

Effect of aberrant values on estimation of heritability

V K BHATIA¹ and AMRIT KUMAR PAUL²

Indian Agricultural Statistics Research Institute, New Delhi 110 012 India

Received: 14 May 2005; Accepted: 26 July 2006

Key words: Aberrant values, Heritability

Heritability is one of the most important genetic parameters widely used in plant and animal breeding genetic improvement studies. In literature several methodologies are available for estimation of heritability for different experimental situations. Unfortunately none of these provide always a valid estimate of heritability and the estimate is so inadmissible that no conclusion can be drawn for the inheritance of the trait under consideration. Further in particular, there is no unique methodology, which is suitable for estimation of heritability in unbalanced situations. Keeping in view this fact, need has been felt that with the help of computer, the sensitivity and robustness of the very widely used genetic parameter, heritability might be studied at length. The sensitivity here refers to that how the estimate of heritability is dependent upon the aberrant or outliers, i.e. whether one or two abnormal data points will really affect the estimate substantially or not. Robustness means that to what extent the estimate of heritability depends upon the underlying statistical distributional assumptions such as that of normality, independence etc. Sandhya Kiran *et al.* (2004) however studied robust method of estimation of heritability in the case of fullsib data.

For studying the influence of aberrant values, its effect can be incorporated in the usual model

$$y_{ij} = \mu + s_i + e_{ij} \quad i=1,2,\dots,s; j=1,2,\dots,n$$

where y_{ij} is the response variable with the sire effect s_i and the random component e_{ij} are independently distributed $N(0, \sigma_s^2)$ and $N(0, \sigma_e^2)$ respectively. The influence of aberrant value, assumed to be fixed in terms of measurement error etc., can be incorporated by considering $y_{ij} = \mu + C + s_i + e_{ij}$ depending upon $C > 0$ or $C < 0$, the influence will be in 'upward' or 'downward' respectively. Its actual effect into the variance component and estimate of heritability is required to be examined. Through simulation studies, the effect of different values of 'C' can be examined to take on the estimates of variance components.

As the heritability is a function of variance component so

Present address: ¹Principal Scientist, ²Scientist

the methods used for estimation of variance component estimation can be used. Some of them are—ANOVA (Henderson 1953), REML (Patterson and Thompson 1971), ML (Harville 1977), and Robust method (Mehrotra 1995).

In the present investigation, the simulation procedures given by Ronningen (1974) have been used. The basic distribution assumed for the random effects is normal distribution, for which standard normal variate $N(0,1)$ generator is used. It is well known that by multiplying a standard normal variate by a given value of standard deviation say σ_s we can get a variable that has $N(0, \sigma_s^2)$ distribution. This technique is precisely used for simulation of data for estimation of heritability. In the present study we considered the parametric value of heritability 0.25 and 0.30. Computer programs for analysis are written in SAS-IML (1990). The heritabilities are computed in the presence of different aberrant values and other conditions and tabulated in Tables 1 and 2.

Table 1. Estimates of heritability obtained by different methods in the presence of outliers for the population parametric value of heritability as 0.25

Extent of outlier	Methods	With inadmissible estimates	Without inadmissible estimates
5-times the normal records	ANOVA	0.1727	0.2568
	REML	0.1924	0.2468
	ML	0.1625	0.2167
10-times the normal records	ANOVA	0.1981	0.2750
	REML	0.2173	0.2750
	ML	0.1862	0.2551
20-times the normal records	ANOVA	0.0502	0.1141
	REML	0.0835	0.1141
	ML	0.0647	0.1221
50-times the normal records	ANOVA	0.0075	0.0600
	REML	0.0312	0.0600
	ML	0.0176	0.0477
100-times the normal records	ANOVA	0.0006	0.0295
	REML	0.0145	0.0295
	ML	0.0042	0.0178

Table 2. Effect of aberrant values on the estimate of heritability with 4 sires having 40 daughters each for the parameter value of $h^2=0.3$. ANOVA estimate=0.3310, Robust estimate=0.3042

