

方法介绍

数据分布类型检验及其在土壤学中的应用

I. Shapiro—Wilk W 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)

W检验法是常用的检验样本是否遵从正态分布的方法之一。其优点是计算比较简单；对于显著性水平在0.1以下的样本判断精度高。缺点是只适用于样品数为50以内的小样本，而当显著性水平大于0.1时则判断精度不足。

W法在运算过程中，须要使用正态性W检验的系数表(以下简称a值表)，在根据计算结果判断分布类型时，须要使用W检验的判断界限值表(以下简称W值表)。好在这两个数表以及W检验的计算方法本身已有详细叙述^[1]，因此本文只着重介绍W检验的计算机程序，以及两种将数表预先录制在磁带上的方法及其程序。

W检验的步骤如下：

1. 先将数据自小至大重新排列，
2. 计算方差

$$S = \sum_{i=1}^n (X_i - \bar{X})^2$$

式中，n为样品数， X_i 为样品测定值， \bar{X} 为样本均值。

3. 计算B值

$$b = a_n(X_n - X_1) + a_{n-1}(X_{n-1} - X_2) + \dots + a_{n-k+1}(X_{n-k+1} - X_k)$$

式中，当n为偶数时 $k = \frac{n}{2}$ ；n为奇数时

$k = \frac{(n-1)}{2}$ 。 a_{n-i+1} ($i = 1, 2, \dots, k$) 值见文献^[1]中附表6。

4. 计算检验的统计量W

$$W = \frac{b^2}{s}$$

5. 将W值与文献^[1]，209页附表7的判断界限值 W_α 相比较，接受或舍弃正态性假设。

6. 将数据变换成对数，按上述步骤进行对数正态性检验。

一、W检验的计算机程序

本程序专为PC-1500机用W法检验样本的正态和对数正态性而编制，稍加修改亦可适用于其它机种。使用本程序时，操作者可有两种选择。一是将预先录制在磁带上的a值表及W值表输入计算机，二是通过键盘将该两表输入计算机。为便于说明，仍以南京地区22个土壤样品中钨的浓度作为计算实例。钨浓度的具体数据参见文献^[2]。两种方法叙述于后。

1. 通过磁带输入数表

先按照说明书要求接好磁带机，并插入录制有a值表的磁带。当计算进行至步骤2(表1)，显示器询问是否由磁带输入数表时，回答是(Y)。接着启动磁带机(为节省磁带机运转时间，最好预先将磁带快进到相应的样品数所在位置)，并输入a值表名称(本例使用英文小写a)。此时计算机即接纳a表数值。输入完毕后，磁带机自行停转。接着显示器要求输入W表。此时应换上录制有W值表的磁带并启动磁带机(为节省运转时间，最好将磁带预先快进到相应位置)，并输入W表名称(本例使用英文大写W)。W表输完后，磁带机自行停转。接着按显示器指示输入样本名称及样品的具体数据。

2. 通过键盘输入数表

当进行至步骤2，显示器询问是否由磁带输入数表时，回答否(N)。于是显示器要求输入a表。操作者即将与样品数相对应的a值(文献^[1]，第207—208页)依次输入。在本例中，样品数为22，a值依次为0.459, 0.3156, ……，0.0122。a值输完后，接着按显示器要求依次输入与样品数相应的W值(文献^[1]，209页)。在本例中，样品数为22，W值依次为0.878, 0.892, 0.911, 0.926, 0.961。W表输完后，再按显示器指示输入样本名称及样品的具体数据。

最后, 计算机印出正态和对数正态检验的显著性水平(W 检验中, 显著性水平分为五级, 即0.5, 0.1, 0.05, 0.02及0.01)。将此两水平加以比较, 即可得出样本遵从何种分布的结论。如规定检验的显著性水平大于0.1时接受假设, 则当正态检验的显著性水平大于0.1, 而对数正态检验的显著性水平不足0.1时, 样本即属正态分布。反之, 则属对数正态分布。当正态和对数正态检验两者处于相同显著性水平, 或均不足0.1时, 则无法得出结论。此时, 应改用其他检验方法。

程序中的有关参数及符号说明如下:

X, 场, 存样品数据, N 个向量;

N, 样品数;

Y, 场, 存运算的中间结果, N 个向量;

Z, 场, 存 a 值表, 25 个向量;

V, 场, 存 W 值表, 5 个向量。

A\$, 样本名称; B\$, 核对否; C\$, 数表名称加样品数; D\$, 数表名称, E\$, 数表由磁带输入否。

S, 样品数; M, a 值表最大向量数;

B, D, S, T, W, 工作单元。

I, J, 循环参数。

印出结果中的符号说明如下:

N, 样品数; SC, 元素名; DA, 样品数据。

NORMAL, 正态检验; LN-NORMAL, 对数正态检验。

M, 均值; SS, 方差; B, B 值; W, W 值,

$W = B^2/S^2$, P, 概率。

二、数表录带方法及程序

计算机在计算过程中往往须要使用统计数表。但在 PC-1500 机的说明书中并没有提供数表的录带程序。因此, 本文提供两种录制数表的程序。当然, 略加修改也可用来录制其它的数据以至文字资料。由于数表的数据量可能很大, 另外, 在实际使用时每次只用一个相应的样品数。因此, 数表是按样品数作为一个维场的格式录制的。对录制的数表量没有限制。现将两种程序分别叙述如下。

1. 程序 SAVE 1

表 1 程序 W TEST 操作方法

步骤	输入		显示	说明
1	DEF	SPACE	NO. OF SAMPLE =	问样本大小
2	22	ENTER	a, W TABLE CLOAD? (Y, N)	印出样品数, 问是否由磁带输入 a 表和 W 表
3	Y	ENTER	FILE NAME(a) =	要求输入 a 表, 启动磁带机转步骤 4
	N	ENTER	DATA OF TABLE a =	要求输入 a 表, 转步骤 6
4	SHIFT A	ENTER	FILE NAME(W) =	a 表输入后, 要求输入 W 表
5	W	ENTER	ELEMENT =	W 表输入后, 问样本名称, 转步骤 22
6	0.459	ENTER	DATA OF TABLE a =	继续输入 a 表
16	0.0122	ENTER	DATA OF TABLE W =	要求输入 W 表
17	0.878	ENTER	DATA OF TABLE W =	继续输入 W 表
21	0.961	ENTER	ELEMENT =	问样本名称
22	SC	ENTER	NO. OF SAMPLE =	问样本大小
23	22	ENTER	DATA =	要求输入数据
24	3.4	ENTER	DATA =	继续输入数据
45	22.9	ENTER	VERIFY? (Y, N)	问是否要核对
46	N	ENTER	ELEMENT =	印出计算结果, 问下一个待计算样本名称, 转步骤 22
	Y	ENTER	CORR, NO =	参见文献[2]

适用于样品数不同时向量相等，且样品数连续不间断的数表。如 W 检验中使用的 W 值表即属此类。在 W 值表中，不同样品数的向量均为 5，且样品数从 3 到 50 是连续不间断的。在本程序第 80 条语句中，“KOLM”字样是使用者任意规定的数表名称。在本文的例子（表 2）中，是用来录制 W 检验的判断界限值表，因此，可以事先将 KOLM 改为 W。

2. 程序 SAVE 2

不论当样品数不同时向量是否相同，亦不论数表中的样品数是否连续，本程序都能适用。因此，本程序有较广的使用范围。在本文的例子中（表 3），是用来

录制 W 检验的系数值表。同样，亦可用来录制 W 值表以及文献[2]所介绍的偏度、峰度检验分位数(SK)表等。读者按照表 2 和表 3 的方法，反复练习数次，当不难掌握。

现将两个录带程序中的有关参数及符号说明如下：

程序 SAVE 1

A，场，存数表，M个向量；

M，数表的向量，在本例即显著性水平数；

N，数表的最大样品数；

CS，数表名称加样品数；

表 2 程序 SAVE 1 操作方法

步骤	输入	显示	说明
1	DEF SPACE	P =	要求输入数表的显示性水平的级数
2	5 ENTER	N =	要求输入数表的最大样品数
3	50 ENTER	N = 1, P = 1, DATA =	要求输入样品数 N = 1 时，显著性水平的第 1 个数值
4	0 ENTER	N = 1, P = 2, DATA =	由于数表是从 N = 3 开始，故输入 0，下同
13	0 ENTER	N = 3, P = 1, DATA =	要求输入 N = 3 时，显著性水平的第一个数值
14	0.753 ENTER	N = 3, P = 2, DATA =	继续输入数表
15	0.756 ENTER	N = 3, P = 3, DATA =	
253	0.974 ENTER		数表输入完毕，程序结束

