Certain Multivariate Problems Arising in Human Genetics (An Argorithm for Multivariate Regression Analysis)

Schull, William J.
Department of Human Genetics, University of Michigan Medical School

Kudo, Akio
Department of Human Genetics, University of Michigan Medical School

CERTAIN MULTIVARIATE PROBLEMS
ARISING IN HUMAN GENETICS

(AN ARGORITHM FOR MULTIVARIATE
REGRESSION ANALYSIS)

By

William J. SCHULL and Akio KUDÔ

Department of Human Genetics
University of Michigan Medical School
Ann Arbor, Michigan

At a given genetic locus, natural selection may operate to favor individuals who are homozygous or heterozygous, that is, who possess two genes which are alike or two genes which are unlike in function. In the former instance, the frequencies of the genes at equilibrium depend upon selection *and* mutation; whereas when the heterozygote is favored, the gene frequencies at which equilibrium occurs are dependent upon selection alone. Thus, the response of these two systems to an increase in mutation would be different—the former being sensitive to change in mutation rate, and the latter not. At a time when man's exposure to mutagenic agents is on the increase, the relative importance of these two systems, in the overall picture of selection, is of considerable concern. Some insight into their importance is afforded by studies on the effect of inbreeding on variables indicative of the general size of individuals. The pertinence of these studies arises as follows : An inevitable consequence of inbreeding is a decay in the frequency of heterozygosity; the rate of this decay is a function of the degree of inbreeding. Now, it is possible to show that the two systems of selection we have described are dissimilarly affected by a decay in heterozygosity. Thus, if we can measure inbreeding and, simultaneously, size in a group of individuals, an evaluation of the importance of the two systems is possible. A conventional measure of inbreeding is the coefficient of inbreeding, F, which may be defined as the probability that, at a given genetic locus, an individual will be homozygous for genes alike in function and origin. Clearly a variety of measures bearing on body size exist. We shall here limit our attention to measurements of growth and development, and, specifically, to measurements on the head.

The statistical problem and the model :— A number of factors other than parental inbreeding are known to influence the measurements under discussion. To cite but a few, we know that age, sex, and nutrition may all be of importance in determining the measurement of a particular child

at a particular time.   Clearly if unambiguous answers are to be obtained
with reference to the genetic question of interest, then the possible effects
of concomitant variation must be allowed for.   Our analysis must also take
into account the fact that head measurements are interdependent.   The gene-
ral problem may, therefore, be stated as follows: What is the effect of
inbreeding on the dimensions of the head when allowance is made for con-
comitant variation and the interdependence of the measures used to describe
head size?

We begin the analysis with the assumption that the distribution of these
four head measurements can be described by the multivariate normal density
function

(1)
$$\frac{1}{\sqrt{(2\pi)^n |\Omega|}} \exp \left[-\frac{1}{2} (x-\mu)' \Omega^{-1} (x-\mu) \right]$$

where μ is the mean vector, Ω is the variance-covariance matrix, and n is
the dimension of the vector of variables; in our particular case $n=4$.   This
model is valid for those children of specified sex, age, A, and inbreeding
coefficient, F.   The effects of concomitant variation, if randomly distributed,
appear on the mean vector, μ, and the variance-covariance matrix Ω, but
need not affect the normality of the distribution.

It is the usual practice to assume linearity of age and consanguinity
effects; that is, that values of these 4 measurements can be represented as
follows:

(2)
$$x_1 = m_1 + a_1 A + b_1 F + \epsilon_1$$
$$x_2 = m_2 + a_2 A + b_2 F + \epsilon_2$$
$$x_3 = m_3 + a_3 A + b_3 F + \epsilon_3$$
$$x_4 = m_4 + a_4 A + b_4 F + \epsilon_4 ,$$

