文章编号：1001-1749(2014)05-0626-08

# Excel 在地球化学数据处理中的高级应用

谭亲平[1,2]，夏 勇[1*]，谢卓君[1,2]，闫 俊[1,2]
（1. 中国科学院 地球化学研究所，贵阳 550002；2. 中国科学院大学，北京 100049）

**摘 要**：Excel 强大的图表绘制和数据计算能力，为地球化学数据的处理提供了便利。在研究地球化学数据处理原理的基础之上，详细解释了三角图解，频率直方图，概率格纸图解法求异常下限，多重分形法计算异常下限，R 型聚类分析和判别分析的具体计算步骤。这些方法有利于地质科技方面的研究。

**关键词**：三角图解；异常下限；直方图；聚类分析；判别分析
**中图分类号**：P 632 **文献标志码**：A **DOI**：10.3969/j.issn.1001-1749.2014.05.21

## 0 引言

在地质工作中，常常需要计算各种数据，绘制各种图表[1-3]，目前尚无统一的软件能够满足地质工作当中所有的数据处理的需求。Excel 为大众软件，地质科技人员较为熟悉，并有一定的操作经验。它的强大的图表绘制和数据计算能力基本能满足地质工作中的数据处理需求。作者以 Microsoft Excel 2010 为界面，以文献和作者在工作中的数据为例，详细解释各种地质数据处理的原理和具体操作步骤，它适用于基层地质科技人员的工作性质和工作条件。本研究涉及了 Excel 的高级用法，需要了解相关的多元统计学知识，并能熟练操作 Excel 才能合理地进行运用。

## 1 三角图解

Microsoft Excel 已经提供了大量的图表类型，但在地学中需要经常应用的三角图解却没有提供。在 Excel 中，可以通过将三角坐标转化为直角坐标的办法来“迂回”实现。三角图解有 $a$、$b$、$c$ 三个轴，它们的值范围都是 0～100，且满足 $a+b+c=100$。通过坐标变换可将三维坐标系$(a,b,c)$转换为二维直角坐标系$(X,Y)$，然后再利用转换得来的直角坐标绘制散点图即可实现三角图解的绘制。

图 1 中直角坐标的原点设为 $BC$ 的中点，任意点 $D$ 在三角图解中的坐标设为$(a,b,c)$。根据几何学原理，任意点 $D$ 在直角坐标中的坐标是：$X=(b+c)/2-b$，$Y=\sin 60^\circ * a$。设 $A$、$B$、$C$ 三个端点的坐标分别为：$A(100,0,0)$、$B(0,100,0)$、$C(0,0,100)$，根据上面的公式计算出它们的直角坐标分别为：$A(0,86.60254)$、$B(-50,0)$、$C(50,0)$。在 Excel 中利用三个端点的直角坐标，每两个点绘制带直线的散点图，就能得到三角图解的外框。将需要处理的数据，根据同样的公式，也转换成直角坐标，然后添加为散点图即可。

下面通过一个实际的例子来说明三角图解的绘制过程。表 1 中的原始数据来自文献[4]。以表 1 中的第 2 排数据为例，在 D2 中输入函数：=100/(A2 + B2 + C2)，即可得到比例因子，再将每个变量乘以比例因子，可在E2:G2中得到归一化后

收稿日期：2014-03-07 改回日期：2014-08-19
基金项目：国家重点基础研究发展计划(973 计划)(2014CB440905)；矿床地球化学国家重点实验室" 十二五" 项目群课题(SKLODG－ZY125－01)
作者简介：谭亲平(1986- )，男，博士，从事构造地球化学研究，E-mail：565310821@qq.com
*通讯作者：夏勇(1960- )，男，博士生导师，从事矿床地球化学研究，E-mail：xiayong@vip.gyig.ac.cn

表 1　三角图解数据表

Tab. 1　Data table of triangular chart

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | | 原始数据 | | 比例 | | 归一化后 | | $x$ 轴 | $y$ 轴 |
| 2 | 12.7 | 14.7 | 20.1 | 2.105263 | 26.737 | 30.947 | 42.316 | 5.684 | 23.154 |
| 3 | 14.6 | 18.3 | 17.0 | 2.004008 | 29.259 | 36.673 | 34.068 | −1.303 | 25.338 |
| 4 | 13.8 | 15.8 | 16.3 | 2.178649 | 30.065 | 34.423 | 35.512 | 0.545 | 26.037 |
| 5 | 11.9 | 11.8 | 21.3 | 2.222222 | 26.444 | 26.222 | 47.333 | 10.556 | 22.901 |

