昆虫田间抽样技术

夏 基 康

南京农业大学植保系农虫教研组

一九八六年六月

引　言

在农业生态系统中，对昆虫及螨类进行抽样调查的目的，主要是了解农业生态系统中生物群落各成份的性质及其在时间和空间的动态状况，为提出有害生物综合治理决策提供依据。为此，植保工作者需要选择一个适宜的方法来取得适当含量的样本，并对所得的数据进行统计与分析。在地球的生命层(或叫生物圈Biosphere)中，昆虫介于细菌、病毒和两栖动物与哺乳动物之中，居于中间层的位置。其种类之繁多，习性之各异和栖息地的复杂状况为任何其他动、植物所难与攀比。因此调查研究昆虫的抽样技术和方法要独特得多。廿余年前，*Morris* (1960) 曾就昆虫的抽样调查形象地形容为“艺术、科学和乏味单调工作的结合”。与一般抽样相比，对昆虫的抽样调查在后两点上表现了共性，而在第1点上却更加体现了出来。

目　录

第一章　基本概念

1.1　计数与计量

使用数字包括两个方面—计数物体的数量和选定测量的等级—计量。如果我们数一下上面这一句句子的字数，共有27个字。这是一个用数字计数物体数量的简单例子。象这样的数字是一个绝对值，它不随计数时间或方法而变更。

然而，我们的一些重要资料，除了点数物体即计数，还来自称量、计算体积和仪器读数。在农业科研活动中，全部计数或计量都包含在一个未知区（范围）之中。在任何试验、调查中，由真值加上误差所组成。

1.2　试验误差

农业调查中的误差分为固定误差与偶然误差两类。

固定误差：系由下列一些因素所造成，（1）因田间操作技术的不一致，如播种不匀，灌溉、中耕除草、防治病虫时间不一和质量上的差别；（2）因工作上的疏忽，如抽样方法设计不当，观察记载错误，整理资料时看错数字，仪器有毛病而事先未予检查发现。这类误差往往是由于人们的活动所造成的。

偶然误差：（1）土壤肥力差，供试材料不一致（种子匀净不一，秧苗大小强弱不一）；（2）由病虫害、鸟兽鼠害以及冰雹等自然灾害引起的不一致；人畜践踏等影响。

固定误差在被察觉时可以避免，而偶然误差难以被消除。它们影响整个试验工作的精确性。我们要尽可能地减少这种误差。

最通常的减少偶然误差的方法就是对同一种事件给予成组的重复计数或计量，然后汇成平均数。

1.3　算术平均数与几何平均数

算术平均数系计数或计量的总和被个数除得的商数。

$$\overline{x}=\frac{\overline{x}_1+\overline{x}_2+\overline{x}_3+\cdots+\overline{x}_n}{n}=\frac{\Sigma\overline{x}}{n} \qquad (1.1)$$

“Σ”音*Sigma*，它表示总和的意义，本身不代表任何数值。

从公式（1.1）可以得出下列两点：

（1）算术平均数是倾中性（*central tendency*）的计算单位，其离均差的总和等于零。

（2）由于它是由一组不同的观察值计算所得，因此不是一个绝对值。

例如：徐州地区所1978年12月检查大豆品种茎杆内豆秆黑潜蝇（*Melanagromyza sojae*）幼虫密度，其中“新海连白花糙”与“铜山平顶黄”两品种10株的平均单株虫量为：

$$\text{“新海连白花糙”}\ (\overline{x}_1)=\frac{11+14+9+22+16+12+5+4+7+5}{10}$$

$$=11.5\ (\text{头/株})$$

“铜山平顶黄” $(\overline{x}_2)=\dfrac{6+5+2+2+2+0+2+1+3+5}{10}$

$=2.8$（头/株）

算术平均数的可靠性依赖于我们所得的各个数值。如16是 0 与32的平均数，也可以是15与17的平均数。对于象 0 与32所取得的平均数16，我们不能或很少能信任其准确性。对于由15、17所取得的平均数16，可以认为是一个良好的估计真值。由此表明，平均数所代表的是一组数值的倾中性，而不能表示这组数值间的变异程度即离散性，在讨论离散性之前须先了解正态分布。