where (ϵ_1, ϵ_2, ϵ_3, ϵ_4) are the error terms, and $m' = (m_1, m_2, m_3, m_4)$ $a' =$
(a_1, a_2, a_3, a_4) and $b' = (b_1, b_2, b_3, b_4)$ are called the vectors of the general
mean, the age effect, and the consanguinity effect.   As a consequence of the
interdependence of the four measurements the correlations among the error
terms ϵ_1, ϵ_2, ϵ_3, ϵ_4, are non-negligible, and hence these data cannot be
meaningfully analyzed by a series of univariate regressions, and we are
obliged to use multivariate normal regression theory.   Thus, we assume ϵ_1,
ϵ_2, ϵ_3, ϵ_4, to be distributed in the multivariate normal distribution with
mean vector zero and a variance-covariance matrix, which is assumed to
be independent of the concomitant variables, age, and inbreeding coefficient.

Before we consider this specific set of observations further, it seems
appropriate to review the statistical theory pertinent to the type of analysis
to be illustrated.   Briefly this theory is as follows:

The model: Consider the q-population, k-variate linear regression model,
namely,

$$y_1 = a_{p11} X_1 + \cdots\cdots + a_{p1n} X_n + \epsilon_1$$

(3)
$$\vdots \qquad\qquad\qquad\qquad p = 1, \cdots\cdots\cdots, q,$$

$$y_k = a_{pk1} X_1 + \cdots\cdots + a_{pkn} X_n + \epsilon_n$$

where $y_1, \cdots, y_k$ are the measurements of interest, $x_1, \cdots, x_n$ are the concomitant variables, a_{pij}'s are the regression coefficients for the ith measurements, y_i, on the jth concomitant variable, x_j, of the pth population, $\epsilon_1, \cdots,$ ϵ_k are the error terms which are distributed in the multivariate normal form with mean vector zero and variance matrix $\sum_p$.

An observation consists of n concomitant variables and k measurements $(x_1, \cdots, x_n, y_1, \cdots, y_k)$. In our particular case, the concomitant variables are the constant 1, age, and the inbreeding coefficient, and the vectors of regression coefficients corresponding to these three concomitants we call the general mean, the age effect, and the consanguinity effect. On the basis of samples of $N_1, N_2, \cdots, N_q$ observations drawn from the first, second, $\cdots, q^{\text{th}}$ population, we wish to make inferences about the parameters a_{pij} and $\sum_p$. Theory suggests the use of the raw sums of squares and cross-products of the samples, $(\sum_{xixj}, \sum_{xiyj} \sum_{yiyj})$, as the material for analysis. We may write these quantities in the pth population as

(4) $$S^{(p)}(x_i x_j), \quad S^{(p)}(x_i y_j), \quad S^{(p)}(y_i y_j) ;$$

collectively they form the matrix

(5) $$S^{(p)} = \begin{bmatrix} S^{(p)}(x, x) & S^{(p)}(x, y) \\ S^{(p)}(x, y)' & S^{(p)}(y, y) \end{bmatrix}, \quad p = 1, 2, \cdots, q$$

of dimension $k + n$, with submatrices $S^{(p)}(x, x)$, $S^{(p)}(x, y)$, $S^{(p)}(y, y)$ corresponding to the cross products of concomitant variables, those of the concomitant variable and measurements, and those of measurements, respectively.

The statistical inferences can be broadened by introducing certain assumptions, namely that

  1) the variances, $\sum^{(1)}, \cdots, \sum^{(p)}$ are homogeneous and/or
  2) some of the regression coefficients are homogeneous. Without loss of generality we can express the latter assumption as follows:

(6) $$a_{1ij} = a_{2ij} = \cdots = a_{pij} \qquad \begin{matrix} i = 1, \cdots, k \\ j = l+1, \cdots, n. \end{matrix}$$

