

A Commentary on the Use of Multivariate Statistical Methods in Anthropometric Research

CHARLES J. KOWALSKI

Department of Oral Biology and The Statistical Research Laboratory,
University of Michigan, Ann Arbor, Michigan 48104

ABSTRACT A critical review of the increasing emphasis being placed on the use of multivariate statistical methods in anthropometric research is given. Particular attention is paid to multivariate techniques for testing hypotheses concerning mean vectors, principal components analysis and the use of discriminant functions, but some more general comments about the proper role of statistics in research are included. It is argued that multivariate techniques often do not allow the effective description and communication of the informational content of a body of data and that much additional research — both from the standpoint of theory and from the standpoint of practice — must be done before multivariate analysis can fulfill the promise it holds for the physical anthropologist.

The need for quantitative measurements to supplement verbal descriptions of body size and shape has been evident since early in the 19th Century when physical anthropology was becoming a distinct discipline, and anthropometry has since occupied the pre-eminent position in the subject. The evolution of techniques ostensibly suited for the analysis of anthropometric measurements has progressed rapidly, and it is now commonplace for physical anthropologists to employ highly sophisticated methods of multivariate statistical analysis¹ in an attempt to gain some insight into morphology, function, heritability, classification, discrimination and growth. And the use of such methods is increasing. It is therefore suggested that the time has come for a critical evaluation of the use of these methods; to point out their limitations as well as their strengths; to recognize problems that remain as well as problems which have been "solved." Half the job has already been done. In a recent paper, Howells ('69a) lucidly presented the case for multivariate analysis in anthropometry and additional favorable discussions have been given by Ashton, Healy and Lipton ('57), Barnard ('35), Bronowski and Long ('51), Harris ('65), Healy ('65), Howells ('57, '68, '69b), Lavelle et al. ('70), Oxnard ('68), Rao ('48a, '61) and Schull ('62) to name just a few.

The present paper is not meant to constitute an "equal time" rebuttal; nor are many of the notions contained herein diametrically opposed to those proffered by Howells. It is intended, simply, to show that there does exist another side to the story and that much research needs to be done before multivariate analysis can fulfill the promise that it holds for the physical anthropologist.

I want to emphasize that this paper does not question statistical theory properly used but, rather, the increasing emphasis on *multivariate* statistical methods in certain contexts. The question of exactly what the formal theory of statistics can contribute to *data analysis* (Tukey, '62) was considered by Tukey and Wilk ('65) and they indicated that data analysis can gain much from formal statistics, *but only if the connection is kept adequately loose*. For those wishing to go a step further, Feller ('69) is a good place to start.

THE OTHER SIDE OF THE STORY

The other side of the story, for me at least, began in a somewhat unexpected

¹ A good introduction to the nature and scope of multivariate analysis is given in Chapter 1 of Kendall ('57). Following Kendall, I find it easier to say what multivariate analysis is not than to say what it is and to prescribe my domain of discussion by enumeration rather than by definition. I exclude bivariate analyses, such as correlation and regression, and enumerate the procedures included in the body of the text.

manner. I was busily engaged in reading Krishnaiah's ('66) collection of papers on multivariate statistical analysis. My intent was to learn something about these methods, methods which I assumed would be of wide applicability and lead to heretofore unachieved levels of insight into the structure of multidimensional phenomena. I was convinced that this was in fact happening for the first 520 pages. Then an article by Kempthorne ('66) appeared: In his introduction he flatly stated,

"It is rare for an experimenter to obtain a single response, and I cannot recall being consulted on an experiment in this class. But I have never used or suggested the procedures of statistical analysis presented so excellently by Dr. Anderson ('58). Also I have yet to see any convincing examples of experimental data in which the standard techniques of multivariate analysis have led to scientific insight."

Never used or suggested their use! Never an example providing scientific insight! And this by one of the most eminent biologically-oriented data analysts of our time. While Kempthorne did hedge a bit (with tongue in cheek?) on the use of factor analysis, "I have no doubt that the psychologists find that the techniques of factor analysis give them insight," one should read the classic paper by Cureton ('39) before assuming that factor analysis (Harman, '67; Horst, '65) is somehow above reproach — even among psychologists.

The root of the conceptual problem lies in honestly interrogating ourselves about what we expect statistical analyses to accomplish for us. If our answer is to provide insightful descriptions of the informational content of the data or, more simply, to form opinions about the underlying situation, then it is certainly possible to obtain such descriptions and to form opinions on the basis of univariate and/or bivariate techniques, which have the advantages of being communicable and relatively easy to understand. The question of what multivariate analysis can provide over and above separate univariate analyses is usually answered by a statement to the effect that it "allows comparisons of several groups or populations to be based on many variables, all treated simultaneously and with due regard for the effects of correlation" (Right-

mire, '70a) or that it "assesses the totality of the data," but this global assessment may introduce as much confusion into the interpretation of the data as it is designed to remove. As Tukey ('62) has said, ". . . multiple-response differences are not simple, are usually not easy to think about, are usually not easy to describe." Some examples may help to illustrate the source of this confusion in particular cases.

Multivariate T^2 and D^2 statistics

The problem of the comparison of treatment effects when the response is multidimensional in character is intrinsically complicated by the fact that there is, in general, no unique linear ordering for vectors (Gnanadesikan and Wilk, '69). The usual way of circumventing this problem is arbitrarily to associate a measure of size with each vector and make the comparisons in terms of *its* size. Examples of this strategy are Mahalanobis' ('30) D^2 , Hotelling's ('31) T^2 and Fisher's ('38) linear discriminant function. These statistics are computed "with due regard for the effects of correlation" but may have little to offer over and above separate univariate analyses when one is called on to interpret the results of his analysis. As stated by Gnanadesikan ('70),

"Unlike the univariate situation, for multiple-response problems, the statistics that are used for purposes of tests of hypotheses . . . are complex derivatives from the data and do not generally have significant value as easily understood summaries of the structure underlying the multiresponse data."

