Indian J. Genet., **48**(3): 275–287 (1988)

SAMPLING DISTRIBUTION AND APPROXIMATIONS TO BIAS AND MEAN SQUARE ERROR OF GENOTYPIC CORRELATION FROM PLANT BREEDING DATA

M. SINGH

International Crops Research Institute for the Semi-Arid Tropics ICRISAT, Patancheru P.O. 502 324

(Received: November 15, 1985; accepted: December 21, 1987)

ABSTRACT

The distributional behaviour using simulation and approximations to bias, mean square error of genotypic correlation when estimated from plant breeding data have been obtained and illustrated with field data from a yield trial on pearl millet. Empirical results have been discussed for the variation in several parameters.

Key words: Genotypic correlation, standard error, bias, simulation, normality.

The large sample variance of genetic correlation coefficient estimated from parent offspring covariance has been discussed [1] and its sampling distributions have been investigated [2, 3] using simulations. Similar results are not available for plants where the genotypic correlation between two characters is obtained from the genotypic variances and covariance. In the case of plants, a number of genotypes are often grown in a replicated trial and two characters are measured on each plot, unlike the parent offspring situation in animals where two traits are measured on the parents and their offspring as well. The purpose of this paper is to provide with the large sample, bias and mean square error (MSE) and study empirically the distribution of the estimate of genotypic correlation from plant breeding data.

MATERIALS AND METHODS

Consider the estimation of the genotypic correlation coefficient between two characters x and y from v genotypes grown in a single environment of r randomized blocks. The bivariate analysis of variance is presented in Table 1.

T_{xx}, T_{yy} and T_{xy} are the corrected sums of squares for characters x and y and the corrected sum of products of x and y due to genotypes. Similarly, E_{xx}, E_{yy}, E_{xy} stand for error components. σ^2_{px}, σ^2_{py}, and σ_{pxy} are related to genotypic population variances σ^2_{gx}, σ^2_{gy} and σ_{gxy} and error variances σ^2_{ex}, σ^2_{ev} and σ^2_{exy}, as

$$\Sigma_p = \Sigma_e + r\,\Sigma_g \qquad (1)$$

Table 1. Bivariate analysis of variance

Source	d.f.	SSP matrix	Expected value of SSP matrix
Blocks	r−1		
Genotypes	q=v−1	$\underset{\sim}{T} = \begin{bmatrix} T_{xx} & T_{xy} \\ T_{xy} & T_{yy} \end{bmatrix}$	$q \begin{bmatrix} \sigma^2_{px} & \sigma_{pxy} \\ \sigma_{pxy} & \sigma^2_{py} \end{bmatrix} = q\underset{\sim}{\Sigma}_p$
Error	$\nu=q(r-1)$	$\underset{\sim}{E} = \begin{bmatrix} E_{xx} & E_{xy} \\ E_{xy} & E_{yy} \end{bmatrix}$	$\nu \begin{bmatrix} \sigma^2_{ex} & \sigma_{exy} \\ \sigma_{exy} & \sigma^2_{ey} \end{bmatrix} = \nu\underset{\sim}{\Sigma}_e$

SSP—sums of squares and products.

The genotypic correlation ρ_G between x and y is defined as

$$\rho_G = \sigma_{gxy}/(\sigma_{gx}\ \sigma_{gy}) \tag{2}$$

The matrices $\underset{\sim}{T}$ and $\underset{\sim}{E}$ are distributed independently as central Wishart distributions $W_2(q, \underset{\sim}{\Sigma}_p)$ and $W_2(\nu, \underset{\sim}{\Sigma}_e)$, respectively [4]. Thus, an unbiased estimate of $\underset{\sim}{\Sigma}_g$ can be written as

$$\underset{\sim}{\Sigma}_g = (1/q)\ \underset{\sim}{T} - (1/\nu)\ \underset{\sim}{E} \tag{3}$$

using

$$E(\underset{\sim}{T}) = q\underset{\sim}{\Sigma}_p,\ E(\underset{\sim}{E}) = \nu\underset{\sim}{\Sigma}_e$$

where E(.) stands for expectation of the variables in parentheses. Using unbiased estimate of σ^2_{gx}, σ^2_{gy} and σ_{gxy} from (3), ρ_G is estimated by

$$r_G = \hat{\sigma}_{gxy}/(\hat{\sigma}_{gx}\ \hat{\sigma}_{gy}) \tag{4}$$

where

$$\hat{\sigma}_{gxy} = (T_{xy}/q - E_{xy}/\nu)/r$$
$$\hat{\sigma}_{gx} = ((T_{xx}/q - E_{xx}/\nu)/r)^{\frac{1}{2}}$$
$$\hat{\sigma}_{gy} = ((T_{yy}/q - E_{yy}/\nu)/r)^{\frac{1}{2}}$$

It is easy to see that r_G is not comparable with the r_g of Brown [3], as their expressions are different.

