

生 物 統 計

(害虫发生预测常用统计方法)

(初 稿)

广 东 农 林 学 院

植保系昆虫学教研组编

一九七六年十月

生 物 鑑 計

(普通生物學與植物學鑑計法)

(第二版)

廣 東 省 科 學 院

植物學系植物學教研室

1981年10月

害虫发生预测常用统计方法

目 录

第一章 导 言

1—1	数量分析的必要	1
1—2	数量分析的对象——数据	2
1—3	调查和试验	2
1—4	总体和样本	3
1—5	什么是统计方法	4
1—6	正确对待统计方法	5

第二章 害虫数据资料的特点和初步整理方法

2—1	害虫数据资料的特点	6
2—2	分组与频数分布表的编制	7
2—3	制作频数分布表的意义	13

第三章 归纳指标

3—1	概 况	16
3—2	表示集中程度的归纳指标	16
3—3	算术平均数	16
3—4	加权平均法	20
3—5	算术平均数的特性和代表性	25
3—6	平均集合度	28
3—7	众 数	30
3—8	表示变异程度的归纳指标	33
3—9	标准差	34
3—10	标准差的特性和功用	37
3—11	相对变异常数——变异系数	39

第四章 百分比和倍数

4—1	百分比概况	43
-----	-------	----

4-2	百分比不可误用	44
4-3	表述害虫种群状况的百分比指标	46
4-4	表示被害程度的百分比指标	48
4-5	虫害损失的表述	52
4-6	表示防治效果的百分比指标	54
4-7	要有分析地对待百分比指标	56
4-8	倍数表述法	58

第五章 统计表与统计图示法

5-1	统计表	59
5-2	图示法概况	65
5-3	图示法的两个基本类型	65
5-4	诱蛾灯资料多边形图纵轴尺度的特殊处理	67
5-5	简单的移动平均法	73
5-6	果蝇频数分布	76
5-7	水稻螟虫诱蛾资料的特征及发蛾最盛日的综合判定	79
5-8	制图注意事项	81

第六章 害虫发生预测式的推求

6-1	相关与回归的基本概念及其应用	83
6-2	直线相关与直线回归	86
6-3	直线相关与直线回归是样应用于害虫预测	95
6-4	关于害虫发生预测式的评论	108

第七章 害虫调查统计法

7-1	概 说	114
7-2	取样和取样技术	115
7-3	调查取样步骤	116
7-4	害虫调查 单位和取样单位的类型	117
7-5	害虫的分布型与取样单位的布置	123
7-6	取样单位数的决定	135
7-7	简捷估英技术简介	148

第一章 导言

1-1 数量分析的必要

伟大领袖毛主席教导我们：“对情况和问题一定要注意到它们的数量方面，要有基本的数量的分析。任何质量都表现为一定的数量，没有数量也就没有质量。我们有许多同志至今不懂得注意事物的数量方面，不懂得注意基本的统计、主要的百分比，不懂得注意决定事物质量的数量界限。一切都是胸中无‘数’，结果就不能不犯错误”。（《党委会的工作方法》）

这段语录教导我们要注意进行数量的分析，以便通过数量抓住质量，把质量表现为数量是加深对客观事物的认识的重要方法。不进行量的关系的研究，就不能得到事物规律性的知识。现代科学发展的趋势是日益要求“定量”，要求掌握决定事物质量（特征）的数量界限，例如，虫害程度要指出害虫密度和被害率，不能只说轻、中、重；杀虫药剂的效力常常要表现为害虫的死亡率及密度、被害率和作物的损失^的降低，而不能只笼统地说药效高或低。虫害发生的预测预报同样要求了解害虫种群数量在时间上和空间上的动态，并进行分析。尽管测报内容分为发生期、发生量、为害程度和为害面积等项，但中心是一个数量问题。以发生期为例，所谓始见、盛发、高峰、或末、终见等无一不是由一定的发生量的界限值决定的，离开了发生量，就谈不上发生期，其他几个方面的预测，也都与害虫发生数量有关，这是很容易理解的。因此，我们可以说，害虫预测的中心课题，是数量分析的问题。

我们这个专题的任务是明确的，那就是学习如何进行昆虫研究中基本的数量分析。要进行数量分析，就要应用一些数学方法，主要是应用统计方法，统计方法应用于昆虫学研究的许多部门，而以昆虫生理——生态学和毒理学——化学保护学应用最多。具体情况，有如下主要项目：

