ANALYSIS OF A COMPLEX OF STATISTICAL VARIABLES INTO PRINCIPAL COMPONENTS[1]

HAROLD HOTELLING

Columbia University

1. INTRODUCTION

Consider n variables attaching to each individual of a population. These statistical variables $x_1, x_2, \ldots, x_n$ might for example be scores made by school children in tests of speed and skill in solving arithmetical problems or in reading; or they might be various physical properties of telephone poles, or the rates of exchange among various currencies. The x's will ordinarily be correlated. It is natural to ask whether some more fundamental set of independent variables exists, perhaps fewer in number than the x's, which determine the values the x's will take. If $\gamma_1, \gamma_2, \ldots$ are such variables, we shall then have a set of relations of the form

$$x_i = f_i(\gamma_1, \gamma_2, \ldots) \qquad (i = 1, 2, \ldots, n) \tag{1}$$

Quantities such as the γ's have been called mental factors in recent psychological literature. However in view of the prospect of application of these ideas outside of psychology, and the conflicting usage attaching to the word "factor" in mathematics, it will be better simply to call the γ's *components* of the complex depicted by the tests.

We shall consider only normally distributed systems of components having zero correlations and unit variances. If we use the symbol E to denote the expectation, or mean value in the population, of the quantity following it, the condition that the means shall be zero is expressed by

$$E\gamma_i = 0.$$

The assumptions of unit variances and zero correlations may be combined in the statement

[1] A study made in part under the auspices of the Unitary Traits Committee and the Carnegie Corporation.

The author is indebted to Professor Truman L. Kelley, who was responsible for the initiation of this study and the propounding of many of the questions to which answers are here attempted; also to Professors L. L Thurstone, Clark V. Hull, C. Spearman, and E. L. Thorndike, who raised some of the further questions treated.

$$E\gamma_i\gamma_j = \delta_{ij} \tag{2}$$

where δ_{ij}, the so-called Kronecker delta, equals unity if i equals j, zero if they are unequal.

If, following the notation of T. L. Kelley, we express the x's in "standard measures," by taking the deviation of each from its mean value and dividing by its standard deviation, we obtain a set of quantities $z_1, z_2, \ldots, z_n$ for which our formulas will be simpler. Confining ourselves to the case in which the functions f_i are linear, the equations (1) then take the form

$$z_i = \Sigma_j a_{ij}\gamma_j, \tag{3}$$

constant terms disappearing because both the z's and γ's have zero means. The summation will be taken from 1 to n; this will include as special cases situations in which there are fewer components than tests, since some of the a_{ij}'s may be zero. However we shall assume in what immediately follows that this is not the case, and that the determinant a of the a_{ij}'s is not zero.

We shall make use of the tensor analysis convention that the repetition of a literal subscript in a term shall, unless otherwise explicitly indicated, denote summation with respect to that subscript from 1 to n. This not only saves writing a large number of summation signs, but has a mnemonic value in helping to indicate what to do next. According to this convention we write (3) in the form:

$$z_i = a_{ij}\gamma_j. \tag{3}$$

Let A_{ij} denote the cofactor of a_{ij} in a, divided by a. Then by the elementary theory of determinants,

$$a_{ij}A_{ik} = \delta_{jk}, \qquad a_{ij}A_{kj} = \delta_{ik}. \tag{4}$$

We may solve (3) for the γ's by multiplying both sides by A_{ik}, summing with respect to i from 1 to n, and using (4). Since $\delta_{jk}\gamma_j$ is a sum consisting of terms which all vanish except γ_k, this gives:

$$\gamma_k = A_{ik}z_i. \tag{5}$$

Let r_{ik} be the correlation between x_i and x_k, equal to unity if $i = k$. This is the same as the correlation between z_i and z_k; and

$$r_{ik} = Ez_iz_k.$$

Here substitute the value for z_i given by (3), and for z_k an expression obtained from (3) by replacing i by k, and j by l. With the help of (2) we then obtain:

$$r_{ik} = a_{i,}a_{kl}E\gamma_{j}\gamma_{l} = a_{i,}a_{kl}\delta_{jl} \qquad (6)$$
$$= a_{i,}a_{k,}.$$

Since $r_{ik} = r_{ki}$ the number of equations (6) is only $\frac{1}{2}n(n+1)$. They are therefore insufficient for determining the n^2 quantities a_{ij} when the correlations between the tests are known. Thus systems of uncorrelated components γ may be chosen, consistently with the observed correlations, in $\infty^{\frac{1}{2}n(n-1)}$ ways. This variety of choices of components corresponds to the $\frac{1}{2}n(n-1)$ degrees of freedom of a rigid rotation in a space of n dimensions.

It might be thought that additional equations for determining the $a_{i,}$ could be obtained with the moments of higher order, or of other parameters of the population. But if we retain our assumption that the x's are linearly compounded of normally distributed components, this is not the case. Indeed, the x's then have a multivariate normal distribution; and every parameter of such a distribution is a function of the means, variances, and covariances, whose available information is fully embodied in the equations (6) and in the assumption of standard measures. If for example we multiply together four such equations as (3) and take the mean value of each side so as to get an equation in the $a_{i,}$'s, this equation will, with the help of (6) and the expressions for the fourth moments of a multiple normal distribution, reduce to an identity.

Various modes of escape from the indeterminateness have been considered. The number n^2 of unknowns $a_{i,}$ may be reduced by supposing that there are fewer than n components, which amounts to setting some of the $a_{i,}$ equal to zero. If carried far enough, this results in fewer equations than unknowns, so that consistency conditions upon $r_{i,}$ may be obtained. A similar situation arises from other arbitrary specializations of the $a_{i,}$, the number of components possibly even exceeding the number of tests. Thus Spearman, putting (in different notation)

$$z_{i} = a_{io}\gamma_{o} + a_{i,}\gamma_{i}, \qquad (7)$$

obtained as consistency conditions the famous tetrad equations,

$$r_{i,}r_{kl} - r_{ik}r_{jl} = 0. \qquad (8)$$

When these are satisfied for every set of different values of the subscripts, the $a_{i,}$ appearing in (7) are determined uniquely except for sign. Systems involving less specialization and leading to different

and fewer consistency conditions have been considered by Truman L. Kelley in "Crossroads in the Mind of Man."[1]

The consistency conditions are of course never satisfied exactly in a sample. Whether the extent of their non-fulfillment in a sample of given size is sufficient to render incredible their fulfillment in the population depends not only on the standard of credibility adopted, but also on the solution of mathematical problems whose study is still incomplete. Apart from this question of sampling, it may well be argued that it is unlikely that the conditions should be fulfilled exactly in the population, and that for sufficiently large samples tetrads such as the left member of (8) may confidently be expected to exceed any assigned multiple of their probable errors, just as the correlation between any two mental or physical measurements is not likely to be exactly zero. This argument is not necessarily conclusive, since small tetrads, even in the infinite population, may like very small correlations be treated as negligible, for economy of thought. It does, however, bring out the special character of the assumption that the number of components is less than the number of tests, as well as of other simplifying particularisations of the a_{ij}.

