

问题讨论

关于土壤元素环境背景值数据统计处理中的一些问题¹⁾

邓 勃 秦 建 侯

(清华大学化学系)

随着环境科学研究和环境保护工作的发展,环境背景值的研究在我国日益受到人们的重视。环境背景值研究数据量很大,要求统计的项目也多,需要用数理统计方法和计算机来处理所得到的大量数据。本文拟就在土壤中元素环境背景值数据统计处理中遇到的一些大家所关心的统计问题,进行初步的探讨。

一、元素含量分布类型检验

土壤中元素含量波动范围较宽,其分布有些遵从正态分布,有些遵从对数正态分布或其他各种形式的偏态分布。对于分布类型的检验,已有多种方法^[1],各有特点,检验效率各不相同,对一组数据各种方法检验的结论有时也不相同。针对具体问题,使用何种检验方法为好,也无公认一致的看法。鉴于这种情况,通常是同时采用多种检验方法,根据各

种方法的检验结论采取“少数服从多数”的方法最后确定分布类型,并依此计算元素环境背景值。很显然,同时使用多种检验方法,势必使计算工作量大增。

诚然,在目前的情况下,要确定统一的检验方法尚无可能,但是否在任何场合都要同时使用多种检验方法而增加不必要的计算工作量呢?根据对于北京地区农业土壤 12 种元素环境背景值数据分布类型检验结果的统计资料(表 1)来看,我们认为,在许多情况下未必都要同时使用多种检验方法。

由表 1 中的数据可以看到,当样本容量 $n \leq 10$,表中所列各种检验方法的检验效率都不高,夏皮罗-威尔克法对各种土壤样本中不同含量范围的元素分布检验结论都是偏态分布,与此相反,偏度-峰度法、柯尔莫哥洛夫法的检验结论又几乎都是正态分布和对数正态分布。这种结论是不符合不同含量范

表 1 分布类型检验结果统计*

检验时 出现次数 样本数	检验方法				夏皮罗-威尔克法				皮尔逊 χ^2 法				偏度-峰度法				柯尔莫哥洛夫法			
	数据状态				原始数据		转换为对数值		原始数据		转换为对数值		原始数据		转换为对数值		原始数据		转换为对数值	
	分布类型				正态		不是对数正态		正态		不是对数正态		正态		不是对数正态		正态		不是对数正态	
	型				正态	不是对数正态	对数正态	不是对数正态	正态	不是对数正态	对数正态	不是对数正态	正态	不是对数正态	对数正态	不是对数正态	正态	不是对数正态	对数正态	不是对数正态
$n \leq 10$					0	74	0	74	—	—	—	—	72	2	70	4	70	4	72	2
$10 < n \leq 20$					50	17	42	25	—	—	—	—	52	15	58	9	59	8	53	14
$20 < n \leq 50$					127	77	135	69	—	—	—	—	6	198	52	152	176	28	176	28
$50 < n \leq 100$					—	—	—	—	56	9	65	0	2	63	20	45	46	19	48	17
$n > 100$					—	—	—	—	58	34	82	10	0	92	0	92	35	57	50	42

* 检验是这样进行的,先对原始数据进行分布类型检验,再将同一组数据转换为对数值再进行分布检验。

1) 中国科学院科学基金资助的课题。

围元素分布多样性的实际情况的。也可从另一方面来理解,即当样本容量小于 10 时进行分布检验并无什么实际意义。表中数据还表明,夏皮罗-威尔克法对于容量 $20 \leq n < 50$ 的样本的分布检验是合适的。皮尔逊 χ^2 法适合用于大样本的分布检验。偏度-峰度法对容量 $n \leq 10$ 或 $n > 100$ 的样本都不适用,对容量 $10 < n < 100$ 的样本,检验效果也不理想。柯尔莫哥洛夫法对容量 $n > 10$ 的样本,检验效果都比较理想。

二、异常值的判断

对于大样本异常值判断,通常采用两倍或三倍标准差准则,对于小样本异常值判断,一般使用狄克松法、格鲁布斯法和 t 法^[1]。在土壤元素环境背景值研究中,过去通常的做法是,同时使用三种检验方法,根据各自的检验结论和“少数服从多数”的原则,最后判定是否为异常值。我们认为,用这样的程序来判断异常值未必是合适的。

