成地科技情报(四)

575

矿产评价的地质统计学方法

成 都 地 质 学 院

1 9 8 3 年 4 月

原　书　序　言

本丛书的第一本是D·G克里格的论著“矿石评价中的对数正态戴维杰斯地质统计学”J·M伦杜现又创作了丛书的第二本，它补充并充实了第一本所介绍的理论。

伦杜是克里格的学生，能够很好地赶上他的老师。本书的叙述是很清晰的，它从大多数地质采矿工程师和地质采矿大学生所熟悉的简单方式出发，然后再扩展开以论述空间统计学的用途和重要性。最后一章介绍了泛克里格法的概念，但并不想详尽地加以论证。这是由于正态克里格法处理了通常问题的绝大部分，尤其对读者有益的是附有大量的工作实例，用以说明书中数学的演进。

解决矿石评价问题方法的基础是由西奇尔和克里格建立的，克里格对早期的发展作了很多的贡献，而后由马西龙加以完善和定型，发展了这门学科（地质统计学）。伦杜从这些前辈的工作中取得其精华，写成简练的形式，人们认为伦杜是有着他自己的重大贡献的。

虽然不是要设计什么理论来解决或阐述某种问题，但对于所有有兴趣于空间相关变量或需要评价矿产的人来说，由于这些理论超越了采矿工程的范围，但又不得不熟悉已经发展成为名为“地质统计学”的理论和原理，对他们说来，这个问题的本身并不是很容易的，而伦杜的书将大大地简化了这个工作。

伦杜自己是很谦虚的，仅宣称本书是为了对地质统计学的理论背景及其实际应用有兴趣的学生和采矿工程师准备的，我的看法是它具有更深远的魅力和重要性。尽管如此，对于那些在实际中需要对矿床进行深入评价的人说来，他确实是编写了一本十分重要的读物。

R·P普洛曼

教授，采矿工程系主任

（约翰内斯堡、威瓦特斯兰大学）

前　言

本书是按照南非出版的数学地质丛书第二本编写的。这套丛书是作为地质统计学克里格法奠基人之一D·G克里格和他的学生们所编著。它适宜于地质、勘探、采矿、资源评价及数学地质人员学习参考之用，也可作为大学有关地质学科的学生作为教材或教学参考书。

参加本书编译的有陆承新、王柏钧、李德清等（其中原序、目录、附录、附录、第一至第六章由王柏钧、李德清执笔，其余部分由陆承新执笔及全书润色），并最后由王柏钧（我院数学地质研究室主任）负责全书的校订和成书。全书图表较多，所有图件均由季恒玉同志清绘，并得到庞加研同志的帮助。由院情报资料室出版和发行。

由于成书仓促，错误之处必然不少，深望读者给予指正。

成都地质学院情报资料室

1983年4月

目　录

第 一 章

导 论

地质统计学为应用统计学的一个分支。它的目的在于对地质观测给予数学描述与分析。地质统计学可以应用于纯粹地质学（例如应用于变质岩中追踪元素的分析），应用于矿产勘探（例如对地球化学勘探数据的分析），以及应用于矿产资源评价等方面。本书的目的在于对矿产评价方面的地质统计学提供一个在实用上的导论。

这些年来，对于矿床中某些值的分布的表示，提出了各种数学模型。比较简单的数学模型都根据这样一个假定：这些值的分析都是随机的。根据这个假定，我们可以应用经典统计学来分析矿床，对于这些矿床来说，假定这样的模型是能适合的。但是，在所有的矿床中人们都知道有这样的区域存在，在这些区域内的值要比其余别的区域的值要高些或低些。还有，如果我们在矿床中的两个邻近的地方取样，则这两个样品值比较接近；而在两个较远的地方取样，则不发生这种情形。这就指出了，在样品值与样品值之间就存在着某种程度的相关关系，而且这种相关关系是样品之间的距离的函数。考虑到这种相关关系的数学模型已经建立起来了；样品值与样品值之间的相关程度我们就用半变异函数来量度。但在这种模型中，我们又要考虑另一种事实：非常邻近的两个样品也很可能表现极不相同的样品值，也就是说，即便距离非常小而相关性却不具备，在值的分布中表现出一种纯粹的随机性。因此，在数学模型中我们必须假定两种变异因素同时存在：相关性与随机性。

