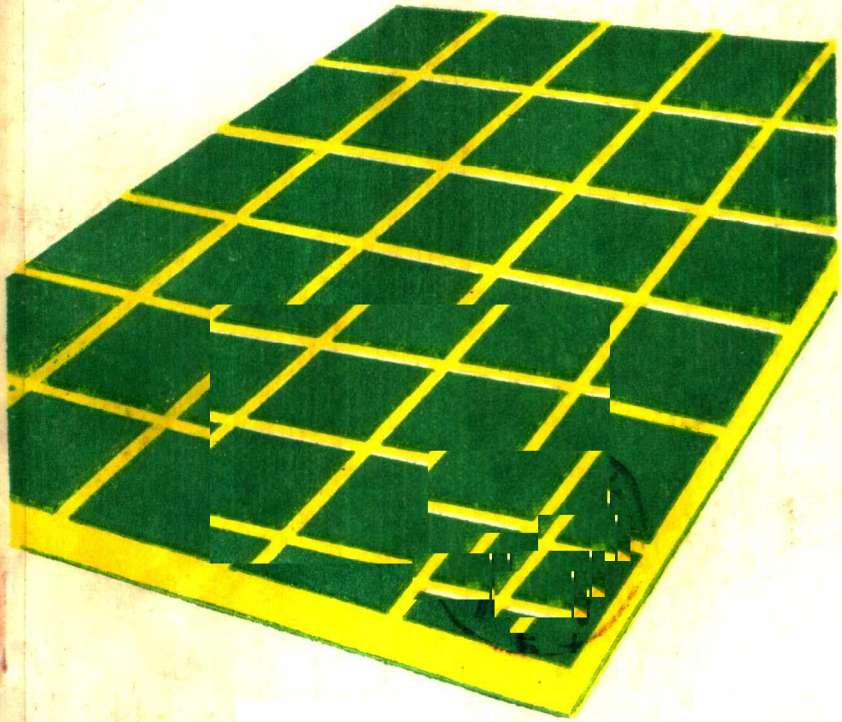
# 农业生物统计手册

NONGYE SHENGWU TONGJI SHOUCE

古华民　编著



技出版社

# 农业生物统计手册

古华民　编著

广东科技出版社

# 农业生物统计手册

NONGYE SHENGWU TONGJI SHOUCE

古华民 编著

广东科技出版社出版
广东省新华书店发行
广东第二新华印刷厂印刷
787×1092毫米 32开本 7.375印张155,000字
1989年7月第1版 1989年7月第1次印刷
印数1— 3,500册
ISBN 7—5359—0461—0
S·50 定价 2.60元

# 前　言

本书主要是介绍农业生物科学试验中经常用到的数理统计方法。对于从事农业生物方面的实际工作者，希望本书能为他们提供常用的统计方法，处理具体工作中的有关统计分析问题，并了解有关方法的统计思想；对于农业生物方面的技术及研究人员，本书将是一本数理统计方面的入门读物，可以为他们进一步阅读更高一级的统计读物提供必要的基础知识。鉴于上述想法，本书不求数学方面的严谨性，读者一般只需具备初中以上的数学知识即可。本书从实际问题出发，着重阐明有关的统计概念、术语，特别是处理实际问题的统计思想；并以实例说明有关方法、具体计算步骤以及方法的应用范围，便于实际工作者在实践中直接应用。其中实例大部分是作者在推广、应用生物统计方法过程中收集到的。本书共分十二章，各章的内容在该章开头都有介绍，这里就不赘述了。由于作者水平有限，书中错漏之处难免，恳请读者多加批评指正。

作　者

# 目 录

# 第一章 引 言

通常在进行了一项科学试验或作了一系列观测，把有关结果记录下来之后，就有一个如何解释这些结果并导出合理结论的问题。有时候结果本身是明显的，解释起来比较容易，所作的结论也是无可争议的。比如，用20只试验用的小白鼠进行药效试验，我们先让这些小白鼠受某种病毒的感染，然后把其中半数小白鼠用某种新药给予治疗，如果这半数小白鼠都获救，而另外半数未经该药物治疗的小白鼠全部死亡，那么得出新药对治疗这种病毒感染是有效的结论是明显的。当然，像这样的结果是无需用到严格的统计分析。但是，在大多数生物科学试验中，上述那种情况是极少的。

