ON THE BIOMETRICAL ANALYSIS OF QUANTITATIVE INHERITANCE[1]

WAI-YUAN TAN and CHENG-HSUAN YUAN[2]

(Received April 16, 1965)

In the analysis of quantitative inheritance as developed by Mather (1949), the partition of genetic components is effected by the conventional least square method using the statistics V_{F_2}, $\bar{V}_{F_3}$ and $V_{\bar{F}_3}$ etc. The estimates of the environmental components for V_{F_2} and $\bar{V}_{F_3}$ and for $V_{\bar{F}_3}$ are given respectively by E_1 and E_2 computed from data taken from the parents and/or F_1. As Mather's method may not be very satisfactory in controlling soil heterogeneity when the block effects are very big, we shall try to undertake another approach in which we allow for more effective statistical control of the above mentioned soil heterogeneity.

Proposed experimental design

The design introduced in this paper somewhat takes on the form of a PBIB combined with a randomized block design. Specifically, it is a group-divisible PBIB with two associates for F_2 and F_3 and a randomized complete block design for the parents and/or F_1. In other words, it is an APBIB (Augmented PBIB) with the parents and/or F_1 being assigned as augmented lines. For illustration, let us assume that we have ba F_3 lines (each F_3 line consisting of r individuals) and ba_1r F_2 individuals. We may then arrange our F_3 and F_2 in a group-divisible PBIB with two associates so that any two individuals in the same block (considered as a group) appear together in one and only one block, while any two individuals in different blocks do not appear together in any block. Let there be p augmented lines (Parents and/or F_1). Then in each block of our design, there are $a+a_1+p$ plots, namely a plots for F_3 line (each line occupies one plot), a_1 plots for F_2 individuals and p plots

(1) Paper No. 39 of the Scientific Journal Series, Institute of Botany, Academia Sinica.

(2) Respectively, Associate Research Fellow and Assistant Research Fellow of the Institute of Botany, Academia Sinica.

The author wishes to thank Dr. H. W. Li, the Director of the Institute, for his encouragements and stimulations during this work. Thanks are also due to Mr. T. S. Weng who assisted in the preparation of the manuscript.

for the augmented lines, with each plot containing r individuals. It is easily seen that this design is connected and that the difference of any two treatments is an estimable function. Based on this design, we shall carry on our analysis in the following sections.

The mathematical model

Let $N_{F_2}=a_1r$ and $N_{F_3}=ar$ denote, respectively, the numbers of F_2 and F_3 individuals in each block. In accordance with Mather (1949)'s assumptions, we may then start our analysis with the following model (fixed model):

$$\begin{aligned} y_{ij(1)} &= \mu+b_i+f_j+e_{ij(1)} & i&=1, 2,\cdots b; & j&=1, 2,\cdots bN_{F_2} & & \\ y_{ij'(2)} &= \mu+b_i+d_{j'}+e_{ij'(2)} & i&=1, 2,\cdots b; & j'&=1, 2,\cdots bN_{F_3} & & \\ y_{ij''k(3)} &= \mu+b_i+a_{j''}+e_{ij''k(3)} & i&=1, 2,\cdots b; & j''&=1, 2,\cdots p; & k&=1, 2,\cdots r \end{aligned} \tag{3.1}$$

with the restriction

$$\sum_{j=1}^{bN_{F_2}} f_i+\sum_{j'=1}^{bN_{F_3}} d_{j'}+br\sum_{j''=1}^{p} a_{j''}=0 \tag{3.2}$$

with $y_{ij(1)}$ denotes the observation of the jth individual of F_2 in the ith block, $y_{ij'(2)}$ the observation of the j'th individual of F_3 in the ith block, and $y_{ij''k(3)}$ the observation of the kth individual of j''th augmented line in the ith block, f_j $d_{j'}$, and $a_{j''}$ (all being unknown constants) denote, respectively, the genetic values of the jth F_2 individual, the j'th F_3 individual, and the j''th augmented line. μ is the unknown population mean and b_i (unknown constants), the effect of the ith block. $e_{ij(1)}$, $e_{ij'(2)}$, and $e_{ij''k(3)}$ are the errors associated with $y_{ij(1)}$, $y_{ij'(2)}$ and $y_{ij''k(3)}$, respectively. It is assumed that $e_{ij(1)}$, $e_{ij'(2)}$ and $e_{ij''k(3)}$ are identically and independently distributed with zero means and the common variance σ^2.

