

电子计算机 在遗传育种 中的应用

黄金龙 孙其信 编著
张爱民 黄铁城

电子计算机在遗传育种中的应用

黄金龙 孙其信 编著
张爱民 黄铁城

农业出版社

电子计算机在遗传育种中的应用

黄金龙 孙其信 编著
张爱民 黄铁城

* * *

责任编辑 徐建华

农业出版社出版 (北京朝阳区枣营路)

新华书店北京发行所发行 通县曙光印刷厂印刷

787×1092mm16开本 38印张 878千字

1991年10月第1版 1991年10月北京第1次印刷

印数 1—900 册 定价 23.50 元

ISBN 7-109-00598-4/S·456

前 言

随着育种理论研究的不断深入,育种学正在向理论学科发展。数量遗传学理论的应用是其重要的标志之一。

育种工作在教育数量遗传学理论时,首先遇到的问题关于及时、准确地处理大量试验数据的问题。由于计算手段,致使一些试验结果不能及时指导育种实践。而现代电子计算机的应用可有效地解决这类问题。

近年来,电子计算机在各行各业得到普及和应用,对我国科学技术的发展起到很大推动作用。本书编写的目的就在于使育种工作者能够借助于电子计算机这一有力工具,有效地将数量遗传等育种学的基础理论应用于育种实践。增加育种工作的预见性,提高育种效率,促进作物育种工作的发展。

本书在编写中尽力突出实用的特点。所介绍的各种方法都配有实例分析,既给出具体的统计分析步骤,又注重于实验结果的分析和解释以及在育种实践中的应用。计算机程序采用通用的BASIC语言编制并在Z-80微型机上实现。很容易移植到其它微型机上。程序编制方法上,采用覆盖技术,尽量节省内存,以便使用者可在微机有限内存里处理更多的数据。程序的内容以数量遗传方法为基础,又根据育种实际作了某些变通,可配合各种不同类型的试验设计。程序所涉及的方法不仅包括一般育种试验结果的统计和常用遗传参数的估计,而且包括近些年才发展起来的多元统计分析在育种中的应用。全书共有35块程序。

本书既可作为育种工作者的实用手册,又可作为各高等农业院校研究生、大学生的教科书或参考书,对其他从事遗传学、育种学研究的科技人员亦有一定的参考价值。

本书的编写是在北京农业大学蔡旭教授、吴仲贤教授指导和鼓励下进行的。编写过程中得到了北京农业大学农学系作物育种教研组、畜牧系动物数量遗传教研组及校计算机室的大力支持。北京师范学院郭平仲教授给予有益的指教。朱虹、王明理、张忠军、李素芳等同志帮助做了不少工作,在此一并致谢。

由于作者水平有限,时间较仓促,书中难免一些缺点和错误,请读者批评指正。

编著者

目 录

第一章 育种的数量遗传学基础知识	1
第一节 加性效应、显性效应和上位性效应	1
第二节 主要育种世代平均值的遗传组成	4
一、不分离世代的平均值组成	4
二、分离世代的平均值组成	4
第三节 各主要世代方差、协方差的遗传组成	7
一、不分离世代	7
二、 F_2 代	7
三、 F_3 代	8
四、 F_2 代个体与 F_3 家系平均值的协方差	10
五、回交世代	11
第二章 植物育种试验的方差分析和协方差分析	14
第一节 单因素单向分组方差分析	14
第二节 单因素双向分组的方差分析	18
一、小区内无重复观察值	18
二、小区内有重复观察值	23
第三节 单因素单向分组的协方差分析	27
第四节 单因素双向分组的协方差分析	29
一、小区内无重复观察值	29
二、小区内有重复观察值	32
第五节 遗传方差和遗传协方差的估计	35
一、单因素单向分组方差分析中遗传方差 σ_g^2 的估计	35
二、单因素双向分组方差分析中遗传方差 σ_g^2 的估计	35
三、单因素单向分组协方差分析中遗传协方差 $COV(g)$ 的估计	35
四、单因素双向分组协方差分析中遗传协方差 $COV(g)$ 的估计	36
第六节 程序使用说明及程序清单	36
一、单因素单向分组方差分析程序: ANOVA2·BAS	36
二、单因素双向分组方差分析程序一: ANOVA1·BAS	37
三、单因素双向分组方差分析程序二: ANOVA3·BAS	37
四、单因素双向分组协方差分析程序: COVAN·BAS	38
第三章 遗传力	48
第一节 遗传力及其估算原理和方法	48
一、遗传力的概念	48
二、遗传决定度和遗传力的估算	49
第二节 遗传力在育种上的应用	50
一、遗传力在育种上应用的几个方面	59
二、正确理解和应用遗传力	61
第三节 程序使用说明及程序清单	62

