

Estimation of heritability by regression using collateral relatives: linear heritability estimation

By WILLIAM G. HILL

Institute of Animal Genetics, Edinburgh EH9 3JN

(Received 26 June 1978)

SUMMARY

The method of heritability estimation proposed by Abplanalp ('linear heritability' estimation) is reviewed and the formulation simplified. It is equivalent to the regression of performance of sibs of an individual, itself excluded, on that of the individual. The method is found to be slightly biased in small samples but, in terms of sampling variance, can make efficient use of the data.

1. INTRODUCTION

Heritability estimates are used for predicting the regression of offspring on parent or response to selection, and are often estimated in that way if records are available on two generations. No assumption needs to be made of linearity of the offspring on parent regression, for a non-linear model can be fitted or the regression computed on only a selected set of parents. For example, it may be possible to check whether the regression in the direction of desired improvement of the trait is the same as in the opposite direction. Theoretical aspects of non-linearity of breeding value on phenotype or of offspring-parent regression have been discussed recently by Nishida & Abe (1974) and Robertson (1977), who give other references. Some experimental evidence of non-linearity of offspring on parent regression has been given by Nishida (1972) and Meyer & Enfield (1975) for *Tribolium*, Robertson (1977) for *Drosophila melanogaster*, and Clayton (1975) and Shalev (1977) for egg production in poultry.

If records are available solely on a single generation, or in populations undergoing directional selection where only a narrow range of parental phenotypes (the independent variates) are used, heritabilities are usually estimated from the intra-class correlation among sibs obtained from an analysis of variance. No tests of linearity are then performed because sums of squares of all observations are computed. Abplanalp (1961) suggested a method, 'linear heritability estimation', to partially overcome this deficiency. Abplanalp's method parallels a selection experiment, except that 'paper' selection among individuals is practised on their own performance and their breeding value estimated from the mean of their sibs. It has received little attention, however, apart from studies by Yamada (1972), Arthur & Abplanalp (1975) and Shalev (1977). The estimator originally proposed by Abplanalp (1961) is biased and a correction to remove bias was given by Arthur & Abplanalp (1975). Their presentation is rather obscure, however, so in this note the methodology of linear heritability estimation is clarified and some of its statistical properties and relationships to other estimators are described.

2. METHODS

Abplanalp (1961) discusses the hierarchical classification, but for simplicity of illustration most emphasis will be given to the one-way classification, for example pair matings giving full-sib families or where there is one progeny per dam giving half-sib families.

Let X_{ij} be the observation on individual j in family i , there being n_i members of family i , with s families and $N = \sum_i n_i$ individuals in all. The usual means are

$$\bar{X}_{i.} = \sum_j X_{ij}/n_i \quad \text{and} \quad \bar{X}_{..} = \sum_i \sum_j X_{ij}/N. \tag{1}$$

It is useful as a basis to consider the random effects model

$$X_{ij} = \mu + a_i + e_{ij}, \tag{2}$$

where a_i and e_{ij} are independently normally distributed with $E(a_i) = E(e_{ij}) = 0$, $\text{var}(a_i) = t\sigma^2$ and $\text{var}(e_{ij}) = (1-t)\sigma^2$, so that X_{ij} is normally distributed with $E(X_{ij}) = \mu$, $\text{var}(X_{ij}) = \sigma^2$ and intra-class correlation t . In this case the regression of a_i on X_{ij} , for example, is linear and equals t , so the intra-class correlation contains all the useful information about the heritability. The intra-class correlation can be estimated by the analysis of variance, maximum likelihood or other procedure assuming the model (2).

Abplanalp (1961) considers the regression of family effects estimated from model (2), $\bar{X}_{i.} - \bar{X}_{..}$, on individual deviations, $X_{ij} - \bar{X}_{i.}$. This regression, even in the normal case, depends on the within family variance in addition to the intra-class correlation, since X_{ij} is included in $\bar{X}_{i.}$. In the appendix to their paper Arthur & Abplanalp (1975) show how to eliminate this bias for the case when all families are of equal size by making corrections to the deviations $\bar{X}_{i.} - \bar{X}_{..}$ using terms in $X_{ij} - \bar{X}_{i.}$. It can be shown, however, that their manipulations simply produce family means from which the individual's own observation is excluded and overall means from which the whole relevant family is excluded. There is then no difficulty in generalizing to families of unequal size. Abplanalp's methodology will therefore be presented in a rather different but somewhat simpler form.

