MIXED MODEL VARIANCE ANALYSIS WITH NORMAL ERROR AND POSSIBLY NON-NORMAL OTHER RANDOM EFFECTS: PART II: THE MULTIVARIATE CASE[1]

By S. N. Roy and Whitfield Cobb

University of North Carolina

3.0. Introduction and summary. The present paper is a continuation of another with similar title and the same overall objectives—confidence bounds on appropriate measures of the dispersion of the distribution from which random (block) effects are drawn in an experiment where fixed (treatment) effects are also under investigation. Specifically, confidence bounds are obtained on the maximum and minimum characteristic roots of the variance matrix of the block effects when the latter are assumed to come from a p-variate normal distribution (without the assumption made in [1], [5] that this variance matrix is proportional to that of the error). When the random block effects are not assumed to be normal, consideration is given to the approximation of an unknown multivariate distribution by means of marginal and conditional quantiles. Then for a rather restricted bivariate case, simultaneous confidence bounds are found for the two interquartile ranges.

Since the ideas and notation of the first paper are presupposed by this one, much duplication is avoided by reference to appropriate sections or steps in the previous article. To facilitate such reference, the numbering is consecutive through both parts.

3.1. The multiresponse model and its statistics. In the previous sections we have considered models in which the observed response was regarded as the sum of a normally distributed error and two or more effects due to treatments and blocks, but only one type of observation was to be made on each experimental unit. Now suppose that these several factors—whether called treatments or blocks and whether represented by fixed effects or random effects—are regarded as influencing more than one observable characteristic of the experimental units. This is the situation sometimes called a multivariate analysis of variance model, but perhaps a better expression would be "multiresponse analysis of variance model." We shall suppose that the response is determined, in a particular experiment, by observing p distinct (but presumably related) characteristics of each experimental unit. Such observations are conveniently arranged as elements of the matrix $\mathbf{Y}$ $(n \times p)$. The experiment might indicate that the observations on one or more of these p characteristics provided no additional information, in which case such characteristics could be dropped from the model. But regardless of how many or which characteristics are to be observed, the structure matrix,

Received August 10, 1959; revised May 25, 1960.

[1] This research was sponsored by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command.

as made specific by a chosen design and field plan, tells how the n experimental units are distributed among the cells of the cross classification, i.e., among the s blocks and t treatments. The p columns of $\mathbf{Y}$ are related to corresponding columns of postulated treatment-effects and block-effects matrices by the same structure matrix, since the different characteristics observed belong to the same experimental units. Thus analogous to (1.1.1), we have for a two-factor multiresponse model

$$\mathbf{Y}(n \times p) = \mathbf{M}(n \times (s + t)) \begin{bmatrix} \mathbf{A}(t \times p) \\ \mathbf{B}(s \times p) \end{bmatrix} + \mathbf{E}(n \times p), \tag{3.1.1}$$

where the structure matrix $\mathbf{M}$ is exactly the same as in Section 1.1. The elements of $\mathbf{E}$ are assumed to be normally distributed and of the nature of errors. As with the components of $\boldsymbol{\varepsilon}$ in Section 1.1, the rows of $\mathbf{E}$ are assumed to be uncorrelated, but we do allow and expect correlation between elements in the same row. We further postulate that every row of $\mathbf{E}$ have a common variance matrix, $\mathbf{\Sigma}(p)$. Then the ith column of $\mathbf{E}$ has as its variance matrix $\sigma_{ii}\mathbf{I}(n)$, where σ_{ii} is the ith element in the principal diagonal of $\mathbf{\Sigma}(p)$. As in the uniresponse model already considered, we regard the treatment effects $\mathbf{A}(t \times p)$ as fixed and the block effects $\mathbf{B}(s \times p)$ as random, but not necessarily normal, and independent of $\mathbf{E}$.

Because the structure matrix $\mathbf{M}$ of (3.1.1) is the same as the structure matrix $\mathbf{M}$ of (1.1.2), the motivation and the actual derivation presented in Section 1.2 are just as relevant here.

The same three orthonormal matrices $\mathbf{L}_0$, $\mathbf{L}_1$, and $\mathbf{L}$, (we recall that $\mathbf{L}_0\mathbf{L}_0' = \mathbf{I}(t-1)$, $\mathbf{L}_1\mathbf{L}_1' = \mathbf{I}(s-1)$, $\mathbf{L}\mathbf{L}' = \mathbf{I}(n-r^*)$, $\mathbf{L}_0\mathbf{L}' = 0$, $\mathbf{L}_1\mathbf{L}' = 0$, but $\mathbf{L}_0\mathbf{L}_1 \neq 0$ in general) are now used for defining three matrices of statistics in terms of the matrix of multiresponse observations:

$$\begin{aligned} \mathbf{U}((t-1) \times p) &\equiv \mathbf{L}_0((t-1) \times n)\mathbf{Y}(n \times p), \\ \mathbf{V}((s-1) \times p) &\equiv \mathbf{L}_1((s-1) \times n)\mathbf{Y}(n \times p), \\ \mathbf{W}((n-r^*) \times p) &\equiv \mathbf{L}((n-r^*) \times n)\mathbf{Y}(n \times p). \end{aligned} \tag{3.1.2}$$

The immediate consequence of these definitions is the desired relation of statistics to unobservables of the model:

$$\begin{aligned} \mathbf{U} &= \mathbf{T}_0^{-1}\mathbf{H}\mathbf{A} + \mathbf{L}_0\mathbf{E}, \\ \mathbf{V} &= \mathbf{T}_1^{-1}\mathbf{H}\mathbf{B} + \mathbf{L}_1\mathbf{E}, \\ \mathbf{W} &= \mathbf{L}\mathbf{E}. \end{aligned} \tag{3.1.3}$$