第四章 模型测验和世代平均值分析	70
第一节 模型测验和联合尺度测验	70
一、模型测验	70
二、联合尺度测验	74
第二节 世代平均值分析	77
一、六参数模式	77
二、五参数模式	81
三、三参数模式	82
第三节 程序使用说明及程序清单	82
一、简单尺度检验程序: ABCD·BAS	82
二、世代平均值分析程序: PAREST·BAS	83
三、联合尺度检验程序: JOINT·BAS	84
第五章 育种工作中的多元回归分析	95
第一节 多元线性回归	95
一、多元线性回归方程	95
二、多元线性回归方程的方差分析	99
三、复相关系数	100
四、多元线性回归分析在育种中的应用	100
第二节 逐步回归	101
第三节 程序使用说明及程序清单	107
第六章 相关和通径分析	115
第一节 遗传相关	115
一、表型相关、遗传相关和环境相关	115
二、遗传相关与间接选择	121
第二节 通径分析	122
一、通径和通径系数的概念	123
二、通径分析——通径系数及多元决定系数的计算	125
三、通径分析在遗传育种上的应用	130
第三节 程序使用说明及程序清单	132
一、简单相关分析程序: SINCOR·BAS	132
二、遗传相关分析程序: PGRA·BAS	132
三、通径分析程序: PATHCO·BAS	133
第七章 配合力分析	146
第一节 配合力的概念及配合力研究方法	146
一、配合力的概念	146
二、配合力研究方法	147
第二节 GRIFFING配合力分析方法	148
一、GRIFFING配合力分析的统计模型	148
二、原始数据的方差分析和配合力分析	149
三、方法1的统计分析	152
四、方法2的统计分析	162
五、方法3的统计分析	168
六、方法4的统计分析	176

第三节 两组亲本杂交的配合力分析	182
一、原始数据的方差分析	184
二、第一种情况的配合力分析	185
三、第二种情况的配合力分析	189
四、第三种情况的配合力分析	192
五、第四种情况的配合力分析	193
六、两组亲本杂交配合力分析方法举例	195
第四节 程序使用说明及程序清单	201
一、GRIFFING配合力分析方法1程序一: GRIFI-10	202
二、GRIFFING配合力分析方法1程序二: GRIFI-A10	205
三、GRIFFING配合力分析方法2程序一: GRIFI-20	206
四、GRIFFING配合力分析方法2程序二: GRIFI-B20	208
五、GRIFFING配合力分析方法3程序一: GRIFI-30	208
六、GRIFFING配合力分析方法3程序二: GRIFI-C30	209
七、GRIFFING配合力分析方法4程序一: GRIFI-40	210
八、GRIFFING配合力分析方法4程序二: GRIFI-D40	210
九、两组亲本杂交的配合力分析程序(1): PQCA-BAS	211
十、两组亲本杂交配合力分析程序(2): PQCAB-BAS	213
第八章 选择指数和相关遗传力	301
第一节 选择指数的基本理论	302
一、选择指数的构造	302
二、选择指数的遗传力及聚合遗传型的预期遗传进展	305
第二节 选择指数的分析方法及其应用	307
第三节 相关遗传力及其应用	315
一、相关遗传力的概念	316
二、相关遗传力的应用	316
第四节 程序使用说明及程序清单	324
一、改进单一目标性状的一般选择指数程序一: SELIND.COM	324
二、改进单一目标性状的一般选择指数程序二: INDEX3.COM	326
三、改进聚合遗传型的综合选择指数程序一: INDEX.COM (适用于重复次数 $R \neq 1$ 的试验)	327
四、改进聚合遗传型的综合选择指数程序二: INDEX5.COM [适用于不设重复($R = 1$)的试验]	328
五、相关遗传力分析程序: CORHH.COM	329
第九章 区域试验的方差分析及品种稳定性	376
第一节 品种区域试验的方差分析	376
一、一年一地试验(品种比较试验)	376
二、一年多地试验	377
三、多年多地试验	383
第二节 品种的稳定性	393
一、EBERHART和RUSSELL稳定性分析模型	393
二、稳定性参数估计	393
三、稳定性参数的显著性检验	395
四、稳定性分析举例	397