Special definitions are needed. Let the mean of the *sibs* of individual j in family i , itself excluded, be

$$\bar{X}'_{i.(j)} = \sum_{\substack{k=1 \\ k \neq j}}^{n_i} X_{ik}/(n_i - 1) = (n_i \bar{X}_{i.} - X_{ij})/(n_i - 1) \tag{3a}$$

and let
$$\bar{X}'_{..(i)} = \sum_{h=1}^s \sum_{\substack{k=1 \\ h \neq i}}^{n_h} X_{hk}/(N - n_i) \tag{3b}$$

be the mean of individuals *unrelated* to members of family i . If family sizes are equal ($n_i = n$ for all i), (3b) reduces to

$$\bar{X}'_{..(i)} = (s\bar{X}_{..} - \bar{X}_{i.})/(s - 1). \tag{3c}$$

Consider first model (2). Then for randomly chosen individuals

$$\text{cov}(\bar{X}'_{i.(j)}, X_{ij}) = t\sigma^2 \tag{4}$$

and the regression of the mean of sibs on individual performance equals t . This relationship (4) provides the motivation for Abplanalp's linear estimator. He proposes that paper selection be practised among individuals on the basis of their own performance (X_{ij}), thereby constructing a selection differential, and the mean of the sibs ($\bar{X}'_{i.[j]}$) of these individuals computed to give a response. Let $q_{ij} = 1$ if the individual is 'selected' and $q_{ij} = 0$, otherwise, with $Q = \sum_i \sum_j q_{ij}$ being the total number selected; for example Q might comprise the top 10% of individuals ranked on their own performance, X_{ij} . (Here and elsewhere summation is over all individuals: $i = 1, \dots, s; j = 1, \dots, n_i$). Then, since for normality,

$$E[\sum_i \sum_j q_{ij}(\bar{X}'_{i.[j]} - \mu)] = tE[\sum_i \sum_j q_{ij}(X_{ij} - \mu)] \tag{5}$$

an estimator of t is given by the ratio of the two quantities within the expectations in (5) although, since it is a ratio, it is not unbiased. Of course (5) cannot be used in practice since μ is an unknown parameter. Abplanalp (1961), as modified by Arthur & Abplanalp (1975), in effect propose that the selection differential be

$$S = \sum_i \sum_j q_{ij}(X_{ij} - \bar{X}'_{..[i]})/Q, \tag{6}$$

i.e. the deviation of observations from unrelated individuals, and the response be

$$R = \sum_i \sum_j q_{ij}(\bar{X}'_{i.[j]} - \bar{X}'_{..[i]})/Q, \tag{7}$$

i.e. the deviation of sibs from unrelated individuals. The linear estimator of intra-class correlation proposed by Abplanalp is then

$$\hat{t}_L = R/S = \frac{\sum_i \sum_j q_{ij}(\bar{X}'_{i.[j]} - \bar{X}'_{..[i]})}{\sum_i \sum_j q_{ij}(X_{ij} - \bar{X}'_{..[i]})}, \tag{8}$$

and it is multiplied by 2 or 4 according to whether family members are full- or half-sibs. Note that the estimator is not strictly linear, since it is a ratio, but the numerator and denominator are both linear in the observations, in contrast to the usual estimator of intra-class correlation from the analysis of variance which is a ratio of sums of squares of observations. Also, because of their connexion through the mean of unrelated individuals, even for the normal model

$$\text{cov}(\bar{X}'_{i.[j]} - \bar{X}'_{..[i]}, X_{ij} - \bar{X}'_{..[i]}) \neq t\sigma^2 \tag{9}$$

and the regression of $\bar{X}'_{i.[j]} - \bar{X}'_{..[i]}$ on $X_{ij} - \bar{X}'_{..[i]}$ does not equal t . This undesirable property cannot be corrected by replacing the unrelated mean $\bar{X}'_{..[i]}$ by the overall mean $\bar{X}_{..}$ in (8) or (9). Thus it is probably best to accept that Abplanalp's estimator is difficult to justify in any rigorous manner, but to examine its behaviour as it stands.