The estimator r_G of ρ_G is obviously not unbiased and may also lie outside the range [−1, 1] because the inequality $|\sigma_{gxy}| < \hat{\sigma}_{gx}\ \hat{\sigma}_{gy}$ may be violated when using data values. The probability of getting negative estimates of σ^2_{gx} has been discussed by Gill and Jensen [5], in which case r_G may take imaginary values, and of getting nonpositive definite genetic covariance matrix by Hill and Thompson [6].

BIAS AND MEAN SQUARE ERROR ESTIMATE OF GENOTYPIC CORRELATION

We use the following results.

Lemma [7]. Suppose X_i has mean θ_i and variances and covariance of the k variates $X_1, X_2, \ldots\ldots X_k$ are of order $n^{-p}(p>0$, in practice taken as unity), and n is the sample size used in constructing X_i for θ_i. Consider a scalar function $g(X_1, \ldots, X_k)$ denoted briefly by g(x). The expectation and MSE of g(x) are

$$E(g(x)) = g(\theta) + (1/2)\,\Sigma\Sigma g''_{ij}(\theta)\,\mathrm{Cov}(x_i, x_j) + o(n^{-p}) \quad (5)$$

$$\mathrm{MSE}\,(g(x)) = \sum_i\sum_j g'_i(\theta) g'_j\,\mathrm{Cov}(x_i, x_j) + o(n^{-p}) \quad (6)$$

where

$$g'_i(\theta) = \partial g(x)/\partial x_i$$
$$g''_{ij}(\theta) = \partial^2 g(x)/\partial x_i\,\partial x_j$$

evaluated at $\theta_1, \theta_2, \ldots, \theta_k$ including the covariance terms between $X_1, \ldots, X_k$ in our approximation given in (5).

Since r_G is a function of variables T_{xx}, T_{yy}, T_{xy}, E_{xx}, E_{yy}, E_{xy}, the lemma is applied to obtain bias and MSE of r_G in terms of the means and, variances and covariances of these variables. Further, we note the following [4].

$$D\begin{pmatrix} T_{xx} \\ T_{yy} \\ T_{xy} \\ E_{xx} \\ E_{yy} \\ E_{xy} \end{pmatrix} = 2q \begin{pmatrix} \sigma^4_{px} & \rho^2_p\sigma^2_{px}\sigma^2_{py} & \rho_p\sigma^3_{px}\sigma_{py} \\ & \sigma^4_{py} & \rho_p\sigma_{px}\sigma^3_{py} \\ \text{Symmetric} & & (1+\rho^2_p)\sigma^2_{px}\sigma^2_{py}/2 \\ & & \\ \sigma^4_{ex} & \rho^2_e\sigma^2_{ex}\sigma^2_{ey} & \rho_e\sigma^3_{ex}\sigma_{ey} \\ & \sigma^4_{ey} & \rho_e\sigma_{ex}\sigma^3_{ey} \\ \text{Symmetric} & & (1+\rho^2_e)\sigma^2_{ex}\sigma^2_{ey}/2 \end{pmatrix}$$

$$\mathrm{Cov}\,(T_z, E_{z'}) = 0 \text{ for } z, z' = xx, yy, xy;$$

where D(.) represents the dispersion (variance-covariance) matrix of the vector variables in parentheses. The covariances between genotype sums of squares or product and error sums of squares or product are zero.

Treating T_{xx}, T_{yy}, T_{xy}, E_{xx}, E_{yy}, E_{xy} as $x_1, x_2, \ldots, x_6$, in the lemma, we obtain after some algebraic simplifications, the following approximations:

$$g(\theta) = \rho_G$$

$$\begin{aligned}\text{bias}\,(r_G) &= E(r_G) - \rho_G = (1/2)\,\Sigma\Sigma g''_{ij}(\theta)\,\mathrm{Cov}\,(x_i, x_j) \\ &= (1/2r^2)\,[(1/q)\,\{(3\rho_G/2)\,(G^2_x+G^2_y)+\rho_p(\rho_p\rho_G G_xG_y - 2\sqrt{(G_xG_y)(G_x+G_y)})\} \\ &\quad + (1/\nu)\,\{(3\rho_G/2)\,(g^2_x + g^2_y) + \rho_e(\rho_e\rho_G g_xg_y - 2\sqrt{(g_xg_y)\,(g_x+g_y)})\}] \end{aligned} \quad (7)$$

$$\begin{aligned}\mathrm{MSE}\,(r_G) &= E(r_G - \rho G)^2 \\ &= \Sigma\Sigma g'_i(\theta)\,\mathrm{Cov}\,(x_i, x_j) \\ &= (1/2r^2)\,[(1/q)\,\{\rho^2_G(G^2_x + G^2_y) + 2((1+\rho^2_p+\rho^2_p\,\rho^2_G)G_xG_y \\ &\qquad -2\rho_p\rho_G\,\sqrt{G_xG_y(G_x+G_y)})\} \\ &\qquad + (1/\nu\,\{\rho^2_G(g^2_x+g^2_x) + 2((1+\rho^2_e+\rho^2_e\rho^2_G)\,g_xg_y \\ &\qquad - 2\rho_G\rho_e\,\sqrt{g_xg_y}\,(g_x+g_y))\}\,] \end{aligned} \quad (8)$$

where $g_z = \sigma^2_{ez}/\sigma^2_{pz}$

$G_z = g_z + r$ for $z = x, y$

$$\mathrm{SE}\,(r_G) = (\mathrm{MSE}\,(r_G))^{\frac{1}{2}} \quad (9)$$

where SE (.) stands for standard error.

