

Linear heritability estimates

By HANS ABPLANALP
*Poultry Husbandry Department,
University of California, Davis*

(Received 16 May 1961)

INTRODUCTION

The use of regression and variance analysis in population genetics is generally accepted as a means of predicting genetic gains from selection. When only one trait is considered these predictions are usually formulated in terms of heritability, h^2 , the linear regression of genotype on measured phenotype, in closed populations, and the selection differential, ΔP . Thus in the simplest case of mass selection in large randomly reproducing populations the anticipated response to selection is

$$\Delta G = \Delta P \cdot h^2$$

Selection usually is by truncation, that is the best of a given generation are retained as parents.

The use of linear regression of genotype on phenotype for prediction of change from selection is only justified when such a linear relationship can reasonably be expected on genetic grounds. This implies, in most instances, the assumption that genes act additively, that gene effects are small, that parental effects on offspring are due to such genes only, and that environmental effects are random for individual offspring. These assumptions will apply throughout this paper.

Heritability estimates based on intraclass correlations have by and large proved useful for predicting gains from selection (Clayton *et al.*, 1957). In some respects, however, the conventional concept of heritability has been found wanting.

In actual selection experiments it is often found that response for a given absolute selection differential is greater in one direction than in the other. Thus for growth rate in mice it has been found that selection response in the direction of small body size is greater than in the direction of large size (Falconer, 1960*a*). Clearly a heritability estimate based on the usual variance analysis is incapable of describing such situations, while conventional selection experiments used under those circumstances involve data from at least two generations.

In this paper a method of estimating heritability and genetic correlation is presented which permits evaluation of genetic gains from selection in either direction on the basis of data from a single generation.

HERITABILITY ESTIMATES

Consider a breeding experiment typical for poultry in which sires (s) are each mated to dams (d), each of which gives rise to offspring (g). The latter are grown

together and measured for a given production trait, x . The mathematical model appropriate for analysis of this experiment is:

$$x_{ijk} = m + s_i + d_{ij} + f_{ijk}$$

where x_{ijk} is the measured performance of the k th offspring from i th sire and j th dam mated to that sire;

m the overall mean common to all observations;

s_i the effect of the i th sire on its progeny, presumed to be mostly genetic;

d_{ij} the effect of the j th dam on its offspring, due primarily to genetic causes, but also to non-genetic maternal effects;

f_{ijk} the effect of environment and unpredictable genetic factors affecting the k th offspring from dam j and sire i .

On this model the least-squares estimates for sire, dam, and offspring effects respectively are:

$$\begin{aligned} \hat{s}_i &= x_{i..} - x_{...} = \text{estimated sire effect,} \\ \hat{d}_{ij} &= x_{ij.} - x_{i..} = \text{estimate of dam effects,} \\ \hat{f}_{ijk} &= x_{ijk} - x_{ij.} = \text{estimate of individual offspring effects.} \end{aligned}$$

Small x 's denote averages taken over all subscripts represented by a dot.

Assuming that all effects with the exception of m are random samples from given distributions (Eisenhart Model 2), we have the following variances of these estimates:

$$\left. \begin{aligned} \text{Var}(\hat{s}) &= \sigma_f^2/dn + \sigma_d^2/d + \sigma_s^2 \\ \text{Var}(\hat{d}) &= \sigma_f^2/n + \sigma_d^2 \\ \text{Var}(\hat{f}) &= \sigma_f^2 \end{aligned} \right\} \quad (1)$$

where σ_s^2 , σ_d^2 and σ_f^2 are variance components due to sire, dam and individual effects respectively.

Using these relations and estimates of variance components the geneticist then proceeds to predict genetic changes from selection in the following generation by means of heritability estimates (in this case intraclass correlations) and known selection differentials.

