

改进三因素统计法数学模型及在矿产 预测中的应用

王全明

(中国地质科学院·北京)

通过对三因素统计法数学模型的改进,使之能够适用于矿产定位
和资源定量预测。并用改进的理论依据与特征分析对比,预测效果相
似,其方法简单、易懂。

关键词: 改进三因素统计法; 矿产预测

问题的引出



工作方法

三因素统计法是近年
生产单位常采用的一种矿
产统计方法。该法的一个
显著特点是对每个地质变
量都从 3 个方面分析,考
查它在模型中的作用大
小,然后通过单元含矿性评价进行预测。
但是,笔者在应用中,对该法数学模型尚有
不同的理解,并作了一定的改进,通过应用
效果显著。下面将改进方案作一简单介绍。

三因素统计法原理 及分析

三因素统计法是传统类比法的一种。它
是通过研究已知成矿单元样品,查明变量之
间的内在联系,确定其成矿、找矿意义,建
立起某种类型成矿单元的定量模型;然后将
模型外推到预测区,再进行预测样品与模型
样品特征的类比。用它的相似程度表示预测
样品成矿的可能性,并据此圈出有利成矿的
各级远景区。

设有 n 个单元, m 个地质变量组成的 $n \times$

m 阶原始数据矩阵

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}$$

所谓三因素,是对每个地质变量都从 3
个方面(某一变量在模型单元中出现“1”时
的次数;对应的总矿石量;所占的平均矿石
量)出发,考查其对模型的贡献,然后将 3
个因素求和,作为该变量的权系数,即

$$R_k = A_k + B_k + D_k \quad k=1, 2, \dots, m$$

(变量数)

式中, R_k 为第 k 个变量的权系数; A_k 为第 k
个变量第 1 个因素的贡献; B_k 为第 k 个变
量第 2 个因素的贡献; D_k 为第 k 个变量第 3 个因
素的贡献。

1. A_k 的求法: 从原始数据矩阵上看,
某地质变量存在(即为 1)的数量越多,表明
该变量对模型越重要,越具普遍性。因此,
设 a_k 为第 k 个变量在所有的模型单元中出
现“1”的频数,即 $a_k = \sum_{i=1}^n x_{ik}, k=1, 2, \dots, m$
(变量数) a_k 越大,表明该变量对模型越重
要,它是反映对单元成矿控制作用大小的一
种度量,为统一量纲,取

$$A_k = a_k / \sum_{j=1}^m a_j \quad k=1, 2, \dots, m$$

其中, $0 \leq A_k \leq 1$, 当 $A_k = 1$ 时, $a_j = 0 (j \neq k)$, 这表明第 k 个变量控矿作用最明显; 当 $A_k = 0$ 时, $a_k = 0$, 表明第 k 个变量控矿作用不明显。

2. B_k 的求法: B_k 表示第 k 个变量存在时所对应的矿床储量与模型单元总储量之比。设 b_k 为第 k 个变量在所有模型中出现“1”时所对应的储量和, 即

$$b_k = \sum_{i=1}^n x_{ik} \cdot g(g_i)$$

式中, x_{ik} 为第 k 个变量在第 i 个模型单元上的反映; $g(g_i)$ 为所对应模型单元储量的单调增函数。 b_k 越大, 表明该变量对控制模型资源方面越重要, 越具普遍性, 取

$$B_k = b_k / \sum_{j=1}^m b_j \quad k=1, 2, \dots, m$$

其中, $0 \leq B_k \leq 1$, 当 $B_k = 1$ 时, $b_j = 0 (j \neq k)$, 则表明第 k 个变量与资源量大小显著呈正相关; 反之, $B_k = 0$ 时, 第 k 个变量不控制资源量大小。

3. D_k 的求法: 令 $d_k = b_k / a_k$, 表示第 k 个变量在所有模型中为“1”时的平均储量。如果某两个变量在模型中存在时的数量相同, 而出现在不同规模的模型单元, 其作用不同。 d_k 越大, 则变量的作用也越大, 取

$$D_k = d_k / \sum_{j=1}^m d_j \quad k=1, 2, \dots, m$$

其中, D_k 的大小与 B_k 相似。

对每个变量都求出 3 个因素的贡献后, 取得该变量的权系数 R_k 。

设预测单元相似性评价模型采用下式:

$$f_i = \sum_{k=1}^m R_k x_{ik} \quad (1)$$

式中, f_i 为第 i 个预测单元与模型单元的联系度; x_{ik} 为第 k 个地质变量在第 i 个单元的取值, $k=1, 2, \dots, m$; R_k 为第 k 个地质变量的变量权。