Moreover, the same interrelations among the vectors of (1.2.9) are preserved among the matrices of (3.1.3): $\mathbf{U}$ is independent of $\mathbf{B}$ though not of $\mathbf{V}$; $\mathbf{V}$ is independent of $\mathbf{A}$ though not of $\mathbf{U}$; $\mathbf{W}$ is independent of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{U}$, and $\mathbf{V}$. But (3.1.3) has another feature which may be worth pointing out: (1) each column of $\mathbf{U}$ depends upon only the corresponding column of the treatment effects and the (normal) errors associated with observing that characteristic of the experimental unit; (2) each column of $\mathbf{V}$ depends upon only the corresponding column of the block effects and the (normal) errors associated with observing that characteristic; (3) each column of $\mathbf{W}$ depends upon only the column of (normal) errors associated with observing that characteristic. Here also we require that the experimental design satisfy the linked block conditions which, in Section 1.4, were shown to be sufficient to reduce $\mathbf{T}_1$ to the scalar matrix $\delta^{-1}\mathbf{I}$.

3.2. Quasi-confidence bounds. In obtaining the quasi-confidence bounds for the uniresponse situation, we wanted $\begin{bmatrix} \mathbf{L}_1 \\ \mathbf{L} \end{bmatrix}$ orthonormal in order to preserve for the $s - 1 + n - r^*$ components of $\begin{bmatrix} \mathbf{L}_1\boldsymbol{\varepsilon} \\ \mathbf{w} \end{bmatrix}$ the same independent normal distribution postulated for $\boldsymbol{\varepsilon}$. Similarly here a consequence of the orthonormality of $\begin{bmatrix} \mathbf{L}_1 \\ \mathbf{L} \end{bmatrix}$ is that $\begin{bmatrix} \mathbf{L}_1\mathbf{E} \\ \mathbf{W} \end{bmatrix}$ consists of $s - 1 + n - r^*$ mutually independent rows, each row having the same p-variate normal distribution postulated for every row of $\mathbf{E}$. One way of demonstrating this is given in Appendix C. Thus it follows that $\mathcal{E}\{\mathbf{E}'\mathbf{L}_1'\mathbf{L}_1\mathbf{E}/(s-1)\} = \mathcal{E}\{\mathbf{W}'\mathbf{W}/(n-r^*)\} = \mathcal{E}\{\mathbf{E}'\mathbf{E}/n\}$. Whereas $\mathbf{w}'\mathbf{w}$ was a positive scalar and $\mathbf{w}'\mathbf{w}/\sigma^2$ had the central chi square distribution, $\mathbf{W}'\mathbf{W}$ is (a.e.) a positive definite symmetric matrix and $\mathbf{W}'\mathbf{W}/(n - r^*)$ has the central Wishart distribution with $n - r^*$ d.f., provided $n - r^* \geqq p$. If $s - 1 \geqq p$, similar statements can be made about $\mathbf{E}'\mathbf{L}_1'\mathbf{L}_1\mathbf{E}$. Moreover, when the above conditions are satisfied, the distribution of the characteristic roots of $\mathbf{E}'\mathbf{L}_1'\mathbf{L}_1\mathbf{E}(\mathbf{W}'\mathbf{W})^{-1}$ is known (cf. pp. 34–35 of [6]) and known to depend only upon the constants $s - 1$, $n - r^*$, and p. The distribution of the maximum of such characteristic roots is now tabulated (cf. [2], [3], [4]). Hence for a chosen $\alpha < 1$ it is possible to find a constant c_α such that

$$\text{(3.2.1)} \qquad \Pr\{\mathrm{ch}_{\max}[\mathbf{E}'\mathbf{L}_1'\mathbf{L}_1\mathbf{E}(\mathbf{W}'\mathbf{W})^{-1}] \leqq c_\alpha\} = 1 - \alpha.$$

This probability statement corresponds to (1.3.1) for the univariate case. It is true regardless of the computed values of $\mathbf{V}$ and regardless of the distribution from which $\mathbf{B}$ is a sample.

Using (3.1.2) to eliminate $\mathbf{L}_1\mathbf{E}$, the inequality within (3.2.1) becomes

$$\text{(3.2.2)} \qquad \mathrm{ch}_{\max}[(\mathbf{V} - \mathbf{T}_1^{-1}\mathbf{HB})'(\mathbf{V} - \mathbf{T}_1^{-1}\mathbf{HB})(\mathbf{W}'\mathbf{W})^{-1}] \leqq c_\alpha.$$

Since for positive definite $\mathbf{M}$ and at least positive semidefinite $\mathbf{N}$, $\mathrm{ch}_{\min}(\mathbf{M}) \cdot \mathrm{ch}_{\min}(\mathbf{N}) \leqq \mathrm{ch}(\mathbf{MN}) \leqq \mathrm{ch}_{\max}(\mathbf{M})\,\mathrm{ch}_{\max}(\mathbf{N})$ (cf. A.1.22 of [6]), (3.2.2) implies

$$\text{(3.2.3)} \qquad \mathrm{ch}_{\max}[(\mathbf{V} - \mathbf{T}_1^{-1}\mathbf{HB})'(\mathbf{V} - \mathbf{T}_1^{-1}\mathbf{HB})] \leqq c_\alpha\,\mathrm{ch}_{\max}(\mathbf{W}'\mathbf{W}).$$

