高等教育教材

铀矿物化探数据处理方法

张锦由　编
黎春华　编
唐声嗤　审

原子能出版社

高等教育教材

铀矿物化探数据处理方法

（修订版）

张锦由
黎春华 编

唐声嗤 审

原子能出版社

图书在版编目(CIP)数据

铀矿物化探数据处理方法/张锦由　黎春华编.—修订版.—北京:原子能出版社,2001.2
ISBN 7-5022-2186-7

Ⅰ.铀… Ⅱ.①张…②黎… Ⅲ.①铀矿-地球物理勘探-数据处理-方法②铀矿-地球化学勘探-数据处理-方法 Ⅳ.P619.140.8

中国版本图书馆 CIP 数据核字(2000)第 82567 号

内　容　简　介

本书按照地质勘查阶段,从理论上和应用上系统地阐述了铀矿地质工作中已成熟的各种物化探数据的处理方法。内容包括:数据的预处理;多元统计分析方法;铀矿测井数据的处理和储量计算的统计学方法;位场数据处理及其它。书内图表较多,实例的构思新颖,叙述简练通畅,涉及面较宽。本书可作为高等学校铀矿勘查地球物理和勘查地球化学专业的试用教材,可供中等专业学校的有关专业选用,也可供从事地质、环境保护等有关专业技术人员参考。本书对于从事气象、生物、水文、医学、管理等统计工作的科技人员也有参考价值。

原子能出版社出版 发行
责任编辑:张　梅
社址:北京市海淀区阜成路 43 号 邮政编码:100037
北京朝阳科普印刷厂印刷 新华书店经销
开本:787×1092mm 1/16 印张:21.125 字数:529 千字
2001 年 5 月北京第 1 版 2001 年 5 月北京第 1 次印刷
印数:1－600
定价:28.00 元

修订版前言

本书是根据1994年3月铀矿地质与采矿教材委员会意见修订再版的。依据社会需求和使用者的意见，删去了原书中第三篇，由黎春华教授重新编写，作为本版书第四篇。内容是空间域和频率域内重磁位场数据处理方法；电阻率测深数学解释；弹性波场的数据处理方法。在序次上也作了调整。

修改过程中得到了唐声[illegible]研究员的支持和帮助，在此表示感谢。

张锦由

2000年11月

第一版前言

本书是根据原核工业部教育司于1983年7月召开的铀矿地质类教材会议审定的编写大纲编写的，可作为高等院校铀矿勘查地球物理和勘查地球化学专业的试用教材。

本书按照地质勘查阶段，以理论联系实际作为指导思想，系统介绍了处理铀矿物化探数据的全过程。

全书共分四篇十三章，主要内容包括铀矿物化探的数据及其预处理；铀矿普查常用数理统计分析方法；铀矿测井数据处理和储量计算；位场数据处理及其它。

全书需讲授100学时左右。不同的课程组可选择相应的重点内容讲授，剩余内容由学生自学或作一般讲授。

讲授本课程前，学生应学习高等数学、线性代数、概率论与数理统计、计算机算法语言和有关铀矿物化探方面的专业课程。

本书由北京铀矿地质研究所唐声喤高级工程师审稿。参加本书审稿会的有成都地质学院贾文懿教授、华东地质学院卢存恒教授、华东地质勘探局269大队郑名寿高级工程师等，与会同志提出了许多宝贵的修改意见。在编写过程中，唐声喤高级工程师给予了热情的支持和诚恳的帮助。对此谨向他们表示衷心的感谢。

书稿的撰写是在原核工业部教育司、原子能出版社、华东地质学院教务处和物探系的具体组织和支持下完成的。王京贵副教授、莫撼讲师、黎春华讲师、刘菁华同志分章看了初稿，并提出了宝贵意见。在此表示真诚的谢意。

限于编者水平，书中缺点和错误在所难免，敬请读者批评指正。

编者　1987.10

目　录

第一篇　铀矿物化探数据及其预处理

第一章　铀矿物化探数据及其概率分布

铀矿物化探数据系指探测铀矿的各种地球物理、地球化学勘查方法的原始观测结果和记录，包括数字记录、文字描述和图形描述。它可分为已知数据和待处理数据，已知数据是在被研究对象为已知的情况下所获得的数据；待处理数据系指为得出被研究对象的某种数学解所需要的数据。本书的目的在于介绍如何对这些数据进行数学处理，得出关于研究对象的某些结论。

