第 23 卷 第 2 期
2004 年 3 月

地 理 研 究
GEOGRAPHICAL RESEARCH

Vol. 23 , No. 2
Mar. , 2004

Zipf 维数和城市规模分布的分维值的关系探讨

谈明洪，范存会
(中国科学院地理科学与资源研究所，北京 100101)

摘要：城市位序-规模理论和分形理论是研究城市系统的重要基础。前者可以较好地刻画城市的规模分布，后者可用来深入地解释城市规模的分布规律。其中，城市规模分布的分维值和 Zipf 维数是这两个基础理论中的重要参数。在研究我国城市规模的分布规律时，理论上可认为分维值和 Zipf 维数的乘积等于 1。但本文认为这种理论上的关系并不能直接套用到统计分析中去，如果城市规模分布的分维值和 Zipf 维数是利用对于样本的 OLS（最小二乘法）估计所得，两者的乘积应等于判定系数（R^2）。最后我们对此结果进行了推导和证明，并对其所具有的理论意义和实践价值进行了简要阐述。

关 键 词：城市规模分布；Zipf 维数；分维
中图分类号：K928.5；O184 **文章编号**：1000-0585(2004)02-0243-06

中国拥有世界 1/5 的人口，是世界上最大的发展中国家。改革开放以来，中国经济持续发展，城市化进程加快。从宏观尺度上认识中国城市系统的发展规律不仅能丰富城市系统理论，对于合理规划未来中国城市发展也具有重大的实践意义。城市位序-规模法则和分形学是研究城市系统发展的重要理论，其中城市规模分布的分维值和 Zipf 维数是这两个基础理论中的重要参数，本文试图对这两个参数之间的关系作重新解释。

1 位序-规模法则及分形理论

Zipf 维数和城市规模分布的分维值分别来自齐夫(G. K. Zipf)公式和帕雷托(Pareto)公式，这两个公式是描述城市位序-规模的重要经验性公式[1]。我国是城市历史悠久的国家，位序-规模法则较好地刻画了城市人口的规模分布[2]。那么，什么是位序-规模法则呢？周一星是这样介绍的：研究对于一个城市的规模和该城市在国家所有城市按人口规模排序中的位序的关系所存在的规律[3]。城市的位序-规模可用下面的公式表示[4]：

$$P_1 \geq P_2 \geq P_3 \geq \cdots \geq P_r \geq \cdots \qquad (1)$$

如果我们将区域城市从大到小按人口规模排序，位序-规模法则常见的表达式为：

$$P_r = P_1 r^{-q} \qquad (2)$$

其相应的自然对数形式为：

$$\ln P_r = \ln P_1 - q \ln r \qquad (3)$$

收稿日期：2003-03-06；**修订日期**：2003-10-22
基金项目：中国科学院知识创新工程项目（KZCX2-310)；欧盟项目（ICA4-CT-2001-10085)
作者简介：谈明洪（1970-），男，博士生，江苏涟水人。主要从事土地利用/覆被变化研究。E-mail：tanmh@igsnrr.ac.cn

式(1)、(2)和(3)中,r 表示某城市在城市系统中的位序;P_r 表示位序为 r 的城市人口规模;P_1 在理论上为首位城市人口;q 为参数,即 Zipf 维数[5]。

对于位序-规模法则,城市地理学中还经常用帕雷托(Pareto)公式[1]来表示:

$$N(P) = AP^{-D} \qquad (D > 0) \tag{4}$$

其相应的自然对数形式为:

