农村科学实验丛书

应用生物统计

杨纪珂　孙长鸣　汤旦林　编著

应用生物统计

杨纪珂　孙长鸣　汤旦林　编著

科学出版社
1983

内 容 简 介

本书为《农村科学实验丛书》之一，它以应用于生物、农业、畜牧及医学等方面的大量例证，比较系统全面地、深入浅出地阐明数理统计的基本原理和方法。主要内容有调查研究的科学方法；多快好省地作科学实验；科学种田实验；科学饲养实验；医、药科学实验；多因素正交实验设计；事物间的相互联系；动物群体生态实验；遗传育种实验等。书中还附有27种统计用表。为生物、农业、畜牧业、医药工作者从事数理统计工作的一本通俗易懂的入门读物。

应 用 生 物 统 计

杨纪珂　孙长鸣　汤旦林 编著

责任编辑　蒋伯宁

科 学 出 版 社 出版

北京朝阳门内大街137号

中国科学院印刷厂印刷

新华书店北京发行所发行　各地新华书店经售

*

1983年1月第 一 版　开本：787×1092 1/32

1983年1月第一次印刷　印张：13 5/8

印数：0001-9,400　字数：310,000

统一书号：13031·2100

本社书号：2864·13-10

定价：1.70元

前　言

生物科学以及与生物科学有关联的农业科学和医学科学，在最近几十年中突飞猛进，其进展之迅速，为其它学科所不及。之所以如此，是跟其它学科如数学、物理学、化学、地学、力学等理论上的渗透与促进，以及近代技术如电子计算机、生物物质分析、电子显微镜等技术上的提高与应用分不开的。在与生物学有关的边缘科学的蓬勃发展中，最为基本而突出的进展，就是包括农学和医学在内的生物学的定量化趋势。不论生物科学的哪个领域，都正在纷纷向定量化的领域迈步前进，不再以过去那种定性地描述各种生物现象为满足。而在生物学的应用方面，单靠定性的指导，在许多场合下已经不能应付实际的需要，只有通过定量的科学实验和调查研究加上定量的数据分析和总结，才能更为有效地解决实际的问题。因此，学习其中方法的要求也与日俱增。有鉴于此，我们准备了这本为有志于学习该方面的知识和想把它们应用在实际工作中的同志们参考之需的读物。其目的一方面想把这门边缘学科，从几个侧面来勾划出它的粗线条轮廓，使读者对它有个鸟瞰，借以开阔眼界，知其所由；另一方面也确实想使这些方法在生产实践中得以应用，例如对在科学种田实验中所需知识的介绍其目的就在于此。

为了使读者们易晓易学，本书力图深入浅出，通俗易懂。

但由于我们才学所限，难免眼高手低，意虽殷而力不能及，谬误之处，更是在所难免。深望读者多加指正，多提宝贵的意见。

作　　者

1978年8月

目　录

第一章　调查研究的科学方法

第一节　什么叫概率

在数学中，把在自然界中无论什么事件出现的可能性称为概率，而且设法用数值去度量它。有人可能要问：一头猪有多重，可以称它的斤数；一只玉米有多长，可以量它的毫米数；你现在的体温，可以测量其摄氏度数，但是某事件出现或发生的可能性有多少，却怎样去度量呢？数学家们想出了好办法，使这种可能性也能度量，并且可以进行科学分析和推导，得出正确的结论。

怎样度量事件发生的可能性呢？先看两种极端的情况。一种情况是某事件必然会发生，譬如说一头猪总有一天要死。另一种情况是某事件决不可能发生，譬如说西天出太阳。在数学上就把类似于一头猪有一天会死这类事件发生的可能性用数字 1 来度量，说成一头猪有一天会死的概率等于 1。为了写起来省事，把概率记成符号 P，写成

$$P(\text{一头猪有一天会死}) = 1$$

把类似于西天出太阳一类事件发生的可能性用 0 来度量，写成

$$P(\text{西天出太阳}) = 0$$

但是，在我们实际所遇到的事件中，属于这两类的是比较少的。大多数是在这绝对可能与绝对不可能两类之间，即它们的概率介乎 1 与 0 之间。事件 s 发生的可能性愈大，其

