

第8章

随机交配群体的遗传分析

王建康

中国农业科学院作物科学研究所

wangjiankang@caas.cn

<http://www.isbreeding.net>

本章的主要内容

- § 8.1 随机交配群体中遗传效应的分解
- § 8.2 随机交配群体的遗传方差和亲子相关
- § 8.3 上位性互作模型的遗传方差分解
- § 8.4 亲属间协方差的一般表示与遗传力估计

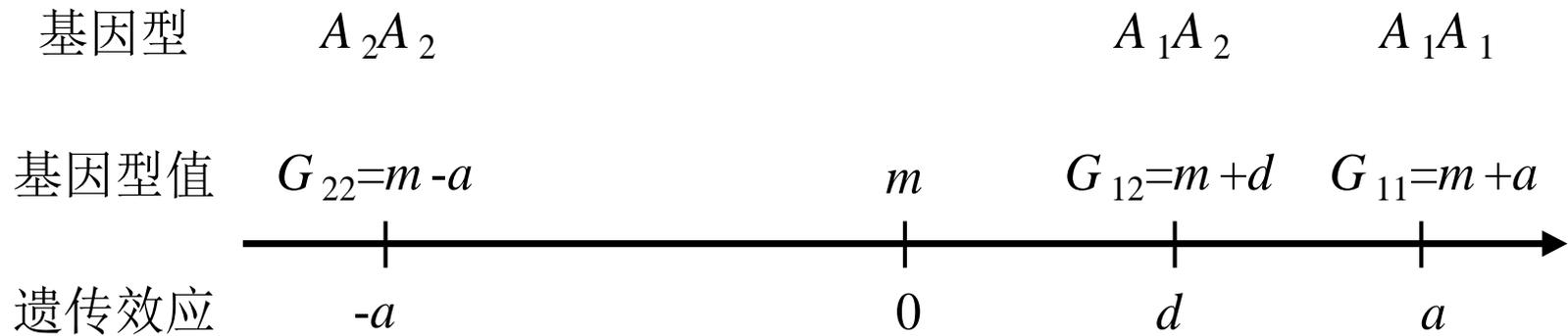
§ 8.1 随机交配群体中遗传效应的分解

- § 8.1.1 表型值和基因型值的分解
- § 8.1.2 等位基因的平均效应
- § 8.1.3 基因替代效应及其回归解释
- § 8.1.4 育种值和显性离差

表型值和基因型值的分解

- 为了便于推广到复等位基因的情形，这一章中一个座位上的等位基因用 A_1 、 A_2 ...表示。基因型 A_iA_j 的个体，在一定环境下的表型值 P_{ij} (Phenotypic value) 是基因型值 G_{ij} (Genotypic value) 和随机环境效应 ε_{ij} 共同作用的结果。
- 可重复的环境效应、以及基因型和宏环境之间的互作，将在第9章介绍。这里我们假定，基因型值是一个未知参数，但可以利用同一基因型多次观测值的平均数进行估计。因此，可将基因型值看作是众多表型构成的一个总体的均值，有时也称为表型均值或表型平均数。

单基因座位加显性效应模型



$$m = \frac{1}{2}(G_{11} + G_{22}) \quad a = \frac{1}{2}(G_{11} - G_{22}) \quad d = G_{12} - m$$

$$G_{11} = m + a \quad G_{12} = m + d \quad G_{22} = m - a$$

- 如果把 m 所在位置看作是坐标原点， A_2A_2 的基因型值相当于在原点基础上减去加性效应 a ， A_1A_2 的基因型值相当于在原点基础上加上显性效应 d ， A_1A_1 的基因型值相当于在原点基础上加上加性效应 a 。与双亲群体的加显性模型一样，图8.1中的加显性效应可正、可负，显性度 d/a 可以有各种可能的取值。

随机交配群体的均值和方差

- 用 p 和 q 分别表示等位基因 A_1 和 A_2 的频率，那么三种基因型 A_1A_1 、 A_1A_2 和 A_2A_2 在Hardy-Weinberg平衡群体中的频率分别为 p^2 、 $2pq$ 和 q^2 。于是，随机交配群体的均值和遗传方差就可分别为：

$$\begin{aligned}\mu &= E(P) = E(G) = p^2 G_{11} + 2pq G_{12} + q^2 G_{22} \\ &= m + (p - q)a + 2pqd\end{aligned}$$

$$\begin{aligned}V_G = V(G) &= p^2 (G_{11} - \mu)^2 + 2pq (G_{12} - \mu)^2 + q^2 (G_{22} - \mu)^2 \\ &= p^2 G_{11}^2 + 2pq G_{12}^2 + q^2 G_{22}^2 - \mu^2\end{aligned}$$

随机交配群体的基因型效应

- 如用随机交配群体的平均数 μ 对各种基因型值进行矫正，对于 A_iA_j 的基因型值，如公下面的公式定义遗传效应，即基因型值与群体均值的离差。

$$G_{ij} = \mu + g_{ij} \quad g_{ij} = G_{ij} - \mu$$

- 这样定义的遗传效应 则是一个群体水平的遗传参数，其取值依赖于等位基因频率，因此包含了群体遗传构成的信息；同时，遗传效应 还满足加权平均数等于0的约束条件。从遗传效应的遗传方差计算公式为：

$$V_G = p^2 g_{11}^2 + 2pqg_{12}^2 + q^2 g_{22}^2$$

老鼠矮化基因座位上三种基因型的6周龄平均体重（g），以及等位基因频率分别为0.9和0.1的随机交配群体的基因型频率、群体均值、群体方差和三种基因型的遗传效应

基因型	++	+pg	pgpg
6周龄平均体重（g）， 即基因型值	14	12	6
等位基因+和pg频率分别为0.9和0.1的随机交配群体			
基因型频率	0.81	0.18	0.01
群体均值	13.56		
遗传效应	0.44	-1.56	-7.56
遗传方差	1.1664		

随机交配群体中后代与亲代的关系

- 在随机交配的情形下，亲本传递给后代的是它携带的一个等位基因、而不是它的基因型。对于存在两个等位基因 A_1 和 A_2 的一个座位，个体 A_1A_1 的每个随机交配后代，只得到该个体的一个等位基因，即 A_1 ；后代的另一个等位基因来自另外一个亲本，它既可能是 A_1 、也可能是 A_2 ；个体 A_1A_1 随机交配后代的基因型有 A_1A_1 和 A_1A_2 两种可能；每种可能的基因型中，只有一个等位基因来自亲本 A_1A_1 ，另一个来自其它亲本。
- 在大群体中，这两个基因的后裔同样概率等于0。随机交配后代的基因型与亲代的基因型尽管有一定联系，但又不完全等同于亲代的基因型。因此，基因型为 A_iA_j 的亲代个体不能把它的所有遗传效应 g_{ij} 都传递给下一代。这就需要对遗传效应作进一步分解，以了解性状在亲子之间的传递规律，并预测选择对后代群体的影响。

等位基因的平均效应

- 一种定义等位基因效应的方法，是利用后代群体的平均表现与随机交配群体均值的离差进行计算。
- 以等位基因 A_1 为例，把它视为配子，与群体中其他配子随机结合产生一个后代群体，其他配子基因型既有 A_1 也有 A_2 ，它们的频率分别为 p 和 q 。因此，配子 A_1 产生后代群体中的基因型有 A_1A_1 和 A_1A_2 两种，频率也分别为 p 和 q 。根据配子 A_1 后代群体的基因型频率，就能得到后代群体的均值为 $pa+qd$ ，从中减去随机交配群体的均值 μ ，就得到等位基因 A_1 的效应。
- 类似地，我们还可以得到等位基因 A_2 平均效应。对于复等位基因，可用同样的方法定义它们的平均效应。