An alternative procedure is to plot R and S from (6) and (7), respectively, for the whole population split up into groups of similar size. Such a plot, given subsequently in section 4, may be useful in illustrating both non-normality and non-linearity.

Relation to the analysis of variance. It is shown in the appendix that the usual estimator of intra-class correlation from the analysis of variance, \hat{t}_v , obtained by

estimating the between and within group components of variance, can be expressed as

$$\hat{t}_V = \frac{\sum_i \sum_j (\bar{X}'_{i..[j]} - \bar{X}'_{..[i]}) X_{ij}}{\sum_i \sum_j (X_{ij} - \bar{X}'_{..[i]}) X_{ij}} \tag{10}$$

when subclass numbers are equal. In (10) the deviations of individual or sibs' mean from the unrelated mean are weighted by the individual observation, rather than by 1 or 0 as in Abplanalp's linear estimator (8).

It is possible to construct another estimator in which summation in (10) is taken only over some of the individuals, i.e. by inserting q_{ij} ($= 1$ or 0) in the numerator and denominator. The properties of this quadratic estimator are little different from those of Abplanalp's and it will not be discussed further.

Hierarchical design. The extension of the linear estimator to this design is immediately obvious. Let X_{ijk} be the observation on the k th individual from the j th dam mated to the i th sire. The deviations needed for a complete analysis are as follows, the definitions following naturally from (3):

- A. Deviation of observation from unrelated, $X_{ijk} - \bar{X}'_{...[i]}$,
- B. Deviation of half-sibs from unrelated, $\bar{X}'_{i..[j]} - \bar{X}'_{...[i]}$,
- C. Deviation of full-sibs from unrelated, $\bar{X}'_{ij..[k]} - \bar{X}'_{...[i]}$,
- D. Deviation of full-sibs from half-sibs, $\bar{X}'_{ij..[k]} - \bar{X}'_{i..[j]}$.

When practising paper selection $q_{ijk} = 1$ or 0 , and the estimator of heritability from half-sibs is obtained as

$$\hat{h}_L^2 = 4 \sum_i \sum_j \sum_k q_{ijk} (\bar{X}'_{i..[j]} - \bar{X}'_{...[i]}) / \sum_i \sum_j \sum_k q_{ijk} (X_{ijk} - \bar{X}'_{...[i]})$$

and the full-sib and full-sib as a deviation from half-sib (dam component) estimators are obtained using C and D, respectively, rather than B in the numerator.

Two or more traits. Estimation of a 'linear' genetic correlation between two traits, or genetic regression of one trait on another, is straightforward. Assume measurements are available on all individuals on two traits, say X_{ij} and Y_{ij} , in a one way classification. A possible method is to rank individuals on X , select the best Q , setting q_{Xij} , say, to unity on these individuals, and estimate the sibs' mean for trait Y ; and then rank on Y and select the best Q , setting q_{Yij} to unity. The quantity

$$\frac{\sum \sum q_{Xij} (\bar{Y}'_{i..[j]} - \bar{Y}'_{..[i]})}{\sum \sum q_{Xij} (\bar{X}'_{i..[j]} - \bar{X}'_{..[i]})}$$

is an estimate of the genetic regression of Y on X , and the product

$$\left\{ \frac{[\sum \sum q_{Xij} (\bar{Y}'_{i..[j]} - \bar{Y}'_{..[i]})] [\sum \sum q_{Yij} (\bar{X}'_{i..[j]} - \bar{X}'_{..[i]})]}{[\sum \sum q_{Xij} (\bar{X}'_{i..[j]} - \bar{X}'_{..[i]})] [\sum \sum q_{Yij} (\bar{Y}'_{i..[j]} - \bar{Y}'_{..[i]})]} \right\}^{\frac{1}{2}}$$

is Abplanalp's 'linear' estimate of the genetic correlation.

