Biometrika (1965), 52, 3 and 4, p. 447
Printed in Great Britain

The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves

By C. RADHAKRISHNA RAO
Indian Statistical Institute, Calcutta

1. Introduction

In an earlier paper (Rao, 1959), the author discussed the method of least squares when the observations are dependent and the dispersion matrix is unknown but an independent estimate is available. The unknown dispersion matrix was, however, considered as an arbitrary positive definite matrix. In the present paper we shall consider a class of problems where the dispersion matrix has a known structure and discuss the appropriate statistical methods.

More specifically the structure of the dispersion matrix results from considering the parameters in the well-known Gauss–Markoff linear model as random variables. Let $\mathbf{Y}$ be a vector random variable with the structure

$$\underset{(p\times 1)}{\mathbf{Y}} = \underset{(p\times m)}{\mathbf{A}}\ \underset{(m\times 1)}{\mathbf{T}} + \underset{(p\times 1)}{\boldsymbol{\epsilon}}, \tag{1·1}$$

where $\mathbf{T}$ and $\boldsymbol{\epsilon}$ are unobservable vector random variables of dimensions m and p respectively and $\mathbf{A}$ is a $p\times m$ matrix of known coefficients with rank equal to m (without loss of generality). Further let

$$\left.\begin{aligned} E(\mathbf{T}) = \boldsymbol{\tau},\quad & E(\boldsymbol{\epsilon}) = 0,\\ D(\mathbf{T}) = e\boldsymbol{\Lambda},\quad & D(\boldsymbol{\epsilon}) = e\boldsymbol{\Sigma},\\ & C(\mathbf{T},\boldsymbol{\epsilon}) = 0,\end{aligned}\right\} \tag{1·2}$$

where E, C and D stand for expectation, covariance and dispersion respectively and e is a known constant. As a consequence of (1·2)

$$D(\mathbf{Y}) = e(\mathbf{A}\boldsymbol{\Lambda}\mathbf{A}'+\boldsymbol{\Sigma}). \tag{1·3}$$

Let $f^{-1}\mathbf{S}$ be an unbiased estimate of $(\mathbf{A}\boldsymbol{\Lambda}\mathbf{A}'+\boldsymbol{\Sigma})$ which is stochastically independent of $\mathbf{Y}$.

Under the set up (1·1)–(1·3), we shall consider the problems of estimating (predicting) linear functions of the random variable $\mathbf{T}$ and linear functions of the parameter $\boldsymbol{\tau}$. The problems of inference are examined under the additional assumptions

$$\left.\begin{aligned} &\mathbf{T}\sim N_m(\boldsymbol{\tau}, e\boldsymbol{\Lambda}),\ \boldsymbol{\epsilon}\sim N_p(0, e\boldsymbol{\Sigma})\\ &\mathbf{S}\sim W_p(f, \mathbf{A}'\boldsymbol{\Lambda}\mathbf{A}+\boldsymbol{\Sigma}),\end{aligned}\right\} \tag{1·4}$$

where N_k denotes a k-dimensional normal distribution with mean and dispersion matrix as indicated within the brackets and W_p represents a p-dimensional central Wishart distribution with degrees of freedom and dispersion matrix as indicated within the brackets. (The symbol $\sim$ is used for 'distributed as'.)

The particular case when $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$, which is of practical interest, is examined in detail. The methods derived are applied to the analysis of polynomial growth curves, and the results are compared with those obtained by using other methods (Rao, 1959; Elston & Grizzle, 1962; Potthoff & Roy, 1964; Elston, 1964).

The model (1·1)–(1·3) is a special case of a mixed effects linear model

$$\mathbf{Y} = \mathbf{A}\boldsymbol{\tau} + \mathbf{B}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \quad \text{with} \quad \mathbf{A} = \mathbf{B}, \tag{1·5}$$

where $\boldsymbol{\tau}$ is a vector of fixed effects, $\boldsymbol{\gamma}$ is a vector of random effects with expectation zero and $\boldsymbol{\epsilon}$ is a random error vector with expectation zero. Such a model in the general case when $\mathbf{A} \neq \mathbf{B}$ has been studied by several authors (Duncan, 1960; Henderson *et al.*, 1959) under simpler assumptions on the covariance structures of $\boldsymbol{\gamma}$ and $\boldsymbol{\epsilon}$. The special case considered in this paper leads to a satisfactory solution for the problem of inference on the unknown parameter $\boldsymbol{\tau}$ under less restrictive conditions on the dispersion matrices of $\boldsymbol{\gamma}$ and $\boldsymbol{\epsilon}$.

2. Estimation of T and $\boldsymbol{\tau}$

We shall present the results on the estimation of linear functions of $\mathbf{T}$ and $\boldsymbol{\tau}$ in a series of lemmas.

LEMMA 1. *Let* $\mathbf{P}'\mathbf{T}$ *be a linear function of* $\mathbf{T}$ *and* $\alpha + \boldsymbol{\beta}'\mathbf{Y}$ *be a linear estimator (predictor) of* $\mathbf{P}'\mathbf{T}$. *Then the optimum values of* α *and* $\boldsymbol{\beta}$, *for which* $E(\mathbf{P}'\mathbf{T} - \alpha - \boldsymbol{\beta}'\mathbf{Y})^2$ *is a minimum, are*

$$\left.\begin{aligned} \alpha^* &= \boldsymbol{\tau}'\mathbf{P} - \boldsymbol{\tau}'\mathbf{A}'\boldsymbol{\beta}^*, \\ \boldsymbol{\beta}^* &= (\mathbf{A}\boldsymbol{\Lambda}\mathbf{A}' + \boldsymbol{\Sigma})^{-1}\mathbf{A}\boldsymbol{\Lambda}\mathbf{P} \end{aligned}\right\} \tag{2·1}$$

and the predictive efficiency $E(\mathbf{P}'\mathbf{T} - \alpha^* - \boldsymbol{\beta}^{*\prime}\mathbf{Y})^2$ *is*

$$e\mathbf{P}'\boldsymbol{\Lambda}\mathbf{P} - e\mathbf{P}'\boldsymbol{\Lambda}\mathbf{A}'(\mathbf{A}\boldsymbol{\Lambda}\mathbf{A}' + \boldsymbol{\Sigma})^{-1}\mathbf{A}\boldsymbol{\Lambda}\mathbf{P}. \tag{2·2}$$

The result quoted in Lemma 1 is simply the linear regression function of $\mathbf{P}'\mathbf{T}$ on $\mathbf{Y}$ and provides a complete answer to the estimation (prediction) of the random variable $\mathbf{P}'\mathbf{T}$ when $\boldsymbol{\tau}$ and the dispersion matrices $\boldsymbol{\Lambda}$ and $\boldsymbol{\Sigma}$ are known. When $\boldsymbol{\tau}$ is unknown the result of Lemma 2 provides a satisfactory solution.