In order to go as far as may reasonably be possible in a given case in expressing the test scores x_i in terms of a smaller number of components, an orderly procedure is required for selecting the components in the order of the definiteness of their existence, or of their importance for our purposes, and rejecting any which prove to be of little importance, or which are not clearly defined by the data. An analogous situation arises in fitting empirical curves. A series of the form

$$y = a + bx + cx^2 + \cdots$$

may be fitted, the number of terms used being limited by the increasing probable errors of the coefficients of higher order, and also by the diminishing contributions to the total variance of y by these higher order terms. If the series is modified so as to consist of orthogonal functions, the successive coefficients have zero intercorrelations. Only those terms should be retained which are significant. Another analogy is the use of regression equations involving more and more variables $x_1, x_2, x_3, \ldots$ to explain or predict y, these being chosen in the order of their contributions to the variance of y.

These analogies suggest that, in choosing among the infinity of possible modes of resolution of our variables into components, we

[1] Stanford University Press, 1928.

begin with a component γ_1 whose contributions to the variances of the x_i have as great a total as possible; that we next take a component γ_2, independent of γ_1, whose contribution to the residual variance is as great as possible; and that we proceed in this way to determine the components, not exceeding n in number, and perhaps neglecting those whose contributions to the total variance are small. This we shall call *the method of principal components.* Its technique will be considered in the subsequent sections.

If $z_1, z_2, \ldots z_n$ be taken as rectangular coordinates in n dimensions, each point represents a possible individual. If, as we assume, the population is normally distributed, the loci of uniform density are concentric, similar, and similarly placed ellipsoids. The method of principal components, we shall see, is equivalent to choosing a set of coordinate axes coinciding with the principal axes of these ellipsoids.

Now since the set of x_i is capable of transformations such as changes of units and other linear transformations, the ellipsoids may be squeezed and stretched in any way. The method of principal components can therefore be applied only if for each x_i there exists a unit of measure of unique importance, and if, furthermore, linear transformations, or at least those which do not correspond to rotations of axes, are unimportant. In other words, a *metric*—a definition of *distance*—must be assumed in the n-dimensional space, and not simply a set of axes; we must use Euclidean, not affine geometry, if the principal axes of the ellipsoids are to possess significance. For various purposes it might well happen that different metrics would be suitable. For example the assumption that all the tests, and all the sets of components to be considered, shall have their chance errors independent of those of the others in the set and of equal variance, provides a unique metric. Other possible metrics might be derived from economic considerations, as by requiring that the component traits shall be of equal market value per unit and must not compete with or complement each other. The particular metric implied by the method of principal components is based on the assumption that the unweighted sum of the variances, where the total variance of each test is taken as unity, is the essential quantity to be analyzed.

Weights may in effect be introduced by changing the units of measure, so as to make the standard deviations of the tests no longer unity. The correlations which appear in our subsequent work would in that case be replaced by covariances, and the 1's in the diagonal of the determinant by the variances. However we shall not treat this

obvious generalization, excepting to discuss in Section 11 a possible criterion for suitable weighting. Analysis of the unweighted sum of variances has somewhat the same sort of validity as the use of an unweighted mean of observations when we do not know what the weights should be.

A question bound to arise is whether the ensuing analysis should be applied to the "raw" correlations or to those corrected for attenuation. This is equivalent to the question whether the unit of measure in the n-space is to be the standard deviation of the true or of the observed scores. If the true scores' standard deviations are to be used as units, the analysis must be based on the corrected correlations, with 1's in the diagonal. This seems for some purposes a reasonable procedure, and is exemplified in Section 5, p. 432, to which the reader may now pass directly if he is interested in learning the method rather than in its theory. If on the other hand the standard deviations of the inexact *observed* scores are taken as units, the analysis must be performed upon a matrix having the reliability coefficients in the principal diagonal, with the raw correlations elsewhere. An advantage of this last method is that the relative influence of the more reliable tests upon the results is in general enhanced.

An easily verified property of the method is that the first of our principal components has a greater mean square correlation with the tests than does any other variable; and that among all variables uncorrelated with the first $q - 1$ principal components ($q = 2, 3, \cdots, n$), that having the greatest mean square correlation with the tests is the qth principal component. The argument is similar to that of the next section, and will not be given explicitly.

2. DERIVATION OF THE METHOD

Upon squaring each side of (3) and taking the mean value, it is evident that the variance of z_i may be written

$$a^2_{i1} + a^2_{i2} + \cdots + a^2_{in},$$

and that the first term is correctly described as the contribution of γ_1 to the variance of z_i. The sum of the contributions of γ_1 to the variances of all the z's is

$$S = a^2_{11} + a^2_{21} + \cdots + a^2_{n1},$$

which in our abbreviated notation may be written

$$S = a_{i1}a_{i1}. \tag{9}$$

Subject to (6), which we rewrite

$$a_{i,}a_{kj} = r_{ik}, \tag{6}$$

our present object is to choose the coefficients $a_{i,}$ so as to make S a maximum. To this end we write

$$2T = S - \lambda_{ih}a_{i,}a_{h,},$$

where the $\lambda_{ih}(= \lambda_{hi})$ are Lagrange multipliers. We put

$$\frac{\partial T}{\partial a_{i1}} = a_{i1} - \lambda_{ih}a_{h1} = 0, \tag{10}$$

$$\frac{\partial T}{\partial a_{ij}} = -\lambda_{ih}a_{hj} = 0 \qquad (j \neq 1) \tag{11}$$

