Heredity (1978), **40** (1), 97-116

THE POWER OF THE CLASSICAL TWIN STUDY

N. G. MARTIN,* L. J. EAVES,* M. J. KEARSEY* and P. DAVIES†
**Department of Genetics and †Department of Mathematical Statistics, University of Birmingham, Birmingham B15 2TT*

Received 5.v.77

Summary

A method based on the non-central chi-square distribution is developed for the calculation of sample sizes required to reject, with given probability, models of variation when they are "wrong". The method is illustrated with reference to simple alternative models of variation in MZ and DZ twins reared together. Simulation of twin experiments finds the empirical power in good agreement with that predicted by the method. Tables are produced showing the sample sizes required for 95 per cent rejection at the 5 per cent level of inappropriate models of variation. For equivalent cases it is always found easier to reject an inappropriate simple genetical model of variation than an inappropriate simple environmental model. For several frequently encountered cases, more than 600 pairs of twins would be required to reject inappropriate alternative models. The optimum proportion of MZ and DZ twins in a sample will vary with the "true" model of variation but is most likely to be between two-thirds and one-half of DZ twin pairs.

The possibility of detecting genetical non-additivity with the classical twin study is investigated by theoretical power calculations and simulation. In the absence of genotype-environment interactions, distributional skewness and mean-variance regression in DZ twins are found to be more powerful tests of directional dominance (or unequal gene frequencies) than the standard model fitting procedure and these tests may be worthwhile in future studies.

1. Introduction

The study of monozygotic and dizygotic twins reared together is the most widely used design for the investigation of continuous variation in humans. What is potentially detectable with classical twin studies and the procedure for estimation and testing models of variation have been discussed extensively in the literature (*e.g.* Jinks and Fulker, 1970; Mather and Jinks, 1971; Eaves and Eysenck, 1975, 1976, 1977; Martin, 1975; Eaves, 1976 and 1977). Briefly, between and within pairs mean squares are calculated for MZ and DZ twins and theoretical models of variation are fitted to the four observed mean squares by the method of weighted least squares, the goodness of fit being judged by the chi-square test.

Models which assume no genotype-environmental covariance (see Eaves, 1976) are usually subsets of the basic model for observed mean squares shown in table 1. This includes sources of variation arising from environmental differences within (E_1) and between (E_2) families, additive gene action (D_R), dominant gene action (H_R), and assortative mating, the parameter A being the correlation between the additive deviations of spouses. The basic model contains five unknown parameters but only four observed statistics. Furthermore, the sum of the first two equations is equal to the sum of the second pair, *i.e.* in the absence of certain kinds of genotype-environment covariation, the total variances of MZ and DZ twins are expected to be equal. Thus there are only three independent equations and the fourth degree of freedom merely tests the equality of the total MZ and DZ variances.

TABLE 1

Basic model for mean-squares of twins reared together

	E_1	E_2	D_R	H_R
MZ between	1	2	$1+\frac{A}{1-A}$	$\frac{1}{2}$
within	1	.	.	.
DZ between	1	2	$\frac{3}{4}+\frac{A}{1-A}$	$\frac{5}{16}$
within	1	.	$\frac{1}{4}$	$\frac{3}{16}$

If a simple explanation in terms of say, two parameters, fails to account for the observations we shall be forced to admit a more complicated model. However, different three parameter models which assume equality of σ^2_{TMZ} and σ^2_{TDZ} are inseparable by a simple statistical test of the model with twin data so we shall be unable to discriminate between alternatives unless one gives clearly nonsensical parameter values.

As a first step in the analysis of any set of twin data we routinely fit the following set of models, each of which is a subset of the basic model: E_1, E_1E_2, E_1D_R, $E_1E_2D_R$, $E_1D_RH_R$. Occasionally, if we have reason to believe that assortative mating is important we note that, in the basic model for twins reared together, E_2 and the term $\frac{1}{2}D_R(A/(1-A))$ have the same coefficients and so are inseparable. Thus, in fitting the $E_1E_2D_R$ model we should really rename E_2 as B (for extra between families variation) where

$$B = E_2+\tfrac{1}{2}D_R\left(\frac{A}{1-A}\right)$$

If we assume $E_2 = 0$ we may then obtain an estimate of A (which may or may not be " sensible "). Alternatively, we may fit an E_1D_RA model by non-linear weighted least squares (see Eaves, 1975) and obtain a direct estimate of A, assuming E_2 is absent. If we have an estimate of the phenotypic marital correlation μ, we may estimate A as $h^2\mu$ and hence estimate how much of $\hat{B}$ is E_2 and how much is extra additive variance produced by the linkage disequilibrium which accrues from assortative mating.

The restricticton to three parameter estimates means that we can never tell directly the relative importance of B and H_R and we can see that it is not possible to separate their contributions if both are present. Eaves (1970a) showed that if we fit the E_1BD_R model when both B and H_R are present then

$$\hat{D}_R = D_R+\tfrac{3}{4}H_R$$

and

$$\hat{B} = B-\tfrac{1}{8}H_R$$

Conversely, if an $E_1D_RH_R$ model is fitted when both B and H_R are present then

$$\hat{D}_R = D_R+6B$$

and

$$\hat{H}_R = H_R-8B$$

Thus, if $B > \frac{1}{8}H_R$ then any estimates of H_R will be negative while if $B < \frac{1}{8}H_R$ any estimates of B will be negative. By fitting these two models and com-

paring the signs of the third parameter estimate we should be able to make some weak inferences about the presence of E_2 (or assortative mating) and H_R. These inferences are discussed in detail by Eaves (1970*a*). It can be seen, however, that this is a poor design for the detection of non-additive genetical variation and this will be confirmed by the power calculations below.

Even where sources of variation are potentially separable by the twin design there is no point in carrying out a twin study if the sample size is too small to give a reasonable probability of discriminating between alternative models of variation. This is the problem of power, the probability of correctly rejecting the null hypothesis when it is false. If the power of a study to detect a given effect is low and in fact we do not find evidence for the effect in our sample then we should be foolish to infer that the effect is not present in the population. Such errors are all too common in the literature (*e.g.* Scarr-Salapatek, 1971; Adams, Ghodsian and Richardson, 1976).

This consideration was first quantified in biometrical genetics by Kearsey (1970) who calculated the sample sizes required to detect dominance with different experimental designs. In an attempt to introduce an element of rational planning to investigations of individual differences, Eaves considered the relative merits of the " minimal data set " designs laid down by Jinks and Fulker (1970) which would allow separation of the major sources of variance. He calculated the proportions of the different relationships which would yield the maximum information in the separation of additive from dominance variance (Eaves, 1969) and E_1 from E_2 variance (Eaves, 1970*b*). Later he went on to consider the power of the different designs and the sample sizes required to detect effects of given magnitude from the minimal data sets (Eaves, 1972). The power of a classical twin study without diagnosis of zygosity of the same sex pairs but which relied on the opposite sex pairs for an estimate of the DZ intra-class correlation, was also considered (Eaves and Jinks, 1972) and it was shown that sample sizes would need to be about three times bigger to achieve the same power as a sample in which zygosity was known. The power of comparisons between estimates of heritability from subsamples was shown to be negligible under such conditions.

