

华南农学院

变数统计学

森林经理学教研组

1958·7·



第一章 緒論

§1. 变数统计的研究对象

变数统计是数理统计的一部分，它是研究总体统计数量变动标识的分佈的，为了说明问题，先要解释下面一些名词。

总体：是客观存在的，在同性质的基础上结合起来的个别单元的集团。这些组成总体的单元称为总体单元。例如：某林区的全体树木可以组成一个总体，其中每一株树木，就是这个总体的个别单元。因为这些个别的单元，都具有一个共同的性质，即它们都是该林区的树木，因此，我们可以把它们看成是同一种类的，变数统计研究这些树木在数量上的某些特性时，是把这些树木的全体，当作一个整体来研究。

数量变动标识：表示总体单元某一方面特性（特征）的名称叫做“标识”，如果这种特性在总体各个单元上的表现不尽相同，这种标识称为“变动标识”，如果这种不同是数量上的不同，这种标识称为“数量变动标识”。例如：某林区的全体树木组成一个总体，其中每一株树在树高，胸高直径，材积等方面特性，往往是不完全相同的，其中有某树木可能是8m高，有的树木是10m高，有的树不胸高直径为6cm，有的只有5cm，因此，树高和胸高直径等特性，都是“数量变动标识”。数量变动标识又叫做变数，变数统计就是研究这种变数的分佈的，变数的分佈可以用表来说明（见表1—1），表中列出了某林区全体树木高度的分佈情况：

树高（标识值） （单位：m）	株数（次数）
10 ——— 12	3
12 ——— 14	200
14 ——— 16	787
16 ——— 18	10
合 计	1000

前面已经讲过，变数统计是数理统计的一部分，而数理统计是一门应用数学，它的研究对象，是受多种自发的偶然因素交织影响下的数量，所以变数统计所研究的数量也应该是受多种自发的偶然因素交织影响下的数量，所谓自发的偶然因素，就是人们还不能控制的因素，这些因素在错综复杂地影响着某一事物，使我们不能预计它的必然结果。例如：某一林区树木的材积，是受气候、土壤、耕作技术等因素错综复杂地影响着，有的因素使材积的生长偏向这一边，有的因素使材积的生长偏向那一边，而这些因素在对材积生长起作用时，并不是由人们来控制着的，所以像材积这样的数量，就是受多种自发的偶然因素交织影响下的数量。有人以为承认偶然现象的存在，就是反科学。在他们看来，一切现象的产生都是必然的，甚至你今天中午给蚊子咬了一口，也是“必然”。这完全是一种不符合事实的论调，事实上偶然现象都是客观存在的，偶然现象如个别的看来，好像是没规律的，但是通过大量的观测后，就可以发现其整体的规律性。

§2. 变数统计发展简史

17世纪是资本主义在各先进国家取得胜利和确立的时期，这时商人的需求与机械的发展，促使数学迅速发展，数理统计也正是从这个时期起，由费尔玛和巴斯加开始，逐渐建立起来的。1654年费尔玛和巴斯加则已成立了一些概率论的基础。最初由于保险事业的发展，促使许多人去研究“概率论”的有关问题，但是当时鼓舞伟大数学家思索的专门问题，却来自骰子和纸牌赌博的贵族的要求。1657年，惠更斯发表了“概率论”的第一篇论文——论赌博中的计标（论机会游戏的演算）。雅各·贝努利（1654—1705）也是一个概率论的早期研究者，他的“猜度术”是在1713年他死后才发表的，概率论的最大发展是与法国大数学家和力学卡拉卜拉斯（1749—1827）的名字分不开的，他先后发表了“论事件原因概率”（1774），“概率的解析理论”（1812），“论概率的哲学”（1814）等重要著作，他在自己的经典著作“分析概率论”里，把他的研究成果作了总结，在这本书里，系统地叙述了概率论的基本原理，建立了观察误差的理论及最小二乘法，他并且把这些基本原理应用到“人口统计”方面。高斯（1777—1855）对概率论也有很大贡献，最小二乘法的基础几乎完全是他成立的。但当时勒让德尔（1752—1833）和拉卜拉斯在这方进行了独立的研究，1837年普哇松（1781—1842）著了“关于民刑审判的概率的研究”，在这里他第一次提出了“大数定理”这个名词。俄国的大数学家П. А. 卓比谢夫（1821—1894），李亚普诺夫（1857—1918），马尔科夫（1856—1922）等证明的极限定理，使数理统计学在工业技

