

# 《数量遗传学》

## 练习题和参考答案

王建康执笔 张鲁燕校对

中国农业科学院作物科学研究所数量遗传课题组

2017年4月6日定稿

## 目 录

第 1 章练习题和参考答案.....	1
第 2 章练习题和参考答案.....	15
第 3 章练习题和参考答案.....	24
第 4 章练习题和参考答案.....	35
第 5 章练习题和参考答案.....	44
第 6 章练习题和参考答案.....	53
第 7 章练习题和参考答案.....	63
第 8 章练习题和参考答案.....	76
第 9 章练习题和参考答案.....	88
第 10 章练习题和参考答案.....	97
第 11 章练习题和参考答案.....	102
第 12 章练习题和参考答案.....	107
第 13 章练习题和参考答案.....	119
附录 A1: EXCEL 中公式的编辑、复制和粘贴.....	125
附录 A2: EXCEL 公式中控制符\$的使用.....	127
附录 A3: EXCEL 中的矩阵运算.....	129
附录 A4: EXCEL 中的描述统计量.....	136
附录 A5: EXCEL 中的数据分析模块.....	138

## 第 1 章练习题和参考答案

**1.1** 已知两个纯合亲本的基因型为  $AA$  和  $aa$ , 计算  $F_1$ 、 $F_2$ 、 $F_3$ 、 $BC_2F_1$  ( $AA$  是轮回亲本) 及  $BC_2F_2$  等 5 个后代群体中的杂合度和基因多样性。

**解答:** 首先根据自交和回交系统中, 基因型频率的变化规律计算每个群体的基因型频率和等位基因频率, 然后根据杂合度和基因多样性的定义, 得到下表的结果。从中还可以看出, 只有  $F_2$  群体的杂合度与多样性之和才等于 1。

群体	基因型频率			等位基因频率		杂合度	多样性
	$AA$	$Aa$	$aa$	$A$	$a$		
$F_1$	0	1	0	0.5	0.5	1	0.5
$F_2$	0.25	0.5	0.25	0.5	0.5	0.5	0.5
$F_3$	0.375	0.25	0.375	0.5	0.5	0.25	0.5
$BC_2F_1$	0.75	0.25	0	0.875	0.125	0.25	0.21875
$BC_2F_2$	0.8125	0.125	0.0625	0.875	0.125	0.125	0.21875

**1.2** 以一个座位上的三个等位基因为例, 证明随机交配大群体中, 群体杂合度等于基因多样性。

**解答:** 假定用  $p$ 、 $q$ 、 $r$  表示一个座位上三个等位基因的频率,  $p+q+r=1$ 。在随机交配大群体中, 根据 HW 平衡定律, 各种基因型的频率可以用下面多项式的展开表示为:

$$(p+q+r)^2 = p^2 + q^2 + r^2 + 2pq + 2pr + 2qr = 1$$

因此,

$$2pq + 2pr + 2qr = 1 - (p^2 + q^2 + r^2)$$

等式左端就是群体的杂合度, 等式右端就是群体的基因多样性, 因此二者在随机交配大群体中是相等的。

**1.3** 不考虑任何家系结构, 证明纯系双亲遗传研究中,  $F_3$  群体在一个座位上三种基因型的理论分离比是 3:2:3,  $P_1BC_2F_2$  群体在一个座位上三种基因型的理论分离比是 13:2:1。

**解答:** 根据自交和回交系统中, 基因型频率的变化规律, 计算得到  $F_3$  群体和  $P_1BC_2F_2$  群体的基因型频率 (见下表)。将频率最小的基因型的比例设为 2, 从  $F_3$  群体的三种基因型频率得到它们的理论分离比为 3:2:3。将频率最小的基因型的比例设为 1, 从  $P_1BC_2F_2$  群体的

三种基因型频率得到理论分离比 13:2:1。

群体	AA	Aa	aa	群体	AA	Aa	aa
F <sub>1</sub>	0	1	0	F <sub>1</sub>	0	1	0
F <sub>2</sub>	0.25	0.5	0.25	P <sub>1</sub> BC <sub>1</sub> F <sub>1</sub>	0.5	0.5	0
F <sub>3</sub>	0.375	0.25	0.375	P <sub>1</sub> BC <sub>2</sub> F <sub>1</sub>	0.75	0.25	0
分离比	3	2	3	P <sub>1</sub> BC <sub>1</sub> F <sub>2</sub>	0.8125	0.125	0.0625
				分离比	13	2	1

1.4 对表 1.1 中各种性状的两种表现型做分离比为 3:1 的适合性检验。

**解答：**利用 EXCEL 计算期望值和  $\chi^2$  值，然后利用 EXCEL 中的  $\chi^2$  分布函数 (CHISQ.DIST.RT 或 CHIDIST) 计算 P 值。所有性状的  $\chi^2$  统计量显著性概率均高于 0.05，说明它们与 3:1 的期望分离比没有显著差异。观测分离比与期望分离比之间的差异是由随机抽样误差造成的。

性状	观测值			期望值		$\chi^2$ 值	P-值
	显性	隐性	总样本	显性	隐性		
圆鼓，皱缩	5474	1850	7324	5493	1831	0.2629	0.6081
黄色，绿色	6022	2001	8023	6017.25	2005.75	0.0150	0.9025
粉红，白色	705	224	929	696.75	232.25	0.3907	0.5319
饱满，收缩	882	299	1181	885.75	295.25	0.0635	0.8010
绿色，黄色	428	152	580	435	145	0.4506	0.5021
无限，有限	651	207	858	643.5	214.5	0.3497	0.5543
长型，短型	787	277	1064	798	266	0.6065	0.4361

1.5 一个纯合亲本的籽粒形状为圆形、籽粒颜色为黄色，另一个纯合亲本的籽粒形状为皱形、籽粒颜色为绿色，它们间的杂交 F<sub>1</sub> 代全部为圆形、黄色，F<sub>2</sub> 的表型观测数为：圆形和黄色 315、圆形和绿色 108、皱形和黄色 101、皱形和绿色 32。

- (1) 对籽粒形状和籽粒颜色两个性状的表型分别做 3:1 分离比检验。
- (2) 对两个性状的 4 种表型做 9:3:3:1 分离比检验。

**解答：**如下表在 EXCEL 中计算期望值和  $\chi^2$  值，进而作分离比适合性检验。上半部分左边是观测值，右边是期望值。中部是单项  $\chi^2$  值，4 个单项  $\chi^2$  值之和就是 9:3:3:1 分离比检验

的  $\chi^2$  值；行和这一列对应的 2 个单项  $\chi^2$  值之和就是粒形性状 3:1 分离比检验的  $\chi^2$  值；列和这一行对应的 2 个单项  $\chi^2$  值之和就是籽粒颜色 3:1 分离比检验的  $\chi^2$  值。从显著性概率可以看出，对每个性状来说，两种表型的分离均符合 3:1 的理论分离比。对两个性状来说，四种表型的分离符合 9: 3: 3:1 的理论分离比。

观测值	黄色	绿色	行和	期望值	黄色	绿色	行和
圆形	315	108	423	圆形	312.75	104.25	417
皱形	101	32	133	皱形	104.25	34.75	139
列和	416	140	556	列和	417	139	

单项 $\chi^2$ 值	黄色	绿色	行和
圆形	0.016187	0.134892	0.086331
皱形	0.101319	0.217626	0.258993
列和	0.002398	0.007194	

性状	$\chi^2$ 值	P-值
籽粒形状	0.345324	0.5568
籽粒颜色	0.009592	0.9220
两个性状	0.470024	0.9254

**1.6** 在孟德尔的豌豆杂交试验中，控制黄色籽粒的基因用  $Y$  表示，其等位基因用  $y$  表示， $Y$  对  $y$  为显性，纯合基因型  $yy$  的籽粒为绿色。现有两个纯系，分别用  $P_1$  和  $P_2$  表示。 $P_1$  的基因型为  $YY$ ，籽粒颜色为黄色。 $P_2$  的基因型为  $yy$ ，籽粒颜色为绿色。

(1) 给出  $P_1$  和  $P_2$  杂交产生  $F_1$  杂种的基因型和籽粒颜色。

**解答：**  $F_1$  杂种的基因型是  $Yy$ ，籽粒颜色全部是黄色。

(2) 给出杂种  $F_1$  与  $P_1$  的回交群体中各种基因型和表型的频率；给出杂种  $F_1$  与  $P_2$  的回交群体中各种基因型和表型的频率。由此说明，为何遗传学研究中测交试验一般选择与隐性亲本做回交。

**解答：** 与  $P_1$  的回交群体中，两种基因型  $YY$  和  $Yy$  的频率各占 0.5，籽粒颜色只有黄色这一种表型。与  $P_2$  的回交群体中，两种基因型  $Yy$  和  $yy$  的频率各占 0.5，籽粒颜色有黄色和绿色两种表型，频率各占 0.5。与隐性亲本的回交，可以观测到表型分离，因此可以从表型分离比来推测控制性状的基因对数和显隐性关系。

(3) 给出杂种  $F_1$  的自交群体（即  $F_2$ ）中，各种基因型和表型的频率。

**解答：** $F_2$  群体中，三种基因型  $YY$ 、 $Yy$ 、 $yy$  的频率分别为 0.25、0.5、0.25，籽粒颜色有黄色和绿色两种表型，频率分别为 0.75 和 0.25。

(4) 利用条件概率公式，计算  $F_2$  群体中黄色籽粒个体基因型为  $YY$  的频率。

**解答：**根据 (3) 的结果， $P\{YY\} = 0.25$ ， $P\{\text{黄色籽粒}\} = 0.75$ ，因此，

$$P\{YY | \text{黄色籽粒}\} = \frac{P\{YY\}}{P\{\text{黄色籽粒}\}} = \frac{0.25}{0.75} = \frac{1}{3}$$

**1.7** 在孟德尔的豌豆杂交试验中，控制黄色籽粒的基因用  $Y$  表示，其等位基因用  $y$  表示， $Y$  对  $y$  为显性，纯合基因型  $yy$  的籽粒为绿色。控制圆型籽粒的基因用  $R$  表示，其等位基因用  $r$  表示， $R$  对  $r$  为显性，纯合基因型  $rr$  的籽粒为皱缩型。已知这两对基因位于不同染色体上。现有两个纯系，分别用  $P_1$  和  $P_2$  表示。 $P_1$  的基因型为  $YYRR$ ，为黄色和圆型籽粒。 $P_2$  的基因型为  $yyrr$ ，为绿色和皱缩型籽粒。

(1) 列出  $F_2$  群体中各种基因型的频率，以及各种基因型对应的表型。

**解答：**下表 1~2 列给出籽粒颜色座位的基因型和频率，下表 3~4 列给出籽粒形状座位的基因型和频率，第 2 列与第 3 列相乘就得到两个座位上 9 种基因型的频率，最后一列给出各种基因型的表型。

籽粒颜色座位	频率	籽粒形状座位	频率	两个座位	频率	表型
$YY$	0.25	$RR$	0.25	$YYRR$	0.0625	黄、圆
$YY$	0.25	$Rr$	0.5	$YYRr$	0.125	黄、圆
$YY$	0.25	$rr$	0.25	$YYrr$	0.0625	黄、皱
$Yy$	0.5	$RR$	0.25	$YyRR$	0.125	黄、圆
$Yy$	0.5	$Rr$	0.5	$YyYRr$	0.25	黄、圆
$Yy$	0.5	$rr$	0.25	$Yyrr$	0.125	黄、皱
$yy$	0.25	$RR$	0.25	$yyRR$	0.0625	绿、圆
$yy$	0.25	$Rr$	0.5	$yyYRr$	0.125	绿、圆
$yy$	0.25	$rr$	0.25	$yyrr$	0.0625	绿、皱

(2)  $F_2$  群体中的表型有哪些？利用概率加法定律，计算各种表型的频率。

**解答：**两个性状同时考虑时， $F_2$  群体中有黄圆、黄皱、绿圆和绿皱四种表型。对应于黄圆的基因型有  $YYRR$ 、 $YYRr$ 、 $YyRR$ 、 $YyRr$  四种，它们的频率之和等于 0.5625。对应于黄皱的基因型有  $YYrr$ 、 $Yyrr$  两种，它们的频率之和等于 0.1875。对应于绿圆的基因型有  $yyRR$ 、

yyRr 两种, 它们的频率之和等于 0.1875。对应于绿皱的基因型有 yyrr 一种, 频率等于 0.0625。

1.8 两个人群中的 MN 血型调查数据如下表:

血型	MM	MN	NN	总人数
群体 I	475	89	5	569
群体 II	233	385	129	747

(1) 计算两个群体中的基因和基因型频率。

**解答:** 下表给出两个群体中的基因和基因型频率, 同时也给出混合群体的基因和基因型频率。

群体	血型			总人数	观测基因频率		观测基因型频率		
	MM	MN	NN		M	N	MM	MN	NN
群体 I	475	89	5	569	0.9130	0.0870	0.8348	0.1564	0.0088
群体 II	233	385	129	747	0.5696	0.4304	0.3119	0.5154	0.1727
混合群体	708	474	134	1316	0.7181	0.2819	0.5380	0.3602	0.1018

(2) 检验这两个群体是否是 HWE 群体。

(3) 如将这两个群体合并形成一个混合群体, 计算混合群体中的基因和基因型频率, 并对混合群体做 HWE 的检验。

**解答:** 首先根据 (1) 的观测等位基因频率, 计算下表的 HWE 期望基因型频率, 进而得到表中的期望样本量, 最后计算  $\chi^2$  统计量和显著性概率。根据显著性概率, 可以认为群体 I 和群体 II 处于 HW 平衡, 但它们的混合群体不满足 HW 平衡。

群体	HWE 期望基因型频率			期望样本量			$\chi^2$ 值	P-值
	MM	MN	NN	MM	MN	NN		
群体 I	0.8336	0.1589	0.0076	474.31	90.39	4.31	0.1341	0.7142
群体 II	0.3245	0.4903	0.1852	242.37	366.26	138.37	1.9555	0.1620
混合群体	0.5156	0.4049	0.0795	678.59	532.82	104.59	16.0374	0.0001

(4) 利用 (2) 和 (3) 得到的检验结果, 说明遗传研究中应该注意的问题。

**解答:** 结构不同的两个 HW 平衡群体, 混合后不再是 HW 平衡群体。遗传研究中应该避免由于结构不同引起的不平衡。基因频率或基因型频率不同的群体, 不宜把它们混合在一起进行遗传研究。

1.9 一个座位上有三个等位基因 A、B 和 C, 在调查的 178 人中, 各种基因型的观测人

数如下表。

基因型	AA	AB	BB	AC	BC	CC
观测人数	17	86	61	5	9	0

(1) 计算该人群中三个等位基因 A、B 和 C 的频率。

**解答：**三个等位基因 A、B 和 C 的频率分别为 0.3511、0.6096、0.0393。

(2) 为什么调查的群体中没有基因型 CC?

**解答：**利用 (1) 计算的等位基因频率，得到随机交配大群体中各种基因型的频率（如下表），基因型 CC 的频率只有 0.0015，约 1/647。在样本量只有 178 的时候，基因型 CC 出现的概率为  $1 - (1 - 0.0015)^{178} \approx 0.2408$ 。因此，这组样本中没有观测到基因型 CC 也就不奇怪了。

基因型	AA	AB	BB	AC	BC	CC
随机交配大群体中的频率	0.1233	0.4281	0.3716	0.0276	0.0479	0.0015

(3) 检验该群体是否处于 HWE。

**解答：**将 (2) 中基因型频率乘以样本量 178，得到下表的观测样本量，进而计算  $\chi^2$  统计量和显著性概率。显著性概率高于 0.05，因此可以认为该群体处于 HWE。

基因型	AA	AB	BB	AC	BC	CC
O	17	86	61	5	9	0
E	21.95	76.19	66.14	4.92	8.53	0.28
(O-E) <sup>2</sup> /E	1.1144	1.2621	0.3989	0.0014	0.0255	0.2753
$\chi^2$ 值	3.0775					
P-值	0.3798					

**1.10** 控制人类 ABO 血型的座位上存在三个等位基因  $I^A$ 、 $I^B$  和  $I^O$ ，基因型  $I^A I^A$  和  $I^A I^O$  的血型为 A 型，基因型  $I^B I^B$  和  $I^B I^O$  的血型为 B 型，基因型  $I^A I^B$  的血型为 AB 型，基因型  $I^O I^O$  的血型为 O 型。某随机交配人群中 3 个决定血型的等位基因  $I^A$ 、 $I^B$  和  $I^O$  的频率分别为 0.7、0.2 和 0.1。

(1) 计算该人群中 6 种不同基因型的频率。

(2) 计算该人群中 A、B、AB 和 O 4 种血型的频率。

**解答：**

基因型频率					
$I^A I^A$	$I^A I^B$	$I^A I^O$	$I^B I^B$	$I^B I^O$	$I^O I^O$
0.49	0.28	0.04	0.14	0.04	0.01
表型频率					
A	B	AB	O		
0.63	0.08	0.28	0.01		

1.11 对一个人群四种血型的调查数据如下表。

血型	A	B	AB	O
人数	2162	738	228	2876

(1) 在 Hardy-Weinberg 平衡的假定下, 利用 EM 迭代算法计算三种等位基因的频率。提示: 从练习 1.10 可以看出, 如果知道了 6 种基因型的观测值或频率, 就可计算三个等位基因的频率。基因型  $I^A I^B$  和  $I^O I^O$  的观测值分别等于 AB 血型和 O 血型的观测值。基因型  $I^A I^A$  和  $I^A I^O$  的观测值需要从 A 血型的观测值分解出来。基因型  $I^B I^B$  和  $I^B I^O$  的观测值需要从 B 血型的观测值分解出来。设定等位基因  $I^A$ 、 $I^B$  和  $I^O$  的一组初始频率  $p_0$ 、 $q_0$ 、 $r_0$ , 在 HWE 下计算六种基因型的期望频率。以 A 血型为例, 两种基因型  $I^A I^A$  和  $I^A I^O$  在整个群体中的频率为  $p_0^2$  和  $2p_0 r_0$ , 由此可以计算基因型  $I^A I^A$  和  $I^A I^O$  在血型 A 中的比例, 根据这个比例把 A 血型样本观测值 2162 分解成两种基因型  $I^A I^A$  和  $I^A I^O$  的观测值。对 B 血型观测值 738 作类似的分解。然后, 重新估计三个等位基因的频率。重复上述过程, 直到等位基因频率收敛为止。

**解答:** 下面的两个表格给出 5 次 EM 迭代算法的计算结果。三个等位基因的初始频率设置为 0.5、0.3 和 0.2。从等位基因频率就可以计算 6 种基因型的频率。在初始基因频率下, 基因型  $I^A I^A$  和  $I^A I^O$  的频率分别为 0.25 和 0.2。

迭代 次数	等位基因频率			基因型频率					
	$I^A$	$I^B$	$I^O$	$I^A I^A$	$I^A I^B$	$I^A I^O$	$I^B I^B$	$I^B I^O$	$I^O I^O$
1	0.5	0.3	0.2	0.25	0.3	0.2	0.09	0.12	0.04
2	0.2991	0.1068	0.5942	0.0894	0.0639	0.3554	0.0114	0.1269	0.3530
3	0.2352	0.0855	0.6793	0.0553	0.0402	0.3196	0.0073	0.1162	0.4614
4	0.2256	0.0841	0.6903	0.0509	0.0379	0.3115	0.0071	0.1161	0.4765
5	0.2243	0.0840	0.6917	0.0503	0.0377	0.3103	0.0071	0.1162	0.4785

然后把血型 A 的观察值 2162 按照 0.25:0.2 的比例分为两部分, 即认为基因型  $I^A I^A$  的观

察值为 1201.11、基因型  $I^A I^O$  的观察值 960.89。类似地，把血型 B 的观察值 738 按照 0.09:0.12 的比例分为两部分，即认为基因型  $I^B I^B$  的观察值为 316.29、基因型  $I^B I^O$  的观察值 421.71。这样就得到第二个表格中所有 6 个基因型的观测值。根据这些观测值重新计算等位基因频率，作为下一次迭代的初始值。

迭代 次数	血型 A 的基因型		血型 B 的基因型		血型 AB	血型 O
	$I^A I^A$	$I^A I^O$	$I^B I^B$	$I^B I^O$	$I^A I^B$	$I^O I^O$
1	1201.11	960.89	316.29	421.71	228	2876
2	434.71	1727.29	60.85	677.15	228	2876
3	319.11	1842.89	43.70	694.30	228	2876
4	303.67	1858.33	42.37	695.63	228	2876
5	301.66	1860.34	42.23	695.77	228	2876

(2) 利用 (1) 得到的等位基因频率估计值，检验该群体是否处于 Hardy-Weinberg 平衡状态。

**解答：**利用 (1) 5 次迭代的基因频率，计算 Hardy-Weinberg 平衡时的四种表型的期望频率，进而得到期望观测值、 $\chi^2$  统计量和显著性概率。 $\chi^2$  统计量的自由度为  $3-1-2=1$ ，其中的 2 表示独立等位基因频率的个数。显著性概率高于 0.05，因此可以认为该群体处于 Hardy-Weinberg 平衡。

血型	A	B	AB	O
期望频率	0.3607	0.1232	0.0377	0.4785
期望样本量	2165.35	739.83	226.20	2872.62
$(O-E)^2/E$	0.0052	0.0045	0.0143	0.0040
$\chi^2$ 值	0.0280			
P-值	0.8672			

**1.12** 利用列联表的独立性检验，说明练习 1.5 中控制籽粒形状和籽粒颜色这两个座位间不存在连锁关系。

**解答：**首先如下表计算行和、列和、行（边缘）频率、列（边缘）频率。

观测值	黄色	绿色	行和	行频率
圆形	315	108	423	0.7608
皱形	101	32	133	0.2392

列和	416	140	556
列频率	0.7482	0.2518	

然后，如下表计算期望频率，即等于对应的行、列边缘频率的乘积。期望频率乘以总样本量 556 得到期望观测值。进而计算  $\chi^2$  统计量和显著性概率。显著性概率高于 0.05，因此可以认为籽粒形状和籽粒颜色这两个性状之间是独立的，控制籽粒形状和籽粒颜色的两个座位间不存在连锁关系。

表型	期望频率		期望观测值		(O-E) <sup>2</sup> /E	
	黄色	绿色	黄色	绿色	黄色	绿色
圆形	0.5692	0.1916	316.49	106.51	0.0070	0.0208
皱形	0.1790	0.0602	99.51	33.49	0.0223	0.0662
$\chi^2$ 值	0.1163					
P-值	0.7330					

**1.13** 现有两个单基因控制性状，一个性状为抗病性，有抗病和感病两种表型，一个性状为分子标记条带，有 A、H、B 三种表型。一个群体中两个性状的 6 种表型观测数据如下表。利用列联表的独立性检验，说明控制这两个性状的基因座位间是否存在连锁不平衡。

抗病性	抗病			感病		
	A	H	B	A	H	B
观测个体数	572	1161	14	3	22	569

**解答：** 首先把观测数据列成一个 3×2 的双向表，如下表计算行和、列和、行（边缘）频率、列（边缘）频率。

标记型	抗病	感病	行和	行频率
A	572	3	575	0.2456
H	1161	22	1183	0.5053
B	14	569	583	0.2490
列和	1747	594	2341	
列频率	0.7463	0.2537		

然后，如下表计算期望频率，即等于对应的行、列边缘频率的乘积。期望频率乘以总样本量 2341 得到期望观测值。进而计算  $\chi^2$  统计量和显著性概率。显著性概率几乎为 0，因此可以认为抗病性和分子标记这两个性状之间不是独立的，控制抗病性的基因和分子标记之间存在紧密的连锁关系。

标记型	期望频率		期望值		$(O-E)^2/E$	
	抗病	感病	抗病	感病	抗病	感病
A	0.1833	0.0623	429.10	145.90	47.59	139.96
H	0.3771	0.1282	882.83	300.17	87.65	257.78
B	0.1858	0.0632	435.07	147.93	407.52	1198.55
$\chi^2$ 值	2139.06					
P-值	0					

**1.14** 已知某疾病基因座位上存在 2 个等位基因 A 和 a，等位基因 a 表现为隐性遗传，在随机交配群体 I 和随机交配群体 II 中的频率分别为 0.2 和 0.4，不考虑其它影响群体结构的因素。

- (1) 计算群体 I 和群体 II 中发病个体所占的比例。
- (2) 计算群体 I 和群体 II 中正常个体携带致病基因的比例。
- (3) 如果将群体 I 和群体 II 按 1:1 比例混合，混合群体是否处于 HW 平衡？为什么？

**解答：**发病个体在两个群体中所占的比例分别为 0.04 和 0.16，正常个体携带致病基因的比例分别为 0.3333 和 0.5714，远高于发病个体的比例。

两个群体按 1:1 比例形成的混合群体中，三种基因型的频率分别为 0.5、0.4、0.1，因此得到的基因频率分别为 0.7 和 0.3。从基因频率得到 HW 平衡群体的基因型频率分别为 0.49、0.42、0.09，显然不等于混合群体的三种基因型频率 0.5、0.4、0.1。因此混合群体不是一个 HW 平衡群体。原因在于群体 I 和群体 II 具有不同的基因频率。

群体	等位基因		基因型			发病比例	携带者比例
	A	a	AA	Aa	aa		
群体 I	0.8	0.2	0.64	0.32	0.04	0.04	0.3333
群体 II	0.6	0.4	0.36	0.48	0.16	0.16	0.5714
1:1 混合群体	0.7	0.3	0.5	0.4	0.1		
混合群体平衡频率			0.49	0.42	0.09		
基因型频率之差			0.01	-0.02	0.01		

**1.15** 已知两个座位之间无连锁，每个座位上有两个等位基因，分别用 A、a 和 B、b 表

示。现将基因型为  $AABB$  和  $aabb$  的个体等量混合，混合群体称为世代 0。计算混合群体的随机交配一代和二代群体中，两个座位间的不平衡度。由此说明，对于不存在连锁的两个基因座位，如果初始群体存在不平衡，这种不平衡随机交配数代后仍然存在。

**解答：**时代 0 群体中，4 种配子  $AB$ 、 $Ab$ 、 $aB$ 、 $ab$  的频率分别为 0.5、0、0、0.5，两个座位间的不平衡度等于  $D_0=0.25$ 。由于  $r=0.5$ ，每随机交配一代，不平衡度将降低一半。因此，随机交配一代和二代的群体中，不平衡度分别为 0.125 和 0.0625。也就是说，即使两个座位之间没有连锁关系，初始群体中的不平衡度在少数几代随机交配后仍然存在。

**1.16** 已知两个座位间的重组率  $r=0.1$ ，每个座位上有两个等位基因，分别用  $A$ 、 $a$  和  $B$ 、 $b$  表示。现将基因型为  $AABB$  和  $aabb$  的个体等量混合，混合群体称为世代 0。计算混合群体的随机交配一代、10 代和 100 代群体中，两个座位间的不平衡度。由此说明，对于存在连锁的两个基因座位，长期的随机交配后，群体中将观测不到连锁不平衡。

**解答：**世代 0 群体中，4 种配子  $AB$ 、 $Ab$ 、 $aB$ 、 $ab$  的频率分别为 0.5、0、0、0.5，两个座位间的不平衡度等于  $D_0=0.25$ 。用  $r$  表示两个座位间的重组率，随机交配  $t$  代的不平衡度为  $D_t = D_0(1-r)^t$ 。由此得到，随机交配一代和 10 代的群体中，不平衡度分别为 0.225 和 0.0872。随机交配 100 代的群体中，不平衡度低于  $10^{-5}$ 。也就是说，即使两个座位之间存在连锁关系，初始群体中的不平衡度在经过很多代随机交配后将不复存在。

**1.17** 座位 A 和 B 上有四个单倍型  $AB$ 、 $Ab$ 、 $aB$ 、 $ab$ ，在一个随机交配群体中的频率分别为  $u$ 、 $s$ 、 $t$ 、 $v$  ( $u+s+t+v=1$ )。四个等位基因的频率分别为  $p_A=u+s$ 、 $p_a=t+v$ 、 $p_B=u+t$ 、 $p_b=s+v$ 。如下表，对座位 A 定义一个变量  $X$ ，对等位基因  $A$  取 1，对等位基因  $a$  取 -1；对座位 B 定义一个变量  $Y$ ，对等位基因  $B$  取 1，对等位基因  $b$  取 -1。

单倍型	观测频率 ( $p$ )	座位 A ( $X$ )	座位 B ( $Y$ )	平衡频率
$AB$	$u$	1	1	$p_A p_B$
$Ab$	$s$	1	-1	$p_A p_b$
$aB$	$t$	-1	1	$p_a p_B$
$ab$	$v$	-1	-1	$p_a p_b$

(1) 证明变量  $X$  的期望和方差分别为  $E(X) = p_A - p_a$  和  $V(X) = 4p_A p_a$ 。

**解答：**根据期望和方差的定义，

$$E(X) = u \times 1 + s \times 1 + t \times (-1) + v \times (-1) = (u + s) - (t + v) = p_A - p_a$$

$$V(X) = u \times 1^2 + s \times 1^2 + t \times (-1)^2 + v \times (-1)^2 - E^2(X)$$

$$= (u + s + t + v) - E^2(X) = 1 - E^2(X)$$

$$= (p_A + p_a)^2 - (p_A - p_a)^2 = 4p_A p_a$$

(2) 证明变量  $Y$  的期望和方差分别为  $E(Y) = p_B - p_b$  和  $V(Y) = 4p_B p_b$ 。

**解答：** 根据期望和方差的定义，

$$E(Y) = u \times 1 + s \times (-1) + t \times 1 + v \times (-1) = (u + t) - (s + v) = p_B - p_b$$

$$V(Y) = 1^2 \times u + (-1)^2 \times s + 1^2 \times t + (-1)^2 \times v - E^2(Y)$$

$$= (u + s + t + v) - E^2(Y) = 1 - E^2(Y)$$

$$= (p_B + p_b)^2 - (p_B - p_b)^2 = 4p_B p_b$$

(3) 证明变量  $X$  与  $Y$  的协方差为  $Cov(X, Y) = 4D$ ，其中  $D = uv - st$ 。

**解答：** 根据 (1) 的结果，

$$E(X) = p_A - p_a = (u + s) - (t + v) = (u - v) + (s - t)$$

$$E(Y) = p_B - p_b = (u + t) - (s + v) = (u - v) - (s - t)$$

因此，

$$E(X)E(Y) = (u - v)^2 - (s - t)^2$$

$$Cov(X, Y) = u \times 1 \times 1 + s \times 1 \times (-1) + t \times (-1) \times 1 + v \times (-1) \times (-1) - E(X)E(Y)$$

$$= u - s - t + v - (u - v)^2 + (s - t)^2$$

$$= (u - u^2) - (s - s^2) - (t - t^2) + (v - v^2) + 2(uv - st)$$

可以验证：

$$(u - u^2) - (s - s^2) - (t - t^2) + (v - v^2)$$

$$= u(1 - u) - s(1 - s) - t(1 - t) + v(1 - v)$$

$$= u(s + t + v) - s(u + t + v) - t(u + s + v) + v(u + s + t)$$

$$= 2(uv - st)$$

因此，

$$\text{Cov}(X, Y) = 4(uv - st) = 4D$$

(4) 利用相关系数的计算公式  $r = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}}$ , 证明变量  $X$  与  $Y$  相关系数的平方等于

公式 1.35 定义的连锁不平衡度。

**解答:** 根据相关系数的定义,

$$r = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{4D}{\sqrt{4p_A p_a \times 4p_B p_b}} = \frac{D}{\sqrt{p_A p_a p_B p_b}}$$

因此,

$$r^2 = \frac{D^2}{p_A p_a p_B p_b}, \text{ 即公式 1.35}$$

(5) 如样本总量为  $n$ ,  $n$  乘以观测频率为观测样本量,  $n$  乘以平衡频率为期望样本量。

证明: 两个座位处于平衡状态的适合检验  $\chi^2$  统计量为  $\chi^2 = nr^2$ , 其中  $r^2$  为公式 1.35 定义的连锁不平衡度, 其平方根等于 (4) 中变量  $X$  与  $Y$  的相关系数。

**解答:** 观测样本分别为  $un$ 、 $sn$ 、 $tn$ 、 $vn$ , 期望样本分别为  $np_A p_B$ 、 $np_A p_b$ 、 $np_a p_B$ 、 $np_a p_b$ 。将它们列于下表, 并计算各个单倍型的  $\frac{(O - E)^2}{E}$ 。

单倍型	观测频率	平衡频率	观测样本 O	期望样本 E	$\frac{(O - E)^2}{E}$
$AB$	$u$	$p_A p_B$	$nu$	$np_A p_B$	$\frac{n(u - p_A p_B)^2}{p_A p_B} = \frac{nD^2}{p_A p_B}$
$Ab$	$s$	$p_A p_b$	$ns$	$np_A p_b$	$\frac{n(s - p_A p_b)^2}{p_A p_b} = \frac{nD^2}{p_A p_b}$
$aB$	$t$	$p_a p_B$	$nt$	$np_a p_B$	$\frac{n(t - p_a p_B)^2}{p_a p_B} = \frac{nD^2}{p_a p_B}$
$ab$	$v$	$p_a p_b$	$nv$	$np_a p_b$	$\frac{n(v - p_a p_b)^2}{p_a p_b} = \frac{nD^2}{p_a p_b}$

因此, 适合度检验的  $\chi^2$  统计量为

$$\chi^2 = nD^2 \left( \frac{1}{p_A p_B} + \frac{1}{p_A p_b} + \frac{1}{p_a p_B} + \frac{1}{p_a p_b} \right) = nD^2 \left[ \frac{(p_A + p_a)(p_B + p_b)}{p_A p_B p_a p_b} \right] = \frac{nD^2}{p_A p_B p_a p_b}$$

利用 (4) 的结果  $r^2 = \frac{D^2}{p_A p_B p_a p_b}$ ，则立即得到  $\chi^2 = nr^2$ 。

## 第 2 章练习题和参考答案

**2.1** 一个座位上，等位基因A突变到a的频率是万分之一，反向突变的频率是十万分之一，三种基因型有相同的适合度。计算平衡群体中三种基因型的频率。

**解答：**这里的  $u=1/10000$ ， $v=1/100000$  或  $0.1u$ ，根据平衡频率公式 2.8，得到等位基因 A 和 a 的平衡频率分别为  $0.1/1.1=0.0909$  和  $1/1.1=0.9091$ ，基因型 AA、Aa、aa 的平衡频率分别为 0.0083、0.1653、0.8264。

**2.2** 一个果蝇突变群体在突变座位上基因型aa的频率为1，野生群体中不存在等位基因a。假定每个世代从野生群体到突变群体的迁移比例为0.01。经过5个世代的迁移之后计算：

- (1) 突变群体中等位基因a的频率。
- (2) 突变群体中纯合突变型aa所占比例的百分数。

**解答：**野生群体中等位基因 A 频率用  $p_m=1$  表示，突变群体中 A 的起始频率用  $p_0=1$  表示，野生群体到突变群体迁移频率  $m=0.01$ ，迁移世代数  $t=5$ 。根据公式 2.13，得到 5 个世代的迁移之后，突变群体中 A 的频率为  $p_5=0.0490$ 。因此，a 的频率为  $q_5=0.9510$ 。

经过 5 个世代的迁移之后，突变群体中基因型 AA、Aa、aa 的频率分别为 0.0024、0.0932、0.9044，突变型 aa 所占比例的百分数为 90.44%，比迁移之前下降了 9.56%。

**2.3** 一个隐性单基因控制的疾病使得个体没有生育能力，该疾病在某人群中的发病率为 1/2000。不考虑其它因素，计算基因频率下降一半需要经过的随机交配世代数。

**解答：**这里的  $q_0=\sqrt{1/2000}=0.0224$ ， $q_n=0.5q_0=0.0112$ 。根据公式 2.26，得到等位基因频率从  $q_0$  下降到  $q_n$  的随机交配世代数大约为 45。

**2.4** 利用表2.1中的数据，分别检验婴儿群体和成人群体是否处于HW平衡，说明两个群体中检验结果存在差异的主要原因。

**解答：**首先如下表计算等位基因的观测频率。

群体	观测值			总样本	等位基因频率	
	AA	AS	SS		A	S
婴儿群体	189	89	9	287	0.8136	0.1864
成人群体	400	249	5	654	0.8020	0.1980

然后如下表计算基因型期望频率、期望样本量、 $\chi^2$ 统计量和显著性概率。

群体	基因型期望频率			期望样本量			$\chi^2$	P 值
	AA	AS	SS	AA	AS	SS		
婴儿群体	0.6619	0.3033	0.0347	189.97	87.05	9.97	0.1434	0.7049
成人群体	0.6432	0.3176	0.0392	420.64	207.71	25.64	25.8363	0.0000

从显著性概率可以看出，婴儿群体与 HW 平衡没有显著差异，但成人群体与 HW 平衡存在极显著的差异。尽管基因频率在两个群体中的差异并不大，但从表 2.1 可以看到，两种纯合基因型的适合度均低于杂合型的适合度。遗传研究表明，该座位其实与婴幼儿的成活率有关，并且这种选择主要发生在婴儿出生之后到成年的这段时间内。因此，婴儿群体仍然可以看作一个无选择随机交配大群体，符合 HW 平衡定律。成人群体则是经过选择的婴儿群体，因此不再符合 HW 平衡定律。

**2.5** 利用表2.1中的数据，不考虑其它改变群体结构的因素，计算选择达到平衡状态时等位基因A和S的频率；根据基因频率计算平衡时三种基因型的频率，并与表2.1婴儿群体基因型频率进行对比。

**解答：**基因型 AA 的选择系数用  $s_1=0.2435$  表示，基因型 SS 的选择系数用  $s_2=0.8014$  表示。根据公式 2.29 得到选择达到平衡状态时等位基因 A 和 S 的频率分别为 0.7670 和 0.2330。平衡时三种基因型的频率分别为 0.5882、0.3575、0.0543。

平衡时，基因型 AS 和纯合基因型 SS 的频率高于表 2.1 中的婴儿群体，纯合基因型 AA 的频率则低于表 2.1 中的婴儿群体。

**2.6** 利用表2.2中的加性模型，推导选择过程中，等位基因a频率在上下代之间的关系。

**解答：**选择前后基因型频率的变化如下表。

基因型	AA	Aa	aa	总和
选择前基因型频率	$p_0^2$	$2p_0q_0$	$q_0^2$	1
适合度	1	$1-0.5s$	$1-s$	
选择后剩余基因型	$p_0^2$	$(2-s)p_0q_0$	$(1-s)q_0^2$	$1-sq_0$
选择后基因型频率	$\frac{p_0^2}{1-sq_0}$	$\frac{(2-s)p_0q_0}{1-sq_0}$	$\frac{(1-s)q_0^2}{1-sq_0}$	1

根据上表最后一行计算选择后等位基因  $a$  的频率为

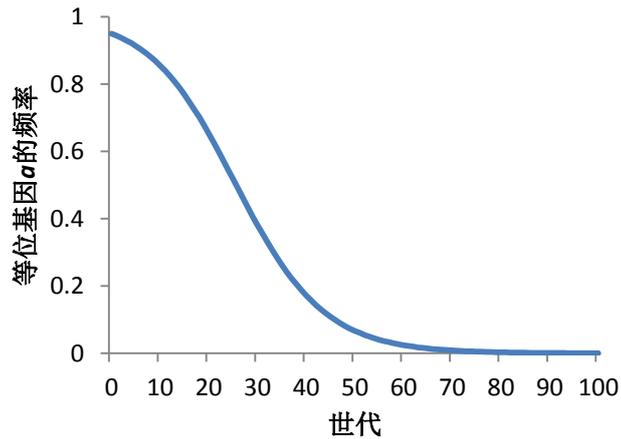
$$q_1 = \frac{(1-0.5s)p_0q_0 + (1-s)q_0^2}{1-sq_0} = \frac{q_0 - 0.5sq_0 - 0.5sq_0^2}{1-sq_0}$$

**2.7** 假定一个起始群体中等位基因  $a$  的频率为 0.95，基因型  $AA$ 、 $Aa$  和  $aa$  的适合度分别为 1、0.9 和 0.8。利用练习 2.6 的结果，计算 10 代随机交配过程中，等位基因  $a$  在每个世代中的频率，绘制基因频率随世代的变化曲线，并与图 2.4A 进行对比。

**解答：** 下表给出前 10 代的随机交配过程中，等位基因  $a$  的频率。

世代	基因 $a$ 的频率
0	0.95
1	0.9441
2	0.9376
3	0.9304
4	0.9225
5	0.9137
6	0.9041
7	0.8935
8	0.8819
9	0.8692
10	0.8555

下图给出 100 代随机交配过程中，等位基因  $a$  频率的变化曲线。这里的选择，不仅对纯合基因型  $aa$  起作用，也对杂合基因型  $Aa$  起作用。与图 2.4A 不利于纯合基因型  $aa$  的选择相比，这里的基因频率下降速度更快，并逐渐趋近于 0。



**2.8** 利用表2.2中的完全显性但选择对基因AA和Aa不利的模型，推导选择过程中等位基因a频率在上下代之间的关系。

**解答：**选择前后基因型频率的变化如下表。

基因型	AA	Aa	aa	总和
选择前基因型频率	$p_0^2$	$2p_0q_0$	$q_0^2$	1
适合度	1-s	1-s	1	
选择后剩余基因型	$(1-s)p_0^2$	$2(1-s)p_0q_0$	$q_0^2$	$1-s(1-q_0^2)$
选择后基因型频率	$\frac{(1-s)p_0^2}{1-s(1-q_0^2)}$	$\frac{2(1-s)p_0q_0}{1-s(1-q_0^2)}$	$\frac{q_0^2}{1-s(1-q_0^2)}$	1

根据上表最后一行计算选择后等位基因 a 的频率为

$$q_1 = \frac{(1-s)p_0q_0 + q_0^2}{1-s(1-q_0^2)} = \frac{(1-s)q_0 + sq_0^2}{1-s(1-q_0^2)}$$