3. PROPERTIES

As a standard for comparison of Abplanalp's linear estimator with other estimators of intra-class correlation it is useful to take as a basis the normal model (2). The bias and sampling variance will both be examined.

Bias. In model (2), for a randomly chosen individual with value X_{ij} ,

$$E(\bar{X}'_{i.[j]} | X_{ij}) = \mu + t(X_{ij} - \mu), \tag{11}$$

but if the population is of finite size and only extreme individuals are chosen such as in Abplanalp's method, (11) no longer holds exactly. For example, the highest ranking individual in the population must be best in its own family, so $\bar{X}'_{i.[j]} < X_{ij}$

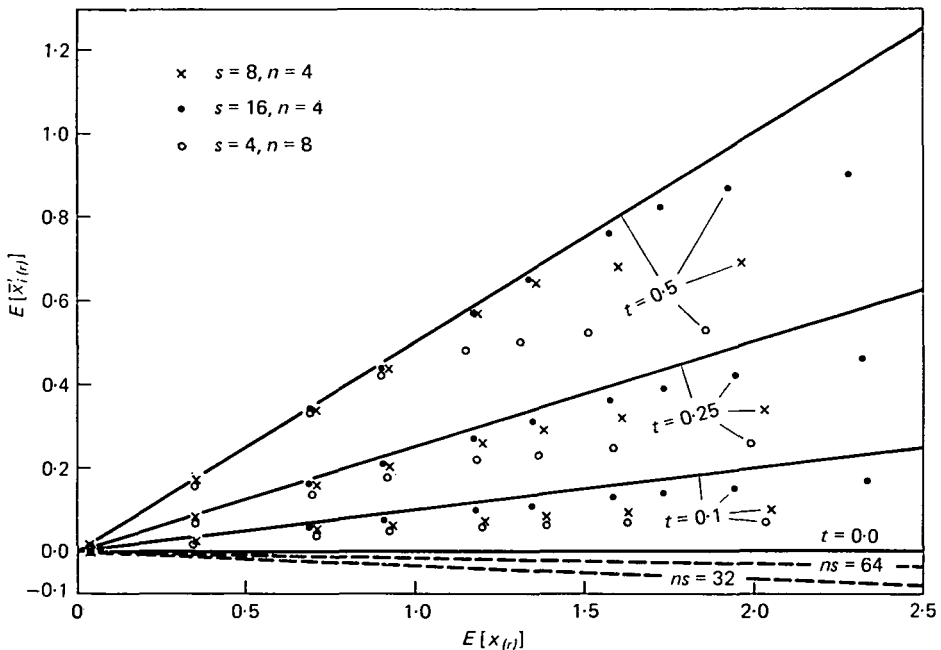


Fig. 1. $E(\bar{X}'_{i.(r)})$ against $E(X_{(r)})$ for individuals ranking $r = 1, 2, 3, 4, 6, 8, 12$ and 16 . For $t = 0$, $E(\bar{X}'_{i.(r)}) = -E(X_{(r)})/(ns - 1)$ and is shown by dotted lines. The solid lines have slope $tE(\bar{X}_{(r)})$. $\mu = 0, \sigma^2 = 1$.

and $E(\bar{X}'_{i.[j]} | X_{ij}) < \mu + t(X_{ij} - \mu)$ for the sibs of this individual. For ranked individuals, $E(\bar{X}'_{i.[j]})$ can be computed by a simple modification of eqn. (18) of Hill (1977); the details are unnecessary here. Some additional notation is useful: if X_{ij} is the observation on the r th ranking individual, let $X_{(r)} = X_{ij}$, $\bar{X}'_{i.(r)} = \bar{X}'_{i.[j]}$ be the mean of its sibs, and $\bar{X}'_{..(r)} = \bar{X}'_{..[j]}$ be the mean of unrelated individuals.