The evaluations of $g'(\theta)$ and $g''(\theta)$ have been given in Appendix I for only nonzero covariance contributing variables.

The estimates of bias and MSE can be obtained by substituting the estimates of ρ_G, ρ_e, ρ_p, g_x, and g_y using the estimate for variance-covariance components given in equation (3).

DISTRIBUTION OF r_G

The first approximation to the expansion of the expression for r_G yields a linear function of the above six variables which are sums (means) of some other random quantities. Thus, using the Central Limit Theorem [8] the distribution of r_G approaches normal distribution as the number of genotypes increases [7]. In order to obtain an idea of the number of genotypes to be considered so that the distribution of r_G approximates reasonably to normal distribution, a simulation study was carried out involving a wide range of parameters. The comparisons between the simulated values of bias and standard errors have also been made with those from the approximations considered in expressions (7) and (8).

Generation of a bivariate normal pair. We carried out simulations with parametric values estimated from the analyses of the real data for various sets of number of genotypes, means, variances and correlations. The RANDU function available in GENSTAT package was used to generate uniform random variables and then standard normal deviates were generated using the relation [9]. If u_1 and u_2 are two independent uniform random variables, then

$$x_1 = (-2 \log_e u_1)^{\frac{1}{2}} \operatorname{Sin}(2\pi u_2)$$

$$x_2 = (-2 \log_e u_1)^{\frac{1}{2}} \operatorname{Cos}(2\pi u_2)$$

are independently normally distributed with zero means and unit variances. In order to generate a bivariate normal pair (Z_1, Z_2) with parameters $(\mu_1, \mu_2, \sigma^2_1, \sigma^2_2, \rho)$ one may use

$$Z_1 = \mu_1 + \sigma_1 y_1' \quad Z_2 = \mu_2 + \sigma_2 y_2$$

where $y_1 = x_1$; $y_2 = \rho x_1 + x_2 (1-\rho^2)^{\frac{1}{2}}$; (μ_1, σ^2_1) and (μ, σ^2_2) are means and variances of Z_1 and Z_2, respectively, and ρ is the correlation coefficient between them.

Generation of a random value of r_G . Let $\underline{\zeta}_x = (\zeta_{x1}, \ldots, \zeta_{xv})'$, $\underline{\zeta}_y = (\zeta_{y1}, \ldots, \zeta_{yv})'$, $\underline{\beta}_x = (\beta_{x1}, \ldots, \beta_{xr})'$, $\underline{\beta}_y = (\beta_{y1}, \ldots, \beta_{yr})'$, σ^2_{ex}, σ^2_{ey}, and ρ_e be parameters representing, respectively, vectors of v genotype means on character x and y, vectors of r block (replications) means for x and y, error variances for x, y and error correlation between these two characters. Several sets of these parameters were taken as estimated values from a real trial (described in last section) in r=3 replications and v=96 genotypes, and selecting for various values of v=10, 20, 30, 40, 50, 60, 70, 90. A random pair (x_{ij}, y_{ij}) of observations on two characters corresponding to i−th genotype and j−th replication was obtained as

$$x_{ij} = \zeta_{xi} + (\beta_{xj} - \bar{\beta}_x) + e_{xij}$$
$$y_{ij} = \zeta_{yi} + (\beta_{yj} - \bar{\beta}_y) + e_{yij}$$

where $\bar{\beta}_z = \sum_{j=1}^{r} \beta_{zj}/r$, z=x, y, i=1, 2, ..., v, j=1,2, ..., r, z=x, y

(e_{xij}, e_{yij}) is a bivariate normal pair with parameters (0, 0, 1, 1, ρ_e) generated using the procedure discussed above. From these (x_{ij}, y_{ij}) i=1, ..., v, j=1, ..., r) vr pairs, we obtained one sample value of genotypic correlation r_G using equation (4). This process was repeated to get N independent values (simulation runs) of r_G. An alternative way to generate r_G values could be with the help of two independent Wishart matrices with parameters in Table 1.

RESULTS

We present 13 sets of values of parameters including number of runs considered in our simulation study (Table 2). In order to examine the normality of distribution of r_G, the simulated skewness and kurtosis of r_G and also of its transformation $z_G = 0.5 \log ((1+r_G)/(1-r_G))$ are presented (Table 3). The cumulative probability distribution of r_G has been compared with that of normal distribution with simulated means and standard error (Table 4) for the cases considered in Table 2.