$$\ln N(P) = \ln A - D\ln P \tag{5}$$

式(4)和(5)中,$N(P)$ 表示大于门槛人口规模的城市数量;P 为城市人口规模;A、D 为系数,下文将对 D 进行具体解释。

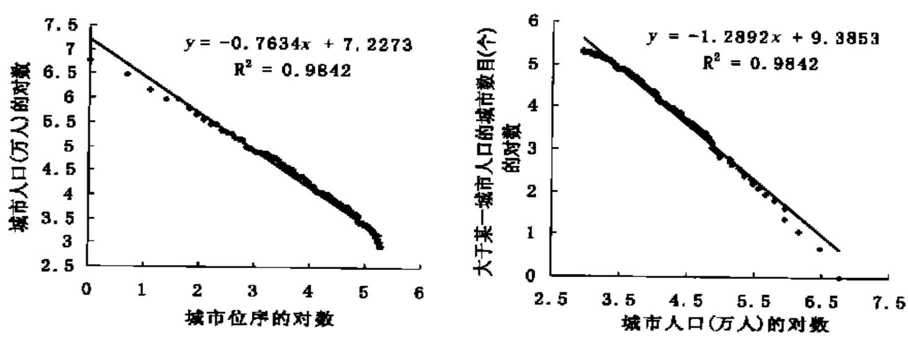
另外,陈彦光等还提出了位序-规模关系的严格表达形式,即三参数 Zipf 模型[5]。此模型的提出为预测城市系统的发展提供了重要工具。

位序-规模法则为描述城市系统的分布规律提供了可能,如要对城市规模分布进行深入解释,可借助于分形理论。分形(fractal)指的是自然界存在的具有自相似性、维数是分数的一类结构破碎的形体[6]。分形研究(fractal studies)是当代理论地理学的前沿[7];Mandelbot、D. Wong、A. S. Fotheringham 和 P. Frankhouser 等先后研究了城市体系的位序-规模法则以及 Pareto 分布与分维的关系[7];陈彦光岳文泽等对分形理论在城市体系中的应用做了深入的探讨,为研究城市系统的规模分布提供了理论基础[5,8]。将分形理论应用于城镇体系研究是分形理论在人文地理学应用中比较成熟的方面[7,8]。研究表明,城镇体系的等级结构存在无标度性,具有分形特征[8],这样,帕雷托公式的系数 D,即函数(4)、(5)中的 D,可被看作是城市规模分布的分维值,对此,已有相应的说明和推导过程[1,9]。

2 问题的提出

函数(5)和函数(3)中的 r 和 $N(P)$、P_r 和 P 是相应的两对变量。人们通常在计算城市规模分布的分维值时,为了研究的方便,理论上常认为函数(5)和函数(3)是互为反函数。所以,在研究我国城市规模分布时,把函数(3)中的 q 当作城市规模分布的分维值(D)的倒数[5,10,11],即:

$$D = 1/q \tag{6}$$



注:左图和右图中的直线均为趋势线;资料来源:《1998 年中国城市统计年鉴》

图 1 1997 年我国城市人口位序-规模双对数图

Fig1. A double-logarithmic plot for rank-size distribution of China urban population in 1997

但是在实际分析中,有时公式(6)并不成立。为了准确地分析城市规模分布的分维值,我们需要对上述的经验结果做重新解释。图 1 是 1997 年我国位于前 200 位、地级或地级

以上城市的人口-规模分布的双对数图。其中，左图是根据函数(3)来布置坐标点的，以城市位序的自然对数为横坐标，以城市人口规模的自然对数为纵坐标；右图是根据函数(5)来布置坐标点的，以城市人口规模的自然对数为横坐标，以超过某一人口规模的城市数目的自然对数为纵坐标。用幂函数曲线分别对这些坐标点进行拟合，幂函数的详细表达公式见图 1。结果显示：Zipf 维数［函数(3)中的 $q = 0.7634$］(图 1 中的左图)和城市规模分布的分维值［函数(5)中的 $D = 1.2892$］(图 1 中的右图)之间存在着这样的关系：

