April 1965

ON THE SIMULTANEOUS STABILIZATION OF VARIANCES
AND COVARIANCES*

Harold Ruben

Technical Report No. 56

University of Minnesota
Minneapolis, Minnesota

*Research supported under a grant from the United States Air Force
Office of Scientific Research - Office of Aerospace Research.

ON THE SIMULTANEOUS STABILIZATION OF VARIANCES
AND COVARIANCES

BY HAROLD RUBEN
University of Minnesota

0. Summary. Analogously to the variance-stabilizing transformations used in analysis of variance problems when heterogeneity of variance is present, it is shown that in at least one instance in analysis of dispersion (MANOVA) problems the observations can be transformed so as to eliminate substantially the effect of variation in the complex of variances and covariances between different groups or treatments. The situation referred to is one in which the correlation matrix is fully specified and is believed to be relatively homogeneous. The more general question whether the same purpose can be achieved in other analysis of dispersion problems is here left open.

The desired result is obtained with the aid of a convergence lemma which is of some intrinsic interest and of rather wide applicability in large-sample theory.

1. Preliminary discussion. Distinctly meager as is our present knowledge of the "robustness", in the sense of Box [2] (effect of departures from the underlying assumptions), of those standard statistical analyses which are collectively labelled "analysis of variance", it can only be described as positively rich relative to our knowledge of the robustness of MANOVA¹. [For a discussion on robustness vis-a-vis univariate data, from the point of view of the nominal

¹ For purposes of discussion, it will be assumed that the methodological appropriateness of MANOVA is not here at issue. I am nevertheless bound to record my view (based on some experience) that the latter can lead to serious interpretive difficulties, and that in exploratory studies it may be more profitable on occasion to sacrifice efficiency for simplicity by analyzing the variation in each characteristic separately, using suitably stringent confidence (or fiducial) and significance levels (etc.) for extra protection.

significance level and Eisenhart's Model I (fixed effects), see Box [3] [4] and Scheffé [9], Ch. 10, and for a discussion of the effect of heterogeneity of variance on power in a one-way layout, see Horsnell [6] .] In point of fact, so far as the published evidence goes, no investigation, theoretical or numerical, seems to have been made to date on the robustness of MANOVA (such an investigation would obviously be a formidable task), although it would perhaps not be unreasonable to believe that the effects of departures from the assumptions, and, more especially, the effect of heterogeneity of covariance matrices, might be more serious for multivariate than for univariate data. Be that as it may, a means of transforming the data so as to effectively stabilize the covariance matrices would be of some theoretical, and possibly even of some practical, interest. Such a transformation would constitute a generalization of the so-called variance-stabilizing transformation (e.g. Scheffé [9], p. 365), which generalization we shall call a dispersion-stabilizing transformation.

It seems clear, however, that the aim mentioned in the previous paragraph cannot be met in general, since the degree of latitude supplied by a single coordinate transformation, consisting of p component transformations is not adequate to control the complex of $p(p + 1)/2$ variances and covariances in a covariance matrix. (Here p stands for the dimensionality of a fundamental observation.) On the other hand, the aim does become a realistic one if the covariance matrices are sufficiently structured, by functional dependence between the elements of a covariance matrix, so as to reduce the effective number of elements in each matrix (i.e., the number of independent elements) to $p' \leq p$. In this article, the special case will be considered where $p' = p$, and the correlation matrix is fully specified and assumed to be constant (at least approximately). Such a situation might occur, for instance, in time series, or more generally stochastic process, analyses with repeated sets of data where a precise mathematical model based on rather extensive experience is to hand and one wishes to compare the process mean functions. Actually, it will appear that

in such circumstances one may go a step further, in that the transformed characteristics can be rendered effectively independent, in a statistical sense.

2. Some convergence results. We begin with some preliminary lemmas which will be used in the sequel to prove the required theorem.

LEMMA 1. If $\underline{x}_1, \underline{x}_2, \dots$ is a sequence of vector-valued random variables such that² for every real constant vector $\underline{\alpha}$

$$\mathcal{L}[\underline{\alpha}' \underline{x}_n] \rightarrow N(\underline{0}, \underline{\alpha}' \underline{V} \underline{\alpha}),$$

where $\underline{V}$ is a constant (non-random) matrix independent of n , then

$$\mathcal{L}[\underline{x}_n] \rightarrow N(\underline{0}, \underline{V}).$$

PROOF. By hypothesis, for every real t and every real vector $\underline{\alpha}$,

$$E e^{it(\underline{\alpha}' \underline{x}_n)} \rightarrow e^{-\frac{1}{2}(\underline{\alpha}' \underline{V} \underline{\alpha})t^2}$$

whence, on replacing $t \underline{\alpha}$ by $\underline{u}$,

$$E e^{i \underline{u}' \underline{x}_n} \rightarrow e^{-\frac{1}{2} \underline{u}' \underline{V} \underline{u}}$$

for every real vector $\underline{u}$. The lemma now follows from the continuity theorem for characteristic functions.