If this assumption is true, we can combine the q sets of regression equations into one by modifying the definition of the concomitant variables as follows:

$$
\begin{aligned}
y_1 = {} & (a_{111}X_{11} + \cdots + a_{q11}X_{q1}) + \cdots + (a_{11(l-1)}X_{1(l-1)} + \cdots \\
& + a_{q1(l-1)}X_{q(l-1)}) + a_{1l}(X_{l1} + \cdots + X_{ql}) + \cdots + \\
& a_{1n}(X_{1n} + \cdots + X_{qn}) + \epsilon_1 \\
\vdots \\
y_k = {} & (a_{1k1}X_{11} + \cdots + a_{qk1}X_{q1}) + \cdots + (a_{1k(l-1)}X_{1(l-1)} + \cdots \\
& + a_{qk(l-1)}X_{(ql-1)}) + a_{kl}(X_{kl} + \cdots + X_{ql}) + \cdots + \\
& a_{1n}(X_{1n} + \cdots + X_{qn}) + \epsilon_k,
\end{aligned}
\tag{7}
$$

where $x_{ij} = x_j$ if the observation belongs to the ith population and $x_{ij} = 0$ otherwise. Thus, we have a regression equation involving $(l-1)\,q + k - l$ concomitant variables and vectors of regression coefficients.

The relation between the matrix of cross products $S^{(p)}$ $(p=1, \cdots, q)$ and the same matrix S based on the above model is the following:

The concomitant variables are classified into two; those having regres-sion coefficients different from population to population, and those where the regression coefficients are the same for all populations. Therefore $S^{(p)}(x, x)$ can be divided into four parts,

$$
\begin{bmatrix}
S^{(p,1)}(x, x) & S^{(p,2)}(x, x) \\
S^{(p,2)}(x, x)' & S^{(p,3)}(x, x)
\end{bmatrix},
\tag{8}
$$

where $S^{(p,1)}(x, x)$, $S^{(p,2)}(x, x)$, $S^{(p,3)}(x, x)$ are respectively, the sums of squares and cross-products of the concomitant variables of both the first categories, the sums of cross-products of the first and the second, and sums of squares and cross-products of both the second categories.

In the same way, $S^{(p)}(x, y)$ is divided into two

$$
\begin{bmatrix}
S^{(p,1)}(x, y) \\
S^{(p,2)}(x, y)
\end{bmatrix}.
\tag{9}
$$

These matrices to be used in the regression analysis, and the ones furnished by the data from each population are connected by the following relations:

$$
(10) \quad S(x, x) =
\begin{pmatrix}
S^{(1,1)}(x, x) & & & & S^{(1,2)}(x, x) \\
& S^{(2,1)}(x, x) & & & S^{(2,2)}(x, x) \\
& & \ddots & & \vdots \\
& & & S^{(q,1)}(x, x) & S^{(q,2)}(x, x) \\
S^{(1,2)}(x, x) & S^{(2,2)}(x, x) & \cdots\cdots\; S^{(q,2)}(x, x) & & \sum_{i=1}^{q} S^{(i,2)}(x, x)
\end{pmatrix}
$$

$$(11) \quad S(x, y) = \begin{matrix} S^{(1,1)}(x, y) \\ \vdots \\ S^{(q,1)}(x, y) \\ \sum_{i=1}^{q} S^{(i,2)}(x, y) \end{matrix}$$

$$(12) \quad S(y, y) = \left[\sum_{i=1}^{q} S^{(i)}(y, y) \right]$$

The matrix

$$(13) \qquad \begin{bmatrix} S(x, x) & S(x, y) \\ S(y, x) & S(y, y) \end{bmatrix}$$

is called the working matrix, and

$$(14) \qquad \begin{bmatrix} S^{(p)}(x, x) & S^{(p)}(x, y) \\ S^{(p)}(y, x) & S^{(p)}(y, y) \end{bmatrix} \qquad (p=1, 2, \cdots, q)$$

is called the data matrix.

If $x_1 = 1$ for all observations, and the regression on x_1 is not of primary interest, the matrix of corrected sums of squares and products can be used instead of the raw sums. The basic arguments remain unchanged.

Estimation:—The material from whence the estimates are derived is the symmetric matrix

$$(15) \qquad S = \begin{bmatrix} S(x, x) & S(x, y) \\ S(x, y)' & S(y, y) \end{bmatrix}$$

The submatrices, $S(x, x)$, $S(x, y)$, and $S(y, y)$ are of dimensions (n, n), (n, k), and (k, k), respectively.