The large samples required to allow comparisons of heritability between populations were confirmed by Klein (1974) who calculated the power of four relationships (offspring-midparent, offspring-single parent, full-sibs, half-sibs) to detect genetic components of given magnitude (see also Klein, DeFries and Finkbeiner, 1973).

So far then, some quite detailed theoretical power calculations have been made for human experimental designs which are seldom (if ever) used but nobody has given much consideration to the power of the classical twin study, the most common design in human biometrical genetics.

Because Eaves (1972) was considering designs in which the main sources of variation were at least potentially separable, his approach was to calculate the sample sizes required to detect given effects with a specified probability. In the classical twin study, we have seen that we can estimate at most three parameters and that these will be confounded with other effects which may be present but cannot be estimated. It will be more useful, therefore, to calculate sample sizes required to discriminate between competing models.

Initially we shall be concerned to answer fairly simple questions such as: " How many pairs do I need for an E_1 model to fail when D_R is also present? ", " What is the best proportion of MZ and DZ pairs to detect E_2 against a background of E_1 and D_R? " and we shall leave the complications of non-additivity and directional effects until later. Our first approach is based on the non-central chi-square distribution and was suggested by power calculations for the rejection of the null hypothesis of Hardy-Weinberg equilibrium (Lewontin and Cockerham, 1959).

2. The power of discriminating between simple models

(i) *Theory*

If we have a set of observed mean squares 0_i, their expected values calculated on the basis of the " true " model of variation F_i and the expected values calculated on the basis of a " false " model, E_i, then we wish, for each i to test the null hypothesis

$$H_0: \mathscr{E}(0_i) = E_i$$

If H_0 is true and the degrees of freedom ν_i are large then $0_i \sim N(E_i, 2E_i^2/\nu_i)$ approximately. We assume H_0 is to be tested by the statistic $\dfrac{\nu_i(0_i-E_i)^2}{2E_i^2}$ which is approximately chi-square with one degree of freedom. The power function of this test is not known for all alternative hypotheses. Consequently we follow Mitra (1958) and consider the limiting power function of the test for large sample sizes and alternative models not too far from the hypothesised model. To express this idea suppose

$$H_1: \mathscr{E}(0_i) = F_i = E_i + \mu_i/\sqrt{\nu_i}$$

where $\mu_i = \sqrt{\nu_i}(F_i - E_i)$ is the deviation between the two models. Given that ν_i is large, then if H_0 is true

$$0_i \sim N(E_i, 2E_i^2/\nu_i)$$

while if H_1 is true then

$$0_i \sim N(F_i, 2F_i^2/\nu_i) \sim N(F_i, (2E_i^2/\nu_i)(1+o(\nu_i^{-\frac{1}{2}}))).$$

where $o(\nu_i^{-\frac{1}{2}})$ denotes a term of the order of $\nu_i^{-\frac{1}{2}}$.
Thus

$$\frac{\sqrt{\nu_i}(0_i-E_i)}{\sqrt{2E_i^2}} \sim N\left(\frac{\sqrt{\nu_i}(F_i-E_i)}{\sqrt{2E_i^2}},\ 1+o(\nu_i^{-\frac{1}{2}})\right)$$

and the asymptotic power function of

$$\frac{\nu_i(0_i-E_i)^2}{2E_i^2}$$

is non-central chi-square with non-centrality parameter $\dfrac{\nu_i(F_i-E_i)^2}{2E_i^2}$ and one degree of freedom. In general, to test

$$H_0: \mathscr{E}(\mathbf{0}) = \boldsymbol{E} \text{ against}$$
$$H_1: \mathscr{E}(\mathbf{0}) = \boldsymbol{F}$$

we use

$$\chi^2_{s-p} = \sum \frac{v_i(0_i - E_i)^2}{2E_i^2} \quad (s \text{ statistics, } p \text{ parameters}) \tag{1}$$

with limiting power function being non central χ^2_{s-p} with non-centrality parameter

$$\lambda = \sum \frac{v_i(F_i - E_i)^2}{2E_i^2}$$

The larger the d.f. v_i, the larger the deviations μ_i may be from E_i before the distribution departs from non-central chi-square.

(ii) *Calculation of required sample size*

Given that we know the true components of variation contributing to the population variance we want to know how many pairs of twins are needed to be (95 per cent) sure of rejecting alternative wrong models (at the 5 per cent level of significance). Our procedure is as follows:

(a) Take the " right " model and obtain the expected " observed mean squares " for given values of population parameters.

(b) Obtain the WLS solutions (for unit total sample size) for the parameters of the " wrong " model.

(c) Obtain the weighted residual sum of squares (as in equation (1) above) for the fit of the " wrong " model and use this as a non-centrality parameter λ' for $s-p$ d.f. where there are s statistics and p parameters in the " wrong " model.

(d) Look up the non-centrality parameter in tables (Pearson and Hartley 1972; Vol. 2, table 25) for required power, *i.e.* $\lambda_{i(0\cdot05, 0\cdot95,\ s-p)}$.

(e) Obtain required sample size

$$N = \frac{\lambda}{\lambda'}$$

To illustrate the procedure we shall work through a specific example, but see Eaves and Eysenck (1975) for a detailed description of model fitting by weighted least squares.

For a total population variance of 1·0, let us suppose that the real components of variation are

$$\Phi = \begin{pmatrix} E_1 \\ V_A \end{pmatrix} = \begin{pmatrix} 0\cdot5 \\ 0\cdot5 \end{pmatrix}$$

i.e. half E_1 and half additive genetic ($V_A = \frac{1}{2}D_R$). We shall consider an experiment where half the total sample is MZ and half DZ.

$$\text{i.e.} \quad N_{mz}/N = N_{dz}/N = 0\cdot5$$

so each weight

$$w_{ii} = \frac{0\cdot5N}{2x_i^2} = \frac{N}{4x_i^2}$$

The " right " E_1V_A model matrix is

$$\boldsymbol{B} = \begin{pmatrix} 1 & 2 \\ 1 & 0 \\ 1 & 1{\cdot}5 \\ 1 & 0{\cdot}5 \end{pmatrix}$$

and the expected " observed " statistics

$$\boldsymbol{x} = \boldsymbol{B\Phi} = \begin{pmatrix} 1{\cdot}50 \\ 0{\cdot}50 \\ 1{\cdot}25 \\ 0{\cdot}75 \end{pmatrix}$$

We now try to fit the " wrong " E_1E_2 model to the data.

