数学地质教程·数学地质教程

李公时　谢国柱　编

中南工业大学出版社

数 学 地 质 教 程

李公时 谢国柱 编

中南工业大学出版社

数学地质教程

李公时　谢圣柱　编

责任编辑：陈五娘
插图责任编辑：刘楷英

*

中南工业大学出版社出版发行
中南工业大学出版社印刷厂印装
湖南省新华书店经销

*

开本：787×1092 1/16　印张：19　字数：486千字
1989年8月第1版　1989年8月第1次印刷
印数：0001—3500

*

ISBN 7-81020-251-0/P·008
定价：3.15元

前　言

数学地质成为地质科学中的一门边缘学科，在国内外发展很快，目前它已经渗透到地质学的各个领域。广大地质工作者时刻感到当代科学技术的飞跃发展，渴望学习和使用数学地质方法，利用电子计算机来为他们作繁锁的处理数据、绘制图件和研究工作。几乎所有院校的地质专业都开设了这门课程来普及数学地质方法。尽管已有许多著作问世，但理想的教材仍然有待于编写，我们希望今后能陆续见到一些好的教材。

我们从1978年开始在地质专业为本科生和硕士生开设数学地质课程，在总结教学经验，不断修改原用教材的基础上，吸取其它教科书的优点，编写了《数学地质教程》。全书共分三篇：应用数理统计篇分五章，比较系统地介绍了数理统计的基本概念、原理和方法，各章配有大量按地质学内容精编的例题和习题。多元统计分析方法篇分七章，详细地介绍了方差、回归、趋势、判别、聚类、因子等分析方法，以及对应分析、典型相关分析、非线性映射分析和最优分割方法，也简略地介绍了马特龙地质统计学，马尔可夫随机过程，杨赤中矿床统计分析以及其它分析方法。各主要方法自成一章，各章配有地质应用举例和10个以上从不同角度理解、训练或拓展介绍内容的习题。数学地质方法篇分三章，比较详细地介绍了地质变量的确定及数值变换方法，地质数据库和数学地质的计算机处理方法。书末附有六个实验的指导书，可根据课时及教学对象的水平层次并行地安排在讲授第二篇或第三篇的内容中。

本书按120学时数编写，适合地质专业本科学生使用，教学时各篇可按32、64、24学时分配学时数。编者希望把本书写成一本便于教学也便于自学的教材。但读者须具备高等数学、线性代数和概率论基本知识。

本书第一篇和第二篇的第一、二、三、五章，以及实验指导书由谢国柱编写，第二篇的四、六、七章和第三篇由李公时编写。在编写过程中曾得到我校陈国珑教授、杨善慈教授的热情支持和帮助，樊钟衡同志也对本书的内容提出很好的建议，谨此一并致谢。

由于作者水平有限，错误之处在所难免。恳请读者批评指正。

编　者

1988年10月

绪　论

地质学是一门古老的自然科学，它起源于人类对自己生活的土地的认识，发展在人们对自然存在规律的不断了解和加深认识之中。最近三十多年来，由于人们在其它领域内取得的成果，使地质学不断得到新的研究方法，形成新的边缘学科，如有机化学地质学、地震地质学、遥感地质学等等，丰富了地质学。六十年代，由于电子计算机的飞速发展和应用使地质学得以与数学广泛地相结合，并逐渐地形成数学地质这一新的边缘学科，使地质学从定性分析的时代逐渐走向定量分析的时代。

数学地质，从广义上说是一门用数学的方法研究地球历史和发展规律、从理论和实践上解决地质领域内各种问题的应用性学科。它是地质学与数学以及电子计算机相结合的产物。

数学地质的形成经历了一百多年的历史。最早使用统计分析方法的是在1840年，英国学者莱伊尔（Charles Lyell）首次用统计分析方法对第三纪地层进行划分。其后又有许多学者陆续地在其它方面使用统计分析方法。