$$D \times q = 0.7634 \times 1.2892 \approx 0.9842 = R^2 \tag{7}$$

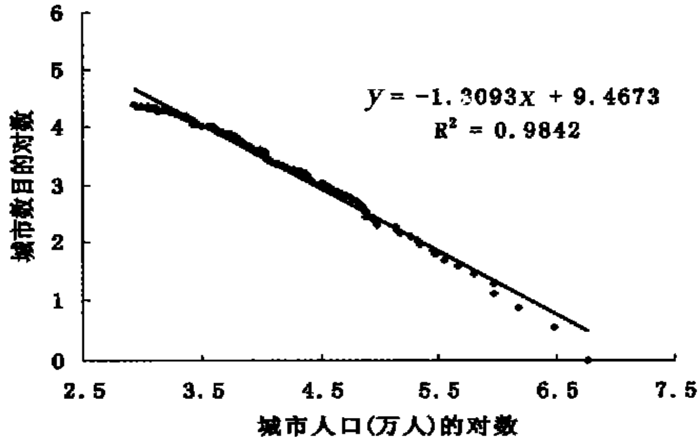
上式中 D 和 q 的乘积并不等于 1，如保留四位小数，正好等于判定系数(R^2)(图 1)。因此，我们可以得出这样的结论：在某些情况下，Zipf 维数和城市规模分布的分维值之间关系并不是倒数关系，相反，它们的乘积等于判定系数(R^2)。下面我们对公式(7)进行推导和说明。

3　Zipf 维数和城市规模分布的分维值关系推导

如果函数(3)和函数(5)互为反函数，那么系数之间应该具备以下关系：

$$1\mathrm{n}A = \frac{\ln P_1}{q} \text{和} D = 1 / q \tag{8}$$

以上是严密的函数-反函数转换过程。但是，函数(3)和函数(5)是利用统计上的OLS(最小二乘估计)方法估计出来的，那么 A，P_1和 D，q 的估计值之间就不具备以上关系，如公式(7)中 Zipf 维数和城市规模分布的分维值相乘不等于 1。



注：图中直线是在固定常数项，使 $\ln A = (\ln P_1) / q = 9.4673$ 的情景下坐标点的趋势线。

资料来源：《1998 年中国城市统计年鉴》

图 2　1997 年我国城市人口规模-频率双对数图

Fig. 2　A double-logarithmic plot for frequency-scale of China urban population in 1997

但是如果我们对函数(5)的估计值进行调整(图 2)，将其常数项固定为定义值：

$$1\mathrm{n}A = \frac{1\mathrm{n} P_1}{q} = 9.4673 \tag{9}$$

则估计结果如下：

$$D \times q = 0.7634 \times 1.3093 \approx 0.9995 \tag{10}$$

式中，D 和 q 值分别来自图 1 中的左图和图 2。如忽略误差因素，显然，D 和 q 的估计值互为倒数关系。出现上述结果的原因就在于利用 OLS 估计系数时，如果考虑残差项对因变量的影响，得出的回归结果必须满足残差平方和最小。

实际上，作为任意两个变量 OLS 估计系数的 D 和 q，都满足 $D \times q = R^2$ 的关系。证明过程如下：

令 1) $$X = c + DY + \varepsilon_2 \tag{11}$$

2) $$Y = a + qX + \varepsilon_1 \tag{12}$$

式中，变量 X 相当于函数(3)和(5)中的 lnr 和 $\ln N(P)$，变量 Y 则相当于函数(3)和(5)中的 $\ln P_r$ 和 $\ln P$，q 和 D 相当于函数(3)和(5)中的 q 和 D。a，q，c，D 分别是相应方程的 OLS (最小二乘估计)估计值，则满足：

$$R_1^2 = D^2 \times S_{YY} / S_{XX} \tag{13}$$

$$R_2^2 = q^2 \times S^{XX} / S_{YY} \tag{14}$$

其中，$$S_{XX} = \sum_i (X_i - \overline{X})^2; \qquad S_{YY} = \sum_i (Y_i - \overline{Y})^2 \tag{15}$$

