

育种实用统计学

(第十二版)

2023.03

北京中农博思科技发展有限公司 版权所有

致读者：

北京中农博思科技发展有限公司曾在京举办免费的育种统计培训班，后考虑外地客户来京不方便及农时、疫情因素，也为了节省大家时间和费用，免费育种培训班就衍变成推出免费《育种实用统计学》、《育种实用统计学视频教程》，将博思公司育种技术研究成果与育种者交流分享，期望推动国内育种统计应用、提升育种选择效率效果。

这是一个免费的版本，欢迎阅读，欢迎购买第 13 版。

自第 13 版起，《育种实用统计学》PDF 文件采用付费购买阅读方式对读者提供。

前言

一、相比第十一版，增加修改了以下内容：

1、遗传力概念澄清

遗传力（Heritability）这个叫法没有问题，比遗传率更符合概念内涵。
遗传力本质是回归。

2、重复力（repeatability）

3、有重复观测值时的 BLUP 模型

4、植物育种采用动物育种 BLUP 方法的技术理论问题探讨

4.1、动物 BLUP 拿到植物上用，可行吗？依据是什么？

4.2、近交系数和亲缘系数

4.3、A 矩阵介绍

4.4、A 矩阵构建规则

4.5、A 矩阵示例

4.6、A 矩阵（系谱）的整理规则

4.7、植物育种要重视育种值和 BLUP 分析方法

4.7.1 遗传力和遗传率的概念困扰

4.7.2 自交和近交的概念困扰

4.7.3 A 矩阵的构建困扰

4.7.4 BLUP 要用到材料系谱，植物系谱不记录自交情况下，适合采用吗？

4.7.5 BLUP 本身真的很难，配合力又很简单

新增修改内容标题，黄底色醒目标注，方便查看。水平有限，有写的不对的地方，欢迎读者批评指正！

前言.....	3
一、试验设计.....	6
(一) 试验设计的原则.....	6
1、重复.....	6
2、随机.....	8
3、局部控制.....	8
(二) 完全区组.....	8
(三) 不完全区组.....	8
1、平衡不完全区组设计 (BIB 设计).....	8
2、格子设计.....	11
3、 α -格子设计.....	13
(四) 增广试验设计.....	15
1、什么是增广试验设计?.....	15
2、增广试验设计在育种中何时采用?.....	15
3、增广随机完全区组设计示例.....	16
4、修饰增广试验设计.....	16
5、增广试验设计的方差分析.....	19
(五) 试验设计的发展趋势.....	19
(六) 试验方法的选用原则.....	19
二、方差分析.....	19
(一) 离差.....	19
(二) 方差 (均方).....	20
(三) 自由度.....	20
(四) F 分布.....	20
(五) F 测验.....	21
(六) 单因素随机完全区组试验方差分析示例.....	21
(七) 方差分析缺区数据分析结果比较.....	24
(八) 方差分析数据“缺区”应对策略讨论.....	24
(九) 缺区方差分析示例.....	26
(十) 固定效应和随机效应.....	28
(十一) 理解方差分析.....	29
(十二) 一次重复的多点试验如何进行方差分析?.....	30
(十三) 区域试验中的变异系数 (CV) 用于评价试验质量.....	32
(十四) 没有重复的多点试验怎么分析.....	32
三、线性模型.....	33
(一) 线性模型之方差分析模型.....	33
(二) 两种方差分析方法.....	33
(三) 混合线性模型.....	34
四、遗传力 (Heritability).....	34
(一) 遗传力定义.....	34
(二) 组内相关法估算广义遗传力.....	35
(三) 实际育种中遗传力育种应用思考.....	35
五、重复力 (Repeatability).....	36
六、选择育种.....	36
(一) 选择差和选择响应.....	36

(二) 差值选择育种法	38
七、配合力	39
(一) 配合力定义	39
(二) 配合力计算示例	39
(四) griffing 配合力分析介绍	44
(四) 理解配合力	45
(五) 配合力育种价值思考	46
1、植物杂交育种中的配合力方法实质	46
2、育种者的配合力幻想	46
八、育种值	46
(一) 育种值定义	46
(二) 理论定义与实际定义辨析	48
(三) 配合力 \neq 基因加性效应(育种值)	48
九、经验式育种的选择技术方法	49
(一) 经验式育种的困惑	49
(二) 经验式育种选择使用的是均值	49
十、BLUP 分析方法	50
(一) BLUP 介绍	50
(二) 单性状 BLUP 育种值计算示例	51
(三) 多性状植物育种 BLUP 分析模型	56
(四) 植物育种 BLUP 模型	57
(五) 植物育种 BLUP 模型中种群效应	57
(六) 有重复观测值时的 BLUP 模型	58
(六) BLUP 植物育种模型用于一般配合力分析	58
(七) 植物育种中育种值应用思路参考	59
(八) 植物育种采用动物育种 BLUP 方法的技术理论问题探讨	59
1、动物 BLUP 拿到植物上用, 可行吗?依据是什么?	59
2、近交系数和亲缘系数	60
3、A 矩阵介绍	61
4、A 矩阵构建规则	61
5、A 矩阵示例	62
6、A 矩阵(系谱)的整理规则	62
7、植物育种要重视育种值和 BLUP 分析方法	63
十一、品种稳定性分析	67
(一) 品种稳定性分析模型	67
(二) 模型分析	68
(三) 品种稳定性分析新模型	69
(四) Shukla 方差计算示例	69
(五) Eberhart-Russell 模型、Shukla 模型、相对平均偏差稳定性分析比较	71
(六) 稳定性选择指标(相对平均偏差 RSD)	72
(七) 稳定性分析策略及建议	73
十二、总结	76

一、试验设计

(一) 试验设计的原则

田间试验设计的原则：重复、随机、局部控制。

1、重复

重复是为了获得更加可靠准确的平均数，并用以估计试验误差，没有重复，就没有平均数（当然，仅有一个观测值，该观测值也可以理解为就是平均数），没有离差，没有方差。如果仅一个观测值，即便存在试验误差，也无法分离。

	A	B
1	1	4
2	CK1	CK2
3	2	5
4	3	1
5	CK3	CK4
6	4	2
7	5	3
8	CK5	CK6

对比法示例：5 个材料 2 次重复

农博士育种家 [间比法]

材料数 5 重复数 2 多排

	A	B
1	CK1	CK2
2	1	5
3	2	4
4	3	3
5	4	2
6	CK3	CK4
7	5	1
8	CK5	CK6

间比法示例：跟对比法在重复方面含义一样。

农博士育种家 [单因素]

区组数 3 每区组内材料数 6

	A	B	C
1	随机区组试验设计方案 (
2	肥力梯度-----		
3	区组1	区组2	区组3
4	3	3	4
5	2	2	2
6	1	6	3
7	4	1	1
8	6	4	5
9	5	5	6

单因素随机完全区组试验示例：3 个区组就是 3 个重复，如果这是一个产量试验，那么每个参试材料，将有 3 个产量观测值。

需要注意一个细节，小区测产，3 个小区，得到了 3 个观测值，由于农业生产的原因，一个小区仅获得了一个产量数据，这个产量数据，无疑会包含仪器测量误差。所谓仪器测量误差，比如，重复 3 次测量一支铅笔的长度，如果获得了不同的测量值，就出现了测量误差，但农业生产面对的显然不是简单的一种仪器重复测量误差情况。一个材料从 3 个小区获得了 3 个观测

值，3个小区，可以是紧挨在一起的一个地块内部的3个区域，或者是3个不同的地块，都是3次重复。

重复3次测量铅笔长度的试验，这样的重复，可以估计测量误差。

单因素随机完全区组试验，3次重复，可以估计出“误差”，但是，这个误差不是简单的重复测量误差，而是包含了重复测量误差，以及地力不同、地力不匀等因素导致的综合误差（这是生物统计一个特殊的地方），可以从中估计出区组效应。

2、随机

非主观等机会安排试验材料。

3、局部控制

一个小区内部，生长条件一致，地力和栽培条件一致。小区内的作物在一致的生长环境下生长。如果不一致，会严重影响试验分析结果。通常认为方形小区，局部控制效果要好。一个小区内部，农艺措施一致很重要。

（二）完全区组

一个区组中包含着所有的试验处理，这种区组称为完全区组。比如，随机完全区组试验。3个区组，那么3个区组中分别存在随机安排的3套一模一样的试验材料。

（三）不完全区组

在科学试验中，由于受到试验条件的限制，有时一个区组中无法容纳全部的试验处理，而只能容纳其中的一部分处理，这种区组称为不完全区组。这样的区组设计称为不完全区组设计。

1、平衡不完全区组设计（BIB设计）

1.1 定义

BIB(Balanced Incomplete Block)，将 v 个处理安排到 b 个区组的一个不完全区组设计称为平衡不完全区组设计，该设计满足下列三个条件：

1. 每个区组都含 k 个不同处理， k 称为区组容量。
2. 每个处理都在 r 个不同区组中出现， r 称为处理重复数。
3. 任一对处理在 λ 个不同区组中相遇， λ 称为相遇数。

从这个定义可以看出：

一个BIB设计中的 v 个处理可以得到公平的比较。

一个BIB设计有五个设计参数 v, k, r, b, λ 。

BIB设计存在的必要条件是：在5个设计参数间同时有下列三个关系式：

$$(1) \quad vr = bk$$

$$(2) r(k-1) = \lambda(v-1)$$

$$(3) b \geq v, r \geq k$$

5 个参数中，任意确定 3 个，即可计算其余 2 个参数，至今人们已经找到很多 BIB 设计。

1.2 BIB 示例

示例：(v=4, k=2, r=3, b=6, λ=1)

区组	1	2	3	4	5	6
处理						
A	y11		y13		y15	
B	y21			y24		y26
C		y32	y33			y36
D		y42		y44	y45	

试验设计解读：

v: 处理数=4，有 ABCD，4 种处理；

k: 区组容量=2，每个区组中有 2 个材料，比如区组 1 中有 y11, y21，区组 2 中有 y32, y42；

r: 处理重复数=3，每种处理存在 3 次重复，

如：A 处理存在 y11, y13, y15, 3 次重复，B 处理存在 y21, y24, y26, 3 次重复；

b: 区组数量=6，有 123456,6 个区组。

λ: 相遇数=1，每两种不同处理，在同一区组相遇数=1，如：A 和 B 两种处理，仅在区组 1 相遇 (y11,y21)，相遇数=1，A 和 C 两种处理，仅在区组 3 相遇 (y13,y33)，相遇数=1。

1.3 BIB 与随机完全区组设计比较

试验 I

区组 1	区组 2	区组 3	区组 4
2	2	1	3
1	1	3	2
4	3	2	4
3	4	4	1

4 个品种 4 次重复的单因素随机完全区组试验设计

农博士育种家 [平衡不完全区组设计(BIB)]																			
试验设计 方差分析 品种区域试验 配合力 遗传模型 数据分析 退出																			
处理(v)		5		区组容量(k)		3		区组(b)		10		重复数 r=6		相遇数 λ=3		定义数据区		分析	
A	B	C	D	E	F	G	H	I	J	K	L								
BIB(v=5 k=3 r=6 b=10 λ=3)																			
		品种1		品种2		品种3		品种4		品种5									
1	区组1	54.2	51	33.1															
2	区组2	28.1	36.6		46.6														
3	区组3	58.7	42.6			38.6													
4	区组4	50.9		18.1	45.4														
5	区组5	60.2		39.1		33.2													
6	区组6	44.3			31.6	14													
7	区组7		25.7	34.5	30.6														
8	区组8		21.3	16		13.1													
9	区组9		30.9		46.6	35.6													
10	区组10			39.4	55.3	44.6													
11	品种效应	10.8600	-1.30000	-8.34000	7.30000	-8.52000													
12	矫正值	48.1900	36.0300	28.9900	44.6300	28.8100													
13																			
14	方差分析																		
15	来源	平方和	自由度	均方和	F														
16	品种	1575.33	4	393.832	7.0084*														
17	区组	2458.50	9																
18	误差	899.152	16	56.1970															
19	总和	4932.98	29																
20																			
21	最小显著差LSD(5%)=9.17425 LSD(1%)=12.6363																		
22	LSD最小显著性差异多重比较																		
23	处理	平均数	5%显著水平	1%显著水平															
24	品种1	49.400	a	A															
25	品种4	42.683	ab	AB															
26	品种2	34.683	bc	BC															
27	品种3	30.033	c	C															
28	品种5	29.850	c	C															

2、格子设计

2.1 简单格子设计

格子设计是一大类设计，以平方格子设计中的简单格子设计为例：

简单格子设计（2个重复，每一重复4个区组，16个材料）（如果是单因素随机完全区组试验设计，就是8个区组算1个重复）

重复 1				
区组 1	1	2	3	4
区组 2	5	6	7	8
区组 3	9	10	11	12
区组 4	13	14	15	16
重复 2				
区组 5	1	5	9	13
区组 6	2	6	10	14
区组 7	3	7	11	15
区组 8	4	8	12	16

格子设计有

平方格子设计：区组内品种数为 p ，供试品种数为 p^2

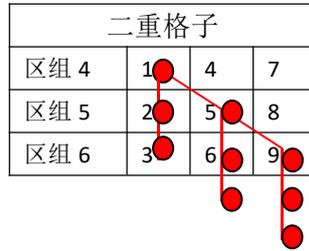
立方格子设计：区组内品种数为 p ，供试品种数为 p^3

矩形格子设计：区组内品种数为 p ，供试品种数为 $p*(p+1)$

植物育种常用的是平方格子设计。

9 个品种，3*3 平方格子设计，示例：
一重、二重、三重、四重格子设计

一重格子				二重格子				三重格子				四重格子			
区组 1	1	2	3	区组 4	1	4	7	区组 7	1	5	9	区组 10	1	6	8
区组 2	4	5	6	区组 5	2	5	8	区组 8	2	6	7	区组 11	2	4	9
区组 3	7	8	9	区组 6	3	6	9	区组 9	3	4	8	区组 12	3	5	7



一重排列方式，就是从左到右，顺序填满第 1 个区组，再折回，顺序填第 2 个区组，不停折回，填满所有为止。表格中，箭头给出填充顺序。

二重排列方式，就是从表格左侧第 1 列，由上到下，顺序填满后，折回，顺序填第 2 列，直到填满。箭头给出填充顺序。

三重排列方式，把二重格子填好的每个数字，想象成弹珠，每一列的珠子用线首尾穿成一个圆环。列之间没有连接。二重格子的对角线位置，是 1-5-9，三个珠子，用一根细棍，顺序穿过 1-5-9，往上挑起，对应放到第一行的 3 个位置上去，之前是 1-4-7，现在是 1-5-9，于是，表格里面所有的数字位置跟着联动，就是三重格子的排列方式了。

四重排列方式，是在三重排列的基础上，用同样的操作，把三重格子对角线 1-6-8，穿起，拉到第一行的位置。

格子设计里面，这种数字滚动和数字联动，是很有意思的现象，用户可以好好理解体会。

2.2 简单格子设计和单因素随机完全区组比较

试验 I

区组 1	A	B	D	C	E	F	G	H	I
区组 2	B	F	H	D	G	C	E	I	A

9 个品种 2 次重复的单因素随机完全区组试验设计

试验 II

重复 1			重复 2				
区组 1	A	B	C	区组 4	A	D	G
区组 2	D	E	F	区组 5	B	E	H
区组 3	G	H	I	区组 6	C	F	I

9 个品种 2 次重复的单因素平方格子设计（简单格子设计）

以上两个试验设计比较

	品种数	重复数	区组数	实际使用小区数
试验 I	9	2	2	18
试验 II	9	2	6	18

两种试验设计在使用相同土地面积的情况下，试验 II 区组数更多，意味着试验 II 的地块误差影响控制的更好，试验数据的准确性更高。

格子设计数据分析结果中，有的会给出相比单因素随机完全区组试验设计相对效率提升了多少的数值，为什么会有提升，从上面两个试验可以看出，简单格子设计等同于在单因素随机完全区组的区组内部又划分了 3 个区组出来，因而，简单格子设计相比单因素随机完全区组，误差更准确。

$$\text{相对效率} = \text{单因素随机完全区组误差均方} / \text{格子设计有效误差均方}$$

格子设计的这种误差准确性的提升，从另一角度解读，就是随着单因素随机完全区组试验规模的扩大，一个区组占地越来越大，这样地块差异导致的区组内的差异，就系统性的增大了，而如果将区组分拆为小区组，自然可以更好的满足地块一致性的试验设计要求。

因此在品种数量较多的情况下，应该选择格子设计，而不是单因素随机完全区组。

2.3 应用示例

The screenshot shows the 'Simple Lattice Design (Simple Lattice)' software interface. At the top, it displays '区组 5' (Block 5) and '重复 2' (Replication 2). Below this is a data table with columns A through L. The data rows (22-31) show values for each block and replication. Below the data table, there are two analysis tables: '随机区组方差分析表' (Random Block ANOVA Table) and '简单格子设计方差分析表' (Simple Lattice Design ANOVA Table). The simple lattice design table shows a significant difference between blocks with a p-value of 4.5944e-05. At the bottom, a note states: '与随机区组设计比较, 效率提高了: 74.3353%' (Compared to random block design, efficiency increased by 74.3353%).

	A	B	C	D	E	F	G	H	I	J	K	L
22		(21)	22	(22)	38	(23)	23	(24)	37	(25)	33	
23												
24		调整后的品种总和										
25												
26		(1)	38.1361	(2)	33.9456	(3)	29.2926	(4)	29.5375	(5)	25.6939	
27		(6)	26.3401	(7)	18.1496	(8)	13.4966	(9)	16.7414	(10)	16.8979	
28		(11)	47.1021	(12)	24.9116	(13)	25.2586	(14)	41.5034	(15)	38.6599	
29		(16)	25.2449	(17)	21.0544	(18)	21.4013	(19)	14.6462	(20)	22.8026	
30		(21)	23.2517	(22)	37.0612	(23)	24.4082	(24)	34.6530	(25)	30.8095	
31												
32												
33												
34	随机区组方差分析表											
35	来源	平方和	自由度	均方和	F							
36	重复	212.180	1									
37	品种	559.280	24	23.3033	0.776433							
38	误差	720.320	24	30.0133								
39	总和	1491.78	49									
40												
41	简单格子设计方差分析表											
42	来源	平方和	自由度	均方和	F							
43	重复	212.180	1									
44	品种(未调)	559.280	24	23.3033								
45	重复内区组	501.840	8	62.7300	4.5944**							
46	区组内误差	218.480	16	13.6550								
47	总和	1491.78	49									
48												
49												
50	调整的品种均方及F测验分析表											
51	来源	平方和	自由度	均方和	F							
52	品种(调整)	644.626	24	26.8594	1.96700							
53	区组内误差	218.480	16	13.6550								
54	与随机区组设计比较, 效率提高了: 74.3353%											
55												
56												
57												

3、 α -格子设计

α -格子设计 (α -lattice) : 不完全区组设计---格子设计---广义格子设计--- α -lattice。

α -格子试验设计，是当前国外育种企业采用较多的一种不完全区组试验设计，基于设计母阵构造设计，主要应用于育种过程中早、中期产量试验阶段的品种筛选。其采用的不完全区组的试验设计方式，提升了试验效率，降低了育种试验成本。

01 农博士育种家 [α-格子设计 (Alpha-Lattice)]								
重复数 (R) 2			重复内区组数 (S) 6					
	A	B	C	D	E	F	G	H
1	Alpha-Lattice设计 (S=6, K=5, R=2, T=30)							
2								
3	重复1							
4		区组 (1) 1	7	13	19	25		
5		区组 (2) 2	8	14	20	26		
6		区组 (3) 3	9	15	21	27		
7		区组 (4) 4	10	16	22	28		
8		区组 (5) 5	11	17	23	29		
9		区组 (6) 6	12	18	24	30		
10	重复2							
11		区组 (7) 1	8	16	21	29		
12		区组 (8) 2	9	17	22	30		
13		区组 (9) 3	10	18	23	25		
14		区组 (10) 4	11	13	24	26		
15		区组 (11) 5	12	14	19	27		
16		区组 (12) 6	7	15	20	28		
17								

与随机完全区组（30个材料，2个区组），做个对比：

02 农博士育种家 [单因素]		
区组数 2		每区组内材料数 30
	A	B
1	随机区组试验设计	
2	肥力梯度-----	
3	区组1	区组2
4	18	6
5	12	2
6	20	19
7	17	14
8	21	17
9	28	8
10	22	7
11	9	10
12	2	1
13	15	4
14	13	26
15	8	18
16	19	3
17	5	21
18	23	23
19	24	22
20	11	12
21	26	25
22	16	5
23	10	27
24	27	30
25	7	9
26	14	11
27	25	29
28	3	28
29	4	16
30	6	13
31	1	24
32	29	15
33	30	20

与随机完全区组试验设计相比，在参试材料很多的情况下，随机完全区组的小区面积急剧扩大，地块地力差异，对试验本身产生了破坏性影响，或者可以表述为，试验设计的局部控制失灵。

参考文献：A New Class of Resolvable Incomplete Block Designs

Author(s): H. D. Patterson and E. R. Williams

Source: Biometrika, Vol. 63, No. 1 (Apr., 1976),

pp. 83-92

农博士软件 QQ 交流群（群号：40082612）中提供该篇文献下载。

（四）增广试验设计

1、什么是增广试验设计？

常规的试验设计，如随机区组、拉丁方、裂区、格子方等，其基本特征是各处理具有相等的重复次数，可统称为“标准”设计。所谓增广设计（augmented design），是指在标准设计的区组中增多试验小区数目的一类设计（任何一种标准设计均可形成其相应的增广设计）。

增广材料一般较多，不安排重复，所以在不同区组中仅出现一次。

增广试验设计可以同时安排有重复的和无重复的处理。重复的处理，可以用于估计试验环境误差，从而可以分离出区组的效应，而安排在不同区组内的增广材料，尽管没有安排重复，可以借鉴区组效应，对获得的观测值进行校正，获得更可靠的试验数据，提高试验的可靠性。

结合实际育种需要，增广试验设计，就很自然的演变为，增广的部分，成为了主要内容，多安排材料，而重复的处理，可以考虑尽可能压缩数量。因而一个增广的随机完全区组试验，增广材料往往是大头。这也是用增广设计代替对比法、间比法的原因。

2、增广试验设计在育种中何时采用？

一般地说，在一个育种周期的开始，往往选系很多，而每一选系的种子较少，难以进行重复试验；以后，随着育种工作的进展，选系数目因逐年淘汰而逐步地由多变少，每一选系的种子数量则因逐年繁殖而逐步地由少变多，因而可以进行重复试验，甚至多地点的重复试验。因此，植物育种试验及其设计可大体分为以下三类：

大量选系的单区试验

试验的选系很多，可能有几百个至几千个；但每一选系的种子量少，在试验中仅种植一个小区。可以应用于这类试验的设计有：增广设计、修饰增广设计和行列增广设计。

较多选系的重复试验

试验的选系已有所减少，但仍然较多，譬如数十个至数百个；但每一选系的种子数量已允许种植两次或两次以上重复。这类试验可应用的设计有各种格子设计和格子方设计。

较少选系的重复试验

这类试验包括育种单位的品种预备试验和国家的品种区域化试验。选系一般在 10-20 个，且种子数量充足，可以在一个或多个地点进行重复试验。这类试验通常应用随机区组设计。

3、增广随机完全区组设计示例

农博士育种家 [增广设计]

区组数 3 每区组内材料数 6 区组随机 增广材料数 23

	A	B	C
1	随机区组试验设计方案 (日期: 2017-1-2)		
2	肥力梯度----->		
3	区组1	区组2	区组3
4	6	Z20	1
5	4	4	3
6	1	3	4
7	Z13	5	Z18
8	3	1	5
9	5	Z8	Z12
10	Z16	2	Z3
11	Z4	6	Z6
12	Z7	Z5	Z15
13	Z10	Z23	2
14	2	Z14	Z9
15	Z1	Z17	Z21
16	Z19	Z11	6
17	Z22	Z2	

这是一个增广随机完全区组试验，在未增广材料之前，试验有 3 个区组，每个区组 6 个材料，也就是说，6 个品种材料，各有 3 次重复。23 个增广材料，随机安排到 3 个区组中，增广材料不设重复。增广材料安排进试验后，试验设计如上图所示。

4、修饰增广试验设计

4.1 修饰增广设计的含义

修饰增广设计最初由 Lin (1983) 等提出。其基本结构是裂区设计，即试验分主区(整区)和副区(裂区)。主区可以应用任何一种标准设计(通常宜用能够消除行、列两个方向的环境差

异的设计，如拉丁方设计、尤顿方设计等。每一主区均须布置成 $3 \times 3 = 9$ 个副区，并使每一副区呈方形或尽可能近于方形，以保证 8 个边缘副区与中心副区最为靠近。中心副区称为对照小区，用于安排对照品种；其余 8 个副区则称为测验小区，均可用于安排增广材料，但为了估计副区的试验误差和校正试验结果，一般每 1 个品种又在 2 个主区的各 1 个边缘副区中重复安排该品种（该小区称为对照副区）。

4.2 修饰增广设计中“修饰”的含义

修饰增广设计主要在两个方面“修饰”了增广设计：

(1) 对照品种的随机排列被“修饰”为规则排列（固定在中心副区上）；

(2) 行、列区组的不规则形状亦被“修饰”为规则的，且在某些行、列的 9 个副区中出现了两个同一品种的副区。

Lin 等认为，对照品种的随机排列固然有利于试验误差的无偏估计，但在校正田间环境的差异上不及规则排列有效。在育种工作的前期，田间试验的主要目的是比较农艺性状，粗略地估计供试材料的基因型值，而不是对系间差异作严格测验（那是需要重复试验的）。增广材料观察值的有效校正往往比误差的无偏估计更为重要。因此，上述两个方面的“修饰”是必要的和可行的。

一般来说，主区为 $b \times b$ 拉丁方的修饰增广设计，总共将有 $9 \times b \times b$ 个小区，其中有对照小区 $b \times b$ 个，对照副区 $2 \times b$ ，测验小区 $b(8b-2)$ 个。

以 5×5 拉丁方的修饰增广设计为例：总共将有 $9 \times 5 \times 5 = 225$ 个小区，其中有对照小区 $5 \times 5 = 25$ 个，对照副区 $2 \times 5 = 10$ ，测验小区 $5(8 \times 5 - 2) = 190$ 个。

4.3 修饰增广设计示例

@农博士育种家 [修饰增广设计]															
处理数 5 <input type="checkbox"/> 字母显示 设计															
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	修饰增广试验设计方案 (日期: 2017-1-2)														
2															
3	E**	1	88	24	141	16	145	23	6	63	17	11	28	150	A**
4	179	B*	126	156	C*	65	D**	D*	29	170	E*	15	89	A*	26
5	46	120	25	13	58	157	85	167	53	104	20	169	189	36	32
6	E**	9	151	122	177	34	113	14	109	C**	178	149	48	D**	139
7	133	E*	4	52	A*	41	54	B*	55	56	C*	59	78	D*	18
8	74	118	162	79	87	A**	B**	116	135	62	182	67	70	112	7
9	73	187	75	81	158	173	38	92	134	40	39	98	8	106	131
10	128	A*	114	180	B*	184	137	C*	168	94	D*	95	190	E*	175
11	84	99	147	12	174	185	93	153	43	3	49	51	161	101	186
12	72	77	110	42	105	102	119	44	121	27	50	111	154	129	127
13	2	C*	5	83	D*	183	132	E*	138	96	A*	22	136	B*	124
14	164	31	115	117	33	165	69	21	160	80	148	163	45	91	57
15	64	103	155	60	90	E**	107	144	47	82	166	181	159	130	68
16	10	D*	146	100	E*	125	71	A*	172	123	B*	97	61	C*	176
17	66	19	171	142	37	143	30	86	76	140	152	188	35	C**	108

这是一个修饰增广试验设计，首先是一个 5 个品种 (A、B、C、D、E) 的拉丁方设计。图中每 9 (3X3) 个单元格，对应拉丁方设计的 1 个单元格。9 个单元格的中心单元格，安排的是对照品种。以最左上角单元格为例。

@农博士育种家 [修饰增广设计]			
处理数 5			
	A	B	C
1	修饰增广试验设计方案 (
2			
3	E**	1	88
4	179	B*	126
5	46	120	25

图中的 9 个单元格，都属于 5*5 拉丁方的 1 个单元格。中心单元格，安排 5 个对照品种中的 1 个，是 B，周围还有 8 个单元格，试验中，5 个对照品种，需要安排 5 个对照，图中 8 个单元格中凑巧安排了 1 个，剩下的 7 个单元格随机安排增广材料。

总结一下：

- 1、图中的 9 个单元格对应 5*5 拉丁方设计的 1 个单元格。
- 2、9 个单元格的中心单元格要安排对照品种，至于是什么，由拉丁方设计决定。
- 3、剩下的 8 个单元格中可能会安排中心单元格的对照，以对照 B 为例，在 5*5 的拉丁方试验设计中，B 一共出现 5 次。其中的 2 个 B 要各安排 1 次重复。
- 4、9 个单元格剩下的可能是 8 个或 7 个，随机安排增广材料。

4.4 修饰增广设计的优势

1、与对比法间比法比较：中心单元格，安排的是对照品种，围绕中心单元格，安排增广材料，可以很好的进行田间比较对照，解决了对比法、间比法安排对照占地过多的问题。

2、修饰增广设计，误差控制有优势，可以对没有重复的增广材料观测值进行科学校正，提高试验可靠性。

5、增广试验设计的方差分析

这部分内容，不在此详细介绍，莫惠栋先生 1990-1991 年曾在《江苏农学院学报》连续发表《植物育种中的试验设计：增广设计》、《植物育种中的试验设计：修饰增广设计》、《植物育种中的试验设计：行列增广设计》等文章，进行了详细介绍。

农博士育种技术理论群（QQ 群号:40082612），提供相关文章 PDF 下载服务，感兴趣的朋友可以自行下载学习。

（五）试验设计的发展趋势

试验设计种类很多，总的发展趋势是，减轻试验工作量，降低试验成本，注重提高试验效率，以尽可能小的试验规模，获取更多的试验信息。

（六）试验方法的选用原则

根据自身工作需要和对试验设计知识掌握的情况，从自身实际出发选择试验设计方法，而不是盲目地采用新方法，尽可能采用试验效率高的方法

当前，植物育种主要用到的试验设计方法：对比法、间比法、单因素随机完全区组、一年多点、多年多点。简单格子设计、 α -格子试验设计逐步会被用户采用。

试验设计方法是统计学家研究的内容，我们侧重应用，试验设计可以极其复杂，比如：交替设计、增广设计、配方设计、回归设计、稳健设计、正交设计、混杂设计、因子设计等等。

明白试验设计的分析原理，知道如何分析试验数据，是更重要的事情。

二、方差分析

（一）离差

离差是观测值与平均数的差。

示例：观测值有 3 个，分别是 1,2,3

平均数=2

观测值 1 的离差: $1 - 2 = -1$

观测值 2 的离差: $2 - 2 = 0$

观测值 3 的离差: $3 - 2 = 1$

离差和=0.