第一节　铀矿物化探数据的特性与分类

一、数据的特性

(一)随机性

地质、地球物理和地球化学的作用过程是复杂的、无法控制的。它产生的地质事件不能事先预言，也不能再现。但是它可能以多大的概率出现是可以确定的。因此地质事件是随机事件，并遵循一定的概率规律。诸如铀矿床及其品位、地球物理特征、地球化学晕的组成成分，以及地质构造及其分布等都是随机事件。在物化探数据中，这种性质依然存在，且与相应的地质事件遵循着同样的概率规律。因此，随机性是物化探数据的固有特性。数据的随机性是某些数据处理方法采用概率模型的原因。

(二)局限性

由于宏大的地质体在地表出露得很少以及目前使用的物化探勘查方法攻深能力有限，致使获得的数据具有一定的局限性，不能反映地质体的全体。所以统计推断和地质推断常常受到限制，且有时要承担一些风险。

(三)混合性

数据的混合性系指样本数据来自多个地质总体的特性。例如自某个花岗岩体测得一组伽玛射线照射量率值，这些值可能来自花岗岩岩体的边缘相、过渡相和中心相的基岩或风化花岗岩，所以伽玛射线照射量率的概率分布常出现多个峰或偏离正态分布，表现出多个总体的特征。

(四)空间性或序列性

在概率论中,依赖于非随机性参数(如测点的空间坐标)的一簇随机变量的全体称为随机函数。它的取值与其空间位置有关。即

$$\omega = f(x、y、z) + \varepsilon$$

式中,x、y、z 为空间坐标;ε 为随机误差。

在地质领域里,这类与空间位置有关的随机变量值是十分普遍的,也是十分重要的。例如,钻孔内伽玛射线的照射量率值与地球化学晕的浓度、岩层标志值等观测值均属此类。应用这类数据时,其顺序是不允许颠倒的。

二、数据的分类

铀矿物化探数据按其获取数据的方法或方式不同分为五种类型。

(一)测量型数据

测量型数据系指连续性的观测值,它们之间不仅能比较大小,而且能定量地表示其间的差异。例如各种仪器的观测值、地球化学元素的分析值等数据均属此类。

(二)计数型数据

计数型数据系指以不连续的个数为计数特征的数据。例如核素的辐射粒子数、异常点数、矿床和矿点数等均属此类。

(三)级序型数据

级序型数据又称等级型数据,是离散型数据的一种。这类数据是按等级划分的具有等级顺序的数据,并且等级之间的级差在绝对量上不一定相等。例如矿床等级数、异常等级数、场晕等级数等数据。

(四)状态型数据

状态型数据系指用逻辑数字"-1,0,+1"表示事物状态的数据。包括二态型数据和三态型数据。通常二态型数据用"1、0"表示"有、无"。例如有矿、无矿;矿致异常与非致异常;是与不是等状态。三态型数据用"-1、0、+1"表示"无、不确定、有"三种状态。

(五)名义型数据

这类数据没有量的概念,只起一种代码作用。它常用于描述不包含相对重要性或相对变化的对象。例如描述岩石类型、电性异常曲线类型、异常性质、构造方向等。若用名义型数据描述放射性异常性质,可以用"1"、"2"、"3"或"A"、"B"、"C"代表"铀异常"、"钍异常"、"铀钍混合异常"进行处理。这里"2"不是2个"1"的和,也不意味着"2"比"1"大,它们只是用来区分研究对象的某种标志或概念的符号。

三、样本构成的条件

随机样本(简称样本)是统计分析的基础,不同的研究目的和统计分析方法对构成样本的数据要求是不同的。概括起来由数据构成的样本需具备如下基本条件。

(一)样本的随机性

自然界中,物化探数据及其空间分布的随机性是构成随机样本的基本性质。为了保证样

本的这一性质，通常的方法是在抽样总体范围内布置简单的矩形测网，在网的结点上或按一定间隔采样或测量，或在数据图上绘制的相等单元内取值。

(二)代表性

样本的代表性取决于构成样本数据的代表性。不同物化探方法采集的样本数据有不同的代表性，而且它们的意义也各异。

在化探工作中，样本的代表性需要考虑两个方面：

1. 样品的采集位置，体积大小和采集对象

由于在不同深度的土壤层中地球化学元素的富集程度和特点各异，致使各层元素的平均含量和变化性(如方差)不同。另外，即是在同一层中取不同粒度的土壤，其平均含量和变异性也有差异。为了避免这种影响，采集样品时需要尽可能保持采样深度和采样对象的一致性。

