王鉴明编著

生物统计学

农业出版社

生物统计学

王鉴明 编著

农业出版社出版（北京朝阳区麦营路）

新华书店北京发行所发行　　农业出版社印刷厂印刷

787×1092 毫米 16 开本　22.25 印张　527 千字

1988 年 3 月第 1 版　1988 年 3 月北京第 1 次印刷

印数 1—3,670册　　定价 4.85 元

ISBN 7-109-00599-2/Q·19

前　言

生物统计在我国生物学和农学上的应用远不如医学和工业。为了满足广大读者的需要，作者将几十年从事生物统计研究、教学中的讲义、科研资料整理和补充成册。本书曾内部印过讲义，在整理时力求不浪费读者的时间和精力去探讨数理统计的严谨性和追求统计公式的来源，而是详尽地交代公式的应用范围和应用条件的限制，务求使读者对公式的灵活运用。书中赞扬简单试验设计和统计分析的巧妙运用，而不主张读者去追求复杂的学院式的应用。

这本书的篇幅，大部用来对生物统计的基本概念和基本方法的详细介绍。应用例子力求多专业性。重大章节相应地加以扩大（如抽样技术、变量分析等）。较新章节破格地加以引入（如生物学测验分析等）。自由度和变量的均衡性和可加性、效应的均衡比较、统计代换的必要性、试验单位或试验小区、区组或区团、局部控制、抽样技术、非参数统计分析、差异显著性测验、变量分析法应用的基本要求等，都加以详细论述和解释。

作者利用在美国留学时期参加了国际生物统计学学会当会员的方便，搜集到大量生物统计学应用在生物学（数量遗传学、数量生态学、数量分类学、生物资源学等）、农学（动植物育种学、植物病理学、昆虫毒理学、经济昆虫学、森林学、种子学、农艺学、作物栽培学、水产学、园艺学、畜牧兽医学、水利学、热带作物学、肥料学、农业经济学等）和地学（土壤学、地质学、气象学、水文学等）各方面的例子选编入本书内，以利读者对公式的灵活运用，有所启发。

这本书由于篇幅和内容的扩大，适用于作为参考书来使用；但在特殊专业培训班的情况下，也可以对书内一些章节，有选择性地作为教材使用，也无不可。这本书虽然引用数学不算多，也不算深；但也要求读者有一定的数理基础。因此读者最好能重温或预习概率论和一般生物统计的读物。高等院校师生和科技人员使用更觉得合适。

亡妻杨瑾英同志生前对我生物统计的教学、科研工作上给以很多的鼓舞、帮助和支持，每念及此不知如何报答她才好，谨在这本书写成而行将出版的时候，为表示我对她的敬仰和感谢，我要说明我是从她那里得到动力来更好完成这本书的写作的。

王鉴明

1985年3月

目　录

编者注：书中计量单位因涉及大量计算问题，未按法定计量单位更改，特此说明。

第一章 绪 论

1.1 生物统计学的发展史及其地位与作用

生物统计学是研究生物群体内个体间的变异性和对生物性状观察过程中的误差进行研究，所以假设世界上不存在变异性和误差，则无所谓统计学，更谈不上生物统计了。生物统计是一门科学，它是统计学的一分支，也是数量生物学的一分支。

统计学一名辞，虽然各国称谓不同，但字根含义则一，德文 Statistik，法文 Statistique，英文 Stutistics，西班牙文Estadistica，捷克文Statistika，葡萄牙文Estatistica，匈牙利文 Statistika，意大利文 Statistica，比利时文Statistiqul，挪威文 Statistikk，波兰文 Statystyka，瑞典文 Statistikln，荷兰文 Statisticln，希腊文Statistiqul，俄文 Статистико，皆源于拉丁文Status。这些都表示政治区域或国家人口而言，阿里士多德第一个提出逻辑推论，Leiluiq是第一个建议实行数学来进行逻辑推论，1713年Jacob Bernoulli才是第一个实现 Leiluiq 的建议。马克思在《资本论》中提到威廉比蒂(1623—1687年) 是经济学之父，在某种程度上是统计学的发明者。十八世纪 (1763年) Thomas Bogeo第一个利用机率定理来归纳理由。1820年法国人Laplace及同时代的Gauss发明常态分布，比利时 Queteler (1796—1874年) 对数理统计贡献很大，十八世纪苏联方面有茹拉夫斯基以及马尔可夫均对统计学上做了重大贡献。