These two sets of equations may be combined in the single form

$$\frac{\partial T}{\partial a_{ij}} = \delta_{1j}a_{i1} - \lambda_{ih}a_{hj} = 0 \tag{12}$$

According to (11), the linear equations

$$\lambda_{ih}x_h = 0 \tag{13}$$

have the $n - 1$ solutions

$$\begin{array}{l} a_{12},\ a_{22},\ \ldots,\ a_{n2} \\ a_{13},\ a_{23},\ \ldots,\ a_{n3} \\ \ldots\ldots\ldots\ldots\ldots \\ a_{1n},\ a_{2n},\ \ldots,\ a_{nn} \end{array}$$

These must all be linearly independent, for otherwise the determinant a would vanish, contrary to hypothesis. Hence the rank of the system (13) is 1. Therefore quantities $\alpha_1, \alpha_2, \ldots, \alpha_n, \beta_1, \ldots \beta_n$, can be found such that $\lambda_{ih} = \alpha_i\beta_h$. But since $\lambda_{ij} = \lambda_{ji}$, it follows that $\alpha_i\beta_h = \alpha_h\beta_i$, so that the β's are proportional to the α's. Hence we put $\beta_i = \epsilon\alpha_i$, and

$$\lambda_{ih} = \epsilon\alpha_i\alpha_h \tag{14}$$

Consequently (10) may be written in the form

$$a_{i1} = \epsilon\alpha_i\alpha_h a_{h1};$$

or, putting

$$\alpha_h a_{h1} = \sqrt{\frac{k}{\epsilon}},$$

in the form

$$\alpha_i = \frac{a_{i1}}{\sqrt{k\epsilon}}.$$

Substituting this in (14) and the result in (12) gives

$$a_{i1}(k\delta_{1j} - a_{h1}a_{hj}) = 0.$$

Since the quantities a_{i1} are not all to be zero, this leads to

$$a_{h1}a_{hj} - k\delta_{1j} = 0 \tag{15}$$

Multiply by a_{mj}, sum for j, and use (6). The result is:

$$r_{hm}a_{h1} - ka_{m1} = 0. \tag{16}$$

Writing out these equations explicitly for $m = 1, 2, \cdots, n$, and dropping the subscript 1 from the unknowns a_{h1}:

$$\begin{aligned}
(1 - k)a_1 + r_{12}a_2 + \cdots + r_{1n}a_n &= 0 \\
r_{12}a_1 + (1 - k)a_2 + \cdots + r_{2n}a_n &= 0 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\
r_{1n}a_1 + r_{2n}a_2 + \cdots + (1 - k)a_n &= 0
\end{aligned} \tag{17}$$

In order that these equations have solutions in which not all unknowns are zero, it is necessary and sufficient that

$$f(k) = \begin{vmatrix} 1 - k & r_{12} & r_{13} \cdots r_{1n} \\ r_{21} & 1 - k & r_{23} \cdots r_{2n} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots\cdots \\ r_{n1} & r_{n2} & r_{n3} \cdots 1 - k \end{vmatrix} = 0 \tag{18}$$

Equations such as (18) were first studied in connection with the perturbations of the planets, and are known as *characteristic equations*. The general theory of such equations[1] shows that all the roots are real, and that if there is a q-fold multiple root, the determinant in (18) has rank $n - q$ when this root is substituted for k.

If we substitute a simple (not a multiple) root of the characteristic equation for k in (17), we obtain a set of linear homogeneous equations of rank $n - 1$. These will have a family of solutions, all of which are proportional to one solution. The factor of proportionality is found by putting $j = 1$ in (15); this shows that the sum of squares, S, which we are trying to maximize, is just equal to one of the roots of the characteristic equation. Since it is to be a maximum, we must obviously use the greatest root.

The problem of finding a component γ_1 which will account for as large as possible a part of the total variance is thus solved by finding

[1] *Cf.* for example G. Kowalewski, Einfuhrung in die Determinantentheorie, 1st ed., Leipzig, 1909, pp. 126 and 274. The proof of the reality of the roots was given by Cauchy in the Philosophical Magazine for 1852, and may be found in the treatment of quadric surfaces in books on solid analytic geometry.

the largest root k_1 of the characteristic equation (18), substituting in (17), finding any solution $a_1, a_2, \ldots, a_n$ of these linear equations, and dividing these last values by the sum of their squares and multiplying by $\sqrt{k_1}$. The resulting quantities are the coefficients of γ_1 in the expression (3) giving the test scores in terms of the independent components. A simplified numerical method is given in §4 below.

If the q largest roots of the characteristic equation are equal to k_1, the determination of $a_{11}, a_{21}, \ldots, a_{n1}$ is not unique, since the rank of the linear equations (17) is then only $n-q$. In this case q linearly independent solutions,

$$\begin{aligned} &a_{11}, a_{21}, \ldots, a_{n1} \\ &a_{12}, a_{22}, \ldots, a_{n2} \\ &\cdots\cdots\cdots\cdots \\ &a_{1q}, a_{2q}, \ldots, a_{nq}, \end{aligned}$$

of (17), can be found. Moreover these solutions may be so chosen as to be "orthogonal" to each other, in the sense that

$$a_{i,}a_{i,m} = \delta_{,m}k_1.$$

They may then be taken as the coefficients of q independent components $\gamma_1, \ldots, \gamma_q$, all of which contribute equally to the total variance.

When the coefficients of the component γ_1 which makes the largest contribution to the total variance, or of the q components which equally make maximum contributions, have been determined, the next problem is to find a component making a maximum contribution to the residual portion of the variance. The argument and procedure are virtually the same as before. If just one component has previously been determined, the subscript i in (10), (11), (12), and (13) takes the values $2, 3, \ldots, n$, and the subscript 1 is replaced by 2. Proceeding in this way we determine the coefficients of $\gamma_1, \gamma_2, \gamma_3, \ldots, \gamma_n$, in the order of the contributions of these components to the sum of the variances of the z's.