LEMMA 2 (Hoeffding and Robbins). Let $\underline{z}_1, \underline{z}_2, \dots$ be a sequence of p -component random variables, $d_1, d_2, \dots$ a sequence of real positive constants, and g a real-valued function defined on E_p . Assume that

$$(a) \mathcal{L}[\underline{z}_n] \rightarrow \mathcal{L}[\underline{z}],$$

$$(b) d_n \rightarrow 0,$$

$$(c) g \text{ has a total differential at the origin.}$$

² All limits, here and subsequently, refer to $n \rightarrow \infty$. Further, as usual,

$\mathcal{L}[\underline{y}_n] \rightarrow \mathcal{L}[\underline{y}]$ means that $\underline{y}_n$ converges in distribution to $\underline{y}$ as $n \rightarrow \infty$.

Then

$$(2.1) \quad \mathcal{L} \left[d_n^{-1} \{ g(d_n z_n) - g(\underline{0}) \} \right] \rightarrow \mathcal{L} \left[\left(\frac{\partial g(\underline{z})}{\partial \underline{z}} \right)'_0 z \right],$$

where

$$\left(\frac{\partial g(\underline{z})}{\partial \underline{z}} \right)'_0 = \left(\frac{\partial g(\underline{z})}{\partial z_1} \right)'_{\underline{z}=\underline{0}} \cdots \left(\frac{\partial g(\underline{z})}{\partial z_p} \right)'_{\underline{z}=\underline{0}}.$$

For a proof of Lemma 2, refer to Hoeffding and Robbins [5]. The proof³ is elementary and consists essentially in showing that the error induced by replacing $d_n^{-1} \{ g(d_n z_n) - g(\underline{0}) \}$ by the linear approximant $\left(\frac{\partial g(\underline{z})}{\partial \underline{z}} \right)'_0 z$ converges in probability to zero as $n \rightarrow \infty$.

LEMMA 3. Let $t_1, t_2, \dots$ be a sequence of p-component random variables, $\underline{\theta}$ a constant (non-random) p-component vector, $\underline{\Sigma}$ a constant symmetric non-negative matrix of size $p \times p$ which is independent of n , and $\underline{f} = (f_1, \dots, f_q)'$ a vector-valued function defined on E_p with range in E_q . Assume that

$$(i) \quad \mathcal{L} [n^{\frac{1}{2}}(t_n - \underline{\theta})] \rightarrow N(\underline{0}, \underline{\Sigma}),$$

$$(ii) \quad \text{each component of } \underline{f} \text{ has a total differential at } \underline{\theta}.$$

Then

$$(2.2) \quad \mathcal{L} \left[n^{\frac{1}{2}} \{ \underline{f}(t_n) - \underline{f}(\underline{\theta}) \} \right] \rightarrow N \left(\underline{0}, \frac{\partial \underline{f}(\underline{\theta})}{\partial \underline{\theta}} \underline{\Sigma} \left(\frac{\partial \underline{f}(\underline{\theta})}{\partial \underline{\theta}} \right)' \right),$$

where $\partial \underline{f}(\underline{\theta}) / \partial \underline{\theta} = (\partial f_i(\underline{\theta}) / \partial \theta_j)$, $i = 1, \dots, q$, $j = 1, \dots, p$.

PROOF. For an arbitrary q-component constant vector $\underline{\alpha}$, define

$$g(\underline{z}) = \underline{\alpha}' \underline{f}(\underline{z} + \underline{\theta}),$$

$$\underline{z}_n = n^{\frac{1}{2}}(t_n - \underline{\theta}), \quad n = 1, 2, \dots$$

We verify directly that conditions (a), (b), (c) of Lemma (2) hold, where here

³ Actually, in the Hoeffding-Robbins lemma $p = 2$, but the proof remains unaffected for $p \neq 2$.