1. 调查观察取样技术。

2. 害虫发生预测技术，重点是害虫时间数列动态分析及发生预测式的推求。

3. 室内和田间试验（特别是杀虫药剂药效试验和生态试验）的设计和分折。

4. 量理统计，即生物测定的设计和分析。

5. 试验结果与调查结果理论模式的拟合。

我们这个专题的侧重点放在调查取样及测报技术上，如以下数章所裁的，关于试验方法，在心保课程中已有讨论，我们尽量减少重复，必要时作一些补充。

1-2 数量分析的对象——数据

数量分析的对象，即要处理的材料，是数字资料，叫做数据（Data）。生物学的原始数据，有一个最重要的属性，就是它属于变数，这是由于生物现象普遍地存在着变异所决定的。生物的性状中，固然有些是固定不变的，例如，昆虫成虫足的数目，一定类群昆虫的跗节数或翅数等，但更普遍的是变异，即使同一个性状，从不同的个体，可以得到不一定相同的观测值。这类存在变异的性状就是变数。例如，昆虫的体重，密度，历期等等，每一个观测值，叫变值，如以 X 表示变数， $X_1, X_2, \dots, X_n$ 每个都表示一定的变值，这样，变数（ X ）是总称，变值（ $X_1, X_2, \dots$ ）是具体的单个变数值，是变数的体现者。

关于变数的特征及其分类等，将于下章讨论。我们先谈数据从何而来。大家通过实践都知道，数据是通过调查观测而来的，在这里，要介绍几个基本概念。

1-3 调查和试验

现代的科学研究所，大别为调查研究 and 试验研究（也叫实验）

两个方面。两者有区别也有联系。两者都是为了搜集某些问题的数据而设置的。调查是了解已经存在着的客观事实的方法。在调查的设计中，常常有意识地抽取典型，划分类型或设置对立面，但对研究对象是不加任何人为控制的，只是如实反映情况，这是调查的特点。试验也是了解客观实际或证明、揭示规律的手段，但必须加入人为控制，首先是控制除试验因子以外的一切条件，尽量做到均衡一致，从而使试验因子的效应能突出地显示出来，便于更深刻地了解和比较。试验可以人为地、暂时地使普遍联系的诸种现象中的某一个或几个现象孤立起来，以便深入分析研究。试验可以把当时当地自然界本来不存在的因素加进去，以便从正面或反面证明某个事实或规律等，一句话，就是人为地驾取材料，迫使规律性的东西透露出来。所有这些，都是试验的特点。

调查需要设计，主要是决定如何取样；试验亦需要设计，主要是决定试验方案，两者都要运用统计的理论和方法，才能合理，才能有效地控制（减少）误差，从而获得尽可能丰富的、确切反映实际的和便于分析的资料。这些资料就是可靠的数据。

调查与试验的联系，表现于两个方面，其一，在设计一项试验之前，必须进行有关问题的历史与现状的调查，将所得的资料作为试验设计的一部分依据；其二，在试验过程中，又常常需要在试验区内进行多种调查，搜集说明试验结果（效应）的资料。由此可见，调查和试验在科学研究中是交织着进行的，很难说哪个更重要。

1-4 总体和样本

在进行调查、搜集数据时，我们要了解总体和样本的概念。我们研究的对象的全部成员称为总体。在生物学上理解，在一定条件下，某一性状的全部观测值称为总体，亦即一个变数的全部变值。总体可分为有限的和无限的，我们经常研究的是有限总体。对总体加以全面考察，称为全面调查；从总体中抽取一部分成员，构成样本，加以考察，称为取样调查。例如我们要调查一块

稻田的三化螟卵块含卵粒数，全田所有卵块是总体，每个卵块都观察，就是全面调查，如果只抽取部分卵块，比方说抽50块或100块来观察，就是取样调查。这50或100个卵块就构成样本。从总体抽取样本的过程叫取样，抽取样本的方式方法，叫取样技术。取样调查的目的，完全是为了通过样本的情况来估计总体。取样之所以成为需要是由于：

1. 在许多情况下，全面调查是不必要的，完全可以而且应该通过“解剖麻雀”而达到调查的目的，从而节省人力、物力。

2. 总体太大，全面观察成为不可能。

3. 有些调查是属于破坏性的，为了减少损失，节约材料，也必须取样，例如调查作物体内的害虫，常要破开植株的一部分或全部。

在昆虫学研究中，通常只能进行取样调查。

样本是用来代表和估计总体的，因此，样本的代表性最为重要，样本与总体必然存在差异，称为取样误差。取样技术如果合理，取样误差可以减少，样本代表性就会增强。关于取样技术，下面我们将有专章讨论。