模型分析与改进

通过对模型的分析, 笔者认为不应把 A_k 、 B_k 、 D_k 放在 1 个等级意义中求和。因为 A_k 是反映变量对单元成矿的控制作用, 是描述变量对单元成矿的重要程度, 仅能说明矿产资源的定位预测, 而 B_k 、 D_k 则是表明变量对资源量变化作用的贡献。因此, 在进行矿产预测时, 笔者根据所研究的目的, 采用以下方法进行计算。

A_k 与上述相同。

B'_k 表示第 k 个变量存在时变量对单元控制作用的相对大小。每个单元中有用地质条件之和记为

$$x_i = \sum_{j=1}^m x_{ij} \\ i=1, 2, \dots, n \\ (\text{单元数})$$

它反映了第 i 个单元的代表性。以单元的不同代表性 x_i 为权, 求各模型单元中变量出现的加权和, 记为

$$b'_k = \sum_{i=1}^n x_{ik} \cdot x_i \\ k=1, 2, \dots, m \\ (\text{变量数})$$

b'_k 值依赖于第 k 个变量与相应模型单元代表性的关系, 只有变量出现的次数多, 而且所出现的单元具有较大代表性, b'_k 值大。可见 b'_k 的大小反映了变量的重要性, 为统一量纲取

$$B'_k = b'_k / \sum_{j=1}^m b'_j$$

其中, $0 \leq B'_k \leq 1$, 当 $B'_k = 1$ 时, $b'_j = 0 (j \neq k)$; 反之, $B'_k = 0$ 时, $b'_k = 0$, 这时变量在单元中不出现, 显然变量不具有控矿作用。

D'_k 的求法: 令 $d'_k = b'_k / a_k$ 表示第 k 个变量在所有模型中出现时的平均控矿作用, d'_k 的大小主要取决于变量所发生单元的代表性。相同发生的变量在代表性较大的单元中出现的变量 d'_k 值大; 相反, 在代表性差的单元中出现的变量 d'_k 值小, 取

表 1

平方和法						三因素法					
变量	变量权	排序	变量	变量权	排序	变量	变量权	排序	变量	变量权	排序
1	0.0134185	27	17	0.0054915	31	1	0.048	27	17	0.035	30
2	0.022603	24	18	0.0312409	15	2	0.059	23	18	0.103	14
3	0.0295168	19	19	0.0069845	30	3	0.094	18	19	0.055	25
4	0.0590932	1	20	0.0127055	28	4	0.1552	2	20	0.032	31
5	0.0496847	5	21	0.0496298	4	5	0.146	4	21	0.152	3
6	0.0282851	21	22	0.048099	7	6	0.056	24	22	0.146	5
7	0.0357509	13	23	0.0317008	16	7	0.106	11	23	0.120	8
8	0.0493251	6	24	0.0311646	17	8	0.132	6	24	0.041	29
9	0.0145598	26	25	0.0396784	10	9	0.046	28	25	0.106	12
10	0.0499683	3	26	0.0506120	2	10	0.162	1	26	0.111	10
11	0.0265712	23	27	0.0214141	25	11	0.102	15	27	0.075	22
12	0.0397007	9	28	0.0297672	18	12	0.113	9	28	0.088	20
13	0.0387167	11	29	0.0267376	22	13	0.088	19	29	0.084	21
14	0.0286125	20	30	0.0117252	29	14	0.104	13	30	0.052	26
15	0.0386709	12	31	0.0322830	14	15	0.094	17	31	0.096	16
16	0.0465306	8				16	0.132	7			

$$D'_k = d'_k / \sum_{j=1}^m d_j = \sum_{j=1}^m \sum_{k=1}^n x_{kj} \cdot x_{kj} = \sum_{j=1}^m r_{1j}$$

按以上方法求出 A'_k 、 B'_k 、 D'_k 三因素的贡献后, 再按下式求变量权

$$R'_k = A'_k + B'_k + D'_k \quad (2)$$

在特征分析中, 采用平方和法求变量权的计算公式为:

$$a_i = \sum_{j=1}^m r_{ij}^2$$

$i=1, 2, \dots, n$ (样品数)

$j=1, 2, \dots, m$ (变量数)

其中, r_{ij} 为乘积矩阵 $R = X'X$ 的元素, 显然

$$r_{ij} = \sum_{k=1}^n x_{ki} \cdot x_{kj}$$

$i, j=1, 2, \dots, m$ (变量数)