1940年至1960年是电子计算机发展的年代，数学地质由于有了能够处理大量数据的可能性而比较活跃，数学方法越来越多地应用于地质学，并从单变量分析发展到多变量分析。

六十年代是数学地质迅速发展的年代，由于电子计算机性能的提高，数学方法和电子计算机开始广泛地应用于地质学。仅1968年至1970年有关数学地质的论文就超过了三千篇，并于1968年在巴黎召开的第二十三届国际地质学会上成立了国际数学地质协会（IAMG），开始出版国际数学地质协会杂志，出版地质计算程序公报等。在这一阶段，由于多元统计分析在地质学中的大量应用，数学地质发展成为一门独立的学科。

我国的数学地质工作开始于六十年代初期，当时只是用个别方法处理一些地层数据，七十年代才扩展到地质学的其它领域。目前，数学地质在地质工作中已得到广泛应用。

数学地质作为数学方法研究地质学基础理论解决地质学中实际问题的地质学分支，目的是定量地研究和解决地质科学中用传统的定性描述和成因推理方法难以解决、或者争论不休的许多模糊不清的理论和实际问题。地质工作中人们经常要收集成堆的资料处理上万的用以表征地质现象的数据，这是凭个人的经验、理解和记忆所不能完全信任的，人们只能凭经验对研究对象作出判断。为了证实自己的推断，人们不惜代价地去收集认为有益的资料。结果，有的得到正确的结论，有的则只能得到违反客观事实的结论。而且，在当今科学还不能作出最完美的精确裁断时，这些错误并不能被人们所认识且继续引导人们去付出代价。数学地质希望用数学的严密性，通过科学的整理、演算和检验方法为所有的地质工作者提供地质解释、分类和推断的共同基础。在这样一个基础上，每一个人可以去考虑自己的思路和方法是否正确，并指导人们根据自己的经验和知识结合研究对象的客观实质去作他们的解释工作。数学地质还不是最完美的方法，也不是万能的方法，更不能取代地质领域内别的研究方法。在地质工作中，在某种意义下，它只是作为一个工具为人们提供进行研究的途径和手段，并把地质工作汇聚到一条逐渐走向正确的道路上。数学地质也象其它现代地质学分支一样，它能否使你得到正确的结论，完全取决于用以计算的原始信息的来源是否正确，不去认真地考

查研究对象、正确地去获取信息数据，而想用数学地质方法取得正确的结论是不可能的，这是每一个从事数学地质的工作者所必须时刻牢记的。

数学地质方法的应用范围十分广泛，几乎渗透到地质学的各个领域。但是数学地质还不能算是一门成熟的学科，摆在数学地质工作者面前的任务还是十分艰巨的。数学地质工作今后的主要任务仍然是加强数学地质的基础理论和方法研究，重点在下面几个方面：

一、继续加强基础理论，包括多元统计分析各种方法的基本原理以及应用条件的研究。

二、继续引进用以解决地质领域内某些特殊问题的数学方法，建立相应的数学模型。

三、建立各种级别的地质数据库和计算机网络，并使之标准化、公共化。

四、改进现有软件的性能，建立多功能综合计算机软件服务系统，并使之智能化。

五、根据数学地质的特点建立健全新的地质工作方法体系。

“科学仅当它成功地利用数学时才达到完善的程度”。数学地质的产生是生产力不断发展、科学技术不断进步的必然结果，可以期望随着数学地质的进一步发展和完善，将给地质科学带来一个新的飞跃，数学地质本身也有着光辉的前景。

