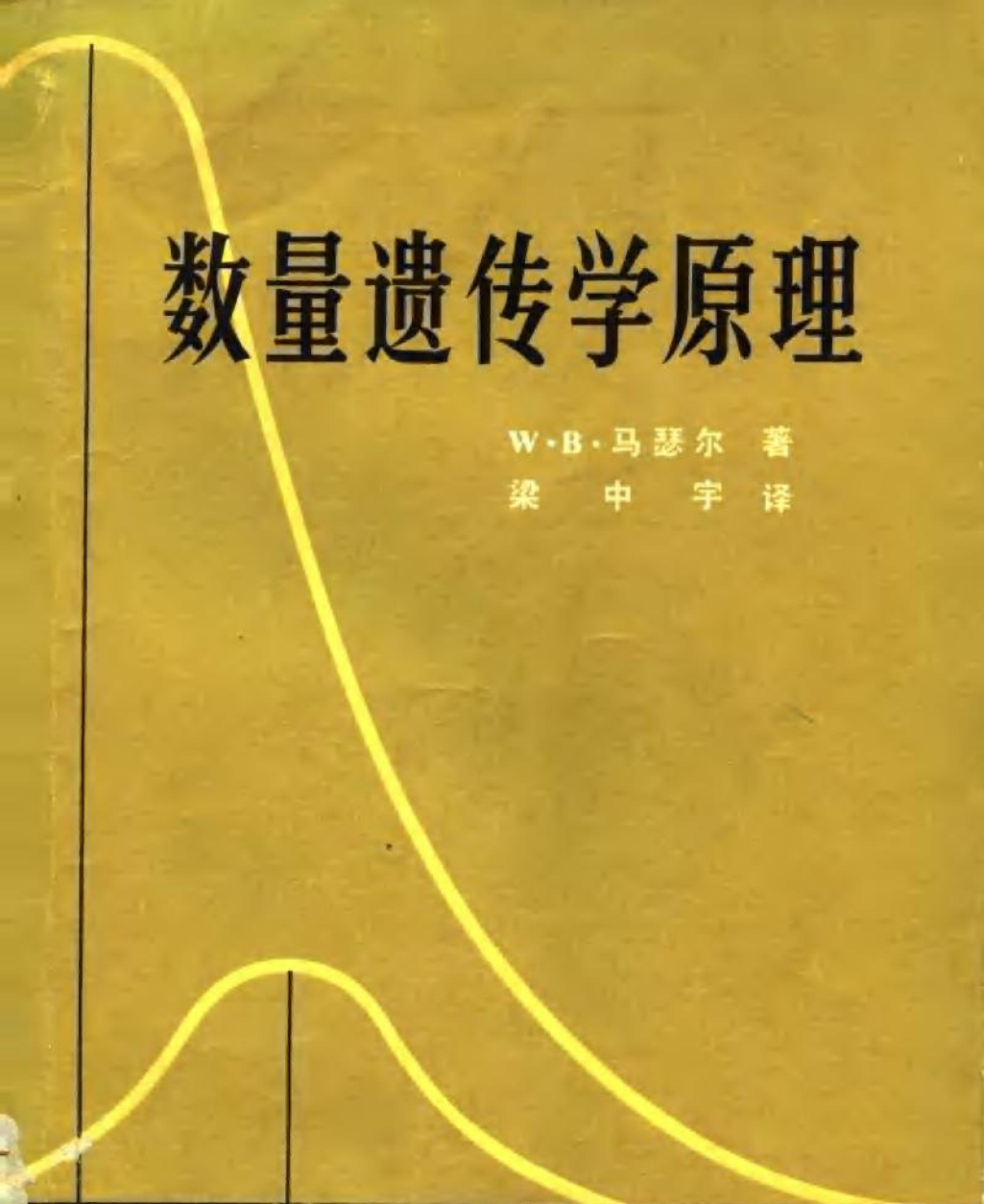


数量遗传学原理

W·B·马瑟尔 著

梁 中 宇 译



人民教育出版社

数量遗传学原理

W. B. 马瑟尔著

梁 中 宇 译

*

人民教育出版社出版

新华书店北京发行所发行

天津人民出版社印刷厂印装

*

开本 850×1168 1/32 印张 $5 \frac{3}{16}$ 字数 130,000

1978年11月第1版 1979年5月第1次印刷

印数 00,001—17,000

书号 14012·021 定价 0.49 元

作 者 序

我打算在这本书中把统计遗传学、群体遗传学以及系统遗传学的主要内容,作为一个整体扼要地讲一下。

我相信有三本书对于数量遗传学各个方面的讲解是很独到的,那就是: K. Mather, 1951: 《生物学中的统计分析》(Methuen 出版); C. C. Li, 1954: 《群体遗传学》(芝加哥大学出版); D. S. Falconer, 1961: 《数量遗传学概论》(Oliver 及 Boyd 出版)(有中译本,杨纪柯、汪安琦译,1965年,科学出版社出版——译者注)。若要更充分地讨论本书的主题,这些著作是应该参考的。

和熟知的Crow氏《遗传学笔记》(Burgess 出版)一样,这本书也有尽量避免乱写课堂笔记的意思,但在每张右面的一页又留有空白,以供教师和学生随自己的意思去作额外的笔记(译本为节约纸张,已省去——编者注)。仿照 Lerner: 《选择的遗传学基础》(Wiley 出版)一书,书中把对主题不关紧要的数学注记和证明分开来放在“栏”内。作习题对于彻底理解本书的内容是很重要的,附录中有这些问题的答案。

感谢下列作者和出版者允许我以修改了的形式使用某些表和图。

附录中表 I-IV 是取自 F. Yates 及 R. A. Fisher 《生物学、农学及医学研究用统计学表》(Oliver 及 Boyd 出版),而在 K. Mather 《生物学中的统计分析》(Methuen 出版)里修改过的。附录表 V 引自 F. Yates 及 R. A. Fisher 《生物学、农学及医学研究用统计学表》(Oliver 及 Boyd 出版)。图 3-5 引自 K. Mather 《生物学中的统计分析》(Methuen 出版)。图 9-1, 9-2, 9-3 以及表