The orthogonality among principal components which we have postulated for multiple roots holds also for simple roots. This follows from (15), which, for $j = 2$, gives $a_{h1}a_{h2} = 0$; and similarly we have for all unequal values of i and j, $a_{h,}a_{h,} = 0$.

Having thus derived the equations for the z's in terms of the γ's, it is desirable to solve these, so as to be able to assign a value to each of these principal components γ in terms of the test scores. This

will make it possible to assign a value for each principal component to any individual upon whom the tests had been made.

The solution is remarkably simple.

In (15), k is the same as k_1. For the ith component we replace the subscript 1 in (15) by i, and k by k_i. This gives

$$a_{hi}a_{hj} = k_i\delta_{ij}, \qquad \text{(not summed for } i\text{)} \tag{19}$$

which simply means that the sum of the products of corresponding elements of two different columns of the determinant a is zero, and that the sum of the squares of the elements of a column is the root of the characteristic equation corresponding to this column.

Recalling that we have denoted the ratio of the cofactor of a_{mj} to a by A_{mj}, we multiply both sides of (19) by A_{mj}, and sum for j. With the help of (4) this gives

$$a_{mi} = k_iA_{mi} \qquad \text{(not summed for } i\text{)} \tag{20}$$

Since (5) gives by a mere change of indices,

$$\gamma_i = A_{mi}z_m,$$

it follows that

$$\gamma_i = \frac{a_{mi}z_m}{k_i} \qquad \text{(not summed for } i\text{)}. \tag{21}$$

3. GEOMETRICAL MEANING

Geometrically the foregoing procedure corresponds to rotating the rectangular axes of $z_1, z_2, \ldots, z_n$ so that the new coordinate axes lie along the principal axes of the ellipsoids of uniform density. The squares of the lengths of the principal axes of one of these ellipsoids are proportional to the k's. These facts are not immediately obvious, but may be easily proved as follows. Let ω be the determinant of the correlation coefficients r_{ij}, and let

$$R_{ji} = R_{ij} = \frac{\text{cofactor of } r_{ij} \text{ in } \omega}{\omega},$$

so that

$$r_{ij}R_{ik} = \delta_{jk}. \tag{22}$$

From the theory of multiple normal distribution, the ellipsoids of uniform density are given by

$$R_{ij}z_iz_j = \text{constant}. \tag{23}$$

The procedure developed in treatises on solid analytic geometry is to solve the equation

$$\begin{vmatrix} R_{11} - \lambda & R_{12} & \cdots & R_{1n} \\ R_{21} & R_{22} - \lambda & \cdots & R_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ R_{n1} & R_{n2} & & R_{nn} - \lambda \end{vmatrix} = 0, \tag{24}$$

to substitute the roots $\lambda_1, \lambda_2, \ldots, \lambda_n$ in the homogeneous linear equations

$$(R_{ij} - \lambda\delta_{ij})l_j = 0, \tag{25}$$

and for each root to solve these equations. Calling the solution corresponding to λ_1

$$l_{11}, l_{21}, \ldots, l_{n1},$$

that corresponding to λ_2

$$l_{12}, l_{22}, \ldots, l_{n2},$$

and so on, the equations of rotation to new rectangular axes $y_1, \ldots, y_n$ are

$$z_i = l_{ij}y_j; \tag{26}$$

the solution of these equations for the y_j has the same coefficients in transposed order, and runs:

$$y_j = l_{ij}z_i.$$

The equation of the ellipsoid in the new coordinates is

$$\lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 = \text{constant} \tag{27}$$

Now multiply both sides of (24) by

$$\omega = \begin{vmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1n} \\ r_{21} & 1 & r_{23} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & r_{n3} & \cdots & 1 \end{vmatrix}$$

By (22), the result is

$$\begin{vmatrix} 1 - \lambda & -\lambda r_{12} & \cdots & -\lambda r_{1n} \\ -\lambda r_{21} & 1 - \lambda & \cdots & -\lambda r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ -\lambda r_{n1} & -\lambda r_{n2} & \cdots & 1 - \lambda \end{vmatrix} = 0$$

Upon dividing each row by $-\lambda$, and setting $k = 1/\lambda$, this reduces to the characteristic equation (18). If we also set $k = 1/\lambda$ in (25), multiply by r_{ih}, and sum for i, the resulting equations have the same coefficients as (17). For simple roots the solutions are therefore the same, apart from a common factor. For multiple roots the solutions

have in both cases the same type of indeterminateness. Since $\lambda_i = 1/k_i$, (27) becomes

$$\frac{y_1^2}{k_1} + \frac{y_2^2}{k_2} + \cdots + \frac{y_n^2}{k_n} = \text{constant}, \tag{28}$$

which shows that the squares of the lengths of the axes are proportional to $k_1, k_2, \ldots, k_n$.

If, instead of the y's or z's, the γ's or other independent and equally variable quantities be taken as rectangular coordinates, the ellipsoids are squeezed and stretched into spheres. Each test is represented by a line through the origin. The correlation between two tests is the cosine of the angle between their lines. L. L. Thurstone has used coordinates equivalent to these.[1]

Not only must the roots of the characteristic equation be real; they must all be positive. If there were a negative root, (28) would represent, not an ellipsoid, but a hyperboloid extending to infinity. Since the density of probability is to be uniform over this locus, the probability of a sample deviating in certain tests from the mean by more than any given amount would be infinite, which is absurd.

According to the original Spearman theory of the mind, one important general factor accounts for the bulk of the variance of mental tests, other components being of minor importance. If this is true, one of the roots of the characteristic equation should be much larger than any of the others. The ellipsoids should be needle-shaped. But if two or more of the roots are equal, there will be a corresponding number of independent components which contribute equally to the variance. In this case the ellipsoids will be figures of revolution. If $n = 3$ and the two largest roots are equal, while the third is very small, the ellipsoids will be thin discs.

In distinguishing among such theories it is of course not the absolute values of the roots that is important, but their ratios, and particularly the ratios among the largest of the roots. The sum of

[1] Multiple Factor Analysis. *Psychological Review*, Vol. XXXVIII, 1931, pp. 406–427.

