Theor Appl Genet (1985) 70:650–654

Estimation of heritability by parent-offspring regression

G. C. J. Fernandez and J. C. Miller, Jr.
Department of Horticultural Sciences, Texas A & M University, College Station, TX 77843, USA

Received January 22, 1985
Communicated by A. L. Kahler

Summary. The implications of bias due to previous inbreeding of parents and genotype × environmental interaction on narrow sense heritability (h^2) estimates by parent-offspring regression are enumerated. To remove the bias caused by genotype × environment interaction, an analysis of covariance model could be used. In special cases, where phenotypic expression is a result of two organisms interacting, such as in symbiotic N_2 fixation, an analysis of covariance model with a test of heterogeneity of slopes is recommended. When host genotype × strain interactions are significant, separate heritability estimates for each strain are suggested to take advantage of genotype × strain interaction, which may be a major factor contributing to the expression of N_2 fixation traits.

Key words: Narrow sense heritability – Analysis of covariance – Standardized heritability estimates – Genotype × environment interaction

Introduction

Heritability in the narrow sense is important to the plant breeder, because the effectiveness of selection depends on the additive portion of genetic variance in relation to total variance (Falconer 1960).

The parent-offspring regression method is commonly used to compute heritability estimates of quantitative characters such as % seed protein, % available methionine (Kelly and Bliss 1975), heading date (Frey and Horner 1957) and dry matter yield (Casler 1982b) in both self- and cross-fertilizing crops. Some of the routinely used parent-offspring combinations in self-fertilizing crops are: F_1/F_2, F_2/F_3, and F_3/F_4 (Kelly and Bliss 1975; Mutschler and Bliss 1981; Smith and Kinman 1965). In cross-fertilizing crops, single parent/halfsibs, and midparent/fullsibs are the most frequently used parent-offspring regression combinations (Falconer 1960). The assumptions for this analysis have been discussed previously (Cockerham 1963; Dudley and Moll 1969; Vogel et al. 1980).

In some cases, plant breeders must evaluate materials for traits that are the result of two organisms interacting, as in biological nitrogen fixation. In many instances, host genotypes behave differentially in association with different rhizobial strains (Fernandez and Miller 1983; Zary 1980). It may not be appropriate to consider the interaction between host genotype and strains as a form of genotype × environment interaction, because the former interaction is the major factor that determines the expression of N_2 fixation traits. Under these circumstances, separate heritability estimates for each strain may be more appropriate than an average estimate.

The objectives of this paper are to review the theory of parent-offspring regression, point out the bias caused by previous inbreeding of parents and genotype × environmental interaction (GE), and suggest approaches to remove the GE bias and to estimate heritability in special cases where host genotype × strain interactions are significant, as in symbiotic N_2 fixation.

Theory and analysis

Statistical model

The statistical model of the parent-offspring regression is

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Where

Y_i = mean of progenies of i^{th} family
β_0 = intercept
β_1 = regression coefficient
X_i = mean of the single parent or the mid-parent of i^{th} family
ε_i = random error, independent and normally distributed with 0 mean and σ^2 variance.

The regression coefficient measures the proportion of parent-offspring covariance (Cov P-O) to the variance of the parent (σ^2_P) ($\beta_1 = \text{Cov P-O}/\sigma^2_P$). The two most commonly used regression estimates are (i) $2\beta_1 = h^2$ and (ii) $\beta_1 = h^2$. The first estimate is appropriate for cross-fertilizing crops when halfsib progenies are regressed on a single parent. The second estimate is applicable to cross-fertilizing crops when fullsibs are regressed on the mid-parent or to self-fertilizing crops when the F_1/F_2 and F_2/F_3 parent-offspring combinations are regressed. The Cov P-O estimates additive variance (σ^2_A) only if the dominance and epistatic components are negligible (Smith and Kinman 1965; Luciano et al. 1965). If this assumption is violated, h^2 estimates are overestimated to some degree and fall between the true h^2 and broad sense (H^2) heritability estimates.

