DOI:10.13758/j.cnki.tr.1984.02.010

方法介绍

数据分布类型检验及其在土壤学中的应用

I. 偏度、峰度检验法及其计算机程序

唐 诵 六

（中 国 科 学 院 南 京 土 壤 研 究 所）

早在1954年，Ahrens[1]就发现岩石、矿物样本中微量金属元素的浓度分布与常量元素不同。它不呈正态分布而呈对数正态分布。而Vistelius[2]认为大多数微量元素的浓度分布呈大的正偏。由于算术平均值和标准差只适用于表达正态分布样本。因而，用算术平均值表示样本的集中性、用算术标准差表示样本的离散度就不适用于非正态分布的微量金属元素。同样，用算术平均数及标准差计算的置信区间也不能反映实际的浓度范围。为了解决这一问题，应当先判断样本属于何种分布类型。如果属于正态分布，可用算术平均数及标准差表示。如果属于对数正态分布则应将原始测定值取对数后计算其平均值及标准差，然后算出几何平均数及标准差来表示。如果既不属正态又不属对数正态，则应将样本作正态化处理。当然，取对数本身也是一种正态化处理。有关判断样本分布类型的若干种方法，读者可参阅其它的著作[3—5]。

由于判断分布类型的计算过程比较复杂，尤其是在样本量多、样品数大的情况下，藉手工或一般计算器往往难于胜任。因此有必要使用计算机。但根据目前条件，工作者不一定具备计算机或微型机。本文的目的，就是介绍几种可供选择的判断方法，并提供专为SHARP PC-1500袖珍计算机编制的Basic语言算法程序。当然，将程序稍作修改，也能用于其它机种。

PC-1500机由于其价格低廉、操作容易，使用者稍经学习便能掌握，因此是目前应用较广的机种。本文介绍偏度、峰度检验法及应用于PC-1500计算机的计算程序。

一、偏度、峰度检验法及分位数表

偏度、峰度法沿用已久，其优点是计算简单，适用于样品数7以上的所有样本，并能提供样本的分布曲线的形态。其具体计算方法可参阅文献5。这里只作简单介绍。第一步，先作正态性检验，即根据样本的原始测定值按公式1至4计算出偏度值和峰度值。

表 1　　偏度、峰度检验分位数表

样品数	偏度概率		峰度概率	
	0.10	0.02	0.10	0.02
	P.1	P.2	P.3	P.4
8	0.99	1.42	3.70	4.53
9	0.97	1.41	3.86	4.82
10	0.95	1.39	3.95	5.00
12	0.91	1.34	4.05	5.20
15	0.85	1.26	4.13	5.32
20	0.77	1.15	4.17	5.36
25	0.71	1.06	4.16	5.30
30	0.66	0.98	4.11	5.21
35	0.62	0.92	4.10	5.13
40	0.59	0.87	4.06	5.04
45	0.56	0.82	4.00	4.94
50	0.53	0.79	3.99	4.88
60	0.49	0.72	3.94	4.78
70	0.46	0.67	3.89	4.68
80	0.43	0.63	3.85	4.58
90	0.41	0.60	3.81	4.48
100	0.39	0.57	3.77	4.39

本表资料由中国科学院系统所吴传义先生提供。

再在偏度、峰度检验的分位数表（表1）中查出与样品数相对应的正态分布的置信水平（概率）。当计算所得的偏度值及峰度值小于表1中所列的概率为0.10的偏度及峰度值时，该样品属于正态分布。若大于概率0.02或大于概率0.10小于0.02的数值时，则样本不属于正态分布。第二步，作对数正态检验，即将原始测定值取对数，按照正态检验的同样公式计算出偏度和峰度值，并由表1查出置信水平。如果满足概率0.1，则属于对数正态分布，否则则不属对数正态。将正态检验和对数正态检验的结果作一比较，不难判定样本的分布类型。有时样本经检验既不属正态又不属对数正态，则提示该样本属偏态分布。当正态检验的偏度值为正值时属正偏，偏度值为负值时属负偏。此时，应