LEMMA 2. *The optimum values of* α *and* $\boldsymbol{\beta}$ *for which*

(i) $E(\mathbf{P}'\mathbf{T} - \alpha - \boldsymbol{\beta}'\mathbf{Y}) = 0$ *for all values of* $\boldsymbol{\tau}$, (2·3)

(ii) $E(\mathbf{P}'\mathbf{T} - \alpha - \boldsymbol{\beta}'\mathbf{Y})^2$ *is a minimum* (2·4)

are

$$\alpha^* = 0, \boldsymbol{\beta}^* = \boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{P} \tag{2·5}$$

and the predictive efficiency is

$$e\mathbf{P}'(\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{P}. \tag{2·6}$$

Consider

$$E(\mathbf{P}'\mathbf{T} - \alpha^* - \boldsymbol{\beta}^{*\prime}\mathbf{Y}) = E[\mathbf{P}'\mathbf{T} - \mathbf{P}'(\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}^{-1}(\mathbf{A}\mathbf{T} + \boldsymbol{\epsilon})] = 0$$

so that the condition (2·3) is satisfied. If α, $\boldsymbol{\beta}$ are any arbitrary values meeting the condition (2·3) then

$$E(\mathbf{P}'\mathbf{T} - \alpha - \boldsymbol{\beta}'\mathbf{Y}) = \mathbf{P}'\boldsymbol{\tau} - \alpha - \boldsymbol{\beta}'\mathbf{A}\boldsymbol{\tau} = 0 \Rightarrow \alpha = 0, \mathbf{P} = \mathbf{A}'\boldsymbol{\beta}.$$

For any such α, $\boldsymbol{\beta}$, it is easy to verify that

$$(\boldsymbol{\beta} - \boldsymbol{\beta}^*)'\boldsymbol{\Sigma}\boldsymbol{\beta}^* = 0$$

so that

$$\begin{aligned} E(\mathbf{P}'\mathbf{T} - \boldsymbol{\beta}'\mathbf{Y})^2 &= E(\boldsymbol{\beta}'\boldsymbol{\epsilon})^2 \\ &= E[\boldsymbol{\beta}^{*\prime}\boldsymbol{\epsilon} + (\boldsymbol{\beta} - \boldsymbol{\beta}^*)'\boldsymbol{\epsilon}]^2 \\ &= E(\boldsymbol{\beta}^{*\prime}\boldsymbol{\epsilon})^2 + E[(\boldsymbol{\beta} - \boldsymbol{\beta}^*)'\boldsymbol{\epsilon}]^2 \\ &\geqslant E(\boldsymbol{\beta}^{*\prime}\boldsymbol{\epsilon})^2 = (\mathbf{P}'\mathbf{T} - \boldsymbol{\beta}^{*\prime}\mathbf{Y})^2, \end{aligned}$$

which proves the required result. The predictive efficiency is

$$E(\boldsymbol{\beta}^{*\prime}\boldsymbol{\epsilon})^2 = e\mathbf{P}'(\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{P}.$$

It is interesting to note that the condition (2·3) yields a predictor which is not only independent of $\boldsymbol{\tau}$ but also of $\boldsymbol{\Lambda}$. Further the predictive efficiency is also independent of $\boldsymbol{\tau}$ and $\boldsymbol{\Lambda}$. In the special case $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$ the optimum value of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta}^* = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{P},$$

which is independent of σ^2 also. The predictive efficiency is $\sigma^2\mathbf{P}'(\mathbf{A}'\mathbf{A})^{-1}\mathbf{P}$ which, however, involves σ^2.

We shall now consider the estimation of a linear parametric function $\mathbf{P}'\boldsymbol{\tau}$ by a linear function $\alpha+\boldsymbol{\beta}'\mathbf{Y}$ of the random variable $\mathbf{Y}$.

LEMMA 3. *The optimum values of* α, $\boldsymbol{\beta}$ *for which*

(i) $E(\mathbf{P}'\boldsymbol{\tau}-\alpha-\boldsymbol{\beta}'\mathbf{Y}) = 0$ *for all values of* $\boldsymbol{\tau}$, (2·7)

(ii) $E(\mathbf{P}'\boldsymbol{\tau}-\alpha-\boldsymbol{\beta}'\mathbf{Y})^2$ *is a minimum* (2·8)

are

$$\alpha^* = 0, \quad \beta^* = \boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{P}. \tag{2·9}$$

The variance of the estimator is

$$e\mathbf{P}'\boldsymbol{\Lambda}\mathbf{P}+e\mathbf{P}'(\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{P}. \tag{2·10}$$

The condition (2·7) implies that $\alpha = 0$ and $\mathbf{A}'\boldsymbol{\beta} = \mathbf{P}$ in which case

$$E(\boldsymbol{\beta}'\mathbf{Y}-\mathbf{P}'\boldsymbol{\tau})^2 = e\mathbf{P}'\boldsymbol{\Lambda}\mathbf{P}+e\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}.$$

Then the problem is that of minimizing the quadratic form $\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}$ subject to the condition $\mathbf{A}'\boldsymbol{\beta} = \mathbf{P}$. In such a case it is well known that the optimum value of $\boldsymbol{\beta}$ is $\boldsymbol{\beta}^*$ as given in (2·9).

It may be noted that *the linear unbiased minimum variance estimator of* $\mathbf{P}'\boldsymbol{\tau}$ *is independent of* $\boldsymbol{\Lambda}$ *and is the same as the best predictor of* $\mathbf{P}'\mathbf{T}$ *as in Lemma* 2, *but the variance of the estimator of* $\mathbf{P}'\boldsymbol{\tau}$ *depends on* $\boldsymbol{\Lambda}$. *In the special case* $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$

$$\boldsymbol{\beta}^* = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{P},$$

which is independent of σ^2 *also. Thus the least-square estimator of* $\mathbf{P}'\boldsymbol{\tau}$ *under the covariance structure* $D(\mathbf{Y}) = \mathbf{A}\boldsymbol{\Lambda}\mathbf{A}'+\sigma^2\mathbf{I}$ *is the same as when the components of* $\mathbf{Y}$ *are uncorrelated.* The expressions for the variances are, however, different.

In the rest of the discussion we shall consider the special case $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$ with an independent unbiased estimator $f^{-1}\mathbf{S}$ of $\mathbf{A}\boldsymbol{\Lambda}\mathbf{A}'+\sigma^2\mathbf{I}$. Lemma 4 gives unbiased estimators of σ^2 and $\boldsymbol{\Lambda}$.

LEMMA 4.

$$E[\text{trace}(\mathbf{I}-\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')(\mathbf{S}+e^{-1}\mathbf{Y}\mathbf{Y}')] = (p-m)(f+1)\sigma^2 \tag{2·11}$$

$$E[(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{S}\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}-f\hat{\sigma}^2(\mathbf{A}'\mathbf{A})^{-1}] = f\boldsymbol{\Lambda}, \tag{2·12}$$

where $\hat{\sigma}^2$ *is the expression within the expectation of* (2·11) *divided by* $(p-m)(f+1)$.