从理论上考虑,用于 t 法检验中的标准差是在事先剔除了怀疑为异常值的数据之后计算出来的,数值变小,因此检验灵敏度提高,这样在有些情况下就可能将本来为正常值的数据判为异常值。格鲁布斯法与 t 法不同之处在于,标准差是由包括被检验值在内的全部数据计算出来的,数值较大,因此检验标准不如 t 法严格。狄克松法是极差型检验,检验不严格,在某些情况下保留异常值的可能性较大。因此,一般说来,在这三种检验方法中,格鲁布斯法严格程度适中,检验效果最好^[1,3]。

实际情况也正是这样。表 2 列出了北京地区农业土壤中各元素环境背景值 42 批小样本($n < 25$)数据 102 个疑为异常值的检验资料。表中 x_n 表示正常值, x_a 表示异常值。

由表 2 中数据看到, t 法判为异常值数目最多, G 法次之, D 法判为异常值的数目最少。而且,凡是 D 法判为异常值的, G 法与 t 法也都判为异常值;凡

是 t 法判为正常值的, D 法和 G 法均亦判为正常值。

理论分析与实验资料表明,若 D 法与 t 法判为异常值的, G 法也一定判为异常值,三种方法的检验结果是相同的;如果 D 法判为正常值而 t 法判为异常值的, G 法可能判为正常值也可能判为异常值。按过去习惯的做法是,当 G 法判为正常值,则最终判为正常值;当 G 法判为异常值,则最终判为异常值,可见最终的结论取决于 G 法的检验结论。因此,不管 D 法与 t 法检验结论是否一致,最终的结论都由 G 法检验结论决定。由此可以引出结论:在小样本的异常值检验中,只要使用格鲁布斯法进行检验就可以了,而不必同时使用三种检验方法,从而增加许多不必要的计算工作量。

三、变异系数的计算

关于变异系数的计算方法,对于正态分布,用(1)式计算

$$C.V. = \frac{\text{标准差}}{\text{平均值}} = \frac{S}{\bar{x}} \quad (1)$$

对于偏态分布,是先将偏态分布正态化^[4],而后按(1)式计算变异系数。对于对数正态分布,计算变异系数的方法颇不一致^[5,6,7],见诸于公开文献者有以下三种计算方法

$$C.V. = \frac{S_G}{G} \quad (2)$$

$$C.V. = \sqrt{\sigma_{\log}^2 e - 1} \quad (3)$$

$$C.V. = (\mu^2 \sigma_{\log}^2 e - 1)^{\frac{1}{2}} \quad (4)$$

式中 μ 、 $\sigma_{\log}$ 、 $\sigma_{\ln}$ 是对数值正态分布的平均值和标准差,在有限次测定中用它们的估计值 $\bar{x}_{\log}$ 、 $S_{\log}$ 、 $S_{\ln}$ 代替。

$$\bar{x}_{\log} = \frac{1}{n} \sum_{i=1}^n \log x_i = \log G,$$

$$S_{\log} = \sqrt{\frac{\sum_{i=1}^n (\log x_i - \bar{x}_{\log})^2}{n - 1}} = \log S_G,$$

表 2 三种方法检验异常值的效果比较

检验方法	狄克松法(D 法)				格鲁布斯法(G 法)				汤姆逊法(t 法)			
	正常值 x_n		异常值 x_a		正常值 x_n		异常值 x_a		正常值 x_n		异常值 x_a	
出现次数	79		23		70		32		10		92	
检验方法	G 法		t 法		D 法		t 法		D 法		G 法	
	x_n	x_a	x_n	x_a	x_n	x_a	x_n	x_a	x_n	x_a	x_n	x_a
出现次数	70	9	10	69	0	23	0	23	70	0	10	60

表 3 冲积性水稻土背景值统计资料 (ppm)

元 素	样本数	全距范围	分布类型	几何平均值	几何标准值	变异系数	95% 置信区间
Cd	152	0.060~0.436	对数正态	0.162	1.45	0.128	0.151~0.170
Zn	151	41.3~109.6	对数正态	79.7	1.23	0.172	77.0~82.4
Mn	153	120.8~650.8	对数正态	128.1	1.49	0.377	263.8~299.5