采用其它方法来验证。计算偏度、峰度及其标准误的公式为：

$$\text{偏度}=\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{\sqrt{\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^3}} \qquad (1)$$

$$\text{偏度的标准误}=\sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}} \qquad (2)$$

$$\text{峰度}=\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^4}{\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^2} \qquad (3)$$

$$\text{峰度的标准误}=\sqrt{\frac{24n(n-1)^2}{(n-2)(n+5)(n^2-9)}} \qquad (4)$$

式中，n为样品数，x_i为样品测定值，$\bar{x}$为样本平均值。

二、偏度、峰度法的计算机程序及其使用方法

本文所提供的计算机程序专为进行上述计算而编制。为便于说明使用方法，在程序的后面(即第380条语句之后)，附有一个实例，系计算机印出的结果。这是南京地区22个土壤样品的钪的分布检验。具体数据为3.40,5.70,9.60,10.3, 11.8, 12.1, 12.2, 12.9, 13.0,13.0,13.3,13.5,13.7,14.9,15.0,15.5, 15.6, 16.5,17.1,17.5,19.6,22.9。操作方法见表2。当使用本程序时，操作者可任意选用下列三种不同的方式：

1. 将预先录制在磁带上的偏度、峰度检验分位数表(表1)输入计算机。数表是以每个样品数作为一维场的形式贮存的。即当操作至步骤2、显示器询问是否由磁带输入数表时，回答Y(是)。此时，计算机将进行计算并印出正态和对数正态检验的结果，包括两