1-5 什么是统计方法

统计方法简单地说是指：搜集、整理、分析和表达（解释）数据的技术，详细地说来是：

1. 提供调查取样和试验设计的依据，使能合理地、正确地搜集数据。

2. 作为数据整理的方法，使之呈现系统的、紧凑的形式，便于进一步的分析工作。

3. 作为数据分析的方法，内容包括：

①. 计算平均数，标准差等“特征数”，进行图示、图解、制表等，以表示资料的特征。

②. 进行不同资料之间的相互比较，即差异显著性的检验。

③. 明确不同资料之间的相互关系，常用的是相关与回归。

分析。

4. 帮助调查与试验结果的判断和解释, 例如理论模式的拟合等。

1-6 正确对待统计方法

在科学研究中, 统计方法是一项重要而有用的工具, 但必须在唯物辩证法的指导下审慎地运用, 统计方法不能代替害虫发生规律本身, 在许多情况下, 不能充分说明事物的内部联系, 它不能代替各方面专业知识和理论对害虫发生机制的解释。因此, 统计分析的结果, 有助于^全面的科学结论, 但却绝不能代替。

再者, 准确的数据比之统计分析方法更为重要, 如果数据不可靠, 即使经过统计处理, 仍然不可能避免谬误。因此, 进行调查和试验必须有严肃的科学态度, 一丝不苟。

第三, 统计方法要结合实际, 灵活运用, 不要不求甚解地盲目套用, 否则会歪曲实际情况。

第二章 害虫数据资料的特点 和初步整理方法

2-1 害虫数据资料的特点

在害虫的调查研究中，就同一事项（性状）往往可以得到多个的观测值或称变值。这样的事项（性状），就叫做变数，已如本章所述。例如虫态历期、卵期、各龄幼虫期、蛹期、成虫寿命、个体之间会不相同，又如害虫体重、体长也是如此，至于诱蛾灯每夜诱获三化螟蛾数、三化螟每个卵块含卵粒数，每卵块孵出蚁螟数，每亩蛾数，每亩卵块数，每个白穗群的白穗数等等，也都是变数。我们在测报工作中，经常收集到和要处理的就是这样的变数资料。

为了要处理变数资料，首先要了解变数的分类。变数可以分为两大类：一类比较简单，整理比较容易，叫做“两级的变数”。一堆变数资料只能分为两部分，非此即彼，中间的情况是没有的，如雌性和雄性，死亡和存活，已孵化和未孵化，卷叶和正常叶，虫蛀果和完好果等等，这类变数通常以比值或百分比来表示，测报中常见的有如下一些两级变异指标：孵化率、化蛹率、羽化率、雌雄比，死亡率、卷叶率、白穗率、寄生率等。

两级变数在昆虫研究中是比较多的。

另一类变数可以叫做多级变数。我们所得的观测值由小到大表现为一系列不同程度的数值。

从另一方面观察，变数又可分为连续性变数和非连续性变数两类。前者如昆虫体重、体长、温度、降水量……凡是属于长度、重量、容量、面积等事项的观测值都是，它的特点是测定时可以有小数，即只能量不能数，因此，又叫做可量变数，后者如昆虫产卵数，被害叶片数，诱蛾灯每夜诱得某种蛾数，三化螟每个卵块孵出蚁螟数，每个白穗群的白穗数，每科水稻的分蘖数，每个禾头的三化螟越冬幼虫数等等，测定时不可能有小数，只能一

一个点英，即只能数不能量，所以又叫质可数变数（注意：原始资料不能有小数，但归纳指标如平均数，是可以有小数的）。

值得注意的是：在昆虫学研究中，大多数变数资料属于可数变数，这是一个重要特点。

处理多级变数资料手续要复杂些，下面将着重讨论它们的整理方法。

2-2 分组和频数分布表的编制

从调查得来的原始资料，如果观测值个数不多，资料的概况容易掌握，当然不需要经过什么整理手续，就可以直接计算归纳指标和分析。但是，如果观测值个数多，一下子看起来好象是一堆乱杂无章的数字，从表面上不容易看出分布趋势的梗概，更不便于作进一步的计算和分析。在这种情况下，整理工作就成为必要的了。由于通常强调观测值个数要多些才能更好地反映实际情况，因而整理工作常是不可少的。