**2.9** 已知一个群体中两个座位上四种配子型AB、Ab、aB、ab的频率分别为0.1、0.4、0.1、0.4，雌雄配子频率无差异。

- (1) 计算两个座位的不平衡度D。
- (2) 计算等位基因A、a、B、b的频率。
- (3) 计算雌雄配子随机结合产生的9种合子的频率。
- (4) 假定座位B与适合度无关，座位A与适合度有关，基因型AA和Aa的适合度为1，aa的适合度为0。计算选择发生之后9种合子的频率。
- (5) 根据(4)得到的基因型频率，计算选择后等位基因A、a、B、b的频率，并比较

选择前后等位基因频率的差异。

**解答：**

(1) 不平衡度  $D=0$ 。

(2) 等位基因  $A$ 、 $a$  的频率均为 0.5；等位基因  $B$ 、 $b$  的频率分别为 0.2 和 0.8。

(3) 雌雄配子随机结合产生 9 种合子的频率见下表第 2 列。

(4) 选择发生之后 9 种合子的频率见下表最后一列。

(5) 根据第 2 列计算等位基因  $A$ 、 $a$ 、 $B$ 、 $b$  的频率分别为 0.5、0.5、0.2 和 0.8。根据最后一列计算等位基因  $A$ 、 $a$ 、 $B$ 、 $b$  的频率分别为 0.6667、0.3333、0.2 和 0.8。可以看出，由于两个座位之间的不平衡度为 0，对基因型  $AA$ 、 $Aa$ 、 $aa$  的选择只改变  $A$  和  $a$  的频率，而不改变  $B$  和  $b$  的频率。

基因型	选择前	选择系数	选择后	标准化频率
$AABB$	0.01	1	0.01	0.013333
$AABb$	0.08	1	0.08	0.106667
$AAbb$	0.16	1	0.16	0.213333
$AaBB$	0.02	1	0.02	0.026667
$AaBb$	0.16	1	0.16	0.213333
$Aabb$	0.32	1	0.32	0.426667
$aaBB$	0.01	0	0	0
$aaBb$	0.08	0	0	0
$aabb$	0.16	0	0	0
和	1		0.75	1

**2.10** 已知一个群体中两个座位上四种配子型  $AB$ 、 $Ab$ 、 $aB$ 、 $ab$  的频率分别为 0.15、0.35、0.05、0.45，雌雄配子频率无差异。

(1) 计算两个座位的不平衡度  $D$ 。

(2) 计算等位基因  $A$ 、 $a$ 、 $B$ 、 $b$  的频率。

(3) 计算雌雄配子随机结合产生的 9 种合子的频率。

(4) 假定座位  $B$  与适合度无关，座位  $A$  与适合度有关，基因型  $AA$  和  $Aa$  的适合度为 1， $aa$  的适合度为 0。计算选择发生之后 9 种合子的频率。

(5) 根据 (4) 得到的基因型频率，计算选择后等位基因  $A$ 、 $a$ 、 $B$ 、 $b$  的频率，并比较

选择前后等位基因频率的差异。

**解答：**

- (1) 不平衡度  $D=0.15 \times 0.45 - 0.35 \times 0.05 = 0.05$ 。
- (2) 等位基因  $A$ 、 $a$  的频率均为 0.5；等位基因  $B$ 、 $b$  的频率分别为 0.2 和 0.8。
- (3) 雌雄配子随机结合产生 9 种合子的频率见下表第 2 列。
- (4) 选择发生之后 9 种合子的频率见下表最后一列。
- (5) 根据第 2 列计算等位基因  $A$ 、 $a$ 、 $B$ 、 $b$  的频率分别为 0.5、0.5、0.2 和 0.8。根据最后一列计算等位基因  $A$ 、 $a$ 、 $B$ 、 $b$  的频率分别为 0.6667、0.3333、0.2333 和 0.7667。可以看出，可以看出，由于两个座位之间不平衡的存在，对基因型  $AA$ 、 $Aa$ 、 $aa$  的选择既改变  $A$  和  $a$  的频率，也改变  $B$  和  $b$  的频率。选择同时增加了等位基因  $A$  和  $B$  的频率，同时降低了等位基因  $a$  和  $b$  的频率。

基因型	选择前	选择系数	选择后	标准化频率
$AABB$	0.0225	1	0.0225	0.03
$AABb$	0.105	1	0.105	0.14
$AAbb$	0.1225	1	0.1225	0.1633
$AaBB$	0.015	1	0.015	0.02
$AaBb$	0.17	1	0.17	0.2267
$Aabb$	0.315	1	0.315	0.42
$aaBB$	0.0025	0	0	0
$aaBb$	0.045	0	0	0
$aabb$	0.2025	0	0	0
总和	1		0.75	1

**2.11** 某随机交配物种中，三种基因型的亲代及这些亲本随机交配产生子代的观测数据如下表。

基因型	$AA$	$Aa$	$aa$
亲代观测值	50	100	50
后代观测值	200	1000	500

- (1) 计算三种基因型的适合度和选择系数。

**解答：**计算结果如下表所示，对于每种基因型，利用后代个数与亲代个数的比值作为绝对适合度的度量。以绝对适合度最高的基因型作为 1，计算其它两种基因型的相对适合度。

1 减去相对适合度即为选择系数。

基因型	AA	Aa	aa
亲代数	50	100	50
后代数	200	1000	500
绝对适合度	4	10	10
相对适合度	0.4	1	1
选择系数	0.6	0	0

(2) 假定起始群体中等位基因A的频率为0.5，不考虑其他改变群体结构的因素，计算选择和随机交配发生一代后等位基因A的频率。

**解答：**未选择群体中，三种基因型AA、Aa、aa的频率分别为0.25、0.5、0.25。分别乘以(1)中计算的基因型适合度，得到选择后三种基因型的比例。除以选择后频率的总和0.4375，得到选择后的基因型频率。选择后等位基因A的频率等于AA的频率加上Aa频率的一半，就得到表中最后一行的0.7143。

基因型	AA	Aa	aa	总和
选择前	0.25	0.5	0.25	1
选择后	0.1	0.5	0.25	0.85
相对频率	0.1176	0.5882	0.2941	1
选择后A的频率	0.4118			

**2.12** 一个自然群体中，座位A观测到两个等位变异，用 $A_1$ 和 $A_2$ 表示。座位B观测到三个，用 $B_1$ 、 $B_2$ 、 $B_3$ 表示。座位C观测到两个，用 $C_1$ 和 $C_2$ 表示。三个座位上12种单倍型的频率如下表。

单倍型编号	座位 A	座位 B	座位 C	频率
1	$A_1$	$B_1$	$C_1$	0.0960
2	$A_1$	$B_1$	$C_2$	0.0040
3	$A_1$	$B_2$	$C_1$	0.0768
4	$A_1$	$B_2$	$C_2$	0.0032
5	$A_1$	$B_3$	$C_1$	0.0192
6	$A_1$	$B_3$	$C_2$	0.0008
7	$A_2$	$B_1$	$C_1$	0.3840
8	$A_2$	$B_1$	$C_2$	0.0160
9	$A_2$	$B_2$	$C_1$	0.3072
10	$A_2$	$B_2$	$C_2$	0.0128
11	$A_2$	$B_3$	$C_1$	0.0768

12	$A_2$	$B_3$	$C_2$	0.0032
----	-------	-------	-------	--------

(1) 计算三个座位上每个等位基因的频率。

**解答:** 单倍型 1~6 的频率之和等于等位基因  $A_1$  的频率, 单倍型 7~12 的频率之和等于等位基因  $A_2$  的频率, 其它等位基因频率的计算与此类似, 得到的频率见下表。

座位	等位基因	频率/杂合度
座位 A	$A_1$	0.2
	$A_2$	0.8
	杂合度	0.32
座位 B	$B_1$	0.5
	$B_2$	0.4
	$B_3$	0.1
	杂合度	0.58
座位 C	$C_1$	0.96
	$C_2$	0.04
	杂合度	0.0768
平均杂合度		0.3256

(2) 计算每个座位的杂合度和三个座位的平均杂合度。

**解答:** 单个座位上, 杂合度等于杂合基因型频率之和, 也就是两个不同等位基因频率乘积之和的两倍。三个座位的平均杂合度等于单个座位杂合度的简单平均。结果见上表。

(3) 哪些是稀有等位变异, 哪些是常见等位变异?

**解答:** 如以频率 0.05 为标准, 只有  $C_2$  是稀有等位变异, 其它均是常见等位变异。如以频率 0.005 为标准, 则这 7 个等位基因均是常见等位变异, 没有稀有等位变异。

(4) 哪些是多态性座位, 哪些是单态性座位?

**解答:** 以频率 0.95 为标准, 座位 C 上等位基因  $C_1$  的频率高于 0.95, 是单态性座位。座位 A 和 B 上, 不存在频率高于 0.95 的等位基因, 是多态性座位。

**2.13** 假定基因型  $AA$ 、 $Aa$ 、 $aa$  的适合度分别为 1、0 和 0.5。随机交配起始群体中, 等位基因的频率用  $p$  和  $q$  表示, 不考虑其它改变群体结构的因素。

(1) 推导选择一代后等位基因 A 的频率和变化量。

**解答:** 选择前后基因型频率的变化如下表。

基因型	AA	Aa	aa	总和
选择前基因型频率	$p^2$	$2pq$	$q^2$	1
适合度	1	0	0.5	
选择后剩余基因型	$p^2$	0	$0.5q^2$	$p^2 + 0.5q^2$
选择后基因型频率	$\frac{p^2}{p^2 + 0.5q^2}$	0	$\frac{0.5q^2}{p^2 + 0.5q^2}$	1

根据上表最后一行计算选择后等位基因 A 的频率为

$$\Delta p = \frac{p^2}{p^2 + 0.5q^2} - p = \frac{-p(1-p)(1-3p)}{2p^2 + q^2}$$

(2) 群体最终达到的平衡点有哪些？并对这些平衡点进行分类。

**解答：**从 (1) 得到的等位基因 A 频率变化公式可以看出，群体达到的平衡点有三个，即  $p=0$ 、 $p=1$  和  $p=1/3$ 。当  $p$  稍大于 0 如  $p=0.01$  时， $\Delta p < 0$ ，选择发生后， $p$  值会逐渐减少并趋近于 0。当  $p$  稍小于 1 如  $p=0.99$  时， $\Delta p > 0$ ，选择发生后， $p$  值会逐渐增加并趋近于 1。因此，平衡点  $p=0$  和  $p=1$  都是稳定平衡点。对平衡点  $p=1/3$  来说，当  $p$  稍微高于  $1/3$  时， $p$  值会逐渐增加并趋近于 1；当  $p$  稍微低于  $1/3$  时， $p$  值会逐渐减少并趋近于 0。因此  $p=1/3$  是不稳定平衡点。

### 第3章练习题和参考答案

3.1 一个随机交配大群体中,某座位上等位基因A的频率为0.8。从中随机抽取5个个体,这5个个体包含等位基因A的个数是一随机变量。

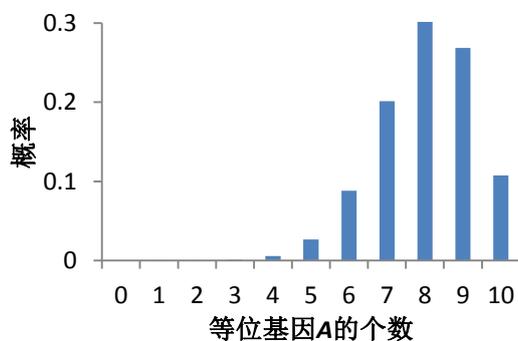
(1) X的可能取值有哪些? 给出所有可能取值的概率。

**解答:** 5个个体共携带10个基因,等位基因A的个数X的取值范围是0~10,服从试验次数等于10、成功概率等于0.8的二项分布 $B(10, 0.8)$ 。每种取值的概率见下表。

等位基因A的个数X	取值概率
0	1.02E-07
1	4.1E-06
2	7.37E-05
3	0.000786
4	0.005505
5	0.026424
6	0.08808
7	0.201327
8	0.30199
9	0.268435
10	0.107374

(2) 绘制随机变量X概率分布的柱形图。

**解答:**



(3) 分别计算X在6到9之间、以及X少于6的概率。

**解答:** 基因个数在6到9之间的概率等于0.8598; 基因个数少于6的概率等于0.0328。

(4) 用Y表示除基因A以外其它等位基因的个数。根据(1)给出的概率,分别计算X

和Y的均值、方差、以及它们之间的协方差和相关系数，并用公式3.11、3.12、3.13进行验证。

**解答：**在 EXCEL 中验证， $E(X)=8$ ， $E(Y)=2$ ， $V(X)=V(Y)=1.6$ ， $Cov(X, Y)=-1.6$ 。

**3.2** 从一个随机交配大群体中随机调查10个个体，对他们的基因型进行鉴定后共发现了13个等位基因A。试估计随机交配大群体的等位基因A频率及估计值的标准差。

**解答：**这里  $k=13$ ， $2N=20$ ，等位基因 A 频率的估计值为  $\hat{p} = \frac{k}{2N} = 0.65$ ，估计值的方差为  $V(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{2N} = 0.0114$ ，因此估计值的标准差为  $\sqrt{0.0114} = 0.1067$ 。

**3.3** 一个随机交配大群体中，等位基因A和a的频率分别为0.9和0.1，现从中随机抽取5个个体。那么，基因型AA、Aa和aa的个体数  $X_1$ 、 $X_2$ 和 $X_3$ 将服从多项分布  $B(n=5, p_1=0.81, p_2=0.18, p_3=0.01)$ 。

(1) 计算基因型AA、Aa和aa的个体数  $X_1$ 、 $X_2$ 和 $X_3$ 所有可能取值的概率。

**解答：**下表给出个体数  $X_1$ 、 $X_2$ 和 $X_3$ 所有可能的取值及取值概率。

$X_1$	$X_2$	$X_3$	取值概率
5	0	0	0.348678
4	1	0	0.387420
4	0	1	0.021523
3	2	0	0.172187
3	1	1	0.019132
3	0	2	0.000531
2	3	0	0.038264
2	2	1	0.006377
2	1	2	0.000354
2	0	3	0.000007
1	4	0	0.004252
1	3	1	0.000945
1	2	2	0.000079
1	1	3	0.000003
0	5	0	0.000189

0	4	1	0.000052
0	3	2	0.000006
0	2	3	0.000000
0	1	4	0.000000
0	0	5	0.000000

(2) 利用 (1) 计算出的概率, 分别计算  $AA$ 、 $Aa$  和  $aa$  的个体数为 0 的概率。

**解答:** 把上表中  $X_1=0$  对应的概率相加, 就得到  $AA$  个体数为 0 的概率。把  $X_2=0$  对应的概率相加, 就得到  $Aa$  个体数为 0 的概率。把  $X_3=0$  对应的概率相加, 就得到  $aa$  个体数为 0 的概率。结果分别为 0.0002、0.3707、0.9510。

(3) 利用 (1) 计算出的概率, 分别计算  $AA$ 、 $Aa$  和  $aa$  的个体数为 2 的概率。

**解答:** 把上表中  $X_1=2$  对应的概率相加, 就得到  $AA$  个体数为 2 的概率。把  $X_2=2$  对应的概率相加, 就得到  $Aa$  个体数为 2 的概率。把  $X_3=2$  对应的概率相加, 就得到  $aa$  个体数为 2 的概率。结果分别为 0.0450、0.1786、0.0010。

**3.4** 以多项分布公式 3.10 中的两个随机变量  $X_1$  和  $X_2$  为例, 根据文中给出的多项分布性质, 我们知道:  $E(X_1) = np_1$ ;  $E(X_2) = np_2$ ;  $V(X_1) = np_1(1-p_1)$ ;  $V(X_2) = np_2(1-p_2)$ ;  $E(X_1 + X_2) = n(p_1 + p_2)$ ;  $V(X_1 + X_2) = n(p_1 + p_2)(1-p_1-p_2)$ 。另外, 对于任意两个随机变量  $X$  和  $Y$ ,  $V(X + Y) = V(X) + V(Y) + 2Cov(X, Y)$ 。利用上述结果证明下面的两个结论。

(1) 随机变量  $X_1$  和  $X_2$  的协方差为  $Cov(X_1, X_2) = -np_1p_2$ 。

**解答:** 根据协方差的定义和多项分布的性质有,

$$\begin{aligned} Cov(X_1, X_2) &= \frac{1}{2}[V(X_1 + X_2) - V(X_1) - V(X_2)] \\ &= \frac{1}{2}[n(p_1 + p_2)(1 - p_1 - p_2) - np_1(1 - p_1) - np_2(1 - p_2)] \\ &= -np_1p_2 \end{aligned}$$

(2) 随机变量  $X_1$  和  $X_2$  的相关系数为  $Corr(X_1, X_2) = -\frac{p_1p_2}{\sqrt{p_1(1-p_1)p_2(1-p_2)}}$ 。

**解答:** 根据相关系数的定义有,

$$\begin{aligned} \text{Corr}(X_i, X_j) &= \frac{\text{Cov}(X_i, X_j)}{\sqrt{V(X_i)V(X_j)}} = \frac{-np_i p_j}{\sqrt{np_i(1-p_i)np_j(1-p_j)}} \\ &= -\frac{p_i p_j}{\sqrt{p_i(1-p_i)p_j(1-p_j)}} \end{aligned}$$

**3.5** 观察一个 $F_2$ 群体中12个单株的花色，红花和白花各发现有6株。

(1) 估计群体中红花表型的频率，以及估计值的标准差。

**解答：**红花表型的频率估计值为 0.5，估计值的标准差为 0.1443。

(2) 利用 $\chi^2$ 统计量检验判断该群体中红花和白花是否服从3:1的分离比。

(3) 利用校正 $\chi^2$ 统计量检验判断该群体中红花和白花是否服从3:1的分离比。

**解答：**(2) 和 (3) 的 3:1 分离比适合性检验如下表。

	红花	白花	$\chi^2$ 值	$P$ 值
观测值	6	6		
期望值	9	3		
$(O-E)^2/E$	1	3	4	0.0455
$( O-E -0.5)^2/E$	0.6944	2.0833	2.7778	0.0956

(4) 利用精确检验判断该群体中红花和白花是否服从3:1的分离比。

**解答：**在 3:1 分离比的零假设下，12 个  $F_2$  单株中的各种可能红花和白花数及取值概率列于下表，按取值概率从小到大的顺序排序，最后一列为累计概率。红花和白花各发现有 6 株的累积概率为 0.0861，高于 0.05 的显著性水平。因此认为抽样群体中红花和白花两种表型与 3:1 的分离比没有显著差异。

红花	白花	取值概率	累积概率
0	12	0.000000	0.000000
1	11	0.000002	0.000002
2	10	0.000035	0.000038
3	9	0.000354	0.000392
4	8	0.002390	0.002782
5	7	0.011471	0.014253
12	0	0.031676	0.045929
<b>6</b>	<b>6</b>	<b>0.040149</b>	<b>0.086079</b>

7	5	0.103241	0.189320
11	1	0.126705	0.316025
8	4	0.193578	0.509603
10	2	0.232293	0.741896
9	3	0.258104	1.000000

(5) 样本量较小时, 应该采用哪种检验方法?

**解答:** 当样本量较小时, 根据 $(O-E)^2/E$  计算的  $\chi^2$  值有较大的偏差, 应该采用矫正方法或精确检验。矫正方法和精确检验得到的显著性概率相差不大, 无矫正  $\chi^2$  检验得到的显著概率偏低, 因此容易得出拒绝零假设的结论。

**3.6** 对一个随机交配群体中的20个个体进行基因型鉴定, 三种基因型AA、Aa、aa分别占7、11、2个。

(1) 估计等位基因A的频率, 以及估计值的标准差。

**解答:** 等位基因 A 的频率估计值为 0.625, 估计值的标准差为 0.0765。

(2) 利用 $\chi^2$ 统计量检验判断该群体是否符合HW平衡。

(3) 利用矫正 $\chi^2$ 统计量检验判断该群体是否符合HW平衡。

**解答:** (2) 和 (3) 的 HW 平衡适合性检验如下表。

	AA	Aa	aa	$\chi^2$ 值	P 值
观测值	7	11	2		
期望频率	0.3906	0.4688	0.1406		
期望值	7.81	9.38	2.81		
$(O-E)^2/E$	0.0845	0.2817	0.2347	0.6009	0.4382
$( O-E -0.5)^2/E$	0.0125	0.1350	0.0347	0.1822	0.6695

(4) 利用精确检验判断该群体是否符合HW平衡。

**解答:** 在 HW 平衡的零假设下, 20 个个体的各种可能基因型观测值及取值概率列于下表, 按取值概率从小到大的顺序排序, 最后一列为累积概率。三种基因型 AA、Aa、aa 分别占 7、11、2 个的累积概率为 0.6473, 远高于 0.05 的显著性水平。因此认为抽样群体与 HW 平衡没有显著差异。

AA	Aa	aa	取值概率	累积概率
12	1	7	0.000050	0.000050

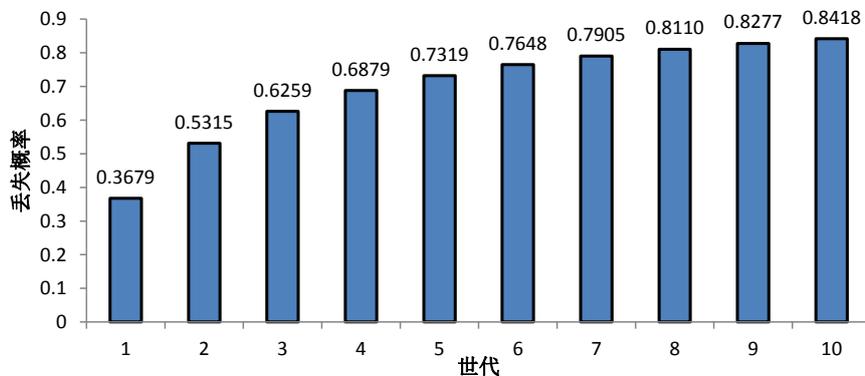
11	3	6	0.002806	0.002856
5	15	0	0.012630	0.015486
10	5	5	0.037038	0.052524
6	13	1	0.110510	0.163034
9	7	4	0.176372	0.339406
<b>7</b>	<b>11</b>	<b>2</b>	<b>0.307850</b>	<b>0.647256</b>
8	9	3	0.352744	1.000000

(5) 样本量较小时，应该采用哪种检验方法？

**解答：**当样本量较小时，根据 $(O-E)^2/E$ 计算的 $\chi^2$ 值有较大的偏差，应该采用矫正方法或精确检验。矫正方法和精确检验得到的显著性概率相差不大，无矫正 $\chi^2$ 检验得到的显著性概率偏低，因此容易得出拒绝零假设的结论。

**3.7** 根据公式3.20，计算随机交配大群体中一个突变基因1~10代的丢失概率，并绘制丢失概率随世代的变化曲线。

**解答：**

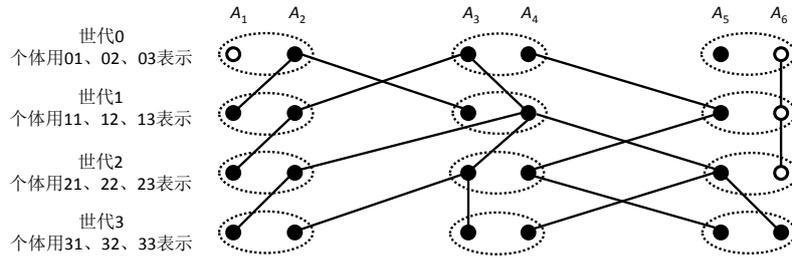


**3.8** 调查结果表明，人类双胞胎的出生比例大约为1:89。某医院每年接收大约300个孕妇。利用泊松分布计算生产双胞胎的孕妇多于6个的概率。

**解答：**此处， $n=300$ ， $p=1/89$ ，生产双胞胎孕妇的个数 $X$ 近似服从 $\lambda=np=3.3708$ 的服从泊松分布。利用EXCEL中的服从泊松分布函数‘POISSON.DIST’得到 $X$ 小于或等于6的累积概率等于0.9442，因此 $X$ 大于6的概率等于 $1-0.9442=0.0558$ 。

**3.9** 下图是大小为三的理想随机交配群体中，基因在三个世代中的传递示意图。世代0为随机交配大群体中抽取的3个个体，它们携带的六个基因用 $A_1\sim A_6$ 表示。实心圆表示等位基

因A，空心圆表示等位基因a。不考虑其它座位。



(1) 计算个体01、02、03的近交系数。

**解答：**个体 01、02、03 来自一个随机交配大群体，它们的近交系数均等于 0。

(2) 判断个体11、12、13携带的两个基因是否后裔同样。

**解答：**从基因传递示意图来看，个体 11、12、13 携带的两个基因均来自亲代的两个不同基因，因此都不是后裔同样基因。

(3) 计算个体11、12、13两两之间的共祖先系数。

**解答：**用  $X$  表示个体 11 的任一基因， $Y$  表示个体 12 的任一基因，则  $X$  为  $A_2$  和  $A_3$  的概率均等于 0.5， $Y$  为  $A_2$  和  $A_3$  的概率均等于 0.5。根据定义，个体 11 和 12 的共祖先系数为：

$$\begin{aligned} P\{X=Y\} &= 0.25 \times P\{(X=A_2) \equiv (Y=A_2)\} + 0.25 \times P\{(X=A_2) \equiv (Y=A_3)\} \\ &\quad + 0.25 \times P\{(X=A_3) \equiv (Y=A_2)\} + 0.25 \times P\{(X=A_3) \equiv (Y=A_3)\} \\ &= 0.25 \times 1 + 0.25 \times 0 + 0.25 \times 0 + 0.25 \times 1 = 0.5 \end{aligned}$$

用  $X$  表示个体 11 的任一基因， $Y$  表示个体 13 的任一基因，则  $X$  为  $A_2$  和  $A_3$  的概率均等于 0.5， $Y$  为  $A_4$  和  $A_6$  的概率均等于 0.5，它们之间的共祖先系数为 0。

用  $X$  表示个体 12 的任一基因， $Y$  表示个体 13 的任一基因，则  $X$  为  $A_2$  和  $A_3$  的概率均等于 0.5， $Y$  为  $A_4$  和  $A_6$  的概率均等于 0.5，它们之间的共祖先系数为 0。

(4) 判断个体21、22、23携带的两个基因是否后裔同样。

**解答：**从基因传递示意图来看，个体 21 携带的两个基因来自亲代的同一个基因  $A_3$ ，因此是后裔同样基因。个体 22 和 23 携带的两个基因均来自亲代的两个不同基因，因此都不是后裔同样基因。

(5) 计算个体31、32、33的近交系数。

**解答：**个体 31 携带的两个基因来着世代 0 的同一个基因  $A_3$ ，因此近交系数为 1。个体 32 携带的两个基因也来着世代 0 的同一个基因  $A_3$ ，因此近交系数也为 1。个体 33 携带的两个基因也来着世代 0 的两个基因  $A_3$  和  $A_4$ ，这两个基因的后裔同样概率为 0，因此个体 33 的

近交系数均也为 0。

**3.10** 对于图3.6的理想群体，假定 $N=2$ ，基础群体中等位基因A、a频率分别为0.7、0.3。

(1) 给出随机遗传漂变Wright-Fisher模型的转移矩阵。

**解答：**转移矩阵由下表给出。

子代群体中等位 基因 A 的个数	亲代群体中等位基因 A 的个数				
	0	1	2	3	4
0	1	0.3164	0.0625	0.0039	0
1	0	0.4219	0.25	0.0469	0
2	0	0.2109	0.375	0.2109	0
3	0	0.0469	0.25	0.4219	0
4	0	0.0039	0.0625	0.3164	1

(2) 给出世代0基因A个数的概率向量。

**解答：**世代 0 基因 A 个数的概率向量为：

基因 A 个数	概率
0	0.0081
1	0.0756
2	0.2646
3	0.4116
4	0.2401

(3) 根据转移矩阵和世代0概率向量，计算随机漂移10个世代后基因A个数的概率向量。

**解答：**随机漂移 10 个世代后，基因 A 个数的概率向量为：

基因 A 个数	概率
0	0.2789
1	0.0135
2	0.0152
3	0.0135
4	0.6789

(4) 如有1000个亚群体，随机漂移10个世代后，大概有多少个亚群体被固定在基因A上？多少个丢失了基因A？多少个基因A和a同时存在？

**解答：**将（3）的概率向量乘以 1000，得到包含不同基因 A 个数的亚群体个数为：

基因 A 个数	概率
0	279
1	14
2	15
3	14
4	679

因此随机漂移10个世代后，大概有679个亚群体被固定在基因A上，279个亚群体中丢失了基因A，43个亚群体中基因A和a同时存在。

（5）随机漂移10个世代后，计算亚群体中基因A频率的均值和方差。

**解答：**在不存在其它影响群体结构的因素时，亚群体中基因 A 频率的均值仍为 0.7。基因 A 个数 0、1、2、3、4 对应的基因频率分别为 0、0.25、0.5、0.75、1，（3）已经给出了它们的取值概率，从中计算出基因频率的方差为 0.2011。

**3.11** 对于图3.6的理想群体，假定基础群体中等位基因A、a的频率分别 $p$ 、 $q$ 。不论亚群体的大小如何，经过足够多代随机漂移后，亚群体被固定在等位基因A、a的频率也分别是 $p$ 、 $q$ 。证明在固定的亚群体中，等位基因A频率的方差等于 $p$ 和 $q$ 的乘积。

**解答：**将长期漂移的两种结果，及其对应的基因 A 频率和发生概率列为下表。

长期漂移的结果	基因 A 的频率	发生概率
A 被固定	1	$p$
A 被丢失	0	$q$
基因 A 频率的均值	$p$	
基因 A 频率的方差	$pq$	

把基因 A 的频率看作一个随机变量，用  $X$  表示。则  $X$  只有 1 和 0 两种取值、对应的取值概率为  $p$  和  $q$ ，计算出  $X$  的均值和方差分别为  $p$  和  $pq$ ，也就是基因 A 频率的均值和方差。

**3.12** 对于图3.6的理想群体，假定基础群体中等位基因A、a频率分别为0.7、0.3，亚群体大小为30。

（1）计算基础群体的基因型频率。

**解答：**三种基因型 AA、Aa、aa 的频率分别为 0.49、0.42、0.09。

(2) 计算20代随机交配后亚群体的近交系数。

**解答：**利用公式 3.34 得到 20 代随机交配后的近交系数为 0.2855。

(3) 如果20代随机交配后合并亚群体形成一个混合群体，计算混合群体的基因频率和基因型频率。

**解答：**结果 20 代随机交配的大量亚群体混合后，混合群体中等位基因  $A$ 、 $a$  频率仍然为 0.7、0.3，利用公式 3.38~公式 3.40 得到三种基因型  $AA$ 、 $Aa$ 、 $aa$  的频率分别为 0.54995、0.30010、0.14995。

(4) 在 (3) 的基础上，计算混合群体中杂合型频率相对于基础群体下降的百分数。

**解答：**混合群体中杂合型频率相对于基础群体下降的百分数为： $(0.42-0.30010)/0.42 = 0.2855$ ，也等于 (2) 计算的近交系数。

(5) 在 (3) 的基础上，计算亚群体中基因 $A$ 频率的方差。

**解答：**利用公式 3.37 得到基因  $A$  频率的方差为 0.05995。

**3.13** 在多个随机漂移群体的混合群体中，观察到一个座位上三种基因型 $AA$ 、 $Aa$ 、 $aa$ 的个体数为42、76、448。估计混合群体的近交系数。

**解答：**根据三种基因型的观测值计算基因型频率，以及混合群体中基因  $A$  的频率  $p$  和因  $a$  的频率  $q$ 。将  $p$  和  $q$  也看作基础群体中的基因频率，根据 HW 平衡定律，参考群体中杂合基因型  $Aa$  的频率为  $2pq$ ，这样就可以利用公式 3.41 计算近交系数。结果列于下表。

基因型	$AA$	$Aa$	$aa$	总样本量
观测值	42	76	448	566
基因型频率	0.0742	0.1343	0.7915	
遗传参数	基因 $A$ 的频率 $p$	基因 $a$ 的频率 $q$	基础群体杂合度	近交系数 $F$
估计值	0.1413	0.8587	0.2427	0.4468

**3.14** 一个随机交配生物种群被地理环境分割成两个亚群体，这两个亚群体分别占种群的25%和75%。调查数据表明，某座位上等位基因 $A$ 在两个亚群体中的频率分别为0.1和0.5。

(1) 计算等位基因 $A$ 在该种群中的平均频率。

**解答：**该种群中，等位基因  $A$  的平均频率等于 0.1 和 0.5 的加权平均数 0.4，权重等于亚群体占种群的比例，即 0.25 和 0.75。

(2) 分别计算这两个亚群体的杂合度。

**解答：**两个亚群体的杂合度分别为 0.18 和 0.5。

(3) 计算该种群在不存在群体结构这一假定下的平均杂合度。

**解答：**平均杂合度等于两个亚群体的杂合度 0.18 和 0.5 的加权平均数 0.42，权重等于亚群体占种群的比例，即 0.25 和 0.75。

(4) 计算亚群体间的分化系数 $F_{ST}$ 。

**解答：**把 (3) 计算的平均杂合度 0.42 看作  $H_s$ ， $H_T=2 \times 0.4 \times 0.6=0.48$  是整个种群在 HW 平衡假定下的杂合度，其中的 0.4 就是 (1) 得到的平均频率。这样，利用公式 3.41 得到的分化系数  $F_{ST}=0.125$ 。

(5) 从基因频率的方差计算分化系数 $F_{ST}$ 。

**解答：**下表给出等位基因 A 频率（用随机变量 X 表示）的均值和方差计算过程。

$$E(X)=0.25 \times 0.1+0.75 \times 0.5=0.4$$

$$V(X)=0.25 \times 0.1^2+0.75 \times 0.5^2-[E(X)]^2=0.03$$

亚群体	基因 A 的频率	权重
亚群体 1	0.1	0.25
亚群体 2	0.5	0.75
基因 A 频率的均值	0.4	
基因 A 频率的方差	0.03	

因此根据公式 3.37 得到的分化系数为： $0.03/(0.4 \times 0.6)=0.125$ ，与 (4) 利用公式 3.41 得到的分化系数相等。

## 第 4 章练习题和参考答案

**4.1** 从一个随机交配大群体中，采用相同的亚群体抽样方法，经过20个世代的遗传漂移后共产生了10个亚群体。随机交配过程中，亚群体之间没有基因交流。在最终的亚群体中调查一个座位上的基因型，得到的基因频率如下表。计算亚群体的有效群体大小。

亚群体	1	2	3	4	5	6	7	8	9	10
基因频率	0.537	0.435	0.118	0.279	0.265	0.161	0.157	0.090	0.254	0.642

**解答：**首先从题目给出的亚群体基因频率，计算 20 个世代遗传漂移后的近交系数。根据第 3 章的知识，近交系数既可以从亚群体间基因频率的方差来计算（公式 3.37），也可以从亚群体平均杂合度相对于 HW 平衡群体的下降程度来计算（公式 3.41）。简单说明如下。

方法 1：利用 10 个亚群体的基因频率，计算基因频率的均值和方差分别为 0.2928 和 0.0312，利用公式 3.37 得到近交系数等于 0.1503（见下表）。

方法 2：利用 10 个亚群体的基因频率，计算它们的杂合度，并得到平均杂合度等于 0.3526。从基因频率的均值计算 HW 平衡群体的杂合度为 0.4150。利用公式 3.41 得到近交系数等于 0.1503（见下表）。

家系	等位基因频率	杂合度
1	0.537	0.4973
2	0.435	0.4916
3	0.118	0.2082
4	0.279	0.4023
5	0.265	0.3896
6	0.161	0.2702
7	0.157	0.2647
8	0.090	0.1638
9	0.254	0.3790
10	0.642	0.4597
平均	0.2938	0.3526      0.4150
方差	0.0312	
近交系数 $F$	0.1503	0.1503

将近交系数  $F_{20}=0.1503$  代入公式 4.2, 得到  $\Delta F = \frac{1}{2N} = 0.0081$ , 从而得到用近交系数

计算的有效群体大小  $N_e = \frac{1}{2\Delta F} = 61.67$ 。

**4.2** 假定一个随机交配群体在世代0、1、2、3的大小分别为1000、1000、10、1000, 即在世代2存在一次瓶颈效应。计算有效群体大小。

**解答:** 利用公式 4.11, 有效群体大小等于 38.83。

**4.3** 假定一个果蝇群体中, 等位基因A、a频率分别为0.7、0.3。从中随机选取1只雄性和10只雌性放在一个封闭的笼子中饲养。现有大量的笼子采用同样的饲养方法, 以后的世代每个笼子都是只留下1只雄性和10只雌性。一个笼子中的群体称为一个家系。

(1) 计算随机交配10代后每个家系的近交系数。

**解答:** 首先利用公式 4.6 计算有效群体大小  $N_e=3.6364$ , 由此计算近交系数的变化量, 即  $\Delta F = \frac{1}{2N_e} = 0.1375$ 。也可利用公式 4.7 直接计算  $\Delta F$ , 得到的结果相同。然后。利用公式 3.34 得到随机交配 10 代的近交系数  $F_{10} = 1 - (1 - \Delta F)^{10} = 0.7722$ 。

(2) 如果10代随机交配后, 将这些家系合并形成一个混合群体, 计算混合群体的基因频率和基因型频率。

**解答:** 混合群体中, 等位基因 A、a 频率仍然为 0.7、0.3。利用公式 3.38~公式 3.39 得到三种基因型 AA、Aa、aa 的频率分别为 0.6522、0.0957、0.2522。

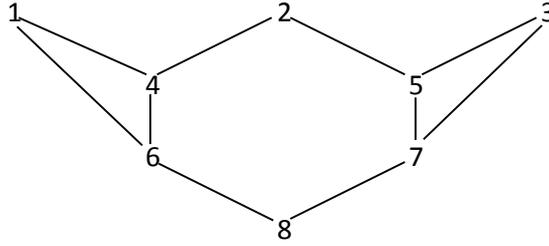
(3) 在 (2) 的基础上, 计算混合群体中杂合型频率相对于基础群体下降的百分数。

**解答:** 基础群体中, 杂合基因型 Aa 的频率为 0.42。10 代随机交配的混合群体中, 杂合基因型 Aa 的频率分别为 0.0957, 相对于基础群体下降了 77.22%, 正好等于 (1) 计算出的近交系数。

(4) 在 (2) 的基础上, 计算亚群体中基因A频率的方差。

**解答:** 利用公式 3.37, 得到亚群体中基因 A 频率的方差等于 0.1622。

**4.4** 在下面的系谱中, 1、2、3表示随机交配大群体中抽取的三个个体。



(1) 计算个体6、7、8的近交系数。

**解答：**个体 1、2、3 是随机交配大群体中抽取的三个个体，因此它们的近交系数均为 0。

个体 6 的后裔同样基因传递路径只有一条，用  $6_{146}$  表示，路径上的亲本数为 2，因此个体 6

的近交系数等于  $F_6 = \left(\frac{1}{2}\right)^2(1 + F_1) = 0.25$ 。个体 7 的后裔同样基因传递路径只有一条，用

$7_{537}$  表示，路径上的亲本数为 2，因此近交系数也等于 0.25。个体 8 的后裔同样基因传递路

径只有一条，用  $8_{642578}$ ，路径上的亲本数为 5，因此个体 8 的近交系数等于  $F_8 = \left(\frac{1}{2}\right)^5 = 0.03125$ 。

(2) 计算加性关系矩阵。

**解答：**个体 1、2、3 是基础群体的三个样本，因此

$$F_1 = F_2 = F_3 = 0,$$

$$f_{11} = \frac{1}{2},$$

$$f_{21} = 0, \quad f_{22} = \frac{1}{2}$$

$$f_{31} = 0, \quad f_{32} = 0, \quad f_{33} = \frac{1}{2}$$

个体 4 的亲本是个体 1 和个体 2，因此

$$f_{41} = \frac{1}{2}(f_{11} + f_{21}) = \frac{1}{4}$$

$$f_{42} = \frac{1}{2}(f_{12} + f_{22}) = \frac{1}{4}$$

$$f_{43} = \frac{1}{2}(f_{13} + f_{23}) = 0$$

$$f_{44} = \frac{1}{2}(1 + F_4) = \frac{1}{2}(1 + f_{12}) = \frac{1}{2}$$

个体 5 的亲本是个体 2 和个体 3，因此

$$f_{51} = \frac{1}{2}(f_{21} + f_{31}) = 0$$

$$f_{52} = \frac{1}{2}(f_{22} + f_{32}) = \frac{1}{4}$$

$$f_{53} = \frac{1}{2}(f_{23} + f_{33}) = \frac{1}{4}$$

$$f_{54} = \frac{1}{2}(f_{24} + f_{34}) = \frac{1}{8}$$

$$f_{55} = \frac{1}{2}(1 + F_5) = \frac{1}{2}(1 + f_{23}) = \frac{1}{2}$$

个体 6 的亲本是个体 1 和个体 4，因此

$$f_{61} = \frac{1}{2}(f_{11} + f_{41}) = \frac{3}{8}$$

$$f_{62} = \frac{1}{2}(f_{12} + f_{42}) = \frac{1}{8}$$

$$f_{63} = \frac{1}{2}(f_{13} + f_{43}) = 0$$

$$f_{64} = \frac{1}{2}(f_{14} + f_{44}) = \frac{3}{8}$$

$$f_{65} = \frac{1}{2}(f_{15} + f_{45}) = \frac{1}{16}$$

$$f_{66} = \frac{1}{2}(1 + F_6) = \frac{1}{2}(1 + f_{14}) = \frac{5}{8}$$

个体 7 的亲本是个体 3 和个体 5，因此

$$f_{71} = \frac{1}{2}(f_{31} + f_{51}) = 0$$

$$f_{72} = \frac{1}{2}(f_{32} + f_{52}) = \frac{1}{8}$$

$$f_{73} = \frac{1}{2}(f_{33} + f_{53}) = \frac{3}{8}$$

$$f_{74} = \frac{1}{2}(f_{34} + f_{54}) = \frac{1}{16}$$

$$f_{75} = \frac{1}{2}(f_{35} + f_{55}) = \frac{3}{8}$$

$$f_{76} = \frac{1}{2}(f_{36} + f_{56}) = \frac{1}{32}$$

$$f_{77} = \frac{1}{2}(1 + F_7) = \frac{1}{2}(1 + f_{35}) = \frac{5}{8}$$

个体 8 的亲本是个体 6 和个体 7，因此

$$f_{81} = \frac{1}{2}(f_{61} + f_{71}) = \frac{3}{16}$$

$$f_{82} = \frac{1}{2}(f_{62} + f_{72}) = \frac{1}{8}$$

$$f_{83} = \frac{1}{2}(f_{63} + f_{73}) = \frac{3}{16}$$

$$f_{84} = \frac{1}{2}(f_{64} + f_{74}) = \frac{7}{32}$$

$$f_{85} = \frac{1}{2}(f_{65} + f_{75}) = \frac{7}{32}$$

$$f_{86} = \frac{1}{2}(f_{66} + f_{76}) = \frac{21}{64}$$

$$f_{87} = \frac{1}{2}(f_{67} + f_{77}) = \frac{21}{64}$$

$$f_{88} = \frac{1}{2}(1 + F_8) = \frac{1}{2}(1 + f_{67}) = \frac{33}{64}$$

**4.5** 从一个随机交配大群体中随机抽取三个个体，按照图4.7的规则近交系统进行半同胞交配。

(1) 计算前5个世代的近交系数。

**解答：**假定世代 0 是随机交配大群体随机抽取三个个体，近交系数  $F_0=0$ ，它们之间的共祖先系数也是 0。因此，它们产生世代 1 群体的近交系数  $F_1=0$ 。利用公式 4.41 得到世代 2~5 的近交系数分别为 0.125、0.2188、0.3047、0.3809。