Table 2. Sets of parameter values used for simulation with three replications

Case	v	N	σ^2_{gx}	σ^2_{gy}	σ^2_{ex}	σ^2_{ey}	ρ_g	ρ_e
1	10	1000	398.8	190796	84.9	287570	0.603	0.381
1a	10	5000	398.8	190796	84.9	287570	0.603	0.381
2	20	1000	289.5	190552	157.4	256736	0.280	0.466
2a	20	4000	289.5	190552	157.4	256736	0.280	0.466
3	30	1000	281.0	149678	145.4	232973	0.333	0.444
3a	30	3000	281.0	149678	145.4	232973	0.333	0.444
4	40	1000	264.0	218849	138.9	253247	0.427	0.353
4a	40	2000	264.0	218849	138.9	253247	0.427	0.353
5	50	1000	310.7	250984	138.1	284323	0.478	0.256
5a	50	2000	310.7	250984	138.1	284323	0.478	0.256
6	60	1000	299.6	254385	131.2	282998	0.484	0.259
7	70	1000	291.9	229793	132.4	279177	0.487	0.215
8	90	1000	283.9	282323	156.1	290868	0.441	0.184

v—No. of genotypes, N—simulation runs
Formulae to estimate:

$$\sigma^2_{gz} = \sum_{i=1}^{v} (\zeta_{zi} - \zeta_z)^2/(v-1), \text{ and } \zeta_z = \sum_{i=1}^{v} \zeta_{zi}/v, \quad z=x,y.$$

The comparison of the simulation results on bias and standard error with those computed using approximations in expressions (7) and (9) are presented in Table 5.

DISCUSSION

Number of simulation runs. The number of runs were varied from N=1000 to see whether this number is satisfactory enough to represent the underlying distributions or any marked differences are occurring in the stimulated parameters due to changes in runs. Other than N=1000 for each number of genotypes considered, we took N=5000, 4000, 3000, 2000, 2000 for v = 10, 20, 30, 40 and 50, respectively, under the consideration that for smaller sample size we need larger simulation runs for adequate representation of the distributions.

Comparing the simulated values of mean, skewness and kurtosis of r_G (Table 3) and standard error (Table 5) for the cases 1 with la, 2 with 2a, 3 and 3a, 4 with 4a, and 5 with 5a, we find a very close agreement between their values except for case 4 with 4a on skewness and kurtosis values. Similar agreement in the cumulative probability values (Table 4) obtained on two different simulation runs can be observed. Thus N=1000 runs can be safely taken for valid comparison.

Normality of r_G. Table 3 shows that values of skewness and kurtosis are insignificant for more than 50 genotypes. Also from cases 4a and 5a their magnitudes are small [10]. Except for the cases with v=10, 20 genotypes, we did not observe

Table 3. Skewness and kurtosis of r_G and its inverse tanh transformation z_G, along with their standard errors (SE) in parentheses[+]

Case No.	Skewness±SE	r_GKurtosis±SE	Skewness	z_GKurtosis
1	−0.76±0.081	0.96±0.161	0.70	0.91
1a	−0.78±0.036	1.15±0.072	0.69	0.90
2	−0.51±0.077	0.95±0.155	−0.25	1.34
2a	−0.66±0.039	1.27±0.077	−0.36	1.81
3	−0.36±0.077	0.70±0.155	0.05	0.63
3a	−0.37±0.045	0.75±0.089	0.06	0.74
4	−0.31±0.077	0.47±0.155	0.75	5.86
4a	−0.18±0.055	0.25±0.109	0.58	3.37
5	−0.23±0.077	0.01±0.155	0.23	0.23
5a	−0.18±0.055	0.09±0.109	0.26	0.22
6	−0.09±0.077	0.05±0.155	—	—
7	−0.03±0.077	0.03±0.155	—	—
8	−0.06±0.077	0.07±0.155	—	—

[+]Standard error of skewness = $(6/N)^{1/2}$ and of kurtosis = $(24/N)^{1/2}$, same standard errors apply for corresponding values on z_G.

— Transformation not done.

any invalid (going outside interval [−1, 1] or imaginary) value of r_G. Thus, with three or more replications, the distribution of r_G approximates normal distribution when the number of genotypes goes beyond 50 and here the significance of genotypic correlation can be tested by comparing the estimate divided by its standard error against the table value of standard normal deviate. We also applied the inverse tanh transformation but we did not get any improvement for the cases 4 and beyond. Although there was a small reduction in skewness values but increase in kurtosis (Table 3) was with similar trend as observed by [3].