注意到

$$b'_k = \sum_{i=1}^n x_{ki} \cdot x_k = \sum_{i=1}^n x_{ki} \cdot \sum_{j=1}^m x_{kj}$$

可见 b'_k 与 a_i 均为 $r_{11}, r_{12}, \dots, r_{1m}$ 的某种和, 差别在于前者为直接求和, 而后者为平方和, 当然它们所反映的变量特点是可比的。此外三因素法中还考虑了 A'_k 和 D'_k , 因此, 其计算的结果与特征分析具有某种相似性。

由表 1 计算结果表明, 两种方法计算变量权对变量的度量是可比的, 排序基本一致, 仅局部相邻的变量排序略有不同, 差异表明两者出发点不同。

模型再改进及资源 定量预测

上述已讨论了三因素统计法的特点, 笔者对该法又重新建立计算系统, 使其能适用于定量预测。

设有 n 个模型单元, 相应资源量为 g_i , $i=1, 2, \dots, n$, 有 m 个控制资源量分布的变量 x_j , $j=1, 2, \dots, m$, 每个变量 x_j 的

取值为

$$x_{ij} = \begin{cases} 1 & \text{当第 } j \text{ 个变量在第 } i \text{ 个单元上} \\ & \text{有反映} \\ 0 & \text{否则} \end{cases}$$

b_k 定义为

$$b_k = \sum_{i=1}^n g(g_i) x_{ik}$$

其中, $g(g_i)$ 为 g_i 的单调增加函数, 如可选择 $g(g_i) = g_i$ 或 $g(g_i) = \ln(g_i)$ 等。

b_k 反映变量 x_k 出现时所对应资源量大小, 可作为变量对资源控制作用的一种度量, 取

$$B_k = b_k / \sum_{j=1}^m b_j \quad k=1, 2, \dots, m$$

因 b_k 值大小依赖于变量出现的次数以及变量所对应单元的资源量, 故 b_k 与变量出现次数呈正相关。为消除变量出现次数的影响及考查变量所出现的单元资源量之间的关系, 下面给出度量这种变量作用的指标 d_k

$$d_k = b_k / a_k$$

其中, $a_k = \sum_{i=1}^n x_{ik}$, $k=1, 2, \dots, m$, 为统一量纲, 取

$$D_k = d_k / \sum_{j=1}^m d_j \quad k=1, 2, \dots, m$$

变量对资源量的变化影响作用的度量——变量权, 记为

$$R_k = B_k + D_k \quad k=1, 2, \dots, m \quad (3)$$

类似可求得单元联系度为

$$f_i = \sum_{k=1}^m R_k \cdot x_{ik} \quad i=1, 2, \dots, n \quad (4)$$

容易理解 f_i 与资源量呈正相关。根据对已知单元计算的联系度 f , 可进行未知单元资源量的估计, 即达到了预测资源目的。

实 例

笔者在1987年应用综合信息矿产评价方法, 对河北北部显示负磁场的中酸性花岗岩

岩体地球物理、地球化学特征以及与该类型岩体有关金矿控矿因素特征进行了研究, 并建立起该类型金矿的综合信息找矿模型, 提取了31个主要控矿变量和矿化显示标志, 划分出35个单元, 其中选取17个模型单元和18个预测单元。在单元中, 出现地质变量赋值为“1”, 不出现者赋值为“0”, 即形成1个 35×31 阶原始数据矩阵。