第三节 程序使用说明及程序清单	402
一、一年多地试验方差分析程序: AOVFR1.COM	402
二、多年多地试验方差分析程序: AOVFR2.COM	403
三、稳定性分析程序: EBERHE.RT	405
第十章 遗传距离测定及其在育种中的应用	433
第一节 遗传距离的概念	433
第二节 主成分分析	435
一、主成分分析的数学描述	435
二、主成分的性质	438
三、主成分的计算	439
第三节 遗传距离测定	439
一、数量性状的选择和调查方法	439
二、方差—协方差分析	439
三、遗传相关矩阵的估算	439
四、计算特征方程 $(R - \lambda I) L' = 0$ (即 $(\Sigma - \lambda I) L' = 0$) 的特征根和特征向量	440
五、计算主成分值	440
六、求性状与主成分的相关系数及性状对主成分的给予率	441
七、计算遗传距离	441
第四节 聚类分析	446
一、系统聚类简介	447
二、8种常用的类间距离的计算方法	448
三、利用谱系图对样本进行分类	452
四、系统分类法的比较	454
第五节 遗传距离应用中的几个问题	455
一、性状的数量化	455
二、性状的选择与调查	455
三、方法问题	455
第六节 遗传距离与聚类分析在育种中的应用	456
一、研究亲本选配和亲本分类	456
二、研究环境分类	460
第七节 程序使用说明及程序清单	460
一、主成分遗传距离估测程序一: PAV.COM	460
二、主成分遗传距离估测程序二: PCH1.COM	461
第十一章 典范相关分析及其在遗传育种中的应用	498
第一节 典范相关分析	498
第二节 典范性状及典范遗传距离	505
一、典范性状及其遗传力	505
二、典范遗传距离	508
三、应用举例	509
第三节 典范选择指数	515
一、间接遗传系数	515
二、典范选择指数	516
第四节 程序使用说明及程序清单	521

一、典范性状及典范遗传距离程序: PAV33.COM	521
二、典范选择指数程序: CANOAN.COM	522
三、典范选择指数程序: CANOSE.COM	525
附录: 矩阵代数	571
附表1 学生氏t值表(两尾)	585
附表2 5%(上)和1%(下)点F值(一尾)表	586
附表3 χ^2 值表(一尾)	592
附表4 Duncan's新复极差测验5%和1%SSR值表	593
附表5 选择率与选择强度的关系	595
参考文献	596

第一章 育种的数量遗传学基础知识

数量遗传学是研究动植物数量性状遗传变异规律的科学。它与育种工作有着密切的联系，是育种工作的重要理论基础。因为动植物的许多性状(尤其是经济性状)多为数量性状，从事动植物新品种的选育，以及现有品种的改良，都需要了解这些重要经济性状的遗传变异规律，以制定合理有效的育种方案。

