the unstandardized values of the orthogonal polynomials. An adjustment for this is made in Table III, where the estimates of β_i are given.)

TABLE II

Individual Tests and S.P. and Correlation Matrices for $b_0, \ldots, b_3$

b	$F_{1,\,15}$	S.P. and correlation matrices			
		b_0	b_1	b_2	b_3
$b_0 = 99.875$	2982[a]	802.7500	−5.6250	−35.8750	−213.7500
$b_1 = 15.687$	59.63[a]	−0.0063	990.4375	−91.4375	105.8750
$b_2 = 0.813$	1.88	−0.1383	−0.3162	84.4375	4.6250
$b_3 = 1.125$	0.26	−0.2221	0.0987	0.0148	1161.7500

[a] Indicates significance at the 1% level.

It is clear from Table II that b_2 and b_3 are of the nature of concomitants, indicating that the degree of the polynomial to be considered is unity. Further, none of the correlations between (b_0, b_1) and the concomitants (b_2, b_3) is high (not significantly different from zero at 5%), showing that covariance adjustment is not profitable. The point estimates of β_0 and β_1 and the confidence intervals based on the observations on b_0 and b_1 only are given in Table III with the corresponding results of Potthoff and Roy. It is seen from

TABLE III

Estimates and Widths of 95% Confidence Interval

Parameter		Using b_0 and b_1	Method of Potthoff and Roy
β_0	Estimate	24.969	25.111
	Width	1.948	1.941
β_1	Estimate	0.7844	0.7665
	Width	0.429	0.471

Table III that slightly more efficient estimates are obtained by the proposed method. It may be noted that a preliminary analysis led us to decisions on the degree of the polynomial and the use of concomitants, on which the final analysis is based. Does it imply that our computed precisions are over-estimated and need some adjustment because of our preliminary examination of data? This situation must be faced in any statistical analysis of real data where the model used for final analysis is partly determined by a preliminary examination of data. But such difficulties in interpretation arise only if we are considering an isolated case. I imagine that in any experimental investigation a series of growth curves will be studied, possibly under different experimental conditions (treatments), involving the estimation of polynomial growth coefficients in each case and very often a comparison of such estimates. It should be possible to decide on a suitable set of concomitants for such a purpose.

For a discussion of the methods of estimation of polynomial growth curves the reader is referred to [6, 9, 10].

REFERENCES

1. Cochran, W. G. (1964). Comparison of two methods of handling covariates in discriminatory analysis. *Ann. Inst. Statist. Math.* **16** 43–53.
2. Cochran, W. G. and Bliss, C. I. (1948). Discriminant functions with covariance. *Ann Math. Statist.* **19** 151–176.
3. Potthoff, R. F. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* **51** 313–326.
4. Rao, C. R. (1946). Tests with discriminant functions in multivariate analysis. *Sankhyā* **7** 407–414.
5. Rao, C. R. (1949). On some problems arising out of discrimination with multiple characters. *Sankhyā* **9** 343–366.
6. Rao, C. R. (1959). Some problems involving linear hypotheses in multivariate analysis. *Biometrika* **46** 49–58.
7. Rao, C. R. (1962). A note on the generalized inverse of a matrix with applications to problems in mathematical statistics. *J. Roy. Statist. Soc. Ser. B* **24** 152–158.

8. Rao, C. R. (1966). Generalized inverse for matrices and its application in mathematical statistics. *Neyman Festschrift Volume, Research Papers in Statistics* (F. N. David, ed.), pp. 263–279. Wiley, London (in press).
9. Rao, C. R. (1965). The theory of least squares when the parameters are stochastic and it application to the analysis of growth curves. *Biometrika* **52** 447–458.
10. Rao, C. R. (1966). Least square theory using an estimated dispersion matrix and its application to measurement of signals. *Proc. 5th Berkeley Symp. Math. Statist. Prob.* (in press).