样品体积对样本代表性的影响，一般是单样体积越大，样品中元素的平均含量越低、方差越小，频率分布逐渐趋于正态分布，样本的代表性越大。这是进行总体对比、类比时必须注意的问题之一。

在勘查地球物理中，无论是放射性勘查方法，还是其他地球物理勘查方法，由于它们的探测深度和范围大，所以其测量值的代表性比一般化探样品大得多。探测深度和探测对象不同，测量值的代表性和意义也各异。

2. 采样点的分布是否均匀合理，数目是否足够

采样点的分布是否合理取决于研究对象的大小。通常大而均质的地质体可稀些，否则可密些。总之，分布要均匀，密度要合适，以能控制和反映整个研究对象的特征，使推断结论不致产生错误为准则。同时为了便于推断解释，采样时还要尽可能分清地质体采样。

采样点数目对样本代表性的影响，可以用数理统计中的估计区间来说明，估计区间越小，则样本代表性越大，反之越小。它与采样点数 n(即样本容量)的关系可由下式估算。即估计区间

$$m_\alpha = t_\alpha \frac{s}{\sqrt{n}}$$

上式表明，对于同一研究对象(此时均方差 s 呈定值)，当概率系数 t_α 一定时，n 越大，估计区间 m_α 越小，估计精度越高，则样本代表性越大。否则代表性越小。换言之，若要求 m_α 一定时，对含量变异性大的元素需要增大样本容量；对变异性小的元素只要取少量样品就有较大代表性了。

(三)样本含义的准确性

样本的含意取决样本中变量的定义，这是不容忽视的。如果变量的定义(如岩石定名)不准确或不严格，则样本的含义就会产生混乱，理解上就会产生差异。这样不仅会给准确选择变量及其参数造成困难，而且难于准确推断研究对象和进行成果解释，有时甚至得出错误的结论。

例如，确定岩石的伽玛射线照射量率底数时，不仅需要明确统计岩石底数的目的，注意岩石定名是否准确统一及仪器类型和测量条件是否一致，而且对同种岩性哪个范围的测量值哪些能参加统计，哪些应当舍弃等都要依据研究目的事先作出规定。否则样本含义的准确性就不能保证，由样本得出的底数及其变化的原因也难于解释。

变量参数的确定和选择是以精心选择和定义变量为前提的，变量确定了，其相应的各种参数才能依据研究目的加以确定。例如，在矿产预测中要进行“定量预测”，首先要查明控矿因素和找矿标志，然后才能依据它们的变化确定找矿远景区的空间位置，给出远景区可能发现矿床的概率以及各种变量的数量特征或有利找矿区间等。这些参数确定的是否准确，直接影响“定量预测”的效果。

(四)数据的可靠程度

样本数据的可靠程度系指数据反映研究对象的某种特性的真实程度。包括一定测试条件下数据的准确度、精确度和数据的可利用程度。

所谓“准确度”系指观测值与真值的符合程度。精确度系指观测值的重复性大小。“可利用程序”系指为了达到某种研究目的，数据满足其要求的程度。

一般来说，一组测量值中，尽管精确度很高，准确度不一定很好，可利用程度也不一定很高。而准确度好，精确度一般就高，但数据的可利用程度不一定很高，其原因是数据的可利用程度与一定的研究目的相关联。也就是说，如果数据的可利用程度高，则对某一项研究目的来说，数据就是可靠的，准确度和精确度也就达到了要求。例如为了划分岩层，只要不同岩层间的物性差异大于各岩层内的物性差异即可，因此精确度不一定要求很高。但是为了满足大面积放射性普查的需要，达到区分岩性与微弱异常的目的，必须满足普查对仪器的精确度要求和对仪器间一致性的要求。而且精确度越高，一致性越好，则数据的可利用程度越高，因此，这时就不要求很高的准确度。但是，对于矿体品位的测定，准确度是必须保证的。

