Reprinted from
BIOMETRICS
THE BIOMETRIC SOCIETY, Vol. 22, No. 1, March 1969

A MULTIVARIATE PALEONTOLOGICAL GROWTH PROBLEM

R. A. REYMENT

Paleontologiska Institutionen, Uppsala Universitet, Box 558, S-75122, Uppsala 1, Sweden

SUMMARY

In a study of the theory of principal component analysis Anderson [1963] showed that the expression:

$$n\{d_i\beta_i'\mathbf{S}^{-1}\beta_i + d_i^{-1}\beta_i'\mathbf{S}\beta_i - 2\},$$

where n is the number of degrees of freedom of the sample covariance matrix **S**, the d_i are the eigenvalues of p-variate **S** and the vectors β_i are hypothetical eigenvectors of unit length, follows the limiting distribution of χ^2 with $p - 1$ D.F. This paper describes the application of this procedure in certain growth studies, the examples used being drawn from the author's biometric studies on fossil and Recent ostracods (Crustacea: Arthropoda).

1. INTRODUCTION

For the purposes of the present discussion it will be taken that the N vectors of a sample, $\mathbf{x}_1, \cdots, \mathbf{x}_N$ (N observation vectors) are p-dimensional random vectors with a common p-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

If one considers two p-variate normally distributed populations with the mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, not necessarily different (i.e. O_1 and O_2 in the diagrams of Figure 1 may be coincident), and the covariance matrices, $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, then if the test of equality of these matrices be applied (cf. Kullback [1959]) and it be concluded that $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, the following situation may prevail in the two-dimensional case: referring to diagram (a) of Figure 1, the ellipsoid Ω_1 is merely a translation of ellipsoid Ω_2. Hence, $AB = EF$ and $CD = GH$, AB is parallel to EF and CD is parallel to GH.

If the homogeneity test for covariance matrices leads to the conclusion that $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$, then one of the following conditions may prevail (for two dimensions): diagram (b) of Figure 1 has AB parallel to EF and CD parallel to GH; $EF > AB$ and $GH > CD$. Hence, ellipsoid Ω_2 is a translation of ellipsoid Ω_1 with magnification. In diagram (c), $AB = EF$, $CD = GH$, AB is not parallel to EF and CD is not parallel to GH. Ellipsoids Ω_1 and Ω_2 have the same shape but are rotated in relation to each other. In diagram (d), $AB \neq EF$, $CD \neq GH$, AB is not parallel to EF, and CD is not parallel to GH. Ellipsoids Ω_1 and Ω_2 are thus differently inflated and their axes are rotated in relation to each other.

For three or more dimensions the situation becomes more complicated. Diagram (e) of Figure 1 illustrates the position for three dimensions. Here, two axes of ellipsoids Ω_1 and Ω_2 are parallel to each other, AB is parallel to EF, but not the remaining axes, owing to rotation about AB and EF. Hence, CD is