英国人哥尔顿 (Galton 1822—1911) 在十九世纪末叶，应用统计方法研究人种特性与遗传，创生物统计学。卡·皮尔逊(K. Plarson 1857—1936)在1906年继续主持哥尔顿实验室，并在 1895 年与哥尔顿共同努力下成立伦敦大学生物统计实验室，1901 年创办 Biometrika权威杂志。生物统计第一篇论文是 1889 年哥氏写的《自然界的遗传》，1877年哥氏已有相关概念，引用其1875年香豌豆的试验资料，皮尔逊创卡方测验，“学生氏”(W. S. Gosset 1876—1937)创小样本学说的 t 值测验法，并与比万氏共创田间试验的棋盘排列法。费歇氏(Fisher, R. A.)继皮尔逊主持哥氏实验室，一生论著甚多，尤以应用生物统计为然，创变量分析法，y测验，判别函数，因z式试验设计，混杂试验设计等。杰出的生物统计学家甚多，印人 Mabeilinrobis P.C. 主要贡献于作物抽象调查，Waecl, A. 于序贯抽象，Finney. 于毒理统计，Mather K. 于生物统计遗传学，Yates F. 于田间试验。

国内生物统计过去也相当早引起注意，应用较为广泛。

生物统计学近年来发展甚速，从中又分支为生物统计遗传学 (群体遗传学)、生态统计学、分类统计学、毒理统计学等等，生物统计学的一些概念如相互作用以及一些方法如t测验，也有所推广。

近年来发展了数学在生物学和农学的应用，已成了一门生物数学的新学科，生物统计

学只是它的一门分科的科学，生物学和农学随着电子计算学和多元分析的发展，数学模拟对生物学和农学的应用逐渐发展起来。

1.2 生物统计的基本符号运算及概念

1.生物统计的基本符号运算及符号的定义

1. $$\sum_{i=1}^{N} X_i \equiv X_1 + X_2 + \cdots + X_N$$

2. $$\sum_{i=1}^{N} (X_i \pm y_i \pm y_i) = \sum_{i=1}^{N} X_i \pm \sum_{i=1}^{N} y_i \pm \sum_{i=1}^{N} y_i$$

3. $$\sum_{i=1}^{N} CX_i = C \sum_{i=1}^{N} X_i$$

4. $$\sum_{i=1}^{N} C = NC$$

5. $$\sum_{i=1}^{N} \sum_{j=1}^{K} X_{ij} \equiv (X_{11} + X_{12} + \cdots + X_{1k}) + (X_{21} + X_{22} + \cdots + X_{2k}) + \cdots + (X_{N1} + X_{N2} + \cdots + X_{Nk})$$

6. $$\sum_{i=1}^{N} \sum_{j=1}^{K} X_{ij} = \sum_{j=1}^{K} \sum_{i=1}^{N} X_{ij}$$

2.生物统计中名词的基本概念

（1）变数（或性状，variable） 一个可以给它以无限多的数值的量数叫变数。如株高、茎径、枯心数、螟害节数、10株蔗头内蔗龟数，30尺行蔗株的黄螟卵数，1月份平均最低温度，7月份最大风速，1两大豆中豆象为害粒数，小区内杂草数，4月份赤眼卵蜂自然寄生率等均可以看成变数。

（2）变员数（相当于在某一特定条件下的性状，variate） 每次所给予的数值是为该变数的变员数。如同一设计的试验进行若干年，每年的试验便是一个变员数。在某种栽培条件下台糖134甘蔗品种的株高也可以看作为一个变员数。

（3）变异(variation) 变员数间彼此差异的现象称做变异。

（4）变异性(variability) 一个量数能给予以不同数值的，它是叫做具有变异性的。

（5）常数(constant) 一个量数不能给予不同数值的，它是叫做没有变异性的。这个量数就叫做常数。

（6）连续变数 变员数按大小排列，任何两个相邻的变员数之间，不管它们差异多么少，我们总可以再指定另一个变员数在其间。连续变数又名可量性状。作物株高201cm与202cm之间，存在有201.1和201.2，在201.1与201.2cm之间又存在有201.11与201.12cm，其间仍存在201.111与201.112cm，如类推。