除此之外，人们还必须考虑矿床的一种特殊的但也很普遍的情况，在这样的矿床中，它的值表现出一种综合的空间变异性，通常看作为一种偏移（drift 或译漂移，或一种趋势，trend）。例如矿体的级别可以随矿的深度而增加，或者随矿体离开中心火山口的距离而减少。Carlicer模型没有对偏移给予适当的表示，但人们已提出了比较复杂的模型，在这种模型中人们对变异考虑有三种原因。所以这种模型是由三种因素组合而成的：决定性因素，相关性因素，随机性因素。决定性因素是用来说明偏移，相关性因素解释了矿床值的规律性变异，而这种规律性变异又不是从偏移而来的；随机性因素所表示的变异就是不能用上述两种因素所能解释的变异。

开始根据只有随机性因素的假定，我们叙述比较简单的模型（第二章）。随后根据在随机性因素上又加上相关性因素的假定，我们详细地叙述这样的模型，这种模型非常普遍地应用于矿床分析中（第三——十一章）。最后，涉及偏移存在的模型，我们仅简短地予以叙述（第十二章）。本书基本上是为采矿工程系的学生们以及对地质统计学理论与应用有兴趣的采矿工程师们所写的。假定读者具有统计学的基础知识。学习第九章的某些部分与第十二章时，需要一些线性代数的知识。关于地质统计学的所有公式，我们都给出了证明，这些证明

在一般初等统计教科书中都不能找到。这些证明的理解，对于实际的应用者来说，并不是非常必需的。读者在初次阅读时，可以略去这些论证，而把注意力集中在数值例题上面。

虽然理论地质统计学者们在本书中找不到新的东西，但在论证公式所应用的方法中，他们会感到兴趣是可以预料的。

第 二 章

经典统计学，随机分布 正态理论与对数正态理论

§ 2.1概述

要确定矿床的特征，通常的方法是先取样，再分析样品的性质，然后从这些性质来推断矿床的特征。这种分析必须应用统计学来实现。目前所应用的两种统计方法，即经典统计方法与空间统计方法。

如果采用经典统计方法来表示样本值的性质，则就假定了样品值都是随机变量的实现。样品的相对位置完全不予考虑，并且假定矿床中的所有样品都有相同的概率被选取。对于富集带和矿化作用的集中区，存在着忽视的倾向。而且，对两个非常靠近的样品具有相接近的样品值这样的事实，也未加以考虑。

反之，如果人们考虑到样本值是随机函数的现实，则就会应用空间统计方法。这时，我们把样本值考虑为矿床的矿化作用中的位置函数，因此必须要考虑样品的相对位置。对样品值之间的相似性，我们定量地把它表示为样品的距离函数；而这种函数就成为空间统计学的基础。

只有在比较少数的情况下才可以应用经典统计学。也就是假定所有样品值在矿床中分布存在着等可能性，这仅仅在样品值是随机地分布的情况下，这个假定才能满足。事实上，矿床中的样品值常常不是随机分布的。地质人员也总是避免纯粹的随机取样，他们常常在规则的网格上或者近乎规则的网格上取样。他们认为这种取样方法比纯粹随机取样方法得到更多的信息。实际中，经典统计学只应该在勘探早期应用；因为这时样品个数比较少而且样品之间的距离比较大。在这种情形下，使用空间统计学的可用信息不够充分，而应用本章的方法是合适的。

§ 2.2 定义和附注

我们用Ω表示所考虑的矿床的区域（即体积和表面积）。在Ω中取一点Z，并取一个以Z为中心的样品W，样品值为$x(z)$，它是位置Z的函数。例如，它可以是矿的品位，矿石的厚度，单位体积中金属的含量，或样品的其它的定量特征。$x(z)$称为区域化变量，它的值是与样品W的大小与位置有关。W定义为$x(z)$的支撑（或译支集）（见图2.1）。