目　　录

第三篇　数学地质方法

第一篇　实用数理统计

数理统计是一门以概率论作为基础理论的数学学科。主要研究怎样合理地搜集、整理、表现和分析资料，并根据这些资料推断整体，指导人们合理地决策。

数理统计是一门应用性很强的学科，已被广泛而深入地应用到工程技术、自然科学和社会科学等各个领域。

数理统计是数学地质基础理论的组成部分。本篇侧重介绍实用数理统计内容。

1　样本与抽样方法

本章主要介绍数理统计中的一些基本术语和基本概念，如母体、子样、子样数字特征等，以及应用数理统计时比较突出的问题——抽样方法。

1.1　母体与子样

1.1.1　母体及其分布

母体　亦称**总体**。是指我们准备加以测量的一个满足指定条件的元素或个体的集合。

例 1　某沉积铁矿共有 1000 个钻孔铁矿层厚度数据，我们要研究铁矿层在 1000 个钻孔中的厚度变化。则 1000 个层厚数据构成一个母体，每个铁矿层厚度数据是个体。

例 2　研究某矿体中 Cu 的品位变化。则整个矿体构成一个母体，矿体的各单位体重中 Cu 含量的重量百分比是个体。

母体可以是有限的。如例 1，母体是 1000 个铁矿层厚度数据。当构成母体的个体为有限个时，称该母体为**有限母体**。

母体也可以是无限的。如例 2，不可能将整个矿体都采样来进行化验取值，特别是当矿体很大时，我们可以认为它是无限的。当构成母体的个体有无限多个时，称该母体为**无限母体**。

母体中的元素常常不是指元素本身，而是指元素的某种数量指标。如例 1 中，母体的元素是指铁矿层的厚度数，常以多少来表示。例 2 中，母体的元素是指Cu的品位数，常以百分数表示。同一母体中元素的取值可以是相同的，也可以是不同的。如1000个铁矿层厚度数据中，1 米厚的有 101 个，2 米厚的有 710 个，3 米厚的有 189 个，不同厚度所占的比率为 $\frac{101}{1000}$，$\frac{710}{1000}$，$\frac{189}{1000}$。研究母体的分布特征时，这些数量指标取不同数值比率的分布称为**母体分布**。

记母体的数量指标为X。从母体中随意地取得的一个个体称为**随机变量**，记为X。显然，随机变量X所有可能取得的数值就是X可能取的不同数值的全体，随机变量X的概率分布列就是X的母体分布。以例1为例，其概率分布列和母体分布为

X	1（米）	2（米）	3（米）
P	$\frac{101}{1000}$	$\frac{710}{1000}$	$\frac{189}{1000}$

母体的数量指标X简记为母体X。母体X的分布可用分布列、分布密度和分布函数具体表示出来。母体的数字特征也可以用母体的平均数（亦称数学期望）、方差和标准差等表示。其记号与概率论中随机变量的相应量的符号同。如分布列记为$P(x)$、分布函数记为$F(x)$、分布密度为$f(x)$。数学期望、方差和标准差分别记为EX、DX和$\sqrt{DX}$或$\sigma[X]$。

1.1.2 子样

子样 亦称**样本**。是指从母体中抽取的一部分个体（x_1，x_2，$\cdots x_n$）。

例3 某矿区在决定合理的采样间距时，沿一穿脉坑道采集100个样品做实验。这100个样品就构成一个分析确定采样间距的子样（x_1，x_2，$\cdots$，x_{100}）。

取得子样的过程称为**抽样**。子样中的每一个个体称为**样品**。子样中个体的个数称为**子样容量**，如例3，该子样的容量为100。子样容量是有限的。

在数理统计中，采用的抽样方法是**随机抽样法**，即子样中每一个个体（样品）是从母体中被任意地取得的。随机抽样分**重复抽样**与**非重复抽样**两种。以例1为例，从1000个厚度数据中抽取一个容量为10的子样。先将这1000个数据列表，如随机地查看某栏内一个厚度数据，记录后从表中划去该数据再随机地查看下一个，直至取得10个数据为止，则这种方法称非重复抽样；如查看某栏内一个厚度数据，记录后并不从表内去掉该数据，又继续随机地查看直至取得10个数据为止，这种方法称为重复抽样。