(二) 方差 (均方)

均方 (MS, mean squares) = 离差平方和 \div 自由度
以上例数据为例,

离差平方和

$$(-1)^2 + (0)^2 + (1)^2 = 2$$

自由度=3-1=2

$$MS = 2 \div 2 = 1$$

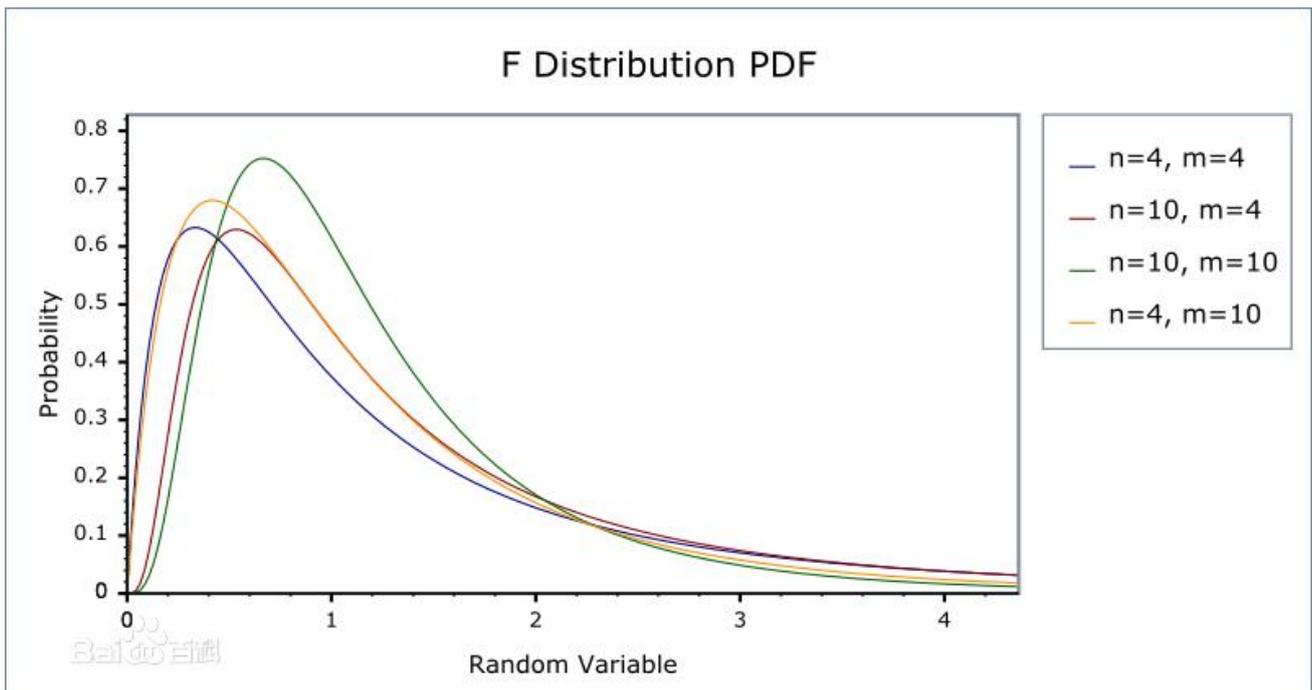
标准差是方差的平方根值。

(三) 自由度

有 n 个观测值的样本, 可以得到 n 个离差, 但是只有 $n-1$ 个离差可以“自由变动”, 因为最后一个离差, 被其余 $n-1$ 个离差“规定”了数值, 因为离差计算使用了平均数的原因, 所以, 自由度是 $n-1$ 。

(四) F 分布

F 分布是 1924 年英国统计学家 R.A.Fisher 提出, 并以其姓氏的第一个字母命名的。



(图片来自百度百科)

F 值的定义:

$$F(v_1, v_2) = S_1^2 / S_2^2$$

F 分布定义：设 X、Y 为两个独立的随机变量，X 服从自由度为 k1 的卡方分布，Y 服从自由度为 k2 的卡方分布，这 2 个独立的卡方分布被各自的自由度除以后的比率这一统计量的分布。简单理解就是两个方差（均方）之比。

从一个正态总体 (μ, σ^2)，按照不同的 v1, v2 自由度进行一系列抽样，就可以得到一系列不同的 F 值，获得一个 F 分布。

根据 F 分布，可以计算得出 F 值表。

(五) F 测验

F 测验，实际上是根据试验需要构建的一个分数。 $F_{(v_1, v_2)} = S_1^2/S_2^2$

试验目的决定了 F 测验的内容，决定了 F 测验的分子和分母。

通常， S_1^2 所包含的变异要比 S_2^2 多一项，这样，获得的 F 值通常 > 1 ，同时分子分母仅有一项差异，好衡量该项差异的影响有多显著。

随机完全区组试验设计的期望均方

变异来源	自由度 DF	均方 MS	模型
区组间	(n-1)	MS_R	$\sigma^2 + k\kappa_\beta^2$
品种间	(k-1)	MS_t	$\sigma^2 + n\sigma_\tau^2$
试验误差	(n-1)(k-1)	MS_e	σ^2

构建

F 测验（区组间/试验误差），可以检测区组间差异是否显著

F 测验（品种间/试验误差），可以检测品种间差异是否显著

从 F 值的计算公式可以看出，只有计算方差的时候，才有可能用到 F 测验，如果试验数据，没有计算方差，仅是计算离差效应，则不会用到 F 测验。

(六) 单因素随机完全区组试验方差分析示例

此处示例没有采用统计教材常用的方差分析的套路，而是仅从离差出发，自然过渡到方差（均方）的思路，希望带给用户一些不一样的理解。

随机完全区组试验统计分析			
	区组 1	区组 2	区组 3
材料 1	10.9	9.1	12.2
材料 2	10.8	12.3	14
材料 3	11.1	12.5	10.5
材料 4	9.1	10.7	10.1
材料 5	11.8	13.9	16.8
材料 6	10.1	10.6	11.8
材料 7	10	11.5	14.1
材料 8	9.3	10.4	14.4

（试验数据来自中国农业出版社《试验统计方法》2000 版 盖钧镒老师，P228 例题）

这是一个单因素随机完全区组试验，如果采用效应分析的思路，进行数据分析，应该怎么做呢？

单因素随机完全区组试验统计分析						
	区组 1	区组 2	区组 3	材料平均数	总平均数	材料效应=

						(材料平均数-总平均数)
材料 1	10.9	9.1	12.2	10.73333	11.5833	-0.84997
材料 2	10.8	12.3	14	12.36667	11.5833	0.783367
材料 3	11.1	12.5	10.5	11.36667	11.5833	-0.21663
材料 4	9.1	10.7	10.1	9.96667	11.5833	-1.61663
材料 5	11.8	13.9	16.8	14.16667	11.5833	2.583367
材料 6	10.1	10.6	11.8	10.83333	11.5833	-0.74997
材料 7	10	11.5	14.1	11.86667	11.5833	0.283367
材料 8	9.3	10.4	14.4	11.36667	11.5833	-0.21663
区组平均数	10.3875	11.375	12.9875			
总平均数	11.5833	11.5833	11.5833		11.5833	材料效应合计=0
区组效应= (区组平均数-总平均数)	-1.1958	-0.2083	1.4042	区组效应合计=0		

注意，材料效应的合计=0，区组效应的合计=0。

如果我们要进行方差分析，那又应该怎么做呢？我们从上面的效应分析入手来进行方差分析，这样就更容易理解方差分析的本质。

材料效应的平方和，均方			
	离差	离差的平方	3 *离差的平方
材料 1	-0.84997	0.722443	2.16733
材料 2	0.783367	0.613663	1.84099
材料 3	-0.21663	0.04693	0.14079
材料 4	-1.61663	2.613503	7.84051
材料 5	2.583367	6.673783	20.02135
材料 6	-0.74997	0.56245	1.68735
材料 7	0.283367	0.080297	0.24089
材料 8	-0.21663	0.04693	0.14079
离差的平方和累加， Σ			34.08
自由度，df			8-1=7
均方，MS (Σ /df)			4.87

为什么乘以 3？材料 1 的离差 -0.84997 是不是由 3 个观测值来的。

区组效应的平方和，均方			
	区组 1	区组 2	区组 3
离差	-1.1958	-0.2083	1.4042
离差的平方	1.429938	0.043389	1.971778
8 *离差的平方	11.4395	0.347111	15.77422
离差的平方和累加， Σ	27.56		
自由度，df	3-1=2		
均方，MS (Σ /df)	13.78		

为什么乘以 8？区组 1 的离差 -1.1958 是不是由 8 个观测值来的。

试验数据的总平方和

观测值	总平均	离差的平方
10.9	11.5833	0.466899
10.8	11.5833	0.613559
11.1	11.5833	0.233579
9.1	11.5833	6.166779
11.8	11.5833	0.046959
10.1	11.5833	2.200179
10	11.5833	2.506839
9.3	11.5833	5.213459
9.1	11.5833	6.166779
12.3	11.5833	0.513659
12.5	11.5833	0.840339
10.7	11.5833	0.780219
13.9	11.5833	5.367099
10.6	11.5833	0.966879
11.5	11.5833	0.006939
10.4	11.5833	1.400199
12.2	11.5833	0.380319
14	11.5833	5.840439
10.5	11.5833	1.173539
10.1	11.5833	2.200179
16.8	11.5833	27.21396
11.8	11.5833	0.046959
14.1	11.5833	6.333779
14.4	11.5833	7.933799
离差的平方和累加, Σ		84.61
自由度, df		24-1=23

总变异平方和=材料平方和+区组平方和+误差平方和（依据方差分析线性模型）
总变异平方和=84.61

总变异	84.61
材料	34.08
区组	27.56
误差平方和 Σ	$84.61 - 34.08 - 27.56 = 22.97$
自由度, df	$23 - 2 - 7 = 14$
均方, MS (Σ/df)	$22.97/14 = 1.64$

方差分析表

变异来源	Df	SS	MS	F	F0.05
区组间	2	27.56	13.78	8.40*	3.74
品种间	7	34.08	4.87	2.97*	2.77
误差	14	22.97	1.64		
总变异	23	84.61			

分析的结果与教科书《试验统计方法》2000版，P228 例题分析结果完全一样。

在算误差平方和的时候，采用了下面的计算方法。

误差平方和 Σ	$84.61 - 34.08 - 27.56 = 22.97$
总变异平方和 - 材料变异平方和 - 区组变异平方和 = 误差平方和	

（七）方差分析缺区数据分析结果比较

比较思路：选取育种常用的单因素随机完全区组试验方差分析，采用同样的试验数据。由于比较部分内容较多，不在此详细列出，感兴趣的读者可以加入农博士 QQ 交流群（群号：40082612）下载 PDF 文档。

SAS9.1、农博士育种家、“缺区估计法”，基于同一数据，三方的方差分析结果比较：

同一数据，缺区情况下分析结果比较

平方和 SS	自由度	SAS9.1	农博士育种家	缺区估计法
总变异	22	1338.246087	1338.246	1402.48
区组间	3	176.554754	176.555	166.84
材料间	5	1028.969420	1028.969	1093.20
误差	14	132.721913	132.722	142.44

比较结果如下：

无论在试验数据完整或者缺失的情况下，农博士育种家软件方差分析结果与 SAS9.1 分析结果一致，该比较采用的是 SAS 的 ANOVA Procedure。缺失数据情况下，SAS9.1 和农博士育种家分析结果一致，“缺区估计法”分析结果明显不同。

思考：

1、国内统计教材很大程度上取材于国外，从教材内容看，对试验数据缺区，国内外教材都教授了“缺区估计”的方法，原理方法是一致的。

2、SAS ANOVN 和农博士育种家，缺区情况下方差分析结果是一致的。由此可以推断，SAS ANOVN 采用的“缺区”应对策略与博思公司自主提出的缺区应对策略是一致，而博思公司不主张采用“缺区估计”的方法。

3、SAS ANOVN 很明显并没有采用“缺区估计”的方法，从前面结果比较可以看出。SAS 并没有采用国外统计教材上的“缺区估计”的方法，这本身就是一个值得思索的问题。

4、在出现试验效应平均数无法计算的情况下，比如，试验某 1 区组数据全部缺失，SAS 采用了区组数量-1 的处理策略，这跟博思公司提出的“效应平均数无法计算，那就不能进行方差分析”的看法是一致的。

（八）方差分析数据“缺区”应对策略讨论

“缺区”，就是试验本应获得的试验数据丢失了。缺区就“不平衡”了，国内外教科书都有主张进行缺区“估计”，计算一个观测值补缺，依据是离差平方和最小的数据拟合，最小二乘法。该方法有其理论根据，但是一个小区在数据缺失的情况下，可以通过计算“获得”数据，终究没有实实在在长出来的产量来的踏实。该方法在缺失一个数据的时候，可以直接计算获得“产量”，缺失两个以上数据，就需要列方程组解方程才能计算得到两个以上“产量”，应用中很不方便。这是一种应对缺区的策略。其他策略还有：列表删除、成对删除、单一插补、多重插补等应对策略，不是很适用于生物统计，不在此讨论。下面就“缺区”情况应对进行简单讨论分析：

讨论如下：主张进行缺区“估计”的理由是由于处理和区组的正交性遭到破坏，这没有异议，但是不是过于重视这种“破坏”了呢？以育种常用的单因素随机完全区组为例，试验中区

组是完全区组，试验本身就是正交的。

方差分析 ANOVA，总平方和可由各部分平方和相加构成，缺区的情况下，这种构成应该还是成立的。分离误差，采用的是先分离区组和品种的平方和，剩余的就是误差平方和，以品种平方和分离为例，计算每一个品种的均值与总均值的离差，也正是此处，出现了前面提到的试验“正交性”由于缺区被破坏了，具体含义也就是，因为缺区，以 3 个区组为例，有的品种均值来自 3 个观测值的平均，有的来自 2 个观测值的平均（1 个缺区），均值间不均衡了，但是，3 个观测值的品种均值，代表了一个品种，2 个观测值的均值也可以代表一个品种，从这个意义上，是不是可以忽略缺区对均值的影响。均值和方差是数据的两部分，在此问题上，生物统计从应用上是不是可以更多重视均值，因为，代表一个品种的是均值。

采用缺区估计填入的方法，缺区“估计”值的填入，不可避免的影响了总平均数，人为增大了总平方和的累加值，增大了总的方差，而自由度又减 1，这对误差的计算有直接影响，想去有问题。进行多重比较的时候，是用实际观测值获得的产量平均数进行比较，还是合并估计产量后计算得到的平均产量进行比较呢，这两个平均数是存在差异的。

缺失的数据，采用估计的方式处理，更多的是为了满足数据模型分析的需要，估计处理，围绕着相应的均值进行，估计处理不能影响试验本身，而当试验中某效应计算需要的观测值全部缺失，方差分析就无法进行了，此时如果还采用缺区估计的处理方式，就等同试验中本应长出来的材料，缺失了，却要用别的材料来算这个材料的表现，试验的意义都不存在了。

缺区情况在育种试验中难免，那么针对缺区情况下还要进行的方差分析，就有必要搞的更清楚一些。很多教材提到缺区，就是笼统的提到，线性模型、GLM 可以处理不均衡的数据，好像线性模型、GLM 什么样的缺区情况都可以处理，这就有问题了，方差分析本身就是一个线性模型，线性模型或者 GLM 构建最小二程方程组解出来的也是一堆试验效应对应的平均数，用回归思想计算得到的平均数，怎么都不会比直接计算的更准确，如果缺区导致试验某个效应对应的平均数无法计算得到，那还在用估计方法进行方差分析，不会因为是采用了线性混合模型就正确和可信。

综上所述，觉得还是不用“估值”直接分析缺区数据来的可靠。

1、缺区，并没有影响**试验材料的客观试验表现**，需要进行相应的平方和和自由度计算调整。

2、F 分布要求遵循正态分布，个别数据缺失，可以理解为对分布没有影响。如果数据缺失严重，就不要进行方差分析，F 测验了，直接进行品种平均数大小排序即可。

3、缺几个自由度就减几个，平方和采用累加的思路，缺了就不加，跟未缺区的时候一样分析，试验分析参数进行对应调整。

4、如果缺区导致试验中某效应的平均数无法计算，那就不能进行方差分析，这可以作为缺区情况下是否能够进行方差分析的一个判断依据。

5、如果考虑采用缺区估计方式填值处理，应在试验数据缺区未影响到效应的平均数计算的情况下进行，否则不宜采用估计方式。

后续进行的多重比较，采用试验中真实获得的品种平均数。

下节给出一个“缺区”分析的例子，供读者参考。

(九) 缺区方差分析示例

还是采用同样的试验数据来进行数据缺区方差分析示例。下图中，材料8区组3数据缺失。

农博士育种家 [单因素]											
试验设计 方差分析 品种区域试验 配合力 遗传模型 数据分析 退出											
材料数		8		区组数		3		指定对照			
	A	B	C	D	E	F	G	H	I	J	K
1	随机区组试验统计分析										
2		区组1	区组2	区组3							
3	材料1	10.9	9.1	12.2							
4	材料2	10.8	12.3	14							
5	材料3	11.1	12.5	10.5							
6	材料4	9.1	10.7	10.1							
7	材料5	11.8	13.9	16.8							
8	材料6	10.1	10.6	11.8							
9	材料7	10	11.5	14.1							
10	材料8	9.3	10.4								
11											
12	变异来源	DF	SS	MS	F	概率	F0.05	F0.01			
13	区组间	2	21.562	10.781	9.240**	0.996	3.810	6.700			
14	材料间	7	39.603	5.658	4.849**	0.993	2.830	4.440			
15	误差	13	15.169	1.167							
16	总变异	22	76.335								
17											
18	最小显著差LSD(5%)=1.90528 LSD(1%)=2.65585										
19	LSD最小显著性差异多重比较										
20	处理	平均数	5%显著水	1%显著水平							
21	材料5	14.167	a	A							
22	材料2	12.367	ab	AB							
23	材料7	11.867	bc	AB							
24	材料3	11.367	bcd	B							
25	材料6	10.833	bcd	B							

农博士育种家软件的计算结果在图中可以看到。

下面我们从头计算，看看过程是怎么样的，以及有什么变化。

随机完全区组试验统计分析			
	区组 1	区组 2	区组 3
材料 1	10.9	9.1	12.2
材料 2	10.8	12.3	14
材料 3	11.1	12.5	10.5
材料 4	9.1	10.7	10.1
材料 5	11.8	13.9	16.8
材料 6	10.1	10.6	11.8
材料 7	10	11.5	14.1
材料 8	9.3	10.4	

单因素随机完全区组试验统计分析						材料效应= (材料平均数-总平均数)
	区组 1	区组 2	区组 3	材料平均数	总平均数	
材料 1	10.9	9.1	12.2	10.73333	11.4609	-0.72757
材料 2	10.8	12.3	14	12.36667	11.4609	0.905767
材料 3	11.1	12.5	10.5	11.36667	11.4609	-0.09423
材料 4	9.1	10.7	10.1	9.966667	11.4609	-1.49423
材料 5	11.8	13.9	16.8	14.16667	11.4609	2.705767

材料 6	10.1	10.6	11.8	10.83333	11.4609	-0.62757
材料 7	10	11.5	14.1	11.86667	11.4609	0.405767
材料 8	9.3	10.4		9.85	11.4609	-1.6109
区组平均数	10.3875	11.375	12.7857			材料效应合计
总平均数	11.5833	11.5833	11.5833		11.4609	=0
区组效应= (区组平均数-总平均数)	-1.0734	-0.0859	1.3248	区组效应合计 =0		

材料效应的平方和，均方			
	离差	离差的平方	3 *离差的平方
材料 1	-0.72757	0.529353	1.58806
材料 2	0.905767	0.820413	2.46124
材料 3	-0.09423	0.00888	0.02664
材料 4	-1.49423	2.232733	6.6982
材料 5	2.705767	7.321173	21.96352
材料 6	-0.62757	0.39384	1.18152
材料 7	0.405767	0.164647	0.49394
材料 8	-1.6109	2.594999	2.594999 * 2=5.189998
离差的平方和累加, Σ			39.603
自由度, df			8-1=7
均方, MS (Σ /df)			5.658

区组效应的平方和，均方			
	区组 1	区组 2	区组 3
离差	-1.0734	-0.0859	1.3248
离差的平方	1.152188	0.007379	1.755095
	8 *	8 *	7 *
离差的平方	9.216978	0.058989	12.28623
离差的平方和 累加, Σ	21.562		
自由度, df	3-1=2		
均方, MS (Σ /df)	10.781		

试验数据的总平方和		
观测值	总平均	离差的平方
10.9	11.4609	0.314575
10.8	11.4609	0.436749
11.1	11.4609	0.130227
9.1	11.4609	5.573705
11.8	11.4609	0.115009
10.1	11.4609	1.851966
10	11.4609	2.13414
9.3	11.4609	4.669357
9.1	11.4609	5.573705
12.3	11.4609	0.70414
12.5	11.4609	1.079792

10.7	11.4609	0.578923
13.9	11.4609	5.949357
10.6	11.4609	0.741096
11.5	11.4609	0.001531
10.4	11.4609	1.125444
12.2	11.4609	0.546314
14	11.4609	6.447183
10.5	11.4609	0.92327
10.1	11.4609	1.851966
16.8	11.4609	28.50631
11.8	11.4609	0.115009
14.1	11.4609	6.965009
离差的平方和累加, Σ		76.335
自由度, df		23-1=22

总变异平方和=材料平方和+区组平方和+误差平方和
总变异平方和=76.335

总变异	76.335
材料	39.603
区组	21.562
误差平方和 Σ	76.335-39.603-21.562=15.169
自由度, df	22-2-7=13
均方, MS (Σ /df)	15.17/13=1.167

方差分析表

变异来源	Df	SS	MS	F	F0.05
区组间	2	21.562	10.781	9.240**	3.81
品种间	7	39.603	5.658	4.849**	2.83
误差	13	15.169	1.167		
总变异	22	76.335			

(十) 固定效应和随机效应

方差分析线性模型可表示为:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

μ 为总体平均数, τ_i 为试验处理效应, ε_{ij} 为随机误差, 具有分布 $N(0, \sigma^2)$ 。

对 τ_i 的不同解释, 产生了固定模型和随机模型, 也就产生了固定效应和随机效应, 生物统计因而变得复杂了, 模型的不同, 仅与 F 测验和统计推断有关, 方差分析计算的 SS、MS、df 都一样。

在育种应用中, 不需受太多固定和随机的困扰, 可以这样简单理解区分这两种效应。

以配合力试验为例, 种植在不同地点的配合力产量试验, 地点效应就是固定效应, 材料

的配合力效应就是随机效应，因为，我们需要对材料的配合力效应进行推断，而不需要对试验地点的效应进行推断，**需要推断的就是随机效应，不需要推断的就是固定效应。严格的说，这样的说法不是统计的标准说法，但是会更好记，好理解一些。**再比如，多年多点配合力试验，年度效应和地点效应都视为固定效应，材料配合力效应视为随机效应。

用统计的术语表述，就是随机模型是通过样本表现去推断总体表现。

（十一）理解方差分析

1、方差分析，本质上是考察试验的各种效应，理解方差分析，应该从试验效应入手，**生物统计和数量遗传学中，各种效应就是各种离差，效应是离差的各种具体应用**，比如，一般配合力效应，品种效应，区组效应。方差只是计算手段，方差分析，采用离差的平方进行计算，可以克服离差合计=0 的计算不便。

2、方差分析中的 F 测验，不是进行方差分析的目的，仅是一个分析环节，可以提供更多一些的统计信息，**不算方差就不需要 F 测验**，比如，效应是否达到 0.05 水平的显著。

3、生物统计，是为农业生产服务的，育种实践中，以产量品比试验为例，在人们的感官层面，衡量比较几个品种的产量高低，就是产量的平均数，**一个产量均值最高的品种，就是最高产的品种**，不会因为方差、误差、F 测验显著与否，影响人们的判断结论，更不可能改变客观事实。生物统计有时会给人一种假象错觉，一个产量（**均值**）最高的品种，去除误差影响后，可能变为不是产量最高的品种了。可能吗？同一个试验，一个产量均值最高的品种，即便需要去除误差因素影响，相关的其余品种同样需要去除误差的影响，会影响产量高低排名吗？一个在不同区组产量（**均值**）最高的品种，不会因为误差影响其产量最高的排名的。**方差、误差，影响到了平均数的核心作用，是统计教学过多强调方差重要性的负面作用。**

4、一个品种的产量平均数最高，那么其与所有品种总平均的离差就是最大的，且是正值，从效应的角度理解，就是其品种正效应是最大的，自然就是参试品种中产量最高的品种。这就指明了育种选择的方向。这才是育种工作中第一重要的大问题。

5、配合力测定试验中，是类似的道理，只不过育种实践中，人们很自然的丢弃了 F 测验，因为，F 测验是基于 F 分布的，是基于方差的检验，而配合力仅是考察效应，考察离差，就够了，而优良基因带来的效应，通常很小，甚至微乎其微，需要长期积累，即便非要转换成方差，进行 F 测验一下，估计也不会显著的，问题是，育种选择是必须进行的。F 测验不显著，在方差分析中，可以考虑不进行多重比较，但育种选择却必须进行下去。

6、线性模型的分析方法越来越多的应用于生物统计和数量遗传学，而线性模型中的效应，即为离差，也就是说，以离差（效应）的分析方法来解决育种实践中的问题，被现代育种越来越接受。