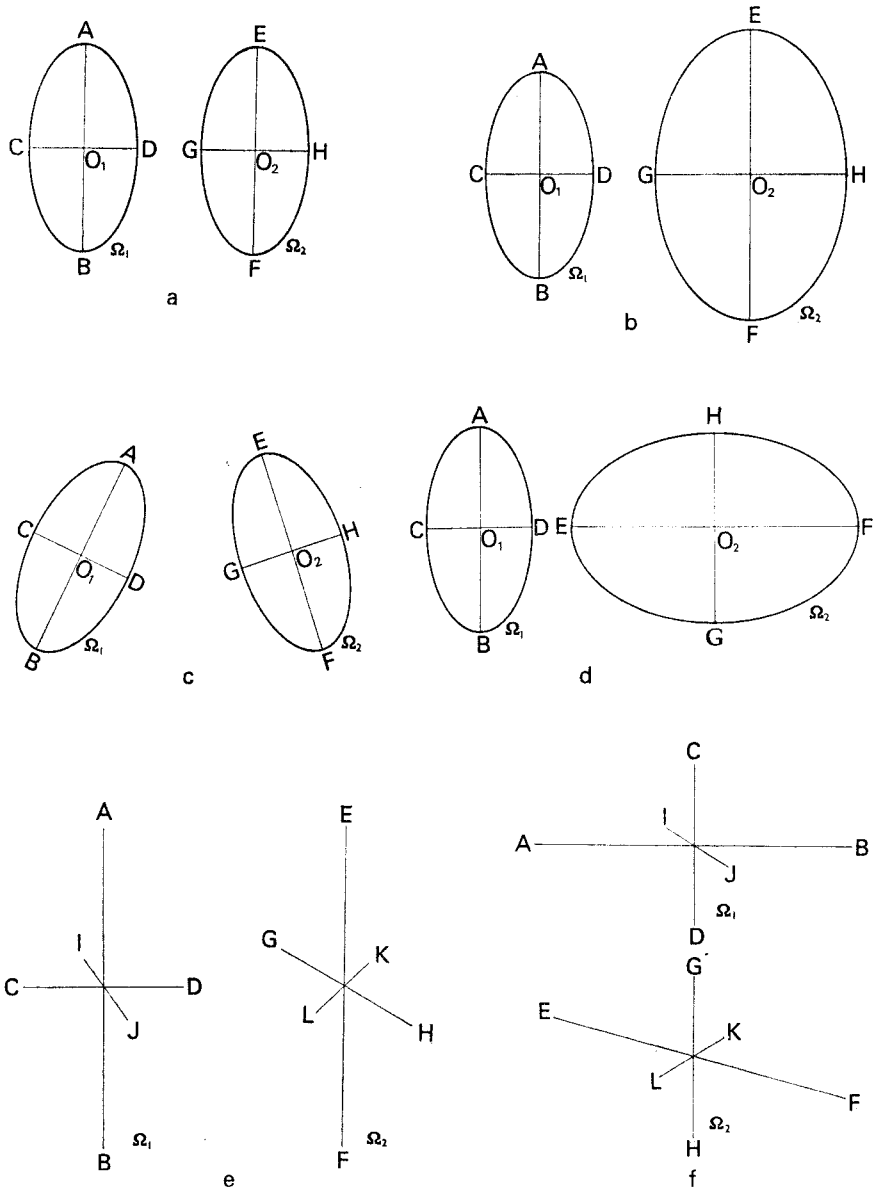

FIGURE 1

not parallel to GH and IJ is not parallel to LK. In this example it has been taken that the ellipsoids have the same shape; thus, $AB = EF$, $CD = GH$, and $IJ = KL$. Diagram (f) indicates the position when the second axes of Ω_1 and Ω_2 are parallel and the first and third axes are rotated about the second axes.

2. THE ANDERSON TEST STATISTIC

The generalized test statistic for homogeneity of covariance matrices in the

form used in Kullback's ([1959] p. 317) discussion is

$$2\hat{I}(H_1 : H_2(*)) = N_1 \log_e (\det S/\det S_1) + N_2 \log_e (\det S/\det S_2), \quad (1)$$

where $NS = N_1S_1 + N_2S_2$ and $N = N_1 + N_2$, H_1 and H_2 are used in the sense of Kullback [1959], S_1 and S_2 are the sample covariance matrices, and N_1 and N_2 the corresponding sample sizes. Expression (1) is approximately distributed as χ^2 with $\frac{1}{2}p(p+1)$ D.F., where p is the number of variables. A better approximation is yielded by the B-distribution (Kullback [1959] p. 317); however, this has been tabulated for only a few dimensions. In using this approximation one requires also

$$\beta^2 = \frac{2p^3 + 3p^2 - p}{12}\left(\sum_{i=1}^{2} N_i^{-1} - N^{-1}\right)$$

and $B^2 = 2\hat{I}(H_1 : H_2(*))$ with $\frac{1}{2}p(p+1)$ D.F.

In order to disclose relative heterogeneity in the orientation of the axes of two ellipsoids of scatter when the generalized test statistic (1) indicates $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$ it is, for $p = 2$ or 3, sufficient to demonstrate that the first two axes are parallel to each other in order to localize the inequality in the covariance matrices to condition (b) of Figure 1.