Bias due to inbreeding of parents

The regression coefficient estimates h^2, only if the inbreeding coefficient of the parents is zero (Smith and Kinman 1965). Failure to consider previous inbreeding of the parent both in self- and cross-fertilizing crops will cause an upward bias of the heritability estimate. Although this bias generally is not very severe in cross-fertilizing crops, it is one of the major causes for the inflated h^2 estimates in self-fertilizing crops. If the inbreeding coefficient of the parents is greater than zero, as with regression of F_3 progeny means on F_2 parent means, the regression coefficient overestimates heritability. This might explain why the h^2 estimates of Kelly and Bliss (1975) and Mutschler and Bliss (1981) were larger than the H^2 estimates. Furthermore, in estimating h^2 of heading date in oat populations, Frey and Horner (1957) did not adjust for the previous inbreeding of parents when they regressed F_3 on F_2, which may account for their inflated h^2 estimates.

Smith and Kinman (1965) suggested an adjustment for inbreeding of parents in the absence of dominance and epistasis. Their estimator of h^2 is $\beta_1/2\, r_{xy}$ where r_{xy} is the coefficient of parentage which measures the degree of genetic relationship between parent (x) and progeny (y) (Kempthorne 1957). The estimator $\beta_1/2\, r_{xy}$ is adjusted for mating systems and for known levels of inbreeding of parents both in self- and cross-fertilizing crops. The coefficient of parentage (r_{xy}) for various parent-offspring relationships under continuous self-pollination is presented in detail by Smith and Kinman (1965).

Scale effects of GE on h^2 estimation

The implications of genotype × environmental interaction on h^2 for parent-offspring regression have been reported in detail (Frey and Horner 1957; Vogel et al. 1980; Casler 1982a). A special effect of GE which arises from changes in scale from parent to progeny is one of the reasons why h^2 estimates are greater than unity (Frey and Horner 1957). Any environmental factor which tends to increase or decrease the range of covariance between parent and progeny could substantially affect heritability estimates. If the scale effect is positive, it will result in overly optimistic expectation of genetic gain. To reduce this scaling effect of GE, Frey and Horner (1957) suggested a method for calculating heritability in standard units, where the regression of progeny mean on parent mean is performed on coded data in terms of standard deviation. Heritability in standard units is equal to the simple linear correlation coefficient between parent mean and offspring mean. The heritability value always falls between 0 and ± 1, whereas regression coefficients sometimes fall outside of this range. The standardized h^2 estimate is equal to the correlation coefficient, $[\gamma = \text{Cov P-O}/(\sigma_P \times \sigma_O)]$ which gives only the degree of linear association between parent mean and progeny mean. In estimation of heritability by parent-offspring regression, β_1 estimates the ratio of Cov P-O to σ^2_P which is equal to h^2. In the case of γ, the numerator is equal to Cov P-O but the denominator is $(\sigma_P \times \sigma_O)$. The γ (standardized h^2) is not a true heritability estimate. In a simple linear regression model, γ (coefficient of correlation) does not have a clear-cut operational interpretation like β_1 (regression coefficient) (Neter et al. 1983). In estimating heritability, γ cannot be substituted for β_1 except when $\sigma_P = \sigma_O$. Theoretically, h^2 estimates should not exceed unity or H^2 estimates, but sometimes this does occur (Kelly and Bliss 1975). Under these circumstances, a transformation to obtain a more normal distribution seems to help, and is better than disregarding errors in the model by using γ (Robinson 1963).