下表是在同一栽培条件下，不同的花生品种产量的比较

表1 同一栽培条件下不同花生品种的亩产量

| 品种 \ 重复 / 产量(斤/亩) | I | II | III | IV | 总和 | 平均 |
|---|---|---|---|---|---|---|
| 韶早红 | 215 | 225 | 217 | 210 | 867 | 216.75 |
| 狮头企 | 237 | 215 | 217 | 205 | 874 | 218.5 |
| 粤油551 | 340 | 325 | 340 | 315 | 1320 | 330 |
| 汕油3号 | 267 | 257 | 250 | 242 | 1016 | 254 |
| 惠花1号 | 257 | 262 | 255 | 255 | 1029 | 257.25 |

试验结果。其中品种狮头企是对照种，表中亩产量是按小区产量折算得到的。

由表1看出，品种粤油551的平均亩产量最高，比对照种狮头企的平均亩产量218.5斤多111.5斤，人们将毫不怀疑地认为粤油551优于对照种，是一个值得大面积推广的良种。象这样有明显差异的试验结果，较易作出肯定的结论。但是，再看看其他品种的情况，我们发现要作出肯定的结论就不那么容易了。比如，韶早红的平均亩产量略低于狮头企的平均亩产量，据此作出韶早红比不上对照种狮头企的断言显然是冒失的。实际上，由四个小区产量来看，有两个小区韶早红的产量却略高于狮头企；而且由表1还看到同一品种在完全相同的栽培条件下，四个小区的产量本身也不一样，说明除品种的不同对产量有影响外，还有其他诸如地力、田间管理、病虫害侵袭等对产量也有影响，这些因素尽管在试验工作中要求做到一致，但是不可避免地存在变异，有些如病虫害侵袭是无法控制的，这些变异大多表现为无法控制的偶然因素综合作用的结果。正是由于这些变异的客观存在，给试验结果的分析带来一定的困难。对于试验或观测结果的分析，如果不把这种变异性考虑进去，那么，我们所作的结论，往往只停留在用诸如"看来……"、"似乎……"、"可能……"这种含糊不清的措辞的表达方式，而这种表达方式是不符合科学试验需要的。尽管由于各种偶然因素影响的客观存在，我们无法作出绝对肯定或绝对否定的判断，但是借助于数理统计方法，就有可能作出具有一定可靠性的合理结论。

生物科学试验处理的对象是生物，而自然界中的生物会受各种各样因素的影响，因此引起变异的原因也是多方面的。在生物试验中，不仅要考虑那些可以控制的试验条件，而且

还要考虑那些无法控制的偶然因素，这样才能通过对试验结果的分析，作出比较全面、比较合理的结论。数理统计方法则是一种能够把本质差异与偶然因素引起的变异区分开来的方法，正是这一原因，数理统计方法已经成为解释与分析乃至安排生物科学试验的不可缺少的工具。

# 第二章　变异性与频率分布

## 第一节　变量、总体与样本

在研究不同地区小麦的蛋白质含量的试验中，测定了几个不同地区小麦的蛋白质含量，得如表2的结果。

表2　不同地区小麦的蛋白质含量

| 地　区 | I | II | III | IV | V |
|---|---|---|---|---|---|
| 蛋白质含量(%) | 11.7 | 12.9 | 16.1 | 16.3 | 19.0 |

在这一试验结果中，小麦蛋白质含量的百分数是一个变量。它反映了各个地区小麦的一个数量特征，它的值可以区分不同地区小麦蛋白质含量的百分数，也是我们分析试验结果的基本数据。上述变量是一个连续变量，因为它可以取到某一范围内的数值中的任何一个值。上述变量取值范围大致为11.7%至19.0%，可以设想有另外一个地区其小麦的蛋白质含量取该范围内的某一数值，或者是取该范围上下的某一数值。因此，该变量是一个连续变量。还有诸如小麦的株高、穗长、水稻叶片的面积、水稻谷粒的重量、含水量等都是连续变量的例子。在一些生物研究中，还有另外一种类型的变量，比如，水稻的穗数、粒数，每株花生的花数、荚果

数等，它们的取值都是整数，不可能取分数值，这类变量称为离散变量。

表2中列出的是五个地区的小麦蛋白质含量，显然它仅仅是种植小麦的所有地区的小麦蛋白质含量的一小部分，如能把所有地区的小麦蛋白质含量都测定出来，那么我们就可以对小麦的蛋白质含量作出一些整体性的结论。这就引出了所谓总体的概念。上述"种植小麦的所有地区"就是一个总体。又如，研究某地区大白猪的体重，则"某地区大白猪的全体"就是一个总体。一般地，把研究对象的全体称为总体，而把总体中的每一个成员称为个体。就统计工作而言，我们的目的不是研究总体中的每个成员本身，而是研究个体的某些特性。比如，"单株水稻的全体"构成一个总体，在这里我们不是研究水稻本身，而只感兴趣于研究水稻单株的株高这一特性，这时的总体往往表现为"水稻单株株高的全体"，因此是一个数据的集合。若以变量X表示水稻单株株高，则总体可以用变量来表示，并且就记为X。为此，统计学中的变量往往指的是某一个总体。