Thus one problem with this approach is that the associated test statistics often turn out to be complicated functions which do not have intuitive value as summaries of the informational content of the data from which they are computed. Another is that their use often results in a substantial loss of power and hence in paradoxes when compared with the results achieved by the corresponding univariate procedures.

The former problem is of considerable practical importance and deserves more study than it apparently has received to date. Recognizing the need to keep measures of distance simple and communicable, Sokal ('61) recommended the use of

Pythagorean distance in a variety of taxonomic problems. However, this choice of metric simply doesn't work (Blackith, '65) when the characters under study exhibit little intra-group variation, i.e., in precisely the case when one might expect discrimination or classification procedures to be most sensitive, and the call has gone out for the employment of "more realistic" — and more complicated — metrics. So, in practice, one is often faced with the choice of using a simple function to communicate nonsense, or using a complicated function and not communicating at all. Clearly the "optimal strategy," if it exists, lies somewhere between these extremes but the prevailing confused theoretical situation needs to be ironed out before much in the way of objective guides to this choice are available. A good discussion is given by Dempster ('69: 219) who concludes, "... no single discriminator or single distance measure can tell the whole story." What this means for "the other side of our story" is that we need more experience with various measures of distance, and with combinations of these measures, before we can assume that we are measuring something with any biological relevance.

As an example of the latter problem, it is not uncommon *in practice* to find significant differences at a given level of significance by applying Student's *t*-test on each individual measurement, whereas the T^2 or D^2 test (Anderson, '58:101) utilizing all the measurements simultaneously fails to indicate significance at the same level. This raises an important point that is, in the author's opinion, often not appreciated by practitioners of multivariate significance tests, namely, that the chance of finding significant differences by a multivariate test is lessened when variables which do not contain information about group differences are included in the analysis. An example of this phenomenon was given by Rao ('66) who discussed the analysis of some anthropometric data involving samples from two supposed bivariate normal populations assumed to have common variances and covariance. Taking each variable on its own, in both cases the differences between the sample means gave rise to a value of *t* which was significant at the

5% level. But when the variables were considered simultaneously, the value of Hotelling's T^2 corresponded to a *P*-value of 12%. Healy ('69) exhibited this paradox — which has become known as "Rao's paradox" — in a very simple form and noted that its consequences are of particular consternation in the context of classification and diagnostic problems, in which many investigators have empirically reached the conclusion that the inclusion of "noise" variables in the analysis can have a deleterious effect on the sensitivity of the procedure. As stated by Rao ('66), "Here is a dangerous situation in which the inclusion of an extra character decreases the discriminatory power of the test." So much for the "shotgun approach," which was characterized succinctly by Koons ('62):

"Although it appears obvious that one chooses variables that 'seem' to belong together for manipulation — whether the choice is by hunch or hypothesis — the temptation of the data-engulfing computer is hard to resist. A shotgun approach certainly yields data; however, one is often hard pressed to interpret the results obtained . . . we can hardly defend our model if we cannot interpret the results it provides."

The problem is compounded by the fact that our knowledge of the power or sensitivity of multivariate tests concerning mean vectors is slight and one can have but little faith in a test of significance if one has no knowledge of its power. One might hope, and many have so conjectured, that the robustness and power of the univariate *t*-test (a good discussion is given in Chapter 10 of Scheffé, '59) would carry over to the multivariate case, but the available evidence tends to indicate that this anticipation is unwarranted. Ito and Schull ('64) and Holloway and Dunn ('69) have investigated the robustness of Hotelling's T^2 with respect to inequality of covariance matrices and their results are not very encouraging. Holloway and Dunn ('69) noted that with unequal sample sizes and covariance matrices we may have either a test with an unreasonably large level of significance and somewhat higher power than we would expect, or a test with a very low level of significance and very low power. In short, we have very little control over the error probabilities of the test. Keep-

ing the sample sizes equal, which is seldom possible in anthropometric research, does tend to keep the actual level of significance close to the nominal level, but does not help in maintaining the power of the test. And, when the covariance matrices are unequal, increasing the number of variables causes a direct increase in the level of significance, i.e., the probability of a type I error, of the test. People have complained long and loud about how the use of separate univariate t-tests muddles up the composite level of significance (though the overall level can often be calculated and almost always bounded) but, unless the covariance matrices are equal, the same sort of phenomenon affects the multivariate test and the greater the number of variables, the greater the distortion of the nominal significance level.

And how can one test the equality of the covariance matrices? Or should one (Box, '53)? The test described by Anderson ('58:247) might be used, but little is known about *its* power or *its* robustness with respect to departures from multivariate normality. Nor does it lead to any ideas of how the matrices differ when significance is reached.

Thus there are a number of problems associated with the use of multivariate tests concerning mean vectors, and other, more subtle, issues could be raised (see, e.g., Chapters 10 and 15 of Dempster, '69). As stated by Tukey and Wilk ('65), "Most multivariate analysis procedures are related, in typical textbooks, to problems of hypothesis testing, usually in a quite artificial framework, as for instance in so-called multivariate analysis of variance. Such procedures provide almost no help whatever in data analysis . . ." So there is indeed another side to the story. The methods of multivariate statistical analysis hold great promise but more needs to be known about their operating characteristics before they can really tell us much about the informational content of our data.

Discriminant functions

Fisher ('38) defined the discriminant function as that linear combination of the elements of the response vector which maximizes the ratio of the variance be-

tween two groups to that within the groups, and Rao ('52) generalized this concept to the case of several groups. These procedures, as is the case with most multivariate procedures, are based on the assumptions of equal covariance matrices and multivariate normality of the response vector. When the covariance matrices are unequal, Smith ('47) has shown that a quadratic discriminant function is appropriate, but extremely little use of non-linear discriminant functions is evident in practice. Han ('68) and Gilbert ('69) have demonstrated that the agreement between the linear discriminant function and the optimal quadratic is often poor and, in general, that it becomes poorer with an increase in the number of variables considered. Another important aspect to this problem is that the estimation of the probabilities of misclassification (see, e.g., Dunn and Varady, '66; Lachenbruch, '67; Lachenbruch and Mickey, '68) may be considerably biased if the linear function is used when the covariance matrices are unequal (Gilbert, '69). From the standpoint of theory, we do now have a considerable accumulation of knowledge regarding the distributions of quadratic forms (Hotelling, '54) but the application of this theory to the discrimination problem has been largely neglected and more comparative studies are necessary if we are to learn which form of the discriminant function should be used in particular situations.

