

方法介绍

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

唐 涌 六

(中国科学院南京土壤研究所)

皮尔逊克方检验法是常用的检验样本是否遵从某一理论分布的方法之一。用来检验样本的正态性有其明显优点。一是计算比较简便,尤其是当样本很大时更为优越,二是它和置信带法[1]一样,具有很高的辨别能力。它的缺点是不适用于样品数不足50的小样本。此外,计算结果受组距的影响,尤其是当样本不很大时,不同的组距会得出不同的累计克方值,然而,它仍不失是一个较好的方法。

一、正态及对数正态性的克方检验法

计算的一般步骤是先由样本的观测值计算统计量作为总体参数的估计量。然后由正态分布的分布函数 $F(X)$ 确定随机变量的概率分布。再由样本的总频数算出各组的理论频数。将实际观测频数与理论频数相比较,即可判断本数据配合正态分布是否适宜。具体计算步骤如下。

1. 确定组距,将原始数据分组,计数每组的实际观测频数 O_i 。

2. 计算各组的标准正态变量 $U_i = (a_i - \bar{X})/S$ 。式中, a_i 为各组的下限值, \bar{X} 为样本均值, S 为样本的标准差。

3. 由正态分布表查得 U_i 的正态分布概率 $F(U_i)$ 。

4. 将相邻两组的 $F(U)$ 值相减,即 $F(U_{i+1}) - F(U_i)$,得出变量出现在该组段内的概率。

5. 将样本总频数乘以各组的正态概率,得到各组的理论频数 E_i 。

6. 将理论频数小于5的组合并。计算累计克方值。

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}, \quad df = n - k$$

式中, n 为合并后的有效组数, df 为自由度, $k=3$ 。

7. 由克方表查得符合正态分布的显著性水平。

8. 将原始数据转换成对数,按同样步骤算出对数正态分布的显著性水平。

9. 比较正态和对数正态性检验的显著性水平,作出结论。

本方法要求分组的有效组数(即频数不足5的组合合并后的组数)大于3。关于本方法的更加详细的原理及计算步骤,可参看有关的著作[2,3]。

二、计算机程序及其使用方法

本程序使用时,要求输入样品数、样本名称及原始观测数据。印出结果包括正态和对数正态检验的全距、均值、标准差、组距、组数、累计克方值及自由度。使用本程序有如下注意事项。要求样品数50以上,分组数30以内。运行中应按照显示器的要求两次输入组距。第一次输入正态检验的组距(表1,步骤78),第二次输入对数正态检验的组距(表1,步骤79)。组距的确定方法,可参照印出的全距 R ,使组距约等于全距的十分之一即可。如在本文的例子中,原始数据的全距为273,组距即定为25;原始数据取对数后全距为4.93(即5.6167-0.6931),组距即定为0.5。此外,当输入的数据中,如有值小于1时,机器将全部数据乘以一千倍再行运算。并在印出结果中印出PPB字样,表示印出的全距、均值、标准差均由原数据乘以一千倍后计算出。

最后,根据印出的累计克方值及自由度,查克方分布表。如在本例中(广东73个土壤样品中铬的浓度ppm数),正态检验的克方值为19.2078,自由度为3,查表得 $P < 0.005$,对数正态检验的克方值为5.1149,自由度为3,查表得 $0.1 < P < 0.2$ 。结论便是广东73个土壤铬的浓度呈对数正态分布,其信度大于0.1。

程序中的有关参数及符号,说明如下:

X, 场, 先存样品观测值, 后存观测值的对数值。

R, 场, 存积分函数, 30个向量。

G, 场, 先存正态检验的各组频数, 后存对数正态检验的各组频数。

F, 场, 先存正态检验、后存对数正态检验的各

```

5:REM PROG.CHI S
  QURE
7:REM N>50, NO. OF
  ITEM<30
10:" ":CLEAR :
  WAIT 0:DIM R(3
    ),G(30),F(30)
    ,H(30)
11:P=1/J(2*PI ):E
  =0:R(15)=.5:U=
    0:C=P:H=.01
12:FOR I=1TO 15
13:FOR J=1TO 10
14:U=U+H:U=C+4*P*
  EXP (-U*U/2):U
  =U+H:C=P*EXP (
    -U*U/2):E=E+(C
    +U)/300
15:NEXT J
16:R(15-I)=.5-E:R
  (15+I)=.5+E
17:NEXT I
20:INPUT "NO. OF S
  AMPLE=";N:DIM
  X(N-1)
21:IF N<50GOTO 20
22:LF 2:LPRINT "N
  =" ;N
25:INPUT "ELEMENT
  =" ;E$
26:LPRINT E$
30:FOR I=0TO N-1
32:INPUT "DATA=";
  X(I)
35:NEXT I:CLS
40:INPUT "VERIFY?
  (Y,N)";B$
45:IF B$="N"GOTO
  100
50:FOR I=0TO N-1
55:LPRINT "DA";I+
  1;"=" ;X(I)
60:NEXT I:LF 1
65:INPUT "CORR.NO
  .=" ;I:GOTO 75
70:GOTO 100
75:INPUT "DATA=";
  X(1-1):GOTO 65
100:GOSUB 700
105:IF Z2<1LPRINT
  "PPB":Z1=1000*
  Z1:Z2=1000*Z2:
  GOTO 115
110:GOTO 130
115:FOR I=0TO N-1
120:X(I)=1000*X(I)
125:NEXT I
127:LPRINT "R=" ;Z2
  ;"-";Z1
130:LF 1:LPRINT "N
  ORMAL":GOSUB 1
  80
140:LF 1:LPRINT "L
  N-NORMAL"
145:FOR I=0TO N-1
150:X(I)=LN (X(I)/
  (10^8))
155:NEXT I
160:GOSUB 700
165:GOSUB 180
170:INPUT "ELEMENT
  =" ;E$
172:LF 2:LPRINT E$
174:INPUT "NO. OF S
  AMPLE=" ;NN
176:IF NN=NGOTO 30
178:END
180:B=0:S=0
185:FOR I=0TO N-1
190:B=B+X(I):S=S+X
  (I)*X(I)
195:NEXT I
200:S=J((S-B*B/N)/
  (N-1)):B=B/N
205:LF 1:LPRINT "M
  =" ;.0001*INT (
  10000*B):
  LPRINT "SD=" ;.
  0001*INT (1000
  0*S)
210:INPUT "INTERVA
  L=" ;DD:LPRINT
  "D=" ;DD
260:M=INT (Z1/DD)-
  INT (Z2/DD)+1:
  LPRINT "NO. OF
  ITEM=" ;M
265:FOR I=0TO M-1
270:K=0
275:FOR A=0TO N-1
280:IF X(A)<INT (Z
  2/DD)*DD+(I+1)
  *DDLET X(A)=X(
  A)*(10^8):K=K+
  1
285:NEXT A:G(I)=K
290:NEXT I
310:F(0)=0:FOR I=0
  TO M-1
312:F(I)=(INT (Z2/
  DD))*DD+DD*I
314:NEXT I
330:FOR I=0TO M-1
332:F(I)=(F(I)-B)/
  S
334:NEXT I
345:F(0)=0
347:FOR I=1TO M-1
348:IF F(I)>36GOTO
  387
350:J=INT (15+5*(F
  (I)):E=.001
355:IF J<0GOTO 390
360:E=R(J):Q=0.2*(
  J-15)
365:IF Q=F(I)GOTO
  387
370:D=1+INT ((F(I)
  -Q)/.02):H=(F(
  I)-Q)/D:Z=H/2:
  T=H/6:C=P*EXP
  (-Q*Q/2)
375:FOR J=1TO D
380:Q=Q+Z:U=C+4*P*
  EXP (-Q*Q/2):Q
  =Q+Z:C=P*EXP (
  -Q*Q/2):E=E+(C
  +U)*T
385:NEXT J:GOTO 39
  0
387:F(I)=.999:GOTO
  391
390:F(I)=E
391:NEXT I
405:FOR I=0TO M-2
410:F(I)=F(I+1)-F(
  I)
415:NEXT I
420:F(M-1)=1-F(M-1
  )
430:FOR I=0TO M-1
435:F(I)=F(I)*N
440:NEXT I
456:I=0:UL=0
457:IF F(I)<5GOSUB
  602
458:I=I+1
459:IF I=M-1GOTO 5
  00