$d_n = n^{-\frac{1}{2}}$ and $\underline{z}$ denotes a random vector distributed as $N(0, \underline{I})$. Lemma (2) gives

$$\mathcal{L} \left[n^{\frac{1}{2}} \{ \underline{\alpha}' \underline{f}(\underline{t}_n) - \underline{\alpha}' \underline{f}(\underline{\theta}) \} \right] \rightarrow \mathcal{L} \left[\underline{\alpha}' \frac{\partial \underline{f}(\underline{\theta})}{\partial \underline{\theta}} \cdot \underline{z} \right],$$

or, since $(\underline{\alpha}' \frac{\partial \underline{f}(\underline{\theta})}{\partial \underline{\theta}}) \underline{z}$ is distributed as

$N(0, \underline{\alpha}' (\frac{\partial \underline{f}(\underline{\theta})}{\partial \underline{\theta}}) \underline{I} (\frac{\partial \underline{f}(\underline{\theta})}{\partial \underline{\theta}})' \underline{\alpha})$, we have

$$\mathcal{L} \left[\underline{\alpha}' \cdot n^{\frac{1}{2}} \{ \underline{f}(\underline{t}_n) - \underline{f}(\underline{\theta}) \} \right] \rightarrow N \left(0, \underline{\alpha}' \cdot \frac{\partial \underline{f}(\underline{\theta})}{\partial \underline{\theta}} \underline{I} \left(\frac{\partial \underline{f}(\underline{\theta})}{\partial \underline{\theta}} \right)' \cdot \underline{\alpha} \right).$$

The theorem now follows from Lemma 1 with $\underline{x}_n = n^{\frac{1}{2}} \{ \underline{f}(\underline{t}_n) - \underline{f}(\underline{\theta}) \}$ and

$$\underline{V} = (\frac{\partial \underline{f}(\underline{\theta})}{\partial \underline{\theta}}) \underline{I} (\frac{\partial \underline{f}(\underline{\theta})}{\partial \underline{\theta}})'.$$

In concluding this section, a few remarks may be in order.

(1) Lemma 1 is surely not new, and is included here for convenience. It is easily shown that the converse of the lemma holds, so that a characterizing condition for the convergence in distribution of $\underline{x}_n$ to a multinormal distribution is the convergence in distribution of every arbitrary real linear combination of the components of $\underline{x}_n$ to a (univariate) normal distribution. In other words, Lemma 1, taken in conjunction with its converse is a limit variant of the well-known characterizing projection property of the multinormal distribution, namely, $\underline{x}$ has a multinormal distribution if and only if $\underline{\alpha}' \underline{x}$ has a (univariate) normal distribution for every real $\underline{\alpha}$ (Anderson [1], p. 37).

(2) Observe in Lemma 3 that q can be less than, equal to, or greater than p , and that no restriction is placed on the rank of $\underline{I}$ nor on the functional independence of the components of $\underline{f}$. In particular, then, either or both of the asymptotic normal distributions of $n^{\frac{1}{2}}(\underline{t}_n - \underline{\theta})$ and $n^{\frac{1}{2}}\{\underline{f}(\underline{t}_n) - \underline{f}(\underline{\theta})\}$ can be degenerate.

(3) The special case of Lemma 3 where $\underline{f}$ is real-valued ($q = 1$) has been widely quoted [e.g., Rao [8], p. 207, Anderson [1], pp. 66-67]. The case where both $\underline{f}$ and $\underline{t}_n$ are real-valued ($q = 1, p = 1$) [e.g., Lehmann [7], p. 274] is, of course, yet more specialized and provides the basis for the variance-stabilizing

transformation. (See comment (5) below.)

These specialized versions of Lemma 3 appear, however, to have been given under unnecessarily restrictive conditions, such as convergence of $\underline{t}_n$ to $\underline{\theta}$ in probability and the existence and continuity of the first or second partial derivatives of f (either locally at $\underline{\theta}$ or globally). We note in particular from condition (ii) of the lemma that neither global continuity of f nor continuity of the first partials (locally or globally) are necessary.

(4) Clearly, Lemma 3 has numerous applications in large-sample theory. A typical such application is that the joint distribution of the $k(k-1)/2$ correlations in random samples from an arbitrary infinite k -variate population with finite fourth-order product moments is asymptotically normal. Here $\underline{t}_n$ denotes the vector of first and second-order sample moments about the origin based on samples of size n , $p = k + k(k+1)/2$, $q = k(k-1)/2$ and the $f_i(\underline{t}_n)$, $i = 1, \dots, k(k-1)/2$ are the sample correlation coefficients. [The asymptotic normality of $\underline{t}_n$ is a consequence of the multivariate Central Limit theorem.]

