

A General Linear Model for the Genotypic Covariance Between Relatives Under Assortative Mating*

A. Gimelfarb**

Department of Statistics, North Carolina State University, Raleigh, NC 27650, USA

Abstract. A linear model for the genotypic covariance between relatives under assortative mating comprising the “classical linear model” and the model of “selective assortative mating” is proposed. The general conditions on the genetical and developmental mechanisms of quantitative characters, as well as on selection and the mating system, on which the model is based, are explicitly stated and discussed. A classification of different relationships is presented and it is shown that these conditions are sufficient to obtain the genotypic covariance between relatives only if the relationship is a combination of descendant-ancestor, full sib, Type 1 and N th uncle-niece relationships. All the “traditional” relationships, i.e., those for which the covariances of the relatives have been obtained in the literature, fall into this category. These conditions also ensure that the regression of the individual’s genotypic value on the genotypic value or phenotype of any of its ancestors is always linear.

Key words: Assortative mating – Covariance between relatives – Linear model

The problem of covariances between relatives is one of the oldest problems in population genetics. Particularly interesting was and still remains the question of how these covariances caused by hereditary mechanisms of quantitative characters are affected by assortative mating based on a quantitative character. Starting from the work by R. A. Fisher (Fisher, 1918; see also Moran and Smith, 1966), this question has been investigated by many authors within the framework of what now is referred to as the “classical linear model”. Some of the authors employed correlation analysis (Wright, 1921a and b; Crow and Felsenstein, 1968; Feldman and Cavalli-Sforza, 1977, 1979; Goldberger, 1978; Nagylaki, 1978), others made use of the path analysis (Morton, 1974; Rao and Morton, 1978; Rice et al., 1978; Cloninger et al., 1979).

* Paper No. 6619 of the Journal Series of the North Carolina Agricultural Research Service, Raleigh, North Carolina. This investigation was supported in part by NIH Research Grant No. GM 11546 from the National Institute of General Medical Sciences

** Present address: Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA

The classical linear model is based on a number of assumptions, sometimes explicitly stated (Nagylaki, 1978; Cloninger et al., 1979), more often implied and, according to a remark by Karlin (1979c), “the discussion of their relevance and limitations is hardly ever done”. One of the most controversial assumption of the classical model is that the correlation between mating individuals (marital correlation) remains the same in any generation of assortative mating.

A different model of “selective assortative mating” appeared in works by Wilson (1973) and Wagener (1976), and has been rigorously investigated by Karlin (1979a, b, c). Changes of the population structure in this model are governed by the “selection mating function” which actually is a complete analog of fertility selection operating on parental pairs in a randomly mating population. The marital correlation arises as a result of such selection and, therefore, does not remain constant but changes along with the changes in population structure. In spite of this attractive feature, the model has an important disadvantage. It is applicable to only those populations where the assortment is weak. Indeed, any assortment of individuals into mating pairs is accompanied in this model by selection. The stronger assortment, the stronger must be the selection and the higher is the “cost” the population has to bear. It can be shown that the average fitness, W_k of a population undergoing selective assortative mating in any generation k satisfies the following inequality:

$$W_k \leq \left(\frac{1 - |\rho_k|}{1 + |\rho_k|} \right)^{1/2},$$

where ρ_k is the marital correlation in generation k . According to reports on the marital correlation of I.Q. in human populations (Vanderberg, 1972), its value is close to 0.5. The corresponding average fitness following from the selective assortative mating model must be less than 0.58, i.e., the “cost” of such assortment must be at least as high as 42%. Even for the assortment of spouses by their heights, in which case the marital correlation is only 0.3 (Vanderberg, 1972), the corresponding “cost” must be at least 27%. It is highly improbable that a human population bears such cost for assortative mating.

It is most likely that the reality is better reflected not by any one of the models but by a combination of them. An attempt is made in this paper to introduce a general linear model encompassing both the “classical linear” and the “selective assortative” models. The main reason for the introduction of the model, as well as the purpose of this paper, was not as much in deriving expressions for covariances between relatives or in providing a method for their derivation, since almost all of these expressions have already been obtained in the literature by different and sometimes more elegant methods, but rather in finding and analyzing the conditions for applicability of these expressions and methods. For this reason, in order to make the analysis simpler, only genotypic covariances are considered. An inclusion into consideration of phenotypic covariances does not affect results of the analysis, and phenotypic covariances between relatives can be easily computed if the genotypic covariances are known.

The model is based on a set of general conditions imposed on hereditary and developmental mechanisms of quantitative characters as well as on selection and the mating system which are explicitly stated and discussed. A classification of

different relationships into three types is made and it is shown that for all "traditional" relationships, i.e., for those that are mainly discussed in the literature, this set of conditions makes it possible to express the covariances between relatives in terms of variances and marital correlations.

It is assumed that in every generation the population goes through two distinct phases. The first phase represents the juvenile individuals, and the second phase is the subpopulation of the reproducing individuals assorted into mating pairs. The transition from the juvenile phase to the reproducing is due to selection and mating assortment. The transition from the reproducing phase of one generation to the juvenile phase of the next generation is due to the hereditary and developmental mechanisms.

A quantitative character of an individual, as usual, is assumed to be a sum of two components: a hereditary component, commonly called the genotypic value of the character, and a nonhereditary or environmental component. An individual will be described by a pair (x, X) , where x stands for the genotypic value and X for the individual's phenotype (quantitative character). Capital letters will be used for the phenotypes whereas small letters are reserved for the genotypic values.

It is assumed for convenience that the quantitative character in question is measured on individuals while they are in the juvenile phase. The same general model is applied to other cases as well, although the resulting formulas may be slightly different.

The following is a list of notations for distributions important in further discussion.

$P_k^0(X)$	the distribution of the character among the juvenile individuals in generation k . Its variance is V_k^0 .
$p_k^0(x)$	the same as the previous, but for the genotypic values. Its variance is v_k^0 .
$P_k(X)$	the distribution of the character among the reproducing individuals in generation k . Its variance is V_k .
$p_k(x)$	the same as the previous, but for the genotypic values. Its variance is v_k .
$P_k(X_i X_j, (i * j))$	the distribution of the character among the mates of the individuals with the character X_j . ($(i * j)$ denotes the condition: individual i mates with individual j).
$p_k(x_i x_j, (i * j))$	the same as the previous, but for the genotypic values.
$P_k(X_i, X_j (i * j))$	the joint distribution of the characters in mating pairs in generation k .
$p_k(x_i, x_j (i * j))$	the same as the previous, but for the genotypic values.