根据原始数据矩阵, 求得3个因素 A_k 、 B'_k 、 D'_k 以及变量权的大小, 见表2。

表 2

变量	a_k	A_k	b'_k	B'_k	d'_k	D'_k	R'_k
1	1	0.004	19	0.005	19.0	0.039	0.048
2	3	0.013	47	0.014	15.7	0.032	0.059
3	9	0.040	108	0.030	12.0	0.024	0.094
4	14	0.060	228	0.066	16.3	0.033	0.155
5	14	0.060	200	0.056	14.3	0.030	0.146
6	2	0.009	36	0.010	18.0	0.037	0.056
7	10	0.040	139	0.040	13.9	0.026	0.106
8	11	0.050	176	0.050	16.0	0.032	0.132
9	2	0.009	30	0.008	15.0	0.030	0.046
10	15	0.070	218	0.063	14.5	0.029	0.162
11	17	0.075	125	0.036	17.9	0.036	0.102
12	9	0.040	141	0.041	15.7	0.032	0.113
13	5	0.020	97	0.028	19.4	0.040	0.088
14	6	0.026	125	0.036	20.8	0.042	0.104
15	7	0.030	111	0.032	15.9	0.032	0.094
16	11	0.050	175	0.050	15.9	0.032	0.132
17	2	0.009	20	0.006	10.0	0.020	0.035
18	8	0.035	125	0.036	15.6	0.032	0.103
19	2	0.009	39	0.010	17.5	0.036	0.055
20	3	0.013	20	0.006	6.67	0.013	0.032
21	14	0.060	216	0.062	15.4	0.032	0.152
22	13	0.060	194	0.056	14.9	0.030	0.146
23	11	0.050	149	0.043	13.5	0.027	0.120
24	4	0.020	66	0.020	16.5	0.033	0.041
25	6	0.026	127	0.037	21.2	0.043	0.106
26	8	0.035	143	0.040	17.9	0.036	0.111
27	5	0.020	80	0.023	16.0	0.032	0.075
28	5	0.020	97	0.028	19.4	0.040	0.088
29	6	0.026	92	0.027	15.3	0.031	0.052
30	2	0.009	33	0.010	16.5	0.033	0.052
31	7	0.030	116	0.033	16.6	0.033	0.096
Σ	227		3463		492.9		

把(2)式代入(1)式中求得单元之间

联系度 f_i ，计算结果见表3。

表 3

单元	联系度 (f)	单元	联系度 (f)	单元	联系度 (f)	单元	联系度 (f)
1	1.936	10	1.966	19	1.742	28	0.604
2	1.862	11	1.844	20	1.274	29	1.842
3	0.847	12	1.341	21	1.383	30	1.266
4	1.045	13	1.864	22	1.022	31	1.194
5	1.507	14	1.936	23	1.453	32	1.441
6	1.673	15	1.936	24	1.744	33	1.576
7	1.872	16	1.844	25	1.585	34	0.895
8	1.833	17	1.560	26	1.877	35	1.065
9	1.117	18	1.246	27	1.675		

对17个模型单元之间的联系度 f_i ($i=1, 2, \dots, n$) 作频率分布直方图 (图1)。

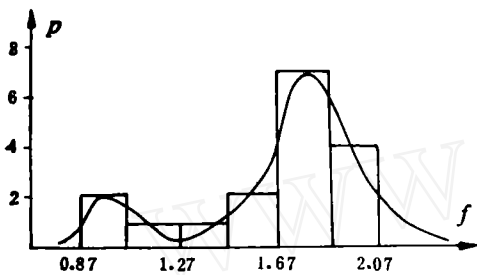


图 1 单元联系度(f)分布直方图

从 f 分布直方图中，可看出它为一近似正态分布形式。对于给定值 α ($\alpha=0.05$)，除

3、4、9模型单元外，其余都满足 $p(f_i) \geq \alpha$ (其中 p 为以 μ 为均值， σ 为方差的正态概率函数)，表明除这3个模型单元外，其它单元同属于相同母体的置信区间内，即两者不属于同一母体。通过对这3个单元的成矿地质环境、矿床特征等因素分析，确与其它类型不同，因此在预测中予以剔除。在18个预测单元中，以19、24、25、26、27、29、33等7个单元所求得联系度 f 值相对较高，与

表 4

变量	变 量 意 义	R_k'
10	单元内发育北东向控矿断裂	0.162
4	地层岩性为斜长片麻岩、角闪斜长片麻岩	0.155
21	矿化距岩体5km之内	0.152
5	地层构造片理化带发育与区域断裂构造一致	0.146
22	单元内脉岩发育、种类多	0.146
8	主要控矿断裂构造为东西或近东西向	0.132
16	单元内发育有负磁场岩体	0.132
23	单元内脉岩与金矿化密切	0.120
26	单元内重砂组合异常有金、多金属矿物显示	0.111
25	单元内重砂组合异常有金、白钨矿、辰砂、辰砂在外带	0.106
14	单元内控矿构造为2组	0.104
18	矿化形成岩体的倾伏端或侧伏端	0.103
11	单元内北西向控矿构造发育	0.102
31	单元内分散流有Cu异常显示	0.10
15	单元内控矿构造有3组以上	0.10