Bias caused by GE on h^2 estimation

One assumption in parent-offspring regression is that environmental correlation among parent and offspring does not exist. If parents and progenies are evaluated in identical environments and in the same replicates, the observed $\text{Cov P-O} = \text{Cov P-O}_{(G)} + \text{Cov P-O}_{(GE)} + \text{Cov P-O}_{(E)}$ (Casler 1982a). Where:

$\text{Cov P-O}_{(G)}$ = the genetic component of Cov P-O
$\text{Cov P-O}_{(GE)}$ = the bias caused by GE
$\text{Cov P-O}_{(E)}$ = the bias caused by environment.

The $\text{Cov P-O}_{(E)}$ will have an expectation of zero when parents and offspring are randomized with respect to each other. If the parents and offspring are evaluated in the same environments, $\text{Cov P-O}_{(GE)}$ remains a portion of $\text{Cov P-O}_{(G)}$ and will bias h^2 estimation.

To remove the $\text{Cov P-O}_{(GE)}$ bias, regression of progeny means from one environment on parent means from another environment has been suggested (Casler and Hovin 1980; Vogel et al. 1980). Because parents and progeny are evaluated in separate environments, genotype × environment interaction bias on heritability cannot exist (Casler 1982a). To compute $\text{COV P-O}_{(GE)}$ and $\text{Cov P-O}_{(E)}$, Casler (1982a) equated expected mean squares and removed these biases from the observed Cov P-O. He then removed the GE bias by using analysis of variance and covariance models and variance and covariance component estimates. When parent and progeny genotypes respond differentially to different environments, i.e. when genotype × environment interaction is significant, Casler's method estimates h^2 free from genotype × environment interaction.

Analysis of covariance – another approach to estimating h^2 free from GE bias

We propose another method using analysis of covariance which is a combination of analysis of variance and regression (Freund and Littel 1981). A detailed description of the assumptions and theoretical aspects of this analysis can be found in most statistical texts (Snedecor and Cochran 1971; Steel and Torrie 1980).

Statistical model

$Y_{ijk} = \beta_0 + \beta_1 X_{ijk} + G_i + E_j + (GE)_{ij} + \varepsilon_{ijk}$
$i = 1 \ldots\ldots\ldots$ n families
$j = 1 \ldots\ldots\ldots$ m environments
$k = 1 \ldots\ldots\ldots\ldots$ r replicates.

Where

Y_{ijk} = F_2 mean of i^{th} family in j^{th} environment in k^{th} replicate
β_0 = intercept
β_1 = regression coefficient of the covariate
X_{ijk} = F_1 mean of i^{th} family in j^{th} environment in k^{th} replicate (the covariate)
G_i = the effect of i^{th} family
E_j = the effect of j^{th} environment.
$(GE)_{ij}$ = the effect of GE.
ε_{ijk} = random error.

Progeny mean (F_2) is the dependent variable and the parent mean (F_1) is the concomitant or covariate in this analysis of covariance model. Like the dependent variable, the covariate is also measured on each experimental unit (Freund and Littel 1981). The classification variables (family and environments) also influence the covariate (F_1 mean); however, the F_1 mean is not confounded within the family. The Cov X-Y estimates the Cov F_1-F_2 which is adjusted for the treatment effects such as G, E, and GE (Snedecor and Cochran 1971; Steel and Torrie 1980). When the environmental effects are considered fixed and manipulated, β_1 is an unbiased estimate of h^2.