The expression (2·11) is equal to

$$\begin{aligned}
&\text{trace}\{(\mathbf{I}-\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')[E(\mathbf{S})+E(e^{-1}\mathbf{Y}\mathbf{Y}')]\}\\
&\quad= \text{trace}[(\mathbf{I}-\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')(\mathbf{A}\boldsymbol{\Lambda}\mathbf{A}'+\sigma^2\mathbf{I})f]\\
&\qquad+\text{trace}[(\mathbf{I}-\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')(\mathbf{A}\boldsymbol{\Lambda}\mathbf{A}'+\sigma^2\mathbf{I}+e^{-1}\mathbf{A}\boldsymbol{\tau}\boldsymbol{\tau}'\mathbf{A}')]\\
&\quad= \text{trace}[(f+1)\sigma^2(\mathbf{I}-\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')]\\
&\quad= \sigma^2(f+1)(p-m).
\end{aligned}$$

Starting from the equation $E(\mathbf{S}) = f(\mathbf{A}\Lambda\mathbf{A}' + \sigma^2\mathbf{I})$ and multiplying both sides by suitable matrices the result (2·12) is deduced.

Lemmas 1-4 contain general results depending only on the first and second order moments of the variables $\mathbf{T}$ and $\boldsymbol{\epsilon}$. We shall now assume specific distributions for the variables involved and deduce some results which are useful for drawing inferences on the unknown variables and parameters. Lemma 5 contains the key results in this direction. We also introduce a new definition which is relevant in the discussion of Lemma 5 (see also Fraser, 1956).

DEFINITION. Let $\mathbf{X}$ be a vector valued random variable whose distribution depends on $\boldsymbol{\theta}$, a vector of unknown parameters $(\theta_1, \ldots, \theta_m)$. Further let $\mathbf{T} = (T_1, \ldots T_k)$ be a vector valued function of $\mathbf{X}$ such that the distribution of $\mathbf{T}$ depends only on $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_s)$, some functions of $\boldsymbol{\theta}$. Then $\mathbf{T}$ is said to be *inference sufficient for* $\boldsymbol{\phi}$ if the conditional distribution of $\mathbf{X}$ given $\mathbf{T}$ depends only on parametric functions of $\boldsymbol{\theta}$ independent of $\boldsymbol{\phi}$.

LEMMA 5. *Let* $\mathbf{T} \sim N_p(\mathbf{A}\boldsymbol{\tau}, e\Lambda)$, $\boldsymbol{\epsilon} \sim N_p(0, e\sigma^2\mathbf{I})$ *and* $\mathbf{S} \sim W_p(f, \mathbf{A}\Lambda\mathbf{A}' + \sigma^2\mathbf{I})$ *be all stochastically independent. Further let* $\boldsymbol{\Theta} = \Lambda + \sigma^2(\mathbf{A}'\mathbf{A})^{-1}$ *and* $G = \text{trace}[(\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')(\mathbf{S} + e^{-1}\mathbf{Y}\mathbf{Y}')]$. *Then*

(i) $\mathbf{Z} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Y} \sim N_m(\boldsymbol{\tau}, e\boldsymbol{\Theta})$ *and* $\mathbf{U} = (\mathbf{A}'\mathbf{A})^{-1}(\mathbf{A}'\mathbf{S}\mathbf{A})(\mathbf{A}'\mathbf{A})^{-1} \sim W_m(f, \boldsymbol{\Theta})$ *are stochastically independent.*

(ii) $\mathbf{Z}$ and $\mathbf{U}$ *are inference sufficient for the parameters* $\boldsymbol{\tau}$ *and* $\boldsymbol{\Theta}$.

(iii) $(\mathbf{T} - \mathbf{Z}) \sim N_m(0, e\sigma^2(\mathbf{A}'\mathbf{A})^{-1})$ *and* $G \sim \sigma^2\chi^2(p-m)(f+1)$ *are stochastically independent.*

It is easy to deduce all the results (i)-(iii). As an example, we demonstrate (ii).

Let $\mathbf{B}$ be a $p \times (p-m)$ matrix of rank $(p-m)$ and with its columns orthonormal to those of $\mathbf{A}$, i.e. $\mathbf{A}'\mathbf{B} = 0$. Then the following are true:

(*a*) $\mathbf{A}'\mathbf{Y} \sim N_m((\mathbf{A}'\mathbf{A})\boldsymbol{\tau}, e\mathbf{A}'\mathbf{A}\boldsymbol{\Theta}\mathbf{A}'\mathbf{A})$ and $\mathbf{B}'\mathbf{Y} \sim N_{p-m}(0, e\sigma^2\mathbf{B}\mathbf{B}')$ and are independent.

(*b*) The random matrix

$$\begin{pmatrix} \mathbf{A}'\mathbf{S}\mathbf{A} & \mathbf{A}'\mathbf{S}\mathbf{B} \\ \mathbf{B}'\mathbf{S}\mathbf{A} & \mathbf{B}'\mathbf{S}\mathbf{B} \end{pmatrix}$$

has Wishart's distribution on f degrees of freedom with the dispersion matrix

$$\begin{pmatrix} \mathbf{A}'\mathbf{A}\boldsymbol{\Theta}\mathbf{A}'\mathbf{A} & 0' \\ 0 & \sigma^2\mathbf{B}'\mathbf{B} \end{pmatrix}$$

so that the exponent of the density in Wishart's distribution consists of the sum of two expressions one involving $\mathbf{A}'\mathbf{S}\mathbf{A}$ and $\mathbf{A}'\mathbf{A}\boldsymbol{\Theta}\mathbf{A}'\mathbf{A}$ and another involving $\mathbf{B}'\mathbf{S}\mathbf{B}$ and $\sigma^2\mathbf{B}'\mathbf{B}$.

Now, writing the joint density of $\mathbf{A}'\mathbf{Y}$, $\mathbf{B}'\mathbf{Y}$ and $\mathbf{A}'\mathbf{S}\mathbf{A}$, $\mathbf{A}'\mathbf{S}\mathbf{B}$, $\mathbf{B}'\mathbf{S}\mathbf{B}$, we find that it is the product of two factors, one involving $\mathbf{A}'\mathbf{Y}$, $\mathbf{A}'\mathbf{S}\mathbf{A}$, $\boldsymbol{\tau}$ and $\boldsymbol{\Theta}$ and another involving σ^2 only. Thus $\mathbf{A}'\mathbf{Y}$, $\mathbf{A}'\mathbf{S}\mathbf{A}$ are inference sufficient for $\boldsymbol{\tau}$ and $\boldsymbol{\Theta}$ and so also are $\mathbf{Z} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Y}$ and $\mathbf{U} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{S}\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}$, which are one to one functions of $\mathbf{A}'\mathbf{Y}$ and $\mathbf{A}'\mathbf{S}\mathbf{A}$.