如果我们取矿体Ω内所有可能的点Z，以及所有可能的样品W，则可以算出变量$x(z)$在Ω内的平均值，这个平均值不依赖于支撑W，而可以用下面的记号来表示，即

$$\mu = E_{\Omega}[x(z)] = x(z)\text{在}\Omega\text{内的平均值} \qquad (2.1)$$

在勘探的初期，分析的主要问题就是对 μ 的估计。为此，我们在点 z_i 处（$i=1,2,\cdots\cdots n$）取相同大小的支撑W的 n 个样品。第i个样品的样品值为 $x(z_i)$。我们的目的就是要用这些样品值 $x(z_i)$ 来计算平均值 μ 的估计值 $\hat{\mu}$，以及平均值的置信区间（本书用符号∧来表示估计值）。这样所得到的估计值将随着 $x(z)$ 的概率分布而定。

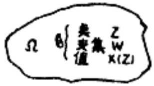

图2.1 矿体Ω中以 z 为中心的样品W之值 $x(z)$

因为在分析中我们已假定所有的样品值是独立的，也就是说样品值 $x(z_i)$ 的位置 z_i 可以忽略不计。于是我们可以用 $x = x(z)$，从而可用

$$x_i = x(z_i) \qquad (2.2)$$

$$\mu = E[x] \qquad (2.3)$$

在这里，对样品值x的分布，我们只考虑两种类型，即正态分布与对数正态分布。其它类型我们可以在别的文献中找到（如Sichel，1973；David，1977；Becken，1964—1966），这些研究已超出本书范围。在大多数实际情况，我们可以假定变量是服从正态分布或者对数正态分布，而应用更为复杂的分布并不适宜。

§2.3 正态分布

2.31 概述

设已有n个样品值 x_i（$i=1,2,\cdots\cdots,n$）。首先，分析的第一步，便是把它们按大小排列而分组，然后计算落在各个组内的样品的个数。这种分组的一个结果已列在表2.1中的第一、第二、第三、三列中。利用这个分组的结果，我们可以画出它的直方图来（见图2.2）。直方图是一个很有用的工具，通过它我们可以直接地看出分布是否对称，样品中是否有异常值，样品值中是否有特殊高的或者特殊低的值。当然直方图的形状是与分组有关系的。

要确定样品值是否服从正态分布，可以应用通常的作图方法。先把样品值的累积频率计算出来（见表2.1中第四、第五两列），然后把这些值画在正态概率纸上（见图2.3），根据正态概率尺度的定义，正态分布变量的累积频率在正态概率纸上应形成一条直线。应用这个原理，如果样品值的累积频数在概率纸上的位置恰好沿着一条直线，或接近于一条直线，我们就有理由认为样品值可以接受正态分布的假定。

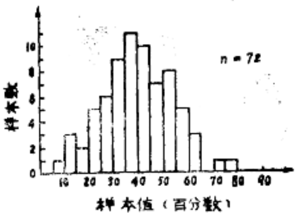

图2·2 样本的值

实际上，除非所取样品的矿化作用有较高的程度（例如某些铁矿床），或者样品值的变化较小（例如成层矿床的厚度，矿化后的比重），样品值的正态分布的假定常常不被满足。如果在一个新的矿体中取样太少而不足以画出有代表性的直方图以及累积频数，则可利用过去的经验来判断是否可以接受正态性的假定。

表 2.1　样品值的百分累积频数分布的计算

组数 \ 列数	1 样品值下限(%)	2 样品值上限(%)	3 组内样品个数	4 样品值的累积频数	5 累积频数(%)
1	5	10	1	1	1,4
2	10	15	3	4	5,6
3	15	20	2	6	8,3
4	20	25	5	11	15,3
5	25	30	6	17	23,6
6	30	35	9	26	36,1
7	35	40	11	37	51,4
8	40	45	10	47	65,3
9	45	50	7	54	75,0
10	50	55	8	62	86,1
11	55	60	5	67	93,1
12	60	65	3	70	97,2
13	65	70	0	70	97,2
14	70	75	1	71	98,6
15	75	80	1	72	100,0

2.3.2 对平均值，方差，和置信区间的估计

样本均值和样本方差估值如下

样本均值　$$\overline{x}=\frac{1}{n}\sum_{i=1}^{n} xi \tag{2.4}$$

样本方差　$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2 \tag{2.5}$$

或者　$$S^2=\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2-n\overline{x}^2\right) \tag{2.6}$$