And what about the normality assumption? Recognizing that this assumption is not always (ever?) tenable in practice, various attempts (e.g., Fix and Hodges, '51; Das Gupta, '64; Kendall, '66) have been made to provide distribution-free or nonparametric methods of discrimination. These have in common the problems of defining "nearness" in terms of distances in multidimensional spaces and the fact that the rules for allocation become extremely complex in precisely the cases which are of the most interest. And how does one choose a metric when, as is so often the case in anthropometric work, some of the variables are continuous and other discrete? Elashoff et al. ('67), Gilbert ('68), Hills ('67), Kurczynski ('70) and Linder ('63) have made valuable contributions towards the definition and im-

plementation of measures of distance between populations on the basis of discrete variables, but mixtures of discrete and continuous variables still present problems (Linhart, '59), and more practical experience with these methods is needed if we are to be able to choose the "correct" metric in a specific situation.

In addition to the choice of metric, the two main conceptual problems in the development of *any* discrimination procedure involve the determination of the dimensionality of the space within which to define the groups and the number of groups one is to have. We have already noted that the introduction of "noise" variables may adversely affect the analysis but we should also note that the discarding of these noise variables from the discriminant function *once they are in* presents a number of difficulties as well. Rao ('46, '48b, '49) developed some tests for hypotheses of the form that "the coefficients of some specified variables in the linear discriminant function are zero" but only recently (Rao, '70) cleared up a number of problems regarding the theory and application of these tests. Unfortunately, some problems of interpretation remain (Bargmann, '70). There are also a number of equally important problems connected with the specification of the number of groups into which the individuals at hand are to be classified. In discriminant analysis we must assume we *know* the number of groups into which our sample is partitioned and the discriminant function serves to allocate a new individual into one or another of these specified groups.² Anderson ('51) gave a comprehensive discussion of the discrimination problem under various degrees of knowledge about the populations under consideration but this elegant theory can be used only when the parameters of the populations are *known* (or are at least estimated from "substantially large samples") and when the investigator is willing to specify a utility function measuring the relative importance of the various costs of misclassification and *a priori* probabilities for each of the groups. Failing to use *a priori* probabilities when they are available can be costly: Morrison ('67:132) gave an example where the use of prior probabilities halved the num-

ber of misclassifications. Other problems of this general type were discussed by Isaacson ('54), Lachenbruch ('66) and Rao ('52). Rao ('52), cited by Howells ('69a) as one of the major contributors to multivariate statistical analysis and, in particular, to the discrimination problem, after discussing the theoretical basis of the solution, stated, "The elegant solution . . . has many limitations so far as the practical applications are concerned," and then went on to list and discuss seven of these limitations. Many of these problems are still with us (Rao, '69). Kendall ('57:37) questioned the rationale behind forcing each new individual into one or another of the existing groups, arguing for an "undecided region" with the observation that, "I prefer a situation in which I may reserve judgement to one which forces me into mistakes even if I know the probability of making them," but procedures of this type are in an early stage of development.

Another problem often overlooked when applying discriminant function theory is that the variables included in the discriminant function often have a marked dependency on age (or some other variable). This leads to the question: What is the proper definition of a sample best linear discriminator in the presence of covariates? Attention has been drawn by Cochran ('64) to two different answers to this question. Other ideas were furnished by Cochran and Bliss ('48), Rao ('62, '69) and Tallis ('70). The point is that we often want to use *shape* variables that are independent of size so that we can classify small and large individuals from the same population into the same group. On the other side of the coin, it is quite possible for two groups to differ significantly in some aspect of shape, and yet for a conventional multivariate test along the discriminant function to indicate that no significant differences exist. As stated by Blackith ('65),

"Even when a group of organisms appears to be homogeneous, there are special prob-

² The problem of using the data to determine whether or not the sample members fall into groups and, if so, to delineate the groups is a problem of "cluster analysis." A good review of cluster analysis is given by Bolshev ('69) and a discussion of the distinctions between discrimination and clustering problems is given by Kendall ('66).

lems connected with the variation of shape within the sample, which have not received the attention they deserve; their existence renders the ordinary multivariate significance tests, like Wilks' criterion [a good development of Wilks' criterion is given by Kendall ('57:106)], irrelevant to the examination of differences of form in living material."

Since it is difficult, both theoretically and practically, to identify shape variables that are in fact independent of size (Burnaby, '66; Mosimann, '70) it appears that either some form of covariance adjustment or an approach along the lines suggested by Tallis ('70), where the response vector $X(t)$ is considered as a multivariate time series, will often be necessary in anthropometric investigations. It seems clear that considerable accuracy may be lost by not taking account of the existing regression on age, but exactly how to account for it remains an open problem.

Before leaving this subject it is perhaps appropriate to admit that the discrimination problem is not even definitely solved to everyone's satisfaction in the univariate case when sorting objects *known* to belong to one or the other of *two normal* populations with the *same known* variance, the costs associated with the two kinds of misclassifications being equal. Robbins ('51) has pointed out that the obvious rule of assigning an object to the population whose mean is nearer (using the usual measure of distance) than that of the other population to the value of the object may not be the best rule. It is anybody's guess what happens when the response is multivariate, when no natural measure of distance is available, when normality is suspect, when one doesn't know how many variables to use nor how many groups there are (let alone their prior probabilities), and when the covariance matrices are unknown and possibly unequal. But the moral should be obvious.