Since this was written Professor Thurstone has kindly sent me a pamphlet he has prepared for class use, in which he uses the same geometric interpretation as in the present section, and discusses the problem from essentially the same standpoint as that taken in §1. His iterative procedure appears to have no relation to that of §4. In June, 1932, Professor Thurstone presented at the Syracuse meeting of the American Association for the Advancement of Science certain of the considerations which have served as a point of departure for this paper.

the roots always equals the number n of the tests, as appears from the form of (18); hence the fraction of the total variance contributed by the ith component is k_i/n.

Developing (18) we have

$$f(k) = (-1)^n(k^n - nk^{n-1} + S_2k^{n-2} - S_3k^{n-3} + \cdots + S_n) = 0, \quad (29)$$

where S_2 is the sum of the two-rowed principal minors in the determinant ω of the correlations, S_3 the sum of the three-rowed principal minors, and so on.

4. ITERATIVE SOLUTION

The explicit calculation of the determinant ω and its principal minors, and the solution of the characteristic equation and of the homogeneous linear equations (17), would be a laborious computation. A vast saving of arithmetical effort is effected by the following iterative method, which yields simultaneously a root and the corresponding coefficients, the roots appearing in order of magnitude, the greatest first. This makes it possible to stop whenever it is evident that all the important principal components have been obtained. Since the sum of the roots is n, the fraction of the total accounted for at any stage is always in evidence. During the calculation of each principal component, an error at any stage is rectified in the next. The risk of serious numerical error is therefore negligible, especially if all the roots are calculated and their sum compared with n.

If numbers $a_1, a_2, \ldots a_n$ proportional to the direction cosines of any line through the origin be substituted in the equations

$$a_i' = r_{i,}a, \qquad (i = 1, 2, \cdots, n) \quad (30)$$

the quantities on the left will be proportional to the direction cosines of a new line through the origin into which we shall consider the original line to have moved. Under this transformation, the invariant lines will be those for which the quantities k exist such that $a_i' = ka_i$. In this case, (30) reduces to (17). Thus, for each invariant line, the direction cosines are proportional to a solution of (17), while k is a root of the characteristic equation. It follows that the invariant lines are the principal axes. Hence, if numbers $a_1, \ldots, a_n$ can be found which, substituted in the right-hand members of (30), give $ka_1, \ldots, ka_n$, these numbers are proportional to the direction cosines of one of the principal axes and to the coefficients of one of

the principal components γ in the expressions for the test scores z_i, while k is the sum of the contributions of this component to the variances of the test scores.

If two or more roots of the characteristic equation are equal, the ellipsoids are figures of revolution, with two or more equal axes. Every line in the plane or hyperplane of the equal axes is invariant under the transformation; the axes themselves may be taken as arbitrary perpendicular lines in this plane or hyperplane.

If with respect to new coordinate axes coinciding with the principal axes of the ellipsoids the direction cosines of a line are proportional to $b_1, b_2, \ldots, b_n$, the transformation

$$b_1' = k_1 b_1,\ b_2' = k_2 b_2,\ \cdots,\ b_n' = k_n b_n, \tag{31}$$

where $k_1, k_2, \ldots, k_n$ are as before the roots of the characteristic equation, will geometrically be the same as (30), since the invariant lines are the principal axes, together perhaps with the lines in the plane or hyperplane determined by two or more equal axes. Algebraically, this amounts to putting.

$$a_i' = a_{ih} b_h', \qquad a_j = a_{jm} b_m.$$

Substituting these expressions in (30) and in the result setting

$$r_{ij} a_{jl} = k_l a_{il} \qquad \text{(not summed for } l\text{)},$$

a relation which is the generalization of (16), we multiply by A_{im}, sum for i, and use (4). The result is (31).

Let the notation be so arranged that

$$k_1 \geqslant k_2 \geqslant \cdots \geqslant k_n.$$

If k_1 is greater than k_2, and if $b_1 \neq 0$, we then have from (31) that each of the ratios $b_2'/b_1', \ldots, b_n'/b_1'$ is numerically less than the corresponding ratios $b_2/b_1, \ldots, b_n/b_1$. When the transformation is repeated, the absolute values of these ratios are further diminished, and with further repetitions approach zero in geometrical progressions. If, however, the q greatest roots are equal, the ratios among the first q direction cosines remain unchanged under the transformation, while the remaining $p - q$ direction cosines approach zero. Thus if we start with any line which does not lie in the hyperplane of the $p - 1$ axes perpendicular to the longest axis, this line will, under iteration of the transformation, approach the longest axis if the greatest root is unique, or, if there are several equal roots greater than the rest, some position which may be taken as the longest axis.

From this it is evident that the coefficients of the greatest principal component of the tests may be obtained with any required accuracy by inserting an arbitrary set of numbers $a_1, \ldots, a_n$ in the right members of (30), multiplying or dividing the resulting numbers $a_1', \ldots, a_n'$ by any constant, again substituting in the right members of (30), and repeating the process until, to the required degree of accuracy, the quantities obtained are multiples of the preceding quantities by a constant k_1. This constant will be the greatest root of the characteristic equation. The process fails only in the infinitely improbable case of the initial values being linearly dependent upon principal components other than the first; the greatest of these components is then approached.

The coefficients $a_{11}, a_{21}, \ldots, a_{n1}$ of the first principal component are found by multiplying each of the quantities $a_1, a_2, \ldots, a_n$ by $\sqrt{k_1/\Sigma a_j^2}$.

After the coefficients of γ_1 have been determined, those of γ_2 are sought. This might be done by starting with arbitrary values orthogonal to the coefficients of γ_1, which would give the desired result if the calculations were carried out exactly. But on account of the practical necessity of working only to a limited number of decimal places, the transformed line will at each stage deviate from the plane perpendicular to the first axis, and will drift off toward this first axis. The tendency could be offset by applying corrections to restore orthogonality each time, but this is excessively laborious. Instead, we revise the matrix of correlations by subtracting $a_{i1}a_{j1}$ from the element in the ith row and jth column. This produces the matrix of covariances of the quantities $z_i - a_{i1}\gamma_1$, which are uncorrelated with γ_1. The iterative process applied to the reduced matrix yields k_2 and the coefficients of γ_2. To obtain γ_3 we apply the iterative process to the further reduced matrix in which the element in the ith row and jth column is $r_{ij} - a_{i1}a_{j1} - a_{i2}a_{j2}$; and so on. When this method is used, it is not even essential to take the initial trial values of the coefficients at any of the later stages precisely orthogonal to the coefficients already determined; round numbers may be chosen at the beginning, and the process will converge to the correct values anyhow.