It should be noted that linear prediction equations of the type discussed do not rely directly on estimates of variance components, but take the general form:

$$\text{Expected displacement} = \left\{ \begin{array}{l} \text{Selection} \\ \text{differential} \end{array} \right\} \left\{ \begin{array}{l} \text{Regression of response} \\ \text{criterion on selection} \\ \text{criterion.} \end{array} \right\}$$

In the context of the paper the term displacement is used to denote genetic or phenotypic changes from selection by truncation in either direction.

While variance components are a convenient means of estimating the regressions involved, it will be shown that they are not necessary for that purpose. Regressions

can be estimated, in principle, from the expected displacements in component parts of the selection criterion (x) as caused by a constant selection differential (ΔP) of the latter. To this end we define the following displacements and partial regressions, within a generation of pedigreed offspring, and in terms of mass selection on x with selection differential, ΔP :

$$\left. \begin{aligned} \Delta S &= \Delta P(\sigma_s^2/\sigma_x^2) = \text{Displacement component due to sire effects,} \\ \Delta D &= \Delta P(\sigma_d^2/\sigma_x^2) = \text{Displacement component due to dam effects,} \\ \Delta F &= \Delta P(\sigma_f^2/\sigma_x^2) = \text{Displacement component due to individual hen effects,} \end{aligned} \right\} (2)$$

where σ_x^2 is the phenotypic variance of x .

It follows easily that

$$\frac{2(\Delta S + \Delta D)}{\Delta S + \Delta D + \Delta F} = \frac{2(\sigma_s^2 + \sigma_d^2)}{\sigma_s^2 + \sigma_d^2 + \sigma_f^2} = h^2,$$

the usual heritability estimate based on sire and dam components of variance.

Unfortunately the expected displacements of sire, dam and hen effects are not directly observable. They can be derived, however, from displacements in respective least-squares estimates, \hat{s}_i , \hat{d}_{ij} and \hat{f}_{ijk} . Thus we have:

$$\left. \begin{aligned} \Delta P_S &= \Delta P \frac{\sigma_s^2 + \sigma_d^2/d + \sigma_f^2/nd}{\sigma_x^2} = \text{Expected displacement in estimates of sire effects } (\hat{s}_i); \\ \Delta P_D &= \Delta P \frac{(\sigma_d^2 + \sigma_f^2/n)}{\sigma_x^2} = \text{Expected displacement in estimates of dam effect } (\hat{d}_{ij}); \\ \text{and} \\ \Delta P_F &= \frac{\Delta P(\sigma_f^2)}{\sigma_x^2} = \Delta F = \text{Expected displacement component due to individual hen effects } (\hat{f}_{ijk}). \end{aligned} \right\} (3)$$

The regressions in the above equations follow directly from the variances of the estimates of sire, dam and hen effects respectively as shown in (1). Furthermore, from (2) and (3) we obtain

$$\Delta S = \Delta P_S - \Delta P_D/d \quad \text{and} \quad \Delta D = \Delta P_D - \Delta P_F/n.$$

Heritability can thus be determined as a simple function of ΔP_S , ΔP_D and ΔP_F , which can be obtained from the data as follows:

$$\left. \begin{aligned} \Delta P_S &= \frac{\sum_{m=1}^q (\hat{s}_{mi.})}{q} = \frac{\sum_{m=1}^q (s_{mi..} - x_{m...})}{q} = \text{Average of estimates of sire effects for the } q \text{ selected individuals.} \\ \Delta P_D &= \frac{\sum_{m=1}^q (\hat{d}_{ij})}{q} = \frac{\sum_{m=1}^q (x_{mij.} - x_{mi..})}{q} = \text{Average of estimates of dam effects for the } q \text{ selected individuals.} \\ \Delta P_F &= \frac{\sum_{m=1}^q (\hat{f}_{mijk})}{q} = \frac{\sum_{m=1}^q (x_{mijk} - x_{mij.})}{q} = \text{Average of estimates of individual hen effects for the selected individuals.} \end{aligned} \right\} (4)$$