Table 4. Probability of [$r_G \leq \rho$] under simulation and normal approximation

Case		ρ											
		−0.1	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
1*	sim	0.00	0.01	0.03	0.05	0.09	0.17	0.29	0.46	0.66	0.83	0.95	1.00
	nor	0.00	0.00	0.01	0.03	0.08	0.18	0.33	0.51	0.69	0.83	0.92	0.97
1a	sim	0.01	0.01	0.02	0.04	0.08	0.16	0.28	0.45	0.64	0.83	0.95	1.00
	nor	0.00	0.00	0.01	0.03	0.08	0.17	0.31	0.50	0.68	0.83	0.97	0.97
2	sim	0.04	0.09	0.19	0.34	0.55	0.75	0.88	0.97	1.00	1.00	.00	.00
	nor	0.03	0.09	0.20	0.37	0.57	0.75	0.88	0.95	0.99	1.00	.00	.00
2a	sim	0.04	0.09	0.18	0.33	0.54	0.75	0.90	0.97	1.00	1.00	.00	.00
	nor	0.03	0.09	0.20	0.37	0.57	0.75	0.88	0.95	0.99	0.99	.00	.00
3	sim	0.01	0.03	0.08	0.21	0.41	0.69	0.88	0.97	1.00	.00	.00	.00
	nor	0.00	0.02	0.08	0.22	0.44	0.69	0.87	0.96	0.99	.00	.00	.00
3a	sim	0.01	0.03	0.07	0.19	0.40	0.67	0.87	0.97	1.00	.00	.00	1.00
	nor	0.00	0.02	0.07	0.20	0.43	0.68	0.87	0.96	0.99	.00	.00	1.00
4	sim		0.00	0.01	0.03	0.15	0.42	0.73	0.95	1.00	.00	.00	1.00
	nor		0.00	0.00	0.03	0.14	0.41	0.74	0.93	0.99	.00	.00	1.00
4a	sim			0.00	0.03	0.15	0.41	0.74	0.95	1.00	.00	1.00	1.00
	nor			0.00	0.03	0.14	0.42	0.74	0.94	0.99	.00	1.00	1.00
5	sim				0.00	0.04	0.21	0.59	0.92	0.99	.00	.00	1.00
	nor				0.00	0.03	0.21	0.61	0.91	0.99	.00	.00	1.00
5a	sim				0.00	0.04	0.21	0.60	0.91	0.99	.00	.00	1.00
	nor				0.00	0.03	0.21	0.60	0.91	0.99	.00	.00	1.00
6	sim				0.00	0.02	0.16	0.57	0.92	0.99	.00	.00	1.00
	nor				0.00	0.02	0.16	0.57	0.91	0.99	.00	.00	1.00
	sim					0.02	0.15	0.58	0.91	1.00	1.00	.00	1.00
	nor					0.01	0.15	0.57	0.92	1.00	1.00	.00	1.00
	sim					0.03	0.32	0.80	0.99	1.00	1.00	.00	1.00
	nor					0.04	0.32	0.81	0.99	1.00	1.00	.00	1.00

* In cases 1, 1a, 2, 2a, the values of r_G outside the range [−1, 1] were observed and these values were excluded from computation.
sim—under simulation, nor—under normality.

Distribution function values [$P(r_G < \rho)$; $\rho = -0.1, 0, (0.1)\ 1$] (Table 4) show a close agreement between simulated values and those on the assumption of normality for $v > 40$, $r > 3$. The results for the cases 1, 1a, 2, 2a are based on the valid values of r_G and hence do not represent the true population of r_G. From the results of Tables 3 and 4, $v \leq 30$ seems to be a low number of genotypes and cannot be recommended for applying the normality assumption when using up to three replications.

Closeness of approximations. Table 5 shows that the amount of bias is small in both simulation and approximation cases. Approximation is producing relatively higher standard errors of r_G than their simulated values. However, this difference decreases with increase in the number of genotypes. The trend of values remains similar under the inverse tanh transformation. The differences between simulated and approximated values of SE of r_G ranged from 0.03 to 0.05 for the number of genotypes between 40 and 90 compared with the true and simulated differences of 0.02 to 0.04 [3] for sample size 200.

Table 5. Bias and standard error (SE) of r_G and z_G using simulation and approximation

Case	ρ_G	100* bias(r_G)		SE(r_G)		SE(z_G)	
		sim	apprx	sim	apprx	sim	apprx
1	0.60	−0.69	−0.25	0.21	0.29	0.39	0.46
1a+		−0.05		0.21		0.39	
2	0.28	−1.38	−1.79	0.20	0.28	0.23	0.30
2a		−1.33		0.20		0.22	
3	0.33	−0.85	−1.05	0.16	0.22	0.18	0.25
3a		−0.32		0.15		0.18	
4	0.43	−0.13	−0.54	0.12	0.17	0.15	0.21
4a		−0.22		0.11		0.15	
5	0.48	−0.31	−0.25	0.09	0.15	0.13	0.19
5a		−0.26		0.09		0.13	
6	0.48	0.11	−0.21	0.09	0.13	0.12	0.17
7	0.49	−0.12	−0.12	0.08	0.12	0.12	0.16
8	0.44	−0.63	−0.12	0.08	0.11	0.10	0.14

+ Other values same as in preceding rows.
sim—simulation, apprx—using approximation.

Behaviour of bias and MSE from approximations with variations in several parameters. Bias and MSE of r_G are presented (Tables 6–8) for a range of number of genotypes, with variations in the number of replications, genotypic correlations and error correlations. Table 6 shows a uniform decrease in the values of bias and MSE with increase in r as well as v.