7、方差分析，方差，对以比较为目的的试验具有实际应用价值，在育种工作中，应用价值没有你想象的那么大。方差的本质是离差，**离差在育种中很重要，决定育种的方向，离差更好用，更实用，更常用，早期自交系的选择，就是差值选择育种。**

(十二) 一次重复的多点试验如何进行方差分析？

一次重复，其实并没有重复，所以无法计算得到试验随机误差，这样的话，按照方差分析表进行方差分析，就进行不下去了，因为试验中的地点效应，品种效应，都需要将随机误差作为F测验的分母，计算F值，再进行F测验。只能想变通的办法进行试验数据分析了。

下图为6个品种在4个地点，每地点内2个区组的一年多点试验，每一地点内有重复，可以估计试验误差。

农博士育种家 [一年多点区域试验统计分析]								
试验设计 方差分析 品种区域试验 配合力 遗传模型 数据分析 数据转换 退出								
A品种数 6			B地点数 4			区组数 2		
	A	B	C	D	E	F	G	H
1	一年多点品种区域试验统计分析							
2								
3			地点1	地点2	地点3	地点4		
4	品种1	区组1	6	7	10	5		
5		区组2	7	8	12	6		
6	品种2	区组1	5	7	13	6		
7		区组2	6	6	11	6		
8	品种3	区组1	6	8	9	7		
9		区组2	4	6	10	5		
10	品种4	区组1	9	9	9	8		
11		区组2	8	9	10	9		
12	品种5	区组1	15	14	13	10		
13		区组2	13	11	15	11		
14	品种6	区组1	10	10	9	10		
15		区组2	11	11	10	10		
16								
17								

如果，本身是一年多点的品种产量试验，地点内没有设重复，则上图的试验就演变成了下面的情形。

农博士育种家 [一年多点区域试验统计分析]								
试验设计 方差分析 品种区域试验 配合力 遗传模型 数据分析 数据转换 退出								
A品种数 6			B地点数 4			区组数 2		
	A	B	C	D	E	F	G	H
1	一年多点品种区域试验统计分析							
2								
3			地点1	地点2	地点3	地点4		
4	品种1	区组1	6	7	10	5		
5								
6	品种2	区组1	5	7	13	6		
7								
8	品种3	区组1	6	8	9	7		
9								
10	品种4	区组1	9	9	9	8		
11								
12	品种5	区组1	15	14	13	10		
13								
14	品种6	区组1	10	10	9	10		
15								
16								
17								

此时用一年多点的试验进行统计分析就无法进行下去了。

解决办法：将地点当成区组，按照单因素随机完全区组进行统计分析。

农博士育种家 [单因素]

试验设计 方差分析 品种区域试验 配合力 遗传模型 数据分析 数据转换 退出

材料数 6 区组数 4

	A	B	C	D	E	F	G	H
1	随机区组试验统计分析							
2		区组1	区组2	区组3	区组4			
3	材料1	6	7	10	5			
4	材料2	5	7	13	6			
5	材料3	6	8	9	7			
6	材料4	9	9	9	8			
7	材料5	15	14	13	10			
8	材料6	10	10	9	10			
9								
10								
11								

从一年多点试验，变为单因素随机完全区组试验，本质上为试验设计的改变。

下面是两种方法分析结果的截图。试验数据不同，分析结果自然不同，但应该更关注分析结果中项目的不同，方差分析表的不同。

农博士育种家 [单因素]

试验设计 方差分析 品种区域试验 配合力 遗传模型 数据分析 数据转换 退出

材料数 6 区组数 4 指定对照

	A	B	C	D	E	F	G	H	I
1	随机区组试验统计分析								
2		区组1	区组2	区组3	区组4				
3	材料1	6	7	10	5				
4	材料2	5	7	13	6				
5	材料3	6	8	9	7				
6	材料4	9	9	9	8				
7	材料5	15	14	13	10				
8	材料6	10	10	9	10				
9									
10	通过了 Levene 方差齐性检验，检验结果：检测F值 =0.878615 < 3.1 (F0.05)								
11									
12	变异来源	DF	SS	MS	F	概率	F0.05	F0.01	
13	区组间	3	25.792	8.597	2.717	0.918	3.290	5.420	
14	材料间	5	97.708	19.542	6.176*	0.997	2.900	4.560	
15	误差	15	47.458	3.164					
16	总变异	23	170.958						
17									
18	最小显著差LSD(5%)=2.68065 LSD(1%)=3.70484								
19	LSD最小显著性差异多重比较								
20	处理	平均数	5%显著水平1%显著水平						
21	材料5	13.000	a	A					
22	材料6	9.750	b	AB					
23	材料4	8.750	bc	B					
24	材料2	7.750	bc	B					
25	材料3	7.500	bc	B					
26	材料1	7.000	c	B					
27									

试验设计及统计分析 数据拟合

北京中农博思科技发展有限公司 荣誉出品

农博士育种家 [一年多点区域试验统计分析]													
试验设计 方差分析 品种区域试验 配合力 遗传模型 数据分析 数据转换 退出													
A品种数		6		B地点数		4		区组数		2		定义数据区 分析	
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	一年多点品种区域试验统计分析												
2													
3			地点1	地点2	地点3	地点4							
4	品种1	区组1	6	7	10	5							
5		区组2	7	8	12	6							
6	品种2	区组1	5	7	13	6							
7		区组2	6	6	11	6							
8	品种3	区组1	6	8	9	7							
9		区组2	4	6	10	5							
10	品种4	区组1	9	9	9	8							
11		区组2	8	9	10	9							
12	品种5	区组1	15	14	13	10							
13		区组2	13	11	15	11							
14	品种6	区组1	10	10	9	10							
15		区组2	11	11	10	10							
16													
17	通过了 Levene 方差齐性检验, 检验结果: 检测F值 =1.70319 < 2.82(F0.05)												
18													
19	各个地点间误差均方同质性测验, 样本卡方值= 1.703(p>0.05) f(0.05)= 2.820, 各个地点间差异不显著, 可以进行方差分析。												
20													
21													
22	变异来源	DF	SS	MS	F[固]	概率[固]	F0.05[固]	F0.01[固]	F[随]	概率[随]	F0.05[随]	F0.01[随]	
23	地点内区组4		3.833	0.958	0.906	0.520	2.870	4.430					
24													
25	A品种	5	191.917	38.383	36.268*	1.000	2.710	4.100	9.200*	1.000	2.900	4.560	
26	B地点	3	68.417	22.806	21.549*	1.000	3.100	4.940	5.466*	0.990	3.290	5.420	
27	AXB	15	62.583	4.172	3.942*	0.998	2.220	3.130	3.942*	0.998	2.220	3.130	
28	误差	20	21.167	1.058									
29	总变异	47	347.917										
30	以下为固定模型基础上的多重比较												
31	A品种间比较												

在种子不多或者经费制约等背景下，将一年多点试验安排成了“没有重复的一年多点”，侧重的可能更多是品种稳定性的考量，这样的安排，有得有失，在进行稳定性分析评价的同时，尽可能利用试验数据，获得更多有用信息，无疑非常重要，采用“随机区组”试验分析方法对“没有重复的一年多点”试验进行数据分析，就很好的解决了这个问题，兼顾了稳定性分析和方差分析，还能省事省力省钱。

(十三) 区域试验中的变异系数 (CV) 用于评价试验质量

变异系数又称离散系数，是标准差与平均数之比，以百分数表示。单因素随机完全区组试验，一年多点试验、多年多点试验，可以计算地点对应的变异系数，用于比较不同地点的试验质量。笼统的说，一个试验中，不同地点的 CV 比较，CV 值低的小区，试验质量高。

1、同一地块的不同区组，CV 相对高的，表明该区组相对其余区组，肥力均匀性差；或者由于管理措施增大了该区组内品种间差异

2、如果试验异地安排区组，区组间 CV 的高低差异，还应该包含品种的环境互作效应。因此，依据 CV 来评价试验质量，要结合试验的具体安排考虑。

CV 也可以用于试验品种，其分析同样要考虑，差异因素是否包含品种的环境互作效应。

(十四) 没有重复的多点试验怎么分析

n 个材料安排在 v 个不同地点，进行产量试验，没有设置重复，可以考虑使用单因素随机区组（增产设计）进行数据分析，操作思路和要求如下：

1、用试验中的对照来充当考察品种，估计试验误差。

2、对照的数量至少 2 个。（如 958，335）

3、每个对照在 v 个地点最好有 1 个观测值，缺区可以，只要别全缺。

4、如果试验安排的时候忘了安排对照，找测试点补齐，补齐要点（同时期、同地块、同管理措施，注意可能需要进行必要的面积换算）

5、可以做方差分析，考察试验中的地点效应是否显著存在、考察品种效应是否显著存在。

6、可以依据地点效应对材料数据进行矫正，更符合材料的真实表现，更科学的评价材料表现。

如果有两年以上试验，客观上就构成了“重复”数据，将两年或多年数据一并分析，就可以考察品种稳定性，适宜性。

稳定性用 RSD 考察，适应性用地点相关的环境指数，或者用线性混合模型考察效应。

（详细内容参照农博士育种技术理论群 40082612，相关文档《没有重复的多点试验怎么分析.pdf》）

三、线性模型

$y=a/(x-b)$ 不是一个线性模型， $y=a*e*b(x-c)^2$ 也不是一个线性模型，都是非线性模型。方差分析模型 ($y_{ij} = \mu + \tau_i + \varepsilon_{ij}$) 属于线性模型。

（一）线性模型之方差分析模型

线性模型是一类统计模型的总称，它包括了线性回归模型、方差分析模型、协方差分析模型和线性混合效应模型等，许多生物、医学、经济、管理、地质、气象、农业、工业、工程技术等领域的现象都可以用线性模型来近似描述，因此，线性模型成为现代统计学中应用最为广泛的模型之一。

线性模型研究的目的往往是寻求因变量与自变量之间客观存在的依赖关系。

线性模型中的方差分析模型，它的自变量往往表示某种效应的存在与否，因而只能取 0,1 两个值，0 表示没有这种效应，1 表示有这种效应。由于比较两个或多个因素效应大小的统计分析在统计学上叫做方差分析，所以，对应地，将这种线性模型叫方差分析（线性）模型。在一些文献中，也将其称为试验设计模型，这是因为它所分析的数据往往跟一个预先安排好的实验相联系。

（二）两种方差分析方法

一个试验设计可以用两种分析方法进行方差分析：

第一种方法是平方和分解法。

将数据总变异平方和按其来源（各种因子和随机误差）进行分解，得到各因子平方和及误差平方和，接下来是基于各因子平方和与误差平方和大小的比较，这种方法叫平方和分解法，就是传统的方差分析方法，这种方法需要的预备知识较少，一般在一些初等统计书中都采用此方法。农业院校教授的就是这种方法。

第二种方法是线性模型的方差分析方法。

既然方差分析模型 ($y_{ij} = \mu + \tau_i + \varepsilon_{ij}$) 是一类线性模型，就可以用一般线性模型的估计与检验的方法应用于这种模型，其实质就是线性回归的分析方法，采用离差平方和最小的最小二乘法，进行数据分析，回归。

对完全区组试验设计，可以直接采用平方和分解法进行方差分析，分析准确性高，对不完全区组试验设计，由于试验效应的混杂，多采用线性模型的方法，求解效应。

（三）混合线性模型

一个线性分析模型，由于又有固定效应，又有随机效应，所以就是“混合”了，效应是混合的。

植物育种值分析模型，就是混合线性模型。

$$Y = \mu + \text{年度效应} + \text{地点效应} + \text{种群效应} + \text{品种育种值} + \sigma$$

y: 数据观测值，

μ : 总体均值

年度效应：试验年度的效应（固定效应）

地点效应：不同试验地点的效应（固定效应）

种群效应：不同种群的效应（固定效应）

品种育种值：品种的加性遗传效应（随机效应）

σ : 误差

四、遗传力（Heritability）

（一）遗传力定义

遗传力这个概念，很多教科书都表述其为数量遗传学最重要的概念，可见其重要性。但是植物育种中，一会出现“遗传力”一会又表述为“遗传率”，反映了对这一概念内涵把握的模糊性，目前大有用“遗传率”代替“遗传力”的迹象。实际育种工作中，在单因素品种比较试验，计算考察试验的遗传力，本身就反应了对遗传力的理解和育种实际应用中存在问题。BLUP MME 巧妙引入了性状遗传力，作为分析计算的重要参数，对遗传力的准确理解，直接跟 BLUP 分析应用挂钩。因此，准确理解和运用遗传力，对植物育种当下尤其重要。

广义遗传力指数量性状基因型方差占表现型方差的比率。

$$h_B^2 = \frac{V_G}{V_P} = \frac{V_A + V_D + V_I}{V_P}$$

狭义遗传力指数量性状加性遗传方差占表现型方差的比率。

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

遗传力（Heritability），反映了一个性状受遗传效应影响有多大，受环境效应影响有多大。遗传力是性状、群体和环境三者的综合体现，群体遗传结构和环境改变，都会影响性状的遗传力。遗传力这个概念，很多教科书都表述其为数量遗传学最重要的概念，可见其重要性。遗传力是对性状而言的，因此，不应该考虑某群体的遗传力，某个体的遗传力。

《植物数量遗传学》（孔繁玲）介绍的世代对比法，亲子回归法等，和《家畜育种学》（张沅）介绍的亲子资料估计遗传力，同胞资料估计遗传力，都是利用亲代和子代数据来计算性状

的遗传力。遗传力作为重要的遗传参数，其估计方法，通常采用约束最大似然法 REML (Restricted Maximum Likelihood)。狭义遗传力是育种值对表型值的决定系数。从实际估计来看，遗传力是育种值对表型值的回归系数。

简单理解，性状遗传力就是性状从亲代向子代传递的能力，因此，**遗传力 (Heritability)** 这个叫法没有问题，比遗传率更符合概念内涵，不能仅从计算公式强调其比率含义。对育种实践而言，遗传力所反映的性状向子代的传递能力，才是育种选择的核心重点。遗传力是对性状而言的，因此，不应该考虑某群体的遗传力，某个体的遗传力。遗传力用更通俗一点的话语来描述就是，以玉米为例，只要是玉米，其产量性状的遗传力大概就是多少数值，其衡量的是这个作物的性状，不是某个品种，也不是某次试验。但是可以通过某次试验来估计这个性状的遗传力。

(二) 组内相关法估算广义遗传力

依据方差分析表，可以直接估算单因素随机完全区组的广义遗传力。

13	变异来源	DF	SS	MS	F	概率	F0.05	F0.01
14	区组间	2	27.561	13.780	8.398*	0.995	3.740	6.510
15	材料间	7	34.080	4.869	2.967*	0.960	2.760	4.280
16	误差	14	22.972	1.641				
17	总变异	23	84.613					
18								
19								

下面是试验的期望均方 (EMS) 构成。

品种	8-1	SS_r	MS_r	$\sigma_e^2 + 3\sigma_v^2$
误差	$(3-1)(8-1)$	SS_e	MS_e	σ_e^2

$$V_E = \sigma_e^2 = 1.641 \quad V_G = \sigma_v^2 = (4.869 - 1.641)/3 = 1.076$$

$$V_P = \sigma_e^2 + \sigma_v^2 = 1.641 + 1.076 = 2.717$$

$$\text{广义遗传力: } h_b^2 = V_G / V_P = 1.076 / 2.717 = 0.396$$

单因素随机完全区组试验适合采用组内相关法，但仅能估算广义遗传力。

(三) 实际育种中遗传力育种应用思考

1、单因素随机区组试验有必要计算产量性状的遗传力吗？没必要。一年多试验同样没有必要。狭义遗传力的计算需要亲代子代数据，这两种试验都不具备这个数据条件，只能依据方差分析得出环境方差 (V_e)，计算出广义遗传力，没有任何育种价值。在没必要的情下去考虑遗传力，徒增困惑烦恼，得不偿失。

2、遗传力的计算复杂，不知道某个性状的遗传力，也查不到权威资料，怎么办？方法很简单：比较子代性状跟亲代的相似性，越像遗传力越高。

3、遗传力是面向育种性状的，这个工作专业搞数量遗传的专家会去考虑，普通育种单位，育种者个人，没必要介入。

但是研究的结论一定要了解：

①狭义遗传力高，表型与基因型的相关程度高，采用混合选择较好。

②狭义遗传力低，表型不能很好的反应上下代间的相似程度，重点应放在系谱法和后裔鉴定上，

必要时也可采用间接选择法或综合选择法。

③有性繁殖作物，狭义遗传力决定着亲子间相似程度，决定着由表型预测育种值的可靠程度，但是无性繁殖作物，是由广义遗传力决定亲属间相似程度。

④性状的狭义遗传力高，说明性状的遗传受环境影响小，而且说明亲子代间的相似程度高，由表型判断基因型的命中率高，选择容易见效，反之，选择不易见效。

⑤性状的狭义遗传力高，可在早代选择，且种植规模可较小，低时，宜在晚代选择，且种植规模宜较大。

⑥产量为低遗传力性状。

五、重复力 (Repeatability)

重复力 (repeatability) 是衡量一个数量性状在同一个个体多次度量值之间的相关程度的指标，该概念来自于动物育种，就是同一个个体性状多次度量值间的相关系数。在数据有重复观测值的情况下，才需要有重复力参数。重复力的估计原理，从效应剖分来看，可以将环境效应剖分为持久性环境效应(E_p)和暂时性环境效应(E_T)。

$$E = E_p + E_T$$

$$P = G + E = G + E_p + E_T$$

假定基因型效应，持久性环境效应和暂时性环境效应都不存在相关，可以将表型方差 (V_p) 剖分为

$$V_p = V_G + V_{EP} + V_{ET}$$

重复力定义为：

$$r_e = \frac{V_G + V_{EP}}{V_p} = \frac{V_G + V_{EP}}{V_G + V_{EP} + V_{ET}}$$

动物育种中，持久性环境效应可以对应养殖场所，如：猪场，多次度量都在一个环境中进行，植物育种中，其实也存在类似情况，比如试验场地固定，材料每年总是种植在同一试验地。

对比遗传力和重复力公式，重复力应该始终大于遗传力。

重复力可以提高 BLUP 分析中估计育种值的准确度。

六、选择育种

(一) 选择差和选择响应

育种的核心问题是“选择”，无论 DH 系、分子标记、基因组选择，还是常规育种，不解决好“育种选择”的理论方法问题，可能都是在烧钱！

但是翻过的数量遗传学教材，对育种选择的理论和方法没有讲透，反而在遗传力上纠缠不清，客观上给实际育种工作增加了困扰。建议读者可以找来数量遗传学教材“选择育种”章节，与本章节对照阅读。

数量遗传学对选择育种给出了选择差、选择响应的定义：

选择差：亲本群体内中选组的表型均值与选择前亲代群体表型均值之差

(对某一数量性状进行选择时，入选群体的平均值将与原始群体平均值产生一定的离差，

即选择差。)

$$S = \mu_0' - \mu_0$$

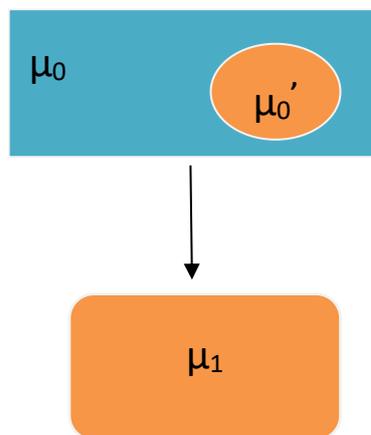
μ_0' 为亲本群体内中选组的表型均值， μ_0 亲代总体的表型均值。

选择响应：实施选择后群体平均值的改变量。

(这些入选亲本后代构成群体的平均值与上述原始群体平均值之差就是遗传进度，也叫遗传增益，也有人称其为遗传获得量、选择响应或选择反应。遗传进度反映了选择效果的大小。)

$$R = \mu_1 - \mu_0$$

μ_1 为中选组子代群体的表型均值， μ_0 选择前亲代总体的表型均值。



μ_0 选择前亲代总体的表型均值。

μ_0' 为亲本群体内中选组的表型均值

μ_1 为中选组子代群体的表型均值

选择响应在数量遗传学中，出现了以下一些相似概念，比如：遗传进度、遗传增益、遗传获得量、选择反应。这本身就说明了在选择育种这个领域存在一些不尽相同的理解。这点需要引起注意。

有的数量遗传学教材从选择育种与遗传力关系的角度来给出育种建议，将育种选择问题弄得很复杂，而实际应用价值不大。

以下观点供参考：

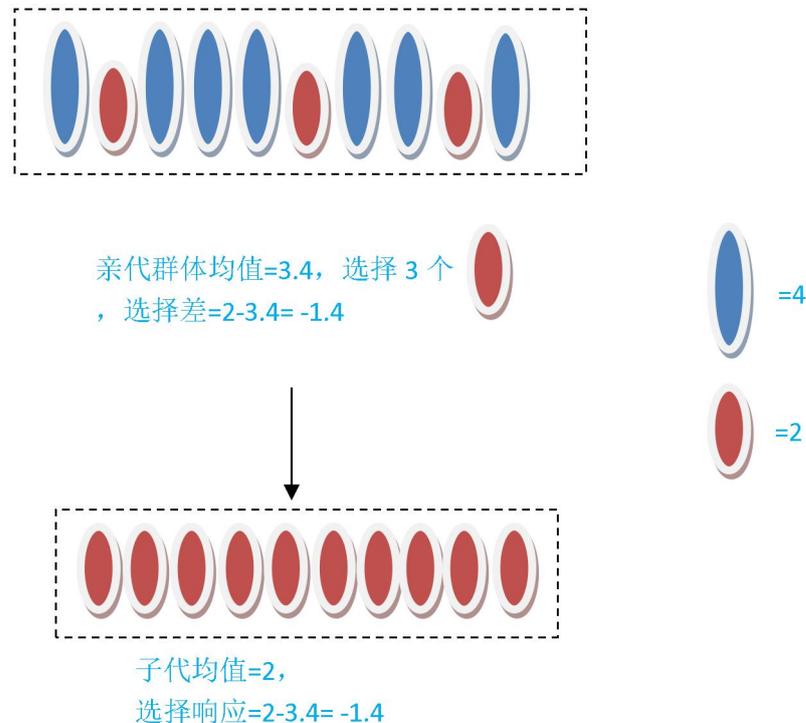
1、育种的选择是带有明确标准或指向的，这一选择标准与被选群体中的遗传方差表现不是面对面的关系，更象是从面中挑选个体（点）。无论表型方差还是遗传方差，或者加性方差，都是一个离散变异幅度，是对群体而言的，而育种选择，常从个体的角度进行，1株、1穗、1朵。植物育种群体中的有益变异，可能影响整个育种结果，因此育种者的眼睛更喜欢盯着与众不同的材料，哪怕仅有1棵，因为这样的材料可能代表着新基因新基因型的出现。而这样与众不同的变异，环境影响以及误差干扰都无法遮盖材料的表现。

2、一个差值是为了选择（S），一个差值是为了检验选择（R）。

3、选择响应，意味着上次选择的结果已经出来了。立足选择后的结果进行育种方向的把握和材料的选择。选择响应和配合力在育种策略上是相似的，无论是什么机理、哪种基因互动，好的选择结果就包含了好的基因。

孔繁玲老师《植物数量遗传学》（中国农业大学出版社 2006 年 6 月第 1 版），在《选择的基本效应》章节提到了“从动植物育种的角度看，选择的主要目标在于改进群体的均值水平，这就是选择响应（或选择进度）”。

选择差、选择响应示意图



图中选择差=-1.4，选择响应也是-1.4，这是最理想的情况，由于环境的影响，即便选择100% 精准，子代均值也很难=2，应该常大于2（含误差），因此选择响应通常小于选择差。

注意事项：

如果苦恼各种数量遗传学概念，就用离差来理解选择育种。

选择前的群体与选择后的群体，产生了离差，而在控制或者忽略环境影响的情况下，离差就对应这两个群体中基因效应的改变，选择就是不断增加基因效应的聚合，而这种聚合，就来自一次又一次的离差积累。

（二）差值选择育种法

思路：

依据世代间育种性状的观测值差值变化，进行材料选择。依据不同性状的期望方向，进行 >0 的增益积累或者 <0 的增益积累。

适用：常规纯系育种；杂交亲本自交系选育。

原理：选择内在改变了群体的基因频率，基因型频率，外在体现在表现型值的变化。在控制好环境影响的情况下，差值反映了选择前与选择后个体或者群体间基因平均效应总和的改变。换

一个角度理解就是基因效应的累积。

示例：产量，越高越好，就是>0 的增益积累；追求矮秆，株高就是<0 的增益积累。

说明：

该方法不是什么新方法，甚至育种工作本身就是这样进行的，之所以这样提，是为了从应用层面方便理解和操作。

七、配合力

（一）配合力定义

一个亲本在其所有杂交组合中表现的平均效应，称为该亲本的一般配合力效应。

$$g_i = y_{i.} - y_{..}$$

$$S_{ij} = y_{ij} - y_{..} - g_i - g_j$$

一般配合力（GCA）反映特定亲本交配效应的平均水平，一般配合力高的亲本，其杂交组合多数表现良好；

特殊配合力（SCA）反映一对特定亲本间杂交的特定配合效应，与该双亲在其他组合中的表现无关。

效应对应的是平均数间的离差。

（二）配合力计算示例

以常用的不完全双列杂交 NC II 配合力试验示例：

试验参数：父本（A），5 个，母本（B）4 个，3 个重复区组，每个区组内部 1 个观测值。

父本母本	区组 1	区组 2	区组 3
	观测值 1	观测值 1	观测值 1
A1B1	31	29	24
A1B2	27	25	26
A1B3	18	18	10
A1B4	14	16	14
A2B1	19	21	21
A2B2	21	17	27
A2B3	14	13	12
A2B4	16	9	14
A3B1	20	14	16
A3B2	16	18	12
A3B3	16	11	14
A3B4	12	9	10
A4B1	26	35	36
A4B2	30	31	34
A4B3	20	18	17

A4B4	26	24	21
A5B1	12	8	11
A5B2	10	10	15
A5B3	7	13	10
A5B4	11	11	8

表格中，所有数据的平均值：17.8

计算一般配合力，根据配合力的定义，跟 A1 有关的数据就是下表中的数据，计算平均值。

A1B1	31	29	24
A1B2	27	25	26
A1B3	18	18	10
A1B4	14	16	14
A1 平均值	21		
A1 一般配合力	$21-17.8= 3.2$		

（把所有跟 A2 有关的观测数据找出来）

A2B1	19	21	21
A2B2	21	17	27
A2B3	14	13	12
A2B4	16	9	14
A2 平均值	17		
A2 一般配合力	$17-17.8= -0.8$		

（把所有跟 A3 有关的观测数据找出来）

A3B1	20	14	16
A3B2	16	18	12
A3B3	16	11	14
A3B4	12	9	10
A3 平均值	14		
A3 一般配合力	$14-17.8= -3.8$		

（把所有跟 A4 有关的观测数据找出来）

A4B1	26	35	36
A4B2	30	31	34
A4B3	20	18	17
A4B4	26	24	21
A4 平均值	26.5		
A4 一般配合力	$26.5-17.8= 8.7$		

（把所有跟 A5 有关的观测数据找出来）

A5B1	12	8	11
A5B2	10	10	15
A5B3	7	13	10
A5B4	11	11	8
A5 平均值	10.5		
A5 一般配合力	$10.5-17.8= -7.3$		

父本 A	一般配合力
A1	3.2
A2	-0.8

A3	-3.8
A4	8.7
A5	-7.3
合计	0

母本 B 的一般配合力怎么计算？

（把所有跟 B1 有关的观测数据找出来）

A1B1	31	29	24
A2B1	19	21	21
A3B1	20	14	16
A4B1	26	35	36
A5B1	12	8	11
B1 平均值	21.53333		
B1 一般配合力	$21.53-17.8= 3.73333$		

（把所有跟 B2 有关的观测数据找出来）

A1B2	27	25	26
A2B2	21	17	27
A3B2	16	18	12
A4B2	30	31	34
A5B2	10	10	15
B2 平均值	21.26667		
B2 一般配合力	$21.26667-17.8=3.46667$		