概率就愈靠近 1；发生的可能性愈小，就愈靠近 0，但是概率不会越出 0 和 1 的区间之外。即

$$0 \leqslant P(s) \leqslant 1 \tag{1.1}$$

怎样才能度量出某一事件发生的概率呢？不妨举几个例子来说明。第一个是生男育女的概率的例子。我们知道，一对夫妇生男育女，究竟生男还是生女在怀孕之前是不知道的。但生男或生女的概率，也就是生男或者生女的可能性，是可以度量的。让我们看一看从生物学的实验所得出的生男育女的规律。一个人的性别取决于她（或他）身上每个细胞内细胞核中的一对性染色体。如果是女的，她的一对性染色体大小形状相同，我们用符号记为 XX；如果是男的，他的一对性染色体是 XY，其中只有一个是X，另一个是Y，比 X小得多。但是在生殖细胞中，染色体的数目要减半。这使女性排出的各个卵子内的性染色体都只含一个X；使男性所排出的各个精子内的性染色体也只含一个，或者是X，或者是Y。因为在精母细胞中X和Y的数目是相等的，所以在男性排出的精子中，含X的精子和含Y的精子在数目上大约相等，应该各占一半。假设每个精子都有机会使一个卵子受精，从而产生一个胎儿的话，那么男胎（XY）和女胎（XX）出现的百分率应该大约各有 50%。这可以用图解表示（图 1.1）。

虽然一对夫妇所生育的子女不多，不能验证上述的理论比例，但是如果把许多对夫妇的子女合起来计算，就可以发现子女的数目相差不多，大约各占 50%。跟理论上的说法基本上是吻合的。因此，在人数比较众多的情况下，预期男胎出现的百分率近于 50%，女胎的也近于 50%。现在如果有个张大嫂怀了孕，不知道是男胎还是女胎，我们就可以有科学根据地说她怀男胎的可能性有一半，怀女胎的可能性也有一半，

写成概率的符号，这就是

$$P(\text{男胎})=0.50 \qquad P(\text{女胎})=0.50$$

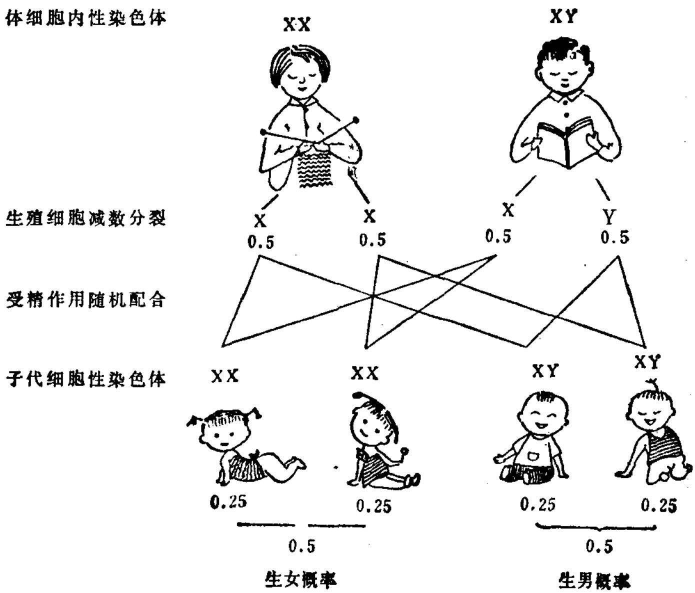

图1.1 男女胎出现的概率

这种概率因为是从科学理论推导出来，然后经过事实或实验证明了的，称为理论概率。

再看一个怎样去估计得病的概率的例子。在旧社会里，卫生差、医院少、药费贵，人们得结核病的很多；一个好好的人，在3、5年内传染上了这种病也很难说。譬如说1940年有人在刘庄地区调查了一万个人，到了1943年再去调查，发现原来健康的人在这三年中有10%传染上了结核病。换句话说，调查结果告诉我们：刘庄1940—1943年间新得结核病

的病例的发生频率是 10%或 0.10。(频率与概率不同的地方在于：频率是事件已发生以后所占总数的百分率，概率则是在事件未发生以前它会发生的可能性的度量。它们在数字上尽管可以相同，但在意义上大有区别。)有了这样一个经验数字，我们就可以推想：如果刘庄每个人传染上结核病的机会都一样，那么在以后三年内，即 1943—1946 年的期间内，任一个刘庄人得结核病的概率或可能性是 0.10，可写成为

$$P(\text{1943—1946年刘庄某人得结核病})=0.10$$

这种概率的度量是根据大量调查得来的频率，它是一个从经验得来的值。象这种概率我们称之为经验概率。

解放后，全国不断开展爱国卫生运动，大家讲究卫生，医院成倍成倍地增加，药费也便宜许多，赤脚医生在农村更是大起作用，结核病的发生频率也就随之而大大降低。譬如说从 1974—1977 年在刘庄再作调查，得到的数字说明该区在三年内新发生结核病的频率已降低到 0.5% 或 0.005。因此可以同样推测一个健康的刘庄人在以后的三年内得结核病的概率已下降到 0.005 以下，可以写成为

