

第11章

随机交配群体的选择与遗传进度

王建康

中国农业科学院作物科学研究所

wangjankang@caas.cn

<http://www.isbreeding.net>

本章的主要内容

- § 11.1 个体选择与遗传进度
- § 11.2 利用亲缘关系的选择与遗传进度
- § 11.3 性状相关和相关遗传进度
- § 11.4 多性状同时选择

§ 11.1 个体选择与遗传进度

- § 11.1.1 选择差和遗传进度
- § 11.1.2 选择比例和选择强度
- § 11.1.3 选择强度和遗传进度
- § 11.1.4 选择对等位基因频率的影响

育种过程中的选择

- 从自然或人工创建的遗传分离群体中，选择合乎人类需要的理想基因型，是动植物育种的主要内容。选择效果的好坏直接影响着育种工作的成败，如何制定有效的选择方案、提高选择效果，一直是育种中关心的重要课题。
- 育种家所能观察到的，通常都是个体的表型值，其中包含遗传和非遗传两方面的因素。这些个体通常都不是孤立的，它们之间存在着各种各样的亲缘关系。各种表型性状通常也不是孤立的，它们之间存在着各种各样的相关关系。
- 如果我们用前面几章的方法，把遗传方差和环境方差从表型方差中分离出来，就能够判断出被选择群体各种性状的变异程度。这样，不但可以预测表型选择所能产生的遗传效应，预测不同亲缘关系和相关性状所能产生的遗传效应，而且可以改进选择方案，提高育种成效。

随机交配群体中的混合选择

- 最简单的选择形式就是根据表型挑选满足一定条件的个体，将它们作为亲本进行互交形成下一世代的育种群体。这样的选择没有利用任何家系或亲本信息，个体表型是唯一的选择标准。因此称为个体选择（individual selection），有时也称混合选择（mass selection）或表型选择（phenotypic selection）。
- 控制数量性状的基因个数较多，个体的基因型不易区分，因此难以研究控制数量性状的基因频率在上下代之间的变化。但是，选择发生前后群体平均数的变化还是可以度量的，这就是遗传进度（genetic gain），也称选择响应。

遗传进度

- 等位基因频率定向改变，是最基本、最直接的选择效应。正是由于基因频率的改变，数量性状的群体均值才会在上下代之间出现差异。
- 所不同的是，控制数量性状的基因个数较多，个体的基因型不易区分，因此难以研究控制数量性状的基因频率在上下代之间的变化。但是，选择发生前后群体平均数的变化还是可以度量的，这就是遗传进度（genetic gain），也称选择响应（response to selection）。

遗传进度的公式表示

- 把中选的亲本个体 (x) 也看作一个群体，它们的随机交配后代 (y) 平均数与未经选择的随机交配群体平均数 μ 之差，就是遗传进度，一般用符号 R 或 ΔG 表示。

$$R = \bar{y} - \mu$$

- 遗传进度反映了经过选择之后，子代从亲代获得的增量，故又称遗传增量或遗传获得量。

选择差

- 为方便推导遗传进度的估计公式，把中选个体构成群体的平均数与未选择群体的平均数之差称为选择差（selection differential），用 S 表示。

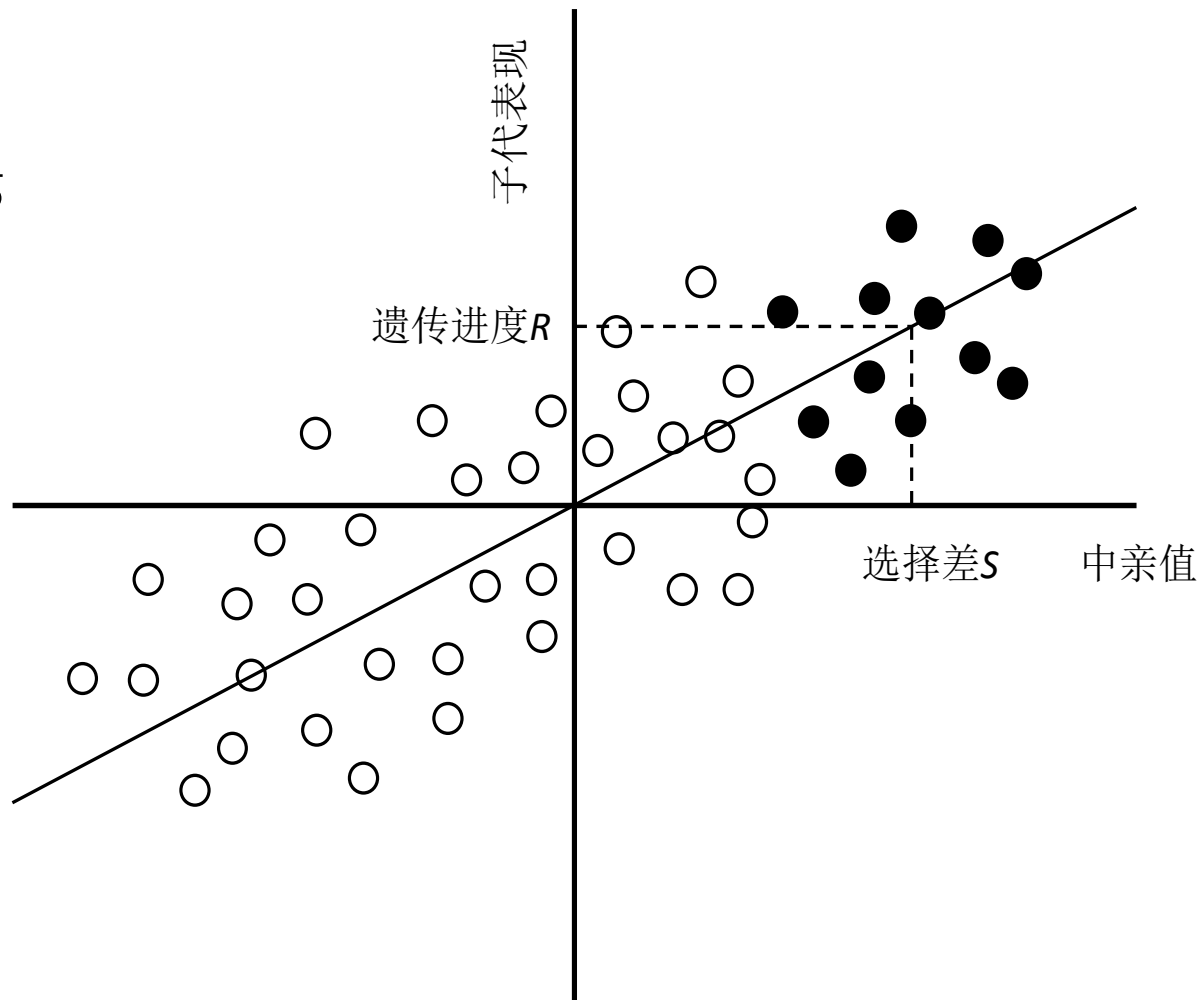
$$S = \bar{x} - \mu$$

选择差和选择响应的关系

$$\frac{\bar{y} - \mu}{\bar{x} - \mu} = b_{OP}$$

$$R = b_{OP} S$$

$$R = h^2 S$$



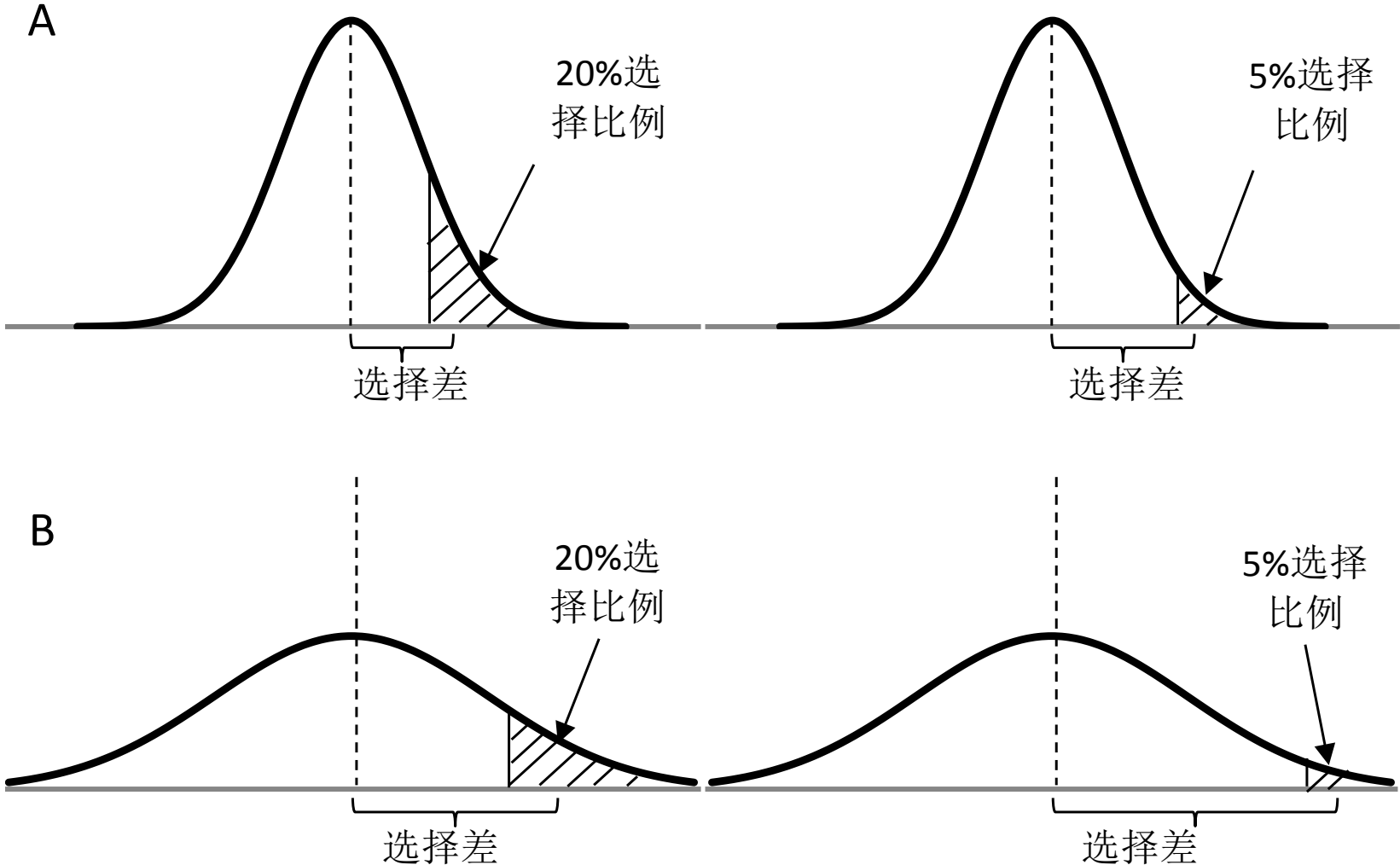
现实遗传进度与现实遗传力

- 在一些实验群体中，遗传进度还可以从上下代的平均表现直接进行计算，这样的遗传进度称为观测或现实遗传进度（observed genetic gain or realized genetic gain）。同时，通过预期遗传进度与观测遗传进度之间的对比，可以对数量遗传分析中的遗传模型进行适合性检验。
- 也可以用观测遗传进度与选择差的比值 R/S 来估计遗传力 h^2 。为便于区分，利用得到的遗传力称为现实遗传力（realized heritability）。