In Fig. 1 examples are plotted of $E(\bar{X}'_{i.(r)})$ against $E(X_{(r)})$ for the best individual (rank 1) down to rank $ns/2$. The same results apply with only a sign change for $r = ns/2 + 1$ to $r = ns$. The bias shown in Fig. 1, i.e. $E(\bar{X}'_{i.(r)}) - tE(X_{(r)})$, is always negative for individuals ranking in the top half of the population, becoming greatest with the most extreme individuals which are necessarily best in their own family. The bias is largest when there are few families and the intra-class correlation is high.

In Abplanalp's linear estimator (8), observations and sib means are expressed as deviations from unrelated individuals, so $E(\bar{X}'_{i(r)} - \bar{X}'_{.(r)})$ is plotted against

$$E(X_{(r)} - \bar{X}'_{.(r)})$$

in Fig. 2. This completely removes bias when $t = 0$, very much reduces the negative bias for extreme ranking individuals, but leaves a positive bias for those not at the extreme. Deviations from the regression line decrease with increase in number of

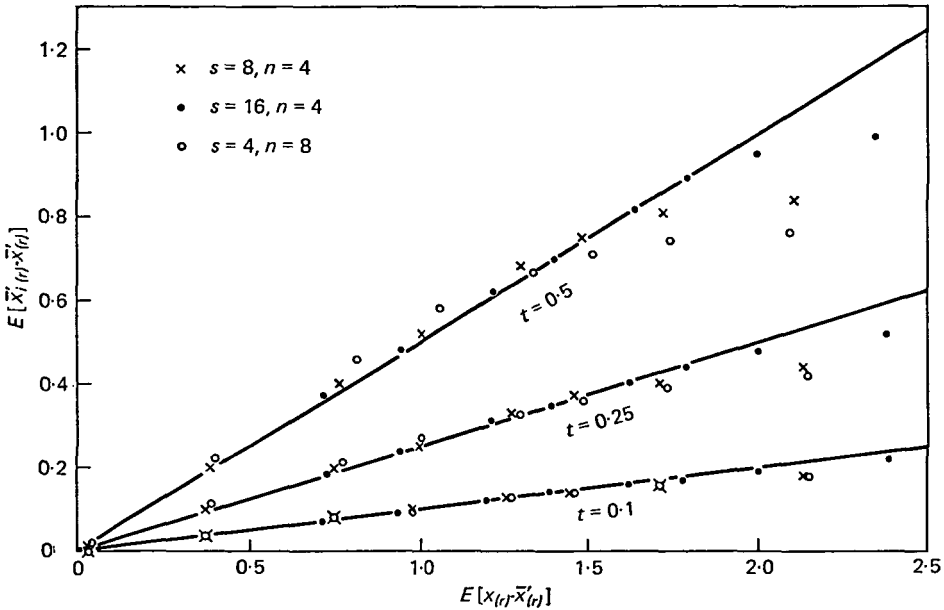


Fig. 2. As Fig. 1, but with correction for unrelated individuals, i.e. $E(\bar{X}'_{i(r)} - \bar{X}'_{.(r)})$ against $E(X_{(r)} - \bar{X}'_{.(r)})$. For $t = 0$, $E(X_{(r)} - \bar{X}'_{.(r)}) = 0$.

families, roughly as $1/s$. It seems, therefore, that in a population of size sufficient to do useful analysis of linearity of regression of sibs on individual performance, bias induced by the ranking procedure will be trivial.

Sampling error. Monte Carlo simulation was used to investigate sampling variances of the linear estimators. Pseudo-random normal deviates were sampled as in (2), ranked and the means of the observations and their 'sibs' computed for several groupings; for example with $s = 32$ and $n = 4$, giving 128 observations, 4 groups each of size 32, consisting of individuals ranking 1–32, 33–64, . . . , were taken. Estimates, such as \hat{t}_L from (8) were computed using each group (i.e. in the first group, $q_{ij} = 1$ if the individual ranked between 1 and 32). By replication, the expectation and standard error (i.e. standard deviation between replicates) of the various estimators were computed.