数量性状都表现为连续性变异，这种连续性变异遗传上是由微效多基因所控制的。所谓“微效多基因”是指控制数量性状表现的基因数目较多，而每个基因的效应较小，且基因效应具有可加性。正是由于“多基因”的“微效”效应造成了数量性状的连续性变异。

数量性状在杂交后的分离世代表现为连续性变异，在其群体中很难找到如质量性状那样明确的分组界限，因此，要研究其遗传特征，就必须采用生物统计学的方法。在数量性状的遗传研究中采用平均数和方差等统计量来描述数量性状的遗传变异规律，也就是说对于一个数量性状的遗传变异情况是从平均数和方差等方面加以讨论。

为了便于理解以后各章的内容，这一章我们将介绍一些育种中常常遇到的数量遗传学的基本概念，作为阅读以后章节的准备知识。

第一节 加性效应、显性效应和上位性效应

在数量遗传学中，把任何个体的数量性状的观测值称为该性状的表型值，用 P 表示。决定某个体数量性状表型的因素除了个体本身的基因型(或遗传型)外，还有个体所处的环境条件。由个体基因型所决定的每一个体的数值部分称为基因型值。用 G 来表示。而环境条件所引起的表型值偏离基因型值的部分称为环境偏差，用 E 表示，即 $E = P - G$ 。因此任一个体数量性状的表型值可以看成是基因型值和环境偏差的代数和，即

$$P = G + E \quad (1.1)$$

1.1式将数量性状的表现型剖分为基因型值和环境偏差两部分。此式称为数量性状的一般遗传模型。一般认为环境偏差 E 服从正态分布：平均数为零，方差为 σ^2 ，即 $E \sim N(0, \sigma^2)$ 。根据 E 的特性，当研究群体足够大时， $\bar{E} = 0$ ，而 $\frac{1}{n} \sum P = G$ ，群体的平均的表型值就等于如上所述基因型值，即

$$\bar{P} = G \quad (1.2)$$

1.1式的表型值可以剖分为基因型值和环境偏差两部分，但仅仅这样的剖分还不能为遗传育种工作提供足够的信息。实际上基因型值还可以进一步剖分。

为了研究基因型值的进一步剖分，我们考虑一对等位基因 $A-a$ ，有3种可能的基

因型, 即AA, Aa, aa。它们的基因型值分别用 G_{AA} , G_{Aa} , G_{aa} 来表示。设两种纯合子AA和aa的平均值为 m , 即 $m = \frac{1}{2}(G_{AA} + G_{aa})$, 则这3种基因型的表现可用图1.1表示。

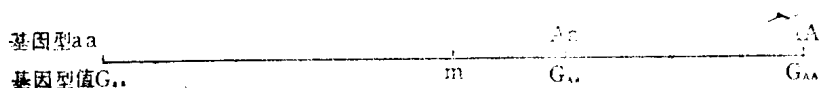


图1.1 单个位点情况下3种基因型的表现

从图1.1看出, m 位于两种纯合基因型中间, 因而又称为中亲值。现在用 a 表示基因型AA与中亲值的差异, 即 $a = G_{AA} - m$, 那么, 基因型aa与中亲值 m 的差异即为 $-a$, 即 $G_{aa} - m = -a$, 又用 d 表示基因型Aa与中亲值 m 的差异, 即 $d = G_{Aa} - m$ 。我们把 a 称为基因的加性效应 (additive effect), d 称为基因的显性效应 (dominant effect)。

在图1.1中, 基因型Aa位于 m 和AA之间, 即 $0 < d < a$, 这时, 基因A对a为部分显性; 而当 $d = a$ 时, 基因A为完全显性; $d = 0$ 时说明没有显性存在。基因A能使基因型值向增大的方向变化。在这种情况下把基因A称为增值基因, 而基因a称为减值基因 (a使基因型值向变小的方向变化)。

从以上讨论可以看出, 显性效应实际上是指基因位点内的互作, 即同一位点内二个等位基因的互作。如果位点间基因的效应是可加的, 即位点间不存在互作, 那么, 当有两对等位基因时, 如A-a和B-b, 则基因型AABB的效应值应该是基因A的效应 a_A 与基因B的效应 a_B 之和。例如, 得到2对基因的4种纯合基因型及其基因型值如下,