Principal components and factor analysis

The basic strategy behind a principal components analysis is to describe the variation of N points in a p -dimensional space by introducing a new set of orthogonal linear coordinates in such a way that

the sample variances of the given points with respect to these new coordinates are in decreasing order of magnitude. Thus the first principal component is such that the projections of the points onto it have maximum variance among all possible linear coordinates; the second principal component has maximum variance subject to being orthogonal to the first; and so on. This technique, due to Pearson ('01) and Hotelling ('33), is perhaps the most widely used multivariate method, both directly as applied to the sample covariance or correlation matrix and indirectly, under various guises collectively referred to as factor analysis.

It is perhaps appropriate to stress at the outset that, although the two techniques of principal component and factor analysis are similar in mathematical structure and are often used more-or-less interchangeably, the strategies behind the two methods are quite different (Kendall, '61:37). In component analysis we begin with the observations and look for components in the hope that we may be able to reduce the dimensions of variation and ascribe to them some biological meaning. We proceed from the data toward a model. In factor analysis we work the other way around; we begin with a model and check to see if it agrees with the data and, if so, we use the data to estimate the parameters of the model. These are quite different notions (Morrison, '67:259) and the distinction between the two approaches should be kept in mind when interpreting the results of one or another of these analyses. While the principal components solution can be viewed as a particular factor solution, there are situations in which it is confusing to attempt to interpret principal components as factors (Cureton, '39). We proceed now to separate discussions of these analyses though many authors confuse the issues by not distinguishing between these two models.

The rationale behind the extraction of principal components seems eminently reasonable and, at first glance at least, the procedure would seem to be useful in the study of morphology since the physical anthropologist is typically interested in determining what measurements or combination of measurements show con-

siderable variation, i.e., measurements which may be used to pinpoint interpersonal differences. The main problem here is that the solution depends on the units of measurement. Anderson ('58:279) says, "Analysis into principal components is most suitable when all the component variables are measured in the same units. If they are not measured in the same units the rationale . . . is questionable; in fact, the analysis will depend on the various units of measurement." So what should one do when, as in a study of craniofacial morphology, angular and linear measurements are collected along with areas, ratios and volumes? What insight will the principal components provide? How can they be interpreted? After giving an example of a principal components analysis, Kendall ('57:26) made the revealing statement:

"The remarkable feature of Stone's work, however, is that he was able to interpret his components. In many cases our principal components do not have an identifiable separate existence and are to be regarded as convenient mathematical artefacts. In others it is arguable whether the components can be given any reality."

Dempster ('69:140), while holding out some hope that future developments may improve the situation, summarized the current state of the art as follows:

"Attempts are sometimes made to regard the factors . . . as hard well-defined variables, but such attempts deserve skeptical scrutiny. It may be that principal component analysis will some day be of use in locating hard underlying factors, such as genetically determined factors. At present, however, the uses are largely descriptive, explanatory, and empirical. Sampling theory and formal procedures for drawing inferences from samples to populations are in an underdeveloped and unsatisfactory state, both for principal components analysis and for methods of factor analysis generally."

It should also be noted that even if all the measurements *are* in the same units, univariate procedures based on ratios, etc., have considerable intuitive appeal and may well behave in a simple way. The early emphasis on indices in anthropometric work may, in this sense, be looked upon as one tradition worth keeping. This is especially evident if one is considering the study of changes of shape with respect to time. There a single measurement like the cephalic index can be

expected to be informative, communicable and relatively easy to handle analytically (Hirschfeld, '70a,b) whereas the changes in the principal component structure, if present, would be formidable to handle, both conceptually and from the point of view of methodology. When considering the problem of measuring change (Harris, '63) it is of considerable importance to keep the structure of the problem as simple as possible. While a number of advances in the analysis of growth data have been made (Bock, '63; Danford, Hughes and McNee, '60; Elston and Grizzle, '62; Geisser, '63; Geisser and Greenhouse, '58; Greenhouse and Geisser, '59; Potthoff and Roy, '64; Rao, '59) a number of problems remain (Horst, '63) not the least of which is the fact that anthropometric data are often too sparse to support the structure assumed by these highly specialized models. Thus Penrose ('54) suggested that the "discredited" coefficient of racial likeness (Pearson, '26) might indeed be more useful than "generalized distances" — which have the advantage of taking the correlation structure of the observations into consideration — in a number of situations. Although the "data-engulfing computer" now makes Penrose's arguments concerning computational difficulty obsolete, he went on to remark: "There are other reasons for not despising simpler indices. It may be doubted whether the anthropological data usually available are always complete enough or sufficiently accurate to justify the application of the most elegant mathematical calculations. . . ."

When we have sufficient data, when the assumptions underlying the use of the technique are satisfied, when we can interpret and communicate the results of the analysis and when multivariate methods contribute insight over and above that which can be achieved using simpler methods their use is certainly justified. But when these criteria are not satisfied it would appear that some case could be made for the employment of simpler procedures. To paraphrase Penrose ('54), there may be a bit of life left in the coefficient of racial likeness.

Turning now to factor analysis, this may be an appropriate place to quote Cureton ('39) since his comments on the

method appear to be relatively unknown among anthropologists:

"Factor theory may be defined as mathematical rationalization. A factor-analyst is an individual with a peculiar obsession regarding the nature of mental ability or personality. By application of higher mathematics and wishful thinking, he always proves that his original fixed idea or compulsion was right or necessary. In the process he usually proves that all other factor-analysts are dangerously insane, and that the only salvation for them is to undergo his own brand of analysis in order that the true essence of their several maladies may be discovered. Since they never submit to this indignity, he classes them all as hopeless cases, and searches about for some branch of mathematics which none of them is likely to have studied in order to prove that their incurability is not only necessary but also sufficient."