Obviously,

$$P_k(X_i, X_j|(i * j)) = P_k(X_j)P_k(X_i|X_j, (i * j)),$$

$$p_k(x_i, x_j|(i * j)) = p_k(x_j)p_k(x_i|x_j, (i * j)).$$

All other distributions appearing in this paper are denoted as $\text{Pr}[\]$, and it should not be difficult to figure out their meaning in each case. Summation symbols suggesting discrete distributions are used only for convenience; all results remain

without changes for continuous distributions in which case the summations may be simply replaced with integrations. It will also be assumed for convenience and without loss of generality that the mean value of the character among juvenile as well as reproducing individuals is zero in any generation.

Let us now formulate the conditions on hereditary and developmental mechanisms, the mating system and selection which will be assumed throughout the paper.

Condition A. The mean genotypic value among the offspring of a mating pair is determined completely by the parental genotypic values as

$$E(z|g_i, g_j) = E(z|x_i, x_j) = \frac{1}{2}(x_i + x_j), \quad (1)$$

where g_i and g_j are parental genotypes. The actually important condition for the model presented in this paper is the bilinearity of the offspring-on-parent regression:

$$E(z|g_i, g_j) = \alpha x_i + \beta x_j.$$

However, the formulae are very cumbersome when coefficients α and β are arbitrary, although they look much simpler in the case of $\alpha = \beta = \frac{1}{2}$. Assumption (1) is satisfied, of course, for all additive characters with Mendelian segregation of the genes. Also, at least as a good approximation, it may hold for nonadditive characters as well, if dominant and epistatic effects in different loci act to counterbalance each other (J. F. Crow, personal communication).

Condition B. For juvenile individuals, the regression of the individual's phenotype on the genotypic value and of its genotypic value on the phenotype are linear as

$$E(X|x) = x, \quad (2a)$$

$$E(x|X) = b_k^0 X, \quad (2b)$$

where b_k^0 is, of course, the heritability h_k^2 of the character among the juvenile individuals in generation k :

$$b_k^0 = h_k^2 = \frac{v_k^0}{V_k^0}. \quad (3)$$

These assumptions are customary in quantitative genetics.

Condition C. For reproducing individuals, the regression of the individual's phenotype on the genotypic value and of its genotypic value on the phenotype are linear:

$$E(X|x) = a_k x, \quad (4a)$$

$$E(x|X) = b_k X. \quad (4b)$$

It can be shown (see Appendix) that in this case

$$\frac{v_k}{V_k} = \frac{b_k}{a_k}. \quad (4c)$$

Since it has been assumed that the quantitative character in question is measured

always on individuals in the juvenile phase, both (4a) and (4b) follow from (2a) and (2b) with $a_k = 1$ and $b_k = b_k^0 = h_k^2$ in the absence of selection. If selection is present, then (4b) still follows from (2b) for any form of phenotypic selection. Assumption (4a), on the other hand, does not follow necessarily from (2a) even for a phenotypic selection, and, therefore, this assumption imposes restrictions on the types of selection admissible in the frame of the linear model presented in this paper. It holds, for example, in the case of selection implied by the selective assortative mating in the models of Wilson (1973), Wagener (1976) and Karlin (1979c) with the selection mating function

$$\exp \left\{ -\frac{1}{2c} (X_i^2 - 2rX_iX_j + X_j^2) \right\}, \quad (4d)$$

in which case the regression coefficient a_k in (4a) is

$$a_k = \frac{S_k}{(1 - h_k^2)V_k + S_k}$$

with

$$S_k = \frac{c(c + V_k)}{c + (1 - r^2)V_k}.$$

Condition D. The environmental component of a juvenile individual is independent of the parental genotypic values:

$$\Pr[Z|z, x_i, x_j] = \Pr[Z|z]. \quad (5)$$

This is another customary assumption of quantitative genetics.

Condition E. For any two reproducing individuals i and j ,

$$\Pr[(i*j)|X_i, X_j, \Omega] = \Pr[(i*j)|X_i, X_j], \quad (6)$$

where Ω stands for any set of conditions. This assumption asserts that the mating in the population is phenotypic assortative, i.e., based exclusively on the phenotypic resemblance of the individuals. It also implies what Karlin (1979c) called the conditional independence assumption, according to which the choice by an individual of its mating partner is not affected by a previous mating experience either of its own or of any of its relatives.

Condition F. The regression of the phenotype of a mating individual on the phenotype of its mate is linear:

$$E(X_i|X_j, (i*j)) = \rho_k X_j, \quad (7)$$

where ρ_k is the “marital correlation” or the correlation coefficient between the characters of mating individuals in generation k . This assumption has been widely used in practically all existing models of phenotypic assortative mating. In classic linear models the marital correlation was assumed to remain the same in any generation, whereas in the selective assortative mating models it changes from generation to generation. In the case of the selection function in the form (4d), for

example, the marital correlation can be computed in any generation as

$$\rho_k = \frac{rV_k}{c + V_k}.$$

Corollaries. *The following corollaries can be inferred (see Appendix) from the above Conditions A–F.*

$$\Pr[x_i|X_i, X_j, x_j, (i * j)] = \Pr[x_i|X_i], \tag{8}$$

$$E(x_i|X_j, (i * j)) = b_k \rho_k X_j, \tag{9a}$$

$$E(X_i|x_j, (i * j)) = a_k \rho_k x_j, \tag{9b}$$

$$E(x_i|x_j, (i * j)) = a_k b_k \rho_k x_j. \tag{9c}$$

Notice, that Conditions A–F are not very restrictive. A model based on these conditions can be applied to a great variety of quantitative characters in populations with quite different hereditary and developmental mechanisms and assorting processes. An exception is assumption (4a) which calls for caution when the model is applied to populations with selection. Linearity of the regressions for genotypic values of any relatives is not assumed. Moreover, as will be shown, the linearity of the descendant-on-ancestor regression follows from Conditions A–F, whereas the question of hereditary mechanisms that provide a linearity of regressions for other relatives remains open.