从母体X随机抽样得到的子样可以用n维随机变量（X_1，X_2，$\cdots$，X_n）表示。现在考察它的概率分布。在重复抽样的情况下，由于每抽取一个个体检查后要放回，母体不发生变化，所以X_1，X_2，$\cdots$，X_n是独立同分布的，每一个随机变量的分布与母体分布相同。对于非重复抽样，在有限母体中抽样时，因每抽取一个个体后，由于不放回母体改变了成分，所以随机变量X_1，X_2，$\cdots$，X_n不相互独立；在无限母体中抽样，每抽取一个个体后并不改变母体的成份，所以随机变量仍然是独立同分布的，并且每一个随机变量的概率分布都是母体的分布。

子样（X_1，X_2，$\cdots$，X_n）是n维随机变量，这是对具体进行一次抽样前而言，在抽样后获得它的一组观察值（x_1，x_2，$\cdots$，x_n），称为**子样值**。为方便起见，子样与子样值常常不加区别统称为子样。

设母体X的分布函数是$F(x)$，则子样（x_1，x_2，$\cdots$，x_n）的概率分布函数

$$F_n(x_1, x_2, \cdots, x_n) = F(x_1)F(x_2)\cdots F(x_n)$$

在母体离散分布的情况下，设母体的分布列为$P(x_j) = P\{X = x_j\}$，$j = 1, 2, \cdots$，则子样的概率分布列

$$\begin{aligned} P_n(x_1, x_2, \cdots, x_n) &= P\{X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n\} \\ &= P(x_1)P(x_2)\cdots P(x_n) \end{aligned}$$

在母体连续分布的情况下，设母体的分布密度为$f(x)$，则子样的概率密度为

$$f_n(x_1, x_2, \cdots, x_n) = f(x_1)f(x_2)\cdots f(x_n)$$

子样的确定的分布，下一章将详细讨论。

1.2 实用抽样方法

1.2.1 抽样原则

用抽样的方法获取研究母体数学特征的资料，其目的在于以最小的代价获得能精确地代表母体的部分资料。因此，抽样方法的选择必须符合**省时、省事、省钱**和取得的子样**具有代表性**的原则。

例1 某矿山有5个开采中段，在开采过程中为进一步研究确定矿体矿化的连续性，需取样做统计分析。该矿山已有各中段沿脉坑道中以2米间距所取的190,500个拣块样的化验结果，现在怎样设计抽样方案。

方案一 按圈定的矿体形态将矿体等分块后，各块重新取样化验获得子样。

方案二 取所有现成的采样化验结果作为子样。

方案三 选取1、3、5中段上若干具有代表性的沿脉坑道中的采样化验结果作为子样。

方案四 根据计算机能处理的数据容量N，产生3组随机数共$3N$个作为理想坐标确定取样点，再在190,500个化验结果中选取N个位置与之靠近的作为子样。

方案一有好的代表性，但不符合省时、省事、省钱的原则，事实上也不可能。方案二虽然省时、省事、省钱，但代表性差且没有考虑计算机能否处理得了。方案三比前二个方案优越，但掺进了人为的干扰因素，也不能说理想。方案四基本符合省时、省事、省钱和具有代表性的原则，在本例中是较好的抽样方法。

抽样，首先必须考虑代表性原则，然后再考虑省时、省事、省钱原则来决定抽样方法和子样容量。

子样的可靠性保证要求采用的抽样方法和子样容量必须与母体的数学特征相适应。分布均匀的母体抽样方法可自由选择，子样容量也可以少些。反之，抽样方法必须考虑母体的数学特征，子样的容量也要求大些。

子样容量的确定将在参数估计章节中详细介绍。现在，我们已有一个初步的印象，就是样本的容量并不是随便地确定的，必须考虑多种因素，简单地用大于某个数，如20或30来划分大小样本是欠妥的。