Of course this quote is over thirty years old, not directed at physical anthropologists, and some progress has since been made. But some nagging questions persist. Consider, for example, the notion of the rotation of factor solutions vis-à-vis the justification of one's "fixed idea." The claim is that the rotated solution is often "easier to interpret," but what exactly does this mean? Has anyone ever found that his first factor (or principal component) was not a "generalized size factor" or has anyone ever *disproved* his prior opinions on the basis of a disagreeable factor solution? Has anyone ever found a *useful* morphological factor which was not previously identified on the basis of biological considerations? Might not the emphasis on complicated multivariate techniques be the result of a quest to find a branch of mathematics that will prove one's sophistication *by* sacrificing interpretability and communicability? Is it not possible to by-pass some of the mystique and concentrate on variables, or combinations of variables, that are suggested by *biological* considerations? Virtually all the meaningful analyses of multidimensional data I have seen have come from experimental insight and scientific theory but, when statistics was indeed helpful, it was generally univariate in character. A definite danger in the indiscriminate employment of a procedure like factor analysis is that it may impede progress in the development of techniques more suited to the problems of the physi-

cal anthropologist. Horst ('65:23) admitted,

"... it is probable that the emphasis on this aspect of factor analysis — namely, its role in classification and taxonomy — has a limited and superficial role and has done much to retard more fundamental applications to the problems of prediction."

Other multivariate procedures

It is not my intention to survey the entire field of multivariate statistical analysis in this paper. I have attempted to concentrate on procedures which are most often used in the anthropometric literature and present "the other side of the story" with respect to their usefulness, applicability and interpretation. However, similar remarks apply to many of the multivariate procedures not considered in detail here. To cite one final example which illustrates quite nicely how questions of interpretation persist throughout much of multivariate analysis, consider the case of canonical correlation analysis (Anderson, '58:288). The mathematical theory behind the technique is quite elegant and applications would appear to be abundant in anthropometry. However, interpretation is quite another matter: As stated by Kendall ('57:82),

"When it comes to the stage of interpretation we shall encounter the same difficulty that we already met in component analysis: of knowing whether our linear functions correspond to anything 'real' or whether they are merely matters of mathematical convenience."

This common thread of problems in the interpretation of certain multivariate analyses was also noted by Dempster ('69:179):

"Canonical correlation analysis has an obvious mathematical appeal; whether or not it is a statistically useful tool is less easily discovered. A similar question was raised when considering principal components analysis . . . the question of meaning and usefulness of the artificial canonical variables remains."

The physical anthropologist usually has an embarrassing profusion of variables and often employs multivariate techniques with the objective of achieving parsimony by introducing a mathematical idealization in an attempt to "reduce the dimensions of the problem." While this approach may in fact occasionally produce the required simplification, it is more usually

the case that the new dimensions defy meaningful interpretation and that no real simplification is realized. These considerations lead to the following question: Is there not something basically wrong with an idealization which *creates* difficult problems rather than serves to avoid them? No one really questions the philosophical foundations of most multivariate techniques nor the mathematics upon which they are based. This does not mean, however, that they are useful in practice, and disciplines which choose to ignore the possibility that mathematical artefacts may not be interpretable in the context of the subject-matter problem leave themselves open to the further question: Is there not something basically wrong with a discipline if difficult problems arise from a supposedly simplifying idealization? Some more general comments concerning the reasons for these difficulties are given in the following section.

Some more general comments

In most anthropological contexts, the main business of statistical data analyses is the *description* and *communication* of the informational content of a body of data. Pictures and graphs play a key role in data analysis but, with multidimensional data, revealing plots of the structure of the raw data cannot easily be made. For example, when three or more variables are involved, there is no convenient graphical representation of a multivariate histogram or empirical cumulative distribution function. Nor can one utilize the examination of residuals (see, e.g., Chapter 3 of Draper and Smith, '66) in order to check on the underlying assumptions, identify outliers, etc. This prevents the investigator from obtaining realistic clues, as he is used to getting in the one or two variable cases, as to what to do, or what models to try (Tukey and Wilk, '65). A case in point is how does one get a feel for whether or not his data are multivariate normal. Marginal normality is certainly not sufficient: There are examples of bivariate distributions having normal marginals where the correlation coefficient can be zero despite a high degree of association between the

variables (Lancaster, '59) and where the correlation coefficient is positive and yet the values of one of the variables decreases as the other increases (Vaswani, '47, '50). Also see Feller ('66:69,99,162). Kempthorne ('66) raised the questions:

"What is the status of goodness-of-fit tests for the multivariate normal distribution? Are there concepts analogous to skewness and kurtosis of univariate distributions, with associated tests of significance? Has anybody since Galton actually looked at a bivariate distribution? How does one look at a five-variate distribution? Although it is true that univariate procedures are dominated by the normal distribution, it was considered relevant, and probably still is, to have classes of non-normal univariate distributions, as exemplified by Pearson's types, but have any classes of non-normal multivariate distributions been developed?"

The present author (Kowalski and Tarter, '69; Kowalski, '70, '71) has attempted to provide some answers to these questions, but they stand a long way from being satisfactorily answered and, until they are, caution must be advised in practice. When the data are multivariate in character, what is available by way of statistical techniques for the assessment of models is very limited from the standpoint of theory and even more so from the standpoint of practice. We know precious little about the operating characteristics of multivariate procedures; they are often based on complex functions of the data which do not provide useful summaries of the informational content of the data; they are often difficult to interpret and communicate; and they do not readily lend themselves to graphical analysis, although Mahalanobis ('70), Wilk and Gnanadesikan ('61, '64), Gnanadesikan and Lee ('70) and Gnanadesikan and Wilk ('70) have made some important contributions in this area.

So there is another side to the story. Most will agree that the field of multivariate statistical analysis is both difficult and important — in theory and in practice. I believe that the current state of the art is such that some consideration should be given to univariate analyses, despite the fact that our data are generally multidimensional in character, and that a concerted effort should be made to restrict the use of formal statistical analyses to situations in which they are

informative, communicable and relatively easy to understand and display. When multivariate analyses do not satisfy these criteria, I question the rationale behind their employment and suggest the use of simpler, data-analysis oriented techniques.

