Zipf's Rank-Size Theory Applied to Some Cities in Indiana and the Ohio River Watershed

Thomas Frank Barton, Indiana University

According to Mayer and Kohn in **Readings in Urban Geography** (the most widely used Urban Geography text in the United States), G. K. Zipf's rank-size theory may be stated as follows:

". . . that the towns of a country are arranged according to size, the Nth town having 1/Nth the population of the first."(1)

Since the writer believes in presenting concepts, theories, principles and other major ideas in the framework of the state and large watershed where an Urban Geography class is held, in this article Zipf's rank-size theory will be applied to: (1) 23 cities in Indiana with populations of over 20,000, (2) eight standard metropolitan statistical areas located in or partially in Indiana, (3) 24 cities in the Ohio River watershed with over 50,000 and (4) 15 cities in the Ohio River valley with populations over 30,000.²

The number of cities in each geographic area to which the theory was applied was selected somewhat arbitrarily. The writer believes this selection to be as scientific if not more so than the attempt to apply the theory. Some discretion was used. Since the number of standard metropolitan statistical areas is small the theory was applied to all eight. Indiana cities with populations of less than 20,000 and Ohio River valley cities with less than 30,000 were not used because fewer cities below these sizes have active, progressive city planning commissions which might use Zipf's rank-size theory as a technique if it were to be proven useful. The application of the theory in the Ohio River watershed was cut off when cities of less than 50,000 were reached because it was obvious that for practical purposes there wasn't sufficient correlation to continue.

Twenty-three Cities in Indiana

According to the 1960 United States census, there are 23 cities in Indiana with populations of 20,000 or more. These statistics give the population of the political cities and do not include the people living in the geographic or agglomerated city outside the precisely drawn legal boundaries.

In Table 1 data about these 23 cities are presented in six columns. In the first the 23 are listed in descending order of size, with their 1960 populations in the second column. The fractions of Zipf's rank-size theory are placed in the third.

In the fourth appears the number of people who should be living in the city if Zipf's theoretical sizes existed in reality. For example, Gary, as the second largest city in Indiana, according to Zipf's theory should have a population of ½ of that of Indianapolis, or 238,129. However, in 1960, Gary had only 178,312 inhabitants, or 59,817 less. These variations from the theoretical size are listed in the fifth column. The — sign indicates fewer people and the + sign more.

TABLE 1

Twenty-three Cities in Indiana

City	1960 Population	Zipf's Rank Size Fractions	Zipf's Theoretical Size	Variation in Population	Percent Variation
1. Indianapolis	476,258				
2. Gary	178,312	1/2	238,129	— 59,817	—25.1
3. Fort Wayne	161,776	1/3	158,753	+ 3,023	+01.9
4. Evansville	141,531	1/4	119,065	+ 22,466	+18.8
5. South Bend	132,445	1/5	95,252	+ 37,193	+39.0
6. Hammond	111,698	1/6	79,376	+ 32,322	+40.7
7. Terre Haute	72,500	1/7	68,037	+ 4,463	+06.6
8. Muncie	68,603	1/8	59,532	+ 9,071	+15.2
9. East Chicago	57,669	1/9	52,918	+ 4,751	+08.9
10. Anderson	49,061	1/10	47,626	+ 1,435	+03.0
11. Kokomo	47,197	1/11	43,296	+ 3,901	+09.0
12. Richmond	44,149	1/12	39,688	+ 4,461	+11.2
13. Lafayette	42,330	1/13	36,635	+ 5,695	+15.5
14. Elkhart	40,274	1/14	34,018	+ 6,256	+18.4
15. Marion	37,854	1/15	31,751	+ 6,103	+19.2
16. New Albany	37,812	1/16	29,766	+ 8,046	+27.0
17. Michigan City	36,653	1/17	28,015	+ 8,638	+30.8
18. Mishawaka	33,361	1/18	26,459	+ 6,902	+26.1
19. Bloomington	31,358	1/19	25,066	+ 6,292	+25.1
20. Laporte	24,157	1/20	23,813	— 2,654	—11.1
21. Logansport	21,106	1/21	22,679	— 1,573	—06.9
22. Columbus	20,778	1/22	21,648	— 970	—04.0
23. New Castle	20,349	1/23	20,707	— 358	—01.7

In the sixth column the variation of the actual number from the theoretical is indicated in percentages.