Table 6. Bias and MSE for variable number of genotypes (v) replications (r)

v/r	100* bias+				100* MSE+			
	3	4	5	6	3	4	5	6
10	−2.70	−2.50	−2.38	−2.31	3.82	3.53	3.37	3.26
20	−1.28	−1.18	−1.13	−1.09	1.81	1.67	1.60	1.55
30	−0.84	−0.78	−0.74	−0.72	1.19	1.10	1.05	1.01
40	−0.62	−0.58	−0.55	−0.53	0.88	0.82	0.78	0.75
50	−0.49	−0.46	−0.44	−0.42	0.70	0.65	0.62	0.60
60	−0.41	−0.38	−0.36	−0.35	0.58	0.54	0.51	0.50
70	−0.35	−0.33	−0.31	−0.30	0.50	0.46	0.44	0.43

+ Based on approximations.
The other parameters were $\sigma^2_{gx}=\sigma_{gxy}=2$, $\sigma^2_{gy}=4$, $\sigma^2_{ex}=\sigma^2_{exy}=1$, $\sigma^2_{ey}=2$.

In the range of the parameter values considered (Table 7), bias increases initially, but starts decreasing after $\sigma_{gxy} = 1.25$. Bias (Table 8) and MSE (Tables 7, 8) show an increasing trend with the values of genotypic and error correlations

Table 7. Bias and MSE for variable values of genotypic covariances (correlations) and number of genotypes (v)

σ_{gxy}	ρ_G	v						
		10	20	30	40	50	60	70
				100*bias				
0.00	0.00	−2.93	−1.39	−0.91	−0.68	−0.54	−0.45	−0.38
0.25	0.08	−3.15	−1.49	−0.98	−0.73	−0.58	−0.48	−0.41
0.50	0.16	−3.34	−1.58	−1.04	−0.77	−0.61	−0.51	−0.44
0.75	0.24	−3.48	−1.65	−1.08	−0.80	−0.64	−0.53	−0.45
1.00	0.32	−3.56	−1.69	−1.10	−0.82	−0.65	−0.54	−0.46
1.25	0.40	−3.56	−1.68	−1.10	−0.82	−0.65	−0.54	−0.46
1.50	0.47	−3.46	−1.64	−1.07	−0.80	−0.64	−0.53	−0.45
1.75	0.55	−3.24	−1.54	−1.01	−0.75	−0.60	−0.50	−0.42
2.00	0.63	−2.90	−1.37	−1.01	−0.67	−0.53	−0.44	−0.38
2.25	0.71	−2.41	−1.14	−0.90	−0.56	−0.44	−0.37	−0.31
				100*MSE				
0.00	0.00	15.46	7.33	4.80	3.57	2.84	2.36	3.02
0.25	0.08	15.03	7.12	4.66	3.47	2.76	2.29	1.96
0.50	0.16	14.29	6.77	4.44	3.30	2.63	2.18	1.86
0.75	0.24	13.27	6.29	4.12	3.06	2.44	2.03	1.73
1.00	0.32	12.01	5.69	3.73	2.77	2.21	1.83	1.57
1.25	0.40	10.54	5.00	3.27	2.43	1.94	1.61	1.38
1.50	0.47	8.92	4.23	2.77	2.06	1.64	1.36	1.16
1.75	0.55	7.22	3.42	2.24	1.67	1.33	1.10	0.94
2.00	0.63	5.50	2.61	1.71	1.27	1.01	0.84	0.72
2.25	0.71	3.86	1.83	1.20	0.89	0.71	0.59	0.50

The other parameters were r=3, $\sigma^2_{gx}=2$, $\sigma^2_{gy}=5$, $\sigma^2_{ex}=1$, $\sigma^2_{ey}=2.5$.

and decreasing over v. For v $\geq$40, biases are quite small and MSE shows small variation with increase in error correlation.

Probability of invalid estimates of ρ_G. The moment estimates of genotypic variances and covariances may lead to invalid estimates of the genotype correlation coefficient, lying outside the interval [−1, 1]. The Hill–Thompson [6] approach can be used to find the probability of getting r_G out of bound and/or heritabilities being negative. For v = 10 and 20, the probability values of observing invalid estimates of r_G using simulation and normal approximations were close: P [$|r_G| > 1$] = 0.080 (simulation) and 0.086 (approximation) for v = 10, 0.001 (simulation) and 0.004 (approximation) for v = 20, 0.000 (simulation) and 0.001 (approximation) for v = 30, and both were zero v $\geq$ 40. This indicates that for genotypes v $\geq$ 40 it is very unlikely to get an invalid value. Here treatment and error degrees of freedom exceed 39 and 78, respectively.