(2) 计算第5个世代相对于第4个的近交系数增长速率，并由此计算近交有效群体大小。

**解答：**  $\Delta F = \frac{F_5 - F_4}{1 - F_4} = 0.1096$ 。利用  $\Delta F = \frac{1}{2N_e}$  得到  $N_e=4.56$ 。

(3) 如果基础群体中一个隐性疾病基因的频率为0.1，5个世代后近交系统中的患病概率是基础群体的多少倍？

**解答：**用  $q=0.1$  表示隐性疾病基因的频率，5 个世代后的近交系数为  $F=0.3809$ 。基础群体中，隐性纯合基因型的频率为  $q^2=0.01$ ，近交群体中隐性纯合基因型的期望频率为  $q^2+q(1-q)F=0.0443$ 。因此，近交群体中的患病概率是基础群体的 4.43 倍。

**4.6** 自然状态下，一个自花授粉植物的异交率为3%。不考虑突变和选择等因素。

(1) 计算自然群体中的近交系数。

**解答：**利用公式 4.43 得到的近交系数等于 0.9417。

(2) 座位1上存在两个等位基因，频率分别为0.2和0.8，计算座位1上群体的杂合度。

**解答：**用  $p$  和  $q$  表示座位 1 上两个等位基因的频率，分别等于 0.2 和 0.8，(1) 得到的近交系数  $F=0.9417$ 。座位 1 的杂合度等于  $2pq(1-F)=0.0186$ 。

(3) 座位2上存在三个等位基因，频率分别为0.2、0.3、0.5，计算座位2上群体的杂合度。

**解答：**用  $p$ 、 $q$ 、 $r$  表示座位 2 上三个等位基因的频率，分别等于 0.2、0.3 和 0.5，(1) 得到的近交系数  $F=0.9417$ 。座位 2 的杂合度等于  $(2pq+2pr+2qr)(1-F)=0.0362$ 。

(4) 计算座位1和2上群体的平均杂合度。

**解答：**平均杂合度等于两个座位上杂合度的平均，即 0.0274。

**4.7** 在以自花授粉为主的一个植物自然群体中，考察200个个体在一个座位上的三种基因型，得到的观测数据分别为109、12、79。不考虑突变和选择等因素，计算该物种的异交

率。

**解答：**从观测数据得到杂合基因型频率的估计值值为 0.06，两个等位基因的频率分别为 0.575 和 0.425。HW 平衡群体中的杂合基因型频率为 0.48875。从杂合基因型频率下降得到的近交系数为 0.8772。从公式 4.43 得到异交率的估计值大约为 6.54%。

**4.8** 今需采集一个玉米地方品种的100粒种子，作为一份种质资源保存起来。计算下列采集方法的有效群体大小。

(1) 成熟时，从田间种植的地方品种群体中，随机选取100个植株，每个植株上随机采集1粒种子。

**解答：**有效群体大小等于 100。

(2) 成熟时，从田间种植的地方品种群体中，随机选取10个植株，每个植株上随机采集10粒种子。

**解答：**在公式 4.46 中，令  $N=100$ ， $n=10$ ，得到  $N_e=30.77$ 。

(3) 成熟时，从田间种植的地方品种群体中，随机选取1个植株，从上面随机采集100粒种子。

**解答：**在公式 4.46 中，令  $N=100$ ， $n=1$ ，得到  $N_e=3.88$ 。

(4) 成熟时，从田间种植的地方品种群体中，随机选取10个植株，分别采集5、5、5、5、10、10、10、10、20、20粒种子。

**解答：**利用公式 4.46，计算每个植株上采集到种子的有效群体大小，相加得到 100 粒种子的有效群体大小为 29.26（见下表）。

采集种子 $N$	植株数 $n$	$N_e$ (公式 4.46)
5	1	2.5
5	1	2.5
5	1	2.5
5	1	2.5
10	1	3.0769
10	1	3.0769
10	1	3.0769
10	1	3.0769
20	1	3.4783

20	1	3.4783
合计	10	29.2642

(5) 从混合收获的大量种子中，随机采集100粒种子。

**解答：**有效群体大小等于 100。

**4.9** 今需对保存多年的一份玉米地方品种进行再生繁殖，从种质库保存的种子中随机抽取100粒在田间种植，假定成活率为100%。计算下列再生繁殖方法的有效群体大小。

(1) 配置50个成对杂交，无反交，成熟时每个杂交组合采集2粒种子。

(2) 配置50个成对杂交，无反交，成熟时混合收获50个杂交组合的所有种子，最后从混合种子中随机抽取100粒。

(3) 将100个植株的花粉混合，用混合花粉进行授粉，成熟时每个植株上采集1粒种子。

(4) 将100个植株的花粉混合，用混合花粉进行授粉，成熟时混合收获100个植株上的所有种子，最后从混合种子中随机抽取100粒。

**解答：**利用公式 4.13 进行计算，结果列于下表。

问题	采集种子数 $N$	雌配子方差 $V_{kf}$	雄配子方差 $V_{km}$	有效群体大小 $N_e$
(1)	100	0	0	200
(2)	100	1	1	100
(3)	100	0	1	133.3333
(4)	100	1	1	100

**4.10** 根据群体遗传学的知识，解释图4.11中观察到的两种再生方法“100-自交-1”和“100-自交-2”在丢失基因数目上表现出来的差异。

**解答：**100-自交-1 没有再抽样误差；100-自交-2 再生繁殖时，200 个中抽取 100 个时存在一次再抽样误差。

**4.11** 在一个近交交配系统中，世代  $t$  的近交系数与前三个世代近交系数的关系是：

$F_t = \frac{1}{16}(3 + 8F_{t-1} + 4F_{t-2} + F_{t-3})$ 。假定世代0、1、2的近交系数均为0。给出0~10个世代的近交系数，并由此计算近交系数的增长速率和该近交交配系统的有效群体大小。

**解答：**利用世代  $t$  与前三个世代近交系数的关系，得到 0~10 世代的近交系数及近交系

数的增长速率如下表，增长速率的计算公式为  $\Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}}$ 。可以看出，世代 F7 之后的

增长速率就稳定在 0.13 附近，利用公式  $\Delta F = \frac{1}{2N_e}$  得到的有效群体大小大约为 3.85。

世代	近交系数	增长速率
F0	0	0
F1	0	0
F2	0	0.1875
F3	0.1875	0.115385
F4	0.28125	0.130435
F5	0.375	0.13125
F6	0.457031	0.129496
F7	0.527344	0.130165
F8	0.588867	0.130048
F9	0.642334	0.130034
F10	0.688843	0.130051

## 第 5 章练习题和参考答案

**5.1** 在大小为30的理想群体中, (1) 计算只有一份拷贝的基因在下一个世代中的丢失概率; (2) 计算只有两份拷贝的基因在下一个世代中的丢失概率。

**解答:** 如基因 A 在 30 个个体中只有一份拷贝, 它的存在频率为 1/60。如用 X 表示该基因在 30 个后代个体中的存在次数, 则 X 服从二项分布  $B(n=60, p=1/60)$ ,  $X=0$  的概率为  $(1-p)^n = 0.3648$ , 这也是基因 A 在下一个世代中的丢失概率。

如基因 A 在 30 个个体中只有两份拷贝, 它的存在频率为  $2/60=1/30$ 。这时的 X 服从二项分布  $B(n=60, p=1/30)$ ,  $X=0$  的概率为  $(1-p)^n = 0.1308$ 。因此, 两份拷贝的基因在下一个世代中的丢失概率等于 0.1308。

**5.2** 如果一个群体的参数  $\theta = 10$ , 利用公式5.15分别计算大小为10、20、50的样本群体中多态性位点的平均数。

**解答:** 公式 5.15 中,  $k$  有 10、20、50 三种取值, 参数  $\theta = 10$ 。对应的  $a = \sum_{i=1}^{k-1} \frac{1}{i}$  和  $E(S) = \theta a$

列于下表。

多态性位点 参数	样本群体大小 $k$		
	10	20	50
$a = \sum_{i=1}^{k-1} \frac{1}{i}$	2.8290	3.5477	4.4792
$E(S) = \theta a$	28.2897	35.4774	44.7921

**5.3** 参阅一本概率统计方面的教材, 证明几何分布  $G(p)$  的均值和方差公式5.3。

**证明:** 假定随机变量  $T$  服从几何分布  $G(p)$ , 其分布概率为

$$P(T=t) = p(1-p)^{t-1} = pq^{t-1}, \text{ 其中 } t=1, 2, \dots, 0 < p < 1$$

因此, 随机变量  $T$  的期望为

$$E(T) = \sum_{t=1}^{\infty} tq^{t-1} = p \sum_{t=1}^{\infty} \frac{dq^t}{dq} = p \frac{d}{dq} \sum_{t=0}^{\infty} q^t = p \frac{d}{dq} \left( \frac{1}{1-q} \right) = p \frac{1}{(1-q)^2} = \frac{1}{p}$$

随机变量  $T^2$  的期望为

$$E(T^2) = \sum_{t=1}^{\infty} t^2 pq^{t-1} = \left( \sum_{t=1}^{\infty} t(t-1) pq^{t-1} + \sum_{t=1}^{\infty} tpq^{t-1} \right)$$

其中,

$$\sum_{t=1}^{\infty} t(t-1) pq^{t-1} = pq \sum_{t=1}^{\infty} \frac{d^2 q^t}{dq^2} = pq \frac{d^2}{dq^2} \sum_{t=0}^{\infty} q^t = pq \frac{d^2}{dq^2} \left( \frac{1}{1-q} \right) = \frac{2pq}{(1-q)^3} = \frac{2q}{p^2}$$

$$\sum_{t=1}^{\infty} tpq^{t-1} = E(T) = \frac{1}{p}$$

因此,  $E(T^2) = \frac{2q}{p^2} + \frac{1}{p}$

$$V(T) = E(T^2) - E^2(T) = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

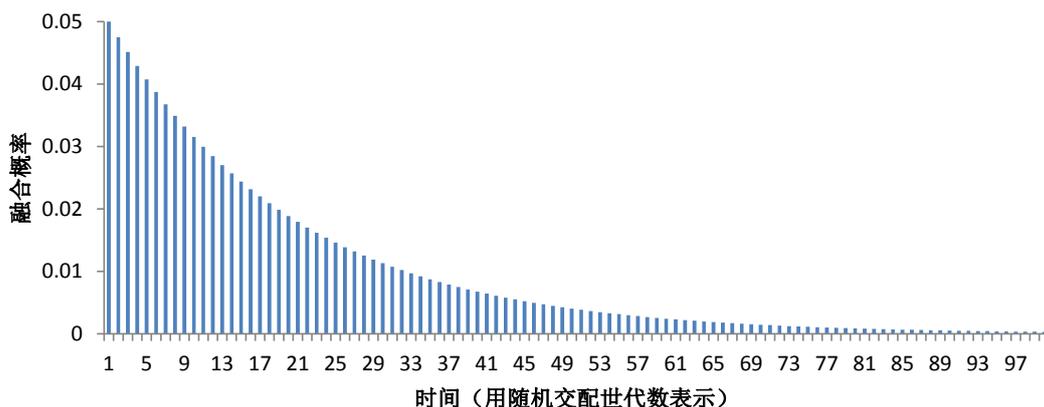
**5.4** 在大小为10的理想群体中, 绘制两个基因融合时间的概率分布柱形图。

**解答:** 群体大小  $N=10$ , 两个基因的融合时间  $T_2$  服从几何分布  $G(p = \frac{1}{2N} = 0.05)$ , 取值

概率为

$$\Pr\{T_2 = t\} = \frac{1}{2N} \left(1 - \frac{1}{2N}\right)^{t-1}, \text{ 其中 } t=1, 2, \dots$$

在 EXCEL 中计算融合时间  $T$  的取值概率, 概率分布柱形图如下:



**5.5** 一个大小为50的理想群体中, 在一个座位上存在4个不同的等位基因, 不考虑突变和选择。大约经过多少个世代, 群体中只存在3个等位基因? 大约经过多少个世代, 群体中只存在2个等位基因? 大约经过多少个世代, 群体被固定到一个等位基因上?

**解答：**在公式 5.10 中，令  $N=50$ ， $k=4、3、2$  对应的  $E(T_k) = \frac{4N}{k(k-1)}$  就是群体中只存在

在 3 个、2 个和被固定到一个等位基因上的期望世代数，即 16.67、33.33、100。

**5.6** 在练习 5.5 的基础上，如果每个世代的突变频率为 0.001，不考虑选择。在最初的 4 个等位基因融合成一个等位基因的过程中，计算突变产生的等位基因个数。在群体达到突变和漂移的平衡状态时，计算群体的近交系数和等位基因的平均个数。

**解答：**利用练习 5.5 的结果和公式 5.14，得到基因融合成一个的分支总长为  $E(T) = 4 \times E(T_4) + 3 \times E(T_3) + 2 \times E(T_2) = 366.67$ ，因此利用公式 5.15 得到突变产生的等位基因的个数为  $E(S) = uE(T) = 0.3667$ 。

这里， $N=50$ ， $u=0.001$ ，因此  $\theta = 4Nu = 0.2$ 。利用公式 5.18 得到的平衡状态近交系数为  $\tilde{F} = \frac{1}{1+\theta} = 0.8333$ 。在公式 5.27 中，令  $n=2N=100$ ， $\theta = 0.2$ ，于是得到等位基因的平均个数  $E(k) = 1.9782$ 。

**5.7** 对同一段长度为 5kb 的基因区域，检测了 6 份遗传材料的 DNA 序列，得到 20 个多态性位点上的碱基如下，其他位置上没有差异。

材料1: GCCTT TATGG CCTGT ATGAG

材料2: ACTAT TAAGG CTTGT TTGAT

材料3: ACCAC TGTCG CCCGT ACGCG

材料4: GTCAT TGTGG TCCTC TTGAG

材料5: GCTTT TATGA CCTTT ATAAG

材料6: ACCAT CATGA CCTTT ATAAT

(1) 计算多态性位点的比例。

**解答：**多态性座位的比例等于  $20/5000=0.4\%$ 。

(2) 计算平均非匹配数。

**解答：**这 20 个多态性位点都是两种碱基的差异，下表给出每个位点上两种碱基的样本量，二者相乘就是该位点上的非匹配数。20 个位点上的非匹配数相加，得到总的非匹配数 132，除以比对次数  $C_6^2 = 15$ ，得到平均非匹配数 8.8。

多态位点	样本量		非匹配数
	$n_1$	$n_2$	
1	3	3	9
2	5	1	5
3	4	2	8
4	4	2	8
5	5	1	5
6	5	1	5
7	4	2	8
8	5	1	5
9	5	1	5
10	4	2	8
11	5	1	5
12	5	1	5
13	4	2	8
14	3	3	9
15	5	1	5
16	4	2	8
17	5	1	5
18	4	2	8
19	5	1	5
20	4	2	8
总的非匹配数			132
平均非匹配数			8.8

(3) 根据公式5.31估计群体参数  $\theta = 4Nu$ ，并用公式5.30计算该估计值的方差和标准差。

**解答：**这里，多态性位点数  $S=20$ ， $a=1+1/2+1/3+1/4+1/5=2.8333$ 。根据公式 5.31 得到  $\theta = 4Nu$  的估计值为  $S/a=8.7591$ 。把这一估计值和  $n=6$  代入公式 5.30，得到估计值的方差为 132.29。

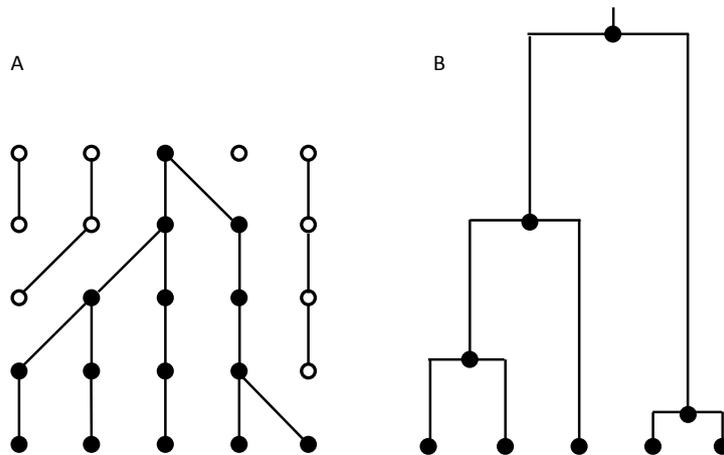
(4) 根据公式5.33估计群体参数  $\theta = 4Nu$ ，并用公式5.32计算该估计值的方差和标准差。

**解答：**根据公式 5.33， $\theta = 4Nu$  的估计值即为非匹配数  $\Pi=8.8$ 。把这一估计值和  $n=6$  代入公式 5.32，得到估计值的方差为 29.92。

(5) 如果这6份材料是随机交配群体的一组随机样本，你觉得中性理论是否可以解释这个座位上的多态性？这个座位上的基因是否受选择的作用？

**解答：**(3) 和 (4) 得到的估计值相差很小，由公式 5.34 计算的统计量将接近于 0。因此可以认为这组样本群体符合中性理论，这个座位没有受到选择的作用。

**5.8** 下图A是一个原始群体中5个基因传递过程的谱系，最终被固定下来的基因用实心圆表示，丢失的基因用空心圆表示。下图B是根据图A绘制的基因融合过程示意图，也称为基因树。请注意，这里的基因树与图5.3的表示方法是等价的。(1) 计算基因树从根部到顶部的长度（单位用世代表示）；(2) 计算基因树中所有分支的总长度（单位用世代表示）。



**解答：**这里， $2N=5$ 。根据公式 5.10 计算融合事件  $T_5$ 、 $T_4$ 、 $T_3$ 、 $T_2$  的期望时间（见下表）。这些期望时间相加，就得到基因树从根部到顶部的长度等于 8 个世代。利用公式 5.14 得到基因树中所有分支的总长度为 20.83 个世代。

融合事件	基因个数 $k$	期望融合时间
$T_5$	5	0.5
$T_4$	4	0.8333
$T_3$	3	1.6667
$T_2$	2	5
$T_1$		8

**5.9** 一个大小为500的理想群体中，每个世代基因的突变频率为 $10^{-3}$ ，不考虑选择。

(1) 计算突变和漂移共同作用下达到平衡状态时群体的固定系数。

**解答：**这里， $N=500$ ， $u=0.001$ ， $\theta=4Nu=2$ 。根据公式 5.18 得到突变和漂变达到平衡状态时群体的固定系数为 $\tilde{F} = 0.3333$ 。

(2) 假定在一个大小为10的样本群体中只发现了两种等位基因，计算这两种等位基因的期望频率。

**解答：**利用(1)的结果有 $\theta=2$ 。根据已知条件，用 $n=2 \times 10=20$ 表示样本群体的基因个数， $k=2$ 表示样本群体中等位基因的个数。首先计算公式 5.28 中的 $S_n(\theta) = 5.1091 \times 10^{19}$ 。用 $n_1$ 和 $n_2$ 表示两个等位基因的样本数，不区分等位基因的顺序， $n_1$ 和 $n_2$ 有10种可能的取值（见下表），取值1和19对应的基因频率为0.05和0.95、取值2和18对应的基因频率为0.1和0.9等等。然后计算公式 5.28 的取值概率，这些取值概率之和等于0.01737。因此得到等位基因数 $k=2$ 的条件概率。

$i=0, n-1$	$\theta+i$	$n_1$	$n_2$	概率	$k=2$ 的条件概率
0	2	1	19	0.005013	0.288571
1	3	2	18	0.002646	0.152301
2	4	3	17	0.001867	0.107507
3	5	4	16	0.001488	0.085669
4	6	5	15	0.00127	0.073105
5	7	6	14	0.001134	0.065272
6	8	7	13	0.001047	0.060251
7	9	8	12	0.000992	0.057113
8	10	9	11	0.000962	0.055382
9	11	10	10	0.000952	0.054828
10	12	总和		0.01737	
11	13				
12	14				
13	15				

14	16
15	17
16	18
17	19
18	20
19	21
$S_n(\theta)$	5.10909E+19

上表最后一列的概率可以理解为，如有很多大小为 10 的样本群体，如果把包含两个等位基因的样本群体挑出来的话，则这些样本群体中，基因频率为 0.95 和 0.05 的样本群体占 28.86%、基因频率为 0.9 和 0.1 的样本群体占 15.23%、...、基因频率为 0.55 和 0.45 的样本群体占 5.54%、基因频率为 0.5 和 0.5 的样本群体占 5.48%。

**5.10** 一个大小为100的理想群体，在长期的随机漂移过程中近交系数接近于1。从目前开始，每个世代从外界迁入10个个体，随机交配的后代群体大小仍维持在100。

- (1) 计算多少个世代后，群体的近交系数下降到0.2以下；
- (2) 计算迁移和漂移共同作用下达达到平衡状态近交系数。

**解答：**根据题意， $N=100$ ， $Nm=10$ ， $m=10/100=0.1$ 。利用公式 5.35 得到 1~20 个世代迁移过程中近交系数的变化（见下表）。可以看出经过 8 个世代的迁移后，群体的近交系数就下降到 0.2 以下。利用公式 5.36 得到迁移和漂变达到平衡状态近交系数为 0.0244。

世代	近交系数 $F$
0	1
1	0.81
2	0.656870
3	0.533454
4	0.433987
5	0.353822
6	0.289213
7	0.237141
8	0.195174
9	0.161350

10	0.134090
11	0.112120
12	0.094413
13	0.080142
14	0.068641
15	0.059371
16	0.051900
17	0.045879
18	0.041026
19	0.037115
20	0.033963
迁移和漂变的平衡状态	0.024390

**5.11** 在一个大群体中，如果随机交配的比例为 $C$ ，全同胞交配的比例为 $1-C$ ，证明平衡近交系数为 $F = \frac{1-C}{1+3C}$ 。[提示：参考公式4.42和公式4.43，把公式4.42中的自交近交系数用全同胞交配的近交系数公式4.39代替，然后令相邻三个世代的近交系数相等即可]

**证明：**利用近交系数公式 4.39，可以知道全同胞交配过程中，上下代近交系数之间的关系是 $F_t = \frac{1}{4}(1+2F_{t-1}+F_{t-2})$ ，比例为 $1-C$ 。随机交配的近交系数 $F_t=0$ ，比例是 $C$ 。因此，混合全同胞交配和随机交配后的近交系数为：

$$F_t = 0 \times C + \frac{1}{4}(1+2F_{t-1}+F_{t-2})(1-C) = \frac{1}{4}(1+2F_{t-1}+F_{t-2})(1-C)$$

上式中令三个相邻世代的近交系数相等，并用 $F$ 表示，于是就得到 $F = \frac{1-C}{1+3C}$ 。

**5.12** 在一个大群体中，如果随机交配的比例为 $C$ ，半同胞交配的比例为 $1-C$ ，证明平衡近交系数为 $F = \frac{1-C}{1+7C}$ 。[提示：参考公式4.42和公式4.43，把公式4.42中的自交近交系数用半同胞交配的近交系数公式4.41代替，然后令相邻三个世代的近交系数相等即可]

**证明：**利用近交系数公式 4.41，可以知道半同胞交配过程中，上下代近交系数之间的关系是 $F_t = \frac{1}{8}(1+6F_{t-1}+F_{t-2})$ ，比例为 $1-C$ 。随机交配的近交系数 $F_t=0$ ，比例是 $C$ 。因此，混合全同胞交配和随机交配后的近交系数为：

$$F_t = 0 \times C + \frac{1}{8}(1 + 6F_{t-1} + F_{t-2})(1 - C) = \frac{1}{8}(1 + 6F_{t-1} + F_{t-2})(1 - C)$$

上式中令三个相邻世代的近交系数相等，并用 $F$ 表示，于是就得到 $F = \frac{1 - C}{1 + 7C}$ 。

## 第 6 章练习题和参考答案

**6.1** 利用表6.1中的英国妇女身高数据。

(1) 计算身高的样本均值、样本方差和标准差。

**解答：**利用 EXCEL 中的积和函数 SUMPRODUCT，得到样本均值和样本方差分别为 63.06 和 7.25。利用平方根函数 SQRT，得到样本标准差为 2.69。这里的样本量很大，样本方差的计算时可以做无偏修正。

(2) 利用  $\chi^2$  统计量检验这组数据是否服从正态分布。

**解答：**把群体看作均值等于 63.06、方差等于 7.25 的正态分布。利用 EXCEL 中的正态分布函数计算每组样本的理论频率，乘以总样本量得到期望观测值，进而计算  $\chi^2$  统计量和显著性概率。最后一组的期望样本量很低，与前一组合并计算， $\chi^2$  统计量只针对 10 组样本，自由度为 7。得到的  $\chi^2$  统计量等于 29.4571，差异显著性概率等于 0.001，低于 0.05。因此样本群体与正态分布之间存在显著的差异。

分组区间	组中值	人数	观测频率	正态分布频率	期望观察值	$(O-E)^2/E$
53-55	54	5	0.0010	0.0013	6.44	0.3208
55-57	56	33	0.0066	0.0108	54.17	8.2708
57-59	58	254	0.0509	0.0537	268.05	0.7362
59-61	60	813	0.1628	0.1565	781.48	1.2711
61-63	62	1340	0.2683	0.2691	1344.08	0.0124
63-65	64	1454	0.2911	0.2732	1364.73	5.8400
65-67	66	750	0.1502	0.1638	818.06	5.6631
67-69	68	275	0.0551	0.0579	289.30	0.7073
69-71	70	56	0.0112	0.0121	60.28	0.3041
71-73	72	11	0.0022	0.0015	7.39	6.3313
73-75	74	4	0.0008	0.0001	0.53	
$\chi^2$ 统计量						29.4571
显著性概率 $P$						0.001

(3) 如果样本量只有原来的 10%，这组数据是否服从正态分布？

**解答：**把群体看作均值等于 63.06、方差等于 7.25 的正态分布，但样本量只有 (2) 的 10%。前两组合并，最后三组合并，得到的  $\chi^2$  统计量等于 2.2775，是 (2) 统计量的约 0.1

倍，自由度为 5。差异显著性概率等于 0.8096，远高于 0.05。因此样本群体与正态分布之间不存在显著的差异。

组中值	观测频率	期望观察值	$(O-E)^2/E$
54	0.5	0.64	
56	3.3	5.42	0.8430
58	25.4	26.80	0.0736
60	81.3	78.15	0.1271
62	134	134.41	0.0012
64	145.4	136.47	0.5840
66	75	81.81	0.5663
68	27.5	28.93	0.0707
70	5.6	6.03	0.0115
72	1.1	0.74	
74	0.4	0.05	
$\chi^2$ 统计量			2.2775
显著性概率 $P$			0.8096

(4) 从 (2) 和 (3) 得到检验结果的差异，说明统计假设检验时“接受零假设”和“拒绝零假设”时应该注意的地方。

**解答：**接收零假设，并不等价于备择假设完全没有可能，只是说明备择假设与零假设之间的差异完全可以是随机抽样误差所引起。拒绝零假设，说明随机抽样误差不是备择假设与零假设之间差异的唯一原因，可能还有其它原因。假设检验针对于有限样本；样本量无穷大时，即使细微的差异都可能是显著的。理论上讲，只要样本量足够大，任何样本群体与理论分布之间的差异都可能达到显著或极显著水平。严格意义上的正态分布是不存在的。但是，对于众多实际群体来说，这无碍于把它们视为正态分布进行参数估计和统计推断。

(5) 在正态总体的假定下，计算总体均值的 95% 和 99% 置信区间。

**解答：**总体均值的 95% 置信区间是 (57.78, 68.33)，99% 置信区间是 (56.12, 69.99)。

**6.2** 图6.2中，亲本 $P_1$ 的方差3.03是表6.2中的三个 $P_1$ 合并起来计算的， $P_2$ 的方差5.64是表6.2中的三个 $P_2$ 合并起来计算的， $F_1$ 的方差8.62根据表6.2中的 $F_1$ 计算。这三个方差的自由度分别是多大？假定这三个群体中的变异都来自随机误差 $\sigma_\varepsilon^2$ ，如用下面的两种方法估计误差方

差  $\sigma_{\varepsilon}^2$ : (1) 将这三个方差相加后除以3; (2) 将这三个方差分别乘以它们的自由度后相加, 然后除以总自由度 (即三个自由度之和)。这两种估计方法你觉得哪一个更好些? 为什么?

**解答:** 三个群体的样本量和自由度如下表。

群体	方差	样本量	自由度
P <sub>1</sub>	3.03	211	210
P <sub>2</sub>	5.64	173	172
F <sub>1</sub>	8.62	168	167
方差的平均	5.7633		
合并方差	5.5481		549

两种方法得到的随机误差  $\sigma_{\varepsilon}^2$  估计值分别为 5.7633 和 5.5481。两种方法得到的估计值在三个群体自由度相同的情况下是相等的。如三个群体自由度不相同, 第 2 种方法更好些。如不做无偏性矫正, 在三个群体误差方差相等的假定下, 第 2 种方法给出的是误差方差的极大似然估计。

**6.3** 下表是Johannsen的菜豆试验中, 第13个家系4种粒重种子的后代平均粒重。

亲代粒重X/cg	20	30	40	50
子代平均粒重Y/cg	47.5	45.0	45.1	45.8

(1) 计算亲代粒重  $X$  与子代平均粒重  $Y$  的方差、以及它们之间的协方差。

**解答:** 利用 EXCEL 中的样本方差函数 VAR, 得到亲代和子代的方差分别为 166.67 和 1.34。利用样本协方差函数 COVARIANCE, 得到亲代和子代之间的协方差为-8.33。

(2) 计算亲代粒重  $X$  与子代平均粒重  $Y$  的相关系数, 并进行显著性检验。

**解答:** 利用 EXCEL 中的相关系数函数 CORREL, 得到亲代和子代之间的相关系数为 -0.5583。根据公式 6.73 得到的  $t$  值为-0.9517、显著性概率为 0.4417, 因此相关系数与 0 之间不存在显著差异。

(3) 将亲代粒重  $X$  视为自变量, 子代平均粒重  $Y$  视为因变量, 建立二者的线性回归关系, 并对回归方程和回归系数进行显著性检验。

**解答:** 利用 EXCEL 中‘数据分析’模块中的‘回归’, 得到的回归方程为  $Y=47.6-0.05X$ 。回归方程的方差分析如下表。显著性检验表明回归系数与 0 之间没有显著差异, 检验结果与(2)利用相关系数的检验结果是一致的。

变异来源	自由度	平方和	均方	$F$ 值	显著概率
回归分析	1	1.25	1.25	0.9058	0.4417
残差	2	2.76	1.38		
总计	3	4.01			

**6.4** 假定某性状是由三对独立遗传、等效加性基因控制的，亲本 $P_1$ 的基因型是 $AABBCC$ ，性状均值为24；亲本 $P_2$ 的基因型是 $aabbcc$ ，性状均值为12。

(1) 不考虑随机误差，亲本  $P_1$  和  $P_2$  杂交  $F_1$  群体的表型是多少？

**解答：** 不考虑随机误差，杂交  $F_1$  群体的表型等于 18。在存在随机误差的情况下，18 可以看作是杂交  $F_1$  群体的均值。

(2) 不考虑随机误差，亲本  $P_1$  和  $P_2$  杂交  $F_2$  群体中有多少种不同的表型？给出这些表型值和它们的存在频率，并计算  $F_2$  群体的方差。

**解答：** 亲本  $P_1$  的基因型  $AABBCC$  中包含 6 个增效基因，性状均值为 24；亲本  $P_2$  基因型  $aabbcc$  包含 0 个增效基因，性状均值为 12。因此，单个增效基因的效应等于 2。如用  $X$  表示增效基因的个数，在  $F_2$  群体中的可能取值是 0~6，服从二项分布  $B(n=6, p=0.5)$ ，基因型值等于  $12+2X$ 。因此得到  $F_2$  群体 7 种基因型值和频率，以及  $F_2$  的群体方差（见下表）。

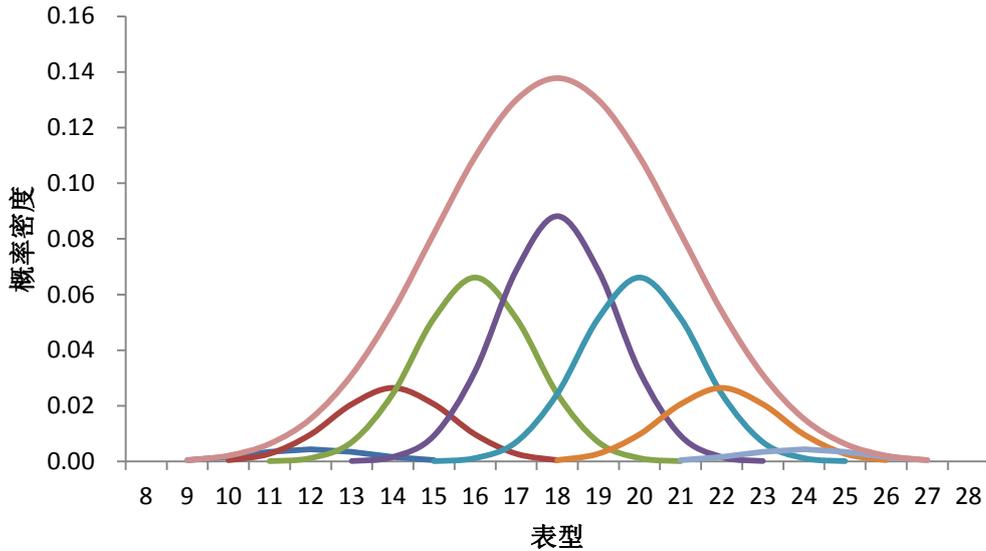
增效基因个数 $X$	基因型值	$F_2$ 群体中的频率
0	12	0.015625
1	14	0.09375
2	16	0.234375
3	18	0.3125
4	20	0.234375
5	22	0.09375
6	24	0.015625
群体均值		18
群体方差		6

(3) 如随机误差为 2，绘制  $F_2$  群体理论分布函数的曲线图。

**解答：**  $F_2$  群体由 7 个正态分布构成，(2) 给出了这些分布的均值和在  $F_2$  群体中的存在比例，它们的方差均为 2。利用 EXCEL 中的正态分布函数 NORM.DIST 计算表型在 9~27

之间的概率密度函数，加权平均数即为  $F_2$  群体的概率密度。进而绘制出  $F_2$  群体理论分布函数的曲线图。

表型	增效基因个数对应的成分分布							$F_2$ 群体
	0	1	2	3	4	5	6	
9	0.0005							0.0005
10	0.0016	0.0005						0.0021
11	0.0034	0.0028	0.0001					0.0063
12	0.0044	0.0097	0.0012					0.0153
13	0.0034	0.0206	0.0070	0.0002				0.0312
14	0.0016	0.0264	0.0243	0.0016				0.0540
15	0.0005	0.0206	0.0515	0.0093	0.0001			0.0820
16		0.0097	0.0661	0.0324	0.0012			0.1095
17		0.0028	0.0515	0.0687	0.0070			0.1299
18		0.0005	0.0243	0.0882	0.0243	0.0005		0.1378
19			0.0070	0.0687	0.0515	0.0028		0.1299
20			0.0012	0.0324	0.0661	0.0097		0.1095
21			0.0001	0.0093	0.0515	0.0206	0.0005	0.0820
22				0.0016	0.0243	0.0264	0.0016	0.0540
23				0.0002	0.0070	0.0206	0.0034	0.0312
24					0.0012	0.0097	0.0044	0.0153
25					0.0001	0.0028	0.0034	0.0063
26						0.0005	0.0016	0.0021
27							0.0005	0.0005



**6.5** 如何理解数量性状的多基因假说？它如何统一生物统计学派和孟德尔学派关于数量性状遗传所产生的分歧？

**解答：**参见‘§6.1.3 数量性状遗传的多基因假说’一节。

**6.6** 若等位基因  $A$  相对于  $a$  为显性，两个亲本的基因型分别为  $AA$  和  $aa$ ，那么  $F_2$  群体中显性个体的基因型既可能为  $AA$  也可能为  $Aa$ ，遗传上常用  $F_2$  个体产生的  $F_3$  家系中是否有显隐性的分离来判断一个显性  $F_2$  个体的基因型是  $AA$  或是  $Aa$ 。

(1) 如果每个  $F_3$  家系仅种植 5 个单株，计算把  $Aa$  基因型误判成  $AA$  的概率是多大？

**解答：**基因型  $Aa$  的自交后代群体中，单个个体出现显性表型的概率等于 0.75。5 个个个体均为显性表型的概率等于  $0.75^5=0.2373$ ，这时就误以为亲本的基因型是  $AA$ 。因此误判概率等于 0.2373。

(2) 如果要保证基因型  $Aa$  被误判成  $AA$  的概率低于 0.05， $F_3$  家系至少要种植多少个单株？

**解答：**用  $n$  表示  $F_3$  家系的大小，整个家系均为显性表型的概率等于  $0.75^n$ 。要求误判概率低于 0.05，其实也就是要求等于  $0.75^n < 0.05$ 。因此得到  $n > 10.41$ 。

(3) 如果要保证误判概率低于 0.01， $F_3$  家系至少要种植多少个单株？

**解答：**用  $n$  表示  $F_3$  家系的大小，整个家系均为显性表型的概率等于  $0.75^n$ 。要求误判概率低于 0.01，其实也就是要求等于  $0.75^n < 0.01$ 。因此得到  $n > 16.01$ 。

**6.7** 随机变量  $X$  服从两点分布（也称 Bernoulli 分布），取 0 的概率为  $1-p$ ，取 1 的概率为  $p$ 。随机变量  $Y$  服从二项分布  $B(n, p)$ ，可以看作  $n$  个独立两点分布的随机变量之和。

(1) 计算  $X$  的均值和方差。

**解答:** 根据两点分布的概率分布, 以及均值和方差的定义, 可以得到

$$E(X) = 0 \times (1-p) + 1 \times p = p$$

$$V(X) = 0^2 \times (1-p) + 1^2 \times p - [E(X)]^2 = p - p^2 = p(1-p)$$

(2) 计算  $Y$  的均值和方差。

**解答:** 设  $X_1, X_2, \dots, X_n$  是  $n$  个独立并具有相同概率分布的两点分布, 则  $Y = X_1 + X_2 + \dots + X_n$ 。

根据期望和方差的性质, 可以得到

$$E(Y) = E(X_1) + E(X_2) + \dots + E(X_n) = np$$

$$V(Y) = V(X_1) + V(X_2) + \dots + V(X_n) = np(1-p)$$

**6.8** 随机变量  $X$  服从区间  $(a, b)$  上的均匀分布, 概率密度函数如下, 计算  $X$  的均值和方差。

$$p(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{其他} \end{cases}$$

**解答:** 根据均匀分布的概率分布, 以及均值和方差的定义, 可以得到

$$E(X) = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \left( \frac{1}{2} x^2 \right) \Big|_a^b = \frac{1}{b-a} \frac{1}{2} (b^2 - a^2) = \frac{1}{2} (b+a)$$

$$\begin{aligned} V(X) &= \int_a^b x^2 \frac{1}{b-a} dx - [E(X)]^2 = \frac{1}{b-a} \left( \frac{1}{3} x^3 \right) \Big|_a^b - \frac{1}{4} (b+a)^2 \\ &= \frac{1}{3} (a^2 + ab + b^2) - \frac{1}{4} (b+a)^2 = \frac{1}{12} (b-a)^2 \end{aligned}$$

