Behavior Genetics, Vol. 4, No. 3, 1974

A Method for Analyzing the Genetic Basis of Covariation

L. J. Eaves[1] and J. S. Gale[1]

Received 30 May 1973—Final 9 Nov. 1973

Covariances between traits can be partitioned into additive and dominance genetic components and between- and within-family environmental components, using a method analogous to that used in the analysis of single traits. The problem arises as to whether all additive genetic components simply reflect a single additive component, in the sense that, given an appropriate rescaling of the breeding values, a single additive genetic component would adequately describe the additive genetic variation. The statistical procedure for testing this hypothesis is discussed in detail. Similar considerations apply to the dominance variation. The approach is applied to twin data given by Loehlin and Vandenberg (1968) on covariation between five of Thurstone's Primary Mental abilities. Although the data do not permit a reliable separation of additive and dominance components, it is shown that a single genetic component will account for almost all the genetic variation and covariation. Unless there is marked linkage disequilibrium, this implies that most of the genetic variation for the five traits can be attributed to the pleiotropic action of genes at a common set of loci.

KEY WORDS: twins; quantitative methods; partitioning; covariance; abilities.

INTRODUCTION

During the last few years, there have been several attempts to provide multivariate extensions of genetic analyses in order to elucidate the struc-

This work is part of a project in Psychogenetics supported by the British Medical Research Council.

[1] Department of Genetics, University of Birmingham, Birmingham, England.

ture of genetic and environmental covariation for multiple measurements (Vandenberg, 1965; Loehlin, 1965; Loehlin and Vandenberg, 1968; Roudabush, 1968; Eaves, 1972*a*, 1973*a*; Eaves and Brumpton, 1972; Gale and Eaves, 1972). While these authors have partitioned covariation into genetic and environmental components, they have not attempted to resolve the genetic components any further, for example, into additive and dominance components. These authors have shown that it will often be necessary to analyze components of mean squares in order to elucidate the genetic system. In such cases, standard multivariate procedures, although sometimes helpful, are not really appropriate for detailed genetic analysis of natural populations.

The biometrical approach, as used in the analysis of individual characters (e.g., Mather and Jinks, 1971), has established a number of principles, which are equally relevant in the multivariate situation. First, raw components estimated from the analysis of variance do not in general have a simple genetic meaning. For example, the genetic component of variation within pairs of dizygotic twins represents both additive and nonadditive variation. It is essential, therefore, to specify the expectations of the components in terms of appropriate genetic parameters, which may then be estimated. Second, since such estimates are correct only if the expectations of components are correctly represented, it is essential to test the validity of the assumptions under which estimates have been made.

In this paper, we shall suggest an approach to the analysis of genetic covariation which takes account of the considerations just mentioned. Our analysis will be in two stages. In the first, we find the simplest genetic model consistent with a given body of data, while leaving open the question of the extent to which different traits are affected by the same genes. We then consider the hypothesis that there are a number of loci controlling the traits studied and that a gene substitution at any of these loci affects all the traits. While this is almost certainly an oversimplification, it may well be the case that the bulk of the observed genetic variation and covariation may be the result of variation at such a single set of loci. We shall apply our approach to data on cognitive abilities given by Loehlin and Vandenberg (1968).

A GENETIC MODEL FOR THE DATA

For a single character in a randomly mating population, we partition the genetic variation into an addditive genetic component (D_R) and a dominance component (H_R). For a polygenic system with two main alleles per

locus these components are defined (Mather and Jinks, 1971) as follows:

$$D_R = \Sigma\, 4u_a v_a d_a^2 + \Sigma\, 4u_a v_a h_a^2 - \Sigma\, 16u_a^2 v_a^2 h_a^2 - \Sigma\, 8u_a v_a (u_a - v_a) d_a h_a$$

$$H_R = \Sigma\, 16u_a^2 v_a^2 h_a^2$$

where $\sum$ indicates summation over all loci affecting the trait and, at a given locus; u_a, v_a are the population frequencies of the two alleles, d_a is the absolute deviation of a homozygote from the mean of the two homozygotes, and h_a is the deviation of the heterozygote from the mean of the two homozygotes.

