

方法介绍

数据分布类型检验及其在土壤学中的应用

V. 偏态样本的正态化及其计算机程序

唐 诵 六

(中国科学院南京土壤研究所)

以前曾介绍过大样本和小样本数据分布类型检验的几种方法[1-4], 这些方法可用来判别样本观测值属于正态还是对数正态分布。然而有些样本经检验可能既不属于正态, 又不属于对数正态分布, 它们是偏态分布。对于这类样本, 就不能简单地用算术均值或几何均值来表示。可以先进行正态化处理, 然后得出均值和标准差。本文介绍的正态化处理, 即是把在原始尺度上呈偏态(正偏或负偏)的频数分布转变成机率尺度, 使原来的偏态分布在转换后的尺度上呈现正态。

由于正态化处理的原理和计算方法已有文献详细叙述过[5,6], 因此, 本文只作简单介绍。本方法只适用于大样本, 步骤是将观测值按大小将样品分组, 计数每组的频数和相对累计频数。根据相对累计频数可

从正态曲线下的面积表查得机率尺度(组段), 并计算出相邻两组间的组距。将频数除以组距即得出该组的正态化后的频数 f 。接着利用公式算出平均组值 $\bar{t}$ 及组值标准差 δt 。

$$\bar{t} = \sum tf / n$$

$$\delta t = \sqrt{\frac{n \sum t^2 f - (\sum tf)^2}{n(n-1)}}$$

式中, n 为正态化后的总频数, 即各组频数的和。 t 为组值, 即相邻两组的组段的均值。最后, 根据平均组值加减组值标准差后所得的 t 值, 即可反算出相应的浓度值。

本程序的键盘操作方法见表 1。在程序运行时要

表 1 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	51 ENTER	ELEMENT =	印出样品数, 问样本名称
3	Zn ENTER	DATA =	印出样本名称, 要求输入数据
4	69 ENTER	DATA =	要求继续输入数据
54	245 ENTER	VERIFY? (Y, N)	问是否要核对
55	N ENTER		转步骤56
	Y ENTER	CORR. NO =	印出输入的数据, 问须修改的样品号
56	ENTER*	INTERVAL =	问组距
57	25 ENTER		印出全部计算结果

* 本例没有须要修正的数据, 如有须修正的数据, 其操作方法可参阅文献[1]。

求使用者输入样品数、样本名称、样品观测值, 答复是否须要核对及须修正的样品号和它的正确数值。在运算中要求输入组距, 组距的确定方法同文献[4], 即令组距约等于全距的十分之一。印出结果部分包括全距、组距、组数、样本均值、标准差、均值加减两倍标准差范围值、正态化处理后的总频数、 $\sum tf$ 值、 $\sum t^2 f$ 值、平均组值、组值标准差。最后印出正态化处理后的均值、均值加减一个标准差范围值、均值加减两个

标准差范围值。

程序中的有关参数及符号, 说明如下:

N, 样品数。

R(30), 场, 存正态积分函数。

X, 场, 存观测值, N个向量。

M, 组数。

FV, 场, 存各组的相对累计频数, M个向量。

W, 场, 存各组相对频数, M个向量。

```

5: REM PROG. NORMA      (M-1), GG(M-1),      -U*U/2): E=E+(C
LIZE                    F(M-1), T(M-1),      +U)/300
7: REM N>50
10: " ": CLEAR ;        TT(M-1)
    WAIT 0: DIM R(3     180: B=0: S=0
    0)
    185: FOR I=0 TO N-1
20: INPUT "NO. OF S     190: B=B+X(I): S=S+X
    AMPLE="; N: DIM      (I)*X(I)
    X(N-1)
    195: NEXT I
21: IF N<50 GOTO 20     200: S=J((S-B*B/N)/
22: LF 2: LPRINT "N     (N-1)): B=B/N
    ="; N
    203: LF 1: LPRINT "O
25: INPUT "ELEMENT     205: LF 1: LPRINT "A
    ="; E$
    M="; .0001*INT
26: LF 1: LPRINT E$     (10000*B):
30: FOR I=0 TO N-1      LPRINT "SD="; .
32: INPUT "DATA=";      0001*INT (1000
    X(I)
    0*S)
35: NEXT I: CLS
207: LPRINT "X+-2S=
40: INPUT "VERIFY?     "; .0001*INT (1
    (Y, N)"; B$
    0000*(B-2*S));
45: IF B$="N" GOTO      "-"; .0001*INT
    100
    (10000*(B+2*S)
50: FOR I=0 TO N-1      )
55: LPRINT "DA"; I+
265: FOR I=0 TO M-1
    1; "="; X(I)
270: K=0
60: NEXT I: LF 1
275: FOR A=0 TO N-1
65: INPUT "CORR. NO    280: IF X(A)<INT (Z
    .="; I: GOTO 75
    2/DD)*DD+(I+1)
70: GOTO 100
    *DD LET K=K+1
75: INPUT "DATA=";
285: NEXT A: G(I)=K
    X(I-1): GOTO 65
290: NEXT I
100: Z1=-10^8: Z2=10
310: FU(0)=0: FOR I=
    ^8
    0 TO M-1
105: FOR I=0 TO N-1
312: FU(I)=(INT (Z2
110: IF Z1<X(I) LET
    /DD))*DD+DD*I
    Z1=X(I)
314: NEXT I
115: IF Z2>X(I) LET
330: FOR I=0 TO M-1
    Z2=X(I)
332: W(I)=G(I)/N
120: NEXT I: LPRINT
334: NEXT I
    "R="; .0001*INT
350: GG(0)=G(0)
    (10000*Z2); "-
355: FOR I=1 TO M-1
    "; .0001*INT (10
360: GG(I)=G(I)-G(I
    000*Z1)
    -1)
125: INPUT "INTERVA
365: NEXT I
    L="; DD: LF 1:
381: P=1/J(2*PI ): E
    LPRINT "D="; DD
    =0: R(15)=.5: U=
130: M=INT (Z1/DD)-
    0: C=P: H=.01
    INT (Z2/DD)+1:
382: FOR I=1 TO 15
    LPRINT "NO. OF
383: FOR J=1 TO 10
    ITEM="; M
384: U=U+H: U=C+4*P*
135: DIM FU(M-1), W(
    EXP (-U*U/2): U
    M-1), WW(M-1), G
    =U+H: C=P*EXP (
    -U*U/2): E=E+(C
    +U)/300
385: NEXT J
386: R(15-1)=.5-E: R
    (15+1)=.5+E
387: NEXT I
400: WW(0)=-4.99
405: FOR I=0 TO M-2
410: J=0
415: J=J+1: IF W(I)<
    R(J) GOTO 425
420: IF J<30 GOTO 41
    5
425: Z=R(J-1): U=0.2
    *J-3.2: C=P*EXP
    (-U*U/2): IF Z>
    W(I) GOTO 450
430: U=U+H: U=C+4*P*
    EXP (-U*U/2): U
    =U+H: C=P*EXP (
    -U*U/2): U=(C+U
    )/300: Z=Z+U
435: IF Z<W(I) GOTO
    430
440: IF Z=W(I) GOTO
    450
445: U=U-(Z-W(I))/U
    *0.02
450: WW(I+1)=U
455: NEXT I
500: F(0)=0: F(M-1)=
    0
501: FOR I=1 TO M-2
502: IF WW(I+1)-WW(I)
    =0 LET WW(I+1)
    =WW(I+1)+.000
    001
503: NEXT I
505: FOR I=1 TO M-2
510: F(I)=GG(I)/(WW
    (I+1)-WW(I))
515: NEXT I
537: F=0: FOR I=0 TO
    M-1
540: F=F+F(I)
542: NEXT I
545: LF 1: LPRINT "S
    IGMA F="; .1*
    INT (10*F)
550: T(0)=0: TT(0)=0
    : T(M-1)=0: TT(M
    -1)=0
555: FOR I=1 TO M-2