截尾选择

- 较大的选择差说明中选个体离群体均值越远，选择的标准越高；选择差越小，说明中选个体离群体均值越近，选择的标准越低。因此，选择差是选择强弱程度的一种反映，一般由育种家来决定。但在实际育种群体中，有时更容易确定的是中选个体的比例或百分数，而不是选择差。
- 因此，需要研究一定选择比例下的选择差和遗传进度。在表型服从正态分布的群体中，截尾选择的选择差可根据选择比例和分布的标准差计算出来。以高值选择为例，所谓截尾选择（truncation selection），就是从育种群体中挑选表型超过一定临界值的所有个体，作为中选个体，低于这个临界值的个体则被淘汰。

截尾选择差、选择比例和表型方差的关系



标准化选择差，即选择强度

- 标准化变换后的选择差，与方差大小没有关系，只依赖于选择比例的高低。我们把标准化后的选择差称为选择强度（intensity of selection），用符号*i*表示。公式右端，*S*为标准化前的选择差（公式11.2）， V_P 为群体的表型方差，开平方后称为表型标准差。

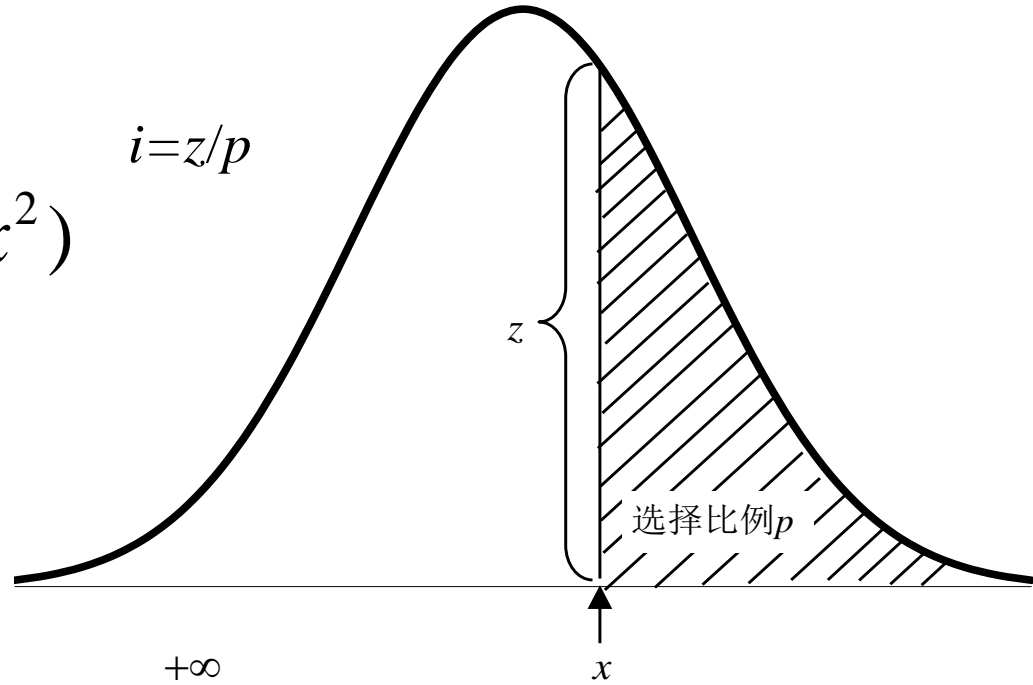
$$i = \frac{S}{\sqrt{V_P}}$$

选择强度的计算方法

$$z = f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$

$$i = z/p$$

$$\int_x^{+\infty} f(t) dt = p$$



$$i = \text{中选群体均值} = \frac{\int_x^{+\infty} tf(t) dt}{p} = \frac{\int_x^{+\infty} tf(t) dt}{\int_x^{+\infty} f(t) dt} = \frac{z}{p}$$

不同选择比例下的选择强度

选择比例p	0.5	0.4	0.3	0.2	0.1
x	0	0.2533	0.5244	0.8416	1.2816
z	0.3989	0.3863	0.3477	0.2800	0.1755
选择强度 <i>i</i>	0.7979	0.9659	1.1590	1.3998	1.7550

选择比例p	0.05	0.01	0.001	0.0001
x	1.6449	2.3263	3.0902	3.7190
z	0.1031	0.0267	0.0034	0.0004
选择强度 <i>i</i>	2.0627	2.6652	3.3671	3.9585

依赖于性别的选择强度和遗传进度

- 有时，选择在性别之间还可能存在差异，有时甚至只能对雌性个体进行选择，或者只能对雄性个体进行选择。不管两种性别的选择强度如何，中选雌亲和中选雄亲对后代的贡献各占一半。
- 因此，选择差等于两种性别选择差（分别用 S_f 和 S_m 表示）的简单平均，选择强度等于两种性别选择强度（分别用 i_f 和 i_m 表示）的简单平均。同样可以利用公式11.5计算后代群体的遗传进度。

$$S = \frac{1}{2}(S_f + S_m) \quad i = \frac{1}{2}(i_f + i_m) \quad R = h^2 S$$

选择强度和遗传进度

$$R = h^2 S = ih^2 \sqrt{V_P}$$

- 这样，在已知遗传力 h^2 和表型方差 V_P 的育种群体中，只要育种家提出一个选择比例，就能从公式11.9计算或表11.1中得到选择强度，代入上面的公式就得到遗传进度的估计值。
- 等价公式：

$$R = i \frac{V_A}{\sqrt{V_P}} \quad R = ih \sqrt{V_A}$$

提高遗传进度的途径分析

(1) 使用较小的选择比例，即提高选择强度

- 当选择比例为10%时，选择强度为1.76；选择比例为1%时，选择强度为2.67。因此，在其它参数一致的情况下，1%的遗传进度相对于10%的遗传进度的倍数等于 $2.67/1.76=1.52$ 。可见，选择比例从10%降低到1%时，遗传进度仅提高了大约50%。同时，采用较高的选择强度，需要增加被选择的个体或家系数。
- 小群体的随机飘变和近交程度较高，如存在近交衰退，也会严重影响下一代群体的表现和遗传变异。一般来说，用于重组下一轮群体的个体不应低于30。如需30个个体互交形成下一轮育种群体，10%的选择比例只需评价300个个体；如要采用1%的选择比例，则需要评价3000个个体。因此在实际育种中，选择强度的提高受群体大小的限制，选择强度不可能被无限地提高。

提高遗传进度的途径分析

(2) 提高加性方差在遗传方差中所占的比例

- 从第7和8章我们看到，不同亲缘关系群体中包含加性方差的倍数是有差别的，有的家系中只有1/2或1/4的 V_A ，但有些家系可有1倍甚至超过1倍的 V_A 。育种群体中，加性方差的倍数越高，遗传力就越大。
- 在动物育种中，可以通过产生全同胞家系的方法，来提高加性方差的倍数。在植物育种中，可以通过控制花粉和产生自交家系的方法，来提高加性方差的倍数。

提高遗传进度的途径分析

(3) 提高加性方差本身的大小

- 加性方差 V_A 本身的提高可以通过引入新的种质来实现。引入新种质的同时，也就引入了新的基因，从而引起遗传方差的增加。
- 新种质刚被引入的一段时间内，可能会由于有利基因与不利基因之间的连锁，造成群体平均数的下降。因此在开始的几个育种周期中，遗传进度可能不一定会高。

提高遗传进度的途径分析

(4) 降低非遗传方差

- 从表型所包含的方差成分来看，随机误差方差 V_{ε} 和互作方差 V_{GE} 任何程度的降低，都会引起遗传力的增加，进而提高选择进度。
- 随机误差方差 V_{ε} 可通过田间试验设计得以控制，如区组设计。在第9章我们看到，通过对目标环境群体的划分，还可以降低基因型 \times 环境互作方差 V_{GE} 。
- 另外，增加表型观测的次数，可以降低表型平均数中的随机误差，进而提高以重复平均数为选择单位的遗传力。