$$P(\text{1978—1981年刘庄某人得结核病})\leqslant 0.005$$

一般说来，经验概率必须积累相当多的数据或资料来定。这些数据或资料从哪里来呢？那就是通过实事求是、没有偏倚的调查和实验。譬如说某奶牛场饲养了一群奶牛。这群奶牛历年所产牛奶的乳脂率平均 3.54%，有 18%的奶牛其乳脂率超过4%。这样，如果不引进群外种公牛的冷冻精液配种，只限于在这闭锁群内配种，那么可以估计某头行将泌乳的年青母牛的乳脂率F超过 3.54%或不到 3.54%的概率都是 0.50，即

$$P(F>3.54\%)=P(F<3.54\%)=0.50$$

它的乳脂率超过4%的概率则是0.18，即

$$P(F \geqslant 4\%) = 0.18$$

但如在调查时把劣牛撇掉不管，结果使平均乳脂率超过4%的牛数占总数的百分率大为提高，从而使这项概率的估计也随着增高。这样就不实事求是，就缺乏代表性。所得数据非但因发生偏倚而归于无效，而且成为这项科研工作中的大害。

第二节 概率的相加和相乘

曾听到过一段有趣的对话：

张：你老兄快做爸爸啦，恭喜恭喜。你猜过没有？将是个女孩，还是个男孩？

李：那谁能猜得中呢？

张：我一定能猜得中！

李：一定吗？

张：我猜呀，不是得女，就是得男。

张、李：哈哈，……！

这虽是个笑话，但其中确实包含有数学的道理。老李的妻子生孩子是一桩将要发生的未知事件。如果只记单胎，不计双胞胎、怪胎及其它不幸事件，那么不是得女就是得男。换句话说，全部可能发生的事件就只有这两种，而且这两桩事件的发生是互不相容的。所谓互不相容，就是不能兼而有之的意思。我们已经从生物学理论知道，在未分娩之前，生女生男的可能性大致各有一半，其概率各为0.5。用符号来表示，就可以写成

$$P(\text{女}) = P(\text{男}) = 0.5$$

老李的妻子生孩子（女或男）这桩事必然要发生，发生的可

能性是百分之百，也就是发生这一事件的概率等于1。写成式子是

$$P(\text{女或男})=1$$

于是不难看到，在生孩子这件事与生男育女这两件事发生的概率之间存在着相加的关系，就是

$$P(\text{女或男})=P(\text{女})+P(\text{男})=0.5+0.5=1$$

其实不单生孩子如此，在农业、工业、医学卫生、科学研究以及日常生活中，无处不遇到概率相加的例子。

譬如，某品种的玉米种子的发芽率从以往的经验知道是80%。如对同一品种的玉米种子用同样方法处理，我们就可以事先猜度任一粒玉米发芽的概率是0.8，不发芽的概率是0.2。发芽与不发芽这两件事也是互不相容的，同样可得：

$$\begin{aligned}P(\text{发芽或不发芽})&=P(\text{发芽})+P(\text{不发芽})\\&=0.8+0.2=1\end{aligned}$$

再如某工厂制造一种拖拉机零件。在1950年，经实事求是的随机抽样调查，其次品率达30%，于是如在其中任抽一件，在未经检验它之前猜度它是次品的概率是0.3，是合格的概率是0.7，可以写成

$$\begin{aligned}P(\text{次品或合格品})&=P(\text{次品})+P(\text{合格品})\\&=0.3+0.7=1\end{aligned}$$

但概率不是一成不变的。这个工厂到了1977年，管理和技术大有进步，这一年的次品率据抽样调查，只有0.1%。于是可以估计，对其中未经检验的零件来说，它属次品的概率是0.001，属合格的概率为0.999，列成式子，是

$$\begin{aligned}P(\text{次品或合格品})&=P(\text{次品})+P(\text{合格品})\\&=0.001+0.999=1\end{aligned}$$

全部互不相容的事件有时不止两种。譬如在一个水库里

放养青、草、鲢、鳙 4 种淡水鱼。放养的比例是青鱼和草鱼各 30%、鲢鱼和鳙鱼各 20%。如果有人去钓鱼，在未钓之前，我们可以估计他如钓到一条鱼的话，钓到青鱼的概率是 0.3、草鱼也是 0.3、鲢鱼是 0.2、鳙鱼也是 0.2（需假定它们的存活率都相同，钓到它们也都同样地容易）。如果湖中只有这 4 种鱼，那么全部的互不相容的事件也只有这 4 种。同样可以得出这样的关系

$$P(\text{青、草、鲢或鳙})=P(\text{青})+P(\text{草})+P(\text{鲢})+P(\text{鳙})=0.3+0.3+0.2+0.2=1$$