基因型	型值
AABB	95
aaBB	90
AAbb	39
aabb	34

由以上数字, 我们可以求出:

在BB背景下,

$$2a_A = G_{AABB} - G_{aaBB} = 95 - 90 = 5$$

或在bb背景下,

$$2a_A = G_{AAbb} - G_{aabb} = 39 - 34 = 5$$

所以, $a_A = 2.5$

$$2a_B = G_{AABB} - G_{AAbb} = 95 - 39 = 56$$

或

$$2a_B = G_{aaBB} - G_{aabb} = 90 - 34 = 56$$

所以, $a_B = 28$

如果位点间效应是可加的, 即位点间不存在互作, 则各种基因型的基因型值可按下式计算:

$$m = \frac{1}{4} (95 + 90 + 39 + 34) = 64.5$$

$$G_{AABb} = m + a_A + a_B = 64.5 + 2.5 + 28 = 95$$

$$G_{aaBB} = m - a_A + a_B = 64.5 - 2.5 + 28 = 90$$

$$G_{AAbb} = m + a_A - a_B = 64.5 + 2.5 - 28 = 39$$

$$G_{aabb} = m - a_A - a_B = 64.5 - 2.5 - 28 = 34$$

可以看出，由上式计算出的基因型值与上面各基因型的实际观测值完全一致，说明位点间的效应是可加的，位点间不存在互作效应。但是，如果再看下面的例子，同样有两对等位基因A-a和B-b，4种纯合基因型及基因型值如下：

基因型	型值	基因型	型值
AABB	24	aaBB	12
AAbb	12	aabb	8

从表中可以看出，在BB背景下，

$$a_A = \frac{1}{2} (G_{AABb} - G_{aaBB}) = \frac{1}{2} (24 - 12) = 6$$

而在bb背景下，

$$a_A = \frac{1}{2} (G_{AAbb} - G_{aabb}) = \frac{1}{2} (12 - 8) = 2$$

同样，在AA背景下和在aa背景下求得的基因型B的效应也有所不同，在AA背景下，

$$a_B = \frac{1}{2} (G_{AABb} - G_{AAbb}) = \frac{1}{2} (24 - 12) = 6$$

而在aa背景下，

$$a_B = \frac{1}{2} (G_{aaBB} - G_{aabb}) = \frac{1}{2} (12 - 8) = 2$$

如果以 $a_A = 6$ ， $a_B = 6$ ， $m = \frac{1}{4} (24 + 12 + 8 + 12) = 14$ 来计算各基因型的型值，则

有，

$$G_{AABb} = 14 + 6 + 6 = 26$$

$$G_{AAbb} = 14 + 6 - 6 = 14$$

$$G_{aaBB} = 14 - 6 + 6 = 14$$

$$G_{aabb} = 14 - 6 - 6 = 2$$

由上面计算的各基因型的型值与表中实际观测情况并不符合，这说明位点A-a和B-b间的效应并不是简单的可以相加的，而是位点间存在互作，这种互作效应称为上位性效应 (epistatic effect)，一般用*i*表示。

实际上，就数量性状而论，我们并不能够把造成连续性变异的各个基因的效应区分开来。如果有2个纯系品种，每个系与中亲值的差异实际上反映了影响该性状表现的所

有基因共同作用的结果。假定这些基因的效应都是简单可加的，则每个系与中亲值的差异实际上就是各基因加性效应 a 的总和，而且 a 可能有正有负。例如二个亲本分别为AABBCC和aabbcc，则前者基因型值 G_{AABBCC} 与中亲值的偏差为 $a_A + a_B + a_C$ ，后者 G_{aabbcc} 与中亲值的偏差为 $-(a_A + a_B + a_C)$ ，而如果纯系品种分别为AAbbcc和aaBBcc，则前者 G_{AAbbcc} 与中亲值的偏差为 $a_A - a_B + a_C$ ，而后者 G_{aaBBcc} 与中亲值偏差为 $-(a_A - a_B + a_C)$ ，如果有 m 个位点控制数量性状的表现，我们用 $[a]$ 表示总的加性效应， $[d]$ 表示总的显性效应， $[i]$ 表示总的上位性效应。因此，数量性状表型值的遗传模型可以进一步表示为