3. INFERENCE ON T AND τ

Test for the model (1·1)-(1·3). In problems of inference on $\mathbf{T}$ and $\boldsymbol{\tau}$, it is relevant first to examine whether the model (1·1)-(1·3) is true. We shall develop a test for this purpose based on $\mathbf{Y}$ and $\mathbf{S}$, using the principle of the likelihood ratio. It is, however, convenient to test the hypothesis about the truth of the model (1·1)-(1·3) in two parts.

First, to examine on the basis of $\mathbf{S}$ whether the dispersion matrix is of the form

$$\mathbf{A}'\Lambda\mathbf{A} + \sigma^2\mathbf{I}.$$

For this we consider the two alternative distributions

$$\mathbf{S} \sim W_p(f, \mathbf{\Gamma}),\ \mathbf{\Gamma} \text{ arbitrary}$$

and

$$\mathbf{S} \sim W_p(f, \mathbf{A\Lambda A'} + \sigma^2\mathbf{I}),\ \mathbf{\Lambda} \text{ and } \sigma^2 \text{ arbitrary.}$$

Maximizing the densities in each case with respect to the arbitrary parameters and taking the ratio, the likelihood ratio test criterion is found to be

$$\Lambda = \frac{[|\mathbf{A'A}|\,|\mathbf{S}|]^{\frac{1}{2}f}}{|\mathbf{A'SA}|^{\frac{1}{2}f}\left[\frac{1}{p-m}\operatorname{trace}\mathbf{S}(\mathbf{I}-\mathbf{A}(\mathbf{A'A})^{-1}\mathbf{A'})\right]^{\frac{1}{2}f(p-m)}}. \tag{3·1}$$

If f is large, the statistic $-2\log_e \Lambda$ can be used as χ^2 on $\frac{1}{2}[(p-m)(p+m+1)-2]$ degrees of freedom.

Secondly, to examine on the basis of **Y** and **S** whether the expectation of **Y** is of the form $\mathbf{A\tau}$, without assuming any structure for the dispersion matrix. A test for this was given in the 1959 paper of the author. The statistic to be used is

$$\frac{(f-p+m+1)}{e(p-m)}[\mathbf{Y'S^{-1}Y} - \mathbf{Y'S^{-1}A(A'S^{-1}A)^{-1}A'S^{-1}Y}], \tag{3·2}$$

which has the variance ratio distribution on $(p-m)$ and $(f-p+m+1)$ degrees of freedom (D.F.).

Confidence intervals for linear functions of T. It has been shown in (iii) of Lemma 5 that

$$(\mathbf{T}-\mathbf{Z}) \sim N_m(0, e\sigma^2(\mathbf{A'A})^{-1})$$

and

$$G \sim \sigma^2\chi^2(p-m)(f+1),$$

where **Z** and G are functions of **Y** and **S** as defined in Lemma 5. Further $\mathbf{T}-\mathbf{Z}$ and G are independently distributed from which it follows that $\mathbf{P'(T-Z)} \sim N_1(0, e\sigma^2\mathbf{P'(A'A)^{-1}P})$ and G are independently distributed. Hence substituting the estimate of σ^2 based on G,

$$\mathbf{P'(T-Z)} \div [e\mathbf{P'(A'A)^{-1}P}v]^{\frac{1}{2}}, \tag{3·3}$$

where $v = [(p-m)(f+1)]^{-1}G$, has a t-distribution on $(p-m)(f+1)$ D.F. Hence a confidence interval for $\mathbf{P'T}$ with $(1-\alpha)$ probability is

$$\mathbf{P'Z} \pm [(e\mathbf{P'(A'A)^{-1}P}v)]^{\frac{1}{2}}t_{\frac{1}{2}\alpha}, \tag{3·4}$$

where $t_{\frac{1}{2}\alpha}$ is the upper $\frac{1}{2}\alpha$ point of the t-distribution on $(p-m)(f+1)$ D.F.

Simultaneous confidence intervals for $\mathbf{P'T}$ where **P** is arbitrary are

$$\mathbf{P'Z} \pm (e\mathbf{P'(A'A)^{-1}P}vmF_\alpha)^{\frac{1}{2}}, \tag{3·5}$$

where F_α is the upper α probability value of the F-distribution on m and $(p-m)(f+1)$ D.F.

Confidence intervals for linear functions of τ. It has been shown in (i) of Lemma 5 that

$$\mathbf{Z} \sim N_m(\boldsymbol{\tau}, e\mathbf{\Theta}), \quad \mathbf{U} \sim W_m(f, \mathbf{\Theta})$$

and further that **Z** and **U** are independent, where **U** and **Z** are as defined in Lemma 5. Hence

$$\mathbf{P'(Z-\tau)} \div (f^{-1}e\mathbf{P'UP})^{\frac{1}{2}}, \tag{3·6}$$

has a t-distribution on f D.F. Thus a $(1-\alpha)$ probability confidence interval for $\mathbf{P'\tau}$ is

$$\mathbf{P'Z} \pm (f^{-1}e\mathbf{P'UP})^{\frac{1}{2}}t_{\frac{1}{2}\alpha}. \tag{3·7}$$

Simultaneous confidence intervals for $\mathbf{P}'\boldsymbol{\tau}$, where $\mathbf{P}$ is arbitrary, are

$$\mathbf{P}'\mathbf{Z} \pm \left[\frac{em\mathbf{P}'\mathbf{U}\mathbf{P}}{f-m+1} F_a\right]^{\frac{1}{2}}, \tag{3·8}$$

where F_a has m and $f-m+1$ D.F.

4. ESTIMATION OF POLYNOMIAL GROWTH CURVES

During recent years there have been a number of papers devoted to the estimation of polynomial growth curves (Rao, 1959; Elston & Grizzle, 1962; Elston, 1964; Potthoff & Roy, 1964) and to the comparison of growth curves (Wishart, 1938; Leech & Healy, 1959; Rao, 1958, 1961). Based on the discussions contained in these papers and in the present paper it is now possible to suggest a systematic approach to this problem.

1. As a first step we replace the measurements at different time points of growth of an individual by *orthogonal polynomial regression coefficients* (O.P.R.C.). Thus if $y_1, \ldots, y_p$ are the measurements on an individual at p time points, then the regression coefficients are

$$b_i = \sum_j y_j \phi_{ij} \quad (i = 0, \ldots, p-1), \tag{4·1}$$

where ϕ_{ij} is the value of the ith degree orthogonal polynomial at the jth time point. Nothing is lost by replacing $y_1, \ldots, y_p$ by the O.P.R. coefficients $b_0, \ldots, b_{p-1}$ as both sets are equivalent. Then the data consist of independent observations (say n in number) on the vector variable $(b_0, \ldots, b_{p-1})$.