提高遗传进度的途径分析

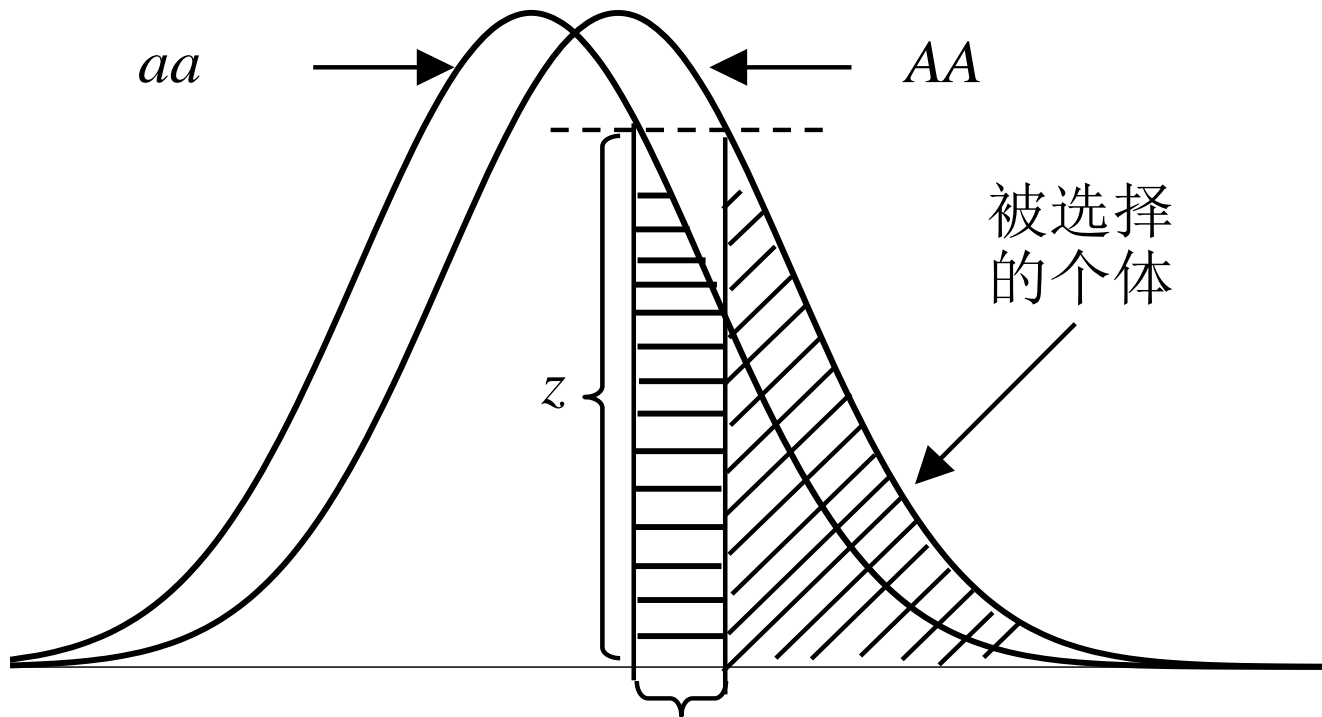
(5) 缩短育种周期

- 以上4个方面，提高的都是单个育种周期的遗传进度。有时，不同选择方法完成一个育种周期的时间存在差异，育种家更关心的可能还是单位时间所能取得的遗传进度。
- 单位周期遗传进度除以时间（一般以年为单位），称为年份遗传进度。育种中的加代，其实就是试图通过缩短育种周期的方法来提高年份遗传进度。

选择对等位基因频率的影响

- 选择前后群体均值的变化，其实是由被选择座位上基因频率的变化引起的。当控制性状的基因较多时，育种群体中难以区分不同个体的基因型，也难以追踪到选择对数量性状座位上等位基因频率的改变。
- 下面在单座位加显性模型的假定下，通过建立不同基因型的选择系数与遗传效应和选择强度之间的近似关系，然后利用第2章的相关结论，从理论的角度简单说明选择对数量性状基因频率的改变量。

一个数量性状座位上，两种纯合基因型的标准化分布曲线



$2a$ 除以表型分布的标准差

基因型 aa 相对于 AA 的适合度

$$1 - s \approx \frac{p - \frac{2a}{\sqrt{V_P}} z}{p} = 1 - \frac{2a}{\sqrt{V_P}} \frac{z}{p} = 1 - \frac{2a}{\sqrt{V_P}} i$$

- 选择系数与选择强度的关系

$$s \approx \frac{2a}{\sqrt{V_P}} i$$

等位基因*a*频率的改变量

- 完全显性 $\Delta q = -sq^2(1-q)$
- 加性模型 $\Delta q = -\frac{1}{2}sq^2(1-q)$

数量性状基因座位的 标准化效应

$$\frac{2a}{\sqrt{V_P}}$$

- 一个座位解释 F_2 群体12.5%的表型变异，如不考虑显性效应，标准化加性效应为0.5。50%选择比例对应的选择强度为0.7979，根据公式11.16得到的选择系数也是0.7979。起始等位基因 a 的频率为0.5，利用公式11.17得到基因频率的改变量为 $\Delta q = -0.0499$ 。
- 大多数数量性状基因，解释的表型变异可能只有百分之几，甚至更低，标准化效应一般为0.05~0.125。对QTL的选择系数在 $s=0.088$ 到 $s=0.334$ 之间。因此，对不利纯合基因型的选择系数较低，有利等位基因的频率提高和被固定是一个缓慢的过程。

§ 11.2 利用亲缘关系的选择与遗传进度

- § 11.2.1 亲缘关系与选择方法
- § 11.2.2 亲缘关系中的遗传力和遗传进度
- § 11.2.3 利用亲缘关系的选择方法
- § 11.2.4 多种亲缘关系的指数选择
- § 11.2.5 指数选择的遗传进度

亲缘关系与选择方法

- 除育种值外，个体的表型中还包含显性离差效应和各种互作效应，当然还有环境效应和随机误差效应，这些效应都不能传递到下一代。因此，个体水平的遗传力较低，选择的效果往往较差。
- 育种群体中，一般都包含各种各样的同胞和亲子等亲缘关系信息。在对一个育种群体进行选择时，除了利用被选择个体自身的表型数据外，有时还有大量的亲缘关系数据可供利用。
- 这些亲缘关系数据可以是来自全同胞、半同胞或自交等家系内其他个体的表型，也可以是来自个体的上一代亲本，甚至是来自个体的后代。

亲缘关系与选择方法

- 有时，利用亲缘信息还可以估计个体的育种值。育种值直接度量了后代群体的平均表现与亲本群体的差异，根据育种值的选择一般都会优于个体选择。
- 在动物育种中，有些性状只能在雌性或雄性个体中表现出来，如奶牛的产奶量。这时，对公牛产奶量的选择只能依赖于后代母牛的产奶量。如何利用各种亲缘关系数据预测育种值，提高选择效率是育种中的一个现实问题。

一个随机交配群体中5个全同胞家系的个体表型数据。只考虑遗传因素，不考虑环境效应、随机误差和抽样误差等非遗传因素

个体编号	全同胞家系编号				
	A	B	C	D	E
1	26	23	22	19	20
2	22	18	19	16	14
3	20	17	15	12	11
4	17	14	12	11	8
5	15	13	12	7	7
总均值	15.6				
总方差= V_A+V_D	24.64				

遗传方差的分解

个体编号	全同胞家系编号				
	A	B	C	D	E
家系平均数	20	17	16	13	12
家系间方差= $1/2 V_A + 1/4 V_D$	8.24				
家系内方差	14.8	12.4	15.6	17.2	22
家系内平均方差= $1/2 V_A + 3/4 V_D$	16.4				
加性方差= V_A	8.32				
显性方差= V_D	16.32				

表型效应的分解

- 表11.2的家系平均数与总平均数的离差，称为家系间效应，用 P_{bF} 表示；表型与家系平均数的离差称为家系内效应，用 P_{wF} 表示。因此，表11.2每个个体的表型 X 与总平均数的离差 P ，就可以分解为家系间效应 P_{bF} 和家系内效应 P_{wF} 两部分之和。

$$P = X - \mu \quad P = P_{bF} + P_{wF}$$

表11.2数据的家系间效应和家系内效应

效应类型		全同胞家系编号				
		A	B	C	D	E
家系间效应 (P_{bF})		4.4	1.4	0.4	-2.6	-3.6
家系内效应 (P_{wF})	1	6	6	6	6	8
	2	2	1	3	3	2
	3	0	0	-1	-1	-1
	4	-3	-3	-4	-2	-4
	5	-5	-4	-4	-6	-5

个体表型选择的结果

- 如果现在要从表11.2的25个个体中，选择5个个体作为亲本随机交配产生下一代群体，这时就会有多种可能的选择方法。
- 一是不考虑表11.2和表11.3的家系结构，只是根据个体的表型（即公式11.18中的 X 或 P ）选择最高的5个，也就是前一节介绍的个体选择或混合选择。显然，个体选择也可以被看作是根据 $P_{bF} + P_{wF}$ 的高低进行选择。**混合选择中选的前4个个体是A1、A2、B1、C1。A3和E1的表型相同，如没有家系信息可用时，只能从二者之中任选一个。**

家系间和家系内选择的结果

- 二是根据家系的平均表现（即公式11.18中的 \bar{P} ）或家系间效应（即公式11.18中的 P_{bF} ）进行选择，这样的选择称为家系间选择（between-family selection），**中选个体为家系A的全部5个个体。**
- 三是根据个体的家系内效应（即公式11.18中的 P_{wF} ）进行选择，这样的选择称为家系内选择（within-family selection），**中选个体为每个家系中表现最高的个体，即A1、B1、C1、D1、E1。**

后代测验的亲本选择

- 此外，根据表11.2的全同胞家系平均数，还可以反过来选择亲本。家系A的平均数最高，如只选择一组全同胞的亲本，中选的是家系A的两个亲本。当然，实际育种中也可以根据半同胞家系的平均表现来选择亲本。利用后代表现的亲本选择方法，称为后代测验（progeny testing）。
- 在有世代重叠的物种中，有时可以把表型选择的中选个体与后代测验的中选亲本混合在一起，形成下一世代的亲本群体。

后代测验的优缺点

- 后代测验对亲本的选择其实就是利用育种值进行选择，预期遗传进度也会比较高。但需要知道的是，后代测验的选择方法需要更长的育种周期。育种中除关注周期遗传进度外，更关注单位时间的遗传进度。有时，后代测验的单位时间遗传进度不一定是最高的。
- 因此，如何评价不同的选择方法，如何利用各种亲缘关系数据获得最好的遗传进度，是育种家经常面临的一个实际问题。

