

基准地价评估中异常样点剔除方法研究

王令超¹, 杨 欣², 宋艳华¹, 樊 鹏¹

(1. 河南省科学院 地理研究所, 郑州 450052; 2. 平顶山学院, 河南 平顶山 467002)

摘要: 分析了均值-方差法剔除异常样点的过程, 指出均值-方差法存在剔除标准难以确定、无法体现级别基准地价的内在联系、破坏样点地价空间分布的连续性、影响宗地地价评估的科学性等缺陷, 提出了改进完善均值-方差法的思路和方法: 将样点地价与其所在位置的土地质量联系起来测算定级总分值理论地价, 并将理论地价作为均值; 计算样点地价的理论标准差, 用理论标准差代替方差。最后, 以理论均值和理论标准差为标准来判断样点地价是否异常, 从而完成异常样点剔除, 并应用实例对改进的均值-方差法进行验证。

关 键 词: 基准地价; 样点地价; 异常样点; 均值; 标准差

中图分类号: F301.3

文献标志码: A

文章编号: 1003-2363(2012)05-0103-04

0 引言

基准地价是指在城市规划区范围内, 对现状利用条件下不同级别或不同均质地域的土地, 按照商业、居住、工业等用途, 分别评估确定的某一估价期日上法定最高土地使用权区域平均价格^[1]。

从基准地价的定义可以看出, 基准地价是特定用途的土地使用权区域平均价格, 这就要求区域内样点地价应分布在合理区间, 为了满足这一要求, 必须对个体样点数据进行检验, 剔除样点地价中异常值。所谓异常值, 是指同一级别或同一均质区域内同一土地利用类型样点中, 由于某些特殊因素而造成地价明显高于或低于其他样点的地价数据, 此类样点不能反映其所在土地的实际地价水平。另外, 在外业数据调查过程中, 同类近似宗地的调查样点数目往往过多(如商业门面房出租样点)、不同区域交易发生的密度相差较大或外业采点的人为倾向, 也往往会造成地价样点价位以及价格在空间分布上不合理。为了使基准地价能够合理的反映一个区域的实际情况, 必须对这些异常样点进行检验、剔除。

基准地价评估中, 在测定样点总体分布类型后, 需要用合适的方法对数据进行检验并剔除异常值。《城镇土地估价规程》规定^[1]: “用t检验法和均值-方差法分别对样本总体为正态和非正态分布的进行异常剔除”、“相同用途的样点地价, 在确定方差检验精度后, 精度以外的数据作为异常数据剔除”。本研究仅对样本总体为非正态分布时, 均值-方差法剔除异常数据存在的缺陷和解决途径进行分析探索。

1 均值-方差法缺陷分析

均值-方差法直观易懂、操作简便, 是基准地价评估中样本总体为非正态分布时常用的删除异常样点的方法, 但是该方法存在一定缺陷, 对基准地价评估具有较大影响。

1.1 均值-方差法剔除异常样点过程

1.1.1 样点地价排序。将级别或均质区域内某一用地类型样点地价由大到小顺序排列: $P_1, P_2, P_3, \dots, P_n$ 。

1.1.2 计算样点地价的平均值 \bar{P} 。

$$\bar{P} = \sum_{i=1}^n P_i / n \text{。} \tag{1}$$

式中: n 为样点个数(下同)。

1.1.3 计算样点方差 S 。

$$S = \sum_{i=1}^n (\bar{P} - P_i)^2 / n \text{。} \tag{2}$$

1.1.4 确定剔除的标准。当 $P_i \in [\bar{P} - k \times S, \bar{P} + k \times S]$ 时, 认为 P_i 为正常样点, 否则, 则为异常值, 予以剔除^[2-3], 其中 k 为系数, 《城镇土地估价规程》没有规定 k 的范围, 在实际应用中由基准地价评估者根据情况确定, 大多数评估中取 $k = 2$ 。