（7）非连续变数（可数性状） 变员数按大小排列，任何两个相差只一个整数个位的相邻变员数间，再不能再指定另一个变员数插入其间，如1与2之间，绝对不可能有1.5的存在，因个体不可能是非整数的。可数性状一般又分为正二项式分布和潘松分布两种，前者在样本内的亚样本大小一定，只含有一定数的变员数，出现某种特殊性状的变员数的机率并不太少，而不出现的机率则并不太大，这种资料称为正二项式分布资料。如100个花粉的发育花粉数就是。后者在样品内的亚样本可以含有无限多个变员数，出现某种特殊性状的变员数的机率非常少，而不出现的机率则非常大，这种资料称为潘松分布资料。

（8）集团与样本 含有全体变员数的群体叫集团；只含有部分变员数的群体叫样本；从集团中抽取部分变员数来组成一个样本叫抽样。抽样技术是根据集团的性质，抽样可靠性的要求和抽样所能付出的代价（如劳力和费用等）而决定的。

序贯抽样：序贯抽样在试验或调查过程中，样本的观察次数决定于上一次或上一期或上一阶段的试验和调查的结果，由于节省了观察次数所以省了物力和人力，同时在计算上亦较简单，结果亦较准确。

双重抽样：双重抽样法特别是应用于调查某一种不易观测或花费颇大才能观测的性状，我们在较小的样本里调查这个性状和与另一个看来是以这一性状有很密切相关的一个简单的性状，然后凭小样本里边的资料求出这两个性状的相关后从而进一步采较大的样本单纯调查那个较简单的性状，便可以估计大样本的不易观测或花费颇大才能观测的那个性状。比如估计生长期中的甘蔗茎重，甘蔗茎体积为简单性状，因为茎体积可以从茎长、茎径计算出来，而茎长、茎径是简单性状，不需破坏性调查。甘蔗茎重量则为复杂性状，因多斩了会增加调查费用和蔗农意见很大。

巢式抽样：从集团里抽样得样本，再从样本里抽样得亚样本，如是类推，可以继续下去抽样，这个抽样方法叫做巢式或簇式抽样。这个抽样方法特点是选择一个适当大小的样本，亚样本以至再低的抽样单位的组合，保持一定准确度；而以最低的经费来完成工作。

阶层抽样：当集团某部分变员数与另一部分变员数显然有异质性时，那么集团之内分有阶层，宜从每个阶层内分别作随机抽样或顺次抽样，这里要注意的是阶层与阶层间的比重。

随机抽样：集团的部分变员数来组成样本是随机率来决定的，适用于当我们实行了局部控制，设尽办法减少试验误差到最大的限度，及因应集团的异质性而分了阶层等等措施之后的那些情况，变量分析的理论是绝对要求变员数的彼此独立和随机，故不采用随机抽样而采用变量分析法是犯了原则性的错误的。

顺次抽样：集团中部分变员数组成样本不是按机率决定，而是按一定的规则顺次抽样，这个办法容易理解，容易管理，容易观察示范，但当材料是有周期性或规律性时，慎不要采用，免产生过大的系统误差。

典型抽样：典型抽样又称主观抽样，即在集团中看中了一些看来能够代表集团的变员数来构成样本，这个方法带有主观成分，但当我们很熟识了集团或集团的全体变员数时来运用它，问题较简便、省人力、物力，但反过来说，容易引起主观因素带来的误差。

（9）样本统计常数与集团统计常数 数学常数为绝对常数，如 $e=2.71828$ $\pi=3.1416$，统计常数乃相对常数，在同一个样本或集团内，统计常数是一定不变的，但不同样本或集团，自有其彼此相异的统计常数；用以描写集团而从集团资料计算出来的称集团

统计常数；用以描写样本而从样本资料计算出来的称样本统计常数。统计常数也可以看成变员数，所以就有由统计常数构成的集团或样本。

(10) 统计群体三种　变员数为不同个体的观测数的群体，变员数为同一个体不同观测数(由于有观测误差的存在)构成的群体(因观测一定有误差)，变员数为统计常数的群体。