几何平均数（*Geometric average*），在统计数值表现在逐个递增或递减情况下，可以用几何平均数表示平均增长（递增）率。用Ga表示，其算式如下：

$$Ga=\sqrt[n]{x_1\cdot x_2\cdot x_3\cdot\cdots\cdot x_n}$$

$$=(x_1\cdot x_2\cdot x_3\cdot\cdots\cdot x_n)^{\frac{1}{n}}$$

$$\log Ga=\frac{\lg x_1+\lg x_2+\lg x_3+\cdots+\lg x_n}{n}=\frac{\Sigma\lg x}{n}$$

$$G_a=arc\lg\left(\frac{\Sigma\lg x}{n}\right)$$

因此，几何平均数是各计数值的对数平均值的反对数。

例如测得豆秆黑潜蝇 1 、 3 、 5 、 7 各日令的平均体长与口咽骨平均长度列于下表，并测算其平均增长率。

日　令	虫体长度 $(L_1)(mm)$	口咽骨长度 $(L_2)(mm)$	每日令的相对增长率		lgx	
			x_1	x_2	x_1	x_2
1	0.5106	0.0344				
3	0.9246	0.0608	1.8108	1.9767	0.2579	0.2959
5	2.5900	0.1069	2.8012	1.7582	0.4473	0.2451
7	2.9710	0.1100	1.1471	1.0290	0.0596	0.0124
Σ			5.7591	4.7639	0.7648	0.5534

$$\therefore\lg G_{a1}=\frac{\Sigma\lg x_1}{n}=\frac{0.7648}{3}=0.2549$$

$$\lg G_{a2}=\frac{\Sigma\lg x_2}{n}=\frac{0.5534}{3}=0.1845$$

$$G_{a1}=arc\lg 0.2549=1.7985$$

$$G_{a2}=arc\lg 0.1845=1.5293$$

表明豆秆黑潜蝇1、3、5、7各日令幼虫的体长与口咽骨长度分别平均增长1.8倍和1.5倍。

若用算术平均数计算，则上述日令间在体长与口咽骨长度的平均增长率分别为

$$\overline{x}_1=\frac{5.7591}{3}=1.9179$$

$$\bar{x}_2 = \frac{4.7639}{3} = 1.5880$$

1.4 正态分布（*Normal distribution*）与*T*分布

假若调查某玉米品种1000株成株期植株高度，并把这些数字按其大小所发生的频率列成坐标图，以横坐标x表示高度，纵坐标y表示株数。我们会发现，这些数值会分布成一个铃型，即多数的数据居于中间，而其它数据以逐渐减少的趋势向中心两侧伸延。这样的数据分布可以用图1的曲线表示，这就是正态分布曲线。

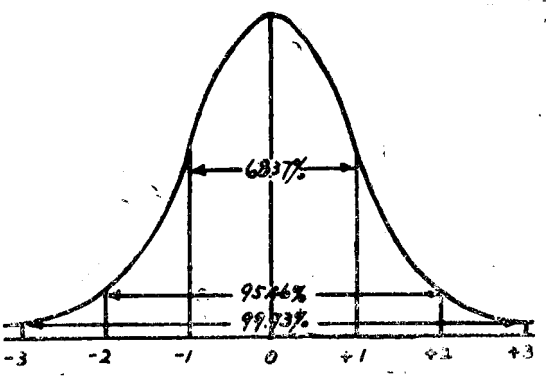

图1 常态分布曲线

正态分布曲线具有两种性质：

（1）它的平均数（$\bar{x}$）位于沿x轴所形成的曲线的中心，其标准差（σ）则表示数据x轴的分散情况。

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} \qquad (1.2)$$

（2）数据频率分布准确。例如：95%的的数据介于$\bar{x}-1.96\sigma$和$\bar{x}+1.96\sigma$之间，而分布在$\bar{x}\pm3\sigma$范围内的将包括99.7%的个体在内。这意味着在1000个植株高度的变数值中，将有50个数值在$\bar{x}\pm1.96\sigma$之外，仅有3个变数值在$\bar{x}\pm3\sigma$之外。