A special case – host genotype × strain interaction

Symbiotic N_2 fixation is a complex phenomenon where the expression of N_2 fixation traits is determined by the host genotype, the rhizobial strain, host genotype × strain interaction, and the environment in which they grow. Recent studies have shown that superior single strains performed better than the commercial "EL" and native strains (Fernandez and Miller 1982). The use of a single strain as opposed to mixed strains in the inheritance studies of seed legumes is recommended (Riley 1979). In special instances, as in symbiotic N_2 fixation, the host genotype × strain interaction is different from genotype × location or genotype × year interactions. Estimating h^2 for N_2 fixation traits that are free from genotype × strain interaction may not be correct. Because any plant breeding selection program for N_2 fixations traits has to be conducted with a given strain, specific heritability estimates might be more appropriate than averaging the estimates. Hence, we are proposing a regression model using the analysis of covariance approach that allows for testing (1) the significance of the linear relationship between parent and offspring means ignoring the effects of strains, (2) the presence of a scaling effect (different intercepts) due to different strains (main effect of strain) assuming a single regression relationship, and (3) whether or not host genotype × strain interaction exists from the significant levels of Type I SS (Freund and Littel 1981). Type I SS, commonly known as the sequential sums of squares, corresponds to a proportion of sums of squares due to the individual variables as they are added sequentially to the regression model. Type I SS are dependent on the ordering of the variables in the model. Type IV SS is refered to as partial or the adjusted sums of squares, where sums of squares of each variable is adjusted for all other variables in the model. Type I SS is useful in testing for heterogeneity of slopes while Type IV SS is used to test the individual effect of a variable after it is adjusted for other variables in the model (Freund and Littel 1981). These SS can be obtained using the SAS General Linear Models procedure (SAS 1982).

Statistical model

For analysis of covariance with estimation of heterogeneity of slopes:

$Y_{ijk} = \beta_{0j} + \beta_{1j} X_{ijk} + G_i + P_k + \varepsilon_{ijk}$
$i = 1 \ldots\ldots\ldots$ n families
$j = 1 \ldots\ldots\ldots$ m strains
$k = 1 \ldots\ldots\ldots\ldots$ k periods.

Where

Y_{ijk} = F_2 mean of i^{th} family with j^{th} strain in k^{th} period
β_{0j} = intercept for j^{th} strain
β_{1j} = regression coefficient for j^{th} strain
X_{ijk} = the covariate, F_1 mean for i^{th} family with j^{th} strain in k^{th} period
G_i = i^{th} family
P_k = k^{th} period
ε_{ijk} = random error

In the case of two strains, two β_1 could be estimated and tested as follows:

H_0: $\beta_{1j} = \beta'_{1j}$ for all $j \neq j'$
H_a: $\beta_{1j} \neq \beta_{1j}$ for all $j \neq j'$

For example, in a genetic study, nine F_1 hybrid families (10 plants/family) and their F_2 progenies

(40 plants/family) of mungbean (*Vigna radiata*) were inoculated with two rhizobial strains, 31Z3 and 41Z2 and were grown in a greenhouse at two different periods (January–March and April–June). At seven weeks after planting, plants were harvested and the nitrogenase activity of the nodules was measured. F_1 and F_2 means were computed for each family inoculated with a specific strain and grown at a given period. The two time periods were considered as two blocks in the analysis.

Analysis one

Heritability estimation by parent-offspring regression without estimation of family × strain interaction was performed using a SAS General Linear Models procedure (SAS 1982). F_1 mean, family, time period and family × period were fitted sequentially in the model and both Type I and Type IV SS were computed and tested for statistical significance. The results of this analysis are presented in Table 1.

Both unadjusted SS (Type I) and adjusted SS (Type IV) for the covariate (F_1 mean) were statistically significant. The effects of family, period and their interactions were not statistically significant. The regression coefficient (β_1) was estimated from the adjusted SS and cross products of F_2 mean and F_1 mean using the 'ESTIMATE' option in the GLM procedure (SAS 1982). The regression coefficient estimate, neglecting strain and host genotype × strain interaction, was 0.34 ± 0.11. Because the F_2 mean was regressed on the F_1 mean, an adjustment for inbreeding of parents was not necessary (Smith and Kinman 1965) and the regression coefficient directly estimated h^2 which was equal to 0.34. However, the standardized heritability computed from the correlation coefficient between the F_2 mean and F_1 mean was 0.57. This example clearly shows how standardized β_1 might overestimate h^2.