术与其他领域中的应用起了很大的作用。近几十年来，数理统计特别是在社会主义的苏联得到更加迅速的发展。苏联学者柯尔哥洛夫，别师傑因，欣斤，罗曼诺夫斯基，格涅坚科，波罗达切夫等对概率论与数理统计学作了很大的贡献。现在“工业技术数理统计学”已成为一门专门的学问，它可用末研究放射性原子的蜕变，赤热金属的电子放射，光电子的放射，方坊中电流的脉动，以及电话，无线电广播，化学的热力学。在工业生产进程中，数理统计的方法也广泛地被用末计标机械，机床及仪器的误差，如今在塑造工业与生产组织中，也都认为数理统计方法是不可缺少的了。

从“数理统计”的理论基础“概率论”开始独立标起，至今会有300多年的历史了，但是数理统计方法被用末研究生物科学的事情，却是在19世纪才开始的。19世纪比利时的数学家兼天文学者 Quetelet (1796—1874) 将生物的或显现象分为系统的次数分配，并且引入常态曲线末说明生物变异中的一些问题，1822—1911年福尔伦色斯·哥洛亭 (Francis Galton) 氏引用“相关”于遗传学方石。1890年科尔洛·皮尔孙 (Karl Pearson) 对生物统计有很大的贡献。他研究的结晶，可以从他所写的生物统计学 (Biometrika) 一书中看到。近四十年末，英人 R. A. Fisher 发现了“重量分析”对生物统计有很大贡献，使近廿年的生物统计 (即变数统计应用在生物学方石) 有很大的发展，尤其是使近廿年来的田间试验突飞猛晋。Fisher 对小样本差异比较与 χ^2 (音开) 比较等也都有一定的贡献，近廿年未对生物统计有贡献的还有 F. Y. Yates 氏，Wisher，Snedecor，Goulden，Limmer，Yule 诸氏。

正当数理统计学日益广泛地被应用在各方石的时候，许多资产阶级的学者过高地估计数理统计的作用，他们把“数理统计”与统计学混为一谈，他们愿证明“统计学”是一门通用的科学，可以研究任何方石的数量规律，企图用“数理统计”末代替统计学。其实“统计学”是一门社会科学，它具有强烈的党性，而“数理统计”则是数学的一个下门，是一门应用数学，它属于自然科学的领域，资产阶级的学者把“数理统计”代替“统计学”的目的，在于掩饰统计学的阶级性，企图掩盖阶级的鸿沟，使劳动人民着不见资产阶级的罪恶统治的必然灭亡的规律，特别是想使无产阶级不去认识自己的历史使命。

在十月革命后的苏联，很快就将统计学提到了它应有的地位，把它当作一门独立的社会科学末看待，革命导师马克思，恩格斯，列宁和斯大林是善于利用统计末研究社会现象的榜样，列宁和斯大林在统计学的理论上有很大的贡献。其实统计学作为一门独立的社会科学是从列宁开始的，但是统计学的研究对象的问题，

就是在苏联也不是很顺利地解决的，在^此一些学者及实际工作者当中，长期存在着一种思想上的混乱，一直到1954年3月16日到26日，通过苏联科学院、苏联中央统计局和苏联高等教育部联合举行的关于“统计学”的科学会议，才给科学的“统计学”的研究对象与方法下了一个正确的定义。参加这次会议的有科学工作者与实际工作者，以及经济学家、工程师、数学家等，事后由苏联科学院院长K. B. 奥斯特洛维季扬诺夫院士作了总结报告（这篇总结报告的原文载于“苏联科学通报”1954年第八期，中文译文载于“统计工作通讯”1954年11月第八期）。总结报告指出：由于主张统计学是通用的科学的人们与主张统计学是一门社会科学的人们之间的斗争的尖锐化，使得某些科学人员忽视了把“数理统计”的方法应用到技术和生产上的问题，其实，只要用得恰当，是应当尽量地加以发愿和鼓励的。