Note the total selection differential from mass selection

$$\Delta P = \frac{\sum_{m=1}^q (x_{mijk} - x_{m\dots})}{q} = \Delta P_S + \Delta P_D + \Delta P_F.$$

As a logical extension of the above development we can define

$\Delta G = 2(\Delta D + \Delta S)$, or $\Delta G = 4\Delta S$, or $\Delta G = 4\Delta D$, as the component of total genetic displacement;

$\Delta M = \Delta D - \Delta S$ as the component of displacement from maternal effects when the latter can reasonably be assumed; and

$\Delta E = \Delta F - \Delta S - \Delta D$ as the component of displacement due to random environmental effects acting on individuals.

This is in accordance with similar definitions based on variance components. It should be noted that the above operations on displacement estimates ΔP_S , ΔP_D and ΔP_F which lead to heritability and other regressions are identical with the calculations applied to corresponding variances, in the usual estimation procedures using variance components.

The present approach to estimation of heritability can easily be extended to cases where selection is on family means or some other known functions of observed performance data. Because the present estimates are simple linear functions of observations it appears appropriate to call them linear heritability estimates. In order for such a linear heritability estimate to be equivalent to one obtained by variance component techniques it would be necessary that the regression of genotype on phenotype be linear. In practice this assumption may not be met. In that case the estimates proposed here are likely to be more realistic than conventional ones in the sense that they are defined directly on changes from selection. They may be visualized as an attempt at fitting a straight line through the origin on one hand and the expected genotype of selected parents on the other.

The present approach to the problem of selection response also has the advantage of relatively few assumptions. Furthermore, we may use it to investigate the symmetry of selection response with data from only one generation of pedigreed individuals.

ESTIMATES OF GENETIC CORRELATION BETWEEN TRAITS

The same method, as applied to one characteristic, lends itself in principle to an investigation of correlated responses in two or more traits. Suppose we are concerned with selection on trait x and wish to predict the consequences of such action for trait y . This can be done by applying mass selection, by truncation, to x at an intensity and direction for which correlation estimates are of interest. For the selected individuals one can then determine average displacement components of y , the correlated trait.

Accordingly we denote the total phenotypic displacement in y due to selection on x as $\Delta P_{y,x}$, and partition it into average displacements from sire effects ($\Delta P_{S_{y,x}}$),

dam effects ($\Delta P_{Dy.x}$) and individual effects ($\Delta P_{Fy.x}$) respectively, based on least-squares estimates of sire, dam and individual effects, in y , for the breeders selected. As in the case of direct response for a single trait we have for the correlated displacement:

$$\Delta P_{y.x} = \Delta P_{Sy.x} + \Delta P_{Dy.x} + \Delta P_{Fy.x},$$

and correlated displacement components derived by operations identical to those used for direct response in x thus are:

$$\Delta F_{y.x} = \Delta P_{Fy.x},$$

$$\Delta D_{y.x} = \Delta P_{Dy.x} - \frac{\Delta P_{Fy.x}}{n},$$

and

$$\Delta S_{y.x} = \Delta P_{Sy.x} - \frac{\Delta P_{Dy.x}}{d}.$$

Also the total correlated displacement in the genotype of y may be defined as

$$\Delta G_{y.x} = 2(\Delta S_{y.x} + \Delta D_{y.x})$$

and the correlated response due to non-genetic factors

$$\Delta E_{y.x} = \Delta F_{y.x} - \Delta D_{y.x} - \Delta S_{y.x}.$$

By this procedure we thus obtain directly the expected genetic change in y as induced by selection on x or vice versa. These estimates may be all that is required in actual experimentation. In order to compare results from different populations, however, one may proceed to calculate genetic correlations. In case of response in y due to selection on x we define the genetic correlation as

$$r_G = \frac{\text{cov } G_{y.x}}{\sigma_{Gx} \sigma_{Gy}}.$$

We now express genetic variances and covariances as functions of selection differentials and genetic displacements. For normally distributed x we have for the selection differential

$$\Delta P_x = \bar{i} \cdot \sigma_x, \tag{5}$$

where \bar{i} is the selection differential for truncation in the standard normal distribution.