A convenient procedure is to divide each of the trial values of any set of coefficients by a fixed one of them. The next value obtained for this coefficient will then be an approximation to the appropriate characteristic number k. The process should be started with trial values of one digit each, the largest of these values corresponding to

variates which are on the whole most highly correlated with the rest, as judged by inspection. Each digit should be accurately determined, by repetition until stationary values are reached, before the calculations are carried to another place.

The labor is sometimes reduced if the deviations of a set of trial a's from the preceding set are multiplied by the rows of the correlation matrix and the results added to the last trial a's to get the next set, instead of multiplying the trial values themselves by the correlations. However, this method does not automatically correct errors unless the values obtained finally are multiplied by the correlations and added.

5. EXAMPLE

Truman L. Kelly (*op. cit.*, p. 100) gives the correlations found in a sample of 140 seventh-grade children among numerous tests. We select the correlations, corrected for attenuation, among: (1) Reading speed, (2) reading power, (3) arithmetic speed, (4) arithmetic power. Curtailed to three places, these are, in the natural order:

$$\begin{vmatrix} 1. & .698 & .264 & .081 \\ .698 & 1. & -.061 & .092 \\ .264 & -.061 & 1. & .594 \\ .081 & .092 & .594 & 1. \end{vmatrix}$$

The correlations being slightly higher on the whole for the first than for the later tests, we take as trial values

$$1, \quad .9, \quad .8, \quad .7.$$

Multiplying by the rows of the matrix of correlations, we have, to two place, 1.90, 1.61, 1.42, 1.33. If we divide each of these values by the first in order to make them comparable with the initial quantities, we obtain as our second approximation:

$$1, \quad .85, \quad .75, \quad .70.$$

Multiplying these by the correlations, we obtain 1.85, 1.57, 1.37, and 1.30. Dividing these by 1.85 gives

$$1, \quad .85, \quad .74, \quad .70.$$

This is close enough to warrant carrying the calculations to one more decimal place. We next obtain 1.846, 1.567, 1.368, 1.304; and upon division by 1.846,

$$1, \quad .849, \quad .741, \quad .707$$

The next trial gives, after division by 1.846,

$$1, \quad .849, \quad .743, \quad .706,$$

values which we adopt after noting that the differences,

$$0 \quad 0 \quad .002, \quad -.001,$$

multiplied by any row of the correlation matrix, produce a total which does not affect the third decimal place. We divide $k_1 = 1.846$ by the sum of the squares of the final trial values, extract the square root, and multiply by the trial values. This gives

$$\begin{aligned} a_{11} &= .816 \qquad k_1 = 1.846 \\ a_{21} &= .693 \\ a_{31} &= .606 \\ a_{41} &= .576 \end{aligned}$$

Subtracting the products $a_{i1}a_{j1}$ from the elements of the correlation matrix we obtain

$$\begin{vmatrix} .330 & .129 & -.233 & -.392 \\ .129 & .517 & -.484 & -.310 \\ -.233 & -.484 & .631 & .243 \\ -.392 & -.310 & .243 & .666 \end{vmatrix}$$

Starting from the trial values $-.6, -1, 1, 1$, we obtain for the second principal component:

$$\begin{aligned} a_{12} &= -.438 \qquad k_2 = 1.465 \\ a_{22} &= -.620 \\ a_{32} &= .674 \\ a_{42} &= .660 \end{aligned}$$

Subtracting the products of these numbers from the reduced matrix we find:

$$\begin{vmatrix} .138 & -.142 & .062 & -.103 \\ -.142 & .133 & -066 & .099 \\ .062 & -.066 & .177 & -202 \\ -.103 & .099 & -.202 & .230 \end{vmatrix}$$

This time we take the trial values $-1, 1, -1, 1.3$, and obtain another root, another set of coefficients, and another reduced matrix. A fourth application of the iterative process completes the resolution of the tests into their principal components. The results are combined in the table below, in which each column corresponds to a principal component, and each of the last four rows to a test. The entries in

the last four rows are the coefficients of the γ's in the expressions for the z's; they are at the same time the correlations of the γ's with the z's.

					Totals
Root	1 846	1 465	521	167	3.999
Percentage of total variance	46½	36½	13	4	100
Reading speed	818	−.438	−.292	.240	
Reading power	.695	− 620	.288	− 229	
Arithmetic speed	608	.674	−.376	−.193	
Arithmetic power	578	660	.459	.143	

The chief component seems to measure general ability; the second, a difference between arithmetical and verbal ability. These two account for eighty-three per cent of the variance. An additional thirteen per cent seems to be largely a matter of speed vs. deliberation. The remaining variance is trivial.

6. SAMPLING ERRORS

The exact distribution of the roots of the characteristic equation, or of their ratio, can be found at once when $n = 2$. Indeed, we have in this case,

$$k_1 = 1 + r, \qquad k_2 = 1 - r;$$

and the distribution of r in samples of N from a normally correlated population is fully known.

For the case of zero correlation in the population, the distribution of r reduces to

$$\frac{1}{\sqrt{\pi}} \frac{\Gamma[\frac{1}{2}(N-1)]}{\Gamma[\frac{1}{2}(N-2)]} (1 - r^2)^{\frac{N-4}{2}} dr.$$

If we put

$$u = \frac{k_1}{k_2} = \frac{1+r}{1-r}, \text{ so that } r = \frac{k_1 - k_2}{k_1 + k_2},$$

this is transformed into

$$\frac{2}{\sqrt{\pi}} \frac{\Gamma[\frac{1}{2}(N-1)]}{\Gamma[\frac{1}{2}(N-2)]} \frac{(4u)^{\frac{N-4}{2}}}{(u+1)^{N-2}} du.$$

For $n > 2$, this same distribution may be used as a close approximation for differentiating between any two roots.