```

460:GOTO 457	N= 73	DA 51= 19
500:J=M-1	Cr	DA 52= 75
501:IF F(J)<5GOSUB	DA 1= 28	DA 53= 34
652	DA 2= 80	DA 54= 47
502:J=J-1	DA 3= 8	DA 55= 15
503:IF J=1GOTO 510	DA 4= 17	DA 56= 31
504:GOTO 501	DA 5= 2	DA 57= 12
510:FOR I=0TO M-1	DA 6= 15	DA 58= 112
512:H(I)=(G(I)-F(I))	DA 7= 12	DA 59= 9
)*G(I)-F(I))	DA 8= 32	DA 60= 23
)/(F(I)+.0000	DA 9= 139	DA 61= 13
00001)	DA 10= 36	DA 62= 18
514:NEXT I	DA 11= 12	DA 63= 40
515:XX=0	DA 12= 17	DA 64= 63
516:FOR I=0TO M-1	DA 13= 12	DA 65= 5
518:XX=XX+H(I)	DA 14= 58	DA 66= 29
520:NEXT I	DA 15= 275	DA 67= 90
522:LF 1;LPRINT "X	DA 16= 193	DA 68= 255
X="; .0001*INT	DA 17= 24	DA 69= 89
(10000*XX)	DA 18= 21	DA 70= 45
523:DF=0	DA 19= 86	DA 71= 8
524:FOR I=0TO M-1	DA 20= 34	DA 72= 14
525:IF G(I)>4LET D	DA 21= 72	DA 73= 38
F=DF+1	DA 22= 75	
526:NEXT I;DF=DF-3	DA 23= 147	R= 2- 275
;LPRINT "DF=";	DA 24= 35	
DF	DA 25= 42	NORMAL
555:RETURN	DA 26= 29	
602:F(I+1)=F(I)+F(I	DA 27= 29	M= 55.6575
+1)	DA 28= 26	SD= 53.4668
603:G(I+1)=G(I)+G(I	DA 29= 55	D= 20
+1)	DA 30= 58	NO.OF ITEM= 14
604:F(I)=0;G(I)=0	DA 31= 72	
606:RETURN	DA 32= 50	XX= 19.2078
652:F(J-1)=F(J)+F(J	DA 33= 65	DF= 3
-1)	DA 34= 28	
653:G(J-1)=G(J)+G(J	DA 35= 48	LN-NORMAL
-1)	DA 36= 28	R= 0.6931- 5.6167
654:F(J)=0;G(J)=0	DA 37= 70	
656:RETURN	DA 38= 57	M= 3.6341
700:Z1=-10^8;Z2=10	DA 39= 34	SD= 0.9224
^8	DA 40= 83	D= 0.5
705:FOR I=0TO N-1	DA 41= 63	NO.OF ITEM= 11
710:IF Z1<X(I)LET	DA 42= 140	
Z1=X(I)	DA 43= 22	XX= 5.1149
715:IF Z2>X(I)LET	DA 44= 36	DF= 3
Z2=X(I)	DA 45= 187	
720:NEXT I;LPRINT	DA 46= 29	
"R="; .0001*INT	DA 47= 79	
(10000*Z2); "-"	DA 48= 50	
; .0001*INT (10	DA 49= 75	
000*Z1).RETURN	DA 50= 94	

表1

键 盘 操 作 步 骤

步 骤	输 入	显 示	说 明
1	DEF SPACE	NO. OF SAMPLE =	问样本大小(程序运行数分钟后显示)
2	73 ENTER	ELEMENT =	印出样本大小, 问元素名称
3	Cr ENTER	DATA =	印出元素名称, 要求输入数据
4	28 ENTER	DATA =	要求继续输入数据
76	38 ENTER	VERIFY? (Y, N)	问是否要核对
77	Y ENTER	CORR. NO =	印出数据, 问须要修改的样品号
	N ENTER		
78	ENTER	INTERVAL =	印出全距、算术平均数、标准差, 要求输入正态检验组距
79	20 ENTER	INTERVAL =	印出正态检验组距、组数、累计克方值、自由度, 印出样本的对数值的全距、均值、标准差。要求输入对数正态检验的组距
80	0.5 ENTER	ELEMENT	印出对数正态检验的组距、组数、累计克方值、自由度。问下一个待算样本名称

组的正态分布概率。

H, 场, 先存正态检验、后存对数正态检验的各组克方值。

P, 常数 $1/\sqrt{2\pi}$ 。N, NN, 样品数。H, 步长。B\$, 印否。E\$, 样本名称。B, C, D, E, K, M, Q, S, U, V, Z, Z1, Z2, DD, DF, UL, XX, 工作单元。A, I, J, 循环参数。

印出结果中的有关符号说明如下:

N, 样品数。Cr, 元素名。DA, 数据。R, 全距。M, 均值。SD, 标准差。D, 组距。XX, 累计克方值。DF, 自由度。NORMAL, 正态检验。LN-NORMAL,

对数正态检验。PPB, 观测值放大一千倍。

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 II。土壤, 16卷4期, 150—152页, 1984。
- (2) 高玉堂, 环境监测常用统计方法, 59—62页, 原子能出版社, 北京, 1980。
- (3) 唐涌六, 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。

《土壤专报》出版预告

中国科学院南京土壤研究所专刊《土壤专报》第39号即将出版, 内容包括: 试拟全国土壤资源评价的原则和方法; 淮北平原白碱土的性态和分类; 根据地表形态指标自动识别和编制土壤图的尝试; 宁夏固源县侵蚀土壤主成分分析研究; 罗马尼亚中部几种主要土壤的发生特点; 罗马尼亚的农业土地评价; 土壤地带性理论的发展; 土壤发生层划分和诊断土壤分类的研究进展。可供土壤、地理、水土保持、农业科学工作者和高等院校有关专业师生参考。

各地新华书店经售, 或径向北京科学出版社订购。

中国科学院南京土壤研究所