(五)数据的统一性

为了保证统计分析结果的可对比性，变量的选择、观测方法、取值区间、数据的取舍标准要统一。因此有的问题事先需要制定某些准则。

总之，在找矿、探矿和矿产预测中，如果样本的上述条件能得到满足，那么用样本推断研究对象，并与专业知识相结合，就可能获得好的效果，否则效果就不会好，甚至导致错误的推断或结论。这就是说，在进行统计分析的过程中，不仅要注意数据处理的全过程，而且要胸怀全局，注意制约整个过程的前提条件。否则再严密、再精确的数学分析及其结论也是没有用的。

四、模　　型

模型是实际事物的一个“缩影”，是自然现象及过程的代表。在地质工作中，模型是一个重要的研究工具，同时也是准确选择变量和获取数据的参照依据。通常模型分三种类型。

(一)物理模型

物理模型是代表自然现象及过程的有形实体。其优点是能将主要的、可控制的地质要素抽取出来，制成变换了尺寸的实体加以有型的模拟研究。例如，伽玛测井中测定换算系数的测井模型；在导电纸上模拟的地球物理电性模型等均属此类。

(二)地质模型

包括地球物理模型、地球化学模型。它是通过直接或间接收集的观测资料并加以检验、抽象成某种概念的系统阐述。在传统地质学中，地质模型可以概括成以下二类。

1. 比例尺模型

比例尺模型系指用缩小比例尺的办法表示地质变量和物化探变量空间变化的图示。例如

各种地质的、物化探的平面图、剖面图、柱状图等都是比例尺模型的例子。

2. 概念模型

概念模型系指各种地质(含地球物理和地球化学)现象及其形成过程在头脑里的反映。它可以是定性的,也可以是半定量或定量的。例如矿体扩散晕内成矿元素含量(或浓度)随远离矿体而下降的现象。又如某铀矿床的成矿条件为一槽(赋存条件)二碳(富集条件)三覆盖(保存条件)等都是概念模型的例子。

(三)数学模型(又称数学模式)

数学模型仅是一组形式法则。它规定变量间的依存关系,这种关系可用数学符号表示成数学方程。常用于描述指定条件下某种地质现象与过程的基本规律。数学模型的例子很多,概括起来可以划分为确定模型与概率统计模型。确定模型中,变量间的关系完全可以准确确定,且能用函数关系表示。概率统计模型不能用普通的函数表示,只能用统计关系式来表示。地质领域里的大多数数学模型在某些重要方面具有不确定性,因此它们主要是概率统计模型。

无论是物理的、地质的还是数学的模型,其目的都是将自然现实加以概括和抽象、简化和组织,以便集中研究自然现象或过程中的一个或几个因素进行模拟、借以解释或解决地质问题。本书的目的之一,就是把地质模型(含地球物理和地球化学模型)转变成数学模型,以使所提问题简单形象,计算方便,而无需像传统的地质学方法那样进行大量而细致的图示及文字描述来分析研究问题。

(四)模型的建立

客观的地质规律是建立各种模型的基础。模型的建立依赖于专业知识和经验的丰富程度以及在参与地质实践活动中解决问题和分析问题的能力。建立模型之后,必须经过实践检验,看其是否同客观情况相符合。当模型显得不够理想时,在检验过程中就要进行修正或提出新的、较好的模型,使模型更符合实际。

第二节　铀矿物化探数据的统计描述

一、几个名词和概念

(一)随机样本、抽样总体和目标总体

抽样总体简称总体,系指被实际抽取样本的研究对象的全体。个体是抽样总体中的一个单位。

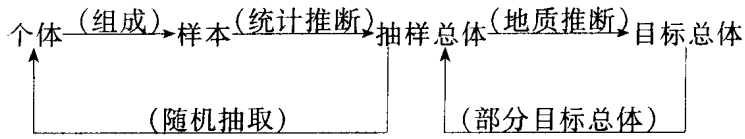
随机样本是从抽样总体中随机抽取的一部分个体。样本中个体的数目称样本容量或样本大小。