1.1.5 归并整理。将剔除异常值后的样点地价进行归并整理, 作为测算基准地价的基础数据。

1.2 均值-方差法的缺陷

1.2.1 剔除标准难以确定。多数情况下方差是一个较大的数值, 有时甚至是平均值的数倍、数十倍, 当 $k = 1$ 时, 往往还会出现所有样点都在 $[\bar{P} - k \times S, \bar{P} + k \times S]$ 内的情况, 达不到剔除异常样点的目的。有时 k 的值通常在 0.1 以下甚至更小时才会出现少数样点不在 $[\bar{P} - k \times S, \bar{P} + k \times S]$ 内的情况。同时, 不同级别在剔除异常样点时所采用的 k 值也多不相同, 这些都造成 k 的取值规律难以把握, 加大了评估过程中人为因素的影响。

收稿日期: 2011-12-05; 修回日期: 2012-06-30
作者简介: 王令超(1962-), 男, 河南陕县人, 研究员, 学士, 主要从事土地评价、土地规划研究, (E-mail) WLCH7758@ yahoo.com.cn。

1.2.2 无法体现级别基准地价的内在联系。《城镇土地估价规程》规定的均值-方差法剔除异常样点是以级别为单元独立进行的^[1],剔除过程中没有考虑级别之间的变化趋势、对比关系和内在联系,在实践中往往会出现应用均值-方差法剔除异常样点后计算的级别基准地价无法满足“级别之间存在正级差,且高级别级差大于低级别级差”的基本原则。

1.2.3 破坏了样点地价空间分布的连续性。从剔除异常样点的过程可以看出,均值-方差法仅考虑样点数据数值的大小,忽略了其空间属性。一些样点尤其是商业门面房出租样点,一般情况下在空间上会呈现连续分布的特点,其地价也会表现出随土地质量变化而相应变化的规律。然而,即使样点地价都真实合理,一般情况下在级别边界附近,处于高级别一侧的样点地价大部分会低于本级别的 $\bar{P} - k \times S$ (正常样点值得下限),而处于低级别一侧的样点地价则大部分会高于本级别的 $\bar{P} + k \times S$ (正常样点值得上限),这样,应用均值-方差法便会将级别边界两侧的大部分样点剔除,从而在级别边界两侧形成一个样点的“真空地带”。在均值 \bar{P} 和方差 S 一定的情况下,“真空地带”的范围与 k 的大小成反比, k 越大,“真空地带”相对越小; k 越小,则“真空地带”相对越大。

1.2.4 影响宗地地价评估。应用均值-方差法剔除异常样点后,保留下来的不同级别样点在地价—总分值构成的二维坐标系内呈阶梯状分布,给容积率修正系数计算、商业路线价及路线价修正系数计算等带来不便,直接影响宗地地价评估。

2 均值-方差法的改进

数理统计中的样点变量一般具有独立性和随机性,但地价样点并非完全随机分布,它们在空间上存在着相关性。根据土地的区位理论,城镇土地因其相对位置不同,在空间上表现出不同的使用价值、不同的经济效益和不同的地租地价,级别越高的土地其地价相应越高^[4-6]。所以,在进行样点检验时,应该把样点地价和土地位置联系起来。把样点地价与土地位置联系起来本质上就是把样点地价与土地质量联系起来,而定级总分值是土地质量的量化表现,因而可以把样点地价与其所在位置的定级总分值联系起来。通过建立数学模型,计算出总分值的理论地价。《城镇土地估价规程》规定的应用样点地价和土地定级单元总分值数学模型评估级别基准地价^[1]就是通过计算定级单元或样点的定级总分值的理论地价来实现的,说明理论地价能够反映定级估价区域的地价分布规律、变化趋势和基本地价水平。因而,可依据样点定级总分值的理论地价进行异常样点的剔除。