(11) 误差　广义的误差包括错误、恒误差和试验误差三种，错误可以避免，但一旦不小心而产生则误差较大，且较难更正。恒误差是由于调查人员观测能力差异或仪表失误所致，误差是恒定的，一旦发现，则易改正或改算。试验误差是人为不能控制不可避免的，来源甚复杂，原因不很明白的，只可以减少不能消灭。生物统计所讨论的误差，主要是这一类的误差，田间的试验误差较大，室内误差较小；物理化学试验误差较小，生物试验的试验误差较大。我们可以从改善技术和采用重复来减少误差，误差大小与重复数目的平方根成反比，重复还有估计误差进而测验差异显著的作用。重复是一个很重要的手段。

(12) 科学方法与统计方法　科学方法包括科学假设，进行试验调查来搜集资料、整理资料、统计分析资料、统计结论，最后科学结论，所以统计方法是包括在科学方法范围内，统计假设为无效假设，只可反证其伪，不能正面证其为真。如测验两个品种产量高低，无效假设应为两个品种产量无差异，而不能为两个品种产量有差异，因在具体试验中A品种比B品种产量高于超过试验误差范围则可认为假想为伪，但纵使A与B品种同一产量也不可能证明假设为真，因A与B产量如真正是有异差仍可以受误差影响使其偶然恰巧做成相同的产量。达到统计结论并未完结科学方法，所以把统计结论当作最后的结论是错误的，统计方法是科学的，但统计方法不能认为代表了全面的科学方法。

1.3　农艺、土化和昆虫生物统计的特点

1.农艺生物统计的特点

(1) 资料绝大多数是可量的，可数的占绝少数，所以大多数资料采用可量资料分析法，资料亦很少需要代换为其他单位。

(2) 一般来说，试验多为田间的，室内较少，所以误差较大，而且较难控制，但由于农艺选种和耕作处理效应相对来说仍算是比较大的，所以一般也不太要求过分灵敏的测验。

(3) 误差一般是试验误差较重要于抽样误差。

(4) 试验单位一般为一块土地面积的概念。

(5) 农艺生物统计重视相关研究，在下列一些问题上：间接鉴定、丰凶预测、性状相关。

(6) 抽样研究重视调查标准决定、作物产量估计、生理生化酶及激素的研究。

2.土化生物统计的特点

(1) 室内和田间都一样多，室内又分化验室、细菌室。此外盆栽试验又介乎田间与室内两者之间的，室内试验多宜注意于抽样误差，田间试验则应多注意试验误差。

(2) 抽样研究重视土壤微生物、需肥诊断、施肥等。

(3) 相关研究重视土地肥力变化规律与轮作施肥关系。

(4) 土壤微生物是可数资料，其他多数是可量资料。

（5）室内试验中，土壤微生物试验误差较大，理化室内试验误差较小。

3.昆虫生物统计的特点

（1）资料绝大部分是可数的居多数，可量的占少数，所以大多数资料采用可数资料分析方法。

（2）在某种情况下，可数资料须进行统计代换后方可分析。

（3）昆虫生物统计学重视研究昆虫统计分布的特点，以便采用对某种害虫相应采取的抽样方法。

（4）昆虫生物统计学重视相关研究来解决害虫发生规律和预测预报的问题，同时要重视时间数例的研究。

（5）昆虫生物统计的试验单位一般与农作物的概念不同，不是一块土地的概念，可能是半叶、一叶、一枝、一株、十株、一行、一段的概念。

1.4 农艺、土化和昆虫生物统计的研究特殊领域

1.农艺生物统计研究的特殊领域

生统遗传学（群体遗传学、数量遗传学）、丰凶预测、间接鉴定、田间试验、作物估产等等。

2.土化生物统计研究的特殊领域

田间肥料试验：重视多因子设计及混杂设计来求因子相互作用及克服土壤差异。

土壤抽样技术：巢式抽样法的应用。

称重问题：微量物体可以利用一定的称重设计来抵消各个物体称重误差和一次称多个物体，提高称重效率，例如我们要称a、b、c、d、e、f、g七个物体，有两种设计：

第一种设计：各物体放在右盘，左盘放法码：

$$
\begin{aligned}
a+b+c+d+e+f+g &= W_1\\
a\quad\ \ +d\quad\ \ +g &= W_2\\
\quad b+c\quad\ \ +g &= W_3\\
a+b\quad\ \ +f\quad\ \ &= W_4\\
\quad b\quad +d+e\quad\ \ &= W_5\\
a\quad +c\quad +e\quad\ \ &= W_6\\
\quad\ \ e+f+g &= W_7\\
\quad c+d\quad +f\quad\ \ &= W_8
\end{aligned}
$$