整理工作的第一步是制成“依次表”，使资料系统化，也就是把各个观测值即变值，按由小至大的顺序在纸上排列起来。如果数字很多，可将每一个数字登记在一个小卡片上，然后按数值大小，由小到大顺序地排列卡片，最后依次抄写在纸上即成。当资料的状况易于掌握时，依次表的制作可以有去。

整理工作的第二步，即核心的工作，是制作“频数分布表”。就是分组归纳各个变值。这个表的结构，其实也很简单，要点是：第一栏是“分组”，好象大家开会或工作时把人员分为若干小组一样。第二栏是“频数”，就是把各个观测值分别归到各组去，计算每组有多少观测值，这个数字叫做频数，有时还设有第三栏：“比率”，即将第二栏绝对频数转变为相对频数，这样就便于各组之间的相互比较，也便于不同频数分布表之间的相互比较。

分组的办法要按资料的性质而定。

最简单的一类分组办法是“按质分组”，例如，按“色泽等

级”、“受害等级”、“昆虫龄态”等分组，举例如下：

例一：某些果树的果实受虫害后，可按受害程度（蛀孔多少、大小和害斑大小等）分级，再据各级果实数列成下面形式的频数分布表：

受害等级 (级值)	各级受害果数 (频数)	比率 (%)
1	20	5.56
2	40	11.11
3	100	27.28
4	120	33.33
5	60	16.67
6	20	5.56
合计	360	100.00

例二，昆虫龄态的频数分布表，在测报上称为发育进度表，通常每隔2~3天调查一次，每次都可以列成一表。根据历次各表中各龄态频数的变化，可看出昆虫发育的趋势，每次调查结果的表结构如下：

龄态	虫数(频数)	比率(%)
1	2	2.9
2	13	18.6
3	30	42.9
4	12	17.1
5	6	8.5
蛹	5	7.1
蛹壳	2	2.9
合计	70	100.0

将历次龄态调查所得的频数分布表合并为一总表，是测报上常用的表格之一，下面是一个三代螟一个世代的发育进度总表，(现行计英方法是只算活虫)：

第二世代三化螟发育进度表 (中山沙朗, 1958)

检查 日期 (月/日)	活 幼 虫					活 蛹	壳	总 活 虫 数	比 率 (%)							总蛹 化率 (%)
	1令	2令	3令	4令	5令				1令	2令	3令	4令	5令	蛹	蛹壳	
5/23	48	4						52	92.3	7.7						
5/27	14	35	12					61	23.0	57.4	19.7					
5/30	20	26	21	6				73	27.4	35.6	28.8	8.2				
6/2		48	7	2				57		84.2	12.5	3.3				
6/5		22	42	6				70		31.4	60.0	8.6				
6/8		6	42	9				57		10.5	73.6	18.9				
6/11		9	35	14	1	1		60		15.0	58.5	23.3	1.7	1.7		1.7
6/14			34	26	1	1		62			54.8	41.9	1.6	1.6		1.6
6/17			43	15		10		68			63.2	22.1		14.7		14.7
6/20			15	41		17		73			20.5	56.1		23.3		23.3
6/23			10	25		11	6	52			19.2	48.1		21.2	11.5	32.7
6/26				14	4	24	11	53				26.4	7.5	45.3	20.8	66.1
7/3			5	6		15	26	49			4.1	12.2		30.6	53.1	83.7
7/6				1		13	33	47				2.1		27.7	70.2	97.9
7/9							10	10							100.0	100.0

例三, 近年在三化螟测报上还常常把剥检得的活蛹, 按色泽状态分级, 以便更精确地计英发蛾盛期。在蛹分级的工作中, 也首先要列出频数分布表。

例如, 三化螟化蛹普查调查, 得活虫 118 头, 按虫态可列出如下频数分布表。

X (指虫态)	f (频数, 指活虫数)	频数比率 (%)
幼 虫	35	29.66
预 蛹	14	11.86
一 级 蛹	27	22.88
二 级 蛹	17	14.41
三 级 蛹	10	8.47
四 级 蛹	9	7.63
五 级 蛹	6	5.08
N = 118		99.99

更常见的频数分布表的分组是“按数量分组”，又可分为两类：

(1) 可数资料：

可数资料整理成频数分布表时，在很多情况下，可以简单的整数的“组值”为依据，例如：

检查 200 个木头，每个木头内越冬三化螟幼虫数的原始资料如下：

1, 1, 0, 0, 2, 0, 0, 1, 0, 2, 1, 0, 1, 1, 0, 1, 0, 0, 3, 0,
 2, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 5, 0,
 1, 0, 0, 0, 0, 4, 2, 0, 0, 3, 0, 4, 1, 3, 1, 4, 0, 1, 2, 0,
 0, 3, 2, 1, 0, 2, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 2, 6,
 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 1, 0, 1, 0, 0, 1,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0,
 1, 1, 1, 0, 0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 0, 0, 1,
 1, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1,
 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0.