等位基因平均效应的计算

等位基因和频率	基因型值及后代群体中的频率			后代群体均值	等位基因效应 α_i
	A_1A_1, a	A_1A_2, d	$A_2A_2, -a$		
A_1, p	p	q		$pa + qd$	$q[a + (q - p)d]$
A_2, q		p	q	$pd - qa$	$-p[a + (q - p)d]$
随机交配群体	p^2	$2pq$	q^2	$(p - q)a + 2pqd$	

$$\alpha_1 = pa + qd - [(p - q)a + 2pqd] = q[a + (q - p)d]$$

$$\alpha_2 = -qa + pd - [(p - q)a + 2pqd] = -p[a + (q - p)d]$$

$$p\alpha_1 + q\alpha_2 = 0$$

基因替代效应

- 育种过程中，当选择有利于某个等位基因时，常意味着有利等位基因对另一个不利等位基因的替换。因此，有必要研究等位基因的替代效应（effect of an allele substitution）。
- 假定我们可以把随机挑选的等位基因 A_2 变为 A_1 ，中选个体的基因型可能是 A_1A_2 也可能是 A_2A_2 ，频率分别为 p 和 q 。把 A_1A_2 变为 A_1A_1 后，基因型值从 d 变为 a ，替换前后的效应变化为 $a-d$ ；把 A_2A_2 变为 A_1A_2 后，基因型值从 $-a$ 变为 d ，替换前后的效应变化为 $a+d$ 。因此得到平均基因替换效应的表达式。

基因替代效应与基因平均效应的关系

替换前		替换后		替换前后的基因型值变化量	两种替换基因型的频率
基因型	基因型值	基因型	基因型值		
A_1A_2	d	A_1A_1	a	$a-d$	p
A_2A_2	$-a$	A_1A_2	d	$d+a$	q

$$\alpha = p(a - d) + q(d + a) = a + (q - p)d$$

$$\alpha_1 = q\alpha \quad \alpha_2 = -p\alpha \quad \alpha = \alpha_1 - \alpha_2$$

老鼠体重的例子

- 例如，在前面老鼠体重的例子中，若一个随机交配群体中基因 pg 的频率 $q=0.1$ ，两种等位基因的平均效应以及等位基因的替代效应如下。

$$\alpha_+ = q[a + (q - p)d] = 0.1 \times [4 + 2 \times (0.1 - 0.9)] = 0.24$$

$$\alpha_{pg} = -p[a + (q - p)d] = -0.9 \times [4 + 2 \times (0.1 - 0.9)] = -2.16$$

$$\alpha = [a + (q - p)d] = [4 + 2 \times (0.1 - 0.9)] = 2.40$$

基因替代效应的回归解释

- 回归分析对基因替代效应给出一个更为明确的解释，并且有利于理解基因型的育种值和显性离差分解。三种基因型值和等位基因 A_1 的数目列于下表：

基因型	频率 f_i	基因型值 Y_i	基因 A_1 的数目 X_i
A_1A_1	p^2	a	2
A_1A_2	$2pq$	d	1
A_2A_2	q^2	$-a$	0
平均	1	$(p - q)a + 2pqd$	$2p$

基因替代效应的回归解释

- 把基因型值视为依变量 Y ，等位基因 A_1 的数目视为因变量 X 。 Y 的加权平均数当然就是随机交配群体的均值， X 的加权平均数等于 $2p$ 。因此得到如下 X 与 Y 之间的协方差及变量 X 的方差。

$$\text{Cov}(X, Y) = \sum f_i X_i Y_i - E(X)E(Y) = 2pq[a + (q - p)d]$$

$$V(X) = \sum f_i X_i^2 - [E(X)]^2 = 2pq$$

- 基因型值 Y 对 A_1 等位基因个数 X 的回归系数和截距为

$$b_{YX} = \frac{\text{Cov}(X, Y)}{V(X)} = a + (q - p)d = \alpha \quad b_0 = \bar{Y} - b_{YX} \bar{X} = \mu - 2p\alpha$$

育种值的定义

- 育种中对亲本的选择，当然是希望中选亲本的后代能够有更加优良的表现。因此在育种中，育种家往往从一个个体所产生后代群体的平均表现去判断它的育种价值，并将其作为亲本选择的重要标准。一个个体的后代群体，相对于整个亲本群体的差异，称为该个体的育种值（breeding value）。
- 前面介绍的基因效应和基因替代效应在实际中难以测量，而育种值是可以测量的。如果一个个体与一个群体中随机挑选的其它个体进行交配，该个体的育种值则定义为该个体后代群体的平均表现与随机交配群体的平均表现之间差异的2倍。之所以要加倍，是由于一个亲本只提供给后代群体一半的基因。

育种值的计算

- 下表给出三种基因型作为亲本，它们的后代均值及其与随机交配群体均值之间离差。表中的后代群体也称半同胞家系。如在一个玉米随机交配群体中，每个植株上收获的种子就可被看成是一个半同胞家系。

亲代基因型	后代群体中的基因型频率			后代群体均值	与随机交配群体均值的离差
	A_1A_1, a	A_1A_2, d	$A_2A_2, -a$		
A_1A_1	p	q		$pa + qd$	$\alpha_1 = \frac{1}{2}A_{11}$
A_1A_2	$\frac{1}{2}p$	$\frac{1}{2}(p + q)$	$\frac{1}{2}q$	$\frac{1}{2}(pa + qd) + \frac{1}{2}(pd - qa)$	$\frac{1}{2}(\alpha_1 + \alpha_2) = \frac{1}{2}A_{12}$
A_2A_2		p	q	$pd - qa$	$\alpha_2 = \frac{1}{2}A_{22}$
随机交配群体	p^2	$2pq$	q^2	$(p - q)a + 2pqa$	

育种值与等位基因平均效应的关系

- 从最后一列给出的结果可以看到，个体的育种值正好等于个体携带两个等位基因的平均效应之和。
- 因此，基因型 A_1A_1 、 A_1A_2 和 A_2A_2 的育种值分别为 $A_{11}=2\alpha_1$ ， $A_{12}=\alpha_1 + \alpha_2$ 和 $A_{22}=2\alpha_2$ 。统一起来，各种基因型的育种值表示为：

$$A_{ij} = \alpha_i + \alpha_j$$

显性离差的定义和计算

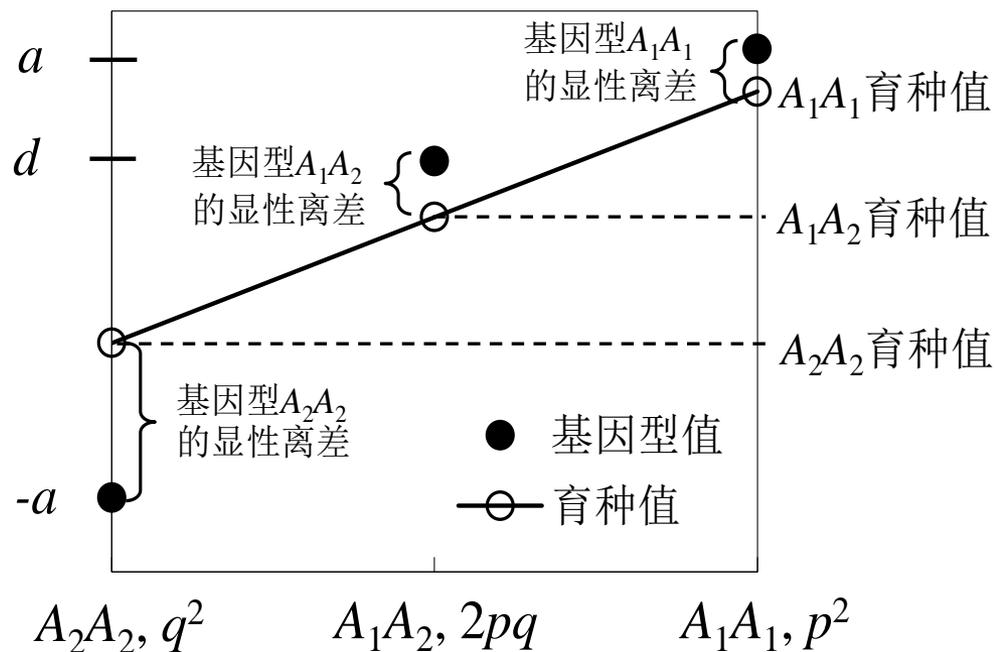
- 在育种值的基础上，我们可以把 A_iA_j 的基因型值进一步分解为

$$G_{ij} = \mu + g_{ij} = \mu + \alpha_i + \alpha_j + \delta_{ij} = \mu + A_{ij} + D_{ij}$$