$$a=\frac{W_1+W_2+W_4+W_6-W_3-W_5-W_7-W_8}{4}$$

$$g=\frac{W_1+W_2+W_3+W_7-W_4-W_5-W_6-W_8}{4}$$

第二种设计：各个“+”号物体放右盘，“-”号物体放左盘。

$$a+b+c+d+e+f+g=W_1'$$

$$
\begin{aligned}
a - b - c + d - e - f + g &= w_2' \\
-a + b + c - d - e - f - g &= w_3' \\
a + b - c - d - e + f - g &= w_4' \\
-a + b - c + d + e - f - g &= w_5' \\
a - b + c - d + e - f - g &= w_6' \\
-a - b - c - d + e + f + g &= w_7' \\
-a - b + c + d - e + f - g &= w_8'
\end{aligned}
$$

$$
a = \frac{w_1' + w_2' + w_4' + w_6' - w_3' - w_5' - w_7' - w_8'}{8}
$$

……

$$
g = \frac{w_1' + w_2' + w_3' + w_7' - w_4' - w_5' - w_6' - w_8'}{8}
$$

3．昆虫生物统计的特殊领域

毒理统计学：应用生物测定(bio—assay)的统计方法。

害虫预测预报：相关回归及时间数列的应用。

害虫群体密度和分布研究：正二项式分布，潘松分布，负二项式分布，核心分布的卡方测验。

害虫防治的田间试验：注意克服喷药试验误差和害虫核心分布的抽样误差。

害虫的损失估计和调查：抽样技术研究。

第二章　试验资料的整理

2.1　试验资料整理的目的和整理过程中的损失

试验原始资料在未整理前是一堆杂乱无章的数字，不易于进行统计分析，更谈不上从其外表看出规律来。试验资料未整理之前，具有最大的信息，保留了资料的原始性和完整性。任何一种资料的整理均对资料的原始性和完整性有所影响，其信息也有不同程度的损失。试验资料整理包含了这么一些意义，那就是：

（1）审核原始资料的可靠性，把不可靠的除去，宁愿另加以估计（缺区产量和混合区产量估计，容后详述）。

（2）原始数据小数位的取舍有助于既便于计算而又维持一定程度的可靠性。

（3）采用简化运算，虽在一定程度上有损于原始资料的原来信息，而对于计算得到大大简化。最后要求没有改变试验结论的前提下，简化运算是有利于生物统计的推广普及的。

（4）整理成次数分配表形式，使人初步得到对试验资料的一些概念般的印象，但对进一步计算其他一切统计常数却有莫大的帮助。制次数分配表过程中一定会引起对试验资料的原来信息的一种损失，但这种损失并不是一定不可避免的，这要看具体情况。如试验资料不多，而我们只希望从中计算出个把统计常数，而我们又有许多计算的方便时，我们应考虑是否一定要制一个次数分配表的必要。在相反情况下，次数分配表的制出是一定必要的，问题是如何制法，才能做到对试验资料的原来面目得到最小的损失，这是关系到决定次数分配表组数的多少或组距的大小问题，组数多或组距小则对试验资料较大程度地保留它的原来面目，减少损失，但计算上未得到较多的方便。反之，组数少或组距大则对试验资料保留它的原来面目较少，牺牲较大，但计算上可得到较大方便。如何去决定组数或组距问题，是制次数分配表的中心问题。

（5）可数资料在某种情况下是要进行统计代换将原来试验资料的单位或数量改变后才得到正确可靠的统计分析结果，这也是试验资料整理的范畴。统计代换是必要的而不是繁琐的试验资料整理方法。