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ 1 & 0 \\ 1 & 2 \\ 1 & 0 \end{pmatrix}$$

and after four iterations, the final solution

$$\hat{\boldsymbol{\Theta}} = (\boldsymbol{A'WA})^{-1}\boldsymbol{A'Wx} = (\boldsymbol{A'WA})^{-1}\boldsymbol{A'WB\Phi} = \begin{pmatrix} \hat{E}_1 \\ \hat{E}_2 \end{pmatrix} = \begin{pmatrix} 0{\cdot}625 \\ 0{\cdot}375 \end{pmatrix}$$

from which we can calculate the expected mean squares and residual S.S. as follows.

M.S.	N_i/N	Weight	" Observed "	Expected
MZ_b	0·5	0·1322	1·50	1·375
MZ_w	0·5	0·6400	0·50	0·625
DZ_b	0·5	0·1322	1·25	1·375
DZ_w	0·5	0·6400	0·75	0·625

$$\text{" } \chi^2_{(2)} \text{ "} = \lambda'_2 = 0{\cdot}0241$$

The non-centrality parameter obtained for unit sample size in this experiment is $\lambda' = 0{\cdot}0241$. From the table we find the non-centrality parameter

$$\lambda_{(0{\cdot}05,\, 0{\cdot}95,\, 2)} = 15{\cdot}443$$

Therefore the sample size required for 95 per cent power of rejecting the " wrong " model at the 5 per cent level of significance in this experiment is

$$N = \frac{\lambda}{\lambda'} = \frac{15{\cdot}443}{0{\cdot}0241} = 640$$

i.e. 320 pairs of MZ twins and 320 pairs of DZ twins.

Since

$$\hat{\boldsymbol{\Theta}} = (\boldsymbol{A'WA})^{-1}\boldsymbol{A'WB\Phi}$$

the matrix $\boldsymbol{R} = (\boldsymbol{A'WA})^{-1}\boldsymbol{A'WB}$ will tell us the contribution of each of the " true " population parameters ($\boldsymbol{\Phi}$) to each of the estimated parameters of the " wrong " model. For this example

$$\boldsymbol{R} = \begin{pmatrix} 1{\cdot}00 & 0{\cdot}25 \\ 0{\cdot}00 & 0{\cdot}75 \end{pmatrix}$$

so

$$\hat{E}_1 = E_1 + \tfrac{1}{4}V_A$$
$$\hat{E}_2 = \tfrac{3}{4}V_A$$

(iii) *Testing the accuracy of the method by a simulation experiment*

It has been shown that the validity of the method depends upon the magnitude of the deviations, μ_i of the " false " expected values E_i from the " true " expected values, F_i.

In order to test this validity we simulated many replicates of the same experimental design and so compared the observed power of rejection at a given level of significance with that expected from the calculation based on non-central chi-square.

It was decided to simulate the above experiment but with a sample size predicted to give 50 per cent power of rejection of the E_1E_2 model at the 5 per cent level of significance. From the table

$$\lambda_{(0\cdot05,\ 0\cdot50,\ 2)} = 4\cdot957$$

so

$$N_{(0\cdot05,\ 0\cdot50,\ 2)} = \frac{4\cdot957}{0\cdot0241} = 206$$

i.e. 103 pairs of MZ and 103 pairs of DZ twins.

The simulation program actually samples individual pairs of twins from a population with specified components of total variance. It was adapted from a program written by one of us (M. J. K.) for the simulation of various aspects of polygenic inheritance and experimental design in biometrical genetics. Genetical variation is simulated on the basis of two alleles at each of 10 loci of equal effect which are in linkage equilibrium. The dominance effects of the loci and frequencies of their alleles can be varied. Two parental genotypes are sampled at random from the population and from these zygotes are sampled—one for MZ twins and two for DZ twins. A common between families environmental deviation is added to the genotypic score of both twins and a separate within families environmental deviation is sampled for each individual thus completing the phenotypic score. The scores for the samples are scaled to have mean 100 and variance 225.

For each set of the 103 twin pairs, MZ or DZ, so generated we computed the within and between pairs mean squares. We simulated 500 experiments each of 103 MZ and 103 DZ pairs drawn from a population in which $E_1 = V_A = 0\cdot5$. To each of the 500 sets of four meansquares we fitted the E_1E_2 model and calculated the residual $\chi^2_{(2)}$.

Pearson and Hartley (1972, Vol. 2, p. 54) give the expectations for the moments of the non-central chi-square distribution. These are given in table 2 with the expected and observed values of the moments for the distribution of the 500 simulated $\chi^2_{(2)}$'s. The observed mean, variance and skewness do not differ significantly from their expected values for the expected non-centrality parameter. The kurtosis is less than expected but, since the variance of the estimate of kurtosis is not available, we cannot say whether significantly so.

TABLE 2

Observed and expected moments for the distribution of the 500 $\chi^2_{(2)}$ values

Statistic	Expectation	Expected	Observed
Mean	$\nu+\lambda$	6·957	6·657
Variance	$2(\nu+2\lambda)$	23·828	21·112
Skewness	$\frac{8(\nu+3\lambda)^2}{(\nu+2\lambda)^3}$	1·346	1·351
Kurtosis	$3+\frac{12(\nu+4\lambda)}{(\nu+2\lambda)}$	4·845	3·561

ν = d.f. = 2.
λ = is the expected non-centrality parameter = 4·957.

As a more direct test of the hypothesis that the residual chi-squares follow the expected non-central chi-square distribution, we have predicted, on the basis of the expected distribution, that 50 per cent of the observed residual chi-squares should be significant at the 5 per cent level, *i.e.* $\chi^2_{(2)} > 5{\cdot}991$. In fact we find the following:

	$\chi^2_{(2)} < 5{\cdot}991$	$\chi^2_{(2)} > 5{\cdot}991$
Observed	257	243
Expected	250	250

$$\chi^2_{(1)} = 0{\cdot}4$$

so the observed power of the test is remarkably close to the predicted.

However, we should like to know whether the observed distribution of the chi-squares is in agreement with the predicted distribution throughout the whole range. From Pearson and Hartley (1972, Vol. 2, Table 24), we can interpolate the 1 per cent, 5 per cent, 95 per cent and 99 per cent points of the expected distribution of $\chi^2_{(\nu,\alpha,\lambda)}$ where $\nu = 2$, $\alpha = 0{\cdot}05$ and the non-centrality parameter $\lambda = 4{\cdot}957$. These points are:

β	1%	5%	50%	95%	99%
χ^2	0·21	0·88	5·99	16·15	21·96

and the observed and expected numbers in the intervals are given in table 3. The observed and expected distributions agree well suggesting that the observed distribution follows the expected throughout the range.