以实际应用的观点来看，从相反的方面来说明上一例子，可能更为重要。假定1000株玉米所得的平均高度，已知为1.75米，$\sigma=0.15$。现取得一个计数值，系在1.45到2.05米范围之外。根据以上所述推理，该数值出现的机会在一百次中将只有五次。象这样小的机遇，该计数值是否来自1000株玉米的？因此应被怀疑。这样的判断就是进行显著性测验的基础。

正态分布的理论是从大量的数据中发展而得，但并不适用于小量的观察值。在实验室内或在田间抽样调查中，我们不可能产生出非常大的观察值。建立在正态分布理论基础上的统计测算可能会从小量观察值中引出虚假（缪误）的结论。这一事实被爱尔兰的化学家高塞脱所发现。1908年他以“学生氏”（*student*）为笔名，发表一篇叫做“平均数的机率误差”的论文。部分依籍于理论上的推导，部分引自小量随机样本。他获得引自正态分布小量样本平均数的理论分布—t分布。

假若我们不应用大量的样本，我们不可能知道真正的标准差（σ）或真正的总体平均数（$\bar{x}$）。但是，我们可以用样本的标准差（s）代替（σ）。当我们进行这一统计分析时，我们必须使用一已知的独立于σ的分布。由高塞脱氏介绍的这个概念，就是众所周知的“学生氏t”。

$$t = \frac{\bar{x}-\mu}{s_{\bar{x}}} \qquad (1.3)$$

式中$\bar{x}$=样本平均数，μ=总体(集团)平均数，$S_{\bar{x}}$为标准误差，即样本平均数的标准差。

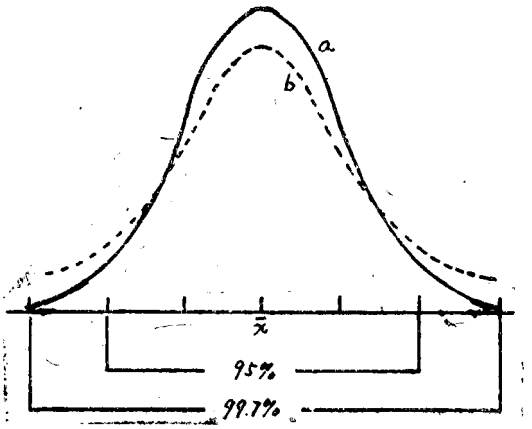

图2 * t 分布曲线

学生氏指出，t 的分布只取决于样本的大小（n）。图2中的 b 曲线（虚线）即 t 分布曲线。从该图可以看出 t 分布曲线与正态分布曲线之间的关系。前者比起后者显得较平坦些，当样本容量增加时则 t 曲线接近正态曲线，而当 n 达无穷大时，则两者相等。在实际应用中，通常对于样本大于30的用正态分布。

了解 t 分布的概念很有必要。这个概念是所有显著性测验（包括来自小样本的两个平均数之间的比较）的基础。有关其具体的应用将在第八章中给以介绍。

1.5 平均数的局限性

在1.3节内曾提到平均数能代表一组计数值的倾中性，但不能表示它们的变异程度（离散性）（*Variation, spread*）。试看下列三组数值：

$$\left.\begin{array}{l}(1)\ 8,8,9,10,11,12,12\\(2)\ 5,6,8,10,12,14,15\\(3)\ 1,2,5,10,15,18,19\end{array}\right\}\bar{x}=\frac{70}{7}=10$$

以上三组数值的平均数 $\bar{x}$ 均为10，但（1）组的各计数倾向（接近）于平均数；（2）组要分散些；（3）组更为分散，两极端值的差数（全距）各组分别为4，10，18。