It would be worthwhile to explain the testing and the estimation of genotypic correlation with the help of real data on pearl millet. Table 9 presents the extracts of multivariate analyses of variance with three characters: grain yield (kg/ha), plant

Table 8. Bias and MSE for variable values of error covariance (correlation) and number of genotypes (v)

σ_{gxy}	ρ_e	v						
		10	20	30	40	50	60	70
				100*bias				
0.0	0.0	0.20	0.09	0.06	0.05	0.04	0.03	0.03
0.2	0.1	−0.33	−0.16	−0.10	−0.08	−0.06	−0.05	−0.04
0.4	0.2	−0.85	−0.40	−0.26	−0.20	−0.16	−0.13	−0.11
0.6	0.3	−1.36	−0.65	−0.42	−0.31	−0.25	−0.21	−0.18
0.8	0.4	−1.87	−0.89	−0.58	−0.43	−0.34	−0.29	−0.24
1.0	0.5	−2.38	−1.13	−0.74	−0.55	−0.44	−0.36	−0.31
1.2	0.6	−2.88	−1.37	−0.89	−0.67	−0.53	−0.44	−0.38
1.4	0.7	−3.38	−1.60	−1.05	−0.78	−0.62	−0.52	−0.44
1.6	0.8	−3.87	−1.83	−1.20	−0.89	−0.71	−0.59	−0.50
1.8	0.9	−4.36	−2.06	−1.35	−1.01	−0.80	−0.66	−0.51
				100*MSE				
0.0	0.0	7.96	3.77	2.47	1.84	1.46	1.21	1.04
0.2	0.1	7.60	3.60	2.36	1.75	1.40	1.16	0.99
0.4	0.2	7.26	3.44	2.25	1.68	1.33	1.11	0.95
0.6	0.3	6.94	3.29	2.15	1.60	1.27	1.06	0.91
0.8	0.4	6.64	3.15	2.06	1.53	1.22	1.01	0.87
1.0	0.5	6.37	3.02	1.98	1.47	1.17	0.97	0.83
1.2	0.6	6.11	2.89	1.90	1.41	1.12	0.93	0.80
1.4	0.7	5.87	2.78	1.82	1.36	1.08	0.90	0.77
1.6	0.8	5.66	2.68	1.76	1.31	1.04	0.86	0.74
1.8	0.9	5.47	2.59	1.70	1.26	1.00	0.83	0.71

The other parameters were r=3, σ^2_{gx}=2, σ^2_{gy}=5, σ_{gxy}=2, σ^2_{ex}=1, σ^2_{ey}=4.

Table 9. Multivariate analysis of variance table for two groups of cultivars (A5141 and A81) and three characters

Source	d.f.	Character	Sum of squares & products matrix		
			yield	height	flowering days
Between 5141A hybrids	47	Yield	54017236		
		Height	509403	13983	
		Bloom time	29857	1169	348
Between 81A hybrids	47	Yield	51872256		
		Height	603221	23822	
		Bloom time	−448	535	219
Error	190	Yield	12493996		
		Height	363348	18791	
		Bloom time	−43490	−200	320

Sources under use only are retained here.

height (cm), and days to bloom on two groups of pearl millet hybrids in a trial conducted at ICRISAT (1983) (K. N. Rai, personal communication). The estimates of parameters were computed for each of the three character pairs at a time. The genotypic, phenotypic and environmental correlations are presented in Table 10. There are two different expressions, r_a [11] and r_b [6], for phenotypic correlation:

$$r_a = \sigma_{pxy}/\sigma_{px}\sigma_{py}$$

where $\sigma_{pxy} = \sigma_{exy} + r\,\sigma_{gxy}$, $\sigma^2_{pz} = \sigma^2_{ez} + r\,\sigma^2_{gz}$, $z = x, y$

and $$r_b = (\sigma_{exy} + \sigma_{gxy})/((\sigma^2_{ex} + \sigma^2_{gx})(\sigma^2_{ey} + \sigma^2_{gy}))^{\frac{1}{2}}.$$

Table 10. Phenotypic, environmental and genotypic correlations, for two groups of cultivars along with standard error

Characters	Groups	Correlations			
		phenotypic		enviornmental	genotypic ± SE
X : Y		(a)	(b)		
Y : H	5141A	0.59***	0.42	0.24	0.90±0.18***
	81A	0.54***	0.40	0.24	0.81±0.18***
Y : B	5141A	0.22	0.03	−0.22	0.52±0.25*
	81A	0.00	−0.11	−0.22	0.19±0.30
H : B	5141A	0.53***	0.31	−0.08	0.77±0.14***
	81A	0.24	0.12	−0.08	0.36±0.19

Y—Grain yield, H—height, and B—days to bloom.
*, **Significant at 5% and 0.1% level, respectively.

The expression r_b is correct while r_a is wrong, and still widely used [12, 14] probably because its significance can be tested easily using t-statistics, while that of r_b is quite complicated. Ahmad and Murthy [15] referred to a personal communication to A. V. R. Sastri for testing the significance of genotypic correlation but the procedure is not available. We do not recommend r_a to be used, as it may imply a conclusion

different than what we obtain from genotypic correlation. For instance, significance of genotypic correlation between yield and flowering time is detected for 5141A group at (P=0.05) although phenotypic correlation (a) is not significant.