在地质领域里,抽样总体通常是无限总体。若按变量个数来划分,有单元总体和多元总体。对于单元总体,个体就是某个元素的一个观测值;对于多元总体,个体是样本中相应元素的一组观测值。因此,一个地质样品与样本的概念不同,总体与地质体的概念也不同。

由于地质工作的特殊性,这里提出目标总体的概念。关于目标总体,科克伦等著的书中写道:"目标总体是我们所关心的总体,地质工作者所希望的就是作出关于它的推断或结论,它正是我们所研究的对象"[32]。为了进一步说明抽样总体和目标总体这两个概念,举例说明如下:

假如研究某个岩体含铀矿的可能性或在什么地方有铀矿体，可对整个岩体按一定比例尺进行伽玛测量(随机性的空间测量)，那么最多能得到整个岩体(目标总体)露出地面的被实际抽取样本的那一部分的全体(抽样总体)，地下的岩体部分测不到(当然也就不是抽样总体)。而且地表的那一部分的全体由于风化作用、污染及地质体的不均匀性等原因与地下的部分也不完全相同。这就是说，抽样总体只能是目标总体的一部分，在形成的时间、空间以及组成成分方面，抽样总体与目标总体也不完全相同。事实上要二者完全一致起来是相当困难的。因此由抽样总体正确推断目标总体，不能只依靠统计理论和方法，还必须依靠研究者专业知识的丰富程度和经验的成熟程度。

应当指出，个体、样本、抽样总体和目标总体之间的内在联系是处理数据的出发点，它们的关系可图示如下：



在地质研究中，由个体观测值组成样本，用样本统计推断抽样总体，进而用专业知识解释、推断(地质推断)目标总体，这是数据处理的最基本的工作步骤。事实上，用样本得出的关于总体的结论是(也只能是)关于抽样总体的统计推断。对目标总体来说，这种推断只能起一种"启发性"或"指示性"的作用。这种作用的大小与二者的相关联程度有关，如果抽样总体与目标总体一致，则关于抽样总体的结论就是目标总体的结论。这时如果目标总体是矿体，则结论也适合这个矿体。

上述步骤是由个别推断一般，由局部推断全体的问题。因此在进行统计推断时，必须使构成样本数据的五个条件得到满足。尽管这样，推断也仍然有一定的风险，这种风险一方面来自地质现象的随机性，另一方面来自抽样总体与目标总体的不一致。这也是地质工作的困难所在，同时也是常被地质工作者忽略而又非遇到不可的普遍问题。

(二)频率分布与概率分布

某一事件 A 在 N 次试验中出现的次数 m，称为频数，而出现次数 m 与试验次数 N 的比 $\left(\frac{m}{N}\right)$，称为频率或相对频数。若在相同条件下多次重复试验，当试验次数 N 足够大时，可以发现，某随机事件(大量重复试验具有统计规律的事件)出现的频率是趋于稳定的。它围绕着某一固定数值作微小波动，这一固定值体现了随机事件的统计规律性。它反映某一事件在某种条件下出现的客观可能性大小。这个某事件出现的客观可能性的固定数值(常数)，称为该事件出现的概率，通常用 p 表示事件 A 在每次试验中发生的概率。记事件 A 出现的概率为 $P\{A\}$。不可能事件的概率为 0，必然事件的概率为 1，而随机事件的概率介于 0 与 1 之间。由贝努里大数定理可知：

$$\lim_{N\to\infty} P\left\{\left|\frac{m}{N}-p\right|<\varepsilon\right\}>1-\delta$$

式中，ε、δ 为任意小的正数。

当试验或观测次数无限增大时，事件出现的频率无限接近其概率 1。这一点正是矿床勘

探中用足够数量的样品的平均品位估计矿体真实平均品位的理论依据。

对 N 个试验结果(N 个观测值)进行统计,用等间距的区间分组作为横轴分度值,用各组观测值在 n 个观测值中出现的频率作为纵轴的高,绘制矩形,即构成频率分布直方图(见图 1-14)。如果从左边开始用直线段依次连接上述直方图中矩形的顶边中点得一条折线,这就是频率分布的折线图。这种以样本观测值为基础得出的频率分布,又称为样本分布或经验分布,记为 $F_N(x)$。假设样本容量无限增大并将各分组区间的长无限减小,则可得出一条反映事件出现的可能性大小的光滑曲线。实际上不可能无限增大样本容量,但通过理论概括可以得出反应总体分布的理论频率曲线,也称总体分布,记为 $F(x)$。二者的关系可以用格利文科定理(依概率收敛)说明:

当 $N\to\infty$ 时,称 $F_N(x)$ 依概率收敛于 $F(x)$,即

$$P\{\lim_{N\to\infty,-\infty<x<\infty}\sup|F_N(x)-F(x)|\geqslant\varepsilon\}=0$$

也就是说,当 N 很大时,样本的分布函数 $F_N(x)$ 实际上将近似等于总体的分布函数 $F(x)$。这也是用样本分布推断总体分布的理论依据。

(三)总体参数与样本统计量的关系

一种概率分布的特征常常用它的特征参数来表示。例如正态分布的集中性、离散性、陡峭性、偏斜性(不对称性)等,可分别用总体的均值、方差、峰度系数、偏度系数来表征。由于它们都是事先未知的,只能用样本的相应特征数来估计。为了区别于样本统计量,称总体特征数为参数,常用希腊字母表示,如 μ、σ、γ_1、γ_2 等。由样本值计算得到的特征数称为样本统计量,常用英文字母表示,如 $\bar{x}$、s、g_1、g_2 等。用样本统计量估计总体参数,称为"参数估计"。由中心极限定理可知:不论原始分布如何,当样本容量增加时,样本平均数的分布近似于正态分布。此时总体参数与样本统计量之间有如下关系。

①样本平均数的平均数等于总体平均数即

$$\mu_{\bar{x}}=\mu$$

②样本平均数的方差等于总体方差除以样本容量 n,即

$$s_{\bar{x}}^2=\frac{\sigma^2}{n},\text{或 }s_{\bar{x}}=\frac{\sigma}{\sqrt{n}}$$

需要注意的是,上述关系只有当 n 足够大时才成立。对于地质变量,一般样本容量在 50 以上时即可以认为成立。

二、样本统计量及其分类

(一)集中性统计量

它反映了分布的集中趋势,可作为大量数据的整体性代表。这类统计量是算术平均数、几何平均数、加权平均数、中位数和众数等。

1. 算术平均数

它是最常用的一种平均数,这是因为样本平均数是总体平均数的无偏估计量。

设 $x_1,x_2,\cdots,x_n$ 为一个样本,则样本的算术平均数

$$\bar{x}=\frac{1}{n}\sum_{i-1}^{n}x_i \tag{1-1}$$

对于样本平均数 $\bar{x}$，有 $E(\bar{x})=\mu$。这就是说 $\bar{x}$ 是 μ 的无偏估计量。

算术平均数的物理意义为各个数值的平衡点或重心。在铀矿物化探工作中，常作为岩石中的元素背景值的估计值或矿体平均品位的估计值。

当 n 很大时，计算 $\bar{x}$ 值就必须分组统计。设分 N 组，把落入每组内的 x_i 值一律用该组的组中值 x_j 来代替。此时

$$\bar{x}=\frac{\sum_{j=1}^{N}f_j^*x_j}{n}=\frac{\sum_{j=1}^{N}f_j^*x_j}{\sum_{j=1}^{N}f_j^*} \tag{1-2}$$

式中，x_j 为第 j 组的组中值；f_j^* 为第 j 组内频数。

用 1-2 式求平均数的方法叫加权平均法。所得平均数叫加权平均数。这里“权”是频数。

2. 几何平均数

设 $x_1,x_2,\cdots,x_n$ 为一个样本，则其几何平均数

$$\bar{x}_L=\sqrt[n]{x_1\cdot x_2\cdot\cdots\cdot x_n} \tag{1-3}$$

计算时一般取对数。即

$$\log\bar{x}_L=\frac{1}{n}\sum_{i=1}^{n}\log x_i \tag{1-4}$$

当 n 很大时，同样可以分组统计，用加权平均法计算

$$\log\bar{x}_L=\frac{1}{n}\sum_{j=1}^{N}f_j^*\log x_j \tag{1-5}$$

$$n=\sum_{j=1}^{N}f_j^*$$

式中，f_j^* 为第 j 组的组内频数；$\log x_j$ 为第 j 组对数值的组中值；N 为分组数。

当求出对数值的平均数后，取反对数就是几何平均数。

3. 中位数

中位数(M_e)是将一个样本的诸值按大小次序排列居于中间的那个数值。例如在样本分布中累积频率等于 50% 时，所对应的自变量值即为中位数。其优点是求法简单，与数列两端的数值变化无关。