- 其中， α_i 是等位基因 A_i 的平均效应， $A_{ij} = \alpha_i + \alpha_j$ 称为基因型 A_iA_j 的育种值， D_{ij} 为遗传效应 g_{ij} 中扣除育种值的部分，称为显性离差（dominance deviation）。为与公式8.1的F无穷模型相区分，上面的公式又称为 F_2 模型（ F_2 model）。

个体育种值和显性离差关系的回归表示

- 基因型值 Y 对等位基因 A_1 数目 X 的回归系数，正好等于基因替换效应，三种基因型的育种值都落在这一条回归直线上。
- 可以将育种值视为图中回归直线上空心圆圈所在的位置，而显性离差为空心圆圈代表的育种值和实心圆圈代表的遗传效应之间的差异。



利用双向表计算育种值和显性离差

配子型及其频率	主效应，即育种值		交互作用，即显性离差	
	A_1, p	A_2, q	A_1	A_2
A_1, p	$\alpha_1 + \beta_1$	$\alpha_2 + \beta_1$	$g_{11} - (\alpha_1 + \beta_1)$	$g_{12} - (\alpha_2 + \beta_1)$
A_2, q	$\alpha_2 + \beta_1$	$\alpha_2 + \beta_2$	$g_{21} - (\alpha_2 + \beta_1)$	$g_{22} - (\alpha_2 + \beta_2)$

- 双因素方差分析中，行效应和列效应之和称为处理的主效应，它们对应于各种基因型的育种值；总效应减去主效应之外为互作效应，它们对应于各种基因型的显性离差。
- 如果 A_1A_2 的基因型值正好位于 A_1A_1 和 A_2A_2 中间，即 $d=0$ ，表中所有交互作用均为0；否则，交互作用不为0。因此，显性效应有时也称为同一个座位上不同等位基因间的互作，是由不同雌雄配子之间相互作用而产生的。

老鼠随机交配群体的均值、等位基因效应和基因替换效应

效应类型	基因 p_g 的频率 q				
	$q=0.1$	$q=0.3$	$q=0.5$	$q=0.7$	$q=0.9$
群体均值	13.56	12.44	11.00	9.24	7.16
基因+的平均效应	0.24	0.96	2.00	3.36	5.04
基因 p_g 的平均效应	-2.16	-2.24	-2.00	-1.44	-0.56
基因+对 p_g 的替代效应	2.40	3.20	4.00	4.80	5.60
基因型++的育种值	0.48	1.92	4	6.72	10.08
基因型+ p_g 的育种值	-1.92	-1.28	0	1.92	4.48
基因型 $p_g p_g$ 的育种值	-4.32	-4.48	-4	-2.88	-1.12
基因型++的显性离差	-0.04	-0.36	-1	-1.96	-3.24
基因型+ p_g 的显性离差	0.36	0.84	1	0.84	0.36
基因型 $p_g p_g$ 的显性离差	-3.24	-1.96	-1	-0.36	-0.04

§ 8.2 随机交配群体的遗传方差和亲子相关

- § 8.2.1 加性方差和显性方差
- § 8.2.2 半同胞家系间的方差与亲子间的协方差
- § 8.2.3 全同胞家系的方差与亲子间的协方差

随机交配群体中不同基因型的频率、基因型值、遗传效应、育种值和显性离差

基因型	频率	基因型值	遗传效应	育种值	显性离差
A_1A_1	p^2	$m+a$	$2q(a-pd)$ 或 $2q(\alpha-qd)$	$2q\alpha$	$-2q^2d$
A_1A_2	$2pq$	$m+d$	$(q-p)a+(1-2pq)d$ 或 $(q-p)\alpha+2pqd$	$(q-p)\alpha$	$2pqd$
A_2A_2	q^2	$m-a$	$-2p(a+qd)$ 或 $-2p(\alpha+pd)$	$-2p\alpha$	$-2p^2d$