（把所有跟 B3 有关的观测数据找出来）

A1B3	18	18	10
A2B3	14	13	12
A3B3	16	11	14
A4B3	20	18	17
A5B3	7	13	10
B3 平均值	14.06667		
B3 一般配合力	$14.06667-17.8= -3.73333$		

（把所有跟 B4 有关的观测数据找出来）

A1B4	14	16	14
A2B4	16	9	14
A3B4	12	9	10
A4B4	26	24	21
A5B4	11	11	8
B4 平均值	14.33333		
B4 一般配合力	$14.33333-17.8= -3.46667$		

母本 B	一般配合力
B1	3.73333
B2	3.46667
B3	-3.73333
B4	-3.46667
合计	0

到这里，父母本的一般配合力计算完成了。

A1B1 特殊配合力应该怎么计算？

根据定义

$$S_{ij} = y_{ij} - y_{..} - g_i - g_j$$

$$S_{A1B1} = y_{A1B1} - y_{..} - g_{A1} - g_{B1}$$

A1B1 观测值	31	29	24
A1B1 均值	28		
总平均	17.8		
A1(GCA)	3.2		
B1(GCA)	3.73333		
A1B1 特殊配合力	28-17.8-3.2-3.73333=3.26667		

其余组合就不一一计算了。

下面用几张连续的贴图，来展示农博士育种家软件的分析结果。

农博士育种家 [不完全双列杂交 (BC II)]

试验设计 方差分析 品种区域试验 配合力 遗传模型 数据分析 退出

父本数 5 母本数量 4 区组数 3 观察值数 1 CK 定义数据区

	A	B	C	D	E	F	G	H	I	J	K	L
1	父本母本	区组1	区组2	区组3								
2		观测值1	观测值1	观测值1								
3	A1B1	31	29	24								
4	A1B2	27	25	26								
5	A1B3	18	18	10								
6	A1B4	14	16	14								
7	A2B1	19	21	21								
8	A2B2	21	17	27								
9	A2B3	14	13	12								
10	A2B4	16	9	14								
11	A3B1	20	14	16								
12	A3B2	16	18	12								
13	A3B3	16	11	14								
14	A3B4	12	9	10								
15	A4B1	26	35	36								
16	A4B2	30	31	34								
17	A4B3	20	18	17								
18	A4B4	26	24	21								
19	A5B1	12	8	11								
20	A5B2	10	10	15								
21	A5B3	7	13	10								
22	A5B4	11	11	8								
23												
24	变异来源	DF	SS	MS	F[固]	概率[固]	F0.05[固]	F0.01[固]	F[随]	概率[随]	F0.05[随]	F0.01[随]
25	区组间	2	7.600	3.800	0.428	0.343	3.240	5.210				
26	处理组合	19	2964.267	156.014	17.554**	1.000	1.920	2.510				
27	父本A	4	1851.600	462.900	52.083**	1.000	2.620	3.860	16.631**	1.000	3.260	5.410
28	母本B	3	778.667	259.556	29.204**	1.000	2.850	4.340	9.325**	0.998	3.490	5.950
29	父本x母本	12	334.000	27.833	3.132**	0.996	2.020	2.690	3.132**	0.996	2.020	2.690
30	误差	38	337.733	8.888								
31	总变异	59	3309.600									
32												
33	杂交组合均值列表:											
34												
35	均值	p2-1	p2-2	p2-3	p2-4							
36	p1-1	28.0000	26.0000	15.3333	14.6667							
37	p1-2	20.3333	21.6667	13.0000	13.0000							
38	p1-3	16.6667	15.3333	13.6667	10.3333							
39	p1-4	32.3333	31.6667	18.3333	23.6667							
40	p1-5	10.3333	11.6667	10.0000	10.0000							

116					
117	母本间比较				
118	最小显著差LSD(5%)=2.20342 LSD(1%)=2.95019				
119	LSD最小显著性差异多重比较				
120	处理	平均数	5%显著水	1%显著水平	
121	B1	21.533	a	A	
122	B2	21.267	a	A	
123	B4	14.333	b	B	
124	B3	14.067	b	B	
125					
126	Duncan新复极差多重比较				
127	P	SSR0.05	SSR0.01	LSR0.05	LSR0.01
128	2	2.89	3.89	2.225	2.994
129	3	3.04	4.06	2.340	3.125
130	4	3.12	4.16	2.402	3.202
131					
132	处理	平均数	5%显著水	1%显著水平	
133	B1	21.533	a	A	
134	B2	21.267	a	A	
135	B4	14.333	b	B	
136	B3	14.067	b	B	
137					
138	父本 x 母本间比较				
139	最小显著差LSD(5%)=4.92699 LSD(1%)=6.59682				
140	LSD最小显著性差异多重比较				
141	处理	平均数	5%显著水	1%显著水平	
142	A4B1	32.333	a	A	
143	A4B2	31.667	a	A	
144	A1B1	28.000	ab	AB	
145	A1B2	26.000	bc	ABC	
146	A4B4	23.667	bcd	BCD	
147	A2B2	21.667	cde	BCDE	
148	A2B1	20.333	def	CDEF	
149	A4B3	18.333	efg	DEFG	
150	A3B1	16.667	fgh	EFGH	
151	A1B3	15.333	ghi	EFGHI	
152	A3B2	15.333	ghi	EFGHI	
153	A1B4	14.667	ghij	FGHI	
154	A3B3	13.667	ghij	GHI	
155	A2B3	13.000	hij	GHI	
156	A2B4	13.000	hij	GHI	
157	A5B2	11.667	ij	HI	
158	A3B4	10.333	i	HI	

育种软件中给出了更多的信息，下节我们会更深入的进行解读。

(四) griffing 配合力分析介绍

B. GRIFFING 1956 年发表了 Concept of general and specific combining ability in relation to diallel crossing systems, 给出了 4 种试验设计，每种设计区分为固定和随机 2 种效应就是 8 种配合力分析方法。

4 种试验这几如下：

配合方式	1.亲本自交+正反交组合	2.亲本自交+正反交组合	3.无亲本自交仅亲本正反交组合	4.无亲本自交仅亲本正交组合
组合数量公式 4 个材料为例	$p*p$ $4*4=16$	$p(p+1)/2$ $4(4+1)/2=10$	$p(p-1)$ $4*(4-1)=12$	$p(p-1)/2$ $4*(4-1)/2=6$
详细组合 (自交)	1X1 1X2 1X3 1X4 2X1 2X2 2X3 2X4	1X1 1X2 1X3 1X4 2X2 2X3 2X4 3X3	1X2 1X3 1X4 2X1 2X3 2X4 3X1 3X2	1X2 1X3 1X4 2X3 2X4 3X4

	3X1	3X4	3X4	
	3X2	4X4	4X1	
	3X3		4X2	
	3X4		4X3	
	4X1			
	4X2			
	4X3			
	4X4			

数学模型:

$$x_{ijkl} = \mu + v_{ij} + b_k + (bv)_{ijk} + e_{ijkl}$$

μ 为群体平均效应, v_{ij} 为第 ij 基因型的效应, b_k 为第 k 个区组效应, $(bv)_{ijk}$ 为第 ij 个基因型与第 k 个块的相互作用, e_{ijkl} 为第 $ijkl$ 个体特有的环境效应。根据配合力的概念, v_{ij} 可以进一步分解为: $v_{ij} = g_i + g_j + s_{ij}$ (不包括反交效应) 和 $v_{ij} = g_i + g_j + s_{ij} + r_{ij}$ (包括反交效应)。

g_i, g_j 为亲本 ij 一般配合力效应, s_{ij} 为亲本 i 和亲本 j 间的特殊配合力效应, r_{ij} 为亲本 i 和亲本 j 间的反交效应。

(GRIFFING 配合力详细资料下载, 可加入农博士育种技术理论群 40082612)

(四) 理解配合力

1、育种实际工作中, 配合力往往用于决定材料的选择, 比如, 早期的配合力测定, 目的是为了日后更容易出品种, 材料选育方向掌控。

2、配合力分析中, 应关注效应(离差), 以产量为例, 产量越高越好, 因此, 如果一个品种分析得到的一般配合力效应是正值, 说明这个品种的产量基因效应在参试群体(参照系)中有优势, 背后的基因可能就有更有利的累积, 而这种优势或者劣势是基于参试群体而言的。因为, 所有参试品种的一般配合力效应合计=0。

3、为了提供多一些分析信息, 育种软件会给出, 方差分析表, F 测验, 多重比较等内容, 注意, 这些信息仅供参考而已, 因为, 配合力分析, 无论 F 测验显著与否, 用户更关心的是配合力的效应, 或者说, 更关心的是哪一些品种的一般配合力效应高或者低。即便 F 测验不显著, 配合力分析还是要做的, 育种的材料选择更是必须进行的。

4、配合力是对交配双方而言的, 一般配合力, 特殊配合力都是基于杂交组合。在交配群体变化的情况下, 如 AXB, 改为 AXC, 配合力会跟着变化, 可是 A 基因没有变化。有的数量遗传学教材中提到的配合力能够反映交配材料的基因加性效应, 是育种值的一半, 就有问题了。配合力仅能说明交配群体之间的一种配合能力关系, 不能简单的把这种能力关系上升到基因层面。

一批材料, 做过 3 年的配合力测定, 如果将这 3 年的数据放在一起分析, 是不是得出的结果准确性更高, 对材料的评价会更客观准确, 但是不同年份对产量是有不同影响的, 应该怎样分析数据呢? 配合力分析模型没有考虑年度、地点这些因素。

（五）配合力育种价值思考

1、植物杂交育种中的配合力方法实质

配合力方法实质是用品种实际表现结果来逆向筛选基因，以产量为例，产量高的，好基因自然就在里面了。至于基因如何发挥的作用，什么模式，什么效应，都包含了。配合力是一种杂交育种的兜底策略，这也是让数量遗传学很尴尬很无奈的地方，非要用基因等理论来解释配合力反而把问题解释复杂了，当然，不是说配合力跟基因没有关系。

配合力测定实质是一个商业育种行为，是已经明确了骨干自交系后，选择与之配对的另一些自交系，因此配合力测定并不是一个单纯的育种选择问题。**从公司行为来说，进行这样的选择是必要的。**公司骨干自交系，决定了与之配套的自交系，固定方向选择在先了。所谓的高配合力自交系，准确理解是高基因效应自交系。跟骨干自交系配，表现好的，就是可以用的材料。于是，想明白了的大公司或小公司，就尽可能扩大组配规模，增加发现好组合的机会就是了。骨干自交系决定了一切。

2、育种者的配合力幻想

育种者难免会有一个冲动或叫幻想，两个不起眼的材料，一杂，没准就能获得一个特别优秀的组合出来！

理论上存在这种可能，如果你一直用配合力的理论，甚至会更加相信这样的事情早晚会发生。配合吗，两个很普通的自交系，配合能力没准就是极高的。如果，理性思考，是不是这样的，杂交组合的基因分别来自父母本各一半，如果父母本的基因加性效应都不高，或者仅有一个亲本基因加性效应高，那么后代组合的基因加性效应就不会高，那后代组合实际表现会特别优异吗？郑单 958 的亲本郑 58 制种产量高达 500-700 公斤，本身就指明了一条育种的技术路线（亲本的基因加性效应要高，要注意不是杂种优势导致的产量高）。

子代的基因加性效应不高，万一特殊配合力特别高呢，提醒您思考的是，即便这个组合的特殊配合力高，您觉得踏实吗？基因的加性效应可以稳定遗传，特殊配合力高可靠吗？为什么大的种业公司特别注重看品种的稳定性，品种不稳，隐患太大，而且，为什么一个组合今年长的好（特殊配合力高），明年一般，甚至不好了呢（特殊配合力低）？

特殊配合力计算中，将环境效应，地点效应，年度效应，种群效应都归结为特殊配合力。

张世煌老师在北京农业 2007 第 5 期的文章《中美两国玉米育种思路和技术水平的比较》中提到：“在过去 80 年，玉米产量的持续高速度增长取决于自交系一般配合力的持续提高。”

八、育种值

植物育种应该重视育种值，并不是否定配合力，而是育种值和配合力搭配实用。

（一）育种值定义

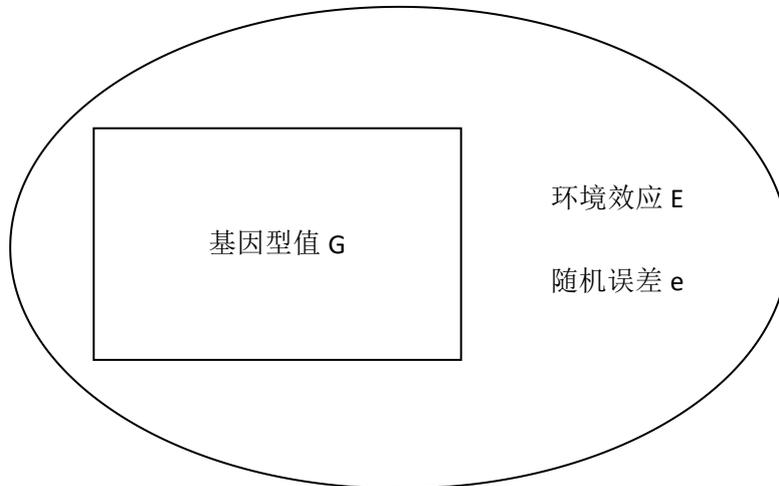
$$\text{表现型值} = \text{基因型值} + \text{环境效应} + \text{随机误差}$$

$$P = G + E + e$$

表现型值：对表现型的测量值称为该个体的表现型值。以 **P** 表示。

基因型值：由个体的基因型所决定的数值称为该个体的基因型值。以 **G** 表示。

表现型值、基因型值间的关系



基因的平均效应：群体内接受了某个基因的所有个体的平均基因型值距群体平均的**离差**称为该基因的平均效应。

育种值

理论定义：一个个体的育种值，就是它所携带的基因的平均效应的总和。

一个群体内基因平均效应、育种值间的关系（育种值理论定义）

基因 A	基因型 AA, Aa 个体求均值，与群体均值离差 = 基因 A 的平均效应	AaBb 个体育种值，是 4 项累加和。
基因 a	基因型 aa, Aa 个体求均值，与群体均值离差 = 基因 a 的平均效应	
基因 B	基因型 BB, Bb 个体求均值，与群体均值离差 = 基因 B 的平均效应	
基因 b	基因型 bb, Bb 个体求均值，与群体均值离差 = 基因 b 的平均效应	
...		
基因 N	基因型 NN, Nn 个体求均值，与群体均值离差 = 基因 N 的平均效应	

如果将图中的群体内每一个个体的育种值求累加和，结果=0.

按照育种值的理论定义，一个个体的育种值是跟所在群体有关的，对个体育种值的计算度量离不开所在群体，否则没有意义。也就是说，无法用理论定义来比较两个群体间的育种值，比如，选择前群体和选择后群体。

但是育种值又是跟加性效应对应，这样就好理解为什么数量遗传学给出了育种值的“实际定义”。

实际定义：如果一个个体与来自群体内的许多个体随机交配，则该个体的育种值为其子代均值与群体平均离差的两倍。

实际育种中，通常不清楚控制性状的基因型，群体内的基因频率、基因型频率也是未知，因此理论育种值是不能够直接度量的，能够知道的只是包含育种值在内的各种遗传效应和环境效应共同作用得到的表型值，因此只能利用统计分析方法，通过表型值和个体间的亲缘关系来对育种值进行估计，这就是**估计育种值（EBV）**。基因型值的加性效应就是育种值。

（二）理论定义与实际定义辨析

- 1、育种值的理论定义定位于一个群体内部，一个群体内，不同个体间由于基因差异存在不同的育种值，但是一个群体内，所有个体的育种值总和=0.
- 2、育种值的实际定义立足于子代群体和亲代群体间的均值离差。
- 3、如果考虑选择育种中的选择响应，在只考虑加性效应，不考虑上位性等其他效应的前提下，估计育种值是选择响应的 2 倍。选择育种，本质上，就是依据选择前群体和选择后群体间，群体均值、基因加性效应的变化在进行育种选择。
- 4、常规育种或者杂交育种选育自交系，依据基因的加性效应选择，就是依据育种值在进行选择，因为只有加性效应可以稳定遗传给子代；一个育种材料内在基因加性效应的改良提升，就是育种值的积累，外在就是群体均值的改变，这也是差值选择育种法的含义。
- 5、BLUP 计算植物育种的育种值，一定意义上可以认为是从育种值理论定义出发计算一个群体内不同个体的育种值，尽管其没有采用基因频率，只要模型设置合理，符合群体效应实际情况。因为其分析模型来自数量遗传学的基本模型，并且符合一个群体内所有个体育种值合计=0.
- 6、BLUP 计算植物育种的育种值，比采用群体间均值差异的方式更加合理、数据更准确，因为群体内个体间亲缘关系，以及群体的环境影响，都被纳入了模型，因此计算得到的育种值更接近真实的育种值。

简单的说，就是选择育种和 BLUP 育种值方法都是利用育种值在进行育种，一个从实际定义出发一个从理论定义出发，一个算的粗糙一些，一个算的更准确一些。

对植物育种来说，只要选择，就可以用育种值。

对杂交育种来说，选自交系，除去早期配合力测定，其余时候用育种值。后期配组合用配合力。

（三）配合力≠基因加性效应（育种值）

配合力试验也是用一个亲本（父本）跟群体的多个亲本（母本）杂交，测量产生子代的均值，来进行计算，好像仅是没有乘以 2，这不就是育种值吗？跟育种值计算一模一样的。不是这样的。

试验数据，是由基因型呈现出来的，尽管数据背后含有基因。如果你对比配合力和 BLUP 育种值模型差别，不难发现，BLUP 模型中的随机效应 a_j 由于 $cov(y, u) = ZG$, $G^{-1} = A^{-1}k$ 的约束，其中 A 是加性遗传相关矩阵，就使得模型中的 a_j 就是基因的加性效应（育种值）。如果是使用育种值，某试验中一个亲本自交系材料育种值是最高的，其理论含义是清晰明确的，就表明该亲本材料相比群体中其它亲本材料聚合了更多的优势有利基因。因此该亲本在本次试验中，传递给子代群体的基因有优势。而且这个结论是在排除了试验地点，年度，等各种固定效应影响后得出的。如果是使用一般配合力，上面这种描述，在运气好的情况下仅是可能性中的情况之一。育种只有加性效应可以稳定遗传给子代。

九、经验式育种的选择技术方法

（一）经验式育种的困惑

经验育种仿佛一夜间被贴上了“落后”的标签，因为“现代商业化育种”来了。

几乎没有哪个育种人能脱离跟“经验育种”的联系，除非他不做育种，那所从事的育种怎么就被称为落后的“经验育种”了，怎么就甚至还带有贬义色彩了呢？

经验式育种，其不足在于天机不可泄露，育种保密性不好保障，因此，往往规模上不去，经验覆盖不过来太多的材料，一个专家选育的规模会小，而没有规模，会影响到出品种，影响到商业化运营。提倡商业化育种的时候，并没有否定经验育种的**经验价值**，在任何一个领域，**谁会否定宝贵的经验积累**。

所以，当在讨论经验育种过时落后的时候，是不是其准确的含义是，经验育种的育种组织形式落后了，不开放，保守，从而限制了育种事业的发展。

“现代商业化育种”究竟先进在什么地方？是不是需要认真思量思量！

经验式育种思想错了吗？经验育种模式、育种思路错了吗？

（二）经验式育种选择使用的是均值

经验式育种是高明的，是有效的，是科学的！

育种本质上是个**选择问题**，可能有的育种者还不认同这样的观点，但其实每天的工作都是围绕着选择展开的，10年出一个品种，也是逐步选择得来的，那进一步讲，支撑育种的核心之一其实是数学，就是生物统计以及数量遗传学，这样说的话，很多人会觉得诧异甚至不好接受，好像觉得自己的育种仅是用到了几个统计试验方法，对比法、间比法、单因素随机区组、一年多点、多年多点，就这么几个简单的试验，不少业内人士，一致的观点就是，前期用点试验设计，育种的后期，用点统计的方法，仅此而已。这跟育种理论中，没有讲清楚生物统计、数量遗传学发挥的作用有关系，就是说，育种中，生物统计、数量遗传学一直发挥着重要的作用，但是育种理论并没有很好很明确的讲清楚，于是，不少的育种者甚至认为生物统计、数量遗传学对育种作用不大，或者没法在育种中去用，不好用，没法用。

育种家的经验育种是高明的，是有效的，是科学的！**因为经验育种，是按照均值的变化在选择**，这代株高降低了，就是这代株高均值低于上一代，这就是基因平均效应的变化，均值的

差值，还是离差，不过是负的离差。每一个育种性状分别往育种家希望的方向走，变化，就都是均值的变化。经验育种一直以来就是这么做的，这样的选择就是依据基因效应在进行高效科学的育种，区别仅是育种家自己明不明白这个道理，这也是生物统计和数量遗传学没有讲清楚的一个地方。生物统计和数量遗传学一直在支撑育种前行，也许这个道理太简单，遗传学家懒得去说，也许这个问题一直没有被从育种选择的角度被提出，困扰着实际育种工作者的思想，但不论怎样，育种从来都是在遵循着生物统计和数量遗传学的基本原理前行的。

育种的本质是选择，无论统计试验也好，还是 DH 系、QTL、MAS、GS 也罢，育种需要好的选择理论、选择技术，正确的选择技术选择方法，甚至比好的材料、好的试验设备工具更重要，决定育种收获的，是育种的选择水平。在关注育种新技术手段的时候，别忽视了育种根本的东西。

十、BLUP 分析方法

(一) BLUP 介绍

1948 年，Henderson 提出了 Best Linear Unbiased Prediction (BLUP) 分析方法，最佳线性无偏估计，为了降低计算难度，Henderson 给出了混合模型方程组 BLUP MME:

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \quad , \quad \text{固定效应 } \mathbf{b}$$

$$\hat{\mathbf{u}} = \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\hat{\mathbf{b}}) \quad , \quad \text{随机效应 } \mathbf{u}$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

化简后

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\mathbf{k} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}, \quad \mathbf{k} = \frac{\sigma_e^2}{\sigma_a^2}$$

并证明了 BLUP MME 的 b,u 解等于 BLUP 的 b,u 解。

BLUP 是最佳线性无偏预测，BLUE 是最佳线性无偏估计，BLUP 对应随机效应（育种值），BLUE 处理固定效应。BLUP，BLUE 分别是两个线性方程组，将它们叠放在一起，组成一个全新的方程组。由于又有固定效应，又有随机效应，所以就是“混合”了，因为都是线性方程组，都来自同一个线性模型，于是就叫“线性混合模型”。准确的说法是混合模型方程组（mixed modal equations MME），这就是 Henderson 1948 年提出的著名 BLUP MME 方法。因为随机效应育种值的预测更有意义，于是 BLUP 就更出名了，BLUE 很少去提。BLUP 方法就是为计算动物育种值才出现的，最初计算机跟不上，解不了大的方程组，使用受限，70 年代，BLUP 改变了动物育种，如今成为世界公认的最好的动物育种值分析方法。90 年代，美国开始在植物育种上采用。

(二) 单性状 BLUP 育种值计算示例

地点	个体	父本	母本	观察值
A	2	-	-	225
A	3	-	-	200
A	4	-	-	255
B	5	1	3	250
B	6	1	3	198
B	7	2	4	245
B	8	2	4	260
B	9	2	4	210

(本例参照张沅老师《家畜育种学》，个别观察值改动)

BLUP 线性模型:

$$y_{ij} = h_i + a_j + e_{ij}$$

y_{ij} 第 i 个地点中个体 j 的观测值

h_i 第 i 个地点的固定效应

a_j 第 j 个个体的育种值

e_{ij} 误差

写成矩阵形式

$$\begin{bmatrix} 225 \\ 200 \\ 255 \\ 250 \\ 198 \\ 245 \\ 260 \\ 235 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \\ a_9 \end{bmatrix} + \begin{bmatrix} e_{12} \\ e_{13} \\ e_{14} \\ e_{25} \\ e_{26} \\ e_{27} \\ e_{28} \\ e_{29} \end{bmatrix}$$

$$y = Xh + Za + e$$

注意：向量 a 中不仅包含了有观测值的个体的育种值，还包含了没有观测值的个体的育种值 a_1 。

$$X'X = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}$$

$$X'Z = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$Z'X = (X'Z)'$$

$$Z'Z = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$X'y = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 225 \\ 200 \\ 255 \\ 250 \\ 198 \\ 245 \\ 260 \\ 210 \end{bmatrix} = \begin{bmatrix} 680 \\ 1163 \end{bmatrix}$$

$$Z'y = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 225 \\ 200 \\ 255 \\ 250 \\ 198 \\ 245 \\ 260 \\ 210 \end{bmatrix} = \begin{bmatrix} 0 \\ 225 \\ 200 \\ 255 \\ 250 \\ 198 \\ 245 \\ 260 \\ 210 \end{bmatrix}$$

个体间的加性遗传相关矩阵

1 2 3 4 5 6 7 8 9

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0.5 \\ 0 & 0 & 1 & 0 & 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 & 0 & 1 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0.5 & 1 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 & 0 & 1 & 0.5 & 0.5 \\ 0 & 0.5 & 0 & 0.5 & 0 & 0 & 0.5 & 1 & 0.5 \\ 0 & 0.5 & 0 & 0.5 & 0 & 0 & 0.5 & 0.5 & 1 \end{bmatrix} \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{matrix}$$

个体间的加性遗传相关矩阵 A 的构建，从系谱关系得出，如材料 1 和材料 5 的“0.5”，因为 5 的父本是 1，所以两个材料间同源相同基因的比例为 50%。材料与自身的相似度为 1。A 矩阵以对角线对称。

其逆矩阵为

$$A^{-1} = \begin{bmatrix} 2 & 0 & 1 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 2.5 & 0 & 1.5 & 0 & 0 & -1 & -1 & -1 \\ 1 & 0 & 2 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 1.5 & 0 & 2.5 & 0 & 0 & -1 & -1 & -1 \\ -1 & 0 & -1 & 0 & 2 & 0 & 0 & 0 & 0 \\ -1 & 0 & -1 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 & 0 & 2 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 & 0 & 0 & 2 & 0 \\ 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

设，已知该性状的遗传力为 $h^2=0.3333$ ，则

$$k = (1-h^2)/h^2 = (1-0.3333)/0.3333 = 2$$

于是，参照 BLUP MME 得到混合模型方程组

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + A^{-1}k \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

$$\begin{bmatrix} 3 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 4 & 0 & 2 & 0 & -2 & -2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 6 & 0 & 3 & 0 & 0 & -2 & -2 & -2 \\ 1 & 0 & 2 & 0 & 5 & 0 & -2 & -2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 3 & 0 & 6 & 0 & 0 & -2 & -2 & -2 \\ 0 & 1 & -2 & 0 & -2 & 0 & 5 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 0 & -2 & 0 & 0 & 5 & 0 & 0 & 0 \\ 0 & 1 & 0 & -2 & 0 & -2 & 0 & 0 & 5 & 0 & 0 \\ 0 & 1 & 0 & -2 & 0 & -2 & 0 & 0 & 0 & 5 & 5 \end{bmatrix} \begin{bmatrix} \hat{h}_1 \\ \hat{h}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \\ \hat{a}_7 \\ \hat{a}_8 \\ \hat{a}_9 \end{bmatrix} = \begin{bmatrix} 680 \\ 1163 \\ 0 \\ 225 \\ 200 \\ 255 \\ 250 \\ 198 \\ 245 \\ 260 \\ 210 \end{bmatrix}$$

解线性方程组，获得解如下：

$$\begin{aligned}
\hat{h}_1 &= 226.349514563107 \\
\hat{h}_2 &= 231.563106796117 \\
\hat{a}_1 &= -0.951456310679609 \\
\hat{a}_2 &= 0.184466019417483 \\
\hat{a}_3 &= -9.41747572815533 \\
\hat{a}_4 &= 10.1844660194175 \\
\hat{a}_5 &= -0.460194174757275 \\
\hat{a}_6 &= -10.8601941747573 \\
\hat{a}_7 &= 6.83495145631069 \\
\hat{a}_8 &= 9.83495145631069 \\
\hat{a}_9 &= -0.165048543689314
\end{aligned}$$