这次科学会议的成就非常巨大，它把“统计学”与“数理统计”联系的科学领域作了明确的划分，批判了过去把“统计学”看作是通用科学而低估了对社会经济问题的统计研究的作用的偏向，以及短视把“数理统计”的方法应用于技术和生产上的偏向。同时，也批判了夸大“数理统计”的作用，企图用“数理统计”来代替“统计学”的偏向，以及低估了“数理统计”应有的作用的偏向。它把这两门科学给予了恰当的地位，澄清了长期以来在这个问题上所存在着混乱思想，这样便使这两门科学在今后将摆脱一切不必要的限制，分别在社会科学和数学的领域内，大踏步前进，并且也帮助广大的统计干部在统一了思想之后，团结起来研究共产主义建设中的迫切问题。

在解放后的中国，也曾经因为批判生物统计方面的资产阶级思想的斗争的尖锐化，使许多人不敢将“数理统计”方法应用到生物科学方面去，自从党提出百花齐放，百家争鸣的方针后，这种不良偏向才开始扭转过来。目前在许多国家中，把数理统计方法应用到生物科学和农业的实践中的范围越来越广泛了，许多苏联的森林学、森林经理学与测树方面的学者，都善于应用数理统计方法研究林业方面的理论与实际工作中的问题，甚至苏联的测树学家认为，如对大量的测树资料不加以数理统计的分析，就会使我们在大堆测树调查资料的数字面前，看不到森林的规律。由此可见，变数统计对于森林调查资料的分析，是起着很重要的作用的。

最后还必须说明，变数统计在我国是一门年轻的科目，几年前以前，我们由苏联的教学计划和一些参考书中，翻译变数统计这个名词时，还认为变数统计即数理统计，最近才肯定，变数统计只是数理统计的一部分，在深度和广度方面，它都远不如数理统计。

§3. 变数统计的内容

上面我们已经讲过，变数统计的研究对象是受多种自发的偶然因素交织影响着的数量，因此，变数统计的内容，就应该包括研究这种数量间的关系与其规律的理论，以及应用这些理论于具体对象的方法。我们是以大量观察的资料着手，然后讨论用来处理偶然现象并体现规律性的数学工具，即概率论，然后，再进一步讨论由 Γ 分资料推断全体的分析方法，研究变数之间的关系及其规律性的方法等。

具体地说，变数统计所研究的问题，所包括的内容，有以下几个方面：

1. 观察资料的整理；
2. 分布曲线的特征数；
3. 概率论；
4. 相关分析；
5. 抽样分析。

上面所列举的5个方面，以后要逐项作详细讲评的，为了避免重复，我们在这里暂不加以详细说明。

最后，还必须指出，变数统计乃是一系列的数学方法，尤其是概率论的原理，佔着很重要的地位，因此，便要求对概率论的基本原理加以注意，概率论是一门较深奥的数学科目，我们只能讨论其中的一些基本原理，因此，所应用的数学基础，仅限于代数和初等微积分，这是我们能了解到的，相信同学们能够全下掌握。

§4. 变数统计在林业上的应用

在林业理论与实际工作中，经常会遇到许多受自发的偶然因素交织影响的数量。在研究和分析这些数量时，便都要应用变数统计。例如：在森林学中，研究森林发育问题，树木的生态问题，以及抚育采伐问题时，需要搜集许多数字资料，如树木的高度，粗度，林木的密度，森林的郁闭度等。在森林利用中，研究木材的强度以及木材利用的途径时，也都要搜集各种数字资料，如木材的年轮数，年轮中夏材百分数，木材的压缩系数，木材的纤维长度，粗度等，在造林学中，研究树木种子的成芽率，造林成活率等问题时，也要搜集果实长度，种子重量，造林株数，成活株

数学数字资料，尤其是在森林经理及测树学中，变数统计的应用更为广泛，因为在这里将要研究和分析许多测量的结果，研究许多数量与数量间的关系和规律性，例如树高，胸径，胸高断面积，材积，形率，形数等数量，所有上述的数字资料，无论是通过观察的方式而获得，或者是通过测量的方式而获得，或者是由实验中得到的，都具有一个共同的特性，即都是受多种自发的偶然因素交织影响而变动的，因此分析和研究这些数字时，以及说明所研究现象的规律性时，就要应用变数统计的方法。