4. 众数

具最大频数(或最大频率分布密度)的自变量值，称众数(M_Q)。

应当指出，在非对称曲线中，中位数在众数和平均数之间(如图 1-1)。在对称曲线中三者重合，如正态频率分布曲线中平均数、中位数和众数就是相等的数。

应该指出，上述统计量都是用于估计总体平均数——即数学期望值的。实践中应用什么样的统计量估计总体分布的中心趋势呢？原则上说，用最接近数学期望值的那个统计量值。但是在有些问题中并不要求有很高的精度，这时可选用计算方便的统计量。例如在铀矿普查中，岩石伽玛射线照射量率底数的确定，就常用中位数来估计总体平均数。

(二)离散性统计量

它反映了数据分布的离散程度，是统计误差大小的量度，常用来反映数据的波动性质。这

类统计量有:极差、方差或均方差,变异系数等。

1. 极差

它是样本值中最大值与最小值之差。常用 d 或 D 表示,即

$$d = \max\{x_i\} - \min\{x_i\}, i = 1,2,\cdots,n \tag{1-6}$$

极差计算简单,但不能充分提供有用的信息。

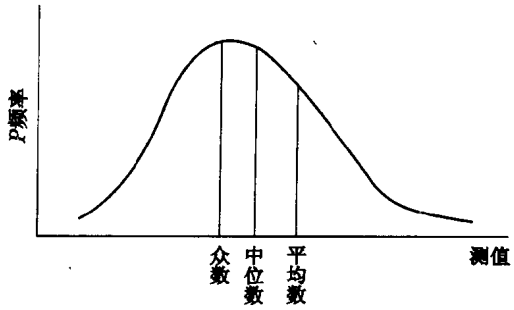


图 1-1 不对称频率分布特征值之间的关系

2. 方差与均方差(标准差)

它们是常用来描述数据波动性的两个统计量。二者能较好地提供有用信息,反映数据的离散程度。

设有 n 个观测值 $x_1, x_2, \cdots, x_n$,组成一个样本,其平均数为 $\bar{x}$(算术平均值),则其方差和均方差分别为:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{1-7}$$

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{1-8}$$

均方差是每个观测值(x_i)与样本平均值($\bar{x}$)之离差平方和的均值再开方,所以叫均方差。

应当指出,当 n 较小($n<30$)时,用

$$s = \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1)} \tag{1-9}$$

计算。这是因为(1-9)式算出的均方差才是总体均方差 σ 的无偏估计量。由(1-8)式计算出的均方差不是 σ 的无偏估计量

当 n 比较大且需要分组时用下式计算

$$s = \sqrt{\frac{1}{n}\sum_{j=1}^{N} f_j^*(x_j - \bar{x})^2}$$

$$= \sqrt{\sum_{j=1}^{N} f_j(x_j - \bar{x})^2} \tag{1-10}$$

式中,把样本值 $x_i(i=1,2,\cdots,n)$分成 N 组;x_j 为第 j 组组中值;f_j^* 为第 j 组频数,f_j 为相应的频率。各组频数和 $\sum_{j=1}^{N} f_j^* = n$;平均值 $\bar{x} = \frac{1}{n}\sum_{j=1}^{N} f_j^* x_j$。

依据矩的定义,均方差是关于平均数 $\bar{x}$ 的二阶中心矩的均方根,二阶中心矩是方差的另一个名称。因方差的单位是平方单位,与被观测对象的单位不相同,故一般用均方差作离散程度的直接量度。这个量度代表的是样本中 n 个数据分布的离散程度,不是个别数据的误差。

由(1-10)式看出,均方差与测量单位有关,这是比较总体均方差时要注意的。

为了计算方便,(1-10)式写成

$$s = \sqrt{\bar{x}^2 - (\bar{x})^2} \tag{1-11}$$