We now proceed to investigate different relationships.

Descendant – Ancestor

First of all let us demonstrate that under Conditions A–F the offspring-on-parent regression $E(z|x_1)$ (Fig. 1) is linear. According to Condition A,

$$E(z|x_1) = \sum_z z \Pr[z|x_1] = \sum_{y_1} E(z|x_1, y_1) P_n(y_1|x_1, (x_1 * y_1)).$$

From this expression, by making use of (1) and (9c), we obtain

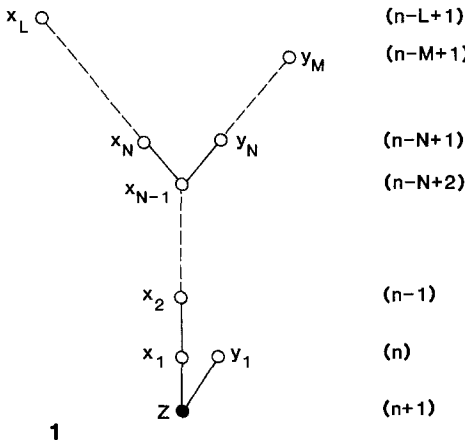


Fig. 1. Descendant-ancestor relationship (generation numbers are indicated in parentheses)

$$E(z|x_1) = \beta_1 x_1, \quad (10)$$

where

$$\beta_1 = \frac{1}{2}(1 + a_n b_n \rho_n). \quad (11)$$

Let us now have a look at the regression $E(z|x_N)$ of the individual's genotypic value on the genotypic value of its N th ancestor (Fig. 1).

$$E(z|x_N) = \sum_z z \Pr[z|x_N] = \sum_{x_1} \sum_z z \Pr[z|x_1, x_N] \Pr[x_1|x_N]. \quad (12)$$

Under the conditional independence assumption (Condition E), given that the parental genotypic value is known, the knowledge of any other ancestor's genotypic value does not provide additional information about the distribution of the genotypic values among the offspring, i.e.,

$$\Pr[z|x_1, x_N] = \Pr[z|x_1].$$

Substitution of this into (12) and reference to (10) yields for the one-step "unfolded" pedigree

$$E(z|x_N) = \beta_1 \sum_{x_1} x_1 \Pr[x_1|x_N].$$

Continuing to "unfold" the pedigree to N th ancestor, we shall end up with the following expression:

$$E(z|x_N) = \beta_N x_N, \quad (13)$$

where

$$\beta_N = \prod_{i=1}^N \beta_1^{(i)}. \quad (14)$$

Coefficients $\beta_1^{(i)}$ are the coefficients of $(i-1)$ th offspring on i th parent regression from (11):

$$\beta_1^{(i)} = \frac{1}{2}(1 + a_{n-i+1} b_{n-i+1} \rho_{n-i+1}). \quad (15)$$

Thus, the regression of the individual's genotypic value on the genotypic value of any of its ancestors is linear under Conditions A–F for any assortative mating. This has also been noticed by T. Nagylaki (personal communication).

Let us define an "ancestral pair" of an individual as a pair of the ancestors whose mating in some previous generation originates a path in the individual's pedigree, and let us have a look at the regression $E(z|x_N, y_N)$ of the individual's genotypic value on the genotypic values of its N th ancestral pair (Fig. 1):

$$\begin{aligned} E(z|x_N, y_N) &= \sum_{x_{N-1}} E(z|x_{N-1}) \Pr[x_{N-1}|x_N, y_N] \\ &= \beta_{N-1} \sum_{x_{N-1}} x_{N-1} \Pr[x_{N-1}|x_N, y_N] \\ &= \frac{1}{2} \beta_{N-1} (x_N + y_N). \end{aligned} \quad (16)$$

(It has been assumed here that $\beta_0 = 1$, and for $N > 1$ coefficient β_{N-1} is the regression coefficient determined by (14).)

Hence, this regression is always bilinear under Conditions A–F. As far as the regression of the individual's genotypic value on the genotypic values of an arbitrary pair of its ancestors is concerned, the bilinearity cannot be ensured any more. Indeed, for $E(z|x_L, y_M)$ we have (Fig. 1):

$$\begin{aligned} E(z|x_L, y_M) &= \sum_{x_N} \sum_{y_N} E(z|x_N, y_N) \Pr[x_N, y_N|x_L, y_M] \\ &= \frac{1}{2}\beta_{N-1}(E(x_N|x_L, y_M) + E(y_N|x_L, y_M)). \end{aligned}$$

If mating in generation $n - N$ is nonrandom, then the expected values of x_N and y_N are functions of both x_L and y_M . Conditions A–F are not sufficient to assure linearity of these functions and they may be nonlinear. If, however, the mating in this generation is random, then

$$\begin{aligned} E(x_N|x_L, y_M) &= E(x_N|x_L) = \beta_{L-N}x_L, \\ E(y_N|x_L, y_M) &= E(y_N|y_M) = \beta_{M-N}y_M, \end{aligned}$$

and (16) is reduced to

$$E(z|x_L, y_M) = \frac{1}{2}\beta_{N-1}(\beta_{L-N}x_L + \beta_{M-N}y_M). \quad (17)$$

Finally, the following may be concluded. Conditions A–F ensure that the regression of the individual's genotypic value on the genotypic value of any of its ancestors is always linear. These conditions are also sufficient for the regression of the individual's genotypic value on the genotypic values of any of its ancestral pairs to be bilinear. They are not sufficient, however, for the regression on the genotypic values of two arbitrary ancestors to be bilinear, except for the case of random mating when this regression is also bilinear.