为什么说，这些数量是受多项自发的偶然因素交织影响而变动的呢？我们再举两个例子来说明，例如：某一树种的树木在一定年令时高度并不一定的，因为树高的生长，是与其周围环境，如土壤，气候，水，周围树木等因素有着密切的关系的，这些因素，都在对树高的生长，起着影响作用，这些影响作用，有的使树高的生长较快，有的使树高的生长较慢，因此，有的使树高的生长偏向一方，有的使树高的生长偏向另一方，所以说这些因素对树高的影响作用，是错综交织的，同时，这些因素在起作用时，并不是由人为的力量来控制的，所以说也是自发的。又如实验的结果，也由于实验的仪器，及实验条件等各方因素起交织的影响的缘故，使产生各种不同的结果。而这些因素在起作用时，也都不是人为的力量所能控制的，因此，也是自发的。

由此可见，上面所列举的林业中的现象，都是受多种自发的偶然因素交织影响的数量。因此，在研究和分析这些现象时，便要应用“变数统计”。应用“变数统计”中所说明的偶然现象的规律性，可以对于尚未获知的数量，以足够的精确程度进行推断，这种推断的精确程度和可靠性，完全可以由研究者预先加以规定。

不过，应该特别注意，决不要把“变数统计”方法作过高的评价，因为在研究具体的对象时，还必须注意到对象本身的规律性，如果完全以抽象的数学方法套在任何对象上，只管数量，毫不考虑这些数量所反映的具体对象时，那就走上了资产阶级数学形式主义的道路。数学是需要的，但数学形式主义是要不得的，正如经验是需要的，而经验主义是要不得的一样。

因此，要正确地应用“变数统计”方法于具体对象时，首先要对具体对象作理论分析，也就是说：质量分析要先于数量分析，我们用变数统计来研究林业理论与实践中的问题时，也必须以林业对象的质量分析为基础，以已知的林业理论为基础，才不致作出错误的结论。并且，还可以进一步用变数统计的方法来发展林业理论，指导实际工作。

由此可见，变数统计在林业理论与实际工作中应用很广泛，并且起着非常积极的作用，因此，我们未来的森林工作者，必须掌握这一宝贵的科学武器。

第二章 观察资料的搜集与整理

§1. 观察资料的搜集

变数统计所分析的资料，是根据预定的目的搜集的，这些资料应该具有足够数量和精度，在搜集资料时所产生的错误，是无法在以后的数理分析中消除的。错误的资料往往会我们得出错误的结论，因此，在搜集资料时的粗枝大叶是极端有害的。实际上，这种粗枝大叶的作风，是对祖国社会主义事业不负责任的表现。

资料的来源，是通过表格的方式和组织专门的调查队而获得，或者是由实验而取得。至于这些搜集资料的具体方法，将在各门专业课程中讲成，我们将在下节讨论搜集资料方法的有关数理上的问题：

(一) · 数字观察资料的单位：

变数统计所研究的，是数量变动标志，这些数量变动标志在每一个总体单元上都要具有一定的具体数量，这些数量，必须用一定的单位来表示，通常所采用的单位有下列几种：

(I) · 自然的单位：即客观事物本身具有的单位，如牛，马，猪，羊，以头数计，鸡，鸭，西瓜，以个计，所有这些单位都是事物本身所具有的。我们只是对这些单位加以“个”“头”“天”“株”等作为单位的名称而已，这些便都是自然单位。

(II) · 人为单位：有些单位不是客观事物本身所具有的单位，而是人为地加以规定的。这种单位又可分为连续单位与不连续单位两种：

(1) · 不连续单位：通常系用升数，例如：衣服以“件”为单位，鞋袜以“双”为单位，苗田中的苗枋与林中的样地都以“块”为单位，年令以“令”为单位等。

(2) · 连续的单位：通常以度量衡为单位，例如长度以公尺，公分为单位，重量以公斤，克为单位，面积以平方公尺为单位，体积以立方公尺为单位，时间以时，分，秒为单位，温度以度为

单位，压力以厘米或毫米为单位等。有时还有更复杂的单位，如力，速度，比重，压缩系数，形数。形率等所系取的单位是复合的单位或百分数等（如速度之单位为“每秒厘米”便是复合单位，“形率”便是体积之比值的百分数）。这类单位，都是人为规定的，如客观的数量，用这些单位来表示时，可能有无数位小数，因此称之为连续单位。