有三种方法估计变异程度：（1）平均离散；（2）方差与标准差；（3）全距。

1.6 平均离散（*The average deviation*）

假若我们不管数字的正负，而将图1.3中的 $x's$ 总加起来，可以得到离差总和，以 n 除之，得平均离差。

$$\text{平均离差 }(AD)=\frac{\Sigma(x-\bar{x})}{n} \qquad (1.4)$$

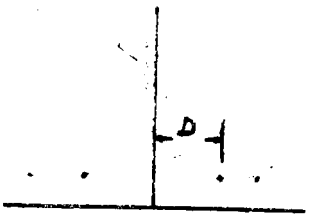

图3 距平均数的差数

用平均离差来计算数据的离散性(变异程度)并不精确，这是因为平均离差会带来“偏性”（*bias*），使数据的精确性超过实际情况（见表1）。

1.7 方差（*variance*）与标准差（*standard deviation*）

方差（V）是由离均差平方的总和数被自由度除之而得。方差的平方根就是标准差。

$$\text{样本的方差 } V=S^2=\frac{\Sigma(x-\bar{x})^2}{n-1} \qquad (1.5)$$

$$\text{样本的标准差 } S=\sqrt{V}=\sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} \qquad (1.6)$$

表 1　平均数、平均离差与离均差的平方

x_1	$x_1-\overline{x}_1$（不管正负）	$x_1-\overline{x}_1$	$(x_1-\overline{x}_1)^2$	x_2	$x_2-\overline{x}_2$（不管正负）	$x_2-\overline{x}_2$	$(x_2-\overline{x}_2)^2$	x_3	$x_3-\overline{x}_3$（不管正负）	$x_3-\overline{x}_3$	$(x_3-\overline{x}_3)^2$
8	2	−2	4	5	5	−5	25	1	9	−9	81
8	2	−2	4	6	4	−4	16	2	8	−8	64
9	1	−1	1	8	2	−2	4	5	5	−5	25
10	0	0	0	10	0	0	0	10	0	0	0
11	1	1	1	12	2	2	4	15	5	5	25
12	2	2	4	14	4	4	16	18	8	8	64
12	2	2	4	15	5	5	25	19	9	9	81
$\overline{x}_1=10$	$\Sigma(x_1-\overline{x}_1)=10$ $AD=10/7=1.43$	$\Sigma(x_1-\overline{x}_1)=0$	$\Sigma(x_1-\overline{x}_1)^2=18$	$\overline{x}_2=10$	$\Sigma(x_2-\overline{x}_2)=22$ $AD=22/7=3.14$	$\Sigma(x_2-\overline{x}_2)=0$	$\Sigma(x_2-\overline{x}_2)^2=90$	$\overline{x}_3=10$	$\Sigma(x_3-\overline{x}_3)=44$ $AD=44/7=6.29$	$\Sigma(x_3-\overline{x}_3)=0$	$\Sigma(x_3-\overline{x}_3)^2=340$

这里用（$n-1$）代之n，这是因为我们用的是样本平均数。统计学家发现，从样本所得的离均差平方的总和数较小于由总体所得的离均差平方的总和数，用自由度（$n-1$）就可以消除这个偏差。例如：现有两个观察值2与4，它们来自已知平均数为2的总体，应用公式（1.2）。

$$\text{总体的方差}\sigma^2=\frac{\Sigma(x-\overline{x})^2}{n}=\frac{(2-2)^2+(4-2)^2}{2}=2$$

$$\text{若我们用样本平均数：}\frac{2+4}{2}=3$$

$$\text{将之代入上式，}\sigma^2=\frac{(2-3)^2+(4-3)^2}{2}=1$$

假如用自由度（2－1）去除，应用公式（1.5）

$$V=\frac{(2-3)^2+(4-3)^2}{(2-1)}=2$$

这里，从样本平均数所得的方差（V）与总体的方差（σ^2）相同。为什么将（$n-1$）叫做自由度呢？现以表2中x_1这组的变数值为例。为了满足$\overline{x}_1=10$这一条件，在7个计数值中只有6个可以自由变动，但最后一个就没有变动的自由了。例如：7个计数值中有6个可以任取8、8、9、10、11、12，最后一个就非取12这个值不可了，否则就不符合$\overline{x}_1=10$的条件。由此可见，自由度等于计数值个数减去制约着它们的常数的个数。本例的常数只有一个（$\overline{x}_1=7$），因此自由度为（7－1＝6）。