The values of ϕ_{ij} as given in statistical tables are not usually standardized. They may be used as such in computing b_i, and at the final stage of estimation of the true regression coefficients, the necessary adjustments can be made by using multiplying factors.

2. The second step is to obtain the sample mean and the corrected sum of products matrix for the observations on $(b_0, \ldots, b_{p-1})$,

$$\left.\begin{array}{l} \bar{b}_0, \ldots, \bar{b}_{p-1}, \\ (S_{ij}) \quad (i,j = 0, \ldots, p-1), \end{array}\right\} \tag{4·2}$$

which may be represented in matrix notation by $\bar{\mathbf{b}}$ and $\mathbf{S}$ respectively.

3. The third step is the most important, which is to examine whether a subset of the O.P.R. coefficients is inference sufficient. This depends on the degree of the polynomial of the assumed growth curve and also on the structure of the true dispersion matrix of $(b_0, \ldots, b_{p-1})$.

In practice, the degree of the polynomial and the covariance structure are unknown and are themselves to be inferred from the available data and hence a careful approach to the problem is needed. If the degree of the polynomial is k, then $b_0, \ldots, b_k$ have (after standardization) as their expected values the true coefficients of the polynomial $\beta_0, \ldots, \beta_k$ and inferences may be drawn on the true values using the observations on $(b_0, \ldots, b_k)$ only. This is the basic approach of Wishart (1938), Elston & Grizzle (1962) and also of Potthoff & Roy (1964). The last authors, however, use a different set of $(k+1)$ reduced values obtained from the original p values, which are arbitrary to some extent, depending on the accuracy and relevance of previous information about the dispersion matrix of the original measurements.

The main departure in the present approach is to examine whether in addition to $b_0, \ldots, b_k$ some of the higher order O.P.R. coefficients $b_{k+1}, \ldots, b_{p-1}$, whose expectations are zero when the degree of the polynomial is k, yield information on $\beta_0, \ldots, \beta_k$ through their correlations with $b_0, \ldots, b_k$. Then we can hope to obtain improved estimates of $\beta_0, \ldots, \beta_k$ by making covariance adjustments choosing some of the coefficients in $(b_{k+1}, \ldots, b_{p-1})$ as *concomitant* to the main coefficients $(b_0, \ldots, b_k)$.

The method developed in the 1959 paper of the author is equivalent to using the entire remaining set $b_{k+1}, \ldots, b_{p-1}$ as concomitants. But this may not be the optimum procedure (see Rao, 1949). The choice of a suitable subset may be more profitable, especially in small samples, when all the correlations between the main and concomitant coefficients are not high. Further p may be large compared to n and consequently the degrees of freedom for the estimation of error will be small if a large number of concomitants are used. Such a situation was faced in the comparison of growth rates discussed in another paper (Rao, 1961), where only a few extra coefficients were computed to begin with and among them only one was chosen for covariance adjustment.

It would indeed be simpler if $(b_0, \ldots, b_k)$ alone were inference sufficient for $(\beta_0, \ldots, \beta_k)$. This is the case if the measurement on an individual at time t can be written

$$y_t = T_0 \phi_0(t) + T_1 \phi_1(t) + \ldots + T_k \phi_k(t) + \epsilon_t, \tag{4·3}$$

where $\phi_0, \ldots, \phi_k$ are orthogonal polynomials, $T_0, \ldots, T_k$ are random variables specific to an individual and ϵ_t are such that

$$\left.\begin{aligned} &E(\epsilon_t) = 0, E(\epsilon_t^2) = \sigma^2, \operatorname{cov}(\epsilon_{t_1}, \epsilon_{t_2}) = 0 \quad (t_1 \neq t_2); \\ &\operatorname{cov}(T_i, \epsilon_t) = 0 \quad \text{for all} \quad i \text{ and } t. \end{aligned}\right\} \tag{4·4}$$

In such a case the dispersion matrix of the coefficients $b_0, \ldots, b_{p-1}$ is of the form

$$\begin{pmatrix} \Lambda & 0 \\ 0 & \sigma^2 \mathbf{I} \end{pmatrix}, \tag{4·5}$$

where Λ is of order $(k+1) \times (k+1)$ and $\mathbf{I}$ is of order $(p-k-1) \times (p-k-1)$.

We can use the test (3·1) developed in §3 for examining the validity of the model (4·3) leading to the covariance structure (4·5) for the coefficients $b_0, \ldots, b_{p-1}$. It may be found that some of the cross-correlations between the sets $(b_0, \ldots, b_k)$ and $(b_{k+1}, \ldots, b_{p-1})$ are not zero as assumed in the model, in which case some of the variables in the set $(b_{k+1}, \ldots, b_{p-1})$ may be chosen for covariance adjustment. We shall illustrate the next steps in the analysis of data through some numerical illustrations.

The estimates $\bar{\mathbf{b}}$, S based on the data of Elston & Grizzle (1962), the correlation matrix **R** derived from S and the F statistic (square of t) for testing the significance of each observed average individually are given in Table 1. Because of symmetry, only the elements of S in the diagonal and above it are shown and in the case of **R**, only elements below the diagonal are shown. The sample size was 20 so that the degrees of freedom for F are 1 and 19.

Table 1. *Analysis of data for ramus height in 20 boys measured at four ages*

$\bar{b}$	$F_{1,19}$	S and R matrices			
$\bar{b}_0 =$ 200·30	8005*	1904·500	136·700	−14·388	−22·800
$\bar{b}_1 =$ 9·33	51·24*	0·1233	645·162	15·874	−124·756
$\bar{b}_2 =$ −0·09	0·18	−0·0827	0·1517	16·978	−3·712
$\bar{b}_3 =$ −0·04	0·01	−0·0629	−0·5812*	−0·1084	69·028

* Indicates significance at 1% level.

It is clear from the preliminary analysis of Table 1 that the degree of the polynomial may be taken to be unity so that b_0 and b_1 are the main coefficients, and a possible candidate for a concomitant is b_3, the correlations between b_3 and (b_0, b_1) being small. Thus we may consider b_0, b_1, b_3 for drawing inferences on β_0, β_1. We shall, however, demonstrate the use of the tests developed in the present paper to support the preliminary conclusions.