在测量和实验中通常遇到的单位都是这类单位，变数统计最常研究的，便是这种单位所表示的数量。

（二）·数字的精确度及有效数字：

测量或实验中所得到的数量，通常都是近似值，例如，我们对某一高度进行测量时，得到的结果为29.3公尺，我们只能认为真实高度在29.25公尺至29.35公尺之间。由于我们测量的仪器只能精确到十分之一公尺的缘故，所以我们只能得到29.3公尺，如果我们使用的仪器可以精确到百分之一公尺时，则我们可能获得29.34公尺，但这仍仅能表示测量的对象的高度在29.335公尺至29.345公尺之间。又例如作某种物理实验时，得到速度为23每秒厘米，也只能说真实的速度在22.5每秒厘米至23.5每秒厘米之间，由此可见通过测量或实验所得的数量常是真实数量的近似值而已。

由于观察资料是大量的，因此，搜集资料时，便往往不是由一个单位进行的，为了使测量或实验的结果具有一致的准确程度起见，便须有一致的了解。因此，应将所求的精确程度预先加以规定。例如，要求精确到百分之一公尺，便表示应在以公尺为单位时，取小数点后的第二位，如要求精确到万哩，便表示应在以哩为单位时，精确到万位数字，个位，十位，百位，千位的数，皆可略而不计，因为在这时，这几位数字已不发生任何效力，故称为无效数字。例如：地球与月球的距离为240,000哩，精确到万哩，则表示真实的距离为235,000哩至245,000哩之间。表示距离的数字，仅为两个数字，其余仅表示位数，在数量的表示上，没有效力，若240,000哩是精确到千哩，则表示其真实数字为239,500~240,500哩之间，有效数字有三个，即240等三个。

登上二哩的高山，不能认为与月球的距离为239,998哩，因为无论精确到万哩或千哩，个位数字皆属于无效数字。

（三）·挑数字的原则

有时，我们所能获得的数字资料比我们所求的精确度更为精确，

这时，为了计标便利起见，可在保证能满足所要求的精确度的原则下，将所获得的数字资料中的有效数字，适当地减少几位。这种处理的过程，就叫做挑数字。

挑数字的原则如下：如所获得的资料比要求的精确程度更精确时，则可按照所要求的精确程度，多取一位数字，如该数字为4及4以下的数字则捨去，如该数字为6及6以上的数字，则将前一位的数字加一，如果该数字为5，则如果前一位数字为奇数时，将前一位数字加一，如果前一位数字是偶数时，则将5捨去，前一位数字不变。例如：所要求的精确程度为十分之一公尺，而所获得的资料为16.14公尺，17.16公尺，18.25公尺，19.35公尺——时，则可挑成16.1公尺，17.2公尺，18.2公尺，19.4公尺等。

我们所以不采取四捨五入的方式，乃是由于四捨五入的方式使捨去与进入的机会不均等，1, 2, 3, 4捨去，5, 6, 7, 8, 9进入，则捨去的机会为 $\frac{4}{9}$ ，进入的机会为 $\frac{5}{9}$ ，因此，容易使资料挑数字之后，较原来的数字为大。

如对于较所要求精确程度更为精确的数字资料进行加，减，乘，除的运算时，应该将所获得的资料挑成比所要求精确程度多一位数字，然后进行加，减，乘，除的运算，最后，将运算结果挑成所要求的精确程度。但在乘除法中，作为乘数和除数的数字通常不变，只对被乘数与被除数进行挑数字（如乘数代表加大几倍，除数代表多少等分时则不挑数字）。这种挑数字的规则是应用于一般的或大量的运算过程中，我们以后的许多运算，就要应用这一原则。如果我们所进行的计标是要求特别精确的少量试验资料，则还应参照误差理论，由于这些理论已超出了我们所讨论的范围，故在这里不讨论它。

§2. 变量数列

(1). 变量数列的概念

我们搜集的大量观察资料，是一堆原始资料，从这些原始材料中，不能看出所研究现象的任何规律性，因此，我们必须将这些材料加以整理。对于数量变动标识的整理，通常是整理成变量数列和相关表。当我们只研究总体单元在某一数量变动标识上的分布情况及各种特性时，我们便把观察资料组成变量数列，当我们研究两个数量变动标识之间的互相关系时，我们便需要组成相关表。关于相关表的组成，我们将在第六章中讲述，现在先讨