现将表1所列三组计数值的材料，列成下表2，从表中可以看出方差与标准差的计算过程。

表2　方差与标准差

x_1	$x_1-\overline{x}_1$	$(x_1-\overline{x}_1)^2$	x_2	$x_2-\overline{x}_2$	$(x_2-\overline{x}_2)^2$	x_3	$x_3-\overline{x}_3$	$(x_3-\overline{x}_3)^2$
8	－2	4	5	－5	25	1	－9	81
8	－2	4	6	－4	16	2	－8	64
9	－1	1	8	－2	4	5	－5	25
10	0	0	10	0	0	10	0	0
11	＋1	1	12	＋2	4	15	＋5	25
12	＋2	4	14	＋4	16	18	＋8	64
12	＋2	4	15	＋5	25	19	＋9	81
$\overline{x}_1=10$		$\Sigma(x_1-\overline{x}_1)^2=18$	$\overline{x}_2=10$		$\Sigma(x_2-\overline{x}_2)^2=90$	$\overline{x}_3=10$		$\Sigma(x_3-\overline{x}_3)^2=340$
$S_1^2=\frac{\Sigma(x_1-\overline{x}_1)^2}{n-1}=\frac{18}{6}=3$ $S_1=\sqrt{\frac{\Sigma(x_1-\overline{x}_1)^2}{n-1}}=\sqrt{3}=1.73$			$S_2^2=\frac{90}{6}=15$ $S_2=3.87$			$S_3^2=\frac{340}{6}=56.67$ $S_3=7.53$		

上述的离均差平方的总和，计算繁复。凡小样本资料可以应用下列简化公式：

$$V=\frac{\Sigma x^2-\frac{(\Sigma x)^2}{n}}{n-1} \tag{1.7}$$

$$S=\sqrt{\frac{\Sigma x^2-\frac{(\Sigma x)^2}{n}}{n-1}} \tag{1.8}$$

怎样从$\Sigma(x-\overline{x})^2$简化计算为$\Sigma x^2-\frac{(\Sigma x)^2}{n}$的，请看下列演算：

$$\begin{aligned}\Sigma(x-\overline{x})^2&=\Sigma(x^2-2x\overline{x}+\overline{x}^2)\\&=\Sigma x^2-2\overline{x}\cdot\Sigma x+n\overline{x}^2\\&=\Sigma x^2-2\overline{x}\cdot n\overline{x}+n\overline{x}^2\\&=\Sigma x^2-2n\overline{x}^2+n\overline{x}^2\\&=\Sigma x^2-n\overline{x}^2\\&=\Sigma x^2-n\left(\frac{\Sigma x}{n}\right)^2\\&=\Sigma x^2-\frac{n(\Sigma x)^2}{n^2}\\&=\Sigma x^2-\frac{(\Sigma x)^2}{n}\end{aligned}$$

以上计算过程的总趋势，是将$\overline{x}$转化为x，然后相消合并。

自从应用计算器和微机以后，原先计算繁复的过程已是“一去不复返”了。但是对于初学者来说，了解标准差计算过程以及不同算式之间的关系仍然很有必要。正如笔算、心算以及算盘不能被计算机完全取代是一样的道理。这里还需要着重指出的，方差是度量变异程度的特征数，又是近代生物统计的一项十分有效的分析法。但是，作为样本变异程度的特征数，其数值水平最好要与样本平均数的数值水平相适应。其度量衡单位也应与平均数单位取得一致。在计算方差时，由于离均差经过平方，数值水平增大很多，度量衡单位也随着平方。以斤、克、头来说，平方斤、平方克、平方头等在概念上不好理解。因此，对于样本变异程度的变量，一般采用方差的平方根即标准差来表示。