育种值和显性离差的期望和协方差

- 育种值和显性离差的加权平均数均为0，它们之间的协方差也为0，即：

$$E(A) = E(D) = E(AD) = 0$$

$$Cov(A, D) = E(AD) - E(A)E(D) = 0$$

- 因此，从遗传效应分解出来的育种值和显性离差，二者之间无相关，这样的分解在统计上称为正交分解。

随机交配群体的加性方差和显性方差

- 在正交分解的条件下，加性方差和显性方差可分别利用下面的公式计算，遗传方差也就等于加性方差与显性方差二者之和。

$$V_A = V(A) = E[A - E(A)]^2 = E(A^2) - [E(A)]^2 = E(A^2)$$

$$V_D = V(D) = E[D - E(D)]^2 = E(D^2) - [E(D)]^2 = E(D^2)$$

$$V_G = V(A + D) = V(A) + V(D) = V_A + V_D$$

加性方差和显性方差的计算公式

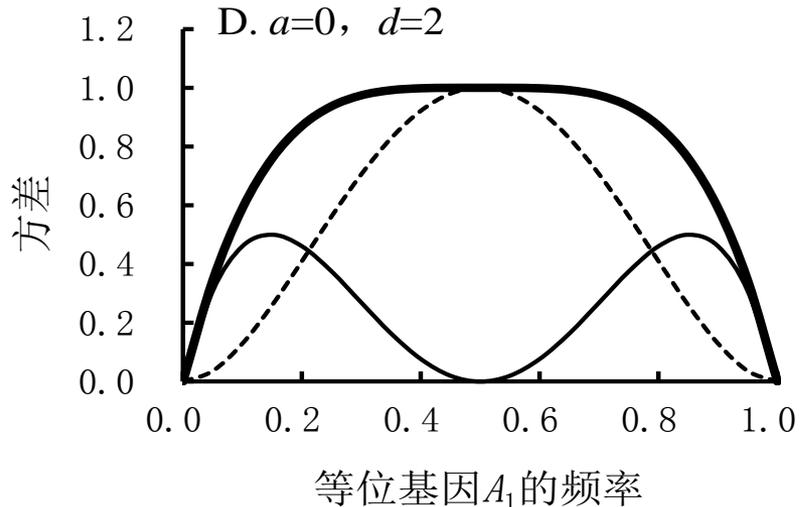
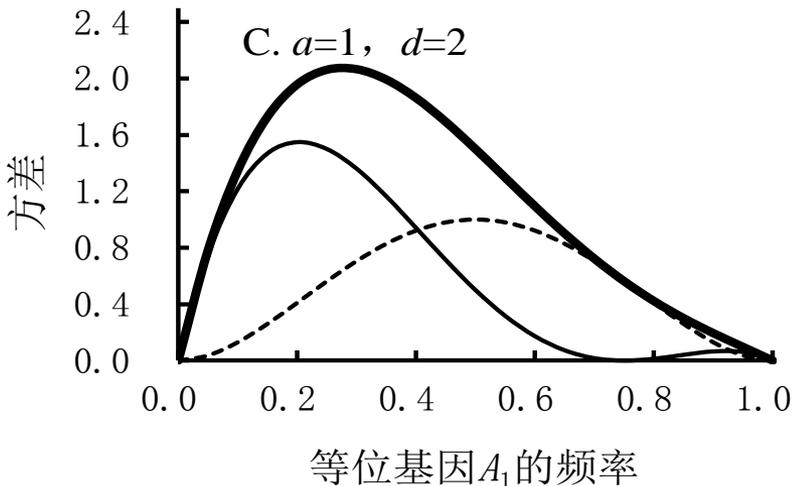
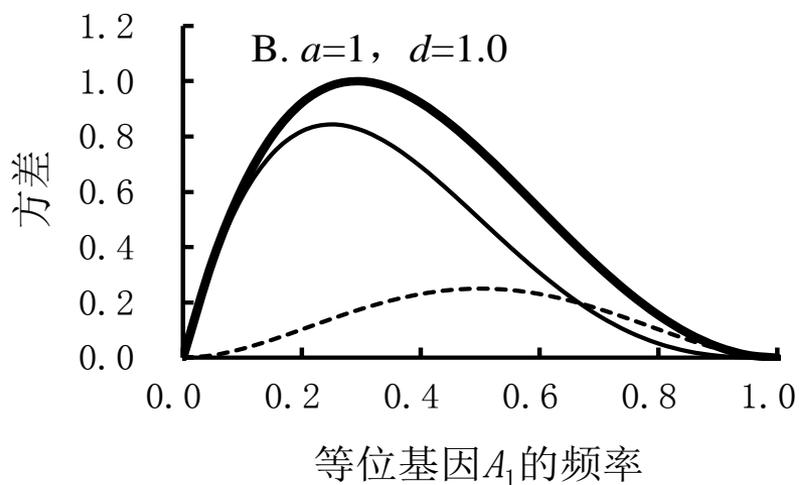
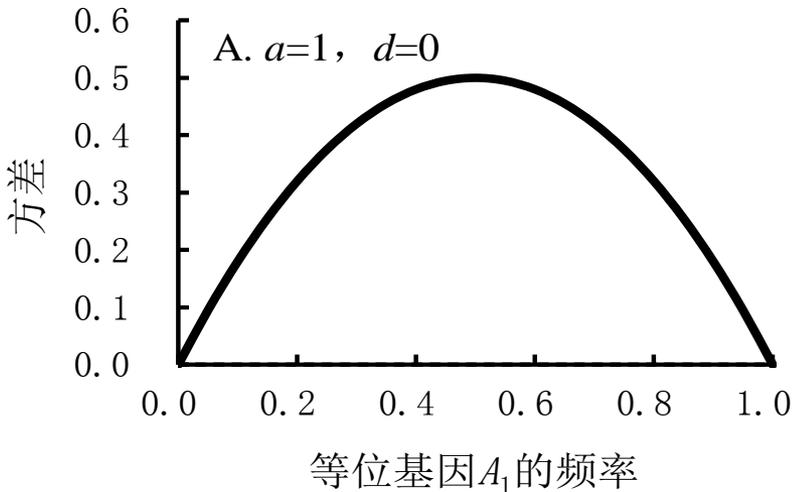
$$\begin{aligned}V_A &= p^2(2q\alpha)^2 + 2pq[(q-p)\alpha]^2 + q^2(-2p\alpha)^2 \\&= [4p^2q^2 + 2pq(q-p)^2 + 4p^2q^2]\alpha^2 = 2pq[4pq + (q-p)^2]\alpha^2 \\&= 2pq\alpha^2 = 2pq[a + (q-p)d]^2\end{aligned}$$

$$V_D = p^2(-2q^2d)^2 + 2pq(2pqd)^2 + q^2(-2p^2d)^2 = 4p^2q^2d^2$$

$$V_G = 2pq[a + (q-p)d]^2 + 4p^2q^2d^2$$

不同遗传模型和等位基因频率的随机交配群体中，加性方差 V_A 、显性方差 V_D 和遗传方差 V_G 的大小

—— 加性方差 - - - - 显性方差 —— 总遗传方差



部分显性遗传模型的育种值

- 表8.8中，等位基因+表现为部分显性，基因型++的育种值总是高于其他两种基因型。说明不论群体的基因频率如何，基因+的平均效应总是高于 pg 的平均效应，基因型++的后代均值总是要高于 $+pg$ 和 $pgpg$ 的后代均值。
- 如果根据育种值进行选择，基因型++就会有更多的机会被选中，后代群体中等位基因+的频率就会不断增加，群体均值也随之提高，直到群体被固定在等位基因+上为止。

超显性遗传模型的育种值

- 从图8.3C的超显性模型我们已经看到，基因频率在0和1之间（不含0和1）取值时，加性方差还可以为0。在图8.3D的无加性模型中，基因频率为0.5的加性方差等于0。在这样的遗传模型中，各种基因型育种值的相对高低并不是一成不变的。
- 假定三种基因型 A_1A_1 、 A_1A_2 、 A_2A_2 的平均表现分别为110、150、90。在单基因加显性模型下， $m=100$ 、 $a=10$ 、 $d=50$ ， $d/a=5$ 为超显性。下表给出5种基因频率随机交配群体的均值、遗传方差、等位基因平均效应、育种值、加性方差、显性离差和显性方差。对于这三种基因型值，容易证明等位基因 A_1 频率 $p=0.6$ 时，随机交配群体的均值达到最高点，即表中给出的126。

随机交配群体的遗传参数

三种基因型 A_1A_1 、 A_1A_2 、 A_2A_2 的平均表现分别是110、150、90

遗传参数	基因 A_1 的频率 p				
	$p=0.4$	$p=0.5$	$p=0.6$	$p=0.7$	$p=0.8$
群体均值	122	125	126	125	122
遗传方差	768	675	576	483	384
基因 A_1 的平均效应	12	5	0	-3	-4
基因 A_2 的平均效应	-8	-5	0	7	16
基因型 A_1A_1 的育种值	24	10	0	-6	-8
基因型 A_1A_2 的育种值	4	0	0	4	12
基因型 A_2A_2 的育种值	-16	-10	0	14	32
加性方差	192	50	0	42	128
基因型 A_1A_1 的显性离差	-36	-25	-16	-9	-4
基因型 A_1A_2 的显性离差	24	25	24	21	16
基因型 A_2A_2 的显性离差	-16	-25	-36	-49	-64
显性方差	576	625	576	441	256

超显性遗传模型下的育种值选择

- 在 $p < 0.6$ 的群体中，等位基因 A_1 有较高的平均表现， A_1A_1 的育种值也较高。如果根据育种值选择亲本的， A_1 的频率就会不断增加，后代群体的均值也随之上升。
- 在 $p > 0.6$ 的群体中，等位基因 A_2 有较高的平均表现， A_2A_2 有较高的育种值。如果根据育种值选择亲本， A_1 的频率就会降低，其结果也是提高了后代群体的均值。
- 当等位基因 A_1 的频率达到0.6时，群体均值达到最高点，两种等位基因的平均效应和三种基因型的育种值均为0。这时，根据育种值的选择不会引起基因频率的任何改变，群体均值在0.6的 A_1 频率上不再发生变化，这就是所谓的选择高原（selection plateau）。
- 达到选择高原的群体，加性方差为0，对选择不再有任何响应。但这时不代表群体中没有遗传方差了，只是说决定选择响应的加性方差没有了。

杂交品种の利用

- 表8.10的例子，一方面说明了育种值和加性方差在选择中的重要性，另一方面也说明了随机交配群体中的选择，有时不一定能使群体均值达到最高的基因型值。
- 试想，如果只选择基因型值最高的 A_1A_2 作亲本，随机交配后代的均值只有125，对应于表8.10频率为0.5的情形。如果根据育种值选择亲本，最高的群体均值也只能达到126，对应于表8.10频率为0.6的情形，远低于 A_1A_2 基因型值150。
- 因此，在存在超显性时，随机交配群体不可能达到最高的可能基因型值。这也从另外一方面说明了植物育种中杂交种选育的重要性，即把两个自交系 A_1A_1 和 A_2A_2 的杂种 F_1 作为品种，杂种 F_1 群体在田间是整齐一致的，同时还实现了150的最高基因型值。