对该示例，我们通过 BLUP 模型分析，可以得到如下信息：

- 1、地点 A 和地点 B 的固定效应。
- 2、父本材料 1 的育种值，这就是动物育种中提到的“公牛的产奶量”。如果我们这个试验性状是关于后代奶牛产奶量的话。
- 3、其余材料 2、3、4、5、6、7、8、9 的育种值。
- 4、1-9 的材料育种值合计不等于 0。

\hat{a}_1	-0.95146
\hat{a}_2	0.184466
\hat{a}_3	-9.41748
\hat{a}_4	10.18447
\hat{a}_5	-0.46019
\hat{a}_6	-10.8602
\hat{a}_7	6.834951
\hat{a}_8	9.834951
\hat{a}_9	-0.16505
合计	5.349514563

如果我们添加一个约束条件，让 1-9 的材料育种值=0，那相应矩阵会发生变化。

$$\begin{bmatrix}
3 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 5 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 4 & 0 & 2 & 0 & -2 & -2 & 0 & 0 & 0 \\
1 & 0 & 0 & 6 & 0 & 3 & 0 & 0 & -2 & -2 & -2 \\
1 & 0 & 2 & 0 & 5 & 0 & -2 & -2 & 0 & 0 & 0 \\
1 & 0 & 0 & 3 & 0 & 6 & 0 & 0 & -2 & -2 & -2 \\
0 & 1 & -2 & 0 & -2 & 0 & 5 & 0 & 0 & 0 & 0 \\
0 & 1 & -2 & 0 & -2 & 0 & 0 & 5 & 0 & 0 & 0 \\
0 & 1 & 0 & -2 & 0 & -2 & 0 & 0 & 5 & 0 & 0 \\
0 & 1 & 0 & -2 & 0 & -2 & 0 & 0 & 0 & 5 & 0 \\
0 & 1 & 0 & -2 & 0 & -2 & 0 & 0 & 0 & 0 & 5
\end{bmatrix}
\begin{bmatrix}
\hat{h}_1 \\
\hat{h}_2 \\
\hat{a}_1 \\
\hat{a}_2 \\
\hat{a}_3 \\
\hat{a}_4 \\
\hat{a}_5 \\
\hat{a}_6 \\
\hat{a}_7 \\
\hat{a}_8 \\
\hat{a}_9
\end{bmatrix}
=
\begin{bmatrix}
680 \\
1163 \\
0 \\
225 \\
200 \\
255 \\
250 \\
198 \\
245 \\
260 \\
210
\end{bmatrix}$$

这是未添加约束条件的方程组，下面是在方程组最后一行添加了约束条件的新方程组。

$$\hat{a}_1 + \hat{a}_2 + \hat{a}_3 + \hat{a}_4 + \hat{a}_5 + \hat{a}_6 + \hat{a}_7 + \hat{a}_8 + \hat{a}_9 = 0$$

$$\begin{bmatrix}
 3 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
 0 & 5 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\
 0 & 0 & 4 & 0 & 2 & 0 & -2 & -2 & 0 & 0 & 0 \\
 1 & 0 & 0 & 6 & 0 & 3 & 0 & 0 & -2 & -2 & -2 \\
 1 & 0 & 2 & 0 & 5 & 0 & -2 & -2 & 0 & 0 & 0 \\
 1 & 0 & 0 & 3 & 0 & 6 & 0 & 0 & -2 & -2 & -2 \\
 0 & 1 & -2 & 0 & -2 & 0 & 5 & 0 & 0 & 0 & 0 \\
 0 & 1 & -2 & 0 & -2 & 0 & 0 & 5 & 0 & 0 & 0 \\
 0 & 1 & 0 & -2 & 0 & -2 & 0 & 0 & 5 & 0 & 0 \\
 0 & 1 & 0 & -2 & 0 & -2 & 0 & 0 & 0 & 5 & 0 \\
 0 & 1 & 1 & -1 & 1 & -1 & 1 & 1 & 1 & 1 & 6
 \end{bmatrix}
 \begin{bmatrix}
 \hat{h}_1 \\
 \hat{h}_2 \\
 \hat{a}_1 \\
 \hat{a}_2 \\
 \hat{a}_3 \\
 \hat{a}_4 \\
 \hat{a}_5 \\
 \hat{a}_6 \\
 \hat{a}_7 \\
 \hat{a}_8 \\
 \hat{a}_9
 \end{bmatrix}
 =
 \begin{bmatrix}
 680 \\
 1163 \\
 0 \\
 225 \\
 200 \\
 255 \\
 250 \\
 198 \\
 245 \\
 260 \\
 210
 \end{bmatrix}$$

蓝色数字为加上了约束条件后的矩阵内容。

计算结果：

$$\begin{aligned}
 \hat{h}_1 &= 226.66999002991 \\
 \hat{h}_2 &= 231.96111665005 \\
 \hat{a}_1 &= -1.05483549351944 \\
 \hat{a}_2 &= -0.208374875373875 \\
 \hat{a}_3 &= -9.59322033898304 \\
 \hat{a}_4 &= 9.79162512462612 \\
 \hat{a}_5 &= -0.651445663010963 \\
 \hat{a}_6 &= -11.051445663011 \\
 \hat{a}_7 &= 6.44107676969093 \\
 \hat{a}_8 &= 9.44107676969093 \\
 \hat{a}_9 &= -0.984845463609174
 \end{aligned}$$

\hat{a}_1	-1.054835494
\hat{a}_2	-0.208374875
\hat{a}_3	-9.593220339
\hat{a}_4	9.791625125
\hat{a}_5	-0.651445663
\hat{a}_6	-11.051445666
\hat{a}_7	6.44107677
\hat{a}_8	9.44107677
\hat{a}_9	-0.984845464
合计	2.129611167

计算结果，用 1-9 的育种值合计=2.129611167 不等于 0。不等于 0，这跟线性方程组的其余约束条件有关。但是，可以看出来，添加了一个 1-9 的材料育种值=0 的额外约束后，1-9 育种值合计从 5.349514563 下降到了 2.129611167。线性模型随机效应部分通常要添加 $\Sigma \hat{u} = 0$ 的约束条件，这样便于材料之间的直接比较。

有几点说明：

1、我们直接采用了遗传力 $h^2=0.3333$ ，这是一个经验值，不是该试验自身的数据。因此会导致 BLUP 数据分析的准确性下降。

$$k = \frac{\sigma_e^2}{\sigma_a^2} = \frac{\sigma_y^2 - \sigma_a^2}{\sigma_a^2} = \frac{1 - \sigma_a^2 / \sigma_y^2}{\sigma_a^2 / \sigma_y^2} = \frac{1 - h^2}{h^2}$$

而如果不采用 h^2 , 那么我们需要知道 σ_e^2 、 σ_a^2 , 误差方差和加性遗传方差。这部分内容就是线性模型的方差分量估计要解决的问题, 目前来看, 解决办法不是很好。有的方法估计出的 σ_e^2 是负数。

2、BLUP 的加性遗传相关矩阵, 反应的是试验材料之间的亲缘关系, 可以通过系谱获得, 在此不做单独讲解, 请用户自行查阅相关书籍。

3、 \hat{a}_1 育种值的获得, 除了示例中介绍的方法, 还有另外一种方法。根据计算获得的子代的育种值, 来直接计算得出, 子代育种值的一半来自父本。

4、从动物育种的实践来看, 采用经验值的做法, 并没有影响到 BLUP 的育种应用。

5、在方程组中, 直接添加固定效应=0 的约束, 本身没有问题, 但是影响了矩阵的对称正定性, 建议采用 harvey 约束条件降维处理。

(三) 多性状植物育种 BLUP 分析模型

BLUP 方法可以应用于多个性状, 以两个性状为例, 给出混合模型方程组。

第一个性状的模型为

$$y_1 = X_1 b_1 + Z_1 a_1 + e_1$$

第二个性状的模型为

$$y_2 = X_2 b_2 + Z_2 a_2 + e_2$$

令

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad X = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad Z = \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix}, \quad a = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \quad e = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

$$y = Xb + Za + e$$

$$E(a) = 0, E(e) = 0$$

令

$$G_0 = \begin{bmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{bmatrix}, \quad R_0 = \begin{bmatrix} r_{11} & r_{12} \\ r_{12} & r_{22} \end{bmatrix} = \begin{bmatrix} r_{11} & 0 \\ 0 & r_{22} \end{bmatrix}$$

g_{11} 是第一个性状的加性遗传方差, g_{22} 是第二个性状的加性遗传方差, g_{12} 为两个性状间的遗传协方差, r_{11} 是第一个性状的误差方差, r_{22} 是第二个性状的误差方差, r_{12} 是两个性状间的误差协方差。通常认为性状间的误差协方差=0。

$$\text{Var}(a) = G = \begin{bmatrix} Ag_{11} & Ag_{12} \\ Ag_{12} & Ag_{22} \end{bmatrix}, \quad A \text{ 为个体间的加性遗传相关矩阵}$$

$$\text{Var}(e) = R = \begin{bmatrix} I r_{11} & I r_{12} \\ I r_{12} & I r_{22} \end{bmatrix}$$

令

$$G_0^{-1} = \begin{bmatrix} g^{11} & g^{12} \\ g^{12} & g^{22} \end{bmatrix}, \quad R_0^{-1} = \begin{bmatrix} r^{11} & r^{12} \\ r^{12} & r^{22} \end{bmatrix}$$

则

$$G^{-1} = \begin{bmatrix} A^{-1}g^{11} & A^{-1}g^{12} \\ A^{-1}g^{12} & A^{-1}g^{22} \end{bmatrix}, \quad R^{-1} = \begin{bmatrix} I r^{11} & I r^{12} \\ I r^{12} & I r^{22} \end{bmatrix}$$

$$\begin{bmatrix} X_1'X_1r^{11} & X_1'X_2r^{12} & X_1'Z_1r^{11} & X_1'Z_2r^{12} \\ X_2'X_1r^{12} & X_2'X_2r^{22} & X_2'Z_1r^{12} & X_2'Z_2r^{22} \\ Z_1'X_1r^{11} & Z_1'X_2r^{12} & Z_1'Z_1r^{11} + A_1^{-1}g^{11} & Z_1'Z_2r^{12} + A_1^{-1}g^{12} \\ Z_2'X_1r^{12} & Z_2'X_2r^{22} & Z_2'Z_1r^{12} + A_2^{-1}g^{12} & Z_2'Z_2r^{22} + A_2^{-1}g^{22} \end{bmatrix} \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} = \begin{bmatrix} X_1'y_1r^{11} + X_1'y_2r^{12} \\ X_2'y_2r^{12} + X_2'y_2r^{22} \\ Z_1'y_1r^{11} + Z_1'y_2r^{12} \\ Z_2'y_2r^{12} + Z_2'y_2r^{22} \end{bmatrix}$$

(四) 植物育种 BLUP 模型

$$Y = \mu + \text{年度效应} + \text{地点效应} + \text{种群效应} + \text{品种育种值} + \sigma$$

Y: 数据观测值,

μ : 总体均值

年度效应: 试验年度的效应

地点效应: 不同试验地点的效应

种群效应: 不同种群的效应

品种育种值: 品种的加性遗传效应

σ : 误差

动物育种发展出动物模型、公畜模型、公畜母畜模型、外祖父模型等, 植物育种可以套用动物模型。

每个观测值可以从构成效应的角度, 剖分为总体均值, 不同年度效应, 不同地点效应, 不同种群效应, 品种的育种值效应。

$$\text{表现型值} = \text{基因型值} + \text{环境效应} + \text{随机误差}$$

$$P = G + E + e$$

如果对照数量遗传学的基本模型, 就好理解为什么有的教材说 BLUP 线性模型特别适合计算育种值。加性遗传模型下, $G=M + A$, M 为群体均值, A 为加性效应, 对应育种值。

(五) 植物育种 BLUP 模型中种群效应

模型中有一个效应, 种群效应 (用 b 表示), 按固定效应处理。

BLUP 既然需要抽取数据中的平均数, 当有明显种群效应存在的时候, 将种群效应以固定效应分离, 更方便开展育种工作。不同种群, 不同血缘, 先从种群的角度考察材料, 再深入到种群内部, 筛选材料。

种群效应示例, 比如 100 个品种进行育种值分析, 其中 30 个材料偏向一个种群, 另外 70 个材料, 没有必要分群。

分群不分群两种情况比较:

在没有进行种群分类的情况下, 将 100 个材料 (b) 等同视为 1 群, 系统给出 (默认种群 0) 的固定效应。 $\sum b_i = 0$, i 的范围, 1-100, 同时, 系统对每个品种育种值 (用 a 表示) 会有一个 $\sum a_i = 0$ 的约束。

100 个材料

种群效应约束条件

育种值约束条件

不分群（默认种群 b0）	$\sum b_i = 0 \quad i(1\sim 100)$	$\sum a_i = 0 \quad i(1\sim 100)$
种群 b1, 种群 b0 (30 个材料组成种群 b1, 其余 70 个材料组成种群 b0)	$\sum b_0 = 0 \quad (70 \text{ 个材料})$ $\sum b_1 = 0 \quad (30 \text{ 个材料})$	$\sum a_i = 0 \quad i(1\sim 100)$

种群效应可以衡量种群之间的水平高低，因此进行育种值的比较，应在种群内进行。

（六）有重复观测值时的 BLUP 模型

当个体存在重复观测值时，观测值 y 可剖分为：

$$y = \sum b + a + p + e$$

b 为固定效应， a 为随机效应（育种值）， p 为随机永久性环境效应。表型方差可进一步分解为：

$$\sigma_y^2 = \sigma_a^2 + \sigma_p^2 + \sigma_e^2$$

其 MME 为：

$$\begin{bmatrix} X'X & X'Z_1 & X'Z_2 \\ Z_1'X & Z_1'Z_1 + A^{-1}k_1 & Z_1'Z_2 \\ Z_2'X & Z_2'Z_1 & Z_2'Z_2 + Ik_2 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} x'y \\ z_1'y \\ z_2'y \end{bmatrix}$$

其中

$$k_1 = \sigma_e^2 / \sigma_a^2 = (1-r) / h^2$$

$$k_2 = \sigma_e^2 / \sigma_p^2 = (1-r) / (r \cdot h^2)$$

$$r = (\sigma_a^2 + \sigma_p^2) / \sigma_y^2 = \text{重复力}$$

$$h^2 = \sigma_a^2 / \sigma_y^2 = \text{遗传力}$$

当观测值存在重复的时候，应该使用有重复观测值的 BLUP 模型。

（六）BLUP 植物育种值模型用于一般配合力分析

BLUP 植物育种值模型可以进行一般配合力分析，计算结果更准确，分析信息更丰富。

同一试验设计，基于相同试验数据，分别采用配合力分析与 BLUP 育种值分析两种方法进行分析，比较两种方法分析结果，为育种材料选择服务。

这部分内容较多，感兴趣的朋友请到，农博士育种技术理论群（40082612），下载《同一试验数据不完全双列杂交（NC II）配合力分析与 BLUP 育种值分析的结果比较》PDF 文档，查看全部内容，在此仅贴出比较结果。给出结论。

同一试验两种方法计算结果比较

材料	一般配合力分析	BLUP 植物育种值分析
2012[201203]	-21.4074	-33.7019
2012[201215]	-14.8519	-21.5344
2012[201202]	-9.1852	-13.6375
2012[201217]	3.2037	4.6452
2012[201216]	11.6481	16.8892
2012[201201]	30.5926	45.9144

如果两种方法分析的不是同一个指标，进行盲目比较，就没有任何意义。植物育种中，一般配合力和育种值，模型有区别，但二者的服务指向都是材料选择这一育种目标，都用于衡量一个自交系材料可稳定遗传给后代的能力(植物育种中，一般配合力可靠有效，就是这个原因)。一般配合力和育种值模型不同，但是相近，实际上，二者想要衡量的是同一个内容，但是一般配合力提供的信息太少，结果数据没有育种值准确。

表中数据比较结果：

1、配合力分析和 BLUP 育种值分析，同一试验、同一数据、同一材料，分析结果不同。

模型不同，分析结果不同，这很好理解，但另一方面，配合力模型，计算过程没有分离试验中的年度效应、地点效应，会导致一般配合力效应值失真。

2、两种方法模型虽不同，但本试验分析结果中，6 个亲本材料的一般配合力排序和育种值排序，顺序相同。这也是个很有意思的情况，一方面部分验证了前面说的，一般配合力和育种值是衡量的同一个内容；另一方面应将出现的排序一样看成是一种概率巧合，如果两者排序不同，你认同哪一种排序结果呢？

分析建议

育种中使用的配合力分析方法，配合力计算，仅是均值的简单相减，没有考虑试验本身的误差、区组间的差异、试验材料间的系谱亲缘关系、年度间差异等，这是配合力分析方法的硬伤，这也是 BLUP 育种值分析模型的诸多优势。

BLUP 育种值分析可以提供试验材料诸多亲本材料育种值估计，这对育种选择也非常有价值。

建议育种者，在进行配合力分析之后，选用同样的试验材料进行 BLUP 育种值分析，两种方法比较，来最终确定材料的育种价值。

（七）植物育种中育种值应用思路参考

植物育种，如果是纯系育种，本质上就是育种值的积累提升，直接用育种值就可以，有的育种者借鉴玉米育种使用的一般配合力，就是这个道理。

如果是非纯系育种，以玉米为例，选自交系早期用配合力，确定材料方向，之后可以考虑用育种值来进行自交系选育，后期，自交系成型后，再用配合力选配组合。自交系不要选成混合种质（AB 群，SS-NSS 都是一个意思），不利于出品种。理想的模式是配组合时，利用特殊配合力，进一步提升组合表现。

（八）植物育种采用动物育种 BLUP 方法的技术理论问题探讨

1、动物 BLUP 拿到植物上用，可行吗?依据是什么？

先说结论：可行。

为了不产生歧义，下面先引用张沅先生《家畜育种学》中的一段文字来说明：

“近交程度 亲缘选配 关键在于亲缘关系的远近。而对个体间亲缘关系远近，可用其下一代的

近交系数予以度量。近交系数这一概念，首先是由 Wright (1921) 提出。据其定义，近交系数是指配子基因效应间的相关，用 F 表示。Malecot (1948) 给出了另一个直观的概念。一般来说，两个基因相同分为两种情况。一是两个基因功能相同，亦即用分子遗传学的话说，2 个基因的核苷酸序列相同。例如，某一特定牛的血型位点 F 上的两个等位基因均为 F 。我们称这两个基因同态相同 (identical by state)。二是 2 个基因因为是同一祖先同一基因的拷贝而相同。它们自然也有相同的功能和核苷酸序列。这样的基因，我们称为同源相同 (identical by descent)。同源相同基因必为同态相同基因，同态相同基因却不一定是同源相同基因。不过我们所考虑的一些甚至大多数同态相同基因也是同源相同基因，因为它们可能源于同一遥远的、甚至几百代前的祖先，只是我们通常只考虑几代的系谱。近交也只是相对于基础群 (base population) 而言。对基础群，我们假设个体间没有亲缘关系，而同态相同基因为非同源相同基因。Cotterman (1940) 对基因做了类似区别，只是用词不同，即以自交纯合基因 (autozygous gene) 表示同源相同基因，而用异源纯合基因 (allozygous gene) 表示同态相同基因。据此，Malecot 将近交系数定义为一个个体同一个基因座上的两个基因为同源相同基因的概率。对于一个特定的基因座，两个基因可能是也可能不是同源相同基因。但对所有基因座，近交系数却意味着可以预期有比例等于 F 的基因座携带同源相同基因。换句话说，在基础群中比例为 F 的杂合基因座已变成了纯合子。对于整个群体，就一个基因座而言，可以进行同样的推理。假设一个 F_2 群体作为基础群，杂合子的频率为 50%。于是，若下一代的近交系数为 40%，即意味着基础群中有 40% 的杂合基因座在子一代变成了纯合基因座，或换言之，即近交群中 70% ($50%+50%*40%$) 的基因座达到了纯合。”

这里解释一个名词“基础群 (base population)”，是加性遗传相关 A 矩阵构建中的一个概念，是互相间没有亲缘关系的一群基础动物，这是一个假定，如果现有基础群间有亲缘关系，那就上溯，用更早的祖先来构建基础群，所以基础群假定动物间没有亲缘关系或者亲缘关系不明。考虑植物的情况，基础群就是基础材料了。互相间不清楚系谱的材料或者就是亲缘关系很远视为无的一组材料，构成了“基础群”。基础群决定了，基础群材料的近交系数=0。跟植物材料多为自交，动物是异交没有关系。也可以理解为基础群内的材料间不存在近交。因此，henderson 的 A 矩阵构建不存在动物和植物的区别，动物 BLUP 育种值分析方法，完全适用于植物育种。

近交系数和亲缘系数，一个个体的近交系数等于其双亲亲缘系数的 1/2。近交系数反映的是材料间的亲缘关系，不能因为一个植物材料是纯合自交就认为其近交系数=1。一个个体与它本身的加性遗传相关为：

$$A_{xx} = 1 + f_x$$

在基础群中， $f_x=0$ 。

2、近交系数和亲缘系数

近交系数的计算公式：

$$F_X = \sum \left[\left(\frac{1}{2} \right)^{N_1+N_2+1} (1 + F_A) \right]$$

F_X ：个体 X 的近交系数；

N_1 ：由父亲到共同祖先所经世代数；

N_2 ：由母亲到共同祖先所经世代数；

F_A ：共同祖先本身的近交系数；

Σ ：所有共同祖先计算值的总和。

亲缘系数的计算公式：

$$R_{SD} = \frac{\sum [(\frac{1}{2})^N (1 + F_A)]}{\sqrt{(1 + F_S)(1 + F_D)}}$$

R_{SD} : 个体 S 和 D 间的亲缘系数;

N: 个体 S 和 D 分别到共同祖先的世代数之和, 即等于 $N_1 + N_2$;

F_S : 个体 S 的近交系数;

F_D : 个体 D 的近交系数;

F_A : 共同祖先的近交系数;

Σ : 所有共同祖先计算值的总和。

一个个体的近交系数等于其双亲亲缘系数 R_{SD} 乘上 $\sqrt{(1 + F_S)(1 + F_D)} / 2$, 在双亲都是没有亲缘关系的自交系的情况下, $F_S = 0, F_D = 0$, 则一个个体的近交系数等于其双亲亲缘系数的 1/2。

3、A 矩阵介绍

个体 i 和 j 之间的加性遗传相关是指在它们的基因组中具有同源相同 (identical by descent, IBD) 基因 (相同的且来自同一祖先的基因) 的比例, 或者说从个体 i 的基因组中随机抽取一个基因与个体 j 的基因组中随机抽取的一个基因同源相同的概率。它也可被理解为个体 i 和 j 的育种值 (加性遗传值) 之间的相关 (称为加性遗传相关)。根据这个定义, A 矩阵中的元素 a_{ij} , 即个体 i 和个体 j 之间的加性遗传相关的计算通式为

$$a_{ij} = \sum [(\frac{1}{2})^{n_i + n_j} (1 + f_A)]$$

其中, n_i 和 n_j 分别为连接个体 i 和个体 j 的一个途径中由 i 和 j 到它们的共同祖先 A 的世代数, f_A 为 A 的近交系数, Σ 表示当连接个体 i 和个体 j 的途径不止一条时, 要对所有途径求和。对比公式会发现, a_{ij} 就等于亲缘系数的分子项。

A 矩阵中对角线元素 a_{ii} , 即个体 i 与其自身的加性遗传相关被定义为:

$$a_{ii} = 1 + f_i$$

f_i 为个体 i 的近交系数。

4、A 矩阵构建规则

Henderson (1976) 给出了 A 矩阵的每一元素, 可以采用以下的递推公式直接构建出 A 矩阵:

$$\begin{aligned} a_{ii} &= 1 + 0.5a_{s_i d_i}, && \text{当 } s_i \text{ 和 } d_i \text{ 均已知} \\ &= 1 && \text{当 } s_i \text{ 和 } d_i \text{ 均未知} \\ a_{ij} &= a_{ji} = 0.5(a_{i s_j} + a_{i d_j}) && \text{当 } s_j \text{ 和 } d_j \text{ 均已知} \\ &= 0.5a_{i s_j} && \text{当 } s_j \text{ 已知, } d_j \text{ 未知} \end{aligned}$$

$$=0.5a_{id_j} \quad \text{当 } d_j \text{ 已知, } s_j \text{ 未知}$$

$$=0 \quad \text{当 } s_j \text{ 和 } d_j \text{ 均未知}$$

S_i (父), d_i (母) 个体 i 的父和母

S_j (父), d_j (母) 个体 j 的父和母

$a_{s_i d_i}$ 表示 s_i 和 d_i 之间的加性遗传相关

a_{is_j} 表示个体 i 和 s_j 之间的加性遗传相关

a_{id_j} 表示个体 i 和 d_j 之间的加性遗传相关

Henderson 同时给出了 A^{-1} 构建的规则, 解决了大型系谱 A 矩阵难以求逆的难题。为 BLUP 实际应用扫清了障碍。

5、A 矩阵示例

系谱数据:

个体	父	母
1	-	-
2	-	-
3	1	-
4	1	2
5	3	4
6	1	4
7	5	6

依照规则, 建立的 A 矩阵

$$A = \begin{bmatrix} 1 & 0 & 0.5 & 0.5 & 0.5 & 0.75 & 0.625 \\ 0 & 1 & 0 & 0.5 & 0.25 & 0.25 & 0.25 \\ 0.5 & 0 & 1 & 0.25 & 0.625 & 0.375 & 0.5 \\ 0.5 & 0.5 & 0.25 & 1 & 0.625 & 0.75 & 0.6875 \\ 0.5 & 0.25 & 0.625 & 0.625 & 1.125 & 0.5625 & 0.84375 \\ 0.75 & 0.25 & 0.375 & 0.75 & 0.5625 & 1.25 & 0.90625 \\ 0.625 & 0.25 & 0.5 & 0.6875 & 0.84375 & 0.90625 & 1.28125 \end{bmatrix}$$

依照规则, 建立的 A^{-1} 矩阵

$$A^{-1} = \begin{bmatrix} 2.33333333 & 0.5 & -0.66666667 & -0.5 & 0 & -1 & 0 \\ 0.5 & 1.5 & 0 & -1 & 0 & 0 & 0 \\ -0.66666667 & 0 & 1.83333333 & 0.5 & -1 & 0 & 0 \\ -0.5 & -1 & 0.5 & 3 & -1 & -1 & 0 \\ 0 & 0 & -1 & -1 & 2.6153846154 & 0.61538461538 & -1.23076923076 \\ -1 & 0 & 0 & -1 & 0.61538461538 & 2.6153846154 & -1.23076923076 \\ 0 & 0 & 0 & 0 & -1.23076923076 & -1.23076923076 & 2.46153846152 \end{bmatrix}$$

6、A 矩阵（系谱）的整理规则

BLUP 分析需要用到材料系谱, 在此给出系谱整理的规则。系谱有文字系谱和数字系谱, 文字系谱需要手工整理, 数字系谱通常由软件完成。

①系谱自祖先依次到后代排列，亲代排在子代前面，顺序必须无误。排在系谱最前面的就是“基础群（base population）”。

②每行系谱有 3 列，分别为：子代 父本 母本，列间用空格隔开，一个材料的亲本不明，用 0 代替。

示例 1:

农 A 0 0

表明材料“农 A”父本母本都不明

示例 2:

农 A 102 0

表明材料“农 A”父本是 102，母本不明

示例 3:

郑单 958 昌 7-2 郑 58

文字系谱转换为数字系谱，从祖先到子代依次用 1234... 替换，数字系谱示例如下：

1 0 0

2 0 0

3 1 0

4 1 2

5 3 4

6 1 4

7 5 6

正确的数字系谱，左 1 列数字大于 2、3 列数字。

A 矩阵和 A^{-1} 是 BLUP 分析理论和方法中的重要内容，对 A 矩阵的正确理解和构建至关重要。想要更系统全面了解 A 矩阵和 A 的逆矩阵，建议阅读 Henderson（1976）A Simple Method for Computing the Inverse of a Numerator Relationship Matrix Used in Prediction of Breeding Values，或者阅读动物育种的相关书籍。