(i) For testing the significance of $\bar{b}_2$ and $\bar{b}_3$ simultaneously we use Hotelling's T^2, which in the present case has a small value. The general formula for T^2 for testing the significance of the observed average coefficients $\bar{b}_{k+1} \ldots, \bar{b}_{p-1}$ is as follows. Let

$$\mathbf{S}_2 = \begin{pmatrix} S_{k+1,k+1} & \cdots & S_{k+1,p-1} \\ S_{p-1,k+1} & \cdots & S_{p-1,p-1} \end{pmatrix}, \tag{4·6}$$

which is the portion of the S matrix relating to the coefficients $b_{k+1}, \ldots, b_{p-1}$. Denote by $\bar{\mathbf{b}}_2$ the column vector of $\bar{b}_{k+1}, \ldots, \bar{b}_{p-1}$. Then

$$T^2 = n\bar{\mathbf{b}}_2' \mathbf{S}_2^{-1} \bar{\mathbf{b}}_2 \tag{4·7}$$

in which case

$$\frac{n-p+k+1}{p-k-1} T^2 \tag{4·8}$$

is a variance ratio on $(p-k-1)$ and $(n-p+k+1)$ D.F.

(ii) For testing the covariance structure which implies the sufficiency of $b_0, \ldots, b_k$ for estimating a kth degree polynomial, the test based on the S matrix of $b_0, \ldots, b_{p-1}$, may be stated as

$$\Lambda = \Lambda_1 \Lambda_2. \tag{4·9}$$

In (4·9)

$$\Lambda_1 = \frac{|\mathbf{S}|}{|\mathbf{S}_1|\,|\mathbf{S}_2|}, \tag{4·10}$$

where $\mathbf{S}_1$ is the submatrix of S corresponding to $b_0, \ldots, b_k$ and

$$\Lambda_2 = \frac{|\mathbf{S}_2|}{a_{k+1} \ldots a_{p-1} s^{p-k-1}}, \tag{4·11}$$

where a_i is the sum of squares of the unstandardized values of the orthogonal polynomial of degree i as given in the statistical tables used, and

$$s = \frac{a_{k+1}^{-1} S_{k+1,k+1} + \ldots + a_{p-1}^{-1} S_{p-1,p-1}}{p-k-1}. \tag{4·12}$$

The test based on Λ_1, using $-(n-1)\log_e \Lambda_1$ as approximately a χ^2 on $(p-k-1)(k+1)$ D.F., is enough to judge the inference sufficiency of $(b_0, \ldots, b_k)$. The test based on Λ_2, using $-(n-1)\log_e \Lambda_2$ as approximately a χ^2 on $\frac{1}{2}(p-k-2)(p-k+1)$ D.F., examines the independence and homoscedasticity of the error terms ϵ_i for an individual growth curve.

The combined statistic

$$\chi^2 = -(n-1)\log_e \Lambda_1 - (n-1)\log_e \Lambda_2 \tag{4·13}$$

can be used as a χ^2 on $[(p-k-1)(k+1) + \frac{1}{2}(p-k-2)(p-k+1)]$ D.F. to test the model (4·3) as a whole for a growth curve. The statistic (4·9) as defined in terms of the S matrix of the coefficients (transformed variables) $b_0, \ldots, b_{p-1}$ is the same as the statistic (3·1) defined in terms of the S matrix of the original measurements.

In the present example choosing $k = 1$, we find that $\Lambda_1 = 0{\cdot}6358$ and $\Lambda_2 = 0{\cdot}9778$, giving

$$\chi_1^2 = -19\log_e(0{\cdot}6358) = 8{\cdot}61$$

on 4 D.F. and

$$\chi_2^2 = -19\log_e(0{\cdot}9778) = 0{\cdot}43$$

on 2 D.F. The value 8·610 for χ^2 on 4 D.F. is high enough to warrant the use of concomitant variables.

Table 2 gives the estimates of β_0 and β_1 without using any concomitants, using b_3 alone and using both b_2 and b_3 as concomitants. The method of covariance adjustment in the general case is explained in Appendix 1. It is seen from Table 2 that it is profitable to use b_3 alone as a concomitant.

Table 2. *Estimate and width of* 95 % *confidence interval*

Parameter	Using only b_0, b_1	Adjusting for concomitants b_3	b_2, b_3
β_0: Estimate*	50·07	50·07	50·05
Width	2·34	2·31	2·49
β_1: Estimate*	0·4665	0·4629	0·4654
Width	0·273	0·227	0·249

* The multiplying factors for b_0 and b_1 are $\frac{1}{4}$ and $\frac{1}{16}$, respectively.

Let us now consider the example of Potthoff & Roy (1964) and compare the estimates given by them with those obtained by the present method. The estimates $\bar{\mathbf{b}}$, S and R are given in Table 3 together with the F statistics as in Table 1. It is clear from the preliminary analysis that the degree of the polynomial may be taken to be unity and because of the low correlations between (b_0, b_1) and (b_2, b_3) covariance adjustment is not profitable. Thus we can draw inferences on β_0, β_1 based on (b_0, b_1) only. The estimates so obtained are compared with the estimates given by Potthoff & Roy (1964) in Table 4. The multiplying factors for (b_0, b_1) are the same as those used in Table 2. It is seen from Table 4 that the method proposed in the present paper provides slightly more efficient estimates. In other situations

Table 3. *Analysis of measurements made in dental study of* 16 *boys*

$\bar{b}_i$	$F_{1,15}$	S and R matrices			
$\bar{b}_0 = 99{\cdot}8750$	2982*	802·7500	−5·6250	−35·8750	−213·7500
$\bar{b}_1 = 15{\cdot}6875$	59·63*	−0·0063	990·4375	−91·4375	105·8750
$\bar{b}_2 = 0{\cdot}8125$	1·88	−0·1383	−0·3162	84·4375	4·6250
$\bar{b}_3 = -1{\cdot}1250$	0·26	−0·2221	0·0987	0·0148	1161·7500

* Indicates significance at 1% level.

Table 4. *Estimate and width of* 95 % *confidence interval*

Parameter	Using b_0, b_1	Method of Potthoff & 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

where the use of concomitants is indicated there may be considerable loss of efficiency by using the method of Potthoff & Roy. For other comments on the method of Potthoff & Roy the reader is referred to Appendix 2.

There is an important point in the statistical analysis which needs some caution. We have not taken into account the fact that the concomitant variables are selected on the basis of certain tests on the observed data and consequently the number and nature of the concomitant variables may vary from sample to sample. The precision is likely to be overestimated if no adjustment is made for such a selection of variables. Any theoretical investigation of this problem is likely to be extremely complicated. Similar problems arise in many other situations involving preliminary tests of significance intended to examine the underlying probability model used in the final statistical analysis. In such problems previous experience will also be a good guide. It may be noted that in research work, one is not faced with an isolated problem of estimating a single growth curve. A series of such curves will be studied under similar conditions involving routine estimation of the polynomial growth coefficients and in some cases a comparison of such estimates. It should be possible to arrive at a suitable set of concomitants for such purposes by an analysis of previous data.