随机交配群体中的半同胞家系

- 随机交配群体中，具有一个共同亲本的后代群体，称为一个半同胞家系（half-sib family）。根据亲本的基因型，可以把半同胞家系分成三类，它们的频率分别等于三种基因型 A_1A_1 、 A_1A_2 、 A_2A_2 在随机交配群体中的频率 p^2 、 $2pq$ 、 q^2 。
- 这三类半同胞家系其实就是计算育种值时各种基因型的后代。根据遗传效应的分解，亲本基因型值与群体均值的离差等于育种值和显性效应之和；半同胞后代与群体均值的离差等于育种值的一半。

亲本表现和半同胞后代家系的平均表现

亲代基因型	频率	基因型值与群体均值的离差效应	子代基因型、基因型值和频率			家系均值与群体均值的离差
			A_1A_1, a	A_1A_2, d	$A_2A_2, -a$	
A_1A_1	p^2	$A_{11} + D_{11}$	p	q	0	$\frac{1}{2}A_{11}$
A_1A_2	$2pq$	$A_{12} + D_{12}$	$\frac{1}{2}p$	$\frac{1}{2}(p + q)$	$\frac{1}{2}q$	$\frac{1}{2}A_{12}$
A_2A_2	q^2	$A_{22} + D_{22}$	0	p	q	$\frac{1}{2}A_{22}$

半同胞家系间和家系内的遗传方差

- 半同胞家系间的方差度量了家系平均数之间的变异程度，称为家系间方差，它包含了1/4的加性方差，不包含显性方差，计算过程如下：

$$V_{\text{bHS}} = \text{Cov}_{\text{HS}} = \text{Cov}\left(\frac{1}{2}A, \frac{1}{2}A\right) = \frac{1}{4} \text{Cov}(A, A) = \frac{1}{4}V_A$$

- 半同胞家系内的平均方差度量了家系内的平均变异，称为家系内方差，它包含了3/4的加性方差和所有的显性方差，计算过程如下：

$$V_{\text{wHS}} = V_G - V_{\text{bHS}} = \frac{3}{4}V_A + V_D$$

半同胞家系的亲子间协方差

- 亲子间的协方差度量了亲本和后代间的相关程度，它包含了1/2的加性方差，不包含显性方差，计算过程如下：

$$Cov_{OP} = Cov(A + D, \frac{1}{2}A) = \frac{1}{2}Cov(A, A) + \frac{1}{2}Cov(D, A) = \frac{1}{2}V_A$$

半同胞家系的亲子间回归系数

- 把后代的表现用O表示、亲代的表现用P表示，亲代的表型方差表示为，

$$V_P = V_G + V_\varepsilon = V_A + V_D + V_\varepsilon$$

- 因此得到亲子之间回归系数的计算公式：

$$b_{OP} = \frac{Cov_{OP}}{V_P} = \frac{\frac{1}{2}V_A}{V_A + V_D + V_\varepsilon}$$

随机交配群体中的全同胞家系

- 随机交配群体中，根据亲本基因型可以把全同胞家系分成6类。这6类家系的遗传结构类似于上一章的6个基本世代 P_1 、 B_1 、 F_1 、 F_2 、 B_2 和 P_2 。
- 它们的频率是由三种基因型 A_1A_1 、 A_1A_2 、 A_2A_2 在随机交配群体中的频率 p^2 、 $2pq$ 、 q^2 决定的，中亲值等于两个亲本基因型值的平均数。

中亲表现和全同胞后代家系的平均表现

家系类型	频率	中亲值	全同胞家系基因型频率			家系均值	家系内方差
			A_1A_1, a	A_1A_2, d	$A_2A_2, -a$		
$A_1A_1 \times A_1A_1$	p^4	a	1	0	0	a	0
$A_1A_1 \times A_1A_2$	$4p^3q$	$\frac{1}{2}(a+d)$	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}(a+d)$	$\frac{1}{4}(a-d)^2$
$A_1A_1 \times A_2A_2$	$2p^2q^2$	0	0	1	0	d	0
$A_1A_2 \times A_1A_2$	$4p^2q^2$	d	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}d$	$\frac{1}{2}a^2 + \frac{1}{4}d^2$
$A_1A_2 \times A_2A_2$	$4pq^3$	$\frac{1}{2}(-a+d)$	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}(-a+d)$	$\frac{1}{4}(a+d)^2$
$A_2A_2 \times A_2A_2$	q^4	$-a$	0	0	1	$-a$	0
随机交配群体			p^2	$2pq$	q^2	$\mu = (p-q)a + 2pqd$	

全同胞家系间的遗传方差

- 全同胞家系间的方差度量了家系平均数之间的变异程度，称为家系间方差，它包含了1/2的加性方差和1/4的显性方差，计算过程如下：

$$\begin{aligned}V_{\text{bFS}} &= \text{Cov}_{\text{FS}} = p^4 \times (a)^2 + 4p^3q \times \left[\frac{1}{2}(a+d)\right]^2 + 2p^2q^2 \times (d)^2 \\ &\quad + 4p^2q^2 \times \left(\frac{1}{2}d\right)^2 + 4pq^3 \times \left[\frac{1}{2}(-a+d)\right]^2 + q^4(-a)^2 - \mu^2 \\ &= pq[a + (q-p)d]^2 + p^2q^2d^2 \\ &= \frac{1}{2}V_A + \frac{1}{4}V_D\end{aligned}$$

全同胞家系内的遗传方差

- 全同胞家系内的平均方差度量了家系内的平均变异，称为家系内方差，它包含了1/2的加性方差和3/4的显性方差，计算过程如下：

$$\begin{aligned}V_{\text{wFS}} &= p^4 \times 0 + 4p^3q \times \frac{1}{4}(a-d)^2 + 2p^2q^2 \times 0 \\ &+ 4p^2q^2 \times \left(\frac{1}{2}a^2 + \frac{1}{4}d^2\right) + 4pq^3 \times \frac{1}{4}(a+d)^2 + q^4 \times 0 \\ &= pq[a + (q-p)d]^2 + 3p^2q^2d^2 \\ &= \frac{1}{2}V_A + \frac{3}{4}V_D\end{aligned}$$

全同胞家系的亲子间协方差

- 亲子间的协方差度量了中亲和后代间的相关程度，它包含了1/2的加性方差，不包含显性方差，计算过程如下：

$$\begin{aligned} Cov_{OP} &= p^4 \times a \times a + 4p^3q \times \frac{1}{2}(a+d) \times \frac{1}{2}(a+d) + 2p^2q^2 \times 0 \times d \\ &\quad + 4p^2q^2 \times d \times \frac{1}{2}d + 4pq^3 \times \frac{1}{2}(-a+d) \times \frac{1}{2}(-a+d) + q^4 \times (-a) \times (-a) - \mu^2 \\ &= pq[a + (q-p)d]^2 = \frac{1}{2}V_A \end{aligned}$$

- 用X和Y表示全同胞家系的两个亲本，后代用O表示，另种计算亲子间协方差的方法是：

$$Cov_{OP} = Cov\left[\frac{1}{2}(X+Y), O\right] = \frac{1}{2}Cov(X, O) + \frac{1}{2}Cov(Y, O) = \frac{1}{2}V_A$$

利用全同胞和半同胞的家系间协方差 估计加显性方差

- 加性方差的估计

$$V_A = 4Cov_{HS}$$

- 显性方差的估计

$$V_D = 4(Cov_{FS} - 2Cov_{HS})$$

全同胞家系的中亲回归系数

- 两个亲本X和Y产生全同胞后代O，中亲及其方差分别表示为：

$$\bar{P} = \frac{1}{2}(X + Y) \quad V(\bar{P}) = \frac{1}{4}[V(X) + V(Y)] = \frac{1}{2}V_P$$

- 因此得到中亲与子代的回归系数公式：

$$b_{O\bar{P}} = \frac{Cov_{O\bar{P}}}{V_{\bar{P}}} = \frac{\frac{1}{2}V_A}{\frac{1}{2}V_P} = \frac{V_A}{V_P}$$