I want to close this discussion by indicating what "the other side of the story" is *not*. It is *not* a signal to physical anthropologists that multivariate analysis is never useful and therefore not worth knowing about. On the contrary, when properly used, multivariate techniques can be of considerable value in certain situations and their scope of applicability promises to expand significantly with future developments in the field. If we are to make reasonable judgments about when to apply these techniques and keep abreast of new developments we must be conversant with "Quotations from an Unfamiliar Bartlett" (Seal, '64:xi). While in order to present "the other side of the story" it was necessary to give views which did not hold much hope for the future of multivariate analysis, these remarks should be tempered in the light of the well-known advantages of multivariate analysis in certain contexts and the expectation that further research will shed additional light on some of the current problems in the field. Both sides of the story have been presented: Papers containing arguments for multivariate analysis were cited in the introduction and questions raised by various investigators — many of whom were responsible for the mathematical development of these methods — were sprinkled liberally throughout the text. In the final analysis the decision of which approach to use in a given situation rests with the individual investigator. In my own view, there is a case to be made for using *both* univariate and multivariate analyses whenever possible. They may not always lead to consistent inferences (*viz.*, Rao's paradox) but even this may be informative of the situation under analysis if it motivates the investigator to go back to the data in an attempt to discover "why" it happened. We should never lose sight of the fact that statistics is a *tool*; and it seems wise to choose the particular tool or tools which are best suited to the job at hand. Nobody ques-

tions the use of multivariate techniques when they can add something to the results of simpler analyses, but everyone *should* question their use when they impede communication among anthropologists and focus attention on mathematical artefacts instead of on biological truths. A handy rule of thumb is suggested by Koons' ('62) observation that, ". . . we can hardly defend our model if we cannot interpret the results it provides."

LITERATURE CITED

- Anderson, T. W. 1951 Classification by multivariate analysis. *Psychometrika*, 16: 31-50.
- 1958 An Introduction to Multivariate Statistical Analysis. John Wiley and Sons, New York.
- Ashton, E. H., M. J. R. Healy and S. Lipton 1957 The descriptive use of discriminant functions in physical anthropology. *Proc. Roy. Soc.*, 146: 552-572.
- Bargmann, R. E. 1970 Interpretation and use of a generalized discriminant function. In: *Essays in Probability and Statistics*. R. C. Bose, I. M. Chakravarti, P. C. Mahalanobis, C. R. Rao and K. J. C. Smith, eds. Univ. North Carolina Press, Chapel Hill, pp. 35-60.
- Barnard, M. M. 1935 The secular variations of skull characters in four series of Egyptian skulls. *Ann. Eugen.*, 6: 352-371.
- Blackith, R. E. 1965 Morphometrics. In: *Theoretical and Mathematical Biology*. T. H. Waterman and H. J. Morowitz, eds. Blaisdell, New York, pp. 225-249.
- Bock, R. D. 1963 Multivariate analysis of variance of repeated measurements. In: *Problems in Measuring Change*. C. W. Harris, ed. Univ. Wisconsin Press, Madison, pp. 85-103.
- Bolshev, L. N. 1969 Cluster analysis. *Bull. Inter. Statist. Inst.*, 43: 409-425.
- Box, G. E. P. 1953 Non-normality and tests on variances. *Biometrika*, 40: 318-335.
- Bronowski, J., and W. M. Long 1951 Statistical methods in anthropology. *Nature*, 168: 794-795.
- Burnaby, T. P. 1966 Growth-invariant discriminant functions and generalized distances. *Biometrics*, 22: 96-110.
- Cochran, W. G. 1964 Comparison of two methods of handling covariates in discriminatory analysis. *Ann. Inst. Statist. Math.*, 16: 43-53.
- Cochran, W. G., and C. I. Bliss 1948 Discriminant functions with covariance. *Ann. Math. Statist.*, 19: 151-176.
- Cureton, E. E. 1939 The principal compulsions of factor-analysts. *Harvard Ed. Rev.* (May): 278-295.
- Danford, M. B., H. M. Hughes and R. C. McNeely 1960 On the analysis of repeated-measurements experiments. *Biometrics*, 16: 547-565.
- Das Gupta, S. 1964 Nonparametric classification rules. *Sankhya, A*, 26: 25-30.
- Dempster, A. P. 1969 Elements of Continuous Multivariate Analysis. Addison-Wesley, Reading.