Then by Lemma 1.2e of [1], (3.2.3) is equivalent to

$$\text{(3.2.4)} \qquad |\mathbf{d}'(\mathbf{V} - \mathbf{T}_1^{-1}\mathbf{HB})\mathbf{e}| \leqq [c_\alpha\,\mathrm{ch}_{\max}(\mathbf{W}'\mathbf{W})]^{\frac{1}{2}}$$

for all unit vectors **d** and **e**. Whence

(3.2.5) $$\mathbf{d}'\mathbf{Ve} - [c_\alpha \operatorname{ch}_{\max} (\mathbf{W}'\mathbf{W})]^{\frac{1}{2}} \leqq \mathbf{d}'\mathbf{T}_1^{-1}\mathbf{HBe} \leqq \mathbf{d}'\mathbf{Ve} + [c_\alpha \operatorname{ch}_{\max} (\mathbf{W}'\mathbf{W})]^{\frac{1}{2}}$$

for all unit vectors **d** and **e**. It is easily seen that $\mathbf{d}'\mathbf{Ve} \leqq \sup (\mathbf{d}'\mathbf{Ve})$ for all unit vectors **d** and **e** including that pair which maximizes $\mathbf{d}'\mathbf{T}^{-1}\mathbf{HBe}$. Similarly $\inf (\mathbf{d}'\mathbf{Ve}) \leqq \mathbf{d}'\mathbf{Ve}$ for all unit vectors **d** and **e** including that pair which minimizes $\mathbf{d}'\mathbf{T}_1^{-1}\mathbf{HBe}$. Applying these arguments to the right and left inequalities, respectively, of (3.2.5) yields

$$\inf (\mathbf{d}'\mathbf{Ve}) - [c_\alpha \operatorname{ch}_{\max}(\mathbf{W}'\mathbf{W})]^{\frac{1}{2}} \leqq \inf (\mathbf{d}'\mathbf{T}_1^{-1}\mathbf{HBe}) \leqq \sup (\mathbf{d}'\mathbf{T}_1^{-1}\mathbf{HBe}) \leqq \sup (\mathbf{d}'\mathbf{Ve}) + [c_\alpha \operatorname{ch}_{\max}(\mathbf{W}'\mathbf{W})]^{\frac{1}{2}},$$

which is equivalent to

(3.2.6) $$\begin{aligned} &[\operatorname{ch}_{\min}(\mathbf{V}'\mathbf{V})]^{\frac{1}{2}} - [c_\alpha \operatorname{ch}_{\max}(\mathbf{W}'\mathbf{W})]^{\frac{1}{2}} \leqq [\operatorname{ch}_{\min}\mathbf{B}'\mathbf{H}'(\mathbf{T}_1\mathbf{T}_1')^{-1}\mathbf{HB}]^{\frac{1}{2}} \\ &\leqq [\operatorname{ch}_{\max}\mathbf{B}'\mathbf{H}'(\mathbf{T}_1\mathbf{T}_1')^{-1}\mathbf{HB}]^{\frac{1}{2}} \leqq [\operatorname{ch}_{\max}(\mathbf{V}'\mathbf{V})]^{\frac{1}{2}} + [c_\alpha \operatorname{ch}_{\max}(\mathbf{W}'\mathbf{W})]^{\frac{1}{2}}. \end{aligned}$$

This inequality, being implied by but not implying the inequality within (3.2.1), would be true with probability not less than $1 - \alpha$. (The same remarks made at the end of Section 1.3 would apply here.) Thus (3.2.6) is the multiresponse analog of (1.3.3). For the large class of designs for which $\mathbf{T}_1\mathbf{T}_1' = \delta^{-2}\mathbf{I}$, the two central members of (3.2.6) may be simplified. And because of the non-negative character of these two, the extreme left member may be replaced by a non-negative upper bound. Thus if we define

(3.2.7) $$\begin{aligned} l_1 &\equiv [\delta^{-2} \operatorname{ch}_{\min}(\mathbf{V}'\mathbf{V})]^{\frac{1}{2}} - [\delta^{-2}c_\alpha \operatorname{ch}_{\max}(\mathbf{W}'\mathbf{W})]^{\frac{1}{2}}, \quad \text{when this is} > 0, \\ l_1 &\equiv 0 \text{ otherwise}, \\ l_2 &\equiv [\delta^{-2} \operatorname{ch}_{\max}(\mathbf{V}'\mathbf{V})]^{\frac{1}{2}} + [\delta^{-2}c_\alpha \operatorname{ch}_{\max}(\mathbf{W}'\mathbf{W})]^{\frac{1}{2}}, \end{aligned}$$

then we have

(3.2.8) $$\Pr\{l_1^2 \leqq \operatorname{ch}_{\min}(\mathbf{B}'\mathbf{H}'\mathbf{HB}) \leqq \operatorname{ch}_{\max}(\mathbf{B}'\mathbf{H}'\mathbf{HB}) \leqq l_2^2\} \geqq 1 - \alpha.$$

This (3.2.8) is in the form of a confidence statement. The bounds, l_1^2 and l_2^2, are computable from the observations, **Y**, and a chosen confidence coefficient, $1 - \alpha$. But the central terms are not explicit parameters or even parametric functions. Hence we call (3.2.8) a quasi-confidence statement. In so far as (3.2.8) is an intermediate step toward confidence bounds on certain parametric functions, it may also be called a preliminary confidence statement.

3.3. Multivariate normal random effects. It has already been stated that the rows of **B** are independently but identically distributed. Suppose now it is further specified that this common distribution be a p-variate normal with (unknown) variance matrix denoted by $\mathbf{\Sigma}_1(p)$. Then regardless of $\mathcal{E}(\mathbf{B})$, **HB** has zero expectation, and $\mathcal{E}(\mathbf{B}'\mathbf{H}'\mathbf{HB}) = (s - 1)\mathbf{\Sigma}_1$. Moreover the distribution of $\operatorname{ch}(\mathbf{B}'\mathbf{H}'\mathbf{BH}\mathbf{\Sigma}_1^{-1})$ is known to depend only on the constants $s - 1$ and p.