Examples of simulation results are given in Table 1. Because of symmetry, the mean and standard error of \hat{t}_L computed from, for example, the highest and lowest ranking groups have the same expectation. Thus to reduce sampling errors due to the finite number of replicates of Monte Carlo simulation, the means and standard

errors shown are the averages of those computed separately for the top and bottom groups. It is clear from Table 1 that an extreme group comprising only part of the total population gives linear estimators which are not much less efficient than the analysis of variance estimator. The intermediate ranking individuals give much poorer estimators. The results also show that designs optimal for estimating heritability by standard intra-class correlation methods are also optimal for estimating linear heritabilities.

Table 1. Monte Carlo simulation of R replicates of normal model with s families each of n individuals and intra-class correlation t .

(Mean and s.e. (= standard deviation among replicates) of analysis of variance (\hat{t}_V) and linear (\hat{t}_L) estimators of t . The linear estimator is given for a range of sizes (Q) of ranked groups, where group 1 comprises individuals 1 to Q , group 2 comprises individuals $Q + 1$ to $2Q$, etc.)

$t\%$	$\hat{t}_V\%$		Q	Rank of group and mean and s.e. of $\hat{t}_L\%$							
				1		2		3		4	
	Mean	s.e.		Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.
Parameters: $s = 32, n = 4, R = 400$											
0	-0.3	7.3	32	-0.3	8.5	-1.3	29.3	—	—	—	—
10	10.1	9.2	32	9.9	10.8	8.8	32.5	—	—	—	—
25	24.1	9.2	64	23.6	10.4	—	—	—	—	—	—
	—	—	32	24.3	11.1	20.7	35.0	—	—	—	—
	—	—	16	24.4	12.5	23.9	19.5	19.9	33.7	58.2	80.2
50	49.2	9.0	32	49.4	11.1	48.0	36.2	—	—	—	—
Parameters: $s = 64, n = 2, R = 400$											
10	10.4	12.4	32	10.5	14.8	10.6	51.9	—	—	—	—
25	25.8	11.4	32	26.2	14.5	29.9	50.2	—	—	—	—
Parameters: $s = 16, n = 8, R = 400$											
10	9.8	7.7	32	10.2	8.8	6.7	24.8	—	—	—	—
25	25.0	9.8	32	25.4	11.3	25.8	30.4	—	—	—	—
Parameters: $s = 128, n = 4, R = 100$											
25	24.7	4.6	256	24.7	4.9	—	—	—	—	—	—
	—	—	128	24.8	5.6	24.6	16.9	—	—	—	—
	—	—	64	24.7	6.1	24.9	10.1	23.1	16.6	28.4	48.7
Parameters: $s = 128, n = 16, R = 100$											
5	4.9	1.3	1024	4.9	1.3	—	—	—	—	—	—
	—	—	512	4.9	1.4	4.9	4.1	—	—	—	—
	—	—	256	4.8	1.6	5.1	2.2	4.8	4.0	5.5	12.7

4. EXAMPLE

Data on egg production of 1685 chickens pedigreed to sire and dam over 2 years were kindly made available by Dr B. Shalev. Individual observations were expressed as deviations from year hatch mean, and the quantities described in section 2 on hierarchical designs computed for each bird. The birds were ranked on individual

observation and split into 8 groups each of size 210, with the residual 5 mid-ranking birds excluded. Plots of group means are shown in Fig. 3, and the linear estimators obtained by taking the bottom and top groups only are as follows:

	Top		Bottom	
	Mean	h_L^2	Mean	h_L^2
Individuals: $\Sigma\Sigma\Sigma q_{ijk}(X_{ijk} - \bar{X}'_{...t(n)})/Q$	19.35	—	-35.20	—
Half-sibs: $\Sigma\Sigma\Sigma q_{ijk}(\bar{X}'_{i..t(n)} - \bar{X}'_{...t(n)})/Q$	1.62	0.34	-1.32	0.15
Full-sibs: $\Sigma\Sigma\Sigma q_{ijk}(\bar{X}'_{ij..t(n)} - \bar{X}'_{...t(n)})/Q$	3.12	0.32	-3.54	0.20
Full within half sibs: $\Sigma\Sigma\Sigma q_{ijk}(\bar{X}'_{ij..t(n)} - \bar{X}'_{i..t(n)})/Q$	1.50	0.31	-2.22	0.25