Now let

$\delta_{ij}=1$ if f_j appears in the ith block
$=0$ otherwise
$\gamma_{ij'}=1$ if $d_{j'}$ appears in the ith block
$=0$ otherwise
$i=1, 2,\cdots b;\quad j=1, 2,\cdots bN_{F_2};\quad j'=1, 2,\cdots bN_{F_3}$.

The matrices, $\triangle=(\delta_{ij})_{b\times bN_{F_2}}$ and $\gamma=(\gamma_{ij'})_{b\times bN_{F_3}}$ then specify the arrangements of F_2 and F_3 in the blocks. We may therefore call them the first and the second incidence matrix. The following relationships can be shown to hold:

$$\begin{aligned} \triangle 1_{bN_{F_2}} &= N_{F_2}\, 1_b, & 1'_b\triangle &= 1'_{bN_{F_2}} \\ \gamma 1_{bN_{F_3}} &= N_{F_3}\, 1_b, & 1'_b\gamma &= 1'_{bN_{F_3}} \end{aligned}$$

where 1_n is an $n\times 1$ column vector consisting of n $1'$s. and $1_n'$ the transpose of 1_n.

Analysis in the sampling theory framework

In the sampling theory framework, we now try to obtain the linear unbiased and mininum-varianced estimators (or LUMV estimators, for short) of f_j, $d_{j'}$, and $a_{j''}$ with the restriction given by (3.2). We will use the following definitions:

$$\mathrm{H}_j=\sum_{i=1}^{b}\delta_{ij}y_{ij(1)} \qquad \text{for} \quad 1\leqq j\leqq b\mathrm{N}_{\mathrm{F}_2}$$

$$\mathrm{S}_{j'}=\sum_{i=1}^{b}\gamma_{ij'}y_{ij'(2)} \qquad \text{for} \quad 1\leqq j'\leqq b\mathrm{N}_{\mathrm{F}_3}$$

$$\mathrm{T}_{j''}=\sum_{i=1}^{b}\sum_{k=1}^{r}y_{ij''k(3)} \qquad \text{for} \quad 1\leqq j''\leqq p$$

$$\mathrm{B}_i=\sum_{j=1}^{b\mathrm{N}_{\mathrm{F}_2}}\delta_{ij}y_{ij(1)}+\sum_{j'=1}^{b\mathrm{N}_{\mathrm{F}_3}}\gamma_{ij'}y_{ij'(2)}+\sum_{j''=1}^{p}\sum_{k=1}^{r}y_{ij''k(3)} \quad \text{for} \quad 1\leqq i\leqq b$$

$$\mathrm{G}=\sum_{i=1}^{b}\mathrm{B}_i$$

$$\mathrm{Q}_{\mathrm{H}j}=\mathrm{H}_j-\sum_{i=1}^{b}\frac{1}{\mathrm{N}}\delta_{ij}\mathrm{B}_i \qquad 1\leqq j\leqq b\mathrm{N}_{\mathrm{F}_2}$$

$$\mathrm{Q}_{\mathrm{S}j'}=\mathrm{S}_{j'}-\sum_{i=1}^{b}\frac{1}{\mathrm{N}}\gamma_{ij'}\mathrm{B}_i \qquad 1\leqq j'\leqq b\mathrm{N}_{\mathrm{F}_3}$$

$$\mathrm{Q}_{\mathrm{T}j''}=\mathrm{T}_{j''}-\sum_{i=1}^{b}\frac{r}{\mathrm{N}}\mathrm{B}_i$$

$$\mathrm{N}=\mathrm{N}_{\mathrm{F}_2}+\mathrm{N}_{\mathrm{F}_3}+rp$$

By Gauss-Markov theorem, it can be easily shown that the LUMV estimator $\hat{\underline{t}}$ of $\begin{pmatrix}\underline{f}\\ \underline{d}\\ \underline{a}\end{pmatrix}$ satisfies

$$\mathrm{C}\hat{\underline{t}}=\begin{pmatrix}\underline{\mathrm{Q}}_{\mathrm{H}}\\ \underline{\mathrm{Q}}_{\mathrm{S}}\\ \underline{\mathrm{Q}}_{\mathrm{T}}\end{pmatrix} \tag{4.1}$$

where $\underline{f}_{b\mathrm{N}_{\mathrm{F}_2}\times 1}$, $\underline{d}_{b\mathrm{N}_{\mathrm{F}_3}\times 1}$ and $\underline{a}_{p\times 1}$ are the column vectors for f_j, $d_{j'}$ and $a_{j''}$, respectively; $\underline{\mathrm{Q}}_{\mathrm{H}_{b\mathrm{N}_{\mathrm{F}_2}\times 1}}$, $\underline{\mathrm{Q}}_{\mathrm{S}_{b\mathrm{N}_{\mathrm{F}_3}\times 1}}$ and $\underline{\mathrm{Q}}_{\mathrm{T}_{p\times 1}}$ are the respective column vectors for

