

Estimation of heritability by offspring-parent regression when observations do not have a common mean

R. L. Fernando and D. Gianola

University of Illinois at Urbana-Champaign, 126 Animal Sciences Laboratory, 1207 West Gregory Drive, Urbana, IL 61801, USA

Received July 6, 1987; Accepted November 11, 1987

Communicated by L. D. Van Vleck

Summary. Offspring-parent regression is a simple method for estimating heritability. This method yields unbiased estimates even when parents are selected. The usual model in offspring-parent regression assumes that observations have the same mean. This assumption, however, is not met in many situations. A method for estimating heritability by offspring-parent regression when observations do not have a common mean is presented. The estimator is distributed as a multiple of a t random variable centered at its parametric value and is unbiased even when the parents are selected. When observations have a common mean, the method reduces to the “usual” regression estimator.

Key words: Offspring-parent regression – Heritability – Genetic parameters

Introduction

Heritability (h^2) is useful to study genetic change of a population undergoing selection (e.g., Falconer 1981) and to choose among alternative breeding programs (e.g., Robertson 1957; Hill 1971). Offspring-parent regression is a widely used estimator of h^2 that is simple to compute and is unbiased even when selection of parents occurs (Falconer 1981). This is the only method that has been proved to be unbiased in the presence of selection.

The usual model for offspring-parent regression assumes that observations have a common mean. In many applications, however, phenotypic values of parents or of offspring are affected by several sources of variation. For example, measurements on parents may be taken on individuals that received different treatments, and offspring

can be classified by sex, birth group, etc. Some researchers have recognized this problem and have used ad hoc methods to adjust the data to have a common mean. The distributional properties of such methods are not well defined. The objective of this paper is to present a method, with well known distributional properties, to estimate heritability by offspring-parent regression when observations do not have a common mean. This method of estimation has the attractive property of the “usual” regression estimator of remaining unbiased in the presence of selection.

Regression estimator

Suppose $(P_1, O_1) \dots (P_n, O_n)$ are observed values of a trait in n unrelated parent-offspring pairs in a random mating population. For many traits, it is reasonable to assume that (P_i, O_i) follow independent bivariate normal distributions with mean vector:

$$E \begin{bmatrix} P_i \\ O_i \end{bmatrix} = \begin{bmatrix} \mu_{pi} \\ \mu_{oi} \end{bmatrix}$$

and variance-covariance matrix:

$$\text{Var} \begin{bmatrix} P_i \\ O_i \end{bmatrix} = \begin{bmatrix} \sigma_p^2 & \sigma_{po} \\ \sigma_{op} & \sigma_o^2 \end{bmatrix} = \Sigma$$

The above assumptions imply that parents are a random sample from a conceptual population. Further, the means of the vectors of phenotypic records can be expressed as

$$E(\mathbf{P}) = \mathbf{W}_p \boldsymbol{\beta}_p \quad \text{and} \quad E(\mathbf{O}) = \mathbf{W}_o \boldsymbol{\beta}_o \quad (1)$$

where \mathbf{W}_p and \mathbf{W}_o are known incidence matrices, and $\boldsymbol{\beta}_p$ and $\boldsymbol{\beta}_o$ are unknown fixed effects affecting parent and offspring records, respectively.

To obtain the regression estimator we use a multivariate linear model for \mathbf{P} and \mathbf{O} :

$$[\mathbf{P}, \mathbf{O}] = [\mathbf{W}_p, \mathbf{W}_o] \begin{bmatrix} \beta_p & \mathbf{0} \\ \mathbf{0} & \beta_o \end{bmatrix} + [\mathbf{e}_p, \mathbf{e}_o] \quad (2)$$

where the vectors of residuals \mathbf{e}_p and \mathbf{e}_o have null expectations. Parents are assumed to be unrelated, so that

$$\text{Var}(\mathbf{e}_p) = \mathbf{I} \sigma_p^2, \quad \text{Var}(\mathbf{e}_o) = \mathbf{I} \sigma_o^2 \quad \text{and} \quad \text{Cov}(\mathbf{e}_p, \mathbf{e}_o) = \mathbf{I} \sigma_{po}$$

where \mathbf{I} is an identity matrix of order n . The residuals can be predicted as

$$[\hat{\mathbf{e}}_p, \hat{\mathbf{e}}_o] = \mathbf{M} [\mathbf{P}, \mathbf{O}] \quad (3)$$

where $\mathbf{M} = [\mathbf{I} - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}']$, $\mathbf{W} = [\mathbf{W}_p, \mathbf{W}_o]$, and $(\mathbf{W}'\mathbf{W})^{-1}$ is a generalized inverse of $\mathbf{W}'\mathbf{W}$.