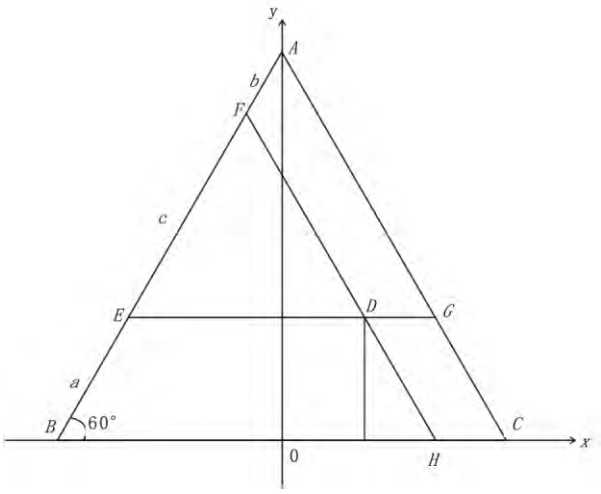


图 1　三角坐标转直角坐标原理图

Fig. 1　Schematic diagram of traingular coordinate transform into rectangular coordinate

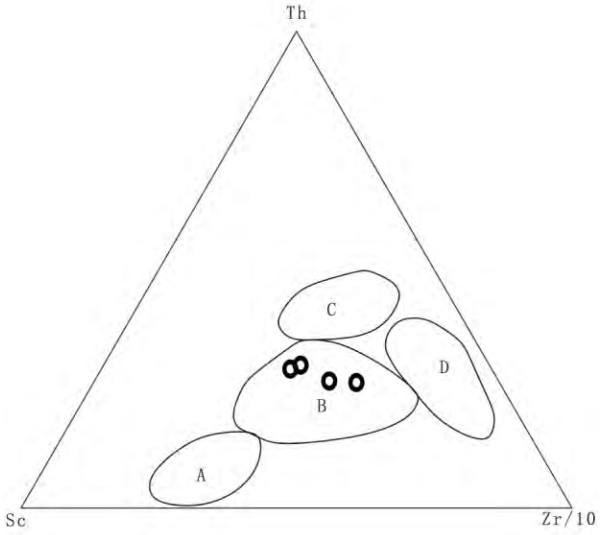


图 2　三角图解应用实例图

Fig. 2　Application example of triangular chart

的值。归一化后使三个变量之和等于 100。在 H2 和 I2 中分别输入函数：=(F2+G2)/2−F2 和=E2*√3/2 得到相对应的直角坐标，将得到的直角坐标添加到前面已生成的三角形外框中，在生成的图中删除直角坐标轴，插入三个文本框，在文本框中填入三个端元符号。图 2 是应用本文的方法得到的图，它与文献[4]中的图重合。该方法的关键为三角坐标向直角坐标的转换原理。

## 2　频率直方图

频率直方图在 Excel 中的绘制过程，可利用实例来说明。假如有表 2 的 A 列中的一组数据需要绘制频率直方图，首先在 B3 和 B7 中分别输入函数：=MIN(A2:A13)和=MAX(A2:A13)，计算出该组数据的极值。根据极值和实际需要确定分组间隔，本次分组间隔为 2。在 C2 中输入极小值整数位，在 C3 中输入：=C2+2，选中 C2:C3，按住鼠标左键将其拖拽至极大值即可得到分组。然后应用“直方图”工具就能绘制频率直方图。首次使用分析工具需加载计算工具：文件－选项－自定义功能区在“开发工具”选项前打勾－确定，然后选择开发工具－加载项－选中“分析工具库”和“规划求解”加载项－确定。工具加载完成后：数据－数据分析－直方图，就能出现频率直方图对话框(图 3)。

表 2　频率直方图数据表

Tab. 2　Data table of frequency histogram

| | A | B | C | D |
|---|---|---|---|---|
| 1 | 原始数据 | 极值 | 分组 | 频率 |
| 2 | 6.5 | 最小值 | 0 | 0 |
| 3 | 9.5 | 0.5 | 2 | 1 |
| 4 | 12.6 | | 4 | 2 |
| 5 | 0.5 | | 6 | 1 |
| 6 | 6.0 | 最大值 | 8 | 2 |
| 7 | 11.7 | 12.6 | 10 | 4 |
| 8 | 8.6 | | 12 | 1 |
| 9 | 7.9 | | 14 | 1 |
| 10 | 2.9 | | | |
| 11 | 9.8 | | | |
| 12 | 8.8 | | | |
| 13 | 3.6 | | | |