General theory tells us the following: the matrix of estimates $\hat{\boldsymbol{\alpha}}_i$ of $\boldsymbol{\alpha}_i$, the regression vector on x_i, are given by

$$(16) \qquad \begin{pmatrix} \hat{\boldsymbol{a}}_1' \\ \vdots \\ \hat{\boldsymbol{a}}_n' \end{pmatrix} = [S(x, x)]^{-1} S(x, y) \text{ and}$$

the estimate of the variance matrix $\sum$ is given by

$$(17) \qquad \hat{\sum} = [S(y, y) - (S(x, y))' S(x, x)^{-1} S(x, y)]/d,$$

where d is some constant (the degrees of freedom), and the variance or covariance matrix of $\hat{a}_i$ and $\hat{a}_j$ is equal to $C_{ij}\sum$, where C_{ij} is the (i, j) element of $S(x, x)^{-1}$.

Tests of significance:—Four types of hypotheses are of interest here. Firstly, we wish to know whether a specific vector of regression coefficients, say $\boldsymbol{a}_i$, differs significantly from a vector all of whose elements are zero. The hypothesis is, then,

   a) $H_0: \boldsymbol{a}_i = 0$

(In this case, the superscript p is not pertinent and is deleted). A test of this hypothesis is afforded by the statistic, $C^{(ii)} \hat{\boldsymbol{a}}_i \hat{\Sigma}^{-1} \hat{\boldsymbol{a}}_i$, which is distributed as χ^2 with d.f., k, when the sample size (N) is large, otherwise as F with d.f., k and $N-k+1$. In the above statistic, $C^{(ii)}$ is the $(i, i)^{\text{th}}$ element of the inverse of $S(x, x)$, $\hat{\boldsymbol{a}}_i$ is the estimate of the vector, $\boldsymbol{a}_i$, and $\hat{\boldsymbol{a}}_i'$ is the transpose of $\hat{\boldsymbol{a}}_i$.

Secondly, we are interested in ascertaining the legitimacy of the assumption that the variance-covariance matrices of the q-populations are equal. The hypothesis is, now,

b) $H_0 : \Sigma^{(1)} = \cdots = \Sigma^{(q)}$

and the appropriate statistic is

$$\prod_{p=1}^{q} \left| \hat{\Sigma}^{(p)} \right| \Big/ \left| \sum_{p=1}^{q} \hat{\Sigma}^{(p)} \right| .$$

Each term in the numerator is the determinant of an estimate of the variance matrix obtained by analyzing each sample separately. The denominator is obtained by analyzing the working matrix as if the hypothesis under test were true.

Thirdly, we are interested in knowing whether a specific vector is homogeneous over the p populations. That is,

c) $H_0 : \boldsymbol{a}_i^{(1)} = \cdots = \boldsymbol{a}_i^{(p)} \quad i = j_1, j_2, \cdots, j_l, \ l \leq n.$

The statistic is now $\left| \hat{\Sigma}^{(A)} \right| \Big/ \left| \hat{\Sigma}^{(H)} \right|$. The numerator is the determinant of the matrix obtained by analyzing the working matrix assuming the variance matrices to be homogeneous; whereas the denominator is based upon the assumptions that the hypothesis is true and the variance matrices are homogeneous.

For the distribution of the statistics appropriate to b) and c), the reader is referred to T. W. Anderson, pages 178–210 and 247–259.

Finally, if we are interested in the comparability of two vectors but cannot assume that the variance matrices are homogeneous, then the following hypothesis and test is in order.

d) $H_0 : \hat{\boldsymbol{a}}_i^{(1)} = \hat{\boldsymbol{a}}_i^{(2)}.$

The statistic is

$$(\hat{\boldsymbol{a}}_i^{(1)} - \hat{\boldsymbol{a}}_i^{(2)})' [C_{(1)}^{ii} \hat{\Sigma}^{(1)} + C_{(2)}^{ii} \hat{\Sigma}^{(2)}]^{-1} (\hat{\boldsymbol{a}}_i^{(1)} - \hat{\boldsymbol{a}}_i^{(2)}),$$

which is distributed as χ^2 with k degrees of freedom if N_1 and N_2 are both large.