```

560: T(I)=F(I)*((WW	10000*TS); "-";	DA 28= 30
(I+1)+WW(I))/2	.0001*INT (100	DA 29= 33
)	00*TL)	DA 30= 36
565: TI(I)=F(I)*((W	695: END	DA 31= 39
W(I+1)+WW(I))/	700: I=1	DA 32= 40
2)*((WW(I+1)+W	705: IF Y>WW(I) GOTO	DA 33= 53
W(I))/2)	715	DA 34= 54
570: NEXT I	710: GOTO 725	DA 35= 54
580: T=0: TT=0	715: I=I+1: IF I=M	DA 36= 56
585: FOR I=1 TO M-2	LET AB=FU(I-1)	DA 37= 59
590: T=T+T(I): TT=TT	+DD*(Y-WW(I-1)	DA 38= 62
+TT(I)	)/(4.99-WW(I-1	DA 39= 64
595: NEXT I	)): RETURN	DA 40= 66
600: LPRINT "SIGMA	720: GOTO 705	DA 41= 68
TF="; .0001*	725: AB=FU(I-1)+DD*	DA 42= 69
INT (10000*T):	(Y-WW(I-1))/(W	DA 43= 114
LPRINT "SIGMA	W(I)-WW(I-1)):	DA 44= 114
TTF="; .0001*	RETURN	DA 45= 116
INT (10000*TT)		DA 46= 136
605: LF J: LPRINT "t		DA 47= 152
BAR="; .0001*		DA 48= 163
INT (10000*T/F		DA 49= 171
)		DA 50= 213
610: ST=J((F*TT-T*T	N= 51	DA 51= 245
)/(F*(F-1)))		
615: LPRINT "St="; .	Zn	R= 8- 245
0001*INT (1000	DA 1= 69	D= 25
0*ST)	DA 2= 73	NO.OF ITEM= 10
650: A=T/F-ST; B=T/F	DA 3= 74	
+ST; C=T/F-2*ST	DA 4= 74	
; D=T/F+2*ST	DA 5= 75	
655: Y=T/F: GOSUB 70	DA 6= 76	ORIGIN
0	DA 7= 76	
660: LF J: LPRINT "N	DA 8= 76	AM= 82.5098
ORMALIZED": LF	DA 9= 78	SD= 47.1259
J: LPRINT "M=";	DA 10= 80	X+-2S=-11.7421- 17
.0001*INT (100	DA 11= 80	6.7616
00*AB): Y=A:	DA 12= 82	
GOSUB 700	DA 13= 88	SIGMA F= 80.7
665: OS=AB; Y=B:	DA 14= 94	SIGMA TF= 24.5727
GOSUB 700	DA 15= 97	SIGMA TTF= 71.4383
670: OL=AB; Y=C:	DA 16= 99	
GOSUB 700	DA 17= 99	t BAR= 0.3044
675: TS=AB; Y=D:	DA 18= 101	St= 0.8957
GOSUB 700	DA 19= 105	
680: TL=AB	DA 20= 105	NORMALIZED
685: LPRINT "X+-Sx=	DA 21= 107	
"; .0001*INT (1	DA 22= 110	M= 89.0493
0000*OS); "-"; .	DA 23= 8	X+-Sx= 58.455- 128
0001*INT (1000	DA 24= 9	.1278
0*OL)	DA 25= 18	X+-2Sx= 24.5016- 2
690: LPRINT "X+-2Sx	DA 26= 20	25.2894
="; .0001*INT (	DA 27= 28	