7、植物育种要重视育种值和 BLUP 分析方法

Henderson 1948 年就提出了 BLUP MME 分析方法，动物育种在 70 年代就采用这一分析方法了，而植物育种，到今天依然没有认识到这一分析方法的科学性和重要性。植物育种以均值为选择依据即本书中的差值选择法，即便采用配合力，仅是均值的简单减法。在动物育种领域，动物育种的 BLUP 计算分析是采用目前最先进的计算机都不一定能够胜任的计算任务，植物育种和动物育种在数据计算分析上存在巨大的反差。这应该引起植物育种者的重视和反思。

困扰或者阻碍 BLUP 分析技术在植物育种领域应用的障碍，大概有以下几点：

7.1 遗传力和遗传率的概念困扰

遗传力这一数量遗传最重要的概念，在植物领域变的变幻莫测，想真正搞清楚遗传力，就去看动物育种学中关于遗传力的描述和计算方法，子代遗传了亲代多少相似性状，就是这一性状的遗传力，所以遗传力的计算，本质上是回归。BLUP 中一个重要的计算参数就是遗传力，比如分析产量，就需要产量性状的遗传力，分析子代获得了多少亲代的育种值。

遗传力用更通俗一点的话语来描述就是，以玉米为例，只要是玉米，其产量性状的遗传力大概就是多少数值，其衡量的是这个作物的性状，不是某个品种，也不是某次试验。但是可以

通过某次试验来估计这个性状的遗传力。

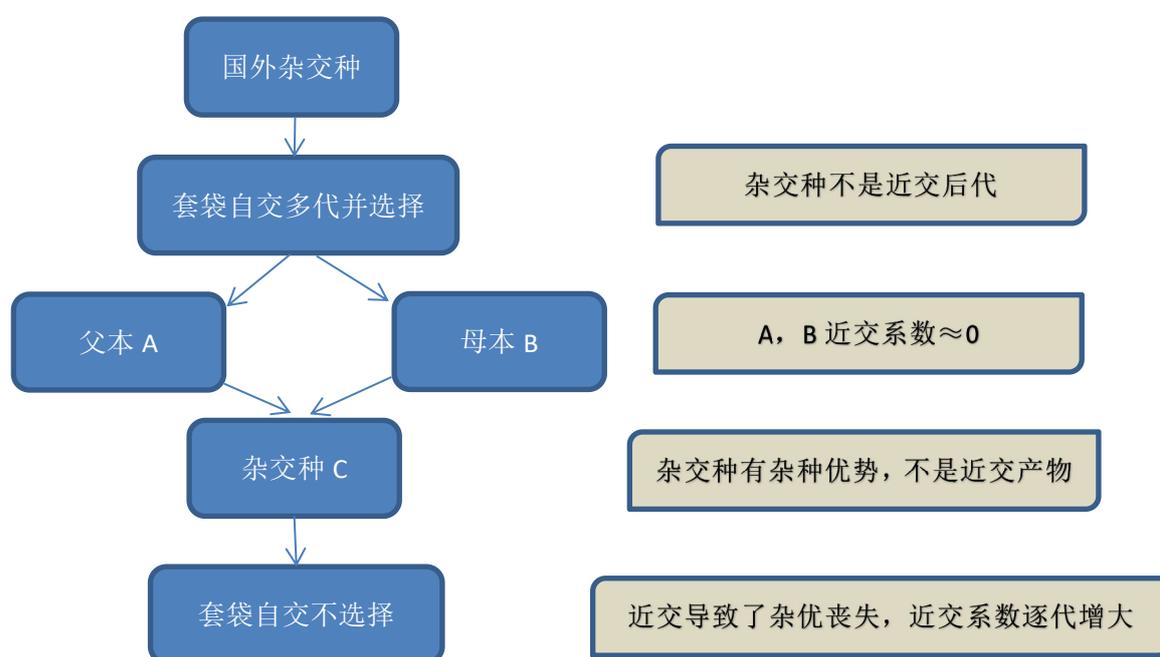
植物育种中，一会会出现“遗传力”一会又变身“遗传率”，反映了对这一概念内涵把握的模糊。

实际育种工作中，在单因素品种比较试验中，计算考察试验的遗传力，其育种价值是什么？遗传力是这样使用的吗？

7.2 自交和近交的概念困扰

植物多自交，动物是异交，自交 \neq 近交，所以泛泛的说：“一个纯合自交系，近交系数=1”，这是有问题的。近交系数是动物育种领域经常使用的统计量，其计算应该基于明确的基础群，而基础群内的个体其近交系数一律设为0。实践中，基础群的个体一般是指那些处于系谱之初的个体，其祖先或无从查考。

近交导致杂种优势丧失。套二环系，采用向两个方向推开的选择策略，就是为了获得后面的杂种优势，推到最后，父本母本分别纯合了，进行杂交，后代获得了明显的杂种优势。不能说此时父母本的近交系数=1，而应该是更趋近于0。这才符合育种实践。



再举个理想化的例子，一个育种者从高山之巅采到了一个纯合的绝世材料，这个材料与世隔绝，从未与同类材料杂交过即不存在近交，人世间就不存在这个材料的任何亲戚，没有任何亲缘关系，无论怎么考虑，怎么计算，这个材料的近交系数都应该=0。不能因为材料是纯合的就说它近交系数=1。

“一个纯合自交系，近交系数=1”，要考虑这句话的适用范围。

类似的另一句话是“自交是近亲交配的极端情况”，在动物育种领域，这句话没有问题，在植物领域，不能这样表述。

7.3 A 矩阵的构建困扰

动物构建加性遗传相关 A 矩阵，假定基础群动物间没有亲缘关系，因此，近交系数 $f_x=0$ ，

一个个体与它本身的加性遗传相关为： $A_{xx} = 1 + f_x$, $A_{00} = 1 + 0 = 1$

所以动物 A 矩阵对角线第一个元素 $a[0,0]$ 一定是 1，到了植物这里，如果认为纯合自交系近交系数 $f_x=1$ ，那这个 $a[0,0]=1+1=2$ 了，这就引发了一系列的问题，我曾经按照 $f_x=1$ 的情况，推导出了“植物 A 矩阵构建的规则”，结果很有意思，构建出的 A 矩阵，除以 2，恰好就等于 Henderson

的 A 矩阵，自己还一度以为，完成了多么了不起的工作。但是后来，还是想清楚了，这是一个弯路。在 A 矩阵的构建上，不存在动植物区别。A 矩阵是个遗传相关矩阵，A 矩阵对角线，是自己跟自己的相关，不就是应该等于 1 吗。非要认为=2，就会被困扰住。

动物是异交，甚至必须异交才能繁殖后代，种族才能维持生存，很多植物靠自交就能繁衍，因此动物比植物更容易发生近交。小麦的自然杂交率一般小于百分之五。小麦自然杂交率在 1% 以下，开花授粉时遇高温干旱会适当提高。杂交率都这么低，何来的 100% 的近交系数， $f_x=1$ 一定是错误的。

Henderson 提到的 a "base" population of animals regarded as noninbred and unrelated，就是考虑动物存在近交可能，通过假定进行了排除。但是并没有排除存在自交的植物使用 BLUP 分析方法。

The animals for which A^{-1} is to be computed are identified by 1, 2, ..., n. They must be ordered so that parents precede their progeny, and 1, ..., b must be a "base" population of animals regarded as noninbred and unrelated. Then the upper left b^2 submatrix of A is I. From that starting point, the upper left submatrix is expanded by order one in succession until the n^2 matrix is formed.

想明白了 A 矩阵“基础群”中植物纯合自交系，近交系数 $f_x=0$ ，而不是=1，就解决了动物育种 BLUP 理论方法在植物领域应用的核心问题。

7.4 植物 BLUP 用共祖先系数矩阵，动物 BLUP 用加性遗传相关矩阵，为什么有这样的差异？两者又是一回事是什么意思？

Rex.Bernardo 在 1994-1996 年发表了多篇文章，如：

Prediction of Maize Single-Cross Performance Using RFLPs and Information from Related Hybrids, Best Linear Unbiased Prediction of the Performance of Crosses between Untested Maize Inbreds
将 Henderson 的 BLUP MME 理论及分析方法应用于植物育种。Bernardo 没有采用动物模型，而是采用了

$$y=X\beta+Z_1a_x+Z_2a_y+Zd+e$$

a_x :母本一般配合力 GCA，对应向量 g_1 ， $V(g_1) = G_1V_{GCA(母)}$

a_y :父本一般配合力 GCA，对应向量 g_2 ， $V(g_2) = G_2V_{GCA(父)}$

d :父母本间特殊配合力 SCA，对应向量 s

上述模型 BLUP MME 构建， G_1, G_2 矩阵采用了共祖先系数矩阵。

模型中的 G_1, G_2 矩阵为协方差矩阵，矩阵对角线元素是自己和自己的相关。从数学角度理解，自己和自己的相关应为 1。如果按照 Henderson 的 A 矩阵构建规则，采用自交系的近交系数=1，一个个体与它本身的加性遗传相关为： $A_{xx} = 1 + f_x = 1 + 1 = 2$ ，那么构建出来的相关矩阵，左上角 [0, 0] 元素值为 2。这就存在明显的问题。采用共祖先系数矩阵，构建出的相关矩阵，左上角 [0, 0] 元素值为 1，就符合了 BLUP MME 的计算要求。因为两个个体的共祖先系数等于二者交配所生后代的近交系数，就等同于将矩阵每个元素-1, $2-1=1$ 。

王建康老师的《应用数量遗传》（第二版）一书，详细介绍了 BLUP 理论方法及其在植物育种中的应用。Henderson 的 BLUP MME 理论方法，在应用到植物育种的时候，唯独将加性遗传相关矩阵 A 替换为了共祖先系数矩阵。

植物育种中，更符合育种实际情况的是，自交系的近交系数=0。 $A_{xx} = 1 + f_x = 1 + 0 = 1$ 。

在自交系的近交系数=1 的设定下，计算得到的共祖先系数矩阵，和自交系的近交系数=0 的设定下，按照 $A_{xx} = 1 + f_x$ 计算的遗传相关矩阵是一致的。

看 $A_{xx} = 1 + f_x$ 这个公式，当用共祖先系数计算的时候，其实仅是计算了 f_x ，而设定自交系的近交系数=1，就把扔掉的 1 又捡回来了。

因此，给出的简单结论是，在设定自交系的近交系数=1 的情况下，得到的共祖先系数矩阵就是自交系的近交系数=0 情况下的遗传相关矩阵。两者是一回事。但是为了学习和理解的正确性，应该逐步纠正植物育种 BLUP MME 中出现的共祖先系数矩阵现象。正确的理解是，植物育种 BLUP MME 分析中，自交系的近交系数=0，上述模型的 G_1, G_2 矩阵分别为母本自交系间的遗传相关矩阵和父本自交系间的遗传相关矩阵。

Rex.Bernardo 这个模型在 1994 年就出现了，实际育种中，父本群母本群间配制组合，往往只能部分实行，未实施的组合可以利用 BLUP 分析技术进行预测，这对植物育种有很大作用。

7.5 BLUP 要用到材料系谱，植物系谱不记录自交情况下，适合采用吗？

适合采用。A 矩阵是依靠系谱构建的，系谱不完善不准确一定会对计算结果产生影响，这是事实。植物的系谱，仅是记录了亲本情况，有的年份没有杂交一直是套袋自交，群体中且有选择淘汰，这些都会对育种产生影响，对计算结果自然也会有影响。但是，想想动物育种，类似的情况其实也会存在，并没有阻碍动物育种采用 BLUP。通常认为亲代传递 1/2 育种值给子代，植物自交环节，并没有引入外源基因，对育种值的传递影响还不好说，极端的情况，材料系谱不明，BLUP 分析中，假定该材料为基础群材料，也没有阻碍 BLUP 应用。另外，分子育种越来越走向实际育种，全基因组育种值估计的一步法（Single-step blup）就是结合了传统系谱和分子亲缘关系来构建的加性遗传相关 H 矩阵，可以理解为用准确的分子亲缘关系来修正校正系谱，来估计育种值，所以没有问题。

7.6 BLUP 本身真的很难，配合力又很简单

配合力半个小时，外行也能看懂，BLUP 就没有这样简单了。动物育种本科生就系统学习 BLUP 了。个人认为，BLUP 是动物育种课程的核心内容。

植物育种学，真的需要增加 BLUP 理论和分析方法了。不能再继续人为以为，植物育种不同于动物育种，植物育种用配合力就很了不起了。这真的不行。植物育种学，植物数量遗传，应该多介绍 BLUP 理论和分析方法，让后来的育种新生力量了解熟悉 BLUP。

差值选择法>配合力分析法>BLUP 育种值分析方法。



张沅老师的《家畜育种学》，张勤老师的《动物遗传育种中的计算方法》，张沅老师、张勤老师的《畜禽育种中的线性模型》，王金玉 陈国宏老师的《数量遗传与动物育种》，于向春 张豪老师主译的《线性模型在动物育种值预测中的应用》等一系列书籍，都值得有志于了解 BLUP 育种值理论和方法的植物育种者学习阅读。

十一、品种稳定性分析

品种稳定性分析是服务于品种推广的一项数据分析工作，一个品种稳定性好，推广起来才放心，一个品种，或者是纯系品种或者是杂交种，在保证没有杂株，基因纯合度 100%的情况下，这个品种应该是“稳定”的，因为，基因始终不变了。

在不同地点种植的情况下，收获有差异的产量结果，是环境的因素造成的结果。这时，我们说，这个品种可能“不稳定”。而一个将要大面积推广的品种，基因的稳定性是必备的条件之一，因此，我们所说的品种稳定性分析，应该针对该品种种植于不同地区的产量差异情况。

一个品种在某地块种植，并设置有 3 次重复，收获了 3 个产量结果，且有不同，我们可以推断引起产量差异的原因，是由于 3 个地块的条件差异导致的产量差异，如：土壤条件差异，栽培措施差异，田间环境差异，而由于是一个地点，我们通常会倾向于认为，气候条件具有影响一致性，从而忽略气候因素对产量结果的影响。

数量遗传学会给出分析模型

$$P(\text{表型值}) = G(\text{基因型值}) + E(\text{环境效应})$$

或者进一步给出这样的模型

$$P(\text{表型值}) = G(\text{基因型值}) + E(\text{环境效应}) + g \times e(\text{交互作用})$$

如果将同一品种同时在两个不同积温带布点进行产量试验，积温高的地点品种产量，常常高出积温低的地点的品种产量，这种农业生产现象普遍存在，而从查阅的文献资料来看，对这样现象的解释，数量遗传学要不笼统的归为环境效应，要不归为基因 x 环境交互作用，或者两者皆而有之，甚至是归为了误差。

如果我们仅在高积温带安排了重复的产量试验，那高积温带比低积温带高出的这部分产量，按照数量遗传学的模型直接就划为了“基因型值”。加之，产量数据需要集中汇总，然后再进行品种稳定性分析，如果不考虑积温等因素的效应影响，可能就得出这个品种不稳定的结论。因为，换个地方，产量有明显差异，在一个地方种的还行，换个地方产量就低了，感觉这个品种“不稳定”。

下面是一些典型的品种稳定性分析方法，我们重点从数据分析模型来进行介绍和比较。

(一) 品种稳定性分析模型

分析模型很多，下面是几种常见常用的。

Eberhart 和 Russell 模型 (1966 年)

$$y_{ij} = \mu_i + \beta_i I_j + \delta_{ij}$$

采用回归系数和校正的离回归方差作为测定稳定性的两个参数，其中， y_{ij} 表示第 i 个品种在第 j 个环境中的平均产量； μ_i 表示第 i 个品种在所有环境中的平均产量； β_i 表示第 i 个品种对各种变化环境反应的回归系数； δ_{ij} 表示第 i 个品种在第 j 个环境的回归离差； I_j 表示第 j 个环境的环境指数。

George C. C. Tai 模型 (1971 年)

$$y_{ijk} = \mu + g_i + l_j + (gl)_{ij} + b_{jk} + e_{ijk}$$

其中， y_{ijk} 表示第 i 个品种在第 j 个环境、第 k 次重复的平均产量； μ 表示所有品种在所有环境、所有重复中总平均产量； g_i 表示第 i 个品种的遗传效应； l_j 表示第 j 个环境的环境效应； $(gl)_{ij}$ 表示第 i 个品种与第 j 个环境的交互作用效应； b_{jk} 表示第 j 个环境内第 k 次重复的效应； e_{ijk} 表示第 i 个品种在第 j 个环境、第 k 次重复的随机误差。

Shukla 模型 (1972 年)

$$y_{ijk} = \mu + d_i + \varepsilon_j + g_{ij} + e_{ijk}$$

其中， y_{ijk} 表示第 i 个品种在第 j 个环境、第 k 次重复的平均产量； μ 表示所有 y_{ijk} 总平均产量； d_i 表示第 i 个品种的遗传效应； ε_j 表示第 j 个环境的环境效应； g_{ij} 表示第 i 个品种与第 j 个环境的交互作用效应； e_{ijk} 表示第 i 个品种在第 j 个环境、第 k 次重复的试验误差。

由于是多个品种在多个地点的试验数据，因此还有一种分析思路是联合回归分析，由 Yates 和 Cockerham(1938 年)提出，当时并没有多少应用。1968 年 Perkins 和 Jinks 又重新提出了这种分析方法，并解决了之前方法存在的问题，使这一方法得到了完善。

AMMI 模型(additive main effects and multiplicative interaction)

(Crossaer al.,1990;Zobel et al., 1988)

$$y_{ge} = \mu + \alpha_g + \beta_e + \sum \lambda_n \gamma_{gn} \delta_{gn} + \theta_{ge}$$

其中， y_{ge} 是在环境 e 中基因型 g 的产量； μ 是总体平均值； α_g 是基因型平均偏差； β_e 是环境的平均偏差； λ_n 是第 n 个主成分分析的特征值； δ_{gn} 是第 n 个主成分的基因型主成分得分； γ_{gn} 是第 n 个主成分的环境主成分得分； n 是模型主成分分析中主成分因子轴的总个数； θ_{ge} 为残差。

GGE 叠图法

区域试验中产量一般分解为以下几个组分：

$$y_{ij} = \mu + \alpha_i + \beta_j + \phi_{ij}$$

式中 μ 为总平均值， α_i 为品种 i 的效应， β_j 为环境 j 的效应， ϕ_{ij} 为品种 i 与环境 j 的相互作用式。将原先产量数据减去各地点的平均产量，由此形成的数据集只含品种主效应 G 和品种-环境互作效应 GE ，合成 GGE 。对 GGE 做单值分解，并以第一和第二主成分近似之。按照第一和第二主成分值将各品种和各地点放到一个平面图上即形成 GGE 叠图。

方法还有很多，不多举了。

(二) 模型分析

国内学者做了不少品种稳定性分析方法比较的研究，并给出了结论和建议：杨涛等^[8]认为“AMMI 模型是一种较为理想的品种稳定性评价方法”；穆培源等^[9]认为，“GGE 叠图是分析品种区试资料的理想方法”；胡希远等^[7]认为“这些模型在数据拟合效果、品种稳定性评判和品种均值差异检验等方面存在明显的差异；没有一个模型对所有试验资料的拟合效果均最佳，而是与具体分析的试验数据有关。不同稳定性分析模型适用于不同的试验数据，模型选择对品种稳定性评价和品种效应差异显著性检验有较大的影响”。

AMMI 模型将方差分析和主成分分析结合在一起，应用于品种稳定性分析，有的教科书将这一模型称为基因型和环境互作的乘积模型^[3]。主成分分析是用主成分代替所有成分分析，方法本身有利有弊，“另外从实际使用效果来看，在有些试验中，第一主成分只可以解释 10-30% 的互作变异，得到的分析结果不理想；AMMI 模型预测结果依赖于互作变异中可重复部分的大小，以及可重复变异的可重复性^[3]”。

从上面列出的几种常用品种稳定性数据分析方法来看，品种的产量差异多用品种和环境的互作来对应，在加性效应不能很好满足结果分析的情况下，就引入非线性效应，目的是用合适的数学方法，解决品种稳定性分析的问题，比如方差分析，回归分析，联合回归分析，主成分分析。纯数学方法数据分析运用痕迹明显，模型生物学意义不明确。

品种稳定性分析，要考虑国内试验数据的不规范性，试验数据甚至不符合统计分析的要

求，而不是简单套用某种稳定性分析方法，是试验数据决定数据分析模型，而不是反过来，否则，即便采用再多漂亮的分析模型，很可能也是毫无意义。

品种稳定性分析，数据分析模型是关键。一个合乎试验本身的数学分析模型，才是数据分析合理性的基础。品种稳定性分析试验，异地布点，多积温带布点，环境差异明显，这是试验本身的特性，因此数据分析模型应该体现出这一特性。品种推广往往首先着眼于地区而不是地点，而地区跟积温带往往是重合的，如果试验中存在明显积温差别，就应该考虑试验的积温效应，其它光效应、土效应、水效应等同理，因此，有必要给出更合理的分析模型。

（三）品种稳定性分析新模型

以玉米为例，农业部国家玉米品种区域试验划分了：极早熟春玉米组，东北早熟春玉米组，东北中熟春玉米组，东华北春玉米组，黄淮海夏玉米组，西北春玉米组，西南春玉米组，东南春玉米组 8 个生态区。

如果试验仅布置在一个生态区，是最简单的事情了，积温因素就基本可以忽略了。试验数据分析直接用单因素随机完全区组分析结果就可以了。

如果试验布置在多个生态区，积温因素就应该作为一个效应单独考虑。栽培措施，如水、肥尽量控制一致，就不用单独考虑。

给出模型：

$$y_{ij} = \mu + \alpha_i + \beta_j + \gamma_k + \sigma$$

μ 为品种平均值， α_i 年度 i 的效应， β_j 为地点 j 的效应， γ_k 为积温 k 的效应。每一个产量观测值的构成，可以剖分为，产量均值、年度效应、地点效应、积温效应、剩余误差。积温出现在模型中，是这个品种稳定性数据分析模型的特点，而这个积温效应，有时被数量遗传学看做品种和环境间的神秘互作，而这样的互作，本身就透露着“不稳定”的色彩，不确定性明显。

但是从品种本身来说，种植在一个日光照充足的地点，品种本身产量就会比寡照地点产量升高，农业工作者，在田间看着禾苗生长的时候，又有谁会怀疑这一点呢？为什么，在室内分析数据的时候，就忽略了这一事实，而宁愿将积温效应归结为品种和环境间的互作，线性互作甚至表现为非线性的互作呢？

一个品种在积温高的地点生长，产量就高，或者将其定义为品种与环境间的某种互作效应，或者将其理解为品种本身的基因效应，如果非要做一个取舍，应该是品种本身的基因效应，因为有的品种即便种在积温高的地点，产量几乎不变。

（四）Shukla 方差计算示例

国内品种区域试验中有在使用 Shukla 模型稳定性分析方法，需要计算 Shukla 方差，让我们先看一组试验数据。

Shukla 模型稳定性分析			
地点	材料	区组 1	区组 2
1	材料 1	750	772
1	材料 2	880	858
1	材料 3	770	544
1	材料 4	685	707
1	材料 5	673	533
2	材料 1	950	974
2	材料 2	862	892

2	材料 3	896	854
2	材料 4	950	914
2	材料 5	933	907
3	材料 1	924	900
3	材料 2	1023	1017
3	材料 3	934	914
3	材料 4	997	969
3	材料 5	760	816
4	材料 1	930	980
4	材料 2	890	938
4	材料 3	890	920
4	材料 4	975	949
4	材料 5	892	870
5	材料 1	759	799
5	材料 2	739	699
5	材料 3	625	683
5	材料 4	800	858
5	材料 5	660	682

每个材料平均数：873.8, 879.8, 803, 880.4, 772.6

每个地点平均数：717.2, 913.2, 925.4, 923.4, 730.4

地点+材料平均数：

(761, 869, 657, 696, 603)

(962, 877, 875, 932, 920)

(912, 1020, 924, 983, 788)

(955, 914, 905, 962, 881)

(779, 719, 654, 829, 671)

Shukla 方差计算公式：

$$\hat{\sigma}_i^2 = \frac{1}{(l-1)(v-1)(v-2)} \left[v(v-1) \sum_j (\bar{y}_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}_{..})^2 - \sum_i \sum_j (\bar{y}_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}_{..})^2 \right]$$

i 对应品种, j 对应地点, l=5, v=5

$$\sum_i \sum_j (\bar{y}_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}_{..})^2 = 53987.36$$

Shukla 方差：

材料 1 = 24.55

材料 2 = 9853.05

材料 3 = 411.4833

材料 4 = 2179.017

材料 5 = 4402.95

材料	材料 1	材料 2	材料 3	材料 4	材料 5	
方差	8661.51	10697.29	19582.67	12920.93	17532.49	
	1	2	5	3	4	排序

Shukla 方差	24.55	9853.05	411.4833	2179.017	4402.95	
	1	5	2	3	4	排序

Shukla 方差用材料在同一地点内部的均值来进行方差的计算，其本质反应的还是试验本身的方差，可以理解为加工后的方差，但是从本试验数据分析结果可以看到，材料间的方差排序和 Shukla 方差排序却不一致了。稳定性分析的结果比较离不开材料间离散程度的比较，而这种比较，由于同属于一个试验，具有材料间离散程度比较的可行性。生物统计中，一个试验的多重比较是立足均值进行的，试验中处理间（材料）离散程度的比较，却没有涉及，而品种稳定性分析，其实质就是一个试验中处理（材料）间离散程度的比较。相对平均偏差（RSD），是一个关于样本离散度的统计量。

（五）Eberhart-Russell 模型、Shukla 模型、相对平均偏差稳定性分析比较

基于同一组试验数据，比较 Eberhart-Russell 模型、Shukla 模型、相对平均偏差（RSD）计算分析结果，为实际育种工作，寻找简单、易用、好用的稳定性分析工具方法。

模拟数据

稳定性分析			
地点	材料	区组 1	区组 2
1	材料 1	750	772
1	材料 2	880	858
1	材料 3	770	544
1	材料 4	685	707
1	材料 5	673	533
2	材料 1	950	974
2	材料 2	862	892
2	材料 3	896	854
2	材料 4	950	914
2	材料 5	933	907
3	材料 1	924	900
3	材料 2	1023	1017
3	材料 3	934	914
3	材料 4	997	969
3	材料 5	760	816
4	材料 1	930	980
4	材料 2	890	938
4	材料 3	890	920
4	材料 4	975	949
4	材料 5	892	870
5	材料 1	759	799
5	材料 2	739	699
5	材料 3	625	683
5	材料 4	800	858
5	材料 5	660	682

材料	材料 1	材料 2	材料 3	材料 4	材料 5	
方差	8661.51	10697.29	19582.67	12920.93	17532.49	

	1	2	5	3	4	排序
Shukla 方差	24.55	9853.05	411.4833	2179.017	4402.95	
	1	5	2	3	4	排序
Eberhart-Russell 模型回归系数	0.872581	0.719150	1.25041	1.02018	1.13768	
	2	1	5	3	4	排序
相对平均偏差 RSD	0.0950332	0.0821096	0.1469489	0.1071331	0.1436707	
	2	1	5	3	4	排序

为了看的更清楚，只保留排序结果的对照表

材料	材料 1	材料 2	材料 3	材料 4	材料 5	
方差	1	2	5	3	4	排序
Shukla 方差	1	5	2	3	4	排序
Eberhart-Russell 模型回归系数	2	1	5	3	4	排序
相对平均偏差 RSD	2	1	5	3	4	排序

相对平均偏差 RSD 和 Eberhart-Russell 模型回归系数的排序一致。Shukla 方差稳定性排序有出入。

结论：评价一个试验中参试品种的稳定性，用相对平均偏差 RSD。

（六）稳定性选择指标（相对平均偏差 RSD）

离差，是观测值偏离均值的差值，存在正负。方差是平均的离差平方和，去除了离差的正负，也放大了离差，放大了数据的离散性。标准差是方差的平方根值，试图把方差又还原回离差的尺度，数据准确性不如原始离差。

由于材料均值的差异，使材料间的方差差异不能准确体现材料的离散程度（稳定性），比如，两个材料，方差相同，稳定性一定相同吗？两个材料，方差相同，一个均值大一些，一个均值小一些，哪一个材料离散程度小，稳定性更好？10（8），表示均值是 10，方差是 8。

$$10(8), 4(8)$$

10(8) 的稳定性优于 4(8)