Some interest may attach to comparing the relative differences in the inflations of ellipsoids Ω_1 and Ω_2. This may be gauged approximately by comparing the eigenvalues of matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$. In order to apply a test due to Anderson [1963] let $\delta_1 > \cdots > \delta_p > 0$ be the p eigenvalues of the positive definite matrix $\boldsymbol{\Sigma}_1$, $|\boldsymbol{\Sigma}_1 - \delta\mathbf{I}| = 0$, and $\gamma_1, \cdots, \gamma_p$ the corresponding normalized eigenvectors which satisfy $\boldsymbol{\Sigma}_1\gamma_i = \delta_i\gamma_i$ and $\gamma_i'\gamma_i = 1$, and let $\lambda_1 > \cdots > \lambda_p > 0$ be the p eigenvalues of the positive definite matrix $\boldsymbol{\Sigma}_2$, $|\boldsymbol{\Sigma}_2 - \lambda\mathbf{I}| = 0$, and $\beta_1, \cdots, \beta_p$ are the corresponding normalized eigenvectors which satisfy $\boldsymbol{\Sigma}_2\beta_i = \lambda_i\beta_i$ and $\beta_i'\beta_i = 1$. If the roots δ_i, respectively λ_i, are different, then $\gamma_i'\gamma_j = 0$ and $\beta_i'\beta_j = 0 (i \neq j; i, j = 1, \cdots, p)$.

For the first eigenvalues and eigenvectors of matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$,

$$\delta_1\gamma_1'\boldsymbol{\Sigma}_1^{-1}\gamma_1 = \delta_1\delta_1^{-1} = 1,$$

and,

$$\delta_1^{-1}\gamma_1'\boldsymbol{\Sigma}_1\gamma_1 = \delta_1^{-1}\delta_1 = 1,$$

and similarly for $\boldsymbol{\Sigma}_2$. Hence, if, say, the first eigenvector of $\boldsymbol{\Sigma}_2$, β_1, has the same direction cosines as the first eigenvector of $\boldsymbol{\Sigma}_1$, γ_1, then $\delta_1\beta_1'\boldsymbol{\Sigma}_1^{-1}\beta_1 = 1$ and $\delta_1^{-1}\beta_1'\boldsymbol{\Sigma}_1\beta_1 = 1$. Anderson ([1963] p. 144) has put forward a procedure for testing the null hypothesis that a given eigenvector (principal component) is a specified vector. This procedure is based on the normal limiting distribution possessed by $\sqrt{n}(\mathbf{c}_1 - \gamma_1)$, where γ_1 is the first eigenvector of a population covariance matrix and $\mathbf{c}_1$ a sample estimate thereof, based on a sample of size $n + 1$. As an approximate test of the hypothesis that a given eigenvector has the same direction as the i-th eigenvector of a very large sample estimate of a covariance matrix it is suggested that the following version of Anderson's test statistic might be applicable:

$$n(d_i\mathbf{b}_i'\mathbf{S}_1^{-1}\mathbf{b}_i + d_i^{-1}\mathbf{b}_i'\mathbf{S}_1\mathbf{b}_i - 2). \quad (2)$$

Here, $n + 1$ is the sample size of $\mathbf{S}_1$, the estimate of $\boldsymbol{\Sigma}_1$, $n = N_1 - 1$, d_i is the sample estimate of δ_i and $\mathbf{b}_i$ that of $\boldsymbol{\beta}_i$, the latter being based on a very large sample. If $\boldsymbol{\beta}_1$ is normalized to have unit length then having the same direction as $\boldsymbol{\gamma}_1$ means $\boldsymbol{\beta}_1 = \boldsymbol{\gamma}_1$.

The approach followed in this paper is to use Anderson's test when the hypothesis of homogeneity of covariance matrices has been rejected by (1), which has the consequence that the correct probabilities for this test are conditional on the rejection of homogeneity at a certain level. It is therefore difficult, on the basis of the tables of chi-square, to make an actual statement of the probability level at which Anderson's test is significant. For this reason a conservative attitude should be adopted in interpreting the results of calculations. The Anderson test statistic has a limiting distribution of chi-square with $p - 1$ D.F.