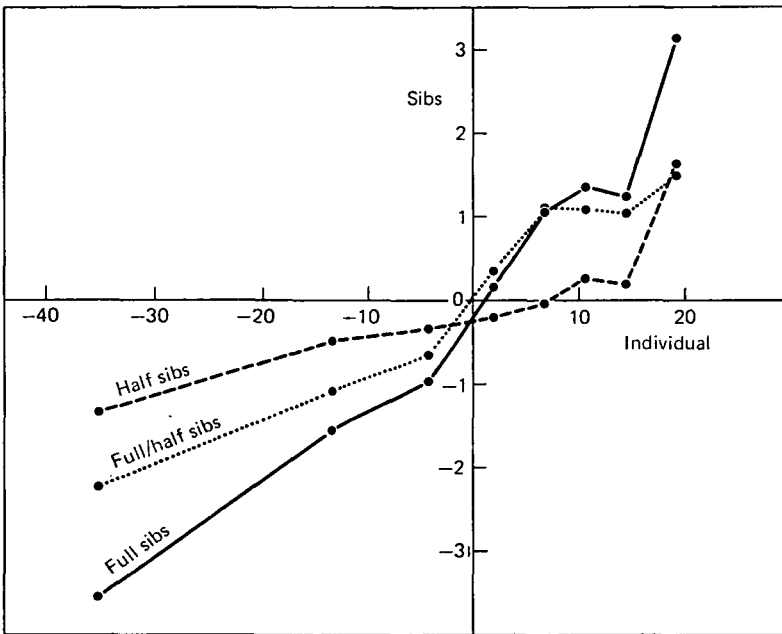


Fig. 3. Plot of sibs against individual egg production in Shalev's data. Individual observations were ranked into eight groups.

The distribution of egg production is skewed downwards, and there is some evidence of non-linearity, the heritability being higher in the up direction, surprisingly perhaps as this is the direction of previous selection. For comparison, from the analysis of variance the estimates of heritability were: 0.16 from half-sibs, 0.22 from full-sibs and 0.28 from full-sibs within half-sibs. One hypothesis which might explain these results is that the genetic distribution is nearly symmetric, while the environmental distribution is skewed downwards, a model discussed by Nishida & Abe (1974). More data would be needed to confirm this, however.

5. COMMENTS

This note has drawn attention to Abplanalp's idea of using the regression of sib on individual performance, whether displayed as a graph (e.g. Fig. 3) or expressed as a ratio of response to paper selection differential, to check on linearity using data on a

single generation. Some further problems remain. The first is methods of testing for non-linearity of sib on individual regression in data such as shown in Fig. 3. The difficulty is that individuals appear both in the abscissa and in the ordinate as many times as they have sibs or sibs falling in different ranked classes, so the error structure is clearly involved. The second area is more genetical: how, to what extent, and under what genetic models, does the sib on individual regression depart from linearity? Work on this problem is being undertaken in Edinburgh by A. Maki-Tanila (personal communication) and will be presented elsewhere.

I am grateful to Marjorie McEwan and Jenny Smith for computational assistance and to Dr Baruch Shalev for use of his data.