可以看出，这类资料的特点：① 变值中“0”值很多；
 ② 较大的变值为数少。上面原始资料，可制成如下的频数分布表：

X (每禾头内的 虫数, 即组值)	f (禾头数 即频数)	比率 (%)
0	134	67.0
1	48	24.0
2	9	4.5
3	4	2.0
4	3	1.5
5	1	0.5
6	1	0.5
N (合计) = 200		100.0

这样的表, 叫做闭式表, 有时因为高组值项频数很少, 但项数多, 零散, 就写成开式如下:

X	f	比率(%)
0	134	67.0
1	48	24.0
2	9	4.5
3	4	2.0
≥ 4	5	2.5
$N = 200$		100.0

有时资料太繁复, 相同的观测值过多, 为了使频数分布表呈压缩的形式, 易于看出分布的特点, 可按“组限”来分组。所谓组限, 就是一定的数量范围。例如:

1956年广州石牌第四世代三化螟53个卵块含卵粒数的资料, 频数分布表可列如下式 (以30粒卵为一组)。

X (每卵块含卵粒数, 即组限)	f (卵块数, 即频数)	比 率 (%)
1 ~ 30	0	0.00
31 ~ 60	4	7.55
61 ~ 90	11	20.75
91 ~ 120	14	26.42
121 ~ 150	11	20.75
151 ~ 180	7	13.21
181 ~ 210	4	7.55
211 ~ 240	2	3.77
$N=53$		100.00

(2) 可量资料

可量资料在昆虫研究和测报工作中遇到较多, 这里只简单地介绍其频数分布表的要点, 不作详和讨论。

就可量资料制作频数分布表时, 依据“组限”来分组, 也是以一定数量范围 (即组限) 作为一组, 每个组有下限和上限, 决定了第一组的下限, 以后各组都可以连续地推定了。可量资料的组限的写法与可数资料不同, 因为资料是连续性的, 为了避免前后两组的上、下限的混同和便于观测值的归属, 应该写成如下方式:

下 限	上 限
0	9.9
10	19.9
20	29.9

组距的大小, 要按资料的具体情况和分组多少而定, 以偶数为好。可能时, 以 10、20、30 等为一组, 适合于一般统计习惯。由于组限整齐, 使人易得深刻印象, 但也不应受习惯限制, 至于组数, 以 6~20 组为适当。

可量资料的频数分布表, 往往还有“组中值”一栏, 是为了方便于统计归纳指标和图示而设立的。组中值就是各组的代表值, 它的计算很简单:

$$\text{组中值} = \text{下限} + \frac{1}{2} \text{组距}$$

对组中值有一个要求，就是等于或极接近于各该组内观测值的平均数。

下面是一个可量资料频数分布的例子。

1957年3月1日在广州石牌称量10白头越冬三化螟幼虫的体重（单位：毫克），把原始资料列成依次表后，以10毫克为组距，分成13个组，频数分布表如下：

组 限	$\bar{x}$ (组中值)	f (频数)	比 率 (%)
10 ~ 15.9	13	1	0.94
16 ~ 21.9	19	2	1.89
22 ~ 27.9	25	8	7.54
28 ~ 33.9	31	22	20.75
34 ~ 39.9	37	20	18.87
40 ~ 45.9	43	23	21.70
46 ~ 51.9	49	8	7.55
52 ~ 57.9	55	6	5.66
58 ~ 63.9	61	9	8.49
64 ~ 69.9	67	1	0.94
70 ~ 75.9	73	2	1.89
76 ~ 81.9	79	3	2.83
82 ~ 87.9	85	1	0.94
		$N=100$	99.99

2-3 制作频数分布表的意义

从本节举出的几个频数分布表，大家就会得出一个初步结论：原来表面上看来很不规则的变数资料，经过整理，制成频数分布表之后，竟然并不是杂乱无章的，而是有它自己的规律的，各