3. BIOLOGICAL DISCUSSION

In some biological work it may be considered useful to analyze homogeneity occurring in covariance matrices to as full a degree as possible. Significant differences in the inflation of covariance matrices, but not orientation, may in the writer's experience result from such fundamental causes as sexual dimorphism and ontogenetic differences in stages in the growth series of ostracod crustaceans, for example.

There is a fair body of evidence to suggest, that differences in the orientation of covariance matrices may derive from inherent growth patterns of organisms. Multivariate statistical studies on insects (Blackith [1960]; Matsuda and Rohlf [1961]), reptiles (Jolicoeur and Mosimann [1960]), mammals (Jolicoeur [1963]), protozoans and crustaceans (Reyment [1961, 1963]; Reyment and Brännström [1962]) appear to offer some support for this interpretation.

The growth interpretation of the principal axes of a covariance matrix of logarithmically (base 10) transformed morphologic variables rests on the concept of the first principal component as an allometric factor and the remaining principal components as shape factors of various kinds. Rao ([1964] p. 344) considers this interpretation of principal components arbitrary and has suggested a development based more definitely on special considerations of size and shape. The investigations accounted for in the above-mentioned publications seem, however, to suggest the existence of fairly persuasive evidence in favor of the growth-allometric properties of the first principal component of logarithmically transformed variables.

Regarding the examples illustrated in Figure 1 as descriptive of biological situations it may be suggested, that case (1a) could occur in the comparison of samples of the same species from the same population of the same age and reared under the same environmental conditions. Case (1b) could occur in the sexual dimorphism of size, without relative growth differences, the samples being drawn from the same population, reared under identical environmental conditions. Case (1c) is indicative of a situation in which different patterns of growth occur, but where there is the same breadth of variation in corresponding variables;

this could take place in special cases of sexual dimorphism and polymorphism. Case (1d) could be found where more complex conditions of sexual dimorphism occur, or where the populations sampled are of different environmental origin, or where different species are involved. These examples are, naturally, by no means exhaustive.

4. EXAMPLES

The two examples given here derive from studies on living and fossil ostracods by the writer (ostracods are microscopic bivalved crustaceans). The computations were performed by means of a double precision FORTRAN program, prepared by the author, on the CD 3600 computer of the University of Uppsala and checked on an IBM 7040 at the University of Kansas. Copies of the program may be obtained from the Geological Survey of Kansas, University of Kansas.

1. *Recent freshwater ostracod.*—This example treats two samples of an African parthenogenetic species of *Strandesia*, drawn from individuals cultured in two markedly different and extreme environments. The dimensions analyzed are length, height, and breadth of carapace, in that order. The respective sample covariance matrices (based on logarithmically [base 10] transformed observations) are:

(a) The sample from the first environment ($N = 365$).

$$\mathbf{S}_1 = \begin{bmatrix} .0003390 & .0002137 & .0003069 \\ .0002137 & .0003393 & .0002552 \\ .0003069 & .0002552 & .0005396 \end{bmatrix}.$$

(b) The sample from the second environment ($N = 908$).

$$\mathbf{S}_2 = \begin{bmatrix} .0005074 & .0002332 & .0002084 \\ .0002332 & .0003311 & .0002448 \\ .0002084 & .0002448 & .0004515 \end{bmatrix}.$$

In terms of formula (3), the eigenvalues, d_i, are: $d_1 = .0009445$, $d_2 = .0001655$, $d_3 = .0001079$. The corresponding eigenvectors, $\mathbf{g}_i$, are:

$$\mathbf{G} = \begin{bmatrix} .525445 & .045020 & -.849636 \\ .481346 & .807700 & .340480 \\ .701580 & -.587872 & .402731 \end{bmatrix}.$$

The eigenvalues of $\mathbf{S}_2$, l_i, are, $l_1 = .0008923$, $l_2 = .0002736$. $l_3 = .0001242$, and the corresponding eigenvectors, $\mathbf{b}_i$, are:

$$\mathbf{B} = \begin{bmatrix} .628000 & -.733152 & .260973 \\ .515343 & .140505 & -.845387 \\ .583130 & .665392 & .466062 \end{bmatrix}.$$

For the application of (1), one has that $\log_e(\det \mathbf{S}_1) = -24.806$, $\log_e(\det \mathbf{S}_2) = -24.220$, and $\log_e(\det \mathbf{S}) = -24.294$. Using (1), $2\hat{I}(H_1 : H_2(*)) = B^2 = 118.43$, which for $\beta^2 = 0.0199$ and six D.F., exceeds the value given in Fisher's table (cf. Kullback [1959] Table 3, p. 380). This result is significant and therefore indicates the distinct possibility that $\mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$.