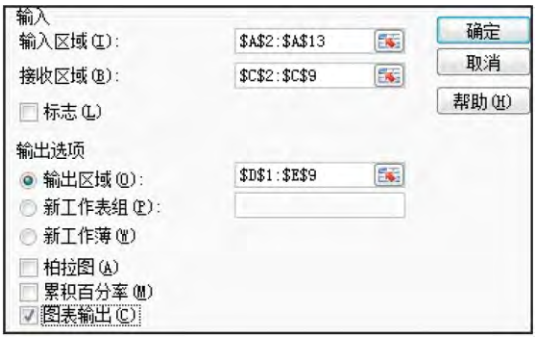


**图 3 频率直方图对话框**

Fig. 3 Dialog box of frequency histogram

**图 3 中“输入区域”需要输入原始数据:=A2:A13,在“接收区域”需要输入分组数据:=C2:C9,输出选项中选中“输出区域”,输入任意空白区域,本文实例中输入:=D1:E9,选择“图表输出”,最后确定,就能得到 D 列数据和相对应的频率直方图。如需要将多组数据在一张图中表示,只要将每组数据经过上面的步骤处理,然后在某一频率直方图中添加数据即可:右击—选择数据—添加。该方法的关键是分析工具库中直方图的使用技巧。**

## 3 概率格纸图解法求异常下限

**概率格纸图解法确定背景值和异常下限,是建立在元素在地质体中呈正态分布或对数正态分布的基础上。应用这种方法时,统计元素各个分组区间内的累积频率,并在概率格纸上绘出各个累积频率分布点的连线,然后根据其在概率格纸上反映的正态分布(或对数正态分布)特点,确定背景值及异常下限。其在 Excel 中的具体做法和步骤如下。**

**图 4 是在 Excel 中绘制的概率格纸,其中纵坐标是正态分布累积频率的反函数值(实际标记为累积频率),横坐标是元素含量的对数值(数据已通过正态检验)。表 3 中 A、B、C 和 D 列是绘制概率格纸横向网格的数据,在 A 和 C 列中假设一组累积频**

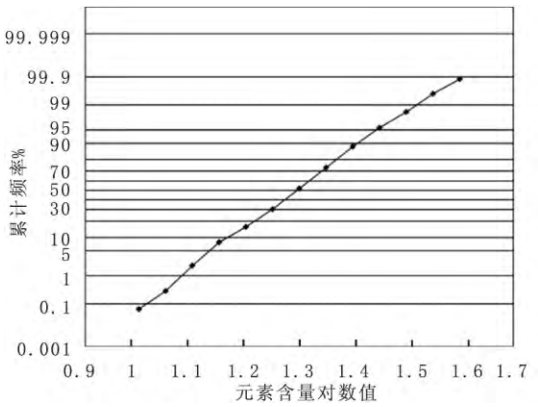


**图 4 元素含量累积频率分布图**

Fig. 4 Chart of cumulative frequency distribution of element content

**表 3 概率格纸法求异常下限数据表**

Tab. 3 Data table of probability ruling paper method to calculate anomaly threshold

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 累积频率/% | 反函数值(y) | 累积频率/% | 反函数值(y) | 原始数据 | 分组 | 频数 | 累积频率/% | 反函数值 |
| 2 | 0.001 | −4.26 | 60 | 0.25 | 0.978 | 1 | 1 | 0.06 | −3.23 |
| 3 | 0.001 | −4.26 | 60 | 0.25 | 1.021 | 1.05 | 4 | 0.30 | −2.74 |
| 4 | 0.1 | −3.09 | 70 | 0.52 | 1.061 | 1.1 | 28 | 2.01 | −2.05 |
| 5 | 0.1 | −3.09 | 70 | 0.52 | 1.064 | 1.15 | 96 | 7.85 | −1.42 |
| 6 | 1 | −2.33 | 80 | 0.84 | 1.068 | 1.2 | 134 | 16.00 | −0.99 |
| 7 | 1 | −2.33 | 80 | 0.84 | 1.068 | 1.25 | 236 | 30.35 | −0.51 |
| 8 | 5 | −1.64 | 90 | 1.28 | …… | 1.3 | 355 | 51.95 | 0.05 |
| 9 | 5 | −1.64 | 90 | 1.28 | 1.509 | 1.35 | 350 | 73.24 | 0.62 |
| 10 | 10 | −1.28 | 95 | 1.64 | 1.515 | 1.4 | 249 | 88.38 | 1.19 |
| 11 | 10 | −1.28 | 95 | 1.64 | 1.525 | 1.45 | 119 | 95.62 | 1.71 |
| 12 | 20 | −0.84 | 99 | 2.33 | 1.526 | 1.5 | 45 | 98.36 | 2.13 |
| 13 | 20 | −0.84 | 99 | 2.33 | 1.540 | 1.55 | 20 | 99.57 | 2.63 |
| 14 | 30 | −0.52 | 99.9 | 3.09 | 1.542 | 1.6 | 5 | 99.88 | 3.03 |
| 15 | 30 | −0.52 | 99.9 | 3.09 | 1.555 | >1.6 | 2 | 100.00 | |
| 16 | 40 | −0.25 | 99.999 | 4.26 | 1.558 | | | | |
| 17 | 40 | −0.25 | 99.999 | 4.26 | 1.610 | 最大 X 轴 | 1.70 | | |
| 18 | 50 | 0.00 | | | 1.623 | 最小 X 轴 | 0.90 | | |
| 19 | 50 | 0.00 | | | | 样品数 | 1644 | | |