用总分值理论地价进行异常样点剔除的基本思路是通过建立样点地价和其总分值的数学模型,测算出总分值的理论地价,将样点总分值对应的理论地价作为样

点地价的理论平均值;计算不同总分值样点地价的理论标准差,以理论平均值和理论标准差为标准来判断样点地价是否异常,从而完成异常样点剔除,该方法的均值和标准差与定级总分值呈有规律的曲线分布,故称该方法为理论曲线法。具体步骤:1)把样点与定级总分值图叠加,获取样点的定级总分值。2)建立样点地价与定级总分值的回归模型。3)根据回归模型的相关系数、模型的内涵以及外延的变化趋势确定最终模型并计算总分值为 x 时的理论地价 D_x 。4)计算总分值为 x 的样点地价的的标准差:

$$b_x = \sqrt{\frac{1}{m} \sum_{t=1}^m (P_{xt} - D_x)^2} \quad (3)$$

式中: b_x 为总分值为 x 的样点地价的的标准差; P_{xt} 为总分值为 x 的第 t 个样点的地价; D_x 为总分值为 x 时的理论地价; m 为总分值为 x 的样点个数。5)为了减少不同总分值之间样点个数差异对 b_x 的影响,建立 b_x 与 x 的回归模型,规范总分值与样点地价标准差的关系,并计算总分值为 x 的理论标准差 B_x 。6)确定正常样点范围。总分值为 x 的正常样点下限:

$$D_{x下} = D_x - a \times B_x \quad (4)$$

总分值为 x 的正常样点上限:

$$D_{x上} = D_x + a \times B_x \quad (5)$$

当 $P_{xt} \in [D_{x下}, D_{x上}]$ 时,认为 P_{xt} 为正常样点,否则,则为异常值,予以剔除,其中 a 为系数。

3 实证分析

在新乡市 2008 年商业基准地价测算中,共调查房屋出租样点 1 324 个,依据 SPSS 软件的相关规定^[7],采用 Shapiro-Wilk (W 检验)检验法对数据进行正态分布检验(表 1),若统计显著性 $S_{ig} > 0.05$,则表明数据服从正态分布,否则,为非正态分布。检验结果表明各级别地价样本总体为均为非正态分布。分别采用均值-方差法和理论曲线法 2 种方法进行剔除异常样点,并对 2 种方法的差异进行了比较分析。

表 1 级别样点总体分布检验表

Tab.1 Inspection of grade samples population distribution

检验量	级别						
	1	2	3	4	5	6	7
样点数	260	215	250	216	197	114	72
S_{ig}	0.006	0.041	0.018	0.009	0.001	0.003	0.002

3.1 均值-方差(标准差)法剔除异常样点

计算出原始样点地价的级别平均值、方差后,发现方差数值过大,难以应用均值-方差剔除异常样点,故采用均值-标准差($k = 1$)确定正常样点地价的范围(表 2)并据此剔除异常样点。

表 2 均值-标准差剔除异常样点统计表							
Tab.2 Statistical results of mean-standard deviation excluding abnormal samples							
统计量	级别						
	1	2	3	4	5	6	7
剔除异常样点前计算级别平均地价/(元·m ⁻²)	3 025.22	2 018.66	1 617.09	1 266.97	1 047.26	919.98	719.56
方差	236 991.43	117 749.98	57 682.85	61 089.14	38 003.73	12 258.03	4 427.38
标准差	486.82	343.15	240.17	247.16	194.95	110.72	66.54
剔除样点/个	87	42	67	55	27	16	9
剔除异常样点后计算级别平均地价/(元·m ⁻²)	3 018.08	2 021.04	1 605.63	1 248.05	1 045.68	923.21	725.89

3.2 理论曲线法剔除异常样点

$R^2 = 0.873\ 1$ 。(7)

通过对多种回归模型的比较分析,确定了理论地价回归模型和理论标准差回归模型。

式中: B_x 为对应于总分值 x 的理论标准差; x 为总分值。同样,从表 3 中可以看出理论标准差与总分值也存在正相关关系。

3.2.1 理论地价回归模型。

$$D_x = 0.001\ 1\ x^3 + 0.331\ 5\ x^2 + 0.081\ 9\ x + 673.6\ ,$$
$$R^2 = 0.861\ 9\ 。$$

(6)

式中: D_x 为对应于总分值 x 的理论地价; x 为总分值。常数项反映的是末级地地价的理论下限,模型中的常数项 673.6 元/㎡ 与新乡市商业用地末级地最低出让地价大体相当。为直观地观察理论地价与总分值的关系,等间距选择了部分总分值计算其理论地价(表 3),从中可以看出理论地价与总分值存在正相关关系。