ACKNOWLEDGEMENTS

The author is thankful to Dr. K. N. Rai for suggesting the problem, to Prof. J. L. Jinks and A. Robertson for helpful and encouraging suggestions, and to Mr. C. P. Jaiswal for secretarial assistance.

REFERENCES

1. E. C. R. Reeve. 1955. The variance of the genetic correlation coefficient. Biometrics, **11:** 357–374.
2. L. D. Van Vleck and C. R. Henderson. 1961. Empirical sampling estimates of genotypic correlations. Biometrics, **17:** 359–371.
3. G. H. Brown. 1969. An empirical study of the distribution of the sample genetic correlation coefficient. Biometrics, **25:** 63–72.
4. N. L. Johnson and S. Kotz. 1972. Distributions in Statistics: Continuous Multivariate Distributions. John Wiley and Sons, Inc., New York: 165–166.
5. J. L. Gill and E. L. Jensen. 1968. Probability of obtaining negative estimates of heritability. Biometrics, **24:** 517–526.
6. W. G. Hill and R. Thompson. 1978. Probabilities of nonpositive definite between group or genetic covariance matrices. Biometrics, **34:** 429–439.
7. M. G. Kendall and A. Stuart. 1969. The Advanced Theory of Statistics. Vol. 1. Charles Griffin and Company Ltd., London: 231–244.
8. C. R. Rao. 1973. Linear Statistical Inference and Its Applications, 2nd ed. Wiley Eastern Private Ltd., New Delhi: 127.
9. G. E. P. Box and M. E. Muller. 1958. A note on the generation of random normal deviates. Annal. Math. Stat., **29:** 610–611.
10. E. S. Pearson and H. O. Hartley. 1976. Biometrika Tables for Statisticians. Vol. I. Biometrika Trust: 67–68, 207–208.
11. H. A. Al-Jibouri, P. A. Miller and H. F. Robinson. 1958. Genotypic and environmental variance in an upland cotton cross of interspecific origin. Agron. J., **50:** 633–637.
12. R. W. Kaw and P. M. Menon. 1972. Association between yield and components in soybean. Indian J. Genet., **32:** 276–280.
13. S. S. Rohewal and M. N. Kappar. 1973. Association analysis in soybean. Indian J. Genet., **33:** 96–100.
14. S. B. S. Tikka. 1975. Interrelationship between yield and components in cluster bean. Indian J. Genet., **35(3):** 340–343.
15. Z. Ahmad and B. R. Murthy. 1972. Inerrelationship between quality components and developmental traits in pearl millet. Indian J. Genet., **32:** 319–324.

APPENDIX I

(i) Expression for $g'_i(\theta) = \partial g(x)/\partial x_i$ at $\underline{x} = \underline{\theta}$

$$\partial g/\partial T_{zz} = -\rho_G/(2rq\sigma^2_{gz})$$

$$\partial g/\partial E_{zz} = -\rho_G/(2r\nu\sigma^2_{gz}) \text{ for } z = x,y$$

$$\partial g/\partial T_{xy} = 1/(rq\,\sigma_{gx}\sigma_{gy})$$

$$\partial g/\partial E_{xy} = 1/(r\nu\sigma_{gx}\,\sigma_{gy})$$

(ii) Expression for $g''_{ij}(\underline{\theta}) = \partial^2 g(x)/\partial x_j\,\partial x_i$

$$\partial^2 g/\partial T^2_{zz} = 3\rho_G/(4q^2r^2\sigma^4_{gz})$$

$$\partial^2 g/\partial E^2_{zz} = 3\rho_G/(4\nu^2r^2\,\sigma^4_{ez}) \text{ for } z = x, y$$

$$\partial^2 g/\partial T_{xx}\partial T_{yy} = \rho_G/(4q^2r^2\,\sigma^2_{gx}\,\sigma^2_{gy})$$

$$\partial^2 g/\partial E_{xx}\partial E_{yy} = \rho_G/(4\nu^2r^2\,\sigma^2_{gx}\,\sigma^2_{gy})$$

$$\partial^2 g/\partial T_{xx}\partial T_{xy} = -1/(2q^2r^2\,\sigma^3_{gx}\sigma_{gy})$$

$$\partial^2 g/\partial T_{yy}\partial T_{xy} = -1/(2q^2r^2\,\sigma_{gx}\sigma^3_{gy})$$

$$\partial^2 g/\partial E_{xx}\partial E_{xy} = -1/(2\nu^2r^2\,\sigma^3_{ex}\sigma_{ey})$$

$$\partial^2 g/\partial E_{yy}\partial E_{xy} = -1/(2\nu^2r^2\,\sigma_{ex}\sigma^3_{gy})$$

$$\partial^2 g/\partial T^2_{xy} = \partial^2 g/\partial E^2_{xy} = 0$$

The cross derivatives between T and E variables are not given as they will have zero covariances as product.