The total genetic variance V_G is

$$\tfrac{1}{2}D_R + \tfrac{1}{4}H_R$$

Should the data give evidence of assortative mating, this formula can be modified appropriately (Jinks and Fulker, 1970). Any bias due to ignoring assortative mating in the data actually analyzed will be discussed later. The model could be extended to include epistasis, but, given the limited resolution in most human studies, this extension is scarcely worthwhile at present.

Our model for trait covariation is analogous to that for the variation of a single trait. We regard some of the genes affecting one trait, say trait i, as though they had a pleiotropic effect on another trait j. Loci are assumed to be in linkage equilibrium.

Thus at a locus with two alleles, we suppose that the homozygote for the increasing allele for trait i deviates from the mean of the two homozygotes by an amount d_{ai}. The corresponding deviation for trait j will be d_{aj}, in cases where the increasing allele for trait i is also the increasing allele for trait j. If this is not so, the deviation will be $-d_{aj}$. For the other homozygote at this locus, we reverse the sign of d_{ai} and d_{aj}. The heterozygote at this locus will deviate by h_{ai} in trait i, h_{aj} in trait j. Then it is easily shown by the usual methods that the contribution of this locus to the population covariation of the two traits is

$$2u_a v_a d_{ai} d_{aj} + 2u_a v_a h_{ai} h_{aj} - 4u_a^2 v_a^2 h_{ai} h_{aj} - 2u_a v_a (u_a - v_a)(d_{ai} h_{aj} + d_{aj} h_{ai})$$

Summing over all loci, we obtain the genetic covariance of the traits

$$W_{Gij} = \tfrac{1}{2}D_{Rij} + \tfrac{1}{4}H_{Rij}$$

where

$$D_{Rij} = \Sigma\, 4u_a v_a d_{ai} d_{aj} + \Sigma\, 4u_a v_a h_{ai} h_{aj} - \Sigma\, 16u_a^2 v_a^2 h_{ai} h_{aj}$$
$$- \Sigma\, 4u_a v_a (u_a - v_a)(d_{ai} h_{aj} + d_{aj} h_{ai})$$

and

$$H_{Rij} = \Sigma\, 16u_a^2 v_a^2 h_{ai} h_{aj}$$

When $i = j$, this becomes the usual

$$V_G = \tfrac{1}{2}D_R + \tfrac{1}{4}H_R$$

We should stress that genetic correlations have very complicated expectations, so that there is every advantage in working with variances and covariances and not calculating genetic correlations.

ESTIMATING THE COMPONENTS OF THE MODEL

Experimental designs appropriate for the estimation of D_R and H_R for a single trait have been considered elsewhere (Jinks and Fulker, 1970; Eaves, 1972*b*). Any design suitable for estimating genetic components of variance will also be suitable for estimating genetic components of covariance.

We estimate components by weighted least squares which, given large samples, will supply approximately maximum likelihood estimates of the unknown components.

We suppose that, in the usual way, observations have been made on a number n of classes, such as monozygotics reared together, full sibs reared apart, cousins. Within a class, we have variation between and within subclasses, e.g., variation between and within pairs of monozygotics reared together. If we have measured p traits, any class supplies both a "between" and a "within" $p \times p$ covariance matrix. These $2n$ matrices, say $\mathbf{S}_1$, $\mathbf{S}_2$, $\mathbf{S}_3, \ldots, \mathbf{S}_{2n}$ are our basic data, which thus consist of

$$2n \times \tfrac{1}{2}p(p+1)$$

separate elements, which we shall write as a column vector y, with $np(p+1)$ rows. We shall refer to y as the vector of raw statistics.

The next step is to calculate the variances and covariances of the raw statistics. We first note that raw statistics derived from different S matrices have covariance zero. The variances and remaining covariances of raw statistics may be found from standard theory (Kendall and Stuart, 1963).

Consider any S matrix, say $\mathbf{S}_m$, with typical element s_{mkl}. Then

$$\mathrm{Cov}(s_{mij}, s_{mab}) = \frac{1}{N_m}\left(\varepsilon s_{mia}\varepsilon s_{mjb} + \varepsilon s_{mja}\varepsilon s_{mib}\right) \qquad (1)$$

where N_m is the number of degrees of freedom of any element in $\mathbf{S}_m$. Although we do not, of course, know the expected values of the s_{mkl}, we shall

obtain estimates of these in the course of the calculation. We shall denote the covariance matrix of the raw statistics as V.