Thus for a chosen $\alpha_1 < 1$ we can find two constants c_1 and c_2 such that

$$(3.3.1)\quad \Pr\{c_1 \leqq \mathrm{ch}_{\min}(\mathbf{B}'\mathbf{H}'\mathbf{H}\mathbf{B}\boldsymbol{\Sigma}_1^{-1}) \leqq \mathrm{ch}_{\max}(\mathbf{B}'\mathbf{H}'\mathbf{H}\mathbf{B}\boldsymbol{\Sigma}_1^{-1}) \leqq c_2\} = 1 - \alpha_1 .$$

Since the non-zero roots, $\mathrm{ch}(\mathbf{M}_1\mathbf{M}_2^{-1})$, are the stationary values of $\mathbf{e}'\mathbf{M}_1\mathbf{e}/\mathbf{e}'\mathbf{M}_2\mathbf{e}$, (A.2.1 of [6]), $\mathrm{ch}_{\max}(\mathbf{B}'\mathbf{H}'\mathbf{H}\mathbf{B}\boldsymbol{\Sigma}_1^{-1}) \leqq c_2$ is equivalent to $\mathbf{e}'\mathbf{B}'\mathbf{H}'\mathbf{H}\mathbf{B}\mathbf{e}/c_2 \leqq \mathbf{e}'\boldsymbol{\Sigma}_1\mathbf{e}$ for all vectors $\mathbf{e}$. Moreover $\inf(\mathbf{e}'\mathbf{B}'\mathbf{H}'\mathbf{H}\mathbf{B}\mathbf{e}) \leqq \mathbf{e}'\mathbf{B}'\mathbf{H}'\mathbf{H}\mathbf{B}\mathbf{e}$ for all $\mathbf{e}$ including that choice which minimizes $\mathbf{e}'\boldsymbol{\Sigma}_1\mathbf{e}$. Thus $\mathrm{ch}_{\min}(\mathbf{B}'\mathbf{H}'\mathbf{H}\mathbf{B})/c_2 \leqq \mathrm{ch}_{\min}(\boldsymbol{\Sigma}_1)$, and similar argument leads to $\mathrm{ch}_{\max}(\boldsymbol{\Sigma}_1) \leqq \mathrm{ch}_{\max}(\mathbf{B}'\mathbf{H}'\mathbf{H}\mathbf{B})/c_1$. Thus the probability statement (3.3.1) leads to

$$(3.3.2)\quad \Pr\{\mathrm{ch}_{\min}(\mathbf{B}'\mathbf{H}'\mathbf{H}\mathbf{B})/c_2 \leqq \mathrm{ch}_{\min}(\boldsymbol{\Sigma}_1) \leqq \mathrm{ch}_{\max}(\boldsymbol{\Sigma}_1) \leqq \mathrm{ch}_{\max}(\mathbf{B}'\mathbf{H}'\mathbf{H}\mathbf{B})/c_1\} \geqq 1 - \alpha_1 ,$$

which is in the form of a confidence statement. What keeps (3.3.2) from being a *bona fide* confidence statement is that its bounds are not actually computable from observations. However, we do have bounds (quasi-confidence bounds) on these bounds. Combining (3.2.8) and (3.3.2) gives

$$(3.3.3)\quad \Pr\{l_1^2/c_2 \leqq \mathrm{ch}_{\min}(\boldsymbol{\Sigma}_1) \leqq \mathrm{ch}_{\max}(\boldsymbol{\Sigma}_1) \leqq l_2^2/c_1\} \geqq (1 - \alpha)(1 - \alpha_1).$$

Because of the postulated independence of $\mathbf{E}$ and $\mathbf{B}$, the respective confidence coefficients of (3.2.8) and (3.3.2) may be multiplied as in (3.3.3).

Confidence bounds on $\mathrm{ch}(\boldsymbol{\Sigma})$ and on σ_1^2 where it was assumed that $\boldsymbol{\Sigma}_1 = \sigma_1^2\boldsymbol{\Sigma}$ have been obtained previously [1], [5]. But (3.3.3) requires no such restrictive assumption. Of course the characteristic roots of a variance matrix are not in themselves easily interpreted parameters like standard deviations of the several variates, but they do constitute a measure of dispersion.

3.4. Marginal and conditional m-tiles. In Section 1 we proposed to replace the unobservable nonnormal (or not-necessarily-normal) block effects variate by a substitute variate taking k unknown values with equal probabilities. We then proceeded to obtain simultaneous confidence bounds on the differences between these successive unknown values and to interpret these $k - 1$ differences as estimates of the differences between the successive odd $2k$-tiles of the population of block effects. In the present section we propose a similar procedure for the multiresponse mixed model. In a practical approach to a multiresponse experiment, the p characteristics to be observed are apt to be selected, one at a time, in the order of their interest or presumed relevance to the factors being studied. Accordingly we adopt the convention that in all multiresponse models, the subscripts $1, 2, \cdots, p$ on column vectors from $\mathbf{Y}$, $\mathbf{A}$, $\mathbf{B}$, $\mathbf{E}$, $\mathbf{U}$, $\mathbf{V}$, or $\mathbf{W}$ will indicate respectively the most important, the next most important, $\cdots$, the least important characteristic. Consistent with the notation introduced in Section 1.5, in a multiresponse model ${}_m\beta_{n_1}$ will denote the n_1th m-tile of the marginal distribution of the first variate or first characteristic of the block-effects factor. The symbol ${}_m\beta_{n_1n_2}$ will denote the n_2th m-tile of the conditional distribution of the second variate or second characteristic given that the first variate lies below the