半同胞和全同胞家系中，家系间方差和家系内方差的构成

遗传参数	亲缘关系	
	半同胞	全同胞
家系间遗传方差中的加性成分	$V_{bA} = rV_A \quad (r = \frac{1}{4})$	$V_{bA} = rV_A \quad (r = \frac{1}{2})$
家系间遗传方差中的显性成分	0	$\frac{1}{4}V_D$
家系间相似性或相关系数	$t = \frac{\frac{1}{4}V_A}{V_P} = \frac{1}{4}h^2$	$t = \frac{\frac{1}{2}V_A + \frac{1}{4}V_D}{V_P} \quad (\frac{1}{2}h^2 \leq t \leq H^2)$
家系间表型方差	$V_{bP} = tV_P$	$V_{bP} = tV_P$
家系内遗传方差中的加性成分	$V_{wA} = (1-r)V_A$	$V_{wA} = (1-r)V_A$
家系内表型方差	$V_{wP} = (1-t)V_P$	$V_{wP} = (1-t)V_P$

家系间遗传力和家系内遗传力

$$h_{bF}^2 = \frac{V_{bA}}{V_{bP}} = \frac{rV_A}{tV_P} = \frac{r}{t} \frac{V_A}{V_P} = \frac{r}{t} h^2$$

$$h_{wF}^2 = \frac{V_{wA}}{V_{wP}} = \frac{(1-r)V_A}{(1-t)V_P} = \frac{1-r}{1-t} \frac{V_A}{V_P} = \frac{1-r}{1-t} h^2$$

家系间选择和家系内选择的遗传进度

$$R_{bF} = ih_{bF}^2 \sqrt{V_{bP}} = ih^2 \sqrt{V_P} \frac{r}{\sqrt{t}} = \frac{r}{\sqrt{t}} R$$

$$R_{wF} = ih_{wF}^2 \sqrt{V_{wP}} = ih^2 \sqrt{V_P} \frac{1-r}{\sqrt{1-t}} = \frac{1-r}{\sqrt{1-t}} R$$

表11.2数据的家系间选择和家系内 选择相对效率

$$h^2 \approx \frac{8.32}{24.64} = 0.3377 \quad r=0.5 \quad t \approx \frac{8.24}{24.64} = 0.3344$$

$$h_{bF}^2 = 0.5049 \quad h_{wF}^2 = 0.2537$$

$$R_{bF} = 0.8646R \quad R_{wF} = 0.6129R$$

利用亲缘关系的指数选择

- 从个体表型的育种值预测 $A = bP = h^2 P$
- 家系间效应和家系内效应均包含着个体育种值的信息。在这两部分效应无相关的条件下，以家系间和家系内遗传力为权重的线性组合，是个体育种值的最优估计。

$$I = h_{bF}^2 P_{bF} + h_{wF}^2 P_{wF} = \left(\frac{r}{t} P_{bF} + \frac{1-r}{1-t} P_{wF} \right) h^2$$

$$I = \mu + P + \frac{r-t}{t(1-r)} P_{bF} = X + \frac{r-t}{t(1-r)} P_{bF}$$

利用表11.2的数据得到的选择指数

公式11.26	A	B	C	D	E
1	30.36	24.39	22.40	16.43	16.43
2	26.36	19.39	19.40	13.43	10.43
3	24.36	18.39	15.40	9.43	7.43
4	21.36	15.39	12.40	8.43	4.43
5	19.36	14.39	12.40	4.43	3.43

公式11.25	A	B	C	D	E
1	3.74	2.23	1.72	0.21	0.21
2	2.73	0.96	0.96	-0.55	-1.31
3	2.22	0.71	-0.05	-1.57	-2.07
4	1.46	-0.05	-0.81	-1.82	-2.83
5	0.95	-0.31	-0.81	-2.83	-3.09

选择指数的方差、与育种值的协方差

$$\begin{aligned} V(I) &= \left(\frac{r}{t}\right)^2 V(P_{bF}) + \left(\frac{1-r}{1-t}\right)^2 V(P_{wF}) = \left(\frac{r}{t}\right)^2 t V_P + \left(\frac{1-r}{1-t}\right)^2 (1-t) V_P \\ &= \left[\frac{r^2}{t} + \frac{(1-r)^2}{1-t}\right] V_P \end{aligned}$$

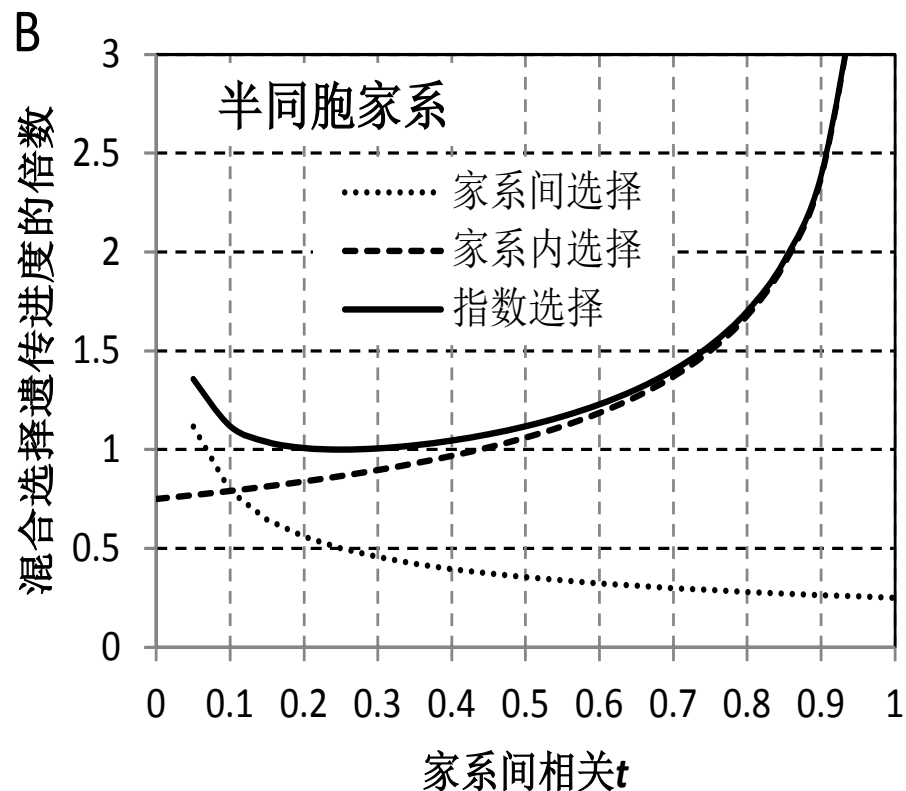
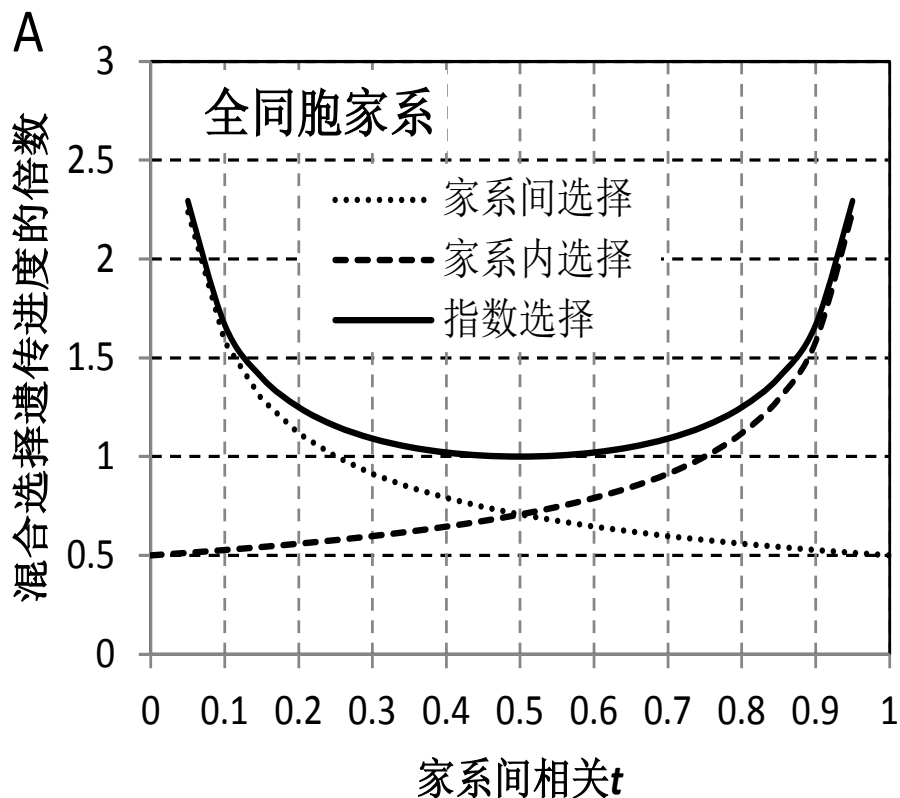
$$\begin{aligned} \text{Cov}(A, I) &= \frac{r}{t} \text{Cov}(A, P_{bF}) + \frac{1-r}{1-t} \text{Cov}(A, P_{wF}) = \frac{r}{t} r V_A + \frac{1-r}{1-t} (1-r) V_A \\ &= \left[\frac{r^2}{t} + \frac{(1-r)^2}{1-t}\right] V_A \end{aligned}$$

选择指数的遗传力和指数选择引起的遗传进度

$$h_I^2 = \frac{\text{Cov}(A, I)}{V(I)} = h^2$$

$$R_I = ih_I^2 \sqrt{V(I)} = \sqrt{\frac{r^2}{t} + \frac{(1-r)^2}{1-t}} R = \sqrt{1 + \frac{(r-t)^2}{t(1-t)}} R$$