率，在 B 和 D 列中计算各自对应的反函数值，比如在 B2 中输入函数：＝NORMSINV(A2%)。函数 NORMSINV(probability)返回标准正态分布累积函数的反函数值。

以纵坐标为 0.1%的横向网格线为例，介绍横向网格的绘制方法。首先绘制带直线的散点图：插入－散点图－带直线和数据标记的散点图－在图中右击－选择数据－添加，出现“编辑数据系列”对话框，在“系列名称”中输入：＝A4（累积频率值 0.1%），在“$X$ 轴系列值”中输入：＝G17：G18（图 4 中 $X$ 轴的极大值和极小值），在“$Y$ 轴系列值”中输入：＝B4：B5，点击“确定”后出现一条横向直线。选中该直线右击，选“添加数据标签”，在直线端点处出现标签数据，选中该标签数据右击，选择“设置数据标签格式”并出现对话框，在“标签包括”里只选择“系列名称”，在“标签位置”中选择“靠左”，关闭后删除右端点的数据标签即可。其他的网格线利用此方法一一添加，并将 $X$ 轴的极大值和极小值固定为 G17：G18 中的数值，删掉 $Y$ 轴坐标即可。

概率格纸绘制好之后，将需要计算的数据添加到概率格纸中，就可以计算异常下限和背景值。表 3 中 E、F、G、H、I 列中的数据是数据的处理过程，其中分组和频数的计算参考上面频率直方图的方法。H 列中累积频率的计算以 H2 和 H3 为例分别输入函数：＝G2/1644＊100，＝G3/1644＊100＋H2。I 列中利用函数 NORMSINV(probability)返回累积频率的反函数值。所有数据计算好之后，将数据添加到已绘制好的概率格纸中，$X$ 轴输入：＝F2：F14（分组），$Y$ 轴输入：＝I2：I14（实际应为反函数值，但 $Y$ 轴标记为累积频率），适当调整后就能得到图 4。图 4 中连线与累积频率为 50%线的交点的横坐标为背景值，连线上累积频率为 97.7%的点横坐标即为异常下限。该方法的关键在于概率格纸的绘制，但绘制好之后可以多次使用，以后只需将新的数据添加在已绘制好的格纸上即可。

## 4　多重分形法计算异常下限

目前利用分形技术进行地球化学异常下限确定的方法主要有：含量－周长法、含量－面积法、含量－距离法、含量－频数法等。这里采用含量－频数法，设分形求和模型：$N(C_i)=kC_i^{-D}(i>0)$，式中 $C_i$ 为元素含量，又称特征尺度，$k$ 为比例常数($k>0$)，$D$ 为一般分维数，$N(C_i)$为当元素含量为 $C_i$ 时所有大于等于 $C_i$ 的元素含量的和数。分形求和模型两边分别取对数得到一元线性回归模型：$\lg N(C_i)=-D\lg(C_i)+\lg(k)$，用最小二乘法求出斜率 $D$ 的量，即为分维数。Excel 中多重分形法的计算过程如下。

表 4 中对 A 列中的数据进行分组并计算频数，C 列对 B 列中的分组数据求对数，如在 C2 中输入：＝LOG10(B2)，F 列中为当元素含量为 $C_i$ 时所有大于等于 $C_i$ 的元素含量的和数，如在 F5 中输入函数：