标准差主要有两个方面的用途：（1）测定研究对象变异程度的大小。凡是其他条件相同，标准差大的，变异程度大；标准差小时，变异程度小。例如比较两个品种穗长这一性状的变异程度，两品种平均穗长相近，则只要分别求出标准差，标准差大的品种，变异程度就大。（2）除以上的用途外，标准差还有更重要的作用，可以用来作为度量试验误差的标准尺度。

1.8 全距（*The Range*）

指一组数列中最大数与最小数的差数。如果将这些数据由大到小进行排列：

$$x_n\cdots\cdots x_{n-1}\cdots\cdots x_{n-2}\cdots\cdots x_1$$

则 $$R = x_n - x_1 \qquad (1.9)$$

与标准差相比较，用全距度量变异程度不那样有效。但由于计算方便，在好些领域（如化学测定）中已成为计算变异程度的通用技术。对这方面感兴趣的读者可以参阅鲍埃（*E.L. Baues*）著“化学家的统计手册”一书。

1.9 变异系数（*Coefficient of variability*）与精密度（*Relative variation*）

对于平均数水平相近的材料，可以应用标准差来比较两个或两个以上样本变异程度的大小。但是对于平均数水平不同的材料，就不能直接用标准差，而要把标准差换算成变异系数再行比较。变异系数是标准差与平均数相比而成的系数，如下式所示：

$$CV = \frac{S}{\bar{x}} \qquad (1.10)$$

实例1：查得“丰产3号”小麦株高平均数为$\bar{x}=113cm$，标准差3.83cm。“墨西哥120”矮杆小麦株高平均数为$\bar{x}=68cm$，标准差为2.55cm（范濂等，1977）。根据3.83cm>2.55cm是否可以认为前者的变异程度大于后者呢？显然不能。因为“丰产3号”植株较高，即使变异程度相同，样准差也较大；后者较矮，标准差相应也较小。应用变异系数进行比较，可以解决这个问题。主要有以下两种用途。

（1）用来比较两个平均数水平不同的材料的变异程度。如将有关数值代入公式(1.10)

“丰产3号”株高$CV = \frac{3.83cm}{113cm} = 3.5\%$

“墨西哥120”株高$CV = \frac{2.55cm}{68cm} = 3.7\%$

由此可见两者变异程度相同。

（2）用来比较同一品种不同性状间的变异程度。由于标准差都带有与平均数相同的度量衡单位，因此对于不同性状之间的变异程度也不能用标准差直接比较。

例如“丰产3号”干粒重的平均数$\bar{x}=41.3$克，标准差是3.54克。这3.54克怎样能与同一品种的株高标准差3.83厘米相比较呢？采用变异系数，就可以解决。同样将有关数值代入公式（1.10）：

“丰产3号”株高$CV = \frac{3.83厘米}{113厘米} = 3.5\%$

“丰产3号”干粒重$CV = \frac{3.54克}{41.3克} = 8.6\%$

由此可见“丰产3号”干粒重的变异程度比株高大得多。

精密度从其计算的来源而言，与变异系数相仿。将试验或调查所得的标准误被与其对应的样本平均数相除，再乘以100即得精密度。其计算公式为：

$$RV = (S_{\bar{x}}/\bar{x})100 \qquad (1.11)$$

实例2 ..R.B.Hammond等（1976）调查大豆上一种叶夜蛾密度，应用RV来比较三种取样方法的精密度。在1英亩（合6.07市亩）田内，每种方法各随机取样20点，所得结果如下表3：

表3 三种抽样方法的精密度比较

日期	取样方法					
	直接法（袋法）		摇株法		网捕法	
	$\bar{x}\pm S_{\bar{x}}$	RV	$\bar{x}\pm S_{\bar{x}}$	RV	$\bar{x}\pm S_{\bar{x}}$	RV
8月17日	5.6±0.41	7	5.55±0.55	10	2.65±0.04	17
8月28日	1.3±0.30	19	1.90±0.26	14	0.70±0.15	21
8月30日	3.8±1.40	11	4.70±0.72	15	1.35±0.30	22