TABLE 3

Distribution of the 500 simulated $\chi^2_{(2)}$'s

Interval %	<1	1-5	5-50	50-95	95-99	>99	Total
Observed	2	21	234	222	18	3	500
Expected	5	20	225	225	20	5	500

$$\chi^2_{(5)} = 3{\cdot}25.$$

This simulation experiment should give us considerable confidence in the ability of the non-central chi-square calculations to predict the power of a given experiment and the sample size required for a given probability of rejection of the sorts of models we are considering. However, outside the range of comparable models, the theoretical distribution obtained by limiting power may not apply to a good approximation if $\mu_i = \sqrt{\nu_i}(F_i - E_i)$ is large. Although we shall calculate sample sizes required for 95 per cent power, the numbers required for lower power (β), or the approximate power obtained by using lower numbers, can easily be obtained by observing that

$$N_{(\beta)} = N_{(0\cdot95)}\left(\frac{\lambda_{(\alpha,\,\beta,\,\nu)}}{\lambda_{(\alpha,\,0\cdot95,\,\nu)}}\right)$$

The values of $\lambda_{(\alpha,\,\beta,\,\nu)}$ for $\alpha = 0{\cdot}05$ can be obtained from table 25 in Pearson and Hartley (1972, Vol. 2).

(iv) *Results*

We shall consider sample sizes required to discriminate between very simple models involving E_1, E_2 (including additive variance due to assortative mating) and V_A. The problem of detection of dominance will be left until the next section. We shall calculate sample sizes required to reject false hypotheses in the following combinations of " true " and " false " models of population variance:

" True " model	" False " model
E_1E_2	E_1
	E_1V_A
E_1V_A	E_1
	E_1E_2
$E_1E_2V_A$	E_1
	E_1E_2
	E_1V_A

For the " true " E_1E_2 and E_1V_A models we shall calculate sample sizes for $E_1 = 0{\cdot}1 - 0{\cdot}9$ (in steps of 0·2) and for the $E_1E_2V_A$ model, all combinations of $E_1 = 0{\cdot}1 - 0{\cdot}9$ (0·2) and $E_2 = 0{\cdot}1 - 0{\cdot}9$ (0·2) with V_A forming the non-zero remainder. For each combination of " true " parameters, calculations will be made for samples comprising a proportion of MZ (pMZ) twins in the range 0·1-0·9 (0·2).

The required sample sizes for the " true " E_1E_2 and E_1V_A models form tables 4a and b respectively and those for the three " false " models fitted to the " true " $E_1E_2V_A$ model are found in table 5.

Where the situation is really E_1E_2 or E_1V_A, the sample sizes required to reject the E_1 model are not very large, provided E_1 does not exceed 70 per cent of the total variance. Against a background of additive genetic variance this becomes easier the larger the proportion of MZ twins in the sample. Since the expectations for MZ and DZ mean-squares are the same when an E_1E_2 model is appropriate, the proportion of MZ pairs in this experiment is immaterial.

TABLE 4

Total number of pairs required for 95 per cent power of rejection of false hypotheses at 5 per cent level when "true" model is (a) E_1E_2, (b) E_1V_A

(a) True model E_1E_2

True model		False Model: E_1 pMZ					False Model: E_1V_A pMZ				
E_1	E_2	0·1	0·3	0·5	0·7	0·9	0·1	0·3	0·5	0·7	0·9
0·1	0·9	22	22	22	22	22	33	45	67	119	385
0·3	0·7	36	36	36	36	36	73	115	164	278	854
0·5	0·5	69	69	69	69	69	298	325	430	696	2055
0·7	0·3	191	191	191	191	191	1491	1229	1485	2289	6534
0·9	0·1	1718	1718	1718	1718	1718	20904	13508	15119	22534	62948

(b) True model E_1V_A

True model		False Model: E_1					False Model: E_1E_2				
E_1	V_A										
0·1	0·9	66	45	34	28	23	388	118	63	40	36
0·3	0·7	108	74	57	46	38	886	313	208	186	303
0·5	0·5	212	145	110	89	75	2181	852	640	670	1344
0·7	0·3	588	402	306	247	207	7026	2914	2356	2683	5955
0·9	0·1	5284	3615	2748	2216	1857	68016	28982	24232	28784	66800

One of the most common problems in the analysis of twin data is deciding whether the E_1E_2 or the E_1V_A model is the more appropriate. For a given E_1 we can see that in the most likely cases it is usually easier to reject an E_1V_A model when the situation is really E_1E_2 than to reject an E_1E_2 model when additive genetic variation forms the non-E_1 variance. A realistic requirement might be to discriminate between the two models when half the variance is E_1 and the remainder is either E_2 or V_A. In this case we should need 430 pairs to reject the E_1V_A model when the remaining variance is E_2 but half as many pairs again—640—to reject the E_1E_2 model when the population variance is really half E_1 and half V_A. Thus for a given sample size, the E_1E_2 model will more often be incorrectly accepted as an adequate hypothesis than the E_1V_A model, *i.e.* the twin method is inherently biased against the detection of genetical rather than cultural variance.

The optimum proportion of MZ twins in the sample varies with the "right" and "wrong" model and the "true" parameter values. In some cases it is better to have a low proportion of MZ pairs, in other cases a high proportion and in others the optimum is equal or near equal proportions. Since rejection of a false E_1E_2 model is the more difficult objective, we should settle on the proportion of MZ pairs which will achieve this most efficiently and, for intermediate heritabilities, this appears to be about equal proportions of MZ and DZ pairs.

Turning now to the rejection of inadequate models for a situation in which the appropriate model is really $E_1E_2V_A$ (table 5) we can see once again, that rejection of the E_1 model is reliable in all cases with fairly small samples. Rejection of the two-parameter models is more difficult and will depend upon the proportions of E_2 and V_A. General trends from the table

TABLE 5

Total number of pairs required for 95% power of rejection of false hypotheses at 5% level when the " true " model is $E_1E_2V_A$

True model			False model													
			E_1 pMZ					E_1E_2 pMZ					E_1V_A pMZ				
E_1	E_2	V_A	0·1	0·3	0·5	0·7	0·9	0·1	0·3	0·5	0·7	0·9	0·1	0·3	0·5	0·7	0·9
0·1	0·1	0·8	57	42	33	27	23	417	126	68	44	42	4126	5086	6943	11315	33124
	0·3	0·6	43	35	30	26	23	497	152	85	59	65	417	521	725	1207	3625
	0·5	0·4	33	30	27	25	23	656	208	123	94	124	134	173	245	417	1285
	0·7	0·2	27	25	24	23	22	1226	422	277	248	414	61	81	117	204	646
0·3	0·1	0·6	89	67	53	44	38	1070	381	257	235	400	6835	6899	9031	14553	42733
	0·3	0·4	63	54	47	42	37	1798	660	466	455	848	645	718	966	1583	4715
	0·5	0·2	47	44	41	39	36	4931	1915	1449	1542	3191	200	242	335	558	1685
0·5	0·1	0·4	162	125	101	85	74	3136	1234	940	1002	2065	11204	9230	11458	17929	51800
	0·3	0·2	102	92	84	77	72	10167	4124	3268	3567	7990	997	963	1233	1963	5732
0·7	0·1	0·2	382	313	265	229	202	15017	6265	5110	5892	13278	16512	11720	13720	20814	58596