表2　　键　盘　操　作　步　骤

步骤	输入	显示	说明
1	DEF SPACE	NO.OF SAMPLE=	问样品数
2	22 ENTER	SK TABLE CLOAD? (Y,N)	问是否由磁带输入数表
3	Y ENTER	FILE NAME(SK)=	要求输入数表名称，转步骤9
	N ENTER	SKTABLE KEY IN? (Y,N)	问是否由键盘输入数表
4	Y ETNRE	P.1=	要求输入数表
	NENTER	ELEMENT=	问无素名称，转步骤10
5	0.77 ENTER	P.2=	继续输入数表
8	5.36 ENTER	ELEMENT=	要求输入元素名称，转步骤10
9	SK ENTER	ELEMENT=	要求输入元素名称
10	SC ENTER	NO.OF SAMPLE=	问样品数
11	22 ENTER	DATA=	要求输入数据
12	13.4 ENTER	DATA=	要求继续输入数据
33	22.9 ENTER	VERIFY? (Y.N)	问是否要核对
34	Y ENTER	CORR.NO=	印出输入的数据供核对，问须改正的样品号，转步骤35
	N ENTER	ELEMENT=	印出计算结果，问下一个待算元素名称
35	1 ENTER	DATA=	问1号样品正确数据
36	3.4 ENTER	CORR.NO=	问尚须改正的样品号
37	ENTER	ELEMENT=	印出计算结果，问下一个待算元素名称

```
 5:REM PROG.SKEWN
   ESS
10:" ":CLEAR :
   WAIT 0:INPUT "
   NO.OF SAMPLE="
   ;S
15:IF 8>SGOTO 10
20:LF 1:LPRINT "N
   =";S
25:B=S-1:DIM X(B)
   ,P(3)
30:INPUT "SK TABL
   E CLOAD?(Y,N)"
   ;Y$
32:IF (Y$="Y")+(Y
   $="N")<>1GOTO
   30
35:IF Y$="N"GOTO
   75
40:IF S>200LET N=
   INT ((S+25)/50
   )*50:GOTO 55
42:IF S>100LET N=
   INT ((S+12)/25
   )*25:GOTO 55
44:IF S>60LET N=
   INT ((S+5)/10)
   *10:GOTO 55
45:IF S<16LET N=S
   :GOTO 55
50:N=INT ((S+2)/5
   )*5
55:LPRINT N:INPUT
   "FILE NAME(SK)
   =";D$
60:C$=D$+STR$ (N)
65:INPUT #C$;M
70:INPUT #C$;P(*)
   :GOTO 100
75:INPUT "SK.TABL
   E KEY IN?(Y,N)
   ";Y$
77:IF (Y$="Y")+(Y
   $="N")<>1GOTO
   75
80:IF Y$="N"GOTO
   100
85:FOR I=0TO 3:P$
   ="P."+STR$ (I+
   1)+"="
90:CLS :PRINT P$;
95:INPUT P(I):
   NEXT I:CLS
100:INPUT "ELEMENT
   =";A$
105:LF 1:LPRINT A$
110:INPUT "NO.OF S
   AMPLE=";Z
115:IF Z=SGOTO 130
120:END
130:FOR I=0TO B
135:INPUT "DATA=";
   X(I)
140:NEXT I:CLS
145:INPUT "VERIFY?
   (Y,N)";B$
147:IF (B$="Y")+(B
   $="N")<>1GOTO
   145
150:IF B$="N"GOTO
   200
155:FOR I=0TO B
160:LPRINT "DA";I+
   1;"=";X(I)
165:NEXT I
170:INPUT "CORR.NO
   .=";I:GOTO 180
175:GOTO 200
180:INPUT "DATA=";
   X(I-1):GOTO 17
   0
200:LF 1:LPRINT "N
   ORMAL":GOSUB 2
   60
205:IF CS<0LF 1:
   LPRINT "NEGATI
   VE SKEW"
210:IF CS>0LPRINT
   "POSSITIVE SKE
   W"
215:LPRINT "AM+-2S
   D=";.0001*INT
   (10000*(Q-2*R)
   );"-";.0001*
   INT (10000*(Q+
   2*R))
220:FOR I=0TO B
225:X(I)=LN X(I)
230:NEXT I
235:LF 1:LPRINT "L
   N-NORMAL":
   GOSUB 260
240:LPRINT "GM=";.
   0001*INT (1000
   0*EXP (Q))
245:LPRINT "GM+-2S
   D=";.0001*INT
   (10000*EXP (Q-
   2*R));"-";.000
   1*INT (10000*
   EXP (Q+2*R))
250:LF 1:GOTO 100
260:T=0:R=0
265:FOR I=0TO B
270:T=T+X(I):R=R+X
   (I)*X(I)
275:NEXT I
280:R=√((R-T*T/S)/
   (S-1))
285:Q=T/S
290:E=0:F=0:G=0:H=
   0
295:FOR I=0TO B
300:E=E+(X(I)-Q):F
   =F+(X(I)-Q)^2:
   G=G+(X(I)-Q)^3
305:H=H+(X(I)-Q)^4
310:NEXT I
315:CS=.0001*INT (
   10000*((G/S)/(
   √((F/S)^3))))
320:CE=.0001*INT (
   10000*((H/S)/(
   (F/S)^2)))
325:LPRINT "M=";.0
   001*INT (10000
   *Q)
330:LPRINT "SD=";.
   0001*INT (1000
   0*R)
335:LPRINT "Cs=";C
   S;"+-";.0001*
   INT (10000*√((
   6*S*(S-1))/((S
   -2)*(S+1)*(S+3
   ))))
340:LPRINT "Ce=";C
   E;"+-";.0001*
   INT (10000*√((
   24*S*(S-1)*(S-
   1))/((S-2)*(S+
   5)*(S*S-9))))
345:IF Y$="N"
   RETURN
350:IF ABS (CS)<P(
   0)LPRINT "P(Cs
   )>0.1":GOTO 36
   5
```

```
355: IF ABS (CS)>P(
     1)LPRINT "P(Cs
     )<0.02":GOTO 3
     65
360: LPRINT "0.02<P
     (Cs)<0.1"
365: IF ABS (CE)<P(
     2)LPRINT "P(Ce
     )>0.1":RETURN
370: IF ABS (CE)>P(
     3)LPRINT "P(Ce
     )<0.02":RETURN
375: LPRINT "0.02<P
     (Ce)<0.1"
380: RETURN
```

```
N= 22

Sc

NORMAL
M= 13.5954
SD= 4.2073
Cs=-0.3241+- 0.490
9
Ce= 3.8797+- 0.952
7
P(Cs)>0.1
P(Ce)>0.1
NEGATIVE SKEW
AM+-2SD= 5.1808- 2
2.01

LN-NORMAL
M= 2.5465
SD= 0.4055
Cs=-1.8245+- 0.490
9
Ce= 6.7693+- 0.952
7
P(Cs)<0.02
P(Ce)<0.02
GM= 12.7633
GM+-2SD= 5.672- 28
.7203
```