Thus, to determine whether k_1 is significantly greater than k_2 in the example worked out in the last section, we compute

$$r = \frac{1.846 - 1.465}{1.846 + 1.465} = .115$$

Since the sample consists of one hundred forty individuals, we may treat this value of r as a sample from a normal distribution of zero mean and standard deviation $1/\sqrt{139} = .085$. Since r is only 1.35 times its standard error, the probability of a greater discrepancy is about .18, and the two roots cannot be called significantly different. The fact that a single component accounts for so much as 46½ per cent of the variance of the four tests tends to support the idea introduced by Spearman that one general factor enters into all tests to a dominating extent; but this argument is considerably weakened by the fact that γ_2 contributes nearly as much variance as γ_1, the magnitudes of the two contributions being indeed not clearly distinguishable in a sample of this size. The contribution of γ_3 is however definitely less than that of γ_2; for

$$r = \frac{1.465 - .521}{1.465 + .521} = .474$$

is some 5.58 times the standard error .085. The third characteristic root exceeds the fourth even more definitely, the ratio of the corresponding value of r to its standard error being 6.1.

Apart from the question of equality of any two roots, it may sometimes be desired to find upper and lower limits such that, corresponding to any given degree of probability, the ratio of the roots may be said to lie between these limits. Such limits may be deduced from the corresponding ones for the correlation coefficient. To obtain these, we may utilize R. A. Fisher's transformation to $z = \tanh^{-1} r$, whose distribution is nearly normal with a variance, $1/(N - 3)$, independent of the population value.[1]

As an example, let us find upper and lower fiduciary limits corresponding to the probability .05 for the ratio k_2/k_3, which has just been seen to differ clearly from unity. From Fisher's table, we find that the value $r = .474$ corresponds to $z = .515$. The table of the normal probability integral shows that a quantity deviates from its mean more than 1.96 times its standard deviation with a probability .05. The deviation

$$1.96\sigma_z = \frac{1.96}{\sqrt{137}} = .117$$

[1] Statistical Methods for Research Workers, Oliver and Boyd, Chap. VI.

is therefore to be added to and subtracted from the sample value .515, giving .398 and .632 as the fiduciary limits for z. Again referring to the table of hyperbolic tangents, we find .378 and .559 as the corresponding values of r. Inserting each of these limits for r in the expression defining u, we have 2.21 and 3.53 as the extreme values of the ratio which can plausibly be assumed, corresponding to the probability .05.

This use of the sampling distribution of r is subject to two qualifications. In the first place, it applies only in comparing two components which are definitely identified, otherwise than by their having a particular order among the entire set of n components determined, such as being the two greatest, or the greatest and least. This situation is common to all problems in which a number of observations are made on a quantity, and two of these observations are examined for the significance of their difference from each other. If the two are selected because of their positions relatively to the others in the sample, they cannot be compared accurately by means of the standard error of the distribution as if they were not so selected. In many examples, however, including the foregoing, this consideration does not materially affect the conclusions to be drawn.

It must also be remembered that the directions as well as the magnitudes of the principal components are subject to sampling errors, and that these errors in determination of the directions will interfere with the accuracy of the foregoing treatment of the k's by transformation into correlation coefficients. That the inaccuracy introduced in this way is very slight is suggested by the geometry. Suppose that a principal plane of an ellipsoid derived from a sample makes a small angle θ with the corresponding principal plane of the population ellipsoid. Then the section of the population ellipsoid made by the sample plane will be an ellipse whose principal axes bear to the corresponding principal axes of the population ellipsoid ratios differing from unity by quantities of order θ^2. Hence if the standard error of θ, like most standard errors, is of order $1/\sqrt{N}$, those of the semiaxes, and consequently those of the k's and of the r's calculated from them, will be of order $1/N$. This suggests that a suitable correction for this kind of error could be made in the foregoing example by changing the standard error used, namely .085, by something of the order of $\frac{1}{140} = .007$.

Essentially the same results may be reached in another way. The k's are, in the population, variances of independent variates. The ratio u of two independent estimates of variance, each based on $N - 2$

degrees of freedom, as for example in independent samples of $N - 1$, has exactly the distribution of u given above. This way of looking at the matter has the further advantage of showing that, for very large samples, $\log k$ may be treated approximately as a normally distributed variate with variance $1/(2N - 4)$. The n values of $\log k$ may then be treated as independent samples from a normal distribution having this variance.

7. MINIMUM NUMBER OF INDEPENDENT COMPONENTS. RELIABILITY COEFFICIENTS

But of still greater importance is the fact that by treating the k's as estimates of variance in n orthogonal directions we may compare them, not only with each other, but also with the variance to be expected on account of the inaccuracy of the tests as revealed by their self-correlations, or reliability coefficients. This is of major interest, for if some of the principal components found contribute so little to the variance that their reality is in doubt, it is possible to assume that the number of independent components is less than the number of variables measured. The following tests may therefore serve as substitutes both for Spearman's use of tetrads and for the more elaborate criteria developed by Kelley in "Crossroads in the Mind of Man."

Upon administering a test twice to the same individuals a measure of its accuracy may be obtained from the correlation of the two sets of scores. This correlation is known as the reliability coefficient; for the ith test we shall denote it by r_{ii}. If the test score is thought of as made up of two parts, a true score, whose variance we shall take as unity, and a random error of variance σ_i^2, it is easy to see that

$$r_{ii} = \frac{1}{1 + \sigma_i^2},$$

whence σ_i^2 may be determined from the data. The random errors are supposed to be independent in the several tests. For the four tests which we have been using as an example the reliability coefficients given by Kelley are, in order

$$.9197, \quad .8942, \quad .9083, \quad .5639.$$

Hence we find

$$\sigma_1^2 = .0873 \quad \sigma_2^2 = .1183 \quad \sigma_3^2 = .1010 \quad \sigma_4^2 = .7734.$$

Since the covariance of the ith and jth tests is the same whether based on the true or the total scores, the correlations will differ in the

two cases, that based on the true score being larger in the ratio $\sqrt{(1 + \sigma_i^2)(1 + \sigma_j^2)}$. If r'_{ij} is the correlation between the observed scores, the correlation between true scores is estimated as

$$r_{ij} = r'_{ij}\sqrt{r_i r_j} = r'_{ij}\sqrt{(1 + \sigma^2_i)(1 + \sigma^2_j)} \qquad (i \neq j; \text{ not summed for } i \text{ or } j), \quad (32)$$

and is known as a correlation corrected for attenuation. The exact sampling distribution of this quantity has never been determined, but for sufficiently large samples it may be used with confidence, subject to the validity of the assumption that the errors of measurement in the several tests are uncorrelated with the test scores and with each other. Since we wish to deal with real quantities as far as possible, we have based our analysis into principal components upon these corrected coefficients r_{ij}.