（6）试验资料有时为了易于表达，特别满足一些展览会的要求，我们有时也会利用试验资料来绘制统计图，也是整理资料工作的一部分。

简化运算是会损失到原来资料的信息，对原始资料舍去一些位数或经过制次数分配表归组的过程，都会带来损失，问题是如何减少损失或纵使损失信息但仍保留正确的结论就是了。

2.2　小数位的取舍

我们观察客观世界事物，一般总是通过衡量和数数把读数记录下来。衡量一个东西要

有一个标尺，然后看看我们要衡量的那件东西是这个标尺的几个单位。原来那件东西有它一定的长度、宽度、重量、大小等一定的数值，这叫做真确数；但我们衡量出来的结果所得的数值只是一个近似数。真确数与近似数之间总存在一定的差异。一般这个差异不致于大于半个标尺单位。因为大半个标尺和小半个标尺总易判别出来的，不会误认为大半个标尺为小半个标尺，也不会误认为小半个标尺为大半个标尺。大半个标尺我们当作一个标准，进一位，误差不会大于半个标尺；同样，小半个标尺我们舍去不要，误差也是小于半个标尺。所以真确数不管是大于抑或是小于近似数，但无论如何它们的差异的绝对值总小于半个标尺，那个东西有多少个标尺的长度、宽度、重量或容量等，我们叫做有效数字有多少个。一般统计学术语这样说：如果近似数的绝对误差界是末位上的半个单位，那么在这个近似数里从第一个不是零的数字起到这个数位止，所有的数字都叫有效数字。所以 3.1 近似数有两个有效数字，而3.100近似数则有四个有效数字，前者的标尺为0.1，而后者的标尺则为0.001。所以310,000,000说成二个有效数字时，写法是不恰当的，应该写成31×10^7或3.1×10^8，如果照310,000,000的写法，那我们应该说它是九个有效数字的。

到底我们要用什么标尺来衡量一件东西才适合呢？这要看我们的要求，标尺过小，衡量时工作量大，资料有效数字多，计算不方便。所以星际距离不用尺，不用里，而用光走一年的时间（光年）来表示距离。我们量细菌的长度和直径不用尺，不用寸，而用微毫尺作单位，因而用大标尺量小东西，量不出来，量不准确。一般作物行距标尺用尺，株距标尺用寸，株高标尺用厘米，既不用米，又不用毫米，就是这个道理。

如果我们没有别种尺，而只有这种尺，标尺小了，原始资料有效数字大了，为了便于计算，我们可以保留较前面几个有效数字。

近似数是具有一定有效数字的，它们之间的加减乘除的运算，主要有下面两个法则：

法则 1　在近似数相加或相减时，小数位数较多的近似数只要比小数位最少的那个加数多留一个位，其余都把它们舍去，在计算结果里应保留的小数位数和原来近似数里小数位最少的那个位数相同。

〔例〕

不正确运算	正确运算
3.145	3.14
2.27	2.27
3.5	3.5
+ 2.8	+ 2.8
11.715	11.71→11.7

法则 2　在两个近似数相乘或相除时，有效数字多的近似数只要比有效数字较少的那个数多保留一个，其余的都把它舍去，在计算结果里从第一个不是零的数字起应保留的数字的个数，和原来近似数里有效数字数少的那个位数相同。

〔例〕

不正确运算	正确运算
3.8654	3.865
× 2.96	× 2.96
231924	23190
347886	34785
+ 77308	+ 7730
11.441584	11.44040→11.44

如果标尺小了，原始资料有效数太大，不便运算，或在近似数运算过程中要决定取舍位数时，应运用四舍五入法，比五大的进一位，比五小的舍去之，五之前遇奇数则进一位，遇偶数则舍去之。13.5及14.5本来有三个有效数字，如只取两个有效数字时，它们都是14。

2.3 捷法运算

统计分析过程中有时要利用到捷法运算来简化计算手续，但注意下列事项：

1.凡在简化运算过程中以某一数加、减、乘或除全数资料数字的任一个则必须以同一数加、减、乘或除其余各数。

2.平均数受到各种简化运算影响（即加、减、乘、除对之均有影响）故最后结果应反简化运算来改正之。同时要注意在反简化运算过程中要把正简化运算过程逆转来运算。

3.标准差只受到乘、除简化运算影响，而不受加、减简化运算所影响，故进行反简化运算时，只须进行乘、除反简化运算。

4.简化运算不影响计算准确性，无损于原始资料的原来面目，但必须进行及简化运算才得最后统计结论。

〔例〕 求0.0238，0.0238，0.0241，0.0241，0.0250，0.0247，0.0241，0.0238，0.0226，0.0232 的平均数和标准差。