We must now write the expected values of the raw statistics in terms of the unknown parameters, such as D_{Rij}, H_{Rij} and corresponding environmental components, which we wish to estimate. We may write these unknown parameters, in the usual way, as a column vector θ. Thus we have

$$\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + \boldsymbol{\varepsilon} \tag{2}$$

where ε is a column vector with $np(p+1)$ rows representing sampling variation of the raw statistics.

The matrix A comprises the coefficients of the unknown parameters in the algebraic expressions for the expected values of the raw statistics. These expressions are readily found from the corresponding expressions for the univariate case (e.g., Jinks and Fulker, 1970; Eaves, 1973*b*); when dealing with a raw statistic based on a pair of traits i and j, we simply replace the univariate genetic components D_R, H_R and environmental components E_1, E_2 by the corresponding multivariate components D_{Rij}, H_{Rij}, E_{1ij}, E_{2ij}.

Recalling that V denotes the covariance matrix of the raw statistics, we have, from standard theory, the maximum likelihood estimates

$$\hat{\boldsymbol{\theta}} = (\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}^{-1}\mathbf{y}$$

In the first instance, we obtain approximate values for the elements of V by using observed rather than expected values in equation (1). This will supply approximate values for the $\hat{\theta}$. We substitute these in

$$\hat{\mathbf{y}} = \mathbf{A}\hat{\boldsymbol{\theta}}$$

The $\hat{y}$ are then improved estimates of the $\mathcal{C}s_{mkl}$ and thus give rise to an improved $\hat{\mathbf{V}}$. We repeat the whole procedure a number of times until stable values for the $\hat{\theta}$ are obtained. These will be sensible provided that our model, given in equation (2), is an adequate description of the data. Provided our raw statistics are based on sufficiently large samples, we test the goodness of fit of the model by calculating.

$$(\mathbf{y} - \hat{\mathbf{y}})'\mathbf{V}^{-1}(\mathbf{y} - \hat{\mathbf{y}}) \tag{3}$$

where $\hat{y}$ is calculated using the final estimates of the $\hat{\theta}$. Given large samples, the expression (3) is distributed as χ^2, with degrees of freedom equal to the number of raw statistics less the number of parameters estimated.

EXAMPLE

The data chosen to illustrate the genetic analysis are twin data given by Loehlin and Vandenberg (1968) for the covariation of five of Thur-

Table I. The Genetic Model for a Single Variable

Mean square	E.M.S.			
	D_R	H_R	E_2	E_1
Between MZ pairs reared together	1	$\frac{1}{2}$	2	1
Within MZ pairs	—	—	—	1
Between DZ pairs reared together	$\frac{3}{4}$	$\frac{5}{16}$	2	1
Within DZ pairs	$\frac{1}{4}$	$\frac{3}{16}$	—	1

stone's Primary Mental Abilities, which form appendices A D of their paper. The authors discuss in detail the structure of their sample. They conclude that their MZ and DZ twins can be regarded as samples from the same population.

Limitations in their experiment are discussed by the authors. Indeed, qualitative and quantitative considerations (see, e.g., Eaves and Jinks, 1972) combine to make this particular experiment a relatively inefficient one for genetic analysis, since we have only 123 pairs of MZ twins and 75 pairs of DZ twins, the members of each pair having been raised together. We may, however, still test very simple hypotheses about the kinds of genetic and environmental influences contributing to variation for the five traits. A relatively complete model for the various *mean squares*, involving additive (D_R) and dominance (H_R) genetic components and also between-family (E_2) and within-family (E_1) environmental components, is given in Table I for variation in a single trait. The model assumes random mating. Although there are four raw statistics and four unknown parameters, it is impossible with the present design to estimate all of the latter, since the model is not of full rank.

We can, however, provide a (not very powerful) test of the adequacy of a simpler model involving the $D_{R/j}$ and the $E_{1/j}$ only, while recognizing that if this model fails, we shall not be able unambiguously to detect the source of the failure. When we attempted to fit this simpler model, we found that solutions were unstable; the residual χ^2 at successive stages were generally significant at the 5% level. Hence we must reject the simpler model. It was noticeable that the variation between DZ pairs was much larger than that expected on the basis of the reduced model. This discrepancy could be due either to assortative mating or to environmental influences common to members of the same pair.