are too numerous to discuss so we shall concentrate on one specific practical example. There are now three twin studies (see Martin and Eysenck, 1976) which suggest that variation for the attitude trait Radicalism is approximately one-third E_1, one-third V_A and one-third E_2 (possibly including assortative mating). We should like to know how many pairs, and of what composition, are required to best reject both the E_1E_2 and E_1V_A models as summaries of this more complex situation. The case corresponding most closely to this in the table comprises $0{\cdot}3E_1$, $0{\cdot}3E_2$ and $0{\cdot}4V_A$ and we can see that the best individual solutions are 455 pairs (70 per cent MZ) to reject the E_1E_2 model at the 5 per cent level in 95 per cent of experiments and 645 pairs (10 per cent MZ) to reject the E_1V_A model. With about 700 pairs (660-718) comprising 30 per cent MZ pairs we should have about a 95 per cent chance of rejecting both two-parameter models and this would be the best compromise design. This would be ideal in a random sample since MZ pairs represent about one-third of all twins in the population. However, they represent nearer two-thirds of the total in most volunteer samples.

These " required number " tables should be a useful tool for workers wishing to plan twin studies with specified power to detect given combinations of genetical and environmental variance. The numbers themselves are sobering for we can see that to achieve 95 per cent power of rejecting an E_1E_2 model when the heritability is actually 50 per cent requires over 600 twin pairs. We should not be too depressed, however, since twin studies of this size are now becoming quite common and $\beta = 95$ per cent is a powerful experiment. From the method given above we can calculate that for 80 per cent power we would need about two-thirds this sample size and for 50 per cent power about one-third this size.

3. The detection of non-additive and directional effects

So far we have only considered the power of models to detect additive genetical variance, E_1 and E_2/assortative mating. We have seen that the prospects for detection of dominance by model fitting to the mean-squares of a classical twin study are poor if there is any E_2 or assortative mating present. This is unfortunate, for a knowledge of the extent and directional properties of non-additive variation allows us to make inferences about the selective history of a trait (Mather, 1973). However, it is possible to detect non-additivity which has a systematic directional effect by means other than model-fitting.

Jinks and Fulker (1970) introduced the test of regressing MZ pair variances on the corresponding pair means in order to detect systematic $G \times E_1$. The same test can be used in DZ twins to detect $G \times E$ and other non-additive effects acting predominantly in one direction. Such directional effects are often reflected in and can therefore be inferred from skewness of the phenotypic distribution.

Skewness in the distribution and mean-variance regression in DZ twins can be produced by (a) systematic $G \times E$ interaction, (b) genetical non-additivity (dominance or epistasis) acting predominantly in one direction, (c) a net inequality in the frequencies of increasing and decreasing alleles. We shall examine each of these in turn. The skewness and mean-variance covariance tests have general properties that extend beyond DZ twins and

have been used in other genetical designs to detect $G \times E$ and epistasis (Perkins and Jinks, 1971; Pooni, Jinks and Cornish, 1977).

(i) *$G \times E$ interaction*

Jinks and Fulker (1970) showed that the regression of pair variances on pair means in MZ twins reared together is a test for systematic $G \times E_1$ or $E_2 \times E_1$ interaction. The regression can contain non-linear as well as linear terms but for practical purposes there seems little point in going beyond quadratic terms. When such interactions are detected the worker will sometimes choose to simplify the analysis by rescaling the data to remove the interaction. Often, however, there may be some sound reason for keeping the chosen scale and accepting the complication of the interaction that this entails (*e.g.* Martin, Eaves and Eysenck, 1977). The power of the regression test for $G \times E_1$ and the rationale of rescaling have been discussed extensively elsewhere (Eaves, Last, Martin and Jinks, 1977). We have found that systematic $G \times E_1$ accounting for as little as 5 per cent of the MZ within pairs variation can be detected with only 95 pairs of twins (Martin, 1977).

Within DZ pairs reared together, $G \times E_2$ interactions may be found as well as the $G \times E_1$ and $E_1 \times E_2$ found in MZ pairs. Thus, even if $G \times E_1$ and $E_2 \times E_1$ interactions are removed by scaling, any remaining mean-variance regression in DZ pairs could be due to systematic $G \times E_2$.

(ii) *Dominance*

There are two tests for directional dominance (or epistasis) in the absence of $G \times E$ which can be applied to data from the classical twin study: sample skewness and regression of DZ pair variances on pair means. These follow closely from the third degree statistics considered by Fisher, Immer and Tedin (1932). We wish to find the power of these two tests, either by theoretical calculation or empirically by simulation, and to compare these with the power of detection of dominance by model fitting to second degree statistics.

In a population with equal gene frequencies and in the absence of $G \times E$, and epistasis, the skewness of the phenotypic distribution due to dominance at a single locus

$$g_1 = \frac{-\frac{3}{4}d_i^2 h_i}{(\sigma_i^2)^{\frac{3}{2}}}$$

where $\sigma_i^2 = \frac{1}{2}d_i^2 + \frac{1}{4}h_i^2$ (ignoring any environmental effects). Note that if h is positive (*i.e.* dominance is in the increasing direction) then the skewness is negative and vice versa. For k loci of equal effect and with alleles of equal frequency

$$g_1 = \frac{-\frac{3}{4}kd_i^2 h_i}{(k\sigma_i^2)^{\frac{3}{2}}}$$

so that the coefficient of skewness is inversely proportional to $k^{\frac{1}{2}}$ and will tend to zero as the number of loci increases. Nevertheless, we shall calculate the power of detection of skewness for the moderately large number of 10

loci which are specified in the simulation program. For total variance $\sigma^2(= k\sigma_i^2 + E_1 + E_2) = 225$, broad heritability $h_B^2 = 0.9$, dominance ratio $h_i/d_i = 0{\cdot}5$ and 10 loci, we calculate $d_i = 6{\cdot}0$ and $h_i = 3{\cdot}0$. So

$$\mathscr{E}(g_1) = \frac{-\frac{3}{4}\sum_{i=1}^{10} d_i^2 h_i}{\sigma^3} = \frac{-810}{3375} = -0{\cdot}240$$

Similarly for

$$h_B^2 = 0{\cdot}5 \quad \text{and} \quad h_i/d_i = 1{\cdot}0, \quad \mathscr{E}(g_1) = -0{\cdot}13$$

The variance of a coefficient of skewness is approximately $6/n$ given a normally distributed population so for sample size 500

$$\sigma_{g_1} = \sqrt{\frac{6}{500}} = 0{\cdot}11$$

and for sample size 1000

$$\sigma_{g_1} = \sqrt{\frac{6}{1000}} = 0{\cdot}08.$$

However, given that we are sampling from non-normal populations we might expect that the expected variances, and hence the theoretical power calculations, are inappropriate. In practice we find that the empirical standard errors (s_{g1}) are close to their expected values so we may be reasonably confident that the non-zero skewness is not appreciably affecting the power calculations.