论变量数列。

按某一个数量变动标识将总体单元分成若干下分（或组），并按递增标识值（变量）的递增或递减次序排列后，所得的数列就称为变量数列（或称分配数列）。

举举例说明如下：

假定有 69 株树，其树高测量数字资料如下：树高（公尺）

16.6	18.3	18.7	18.8	19.0	19.5	19.6	19.8	19.8	19.9
19.9	20.0	20.3	20.6	20.8	20.8	21.0	21.1	21.2	21.4
21.5	21.5	21.5	21.8	22.1	22.2	22.2	22.3	22.3	22.5
22.5	22.6	22.6	22.6	22.8	23.0	23.1	23.1	23.1	23.5
23.5	23.5	23.6	23.7	23.8	23.9	23.9	23.9	23.9	23.9
24.0	24.5	24.6	24.7	24.8	24.8	24.8	25.0	25.3	25.3
25.9	26.2	26.3	26.4	26.6	26.6	27.1	27.2	27.9	

我们从这些原始资料中，很难看出其中有什么规律性，但是如果我们把总体单元按树高这一标识，分成为若干组，以两公尺为一组，则可得如下数列：

树 高 (公尺)	树 木 (株数)
16 — 18	1
18 — 20	10
20 — 22	13
22 — 24	26
24 — 26	11
26 — 28	8
共 计	69

表(2—1)

经过这样组织了之后，便可以看到这 69 株树分配在 22—24 公尺之间的最多。由此处起，较高的树木较少，较矮的树木也较少，由此可见由 69 株树所组成的总体，树高这方面的水平约在 22—24 公尺之间。

组成这样的数列之后，还可以使我们作进一步的数理分析。这一章，我们将在以下再讨论。

原始材料经过整理之后，各总体单元在该标识上的标识值即已消失。例如，这些数列中，我们只能看到 18—20 公尺的树木有 10 株，至于这 10 株树木各为若干公尺高，则由数列中无法看出。

在这种情况下，我们可用各该组组中数作为该组全部总体单元的变量。于是，最后所获得的材料，便变成如下的形状：

树 高 (公尺)	树 木 (株数)
1 7	1
1 9	1 0
2 1	1 3
2 3	2 6
2 5	1 1
2 7	8
	6 9

表(2-2)

这样的数列便是变量数列。

(2)·变量数列的种类

变量数列可分为分组数列与组距数列两种：

(一)·分组数列

如果按照数量变动标识分组之后，每组内的总体单元在该数量变动标识上的数值完全相同，这样的变量数列，则称为分组数列。例如：下列某林地一公分心材年令数的分配，就是分组数列：

年轮数	株 数
1	1
2	1
3	1
4	2
5	6
6	2
7	5
8	8
9	4
1 0	6
1 1	1
1 2	5
1 3	2
1 4	0
1 5	1
合 计	4 5

(二)·组距数列

如果按照数量变动标识分组并理之后，每组内的总体单元在该数量变动标识上的数值并不完全相同，因此，必须以“由----至----”的形式来表示各该组总体单元标识值的变动范围，并以各该组的组中类作为各该组全下总体单元共同变量时，则所组成的变量数列，便称为组距数列。例如：在上节所举的树高的变量数列中（见表2-1），每组内总体单元在树高这一标识上的数值并不相同，因此，必须以16—18，18—20，20—22----的形式来表示各该组内总体单元之标识值的变动范围，并以17，19，21----作为各该组内全下总体单元的共同变量（见表2-2），所以树高的变量数列，便是组距数列。

由于我们在作进一步的分析时，对于分组数列和组距数列，是采取不同的方式，因此，我们应将它们加以区别。

(三)·有关变量数列的几个概念

今后，我们所研究的，主要是数量变动标识^的，而且为了寻出其中所存在的规律性起见，必须将数量变动标识上的大量材料组成变量数列，以便作进一步的研究。因此，先将几个有关变量数列的概念说明如下：