We extend our model, therefore, by the addition of $E_{2/j}$ components. For this model, the residual χ^2, with 15 df, converged in four iterations to a

stable value of 10.8 ($0.7 < P < 0.8$). Estimates of components are given in Table II.

Thus a model including only additive and the two kinds of environmental effects gives an excellent fit to the data. This is not as helpful as it might seem, since with the present experimental design, dominance would not lead to failure of the model. In fact, if dominance is present, our supposed estimates of D_{Rij} and E_{2ij} are really

$$\hat{D}_{Rij} + \tfrac{3}{4}\hat{H}_{Rij} \qquad \text{and} \qquad \hat{E}_{2ij} - \tfrac{1}{8}\hat{H}_{Rij}$$

respectively. However, if the H_{Rij} were large relative to the E_{2ij}, our model should yield significant *negative* estimates of the E_{2ij} rather frequently. This has not happened. We shall proceed, therefore, on the tentative basis that our $\hat{D}_{Rij}$ are reasonable estimates of additive effects. At least, this will illustrate the procedure that we think should be followed when more extensive data become available.

If assortative mating is present, any supposed $\hat{E}_{2ij}$ is really

$$\hat{E}_{2ij} + \tfrac{1}{2}\frac{A}{1-A}\hat{D}_{Rij}$$

where A is the correlation between the breeding values of the spouses, but

Table II. Estimates of Variance and Covariance Components

ij	$\hat{D}_{Rij}$	$\hat{E}_{2ij}$	$\hat{E}_{1ij}$
11	3137.61[a]	321.14	370.53[a]
12	848.66[b]	513.17[c]	96.97
13	1671.45[c]	193.44	2.16
14	1063.20[c]	54.46	0.00
15	712.47[c]	249.86	30.81
22	663.57[b]	667.98[a]	159.53[a]
23	242.55	438.22[c]	20.27
24	520.17[b]	193.53	58.20[a]
25	228.79	375.56[a]	69.39[a]
33	2426.95[a]	837.22[b]	451.02[a]
34	963.50[c]	−171.66	31.25
35	571.00[b]	85.07	16.34
44	1095.19[a]	−19.39	195.39[a]
45	224.36	171.95[b]	44.17[c]
55	184.57	354.99[a]	126.45[a]

[a] Significant at the 0.1% level.
[b] Significant at the 5% level.
[c] Significant at the 1% level.

our $\hat{D}_{R_{ij}}$ are still estimates of the corresponding $D_{R_{ij}}$ given dominance absent. Thus assortative mating presents a much less serious problem than dominance in the present context.

TESTING THE SIMPLEST GENETIC MODEL FOR COVARIATION

We now ask: do all the five traits have a common genetic control? In other words, are all five traits the pleiotropic expression of a single set of loci? We should note, however, that apparent pleiotropy would arise with more than one set of loci if these were in marked linkage disequilibrium. Since we cannot distinguish these two situations, we shall refer to both as the "single-set" hypothesis.

One approach, which has appeared in the literature but is agreed to be unsatisfactory, is to convert the $\hat{D}_{R_{ij}}$ into correlation coefficients and carry out a principal components analysis on these coefficients. Should a single component account for most of the variation, this would be regarded as evidence for the single-set hypothesis. Apart from the often mentioned mathematical objections to this procedure, it is extremely difficult to interpret the procedure in genetic terms. For example, the correlation coefficients have no simple genetic meaning.

We shall propose a rather different approach, in which we attempt to fit a simple model to the $\hat{D}_{R_{ij}}$. Extensions for cases where $H_{R_{ij}}$ are detected will be discussed below. We ask: are all the $D_{R_{ij}}$ really the same D_R, any differences being merely a result of the scales on which the different traits were measured? If this is so, we shall find that after an appropriate rescaling of the breeding values only D_R is required to account for all the additive variation and covariation. Let the breeding values for the five traits be (linearly) rescaled by multiplying by $z_1, z_2, \ldots, z_5$ respectively. Then we have

$$\varepsilon\hat{D}_{R_{ij}} = z_i z_j D_R$$

The z_i and D_R must now be estimated, account being taken of the differing precisions of the various $\hat{D}_{R_{ij}}$. Finally, an approximate test of significance of this single-set model is calculated. It would, in principle, be possible to fit a model of this kind (taken in conjunction with environmental components) to the original raw statistics, but this would be complicated and the gain probably marginal.