- Draper, N. R., and H. Smith 1966 Applied Regression Analysis. John Wiley and Sons, New York.
- Dunn, O. J., and P. D. Varady 1966 Probabilities of correct classification in discriminant analysis. *Biometrics*, 22: 908-924.
- Elashoff, J. D., R. M. Elashoff and G. E. Goldman 1967 On the choice of variables in classification problems with dichotomous variables. *Biometrika*, 54: 668-670.
- Elston, R. C., and J. E. Grizzle 1962 Estimation of time-response curves and their confidence bands. *Biometrics*, 18: 148-159.
- Feller, W. 1966 An Introduction to Probability Theory and Its Applications. Vol. II. John Wiley and Sons, New York.
- 1969 Are life scientists overawed by statistics? *Scientific Res.* (Feb.): 24-29.
- Fisher, R. A. 1938 The statistical utilization of multiple measurements. *Ann. Eugen.*, 8: 376-386.
- Fix, E., and J. L. Hodges 1951 Discriminatory Analysis. Proj. Rep. 21-49-004, Nos. 4 and 11. U. S. Air Force School of Aviation Medicine, Randolph Field, San Antonio.
- Geisser, S. 1963 Multivariate analysis of variance for a special covariance case. *J. Am. Statist. Assn.*, 58: 660-669.
- Geisser, S., and S. W. Greenhouse 1958 An extension of Box's results on the use of the F distribution in multivariate analysis. *Ann. Math. Statist.*, 29: 885-891.
- Gilbert, E. S. 1968 On discrimination using qualitative variables. *J. Am. Statist. Assn.*, 63: 1399-1412.
- 1969 The effect of unequal variance-covariance matrices on Fisher's linear discriminant function. *Biometrics*, 25: 505-515.
- Gnanadesikan, R. 1970 S. N. Roy's interests in and contributions to the analysis and design of certain quantitative multiple response experiments. In: *Essays in Probability and Statistics*. R. C. Bose, I. M. Chakravarti, P. C. Mahalanobis, C. R. Rao and K. J. C. Smith, eds. Univ. North Carolina Press, Chapel Hill, pp. 293-310.
- Gnanadesikan, R., and E. T. Lee 1970 Graphical techniques for internal comparisons amongst equal degree of freedom groupings in multi-response experiments. *Biometrika*, 57: 229-237.
- Gnanadesikan, R., and M. B. Wilk 1969 Data analytic methods in multivariate statistical analysis. In: *Multivariate Analysis — II*. P. R. Krishnaiah, ed. Academic Press, New York, pp. 593-638.
- 1970 A probability plotting procedure for general analysis of variance. *J. Roy. Statist. Soc.*, B, 32: 88-101.
- Gould, A. L. 1969 A regression technique for angular variates. *Biometrics*, 25: 683-700.
- Greenhouse, S. W., and S. Geisser 1959 On methods in the analysis of profile data. *Psychometrika*, 24: 95-112.
- Han, C. P. 1968 A note on discrimination in the case of unequal covariance matrices. *Biometrika*, 55: 586-587.
- Harman, H. H. 1967 *Modern Factor Analysis*. Univ. Chicago Press, Chicago.
- Harris, C. W., ed. 1963 *Problems in Measuring Change*. Univ. Wisconsin Press, Madison.
- Harris, J. E. 1965 Craniofacial growth and malocclusion: a multivariate approach to the study of the skeletal contribution to class II malocclusion. *Trans. Europ. Orthodont. Soc.*: 1-17.
- Healy, M. J. R. 1965 Descriptive uses of discriminant functions. In: *Mathematics and Computer Science in Biology and Medicine*. H.M.S.O. London, pp. 93-102.
- 1969 Rao's paradox concerning multivariate tests of significance. *Biometrics*, 25: 411-413.
- Hills, M. 1967 Discrimination and allocation with discrete data. *Appl. Statist.*, 16: 237-250.
- Hirschfeld, W. J. 1970a Time series and exponential smoothing methods applied to the analysis and prediction of growth. *Growth*, 34: 129-143.
- 1970b A comparison of regression with time series-exponential smoothing predictions of craniofacial growth. *Growth*, 34: 431-435.
- Holloway, L. N., and O. J. Dunn 1967 The robustness of Hotelling's T^2 . *J. Am. Statist. Assn.*, 62: 124-136.
- Horst, P. 1963 Multivariate models for evaluating change. In: *Problems in Measuring Change*. C. W. Harris, ed. Univ. Wisconsin Press, Madison, pp. 104-121.
- 1965 *Factor Analysis of Data Matrices*. Holt, Rinehart and Winston, New York.
- Hotelling, H. 1931 The generalization of Student's ratio. *Ann. Math. Statist.*, 2: 360-378.
- 1933 Analysis of a complex of statistical variables into principal components. *J. Ed. Psych.*, 24: 417-441, 498-520.
- 1954 Multivariate analysis. In: *Statistics and Mathematics in Biology*. O. Kempthorne, T. Bancroft, J. Gowen and J. L. Lush, eds. Hafner, New York, pp. 67-80.
- Howells, W. W. 1957 The cranial vault: factors of size and shape. *Am. J. Phys. Anthropol.*, 15: 19-48.
- 1968 Measurement and analysis in physical anthropology. In: *Handbook of Measurement and Assessment in Behavioral Sciences*. D. K. Whitla, ed. Addison-Wesley, Reading, pp. 393-418.
- 1969a The use of multivariate techniques in the study of skeletal populations. *Am. J. Phys. Anthropol.*, 31: 311-314.
- 1969b Criteria for selection of osteometric dimensions. *Am. J. Phys. Anthropol.*, 30: 451-457.
- Isaacson, S. L. 1954 Problems in classifying populations. In: *Statistics and Mathematics in Biology*. O. Kempthorne, T. Bancroft, J. Gowen and J. L. Lush, eds. Hafner, New York, pp. 107-117.
- Ito, K., and W. J. Schull 1964 On the robustness of the T^2_0 test in multivariate analysis of variance when the variance-covariance matrices are not equal. *Biometrika*, 51: 71-82.
- Kempthorne, O. 1966 Multivariate responses in comparative experiments. In: *Multivariate Analysis*. P. R. Krishnaiah, ed. Academic Press, New York, pp. 521-540.
- Kendall, M. G. 1957 *A Course in Multivariate Analysis*. Hafner, New York.