Computational logic :—To put into practice the analysis here outlined was, in pre-computer days, a formidable undertaking. Today and particularly as a consequence of the development of more flexible algorithmic languages, such as MAD (Michigan Algorithm Decoder), the computations

are no longer so formidable a task. In fact, the procedures here outlined have been programmed in MAD. Certain aspects of this programming are, we believe, widely applicable.

To derive the estimates and the data required for tests of significance of hypotheses b.) and c.) the calculations of the denominators and the numerators can be easily included in one program. To do this, we use, as input, the elements of the data matrix arrangde in the order

$$(S_{11}, S_{12}, \cdots, S_{1m}, S_{22}, \cdots, S_{2m}, \cdots, S_{mm}).$$

There are two important issues here. The first one is the reordering of the input data. In the data matrix, if the subscripts assigned to the concomitant variables are smaller than those assigned to the measurements, and if the equality of the regression vectors is assumed, then the matrix of the sum of squares and products of the concomitant variables can be split into four submatrixes as in (8). The nature of the reordering may vary from one computation to the next. For example, if the consanguinity effect is assumed to be common for all populations but the age effect is not, then age should be the first, and the inbreeding coefficient the second variable. If we assume the contrary, then the inbreeding coefficient should be the first variable, and the age the second. A single control card, preceding the input data, serves to indicate the permutation to be applied to the subscripts.

The second issue is the formation of the working matrix, shown in (15). Again, a single control card is sufficient to specify completely this task. The control card must, however, indicate

1) number of populations.
2) number of concomitant variables whose regression vectors differ from population to population (l in (7)).
3) number of concomitant variables in the working regression equation, or equivalently the dimension of $S(x, x)$ in (13).
4) dimension of the working matrix.

After construction of the working matrix, the next step is to calculate the estimates of the parameters, and the determinants of the variance matrix. These calculations are straightforward, and we need only assert that the three sets of estimates, namely, the matrix of the coefficients, (C_{ij}), the regression estimates and the estimate of the variance matrix need not be computed separately. The well-known Gauss-Doolittle method will compute these three sets of estimates simultaneously (see Rao, 1955, Appendix).

In the program which we use other computations have been added to assist in the evaluation of the probability level of the test statistic, and the statistic needed for test a) is also computed. Still another program has been written which can compute the statistic for the test a) and d), as

well as others.

We are not, as yet, prepared to make definite statements with respect to the memory capacity required by the program we have been discussing, for we feel that a number of refinements can still be made. At present, and as a MAD program for the IBM 704, the following restriction holds:

$$2n + k \leq 40,$$

where n is the number of concomitant variables and k the number of measurements.

The data:—In the course of a study on the genetic effects of the atomic bombs in Hiroshima and Nagasaki, some 71,280 children were examined in these two cities between 1948 and 1953. Of this number, 4,598 children were born to parents who reported a consanguineous marriage at the time of pregnancy registration and had received no or inappreciable amounts of radiation. The findings in these children at birth, and, in a subsample, at nine months of age, as contrasted to the remainder, were sufficiently provocative to indicate the need for further studies (see Schull, 1958; Morton, 1958). Between 1958 and 1960, comprehensive follow-up studies were undertaken on these children of consanguineous marriages and a suitable control group. The nature of these follow-up studies have been described elsewhere (Schull and Neel, 1961). In all 9,382 children, approximately half of whom were of consanguineous parèntage, were selected for reexamination. Of this number, about 6,700 were still alive and available for examination in Hiroshima or Nagasaki in the years mentioned. As one facet of

Table 1. The distribution by sex, age, and parental relationship of children seen in Hiroshima as part of the Child Health Survey.