Although our approach bears an obvious relation to principal components analysis in this simple case, even here the results are slightly different, as will be seen later. Our method appears to extend more readily

than others to at least some of the more complicated cases encountered in practice. For example, if $H_{R_{ij}}$ are detected we can attempt to fit a single D_R and a single H_R to the $D_{R_{ij}}$ and the $H_{R_{ij}}$ taken together using the *same* five z_i for both. Only if this model failed would we fit separate z_i for additive and dominance components, which would be roughly equivalent to separate principal components analyses of the additive and dominance effects. Similar considerations apply if we have data from different populations. Testing the significance of departures from the various models is also quite straightforward, provided, of course, that samples are large.

For brevity, we shall write D_{ij} as short for $D_{R_{ij}}$ and D as short for D_R. We write p for the number of z_i (five in our case). Let $R_{ij,mn}$ be the element of the information matrix (of the D_{ij}) corresponding to $\hat{D}_{ij}$ and $\hat{D}_{mn}$. We minimize

$$F = \sum_{\substack{ij \\ i \leq j}} \sum_{\substack{mn \\ m \leq n}} R_{ij,mn}(\hat{D}_{ij} - \varepsilon\hat{D}_{ij})(\hat{D}_{mn} - \varepsilon\hat{D}_{mn})$$

for variation in the p z_i and in D.

Given sufficiently large samples, F is distributed as χ^2 and the logarithmic likelihood L is $-\frac{1}{2}$ (this χ^2), so that we should obtain approximate maximum likelihood estimates, given our present data.

One important point is the restraint which must be imposed on the z_i. Our estimation will be subject to the condition

$$\Sigma\, \hat{z}_i^2 = 1$$

so that we have p independent parameters. Writing ϕ_i $(i = 1, 2, \ldots, p)$ for the independent unknown parameters (four of the z_i and D in the present case) we must solve the equations

$$\frac{\partial F}{\partial \varphi_i} = 0 \qquad (i = 1, 2, \ldots, p)$$

We may take any $(p - 1)$ of the z_i as free parameters without affecting the final result; we follow the usual procedure of taking the "last," z_p, as the dependent parameter. Then, writing

$$\frac{\partial^* F}{\partial \varphi_i}$$

for the formal derivative of F with respect to any ϕ_i, including z_p, obtained by regarding all the z_i as free parameters, we have

$$\frac{\partial F}{\partial z_k} = \frac{\partial^* F}{\partial z_k} + \frac{\partial^* F}{\partial z_p}\frac{\partial z_p}{\partial z_k} = \frac{\partial^* F}{\partial z_k} + \frac{\partial^* F}{\partial z_p}\left(\frac{-z_k}{z_p}\right) \qquad (k = 1, 2, \ldots, p-1) \quad (4)$$

where the "starred" derivatives are obtained from the usual rules for differentiation of matrix products, as follows.

Let the $D_{ij}(i \leq j)$ be written in any convenient order. We denote the rth D_{ij} in this list by the symbol θ_r. Thus we wish to minimize

$$F = \sum_r \sum_s R_{rs}(\hat{\theta}_r - \varepsilon\hat{\theta}_r)(\hat{\theta}_s - \varepsilon\hat{\theta}_s)$$

which we may write in matrix form

$$(\hat{\boldsymbol{\theta}} - \varepsilon\hat{\boldsymbol{\theta}})'\mathbf{R}(\hat{\boldsymbol{\theta}} - \varepsilon\hat{\boldsymbol{\theta}})$$

We now have

$$\frac{\partial^* F}{\partial \varphi_r} = 2(\hat{\boldsymbol{\theta}} - \varepsilon\hat{\boldsymbol{\theta}})'\mathbf{R}\frac{\partial}{\partial \varphi_r}(\hat{\boldsymbol{\theta}} - \varepsilon\hat{\boldsymbol{\theta}})$$

Then the $\partial F/\partial z_k$ follow at once from (4); $\partial F/\partial D$ is, of course, the same as $\partial^* F/\partial D$.