(一)·变量：这个概念我们曾经在前节说过，即各总体单元在数量变动标识上的具体数值，不过，当组成组距数列之后，则各总体单元的变量，便应用各该组的组中类的数值来表示（见表2-1与表2-2），在分组数列中，仍维持各总体单元原有的变量（见表2-3）

(二)·次数（或称频数）：即为相当于各该变量的总体单元的数目。例如：在树高数列中（见表2-2），变量17公尺的次数为一，变量为19公尺的次数为10，变量为21，23，25，27的相应次数分别为13，26，11，8等，在年轮数的变量数列中，变量1，2，3，4，5----等的相应次数各为1，1，1，2，6----等。

在变量数列中，次数可用各组内总体单元的个数的绝对数来表示，称之为绝对次数。也可用各组内总体单元数对于全下总体单元数之比来表示，称之为相对次数。

例如：上述的树高变量数列如用相对次数表示时，则成为下列形状（参看表2-4）。

变量 (公尺)	相对次数
17	0.015
19	0.145
21	0.188
23	0.377
25	0.159
27	0.116
共 计	1.000

其中，变量17的绝对次数为1，故其相对次数为 $\frac{1}{69} = 0.015$ ，其余依此类推。

变量数列，实际上便是由变量与次数两个因素组成，所以称它为变量数列的两个要素。

(3)·组限：在组距数列中，每一组中的总体单元在数量变动标尺上的具体数值并不完全一致，而是在一定范围内变动的。这种变动的限度称为组限。在组变动的最小限度称为下限，最大限度称为上限。例如表2-1所举的树高数列中，变量17表示该组的总体单元在树高这个标尺上的变动范围为16—18公尺，这便是该组的两个组限，其中16为下限，18为上限。

(4)·组中值：在组距数列中，每组的中点，所具有的数值便是组中值。显然，每组的组中值便等于各该组的上限与下限的和被二除，所得的结果。例如在表2-1的树高的数列中，16—18公尺之一组的组中值便是 $\frac{16+18}{2} = 17$ ，在组距数列中，我们可以用组中值代替各该组所有各总体单元的实际数量，乃是以各组内的变量是均匀地分布于各该组的假定为基础的，因此，组中值可以看作是各该组在某一标尺上的平均数。

(5)·组距：组与组之间的距离称为组距。在分组数列中，组距就是相邻两组数量之差，在组距数列中，组距便是相邻二组上限或下限之差。例如在表2-3的分组数列中各组的组距均为1，在表2-1的组距数列中，各组距皆为二公尺。应当指出，组距不一定是完全相等的。

(4)·变量数列的编制

编制变量数列，应快考虑到两项原则，第一，在理论上，在

依所编的变量数列能正确地反映具体对象所存在的实际规律。第二，在技术上应使每一个总体单元在该标识上所表现的数值都能而且仅能分在一个组中。也就是说：既不要使任一数值无组可分，也不致使同一数值分在两个组中，现在分别来讨论这两个原则。

第一，欲使变量数列能正确地反映出具体对象所存在着的实际规律，首先须对对象作理论的分析。例如，苏联特列买雅可夫教授在求树木的平均直径时，认为必须使最大的直径与最小的直径的比值不超过4。如最低一组为12公分，则最高一组不应超过48公分。求平均树高时，最大的树高与最小树高的比值不应超过3。如最小一组为10公尺，最大一组则不应超过30公尺。如果数列的最大值与最小值的差异太大，则最好分为两个数列来研究，否则平均数就不可靠。

这些系数，并没有数理上的根据，但却具有实际的意义。因此，在研究森林对象时，正确编制变量数列，还必须以林业的理论为基础。

这个问题，因为涉及到林业理论的基础，故不作进一步的讨论，但应将这一原则作为重要原则提示。

第二：在技术方面的考虑，如果我们所研究的数量变动标识是用不连续单位所表示的数量，则可以编制分组数列，也可以编制组距数列，但所研究的数量变动标识是用连续单位所表示的数量时，则只能编制组距数列，而不能编制分组数列。因为这对各总体单元的标识值是连续的，只能把一定范围内的变量作为一组，如编制分组数列，便将使某些总体单元无组可分。