Employing the larger sample as a close approximation to the population, the application of (2) gives:

$$\chi^2 = 364[(0.0009445)(1235.0164) + (1/0.0009445)(0.0009238) - 2] = 52.755.$$

The test (1) indicates, that there is good evidence for the possibility of heterogeneity in the covariance matrices. Test (2), with respect to the first eigenvectors, indicates the strong possibility that these are not collinear; hence with respect to these axes, it may be concluded, that the ellipsoids of scatter are probably differently oriented. The same test was applied to the second eigenvector, with d_2 replacing d_1, and b_2 replacing b_1. This time a value of $\chi^2 = 91.292$ was obtained, which for 2 D.F. is highly significant. Thus, the first two principal axes of the ellipsoids of scatter are probably differently oriented.

Concerning the biological significance of these results, it is noted, that inasmuch as the species analysed is parthenogenetic, it should be possible to identify the variational differences found in relation to the reaction of the organisms to environment, on the one hand, and to those arising from genetic variation, on the other. This latter should, ideally, agree in both samples. The results of the foregoing calculations would appear to suggest that all three principal components (eigenvectors) are correlated with the effects of the environment on the size and shape of the carapace.

2. *Males and females of a fossil marine ostracod.*—The material here studied derives from a single borehole sample and it is assumed that all carapaces belonged to individuals of a species of *Buntonia* living under identical environmental conditions. The covariance matrices of adults and final instars are:

(a) Males ($N = 66$).

$$\mathbf{S}_1 = \begin{bmatrix} .0008556 & .0002544 & .0002349 \\ .0002544 & .0001704 & .0001325 \\ .0002349 & .0001325 & .0001710 \end{bmatrix}.$$

(b) Females ($N = 970$).

$$\mathbf{S}_2 = \begin{bmatrix} .0030804 & .0006507 & .0006158 \\ .0006507 & .0010193 & .0006158 \\ .0006158 & .0006158 & .0008827 \end{bmatrix}.$$

In terms of formula (2), the eigenvalues d_i are: $d_1 = .0010223$, $d_2 = .0001371$, $d_3 = .0000377$. The corresponding eigenvectors, $\mathbf{g}_i$, are:

$$\mathbf{G} = \begin{bmatrix} .901046 & -.432009 & .038540 \\ .315396 & .591635 & -.741953 \\ .297728 & .680689 & .669344 \end{bmatrix}.$$

The eigenvalues of $\mathbf{S}_2$, l_i, are: $l_1 = 0.0034969$, $l_2 = 0.0011543$, $l_3 = 0.0003311$, and the corresponding eigenvectors, $\mathbf{b}_i$, are:

$$\mathbf{B} = \begin{bmatrix} .906776 & -.421449 & -.011784 \\ .309346 & .684066 & -.660583 \\ .286466 & .595345 & .750661 \end{bmatrix}.$$

For the application of (1), one has that $\log_e(\det \mathbf{S}_1) = -25.967784$, $\log_e(\det \mathbf{S}_2) = -20.433091$ and $\log_e(\det \mathbf{S}) = -20.592684$. Using (1), $2\hat{I}(H_1 : H_2(*)) = B^2 = 194.736$, which for $\beta^2 = 0.100422$ and 6 D.F. is highly significant. This therefore offers strong evidence in favor of the hypothesis $\mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$. Application of formula (2) gives, for the first eigenvector, $\chi^2 = 0.0828$, which for 2 D.F. is not significant. Hence, the test offers no support for the hypothesis that the first principal axes of the scatter ellipsoids are differently oriented. For the second eigenvectors, $\chi^2 = 2.0496$, which for 2 D.F. is not significant. Hence, the heterogeneity occurring in these matrices is connected mainly with the differing inflations of the scatter ellipsoids and also possibly to rotation about the first principal axes, which appear to be parallel to each other. These results accord reasonably well with what one may find in studies of the sexual dimorphism of certain groups of ostracods. This is a three-dimensional example of case (b) of Figure 1, with the males corresponding to ellipsoid Ω_1 and the females to ellipsoid Ω_2.