$\underset{-}{Q}_{H\,j}$, $\underset{-}{Q}_{S\,j'}$, and $\underset{-}{Q}_{T\,j''}$. C is a $(bN_{F_2}+bN_{F_3}+p)\times(bN_{F_2}+bN_{F_3}+p)$ matrix obtained from

$$E\begin{pmatrix}\underset{-}{Q}_H\\ \underset{-}{Q}_S\\ \underset{-}{Q}_T\end{pmatrix}=C\underset{-}{t}$$

To obtain C let us now index the subscripts of the observations systematically so that the first set appears in the first block and the second set in the second block, and so on. Then, taking expectation of $\begin{pmatrix}\underset{-}{Q}_H\\ \underset{-}{Q}_S\\ \underset{-}{Q}_T\end{pmatrix}$, we obtain

$$C=\begin{pmatrix}H & A & B\\ A' & G & C\\ B' & C' & M\end{pmatrix},$$

Where

$$H=(I_{N_{F_2}}+\left(-\frac{1}{N}\right)1_{N_{F_2}}1'_{N_{F_2}})*I_b$$

$$G=(I_{N_{F_3}}+\left(-\frac{1}{N}\right)1_{N_{F_3}}1'_{N_{F_3}})*I_b$$

$$M=br\ \left(I_p-\frac{r}{N}\ 1_p1'_p\right)$$

$$A=\left(-\frac{1}{N}\right)1_{N_{F_2}}1'_{N_{F_3}}*I_b$$

$$B=\left(-\frac{r}{N}\right)1_{N_{F_2}}1'_p*1_b$$

$$C=\left(-\frac{r}{N}\right)1_{N_{F_3}}1'_p*1_b$$

$*$denotes "the Kronecker product" (see Marcus 1960)

It is easy to see that

Rank $C=b\ (N_{F_2}+N_{F_3})+p-1$

By imposing the restriction (3.2), we now solve (4.1) to obtain $\hat{\underset{-}{t}}$. Subtracting from each of last p rows of C, the row vector, $\left(-\frac{r}{N}\right)1_p(1'_{N_{F_2}}*1'_b, 1'_{N_{F_3}}*1'_b, 1'_p)$ (4.1) then becomes

$$C^*\hat{\underset{-}{t}}=\begin{pmatrix}\underset{-}{Q}_H\\ \underset{-}{Q}_S\\ \underset{-}{Q}_T\end{pmatrix}$$

where

$$C^*=\begin{pmatrix}H & A & B\\ A' & G & C\\ O & O & brI_p\end{pmatrix}$$

Now

$$(C^*)^{-1}=\begin{pmatrix} (I_{N_{F_2}}+\frac{1}{rp}1_{F_2}1'_{N_{F_2}})*I_b, & \frac{1}{rp}1_{N_{F_2}}1'_{N_{F_3}}*I_b, & \frac{1}{brp}1_{N_{F_2}}1'_p*1_b \\ \frac{1}{rp}1_{N_{F_3}}1'_{N_{F_2}}*I_b, & (I_{N_{F_3}}+\frac{1}{rp}1_{N_{F_3}}1'_{N_{F_3}})*I_b, & \frac{1}{brp}1_{N_{F_3}}1'_p*1_b \\ \underline{O} & \underline{O} & \frac{1}{rp}I_p \end{pmatrix}$$

So

$$\hat{\underline{t}}=(C^*)^{-1}\begin{pmatrix}\underline{Q}_H\\ \underline{Q}_S\\ \underline{Q}_T\end{pmatrix}$$

or

$$f_{N_{F_2}(i-1)+j}=Q_{HN_{F_2}(i-1)+j}+\frac{1}{rp}\sum_{j'=1}^{N_{F_2}}Q_{HN_{F_2}(i-1)+j'}+\frac{1}{rp}\sum_{j'=1}^{N_{F_3}}Q_{SN_{F_3}(i-1)+j'}$$

$$+\frac{1}{brp}\sum_{j''=1}^{p}Q_{Tj''},\quad 1\leqq i\leqq b,\quad 1\leqq j\leqq N_{F_2}$$