总之，抽样是一项基础的复杂的工作，在可能的条件下，抽样点的分布应相对地均匀些，子样容量也尽可能大些，子样的可靠性就大些。

1.2.2 常用抽样方法

抽样的方法多种多样，这些方法可分为两大类，即随机抽样方法和非随机抽样方法。随机抽样方法包括简单随机抽样、分层随机抽样、系统随机抽样、多段抽样和多相抽样等。非随机抽样包括定额抽样、控制抽样、判断抽样和自由抽样等。下面分别介绍这些方法。

1.2.2.1 随机抽样方法

无论抽样的方法如何，凡能使母体的每一个个体均有同等被选中机会者，都属随机抽样。均等机会的实现有三个条件：（一）各个体取得的概率必须相同；即当一个体被选取后

其它个体被选取的概率不变。（二）各个体被选的可能性必须均等，抽样过程无外界因素的影响。如人为的干扰等。（三）各个体彼此独立，互不影响；即一个体被选后不会导致其它个体带入或不能被选。设计随机抽样方法时应充分考虑这三个条件。随机抽样方法比较客观地取得母体的代表子样。是现代统计学最重要的抽样方法。

（1）简单随机抽样法。

抽样方法使得母体的各个体被选机会均等，且依子样大小的各组合被选机会也均等时称之为简单随机抽样法。

例2 某沉积盆地已按等间距网格取得某地层的1000个样品分析内碎屑不同粒径百分比含量数据，拟取100个样品数据研究沉积物的来源区。按简单随机抽样方法，首先将1000个原始数据从1至1000编码，然后产生一组其数值在1至1000内变化的随机数，最后任意地在随机数表中取一数为起点，顺序取100个数，编码与这些数对应的原始数据即为所抽得的简单随机子样。

例2中随机数表的产生是任意的，取得的100个随机数也是任意地抽得的。因此，母体的各个体具有相同的被选机会，其概率为$\frac{100}{1000}$。而且当将样品按100个组合时，1000个样品按每100个组成一组的所有可能组合，也均有被选的同等概率。

（2）分层随机抽样法

将母体依某种标准划分为若干层次（或类别），再从各层分别随机抽样构成子样的抽样方法称为分层随机抽样法。

续例2，如地层中包括泥岩、砂岩和灰岩层，则上述1000个样品的内碎屑不同粒径百分比含量数据是由泥岩中的、砂岩中的、灰岩中的内碎屑不同粒径百分比含量组成的，混合起来考虑并不能很好地表现沉积盆地的沉积环境特征，找到沉积盆地的蚀源区。现分别按泥岩层、砂岩层和灰岩层来考虑。如先研究砂岩层的内碎屑不同粒径百分比含量数据，再研究泥岩层和灰岩层的，最后综合起来分析，就可以比较准确地了解到各岩性时期的沉积环境特征，分析效果就好得多。

分层随机抽样是当对母体的某一部分特别感兴趣，或分类调查更能表现全体的某种特征；或由分类调查可得到均一性较高的分层母体时所采用的随机抽样方法。当母体离散程度较大时应考虑采用分层抽样方法，将母体分为若干类，同类的个体相似性则较大。另外，分层抽样可提高抽样的效率。因为分层子样可以以较少的子样容量获得必要精度的估计量。

分层抽样须在抽样前将母体划分为若干层，各分层界线必须分明。反之，则不应分层。分层后从各层抽取的子样构成**再抽子样**。各再抽子样构成母体样本。

采用分层抽样法当各分层再抽子样需联合构成母体样本时，各分层抽样的容量可由以下方法确定：

①**同数定分**；设n_1，n_2，…为各层子样容量，当$n_1=n_2=\cdots$时为同数定分。即不论各层情况如何，一律抽取容量相等的子样。

②比例定分；设$f=\frac{n}{N}$为母体的抽样分数，$f_1=\frac{n_1}{N_1}$，$f_2=\frac{n_2}{N_2}$，…为各分层的抽样分数，当$f_1=f_2=\cdots$时为**比例定分**。即各层子样容量n_i依各层的大小N_i而定，分层大，子样也大。