ACKNOWLEDGMENTS

Professors T. W. Anderson, M. S. Bartlett and C. R. Rao and Drs. H. Seal and T. W. Burnaby have kindly contributed advice and read the paper. Any inaccuracies occurring are naturally my responsibility. Helpful comments were contributed by the referees. The paper was read at the 36th Session of the International Statistical Institute, Sydney, Australia in September, 1967.

The work was supported by contracts 2320-17/7242 and 2320-16-7330 of the Swedish Natural Science Research Council.

UN PROBLEME A PLUSIEURS VARIABLES EN CROISSANCE PALEONTOLOGIQUE

RESUME

Dans une étude de la théorie de l'analyse en composantes principales, Anderson montra [1963] que l'expression:

$$n\{d_i \beta_i' \mathbf{S}^{-1} \beta_i + d_i^{-1} \beta_i' \mathbf{S} \beta_i^{-2}\}$$

suivait la distribution du χ^2 à $p - 1$ ddl (n est le nombre de degrés de liberté de la matrice $\mathbf{S}$ de variance-covariance observée; les d_i sont les valeurs propres de $\mathbf{S}$ et les β_i sont les vecteurs propres (unitaires correspondants). Cet article décrit l'application de cette méthode à des études de croissance; les exemples utilisés sont tirés d'études biométriques de l'auteur sur les ostracods fossiles et récents (crustacea: arthropoda).

REFERENCES

Anderson, T. W. [1958]. *An Introduction to Multivariate Statistical Analysis.* Wiley, New York.

Anderson, T. W. [1963]. Asymptotic theory for principal component analysis. *Ann. Math. Statist. 34,* 122–48.

Blackith, R. E. [1960]. A synthesis of multivariate techniques to distinguish patterns of growth in grasshoppers. *Biometrics 16*, 28–40.

Burnaby, T. P. [1966]. Growth invariant discriminant functions and generalized distances. *Biometrics 22*, 96–110.

Jolicoeur, P. [1963]. The degree of robustness in *Martes americana*. *Growth 27*, 19–27.

Jolicoeur, P. and Mosimann, J. E. [1960]. Size and shape variation in the painted turtle. A principal component analysis. *Growth 24*, 335–54.

Kullback, S. [1959]. *Information Theory and Statistics*. Wiley, New York.

Matsuda, K. and Rohlf, F. J. [1961]. Studies of relative growth in Gerridae (5). Comparison of two populations. (Heteroptera: Insecta) *Growth 25*, 211–7.

Pearce, S. C. and Holland, D. A. [1960]. Some applications of multivariate methods in Botany. *Appl. Statist. 9*, 1–7.

Rao, C. R. [1964]. The use and interpretation of principal component analysis in applied research. *Sankhyā A 26*, 329–58.

Reyment, R. A. [1961]. Quadrivariate principal component analysis of *Globigerina yeguaensis*. *Stockh. Contr. Geol. 8*, 17–26.

Reyment, R. A. [1963]. Studies on Nigerian Upper Cretaceous and Lower Tertiary ostracoda. III. Stratigraphical, paleontological and biometrical conclusions. *Stockh. Contr. Geol. 14*, 1–144.

Reyment, R. A. and Brännström, B. [1962]. Certain aspects of the physiology of *Cypridopsis* (Ostracoda, Crustacea). *Stockh. Contr. Geol. 9*, 207–43.

Reyment, R. A. and Naidin, D. P. [1962]. Biometric study of *Actinocamax verus* s. 1. from the Upper Cretaceous of the Russian Platform. *Stockh. Contr. Geol. 9*, 147–206.

Received March 1968, *Revised August* 1968