相同强度下，三种选择方法相对于混合选择的遗传进度



多种亲缘关系数据的利用

- 实际育种群体中，亲缘关系的种类有很多种。除上一节介绍的全同胞和半同胞家系外，还有亲子关系、以及稍远一点的全同胞关系和亲子关系等等。
- 现假定有 m 个与个体育种值估计有关的观测数据，这些数据可以来自不同的亲缘关系个体，它们与群体均值的离差分别用 P_1 、 P_2 、 \dots 、 P_m 表示。这些数据有时会比个体表型 P 包含更多的遗传信息，因此可以获得更准确的个体育种值估计。
- 同时，对于没有表型的个体，我们也可以对它的育种值进行估计并进行选择。

多种亲缘关系的指数选择

- 利用多种亲缘关系数据的最优选择指数，是满足一定条件的线性组合，用下面的公式表示最优选择指数 I 。

$$I = b_1 P_1 + b_2 P_2 + \cdots + b_m P_m$$

选择指数的矩阵表示

- 推广到 n 个个体，每个个体都有一个选择指数，联合起来构成一个阶为 n 的向量或阶为 $n \times 1$ 的矩阵，用 \mathbf{I} 表示；每个个体都有对应的 m 种亲缘关系信息 $P_1、P_2、\dots、P_m$ ，联合起来构成一个阶为 $n \times m$ 的矩阵，用 \mathbf{P} 表示； m 个系数构成的参数向量用 \mathbf{b} 表示。这时就得到选择指数的矩阵表达式。

$$\mathbf{I} = \mathbf{b}^T \mathbf{P}$$

待估计个体的育种值

- 为避免与亲缘关系数据中包含的育种值混淆，同时又与后两节的内容保持一致，用 H 表示待估计个体的育种值。
- 使得前面公式给出最好的育种值估计，就要求选择指数 I 与待估育种值 H 之间的离差平方和最小。对于育种群体中的 n 个个体来说，即要求下面的离差平方和最小，其中 \mathbf{H} 表示 n 个个体待估育种值构成的列向量。

$$\text{离差平方和} = \|\mathbf{I} - \mathbf{H}\|^2 = \|\mathbf{b}^T \mathbf{P} - \mathbf{H}\|^2$$

- 如把 H 看作因变量、 I 看作自变量进行回归分析，前面的离差平方和，与多元回归模型公式6.74或公式6.75的离回归平方和是完全一致的。

选择指数中系数的估计

- 类似正规方程的公式6.76，得到选择指数中系数向量**b**的正规方程，其中，*P*和*H*均用群体平均数的离差表示。 $\mathbf{P}^T\mathbf{P}$ 就是变量*P*之间的方差协方差矩阵，用 $\text{Cov}(\mathbf{P}, \mathbf{P})$ 表示； $\mathbf{P}^T\mathbf{H}$ 就是变量*P*与待估育种值*H*之间的协方差向量，用 $\text{Cov}(\mathbf{P}, \mathbf{H})$ 表示。这样，得到另外一个等价的、用向量和矩阵表示的正规方程。

$$\mathbf{P}^T\mathbf{P}\mathbf{b} = \mathbf{P}^T\mathbf{H}$$

$$\text{Cov}(\mathbf{P}, \mathbf{P})\mathbf{b} = \text{Cov}(\mathbf{P}, \mathbf{H})$$

选择指数中系数的估计

- 由于育种值与非育种值之间的协方差为0，因此等于亲缘信息 P_i 中的育种值 A_i 与待估育种值 H 的协方差。这样就得到正规方程公式11.35的另一种形式及其展开形式。

$$\mathbf{Cov}(\mathbf{P}, \mathbf{P})\mathbf{b} = \mathbf{Cov}(\mathbf{A}, \mathbf{H})$$

$$\begin{bmatrix} \text{Cov}(P_1, P_1) & \text{Cov}(P_1, P_2) & \cdots & \text{Cov}(P_1, P_m) \\ \text{Cov}(P_2, P_1) & \text{Cov}(P_2, P_2) & \cdots & \text{Cov}(P_2, P_m) \\ \vdots & \vdots & \cdots & \vdots \\ \text{Cov}(P_m, P_1) & \text{Cov}(P_m, P_2) & \cdots & \text{Cov}(P_m, P_m) \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} \text{Cov}(A_1, H) \\ \text{Cov}(A_2, H) \\ \vdots \\ \text{Cov}(A_m, H) \end{bmatrix}$$

全同胞家系间和家系内两种亲缘关系数据的选择指数

$$P_1 = P_{bF} \quad P_2 = P_{wF}$$

$$\text{Cov}(P_1, P_1) = V_{bP} = tV_P \quad \text{Cov}(P_2, P_2) = V_{wP} = (1-t)V_P$$

$$\text{Cov}(P_1, P_2) = \text{Cov}(P_2, P_1) = 0$$

$$\text{Cov}(A_1, H) = V_{bA} = rV_A \quad \text{Cov}(A_2, H) = V_{wA} = (1-r)V_A$$

全同胞家系间和家系内两种亲缘信息的选择指数

$$\begin{bmatrix} tV_P & 0 \\ 0 & (1-t)V_P \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} rV_A \\ (1-r)V_A \end{bmatrix}$$

- 上面的结果与公式11.24完全相同。

个体与他的一个亲缘关系个体

- 假定在一个育种群体中，每个个体除了自身表型外，还有相同类型亲缘关系个体的表型。两种信息分别用 P_1 和 P_2 表示，需要估计每个个体的育种值 H 。

$$\text{Cov}(P_1, P_1) = V_P \quad \text{Cov}(P_2, P_2) = V_P$$

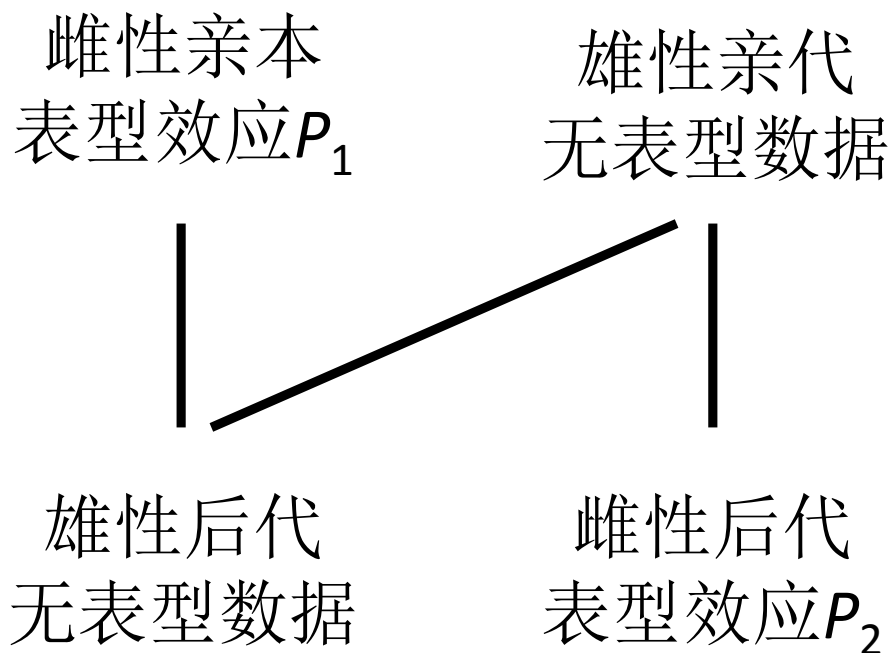
$$\text{Cov}(P_1, P_2) = \text{Cov}(P_2, P_1) = tV_P$$

$$\text{Cov}(A_1, H) = V_A = h^2V_P \quad \text{Cov}(A_2, H) = V_{wA} = rV_A = rh^2V_P$$

$$\begin{bmatrix} V_P & tV_P \\ tV_P & V_P \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} h^2V_P \\ rh^2V_P \end{bmatrix} \quad b_1 = \frac{h^2(1-rt)}{1-t^2} \quad b_2 = \frac{h^2(r-t)}{1-t^2}$$

无表型雄性个体的系谱

- 在另一个育种群体中，只有雌性才有表型数据，亲缘关系如下图，现要估计雄性后代个体的育种值。



无表型雄性个体的育种值估计

$$\text{Cov}(P_1, P_1) = V_P \quad \text{Cov}(P_2, P_2) = V_P$$

$$\text{Cov}(P_1, P_2) = \text{Cov}(P_2, P_1) = 0$$

$$\text{Cov}(A_1, H) = \frac{1}{2}V_A \quad \text{Cov}(A_2, H) = \frac{1}{4}V_A$$

$$I = \frac{1}{2}h^2P_1 + \frac{1}{4}h^2P_2$$

三种亲缘信息数据的选择指数

- 在表11.2的全同胞家系中，假定还有每个全同胞家系的中亲值数据。中亲与群体均值的离差用 P_1 表示， $P_2=P_{bF}$ ， $P_3=P_{wF}$ ，它们包含的育种值分别用 A_1 、 A_2 和 A_3 表示。现要估计全同胞后代家系中个体的育种值 H 。

$$Cov(P_1, P_1) = \frac{1}{2}V_P \quad Cov(P_1, P_2) = \frac{1}{2}V_A = \frac{1}{2}h^2V_P \quad Cov(P_1, P_3) = 0$$

$$Cov(P_2, P_2) = V_{bP} = tV_P \quad Cov(P_2, P_3) = 0$$

$$Cov(P_3, P_3) = V_{wP} = (1-t)V_P$$

三种亲缘信息数据的选择指数

- 数据中的育种值 A 与待估育种值 H 的斜方差

$$\text{Cov}(A_1, H) = \frac{1}{2}V_A \quad \text{Cov}(A_2, H) = V_{bA} = rV_A$$

$$\text{Cov}(A_3, H) = V_{wA} = (1-r)V_A$$

$$\begin{bmatrix} \frac{1}{2}V_P & \frac{1}{2}h^2V_P & 0 \\ \frac{1}{2}h^2V_P & tV_P & 0 \\ 0 & 0 & (1-t)V_P \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}V_A \\ rV_A \\ (1-r)V_A \end{bmatrix}$$