表 4　多重分形法数据表

Tab.4　Data table of multifractal method

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | 数据 | 分组 | lg(*Ci*) | 分组区间 | 频数 | $N(C_i)$ | $\lg N(C_i)$ |
| 2 | 0.978 | 1 | 0.000 | ＜1 | 1 | 1644 | 3.22 |
| 3 | 1.021 | 1.05 | 0.021 | 1～1.05 | 4 | 1643 | 3.22 |
| 4 | 1.061 | 1.1 | 0.041 | 1.05～1.1 | 28 | 1639 | 3.21 |
| 5 | 1.064 | 1.15 | 0.061 | 1.1～1.15 | 96 | 1611 | 3.21 |
| 6 | 1.068 | 1.2 | 0.079 | 1.15～1.2 | 134 | 1515 | 3.18 |
| 7 | 1.068 | 1.25 | 0.097 | 1.2～1.25 | 236 | 1381 | 3.14 |
| 8 | …… | 1.3 | 0.114 | 1.25～1.3 | 355 | 1145 | 3.06 |
| 9 | 1.526 | 1.35 | 0.130 | 1.3～1.35 | 350 | 790 | 2.90 |
| 10 | 1.540 | 1.4 | 0.146 | 1.35～1.4 | 249 | 440 | 2.64 |
| 11 | 1.542 | 1.45 | 0.161 | 1.4～1.45 | 119 | 191 | 2.28 |
| 12 | 1.555 | 1.5 | 0.176 | 1.45～1.5 | 45 | 72 | 1.86 |
| 13 | 1.558 | 1.55 | 0.190 | 1.5～1.55 | 20 | 27 | 1.43 |
| 14 | 1.610 | 1.6 | 0.204 | 1.55～1.6 | 5 | 7 | 0.85 |
| 15 | 1.623 | 1.65 | 0.217 | ＞1.6 | 2 | 2 | 0.30 |

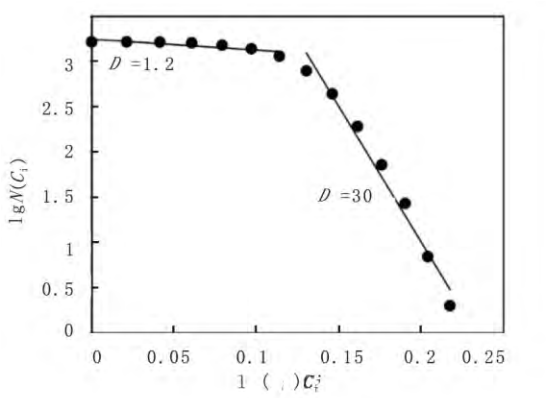


图 5 元素含量－频数双对数曲线

Fig. 5 Chart of element content frequency double logarithmic curve

＝E5＋F6，G 列对 F 列中的数据求对数，如在 G2 中输入：＝LOG10(F2)。然后绘制散点图：$X$ 轴中输入：＝C2:C15，$Y$ 轴中输入：＝G2:G15，其散点大致分布在两段直线上，同时在图中可确定两段直线的分界点。根据分界点，两段直线在表 4 中的坐标数据分别为：$X$＝C2:C8，$Y$＝G2:G8 和 $X$＝C9:C15，$Y$＝G9:G15。确定好两段直线的坐标数据之后，重新绘制散点图，分别输入两段散点的 $X$、$Y$ 坐标，生成两段散点，然后分别选中一段散点，右击弹出下拉菜单，选择“添加趋势线”，弹出“设置趋势线格式”对话框，在“趋势线选项/回归分析类型”中选择“线性”，同时在“显示公式”前的方框中打勾，最后关闭即可。该方法的关键在于理解多重分形法的原理。

## 5 R 型聚类分析

R 型聚类分析是根据样品的多种变量的测定数据进行数字分类，定量地确定变量之间的亲疏程度[5]。进行数字分类，需要选择合适的数量指标，以此衡量样品之间的亲疏程度。数量指标主要有距离系数，相似系数和相关系数。本研究用相关系数作相似性统计量的逐步计算形成法做 R 型聚类分析，将通过一个实际例子，说明在 Excel 中聚类分析的计算步骤。

由于各变量的单位，量级和数值变动范围的差异很大，计算中往往突出了那些绝对值较高的变量。因此，在进行聚类分析之前需将各个变量换算成一致的相对值。常用的变换的方法有标准化和正规化，本例中选择标准化。例如表 5 中 B2:G7为原始数据，在第 8 和第 9 行中分别计算均值和标准差，比如分别在 B8、B9 中输入函数：＝AVERAGE(B2:B7)和＝STDEV(B2:B7)，然后选中拖动即可全部算出。在 B11:G16 中计算出标准化数据，比如在 B11 中输入函数：＝(B2－B8)/B9。标准化之后的数据，均值为 0，标准差为 1。数据标准化之后即可计算相关矩阵：数据－数据分析－选中“相关系数”－确定，然后弹出“相关系数”对话框(图 6)，在“输入区域”选中 B10:G16，在“分组方式”中选择逐列，选中“标志位于第一行”，在“输出区域”中选择 A19，确定之后会得到表 5 中 A9:G25 的相关矩阵。从相关矩阵中可以看出 Cu 和 Co 相关系数最大，因此首先将 Cu 和 Co 连为一组。逐步加权平均，即新的 CuCo＝(Cu＋Co)/2，计算修正数据，将得到的新数据 CuCo 替换 Cu 和 Co，并与其他数据一起，计算新的相关矩阵，直到所有元素均已分组完成(表 6)。最后在绘图软件中根据上面的计算结果，绘制出谱系图即可(图 7)。