- 后代与中亲的回归系数是与单个亲本回归系数的2倍。因此，后代与中亲的回归系数要远高于与单个亲本的回归系数。在 § 6.3.4 中，后代子女平均身高与父母平均身高的回归系数为0.65，远高于儿子与父亲身高的回归系数0.516。

§ 8.3 上位性互作模型的遗传方差分解

- § 8.3.1 两个座位的互作模型和上位性方差
- § 8.3.2 上位性对保持加性方差的作用

两个座位间的互作

- 上位性 (epistasis) 表示不同座位上等位基因之间的相互作用，上位性效应 (epistatic effect) 源于座位间效应的非可加性。换句话说，如果遗传效应不等于单个座位上效应的总和 (总体不等于各部分之和)，则存在上位性效应。

两个座位间的互作模型

- 假定两个基因座位A和B影响同一性状，一个个体在座位A的基因型为 A_iA_j ，在座位B的基因型为 B_kB_l ， $A_iA_jB_kB_l$ 的基因型值 G_{ijkl} 用上位性模型表示为：

$$\begin{aligned}G_{ijkl} &= \mu + (\alpha_i + \alpha_j + \delta_{ij}) + (\alpha_k + \alpha_l + \delta_{kl}) + I_{ijkl} \\ &= \mu + A_{ij} + B_{kl} + I_{ijkl}\end{aligned}$$

- 其中， A_{ij} 为座位A产生的遗传效应， B_{ij} 为座位B产生的遗传效应。单个座位上的效应仍然按照加显性模型分解， α_i 、 α_j 、 α_k 、 α_l 表示等位基因的平均效应； δ_{ij} 表示座位A上的显性离差， δ_{kl} 表示座位B上的显性离差； I_{ijkl} 表示两个座位之间的上位性离差或上位性效应，计算公式是：

$$I_{ijkl} = G_{ijkl} - (\mu + A_{ij} + B_{kl}) = G_{ijkl} - G_{ij..} - G_{..kl} + \mu$$

重叠显性上位作用下的上位性效应

等位基因的频率均为0.5，两个座位间的连锁不平衡度为0

座位 A	座位 B			行平均 $G_{ij..}$
	$B_1B_1, \frac{1}{4}$	$B_1B_2, \frac{1}{2}$	$B_2B_2, \frac{1}{4}$	
$A_1A_1, \frac{1}{4}$	$G_{1111} = 1$	$G_{1112} = 1$	$G_{1122} = 1$	$G_{11..} = 1$
	$I_{1111} = -\frac{1}{16}$	$I_{1112} = -\frac{1}{16}$	$I_{1122} = \frac{3}{16}$	
$A_1A_2, \frac{1}{2}$	$G_{1211} = 1$	$G_{1212} = 1$	$G_{1222} = 1$	$G_{12..} = 1$
	$I_{1211} = -\frac{1}{16}$	$I_{1212} = -\frac{1}{16}$	$I_{1222} = \frac{3}{16}$	
$A_2A_2, \frac{1}{4}$	$G_{2211} = 1$	$G_{2212} = 1$	$G_{2222} = 0$	$G_{22..} = \frac{3}{4}$
	$I_{2211} = \frac{3}{16}$	$I_{2212} = \frac{3}{16}$	$I_{2222} = -\frac{9}{16}$	
列平均 $G_{..ij}$	$G_{..11} = 1$	$G_{..12} = 1$	$G_{..22} = \frac{3}{4}$	总平均 $\mu = \frac{15}{16}$

计算上位性效应的一般方法

- 前面计算互作效应的方法具有普遍性。对于任意基因频率的随机交配群体，只要两个座位之间的连锁不平衡度为0，都可以使用。
- 例如，等位基因 A_1 和 A_2 的频率分别为 p 和 q ，等位基因 B_1 和 B_2 的频率分别为 u 和 v ，只需要把表中座位A的三种基因型频率用 p^2 、 $2pq$ 、 q^2 代替，座位B的三种基因型频率用 u^2 、 $2uv$ 、 v^2 代替即可。

加性、显性和上位性方差

- 在两个座位之间无连锁不平衡的假定下，9种基因型的联合频率等于边缘频率的乘积。这样，行平均数之间的方差就是座位A的遗传方差、列平均数之间的方差就是座位B的遗传方差、互作效应之间的方差就是上位性方差。
- 这样的分解也满足正交性。同时，利用座位A的三种边缘平均数，就可以如单基因座位进行加显性方差的分解，从而计算座位A的加显性方差；同时，利用座位B的三种边缘平均数，计算座位B的加显性方差。

上位性模型的遗传效应和遗传方差分解

- 考虑所有控制数量性状基因的效应，下面的公式给出遗传效应的一般分解模型，其中 A 表示来自所有座位育种值的总和， D 表示显性离差的总和， I 表示上位性离差（epistatic deviation）或互作离差（interaction deviation）的总和。

$$G = A + D + I$$

- 随机交配群体中，育种值、显性离差和上位性离差的加权均值都是0，且三者之间无相关。因此，总的遗传方差等于加性方差、显性方差和上位性方差三部分之和。

$$V_G = V_A + V_D + V_I$$

F₂群体常见上位互作类型中，不同方差成分占遗传方差的比例

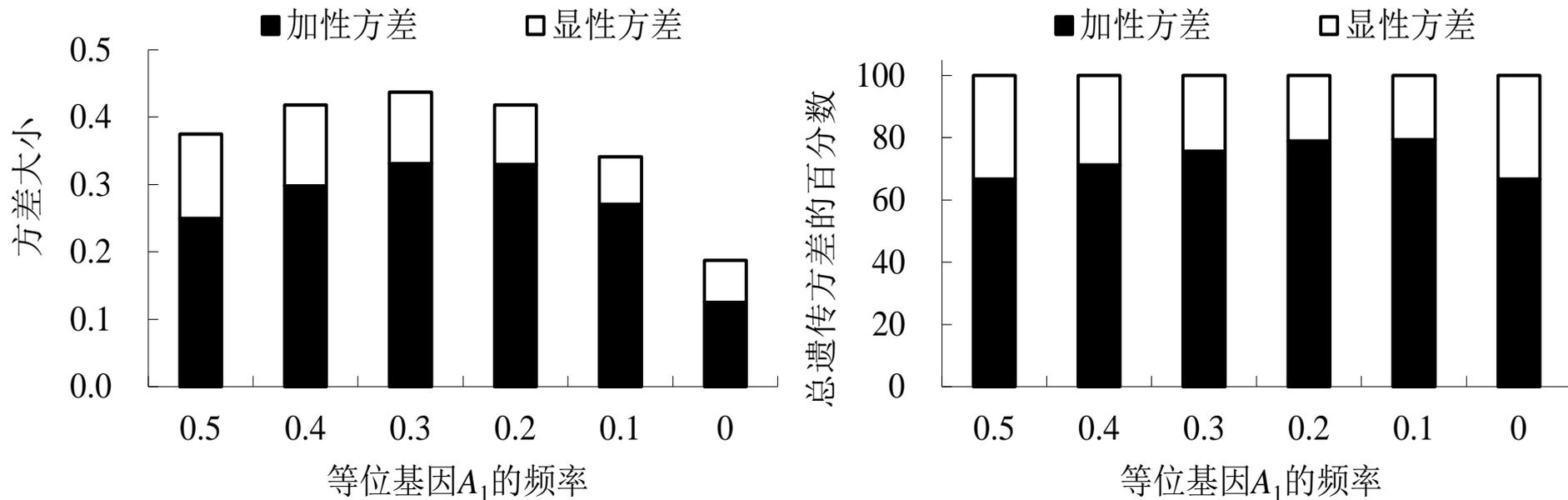
F ₂ 分离比	表型均值（或基因型值）				V _A /V _G	V _D /V _G	V _I /V _G
	A ₁ *B ₁ *	A ₁ *B ₂ B ₂	A ₂ A ₂ B ₁ *	A ₂ A ₂ B ₂ B ₂			
9:3:3:1	3	2	1	0	0.6667	0.3333	0
12:3:1	3	3	1	0	0.6431	0.3216	0.0353
10:3:3	3	2	1	3	0.3270	0.1635	0.5094
9:6:1	3	2	2	0	0.6289	0.3145	0.0566
9:4:3	3	2	1	2	0.5128	0.2564	0.2308
15:1	3	3	3	0	0.2667	0.1333	0.6
13:3	3	3	1	3	0.5128	0.2564	0.2308
10:6	3	2	2	3	0.2667	0.1333	0.6
9:7	3	1	1	1	0.5714	0.2857	0.1429