$$P = m + g + E = m + [a] + [d] + [i] + E \quad (1.3)$$

注意，这里以 g 表示基因型效应，即 $g = G - m$ ，而1.1式中的 G 为基因型值。

以上我们将基因型值进一步剖分为加性效应，显性效应和上位性效应，这在遗传育种研究中更具有实际意义。在育种中，从杂种后代进行选择以及决定是利用纯系品种还是利用杂交种时，都会考虑到不同的基因效应。

第二节 主要育种世代平均值的遗传组成

我们将以一对等位基因为例，按1.2式的表型值遗传模型来说明各世代平均值的遗传组成，在计算时假定上位性效应 i 不存在，且加性效应 a 与显性效应 d 相互独立。

一、不分离世代的平均值组成

从图1.1可以看出亲本 $P_1(AA)$ ， $P_2(aa)$ ，和 F_1 的平均值(用 $\bar{P}_1$ 、 $\bar{P}_2$ 和 $\bar{F}_1$ 表示)分别为：

$$\bar{P}_1 = m + a$$

$$\bar{P}_2 = m - a$$

$$\bar{F}_1 = m + d$$

二、分离世代的平均值组成

(一) F_2 代

对于一对基因来说， F_2 代的基因型及其相应的频率为：

基因型	频率	基因型值
AA	$\frac{1}{4}$	$m + a$
Aa	$\frac{1}{2}$	$m + d$
aa	$\frac{1}{4}$	$m - a$

所以后代平均值的遗传组成为：

$$\bar{F}_2 = \frac{1}{4}(m + a) + \frac{1}{2}(m + d) + \frac{1}{4}(m - a)$$

$$= m + \frac{1}{2}d$$

(二) F_3 代

F_2 代个体自交产生 F_3 代，其基因型组成为：

$$\frac{1}{4}AA + \frac{1}{2}\left(\frac{1}{4}AA + \frac{1}{2}Aa + \frac{1}{4}aa\right) + \frac{1}{4}aa = \frac{3}{8}AA + \frac{1}{4}Aa + \frac{3}{8}aa$$

F_3 代的基因型及其相应的频率为：

基因型	频率	型值
AA	$\frac{3}{8}$	$m + a$
Aa	$\frac{1}{4}$	$m + d$
aa	$\frac{3}{8}$	$m - a$

因此， F_3 代的平均值遗传组成为：

$$\begin{aligned}\bar{F}_3 &= \frac{3}{8}(m + a) + \frac{1}{4}(m + d) + \frac{3}{8}(m - a) \\ &= m + \frac{1}{4}d\end{aligned}$$

(三) F_n 代

我们知道每自交一代杂合体减少一半，即 F_2 代时杂合体的频率为 $\frac{1}{2}$ ， F_3 代的杂合体的频率为 $(\frac{1}{2})^2$ ，而 F_n 代则为 $(\frac{1}{2})^{n-1}$ 。而 F_n 代时纯合体比例为 $[1 - (\frac{1}{2})^{n-1}]$ ，每一种纯合体的频率为 $\frac{1}{2}[1 - (\frac{1}{2})^{n-1}]$ ，因此，可以得到不同世代平均值遗传组成的一般公式为：

$$\bar{F}_n = m + \left(\frac{1}{2}\right)^{n-1}d$$

(四) 回交一代

有2个回交一代，设以 B_1 表示 $F_1 \times P_1$ ，以 B_2 表示 $F_1 \times P_2$ ， B_1 世代的基因型及其频率为：

$$\begin{array}{c}Aa \times AA \\ \downarrow \\ B_1: \frac{1}{2}AA + \frac{1}{2}Aa\end{array}$$