Age (months)	Males Parental relationship				Females Parental relationship			
	Unrelated	2nd C.	1½ C.	1st C.	Unrelated	2nd C.	1½ C.	1st C.
60–65	13	8	9	15	23	12	2	11
66–71	54	9	6	15	50	8	11	12
72–77	62	16	10	24	50	13	7	28
78–83	62	10	6	28	46	8	8	21
84–89	57	10	8	37	61	14	13	22
90–95	75	9	16	29	64	14	14	31
96–101	69	14	12	39	60	17	14	34
102–107	79	13	13	32	71	11	9	35
108–113	75	16	9	37	63	12	18	46
114–119	100	14	15	32	86	10	17	42
120–125	83	9	8	38	59	19	7	32
126–131	18	6	2	14	25	4	5	11
132–138	1	—	—	—	2	—	—	—
Total	749	134	114	340	660	142	125	325

the stuidy of these children, detailed anthropometric measurements were obtaned. We wish to make infcrences regarding the effect of inbreeding on a subset of these anthropometric measurements, namely, those derived from the head. Our problem, then, is to ascertain whether or not inbreeding alters, in a predictable fashion, the girth, height, breadth, and length of the head. Since our purposes are essentially illustrative. we shall, for convenience, restrict our attention to just those children who were examined in Hiroshima.

In Table 1, we present the distribution of the children whose measurements are to be analyzed by age, sex, and consanguinity. In Tables 2–6, we present in numerical fcrm the logic previously set out in more general terms. Table 2 is, in effect, the raw data ; here, it consists of the corrected

Table 2. Data matrices consisting of corrected sums of squares and products of age, inbreeding coefficient, and the four head measurements. Age is measured in months, F as the numerator of a fraction whose denominator is 64, and all head measurements are in mm.

Age	F	Head			
		Circum.	Length	Breadth	Height
HIROSHIMA MALES					
428164.7461	−607.6191	85927.0000	27225.5000	19650.5000	12990.5000
	3789.9561	−2810.0078	−1164.0586	−796.7480	−898.0977
		211224.0000	86032.0000	39900.0000	31111.0000
			80705.0000	−11873.0000	11172.5000
				47039.2500	9317.0000
					23305.5000
HIROSHIMA FEMALES					
425391.2500	1645.1699	137599.0000	50443.5000	20928.7500	5295.2500
	3526.1661	−1175.5000	−290.41418	−1044.1914	−125.7090
		200872.0000	0993.0000	32873.0000	25470.0000
			69904.5000	−7996.2500	9020.0000
				132244.0000	8627.5000
					23268.2500

sums of squares and products of age, inbreeding, and the four head measurements for the sexes separately. Table 3 indicates the method of construction of the working matrix appropriate to a specific hypothesis, namely, that age effects are different in the sexes whereas the inbreeding effects are not. Table 3 further illustrates the reduction of this matrix by the Gauss-Doolittle method to obtain the vectors of the estimates of the regression coefficients, a matrix of coefficients which when combined with the variancecovariance matrix will yield a matrix whose elements are the variances and covariances of the regression estimates, and finally, the residual matrix. In Tables 4–6, we give the results of analyzing the data in Table 2 under several different assumptions. In Table 4, the sexes are analyzed separately both with res-

　　　　　William J. Schull and Akio Kudô

Table 3. Formation of the working matrix: Age effects are treated
separately for the sexes, but inbreeding effects are pooled.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)			
(1)	428164.7461	0	−607.6191	85927.0000	27225.5000	19650.5000	12990.5000	1	0	0
(2)	0	425391.2500	1645.1699	137599.0000	50443.5000	20928.7500	5295.2500	0	1	0
(3)	−607.6191	1645.1699	7316.1222	−3985.5078	−1454.4727	−1840.9394	−1023.8067	0	0	1
(4)	85927.0000	137599.0000	−3985.5078	412096.0000	167025.0000	72773.0000	56581.0000			
(5)	27225.5000	50443.5000	−1454.4727	167025.0000	150609.5000	−19869.2500	20192.5000			
(6)	19650.5000	20928.7500	−1840.9394	72773.0000	−19869.2500	179283.2500	17944.5000			
(7)	12990.5000	5295.2500	−1023.8067	56581.0000	20192.5000	17944.5000	46573.7500			

In the construction of the above matrix, the first row (or column) is transferred from the first
row of the male data matrix; the second row (column) stems from the first row of the female
data matrix; the remaining elements are obtained by summing corresponding elements from the
male and female data matrices, for example, the element 3, 3 in the above is the sum of 3789.9561
and 3526.1661.