Let \mathbf{S} be a 2×2 matrix of sums of squares and cross-products of predicted residuals.

$$\mathbf{S} = \begin{bmatrix} \hat{\mathbf{e}}_p' \hat{\mathbf{e}}_p & \hat{\mathbf{e}}_p' \hat{\mathbf{e}}_o \\ \hat{\mathbf{e}}_o' \hat{\mathbf{e}}_p & \hat{\mathbf{e}}_o' \hat{\mathbf{e}}_o \end{bmatrix}$$

The regression of offspring on parent (α) is estimated as

$$\hat{\alpha} = s_{12}/s_{11} \quad (4)$$

Distribution of the regression estimator

In order to derive the distribution of $\hat{\alpha}$, let \mathbf{Z} be a $2 \times (n-r)$ matrix of standard normal, independent random variables where r is the rank of \mathbf{W} , and let \mathbf{C} be a 2×2 lower triangular matrix such that

$$\Sigma = \mathbf{C}\mathbf{C}' \quad (5)$$

The matrix

$$\mathbf{Q} = \mathbf{C}\mathbf{Z}\mathbf{Z}'\mathbf{C}' \quad (6)$$

and the matrix \mathbf{S} have the same Wishart distribution with $n-r$ degrees of freedom and parameter matrix Σ (Anderson 1984). The matrix $\mathbf{Z}\mathbf{Z}'$ can be written as

$$\mathbf{Z}\mathbf{Z}' = \mathbf{B}\mathbf{B}' \quad (7)$$

where \mathbf{B} is a 2×2 lower triangular matrix; this relationship is known as the Bartlett decomposition of $\mathbf{Z}\mathbf{Z}'$ (Anderson 1984). Diagonal elements of \mathbf{B} are independent chi random variables with element i having $n - [(r-1) + i]$ degrees of freedom; the off-diagonal element of \mathbf{B} is a standard normal random variable independent of the diagonal elements. Now, \mathbf{Q} in (6) can be written in terms of [7] as

$$\mathbf{Q} = \mathbf{C}\mathbf{Z}\mathbf{Z}'\mathbf{C}' = \mathbf{C}\mathbf{B}\mathbf{B}'\mathbf{C}' = \mathbf{X}\mathbf{X}' \quad (8)$$

where $\mathbf{X} = \mathbf{C}\mathbf{B}$ is a lower triangular matrix. Because the matrix \mathbf{Q} has the same distribution as \mathbf{S} , the distribution

of $\hat{\alpha}$ is identical to that of

$$q_{12}/q_{11} = x_{11}x_{21}/x_{11}^2 = x_{21}/x_{11} \quad (9)$$

From (8), we have that

$$x_{11} = c_{11}b_{11} \quad (10)$$

$$x_{21} = c_{21}b_{11} + c_{22}b_{21} \quad (11)$$

and

$$x_{21}/x_{11} = (c_{21}/c_{11}) + (c_{22}/c_{11})(b_{21}/b_{11}) \quad (12)$$

Also, from (5) it follows that

$$c_{11} = \sigma_p \quad (13)$$

$$c_{21} = \sigma_{po}/\sigma_p \quad (14)$$

$$c_{22} = (\sigma_o^2 - \sigma_{po}^2/\sigma_p^2)^{1/2} \quad (15)$$

Substituting (13)–(15) into (12) gives

$$x_{21}/x_{11} = \alpha + (\sigma_o^2/\sigma_p^2 - \alpha^2)^{1/2}(b_{21}/b_{11}) \quad (16)$$