REFERENCES

ABPLANALP, H. (1961). Linear heritability estimates. *Genetical Research* **2**, 439-448.
 ARTHUR, J. A. & ABPLANALP, H. (1975). Linear estimates of heritability and genetic correlation for egg production, body weight, conformation and egg weight in turkeys. *Poultry Science* **54**, 11-23.
 CLAYTON, G. A. (1975). Normality of egg production in poultry. *British Poultry Science* **16**, 431-439.
 FALCONER, D. S. (1960). *Introduction to Quantitative Genetics*. Oliver and Boyd, Edinburgh.
 HILL, W. G. (1977). Order statistics of correlated variables and implications in genetic selection programmes. II. Response to selection. *Biometrics* **33**, 703-712.
 MEYER, H. H. & ENFIELD, F. D. (1975). Experimental evidence on limitations of the heritability parameter. *Theoretical and Applied Genetics* **45**, 268-273.
 NISHIDA, A. (1972). Some characteristics of parent-offspring regression in body weight of *Mus musculus* at different ages. *Canadian Journal of Genetics and Cytology* **14**, 293-303.
 NISHIDA, A. & ABE, T. (1974). The distribution of genetic and environmental effects and the linearity of heritability. *Canadian Journal of Genetics and Cytology* **16**, 3-10.
 ROBERTSON, A. (1977). The non-linearity of offspring-parent regression. In *Proceedings of International Conference of Quantitative Genetics* (ed. O. Kempthorne and E. Pollak), pp. 297-304. Iowa State University Press.
 SHALEV, B. A. (1977). Reasons for failure of response to selection for egg production in laying hens. Unpublished Ph.D. thesis, University of Reading.
 YAMADA, Y. (1972). Linear heritability and genetic correlation estimates. *Japanese Poultry Science* **9**, 90-95.

APPENDIX

An alternative formulation of the analysis of variance (ANOVA). From (4),

$$\text{Cov}(\bar{X}'_{i..(j)}, X_{ij}) = E[(\bar{X}'_{i..(j)} - \bar{X}'_{..(i)}) X_{ij}] = t\sigma^2$$

and

$$V(X_{ij}) = E[(X_{ij} - \bar{X}'_{..(i)}) X_{ij}] = \sigma^2.$$

These formulae suggest estimators of between group (family) variance and of total (phenotypic) variance as

$$(\hat{t}\sigma^2) = \sum_i \sum_j (\bar{X}'_{i..(j)} - \bar{X}'_{..(i)}) X_{ij} / N, \tag{A 1}$$

$$\hat{\sigma}^2 = \sum_i \sum_j (X_{ij} - \bar{X}'_{..(i)}) X_{ij} / N \tag{A 2}$$

and of the intra-class correlation as the ratio of terms in (A 1) and (A 2), given by (10). This estimator (10) turns out to be the same as the usual ANOVA estimator in a balanced design.

Let $B = n\sum_i(\bar{X}_{i.} - \bar{X}_{..})^2/(s-1)$ and $W = \sum_i\sum_j(X_{ij} - \bar{X}_{i.})^2/[s(n-1)]$ be the usual between and within group mean squares in ANOVA. Expanding (A 1) using (3), gives

$$\begin{aligned} \sum_i\sum_j(\bar{X}'_{i.[j]} - \bar{X}'_{..[i]})X_{ij} &= \sum_i\sum_j\left[\frac{n\bar{X}_{i.} - X_{ij}}{n-1} - \frac{s\bar{X}_{..} - \bar{X}_{i.}}{s-1}\right]X_{ij} \\ &= \sum_i\sum_j\left[\frac{s(\bar{X}_{i.} - \bar{X}_{..})}{s-1} - \frac{X_{ij} - \bar{X}_{i.}}{n-1}\right]X_{ij} \\ &= \frac{sn\sum_i(\bar{X}_{i.} - \bar{X}_{..})^2}{s-1} - \frac{s\sum_i\sum_j(X_{ij} - \bar{X}_{i.})^2}{s(n-1)} = s(B - W). \end{aligned} \tag{A 3}$$

Similarly, from (A 2),

$$\sum_i\sum_j(X_{ij} - \bar{X}'_{..[i]})X_{ij} = s[B + (n-1)W]. \tag{A 4}$$

The ratio of expressions in (A 3) and (A 4), $(B - W)/[B + (n-1)W]$ is the usual estimator of intra-class correlation.

The methodology can be extended to hierarchical or cross classified designs. For example, in the balanced hierarchical design, the analysis of variance estimator of the intra-class correlation among sires can be written

$$\sum_i\sum_j\sum_k(\bar{X}'_{i.[j]} - \bar{X}'_{...[i]})X_{ijk}/\sum_i\sum_j\sum_k(X_{ijk} - \bar{X}'_{...[i]})X_{ijk}.$$