其中 $S=\sqrt{S^2}$ 是样本母体标准差的估值。矿体的均值估计为

$$\widehat{\mu}=\overline{x} \tag{2.7}$$

具有方差　$$v(\widehat{\mu})=s^2/n \tag{2.8}$$

如果有　$P\{\mu<\mu_p\}=P$

$P\{\mu>\mu_{1-p}\}=P$

则μ_p称为真正平均值的置信下限，而μ_{1-p}称为真正平均值μ的置信上限。于是平均值μ落在μ_p与μ_{1-p}之间的概率为1-2p，即

$P\{\mu_p<\mu<\mu_{1-p}\}=1-2P$

而这个区间（μ_P，μ_{1-P}）称为概率为1—2P的中心置信区间。

如果n＞25，对于矿体的平均值，则可应用下面的近似公式来计算概率为68%与概率为95%的中心置信区间（P＝16%与P＝2.5%）

68%的中心置信区间：$\left(\bar{x}-\frac{S}{\sqrt{n}} \quad \bar{x}+\frac{S}{\sqrt{n}}\right)$ (2.9)

95%的中心置信区间：$\left(\bar{x}-\frac{2S}{\sqrt{n}} \quad \bar{x}+\frac{2S}{\sqrt{n}}\right)$ (2.10)

如果n＜25，则必须应用下面的公式来计算置信下限μ_P与置信上限μ_{1-P}：

置信下限：$\mu_P=\hat{\mu}+t_{1-P}\frac{S}{\sqrt{n}}$ (2.11)

置信上限：$\mu_{1-P}=\hat{\mu}+t_{1-P}\frac{S}{\sqrt{n}}$ (2.12)

其中t_{1-P}为自由度f＝n-1的学生氏t变量的临介值，也就是使得变量t小于t_{1-P}的概率为1-P。即

$$P\{t<t_{1-P}\}=1-P$$

这个t_{1-P}的值可以从所有的统计教科书中找到（例如，Fraser，1958，P289）。为了方便起见，我们将有些t_{1-P}的值列在表2.2中。

例如，对于P＝5%，n＝10，我们得到

$$1-P=95\%, \qquad f=n-1=9;$$

因此，查表2.2，$t_P=-t_{1-P}=-1.833$

2.3.3 平均值与标准差的图形估计法

如果提供的样品个数相当多，则母体的平均值和方差可以用图形法来估计（见图2.3）。读者需要记住下列的概率：

$$P\{x<\mu\}=50\%$$

$$P\{x<\mu-2\sigma\}=0.02275$$

$$P\{x<\mu+2\sigma\}=0.97725$$

其中设x为正态变量，μ为其平均值，σ为其方差。

例如，从图2.3中，我们估计

$$\hat{\mu}=39.5\%$$

$$\hat{\sigma}=(67.5\%-11.5\%)/4=14.0\%$$

由这些估计值我们不能得到平均值μ的精

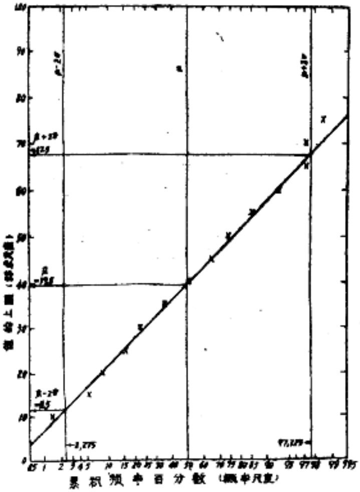

图2.3 正态变量的累积频数分布

表 2.2　t 分布中的 t_P 值

P / F = n − 1	80%	90%	95%	97,5%
1	0,376	3,078	6,134	12,06
2	1,061	1,886	2,920	4,303
3	0,978	1,638	2,353	3,182
4	0,941	1,533	2,132	2,776
5	0,920	1,476	2,015	2,571
6	0,906	1,440	1,943	2,447
7	0,896	1,415	1,895	2,365
8	0,889	1,397	1,860	2,306
9	0,883	1,383	1,833	2,262
10	0,879	1,372	1,812	2,228
11	0,876	1,363	1,796	2,201
12	0,873	1,356	1,782	2,179
13	0,870	1,350	1,771	2,160
14	0,868	1,345	1,761	2,145
15	0,866	1,341	1,753	2,131
16	0,865	1,337	1,746	2,120
17	0,863	1,333	1,740	2,110
18	0,862	1,330	1,734	2,101
19	0,861	1,328	1,729	2,093
20	0,860	1,325	1,725	2,086
21	0,859	1,323	1,721	2,080
22	0,858	1,321	1,717	2,074
23	0,858	1,319	1,714	2,069
24	0,857	1,318	1,711	2,064
25	0,856	1,316	1,708	2,060
30	0,854	1,310	1,697	2,042
40	0,851	1,303	1,784	2,021
50	0,849	1,298	1,676	2,009
100	0,845	1,290	1,660	1,984
∞	0,842	1,282	1,645	1,960