The only random variables in (16) are the elements b_{21} and b_{11} from the Bartlett decomposition. The ratio $(n-r)^{1/2}b_{21}/b_{11}$ has a t distribution with $n-r$ degrees of freedom because b_{21} and b_{11} are independent random variables with a standard normal distribution and a chi distribution with $n-r$ degrees of freedom, respectively. Because the distribution of $\hat{\alpha}$ is identical to that of x_{21}/x_{11} ,

$$\hat{\alpha} \sim \alpha + (\sigma_o^2/\sigma_p^2 - \alpha^2)^{1/2} t[n-r]/(n-r)^{1/2} \quad (17)$$

where $t[n-r]$ is a t random variable with $n-r$ degrees of freedom. For the situation considered here, $\sigma_p^2 = \sigma_o^2 = \sigma^2$ is the phenotypic variance, and $\sigma_{po}/\sigma_p^2 = \alpha = h^2/2$. Thus, (17) can be written as

$$\hat{\alpha} \sim h^2/2 + (1 - h^4/4)^{1/2} t[n-r]/(n-r)^{1/2} \quad (18)$$

From (18),

$$E(\hat{\alpha}) = h^2/2 = \alpha$$

and

$$\text{Var}(\hat{\alpha}) = (1 - h^4/4)/(n-r-2) \quad (19)$$

for $n-r > 2$ because

$$\text{Var}(t[n-r]) = (n-r)/(n-r-2)$$

for $n-r > 2$. Because heritability is estimated as $2\hat{\alpha}$, the estimator of heritability is also distributed as $2\hat{\alpha}$.

When more than one offspring per parent are recorded, the offspring means are regressed on the parental records. For example, if s half-sib records are available for each sire, the variance of the offspring mean is:

$$\sigma_o^2 = \sigma^2 [(h^2/4) + (1 - h^2/4)/s] \quad (20)$$

As before, $\text{Var}(p_i) = \sigma^2$, and the regression of offspring mean on sire, $\alpha = \sigma_{po}/\sigma_p^2$, is equal to $h^2/2$, where σ_{po} is the

covariance between the sire record and the offspring mean. Substituting (20) in (17) and rearranging gives

$$\hat{\alpha} \sim h^2/2 + [h^2(1-h^2)/4 + (1-h^2/4)/s]^{1/2} t[n-r]/(n-r)^{1/2} \quad (21)$$

and the variance of estimator is

$$\text{Var}(\hat{\alpha}) = [h^2(1-h^2)/4 + (1-h^2/4)/s]/(n-r-2) \quad (22)$$

for $n-r > 2$.

When regression is on mid-parent, $h^2 = \alpha$ and the distribution of the regression estimator is

$$\hat{\alpha} \sim h^2 + [h^2(1-h^2) + (2-h^2)/s]^{1/2} t[n-r]/(n-r)^{1/2} \quad (23)$$

where s is the number of offspring per family and n is the number of families. The variance of the estimator is

$$\text{Var}(\hat{\alpha}) = [h^2(1-h^2) + (2-h^2)/s]/(n-r-2) \quad (24)$$

for $n-r > 2$.

Distribution of the estimator under selection of parents

The regression estimator (4) remains unbiased even when parents are selected. To demonstrate this it is sufficient to show that

$$E(\hat{\alpha} | \mathbf{P}) = \alpha \quad (25)$$

because the unconditional expectation of $\hat{\alpha}$ is the expected value of (25) taken over the sample space of \mathbf{P} . When parents are selected, the sample space of \mathbf{P} is altered. However, if (25) is true, the unconditional expectation is also α , irrespective of the sample space of \mathbf{P} .

To prove that (25) is true, write $\hat{\alpha}$ given by (4) as

$$\hat{\alpha} = \hat{\mathbf{e}}_p' \hat{\mathbf{e}}_o / (\hat{\mathbf{e}}_p' \hat{\mathbf{e}}_p) \quad (26)$$

Now, (25) can be written as

$$\begin{aligned} E(\hat{\alpha} | \mathbf{P}) &= E[\hat{\mathbf{e}}_p' \hat{\mathbf{e}}_o / (\hat{\mathbf{e}}_p' \hat{\mathbf{e}}_p) | \mathbf{P}] = \hat{\mathbf{e}}_p' E(\mathbf{MO} | \mathbf{P}) / (\hat{\mathbf{e}}_p' \hat{\mathbf{e}}_p) \\ &= \hat{\mathbf{e}}_p' \mathbf{M} [\mathbf{W}_o \boldsymbol{\beta}_o + \text{Cov}(\mathbf{O}, \mathbf{P}') \text{Var}^{-1}(\mathbf{P}) \\ &\quad \cdot (\mathbf{P} - \mathbf{W}_p \boldsymbol{\beta}_p)] / (\hat{\mathbf{e}}_p' \hat{\mathbf{e}}_p) \end{aligned} \quad (27)$$