$(n_1 + 1)$th and not below the $(n_1 - 1)$th m-tile of its marginal distribution. Similarly ${}_m\beta_{n_1 n_2 n_3}$ will denote the n_3th m-tile of the conditional distribution of the third variate given that the first variate lies below the $(n_1 + 1)$th m-tile and not below the $(n_1 - 1)$th m-tile of its marginal distribution and given that the second variate lies below the $(n_2 + 1)$th m-tile and not below the $(n_2 - 1)$th m-tile of its conditional distribution, etc. As in Section 1.5 we consider only even values of m, say $2k$, and only odd values for n_i. Using b_i for any element in the ith column of $\mathbf{B}(s \times p)$, then no matter what the distribution of b_i

$$\Pr\{{}_{2k}\beta_{n_1-1} \leqq b_1 < {}_{2k}\beta_{n_1+1}\} = 1/k,$$

$$(3.4.1)\quad \begin{aligned} &\Pr\{{}_{2k}\beta_{n_1 n_2-1} \leqq b_2 < {}_{2k}\beta_{n_1 n_2+1} \mid {}_{2k}\beta_{n_1-1} \leqq b_1 < {}_{2k}\beta_{n_1+1}\} = 1/k, \\ &\Pr\{{}_{2k}\beta_{n_1 n_2 n_3-1} \leqq b_3 < {}_{2k}\beta_{n_1 n_2 n_3+1} \mid {}_{2k}\beta_{n_1-1} \leqq b_1 < {}_{2k}\beta_{n_1+1} \text{ and} \\ &\qquad {}_{2k}\beta_{n_1 n_2-1} \leqq b_2 < {}_{2k}\beta_{n_1 n_2+1}\} = 1/k; \qquad \text{etc.} \end{aligned}$$

Combining p such probability statement yields

$$(3.4.2)\quad \begin{aligned} &\Pr\{{}_{2k}\beta_{n_1-1} \leqq b_1 < {}_{2k}\beta_{n_1+1}\,;\, {}_{2k}\beta_{n_1 n_2-1} \leqq b_2 < {}_{2k}\beta_{n_1 n_2+1}\,;\, \cdots ; \\ &\qquad {}_{2k}\beta_{n_1\cdots n_p-1} \leqq b_p < {}_{2k}\beta_{n_1\cdots n_p+1}\} = 1/k^p. \end{aligned}$$

In Section 1.5 the presumably continuous but unknown distribution of the random block effects was approximated by the discrete distribution whose equally probable values were the odd $2k$-tiles of the unknown distribution. It was as if the data were classified into k classes with class boundaries at the even $2k$-tiles and with the odd $2k$-tiles used as class marks. This scheme may be extended to bivariate and even p-variate distributions. The data of a bivariate distribution may be classified into k^2 classes where the class boundaries are the even $2k$-tiles of the marginal distribution of the first variate and the conditional distributions of the second variate given the class of the first variate. E.g., there might be four classes defined as follows: (1) ${}_4\beta_0 \leqq b_1 < {}_4\beta_2$, ${}_4\beta_{10} \leqq b_2 < {}_4\beta_{12}$; (2) ${}_4\beta_0 \leqq b_1 < {}_4\beta_2$, ${}_4\beta_{12} \leqq b_2 < {}_4\beta_{14}$; (3) ${}_4\beta_2 \leqq b_1 < {}_4\beta_4$, ${}_4\beta_{30} \leqq b_2 < {}_4\beta_{32}$; (4) ${}_4\beta_2 \leqq b_1 < {}_4\beta_4$, ${}_4\beta_{32} \leqq b_2 < {}_4\beta_{34}$. As class marks for these classes we would take the following pairs: (1) $b_1 = {}_4\beta_1$, $b_2 = {}_4\beta_{11}$; (2) $b_1 = {}_4\beta_1$, $b_2 = {}_4\beta_{13}$; (3) $b_1 = {}_4\beta_3$, $b_2 = {}_4\beta_{31}$; (4) $b_1 = {}_4\beta_3$, $b_2 = {}_4\beta_{33}$. A bivariate which takes on just these four pairs of values we denote by $({}_4\beta_1, {}_4\beta_2)$, and when it is used in place of the presumably continuous bivariate (b_1, b_2), we call the former the "substitute variate". Similarly $({}_{2k}b_1, {}_{2k}b_2)$ will denote a substitute variate which takes on k^2 equally probable pairs of values, viz.,

$$(3.4.3)\qquad \Pr\{{}_{2k}b_1 = {}_{2k}\beta_{2m-1}\,;\, {}_{2k}b_2 = {}_{2k}\beta_{2m-1,2n-1}\} = 1/k^2$$

for $m = 1, 2, \cdots, k$ and $n = 1, 2, \cdots, k$. For a trivariate situation the distribution of the substitute variate would be defined by

$$(3.4.4)\quad \Pr\{{}_{2k}b_1 = {}_{2k}\beta_{2m-1}\,;\, {}_{2k}b_2 = {}_{2k}\beta_{2m-1,2n-1}\,;\, {}_{2k}b_3 = {}_{2k}\beta_{2m-1,2n-1,2q-1}\} = 1/k^3,$$

for $m, n, q = 1, 2, \cdots, k$. The extension to a larger number of characteristics is now obvious and need not be written.

These marginal and conditional m-tiles provide a means of distinguishing different kinds of interrelatedness or dependence in a multivariate distribution. For simplicity of discussion, we shall consider $p = 2$ and $k = 2$. Then it is possible to classify bivariate distributions into four types, the first being the most general and including all distributions not qualifying for the other types. We call type 2 those distributions in which conditional distributions of the second variate, given different values of the first variate, all have the same dispersion as measured by the interquartile difference; but the variates are "dependent" in the sense (i) that medians of the conditional distributions of the second variate are different for different values of the first variate. In type 3, on the other hand, conditional distributions of the second variate, given different values of the first variate, all have the same median; but the variates are "dependent" in the sense (ii) that conditional distributions of the second variate have different dispersions as measured by the interquartile difference. In type 4 the variates are not "dependent" in either sense (i) or sense (ii). A bivariate normal of type 4 would consist of two independent normal distributions.