17-1 引自 D. S. Falconer 《数量遗传学概论》(Oliver 及 Boyd 出版)。图 12-3 要感谢 Th. Dobzansky 及《演化》的编辑。图 18-1 则引自 I. M. Lerner 所著, 由芝加哥大学出版的《群体遗传学和动物的改良》。最后还要感谢 Rosalyne Spurway 女士对我未加过工的笔记准备了出版用的插图。

哲 学 博 士 Wharton B. Mather

1963 年 12 月于 Queensland 大学动物学系
遗传学实验室

目 录

作者序

第一篇 统计遗传学	1
第一章 基本概念	1
第二章 生物统计常数	5
第三章 显著性测验	15
第四章 方差分析	29
第五章 回归	36
第六章 协方差分析	47
第七章 相关	51
第八章 最大似然法	56
第二篇 群体遗传学	65
第九章 Hardy-Weinberg 定律	65
第十章 迁移及非频发突变	77
第十一章 频发突变	81
第十二章 选择	86
第十三章 突变及选择的联合影响	102
第十四章 理想化小群体	107
第十五章 非理想群体	116
第十六章 平稳基因频次分布	120
第三篇 生统遗传学	122
第十七章 表现型分量及其方差	122
第十八章 选择	140
第十九章 近交与杂交	148
第二十章 相关性状	153
附录 1. 附表	155
附录 2. 问题答案	163

第一篇 统计遗传学

第一章 基本概念

A. 引言

可以把统计学看作是对实验的数学处理,它的目的有两方面:

1. 把数据简化为少数简单常数;
2. 评定这些常数的显著性或重要性。

实验科学的最终目的是由特殊去论证一般,因此实验科学工作者涉及更多的是归纳逻辑,而非演绎逻辑的方法。

B. 总体及样本

统计学的出发点是样本取自总体。总体又可称作特征数,而样本则由所谓统计量的生物统计常数来定义。统计量在取样中会受到机率误差,但是正如我们将会见到的,如果一旦知道了统计量,就有可能在给定的概率下去指出特征数所在的范围。习惯上以希腊字母记特征数,而以拉丁字母去记统计量。