三种亲缘信息数据的选择指数

- 把表型方差、遗传力、加性方差和家系间相关的估计值，以及 $r=0.5$ 带入上面的方程组，就得到最终的选择指数。从三个系数的估计值可以看出，家系效应在选择指数中的重要性远高于中亲效应和家系内效应。

$$\begin{bmatrix} 12.32 & 4.16 & 0 \\ 4.16 & 8.24 & 0 \\ 0 & 0 & 16.4 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 4.16 \\ 4.16 \\ 4.16 \end{bmatrix} \quad \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 0.2016 \\ 0.4031 \\ 0.2537 \end{bmatrix}$$

$$I = 0.2016P_1 + 0.4031P_2 + 0.2537P_3$$

选择指数的方差及其与育种值的协方差

$$\begin{aligned} V_I &= V(I) = V(\mathbf{b}^T \mathbf{P}) = \mathbf{b}^T \mathbf{Cov}(\mathbf{P}, \mathbf{P}) \mathbf{b} \\ &= \mathbf{b}^T \mathbf{Cov}(\mathbf{P}, \mathbf{H}) = \mathbf{b}^T \mathbf{Cov}(\mathbf{A}, \mathbf{H}) \end{aligned}$$

$$\begin{aligned} \mathbf{Cov}(I, H) &= \mathbf{Cov}(\mathbf{b}^T \mathbf{P}, H) \\ &= \mathbf{b}^T \mathbf{Cov}(\mathbf{P}, \mathbf{H}) = \mathbf{b}^T \mathbf{Cov}(\mathbf{A}, \mathbf{H}) = V_I \end{aligned}$$

选择指数的遗传力及其与育种值的相关系数

$$h_I^2 = b_{IH} = \frac{\text{Cov}(H, I)}{V(H)} = \frac{V_I}{V_A}$$

$$r_{IH} = \frac{\text{Cov}(I, H)}{\sqrt{V_I V_H}} = \frac{V_I}{\sqrt{V_I V_A}} = \sqrt{\frac{V_I}{V_A}}$$

$$r_{IH}^2 = \frac{V_I}{V_A} = h_I^2$$

指数选择引起的遗传进度

$$R_I = ih_I^2 \sqrt{V_I} = ir_{IH} \sqrt{V_A} = ih \sqrt{V_A} \frac{r_{IH}}{h} = \frac{r_{IH}}{h} R$$

表11.2的数据为例

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \frac{r}{t} h^2 \\ \frac{1-r}{1-t} h^2 \end{bmatrix} = \begin{bmatrix} 0.5049 \\ 0.2537 \end{bmatrix} \quad \mathbf{Cov}(\mathbf{A}, \mathbf{H}) = \begin{bmatrix} rV_A \\ (1-r)V_A \end{bmatrix} = \begin{bmatrix} 4.16 \\ 4.16 \end{bmatrix}$$

$$V_I = \mathbf{b}^T \mathbf{Cov}(\mathbf{A}, \mathbf{H}) = 3.1554 \quad h_I^2 = \frac{V_I}{V_A} = 0.3793$$

$$r_{IH} = \sqrt{\frac{V_I}{V_A}} = 0.6158 \quad \frac{r_{IH}}{h} = 1.0598$$

- 选择指数的方差等于3.1554，遗传力等于0.3793，选择指数与育种值之间的相关系数等于0.6158。指数选择相对于混合选择的效率为1.0598，与公式11.30中R前面的系数给出的结果完全相等。

§ 11.3 性状相关和相关遗传进度

- § 11.3.1 遗传相关及相关系数的分解
- § 11.3.2 相关遗传进度的估计
- § 11.3.3 多个相关性状的指数选择
- § 11.3.4 相关遗传进度的育种应用

表型性状相关的遗传基础

- 许多表型性状在遗传或育种群体中，都表现出一定程度的相关。这时，对一个性状的选择，势必会影响到其它相关的性状。性状在表型上的相关既有遗传因素、也可能有环境因素，遗传研究中自然更关心的是遗传因素。
- 引起相关的遗传因素主要有两个，即一因多效和紧密连锁。一个基因同时影响到两个或多个性状，这样的现象在遗传上称为一因多效（pleiotropy）。控制两个性状的基因之间紧密连锁，也会引起这两个性状的相关。
- 一因多效和紧密连锁在遗传上是有本质区别的，一因多效引起的相关不受连锁不平衡的影响，连锁引起的相关随着连锁不平衡的降低而逐渐减弱。但是在很多情况下，要想从遗传上把一因多效和紧密连锁区分开来，还是相当困难的。

表型性状相关的遗传基础

- 另外需要注意的是，表型相关衡量的是所有基因的整体效应。当不同基因有着不同方向的一因多效、基因之间既存在相引又存在互斥连锁时，也许有时在表型上根本观测不到两个性状的相关。
- 同时，与方差和遗传力类似，相关也是一个群体概念。同样两个性状，它们在不同群体中的相关程度，甚至相关的方向都会有所差异。对于多基因控制的数量性状来说，表型上的相关还与基因数目、一因多效的方向、基因之间的连锁状态等因素有关。

表型相关和遗传相关

- 观察到的一般只是表型相关（phenotypic correlation），为研究相关性状的遗传进度，需要知道两个性状的育种值引起的遗传相关（genetic correlation），或称加性遗传相关（additive genetic correlation）。
- 另外，性状的显性离差之间也可能存在相关，上位性离差之间也可能存在相关。但它们与环境效应一样，不会传递给下一代，也不会影响遗传进度的计算。因此把育种值之外的效应，统称为非加性效应（non-additive effect），非加性效应与环境效应引起的相关，笼统地叫做环境相关（environmental correlation）。
- 没有特别说明的话，下面的遗传相关一般指的是育种值（即加性效应）引起的相关，环境相关指的是所有非加性效应和非遗传因素引起的相关。

遗传相关和环境相关的计算

- 性状1和性状2的遗传相关和环境相关

$$r_{A12} = \frac{\text{Cov}_{A12}}{\sqrt{V_{A1}V_{A2}}} \quad r_{E12} = \frac{\text{Cov}_{E12}}{\sqrt{V_{E1}V_{E2}}}$$

性状1和性状2表型相关的分解

$$\begin{aligned}r_{P12} &= \frac{\text{Cov}_{P12}}{\sqrt{V_{P1}V_{P2}}} = \frac{\text{Cov}_{A12} + \text{Cov}_{E12}}{\sqrt{V_{P1}V_{P2}}} \\&= \frac{\text{Cov}_{A12}}{\sqrt{V_{P1}V_{P2}}} \times \sqrt{\frac{V_{A1}V_{A2}}{V_{A1}V_{A2}}} + \frac{\text{Cov}_{E12}}{\sqrt{V_{P1}V_{P2}}} \times \sqrt{\frac{V_{E1}V_{E2}}{V_{E1}V_{E2}}} \\&= \sqrt{h_1^2 h_2^2} \times r_{A12} + \sqrt{(1-h_1^2)(1-h_2^2)} \times r_{E12} \\&= \sqrt{h_1^2 h_2^2} \times r_{A12} \quad (\text{当环相关等于0时})\end{aligned}$$

水稻抽穗期、株高的方差分析

均方期望值中的 r 表示重复次数，此例中的重复数 $r=3$

变异来源	自由度	均方		期望均方
		抽穗期	株高	
重复间	2	7.61	206.30	
家系间	49	213.52	191.32	V_E+rV_A
家系内	98	14.33	5.82	V_E

水稻抽穗期和株高的协方差分析

协均方期望值中的 r 表示重复次数，此例中的重复数 $r=3$

变异来源	自由度	协均方	期望协均方
		抽穗期×株高	
重复间	2	65.50	
家系间	49	146.16	$\text{Cov}_{E12} + r\text{Cov}_{A12}$
家系内	98	0.74	Cov_{E12}

水稻抽穗期和株高的方差成分、协方差成分和遗传力，以及它们之间协方差和相关系数

性状	遗传方差	环境方差	表型方差	遗传力
抽穗期	66.4	14.33	80.73	0.8225
株高	61.83	5.82	67.65	0.9140
协方差	48.47	0.74	49.21	
相关系数	0.7565	0.0810	0.6659	

- 本例中，非遗传效应引起的相关很小，表型相关系数基本上反映了遗传相关系数。
- 对于两个遗传相关的性状，对一个性状的选择势必会引起另一个性状的变化。当一个性状遗传力较低时，通过高度相关、遗传力较高性状的选择，获得的间接遗传进度还有可能高于直接选择的遗传进度。

两个性状的直接遗传进度

$$R_1 = i_1 h_1 \sqrt{V_{A1}} \quad R_2 = i_2 h_2 \sqrt{V_{A2}}$$

- 对两个被选择的性状（称性状1和性状2），上面的公式是对它们进行直接选择的遗传进度。
- 可以看出，如果性状（如产量等）的遗传力较低时，直接选择的效果并不好。

性状1选择对性状2的相关遗传进度

- 为了与直接选择的遗传进度相区别，把由性状1直接选择得到性状2的遗传进度称为相关遗传进度（correlated genetic gain），有时也称间接遗传进度（indirect genetic gain），用 R_2^* 表示。
- 为得到性状1选择对性状2产生的间接遗传进度，首先计算性状2育种值对性状1育种值的回归系数。

$$b = \frac{Cov_{A12}}{V_{A1}} = r_{A12} \sqrt{\frac{V_{A2}}{V_{A1}}}$$

性状1选择对性状2的相关遗传进度

- 性状1遗传进度 R_1 乘以回归系数，正好就是性状2的遗传进度。由此得到性状2间接遗传进度的计算公式。

$$R_2^* = bR_1 = r_{A12} \sqrt{\frac{V_{A2}}{V_{A1}}} \times i_1 h_1 \sqrt{V_{A1}} = i_1 r_{A12} h_1 \sqrt{V_{A2}}$$

$$R_2^* = i_1 r_{A12} h_1 h_2 \sqrt{V_{P2}}$$

相关遗传力 (co-heritability)

$$h^2 = r_{A12} h_1 h_2$$

$$R_2^* = i_1 r_{A12} h_1 h_2 \sqrt{V_{P2}} = ih \sqrt{V_{P2}}$$

相关遗传进度与直接遗传进度的比值

$$\frac{R_2^*}{R_2} = r_{A12} \frac{i_1 h_1}{i_2 h_2}$$

- 可以看出，在相同选择强度的情况下（即 $i_1=i_2$ ），如果性状1和2之间有较高的遗传相关，同时性状2有较低的遗传力、性状1有较高的遗传力，即满足 $r_{A12} h_1 > h_2$ 的条件时，上面的比值大于1，选择性状1而对性状2产生的间接遗传进度，就会超过对性状2进行直接选择产生的遗传进度。