After having obtained the expression for the descendant-on-ancestor regression, it is not difficult to compute the covariance. For the genotypic value covariance of offspring in the juvenile phase and of an ancestor in the reproducing phase, it follows from (13) that

$$\begin{aligned} \text{cov}_g^{(D-A)} &= \sum_{x_N} x_N E(z|x_N) p_{n-N+1}(x_N) \\ &= \beta_N v_{n-N+1}, \end{aligned} \quad (18)$$

where v_{n-N+1} is the genotypic value variance in generation $n - N + 1$, and we may write referring to (14) and (11) the following expression for the descendant-ancestor genotypic covariance:

$$\text{cov}_g^{(D-A)} = \left(\frac{1}{2}\right)^N \prod_{i=1}^N (1 + a_{n-i+1} b_{n-i+1} \rho_{n-i+1}) v_{n-N+1}. \quad (19)$$

The substitution into this of $N = 1$ yields the formula for the offspring-parent genotypic covariance:

$$\text{cov}_g^{(\text{OFFSPR-PARENT})} = \frac{1}{2} (1 + a_n b_n \rho_n) \frac{b_n}{a_n} V_n. \quad (20)$$

For $N = 2$ from (19) follows the formula for the offspring-grandparent genotypic covariance:

$$\text{COV}_g^{(\text{OFFSPR-GR. PARENT})} = \frac{1}{4}(1 + a_n b_n \rho_n)(1 + a_{n-1} b_{n-1} \rho_{n-1}) \frac{b_{n-1}}{a_{n-1}} V_{n-1}. \quad (21)$$

Step Offspring – Step Parent (Fig. 2)

In a population with random mating, correlations of the genotypic values occur only for “blood” relatives because of their connections through common ancestors. Under assortative mating, however, correlations of genotypic values may also occur for “nonblood” relatives due to the “step” connections through intermatings within their pedigree. Step offspring – step parent is an example of relatives with no “blood” connections, although they are connected due to the mating between the natural and step parents. The regression $E(z|x_3)$ of the step offspring genotypic value on that of its step parent can be determined as follows.

$$\begin{aligned} E(z|x_3) &= \sum_z z \Pr[z|x_3, (2 * 3)] \\ &= \sum_{x_1} \sum_{x_2} E(z|x_1, x_2) \Pr[x_1, x_2|x_3, (1 * 2), (2 * 3)] \\ &= \frac{1}{2} \left(\sum_{x_1} x_1 \Pr[x_1|x_3, (1 * 2), (2 * 3)] + E(x_2|x_3, (2 * 3)) \right). \end{aligned} \quad (22)$$

Taking into consideration Condition E, the first term in the parentheses may be rewritten as

$$\begin{aligned} E(x_1|x_3, (1 * 2), (2 * 3)) &= \sum_{x_2} \sum_{x_1} x_1 \Pr[x_1|X_2, (1 * 2)] \Pr[X_2|x_3, (2 * 3)] \\ &= b_n \rho_n E(X_2|x_3, (2 * 3)). \end{aligned} \quad (23)$$

By substituting (23) into (22) and referring to (9b) and (9c), we obtain

$$E(z|x_3) = \beta_1^s x_3, \quad (24)$$

where

$$\beta_1^s = \frac{1}{2}(1 + \rho_n) a_n b_n \rho_n. \quad (25)$$

Thus, the regression of the step offspring’s genotypic value on the genotypic value of its step parent is always linear under Conditions A – F for any assortative mating. It is also not difficult to demonstrate by following the same steps as in the previous section that the regression of the individual’s genotypic value on the genotypic value of any of its step ancestors is always linear under Conditions A – F. Notice, however, that although the regression $E(z|x_1, x_2)$ of the offspring’s genotypic value on the genotypic values of both of its natural parents has been assumed to be bilinear (Condition A), the regressions $E(z|x_1, x_3)$ and $E(z|x_2, x_3)$ are not necessarily bilinear.

Given the expression for the regression $E(z|x_3)$, the expression for the step offspring-step parent genotypic covariance is straightforward:

$$\text{cov}_g^{(\text{S.O.-S.P.})} = \frac{1}{2}(1 + \rho_n) b_n^2 \rho_n V_n. \quad (26)$$

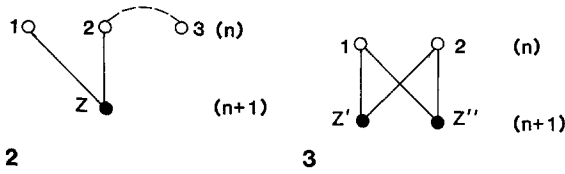


Fig. 2. Step parent-offspring
Fig. 3. Full sibs

Full Sibs (Fig. 3)

According to Condition A, we have for the genotypic covariance of full sibs:

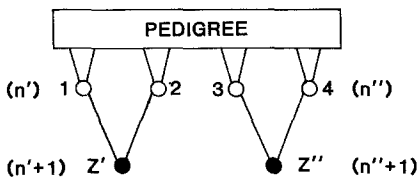
$$\begin{aligned}
 cov_g^{(SIB)} &= \sum_{z'} \sum_{z''} z' z'' Pr[z', z'' | SIB] \\
 &= \sum_{g_1} \sum_{g_2} E(z' | g_1, g_2) E(z'' | g_1, g_2) Pr[g_1, g_2 | (1 * 2)] \\
 &= \sum_{x_1} \sum_{x_2} E(z' | x_1, x_2) E(z'' | x_1, x_2) p_n(x_1, x_2 | (1 * 2)) \\
 &= \frac{1}{4} \sum_{x_1} \sum_{x_2} (x_1 + x_2)^2 p_n(x_1, x_2 | (1 * 2)).
 \end{aligned}
 \tag{27}$$

Reference to (4c) and (9c) yields

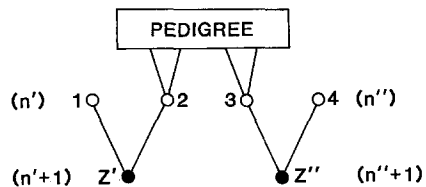
$$cov_g^{(SIB)} = \frac{1}{2} (1 + a_n b_n \rho_n) \frac{b_n}{a_n} V_n.
 \tag{28}$$

Relationships other than descendant-ancestor or full sibs may be depicted as in Fig. 4, where Z' stands for the phenotype of one of the relatives and Z'' for the phenotype of the other. The relatives need not belong to the same generation, and $n' + 1$ denotes the generation of the first of them and $n'' + 1$ of the second. Individuals 1 and 2 represent the parental pair of the first relative and individuals 3 and 4 are the parents of the second. Pedigree refers to all connections either “blood” or “step” existing between individuals 1, 2, 3, 4 except for the connections due to the matings between 1 and 2 and between 3 and 4. Thus, for example, in the case of step sibs (Fig. 8) the connection between individuals 2 and 3 due to their mating is regarded as belonging to the Pedigree, whereas connections 1–2 and 3–4 do not belong to the Pedigree.