3.5. Confidence bounds in a simple bivariate case. In Section 3.2 we obtained preliminary or quasi-confidence bounds on $\text{ch}(\mathbf{B}'\mathbf{H}'\mathbf{H}\mathbf{B})$ regardless of the distribution of $\mathbf{B}$, and then in Section 3.3 we used these quasi-confidence bounds as a preliminary stage in finding genuine confidence bounds on $\text{ch}(\mathbf{\Sigma}_1)$ when $\mathbf{B}$ was assumed to be normal, each row of $\mathbf{B}$ having $\mathbf{\Sigma}_1$ as its variance matrix. But if $\mathbf{B}$ is not normal, or not assumed to be normal prior to the experiment, is it possible to use the quasi-confidence bounds to obtain a confidence statement about the marginal and conditional m-tiles described in Section 3.4? The answer is yes. At least for the simplest multiresponse model—$p = 2$, $k = 2$, type 4—perhaps also for others, it is feasible as well as possible.

For $p = 2$, the quadratic formula may be used to find explicitly

$$\text{ch}_{\min}(\mathbf{B}'\mathbf{H}'\mathbf{H}\mathbf{B}) \equiv \lambda_1 \quad \text{and} \quad \text{ch}_{\max}(\mathbf{B}'\mathbf{H}'\mathbf{H}\mathbf{B}) \equiv \lambda_2 :$$

$$\begin{aligned}\lambda_i = {} & \tfrac{1}{2}(\mathbf{b}_1'\mathbf{H}'\mathbf{H}\mathbf{b}_1 + \mathbf{b}_2'\mathbf{H}'\mathbf{H}\mathbf{b}_2) \\ & \pm \tfrac{1}{2}[(\mathbf{b}_1'\mathbf{H}'\mathbf{H}\mathbf{b}_1 - \mathbf{b}_2'\mathbf{H}'\mathbf{H}\mathbf{b}_2)^2 + 4(\mathbf{b}_1'\mathbf{H}'\mathbf{H}\mathbf{b}_2)^2]^{\frac{1}{2}}.\end{aligned} \tag{3.5.1}$$

Next we set the two λ_i of (3.5.1) equal respectively to l_i^2 of (3.2.7) to obtain the two extreme conditions permitted by the quasi-confidence statement (3.2.8). Algebraic simplification results in

$$\begin{aligned}(\mathbf{b}_1'\mathbf{H}'\mathbf{H}\mathbf{b}_2)^2 - {} & (\mathbf{b}_1'\mathbf{H}'\mathbf{H}\mathbf{b}_1)(\mathbf{b}_2'\mathbf{H}'\mathbf{H}\mathbf{b}_2) \\ & + l_i^2(\mathbf{b}_1'\mathbf{H}'\mathbf{H}\mathbf{b}_1 + \mathbf{b}_2'\mathbf{H}'\mathbf{H}\mathbf{b}_2) - l_i^4 = 0.\end{aligned} \tag{3.5.2}$$

For type 4 there are no regression-like parameters to be found—only the separate measures of dispersion. For $k = 2$ there are merely ${}_4\beta_3 - {}_4\beta_1$ and ${}_4\beta_{13} - {}_4\beta_{11}$ since ${}_4\beta_{31} = {}_4\beta_{11}$ and ${}_4\beta_{33} = {}_4\beta_{13}$. Now replacing the unknown variate (b_1, b_2)

by the substitute variate $({}_4b_1, {}_4b_2)$ means that the s rows of $\mathbf{B}$ are replaced by so many, say s_{11}, rows equal to $({}_4\beta_1, {}_4\beta_{11})$, so many, say s_{13}, rows equal to $({}_4\beta_1, {}_4\beta_{13})$, etc., where $s_{11} + s_{13} + s_{31} + s_{33} = s$. When these replacements are made in the quadratic and bilinear forms occurring in (3.5.2), Lemma A.10 may be used, and (3.5.2) becomes

$$
\begin{aligned}
&s^{-2}[(s_{11} + s_{13})(s_{31} + s_{33})(s_{11} + s_{31})(s_{13} + s_{33}) - (s_{11}s_{33} - s_{13}s_{31})^2] \\
&\times ({}_4\beta_3 - {}_4\beta_1)^2({}_4\beta_{13} - {}_4\beta_{11})^2 - l_i^2 s^{-1}[(s_{11} + s_{13})(s_{31} + s_{33})({}_4\beta_3 - {}_4\beta_1)^2 \\
&\qquad + (s_{11} + s_{31})(s_{13} + s_{33})({}_4\beta_{13} - {}_4\beta_{11})^2] + l_i^4 = 0.
\end{aligned} \tag{3.5.3}
$$

For a given partition of s and for a given l_i, (3.5.3) may be regarded as determining a locus in the plane of $({}_4\beta_3 - {}_4\beta_1)^2$ and $({}_4\beta_{13} - {}_4\beta_{11})^2$. Note that the coordinates to be used are the squares of these interquartile differences. All possible partitions of s thus determine a finite family of conics for each value of l_i. As in Section 1.6 some of these loci are called bounding because they (together with the two coordinate axes) would inclose a region of the first quadrant. If s is so partitioned that any three components are zero, (3.5.3) becomes a contradiction. If s_{11} and s_{13} or s_{31} and s_{33}, are zero and no other component is, the locus of (3.5.3) is a line parallel to one coordinate axis. These are the cases in which (3.5.3) does not correspond to bounding loci. But if s_{11} and s_{33} or s_{13} and s_{31} are zero and no other component is, the locus of (3.5.3) is a line through the first quadrant making equal intercepts. If $s_{11}s_{33} = s_{13}s_{31} \neq 0$, the locus of (3.5.3) is two lines parallel to the coordinate axes and intersecting in the first quadrant. For all other possible partitions of s, the locus of (3.5.3) is a rectangular hyperbola whose asymptotes are parallel to the coordinate axes and whose center is in the first quadrant. The inner boundary of this set of bounding loci is itself a locus of (3.5.3), viz., the straight line whose equation is