Since we can obtain both positive and negative values of g_1, we shall only accept that there is significant (5 per cent level) skewness if

$$x = \frac{\hat{g}_1}{\sigma_{g_1}} > 1{\cdot}96 \quad \text{or} \quad x < -1{\cdot}96$$

For given sample size we want to find the probability that the null hypothesis, $x = 0$ will be rejected.

The power of the test is the probability that $\hat{\theta}$ (here g_1) $> 1{\cdot}96\sigma$ or $\hat{\theta} < -1{\cdot}96\sigma$ given that $\theta = 0$. Hence the power of the test is the area under the curve for a variable $\hat{\theta}$ which is $N(\theta, \sigma^2)$ corresponding to values of $\hat{\theta} > 1{\cdot}96\sigma$ or $\hat{\theta} < -1{\cdot}96\sigma$. If $c = (\hat{\theta} - \theta)/\sigma$ so that c is $N(0, 1)$ when $\hat{\theta} = 1{\cdot}96\sigma$ then $c = 1{\cdot}96 - x$, or when $\hat{\theta} = -1{\cdot}96\sigma$, $c = -1{\cdot}96 - x$. Thus the power of the test is the area under the curve for a variable c which is $N(0, 1)$ for values of $c > 1{\cdot}96 - x$ or $c < -1{\cdot}96 - x$. If $\theta \neq 0$ then $c > 1{\cdot}96 - x$ is the probability that we detect $\hat{\theta} > 0$ on a two-tail test and $c < -1{\cdot}96 - x$ is the probability that we detect $\hat{\theta} < 0$.

In the present example

$$x = \frac{-0{\cdot}240}{0{\cdot}08} = -3{\cdot}0$$

so the probability of detecting significant negative skewness as a two-tail (5 per cent) test is the integral from $-\infty$ to $-1{\cdot}96 - x\,(= 1{\cdot}04)$, *i.e.* P = 0·85. The chance of (wrongly) detecting significant positive skewness is less than

one in a million. If we require the power of the test to be 0·95, then we want $-1{\cdot}96-x = 1{\cdot}65$, *i.e.* $x = -3{\cdot}61$ whence

$$s_x = \frac{-0{\cdot}24}{-3{\cdot}61} = 0{\cdot}0664$$

so we require $n = 1364$, or nearer 700 pairs of twins.

Thus it is possible to calculate from theoretical considerations the power of the test for skewness (and so for dominance) for given sample size and hence the sample size required for given power. We can compare the theoretically derived power with that found empirically by simulation.

In the mean-variance regression test, on the other hand, the regression coefficient b is not dependent on k, the number of loci. It can be shown that, for equal allele frequencies and no $G \times E$, the contribution of a single locus to the expected regression coefficient of DZ pair variances on pair means

$$b_i = \frac{\frac{-7}{32} d_i^2 h_i}{\frac{3}{8} d_i^2 + \frac{5}{32} h_i^2} \quad \text{(ignoring environmental effects)}$$

and that for loci of equal effect and alleles of equal frequency this will be a constant, independent of the number of loci. Similar considerations apply if the mean-variance correlation is chosen as the summary statistic.

Table 6 shows the results of computer-simulated sampling of twins from populations with equal gene frequencies but of differing broad heritabilities (h_B^2), different degrees of dominance (h_i/d_i) and different proportions of E_1 and E_2 (E_1/V_E where $V_E = E_1 + E_2$). For each combination of values of these three parameters 25 replicates of 500 MZ and 500 DZ twin pairs have been sampled.

The theoretical value of $g_1(\mathscr{E}(g_1))$ is given for each combination of heritability and dominance ratio. The expected power of detection of significant g_1 is given separately for MZ (based on $n = 500$) and DZ ($n = 1000$) twins. The mean g_1 ($\bar{g}_1$) and its observed standard error from the 25 replicates (s_{g1}) is given for both MZ and DZ twins. The number of replicates out of 25 where significant (5 per cent, two-tail test) g_1, is detected is given: the theoretical s.e. of 0·08 for DZ and 0·11 for MZ is used in judging the significance. It can be seen that the observed numbers of significant g_1's (β_{gi}) agree well with the theoretical power of the tests ($\mathscr{E}(\beta_{g1})$).

For each replicate the within pair variances were regressed on pair means for both MZ and DZ twins. Both linear and second degree regressions were carried out but the number of significant quadratic regressions was at chance level in MZ twins and little above chance level in DZ twins so only the number of significant (5 per cent) linear regression coefficients out of 25 replicates is given for MZ and DZ twins (β_b). These constitute empirical estimates of the power of this test of directional non-additive genetical effects for different population parameter values.

Finally, for each replicate of 500 pairs of MZ and 500 pairs of DZ twins, mean-squares were calculated and two and three parameter models were fitted to the data. The E_1E_2 model failed in all but one of the 200 replicates

TABLE 6

The results of simulation of 25 replicates of samples of 500 MZ and 500 DZ twin pairs of three tests for dominance: skewness, DZ mean-variance regression and model fitting. See text for explanation

				MZ					DZ					E_1D_R model	$E_1D_RH_R$ model
h_B^2	h_i/d_i	E_1/V_E	$\varepsilon(g_1)$	$\tilde{g}_1$	$S\hat{g}_1$	$\beta(\hat{g}_1)$	$\varepsilon(\beta g_1)$	$\beta(b)$	$\tilde{g}_1$	$S\hat{g}_1$	$\beta(\hat{g}_1)$	$\varepsilon(\beta g_1)$	$\beta(b)$	fails	$\lvert \hat{H}_R \rvert > 1{\cdot}65$
0·9	1·0	1·0	−0·31	−0·28	0·08	17	0·81	2	−0·31	0·07	24	0·97	21	12	20+
		0·5	−0·31	−0·32	0·09	22	0·81	2	−0·30	0·06	25	0·97	21	2	7+
	0·5	1·0	−0·24	−0·27	0·08	18	0·59	1	−0·24	0·08	21	0·85	14	5	7+
		0·5	−0·24	−0·24	0·09	15	0·59	1	−0·22	0·08	20	0·85	13	3	2+ 5−
0·5	1·0	1·0	−0·13	−0·13	0·12	6	0·21	4	−0·14	0·07	12	0·36	4	2	3+
		0·5	−0·13	−0·13	0·09	6	0·21	1	−0·13	0·07	8	0·36	7	18	22−
	0·5	1·0	−0·10	−0·12	0·08	3	0·15	1	−0·09	0·05	3	0·24	0	2	1+ 2−
		0·5	−0·10	−0·07	0·08	1	0·15	2	−0·09	0·09	4	0·24	3	20	25−

listed in table 6. However, for each combination, the number of replicates out of 25 in which the E_1D_R model failed is given. The three parameter models nearly always fit since the total variances of MZ and DZ twins are equal within sampling error. For the $E_1D_RH_R$ model, the number of cases in which $|\hat{H}_R| > 1{\cdot}65\, s_{HR}$ is shown, divided into significant positive estimates of H_R and significant negative estimates which indicate the presence of $E_2 > \frac{1}{8}H_R$.