Three types of relatives may be distinguished depending on the connections through the Pedigree.



4
Fig. 4. Relationship of a general type



5
Fig. 5. Type 1 relationship

Type 1 relatives are those that are connected through only one of their parents, as shown in Fig. 5. Half sibs, step sibs and first cousins are examples of this type.

Relatives are of Type 2 if both parents of one of them are connected to only one of the parents of the other, as shown in Fig. 6. An example of this type is the uncle-niece relationship.

Type 3 relatives are those connected through both of their parents. This type is well represented by Fig. 4, and double first cousins provide the best familiar example.

This classification of relatives into three types is similar to the classification by Cotterman (1960) into unilineal and bilineal relatives. Noninbred, “blood” relatives have been termed bilineal if they are connected by two or more completely separate (link-disjoint) pedigree paths. Otherwise they are unilineal. Certainly, unilineal relatives are always of either Type 1 or Type 2, and bilineal are always of Type 3. Notice, however, that the Cotterman’s classification refers to only noninbred, “blood” relatives, whereas for our classification it is irrelevant whether the relatives are inbred or not and whether they are “blood” or “step” relatives.

Let us now bring into consideration the following covariances: the covariance of the genotypic values of individuals i and j ,

$$a_{ij} = \sum_{x_i} \sum_{x_j} x_i x_j \Pr[x_i, x_j | \text{PED}, (1 * 2), (3 * 4)], \tag{29}$$

and the partial covariances between individuals i and j due strictly to their connections through the Pedigree,

$$c_{ij} = \sum_{x_i} \sum_{x_j} x_i x_j \Pr[x_i, x_j | \text{PED}], \tag{30a}$$

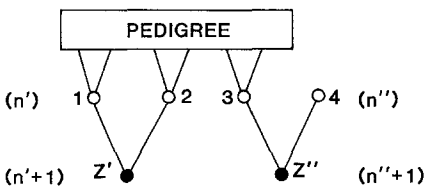
$$c_{i(j)} = \sum_{x_i} \sum_{X_j} x_i X_j \Pr[x_i, X_j | \text{PED}], \tag{30b}$$

$$c_{(i)j} = \sum_{X_i} \sum_{X_j} X_i X_j \Pr[X_i, X_j | \text{PED}]. \tag{30c}$$

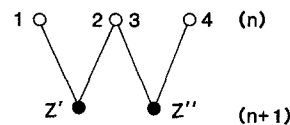
Notice that the above covariances are defined for the individuals in the reproducing phase. In a population with selection they may differ from the covariances for the same individuals when they are in the juvenile phase.

The genotypic covariance of relatives, $\text{cov}_g^{(\text{REL})}$ is the covariance of the joint distribution $\Pr[z', z'' | \text{REL}]$, and the following expression is straightforward:

$$\text{cov}_g^{(\text{REL})} = \sum_{\{x\}} E(z' | x_1, x_2) E(z'' | x_3, x_4) \Pr[x_1, x_2, x_3, x_4 | (1 * 2), (3 * 4), \text{PED}], \tag{31}$$



6
Fig. 6. Type 2 relationship



7
Fig. 7. Half sibs

($\{x\}$ denotes the set $\{x_1, x_2, x_3, x_4\}$). According to Condition A, from (31) follows

$$\text{cov}_g^{(\text{REL})} = \frac{1}{4}(a_{13} + a_{23} + a_{14} + a_{24}). \quad (32)$$

The problem now is to express the covariances a_{ij} in (32) through the covariances c_{ij} , $c_{i(j)}$ or $c_{(ij)}$. It is clear that Conditions A – F are not sufficient to enable us to do it for any arbitrary relationship. Let us therefore proceed to investigate relatives of different types.

Type 1 (Fig. 5)

Using (9a), covariances a_{ij} for relatives of this type are obtained as follows:

$$a_{23} = c_{23}, \quad (33a)$$

$$\begin{aligned} a_{13} &= \sum_{X_2} \sum_{x_1} \sum_{x_3} x_1 x_3 \Pr[x_1|X_2, (1 * 2)] \Pr[X_2, x_3|\text{PED}] \\ &= b_{n'} \rho_{n'} c_{(2)3}, \end{aligned} \quad (33b)$$

$$\begin{aligned} a_{24} &= \sum_{X_3} \sum_{x_2} \sum_{x_4} x_2 x_4 \Pr[x_4|X_3, (3 * 4)] \Pr[x_2, X_3|\text{PED}] \\ &= b_{n''} \rho_{n''} c_{2(3)}, \end{aligned} \quad (33c)$$

$$\begin{aligned} a_{14} &= \sum_{X_2} \sum_{X_3} \sum_{x_1} \sum_{x_4} x_1 x_4 \Pr[x_1|X_2, (1 * 2)] \Pr[x_4|X_3, (3 * 4)] \Pr[X_2, X_3|\text{PED}] \\ &= b_{n'} \rho_{n'} b_{n''} \rho_{n''} c_{(2)3}. \end{aligned} \quad (33d)$$

Substituting these expressions into (32), we shall obtain an expression for the genotypic covariance of different Type 1 relatives in terms of the partial covariances of individuals 2 and 3 due to the Pedigree.