Table 3 (continued). The first step in the preceding matrix by the
Gauss-Doolittle Method.

1	0	−.001419	.200687	.063587	.045895	.030340	.00000234	0	0
0	425391.2500	1645.1699	137599.0000	50443.5000	20928.7500	5295.2500	0	1	0
0	1645.1699	7315.2600	−3863.5665	−1415.8360	−1813.0527	−1005.3715	.00142183	0	1
0	137599.0000	−3863.5774	394851.5682	161561.1599	68829.3803	53973.9748			
0	50443.5000	−1415.8397	161561.1961	148878.3121	−18619.7357	19366.4783			
0	20928.7500	−1813.0553	68829.4001	−18619.7337	178381.3903	17348.3038			
0	5295.2500	−1005.3732	53973.9755	19366.4731	17348.3010	46179.6182			

At the first step　(i) adjust the scale of the first row so that the first element in the row
becomes one.
(ii) eliminate the first element of all the remaining rows.

At the nth step　(i) adjust the scale of the nth row so that the nth element in the nth row
becomes one.
(ii) eliminate the nth element of all the remaining rows.

Table 3 (continued). Finally, we arrive, after the third sweep-out,
at the following matrix;

		(1)	(2)	(3)	(4)			
(1)	1 0 0	.19983327	.06327372	.04552697	.03014077	.00000233	−.00000000	.00000019
(2)	0 1 0	.32579052	.11943382	.05020101	.01299077	—	.00000235	−.00000053
(3)	0 0 1	−.60142051	−.22040574	−.25913529	−.14035640	—	—	.00013682
	0 0 0	347699.5140	144275.6777	60920.6103	51644.1854			
	0 0 0	—	142541.6085	−24017.9642	18512.4583			
	0 0 0	—	—	176860.9296	16821.9506			
	0 0 0	—	—	—	45969.7197			

The submatrix of dimensions 3×4 in the upper middle is the matrix of regression vectors. The
submatrix of dimensions 4×4 in the lower middle is the matrix of residuals which is distributed
in the Wishart distribution with degrees of freedom equal to the sum of the degrees of freedom
of the data matrices minus 3. The unbiased estimate of the variance matrix can be obtained by
dividing the elements of matrix of residuals by the degrees of freedom. The submatrix of di-
mensions 3×3 in the upper right is the matrix of coefficients to the matrix of variances and
covariances of the estimates of the regression vectors   This is identical to the inverse of the
submatrix formed by the first three rows and columns of the working matrix.

Table 4. The results of the analysis of the data in Table 2. The three matrices for each sex are the regression estimates, their coefficients, and the estimates of variance and covariance. D.F. are the degrees of freedom on which the estimates of variance and covariance are based. χ^2 is the value of the statistic for testing the hypothesis that the regression vector on inbreeding is not significantly different from a zero vector. $\triangle$ is the generalized variance.

	Head circum.	Head breadth	Head length	Head height	Age	F
HIROSHIMA MALES						
Age	0.199680	0.063165	0.045607	0.030010	.000002332	.000000373
F	−0.709422	−0.297016	−0.202914	−0.232156	—	.000263914
Head circum.	145.730362	60.523372	26.867191	21.153207		
Head length		59.665825	−10.129644	7.651901	D.F. = 1318	
Head breadth			34.887240	6.481266	χ^2 = 18.30	
Head height				17.228490	$\triangle$ = 1176553.5	
HIROSHIMA FEMALES						
Age	0.325341	0.119115	0.050435	0.012609	.000002354	−.000001095
F	−0.485156	−0.137934	−0.319658	−0.041533	—	.000284106
Head circum.	124.527719	51.593892	20.462311	18.964172		
Head length		51.125597	−8.513374	6.702896	D.F. = 1249	
Head breadth			104.767554	6.661529	χ^2 = 9.10	
Head height				18.571868	$\triangle$ = 5281988.1	