表 5 聚类分析数据表

Tab. 5 Data table of R cluster analysis

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | 元素 | Ni | Co | Cu | Cr | S | As |
| 2 | | 3.28 | 2.44 | 2.20 | 3.07 | 3.91 | 0.60 |
| 3 | | 3.37 | 1.90 | 0.78 | 3.50 | 2.77 | 1.15 |
| 4 | | 2.87 | 1.42 | 0.10 | 2.92 | 2.63 | 0.48 |
| 5 | 数据 | 3.44 | 2.44 | 2.18 | 3.38 | 3.02 | 1.57 |
| 6 | | 3.25 | 1.97 | 1.11 | 3.50 | 1.73 | 0.10 |
| 7 | | 3.02 | 1.64 | 0.77 | 3.32 | 2.02 | 0.60 |
| 8 | 均值 | 3.21 | 1.97 | 1.19 | 3.28 | 2.68 | 0.75 |
| 9 | 标准差 | 0.22 | 0.41 | 0.84 | 0.24 | 0.77 | 0.52 |
| 10 | 元素 | Ni | Co | Cu | Cr | S | As |
| 11 | | 0.34 | 1.14 | 1.20 | −0.90 | 1.60 | −0.28 |
| 12 | | 0.74 | −0.17 | −0.49 | 0.93 | 0.12 | 0.76 |
| 13 | | −1.53 | −1.34 | −1.30 | −1.52 | −0.07 | −0.52 |
| 14 | 标准化 | 1.10 | 1.14 | 1.17 | 0.41 | 0.44 | 1.57 |
| 15 | | 0.20 | 0.01 | −0.09 | 0.91 | −1.23 | −1.24 |
| 16 | | −0.85 | −0.78 | −0.50 | 0.16 | −0.86 | −0.28 |
| 17 | 均值 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 18 | 标准差 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 19 | | Ni | Co | Cu | Cr | S | As |
| 20 | Ni | 1.00 | | | | | |
| 21 | Co | 0.85 | 1.00 | | | | |
| 22 | Cu | 0.75 | 0.98 | 1.00 | | | |
| 23 | Cr | 0.64 | 0.24 | 0.15 | 1.00 | | |
| 24 | S | 0.34 | 0.61 | 0.60 | −0.46 | 1.00 | |
| 25 | As | 0.60 | 0.45 | 0.41 | 0.24 | 0.42 | 1.00 |

表 6　R 型聚类分析连结顺序表

Tab. 6　Link sequence table of R cluster analysis

| 连结顺序 | 连结的元素 | | 替代数据 | 相关性水平 |
|---|---|---|---|---|
| 1 | Co | Cu | CoCu=(Co+Cu)/2 | 0.979514116 |
| 2 | S | As | SAs=(S+As)/2 | 0.876545651 |
| 3 | Ni | CoCu | CoCuNi=(CoCu*2+Ni)/3 | 0.799764994 |
| 4 | CoCuNi | SAs | CoCuNiSAs=(CoCuNi*3+SAs*2)/5 | 0.659007309 |
| 5 | CoCuNiSAs | Cr | | 0.112335175 |


图 6　相关矩阵对话框

Fig. 6　Dialog box of correlation matrix

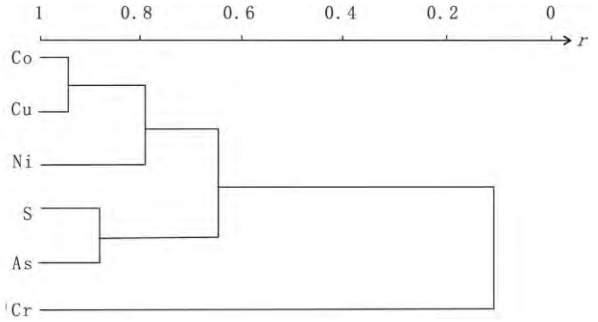


图 7　R 型聚类分析谱系图

Fig. 7　Pedigree chart of R cluster analysis

## 6　判别分析

若有两个母体，按照某种准则，把它们的 P 种特征(变量)组合成一个综合指标(关系式)，称为判别函数，使两类母体的区分率达到最好[6]。判别函数中最简单的形式是线性函数。有 $p$ 个变量 $n$ 个样品的两个母体的判别函数可以写成：