Half Sibs (Fig. 7)

In this case, since 2 and 3 actually represent the same individual,

$$\begin{aligned} n' &= n'' = n, \\ c_{(23)} &= V_n, \\ c_{23} &= v_n = \frac{b_n}{a_n} V_n, \\ c_{2(3)} &= c_{(2)3} = b_n V_n, \end{aligned}$$

and for the genotypic covariance of half sibs we have, using (33) and (32),

$$\text{cov}_g^{(\text{H.SIB})} = \frac{1}{4} \left(\frac{1}{a_n} + 2b_n \rho_n + b_n \rho_n^2 \right) b_n V_n. \quad (34)$$

Step Sibs (Fig. 8)

In the case of step sibs,

$$n' = n'' = n,$$

$$\begin{aligned}
 c_{(23)} &= \rho_n V_n, \\
 c_{23} &= b_n^2 c_{(23)} = b_n^2 \rho_n V_n, \\
 c_{2(3)} &= c_{(2)3} = b_n c_{(23)} = b_n \rho_n V_n.
 \end{aligned}$$

The genotypic covariance of step sibs is therefore

$$\text{cov}_g^{(S.SIB)} = \frac{1}{4}(1 + \rho_n)^2 b_n^2 \rho_n V_n. \tag{35}$$

It is easy to see that for all Type 1 relationships, other than half sibs or step sibs, in which individuals 2 and 3 do not have a descendant-ancestor relationship,

$$\begin{aligned}
 c_{2(3)} &= a_{n'} c_{23}, \\
 c_{(2)3} &= a_{n'} c_{23}, \\
 c_{(23)} &= a_{n'} a_{n''} c_{23} + \text{cov}_e^{(2,3)},
 \end{aligned}$$

where $\text{cov}_e^{(2,3)}$ is the environmental covariance of individuals 2 and 3. Therefore, for the genotypic covariance of the relatives we have

$$\begin{aligned}
 \text{cov}_g^{(Type\ 1)} &= \frac{1}{4}(1 + a_{n'} b_{n'} \rho_{n'}) (1 + a_{n''} b_{n''} \rho_{n''}) c_{23} \\
 &\quad + \frac{1}{4} b_{n'} \rho_{n'} b_{n''} \rho_{n''} \text{cov}_e^{(2,3)}.
 \end{aligned} \tag{36}$$

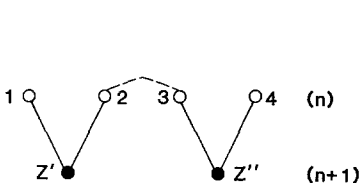
It is important to remember that c_{23} and $\text{cov}_e^{(2,3)}$ are defined as the covariances between individuals 2 and 3 in the reproducing phase. When there is no selection in the population, they, of course, are the same as the covariances between these individuals in the juvenile phase. When, however, there is selection, they may differ. If (as a result of previous computations, for example) the covariances are determined in the juvenile phase, they need to be transformed into the covariances in the reproducing phase. There is not a unique transformation for any arbitrary selection. The transformation will depend, of course, on the form of the distributions and the form of selection. Feldman and Cavalli-Sforza (1977) have computed c_{23} for some relationships in the case of normalizing selection.

As an example, let us apply formula (36) to the first cousin relationship (Fig. 9).

First Cousins (Fig. 9)

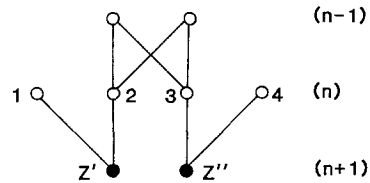
For this relationship,

$$n' = n'' = n,$$



8

Fig. 8. Step sibs



9

Fig. 9. First cousins

$$c_{23} = \text{cov}_g^{(\text{SIB})} = \frac{1}{2}(1 + a_{n-1}b_{n-1}\rho_{n-1})\frac{b_{n-1}}{a_{n-1}}V_{n-1},$$

$$\text{cov}_e^{(2,3)} = \text{cov}_e^{(\text{SIB})}. \tag{37}$$

From (36) then follows

$$\text{cov}_g^{(\text{F, COUSIN})} = \frac{1}{8}(1 + a_n b_n \rho_n)^2(1 + a_{n-1}b_{n-1}\rho_{n-1})\frac{b_{n-1}}{a_{n-1}}V_{n-1} + \frac{1}{4}b_n^2\rho_n^2\text{cov}_e^{(\text{SIB})}. \tag{38}$$

Let us now turn to Type 2 relationships.

Type 2 (Fig. 6)

Not too much can be made out of expression (32) for the covariance of relatives of this type, unless some additional assumptions are made. Let us assume that only “blood” connections exist between individuals 1 and 3 as well as between 2 and 3. In this case

$$a_{14} = \sum_{x_1} \sum_{x_4} x_1 x_4 \Pr[x_1, x_4 | \text{PED}, (1 * 2), (3 * 4)]$$

$$= \sum_{x_3} \sum_{x_1} \sum_{x_4} x_1 x_4 \Pr[x_1, x_3 | \text{PED}, (1 * 2)] \Pr[x_4 | x_3, (3 * 4)]$$

$$= a_{n'} \cdot b_{n'} \cdot \rho_{n'} \cdot a_{13}. \tag{39a}$$

Analogously,

$$a_{24} = a_{n'} \cdot b_{n'} \cdot \rho_{n'} \cdot a_{23}. \tag{39b}$$

Therefore, in this case of Type 2 relatives we have

$$\text{cov}_g^{(\text{Type 2})} = \frac{1}{4}(a_{13} + a_{23})(1 + a_{n'} \cdot b_{n'} \cdot \rho_{n'}). \tag{40}$$

For $(a_{13} + a_{23})$ we may write the following expression

$$a_{13} + a_{23} = \sum_{x_1} \sum_{x_2} (x_1 + x_2) \sum_{x_3} x_3 \Pr[x_3 | x_1, x_2, \text{PED}] \Pr[x_1, x_2 | \text{PED}, (1 * 2)]. \tag{41}$$