表 4 北京地区土壤中镍背景值的变异系数的计算结果

算术平均值	几何平均值	标准差	几何标准差	变异系数计算方法*			
				(1) 式	(2) 式	(3) 式	(4) 式
20.28	19.80	4.35	1.25	0.214	0.063	0.224	0.126
20.68	20.11	4.80	1.27	0.232	0.063	0.242	0.135
24.44	24.09	4.12	1.19	0.169	0.049	0.175	0.105
24.87	24.52	4.26	1.19	0.171	0.049	0.175	0.105
29.47	28.63	7.49	1.29	0.254	0.045	0.259	0.162
30.95	30.25	6.86	1.25	0.222	0.041	0.226	0.144
26.91	26.26	6.30	1.25	0.234	0.048	0.226	0.138
28.73	27.08	9.48	1.46	0.330	0.054	0.392	0.239
20.45	20.02	4.22	1.23	0.206	0.061	0.209	0.117
25.92	25.43	4.71	1.23	0.182	0.048	0.209	0.129
24.75	24.74	0.83	1.03	0.034	0.042	0.030	0.018
28.13	27.91	3.93	1.15	0.140	0.041	0.140	0.088
22.34	22.40	5.18	1.29	0.232	0.058	0.259	0.150
22.80	22.14	5.24	1.29	0.230	0.058	0.259	0.150
20.39	19.94	4.28	1.24	0.210	0.062	0.218	0.122
25.15	24.79	4.29	1.19	0.171	0.048	0.175	0.106
27.01	25.80	7.93	1.38	0.294	0.053	0.331	0.199
27.10	26.59	4.86	1.23	0.179	0.046	0.209	0.129
18.34	17.91	3.84	1.25	0.209	0.070	0.226	0.122
27.40	27.23	3.24	1.13	0.118	0.041	0.123	0.076
27.69	27.50	3.45	1.13	0.125	0.041	0.123	0.077
22.27	21.62	5.40	1.29	0.242	0.060	0.259	0.148
23.07	22.31	5.82	1.31	0.252	0.059	0.275	0.159
23.43	22.73	5.77	1.28	0.246	0.056	0.251	0.146

* 第一列是用算术平均值和标准差计算的,第二、三、四是用几何平均值与几何标准差计算的。

G 为几何平均值, S_G 为几何标准差。

从理论上考虑,变异系数是表示一组测定数据的相对离散性,随标准差的增大而增大,随平均值增大而减小。如果用(4)式计算对数正态分布数据的变异系数,所得到的变异系数则不能反映它的这一属性。表3中列出了几个元素的背景值资料,表中的数据表明,Cd与Mn的标准差值很接近,而Mn的平均值远大于Cd,按(4)计算的Mn变异系数反而大于Cd,对Zn的情况也类似。很显然,这一结论同变异系数的属性是不相符合的。(3)式中不包含平均值项,计算的变异系数与平均值无关,这也是不合理的。

从实际情况考虑,若一组数据既遵从正态分布又遵从对数正态分布,用不同方法分别计算其变异系数,只要对计算结果进行适当换算,不同方法计算得到的变异系数理应是接近的,得到的结论应该是一致的。我们统计了北京地区农业土壤的测定数据,其中有24组数据既遵从正态分布又遵从对数正态分布。表4中列出了这些数据资料,表中的变异系数是分别按(1)、(2)、(3)和(4)式计算出来的。

表4中的数据表明,用(2)和(4)式计算的变异系数与按(1)式计算的相差很大,而用(3)式计算的变异系数与按(1)式计算的很相近。由此可见,用(2)和(4)式计算遵从对数正态分布数据的变异系

表 5 平均值与标准差的换算结果 (ppm)

元素	土 层	样本数	分布类型	变 换 前		变 换 后		全距范围
				平均值	标准差	平均值	标准差	
Cd	表土	117	对数正态	0.117	1.38	0.123	0.02	0.050~0.213
	底土	114	偏态	0.125	0.03	0.125	0.03	0.050~0.213
Cr	表土	116	对数正态	51.80	4.19	52.59	9.22	36.00~79.20
	底土	115	正态	51.11	8.47	51.11	8.47	27.30~73.80
As	表土	121	对数正态	7.95	1.30	8.23	2.20	4.12~14.48
	底土	120	正态	8.27	2.24	8.27	2.24	2.40~16.30
Zn	表土	120	对数正态	47.31	1.34	49.38	14.8	25.32~86.50
	底土	117	偏态	52.88	14.2	52.88	14.2	24.95~90.00
Ni	表土	120	正态	20.45	4.20	20.45	4.20	9.70~31.20
	底土	119	对数正态	20.16	1.25	20.66	4.67	9.70~31.20