上表表明，在由三种抽样方法取得的数据之间，其变异程度都在可允许的范围内。换言之，与直接取样法相比较，摇株法与网捕法均较精确。因为*Southwood*（1966）曾提出，凡大面积（较为粗放的）抽样调查的RV值平均为25，小样本（较细致的）抽样调查所得的RV值平均为10，都属精确。上述三种方法调查所得的$\overline{RV}$值均≤25。表明应用摇株法与网捕法可以和较为费工的袋法一样取得良好的结果。

*Hammond*等进一步应用相对纯净精密度(*Relative net precision*)分析三种抽样方法所化费的人工（取样与数虫量的时间）与精密度之间的关系，比较结果列下表4。

表4 三种抽样方法的相对纯净精密度

抽样方法	人工* C	平均精密度 $\overline{RV}$	相对纯净精密度**
直接法（袋法）	10.00	12.33	0.80
摇株法	1.00	13.00	7.70
网捕法	1.00	20.00	5.00

* 20个样点所需的近似人工（人一小时）。

** 相对纯净精密度$=\dfrac{1}{\overline{RV}\cdot C}\times 100$ (1.12)

上表中以摇株法的相对纯净精密度值最大，这表明摇株法的经济效益最高。

应用方差与平均数两者之间的关系，还可以用来分析昆虫田间分布型的特征，从中推导出很有意义的概念。这将在第三章中进行介绍。

第二章　昆虫空间分布型

2.1　空间分布型

提到“分布”两字，人们一般会将这个名词与“生存”、“发生”相联系。如水稻分布指的是从南纬几度到北纬几度可以种植，在这范围内适宜于水稻的生存和繁殖。又如昆虫分布则常指它的发生而言。棉红铃虫在我国除新疆、西藏以外都有发生，这就是红铃虫在全国的分布状况。本章所讲的昆虫空间分布型（包括田间、温室、仓库）与上述的大区（地理）分布有原则区别。所谓昆虫分布型，就是指某种昆虫种群在所发生的生境或生活小区内，在特定时间内的空间分布结构。棉花、水稻、大豆等作物的植株均匀地遍布于大田，可以称之为均匀的分布型。在这些作物上生活着的各种昆虫，它们的个体在田间是密集的还是分散的？则要联系具体的虫种进行分析。小地蚕卵散产在杂草上、土缝内或土块上，孵化的幼虫寻觅杂草作食料。由于杂草在田间分布得不均匀，因此小地蚕幼虫在田间分布是不均匀的聚集状态。水稻三化螟、二化螟和玉米螟成虫产的卵呈块状。这三种害虫的卵块在田间均呈随机分布，由卵块孵出的数十条幼虫以及受害的植株则呈核心状分布。可见昆虫种群在田间的分布状况因虫种及其虫态而呈现多样。如何取得经济有效的样本材料，并从中确切地分析推导总体的规律性，对于掌握昆虫分布型的基本概念和正确的抽样技术很为必要。

昆虫田间分布型在生物统计学上称谓集团分布。根据种群密度变异程度和性质，可以用数理统计推导的方法，将之归纳、分型并进行检验。在具体运用这些方法以前，需要先从经典的频次分布谈起。

2.2　频次分布

一个变量（*a variable quantity*）可以是连续性的，在某一范围内能呈现为任何值。也可以是间断性的，它只能是整数值（*integral value*）。连续性变数（*continuous variables*）通常可以测量，如高度、长度和体重；而间断性变数通常为计数，如毛的数目、花的瓣数，在一个抽样单位内的虫数和病斑数目等。

当测量或计数大量的变数时，可以用次数分布图进行概括，先将数字一一排列起来，然后归入各频次组（*frequency classes*）。每个频次组的 f 值表示含有相同虫量的抽样单位数。在每个频次组所含整数在一个以上时，频次组间不能重叠，组距又须相等。如1—5，6—10，11—15等，这里组距为5。