The genetic displacement in x due to selection on x ,

$$\Delta G_x = \frac{\sigma_{Gx}^2}{\sigma_x^2} \Delta P_x,$$

where σ_{Gx}^2 is the genetic variance in x .

Substituting σ_x^2 in (5) we then have

$$\sigma_{Gx}^2 = \frac{\Delta G_x \Delta P_x}{\bar{i}^2},$$

and similarly

$$\sigma_{Gy}^2 = \frac{\Delta G_y \Delta P_y}{\bar{i}^2}.$$

For the genetic displacement in y due to selection on x we have

$$\Delta G_{y.x} = \frac{\text{cov } G_{y.x}}{\sigma_x^2} \Delta P_x$$

hence

$$\text{cov } G_{y.x} = \frac{\Delta G_{y.x} \Delta P_x}{\bar{v}^2}$$

and

$$r_{G_{y.x}} = \frac{\Delta G_{y.x}}{\Delta P_y} \bigg/ \sqrt{\left[\frac{\Delta G_x \Delta G_y}{\Delta P_x \Delta P_y} \right]}$$

Alternately we can define genetic correlation on the basis of genetic change in x due to selection on y as

$$r_{G_{x.y}} = \frac{\Delta G_{x.y}}{\Delta P_x} \bigg/ \sqrt{\left[\frac{\Delta G_x \Delta G_y}{\Delta P_x \Delta P_y} \right]}$$

From what has been said earlier it should be clear that this definition further depends on the direction in which the primary variable is selected.

The present method of estimating genetic correlations deals with a problem for which adequate statistical tools have not been provided by genetic theory so far. According to the conventional statistical formulation of correlated response the genetic correlation between characters is symmetrical, that is, independent of the fact which of the two traits is subjected to selection. With the present method we may subject this assumption to an empirical test, by comparing a given correlation estimate, when obtained from genetic changes in y as induced by selection on x , with comparable changes in x brought about by selection on y . Again limited empirical information from selection experiments suggests that such alternate correlations may be different from each other (Falconer, 1960).

The present method of estimating selection response with data from one generation can readily be extended to other experimental designs than the one considered here, provided such selection has operational meaning. In general this would imply the assumption of selection on unknown random effects for which the experiment is capable of providing least-squares estimates. The present estimation technique can also be adapted to situations where dominance is assumed, in a manner analogous to corresponding analyses of variance.

It should be noted, however, that the present method of partitioning the selection differential (ΔP_x) according to assumed genetic effects breaks down, when the total selection differential becomes small. Empirical results obtained so far suggest that selection differentials of about one standard deviation (σ_x) are giving results comparable to those obtained from conventional variance components techniques. Clearly the method of estimation proposed here is only of interest, when the assumption of linear regression of genotype on phenotype is violated. Under such circumstances non-linearity of regressions would bias predictions, primarily with large selection differentials. But, for relatively weak selection, the conventional intraclass correlations and offspring parent regressions might be preferable to the ones proposed here.*

* It has been pointed out by a referee that the optimum selection intensity of estimating linear heritability for a given set of data must be close to 0.27, the known optimum selection intensity for an ordinary selection programme. (See Robertson, 1957.)

Nothing has been said so far about the statistical properties of the proposed estimates. A derivation of exact distribution functions for either heritability or genetic correlation estimates has not been possible, so far. However, approximate sampling distributions for population structures of particular interest might be obtained on high-speed computers. So far, we have derived heritability estimates from a number of small populations. The results indicate relatively good agreement for heritability based on intraclass correlations and the present estimates, as illustrated by the numerical example given below. Genetic correlations based on populations of only few parents, however, have shown large sampling variance. Thus, in case of relatively small populations there remains the question as to the usefulness of the proposed estimation technique for correlation analysis. However, in case of poultry populations where data are often available from many and large populations it may be desirable to gain further insight into the genetic situation at the expense of some precision of estimates. In such situations the present method of estimation may be useful. Also it requires much less computational work than comparable variance analyses and can thus serve for a quick appraisal of the genetic situation.