It can be seen that no amount of genetical non-additivity causing distributional skewness produces mean-variance regressions in MZ twins at above the level of chance—almost exactly 5 per cent of the regression tests in MZ twins are significant at the 5 per cent level. The sampling skewness coefficients agree very well with the predicted values, and will be detected with greater efficiency if in DZ twins we regard the sampling size as nearer twice the number of pairs than in MZ twins where this assumption is less justified.

For the case of high heritability and complete dominance the power of the test for skewness and the regression test in DZ twins is very high. Where there is no E_2 the model fitting approach also finds significant positive H_R with great efficiency. However, as soon as there is even a small amount of E_2 (here only 5 per cent of the total variance) the detection of H_R drops markedly. This is because, as we have already mentioned, in the $E_1D_RH_R$ model

$$\hat{H}_R = H_R - 8E_2.$$

Thus if there is any E_2 present there is considerable advantage to be gained in detecting directional dominance by using the skewness and regression tests, provided there is no $G \times E_2$ interaction. We predict from theory and it can be seen in the simulations that the skewness and regression tests are unaffected by the proportions of E_1 and E_2 which comprise the environmental variance V_E.

When the heritability is high and dominance is intermediate the power of the skewness test is about 80 per cent, of the DZ regression test about 60 per cent and of the detection of H_R by model fitting when no E_2 is present, about 30 per cent. For intermediate heritability, even when there is complete dominance there is some hope of detecting significant skewness but the power of all the other tests is little above chance level. However, we must emphasise again that, unlike the regression test, the power of the skewness test is dependent on the number of loci and will fall off as the number of loci increases.

We may examine a little more closely the power of the model fitting approach to detect dominance. In table 7 are listed the " true " values of each of the eight combinations of parameters we have simulated. The expected non-centrality parameters (calculated as shown above) for each combination based on sample size of 1000 pairs (50 per cent MZ) are given with the observed percentage powers (from the 25 replicates listed in table 6). It was found that, making allowance for the small number of replicates, the observed and expected powers correspond quite well.

Having calculated the non-centrality parameter we can obtain the sample sizes required to reject the E_1V_A model at the 5 per cent level in 95 per cent of experiments and these too are listed in table 7. Even for the most extreme case of 90 per cent broad heritability, complete dominance and no E_2 we

TABLE 7

Observed and expected power for rejection of E_1V_A *model in selected cases. Required sample size for 95 per cent rejection and contributions of population parameters to estimates are also given*

h_B^2	h_i/d_i	E_1/V_E	E_1	E_2	V_A^*	V_D^*	$\lambda_{(0\cdot05,\ \beta,1000)}$	Obs. β%	$N_{(\beta=0\cdot95)}$
0·9	1·0	1·0	22·5	0	135	67·5	4·6	48	3330
		0·5	11·25	11·25	135	67·5	2·2	8	7013
	0·5	1·0	22·5	0	180	22·5	0·5	20	29077
		0·5	11·25	11·25	180	22·5	0·0	12	—
0·5	1·0	1·0	112·5	0	75	37·5	0·9	8	16977
		0·5	56·25	56·25	75	37·5	6·0	72	2581
	0·5	1·0	112·5	0	100	12·5	0·1	8	151600
		0·5	56·25	56·25	100	12·5	10·1	80	1525

* $V_A = \frac{1}{2}D_R$, $V_D = \frac{1}{4}H_R$.

shall need 3330 pairs (half MZ and half DZ) to reject the E_1V_A model in 95 per cent of cases, and this rises to nearly 30,000 pairs with intermediate dominance. Eaves (1972) was able to show that, using optimal proportions of either of his " minimal data sets ", all four parameters E_1, E_2, D_R, H_R can be detected at the 5 per cent level of significance with 95 per cent certainty employing total sample sizes which are smaller than those required with the classical twin design to merely reject the E_1D_R model, underlining the inefficiency of the twin study for this purpose.

The presence of E_2 complicates the situation. When $h_B^2 = 0{\cdot}90$, $h_i/d_i = 1{\cdot}0$ and $E_1/V_E = 0{\cdot}5$ we see that $V_D = 67{\cdot}5$ is approximately equal to $8E_2 = 8 \times 11{\cdot}25 = 92$ so $|\hat{H}_R| = H_R - 8E_2$ is smaller than H_R and the E_1V_A model becomes more difficult to reject. In other cases, however, $|\hat{H}_R| > H_R$ and the E_1V_A model becomes easier to reject although the $\hat{H}_R$'s will be mainly negative reflecting the presence of E_2. In the eight cases simulated it was found that $\hat{E}_1$ was largely unconfounded with the other sources of variation whereas $\hat{V}_A$ contained any E_2 and V_D variation which was present.

In the case of $h_B^2 = 0{\cdot}9$, $h_i/d_i = 0{\cdot}5$ and $E_1/V_E = 0{\cdot}5$, there is a singularity since $\hat{V}_A = V_A + V_D + E_2$ and this leads to a perfect fit of the E_1V_A model.

In conclusion it seems that in the absence of $G \times E$ the regression of variance on mean for DZ pairs has a reasonable chance of detecting genetical non-additivity against a background of high heritability. However, against a background of intermediate heritability, even high levels of dominance have little chance of detection. Thus, for the sample size on which these calculations are based, which is not unrealistic there would be some point in looking at DZ mean-variance regressions in a trait like IQ thought to have high broad heritability and a reasonable degree of dominance (Eaves, 1973, 1975). However, there would be little chance of success with traits like personality which usually exhibit intermediate heritability and in which there is no evidence for dominance variation (Eaves and Eysenck, 1975; 1976; 1977; Martin and Eysenck, 1976).

(iii) *Unequal allele frequencies*

A net inequality in allele frequencies predominantly in either the increasing or decreasing direction will produce skewness in the phenotypic

distribution but, once again this will tend to disappear as k increases. However, for a single locus, with no $G \times E$ and no dominance the covariance of DZ mean and variance will be

$$-\tfrac{1}{2}uvd_i^3(u-v)$$

(where u is the frequency of the increasing allele and $u+v = 1$) and the DZ mean variance regression coefficient will be independent of k. Just as dominance in the increasing direction produces a negative regression, the increasing allele being more frequent will also produce a negative mean-variance regression.

Table 8 shows the results of simulation of samples of twins from populations with only additive and specific environmental variance but with differing gene frequencies and heritabilities. In a population with high heritability, a moderate difference (0·7 to 0·3) in frequencies of increasing and decreasing alleles will be detected by the test for skewness in about 50 per cent of cases and by a linear mean-variance regression in DZ twins in about 25 per cent of samples of this size. The same is true for extreme differences in gene frequencies in a population with intermediate heritability. In the less likely case of high heritability and extremely unequal gene frequencies, samples of this size will detect the effect with near complete certainty.