Table 1. Heritability estimation by analysis of covariance unadjusted for the effects of strain and genotype × strain interaction

Source	df	Type I SS	Type IV SS	R^2
F_1 mean	1	2.063**	0.482**	0.75
Family	8	0.435	0.469	
Period	1	0.161	0.145	
Family × period	8	0.126	0.126	
Error	17	0.896		
	β_1		γ^a	
h^2	0.34 ± 0.11		0.56 ± 0.11	

[a] Standardized heirtability estimate
** Significant at 0.01 level

Table 2. Heritability estimation by analysis of covariance with heterogeneity of slopes adjusted for the effects of strain and genotype × strain interaction

Source	df	Type I SS	Type IV SS	R^2
F_1 mean	1	2.063***	0.544**	0.92
Family	8	0.435	0.275	
Period	1	0.161*	0.108*	
Family × period	8	0.126	0.099	
Strain	1	0.058	0.424**	
F1 mean × strain	1	0.478***	0.478***	
Error	15	0.359		
	β_{1j}	S.E		
h_j^{2a}	0.70	0.11		
$h_j^{2,b}$	0.13	0.11		

[a] Heritability estimate for strain 31Z3
[b] Heritability estimate for strain 41Z2
***, **, and * Significant at 0.001, 0.01 and 0.05 levels

Analysis two

The variables F_1 mean, family, period, family × period, strain and F_1 mean × strain were sequentially fitted and the significance levels of both Type I and Type IV SS were examined. The results of heritability estimation by analysis of covariance with host genotype × strain interaction are presented in Table 2. The Type I SS for strain was not significant, indicating that the sums of squares due to different intercepts with different strains assuming a single regression relationship, was not significant. Heterogeneity of slopes was evident, because the Type I SS due to F_1 mean × strain was significant. Hence, estimating two regression coefficients, one for each strain, should be more appropriate than an average estimate. Based on the significance of Type IV SS, the effects of family and family × period interaction were not significant; however, the adjusted sums of squares for F_1 mean, period, strain and F_1 mean × strain were significant.

The results indicated that the regression coefficients are different for different rhizobial strains; however, the main effects of rhizobial strains were not significant. These results also indicated that selecting genotypes for higher plant specific activity in the mungbean population will be more efficient with 31Z3 strain inoculation than with strain 41Z2. Hence, more emphasis should be given to the specific genotype × rhizobial strain combination in estimating h^2.

Conclusion

In summary, heritability estimation by the parent-offspring regression method is inflated when previous

inbreeding of the parents is not considered and GE effects are not removed from the h^2 estimates. These inflated heritability estimates may result in overly optimistic expected gains from selection.

To adjust for the bias caused by inbreeding of parents, both in self- and cross-fertilizing crops, the regression coefficient should be adjusted for the mating systems and known levels of inbreeding of the parents. Different environments which tend to decrease or increase the magnitude of covariances between parents and offspring also bias the h^2 estimates. Standardized parent-offspring regression has been suggested as a method for removing the scale effect of GE on h^2. Heritability in standardized units simply represents the correlation coefficient between parent and offspring, does not have clearcut interpretation in regression, and frequently overestimates h^2.

When the parents and progenies are evaluated in the same environment, the observed Cov P-O is biased because it also includes the Cov P-$O_{(GE)}$. To eliminate GE bias in heritability estimates, progeny means from one environment should be regressed on parent means from separate environments. However, the scaling effects of different environments on parent and progeny might inflate h^2 estimation. Casler (1982a) used the analysis of variance approach to estimate variance and covariance, remove the GE bias from the covariances between parent and progeny, and estimate unbiased h^2. The analysis of covariance approach also can be used as a tool for taking advantage of GE effects in the expected Cov P-O when the target environment can be manipulated. This analysis can be efficiently performed by using a popular statistical package such as SAS.

In special cases, as in symbiotic N_2 fixation, the effect of host genotype × strain interaction is different from that of genotype × environment interaction. Analysis of covariance with estimation of heterogeneity of slopes can be used to determine whether the genotype × strain interaction is significant. Under these circumstances, we suggest that separate h^2 estimates for each strain might be more appropriate than averaging the estimates.