例如对平均数的记法有 μ 及 $\bar{x}$, 而对标准离差的则有 σ 及 s 。在本书中系用 S 而不是 Σ 去作总和的记法。

C. 概率

概率是统计学及遗传学中的基本概念之一,可用下述方式去定义。如在一系列 n 个事件中,有 m 个是以一种特殊方式出现的,则此特殊事件的概率 $P = \frac{m}{n}$ 。例如,果蝇中 e/e 呈黑檀色(ebony), E/e 及 E/E 呈野生型。在 $E/e \times e/e$ 的杂交中,配子为 E, e 及 e, e 。由 E/e 而来的配子为 e 的概率是 $1/2$, 来自 e/e 的配子为 e 的

概率是 1, 因此后代是黑檀色果蝇的概率为一半。

两独立事件同时发生的概率等于每个事件各自发生时概率的乘积。另一方面, 两独立事件单一发生的概率则为它们各自发生时概率的和。

例(a) 在下表中:

<i>AB</i>	<i>aB</i>	<i>Ab</i>	<i>ab</i>
9	3	3	1

有 3/16 的 *aB*, 3/16 的 *Ab*, 但有 6/16 的 *aB* 或 *Ab*。

例(b) *E/e* × *e/e* 杂交而得 5 只果蝇的家系中, 全部都是黑檀色果蝇的概率

$$P = 1/2 \times 1/2 \times 1/2 \times 1/2 \times 1/2 = \left(\frac{1}{2}\right)^5$$

全部都是野生型的概率同样也是

$$P = \left(\frac{1}{2}\right)^5 = \frac{1}{32}$$

这正是 $\left(\frac{1}{2} + \frac{1}{2}\right)^5$ 二项式展开时最前及最后两项。而在家系中有些是野生型, 又有些是黑檀色果蝇的概率, 则可由 $\left(\frac{1}{2} + \frac{1}{2}\right)^5$ 展开时的其它项来给出。概括起来就是

$$(p + q)^k = p^k + \dots \frac{k! p^r q^{k-r}}{r!(k-r)!}$$

其中 *p* 是第一类个体的期望概率, *q* 是另一类个体的期望概率, *k* 是家系里的个体数, *r* 是在一个特定家系里的第一类个体数。

例 *E/e* × *E/e* 杂交得 5 只果蝇的家系中, 4 只是黑檀色果蝇, 1 只是野生型的概率是多少?

由于 $E/e \times E/e$ 得 $1E/E:2E/e:1e/e$, 亦即 3 只野生型: 1 只黑檀色, 因此野生型的概率是 $3/4$, 而黑檀色的概率是 $1/4$ 。

$$\begin{aligned} \therefore P &= \frac{5! \left(\frac{1}{4}\right)^4 \left(\frac{3}{4}\right)^1}{4! \times 1!} \\ &= \frac{15}{1024} \\ &= 0.01465 \end{aligned}$$

D. 总体分析

1. 分组

在一群有机体中测量某些数量性状时, 会得出许多不同数字。若先把这些数字选拣安排在由某种任意数, 例如 5%, 所分别开来的组中, 就可用这种方法去得出测量结果的频次分布, 每组的中值则称为组中价。

2. 频次曲线

如以组价作横坐标, 各组个体数作纵坐标, 并按分组得知的情況去制图, 则得频次曲线。用这种方法可以对数据取得更清晰的概念。

3. 常态曲线

频次曲线有接近数学工作者们称之为常态曲线的趋向。常态曲线是在 k 为无穷大, 且 p 及 q 均不等于 $\frac{1}{k}$ 或较 $\frac{1}{k}$ 更小时, $(p+q)^k$ 的展开。这里, 我们直接提出了常态曲线与概率以及二项式展开的联系。

常态曲线的公式是:

$$m = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

其中 $m = x$ 组的频率, $e =$ 自然对数的底 $= 2.71828$, 而 $\pi = \frac{22}{7}$ 。

注意在公式中既无 p, q , 也没有 k , 但有两个新的特征数 μ 及 σ 。同时也可注意 p 及 q 是由假设固定下来的 (在前面有关概率的例子中, $p = \frac{1}{4}, q = \frac{3}{4}$), 但这两个特征数却不由假设来固定, 而能从数据中去估计出来。其中 μ 称作平均数, σ 则称为标准离差。它们的图象可参阅(图 1-1)。