当然也可以从区间  $(0, 1)$  上均匀分布的均值和方差进行计算。如  $U$  是  $(0, 1)$  上均匀分布, 则线性变换  $(b-a)U + a$  是区间  $(a, b)$  上的均匀分布,  $E(U) = \frac{1}{2}$ ,  $V(U) = \frac{1}{12}$ 。因此,

$$E(X) = (b-a)E(U) = \frac{1}{2} (b-a)$$

$$V(X) = (b-a)^2 V(U) = \frac{1}{12} (b-a)^2$$

**6.9** 两点分布总体  $X$  的概率函数可以用  $p(x) = (1-p)^{1-x} p^x$  表示, 其中  $x$  等于 0 或 1。

设  $X_1, X_2, \dots, X_n$  是一组简单随机样本。计算总体参数  $p$  的 MLE。

**解答：**记  $X_{\cdot} = \sum_{i=1}^n X_i$ ，则对数似然函数为，

$$\begin{aligned} L(p) &= \ln \prod_{i=1}^n (1-p)^{1-X_i} p^{X_i} = \sum_{i=1}^n [(1-X_i)\ln(1-p) + X_i \ln p] \\ &= (n - X_{\cdot})\ln(1-p) + (X_{\cdot})\ln p \end{aligned}$$

对分布参数  $p$  求对数得到，

$$\frac{dL(p)}{dp} = (n - X_{\cdot})\frac{-1}{1-p} + (X_{\cdot})\frac{1}{p} = \frac{-np + X_{\cdot}}{p(1-p)}$$

令对数函数等于 0，得到  $p$  的 MLE 为，

$$\hat{p} = \frac{X_{\cdot}}{n} = \bar{X}$$

因此，总体参数  $p$  的 MLE 等于样本均值。

**6.10** 假定变量  $X$  与  $Y$  有线性关系，大小为  $n$  的样本用模型  $Y_i = a + bX_i + \varepsilon_i$  表示， $i=1, 2, \dots, n$  表示样本。用  $\bar{X}$  和  $\bar{Y}$  表示样本均值，对样本进行中心化变换，即令  $X_i^* = X_i - \bar{X}$ ， $Y_i^* = Y_i - \bar{Y}$ 。中心化变换后得到的模型中只有回归系数，即  $Y_i^* = bX_i^* + \varepsilon_i$ 。计算中心化变换后模型中回归系数的 LSE。

**解答：**中心化变换不改变两个变量的方差和协方差，模型  $Y_i^* = bX_i^* + \varepsilon_i$  中回归系数的与公式 6.71 完全相同。

**6.11** 在  $n$  个不同基因型构成的遗传群体中，繁殖后的基因型保持不变，基因型  $i$  的均值用  $G_i$  表示，它们之间的方差用  $\sigma_G^2$  表示。假定每个亲代基因型只繁殖一个后代个体，亲代和子代的表型  $X_i$  和  $Y_i$  分别用线性模型  $X_i = G_i + \varepsilon_i$  和  $Y_i = G_i + \delta_i$  表示， $\varepsilon_i, \delta_i \sim N(0, \sigma_\varepsilon^2)$  且相互独立。 $\sigma_p^2 = \sigma_G^2 + \sigma_\varepsilon^2$  表示表型方差。计算子代表型对亲代表型的回归系数，以及它们之间的相关系数。

**解答：**表型  $X_i$  和  $Y_i$  的方差均等于遗传方差  $\sigma_G^2$  加上误差方差，即，

$$V(X) = V(Y) = \sigma_G^2 + \sigma_\varepsilon^2 = \sigma_p^2$$

表型中的遗传效应和误差效应相互独立，亲代和子代表型中的误差效应也相互独立，即，

$$\text{Cov}(G_i, G_i) = \sigma_G^2, \quad \text{Cov}(G_i, \delta_i) = \text{Cov}(G_i, \varepsilon_i) = \text{Cov}(\varepsilon_i, \delta_i) = 0$$

因此，表型  $X_i$  和  $Y_i$  之间的协方差等于遗传方差  $\sigma_G^2$ ，即，

$$\text{Cov}(X, Y) = \text{Cov}(G_i, G_i) + \text{Cov}(G_i, \delta_i) + \text{Cov}(G_i, \varepsilon_i) + \text{Cov}(\varepsilon_i, \delta_i) = \sigma_G^2$$

根据公式 6.71 和公式 6.72，立即得到回归系数和相关系数均等于  $\frac{\sigma_G^2}{\sigma_P^2}$ ，即均等于广义遗传

力  $H^2$ 。

**6.12** 在一个随机交配群体中，等位基因  $A$  和  $a$  的频率为  $p$  和  $1-p$ 。对  $N$  个个体的调查发现，三种基因型  $AA$ 、 $Aa$ 、 $aa$  的个体数分别为  $D$ 、 $H$ 、 $R$ 。证明  $\frac{1}{N}(D + \frac{1}{2}H)$  是等位基因频率  $p$  的 MLE [提示：基因型的观察值  $D$ 、 $H$ 、 $R$  服从频率分别为  $p^2$ 、 $2p(1-p)$ 、 $(1-p)^2$  的多项分布]。

**解答：**根据 HW 平衡定律，三种基因型  $AA$ 、 $Aa$ 、 $aa$  的频率分别为  $p^2$ 、 $2p(1-p)$ 、 $(1-p)^2$ 。它们的观测个体数  $D$ 、 $H$ 、 $R$  服从多项分布。根据公式 3.10，多项分布的概率分布为，

$$P(D, H, R) = \frac{N!}{D!H!R!} [p^2]^D [2p(1-p)]^H [q^2]^R$$

$$P(D, H, R) = \frac{2^H N!}{D!H!R!} p^{2D+H} (1-p)^{H+2R}$$

对数似然函数为，

$$L(p) = \ln P(D, H, R) = C + (2D + H) \ln p + (H + 2R) \ln(1-p)$$

对数似然函数的导数为，

$$\frac{dL(p)}{dp} = \frac{(2D + H) - p(2N)}{p(1-p)}, \quad \text{其中 } N=D+H+R$$

令对数函数等于 0，得到  $p$  的 MLE 为，

$$\hat{p} = \frac{2D + H}{2N} = \frac{1}{N} \left( D + \frac{1}{2} H \right)$$

也就是说， $\frac{1}{N} \left( D + \frac{1}{2} H \right)$  是等位基因频率  $p$  的 MLE。

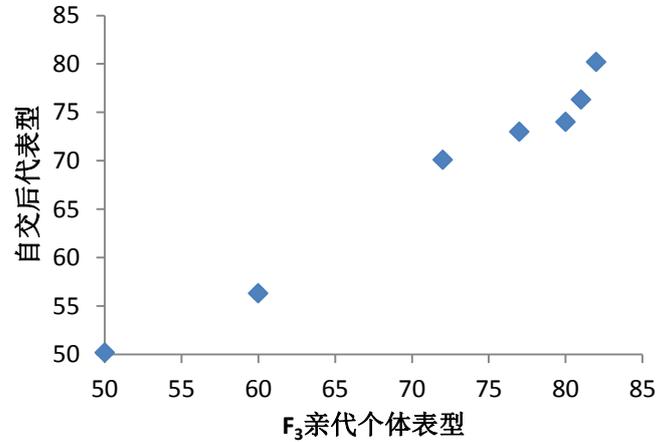
**6.13** 下表是烟草杂交试验中，8 个  $F_3$  个体的花冠长度 ( $X$ ) 及它们自交后代的平均花冠长度 ( $Y$ ) 数据。

$F_3$ 个体花冠长度 ( $X$ )	46	50	60	72	77	80	81	82
----------------------	----	----	----	----	----	----	----	----

后代花冠长度平均数 (Y)	53.5	50.2	56.3	70.1	73.0	74.0	76.3	80.2
---------------	------	------	------	------	------	------	------	------

(1) 绘制  $X$  与  $Y$  的散点图。

**解答：**在 EXCEL 中绘制下面的散点图。



(2) 计算  $X$  与  $Y$  的相关系数，并进行显著性检验。

**解答：**利用 EXCEL 中的相关系数函数 CORREL，得到亲代和子代之间的相关系数为 0.9731。根据公式 6.73 得到的  $t$  值为 10.35、显著性概率为  $4.77 \times 10^{-5}$ ，因此相关系数与 0 之间存在极显著的差异。

(3) 将  $X$  视为自变量，将  $Y$  视为因变量，建立二者的线性关系，并对回归方程和回归系数进行显著性检验。

**解答：**利用 EXCEL 中‘数据分析’模块中的‘回归’，得到的回归方程为  $Y=13.69-0.7739X$ 。回归方程的方差分析如下表。显著性检验表明回归系数与 0 之间存在极显著差异，检验结果与 (2) 利用相关系数的检验结果是一致的。

变异来源	自由度	平方和	均方	$F$ 值	显著概率
回归分析	1	884.04	884.04	107.04	4.77E-05
残差	6	49.56	8.26		
总计	7	933.6			

## 第 7 章练习题和参考答案

**7.1** 下表是20个玉米自交系与同一个测验种的杂交试验中，每个杂种F<sub>1</sub>群体3个单株的株高（cm）调查数据。给出这组数据的方差分析表；并利用方差分析的结果，估计株高性状的误差方差和遗传方差；在方差估计的基础上，估计单株株高的广义遗传力、重复平均数的广义遗传力。

自交系	植株 1	植株 2	植株 3	自交系	植株 1	植株 2	植株 3
L01	305	280	290	L11	260	280	285
L02	280	305	315	L12	310	310	310
L03	270	255	270	L13	310	335	330
L04	270	260	290	L14	335	310	330
L05	310	315	325	L15	260	245	255
L06	290	290	300	L16	320	305	300
L07	255	235	235	L17	300	298	278
L08	287	300	310	L18	310	305	278
L09	320	320	275	L19	295	270	275
L10	270	265	260	L20	290	300	280

**解答：**可以利用 EXCEL‘数据分析’模块中的‘单因素方差分析’，也可根据方差分析的线性模型（公式 7.8）直接在 EXCEL 中进行计算。得到的方差分析表如下，基因型间的差异达到极显著水平。

变异来源	自由度	平方和	均方	F 值	P 值
基因型间	19	29732.32	1564.86	9.60	1.44E-09
随机误差	40	6521.33	163.03		
总计	59	36253.65			

在上表均方的基础上，根据公式 7.18 和公式 7.19 计算误差方差和基因型方差的估计值，进而根据公式 7.24 和公式 7.25 计算单株和重复平均数的遗传力。结果如下表。

遗传参数	遗传方差	误差方差	单株遗传力	重复平均数遗传力
估计值	467.28	163.03	0.7413	0.8958

**7.2** 利用表 7.3 环境 I 的数据，假定两个重复安排在环境条件相对一致的两个区组中。给出表型数据的方差分析表；并利用方差分析的结果，估计直链淀粉含量在环境 I 的误差方差和遗传方差；在方差估计的基础上，估计直链淀粉含量小区水平的广义遗传力、重复平均数的广义遗传力。

**解答：**可以利用 EXCEL‘数据分析’模块中的‘无重复双因素方差分析’，也可根据双因素

无互作的方差分析线性模型直接在 EXCEL 中进行计算。得到的方差分析表如下，基因型间和区组间的差异均达到极显著水平。

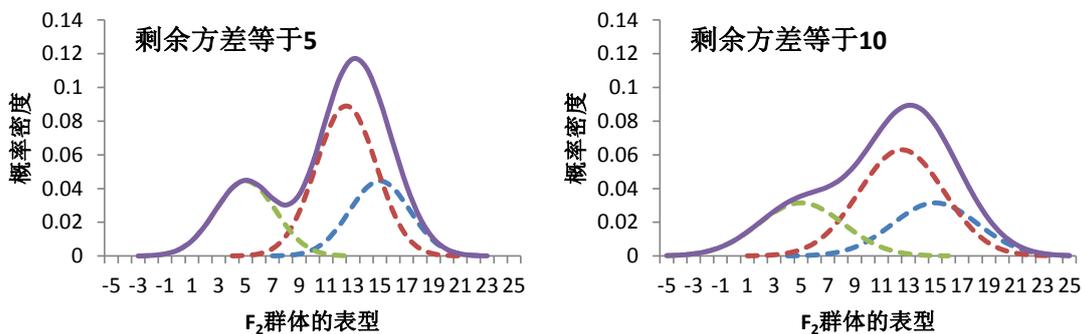
变异来源	自由度	平方和	均方	F 值	P 值
基因型间	9	23.6545	2.6283	40.06	3.39E-06
区组间	1	1.6245	1.6245	24.76	0.000764
随机误差	9	0.5905	0.0656		
总计	19	25.8695			

在上表均方的基础上，仍然可以利用公式 7.18 和公式 7.19 计算误差方差和基因型方差的估计值，进而根据公式 7.24 和公式 7.25 计算单株和重复平均数的遗传力。结果如下表。

遗传参数	遗传方差	误差方差	单株遗传力	重复平均数遗传力
估计值	1.2813	0.0656	0.9513	0.9750

**7.3** 在单基因座位加显性模型下，两个亲本的基因型分别为 AA 和 aa，该座位上群体平均数、加性效应和显性效应分别为  $m=10$ 、 $a=5$ 、 $d=0$ ，其它效应（包括遗传和误差）服从均值为 0、方差为  $\sigma^2$  的正态分布。当  $\sigma^2=5$  或 10 时，绘制  $F_2$  群体的理论分布曲线图。

**解答：**三种基因型 AA、Aa、aa 服从均值分别为  $m+a$ 、 $m+d$ 、 $m-a$ 、方差为  $\sigma^2$  的正态分布，在  $F_2$  群体中的比例分别为 0.25、0.5、0.25。因此， $F_2$  群体可以看作三种基因型 AA、Aa、aa 按比例 0.25、0.5、0.25 构成的混合分布。利用 EXCEL 中的正态分布函数 NORM.DIST，计算表型在 -5~25 之间的三种基因型的概率密度函数，加权平均即为  $F_2$  群体的概率密度函数。当  $\sigma^2=5$  或 10 时， $F_2$  群体的理论分布曲线图如下。



**7.4** 下表是 East (1911) 玉米穗长 (cm) 的杂交试验中亲本、 $F_1$  和  $F_2$  世代的次数分布。利用 Castle-Wright 公式计算有效因子个数。

群体	穗长的组中值/cm																				
	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21				

P <sub>1</sub>	4	21	24	8												
F <sub>1</sub>				1	12	12	14	17	9	4						
P <sub>2</sub>								3	11	12	15	26	15	10	7	2
F <sub>2</sub>		4	5	22	56	80	145	129	91	63	27	17	6	1		

**解答：**首先在 EXCEL 中计算分组数据的频率，利用积和函数 SUMPRODUCT 计算每个群体的均值和方差，利用 P<sub>1</sub>、F<sub>1</sub> 和 P<sub>2</sub> 的方差计算合并方差作为误差方差的估计值。最后利用公式 7.61 计算有效因子数，结果如下表。

群体	均值	方差
P <sub>1</sub>	6.6316	0.6537
F <sub>1</sub>	12.1159	2.2764
P <sub>2</sub>	16.8020	3.5251
F <sub>2</sub>	12.6780	3.9737
合并误差方差		2.4246
有效因子 <i>k</i>		8.3462

**7.5** 在单基因座位加显性模型下，群体平均数中亲值、加性效应和显性效应分别用 *m*、*a* 和 *d* 表示，2 个亲本的基因型分别为 AA 和 aa，计算回交群体 B<sub>1</sub> 和 B<sub>2</sub> 的群体平均数和遗传方差。

**解答：**下表左半部分给出三种基因型 AA、Aa、aa 在回交群体 B<sub>1</sub> 和 B<sub>2</sub> 中的频率。它们的均值和方差的计算如下：

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

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

$$V_{B_1} = \frac{1}{2}a^2 + \frac{1}{2}d^2 - \left[\frac{1}{2}(a+d)\right]^2 = \frac{1}{4}(a-d)^2$$

$$V_{B_2} = \frac{1}{2}d^2 + \frac{1}{2}a^2 - \left[\frac{1}{2}(a-d)\right]^2 = \frac{1}{4}(a+d)^2$$

群体	AA	Aa	aa	均值 ( $\bar{X}$ )	遗传方差 ( $V_x$ )
	$\frac{m+a}{2}$	$\frac{m+d}{2}$	$\frac{m-a}{2}$		
B <sub>1</sub>	$\frac{1}{2}$	$\frac{1}{2}$	0	$m + \frac{1}{2}(a+d)$	$\frac{1}{4}(a-d)^2$
B <sub>2</sub>	0	$\frac{1}{2}$	$\frac{1}{2}$	$m - \frac{1}{2}(a-d)$	$\frac{1}{4}(a+d)^2$

7.6 在单基因座位加显性模型下, 中亲值、加性效应和显性效应分别用  $m$ 、 $a$  和  $d$  表示, 2 个亲本的基因型分别为  $AA$  和  $aa$ 。

(1) 计算  $F_2$  群体和  $F_3$  混合群体的均值和方差。

**解答:** 下表左半部分给出三种基因型  $AA$ 、 $Aa$ 、 $aa$  在  $F_2$  群体和  $F_3$  混合群体中的频率。它们的均值和方差的计算如下:

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

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

$$V_{F_2} = \frac{1}{4}a^2 + \frac{1}{2}d^2 + \frac{1}{4}a^2 - \left[\frac{1}{2}d\right]^2 = \frac{1}{2}a^2 + \frac{1}{4}d^2$$

$$V_{F_3} = \frac{3}{8}a^2 + \frac{1}{4}d^2 + \frac{3}{8}a^2 - \left[\frac{1}{4}d\right]^2 = \frac{3}{4}a^2 + \frac{3}{16}d^2$$

群体	$AA$	$Aa$	$aa$	均值 ( $\bar{X}$ )	遗传方差 ( $V_x$ )
	$m+a$	$m+d$	$m-a$		
$F_2$ 群体	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$m + \frac{1}{2}d$	$\frac{1}{2}a^2 + \frac{1}{4}d^2$
$F_3$ 混合群体	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{3}{8}$	$m + \frac{1}{4}d$	$\frac{3}{4}a^2 + \frac{3}{16}d^2$

(2) 计算  $F_{2:3}$  家系平均数的方差和家系的平均方差。

**解答:**  $F_{2:3}$  家系群体中, 三种基因型的频率等于它们在  $F_2$  群体中的频率。基因型  $AA$  的自交家系中, 基因型全部为  $AA$ , 家系平均数等于  $m+a$ , 家系内的遗传方差为 0。基因型  $Aa$  的自交家系中, 三种基因型的频率等于它们在  $F_2$  群体中的频率, 家系平均数等于  $m + \frac{1}{2}d$ , 家系内的遗传方差为  $\frac{1}{2}a^2 + \frac{1}{4}d^2$ 。基因型  $aa$  的自交家系中, 基因型全部为  $aa$ , 因此家系平均数等于  $m-a$ , 家系内的遗传方差为 0。将这些结果列于下表, 并计算  $F_{2:3}$  家系平均数的方差以及家系内的平均方差, 即,

$$V_{bF_{2:3}} = \frac{1}{4}a^2 + \frac{1}{2}\left(\frac{1}{2}d\right)^2 + \frac{1}{4}(-a)^2 - \left(\frac{1}{4}d\right)^2 = \frac{1}{2}a^2 + \frac{1}{16}d^2$$

$$V_{wF_{2:3}} = \frac{1}{4} \times 0 + \frac{1}{2}V_{F_2} + \frac{1}{4} \times 0 = \frac{1}{4}a^2 + \frac{1}{8}d^2$$

基因型	$AA$	$Aa$	$aa$	家系间/家系内方差
$F_2$ 群体中的频率	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	

F <sub>2:3</sub> 家系平均数	$m+a$	$m+\frac{1}{2}d$	$m-a$	$\frac{1}{2}a^2 + \frac{1}{16}d^2$
F <sub>2:3</sub> 家系内方差	0	$\frac{1}{2}a^2 + \frac{1}{4}d^2$	0	$\frac{1}{4}a^2 + \frac{1}{8}d^2$

(3) 计算 F<sub>2</sub> 群体中个体与其 F<sub>3</sub> 家系平均数之间的协方差。

**解答：**将 F<sub>2</sub> 群体和 F<sub>2:3</sub> 家系中三种基因型的频率、基因型值和家系平均数列于下表。

F<sub>2</sub> 群体的均值  $m + \frac{1}{2}d$ ，F<sub>2:3</sub> 家系群体的均值与 F<sub>3</sub> 混合群体相同，即等于  $m + \frac{1}{4}d$ 。协方差的

的计算过程如下：

$$\begin{aligned} \text{Cov}_{F_{2,3}} &= \frac{1}{4} \times a \times a + \frac{1}{2} \times d \times \left(\frac{1}{2}d\right) + \frac{1}{4} \times (-a) \times (-a) - \left(\frac{1}{2}d\right) \times \left(\frac{1}{4}d\right) \\ &= \frac{1}{2}a^2 + \frac{1}{8}d^2 \end{aligned}$$

基因型	AA	Aa	aa	均值
F <sub>2</sub> 群体中的频率	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	
F <sub>2</sub> 群体的基因型值	$m+a$	$m+d$	$m-a$	$m + \frac{1}{2}d$
F <sub>2:3</sub> 家系的均值	$m+a$	$m + \frac{1}{2}d$	$m-a$	$m + \frac{1}{4}d$
F <sub>2</sub> 与 F <sub>2:3</sub> 家系的协方差	$\frac{1}{2}a^2 + \frac{1}{8}d^2$			

**7.7** 在单基因座位加显性模型下，中亲值、加性效应和显性效应分别用  $m$ 、 $a$  和  $d$  表示，

2 个亲本的基因型分别为 AA 和 aa。

(1) 计算 F<sub>4</sub> 混合群体的均值和方差。

**解答：**下表左半部分给出三种基因型 AA、Aa、aa 在 F<sub>4</sub> 混合群体中的频率。均值和方差的计算如下：

$$\bar{F}_4 = \frac{7}{16}(m+a) + \frac{1}{8}(m+d) + \frac{7}{16}(m-a) = m + \frac{1}{8}d$$

$$V_{F_4} = \frac{7}{16}a^2 + \frac{1}{8}d^2 + \frac{7}{16}a^2 - \left(\frac{1}{8}d\right)^2 = \frac{7}{8}a^2 + \frac{7}{64}d^2$$

群体	AA	Aa	aa	均值 ( $\bar{X}$ )	遗传方差 ( $V_X$ )
	$m+a$	$m+d$	$m-a$		

F <sub>4</sub> 混合群体	$\frac{7}{16}$	$\frac{1}{8}$	$\frac{7}{16}$	$m + \frac{1}{8}d$	$\frac{7}{8}a^2 + \frac{7}{64}d^2$
---------------------	----------------	---------------	----------------	--------------------	------------------------------------

(2) 计算 F<sub>2,4</sub> 家系平均数的方差和家系的平均方差。

**解答:** F<sub>2,4</sub> 家系群体中, 三种基因型的频率等于它们在 F<sub>2</sub> 群体中的频率。基因型 AA 的自交家系中, 基因型全部为 AA, 家系平均数等于  $m+a$ , 家系内的遗传方差为 0。基因型 Aa 的自交家系中, 三种基因型的频率等于它们在 F<sub>3</sub> 群体中的频率, 家系平均数等于  $m + \frac{1}{4}d$ , 家系内的遗传方差为  $\frac{3}{4}a^2 + \frac{3}{16}d^2$ 。基因型 aa 的自交家系中, 基因型全部为 aa, 因此家系平均数等于  $m-a$ , 家系内的遗传方差为 0。将这些结果列于下表, 并计算 F<sub>2,3</sub> 家系平均数的方差以及家系内的平均方差, 即,

$$V_{bF_{2,4}} = \frac{1}{4}a^2 + \frac{1}{2}\left(\frac{1}{4}d\right)^2 + \frac{1}{4}(-a)^2 - \left(\frac{1}{8}d\right)^2 = \frac{1}{2}a^2 + \frac{1}{64}d^2$$

$$V_{wF_{2,4}} = \frac{1}{4} \times 0 + \frac{1}{2}V_{F_3} + \frac{1}{4} \times 0 = \frac{3}{8}a^2 + \frac{3}{32}d^2$$

显然, 这两种方差之和就等于 (1) 中 F<sub>4</sub> 混合群体的方差。

基因型	AA	Aa	aa	家系间/家系内方差
F <sub>2</sub> 群体中的频率	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	
F <sub>2,4</sub> 家系平均数	$m+a$	$m + \frac{1}{4}d$	$m-a$	$\frac{1}{2}a^2 + \frac{1}{64}d^2$
F <sub>2,4</sub> 家系内方差	0	$\frac{3}{4}a^2 + \frac{3}{16}d^2$	0	$\frac{3}{8}a^2 + \frac{3}{32}d^2$
F <sub>2</sub> 与 F <sub>2,3</sub> 家系的协方差	$\frac{1}{2}a^2 + \frac{1}{8}d^2$			

(3) 计算 F<sub>2</sub> 群体中个体与其 F<sub>4</sub> 家系平均数之间的协方差。

**解答:** 将 F<sub>2</sub> 群体和 F<sub>2,4</sub> 家系中三种基因型的频率、基因型值和家系平均数列于下表。

F<sub>2</sub> 群体的均值  $m + \frac{1}{2}d$ , F<sub>2,4</sub> 家系群体的均值与 F<sub>4</sub> 混合群体相同, 即等于  $m + \frac{1}{8}d$ 。协方差

计算过程如下:

$$\begin{aligned} Cov_{F_{2,3}} &= \frac{1}{4} \times a \times a + \frac{1}{2} \times d \times \left(\frac{1}{4}d\right) + \frac{1}{4} \times (-a) \times (-a) - \left(\frac{1}{2}d\right) \times \left(\frac{1}{8}d\right) \\ &= \frac{1}{2}a^2 + \frac{1}{16}d^2 \end{aligned}$$

基因型	AA	Aa	aa	均值
-----	----	----	----	----

F <sub>2</sub> 群体中的频率	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	
F <sub>2</sub> 群体的基因型值	$m+a$	$m+d$	$m-a$	$m + \frac{1}{2}d$
F <sub>2:4</sub> 家系的均值	$m+a$	$m + \frac{1}{4}d$	$m-a$	$m + \frac{1}{8}d$
F <sub>2</sub> 与 F <sub>2:4</sub> 家系的协方差	$\frac{1}{2}a^2 + \frac{1}{16}d^2$			

7.8 假定两个座位上 9 种基因型的平均表现有下表所示的 4 种情形。对每种情形，计算座位 A 和 B 的加显性效应，以及两座位间的上位性互作效应。

基因型	情形 I	情形 II	情形 III	情形 IV
AABB	6	5	5	5
AABb	5	5	5	8
AAbb	4	4	5	7
AaBB	5	5	5	6
AaBb	4	5	5	9
Aabb	3	4	5	4
aaBB	4	4	5	3
aaBb	3	4	5	2
aabb	2	3	1	1

解答：根据表 7.12，9 种基因型值与遗传参数的关系如下：

$$\begin{bmatrix} \mu_{AABB} \\ \mu_{aabb} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & -1 & 0 & 0 & 0 & -1 & 0 \\ 1 & -1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} m \\ a_A \\ d_A \\ a_B \\ d_B \\ aa \\ ad \\ da \\ dd \end{bmatrix}$$

利用 EXCEL 中的逆矩阵函数 MINVERSE，得到遗传参数的：

$$\begin{bmatrix} m \\ a_A \\ d_A \\ a_B \\ d_B \\ aa \\ ad \\ da \\ dd \end{bmatrix} = \begin{bmatrix} 0.25 & 0 & 0.25 & 0 & 0 & 0 & 0.25 & 0 & 0.25 \\ 0.25 & 0 & 0.25 & 0 & 0 & 0 & -0.25 & 0 & -0.25 \\ -0.25 & 0 & -0.25 & 0.5 & 0 & 0.5 & -0.25 & 0 & -0.25 \\ 0.25 & 0 & -0.25 & 0 & 0 & 0 & 0.25 & 0 & -0.25 \\ -0.25 & 0.5 & -0.25 & 0 & 0 & 0 & -0.25 & 0.5 & -0.25 \\ 0.25 & 0 & -0.25 & 0 & 0 & 0 & -0.25 & 0 & 0.25 \\ -0.25 & 0.5 & -0.25 & 0 & 0 & 0 & 0.25 & -0.5 & 0.25 \\ -0.25 & 0 & 0.25 & 0.5 & 0 & -0.5 & -0.25 & 0 & 0.25 \\ 0.25 & -0.5 & 0.25 & -0.5 & 1 & -0.5 & 0.25 & -0.5 & 0.25 \end{bmatrix} \times \begin{bmatrix} \mu_{AABB} \\ \mu_{aabb} \end{bmatrix}$$

然后就可以利用矩阵乘积函数 **MMULT** 计算任意一组基因型值对应的遗传参数。对于题目中给出的 4 组基因型值，对应的遗传参数列于下表中。

遗传效应	情形 I	情形 II	情形 II	情形 IV
<i>m</i>	4	4	4	4
<i>a<sub>A</sub></i>	1	0.5	1	2
<i>d<sub>A</sub></i>	0	0.5	1	1
<i>a<sub>B</sub></i>	1	0.5	1	0
<i>d<sub>B</sub></i>	0	0.5	1	1
<i>aa</i>	0	0	-1	-1
<i>ad</i>	0	0	-1	1
<i>da</i>	0	0	-1	1
<i>dd</i>	0	0	-1	3

**7.9** 假定控制某一性状的基因有两对，它们之间不存在连锁。选择练习 7.8 平均基因型效应中的任何一种情形，计算亲本 **P<sub>1</sub>**: *AABB* 和亲本 **P<sub>2</sub>**: *aabb* 杂交产生的 **B<sub>1</sub>**、**B<sub>2</sub>**、**F<sub>2</sub>**、**DH** 和 **RIL** 群体的均值和遗传方差。

**解答：**以练习 7.8 中的情形 IV 为例，下表给出 9 种基因型在不同群体中的频率。根据这些频率以及基因型值，就可以得到每个群体的均值和方差。

基因型	基因型值	基因型的频率				
		B <sub>1</sub> 群体	B <sub>2</sub> 群体	F <sub>2</sub> 群体	DH 群体	RIL 群体
<i>AABB</i>	5	0.25	0	0.0625	0.25	0.25
<i>AABb</i>	8	0.25	0	0.125	0	0
<i>AAbb</i>	7	0	0	0.0625	0.25	0.25

<i>AaBB</i>	6	0.25	0	0.125	0	0
<i>AaBb</i>	9	0.25	0.25	0.25	0	0
<i>Aabb</i>	4	0	0.25	0.125	0	0
<i>aaBB</i>	3	0	0	0.0625	0.25	0.25
<i>aaBb</i>	2	0	0.25	0.125	0	0
<i>aabb</i>	1	0	0.25	0.0625	0.25	0.25
群体均值		7	4	5.75	4	4
群体方差		2.5	9.5	7.4375	5	5

**7.10** 假定控制某一性状的基因有两对，它们之间的重组率  $r=0.1$ 。选择练习 7.8 平均基因型效应中的任何一种情形，计算亲本  $P_1: AABB$  和亲本  $P_2: aabb$  杂交产生的  $B_1$ 、 $B_2$ 、 $F_2$ 、DH 和 RIL 群体的均值和遗传方差。

**解答：**以练习 7.8 中的情形 IV 为例，下表给出 9 种基因型在不同群体中的频率。根据这些频率以及基因型值，就可以得到每个群体的均值和方差。

基因型	基因型值	基因型的频率				
		$B_1$ 群体	$B_2$ 群体	$F_2$ 群体	DH 群体	RIL 群体
<i>AABB</i>	5	0.45	0	0.2025	0.45	0.4167
<i>AABb</i>	8	0.05	0	0.045	0	0
<i>AAbb</i>	7	0	0	0.0025	0.05	0.0833
<i>AaBB</i>	6	0.05	0	0.045	0	0
<i>AaBb</i>	9	0.45	0.45	0.41	0	0
<i>Aabb</i>	4	0	0.05	0.045	0	0
<i>aaBB</i>	3	0	0	0.0025	0.05	0.0833
<i>aaBb</i>	2	0	0.05	0.045	0	0
<i>aabb</i>	1	0	0.45	0.2025	0.45	0.4167
群体均值		7	4.8	5.83	3.2	3.3333
群体方差		3.7	14.86	10.0311	4.36	4.5556

**7.11** 下表是两个座位上 4 种纯合基因型的 6 次重复观测数据。

基因型	重复观测值					
	重复 1	重复 2	重复 3	重复 4	重复 5	重复 6

<i>AABB</i>	27.2	23.1	17.4	19.1	25.9	21.1
<i>AAbb</i>	29.2	27.3	25.1	31.1	26.9	33.3
<i>aaBB</i>	21.5	26.2	31.9	23.2	25.7	28.1
<i>aabb</i>	10.8	7.9	8.2	6.3	10.7	11.6

(1) 把重复平均数作为基因型值的估计, 给出两个座位的加性效应和加加互作效应的估计值。

**解答:** 四种基因型均值与遗传参数的关系是:

$$\mu_{AABB} = m + a_A + a_B + aa = 22.3$$

$$\mu_{AAbb} = m + a_A - a_B - aa = 28.82$$

$$\mu_{aaBB} = m - a_A + a_B - aa = 26.1$$

$$\mu_{aabb} = m - a_A - a_B + aa = 9.25$$

从而得到两个座位的加性效应分别为 3.94 和 2.58, 加加互作效应为-5.84。

(2) 利用两因素 2×2 的方差分析, 对两个座位的遗传效应和互作效应进行显著性检验, 并估计两个座位的遗传方差及它们之间的上位性互作方差。

**解答:** 把 4 种基因型的均值列为一个 2×2 双向表, 然后计算行平均、列平均、行效应、列效应以及互作效应。

基因型	<i>BB</i>	<i>bb</i>	行平均	行效应	互作效应	<i>BB</i>	<i>bb</i>
<i>AA</i>	22.3	28.82	25.56	3.94	<i>AA</i>	-5.84	5.84
<i>aa</i>	26.1	9.25	17.68	-3.94	<i>aa</i>	5.84	-5.84
列平均	24.2	19.03	21.62				
列效应	2.58	-2.58					

根据上表的效应, 得到下面的方差分析表。其中, 4 种基因型与总平均数的离差平方和、再乘以重复次数 6, 就是表中基因型的平方和, 自由度等于 3。基因型的平方和, 进一步分解为 3 个自由度为 1 的效应平方和, 分别对应于座位 A 主效应、座位 B 主效应和两个座位之间的加加互作效应。上表中两个行效应的平方和、再乘以 12, 就是座位 A 效应的平方和。两个列效应的平方和、再乘以 12, 就是座位 B 效应的平方和。4 个互作效应的平方和、再乘以 6, 就是两个座位互作效应的平方和。

变异来源	自由度	平方和	均方	F 值	P 值	方差估计
基因型	3	1352.05	450.68	43.42	6.1E-09	73.38

座位 A	1	372.88	372.88	35.93	7.3E-06	30.21
座位 B	1	160.17	160.17	15.43	0.00083	12.48
互作	1	819.00	819.00	78.91	2.2E-08	134.77
随机误差	20	207.58	10.38			10.38
总和	23	1559.63				

方差分析表明，两个座位的遗传效应和互作效应均达到极显著性水平。用  $MS_A$ 、 $MS_B$ 、 $MS_{AB}$ 、 $MS_\varepsilon$  分别表示座位 A 主效应、座位 B 主效应、互作效应、随机误差的均方，用  $\sigma_A^2$ 、 $\sigma_B^2$ 、 $\sigma_{AB}^2$ 、 $\sigma_\varepsilon^2$  分别表示座位 A 主效应、座位 B 主效应、互作效应、随机误差的方差。固定效应模型下，均方的期望与方差之间的关系如下：

$$E(MS_\varepsilon) = \sigma_\varepsilon^2$$

$$E(MS_{AB}) = \sigma_\varepsilon^2 + r\sigma_{AB}^2, \text{ 其中 } r \text{ 是重复观测次数}$$

$$E(MS_B) = \sigma_\varepsilon^2 + 2r\sigma_B^2, \text{ } r \text{ 前面的系数 } 2 \text{ 可以看作是 A 座位上基因型的个数}$$

$$E(MS_A) = \sigma_\varepsilon^2 + 2r\sigma_A^2, \text{ } r \text{ 前面的系数 } 2 \text{ 可以看作是 B 座位上基因型的个数}$$

因此可以得到各种方差成分的无偏估计，列于上表最后一列。如需估计 4 种基因型之间的遗传方差  $\sigma_G^2$ ，仍要从基因型的均方和误差的均方进行计算，即

$$E(MS_G) = \sigma_\varepsilon^2 + r\sigma_G^2, \quad \hat{\sigma}_G^2 = \frac{1}{r}(MS_G - MS_\varepsilon)$$

**7.12** 下表是两个座位上 9 种基因型的 4 次重复观测数据。杂合基因型可利用两个纯合基因型的杂交来产生，如  $AABB$  和  $AAbb$  杂交就得到基因型  $AABb$ ， $AAbb$  和  $aaBB$  杂交就得到基因型  $AaBb$ 。

基因型	重复观测值			
	重复 1	重复 2	重复 3	重复 4
$AABB$	55	51	52	54
$AABb$	76	72	72	74
$AAbb$	11	7	7	9
$AaBB$	64	60	60	62
$AaBb$	39	35	35	38
$Aabb$	57	53	54	56
$aaBB$	43	39	40	42
$aaBb$	96	92	93	95

*aabb*    21    17    18    20    

(1) 把重复平均数作为基因型值的估计, 给出两个座位的加性效应、显性效应及各种互作效应的估计值。

**解答:** 与练习 7.8 类似, 从表型平均数得到遗传效应的估计值如下:

基因型	表型平均数	遗传效应	估计值
<i>AABB</i>	53	<i>m</i>	30.375
<i>AABb</i>	73.5	<i>a<sub>A</sub></i>	0.375
<i>AAbb</i>	8.5	<i>d<sub>A</sub></i>	27.875
<i>AaBB</i>	61.5	<i>a<sub>B</sub></i>	16.625
<i>AaBb</i>	36.75	<i>d<sub>B</sub></i>	53.375
<i>Aabb</i>	55	<i>aa</i>	5.625
<i>aaBB</i>	41	<i>ad</i>	-10.625
<i>aaBb</i>	94	<i>da</i>	-13.375
<i>aabb</i>	19	<i>dd</i>	-74.875

(2) 利用两因素 3×3 的方差分析, 对两个座位的遗传效应和互作效应进行显著性检验, 并估计两个座位的遗传方差及他们之间的上位性互作方差。

**解答:** 把 9 种基因型的表型平均数列成一个 3×3 双向表, 然后计算行平均、列平均、行效应、列效应以及互作效应。

主效应	<i>BB</i>	<i>Bb</i>	<i>bb</i>	行平均	行效应	互作效应	<i>BB</i>	<i>Bb</i>	<i>bb</i>
<i>AA</i>	53	73.5	8.5	45	-4.14	<i>AA</i>	5.31	9.56	-14.86
<i>Aa</i>	61.5	36.75	55	51.08	1.94	<i>Aa</i>	7.72	-33.28	25.56
<i>aa</i>	41	94	19	51.33	2.19	<i>aa</i>	-13.03	23.72	-10.69
列平均	51.83	68.08	27.5	49.14					
列效应	2.69	18.94	-21.64						

根据上表的效应, 得到下面的方差分析表。其中, 9 种基因型与总平均数的离差平方和、再乘以重复次数 4, 就是表中基因型的平方和, 自由度等于 8。基因型的平方和, 进一步分解为 3 个自由度为 2 的效应平方和, 分别对应于座位 A 主效应、座位 B 主效应和两个座位之间的加加互作效应。上表中三个行效应的平方和、再乘以 12, 就是座位 A 效应的平方和。三个列效应的平方和、再乘以 12, 就是座位 B 效应的平方和。9 个互作效应的平方和、再乘以 4, 就是两个座位互作效应的平方和。

变异来源	自由度	平方和	均方	F 值	P 值	方差估计
基因型	8	22350.56	2793.82	787.81	6.2E-30	697.57
座位 A	2	308.72	154.36	43.53	3.6E-09	12.57
座位 B	2	10012.72	5006.36	1411.72	4.8E-28	416.90
交互	4	12029.11	3007.28	848.01	5.9E-28	750.93
随机误差	27	95.75	3.55			3.55
总和	35	22446.31				

方差分析表明，两个座位的遗传效应和交互效应均达到极显著性水平。用  $MS_A$ 、 $MS_B$ 、 $MS_{AB}$ 、 $MS_\varepsilon$  分别表示座位 A 主效应、座位 B 主效应、交互效应、随机误差的均方，用  $\sigma_A^2$ 、 $\sigma_B^2$ 、 $\sigma_{AB}^2$ 、 $\sigma_\varepsilon^2$  分别表示座位 A 主效应、座位 B 主效应、交互效应、随机误差的方差。固定效应模型下，均方的期望与方差之间的关系如下：

$$E(MS_\varepsilon) = \sigma_\varepsilon^2$$

$$E(MS_{AB}) = \sigma_\varepsilon^2 + r\sigma_{AB}^2, \text{ 其中 } r \text{ 是重复观测次数}$$

$$E(MS_B) = \sigma_\varepsilon^2 + 3r\sigma_B^2, \text{ } r \text{ 前面的系数 } 3 \text{ 可以看作是 A 座位上基因型的个数}$$

$$E(MS_A) = \sigma_\varepsilon^2 + 3r\sigma_A^2, \text{ } r \text{ 前面的系数 } 3 \text{ 可以看作是 B 座位上基因型的个数}$$

因此可以得到各种方差成分的无偏估计，列于上表最后一列。如需估计 4 种基因型之间的遗传方差  $\sigma_G^2$ ，仍要从基因型的均方和误差的均方进行计算，即

$$E(MS_G) = \sigma_\varepsilon^2 + r\sigma_G^2, \quad \hat{\sigma}_G^2 = \frac{1}{r}(MS_G - MS_\varepsilon)$$