Now everything depends on the regression of x_3 on x_1 and x_2 due to their connections through the Pedigree. If this regression is nonlinear, then again not too much can be concluded about the covariance of the relatives, except that it will depend on higher than the second moments of the distribution $p_{n'}(x_1, x_2 | (1 * 2))$, i.e., the covariance of relatives cannot be obtained in terms of only variances and marital correlations. If, however the regression of x_3 on x_1 and x_2 is bilinear:

$$\sum_{x_3} x_3 \Pr[x_3 | x_1, x_2, \text{PED}] = \alpha_{31}x_1 + \alpha_{32}x_2, \tag{42}$$

then the expression for $\text{cov}_g^{(\text{Type 2})}$ is readily derived:

$$\text{cov}_g^{(\text{Type 2})} = \frac{1}{4}(1 + a_{n'} \cdot b_{n'} \cdot \rho_{n'})(1 + a_{n'} \cdot b_{n'} \cdot \rho_{n'})(\alpha_{31} + \alpha_{32})v_{n'}. \tag{43}$$

There is only one kind of a relationship for which Conditions A – F ensure that the

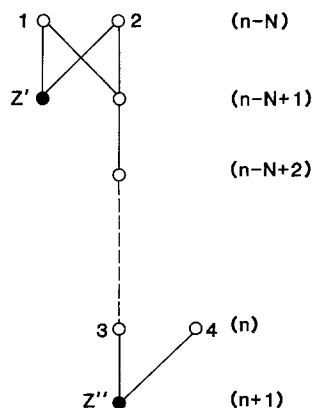


Fig. 10. *N*-th uncle-niece

regression (42) is bilinear. This is when individuals 1 and 2 form an *N*th ancestral pair of individual 3. The relatives in this case are an individual and a descendant of its sib. The relationship (Fig. 10) may be named “*N*th uncle-niece”. This relationship has been considered and a formula for the correlation between the relatives has been given in Cloninger et al. (1979). It follows from (15) that in this case

$$\alpha_{31} = \alpha_{32} = \beta_{N-1},$$

where β_{N-1} is the descendant-on-ancestor regression coefficient (14). The genotypic covariance of the relatives is expressed then as

$$\text{cov}_g^{(N\text{th UNCLE-NIECE})} = \frac{1}{4}(1 + a_n b_n \rho_n)(1 + a_{n-N} b_{n-N} \rho_{n-N}) \beta_{N-1} v_{n-N}. \quad (44)$$

An expression for the uncle-niece genotypic covariance follows from (44) when $N = 1$:

$$\text{cov}_g^{(\text{UNCLE-NIECE})} = \frac{1}{4}(1 + a_{n-1} b_{n-1} \rho_{n-1})(1 + a_n b_n \rho_n) \frac{b_{n-1}}{a_{n-1}} V_{n-1}. \quad (45)$$

For Type 2 relatives other than *N*th uncle-niece, Conditions A – F do not ensure bilinearity of the regression of x_3 on x_1 and x_2 and, therefore, they are not sufficient to compute the covariance to relatives. It is obvious, that these conditions are also not sufficient to compute the covariance of Type 3 relatives.

Conclusion

It may be concluded that for any relationship between two individuals which is a combination of descendant-ancestor, full sib, Type 1 and *N*th uncle-niece relationships, Conditions A – F are sufficient to obtain the genotypic covariance of the relatives in terms of marital correlations and variances. (Notice that all the “traditional” relationships, i.e., those for which the covariances of the relatives have been obtained in the literature, fall into this category). These conditions also ensure that the regression of the individual’s genotypic value on the genotypic value of any of its ancestors (either “blood” or “step”) is always linear. This is in

accordance with a result of Bulmer (1976), who has demonstrated that in the case of additive polygenic characters controlled by a large number of loci, the joint distribution of the parent-offspring genotypic values is a bivariate normal. Linearity of the individual's genotypic value regression on the genotypic value of any other of its relatives is not ensured by Conditions A – F. However, when the covariance between relatives can be obtained under these Conditions, it is irrelevant whether this regression is linear or nonlinear. For a relationship which is not a combination of those listed above, Conditions A – F are not sufficient to obtain the covariance between the relatives in terms of only variances and marital correlations.

Appendix

Lemma. For any three events A, B, C

$$\Pr[A|B, C] = \frac{\Pr[A|B]\Pr[C|A, B]}{\Pr[C|B]}. \quad (\text{A1})$$

This may be easily verified by multiplying both the numerator and denominator by $\Pr[B]$.

Corollary 8.

$$\Pr[x_i|X_i, X_j, x_j, (i * j)] = \Pr[x_i|X_i]. \quad (\text{A2})$$

According to the Lemma,

$$\Pr[x_i|X_i, X_j, x_j, (i * j)] = \Pr[x_i|X_i] \frac{\Pr[X_j, x_j, (i * j)|x_i, X_i]}{\Pr[X_j, x_j, (i * j)|X_i]},$$

or, using an obvious transformation,

$$= \Pr[x_i|X_i] \frac{\Pr[(i * j)|X_i, X_j, x_i, x_j]}{\Pr[(i * j)|X_i, X_j, x_j]} \frac{\Pr[X_j, x_j|x_i, X_i]}{\Pr[X_j, x_j|X_i]}. \quad (\text{A3})$$

The second factor in (A3) cancels because its numerator and denominator are the same according to Condition E. The numerator of the third factor represents the probability of an individual having phenotype X_j and genotypic value x_j , given that some other individual without any connections to the first one has phenotype X_i and the genotypic value x_i . Obviously,

$$\Pr[X_j, x_j|x_i, X_i] = \Pr[X_j, x_j].$$

For the same reason, for the denominator of the third factor in (A3) we have

$$\Pr[X_j, x_j|X_i] = \Pr[X_j, x_j].$$

Hence, the third factor in (A3) also cancels, and (A3) is reduced to (A2).