$$d_{N_{F_3}(i-1)+j'}=\frac{1}{rp}\sum_{j'=1}^{N_{F_2}}Q_{HN_{F_2}(i-1)+j}+Q_{SN_{F_3}(i-1)+j'}+\frac{1}{rp}\sum_{j'=1}^{N_{F_3}}Q_{SN_{F_3}(i-1)+j'}$$

$$+\frac{1}{brp}\sum_{j''=1}^{p}Q_{Tj''},\quad 1\leqq i\leqq b,\quad 1\leqq j'\leqq N_{F_3}$$

$$a_{j''}=\frac{1}{br}Q_{T_{j''}},\qquad 1\leqq j''\leqq p$$

The variance and covariance matrix of $\hat{\underline{t}}$ is given by

$$V_{\hat{\underline{t}}}=\begin{pmatrix} V_{11} & V_{12} & V_{13}\\ V_{21} & V_{22} & V_{23}\\ V_{31} & V_{32} & V_{33}\end{pmatrix}$$

where

$$V_{11}=(I_{N_{F_2}}+\frac{1}{rp}1_{N_{F_2}}1'_{N_{F_2}})*I_b+\left(-\frac{N+rp}{Nbrp}\right)1_{N_{F_2}}1'_{N_{F_2}}*1_b1'_b$$

$$V_{22}=(I_{N_{F_3}}+\frac{1}{rp}1_{N_{F_3}}1'_{N_{F_3}})*I_b+\left(-\frac{N+rp}{Nbrp}\right)1_{N_{F_3}}1'_{N_{F_3}}*1_b1'_b$$

$$V_{33}=\frac{1}{rp}\left(I_p-\frac{r}{N}1_p1'_p\right)$$

$$V_{12}=V'_{21}=\frac{1}{rp}1_{N_{F_2}}1'_{N_{F_3}}*I_b+\left(-\frac{N+rp}{Nbrp}\right)1_{N_{F_2}}1'_{N_{F_3}}*1_b1'_b$$

$$V_{13}=V'_{31}=(-\frac{1}{rp})1_{N_{F2}}1'_p*1_b$$

$$V_{23}=V'_{32}=\left(-\frac{1}{N_b}\right)1_{N_{F_3}}1'_b*1_b$$

Reexamination of Mather's Statistics

On replacing every individual of the F_2, F_3, and their parents and/or F_1, respectively, by their estimates, we obtain a set of observations clear of soil effects. We now obtain the corresponding second order statistics and their expectations as those given in Mather, based on this set of observations. Assuming that the errors e_{ij} are normally distributed, we have, from (4.2)

$$\hat{\underline{f}}\sim N(\underline{f},\sigma^2[(I_{N_{F_2}}+\frac{1}{rp}1_{N_{F_2}}1'_{N_{F_2}})*I_b+\left(-\frac{N+rp}{Nbrp}\right)1_{N_{F_2}}1'_{N_{F_2}}*1_b1'_b])$$

and

$$\hat{\underline{d}}\sim N(\underline{d},\sigma^2[(I_{N_{F_3}}+\frac{1}{rp}1_{N_{F_3}}1'_{N_{F_3}})*I_b+\left(-\frac{N+rp}{Nbrp}\right)1_{N_{F_3}}1'_{N_{F_3}}*1_b1'_b]).$$

Define now

$$V_{F_2(c)}=\frac{1}{bN_{F_2}-1}\sum_{i=1}^{bN_{F_2}}(\hat{f}_i-\bar{\hat{f}})^2=\frac{1}{bN_{F_2}-1}\hat{\underline{f}}'(I_{N_{F_2}}*I_b-\frac{1}{bN_{F_2}}1_{N_{F_2}}1'_{N_{F_2}}*1_b1'_b)$$

$$\bar{V}_{F_3(c)}=\frac{1}{ba}\sum_{i=1}^{ab}[\frac{1}{r-1}\sum_{j=1}^{r}(\hat{d}_{ij}-\bar{\hat{d}}_{i.})^2]=\frac{1}{ba(r-1)}\hat{\underline{d}}'\{[(I_r-\frac{1}{r}1_r1'_r)*I_a]*I_b\}\hat{\underline{d}}$$

$$V_{\bar{F}_3(c)}=\frac{1}{ba-1}\sum_{i=1}^{ba}(\bar{\hat{d}}_{i.}-\bar{\hat{d}}_{..})^2=\frac{1}{(ba-1)r^2}\hat{\underline{d}}'\{[(1_r1'_r)*I_a]*I_b$$

$$-\frac{1}{ba}(1_r1'_r*1_a1'_a)*1_b1'_b\}\hat{\underline{d}}$$

where $\hat{d}_{ij}$ is the jth $\hat{d}$ in the ith line.