A NUMERICAL EXAMPLE

From a large population of pedigreed SCWL hens a random sample of records was drawn so that twenty mated dams were each represented by five offspring. The dams themselves had been mated to individual sires in five groups of four, respectively. We shall now consider part-time egg production records to 40 weeks of age (x) and 40 week egg weight (y) for the 100 hens sampled.

Table 1 shows egg production and egg-weight records of the twenty hens with highest egg number (x), ranked according to the latter. Also given are least-squares estimates of sire effects (\hat{s}), dam effects (\hat{d}) and individual hen effects (\hat{f}) for both traits, based on the entire sample of 100 hens. Average phenotypic displacements in sire effects, dam effects and individual effects are shown in the bottom row of the table; thus, for example,

$$\Delta P_{Sx} = 2.13 \text{ eggs, etc.,}$$

and correlated displacements in y due to selection on x are

$$\Delta P_{Sy.x} = -0.13 \text{ grams egg weight, etc.}$$

Results comparable to those shown in Table 1 were also computed (but are not shown here) for the twenty hens with highest egg weights from which average phenotypic displacements for egg weight (y) and correlated phenotypic displacements in egg number (x) were obtained.

We may now calculate displacement components due to sires, dams and individuals respectively for both direct response to selection and correlated responses. For egg number we thus have the sire component

$$\Delta S_x = \Delta P_{Sx} - \Delta P_{Dx}/d = 2.13 - 4.91/4 = 0.90 = \Delta G/4,$$

Table 1. Selected breeding animals and corresponding phenotypic displacement

Selection criterion	Egg number to 40 weeks			Egg weight (g.) at 40 weeks		
	Sire effects $x_{i..} - x_{...}$	Dam effects $x_{ij.} - x_{i..}$	Individual hen effects $x_{ij\mu} - x_{ij.}$	Sire effects $y_{i..} - y_{...}$	Dam effects $y_{ij.} - y_{i..}$	Individual hen effects $y_{ij\mu} - y_{ij.}$
39.06	8.51	10.20	20.35	-0.74	-0.53	1.44
33.81	-1.57	4.59	30.79	0.82	0.34	2.98
29.06	8.51	7.65	12.90	-0.74	-1.84	0.75
29.01	2.23	2.13	28.91	-1.13	-2.29	2.83
26.81	-4.60	2.73	28.68	-0.01	-0.29	0.46
25.04	2.23	4.28	18.53	-1.13	1.54	-1.85
24.31	-4.55	10.85	18.01	1.07	-0.01	2.50
23.06	-1.57	6.48	18.15	0.82	-1.23	-2.42
21.31	-1.57	4.39	18.49	0.82	0.74	-1.00
20.31	8.51	10.20	1.60	-0.74	-0.53	-1.17
19.81	8.51	10.20	1.10	-0.74	-0.53	3.45
19.06	2.23	4.62	12.21	-1.13	-0.17	-1.53
19.06	8.51	-6.56	17.11	-0.74	-0.69	-2.40
18.81	8.51	-11.31	21.61	-0.74	3.04	-0.12
18.81	-4.55	2.15	21.21	1.07	-0.05	2.16
18.31	8.51	7.65	2.15	-0.74	-1.84	1.14
17.31	-4.60	2.73	19.18	-0.01	-0.29	0.84
17.06	-4.55	10.85	10.76	1.07	-0.01	1.11
16.31	8.51	7.65	0.15	-0.74	-1.84	-2.86
16.31	-4.55	10.85	10.01	1.07	-0.01	-6.50
Average						
22.63	2.13	4.91	15.59	-0.13	-0.32	-0.31
(ΔP_z)	(ΔP_{sz})	(ΔP_{Dz})	(ΔP_{Fz})	($\Delta P_{Sv,z}$)	($\Delta P_{Dv,z}$)	($\Delta P_{Fv,z}$)