频次分布型可以用数字分布图表示，但若采用矩形图则更为清晰（见图4）。在矩形图内，每个频次组由柱表示，各个柱的面积与频次成比例。如果柱间之底相等（如组距相同），则表示柱的高度与频次成比例。其它矩形图见图5—6。（见2.2.5节）。

实例3．在频次分布中，数字的整理与平均数、方差的计算。

调查花生田土内蛴螬在各个抽样单位中的数量。取得一个大的随机样本，该样本含有80个抽样单位，每个抽样单位为1M^2。在每个抽样单位内的幼虫数为：

4，0，2，4，0，2，11，2，5，6
2，2，1，3，3，1，1，3，1，1

2，1，4，2，1，1，4，3，1，2
5，3，3，5，3，3，2，2，4，1
5，0，3，1，2，6，0，3，6，0
2，1，5，5，0，6，5，2，0，0
6，4，6，6，2，5，2，10，3，2
0，3，1，1，1，4，5，1，4，5

从上述各数中，找出最小值为0，最大值为11。列出由0至11的各个 x 值与相应的 f 值。取得频次分布如下：

x	0	1	2	3	4	5	6	7	8	9	10	11
f	9	16	16	12	8	10	7	0	0	0	1	1

$\Sigma f = n = 80$

这里 x 值表示每抽样单位内的虫数，f 值为在抽样单位中出现该虫数的频次，即具相同虫数的抽样单位数。总的抽样单位数（n）为：$\Sigma f = n = 80$。

在本例中，蚜螨的总数为 Σfx，因此该样本的算术平均数（$\bar{x}$）为：

$$\bar{x} = \frac{\Sigma fx}{n} = \frac{229}{80} = 2.8625$$

该样本的方差：（S^2）

$$S^2 = \frac{\Sigma fx^2 - \bar{x}\,\Sigma fx}{n-1}$$

$$= \frac{1039 - 655.5}{79}$$

$$= 4.8544$$

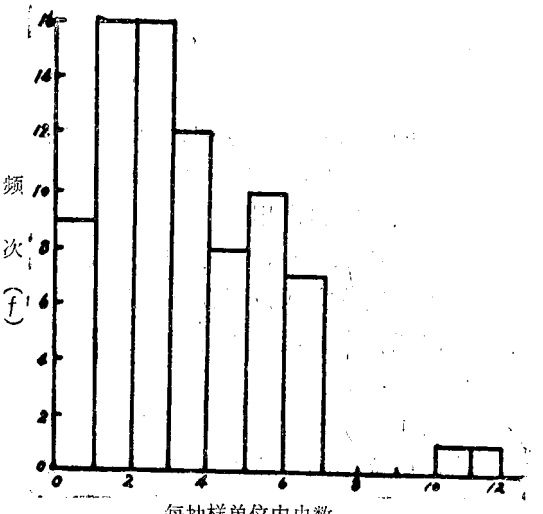

图4　例3各数的频次分布

四个常用的频次分布

已知的算术频次分布可以用来作为从种群中获得样本的模式。如果计数的频次分布适合于这些模式中的一个，那么：

（1）可以用该模式来描述种群的空间分布（见下一节昆虫种群分布型）。

（2）可以算得种群参数的误差（见第七章样本平均数的精密度）。

（3）对于种群密度在时间和空间上的改变进行比较（见第八章样本间的比较）。

（4）测出环境因子的效应。

在种群的方差（σ^2）与算术平均数（μ）之间的关系，通常与下面这三种数学分布的模式相符合。

（1）正二项式：当方差明显地小于平均数时（$\sigma^2 < \mu$），近似于这一模式。

（2）泊松式：为方差近似于平均数（$\sigma^2 = \mu$）时的唯一模式。

（3）负二项式：当方差显著大于平均数时，（$\sigma^2 > \mu$），有若干种模式相近似，其中