## 第 8 章练习题和参考答案

**8.1** 对于表8.9给出的基因型值分解结果，验证  $E(A) = 0$ 、 $E(D) = 0$ 、 $E(AD) = 0$ 。

**解答：**将表 8.9 中不同基因型的频率、育种值和显性离差列于下表。

基因型	频率	育种值 $A_{ij}$	显性离差 $D_{ij}$
$A_1A_1$	$p^2$	$2q\alpha$	$-2q^2d$
$A_1A_2$	$2pq$	$(q-p)\alpha$	$2pqd$
$A_2A_2$	$q^2$	$-2p\alpha$	$-2p^2d$

因此，

$$E(A) = p^2(2q\alpha) + 2pq[(q-p)\alpha] + q^2(-2p\alpha) = 0$$

$$E(D) = p^2(-2q^2d) + 2pq(2pqd) + q^2(-2p^2d) = 0$$

$$E(AD) = p^2[2q\alpha \times (-2q^2d)] + 2pq[(q-p)\alpha \times (2pqd)] \\ + q^2[(-2p\alpha) \times (-2p^2d)] = 0$$

**8.2** 已知一个老鼠矮化基因座位上的两个等位基因为+和pg，三种基因型的六周龄平均体重（g）如下表。

基因型	++	+pg	pg pg
平均体重/g	14	12	6

在等位基因 pg 频率为 0.2 的随机交配群体中，计算：

(1) 群体平均体重和遗传方差。

**解答：**利用公式 8.2 和公式 8.3，群体平均体重和遗传方差分别为 13.04 和 2.9184。

(2) 三种基因型的遗传效应。

**解答：**利用公式 8.4，将基因型值减去群体平均体重，得到 3 种基因型的遗传效应分别为 0.96、-1.04 和-7.04。

(3) 等位基因+和 pg 的平均效应。

**解答：**利用公式 8.6 和公式 8.7，得到两个等位基因的平均效应分别为 0.56 和-2.24。

(4) 三种基因型的育种值和显性离差。

**解答：**将基因型中两个等位基因的平均效应相加，得到 3 种基因型的育种值分别为 1.12、-1.68 和-4.48。从基因型的遗传效应中减去育种值，得到 3 种基因型的显性离差分别为-0.16、

0.64 和-2.56。

(5) 群体中的加性方差  $V_A$ 、显性方差  $V_D$ 、总遗传方差  $V_G$ 。

**解答:** 三种基因型的频率分别为 0.64、0.32 和 0.04。利用 3 种基因型的育种值 1.12、-1.68 和-4.48，得到加性方差等于 2.5088；利用 3 种基因型的显性离差-0.16、0.64 和-2.56，得到显性方差等于 0.4096。总遗传方差等于二者之和，即 2.9184，也正好等于 (1) 计算出的遗传方差。

将上面的结果整理如下表。遗传参数  $m$ 、 $a$ 、 $d$  由加显性模型 8.1 估计，等位基因的频率效应由公式 8.6 和公式 8.7 估计。根据表中的基因型值和频率就能计算群体均值和遗传方差；从基因型值减去群体平均体重估计 3 种基因型的遗传效应；将基因型中两个等位基因的平均效应相加，得到 3 种基因型的育种值；从基因型的遗传效应中减去育种值，得到 3 种基因型的显性离差；利用三种基因型的频率和育种值，得到加性方差；利用三种基因型的频率和显性离差，得到显性方差。

基因型	++	+pg	pgpg	$m$	$a$	$d$	$p$	$q$	$\alpha_+$	$\alpha_{pg}$
基因型值	14	12	6	10	4	2	0.8	0.2	0.56	-2.24
频率	0.64	0.32	0.04							
群体均值	13.04									
遗传方差	2.9184									
遗传效应	0.96	-1.04	-7.04							
育种值	1.12	-1.68	-4.48							
显性离差	-0.16	0.64	-2.56							
加性方差	2.5088									
加性方差	0.4096									
遗传方差	2.9184									

**8.3** 一个座位上 3 种基因型的平均表现如下表。

基因型	$A_1A_1$	$A_1A_2$	$A_2A_2$
平均表现	110	150	90

在等位基因  $A_1$  频率为 0.2 的随机交配群体中，计算：

- (1) 群体均值和遗传方差。
- (2) 三种基因型的遗传效应。

- (3) 等位基因  $A_1$  和  $A_2$  的平均效应。
- (4) 三种基因型的育种值和显性离差。
- (5) 群体中的加性方差  $V_A$ 、显性方差  $V_D$ 、总遗传方差  $V_G$ 。

**解答：**计算方法同练习 8.2，结果整理在下表中。

基因型	$A_1A_1$	$A_1A_2$	$A_2A_2$	$m$	$a$	$d$	$p$	$q$	$\alpha_1$	$\alpha_2$
基因型值	110	150	90	100	10	50	0.2	0.8	32.00	-8.00
频率	0.04	0.32	0.64							
群体均值	110									
遗传方差	768									
遗传效应	0	40	-20							
育种值	64	24	-16							
显性离差	-64	16	-4							
加性方差	512									
加性方差	256									
遗传方差	768									

**8.4** 一个座位上三种基因型的平均表现如下表。

基因型	$A_1A_1$	$A_1A_2$	$A_2A_2$
平均表现	110	150	90

什么样的随机交配群体具有最高的平均表现？并对此随机交配群体，计算：

- (1) 群体均值和遗传方差。
- (2) 三种基因型的遗传效应。
- (3) 等位基因  $A_1$  和  $A_2$  的平均效应。
- (4) 三种基因型的育种值和显性离差。
- (5) 群体中的加性方差  $V_A$ 、显性方差  $V_D$ 、总遗传方差  $V_G$ 。

**解答：**根据练习 8.3，遗传参数  $m=100$ 、 $a=10$ 、 $d=50$ 。用  $p$  表示等位基因  $A_1$  的频率， $q=1-p$  表示等位基因  $A_2$  的频率。则公式 8.2 给出的群体均值为：

$$\begin{aligned}\mu &= 100 + (2p - 1) \times 10 + 2(p - p^2) \times 50 = 90 + 120p - 100p^2 \\ &= 126 - 100(0.6 - p)^2\end{aligned}$$

因此，当基因频率  $p=0.6$ ，有最大值 126。其它结果整理在下表中。

基因型	$A_1A_1$	$A_1A_2$	$A_2A_2$	$m$	$a$	$d$	$p$	$q$	$\alpha_1$	$\alpha_2$
基因型值	110	150	90	100	10	50	0.6	0.4	0	0
频率	0.04	0.32	0.64							
群体均值	126									
遗传方差	576									
遗传效应	-16	24	-36							
育种值	0	0	0							
显性离差	-16	24	-36							
加性方差	0									
加性方差	576									
遗传方差	576									

**8.5** 下表给出随机交配群体中，一个座位上三种基因型的频率、基因型值和等位基因数目（即表 8.4）。

基因型	频率 $f$	基因型值 $Y$	基因 $A_1$ 的数目 $X$
$A_1A_1$	$p^2$	$a$	2
$A_1A_2$	$2pq$	$d$	1
$A_2A_2$	$q^2$	$-a$	0

(1) 计算  $X$  和  $Y$  的均值。

**解答：** 根据均值的定义，并利用条件  $p+q=1$ ，得到：

$$E(X) = p^2 \times 2 + 2pq \times 1 + q^2 \times 0 = 2p(p+q) = 2p$$

$$E(Y) = p^2 \times a + 2pq \times d + q^2 \times (-a) = (p-q)a + 2pqd$$

(2) 计算  $X$  的方差。

**解答：** 根据方差的定义得到：

$$V(X) = p^2 \times 2^2 + 2pq \times 1^2 + q^2 \times 0^2 - [E(X)]^2 = 4p^2 + 2pq - 4p^2 = 2pq$$

(3) 计算  $Y$  与  $X$  之间的协方差。

**解答：** 根据协方差的定义得到：

$$\begin{aligned} \text{Cov}(X, Y) &= p^2 \times 2 \times a + 2pq \times 1 \times d + q^2 \times 0 \times (-a) - E(X)E(Y) \\ &= 2pqa + 2pq(1-2p)d = 2pq[a + (q-p)d] = 2pq\alpha \end{aligned}$$

(4) 计算  $Y$  对  $X$  的一元回归方程。

**解答：**根据协方差的定义得到：

$$b_{YX} = \frac{\text{Cov}(X,Y)}{V(X)} = \frac{2pq\alpha}{2pq} = \alpha$$

**8.6** 下表给出随机交配群体中，雌雄配子随机结合后代基因型值的  $2 \times 2$  双向表。根据方差分析原理，计算表中的行列平均数和行列效应，并由此计算三种基因型的育种值和显性离差。

雌配子及其频率	雄配子及其频率		行平均	行效应
	$A_1, 0.7$	$A_2, 0.3$		
$A_1, 0.7$	14	12		
$A_2, 0.3$	12	6		
列平均				
列效应				

**解答：**将两等位基因的频率看着编辑频率，计算 4 种基因型的频率（相当于把杂合基因型  $A_1A_2$  分成两半，分别对应于两种雌雄配子的结合类型）。

联合频率	$A_1$	$A_2$	边际频率
$A_1$	0.49	0.21	0.7
$A_2$	0.21	0.09	0.3
边际频率	0.7	0.3	

等位基因  $A_1$  所在行的平均数，等于基因型值 14 和 12 以边际频率 0.7 和 0.3 为权的加权平均，即 13.4。等位基因  $A_2$  所在行的平均数，等于基因型值 12 和 6 以边际频率 0.7 和 0.3 为权的加权平均，即 10.2。行平均数与行平均数相等。总平均数等于 4 种基因型值 14、12、12、6 以联合频率 0.49、0.21、0.21、0.09 为权的加权平均，即 12.44。

行、列平均数与总平均数的离差即为行、列效应，均等于 0.96 和 -2.24。从基因型值与总平均数的离差中减去行、列效应之后，得到的就是显性离差效应。4 种显性离差分别为 -0.36、0.84、0.84 和 -1.96。上述结果综合起来列于下表。

基因型值	$A_1$	$A_2$	行平均	行效应	显性离差	$A_1$	$A_2$
$A_1$	14	12	13.4	0.96	$A_1$	-0.36	0.84
$A_2$	12	6	10.2	-2.24	$A_2$	0.84	-1.96
列平均	13.4	10.2	12.44				
列效应	0.96	-2.24					

**8.7** 假定一个座位上三种基因型  $A_1A_1$ 、 $A_1A_2$ 、 $A_2A_2$  的平均表现分别为 10、10 和 6。在

$A_1$  和  $A_2$  频率分别为 0.7 和 0.3 的随机交配群体中，下表给出了三种基因型频率，以及三类半同胞家系后代的基因型频率。

基因型	基因型值	频率	半同胞后代基因型的频率			半同胞家系的均值	半同胞家系内的方差
			$A_1A_1$	$A_1A_2$	$A_2A_2$		
$A_1A_1$	10	0.49	0.7	0.3	0		
$A_1A_2$	10	0.42	0.35	0.5	0.15		
$A_2A_2$	6	0.09	0	0.7	0.3		

(1) 根据上表的三种基因型值和频率，计算随机交配群体的遗传方差。

**解答：**该随机交配群体的均值等于 9.64，遗传方差等于 1.3104。

(2) 根据上表的三种基因型值和家系内的基因型频率，计算家系均值和家系内的方差，即上表的最后两列。

**解答：**基因型  $A_1A_1$  半同胞家系的均值和家系内方差分别为 10 和 0。基因型  $A_1A_2$  半同胞家系的均值和家系内方差分别为 9.4 和 2.04。基因型  $A_2A_2$  半同胞家系的均值和家系内方差分别为 8.8 和 3.36。

(3) 根据 (2) 计算出的家系均值，计算家系间方差；根据 (2) 计算出的家系内的方差，计算家系内的平均方差；验证这两部分方差之和等于 (1) 计算出的遗传方差。

**解答：**三种家系平均数 10、9.4、8.8 以频率 0.49、0.42、0.09 为权重的方差，即为家系平均数的方差，即 0.1512。三种家系内方差 0、2.04、3.36 以频率 0.49、0.42、0.09 为权的加权平均数，即为家系内的平均方差，即 1.1592。二者之和等于 1.3104，正好等于 (1) 得到的遗传方差。

(4) 计算亲本与半同胞家系均值之间的协方差，并说明协方差与家系间方差的关系。

**解答：**从公式 8.21 和公式 8.23 可以看到，家系间的方差等于  $\frac{1}{4}V_A$ ，亲本与半同胞家系之间的协方差等于  $\frac{1}{2}V_A$ 。这里，亲本与半同胞家系均值之间的协方差等于 0.3024，正好等于家系间方差 0.1512 的两倍。

将上面的结果整理如下表。

基因型	基因型值	频率	半同胞家系基因型的频率			家系均值	家系内方差
			$A_1A_1$ , 10	$A_1A_2$ , 10	$A_2A_2$ , 6		
$A_1A_1$	10	0.49	0.7	0.3	0	10	0
$A_1A_2$	10	0.42	0.35	0.5	0.15	9.4	2.04
$A_2A_2$	6	0.09		0.7	0.3	8.8	3.36

均值	9.64	9.64
方差	1.3104	0.1512    1.1592
协方差	0.3024	

**8.8** 假定一个座位上三种基因型  $A_1A_1$ 、 $A_1A_2$ 、 $A_2A_2$  的平均表现分别为 10、10 和 6。在  $A_1$  和  $A_2$  频率分别为 0.7 和 0.3 的随机交配群体中，下表给出了 6 种全同胞家系的中亲值。

全同胞家系	中亲值	频率	全同胞后代基因型的频率			全同胞家系的均值	全同胞家系内的方差
			$A_1A_1$	$A_1A_2$	$A_2A_2$		
$A_1A_1 \times A_1A_1$	10						
$A_1A_1 \times A_1A_2$	10						
$A_1A_1 \times A_2A_2$	8						
$A_1A_2 \times A_1A_2$	10						
$A_1A_2 \times A_2A_2$	8						
$A_2A_2 \times A_2A_2$	6						

(1) 计算各种全同胞家系的频率、以及全同胞后代的基因型频率，即上表中亲值后面两列。

(2) 在 (1) 的基础上，计算全同胞家系的均值和方差，即上表的最后两列。

(3) 根据 (2) 计算出的家系均值，计算家系间方差；根据 (2) 计算出的家系内的方差，计算家系内的平均方差。

(4) 计算中亲与全同胞家系均值之间的协方差。

**解答：** 计算过程省略，结果如下表。

基因型	中亲	频率	全同胞家系内基因型的频率			家系均值	家系方差
			$A_1A_1$ , 10	$A_1A_2$ , 10	$A_2A_2$ , 6		
$A_1A_1 \times A_1A_1$	10	0.2401	1	0	0	10	0
$A_1A_1 \times A_1A_2$	10	0.4116	0.5	0.5	0	10	0
$A_1A_1 \times A_2A_2$	8	0.0882	0	1	0	10	0
$A_1A_2 \times A_1A_2$	10	0.1764	0.25	0.5	0.25	9	3
$A_1A_2 \times A_2A_2$	8	0.0756	0	0.5	0.5	8	4
$A_2A_2 \times A_2A_2$	6	0.0081	0	0	1	6	0
均值	9.64					9.64	
方差						0.4788	0.8316
协方差	0.302						

**8.9** 随机交配可以降低连锁不平衡度 (LD), 从而检测到比较紧密的连锁。例如, 美国的 IBM 群体是以玉米自交系‘B73’作母本、‘Mo17’作父本杂交, 自  $F_2$  开始随机交配 4 个世代后再连续自交而产生的重组近交家系。假定 2 个座位间一次交换的重组率为  $r$ , 每个座位上只有 2 个等位基因, 分别用  $A, a$  和  $B, b$  表示, 2 个亲本的基因型分别为  $AABB$  和  $aabb$ , 杂种  $F_1$  的基因型为  $AB/ab$ 。

(1) 计算  $F_1$  产生配子的 LD, 并视为  $F_1$  随机交配一代 (即  $F_2$ ) 的 LD。

**解答:** 杂种  $F_1$  的基因型是  $AB/ab$ , 两种亲本型配子  $AB$  和  $ab$  的频率均为  $\frac{1}{2}(1-r)$ , 两种重组型配子  $Ab$  和  $aB$  的频率均为  $\frac{1}{2}r$ , 连锁不平衡度记为  $D_0$ , 与重组率的关系为:

$$D_0 = \left[\frac{1}{2}(1-r)\right]^2 - \left(\frac{1}{2}r\right)^2 = \frac{1}{4}(1-2r)$$

(2) 如果在  $F_2$  群体内再次随机交配, 计算  $F_2$  随机交配 1 代群体中的 LD 值。

**解答:**  $F_2$  随机交配 1 代群体中的连锁不平衡度记为  $D$ , 与  $D_0$  和重组率的关系为:

$$D = D_0(1-r) = \frac{1}{4}(1-2r)(1-r)$$

(3) 已知各种基因型连续自交产生的 RIL 群体中 4 种纯合基因型  $AABB$ 、 $AAbb$ 、 $aaBB$  和  $aabb$  的频率如下表, 将  $F_2$  随机交配 1 代群体中各种基因型的频率填在下表第 2 列。计算  $F_2$  随机交配 1 代后再连续自交产生的 RIL 群体中 4 种基因型的频率, 并填在下表最后一行。

亲代基因型	频率	连续自交多代的基因型, 已知 $R = \frac{2r}{1+2r}$ , $r$ 为一次交换的重组率			
		$AABB$	$AAbb$	$aaBB$	$aabb$
$AABB$	1	0	0	0	0
$AAbb$	0.5	0.5	0	0	0
$Aabb$	0	1	0	0	0
$AaBB$	0.5	0	0.5	0	0
$AB/ab$	$\frac{1}{2}(1-R)$	$\frac{1}{2}R$	$\frac{1}{2}R$	$\frac{1}{2}R$	$\frac{1}{2}(1-R)$
$Ab/aB$	$\frac{1}{2}R$	$\frac{1}{2}(1-R)$	$\frac{1}{2}(1-R)$	$\frac{1}{2}(1-R)$	$\frac{1}{2}R$
$Aabb$	0	0.5	0	0.5	0
$aaBB$	0	0	1	0	0
$aaBb$	0	0	0.5	0.5	0
$aabb$	0	0	0	0	1
累计频率					

解答:  $F_2$  随机交配 1 代群体中的连锁不平衡度为  $D = \frac{1}{4}(1-2r)(1-r)$ , 群体中两种亲本型配子  $AB$  和  $ab$  的频率均为  $\frac{1}{4}+D$ , 两种重组型配子  $Ab$  和  $aB$  的频率均为  $\frac{1}{4}-D$ 。这 4 种配子的随机结合, 就得到下表 16 种结合方式得到的基因型及其频率。

配子和频率	$AB, \frac{1}{4}+D$	$Ab, \frac{1}{4}-D$	$aB, \frac{1}{4}-D$	$ab, \frac{1}{4}+D$
$AB, \frac{1}{4}+D$	$AB/AB, (\frac{1}{4}+D)^2$	$AB/Ab, \frac{1}{16}-D^2$	$AB/aB, \frac{1}{16}-D^2$	$AB/ab, (\frac{1}{4}+D)^2$
$Ab, \frac{1}{4}-D$	$Ab/AB, \frac{1}{16}-D^2$	$Ab/Ab, (\frac{1}{4}-D)^2$	$Ab/aB, (\frac{1}{4}-D)^2$	$Ab/ab, \frac{1}{16}-D^2$
$aB, \frac{1}{4}-D$	$aB/AB, \frac{1}{16}-D^2$	$aB/Ab, (\frac{1}{4}-D)^2$	$aB/aB, (\frac{1}{4}-D)^2$	$aB/ab, \frac{1}{16}-D^2$
$ab, \frac{1}{4}+D$	$ab/AB, (\frac{1}{4}+D)^2$	$ab/Ab, \frac{1}{16}-D^2$	$ab/aB, \frac{1}{16}-D^2$	$ab/ab, (\frac{1}{4}+D)^2$

将上表相同基因型的频率合并, 得到下表第 2 列给出的 10 种基因型频率。

亲代基因型	频率	连续自交多代的基因型			
		$AABB$	$AAbb$	$aaBB$	$aabb$
$AABB$	$\frac{1}{16}(1+4D)^2$	1	0	0	0
$AABb$	$\frac{1}{8}(1-16D^2)$	0.5	0.5	0	0
$AAbb$	$\frac{1}{16}(1-4D)^2$	0	1	0	0
$AaBB$	$\frac{1}{8}(1-16D^2)$	0.5	0	0.5	0
$AB/ab$	$\frac{1}{8}(1+4D)^2$	$\frac{1}{2}(1-R)$	$\frac{1}{2}R$	$\frac{1}{2}R$	$\frac{1}{2}(1-R)$
$Ab/aB$	$\frac{1}{8}(1-4D)^2$	$\frac{1}{2}R$	$\frac{1}{2}(1-R)$	$\frac{1}{2}(1-R)$	$\frac{1}{2}R$
$Aabb$	$\frac{1}{8}(1-16D^2)$	0	0.5	0	0.5
$aaBB$	$\frac{1}{16}(1-4D)^2$	0	0	1	0
$aaBb$	$\frac{1}{8}(1-16D^2)$	0	0	0.5	0.5
$aabb$	$\frac{1}{16}(1+4D)^2$	0	0	0	1
累计频率		$\frac{1}{4}+D(1-R)$	$\frac{1}{4}-D(1-R)$	$\frac{1}{4}-D(1-R)$	$\frac{1}{4}+D(1-R)$

亲代中 10 种基因型的频率与 RIL 后代基因型所在列频率的积和, 则为 RIL 群体中 4 种

基因型的频率。列于上表最后一行。基因型 *AABB* 和 *aabb* 有相同的频率，具体计算过程如下：

$$\begin{aligned} & \frac{1}{16}(1+4D)^2 + \frac{1}{8}(1-16D^2) + \frac{1}{16}(1+4D)^2(1-R) + \frac{1}{16}(1-4D)^2R \\ &= \frac{1}{8}(1+4D)^2 + \frac{1}{8}(1-16D^2) - \frac{1}{16}(1+4D)^2R + \frac{1}{16}(1-4D)^2R \\ &= \frac{1}{8}(2+8D) - \frac{1}{16}R[(1+4D)^2 - (1-4D)^2] \\ &= \frac{1}{4} + D - RD = \frac{1}{4} + D(1-R) \end{aligned}$$

基因型 *AAbb* 和 *aaBB* 有相同的频率，具体计算过程如下：

$$\begin{aligned} & \frac{1}{16}(1-4D)^2 + \frac{1}{8}(1-16D^2) + \frac{1}{16}(1-4D)^2(1-R) + \frac{1}{16}(1+4D)^2R \\ &= \frac{1}{8}(1-4D)^2 + \frac{1}{8}(1-16D^2) - \frac{1}{16}(1-4D)^2R + \frac{1}{16}(1+4D)^2R \\ &= \frac{1}{8}(2-8D) - \frac{1}{16}R[(1-4D)^2 - (1+4D)^2] \\ &= \frac{1}{4} - D + RD = \frac{1}{4} - D(1-R) \end{aligned}$$

上面的结果对于  $F_2$  随机交配多代之后、再连续自交衍生的 RIL 群体也是适用的，只不过把上面的不平衡度  $D$  用随机交配多代的不平衡度  $D = \frac{1}{4}(1-2r)(1-r)^t$  代替即可。

**8.10** 假定两对独立遗传基因的 9 种基因型的平均表现如下表。这些表现型在 *AABB* 和 *aabb* 的杂种  $F_2$  群体中将出现 15:1 的分离比。

座位 A	座位 B		
	<i>BB</i>	<i>Bb</i>	<i>bb</i>
<i>AA</i>	13	13	13
<i>Aa</i>	13	13	13
<i>aa</i>	13	13	1

以 *AABB* 和 *aabb* 为亲本，不考虑其他遗传因素，计算：

(1) 两个座位上的加显性效应、以及两个座位间的各种上位型互作效应。

**解答：**计算方法见练习 7.8，这里只给出结果。两个座位上的加显性效应均等于 3，4 种互作效应均等于-3。

(2) 杂种  $F_2$  群体的平均表现和遗传方差。

(3) 杂种  $F_3$  群体的平均表现和遗传方差。

(4) 杂种  $F_1$  连续自交衍生的重组近交家系群体的平均表现和遗传方差。

**解答：**将三个群体中，9 种基因型的频率列于下表。根据 9 种基因型值和频率计算群体

的均值和方差，结果列于下表最后两行。

座位 A	座位 B	基因型值	基因型频率		
			F <sub>2</sub> 群体	F <sub>3</sub> 群体	RIL 群体
AA	BB	13	0.0625	0.140625	0.25
AA	Bb	13	0.125	0.09375	0
AA	bb	13	0.0625	0.140625	0.25
Aa	BB	13	0.125	0.09375	0
Aa	Bb	13	0.25	0.0625	0
Aa	bb	13	0.125	0.09375	0
aa	BB	13	0.0625	0.140625	0.25
aa	Bb	13	0.125	0.09375	0
aa	bb	1	0.0625	0.140625	0.25
均值			12.25	11.3125	10
方差			8.4375	17.402344	27

**8.11** 假定两对独立遗传基因的 9 种基因型的平均表现如下表。这些表现型在 *AABB* 和 *aabb* 的杂种 F<sub>2</sub> 群体中将出现 9:3:4 的分离比。

座位 A	座位 B		
	BB	Bb	bb
AA	3	3	2
Aa	3	3	2
aa	1	1	1

在亲本 *AABB* 和 *aabb* 产生的 F<sub>2</sub> 群体中，利用方差分析的方法计算：

(1) 座位 A 和座位 B 的主效应。

**解答：**下表上半部分给出 9 中基因型的频率。边际频率分别对应于两个座位上的三种基因型。基因型 AA 所在行的三种基因型值与座位 B 基因型频率的加权平均，即为基因型 AA 的行平均。基因型 BB 所在列的三种基因型值与座位 A 基因型频率的加权平均，即为基因型 BB 的列平均。其它行、列平均数的计算方法与此类似。

表中 9 种基因型值与 9 种基因型频率的的加权平均，即为总平均。行、列平均数与总平均的离差，即为行、列效应，分别对应于座位 A 和座位 B 的主效应。

联合频率	BB	Bb	bb	边际频率
------	----	----	----	------

AA	0.0625	0.125	0.0625	0.25
Aa	0.125	0.25	0.125	0.5
Aa	0.0625	0.125	0.0625	0.25
边际频率	0.25	0.5	0.25	

基因型值	BB	Bb	bb	行平均	行效应
AA	3	3	2	2.75	0.4375
Aa	3	3	2	2.75	0.4375
aa	1	1	1	1	-1.3125
列平均	2.5	2.5	1.75	2.3125	
列效应	0.1875	0.1875	-0.5625		

(2) 座位 A 和座位 B 之间的互作效应。

**解答：**在 (1) 中，每种基因型值与总平均的离差可以看作是总效应，从中减去行、列效应，得到的就是两个座位的互作效应。结果见下表。

互作效应	BB	Bb	bb
AA	0.0625	0.0625	-0.1875
Aa	0.0625	0.0625	-0.1875
aa	-0.1875	-0.1875	0.5625

(3) 座位 A 的方差和座位 B 的方差。

**解答：**以座位 A 三种基因型的频率 0.25、0.5、0.25 为权重，得到行效应的方差等于 0.5742，即为座位 A 的方差。以座位 B 三种基因型的频率 0.25、0.5、0.25 为权重，得到列效应的方差等于 0.1055，即为座位 B 的方差。

(4) 座位 A 与座位 B 之间的互作方差。

**解答：**以 9 种基因型的频率为权重，得到 (2) 互作效应的方差等于 0.0352，即为座位 A 和座位 B 的互作方差。

(5) 互作方差占遗传方差的比例。

**解答：**以 9 种基因型的频率为权重，得到基因型值的方差等于 0.7148，即为总遗传方差，正好等于 (3) 中座位 A 方差和座位 B 方差以及 (4) 中互作方差这三者之和。互作方差占总遗传方差的 4.92%。

## 第 9 章练习题和参考答案

**9.1** 下表是4个基因型在三个环境下的平均表现，不考虑随机误差。计算基因型效应、环境效应和互作效应；计算各种方差成分及占总方差的百分数。

基因型	环境 1	环境 2	环境 3
基因型 1	13	14	15
基因型 2	18	19	26
基因型 3	17	21	13
基因型 4	15	24	27

**解答：**略去计算过程，基因型效应、环境效应和互作效应的估计值如下表。

基因型	表型平均			行平均	行效应	互作效应		
	环境 1	环境 2	环境 3			环境 1	环境 2	环境 3
基因型 1	13	14	15	14	-4.5	1.75	-1	-0.75
基因型 2	18	19	26	21	2.5	-0.25	-3	3.25
基因型 3	17	21	13	17	-1.5	2.75	3	-5.75
基因型 4	15	24	27	22	3.5	-4.25	1	3.25
列平均	15.75	19.5	20.25	18.5				
列效应	-2.75	1	1.75					

假定 4 种基因型有相同的频率或权重，3 个环境也具有相同的频率或权重。利用 EXCEL 中的样本方差函数 VAR.S 得到的基因型方差、环境方差、互作方差如下。总方差等于三个方差成分之和，进而计算各种方差成分占总方差的百分数。结果见下表。

效应类型	方差估计	百分数
基因型间	13.67	47.31
环境间	5.81	20.12
基因型与环境互作	9.41	32.57
总效应	28.89	

**9.2** 下表是20个玉米自交系（用L1~L20表示）在干旱和非干旱两种环境条件下、三次重复的吐丝期（日）观测数据。

自交系编号	干旱环境			非干旱环境		
	重复 I	重复 II	重复 III	重复 I	重复 II	重复 III
L1	101	90	91	89	89	93
L2	82	85	87	84	84	83
L3	86	85	83	80	84	88
L4	85	85	87	83	84	83
L5	80	82	81	81	82	83

L6	95	98	95	89	94	95
L7	84	85	85	81	84	85
L8	86	85	87	84	85	83
L9	87	89	91	85	88	87
L10	84	85	89	82	85	85
L11	82	85	83	81	85	82
L12	83	85	87	83	87	83
L13	89	87	94	89	92	88
L14	90	92	93	89	91	90
L15	95	89	95	89	90	91
L16	82	85	87	84	85	87
L17	91	95	92	89	94	91
L18	88	90	89	89	91	88
L19	84	85	87	82	85	85
L20	88	92	101	89	90	93

(1) 列出 20 种基因型和两种环境吐丝期性状的 GE 双向表, 在此基础上计算基因型的主效应和环境主效应, 以及基因型和环境互作效应。

**解答:** 下表第 2、3 列给出 20 种基因型和两种环境吐丝期性状的 GE 双向表。行平均其实是自交系在环境和重复间的平均表现, 与总平均数的离差即为基因型的主效应。列平均其实是环境在自交系和重复间的平均表现, 与总平均数的离差即为环境的主效应。GE 双向表中, 每个自交系在每个环境的重复平均数与总平均数的离差可以看作总效应, 从中减去基因型和环境的主效应, 剩下部分就是互作效应。

自交系	重复平均数		行平均	自交系效应	互作效应	
	干旱环境	非干旱环境			干旱环境	非干旱环境
L1	94.00	90.33	92.17	4.88	1.12	-1.12
L2	84.67	83.67	84.17	-3.12	-0.22	0.22
L3	84.67	84.00	84.33	-2.95	-0.38	0.38
L4	85.67	83.33	84.50	-2.78	0.45	-0.45
L5	81.00	82.00	81.50	-5.78	-1.22	1.22
L6	96.00	92.67	94.33	7.05	0.95	-0.95
L7	84.67	83.33	84.00	-3.28	-0.05	0.05
L8	86.00	84.00	85.00	-2.28	0.28	-0.28
L9	89.00	86.67	87.83	0.55	0.45	-0.45
L10	86.00	84.00	85.00	-2.28	0.28	-0.28
L11	83.33	82.67	83.00	-4.28	-0.38	0.38
L12	85.00	84.33	84.67	-2.62	-0.38	0.38
L13	90.00	89.67	89.83	2.55	-0.55	0.55
L14	91.67	90.00	90.83	3.55	0.12	-0.12
L15	93.00	90.00	91.50	4.22	0.78	-0.78
L16	84.67	85.33	85.00	-2.28	-1.05	1.05
L17	92.67	91.33	92.00	4.72	-0.05	0.05

L18	89.00	89.33	89.17	1.88	-0.88	0.88
L19	85.33	84.00	84.67	-2.62	-0.05	0.05
L20	93.67	90.67	92.17	4.88	0.78	-0.78
列平均	88.00	86.57	87.28			
环境效应	0.72	-0.72				

(2) 不考虑重复的区组效应，对吐丝期进行多环境联合方差分析，并利用联合方差分析的结果，估计吐丝期性状平均表现（即环境和重复平均数）的遗传力。

**解答：**利用线性模型公式 9.7 进行方差分析。各种变异的自由度列于下表第 2 列。基因型的自由度为 19、环境的自由度为 1、互作的自由度为 19。总自由度为 119，因此得到误差自由度等于 80。各种变异的自由度列于下表第 2 列。基因型平方和等于（1）中 20 种基因型主效应的平方和乘以环境数 2 和重复数 3。环境平方和等于（1）中 2 种环境主效应的平方和乘以基因型 20 和重复数 3。互作平方和等于（1）中 40 种互作效应的平方和乘以重复数 3。总平方和等于 120 个观测值与总平均的离差平方和。

第 3 列的平方和除以第 2 列的自由度，得到第 4 列给出的均方。从均方就可以计算 F 值、并进行显著性检验（略）。利用表 9.4 的期望均方，得到各种方差成分的估计值，列于下表第 5 列。本例中的互作效应不显著，它的无偏估计还是一负值，因此可以认为互作方差为 0。利用公式 9.24 得到环境和重复平均数的遗传力为 0.9336。

变异来源	自由度	平方和	均方	方差估计值	遗传力
基因型	19	1670.70	87.93	13.68	0.9336
环境	1	61.63	61.63	0.93	
互作	19	48.70	2.56	-1.09	
机误	80	467.33	5.84	5.84	
总和	119	2248.37			

**9.3** 对于练习 9.2 中的数据，现假定每个重复是被安排在一个土壤条件相对一致的区组内。因此，方差分析时需要考虑区组效应。考虑区组、基因型、环境、以及基因型和环境互作这 4 种变异来源，对于练习 9.2 中的数据作联合方差分析，并利用联合方差分析的结果，估计吐丝期性状平均表现（即环境和重复平均数）的遗传力。

**解答：**利用线性模型公式 9.11 进行方差分析。首先说明一下区组效应的估计。每个环境中 有 3 个区组，因此共有 6 个区组。下表中的列平均即为这 6 个区组的平均数，它们与所在环境的总平均之间的离差，即为区组效应。这其实也是公式 9.11 中将区组效应记为  $B_{k/j}$  的含义所在。因此，多环境试验中，区组效应等于每个区组内所有观测值平均数与所在环境内观测值平均数的离差。离差计算时，不能用总平均。更不能在环境之间计算区组 I、区组 II、

区组 III 的效应。

自交系	干旱环境			非干旱环境		
	重复 I	重复 II	重复 III	重复 I	重复 II	重复 III
L1	101	90	91	89	89	93
...	...	...	...	...	...	...
L20	88	92	101	89	90	93
列平均	87.1	87.7	89.2	85.1	87.45	87.15
环境平均	88			86.5667		
区组效应	-0.9	-0.3	1.2	-1.4667	0.8833	0.5833

各种变异的自由度列于下表第 2 列。环境内区组的自由度为 4、基因型的自由度为 19、环境的自由度为 1、互作的自由度为 19。总自由度为 119，因此得到误差自由度等于 76。各种变异的自由度列于下表第 2 列。环境内区组平方和等于上表中 6 种区组效应的平方和乘以基因型数 20。其它平方和与练习 9.2 相同。

第 3 列的平方和除以第 2 列的自由度，得到第 4 列给出的均方。从均方就可以计算  $F$  值、并进行显著性检验（略）。利用表 9.5 的期望均方，得到各种方差成分的估计值，列于下表第 5 列。本例中的互作效应不显著，可以认为互作方差为 0。利用公式 9.24 得到环境和重复平均数的遗传力为 0.9469。

变异来源	自由度	平方和	均方	方差估计值	遗传力
环境内区组	4	112.23			
基因型	19	1670.70	87.93	13.88	0.9469
环境	1	61.63	61.63	0.95	
互作	19	48.70	2.56	-0.70	
机误	76	355.10	4.67	4.67	
总和	119	2248.37			

**9.4** 下表是粳稻品种‘Asominori’与籼稻品种‘IR24’杂交产生的 20 个 RIL 家系（用 R1~R20 表示）在四个环境下两次重复的粒长（mm）数据，排在前两行的是两个亲本的表型观测值。

基因型	E1		E2		E3		E4	
	重复 1	重复 2						
Asominori	5.3	5.2	5.2	5.2	5.2	5.2	5.3	5.2
IR24	6.0	6.2	6.2	6.1	6.2	6.1	6.1	6.1
R1	5.4	5.5	5.5	5.4	5.4	5.4	5.6	5.4
R2	6.3	6.4	6.2	6.1	6.2	6.3	6.5	6.3
R3	5.3	5.4	5.4	5.3	5.3	5.2	5.3	5.4
R4	5.4	5.4	5.4	5.5	5.5	5.4	5.4	5.3
R5	5.3	5.4	5.4	5.3	5.4	5.3	5.3	5.3
R6	5.4	5.5	5.5	5.3	5.4	5.4	5.4	5.3
R7	5.5	5.5	5.5	5.4	5.3	5.5	5.5	5.4
R8	5.7	5.7	5.7	5.7	5.7	5.7	5.8	5.6

R9	5.1	5.2	5.2	5.2	5.2	5.1	5.1	5.1
R10	5.1	5.3	5.2	5.2	5.3	5.2	5.4	5.3
R11	5.0	4.8	4.7	4.9	4.8	4.8	5.5	5.3
R12	5.7	5.7	5.8	5.7	5.6	5.7	5.7	5.7
R13	5.0	4.9	4.9	5.0	4.9	5.1	4.9	4.9
R14	5.8	5.9	5.9	5.9	5.9	5.9	6.0	5.9
R15	5.4	5.6	5.5	5.5	5.6	5.5	5.5	5.6
R16	4.9	4.8	4.9	4.8	4.6	4.9	4.9	4.9
R17	5.6	5.6	5.7	5.6	5.5	5.7	5.5	5.4
R18	6.0	6.1	6.1	5.9	6.0	6.0	5.9	5.8
R19	5.0	4.9	4.9	5.1	5.1	4.9	5.1	5.0
R20	6.1	5.9	6.1	5.9	6.0	6.1	6.0	6.1

(1) 不考虑排在最前面的两个亲本，列出 20 种 RIL 基因型和 4 种环境粒长的 GE 双向表，在此基础上计算基因型的主效应和环境主效应，以及基因型和环境互作效应。

**解答：**下表第 2~5 列给出 20 种基因型和 4 种环境粒长的 GE 双向表。行平均其实是 RIL 在环境和重复间的平均表现，与总平均数的离差即为基因型的主效应。列平均其实是环境在 RIL 家系和重复间的平均表现，与总平均数的离差即为环境的主效应。GE 双向表中，每个 RIL 在每个环境的重复平均数与总平均数的离差可以看作总效应，从中减去基因型和环境的主效应，剩下部分就是互作效应。

RIL 家系	重复平均数				行平均	基因型效应	互作效应			
	E1	E2	E3	E4			E1	E2	E3	E4
R1	5.45	5.45	5.4	5.5	5.450	-0.011	-0.001	0.006	-0.034	0.029
R2	6.35	6.15	6.25	6.4	6.288	0.826	0.061	-0.131	-0.021	0.091
R3	5.35	5.35	5.25	5.35	5.325	-0.136	0.024	0.031	-0.059	0.004
R4	5.4	5.45	5.45	5.35	5.413	-0.049	-0.014	0.044	0.054	-0.084
R5	5.35	5.35	5.35	5.3	5.338	-0.124	0.011	0.019	0.029	-0.059
R6	5.45	5.4	5.4	5.35	5.400	-0.061	0.049	0.006	0.016	-0.071
R7	5.5	5.45	5.4	5.45	5.450	-0.011	0.049	0.006	-0.034	-0.021
R8	5.7	5.7	5.7	5.7	5.700	0.239	-0.001	0.006	0.016	-0.021
R9	5.15	5.2	5.15	5.1	5.150	-0.311	-0.001	0.056	0.016	-0.071
R10	5.2	5.2	5.25	5.35	5.250	-0.211	-0.051	-0.044	0.016	0.079
R11	4.9	4.8	4.8	5.4	4.975	-0.486	-0.076	-0.169	-0.159	0.404
R12	5.7	5.75	5.65	5.7	5.700	0.239	-0.001	0.056	-0.034	-0.021
R13	4.95	4.95	5	4.9	4.950	-0.511	-0.001	0.006	0.066	-0.071
R14	5.85	5.9	5.9	5.95	5.900	0.439	-0.051	0.006	0.016	0.029
R15	5.5	5.5	5.55	5.55	5.525	0.064	-0.026	-0.019	0.041	0.004
R16	4.85	4.85	4.75	4.9	4.838	-0.624	0.011	0.019	-0.071	0.041
R17	5.6	5.65	5.6	5.45	5.575	0.114	0.024	0.081	0.041	-0.146
R18	6.05	6	6	5.85	5.975	0.514	0.074	0.031	0.041	-0.146
R19	4.95	5	5	5.05	5.000	-0.461	-0.051	0.006	0.016	0.029
R20	6	6	6.05	6.05	6.025	0.564	-0.026	-0.019	0.041	0.004