No higher mathematics than arithmetic is used in this presentation because the writer does not wish to limit the number of potential readers.

A brief study of the statistics in Table 1 reveals the following generalizations about Indiana's 23 largest cities:

1. The range of variation is from a minus 25.1 per cent (in the case of Gary, the second largest city) to a plus 40.7 percent (in the case of Hammond, the sixth largest) resulting in a total variation of 65.8 percent.

2. The 23rd largest city, New Castle, comes the nearest to having the same population as the theory would indicate. New Castle's actual size is only a minus 1.7 percent smaller. Fort Wayne, the third largest city, is only 1.9 percent larger.

3. In seven cities the percentage of variation is greater than 25 and in seven others it is less than ten. The latter seven cities rank in 3rd, 7th, 9th, 10th, 21st, 22nd and 23rd positions in descending order of size.

4. The 1960 population census of the 23 largest cities in Indiana do not support Zipf's rank-size theory. Consequently, city planners in Indiana could not rely on applying Zipf's theory as a practical technique.

5. The observation and generalization that Mark Jefferson made over a quarter of a century ago that many countries (political areas) have one large city which is two or three times numerically greater than the second is substantiated by Indianapolis' being in 1960 approximately three (2.67) times larger than Gary, the second largest.

Because of insufficient correlation between the actual population sizes of these cities and their theoretical sizes, a map was not made to show the distribution of Indiana's 23 largest political cities.

Standard Metropolitan Statistical Areas

The same method used in examining the 23 largest cities in Indiana was applied to the eight standard metropolitan statistical areas in or partially in the state. A brief study of the data in Table 2 reveals the following generalizations:

TABLE 2

Standard Metropolitan Statistical Areas in Indiana

City	1960 Population	Zipf's Rank Size Fractions	Zipf's Theoretical Size	Variation in Population	Percent Variation
1. Indianapolis	697,567				
2. Gary-Hammond-East Chicago	573,548	1/2	348,784	+224,764	+64.4
3. South Bend	238,614	1/3	232,522	+ 6,092	+ 2.6
4. Fort Wayne	232,196	1/4	174,392	+ 57,804	+33.1
5. *Evansville	165,794	1/5	139,513	+ 26,281	+18.8
6. *Louisville	114,192	1/6	116,261	— 2,069	— 1.8
7. Muncie	110,938	1/7	99,652	+ 11,286	+11.3
8. Terre Haute	108,458	1/8	87,196	+ 21,262	+24.4

* The statistics given in the second column after Evansville and Louisville refer to the population in the counties of Indiana which are a part of the Evansville and Louisville Standard Metropolitan Statistical Areas.

1. The range of variation is from a negative 1.8 percent to a positive 64.4 percent resulting in a total of 66.2 percent. This range is similar to the total range when the theory is applied to Indiana's 23 cities but the variation is more on the positive side.

2. The sixth largest standard metropolitan statistical area of Louisville is only 1.8 percent smaller than the theory indicates its size should be. And the standard metropolitan statistical area of South Bend is only 2.6 percent larger.

3. In sharp contrast with the small percent variation between the actual and the theoretical sizes of the Louisville and South Bend standard metropolitan statistical areas are those of (1) Gary-Hammond-East Chicago which is approximately two-thirds larger than it should be and (2) Fort Wayne, about a third larger.