利用表11.7和表11.8数据，计算性状 2（单株穗重）的遗传进度

$$r_{A12} = 0.76 \quad h_1^2 = 0.82 \quad h_2^2 = 0.91$$

$$V_{P2} = 67.65 \quad V_{A2} = V_{P2}h_2^2 = 61.83$$

$$R_2^* = 2.06 \times 0.76 \times \sqrt{0.82 \times 0.91 \times 67.65} = 11.11$$

$$R_2 = 2.06 \times \sqrt{0.91 \times 61.83} = 15.49$$

多个相关性状的选择指数

- 假定现有 m 个相关性状的表型观测数据，它们与群体均值的离差分别用 P_1, P_2, \dots, P_m 表示。目标性状的表型效应 P 也是观察值与群体均值的离差，它可以包含在 m 个相关性状中，也可以不含其中。我们的目标是如何利用 m 个相关性状的信息，实现对目标性状的最优选择，以获得最高的遗传进度。用下面的公式表示多个相关性状的最优选择指数。

$$I = b_1 P_1 + b_2 P_2 + b_3 P_3 + \dots + b_m P_m$$

- 与公式11.31是完全相同的，所不同的只是效应代表的含义有所差异。公式11.31中的效应来自不同的亲缘关系个体，这里的效应则来自同一个个体在不同性状上的表型。

4个水稻性状的表型相关和表型方差

性状	表型相关和表型方差			
	抽穗期	株高	百粒重	穗粒重
抽穗期	1	0.67	-0.38	0.64
株高	0.67	1	-0.24	0.73
百粒重	-0.38	-0.24	1	-0.05
穗粒重	0.64	0.73	-0.05	1
表型方差	80.73	61.83	15.45	25.24

4个水稻性状的遗传相关和遗传方差

性状	遗传相关和遗传方差			
	抽穗期	株高	百粒重	穗粒重
抽穗期	1	0.76	-0.63	0.95
株高	0.76	1	-0.44	0.94
百粒重	-0.63	-0.44	1	-0.55
穗粒重	0.95	0.94	-0.55	1
遗传方差	66.40	49.21	10.45	15.24

利用4个相关性状的性状1选择指数

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} 80.73 & 47.34 & -13.42 & 28.89 \\ 47.34 & 61.83 & -7.42 & 28.84 \\ -13.42 & -7.42 & 15.45 & -0.99 \\ 28.89 & 28.84 & -0.99 & 25.24 \end{bmatrix}^{-1} \begin{bmatrix} 66.40 \\ 43.44 \\ -16.60 \\ 30.22 \end{bmatrix} = \begin{bmatrix} 0.54 \\ -0.09 \\ -0.61 \\ 0.66 \end{bmatrix}$$

$$I = 0.54P_1 - 0.09P_2 - 0.61P_3 + 0.66P_4$$

相关遗传进度的育种应用

- 育种中一些目标性状的改良可以通过其他相关性状的选择来实现。这样的选择又称相关选择（correlated selection）或间接选择（indirect selection），这些受选择的非育种目标性状，有时也称次级性状（secondary trait）。
- 显然，当目标性状遗传力较低或难以准确度量时，直接选择的效果就会很差。这时可以考虑选择与目标性状高度相关、同时遗传力又较高的性状，通过相关遗传进度来实现对目标性状的改良。

相关遗传进度的育种应用

- 例如，单株产量在育种早期世代的遗传力很低、测量误差大，直接选择的效果不好。而一些与光合作用有关的生理性状可能与产量存在一定程度的正相关、又易于精确测量。因此，在早期分离群体中，可以通过对这些性状的选择，来间接提高最终的单株产量和总产量。
- 产量构成性状的遗传力一般都远高于产量本身的遗传力，同时又与产量存在一定的正相关。因此，在育种早期世代，还可以通过对产量构成性状进行选择，以间接提高最终的产量。
- 另外，当目标性状的表型鉴定成本较高时，也可以考虑选择与目标性状高度相关、同时又容易测量的性状，以提高育种的成本收益。

相关遗传进度的育种应用

- 早期世代的选择一般在育种试验田里进行，但最终品种需要种植在农民的土地上。如果把产量在育种试验田中的表现看作性状1，在农民土地里的表现看作性状2。育种家选择的是性状1，但其目标其实是希望在性状2上取得较高的遗传进度。
- 要达到这一目的，就要求性状1和2之间存在较高的正向遗传相关。如果产量在育种试验田的表现与在农民土地上的表现没什么关系，即性状1和2之间没有太大的遗传相关，其结果必然是，育种试验田中表现很好的基因型，将其种植到农民的土地上却不一定有很好的产量表现。

相关遗传进度的育种应用

- 其实，我们还可以把育种试验田看作一种环境，农民的土地和种植方式看作另外一种环境。只要这两个环境之间有较高的相似性，就可以期望一个育种性状在这两个环境间的表现高度相关。
- 因此，育种家需要对选育品种的最终种植环境进行仔细分析，种植育种材料和进行选择的环境条件，要尽可能接近未来品种的种植环境。只有这样，育种过程中选择产生的遗传进度，才能在未来的农业生产中体现出来。

§ 11.4 多性状同时选择

- § 11.4.1 多性状同时选择的最优指数
- § 11.4.3 最优选择指数的遗传进度
- § 11.4.3 选择指数的应用以及与其他类型的指数

多性状选择方法

- 单项选择（tandem selection）：对于所要改良的性状，一个世代一个世代地依次进行选择、改良，改良完一个性状后再来改良下一个。
- 独立水平选择（independent culling level）：每一个世代中，对所有目标性状同时进行选择，对每一个性状制订出一个独立的中选（或淘汰水平），某个性状达到它的水平就予以保留，否则就被淘汰。
- 指数选择（index selection）：每一个世代中，利用所有性状的遗传力、遗传和表型相关关系、以及经济权重，构造出一个综合选择指数，对这一指数进行单项指标的选择。

多性状的综合改良

- 构成优良品种的条件是综合性的，除产量外，还要注意到早熟、优质、抗逆、适应性等特性。这些特性在育种实践中的重要性，也因作物、生态区域、育种材料等而有所不同，应分轻重缓急综合考虑。特别是一些性状间存在着负相关时，尤应如此。
- 在考虑对多个性状的综合改良时，就会牵涉到多个性状同时选择（multiple trait selection）的问题。

综合改良的育种目标

- 现有 m 个有待同时改良的性状，表型观测值与群体均值的离差分别用 P_1, P_2, \dots, P_m 表示。这些性状在加性遗传效应和环境效应上，可以存在各种各样的相关关系。
- 这些性状的育种值分别用 A_1, A_2, \dots, A_m 表示，权重分别用 w_1, w_2, \dots, w_m 表示。性状的权重可以用它们在育成品种中的相对重要性表示，也可以用这些性状的经济价值表示。
- 例如，在大豆育种中，需要同时提高籽粒油分和蛋白质含量，如果高油分的重要性超过高蛋白，这时可把油分含量的权重设为1、蛋白质含量的权重设为0.6或其他小于1的数值。如果大豆籽粒加工后，豆油制品的市场价格是5元/kg、豆蛋白制品的市场价格是10元/kg，这时可把两个性状的权重分别设为5和10。

综合改良的育种目标

- 多个性状的同时改良，其实是期望在下面的综合指标 H 上取得最大增益。多性状同时选择的问题，就变成如何利用 m 个相关性状的信息，构造出一个最优选择指数 I ，使得选择指数 I 与综合指标 H 之间有最高的遗传相关关系。

$$H = w_1A_1 + w_2A_2 + \cdots + w_mA_m = \mathbf{w}^T \mathbf{A}$$

$$I = b_1P_1 + b_2P_2 + b_3P_3 + \cdots + b_mP_m = \mathbf{b}^T \mathbf{P}$$

- 由于选择指数 I 与综合指标 H 之间的高度遗传相关，利用指数 I 进行选择时，就能保证在指标 H 上取得最好的遗传或经济效益。

多性状选择指数的计算

- 因此，选择指数系数的估计问题，就变成求解离差平方和的最小二乘估计，也是综合指标 H 对表型数据多元回归模型的系数估计问题。

$$\text{离差平方和} = \|\mathbf{I} - \mathbf{H}\|^2 = \|\mathbf{b}^T \mathbf{P} - \mathbf{w}^T \mathbf{A}\|^2$$

$$\mathbf{P}^T \mathbf{P} \mathbf{b} = \mathbf{P}^T \mathbf{A} \mathbf{w}$$

多性状选择指数的计算

- 正规方程中， $\mathbf{P}^T\mathbf{P}$ 就等于表型方差协方差矩阵， $\mathbf{P}^T\mathbf{A}$ 就等于表型与育种值的协方差矩阵，又等于育种值方差协方差矩阵。这样就得到另外一个更常用的估计系数向量 \mathbf{b} 的公式。

$$\mathbf{P}^T\mathbf{P} = \mathbf{Cov}(\mathbf{P}, \mathbf{P}) \quad \mathbf{P}^T\mathbf{A} = \mathbf{Cov}(\mathbf{P}, \mathbf{A}^T) = \mathbf{Cov}(\mathbf{A}, \mathbf{A})$$

$$\mathbf{Cov}(\mathbf{P}, \mathbf{P})\mathbf{b} = \mathbf{Cov}(\mathbf{A}, \mathbf{A})\mathbf{w}$$

$$\mathbf{b} = [\mathbf{Cov}(\mathbf{P}, \mathbf{P})]^{-1} \mathbf{Cov}(\mathbf{A}, \mathbf{A})\mathbf{w}$$