所以在进行一个试验中两个材料间离散程度比较的时候，不仅需要考虑离差或者方差，还要考虑均值。

统计学中，平均偏差、相对平均偏差的定义如下：

$$\text{平均偏差 adev} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

相对平均偏差（RSD, relative standard deviation）：平均偏差与算术平均值的比值。

$$\text{相对平均偏差 RSD} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n\bar{x}}$$

相对平均偏差 RSD，适合用于一个试验中材料间的离散程度比较。

在样本数据中存在大离散的情况下，大离散，其为方差做出的贡献明显比离散多，会导致用方差衡量材料间的离散程度失真，离散衡量数据离散程度会比方差可靠。Shukla 方差采用类似方差计算的套路，跟方差的排名结果通常会保持一致。

对样本群体，使用平均偏差（adev），会比方差更有效比较样本间的离散程度，但需要将样本均值考虑进来，这样可以更进一步提高分辨准确性。品种稳定性分析，需要找到合适的比较材料间离散程度的统计量，相对平均偏差（RSD）就非常适合。

一个试验中两个试验安排相同的材料离散程度的比较有以下情况：

1. 两个材料均值一样，离差绝对值累加和大的材料，离散程度大。
如：5（3），5（6）； $3/5 < 6/5$
2. 两个材料离差绝对值累加和一样，均值小的离散程度大。
如：5（8），4（8）； $8/5 < 8/4$
3. 两个材料离差绝对值累加和都大于均值。
如：5（7），3（4）； $7/5 > 4/3$
4. 两个材料离差绝对值累加和都小于均值。
如：5（3），3（2）； $3/5 < 2/3$
5. 两个材料，一个离差绝对值累加和大于均值，一个离差绝对值累加和小于均值，离差绝对值累加和大于均值的离散程度自然更大。
如：5（7），3（2）； $7/5 > 2/3$

以上 5 种情况，相对平均偏差都可以定量比较出两个材料的离散程度大小。如果仅用方差、标准差、离差绝对值累加和、极差、平均偏差都不能进行准确比较，CV 源自方差，也会跟方差保持类似的排名趋势。

相对平均偏差越小越好，越小说明数据的离散度越低，品种越稳定。

相对平均偏差利用的是试验数据的离差，计算简便。

相对平均偏差，由于除了观测值数量，因而是材料观测值的单个表现或者平均表现，所以可以用于观测值数量相同或者不同的材料间的比较，这就拓展了应用范围。

博思公司开发了在线 RSD 计算程序，感兴趣种业朋友免费使用。

<http://www.nbs.net.cn/rsd.htm>

（七）稳定性分析策略及建议

一个品种种植在不同的积温带，产量表现高低不同，甚至差异很大，以前我们说，这个品种稳定性差，但可能是该品种在有的积温带生长表现更优异而已，更适合在一些积温带推广。

设想一个理想化的试验，每一个参试品种纯度 100%，无一棵杂株，每一个品种内部基因型 100%一致，种植在同一个积温带的不同地点，农艺措施 100%一致，产量数据依旧会有差异，引发因素包括：土壤条件，小环境，小气候。因为排除了基因因素，人为因素。即便种植在同一个积温带的同一地点，产量数据同样会有差异。这样一个试验，稳定性分析可以用相对平均偏差（RSD）指标，衡量同一品种发生的偏差。而上面这个试验，现在如果安排在不同的积温带另做试验，将试验中的积温效应单独分离考虑后，剩下的部分，与前一个试验没有什么本质

不同。所以，品种稳定性分析，不应该是一个复杂的问题。

实际生产中，如果参试单位将混有杂株的品种，用于区试布点；赶时间，将尚未完全纯合的品种用于区试；实际农艺措施不可能不存在差异；制种混杂；加工精选混杂；包衣包装混杂；运输混杂；都将品种稳定性问题复杂化了，而这就不是数据分析所能解决的问题了。

品种稳定性分析无论采用何种分析模型，应以均值为第一考虑因素，然后再考虑数据离散性。稳定性分析可以进一步区分为稳定性和适宜性问题。

品种稳定性分析中，相对平均偏差（RSD，Relative Standard Deviation）可以度量稳定性试验中材料间的数据离散程度，比方差、Shukla 方差度量分辨效果更好。用方差不如用 CV，用 CV 不如用相对平均偏差 RSD。

品种稳定性分析建议采用相对平均偏差，有以下一些考虑：

1、计算简单，基于离差绝对值计算，先求离差和，再求平均，再除以平均数。

$$\text{相对平均偏差 RSD} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n\bar{x}}$$

2、可供使用的稳定性分析方法模型，很多很多，俞世蓉在《作物的品种适应性和产量稳定性》《作物杂志》1991 一文中下面一段介绍：“自 Yates 和 Cochran(1938)首次提出用回归分析法研究基因型 X 环境的互作效应以来，至今 50 多年间，还没有找到一个很好的统计模型能把适应性参数和稳定性参数分解清楚，所以现在都笼统称为稳定性参数。迄今为止，不少工作者已提出了几十种的稳定性或适应性的参数估计方法”。

品种稳定性分析方法真的很多，但问题好像还是没有得到很好的解决。使用相对平均偏差 RSD 进行品种稳定性分析也是博思公司针对稳定性分析提出的一种解决办法。但是相对平均偏差 RSD 不是博思公司提出的，这是一个基本的统计量，适用于样本间离散程度比较。而稳定性分析的本质就是一种比较。同一试验中的品种稳定性分析，就是样本间离散程度比较。方差校正、方差变换、回归、主成分分析，都试图解决稳定性分析问题，效果不好带有必然性。

3、品种区域试验，难免缺区，相对平均偏差 RSD 的计算应用，不受缺区影响，缺区只要调整观测值个数 n 即可；而常用的稳定性分析方法，缺区会带来计算分析困难，用缺区估计的方法又会影响结果的准确性，建立在方差基础上的稳定性分析方法，本就不如离差衡量稳定性精确，所以实际应用效果不会很理想。

4、建立在方差基础上的稳定性分析方法，多设有各种模型假设，而这些假设，实际不少难以成立，这也导致了实际应用效果不理想。

5、基于回归思想建立的稳定性分析方法，在品种稳定性分析中，本身存在自变量依变量不独立的问题。主成份分析应用于稳定性分析又是考虑其中的主成分部分，实际应用效果不会理想。

6、品种适宜性

俞世蓉在《作物的品种适应性和产量稳定性》《作物杂志》1991 中提到了品种稳定性分析包含了稳定性参数和适应性参数，这是品种区域试验要解决的两个问题。

Eberhart-Russell 模型，提出了环境指数。

环境指数 = 该环境中所有品种的均值-试验总平均

环境指数（高），表明该环境对品种表现有利，因为该环境种植品种产量均值（高）

环境指数（低），表明该环境对品种表现不利，因为该环境种植品种产量均值（低）

环境指数，可以作为品种适宜推广地区参考依据，主要体现的是（地点对品种）的影响。

环境指数

71	环境地点	环境均值	环境指数(排名)							
72	地点3	925.400	83.4800	1						
73	地点4	923.400	81.4800	2						
74	地点2	913.200	71.2800	3						
75	地点5	730.400	-111.520	4						
76	地点1	717.200	-124.720	5						

地点3的环境指数高，说明，试验中的品种在地点3种植，普遍表现为高产，所以，品种1-5，如果种植在地点3，有望高产。

环境指数，计算的是试验中的地点效应，是地点均值跟总平均的离差。因此，Eberhart-Russell 模型中的环境指数，可以脱离 Eberhart-Russell 模型，单独拿出来，应用于品种适宜性分析，解决的是品种区域试验中的品种推广适应性问题，这依旧是方差分析的效应套路。Eberhart-Russell 模型稳定性分析，回归系数用于稳定性评价，对应 RSD，环境指数用于品种适应性评价。

7、品种区域试验，通常数据多，地点多，多年试验数据就更多了，因此方差同质性检验通常为异质，F 测验经常会显著，Eberhart-Russell 模型在分析的时候，建立了4个不同的方差分析表，可谓极其复杂。但是，结合育种实际应用，品种稳定性参数来自回归系数 b，回归方程中，自变量 x 是各地点的产量均值，依变量 y 是品种在各地点的产量均值，可以直接获得回归系数 b；环境指数，更简单，是试验中的地点效应，是地点均值跟总平均的离差，这两个关键参数，跟方差分析表，跟 F 测验没有任何关系。各种稳定性分析方法，建立在方差基础上，依据各自的模型，进行各种方差的分解，然后构建出可以进行 F 测验的 F 值，这基本是个套路。但是，从实际育种、实际应用来看，F 测验无论是否显著，联合方差分析无论是否同质，一个试验中的均值排序最重要，试验中稳定性排序最重要。

结论：品种稳定性分析，用 RSD 解决稳定性问题，用环境指数或者稳定性分析模型（ $Y = \mu + \text{年度效应} + \text{地点效应} + \text{积温效应} + \sigma$ ），解决适宜性问题。

参考文献

- [1] 盖钧镒 试验统计方法 北京：中国农业出版社，2003.2
- [2] 唐启义 DPS 数据处理系统 北京：科学出版社，2013
- [3] 王金玉 陈国宏 主编 数量遗传与动物育种 南京：东南大学出版社，2004.8
- [4] 翟虎渠 王建康 应用数量遗传 北京：中国农业科学技术出版社，2007.3
- [5] 张勤 动物遗传育种中的计算方法 北京：科学出版社，2007
- [6] 翟虎渠 应用数量遗传 北京：中国农业出版社，2001.9
- [7] 刘继麟 主编 玉米育种学 北京：中国农业出版社，2001.8

- [8] 朱军 遗传学（第三版） 北京：中国农业出版社，2003.6
- [9] 孔繁玲 植物数量遗传学 北京：中国农业大学出版社，2007.1
- [10] 张沅 家畜育种学 北京：中国农业出版社，2006.1
- [11] 王松桂等 线性模型引论 北京：科学出版社，2014.4
- [12] 王松桂等 线性统计模型（线性回归与方差分析） 北京：高等教育出版社，2010.2
- [13] 胡希远等，作物品种稳定性分析不同模型比较. 麦类作物学报，2009, 29(1):110-117
- [14] 杨涛等，三种评价品种稳定性方法的比较. 贵州农业科学，2006, 34(1):28-31
- [15] 穆培源等，作物品种稳定性分析方法的研究进展. 新疆农业科学，2003, 40(3):142-144
- [16] 刘大群 王恒立 品种稳定性评价方法的比较和分析. 作物学报，1988,14(4):290-295
- [17] 严威凯等 GGE 叠图法—分析品种 X 环境互作模式的理想方法. 作物学报，2001,27(1):21-28
- [18] 俞世蓉 作物的品种适应性和产量稳定性《作物杂志》，1991
- [19] 胡秉民 耿旭 作物稳定性分析法 北京科学出版社，1993
- [20] Kwanchai A.Gomez Arturo A.Gomez Statistical Procedures for Agricultural Research,2nd Edition Philippines: John Wiley & Sons,Inc. 1984
- [21] Arnel R. Hallauer Marcelo J.Carena J.b.Miranda Filho Quantitative Genetics in Maize Breeding, Springer New York Dordrecht Heidelberg London,2010

十二、总结

总结是想写一些对生物统计和数量遗传的总体印象，对自己是个知识梳理，对学习和应用生统的朋友也许还有些帮助。

1、生物统计和数量遗传，在育种应用中，离差是更有实用价值的工具，方差不是，因为离差有方向性，这种方向性是育种选择的基础。这从育种实践已经证明了。

2、F 测验显著，其实真没有什么太大的用处，对育种来说，如果只告诉我有效应存在，而且达到显著水平，能有什么帮助呢，育种者往往面对的是，微小的有利基因累积实现了没有，这样的累积往往是达不到显著水平的，不管显著不显著，都要做出选择。

3、数据分析有很多方法，生物统计和数量遗传的实际运用要把握住为生产和科研服务的方向，不见的越新的方法就越好，模型是为育种需要服务的，数据分析也是为育种需要服务的，试验设计方法也是，所以在育种实战中，要挑选适合的试验方法，适合的数据模型，适合的数据分析方法。

4、对育种来说，离差（效应），目前来看，是更适用的工具。

5、数量遗传学有统一的度量基因型值的尺度， $(-a, d, a)$ ，由于是从基因的角度切入，且需要知道基因频率 (p, q) ，这样与育种实际就很难结合到一起，因为实际育种多是不清楚材料的基因情况，基因频率就更不得而知，这样就使数量遗传的很多结论理论被束之高阁，仅有理论意义。配合力是一种育种方法策略，跟数量遗传甚至可以说关系不大。估计育种值在动物育种领域广泛采用，才使数量遗传真正发挥了育种作用，可惜在植物育种领域仍有待育种者认识到其育种价值。植物育种的 BLUP 育种值模型价值，会逐步被育种者发现和理解。

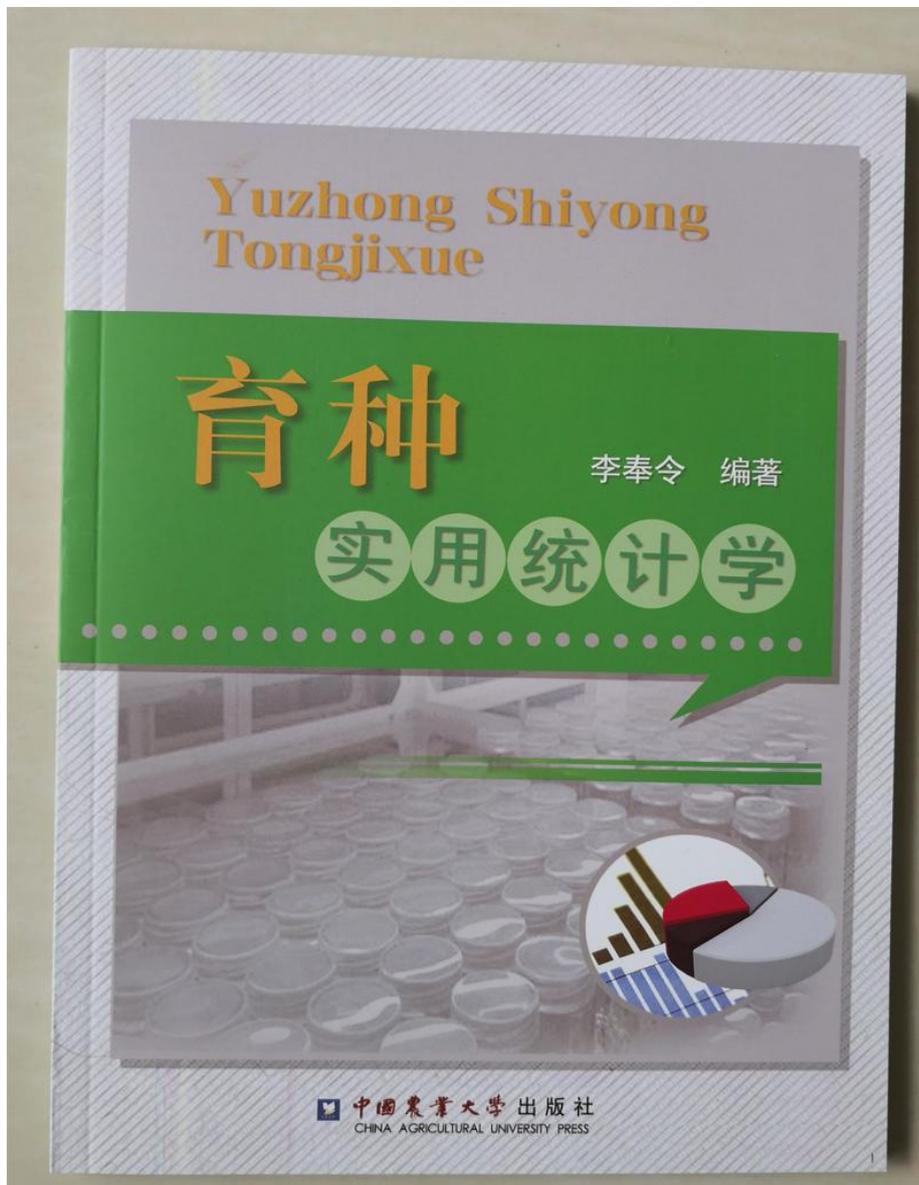
参考文献

- [1]张沅 .2001. 家畜育种学 . 北京: 中国农业出版社
- [2]盖钧镒 .2000. 试验统计方法 . 北京: 中国农业出版社
- [3]于向春张豪主译 .2016. 线性模型在动物育种值预测中的应用 (第3版). 北京: 中国农业出版社
- [4]刘榜 .2007. 家畜育种学 . 北京: 中国农业出版社
- [5]王松桂 史建红 尹素菊 吴密霞 .2004. 线性模型引论 . 北京: 科学出版社
- [6]孔繁玲 .2006. 植物数量遗传学 . 北京: 中国农业大学出版社
- [7]王金玉 陈国宏 .2004. 数量遗传与动物育种 . 南京: 东南大学出版社
- [8]张勤 .2007. 动物遗传育种中的计算方法 . 北京: 科学出版社
- [9]张沅 张勤 .1993. 《畜禽育种中的线性模型》 . 北京: 北京农业大学出版社
- [10]俞世蓉 郭蔼平 .1979.03. 田间试验资料统计分析讲话 . 《天津农业科学》
- [11]王松桂 陈敏 陈立萍 .1999. 线性统计模型: 线性回归与方差分析 . 北京: 高等教育出版社
- [12]朱军 .2000. 线性模型分析原理 . 北京: 科学出版社
- [13]朱军 .2002. 遗传学 . 北京: 中国农业出版社
- [14]盖钧镒 章元明 王健康 .2003. 植物数量性状遗传体系 . 北京: 科学出版社
- [15]唐启义 .2013. DPS 数据处理系统 . 北京: 科学出版社
- [16]刘纪麟 .2001. 玉米育种学 . 北京: 中国农业出版社
- [17]翟虎渠 王健康 .2007. 应用数量遗传 . 北京: 中国农业科学技术出版社
- [18]顾万春 .2004. 统计遗传学 . 北京: 科学出版社
- [19]茆诗松 汤银才 .2012. 贝叶斯统计 . 北京: 中国统计出版社
- [20]翟虎渠 .2001. 应用数量遗传 . 北京: 中国农业出版社
- [21]袁志发 .2011. 群体遗传学、进化与熵 . 北京: 科学出版社
- [22]唐守正 郎奎建 李海奎 .2009. 统计和生物数学模型计算: ForStat 教程 . 北京: 科学出版社
- [23]齐明 .2009. 林木遗传育种中试验统计法新进展 . 北京: 中国林业出版社
- [24]徐礼文 .2011. 线性模型的预测理论及其应用 . 中国水利水电出版社
- [25]莫惠栋.1990. 植物育种中的试验设计: 增广设计. 江苏农学院学报
- [26]莫惠栋.1990. 植物育种中的试验设计: 修饰增广设计. 江苏农学院学报
- [25]莫惠栋.1991. 植物育种中的试验设计: 行列增广设计. 江苏农学院学报
- [27]杨德.2000. 试验设计与分析[M] . 北京: 中国农业出版社
- [28]王健康.2017. 数量遗传学 . 北京: 科学出版社
- [29]A New Class of Resolvable Incomplete Block Designs Author(s): H. D. Patterson and E. R. Williams
Source: Biometrika, Vol. 63, No. 1 (Apr., 1976), pp. 83-92
- [30]Comparing efficiency of RCBD and alpha lattice designs Author(s):Abdisa Gurmesssa Source:LAP
LAMBERT Academic Publishing,2011

买过书的朋友或者收费视频教程的朋友, 学习中有不明白的地方欢迎来问我, 内容不限于你购买的课程。跟统计, 跟育种有关都可以。李奉令



这是正式出版的书，有需要的读者可以联系博思公司。



内 容 简 介

本书从为植物育种服务的角度,梳理了生物统计学、数量遗传学与实际育种应用结合的试验设计和数据分析方法,分析了遗传力错误概念、离差(效应)的育种价值、方差的育种价值、数据缺区应对策略、育种值的育种价值、用 BLUP 育种值提升配合力方法选择效果等内容,给出了植物育种用育种值的育种思路。主要内容包括:试验设计部分,侧重介绍 BIB 设计、 α -格子设计、增广设计;方差分析部分侧重从效应的角度理解方差分析,总结提出了方差分析三句话;配合力部分讨论了配合力方法实质;BLUP 育种值部分提出了植物育种值分析线性混合模型;稳定性分析部分比较了 Eberhart-Russell 模型、Shukla 模型等多种稳定性分析方法,分析了相对平均偏差(RSD)在稳定性分析中的应用价值。

图书在版编目(CIP)数据

育种实用统计学/李奉令编著. —北京:中国农业大学出版社,2018.8
ISBN 978-7-5655-2059-4

I. ①育… II. ①李… III. ①植物育种-生物统计 IV. ①S33

中国版本图书馆 CIP 数据核字(2018)第 169388 号

书 名 育种实用统计学
作 者 李奉令 编著

策划编辑	梁爱荣	责任编辑	韩元凤
封面设计	郑 川		
出版发行	中国农业大学出版社		
社 址	北京市海淀区圆明园西路 2 号	邮政编码	100193
电 话	发行部 010-62818525,8625	读者服务部	010-62732336
	编辑部 010-62732617,2618	出版部	010-62733440
网 址	http://www.caupress.cn	E-mail	cbsszs@cau.edu.cn
经 销	新华书店		
印 刷	涿州市星河印刷有限公司		
版 次	2018 年 8 月第 1 版		2018 年 8 月第 1 次印刷
规 格	787×980 16 开本		7.5 印张 135 千字
定 价	68.00 元		

图书如有质量问题本社发行部负责调换

农博士 BLUP 育种值分析软件

Ver 1.13

The screenshot shows a terminal window titled "农博士分子平台" (Nbs Platform). The interface includes a top navigation bar with "命令" (Command), "设置" (Settings), and "登录" (Login) options. Below the navigation bar, there are two tabs: "BLUP (育种值)" and "BLUP (配合力)". The main terminal area displays a series of commands for GWAS analysis, including file paths like //test.ped and //test.map, and plink commands such as plink --file test, plink --file test --missing, and plink --bfile plink --geno 0.02 --make-bed --out q1. The terminal also shows the status of each command (1, 2, 3) and a progress bar at the bottom. The status bar at the bottom of the window indicates the file path "E:\d\nbs_la\lib\i386-win32\命令.txt" and the version "2023.5.18, Ver 1.13".

功能一、BLUP 分析功能，有两种模式供用户使用：分别是**基于传统系谱的 BLUP 分析**和**基于分子数据的 ssGBLUP 分析（一步法）**。模型会自动跟随数据调整，不需要人工干预，有分子数据就指定数据文件，软件就会按照 ssGBLup 进行分析，没有分子数据就按照传统 BLUP 运行。该功能目前仅在 windows 版本中实现。BLUP 分析功能提供 BLUP (育种值) 和 BLUP (配合力) 分析两种用法。

功能二、平台调用功能。分子数据分析中，用户需要学习熟悉种类繁多的各类软件，其中有些软件需要在 linux 系统环境中运行操作，很多软件，要求用户具备编程能力，进行一次分子数据分析工作，要用到多种软件，才能解决问题。下载软件，搜索软件安装维护知识，学习熟悉不同软件编程知识，占据了用户大量的学习精力，本平台软件有望打破这一困境，解决一些这方面的问题，把用户从繁琐的软件学习和操作中解放出来，精力专注于分子数据分析本身。本平台软件 (Nbs) 通过调用各种分子数据分析使用到的软件，例如：plink，并对中间分析结果进行针对性编程适配，提供分析工具，支持统计及图示分析结果，并可以类似 windows 和 linux 系统下的批处理方式，以文件模板方式，连续执行命令。本平台软件 (Nbs) 支持调用运行 python, R 语言命令文件。

专业的 BLUP (系谱) 育种值分析：**不测序也能分析育种值，省钱实用。**

BLUP（配合力）线性混合模型分析

ssGBLUP（一步法）育种值分析

免费下载试用：<http://www.nbs.net.cn/fz.htm>

农博士试验设计及统计分析系统

(软著登字第 12181103 号)

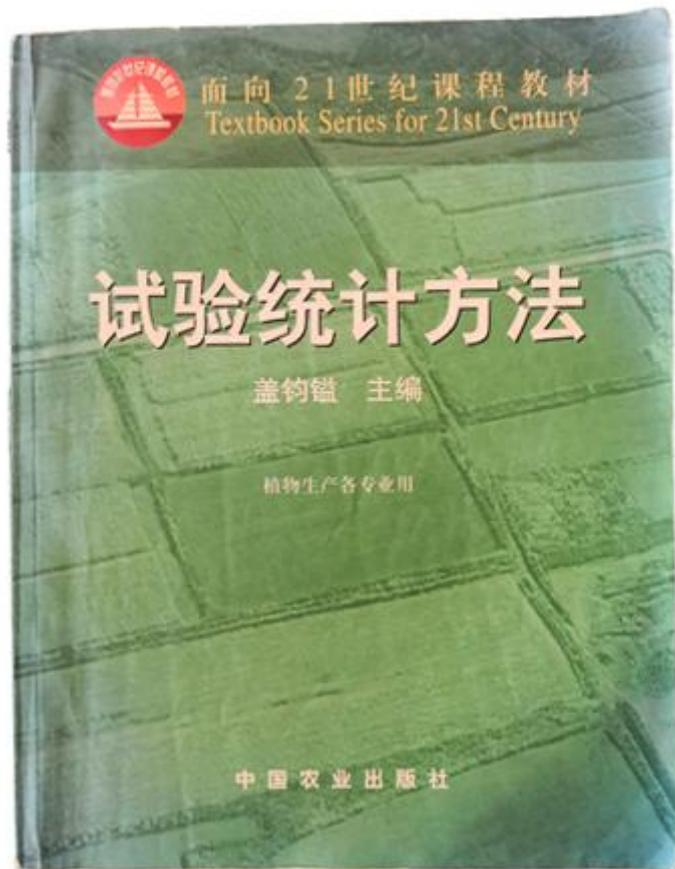
Ver 1.20

发布日期：2024-02-28

考虑育种客户使用需要，将农博士育种家软件---试验设计及统计分析模块，拆分为一个独立运行统计软件。

依据权威生物统计教材开发，软件安装包配有多个书中例题数据，可作为标准数据模板使用。

 p270条区.xls	2008-04-25 16:25	XLS 工作表	6 KB
 p263裂区.xls	2008-04-25 16:24	XLS 工作表	6 KB
 p249随机区组两因素.xls	2008-04-25 16:23	XLS 工作表	5 KB
 p228随机区组单因素.xls	2008-04-25 16:21	XLS 工作表	6 KB
 p227间比法.xls	2008-04-25 16:21	XLS 工作表	6 KB
 p225对比法.xls	2008-04-25 16:20	XLS 工作表	5 KB
 p122两向分组重复观察值.xls	2008-04-25 16:20	XLS 工作表	5 KB
 p119两向分组单个观察值.xls	2008-04-25 16:19	XLS 工作表	5 KB
 p116组内分亚组巢式试验.xls	2008-04-25 16:18	XLS 工作表	6 KB
 p113单向分组观察值不等.xls	2008-04-25 16:18	XLS 工作表	5 KB
 p111单向分组完全随机.xls	2008-04-25 16:17	XLS 工作表	5 KB



试验设计、试验分析（单因素随机区组，两因素随机区组，简单格子设计，alpha-格子设计，增广设计等）、品种区域试验（一年多点、多年多点、稳定性分析、品种性状综合评价、shukla 模型稳定性分析、Eberhart-Russell 稳定性分析）、配合力分析（Griffing 完全双列杂交 4 种方式，不完全双列杂交 NCII、不完全双列杂交 NCII 多性状）、遗传力估计（广义，狭义）、聚类分析、关联分析、通径分析、数据拟合、PCA（主成分）分析、AMMI 模型分析、BP 神经网络分析、GGE 双标图、Q-Q 图等功能。

农博士试验设计及统计分析系统，具有以下几个特点：

1、软件联系育种实际需要，为客户自动解决统计方法适用问题：很多统计分析方法，在育种实际应用中是需要选择性使用和联系育种实际情况而决定如何使用的，可以上升为育种中的统计应对策略也可以简化为一个软件参数选项，我们在该款统计软件中，充分考虑了可能遇到的数据情况进行了策略

性的编程处理。比如一年多点的布点品种试验仅有一个观测值数据，软件自动按照单因素随机区组试验进行方差分析。又比如，育种数据常有缺区情况，采用统计参数调整的方式应对缺区分析（实测发现 SAS9.1 就是这样分析的，点击查看详情：[农博士育种家与 SAS 统计软件数据分析结果比较.pdf](#)），而不是采用缺区估值方式处理等。2018 年，我公司正式出版的《育种实用统计学》一书，对育种中可能遇到的类似情况进行了更多的分析并给出了应对策略。免费电子版 PDF 文件（《[育种实用统计学第 12 版](#)》）尽管内容不全面，但可以参照。

2、软件育种专业性强：这是一款为育种工作服务的统计软件，软件中生物统计类试验设计及分析方法，数量遗传学分析方法都是围绕育种工作而设计。不是单纯的为了统计而统计。单因素随机区组设计，间比法设计，格子设计，alpha-格子设计，增广设计等都是育种常用试验设计及数据分析方法；品种区域试验（一年多点、多年多点、RSD 稳定性分析、品种性状综合评价、shukla 模型稳定性分析、Eberhart-Russell 稳定性分析，着眼育种中的品种布点品种推广；配合力分析(Griffing 完全双列杂交 4 种方式，不完全双列杂交 NCII、不完全双列杂交 NCII 多性状)、遗传力估计（广义，狭义）着眼育种中的材料选育，品种创制。

其中 Griffing 完全双列杂交 4 种分析方法

依据 Griffing（1956）年发表论文开发，保证软件分析结果忠实原著。

CONCEPT OF GENERAL AND SPECIFIC COMBINING ABILITY IN RELATION TO DIALLEL CROSSING SYSTEMS

By B. GRIFFING*

[Manuscript received June 5, 1956]

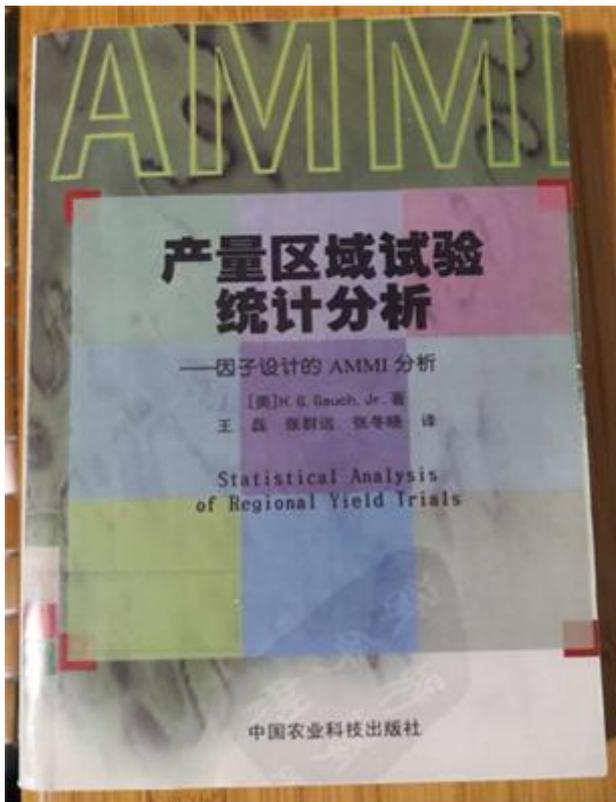
Summary

A detailed examination of the concept of combining ability in relation to diallel crossing systems is made. Eight different analyses are presented. These result from a consideration of four different diallel crossing systems together with two alternative assumptions with regard to the sampling nature of the experimental material. A numerical example is given.