$$
\begin{aligned}
&({}_4\beta_3 - {}_4\beta_1)^2 + ({}_4\beta_{13} - {}_4\beta_{11})^2 = 4sl_1^2/(s^2 - 1) \qquad \text{or} \\
&({}_4\beta_3 - {}_4\beta_1)^2 + ({}_4\beta_{13} - {}_4\beta_{11})^2 = 4l_1^2/s,
\end{aligned} \tag{3.5.4}
$$

depending upon whether s is odd or even. But there is no unique outer boundary among the bounding loci of (3.5.3). By comparing the various types of bounding loci enumerated above, it is easy to pick out the four segments of three members of the family (3.5.3) which constitute the outer boundary. Thus the confidence region is given by

$$
\left\{
\begin{aligned}
&\text{(i)}\ ({}_4\beta_3 - {}_4\beta_1)^2 + ({}_4\beta_{13} - {}_4\beta_{11})^2 \leqq sl_2^2/(s - 1) \\
&\qquad \text{for}\ \ 0 \leqq ({}_4\beta_3 - {}_4\beta_1)^2 \leqq s(s - 3)l_2/2(s - 1)(s - 2) \\
&\qquad \text{and for}\ \ sl_2^2/2(s - 2) \leqq ({}_4\beta_3 - {}_4\beta_1)^2 \leqq sl_2^2/(s - 1), \\
&\text{(ii)}\ 0 \leqq ({}_4\beta_{13} - {}_4\beta_{11})^2 \leqq sl_2^2/2(s - 2) \quad \text{for} \\
&\qquad\quad s(s - 3)l_2/2(s - 1)(s - 2) \leqq ({}_4\beta_3 - {}_4\beta_1)^2 \leqq sl_2^2/2(s - 2); \\
&\text{(iii)}\ ({}_4\beta_3 - {}_4\beta_1)^2 + ({}_4\beta_{13} - {}_4\beta_{11})^2 \geqq 4sl_1^2/(s^2 - 1) \quad \text{for } s \text{ odd}, \\
&\text{or} \\
&\text{(iv)}\ ({}_4\beta_3 - {}_4\beta_1)^2 + ({}_4\beta_{13} - {}_4\beta_{11})^2 \geqq 4l_1^2/s \quad \text{for } s \text{ even},
\end{aligned}
\right. \tag{3.5.5}
$$

where (i) and (ii) exhibit the outer boundary, (iii) or (iv), the inner boundary.

Associated with each partition of s is the *a priori* probability

$$\frac{s!}{2^{2s}s_{11}!s_{13}!s_{31}!s_{33}!}. \tag{3.5.6}$$

In the present case the total *a priori* probability associated with all bounding loci is easily found by subtracting from unity the probability of the non-bounding loci:

$$\gamma = 1 - \frac{4}{4^s} - \frac{2^s - 2}{4^s} - \frac{2^s - 2}{4^s} = 1 - 2^{1-s}. \tag{3.5.7}$$

It is interesting to note that here for $p = 2$ and $k = 2$ (where there are 4 equiprobable values) γ is exactly the same as in the uniresponse case with $k = 2$ (where there were 2 equiprobable values).

Now since **B** and **E** are independent by hypothesis, we may multiply the $1 - \alpha$ of the quasi-confidence statement (3.2.8) by the γ of (3.5.7) to obtain a lower bound on the final confidence coefficient. The final confidence statement says that with probability $\geqq (1 - \alpha)\gamma$, the interquartile differences ${}_4\beta_3 - {}_4\beta_1$ and ${}_4\beta_{13} - {}_4\beta_{11}$ are positive square roots of the coordinates of some point lying in the first quadrant region defined by (3.5.5) and the $({}_4\beta_3 - {}_4\beta_1)^2 = 0$ and $({}_4\beta_{13} - {}_4\beta_{11})^2 = 0$ axes. Of course here, as in the uniresponse model, there is an element of approximation due to replacing the unknown variate by the k^2-valued substitute variate. But presumably here too the degree of approximation can be improved by increasing k.

APPENDIX C

Variances and Covariances for a Matrix of Normal Variates

It is a well established custom to exhibit the n variances and $\binom{n}{2}$ covariances of a set of n normal variates as elements of an $n \times n$ matrix, displaying each covariance twice. Thus we say a stochastic vector has a variance-covariance matrix. If the elements of an $m \times n$ matrix were first written as components of a single vector, then the mn variances and $\binom{mn}{2}$ covariances of those elements could be displayed as elements of an $mn \times mn$ symmetric matrix. But it would certainly be desirable to arrange these variances and covariances in such a way that properties of the rows (or columns) of the original matrix are readily apparent from this larger matrix. To facilitate this systematization we define *ad hoc* a special vector and list its properties.

(C.1) $\mathbf{h}'(4) \equiv (1, 0, 0, 1)$, $\mathbf{h}'(9) \equiv (1, 0, 0, 0, 1, 0, 0, 0, 1)$, and for any positive integer m, $\mathbf{h}(m^2)$ will denote a column vector with m^2 components, m of which (including the first and last components) are unity with m zeros between successive unities.

(C.2) $\mathbf{h}'(m^2)\mathbf{h}(m^2) = \mathbf{h}'(m^2)\mathbf{j}(m^2) = m$.

(C.3) $\mathbf{h}(m)\mathbf{h}'(m) = \mathbf{h}(m) \cdot \times \mathbf{h}'(m^2)$, where $\cdot \times$ indicates the left direct product as in Section 0.2.

(C.4) $[\mathbf{A}(m \times n) \cdot \times \mathbf{I}(n)]\mathbf{h}(n^2) = \mathbf{a}(mn)$, where the m elements in the jth column of $\mathbf{A}$ become respectively the $(jm - m + 1)$th through jmth components of $\mathbf{a}$ for $j = 1, 2, \cdots, n$.