者的偏度、峰度值及其标准误；偏度、峰度的置信水平；并印出样本属何种偏斜；样本均值、标准差以及均值加减两倍标准差的范围值。操作者根据两种检验的置信水平，即可看出样本属何种分布。

2. 通过键盘输入偏度、峰度检验的分位数表。即当显示器询问是否由磁带输入数表时，回答N(否)。则此时进入步骤3，显示器询问是否由键盘输入数表。回答Y。接着将表1中与样品数相对应的偏度、峰度检验值(P. 1至P. 4)依次输入计算机。用本法所得的结果与上法完全相同。只是当样品数在表1中查找不到时(样品数为11,13,14等)，可输入相邻的较小的样品数的数据。如本例中，样品数为22，则输入样品数为20的数值即0.77，1.15，4.17，5.36。

3. 不输入偏度、峰度检验分位数表。即当进入第2，3步骤时，均回答N。此时，计算机只印出偏度、峰度值而不印出置信水平。操作者须自行查表。在本例中，偏度的正态检验值为-0.3241，其绝对值为0.3241，小于表1中样品数为20、概率为0.1的偏度值0.77，即$P>0.1$。而对数正态检验的偏度值计算所得为-1.8245，其绝对值为1.8245，大于表1中样品数为20、概率为0.02的偏度检验值1.15，即对数正态置信水平为$P<0.02$。按同样方法判断峰度概率，最后可得出样本属于正态分布的结论。因而样本的均值应为13.6，标准差为4.21，95%复盖范围值为5.18至22.0。

此外，本程序中有核对改错部分。在本例中，第12步误将13.4输入，经印出发现后，于35、36步骤中将正确数据3.4输入，而错误即行取消。

程序中的有关参数与符号，说明如下：

X，场，存Xi，S个向量；P，场，存分位数表，4个向量；

S，样品数；B=S-1；

A$，元素名；B$，核对否；C$，数表名称加样品数；D$，数表名称；Y$，输入数表否；

Q，均值；R，标准差；CS，偏度；CE，峰度；

I，循环参数；

E、F、G、H、R、T，工作单元。

计算机的印出结果中，有关符号说明如下：

N，样品数；M，均值；SD，标准差；Cs，偏度；Ce，峰度；P，概率；AM，算术平均数；GM，几何平均数；

NORMAL，正态检验；LN-NORMAL，对数正态检验；

NEGATIVE SKEW，负偏；POSSITIVE SKEW，正偏。

参考文献

〔1〕 Ahrens, L.H., Geochim. Cosmochim. Acta, 5, 49—73, 1954.

〔2〕 Vistelius, A.B., J.Geol., 68：1—22, 1960.

〔3〕 唐诵六，南京地区土壤中重金属浓度的概率分布。环境中若干元素的自然背景值及其研究方法，9-15页，科学出版社，1982。

〔4〕 杨国治、杨学义，土壤背景值的频数分布与统计方法。环境科学，4(3)，20—25，1983年。

〔5〕 薛仲三，医学统计方法与原理，104—120页，人民卫卫出版社，1978。