式(13)、(14)和(15)的详细过程请见参考文献[12]。由(14)得：

$$R_2^2 = q^2 \times S^{XX} / S_{YY} = (q^2 \times D^2) / R_1^2 \tag{16}$$

于是：$$R_2^2 \times R_1^2 = D^2 \times q^2 \tag{17}$$

由于 X 和 Y 是相同的变量，判定系数(R^2)必然相等，有：

$$R^4 = R_2^2 \times R_1^2 = D^2 \times q^2 \tag{18}$$

然后有：$$D \times q = R^2 \tag{19}$$

即，公式 (3) 和公式 (5) 中的 q、D 相乘等于判定系数 R^2。

4 研究意义

研究 Zipf 维数和城市规模分布的分维值具有重要的实践意义，有利于正确把握城市规模（人口或用地规模）分布的规律和预测城市的发展规模[13～15]。本文对 1990～1996 年的中国城市建成区面积（包括县级市）位于全国前 200 位的城市进行统计分析，并对这些坐标点按照城市建城区用地面积的大小进行排列，建立回归曲线（图 3）。由于数据太多，在图上不易表达，只选用 1990、1993 和 1996 年资料绘图。可以看出中国城市建成区的土地利用的位序-规模曲线具有幂函数曲线的特征。用幂函数对 1990、1993、1994、1996 这四年数据进行拟合，判定系数(R^2)都在 0.97 以上(表 1)，最高的是 1994 年，判定系数值达到 0.9836。这可以证明：中国城市建城区用地规模分布符合位序-规模法则。当探讨 1990～1996 年城市建城区用地规模分布的特点时，我们如果用公式(7)来计算分维值，就会得出这样的规律：在此期间，中国城市建城区用地规模分布的分维值有逐渐增大的趋势(表 1)，分维值

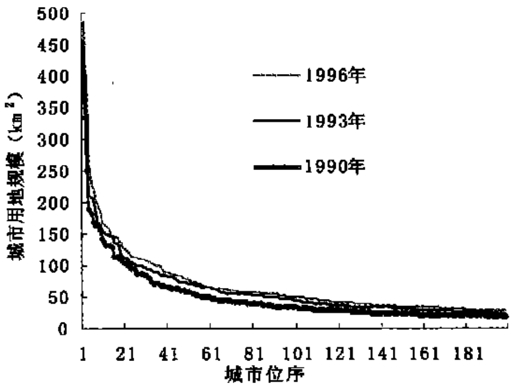


图 3 1990、1993 和 1996 年中国城市建成区用地位序-规模图

Fig. 3 Rank-size distribution of urban built-up area in China in 1990, 1993, and 1996

从 1990 的 1.4635 增加到 1996 年的 1.6376 (表 1) 。相反，如果用公式(6) 来计算城市建城区用地规模分布的分维值，分维值从 1990 年的 1.4966 增加到 1994 年的 1.6906，但 1996 年又变小了，分维值降为 1.6689。这样我们就得不出我国城市建城区用地规模分布的分维值有逐渐增大的趋势的，这样的规律就被遮蔽了。

弄清 Zipf 维数和城市规模分布的分维值之间的关系也具有理论意义。它不仅有利于我们深刻理解齐夫公式和帕雷托公式，对计算城市规模分布的分维值和预测城市的发展也有一定意义。可以防止出现不必要的误差，使分维值偏大。特别是当判定系数比较小的时候，这种偏大就会表现得更明显 (表 1) 。

表 1　1990～1996 年中国城市建成区用地位序-规模分析表

Tab.1　Rank-size analysis results of urban built-up areas for 200 cities between 1990 and 1996 in China

年份	位序-规模表达式 ($S = S_0 R_i^q$)	判定系数 (R_2)	Zipf 维数	公式 6 计算的建城区用地规模分布的分维值 D	公式 7 计算的建城区用地规模分布的分维值 D
1990	$Y = 701.68 X^{-0.6682}$	0.9779	0.6682	1.4966	1.4635
1993	$Y = 700.45 X^{-0.6092}$	0.9739	0.6092	1.6415	1.5988
1994	$Y = 665.18 X^{-0.5915}$	0.9836	0.5915	1.6906	1.6330
1996	$Y = 751.38 X^{-0.5992}$	0.9813	0.5992	1.6689	1.6376

注：此表对 1990～1996 年中的偶数年、城市建成区面积位于全国前 200 位的城市建区面积（包括一些面积较大的县级市）进行了统计分析。《中国城市统计年鉴》从 1997 年起，未收集县级市的建成区资料，因此，1997 年以后的数据未作统计。1993～1994《中国城市统计年鉴》中缺少 1992 年城市用地面积统计资料，故未对 1992 年进行分析。