列平均	5.463	5.455	5.445	5.483	5.461
环境效应	0.001	-0.006	-0.016	0.021	

(2) 不考虑重复的区组效应，对粒长进行多环境联合方差分析，并利用联合方差分析的结果，估计粒长性状平均表现（即环境和重复平均数）的遗传力。

**解答：**利用线性模型公式 9.7 进行方差分析。各种变异的自由度列于下表第 2 列。基因型的自由度为 19、环境的自由度为 3、互作的自由度为 57。总自由度为 159，因此得到误差自由度等于 80。各种变异的自由度列于下表第 2 列。基因型平方和等于（1）中 20 种基因型主效应的平方和乘以环境数 4 和重复数 2。环境平方和等于（1）中 4 种环境主效应的平方和乘以基因型 20 和重复数 2。互作平方和等于（1）中 80 种互作效应的平方和乘以重复数 2。总平方和等于 160 个观测值与总平均的离差平方和。

第 3 列的平方和除以第 2 列的自由度，得到第 4 列给出的均方。从均方就可以计算  $F$  值、并进行显著性检验（略）。利用表 9.4 的期望均方，得到各种方差成分的估计值，列于下表第 5 列。利用公式 9.24 得到环境和重复平均数的遗传力为 0.9884。

变异来源	自由度	平方和	均方	方差估计值	遗传力
基因型	19	22.9548	1.2081	0.1501	0.9884
环境	3	0.0303	0.0101	6.8E-05	
互作	57	0.8048	0.0141	0.0034	
机误	80	0.5900	0.0074	0.0074	
总和	159	24.3798			

**9.5** 对于练习 9.4 中的数据，现假定每个重复是被安排在一个土壤条件相对一致的区组内。因此，方差分析时需要考虑区组效应。考虑区组、基因型、环境、以及基因型和环境互作这 4 种变异来源，对于练习 9.4 中的数据作联合方差分析，并利用联合方差分析的结果，估计粒长性状平均表现（即环境和重复平均数）的遗传力。方差分析时不考虑两个亲本。

**解答：**利用线性模型公式 9.11 进行方差分析。首先说明一下区组效应的估计。每个环境中 有 2 个区组，因此共有 8 个区组。下表中的列平均即为这 8 个区组的平均数，它们与所在环境的总平均之间的离差，即为区组效应。这其实也是公式 9.11 中将区组效应记为  $B_{k/j}$  的含义所在。因此，多环境试验中，区组效应等于每个区组内所有观测值平均数与所在环境内观测值平均数的离差。离差计算时，不能用总平均。更不能在环境之间计算区组 I、区组 II 的效应。

RIL 家系	E1		E2		E3		E4	
	重复 1	重复 2						
R1	5.4	5.5	5.5	5.4	5.4	5.4	5.6	5.4

...	...	...	...	...	...	...	...	...
R20	6.1	5.9	6.1	5.9	6	6.1	6	6.1
列平均	5.45	5.475	5.475	5.435	5.435	5.455	5.515	5.45
环境平均	5.4625		5.455		5.445		5.4825	
区组效应	-0.0125	0.0125	0.02	-0.02	-0.01	0.01	0.0325	-0.0325

各种变异的自由度列于下表第 2 列。环境内区组的自由度为 4、基因型的自由度为 19、环境的自由度为 3、互作的自由度为 57。总自由度为 159，因此得到误差自由度等于 76。各种变异的自由度列于下表第 2 行。环境内区组平方和等于上表中 8 种区组效应的平方和乘以基因型数 20。其它平方和与练习 9.4 相同。

第 3 列的平方和除以第 2 列的自由度，得到第 4 列给出的均方。从均方就可以计算  $F$  值、并进行显著性检验（略）。利用表 9.5 的期望均方，得到各种方差成分的估计值，列于下表第 5 列。利用公式 9.24 得到环境和重复平均数的遗传力为 0.9884。

变异来源	自由度	平方和	均方	方差估计值	遗传力
环境内区组	4	0.0685			
基因型	19	22.9548	1.2081	0.1502	0.9884
环境	3	0.0303	0.0101	8.1E-05	
互作	57	0.8048	0.0141	0.0036	
机误	76	0.5215	0.0069	0.0069	
总和	159	24.3798			

**9.6** 下表是 14 个棉花品种在 8 个环境下的平均产量 ( $\text{kg}/\text{hm}^2$ )，试根据最后一列的回归系数，对 14 个品种的环境稳定程度进行分类。

品种	环境								平均	回归系数
	E1	E2	E3	E4	E5	E6	E7	E8		
1	761	962	912	955	779	826	1048	416	832.38	1.03
2	869	877	1020	914	719	671	986	405	807.63	0.99
3	657	875	924	905	654	783	909	384	761.38	1.01
4	696	932	983	962	829	825	997	405	828.63	1.05
5	603	920	788	881	671	866	989	526	780.50	0.78
6	741	921	992	1018	744	889	863	376	818.00	1.04
7	686	905	1029	979	800	881	858	463	825.13	0.90
8	776	1030	1003	972	783	877	1053	498	874.00	0.99
9	819	863	1047	939	729	699	1064	477	829.63	0.98
10	481	814	845	697	487	702	954	277	657.13	1.16
11	384	542	837	702	517	573	776	266	574.63	0.96
12	756	937	1138	1004	739	945	932	401	856.50	1.15
13	745	817	893	938	646	897	1165	402	812.88	1.12
14	791	947	923	831	781	785	992	455	813.13	0.85
环境指数	697.5	881.6	952.4	906.9	705.6	801.4	970.4	410.8		

**解答：**估计值在 0.95~1.05 之间时，认为回归系数与 1 无显著差异。这样的基因型包括品种 1~4、6、8、9 和 11，它们对环境的反应等于所有基因型的平均反应，具有中等水平的环境稳定性。估计值低于 0.95 时，认为回归系数显著低于 1。这样的基因型包括品种 5、7 和 14，它们具有较高的环境稳定性。估计值高于 1.05 时，认为回归系数显著高于 1。这样的基因型包括品种 10、12 和 13，它们具有较低的环境稳定性。

**9.7** 下表是 10 个黄花烟草品系在 8 个环境中的株高平均数，单位为 cm。利用适当的软件（如 QTL IciMapping 软件的 AOV 功能，Meng et al., 2015）开展 AMMI 分析，绘制 AMMI 双标图。

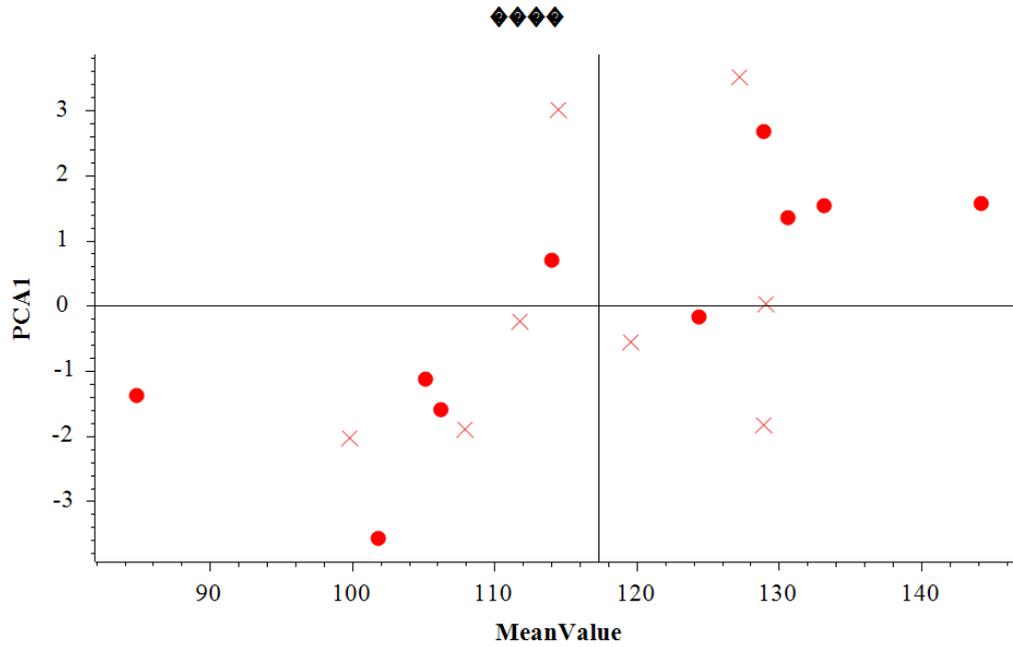
基因型	环境 1	环境 2	环境 3	环境 4	环境 5	环境 6	环境 7	环境 8
基因型 1	87.00	75.75	77.25	70.00	86.87	81.25	99.12	101.75
基因型 2	106.50	96.37	98.87	93.75	112.62	99.62	113.12	120.62
基因型 3	106.87	101.62	94.37	90.00	100.62	87.25	113.87	120.25
基因型 4	109.37	96.25	108.37	85.62	110.12	95.12	121.12	124.00
基因型 5	114.75	106.75	107.37	94.00	128.50	111.12	127.12	122.75
基因型 6	130.62	118.12	118.50	115.25	132.00	124.37	131.00	125.25
基因型 7	132.25	112.62	121.50	101.50	148.00	133.00	143.25	139.62
基因型 8	127.75	120.00	128.00	117.37	149.85	135.12	144.87	142.37
基因型 9	130.75	116.62	126.00	111.62	144.62	130.25	148.25	136.87
基因型 10	149.87	135.00	137.37	119.25	159.25	148.12	149.25	155.62

**解答：**每个基因型和每个环境的均值、以及 AMMI 分析的第 1 和 2 主成分得分如下表。

双标图附后。

基因型/环境	均值	第 1 主成分 PCA1	第 2 主成分 PCA2
G1	84.8737	-1.3789	-0.4677
G2	105.1837	-1.1254	0.5663
G3	101.8562	-3.5701	-0.0993
G4	106.2462	-1.5934	-2.2478
G5	114.045	0.7059	0.0212
G6	124.3887	-0.1654	3.4142
G7	128.9675	2.673	-1.4199
G8	133.1662	1.5332	0.4165
G9	130.6225	1.3495	-0.2302
G10	144.2162	1.5717	0.0468
E1	119.573	-0.5584	0.3507
E2	107.91	-1.8955	1.063
E3	111.76	-0.2337	-0.7737
E4	99.836	-2.0323	2.6675
E5	127.245	3.5109	-0.0826
E6	114.522	3.0056	1.0374

E7	129.097	0.0241	-1.7009
E8	128.91	-1.8208	-2.5614
Score	117.3566	0.5692	0.2017



## 第 10 章练习题和参考答案

**10.1** 遗传设计和环境设计的主要区别是什么？它们分别起什么作用？开展一项遗传交配设计的大概过程是什么？

**解答：**参看 10.1 节的有关内容。

**10.2** 通过QTL定位研究，发现一个控制数量性状座位上三种基因型AA、Aa、aa的均值分别为7、6、3。假定误差方差为1，不考虑其它遗传因素。以AA和aa的杂交F<sub>2</sub>为参照群体。

(1) 计算加性方差、显性方差、狭义遗传力和广义遗传力。

**解答：**该座位上的加性效应  $a=2$ ，和显性效应  $d=1$ 。因此，F<sub>2</sub> 群体中，

$$\text{加性方差 } V_A = a^2/2 = 2$$

$$\text{显性方差 } V_D = d^2/4 = 0.25$$

$$\text{遗传方差 } V_G = V_A + V_D = 2.25$$

$$\text{表型方差 } V_P = V_A + V_D + V_e = 3.25$$

进而得到，

$$\text{狭义遗传力 } h^2 = V_A/V_P = 0.6154$$

$$\text{广义遗传力 } h^2 = V_G/V_P = 0.6923$$

(2) 计算半同胞家系间和家系内的遗传方差。

**解答：**半同胞家系间的遗传方差等于  $V_A/2=1$ ，家系内的遗传方差等于  $2.25-1=1.25$ 。

(3) 计算全同胞家系间和家系内的遗传方差。

**解答：**全同胞家系间的遗传方差等于  $V_A/4+V_D/4=0.5625$ ，家系内的遗传方差等于  $2.25-0.5625=1.6875$ 。

**10.3** 下表是4个父本与4组不同母本的NCI交配设计，每组母本包含8个个体，每个杂交组合的后代有两个观测数据。参照群体的近交系数为0。通过固定效应模型的方差分析，估计参照群体的加性方差、显性方差和遗传力。

重复	父本	1	2	3	4	5	6	7	8
重复 I	1	5.4	6.4	6.6	4.2	8.6	6.0	6.4	5.7
	2	4.7	7.0	4.2	6.4	4.8	4.9	6.2	4.8
	3	3.8	7.2	4.9	4.3	7.2	4.0	4.8	4.1
	4	6.2	4.2	5.6	3.0	6.0	3.2	4.4	3.3
重复 II	1	6.6	7.9	7.5	5.4	8.5	7.4	5.8	6.3
	2	6.8	6.5	7.1	6.2	5.7	5.7	5.5	5.0
	3	6.8	7.6	6.0	3.5	5.9	4.6	5.8	4.8

4      6.6      4.2      5.6      2.0      8.0      4.9      5.7      2.8

**解答：**方差分析的结果如下表。

变异来源	自由度	平方和	均方	F 值	P 值
父本间	3	27.6850	9.2283	12.2382	1.69E-05
母本间/父本	28	76.0025	2.7144	3.5997	0.000318
剩余	32	24.1300	0.7541		
总和	63	127.8175			

固定效应和随机效应模型下，各种方差成分和遗传力的估计如下表。

	剩余方差	父本方差	母本方差	$V_A$	$V_D$	$h^2$	$H^2$
固定模型	0.7541	0.5296	0.9802	2.1186	1.8021	0.4532	0.8387
随机模型	0.7541	0.4071	0.9802	1.6285	2.2921	0.3484	0.8387

**10.4** 把练习10.3看作4个父本与8个母本的NCII交配设计。参照群体的近交系数为0。通过固定效应模型的方差分析，估计参照群体的加性方差、显性方差和遗传力。

**解答：**方差分析的结果如下表。

变异来源	自由度	平方和	均方	F 值	P 值
父本间	3	27.6850	9.2283	12.2382	1.69E-05
母本间	7	40.6075	39.9747	53.0124	5.17E-24
互作	21	35.3950	1.6855	2.2352	0.019551
剩余	32	24.1300	0.7541		

固定效应和随机效应模型下，各种方差成分和遗传力的估计如下表。

	剩余方差	父本方差	母本方差	互作方差	$V_A$	$V_D$	$h^2$	$H^2$
固定模型	0.7541	0.5296	4.9026	0.4657	10.8644	1.8628	0.8059	0.9441
随机模型	0.7541	0.4714	4.9026	0.4657	10.7480	1.8628	0.8042	0.9436

**10.5** 把练习10.3看作32个随机配对杂交产生的全同胞家系。参照群体的近交系数为0。通过固定效应模型的方差分析，估计全同胞家系间和家系内的方差。

**解答：**方差分析的结果如下表。

变异来源	自由度	平方和	均方	F 值	P 值
组合间	31	103.6875	3.3448	4.4357	3.26E-05
剩余	32	24.1300	0.7541		
总和	63	127.8175			

全同胞家系间和家系内的方差估计值分别为 0.7541 和 1.2953。

**10.6** 通过QTL定位研究，发现一个控制数量性状座位的中亲值 $m$ 、加性效应 $a$ 、显性效应 $d$ 分别为10、2、2。以 $P_1$  (AA) 和 $P_2$  (aa) 的杂交 $F_2$ 为参照群体，不考虑其它遗传和非遗传因素，下表给出三种基因型的均值、频率，以及参照群体的总平均和遗传方差。计算 $F_2$ 的三种测交家系的均值、以及这些家系的4个线性组合；同时计算每列的均值和方差。

F <sub>2</sub> 个体	均值	频率	L <sub>1</sub>	L <sub>2</sub>	L <sub>3</sub>	L <sub>1</sub> +L <sub>2</sub>	L <sub>1</sub> -L <sub>2</sub>	L <sub>1</sub> +L <sub>2</sub> +L <sub>3</sub>	L <sub>1</sub> +L <sub>2</sub> -2L <sub>3</sub>
AA	12	0.25							
Aa	12	0.5							
aa	8	0.25							
总平均	11								
遗传方差	3								

**解答:** 首先从 F<sub>2</sub> 测交家系的基因型构成, 计算三种测交家系 L<sub>1</sub>、L<sub>2</sub> 和 L<sub>3</sub> 的均值。然后, 计算 4 个线性组合的均值。在这些家系均值的基础上, 考虑三种基因型 AA、Aa、aa 的频率 0.25、0.5、0.25, 计算下表最后两行的总平均和遗传方差

F <sub>2</sub> 个体	均值	频率	L <sub>1</sub>	L <sub>2</sub>	L <sub>3</sub>	L <sub>1</sub> +L <sub>2</sub>	L <sub>1</sub> -L <sub>2</sub>	L <sub>1</sub> +L <sub>2</sub> +L <sub>3</sub>	L <sub>1</sub> +L <sub>2</sub> -2L <sub>3</sub>
AA	12	0.25	12	12	12	24	0	36	0
Aa	12	0.5	12	10	11	22	2	33	0
aa	8	0.25	12	8	10	20	4	30	0
总平均	11		12	10	11	22	2	33	0
遗传方差	3		0	2	0.5	2	2	4.5	0

**10.7** 从一个 F<sub>2</sub> 群体中随机抽取 8 个个体分别与双亲和 F<sub>1</sub> 进行测交, 下表是每个测交组合产生后代的两次重复观测数据。

F <sub>2</sub> 个体	重复 I			重复 II		
	L <sub>1</sub>	L <sub>2</sub>	L <sub>3</sub>	L <sub>1</sub>	L <sub>2</sub>	L <sub>3</sub>
1	10.43	9.33	13.63	12.68	11.20	11.32
2	11.89	11.74	10.55	10.52	10.07	10.41
3	12.11	7.38	9.88	11.33	7.99	9.80
4	10.17	9.75	10.67	11.61	9.62	11.91
5	10.20	12.25	10.31	12.01	11.81	10.68
6	13.45	9.70	10.35	10.70	8.52	10.95
7	10.68	9.15	10.75	11.86	8.72	9.59
8	13.94	9.98	11.34	12.27	10.30	10.39

(1) 利用方差分析估计误差方差。

**解答:** 利用两因素有交互方差分析模型。这里要求估计误差方差, 不对效应做进一步分解。每个 F<sub>2</sub> 个体与每个测交亲本的组合效应等于重复平均数减去总平均数。每个观测数据与对应的组合效应的离差, 就是误差效应。从而得到下表。

F <sub>2</sub> 个体	F <sub>2</sub> 个体 × 测交亲本效应			重复 I 误差效应			重复 II 误差效应		
	L1	L2	L3	L1	L2	L3	L1	L2	L3
1	0.807	-0.483	1.727	-1.125	-0.935	1.155	1.125	0.935	-1.155
2	0.457	0.157	-0.268	0.685	0.835	0.070	-0.685	-0.835	-0.070

3	0.972	-3.063	-0.908	0.390	-0.305	0.040	-0.390	0.305	-0.040
4	0.142	-1.063	0.542	-0.720	0.065	-0.620	0.720	-0.065	0.620
5	0.357	1.282	-0.253	-0.905	0.220	-0.185	0.905	-0.220	0.185
6	1.327	-1.638	-0.098	1.375	0.590	-0.300	-1.375	-0.590	0.300
7	0.522	-1.813	-0.578	-0.590	0.215	0.580	0.590	-0.215	-0.580
8	2.357	-0.608	0.117	0.835	-0.160	0.475	-0.835	0.160	-0.475

根据上表中的效应，计算下表的平方和。组合平方和等于上表组合效应的平方和乘以重复数 2，所有误差效应平方和就是剩余平方和。因此，误差方差的估计值位 0.8924。

变异来源	自由度	平方和	均方	F 值	P 值
F <sub>2</sub> 个体 × 测交亲本的组合	23	65.7768	2.8599	3.2048	0.0031
剩余项	24	21.4171	0.8924		
总和	47	87.1938			

(2) 利用NCIII设计的分析方法，估计F<sub>2</sub>群体的加性方差、显性方差、狭义遗传力和广义遗传力等遗传参数。

(3) 利用TTC设计的分析方法，估计F<sub>2</sub>群体的加性方差、显性方差、狭义遗传力和广义遗传力等遗传参数。

**解答：**利用两因素有互作方差分析模型。这里要求估计误差方差，不对效应做进一步分解。每个 F<sub>2</sub> 个体与每个测交亲本的组合效应等于重复平均数减去总平均数。每个观测数据与对应的组合效应的离差，就是误差效应。从而得到下表。

F <sub>2</sub> 个体	L1	L2	L3	L1+L2	L1-L2	L1+L2+L3	L1+L2-2L3
1	11.555	10.265	12.475	21.820	1.290	34.295	-3.130
2	11.205	10.905	10.480	22.110	0.300	32.590	1.150
3	11.720	7.685	9.840	19.405	4.035	29.245	-0.275
4	10.890	9.685	11.290	20.575	1.205	31.865	-2.005
5	11.105	12.030	10.495	23.135	-0.925	33.630	2.145
6	12.075	9.110	10.650	21.185	2.965	31.835	-0.115
7	11.270	8.935	10.170	20.205	2.335	30.375	-0.135
8	13.105	10.140	10.865	23.245	2.965	34.110	1.515
均值	11.616	9.844	10.783	21.460	1.771	32.243	-0.106
方差	0.5027	1.7458	0.6549	1.8858	2.6113	3.2182	3.1503

利用上表中的方差与遗传方差的关系，得到不同设计的遗传方差估计，进而计算遗传力。

结果如下表。

交配设计	加性方差	显性方差	误差方差	$h^2$	$H^2$
NCIII	1.8858	1.3056	0.8924	0.4618	0.7815
TTC	1.4303	1.3056	0.8924	0.3942	0.7541

**10.8** 下表是10个双亲RIL家系和10个永久F<sub>2</sub>在某地点的三次重复产量数据。利用表中数据的方差分析结果，估计F<sub>2</sub>参照群体的加性方差、显性方差、狭义遗传力和广义遗传力等遗传参数。

传参数。

RIL 家系	重复 I	重复 II	重复 III	永久 F <sub>2</sub> 家系	重复 I	重复 II	重复 III
1	4.09	4.19	4.01	1	7.46	8.10	7.80
2	1.35	2.12	2.81	2	7.39	8.12	6.33
3	3.89	3.20	3.15	3	6.95	6.36	5.45
4	3.83	4.73	4.33	4	6.87	8.88	8.93
5	3.66	3.71	3.89	5	7.46	7.23	7.11
6	4.27	3.17	4.32	6	7.19	5.99	8.10
7	3.73	3.65	1.88	7	6.99	6.33	6.40
8	3.33	3.30	3.38	8	7.51	6.29	7.62
9	4.06	3.90	4.29	9	6.48	7.14	5.85
10	2.89	1.55	2.12	10	4.73	5.41	5.70

**解答：**两个群体的方差分析结果如下表。

群体类型	均方		F 值	方差估计值		广义遗传力	
	基因型	随机误差		基因型	随机误差	一次重复	重复平均数
RIL	1.8088	0.2939	6.1544	0.5050	0.2939	0.6321	0.8375
永久 F <sub>2</sub>	2.0741	0.5357	3.8719	0.5128	0.5357	0.4891	0.7417

RIL 群体的遗传方差是加性方差的 2 倍，因此得到加性方差的估计为 0.2525。永久 F<sub>2</sub> 群体等于加性方差与显性方差之和，因此得到显加性方差的估计值 0.2603。永久 F<sub>2</sub> 群体误差方差的估计值 0.5357。由此得到狭义和广义遗传力为 0.2408 和 0.4891。

## 第 11 章练习题和参考答案

**11.1** 一个随机交配群体中，某性状的遗传力 $h^2=0.37$ ，表型方差 $V_p=10.7$ 。如根据个体的表型进行截尾选择，计算不同选择比例下的选择强度、选择差和预期遗传进度，并将计算结果填在下面的表格中。

选择比例 $p/\%$	选择强度 $i$	选择差 $S$	预期遗传进度 $R$
1			
5			
25			
50			

**解答：**结果如下，过程略去。

选择比例 $p/\%$	选择强度 $i$	选择差 $S$	预期遗传进度 $R$
1	2.6652	8.7181	3.2257
5	2.0627	6.7473	2.4965
25	1.2711	4.1579	1.5384
50	0.7979	2.6099	0.9657

**11.2** 遗传进度或选择响应的估计公式为 $R = ih\sqrt{V_A}$ ，其中  $i$  表示选择比例为  $p$  时的选择强度， $h$  为遗传力的平方根， $V_A$  是加性方差。依据此公式，说明育种中提高遗传进度的途径有哪些？

**解答：**参看 §11.1.3。

**11.3** 一个双亲  $F_2$  群体中，存在控制某性状的三个独立遗传基因座位，加性效应均为 1，显性效应分别为 1、0.5 和 0。假定中亲值等于 10，表型随机误差方差等于 4。现从 1000 个  $F_2$  个体中，根据个体表型选择最高的 30 个个体。

(1) 计算  $F_2$  群体的均值、加性方差、表型方差、狭义遗传力和广义遗传力。

**解答：** $F_2$  群体的均值等于中亲与三个显性效应一半之和，即  $10+0.5+0.25+0=10.75$ 。根据已知的加、显性效应，首先计算每个座位上的加、显性方差，相加就得到性状的加、显性方差，分别等于 1.5 和 0.3125。表型方差等于加、显性方差与误差方差之和，即 5.8125。从各种方差成分，得到狭义遗传力和广义遗传力分别为 0.2581 和 0.3118。

(2) 计算选择差和中选个体的均值。

**解答：**选择比例等于  $30/1000=0.03$ ，因此选择强度等于 2.2681。利用公式 11.6 得到选

择差等于 5.4681，利用 11.2 得到中选个体的平均表型等于 16.2182。

(3) 计算遗传进度和中选个体随机交配后代的期望均值。

**解答：**利用 (1) 和 (2) 的结果，得到的遗传进度为 1.4111。中选个体随机交配后代的期望均值等于 F<sub>2</sub> 的群体均值加上遗传进度，即 12.1612。

**11.4** 一个遗传群体中，个体的表型相对于群体均值用  $P_1$  表示，亲属表型相对于群体均值用  $P_2$  表示，其中包含的育种值用  $A_1$  和  $A_2$  表示。表型和育种值的方差协方差如下。试构建个体育种值  $A_1$  的最优选择指数。

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

$$\text{Cov}(A_1, A_1) = h^2V_p, \quad \text{Cov}(A_2, A_1) = rV_A = rh^2V_p$$

**解答：**这里的正规方程是，

$$\begin{pmatrix} V_p & tV_p \\ tV_p & V_p \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} h^2V_p \\ rh^2V_p \end{pmatrix}, \quad \text{即} \begin{pmatrix} 1 & t \\ t & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} h^2 \\ rh^2 \end{pmatrix}$$

由于

$$\begin{pmatrix} 1 & t \\ t & 1 \end{pmatrix}^{-1} = \frac{1}{1-t^2} \begin{pmatrix} 1 & -t \\ -t & 1 \end{pmatrix}$$

因此

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

最优选择指数为

$$I = b_1P_1 + b_2P_2 = \frac{1}{1-t^2} [(1-rt)h^2P_1 + (r-t)h^2P_2]$$

把个体表型  $P_1$  的权重矫正为 1，则最优选择指数为，

$$I = P_1 + \frac{r-t}{1-rt} P_2$$

**11.5** 下表是一个玉米随机交配群体中，油分含量和蛋白含量的方差协方差估计值。

性状名称	表型方差协方差矩阵	加性遗传方差协方差矩阵
------	-----------	-------------

	油分含量	蛋白质含量	油分含量	蛋白质含量
油分含量	287.5	477.4	128.7	160.6
蛋白质含量	477.4	935.0	160.6	254.6

(1) 计算两个性状的遗传力。

**解答：**油份含量和蛋白含量的遗传力分别为 0.4477 和 0.2723。

(2) 计算两个性状的表型相关系数和遗传相关系数。

**解答：**油份含量和蛋白含量的表型相关系数等于 0.9208，遗传相关系数等于 0.8872。

(3) 如把加性效应之外的遗传效应也归入环境方差，计算两个性状的环境方差协方差矩阵，以及环境相关系数。

**解答：**从表型方差或协方差中，减去加性方差或协方差，就得到环境方差或协方差。结果如下。油份含量和蛋白含量的环境相关系数等于 0.9638。

性状名称	环境方差协方差	
	油分含量	蛋白质含量
油分含量	158.8	316.8
蛋白质含量	316.8	680.4

(4) 计算蛋白质含量选择比例为 5% 的蛋白质含量遗传进度。

**解答：**直接选择比例为 5% 时，蛋白质含量的遗传进度等于 17.17。

(5) 计算油分选择比例为 5% 的蛋白质含量遗传进度，并与之前的直接进度相比较。

**解答：**对油分的选择比例为 5%，蛋白质含量的间接遗传进度等于 19.54。油份含量有较高的遗传力、与蛋白含量的遗传相关也比较高。因此，对蛋白质含量来说，间接选择的遗传进度超过了直接选择的遗传进度。

**11.6** 利用练习 11.5 的数据，构造改良油分含量的最优选择指数，并计算指数选择相对于个体选择的优势。

**解答：**油份含量和蛋白含量的表型分别用  $P_1$  和  $P_2$  表示，其中包含的育种值用  $A_1$  和  $A_2$  表示。改良油份含量最优选择指数的正规方程表示为

$$Cov(\mathbf{P}, \mathbf{P})\mathbf{b} = Cov(\mathbf{A}, \mathbf{A}_1)$$

其中，

$$Cov(\mathbf{P}, \mathbf{P}) = \begin{pmatrix} 287.5 & 477.4 \\ 477.4 & 935 \end{pmatrix}, \quad Cov(\mathbf{P}, \mathbf{P})^{-1} = \begin{pmatrix} 0.0229 & -0.0117 \\ -0.0117 & 0.0070 \end{pmatrix},$$

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \quad \text{Cov}(\mathbf{A}, A_1) = \begin{pmatrix} 128.7 \\ 160.6 \end{pmatrix}$$

因此

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 0.0229 & -0.0117 \\ -0.0117 & 0.0070 \end{pmatrix} \begin{pmatrix} 128.7 \\ 160.6 \end{pmatrix} = \begin{pmatrix} 1.0675 \\ -0.3733 \end{pmatrix}$$

改良油份含量的最优选择指数为  $I=1.0675P_1-0.3733P_2$ 。最优选择指数遗传进度是个体选择的 1.1594 倍。具体计算过程如下。

遗传参数	估计值
油份含量遗传力平方根 $h$	0.6691
选择指数的方差 $V_I$ (公式 11.38)	77.4388
选择指数的遗传力 $h_I^2$ (公式 11.40)	0.6017
选择指数与油份含量的遗传相关 $r_{IH}$ (公式 11.41)	0.7757
相对遗传进度 $r_{IH}/h$ (公式 11.42)	1.1594

**11.7** 利用练习 11.5 的数据，构造改良蛋白质含量的最优选择指数，并计算指数选择相对于个体选择的优势。

**解答：**油份含量和蛋白含量的表型分别用  $P_1$  和  $P_2$  表示，其中包含的育种值用  $A_1$  和  $A_2$  表示。改良蛋白质含量最优选择指数的正规方程表示为

$$\text{Cov}(\mathbf{P}, \mathbf{P})\mathbf{b} = \text{Cov}(\mathbf{A}, A_2)$$

其中，

$$\text{Cov}(\mathbf{P}, \mathbf{P}) = \begin{pmatrix} 287.5 & 477.4 \\ 477.4 & 935 \end{pmatrix}, \quad \text{Cov}(\mathbf{P}, \mathbf{P})^{-1} = \begin{pmatrix} 0.0229 & -0.0117 \\ -0.0117 & 0.0070 \end{pmatrix},$$

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \quad \text{Cov}(\mathbf{A}, A_2) = \begin{pmatrix} 160.6 \\ 254.6 \end{pmatrix}$$

因此

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 0.0229 & -0.0117 \\ -0.0117 & 0.0070 \end{pmatrix} \begin{pmatrix} 160.6 \\ 254.6 \end{pmatrix} = \begin{pmatrix} 0.6996 \\ -0.0849 \end{pmatrix}$$

改良蛋白质含量的最优选择指数为  $I=0.6996P_1-0.0849P_2$ 。

最优选择指数遗传进度是个体选择的 1.1440 倍。具体计算过程如下。

遗传参数	估计值
蛋白质含量遗传力平方根 $h$	0.5218
选择指数的方差 $V_I$ (公式 11.38)	90.7382

选择指数的遗传力 $h_I^2$ (公式 11.40)	0.3564
选择指数与蛋白质含量的遗传相关 $r_{IH}$ (公式 11.41)	0.5970
相对遗传进度 $r_{IH}/h$ (公式 11.42)	1.1440

**11.8** 利用练习 11.5 的数据，假定油分和蛋白质的经济权重分别为 1 和 0.6，构造提高总收益的最优选择指数；假定利用此指数开展选择工作，在选择比例为 5% 的条件下，计算总收益的遗传进度、以及油分和蛋白两个性状的遗传进度。

**解答：**油份含量和蛋白含量的表型分别用  $P_1$  和  $P_2$  表示，其中包含的育种值用  $A_1$  和  $A_2$  表示。改良总收益最优选择指数的正规方程表示为

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

其中，

$$\text{Cov}(\mathbf{P}, \mathbf{P}) = \begin{pmatrix} 287.5 & 477.4 \\ 477.4 & 935 \end{pmatrix}, \quad \text{Cov}(\mathbf{P}, \mathbf{P})^{-1} = \begin{pmatrix} 0.0229 & -0.0117 \\ -0.0117 & 0.0070 \end{pmatrix},$$

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \quad \text{Cov}(\mathbf{A}, \mathbf{A}) = \begin{pmatrix} 128.7 & 160.6 \\ 160.6 & 254.6 \end{pmatrix}, \quad \mathbf{w} = \begin{pmatrix} 1 \\ 0.6 \end{pmatrix}$$

因此

$$\begin{aligned} \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} &= \begin{pmatrix} 0.0229 & -0.0117 \\ -0.0117 & 0.0070 \end{pmatrix} \begin{pmatrix} 128.7 & 160.6 \\ 160.6 & 254.6 \end{pmatrix} \begin{pmatrix} 1 \\ 0.6 \end{pmatrix} \\ &= \begin{pmatrix} 1.0675 & 0.6996 \\ -0.3733 & -0.0849 \end{pmatrix} \begin{pmatrix} 1 \\ 0.6 \end{pmatrix} = \begin{pmatrix} 1.4873 \\ -0.4243 \end{pmatrix} \end{aligned}$$

改良总收益的最优选择指数为  $I=1.4873P_1-0.4243P_2$ 。

在选择比例为 5% 的条件下，总收益的遗传进度等于 29.30，油份和蛋白的遗传进度分别为 17.90 和 19.00。具体计算过程如下。

遗传参数	估计值
选择指数的方差 $V_I$ (公式 11.61)	201.7875
选择指数的遗传进度 $R_H$ (公式 11.64)	29.30123
选择指数与油份含量的遗传协方差 $\text{Cov}(I, A_1)$ (公式 11.65)	123.2803
选择指数与蛋白质含量的遗传协方差 $\text{Cov}(I, A_2)$ (公式 11.65)	130.8453
油份含量的遗传进度 $R_1$ (公式 11.67)	17.90133
蛋白质含量的遗传进度 $R_2$ (公式 11.67)	18.99984

## 第 12 章练习题和参考答案

**12.1** 纯系品种选育过程中，什么样的 $F_1$ 需要与哪个亲本回交之后，再进行连续自交和选择？如何确定一个适当的回交代数？

**解答：**参看 §12.1。

**12.2** 通过QTL定位研究，发现一个控制数量性状座位上的三种基因型AA、Aa、aa的均值分别为7、6、3。现将基因型为AA和aa的两个亲本杂交，不考虑选择、以及其它遗传和非遗传因素。

(1) 计算 $F_2$ 的群体均值和遗传方差。

**解答：**三种基因型 AA、Aa、aa 的均值分别为 7、6、3，在  $F_2$  群体中的频率分别为 0.25、0.5、0.25，因此得到  $F_2$  的群体均值和遗传方差分别为 5.5 和 2.25。

(2) 计算 $F_3$ 的群体均值和遗传方差，并给出相对于 $F_2$ 的近交衰退程度。

**解答：**三种基因型 AA、Aa、aa 在  $F_3$  群体中的频率分别为 0.375、0.25、0.375，因此得到  $F_3$  的群体均值和遗传方差分别为 5.25 和 3.1875。

与  $F_2$  相比较， $F_3$  的群体均值下降了 0.25。这里的显性效应  $d=1$ ，基因频率  $p=q=0.5$ ，近交系数  $F=0.5$ ，根据公式 12.17 得到的近交衰退也正好等于 0.25。

(3) 计算重组近交家系的群体均值和遗传方差，并给出相对于 $F_2$ 的近交衰退程度。

**解答：**三种基因型 AA、Aa、aa 在重组近交家系群体中的频率分别为 0.5、0、0.5，因此得到重组近交家系的群体均值和遗传方差分别为 5 和 4。

与  $F_2$  相比较，重组近交家系的群体均值下降了 0.5。这里的显性效应  $d=1$ ，基因频率  $p=q=0.5$ ，近交系数  $F=1$ ，根据公式 12.17 得到的近交衰退也正好等于 0.5。

**12.3** 练习12.2中，现将基因型为AA和aa的两个亲本杂交，然后通过连续自交产生了一个重组近交家系群体。不考虑选择、以及其它遗传和非遗传因素。这些近交系之间的杂交 $F_1$ 有哪些类型？给出每种类型的频率、中亲值和 $F_1$ 杂种的表现，并计算 $F_1$ 杂种的群体均值和遗传方差。

**解答：**通过连续自交产生的重组近交家系群体中，两种纯合基因型 AA 和 aa 的频率各占 0.5。根据两个自交系亲本的基因型，可以将它们之间的杂交组合分为 4 类，各占 0.25。利用两个亲本自交系的基因型值，就能计算中亲值。利用两个亲本自交系的基因型，可以推测杂种  $F_1$  的基因型，进而得到杂种  $F_1$  的表现。

交配类型	频率	第一亲本	第二亲本	中亲值	F <sub>1</sub> 表现
AA×AA	0.25	7	7	7	7
AA×aa	0.25	7	3	5	6
aa×AA	0.25	3	7	5	6
aa×aa	0.25	3	3	3	3
均值					5.5
方差					2.25

根据 4 类 F<sub>1</sub> 杂种的频率和平均表现，得到的群体均值和遗传方差分别为 5.5 和 2.25，正好等于练习 12.2 中 F<sub>2</sub> 的群体均值和遗传方差。

**12.4** 下表给出的是一个随机交配群体中，三个座位上的等位基因频率和加显性遗传效应。不考虑其它遗传因素，假定总的中亲值为10，三个座位间的不平衡度为0。计算近交系数分别为0、0.1、0.2、0.5和1的群体均值。

座位	<i>p</i>	<i>q</i>	<i>a</i>	<i>d</i>
A	0.6	0.4	2	1
B	0.3	0.7	3	2
C	0.8	0.2	1	2

**解答：**根据公式 12.13 和公式 12.17，计算每个座位对群体均值和近交衰退的贡献，结果如下表最后两列。

座位	<i>p</i>	<i>q</i>	<i>a</i>	<i>d</i>	$(p-q)a+2pqd$	$-2pqd$
座位 A	0.6	0.4	2	1	0.88	-0.48
座位 B	0.3	0.7	3	2	-0.36	-0.84
座位 C	0.8	0.2	1	2	1.24	-0.64
总和					1.76	-1.96

利用公式 12.15 得到不同近交系数的群体均值，结果如下。

近交系数	0	0.1	0.2	0.5	1
群体均值	11.76	11.564	11.368	10.78	9.8

**12.5** 通过QTL定位研究，发现了两个控制数量性状的座位。两个座位之间无连锁关系，下表给出9种基因型值。不考虑其它遗传因素和误差效应。

基因型	AABB								
平均表现	2.1	9.8	1.5	4.7	6.8	9.6	3.5	8.1	9.3

(1) 计算两个座位的加显性效应、以及它们之间的各种互作效应。

**解答：**两个座位的加显性效应、以及它们之间的各种互作效应如下表。

遗传参数	估计值
<i>m</i>	4.1
<i>a</i> <sub>1</sub>	-2.3
<i>d</i> <sub>1</sub>	3.05

$a_2$	-1.3
$d_2$	4.85
$aa$	1.6
$ad$	3.15
$da$	-1.15
$dd$	-5.2

(2) 利用4个重组近交系AABB、AAbb、aaBB、aabb作为父母本，配置16个F<sub>1</sub>杂交组合。给出每个组合的中亲值、基因型值、中亲优势和高亲优势。

**解答：**16种F<sub>1</sub>杂交组合的中亲值、基因型值、中亲优势和高亲优势如下表。

亲本基因型		亲本基因型值		中亲值	杂种 F <sub>1</sub>	中亲优势	高亲优势
亲本 1	亲本 2	亲本 1	亲本 2				
AABB	AABB	2.1	2.1	2.1	2.1	0	0
AABB	AAbb	2.1	1.5	1.8	9.8	8	7.7
AABB	aaBB	2.1	3.5	2.8	4.7	1.9	1.2
AABB	aabb	2.1	9.3	5.7	6.8	1.1	-2.5
AAbb	AABB	1.5	2.1	1.8	9.8	8	7.7
AAbb	AAbb	1.5	1.5	1.5	1.5	0	0
AAbb	aaBB	1.5	3.5	2.5	6.8	4.3	3.3
AAbb	aabb	1.5	9.3	5.4	9.6	4.2	0.3
aaBB	AABB	3.5	2.1	2.8	4.7	1.9	1.2
aaBB	AAbb	3.5	1.5	2.5	6.8	4.3	3.3
aaBB	aaBB	3.5	3.5	3.5	3.5	0	0
aaBB	aabb	3.5	9.3	6.4	8.1	1.7	-1.2
aabb	AABB	9.3	2.1	5.7	6.8	1.1	-2.5
aabb	AAbb	9.3	1.5	5.4	9.6	4.2	0.3
aabb	aaBB	9.3	3.5	6.4	8.1	1.7	-1.2
aabb	aabb	9.3	9.3	9.3	9.3	0	0