TABLE 8

The results of simulation of 25 replicates of samples of 500 MZ and 500 DZ twin pairs from a population with no dominance or E_2 but unequal gene frequencies. See text for explanation

			MZ					DZ				
h_B^2	u	$\varepsilon(g_1)$	$\bar{g}_1$	$S\hat{g}_1$	$\beta(\hat{g}_1)$	$\varepsilon\beta(g_1)$	$\beta(b)$	$\bar{g}_1$	$S\hat{g}_1$	$\beta(\hat{g}_1)$	$\varepsilon(\beta g_1)$	$\beta(b)$
0·9	0·5	0·00	0·00	0·10	0	0·00	0	0·00	0·09	0	0·00	1
	0·7	−0·17	−0·23	0·13	15	0·34	0	−0·18	0·09	15	0·57	7
	0·9	−0·51	−0·48	0·06	25	1·00	1	−0·49	0·07	25	1·00	25
0·5	0·5	0·00	−0·01	0·10	0	0·00	0	−0·01	0·06	0	0·00	2
	0·7	−0·07	−0·09	0·09	2	0·08	3	−0·05	0·07	2	0·14	1
	0·9	−0·21	−0·21	0·07	12	0·48	0	−0·21	0·08	17	0·75	6

As we predicted, unequal gene frequencies have no effect on the mean-variance regression in MZ twins although their phenotypic distribution is skewed.

Further simulation studies would determine the power of the classical twin study to detect more complex combinations of unequal gene frequencies and directional dominance.

Acknowledgments.—We are grateful to Professor J. L. Jinks for helpful discussions. This research was supported by an MRC programme grant in psychogenetics.

4. REFERENCES

ADAMS, B., GHODSIAN, M., AND RICHARDSON, K. 1976. Evidence for a low upper limit of heritability of mental test performcane in a national sample of twins. *Nature, 263*, 314-316.

EAVES, L. J. 1969. The genetic analysis of continuous variation: a comparison of experimental designs applicable to human data. *Brit. J. Math. Psychol. 22*, 131-147.

EAVES, L. J. 1970*a*. *Aspects of human psychogenetics*. Ph. D. Thesis, University of Birmingham.

EAVES, L. J. 1970*b*. The genetic analysis of continuous variation: a comparison of designs applicable to human data. II. Estimation of heritability and comparison of environmental components. *Brit. J. Math. Stat. Psychol., 23*, 189-198.

EAVES, L. J. 1972. Computer simulation of sample size and experimental design in human psychogenetics. *Pyschol. Bull. 77*, 144-152.

EAVES, L. J. 1973. Assortative mating and intelligence: an analysis of pedigree data. *Heredity, 30*, 199-210.

EAVES, L. J. 1975. Testing models for variation in intelligence. *Heredity, 34*, 132-136.

EAVES, L. J. 1976. A model for sibling effects in man. *Heredity, 36*, 205-214.

EAVES, L. J. 1977. Inferring the causes of human variation. *J. Roy. Stat. Soc.*, A., *140*, 324-365.

EAVES, L. J., AND EYSENCK, H. J. 1975. The nature of extraversion: a genetical analysis. *Journal of Personality and Social Psychology, 32*, 102-112.

EAVES, L. J., AND EYSENCK, H. J. 1976. Genetic and environmental components of inconsistency and unrepeatability in twins' responses to a neuroticism questionnaire. *Behav. Genet., 6*, 145-160.

EAVES, L. J., AND EYSENCK, H. J. 1977. *A genetic model for psychoticism. Behaviour Research and Therapy*. Advanced, *1*, 5-26.

EAVES, L. J., AND JINKS, J. L. 1972. Insignificance of evidence for differences in heritability of IQ between races and social classes. *Nature*, 240, 84-88.

EAVES, L. J., LAST, K., MARTIN, N. G., AND JINKS, J. L. 1977. A progressive approach to non-additivity and genotype-environmental covariation in the analysis of human differences. *Brit. J. Math. Stat. Psychol., 30*, 1-42.

FISHER, R. A., IMMER, F. R., AND TEDIN, O. 1932. The genetical interpretation of statistics of the third degree in the study of quantitative inheritance. *Genetics, 17*, 107-124.

JINKS, J. L., AND FULKER, D. W. 1970. Comparison of the biometrical genetical, MAVA and classical approaches to the analysis of human behavior. *Psychol. Bull., 73*, 311-349.

KEARSEY, M. J. 1970. Experimental sizes for detecting dominance. *Heredity, 25*, 529-542.

KLEIN, T. W. 1974. Heritability and sample size: statistical power, population comparisons and sample size. *Behavior Genetics, 4*, 171-189.

KLEIN, T. W., DEFRIES, J. C., AND FINKBEINER, C. T. 1973. Heritability and genetic correlation: standard errors of estimates and sample size. *Behavior Genetics, 3*, 355-364.

LEWONTIN, R. C., AND COCKERHAM, C. C. 1959. The goodness-of-fit test for detecting natural selection in random mating populations. *Evolution*, 13, 561-564.

MARTIN, N. G. 1975. The inheritance of scholastic abilities in a sample of twins. II. Genetical analysis of examination results. *Ann. Human Genet., 39*, 219-229.

MARTIN, N. G. 1977. *The classical twin study in human behavior genetics*. Ph.D. Thesis, University of Birmingham.

MARTIN, N. G., EAVES, L. J., AND EYSENCK, H. J. 1977. Genetical, environmental and personality factors influencing the age of first sexual intercourse in twins. *J. Biosoc. Sci., 9*, 91-97

MARTIN, N. G., AND EYSENCK, H. J. 1976. Genetical factors in sexual behavior. Chapter in H. J. Eysenck: *Sex and Personality*. Open Books, London.

MATHER, K. 1973. *Genetical structure of populations*. Chapman and Hall, London.

MATHER, K., AND JINKS, J. L. 1971. *Biometrical Genetics*, 2nd Ed. Chapman and Hall, London.

MITRA, S. K. 1958. On the limiting power function of the frequency chi-square test. *Ann. Mathematical Statistics, 29*, 1221-1233.

PEARSON, E. S., AND HARTLEY, H. O. (eds). 1972. *Biometrika Tables for Statisticians*, Vol. 2. Cambridge University Press.

PERKINS, J. M., AND JINKS, J. L. 1971. Analysis of genotype x environment interaction in triple test cross data. *Heredity, 26*, 203-209.

POONI, H. S., JINKS, J. L., AND CORNISH, M. A. 1977. The causes and consequences of non-normality in predicting the properties of recombinant inbred lines. *Heredity, 38*, 329-338.

SCARR-SALAPATEK, S. 1971. Race, social class and IQ. *Science, 174*, 1285-1295.