If the number of independent components of the n true scores is less than n, the scatter diagram of the true scores will lie in a flat space of smaller dimensionality immersed in the n-dimensional space. The scatter diagram of the observed scores will however be n-dimensional in character, since the scatter diagram of the errors of measurement is n-dimensional. The scatter diagram corresponding to the correlations corrected for attenuation would in this case be of the smaller dimensionality if calculated from the whole population; but on account of the fluctuations of the errors of measurement this would not in general be true for samples. We are therefore interested in comparing the variances of the principal components we find, and particularly the least of these variances, with those to be expected on the basis of the reliability coefficients.

Now by (21), we have upon omitting the subscript i,

$$\gamma = \frac{a_1 z_1 + a_2 z_2 + \cdots + a_n z_n}{k}. \quad (33)$$

The variance of this expression which results from errors of measurement is

$$\sigma'^2 = \frac{a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2}{k^2}.$$

On the hypothesis that γ has no real existence, but arises purely from errors of measurement, its variance, which we have taken as unity, should differ from σ'^2 only on account of fluctuations in these errors. On this hypothesis the equation

$$k = \bar{k}$$

where

$$\bar{k} = \sqrt{a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2},$$

should fail to be satisfied only in so far as k and $\bar{k}$ are affected by random sampling errors. These two quantities appear to be uncorrelated with each other; for although the k's are derived from correlation coefficients corrected for attenuation with the help of the data from which the $\bar{k}$'s are deduced, still this process is analogous to that in Fisher's analysis of variance of subtracting from the total variance the intraclass variance before comparing with the interclass variance, which is independent of this difference. If we rely upon this analogy, we may compare k with $\bar{k}$ simply by adding their variances, provided the samples are large enough to allow the difference to be treated as normally distributed.

Instead of comparing k with $\bar{k}$ directly, we might use any function of k and the corresponding function of $\bar{k}$. In choosing among such functions as k, k^2, $\sqrt{k}$, and $\log k$, it is to be recalled that k is of the nature of a variance, as pointed out at the end of the last section. In comparing estimates of variance, the logarithm is used by R. A. Fisher because its standard error is independent of the value of the variance in the population, and depends only on the number of cases in the sample—or, more generally, upon the number of degrees of freedom on which the estimate of variance is based. Against this advantage must be set the fact that the square root of an estimate of variance has a more nearly normal distribution than either the estimate of variance itself or its logarithm, and the use of the standard error presupposes a normal distribution. Further, $\log \bar{k}$ has a standard error which may be found approximately from the definition of $\bar{k}$ above, and which lacks the property of depending only on the number of cases, the a's and σ's being involved in it also. The advantage of the logarithm is thus lost when a comparison is made with $\bar{k}$, but the gain in accuracy in the use of $\sqrt{k}$ persists, though the higher moments of the various functions of $\bar{k}$ have not yet been investigated. Consequently we shall make our comparisons in terms of the square roots.

Since

$$\sqrt{\bar{k}} = (\Sigma a_i^2\sigma_i^2)^{1/4},$$

we have, apart from terms of higher order, the following relation between deviations of sample from population values:

$$\delta\sqrt{\bar{k}} = \tfrac{1}{2}(\Sigma a_i^2\sigma_i^2)^{-3/4}(\Sigma a_i^2\sigma_i\delta\sigma_i).$$

The mean value of this expression is zero, provided the estimate of σ_j^2 is without bias. This will be the case if σ_j^2 is calculated as the ratio of the sum of the squares of the differences between test and retest to $2N$; this method of calculation appears to be approximately equivalent to that from reliability coefficients. Neglecting any bias, then, the variance of $\sqrt{k}$ will be the mathematical expectation of the square of the above expression. Since the errors of measurement of the tests are independent of each other,

$$E\delta(\sigma_i)\delta(\sigma_j) = 0, \text{ if } i \neq j,$$

while the usual formula for the variance of the standard deviation gives

$$E(\delta\sigma_j)^2 = \frac{\sigma_j^2}{2N}.$$

Making these substitutions after squaring $\delta\sqrt{k}$, we obtain finally,

$$\sigma^2_{\sqrt{k}} = \frac{\Sigma a_j^4 \sigma_j^4}{8Nk^3}.$$

To this we add in each case

$$\sigma^2_{\sqrt{k}} = \frac{k}{2N}$$

to obtain the variance of $\sqrt{k} - \sqrt{\bar{k}}$. The results for the four tests based on one hundred forty cases which we have been considering are given in the following table. The variances found for $\sqrt{k}$ were all considerably higher than those for $\sqrt{\bar{k}}$, the ratios to the latter ranging from 2½ to sixty-seven.

Principal component	k	$\sqrt{k}$	$\sqrt{\bar{k}}$	Ratio of difference to standard error
1	1.846	1 359	.801	6 72
2	1.465	1.210	.814	5.28
3	.521	.722	.665	1.31
4	.167	.406	.413	−.28

The value for the fourth component is actually less than the value to be expected on the basis of errors of measurement, while that for the third component does not significantly exceed the expected value. For the other two, however, the excess is decidedly significant.

We conclude that the true scores, if we could find them, would display a scatter diagram of at least two dimensions, *i.e.* that there are at least two genuine independent components; but this experimental material supplies no evidence for more than two independent components. Thus we can definitely affirm the existence of two such components, though we cannot distinguish definitely between them with one hundred forty individuals. In the scatter diagram, we can fix with some definiteness the plane of these two leading components, but on account of their nearly equal contributions to the variance we cannot be at all sure of their directions within the plane. The ellipses of the scatter diagram are too nearly circular for this. It is possible, but far from certain on this evidence, that they are really three-dimensional, close in form to oblate spheroids.

(*To be concluded in October issue.*)