THE FOUR BASIC PROPERTIES OF RANK-SIZE HIERARCHICAL DISTRIBUTIONS: THEIR CHARACTERISTICS AND INTERRELATIONSHIPS

Barclay G. Jones and Blane D. Lewis
Department of City and Regional Planning
106 West Sibley Hall
Cornell University, Ithaca, NY 14853

ABSTRACT In sets of empirical data that can be described by a skewed distribution function such as the rank size, relationships exist among the total observations, the relative size of the largest observation, the number of categories, and the slope of the distribution. We consider only one phenomenon, the distribution of urban population into urban centers of various sizes. The properties become the total urban population, the size of the largest center, the number of places, and the slope characterizing the distribution. We present an analytical model which defines these relationships for various sizes of urban populations and various slopes. Numerical examples demonstrate the results of the analysis. The advantages of using the Principal Axis to measure slope are discussed, and we demonstrate this method. The model shows clearly how the proportion of the urban population in the largest center declines and also how the number of smaller centers grows dramatically as total urban population increases. Implications are discussed for the nature of changes to be expected in those areas of the world where urbanization is still taking place, such as Asia, Latin America, and Africa.

1. INTRODUCTION
It is a characteristic of various phenomena, particularly in the biological and social sciences, that the distribution of their occurrences, sizes, and other attributes are highly skewed to the right (Aitchison and Brown 1957; Silberman 1967). Scholars have found a number of different distributions useful in the analysis of these phenomena such as the negative binomial, log-normal, Gamma, Yule, and similar functions. Since it was first fully developed and described by Lotka (1925), the rank-size distribution has proven to be one of the most useful of them because of its simplicity and ease of application. It is the purpose of this paper to analyze the four basic properties of this function and demonstrate their relationships. The only empirical distribution to which we will refer is that of cities ordered by size of population.

Since Auerbach (1913) first applied it to concentrations of human populations and Zipf (1949) drew attention to it in a general discussion of a wide variety of phenomena, the rank size distribution of cities has generated a significant amount of interest among scholars, and a very large number of works has accumulated. It is difficult to isolate those studies that deal exclusively with cities from the larger literature that treats city sizes along with other skewed phenomena or focuses on other phenomena entirely, and this brief review will not attempt to make that distinction. Theoretical work has focused on the search for possible explanations of how observed size distributions are generated. Stochastic process models (R. Vining 1954; Simon 1957; Ijiri and Simon 1964) have achieved much prominence, although hierarchy models (Beckmann 1958; Beckmann and McPherson 1970; Parr 1969) and growth models (Zipf 1949; Darwin 1953; Davis...
and Swanson 1972) have also had great influence. Richardson (1973) and Carroll (1982), among others, review the theoretical literature.

Empirical studies have demonstrated the prevalence of the phenomenon and variations in the characteristics of urban hierarchies in different regions and in different sized sets of urban centers. Many of these studies have stressed testing relationships between variations in rank size distributions and various indices of urbanization and economic development. Despite early pessimism by Berry (1961) concerning a possible relationship between city size distributions and economic development, a voluminous literature has accumulated. Extensive reviews have been provided by Farid (1978) and Carroll (1982).

A number of recent studies concentrate on the properties of rank size distributions as urban systems evolve over time. D. Vining (1977) has argued, for example, that the city size distribution of a growing urban system is determined by the ratio of the average rate of growth in the size of cities to the rate of growth in the number of cities. Malecki (1980) studied the effect of the variation in growth rates among cities of different sizes in the American Midwest between 1940 and 1970 and notes an increasing concavity in the distribution over time. As a final example, Parr (1985) suggested that the value of the rank size parameter exhibits an inverted U-shaped pattern when graphed as a function of time.

A closely related topic, which has generated a considerable literature in itself, concerns the size of the largest center in relation to the rest of the set of urban centers. Since Jefferson (1939) first identified the condition where the size of the largest city is inordinately large compared to other cities, primacy has been an issue that has claimed much attention and caused considerable concern. Jefferson measured primacy by the ratios of the populations of the second and third largest cities to the first. Many definitions and measures have emerged since. The more recent Davis index (Davis 1976), which uses the ratio of the population of the largest city to the sum of the populations of the four largest, has gained widespread use as a measure of primacy. These measures use only information about the very top of the urban hierarchy and none of the other data about the set of urban centers. Rosen and Resnick (1980) incorporate more information about the distribution by using as the denominator in their measure the sum of the population of the fifty largest cities. These measures are dependent on the slope of the portion of the distribution that is considered, and Sheppard (1982) notes that rank size relationships with different slopes should not necessarily have different levels of primacy. He offers an alternative based on El Shakhs (1972) which is independent of the slope, however, he too considers only the top end of the size distribution in calculating the index.

In this paper we will consider four general properties of rank size distributions: the number of observations (or occurrences) they contain; the relative size of the largest observation (or most frequent occurrence); the number of classifications or categories of observations (or occurrences); and the slope of the distribution. We will refer to one phenomenon of interest, city size distributions, as urban populations increase and explore some of the implications for parts of the world where urbanization continues to be a major phenomenon. The total urban population is the first topic. The changing relative size of the largest center as urban populations increase is treated next. Third, we consider the ways the number of places change as urban populations increase. Changing slopes of hierarchical distributions of urban centers are next considered as urban populations expand. Finally, we treat the condition of primacy and devise a measure that incorporates all of the available information about the set of urban centers under consideration.