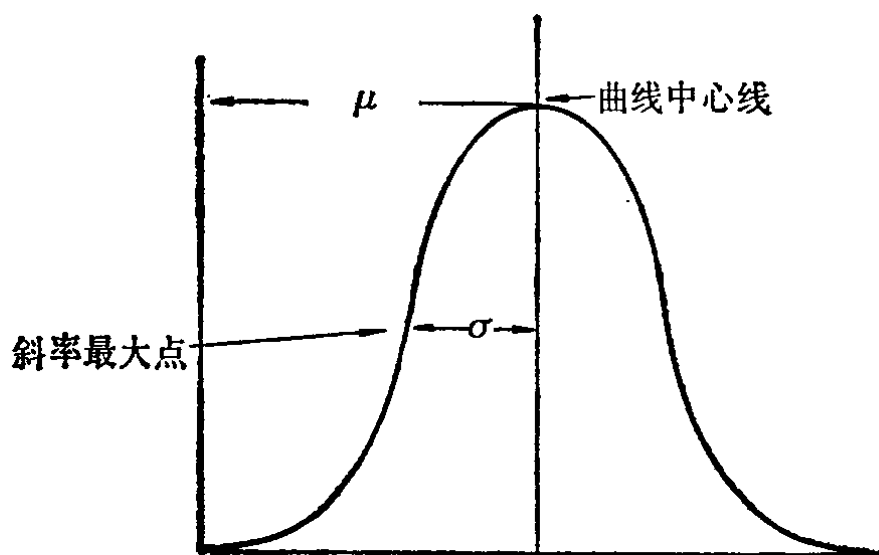


图 1-1

问 题

1. 对隐性蓝眼是杂合的棕眼双亲:
 - a. 生出五个小孩在姊妹顺序上为蓝眼、棕眼、蓝眼、蓝眼、棕眼的概率如何?
 - b. 五个小孩中有三个蓝眼及两个棕眼(不计姊妹顺序)的概率如何?
2. 某人的祖母及外祖父都曾患过一种由于纯合隐性所引起的疾病, 试问此人:
 - a. 患这种病的概率?
 - b. 是传递者的概率?
3. 不计互换时,
 - a. 一个人全部来自父亲的常染色体均自祖父遗传而来的概率?
 - b. 全部来自母亲的常染色体均自外祖父遗传而来的概率?

第二章 生物统计常数

A. 平均数及标准离差

1. 公式

a. 平均数 $\bar{x} = \frac{Sx}{n}$

b. 标准离差——离散性的测量

最简单的离散性测量,似乎是不管 x 的正负而等于 $\frac{S(x-\bar{x})}{n}$

的,观测值对平均数的平均离差。但这种正负号的困难,可由在总和之前,把所有观测值与平均数的离差(原文仅指观测值是不妥当的——译者注)都予以平方,而在除以 n 后,再取其平方根而得到克服。更由于在下一节中会提到的原因,分母还应是 $n-1$ 而不是 n 。这就使我们得出了对标准离差的一种估计,

$$s = \sqrt{\frac{S(x-\bar{x})^2}{n-1}}$$

其中 x 为个体测量值, n 为测量值的个数。

$s^2 = \text{方差} = V = \text{与平均数离差的平方和}(SS)$

被 $n-1$ 或 N (自由度数)除,也就是平均平方(MS)。

涉及成对而非单独的观测时,我们会遇到 x 及 y 的协方差,

$$W_{xy} = \frac{S[(x-\bar{x})(y-\bar{y})]}{n-1}$$

其中 $S(x-\bar{x})(y-\bar{y})$ 称作 x 与 y 的叉积和。

有些时候,如使用计算机时,以下的公式是很方便的,

$$S(x-\bar{x})^2 = Sx^2 - \frac{1}{n}S^2x$$

其中 Sx^2 称作平方和，而 S^2x 则为和的平方。它的正确性可从以下的论证看出来。（见 2-1 栏）

2-1 栏

$$S(x - \bar{x})^2 = Sx^2 - \frac{1}{n}S^2x \text{ 的证明}$$

如果引进一个任意暂定的平均数 x_m ，则有

$$(x - \bar{x}) = (x - x_m) - (\bar{x} - x_m)$$

两端平方时

$$(x - \bar{x})^2 = (x - x_m)^2 - 2(x - x_m)(\bar{x} - x_m) + (\bar{x} - x_m)^2$$