Because $\text{Cov}(\mathbf{O}, \mathbf{P}') = \mathbf{I}\sigma_{po}$, $\text{Var}(\mathbf{P}) = \mathbf{I}\sigma_p^2$, $\mathbf{M}\mathbf{W}_p = \mathbf{0}$ and $\mathbf{M}\mathbf{W}_o = \mathbf{0}$, (27) becomes

$$E(\hat{\alpha} | \mathbf{P}) = \hat{\mathbf{e}}_p' \mathbf{M} \mathbf{P} (\sigma_{po}/\sigma_p^2) / (\hat{\mathbf{e}}_p' \hat{\mathbf{e}}_p) = \sigma_{po}/\sigma_p^2 = \alpha$$

where \mathbf{M} is defined previously. This proves (25) and, therefore, α is unbiased even when the parents are selected. The sampling distribution of the estimator, however, will be altered due to selection, to an extent depending on the form of selection.

Numerical example

Hypothetical weaning weights (arbitrary units) for eight parent-offspring pairs of beef cattle raised under two dif-

ferent management systems are given in Table 1. The incidence matrices corresponding to these data are

$$\mathbf{W}_p = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{W}_o = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

and the matrix \mathbf{W} is

$$\mathbf{W} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

The residuals $\hat{\mathbf{e}}_p$ and $\hat{\mathbf{e}}_o$ can be calculated as

$$[\hat{\mathbf{e}}_p, \hat{\mathbf{e}}_o] = [\mathbf{P}, \mathbf{O}] - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'[\mathbf{P}, \mathbf{O}]$$

where \mathbf{P} and \mathbf{O} are the weaning weights of the parents and offspring, respectively, from Table 1. The regression estimate of offspring on parent, taking into account the different management systems, is

$$\hat{\alpha} = \hat{\mathbf{e}}_p' \hat{\mathbf{e}}_o / (\hat{\mathbf{e}}_p' \hat{\mathbf{e}}_p) = 1.06/7.63 = 0.14$$

and

$$\hat{h}^2 = 2\hat{\alpha} = 0.28$$

An estimate of the variance of $\hat{\alpha}$ can be obtained by substituting \hat{h}^2 for h^2 in (19). Thus, $\hat{\text{Var}}(\hat{\alpha}) = 0.33$. The "usual" regression estimate of h^2 for this data is 0.15.

Table 1. Data for numerical example

Parents		Offspring	
MS ^a	WW ^b	MS ^a	WW ^b
1	3.52	1	5.53
1	4.62	1	4.31
1	6.70	2	6.87
1	3.03	2	6.07
1	4.31	1	5.73
2	5.71	1	4.72
2	5.63	2	6.02
2	5.70	2	4.89

^a Management system

^b Weaning weight in arbitrary units

Conclusion

When all observations have the same mean, the estimator presented here reduces to the "usual" offspring-parent regression estimator. The distribution and variance of the "usual" estimator are given by (18) and (19), with $r = 1$, in agreement with Kendall and Stuart (1977). The variance of the "usual" regression estimator has been given incorrectly by others (e.g., Pirchner 1969; Falconer 1981) with $(n - 2)$ in the denominator of (19). As with the "usual" offspring-parent regression estimator, this generalized regression estimator remains unbiased with the selection of parents. The numerator and denominator however, will be biased and the distribution of the estimator will be altered when selection occurs.

References

- Anderson TW (1984) An introduction to multivariate statistical analysis. Wiley and Sons, New York
- Falconer DS (1981) Introduction to quantitative genetics. Longman, New York
- Hill WG (1971) Investment appraisal for national breeding programmes. *Anim Prod* 13:37-50
- Kendall M, Stuart A (1977) The advanced theory of statistics, vol 1. Distribution theory. Macmillan, New York
- Pirchner F (1969) Population genetics in animal breeding. Freeman, San Francisco
- Robertson A (1957) Optimum group size in progeny testing and family selection. *Biometrics* 13:442-450