Another domain of application concerns situations in which one is interested in the asymptotic distribution of a function of $\underline{t}_n$ which is not given in explicit form, as in Lemma 3, but rather implicitly, as in a set of estimation equations $\underline{g}(\underline{t}_n, \underline{y}) = \underline{0}$ (e.g. likelihood equations), $\underline{t}_n$ representing functions defined on the sample space and $\underline{y}$ a vector estimator of specified parametric functions. Subject to suitable regularity conditions on $\underline{g}$, Lemma 3 can then still be applied with $\underline{f}(\underline{t}_n)$ replaced by $\underline{y}$.

(5) For purposes of comparison with the main result in Section 3, note that for $p = 1 = q$, the variance of the asymptotic distribution on the right of (2.2) reduces to $[df(\theta)/d\theta]\sigma^2$, on replacing $\underline{t}$ by σ^2 . Assume that σ is a positive continuous function of θ , the mean of the asymptotic distribution on the right of (2.2). The latter variance is 'stabilized' (to the value 1) by setting

$$(2.3) \quad \left[\frac{df(\theta)}{d\theta} \right]^2 \sigma^2(\theta) = 1,$$

and any solution of (2.3), such as

$$(2.4) \quad f(t_n) = \int^{t_n} \frac{d\theta}{\sigma(\theta)}$$

is a variance-stabilizing transformation.

3. A dispersion-stabilizing transformation. In this section, a dispersion-stabilizing transformation will be obtained with the aid of Lemma 3 for the special case where the correlation matrix $\underline{P}$ is specified and constant. It turns out that the component transformations are then linear combinations of variance-stabilizing transformations in which the coefficients are the elements of an orthogonal diagonalizing matrix for $\underline{P}$. This result is expressed in the theorem which follows.

THEOREM. Let $\{\underline{t}_n, n = 1, 2, \dots\} = \{(t_{1n} \dots t_{pn})', n = 1, 2, \dots\}$ be a sequence of p -component random vectors, $\underline{\theta} = (\theta_1, \dots, \theta_p)'$ a constant (non-random) p -component vector and $\underline{\Phi}(\underline{\theta})$ a constant symmetric positive definite matrix of size $p \times p$ which is independent of n . Let $\sigma_{jj}(\theta_j)$ be the j th diagonal element of $\underline{\Phi}(\underline{\theta})$, and define the matrices $\underline{\Omega}(\underline{\theta})$, $\underline{P}$, $\underline{\Lambda}$, $\underline{C}$ by

$$\underline{\Omega}(\underline{\theta}) = \begin{pmatrix} \sigma_{11}^{-\frac{1}{2}}(\theta_1) & 0 & \dots & 0 \\ 0 & \cdot & & \cdot \\ \vdots & & \ddots & \\ \vdots & & & \cdot \\ 0 & \dots & \dots & \sigma_{pp}^{-\frac{1}{2}}(\theta_p) \end{pmatrix},$$

$$\underline{\Phi}(\underline{\theta}) = \underline{\Omega}(\underline{\theta}) \underline{P} \underline{\Omega}(\underline{\theta}),$$

$$\underline{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \cdot & & \cdot \\ \vdots & & \ddots & \\ \vdots & & & \cdot \\ 0 & \dots & \dots & \lambda_p \end{pmatrix},$$

$$\underline{C} \underline{P} \underline{C}' = \underline{\Lambda},$$

where $\underline{P}$ is independent of $\underline{\theta}$, the λ_j are the eigenvalues of $\underline{P}$, and $\underline{C}$ is orthogonal. Assume that

- (i) $\sigma_{jj}(\theta_j)$ is continuous ($j = 1, 2, \dots, p$),
- (ii) $\mathcal{L}\left[n^{\frac{1}{2}}(\underline{t}_n - \underline{\theta})\right] \rightarrow N(0, \underline{\Delta}(\underline{\theta}))$.

Then

$$\mathcal{L}\left[n^{\frac{1}{2}}\{\underline{C} \underline{h}(\underline{t}_n) - \underline{C} \underline{h}(\underline{\theta})\}\right] \rightarrow N(0, \underline{\Delta}) ,$$

where

$$\underline{h}(\underline{t}_n) = \begin{pmatrix} \int^{t_{1n}} \sigma_{11}^{-\frac{1}{2}}(\theta_1) d\theta_1 \\ \vdots \\ \int^{t_{pn}} \sigma_{pp}^{-\frac{1}{2}}(\theta_p) d\theta_p \end{pmatrix} .$$

PROOF. Define

$$(3.2) \quad \underline{f}(\underline{t}_n) = \underline{C} \underline{h}(\underline{t}_n) .$$