如果事件互不相容，但是不包括余部的事件，那么这桩或者那桩事件发生的概率等于各桩事件发生的概率之和，不过这时所得的值小于 1。在上例中，如只问钓到青鱼或草鱼的概率，则有

$$P(\text{青或草})=P(\text{青})+P(\text{草})=0.3+0.3=0.6$$

也就是说，钓到一条青鱼或草鱼的概率是 0.6。同样，钓到鲢鱼或鳙鱼的概率是

$$P(\text{鲢或鳙})=P(\text{鲢})+P(\text{鳙})=0.2+0.2=0.4$$

钓到青鱼、草鱼或鲢鱼的概率是

$$P(\text{青或草或鲢})=P(\text{青})+P(\text{草})+P(\text{鲢})=0.3+0.3+0.2=0.8$$

于是可以更概括地总结出一条规律：几桩互不相容的事件发生的总概率等于各桩事件发生概率之和。这就是概率相加定律。把全部互不相容的可能事件发生的概率总加起来，应等于 1，是概率相加定律中的一种特殊情况。

如果事件并非互不相容，这个定律就不再适用。譬如某城市中学生患砂眼的有 3%、近视眼有 15%。有了近视的人还可能患砂眼，患砂眼的人也不一定不近视。因此，这两桩

事不是互不相容的，概率相加定律就不适用，而必须采用概率相乘定律。现假定这两桩事件的发生并无关联性，可以设想：在那15%近视的学生中也可能有3%患砂眼，或者在那3%患砂眼的学生中有15%属近视。不论怎样，既是近视又患砂眼的学生应该占总数的15%×3%＝45/10,000或0.45%。如从概率的角度看，从学生总名册中任意点一人，他是近视的概率为0.15，他是砂眼患者的概率为0.03，即

$$P(\text{近视})=0.15,\quad P(\text{砂眼})=0.03$$

他既是近视眼又是砂眼患者的概率为0.0045，即

$$P(\text{既是近视又是砂眼})=P(\text{近视})\times P(\text{砂眼})=0.15\times 0.03=0.0045$$

于是我们又得出一条有关概率的定律：一些互无关联的事件依次或一齐出现的概率等于它们各自发生的概率之积，这就是概率相乘定率。这里互无关联4字很是要紧。如果两桩事件的发生彼此有些关系，这个相乘定律就归于无效。举个例，大家知道胸围跟体重有关，体重大的人胸围往往也要大些。以成年汉族男子而论，如已知胸围超过83厘米的概率为0.05，体重超过75公斤的概率为0.01。如果错用相乘定律，那么，胸围既超过83厘米、体重又超过75公斤的概率为0.05×0.01＝0.0005，只有万分之五了。实际上这1%体重超过75公斤的胖子的胸围都超过了83厘米，在其中任抽一人，这两种事件兼而有之的概率就等于0.01，远超过0.005。对于这类互有关联的事件，我们可以使用条件概率来表达。

如果汉族成年男子凡是体重超过75公斤的胸围都超过83厘米，则在体重$W\geqslant 75$公斤的条件下胸围$C\geqslant 83$厘米的概率等于1。如把这句话写成为一个数学等式，它就是

$$P(C \geqslant 83\text{厘米} | W \geqslant 75\text{公斤}) = 1$$

其中一短竖代表"在…的条件下"。此外，体重超过 75 公斤的概率为 0.01，这句话可以写成

$$P(W \geqslant 75\text{公斤}) = 0.01$$

从常识就可以判断这两桩事件兼而有之的概率也是 0.01，可以写成

$$P(C \geqslant 83\text{厘米}, W \geqslant 75\text{公斤}) = 0.01$$

中间的"，"代表"既是…，又是…"。显然，在以上三项概率之间存在着另一种相乘的关系，即：

$$P(C \geqslant 83\text{厘米}, W \geqslant 75\text{公斤}) = P(W \geqslant 75\text{公斤}) \times P(C \geqslant 83\text{厘米} | W \geqslant 75\text{公斤})。$$

如用符号 A 记事件 $C \geqslant 83$厘米，用符号 B 记事件 $W \geqslant 75$公斤，则上式可简写成为

$$P(A, B) = P(B)P(A|B) \tag{1.2}$$

这就成了一个通用公式，在这个公式中还包含有概率相乘定律。请看，如果事件 A 与事件 B 的发生互无关联，显然不论 B 的条件是什么，对 A 的发生与否都无影响。此时 $P(A|B) = P(A)$，代入上式得

$$P(A, B) = P(A)P(B) \tag{1.3}$$

这就是概率相乘定律的简写。

$P(A|B)$ 跟 $P(A)$ 相等与否还可以用来验证 A 与 B 关联性的有无。在上例中

$$P(C \geqslant 83\text{厘米} | W \geqslant 150\text{公斤}) = 1$$

$$P(C \geqslant 83\text{厘米}) = 0.05$$

两者不相等，所以验证出在体重与胸围之间具有关联性。

概率的相乘与相加还可以联合起来运用。譬如，调查有 3 个孩子的家庭，按子女排行次序，可以有 8 种不同的排行：