$$y = c_1x_1 + c_2x_2 + \cdots\cdots + c_px_p$$

根据公式推导[7]，得到以下方程组：

$$\begin{cases} S_{11}c_1 + S_{12}c_2 + \cdots\cdots + S_{1p}c_p = \bar{x}_1(A) - \bar{x}_1(B) \\ S_{21}c_1 + S_{22}c_2 + \cdots\cdots + S_{2p}c_p = \bar{x}_2(A) - \bar{x}_2(B) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ S_{p1}c_1 + S_{p2}c_2 + \cdots\cdots + S_{pp}c_p = \bar{x}_p(A) - \bar{x}_p(B) \end{cases}$$

解线性方程组即可求出 $c_1$、$c_2$、…、$c_p$。其中 $S_{kf} = \Sigma\{x_{ki}(A) - \bar{x}_k(A)\} * \{x_{fi}(A) - \bar{x}_f(A)\} + \Sigma\{x_{ki}(B) - \bar{x}_k(B)\} * \{x_{fi}(B) - \bar{x}_f(B)\}$，$(k, f = 1, 2, \cdots\cdots, p; i = 1, 2, \cdots, n)$。罗先熔等[8]提到，当 $k=f$ 时 $S_{kf}$ 代表每个变量的方差和，当 $k \neq f$ 时 $S_{kf}$ 代表两两变量之间的协方差和。方差/协方差的定义为同一/不同变量之间的偏差乘积的平均数。但是根据 $S_{kf}$ 表达式，$S_{kf}$ 应为方差或协方差的 $n$ 倍($n$ 为样品数)的和。实际胡以铿[7]在《地球化学中的多元分析》中的计算实例中 $S_{if}$ 也是方差或协方差的 $n$ 倍的和。

作者以文献[7]中的实例来说明在 Excel 中判别分析的计算过程。表 7 中第 3 到第 9 行为原始数据，共 7 个样品($n=7$)，在第 10 行计算每个变量的均值，比如在 B10 中输入：=AVERAGE(B3:B9)，在第 11 行中计算两组样品同一变量的均值差，比如在 B11 中输入：=B10－E10，在第 12 行中计算每个变量方差的 7 倍，比如在 B12 中输入：=VAR.P(B3:B9)*7，在第 13 行中计算两组样品中同一变量方差的加和，比如在 B13 中输入：=B12+E12，其中 B13，C13，D13 分别代表 $S_{11}$、$S_{22}$、$S_{33}$，同理在 B14、B15 中分别计算协方差的 7 倍以及协方差和，分别输入：=COVARIANCE.P(B3:B9,C3:C9)*7 和 =B14+E14，其中 B15、C15、D15 分别代表 $S_{12}$、$S_{13}$、$S_{23}$。经过上面方差和协方差的计算得到如下方程组：

$$\begin{cases} 2.2909c_1 + 0.0947c_2 + 0.2176c_3 = 0.5571 \\ 0.0947c_1 + 1.7140c_2 + 0.1873c_3 = -0.17 \\ 0.2176c_1 + 0.1873c_2 + 0.5541c_3 = -0.2043 \end{cases}$$

将方程组的系数在第 16、17、18 行中按顺序排列，在 E16、E17、E18 中分别输入函数：=B16*B19+C16*C19+D16*D19，=B17*B19+C17*C19+D17*D19，=B18*B19+C18*C19+D18*D19，此时 E16、E17、E18 中的值均显示为零。然后利用“规划求解”解线性方程组：数据－规划求解－

弹出“规划求解参数”对话框(图 8),在“设置目标”中输入:=E16,在目标值中输入:0.5571,在“更改可变单元格”中输入:=B19:D19,在“遵守约束”中点击“添加”,在弹出对话框中分别输入:E17=F17,E18=F18,最后把“使无约束变量为非负数”前的“勾号”去掉,确定即可得到 B19:D19 中的三个解,同时 E16、E17、E18 中的值也变为非零。获得方程组的解之后,就可以列出判别函数,判别显著性检验以及对未知样品的判别,本文略。本文计算结果与文献[7]中实例计算中的微有差别,这是在计算过程中保留有效数字个数不同造成的(本文计算过程中保留 9 位有效数字,实际在表 7 中显示 4 位)。该方法的难点为判别分析的原理,方差/协方差的计算以及利用“规划求解”解方程组技巧。