表 3 程序 SAVE 2 操作方法

步骤	输入	显示	说明
1	DEF SPACE	FILE NAME =	要求输入数表名称
2	SHIFT A ENTER	MAX DIMENSION =	要求输入数表的最大向量
3	25 ENTER	SAVING NUMBER =	要求输入样品数
4	3 ENTER	N = 3, P = 1, DATA =	要求输入 N = 3 的第 1 个 a 值
5	0.7071 ENTER	N = 3, P = 2, DATA =	要求输入 N = 3 第 2 个 a 值 (N = 3 时，只有 1 个 a 值)
6	ENTER	SAVING NUMBER =	要求输入样品数
7	4 ENTER	N = 4, P = 1, DATA =	要求输入 N = 4 时 a 表的第 1 个数值
8	0.6872 ENTER	N = 4, P = 2, DATA =	要求输入第 2 个 a 值
9	0.1677 ENTER	N = 4, P = 3, DATA =	
10	ENTER	SAVING NUMBER =	如此往复输入数表
...			
722	0.0035 ENTER	SAVING NUMBER =	数表已输入完毕
723	ENTER		

```

5:REM PROG.W TES      140:FOR J=0TO N-1    290:T=0:S=0
T
10:REM a,W TABLE     145:INPUT "DATA=";  295:FOR J=0TO N-1
NEEDED                X(J)
15:" ":CLEAR ;        150:NEXT I:CLS      300:T=T+X(I):S=S+X
WAIT 0:INPUT "        (I)*X(I)
NO.OF SAMPLE="      155:INPUT "VERIFY?  305:NEXT I
;N:LPRINT "N="      (Y,N)";B$
;N:LF 1              157:IF (B$="Y")+ (B  310:S=S-T*T/N
20:IF N>50GOTO 15    $="N")<>16GOTO    315:LF 1:LPRINT "M.
25:M=25:DIM X(N-1   155                    =";.0001*INT (
),Y(N-1),Z(M-1     160:IF B$="N"GOTO  10000*T/N):
),U(4)              165:FOR I=0TO N-1  LPRINT "SS=";
30:INPUT "a,W TAB    170:LPRINT "DA";I+  .0001*INT (1000
LE CLOAD?(Y,N)      1;="";X(I)         *S)
";E$
35:IF (E$="Y")+ (E   175:NEXT I        320:LPRINT "B=";.0
$="N")<>16GOTO      180:INPUT "CORR.NO  001*INT (10000
30                    .="";I:GOTO 190  *B)
40:IF E$="Y"GOTO     185:GOTO 195        325:W=B*B/S
75
45:FOR J=0TO INT     190:INPUT "DATA=";  330:LPRINT "W=";.0
(N/2)-1              X(I-1):GOTO 18    0001*INT (1000
50:INPUT "DATA OF   0
TABLE a=";Z(I)
55:NEXT I
60:FOR I=0TO 4
65:INPUT "DATA OF  205:IF X(J)<=X(J+1  335:IF W<U(0)
TABLE W=";U(I)      )GOTO 215          LPRINT "0.01>P
70:NEXT I:GOTO 11   210:D=X(J):X(J)=X(  ";GOTO 365
5
75:INPUT "FILE NA   215:NEXT J        LPRINT "0.02>P
ME(a)=";D$          220:NEXT I        >0.01":GOTO 36
80:C$=D$+STR$(N)    225:LF 1:LPRINT "N  5
85:INPUT #C$;M      ORMAL":GOSUB 2    340:IF W<U(1)
90:INPUT #C$;Z(*)   55
240:NEXT I          LPRINT "0.02>P
95:INPUT "FILE NA   230:FOR J=0TO N-1  >0.02":GOTO 36
ME(W)=";D$          235:X(I)=LN X(I)  5
100:C$=D$+STR$(N)  240:NEXT I        LPRINT "0.05>P
105:INPUT #C$;M     245:LF 1:LPRINT "I  5
110:INPUT #C$;U(*)  N-NORMAL":
115:LF 1:INPUT "EL  GOSUB 255
MENT=";A$           250:GOTO 115
120:LPRINT A$       255:FOR I=0TO INT  LPRINT "0.1>P>
125:INPUT "NO.OF S  (N/2-1)          0.05":GOTO 365
AMPLE=";S           260:Y(I)=X(N-1-I)-  LPRINT "0.5>P>
130:IF S=NGOTO 140 X(I)
135:END             265:NEXT I        0.1":GOTO 365
305:RETURN          270:B=0
5:REM PROG.SAVE     275:FOR J=0TO INT  360:IF W>U(4)
I
10:" ":CLEAR ;      (N/2-1)          LPRINT "P>0.5"
WAIT 0:INPUT "      280:B=B+Y(I)*Z(I)  365:RETURN
FILE NAME=";D$
30:INPUT "MAX.DIM   285:NEXT I
ENSION=";M:DIM
A(M-1)
40:CLS :INPUT "SA   290:T=0:S=0
UING NUMBER=";
N:GOTO 50
45:END
50:J=0
60:CLS :B$="N"+
STR$(N)+",P="
+STR$(J+1)+",
DATA="
70:CLS :PRINT B$;
80:INPUT A(J):
GOTO 100
90:GOTO 110
100:J=J+1:IF J<M
GOTO 60
110:C$=D$+STR$(N)
120:PRINT #C$;M
130:PRINT #C$;A(*)
140:GOTO 40

```

A\$, 工作单元,
I, J, 循环参数。

程序 SAVE 2

A, 场, 存数表, M个向量,
M, 数表的最大向量数;
N, 样品数;
D\$, 数表名称;
C\$, 数表名称加样品数;

B\$, J, 工作单元。

参考文献

- [1] 高玉堂: 环境监测常用统计方法, 原子能出版社, 北京, 1980。
- [2] 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(2): 66-69, 1984。