Table 2. Components of displacement, heritabilities and genetic correlations from mass selection on egg number and egg weight in either direction

Selection method	Displacement components in egg number due to			Displacement components in egg weight due to			Heritability estimates	Estimates of genetic correlations
	sire	dam	individual	sire	dam	individual		
High egg number	0.90* (ΔS_x)	1.79 (ΔD_x)	15.60 (ΔF_x)	-0.05 ($\Delta S_{y,x}$)	-0.26 ($\Delta D_{y,x}$)	-0.31 ($\Delta F_{y,x}$)	0.29 (h_x^2)	-0.63 ($r_{oy,x}$)
Low egg number	-0.29	1.23	-19.44	0.01	0.23	-0.07	-0.10	-1.07
High egg weight	-1.26 ($\Delta S_{x,y}$)	-0.78 ($\Delta D_{x,y}$)	-5.48 ($\Delta F_{x,y}$)	0.32 (ΔS_y)	0.26 (ΔD_y)	2.26 (ΔF_y)	0.41 (h_y^2)	-0.68 ($r_{ox,y}$)
Low egg weight	1.04	1.25	-3.21	-0.03	-0.26	-2.90	0.18	-1.77

* Corresponding symbols are given in brackets below numerical values; the same symbols are used for downward selection as in upward selection.

the dam component

$$\Delta D_x = \Delta P_{Dx} - \Delta P_{Fx}/n = 4.91 - 15.60/5 = 1.79 = \Delta G/4,$$

and the individual component

$$\Delta F_x = \Delta P_{Fx} = 15.60 = \Delta G/2 + \Delta E.$$

Finally we obtain the total genetic displacement $\Delta G_x = 2(\Delta S_x + \Delta D_x) = 5.38$; and the displacement due to environmental effects $\Delta E_x = 12.91$; from which we can calculate a heritability

$$h_x^2 = \frac{2(\Delta S_x + \Delta D_x)}{\Delta G_x + \Delta E_x} = \frac{5.38}{18.29} = 0.29.$$

A summary of results obtained for the present example is given in Table 2. Also included in the summary are heritabilities and genetic correlations for selection of either x or y in the downward direction at a selection intensity of one in five.

A comparison of the heritabilities and genetic correlation estimates in Table 2 with corresponding estimates derived from variances shows them in good agreement, in the sense that the variance estimates lie between the two displacement estimates derived for up and down selections respectively. Thus heritability based on variance components is 0.25 for egg number, and 0.31 for egg size, and the genetic correlation is -0.97 .

SUMMARY

A method for obtaining linear estimates of heritability and genetic correlation is given. It is based, essentially, on selecting, from a pedigreed population, prospective parents for which estimates of average genotype and phenotype values are obtained; a regression of genotype on phenotype value is then determined from a straight line fitted through points representing the population mean genotype and phenotype on one hand and the mean genotype and phenotype of selected parents on the other.

The method permits an evaluation of asymmetry in response for a trait selected in both directions, as well as asymmetry in correlated response to selection of two different traits, with data from a single pedigreed population.

REFERENCES

- CLAYTON, G. A., MORRIS, J. A. & ROBERTSON, A. (1957). An experimental check on quantitative genetical theory. *J. Genet.* **55**, 131-151.
- FALCONER, D. S. (1960*a*). Selection of mice for growth on high and low planes of nutrition. *Genet. Res.* **1**, 91-113.
- FALCONER, D. S. (1960*b*). The genetics of litter size in mice. *J. cell. comp. Physiol.* **56** (suppl. 1), 153-167.
- ROBERTSON, A. (1957). Optimum group size in progeny testing and family selection. *Biometrics*, **13**, 442-450.