数是不正确的。用(3)式计算的变异系数虽然与按(1)式计算结果都相近似,而正如前面指出的,用(3)式计算的变异系数与平均值的大小是无关的,这是不符合变异系数的属性的。由上述理论分析与实验数据可以看到,用各式计算对数正态分布数据的变异系数均不尽合理。

四、不同分布下平均值和标准差的换算

在土壤环境背景值研究中,为了判断样点是否受到污染,需对表土层和底土的环境背景值进行成对 t 检验。在编制分级统计图时,分级确当与否,直接关系到图的科学性、直观性与它的使用价值,因此需用显著性检验法来检验分级的合理性。此种显著性检验分级法包括方差分析和多重比较。上述这些统计检验都是基于被检数据都遵从正态分布。事实上在有些情况下,被检验数据并不都遵从正态分布,例如北京地区农业土壤中镍背景值,潮土表土层遵从正态分布,底土的镍背景值遵从对数正态分布;而对于棕壤土,表土层镍背景值遵从正态分布,底土镍背景值遵从偏态分布。为了进行表层和底层土背景值差异性统计检验,就要求对非正态分布的平均值和标准差变换为相应于正态分布的平均值和标准差,否则检验就无法进行。但是现在在一些文献中,不进行这种必要的变换,就将不同分布的平均值、标准差直接用于成对 t 检验,这种做法是不合理的。

非正态分布平均值与标准差转换为相应于正态分布的平均值与标准差是这样来实现的:若为偏态分布,则需进行偏态分布正态化^[4];若为对数正态分布,则用(1)、(3)和(5)式换算。

$$\bar{x}_{1n} = \ln \bar{x} - \frac{1}{2} \sigma_{1n}^2 \quad (5)$$

式中 $\bar{x}$ 是算术平均值, $\bar{x}_{1n}$ 、 σ_{1n} 是将测定值转换为对数值 $\ln x$ 之后求得平均值和标准差。表5列出

了研究北京地区潮土中一些元素背景值时非正态分布平均值与标准差的换算结果。

表5中变换前的平均值与标准差系指几何平均值与几何标准差。由表5看到,表土与底土各相应元素的全距范围是相近的,甚至是相同的,表明数据的离散范围与程度是接近的,因此,其标准差理应是接近的。然而从表5看到,变换之前,表土与底土元素背景值的标准差表观值相差较大,经过变换之后,表土与底土元素背景值标准差更趋于接近了,这是合理的。从实际情况考虑,不难想象,表土与底土所处环境条件基本上相近,且表土没有受到污染,表土与底土的标准差接近是合乎客观实际的。上述事实说明,利用(1)、(3)与(5)式将对数正态分布平均值与标准差换算为相应于正态分布的平均值与标准差是合理的。

五、结论

(1) 对于分布类型检验,样本容量 $n \leq 10$,用已试验过的各种方法进行分布检验效果都不好。样本容量 $20 < n < 50$,以夏皮罗-威尔克法与柯尔莫哥洛夫法检验效果较好; $n \geq 50$,用 χ^2 法与柯尔莫哥洛夫法检验效果较好。偏度-峰度法对各种容量的样本检验效果都不理想。

(2) 对小样本的异常值检验,没有必要同时使用多种检验方法,只需用格鲁布斯法检验就可以了。

(3) 当数据遵从对数正态分布,用关系式(2)(3)(4)计算变异系数均不尽合适,而以变换后的平均值和标准差计算变异系数更为合理。

(4) 对遵从不同分布的数据进行成对 t 检验时,应将非正态分布的平均值和标准差变换为相应于正态分布的平均值和标准差,而后进行 t 检验。

本文采用的数据源自北京市农业土壤及主要粮

食作物中污染元素环境背景值研究协作组, 特此致谢。

参 考 文 献

- [1] 尹启后, 赖维平, 重庆环境保护, (4), 79(1982).
- [2] 邓勃, 数理统计方法在分析测试中的应用 48-63 页, 化学工业出版社, 1984 年.