③标准差比例定分；设各层的标准差为σ_1，σ_2，…，当$\frac{n_1}{\sigma_1}=\frac{n_2}{\sigma_2}=\cdots$时为**标准差比例定分**。

例3 若从某母体分三层共抽取容量为300个的子样，已知三分层的标准差为$\sigma_1=5$，$\sigma_2=10$，$\sigma_3=15$，则按标准差比例定分的各分层再抽子样容量为

$$n_1=\frac{\sigma_1}{\sigma_1+\sigma_2+\sigma_3}\times n=\frac{5}{30}\times 300=50$$

$$n_2=\frac{\sigma_2}{\sigma_1+\sigma_2+\sigma_3}\times n=\frac{10}{30}\times 300=100$$

$$n_3=\frac{\sigma_3}{\sigma_1+\sigma_2+\sigma_3}\times n=\frac{15}{30}\times 300=150$$

用标准差比例定分，可先作小规模试验抽样估计出各层的标准差。由于该法没有考虑分层大小的差别，采用此法时应尽可能使各分层大小相同。

④**最佳定分**；当$\frac{n_1}{N_1\sigma_1}=\frac{n_2}{N_2\sigma_2}=\cdots$时为**最佳定分**。最佳定分同时考虑了按分层大小比例定分和按分层标准差比例定分的原则，所求得的n_i可同时与N_i和σ_i成比例。

（3）系统随机抽样法

按母体中个体的连续序列，每隔一定间隔选取一个样品的抽样方法称为**系统随机抽样法**。

例4 某母体由1000个个体构成，现取容量为100的子样，先将母体的1000个个体顺序排号，再每隔10号抽取一个样品，抽样的起点任意确定。

系统随机抽样设计和抽样过程都很简单，当母体无周期性现象时应用此法，效果较好。当母体具周期性变化特性时，用此法抽样各周期内应有充分多的样品。否则，子样不具代表性。

例5 母体连续分布时，间隔取样称为**离散化抽样**。依据奈奎斯特取样定理，当取样频率等于或大于信号所含最高频率的两倍时，即取样间隔等于或少于信号中最高频率成分的半周期时，抽样过程保存了信号中的全部信息。

（4）多段随机抽样法

抽样按阶段进行，先抽取个体的集合体作为子样个体，再在子样个体中抽取基本个体构成子样的抽样方法称为**多段随机抽样法**。

例6 某地在进行成矿远景区定量预测时，先将整个研究区划分为预测基本单元，并抽取若干单元做为控制区，再在被抽取的基本单元内抽取地质变量研究矿产的分布规律。

多段抽样广泛应用于母体基本单位可按区域划分的资料。抽样时，第一阶段只须根据区域目录抽取子样，然后再对区域子样包含的基本单位分别进行抽样调查。由于各阶段抽样均用随机抽样法，该方法符合随机抽样原则。

（5） 多相随机抽样法

从不同的角度分别抽样研究同一对象的抽样方法称之为**多相随机抽样法**。

续例2 盆地内地层的1000个样品的内碎屑，某粒径百分比含量数据已经抽取容量为

100的子样分析沉积物的来源区。现重新从中抽取容量为500个的子样研究沉积区的环境特征。或者重新检测这500个样品的其它指标，如生物化石含量、种类、以及氧化物的百分比含量等，再来研究该沉积区的沉积环境特征。于是，研究同一沉积盆地对这1000个样品的抽样构成了多相随机抽样。

多相随机抽样，可得到研究对象从不同角度观察的数学特征，互相佐证，使研究更加深入可靠。

1.2.2 非随机抽样方法

非随机抽样是指抽样时母体X各个体x_i被抽取的机会受到限制，不能应用概率法则计算子样统计测定数的可靠性。由于非随机抽样具有一定寓意，仅选择若干典型个体作为代表，抽取的子样称为**计划子样**。一般来说，计划子样的代表性较差。基于某种特定的场合和需要，非随机抽样也经常在大规模调查时使用。下面介绍几种常用的非随机抽样方法。