**图 8 规划求解对话框**

Fig. 8 Dialog box of solution of programming

## 7 结论

Excel 为大众软件,除本研究中提到的应用外,微量/稀土配分图,散点图,相关性计算,正态检验,样品化验数据误差的检验,一次趋势面分析等等都可以用 Excel 来实现。因此熟练使用 Excel 基本能满足地球化学的数据处理需求。地球化学数据处理多种多样,有时需借助好几种软件甚至收费的软件才能实现各种计算需求,每一种软件又需要花费一定的时间去掌握和熟练。熟练使用 Excel 不失为较为理想地选择,对地质人员较为适用。

**表 7 判别分析数据表**

Tab. 7 Data table of discriminant analysis

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | | | 有矿样品 | | | 无矿样品 | |
| 2 | 变量(序号) | Cu(1) | Ag(2) | Bi(3) | Cu(1) | Ag(2) | Bi(3) |
| 3 | | 2.58 | 0.9 | 0.95 | 2.25 | 1.98 | 1.06 |
| 4 | | 2.9 | 1.23 | 1 | 2.16 | 1.8 | 1.06 |
| 5 | | 3.55 | 1.15 | 1 | 2.33 | 1.74 | 1.1 |
| 6 | 数据 | 2.35 | 1.15 | 0.79 | 1.96 | 1.48 | 1.04 |
| 7 | | 3.54 | 1.85 | 0.79 | 1.94 | 1.4 | 1 |
| 8 | | 2.7 | 2.23 | 1.3 | 3 | 1.3 | 1 |
| 9 | | 2.7 | 1.7 | 0.48 | 2.78 | 1.7 | 1.48 |
| 10 | 均值 | 2.9029 | 1.4586 | 0.9014 | 2.3457 | 1.6286 | 1.1057 |
| 11 | 均值差 | 0.5571 | −0.1700 | −0.2043 | | | |
| 12 | 方差*7 | 1.3169 | 1.3613 | 0.3831 | 0.9740 | 0.3527 | 0.1710 |
| 13 | 方差和 | 2.2909 | 1.7140 | 0.5541 | | | |
| 14 | 协方差*7 | 0.1958 | 0.0431 | 0.1164 | −0.1011 | 0.1746 | 0.0709 |
| 15 | 协方差和 | 0.0947 | 0.2176 | 0.1873 | | | |
| 16 | | 2.2909 | 0.0947 | 0.2176 | 0.5571 | 0.5571 | |
| 17 | 方程组系数 | 0.0947 | 1.7140 | 0.1873 | −0.1700 | −0.1700 | |
| 18 | | 0.2176 | 0.1873 | 0.5541 | −0.2043 | −0.2043 | |
| 19 | 方程组解 | 0.2896 | −0.0649 | −0.4605 | | | |

**参考文献：**

[1] 蒋敬业，程建萍，祁士华，等. 应用地球化学[M]. 武汉：中国地质大学出版社，2006.

[2] H. E. 霍克斯，J. S. 韦布. 矿产勘查的地球化学[M]. 谢学锦，译，廊坊：地质科学院物探研究所，1974.

[3] 伍宗华，古平. 隐伏矿床的地球化学勘查[M]. 北京：地质出版社，2000.

[4] 李明欣，梁斌，王全伟，等. 川西龙泉山白垩系泥质岩的元素地球化学特征[J]. 高校地质学报，2013，19(2)：346－354.

[5] 春乃芽. 利用 Excel 实现 R 型聚类分析[J]. 物探与化探，2007，31(4)：374－376.

[6] 春乃芽. 利用 Excel 实现判别分析[J]. 物探化探计算技术，2007，29(6)：560－564.

[7] 胡以铿. 地球化学中的多元分析[M]. 武汉：武汉地质学院地球化学教研室，1984.

[8] 罗先熔，文美兰，欧阳菲，等. 勘查地球化学[M]. 北京：冶金工业出版社，2007.

## Advanced application of Excel in geochemical data processing

TAN Qin-ping[1,2], XIA Yong[1*], XIE Zhuo-jun[1,2], YAN Jun[1,2]

(1. State Key Laboratory of Ore Deposit Geochemistry, Institute of Geochemistry, Chinese Academy of Sciences, Guiyang 550002, China;
2. University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract**: The Excel functions of chart drawing and data computing facilitate geochemical data processing. With studying in the principle of geochemical data processing this paper explains the calculation steps of the triangular diagram, the frequency histogram, the probability ruling paper graphical method for anomaly threshold, the multifractal method to calculate anomaly threshold, R cluster analysis, and discriminant analysis. These methods are entirely applicable to geochemical data processes.

**Key words**: triangular diagram; anomaly threshold; histogram; cluster analysis; discriminant analysis