4. The 1960 populations of these eight standard metropolitan statistical areas do not support Zipf's theory.

The Ohio River Watershed

Although Zipf's theory in Urban Geography has been applied to political units, the writer applied the theory to the 24 largest cities in the Ohio River watershed which had populations of over 50,000 in 1960. Most of Indiana's territory is in the Ohio River watershed.

Data in Table 3 substantiates the following generalizations:

1. Thirteen of these 24 cities have over 90 percent more people than they should have (column 5 of Table 3). Or, one may say that these 13 cities are almost twice as large as they should be.

TABLE 3

Cities in Ohio River Watershed over 50,000

	City	1960 Population	Zipf's Rank Size Fractions	Zipf's Theoretical Size	Variation in Population	Percent Variation
1.	Pittsburgh, Penn.	604,332				
2.	Cincinnati, Ohio	502,550	1/2	302,166	+200,384	+66.3
3.	Indianapolis, Ind.	476,258	1/3	201,444	+274,814	+73.3
4.	Columbus, Ohio	471,316	1/4	151,083	+320,233	+47.2
5.	Louisville, Ky.	390,639	1/5	120,866	+269,773	+44.8
6.	Dayton, Ohio	262,332	1/6	100,722	+161,610	+62.3
7.	Nashville, Tenn.	170,874	1/7	86,333	+ 84,541	+97.9
8.	Youngstown, Ohio	166,689	1/8	75,542	+ 91,147	+82.9
9.	Evansville, Ind.	141,543	1/9	67,148	+ 74,395	+90.2
10.	Chattanooga, Tenn.	130,009	1/10	60,433	+ 69,576	+86.9
11.	Canton, Ohio	113,631	1/11	54,939	+ 58,692	+93.6
12.	Knoxville, Tenn.	111,827	1/12	50,361	+ 61,466	+81.9
13.	Charlestown, W. Va.	85,796	1/13	46,487	+ 39,309	+84.5
14.	Huntington, W. Va.	83,627	1/14	43,167	+ 40,450	+93.7
15.	Springfield, Ohio	82,723	1/15	40,289	+ 42,434	+94.9
16.	Terre Haute, Ind.	72,500	1/16	37,771	+ 34,729	+91.9
17.	Hamilton, Ohio	72,354	1/17	35,549	+ 36,805	+96.6
18.	Muncie, Ind.	68,603	1/18	33,574	+ 35,029	+95.8
19.	Lexington, Ky.	62,810	1/19	31,807	+ 31,003	+97.5
20.	Covington, Ky.	60,376	1/20	30,217	+ 30,159	+99.8
21.	Warren, Ohio	59,648	1/21	28,778	+ 30,870	+93.2
22.	Kettering, Ohio	54,462	1/22	27,470	+ 26,992	+98.3
23.	Johnstown, Penn.	53,949	1/23	26,275	+ 27,674	+94.9
24.	Wheeling, W. Va.	53,400	1/24	25,181	+ 28,219	+89.2

2. Another five cities are over 80 percent larger.

3. Twenty-two of the 24 cities are over 60 percent larger.

Ohio River Valley

The statistics in Table 4 concerning the 15 cities in the Ohio valley with populations of over 30,000 substantiate the following generalizations:

1. The range is from a minus 40.7 percent to a plus 93.9 percent resulting in a total of 134.6.

2. Thirteen of the 15 cities are smaller than the theory would indicate. Nine of these are over 30 percent smaller.

TABLE 4

Cities Along the Ohio River

City	1960 Population	Zipf's Rank Size Fractions	Zipf's Theoretical Size	Variation in Population	Percent Variation
1. Pittsburgh, Penn.	604,332				
2. Cincinnati, Ohio	502,550	1/2	302,166	+ 200,384	+ 66.3
3. Louisville, Ky.	390,639	1/3	201,444	+ 189,195	+ 93.9
4. Evansville, Ind.	141,543	1/4	151,083	— 9,540	— 6.3
5. Huntington, W. Va.	83,627	1/5	120,866	— 37,239	—30.8
6. Covington, Ky.	60,376	1/6	100,722	— 40,346	—40.1
7. Wheeling, W. Va.	53,400	1/7	86,333	— 32,933	—38.1
8. Parkersburg, W. Va.	44,797	1/8	75,542	— 30,745	—40.7
9. Owensboro, Ky.	42,471	1/9	67,148	— 24,677	—36.8
10. New Albany, Ind.	37,812	1/10	60,433	— 22,621	—37.4
11. Paducah, Ky.	34,479	1/11	54,939	— 20,460	—37.2
12. Portsmouth, Ohio	33,637	1/12	50,361	— 16,724	—33.2
13. Steubenville, Ohio	32,495	1/13	46,487	— 13,992	—30.1
14. Ashland, Ky.	31,283	1/14	43,167	— 11,884	—27.5
15. Newport, Ky.	30,070	1/15	40,289	— 10,219	—25.4

3. The only close correlation between actuality and the theory is Evansville, the fourth largest city in the valley, which has a minus variation of 6.3 percent.

Falacies of Applying Zipf's Theory

There are at least two major falacies in applying Zipf's theory to any political or geographic area in the United States. First, the United States census gives only the population of the political city. And often, if not usually so, the geographic or "agglomerated"[3] city is much larger than the political one in both areal size and population. Many cities have not expanded their political boundaries regularly so as to include all the urban population. If all the people clustered in a settlement are not recorded, is it not foolish and scientifically unsound to attempt to compare the population of these settlements or cities and apply Zipf's theory or other theories?

Over 30 years ago before this theory received wide attention in America's Urban Geography field, many geographers in the United States abandoned the practice of attempting to compare the sizes of cities or to characterize the growth of an individual or a group of cities because different methods and rates of annexation resulted in statistics that did not warrant comparisons.

Second, not only is there no uniform method of computing all of the people living in a cluster settlement or geographic city, but because of the proliferation of political units around large cities (such as independent legally-established villages, towns and cities), the population in a cluster settlement is often fragmented and listed in the census under different names. This is especially true of the largest settlements. For example, the number of people living in the cluster settlement of

Indianapolis in 1960 was much larger than that of the political city for the same year. The surface area of some political cities is completely surrounded by political boundaries of other legal settlements.

In Conclusion

1. Zipf's theory is not substantiated by an examination of: (1) the 23 largest cities in Indiana, (2) the eight standard metropolitan statistical areas, (3) the 24 largest cities in the Ohio River watershed or (4) the 15 largest in the Ohio River valley.

2. However the generalizations based on the data in the four tables do not necessarily discredit Zipf's theory because is not his theory based on the assumption that all the people clustered in a single settlement are recorded?

3. Because of the method of collecting census statistics and the political fragmentation of single cluster settlements, Zipf's theory has a limited or very little value in functional planning in the large cities in Indiana and the Ohio River watershed.

4. By applying Zipf's theory to a group of cities and supplying the type of data presented in the tables of this article one provides a good teaching technique because it stimulates class discussion and the students speculate about the findings.

5. The writer believes that geographers and others who study cities and their problems should draft a list of criteria by which the cluster settlement (geographic or agglomerated city) could be delimited and its boundaries drawn before the 1970 census is taken.

Literature Cited

1. MAYER, HAROLD M. and KOHN, CLYDE F., **Readings in Urban Geography.** Chicago: the University Press, 1959, p. 229. Also refer to Zipf, George Kingsley, **Human Behavior and the Principle of Least Effort.** Addison-Wesley Press, Cambridge, Mass., 1949, p. 75.

2. U. S. Bureau of the Census, **U. S. Census of Population: 1960, Vol. 1, Characteristics of the Population.** U. S. Government Printing Office, Washington, D.C., 1963.

3. Agglomerated is a term used in the **United Nations' Demographic Yearbook,** 1962, to designate the population of the political city and adjacent areas.