因此， B_1 世代平均值的遗传组成为：

$$\bar{B}_1 = \frac{1}{2}(m + a) + \frac{1}{2}(m + d)$$

$$= m + \frac{1}{2}a + \frac{1}{2}d$$

B₂世代基因型及其频率为:

$$\begin{array}{c} Aa \times aa \\ \downarrow \\ B_2: \frac{1}{2}Aa + \frac{1}{2}aa \end{array}$$

因此, B₂世代平均值的遗传组成为:

$$\begin{aligned} \bar{B}_2 &= \frac{1}{2}(m-a) + \frac{1}{2}(m+d) \\ &= m - \frac{1}{2}a + \frac{1}{2}d \end{aligned}$$

上述对于一对等位基因所得到的平均值遗传组成公式可推广到多个位点的情况, 这时可用总效应[a]、[d]代替一个位点的效应a和d, 即可得到相应世代平均值的遗传组成:

$$\begin{aligned} \bar{P}_1 &= m + [a] & \bar{P}_2 &= m - [a] \\ \bar{F}_1 &= m + [d] & \bar{F}_2 &= m + \frac{1}{2}[d] \\ \bar{F}_3 &= m + \frac{1}{4}[d] & \bar{F}_n &= m + \left(\frac{1}{2}\right)^{n-1}[d] \\ \bar{B}_1 &= m + \frac{1}{2}[a] + \frac{1}{2}[d] \\ \bar{B}_2 &= m - \frac{1}{2}[a] + \frac{1}{2}[d] \end{aligned}$$

为方便起见, 表1—1列出了各主要育种世代平均值的遗传组成, 表中数值为各效应的系数。

表1—1 各主要育种世代平均值的遗传组成

世 代	基 因 型 值		
	m	[a]	[d]
P ₁	1	1	0
P ₂	1	-1	0
F ₁	1	0	1
F ₂	1	0	$\frac{1}{2}$
B ₁	1	$\frac{1}{2}$	$\frac{1}{2}$
B ₂	1	$-\frac{1}{2}$	$\frac{1}{2}$
F ₃	1	0	$\frac{1}{4}$
F ₄	1	0	$\frac{1}{8}$
F _n	1	0	$\left(\frac{1}{2}\right)^{n-1}$

第三节 各主要世代方差、协方差的遗传组成

一、不分离世代

纯系亲本 P_1 和 P_2 及其 F_1 代都是基因型一致的群体，通常称为不分离世代。考虑1.1式中，在基因型值 G 和环境偏差 E 相互独立时，表型方差为基因型值的方差和环境偏差的方差（或分别简称为基因型方差和环境方差）之和，即

$$V_P = V_G + V_E \quad (1.3a)$$

P_1 、 P_2 和 F_1 三个群体各自都具有同样的基因型，所以其基因型方差为0，它们的表型方差 V_{P_1} 、 V_{P_2} 、 V_{P_3} 和 V_F 就等于环境方差，所以，在估计基因型方差时，不分离世代的表型方差可以作为分离世代环境方差的估值。

在1.2式中，基因型效应又进一步分解为加性效应，显性效应和上位性效应，因此由基因的加性效应引起的方差称为加性方差（additive variance），由基因的显性效应引起的方差称为显性方差（dominant variance），而由上位性效应引起的方差称为上位性方差（epistatic variance）或互作方差（interaction variance）。加性方差是可以固定遗传的，而显性方差则不能。在下面关于各分离世代方差的讨论中，我们先以一对基因为例，并假定位点间互作不存在。