We write $\mathbf{T}$ for the column vector whose elements are $\partial F/\partial D$ and the $(p - 1)$ $\partial F/\partial z_k$. The matrix of maximum likelihood "scores," in Fisher's (1946) terminology (see also Bailey, 1961, Appendix I), is

$$\mathbf{S} = -\tfrac{1}{2}\mathbf{T}$$

given the qualifications about sample size discussed earlier.

Again, using the "star" notation for formal differentiation when all parameters are regarded as free, we have

$$\frac{\partial^2 F}{\partial z_k \partial z_l} = \frac{\partial^{*2} F}{\partial z_k \partial z_l} + \frac{\partial^{*2} F}{\partial z_k \partial z_p}\frac{\partial z_p}{\partial z_l} + \frac{\partial^{*2} F}{\partial z_l \partial z_p}\frac{\partial z_p}{\partial z_k} + \frac{\partial^* F}{\partial z_p}\frac{\partial^2 z_p}{\partial z_k \partial z_l} + \frac{\partial^{*2} F}{\partial z_p{}^2}\frac{\partial z_p}{\partial z_k}\frac{\partial z_p}{\partial z_l}$$

In this expression

$$\frac{\partial z_p}{\partial z_l} = -\frac{z_l}{z_p}\frac{\partial^2 z_p}{\partial z_k \partial z_l} = -\frac{z_k z_l}{z_p{}^3}$$

Also,

$$\frac{\partial^2 F}{\partial D \partial z_l} = \frac{\partial^{*2} F}{\partial D \partial z_l} + \frac{\partial^{*2} F}{\partial D \partial z_p}\left(-\frac{z_l}{z_p}\right)$$

and

$$\frac{\partial^2 F}{\partial D^2} = \frac{\partial^{*2} F}{\partial D^2}$$

The "starred" derivatives are given by

$$\frac{\partial^{*2} F}{\partial \varphi_r \partial \varphi_s} = 2\left[\frac{\partial}{\partial \varphi_s}(\hat{\boldsymbol{\theta}} - \varepsilon\hat{\boldsymbol{\theta}})'\right]\mathbf{R}\frac{\partial}{\partial \varphi_r}(\hat{\boldsymbol{\theta}} - \varepsilon\hat{\boldsymbol{\theta}}) + 2(\hat{\boldsymbol{\theta}} - \varepsilon\hat{\boldsymbol{\theta}})'\mathbf{R}\frac{\partial^2}{\partial \varphi_r \partial \varphi_s}(\hat{\boldsymbol{\theta}} - \varepsilon\hat{\boldsymbol{\theta}})$$

We write **J** for $(\frac{1}{2})$ times the $p \times p$ matrix of "unstarred" second derivatives. Then **J** is, approximately, the matrix of information *observed*.

In carrying out the computations, the first step is to evaluate **R**. This is done, in the usual way, by taking the covariance matrix of the estimates of all the genetic and environmental components in the first analysis, namely the covariance matrix

$$(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}$$

striking out all rows and columns except those pertaining to the variances and covariances *inter se* of the $\hat{D}_{ij}$ and inverting what remains.

The equations

$$\mathbf{S} = \mathbf{0}$$

must be solved numerically. Since we are attempting a kind of weighted components analysis, appropriate trial values $\phi_{(0)}$ for the unknown parameters are for D the first eigenvalue of the matrix of the $\hat{D}_{ij}$ and for the z_i the elements of the corresponding eigenvector. We now evaluate **S** and **J** using these trial values and find the improved estimates

$$\hat{\boldsymbol{\varphi}}_{(1)} = \hat{\boldsymbol{\varphi}}_{(0)} + \mathbf{J}^{-1}\mathbf{S}$$

The process can be repeated, using the improved estimates as trial values to obtain still better estimates. We continue to repeat the process until we obtain stable estimates. However, since some of our expressions will, in practice, involve the dependent parameter z_p, care must be taken when obtaining this parameter as

$$\hat{z}_p = \left(1 - \sum_{i=1}^{p-1} \hat{z}_i^2\right)^{1/2}$$

We take the sign which gives best agreement between observed and expected D_{ij}. This can be done without difficulty unless there is a gross failure of the model. Again, although there are presumably maxima in the likelihood surface corresponding roughly to the *other* eigenvalues, we should not be in any serious danger of converging on such irrelevant maxima, except perhaps in cases where the model is clearly inappropriate.