3.2.2 理论标准差回归模型。

$$B_x = 0.000\ 6\ x^3 + 0.020\ 4\ x^2 + 2.102\ 3\ x + 64.384\ ,$$

3.2.3 正常样点范围及异常样点剔除。为便于与均值-方差(标准差)法进行比较, a 值取 1,则总分值为 x 的正常样点下限、上限分别为:

$$D_{x下} = D_x - a \times B_x =$$
$$0.000\ 5\ x^3 + 0.311\ 1\ x^2 - 2.020\ 4\ x + 609.216\ ;$$
$$D_{x上} = D_x + a \times B_x =$$
$$0.001\ 7\ x^3 + 0.351\ 9\ x^2 + 2.184\ 2\ x + 737.984\ 。$$

(8)

(9)

根据(8),(9)式和样点总分值计算样点正常值的上下限,并以此确定样点是否异常,从而完成异常样点剔除(表 4)。

表 3 理论地价与总分值关系表								
Tab.3 The relation between theoretical land price and total value								
量	总分值							
	80.00	70.00	60.00	50.00	40.00	30.00	20.00	10.00
理论地价/(元·m ⁻²)	3 364.95	2 680.98	2 109.51	1 643.95	1 277.68	1 004.11	816.64	708.67
理论标准差	670.33	517.31	393.56	295.50	219.52	162.01	119.39	88.05
正常样点下限/(元·m ⁻²)	2 694.62	2 163.68	1 715.95	1 348.45	1 058.16	842.09	697.25	620.62
正常样点上限/(元·m ⁻²)	4 035.28	3 198.29	2 503.08	1 939.44	1 497.19	1 166.12	936.03	796.72

表 4 理论曲线剔除异常样点统计表							
Tab.4 Statistical results of theoretical curve excluding abnormal samples							
项目	级别						
	1	2	3	4	5	6	7
调查样点/个	260	215	250	216	197	114	72
剔除样点/个	63	69	57	65	60	30	20
剔除异常样点后计算地价/(元·m ⁻²)	3 009.65	2 097.53	1 580.77	1 257.35	1 053.09	883.57	733.68

3.3 两种剔除方法对比分析

3.3.1 异常样点数量不同。均值-标准差法剔除异常样点 303 个,理论曲线法剔除异常样点 364 个,后者明显多于前者。

3.3.2 异常样点分布区域不一致。以级别为单元进行异常样点剔除时,一些样点属于正常值,若将样点和其质量属性联系起来,并将级别与级别也联系起来从整体考察样点,则这些样点可能就属于异常样点,均值-标准差剔除的异常样点 47.23% 分布在级别界限两侧,而理论曲线法剔除的异常样点仅有 19.11% 分布在级别界限两侧,表明理论曲线法剔除异常样点更倾向于整个定级估价范围。

3.3.3 对级别平均值的影响不同。均值-标准差法剔除异常样点后的级别平均值与原始样点计算的级别平均值比较接近,差别最大的四级地相差也不过 1.5%;理论曲线法剔除异常样点后的级别平均值与原始样点计算的级别平均值之间差异增大,差别最大的六级地可达 4.12%。表明均值-标准差法剔除异常样点对计算级别

平均值的作用较小, 相比而言, 理论曲线法剔除异常样点则对计算级别平均值有较大影响。实际上均值-标准差法剔除的异常样点, 一般情况下既有数值较大的样点, 也有数值较小的样点, 所以其对平均值影响较小, 只有在样点总数较少, 且异常样点的数值为极端值时, 才会对平均值的计算有较大影响; 对级别而言, 理论曲线法剔除的异常样点并非都对称分布在其平均值两侧, 故其对级别平均值的影响要大于均值-标准差法。

3.3.4 对级别之间的对比关系影响不同。均值-标准差法剔除异常样点是以级别为单元进行的, 无法兼顾级别之间的对比关系; 而理论曲线法剔除异常样点是从整体出发, 以地价与整个估价区域土地质量的关系为基础进行剔除, 若样点偏离地价与土地质量的变化规律或趋