二、 F_2 代

如果有两个亲本 $P_1(AA)$ 和 $P_2(aa)$ 杂交， F_1 为 Aa ，则 F_2 代群体的基因型，基因型频率及型值如表1—2。

表1—2 计算 F_2 代群体方差

基 因 型	型 值	频 率	(型值) ²	(型值) ² ×频率
AA	a	$\frac{1}{4}$	a^2	$\frac{1}{4}a^2$
Aa	d	$\frac{1}{2}$	d^2	$\frac{1}{2}d^2$
aa	-a	$\frac{1}{4}$	a^2	$\frac{1}{4}a^2$
总		1		$\frac{1}{2}a^2 + \frac{1}{2}d^2$

根据方差的计算公式，

$$\begin{aligned} V &= \sum_i (x_i - \bar{x})^2 f_i \\ &= \sum_i f_i x_i^2 - \bar{x}^2 \end{aligned} \quad (1.3b)$$

式中 f_i 为 x_i 出现的频率，而 F_2 群体的平均数 $\bar{x}$ 为 $(\frac{1}{4}a + \frac{1}{2}d - \frac{1}{4}a) = (\frac{1}{2}d)$ 。

按表1—2， F_2 代群体基因型方差为：

$$\begin{aligned}
 V_{(1G)F_2} &= \frac{1}{4}a^2 + \frac{1}{2}d^2 + \frac{1}{4}a^2 - \left(\frac{1}{2}d\right)^2 \\
 &= \frac{1}{2}a^2 + \frac{1}{2}d^2 - \frac{1}{4}d^2 \\
 &= \frac{1}{2}a^2 + \frac{1}{4}d^2
 \end{aligned} \tag{1.4}$$

如果有多个位点，且位点间的效应是可加的，则：

$$V_{(1G)F_2} = \frac{1}{2}\sum_i a_i^2 + \frac{1}{4}\sum_i d_i^2 \tag{1.5a}$$

式中 $\sum a^2$ 为加性方差， $\sum d^2$ 为显性方差，如果用D表示加性方差 $\sum a^2$ ，H表示显性方差 $\sum d^2$ ，则1.5a式变为

$$V_{(1G)F_2} = \frac{1}{2}D + \frac{1}{4}H \tag{1.5b}$$

在 F_2 代的表型方差中，除基因型方差外，还包括环境方差，现在用E表示 F_2 代的环境方差，则 F_2 代的表型方差为

$$V_{1F_2} = \frac{1}{2}D + \frac{1}{4}H + E \tag{1.6}$$

三、 F_3 代

由 F_2 代个体自交产生 F_3 家系， F_3 代方差计算如下。

表1—3 F_3 家系方差的计算

F_2 基因型	F_2 频率	F_3 家系平均数 x_i	F_3 家系频率 f_i	$x_i^2 \times f_i$	F_3 家系方差*
AA	$\frac{1}{4}$	a	$\frac{1}{4}$	$\frac{1}{4}a^2$	0
Aa	$\frac{1}{2}$	$\frac{1}{2}d$	$\frac{1}{2}$	$\frac{1}{2}\left(\frac{1}{2}d\right)^2$	$\frac{1}{2}a^2 + \frac{1}{4}d^2$
aa	$\frac{1}{4}$	-a	$\frac{1}{4}$	$\frac{1}{4}(-a)^2$	0
总			1	$\frac{1}{2}a^2 + \frac{1}{8}d^2$	

* 由 F_2 的个体aa和AA产生的后代都是不分离世代，其遗传方差等于零，而由 F_2 的Aa个体产生的 F_3 家系，其方差就等于 F_2 代的方差，即为 $\frac{1}{2}a^2 + \frac{1}{4}d^2$ 。

从上表及 F_3 家系的平均值 $\frac{1}{4}d$ ，可得 F_3 家系平均值的遗传方差 $V_{(1G)F_3}$

$$\begin{aligned}
 V_{(1G)F_3} &= \sum_i f_i x_i^2 - \bar{x}^2 \\
 &= \frac{1}{4}a^2 + \frac{1}{2}\left(\frac{1}{2}d\right)^2 + \frac{1}{4}(-a)^2 - \left(\frac{1}{4}d\right)^2
 \end{aligned}$$