上位性对保持加性方差的作用

- 有时，由于基因的固定，上位性方差还可能会转化成加性方差。考虑到频率对方差的影响，就容易理解为何一个座位的固定有可能会提高方差。
- 以15:1重叠显性上位为例， F_2 群体的表型只有两种，假定基因型 $aabb$ 的平均表现为1，其它基因型为0，它们的频率分别为 $1/16$ 和 $15/16$ 。因此， F_2 群体的均值为 $1/16$ 、方差为 $1/16 - 1/256$ 。
- 如果群体在座位A被固定在等位基因 a 上，座位B的两个等位基因频率仍为0.5，这时仍然有1和0两种表型，但它们的频率变分别为 $1/4$ 和 $3/4$ 。因此，群体的均值为 $1/4$ 、方差为 $1/4 - 1/16$ ，均高于4个等位基因频率均为0.5的群体均值和方差。

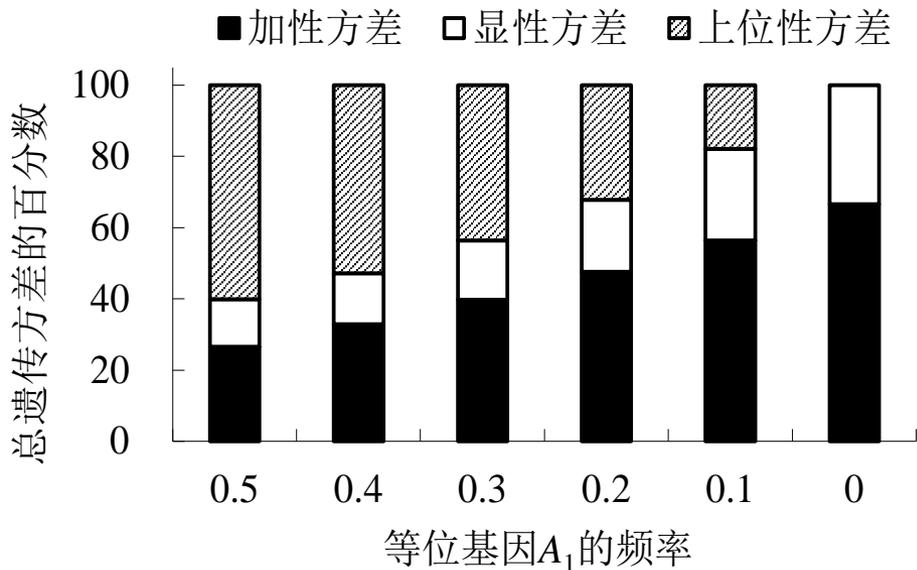
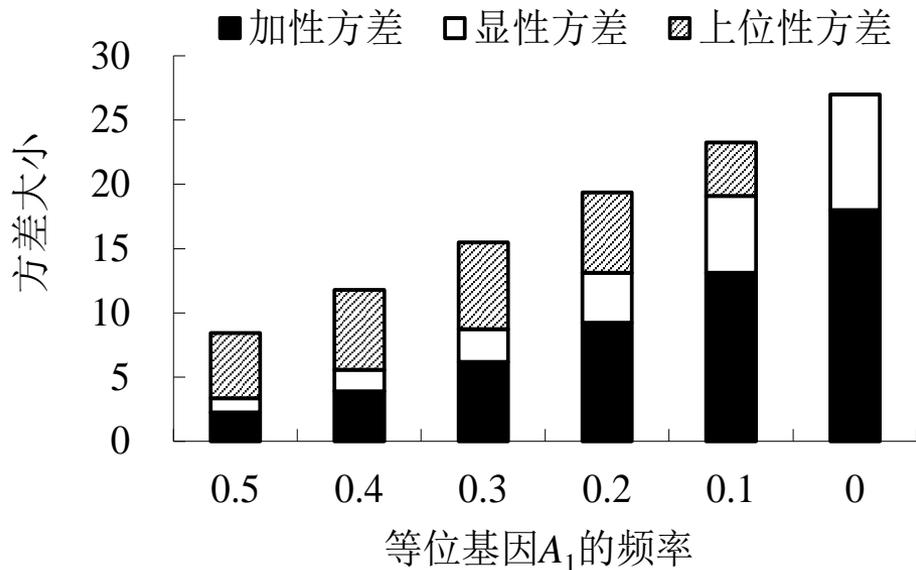
完全显性遗传模型（9:3:3:1）中不同遗传方差成分

- $a_A=d_A=a_B=d_B=1$ ；等位基因 B_1 和 B_2 等于0.5并保持不变。
- 当基因频率均为0.5时， F_2 群体中， $V_G=0.38$ ， $V_A=0.25$ ， $V_D=0.13$ ；当座位A上的基因被固定在 A_2 时， $V_G=0.19$ ， $V_A=0.13$ ， $V_D=0.06$ ，加性方差和显性方差均减少50%。



重叠显性上位遗传模型（15:1）中 不同遗传方差成分

- 假定基因型值分别为13和1，当基因频率均为0.5时， F_2 群体中， $V_G=8.44$ ， $V_A=2.25$ ， $V_D=1.13$ ， $V_I=5.06$ ；当座位A上的基因被固定在 A_2 时， $V_G=27$ ， $V_A=18$ ， $V_D=9$ ， $V_I=0$ 。加性方差从2.25增加到18，显性方差从1.13增加到9。



上位性对保持加性方差的作用

- 并非所有的上位性遗传模型，随着基因的固定，加性方差和显性方差都随着增加。以9:7上位性模型为例，假定基因型值也分别为13和1。当基因频率均为0.5时， F_2 群体中， $V_G=35.44$ ， $V_A=20.25$ ， $V_D=10.13$ ， $V_I=5.06$ ；当座位A上的基因被固定在 A_1 时， $V_G=27$ ， $V_A=18$ ， $V_D=9$ ， $V_I=0$ 。加性方差从20.25下降到18，显性方差从10.13下降到9。尽管加性方差和显性方差随着基因的固定在下降，但下降的幅度要比不存在上位性效应时的50%要小很多。
- 上面的这些例子说明，尽管上位性方差本身不是很大，但对加性方差的保持来说起重要作用。维持在一定水平上的加性方差，是育种中各种选择方法取得遗传效果的基础。上位性的存在，也许是对长期选择有效性的一个较合理的解释。