WW,场,存各组组段,M个向量。
 G,场,存各组累计实际频数,M个向量。
 GG,场,存各组实际频数,M个向量。
 F,场,存正态化后的各组频数,M个向量。
 T,场,存各组组值,M个向量。
 TT,场,存各组组值的平方,M个向量。
 P,常数 $1/\sqrt{2\pi}$ 。H,步长。
 I, J, 循环参数。
 B,C,D,E,F,K,P,S,U, V, Z, AB, ST, OS,
 OL,TS,TL, 工作单元。
 计算机印出结果中的有关符号,说明如下:
 N, 样品数。 Zn, 元素名,本例为锌。
 DA, 数据,即输入的样品观测值。
 R, 全距。 D, 组距。
 NO. OF ITEM, 组数。ORIGIN,正态化前。
 AM, 均值。 SD, 标准差。
 $X + -2S$, 均值加减两倍标准差。
 SIGMA F, 正态化后总频数。
 SIGMA TF, Σtf 值。
 SIGMA TTF, Σt^2F 值。
 t BAR, 平均组值。 St, 组值标准差。

NORMALIZED, 正态化后。 M, 均值。
 $X + -Sx$, 均值加减一个标准差。
 $X + -2Sx$, 均值加减两个标准差。

参 考 文 献

- [1] 唐涌六, 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。土壤, 16 卷 2 期, 66—69 页, 1984 年。
- [2] 唐涌六, 数据分布类型检验及其在土壤学中的应用 II. Shapiro—Wilk W 检验法及其计算机程序。土壤, 16 卷 3 期, 112—115 页, 1984。
- [3] 唐涌六, 数据分布类型检验及其在土壤学中的应用 III. Vistelius 置信带法及其计算机程序。土壤, 16 卷 4 期, 150—152 页, 1984。
- [4] 唐涌六, 数据分布类型检验及其在土壤学中的应用 IV. 克方检验法及其计算机程序。土壤, 16 卷 5 期, 192—195 页, 1984。
- [5] 薛仲三, 医学统计方法和原理, 150—152 页, 人民卫生出版社, 1978。
- [6] 杨国治, 环境中若干元素的自然背景值及其研究法, 85—87 页, 科学出版社, 1982。

(上接第213页)

表 7

几种水稻土在不同时期的呼吸强度和氧化还原电位

地 点	土 壤	作 物	呼吸强度(克/厘米 ² ·天)				氧化还原电位(毫伏)		
			分 蘖 期 (苗 期)	孕 穗 期 (开花期)	抽 穗 期 (结荚期)	成 熟 期	分 蘖 期 (苗 期)	孕 穗 期 (开花期)	抽 穗 期 (结荚期)
东 莞 县	坭 肉 田	水 稻	0.71	0.95	—	—	—	~1.5	284
	坭 骨 田		0.71	0.55	—	—	—	~2	198
	坭 肉 田	花 生	3.3	5.7	4.2	—	564	538	560
	坭 骨 田		1.4	3.4	4.1	—	504	459	535
南 海 县	坭 肉 田	水 稻	2.7	4.7	4.6	0.4	53	388	512
	坭 骨 田		2.3	2.2	0.3	0.3	64	352	452

注: 括号内为花生的生育期。

施有机肥(或与绿肥轮作)是培育该土壤的基本措施, 同时还要注意深耕松土, 促进土体向良好的结构体方向发展。

3. 油坭田: 土体各层中均含有较多黑色腐殖化、半腐殖化的有机残体, 但因土体通透性能差, 处于淹水时间较长, 所以有机质分解十分缓慢, 有机胶结物质随之不足。另外, 因土壤干湿胀缩作用较差, 作物根系生长发育差, 致使土壤结构体发育不良。这类土壤应加强排水和疏松土壤措施, 改善土壤通透性能。有条件的地区, 应推行水旱轮作, 为改善土壤结构体状况创造良好的条件。

参 考 文 献

- [1] 钟功甫、李次尼编著, 珠江三角洲, 13 页, 商务印书馆, 1960 年。
- [2] 龚子同等, 珠江三角洲土壤, 土壤专刊, 第 36 号, 科学出版社, 1964。
- [3] J. W. Dickson and P. Smart, Some Interactions Between Stress and Microstructure of kaolin (in Modification of Soil Structure, Edited by W. W. Emerson et al.), p. 57., copyright by John Wiley & Sons Ltd, 1978。
- [4] 彭拉姆帕鲁马著(刘志光等译), 渍水土壤的化学与水稻生长的关系, 45 页, 科学出版社, 1966。