生物研究中往往要了解的是总体的种种性质，如能对总体中的每一个成员即个体都加以研究，比如逐个测定每一个个体的某种特性，则对总体的该种特性就可以获得全面的了解。但是由于总体中的个体数往往很多，甚至是无限多个，事实上无法把所有个体都加以测定；另外，有些情况总体中个体数目并不很多，但测定其某种数量指标时具有破坏性。例如，测定小鸡服用某种药物的致死量，当致死量被测定时，小鸡已死亡。在这一情况下，我们仍然不能对所有个体一一测定。为此，一般在试验研究中，我们常在总体中抽取一部

分个体进行试验研究，然后根据这一部分个体的性质来估计、推断总体的性质。这一部分个体常称为样本。这一部分个体的数目称为样本的容量。

生物研究的目的往往是要了解总体的种种性质，但是由于各种原因，客观上又只允许我们进行一定数量的试验或观测。所以，在很多情况下，我们需要通过对样本的研究来分析、估计、推断总体的种种性质，这些就构成了生物统计方法的基本内容。

## 第二节　变异性与频率分布

引言中曾述及在生物研究的试验或观测结果中不可避免地带有偶然因素引起的变异，影响了我们对试验或观测结果的分析。为此必须首先揭示变异性的内在规律性。下面的表3是观测水稻品种中青二单株株高的数据。

表3　57个单株水稻株高（厘米）

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 108 | 99 | 93 | 105 | 103 | 101 | 100 | 102 | 101 | 98 |
| 95 | 96 | 94 | 96 | 97 | 101 | 97 | 98 | 91 | 102 |
| 98 | 93.5 | 97 | 95 | 96 | 99 | 97 | 95 | 99 | 101 |
| 95 | 97 | 102 | 97 | 92.5 | 100 | 98 | 100 | 103 | 99 |
| 100 | 99 | 99 | 100 | 97 | 96 | 98 | 100 | 99 | 101 |
| 101 | 102 | 99 | 96 | 102 | 90 | 94 | | | |

为了了解某水稻单株株高这一特性，我们测定了57个该品种单株的株高，所得数据如表3。此时该品种水稻单株株高的全体就是一个总体，这57个单株样高就是一个由该总体

获得的样本，样本容量是57。由这57个数据看出，尽管是同一品种在同一栽培条件下种植的水稻，但其株高不尽相同，呈现出一定的波动性，这就是各种无法控制的偶然因素引起的变异性。由于这种变异性给我们从表3所列数据提取有关的信息带来一定的困难。统计工作就是要把这些表面上看来是杂乱无章的数据，经过加工整理，揭示其内部所依循的规律性，并进而由样本所获得的信息去估计、推断总体的有关性质。比如，我们可以先把表3的数据按一定的方式整理成表4，即把表3的数据按由小到大的顺序加以排列得表4。

表4　表3数据按大小顺序排列

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 90 | | | | | | | |
| 91 | | | | | | | |
| 92.5 | | | | | | | |
| 93 | 93.5 | | | | | | |
| 94 | 94 | | | | | | |
| 95 | 95 | 95 | 95 | | | | |
| 96 | 96 | 96 | 96 | 96 | | | |
| 97 | 97 | 97 | 97 | 97 | 97 | 97 | |
| 98 | 98 | 98 | 98 | 98 | | | |
| 99 | 99 | 99 | 99 | 99 | 99 | 99 | 99 |
| 100 | 100 | 100 | 100 | 100 | 100 | | |
| 101 | 101 | 101 | 101 | 101 | 101 | | |
| 102 | 102 | 102 | 102 | 102 | | | |
| 103 | 103 | | | | | | |
| 105 | | | | | | | |
| 108 | | | | | | | |