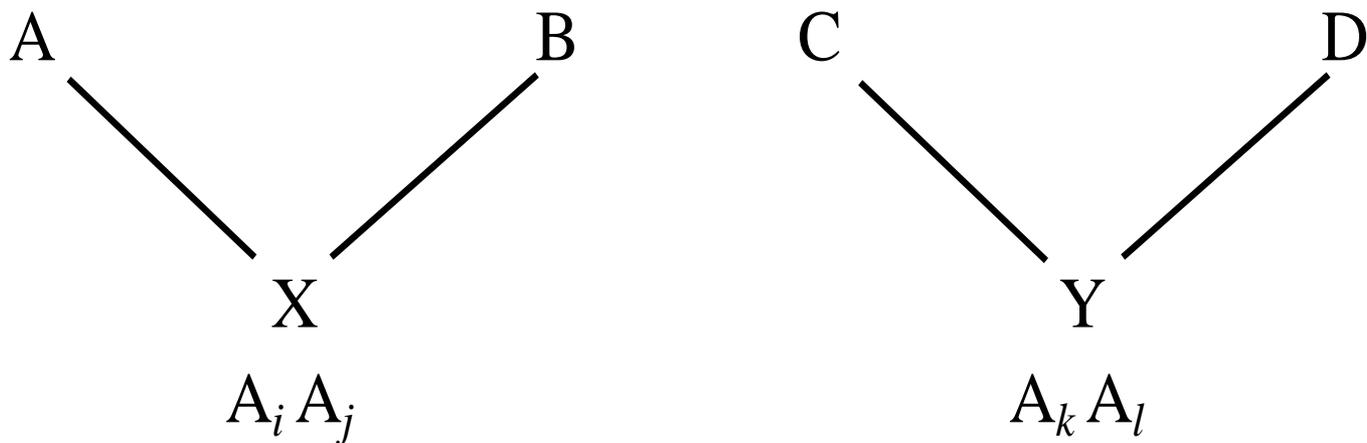
§ 8.4 亲属间协方差的一般表示与遗传力估计

- § 8.4.1 亲属间协方差的一般表示方式
- § 8.4.2 随机交配群体的遗传力估计

亲属间的相关

- 谚语“龙生龙，凤生凤”（like begets like）也许是人们最早通过观测遗传学上亲属间相似性得到的结果，也就是说，较近的亲属（如親子）要比较远的亲属（如叔侄）有较大的相似性。
- 亲属间的协方差，度量了亲属间的相似性。没有祖先关联的个体间不存在相似性，即协方差为0。
- 有时非遗传因素也会影响个体间的相似性，植物上可通过随机设计排除非遗传因素的影响。这里我们认为非遗传因素的效应在个体间是独立的。

个体X和Y的系谱



- X的两个亲本与Y的两个亲本之间的共祖先程度越高，X和Y之间的相关程度越高。

X和Y育种值之间的协方差

$$V(\alpha_i + \alpha_j) = V(A) = V_A \quad V_A = 2V(\alpha_i)$$

- X和Y育种值之间协方差的一般表示为

$$\begin{aligned} Cov_\alpha &= P(A_i \equiv A_k)Cov(\alpha_i, \alpha_k) + P(A_i \equiv A_l)Cov(\alpha_i, \alpha_l) \\ &\quad + P(A_j \equiv A_k)Cov(\alpha_j, \alpha_k) + P(A_j \equiv A_l)Cov(\alpha_j, \alpha_l) \\ &= 4f_{XY}V(\alpha_i) = 2f_{XY}V_A \end{aligned}$$

X和Y显性离差之间的协方差

- X和Y显性离差之间协方差的一般表示为

$$\begin{aligned}Cov_{\delta} &= P(A_i \equiv A_k, A_j \equiv A_l)Cov(\delta_{ij}, \delta_{kl}) \\ &\quad + P(A_i \equiv A_l, A_j \equiv A_k)Cov(\delta_{ij}, \delta_{kl}) \\ &= (f_{AC}f_{BD} + f_{AD}f_{BC})V_D\end{aligned}$$

亲属间协方差的一般表示

- 用 r 表示加性方差前面的系数， u 表示显性方差前面的系数，得到个体X和Y协方差的一般表示：

$$\text{Cov}(X, Y) = rV_A + uV_D$$

$$r = 2f_{XY} \quad u = f_{AC}f_{BD} + f_{AD}f_{BC}$$

- 在上位性模型下，个体之间的协方差中还会同时包含二阶乃至更高阶的加加、加显和显显互作方差。这些互作方差的系数都可以从加性方差的系数 r 和显性方差的系数 u 计算出来，即：

$$\text{Cov}(X, Y) = rV_A + uV_D + r^2V_{AA} + ruV_{AD} + u^2V_{DD} + \dots$$

常见亲属间的协方差

个体 X 和 Y 的类型	V_A 的系数 (公式 8.35)	V_D 的系数 (公式 8.36)
亲代和子代	$\frac{1}{2}$	0
半同胞		
非近交共同亲本	$\frac{1}{4}$	0
任意亲本 F	$\frac{1+F}{4}$	0
全同胞		
非近交共同亲本	$\frac{1}{2}$	$\frac{1}{4}$
任意亲本 F_A 和 F_B	$\frac{2+F_A+F_B}{4}$	$\frac{(1+F_A)(1+F_B)}{4}$

随机交配群体的方差成分及其来源

方差成分	符号	方差来源
表型方差	V_P	表型值
基因型方差	V_G	基因型值（遗传效应）
加性方差	V_A	育种值
显性方差	V_D	显性离差
上位性方差	V_I	上位性离差
基因型×环境互作方差	V_{GE}	基因型×环境互作效应
随机误差方差	V_ε	随机误差

随机交配群体中育种值的作用

- 在一个无选择的HW平衡群体中，群体均值和遗传方差在世代间保持恒定不变。
- 根据定义，育种值越高的亲代，后代的平均表现也就越高。因此，如果能够根据育种值选择亲本，就可以有效改变后代群体的平均表现。

随机交配群体的狭义遗传力

- 育种中首先观测到的是亲本个体的表型，正是由于表型中包含育种值的信息，因此根据表型的选择也能改变后代的平均表现。
- 后代群体均值被改变的程度，既取决于随机交配群体中加性方差的大小，同时还取决于群体的表型方差。狭义遗传力（heritability in the narrow sense）定义为由育种值决定的加性方差占表型方差的比例，即：

$$h^2 = \frac{V_A}{V_P}$$

狭义遗传力的作用

- 知道了前面公式中的遗传力，就可以预测根据表型选择亲本后，后代群体平均表现相对于亲代的变化量，即预测遗传进度（详见第11和12章）。
- 因此，狭义遗传力可能是传统数量遗传中最重要的一个遗传参数，利用各种遗传交配设计对狭义遗传力进行估计是数量遗传学的重要内容（详见第10章）。

狭义遗传力与回归系数和相关系数

- 遗传力正好等于育种值对表型的回归系数，也等于后代与中亲值的回归系数，即：

$$h^2 = b_{AP} = b_{OP} = \frac{Cov(A, P)}{V_P}$$

- 亲本育种值与亲本表型之间的相关系数，正好等于狭义遗传力的平方根，即：

$$r_{AP} = \frac{Cov_{AP}}{\sqrt{V_A V_P}} = \frac{Cov_{AP}}{V_P} \sqrt{\frac{V_P}{V_A}} = h^2 / h = h$$

狭义遗传力的预测作用

- 个体的表型用 P 表示，它的育种值 A 就可以利用公式8.40给出的关系进行预测。育种值的另外一个含义是后代的平均表现，预测公式也代表了后代群体相对于亲代的差异，这其实就是第11章要介绍的遗传进度。

$$A = b_{AP}(P - \mu) = h^2(P - \mu)$$

狭义遗传力的群体特异性

- 与广义遗传力一样，狭义遗传力也是群体水平的遗传参数。对于同一个性状，利用不同群体、不同环境的表型鉴定数据，可能得到不同的估计值。从Falconer and Mackay (1996) 第162页表10.1来看，有些性状的狭义遗传力估计值还不到0.1，但有的可能会达到0.6以上。
- 影响遗传力的因素有很多，既有性状的遗传基础，又有环境和随机误差，还依赖于所采用的遗传群体。

狭义遗传力的群体特异性

- 同时，方差和协方差在统计学上被称为二阶统计量，二阶统计量的估计误差要远高于如样本均值、中位数等一阶统计量。遗传力是两个二阶统计量的比值，与相关系数类似，有着复杂的统计分布和较高的抽样误差。
- 因此，尽管遗传力是十分重要的遗传参数，但要得到遗传力的准确估计或正确估计还是相当不容易的。