pect to the effects of age *and* inbreeding. In Table 5, it is assumed that the inbreeding effects are common in the two sexes whereas the age effects are not. Finally, in Table 6, we assume age effects are common, but inbreeding effects are not. It is interesting, perhaps, to point out at this juncture that the average time spent in the calculations for each of the above analyses was less than ten seconds. While it is not our purpose to dwell ont he biological findings, it should be noted that there is a significant depression of the four head measurements with inbreeding, and that the residual matrix is larger in the case of the female than in the male. With

Table 5. Analysis of the data in Table 2 under the assumption that the inbreeding effects are equal in the two sexes.

	Head circum.	Head length	Head breadth	Head height	Age of male	Age of female	F
Age of male	0.199833	0.063274	0.045527	0.030141	0.000002332	−0.000000000	0.000000194
Age of female	0.325791	0.119434	0.040201	0.012991	—	0.000002347	−0.000000529
F	−0.601421	−0.220406	−0.259135	−0.140356	—	—	0.000136815
Head circum.	135.397007	56.182117	23.722979	20.110664			
Head length		55.506857	−9.352790	7.208901			
Head breadth			68.871078	6.550604			
Head height				17.900981			
	D.F.=2568		χ^2=23.23			$\triangle$=3429757.1	

Table 6. Analysis of the data in Table 2 assuming equality of age effects in the sexes, but not the inbreeding effects.

	Head circum.	Head length	Head breadth	Head height	F of male	F of female	Age
F of male	−0.699390	−0.292549	−0.202529	−0.233546	.0002633885	−0.000000082	.000000186
F of female	−0.455724	−0.124829	−0.318527	−0.045609	—	.000283845	−.000000544
Age	0.262257	0.091027	0.048011	0.021345	—	—	.000001170
Head circum.	136.672049	56.740385	23.791947	19.898766			
Head length		55.748727	−9.317169	7.106532			
Head breadth			68.863329	6.559442			
Head height				17.900296			
	D.F.=2568		χ^2=8.72		△=3481591.2		

respect to this latter observation, we would point out that because of the large sample size biologically meaningless differences in the variance matrices are apt to be statistically highly significant, and that under these circumstances test procedure d) appears to be more appropriate than b).

Summary: Considerable concern presently exists with respect to the genetic damage accruing to mankind as a consequence of his increasing exposure to mutagenic agents. Information pertinent to the magnitude of this risk is afforded by studies on the effect of inbreeding on a number of variables indicative of general fitness. These latter variables tend to be interdependent measures of the same phenomenon; hence, traditional analyses of the univariate variety are inappropriate and must be supplanted by multivariate procedures. In the past, multivariate analysis has been hampered by the cumbersomeness of the computations, not to mention certain interpretative ambiguities in the earlier tests. High-speed computers have, in the main, overcome the computational difficulties, and in so doing, have opened up opportunities for the devising of less ambiguous and more broadly applicable statistical procedures. The relevance of these procedures to a broad class of genetic problems, including the radiation hazard, is briefly indicated in this paper.

References

1. Anderson, T. W. 1958. *An Introduction to Multivariate Statistical Analysis.* John Wiley and Sons.
2. Morton, N. 1958. Empirical risks in consanguineous marriages: Birthweight, gestation time, and measurements of infants. *Am. J. Human Genet.,* 10: 344–349.
3. Rao, C. R. 1955. Analysis of dispersion for multiple classified data with unequal numbers in cells. *Sankya,* 15: 253–280.
4. Schull, W. J. 1958. Empirical risks in consanguineous marriages: Sex ratio, malformation, and viability. *Am. J. Human Genet.,* 10: 294–343.
5. Schull, W. J. and J. V. Neel. 1961. The Child Health Survey: A Genetic Study in Japan. *UN/ WHO Seminar on the Use of Vital and Health Statistics for Genetic and Radiation Studies,* (In press)