The University of Birmingham

DEPARTMENT OF GENETICS
The University of Birmingham, P.O. Box 363, Birmingham B15 2TT
Telephone 021 - 472 1301 Ext. 2030

Head of Department: Professor J.L. Jinks DSC FI Biol FRS

12th November 1975

Owen White Esq.,
Department of Psychology,
Institute of Psychiatry,
De Crespigny Park,
Denmark Hill,
London SE5 8AF

Dear Owen,

I have mentioned the Joereskog package to Hans, who says there is sufficient money left in the Tobacco fund if needed. Has he talked to you about the possibility of an SRC grant for your twin research?

I thought you would like to see the outcome of my jottings on the train last Saturday, in relation to the problem about fitting multivariate genetical models. I am satisfied that the Joereskog approach comes closest to the right one, but as you will see, I think a couple of modifications (generalisation?) might be needed to estimate the parameters of a genetical model. Obviously, I'd appreciate your comments.

We will assume orthogonal factors throughout. This is necessary for the treatment of assortative mating, though it may not be necessary if mating is random. We start with a simple case, but this should not obscure the generality of the model.

Suppose, for illustration only, that we have data on MZ and DZ twins measured on p variables. For each kind of twin we have two p x p matrices, the matrix of mean products between pairs and that within pairs. These may be written: $\underset{\sim}{S}_{bmz}$, $\underset{\sim}{S}_{wmz}$, $\underset{\sim}{S}_{bdz}$, $\underset{\sim}{S}_{wdz}$, (the subscripts b and w denote the between and within pairs matrices respectively).

We recall that for the univariate case the expectations of mean squares can be written in terms of a biometrical-genetical model. If, for example, all gene action is additive (D_R) mating is random (A=0) and all environmental variation is due to within-family (E_1) effects we may write:

Univariate mean square	Expectation
S_{bmz}	$D_R + E_1$
S_{wmz}	E_1
S_{bdz}	$\frac{1}{4}D_R + E_1$
S_{wdz}	$\frac{1}{4}D_R + E_1$

For the multivariate case we may write a factorial analogue of this model by recognising that:

$$\underset{\sim}{\Sigma}_{D_R} = \underset{\sim}{\Delta}\,\underset{\sim}{\Delta}' + \theta^2$$

and $$\underset{\sim}{\Sigma}_{E_1} = \underset{\sim}{H}\underset{\sim}{H}' + \mathfrak{I}^2$$

where $\underset{\sim}{\Sigma}_{D_R}$ is the covariance matrix of additive genetical effects for a random mating population, $\underset{\sim}{\Sigma}_{E_1}$ represents the covariance matrix of within-family environmental effects.

In order to write the expectations of the $\underset{\sim}{\Sigma}$'s in a form suitable for ML estimation it may be necessary to modify Joereskog's procedure slightly. In general we may write:

$$\underset{\sim}{\Sigma} = \underset{\sim}{B}\,[\underset{\sim}{\Delta}\,\underset{\sim}{\Upsilon}\,\underset{\sim}{\Delta}' + \underset{\sim}{\psi}\,\underset{\sim}{\varrho}\,\underset{\sim}{\psi}]\,\underset{\sim}{B}'$$

The similarity with the general model for the analysis of covariance structures is fairly clear, the only additional component is $\underset{\sim}{\varrho}$ which will represent the different scaling of the $\underset{\sim}{\psi}$ in the expectation of different $\underset{\sim}{\Sigma}$'s.

$\underset{\sim}{\Sigma}$ is the expected value of a particular mean-products matrix (bmz, wmz etc.

$\underset{\sim}{B}$ is a matrix of coefficients

$\underset{\sim}{\Delta}$ contains the genetical and environmental factor loadings

$\underset{\sim}{\psi}$ contains the genetical and environmental specific variances

$\underset{\sim}{\Upsilon}$ and $\underset{\sim}{\varrho}$ are matrices whose coefficients depend on genetical theory. When genotypic and environmental effects are independent they will be diagonal. The diagonal elements of Υ & ϱ may not be unity (is this a problem for Joereskog's approach?)

With m genetical factors and n environmental factors we have, for our D_R, E_1 factor model,

$$\underset{2p \times (m+n)}{\underset{\sim}{\Delta}} = \begin{pmatrix} \underset{\sim}{\Delta} & \cdot \\ \cdot & \underset{\sim}{H} \end{pmatrix}$$

The same $\underset{\sim}{\Delta}$ is expected to apply to all $\underset{\sim}{\Sigma}$, given that our model fits, the contributions of the factors are only going to vary between different relatives by a scalar proportion, defined in $\underset{\sim}{\Upsilon}$ & $\underset{\sim}{\varrho}$.

For a model involving just D_R and E_1 effects B will be the same throughout:

$$\underset{\sim}{B} = (\underset{\sim}{I}, \underset{\sim}{I})$$

where $\underset{\sim}{I}$ is the p x p identity matrix [For models involving additional genetical or environmental effects e.g. E_2, H_R, $\underset{\sim}{B}$ will have additional p x p $\underset{\sim}{I}$ matrices for every $\underset{\sim}{\Sigma}$].

$$\underset{\sim}{\psi} = \begin{pmatrix} \underset{\sim}{\zeta} & \cdot \\ \cdot & \underset{\sim}{\eta} \end{pmatrix}$$

$2p \times 2p = \quad 2p \times 2p$

Most ingenuity will go into specifying $\underset{\sim}{\gamma}$ and $\underset{\sim}{\varrho}$. These will depend on the model being fitted and will vary from one kind of relationship to the next.