$$\bar{\hat{d}}_{i.}=\frac{1}{r}\sum_{j=1}^{r}\hat{d}_{ij},\qquad \bar{\hat{d}}_{..}=\frac{1}{ba}\sum_{i=1}^{ba}\bar{\hat{d}}_{i.}$$

Then the expected value of $\bar{V}_{F_2(c)}$ is given by

$$E(V_{F_2(c)})=\frac{1}{2}D+\frac{1}{4}H+\left[1+\frac{N_{F_2}}{rp(bN_{F_2}-1)}(b-1)\right]\sigma^2 \tag{5.1}$$

Similary, the corresponding $\bar{V}_{F_3(c)}$ and $V_{\bar{F}_3(c)}$ will have expectations respectively given by

$$E(V_{\bar{F}_3(c)})=\frac{1}{2}D+\frac{1}{16}H+\left(\frac{p(ab-1)+a(b-1)}{rp(ba-1)}\right)\sigma^2 \tag{5.2}$$

and

$$E(\bar{V}_{F_3(c)})=\frac{1}{4}D+\frac{1}{8}H+\sigma^2 \tag{5.3}$$

From the analysis of variance table, we have the minimum variance unbiased estimator $\hat{\sigma}^2$ of σ^2, as given by

$$\hat{\sigma}^2 = \frac{SSE}{df_E} \tag{5.4}$$

Analysis of variance table

Sources of variation	Degrees of freedom	Sum of squares	Mean square	Expected value
Due to adjusted values of F_2, F_3 and parents and/or F_1	$df_1 = b(N_{F_2} + N_{F_3}) + p - 1$	$SS_1 = \sum \hat{f}_j Q_{Hj} + \sum \hat{d}_{j'} Q_{Sj'} + \sum \hat{a}_{j''} Q_{Tj''}$		
Blocks	$df_2 = b - 1$	$SS_2 = N \sum_{i=1}^{b} (\bar{B}_i - \bar{B})^2$		
Residual	$df_E = b(N - N_{F_2} - N_3) + p - 1$	$SS_E = SS_T - SS_1 - SS_2$	$\frac{SS_E}{Df_E}$	σ^2
Total	$bN - 1$	SST		

Using (5.1), (5.2), (5.3) and (5.4), we can obtain the estimates of D and H by applying the usual least square method, as given in Mather.

SUMMARY

1. In this paper another approach is undertaken to control the enviromental effects in the studies of the quantitative inheritance. Our approach differs from that of Mather in that, instead of computing E_1 and E_2, we first obtain the LUMV estimators of the genetic values of the F_2 and F_3 individuals by the conventional least square method; the corresponding statistics V_{F_2}, $V_{\bar{F}_3}$ and $\bar{V}_{F_3}$, etc, as those given in Mather, are then obtained to effect the partition of the genetic variances, based on this set of LUMV estimators.
2. In section (2) a design which is called APBIB in this paper is proposed for the studies of the quantitative inheritance. Specifically this design is a group-divisible PBIB with two associates for the F_2 and F_3 individuals combined with a randomized complete block design for the parents and/or F_1 individuals (called augmented lines). In essence this design is equivalent to that given in Mather, even though neither the nature nor the properties of this design have been revealed in Mather's book.

數量遺傳之最小二乘方統計分析法

譚 外 元　　袁 宸 宣

1. 在應用 Mather 氏之變方區分法以分析數量性狀之遺傳時常因地力變異較大而不能

予以有效的控制。本文茲提供另一方法以分析數量遺傳之資料。此項方法與 Mather 方法不同之點在於其不需計算 E_1 與 E_2，而直接由最小二乘方以求得 F_2 及 F_3 個體之線性不偏最佳 (變方最小) 估值。次將此等估值取代 F_2 及 F_3 個體之觀測值再估算 V_{F_2}，$V\bar{_{F_3}}$ 及 $\bar{V}_{F_3}$ 等，以進行變方成份之區分。

2. 在本文第二節中，作者特提供一試驗設計，應用於數量遺傳之研究，此設計在本質上係一可分組之二次 PBIB (對 F_2 及 F_3 個體而言)與一完全逢機區集(對親本及 F_1 而言)設計之組合，故特稱之為 APBIB。本文之分析即以此設計為基礎。

REFERENCES

K. MATHER Biometrical Genetics, Mathuen 1949

M. MARCUS Basic Theorems in Matrix theory. USDC, Applied Mathematies 57. 1960