RESULTS

Starting with the $\hat{D}_{R_{ij}}$ obtained from Loehlin and Vandenberg's data, stability (to the tenth significant figure) was obtained after six iterations. Trial values and final estimates, together with their approximate standard errors (obtained, for the free parameters, as the square root of diagonal ele-

ments of $\mathbf{J}^{-1}$), are given in Table III. The standard error of $\hat{z}_5$ is estimated by

$$\frac{1}{\hat{z}_5}\left[\sum_{i=1}^{4}\sum_{j=1}^{4}\hat{z}_i\hat{z}_j\,\mathrm{Cov}(\hat{z}_i,\hat{z}_j)\right]^{1/2}$$

From these, values for expected $\hat{D}_{Rij}$ were calculated. These are given in Table IV.

The approximate χ^2_{10} testing the goodness of fit of our simple model proved to be

$$20.41, \qquad P = 2.6\%$$

indicating that the fit is not completely satisfactory.

On the other hand, there is no doubt that the bulk of the genetic variation can be accounted for in terms of our simple model. For on testing the null hypothesis that the $\hat{D}_{Rij}$ are, jointly, not significantly different from zero, we obtained

$$\chi^2_{15} = 44.58, \qquad P = 0.01\%$$

so that the reduction in χ^2 obtained on fitting our simple model is

$$\chi^2_5 = 24.17, \qquad P = 0.02\%$$

In view of this, there is little point, in the present case, in attempting to fit a more complicated model, such as one involving additional loci with effects on some of the characters only or perhaps with effects specific to particular characters. Since an improved fit would, in the present case, be obtained using any one of a number of distinct genetic models, it is unlikely that any decision could be made as to which of these models is correct.

Table III. Estimates of Parameters of the Single-Set Model

Parameter	Trial value	Final value	Standard error
D_R	5387.9235	5349.2000	1164.3186
z_1	0.7076	0.6983	0.0863
z_2	0.2015	0.2307	0.0854
z_3	0.5599	0.5355	0.1164
z_4	0.3348	0.3772	0.0871
z_5	0.1816	0.1736	0.0557

Table IV. Expected Values of $\hat{D}_{Rij}$ Obtained by Fitting the Single-Set Model

Scale	N	V	S	W	R
N	2608.39	861.74	2000.28	1408.97	648.46
V		284.70	660.84	465.49	214.23
S			1533.94	1080.49	497.28
W				761.08	350.28
R					161.21

DISCUSSION

In the present case, our results are substantially the same as those reached by Loehlin and Vandenberg, who found that a single factor would account for most of the genetic correlation between the traits studied. We suggest, however, that even in the present case our approach has advantages. In that our simple model is formulated in precise genetic terms, it is easy to see which hypothesis is actually being tested. Our method also provides an approximate test of significance of departures from the model.

We now consider briefly more general situations. We have shown that, given an adequate experimental design, the critical separation of genetic variation into additive and dominance effects may be carried out as readily in the multivariate as in the univariate case.

Analysis of unresolved genetic components, in cases where the single-factor model failed, would not enable us to distinguish between the following two distinct situations. In the first, there are two or more separate sets of loci affecting at least some of the traits. In the second, there is just one set of loci, but the degree of dominance varies from one character to another. In this latter case, a model with one D_R, one H_R but the same set of z_i both for dominance and for additive effects would fail, whereas a model with one D_R, one H_R, but two different sets of z_i (one for additive effects, one for dominance effects) should fit the data. In such a case, attempts to explain the data in terms of orthogonal factors would lead to purely formal results, with a misleading genetic interpretation. Although, in factor terms, the correct result would be two oblique factors, one for additive effects and one for dominance, we would not arrive at this conclusion by oblique rotation to simple structure, so that there is no known method by which the factor approach can be made to yield the correct result.