（1）定额抽样法

当母体的离散程度较高，适用分层抽样方法抽样时，抽样过程中抽样者可视可能性和方便的情况决定取舍，但必须保持各层间样品数额的比例定额的抽样方法称为**定额抽样法**。

例7 某地出露的岩浆岩包括花岗岩50%、石英斑岩30%、玄武岩20%，现从三种岩体中抽样100个研究某元素在岩浆岩中的赋存规律。

方案一 由于该地区岩浆岩有三种类型，按分层抽样的方法设计抽样并限额花岗岩类样品数必须占子样容量的50%，石英斑岩类约占30%，玄武岩类必须占20%。采样时需随时注意保持定额比例。如已采得花岗岩样50个则不必再采集花岗岩样。

方案二 抽样者无需详细准备目录，设计随机数表，确定抽样个体，再取得样品。可根据时间安排和地区分布情况便利地自由抽取，但必须保持各类样品的定额比例，如石英斑岩类只采30个，玄武岩类只采20个，其它都采花岗岩样。

方案一采用分层抽样方法，只限定了各类样品的定额比例，与分层随机抽样差别不大。方案二采样地点和路线是根据便利工作确定的，虽经济省时，但代表性差。二者都属定额抽样，这在实际中是常用的抽样方法。

定额抽样其子样的代表性在很大程度上取决于抽样者对抽样原理、抽样方法和母体特性的了解程度。有经验的抽样者可用此法方便地得到较好的子样，反之，子样没有什么意义。

（2）控制抽样法

抽样时先估计出某项指标在单位范围内的平均数，然后确定一个该项指标等于平均数的单位范围作为抽样范围，再在此被控制的范围内抽取子样的抽样方法称为**控制抽样法**。

例8 已知某区域内平均每100条石英脉中有10条含矿，现研究含钨石英脉的分布情况。在无控制的情况下，有可能全部在矿区内抽得子样，也可能全部在贫矿区抽得子样，结果是很不相同的。为得到具有代表性的子样现采用控制抽样法抽样。

先以是否含矿作为控制指标随机地抽取100条石英脉，如其中含矿石英脉多于或少于10条，则需重新抽取。若该100条石英脉中刚好有10条含矿，则再从此100条脉中抽取研究含钨石英脉分布情况的子样。

控制抽样法是避免其它客观因素影响的一种常用抽样方法。就现状而论，由于地质工作的主要目的是找寻矿产资源，在富矿地区特别是矿床密集区或矿区范围内积累的原始资料就

比较多，人们抽样时很自然地在这些地区抽取较多的样品。于是，分析的结果并不能客观地反映区域母体的特征。特别是在完全利用现有资料进行分析时，利用控制抽样法抽样往往能得到具有较好代表性的母体子样。

（3）判断抽样法

基于对调查对象掌握的情况，有选择地抽取子样的方法称为**判断抽样法**。

例9 沿某河道做重砂取样研究某砂矿物的分布，由于财力和人力所限，只能抽少量样品来进行研究，问如何设计抽样。

显然，不能按随机抽样方法在整个河道范围系统地布置抽样。于是，根据重砂富集规律，分别从各支流入河口及若干典型的滩头布置采集砂样。

例10 某地计划大规模地进行矿产普查，为使普查工作能取得较好效果，先选择一小片边远地质空白区做普查抽样试验，从中摸索经验，训练人员，指导整个普查。

例9由于有重砂分布规律的经验，有选择地抽取小量样品便达到了目的。这是判断抽样法的一大优点。例10由于选择了最困难的地区作指导性抽样试验，对全面展开普查工作有重要意义。