The method of estimation of the polynomial coefficients described above can be extended to the comparison of growth curves under different experimental conditions (treatments) using any suitable design. The first step is to replace the observations on each individual growth curve by the O.P.R.C. The number of O.P.R.C.'s may be smaller than the number of observations on each curve. The second step is to obtain an analysis of dispersion considering the O.P.R.C.'s as multiple measurements, appropriate to the design of experiment used. We thus obtain dispersion matrices due to treatments and due to error which are used in further analysis. The main interest of analysis of data is to estimate the differences in growth curves (or the O.P.R.C.'s) caused by the treatments. Then we have the problem of examining which of the O.P.R.C.'s (or their functions) are unaffected by treatments and the possibility of using them as concomitants in estimating the differences in other O.P.R.C.'s (or their functions). Such questions can be examined by analysis of dispersion.

An interesting example of such an analysis is due to Leech & Healy (1959), who use a function of the O.P.R.C.'s providing a graduation of the initial measurement of a growth curve (at time point zero) as a concomitant. This is justified by the fact that individuals are assigned at random to different treatments and therefore the initial differences have expectation zero. For some comments on this analysis and a general approach in such cases the reader is referred to a paper by the author (Rao, 1961).

Note added in proof. Since submitting this paper, further results have been obtained on the selection of concomitants and on methods of inference allowing for variation in the concomitants. These will be published in the Proceedings of the Fifth Berkeley Symposium.

APPENDIX 1. ADJUSTMENT FOR CONCOMITANT VARIATION

Let $\mathbf{X}$ be $n\times r$ matrix of main variables and $\mathbf{Z}$ be $n\times q$ matrix of concomitant variables. Let

$$\underset{n\times r}{E(\mathbf{X})} = \underset{n\times m}{\mathbf{B}}\ \underset{m\times r}{\boldsymbol{\xi}},\quad \underset{n\times q}{E(\mathbf{Z})} = \mathbf{0} \tag{A 1·1}$$

while

$$E(\mathbf{X}|\mathbf{Z}) = \mathbf{B}\boldsymbol{\xi}+\mathbf{Z}\boldsymbol{\rho}. \tag{A 1·2}$$

The variables in each row of $\mathbf{X}$ may be dependent while the sets of variable in the rows of $\mathbf{X}$ are independent. The problem is to estimate parametric functions of $\boldsymbol{\xi}$ and the conditional dispersion matrix $\boldsymbol{\Lambda}$ of the variables in a row, which is assumed to be the same for all the rows.

This is not a new problem since for given $\mathbf{Z}$ the expression on the right-hand side of (A 1·2) involves only known coefficients as multipliers of unknown parameters $\boldsymbol{\xi}$ and $\boldsymbol{\rho}$. Thus we have the Gauss–Markoff model for the multivariate case

$$E(\mathbf{X}) = \mathbf{C}\boldsymbol{\tau}, \tag{A 1·3}$$

where $\mathbf{C} = (\mathbf{B} \vdots \mathbf{Z})$ and $\boldsymbol{\tau}' = (\boldsymbol{\xi}' \vdots \boldsymbol{\rho}')$. The least-squares estimate of $\boldsymbol{\tau}$ is

$$\hat{\boldsymbol{\tau}} = \begin{pmatrix} \hat{\boldsymbol{\xi}} \\ \hat{\boldsymbol{\rho}} \end{pmatrix} = (\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{X} \tag{A 1·4}$$

and that of $\boldsymbol{\Lambda}$ is

$$(n-r-q)^{-1}[\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{X}]. \tag{A 1·5}$$

We are interested in the simultaneous estimation of linear parametric functions $\mathbf{P}'\boldsymbol{\xi}$. The estimate of $\mathbf{P}'\boldsymbol{\xi}$ is $\mathbf{P}'\hat{\boldsymbol{\xi}}$ where $\hat{\boldsymbol{\xi}}$ is obtained from the formula (A 1·4). The dispersion matrix of $\mathbf{P}'\hat{\boldsymbol{\xi}}$ is

$$(\mathbf{P}'\mathbf{E}\mathbf{P})\,\boldsymbol{\Lambda}, \tag{A 1·6}$$

where $\mathbf{P}'\mathbf{E}\mathbf{P}$ is a constant and $\mathbf{E}$ is the $(m \times m)$ submatrix obtained by omitting the last q columns and the last q rows of $(\mathbf{C}'\mathbf{C})^{-1}$. Since we have an estimate of $\boldsymbol{\Lambda}$, the inference on the parametric functions $\mathbf{P}'\boldsymbol{\xi}$ follows on the standard lines when the row vectors in $\mathbf{X}$ given $\mathbf{Z}$ have an r-variate normal distribution.

Alternative expressions for the estimates of $\boldsymbol{\xi}$ and $\boldsymbol{\Lambda}$, suitable for desk calculators, are as follows. Define

$$\hat{\boldsymbol{\xi}}_X = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{X}, \quad \hat{\boldsymbol{\xi}}_Z = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{Z}. \tag{A 1·7}$$

Then the equation for $\boldsymbol{\rho}$ alone can be written

$$(\mathbf{Z}'\mathbf{Z} - \mathbf{Z}'\mathbf{B}\hat{\boldsymbol{\xi}}_Z)\,\boldsymbol{\rho} = \mathbf{Z}'\mathbf{X} - \hat{\boldsymbol{\xi}}_Z'\mathbf{B}'\mathbf{X}. \tag{A 1·8}$$

Let $\hat{\boldsymbol{\rho}}$ be a solution of (A 1·8). Then the solution for $\boldsymbol{\xi}$ can be written

$$\hat{\boldsymbol{\xi}} = \hat{\boldsymbol{\xi}}_X - \hat{\boldsymbol{\xi}}_Z\hat{\boldsymbol{\rho}}. \tag{A 1·9}$$

The estimate of $(n-m-q)\,\boldsymbol{\Lambda}$ is

$$(\mathbf{X}'\mathbf{X} - \hat{\boldsymbol{\xi}}_X'\mathbf{B}'\mathbf{X}) - \hat{\boldsymbol{\rho}}'(\mathbf{Z}'\mathbf{X} - \boldsymbol{\xi}_Z'\mathbf{B}'\mathbf{X}). \tag{A 1·10}$$

The dispersion matrix of $\mathbf{P}'\hat{\boldsymbol{\xi}}$ is $c\boldsymbol{\Lambda}$ where

$$c = \mathbf{P}'(\mathbf{B}'\mathbf{B})^{-1}\mathbf{P} + \mathbf{P}'\hat{\boldsymbol{\xi}}_Z(\mathbf{Z}'\mathbf{Z} - \hat{\boldsymbol{\xi}}_Z'\mathbf{B}'\mathbf{Z})^{-1}\boldsymbol{\xi}_Z'\mathbf{P}. \tag{A 1·11}$$

The set up in terms of the O.P.R. coefficients in the illustrative examples is a special case of the model (A 1·1)–(A 1·2) with $\mathbf{X}$ as the matrix of the main O.P.R. coefficients on n individuals and $\mathbf{Z}$ as the matrix of the concomitant coefficients. The matrix $\mathbf{B}$ consists of a single column of n unit elements and $\boldsymbol{\xi}$ consists of a single row of $(k+1)$ parameters, viz. the expected values of the main O.P.R. coefficients.