在编制组距数列时有两种方法，第一种是上限排斥法，上节所举的树高数列（表2-1），便是按照上限排斥法编制的。这时恰巧等于上限的数量的总体单元，之排斥于该组之外，亦即应分至上邻组中。因为这种数列中，某一组的上限与其上邻组的下限具有相同的数量，如果不加以统一的规定，则可能使同一数值，有时分到某一组中，有时又分到另一组中，则同一材料便可能作出不同的结论。因此，应该有统一的了解。

第二种方法是间隔法：这时任一组的上限皆与其上邻组的下限间存在一定的间隔，并不具有相同的数量，这种编制变量数列的方法，便是间隔法。例如：下列果园亩数亩株数的变量数列，便是用间隔法编制的组距数列（见表2-5）。

云 苗 株 数	苗 床 数
0 — 99	1
100 — 199	2
200 — 299	5
300 — 399	10
400 — 499	14
500 — 599	12
600 — 699	8
700 — 799	3
	55

表(2-5)

如果所研究的数量变动标识是用连续的单位所表示的数量，则应该用上限排除法编制变量数列，如果所研究的数量变动标识是不连续的单位所表示的数量，则应该用间隔法编制变量数列。

变量数列的编制是一件比较困难的工作，如应该选择分组数列呢？还是应该选择组距数列呢？如选择组距数列，组距应多大？如何规定组距和组限等。有时由于研究对象本身所具有规律性，还需使各组的组距不尽相等，亦即还需编制异距数列。当然，等距数列在计算时方便得多，但不可为了计算方便而忽视现象本身的规律性，现在编制变量数列最一般的科学方法，还没有被发现，因此，我们主要是依靠林业知识和经验来解决上述的问题。

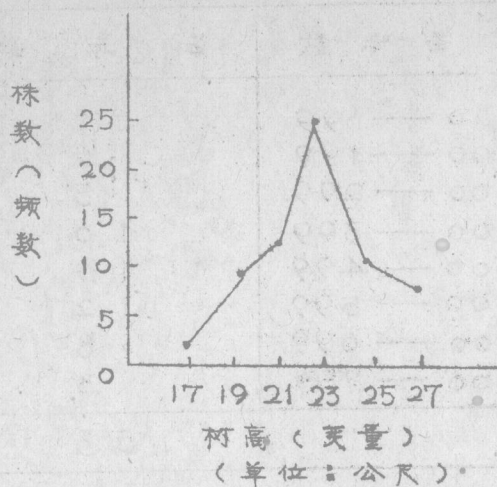
(5) 变量数列的图示法

变量数列也可以用图形表示，这样，可以使具体对象在这个数量变动标识上的特性更加明显地表现出来。

由于变量数列可以把总体单元在各个变量上的分配情况表示出来的原故，所以变量数列又称为分配数列。而把变量数列或分配数列绘成图形之后，这种图形便称为次数分配图或频数分配图。

频数分配图的编制，是以直角坐标的横轴表示变量，以纵轴表示各变量相应的次数。每一变量皆有相应的次数，所以根据每一变量与其相应的次数的数字，便可在直角坐标上，决定其点的位置，用这样的方式，可以求出许多点，将这些点连接起来后，所得的图形便是频数分配图。

例如：上述树高的变量数列（见表2-2）便可绘成如下的次数（频数）分配图：



在这个图形中，变量数列表现为一折线，如果总次数（总体单元总数）很大，而组距很小，便可获得一个平滑的曲线，实际上，我们所研究的变量既属于大量现象，则其分配规律常可用曲线表示。

在以下的分析当中，为了便于了解起见，我们将常利用这种频数分配曲线图形来说明问题。

变量数列还可以用方程式来表示，这时，变量表示自变数，次数表示依变数。

§3. 累积次数数列

有时，我们想了解在某种变量以下之总体单元数目，这时，我们便可组成累积次数数列。例如，上述树高的变量数列，如欲组成累积次数数列，可按照下列方式进行：

表(2-6)

树高分组 (单位:公尺)	次 数	树高小于, (公尺)	累积次数
16 — 18	1	18	1
18 — 20	10	20	11
20 — 22	13	22	24
22 — 24	26	24	50
24 — 26	11	26	61
26 — 28	8	28	69
合 计	69	—	—

由累积次数数列中，便可很清楚地看到，树高在18公尺以下者一株，在20公尺以下者11株，在22公尺以下者24株-----等。