(C.5) $[\mathbf{I}(m) \cdot \times \mathbf{h}'(n^2)][\mathbf{a}(mn) \cdot \times \mathbf{I}(n)] = \mathbf{A}(m \times n)$, where the $(jm - m + 1)$th through jmth components of $\mathbf{a}$ become respectively the elements of the jth column of $\mathbf{A}$ for $j = 1, 2, \cdots, n$.

Now starting with a matrix $\mathbf{A}(m \times n)$ of normal variates we collapse it into the vector $\mathbf{a}(mn)$, defined as in (C.4), which separates the elements in the same row of $\mathbf{A}$ but keeps consecutive the elements in the same column. Hence the covariance matrix $\mathbf{\Sigma}(mn)$ will have the following pattern. The variances of the consecutive elements in the jth column of $\mathbf{A}$ will be consecutive elements along the principal diagonal of $\mathbf{\Sigma}$ from the $(jm - m + 1)$th through the jmth row for $j = 1, 2, \cdots, n$. The $\binom{m}{2}$ covariances of elements in the jth column of $\mathbf{A}$ will appear (twice) as the nondiagonal elements of the $m \times m$ principal submatrix in the $(jm - m + 1)$th through the jmth rows of $\mathbf{\Sigma}$. The $\binom{n}{2}$ covariances of the ith and jth elements within the kth row will appear (twice) as the kth diagonal element in the $m \times m$ submatrix lying in the $(im - m + 1)$th through imth rows (columns) and $(jm - m + 1)$th through jmth columns (rows). The $mn(m - 1)(n - 1)/2$ covariances of elements not in either the same row or same column of $\mathbf{A}$ will appear (twice) as the nondiagonal elements of these nonprincipal submatrices.

(C.6) If and only if each column of $\mathbf{A}$ has the identical variance-covariance matrix $\mathbf{\Sigma}_c(m)$ and the n columns are independent, then $\mathbf{\Sigma}(mn) = \mathbf{\Sigma}_c(m) \cdot \times \mathbf{I}(n)$.

(C.7) If and only if each row of $\mathbf{A}$ has the identical variance-covariance matrix $\mathbf{\Sigma}_r(n)$ and the m rows are independent, then $\mathbf{\Sigma}(mn) = \mathbf{I}(m) \cdot \times \mathbf{\Sigma}_r(n)$.

Now suppose we make a transformation of the original matrix $\mathbf{A}$, obtaining $\mathbf{B}(q \times n) \equiv \mathbf{C}(q \times m)\mathbf{A}$, and want to know the variances and covariances of the elements of $\mathbf{B}$. Applying (C.4) to $\mathbf{B}$ gives $\mathbf{b}(qn) = [\mathbf{CA} \cdot \times \mathbf{I}(n)]\mathbf{h}(n^2)$ which, by a property of Kronecker products, can be written $\mathbf{b} = [\mathbf{C} \cdot \times \mathbf{I}(n)] \cdot [\mathbf{A} \cdot \times \mathbf{I}(n)]\mathbf{h}(n^2)$, which is easily recognized as $\mathbf{b} = [\mathbf{C} \cdot \times \mathbf{I}(n)]\mathbf{a}$. Thus premultiplying the matrix $\mathbf{A}(m \times n)$ by the matrix $\mathbf{C}(q \times m)$ corresponds to premultiplying the vector $\mathbf{a}(mn)$ by the matrix $[\mathbf{C} \cdot \times \mathbf{I}(n)]$. Thus the variance-covariance matrix of $\mathbf{B}$ can be found as easily as that of a vector.

(C.8) If $\mathbf{\Sigma}(mn)$ is the variance-covariance matrix of the matrix $\mathbf{A}(m \times n)$, then the variance-covariance matrix of $\mathbf{C}(q \times m)\mathbf{A}$ is $[\mathbf{C} \cdot \times \mathbf{I}(n)]\mathbf{\Sigma}[\mathbf{C}' \cdot \times \mathbf{I}(n)]$.

If the conditions of (C.6) are satisfied, the above matrix reduces to $\mathbf{C\Sigma}_c\mathbf{C}' \cdot \times \mathbf{I}(n)$. If the conditions of (C.7) are satisfied, the same matrix becomes $\mathbf{CC}' \cdot \times \mathbf{\Sigma}_r$. Comparing this latter form with (C.7) we obtain the following conclusion:

(C.9) An orthonormal (or orthogonal) transformation applied to a matrix whose rows are uncorrelated yields a new matrix whose rows are uncorrelated, and if each row of the original matrix has a common variance-covariance matrix, it will be the variance-covariance matrix of every row of the new matrix.

REFERENCES

A more complete list of references may be found at the end of Part I. Only those items specifically referred to in Part II are listed here.

[1] Gnanadesikan, R., "Contributions to Multivariate Analysis Including Univariate and Multivariate Variance Components Analysis and Factor Analysis," Institute of Statistics, University of North Carolina, Mimeo Series No. 158 (1956).

[2] Heck, D. L., "Some Uses of the Distribution of the Largest Root in Multivariate Analysis," Institute of Statistics, University of North Carolina, Mimeo Series No. 194 (1958).

[3] Pillai, K. C. S., "On the distribution of the largest or the smallest root of a matrix in multivariate analysis," *Biometrika*, Vol. 43 (1956), pp. 122–127.

[4] Pillai, K. C. S., *Concise Tables for Statisticians*, The Statistical Center, University of the Philippines, Manila, 1957.

[5] Roy, S. N. and Gnanadesikan, R., "Some contributions to ANOVA in one or more dimensions: II," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 318–340.

[6] Roy, S. N., *Some Aspects of Multivariate Analysis*, John Wiley and Sons, New York, 1957.