The theory and analysis under the general set up (A 1·1)–(A 1·2) is useful in problems involving comparison of growth curves under different experimental conditions, etc.

Appendix 2. Comment on Potthoff & Roy's approach

Under the title, 'A generalized multivariate analysis of variance model useful especially for growth curve problems' Potthoff & Roy (1964) considered a linear model for the observations which appeared to be different from the Gauss–Markoff model and suggested its use in the estimation of polynomial growth curves. The model mentioned is

$$E(\mathbf{X}_0) = \mathbf{B}\boldsymbol{\xi}\mathbf{A}, \tag{A 2·1}$$

where $\mathbf{B}$ and $\mathbf{A}$ are known matrices of orders $n \times m$ and $p \times q$ and $\boldsymbol{\xi}$ is the matrix of unknown parameters and of order $m \times p$. The different rows of $\mathbf{X}_0$ are distributed mutually independently while the q elements in any row follow a q-variate normal distribution with the same dispersion matrix.

Construct a $q \times q$ non-singular matrix $\mathbf{H} = (\mathbf{H}_1 \vdots \mathbf{H}_2)$ such that $\mathbf{A}\mathbf{H}_2 = 0$ and the columns of $\mathbf{H}_1$ form a basis of the vector space generated by the rows of $\mathbf{A}$. Such a matrix $\mathbf{H}$ is not necessarily unique. Let r be the number of columns in $\mathbf{H}_1$. Multiplying both sides of (A 2·1) by $\mathbf{H}$, we find

$$E(\mathbf{X}_0\mathbf{H}_1) = \mathbf{B}\boldsymbol{\xi}\mathbf{A}\mathbf{H}_1, \quad E(\mathbf{X}_0\mathbf{H}_2) = 0. \tag{A 2·2}$$

The rank of $\mathbf{A}\mathbf{H}_1$ is evidently r and hence $\boldsymbol{\xi}\mathbf{A}\mathbf{H}_1$ can be replaced by a matrix $\boldsymbol{\eta}$ of independent parameters and of order $m \times r$ so that the set up (A 2·1) is equivalent to

$$E(\mathbf{X}) = \mathbf{B}\boldsymbol{\eta}, \quad E(\mathbf{Z}) = 0 \tag{A 2·3}$$

where $\mathbf{X} = \mathbf{X}_0\mathbf{H}_1$ and $\mathbf{Z} = \mathbf{X}_0\mathbf{H}_2$. Thus the problem reduces to that of the model (A 1·1).

When the rank of **A** is p, the matrix $\mathbf{H}_1$ can also be chosen as $\mathbf{G}^{-1}\mathbf{A}'$ where **G** is any positive definite matrix leading to the equations

$$E[\mathbf{X}_0\mathbf{G}^{-1}\mathbf{A}'(\mathbf{AG}^{-1}\mathbf{A}')^{-1}] = \mathbf{B}\xi, \quad E(\mathbf{X}_0\mathbf{H}_2) = 0, \tag{A 2·4}$$

where $\mathbf{H}_2$ is the same as in (A 2·2). The model (A 2·4) is of the same form as (A 2·1).

Let us suppose that the conditional expectation of the main variables given the concomitants $\mathbf{X}_0\mathbf{H}_2$ has the Gauss–Markhoff model involving the original parameters ξ and the regression coefficients on the concomitants as in (A 1·2). Then the analysis of Appendix A1 is applicable and the inference on ξ is the same whatever may be the transformations used such as **H** in (A 2·2) and **G**, $\mathbf{H}_2$ in (A 2·4).

In the method proposed by Potthoff & Roy the information supplied by the concomitants $\mathbf{X}_0\mathbf{H}_2$ is completely ignored and further, some arbitrariness is introduced in the choice of **G**. They recommend the choice of **G** based on *a priori* or some previous information. But such a procedure is not justified unless its applicability to given data is established or the information supplied by the concomitants is shown to be negligible (by an appropriate test) when a particular choice of **G** is made.

Under the general set up such as (A 2·1) for expectations and a general dispersion matrix for the row variables, only two extreme types of discussions are possible. One is to ignore the concomitants hoping that a suitable choice of **G** has been made as done by Potthoff & Roy. Another is to make adjustment for all the concomitant variables and thus avoid all arbitrariness, as done by Rao (1959). Any discussion of other possibilities leading to a selection of concomitants would not be easy under the general set up since the representation of the concomitants in (A 2·2) or (A 2·4) is arbitrary. In the example of polynomial growth curves discussed in the present paper, there is a natural transformation leading to main and concomitant variables in the form of polynomial regression coefficients for individual growth curves. In such a case the problem is to select a subset from the concomitants for covariance adjustment. This can be done as illustrated in the present paper, although certain refinements may be necessary to allow for the fact that the number and the set of the concomitants chosen may vary from sample to sample.

REFERENCES

DUNCAN, O. B. (1960). On the optimum estimation of a missile trajectory from several tracking systems. *R.C.A. systems analysis Technical report No.* 12, *AFMTC-TR* 16, Patrick Air Force Base, Florida.

ELSTON, R. C. & GRIZZLE, J. E. (1962). Estimation of time-response curves and their confidence bands. *Biometrics*, 18, 148–59.

ELSTON, R. C. (1964). On estimating time response curves. *Biometrics*, **20**, 643–7.

FRASER, D. A. S. (1956). Sufficient statistics with nuisance parameters. *Ann. Math. Statist.* **27**, 838–42.

HENDERSON, C. R., KEMPTHORNE, O., SEARLE, S. R. & VON KROSIGK, C. M. (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics*, 15, 192–218.

LEECH, F. B. & HEALY, M. J. R. (1959). The analysis of experiments on growth rate. *Biometrics*, **15**, 98–106.

POTTHOFF, RICHARD, F. & ROY, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, **51**, 313–26.

RAO, C. RADHAKRISHNA (1949). On some problems arising out of discrimination with multiple characters. *Sankhyā*, **9**, 343–66.

RAO, C. RADHAKRISHNA (1958). Some statistical methods for the comparison of growth curves. *Biometrics*, **14**, 1–17.

RAO, C. RADHAKRISHNA (1959). Some problems involving linear hypotheses in multivariate analysis. *Biometrika*, **46**, 49–58.

RAO, C. RADHAKRISHNA (1961). Some observations on multivariate statistical methods in anthropological research. *Bull. Int. Stat. Inst.* 38, (Part 4), 99–109.

WISHART, JOHN (1938). Growth rate determination in nutrition studies with the bacon pig, and their analysis. *Biometrika*, **30**, 16–28.