判断抽样基于对研究对象须具有一定的知识，因此，必须由有专业知识的人来做。

（4）自由抽样法

无约束的非随机抽样称为**自由抽样**。抽样时不考虑定额比例、控制条件，也不考虑随机性要求。

例11 某人在研究钾在岩浆作用中的地球化学行为时，从若干岩体的分析结果抽出一些来组成一个子样。

由于没有足够的资金和时间大量地收集资料、系统地设计和实施抽样，该子样的代表性是很差的，但在作为初步了解这一特定场合，它仍属正常的子样。

自由抽样法取得的子样只能对母体做大致的估计，当人力物力允许时，应考虑用其他的抽样方法。不得不用此法时，也应尽可能地谋求子样的代表性。

抽样是数理统计的基础工作也是最重要的工作。同样，也是数学地质的最重要的基础工作。子样的可靠性直接关系到研究结果的正确与否，这就是本书以较大篇幅介绍抽样方法的原因。

1.3 子样的数字特征

数字特征 凡反映某批数据的主要状况的数字都称为**统计特征数**，简称**特征数**或**数字特征**。子样的数字特征是刻划子样中数据的某种特征的指标。主要有两大类：一类表示数据的集中位置，包括**平均数**、**中位数**和**众数**等；一类表示数据的离散程度，包括**方差**、**均方差**和**极差**等。子样平均数和方差是最重要的特征数。

1.3.1 子样的平均数和子样方差

由子样值 $(x_1, x_2, \cdots, x_n)$ 可定义

子样平均（数） $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

子样方差 $S^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2$

子样k阶原点矩　$a_k=\frac{1}{n}\sum_{i=1}^{n}x_i^k$

子样k阶中心矩　$b_k=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^k$

例1　有一容量为10的子样（2，1，3，4，1，5，5，5，1，3），其平均数和方差为

$$\bar{x}=\frac{1}{10}(2+1+3+4+1+5+5+5+1+3)=\frac{30}{10}=3$$

$$S^2=\frac{1}{10}[(2-3)^2+(1-3)^2+(3-3)^2+(4-3)^2+(1-3)^2+(5-3)^2$$

$$+(5-3)^2+(5-3)^2+(1-3)^2-(3-3)^2]=\frac{26}{10}=2.6$$

当子样用频数分布给出时

子样平均（数）　$\bar{x}=\frac{1}{n}\sum_{i=1}^{l}m_ix_i^*$

子样方差　$S^2=\frac{1}{n}\sum_{i=1}^{l}m_i(x_i^*-\bar{x})^2$

子样k阶原点矩　$a_k=\frac{1}{n}\sum_{i=1}^{l}m_ix_i^{*k}$

子样k阶中心矩　$b_k=\frac{1}{n}\sum_{i=1}^{l}m_i(x_i^*-\bar{x})^k$

例2　从母体中抽得容量为10的子样同例1，其频数分布为

x_i^*	1	2	3	4	5
m_i	3	1	2	1	3

其平均数和方差为

$$\bar{x}=\frac{1}{10}(3\times1+1\times2+2\times3+1\times4+3\times5)=\frac{30}{10}=3$$

$$S^2=\frac{1}{10}[3\times(1-3)^2+1\times(2-3)^2+2\times(3-3)^2+1\times(4-3)^2+3\times(5-3)^2]=\frac{26}{10}=2.6$$

设母体X的平均数u，方差σ^2。由切比雪夫大数定律，当$n\to\infty$时，$\bar{x}$依概率收敛于u，即对任意$\varepsilon>0$，

$$\lim_{n\to\infty}P\{|\bar{x}-u|<\varepsilon\}=1$$

此结果表明n很大时，可用一次抽样后所得的子样平均数$\bar{x}$近似表示母体平均数u。

对X_1^2，X_2^2，…，利用切比雪夫大数定律可得$\frac{1}{n}\sum_{i=1}^{n}x_i^2$依概率收敛于$EX^2$，再利用依概率收敛的性质可得　$S^2=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\bar{x}^2$依概率收敛于　$EX^2-u=\sigma^2$，即对任意$\varepsilon>0$