势, 则样点就会被剔除, 从而使保留下来的样点能够反映地价与土地质量的关系, 如二级地剔除的异常样点大多位于其低值区, 而六级地剔除的异常样点大多位于其高值区。这样, 通过异常样点剔除使级别之间对比关系更加科学合理。

3.3.5 级差分布不同。两种方法剔除异常样点后的级别平均值的级差不同(表 5), 在均值-标准差法剔除异常样点后的平均值级差中 6~7 级级差大于 5~6 级级差, 且 6~7 级级差与 4~5 级级差十分相近, 这种分布形态不能科学合理反映商业用地土地质量的变化规律; 理论曲线法剔除异常样点后的平均值级差体现出“级别之间存在正级差, 且高级别级差大于低级别级差”的基本规律, 且级差具有渐变过渡的特点。

表 5 级别基准地价级差对比表

Tab.5 The comparison between the standard land price differential and the land grade

基准地价级差	级别					
	1~2	2~3	3~4	4~5	5~6	6~7
均值-标准差法剔除异常样点后/(元·m ⁻²)	997.04	415.41	357.58	202.37	122.47	197.32
理论曲线法剔除异常样点后/(元·m ⁻²)	912.12	516.76	323.42	204.26	169.52	149.89

3.3.6 与最终评估结果吻合程度不同。2008 年新乡市基准地价更新调整中采用了房屋出租、房屋买卖、柜台出租、土地使用权出让、土地使用权转让等多种方法进行测算, 经专家论证、社会各界代表听证, 综合各种方法测算结果, 最终确定了商业用地级别基准地价调整结果

(表 6)。将两种测算结果分别减去最终结果, 再与最终结果相比可得到差异程度(表 6), 可以看出, 除四级地外, 理论曲线法测算的结果与最终结果的差异程度较小, 均值-标准差法测算的结果与最终结果的差异程度相对较大, 说明理论曲线法测算的结果更接近实际。

表 6 两种方法计算结果与最终结果的差异程度

Tab.6 The difference between the calculated results of two methods and the final result

结果与差异	级别						
	1	2	3	4	5	6	7
最终确定的基准地价/(元·m ⁻²)	3 000.00	2 100.00	1 580.00	1 250.00	1 050.00	880.00	730.00
均值-标准差法与最终结果的差异程度/%	0.60	-3.76	1.62	-0.16	-0.41	4.91	-0.56
理论曲线法与最终结果的差异程度/%	0.32	-0.12	0.05	0.59	0.29	0.41	0.50

4 结论

理论曲线法剔除异常样点保持了样点空间分布的连续性, 体现了级别基准地价各级别之间的内在联系, 使级别之间的对比关系更加科学合理, 使评估结果更加符合实际。理论地价模型、理论标准差模型的确定和建立不能只依据 R^2 的大小, 而应综合考虑 R^2 、模型的内涵以及外延的变化趋势来确定, 在确定理论地价模型时还要考虑常数项的大小和意义。标准差倍数 a 应根据样点的分布情况确定, 样点离散程度大 a 相应较大, 样点离散程度小则 a 也较小; 在具体剔除过程中, 可选择大、中、小多个数值进行比较, 确定合理的 a 值。模型的科学性和可靠性与样点数量密切相关, 故样点较少时不宜采用理论曲线法。

参考文献:

[1] 国家质量监督检验检疫总局. 城镇土地估价规程(GB/T18508-2001)[S]. 北京: 中国标准出版社, 2002.

[2] 王平, 唐旭, 祝国瑞. 基于回归模型与优化算法的样本检验方法探讨[J]. 昆明理工大学学报(理工版), 2004, 29(3): 17-20.

[3] 陈幼松, 杨位钦. 实用数理统计方法及应用详解[M]. 北京: 北京科学技术出版社, 1998: 83-134.

[4] 陈思源, 曲福田, 倪绍祥, 等. GIS 空间分析支持下的城市地价分布研究[J]. 南京农业大学学报, 2005, 28(3): 119-122.

[5] 张洪, 金杰. 中国省会城市地价空间变化实证研究[J]. 中国土地科学, 2007, 21(1): 24-30.

Key words: standard land price; sample land price; abnormal sample; mean; standard deviation