总和起来，并因 $\bar{x}$ 及 x_m 都是常数，则有

$$\begin{aligned} S(x - \bar{x})^2 &= S(x - x_m)^2 - 2S[(x - x_m)(\bar{x} - x_m)] + S(\bar{x} - x_m)^2 \\ &= S(x - x_m)^2 - 2(\bar{x} - x_m)S(x - x_m) + n(\bar{x} - x_m)^2 \end{aligned}$$

另据 $\bar{x}$ 的定义

$$\begin{aligned} n(\bar{x} - x_m) &= n\bar{x} - nx_m \\ &= S(x) - nx_m \\ &= S(x - x_m) \end{aligned}$$

用它去代换前式中的 $(\bar{x} - x_m)$ ，并把 x_m 取成 0 时，就可得

$$\begin{aligned} S(x - \bar{x})^2 &= S(x - x_m)^2 - \frac{2}{n}S^2(x - x_m) + \frac{1}{n}S^2(x - x_m) \\ &= S(x - x_m)^2 - \frac{1}{n}S^2(x - x_m) \\ &= Sx^2 - \frac{1}{n}S^2x \end{aligned}$$

c. 二项式的平均数及标准离差

$\mu = p$ 或 q ，由此得组平均数令人感兴趣的是 $\sigma = \sqrt{\frac{pq}{k}}$ ，其中

p 及 q 是比例， k 是家系里的个体数。（见 2-2 栏）

$\mu = p \text{ 或 } q, \sigma = \sqrt{\frac{pq}{k}}$ 的证实

公式的真实性可以从估计某些数据的 $\bar{x}$ 及 s 时，显然有 $\bar{x} \rightarrow p \text{ 或 } q, s \rightarrow \sqrt{\frac{pq}{k}}$ 而看出来。仍以 E/e 与 e/e 果蝇回交时，个体有一半的机会是黑檀色果蝇为例。在有 5 个体的家系中，根据上述公式可得：

$\mu = p = 0.5$

$\sigma^2(\text{方差}) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{5} = 0.05$

也可以从下面列举的数据，按每家有 5 只果蝇的 n (原文误作 k ——译者注)个家系去估计 $\bar{x}$ 及 s

野生型的 比例(x)	发现频 次(a)	ax	ax^2
0.0	0	0.0	0.00
0.2	2	0.4	0.08
0.4	5	2.0	0.80
0.6	6	3.6	2.16
0.8	3	2.4	1.92
1.0	1	1.0	1.00
17		9.4	5.96

$\bar{x} = \frac{Sax}{n} = \frac{9.4}{17} = 0.55$ 与 $\mu = 0.5$ 是近似的。

注：实际数系由比例乘频次得出。

$$\begin{aligned}
 S(x-\bar{x})^2 &= Sx^2 - \frac{1}{n}S^2x \\
 &= 5.96 - \frac{(9.4)^2}{17} \\
 &= 0.76
 \end{aligned}$$

$$s^2 = V = \frac{S(x-\bar{x})^2}{n-1} = \frac{0.76}{16} = 0.05$$

这和 $\sigma^2 = 0.05$ 是相似的。

μ 与 $\bar{x}$ 以及 σ^2 与 s^2 显然是相似的, 因此就说明了上述公式的真实性。

2. 用 N 代替 n (Bessel 氏修正)

选择 $n-1$, 亦即自由度 N 去代替 n 作为求方差的分母的原因, 可以这样来解释:

由于方差是对曲线“离散性”的一种测度, 因此离散性愈大, x 值两两之间的差别也就愈大。两值间有一个差别, 因此对 n 个值就有 $n-1$ 个差别或自由度。

每有一个参数被估计出来, 就会失去一个自由度, 这是在实际工作中的一条规则。例如计算 s 时平均数就是已经估计出来了的。

3. Sheppard 对分组的修正

数据分组虽然在减少计算劳动上是很有用的 (例如求平均数时, 可以频次乘组价的相加去代替冗长的相加), 但却微微地增加了曲线的离散, 以致标准离差增大而导致不准确。Sheppard 修正是在估计标准离差时, 在开平方之前, 先减去组距的 $1/12$ 。

B. 平均数的标准误差

平均数的标准误差测度着由整个总体抽取一些样本所计算出来的平均数的离散性。其公式为:

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} \text{ (见 2-3 栏的推导)}$$

注：由此可以看出， n (样本中个体数) 增加则平均数的标准误差减小，因此增加样本含量就可减少平均数的标准误差。

2-3 栏

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} \text{ 的证明}$$

令 x 和 y 是成对的测量值，并令样本含量为 n 对，则

$$V_x = \frac{S(x - \bar{x})^2}{n-1}, \text{ 而 } V_y = \frac{S(y - \bar{y})^2}{n-1}$$

以 $x+y$ 代替 x 时，

$$V_{x+y} = \frac{S[x+y - \overline{(x+y)}]^2}{n-1}$$

我们要求它呈一种完全平方的形式。由于

$$\overline{x+y} = \frac{S(x+y)}{n} = \frac{Sx}{n} + \frac{Sy}{n} = \bar{x} + \bar{y}$$

因此
$$\frac{1}{n-1} S[x+y - \overline{(x+y)}]^2$$

$$= \frac{1}{n-1} S[(x - \bar{x}) + (y - \bar{y})]^2$$

$$= \frac{1}{n-1} S(x - \bar{x})^2 + \frac{2}{n-1} S[(x - \bar{x})(y - \bar{y})] + \frac{1}{n-1} S(y - \bar{y})^2,$$

或
$$V_{x+y} = V_x + \frac{2}{n-1} S[(x - \bar{x})(y - \bar{y})] + V_y$$

x 与 y 独立时便有

$$V_{x+y} = V_x + V_y$$

这点可以从以下的论证而看出来：

情况(i)

与大 x 值伴同的是大的或小的 y 值时，得到正 $x - \bar{x}$ 值与

正或负的 $y - \bar{y}$ 值相伴, 它们相乘再加起来就会得零。

$(x - \bar{x})$		$(y - \bar{y})$			
+	×	+	=	+	
+	×	-	=	-	
+	×	+	=	+	
+	×	-	=	-	
	•			•	
	•			•	
	•			•	
					0

情况(ii)

与小 x 值伴同的是大的或小的 y 值时, 得到负 $x - \bar{x}$ 值与正或负的 $y - \bar{y}$ 值相伴, 它们相乘再加起来时也会得零。

$(x - \bar{x})$		$(y - \bar{y})$			
-	×	+	=	-	
-	×	-	=	+	
-	×	+	=	-	
-	×	-	=	+	
	•			•	
	•			•	
	•			•	
					0

如 x 与 y 相关, 则有正或负的 $\frac{2}{n-1} S[(x - \bar{x})(y - \bar{y})]$ 。

情况(i)

伴同大 x 值的主要是大的 y 值时, 相乘再加起来会得到

正值。正如

$(x-\bar{x})$		$(y-\bar{y})$		
+	×	+	=	+
+	×	+	=	+
+	×	-	=	-
+	×	+	=	+
	•			•
	•			•
	•			•
				+

情况(ii)

与小 x 值相伴的主要是小 y 值时, 相乘再加起来, 也会得出正值。

$(x-\bar{x})$		$(y-\bar{y})$		
-	×	-	=	+
-	×	-	=	+
-	×	+	=	-
-	×	-	=	+
	•			•
	•			•
	•			•
				+

情况(iii)

伴同大 x 值的主要是小 y 值, 相乘再加起来时, 得出负值。

$(x-\bar{x})$		$(y-\bar{y})$		
+	×	-	=	-
+	×	-	=	-
+	×	+	=	+
+	×	-	=	-
	•			•
•	•			•
•	•			•

情况(iv)

伴同小 x 值的主要是大 y 值，相乘再加起来时，也得负值。

$(x-\bar{x})$		$(y-\bar{y})$		
-	×	+	=	-
-	×	+	=	-
-	×	-	=	+
-	×	+	=	-
	•			•
•	•			•
•	•			•

概括起来就是 $V_{x+y} = V_x + V_y + 2W_{xy}$ ，亦即，方差是可加的，但标准离差却不。这点在生统遗传学中很重要(见第三篇)。正如， $4 = 2 + 2$ ，但 $\sqrt{4} = 2 \neq \sqrt{2} + \sqrt{2}$ 。

前面已曾看到，独立时上式可简化为 $V_{x+y} = V_x + V_y$ 。而 $V_{x-y} = V_x + V_y - 2W_{xy}$ 在 x 与 y 独立时，也同样可简化为