(3) 在(2)的基础上，计算单个亲本、中亲、F<sub>1</sub>表现、中亲优势和高亲优势这5个参数之间的相关系数矩阵。

**解答：**利用 EXCEL 中的相关系数函数 CORREL，得到的相关系数矩阵如下表。

相关系数	亲本 1	亲本 2	中亲值	杂种 F <sub>1</sub>	中亲优势	高亲优势
亲本 1	1.0000	0.0000	0.7071	<b>0.3278</b>	-0.2647	-0.4239
亲本 2		1.0000	0.7071	<b>0.3278</b>	-0.2647	-0.4239
中亲			1.0000	<b>0.4636</b>	-0.3743	-0.5995
杂种 F <sub>1</sub>				<b>1.0000</b>	<b>0.6481</b>	<b>0.3232</b>
中亲优势					1.0000	0.8535
高亲优势						1.0000

(4) 根据(3)计算的相关系数矩阵，说明植物育种中如何更好地预测优良杂交组合。

**解答：**杂交种的育种目标当然是提高杂种 F<sub>1</sub> 的表现。从(3)的相关系数矩阵可以看出，单个亲本、中亲、中亲优势和高亲优势均与杂种 F<sub>1</sub> 的表现有正的相关关系。但相关程度有

很大差异，按照从高到低的顺序是中亲优势 ( $r=0.6481$ )、中亲值 ( $r=0.4636$ )、单亲值 ( $r=0.3278$ )、高亲优势 ( $r=0.3232$ )。因此，这里的中亲优势是优良杂交组合预测的最好方法，其次是中亲值。值得一提的是，这些结论是从题目给出的基因型值得出的。对于其它不同的基因型值，可能得到不完全一样的结论。

**12.6** 在练习12.5的基础上，如将4种纯合基因型看作4个自交系亲本、同时又将它们看作4个测交亲本进行杂交，16个杂种 $F_1$ 的基因型值列成下面的双向表。表中的行平均可以看作自交系亲本的一般配合力(GCA)。下表后半部分还同时给出了每个测交亲本的均值和标准差、测交表现与自交系亲本GCA的相关系数。最后一行的综合指标等于测交亲本的均值与标准差之和。

自交系亲本	测交亲本				GCA
	<i>AABB</i>	<i>AAbb</i>	<i>aaBB</i>	<i>aabb</i>	
<i>AABB</i>	2.1	9.8	4.7	6.8	5.85
<i>AAbb</i>	9.8	1.5	6.8	9.6	6.925
<i>aaBB</i>	4.7	6.8	3.5	8.1	5.775
<i>aabb</i>	6.8	9.6	8.1	9.3	8.45
均值	5.85	6.925	5.775	8.45	
标准差	2.8236	3.3492	1.7880	1.1057	
与 GCA 的相关系数	0.5568	0.0781	0.9458	0.7258	
综合指标	8.6736	10.2742	7.5630	9.5557	

(1) 根据GCA判断最优的自交系亲本。

**解答：**在这4个自交系亲本中，基因型*aabb*的GCA最高，因此是最优的自交系亲本。但是，这个自交系产生4个杂种 $F_1$ 的最高表现是9.6，而所有16个杂种 $F_1$ 的最高表现是9.8。因此，特殊配合力最高的组合不一定存在于一般配合力最高的亲本自交系中。杂交种选育中，既要注重一般配合力的选择，又要注重特殊配合力的选择。

(2) 根据表中数据说明应该利用什么样的标准来选择最优测交亲本。

**解答：**从表中数据可以看到，基因型*aabb*作为测交亲本的杂种后代群体均值最高，基因型*AAbb*的测交后代群体的变异程度最高。一个育种群体的价值，既依赖于群体的均值、又依赖于群体的变异程度。最后一行的综合指标，既考虑了群体的均值、又考虑了群体的变异程度。如以这个综合指标为标准的话，则基因型*AAbb*是最好的测交亲本。利用这样的自交系作测交亲本，能够更好地反映出亲本自交系之间的遗传差异，同时又有利于选择到优异的杂交种。本例中，如用*AAbb*做测交亲本，就可能从4个自交系亲本中选育出具有最高表现9.8的杂交种*AABb*。

基因型  $aaBB$  的后代虽然与自交系的 GCA 相关程度最高，但杂种后代群体的均值和方差都不是最高的，最高表现 9.8 的杂种  $AABb$  并不存在于它的测交后代中。因此，测交后代与 GCA 的相关程度，不宜作为测交亲本优劣的选择标准。

12.7 下表是7个纯系亲本、49个杂交组合的两次重复观测数据。考虑两因素带交互、固定效应的方差分析模型。

父本	重复 I							重复 II						
	A	B	C	D	E	F	G	A	B	C	D	E	F	G
A	22.8	14.4	27.2	17.2	18.3	16.2	18.6	24.2	16.2	30.8	27.0	20.2	16.8	14.4
B	15.4	17.2	14.8	18.6	15.2	17.0	14.4	16.5	18.8	14.6	18.6	15.3	15.2	14.8
C	31.8	21.0	24.8	24.6	19.2	29.8	12.8	30.4	23.0	21.2	25.4	20.0	28.4	14.2
D	16.2	11.4	16.8	18.4	12.4	16.8	12.6	17.8	13.0	16.3	18.0	14.2	14.8	12.2
E	14.6	12.2	15.2	15.2	15.2	18.0	10.4	18.8	13.6	15.4	13.8	15.2	16.0	12.2
F	20.2	14.2	18.6	22.2	14.3	20.2	9.0	23.4	14.0	14.8	17.0	17.3	22.6	10.2
G	14.0	12.2	13.6	13.8	15.6	15.6	11.4	16.6	9.2	16.2	14.4	15.6	11.0	10.6

(1) 给出父本一般配合力、母本一般配合力和特殊配合力的估计值。

**解答：**把重复平均数列为下面的双向表，并计算行平均、列平均、行效应和列效应。行效应即为自交系作为父本的一般配合力，列效应即为自交系作为母本的一般配合力。

父本	母本							行平均	行效应
	A	B	C	D	E	F	G		
A	23.50	15.30	29.00	22.10	19.25	16.50	16.50	20.31	3.15
B	15.95	18.00	14.70	18.60	15.25	16.10	14.60	16.17	-0.99
C	31.10	22.00	23.00	25.00	19.60	29.10	13.50	23.33	6.17
D	17.00	12.20	16.55	18.20	13.30	15.80	12.40	15.06	-2.10
E	16.70	12.90	15.30	14.50	15.20	17.00	11.30	14.70	-2.46
F	21.80	14.10	16.70	19.60	15.80	21.40	9.60	17.00	-0.16
G	15.30	10.70	14.90	14.10	15.60	13.30	11.00	13.56	-3.60
列平均	20.19	15.03	18.59	18.87	16.29	18.46	12.70	17.16	
列效应	3.03	-2.13	1.43	1.71	-0.88	1.30	-4.46		

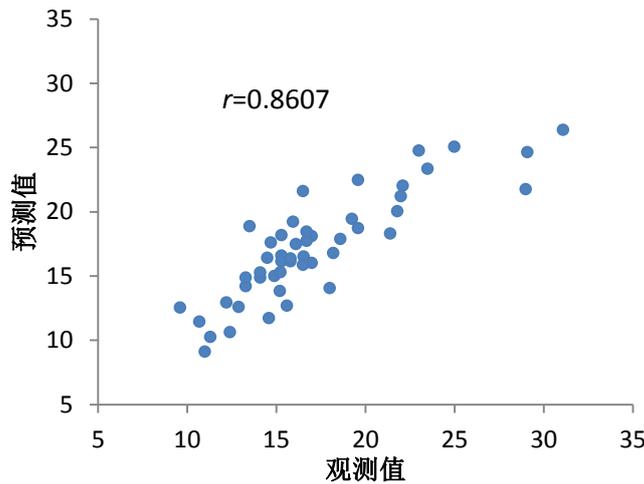
父本和母本之间的交互，即为特殊配合力，结果如下表。

父本	母本						
	A	B	C	D	E	F	G
A	0.16	-2.87	7.26	0.08	-0.18	-5.10	0.65
B	-3.25	3.96	-2.90	0.72	-0.05	-1.37	2.89
C	4.74	0.80	-1.76	-0.04	-2.85	4.48	-5.37
D	-1.10	-0.73	0.05	1.43	-0.89	-0.56	1.80
E	-1.03	0.33	-0.83	-1.91	1.38	1.00	1.06
F	1.77	-0.77	-1.73	0.89	-0.32	3.10	-2.94
G	-1.29	-0.72	-0.09	-1.17	2.92	-1.55	1.90

(2) 利用亲本的一般配合力，预测这49个组合的表现，绘制预测值与观察值重复平均数的散点图，并计算二者之间的相关系数。

**解答：**利用(1)计算出的一般配合力和特殊配合力，得到这49个组合的预测值、以及预测值与观测值的散点图如下，预测值与观察值之间的相关系数标于图中。

父本	母本						
	A	B	C	D	E	F	G
A	23.34	18.17	21.74	22.02	19.43	21.60	23.34
B	19.20	14.04	17.60	17.88	15.30	17.47	19.20
C	26.36	21.20	24.76	25.04	22.45	24.62	26.36
D	18.10	12.93	16.50	16.77	14.19	16.36	18.10
E	17.73	12.57	16.13	16.41	13.82	16.00	17.73
F	20.03	14.87	18.43	18.71	16.12	18.30	20.03
G	16.59	11.42	14.99	15.27	12.68	14.85	16.59



(3) 给出父本一般配合力、母本一般配合力和特殊配合力方差的估计值。

**解答：**方差分析的部分结果如下表，利用固定效应模型得到的方差估计值列于下表最后一列。

来源	自由度	平方和	均方	方差估计(固定模型)
父本配合力	6	1013.36	168.89	11.83
母本配合力	6	574.87	95.81	6.61
特殊配合力	36	555.59	15.43	6.08
误差	49	160.78	3.28	3.28
总和	97	2304.59		

**12.8** 下表是练习12.7的重复平均数，最后一列为每个亲本所在行和列的平均，最后一行为总平均。不区分亲本的一般配合力。

重复平均	A	B	C	D	E	F	G	行列平均
A	23.50	15.30	29.00	22.10	19.25	16.50	16.50	20.25
B	15.95	18.00	14.70	18.60	15.25	16.10	14.60	15.60
C	31.10	22.00	23.00	25.00	19.60	29.10	13.50	20.96
D	17.00	12.20	16.55	18.20	13.30	15.80	12.40	16.97
E	16.70	12.90	15.30	14.50	15.20	17.00	11.30	15.49
F	21.80	14.10	16.70	19.60	15.80	21.40	9.60	17.73
G	15.30	10.70	14.90	14.10	15.60	13.30	11.00	13.13
总平均	17.16							17.16

(1) 给出7个亲本的一般配合力以及49个组合特殊配合力的估计值。

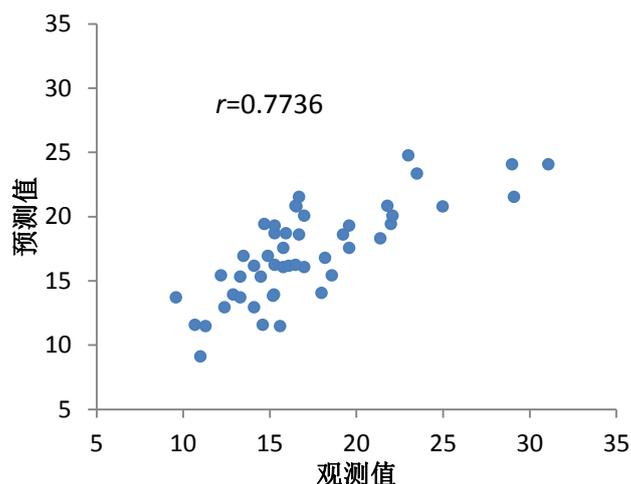
**解答：**题目中的行列平均与总平均数的离差，即为亲本的一般配合力，结果列于下表最后一列。每个组合的重复平均数与总平均数的离差、减去相应的一般配合力之后，即为特殊配合力，结果列于下表。

亲本	SCA							GCA
	A	B	C	D	E	F	G	
A	0.16	-3.39	4.95	2.04	0.67	-4.32	0.28	3.09
B	-2.74	3.96	-4.70	3.19	1.32	-0.07	3.03	-1.56
C	7.05	2.60	-1.76	4.23	0.31	7.57	-3.43	3.80
D	-3.06	-3.21	-4.22	1.43	-2.00	-1.74	-0.54	-0.19
E	-1.88	-1.03	-3.99	-0.80	1.38	0.94	-0.16	-1.67
F	0.98	-2.07	-4.83	2.06	-0.26	3.10	-4.10	0.57
G	-0.92	-0.87	-2.03	1.16	4.14	-0.40	1.90	-4.03

(2) 利用一般配合力，预测这49个组合的表现，绘制预测值与观察值重复平均数的散点图，并计算二者的相关系数。

**解答：**利用(1)计算出的一般配合力和特殊配合力，得到这49个组合的预测值、以及预测值与观测值的散点图如下，预测值与观察值之间的相关系数标于图中。

父本	母本						
	A	B	C	D	E	F	G
A	23.34	18.69	24.05	20.06	18.58	20.82	16.22
B	18.69	14.04	19.40	15.41	13.93	16.17	11.57
C	24.05	19.40	24.76	20.77	19.29	21.53	16.93
D	20.06	15.41	20.77	16.77	15.30	17.54	12.94
E	18.58	13.93	19.29	15.30	13.82	16.06	11.46
F	20.82	16.17	21.53	17.54	16.06	18.30	13.70
G	16.22	11.57	16.93	12.94	11.46	13.70	9.10



**12.9** 下表是练习12.7的重复平均数，但不包含自交，最后一列为每个亲本所在行和列的平均，最后一行为总平均。不区分父母本的一般配合力。

重复平均	A	B	C	D	E	F	G	行列平均
A		15.30	29.00	22.10	19.25	16.50	16.50	19.71
B	15.95		14.70	18.60	15.25	16.10	14.60	15.20
C	31.10	22.00		25.00	19.60	29.10	13.50	20.62
D	17.00	12.20	16.55		13.30	15.80	12.40	16.76
E	16.70	12.90	15.30	14.50		17.00	11.30	15.54
F	21.80	14.10	16.70	19.60	15.80		9.60	17.12
G	15.30	10.70	14.90	14.10	15.60	13.30		13.48
总平均	16.92							16.92

(1) 给出7个亲本一般配合力及42个组合特殊配合力的估计值。

**解答：**题目中的行列平均与总平均数的离差，即为亲本的一般配合力，结果列于下表最后一列。每个组合的重复平均数与总平均数的离差、减去相应的一般配合力之后，即为特殊配合力，结果列于下表。

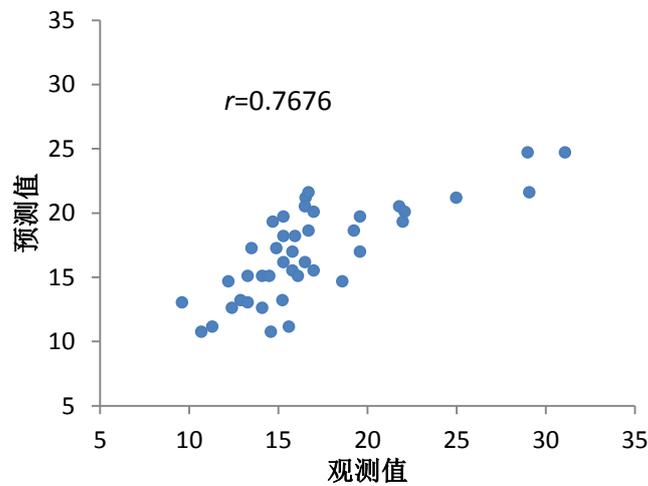
亲本	SCA							GCA
	A	B	C	D	E	F	G	
A		-2.90	4.29	2.02	0.64	-4.00	0.36	3.35
B	-2.25		-4.60	3.93	2.05	1.01	3.87	-2.06
C	6.39	2.70		3.83	-0.11	7.50	-3.74	4.44
D	-3.08	-2.47	-4.62		-1.78	-1.17	-0.21	-0.19
E	-1.91	-0.30	-4.41	-0.58		1.50	0.16	-1.65
F	1.30	-0.99	-4.90	2.63	0.30		-3.43	0.24
G	-0.84	-0.03	-2.34	1.49	4.46	0.27		-4.12

(2) 利用一般配合力，预测这42个组合的表现，绘制预测值与观察值重复平均数的散点图，并计算二者的相关系数。

**解答：**利用(1)计算出的一般配合力和特殊配合力，得到这42个组合的预测值、以及

预测值与观测值的散点图如下，预测值与观察值之间的相关系数标于图中。

父本	母本						
	A	B	C	D	E	F	G
A		18.20	24.71	20.08	18.61	20.50	16.14
B	18.20		19.30	14.67	13.20	15.09	10.73
C	24.71	19.30		21.17	19.71	21.60	17.24
D	20.08	14.67	21.17		15.08	16.97	12.61
E	18.61	13.20	19.71	15.08		15.50	11.14
F	20.50	15.09	21.60	16.97	15.50		13.03
G	16.14	10.73	17.24	12.61	11.14	13.03	



**12.10** 下表是练习12.7的重复平均数，但不包含反交，最后一列为每个亲本所在行和列的平均，最后一行为总平均。不区分父母本的一般配合力。

重复平均	A	B	C	D	E	F	F	行列平均
A	23.50	15.30	29.00	22.10	19.25	16.50	16.50	20.71
B		18.00	14.70	18.60	15.25	16.10	14.60	16.32
C			23.00	25.00	19.60	29.10	13.50	22.11
D				18.20	13.30	15.80	12.40	17.95
E					15.20	17.00	11.30	15.76
F						21.40	9.60	18.36
F							11.00	12.49
总平均	17.67							17.67

(1) 给出7个亲本一般配合力及28个组合特殊配合力的估计值。

**解答：**题目中的行列平均与总平均数的离差，即为亲本的一般配合力，结果列于下表最后一列。每个组合的重复平均数与总平均数的离差、减去相应的一般配合力之后，即为特殊配合力，结果列于下表。

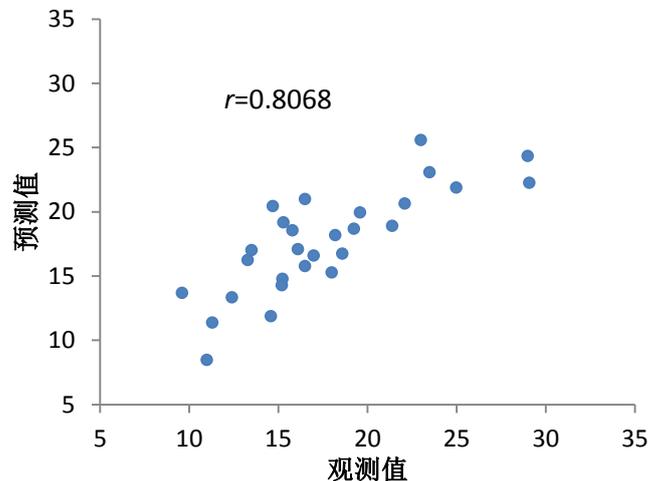
亲本	SCA	GCA
----	-----	-----

	A	B	C	D	E	F	G	
A	0.43	-3.87	4.68	1.48	0.58	-4.48	0.74	2.70
B		2.73	-5.72	1.88	0.48	-0.98	2.74	-1.20
C			-2.57	3.13	-0.32	6.87	-3.51	3.95
D				0.03	-2.92	-2.73	-0.91	0.25
E					0.92	0.41	-0.07	-1.70
F						2.50	-4.08	0.61
G							2.54	-4.61

(2) 利用一般配合力，预测这28个组合的表现，绘制预测值与观察值重复平均数的散点图，并计算二者的相关系数。

**解答：**利用(1)计算出的一般配合力和特殊配合力，得到这28个组合的预测值、以及预测值与观测值的散点图如下，预测值与观察值之间的相关系数标于图中。

父本	母本						
	A	B	C	D	E	F	G
A	23.07	19.17	24.32	20.62	18.67	20.98	15.76
B		15.27	20.42	16.72	14.77	17.08	11.86
C			25.57	21.87	19.92	22.23	17.01
D				18.17	16.22	18.53	13.31
E					14.28	16.59	11.37
F						18.90	13.68
G							8.46



**12.11** 下表是练习12.7的重复平均数，但不包含反交和自交，最后一列为每个亲本所在行和列的平均，最后一行为总平均。不区分父母本的一般配合力。

重复平均	A	B	C	D	E	F	G	行列平均
A		15.30	29.00	22.10	19.25	16.50	16.50	19.78
B			14.70	18.60	15.25	16.10	14.60	15.76
C				25.00	19.60	29.10	13.50	21.82

D	13.30	15.80	12.40	17.87
E		17.00	11.30	15.95
F			9.60	17.35
G				12.98
总平均	17.36			17.36

(1) 给出7个亲本一般配合力及21个组合特殊配合力的估计值。

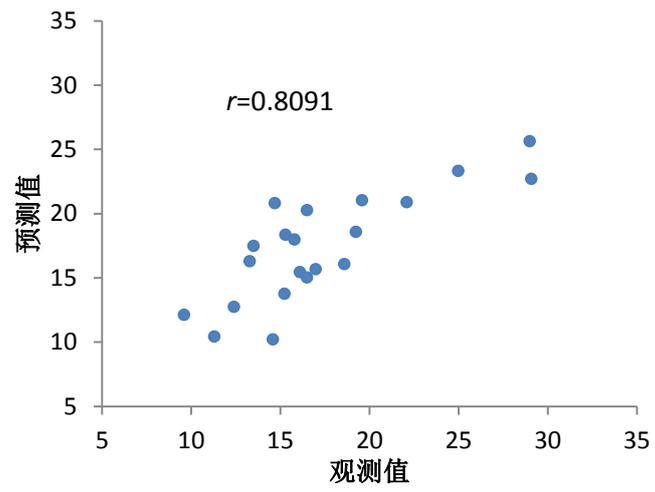
**解答：**题目中的行列平均与总平均数的离差，即为亲本的一般配合力，结果列于下表最后一列。每个组合的重复平均数与总平均数的离差、减去相应的一般配合力之后，即为特殊配合力，结果列于下表。

亲本	SCA							GCA
	A	B	C	D	E	F	G	
A	-3.04	3.39	1.23	0.68	-3.75	1.49	2.90	
B		-6.09	2.55	1.50	0.67	4.41	-1.92	
C			1.68	-1.42	6.40	-3.96	5.35	
D				-2.98	-2.16	-0.32	0.61	
E					1.34	0.88	-1.69	
F						-2.50	-0.01	
G							-5.25	

(2) 利用一般配合力，预测这21个组合的表现，绘制预测值与观察值重复平均数的散点图，并计算二者的相关系数。

**解答：**利用(1)计算出的一般配合力和特殊配合力，得到这21个组合的预测值、以及预测值与观测值的散点图如下，预测值与观察值之间的相关系数标于图中。

父本	母本						
	A	B	C	D	E	F	G
A		18.34	25.61	20.87	18.57	20.25	15.01
B			20.79	16.05	13.75	15.43	10.19
C				23.32	21.02	22.70	17.46
D					16.28	17.96	12.72
E						15.66	10.42
F							12.10
G							



## 第 13 章练习题和参考答案

**13.1** 三个标记座位 A、B 和 C 在染色体上的排列顺序为 A-B-C，它们之间的重组率用  $r_{AB}$ 、 $r_{BC}$  和  $r_{AC}$  表示。假定染色体区间 A-B-C 上不存在交换干涉，证明三个成对重组率满足关系式  $r_{AC} = r_{AB} + r_{BC} - 2r_{AB}r_{BC}$ 。

**解答：** 区间 A-C 没有观察到交换发生的概率等于  $1-r_{AC}$ 。区间 A-C 没有观察到交换可以分解成相互独立的两个事件，一是区间 A-B 和区间 B-C 上都没有观察到交换，发生概率等于  $(1-r_{AB})(1-r_{BC})$ ；二是区间 A-B 和区间 B-C 上都观察到交换，发生概率等于  $r_{AB}r_{BC}$ 。因此有，

$$1 - r_{AC} = (1 - r_{AB})(1 - r_{BC}) + r_{AB}r_{BC}$$

也就是

$$r_{AC} = r_{AB} + r_{BC} - 2r_{AB}r_{BC}$$

**13.2** 在本章作为实例的大麦 DH 群体中，前 45 个家系在标记 Act8A 上的标记型和粒重数据见下表。利用单标记分析检验标记 Act8A 与粒重 QTL 之间是否存在显著的连锁关系。

1~15	Act8A	0	2	-1	2	0	2	2	0	0	2	2	2	0	2	2
	粒重	41	40	40	40	41	46	40	44	42	46	44	41	43	42	47
16~30	Act8A	0	0	0	2	2	0	0	2	0	0	0	2	0	2	2
	粒重	44	41	41	44	42	44	45	40	41	41	48	47	42	44	43
31~45	Act8A	0	0	2	0	2	0	2	0	0	0	0	2	0	0	2
	粒重	40	39	41	46	42	41	40	47	42	41	43	46	43	44	41

**解答：** 对这 45 个家系根据基因型进行分组，标记型为 0 的家系有 24 个、标记型为 2 的家系有 20 个、一个家系的标记型缺失。在 EXCEL 中计算两组样本的均值、方差和标准差。在两种样本具有相同方差的假定下，计算合并方差。利用合并方差计算两组样本均值之差的标准差，两组样本均值的差异绝对值除以该标准差，就得到  $t$  值，其自由度为  $24+20-2=42$ 。因此得到显著概率 0.8537。

统计量	标记型 0	标记型 2
样本量	24	20
均值	42.67	42.80
方差	5.01	6.38
标准差	2.24	2.53
合并方差	5.63	
样本均值之差的标准差	0.7185	
t 值	0.1856	

P 值	0.8537
-----	--------

读者也可以利用 EXCEL 的  $t$ -检验函数 TTEST, 直接得到显著概率 0.8537。因此, 座位 Act8A 上两种标记型的平均粒重之间不存在显著差异。说明该标记与粒重 QTL 之间不存在连锁关系。

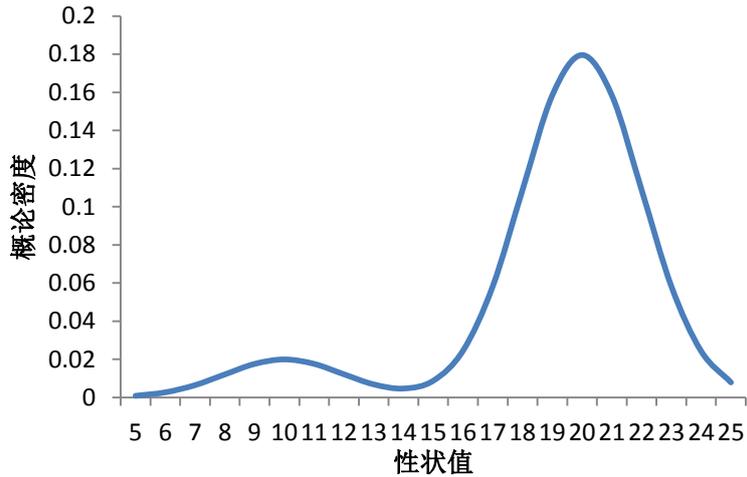
**13.3** 亲本  $P_1$  和  $P_2$  在一个标记座位和一个 QTL 上的基因型分别为  $MMQQ$  和  $mmqq$ , 标记与 QTL 间的重组率  $r=0.1$ 。基因型  $QQ$  和  $qq$  服从正态分布, 均值分别为 20 和 10, 方差均为 4。计算 DH 群体中, 两种标记型下  $QQ$  和  $qq$  的频率; 根据两种 QTL 的基因型频率, 计算每种标记型的群体平均数, 并将结果整理在下面的表格中; 绘制标记型  $MM$  的表型混合分布曲线图。

标记型	$QQ$	$qq$	标记型的群体均值
$MM$			
$mm$			

**解答:** 同时考虑标记和 QTL 两个座位, 四种基因型  $MMQQ$ 、 $MMqq$ 、 $mmQQ$ 、 $mmqq$  的频率等于 F1 产生 4 中配子  $MQ$ 、 $Mq$ 、 $mQ$ 、 $mq$  的频率, 即分别为 0.45、0.05、0.05、0.45。标记型  $MM$  中, 两种 QTL 基因型  $QQ$  和  $qq$  的频率分别为 0.9 和 0.1。标记型  $mm$  中, 两种 QTL 基因型  $QQ$  和  $qq$  的频率分别为 0.1 和 0.9。由此得到两种标记型的均值分别为 19 和 11。

标记型	$QQ$	$qq$	标记型的群体均值
$MM$	0.9	0.1	19
$mm$	0.1	0.9	11

标记型  $MM$  的表型是两种 QTL 基因型  $QQ$  和  $qq$  按照比例 0.9 和 0.1 构成的混合分布。QTL 基因型  $QQ$  和  $qq$  分别服从正态分布  $N(\mu = 20, \sigma^2 = 4)$  和  $N(\mu = 10, \sigma^2 = 4)$ 。在 EXCEL 中, 利用正态分布函数 NORM.DIST 计算表型取值在 5~25 之间的概率密度。混合分布曲线图如下。



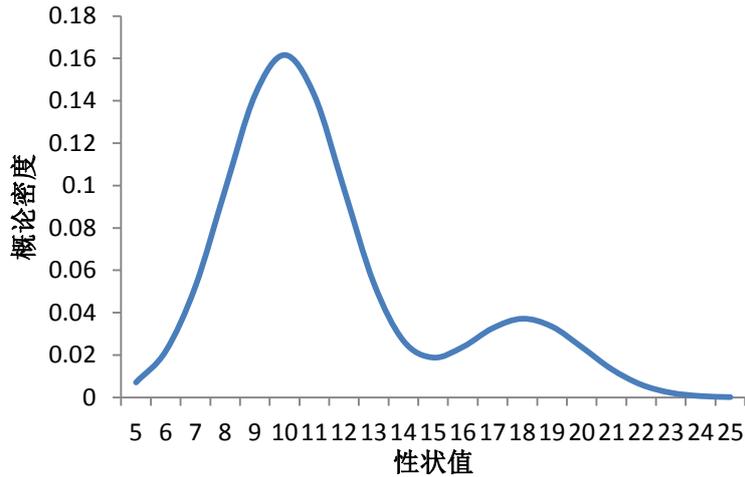
**13.4** 亲本  $P_1$  和  $P_2$  在一个标记座位和一个 QTL 上的基因型分别为  $MMQQ$  和  $mmqq$ ，标记与 QTL 间的重组率  $r=0.1$ 。QTL 基因型  $QQ$ 、 $Qq$  和  $qq$  的服从正态分布均值分别为 20、18 和 10、方差均为 4。计算  $F_2$  群体中，三种标记型下  $QQ$ 、 $Qq$  和  $qq$  的频率；根据三种 QTL 的基因型频率，计算每种标记型的群体平均数，并将结果整理在下面的表格中；绘制标记型  $mm$  的表型混合分布曲线图。

标记型	$QQ$	$Qq$	$qq$	标记型的群体均值
$MM$				
$Mm$				
$mm$				

**解答：**首先从  $F_1$  的 4 种配子  $MQ$ 、 $Mq$ 、 $mQ$ 、 $mq$  的频率，计算 9 种基因型的频率。然后计算每种标记型中，三种 QTL 基因型的频率，并计算标记型的群体均值，结果如下。

标记型	$QQ$	$Qq$	$qq$	标记型的群体均值
$MM$	0.81	0.18	0.01	19.54
$Mm$	0.09	0.82	0.09	17.46
$mm$	0.01	0.18	0.81	11.54

标记型  $mm$  的表型是三种 QTL 基因型  $QQ$ 、 $Qq$  和  $qq$  按照比例 0.01、0.18 和 0.81 构成的混合分布。三种 QTL 基因型分别服从正态分布  $N(\mu = 20, \sigma^2 = 4)$ 、 $N(\mu = 18, \sigma^2 = 4)$  和  $N(\mu = 10, \sigma^2 = 4)$ 。在 EXCEL 中，利用正态分布函数 NORM.DIST 计算表型取值在 5~25 之间的概率密度。混合分布曲线图如下。



**13.5** 假定标记  $A/a$  和  $B/b$  决定的区间上存在一个影响数量性状的 QTL, 两个等位基因用  $Q$  和  $q$  表示, 基因型  $QQ$  和  $qq$  的平均表现分别为 20 和 15, QTL 与标记 A 的重组率为 0.05、与标记 B 的重组率为 0.03。计算亲本  $AAQQBB$  与  $aaqqbb$  衍生的 DH 群体中, 各种标记型下  $QQ$  和  $qq$  的频率; 根据两种 QTL 的基因型频率, 计算每种标记型的群体平均数。并将结果整理在下面的表格中。

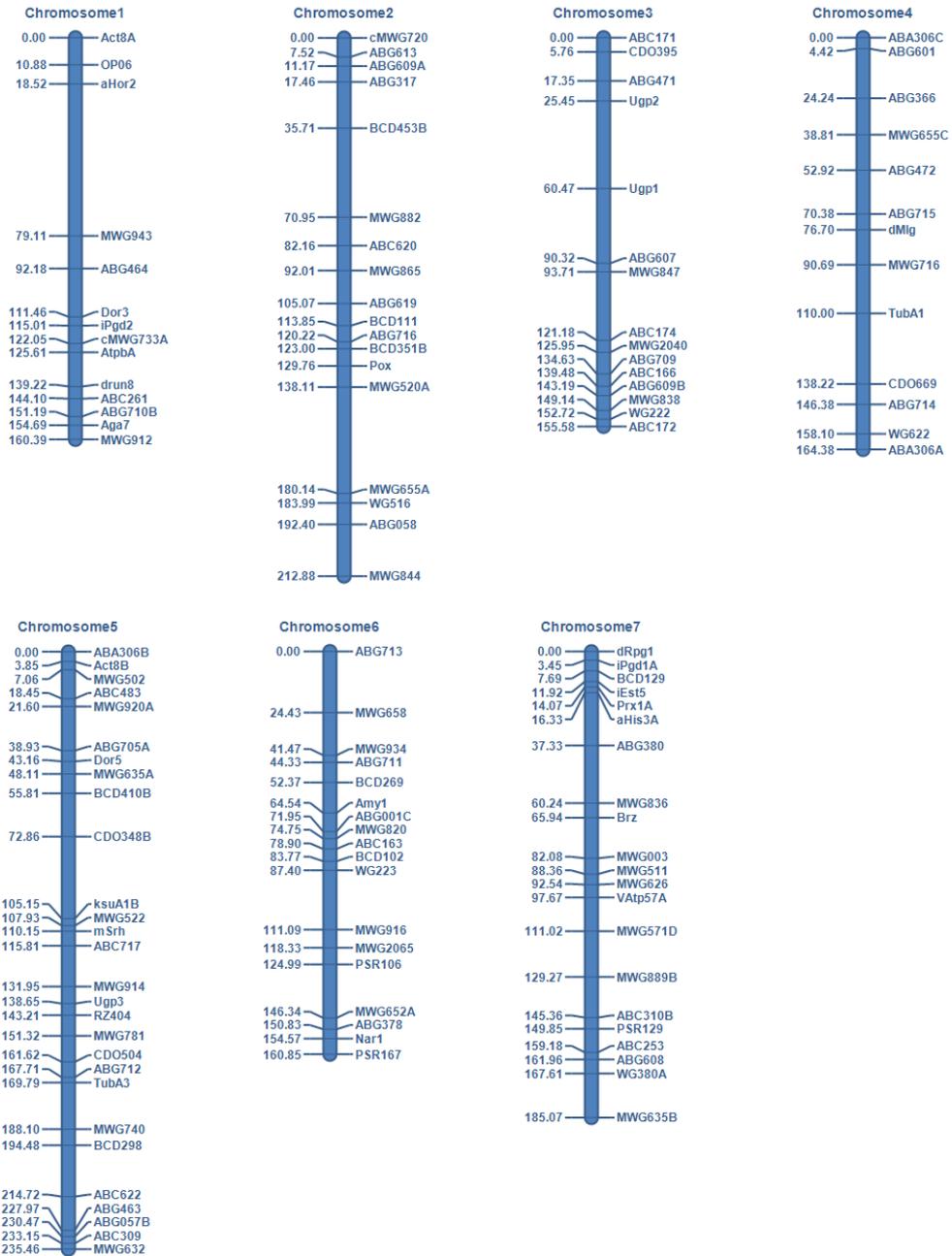
标记型	$QQ$	$qq$	标记型的群体均值
$AABB$			

**解答:** 利用表 13.3 中的公式, 在 EXCEL 中计算各种区间标记型下两种 QTL 基因型的频率, 进而计算标记型的均值, 结果如下。

标记型	$QQ$	$qq$	标记型的群体均值
$AABB$	0.9984	0.0016	19.9919
$AAbb$	0.3701	0.6299	16.8506
$aaBB$	0.6299	0.3701	18.1494
$Aabb$	0.0016	0.9984	15.0081

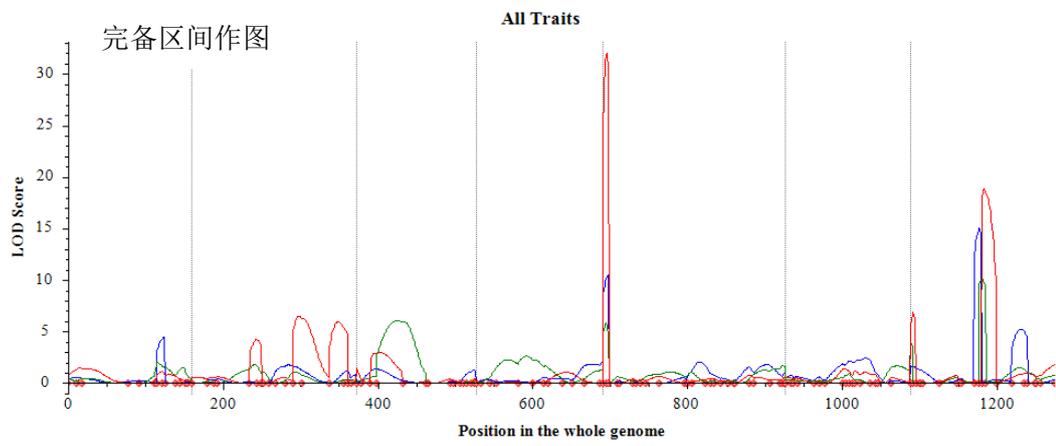
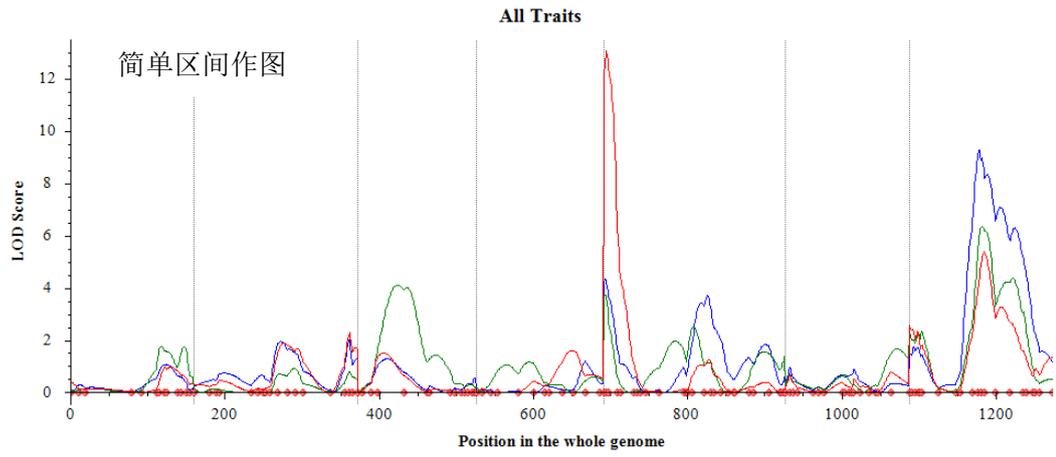
**13.6** 利用 QTL IciMapping 软件附带的大麦 DH 群体, 练习使用连锁图谱构建的 MAP 功能, 并绘制大麦的 7 条连锁图谱。

**解答:** 利用该群体得到的大麦 7 条连锁图如下。



**13.7** 利用 QTL IciMapping 软件附带的大麦 DH 群体，练习使用数量性状基因定位的 BIP 功能，并绘制简单区间和完备区间作图结果的 LOD 曲线图。

**解答：**利用该群体，简单区间和完备区间作图结果的 LOD 曲线如下。



## 附录 A1: Excel 中公式的编辑、复制和粘贴

**问题:** 假定变量  $X$  的取值是  $1, 2, \dots, 10$ , 现要计算  $\ln(1+X)$ 、 $X^2$  和  $\sqrt{X+0.5}$ 。

**解答:** 打开 EXCEL, 把一个空白工作表的第一行作为标题行, 在单元格 A2:A11 中, 写入  $1, 2, \dots, 10$  这 10 个数字。根据下图所示, 依次在三个单元格 B2、C2、D2 中编辑所要计算的三个公式。

**步骤 1:** 在单元格 B2 中写入公式‘=LN(1+A2)’, 其中, 等号‘=’表示该单元格将是通过等号后公式计算出来的数值, 符号‘LN’在 EXCEL 中表示自然对数函数, 公式中的 A2 代表单元格 A2 中的数值, 这里为 1。