Sire	Changed value		ANOVA estimate	Robust estimate
	From	To		
1	1.6600	16.60	0.1937	0.2990
	1.6600	166.00	0.0133	0.2990
	1.6600	1660.00	0.0012	0.2990
2	0.6770	6.77	0.1084	0.3665
	0.6770	67.70	-0.0674	0.4291
	0.6770	677.00	-0.0081	0.4324
3	0.6770	6770.00	-0.0008	0.4379
	-0.9070	-9.07	0.1159	0.1981
	-0.9070	-90.70	-0.0145	0.1877
4	-0.9070	-907.00	-0.0021	0.1877
	-0.0084	-0.08	0.3306	0.3040
	-0.0084	-0.84	0.3218	0.3256
	-0.0084	-8.40	-0.1965	0.3084
	-0.0084	-84.00	-0.0103	0.3013
	Combination of two observations			
	1.6600	16.60		
	0.6770	6.77	0.0944	0.3438
	1.6600	166.00		
	0.6770	67.70	-0.0231	0.4144
	1.6600	1660.00		
	0.6770	677.00	-0.0236	0.4228
	Combination of two or more observations			
	1.6600	1660.00		
	0.6770	677.00		
	-0.9070	-907.00		
	-0.0084	-84.00	0.0177	0.3300
	0.6770	677.00		
	-0.9070	-907.00		
	-0.0084	-84.00	0.0261	0.3192
	-0.9070	-907.00		
	-0.0084	-84.00	-0.0084	0.1826
	1.6600	1660.00		
	-0.9070	-907.00	0.0282	0.2113
1.6600	1660.00			
-0.0084	-84.00	0.0045	0.3280	
0.6770	667.00			
-0.9070	-907.00	0.0274	0.3280	
0.6770	677.00			
-0.0084	-84.00	-0.0001	0.4308	

It is well known that the variance components are very much sensitive to abnormal observations. It is being seen that even one single observation in the data set affects significantly to the estimate of variance component and

heritability. To study the influence of aberrant value, the aberrant value(s) were created deliberately by changing the normal value to varying degree of aberrant values. Table 1 showed that in the presence of aberrant value none of the classical methods give the reliable, correct and precise estimate of heritability. However, when the aberrant value is not far of from the rest of the data set then one can depend upon to a certain extent on the procedure of REML. In the case of extreme observations even REML also fails to yield estimate close to the parametric values. This can easily be seen from Table 1. Since REML did not result always a reliable and accurate estimate of heritability, the application of robust method of estimation of variance component was also carried out and the result thus obtained are presented in Table 2. From the data it is clearly seen that in most of the situations robust estimate is very close to the parametric value of heritability.

SUMMARY

The methods of ANOVA and REML failed miserably in almost all the cases and it is thus concluded that for most of the practical situations having presence of aberrant values one should always use robust method of estimation of variance components, which yielded accurate and reliable estimates of heritability.

REFERENCES

Harville D A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 72: 320-40.

Henderson C R. 1963. Selection index and expected genetic advance, *Statistical Genetics and Plant Breeding*. National Research Council Publication 98L, National Academy of Sciences. pp. 141-63.

Mehrotra D V. 1995. Robust elementwise estimation of a dispersion matrix. *Biometrics* 51: 1344-51.

Patterson H D and Thompson R. 1971. Recovery of inter block information when block sizes are unequal. *Biometrika* 58: 545-54.

Ronningen K. 1974. Monte-Carlo Simulation of Statistical-Biological Models which are of interest in Animal Breeding. *Acta Agriculturae Scandinavica* 24: 135-42.

Sandhya Kiran P, Bhatia V K and Rao A R. 2004. Robust method of estimation of heritability. *Journal of Indian Society of Agricultural Statistics* 57 (special volume): 116-28.

SAS-IML. 1990. SAS documentation, SAS Institute Inc. Cary, NC, USA.