表 5

变量	b_k	B_k	d_k	D_k	R_k	变量	b_k	B_k	d_k	D_k	R_k
1	4.050	0.082	4.05	0.232	0.314	16	2.610	0.053	0.237	0.014	0.067
2	3.560	0.072	1.78	0.102	0.174	18	-0.090	-0.002	-0.011	0.0006	-0.003
3	-0.446	-0.009	-0.05	-0.0028	-0.012	19	1.750	0.035	0.875	0.050	0.085
6	5.461	0.110	2.731	0.157	0.267	24	7.611	0.153	1.903	0.109	0.262
7	-2.90	-0.058	-0.290	-0.017	-0.075	25	5.551	0.112	0.925	0.053	0.180
8	0.475	0.010	0.043	0.0025	0.012	26	5.271	0.106	0.659	0.038	0.144
9	2.150	0.042	1.075	0.061	0.103	27	-1.139	-0.024	0.227	-0.053	-0.037
11	2.97	0.060	0.424	0.024	0.084	28	6.672	0.134	1.334	0.076	0.210
12	-2.019	-0.041	-0.224	-0.013	-0.054	29	-5.61	-0.113	-1.122	-0.064	-0.177
13	1.710	0.034	0.342	0.02	0.084	30	3.58	0.072	1.790	0.103	0.175
14	-0.649	-0.013	-0.108	0.006	-0.019	31	5.871	0.118	0.840	0.048	0.166
15	3.280	0.046	0.069	0.027	0.093	Σ	49.677		17.44		

模型单元中型矿田的联系度近似,表明这些预测单元成矿条件较好,应引起重视。

依据(2)式所求出的变量权 R_i^* ,从中选出15个 $R_i^* \geq 0.1$ 的地质变量列于表4。

表4基本反映了本区该类型金矿的主要控矿因素和矿化标志。

表 6

单 元	联 系 度 (f)	估 计 值 (t)	模型单元 储量对数值
1	1.968	40.000	4.05
2	1.205	5.590	1.411
5	1.263	6.500	0.070
6	1.258	6.415	2.080
7	-0.044	0.220	-2.300
8	0.605	1.170	-0.510
10	0.542	1.011	1.500
11	0.599	1.171	-1.390
13	0.339	0.600	0.390
14	0.079	0.300	0.990
15	0.079	0.306	0.550
16	0.057	0.289	-0.090
17	0.059	0.291	-1.390
18	1.592	15.185	
19	1.318	7.49	
20	1.678	18.958	
21	1.809	26.582	
23	0.896	2.394	
24	1.102	4.289	
25	0.310	0.556	
26	0.328	2.115	
27	0.198	0.903	
29	1.636	17.011	
30	1.023	3.499	
32	1.129	4.599	
33	1.189	5.369	

在应用(1)、(2)式对成矿单元进行定位预测的基础上,笔者剔除了3、4、9号模型单元和不在母体置信区间范围内的22、28、31、34、35号预测单元。在重新筛选出14个模型单元中,选出23个变量形成 14×23 阶原始数据矩阵,再根据(3)、(4)式建立矿产资源定量预测模型,并对12个预测矿田单元进行预测,其计算结果见表5和表6。

将模型单元联系度(f_i)与资源量对数值 $\ln(g_i)$ 建立一元回归定量预测方程

$$\ln(g_i) = -1.387 + 2.58f_i$$

$$i=1, 2, \dots, n \text{ (单元数)}$$

其中, g_i 为第 i 个单元所对应的资源量; f_i 为第 i 个单元所对应的联系度(预测曲线见图2)。该方程的相关系数(r)为0.80,可知 $\ln(g_i)$ 与 f_i 相关性显著,证明建立的一元回归定量预测方程可靠。

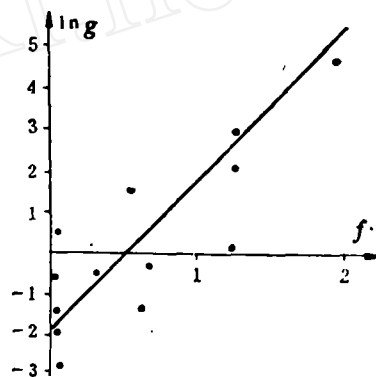


图 2 三因素预测曲线

本文撰写过程中,得到长春地质学院王世称、成秋明先生的指导,在此致以谢意。

An Improved Mathematical Model of Three-Factor

Statistical Method and Its Application to

Mineral Resources Prediction

Wang Quanming

The three-factor statistical method, being a multivariate analysis one, is mainly used for studying the distribution of each variable for a group of samples, the representativeness of each sample and the relationship between above two factors. For this method a discussion and some improvements are made in this paper. Application of this method to mineral resources prediction has been achieved satisfactory results.