**步骤 2:** 在单元格 C2 写入公式‘=A2^2’, 符号‘^’在 EXCEL 表示幂函数。

**步骤 3:** 在单元格 D2 写入公式‘=SQRT(A2+0.5)’, 符号‘SQRT’在 EXCEL 表示平方根函数。

**注意:** EXCEL 公式中的大小写是等价的, 也就是不区分大小写。如平方根函数也可写为‘Sqrt’或‘sqrt’等。公式最好在英文输入的方式下进行编辑, 要使用半角来编辑字母和括号。

步骤1: 在单元格B2写入公式 ‘=LN(1+A2)’

MDETERM				
	A	B	C	D
1	X	$\ln(1+X)$	$X^2$	
2	1	=LN(1+A2)		
3	2			
4	3			
5	4			
6	5			
7	6			
8	7			
9	8			
10	9			
11	10			
12				

步骤2: 在单元格C2写入公式 ‘=A2^2’

MDETERM				
	A	B	C	D
1	X	$\ln(1+X)$	$X^2$	
2	1	0.693147	=A2^2	
3	2			
4	3			
5	4			
6	5			
7	6			
8	7			
9	8			
10	9			
11	10			
12				

步骤3: 在单元格D2写入公式 ‘=SQRT(A2+0.5)’

MDETERM					
	A	B	C	D	E
1	X	$\ln(1+X)$	$X^2$	$\sqrt{X+0.5}$	
2	1	0.693147	1	=SQRT(A2)	
3	2				
4	3				
5	4				
6	5				
7	6				
8	7				
9	8				
10	9				
11	10				
12					

然后对前面编辑出的公式进行复制和粘贴。如下图左所示, 选择单元格 B2 和 C2, 按下‘Ctrl+C’组合键, 选择单元格 B3:D11, 按下‘Ctrl+V’组合键即可。也可如下图右所示, 选择单元格 B2 和 D2, 将鼠标移动到选择区域的右下角, 当看到‘+’符号时, 拖动鼠标从第 2 行一直到第 11 行即可。

可以看到，在粘贴公式时，公式中的单元格会随粘贴位置而变化。当粘贴到当前位置的第  $i$  行、第  $j$  列的单元格时，公式中的单元格也会随之加上  $i$  行、加上  $j$  列。因此，本例中的公式只需要对单元格 A2 编辑一次，利用拷贝和粘贴，就能得到单元格 A3:A11 对应的三个函数值。

步骤4：如下图左所示，选择单元格B2和D2，按下Ctrl+C组合键，选择单元格B3:D11，按下Ctrl+V组合键即可；也可如下图右所示，选择单元格B2和D2，将鼠标移动到选择区域的右下角，当看到‘+’符号时，拖动鼠标从第2行一直到第11行即可

B3		=LN(1+A3)			
A	B	C	D	E	
1	X	$\ln(1+X)$	$X^2$	$\sqrt{X+0.5}$	
2	1	0.693147	1	1.224745	
3	2	1.098612	4	1.581139	
4	3	1.386294	9	1.870829	
5	4	1.609438	16	2.12132	
6	5	1.791759	25	2.345208	
7	6	1.94591	36	2.54951	
8	7	2.079442	49	2.738613	
9	8	2.197225	64	2.915476	
10	9	2.302585	81	3.082207	
11	10	2.397895	100	3.24037	

B2		=LN(1+A2)			
A	B	C	D	E	
1	X	$\ln(1+X)$	$X^2$	$\sqrt{X+0.5}$	
2	1	0.693147	1	1.224745	
3	2	1.098612	4	1.581139	+
4	3	1.386294	9	1.870829	
5	4	1.609438	16	2.12132	
6	5	1.791759	25	2.345208	
7	6	1.94591	36	2.54951	
8	7	2.079442	49	2.738613	
9	8	2.197225	64	2.915476	
10	9	2.302585	81	3.082207	
11	10	2.397895	100	3.24037	

说明：上图中，如选中区域 B2:D11 中的任意一个单元格，看到的是一个公式。有时可能只想要公式计算出的结果、而不再需要公式。这时可以选中区域 B2:D11，按下‘Ctrl+C’组合键。选中存放数值的一个起始位置，在 EXCEL 的菜单中激活下图的选择性粘贴对话框中，选择‘数值(V)’这一选项，点击‘确定’即可。



## 附录 A2: Excel 公式中控制符\$的使用

**问题:** 假定变量  $X$  的取值是  $1, 2, \dots, 10$ , 根据公式  $\ln(1+X)$ 、 $X^2$  和  $\sqrt{X+0.5}$  共得到 30 个数值, 排列成 10 行 3 列。现要计算行平均、列平均、总平均、行平均与总平均的离差、列平均与总平均的离差、以及每个数值与行平均、列平均、总平均的离差。

**解答:** 首先把这 30 个数值存放在单元格 B2:D11 中。

在单元格 E2 输入公式‘=AVERAGE(B2:D2)’, 符号 AVERAGE 在 EXCEL 表示计算多个数值平均数的函数, 括号内的 B2:D2 指明对哪些数值计算平均数; 将单元格 E2 的公式拷贝粘贴到单元格 E3:E11, 就得到其它 9 个行平均数。

在单元格 B12 输入公式‘=AVERAGE(B2:B11)’, 并将单元格 B12 的公式拷贝粘贴到单元格 C12:D12, 就得到其它 2 个行平均数。

在单元格 E12 输入公式‘=AVERAGE(B2:D11)’, 就得到这 30 个数值的总平均数。实现过程, 参看下图。

	A	B	C	D	E	F
1	X	Ln(1+X)	X <sup>2</sup>	√X+0.5	行平均	
2	1	0.693147	1	1.224745	=AVERAGE	
3	2	1.098612	4	1.581139	2.226584	
4	3	1.386294	9	1.870829	4.085708	
5	4	1.609438	16	2.12132	6.576919	
6	5	1.791759	25	2.345208	9.712322	
7	6	1.94591	36	2.54951	13.49847	
8	7	2.079442	49	2.738613	17.93935	
9	8	2.197225	64	2.915476	23.03757	
10	9	2.302585	81	3.082207	28.79493	
11	10	2.397895	100	3.24037	35.21276	
12	列平均	1.750231	38.5	2.366942	14.20572	
13						
14						

参看下图左, 在单元格 F2 输入公式‘=E2-\$E\$12’, 单元格 E12 中的两个\$符号, 可以保证公式在拷贝粘贴过程中, 单元格 E12 不发生改变, 也就是起到冻结这个单元格的作用。将单元格 F2 的公式拷贝粘贴到单元格 F3:F11, 就得到其它 9 个行平均数与总平均数的离差。单元格 F2:F11 给出的离差就是单元格 E2:E11 这 10 个行平均减去总平均 (单元格 E12)。

类似地参看下图右，计算 3 个列平均数与总平均数的离差。

MDETERM						
	A	B	C	D	E	F
1	X	Ln(1+X)	X <sup>2</sup>	√X+0.5	行平均	行平均-总平均
2		1 0.693147	1	1.224745	0.972631	=E2-\$E\$12
3		2 1.098612	4	1.581139	2.226584	-11.97914044
4		3 1.386294	9	1.870829	4.085708	-10.12001646
5		4 1.609438	16	2.12132	6.576919	-7.628804725
6		5 1.791759	25	2.345208	9.712322	-4.493401694
7		6 1.94591	36	2.54951	13.49847	-0.707250842
8		7 2.079442	49	2.738613	17.93935	3.7336273
9		8 2.197225	64	2.915476	23.03757	8.831842698
10		9 2.302585	81	3.082207	28.79493	14.58920655
11		10 2.397895	100	3.24037	35.21276	21.00703106
12	列平均	1.750231	38.5	2.366942	14.20572	
13	列平均-总平均	-12.4555	24.29428	-11.8388		

每个数值与行平均、列平均、总平均的离差，其实是要计算每个数值减去它所在的行平均数、减去它所在的列平均数、最后加上总平均数。假定把这 30 个离差保存在单元格 B15:D24 中。如下图左，首先在单元格 B15 输入公式‘=B2-\$E2-B\$12+\$E\$12’。这个公式中，\$E2 将保证公式在同一行拷贝粘贴时，单元格 E2 不发生变化；B\$12 将保证公式在同一列拷贝粘贴时，B12 不发生变化；\$E\$12 将保证公式在任何地方拷贝粘贴时，E12 都不发生变化。将单元格 B15 拷贝粘贴到所有的单元格 B15:D24 中即得到想要的离差。

如检查单元格 D24（下图右），它的计算公式为‘=D11-\$E11-D\$12+\$E\$12’，正好是最后一个数值减去它所在的行平均数、减去它所在的列平均数、最后加上总平均数。

MDETERM						
	A	B	C	D	E	F
1	X	Ln(1+X)	X <sup>2</sup>	√X+0.5	行平均	行平均-总平均
2		1 0.693147	1	1.224745	0.972631	-13.23309346
3		2 1.098612	4	1.581139	2.226584	-11.97914044
4		3 1.386294	9	1.870829	4.085708	-10.12001646
5		4 1.609438	16	2.12132	6.576919	-7.628804725
6		5 1.791759	25	2.345208	9.712322	-4.493401694
7		6 1.94591	36	2.54951	13.49847	-0.707250842
8		7 2.079442	49	2.738613	17.93935	3.7336273
9		8 2.197225	64	2.915476	23.03757	8.831842698
10		9 2.302585	81	3.082207	28.79493	14.58920655
11		10 2.397895	100	3.24037	35.21276	21.00703106
12	列平均	1.750231	38.5	2.366942	14.20572	
13	列平均-总平均	-12.4555	24.29428	-11.8388		
14						
15		=B2-\$E2	-52.6784	-16.3206		
16		-17.0839	-50.9323	-17.2181		
17		-18.6554	-47.7914	-18.7875		
18		-20.9234	-43.2826	-21.0283		
19		-23.8765	-37.418	-23.9398		
20		-27.5085	-30.2042	-27.5216		
21		-31.8159	-21.6451	-31.7734		
22		-36.7963	-11.7433	-36.6948		
23		-42.4483	-0.50065	-42.2854		
24		-48.7708	12.08152	-48.5451		

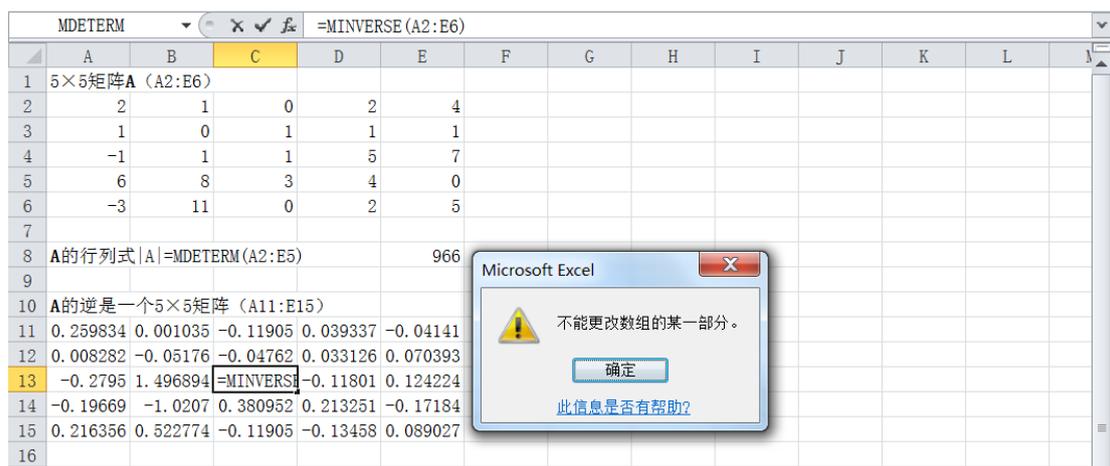
说明：灵活使用控制符\$，可以有效利用 EXCEL 开展的数据分析。

## 附录 A3: Excel 中的矩阵运算

### A3.1 三键组合‘Ctrl+Shift+Enter’的使用

当一个函数返回多于一个的结果数值时，就需要使用三键组合‘Ctrl+Shift+Enter’。使用这个组合时，先选中返回数值所要存放的整个区域，然后按下‘Ctrl’键不动、再按下‘Shift’键不动、最后按下‘Enter’键，即可获得计算结果。

**注意：**当一个函数的返回值保存在多个单元格的时候，试图修改、删去部分单元格都是无效的，这时会弹出下面的信息框。出现这种情况时，可以使用三键组合‘Ctrl+Shift+Enter’，也可以使用‘Esc’键来消除。如要修改或删除结果，需要选中结果所在的所有单元格，才能进行修改或删除。



### A3.2 矩阵加减法

维数（即行数和列数）相同的两个矩阵可以进行加减，定义为相应元素的加减。如下图所示，**A** 是一个 4×3 矩阵，存放在单元格 A2:C5 中；**B** 也是一个 4×3 矩阵，存放在单元格 A8:C11 中。现要计算 **A+B** 和 **A-B**。

假定把 **A+B** 存放在单元格 E2:G5 中，选中存放 **A+B** 的所有单元格 E2:G5，输入公式‘=A2:C5+A8:C11’，当然 A2:C5 就是矩阵 **A**，A8:C11 就是矩阵 **B**。然后使用三键组合‘Ctrl+Shift+Enter’，即得到 **A+B** 的所有元素。

与 **A+B** 的计算类似，假定把 **A-B** 存放在单元格 E8:G11 中，选中存放 **A-B** 的所有单元

格 E8:G11，输入公式‘=A2:C5-A8:C11’，使用三键组合‘Ctrl+Shift+Enter’立即得到 **A-B** 的所有元素。

	A	B	C	D	E	F	G	H
1	4×3矩阵A (A2:C5)				A+B是4×3矩阵			
2	1	1	0		=A2:C5+A8:C11	3	3	
3	1	0	1		4	-2	2	
4	-1	1	1		-1	2	2	
5	1	-1	0		1	-2	0	
6								
7	4×3矩阵B (A8:C11)				A-B是4×3矩阵			
8	1	2	3		0	-1	-3	
9	3	-2	1		-2	2	0	
10	0	1	1		-1	0	0	
11	0	-1	0		1	0	0	
12								

### A3.3 矩阵乘法

矩阵 **A** 的列数等于矩阵 **B** 的行数，这时可以进行矩阵乘法，乘积矩阵 **AB** 的行数等于矩阵 **A** 的行数，列数等于矩阵 **B** 的列数，每个元素定义 **A** 的行向量与 **B** 的列向量的积和。如下图所示，**A** 是一个 4×3 矩阵，存放在单元格 A2:C5 中；**B** 也是一个 3×5 矩阵，存放在单元格 A8:E10 中。现要计算乘积矩阵 **AB**。

**AB** 是一个 4×5 矩阵，假定存放在单元格 A13:E16 中。选中存放 **AB** 的所有单元格 A13:E16，输入公式‘=MMULT(A2:C5, A8:E10)’，其中 MMULT 是 EXCEL 中计算矩阵乘积的函数，A2:C5 就是矩阵 **A**，A8:E10 就是矩阵 **B**，两个矩阵之间用逗号分隔。然后使用三键组合‘Ctrl+Shift+Enter’，即得到 **AB** 的所有元素。

MDETERM						
=MMULT(A2:C5,A8:E10)						
	A	B	C	D	E	F
1	4×3矩阵A (A2:C5)					
2	1	1	0			
3	1	0	1			
4	-1	1	1			
5	1	-1	0			
6						
7	3×5矩阵B (A8:C11)					
8	1	2	3	0	5	
9	3	-2	1	2	-4	
10	0	1	1	1	3	
11						
12	AB是一个4×5矩阵 (A13:E16)					
13	=MMULT(A2:C5,A8:E10)	0	4	2	1	
14	1	3	4	1	8	
15	2	-3	-1	3	-6	
16	-2	4	2	-2	9	
17						

### A3.4 矩阵的转置

矩阵  $\mathbf{A}$  行列元素对调得到的矩阵，称为  $\mathbf{A}$  的转置，用  $\mathbf{A}^T$  表示。如下图所示， $\mathbf{A}$  是一个  $3 \times 4$  矩阵，存放在单元格 A2:D4 中。现要计算转置矩阵  $\mathbf{A}^T$ 。

$\mathbf{A}^T$  是一个  $4 \times 3$  矩阵，假定存放在单元格 A7:C10 中。选中存放  $\mathbf{A}^T$  的所有单元格 A7:C10，输入公式‘=TRANSPOSE(A2:D4)’，其中 TRANSPOSE 是 EXCEL 中计算矩阵转置的函数，A2:D4 就是矩阵  $\mathbf{A}$ 。然后使用三键组合‘Ctrl+Shift+Enter’，即得到  $\mathbf{A}^T$  的所有元素。

MDETERM					
=TRANSPOSE(A2:D4)					
	A	B	C	D	E
1	3×4矩阵A (A2:D4)				
2	1	1	0	2	
3	1	0	1	1	
4	-1	1	1	5	
5					
6	A的转置是一个4×3矩阵 (A7:C10)				
7	=TRANSPOSE(A2:D4)	1	-1		
8	1	0	1		
9	0	1	1		
10	2	1	5		
11					

### A3.5 正方矩阵的行列式和逆矩阵

如下图所示， $\mathbf{A}$  是一个  $5 \times 5$  的正方矩阵，存放在单元格 A2:E6 中。现要计算  $\mathbf{A}$  的行列

式 $|A|$ 和逆矩阵 $A^{-1}$ 。

正方形矩阵  $A$  的行列式 $|A|$ 是一个数值，假定存放在单元格 E8 中。对单元格 E8，输入公式‘=MDETERM(A2:E6)’，其中 MDETERM 是 EXCEL 中计算行列式的函数，A2:E6 就是正方形矩阵  $A$ ，按下‘Enter’键立即得到  $A$  的行列式 $|A|$ 。

$A^{-1}$  是一个  $5 \times 5$  正方形矩阵，假定存放在单元格 A11:E15 中。选中存放  $A^{-1}$  的所有单元格 A11:E15，输入公式‘=MINVERSE(A2:E6)’，其中 MINVERSE 是 EXCEL 中计算逆矩阵的函数，A2:E6 就是矩阵  $A$ 。然后使用三键组合‘Ctrl+Shift+Enter’，即得到  $A^{-1}$  的所有元素。

MDETERM						=MINVERSE (A2:E6)
	A	B	C	D	E	F
1	5×5矩阵A (A2:E6)					
2	2	1	0	2	4	
3	1	0	1	1	1	
4	-1	1	1	5	7	
5	6	8	3	4	0	
6	-3	11	0	2	5	
7						
8	A的行列式  A =MDETERM(A2:E5)					966
9						
10	A的逆是一个5×5矩阵 (A11:E15)					
11	=MINVERSE	0.001035	-0.11905	0.039337	-0.04141	
12		0.008282	-0.05176	-0.04762	0.033126	0.070393
13		-0.2795	1.496894	-0.14286	-0.11801	0.124224
14		-0.19669	-1.0207	0.380952	0.213251	-0.17184
15		0.216356	0.522774	-0.11905	-0.13458	0.089027
16						

### A3.6 利用矩阵运算计算最小二乘估计

假定有下面的一个线性模型，现要利用 EXCEL 计算参数  $b_0$ 、 $b_1$ 、 $b_2$  的最小二乘估计值

$$\hat{\mathbf{b}} = (\mathbf{X}\mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}.$$

$$\mathbf{y} = \begin{bmatrix} 116.30 \\ 98.45 \\ 117.68 \\ 116.00 \\ 109.16 \\ 111.78 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & 1 \\ 1 & 0.5 & 0.5 \\ 1 & -0.5 & 0.5 \\ 1 & 0 & 0.5 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$$

把观测向量  $\mathbf{y}$  存放在单元格 A2:A7 中，把数据矩阵  $\mathbf{X}$  存放在单元格 B2:D7 中。利用下图所示的 5 个步骤，就可以通过矩阵运算得到参数向量  $\mathbf{b}$  的最小二乘估计。**步骤 1** 计算转置矩阵  $\mathbf{X}^T$ ，存放单元格 A10:F12 中；**步骤 2** 计算  $\mathbf{X}^T$  与观测向量  $\mathbf{y}$  的乘积，存放单元格 G10:G12 中；**步骤 3** 计算  $\mathbf{X}^T\mathbf{X}$ ，存放单元格 A15:C17 中；**步骤 4** 计算  $(\mathbf{X}^T\mathbf{X})^{-1}$ ，存放单元格 E15:G17 中；**步骤 5** 计算  $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ ，即参数向量  $\mathbf{b}$  的最小二乘估计，存放单元格 B20:B22 中。

MDETERM $=\text{MMULT}(\text{E15:G17}, \text{G10:G12})$										
	A	B	C	D	E	F	G	H	I	J
1	$\mathbf{y}$	$\mathbf{X}$ (B2:D7) 是一个 $6 \times 3$ 矩阵				$\mathbf{y}$ (A2:A7) 是一个 $6 \times 1$ 矩阵				
2	116.30	1	1	0	$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$					
3	98.45	1	-1	0	回归系数的最小二乘估计是 $\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$					
4	117.68	1	0	1						
5	116.00	1	0.5	0.5						
6	109.16	1	-0.5	0.5						
7	111.78	1	0	0.5						
8										
9	步骤 1: $\mathbf{X}^T$ (A10:F12) 是一个 $3 \times 6$ 矩阵				步骤 2: $\mathbf{X}^T\mathbf{y}$ (G10:G12) 是一个 $3 \times 1$ 矩阵					
10	1	1	1	1	1	1	669.37			
11	1	-1	0	0.5	-0.5	0	21.27			
12	0	0	1	0.5	0.5	0.5	286.15			
13										
14	步骤 3: $\mathbf{X}^T\mathbf{X}$ (A15:C17) 是一个 $3 \times 3$ 矩阵			步骤 4: $(\mathbf{X}^T\mathbf{X})^{-1}$ (E15:G17) 是一个 $3 \times 3$ 矩阵						
15	6	0	2.5	0.411765			0	-0.58824		
16	0	2.5	0	0			0.4	0		
17	2.5	0	1.75	-0.58824			0	1.411765		
18										
19	步骤 5: $\mathbf{b}$ (B20:B22) 是一个 $3 \times 1$ 矩阵，表示回归系数的最小二乘估计									
20	b0	$=\text{MMULT}(\text{E15:G17}, \text{G10:G12})$								
21	b1	8.508								
22	b2	10.22941								
23										

### A3.7 利用函数 LINEST 计算最小二乘估计

用  $\mathbf{y} = \mathbf{X}\mathbf{b}$  表示一个线性回归模型，EXCEL 中的 LINEST 函数能够同时给出参数向量  $\mathbf{b}$  的最小二乘估计以及与回归显著性检验有关的统计量。这个函数中包含四组变量假定用 ‘LINEST(变量 1, 变量 2 变量 3, 变量 4)’ 表示。变量 1 指定观测数据向量  $\mathbf{y}$  所在的区域。变量 2 指定设计矩阵  $\mathbf{X}$  所在的区域。变量 3 是一个逻辑变量，等于 ‘True’、或 1、或省略时，除参数向量  $\mathbf{b}$  外还将估计截距项；否则，将截距项强制设为 0。变量 4 也是一个逻辑变量，等于 ‘True’、或 1 时，除输出参数向量  $\mathbf{b}$  外的估计值外，还将输出与回归显著性检验有关的统计量；等于 ‘False’、或 0、或省略时，仅输出参数向量  $\mathbf{b}$  外的估计值。

下图上部显示观测向量  $\mathbf{y}$  和设计矩阵  $\mathbf{X}$  在工作表中的位置。中部给出输出结果的解释，

LINEST 返回结果的第一行是回归系数，按照  $b_n$ 、 $b_{n-1}$ 、...、 $b_2$ 、 $b_1$ 、 $b_0$  的顺序排列 ( $n$  表示回归参数的个数，等于设计矩阵  $\mathbf{X}$  的列数， $b_0$  是常数项或截距项)。第二行是回归系数估计值的标准差，顺序与回归系数估计值的顺序相同。第三行包含两个数值，一个是回归方程的判定系数  $R^2$ ，一个是  $\mathbf{Y}$  估计值的标准差。第 4 行包含两个数值，一个是回归方程显著性检验的  $F$  统计量，一个是误差效应的自由度。第 5 行包含两个数值，一个是回归平方和，一个是剩余平方和。

下图的下部给出计算结果。观测向量是区域 B2:B8 的 7×1 矩阵。设计矩阵  $\mathbf{X}$  的第一列元素全部为 1，因此对应的回归参数其实是常数项  $b_0$ 。包含常数项在内，这里的回归系数有  $b_2$ 、 $b_1$ 、 $b_0$  共三个。设计矩阵是区域 D2:E8 的 7×2 矩阵。返回的结果是一个 5×3 矩阵，这里的 5 表示 LINEST 返回结果的列数，这里的 3 表示包含常数项的回归系数的个数。

MDETERM		=LINEST(B2:B8, D2:E8, TRUE, TRUE)							
	A	B	C	D	E	F	G	H	I
1	群体	均值	总平均	加性效应	显性效应				
2	P1	116.30	1	1	0				
3	P2	98.45	1	-1	0				
4	F1	117.68	1	0	1				
5	B1	116.00	1	0.5	0.5				
6	B2	109.16	1	-0.5	0.5				
7	F2	111.78	1	0	0.5				
8	F2	101.52	1	0	0.25				
9									
10	最小二乘函数LINEST(B2:B8, C2:E8, FALSE, TRUE)输出结果的解释								
11	回归系数			bn	bn-1	...	b2	b1	b0
12	回归系数的标准差			SEn	SEn-1		SE2	SE1	SE0
13	判断系数和Y估计值的标准差			R2	SEy				
14	F统计量和误差自由度			F	df				
15	回归平方和及剩余平方和			SSreg	SSresid				
16									
17				显性效应	加性效应	总平均			
18				b2	b1	b0			
19	回归系数			=LINEST(F	8.508	105.4694			
20	回归系数的标准差			4.521282	2.446749	2.300655			
21	判断系数和Y估计值的标准差			0.825843	3.86865	#N/A			
22	F统计量和误差自由度			9.483886	4	#N/A			
23	回归平方和及剩余平方和			283.8803	59.86582	#N/A			
24									

假定返回的结果存放在单元格 D19:F23 表示的 5×3 矩阵中。选中存放结果的所有单元格 D19:F23，输入公式‘=LINEST(B2:B8, D2:E8, TRUE, TRUE)’，其中 LINEST 是 EXCEL 中计算最小二乘估计的函数，B2:B8 就是 7×1 观测向量  $\mathbf{y}$ ，D2:E8 就是 7×2 设计矩阵  $\mathbf{X}$ ，后面的两个逻辑参数 TRUE，表示除回归系数  $b_2$ 、 $b_1$  外还将返回截距项  $b_0$ ，同时还将输出与回归

显著性检验有关的统计量。然后使用三键组合‘Ctrl+Shift+Enter’，即得到函数 LINEST 的所有结果。

如愿意将常数项  $b_0$  也视为一个回归系数，这时的设计矩阵是区域 C2:E8 的  $7 \times 3$  矩阵。在利用 LINEST 函数时，要把常数项要强制设为 0。返回的结果仍是一个  $5 \times 3$  矩阵。只是把公式修改为‘=LINEST(B2:B8, C2:E8, FALSE, TRUE)’，就可以得到与之前完全相同的结果。

## 附录 A4: Excel 中的描述统计量

如下图所示, 20 个样本观测值保存在单元格 A2:B21, 其中包含 3 个缺失数据, 用‘\*’表示。

利用 COUNT 函数计算样本量 (不含缺失值)

利用 MIN 函数计算最小样本观测值

利用 MAX 函数计算最大样本观测值

利用 MEDIAN 函数计算样本观测值的中位数

利用 VAR 函数计算样本方差

利用 STDEV 函数计算样本标准差

利用 SKEW 函数计算样本的偏度

利用 KURT 函数计算样本的峰度

上述函数都只有一个返回值, 只需要在存放结果的单元格中写入相应的公式, 公式中指明样本观测值所在的单元格范围即可。图中还同时给出 2003 版和 2007 版函数的对照。

MDETERM		=FREQUENCY(A2:B21, C15:C20)					
	A	B	C	D	E	F	G
1	观测值 (A2:B21)		描述统计量	2003版函数使用方法	本例的结果	2007版函数使用方法	本例的结果
2	5.85	5.45	样本量	=COUNT(A2:B21)	37	同2003版	37
3	5.43	5.92	最小值	=MIN(A2:B21)	5.43	同2003版	5.43
4	5.60	5.74	最大值	=MAX(A2:B21)	6.4	同2003版	6.4
5	6.40	5.94	中位数	=MEDIAN(A2:B21)	5.85	同2003版	5.85
6	6.08	6.06	均值	=AVERAGE(A2:B21)	5.872972973	同2003版	5.872972973
7	6.10	6.37	方差	=VAR(A2:B21)	0.09311036	=VAR.S(A2:B21)	0.09311036
8	*	6.03	标准差	=STDEV(A2:B21)	0.305139903	=STDEV.S(A2:B21)	0.305139903
9	5.95	5.81	偏度	=SKEW(A2:B21)	0.262949197	同2003版	0.262949197
10	5.51	5.48	峰度	=KURT(A2:B21)	-1.093919853	同2003版	-1.093919853
11	5.48	5.59					
12	5.63	5.57	次数分布函数Frequency				
13	6.23	*	分组	组中值	频率		
14	6.05	5.65	5.3				
15	5.85	6.32	5.5	5.4	=FREQUENCY(A2		
16	6.18	5.60	5.7	5.6	8		
17	5.63	6.26	5.9	5.8	7		
18	5.81	6.39	6.1	6	9		
19	5.77	5.45	6.3	6.2	3		
20	6.38	*	6.5	6.4	5		
21	5.95	5.79					
22							

利用 FREQUENCY 函数, 可以计算样本观测值在不同分组中的频率。具体作法如下。在上面的 27 个观测值中, 最小值是 5.43, 最大值是 6.3。假定把这些观测值分成 6 组, 小于 5.5 是第 1 组, 5.5~5.7 是第 2 组等等, 见上图单元格 C15:C20。组距为 0.2, 组中值见单元

格 D15:D20。现在想要知道落入这 6 个组的样本数。

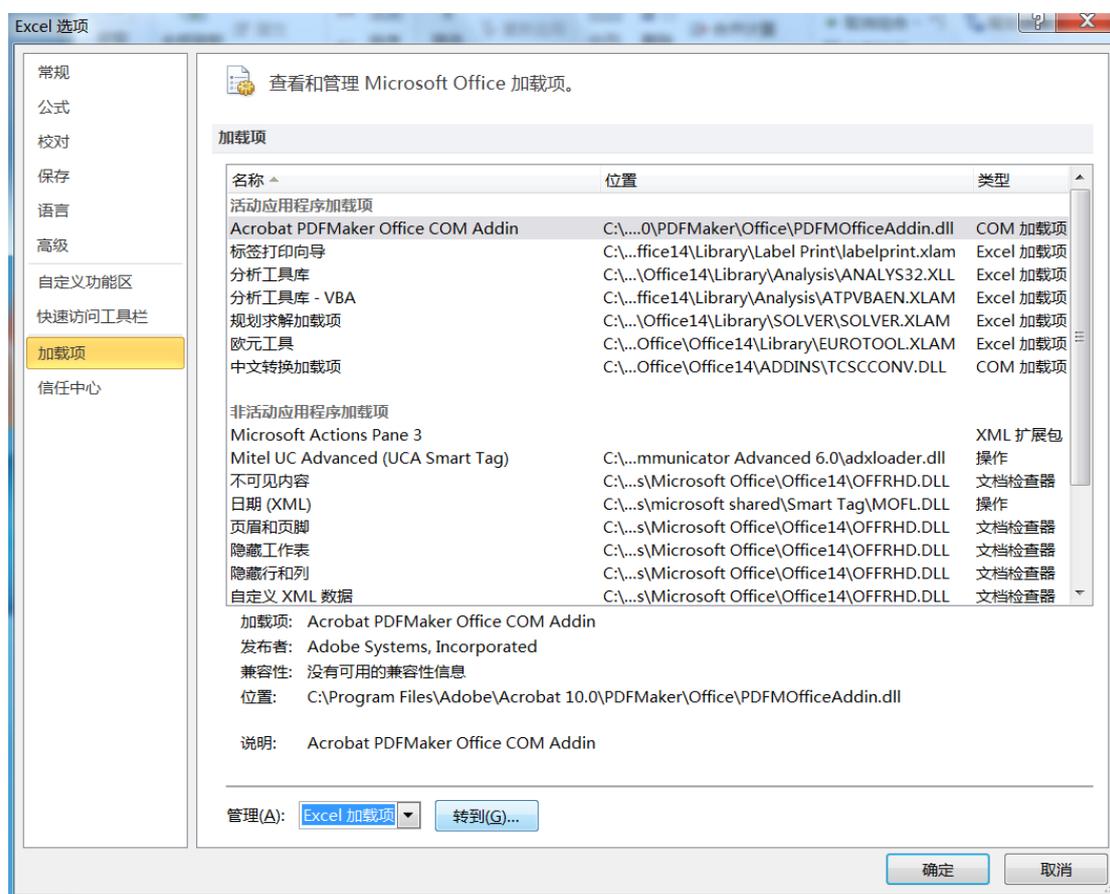
假定返回的结果存放在单元格 E15:E20 中，选中存放结果的所有单元格 E15:E20，输入公式‘=FREQUENCY(A2:B21, C15:C20)’，其中 FREQUENCY 是 EXCEL 中计算次数分布的函数，A2:B21 是待分组的观测值，C15:C20 是分组方法。然后使用三键组合‘Ctrl+Shift+Enter’，即得到落入 6 个组的样本数。将分组样本数除以总样本量，就得到频率。根据次数或频率，就可以在 EXCEL 中绘制次数分布图。

## 附录 A5: Excel 中的数据分析模块

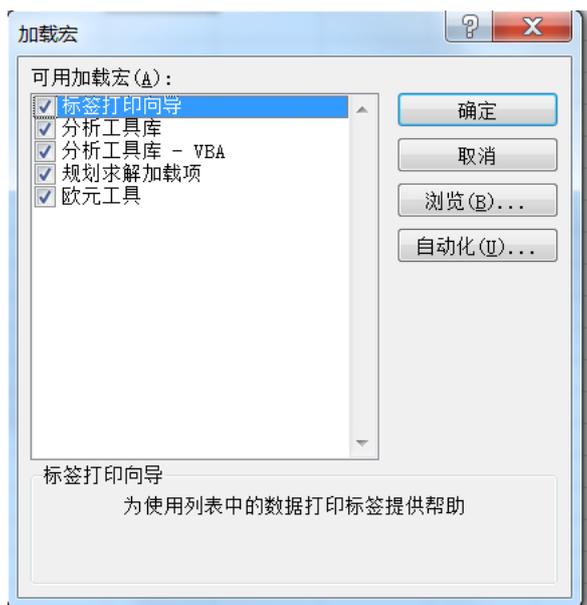
利用 EXCEL 的‘数据分析’模块，可以开展很多基本的统计分析，如方差分析、协方差分析、回归分析、 $t$  检验等等，甚至开展简单的蒙特卡洛模拟试验。第一次使用时，需要首先从 EXCEL 选项中加载这一模块。加载这一模块的步骤如下。

(1) 从‘文件’菜单中打开‘EXCEL 选项’对话框，如下图

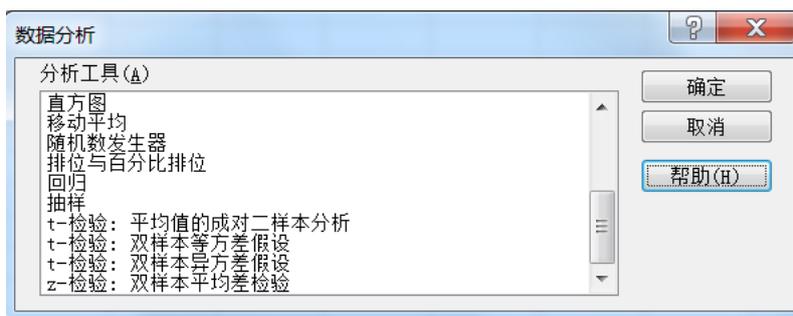
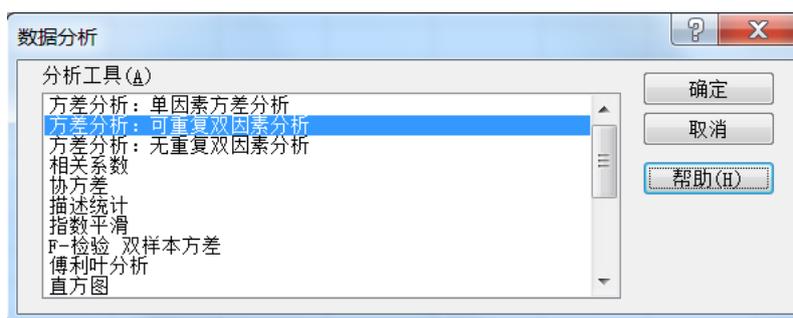
(2) 点击窗口下方的‘转到(G)...’按钮打开‘加载宏’对话框



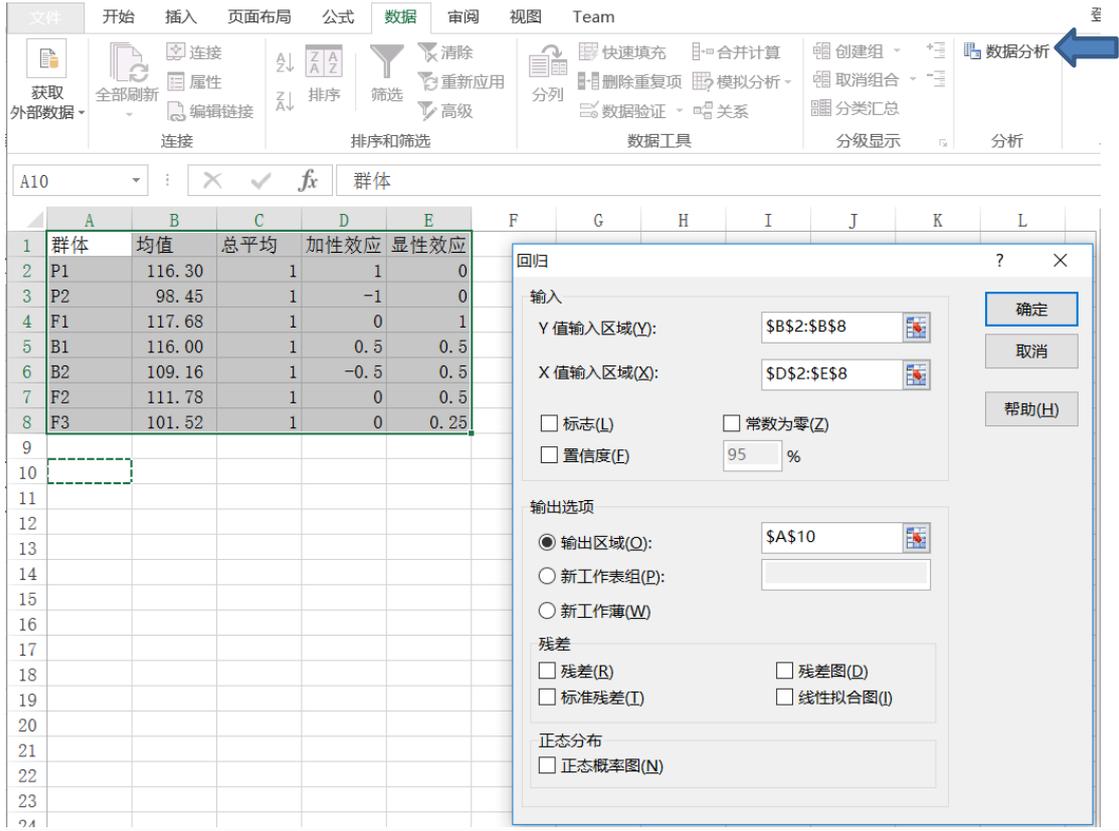
(3) 在‘加载宏’对话框中，如下图选中‘分析工具库-VBA’后，然后点击‘确定’



这样，在菜单‘数据’下，就能看到一个‘数据分析’按钮。点击这一按钮，就可以开展下图所示的近 20 种分析方法。



以附录 A3.7 的数据为例，说明分析工具中的‘回归’这一模块的使用。在‘数据’菜单中，点击‘数据分析’按钮以打开‘数据分析’模块；点击‘回归’可以看到下图的窗口。在‘回归’对话框中，点击‘Y 值输入区域(Y):’右边的输入框，用鼠标选中单元格 B2:B8；点击‘X 值输入区域(X):’右边的输入框，用鼠标选中单元格 D2:E8；点击‘输出区域(O):’右边的输入框，用鼠标选中单元格 A10；点击‘确定’按钮。



经过上述操作，点击‘确定’按钮后就得到下图单元格 A10:I28 的结果，这些结果在大多数情况下已经足够了。

群体	均值	总平均	加性效应	显性效应
P1	116.30	1	1	0
P2	98.45	1	-1	0
F1	117.68	1	0	1
B1	116.00	1	0.5	0.5
B2	109.16	1	-0.5	0.5
F2	111.78	1	0	0.5
F3	101.52	1	0	0.25

SUMMARY OUTPUT								
回归统计								
Multiple R	0.908759004							
R Square	0.825842928							
Adjusted R Square	0.738764392							
标准误差	3.868650337							
观测值	7							
方差分析								
	df	SS	MS	F	Significance F			
回归分析	2	283.8803211	141.9401606	9.483886233	0.030330686			
残差	4	59.86582171	14.96645543					
总计	6	343.7461429						
	Coefficients	标准误差	t Stat	P-value	Lower 95%	Upper 95%	下限 95.0%	上限 95.0%
Intercept	105.4693902	2.300655281	45.84319567	1.35418E-06	99.08174715	111.8570333	99.08174715	111.8570333
X Variable 1	8.508	2.446749307	3.477266746	0.02541505	1.714734863	15.30126514	1.714734863	15.30126514
X Variable 2	11.85609756	4.521281666	2.622286873	0.058658499	-0.696992789	24.40918791	-0.696992789	24.40918791