5　结束语

本文分析了 Zipf 维数和城市规模分布的分维值的关系，认为如果城市规模分布的分维值和 Zipf 维数是利用对于样本的最小二乘法估计所得，两者的乘积应等于判定系数 (R^2) 。这个结论不但有助于理解齐夫公式和帕雷托公式，对预测未来城市规模也有实践意义。

致谢：本文得到了李秀彬研究员、吕昌河研究员的多次指导和帮助，深表谢意！

参考文献：

[1] 仵宗卿,戴学珍,杨吾扬. 帕雷托公式重构及其与城市体系演化. 人文地理,2000,**15**(1):15～19.
[2] David T Herbert,Colin J Thomas. Urban Geography. John Wiley & Sons Ltd.,1982. 91～96.
[3] Carlos M Urzua. A simple and efficient test for Zipf 's law. Economics Letters,2000,**66**:257～260.
[4] 周一星. 城市地理学. 北京商务印书馆,1995. 255～275,287～294.
[5] 陈彦光,刘继生. 城市系统的异速生长关系与位序-规模法则—对 Steindl 模型的修正与发展. 地理科学,2001,**21**(5):412～416.
[6] 管驰明,陈干,贾玉连. 乡村聚落群结构分形性特征研究. 地理学与国土研究. 2001,**17**(2):57～62.
[7] 刘继生,陈彦光. 城市地理学的分形研究的回顾与前瞻. 地理科学,2000,**20**(2):166～171.
[8] 岳文泽,徐建华等. 分形理论在人文地理学中的应用研究. 地理学与国土研究,2001,**17**(2):51～56.
[9] 刘继生,陈彦光. Davis 规律与 Beckmann 模型的数理等价性—城市体系等级结构的宏观—微观对称性分析. 经济地理,2001,**21**(3):231～234.

[10] 陈勇,陈嵘,艾南山,李后强. 城市规模的分形研究. 经济地理,1993,**13**(3):48～53.
[11] 陈彦光,罗静. 豫南区域经济支持系统的空间结构特征. 信阳师范学院学报(自然科学版),2001,**14**(2):204～208.
[12] William H Greene. Econometric Analysis. Prentice Hall, Inc,2000: 238.
[13] Zanntte H D, Manrubia S C. Role of intermittency in urban development: a model of large-scale city formation. Physical Review Letters. 1997,**79**(3):523～526.
[14] Marsili M, Zhang Y C. Inteacting individuals leading to Zipf's law. Physical Review Letters. 1998,**80**(12):2741～2744.
[15] Schweitzer F, Steinbrink J. Estimation of megacity growth, simple rules versus complex phenomena. Applied Geography 1998, **18**(1): 69～81.

Relationship between Zipf dimension and fractal dimension of city-size distribution

TAN Ming-hong, FAN Cun-hui
(Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing, 100101, China)

Abstract: China has a long history of urban development and has numerous cities. Macroscopically the congnition of development law of China's urban system provides an ideal source for enrichment of urban system theoretical research. Furthermore, China, as a developing country on the fast track, is experiencing rapid urban growth. A better understanding of China's urban system development will be helpful for predicting the city scale (such as, urban land scale, urban population scale) and macro-planning of China's urban growth.

The urban rank-size and the fractal theories are the important bases for studying the urban system. The former can perfectly depict the distribution of urban size, and the latter can be employed to explain the characteristics of distribution of urban size. At the same time, the fractal dimension and the Zipf dimension are the basic parameters of the two theories. But when studying the urban rank-size rule and the fractal, some scholars theoretically believe there is the relationship between Zipf dimension and the fractal dimension of distribution of urban size, that is, their product equals 1. But, we think, if D and q are the results of OLS (Ordinary Least Square) estimation, then their product should be R^2 (R^2, the coefficient of determination). Then, this paper deduces and proves this result.

To study the relationship between Zipf dimension and the fractal dimension of distribution of urban size is helpful to understanding the rule of urban rank-size. Moreover, this study is also necessary for grasping urban development in the future characteristics of urban system evolution in the past and predicting in the urban development future because China is experiencing rapid urbanization process.

Key words: city-size distribution; Zipf dimension; fractal dimension