我们可以把原来数字(x)改为10000x—200，即以10000乘之再减去200，得：

38，38，41，41，50，47，41，38，26，32。（如果其中有一数为0.0176则简化后为24）

$$\text{假平均数}=\bar{x}^1=\frac{\sum_{i=1}^{10}(10000x_i-200)}{10}=\frac{\sum_{i=1}^{10}x_i^1}{10}=\frac{38+38+\cdots+26+32}{10}$$

$$=39.2$$

$$\text{真平均数}=\bar{x}=\frac{\bar{x}^1+200}{10000}=\frac{39.2+200}{10000}=0.02392\rightarrow 0.0239$$

$$\text{假标准差}=s^1=\sqrt{\frac{\sum_{i=1}^{10}(10000x_i-200)^2-\frac{\left[\sum_{i=1}^{10}(10000x-200)\right]^2}{10}}{10-1}}$$

$$=\sqrt{\frac{38^2+\cdots+32^2-\frac{(392)^2}{10}}{10-1}}$$

$$=\sqrt{\frac{15784-15366.4}{9}}$$

$$=\sqrt{46.4}=6.81$$

$$\text{真标准差}=s=s^1\times\frac{1}{10000}=6.81\times\frac{1}{10000}=0.000681$$

2.4 改变资料为其他单位或量数或等级或符号

1.统计代换

目的在改变原始资料使到处理效果为可加性与消除均数与变异的相关性，统计分析趋于严密。

（1）正二项式资料的统计代换　一般可数资料如害虫死亡率，花粉发育率等进行变量分析时应先进行角度代换或反正弦代换才能达到正确的统计分析。

$$\phi=\sin^{-1}\sqrt{p}$$

ϕ代表经代换后的角度，p代表代换前资料的百分数。

$$0<p<100\%\qquad 0<\phi<90^\circ$$

（2）潘松分布资料的统计代换：可数资料如每小区金针虫数，每晚捕蛾灯下捕蛾数等。

i.均数与变异无明显关系：如每小区金针虫数。

$$x^1=\sqrt{x}\ \text{或}\ \sqrt{x+\frac{1}{2}}\ \text{或}\ \sqrt{x+1}$$

当x（小区金针虫数）>10，$x^1=\sqrt{x}$

$$x<10,\ x^1=\sqrt{x+\frac{1}{2}}\ \text{或}\ \sqrt{x+1}$$

ii.均数与变异成比例：如每晚捕蛾灯捕蛾数。

田区别	氟矽酸铝(A)	氟矽酸钠(B)	A－B	等级
1	28(区内螟数)	22	6	2.5
2	29	19	10	6
3	36	23	13	9
4	45	34	11	7
5	26	45	－19	－11.5
6	57	63	－6	－2.5
7	49	55	－6	－2.5
8	35	43	－8	－5
9	35	82	－47	－14
10	38	45	13	9
11	50	50	0	—
12	13	32	－19	－11.5
13	46	59	－13	－9
14	39	72	－33	－13
15	68	74	－6	－2.5
总和				＋33.5　－71.5

$x^1 = \log x$　　当全体x不为0

$x^1 = \log(x+1)$　　当其中有些x为0

统计代换还有其他的，容后再述。

2.将资料改变为等级来运算

（1）有局部控制的二样本问题（成对的两个处理比较）

〔例1〕 甘蔗条螟防治法比较试验

n（比较对数）=15，r_0（绝对值较小的等级和）=33.5，

查威柯崇氏成对等级资料表，得 $r_{.05}=25$，$r_0 > r_{.05}$，$P>0.05$。结论：两种药剂防螟效果无显著差异。

威柯崇氏成对等级资料表

重复次数 (n)	r	P 机率
7	0	0.016
7	2	0.047
8	0	0.0078
8	2	0.024
8	4	0.055
9	2	0.0092
9	3	0.019
9	6	0.054
10	3	0.0098
10	5	0.019
10	8	0.049
11	5	0.0093
11	7	0.018
11	11	0.053
12	7	0.0093
12	10	0.021
12	14	0.054
13	10	0.0105
13	13	0.021
13	17	0.050
14	13	0.0107
14	16	0.021
14	21	0.054
15	16	0.0103
15	19	0.019
15	25	0.054
16	19	0.0094
16	23	0.020
16	29	0.053

〔例2〕 东方果蝇以含P^{32}0.34微毫居里/毫升的胡萝卜培养液喂饲后各时期在三龄幼虫及蛹的电离辐射（每分钟脉冲）调查如下资料，试以变成等级资料来分析其差异的显著性。