I. INTRODUCTION

AMMI 模型分析开发，依据 H. G. Gauch, Jr. 著的

《产量区域试验统计分析——因子设计的 AMMI 分析》

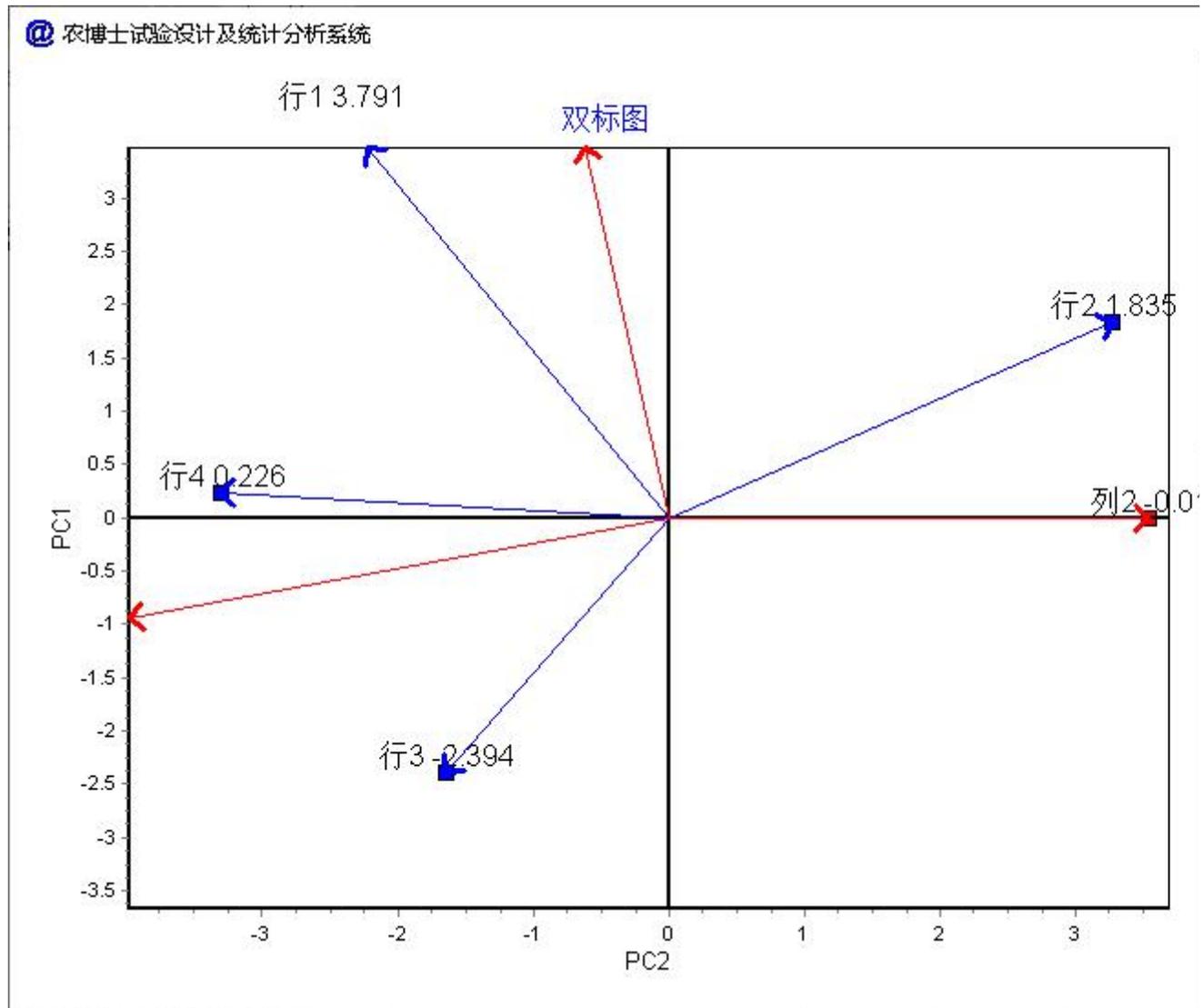


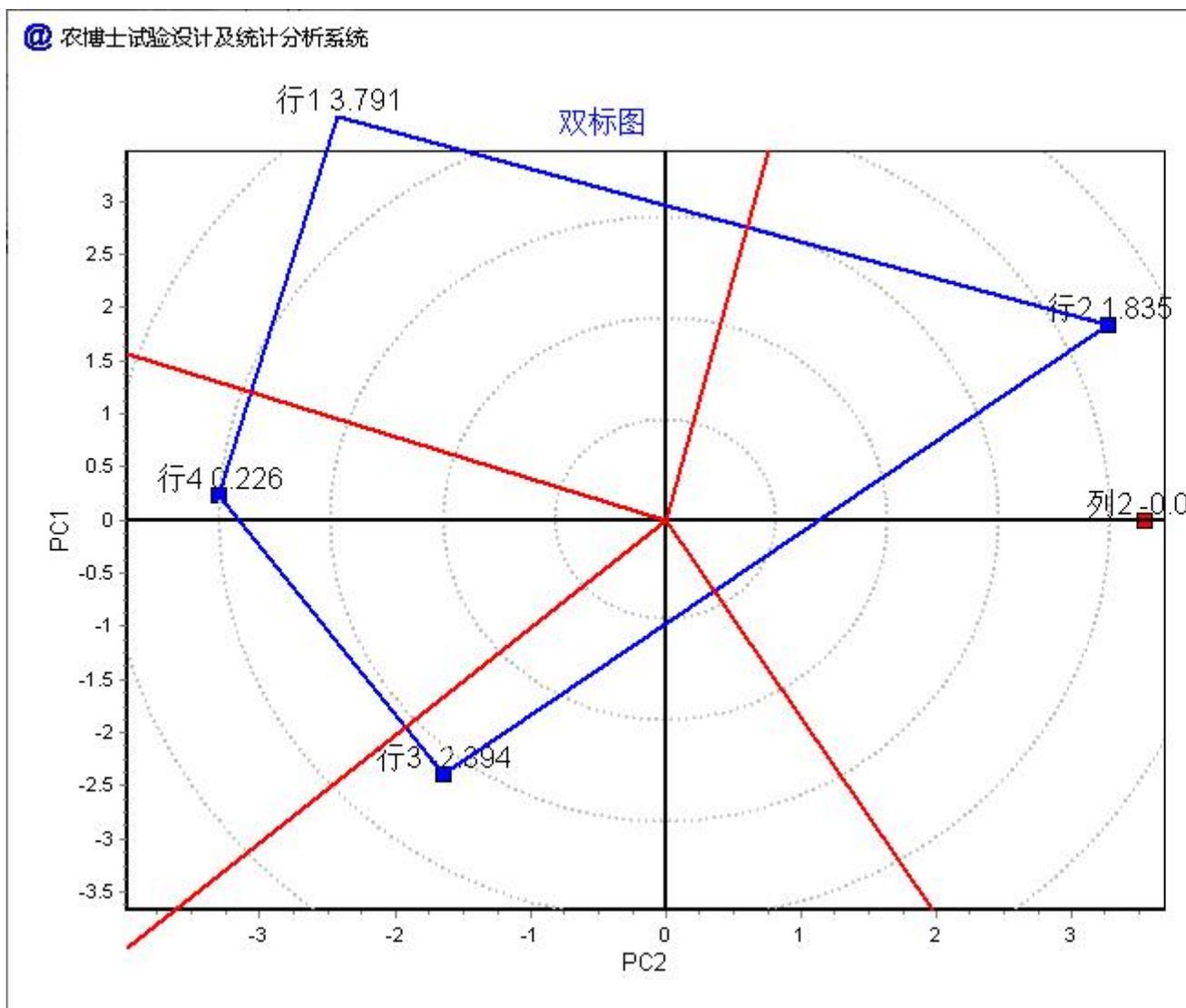
3、使用及操作简单：软件使用简易：只需三步，①设置试验参数，②依照软件模板格式整理好自己数据，③点击分析按钮，出方差分析等分析结果。

4、功能多而实用：除以上方法外，还有聚类分析、关联分析、通径分析、数据拟合、PCA（主成分）分析、BP 神经网络分析，将一些常用的热门的

数据分析方法也纳入育种数据分析技术体系，供用户使用。

GGE 双标图，部分功能免费使用。依据严威凯博士的《Crop Variety Trials Data Management and Analysis》中文书名《农作物品种试验数据管理与分析》开发。





5、一款支持脚本编程的国产统计软件：国外很多数学、统计类软件支持脚本编程，如 EXCEL 就支持 VBA 编程，且功能强大，SAS 也支持编程，为了探索编程技术和具体实现，就尝试在软件中加入了类似的功能，基本达到了相同效果，尽管这不符合国人使用软件的习惯，但是对扩展软件已有功能，是有大帮助的。需要和感兴趣的用户可以了解使用。



软件具体操作方法，见软件包中的软件操作说明书。以下表格中列出的功能均已实现，一些新增功能可能未出现在表格中。

购买者可免费获赠最新版《育种实用统计学》PDF 版本文件

软件功能项目	未注册 使用	注册 账户	付费 账户
试验设计			
对比法	√	√	√
间比法	√	√	√
完全随机	√	√	√
拉丁方	√	√	√
裂区	√	√	√
条区	√	√	√
随机区组		√	√
单因素		√	√
两因素		√	√
平衡不完全区组设计(BIB)		√	√
简单格子设计(Simple Lattice)		√	√
α-格子设计(Alpha-Lattice)		√	√
增广设计---增广随机完全区组设计		√	√
增广设计---修饰增广设计		√	√
试验分析			
成对数据比较		受限试用	√
对比法	√	√	√
对比法 (BS)		√	√
间比法	√	√	√
间比法 (BS)		√	√
完全随机---单因素试验		受限试用	√
完全随机---组内分亚组巢式试验		受限试用	√
完全随机---二因素组合内无重复观察值		受限试用	√
完全随机---二因素组合内有重复观察值		受限试用	√
拉丁方		受限试用	√
裂区		受限试用	√
条区		受限试用	√
随机区组---单因素		受限试用	√
随机区组---二因素		受限试用	√
平衡不完全区组设计(BIB)		受限试用	√
简单格子设计(Simple Lattice)		受限试用	√
α-格子设计(Alpha-Lattice)		受限试用	√
增广试验---增广随机完全区组设计		受限试用	√
增广试验---修饰增广设计		受限试用	√
品种区域试验			
一年多点区域试验统计分析		受限试用	√
多年多点区域试验统计分析		受限试用	√
品种稳定性分析		受限试用	√
品种性状综合评价		受限试用	√
Shukla 模型稳定性分析		受限试用	√
Eberhart-Russell 稳定性分析		受限试用	√
配合力			
亲本自交+正反交组合		受限试用	√
亲本自交+正交组合		受限试用	√
无亲本自交 仅亲本正反交组合		受限试用	√
无亲本自交 仅亲本正交组合		受限试用	√
不完全双列杂交(NCII)		受限试用	√
遗传			
遗传模型 加性-显性尺度测验 ABC		受限试用	√
广义遗传力估计		受限试用	√
狭义遗传力估计		受限试用	√
数据分析			
聚类分析		受限试用	√
关联分析		受限试用	√
通径分析		受限试用	√
数据拟合		受限试用	√
Q-Q 图	√	√	√
PCA主成分分析		受限试用	√
AMMI模型分析		受限试用	√
BP神经网络		受限试用	√
GGE双标图		√	√

软件使用定价：500 元。

自付费之日起，账户可连续使用 16 个月，在校学生，可连续使用 24 个月，期满停用，可续费，

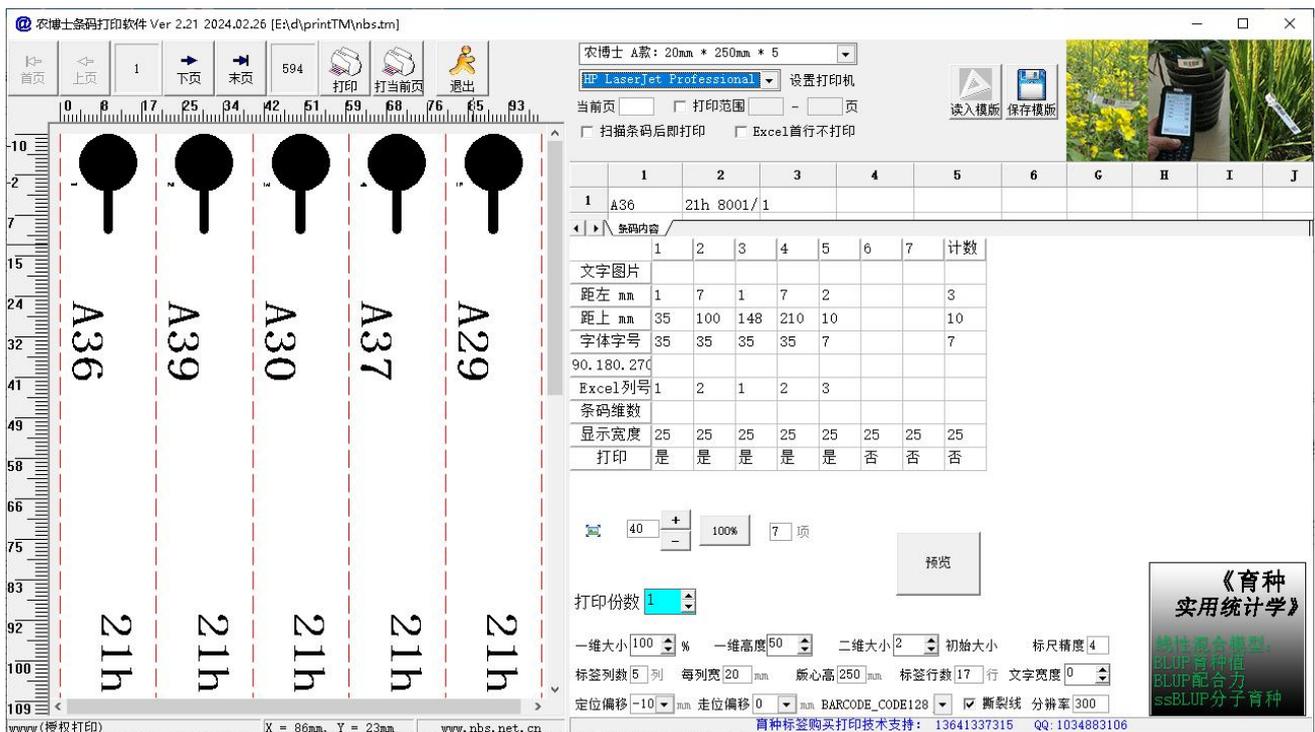
长期不限时使用定价 5000 元，欢迎咨询购买。

买软件另向购买者免费赠送最新版《育种实用统计学》PDF 版本。

免费下载试用：http://www.nbs.net.cn/nbs_tj.htm

下面是中农博思公司别的软件，欢迎下载试用：

1、农博士条码打印软件，设计打印育种挂签的利器。专为育种专用标签设计开发，所见即所得，设计轻松，好上手。农博士育种标签用户**免费使用**。实现了国外同类软件的国产替代。目前在玉米等作物育种中，用户普遍使用。



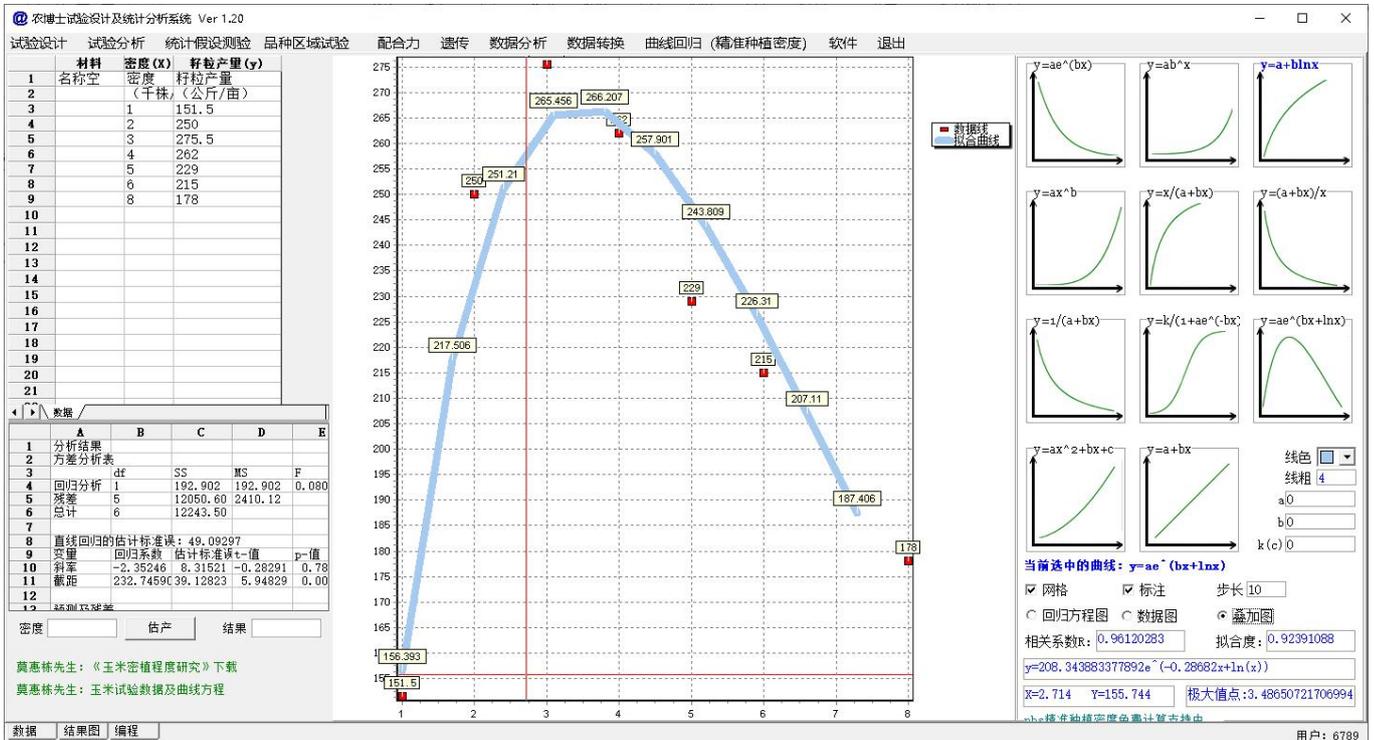
2、NBS 田间数据采集助手，android 版田间采集 APP，简单易用。农博士育种标签用户**免费使用**。

3、农博士育种家软件（NBSbreeding）



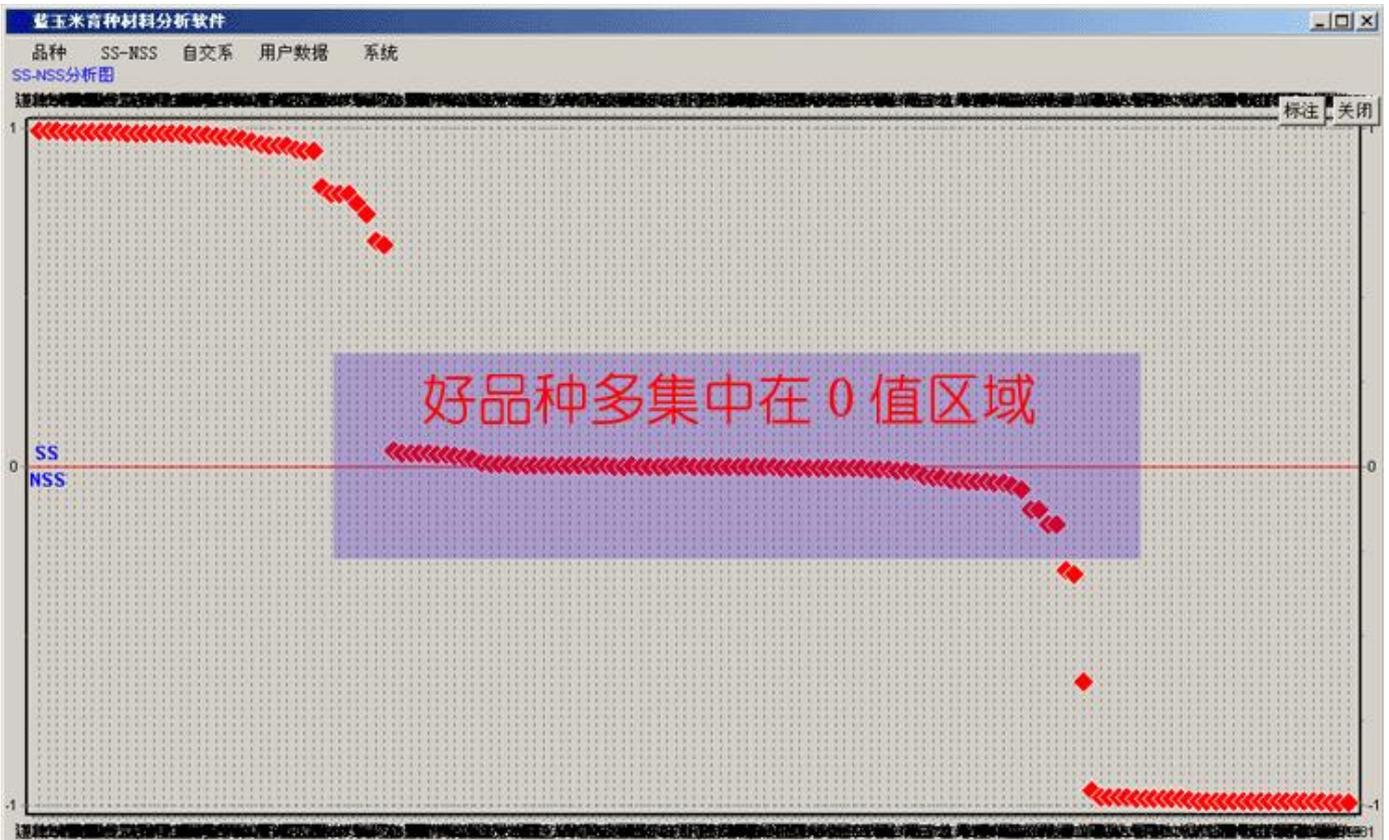
农博士育种家（NBSbreeding）诞生于2004年，做贴合中国育种实际需要的育种数据采集管理分析系统，可订制开发，成熟稳定。

4、农博士试验设计及统计分析系统



更适合植物育种使用的专业统计分析软件

6、蓝玉米育种材料分析软件（玉米）



- 7、农博士育种材料分析软件（水稻）
- 8、农博士育种材料分析软件（小麦）
- 9、农博士数据采集汇总工具软件
- 10、农博士布局图工具软件

农博士布局图工具 Ver1.00 2023.6.28

1	1	81
	2	82
	3	83
	4	84
	5	85
	6	86
	7	87
	8	88
	9	89
	10	90
	11	91
	12	92
	13	93
	14	94
	15	95
	16	96
	17	97
	18	98
	19	99
	20	100
2	21	101
	22	102
	23	103
	24	104
	25	105
	26	106
	27	107
	28	108
	29	109
	30	110
	31	111
	32	112
	33	113
	34	114
	35	115
	36	116
	37	117
	38	118
	39	119
	40	120
	41	121

地块 自定义 | 参数

地块信息

长 30 米

宽 20 米

留边行 4 行

小区信息

数量 20 填满

每小区内株行数 20

株行距 0.2 米 = 像

素点个数 15

小区间保护行数量 1

标记保护行 登录

生成 + -

读入布局 保存布局

读入EXCEL 导出为图片

(长度: 米, 面积: 平方米)

地块面积: 600.0000

使用面积: 468.1200

差值: 54.4400

地块边行面积: 77.4400

小区间保护行面积: 29.7200

小区长度: 5.4800

小区宽度: 4.0000

以小区为单位的行数: 4

以小区为单位的列数: 5

以小区为单位的行列表数: 0

小区间行保护行数量: 15

小区间列保护行数量: 16

http://www.nbs.net.cn 010-88435130

...更多信息访问我们官网：<http://www.nbs.net.cn>

扫描二维码，加入农博士软件 QQ 交流群（群号：40082612）



扫描二维码，关注农博士育种信息化微信公众号（公众号：nbsbreeding）



育种实用统计学免费视频教程网址：

<http://www.nbs.net.cn/yk.htm>

欢迎加入 **育种实用统计学** 微信群。

软件、书购买和使用答疑及统计软件交流探讨等事宜，请扫描下方二维码



扫一扫上面的二维码图案，加我为朋友。

买过书的朋友或者收费视频教程的朋友，学习中有不明白的地方欢迎来问我，内容不限于你购买的课程。跟统计，跟育种有关都可以。李奉令



北京中农博思科技发展有限公司
<http://www.nbs.net.cn>