确的置信区间，但在实际中，上面的（2.11）与（2.12）式将被应用，在其中是用$\hat{\sigma}$来代替S。

2.3.4 例题

在矿体中取了8个样品，认为样品值是服从正态分布的，样品值列在表2.3中。

表2·3 样品值

i	1	2	3	4	5	6	7	8
X_i	1.2	2.0	1.6	1.7	2.5	1.9	1.5	2.1

(1) 画直方图，其中组距为0.2。直方图在图2.4中。

(2) 估计矿体的平均值以及概率为90%的中心置信区间。这些都可以应用表2.4求得。

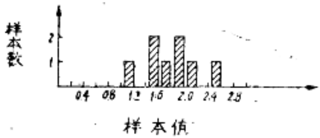

图2.4 在表2.3中的样品值的直方图

表 2.4 均值与方差的计算

i	xi	xi^2
1	1.2	1.4
2	2.0	4.0
3	1.6	2.56
4	1.7	2.89
5	2.5	6.25
6	1.9	3.61
7	1.5	2.25
8	2.1	4.41
和数	14.5	27.41

矿体的平均值为

$$\hat{\mu} = \bar{x} = 1.812$$

若给定$1-2P=0.90$，则有$P=0.05$，对于$n=8$，则从t分布表2.2中查到$t_{0.95}=1.895$。因

此，95%的中心置信区间的上限与下限为

$$上限=\mu_{1-P}=1.812+1.895\frac{0.402}{\sqrt{8}}=2.08$$

$$下限=\mu_P=1.812-1.895\frac{0.402}{\sqrt{8}}=1.54$$

这就是说，平均值μ有90%的机会落在1.54与2.08之间，有5%的机会小于1.54。

§2.4 对数正态分布

2.4.1 概述

在大多数情况下，低品位矿床的样品值分布就不是对称的，而且具有正的偏度（见图2.5）。这些具有偏度的分布通常可以用两个参数的或者三个参数的对数正态分布很好地来表示。

设x为服从对数正态分布的变量。如果$\log_e x$服从正态分布，则我们称x为二参数的对数正态分布的变量。如果$\log_e(x+\beta)$服从正态分布，则我们称x为三参数的对数正态分布的变量，其中β为一正的常数。

图2.6表示，在三参数对数正态分布中变量x与参数β对直方图的影响：x为具有正偏度的分布，$\log_e x$为具有负偏度的分布，而$\log_e(x+\beta)$为正态分布。

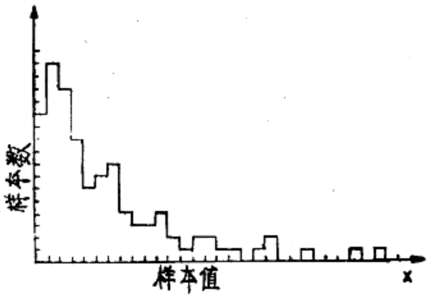

图2.5 具有正偏度的直方图

两个参数的对数正态分布变量的累积频数如果画在对数正态格纸上，则是一条直线。如果变量是三参数的对数正态变量，则它是累积频数曲线将表示一种低数值的超越（excess of low values）。我们在表2.5中列举了具有三参数对数正态分布的变量，它的曲线表示在图2.7中（见曲线C）。

2.4.2 参数估计与置信区间

三参数的对数正态分布律，完全由三个参数所决定，即常数β，$\log_e(x+\beta)$的平均值，以及$\log_e(x+\beta)$的方差。如果我们已经有了几个样品值$x_i(i=1,2,\cdots,n)$，则可以估计这三参数，从而可以计算出矿床平均值μ的估计值$\hat{\mu}$以及μ的置信区间。