Generally, $\underset{\sim}{\gamma}$ and $\underset{\sim}{\varrho}$ will have the same form. When mating is random they will consist simply of a diagonal matrix of the parameter coefficients. each coefficient being repeated for each factor.

Thus, for the D_R, E_1 model

$$\underset{\sim}{\gamma} = \begin{pmatrix} \underset{\sim}{\alpha} & \cdot \\ \cdot & \underset{\sim}{I} \end{pmatrix}$$

where $\underset{\sim}{I}$ is the nxn identity matrix corresponding to the coefficients of the n 'E_1' factors in every $\underset{\sim}{\Sigma}$ and $\underset{\sim}{\alpha}$ is an mxm diagonal matrix whose coefficients will depend on the contribution of the D_R factors to each $\underset{\sim}{\Sigma}$.

When mating is random all the elements of $\underset{\sim}{\alpha}$ will be the same for a given $\underset{\sim}{\Sigma}$.

Below are given the contributions of $\underset{\sim}{\alpha}$ for each of the four $\underset{\sim}{\Sigma}$'s in a randomly mating population.

Statistic	Random mating $\underset{\sim}{\alpha}$
bmz	$\underset{\sim}{I}$
wmz	$0\underset{\sim}{I}$
bdz	$\frac{3}{4}\underset{\sim}{I}$
wdz	$\frac{1}{4}\underset{\sim}{I}$

When mating is assortative $\underset{\sim}{\zeta}$ will be modified by an amount depending on the assortative mating parameter A. Since assortative mating may be different for each factor we can only give a typical element of $\underset{\sim}{\alpha}$, corresponding to the ith genotypic factor.

Statistic	α_{ii}
bmz	$\frac{1}{1}-A_i$
wmz	0
bdz	$(3+A)/4(1-A)$
wdz	$\frac{1}{4}$

The modifications of α_{ii} for other kinds of relatives follow from genetical theory.

If there is genotype-environmental covariation $\underset{\sim}{\rho}$ and $\underset{\sim}{\gamma}$ will no longer be diagonal, since we would have to allow for correlations between the factors contributing to $\underset{\sim}{\Delta}$ and $\underset{\sim}{H}$.

Just to give some clarity to the model, I give the expectations in full for a simple example of three variables, with a single genetical (D_R) and single environmental (E_1) factor, assuming random mating, for the mean product matrix within DZ twins.

$$\underset{\sim}{\Sigma}_{wdz} = \begin{pmatrix} 1 & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & 1 \end{pmatrix} \left[\begin{pmatrix} \Delta_{11} & \cdot \\ \Delta_{21} & \cdot \\ \Delta_{31} & \cdot \\ \cdot & H_{11} \\ \cdot & H_{21} \\ \cdot & H_{31} \end{pmatrix} \begin{pmatrix} \frac{1}{4} & \cdot \\ \cdot & 1 \end{pmatrix} \begin{pmatrix} \Delta' & \cdot \\ \cdot & H' \end{pmatrix} + \begin{pmatrix} \delta_1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \delta_2 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \delta_3 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \eta_1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \eta_2 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \eta_3 \end{pmatrix} \begin{pmatrix} \frac{1}{4} & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \frac{1}{4} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \frac{1}{4} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix} \begin{pmatrix} \delta & \cdot \\ \cdot & \eta \end{pmatrix} \right] \begin{pmatrix} \underset{\sim}{I} \\ \underset{\sim}{I} \end{pmatrix}$$

$$\underset{\sim}{B} \left[\underset{\sim}{\Lambda} \; \underset{\sim}{\gamma} \; \underset{\sim}{\Lambda}' + \underset{\sim}{\psi} \; \underset{\sim}{\rho} \; \underset{\sim}{\psi} \right] \underset{\sim}{B}'$$

This multiplies out to:

$$\tfrac{1}{4}\underset{\sim}{\Delta}\underset{\sim}{\Delta}' + \tfrac{1}{4}\underset{\sim}{\delta}^2 + \underset{\sim}{H}\underset{\sim}{H}' + \underset{\sim}{\eta}^2$$

which is the multivariate analogue of $\frac{1}{4}D_R + E_1$

Whether or not the model (or some more elaborate version of it) can be fitted to any data will depend on whether the existing Joereskog routine can already handle, or can be made to handle:

1. γ not being a "correlation matrix" sensu strictu, but a scaling matrix whose leading diagonal elements are not necessarily unity.
2. The interposition of $\underset{\sim}{\rho}$ in the definition of the specificivariances;
3. The fitting of a model to several independent $\underset{\sim}{\Sigma}$'s separately.

My guess is that the numerical analysis wouldn't be too bad anyhow, we could either use generalised least squares, or an adaptation of the equations for M.L.factor analysis. I haven't much idea about the distribution of the test statistic, though! I don't need to stress the importance of the question. It would be a valuable breakthrough in behaviour genetic analysis to be able to estimate the parameters of such a model.

If you see John Rust, tell him I've read his "evoked potential" paper and found it very clear. It's nice to see the gospel spreading!

Best wishes,

Yours sincerely,

Linda

Lindon Eaves