Corollary 9. As an example, let us consider relation (9c), since the proofs of (9a) and (9b) follow along the same path.

$$E(x_i|x_j, (i * j)) = a_k b_k \rho_k x_j. \quad (\text{A4})$$

The left side of this can be expanded as

$$E(x_i|x_j, (i * j)) = \sum_{x_i} \sum_{X_j} \sum_{x_i} x_i \Pr[x_i|X_i, X_j, x_j, (i * j)] \Pr[X_i, X_j|x_j, (i * j)].$$

From (A2) follows

$$E(x_i|x_j, (i * j)) = \sum_{x_i} \sum_{x_j} \sum_{x_i} x_i \Pr[x_i|X_i] \Pr[X_i, X_j|x_j, (i * j)],$$

or, according to Condition C, (4b),

$$= b_k \sum_{x_i} \sum_{x_j} X_i \Pr[X_i, X_j|x_j, (i * j)],$$

which can be transformed as

$$= b_k \sum_{x_i} \sum_{x_j} X_i \Pr[X_i|X_j, x_j, (i * j)] \Pr[X_j|x_j, (i * j)]. \tag{A5}$$

According to the Lemma, the first distribution in (A5) can be represented as

$$\Pr[X_i|X_j, (i * j)] = \frac{\Pr[x_j|X_i, X_j, (i * j)]}{\Pr[x_j|X_j, (i * j)]}. \tag{A6}$$

It follows from (A2) that the numerator and denominator in (A6) are the same, and, therefore,

$$\Pr[X_i|X_j, x_j, (i * j)] = \Pr[X_i|X_j, (i * j)]. \tag{A7}$$

Considering the second distribution in (A5), notice that the information that an individual in the reproducing phase has an unspecified mate, does not affect the probability for the individual to have a particular phenotype, given its genotypic value, i.e.,

$$\Pr[X_j|x_j, (i * j)] = \Pr[X_j|x_j]. \tag{A8}$$

The substitution of (A7) and (A8) into (A5) yields

$$E(x_i|x_j, (i * j)) = b_k \sum_{x_i} \sum_{x_j} X_i P_k(X_i|X_j, (i * j)) \Pr[X_j|x_j].$$

References to (4a) and (7) conclude the proof of (A4).

Proof of (4c).

$$\frac{v_k}{V_k} = \frac{b_k}{a_k}. \tag{A9}$$

Consider the expectation of the product xX :

$$E(xX) = \sum_x \sum_X xX \Pr[x, X].$$

This can be written as

$$E(xX) = \sum_x \sum_X xX \Pr[x|X] P_k(X), \tag{A10}$$

or as

$$E(xX) = \sum_x \sum_X xX \Pr[X|x] p_k(x). \tag{A11}$$

Using (4b), from (A10) follows

$$E(xX) = b_k V_k. \tag{A12}$$

At the same time, using (4a), from (A11) follows

$$E(xX) = a_k v_k. \tag{A13}$$

Then, (A9) follows directly from (A12) and (A13).

Acknowledgments. J. Crow and W. Engels provided valuable help and criticism. My contacts with T. Nagylaki were very productive. Comments on the manuscript by C. Cockerham and B. Weir were very useful. Discussions with Prof. Sewall Wright were particularly useful and inspiring. I also wish to thank both referees for their detailed and important comments.

References

- Bulmer, M. G.: Regression between relatives. *Genet. Res.* **28**, 199–203 (1976)
- Cloninger, C. R., Rice, J., Reich, T.: Multifactorial inheritance with cultural transmission and assortative mating. II. A general model of combined polygenic and cultural inheritance. *Am. J. Hum. Genet.* **31**, 176–198 (1979)
- Cotterman, C. W.: Relationship and probability in Mendelian populations. (Unpublished notes) 1960
- Crow, J. F., Felsenstein, J.: The effect of assortative mating on the genetic composition of a population. *Eugenics Quart.* **15**, 85–97 (1968)
- Feldman, M. W., Cavalli-Sforza, L. L.: Quantitative inheritance, stabilizing selection, and cultural evolution. In: *Proceedings of the International Conference on Quantitative Genetics* (Pollak, H., Kempthorne, O., eds.), pp. 761–777, 1977
- Feldman, M. W., Cavalli-Sforza, L. L.: Aspects of variance and covariance analysis with cultural inheritance. *Theoret. Population Biology* **15**, 276–307 (1979)
- Fisher, R. A.: The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edinburgh* **52**, 399–433 (1918)
- Goldberger, A. S.: Models and methods in the IQ debate: Part I, Revised. Social System Research Institute, University of Wisconsin 1978
- Karlin, S.: Models of multifactorial inheritance: I. Multivariate formulations and basic convergence results. *Theoret. Population Biology* **15**, 308–355 (1979a)
- Karlin, S.: Models of multifactorial inheritance: II. The covariance structure for a scalar phenotype under selective assortative mating and sex-dependent symmetric-parental transmission. *Theoret. Population Biology* **15**, 356–393 (1979b)
- Karlin, S.: Models of multifactorial inheritance: III. Calculation of covariance of relatives under extended selective mating mechanisms. *Theoret. Population Biology* **15**, 394–424 (1979c)
- Moran, P. A. P., Smith, C. A. B.: Commentary on R. A. Fisher's paper on "The correlation between relatives on the supposition of Mendelian inheritance". *Eugenics Lab. Memoirs No. 46* (1966).
- Morton, N. E.: Analysis of family resemblance: I. Introduction. *Am. J. Hum. Genet.* **26**, 318–330 (1974)
- Nagylaki, Th.: The correlation between relatives with assortative mating. *Ann. Hum. Genet.* **42**, 131–137 (1978)
- Rao, D. C., Morton, N. E.: IQ as a paradigm in genetic epidemiology. In: *Genetic epidemiology* (Morton, N. E., Chung, C. K., eds.), pp. 145–193, 1978
- Rice, J., Cloninger, C. R., Reich, T.: Multifactorial inheritance with cultural transmission and assortative mating. I. Description and basic properties of the unitary models. *Am. J. Hum. Genet.* **30**, 618–643 (1978)
- Vanderberg, S. G.: Assortative mating, or who marries whom? *Behavior Genet.* **2**, 127–157 (1972)
- Wagener, D. K.: Preferential mating: Nonrandom mating of a continuous phenotype. *Theoret. Population Biology* **10**, 185–204 (1976)
- Wilson, S. R.: The correlation between relatives under the multifactorial model with assortative mating: I. The multifactorial model with assortative mating. *Ann. Hum. Genet.* **37**, 289–304 (1973)
- Wright, S.: Systems of mating: III. Assortative mating based on somatic resemblance. *Genetics* **6**, 144–161 (1921a)
- Wright, S.: Systems of mating: V. General considerations. *Genetics* **6**, 167–178 (1921b)

Received March 23/Revised July 20, 1981