把表3数据整理成表4就容易看出，水稻单株株高大致在90—108厘米之间，而且株高在95—102厘米范围内的比较

集中。另外，还可看出不同株高出现的株数，比如株高为99厘米的株数最多，达8株，其次是株高为97厘米的株数达7株，等等。表4数据还可以用图直观地表示出来，如图1。

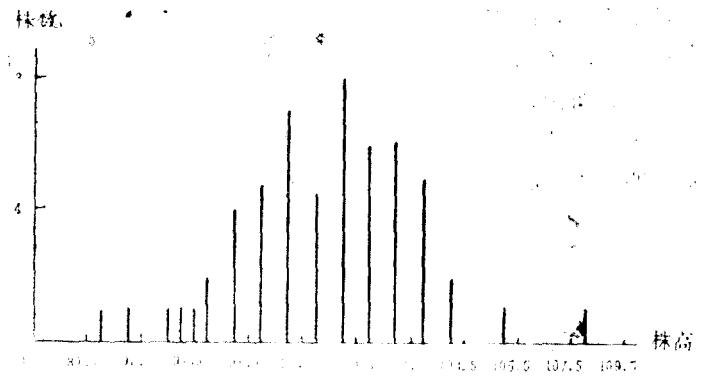



图1　表4数据之分布

图1虽能直观地表现表4数据的一些情况，但其规律性不明显，数据出现的次数起伏不定，给出的信息不多。加上水稻株高为一连续变量，为了更清楚地显示数据波动的规律性，通常采用分组的办法作出直方图。首先把株高这一变量的取值范围按1厘米的区间间隔划分为等分区间，然后计算落入各区间的株数，即所谓频数。为了便于计算频数，通常要使分点取得比原测量精度高一位，并且约定，当株高的数值恰好等于分点时，计算频数则把该株高的株数归入左边那一个区间，于是我们可得如下的频数分布表。

表 5　表 3 数据按区间长度为 1 厘米的频数分布

| 区间间隔 | 频数 |
|---|---|
| 89.5—90.5 | 1 |
| 90.5—91.5 | 1 |
| 91.5—92.5 | 1 |
| 92.5—93.5 | 2 |
| 93.5—94.5 | 2 |
| 94.5—95.5 | 4 |
| 95.5—96.5 | 5 |
| 96.5—97.5 | 7 |
| 97.5—98.5 | 5 |
| 98.5—99.5 | 8 |
| 99.5—100.5 | 6 |
| 100.5—101.5 | 6 |
| 101.5—102.5 | 5 |
| 102.5—103.5 | 2 |
| 103.5—104.5 | 0 |
| 104.5—105.5 | 1 |
| 105.5—106.5 | 0 |
| 106.5—107.5 | 0 |
| 107.5—108.5 | 1 |
| | 总和57 |

此时，实际上我们把数据分为19组，每组组距为 1 厘米。可按表 5 作直方图，见图 2 。

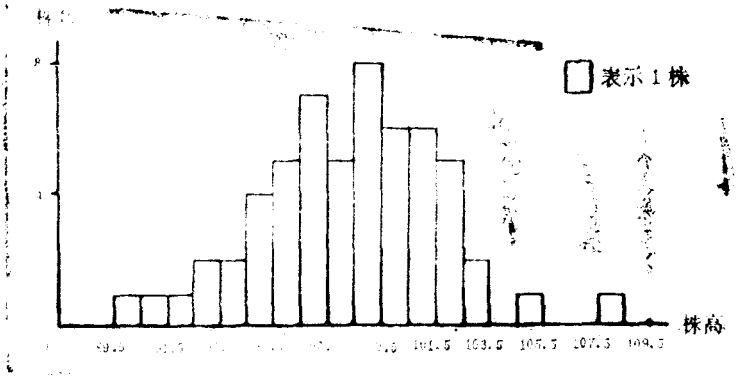



图 2　按表 5 所作直方图

（区间长度表示 1 厘米）

我们在作直方图时是以横轴表示株高，并取等分区间，区间长度表示 1 厘米作为一个单位长度；而纵轴则表示株数，单位与横轴一致，一个单位长度表示一株。例如，株高在90.5—91.5的有 1 株，则以该区间为底，高为 1 个单位长度作矩形，于是矩形面积实际上就是落入该区间的株数即频数。又如，株高在95.5—96.5之间的有 5 株，则以该区间为底，5 个单位长度为高作矩形，该矩形面积就表示株高落入该区间的频数，余类似。这样所作直方图就是频数分布直方图。

图 2 虽比图 1 更能反映数据波动的情况，但是规律还不够突出，其中仍有间隙。比如，103.5—104.5之间，105.5—107.5之间频数均为 0 。另外，还有一些组的频数比较小。为了提高各组的频数，可采用如下方法。

1.通过收集更多的数据可以增加各组的频数。当然，我们无法用本例来说明这一点。但是，可以设想若我们测定更