- [3] 吴传义, 数理统计与管理, (2), 20(1984).
- [4] 杨国治, 环境中若干元素的自然背景值及其研究方法, 82 页, 科学出版社, 1982 年.
- [5] 吴开治, 环境科学与技术, (1), 51(1984).
- [6] 高玉堂, 环境监测常用统计方法, 33 页, 原子能出版社, 1980 年.
- [7] A. A 别乌斯等著, 朱颜明等译, 环境地球化学, 226-239 页, 科学出版社, 1982 年.

高层建筑环境问题探讨

韩 伟 强

(上海城市建设学院设计所)

当前高层建筑迅速发展, 在城市规划、建筑设计、结构设计等诸方面给人们带来了新的课题。大量新建高层给人们的工作、生活、娱乐带来了极大的影响。国内外从事城市规划、建筑设计的专家们对高层建筑的前景有着不同的看法。但是从世界范围来看, 城市人口不断增加, 科学技术迅速进步, 特别是在人口比较密集的城市中, 高层建筑在一定时期, 一定范围内还将迅速发展。然而, 大量建造高层建筑以后所带来的一系列环境问题是决不能忽视的。本文就高层建筑的环境问题及如何解决这个问题加以探讨。

一、高层建筑迅速发展的原因

在本世纪中, 高层建筑的出现如雨后春笋遍及世界各地, 其主要原因是二次世界大战以后, 房屋破坏严重, 战后经济恢复, 发展迅速, 城市人口激增, 房屋用地越来越紧张, 而高层建筑则占地面积较少, 在一定程度上又能较多地增加建筑面积, 扩大城市空地, 有利城市绿化。高层建筑还可使道路、管线设备相对集中, 节约市政设施费用, 此外, 建筑的高低结合可丰富城市面貌。现代社会生产力高度发展, 科学技术日新月异, 大型电子计算机又为高层建筑提供了科学的计算基础。在社会生活和工作方面, 现代生活节奏之快, 工作效率之高是以往任何时代都不能比拟的。这种快节奏, 高效率的生活和工作使得人们希望有集中的办公地点, 居住区与生活区尽可能靠近, 以便节省上、下班路途的往返时间。

二、当前高层建筑的发展动向及存在问题

在国内, 从七十年代起高层建筑在一些大城市

如上海、北京、广州等地先后迅速发展。但我国目前高层建筑的功能主要还是住宅、旅馆、办公楼等类型的建筑, 单体功能比较单一。如上海 1984 年完工的 28 层上海联谊大厦主要作为外商的办公场所。1983 年完工的广州白天鹅宾馆共 34 层, 主要是用来接待国内外游客, 属旅游宾馆。上海于 1979 年完工的漕溪北路 16 层住宅是居住建筑。当然也有少量其它类型的高层建筑, 如 1985 年完工的上海交通大学图书馆是我国第一所全开架的高层图书馆。在国外, 高层建筑发展较早, 目前, 高层建筑正在向综合性发展, 如 1986 年完工的 100 层美国芝加哥约翰汉考克中心大厦要在 1.4 公顷的土地上安排 260000 m² 的建筑面积, 其中六到十二层为可放 1200 辆车的汽车库, 六层以下及四十四到四十五层的高空门厅内设有商店、游泳池、棒球房等公共设施, 十三到四十一层为办公室, 四十六到九十二层是有七百个单元的公寓, 九十三到九十七层为餐厅、礼堂及电视台。在这里, 人们生活和工作问题都可在大楼内得到解决。1982 年完工的 64 层香港合和中心大厦, 总建筑面积为 77900 m², 总高度为 216m, 主体为圆形结构。七十年代建造的日本东京“阳光城”采用“城中之城”的设计方案。这是一个以某个高层建筑为主体的巨大建筑群, 它区别于单体的高层综合型建筑, 而是包括四组建筑: 阳光大厦 60 层; 舶来品商场 11 层; 文化会馆 12 层; 旅馆 36 层; 总建筑面积共达 587000 m², 满足了该地区人民的生活、工作、娱乐等各种需要。

尽管这样, 从目前来看, 国内外高层建筑的环境还存在着许多问题, 如久住高层建筑的人容易得心脏病、咽喉炎、眼睛也容易患炎症。经常在高层建筑