Then⁴

$$\begin{aligned} \frac{\partial \underline{f}(\underline{\theta})}{\partial \underline{\theta}} \underline{\Delta}(\underline{\theta}) \left(\frac{\partial \underline{f}(\underline{\theta})}{\partial \underline{\theta}} \right)' &= \underline{C} \underline{\Omega}^{-1}(\underline{\theta}) \cdot \underline{\Omega}(\underline{\theta}) \underline{P} \underline{\Omega}(\underline{\theta}) \cdot \underline{\Omega}^{-1}(\underline{\theta}) \underline{C}' \\ &= \underline{C} \underline{P} \underline{C}' \\ (3.3) \quad &= \underline{\Delta} . \end{aligned}$$

The theorem now follows directly from Lemma 3.

Observe finally that the dispersion-stabilizing transformation (3.2) is

$$(3.2') \quad f_i(t_n) = \sum_{j=1}^p c_{ij} \cdot \int^{t_{jn}} \frac{d\theta_j}{\sigma_{jj}^{\frac{1}{2}}(\theta_j)} \quad (i = 1, 2, \dots, p) ,$$

where $\underline{C} = (c_{ij})$, $i, j = 1, 2, \dots, p$, and that $\int^{t_{jn}} d\theta_j / \sigma_{jj}^{\frac{1}{2}}(\theta_j)$ is the variance-

⁴ Note that (3.3) is the multivariate extension of (2.3).

stabilizing transformation for the j th characteristic. For $p = 1$, (3.2') reduces to the usual variance-stabilizing transformation (2.4). For $p = 2$, we also obtain a particularly simple result, in that here $\underline{C}$ is independent of $\underline{P}$. Thus, since for $p = 2$,

$$\underline{C} = \begin{pmatrix} 2^{-\frac{1}{2}} & 2^{-\frac{1}{2}} \\ 2^{-\frac{1}{2}} & -2^{-\frac{1}{2}} \end{pmatrix},$$

(3.2') becomes

$$(3.4) \quad \begin{aligned} f_1(t_{1n}, t_{2n}) &= 2^{-\frac{1}{2}} \left(\int_{t_{1n}}^{t_{1n}} \frac{d\theta_1}{\sigma_{11}^{\frac{1}{2}}(\theta_1)} + \int_{t_{2n}}^{t_{2n}} \frac{d\theta_2}{\sigma_{22}^{\frac{1}{2}}(\theta_2)} \right), \\ f_2(t_{1n}, t_{2n}) &= 2^{-\frac{1}{2}} \left(\int_{t_{1n}}^{t_{1n}} \frac{d\theta_1}{\sigma_{11}^{\frac{1}{2}}(\theta_1)} - \int_{t_{2n}}^{t_{2n}} \frac{d\theta_2}{\sigma_{22}^{\frac{1}{2}}(\theta_2)} \right). \end{aligned}$$

As an example, if t_{1n} and t_{2n} are each Poisson variates (as in many counting problems), for which the variance-stabilizing transformation is the square root transformation, (3.4) reduces to

$$(3.5) \quad \begin{aligned} f_1(t_{1n}, t_{2n}) &= 2^{\frac{1}{2}} (t_{1n}^{\frac{1}{2}} + t_{2n}^{\frac{1}{2}}), \\ f_2(t_{1n}, t_{2n}) &= 2^{\frac{1}{2}} (t_{1n}^{\frac{1}{2}} - t_{2n}^{\frac{1}{2}}). \end{aligned}$$

REFERENCES

- [1] ANDERSON, T. W. (1958). An Introduction to Multivariate Statistical Analysis. Wiley, New York.
- [2] BOX, G. E. (1953). Non-normality and tests on variances. Biometrika 40 318-335.
- [3] BOX, G. E. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. Ann. Math. Stat. 25 290-302.
- [4] BOX, G.E. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effect of inequality of variance and of correlation of errors in the two-way classification. Ann. Math. Stat. 25 484-478.
- [5] HOEFFDING, WASSILY and ROBBINS, HERBERT. The central limit theorem for dependent random variables. Duke Math. J. 15 773-780.
- [6] HORSNELL, G. (1953). The effect of unequal group variances on the F-test for the homogeneity of group means. Biometrika 40 128-136.
- [7] LEHMANN, E. L. (1959). Testing Statistical Hypotheses. Wiley, New York.
- [8] RAO, C.R. (1952). Advanced Statistical Methods in Biometric Research. Wiley, New York.
- [9] SCHEFFÉ, H. (1959). The Analysis of Variance. Wiley, New York.