- 利用 \mathbf{I} 进行选择时，可以在综合指标 \mathbf{H} 上取得最高的遗传增益，有时又称最优选择指数（optimum selection index）。

4个水稻性状的最优选择指数

$$\mathbf{Cov}(\mathbf{P}, \mathbf{P}) = \begin{bmatrix} 80.73 & 47.34 & -13.42 & 28.89 \\ 47.34 & 61.83 & -7.42 & 28.84 \\ -13.42 & -7.42 & 15.45 & -0.99 \\ 28.89 & 28.84 & -0.99 & 25.24 \end{bmatrix}$$

$$\mathbf{Cov}(\mathbf{A}, \mathbf{A}) = \begin{bmatrix} 60.40 & 43.44 & -16.60 & 30.22 \\ 43.44 & 49.21 & -9.98 & 25.74 \\ -16.60 & -9.98 & 10.45 & -6.94 \\ 30.22 & 25.74 & -6.94 & 15.24 \end{bmatrix}$$

4个水稻性状的最优选择指数

$$\mathbf{Cov}(\mathbf{P}, \mathbf{P})^{-1} \mathbf{Cov}(\mathbf{A}, \mathbf{A}) = \mathbf{B} = \begin{bmatrix} 0.54 & 0.03 & -0.04 & 0.16 \\ -0.09 & 0.60 & 0.09 & 0.16 \\ -0.61 & -0.31 & 0.67 & -0.22 \\ 0.66 & 0.28 & -0.31 & 0.22 \end{bmatrix}$$

- 矩阵 \mathbf{B} 的4个列向量分别对应于抽穗期、株高、百粒重和穗粒重这4个性状的选择指数。例如，矩阵 \mathbf{B} 第1列为抽穗期单独作为育种目标的选择指数向量，已在 § 11.3.3 中给出。

两种权重向量的最优选择指数

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \mathbf{B} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.70 \\ 0.76 \\ -0.47 \\ 0.87 \end{bmatrix}$$

- 4个性状等权重

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \mathbf{B} \begin{bmatrix} -0.5 \\ 0 \\ 0.3 \\ 1 \end{bmatrix} = \begin{bmatrix} -0.12 \\ 0.23 \\ 0.29 \\ -0.20 \end{bmatrix}$$

- 抽穗期权重设为-0.5
- 株高权重设为0
- 穗粒重权重设为1
- 百粒重权重设为0.3

最优选择指数的遗传进度

- 利用最优指数进行选择之后，我们关心的遗传进度有两种类型。一是对综合指标 H 产生的遗传进度，二是对单个构成性状产生的遗传进度。
- 首先考虑综合指标 H 的遗传进度，用 R_H 表示。用 S 表示最优指数 I 的选择差，即中选个体的平均指数与未选择群体的平均指数之差。这样，如果能够估计出综合指标 H 对选择指数 I 的回归系数 b_{HI} ，显然就可以利用下面的公式计算 H 的遗传进度。

$$R_H = b_{HI} S_I$$

综合指标 H 对选择指数 I 的回归系数 b_{HI}

$$V_I = V(\mathbf{b}^T \mathbf{P}) = \mathbf{b}^T \mathbf{Cov}(\mathbf{P}, \mathbf{P}) \mathbf{b} = \mathbf{b}^T \mathbf{Cov}(\mathbf{A}, \mathbf{A}) \mathbf{w}$$

$$\begin{aligned} \text{Cov}(I, H) &= \text{Cov}(\mathbf{b}^T \mathbf{P}, \mathbf{A}^T \mathbf{w}) = \mathbf{b}^T \mathbf{Cov}(\mathbf{P}, \mathbf{A}) \mathbf{w} \\ &= \mathbf{b}^T \mathbf{Cov}(\mathbf{A}, \mathbf{A}) \mathbf{w} = V_I \end{aligned}$$

$$b_{HI} = \frac{\text{Cov}(I, H)}{V_I} = 1 \quad R_H = b_{HI} S = S = i \sqrt{V_I}$$

单个性状的遗传进度

- 与综合指标 H 的计算方法类似，如果能够得到每个性状的育种值对最优选择指数 I 的回归系数，也就可以计算每个性状的遗传进度。

$$\begin{aligned}\text{Cov}(I, A_j) &= \text{Cov}(b_1P_1 + b_2P_2 + \cdots + b_mP_m, A_j) \\ &= \sum_{i=1}^m \text{Cov}(b_iP_i, A_j) = \sum_{i=1}^m b_i \text{Cov}(P_i, A_j) = \sum_{i=1}^m b_i \text{Cov}(A_i, A_j)\end{aligned}$$

单个性状的遗传进度

- 与综合指标 H 的计算方法类似，如果能够得到每个性状的育种值对最优选择指数 I 的回归系数，也就可以计算每个性状的遗传进度。

$$b_{A_j, I} = \frac{\text{Cov}(I, A_j)}{V_I}$$

$$R_j = b_{A_j, I} S_I = \frac{\text{Cov}(I, A_j)}{V_I} i \sqrt{V_I} = \frac{i}{\sqrt{V_I}} \text{Cov}(I, A_j)$$

综合指标 H 的方差、遗传力

$$V_H = V(\mathbf{A}^T \mathbf{w}) = \mathbf{w}^T \mathbf{Cov}(\mathbf{A}, \mathbf{A}) \mathbf{w}$$

$$r_{IH} = \frac{\text{Cov}(I, H)}{\sqrt{V_I V_H}} = \frac{\text{Cov}(I, H)}{V_I} \sqrt{\frac{V_I}{V_H}} = \sqrt{\frac{V_I}{V_H}}$$

$$h_H^2 = r_{IH}^2 = \frac{V_I}{V_H} \quad R_H = i r_{IH} \sqrt{V_H} = i h_H \sqrt{V_H}$$

4个性状权重均为1的最优选择指数

- 假定选择出指数最高的10%个体，即选择强度 $i=1.755$ 。
- 从公式11.61得到选择指数 I 的方差 $V_I=234.74$ ，从公式11.64得到综合指标 H 的遗传进度为26.89。
- 选择指数 I 对4个构成性状育种值的协方差分别为113.11、94.60、-30.01、57.03，从公式11.67得到抽穗期、株高、百粒重、穗粒重的遗传进度分别为12.97、10.84、-3.43、6.53。因此，如果性状之间有比较复杂的相关关系，在利用最优指数进行选择时，正的权重不一定代表在这个性状上就一定会有正的遗传进度。
- 从公式11.68得到综合指标 H 的方差等于273.08，选择指数 I 的方差等于234.74。因此，选择指数的遗传力等于0.86，选择指数与综合指标的相关系数等于0.93。

选择指数的育种应用

- 前面介绍的最优选择指数主要是Hazel和Lush（1942）以及Hazel（1943）的工作。最优选择指数自然有其理论上的优越性，但从前面的例子来看，计算最优选择指数的过程还相当繁琐。
- 一方面，最优选择指数依赖于性状的权重，育种家对大多数目标性状的相对重要性一般都会心里有数，但要给出一个具体的权重数值可能还很困难。另外，性状的权重还可能因时间而发生变化。当一个性状经过一段时间的有效改良后，它的重要性可能就会下降，原来不太重要的性状随着时间和环境的变化可能会变得重要起来。

选择指数的育种应用

- 同时，选择既引起单个性状遗传方差和遗传力的改变，又会引起性状之间相关关系的改变。在多基因的情况下，同时选择会降低性状间的相关，并使得原来的正相关逐步变为负相关。构建最优选择指数时，需要利用育种群体的遗传参数。选择指数的优良性质是针对特定群体而言。
- 一个育种群体构建出的最优指数不一定适用于其它群体，不同的育种群体需要构建出不同的指数，以用于选择。

选择指数的育种应用

- 同时还要注意，单个性状的遗传进度不一定与权重的方向完全一致。当构成性状较多、性状间相关关系比较复杂时，构成性状不一定能够在预期的方向上发生改变。
- 尽管具有理论上的优越性，但由于上述的这些原因，最优选择指数只是在少数大型动物的育种过程中得以应用。植物育种群体具有个体多、单个个体经济价值低、世代短、选择强度高等特点，最优指数选择并没有得到普遍的应用。很多时候，育种家采用的选择方法仍以独立水平选择为主，但同时注重家系间和家系内两种水平的选择。

其他类型的指数

- 最优选择指数的构建，需要估计遗传协方差矩阵。育种中真正要利用它开展选择并不是一件容易的事。同时，选择后的群体要重建最优选择指数。鉴于此，人们随后又提出了一些计算简便的指数。
- Williams (1962) 提出一个基础指数 (base index)，它直接用性状的经济权重表示出来。 w_1, w_2, \dots, w_m 为 m 个性状的经济权重，由育种家根据目标性状的重要性来确定； P_1, P_2, \dots, P_m 表示表型观测值与群体均值的离差。

$$I = w_1 P_1 + w_2 P_2 + \dots + w_m P_m$$

- 在基础指数中，一般把最重要性状的权重设为1。根据相对重要性，其他性状的权重在0和1之间取值。

其他类型的指数

- 最优指数和基础指数都需要知道性状的经济权重，Elston（1963）提出一个无权重的乘积指数，见公式11.73。其中，对应于第*i*个性状的最低可接收的性状值， X_1, X_2, \dots, X_m 表示表型观测值。

$$I = (X_1 - l_1)(X_2 - l_2) \cdots (X_m - l_m) = \prod (X_i - l_i)$$

其他类型的指数

- 在正规方程（公式11.59）中，如果忽略不同性状之间的表型协方差和遗传协方差，表型方差协方差就变成一个对角阵，对角线元素为性状的表型方差；遗传方差协方差也变成一个对角阵，对角线元素为性状的加性方差。公式11.74给出这种情况下的选择指数，从而避免了协方差的估计和计算逆矩阵。

$$I = w_1 h_1^2 P_1 + w_2 h_2^2 P_2 + \cdots + w_m h_m^2 P_m$$