Acknowledgement. These studies were supported in part by a grant under A.I.D. PASA AG/TAB 610-9-76 (USDA-CSRS-59-2481-0-5-001-0). Texas Agricultural Experiment Station Technical Article 20147.

References

Casler MD, Hovin AW (1980) Genetics of vegetative stand establishment characters in reed canary grass clones. Crop Sci 20:511–515

Casler MD (1982a) Genotype × environment interaction bias to parent-offspring regression heritability estimates. Crop Sci 22:540–542

Casler MD (1982b) Parent-offspring regression in reed canary grass: methods for parent and offspring evaluation and their effects on heritability estimate. Can J Genet Cytol 24: 467–473

Cockerham CC (1963) Estimation of genetic variances. In: Hanson WD, Robinson HF (eds) Statistical genetics and plant breeding. Natl Acad Sci Nat Res Council Publ 982. Washington DC, pp 53–94

Dudley JW, Moll RH (1969) Interpretation and use of estimates of heritability and genetic variances in plant breeding. Crop Sci 9:257–261

Falconer DC (1960) Introduction to quantitative genetics. Ronald Press, New York

Fernandez GCJ (1983) Heritability of and inter-relationship among N_2 fixation variables in two mungbean populations. MS Thesis, Texas A & M University

Fernandez GCJ, Miller JC Jr (1982) Influence of host genotype × *Rhizobium* strain on nitrogen fixation in mungbean. HortScience 17:152

Fernandez GCJ, Miller JC Jr (1983) Interrelationships among N_2 fixation variables in mungbean (*Vigna radiata* (L.) Wilczek). HortScience 18:84

Freund RJ, Littell RC (1981) SAS for linear models. A guide for the ANOVA and GLM procedures. SAS Inst Inc, Cary, NC

Frey KJ, Horner T (1957) Heritability in standard units. Agron J 49:59–62

Kelly JD, Bliss FA (1975) Heritability estimates of percentage seed protein and available methionine and correlations with yield in dry beans. Crop Sci 15:753–757

Kempthorne O (1957) An introduction to genetics statistics. Wiley and Sons, New York

Luciano A, Kinman ML, Smith JD (1965) Heritability of self-incompatibility in the sunflower (*Helianthus annus*). Crop Sci 5:529–532

Mutschler MA, Bliss FA (1981) Inheritance of bean seed globulin content and its relationship to protein content and quality. Crop Sci 21:289–294

Neter J, Wasserman W, Kutner MH (1983) Applied linear regression models. Richard D Irwin, Homewood, IL

Riley R (1979) Breeding for symbiotic nitrogen fixation. In: Sneep J, Hendriksen AJT (eds) Plant breeding perspectives. PUDOC, Wageningen, pp 341–348

Robinson P (1963) Heritability: a second look. In: Hanson WD, Robinson HF (eds) Statistical genetics and plant breeding. Natl Acad Sci Nat Res Council Publ 982. Washington DC, pp 609–611

SAS (1982) SAS user's guide: statistics. SAS Inst Inc, Cary NC, pp 584

Smith JD, Kinman ML (1965) The use of parent-offspring regression as an estimator of heritability. Crop Sci 5: 595–596

Snedecor GW, Cochran WG (1971) Statistical methods. Iowa State University Press, Ames IA

Steel RG, Torrie JH (1980) Principles and procedures of statistics. McGraw Hill Book Co, New York

Vogel KP, Haskins FA, Gorz HJ (1980) Parent-progeny regression: Inflation of heritability estimates by environmental covariances. Crop Sci 20:580–582

Zary KW (1980) The genetics and heritability of quantitative differences in N_2 fixation in cowpea (*Vigna unguiculata* (L.) Walp). PhD Dissertation, Texas A & M University