- 1966 Discrimination and classification. In: *Multivariate Analysis*. P. R. Krishnaiah, ed. Academic Press, New York, pp. 165–185.
- Koons, P. B. 1962 Canonical analysis. In: *Computer Applications in the Behavioral Sciences*. H. Boriko, ed. Prentice-Hall, Englewood Cliffs, N. J., pp. 266–279.
- Kowalski, C. J. 1970 The performance of some rough tests for bivariate normality before and after coordinate transformations to normality. *Technometrics*, 12: 517–544.
- 1971 On the effects of non-normality on the distribution of the sample product-moment correlation coefficient. *Appl. Statist.*, in press.
- Kowalski, C. J., and M. E. Tarter 1969 Coordinate transformations to normality and the power of normal tests for independence. *Biometrika*, 56: 139–148.
- Krishnaiah, P. R., ed. 1966 *Multivariate Analysis*. Academic Press, New York.
- Kurczynski, T. W. 1970 Generalized distance and discrete variables. *Biometrics*, 26: 525–534.
- Lachenbruch, P. A. 1966 Discriminant analysis when the initial samples are misclassified. *Technometrics*, 8: 657–662.
- 1967 An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. *Biometrics*, 23: 639–645.
- Lachenbruch, P. A., and M. R. Mickey 1968 Estimation of error rates in discriminant analysis. *Technometrics*, 10: 1–11.
- Lancaster, H. O. 1959 Zero correlation and independence. *Aust. J. Statist.*, 1: 53–56.
- Lavelle, C. B. L., R. M. Flinn, T. D. Foster and M. C. Hamilton 1970 An analysis into age changes of the human dental arch by a multivariate technique. *Am. J. Phys. Anthropol.*, 33: 403–411.
- Linder, A. 1963 Discriminatory analysis with qualitative variables. *Metrika*, 6: 76–83.
- Linhart, H. 1959 Techniques for discriminant analysis with discrete variables. *Metrika*, 2: 138–149.
- Mahalanobis, P. C. 1930 On tests and measures of group divergence. *J. Asiatic Soc. Bengal*, 2: 541–558.
- 1970 Extensions of fractile graphical analysis to higher dimensional data. In: *Essays in Probability and Statistics*. R. C. Bose, I. M. Chakravarti, P. C. Mahalanobis, C. R. Rao and K. J. C. Smith, eds. Univ. North Carolina Press, Chapel Hill, pp. 397–406.
- Morrison, D. F. 1967 *Multivariate Statistical Methods*. McGraw-Hill, New York.
- Mosimann, J. E. 1970 Size allometry: size and shape variables. *J. Am. Statist. Assn.*, 65: 930–945.
- Oxnard, C. E. 1968 Primate evolution—a method of investigation. *Am. J. Phys. Anthropol.*, 28: 289–302.
- Pearson, K. 1901 On lines and planes of closest fit to systems of points in space. *Phil. Mag.*, 2: 559–572.
- 1926 On the coefficient of racial likeness. *Biometrika*, 18: 105–117.
- Penrose, L. S. 1954 Distance, size and shape. *Ann. Eugen.*, 18: 337–343.
- Potthoff, R. F., and S. N. Roy 1964 A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, 51: 313–326.
- Rao, C. R. 1946 Tests with discriminant functions in multivariate analysis. *Sankhya*, 7: 404–414.
- 1948a The utilization of multiple measurements in problems of biological classification. *J. Roy. Statist. Soc., B*, 38: 159–193.
- 1948b Tests of significance in multivariate analysis. *Biometrika*, 35: 58–79.
- 1949 On some problems arising out of discrimination with multiple characters. *Sankhya*, 9: 343–364.
- 1952 *Advanced Statistical Methods in Biological Research*. John Wiley and Sons, New York.
- 1959 Some problems involving linear hypotheses in multivariate analysis. *Biometrika*, 46: 49–58.
- 1961 Some observations on multivariate statistical methods in anthropological research. *Bull. Inst. Internat. Statist.*, 38: 99–111.
- 1962 Use of discriminant and allied functions in multivariate analysis. *Sankhya*, 24: 149–154.
- 1966 Covariance adjustment and related problems in multivariate analysis. In: *Multivariate Analysis*. P. R. Krishnaiah, ed. Academic Press, New York, pp. 87–103.
- 1969 Recent advances in discriminatory analysis. *J. Indian Agric. Statist.*, 21: 3–15.
- 1970 Inference on discriminant function coefficients. In: *Essays in Probability and Statistics*. R. C. Bose, I. M. Chakravarti, P. C. Mahalanobis, C. R. Rao and K. J. C. Smith, eds. Univ. North Carolina Press, Chapel Hill, pp. 587–602.
- Rightmire, G. P. 1970a Iron age skulls from Southern Africa re-assessed by multiple discriminant analysis. *Am. J. Phys. Anthropol.*, 33: 147–167.
- 1970b Bushman, Hottentot and South African Negro crania studied by distance and discrimination. *Am. J. Phys. Anthropol.*, 33: 169–195.
- Robbins, H. 1951 Asymptotically subminimax solutions of compound statistical decision problems. *Proc. Second Berkeley Symp. in Math. Stat. and Prob.* Univ. California Press, Berkeley, pp. 131–148.
- Scheffé, H. 1959 *The Analysis of Variance*. John Wiley and Sons, New York.
- Schull, W. J. 1962 The role of statistics in dentistry. In: *Genetics and Dental Health*. C. J. Witkop, ed. McGraw-Hill, New York, pp. 70–79.
- Seal, H. 1964 *Multivariate Statistical Analysis for Biologists*. Methuen, London.
- Smith, C. A. B. 1947 Some examples of discrimination. *Ann. Eugen.*, 13: 272–282.
- Sokal, R. R. 1961 Distance as a measure of taxonomic similarity. *Syst. Zool.*, 10: 70–79.
- Tallis, G. M. 1970 Some extensions of discriminant function analysis. *Metrika*, 15: 86–91.

- Tukey, J. W. 1962 The future of data analysis. *Ann. Math. Statist.*, 33: 1-67.
- Tukey, J. W., and M. B. Wilk 1965 Data analysis and statistics: Techniques and approaches. *Proc. Symp. Inf. Proc. in Sensory Light Syst.* California Inst. Tech., Pasadena, pp. 7-27.
- Vaswani, S. P. 1947 A pitfall in correlation theory. *Nature*, 160: 405-406.
- 1950 Assumptions underlying the use of the tetrachoric correlation coefficient. *Sankhya*, 10: 269-276.
- Wilk, M. B., and R. Gnanadesikan 1961 Graphical analysis of multiresponse experimental data using ordered distances. *Proc. Nat. Acad. Sci.*, 47: 1209-1212.
- 1964 Graphical methods for internal comparisons in multiresponse experiments. *Ann. Math. Statist.*, 35: 613-631.