Extension to the case where we must fit two or more D_R, or two or more H_R, is more difficult. Here we encounter the same difficulties with

regard to nonuniqueness of solutions as are found in standard factor analysis. We may note that this nonuniqueness is, in fact, a mathematical reflection of a real limitation in the inferences that can be made about the genetics of covariation for an organism in which breeding experiments are not possible. Consider the following very simple artificial example. Two traits A and B are genetically correlated with one another, but genetically uncorrelated with two other traits C and D, which in turn are genetically correlated with one another. It is natural to interpret this result as the outcome of two sets of loci, one for A,B and one for C,D. On the other hand, we can explain the results equally well by postulating one set of loci, half of which affect A, B, C, D in a consistent manner, whereas for the other half, increasing alleles for A,B are decreasing alleles for C,D. If the first interpretation is correct, a two-factor simple structure representation would correctly reflect the genetic situation. If the second interpretation is correct, then the genetic situation should be represented by one general factor and one bipolar factor. The interpretations could be distinguished only if we were in the position to carry out an appropriate selection experiment. Otherwise, as we have discussed elsewhere (Gale and Eaves, 1972), we have to postulate consistency of gene action if we are to identify factors with sets of loci in cases where a single-factor model fails. The same problem will arise in the present context if we are required to fit more than one D_R, or more than one H_R.

ACKNOWLEDGMENT

Computations were carried out on the University of Birmingham KDF-9 computer.

REFERENCES

Bailey, N. T. J. (1961). *Introduction to the Mathematical Theory of Genetic Linkage*. Oxford University Press, London.

Eaves, L. J. (1972*a*). The multivariate analysis of certain genotype-environment interactions. *Behav. Genet.* **2**: 241–244.

Eaves, L. J. (1972*b*). Computer simulation of sample size and experimental design in human psychogenetics. *Psychol. Bull.* **77**: 144–152.

Eaves, L. J. (1973*a*). The structure of genotypic and environmental covariation for personality measurements: An analysis of the P.E.N. *Brit. J. Soc. Clin. Psychol.* **12**: 275–282.

Eaves, L. J. (1973*b*). Assortative mating and intelligence: An analysis of pedigree data. *Heredity* **30**: 199–210.

Eaves, L. J., and Brumpton, R. J. (1972). Factors of covariation in *Nicotiana rustica*. *Heredity* **29**: 151–175.

Eaves, L. J., and Jinks, J. L. (1972). Insignificance of evidence for differences in heritability of I.Q. between races and social classes. *Nature* **240**: 84–88.

Fisher, R. A. (1946). A system of scoring linkage data, with special reference to pied factors in mice. *Am. Naturalist* **80**: 568–578.

Gale, J. S., and Eaves, L. J. (1972). Variation in wild populations of *Papaver dubium*. V. The application of factor analysis to the study of variation. *Heredity* **29**: 135–149.

Jinks, J. L., and Fulker, D. W. (1970). A comparison of the biometrical genetical, MAVA, and classical approaches to the analysis of human behavior. *Psychol. Bull.* **73**: 311–349.

Kendall, M. G., and Stuart, A. (1963). *The Advanced Theory of Statistics*, 2nd ed., Vol. 1, Griffin and Company, London.

Loehlin, J. C. (1965). A heredity-environment analysis of personality inventory data. In Vandenberg, S. G. (ed.), *Methods and Goals in Human Behavior Genetics*, Academic Press, New York, pp. 163–170.

Loehlin, J. C., and Vandenberg, S. G. (1968). Genetic and environmental components in the covariation of cognitive abilities: An additive model. In Vandenberg, S. G. (ed.), *Progress in Human Behavior Genetics*, Johns Hopkins, Baltimore, pp. 261–278.

Mather, K., and Jinks, J. L. (1971). *Biometrical Genetics: The Study of Continuous Variation*, Chapman and Hall, London.

Roudabush, G. E. (1968). Analyzing dyadic relationships. In Vandenberg, S. G. (ed.), *Progress in Human Behavior Genetics*, Johns Hopkins, Baltimore, pp. 303–333.

Vandenberg, S. G. (1965). Multivariate analysis of twin differences. In Vandenberg, S. G. (ed.), *Methods and Goals in Human Behavior Genetics*, Academic Press, New York, pp. 29–43.